12-2018

Decadal Land Surface Phenology and Water Quality in the Headwaters Illinois River Watershed

Justin Ray Rollans
University of Arkansas, Fayetteville

Decadal Land Surface Phenology and Water Quality in the Headwaters Illinois River Watershed

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Geography

by

Justin Rollans
Louisiana Tech University
Bachelor of Science in Geographic Information Sciences, 2016

December 2018
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Jason A. Tullis, Ph.D.
Thesis Director

_____
Fred Limp Jr., Ph.D.
Committee Member

_____
J. Van Brahana, Ph.D.
Committee Member

Abstract

Over 25 percent of the world's population either lives on or obtains water from karst aquifers. The complex interactions between subsurface karst geologic features, the constant motion of the plant life cycle, and significant water resource demand all suggest the need to better define those interactions. The relationship of historical land surface phenology and water quality in karst topography were investigated in the Headwaters Illinois River watershed in Northwest Arkansas (NWA). This area represents high vulnerability to surface water and groundwater contamination, with both natural and anthropogenic processes such as over application of broil litter for enhanced cattle browse, affecting groundwater quality. Land surface phenology patterns influenced by these processes were identified using Landsat satellite imagery and object-based image analysis (OBIA). A normalized difference vegetation index (NDVI) time series was produced using Google Earth Engine for all passes over the study area that meet atmospheric and data quality criteria over two decades from 1999 to 2018. Analysis of NDVI and ancillary data over time allowed insight into vegetation health norms, deviation from those norms, and human impact upon regional vegetation. OBIA techniques were used to segment vegetation index time series pixels into polygons based on adjacency and similarity. Resulting polygons were categorized using an unsupervised clustering approach, and were labeled based on visual and expert interpretation of the study area. The relation of the image analysis results to groundwater quality was determined using data organized by hydrologic catchments within the study area. Comparison of the decadal water quality data and NDVI image analysis resulted in meaningful temporal patterns within the datasets but showed a near 0 slope for NDVI and water quality metrics. Future LSP studies should consider areas with greater spatial and temporal availability of water quality metrics and variable surface/groundwater interactions.

Acknowledgements

I would like to acknowledge the people that have helped me along this thesis journey. Thank you Dr. Jason Tullis for all of the time you took to help me through the struggles of data collection, analysis, and presentation. I would also like to thank my other committee members Dr. John Brahana, for all of the karst geology insight and recommendations. Also thank you to Dr. Fred Limp for being there in a moral support role. Thank you as well to Panagiotis Giannakis for endless amounts of help and being my go to person in the bull pen.

Table of Contents

Table of Figures

Table of Tables

Introduction

Northwest Arkansas (NWA) is located in the southwestern portion of the physiographic region known as the Ozark Plateau or Ozark Mountains. Topography of this area is one of uplifted and incised plateaus that have been intensively weathered and eroded into rounded hills and broad valleys. The interactions between those rocks and water generally creates what is known as a karst terrain (Monroe, 1970). The technical definition of karst is: an area where drainage moves rapidly from the surface to the subsurface. The topography is chiefly formed by the dissolving of rock by the focused flow of groundwater, which dissolves the rock and creates open voids. Karst may be characterized by sinkholes, sinking streams, closed depressions, subterranean drainage, and caves (Monroe, 1970). These characteristics of karst terranes allow for quick infiltration of precipitation vertically to the water table, and from there, horizontally along permeability contrasts that perch and confine the groundwater until it resurges to the surface as springs or baseflow to springs. Karst aquifers are highly susceptible to contamination in most regions owing to their open nature, and general lack of filtration (Leidy & Morris, 1990). Processes occurring upon the surfaces within these terranes can inherently have an effect on nutrient levels and overall quality of the groundwater. (Leidy & Morris, 1990; Peterson et al., 1998; Jiang et al., 2008).

The water quality in these type of terrain can have multiple different effects on plant life, wildlife, and human life, especially where all three interact heavily. Specifically, nitrate (NO_3) and phosphorus (P) are twp of the most common excess nutrients found in surface and ground water near areas where active livestock agriculture and urbanization is prevalent (Brion et al., 2010). NO_3 and P are both essential nutrients when it comes to plant growth, but when higher levels are present in aquatic systems they can cause eutrophication from higher levels of

algal and organic material growth (Brion et al., 2010). In karstic systems where this is apparent in surface water these nutrient levels can almost directly translate to groundwater in that area which can prove problematic in some cases, especially for people using shallow mantled wells in the area, for various water uses.

The interflow zone is a stratigraphic zone within unsaturated zones of karst terrain that has been found to have a major impact on denitrification and microbial processing of potential contaminants (Brahana, 2011). Interflow just refers to overland flow infiltrating the unsaturated zone and then returning to the surface before reaching the saturated zone. This is important to this study because that same zone is where the root structures of plants will also reside in the stratigraphic make up of karst terrain. Those root structures contribute to the microbial structure of the interflow and unsaturated zones. Studying an area's land surface phenology (LSP) over time can provide insight into the overall history of health, growth, and decline of these root systems based on satellite image analysis of spectral indices.

LSP refers to the phenologic cycles of the vegetation upon the surface of the earth. Vegetation phenology cycles are defined as the relatively predictable diurnal, seasonal, and annual cycle of vegetation development and senescence (Jensen, 2016). LSP is quantified by recording remotely sensed spectral information, usually from a satellite platform. Normalized Difference Vegetation Index (NDVI) is a remote sensing vegetation index that was developed to give scale to overall vegetation health based on chlorophyll content of green biomass (Jensen, 2016). LSP curves are usually made up of NDVI values from a various possible range of temporal scales; average monthly, average weekly, etc. More specifically this is information that is recorded by measuring the amount of reflected energy, within the red and near infrared portion of the electromagnetic spectrum, from the earth's surface. These curves can be affected annually

by multiple different weather and moisture related factors, but when analyzed over large periods of time, such as decades, these curves would, on average, be the same if there was no human factor for change upon the earth's surface. Since almost seven and a half billion people inhabit the earth, they're a constant factor of change.

Change is the main factor when it comes to the results of these type of land surface studies. What particular parameter or value is used to assess change will vary throughout the scientific community. The majority of studies that try and define land surface changes as a cause for various phenomena analyze land-use or land-cover parameters to quantify and monitor changes upon the surface of the earth (Jiang et al., 2008; Brion et al., 2010; Chunhao & Yingkui, 2014; Öztürk et al., 2013). Land-use is defined as the specific anthropogenic processes that land is being used for, while land-cover is more loosely defined as what is physically covering the surface of the land whether natural or anthropogenic. Many land-use/land-cover (LULC) classification systems have been standardized over the years, such as: American Planning Association's *Land-Based Classification System (LBCS)*, U.S. Geologic Survey's *Land Use/Land Cover Classification System for Use with Remote Sensor Data*, U.S. National Land Cover Dataset (NLCD), and many more (Jensen 2016).

Karst ground and surface waters have become more heavily monitored in recent history but still rely on old fashion techniques of change detection over large and usually incomplete temporal scales. Higher quality analysis at a smaller temporal scale is possible with more complete water data. This leads to the following research questions that need to be answered:

1. What is the optimal workflow to monitor influences upon karst water systems, using remote data collection as much as possible?

2. How can publicly available satellite data be leveraged to more accurately and regularly assess anthropogenic processes that affect sensitive environments?

In support of these questions, the following hypotheses were addressed:

1. Research hypothesis: meaningful LSP classes will be derived from the NDVI time series image analysis.
2. Null hypothesis: There is no relationship between LSP classes of NDVI time series and surface water nutrient levels.

The goal of this study is to determine if there is a relationship between water quality and historical land surface changes based on two decades' worth of Landsat-derived land surface phenology classes, within the Headwaters Illinois River Watershed (HWIRW). These land surface phenology classes will take the place of LULC classes traditionally used for these types of comparative studies. The reasoning behind this substitution is that many land-use specific studies fall short of propagating error induced by studies that involve larger spatial scales, due to the amount of ground referencing needed to accurately represent the study area's land-use (Fisher et al., 2017). Land-cover classification and quantification methods are for the most part quite general and aren't easily normalized for reproduction of results.

NDVI values over time provide a mathematical normalized dataset that can be applied worldwide, for many types of comparisons, because the values are recorded the exactly same way depending on which specific sensor is used. These values eliminate human induced error, as mentioned in the LULC studies, for the data upon which the change will be assessed. Along with other accuracy optimization steps, error can be minimized. It is expected useful LSP classes will be derived from the NDVI time series image analysis in relation to water quality in the HWIRW.

Literature Review

Changes of many different types have made NWA what it is today; of particular interest among these is the fast rate of population growth, between 1999-2018. According to a March 2017 article by, Jeff Della Rosa of talkbusiness.net, he quotes U.S. Census Bureau statistics regarding population growth of the 382 designated metropolitan statistical areas (MSA) in the U.S. of which NWA MSA was ranked 22nd in population growth, as of 2016 (Rosa, 2017). The NWA MSA is also the fastest growing MSA within the state of Arkansas, as of 2016 (Rosa, 2017). With this rapid rate of population growth, it is presumable that the land surface of NWA has made some drastic changes since 1999. Responding to a population increase of a little more than 60 percent in less than a twenty-year time period can put a major strain on natural resources within the area, particularly if they are not managed properly. In addition to the area's urban sprawl there has also been a sprawl of livestock agriculture.

According to the USDA's Economic Research Service 2017 report (as cited in University of Arkansas Division of Agriculture, 2017) Arkansas's top agricultural commodity in terms of cash receipts was broilers, with cattle ranking fourth, and pigs ninth. These are the three main types of livestock agriculture in NWA. Areas that once were farms on the outskirts of the city have now been sold to make room for expanding urban areas, which in turn means other more rural locations have been converted into land used for livestock agriculture. With more livestock come more livestock waste and more waste disposal, particularly with poultry livestock.

One of the most vital and vulnerable natural resources in NWA is water. Water in NWA serves multiple functions for its residents, as it does in most of the world, but due to the geologic terrain of NWA surface water and groundwater quality can be highly influenced by the way

other natural resources of the area are managed (Brion et al., 2010). Regional scale planning could benefit from an empirical method of determining what type of anthropogenic processes affect different regions based on geographical observations. Moderate resolution NDVI data derived LSP classes can help more accurately assess how certain instances of increase and decrease in recent history affect other environmental processes on local scale (Morisette et al., 2009).

Water in Karst Terrain

Surface water is defined as, water on the earth's surface that is fed by groundwater discharge and overland storm flow (White, 2002). This type of water in karst terrain plays two key roles in what is, essentially, a three phase water cycle. This three phase water cycle consists of recharge, groundwater flow, and discharge (White, 2002). The recharge and discharge phase of the karst water cycle is where surface water quality values are most valuable for scientific analysis. According to William White's 2002 journal article, precipitation results in overland storm flow; that overland storm flow of surface water can recharge a karst aquifer in three ways, allogenic recharge, diffuse infiltration, and internal runoff (2002). The HIRW resides in two different physiographic regions, the Boston Mountains to the south and the majority residing within the Springfield Plateau to the north (Figure 1). Due to the slightly different lithology and topography present in each region, different types of recharge are more prevalent between regions (Kresse et al., 2014). Internal runoff is the quickest type of recharge because it happens when overland flow goes directly from surface to aquifer via sinkholes or large solution dissolved fractures (White, 2002).

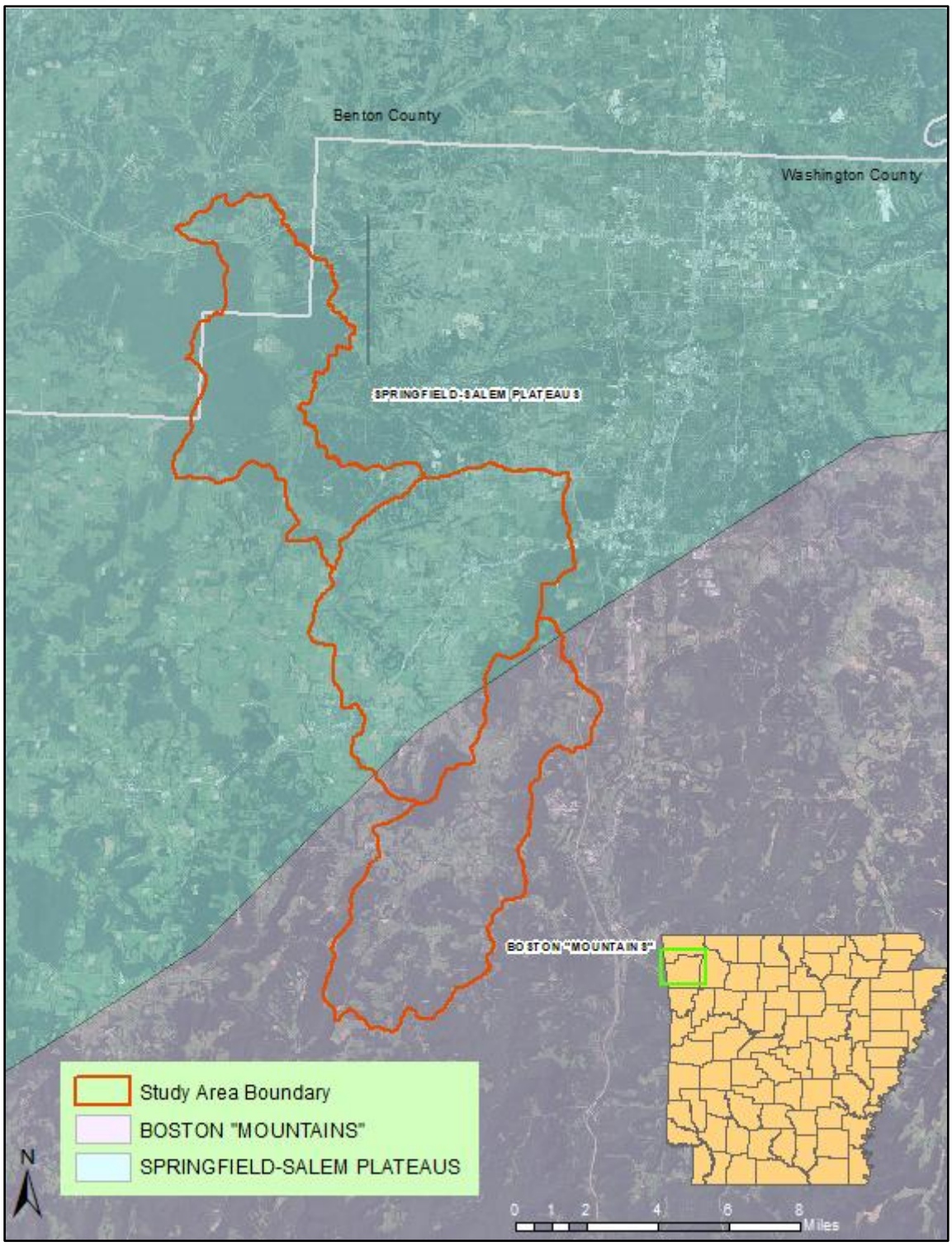


Figure 1: HWIRW Physiographic Regions

This map shows the 10-digit hydrologic unit code (HUC 12) in the Headwaters Illinois River Watershed (HWIRW; outlined in orange) and how it falls within two physiographic regions, as delineated by US Geological Survey (USGS; data courtesy Arkansas GIS Office [AGIO], 2017).

This type of recharge will only be prevalent in areas that contain those karst features and, under normal rainfall conditions, will be restricted to one aquifer (White, 2002). These karst features are present in both physiographic regions within the study area but the majority of recharge does not happen by this mechanism. Even though it is not the main mechanism of recharge it is the most vulnerable to contaminants, because of direct infiltration from surface to aquifer. When water bypasses the interflow zone it has bypassed all potential for denitrification or microbial processing of contaminants within the water (Brahana, 2011).

The HWIRW is dominated by diffuse infiltration in the less steep parts of the topography. Diffuse Infiltration is defined as overland flow moving from the epikarst to the aquifer via soil and fractures, then entering the matrix permeability of the underlying carbonate bedrock (White, 2002). This recharge mechanism gives water the most interaction with the interflow zone because of the lower level of channelization of overland flow, allowing the water to infiltrate more locally to original rainfall locations. The amount of fracturing, conduits, and flow-path diversity within the interflow zone, are heavily influenced by the root systems of vegetation upon the surface (Brahana, 2011). This means that more diversity in flow-path can potentially mean more nutrient attenuation within the interflow zone, leading to better quality water reaching the aquifer.

