On the Nature of Freshwater Flow throughout the Woodville Karst Plain

Livia Nason and Seth Bassett

April 24, 2018

Abstract

Karst aquifers exhibit complex relationships between stored groundwater, infiltration of precipitation, and spring flow. The exact relationship between precipitation and conduit flow, as well as the interconnectedness of the springs in the Woodville Karst Plain both remain unknown and advancing our understanding of these phenomena is the goal of this paper. We analyze available flow velocity data collected from three Florida springs as well as precipitation data collected in the Woodville Karst Plain over three years from 2011 to 2013. The analysis is conducted using clustering methods and various correlation tests in R and is presented in the form of plots using R, while other results are mapped using ArcGIS. We discover certain periods during which conduit flow is highly affected by rainfall, and others where rainfall does not affect flow at all. We study these periods and present a set of correlations between measured flow parameters and precipitation which may indicate that under certain conditions, rainfall may be positively correlated with conduit flow. We discover certain drainage basins within the study area that are more strongly correlated with flow data than others. We hope to identify these basins and possibly discover a new mechanism that drives velocity variations. This new mechanism may offer a way to better study freshwater flow throughout the Floridan Aquifer.

1 Introduction

The Floridan Aquifer System is home to more than 700 discharging springs; 33 of which are first magnitude springs which discharge at least 2.83 m^3 of freshwater per second (Scott 2004). The development of the Floridan Aquifer has led to the formation of underground conduit systems that stretch from inland springs southward to the coastline where submarine springs are present (Davis 2013). The connectedness of the inland freshwater systems to the submarine springs is a cause for concern since salty seawater often finds its way into our freshwater resources. According to the Florida Geological Survey (FGS), mixing less than 1% of saltwater by volume with freshwater supplies can render public drinking water non-potable. Because the freshwater spring and conduit systems are so incredibly sensitive to the impacts of precipitation and seawater on the system, it is important to study and model these effects to better understand the "complex mechanics" of our freshwater systems (Xu 2016).

One way of studying the flow of freshwater through underground conduits is though dye tracing. Dye tracing is a method of detecting and connecting karst windows, and has been used by geologists for decades (Wilson 1967). Groundwater tracing is among the oldest diagnostic methods and is considered the most accurate of methods for studying karst hydrogeology (Kresic 2013). This method of tracking groundwater flows involves injecting fluorescent liquid into a body of water to determine either how long it will take to travel to a specific destination or how far it can travel in a specific amount of time. Dye tracing was first used with a modern fluorometer, which accurately measures fluorescent dye concentrations, in 1962 in the Tittabawassee River by the University of Michigan (Wilson 1967). The method had also been used around the same time by the USGS in nearly 30 states across the country and the procedures have hardly changed since then (Wilson 1967). The amount of dye used in these experiments can range from a few grams to a few thousand kilograms depending on the conditions of the experiment such as flow velocity, waterbody discharge, and expected travel distance. The amount of dye to be used is computed to minimize coloration of the stream and to minimize cost of the materials, but to still produce a detectable result at each destination.

However, dye traces are still expensive due to the cost of dye, personnel time, and laboratory testing needed. One possible alternative is to utilize NEXRAD data and conduit flow meters to establish correlations between precipitation and changes in spring flow. This paper explores this potential new method in an attempt to assess whether or not it is a viable method for establishing connections within a karst aquifer.

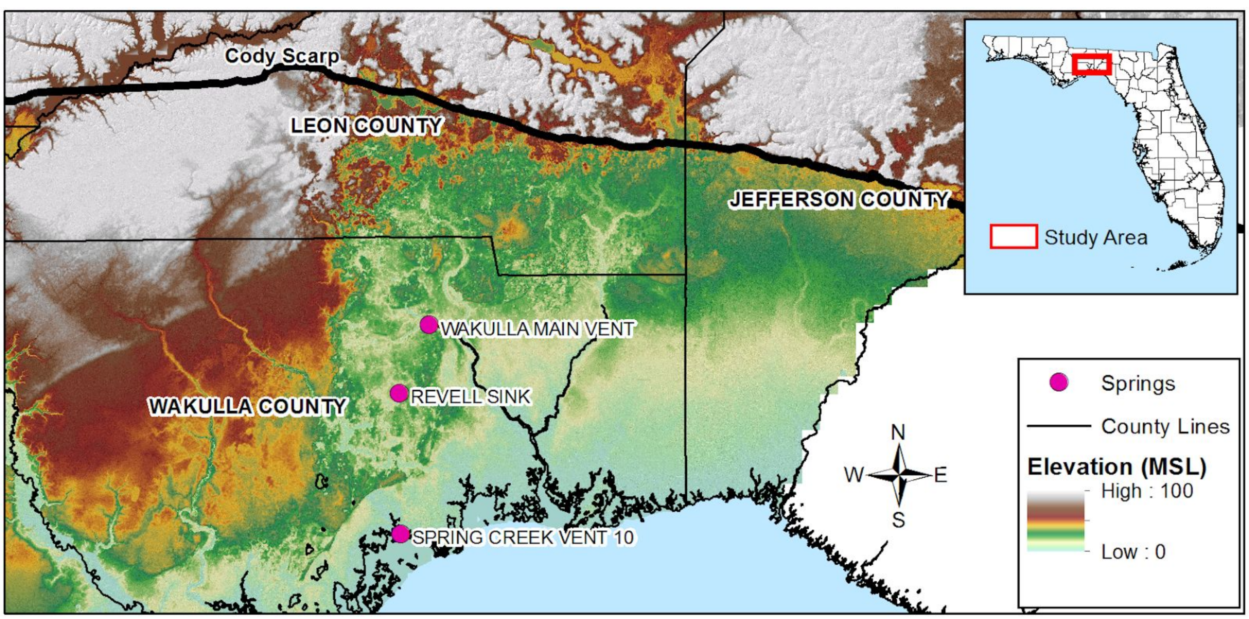

Figure 1. The map of the study area. The region of interests includes several counties in the Florida Panhandle, including the drainage basins of the Ochlockonee, Wakulla, and St. Marks Rivers. The red dots mark the approximate locations of the springs considered in this work. The study area with respect to Florida is shown in the inset in the top right section of the figure, outlined in red.

1.1 Study Area

The study area of this project is a portion of the North Florida panhandle that contains the Woodville Karst Plain (WKP), reaches around the coast of the Big Bend, and stretches into Southern Georgia's Red Hills Region, including the Cody Scarp (Figure 1). The Cody Scarp is an escarpment that cuts across the Florida Panhandle and is a remnant shoreline of the Gulf of Mexico. The scarp separates the Gulf Coast Lowlands from Florida's Northern Highlands, which have maximum elevations of 24 meters and 70 meters, respectively. The Red Hills is a region of Southern Georgia and the Florida Panhandle that lies just north of the Cody Scarp acting as the northern boundary of the Woodville Karst Plain. The Red Hills area is one of the highest recharge areas for the Florida Aquifer system consisting of elevated confining red clay hills.

The Woodville Karst Plain is a highly karstified landscape with a surface layer of sand about 20 feet in depth overlaying a porous limestone that resembles the structure of Swiss cheese. Water can easily navigate through the unconsolidated limestone, dissolving the rock layer and eventually sculpting out underground rivers called conduits. The underground conduits surface at springs, which discharge groundwater out of their spring vents at varying flow rates.

