Eutrophication in Jordan Lake: Using Specific Conductance During Storm Events to Characterize Nutrient Transport from Varying Landscapes

Will Hamilton

Approved:

X_____

Diego Riveros-Iregui, Faculty Advisor

Joseph Delesantro, Mentor & Reader

ABSTRACT

As the North Carolina Triangle region experiences rapid growth and development, water quality becomes an increasingly pertinent issue. Increased nutrient loading in the Jordan Lake drinking reservoir causes toxic cyanobacterial blooms that pose public health concerns. The Jordan Lake Nutrient Study seeks to determine how land use within this region affects the transport of nutrients and other contaminants. Over the course of two years, our team collected water level, streamflow, and specific conductance observations at five-minute intervals using automated sensors. Specific conductance indicates the concentration of dissolved ions present in water in μS/cm. Typically, this metric is high in ground water and discharge yet almost absent in rainfall, providing a measurement of contaminants traveling over the landscape that excludes contaminants added by rainfall. We examined changes in contaminant concentrations under changing flow conditions. Storm events are one such change because they cause a sharp increase in flow followed by a gradual recession. Specific conductance exhibits a delayed response to these changing flow conditions that we characterized as dilution (decreased specific conductance) or enrichment (increased specific conductance). We used a hidden Markov model to iteratively filter out storm behavior from intermittent flow conditions. This model, paired with a peak identification function in R, allowed us to distinguish single storm events in large flow datasets. For each storm, we performed hysteresis analyses by plotting flow versus specific conductance and assessing the response of specific conductance. Our preliminary results suggest distinct behaviors of specific conductance between sites. During storms, landscapes with more intense development and flashier (lower duration and higher change in flow) storms often showed dilution, while forested, less developed sites exhibited either enrichment or dilution. These findings support land use as a driving factor of contaminant transport and prompt further investigation into this relationship using a wider range of land uses.

INTRODUCTION

Eutrophication is the process by which excess nutrients in a body of water lead to prolific growth of algae. When the algae die, much of the dissolved oxygen is depleted during decomposition, leading to hypoxia and massive death of aquatic organisms (US EPA 1996a). In addition to ecological impacts, toxins released by cyanobacteria in these algal blooms may negatively impact human health (Codd, 2000); remediation steps taken by municipalities to reduce and counteract the negative consequences of eutrophication pose economic challenges to the impacted communities (Carpenter et al., 1998). Due to these ecological, public health, and economic concerns, eutrophication has been the focus of many studies since the 1960s (Vollenweider; 1970; Schindler, 1974; Carpenter et al., 1999; Smith, 2003; Søndergaard et al., 2003; Conley et al., 2009). National organizations in countries such as Australia and the United States as well as international organizations including the Organization for Economic Cooperation and Development and the World Health Organization recognize and address eutrophication as a crucial issue (OECD, 1982; US EPA 1996a; WHO, 1997; Smith et al., 1999; Codd, 2000). Such studies and organizational bodies seek to explain driving mechanisms behind eutrophication in an effort to devise more effective prevention and mitigation strategies.

The primary nutrients responsible for eutrophication are excess Nitrogen (N) and Phosphorus (P) (Smith et al., 1999). N and P are limiting factors of biological productivity in both aquatic and terrestrial ecosystems and are introduced into streams via industrial processes in the form of

wastewater, fertilizers from agriculture, and runoff from urban environments (Carpenter et al., 1998; Smith et al., 1999; Codd, 2000). While both point and nonpoint sources contribute to nutrient runoff, varying land use practices on the nonpoint scale are the primary contributor (Smith et al., 1999). Headwater streams are the origin for just under half of the Nitrogen in larger order streams in the United States (Royer et al., 2004; Alexander et al., 2007). These nutrients mobilize via discharge during precipitation events (Sharpley et al., 1994; Hessen et al., 1997; Paerl, 2006). Storms in particular cause a major flux in runoff that moves nutrients across the landscape and into streams to varying degrees depending on storm intensity and duration (US Army, 1977; Rimer et al., 1978; Jarvie et al., 2006). All of these studies, however, sampled streams ranging from an hourly to weekly rate. As most storms last only a few hours, the resolution at which they characterized stream behavior during storms was limited to just a few measurements per storm. While these studies did find that storm discharge was responsible for transport of heavy nutrient loads, they could not determine the mechanisms of that transport. Historically, discharge carrying heavy nutrient loads were attributed to agricultural land use practices (Sharpley et al., 1994; Hessen et al., 1997). While agriculture remains a major culprit in the transport of nutrients, other land use practices have since been linked to eutrophication. Human activities such as urban development and population growth have also been connected to the mobilization of nutrients into streams (Jarvie et al., 2006; Paerl, 2006; Mörth et al., 2007; Savage et al., 2010).

There have been many studies of land use and eutrophication of major rivers and coastal drainage areas (Ning et al., 2002; Paerl, 2006; Mörth et al., 2007; Savage et al., 2010). Many such studies attempt to model nutrient loading in watersheds to give better insight into effective mitigation strategies, yet a common limitation exists (Mörth et al., 2007)—how do nutrients travel from their points of origin? Due to the important role that headwaters play as a source of nutrients, understanding how nutrients move from headwaters over a variety of landscapes would allow us to better model nutrient loading from origin to sink. While various studies have worked to characterize nutrient flux over a single land use such as agricultural fields or urban areas (Lawler et al., 2006; Alexander et al., 2007), little research details the effects of different land use practices on nutrient transport in headwater streams.

The Piedmont region of North Carolina offers a unique landscape to study eutrophication because the region sits at the headwaters of multiple drainage basins leading to the Atlantic Ocean including the Catawba, Broad, and Yadkin Pee-Dee River Basins. The Piedmont is actively undergoing rapid population growth and development, which takes place adjacent to still undeveloped, forested land. Chapel Hill, North Carolina is one such city in the Piedmont that possesses a wide gradient of land uses within and around the growing city's limits. This includes urban and residential land use within the city, with ample forested lands just outside the town. Chapel Hill contains headwater streams that make up the Jordan Lake watershed. The North Carolina government continues to combat water quality threats from eutrophication at Jordan Lake, and an effective mitigation strategy has yet to be achieved (NC DEQ, 2009).

The UNC Nutrient Management Study is a multi-year, cross-disciplinary study that aims to determine effective strategies for combating eutrophication in Jordan Lake. The primary goals of this study are to characterize nutrient transport across varying land uses in the Jordan Lake watershed and to gain a better understanding of nutrient transport from nonpoint sources across the landscape. These goals not only involve assessing movement of nutrients across a variety of landscapes, but also during a variety of flow conditions. For data collection, the study uses in-

situ sensors that measure water-level, streamflow, and specific conductance in 5-minute intervals. Data collection began in the summer of 2017 and will continue through the summer of 2019. With 10 study sites spread across Chapel Hill and Durham, the dataset has quickly grown to include roughly 120,000 points per site in one year.

In this thesis, I focused on storm events: when flow conditions changed and nutrients were more likely to mobilize. Elevated flow caused by storm events has been linked to the transportation of nutrients throughout the watershed (Jarvie et al., 2006; Butturini et al., 2008; Vaughan et al., 2017). Most studies analyzing stream behavior during storm events do so with fewer events and/or fewer sites (Kobayashi et al., 1999; Lawler et al., 2006; Vaughan et al., 2017). Because the Jordan Lake Nutrient Management Study includes two years of high-resolution data at ten sites, I developed an automated method to extract hundreds of storms from hundreds of thousands of points worth of data. This storm isolation method was designed for general use that may be applied to large flow datasets beyond this study. Once I isolated storms, I characterized the relationship between flow and specific conductance during storm events across landscapes with varying degrees of urban development. Through this study, I aim to develop methodology that improves our use of high temporal resolution data, and improve understanding of how urban landscapes alter transport and nutrient loading.

METHODS

Site Description

Land use was characterized for headwater watersheds in the Jordan Lake Watershed. Four study watersheds where selected for this study, spanning a gradient in land uses, which were assessed using the 2011 National Land Cover Database (Figure 1 & Table 1). Each sampling site is located at the pour (drainage) point of the watershed. The Tallyhoe watershed was mostly forested, with low road density and low development. Rogers Road was moderately developed with higher road density, and was the only site where the majority of residents used septic systems for waste management. The Burlage watershed was more developed with a moderate to high road density and was connected to sanitary sewer pipelines. Of the four sites, Booker Trib was the most intensely developed watershed, with the highest road density and impervious surface cover.

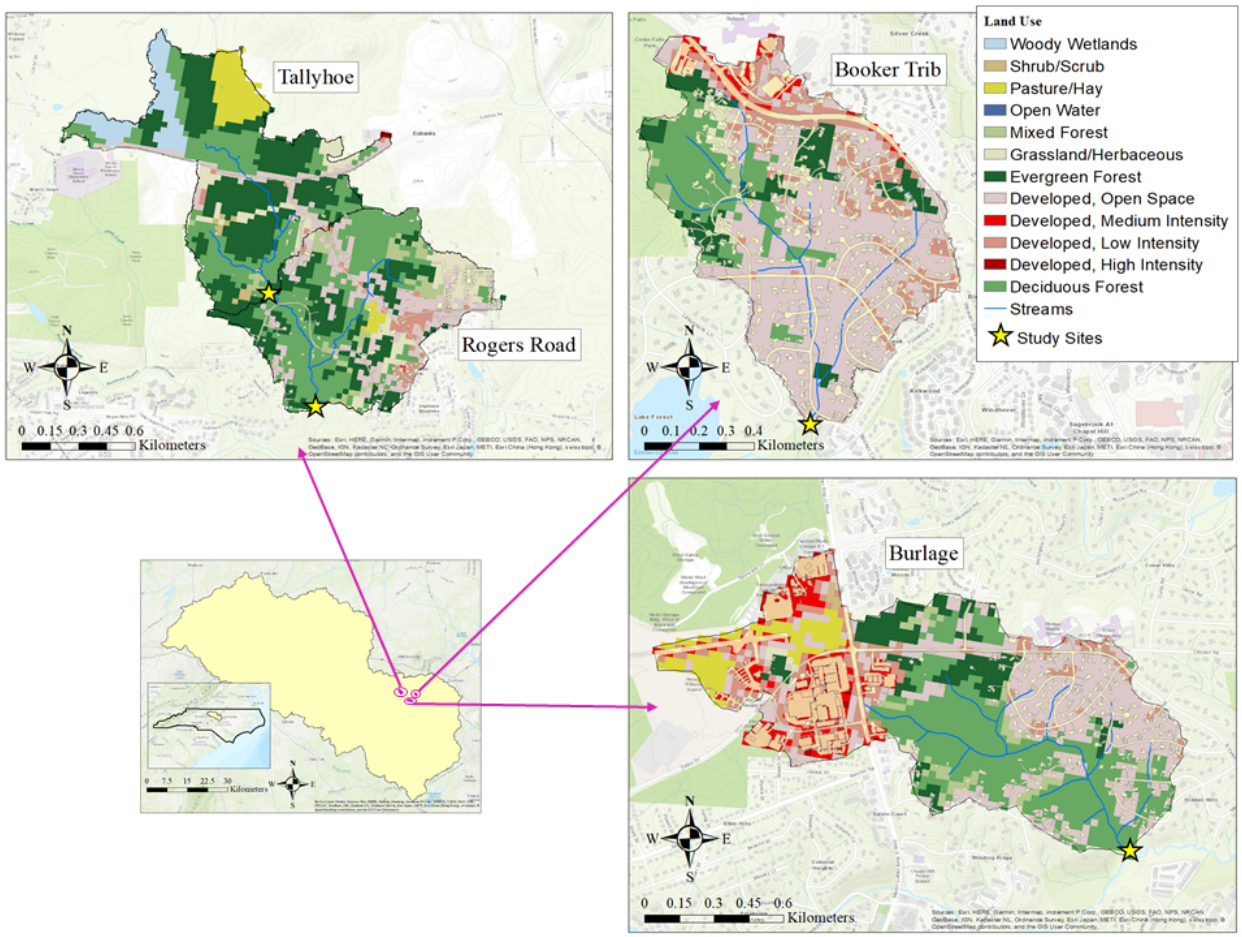

Figure 1. *The four study sites are all sub-watersheds of the larger Jordan Lake Watershed (bottom-left). Each map illustrates the land uses within each watershed, with a star representing the location of the sampling site.*