The karst water cycle is just that, it is a recycling process that goes from recharge, to groundwater flow, and in the majority of cases that same water will be discharged into another surface water unit. Discharge is the other component, along with recharge, of surface water quality. Rates of discharge, or volume of water in a stream flowing past a single point, can be closely related to water quality in karstic systems (Davis & Schumacher, 1992). According to some studies the concentration of contaminants can vary between the different levels of

discharge based on contaminant source. In the 1998 published report by Peterson et al., shows that if the contamination source does not come from runoff, but instead from something such as a wastewater treatment plant that is directly putting their discharge into the stream, contaminant concentrations will be lower in times of high discharge rates and higher during times of low discharge rates (1998). If the main source of contaminants comes from runoff, or overland flow, then concentration of contaminants will more highly correlate with rates of discharge (Peterson et al., 1998). The Illinois River receives most of its contaminants from overland flow that is channelized into its many tributaries (Brahana, 2011; Brion et al., 2010).

In the HWIRW most of the cyclic water processing occurs in the higher elevations of the watershed, where the various groundwater discharge and overland flow fed tributaries flow downslope into the Illinois River. The riverbed is the lowest point within the watershed meaning all discharge from the surrounding higher elevation land will flow into the river at some point. Streambeds in karst terrains, including the Illinois River, are characterized by the fact that they can gain and lose water to groundwater sources. This is due to the fractured formation of the underlying bedrock which allows groundwater inflow from aquifers that are mantled at higher elevations, and for outflow into shallow flowing aquifers beneath the surface of the stream bed (Kalkhoff, 1993). This relationship between recharge and discharge in relation to surface water provides the insights needed to understand how the far more complex karst groundwater systems can mitigate excess nutrients that are common in areas with a major livestock agriculture presence and urbanization.

Potential Factors of Change

Rapid urbanization and livestock agriculture are known land use types that can be harmful to karst water systems (Chunhao & Yingkui, 2014; Jiang et al., 2008; Brion et al., 2010). When an area is urbanized it goes from rural land with great biodiversity and little impervious surfaces to a majority of its surfaces being impervious and less biodiversity (Barbec, Schulte, & Richards, 2002). M.K. Ridd (1995) presented a study focused on exploring the possibilities of a vegetation-impervious surface-soil (V-I-S) that relied on remote sensing data (1995). The goal of Ridd's exploratory model was to characterize urban and near-urban environments on a universal scale and to be able to compare the way those environments change within and between cities (Ridd, 1995). He realized that the majority of the urban environmental change remote sensing literature was based on traditional spectral classifications methods for defining land-use (Ridd, 1995). This highlights the objectiveness of using multi-spectral index values as defining characteristics of a landscape instead of a subjective system of classification definitions. The study also addressed the great amount of variability between types of surfaces within an urban and urbanizing area and how they impact different ecological parameters, such as storm water runoff (Ridd, 1995). Ridd breaks urban environments down into three fundamental surface categories: impervious material, green vegetation, and exposed soil (1995). These three surface types will react to water differently. Obviously, impervious surface do not allow water to infiltrate into the soil causing the water to flow the surfaces gradient which usually is constructed in a manner that will remove the water from the surface in the fastest way possible (Barbec, Schulte, & Richards, 2002). This forces overland flow of large areas to become channelized in runoff streams which limits the amount of natural surfaces the water gets to interact with, therefore causing all pollutants to be more concentrated in runoff streams. Areas dominated by

green vegetation within an urban setting usually are the smallest, proportionally, to the other surface types (Barbec, Schulte, & Richards, 2002; Ridd, 1995). These areas react to water similarly to the rural, more vegetated areas surrounding urban areas. Rural areas tend to not centralize water movement as much urban areas, which allow a better percentage of surface water to be absorbed and cleaned in the natural way of that area's terrain (Barbec, Schulte, & Richards, 2002). Depending upon how urban watershed is designed; vegetated urban areas' potential to absorb and clean overland flow from surrounding impervious surfaces can be over used. Overloading these areas, especially in karst terrain, with this type of runoff will exceed the surfaces drainage carrying capacity and cause pollutants to be more concentrated (White, 2002). Exposed soils can vary widely in composition in urbanizing areas. This surface type can usually be found on the peripherals of the established urban zone because of higher rates of construction. One thing remaining common among soil compositions, just at varying magnitudes, is their susceptibility to compaction and erosion when root systems are removed. When soils remain exposed for long periods of time, especially in regions of drastic topography change, erosion can cause many different problems with surface water behavior and quality (Barbec, Schulte, & Richards, 2002). This causes new sediment and potential pollutants to be transported into stream beds which can shift the course of the stream and again add to the carrying capacity of pollutants. Barbec, Schulte, & Richards (2002) found that exposed soils of urbanizing areas tend to act like impervious surfaces in relation to runoff. They attributed that to the compaction and change in profile of the soils in construction areas; where exposed soil surfaces are typically found (Barbec, Schulte, & Richards, 2002). By compacting the soil and removing the root system once covering the surface infiltration rates are adversely affected by a reduced conduit into the soil

and plant uptake. This evidence shows why and how urban surfaces generally do not improve the natural process of cleaning and absorbing surface water.

Where present, livestock agriculture is a major anthropogenic contributor to land surface change. According to a regional food assessment report, prepared by Karp & Sandusky of Karp Resources, 99% of NWA's agriculture production comes from livestock and poultry (2014). These agricultural practices require generally open land and happen in rural areas. This is one reason why the HWIRW was selected as a study area; there are rural areas in the southern and northern portions and urban areas in the central portion. Livestock and poultry production impacts surfaces in different ways especially when the industry is expanding and contributing to deforestation of the surrounding land surface. Brion et al. argues that this transition affects the natural land surface characteristics, water balance, and hydrologic cycle of a region (as cited in LeBlanc et al., 2007; Tong & Chen, 2002). Broilers, the number one commodity produced in Arkansas, produce a large amount of litter that is predominantly high in NO_3 and P content due to their inability to fully utilize these nutrients from their feed (Brion et al, 2010). Due to this high concentration of nutrients, broiler litter is repurposed for improving low nutrient soils as a fertilizer. This is where livestock and poultry production cross paths. The fourth biggest commodity as of 2016 produced by the state of Arkansas was cattle (University of Arkansas Division of Agriculture, 2017). In the spirit of cost efficiency cattle farmers, especially in NWA, try to rely as little as possible on supplemental feeding and rely more heavily on grazing. To improve grazing habitats it is general practice to apply fertilizer in order for the soil to allow plants to express their full biomass potential. The Ozarks in general are characterized by having naturally rocky soil which makes seed to soil contact difficult for seed germination and nutrient retention over time. This causes cattle farmers in the area to use a higher amount of fertilizer, in

the form of broil litter, to improve the soil quality for grazing potential (Brion et al., 2010; Brahana, 2011). Traditionally practice was to apply broil litter every spring to promote growth during the green-up phase of the phenologic cycle, regardless of the nutrient levels within the soil (Brion et al., 2010). Arkansas Natural Resource Commission (ANRC) got legislation passed, that went into effect Jan. 1, 2010, developing rules governing the Arkansas Soil Nutrient and Poultry Litter Application and Management Program; known as Title 22 (2009). The ultimate goal of this provision was to enhance and protect surface water quality within at risk watersheds, while maximizing surrounding soil fertility and proper plant growth (Arkansas Natural Resources Commission [ANRC], 2009). This set of rules designated nutrient surplus areas (Figure 2) within the state which includes almost all of NWA, including the HWIRW (ANRC, 2009).

While this legislation finally made it mandatory, in certain areas, for a soil test to be conducted before organic or inorganic fertilizer application; surface water quality problems are still present. Cattle are also shown to have derogatory effects on biodiversity especially in riparian habitats where there is a higher concentration of desired plants, water, and shade causing over use of these habitats (Brion et al., 2010). The geographic relationships of agricultural practices, prevalent in NWA, make it an interesting and diverse area to better understand the complexity of those relationships to the natural environment.

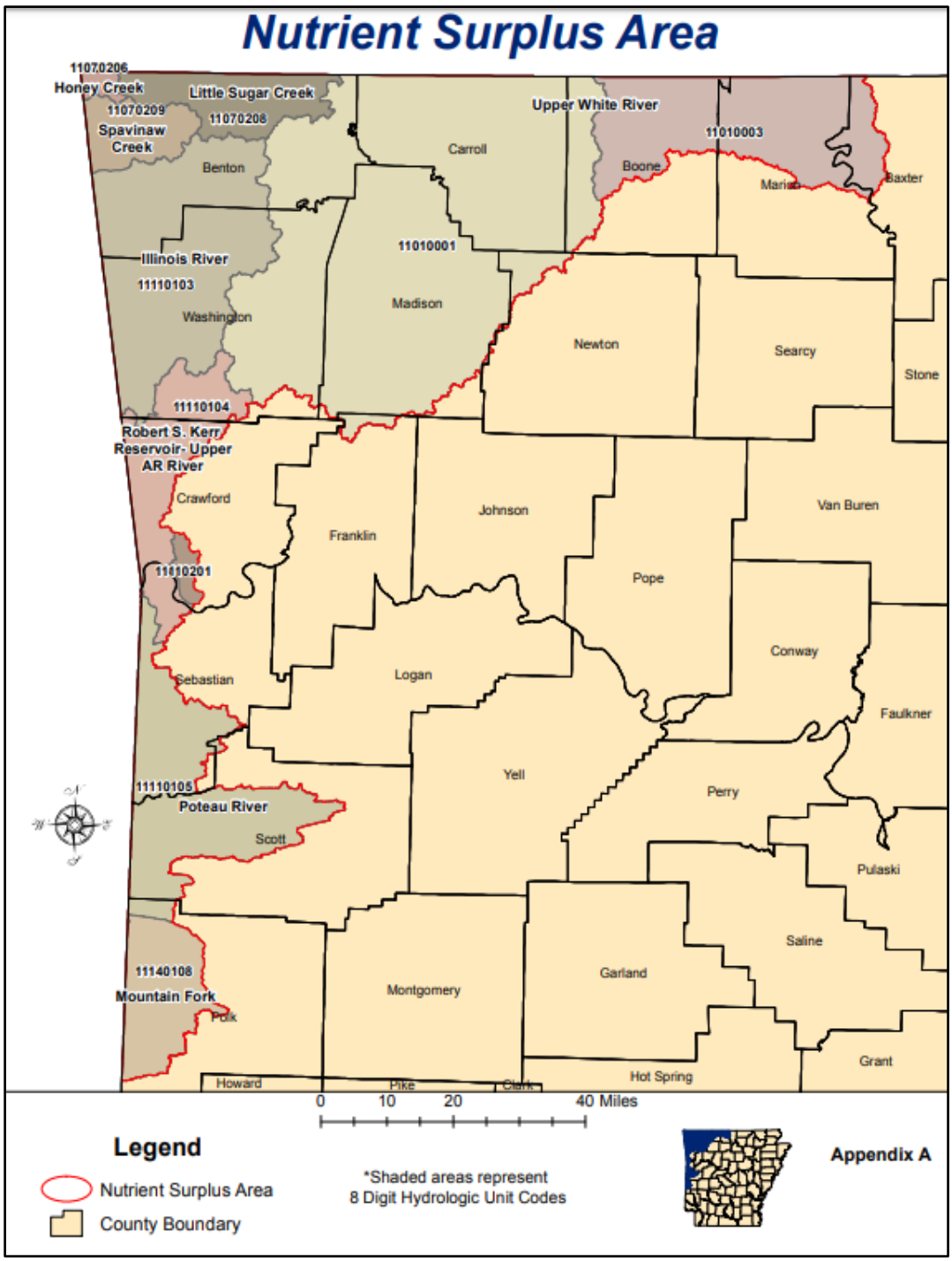


Figure 2: Nutrient Surplus Area

Nutrient surplus information indicating a region that is susceptible to environmental regulations regarding organic and inorganic nitrogen application rates. Courtesy of Arkansas Natural Resource Commission (ANRC, 2009).

Land-Surface Phenology

Phenology is defined as the study of recurring life-cycle events that are initiated and driven by environmental factors (Morisette et al., 2009). LSP is consulted as the independent variable of this study because of its basic cyclic nature, meaning it is a constant phenomenon that varies only slightly annually. This is especially true in settings where humans are of no impact to the land surface. Any large deviation from that cyclic pattern, in most cases, makes for a quick indicator of human impact to an areas land surface characteristics. A Fried et al. white paper, defines LSP specifically as the seasonal pattern of variation in the properties of vegetated land surfaces on the regional or global scale, which is typically characterized using satellite remote sensing products (as cited in Morisette et al., 2009). LSP is useful to many different types of studies because of its utilization of quantitative NDVI measurements to make qualitative definitions of biophysical phenomena on many scales. NDVI specifically normalizes radiant flux recorded in the red and near infrared portions of the electromagnetic spectrum which is reacting to the amount of chlorophyll present in greed biomass (Jensen, 2016). Chlorophyll levels in leaves of green biomass will allude to the overall health and phenologic progress of the plants within the spatial scale. This means that substances that do not contain chlorophyll, such as impervious urban surfaces, will react negatively in these portions of the electromagnetic spectrum and will be easily classified. Most studies aiming to derive LSP classes for an area use NASA's moderate-resolution imaging spectrometer (MODIS) NDVI bands which have a 250 meter spatial resolution (Morisette et al., 2009; Jensen, 2016). Landsat 5's Thematic Mapper (TM) and Landsat 8's Operational Land Imager (OLI) have a much higher spatial resolution at 30 meters. This higher resolution will help highlight patterns on a more localized scale for environmental parameters (Morisette et al., 2009). In a 2010 Gu et al. presented a study about

phenological classification of the contiguous United States for multiple sensors. Their methods of deriving LSP metrics are useful because of the replicable nature of NDVI parameters across sensors. Gu et al. used nine phenological metrics in their analysis: start-of-season time (SOST), start-of-season NDVI (SOSN), end-of-season time (EOST), end-of-season NDVI (EOSN), maximum NDVI (MAXN), maximum NDVI time (MAXT), duration of season (DUR), amplitude of NDVI (AMP), and seasonal time integrated NDVI (TIN) (Gu et al., 2010). They derived these values from the MODIS NDVI band and Advanced Very High Resolution Radiometer (AVHRR) data, for a time period of five years. Their method of establishing "normal" phenology conditions is useful to my study. Gu et al. took the median value from all of the metrics and set that as their baseline (2010). M.K. Ridd's 1995 study, published in the International Journal of Remote Sensing, developed a model that used remotely sensed spectral information to interpret urban ecosystems relationships with vegetation surface soils and impervious surface soils. Estes & Star point out the importance of utilizing quantitative information of environmental parameters to determine how to qualitatively describe those parameters (as cited in Ridd, 1995). This method of describing data takes out the variability of describing environmental impacts in terms of land use which are human related and interpreted. Studies focused of biophysical actions should consult objective quantitative measures when defining classifications and distinctions between said classifications (Ridd, 1995). Since Ridd's study was focused on showing the more discrete relationships within complex urban environments through quantitative means, it will apply directly to my study due to the change in environments throughout the watershed and significant intimacy of surface water bodies to pollution sources in karst terrain.

