SURFACE WATER HYDROLOGIC MODELING USING REMOTE SENSING DATA FOR NATURAL AND DISTURBED LANDS

by

MULUKEN EYAYU MUCHE

B.S., Jimma University, 2002
M.S., University of Bremen, 2007
M.Eng., Boise State University, 2011

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biological and Agricultural Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Abstract

The Soil Conservation Service-Curve Number (SCS-CN) method is widely used to estimate direct runoff from rainfall events; however, the method does not account for the dynamic rainfall-runoff relationship. This study used back-calculated curve numbers (CNs) and Normalized Difference Vegetation Index (NDVI) to develop NDVI-based CNs (CN_{NDV}) using four small northeastern Kansas grassland watersheds with average areas of 1 km^2 and twelve years (2001–2012) of daily precipitation and runoff data. Analysis indicated that the CN_{NDVI} model improved runoff predictions compared to the SCS-CN method. The CN_{NDVI} also showed greater variability in CNs, especially during growing season, thereby increasing the model's ability to estimate relatively accurate runoff from rainfall events since most rainfall occurs during the growing season. The CN_{NDVI} model was applied to small, disturbed grassland watersheds to assess the model's ability to detect land cover change impact for military maneuver damage and large, diverse land use/cover watersheds to assess the impact of scaling up the model. CN_{NDVI} application was assessed using a paired watershed study at Fort Riley, Kansas. Paired watersheds were identified through k-means and hierarchical-agglomerative clustering techniques. At the large watershed scale, Daymet precipitation was used to estimate runoff, which was compared to direct runoff extracted from stream flow at gauging points for Chapman (grassland dominated) and Upper Delaware (agriculture dominated) watersheds. In large, diverse watersheds, CN_{NDVI} performed better in moderate and overall flow years. Overall, CN_{NDVI} more accurately simulated runoff compared to SCS-CN results: The calibrated model increased by 0.91 for every unit increase in observed flow (r = 0.83), while standard CN-based flow increased by 0.506 for every unit increase in observed flow (r = 0.404). Therefore, CN_{NDVI} could help identify land use/cover changes and disturbances and spatiotemporal changes in

runoff at various scales. CN_{NDVI} could also be used to accurately estimate runoff from precipitation events in order to instigate more timely land management decisions.

SURFACE WATER HYDROLOGIC MODELING USING REMOTE SENSING DATA FOR NATURAL AND DISTURBED LANDS

by

MULUKEN EYAYU MUCHE

B.S., Jimma University, 2002
M.S., University of Bremen, 2007
M.Eng., Boise State University, 2011

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Biological and Agricultural Engineering
College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Stacy L. Hutchinson

Copyright

.

Abstract

The Soil Conservation Service-Curve Number (SCS-CN) method is widely used to estimate direct runoff from rainfall events; however, the method does not account for the dynamic rainfall-runoff relationship. This study used back-calculated curve numbers (CNs) and Normalized Difference Vegetation Index (NDVI) to develop NDVI-based CNs (CN_{NDV}) using four small northeastern Kansas grassland watersheds with average areas of 1 km^2 and twelve years (2001–2012) of daily precipitation and runoff data. Analysis indicated that the CN_{NDVI} model improved runoff predictions compared to the SCS-CN method. The CN_{NDVI} also showed greater variability in CNs, especially during growing season, thereby increasing the model's ability to estimate relatively accurate runoff from rainfall events since most rainfall occurs during the growing season. The CN_{NDVI} model was applied to small, disturbed grassland watersheds to assess the model's ability to detect land cover change impact for military maneuver damage and large, diverse land use/cover watersheds to assess the impact of scaling up the model. CN_{NDVI} application was assessed using a paired watershed study at Fort Riley, Kansas. Paired watersheds were identified through k-means and hierarchical-agglomerative clustering techniques. At the large watershed scale, Daymet precipitation was used to estimate runoff, which was compared to direct runoff extracted from stream flow at gauging points for Chapman (grassland dominated) and Upper Delaware (agriculture dominated) watersheds. In large, diverse watersheds, CN_{NDVI} performed better in moderate and overall flow years. Overall, CN_{NDVI} more accurately simulated runoff compared to SCS-CN results: The calibrated model increased by 0.91 for every unit increase in observed flow (r = 0.83), while standard CN-based flow increased by 0.506 for every unit increase in observed flow (r = 0.404). Therefore, CN_{NDVI} could help identify land use/cover changes and disturbances and spatiotemporal changes in

runoff at various scales. CN_{NDVI} could also be used to accurately estimate runoff from precipitation events in order to instigate more timely land management decisions.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to express deepest appreciation to my major advisor Dr. Stacy Hutchinson for her continuous guidance, support, encouragement, and patience throughout my study. I would like to express my sincere gratitude to Dr. Shawn Hutchinson, my supervisory committee member for providing NDVI data, his guidance and valuable comments. Many thanks to my supervisory committee members: Dr. Phil Barnes and Dr. Aleksey Sheshukov for their valuable comments and continuous encouragement. My gratitude also goes to my outside Chairperson, Dr. Nathan Nelson, for his time, comments and suggestions.

I would also like to express my gratitude to the following people who helped me in various ways during my study period: Brandon Lantz, Seely McCarty, Yan Donglin, Ray Clotfelter, Tamara Robinson, Dr. Kyle Douglas-Mankin, Dr. Tim Keane, Dr. Trisha Moore, Dr. Katherine Burke, Vladmir Karimov, Barb Moore, Temesgen Jemaneh, Ginger Pugh, Ameha Gebreiyesus, Tesfaye Tadesse and other friends.

Last but not least, I would like to express my special thanks to my wife (Dr. Sumathy Sinnathamby), my daughter (Sophia) and the rest of my family for their endless support throughout the way. Without their support this work would have not been possible.

Dedication

To my dearest daughter Sophia!

You provided the strength and inspiration necessary for me to complete this project.

Abbreviation

AGNPS	Agricultural Non-Point Source Pollution Model
AVHRR	Advanced Very High Resolution Radiometer
cfs	Cubic feet per second
CN	Curve Number
CN_{NDVI}	Curve Number based on Normalized Difference Vegetation Index
DEM	Digital Elevation Model
EPIC	Erosion Productivity Impact Calculator
ET	Evapotranspiration
ETM+	Enhanced Thematic Mapper Plus
GIS	Geographical Information System
HEC-HMS	Hydrologic Engineering Center-Hydrologic Modeling System
HUC	Hydrologic Unit Code
LTER	Long-term Ecological Research
MODIS	Moderate Resolution Imaging Spectroradiometer
NAD83	North American Datum of 1983
NASA LPDAAC	National Aeronautics and Space Administration Land Processes Distributed Active Archive Center
NCDC	National Climatic Data Center
NDVI	Normalized Difference Vegetation Index
NOAA	The National Oceanic and Atmospheric Administration
NLCD	National Land Cover Dataset
NRCS	Natural Resources Conservation Service (USDA)
PBIAS	Percent bias (%)
QA	Quality Assurance
R	Pearson Correlation Coefficient
r^2	Coefficient of determination
SCS	Soil Conservation Service

SDS	Science Data Sets
SSURGO	Soil Survey Geographic Database
SNOTEL	Snow Telemetry
SWAT	Soil and Water Assessment Tool
TM	Thematic Mapper
USDA	United States Department of Agriculture
USGS	United States Geological Survey
UTM	Universal Transverse Mercator coordinate system
VI	Vegetation Indices

Chapter 1 - Introduction

1.1 General Background

Water is one of the most essential natural resources on the earth, but its availability and quality is adversely affected by increasing agricultural, domestic, and industrial demands. Water demands vary depending on the climate of an area and traditional uses of water (Davie, 2008). As the world's population grows rapidly, water demands are also increasing. Intensified urbanization and the degradation of ecosystems exacerbate adverse effects on water quality and quantity (Lal and Shukla, 2004). Water experiences cyclical continuous movement on, above, and below the surface of the earth; the sun is the primary cause for this movement (Perlman, 2014). Figure 1-1 shows the hydrologic cycle, which involves an exchange of energy that drives precipitation, evapotranspiration, infiltration, and runoff generation in the atmosphere, hydrosphere, biosphere, and lithosphere. Quantifying components of the hydrologic cycle at various scales provides invaluable assistance for water resource planning and management and the decision-making process on water usage and quality.

Hydrologists obtain measurable information and knowledge about the water cycle and its components in order to increase understanding of hydrologic physical processes and make appropriate and timely water quantity and quality-related decisions. Hydrologic processes and estimations of hydrologic components, mainly precipitation and runoff, comprise the foundation of effective water resource and environmental planning and management. Fedak (1999) stated that models enable hydrologists to understand complex problems by simulating and predicting hydrologic behavior. However, adaptations of models' results need precautions since models depend on assumptions, input, and parameter estimates. Quantification of hydrologic components also requires understanding of the factors that determine those hydrologic processes,

their inter-relationships, and the methods used to determine the relationship. For example, the amount of runoff produced from precipitation events is dependent upon hydrologic characteristics such as area and shape of watershed, vegetation, topography, soil, and rainfall volume and intensity. Accurate runoff predictions require thorough understanding of these factors and their relationship.

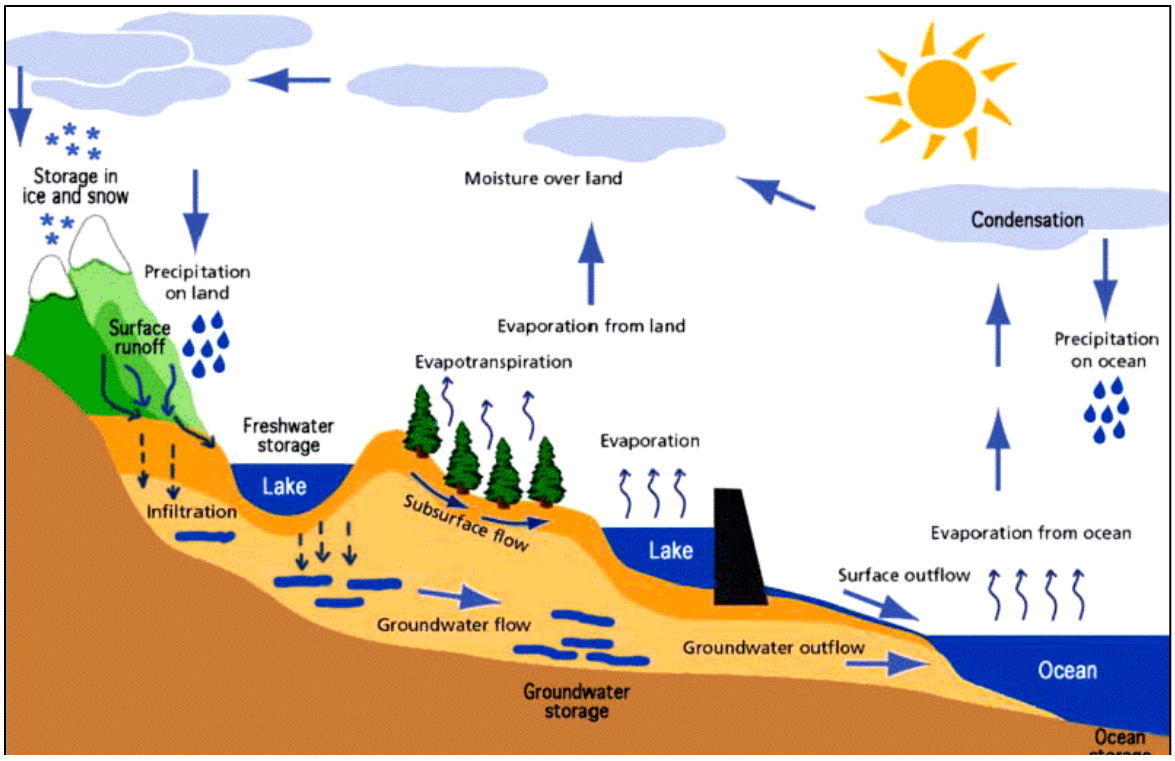


Figure 1-1: Hydrologic cycle with components that describe water movement derived by sun and wind energy (Adopted from Enock, 2011)

Hydrologic models have been used to quantify and predict the amount of water, nutrients, sediment, and pollutant loads in the system. However, the accuracy of hydrologic models typically depends on how well the input data represent precipitation, soil, land use/cover, land management, topography, and other factors that affect physical processes. In addition, determining the empirical and numerical relationships of various components of hydrologic processes is a critical but daunting issue in modeling. These empirical and numerical

relationships, also called parameterizations, relate hydrologic variables to each other and describe specific physical mechanisms in hydrologic processes.

Although hydrologic models are simplified representations of a complex system, they could quantify hydrologic processes more appropriately with proper model setup, input, and parameterization. Hydrological models have potential to effectively estimate annual and seasonal water yields, predict flood peak for improved structural design to prevent floods, conserve water for better management during drought periods, estimate sedimentation and chemical transport, and estimate crop yields from agricultural areas with different weather and climatic conditions. In order to accomplish these purposes, however, components of hydrological processes must be quantified at different scales which fits the objective of specific study.

When any process is defined or parameterized, that process must be considered at a given scale. A watershed model holistically considers hydrological processes at the watershed scale when compared to other models that simulate at small or field scale (Daniel et al., 2011; Oogathoo, 2006). A watershed is defined as an entire upstream area that contributes to the stream system that flows through a given outlet point. Watershed size depends on the outlet point in the drainage system under consideration, which is typically the entrance point to the large water body, reservoir, or any point where it changes from a tributary to a river. In most cases, river flow is measured at that outlet point. Watersheds, which are naturally delineated by land surface topography, are natural systems boundaries for mass and energy flux. A watershed is also the basic organizing unit for a wide range of scientific, engineering, and water management activities that quantify water and sediment fluxes into and out of a specified system (Genereux, 2005). Watersheds integrate hydrologic processes within their boundaries, and watershed data is often extrapolated to a larger extent with appropriate precautions (Dahlgren, 1999). However,

when hydrological processes are defined or parameterized at the watershed scale, its spatial variability must be included because most hydrologic processes are highly variable in space (Daniel et al., 2011). Therefore, parameterization of a watershed model requires additional expertise and precaution, especially in semidistributed and distributed models because these models account for spatial variability of hydrologic processes, input, boundary conditions, and watershed characteristics (Refsgaard, 1996). The study of hydrologic processes must consider spatial and temporal variability of topography, soil, and vegetation in the study area in order to improve the representation of runoff generating processes.

Variability in land use/cover, soil physicochemical properties, and climatic variables such as precipitation, humidity, temperature, solar radiation, and wind speed and direction cause spatiotemporal heterogeneity in hydrologic processes. The changing climate, occurrences of variable extreme events, and landscape disturbances contribute to spatiotemporal variability in hydrologic processes at various scales. As the watershed area decreases in size, precipitation variability could increasingly impact model outputs, especially peak flows (Mandapaka et al., 2009). Changes in amount, frequency, and intensity of precipitation directly affect the magnitude and timing of runoff as well as flood and drought intensities (Arnell et. al, 1996; Immerzeel, 2008). In addition to precipitation, landscape and topography are also very significant variables that require pertinent representation in models. For example, change in land use affects hydrologic processes because it alters evapotranspiration (ET), the ability of soil to hold water, and the ability of vegetation to intercept precipitation. Changes in land use also modify runoff pathway and timing. Spatiotemporal variability of components in the hydrologic cycle incites an increasing demand for simple, accurate, reliable methods to measure or estimate hydrologic variables and define suitable relationships of rainfall and runoff.

Current and future advancements in remote sensing and geographic information science (GIS) provide opportunity to quantify and predict components of hydrologic processes at various scales by providing a means of observing and assessing hydrologic variables over large areas (Schmugge, 1987; Schmugge et al., 2002), thereby improving input data and parameterization of hydrologic models. Remote sensing allows spatial variability of parameters and processes in hydrology to be defined at fine spatial and temporal scales that are typically present in the natural environment. Remote sensing techniques also help quantify temporal variability if observations are made repeatedly. Remote sensing is applied in various hydrologic studies, including estimation of soil moisture, runoff, ET, snow pack and melt, precipitation, land surface temperature, vegetation covers, and vegetation phenology (Hong and Adler, 2008; Jackson et al., 1996; Pockrandt, 2014; Schmugge et al., 2002; Schnur et al., 2010). The remote sensing and GIS technology allows researchers to understand integral variables in the land surface water balance at various spatial and temporal resolutions. GIS allows those variables to be analyzed and mapped at different spatial scales, consequently facilitating water management and decision-making processes.

Although computational capability of computing equipment has tremendously increased over the time in recent decades, quantifying hydrologic processes and accounting for watersheds' variability especially in spatial cases such as changes in climate, occurrences of variable extreme events, and landscape disturbances on hydrologic processes cannot be done with available traditional data. This is difficult mainly due to unavailability of long-term hydrologic records, higher natural variability of most hydrologic systems, difficulties in controlling catchment land-use changes, limited controlled small-scale experimental studies and the challenges involved in extrapolating model results to other systems (DeFries and Eshleman, 2004). However, these

problems can be resolved by using advancements in remote sensing and GIS (DeFries and Eshleman, 2004; Lambin et al., 2001) to implement hydrologic models, especially at the watershed scale. This study devised methods to improve the development and analysis of surface hydrologic modeling using remote sensing, GIS, and statistical techniques.

1.2 Research Agenda

1.2.1 Statement of Problem

Land surface and its transformation plays a fundamental role in modulating the atmospheric, geomorphic, hydrologic, and ecological processes on or near the earth's surface; therefore, understanding land surface could provide valuable information regarding natural process functions (Wilson, 2012). Land use/cover change and/or disturbances affect the water cycle because it changes the ET, soil's ability to hold water, and vegetation's ability to intercept precipitation. Changes in land use/cover also modify the pathways of water and the surface roughness that influences the timing of runoff and river flow.

Surface runoff is a vital hydrologic component that is widely studied in water resources. Understanding and predicting the rainfall-runoff relationship is crucial for water resource planning and management and understanding processes of erosion, sediment transport, and contaminant loading. However, the rainfall-runoff process is complex, dynamic, and nonlinear (Fan et al., 2013; Song et al., 2011). The amount of runoff generated from a rainfall event is dependent on hydrologic conditions of an area as well as the amount, intensity, and duration of the rainfall. Hydrologic conditions of a watershed are primarily affected by land use/cover, soil physicochemical properties, antecedent moisture condition, and local and regional temperature. Land surface topography, land use/cover, soil physical properties, and compaction are the main determinants in hydrologic processes, especially for surface and groundwater processes. Direct

runoff from rainfall events is a fundamental hydrologic concept for flood peak estimation and structure design (Hawkins et al., 2008). However, reliable predictions of amount and rate of runoff from land surface is becoming increasingly difficult and time-consuming as watershed's complexity increases.

The volume of direct runoff in ungauged rural catchments widely estimated using the curve number method (Boughton, W., 1989). The Curve Number (CN) method, also referred as the SCS-CN method, was developed in the 1950s by the U.S. Department of Agriculture (USDA), formerly the Soil Conservation Service (SCS) and currently the Natural Resources Conservation Service (NRCS), in order to implement public law 566 (Hawkins et al., 2008; Hawkins, 2014). Public law 566 is a Watershed Protection and Flood Prevention Act, enacted on August 4, 1954, that requires upstream flood prevention and watershed condition improvement, primarily within the small watershed scale (Woodward et al., 2002). The SCS-CN method was developed as a uniform procedure the USDA-SCS could apply nationwide based on available data (Woodward et al., 2002). The SCS-CN method has been widely applied in many hydrologic models, such as Soil and Water Assessment Tool (SWAT), Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS), Erosion Productivity Impact Calculator (EPIC), and Agricultural Non-Point Source Pollution Model (AGNPS) (Kousari et al., 2010).

The SCS-CN method is based on land use/cover, hydrologic soil group (HSG), and hydrologic condition of the watershed. It expresses runoff depth as a function of rainfall depth, hydrologic storage, and initial abstraction. Since a watershed is usually a combination of different land uses and soil conditions, an area weighted average CN for the entire watershed or sub-watershed is often used based on SCS-CN lookup tables and seasonal or dynamic land use changes is not well represented. Dynamic land use/cover changes and hydrologic conditions are

not precisely accounted for in most hydrologic studies due to limited spatial and temporal data availability and their time-consuming nature. However, because spatiotemporal changes in a landscape significantly influence runoff estimation, the changes should be accounted for in order to improve water resources management. Although the SCS-CN method is widely applied and incorporated in numerous hydrologic models (Kousari et al., 2010), it does not adequately reflect detailed spatiotemporal variability of the rainfall-runoff relationship because of limited spatial and temporal data availability. The first part of this study more efficiently accounted for spatiotemporal variability of the rainfall-runoff relationship than the standard CN method (SCS-CN) because CN based on the Normalized Difference Vegetation Index (NDVI) was utilized. NDVI is remotely sensed product of Moderate Resolution Imaging Spectroradiometer (MODIS).

In hydrologic studies the understanding of hydrologic impacts of watershed management decisions or disturbances are derived from controlled, experimental manipulations of land cover and pre- and post-manipulation observations of hydrologic processes at the watershed scale so that watershed is used as an observable scale of catchment experiments with precipitation inputs and stream discharge outputs (DeFries and Eshleman, 2004; Tollan, 2002). The paired watershed approach, a commonly used hydrologic approach, captures the effects of climate and hydrologic processes due to land use/cover changes in watersheds without measuring all components throughout the study area (Andréassian, 2004). A paired watershed study is especially useful for uncontrolled and landscape disturbances and significant localized changes in soil properties and compaction, such as from military maneuvers. However, no consistent, statistically sound criterion and method exist to process paired watershed selection; many studies use subjective professional judgments to select paired watersheds for their studies. Therefore, the second part of this study utilized a statistically sound technique with topographic and geologic parameters for

paired watershed selection since selected topographic and geologic parameters determine hydrologic processes in small watersheds. This is a useful method for determining hydrologic process disturbances.

1.2.2 Study Goals and Objectives

One of the principal applications of hydrology is forecasting and predicting runoff volumes and flood peaks due to large rain and snowmelt events (Dingman, 2008). Systematic approaches to quantifying hydrologic processes with various landscape and watershed management and the spatiotemporally variability of land use/cover, soil physicochemical properties, and climatic forcing are essential in order to address the ever increasing importance of estimating available water for numerous needs and dealing with extreme events that could potentially affect resources and security.

The overall goal of this study was to improve surface water hydrologic modeling by developing a dynamic method to estimate runoff and identify a systematic approach for paired watershed analysis. This dissertation has three main component studies: curve number based on Normalized Difference Vegetation Index (CN_{NDVI}) development, application of developed CN_{NDVI} in similar small but disturbed land use, and relatively diverse land use larger watersheds, and statistically paired watershed selection.

The first goal of this study was to assess the applicability of satellite data for capturing spatiotemporal changes of hydrologic conditions in order to improve estimation of surface runoff. The hypothesis of the study was that a CN_{NDVI} would provide a more accurate spatiotemporal prediction of runoff than the standard SCS-CN method. The specific objective of this study was to develop a model that estimates CN_{NDVI} as a surrogate of spatial and temporal changes of hydrologic conditions in order to accurately capture the spatiotemporal relationship of

rainfall and runoff. This study was based on12 years of recorded rainfall and runoff data from four small watersheds in the Konza Prairie, Kansas and NDVI for the respective time period.

Regression analysis was used to develop the model to calculate CN using NDVI, and model-based runoff was calibrated and validated using observed (recorded) runoff compared to standard CN-based runoff. The study also addressed the following research questions: How can GIS and remote-sensed products be jointly used to develop periodic and seasonal curve numbers? Are there any temporal and seasonal changes in developed CN? If there are changes, what are the patterns of the changes? Can NDVI provide relatively accurate estimates of rainfall-runoff relationships by developing a CN_{NDVI}? Can CN_{NDVI} provide better runoff estimates than the SCS-CN method?

The second goal of the study was to determine a statistical approach to process paired watershed selections and apply the CN_{NDVI} on small, disturbed grassland watersheds. The specific objectives were to conduct a paired watershed selection by devising a statistical method to select watersheds based on topographic and geologic parameters that dominate hydrologic processes in small watersheds and to apply CN_{NDVI} in disturbed grassland. The paired watershed chapter (Chapter 4) addresses the following research questions: Can statistical techniques be applied to select paired watersheds in order to study hydrological responses? Is using topographic and geologic parameters scientifically sound and statistically feasible? Can CN_{NDVI} be applicable in disturbed lands?

The third goal of this study was to evaluate the application of a previously developed CN_{NDVI} model on relatively diverse land use/cover and large watersheds. The hypothesis was that Daymet precipitation and the CN_{NDVI} model could estimate runoff accurately than standard CN methods for large watersheds with various land use/cover. The specific objectives were to

assess the application of Daymet precipitation and CN_{NDVI} to estimate runoff in relatively large watersheds and to compare CN_{NDVI} runoff of the two watersheds to direct runoff component of stream flow at United States Geological Survey (USGS) outlets of the watersheds. The research questions associated with this goal were: Can the CN_{NDVI} model be applied at a larger scale and various land cover areas? How much spatiotemporal variability in runoff can be addressed using Daymet precipitation and CN_{NDVI}? This study used Daymet data to account for spatial variability of precipitation. The Daymet data, available with 1 km × 1 km spatial resolution, was compared to nearby National Climate Data Center (NCDC) gauge precipitation to assess relative accuracy of the data. The analysis compared and validated Daymet precipitation using NCDC data (within 1 km radius from NCDC points) to use precipitation in rainfall-runoff modeling. The study was done at HUC10 watershed scale.

Overall, the methods explored in this research (CN_{NDVI} development and paired watershed techniques based on statistics) provided a better alternative for studying natural and anthropogenic impacts on hydrologic responses at the watershed scale. The study was applied in two watersheds with distinct land uses in Kansas, but a similar approach could be adapted to various land use categories. The method can also alleviate difficulties in studying environmental impacts of natural and/or disturbed lands and also provides a mechanism to study the effectiveness of conservation practices at larger watershed scales.

References

Andréassian, V. (2004). Waters and forests: From historical controversy to scientific debate. *Journal of Hydrology* 291(1): 1–27.

Arnell, N., Bates, B., Lang, H., Magnuson, J., and Mulholland, P. (1996*). Hydrology and freshwater ecology*. New York: Cambridge University Press.

Boughton, W. (1989). A review of the USDA SCS curve number method. *Soil Res*. 27:511–523.

Dahlgren, R. (1999). Watershed scale. Retrieved from http://lawr.ucdavis.edu/classes/ssc219/biogeo/watersh.htm

Daniel, E. B., Camp, J. V., LeBoeuf, E. J., Penrod, J. R., Dobbins, J. P., and Abkowitz, M. D. (2011). Watershed Modeling and its Applications: A State-of-the-Art Review. *The Open Hydrology Journal* 5:26–50.

Davie, T. (2008). *Fundamentals of hydrology.* New York: Taylor & Francis.

DeFries, R. and Eshleman, K. N. (2004). Land-use change and hydrologic processes: A major focus for the future. *Hydrological Processes* 18(11): 2183–2186.

Dingman, S. L. (2008). *Physical hydrology* (2nd ed.). Illinois: Waveland press.

Enock, F. (2011). The hydrosphere. Retrieved from https://geofreekz.wordpress.com/the-hydrosphere/

Fan, F., Deng, Y., Hu, X., and Weng, Q. (2013). Estimating composite curve number using an improved SCS-CN method with remotely sensed variables in Guangzhou, China. *Remote Sensing* 5(3): 1425–1438.

Fedak, R. (1999). Effect of spatial scale on hydrologic modeling in a headwater catchment Thesis. Master of Science. Virginia Polytechnic Institute and State University.

Genereux, D., Jordan, M., and Carbonell, D. (2005). A paired-watershed budget study to quantify interbasin groundwater flow in a lowland rain forest, Costa Rica. *Water Resources Research* 41(4). doi:10.1029/2004WR003635

Hawkins, R. H., Ward, T. J., Woodward, D. E., and Van Mullem, J. A. (2008). *Curve number hydrology*: *State of the practice*. Maryland: ASCE Publication.

Hawkins, R. H. (2014). Curve number method: Time to think anew? *Journal of Hydrologic Engineering* 19(6): 1059–1059.

Hong, Y. and Adler, R. F. (2008). Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *International Journal of Remote Sensing* 29(2): 471–477. doi:10.1080/01431160701264292

Kousari, M. R., Malekinezhad, H., Ahani, H., and Zarch, M. A. A. (2010). Sensitivity analysis and impact quantification of the main factors affecting peak discharge in the SCS curve number method: An analysis of Iranian watersheds. *Quat. Int.* 226:66–74.

Immerzeel, W. (2008). Historical trends and future predictions of climate variability in the Brahmaputra basin. *International Journal of Climatology* 28(2): 243–254.

Jackson, T., Schmugge, J., and Engman, E. (1996). Remote sensing applications to hydrology: Soil moisture. *Hydrological Sciences Journal* 41(4): 517–530.

Lal, R. and Shukla, M. K. (2004). *Principles of soil physics*. New York: Marcel Dekker, Inc.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., . . . Folke, C. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change* 11(4): 261–269.

Mandapaka, P. V., Krajewski, W. F., Mantilla, R., and Gupta, V. K. (2009). Dissecting the effect of rainfall variability on the statistical structure of peak flows. *Advances in Water Resources* 32(10): 1508–1525.

Oogathoo, S. (2006). Runoff Simulation in the Canagagigue Creek Watershed using the MIKE SHE Model. Department of Bioresource Engineering, M.Sc. Thesis. McGill University, Montreal, Quebec. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.119.6517&rep=rep1&type=pdf

Perlman, H. (2014). The water cycle. Retrieved from http://water.usgs.gov/edu/watercycle.html

Pockrandt, B. R. (2014). A multi-year comparison of vegetation phenology between military training lands and native tallgrass prairie using TIMESAT and moderate-resolution satellite imagery. Thesis. Master of Arts. Kansas State University.

Refsgaard, J. C. (1996). Terminology, modelling protocol and classification of hydrological model codes. In: Abbott M. B., Refsgaard J. C., editors. Distributed hydrological modeling. *Water Science and Technology Library* 22:17-39.

Schmugge, T. (1987). Remote sensing applications in hydrology. *Reviews of Geophysics* 25(2): 148–152.

Schmugge, T. J., Kustas, W. P., Ritchie, J. C., Jackson, T. J., and Rango, A. (2002). Remote sensing in hydrology. *Advances in Water Resources* 25:1367–1385.