Watershed Site Name	Area (km^2)	ISC	Road Density (m/km^2)	Parcel Density (*per* km^2)	Sewer Density (m/km^2)
Tallyhoe	.95	2.6%	4543	41	0
Rogers Road	.99	9.0%	5778	243	1750
Burlage	1.53	20%	7870	167	5598
Booker Trib	.98	24%	9557	307	7178

Table 1. *Four study sites with varying land use practices. ISC stands for impervious surface cover.*

Date Collection

Data collection at these sites started in June, July, and September of 2017 using in-situ sensors, and continued in 5-minute intervals to present day. We used a HOBO U24 probe to measure specific conductance, using temperature and electrical conductivity, and a HOBO U20 for water level (Onset, Cape Cod, Massachusetts). These water level measurements were used in combination with velocity profiling by electromagnetic velocity sensors (Marsh-McBirney Flo-Mate, Frederick, Maryland), dilution gauging, and acoustic doppler profiling (SonTek IQ+, San Diego, California) to develop level-discharge rating curves. These rating curves gave us estimates of flow in 5-minute intervals.

Storm Isolation

I isolated storm events from our dataset using streamflow observations. Storm events generate a unique spike followed by a gradual decline in flow that can be isolated from non-storm behavior. Previous studies have automated some parts of the isolation process (Hammond & Han, 2006; Vaughan et al., 2017), yet delimitation of the storm's end has posed major challenges since the recession is often gradual, and, thus, difficult to pick out from base flow conditions. The high-resolution data also proved difficult to process due to the large amount of noise generated by 5-minute interval measurements.

The statistical software R was used to process data and build a repeatable and automated filter for delineating peaks in flow as storm events. I first reduced noise using locally weighted smoothing (LOESS, stats package), and performed an estimated baseflow-quickflow separation (EcoHydRology package). Isolating quickflow allowed storm detection to be more sensitive to changes in flow due to discharge as opposed to other processes such as diurnal fluctuations brought on by evapotranspiration or sensor error. Peaks were identified in quickflow, using the peakpick function (peakPick package), which outputs local maximums and ignores short spikes that could be attributed to noise. To prevent flashy storms from being classified as noise, I normalized quickflow down to values between 0 and 1 by dividing each point by the maximum flow. Two quickflow peaks needed to be to be at least 48 points (4 hours) apart to be considered separate storms. A point also needed to be at least three standard deviations above the mean of the surrounding 20 data points to be a peak. To prevent noise from being detected as peaks, I set a threshold for discharge, below which a peak could not be considered a storm. The discharge threshold was set at three times the mean of all discharge values at a given site. Thresholds

below this value resulted in an exponential increase in peak detection at all sites, which were attributed to noise.

From the identified storm peaks, I used a hidden Markov model to identify the start and end of each storm. The hidden Markov model has been used to evaluate rainfall occurrence (Gabriel & Neumann, 1962; Hughes et al., 1999); however, we are among the first to apply the model to measurements of flow to find storm events. When implementing the model, I differentiated between non-storm, storm, and flux states, where flux marked fluctuation in quickflow that was not necessarily a storm. At each point, the model predicted the probability of transitioning from one state to another. For example, during quickflow fluctuation the model predicted the probability that each measurement of flow would transition from a flux to a storm point, providing an estimation of the point at which the storm began. Each storm event was, thus, padded on either side with the flux state—the period of fluctuation in quickflow before and after the storm. I marked the first flux state point as the beginning of the storm and final flux point as the end of each storm (Figure 2). Sites with less impervious surface cover, however, generated less distinct changes in flow during storms. Tallyhoe, especially demonstrated smaller changes in quickflow during storms, so I adjusted the method at Tallyhoe to only mark storm states as storms. Altogether, the four study sites generated 189 storm events.

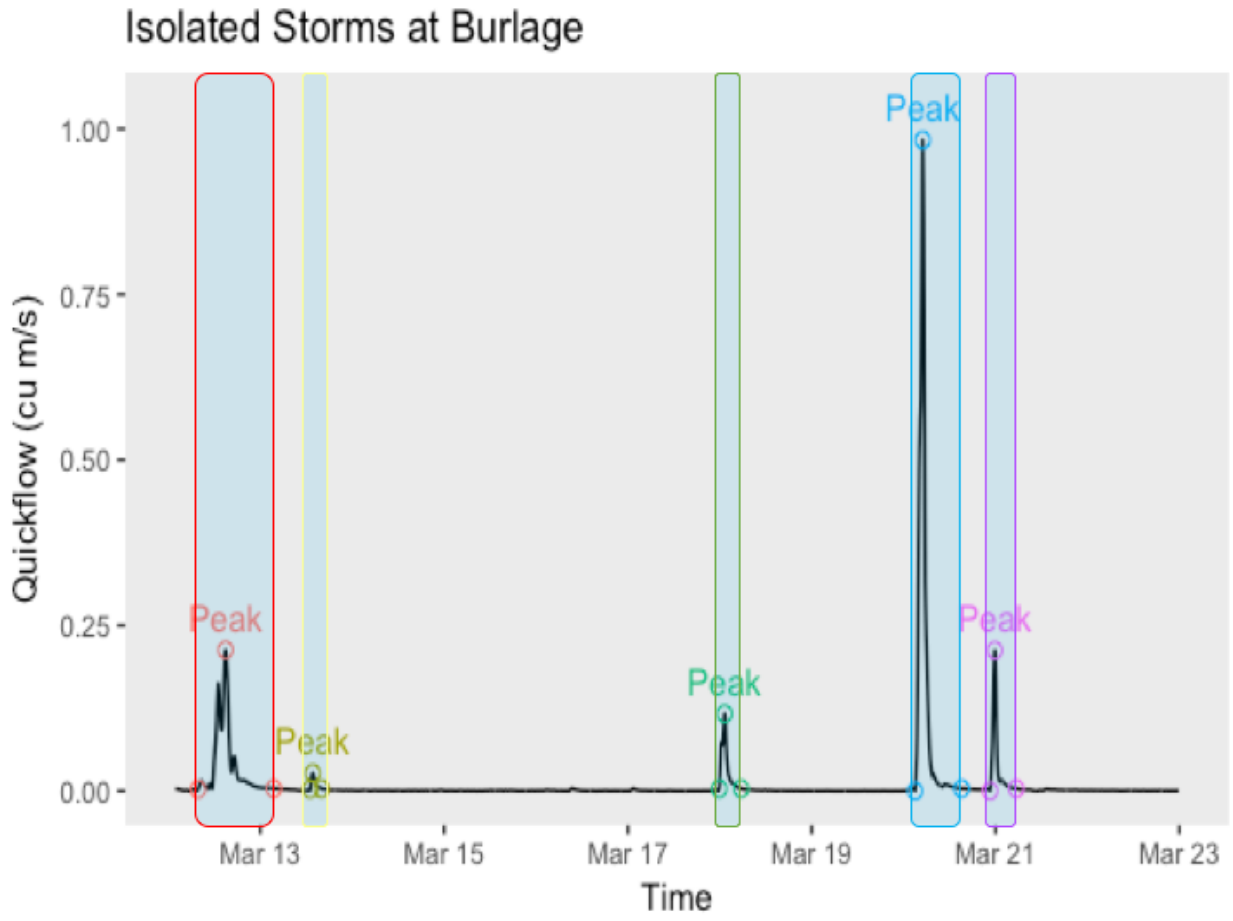

Figure 2. *Time series of quickflow in cubic meters/second at Burlage. Each shaded region marks an individual storm outputted by the model.*

Date Analysis

The start and end times for each storm were used to isolate storm data for analysis of flow versus specific conductance. Flow and specific conductance variables exhibit a hysteretic relationship because specific conductance shows a delayed response to changes in flow (Kobayashi et al., 1999; Clark et al., 2016). To characterize the relationship between the two variables, specific conductance was plotted against flow for each storm event to generate the even hysteresis curve. For each hysteresis, I determined the height—range between the maximum and minimum specific conductance—and the width—range between maximum and minimum flow. I then used the width and height to calculate the normalized width and height magnitude of each storm using the following equations:

$$\text{Normalized Flow (Q): } \frac{Q_{max} - Q_{min}}{Q_{max}}$$

$$\text{Normalized Specific Conductance (SC): } \frac{SC_{max} - SC_{min}}{SC_{max}}$$

To evaluated the behavior of each storm hysteresis, I computationally determined between enrichment and dilution. Enrichment (increase in specific conductance) took place if the third quartile value of specific conductance during a storm was greater than the specific conductance at the start of the storm. Dilution (decrease in specific conductance) took place if the third quartile specific conductance value during the storm was lower than at the start of the storm. I verified this method by looking at five hystereses from each site and assessing the direction by visual inspection. If a hysteresis demonstrated anticlockwise directionality, it was considered to be enriching, and clockwise behavior was labeled as dilution (Vaughan et al., 2017). Four out of five storms at Burlage and Booker Trib showed agreement between the computationally determined response with the visual verification, three of five agreed at Rogers Road, and all storms at Tallyhoe were a match. The computational method was effective for single peak storms that yielded a circular hysteresis. Hystereses with more complex shapes and multiple peaks confounded the results. Rogers Road especially, produced more of these complex storms that proved difficult to analyze.

Many storm events during the winter had highly variable measurements of specific conductance as a result of salts and brines laid out on roads to improve driving conditions. Runoff from roads carried these salts and caused major spikes in specific conductance measurements. I did not account for road salts in my analysis of nutrient movement; thus, storms that demonstrated these irregular spikes in specific conductance during the winter months were removed. I identified storms to remove by visually choosing a threshold at each study site above which specific conductance was attributed to road salts. Sites needed different thresholds because they showed different average specific conductance measurements. Forested sites, for instance, exhibited much lower specific conductance measurements during storms, and received fewer road salt inputs. Therefore, these sites needed a lower specific conductance threshold to catch and eliminate winter storms impacted by road salts. I also removed two non-winter storms from both Burlage and Rogers Road that showed extraordinarily high specific conductance measurements. By removing these outliers, I reduced my dataset by 17.4% to a total of 162 storms.