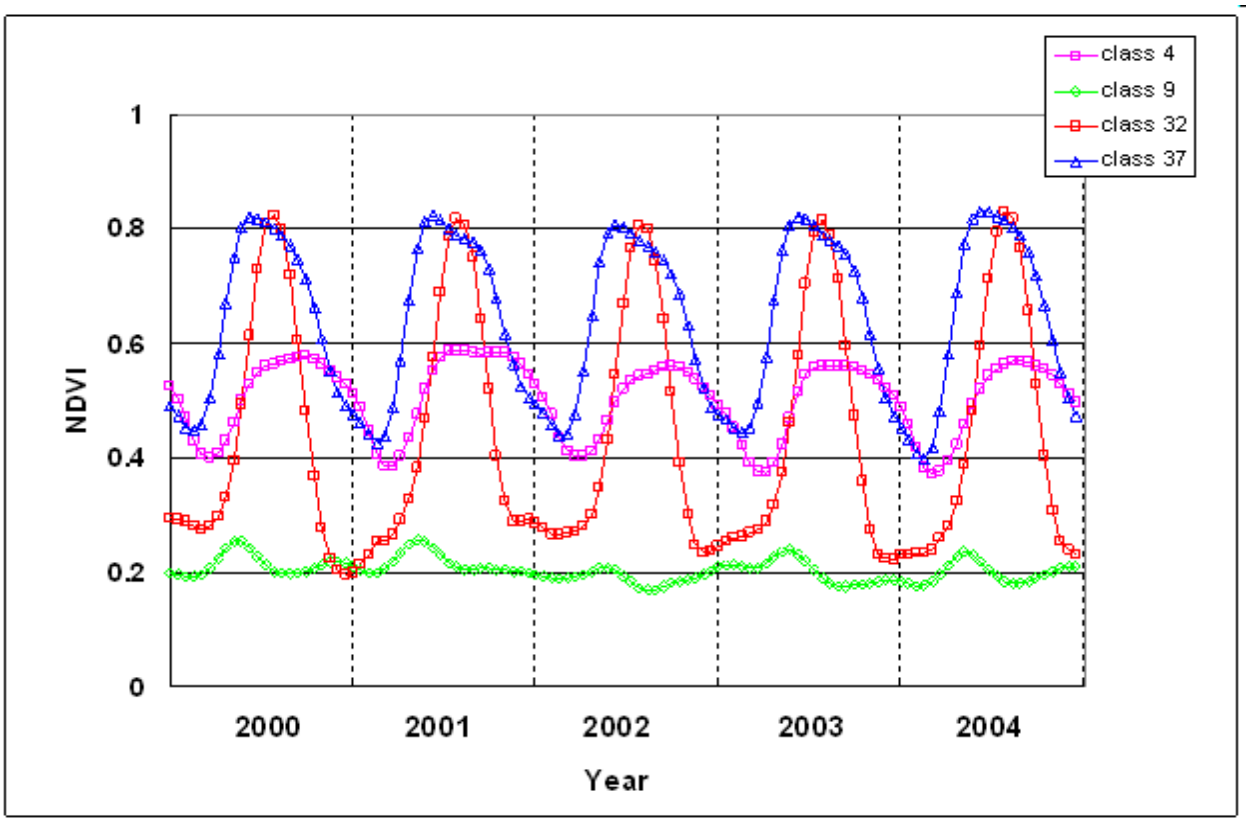


Figure 3: NDVI Phenology Curves

Examples of NDVI curves of different classes or clusters defined by Gu et al (2010). This is one way to visualize the annual and inter-annual land surface phenological cycles associated with spectrally defined classes. Class 9 includes the cycle of evergreen trees.

These relationships have a spatial component at their core. As shown in Griffith et al. (2002), LULC and landscape features such as slope are the main contributors to water quality and even more so with areas dominated by agriculture (2002). That study also emphasizes that water quality and other environmental quality parameters can be analyzed more cost effectively and efficiently on a landscape-level within ecoregions, given the continual advancements in remote sensing technologies (Griffith et al., 2002). These remote sensing assessments are all based upon how the data gathered on certain environmental variables relates to either *in-situ* or remotely sensed data of other environmental variables. Ecoregions are geographically bounded regions based on similar factors such as physiography, vegetation, land use, geology, and soils

(Griffith et al., 2002). Analysis that is high in temporal resolution is key to account for phenological, interannual, and anthropogenic changes according to Hobbs 1990 (as cited in Griffith et al., 2002). When temporal specificity and objectively defined data points are used simultaneously to assess environmental management practices of a region, a much better picture can be drawn to use for reference in decision making processes affecting that region. Maintaining full temporal scale within data sets can allow major environmental indicators to be shown, especially when referring to LSP data sets. This is because phenologic cycles, a.k.a seasonality, have been shown to correlate with surface water quality parameters in different ways (Johnson et al., 1997). The Michigan study area analyzed in this study was shown to have different influential factors for water quality based on the season (Johnson et al., 1997). It was shown that landscape structures such as slope, and land use patch density were more influential factors on total nitrogen in surfaces waters during the summer time, but not so much during autumn (Johnson et al., 1997). They attributed the seasonality of this relationship to the varying average discharge rates recorded, which is directly influenced by the landscape of the catchment and the mechanisms of specific LU type density (Johnson et al., 1997). Phosphorus and total suspended solid concentrations weren't shown to be as seasonal when correlated with other dependent variables of the study, on a catchment level, but were correlated to variables such as LU during the summer on a sub-catchment level (Johnson et al., 1997). Sensitivity indicators are rooted in temporal responses to change within an ecosystem, and are shown when a particular stress is readily responded to by an ecosystem (Kelly & Harwell, 1990). Empirical analysis of NDVI for LSP classification and the linear relationship to surface water nutrient loads will be an ecosystem sensitivity indicator within the HWIRW, because of this studies preservation of a full temporal scale for the past twenty years (Griffith et al., 2002; Kelly & Harwell, 1990). When this

data is reduced to a parcel by parcel level while still maintaining temporal range, it allows planners and others to accurately identify useful indicators of environmental change.

Image Segmentation for Data Reduction

Traditional pixel based methods for time series LULC or LSP classification worked well when spatial resolution was much coarser, due to the ease of generalization and smaller amount of data contained within a raster. Once sensors became more accurate and powerful, these pixel based methods fell short of accurately classifying different remote sensing products due to the higher spectral complexity produced by these higher resolutions (Jensen, 2016). When that higher resolution is reached it becomes more necessary when evaluating and classifying a pixel to take into account the pixels around it for contextual information (Jensen, 2016). Thus geographic object-based image (GEOBIA) was developed, to facilitate growing sensor technology and product quality (Jensen, 2016). GEOBIA also provides an approach for spatial dimensionality reduction for large raster datasets, similar to those often needed in LSP related studies (Bunker et al., 2016; Jensen 2016). These algorithms are extremely useful for environmental studies in regards to their mathematical evaluation of raster pixel values based on set criterion that are geographic in nature. Image segmentation is the basis for GEOBIA, because it breaks the raster image up into homogenous sections known as, image objects (Jensen, 2016). While there are multiple image segmentation algorithms, Jensen 2016 shows that most of them belong to two classes: edge-based and area-based. These algorithms do not consult spectral and spatial information simultaneously. Baatz and Schape addressed this problem in their 2000 study aimed at an optimization approach for high quality multi-scale image segmentation; in which they developed a color criterion and a shape or spatial criterion that would be evaluated for each individual pixel and their neighbors (as cited in Jensen, 2016). The only part of this approach to

spatial dimensionality reduction that is anecdotal is the weighting of criterion because of relevance. The process of image segmentation for remote sensing purposes has this basic process according to Jensen (2016):

> The user specifies the spectral (color) and spatial shape parameters (compactness and smoothness) criteria and the neighborhood function logic (Definiens, 2007). A specially designed heuristic algorithm then applies these criteria to individual pixels in the scene and, in effect, grows homogenous regions (or, if you like, regions with specified amounts of heterogeneity). Once a segment patch exceeds the user-specified parameters, it stops growing. The final result is a new segmented image consisting of image objects (patches) that contain relatively homogeneous spectral and spatial characteristics (Jensen et al., 2006).

The reason image segmentation works well for utilization of remote sensing datasets is because of the incorporation of spatial aspects of raster based data, instead of just the spectral values within pixels (Jensen, 2016). The image objects that are produced by image segmentation are based on geographic principals of adjacency and shape, who in turn allow the classification of those objects to be applied in a geographic nature.

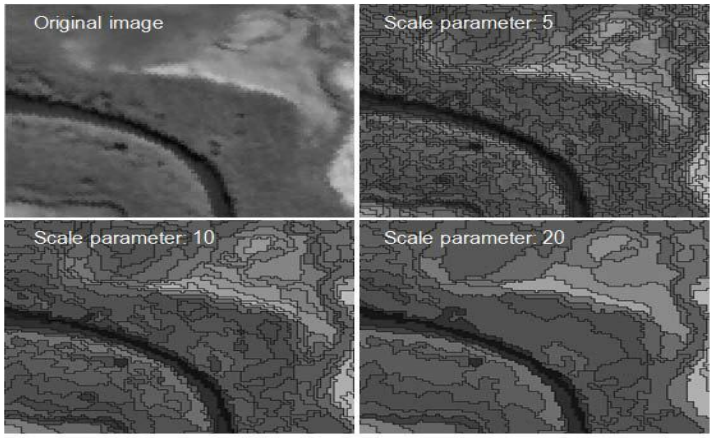


Figure 4: Scale Parameters and Region Growing

This image shows how different scale parameters will affect the region growing process based on a user specified level of homogeneity within the image objects. The higher the scale parameter the looser the criteria for homogeneity. Adapted from Kamal & Phinn (2011).

Environmental studies for the most part are geared towards understanding how certain forces in nature or in human activity, will affect the natural processes that interact with each other in particular areas. A lot of these relationships have spatial components that are defined for different landscape-ecology metrics (Jensen, 2016). These metrics can be calculated within each image object produced by the segmentation process. For example, Strasser & Lang wanted to develop a semi-automated method of classifying and mapping complex riparian forested habits by using high quality WorldView-2 satellite imagery (2014). The aim of their paper was to develop a tree species identification product using image segmentation techniques and then classify that image based on widely adopted GEOBIA classification schemes, in an effort to develop a hierarchy of forest type relationships within the riparian zone (Strasser & Lang, 2014). Since the information within the image objects contain spectral and spatial information of the area it allowed the products to be utilized in known forest habitat relationship models (Strasser & Lang, 2014). This shows GEOBIA classification can be very effective at fully representing complex relationship within environments. Unit-less scale factors regulate the amount of homogeneity allowed in the region growing portion of the image segmentation process (Bisquert, Begue, & Deshayes, 2014). These scale factors should be set based on the amount of desired spatial detail of the independent variable, in this case spectral homogeneity (Bisquert, Begue, & Deshayes, 2014).

Methods

Study Area

The HWIRW is a hydrologic unit code (HUC) 10 watershed, as delineated by the USGS, located in the northwest region of Arkansas. This study area encompasses just over

73,000 acres of mostly rural land with an estimated population density of 110 persons per square mile, as of the year 2000 (CAST, 2006). That number has surely increased in the past eighteen years, as the entire NWA region has grown exponentially in that same time frame (Rosa, 2017). As described in its name, the HWIRW is where the Illinois River begins its 145 mile route, flowing northwest and on into the Ozark Mountains of Oklahoma. The Illinois River is the main surface water body within the watershed and has several tributaries that flow into it throughout the 113 square mile catchment (Center for Advanced Spatial Technologies [CAST], 2006). Its geographic location and attributes make it an interesting and suitable area for this type of study, because it offers multiple different characteristics of geology, geography, land cover, topography, and much more. This will allow the results from this study to be applied or referenced for numerous other, future studies. The HWIRW is a primarily rural area within the Ozark region of the state of Arkansas. Natural make-up of this area is characterized by unique karst geologic features and dominated by an oak-hickory forest composition. This study area is surrounded by the Ozark National Forest (ONF) and contains many of the same natural characteristics. ONF is made up of over a million acres of mainly forested land and that is within the Boston Mountain range and the southern end of the Springfield Plateau (United States Forest Service [USFS], 2018).

These physiographic regions were part of one geologic unit that uplifted together to form plateaus; the cause of their topography in present day, is the easily water soluble limestone and dolomite rock that make up the vast majority of the regions (USFS, 2018; Monroe, 1970). The rivers and streams that were inevitably formed by these erosion patterns are what gave the

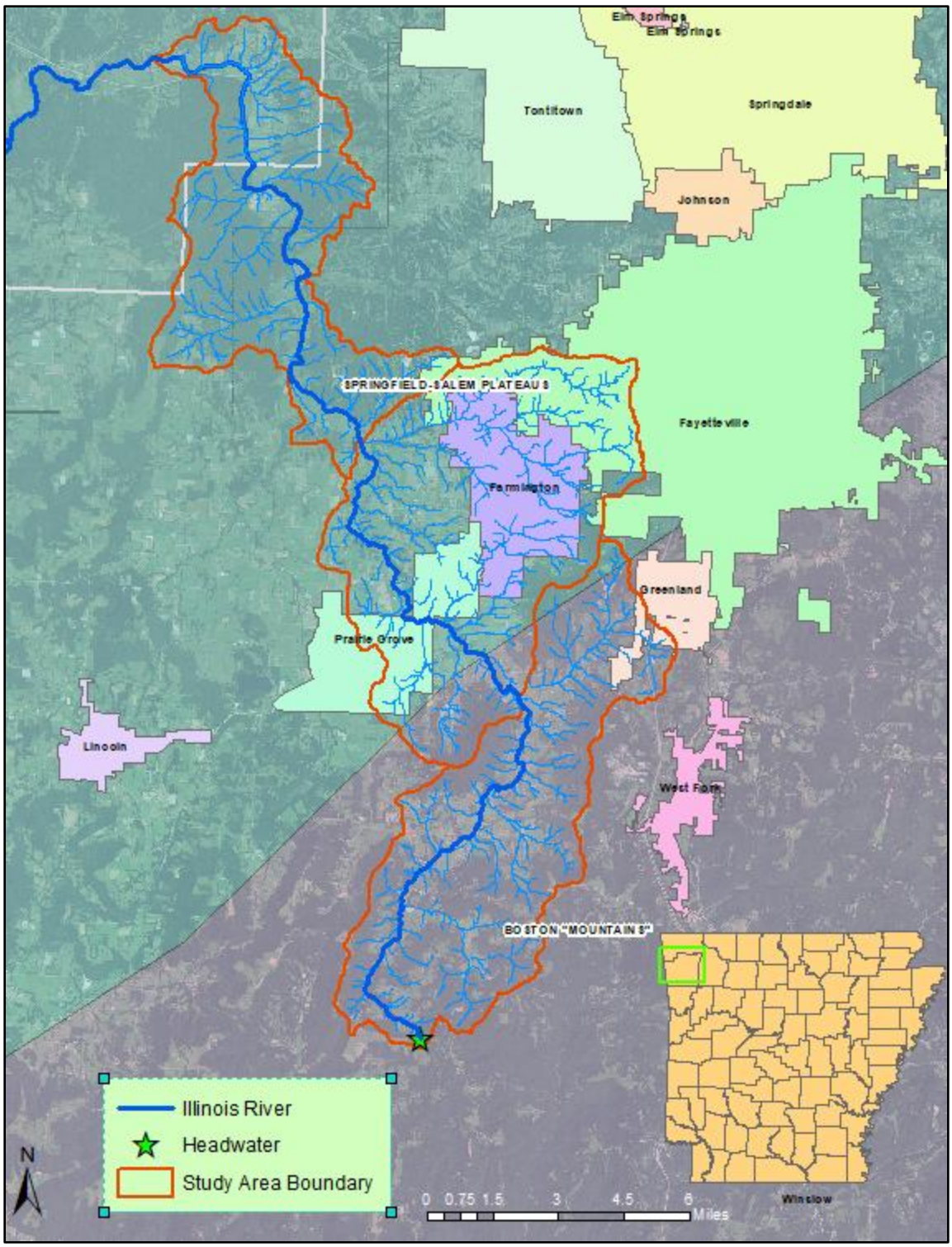


Figure 5: Study Area Overview

This is a map showing the overall layout of the study area (outlined in orange) and surrounding areas. It shows how the Illinois starts in a rural area, flows north through a more urban area, and on into another rural area. Data courtesy of AGIO (2017).

Boston Mountains and Springfield Plateau their steep valleys and exposed groundwater springs that characterize the region today (USFS, 2018). Once livestock farming arrived with the Scotch-Irish frontiersmen of early America the landscape in the Ozarks, including the HWIRW, began to change to facilitate these farming practices. Due to the rocky soil and steep topography of the region it has never been susceptible to growing massive amounts of food crops, but it is, however, highly suited for the growing and sustaining of livestock due to its natural grasses and abundance of natural sources of water. The expansion of these practices as history progressed led the shift from upland oak-hickory forest to maintained pastures for livestock grazing. This was especially prevalent in the flatter areas, such as the HWIRW, near rivers where people originally settled many years ago (Rafferty, 1980). Ever expanding populations within the NWA region have caused rapid land cover change, especially in the past twenty years. Population growth has also directly correlated with expansion of livestock agriculture in the region (Brahana, 2011; Brion et al., 2010). The Illinois River has also been the source of controversy between Arkansas and Oklahoma, since the late 1990's- early 2000's (Haggard & Sorens, 2006). Essentially Oklahoma was blaming Arkansas for high levels of phosphorus, causing eutrophication, within their portion of the Illinois River. This was attributed to higher levels of municipal effluent with the HWIRW that then carried downstream into Oklahoma (Haggard & Sorens, 2006). The controversy went all the way to the U.S. Supreme Court (Arkansas versus Oklahoma, 503 US 91) which lead to a ruling that would allow the EPA to require upstream states to be in accordance with downstream states water quality standards, at the river's border crossing (Haggard & Sorens, 2006). In an effort to streamline and enhance the quality of monitoring these parameters this study has developed a possible methodology using remote sensing capabilities.