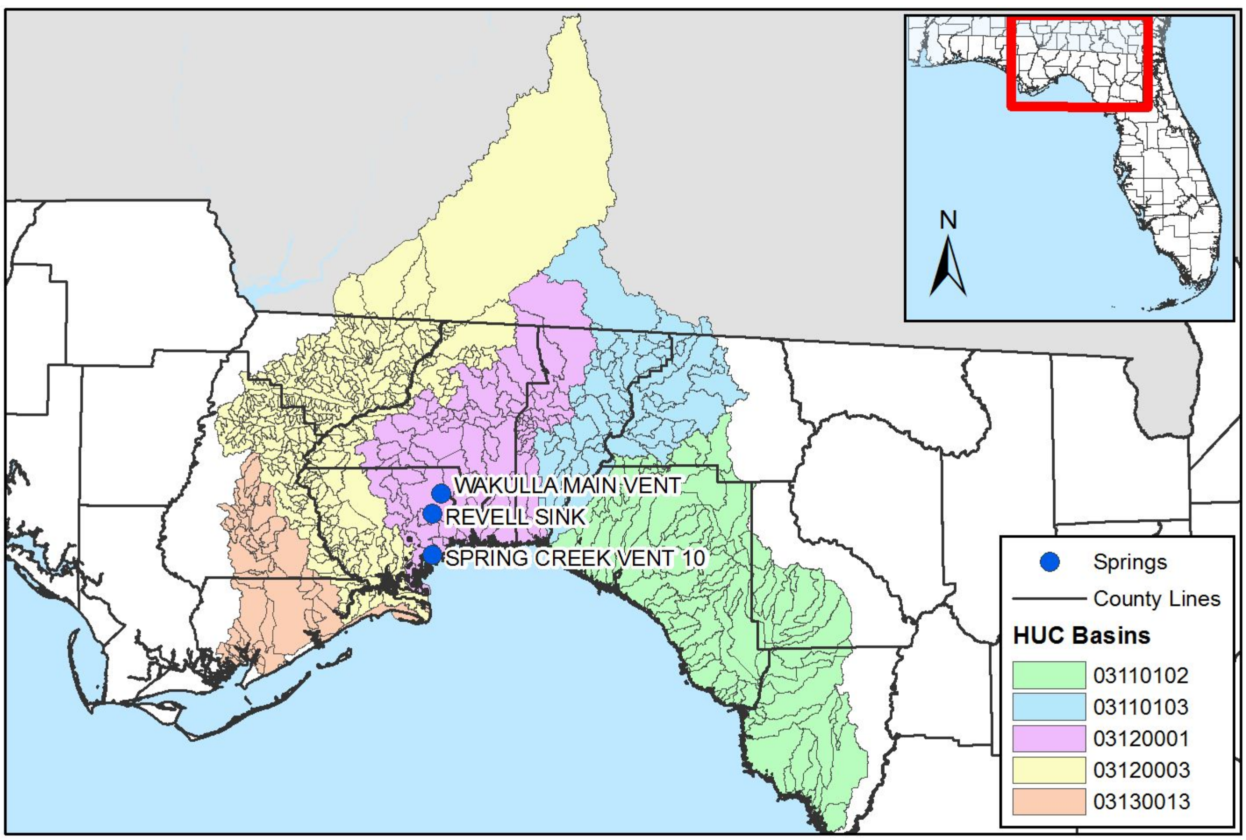

Figure 2. The drainage basins and the watersheds in the study area. The HUC watersheds of interest, 03120001 and 03120003, are shaded in pink and yellow. Locations of springs of importance to the present study are marked with blue dots.

A first magnitude spring discharges more than 100 cubic feet of water per second, while second magnitude springs discharge anywhere from 10-100 cubic feet of water per second.

This project is focused on the flows of two first magnitude springs, Wakulla Spring and Spring Creek Springs, and a karst window located between the two springs called Revell Sink. To study how rainfall affects each of these springs, we take a look at the specific watersheds or drainage basins whose collected rainwater has an effect on these sources. Watersheds are synonymous with hydrologic units which are classified by the US Geological Survey (USGS) into hierarchical Hydrologic Unit Codes (HUC). In this study we focus on the basins within the hydrologic units that include the WKP, classified as HUCs 03120001 and 03120003. The two 8-digit HUC watersheds are split up further into drainage basins for each body of water in the study area (Figure 2). Using the size and shape of each of 321 individual basins, the total amount of rainfall that falls into each basin, based on weather radar data, can be calculated and used to determine the impact of each basin on spring discharge. The study area defined by the basins has a total area of about 3663 square miles.

1.1.1 Wakulla Spring

The Wakulla Spring vent is located 14 miles south of Tallahassee, Florida in the Ed Ball Wakulla Spring State Park (Figure 3a) (Stone 1989). The main vent is 300 feet in diameter and discharges an average of 400 cubic feet of water per second, feeding the Wakulla River which flows southward into the Gulf of Mexico (Stone 1989). The Wakulla Springs area was first acquired in the Forbes Purchase of 1811, which was the purchase of 1.2 million acres of land in the Florida Panhandle from the local Native Americans by the Spanish.

The state park itself has a complex ownership history, but has been retained by only six owners from 1860 to 1934, when it was purchased by the Wakulla Springs Development Corporation directed by Edward Ball (Stone 1993). In 1930, multiple mastodon bones were discovered within the spring which permitted Dr. Herman Gunter of the FGS to eventually excavate a full skeleton. In 1955, deep cave divers struggled to be allowed to dive deep into the spring, but upon discovering more mastodon bones, the Ball management endorsed further exploration which has led to the conduit system being one of the most well explored underground cave systems in the world.

1.1.2 Spring Creek Springs Complex

Spring Creek Springs is a system of up to 14 individual submarine springs located in the Gulf of Mexico on the southern border of Wakulla County, Florida (Davis 2013). From 1972-1974, the US Geological Survey collected initial water quality data on the first 8 springs discovered in the Spring Creek group, named springs 1-8. Later, in 1997, the Florida Geological Survey returned to the site to collect additional background data, and discovered 3 more springs, labeled springs

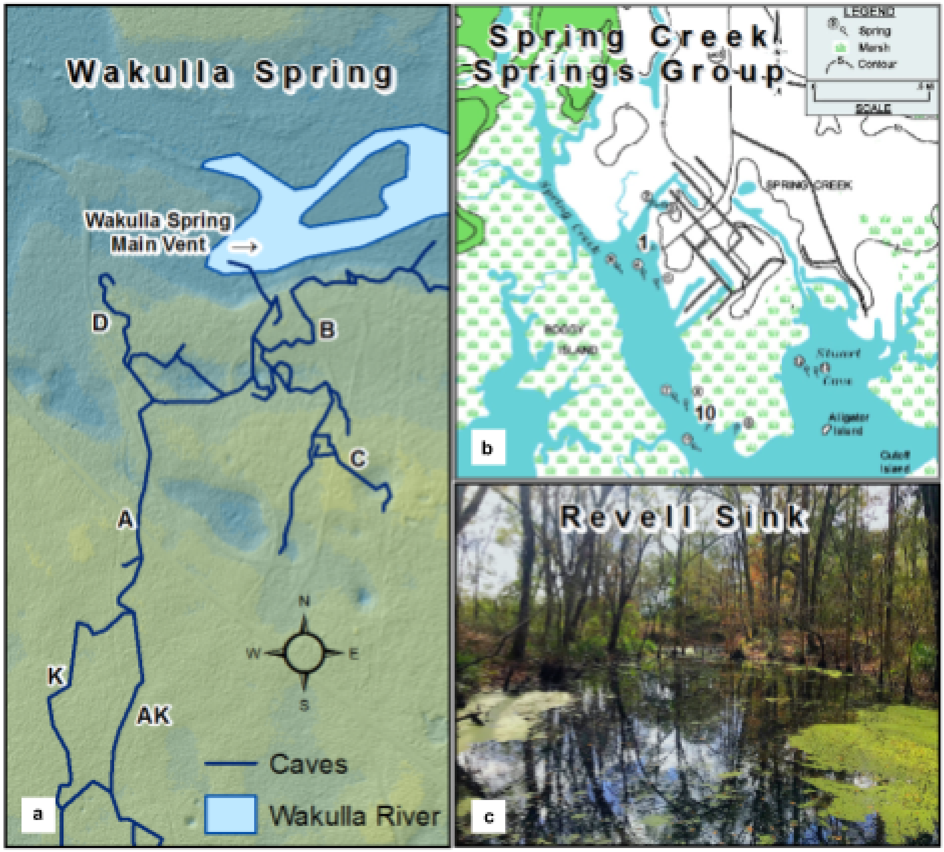

Figure 3. The locations of the conduits, vents, and the general appearance of sources used in the present study. (**a**) A map of the Wakulla Spring system. The conduits are shown with blue lines and are marked with individual labels. The location of the main spring vent, and the direction of flow, that feeds the Wakulla River is identified with an arrow. (**b**) A map of the Spring Creek Springs Group showing the location of the spring's 14 vents, with emphasis on vents 1 and 10. (**c**) A general view of Revell Sink.

9-11. And in 1998 two more springs were located further west and were named springs 12 and 13 (Figure 3b) (Lane 2001).

According to Fleury et al (2007), Spring Creek Springs is a type 2 submarine spring, and is characterized by the following four conditions. First, Spring Creek connects to a well developed karst network of underground conduits that are joined together by vertical tunnels. Next, the conduits of Spring Creek are too large for the volume of freshwater being discharged, which can often result in the inability to prevent seawater from entering the system. Also, the Spring Creek system connects to springs throughout the WKP through well developed conduit networks that drain vast recharge areas and have large freshwater storage capacities. Lastly, Spring Creek has a high average groundwater discharge rate and a strong seasonal variability, as well as a low salinity during periods of high flow and a high salinity during periods of low flow.