RESULTS

During storms, specific conductance exhibited a hysteretic response to changes in flow that were characterized as enrichment (increase followed by a slow decline) or dilution (decrease followed by a gradual climb). The normalized width and height from the storm hystereses were plotted, and grouped by specific conductance response (Figure 3) and relative storm flashiness, which describes the intensity of change in flow (e.g. flashier storms have a higher change in flow over a shorter period of time) (Figure 4). Boxplots were also generated illustrating magnitude of flow (Figure 5) and magnitude of specific conductance (Figure 6) grouped into different boxplots by relative flashiness. The maximum specific conductance of each storm was also plotted in chronological order of occurrence, and each storm was grouped into the season in which it occurred (Figure 7). Last, the same plot was generated without excluding outlier storms (Figure 8).

Plots of normalized magnitudes of storms (Figure 3) suggested a strong relationship between change in flow and specific conductance during storms. Change in specific conductance and flow at the two most developed sites–Booker Trib and Burlage–was highly correlated, but the variance in specific conductance explained by flow became much less significant at less developed sites. Linear regression between these two variables yielded an R-squared of .732 at Booker Trib and .773 at Burlage, and p-values less than .001 at both sites. At the less developed Rogers Road, the R-squared fell to .132, with a p-value less than .01. The least developed site—Tallyhoe—showed an even smaller R-squared of .094 and p-value less than .05. Specific conductance and flow did not exhibit a linear relationship at these sites with less urban development. While all sites demonstrated statistically significant p-values from linear regression, only Burlage and Booker Trib generated R-squared values supporting a linear relationship between the two variables.

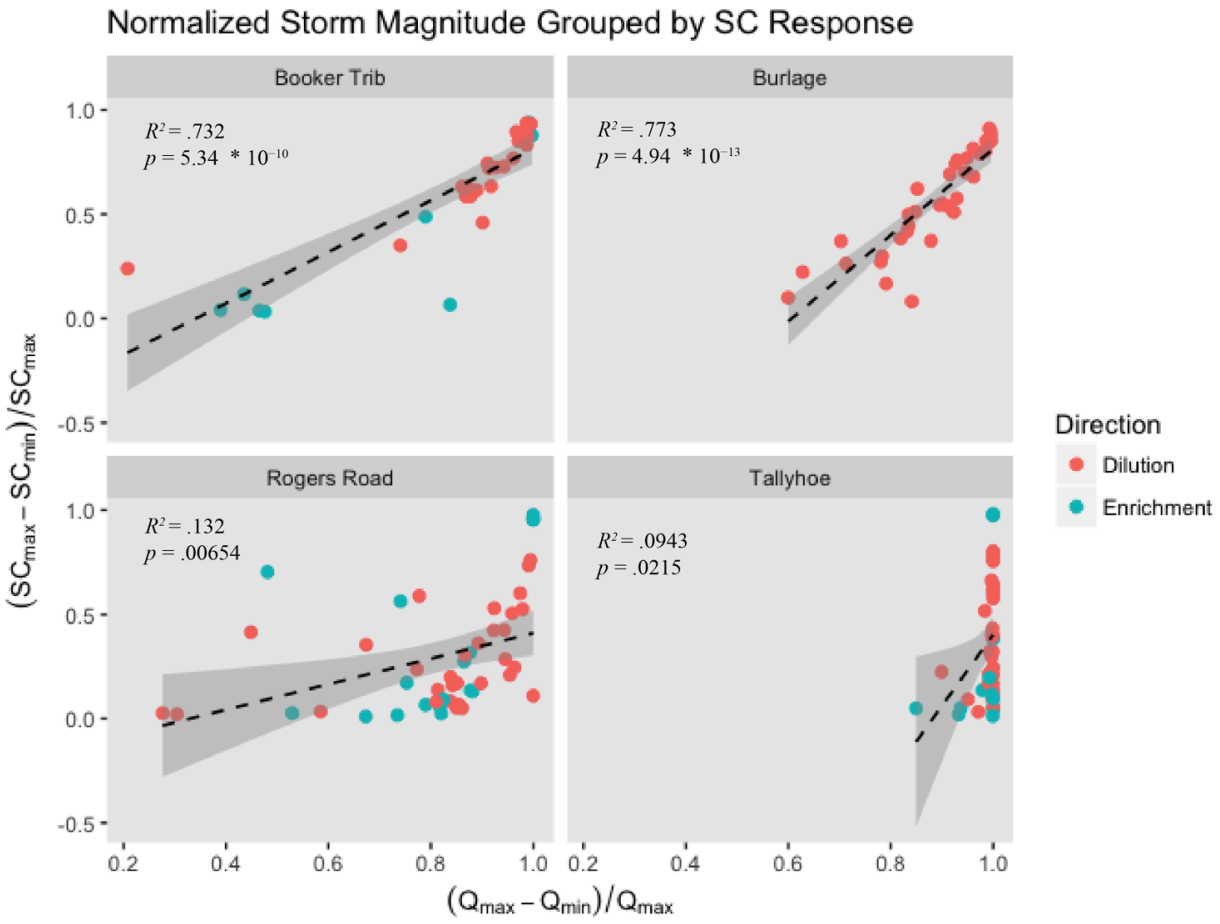

Figure 3. *Normalized flow magnitude versus normalized specific conductance magnitude of each storm at each study site. Shaded region indicates a 95% confidence interval. Color of each storm is determined by whether dilution or enrichment occurred during that storm.*

Discrepancy amongst sites also appears when grouping storms according to specific conductance response—less developed sites exhibited dilution more than enrichment, while more developed sites diluted more often (Figure 3). Burlage consistently exhibited dilution during its 37 storms. Booker Trib, on the other hand, exhibited enrichment in 25.8% of storms (n=31) and dilution in 74.2%. Tallyhoe and Rogers Road exhibited enrichment most frequently. Tallyhoe enriched during 30.4% of storms (n=46) and diluted in 69.6%, and Rogers Road exhibited enrichment in 35.4% (n=48) and dilution in 64.6%. At Booker Trib and Rogers Road enrichment mostly occurred during storms with little change in flow and specific conductance. Tallyhoe mostly enriched during storms with little variation in specific conductance.

Storms at more developed sites with higher flow and specific conductance magnitudes also exhibited a faster and more drastic change in flow (higher flashiness) (Figure 4). Less developed sites demonstrated no apparent relationship between flashiness and variation in flow. This discrepancy amongst sites was supported by boxplot representations that compared relative flashiness to normalized specific conductance and flow magnitude (Figures 5&6). In both

boxplots, flow and specific conductance magnitude increased with relative flashiness at Booker Trib and, more weakly, at Burlage. Less urbanized sites showed no apparent relationship between flashiness and specific conductance or flow.

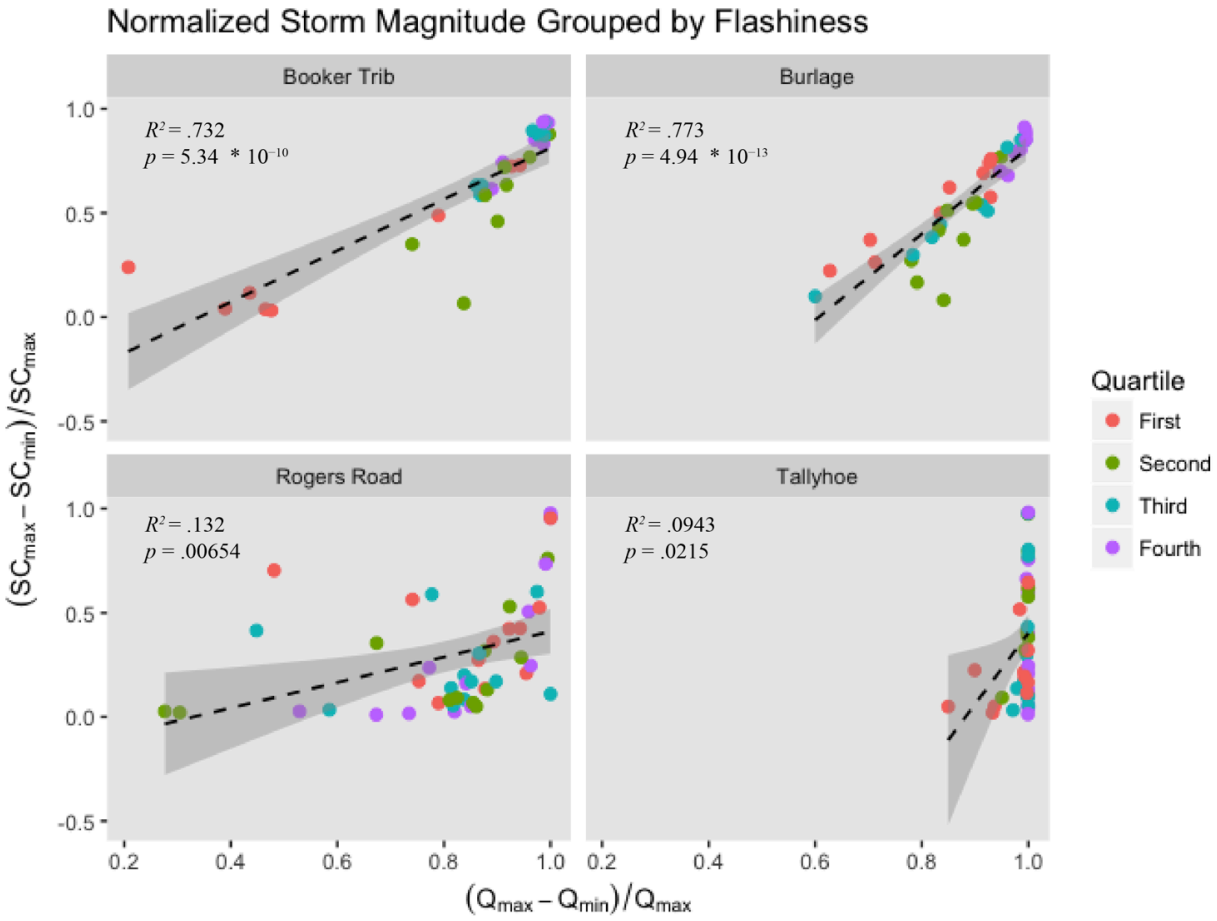

Figure 4. *Normalized flow magnitude versus normalized specific conductance magnitude with a regression line. The shaded region marks a 95% confidence interval. The color of each storm is determined by the quartile of flashiness it falls under for that study site (e.g. first quartile storms were the least flashy at that site).*