All of these factors together make the HWIRW a more than qualified study area for this type of environmental monitoring study.

Data

All of the input data used for this study was retrieved from open source web services. This input data included the Landsat 8 and Landsat 5 satellite images from January 1, 1999 through March 15, 2018, as well as total inorganic nitrogen and total phosphorus levels at two separate locations within the HWIRW for the same temporal scale (Appendix 1). Satellite imagery data was collected using a cloud based platform called Google Earth Engine (GEE), and the water quality information was harvested from Arkansas Department of Environmental Quality (ADEQ) database.

GEE is a cloud based geospatial analysis platform that has been developed within the past couple of years, in an effort to provide high levels of computational power for large geospatial projects in an open source format (Gorelick et al., 2017). Using the enormous amount of computational power provided by Google allows users to analyze incredibly large datasets for different geospatial purposes (Gorelick et al., 2017). GEE also provides access to multiple different data collections that are sourced from multiple different public and private entities, such as USGS and NASA. This eliminates the hassle and confusion of having to consult multiple different data sources when attempting to gather data for analysis. Gorelick et al. (2017), gives this basic overview of the GEE platform:

> Earth Engine consists of a multi-petabyte analysis-ready data catalog co-located with a high-performance, intrinsically parallel computation service. It is accessed and controlled through an Internet-accessible application programming interface (API) and an associated web-based interactive development environment (IDE) that enables rapid prototyping and visualization of results.

> The data catalog houses a large repository of publicly available geospatial datasets, including observations from a variety of satellite and aerial imaging systems in both optical and non-optical wavelengths, environmental variables, weather and climate forecast and hindcast, land cover, topographic and socio-economic datasets. All of this data is preprocessed to a ready-to-use but information-preserving form that allows efficient access and removes many barriers associated with data management.

In the case of this study GEE was used to query Landsat 5 TM and Landsat 8 OLI surface reflectance image collections in order to retrieve NDVI products from those sensors for the temporal scale. As stated in the quote above, the IDE was used to import data collections that would be needed to develop the desired NDVI data products. The two surface reflectance image collections along with USGS Watershed Boundary datasets were imported by using the search bar at the top of the IDE window which provided detailed metadata from the entity that provided the data. GEE allowed the data to be harvested in a format and scale range that was applicable to this study. The HWIRW only covers around 113 square miles while a Landsat scene covers over 12,000 square miles. Normally an analyst would have to find which row/path combination contained their study area and download that entire scene(s) for whatever temporal scale they desired. When trying to analyze a full range of data for a twenty year temporal scale, that method would produce a massive amount of un-wanted data, that would have to be manually reduced using other GIS and image processing techniques. Using GEE's geospatial analysis tools, that 12,000 square mile scene could be filtered to the spatial bounding box of the HUC10 code 1111010304, which is the intended study area of the HWIRW. That bounding box was retrieved based on filtering the metadata within the USGS feature collection for watershed boundaries. Additionally a 300 meter buffer was applied to the spatial bounding box of the HWIRW, using a built-in buffer function within GEE's IDE, in an effort to preserve spectral information within the pixels that are in close proximity to the study area. This is necessary for a more robust analysis of spatial relationships with in the image segmentation process applied later in the study

(Jensen, 2016). Filtering the metadata of the Landsat image collections is also necessary to obtain all images within the collection that fall within the desired temporal scale of the previous twenty years. Each image within the collection has a time stamp for the date of capture. GEE's built-in 'filterDate' function requires two dates, in year-month-day format, the first argument is the start date and the second argument is the end date. The Landsat Surface Reflectance image collection contains multi-band raster images that correspond with the bands of each sensor; the TM for Landsat 5 and the OLI for Landsat 8. Developing NDVI values requires applying a normalized difference equation using the red and NIR band values for each pixel in an image. GEE has a built-in 'normalizedDifference' function that can be applied to two selected bands within a multi-band raster. The 'normalizedDifference' function is known as a spectral transformation, according to GEE's API, that can be applied to a single multi-band raster (Gorelick et al., 2017). To retrieve proper NDVI values the NIR band should be listed as the first argument and the red band should be listed as the second argument. The purpose of this study, though, is to analyze NDVI trends over the past twenty years. This requires multiple images to be analyzed but the built-in 'normalizedDifference' function only analyzes one raster at a time. Functions and mapping are how to get around this. Custom functions in GEE are used to make coding more efficient and organized (SEVIR-Mekong, 2017). Functions used in GEE are just sections of code that may be used heavily in different projects, or any section of code that is lengthy would make repetition of that section highly inefficient. They are also necessary when an iterative process called "mapping" is desired to be used. The GEE doc called 'map' is a JavaScript iterator that can apply a function to an image or feature collection within GEE. This study applied an NDVI function to each Landsat sensor's series of images by using the 'map' doc. They had to be separate because of the different band designations between TM and OLI.

Table 1: Landsat 5 Band Designations

Band designations and electromagnetic (EM) spectrum ranges for each band captured by Landsat 5 TM. Adapted from USGS (2017).

Band	Wavelength	Description
1	0.45μm – 0.52μm	Band 1 (blue) surface reflectance
2	0.52μm – 0.60μm	Band 2 (green) surface reflectance
3	0.63μm – 0.69μm	Band 3 (red) surface reflectance
4	0.77μm – 0.90μm	Band 4 (near infrared) surface reflectance
5	1.55μm – 1.75μm	Band 5 (shortwave infrared 1) surface reflectance
6	10.40μm – 12.5μm	Band 6 (brightness temperature)
7	2.08μm – 2.85μm	Band 7 (shortwave infrared 2) surface reflectance

Once the functions were developed for the different Landsat sensors, they were mapped to the image collections. The 477 images that were filtered out by the date range metadata filter, are now single band images that contain NDVI values between -1 and 1 for each pixel. The NDVI images were now ready for export but GEE did not have a built-in doc for multiple image export without applying what's known as a 'reducer'. GEE 'reducers' are used to reduce image collections to a single raster layer that contains statistical information the pixels in each image of a collection (SEVIR-Mekong, 2017). An example of this would be calculating the mean or median pixel values of an image collection to summarize a temporal range of values. This study wants to maintain the NDVI pixels values recorded in each of the 477 images, instead of reducing them to an average for each year or bi-monthly, like previous methods of data reduction in similar studies (Bunker et al., 2016).

Table 2: Landsat 8 Band Designations

Band designations and EM spectrum ranges for each band captured by Landsat 8 OLI. Adapted from USGS (2017).

Band	Wavelength	Description
1	0.435μm – 0.451μm	Band 1 (ultra blue) surface reflectance
2	0.452μm – 0.512μm	Band 2 (blue) surface reflectance
3	0.533μm – 0.590μm	Band 3 (green) surface reflectance
4	0.636μm – 0.673μm	Band 4 (red) surface reflectance
5	0.851μm – 0.879μm	Band 5 (near infrared) surface reflectance
6	1.566μm – 1.651μm	Band 6 (shortwave infrared1)
7	2.107μm – 2.294μm	Band 7 (shortwave infrared 2) surface reflectance

This was accomplished in consultation with the GEE developers forum, where users and developers can interact openly in sharing of code and techniques. Rodrigo E. Principe developed a function that would utilize an iterative 'for' loop to query the metadata needed for the function to designate between images within a collection and export them out individually. This study utilized the export function authored by Principe to export all images within each Landsat Surface Reflectance image collections, which have been clipped to the buffered study area bounding box, and calculated for NDVI in each pixel. The images are exported to a Google Drive account where they are then moved to a file where they can be interpreted manually for atmospheric pollution that would affect trend analysis for the time series. The Landsat Surface Reflectance image collections provided by the USGS were held to a threshold for atmospheric pollution based on the LEDAPS atmospheric correction algorithm and the CFMASK cloud,

snow, shadow, and water masking algorithm (United States Geologic Survey [USGS], 2017). These thresholds were essentially based on a percentage of pollution present within an entire Landsat scene (USGS, 2017). Since this study is cutting down each scene to the buffered bounding bow of the HWIRW, manual interpretation of each image was needed to ensure that atmospheric pollution was not just present in that small section of the entire scene, which may not have caused the scene to exceed the threshold. Once they were in a folder on the computer each image was manually interpreted for atmospheric pollution such as clouds and weather events. If any pollution was noticed, the image was then deleted and not utilized for time series analysis of phenologic cycles. After eliminating duplicates and unqualified images the original set of 477 images was reduced down to 122.

Water quality information from the HWIRW was collected from the historical database that is managed and populated by the ADEQ. The database was accessible from their website and it includes multiple different water quality parameters that were categorized into ten parameter groups: bacteriological, biological, field, metals, organics, nutrients, surrogate, solids, wet chemistry, and anions, minerals, TDS. The parameter group relevant to this study is the nutrient group. Regions dominated by livestock agriculture are prone to contributing to excess nutrients into surrounding surface waters, which causes eutrophication and other environmental problems (Leidy & Morris, 1990; ANRC, 2009; Brion et al., 2010). The nutrient parameter group includes levels of: Ammonia-nitrogen, inorganic nitrogen, orthophosphate, and total phosphorus (ADEQ). Inorganic nitrogen and total phosphorus are shown to be the biggest contributors to environmental problems as well as being the most easily attenuated by plant root systems and the interflow zone. This is the reason why this study analyzed these parameter trends for the past twenty years. One of the biggest limitations encountered in this study was the

lack of multiple locations with nutrient levels for the past twenty years. Most locations seemed to only be recorded on a need to know basis for various reasons. Only two locations in the HWIRW had recorded nutrient levels on a regular basis for the past twenty years. Illinois River at Savoy (ARK0040), which is located in the middle portion of the study area just north of Farmington, and Illinois River south of Siloam Springs (ARK0006), which is located near the Arkansas/Oklahoma border. ARK0006 was chosen for two reasons; because of the availability data for the temporal scale, and its location downstream of the study area and near the Oklahoma line where quality must be closely monitored (Figure 8). The data from those sites were exported from ADEQ's online database in a CSV format to be useable in Excel for regression analysis.

Image Segmentation

The image segmentation part of this study was conducted in two steps. First the images that were processed and exported from GEE had to be prepped for seamless use within Trimble's eCognition. When images are exported from GEE the data contained within the pixels are in the same data structure as the image collection that was analyzed. In this case the Landsat 5 and 8 surface reflectance images were in a 32-bit floating point data structure. That means that all of the values contained within the pixels would fall into a range of -3.402823466E+38 through 3.402823466E+38 (ESRI). The same 32-bit floating point data structure was maintained after NDVI values were calculated for all pixels in the image collection. While this structure provides a high level of precision it also creates a large amount of data to be stored and would slow the segmentation process within eCognition. To facilitate computational speed, the raster images were processed in a few ArcGIS

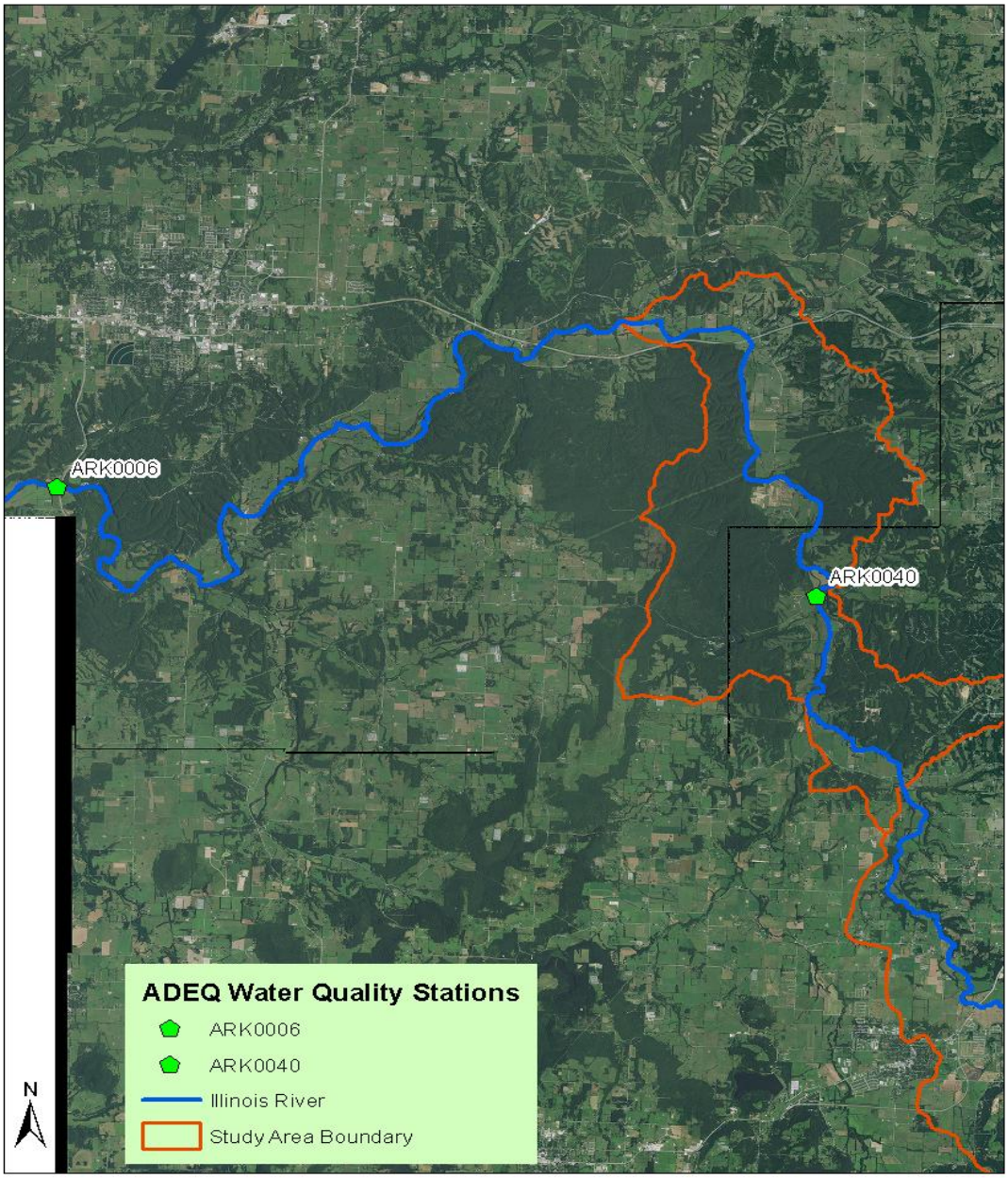


Figure 6: Water Quality Monitoring Stations

A map showing the locations of the two water quality monitoring sites that were documented and analyzed in this study. ARK0006 is the Illinois River sout of Siloam Springs and upstream is ARK0040, the Illinois River at Savoy. Data courtesy of AGIO (2017).

Table 3: Raster Data Structure

This table shows the different bit depth types and the range of values each type is capable of storing (ESRI).

Bit depth	Range of values that each cell can contain
1 bit	0 to 1
2 bit	0 to 3
4 bit	0 to 15
Unsigned 8 bit	0 to 255
Signed 8 bit	-128 to 127
Unsigned 16 bit	0 to 65535
Signed 16 bit	-32768 to 32768
Unsigned 32 bit	0 to 42949967295
Signed 32 bit	-2147483648 to 2147483648
Floating-point 32 bit	-3.402823466e+38 to 3.402823466e+38
Unsigned 64 bit	0 to 18446744073709551616

models to change their data structures from 32-bit floating point to 16-bit unsigned integers. In the first model, the file containing all of the exported and vetted NDVI images were sent through an iterator that first added 1 to all of the pixel values using the “Plus” tool. This was done because all of the NDVI values ranged from -1 to 1, so when 1 was added to all of the values it ensured that no values would lie within the negative range. Then those values were multiplied by 5000 using the “Times” tool. This was done to maintain more precision within the range of NDVI values. Those images were then converted to integers through truncation using the “Int” tool. These three steps of data manipulation converted all of the pixels values within each image to a 16-bit unsigned integer data structure (Figure 10).