1.2 Background

Within the Floridan Aquifer system, conduits act as a series of pipes hundreds of feet below the land surface of Florida. The conduits branch out from each other in various places and at different angles allowing the springshed to extend past the Florida-Georgia border. With a springshed this large, an average of 800 cubic feet of groundwater per second (cfs) rush into the springs system (Davis 2013). A long period of low rainfall changes the efficiency of a spring's flow. A long drought period will prevent Spring Creek from being able to discharge freshwater. Because of the low pressure within the spring, seawater rushes into the conduit system at anywhere between 0 - 700 cfs, blocking freshwater from discharging and is known as seawater intrusion. When this happens, the northern Wakulla Spring captures the flow from Spring Creek and continues discharging at around 800 cfs, even during low rainfall periods.

A period of high rainfall has the potential to restore and reestablish the proper flow of the springs system by forcing the heavier, denser saltwater out of the Spring Creek vents.. During a period of such high rainfall, both the Wakulla Spring and Spring Creek systems can discharge up to 3,000 cubic feet of freshwater per second. After a period of high rainfall ends, low rainfall conditions slowly return to the spring system and the discharges at each of the springs begin to diminish. Wakulla Spring lowers its rate of discharge to less than 400 cfs and the Spring Creek system slows down to less than 1000 cfs. As the period of low rainfall lengthens, the system returns to its drought-like conditions. Wakulla Spring will capture the contributing freshwater from Spring Creek at around 800 cfs. The process detailed above occurs multiple times each year. A single large rainfall event has the potential to elevate a spring's discharge for weeks at a time, and multiple consecutive high rainfall events can keep a spring discharging for months.

1.3 Current Approach

The study data includes NEXRAD precipitation data recorded over the course of 3 years (2011-2013), as well as hourly flow velocity data from flow meters located throughout the southern Woodville Karst Plain conduit systems, over the same time period, to detect correlations between the data sets. Other types of data that are used in addition to current flow velocity include average temperature, conductivity, and direction of the flow over the course of this period.

Studies have show that an increase in rainfall is linked to an increase in spring discharge. But certain erratic "spikes" in conduit flow velocity data may or may not be caused solely by rainfall. We will be investigating to determine whether these spikes in average flow speed are a result of various periods of rainfall or another phenomenon altogether. The initial inquiry is based off of a portion of recorded data in which a destructive amount of precipitation drained into the Floridan Aquifer system and an anomaly occurred just a few months afterward. In June of 2012, Tropical Storm Debby dropped an average of 12 inches of rainfall across the Florida Panhandle, with a maximal amount of rainfall at 28 inches in Wakulla County (Kimberlain 2013). The runoff from the storm water caused a surge in flow velocity within the conduit system. When the stormwater had completed draining a few months later in November, the velocity appears to spike up to five times greater than the activity in the spring caused by the tropical storm. This study began as an attempt to determine the cause of these spikes in flow velocity at Spring Creek, but has developed into a more general, yet more detailed, study of the spring system's relationship with heavy rainfall, periods of drought, and everything in between.

To study the relationship between rainfall and flow meter data, we calculate and analyze correlations between the two data sets. Calculating these correlations properly requires that each significant rainfall event lines up with the induced change in flow data appropriately. To achieve this merger between the datasets, the amount of time it takes for rainfall to effect the flow data, or lag time, must be determined and the rainfall data must be shifted accordingly. Because natural data often has a seasonal trend, rainfall does not have the same effect on spring discharge throughout the year. Therefore, it is imperative that small fragments of the study's timeline, periods when the relationship between rainfall and discharge is clear be individually selected, so that periods with weak relationships are not used in critical calculations. Next, correlations are calculated between the individually chosen study periods and lagged rainfall amounts during that period, to confirm that the source data is correlated with the rainfall data. Finally, correlations between the spring data and each individual basin's rainfall are then determined and analyzed. The values of these correlations will be plotted in a GIS application to be visualized so that inferences about the data and analyses of the results can be made.

2 Methods

This section will first discuss the two types of data that were utilized throughout the study, as well as the various statistical methods that are used to discover trends and correlations and determine their significance. Data manipulation and analysis were done using SQL and R, while data visualization was done in Esri's ArcGIS.

2.1 Available Data

Two types of data were used throughout the project to discover previously unknown relationships between precipitation and conduit flow meter data. The first dataset presents rainfall measurements throughout the study area, over the course of three years from 2011-2013. The precipitation data is retrieved from the NOAA's Next Generation Weather Radar (NEXRAD). The data consists of polygons covering the study area (Figure 3); each of which delineate the amount of precipitation that fell within each polygon in inches. The data is sampled in two different ways. The N1P dataset records, at 5 minute intervals, the total amount of precipitation over the previous hour. Because of the small time intervals, analyses can be done in great detail and with a high accuracy. The drawback to using such detailed data is the performance cost. The speed at which calculations and analyses are performed increases

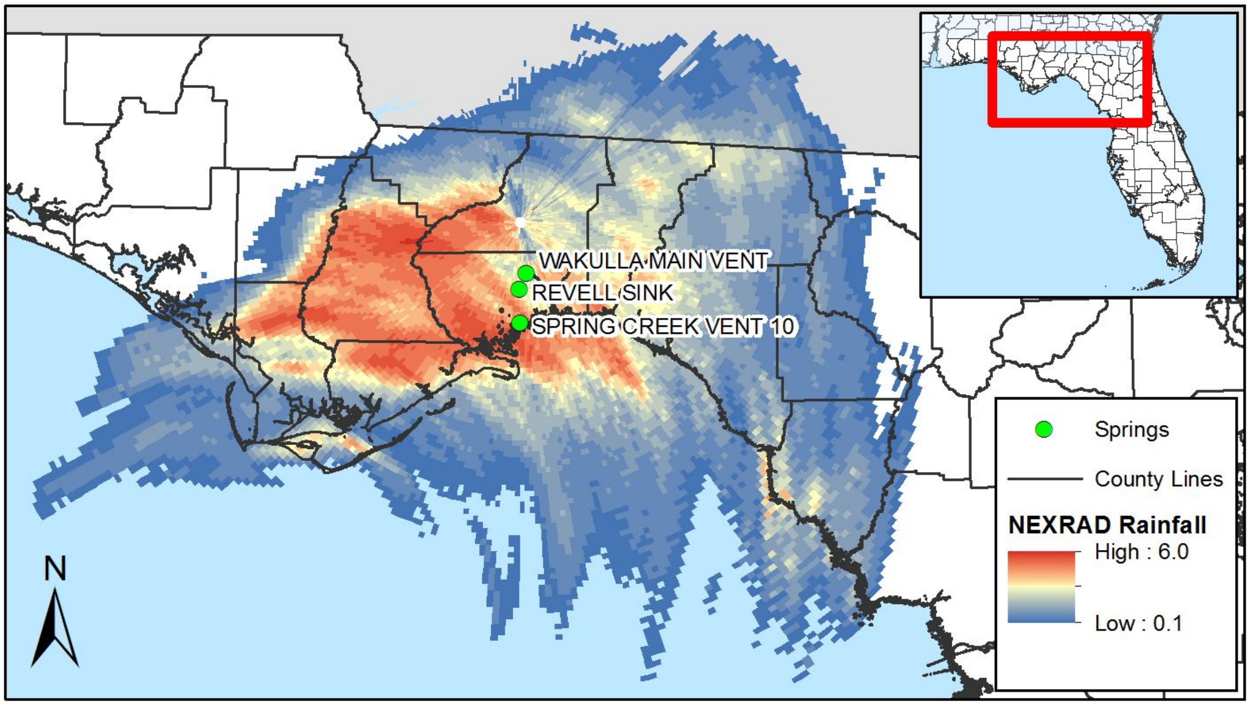

Figure 4. NEXRAD precipitation data for the study area is shown on a scale of blue to red. To study the amount of precipitation per basin, the NEXRAD polygons must be intersected with the basin polygons (Figure 2) to determine how much rain has fallen into each basin. This data, extracted from the NEXRAD dataset, was recorded at 2:11 PM on 25 June 2012 when Tropical Storm Debby dropped 6 inches of rain across the Woodville Karst Plain.

Basin	timestamp	Area [m^2]	Volume [in*m^2]	Rainfall [in]
235	2011-01-01 17:00:00	1.00E+07	8.35E+04	0.008333333
312	2011-01-01 17:00:00	1.02E+07	3.35E+05	0.032865606
489	2011-01-01 17:00:00	5.95E+06	2.25E+05	0.037815189
312	2011-01-01 18:00:00	1.02E+07	4.01E+05	0.039338204
489	2011-01-01 18:00:00	5.95E+06	2.47E+05	0.041449241
1	2011-01-01 19:00:00	3.04E+09	1.59E+04	0.008333333
274	2011-01-01 19:00:00	1.27E+07	1.06E+05	0.008333333
312	2011-01-01 19:00:00	1.02E+07	4.01E+05	0.039338204
489	2011-01-01 19:00:00	5.95E+06	2.47E+05	0.041449241
1	2011-01-01 20:00:00	3.04E+09	3.57E+07	0.011738087

Table 1. A sample from the N3P NEXRAD dataset. The basin column shows the basin number which is collecting rainfall, the timestamp column shows the end of the 3 hour measuring period, while the area and volume variables allowed us to calculate the amount of rainfall into each basin in inches.

substantially with the size of the dataset, so oftentimes a more broad, less detailed dataset is used for more computationally intensive calculations. The N3P data records in hourly increments, the total rainfall summed over the previous three hours. For the purpose of this paper, all flow meter data was averaged or summed to hourly intervals.