Schnur, M., Xie, H., and Wang, X. (2010). Estimating root zone soil moisture at distant sites using MODIS NDVI and EVI in a semi-arid region of southwestern USA. *Ecological Informatics* 5(5): 400–409. doi:10.1016/j.ecoinf.2010.05.001

Song, X., Kong, F., Zhan, C., and Han, J. (2011). Hybrid optimization rainfall-runoff simulation based on Xinanjiang model and artificial neural network. *Journal of Hydrologic Engineering* 17(9): 1033–1041.

Tollan, A. (2002). Land-use change and floods: What do we need most, research or management? *Water Science & Technology* 45(8): 183–190.

Wilson, J. P. (2012). Digital terrain modeling. *Geomorphology* 137(1): 107–121.

Woodward, D. E., R. H. Hawkins, A. T. Hjelmfelt, J. A. VanMullem, and Q. D. Quan (2002). Curve number method: Origins, applications, and limitations. U.S. Geological Survey Advisory Committee on Water Information - Second Federal Interagency Hydrologic Modeling Conference. July 28–August 1, (2002), Las Vegas, Nevada.

Chapter 2 - Literature Review

This study encompassed a multidisciplinary approach to study watersheds, such as surface water hydrology, remote sensing, and GIS. The major goal of the study was to improve surface water modeling through remote sensing and GIS. The study developed CN_{NDVI} and systematic paired watershed selection methods. This literature review provides a general overview of surface water modeling, especially runoff prediction pertaining to remote sensing and GIS. A broad overview of watershed modeling, land use/cover changes and management, remote sensing and GIS, the importance of spatiotemporal variability in hydrology, rainfall-runoff modeling, and the CN method are also discussed. In addition, data acquisition and uncertainty related to hydrologic modeling are addressed, and the paired watershed, a widely applied practice to study hydrologic responses of different treatments, is discussed.

2.1 Modeling the Watershed

Detailed field studies emerged in the 1960s in order to broaden understanding of the physical processes by which water enters streams; conceptual and mathematical models were used to comprehend and quantify those processes (Hubbart, 2008). Watershed modeling attempted to understand the components of hydrologic processes as well as water movements and interactions with vegetation and physicochemical properties of sol at the surface, subsurface, and interface. Watershed models are valuable tools that provide valuable information for water resource decisions. Jeon et al. (2014) described watershed modeling as a rational, economical, and useful approach for water resource planning, design, and decision making. Moriasi et al. (2007) portrayed watershed models as powerful tools to quantify hydrologic processes and simulate the effects of hydrologic processes within the watershed which increase management efficiency of soil and water resources. Watershed scale models such as the AGNPS, the

Kinematic Runoff and Erosion Model (KINEROS), Hydrologic Simulation Program-FORTRAN (HSPF), MIKE SHE, and SWAT were developed to address water quality issues (Borah and Bera, 2003).

Hydrologic models can be conducted from field-scale to watershed-scale levels with various sizes to quantify the amount of water in each hydrologic process component, assess flood and drought occurrences, and estimate the amount of sediment and pollutant loads (Garen and Moore, 2005; Parajuli and Ouyang, 2013). A watershed is an area of land from which all runoff from rainfall excess and/or snow melt flows to a certain point. It can be considered a closed boundary in order to quantify water as input, output, and storage in watershed models. GIS advancement, remote sensing, and data computing capabilities of computers increase the applicability of hydrologic models at various scales.

Hydrologic models have varied from lumped to distributed models throughout the conceptualization and parameterization evolution of hydrologic processes in catchment models over several decades of research (Todini, 2007). Watershed models can be categorized spatially as lumped, semi-distributed, or distributed models. In the lumped watershed approach, a watershed is often considered as a single unit, and watershed parameters and variables are averaged over this unit. For semi-distributed or distributed model approaches, the models account for spatial variability of hydrologic processes and parameters, and variables are allowed to vary. Advancements of digital computation and increasingly improving resolution data from remote satellites have caused rapid development and acceptance of distributed hydrologic models. Temporal variability is often included in event-based or continuous process watershed models. Event-based models consider individual rainfall events, focusing on infiltration and

surface runoff; continuous process models account for all runoff components while considering soil moisture redistribution between storm events (Melone et al., 2005).

Watershed models utilize mathematical expressions or conceptual categorization to quantify and predict components of hydrologic processes. These models also use precipitation, land use/cover, impervious areas, slope, soil type, drainage area, and other topographic characteristics as input data to characterize variability and quantify components of hydrologic processes (US EPA Ecosystems Research, 2013). Watershed model accuracy depends on complex hydrologic process representation and how well inputs are parameterized. Land use/cover is a crucial variable used in watershed models to characterize spatial variability.

2.2 Land Use/Cover Changes and Management

Land is one of Earth's foremost natural treasures, but its use and availability depends on population growth and suitable land management. Land use/cover changes are induced primarily by population growth (DeFries et al., 2004; Geist and Lambin, 2002; Lambin et al., 2001; Mul, 2009) and the changing climate (Millennium Ecosystems Assessment, 2005). As the population expands, more land is changed from natural systems to agricultural, residential, and industrial areas (Barbosa et al., 2012; De Fries et al., 2004; DeFries and Eshleman, 2004; Isik et al., 2013). In addition, the changing climate may exacerbate land use/cover changes that affect hydrologic processes.

Changes in land use/cover affect local, regional, and global hydrologic processes and water management activities (Foley et al., 2005; Schilling et al., 2008; Mao and Cherkauer, 2009; Elfert and Bormann, 2010; Ghaffari et al., 2010). Land owner's and manager's choices about land use and land cover patterns also exacerbate the effects of climate change. The impacts of land use/cover change on surface and near-surface landscape affect runoff generation

processes, including increasing direct runoff volume by interfering natural infiltration rates or surface storage or both. Land use changes can also significantly alter leaf area index (LAI) and ET (Mao and Cherkauer, 2009), soil moisture content and infiltration capacity (Fu et al., 2000; Costa et al., 2003), subsurface flow (Tu, 2009), ground water recharge, and as well as cause lag effects in stream flow. Land use change can also alter the presence and extent of perched water tables and alter the generation of overland flow (Germer, 2010). Infiltration capacity of the soil determines the amount and time of rainfall excess distribution that is available for runoff and surface storage (Garen and Moore, 2005). Zimmermann et al. (2006) examined how land use influences soil hydraulic conductivity and infiltrability, thereby affecting other components of hydrologic processes, such as surface runoff.

Quantification of hydrologic processes at field and watershed scales is essential for understanding hydrologic responses of land management activities, including natural and anthropogenic disturbances. Various methodologies have been implemented to attempt to understand the effects of land use and land cover change on hydrologic processes. Most current understanding of hydrological impacts of land use/cover changes were derived from paired watershed comparisons of controlled and experimental watersheds (Andréassian, 2004; Hong and Adler, 2008; King et al., 2008). Hydrological models, which assess the impact of land use changes, are benefiting from digital computation advancements and increasing availability of high resolution, remotely sensed data.

2.3 Remote Sensing in Hydrological Modeling

The invention and development of GIS and remote sensing have significantly improved hydrological modeling and advanced the physical understanding of hydrology. Schmugge et al. (2002) defined remote sensing as the process of inferring surface parameters from measurements

of upwelling electromagnetic radiation from the land surface. Remote sensing is applied in environmental, hydrological and climatological studies, including estimation of soil moisture, ET, snow pack and melt, precipitation, land surface temperature, and vegetation covers (Khanbilvardi e al., 2014). GIS analyzes spatial information, manages georeferenced data such as the digital elevation model (DEM), creates interactive queries, and presents model results in maps. It also helps formulate spatially distributed hydrological models (Mitas and Mitasova, 1998).

Distributed hydrologic models require spatial distribution of meteorological and geographical elements such as temperature, precipitation, humidity, solar radiation, and other observation data as their main inputs or forcing parameters. Because traditional hydrologic data are point/field measurements, hydrologic analysis are limited by spatial data availability. Satellite remote sensing data have emerged as a viable alternative or supplement to in situ observations due to its availability for implementation and calibration of hydrologic models over vast ungauged regions. Distributed hydrologic model demands are often met by integrating GIS and remote sensing products. Landsat (mainly TM and ETM+), Satellite Pour l'Observation de la Terre (SPOT), MODIS, National Oceanic and Atmospheric Administration - Advanced Very High Resolution Radiometer (NOAA–AVHRR), IKONOS, and QuickBird are commonly used remote sensing products in data acquisition tasks. Scientists monitor patterns of land cover change over space and time at regional, national, and global scales using satellite and other remotely sensed data (Slonecker, et al., 2013).

Remote sensing data is widely applied in hydrologic modeling, especially to quantify land use/cover changes (Alexakis et al., 2013; Bhaduri et.al, 2000; Li, et. al, 2013; Zacharias et. al, 2004). Zacharias et al. (2004) used remote sensing and modeling techniques to quantify land

use alterations and associated hydrologic impacts. Charoenhirunyingyos et al. (2011) proposed remote sensing data as an emerging alternative to estimate soil hydraulic parameters. Band et al. (1993) used GIS processing of remote sensing data to study ecosystem processes at a watershed scale. In addition to use in many water quality studies, remote sensing techniques have been used to map snow and soil moisture and to monitor crop development (Schmugge et al., 2002). Remotely sensed data have also been used to estimate land surface water balance variables such as precipitation, ET, snow and ice, soil moisture, and terrestrial water storage variations at various spatial and temporal resolutions and accuracies (Tang et al., 2009).

2.3.1 Normalized Difference Vegetation Index and Quality Check

The Normalized Difference Vegetation Index (NDVI) is an index that measures "greenness" of vegetation. NDVI is calculated as the normalized difference between near-infrared (841–876 nm) and red (620–670 nm) bands as

$$NDVI = \frac{NIR-Red}{NIR+Red} \qquad (2\text{-}1)$$

NDVI is one of the most commonly used vegetation indices (VIs) to monitor vegetation conditions and display land use/cover and its changes. NDVI has been successfully applied worldwide to study temporal and spatial trends and variation in vegetation distribution, productivity, and dynamics and to monitor habitat degradation, fragmentation, and ecological effects. In addition, NVDI has been applied to monitor climatic disasters such as drought (Singh et al., 2003), fire (Maselli et al., 2003), flood (Wang et al., 2003), and frost (Tait and Zheng, 2003). NDVI also has been used as a pseudo indicator of soil moisture conditions (Narasimhan et al., 2005) and derived LAI to monitor potential ET (Zhou et al., 2006). Global usage of NDVI suggests that any application can be adapted to various regions.

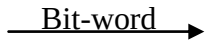
MODIS provides NDVI globally over land at 16-day composite periods (MOD13Q1). The 16-day composite image is less noisy than daily data. MODIS NDVI products are already calibrated and atmospherically corrected bi-directional surface reflectance that has been masked from cloud and cloud shadows, land/water, and aerosol products. However, persistent cloud cover is a potential threat during the 16-day period. Therefore, pixel reliability ranking and VI quality imagery, which are 16 days 250 m as that of other VIs, QA Science Datasets (SDS) can be used to assess the quality assurance of MOD13Q1 products.

Detail quality assessment of MODIS VI products is described in NASA LP DAAC (2013), and NDVI quality is summarized in the pixel reliability, which is represented by numbers ranking from -1 to 4 to depict pixel level data quality (Table 2-1). VI quality pertains to acquired and processed conditions of each pixel (Table 2-2).

Table 2-1: Pixel realization rank values and descriptions (Adopted from NASA LP DAAC, 2013)

Pixel Reliability Rank	Summary QA	Description
-1	Fill/No data	Not processed
0	Good data	Use with confidence
1	Marginal data	Useful but look at other QA information
2	Snow/Ice	Target covered with snow/ice
3	Cloudy	Target not visible, covered with cloud
4	Estimated	Based on MODIS historic time-series. All products are gap-filled, indicating whether or not the value was interpolated from long-term averages

VI QA bit values can be converted to 16-bit binary numbers for interpretation in which the binary bit-string must be read from right to left, as shown in the following bit-string direction. Individual bits within a bit-field are read from left to right, as shown via the bit-word arrow (NASA LP DAAC, 2013). The following values were common at the Konza Prairie Biological Station, Kansas, for this study; so that it is described as an example of interpretation.

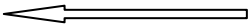


2116 = 0|0|001|0|0|0|01|0001|00

2120 = 0|0|001|0|0|0|01|0010|00

2185 = 0|0|001|0|0|0|10|0010|01

2189 = 0|0|001|0|0|0|10|0011|01



Bit-string direction for interpretation

Table 2-2: Bit-no, parameters, bit-words, and interpretations (Adopted from NASA LP DAAC, 2013)

Bit-No	Parameter	Bit-Word	Interpretation
0-1	VI Quality	00	VI produced with good quality
		01	VI produced but check other QA
		10	Pixel produced but most probably cloudy
		11	Pixel not produced due to reasons other than clouds
2-5	VI usefulness	0000	Highest quality
		0001	Low quality
		0010,0001,0010,0100, 1000,1001, 1010	Decreasing quality
		1100	Lowest quality
		1101	Quality so low it is not useful
		1110	L1B data faulty
		1111	Not useful for any other reason
6-7	Aerosol quantity	01	Low
		10	Intermediate
8	Adjacent cloud detected	0	No
9	Atmospheric BRDF Correction	0	No
10	Mixed clouds	0	No
11-13	Land-water mask	001	Land (Nothing else but land)
14	Possible snow/ice	0	No
15	possible shadow	0	No

2.4 Spatiotemporal Variability in Hydrology

Spatiotemporal heterogeneity and dynamic landscapes are the products of complex interactions of abiotic and biotic processes at different scales (Schroeder, 2006). Spatiotemporal variability of hydrologic processes due to land use/cover changes, soil physicochemical properties, and climatic variables such as precipitation, humidity, temperature, solar radiation, and wind speed and direction are the main factors to affect hydrologic processes. In addition to examining a variable's spatial and/or temporal variability, the scale at which the variable is applied in the study should also be considered. Spatiotemporal variability is an important part in rainfall-runoff modeling because runoff variability depends on factors such as the amount and intensity of precipitation, watershed topographic characteristics, land use/cover of the watershed, and soil physicochemical properties. Figure 2-1 shows the spatial and temporal scale of topographic, soil, vegetation, and atmospheric phenomenon. At the spatial scale hydrologic processes range from micro level to regional basin level; at the temporal scale processes range from minutes to years. As shown in Figure 2-1, for instance, the infiltration/ soil moisture related processes vary from minutes to daily at temporal scale, and with micro level at spatial scale. Infiltration and soil moisture are the primary factors in runoff generation processes. Accounting the spatial and temporal variability of topographic, soil, vegetation and weather conditions is crucial to accurately estimate runoff from rainfall events.

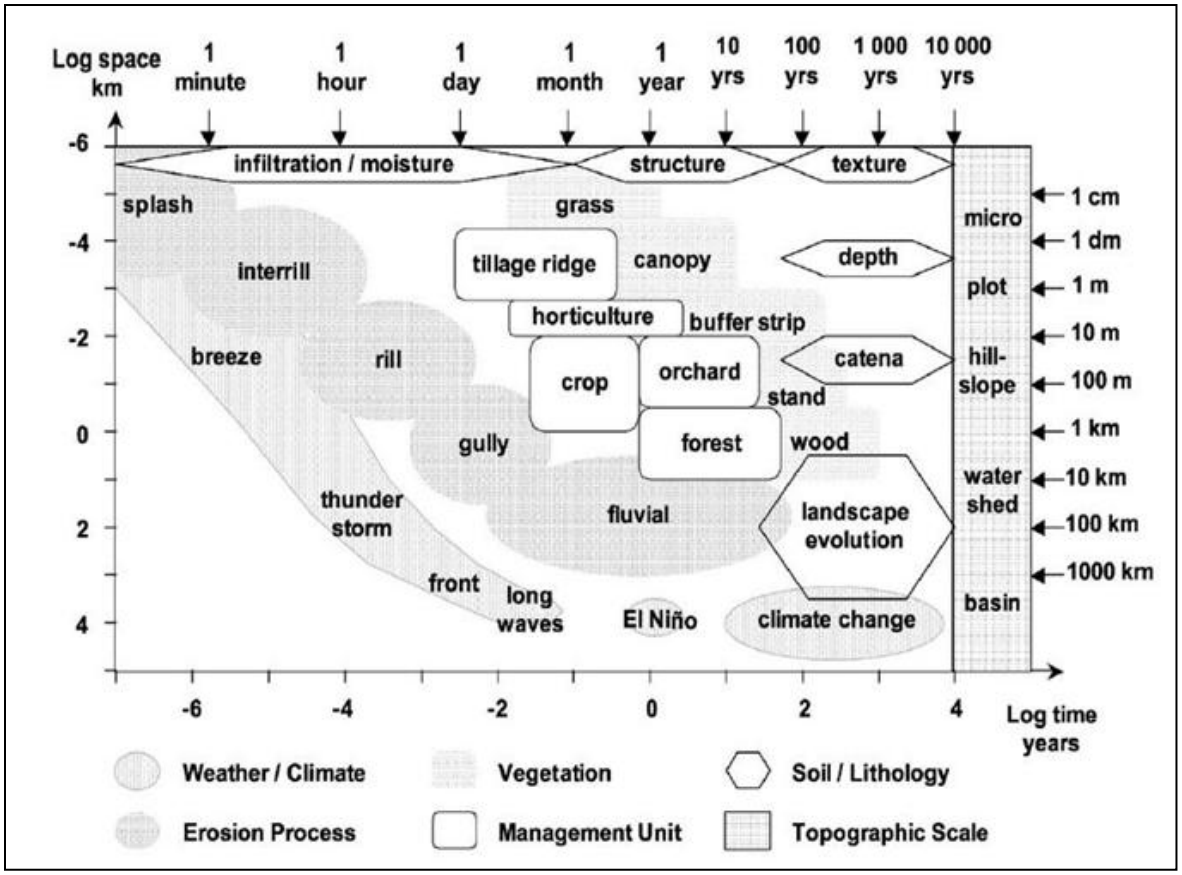


Figure 2-1: Spatial and temporal scale of factors affecting hydrologic and soil erosion processes (adopted with permission from Renschler and Harbor, 2002)

Blöschl and Sivapalan (1995) clearly explained the spatial versus temporal scale of precipitation and runoff, as shown in Figure 2-2, which depicts water movement at various scales. The figure shows that precipitation varied from several minutes to more than a day at the temporal scale, and precipitation ranged from meters (i.e., cumulus convection) to thousands of kilometers (e.g., frontal systems) in the spatial scale. Similar to other hydrologic processes, runoff processes operate in response to precipitation and at similar length scale (Blöschl and Sivapalan, 1995); however, the amount differs based on spatial heterogeneity and temporal scales vary on dominant runoff mechanisms. For example, at a small spatial scale (e.g., 1 km^2) with high rainfall intensities, infiltration excess or Horton overland flow is very fast (<30

minutes), but saturation excess or saturation overland flow is slower because saturated layers require time to accumulate (Blöschl and Sivapalan, 1995).

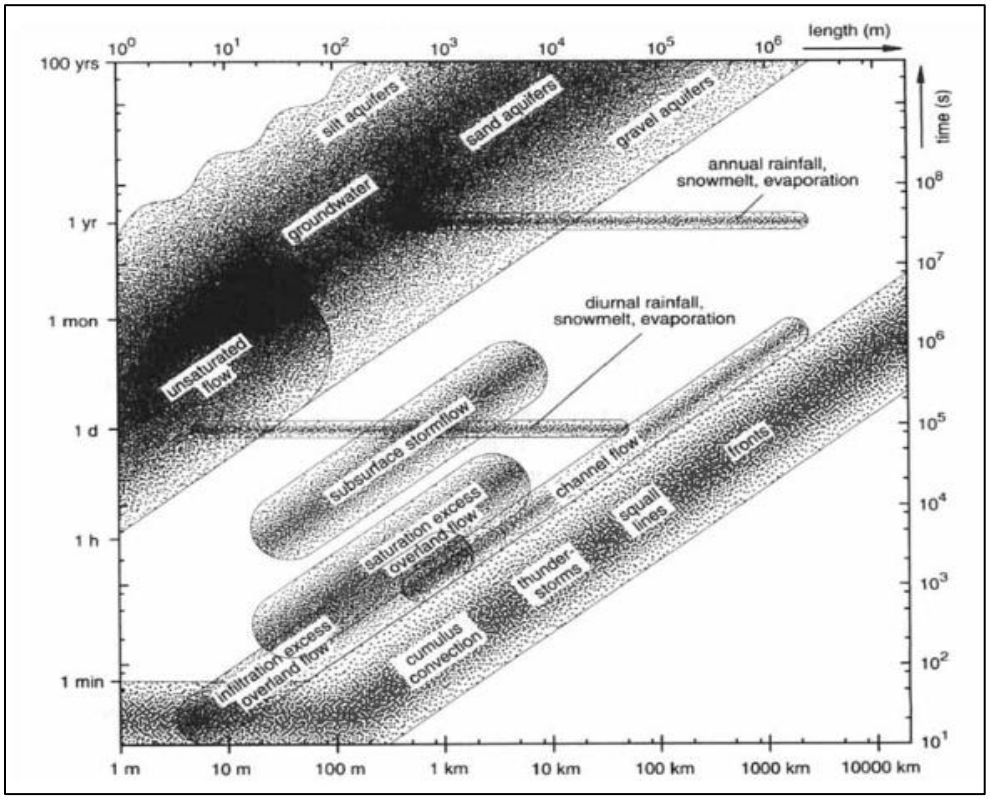


Figure 2-2: Schematic relationship between spatial and temporal process scales in hydrology (adopted with permission from Blöschl and Sivapalan, 1995)

2.5 Rainfall-Runoff Modeling

The rainfall and runoff relationship is a highly nonlinear, complex relationship that exhibits a high degree of spatial and temporal variability (Isik et al., 2013; Solomatine and Dulal, 2003). This relationship is dependent upon factors such as soil type, antecedent soil moisture, evaporation, infiltration, distribution and / or duration of rainfall, slope, and catchment size (Critchley et al., 1991; Isik et al., 2013). However, rainfall-runoff models are widely accepted as simplified, appropriate methods to estimate the amount of runoff from precipitation events by approximating all other components of hydrologic processes as losses (Dingman, 2008; Wagener and McIntyre, 2005). According to Beven (2012), rainfall-runoff models are categorized as

perceptual, conceptual, and procedural. Beven (2012) and Džubáková (2010) stated that perceptual models summarize modeler perceptions of how the catchment responds to rainfall under various conditions, conceptual models include a mathematical description of the hydrologic process, and procedural models contain the code of techniques of numerical analysis that run on the modeling computer.

Various rainfall-runoff models have been developed and used, including the SCS-CN method, the rational method, the time-area method, and the unit hydrograph method (Dingman, 2008). The SCS-CN method is a popular rainfall-runoff model that creates a relationship between rainfall events and runoff using an empirical number that depends upon land cover, hydrologic soil group, and hydrologic conditions of the area. The rational method is a simple method that determines peak discharge of watershed runoff based on drainage area, rainfall intensity, and runoff coefficient. The rational method assumes that rainfall occurs uniformly over the drainage area; peak rate of runoff can be reflected by rainfall intensity, and the frequency of occurrence for peak discharge is identical to the frequency of rainfall producing that event (Beven, 2012). The time-area method is the first distributed model developed based on the time step, which considers the lag in time of runoff from a subdivided catchment area to an outlet. The time-area method transforms an effective storm hyetograph into a runoff hydrograph (Ponce, 1989). It assumes that the outflow hydrograph results from the direct runoff to the outlet at uniform velocity. The unit hydrograph is the runoff hydrograph generated by an excess precipitation of one unit (1 mm) that flows towards the catchment outlet. Detailed descriptions of the area-time method, unit hydrograph, and the SCS-CN method are presented in Ponce (1989) and Bedient et al. (2008).

2.6 Curve Number Method

The Curve Number (CN) method was developed in the 1950s by USDA, formerly Soil Conservation Service (SCS) and currently Natural Resources Conservation Service (NRCS), to meet the needs of the agency in implementing Public Law 566 (Hawkins et al., 2008; Hawkins, 2014) and is referred to as SCS-CN method. It is a widely used means of estimating storm runoff from rainfall events and designing water management structures. Its simplicity, ease of use, predictability, stability, reliance on only one variable (which is CN to predict the rainfall-runoff relationship), and responsiveness to major runoff producing watershed properties such as soil, land use, surface condition and antecedent moisture condition are among the perceived advantages of CN method (Ponce and Hawkins, 1996; Garen and Moore, 2005, Hong and Adler, 2008; Hawkins, 2014). Though it was developed to estimate the amount of direct runoff from event precipitations, the method is incorporated into several continuous hydrologic models such as Soil and Water Assessment Tool (SWAT), Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS), Erosion Productivity Impact Calculator (EPIC), and Agricultural Non-Point Source Pollution Model (AGNPS) (Kousari et al., 2010).

The following steps were used to derive the SCS-CN equation (Hawkins et al., 2008; Mishra and Kumar, 2002):

1. The original equation assumed the basic water budget at watershed level in depth units as

$$P = Q + F \qquad (2\text{-}2)$$

where P is rainfall depth, Q is runoff depth, and F is all losses. Initial abstraction (I_a) was not included in the original equation.

2. Maximum potential losses to runoff (S) were incorporated as a limiting factor for runoff.

$$S = \lim_{p \to \infty}(P - Q) = \lim_{p \to \infty}(F) \qquad (2\text{-}3)$$

3. Proportionality equation hypothesis was used to explain the runoff fraction (*Q/P*).

$$\frac{Actual\ losses}{Potential\ losses} = \frac{F}{S} = \frac{Q}{P} \qquad (2\text{-}4)$$

Hawkins et al. (2008) noted that the hypothesis is true when $P = 0$ and $P = \infty$; however, its validity in the interval is questionable.

4. The following equation was derived from equation in step 2-2 and 2-4:

$$Q = \frac{P^2}{(P+S)} \qquad (2\text{-}5)$$

5. Through USDA-SCS interagency review, the initial abstraction (I_a) was not incorporated, and I_a was determined to 20% of the maximum retention potential. Effective rainfall (P_e) was used instead of total rainfall, resulting in the following equations:

$$I_a = 0.2S \qquad (2\text{-}6)$$
$$P_e = P - I_a \qquad (2\text{-}7)$$

6. Based on steps 2-5 and 2-6,

$$Q = \frac{{P_e}^2}{P_e + S} \qquad (2\text{-}8)$$

$$Q = \frac{(P-I_a)^2}{(P-I_a+S)} \qquad (2\text{-}9)$$

7. The following equation was developed from steps 2-6 and 2-9:

$$Q = \frac{(P-0.2S)^2}{(P+0.8S)} \qquad P \geq 0.2S \qquad (2\text{-}10)$$
$$otherwise \quad Q = 0$$

8. Maximum retention potential (*S*) was transformed to CN, which was determined to be 0–100. CN is a dimensionless empirical number that determines the relationship of rainfall event and direct runoff obtained using lookup tables developed by USDA-SCS based on land cover, HSG, hydrologic condition based on ground cover, and antecedent moisture

conditions. The relationship between CN and *S* was developed based on recorded events of rainfall and runoff of research watersheds in the United States.

$$S = \frac{25400}{CN} - 254 \quad or \quad CN = \frac{25400}{S+254} \tag{2-11}$$

where *S* is in millimeter (mm). When the depth unit of *S* changes, the numbers in the equation change accordingly. Back-calculated CNs in this study were calculated from 12 years of precipitation event and runoff depth records.

The CN method has been adapted to areas with various land use/cover and climatic conditions throughout the world and successfully applied to situations ranging from simple runoff calculations and land use change assessment to comprehensive hydrologic/water quality simulations (Auerswald and Haider, 1996; Garen and Moore, 2005; Ponce and Hawkins, 1996). CN application can be grouped into three distinctly different modes: determination of runoff volume for a given return period, determination of direct runoff for individual events, and process model or infiltration model or a soil moisture-CN relationship (Van Mullem, 1989). Although the CN method has been accepted and applied worldwide, studies have indicated that the method should be evaluated and calibrated at local and regional scales (Hawkins and Cate, 1998; Van Mullem, 1989). Remote sensing and GIS have improved CN research by defining land use/cover and enabling data processing and mapping, but CN application still requires SCS-CN lookup tables to determine the CN.

2.7 Data Acquisition and Uncertainty

Comprehensive scientific work requires statements of accuracy in order to understand and communicate reliability of results. If accuracy is known objectively, then it can be expressed as error; if it is not known objectively, then the results are uncertain (Gahegan and Ehlers, 2000; Hunter and Goodchild, 1993). Uncertainty is the degree, not the actual value, of discrepancy

between geographical data in GIS and the geographical reality these data represent (McMaster and Usery, 2004). Uncertainty in hydrologic prediction continues to challenge modelers despite significant recent developments in computational power and distributed hydrologic modeling (Liu and Gupta, 2007). A model cannot realistically be free of uncertainty and bias, but reducing the bias and uncertainty must be a priority in planning and implementing hydrologic models for prediction. Every aspect of modeling is affected by uncertainty and error from sampling, measurements, and prediction procedures, depending on spatial and temporal scales of the study.

No spatial model is a perfect representation of reality; modeling is the only means of addressing complex environmental problems in large-scale studies (Klepper, 1997; Li and Wu, 2006; Petersen, 2000). Each model and transformation process contributes to the overall uncertainty within the data (Gahegan and Ehlers, 2000). Those uncertainties could originate from model inputs such as precipitation, land use/cover, soil physicochemical properties, and other climatic forcing. Uncertainties could also be a result of inaccurate representation of the processes and spatial and temporal scales of data representation. Liu and Gupta (2007) asserted that uncertainty must be cohesively and systematically understood, quantified, and reduced in hydrologic modeling.

2.7.1 Weather and Climate Data

Weather and climate forecast predictability is determined by uncertainty projections in initial conditions and model formulation onto flow-dependent instabilities of the chaotic climate attractor that embraces errors and uncertainties (Palmer, 2000). Depending on the model type, weather and climatic data are used as input for predicting and quantifying hydrologic processes. Weather and climate data are often measured and recorded at points, for spatial modeling purposes point values are usually interpolated to a relatively larger area or scale (Li et al., 2013;

Merz et al., 2006). Hydrologic processes occur within a wide range of scales and span approximately eight orders of magnitude in space and time (Klemes, 1983). For example, precipitation occurrence scale can range from1 m (cumulus convection) to 1000 km (frontal systems). Most hydrologic studies, however, use only point measurements in which transformation of information across scales, referred to as scaling, introduces most uncertainties and biases at different levels. Rainfall is the one of the main source of uncertainty in most hydrologic models due to its variability and its use as the main climatic forcing input (Srinivasan, 2005; Tuppad, 2006).