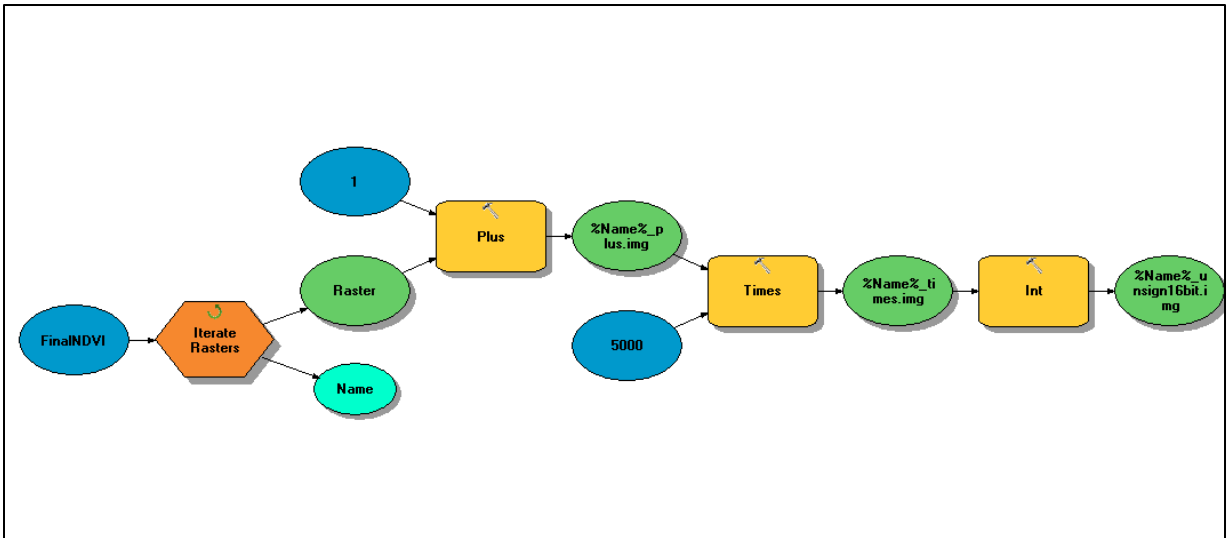


Figure 7: Data Conversion Model

Structure of the ArcGIS model that was used to convert all of the rasters to a 16-bit data structure.

Next the newly converted images needed to be composited into one multi-band raster in order for the image segmentation to consult the full temporal range of data for each pixel when analyzing spectral homogeneity during the "region growing" process (Baatz & Schape, 2000; Jensen, 2016). Using the "Composite Bands" tool each newly converted image was composited as a single NDVI image per band in chronological order. This created a 122 band 16-bit unsigned raster image that was now ready to be segmented within eCognition (Figure 11). Once the images were converted to a more computational friendly format and composited together, the multi-band raster was then uploaded to eCognition to conduct image segmentation. In eCognition the composited image was subjected to a custom "rule-set" that was formulated specifically, for this study. Since this study is based upon spatial relationships of NDVI values over time, the spatial criterion was considered to be more relevant. Spatial criterion in eCognition is addressed by changing the scale factor of the rule set (Definiens, 2007).

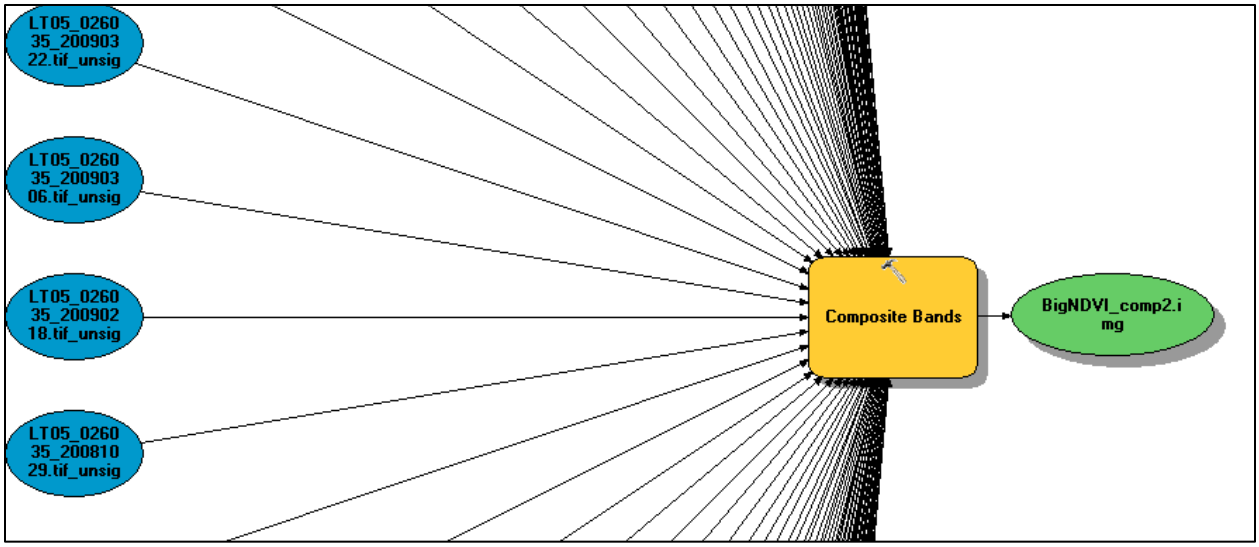


Figure 8: Composite Raster Model

Structure of the ArcGIS model that was created to composite all 122 converted NDVI images into one muti-band raster for image segmentation analysis. Each black line coming into the "Composite Band" box is one of the 122 images.

To assess what a proper scale factor should look like for this study, multiple scale factors were tested for the average number of pixels contained within the objects. This number was decided to be the number of pixels that represents the average real estate parcel size within the study area. Since this study is focused on finding phenological patterns that effect surface water quality, roots of the potential problems or positives need to be on a parcel level so that patterns could better located. This means that the average size of the image objects contained in the image segmentation scene need to be equal to the average size of real estate parcels within the HWIRW. The median parcel size, 0.70 acres, was also consulted, but it was determined that the high concentration of small parcels in the urban areas would cause larger rural parcels to be over represented. Rural parcels are the main focus for monitoring antrhropogenic changes. Washington and Benton county assessor's office publish real estate parcel datasets to the Arkansas Geographic Information Office's (AGIO) web server, where data for the state of

Arkansas can be accessed publicly. Using these datasets, all of the real estate polygon shapefiles were selected, using ArcMap 10.5.1, by their location being within the study area boundary. Once selected the polygons were exported as their own shapefile and the acreage information was then queried from the attribute table; using the field statistics tool, it was shown that the average parcel size within the HWIRW was around 4.5 acres. Since eCognition object statistic tools work in pixels, 4.5 acres needed to be translated to number of pixels. A pixel from the Landsat Surface Reflectance image collection will have a resolution of 30m for an NDVI image. That means that each square pixel is 30m by 30m, therefore the area of the pixel in 900 square meters. There are 4046.86 square meters in one acre, so multiply that by the average real estate parcel size and you get 18,210.87 square meters. That number divided by 900 is equal to 20.23, which is the number of Landsat pixels within the average real estate parcel for the study area. Once this average pixel number was reached the different scale factors could be tested until the average pixel number within each image object was reached. Scales of 10, 20, and 30 were tested based on the Image Object Information statistical tool that would display the average number of pixels in each image object within the scene. Scale 30 was the scale factor that most accurately represented the desired average size of the image objects within the scene, with an average pixel number of 20.83 (Figure 12). The newly generated 41,238 image objects were then exported as polygon shapefiles, with mean NDVI values of all the pixels for each of the 122 dates per polygon.

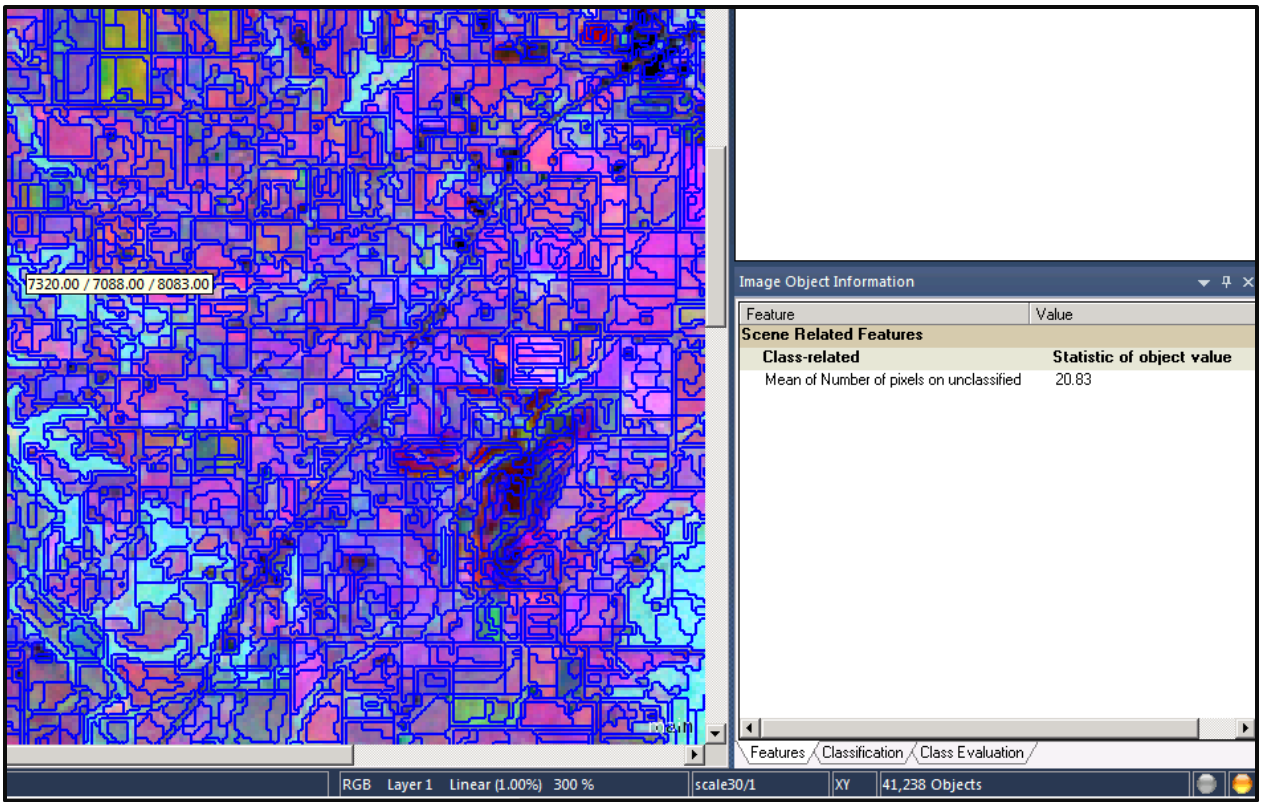


Figure 9: OBIA Example

Multilayer image segmentation from Trimble's image segmentation software, eCognition. The colorful parts on the left are the polygons that were generated by the image segmentation process at a scale level of 30. The Image Object Information box on the right shows the statistic calculated for each object, the average number of pixels.

Clustering

To further understand the spatial patterns that might exist within the data, clustering was conducted based on similar logic described by Bunker et al. (2016). While a number of clustering options are available, Bunker et al. (2016) incorporated a combination of ISODATA (using ERDAS Imagine) and a cluster optimization technique based on transformed divergence. In the present study, the popular RStudio program within the R statistical computing environment was incorporated. RStudio was developed to be an IDE for the rapidly growing community of analyst using the open source R statistical language for a bevy of different forms of analysis (RStudio). Within RStudio there are different "packages" that contain multiple

functions that can be applied to different datasets. This study utilized the K-means Clustering function within the version 3.4.4 *stats package* of RStudio. K-means Clustering is defined within the help section of the *stats package* as follows:

> The data given by x are clustered by the *k*-means method, which aims to partition the points into *k* groups such that the sum of squares from points to the assigned cluster centres is minimized. At the minimum, all cluster centres are at the mean of their Voronoi sets (the set of data points which are nearest to the cluster centre).

In the case of this study, the *x* mentioned in the definition of the function, would be the attribute table that comes from the newly exported polygon shapefile containing all of the NDVI values. There are multiple K-means algorithms in the world of statistics but the default algorithm used in RStudio's kmeans function was developed by J.A. Hartigan and M.A. Wong in an optimized form, in 1979. The main parameter set by the analyst when using K-means as a clustering method is setting how many clusters the data points will be separated into. To find the optimal number of clusters needed for a dataset this study used the "Elbow Method" of optimization. The "Elbow Method" is used to ensure that the dataset can be sufficiently represented without introducing unwanted 'noise' or over representation by the presence of too many clusters. Anand (2017) gives this explanation of what the "Elbow Method" (Figure 13) is:

> The elbow method looks at the percentage of variance explained as a function of the number of clusters: One should choose a number of clusters so that adding another cluster doesn't give much better modeling of the data. More precisely, if one plots the percentage of variance explained by the clusters against the number of clusters, the first clusters will add much information (explain a lot of variance), but at some point the marginal gain will drop, giving an angle in the graph. The number of clusters is chosen at this point, hence the "elbow criterion". This "elbow" cannot always be unambiguously identified (Anand, 2017).

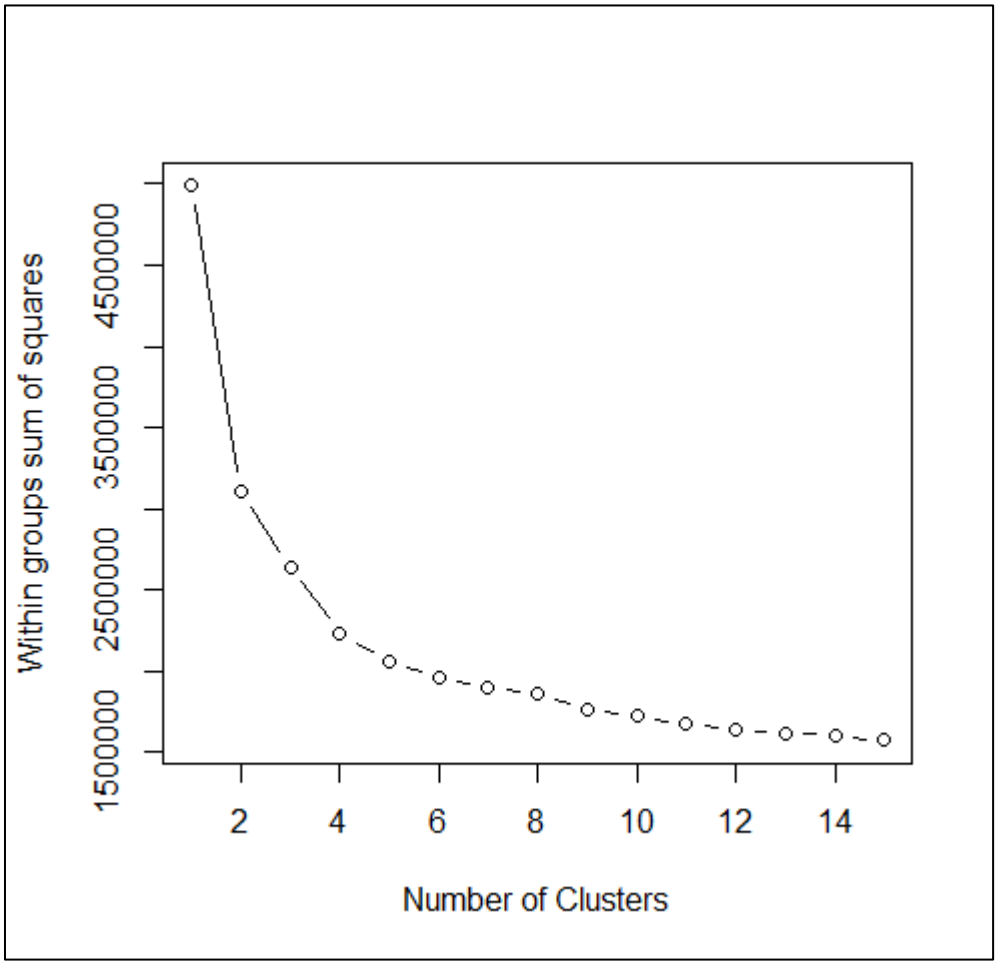


Figure 10: Elbow Method of Cluster Optimization

This graph shows the results of the cluster optimization process known as the "Elbow Method" (S Anand, 2017). The optimal number is located at the point where the line begins to flatten out, which is near the number 5.