The second dataset used in this paper is the Falmouth 2D-ACM sensor data. The raw Falmouth data were originally recorded at 15 minute intervals and thus was unsuitable for correlation analysis with NEXRAD data. To enable such analysis, we averaged the Falmouth data over one hour long intervals. A total of nine Falmouth sensors were installed within selected conduit channels to collect information on the status of nearby springs. The Falmouth sensors collect data regarding: current velocity (cm/s), conductivity (S/m), flow direction (degrees), temperature (degrees Celsius), and pressure (dBar) (Falmouth). Because of the extreme conditions within the conduits, oftentimes, sensor data may feature implausible recorded values or periods of missing data that show when a sensor has failed and needed reparations. Due to the depths at which they are installed, the sensors have been known to record inaccurate measurements of temperature and pressure, or stop recording any parameters altogether until the sensor is repaired.

Falmouth sensors are located in Wakulla Spring conduits AD, AK, K, B, C, and D, as well as within Revell Sink and Spring Creek vents 1 and 10. Although Spring Creek 1 is the main vent at Spring Creek, the Falmouth sensor in this conduit failed at the end of the study's first year (2011), and could not be recovered or repaired (FGS). Instead, Spring Creek vent 10

SiteName	timestamp	Flow Speed [cm/s]	Water Temperature [°C]	Source Conductivity [S/m]	Water Pressure [dBars]	Flow Direction [Degrees]
Spring Creek 10 (Deep)	2011-01-01 0:00:00	10.887	10.8238	31.6748	16.625	224.8401
Spring Creek 10 (Deep)	2011-01-01 1:00:00	11.651	10.7513	31.586	16.7575	229.2233
Spring Creek 10 (Deep)	2011-01-01 2:00:00	10.482	10.7455	31.6093	16.8525	228.4262
Spring Creek 10 (Deep)	2011-01-01 3:00:00	7.888	11.0905	32.1583	16.9275	244.4584
Spring Creek 10 (Deep)	2011-01-01 4:00:00	7.339	11.1158	32.13	16.96	234.8346
Spring Creek 10 (Deep)	2011-01-01 5:00:00	7.503	11.3635	32.6115	16.915	248.6182
Spring Creek 10 (Deep)	2011-01-01 6:00:00	8.027	11.4043	32.8	16.825	251.2683
Spring Creek 10 (Deep)	2011-01-01 7:00:00	7.522	11.3988	32.866	16.62	246.7203
Spring Creek 10 (Deep)	2011-01-01 8:00:00	6.85	11.395	32.8738	16.425	236.128
Spring Creek 10 (Deep)	2011-01-01 9:00:00	6.911	11.407	32.7485	16.205	235.1219

Table 2. A sample of data collected from the 2D-ACM Falmouth Sensor located in Spring Creek vent 10. The table entries show parameters measured at the source for select measuring periods. The data is represented by hourly measurements of the five variables that follow. This sample shows the status of Spring Creek on January 1, 2011. Spring Creek vent 10 had an average speed of 9 cm/s and an average temperature of 11°C. The conductivity of the water is relatively high in the low thirties, and water within the conduit appears to be flowing South-West.

data is used throughout the study to assess the status of Spring Creek. Multiple sensors located within the conduits spreading out from the main Wakulla Spring vent are used in the study to represent different aspects of the conduit system. For example, data recorded in Wakulla conduit D is used primarily to represent the flow of groundwater. Because conduit D has nearly constant values for temperature, flow direction, and speed throughout the study period, we know that its main source of water must be groundwater and that its parameters are only slightly affected by rainfall or drought events. On the other hand, Wakulla conduit AK is highly affected by changes in both rainfall and the flow at Spring Creek, so it is used to represent the discharge of Wakulla Spring into the Wakulla River.

2.2 Analysis Methods

In this paper we are investigating to determine if connecting conduit flow and NEXRAD data can replace dye tracing as a method of showing connections in karst. To verify that two conduits are connected, individual aspects of their flow patterns are investigated and analyzed using various statistical methods, including distributed lag analysis, clustering analysis, and frequency analysis.

2.2.1 Lag Analysis

Because it takes time for rainfall to reach the aquifer system, a lag can be found in the flowmeter time series data. To calculate significant correlations between varying time series data sets, the lag between them needs to be determined and removed from the problem, allowing significant events to line up appropriately and be analyzed properly. Autocorrelation is the correlation between a time series data set and a lagged copy of itself, and is a function of the lag value. Autocorrelation is often used to determine if past values in the time series have an influence on present values or can predict future values, as well as helping to determine whether the data set is stationary (Prabhakaran 2017). A time series that is considered stationary abides by a few conditions including having a constant average value throughout the time series, or having no trend, having a variance that does not increase over time, and has minimal seasonality. A cross correlation function computes the correlation between two time series (Bourke 1996). One of two data sets is shifted past the other, and for each shift of one time unit, a new correlation coefficient is calculated. The correlation coefficient will be highest for the shift, or amount of lag, which best aligns the structures of the two signals. When studying the effect of precipitation on groundwater flow, it is important to take into account the lag between the two signals, since a spike in rainfall data might not optimally line up with the change it causes in the flow data. An example is shown in Figure 5.

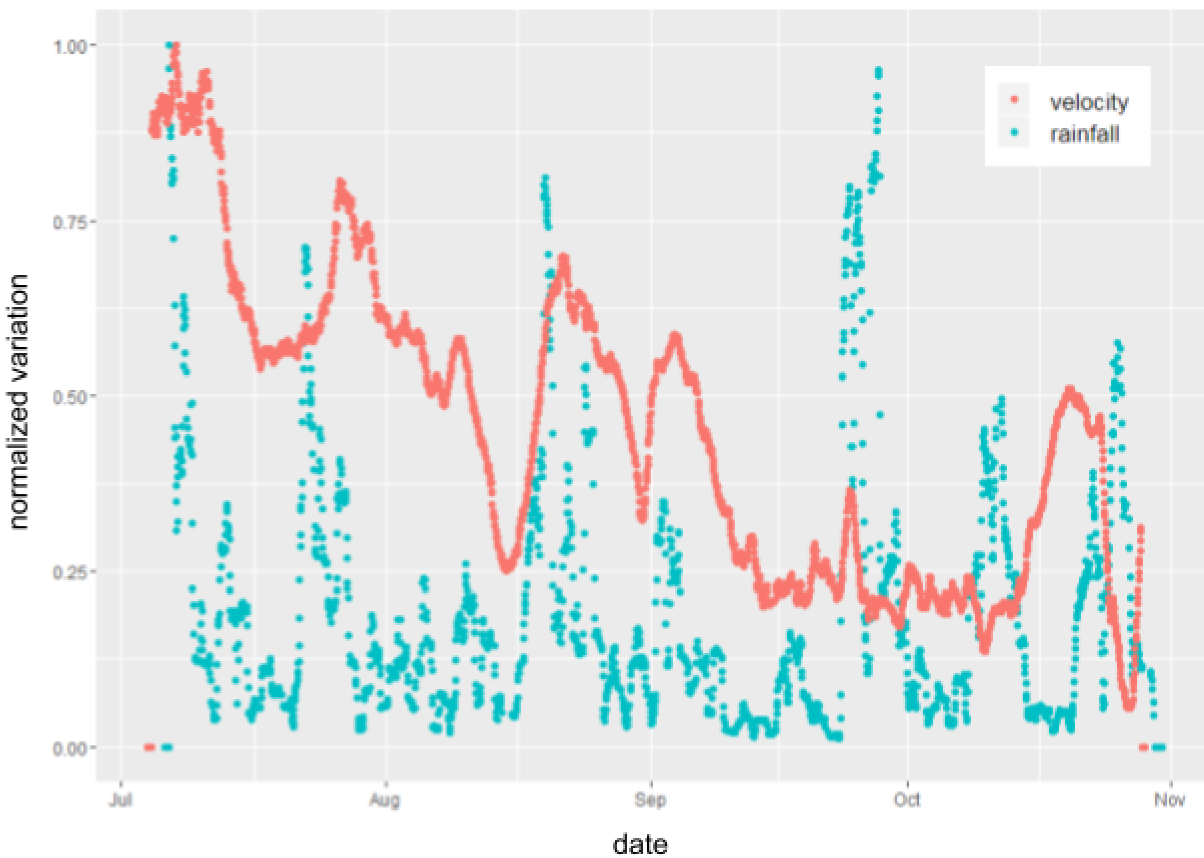

Figure 5. Variation of flow velocity and precipitation for Period 6 in conduit AK. The data illustrates the cause-and-effect relationship between flow velocity and precipitation. The data featured in this plot are measurements of both rainfall and velocity normalized by the range of the data during the measuring period to mitigate the dramatically different scales of the two data sets. In visualizing this data, it becomes obvious that a spike in rainfall is soon followed by a spike in flow velocity within conduit AK at Wakulla Spring.