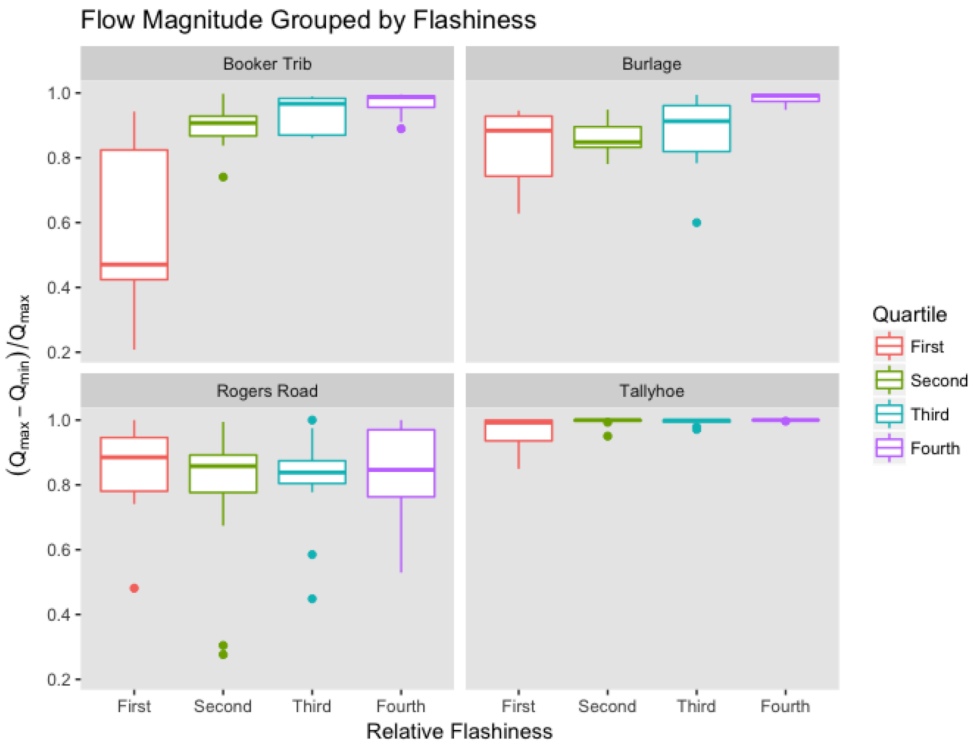

Figure 5. *Magnitude of flow of each storm grouped into boxplots by relative flashiness.*

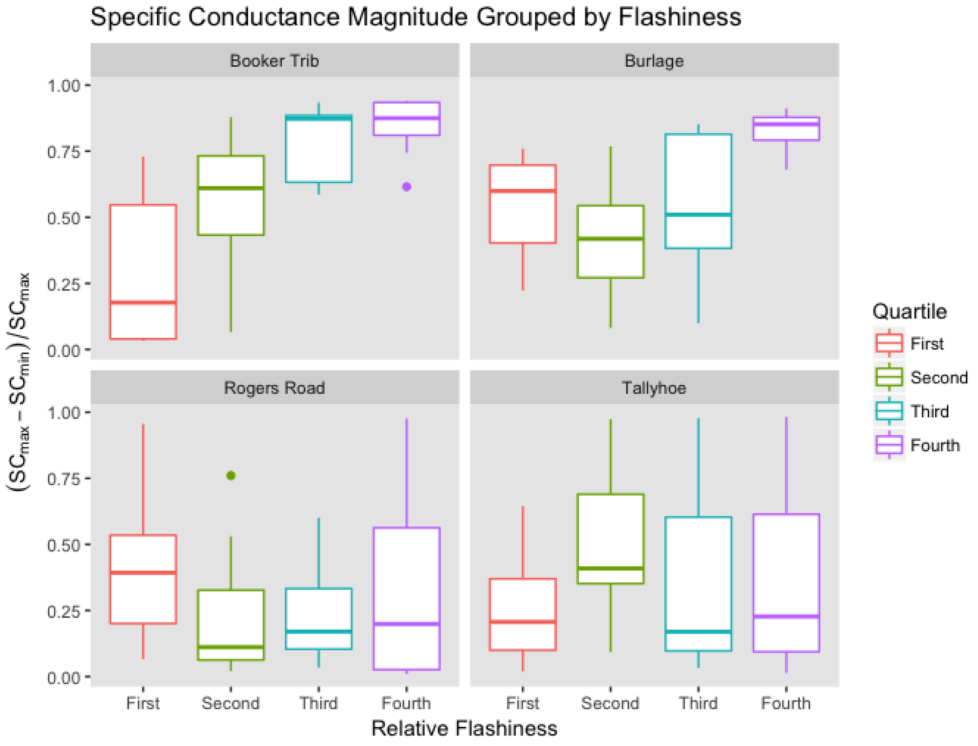

Figure 6. *Magnitude of specific conductance of each storm grouped into boxplots by relative flashiness.*

Plotting the maximum specific conductance for each storm (Figure 7) indicated that Booker Trib and Burlage had storms with higher specific conductance maximums ranging from 150-400 µS/cm. Storms at Rogers Road and Tallyhoe varied less and showed lower measurements of specific conductance, ranging from 50-250 µS/cm. Including outlier storms revealed that winter storms especially showed high specific conductance measurements (Figure 8). Winter storms at both Burlage and Booker Trib showed maximum specific conductance measurements at least several hundred µS/cm above almost all other storms at that site. Rogers Road and Tallyhoe also showed several winter storms with higher maximum specific conductance measurements than other storms, but only 100-200 µS/cm greater.

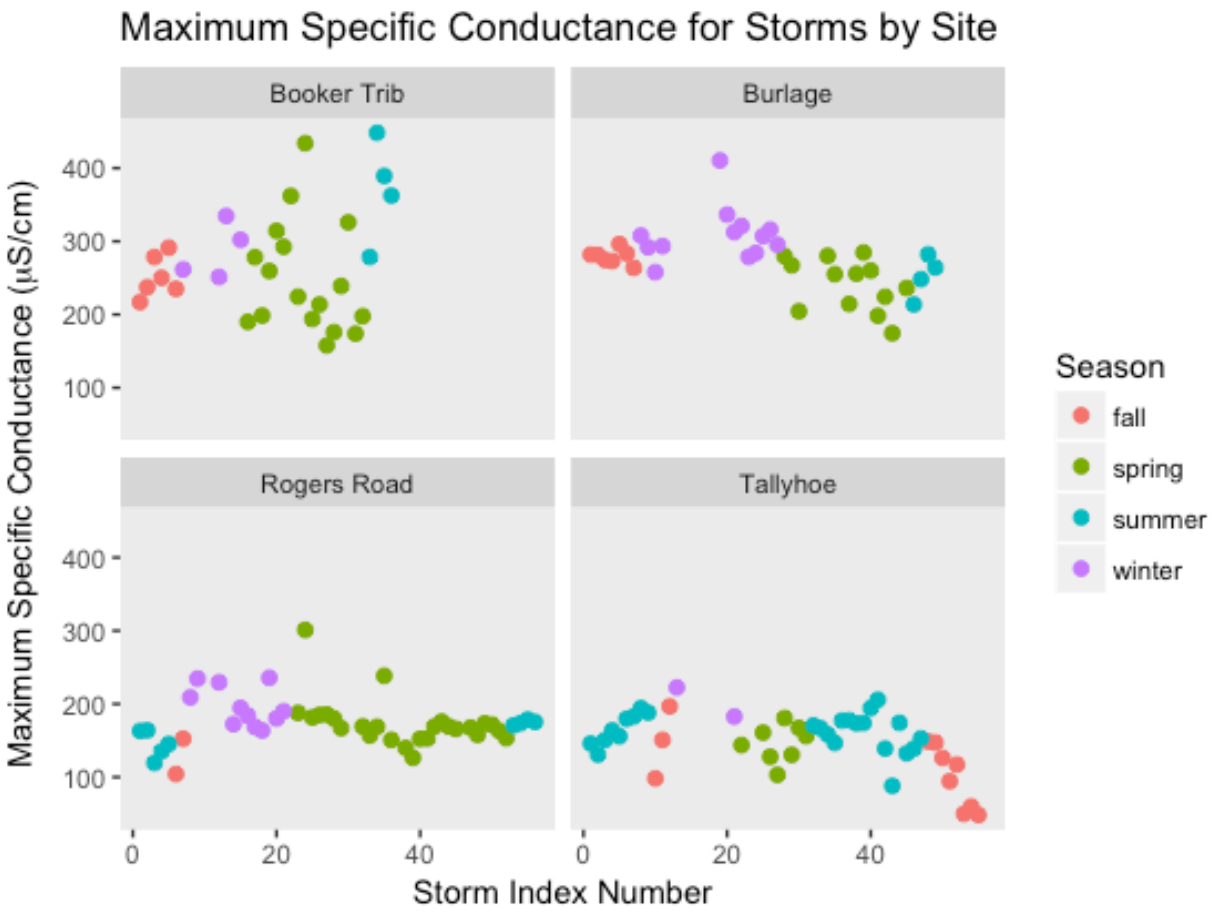

Figure 7. *Maximum specific conductance measurement from each storm. Storms are plotted in chronological order from left to right. Color is based off of the season in which that storm took place.*

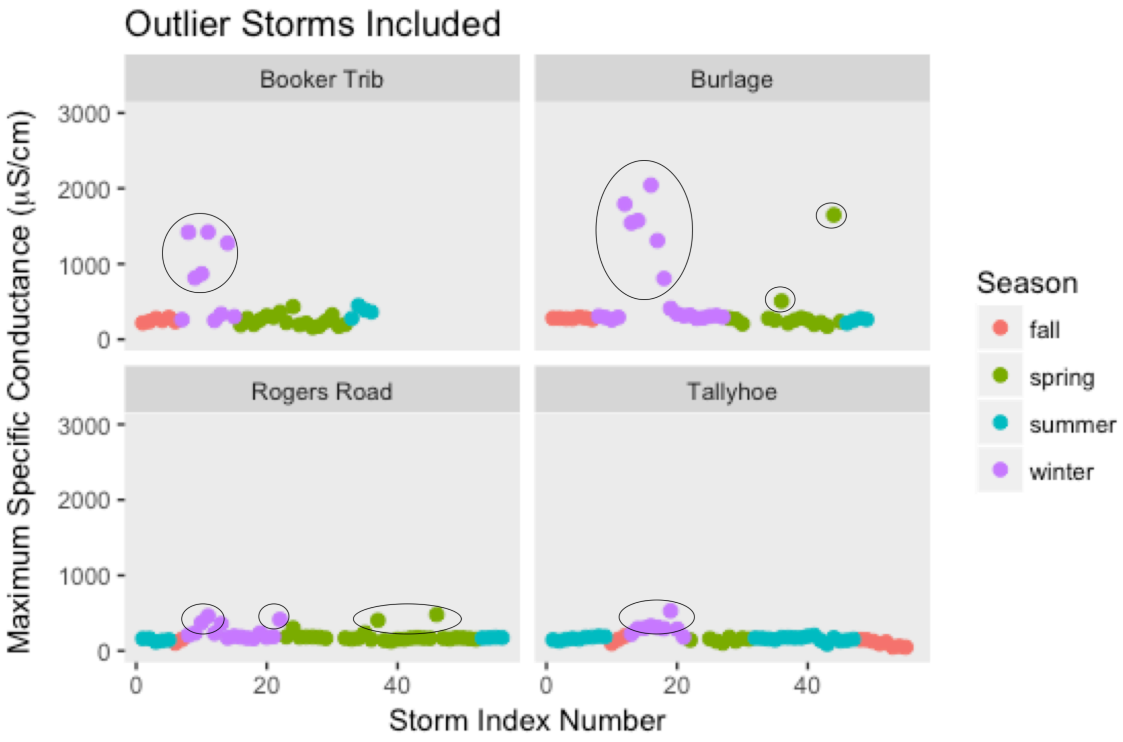

Figure 8. *Maximum specific conductance measurements from all observed storms. Storms inside black circles were excluded from analyses due to abnormally high specific conductance measurements.*