2.7.2 Satellite Data Uncertainty

GIS has provided hydrologists new platforms for data management and visualization in hydrologic applications because it can process DEM and other remote sensing related data (Sui and Maggio, 1999; Zhang et al., 2003). Gahegan and Ehlers (2000) determined that GIS provides a whole series of tools with which data can be manipulated, but it does not offer control over misuse. Spatial and/or temporal scale inconsistencies of input data is one of most often misused components in software such as GIS. Scale issues in hydrology manifest in various space-time scales because physical processes are described by mathematical relationships (Gupta et al., 1986). Changing the scale without understanding the effects of such an action can result in the representation of processes or patterns that are different from intended (McMaster and Usery, 2004). Scaling and its effects on hydrologic modeling are linked to heterogeneity (Blöschl and Sivapalan, 1995). Heterogeneity that affects scaling is small at small scales and large at large scales; the greater the degree of heterogeneity, the smaller the scale must be to represent variability. Hydrologic parameters are often used as lumped parameters in order to represent

entire watersheds, but watershed characteristic data is collected only at a limited number of field locations (Singh, 1995).

2.8 Paired Watershed Studies

The principle of the paired-watershed design is simple and remains the reference for all studies of the impact of land use /cover changes on hydrology (Andréassian, 2004; Hewlett, 1971; Hewlett, 1982). Paired watershed design is based on two watersheds (control and treatment) similar in terms of size, morphology, geology, climatic forcing, and land use (Andréassian, 2004), with the assumption that their similarities will cause both watersheds to respond similarly if all conditions remain identical. However, the pretreatment period must be monitored in order to effectively study hydrologic responses of land use/cover changes or any other treatment.

The paired watershed approach captures the effects of climate and hydrologic differences due to land use/cover changes in watersheds and eliminates the need to measure all change components throughout the study area (Andréassian, 2004). This approach, which has been used to study the impact of land use change in hydrology (DeFries and Eshleman, 2004; Schilling and Spooner, 2006; King et al., 2008; Prokopy et al., 2011), also provides high quality experimental data that could advance understanding of the hydrologic response of watersheds to land use changes (DeFries and Eshleman, 2004). The paired watershed approach has been extensively used to investigate the effects of forestry practices on water yield, streamflow, and aquatic habitat (Bishop et al., 2005; Skaugest, 2005), as well as to document best management practices (BMP) effects (Spooner et al., 1985), organic carbon loss (Veum et al., 2009), soil erosion (Ricker et al., 2008), and non-point source pollution (Prokopy et al., 2011) and to assess conservation practices (King et al., 2008).

2.8.1 Clustering Techniques

Clustering techniques are commonly used in paired watershed selection. Two methods of clustering, k-means and hierarchical clustering, are the most widely used methods. K-means clustering is a partition clustering algorithm that clusters data into *k* groups in order to minimize the within-group sum-of-squares (Cao et al., 2013; Martinez and Martinez, 2008; Martinez and Martinez, 2004). Divisive and agglomerative hierarchical clustering are popular in gene expression studies and proteomics (Martinez and Martinez, 2008; Martinez and Martinez, 2004; Varshavsky et al., 2008). The divisive method starts with one large group and successively split the groups until there is one observation per group which is computationally inefficient (Martinez and Martinez, 2008; Martinez and Martinez, 2004). The agglomerative method, the most widely used hierarchical method, however, starts with *n* groups (one observation per group) and successively merges most similar groups until only one group left (Lu and Wan, 2013; Bettegoni et al., 2006; Martinez and Martinez, 2008). Agglomerative clustering has been applied to study robustness of fish assemblages (Singh et al., 2011), identify protein complexes (Yu et al., 2011), and incorporate social context variables in paired watershed study design in order to study non-point source program effectiveness (Prokopy et al., 2011).

Each of these well-known clustering techniques has unique advantages and disadvantages. When a similar pattern exists, hierarchical clustering cannot determine distinct clusters; actual expression patterns become less effective when cluster size increases. However, k-means clustering requires a specified number of clusters in advance, and it is sensitive to outliers. A combination of both clustering methods is a novel technique that allows utilization of features and merits of both methods.

References

Alexakis, D., Grillakis, M., Koutroulis, A., Agapiou, A., Themistocleous, K., Tsanis, I., and Aristeidou, K. (2014). GIS and remote sensing techniques for the assessment of land use change impact on flood hydrology: The case study of Yialias Basin in Cyprus. *Natural Hazards and Earth System Science* 14(2): 413–426.

Andréassian, V. (2004). Waters and forests: From historical controversy to scientific debate. *Journal of Hydrology* 291(1): 1–27.

Auerswald, K. and Haider, J. (1996). Runoff curve numbers for small grain under German cropping conditions. *Journal of Environmental Management* 47(3): 223–228.

Band, L.E., Patterson, P., Nemani R., and Running S.W. (1993). Forest ecosystem processes at the watershed scale: incorporating hillslope hydrology, *Agricultural and Forest Meteorology* 63(1–2): 93–126.

Barbosa, A. E., Fernandes, J. N., and David, L. M. (2012). Key issues for sustainable urban stormwater management. *Water Research* 46(20): 6787–6798. doi:10.1016/j.watres.2012.05.029

Bedient, P. B., Huber W. C., and Vieux, B. E. (2008). *Hydrology and Floodplain Analysis* (4th ed.). New Jersy: Prentice-Hall Publishing Co.

Beven, K. J. (2012). *Rainfall-runoff modelling: The primer* (2nd ed.). New Jersy:Wiley-Blackwell.

Bhaduri, B., Harbor, J., Engel, B., and Grove, M. (2000). Assessing watershed-scale, long-term hydrologic impacts of land-use change using a GIS-NPS model. *Environmental Management* 26(6): 643–658.

Bishop, P. L., Hively, W. D., Stedinger, J. R., Rafferty, M. R., Lojpersberger, J. L., Bloomfield, and J. A. (2005). Multivariate analysis of paired watershed data to evaluate agricultural best management practice effects on stream water phosphorus. *J. Environ. Qual.* 34 (3): 1087–1101.

Blöschl, G., and M. Sivapalan (1995). Scale issues in hydrological modeling—A review, *Hydrol. Processes* 9:251–290.

Borah, D. K. and Bera, M. (2003). Watershed-scale hydrologic and nonpointsource pollution models: Review of mathematical bases. *T ASAE* 46(6): 1553–66.

Cao, J., Z. Wu, J. Wu, and W. Liu (2013). Towards information-theoretic K-means clustering for image indexing, *Signal Process* 93:2026–2037.

Charoenhirunyingyos, S., Honda, K., Kamthonkiat, D., and Ines, A. V. (2011). Soil moisture estimation from inverse modeling using multiple criteria functions. *Computers and Electronics in Agriculture* 75(2): 278–287.

Colombo R., Busetto L., Fava F., Di Mauro B., Migliavacca M., Cremonese E., GalvagnoM., Rossini M., Meroni M., Cogliati S., Panigada C., Siniscalco C., and di Cella U.M. (2011). Phenological monitoring of grassland and larch in the Alps from Terra and Aqua MODIS images. *Italian Journal of Remote Sensing* 43:83–96. doi: http://dx.doi.org/10.5721/ItJRS20114336

Costa, M. H., Botta, A., and Cardille, J. A. (2003). Effects of large-scale changes in land cover on the discharge of the Tocantins River, Amazonia., *J. Hydrol*. 283:206–217.

Critchley, W., K. Siegert, and C. Chapman (1991). A manual for the design and construction of water harvesting schemes for plant production. Retrieved from http://www.fao.org/3/a-u3160e/index.html

DeFries, R. and N. Eshleman (2004). Land-use change and hydrologic processes: a major focus for the future, *Hydrol. Process*. 18:2183–2186.

DeFries, R., Foley, J., and Asner, G. (2004). Land-use choices: Balancing human needs and ecosystem function. *Frontiers in Ecology and the Environment* 2(5): 249–257. doi:10.1890/1540-9295(2004)002(0249:LCBHNA)2.0.CO 2

Dingman, S. L. (2008). *Physical hydrology* (2nd ed.). Illinois: Waveland press.

Džubáková, K. (2010). Rainfall-runoff modelling: Its development, classification and possible applications. *Acta Geographica Universitatis Comenianae* 54(2): 173–181.

Elfert, S. and Bormann, H. (2010). Simulated impact of past and possible future land use changes on the hydrological response of the Northern German lowland "Hunte" catchment, *J. Hydrol.* 383:245–255.

Foley, J. A., Defries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., . . . Snyder, P. K. (2005). Global consequences of land use. *Science* (New York, NY) 309(5734): 570–574. doi:309/5734/570 (pii)

Fu, B., Chen, L., Ma, K., Zhou, H., and Wang, J. (2000). The relationships between land use and soil conditions in the hilly area of the loess plateau in northern Shaanxi, China. *CATENA* 39(1): 69–79.

Gahegan, M. and Ehlers, M. (2000). A framework for the modelling of uncertainty between remote sensing and geographic information systems. *ISPRS Journal of Photogrammetry and Remote Sensing* 55(3): 176–188.

Garen, D. C. and D. S. Moore. (2005). Curve number hydrology in water quality modeling: uses, abuses, and future directions. *Journal of the American Water Resources Association (JAWRA)* 41(2): 377–388.

Geist, H. J. and E.F. Lambin. (2002). Proximate causes and underlying driving forces of tropical deforestation. *Bioscience* 52(2): 143. doi:10.1641/0006-3568(2002)052(0143:PCAUDF)2.0.CO 2

Germer, S., Christopher, N., Krusche, A.V., and Elsenbeer, H. (2010). Influence of land-use change on near-surface hydrological processes: Undisturbed forest to pasture. *Journal of Hydrology* 380(3–4): 473–480. doi:10.1016/j.jhydrol.2009.11.022

Ghaffari, G., Keesstra, S., Ghodousi, J., and Ahmadi, H. (2010). SWAT-simulated hydrological impact of land-use change in the Zanjanrood Basin, Northwest Iran., *Hydrol. Process*. 24(7): 892–903.

Gupta, V. K., Rodriguez-Iturbe, I., and Wood, E. F. (1986). Preface. In Scale Problems in Hydrology. Dordrecht: D. Reidel Publishing Company.

Hawkins, R. and Cate, A. (1998). Secondary effect in curve number rainfall runoff. *Presented at Water Resources Engineering* 98, Memphis Tennessee, August 3-7. (abstract)

Hawkins, R. H. (2014). Curve number method: Time to think anew*? Journal of Hydrologic Engineering* 19(6): 1059–1059.

Hawkins, R. H., Ward, T. J., Woodward, D. E., and Van Mullem, J. A. (2008). Curve number hydrology: State of the practice. Maryland: *ASCE Publication.*

Hewlett, J. D. (1971). Comments on the catchment experiment to determine vegetal effects on water yield. *Water Resources Bulletin* 7(2): 376–381.

Hewlett, J. D. (1982). *Principles of forest hydrology*. Georgia: University of Georgia Press.

Hong, Y. and Adler, R. F. (2008). Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *International Journal of Remote Sensing* 29(2): 471–477. doi:10.1080/01431160701264292

Hubbart, J. (2008). History of hydrology. Retrieved from http://www.eoearth.org/view/article/153525/

Hunter, G. J. and Goodchild, M. (1993). Managing uncertainty in spatial databases: Putting theory into practice. *Journal of the Urban and Regional Information systems association*, 55–62.

Isik, S., Kalin, L., Schoonover, J. E., Srivastava, P., and Graeme Lockaby, B. (2013). Modeling effects of changing land use/cover on daily streamflow: An artificial neural network and

curve number based hybrid approach. *Journal of Hydrology* 485(0): 103–112. doi:http://dx.doi.org/10.1016/j.jhydrol.2012.08.032

Jeon, J., Lim, K. J., and Engel, B. A. (2014). Regional calibration of SCS-CN L-THIA model: Application for ungauged basins. *Water* 6(5): 1339–1359.

Khanbilvardi, R., Lakhankar, T., Krakauer, N., Nazari, R., and Powell A. (2014). Remote sensing data and information for hydrological monitoring and modeling. In: Eslamian S., editor. Handbook of engineering Hydrology: Modeling, Climate Change, and Variability. CRC Press: 501-516.

King, K. W., Smiley, P. C., Baker, B. J., and Fausey, N. R. (2008). Validation of paired watersheds for assessing conservation practices in the upper Big Walnut Creek watershed, Ohio. *Journal of Soil and Water Conservation* 63(6): 380–395. doi:10.2489/jswc.63.6.380

Klemes, V. (1983). Conceptualization and scale in hydrology. *Journal of Hydrology* 65(1–3): 1–23.

Klepper, O. (1997). Multivariate aspects of model uncertainty analysis: Tools for sensitivity analysis and calibration. *Ecological Modelling* 101(1): 1–13.

Kousari, M. R., Malekinezhad, H., Ahani, H., and Zarch, M. A. A. (2010). Sensitivity analysis and impact quantification of the main factors affecting peak discharge in the SCS curve number method: An analysis of Iranian watersheds. *Quat. Int*. 226:66–74.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., . . . Folke, C. (2001). The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change* 11(4): 261–269.

Li, H. and J. Wu (2006). Uncertainty analysis in ecological studies. In J. Wu, K.B. Jones, H. Li, and O. L. Loucks (eds), Scaling and Uncertainty Analysis in Ecology: Methods and Applications: 43–64.

Li, Q., Cai, T., Yu, M., Lu, G., Xie, W., and Bai, X. (2013). Investigation into the impacts of land-use change on runoff generation characteristics in the Upper Huaihe River basin, China. *Journal of Hydrologic Engineering* 18(11): 1464–1470.

Liu, Y., and H. V. Gupta (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resour. Res.* 43 (7):W07401. doi:10.1029/2006WR005756.

Lu, Y. and Y. Wan (2013). PHA: A fast potential-based hierarchical agglomerative clustering method, *Pattern Recognition* 46:1227–1239.

Mao, D. and Cherkauer, K. A. (2009). Impacts of land-use change on hydrologic responses in the Great Lakes region, *J. Hydrol*. 374(1–2): 71–82.

Melone, F., Barbetta, S., Diomede, T., Peruccacci, S., Rossi, M., Tessarolo, A., et al. (2005). Review and selection of hydrological models – integration of hydrological models and meteorological inputs. Contract No 12. Gennaio, Italy.

Maselli F., S. Romanelli, L. Bottai, and G. Zipoli. (2003). Use of NOAA-AVHRR NDVI images for the estimation of dynamic fire risk in Mediterranean areas. *Remote Sensing of Environment* 86:187–197.

Merz, R., G. Blöschl, and J. Parajka (2006). Spatio-temporal variability of event runoff coefficients, *J. Hydrol.* 331:591–604. doi:10.1016/ j.jhydrol.2006.06.008

Millennium Ecosystems Assessment (2005). Ecosystems and human well-being Island Press Washington, D.C.

Martinez, W. L. and A. R. Martinez (2008). *Computational Statistics Handbook with MATLAB*, (2nd ed.).New York: Chapman and Hall.

Martinez, W. L. and A. R. Martinez (2004). *Exploratory Data analysis with MATLAB: Computer Science and Data Analysis Series*. New York: Chapman and Hall.

Mishra, S. K. and V. P. Singh (2003). SCS-CN method. In Mishra, S. K. and V. P. Singh (eds), Soil conservation Service Curve Number (SCS-CN) Methodology (84–146), Water Science and Technology Library, 42.

Mitas, L. and Mitasoca, H. (1998). Distributed soil erosion simulation for effective erosion prevention. *Water Resources Research* 34(3): 505–516.

McMaster, R. B. and Usery, E. L. (2004). *A research agenda for geographic information science.* Florida: CRC Press.

Moriasi, D., J. Arnold, M. Van Liew, R. Bingner, R. Harmel, and T. Veith (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50(3): 885–900.

Mul, M. L. (2009). Understanding hydrological processes in an ungauged catchment in sub-saharan Africa UNESCO-IHE, Institute for Water Education. Ph.D. Dissertation. Delft University of Technology, Delft, Netherlands. Retrieved from http://www.samsamwater.com/library/Understanding_hydrological_processes_in_an_ungauged_catchment.pdf

Narasimhan, B., R. Srinivasan, J. G. Arnold, and Di Luzio. M. (2005). Estimation of long-term soil moisture using a distributed parameter hydrologic model and verification using remotely sensed Data. Trans. ASABE 48(3): 1101−1113.

"NASA LP DAAC" (2013). MODIS Land Products Quality Assurance Tutorial: Part 2, How to interpret and use MODIS QA information in the Vegetation Indices product suite https://lpdaac.usgs.gov/sites/default/files/public/modis/docs/MODIS_LP_QA_Tutorial-2.pdf

Palmer, T. N. (2000). Predicting uncertainty in forecasts of weather and climate. *Reports on Progress in Physics* 63(2): 71.

Parajuli, P. B. and Ouyang, Y. (2013). Watershed-scale hydrological modeling methods and applications. INTECH Open Access Publisher.

Petersen, A. C. (2000). Philosophy of climate science. *Bulletin of the American Meteorological Society* 81(2): 265–271.

Ponce, V. M. (1989). *Engineering Hydrology: Principles and Practices*. New Jersey: Prentice Hall.

Ponce, V. M. and Hawkins, R. H. (1996). Runoff curve number: Has it reached maturity? *Journal of Hydrologic Engineering* 1(1): 11–19.

Prokopy, L., Z. A. Goecmen, J. Gao, S. Allred, and J. Bonnell (2011). Incorporating Social Context Variables Into Paired Watershed Designs to Test Nonpoint Source Program Effectiveness, *J. Am. Water Resour. Assoc*. 47:196–202.

Renschler, C. S. and J. Harbor (2002). Soil erosion assessment tools from point to regional scales—the role of geomorphologists in land management research and implementation. *Geomorphology* 47:189–209.

Ricker, M. C., B. K. Odhiambo, and J. M. Church (2008). Spatial analysis of soil erosion and sediment fluxes: a paired watershed study of two Rappahannock River tributaries, Stafford County, Virginia, *Environ. Manage*. 41:66–778.

Schilling, K. E., Jha, M. K., Zhang, Y. K., Gassman, P. W., and Wolter, C. F (2008). Impact of land use and land cover change on the water balance of a large agricultural watershed: Historical effects and future directions, *Water Resour. Res*. 44:W00A09. doi:10.1029/2007WR006644

Schilling, K.E. and Spooner, J. (2006). Effects of watershed-scale land use change on stream nitrate concentrations. *Journal of Environmental Quality* 35:2132–2145.

Schmugge, T. J., Kustas, W. P., Ritchie, J. C., Jackson, T. J., and Rango, A. (2002). Remote sensing in hydrology. *Advances in Water Resources* 25(8): 1367–1385.

Schroeder, B. (2006). Pattern, process, and function in landscape ecology and catchment hydrology - how can quantitative landscape ecology support predictions in ungauged basins? *Hydrology and earth system sciences* 10(6): 967–979.

Singh, V. P. (1995). Watershed modeling. In computer models of watershed hydrology, V. P. Singh, ed., Water Resources Publications, Littleton, CO, 1–22.

Singh, W., Hjorleifsson, E. and G. Stefansson. (2011). Robustness of fish assemblages derived from three hierarchical agglomerative clustering algorithms performed on Icelandic groundfish survey data. *ICES Journal of Marine Science* 68:189–200.

Singh, R. P., Roy, S., and Kogan, F. (2003). Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India. *International Journal of Remote Sensing* 24:4393–4402.

Skaugest, A. E., J. Li, K. Cromack, R. E. Gresswell, and M. Adams. (2005). Hinkle. Creek Paired Watershed Study. Fish and Wildlife Habitat in Managed Forests Research Program. Progress Reports, FY 2005.

Slonecker, E. T., Barnes, C., Karstensen, K., Milheim, L. E., and Roig-Silva, C. M. (2013). Consequences of Land use and Land Cover Change: U.S. Geological Survey Fact Sheet 2013–3010. Retrieved from http://pubs.usgs.gov/fs/2013/3010/

Solomatine, D. P. and Dulal, K. N. (2003). Model trees as an alternative to neural networks in rainfall—runoff modelling. *Hydrological Sciences Journal* 48(3): 399–411.

Spooner, J., R. P. Maas, S. A. Dressing, M. D. Smolen, and F. J. Humenik. (1985). Appropriate designs for documenting water quality improvements from agricultural NPS control programs. *In* Perspectives on nonpoint source pollution: Proceedings of a national conference, Kansas City, Missouri, May 19–22, 1985. US Environmental Protection Agency, Office of Water Regulations and Standards, Washington, D.C.

Srinivasan, R. (2005). *Advanced SWAT BASINS training manual*, Texas A&M University, Blackland Research Center.

Sui, D. Z. and Maggio, R. C. (1999). Integrating GIS with hydrological modeling: practices, problems, and prospects. *Computers, environment and urban systems* 23(1): 33–51.

Tait, A. and Zheng, X. G. (2003). Mapping frost occurrence using satellite data. *Journal of American Meteorological Society*, 42:193–203.

Tang, Q. H., Gao, H. L., Lu, H., and Lettenmaier, D. P. (2009). Remote sensing: hydrology, Progr. Phys. Geogr. 33:490–509.

Todini, E. (2007). Hydrological catchment modelling: Past, present and future. *Hydrology and Earth System Sciences* 11(1): 468–482.

Tu, J. (2009). Combined impact of climate and land use changes on streamflow and water quality in eastern Massachusetts, USA, *J. Hydrol*. 379:268–283.

Tuppad, P. (2006). Hydrologic modeling response to NEXRAD and raingage spatial variability and strategic watershed management. . Dissertation, Doctor of Philosophy, Kansas State University.

US EPA Ecosystems Research Athens GA. (2013). Watershed models. Retrieved from http://www.epa.gov/athens/wwqtsc/html/watershed_models.html

Van Mullem, J. A. (1989). "Runoff and peak discharges using Green-Ampt infiltration model." *J. Hydr. Engrg.* ASCE 117(3): 354–370.

Varshavsky, R., Horn, D., and M. Linial. (2008). Global considerations in hierarchical clustering reveal meaningful patterns in data. *PLoS ONE* .Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0002247.

Veum, K. S., K. W. Goyne, P. P. Motavalli, and R. P. Udawatta (2009). Runoff and dissolved organic carbon loss from a paired-watershed study of three adjacent agricultural Watersheds, Agric., *Ecosyst. Environ*. 130:115–122.

Wagener, T. and McIntyre, N. (2005). Identification of rainfall–runoff models for operational applications. *Hydrological Sciences* 50(5): 735–751.

Wang Q., M. Watanabe, S. Hayashi, and S. Murakami (2003). Using NOAAAVHRR data to assess flood damage in China. *Environ. Monit. Assess*. 82:119–148.

Zacharias, I., E. Dimitriou, and Th. Koussouris (2004). Quantifying Land-Use Alterations and Associated Hydrologic Impacts at a Wetland Area by Using Remote Sensing and Modeling Techniques, *Environmental modeling & assessment* 9:23–32.

Zhang, X., Friedl, M. A., Schaaf, C.B., Strahler, A.H., Hodges, J. C. F., Gao, F., Reed, B. C., and Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment.* 84:471–475.

Zhou, M. C., Ishidaira, H., Hapuarachchi, H. P., Magome, J., Kiem, A. S., and Takeuchi, K. (2006). Estimating potential evapotranspiration using Shuttleworth-Wallace model and NOAA-AVHRR NDVI data to feed a distributed hydrological model over the Mekong River basin. *Journal of Hydrology* (1–2):151–173.

Zimmermann, B., Elsenbeer, H., and De Moraes, J. M. (2006). The influence of land-use changes on soil hydraulic properties: Implications for runoff generation. *Forest Ecology and Management* 222(1): 29–38.

Yu, Y., Yoon, S. O., Poulogiannis, G., Yang, Q., Ma, X. M., and J. Villén. (2011). Quantitative phosphoproteomic analysis identifies the adaptor protein Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. *Science*. 332:1322–1326.

Chapter 3 - Curve Number Development using Normalized Difference Vegetation Index

Abstract

The Curve Number (CN) is a widely applied method to estimate runoff from rainfall events. It has been adapted to various areas with different land use/cover and climatic conditions in many parts of the world, and successfully applied to situations ranging from simple runoff calculations and land use change assessment to comprehensive hydrologic/water quality simulations. However, the SCS-CN does not consider seasonal or dynamic land-use changes. This study used regression analysis to develop an NDVI-based CN (CN_{NDVI}) using Moderate Resolution Imaging Spectroradiometer – Normalized Difference Vegetation Index (MODIS-NDVI (MOD13Q1). The MODIS-NDVI was derived for every 16 days at 231 meter spatial resolution with NAD83/ UTM Zone 14 projection. Rainfall and runoff data collected from 2001 to 2012 from four small watersheds in Konza Prairie Long-Term Ecological Research, Kansas were used to develop, calibrate, and validate the model. A total of 398 and 201 data points were used for calibration and validation respectively. Results showed that the flow based on the calibrated model performed significantly better than the standard CN (SCS-CN) flow. The pairwise comparison of the calibrated flow and observed flow did not show a statistical difference (*p-Value* = 0.6622); however the standard CN based flow showed statistical differences with both observed flow and calibrated flow (*p-Value* < 0.0001). Calibrated flow increased by 0.91 for every unit increase in observed flow, while the standard CN based flow increased by 0.506 for every unit increase in observed flow. The calibrated flow was highly correlated to the observed flow (r = 0.83) with the standard CN was less correlated to measured flow (r = 0.404). The validated flow was better correlated to the measured flow than the

literature CN based flow (for validated model *r* ranges from 0.52 to 0.63 for different options, while *r* = 0.44 for literature CN), indicating that the calibrated improved the prediction of runoff compared to the existing method. These findings suggest that the CN_{NDVI} could be used to provide better estimates of surface runoff from the precipitation events in order that more timely land management decisions can be made.

3.1 Introduction

Complex interactions of biotic and abiotic processes at various scales result in spatiotemporally heterogeneous and dynamic landscapes with multiple ecological processes (Schroeder, 2006). Geological characteristics that affect the location, movement, and chemistry of water are extremely complex, but water movement in seemingly diverse landscapes demonstrates commonalities that allow the use of models (Wolock et al., 2004). Landscape heterogeneity results in spatiotemporal variability of hydrologic conditions and scale-dependent flow and transport properties both at field and catchment-scale hydrologic responses (Troch et al., 2009). Landscape characteristic changes due to natural and/or anthropogenic impacts determine water quantity and quality parameters in watersheds and stream channels. However, hydrologic responses due to landscape characteristic interactions are not the same as responses produced by the same variables without interaction as variability increases (Mohamoud, 2004).

Quantifying water movement and available water in a landscape aids drought and flood occurrence estimations, and understanding water quantity and quality of a landscape is a crucial component of ecosystem functions to assess the productivity, abundance, nutrient cycling, and ecological response of a landscape to natural and/or other disturbances. Stream flow trends and vegetation-hydrology interactions have been used to understand physical, hydrologic, and ecological factors and their interactions relative to channel evolution (Lenhart et al., 2013).

Long-term monitoring of stream discharge and sediment and the estimation of those parameters using various hydrologic models increase understanding of the overall hydrologic system.

Hydrologic models are necessary tools for estimating water quantity and quality parameters, resulting in better environmental resources management (Zhang and Savenije, 2005). Hydrologic models have been widely used to predict complex behaviors associated with environmental systems management and land use/cover changes across the landscape (Isik et al., 2013). Watershed modeling is a rational, economical, and useful approach for pertinent water resource management and decision making (Jeon et al., 2014). The SCS-CN method is a commonly used empirical method to simulate surface runoff of a rainfall event at the watershed scale, thereby representing combined hydrologic effects of soil, land use, land management, hydrologic condition, and antecedent soil moisture. CN-based runoff estimation predicts direct runoff from individual rainfall events. Runoff estimates based on CN are used to determine soil erosion, pollutant loading, and the amount of nutrients and pesticides transported in water systems from the landscape (Garen and Moore, 2005).

3.1.1 Background of SCS-CN Method

The Curve Number (CN) method, also referred to as the SCS-CN method, was developed in the 1950s by USDA, formerly Soil Conservation Service (SCS) and currently Natural Resources Conservation Service (NRCS), to meet the needs of the agency in implementing Public Law 566 (Hawkins et al., 2008; Hawkins, 2014). Public law 566, a Watershed Protection and Flood Prevention Act enacted on August 4, 1954 (Woodward et al., 2002), provided for upstream flood prevention and watershed condition improvement. It primarily targeted flood prevention and alleviating sediment problems at the small watershed scale. The CN method is relatively simple and widely used to calculate event rainfall-runoff volume relationships

(Hawkins, 2014). Through time, a good deal has been learned about the CN method; new applications and developments have emerged; and also insights to general rainfall-runoff hydrology have been gained through its application (Hawkins et al., 2008).

The CN is determined based on land use, hydrologic soil group (HSG), land treatments or practices, and the hydrologic condition of the area which accounts for a combination of factors that affect infiltration and runoff. It has been adapted to various areas with different land use/cover and climatic conditions in many parts of the world and successfully applied to situations ranging from simple runoff calculations and land use change assessment to comprehensive hydrologic/water quality simulations (Auerswald and Haider, 1996; Garen and Moore, 2005; Ponce and Hawkins, 1996). It is a widely used means of estimating storm runoff from rainfall events for designing water management structures. Its simplicity, predictability, stability, reliance on only one variable (CN) to predict the rainfall-runoff relationship, and responsiveness to major runoff producing watershed properties (soil type, land use, surface condition and antecedent moisture condition) are among the perceived advantages of CN method (Ponce and Hawkins, 1996).

Although the method was designed for a single storm event and originated as an empirical, event-based procedure for flood hydrology, the CN method has been adapted and used in various hydrologic models to simulate runoff behavior of ordinary and large rainfall events and daily time series (Garen and Moore, 2005; "*SCS Curve Number Method*", 2015). The CN method is widely used for rainfall-runoff modeling in continuous hydrologic simulation models (Kannan et al., 2007) and is the foundation of hydrology algorithms in most simulation models developed by the USDA for hydrology, soil erosion, and nonpoint source water quality (Garen and Moore, 2005).