2.2.2 Multi-state systems and Clustering

A multi-state system is a system that can exist in multiple states and can transition from one state to another, while only existing in one state at a time. Determining the number of states a system can exist in can be challenging but is useful in studying the system in more detail. Once the various states have been determined and classified, instead of studying the entire data set as a whole, the data from the system can be split up using these states and each group can be analyzed individually. One method for identifying a system's various states is using cluster analysis. Clustering is the process of organizing data points into groups where points in the same group are similar in some way and points in different groups are dissimilar in some way (Pang-Ning 2013). The two most common clustering methods are partitional and hierarchical clustering. In hierarchical clustering, similar points create subclusters, and similar subclusters create clusters. These clusters and subclusters are nested groups that can be represented by a tree diagram or dendrogram. Partitional clustering is the division of a data set into non-overlapping clusters where each data point belongs to the center of that cluster. The most common partitional clustering method is the *k*-means algorithm, in which a number *k* of initial centers are selected from the data set at random and are moved to the average position of their closest data points. For each iteration of the algorithm, each clusters' points are averaged and the centers are moved to the average location, and when the cluster centers are no longer updated, the algorithm has converged.

2.2.3 Correlation Analysis

After the data is clustered, we can analyze each state of the system at a time. We attempted to calculate correlations to identify dependencies between the datasets, but discovered that tidal cycles in the Gulf of Mexico were causing swift drastic changes in flow velocity multiple times a day. Because a Spearman correlation algorithm is unable to discern tidal periods, the data must be smoothed to be analyzed in this way. To smooth the effect of high and low tides on flow velocity, a moving average of the data is used. A moving, or rolling, average is calculated by averaging a specific number of past observations. For the purpose of this project, each moving average was calculated using data collected from the previous 24 hours. Upon discovering significant correlations in the data, we began to focus on calculating correlations with respect to individual drainage basins (Figure 2) to determine which basins contribute most effectively to the conduit system.

3 Results

In this section, the results of the study are laid out and discussed in detail. Different aspects of the WKP were analyzed throughout this project and were used to draw conclusions about the status of Florida's springs and their interconnectedness. We hypothesize that there are explicit periods where rainfall has a notable influence on conduit flow, and that within these periods rainfall and flow velocity will be highly correlated while, during other periods, there will be no certain dependence.

3.1 Clustering Results

We applied the clustering algorithm to the flow meter data collected at Spring Creek 10. This choice is motivated by a hypothesis of Davis (2013) who speculated that the speed and direction of the flow in that spring strongly influences not only the flow within the conduit system but also the conditions in other springs of the system.

In clustering, we used conductivity and temperature since those measurements displayed the smallest amount of short term variability. This is illustrated in Figure 6 which shows the run of average speed, conductivity, and temperature for Spring Creek vent 10 over a period from July to November of 2013. It is evident from the figure that conductivity and temperature within the conduits vary more smoothly than velocity.

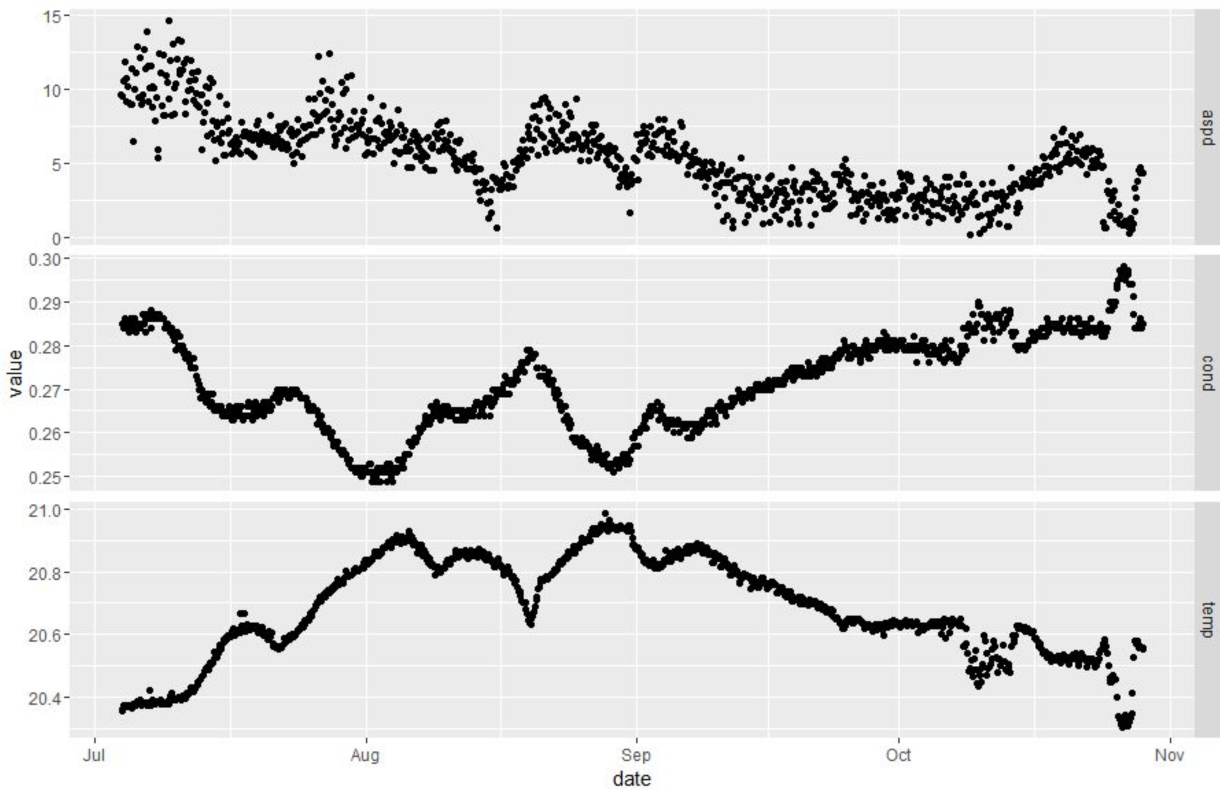

Figure 6. The variability of average speed, conductivity, and temperature at Spring Creek 10 for a select range of dates. Note that the velocity shows a far greater short-term variability than either the conductivity or temperature. This justifies the selection of conductivity and temperature for correlation analysis between rainfall and source flow.

Figure 7 shows the results of the clustering procedure. The *k*-means algorithm converged with a result producing 2 clusters. Cluster one contains 49.8% of the data set and is characterized by low conductivity and an intermediate range of temperature values. Cluster two makes up the remaining 50.2% of the data set and the population is characterized by high conductivity and a wide range of temperature values. Giving the algorithm a value of $k = 3$ clusters would cause the previously second cluster to further split into two clusters. Cluster one is comprised of data points with low conductivity and medium temperatures, cluster two would consist of high conductivity and low temperature values, and cluster three would contain observations with both high conductivity and high temperatures. We concluded that the high conductivity values were clustered separately into high and low temperatures due to seasonal temperature changes, so the data set grouped using $k = 2$ clusters was instead used for analysis.

Using the results from the clustering algorithm, we are able to color the data points of the time series by their clusters to see when the system exists in state 1 and when it exists in state 2, such as in Figure 8. After coloring the plot is is easy to identify the periods that are well suited for analysis. Figure 8 shows the extents of periods 5 and 6 in red, while the segments that are not used in correlation analysis are shown in blue. The clustering algorithm identified a total of 6 periods (Table 3) throughout the duration of the study that can be used to calculate correlations.