DISCUSSION

These findings suggest that during small, low-intensity storms, enrichment of dissolved ions tends to occur, while during large, flashier storms dilution of dissolved ions tends to occur. These patterns apply most strongly to the most developed site (Booker Trib), and apply less well to the less developed sites (Rogers Road and Tallyhoe) where flashier storms still showed enrichment. This lack of consistency between sites is likely due to higher potential for infiltration across landscapes draining into less developed sites. As water drains over the surface of a landscape into the stream, it transports contaminants on the surface into streams, but water that infiltrates also carries subsurface pollutants (Pitt et al., 1999). If sites with high potential for infiltration received significant inputs from groundwater sources, then streams at those sites may exhibit increased enrichment due to inputs from underground sources. In particular, Rogers Road may have exhibited this behavior because septic systems are the primary sewage treatment of residents in that drainage area, and contaminants likely leached from those systems into the groundwater (Swartz et al., 2006). The stream at Rogers Roads also enriched most often—35.4% of storms—which supports the theory that the site received contaminant input from surrounding septic systems. Landscapes draining into Booker Trib and Burlage, on the other hand, have greater impervious surface cover, suggesting that stormwater primarily came from surface runoff, leading to dilution. Several studies have connected stormwater infiltration to contamination of groundwater (Ku & Simmons, 1986; Pitt et al., 1999; Fischer et al., 2003), and interactions between surface and subsurface water also lead to exchange of contaminants (Sophocleous, 2002). While these surface-groundwater interactions are complex and difficult to

quantify, the results of this study make clear the need to further investigate those relationships during storms for watersheds with higher infiltration potential. Landscapes with more impervious surfaces, however, seem to mostly receive surface inputs during storms.

Specific conductance and flow exhibit a more direct relationship in sites with higher development. At Booker Trib and Burlage, the magnitude of specific conductance and flow exhibited a strong linear relationship, indicating that storms with higher ranges of flow would be more likely to have higher ranges in specific conductance. Again, this discrepancy is likely due to a more direct transport of contaminants over the surface of the landscape at sites with higher impervious surface cover than sites with more potential for infiltration. This discrepancy, too, prompts further investigation on interactions between groundwater and surface water, and supports that streams in watersheds with more impervious surfaces are more dependent on stormwater inputs from the surface.

The higher specific conductance measurements during storms indicate that nutrients mobilize in surface discharge over more developed landscapes. Sites with more urban development contribute more contaminants as indicated by the higher specific conductance measurements. These findings are consistent with other studies investigating stormflow discharge across varying land uses (Rimer et al., 1978; Lee & Bang, 2000). While streams in developed watersheds may receive some nutrient inputs from groundwater, the higher impervious surface cover supports that most nutrients enter the stream as surface runoff. As storms result in surface runoff, they serve as the primary contributor of nutrient transport in urban watersheds (Lee & Bang, 2000; Lawler et al., 2006; Butturini et al., 2008; Vaughan et al., 2017). The Jordan Lake Nutrient Study should continue investigation into storm events as a major transporter of nutrients, and the larger study will analyze movement of specific contaminants during storm events. By determining the flow paths of specific nutrients throughout the landscape, the larger study will provide a better picture of nutrient transport across the larger watershed, and contribute information needed for development of effective mitigation strategies.

Winter storms impacted by roads salts and brine introduced new variables that were not covered in this study, but generated significant specific conductance responses, especially at more developed sites. The large change in specific conductance at sites with more urban development compared to noticeable but less drastic changes at more forested sites merits attention. Booker Trib and Burlage are both more likely to receive direct inputs from roads treated with road salts and brine, and their higher degree of impervious surface cover could also facilitate transport of those salts. While it is possible that Rogers Road and Booker Trib also receive salt inputs, fewer impervious surfaces and fewer treated roads within those sub-watersheds are likely to reduce transport and volume of salts reaching the streams. It has been shown that as percent of impervious surfaces increases, up to half of applied salts can enter surface water discharge (Marsalek, 2003). While road salts were not the focus of this study, the impacts of large amounts of salt inputs applied across urban to rural transition landscapes, including contamination of groundwater, reduced biodiversity, and poorer drinking water quality, may warrant further examination (Crowther & Hynes, 1977; EC & HC, 2001; Marsalek, 2003).

Further analyses could be performed on the specific conductance and flow hystereses to obtain more information. Other studies have utilized a new hysteresis index for interpreting hysteresis of flow versus a response variable such as turbidity or nitrate yield during storms (Lloyd et al., 2016; Vaughan et al., 2017). This index can be used to determine the direction of a hysteresis,

which would help to validate and/or improve the accuracy of my results. Other studies have also used a flush index that compares the initial contaminant concentration with the contamination during peak discharge of the storm (Butturini et al., 2008; Vaughan et al., 2017). While this flush index is fairly similar to my own, it could also be used to improve and/or validate my results. Future studies could also improve upon the storm isolation method developed in this study. Currently, more forested sites challenge the storm isolation method because storms at those sites generate less distinct changes in quickflow. More parameters could be added to the hidden Markov model to dictate necessary probabilities for transitions between different states. For instance, the model could be set to 90% confidence needed for a state to change from non-storm to flux state. By introducing these parameters, the storm isolation method could be made more sensitive to smaller changes in flow that indicate storm events at sites with less urban development.

CONCLUSION

I sought to improve understanding of nutrient transport over urban landscapes. My findings support that different landscapes transport nutrients in different ways with the greatest determining factors being amount of urban development and degree of impervious surface cover. The higher specific conductance measurments from more developed sites indicate that more development leads to higher contributions of contaminants. The different responses in specific conductance among the study sites begs the question–what causes streams to enrich versus dilute during storms? My results show that streams surrounded by urban development tend to dilute more often than enrich, but the downstream effects of this response are still unclear. My results leave unanswered questions about the mechanisms causing eutrophication in Jordan Lake. The UNC Nutrient Study, however, will continue to conduct analyses using many more sites and several other variables including individual nutrient concentrations which will be better informed by this work. Although imperfect, the storm isolation method will prove useful for future analyses carried out on storms in the larger project and in future studies.

AKNOWLEDGEMENTS

Thank you to Diego Riveros-Iregui and Joseph Delesantro for their continuous support and mentorship throughout this project, as well as Matt Jansen for his advice and contributions towards the statistical analyses of this research. I am grateful to the Megan Hughes and the IDEA program for funding and initiating my research, as well as Sarah Steele Danhoff for further funding this thesis. I thank the UNC Environment, Energy, and Ecology Program for overseeing this thesis and oral defense. I am grateful to Kriddie Whitmore, Rhyan Stone, Tianzhen Nie, and previous members of Carbonshed Lab for collecting the data I used in this project.

REFERENCES

Alexander, R. B., Boyer, E. W., Smith, R. A., Schwarz, G. E., & Moore, R. B. (2007). The Role of Headwater Streams in Downstream Water Quality 1. JAWRA Journal of the American Water Resources Association, 43(1), 41-59.

Bache, S. M., and Wickham, H. (2014). magrittr: A Forward-Pipe Operator for R. R package version 1.5. https://CRAN.R-project.org/package=magrittr

Butturini, A., Alvarez, M., Bernal, S., Vazquez, E., & Sabater, F. (2008). Diversity and temporal sequences of forms of DOC and NO3- discharge responses in an intermittent stream: Predictable or random succession?. Journal of Geophysical Research: Biogeosciences, 113(G3).

Carpenter, S. R., Bolgrien, D., Lathrop, R. C., Stow, C. A., Reed, T., & Wilson, M. A. (1998). Ecological and economic analysis of lake eutrophication by nonpoint pollution. Australian Journal of Ecology, 23(1), 68-79.

Carpenter, S. R., Ludwig, D., & Brock, W. A. (1999). Management of eutrophication for lakes subject to potentially irreversible change. Ecological applications, 9(3), 751-771.

Clark, E. V., Greer, B. M., Zipper, C. E., & Hester, E. T. (2016). Specific conductance–stage relationships in Appalachian valley fill streams. Environmental Earth Sciences, 75(17), 1222.

Codd, G. A. (2000). Cyanobacterial toxins, the perception of water quality, and the prioritisation of eutrophication control. Ecological engineering, 16(1), 51-60.

Conley, D. J., Paerl, H. W., Howarth, R. W., Boesch, D. F., Seitzinger, S. P., Havens, K. E., … & Likens, G. E. (2009). Controlling eutrophication: nitrogen and phosphorus. Science, 323(5917), 1014-1015.

Crowther, R. A., & Hynes, H. B. N. (1977). The effect of road deicing salt on the drift of stream benthos. Environmental Pollution (1970), 14(2), 113-126.

Daniel, T. C., Sharpley, A. N., Edwards, D. R., Wedepohl, R., & Lemunyon, J. L. (1994). Minimizing surface water eutrophication from agriculture by phosphorous management. Journal of Soil and Water Conservation, 49(2), 30.

Environment Canada & Health Canada. (2001). Priority substances list assessment report: Road Salts. Ottawa, Ontario.

Fischer, D., Charles, E. G., & Baehr, A. L. (2003). Effects of stormwater infiltration on quality of groundwater beneath retention and detention basins. Journal of Environmental Engineering, 129(5), 464-471.

Fuka, D. R., Walter, M. T., Archibald, J. A., Steenhuis, T. S., and Easton, Z. M. (2014). EcoHydRology: A community modeling foundation for Eco-Hydrology. R package version 0.4.12. https://CRAN.R-project.org/package=EcoHydRology

Gabriel, K., and Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. Quarterly Journal of the Royal Meteorological Society, 88(375), 90–95. https://doi.org/10.1002/qdj.49708837511

Grolemund, G., and Wickham, H. (2011). Dates and Times Made Easy with lubridate. Journal of Statistical Software, 40(3), 1-25. URL http://www.jstatsoft.org/v40/i03/.

Hammond, M., & Han, D. (2006). Recession curve estimation for storm event separations. Journal of Hydrology, 330(3-4), 573-585.

Hessen, D. O., Hindar, A., & Holtan, G. (1997). The significance of nitrogen runoff for eutrophication of freshwater and marine recipients. Ambio, 312-320.

Hughes, J. P., Guttorp, P., and Charles, S. P. (1999). A non-homogeneous hidden markov model for precipitation occurrence. Journal of the Royal Statistical Society. Series C: Applied Statistics, 48(1), 15-30.

Jarvie, H. P., Neal, C., & Withers, P. J. (2006). Sewage-effluent phosphorus: a greater risk to river eutrophication than agricultural phosphorus?. Science of the total environment, 360(1-3), 246-253.

Kobayashi, D., Ishii, Y., & Kodama, Y. (1999). Stream temperature, specific conductance and runoff process in mountain watersheds. Hydrological Processes, 13(6), 865-876.

Ku, H. F., & Simmons, D. L. (1986). Effect of urban stormwater runoff on ground water beneath recharge basins on Long Island, New York (No. 85-4088). US Geological Survey.