Although the CN model is widely used and applicable, some watersheds were found to perform quite differently from basic CN runoff response patterns, leading to significant differences between the model and reality (Hawkins, 2014). For example, several studies (Baltas et al., 2007; Hjelmfelt, 1991; Soulis et al., 2009; Yuan et al., 2014) found that estimated runoff was inaccurate in semiarid watersheds in southeastern Arizona due to high retention capacity of the soil. Garen and Moore (2005) stated that the CN method is beset with many problems, issues, and misinterpretations that weaken its use for accurately estimating the amounts, paths, and source areas upon which erosion and water quality predictions depend. Many unresolved issues in the empirical CN-based rainfall-runoff relationship occur because the method is simplified but the relationship is dependent upon multiple watershed-related factors, such as land use/cover and soil moisture at the time of the rainfall event. The rainfall-runoff relationship is also highly affected by rainfall volume, intensity, and frequency, which are not accurately accounted for in CN determination; however, the 5 days prior rainfall in the form of antecedent moisture condition is to some extent incorporated. Seasonal changes in rainfall-runoff relationship must be accounted for in order to accurately estimate runoff for water management planning and decision making.

3.1.2 Rationale

The CN is traditionally determined from land-use, HSG, hydrologic condition, and antecedent soil moisture condition of the watershed using SCS lookup tables (Figure 3-1). Since a watershed is usually a combination of different land-use/cover and soil conditions, an area weighted average curve number for the entire watershed is often calculated. However, the standard CN approach does not reflect the temporal variability of hydrologic conditions and cannot capture seasonal or dynamic land-use changes, which are essential for reflecting

seasonality of hydrologic conditions, particularly in vegetated systems such as rangeland or prairie grasses that do not demonstrate significant variability in the lookup tables. The incorporation of dynamic land-use changes is essential in any hydrologic process study because they affect the water cycle, especially the ET and vegetation's ability to intercept precipitation. In addition, change in land use modifies the pathway and surface roughness, which then affects the timing of runoff. All these dynamic land use effects ultimately alter the runoff.

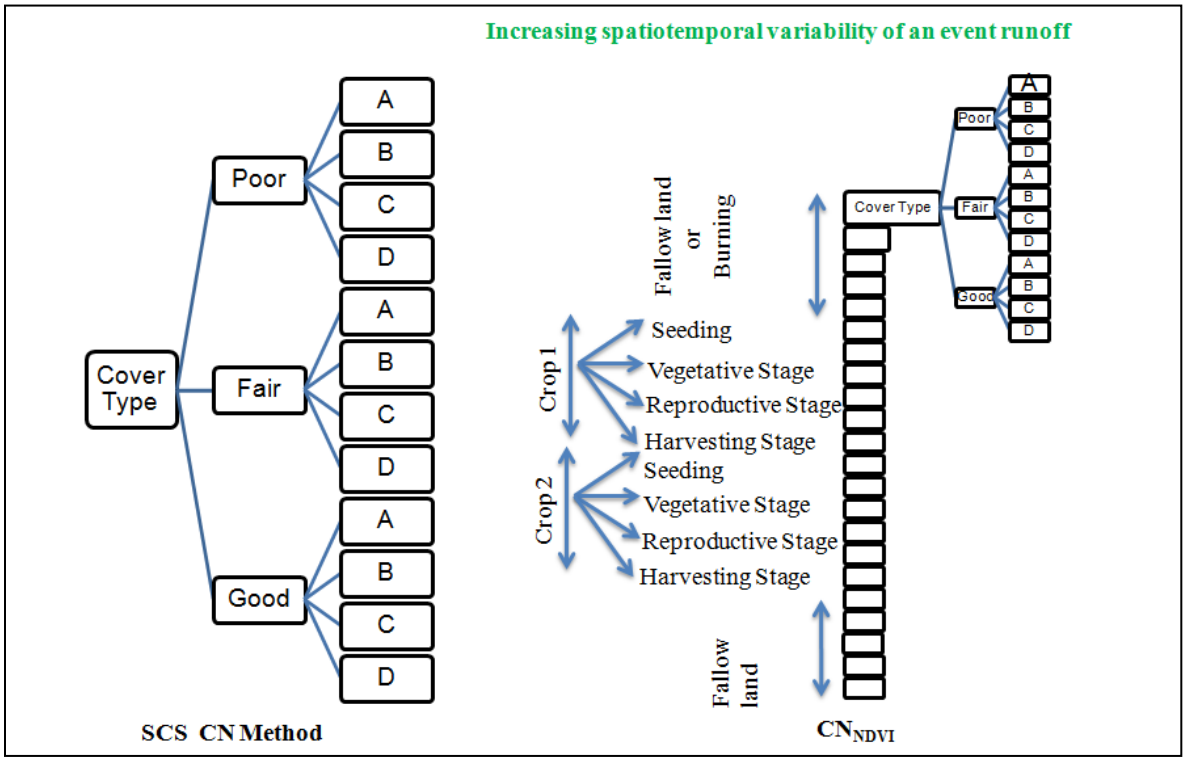


Figure 3-1: Schematic comparison of SCS-CN and CN_{NDVI} depicting spatiotemporal changes of event runoff

The objective of developing CN_{NDVI} was to capture the spatiotemporal variability of runoff from each rainfall event. This study developed a method to derive CN_{NDVI} using satellite data in order to account the seasonal effect of land use/cover changes and capture the temporal variability of hydrologic conditions. Figure 3-1 compares SCS-CN and CN_{NDVI}, demonstrating how CN_{NDVI} can account for spatiotemporal variability of annual land cover in order to capture

the spatiotemporal variability of the rainfall-runoff relationship. The specific objective of this study was to develop a model that estimates CN in order to accurately capture the spatiotemporal relationship of rainfall and runoff based on NDVI as a surrogate of spatial and temporal changes of hydrologic conditions.

3.2 Study Area

This study utilized 12 years of rainfall and runoff data from four small watersheds in northeast Kansas on the Konza Prairie Long-Term Ecological Research (LTER) site (Figure 3-2 and Table 3-1). Konza Prairie LTER is a comprehensive ecological research, education, and outreach program centered in one of the most productive grasslands in North America, the tallgrass prairie (Macpherson, 1996). The Konza Prairie provides an array of burning and grazing (especially bison) treatments to facilitate research and evaluate the effects of fire and grazing on plant composition, primary production, consumer density and diversity, nutrient dynamics, soil chemistry, and hydrology (Konza Prairie LTER Data Catalog, 2015). The Konza Prairie experiences a temperate midcontinental climate, with annual temperatures ranging from an average low of -3 ^{o}C (-9–3 ^{o}C) in January to an average high of 27 ^{o}C (20–33 ^{o}C) in July. Annual average precipitation is 835 mm, 75% of which occurs during the growing season, April through October (Konza Prairie LTER Data Catalog, 2015).

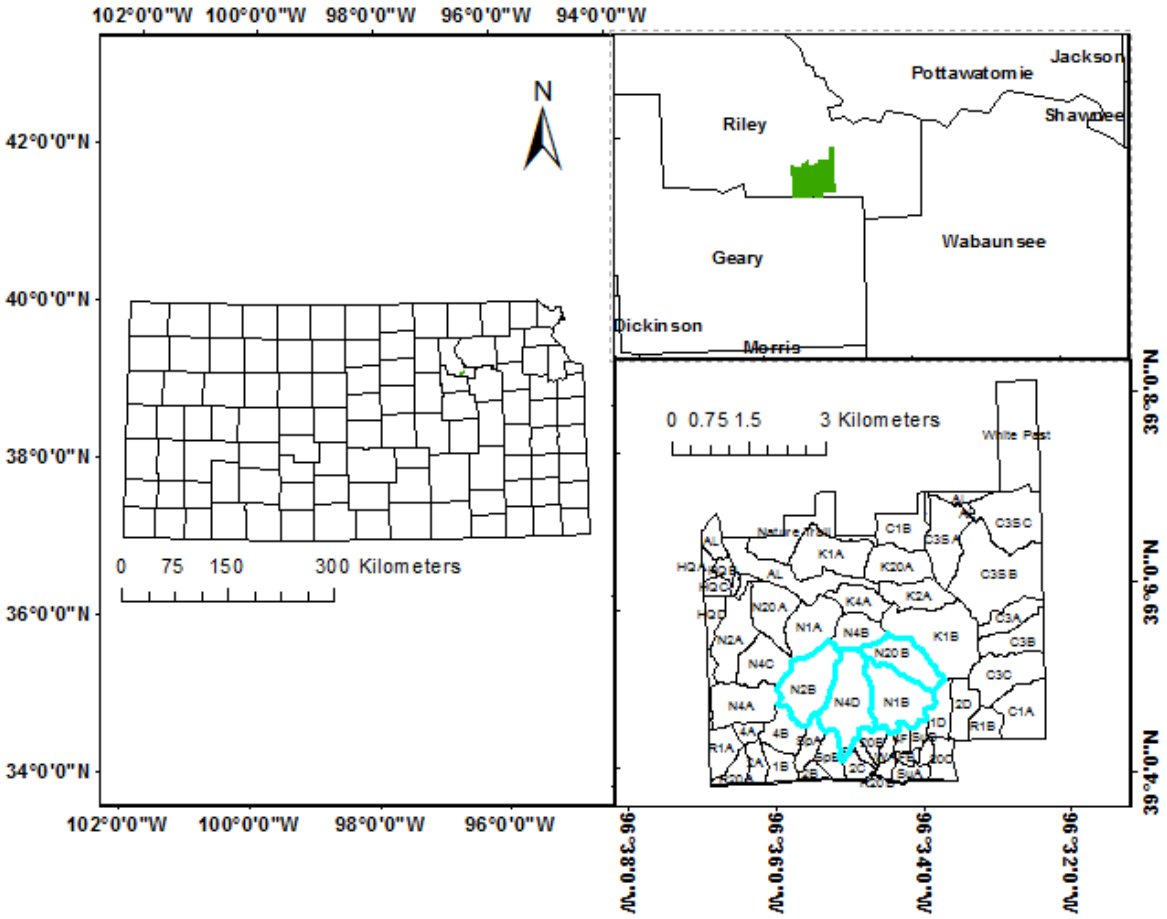


Figure 3-2: Konza Prairie Biological Station and watersheds used (aqua color) in this study

Konza watersheds are categorized based on grazing and burning treatments. Watersheds used in this study were N1B, N4D, N2B, and N20B, where N signifies grazing by native herbivores, the number indicates the years between burnings in spring, and B and D indicate replicates. Table 3-1 shows the watershed area, percentage of land cover, and HSG of the four small watersheds.

Table 3-1: Konza Prairie Biological Station, KS, study watersheds' land cover and hydrologic soil group

Watershed	Area (km^2)	Hydrologic Soil Group (%)			Land Cover (%)		
		B	C	D	Grassland	Forest	Other
N1B	1.2	8.6	90.3	1.1	96.0	3.7	0.2
N2B	1.2	8.6	89.9	1.5	88.5	11.3	0.2

N4D	1.4	7.0	92.0	1.0	95.3	4.3	0.4
N20B	0.8	7.0	91.2	1.8	95.6	3.8	0.5

3.3 Model Development

The CN_{NDVI} model was developed using regression analysis between the back-calculated observed CN and NDVI developed using remotely sensed data. In this study NDVI was assumed to represent land use/cover changes and hydrologic conditions of the watershed for predicting the rainfall-runoff relationship. Because NDVI is a measure of vegetation health, a negative relationship between the CN and NDVI was expected due to the fact that high CNs indicate poor hydrologic conditions. Increases in NDVI suggest healthier vegetation and improved hydrologic condition, thereby indicating better infiltration and lower CNs and vice versa (Figure 3-3). The model was developed based on the assumption that the 16-day NDVI interval reflected hydrologic conditions of the landscape better than the SCS-CN method.

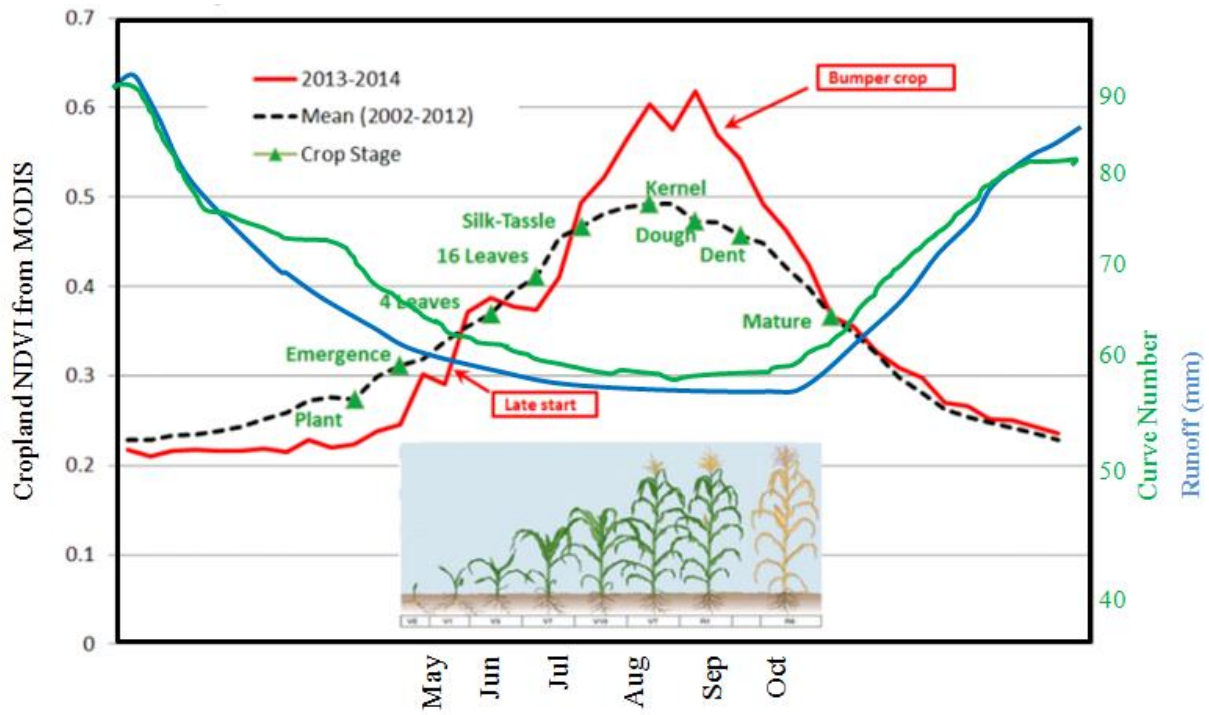


Figure 3-3: Graphical representation of expected relationship between vegetation phenology, change in NDVI, and resulting CN and runoff (adopted and modified from Reynolds, 2014)

3.3.1 Observed CN: Back-Calculated Curve Number

Twelve years of daily precipitation and runoff data from four small Konza watersheds were used to back-calculate observed CNs based on the SCS-CN method (Table 3-2). Rain gauges in the watersheds operated from April 1 to October 31. Prior to 2010, rainfall data was measured using Belfort Weighing rain gauges at each study watershed (LTER Methods Manual). Starting in March 2010, precipitation at Konza headquarters, approximately 3 km from the watersheds, was measured using an OTT Pluvio rain gauge. Watershed discharge was measured at 5 minutes intervals using triangular-throated flumes with pressure transducers at the base of each catchment and processed into a daily volume (from midnight to midnight). The volume was converted to depth of runoff based on each watershed area, and rainfall and runoff were reported in mm/day. Since the watersheds were small, "no lag time" was assumed between rainfall and runoff at the watershed outlets.

Table 3-2: Rainfall and runoff data used in each study year watershed at Konza Prairie Biological Station, KS

Watershed	Year
N1B	2001–2006 and 2008–2012
N2B	2001–2012
N4D	2001–2007 and 2012
N20	2002–2011

The CN was back-calculated based on the SCS-CN equation (Equation 3-1). Equation 3-3 was derived from Equation 3-2 in order to calculate potential maximum retention (S) from observed rainfall and runoff data (Hawkins, 1979; Soulis et.al., 2009). The National Engineering Handbook, Section 4 (NEH4) suggested the initial abstraction (I_a) in the SCS-CN equation (3-4) be 20% of the maximum retention (S) with limited explanation (Hawkins et al., 2008; USDA, SCS, 1986). In an extensive study using 97 small watersheds, Hawkins and Khojeini (2000)

found that the group median value for the ratio of I_a/S varied from 0 to 0.0966 for "ordered" data and was equal to 0 for all cases with "natural" data. However, this study used the original SCS I_a/S ratio value of 0.2 (USDA, Soil Conservation Service, 1954 et seq.) since the initial abstraction in the study area could be substantial because the landscape is dominated by dense grassland, thereby increasing the initial abstraction.

$$Q = \frac{(P-I_a)^2}{(P-I_a)+S} \quad (3\text{-}1)$$

$$Q = \frac{(P-(0.2S))^2}{P+(0.8S)} \quad (3\text{-}2)$$

$$S = 5\left(P + 2Q - \sqrt{4Q^2 + 5PQ}\right) + 254 \quad (3\text{-}3)$$

$$CN = \frac{25400}{S+254} \quad (3\text{-}4)$$

Where Q = runoff depth (mm),
P = precipitation depth (mm),
CN=Curve Number
I_a = 0.2*S, initial abstraction, and
S= potential maximum retention.

After calculating the CN, adjustments for antecedent moisture conditions (AMCs) were made based on SCS AMC criteria (Huffman et al., 2013; Hawkins et al., 2008; Hawkins and Cate, 1998). First, the back-calculated CN were calculated using AMC II conditions and each event CN was adjusted based on the criteria in Table 3.3, which lists AMC classifications based on rainfall amount (Huffman et al., 2013). The CN was back-calculated for all data, including zero value rainfall and runoff data, in order to implement the 5-day AMC criteria. Values were filtered to remove zero rainfall and runoff and negative potential maximum retention (*S*) values.

After filtering, 599 data points were used for model development, calibration, and validation of the CN_{NDVI} model.

Table 3-3: Classification of antecedent moisture conditions

AMC	Total 5-day Antecedent Rainfall (mm)	
	Dormant Season	Growing Season
I	<13	<36
II	13 - 28	36 - 53
III	>28	>53

Data-defined CN is dependent on rainfall depth in addition to landscape characteristics, and a distinct bias for high CN at lower rainfall depth was evident (Hawkins et al, 2008). In this study, the back-calculated CN response to rainfall depth was assessed; a majority of the data was associated with low rainfall depths, so the back-calculated CNs were in the high range. Figure 3-4 shows CN response to rainfall depth based on observed runoff used for model development. Assessing the dominating behavior of CN response to rainfall depth was crucial for determining whether the CN method was appropriate for watersheds in this study. In general, three dominating behaviors of CN response for rainfall were identified: standard (characterized by decreasing CN with increasing P but approaching a constant or near-stable value asymptotically at higher rainfalls), complacent (characterized by declining CN with increasing rainfall but not approaching a fixed equilibrium value), and violent (a pattern with complacent behavior with declining CN with increasing rainfall at low rainfalls but with sudden change to a much higher runoff response at some threshold elevated rainfall depth) (D'Asaro et.al, 2014; Hawkins et al, 2008; Hawkins, 1993). The standard behavior is typically most common, as demonstrated in a majority of agricultural, urban, and rangeland settings where rainfall excess occurs due to infiltration processes (Hawkins et al, 2008; Hawkins, 1993). The rainfall-runoff relationship of standard response behavior can be best explained by the CN method, as proven by the CN-

rainfall data relationship from this study (Figure 3-4). Figure 3-4 shows that higher CNs were calculated for low rainfall depths; therefore, since most rainfall depths were in the low range, rainfall was the main factor in determining CN.

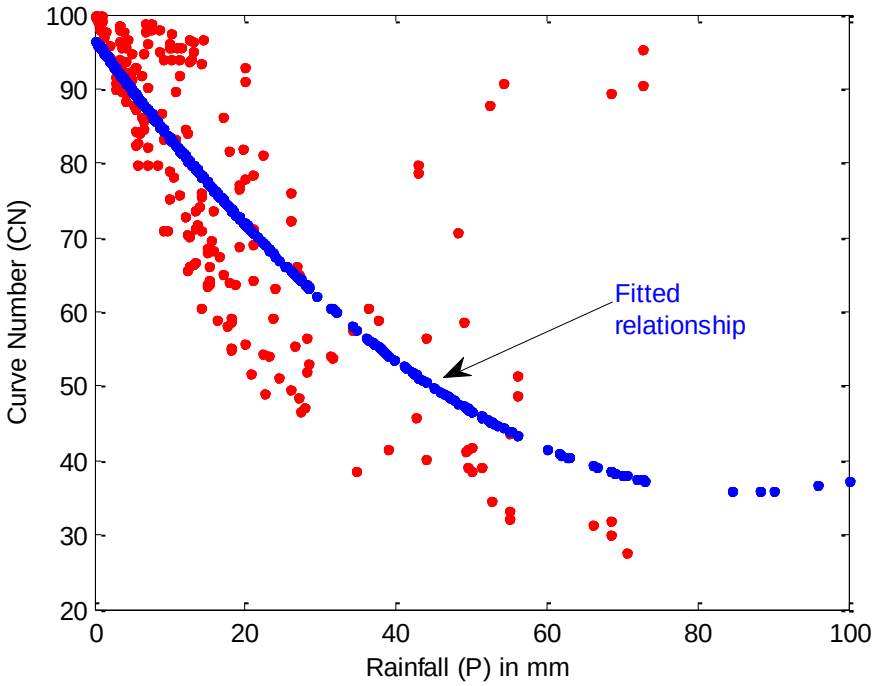


Figure 3-4: CN response to rainfall depth for observed records in Konza Prairie Biological Station, KS, used for model development

3.3.2 Land cover change: Normalized Difference Vegetation Index

Moderate Resolution Imaging Spectroradiometer – Normalized Difference Vegetation Index (MODIS-NDVI) 16-day composite grid data (MOD13Q1) was used to estimate CN based on 250 m, 16-day resolution NDVI with 231 m spatial resolution when re-projected from sinusoidal projection to UTM Zone 14 North using NAD 83 projection. MODIS-NDVI was derived from the reflectance of near-infrared (841–876 nm) and red (620–670 nm) bands (Equation 5).

$$NDVI = \frac{NIR - Red}{NIR + Red} \qquad (3\text{-}5)$$

NDVI is a VI used to monitor vegetation conditions and display land cover and its changes. It is useful for assessing vegetation health and allows fast visual examination of vegetation density and photosynthetically active regions (Huete et al., 1999). NDVI relies on the absorption of red radiation by chlorophyll and other leaf pigments and the strong scattering or reflection of near-infrared radiation by foliage (Beck et al., 2006). Seasonal variations of NDVI are closely related to vegetation phenology, such as green-up, peak, and offset of development (Beck et al., 2006; McCloy and Lucht, 2004). Most monitoring of large-scale vegetation activity is based on NDVI (Beck et al., 2006).

NDVI ranged from 0 to 255 (theoretical range) in this analysis, which is an 8-bit unsigned integer. The 8-bit unsigned integer format was used to capture CN variability with comparable NDVI and CN ranges. Rescaling NDVI from 0 to 255 (8-bit unsigned integer) resulted in values from -1 to 1 (floating point). Low NDVI values (approximately 0.1 and below) corresponded to barren areas of rock, sand, or snow. Values between 0.2 and 0.3 represented sparse vegetation such as shrub and grassland, while high values (approximately 0.6–0.9) indicated dense vegetation such as temperate and tropical rainforests (Jesslyn, 2015).

NDVI has been widely used in phenological studies (Boschetti et al., 2000; Colombo et al., 2011; Soudani et al., 2008; Hmimina et al., 2013) because it is more sensitive to small increases in the amount of photosynthetic vegetation (Soudani et al., 2006; Soudani et al., 2008; Sesnie et al., 2012) than other VIs. The link between NDVI and vegetation productivity has been well established theoretically and empirically and well documented: Studies have shown that NDVI differentiated vegetative systems at various scales worldwide (Soriano and Paruelo, 1992; Paruelo et al., 2001) and improved impact assessments of disturbances such as drought (Singh et al., 2003), fire (Maselli et al., 2003), flood (Wang et al., 2003), and frost (Tait and Zheng, 2003).

NDVI has also been successfully applied to study temporal and spatial trends and variation in vegetation distribution and productivity and dynamics and to monitor habitat degradation and fragmentation, as well as the ecological effects of climatic disasters such as drought or fire. Narasimhan et al. (2005) used NDVI as a pseudo indicator of soil moisture condition. Zhou et al. (2006) used NDVI-derived LAI to monitor potential ET. Global usage of NDVI suggests that any application can be easily adapted to other regions. Enhanced vegetation index (EVI) could be an alternative for NDVI; a comparison of MODIS NDVI and EVI to predict crop-related land use/cover produced similar output with equivalent accuracies (Wardlow and Egbert, 2010).

Based on rainfall and runoff data availability in the study watersheds, corresponding MODIS NDVI 16-day composite grid data (MOD13Q1) in Hierarchical Data Format (HDF) format were gathered from January 2001 through December 2012 (12 years) from the NASA Earth Observing System Data and Information System (EOSDIS) data gateway. Constrained view angle-maximum NDVI value (CV-MVC) compositing technique was used in MOD13Q1 to reduce atmospheric and cloud effects and constrain view angles (Huete et al., 2002; Hutchinson et al., 2015). These NDVI products were already calibrated and atmospherically corrected bi-directional surface reflectance that have been masked from cloud and cloud shadows, land/water, and aerosol products. NDVI data were extracted to the extent of the study area boundary and averaged for each period for the linear regression developed to estimate CN.

3.3.3 CN_{NDVI} Regression Model

Regression analysis, frequency analysis, and screening time series are the most common statistical methods of analyzing hydrologic data (Oosterbaan and Nijland, 1994). Regression analysis was developed to detect the presence of mathematical relationship between two or more variables (Helsel and Hirsch, 2002; Mendenhall and Sincich, 2012; Oosterbaan and Nijland,

1994). In this study regression analysis was used to mathematically model relationships between the response variable (CN_{NDVI}) and predictor variable (NDVI). The overall procedure to develop the NDVI-based curve model is described in Figure 3-5.

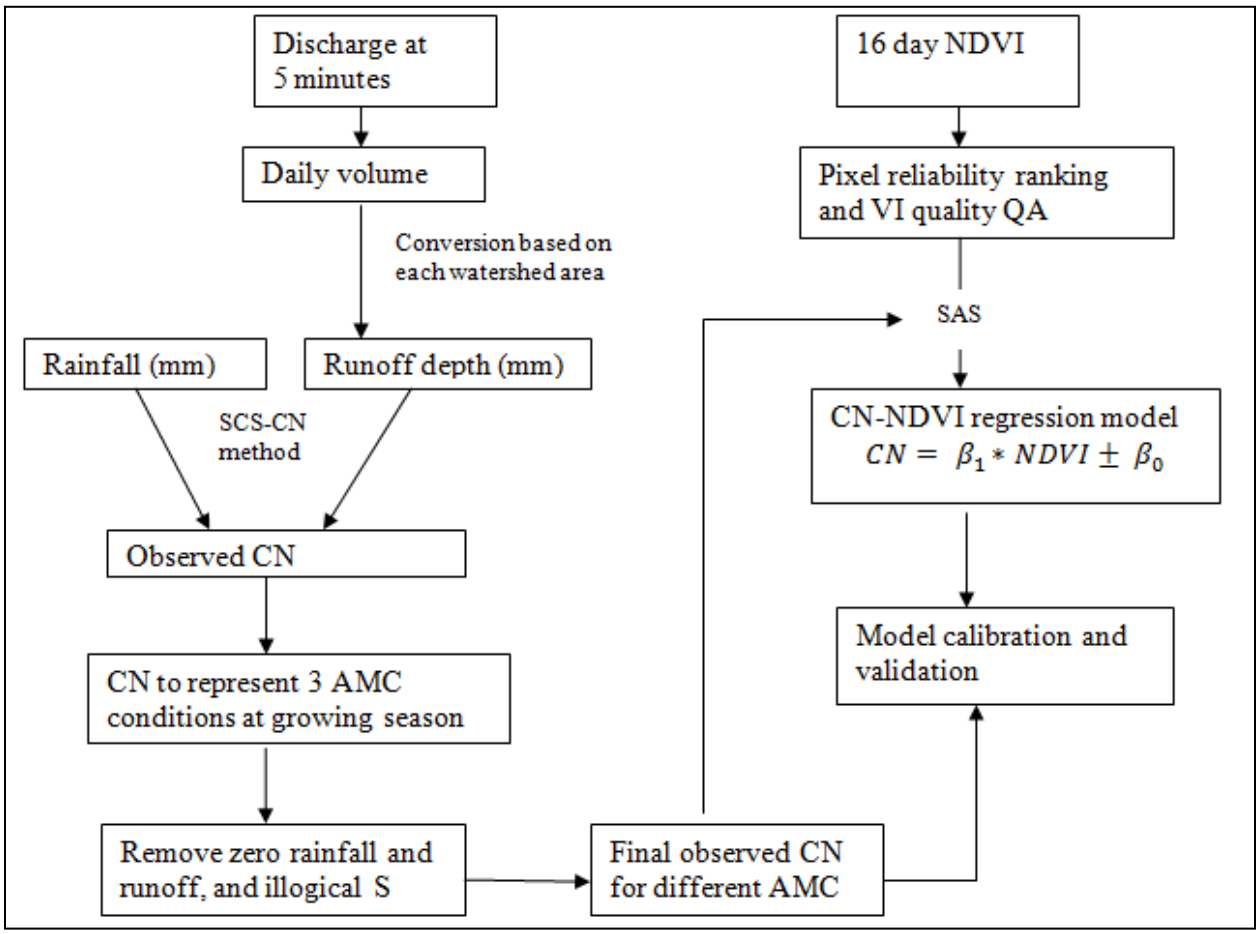


Figure 3-5: Schematic diagram of the overall method of NDVI based CN development using rainfall, runoff, and NDVI imagery

One of the problems in regression analysis is caused by the type of data for regression which can be either observational (where the independent variables are uncontrolled) or experimental (where the independent variables can be controlled) (Mendenhall and Sincich, 2012). Moreover, regression analysis is dependent upon the amount of information; the more data, the better the regression analysis is; however uncontrollable in the case of observational studies. Mendenhall and Sincich (2012) suggested that the number of data should be greater than

or equal to 10 times the number of β parameters as a good rule of thumb to ensure a sufficiently large sample. β parameters are coefficients of the independent variables and their interaction plus the *y*-intercept; so that we have two β parameters in the case of linear regression. This would suggest that a minimum of 20 data points are necessary for good model development. Therefore, more than 390 points were used for model development, calibration, and validation in this study.

The least squares estimation of parameters for the linear regression model was conducted using SAS 9.3 statistical software. The linear regression analysis fits the following linear regression model:

$$CN = \beta_1 * NDVI + \beta_0 \quad . (3\text{-}6)$$

Where: CN = Curve Number (response variable)

NDVI = Normalized Difference Vegetation Index (the predictor variable)

β_1 = the slope of the regression line which shows the decrease in the mean of CN of every 1-unit NDVI value increase.