After running the cluster optimization code in RStudio, using the attribute table produced by eCognition, it was determined that five clusters would be an optimal representation of the data (Appendix 2). Once the cluster number was decided, then the actual *kmeans* function could be applied to the dataset in order to produce the clusters of data (Appendix 2). The function produced a cluster designation of 1-5 for each of the 41,238 polygons. This cluster designation was then added to the attribute table as a new column, so that further analysis and representation

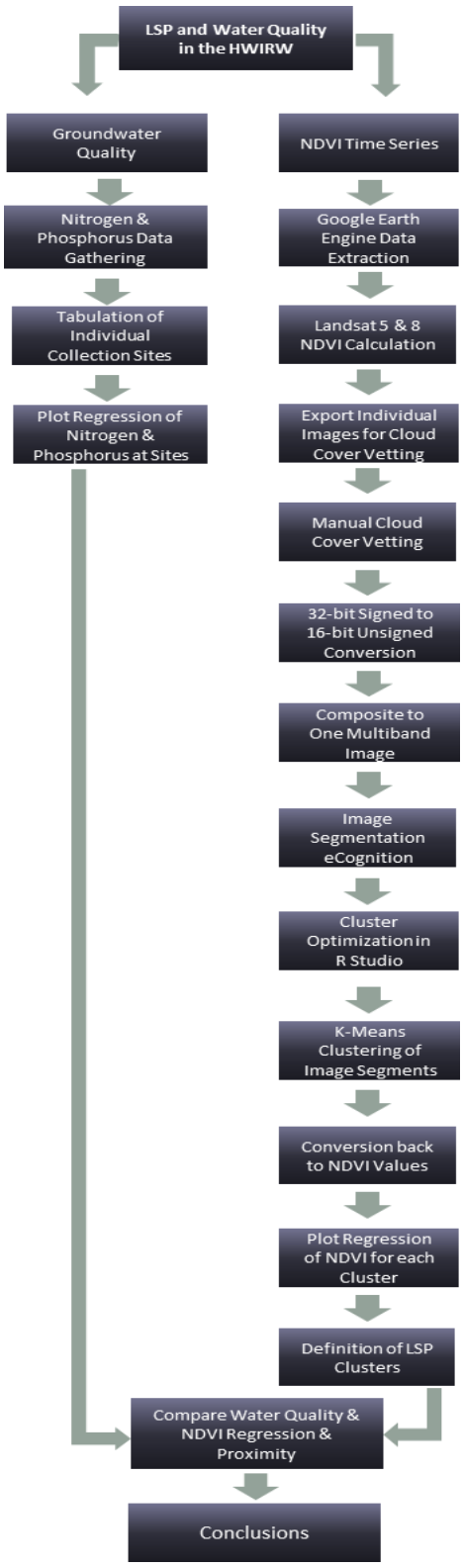


Figure 11: Methodology Flow Chart

of the clusters would be easier. RStudio was also used to convert the NDVI values back into the normal range of -1 to 1 (Appendix 2). This was done using a 'for' loop that subtracted 1 and divided by 5000 for all values. This was applied after the attribute tabular data was split up into separate tables based on cluster designation.

Results and Discussion

GEOBIA showed to be an effective means of data reduction and brought a new prospective to medium resolution satellite time series analysis, which translated the results of the segment trends to closely mirror the earth's surface. The results of the image segmentation produced 41,238 polygons within the bounding box of the study area, with 12,714 of those being within the geographic boundaries of the study area. By making sure that the scale factor would produce, on average, a polygon that was the same size as the average real estate parcel, studying the different clusters after unsupervised classification helped make more accurate conclusions on a parcel level.

Unsupervised K-means clustering in RStudio was conducted upon the entire dataset of 41,238 polygons produce from image segmentation. This was done to maintain consistency of full dataset analysis from segmentation to clustering. The polygons were reduced to the geographic bounds of the study area and analyzed by cluster, after full dataset analysis in RStudio. K-means clustering was optimized to do five cluster designations. Cluster 1 contained 3920 polygons, Cluster 2 contained 280 polygons, Cluster 3 contained 3131 polygons, Cluster 4 contained 5383 polygons, and Cluster 5 contained 1230 polygons.

Using Microsoft Excel, water quality and NDVI trends within clusters were analyzed graphically and developing trend lines for all datasets. All of the polygons within the cluster

were averaged for each date and those numbers were represented as markers on the graphs. It was interesting to find that all of the clusters, except cluster 5, had the exact same slope and just slightly different intercepts for the time series. This made their graphs look almost identical. Also to get a better representation of each cluster, a total average was taken of all date averages within the clusters. This number was used in accompany with the slope factor to define each clusters spatial relationship to water quality based on up or downstream from ARK0040.

Table 4: Cluster Statistics Breakdown

Cluster designations and the corresponding number of polygons, percentages in north, middle, and south subcatchments, total average NDVI, and slope of NDVI change from 1999-2018.

Designation	Number of Polygons	% in North	% in Middle	% in South	Total Average NDVI	Slope Factor
Cluster 1	3920	35	23	42	0.567	0.000002x
Cluster 2	280	25	56	19	0.46	0.000002x
Cluster 3	3131	17	58	25	0.529	0.000002x
Cluster 4	5383	17	53	30	0.549	0.000002x
Cluster 5	1230	19	62	19	0.493	0.000003x

Water quality was analyzed in terms of inorganic nitrogen and total phosphorus for each of the two locations. ARK0040 was used as a determining point within the study area, since it is directly downstream from the more urban middle section. ARK0006 was used to determine how well the study area contributed to the overall attenuation of nutrients after flowing through predominantly rural forested area after ARK0040. For both inorganic nitrogen and total phosphorus, the levels in milligrams per liter were represented on the y-axis while the date range of January 12th, 1999 through March 6th, 2018 was represented on the x-axis. Once the points were plotted a linear regression line was fit to each dataset. Inorganic nitrogen levels rose by an overall rate of 0.0037 for the past 20 years at the ARK0040 location, while over the same period of time inorganic nitrogen levels only rose by an overall rate of 0.00003 at the downstream

location of ARK0006 (Appendix 4). Total phosphorus showed a decreasing trend at both location but at varying severity. At ARK0006, total phosphorus decreased at a rate of -0.001, while at the ARK0040 location it only decreased at a rate of -0.0001 (Appendix 5).

Definitions of Clusters

Cluster 1 was shown to be the best cluster for designating areas that maintained high levels of NDVI throughout the time series. This is shown by the majority of the polygons being located in the mostly forested northern and southern regions of the study area where it would be assumed to maintain NDVI levels better (Figure 15). By using that contextual information it can be assumed that cluster 1 polygons located in places that are as obvious for maintaining the average NDVI, are in fact doing so. Also based of the limited information available, it is shown that cluster 1 polygons were the biggest helpers for nutrient attenuation. This is shown because the majority of the northern section is covered by cluster 1 polygons, and water coming down the Illinois from ARK0040 was shown to on average have elevated level of nutrients but once it reached ARK0006 those levels were lower. This was especially true with inorganic nitrogen levels. So it can be deducted that the portion of land upstream from ARK0040, the middle section of the study area, contributed to the higher levels of nutrients over time. Cluster 2 was shown to have the lowest total average NDVI and it was also the smallest cluster, number wise (Figure 15). This was interpreted to mean that cluster 2 polygons were

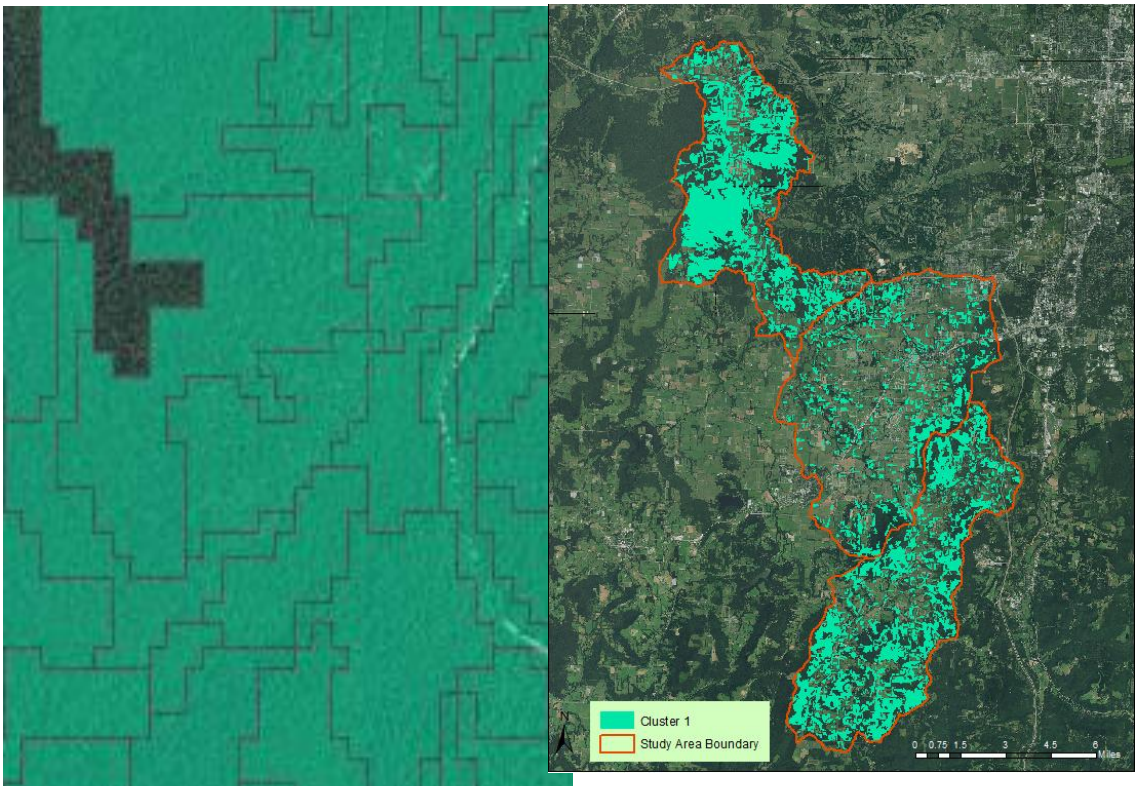


Figure 12: Cluster 1 Definition

Light blue polygons on the left represent areas classified in cluster 1 and show how well they correspond with forested areas. The cluster distribution on the right is displayed within the HUC 12: HWIRW study area.

effective at designating natural covered areas that remained naturally covered but may have had anthropogenic factors decreasing NDVI around them. This is shown because the majority of the cluster 2 polygons were located in the more urban middle section of the study area, but on the fringes of the actual municipal areas (Figure 15). Also with only 25% of cluster 2 polygons residing in the northern portion of the study area it can be concluded that they do not contribute highly to nutrient attenuation.

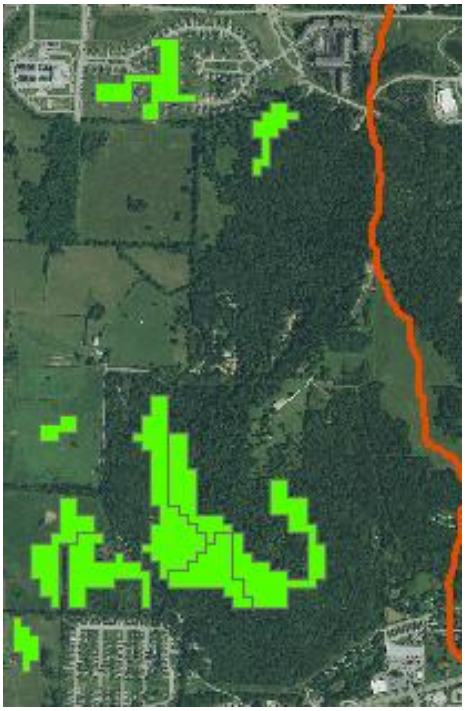


Figure 13: Cluster 2 Definition
The polygons in green represent those belonging to Cluster 2. It shows how they could possibly be affected by the anthropogenic nature of the surrounding land and how they are located just outside of urban environments.

Clusters 3 and 4 were very similar because they account for the majority of the polygons within the study area. Both clusters seemed to designate pasture or grasslands very well which makes up the major land cover type within the study area. Cluster 3 has a favorability to being present in areas with slightly worse NDVI values over time, whereas the larger cluster 4 maintained a higher level of NDVI throughout the time period. 17 percent of each cluster 3 and 4 are located in the northern region of the study area while over 50 percent of each is located in the middle portion of the study area. It is believed to be like this because farms grew outwards from the city centers throughout history therefore the lands closer to the city have been cleared and developed for livestock farming for a longer period of time. This goes to show that polygons

with either of these designations are also not helpful when it comes to helping with nutrient attenuation, because of how spread out they are throughout the study area.

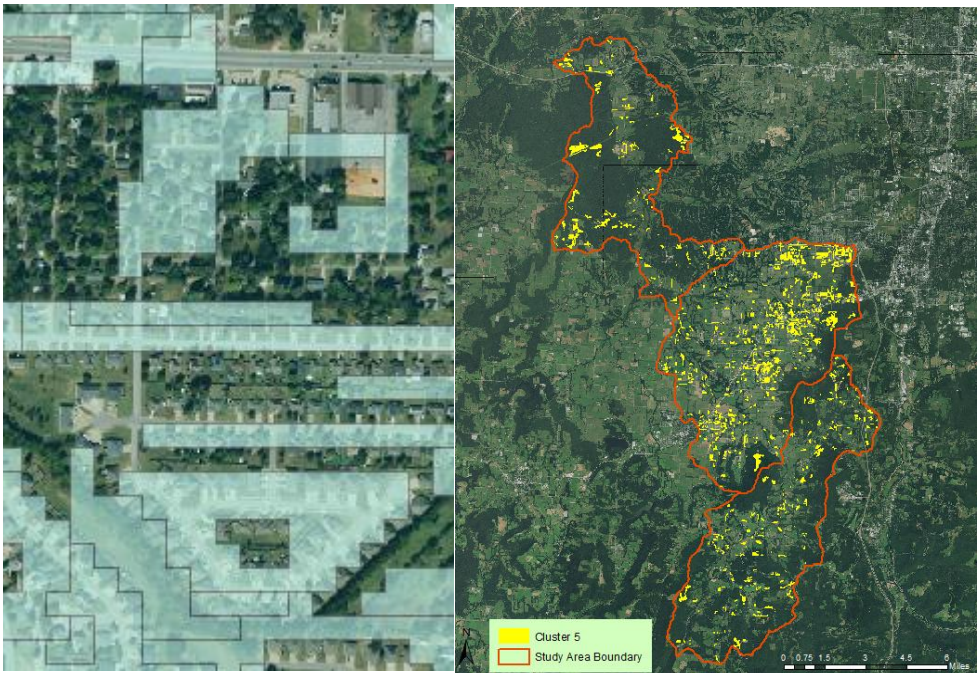


Figure 14: Cluster 5 Definition

In this figure the blue polygons, on the left, represent those belonging to cluster 5. This shows how well cluster 5 designated phenological cycle associated with manmade materials within urban areas. The image on the right shows the distribution of cluster 5 polygon within the HUC 12: HWIRW study area.

Cluster 5 performed extremely well at designating areas that would be considered urban or of manmade materials (Figure 16). It was the second smallest cluster and also had the second lowest total average NDVI. The interesting thing was that it had the highest rate of overall NDVI growth over time. It is believed this higher growth rate is caused by a rise in “green” city planning within the region (Northwest Arkansas Regional Planning Commission). In the past if a neighborhood was going in all the trees and natural flora would be wiped out and only replaced with the grass in the yard as far as NDVI goes. With the new planning requirements more of that NDVI level was maintained but the cluster still designated the manmade materials present

therefore making the cluster have a lower total NDVI average. It is known that manmade materials, impervious surfaces, and urban areas in general are detrimental to water quality, therefore based on the classifications reached in this study, polygons with a cluster 5 designation do not help with nutrient attenuation on average (Griffith et al., 2002; Brion et al., 2010; Ridd, 1995).