Lawler, D. M., Petts, G. E., Foster, I. D., & Harper, S. (2006). Turbidity dynamics during spring storm events in an urban headwater river system: The Upper Tame, West Midlands, UK. Science of the Total Environment, 360(1-3), 109-126.

Lee, J. H., & Bang, K. W. (2000). Characterization of urban stormwater runoff. Water research, 34(6), 1773-1780.

Lloyd, C. E. M., Freer, J. E., Johnes, P. J., & Collins, A. L. (2016). Testing an improved index for analysing storm discharge–concentration hysteresis. Hydrology and Earth System Sciences, 20(2), 625-632.

Marsalek, J. (2003). Road salts in urban stormwater: an emerging issue in stormwater management in cold climates. Water Science and Technology, 48(9), 61-70.

Mörth, C. M., Humborg, C., Eriksson, H., Danielsson, Å., Medina, M. R., Löfgren, S., … & Rahm, L. (2007). Modeling riverine nutrient transport to the Baltic Sea: a large-scale approach. Ambio, 124-133.

Ning, S. K., Jeng, K. Y., & Chang, N. B. (2002). Evaluation of non-point sources pollution impacts by integrated 3S information technologies and GWLF modelling. Water Science and Technology, 46(6-7), 217-224.

North Carolina Department of Environmental Quality (NC DEQ). (2009). Jordan Lake Nutrient Strategy. Retrieved 2019, from https://deq.nc.gov/about/divisions/water-resources/water-planning/nonpoint-source-planning/jordan-lake-nutrient

OECD, 1982. Eutrophication of waters, monitoring, assessment and control. Organisation for Economic Cooperation and Development, Paris.

Paerl, H. W. (2006). Assessing and managing nutrient-enhanced eutrophication in estuarine and coastal waters: interactive effects of human and climatic perturbations. Ecological Engineering, 26(1), 40-54.

Pitt, R., Clark, S., & Field, R. (1999). Groundwater contamination potential from stormwater infiltration practices. Urban water, 1(3), 217-236.

Rimer, A. E., Nissen, J. A., & Reynolds, D. E. (1978). Characterization and impact of stormwater runoff from various land cover types. Journal (Water Pollution Control Federation), 252-264.

Royer, T. V., Tank, J. L., & David, M. B. (2004). Transport and fate of nitrate in headwater agricultural streams in Illinois. Journal of Environmental Quality, 33(4), 1296-1304.

Savage, C., Leavitt, P. R., & Elmgren, R. (2010). Effects of land use, urbanization, and climate variability on coastal eutrophication in the Baltic Sea. Limnology and Oceanography, 55(3), 1033-1046.

Schindler, D. W. (1974). Eutrophication and recovery in experimental lakes: implications for lake management. Science, 184(4139), 897-899.

Session Law (2016) (NC) s. 14.13 (US).

Smith, V. H. (2003). Eutrophication of freshwater and coastal marine ecosystems a global problem. Environmental Science and Pollution Research, 10(2), 126-139.

Smith, V. H., Tilman, G. D., & Nekola, J. C. (1999). Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. Environmental pollution, 100(1-3), 179-196.

Søndergaard, M., Jensen, J. P., & Jeppesen, E. (2003). Role of sediment and internal loading of phosphorus in shallow lakes. Hydrobiologia, 506(1-3), 135-145.

Sophocleous, M. (2002). Interactions between groundwater and surface water: the state of the science. Hydrogeology journal, 10(1), 52-67.

Swartz, C. H., Reddy, S., Benotti, M. J., Yin, H., Barber, L. B., Brownawell, B. J., & Rudel, R. A. (2006). Steroid estrogens, nonylphenol ethoxylate metabolites, and other wastewater contaminants in groundwater affected by a residential septic system on Cape Cod, MA. Environmental science & technology, 40(16), 4894-4902.

US Army (1977). Storage, Treatment, Overflow, Runoff Model "STORM"-Users Manual. Corps of Engineers, U.S. Army.

US EPA, (1996a). Environmental Indicators of Water Quality in the United States (US EPA 841-R-96-02). Office of Water (4503F), US Government.

Vaughan, M. C., Bowden, W. B., Shanley, J. B., Vermilyea, A., Sleeper, R., Gold, A. J., … & Birgand, F. (2017). High- frequency dissolved organic carbon and nitrate measurements reveal differences in storm hysteresis and loading in relation to land cover and seasonality. Water Resources Research, 53(7), 5345-5363.

Visser, I., and Speekenbrink, M. (2010). depmixS4: An R Package for Hidden Markov Models. Journal of Statistical Software, 36(7), 1-21. URL http://www.jstatsoft.org/v36/i07/.

Vollenweider, R. A. (1970). Scientific fundamentals of the eutrophication of lakes and flowing waters, with particular reference to nitrogen and phosphorus as factors in eutrophication. Paris: OECD.

Westermark, P. O. (2015). peakPick: Peak Picking Methods Inspired by Biological Data. R package version 0.11. https://CRAN.R-project.org/package=peakPick

WHO (1997). Report of the Working Group on Chemical Substances in Drinking Water, Geneva, 22–26 April 1997. Section 5.2. Microcystin-LR. World Health Organization, Geneva

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., François, R., Henry, L., and Müller, K. (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.6. https://CRAN.R-project.org/package=dplyr

Zeileis, A., Hornik, K., and Murrell, P. (2009). Escaping RGBland: Selecting Colors for Statistical Graphics. Computational Statistics & Data Analysis, 53(9), 3259-3270. doi:10.1016/j.csda.2008.11.033

APPENDIX

Storm Isolation

```
#Isolates storms from flow dataset and outputs csv with start and end times o
f storms
#Will Hamilton
#wamilton@ad.unc.edu
#February 18, 2019

library(pacman)
p_load(tidyverse, EcoHydRology, magrittr,
       lubridate, peakPick, depmixS4, colorspace)

#run one site at a time

data <- read.csv("Data/BG_10_19_18.csv")
#data <- read.csv("Data/BT_10_19_18.csv")
#data <- read.csv("Data/RR_10_19_18.csv")
#data <- read.csv("Data/TH_12_9_18.csv")

#Format the Timestamp
data$TIMESTAMP <- ymd_hms(data$TIMESTAMP)
num.time <- as.numeric(as.POSIXct(data$TIMESTAMP, tz="UTC"))

#loess smoothing
Sm <- loess(data$Q~ num.time, span = 0.00025, Data=data,na.rm=TRUE)
Sm.predict <- predict(Sm , Data=Data)

#copy data into dataCopy, supplies Baseflow with timestamps
dataCopy <- c()
```

```r
dataCopy$Q <- data$Q
dataCopy$Lvl_m <- data$Lvl_m
dataCopy$TIMESTAMP <- data$TIMESTAMP
dataCopy <- as.data.frame(dataCopy)
dataCopy <- na.omit(dataCopy)


#loess baseflow separation
loessBaseflow <- BaseflowSeparation(Sm.predict, filter_parameter = .91, passe
s = 3)
loessBaseflow$TIMESTAMP <- dataCopy$TIMESTAMP

#FIND PEAKS

#eliminate quickflow low peaks and find storm peaks
quickFlow <- loessBaseflow$qft
quickFlow <- pmax(quickFlow, 3*mean(loessBaseflow$qft))
#scale so we capture huge spikes (only changes result for BT b/c captures hug
e peaks)
quickFlow <- quickFlow/max(quickFlow)
peaks <- peakpick(quickFlow, 48, peak.min.sd=3)
stormEvents <- c()
stormEvents$peakFlow <- loessBaseflow$qft[which(peaks)]
stormEvents$peakTime <- loessBaseflow$TIMESTAMP[which(peaks)]


#HIDDEN MARKOV

#we will run the model by month
num.month <- as.character(month(loessBaseflow$TIMESTAMP))

loessBaseflow$TIMESTAMP <- ymd_hms(loessBaseflow$TIMESTAMP)
dataCopy$TIMESTAMP <- ymd_hms(dataCopy$TIMESTAMP)

#delete any month where the stream was dry
sub <- unique(num.month)
for (k in sub){
  if(max(dataCopy$Q[num.month==k]) == 0){
    sub <- sub[sub != k]
  }
}

loessBaseflow$num.month <- as.character(month(loessBaseflow$TIMESTAMP))

#copy loessBaseflow into master for months that we want to use in model
master <- loessBaseflow
master$num.month <- as.character(month(master$TIMESTAMP))
master <- master[master$num.month %in% sub,]
master$qft.cut <- master$qft
```

```r
master$qft.cut[master$qft.cut>quantile(master$qft.cut,0.999)] <- quantile(mas
ter$qft.cut,0.999) #brings points down to 99.9 percentile


n <- length(sub)

#Run Hideen Markov Model
hmm <- depmix(response=qft.cut~num.month,
              data=master,
              nstates=3,
              family=gaussian("log"),
              respstart=c(0,rep(0,n-1),0, #interecept, month #, sd
                          0,rep(0,n-1),.01,
                          0,rep(0,n-1),1))
summary(hmm)

fm <- fit(hmm,emcontrol=em.control(random.start=TRUE))

out <- summary(fm)

loessBaseflow$state <- NA
loessBaseflow$state[loessBaseflow$num.month %in% sub] <- posterior(fm)[,1]

#formats storm state into asscending order based on quickflow (nonstorms will
be lowest state)
loessCopy <- loessBaseflow
loessCopy <- loessCopy[!is.na(loessCopy$state),]
grouped <- loessCopy %>% group_by(state) %>%
  summarise(qftVal = mean(qft))
grouped <- grouped[order(grouped$qftVal, decreasing=F),]
loessBaseflow$state <- match(loessBaseflow$state, grouped$state)


#plot HMM
plot(loessBaseflow$TIMESTAMP,loessBaseflow$qft,col="grey80",main="HMM CUT Lab
els",type="l")
points(loessBaseflow$TIMESTAMP,loessBaseflow$qft,col=loessBaseflow$state,pch=
20,cex=0.6)

#END TIMES

#for each peak time go forward and find timestamp of the first nonstorm state
endTimes <- c()
findEndTime <- function(peakTime){
  tempTime <- loessBaseflow$TIMESTAMP[match(peakTime, loessBaseflow$TIMESTAMP
):length(loessBaseflow$state)]
  endTimes <- c(endTimes, tempTime[match(1, loessBaseflow$state[match(peakTim
e, loessBaseflow$TIMESTAMP):length(loessBaseflow$state)])])
}
```

```r
#save results in endTimes, then properly format the timestamps
endTimes <- sapply(stormEvents$peakTime, findEndTime)
endTimes <- loessBaseflow$TIMESTAMP[match(endTimes, loessBaseflow$TIMESTAMP)]

#save results in stormEvents
stormEvents$endTime <- c()
stormEvents$endFlow <- c()
stormEvents$endTime <- endTimes
stormEvents$endFlow <- loessBaseflow$qft[match(endTimes, loessBaseflow$TIMEST
AMP)]

#get rid of duplicate endTimes b/c those will be the same storm
stormEvents <- as.data.frame(stormEvents)
stormEvents <- stormEvents %>% group_by(endTime) %>% dplyr::filter(peakFlow =
= max(peakFlow))

#START TIMES

#for each peak time go backward and find timestamp of last nonstorm state
startTimes <- c()
findStartTime <- function(peakTime){
  tempTime <- loessBaseflow$TIMESTAMP[match(peakTime, loessBaseflow$TIMESTAMP
):1]
  startTimes <- c(startTimes, tempTime[match(1, loessBaseflow$state[match(pea
kTime, loessBaseflow$TIMESTAMP):1])])
}

#save results in startTimes, then properly format the timestamps
startTimes <- sapply(stormEvents$peakTime, findStartTime)
startTimes <- loessBaseflow$TIMESTAMP[match(startTimes, loessBaseflow$TIMESTA
MP)]

#save results in stormEvents
stormEvents$startTime <- c()
stormEvents$startFlow <- c()
stormEvents$startTime <- startTimes
stormEvents$startFlow <- loessBaseflow$qft[match(startTimes, loessBaseflow$TI
MESTAMP)]

#get rid of duplicate startTimes b/c those will be the same storm
stormEvents <- as.data.frame(stormEvents)
stormEvents <- stormEvents %>% group_by(startTime) %>% dplyr::filter(peakFlow
== max(peakFlow))

#VISUALIZATION
#color palette comes out weird the first time, so I run points twice the firs
t time to get the colors right
plot(loessBaseflow$TIMESTAMP[50000:55000], loessBaseflow$qft[50000:55000], "l
```

```
")
points(stormEvents$startTime, stormEvents$startFlow, col=palette(rainbow(6)))
points(stormEvents$startTime, stormEvents$startFlow, col=palette(rainbow(6)))
points(stormEvents$endTime, stormEvents$endFlow, col=palette(rainbow(6)))
points(stormEvents$peakTime, stormEvents$peakFlow, col=palette(rainbow(6)))

#plot example of storm isolation
print(ggplot(stormEvents[35:39,])+
        ggtitle("Isolated Storms at Burlage")+
        geom_line(data=loessBaseflow[51850:55000,], aes(x=TIMESTAMP, y=qft))+
        geom_point(aes(x=peakTime, y=peakFlow, color=as.factor(peakTime)), sh
ape=1, size=2)+
        geom_point(aes(x=startTime, y=startFlow, color=as.factor(peakTime)),
shape=1, size=2)+
        geom_point(aes(x=endTime, y=endFlow, color=as.factor(peakTime)), shap
e=1, size=2)+
        geom_text(aes(x=peakTime, y=peakFlow, label="Peak", color=factor(peak
Time)), nudge_y = .05)+
        theme(panel.grid = element_blank(), legend.position = "none")+
        labs(x="Time", y="Quickflow (cu m/s)"))


#OUTPUT
# write_csv(stormEvents, "Data/stormsBG.csv")
# write_csv(stormEvents, "Data/stormsBT.csv")
# write_csv(stormEvents, "Data/stormsRR.csv")
# write_csv(stormEvents, "Data/stormsTH.csv")
```