β_0 = The y-intercept at which the CN value for zero NDVI.

Simple linear regression is an approach to model the relationship of one response variable (i.e., CN) using a single independent variable (i.e., NDVI), with the assumption that the two variables are linearly related and a single predictor observation provides a distinct criterion variable. Because one or more rainfall events could occur (as shown in Figure 3-6) in a 16-day NDVI period, resulting in several back-calculated CNs associated with a single NDVI value, all back-calculated CNs occurring within one 16-day NDVI window were averaged to create a unique CN for each NDVI period. Averaging back-calculated CN based on the same NDVI period reduced variation related to the model and provided normally distributed residuals (Figure 3-8 and Appendix A), thereby improving model performance.

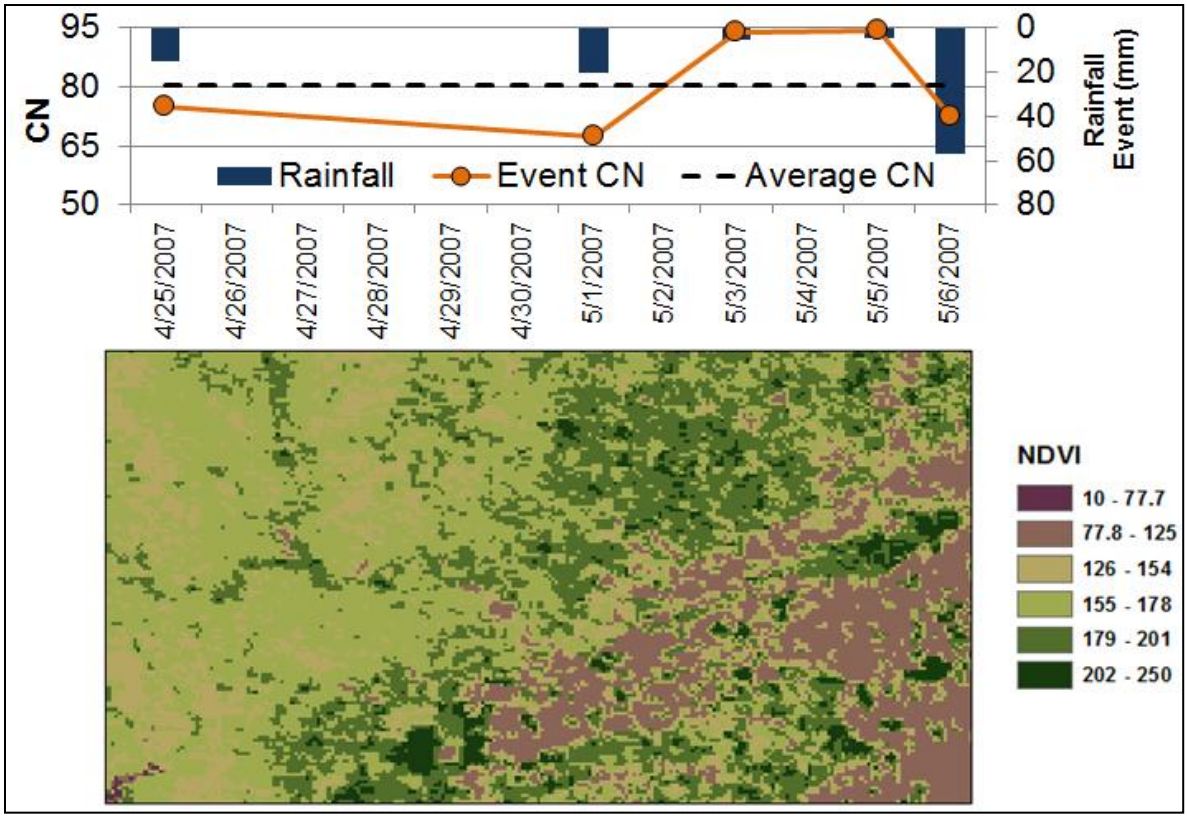


Figure 3-6: Diagram depicting the back-calculated CN versus NDVI relationship that shows presence of multiple events in each NDVI period

Residual analysis was conducted to assess whether the assumptions of the regression analysis were met. Those assumptions were that the error term was (roughly) normally and (approximately) independently distributed with a mean zero and constant and all pairs of error terms were uncorrelated (Mendenhall and Sincich, 2012; Engineering Statistics Handbook) and that the assumptions were interrelated. The normality assumption was checked, outliers were detected, and influential observations were identified based on standardized residual, studentized residual, and leverage. An observation was considered an outlier and removed when the standardized residual was larger than three times the standard deviation (in absolute value) and when the studentized residual was larger than 3 (in absolute value).

3.3.4 Model Calibration and Validation

As mentioned, the CN_{NDVI} model was developed by averaging the back-calculated CN with the same NDVI period if more than one event occurred in the same period in order to meet the requirements of valid regression analysis. Averaging was done to meet one-to-one relationships of linear regression in order to reduce residual errors and meet normally distributed error assumptions. However, calibration and validation processes were conducted using whole set of data without averaging the back-calculated CN and without removing extreme values in order to improve model efficiency. Two-thirds of the data was used for calibration, and the other one-third of data was used for validation. The data was divided based on the amount of precipitation for each event. As previously shown in Figure 3-4, the rainfall-runoff relationship (back-calculated CN) was dependent upon the amount of precipitation at each event, so the calibration and validation data were expected to have similar precipitation distribution. Data used for calibration represented 66% (398) of the complete data set, with 71% of low precipitation and 63% of high precipitation events. The validation data (201) accounted for 34% of the complete data, with 29% of low precipitation and 37% of high precipitation events. The assumption was made that the rainfall-runoff relationship was similar throughout the data period so that the calibration and validation data division could provide a better statistical analysis and inference. For calibration and validation, the CN was calculated from the NDVI based on the regression model developed. Figure 3-7 plots the data used for calibration and validation. Calibration was stopped when r and r^2 reached to satisfactory.

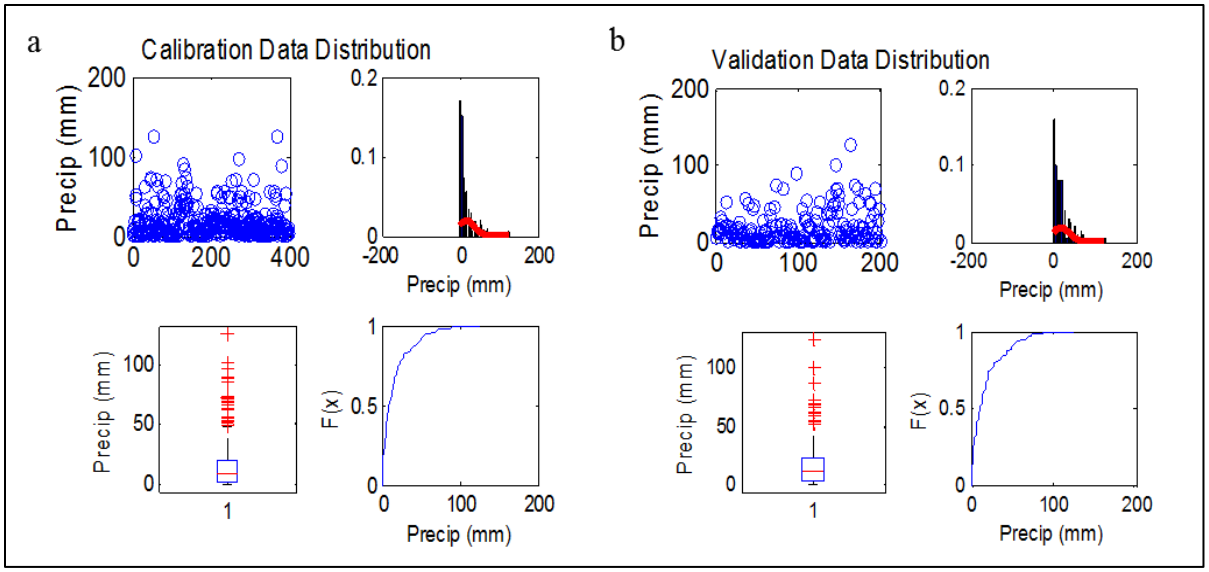


Figure 3-7: Konza watersheds precipitation data distribution for a) calibration and b) validation

Note: From top-left: scatter plot, histogram, boxplot, and quantile of residuals

Model calibration was conducted using MATLAB considering the relationship of CN and precipitation depth, runoff depth and precipitation depth, and CN and runoff depth comparisons of the observation and model. Coefficient of correlation (r) and coefficient of determination (r^2) and different plots were used in assessment. The coefficient of correlation measured the strength of the linear relationship between the predictor and the response variable. The coefficient of determination indicated how well the data fits a statistical model, allowing determination of prediction certainty using the model. Graphical plots were used to verify whether or not the model output imitated the observed distribution. Figure 3-10 and statistical parameters were used in the calibration process.

The model was calibrated in two phases. The first phase used the observed CN (back-calculated), rainfall with relatively high values but low runoff. The second phase used various combinations of rainfall, CN, and observed flow. After initial satisfactory calibration, the model

was adjusted to account for theoretical bounds of CN; the CN could not be greater than 98 according to literature values for the method.

Validation of the model occurred in three alternatives to highlight the importance of high rainfall events in the rainfall-runoff relationship. The first alternative used all data designated for validation. The second alternative removed the most extreme data point generated from the highest rainfall point that yielded almost no runoff. The third alternative removed two additional data points.

3.3.5 Vegetation Indices Quality Assessment

Because clouds can obstruct the quality of NDVI and other VI products and introduce noise into the data, Huete et al. (2002) and Hutchinson et al. (2015) developed the constrained view angle-maximum NDVI value (CV-MVC) compositing technique used in MOD13Q1 to reduce atmospheric and cloud effects and constrain view angles. However, persistent cloud cover was still a possibility during the 16-day study period. In order to assess the quality assurance of MOD13Q1 products, pixel reliability ranking and VI quality imagery (16-day, 250 m similar to other VIs) SDS were used in this study.

In this study, the pixel reliability ranking and VI quality pixel bit numbers were used as described in NASA LP DAAC (2013) guidance to monitor NDVI quality. The purpose of VI quality analysis is to find any cloud contamination. NDVI quality was assessed for the whole NDVI products covering all watersheds used to develop the CN_{NDVI} model. Even though, the pixel reliability rank and VI quality data were processed (Appendix Table A-1) for all used NDVI products; the pixel realization values were first used to determine the quality of NDVI as it is recommended in the guidance.

Pixel reliability is represented by numbers ranking from -1 to 4 that depict the pixel level data quality (Table 2-1). First, the percentages of the area with similar pixel reliability rank were estimated for the four small watersheds used to develop the model. The total area of four watersheds was considered as single study area as they are located next to each other (Figure 3-1). If the area covering all four watersheds contains pixel reliability not equal to 0 which means the data is good and use confidence, then the VI Quality bit numbers were considered for further analysis. During VI Quality analysis, the VI Quality values were changed to 16-bit numbers for appropriate interpretation. For interpretation, the binary bit-string were read from right to left and the individual bits within a bit-field were read from left to right as described in NASA LP DAAC (2013) (or in literature review of this dissertation[Chapter 2]). Detailed description on VI quality interpretation can be seen in NASA LP DAAC (2013). Pixel reliability rankings and VI quality values for study watersheds at the Konza Prairie Biological Station are included in Appendix A and Table A-3.

The VI quality QA values extracted for the entire Konza Prairie area, the data quality description based on NASA LP DAAC (2013), and the summary description are provided in Table 3-4. The percentage areas of each VI quality value for the study area are summarized inTable A-3 (Appendix A).

Table 3-4: VI quality Bit-No, parameter, bit-word, and their interpretation in Konza Prairie watersheds (Adopted and modified from NASA LP DAAC, 2013)

16-bit binary	Imagery value	**Data quality description based on NASA LP DAAC (2013) guidance**	**Summary description**
0|0|001|0|0|0|01|0001|00	2116	VI produced with good quality, low quality VI usefulness, low aerosol quantity, no adjacent cloud detected, no atmospheric BRDF correction, no mixed clouds, only land, no possible snow/ice, possible shadow	Good quality data and no noise*
0|0|001|0|0|0|01|0010|00	2120	VI produced with good quality, decreasing quality VI usefulness, low aerosol quantity, no adjacent cloud detected, no atmospheric BRDF correction, no mixed clouds, only land, no possible snow/ice, possible shadow	Good quality data with decreasing quality VI usefulness and no noise*
0|0|001|0|0|0|10|0010|01	2185	VI produced but check other QA, decreasing quality VI usefulness, intermediate aerosol quantity, no adjacent cloud detected, no atmospheric BRDF correction, no mixed clouds, only land, no possible snow/ice, possible shadow	Decreasing VI usefulness with intermediate aerosol and no clouds, snow or shadow

*Aerosols, adjacent and mixed clouds, possible snow/ice, and shadow were considered to be noise.

3.4 Results and Discussion

The CN_{NDVI} model proved to be a promising alternative to a standard CN method by providing a mechanism to address spatiotemporal variability of the rainfall-runoff relationship in the CN method. Results showed that the regression model based on the 16-day NDVI composite data provided a better prediction, even with no calibration, than the standard CN. The CN_{NDVI} method could be appealing to hydrologists because it uses only one variable, CN, to estimate runoff from rainfall events.

3.4.1 Vegetation Indices/Quality Assurance Assessment

Based on pixel reliability rank (Table A-1), all NDVI data used for model development were categorized as good data (confidence 92.74%) and marginal data (useful but look at other QA information, 7.26 %) for the four Konza watersheds (Table 3-5, Appendix A, Table A-3). Among marginal data, A majority of the data had a value of 2116 (binary value 0|0|001|0|0|0|01|0001|00, meaning that VI produced with good quality, no mixed clouds, only land, no possible snow/ice, but possible shadow) or 2120 (binary value 0|0|001|0|0|0|01|0010|00, meaning that VI produced with good quality, low aerosol quantity, no adjacent or mixed cloud, only land, no possible snow/ice, but possible shadow). Therefore, 7.26 % of marginal data (useful based on pixel reliability ranking) had no noise issues (no adjacent/mixed clouds, no possible snow and shadow, and no aerosols) based on VI quality QA. Based on these interpretations of values, no significant cloud contamination was present in all NDVI products used in model development.

Table 3-5: Pixel reliability rank percentage with good data (pixel reliability rank = 0, meaning the data can be used with confidence)

Pixel reliability rank	%
Percentage of NDVI temporal products with good data for all watersheds (i.e., 100% of area)	92.74
NDVI temporal products with <100 % and ≥90% of area with good data; the remaining is useful data but need to look at other VI quality QA	2.42
NDVI product with <90% of area with good data; the remaining is useful data but need to look at other VI quality QA	4.84

3.4.2 Model Output

Regression analysis mathematically modeled relationships between the CN (response variable) and NDVI (predictor variable). The CN_{NDVI} regression model is given in Equation 3-7 and Figure 3.8.

$$CN = -0.117 * NDVI + 97.3 \qquad (3\text{-}7)$$

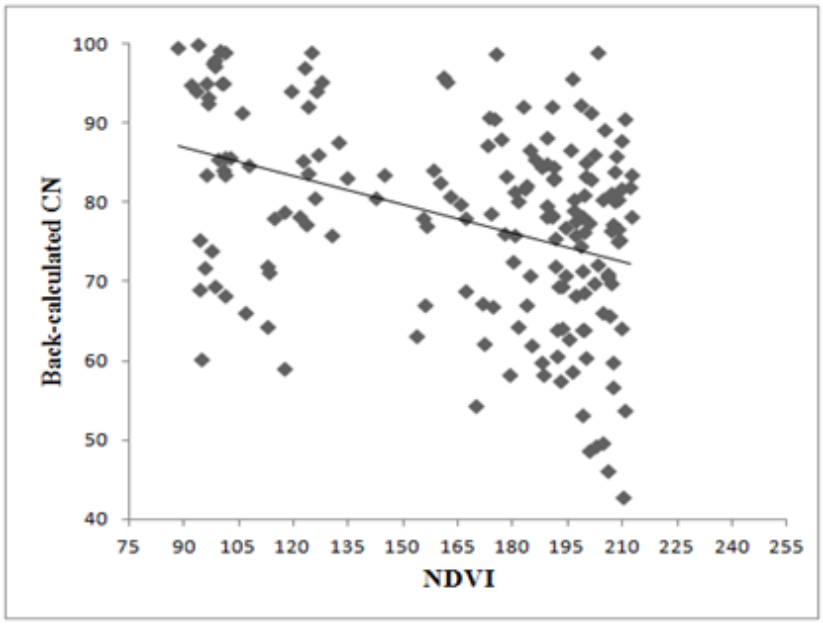


Figure 3-8: NDVI versus back-calculated CN relationship based on rainfall and runoff data and 16-day NDVI for Konza Prairie Biological Station, KS, study watersheds

Output of residual analysis showed that residual errors were normally distributed and the model met the assumptions of regression (Figure 3-9, Appendix A). The scatterplot, quantile

plot, and histogram of the residuals (Figure 3-9) showed the normality of residuals with zero mean value. Precipitation and runoff, which are observational data, are not normally distributed (usually skewed to the right or positive skewness), so the expectation of perfect normality in residual distribution of observational analysis, especially weather-related data such as rainfall and runoff, would be unreasonable because of significant extreme event occurrences. As described in the methods, regression analysis performs better in experimental studies because the independent variable is controlled. Nevertheless, analysis of this study showed that the residuals were normally distributed, so the diagnostics of the residual analysis outcome was sufficient for observational study.

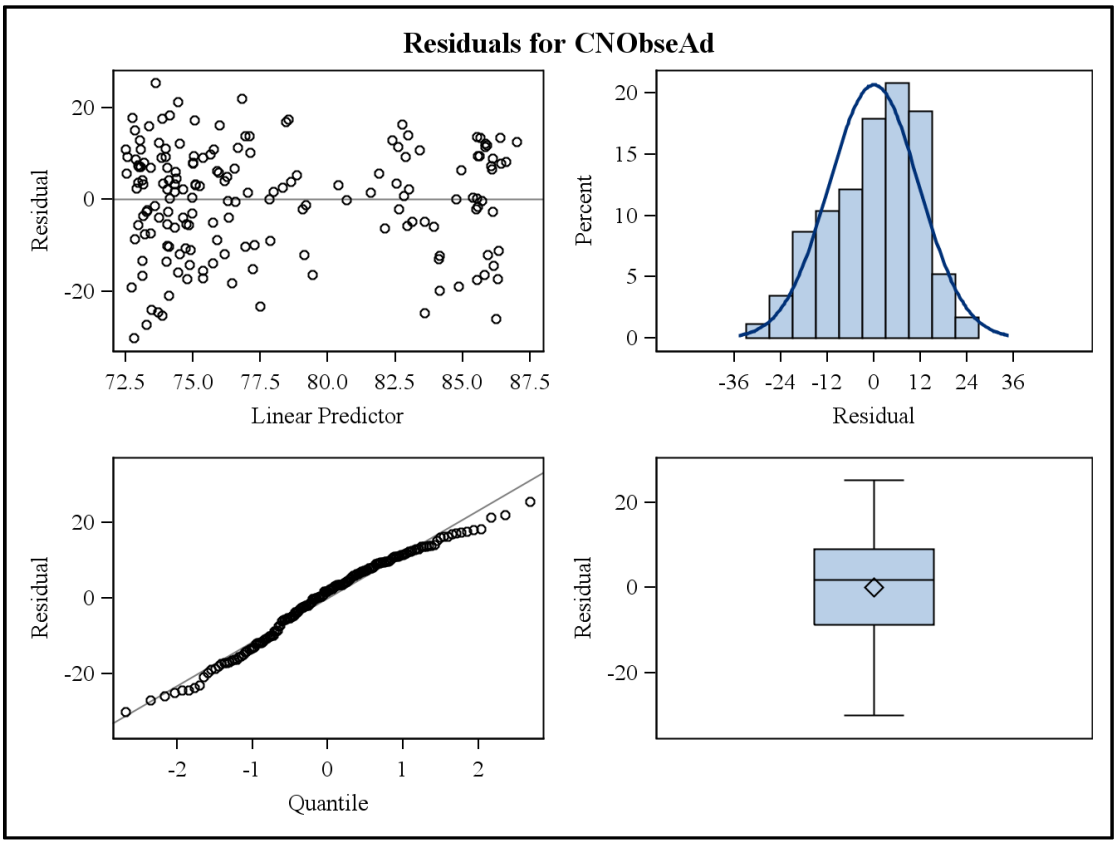


Figure 3-9: Residual analysis output of developing CN based on Normalized Difference Vegetation Index (NDVI)

Note: From top-left: the scatter plot, histogram, boxplot, and quantile of residuals

Statistical analysis of the model relating CN and NDVI is shown in Table 3-6. The *p*-value for the overall model was less than 0.05, proving that the model was statistically significant, with the coefficient of determination (r^2) of 0.1445. Many researchers prefer the *p*-value approach for statistical decision making (Mendenhall and Sincich, 2012). The *p*-value indicates the probability of making an incorrect decision, which was less than 0.01% according to output in the model for this study. The low r^2 indicates that the rainfall-runoff relationship is much more complicated than prediction using a single variable; however, as explained in Sections 3.4.3–3.4.6, it is more accurately determines the rainfall-runoff relationship than the existing SCS-CN method. As shown in Figure 3-11, the regression model performed better than the standard CN-based flow even without calibration, and improved after calibration.

Table 3-6: Regression analysis output of SAS to model CN using NDVI as predictor

Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	1	3903.13	3903.13	28.89	<.0001
Error	171	23101	135.09216		
Corrected Total	172	27004			

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	97.295	3.72116	26.15	<.0001
NDVI	1	-0.117	0.02171	-5.38	<.0001

3.4.3 Model Calibration

Model calibration was accomplished in two phases. The first phase used the observed CN (back-calculated), and the second phase used combinations of rainfall, CN, and observed flow. Calibration was stopped when $r = 0.85$ and $r^2 = 0.72$ as the plot in figure 3-10 believed

satisfactory. Figure 3-10 shows the final plots of rainfall versus CN, rainfall versus runoff, and runoff versus CN.

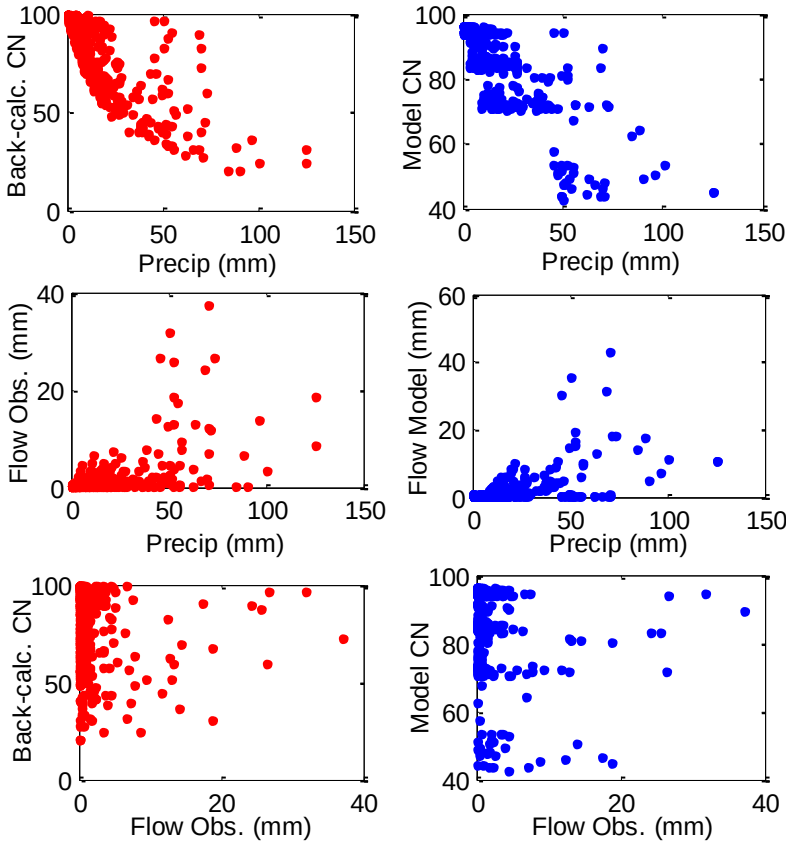


Figure 3-10: Plots to assess model performance during calibration

Note: Red plots are based on observed data, and blue plots are based on model output.

After satisfactory calibration output was found based on statistical measurements, the NDVI and model CN were used and remodeled using SAS to obtain the calibrated regression model of CN based on NDVI. The calibrated model is shown in Equation 3-8.

$$CN = -0.1251 * NDVI + 104.65 \quad (3\text{-}8)$$

The model was adjusted to account for theoretical bounds of CN; the CN cannot be greater than 98 according to the literature values based on the concept of the method (Equation 3-9). CN value of 98 indicates that almost all rainfall in an event would be runoff.

$$CN = -0.108 * NDVI + 98 \quad (3\text{-}9)$$

Table 3-7 shows pairwise comparisons of observed (measured), literature CN based, and calibrated model flow to assess the overall distribution of flow. Observed and calibrated model flows did not show statistical difference (p-value = 0.6622), but flow based on literature was statistically different from observed and calibrated model flows (p-value <0.0001).

Table 3-7: Pairwise comparison between observed, literature CN, and calibrated model flow

Differences of FlowGroup Least Squares Means						
FlowGroup	_FlowGroup	Estimate	Standard Error	DF	t Value	Pr > |t|
FCal	FLit	-1.7644	0.3385	1191	-5.21	<.0001
FCal	FObs	0.1479	0.3385	1191	0.44	0.6622
FLit	FObs	1.9123	0.3385	1191	5.65	<.0001

Figure 3-11 plots flows based on observed CN values with respect to standard (literature) CN, regression model CN, and calibrated model CN. The regression model flow fit the observed flow better than flow based on literature CN, and flow based on the calibrated model (slope = 0.91) was better compared to the literature (standard)-based CN flow (slope = 0.51).

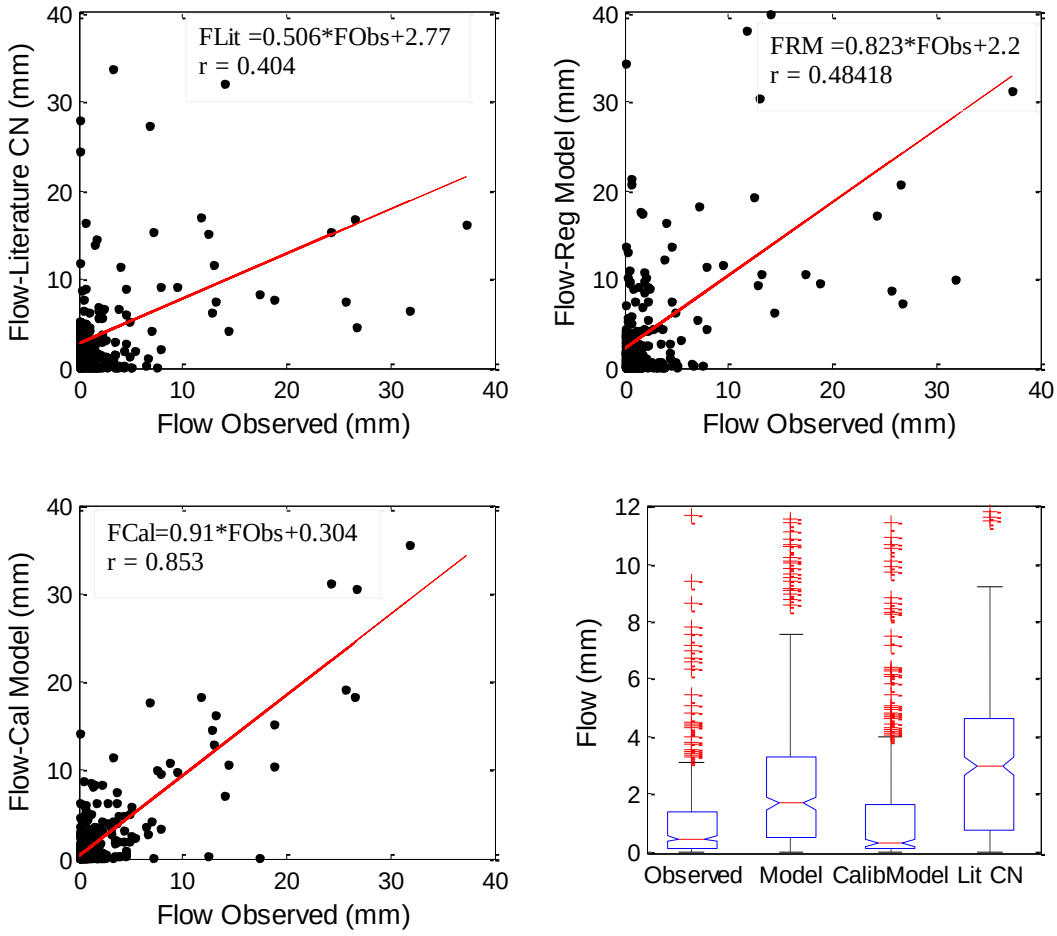


Figure 3-11: Calibration: literature CN, regression model, and calibrated model flows compared to observed flow at Konza Prairie Biological Station

Note: The box plot is zoomed to show a majority of the data distribution, thereby excluding some relatively extreme observations from the box plot.

The calibrated model accounted for approximately 91% of observed flow, while literature CN-based calculations explained less than 51% of observed flow. The error component, approximated by the *y*-intercept, decreased in the calibrated model compared to the literature-based flow. The *y*-intercept measured the lag or lead between the model and observed data with lower *y*-intercepts representing better model fits. The calibrated model exhibited the lowest *y*-intercept (0.304) compared to literature CN-based flow (2.77). As shown in the top-right of the figure, the flow regression model performed better than the standard CN flow, and calibration

improved model performance. The slope, intercept, and coefficient of correlation showed that the calibrated model performed very well.

The coefficient of correlation (*r*), which measures the strength of the linear relationship between the observed flow and estimated flow, supported the above analysis. Analysis showed that the *r* value was 0.40 for the literature CN flow, 0.48 for the uncalibrated regression model flow, and 0.85 for calibrated model flow with observed (measured) flow, respectively. These values indicate that the calibrated model improved runoff prediction compared to the standard method. Box plots of the flow also supported the analysis. As shown in the box plots in Figure 3-11, a majority of the calibrated flow had similar distribution to the observed flow. This study showed that the CN_{NDVI} provided more statistically acceptable results of CN than the standard CN.