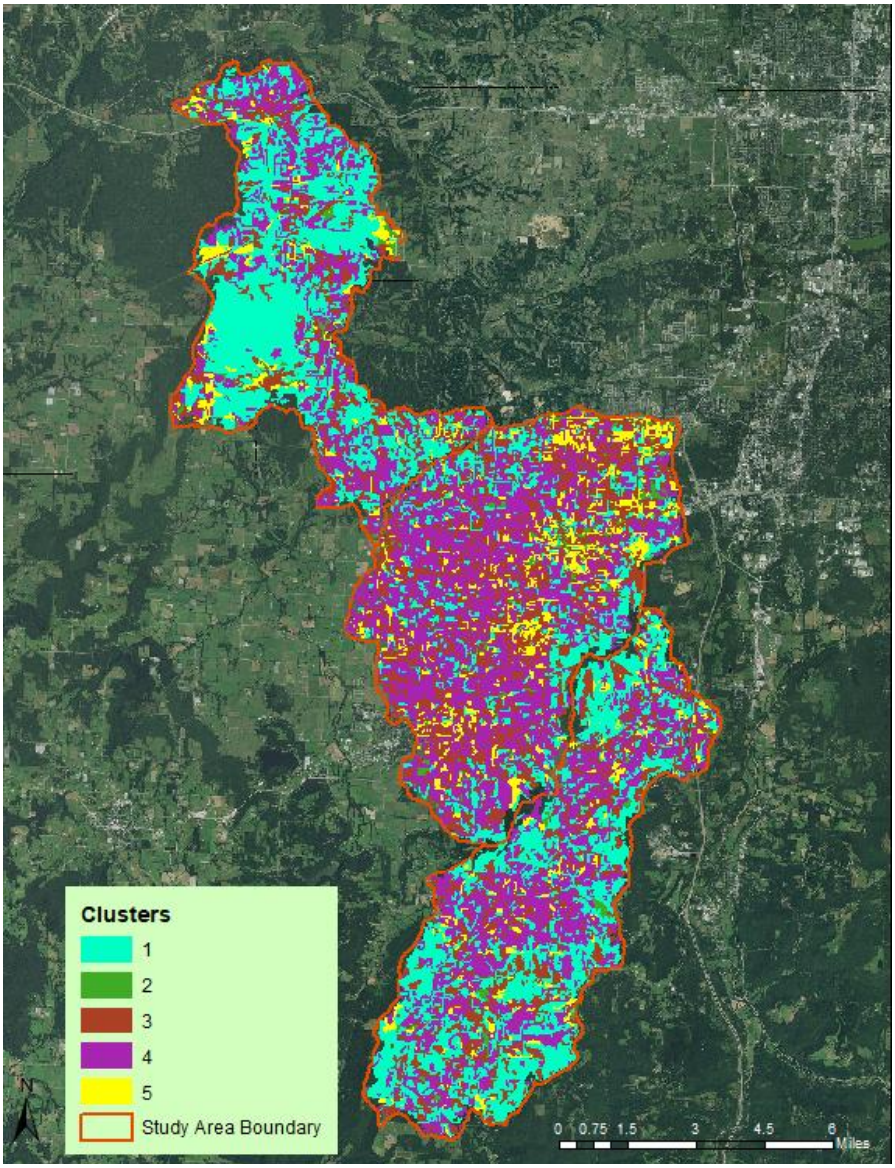


Figure 15: Cluster Distribution

Distribution of all clusters throughout the HUC 12: HWIRW study area.

Limitations and Future Study

This study encountered many limiting factors along the way. One of the biggest ones was not being able to find water quality information that was maintained regularly, for the temporal scale of 20 years. It would have been optimal to have one water quality station that was documented regularly, near the headwaters in the southern portion of the study area. This could have acted as somewhat of a control point to better enumerate the contributions of nutrients by the land within the study area. Complicated online database structures also presented a problem when harvesting water quality information. The only format produced were .txt files requiring further manipulation; these were copied and pasted into Microsoft Excel and subsequently formatted for analysis. Many other spectral indices and classification formats could have been tested to see how well they lined up with what was actually happening on the earth's surface in relation to water. Applying this methodology could also lead to labor intensive soil sampling and other in-situ data collection in order to attribute changes to specific properties. This methodology is intended to get you closer to the area of interest.

Building upon this study's general framework of image segmentation for data reduction, open source data harvesting and manipulation, and correlation to water quality values, there are some things that could be done that would greatly expand upon these relationships. In future studies there should be analyzing of different spectral indices other than just NDVI. Things like Normalized Difference Water Index (NDWI), Leaf Area Index (LAI), Enhanced Vegetation Index (EVI), and many other spectral remote sensing products should be compared to water quality values in order to possibly find relations between them. Water quality could also be assessed in terms of groundwater in future studies of regions in karst terrains. As of now there is

not a large stock of historical water quality information that can be queried for study but initiatives could be started in certain places to maintain that information in the name of science.

Conclusion

An efficient workflow was determined for monitoring and categorizing NDVI variations on a decadal scale. Using publicly available data and cloud-based processing, manipulation, and categorization, NDVI information from the past 20 years was effectively categorized by unsupervised classification techniques after innovative OBIA-based spatial data reduction. While meaningful LSP classes were successfully defined for the study area, anthropogenic effects to vulnerable water systems were hard to pinpoint, potentially due in part to temporal and spatial availability of water quality metrics. The slopes of each cluster's NDVI over time did not correlate positively with the water quality metrics as expected based on cluster definitions. Remote sensing has evolved to become utilized in a wide range of disciplines, which further advances the interconnectivity the natural resource management community and beyond.

Works Cited

Arkansas Natural Resources Commission. (2009). *Rules Governing the Arkansas Soil Nutrient and Poultry Litter Application and Management Program* (ANRC Publication No. 138 -Title 22). Little Rock, AR.

Arkansas GIS Office. (2017). *Data*. Retrieved from https://gis.arkansas.gov/data/

Baatz, M. & Schape, A. (2000). Multiresolution Segmentation: an Optimization Approach for High Quality Multi-scale Image Segmentation. *Angewandte Geographische Informationsverarbitung, 12*, 12-23.

Bisquert, M., Begue, A., & Deshayes, M. (2014). Object-based delineation of homogeneous landscape units at regional scale based on MODIS time series. *International Journal of Applied Earth Observation and Geoinformation, 37(2015),* 72-82.

Brabec, E., Schulte, S., & Richards, P.L. (2002). Impervious Surfaces and Water Quality: A Review of Current Literature and Its Implications for Watershed Planning. *Journal of Planning Literature, 16(4),* 499-514.

Brahana, Van, (2011), Ten relevant karst hydrogeologic insights gained from 15 years of *in situ* field studies at the Savoy Experimental Watershed; *in* Kuniansky, E.L., ed., U.S. Geological Survey Karst Interest Group Proceedings, Fayetteville, Arkansas, April 26-29, 2011, U.S. Geological Survey Scientific Investigations Report 2011-5031, p. 132-141.Brion, G., Brye, K., Haggard, B., West, C., Brahana, V. (2010). Land-use effects on water quality of a first-order stream in the Ozark-Highlands, mid-southern United States: River Research and Applications. DOI: 10.1002/rra.1394, 19 p.

Bunker, B.E., Tullis, J.A., Cothren, J.D., Casana, J., Aly, M.H. (2016) Object-based Dimensionality Reduction in Land Surface Phenology Classification. *AIMS Geosciences, 2(4),* 302-328.

Chunhao, Z. & Yingkui, L. (2014). Long-Term Hydrological Impacts of Land Use/Land Cover Change From 1984 to 2010 in the Little River Watershed, Tennessee. *International Soil and Water Conservation Research, 2(2),* 11-21.

Center for Advanced Spatial Technologies. (2006). *Arkansas Watershed Information System*. Retrieved from http://watersheds.cast.uark.edu/viewhuc.php?hucid=1111010301

Davis, J.V. & Schumacher, J.G. (1992). Water-Quality Characterization of the Spring River Basin, Southwestern Missouri and Southeastern Kansas. *USGS: Water-Resources Investigations Report, 90,* 4176.

Definiens. (2007). *eCognition Developer 7 Reference Book*. Munich, DE: Definiens AG, p.195

ESRI. *Bit depth capacity for raster dataset cells*. Retrieved from http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/bit-depth-capacity-for-raster-dataset-cells.htm

Fisher, J. R. B., Acosta, E. A., Dennedy-Frank, P. J., Kroeger, T., & Boucher, T. M. (2017). Impact of satellite imagery spatial resolution on land use classification accuracy and modeled water quality. *Remote Sensing in Ecology and Conservation*.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202(2017),* 18-27.

Griffith, J.A., Martinko, E.A., Whistler, J.L., Price, K.P. (2002). Interrelationships Among Landscapes, NDVI, and Stream Water Quality in the U.S. Central Plains. *Ecological Applications, 12(6),* 1702-1718.

Gu, Y., Brown, J.F., Tomoaki, M., van Leeuwen, W.J.D, Reed, B.C. (2010). Phenological Classification of the United States: A Geographic Framework for Extending Multi-Sensor Time-Series Data. *Remote Sensing, 2,* 526-544.

Haggard, B.E. & Sorens, T.S. (2006). Sediment phosphorus release at a small impoundment on the Illinois River, Arkansas and Oklahoma, USA. *Ecological Engineering, 28(2006),* 280-287.

Hartigan, J.A. & Wong, M.A. (1979). Algorithm AS 136: A K-means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(1),* 100-108.

Jensen, J.R. (2016). *Introductory Digital Image Processing: A Remote Sensing Perspective*. Glenview, IL: Pearson Education.

Jensen, J.R., Garcia-Quijano, M., Hadley, B., Im, J., Wang, Z., Nel, A.L., Teixeira, E., Davis, B.A. (2006). Remote Sensing Agricultural Crop Type for Sustainable Development in South Africa. *Geocarto International, 21(2),* 5-18.

Jiang, Y., Zhang, C., Yuan, D., Zhang, G., He, R. (2008). Impact of land use change on groundwater quality in a typical karst watershed of southwest China: a case study of the Xiaojiang watershed, Yunnan Province. *Hydrogeology Journal, 16,* 727-735.

Johnson, L., Richards, C., Host, G., Arthur, J. (1997). Landscape Influences on Water Chemistry in Midwestern Stream Ecosystems. *Freshwater Biology, 37,* 113-132.

Kalkhoff, S.J. (1993) Using A Geographic Information System to Determine the Relation Between Stream Quality and Geology in the Roberts Creek Watershed, Clayton County, Iowa. *AWRA: Water Resources Bulletin, 29(6),* 986-996.

Kamal, M. & Phinn, S. (2011) Hyperspectral Data for Mangrove Species Mapping: A Comparison of Pixel-Based and Object-Based Approach. *Remote Sensing, 3(10),* 2222-2242.

Karp Resources. (2014). *Northwest Arkansas Regional Food Assessment*. New York, NY: Karp, K. & Sandusky, E.

Kelly, J.R. & Harwell, M.A. (1990). Indicators of Ecosystem Recovery. *Environmental Management, 14(5),* 527-545.

Kresse, T.M., Hays, P.D., Merriman, K.R., Gillip, J.A., Fugitt, D.T., Spellman, J.L., Nottmeier, A.M., Westerman, D.A., Blackstock, J.M., Battreal, J.L. (2014). Aquifers of Arkansas -- Protection, Management, and Hydrologic and geochemical Characteristics of Groundwater Resources in Arkansas. *Scientific Investigations Report, 5149*.

LeBlanc, R.T., Brown, R.D., FitzGibbon, J.E. (1997). Modeling the effects of land use changes on the water temperature in unregulated urban streams. *Journal of Environmental Management, 49,* 446-469.

Leidy, V.A. & Morris, E.E. (1990). Hydrogeology and quality of ground water in the Boone formation and Cotter dolomite in karst terrain of northwestern Boone County, Arkansas. *USGS: Water-Resources Investigations Report, 90,* 4066.

Monroe, W.H. (1970). A glossary of karst terminology. *USGS: Water Supply Paper, 1899,* 26.

Morisette, J.T., Richardson, A.D., Knapp, A.K., Fisher, J.I., Graham, E.A., Abatzoglou, J., Wilson, B.E., Breshears, D.D., Henebry, G.M., Hanes, J.M., Liang, L. (2009). Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. *Frontiers in Ecology and the Environment, 7,* 253-260.

Northwest Arkansas Regional Planning Commision. *NWA Open Space Plan*. Retrieved from http://nwarpc.org/environment/nwa-open-space-plan/

Öztürk, M., Copty, N. K., & Saysel, A. K. (2013). Modeling the impact of land use change on the hydrology of a rural watershed. *Journal of Hydrology, 497*, 97-109.

Peterson, J.C., Adamski, J.C., Bell, R.W., Davis, J.V., Femmer, S.R., Freiwald, D.A., Joeseph, R.L. (1998). Water quality in the Ozark Plateaus, Arkansas, Kansas, Missouri, and Oklahoma, 1992-1995. *U.S. Geological Survey Circular 1158*. Retrieved from http://pubs.usgs.gov/circ/circ1158

Rafferty, M.D. (1980) *The Ozarks, Land and Life*. Norman, OK: University of Oklahoma.

Ridd, M.K. (1995). Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystems analysis through remote sensing: comparative anatomy for cities. *International Journal of Remote Sensing, 16(12),* 2165-2185.

Rosa, J.D. (2017, March 23). *Census estimate shows Northwest Arkansas continues to lead state in pace of population growth*. Retrieved from https://talkbusiness.net/2017/03/census-estimate-shows-northwest-arkansas-continues-to-lead-state-in-pace-of-population-growth/

RStudio. *Our Mission*. Retrieved from https://www.rstudio.com/about/

S Anand. (2017, February 9). Finding Optimal Number of Clusters. Retrieved from https://www.r-bloggers.com/finding-optimal-number-of-clusters/

Strasser, T. & Lang, S. (2014) Object-based class modeling for multi-scale riparian forest habitat mapping. *International Journal of Applied Earth Observation and Geoinformation, 37(2015)*, 29-37.

SERVIR-Mekong. (2017). Training Resources. *Exercise 3: Writing Custom Functions and Mapping across Image Collections*. Retrieved from https://developers.google.com/earth-engine/ttt

Tong, S.T.Y., & Chen, W.L. (2002). Modeling the relationship between land use and surface water quality. *Journal of Environmental Management, 66*, 377-393.

United States Geologic Survey. (2017). *USGS Landsat 5 Surface Reflectance Tier 1*. Retrieved from https://code.earthengine.google.com/dataset/LANDSAT/LT05/C01/T1_SR

University of Arkansas Division of Agriculture. (2017). Arkansas Agriculture Profile. *Pocket Facts 2017*, 5. Retrieved from https://division.uaex.edu/docs/2017%20AR%20Ag%20profile.pdf

United States Forestry Service. (2018). Ozark-St. Francis National Forests. *Forest Information*. Retrieved from https://www.fs.usda.gov/detail/osfnf/about-forest/?cid=fsm8_042910

White, W.B. (2002). Karst Hydrology: recent developments and open questions. *Engineering Geology, 65(2002),* 85-105.