Data Concatenation

```
#Pulls in flow and specific conductance data for each storm; outputs result i
n csv
#Will Hamilton
#wamilton@ad.unc.edu
#February 18, 2019

library(pacman)
p_load(tidyverse, EcoHydRology, magrittr, animation,
       lubridate, peakPick, depmixS4, colorspace, RColorBrewer)

#determines approximate season given month
season <- function(month){
  if(month >= 3 && month <= 5){
    return("spring")
  }
  if(month >= 6 && month <= 8){
    return("summer")
  }
  if(month >= 9 && month <= 11){
    return("fall")
```

```
  }
  return("winter")
}

#BG
{
  stormsBG <- read_csv("Data/stormsBG.csv")
  flowData <- read.csv("Data/BG_10_19_18.csv")
  flowData$TIMESTAMP <- ymd_hms(flowData$TIMESTAMP)
  storms <- stormsBG
  initials <- "BG"
  dataEC <- read.csv("Data/BG_WL_EC_2019-2-13_DLScor.csv")
  dataEC$Date_Time <- ymd_hms(dataEC$Date_Time)
  num.time <- as.numeric(as.POSIXct(dataEC$Date_Time, tz="UTC"))
  winterCond <- 500
  #remove specific conductivity outliers
  remEC <- c(81467)
  dataEC <- dataEC[-remEC,]
}

#BT
{
  stormsBT <- read_csv("Data/stormsBT.csv")
  flowData <- read.csv("Data/BT_10_19_18.csv")
  flowData$TIMESTAMP <- ymd_hms(flowData$TIMESTAMP)
  storms <- stormsBT
  initials <- "BT"
  dataEC <- read.csv("Data/BT_WL_EC_2019-2-13_DLScor.csv")
  dataEC$Date_Time <- ymd_hms(dataEC$Date_Time)
  num.time <- as.numeric(as.POSIXct(dataEC$Date_Time, tz="UTC"))
  winterCond <- 600
  #remove flow outliers
  remFlow <- c(38218, 63895, 86525, 47123, 65523)
  flowData <- flowData[-remFlow,]
  #remove specific conductivity outliers
  remEC <- c(79163, 96321, 96467)
  dataEC <- dataEC[-remEC,]
}

#RR
{
  stormsRR <- read_csv("Data/stormsRR.csv")
  flowData <- read.csv("Data/RR_10_19_18.csv")
  flowData$TIMESTAMP <- ymd_hms(flowData$TIMESTAMP)
  storms <- stormsRR
  initials <- "RR"
  dataEC <- read.csv("Data/RR_WL_EC_2019-2-13_DLScor.csv")
  dataEC$Date_Time <- ymd_hms(dataEC$Date_Time)
  num.time <- as.numeric(as.POSIXct(dataEC$Date_Time, tz="UTC"))
  winterCond <- 350
```

```
}

#TH
{
  stormsTH <- read_csv("Data/stormsTH.csv")
  flowData <- read.csv("Data/TH_12_9_18.csv")
  flowData$TIMESTAMP <- ymd_hms(flowData$TIMESTAMP)
  storms <- stormsTH
  initials <- "TH"
  dataEC <- read.csv("Data/TH_WL_EC_2019-2-13_DLScor.csv")
  dataEC$Date_Time <- ymd_hms(dataEC$Date_Time)
  num.time <- as.numeric(as.POSIXct(dataEC$Date_Time, tz="UTC"))
  winterCond <- 275
}


#PLOT ALL STORMS

{
  IDX <- 1
  allStorms <- c()
  allStorms$flow <- c()
  allStorms$specCond <- c()
  allStorms$TIMESTAMP <- c()
  allStorms$Num <- c()
  allStorms$season <- c()
  allStorms <- as.data.frame(allStorms)
  thisStorm <- c()
}

while(IDX<= length(storms$peakFlow)){
  hyst <- c()
  #obtain flow data and timestamps
  hyst$flowTIME <- flowData$TIMESTAMP[match(storms$startTime[IDX], flowData$T
IMESTAMP):match(storms$endTime[IDX], flowData$TIMESTAMP)]
  hyst$flowTIME <- unique(hyst$flowTIME)#get rid of duplicates
  hyst$flow <- flowData$Q[match(hyst$flowTIME, flowData$TIMESTAMP)]

  #obtain conductivity data and timestamps
  hyst$condTIME <- dataEC$Date_Time[match(storms$startTime[IDX], dataEC$Date_
Time):match(storms$endTime[IDX], dataEC$Date_Time)]
  hyst$condTIME <- unique(hyst$condTIME)#get rid of duplicates
  hyst$specCond <- dataEC$EC_SpecCond[match(hyst$condTIME, dataEC$Date_Time)]

  #in case timestamps differ (only takes data where the timestamps are the sa
me)
  intersect <- intersect(hyst$flowTIME, hyst$condTIME)
  hyst$flow <- hyst$flow[hyst$flowTIME %in% intersect]
  hyst$flowTIME <- hyst$flowTIME[hyst$flowTIME %in% intersect]
```

```
  hyst$specCond <- hyst$specCond[hyst$condTIME %in% intersect]

  hyst$Num[1:length(hyst$specCond)] <- IDX

  #remove winter salt storms
  if(max(hyst$specCond, na.rm = TRUE) > winterCond){
    print(paste("removed storm:", as.character(IDX), " at month:", as.charact
er(month(hyst$flowTIME[1])), sep= " "))
    IDX <- IDX + 1
    next
  }

  #obtain desired data from this storm
  thisStorm$flow <- hyst$flow
  thisStorm$specCond <- hyst$specCond
  thisStorm$TIMESTAMP <- hyst$flowTIME
  thisStorm$Num <- thisStorm$flow
  thisStorm$season <- thisStorm$flow
  thisStorm$Num[1:length(thisStorm$Num)] <- IDX
  thisStorm$season[1:length(thisStorm$season)] <- season(as.numeric(month(hys
t$flowTIME[1])))

  #concatenate into larger dataframe with all storms
  thisStorm <- as.data.frame(thisStorm)
  allStorms <- rbind(allStorms, thisStorm)
  thisStorm <- as.list(thisStorm)

   IDX <- IDX+1
}


write_csv(allStorms, paste("Data/", "combined", initials, ".csv", sep=""))

#Combines site storm data and outputs combined dataframe as csv
#Will Hamilton
#wamilton@ad.unc.edu
#February 18, 2019

library(pacman)
p_load(tidyverse, EcoHydRology, magrittr, animation,
       lubridate, peakPick, depmixS4, colorspace, RColorBrewer)


stormsBG <- read_csv("Data/combinedBG.csv")
stormsBG$site <- "Burlage"


stormsBT <- read_csv("Data/combinedBT.csv")
stormsBT$site <- "Booker Trib"
```

```
stormsRR <- read_csv("Data/combinedRR.csv")
stormsRR$site <- "Rogers Road"


stormsTH <- read_csv("Data/combinedTH.csv")
stormsTH$site <- "Tallyhoe"

allSites <- rbind(stormsBG, stormsBT, stormsRR, stormsTH)

write_csv(allSites, "Data/allSiteStorms.csv")
```