3.4.4 Model Validation

The model was validated with 34% of the total available data, as described above and shown in Figure 3-12. In the figure the blue trend line shows results with all data designated for validation, the red trend line shows results after removing the most extreme data point generated from the highest rainfall point yielding almost no runoff, and the black trend line shows results after removing two additional data points. All three validation options demonstrated improved results compared to literature-based CN flow. In the first case the validated flow explained 87.3% of observed flow, with a moderately strong relationship ($r = 0.52$). In the second case the validated flow accounted for 89.3% of the observed flow, with a stronger relationship ($r = 0.641$), and in the last case the observed flow over-predicted the validated flow by 13.6%. The validated flow and observed flow had the strongest relationship ($r = 0.693$). However, the

literature CN-based flow accounted for 48.7% of the observed flow, with weak relationship (r = 0.44). In general, NDVI-based estimation of CN offers better rainfall-runoff prediction.

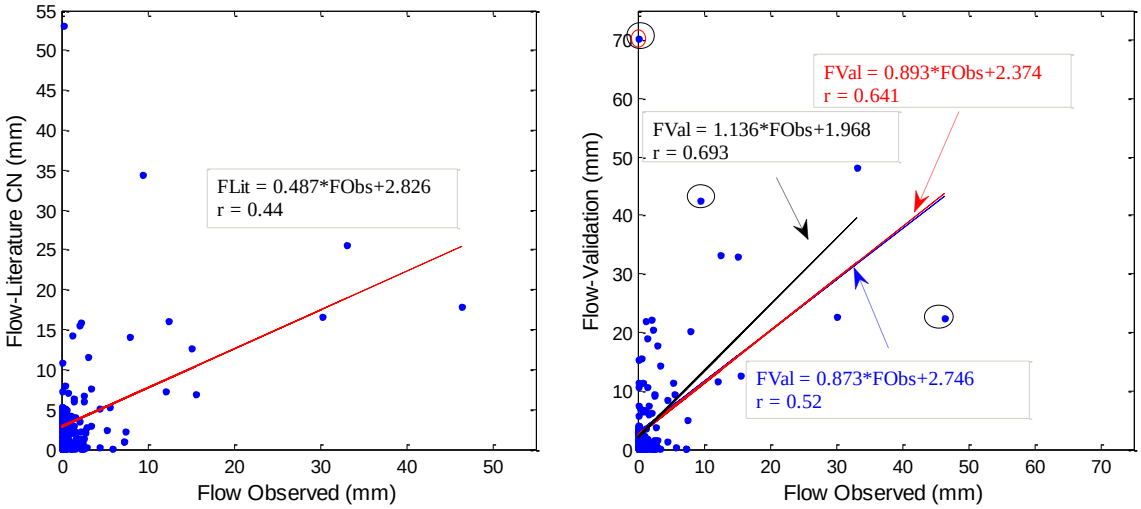


Figure 3-12: Validation: literature CN and calibrated model flows compared to observed flow at Konza Prairie Biological Station

Note: In the validation versus observed plot, the line and the equation in blue is validation data with no removal; the line and equation in red is validation by removing one event data in red circle, assumed to be an outlier; and the line and equation in black is validation by removing data from three events in the black circle.

The box plot shown in Figure 3-13 supported the analysis; validation flow was better than the literature CN (SCS-CN) flow compared to the observed flow.

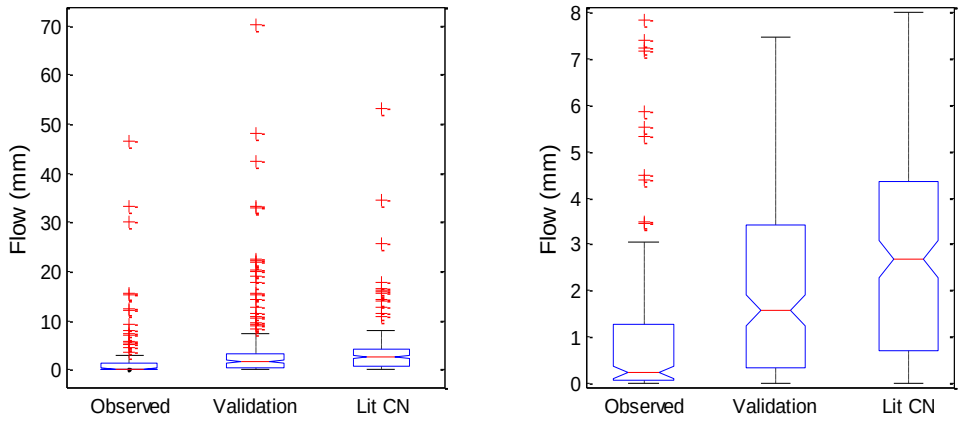


Figure 3-13: Observed, validation, and literature flow for watersheds at Konza Prairie Biological Station

Note: The second figure shows the zoomed portion of the first figure to highlight a majority of the data.

This analysis contained limitations, however, primarily due to the complex nature of the rainfall-runoff relationship. Davie (2008) defined runoff as the end-product of precipitation after all other hydrologic processes, meaning that the amount of runoff is dependent upon watershed hydrologic characteristics and other processes. This model attempted to incorporate the effects of various factors into one parameter, NDVI. In addition, possible errors in precipitation and runoff estimations for extreme events (low and high rainfall and runoff events) could be present; data from this study was primarily in the lower range. Harmel et al. (2006) compiled detailed uncertainties in streamflow using various measurement techniques for small watersheds. Those uncertainties may be associated with preciseness of the measurement techniques and environmental conditions such as wind effects. Another source of uncertainty could be the initial abstraction percentage in back-calculating CN for the analysis. Hawkins et al. (2008) analyzed and documented the unsettled nature of initial abstraction and provided a range of initial abstraction values based on the data source and type of analysis (natural and ordered data). Moreover, the coarse spatial scale of NDVI could also add errors and uncertainty. Even so, based on the analysis, NDVI provided a better prediction of flow than the standard literature-based CN method.

3.4.5 Spatiotemporal Variability of Curve Number

Studies have shown that hydrologic processes, soil moisture, infiltration, and other abstractive losses of rainfall have significant spatial and temporal variability (Gonzalez et al., 2014; Hosseini and Saradjian, 2011; Ponce and Hawkins, 1996); however, the standard CN method does not account for temporal variability of those processes (Ponce and Hawkins, 1996). In this analysis, MODIS-NDVI captured the effects of those processes on a 16-day cycle

(Figures 3-14, 3-15, and 3-17). Figure 3-14 shows the 11-year average CN for the Konza Prairie based on the time period of MODIS-NDVI data acquisition.

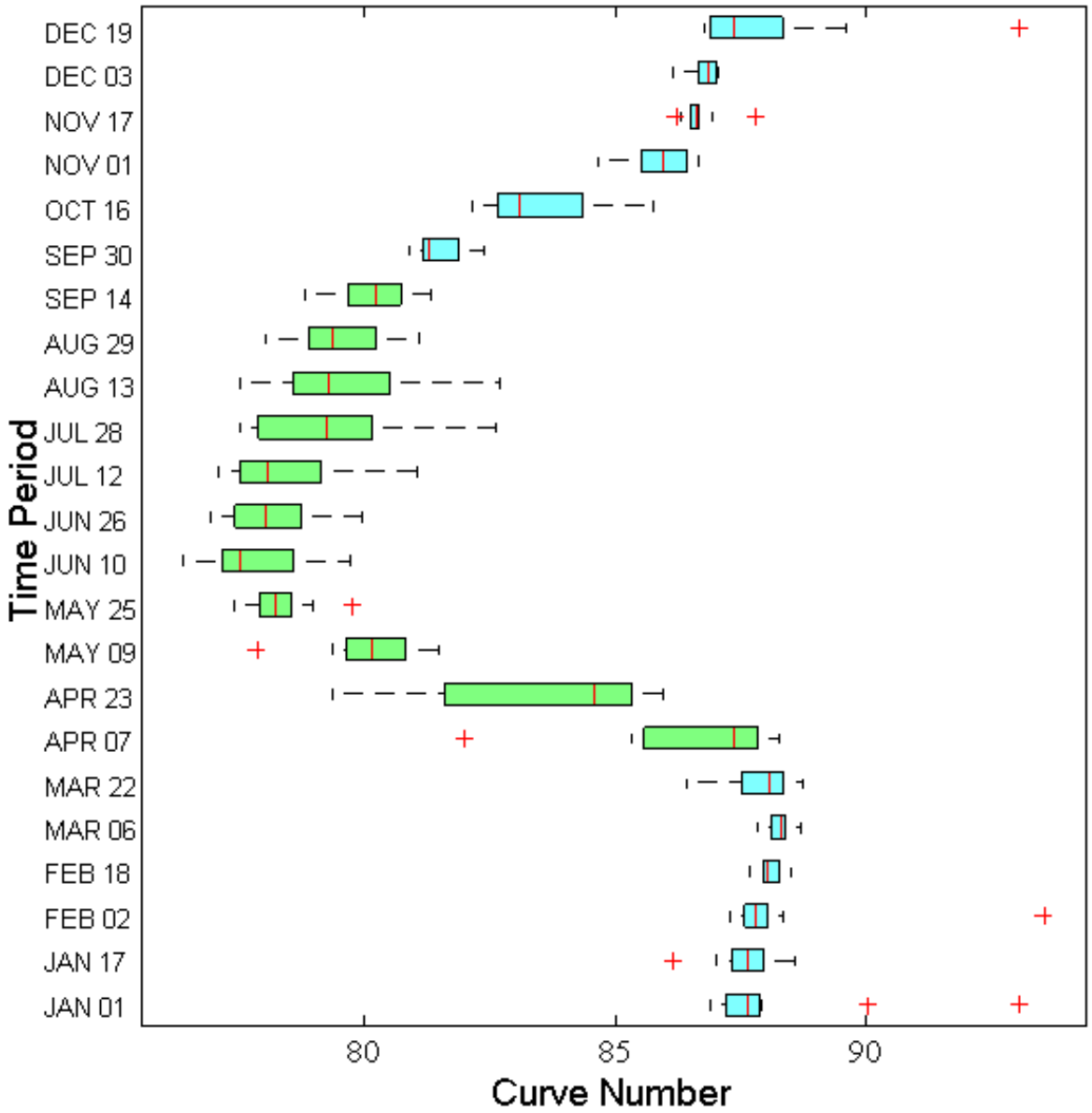


Figure 3-14: Konza Prairie CN_{NDVI} box plot for each time period of Normalized Difference Vegetation Index

Note: Green in the figure shows the growing season.

CN_{NDVI} demonstrated a vivid change during the beginning of the growing season (late March to April), with the lowest CN_{NDVI} around June 10. CN_{NDVI} standard deviation was greatest during the spring (April 23), which was expected based on varying spring weather and changes

of the landscape frost; and wider during the growing season in general when vegetation health varies based on various weather conditions (e.g., available soil moisture). Anandhi et al. (2013) highlighted the general distribution percentage of late spring frost from March 21 to May 20 from 1980 to 2009, supporting the variability of CN_{NDVI} in the growing season and providing opportunity to calculate relatively accurate runoff from rainfall events since most rainfall occurs during that season. The wide ranges of NDVI could be attributed to grazing and burning activities conducted in the growing season, especially during spring and fall, since the Konza Prairie is a long-term ecological research site with controlled grazing and burning practices.

In general the CN_{NDVI} was higher from November 01 to March 22 due to the lower NDVI so that less water could be held by the vegetation and creates a higher runoff potential. Also, during this season the soil profile is likely full of water due to lower ET. Occurrence of Snow could also contribute for lower NDVI in addition to barren land especially during December and January. In other hand, during growing season ET will be higher, thus infiltration high and runoff reduced. This process lead lower CN during the growing season.

For example, if a 50 mm rainfall event occurred and the CN increased from 75 to 77, the runoff would increase by 17.9%; if the CN increased by 5 (from 75 to 80), the runoff would increase by 47%. Runoff changes depend on the amount of rainfall and CN ranges. Figure 3-15 illustrates how CN_{NDVI} varies seasonally. Runoff-generating mechanisms vary by season due to landscape characteristics and available water in the soil. In general, CN_{NDVI} was lower during summer (June through August) and higher in winter (December through February), but there was a relatively wider range of CN_{NDVI} in the spring (March through May) and autumn (September through November). CN_{NDVI} variability in the spring and summer could reflect variability in the relationship of rainfall events and runoff occurrences in seasons in which most precipitation

occurs. During late autumn, winter, and early spring the runoff-generating mechanism is mainly dominated by available soil moisture, low surface temperature, and low ET that affect the proportion of runoff and infiltration during rainfall events. However during peak growing season, vegetation intercepts significant amounts of water and increases the porosity of the upper soil profile that increases infiltration.

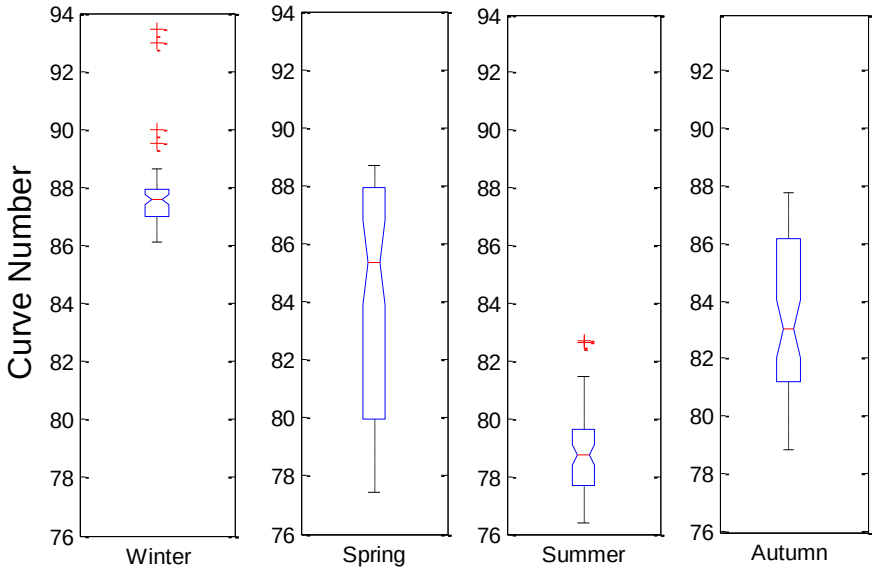


Figure 3-15: Konza Prairie Biological Station, KS, seasonal CN_{NDVI} averaged based on time period of Normalized Difference Vegetation Index (NDVI)

Figure 3-16 shows CN based on SCS lookup tables using land cover, hydrologic condition, and HSG. A majority of the area had a CN value of 77. Similar land cover and HSG were used throughout the year with some adjustments related to the hydrologic condition. It had lower values due to the riparian soil categorized as HSG A with low runoff potential. However, CN_{NDVI} demonstrated significant variability in spatial and temporal conditions (Figure 3-17).

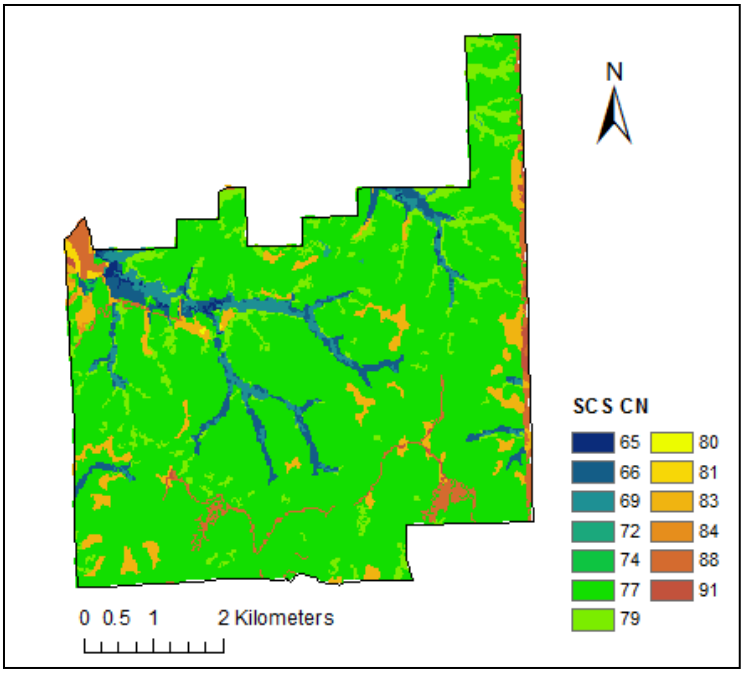


Figure 3-16: Konza Prairie SCS-CN based on NLCD land cover, SSURGO hydrologic condition, and hydrologic condition using SCS lookup table

CN_{NDVI} captured CN variability based on each NDVI period so that relatively accurate runoff could be estimated from rainfall events. Figure 3-17 shows CN_{NDVI} maps the Konza Prairie during five periods in 2010. As shown in the figure, CN_{NDVI} changed throughout those selected time periods. CN_{NDVI} decreased from March to July and then increased throughout the following periods. The map of May 9–24 supports the plot in Figure 3-17, showing drastic decreases of CN during the beginning of the growing season and wide variable CN values throughout the season. However, the period between March and November did not demonstrate significant CN variability.

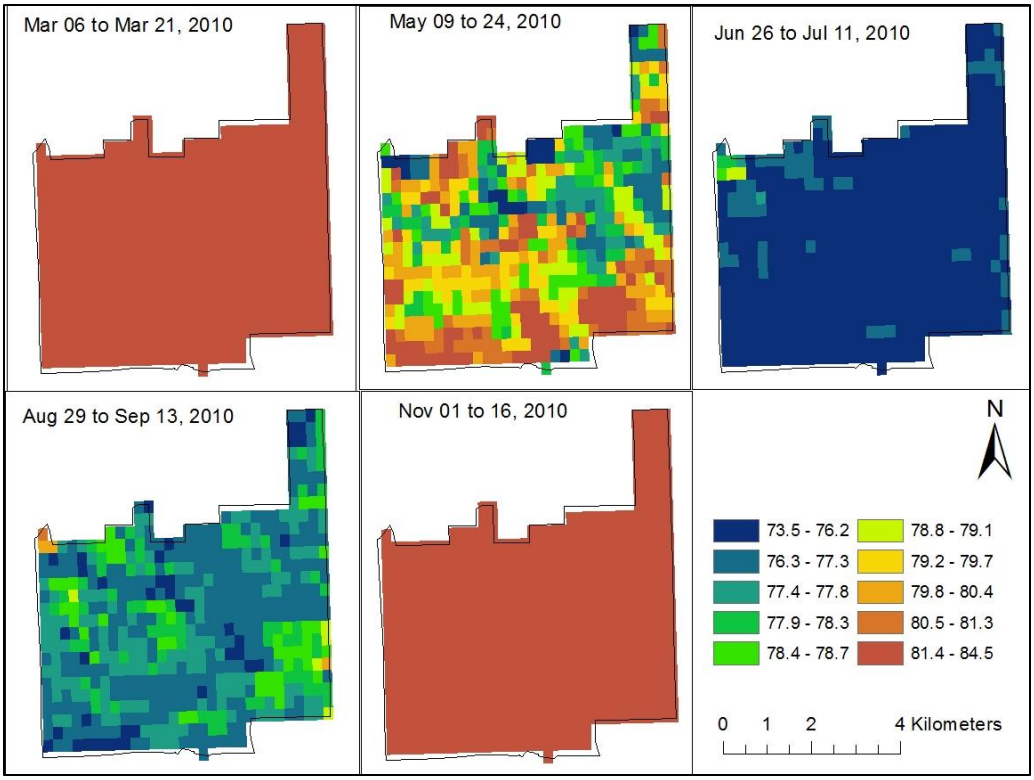


Figure 3-17: Konza CN_{NDVI} maps of selected periods in 2010 to assess the spatiotemporal changes in curve number

VI quality of the spatial maps in Figure 3-18 and Figure 3-19 were analyzed to assess whether or not cloud contamination was present in the study. Figure 3-18 shows the pixel reliability ranking for the five time periods of NDVI imagery. The imagery from March 6–21, 2010, shows that the entire area was categorized under marginal classification. A similar situation was observed from May 9–24, 2010. Further VI quality analysis was done for those two time periods of NDVI imagery.

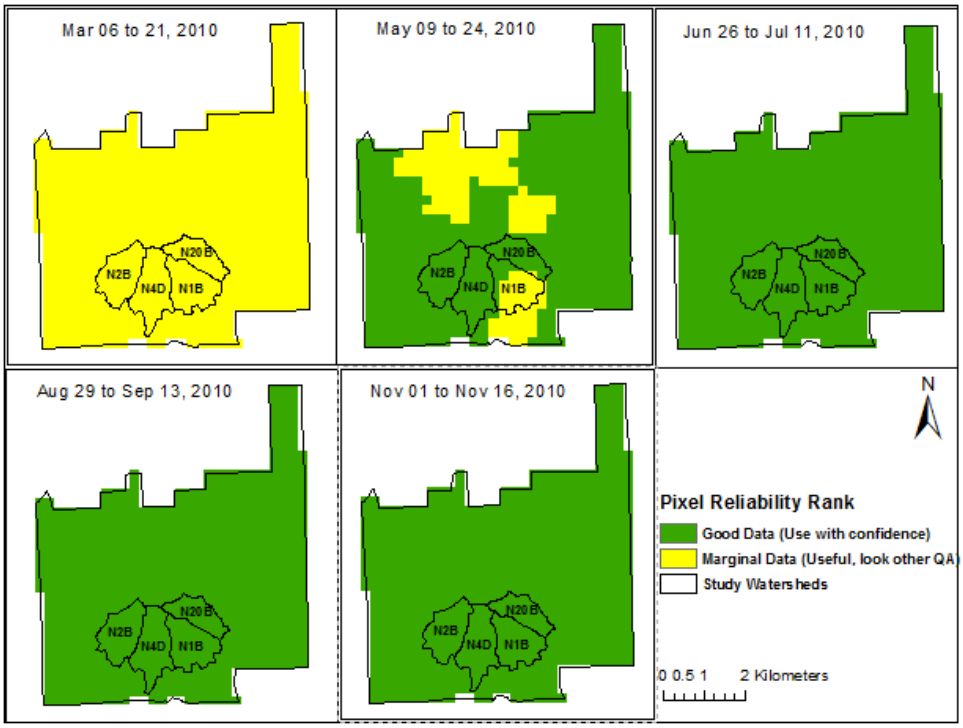


Figure 3-18: Pixel reliability ranking in Konza Prairie watersheds for corresponding NDVI periods of CN_{NDVI} in Figure 3-16

Figure 3-19 shows the pixel reliability rank and VI quality map of those two time periods with marginal data (pixel reliability is 1 [i.e., useful but look at other QA information]). According to the interpretation of VI quality described in Sections 3.3.5 and 3.4.5, the two imageries were not cloud contaminated, although they demonstrated decreasing quality. As also shown in Appendix A, Table 1, March 6 imagery was not used in model development because no rainfall and runoff data were used for model development. Imagery from May 09, 2010, was used in model development in which majorities of the watershed data were under good data pixel reliability classification (Figure 3-18).

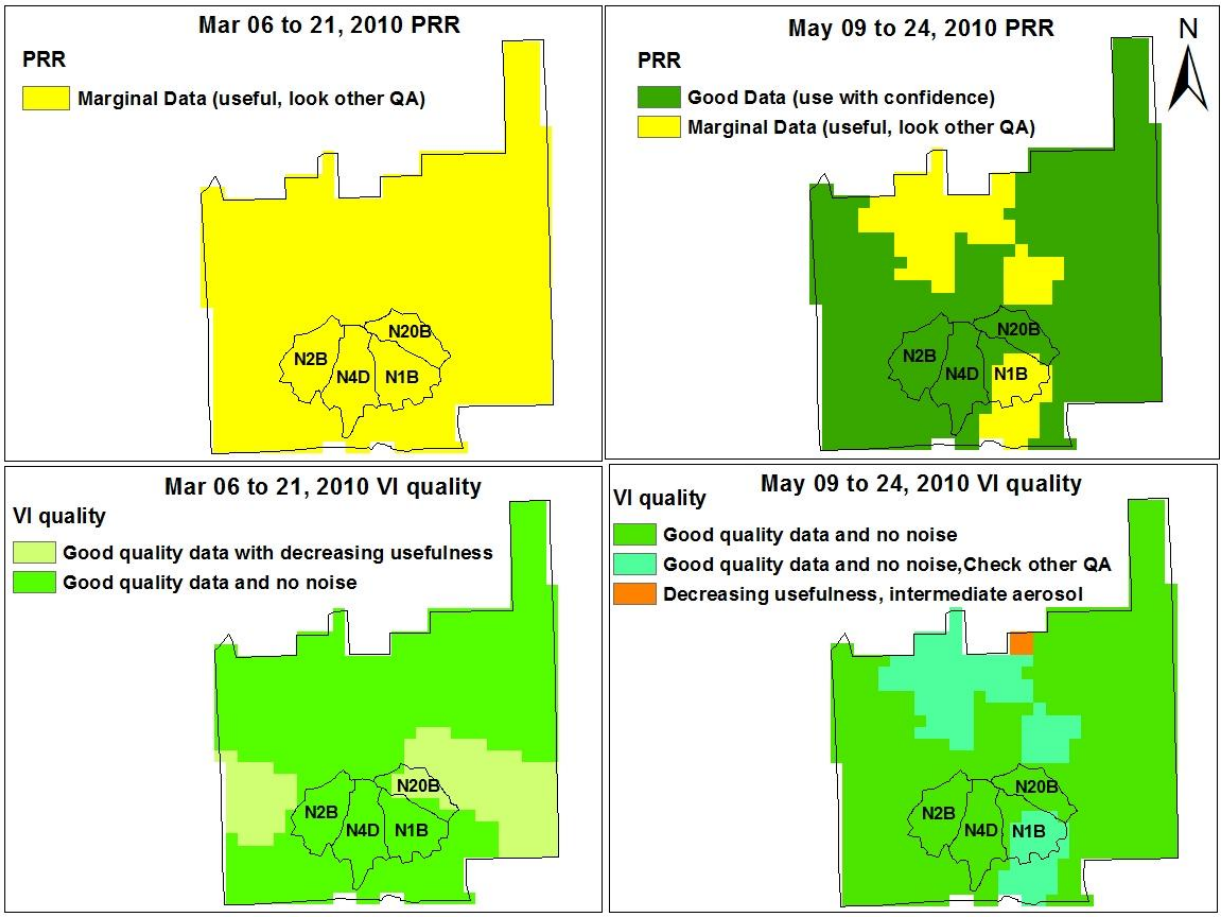


Figure 3-19: Pixel reliability rank (upper) and VI quality (lower) of MODIS (MOD13Q1) of two NDVI periods at Konza Prairie Bilogical Station

Note: No noise means no aerosols, no mixed clouds, no snow/ice, and no shadow.

3.5 Summary and Conclusion

CN prediction using any variable is more complicated than prediction via linear regression. However, in this study, NDVI was proven to be a better predictor than the existing standard CN method. In order to evaluate the applicability of NDVI to estimate CN_{NDVI} for 16 days, this study developed a regression model based on the assumption that NDVI and CN have an inverse relationship. Back-calculated CN using similar period runoff and precipitation from four watersheds in at the Konza Prairie Long-Term Ecological Research Station was used as observed CN for calibration and validation. During calibration the slope and intercepts were improved based on coefficient of determination (r^2) performances and comparison of model plots

and observed data. Results showed that derived models behaved satisfactorily during calibration and validation even though the rainfall event-runoff relationship is complex. The model also produced better CN than popular literature CN values for runoff predictions of a rainfall event.

The method used in this study could be adjusted (based on local data) and used in any watershed or climatic conditions in order to develop the CN_{NDVI} method. Improved rainfall and runoff measurement techniques and issues related to improved satellite data (e.g., higher resolution NDVI to capture land use/cover variability on the ground) could improve regression model development. This study showed that CN could be approximated using NDVI to capture the rainfall-runoff relationship more accurately than the standard method even though CN depends on factors, such as land use, soil physicochemical properties, and intensity, duration, and frequency of precipitation of the watershed, which reflects the complexity of the rainfall-runoff relationship. This method could help alleviate the static nature of the standard method and reflect a fairly accurate relationship of rainfall and runoff that could be improved since the resolution of satellite data is higher. The following chapter specifies how the developed model could perform in similar land use/cover with mechanical disturbances of land cover.

References

Anandhi, A., Perumal, S., Gowda, P. H., Knapp, M., Hutchinson, S., Harrington Jr, J., Murray, L., Kirkham, M.B. and Rice, C. W. (2013). Long-term spatial and temporal trends in frost indices in Kansas, USA. *Climatic Change* 120(1–2): 169–181.

Auerswald, K., Haider, J. (1996). Runoff Curve Numbers for Small Grain Under German Cropping Conditions. *Journal of Environmental Management* 47(3): 223–228.

Baltas, E., Dervos, N., and Mimikou, M. (2007). Technical note: Determination of the SCS initial abstraction ratio in an experimental watershed in Greece. *Hydrology and Earth System Sciences* 11(6): 1825–1829.

Beck, P. S., C. Atzberger, K. A. Høgda, B. Johansen, and A. K. Skidmore. (2006). Improved monitoring of vegetation dynamics at very high latitudes: A new method using MODIS NDVI. *Remote Sensing of Environment* 100(3): 321–334.

Boschetti, M., Stroppiana, D., Brivio, P. A. and Bocchi, S. (2009). Multi-year monitoring of rice crop phenology through time series analysis of MODIS images. *International Journal of Remote Sensing* 30:4643–4662. doi: http://dx.doi.org/10.1080/01431160802632249

Colombo, R., Busetto, L., Fava F., Di Mauro, B., Migliavacca, M., Cremonese, E., Galvagno, M., Rossini, M., Meroni, M., Cogliati, S., Panigada, C., Siniscalco, C., and di Cella U. M. (2011). Phenological monitoring of grassland and larch in the Alps from Terra and Aqua MODIS images. *Italian Journal of Remote Sensing* 43:83–96. doi: http://dx.doi.org/10.5721/ItJRS20114336

Davie, T. (2008). *Fundamentals of hydrology*. New York: Taylor & Francis.

D'Asaro, F., Grillone, G., and Hawkins, R. H. (2014). Curve Number: empirical evaluation and comparison with Curve Number handbook tables in Sicily. *J.Hydrol.Eng*. 19 (12); doi: 10.1061/(ASCE)HE.1943-5584.0000997

"Engineering Statistics Handbook." (2015). Are the model residuals well behaved? National Institute of Standards and Technology. --- Web. January 28, 2015. http://www.itl.nist.gov/div898/handbook/pri/section2/pri24.htm

Garen, D. C. and D. S. Moore. (2005). Curve number hydrology in water quality modeling: uses, abuses, and future directions. *Journal of the American Water Resources Association (JAWRA)* 41(2): 377–388.