Appendix 1

Filtered_Exported_ndvi_timeseries *

Get Link | Save | Run | Reset

```
Imports (4 entries)
  var ls5: ImageCollection "USGS Landsat 5 Surface Reflectance Tier 1"
  var ls8: ImageCollection "USGS Landsat 8 Surface Reflectance Tier 1"
  var watersheds: Table "HUC08: USGS Watershed Boundary Dataset of Subbasins"
  var watersheds1: Table "HUC10: USGS Watershed Boundary Dataset of Watersheds"
var studyArea = watersheds1.filterMetadata('huc10', 'equals', '1111010304');

var buf300 = function(feature) {
  return feature.buffer(300);
};

var studyAreaBuf = studyArea.map(buf300);


var timeseries8 = ls8.filterDate('2015-1-2', '2018-3-30');
var timeseries8_SA = timeseries8.filterBounds(studyAreaBuf);

//var stNDVI5 = function(image) {return image.normalizedDifference(['B4', 'B3']);};
var stNDVI8 = function(image) {return image.normalizedDifference(['B5', 'B4']);};

//var final5 = timeseries5_SA.map(stNDVI5);
var final8 = timeseries8_SA.map(stNDVI8);

//This code uses the .map function to apply a function (stNDVI variable) to each
//image in the collection.  The function I created will return, for each date within the temporal range, a
//calculated NDVI.

/*
 * Author: Rodrigo E. Principe
 * License: Apache 2.0

PURPOSE:
This function Exports all images from one Collection
PARAMETERS:
col = collection that contains the images (ImageCollection) (not optional)
folder = the folder where images will go (str) (not optional)
scale = the pixel's scale (int) (optional) (defaults to 1000) (for Landsat use 30)
type = data type of the exported image (str) (option: "float", "byte", "int", "double") (optional) (defaults to "float")
nimg = number of images of the collection (can be greater than the actual number) (int) (optional) (defaults to 500)
maxPixels = max number of pixels to include in the image (int) (optional) (defults to 1e10)
region = the region where images are on (Geometry.LinearRing or Geometry.Polygon) (optional) (defaults to the image footprint)
```

Appendix 1 (cont'd)

```
Be careful with the region parameter. If the collection has images
in different regions I suggest not to set that parameter
EXAMPLE:
ExportCol(myLandsatCol, "Landsat_imgs", 30)
*/

var ExportCol = function(col, folder, scale, type,
                         nimg, maxPixels, region) {
    type = type || "float";
    nimg = nimg || 500;
    scale = scale || 1000;
    maxPixels = maxPixels || 1e10;

    var colList = col.toList(nimg);
    var n = colList.size().getInfo();

    for (var i = 0; i < n; i++) {
      var img = ee.Image(colList.get(i));
      var id = img.id().getInfo();
      region = region || img.geometry().bounds().getInfo()["coordinates"];

      var imgtype = {"float":img.toFloat(),
                     "byte":img.toByte(),
                     "int":img.toInt(),
                     "double":img.toDouble()
                    }

      Export.image.toDrive({
        image:imgtype[type],
        description: id,
        folder: folder,
        fileNamePrefix: id,
        region: region,
        scale: scale,
        maxPixels: maxPixels})
    }
  };

// ExportCol(final5, "FinalNDVI", 30, "float", 500, 1e10, studyAreaBuf);
 ExportCol(final8, "FinalNDVI", 30, "float", 500, 1e10, studyAreaBuf);
```

Appendix 2

Cluster optimization

```
jdata<-Export_OutputNDVI

jdata<-jdata[,-98]
jdata<-jdata[,-2]
jdata<-jdata[,-1]
scale_jdata<-scale(jdata)


set.seed(13)

wss <- (nrow(scale_jdata)-1)*sum(apply(scale_jdata,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(scale_jdata,
                                     centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
     ylab="within groups sum of squares")
```

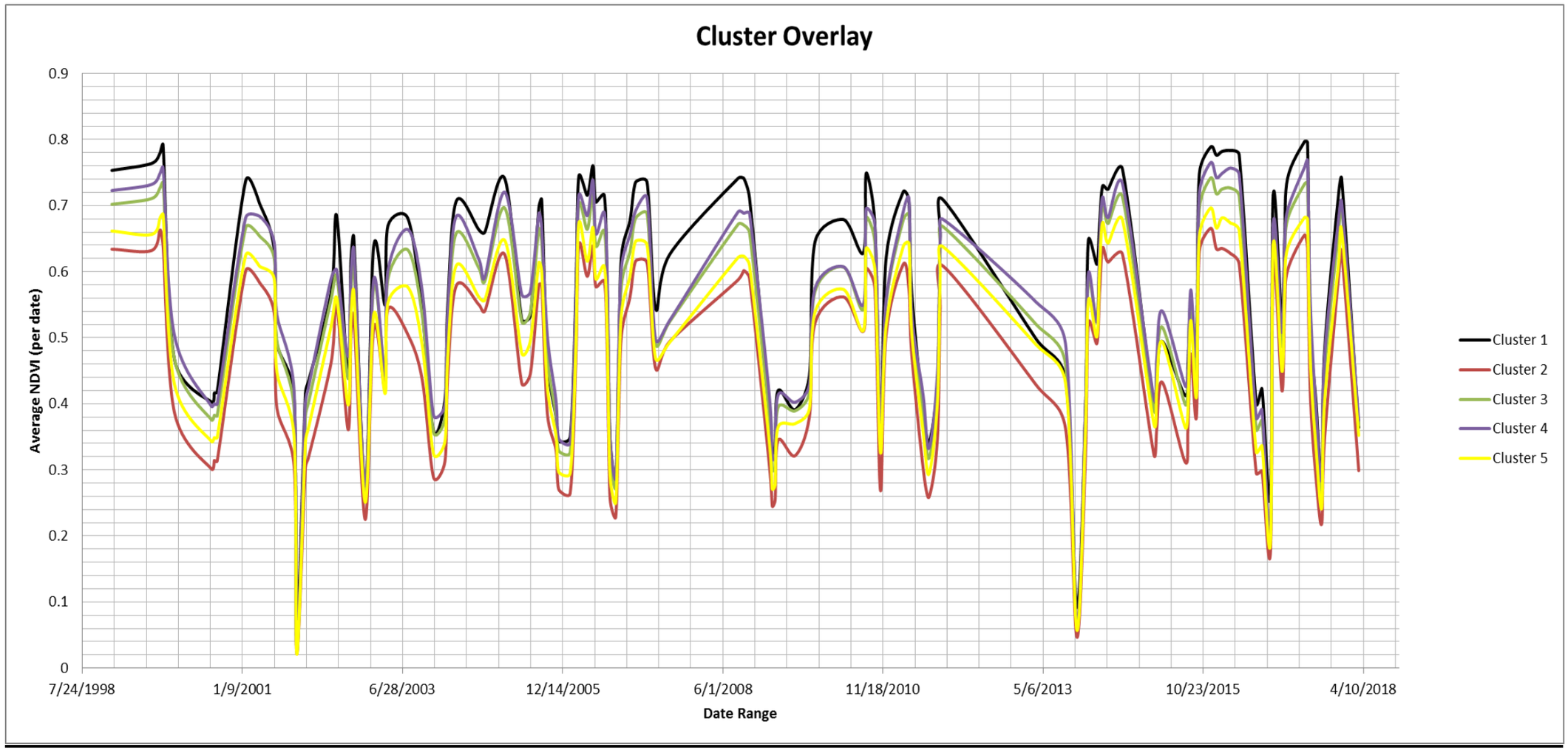
K-means Clustering

```
jdata<-Export_OutputNDVI

jdata<-jdata[,-98]
jdata<-jdata[,-2]
jdata<-jdata[,-1]

set.seed(13)
OptNDVIclust2 <- kmeans(jdata, 5, nstart = 25, iter.max = 500)

```

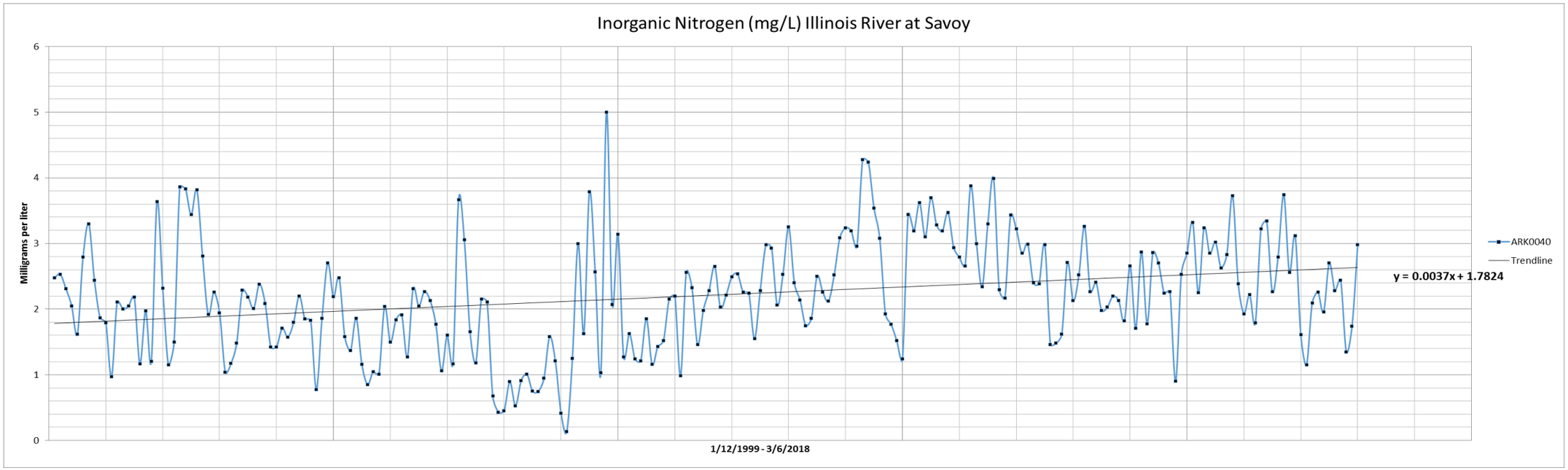
Data Transform

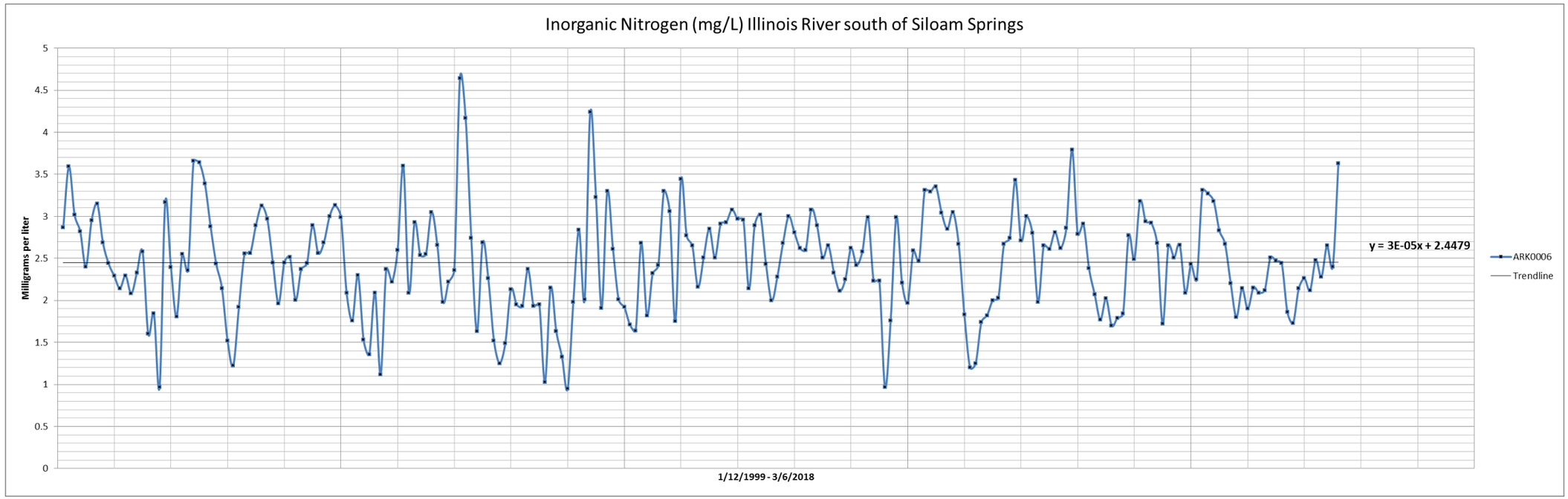
```
trueNDVIr_clust5<-cluster5

for (i in 1:1230){
  print(i)
  for(j in 1:121){
    trueNDVIr_clust5[i,j]<-trueNDVIr_clust5[i,j]/5000-1

  }
}
```

Appendix 3

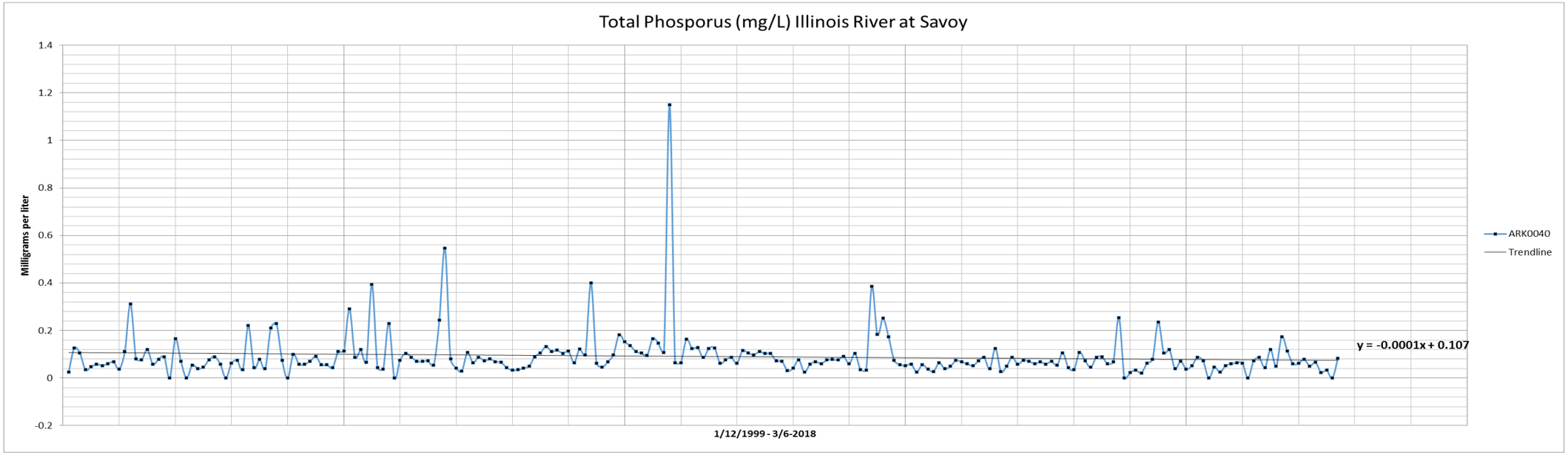

Cluster Overlay
Average NDVI (per date)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
7/24/1998
1/9/2001
6/28/2003
12/14/2005
6/1/2008
11/18/2010
5/6/2013
10/23/2015
4/10/2018
Date Range
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5

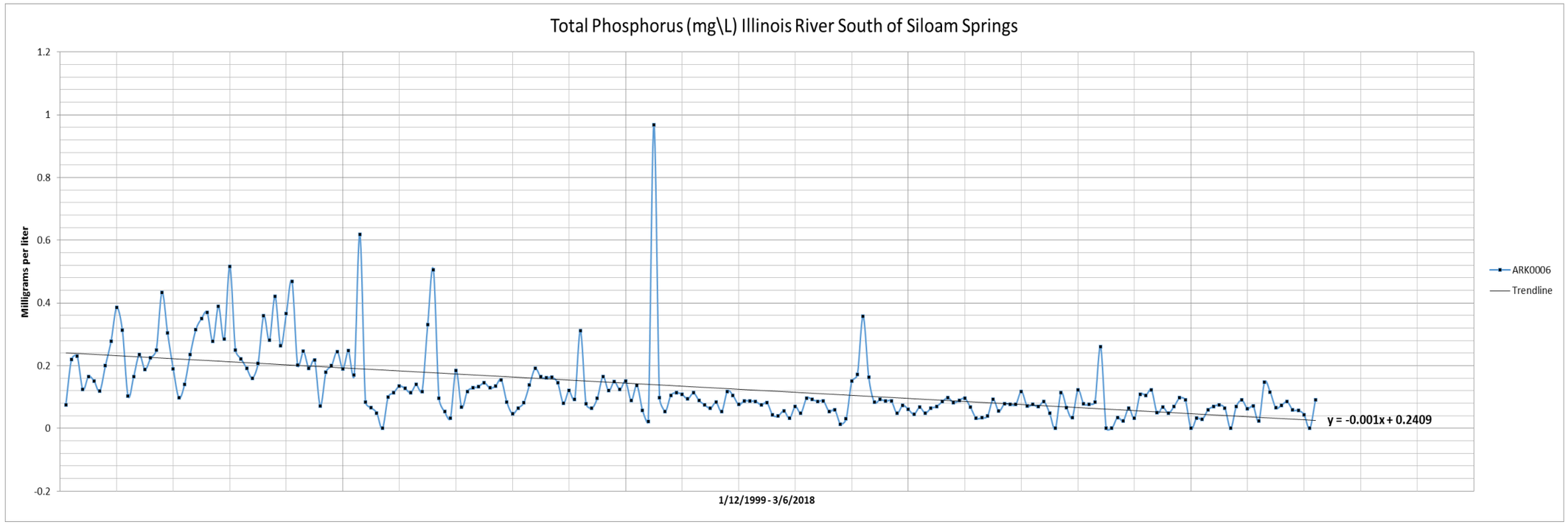
Appendix 4


Inorganic Nitrogen (mg/L) Illinois River at Savoy
Milligrams per liter
0
1
2
3
4
5
6
1/12/1999 - 3/6/2018
ARK0040
Trendline
y = 0.0037x + 1.7824


Inorganic Nitrogen (mg/L) Illinois River south of Siloam Springs
Milligrams per liter
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
1/12/1999 - 3/6/2018
y = 3E-05x + 2.4479
ARK0006
Trendline

Appendix 5


Total Phosporus (mg/L) Illinois River at Savoy
Milligrams per liter
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
1/12/1999 - 3/6-2018
ARK0040
Trendline
y = -0.0001x + 0.107


Total Phosphorus (mg\L) Illinois River South of Siloam Springs
Milligrams per liter
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
1/12/1999 - 3/6/2018
ARK0006
Trendline
y = -0.001x + 0.2409