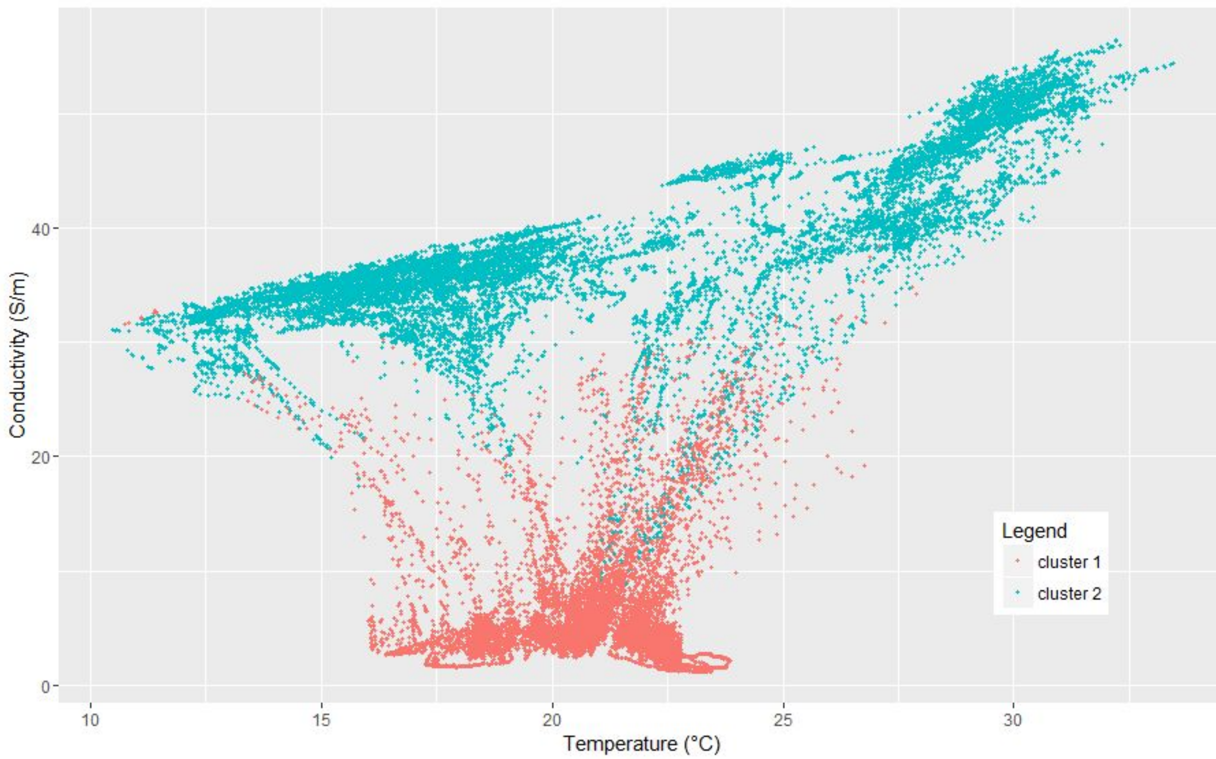

Figure 7. Distribution of conductivity and temperature measurements from Spring Creek 10 analyzed in the present study. The data points are colored according to the results of *k*-means clustering analysis. Two distinct groups of measurements were identified. The first group of measurements, shown in red, is characterized by low conductivity and intermediate temperatures. In contrast, the second group, shown in blue, is characterized by high conductivity and a broad range of temperatures.

Period	StartDate	EndDate	Season
1	2011-01-28	2011-05-18	cold
2	2012-03-17	2012-04-15	cold
3	2012-04-23	2012-05-12	cold
4	2012-06-26	2012-11-02	hot
5	2013-02-27	2013-06-24	med
6	2013-07-04	2013-10-29	hot

Table 3. The start and end dates of the six periods identified by the clustering algorithm for analysis.

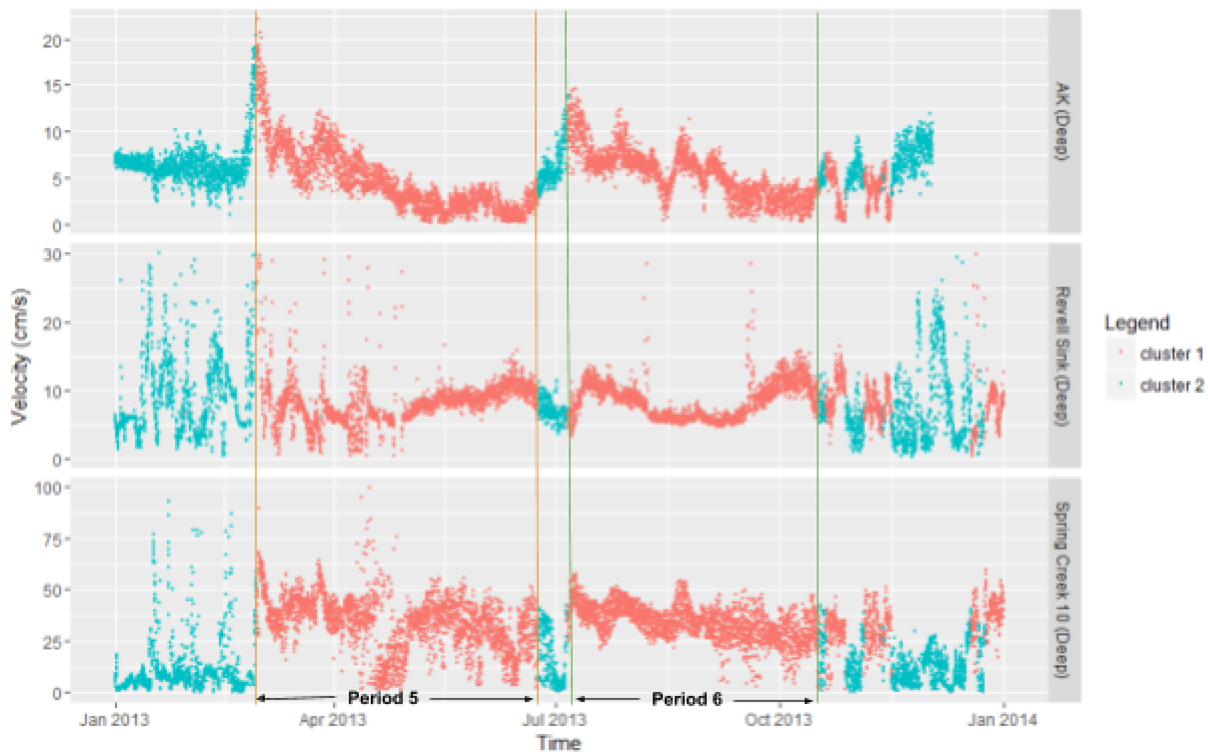

Figure 8. Velocity measurements for Period 5 and Period 6. The data are shown for Wakulla Spring (top panel), Revell Sink (middle panel), and Spring Creek (bottom panel). Cluster 1 data are shown with red points, while Cluster 2 data are colored blue. The range of Period 5 data and Period 6 data are marked with vertical orange and green lines, respectively.

3.2 Global Correlations

Upon splitting the data into analysis periods, we take a look at rainfall data to see how it aligns with changes in conduit flow. Because Wakulla conduit AK has the lowest variability and fewest erratic peaks in velocity, as seen in Figure 8, it is used to calculate the initial correlations. R's Cross-Correlation function (CCF) tells us what amount of lag time maximizes the correlation coefficient between rainfall and each of the three testing variables: velocity, temperature, and conductivity. Looking at Period 1, no significant correlation is found between rainfall and neither temperature nor velocity in AK. Though we do see a significant negative correlation between rainfall and conductivity, -0.36, when rainfall is moved ahead 30 hours. In Period 2, we see two significant correlations between conduit flow and rainfall. Velocity in AK is positively correlated with rainfall, 0.79, at a lag of 28 hours, while temperature is also correlated with rainfall, 0.33, with a lag time of 0 hours. During Period 3 we see significant correlations for each of the three testing variables. Velocity provides a positive correlation of 0.69 with a lag of 10 hours, temperature gives a positive correlation of 0.62 at a lag of 25 hours, and conductivity yields a negative correlation of -0.50 at a lag of 23 hours. Data from Period 4 did not provide any reasonable or significant correlation values. Temperature values in Period 5 are found to have a slight positive correlation with rainfall, 0.31, after an 11 hour lag time. Period 6 velocity appears to be positively correlated with rainfall, 0.67 with no lag time, and temperature values are found to be significantly negatively correlated with rainfall with a 14 hour lag time. The significance of each result was decided based on the validity of the correlation coefficients and the plausibility of the lag time. Additional results are outlined in Table 4.