Gonzalez, A., M. Temimi, and R. Khanbilvardi. (2014). Adjustment to the Curve Number (NRCS–CN) to Account for the Vegetation Effect on the Hydrological Processes. *Hydrological Sciences Journal* (just-accepted).

Harmel, R. D., R. J. Cooper, R. M. Slade, R. L. Haney, and J. G. Arnold. (2006). Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *American Society of Agricultural and Biological Engineers* 49(3): 689–701.

Hawkins, R. and A. Cate. (1998). Secondary effect in curve number rainfall runoff. *Powerpoint Representation at Water Resources Engineering* 98.

Hawkins, R. H. (1979). Runoff curve numbers from partial area watersheds. *Journal of Irrigation and Drainage Engineering* 105(ASCE 15024 Proceeding).

Hawkins, R. H. (1993). Asymptotic determination of runoff curve numbers from data. *Journal of Irrigation and Drainage Engineering* 119(2): 334–345.

Hawkins, R. H. (2014). Curve Number Method: Time to Think Anew? *Journal of Hydrologic Engineering* 19(6): 1059–1059.

Hawkins, R. H. and A. V. Khojeini. (2000). Initial abstraction and loss in the curve number method. *Hydrology and Water Resources in Arizona and the Southwest* 30:(0272–6106).

Hawkins, R. H., T. J. Ward, D. E. Woodward, and J. A. Van Mullem (2008). Curve number hydrology: State of the practice. *ASCE publication.*

Helsel, D. R. and R. M. Hirsch (2002). Statistical methods in water resources, U.S. Geological Survey, Reston Va. Hjelmfelt Jr., A. T. (1991). Investigation of curve number procedure. *Journal of Hydraulic Engineering* 117(6): 725–737.

Hmimina, G., Dufrene, E., Pontailler, J. Y., Delpierre, N., Aubinet, M., Caquet, B., de Grandcourt, A., Burban, B., Flechar, C., Granier, A., Gross, P., Heinesch, B., Longdoz, B., Moureaux, C., Ourcival, J. M., Rambal, S., Saint Andre, L., and Soudani, K. (2013). Evaluation of the potential of MODIS satellite data to predict vegetation phenology in different biomes: An investigation using ground-based NDVI measurements. *Remote Sensing of Environment* 132:145–158. doi: http://dx.doi.org/10.1016/j.rse.2013.01.010

Hosseini, M. and M. Saradjian. (2011). Multi-index-based soil moisture estimation using MODIS images. *International Journal of Remote Sensing* 32(21): 6799–6809.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83(1e2): 195e213.

Huete, A., Justice C., and Leeuwen W. V. (1999). MODIS vegetation index (MOD 13): Algorithm theoretical basis document.

Huffman, R. L., D. D. Fangmeier, W. J. Elliot, and S. R. Workman (2013). Infiltration and runoff. In Huffman, R. L., D. D. Fangmeier, W. J. Elliot, and S. R. Workman (eds), *Soil and Water Conservation Engineering* (115–143), 7th ed. ASABE.

Hutchinson. J. M. S., A. Jacquin , S. L. Hutchinson, and Verbesselt. J. (2015). Monitoring vegetation change and dynamics on U.S. Army training lands using satellite image time series analysis. *Journal of Environmental Management* 150:355–366.

Isik, S., L. Kalin, J. E. Schoonover, P. Srivastava, and B. G. Lockaby (2013). Modeling effects of changing land use/cover on daily streamflow: An Artificial Neural Network and curve number based hybrid approach. *Journal of Hydrology* 485:103–112.

Jesslyn F. B. (2015). Remote Sensing Phenology. USGS Phenology Fact Sheet. U.S. Department of the Interior | U.S. Geological Survey URL: http://phenology.cr.usgs.gov/index.php

Jeon, J., K. Lim, and B. Engel. (2014). Regional Calibration of SCS-CN L-THIA Model: Application for Ungauged Basins. *Water* 6(5): 1339–1359.

Kannan, N., C. Santhi, J. R. Williams, and J. G. Arnold (2007). Development of a continuous soil moisture accounting procedure for curve number methodology and its behaviour with different evapotranspiration methods. *Hydrological Processes* 22(13): 2114–2121.

"Konza Prairie Long-Term Ecological Research (LTER) Data Catalog." (2015). Konza Prairie. Kansas State University, May 31, 2012. Web. January 28, 2015. http://www.konza.ksu.edu/knz/pages/data/..%5C..%5Cdocs%5CDC.pdf

"Konza Prairie Long-Term Ecological Research (LTER), Methods Manual." (2011). Konza Prairie. Kansas State University, March 8, 2012. Web. January 28 2015. http://www.konza.ksu.edu/knz/pages/data/..%5C..%5Cdocs%5CMM.pdf

Lenhart, C., M. Titov, J. Ulrich, J. Nieber, and B. Suppes. (2013). The role of hydrologic alteration and riparian vegetation dynamics in channel evolution along the lower Minnesota River. *Trans.ASABE* 56:549–561.

Macpherson, G. (1996). Hydrogeology of thin limestones: The Konza Prairie long-term ecological research site, northeastern Kansas. *Journal of Hydrology* 186(1): 191–228.

Maselli, F., Romanelli, S., Bottai, L., and Zipoli, G. (2003). Use of NOAA–AVHRR NDVI images for the estimation of dynamic fire risk in Mediterranean areas. *Remote Sensing of Environment.* 86:187–197.

McCloy, K. R. and W. Lucht. (2004). Comparative evaluation of seasonal patterns in long time series of satellite image data and simulations of a global vegetation model. *Geoscience and Remote Sensing, IEEE Transactions* 42(1): 140–153.

Mendenhall, W. and T. Sincich. (2012). *A Second Course in Statistics: Regression Analysis*. 7th Ed., Pearson education.

Maselli F., S. Romanelli, L. Bottai, and G. Zipoli. (2003). Use of NOAA-AVHRR NDVI images for the estimation of dynamic fire risk in Mediterranean areas. *Remote Sensing of Environment* 86:187–197.

Mohamoud, Y. (2004). Comparison of hydrologic responses at different watershed scales, Ecosystems Research Division, USEPA, September 2004.

Narasimhan, B., R. Srinivasan, J. G. Arnold, and Di Luzio. M. (2005). Estimation of long-term soil moisture using a distributed parameter hydrologic model and verification using remotely sensed data. Transactions of the ASAE 48(3): 1101−1113.

"NASA LP DAAC" (2013). MODIS Land Products Quality Assurance Tutorial: Part 2, How to interpret and use MODIS QA information in the Vegetation Indices product suite https://lpdaac.usgs.gov/sites/default/files/public/modis/docs/MODIS_LP_QA_Tutorial-2.pdf

Oosterbaan, R. J. and H.J. Nijland (1994). Determining the saturated hydraulic conductivity. Chapter 12. In Ritzema, H. P. (ed), Drainage principles and applications, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, 1125.

Paruelo, J. M., Jobbágy, E. G., and Sala, O. E. (2001). Current distribution of ecosystem functional types in temperate South America. *Ecosystems* 4:683–698.

Pockrandt, B. R. (2014). A multi-year comparison of vegetation phenology between military training lands and native tallgrass prairie using TIMESAT and moderate-resolution satellite imagery. *MA Thesis*. Kansas State University URL: http://hdl.handle.net/2097/17320

Ponce, V. M. and R. H. Hawkins. (1996). Runoff curve number: Has it reached maturity? *Journal of Hydrologic Engineering* 1(1): 11–19.

Reynolds, C. (2014). Record Wheat Yields Expected for South Africa's 2014/15 Wheat Crop. *Commodity Intelligence Report* USDA, Foreign Agricultural Services.

Schroeder, B. (2006). Pattern, process, and function in landscape ecology and catchment hydrology - how can quantitative landscape ecology support predictions in ungauged basins? *Hydrology and earth system sciences* 10(6): 967–979.

"SCS Curve Number Method." (2015). Long-Term Hydrologic Impact Analysis (L-THIA). College of Engineering, Purdue University. --, Web. January 28, 2015. https://engineering.purdue.edu/mapserve/LTHIA7/documentation/scs.htm

Singh, R. P., Roy, S., and Kogan, F. (2003). Vegetation and temperature condition indices from NOAA AVHRR data for drought monitoring over India. *International Journal of Remote Sensing* 24:4393–4402.

Sesnie, S. E., Dickson, B. G., Rosenstock, S. S., and Rundall, J. M. (2012). A comparison of Landsat TM and MODIS vegetation indices for estimating forage phenology in desert bighorn sheep (Ovis canadensis nelsoni) habitat in the Sonoran Desert, USA. *International Journal of Remote Sensing* 33:276–286. doi: http://dx.doi.org/10.1080/01431161.2011.592865

Soriano, A. and Paruelo, J. M. (1992). Biozones: vegetation units defined by functional characters identifiable with the aid of satellite sensor images. *Glob. Ecol. Biogeogr*. Lett. 2:82–89.

Soudani, K., Francois C., le Maire, G., Le Dantec, V., and Dufrene, E. (2006). Comparative analysis of IKONOS, SPOT, and ETM+ data leaf area index estimation in temperature coniferous and deciduous forest stands. *Remote Sensing of Environment* 102:161–175. doi: http://dx.doi.org/10.1016/j.rse.2006.02.004

Soudani K., le Maire G., Dufrene E., Francois C., Delpierre N., Ulrich E., Cecchini S. (2008). Evaluation of the onset of green-up in temperate deciduous broadleaf forests derived from Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Remote Sensing of Environment* 112:2643–2655. doi: http://dx.doi.org/10.1016/j.rse.2007.12.004

Soulis, K., J. Valiantzas, N. Dercas, and P. Londra. (2009). Analysis of the runoff generation mechanism for the investigation of the SCS-CN method applicability to a partial area experimental watershed. *Hydrol Earth Syst Sci*. 6:373–400.

Tait, A. and Zheng, X. G. (2003). Mapping frost occurrence using satellite data. *Journal of American Meteorological Society* 42:193–203.

Troch, P. A., Carrillo, G. A., Heidbüchel, I., Rajagopal, S., Switanek, M., Volkmann, T. H., and Yaeger, M. (2009). Dealing with landscape heterogeneity in watershed hydrology: A review of recent progress toward new hydrological theory. *Geography Compass* 3(1): 375–392.

USDA, Soil Conservation Service. (1954 et seq). "National Engineering Handbook, Section 4 Hydrology." ca 400pp

USDA, Soil Conservation Service. (1986). Urban hydrology for small watersheds. SCS Technical Release 55, U.S. Government Printing Office, Washington, D.C.

Wang Q., M. Watanabe, S. Hayashi, and S. Murakami (2003). Using NOAAAVHRR data to assess flood damage in China. *Environ. Monit. Assess*. 82:119–148.

Wardlow, B. D. and S. L. Egbert (2010). A comparison of MODIS 250-m EVI and NDVI data for crop mapping: a case study for southwest Kansas. *International Journal of Remote Sensing* 31(3): 805–830.

Wolock, D. M., Winter, T. C., and McMahon, G. (2004). Delineation and evaluation of hydrologic-landscape regions in the united states using geographic information system tools and multivariate statistical analyses. *Environmental Management* 34(1): S71–S88.

Woodward, D. E., R. H. Hawkins, A. T. Hjelmfelt, J. A. VanMullem, and Q. D. Quan (2002). Curve number method: Origins, applications, and limitations. U.S. Geological Survey Advisory Committee on Water Information - Second Federal Interagency Hydrologic Modeling Conference. July 28–August 1, (2002), Las Vegas, Nevada.

Yuan, Y., W. Nie, S. C. McCutcheon, and E. V. Taguas. (2014). Initial abstraction and curve numbers for semiarid watersheds in Southeastern Arizona. *Hydrological Processes* 28(3): 774–783.

Zhang, G. and H. Savenije. (2005). Rainfall-runoff modelling in a catchment with a complex groundwater flow system: application of the Representative Elementary Watershed (REW) approach. *Hydrology and Earth System Sciences Discussions* 9(3): 243–261.

Zhou, M. C., Ishidaira, H., Hapuarachchi, H. P., Magome, J., Kiem, A. S., and Takeuchi, K. (2006). Estimating potential evapotranspiration using Shuttleworth-Wallace model and NOAA-AVHRR NDVI data to feed a distributed hydrological model over the Mekong River basin. *Journal of Hydrology* (1–2):151–173.

Chapter 4 - Paired Watershed Selection and Application of NDVI-based Curve Number in Disturbed Land

Abstract

Paired watershed design has been used to study the impact of land use changes on hydrology. It is based on the idea that watersheds located in close proximity will respond similarly to the climatic variables and land use/cover changes. Even though it has been widely used, the method lacks objective criteria for the selection of paired watersheds. This study focuses on statistical methods to select paired watersheds to study land use change impacts on hydrological processes and the potential to use the CN_{NDVI} for assessing runoff from small watersheds.

A combined K-means and hierarchical-agglomerative clustering techniques were applied in this study to identify hydrologically homogeneous paired watersheds on Fort Riley Military Installation, KS. Clustering techniques were done using seven topographic variables (total stream length, drainage density, ruggedness, total mean slope, no flat area slope, percent flat area, curvature) and one soil parameter (percent clay in the upper layer). These parameters play important roles in hydrologic processes by revealing surface runoff behavior, indicating closeness of spacing of channel network, measuring terrain heterogeneity, affecting overland and subsurface flow velocity that determine the rate of runoff, converging/diverging flow and soil moisture and influencing infiltration, respectively. In order to eliminate the domination of one variable due to the different magnitudes of parameters, Z-score standardization was done to improve generalizability of the results. The K-means clustering techniques led to six clusters for

optimal classification with a silhouette value of 0.42. The watersheds in each cluster were paired with higher Cophenetic correlation coefficient with values greater than 0.80 revealed the validity of the agglomerative hierarchical clustering technique. The performance of CN_{NDVI} was further assessed using two combinations of paired watersheds in low, medium and high intensities. The results showed that CN_{NDVI} were able to predict the land cover impact and differentiate the maneuver difference in intensities. The results also revealed that CN_{NDVI} model performs better than the standard SCS CN at smaller scale watershed.

4.1 Introduction and Background

Human transformation of the earth's land surface to feed, shelter, and accommodate the increasing human population has become a concern for this century (DeFries and Eshelman, 2004; DeFries et al., 2004). Expanding agricultural areas, growing developed areas, and deforestation are expected to accelerate due to an increasing population worldwide. Anthropogenic land transformation has multiple consequences, including streamflow alteration with high peak flows and reduced recharge (Storck et al., 1998; Rose and Peters, 2001), altered atmospheric circulation with reductions in precipitation, ET, and cloudiness (Werth and Avissar, 2002), and habitat fragment and species losses. Numerous studies have investigated the effect of human land transformation on climate, habitat loss, and geomorphic, hydrologic processes, and ecological processes (DeFries et al., 2004; DeFries and Eshelman, 2004; Gaston et al., 2003; Jetz et al., 2007; Sala et al., 2000; Vitousek et al., 1997; Wilson, 2012), indicating that understanding the consequences of land use change, especially on hydrological processes, should be a primary research focus.

Consequences of land use change on hydrology have been studied via modeling, experimental watersheds, and measurements, especially at the watershed scale. A watershed is

the basic organizing unit for a wide range of scientific, engineering, and water management activities (Genereux et al., 2005). When studies are done at the watershed scale, results can be extrapolated to a larger scale with appropriate precautions and used for management and planning activities. Watersheds with controlled and experimental manipulated land surface, together with pre- and post-manipulation observations, have been used to understand land-use effects on hydrological process (Nobert and Jeremiah, 2012; Schilling, 2002). However, identifying and quantifying hydrological consequences of land-use change is still challenging due to the short length of hydrological records, highly variable hydrological systems, and difficult control of land use changes in real system (DeFries and Eshelman, 2004).

Current understanding of land use effects on hydrology are based on paired watershed studies (DeFries and Eshleman, 2004; Schilling and Spooner, 2006; King et al., 2008; Prokopy et al., 2011). The paired watershed approach captures the effects of land cover, climate, and watershed hydrological variability, and it eliminates the need to measure all components that cause changes throughout watersheds under study (Andréassian, 2004). The paired watershed approach engages two similar watersheds (control and treatment) that are synchronously monitored during calibration (pretreatment) and post-treatment periods (Clausen and Spooner, 1993). Watershed similarity leads to the belief that both watersheds respond similarly to climatic forcing, land use/cover, and soil physic-chemical property changes, so that a conclusion can be made on the impact of the treatment and control watersheds. Paired watershed studies provide high quality experimental data that seeks to advance the understanding of the hydrologic response of watersheds to land use changes (DeFries and Eshleman, 2004). Paired watersheds have also been used to study environmental conditions such as organic carbon loss (Veum et al., 2009), soil erosion (Ricker et al., 2008), non-point source pollution (Prokopy et al., 2011),

conservation practices (King et al., 2008), and forest management practices (Andréassian, 2004; Ford et al., 2011).

Despite relatively wide use of paired watershed applications, no consistent statistically sound criteria and methods exist for selecting paired watersheds. Limited paired watershed selection has been done based on subjective professional judgments. Other studies have used clustering techniques, such as k-means (Burn and Boorman, 1993; Nathan and Mcmahon, 1990; Ssegane et al., 2012; Razavi, 2013), canonical correlation (Cavadias et al., 2001; De Prinzio, 2011), hierarchical clustering (Kahya et al, 2008), and fuzzy clustering algorisms (Rao and Srinivas, 2006) for watershed selection.

Selected watersheds should be representative of the region so that study results are transferrable to other (ungauged) watersheds (Blöschl and Sivapalan, 1995). Watershed representativeness and similarity can be determined with respect to vegetative cover, soil properties, and topographic and climatic conditions (Chang, 2006; Ssegane, 2013), especially in hydrological studies. The paired watershed approach utilizes land surface topography, vegetative cover, and soil physiochemical properties to determine hydrological processes, especially surface and groundwater processes. This study devised a paired watershed selection method based on soil, topographic parameters, and clustering techniques. The objectives of the study were to conduct paired watershed selection by devising a statistical method based on topographic and soil parameters that dominate hydrological processes in small watersheds and apply developed CN_{NDVI} to study the impact of military maneuvers on runoff generation using paired watersheds with various training intensities.

4.2 Study Area

The study area, Fort Riley military base, is located in the Flint Hills of northeastern Kansas and covers approximately 411 km^2 in Clay, Geary, and Riley counties (Figure 4-1). The elevation of the area ranges from 313 m to 419 m. Fort Riley experiences a typical midcontinental climate characterized by large seasonal temperature contrasts with cold, relatively dry winters and hot summers. Average monthly temperature ranges from -2 ^oC in the winter to 26 ^oC in the summer (Althoff et al., 2005), while average precipitation varies from 812.8 mm to 863.6 mm (USDA, 2007). Farnum Creek, Honey Creek, Little Arkansas Creek, Madison Creek, Sevenmile Creek, Timber Creek, Threemile Creek, and Wildcat Creek are located in Fort Riley. Downstream flooding due to increased flow in these creeks, especially Wildcat Creek, is a major concern (Bunger, 2013).

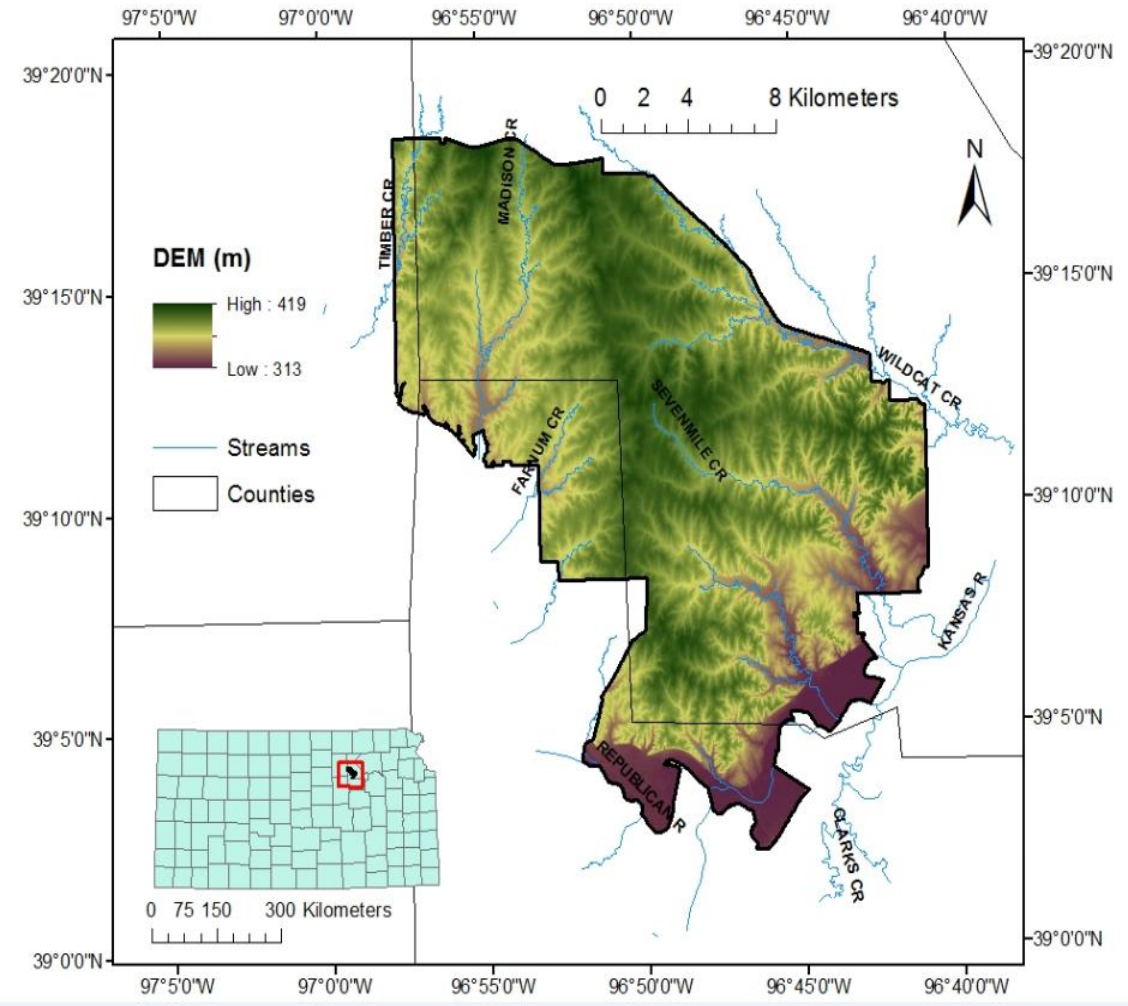


Figure 4-1: Elevation range of Fort Riley military installation located in Northeast Kansas with counties and State map

Fort Riley land use (Figure 4-2) is dominated by grassland (63% of total area); other land cover includes forest (20%), developed (11%), and crop areas (3.5%). Forest land is mainly located along the Kansas and Republican rivers. The Fort Riley area is dominated by Wymore series soils (USDA Soil Conservation Service, 1975; Althoff et al., 2009a), which consists of moderately well-drained silty clay loam and silty clay soils. Based on SSURGO, a majority of the area is C and D HSGs (Figure 4-2), indicating that the area is prone to high runoff potential compared to type A and B soils (Clausen and Spooner, 1993).

Fort Riley military base was established in 1853 and currently serves as a combat training ground for mortar and artillery fire, maneuver training, and mechanized infantry units (Althoff et al., 2007; Pockrandt, 2013; Quist et al., 2003). Fort Riley is divided into training areas with average areas of 2.76 km^2 (0.43 km^2 minimum and 6.46 km^2 maximum) and maneuver areas (larger than training areas). Activities at the base affect vegetation cover, biotic component compositions, and soil physicochemical properties (Althoff et al., 2009a). These activities have been shown to detrimentally affect ecosystems.

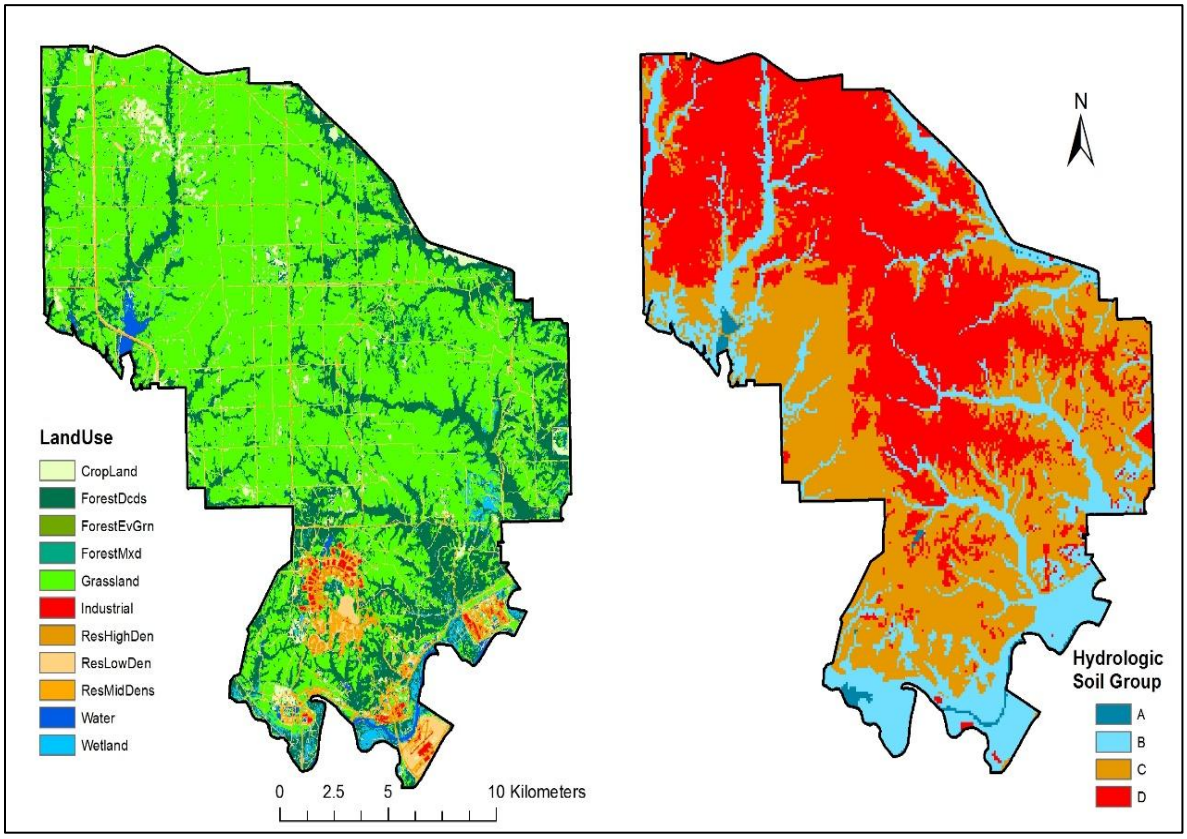


Figure 4-2: National Agricultural Statistics Service- Cropland Data Layer (NASS CDL) 2010 land use/cover and Soil Survey Geographic Database (SSURGO) hydrologic soil group of Fort Riley, KS

4.2.1 Military Training Maneuver and Landscape Impacts

Military training maneuvers at Fort Riley include off-road use of large tracked and wheeled vehicles. Impacts of military maneuvers on the land surface have been reported in many studies (Althoff and Thien, 2005; Althoff et al., 2007; Althoff et al., 2009a; Althoff et al., 2009b; Anderson et al., 2005, 2006; Bhat et al., 2005; Foster et al., 2006; Haugen et al., 2003; Kun et al., 2009; Perkins et al., 2007; Thurow et al., 1993). Impacts include removal/damage of ground vegetation cover, alteration of soil physical properties, increased soil exposure, and accelerated soil erosion (Howard et al., 2011; Quist et al., 2003). Maneuvers have also been shown to reduce native flora and fauna and spread invasive species (Milchunas et al., 2000; Prosser et al., 2000). Perkins et al. (2007) found high correlations between training intensity and ecological disturbances. Althoff et al. (2009a) reported reduction of vegetative biomass with repeated tank traffic at Fort Riley. Althoff and Thien (2005) found loss of biodiversity due to military training activities.

Vegetation destruction has been reported as the primary effect of military maneuvers (Althoff et al., 2009a; Anderson et al., 2005). Destruction of vegetation can cause significant secondary effects, including soil loss through accelerated wind and water erosion. Althoff et al. (2007, 2009a, and 2009b) reported substantial changes in landscape conditions and physiochemical soil properties due to military activities. The authors reported that mechanized military training causes landscape-scale soil disturbances that affect soil quality through displacement and compaction. Quist et al. (2003) also reported that intensive mechanized military maneuvers remove vegetation cover, increase bare and compacted soil, and shift the composition of plants. Changes in soil physical properties and land use/cover disturbances affect hydrological processes, specifically the rainfall-runoff relationship.

Althoff and Thien (2005) and Perkins et al. (2007) found reduced mean soil pore sizes and increased soil bulk density in training areas compared to undisturbed areas. Disturbed pore structure of the soil alters infiltration and may cause poor soil aeration. These changes can affect vegetation growth by inhibiting root growth, nutrient uptake, and seedling emergence. Displacement of organic-rich topsoil and compaction resulting in reduced infiltration and water-holding capacity have been reported as common effects of military training activities (Grantham et al., 2001; Prosser et al., 2000; Raper, 2005). Thurow et al. (1993) reported a decline in the infiltration rate in wet-tracked military maneuver conditions. The reduced and/or lost vegetation cover results in less interception and dissipation of raindrop energy directly into the ground. Dissipation of raindrop energy and reduced infiltration and interception create an erosion-prone environment. These studies showed that military maneuvers can alter regional hydrological processes. The degree of hydrological impact due to military maneuvers and rates of subsequent recovery of the system are dependent upon site characteristics such as vegetation type, soil texture, soil type, soil moisture content at the time of impact, vehicle types and maneuvers, and local climatic characteristics (dry and wet). The net effect of soil and vegetation degradation and recovery associated with a particular activity can be studied by measuring the change in hydrological process, especially change in infiltration rate and runoff (Thurow et al., 1993).