Data Visualization and Cleaning

```
#Plots flow and specific conductance of different storms
#Will Hamilton
#wamilton@ad.unc.edu
#February 18, 2019

library(pacman)
p_load(tidyverse, EcoHydRology, magrittr,
       lubridate, peakPick, depmixS4, colorspace, RColorBrewer)

#BG
allStorms <- read_csv("Data/combinedBG.csv")
site <- "Burlage"

#BT
allStorms <- read_csv("Data/combinedBT.csv")
site <- "BookerTrib"

#RR
allStorms <- read_csv("Data/combinedRR.csv")
site <- "RogersRoad"

#TH
allStorms <- read_csv("Data/combinedTH.csv")
site <- "TallyHo"


#plot flow vs spec cond hystereses by season
{
  seasonalStorms <- allStorms %>% filter(season == "summer")
  stormCount <- length(unique(seasonalStorms$Num))
  getPalette <- colorRampPalette(brewer.pal(8, "Accent"))
  g <- ggplot(seasonalStorms, mapping = aes(x=flow, y=specCond, color=as.fact
or(Num))) +
    geom_point(size=3) +
    scale_color_manual(values=getPalette(stormCount))+
    ggtitle(paste(site, "storm number:", "\n", "month:", as.character(month(a
```

```
llStorms$TIMESTAMP[1])), sep=" ")) +
    labs(x="Flow", y ="Specific Conductance")+
    facet_wrap(~season, nrow=2)
  plot(g)
}

#plot hysteresis of single storm
{
  singleStorm <- allStorms %>% filter(Num == 44)
  g <- ggplot(singleStorm) +
    scale_colour_gradient(low="blue", high="yellow")+
    geom_point(aes(x=flow, y = specCond, color=TIMESTAMP), size=3) +
    ggtitle(paste(site, "storm", "\n", "month:", as.character(month(singleSto
rm$TIMESTAMP[1])), sep=" ")) +
    labs(x="Flow", y ="Specific Conductance")
  plot(g)
}



#spec cond and flow plotted separately for that storm
{
  g2 <- ggplot(singleStorm) +
    geom_line(aes(x=TIMESTAMP, y=specCond, color = "Specific Conductance"))+
    geom_line(aes(x=TIMESTAMP, y=flow*mean(specCond, na.rm=T)*2, color = "Flo
w"))+
    scale_y_continuous(sec.axis = sec_axis(~./(2*mean(singleStorm$specCond, n
a.rm=T)), name="Flow"))+
    ggtitle(paste(site, "storm number:", "\n", "month:", as.character(month(s
ingleStorm$TIMESTAMP[1])), sep=" ")) +
    labs(x="Time", y ="Specific Conductance")
  plot(g2)
}


#all storms at once
{
  g <- ggplot(allStorms, mapping = aes(x=flow, y=specCond, color=Num)) +
    geom_point(size=3) +
    scale_color_gradient(low="blue", high="yellow")+
    ggtitle(paste(site, "storm number:", "\n", "month:", as.character(month(a
llStorms$TIMESTAMP[1])), sep=" ")) +
    labs(x="Flow", y ="Specific Conductance")+
    facet_wrap(~season, nrow=2)
  plot(g)
}

#Finds index of outlier points to be deleted
#once a point is found, write the index in remove list in EC_Data.R
```

```r
#Will Hamilton
#wamilton@ad.unc.edu
#February 18, 2019

library(pacman)
p_load(tidyverse, EcoHydRology, magrittr, pracma,
       lubridate, peakPick, depmixS4, colorspace, RColorBrewer)

#choose site
flowData <- read.csv("Data/BG_10_19_18.csv")
# flowData <- read.csv("Data/BT_10_19_18.csv")
# flowData <- read.csv("Data/RR_10_19_18.csv")
# flowData <- read.csv("Data/TH_12_9_18.csv")

#**look for outliers visually in dataVis.R**

#flow outliers
View(singleStorm) #create singleStorm in dataVis.R
match(singleStorm$flow[88], flowData$Q)

#EC outliers
View(singleStorm)
match(singleStorm$specCond[13], dataEC$EC_SpecCond)
```

Result Statistics

```r
#Calculates and plots statistics of all storms for all sites
#Will Hamilton
#wamilton@ad.unc.edu
#February 18, 2019

library(pacman)
p_load(tidyverse, EcoHydRology, magrittr, pracma,
       lubridate, peakPick, depmixS4, colorspace, RColorBrewer)

#get rid rows with na flow and spec cond values
{
  allSites <- read_csv("Data/allSiteStorms.csv")
  allSites <- allSites[!is.na(allSites$flow),]
  allSites <- allSites[!is.na(allSites$specCond),]
}

#Finds number of storms at each site and the average duration
#(in points with each point being 5 minutes) of storms at each site
stormLength <- allSites %>% group_by(site, season, Num) %>%
  summarize(duration = length(Num))
print(ggplot(stormLength, aes(x=Num, y=duration/12, color=season))+
        ggtitle("Duration of Storms by Site")+
        labs(y="Duration(hours)", x="Storm Index Number")+
        geom_point(size=3)+
```

```
        facet_wrap(~site, nrow=2))
print(allSites %>% group_by(site, Num) %>%
  summarize(duration = length(Num)) %>%
  summarize(avgDuration = mean(duration), numStorms = length(unique(Num))))


#Determine and plot maximum specific conductance of each storm
peakSC <- allSites %>% group_by(site, season, Num) %>%
  summarize(maxSC = max(specCond, na.rm=T))
print(ggplot(peakSC, aes(x=Num, y=maxSC, color=season))+
        ggtitle("Maximum Specific Conductance for Storms by Site")+
        labs(y=expression(paste("Maximum Specific Conductance (", mu, "S/cm)"
, sep=""")), x="Storm Index Number")+
        geom_point(size=2)+
        facet_wrap(~site, nrow=2)+
        scale_color_discrete(name="Season")+
        theme(panel.grid = element_blank(),
        plot.background = element_rect(fill = "lightcyan1"), legend.backgroun
d = element_rect(fill="lightcyan1")))

#Outlier storm comparison for maximum specific conductance
peakSC <- allSites %>% group_by(site, season, Num) %>%
  summarize(maxSC = max(specCond, na.rm=T))
print(ggplot(peakSC, aes(x=Num, y=maxSC, color=season))+
        ggtitle("Maximum Specific Conductance for Storms by Site")+
        labs(y=expression(paste("Maximum Specific Conductance (", mu, "S/cm)"
, sep=""")), x="Storm Index Number")+
        geom_point(size=2)+
        ylim(c(0,3000))+
        facet_wrap(~site, nrow=2)+
        scale_color_discrete(name="Season")+
        theme(panel.grid = element_blank()))
{
  allWinterSites <- read_csv("Data/allSiteWinterStorms.csv")
  allWinterSites <- allWinterSites[!is.na(allWinterSites$flow),]
  allWinterSites <- allWinterSites[!is.na(allWinterSites$specCond),]
}
peakWinterSC <- allWinterSites %>% group_by(site, season, Num) %>%
  summarize(maxSC = max(specCond, na.rm=T))
print(ggplot(peakWinterSC, aes(x=Num, y=maxSC, color=season))+
        ggtitle("Outlier Storms Included")+
        labs(y=expression(paste("Maximum Specific Conductance (", mu, "S/cm)"
, sep=""")), x="Storm Index Number")+
        geom_point(size=2)+
        ylim(c(0,3000))+
        facet_wrap(~site, nrow=2)+
        scale_color_discrete(name="Season")+
        theme(panel.grid = element_blank()))
```

```r
#Determine direction, flashiness, and magnitude of hysteresis for each storm

#takes two SC values and outputs direction of hysteresis depending on which i
s greater
scBehav <- function(firstSC, secondSC){
    if(is.na(firstSC) || is.na(secondSC)){return(NA_character_)} #don't want
to consider NAs
    if(firstSC > secondSC){return("Enrichment")} #min first (these are enrich
ing)
    if(firstSC < secondSC){return("Dilution")} #max comes first (dilution)
    else{return(0)} #this should never happen
}

#intakes values and values of each quartile. Outputs quartile that value fall
inside
quartiles <- function(value, fstQuart, sndQuart, thrdQuart, fthQuart){
  if(is.na(value)){return(NA_character_)}
  if(value <= fstQuart){return("First")}
  if(value <= sndQuart){return("Second")}
  if(value <= thrdQuart){return("Third")}
  if(value <= fthQuart){return("Fourth")}
}

thirdQuartile <- allSites %>% group_by(site,Num, season) %>%
  summarize(startTime = TIMESTAMP[1],
            endTime = TIMESTAMP[length(TIMESTAMP)],
            maxSC = max(specCond, na.rm=T),
            maxSCTime = TIMESTAMP[match(maxSC, specCond)],
            minSC = min(specCond, na.rm=T),
            minSCTime = TIMESTAMP[match(minSC, specCond)],
            totQ = trapz(as.numeric(TIMESTAMP)/1000, flow), #finds total flow
            maxQ = max(flow, na.rm=T),
            flash = maxQ/totQ, #determine flashiness
            maxQTime = TIMESTAMP[match(maxQ, flow)],
            minQ = min(flow, na.rm=T),
            minQTime = TIMESTAMP[match(minQ, flow)],
            width =(maxQ-minQ)/(maxQ), # normalize width
            height = (maxSC-minSC)/(maxSC), #normalize height
            thirdQuartTIME = round_date(quantile(TIMESTAMP, .75), "5 minutes"
), #start time of last quarter of storm
            thirdQuartSC = specCond[TIMESTAMP == thirdQuartTIME], #SC at the
beginning of the last quarter of the storm
            startSC = specCond[1], #first SC measurement of each storm
            Direction = scBehav(thirdQuartSC, startSC)) #if start SC is great
er then marked as dilution

#determine quartile of flashiness of each storm relative to other storms at t
hat site
```

```
{
  thirdQuartile <- thirdQuartile %>% group_by(site) %>%
    mutate(fstQuart = quantile(flash, .25, na.rm = T),
           sndQuart = quantile(flash, .50, na.rm = T),
           thrdQuart = quantile(flash, .75, na.rm = T),
           fthQuart = quantile(flash, 1.0, na.rm = T),
           rSquared = summary(lm(height~width))$adj.r.squared)
  thirdQuartile <- thirdQuartile %>% group_by(site, Num) %>%
    mutate(Quartile = quartiles(flash, fstQuart, sndQuart, thrdQuart, fthQuar
t))
  thirdQuartile$Quartile <- factor(thirdQuartile$Quartile, c("First", "Second
", "Third", "Fourth"))
  }



#normalized magnitude with coloring by direction
print(ggplot(thirdQuartile, aes(x=width, y=height, color=Direction))+
        ggtitle("Normalized Storm Magnitude Grouped by SC Response")+
        labs(x=expression((Q[max]-Q[min])/Q[max]), y=expression((SC[max]-SC[m
in])/SC[max]))+
        geom_point(size=2)+
        facet_wrap(~site, nrow=2)+
        geom_smooth(method="lm", aes(x=width, y=height, color=NULL), color="b
lack", linetype="dashed", size=.6)+
        theme(panel.grid = element_blank()))

#normalized magnitude with coloring by flashiness
print(ggplot(thirdQuartile, aes(x=width, y=height, color=Quartile))+
        ggtitle("Normalized Storm Magnitude Grouped by Flashiness")+
        labs(x=expression((Q[max]-Q[min])/Q[max]), y=expression((SC[max]-SC[m
in])/SC[max]))+
        geom_point(size=2)+
        facet_wrap(~site, nrow=2)+
        geom_smooth(method="lm", aes(x=width, y=height, color=NULL), color="b
lack", linetype="dashed", size=.6)+
        theme(panel.grid = element_blank()))

#flash vs SC mag
print(ggplot(thirdQuartile, aes(x=Quartile, y=height, color=Quartile))+
        ggtitle("Specific Conductance Magnitude Grouped by Flashiness")+
        labs(x="Relative Flashiness", y=expression((SC[max]-SC[min])/SC[max])
)+
        geom_boxplot()+
        facet_wrap(~site, nrow=2)+
        theme(panel.grid = element_blank()))

#flash vs flow mag
print(ggplot(thirdQuartile, aes(x=Quartile, y=width, color=Quartile))+
```

```
        ggtitle("Flow Magnitude Grouped by Flashiness")+
        labs(x="Relative Flashiness", y=expression((Q[max]-Q[min])/Q[max]))+
        geom_boxplot()+
        facet_wrap(~site, nrow=2)+
        theme(panel.grid = element_blank()))
```