site	period	x	y	cor	lag	sig
AK	1	rain	aspd	0.523	128	
AK	1	rain	temp	-0.244	82	
AK	1	rain	cond	-0.361	30	1
AK	2	rain	aspd	0.798	28	1
AK	2	rain	temp	0.327	0	1
AK	2	rain	cond	0.619	40	
AK	3	rain	aspd	0.699	10	1
AK	3	rain	temp	0.619	25	1
AK	3	rain	cond	-0.503	23	1
AK	4	rain	aspd	0.308	131	
AK	4	rain	temp	0.084	0	
AK	4	rain	cond	0.496	193	
AK	5	rain	aspd	-0.365	-12	
AK	5	rain	temp	0.306	11	1
AK	5	rain	cond	-0.015	48	
AK	6	rain	aspd	0.676	0	1
AK	6	rain	temp	-0.454	14	1
AK	6	rain	cond	0.069	35	

site	period	x	y	cor	lag	sig
Rev	1	rain	aspd	0.631	68	1
Rev	1	rain	temp	-0.552	24	1
Rev	1	rain	cond	-0.525	24	1
Rev	2	rain	aspd	0.562	0	1
Rev	2	rain	temp	-0.778	52	1
Rev	2	rain	cond	0.794	25	
Rev	3	rain	aspd	-0.344	30	
Rev	3	rain	temp	0.169	21	1
Rev	3	rain	cond	-0.346	27	1
Rev	4	rain	aspd	0.202	21	1
Rev	4	rain	temp	-0.081	-42	
Rev	4	rain	cond	0.124	35	
Rev	5	rain	aspd	-0.089	-29	
Rev	5	rain	temp	0.460	21	1
Rev	5	rain	cond	-0.112	95	1
Rev	6	rain	aspd	-0.067	12	
Rev	6	rain	temp	-0.623	12	1
Rev	6	rain	cond	0.359	0	

site	period	x	y	cor	lag	sig
SC	1	rain	aspd	0.351	70	1
SC	1	rain	temp	-0.628	149	
SC	1	rain	cond	-0.189	-14	
SC	2	rain	aspd	-0.480	53	
SC	2	rain	temp	-0.661	12	1
SC	2	rain	cond	-0.028	6	1
SC	3	rain	aspd	0.104	0	1
SC	3	rain	temp	0.631	37	1
SC	3	rain	cond	0.624	38	
SC	4	rain	aspd	0.051	0	
SC	4	rain	temp	0.105	35	1
SC	4	rain	cond	-0.129	60	1
SC	5	rain	aspd	-0.344	31	
SC	5	rain	temp	0.503	6	1
SC	5	rain	cond	0.354	4	
SC	6	rain	aspd	0.211	76	1
SC	6	rain	temp	0.234	34	1
SC	6	rain	cond	0.185	12	

Table 4. Results of global correlation analysis. The table is split by site name, and is further organized by period and then by variable *y*. The significant observations are also colored by period number. The table shows that only about 50% of the results were deemed significant based on correlation value and amount of lag in hours.

3.3 Geospatial Correlations

After examining the global correlations above, we began to look into how the spatial distribution of rainfall in the study area affects conduit flow. We ran correlation analyses across the six individual study periods identified by the clustering procedure. We correlated each of the three testing variables, velocity, temperature, and conductivity, at each of the three test sites with each of the 321 drainage basins in the study area. To calculate these correlations, we had also applied certain amounts of lag time to the rainfall data ranging from 6 to 20 hours, to account for the time between rainfall and change in conduit flow.

After calculating the correlation coefficients between conduit velocity and rainfall into each basin, we are able to color each basin on a map of the study area by the value of the resulting correlation. The map shown in Figure 9 helps to visualize the results of this procedure. Figure 9 illustrates the correlations between each basin's precipitation data and Wakulla conduit AK temperature data during Period 6. The legend shows that the correlation results range from -1 (blue) to 1 (red) meaning that the algorithm had computed both perfectly positive correlations and perfectly negative correlations with regard to certain basins. The drainage basins in the figure are colored by correlation value, where dark red basins represent strong positive correlations, dark blue basins represent strong negative correlations, and yellow basins represent weak correlations. After illustrating the results we can see the distribution of correlation values. Negatively correlated basins, colored in shades of blue, appear west of the Ochlockonee River, colored green, and far to the west of the springs' locations, portrayed by pink dots. Positively correlated basins, shown in shades of red, also occur to the west and north of the Ochlockonee River which are likely to be false positives. The more meaningful results appear between the locations of Revell Sink and Spring Creek, where rainfall in the basins are positively correlated with temperature changes in Conduit AK.

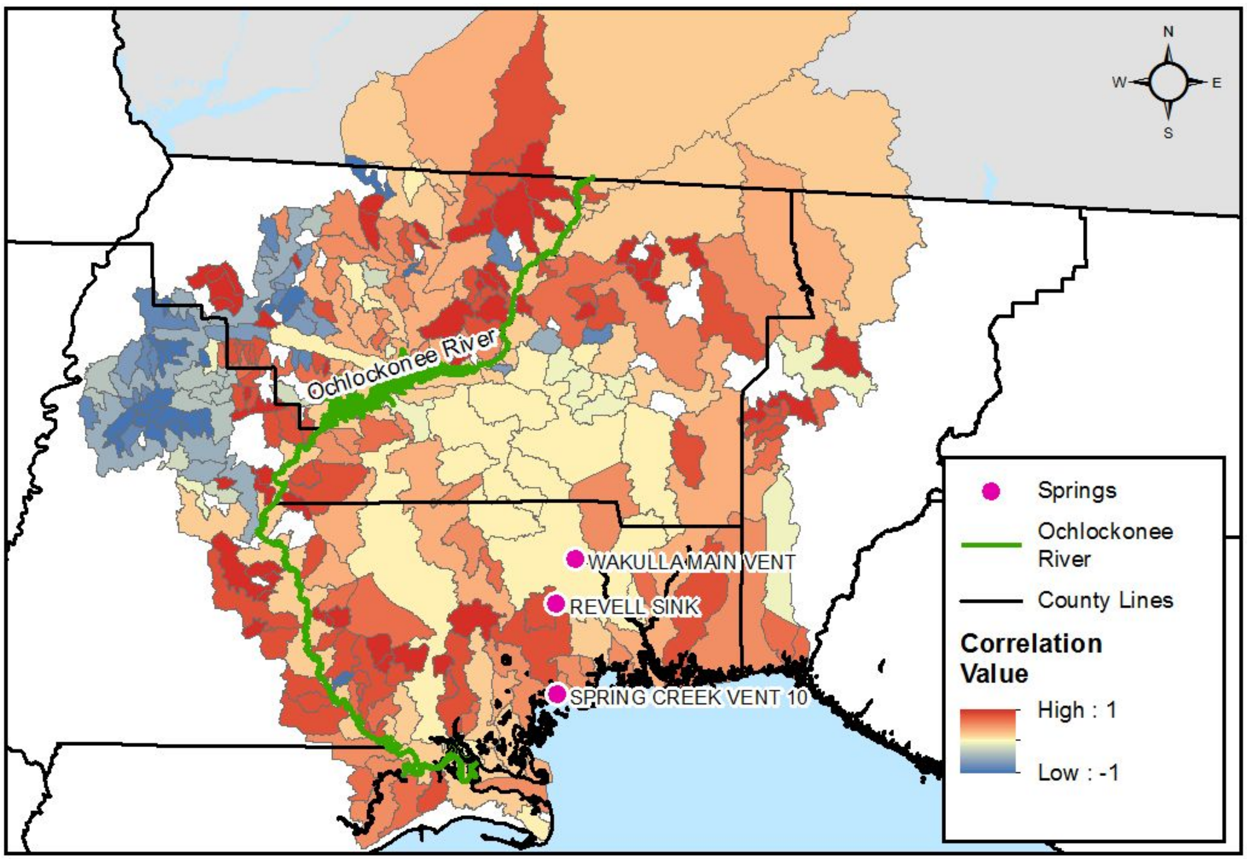

Figure 9. Spatial distribution of correlations between rainfall in each of the 321 basins and temperature at Wakulla conduit AK during Period 6. Drainage basins are colored by their correlation value ranging from -1 (blue), to 0 (yellow), to 1 (red), while spring locations are depicted with pink dots. Note how the negatively correlated basins appear farther from the springs' locations, west of the Ochlockonee River colored in green. Basins with high positive correlations that surround the river are likely to be false positives. In the area surrounding Revell Sink and Spring Creek, we see significant positive correlations.

4 Discussion

The results of this paper are discussed below. Results were obtained with regards to clustering analysis, global correlation analysis, as well as geospatial correlation analysis.

4.1 Clustering

The clustering algorithm found that the conduit system can exist in two states. The first state, or cluster, is characterized by low conductivity and medium temperatures. The low conductivity tells us that the Falmouth sensor detects low salinity and therefore, freshwater is in the conduit. The medium temperatures shows us that groundwater is flowing, rather than surficial water, since surface water is affected by seasonal temperatures. We can immediately conclude that when the system is in state 1, Spring Creek is flowing South and is steadily discharging freshwater into the Gulf of Mexico. Observations in cluster 2, or state 2, typically have high conductivity values and cover a wide range of temperature values. From the high conductivity values, we know that the Falmouth sensor is detecting salt water, and the wide range of temperatures tells us that the water temperature has been affected by seasonal change, and therefore must be surface water. So, when the system is in state 2, we can infer that Spring Creek is not discharging enough fresh groundwater to keep the pressure up at the spring vent and that seawater has instead begun to flow into the conduit, due to its higher density. As seawater flows North through the conduit, the sensor detects a spike in conductivity, as well as a change in water temperature from groundwater to seasonal. When Spring Creek is in this state, it is considered not flowing.

For the purpose of the study, the two states are referred to as “Flowing” and “Not Flowing”, in reference to the status of Spring Creek. After organizing the data into their appropriate clusters, we also classified the data collected at Wakulla Spring and Revell Sink into the same clusters. This allowed us to visualize the periods during which Spring Creek is “flowing” in all of the conduits, such as in Figure 8, so that we can compare the effect of Spring Creek’s flow on the other springs in the study. A total of six periods where Spring Creek is flowing were identified using the clustering mechanism and are laid out in Table 3. The table shows that periods 1, 4, 5, and 6 are each about four months long, while periods 2 and 3 are approximately one month. Also, periods 1, 2, and 3 occur during the winter months, while periods 4 and 6 occur during the summer, and period 5 starts in the winter and ends in the summer. The clustering method recognizes the data in Figure 4 as one period, period six.. The figure provides some initial evidence for the existence of correlations between rainfall and source activity.

4.2 Correlations

The results of the cross correlation function, outlined in Table 4, prove that there is a period of lag time between the moment rain has begun to fall to the moment conduit flow is affected by it.

The correlation values and lag times appear to vary depending on the time of year, how much rain has fallen, and where the rain has fallen. The seasonality of the temperature data causes certain periods to generate significantly positive or significantly negative correlations. In the winter months, rainfall may cause the temperature within the conduits to decrease, yielding a negative correlation, while summer rains will increase the water temperature, yielding positive correlations. Furthermore, heavy rains may quickly change the speed, temperature, or conductivity within the conduit potentially resulting in lower lag times or stronger correlations, while light precipitation may result in longer lag times, and weaker correlations. Lastly, rain that falls into the drainage basin bearing a certain spring will induce higher correlations and lower lag times at that spring, while rain falling into a distant basin that feeds a potentially undiscovered sinkhole or spring will affect the conduit system after longer lag times, and with potentially lower correlations.

4.2.1 Global Results

The results in Table 4 show that there are only certain cases where a meaningful or significant correlation can be calculated between conduit flow parameters and the total precipitation fallen across the study area. This could be due to the intensity or location of each rainfall event or a number of other external factors that complicate the nature of the system. In taking a closer look at the lag times of significant observations in the results table, a possible trend can be spotted. On three separate occasions, the amounts of lag needed to correlate both temperature and conductivity with rainfall seem to be more similar than the lag time correlating velocity with rainfall. This relationship can be seen in conduit AK during Period 3 and in Revell Sink during Periods 1 and 3. In each of these cases, the amount of time it takes for rainfall to have an effect on velocity is dramatically different from, either longer or shorter than, the time it takes to affect conduit temperature or conductivity, which have very similar lag times. With more conclusive results across the board, we might be able to determine the significance of this trend.

4.2.2 Geospatial Results

The results of the geospatial correlations are somewhat significant, but are generally unreliable. One would expect that the basins which harbor the springs in question would show higher and stronger correlations, than those farther away in the study area. Though, according to the results, many basins surrounding the Ochlockonee River show positive correlations, where no correlation should be possible. These results are likely to be false positives which may have been correlated by coincidence. This procedure may have also delivered some significant results. Precipitation in basins near Spring Creek and Revell Sink seem to be positively correlated with temperature changes in conduit AK, confirming that this procedure is able to detect significant results. With more advanced data cleaning and a more intricate correlation procedure, this algorithm may be able to accurately detect undiscovered springs in the Woodville Karst Plain.

5 Conclusions

This study examined three continuous years of data totaling 158,000 rows of flow meter data and 975,000 rows of NEXRAD precipitation data at a temporal resolution of one hour, over a geographical area of 3,663 square miles. The results of this study provide strong evidence for the existence of a dependence of conduit activity on rainfall, though the analysis methods applied in this paper failed to verify that significant correlations could be calculated. We have concluded that these methods alone cannot replace dye tracing. We attribute this to the possibility that the raw data, either NEXRAD, Falmouth, or both, may not have been cleaned properly before analysis. It is also possible that the short term variance in each of the variables may have caused discrepancies between the data sets that even accounting for a period of lag could not resolve.

Throughout the course of the project, we found that the conduit system can exist in one of two states, where the freshwater in the Spring Creek Springs Group is either “flowing” or “not lowing”. Upon further analyzing these two states, we found that rainfall seems to affect conduit flow to a greater extent while Spring Creek is flowing, as opposed to when it is not flowing. We were able to determine the periods throughout the span of the study during which Spring Creek was flowing, and analyze the effect of rainfall on conduit flow during each one individually. These results, shown both in Table 4 are quite inconclusive. Few of the results seem to be significant, while a majority of the results are implausible, due to either the correlation value or the amount of lag at which the highest correlation was found. We continued on to analyze the relationships between conduit flow and each individual drainage basin and discovered more insignificant results. Basins far from the springs, surrounding the Ochlockonee River, seem to show stronger correlations, both positive and negative, than those basins nearer to the springs in the study which we expected to be correlated.

In conclusion, these methods, as they are, cannot replace the capabilities of dye tracing. Although, clustering has proved to be an effective technique for classifying flow states in the system. The global correlation method proved to be of moderate use, but more work is needed to refine the technique. Unfortunately, the geospatial correlation method was unable to sufficiently delineate basins contributing to changes in conduit flow. As such, much more work and refinement is needed before these methods can effectively replace dye tracing.

6 References

Bourke, P. *Cross Correlation*, August 1996.

Chicken, E., and P. Chalise. *Conduit Prevalence in the Woodville Karst Plain*. *Sinkholes and the Engineering and Environmental Impacts of Karst*, 2008, 303–12.

Davis, H. *The Interconnectedness Between Spring Creek Springs and Wakulla Spring and the Effects on Flow at Wakulla Spring*. 2011.

Davis, H., and R. Verdi. *Groundwater Cycling Between a Submarine Spring and an Inland Fresh Water Spring*. *Groundwater* 52, no. 5 (2013): 705–16.

Falmouth Scientific, Inc. *2D-ACM User Manual.* Falmouth Scientific, Inc., November 2006.

Fleury, P. *Submarine Springs and Coastal Karst Aquifers: A Review*. Translated by P. Fleury, M. Bakalowicz, and G. de Marsily. *Journal of Hydrology* v. 339, no. no. 1–2 (June 10, 2007): pp. 79–92.

Kimberlain, T.B. *Tropical Cyclone Report: Tropical Storm Debby*, January 7, 2013.

Kresic, N. *Water in Karst: Management, Vulnerability, and Restoration*. New York: McGraw Hill, 2013.

Lane, E. *The Spring Creek Submarine Springs Group, Wakulla County, Florida*. Special Publication. Tallahassee, Fl: Florida Geological Survey, 2001.

Pang-Ning, T., M. Steinbach, A. Karpatne, and V. Kumar. *Introduction to Data Mining*. 2nd ed. Pearson, 2013.

Prabhakaran, S. *Time Series Analysis*. *R-Statistics.co* (blog), 2017.

Scott, T.M., G.H. Means, R.P. Meegan, R.C. Means, S.B. Upchurch, R.E. Copeland, J. Jones, T. Roberts, and A. Willet. *Springs of Florida.* Bulletin. Tallahassee, Fl: Florida Geological Survey, 2004.

Stone, W.C., ed. *The Wakulla Springs Project*. Huntsville, AL: The U.S. Deep Caving Team, 1993.

US Geological Survey. *Watersheds, Hydrologic Units, Hydrologic Unit Codes, Watershed Approach, and Rapid Watershed Assessments*. US Geological Survey, June 2007.

Wilson, J.F. *Time-Of-Travel Measurements and Other Applications of Dye Tracing*. Hydrologic aspects of the utilization of water: International Association of Scientific Hydrology, 76 (1967).

Xu, Z., S. Bassett, B. Hu, and S. Dyer. *Long Distance Seawater Intrusion through a Karst Conduit Network in the Woodville Karst Plain, Florida*, 2016.