4.3 Materials and Methods

4.3.1 Watershed Delineation

Watersheds of approximately 1 km^2 were delineated using ArcGIS model builder application (Figure 4-3) to capture watersheds with relatively similar military training intensity in order to study the hydrological impact of maneuvers. Hydrology tools in Spatial Analyst Tools of ArcGIS 10 were used to process watersheds from a 3 m DEM that was downloaded from the

USGS “Seamless” archive (http://seamless.usgs.gov). The DEM was filled to remove sinks in the surface raster and reduce DEM imperfections prior to conducting a flow direction that created a raster layer with values indicating flow direction from the highest point down-slope to the lowest point. Accumulated flow was computed based on flow direction used to determine the stream network. Conditional evaluation on each cell of accumulated flow was conducted to determine the average size of the watersheds to determine the contributing area above a set of cells using the watershed tool. The raster watersheds were converted to vector data format.

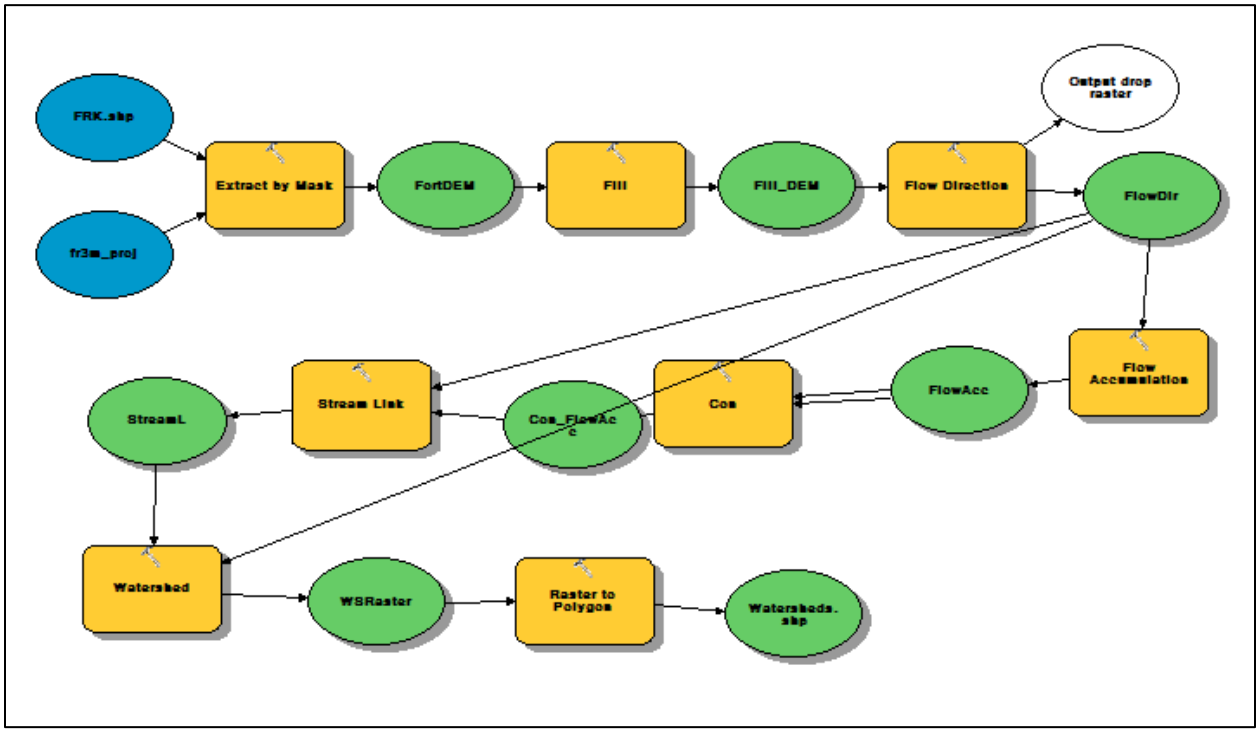


Figure 4-3: ArcGIS model builder to delineate watersheds in Fort Riley with average areas of 1 km^2

4.3.2 Topographic Parameter Selection and Cluster Analysis

Various topographic and soil variables were used in paired watershed selection because hydrological processes in any landscape are complex and defining the optimum governing equation of all processes using one or two variables is difficult. A total of 8 watershed topographic parameters were assessed to conduct cluster analysis and apply to paired watershed

selection analysis (Table 4-1). The first seven parameters were derived from the 3 m DEM, and the percent clay of the top zone of the soil was acquired from the SSURGO database. These variables were selected as the most determinant parameters in the hydrological processes, particularly to the study area since topographic characteristics affect hydrologic processes at field and small-scale watersheds (Christopher et al., 2010; Horton, 1932; Olaya, 2009; Riley et al., 1999; Strahler, 1957; Wilson, 2012). A brief description of each selected topographic parameters is given in Table 4-1 and the following sections.

Total stream length*:* The most important hydrologic feature of the basin is stream length, which is skewed to the right (Strahler, 1957) and denotes surface characteristics. Relatively small stream lengths reveal the surface area with steep slopes and fine texture; long stream lengths denote flat surfaces (Christopher et al., 2010). Stream length is generally inversely proportional to the stream order.

Drainage density: Drainage density is a linear scale of landform elements in stream-eroded topology (Horton, 1932), revealing the closeness of channel spacing. Drainage density is the ratio between stream length to total watershed area (Strahler, 1957), expressed in terms of mi. mi/sq. or km. km/sq. Based on drainage density, drainage basins can be classified as low density basin ($D < 12$), medium density basin ($D = 12–16$), and high drainage density ($D > 16$) (Strahler, 1964). Low drainage density has been widely noticed in areas with highly permeable subsoil material under dense vegetation. High drainage density occurs in weak or impermeable, sparse vegetation and mountainous areas (Christopher et al., 2010). A high drainage density value indicates a relatively high density of streams and a rapid storm response with high runoff yield.

Ruggedness: Ruggedness index indicates the instability of a land surface (Olaya, 2009; Strahler, 1957). Ruggedness is calculated using the range of elevation divided by the square root of the watershed area. Ruggedness provides a measure of terrain heterogeneity (Riley et al., 1999).

Flat area, mean Slope of total area, and mean slope of no-flat area: Slope is one of the most important factors governing the rate of runoff. Runoff flows faster on a steep slope; a high slope results in decreased response time and high peaks at downstream locations, while flat slope results in extended response time and high infiltration. In order to include impact of sloped and flat areas, both parameters were studied individually. Mean slope of total area was used to study the combined effect. Slope can be calculated as a percent of slope or degree of slope using ArcGIS.

Mean curvature: Curvature describes the curve of a surface at a particular point in the landscape. Profile curvature affects the acceleration or deceleration of flow across the surface, converging/diverging flow, and soil moisture.

Mean clay of upper layer: The type of soil in the upper soil layer significantly affects runoff due to its infiltration rate. Pore size and pore distribution affects infiltration process. Mean clay in the upper soil layer causes variation in pore size and distribution as it changes with wetting and drying.

Table 4-1: Topographic parameters used in paired watershed analysis

	Variable	Range	Comment	Significance in Hydrological Process
1	Total stream length	0.6–14.9 km	Streams are delineated based on a 3 m DEM	Distance of stream from its confluence with another stream or water body, reveals surface runoff behavior
2	Drainage density	0.5–46.7 km/km^2	Total stream length divided by the area of the watershed	Indicates closeness of spacing of channel network
3	Ruggedness	0.00–0.2 km^2	The range of elevation divided by the square root of area of the watershed (Olaya, 2009)	Measures terrain heterogeneity
4	Flat area	21.4%–79.9%	Derived from a 3 m DEM	Affects overland and subsurface flow velocity and determines the rate of runoff
5	Mean slope of total area	1.5–10.8 degree		
6	Mean Slope of no-flat area	6.5–13.4 degree		
7	Mean curvature	-0.02–0.01 one hundredth (1/100) of a z-unit	The second derivative of a surface or the slope of the slope, profile curvature processed for this analysis.	Profile curvature affects the acceleration or deceleration of flow across the surface, converging/diverging flow and soil moisture
8	Mean Clay of Upper Layer	19.5-36.5%	Based on the SSURGO soil database which is the % of the upper 50 cm of the soil	Affects infiltration

Clustering methods utilized variability of parameters in order to categorize the appropriate groups based on them and to match each paired watershed (Martinez and Martinez, 2008; MathWorks, 2012). However, the scale and magnitude difference among parameters caused the clusters to be dominated by wider and larger magnitude parameters, such as percent

flat area (Table 4-1). The eight variables had unique units and ranges of magnitude, as listed in Table 4-1. Histograms and boxplots were used to identify distribution and variability of selected parameters. For example, mean clay in the upper layer ranged from 19.50 to 36.50, and ruggedness ranged from 0.00 to 0.19. Therefore, if the data were not standardized, mean clay values could have higher role in pairing watersheds statistically. To alleviate that problem, even though the study required nonparametric analysis, standardization of data was done using z-score function available in MATLAB (MathWorks, 2012; Martinez and Martinez, 2008). Standardization of z-scores allows normalization of data. Mathworks (2012) describes z-score as a measure of the distance of data based on the mean in terms of the standard deviation. Z-score standardization returns each variable value to a mean centered at zero and scaled to have a standard deviation of 1, normalizing the data and equalizing variance for positively and negatively skewed variables. The z-score standardization function in MATLAB uses Equation (1). Z-scores were computed using the mean and standard deviation for each parameter to standardize the observations (Martinez and Martinez, 2008).

$$z = \frac{x-\mu}{s} \qquad (4\text{-}1)$$

where z is the standardized value, x is actual observation data value, μ is the mean value of each variable, and s is the standard deviation of data.

Interdependence of variables was checked using Pearson's correlation technique and a variance-covariance matrix. Pearson's correlation values were plotted into a correlation matrix with pseudo color for easy results interpretation. The correlation matrix was also used to remove redundancy if one or more variables had a strong relationship with the remaining. Variance-covariance matrix of the parameters was then calculated to validate correlation matrix results.

Two methods of clustering, k-means and agglomerative hierarchical clustering, were used for cluster analysis using the Statistics Toolbox of MATLAB. Both techniques utilize measures of distance using the pairwise distance between pairs of objects (pdist) function that measures dissimilarity between every pair of watersheds in a dataset (MathWorks, 2012; Martinez and Martinez, 2008). For example, in k-means, cluster centroids are determined based on the specified number of clusters so that the pdist function can determine the distance using a combination of variables from each centroid. Both methods are well-known clustering techniques; there are advantages and disadvantages for each method. When a similar pattern exists, hierarchical clustering cannot determine distinct clusters. The actual expression patterns become less effective when clusters sizes increase. In the case of K-means clustering, it requires a specified number of clusters in advance. It is also sensitive to outliers. This study used a combination of K-means and agglomerative clustering techniques to minimize the weaknesses and take advantage of the strengths of the two methods. K-means clustering was first used to reduce the large number of watersheds (316 watersheds), and then agglomerative hierarchical clustering was used to pair individual watersheds within the groups.

4.3.2.1 K-means

K-means clustering is a partition clustering algorithm (Cao et al., 2013; Martinez and Martinez, 2008; Martinez and Martinez 2004) that clusters data into *k* groups so that the within-group sum-of-squares is minimized (Martinez and Martinez, 2008; Martinez and Martinez, 2004). K-means clustering uses a two-step basic procedure: 1) assign observations to the closest group, and 2) calculate the cluster centroid using the assigned objects. It uses a two-phase iterative algorithm (i.e., batch updates and online updates) to minimize the sum of point-to-centroid distances (MathWorks, 2012). K-means is an inbuilt function in the Statistics Toolbox

of MATLAB. The following is the detailed k-means algorithm procedure (Martinez and Martinez, 2008):

i. Obtain a partition of *k* groups using silhouette statistics.

ii. Take each data point ($\boldsymbol{x}_i$) and calculate the Euclidean distance between it and every cluster centroid.

iii. Here $\boldsymbol{x}_i$ is in the r^{th} cluster, n_r is the number of points in the r^{th} cluster, and d^2_{ir} is the Euclidean distance between $\boldsymbol{x}_i$ and the centroid of cluster *r*. If there is a group *s* such that $\frac{n_r}{n_r-1} d_{ir}^2 > \frac{n_s}{n_s+1} d_{is}^2$, then move $\boldsymbol{x}_i$ to cluster *s*.

iv. If there are several clusters that could satisfy the above inequality, then move $\boldsymbol{x}_i$ to the group that has the smallest value for $\frac{n_s}{n_s+1} d_{is}^2$.

v. Repeat steps *ii* through *iv* until no more changes are made.

Using this method, similar watersheds were grouped within a group, while others were grouped into a different group. The optimum number of groups was specified using mojena-upper tail rule (Mojena, 1977) and silhouette statistics (Kaufman and Rousseeuw, 1990; Martinez and Martinez, 2008; MathWorks, 2012). The upper tail rule of mojena (Mojena, 1977) was used to determine the number of clusters based on raw fusion levels (Martinez and Martinez, 2008). The possible silhouette value ranged from -1 to 1, where the greater silhouette value was assumed to be a better cluster. The silhouette value was used to validate the upper tail rule of mojena results because various numbers of possible clusters may exist in unsupervised learning techniques. Use of both methods helps determine the most optimal number of clusters.

4.3.2.2 Agglomerative Hierarchical Clustering

Agglomerative hierarchical clustering merges the closest pair of watersheds within a group based on clustering distance. The Euclidean distance technique of pdist function in the

Statistics Toolbox of MATLAB was used to calculate the pairwise distance between watersheds (MathWorks, 2012). Then the centroid linkage method, a method that calculates the distance between two clusters as the distance between centroids, was used to perform agglomerative clustering (MathWorks, 2012). A dendrogram was used to show links between the best pairs; the tree was not a single set of clusters but rather a multilevel hierarchy of related watersheds (MathWorks, 2012). Cophenetic correlation coefficient was used to assess the correlation between clusters.

4.3.3 Application of CN_{NDVI} for Runoff Estimation Fort Riley

Fort Riley military installation is in close proximity to the Konza Prairie Biological Station with similar land cover but very different management and disturbance processes. The Konza Prairie Biological Station and Fort Riley have similar hillslope processes with similar elevation range, geologic formations, and soils. A majority of land cover for both areas is grassland; and they experience identical climatic conditions.

The CN_{NDVI} model (Equation 4-2) was initially developed using four small northeastern Kansas grassland watersheds with the assumption that it was a natural system. A detailed description of CN_{NDVI} development was included in Chapter 3. CN_{NDVI} was applied to test the ability of the developed method to capture the hydrological impact of various training intensities. Runoff was estimated with uniform rainfall during the year 2010 to eliminate rainfall effect and to test only CN_{NDVI} predictability. The assumption was made that all eleven periods received similar rainfall (75 mm). Eleven consecutive CN_{NDVI} during 2010 were used for analysis.

$$NDVI\_CN = -0.108 * NDVI + 98 \qquad (4\text{-}2)$$

where *NDVI_CN* is curve number based on NDVI and *NDVI* is NDVI with value ranges 0–255.

In order to compare the effect of low, medium, and high maneuver intensities on runoff, paired watersheds with various training intensities were selected (Figure 4-4). Watersheds 99 (moderate intensity), 310 (high intensity), 361 (low intensity), 185 (high intensity), 262 (low intensity), and 112 (moderate intensity) were selected for analysis. ISCO samplers (Teledyne ISCO, Lincoln, Nebraska) were used to study rainfall-runoff processes at Fort Riley (Figure 4-4). Water samplers were installed at the outlets of five watersheds (one in the northeast, two in the center, and two in the west part of Fort Riley) to record surface runoff from rainfall events based on accessibility and range of training intensity. The watershed runoff contributing area was delineated using ArcGIS based on a 3m DEM (Table 4-2). Paired watersheds of two ISCO-installed watersheds with various intensities of runoff were calculated to study the impact of land cover change due to maneuver damage.

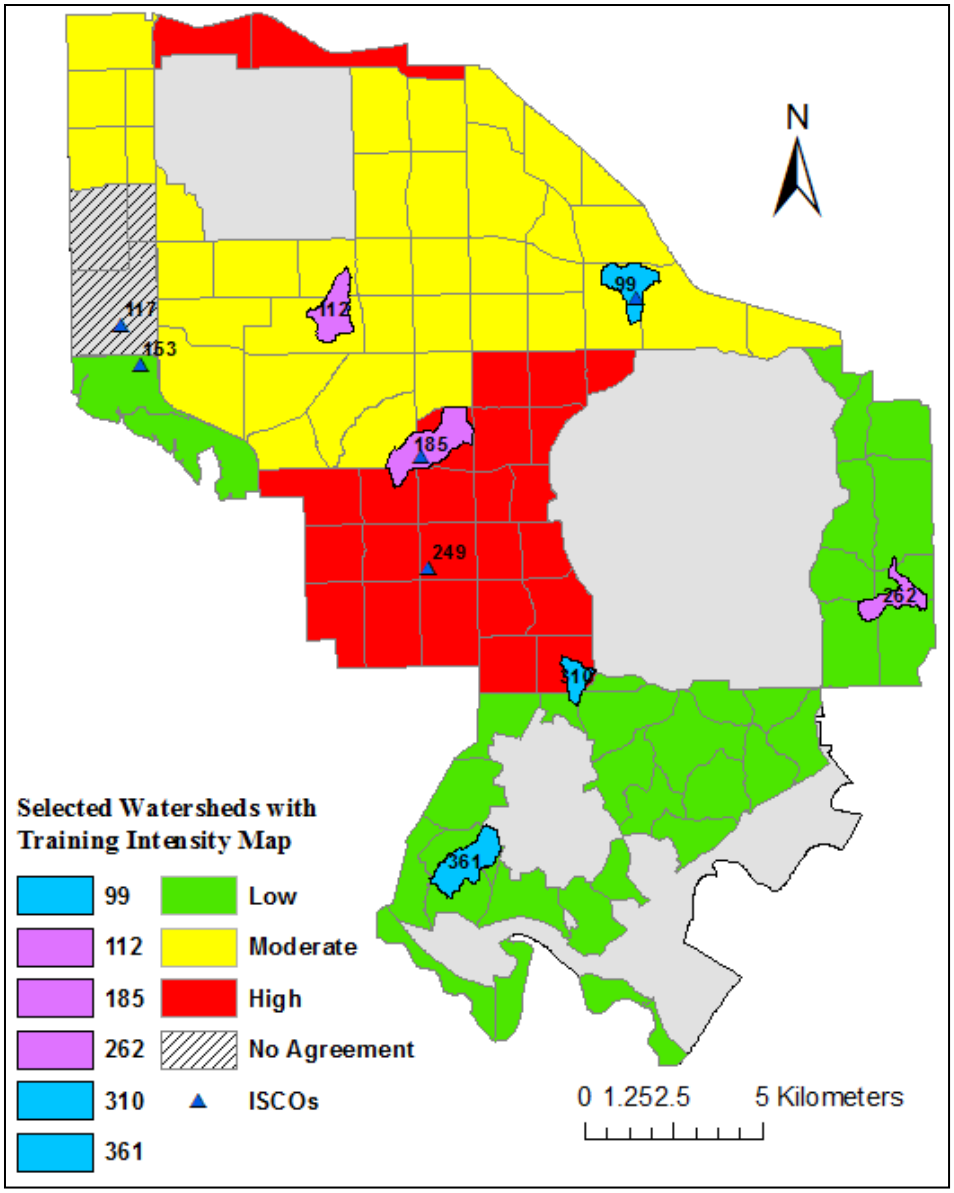


Figure 4-4: Watersheds for analysis embedded with Fort Riley's training intensity map (from P. Denker (pers. Comm.), J. M. S. Hutchinson (pers. Comm.), and Johnson et al., (2010))

Note: Similar colors of watersheds are in the same group for runoff increase comparison due to low, moderate, and high training intensities.

ISCO samplers were installed in the spring of 2012; one ISCO in the west was reinstalled in a different cross section of the nearby area in the spring of 2013 since no runoff was recorded during the period due to very small contributing area for the watershed at the particular cross

section. Stream geomorphological parameters, mainly the cross section (Figure 4-5) and the long profile at the site, were surveyed.

Table 4-2: Watersheds with the training area location delineated using 3-meter DEM, Fort Riley military installation

Watershed	Training Area	Maneuver Intensities	Area (km^2)
99	90	Moderate	0.27
117	65	No Agreement	0.81
153	64	Low	0.13
185	56	High	1.33
249	52	High	1.01

ISCO-recorded runoff was processed and compared to CN_{NDVI} runoff. Manning's method was used to convert stage depth into discharge and estimate the amount of runoff from rainfall events. Manning's equation was used to calculate discharge using the cross section of the stream at the ISCO location, hydraulic radius, and the slope of the channel (Equation 4-3 and 4-4). Figure 4-6 shows the cross sections of small streams where the ISCOs were set up.

$$V = \frac{K_n}{n} R^{2/3} S^{1/2} \quad (4\text{-}3)$$

$$Q = V * A \quad (4\text{-}4)$$

where V is cross-sectional average velocity (m/s), Q is discharge (m^3/s), R is A/P (hydraulic radius (m)), P is wetted perimeter (m), A is the cross-sectional area (m^2), S is the slope of hydraulic gradient (m/m), K_n is 1.49 for English units or 1 for SI units ($\frac{m^{1/3}}{s}$), and n is Manning's resistant coefficient or Manning's roughness coefficient.

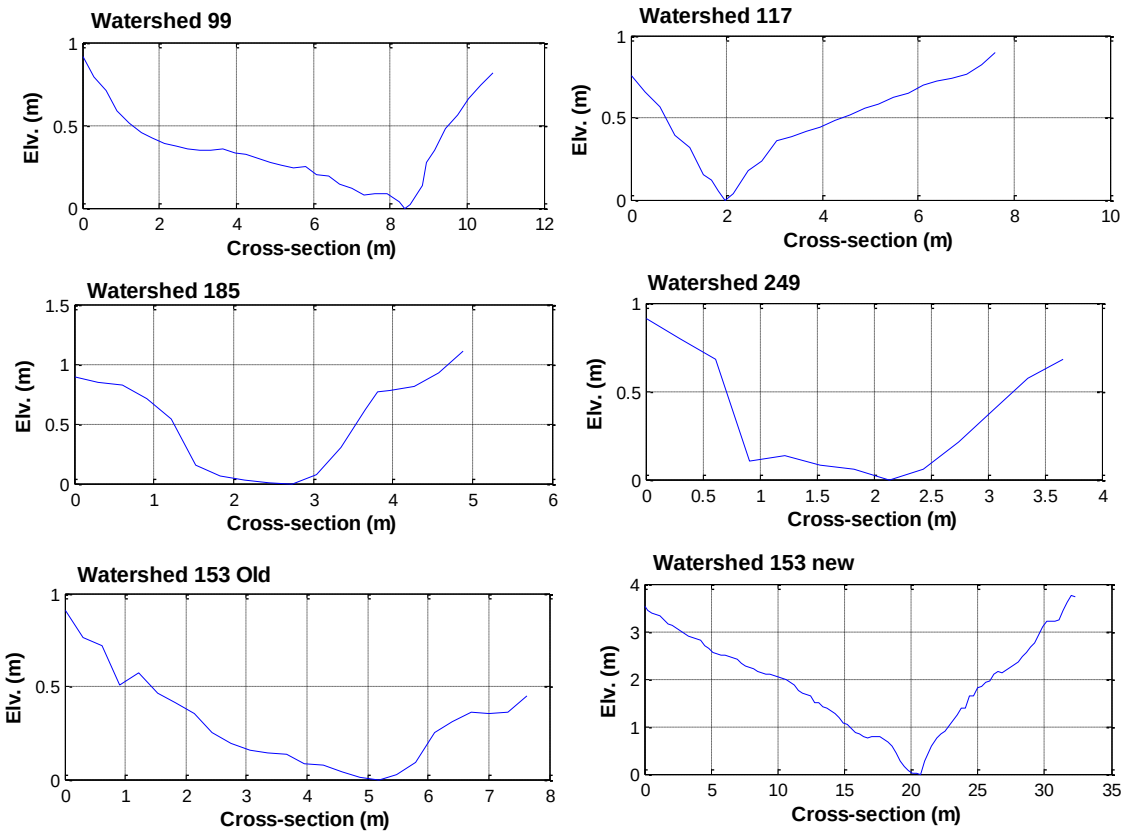


Figure 4-5: Stream cross section at ISCO water sampler sites at small watersheds at Fort Riley

The ISCOs recorded runoff depth during rainfall events every 5 minutes. For each watershed, the relationship between depth versus cross-sectional area at the ISCO site (Figure 4-6) and depth versus discharge (Figure 4-7) were developed by dividing each cross-sectional area into smaller grids in order to capture possible runoff amounts at smaller depths. Cross-sectional area (*A*) was estimated based on cross-sectional survey data. Gridded maps of the cross section were used to estimate the area based on a 0.03048 m (0.1 ft.) progression of depth from the thalweg of the cross-section. The depth versus cross-sectional area mathematical relationship was developed using a two-order polynomial.

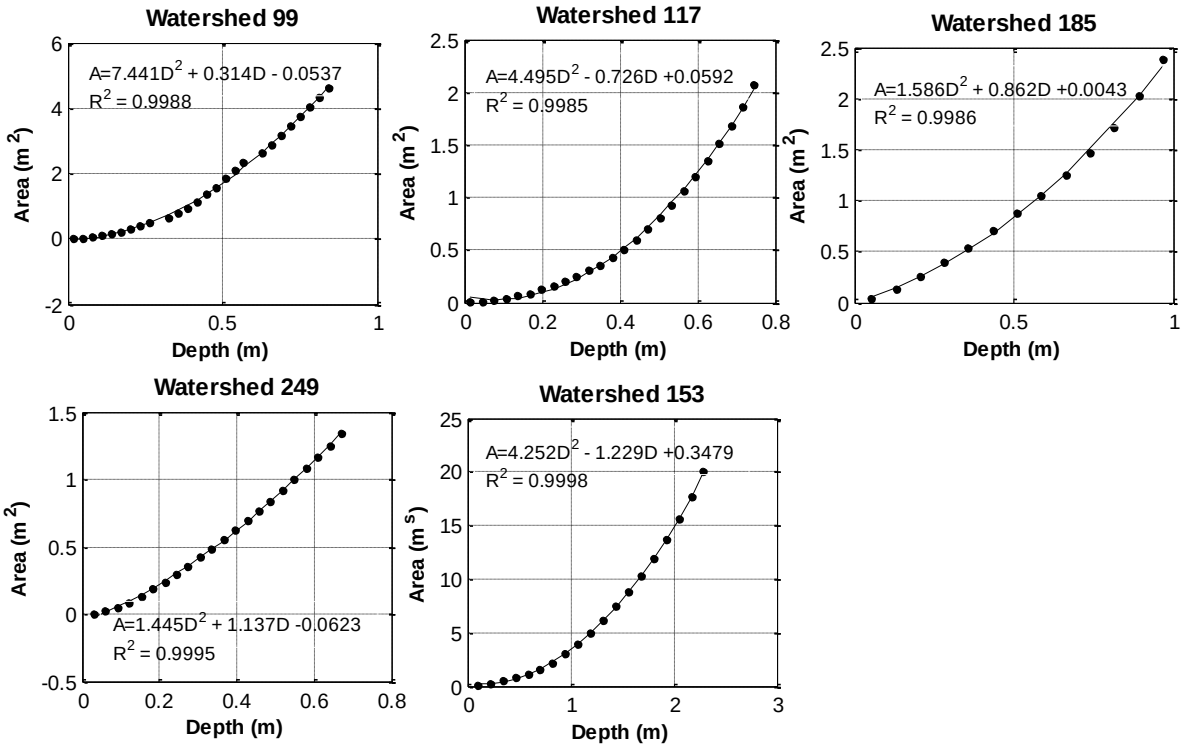


Figure 4-6: Depth versus cross-sectional area at the ISCO water sampler of small watersheds at Fort Riley

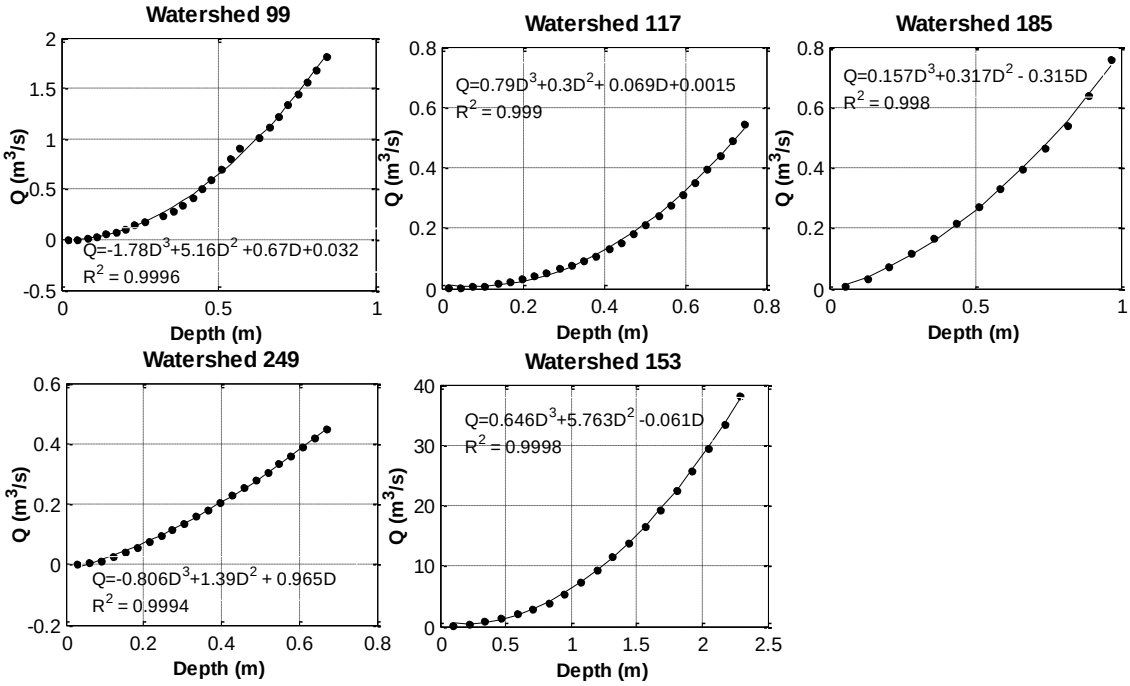


Figure 4-7: Depth versus discharge at each cross section of small watersheds at Fort Riley

4.4 Results and Discussion

Combination of k-means and agglomerative clustering techniques provided a statistical mechanism to conduct paired watershed selection that could improve water quality and quantity studies. The advancement of GIS and remotely sensed data increases the applicability of hydrological studies at various scales. This section describes watershed delineation, watershed parameters selection, paired watershed selection, and application of CN_{NDVI} developed in the previous chapter on a relatively undisturbed system to be applied in a disturbed area with QA analysis.

4.4.1 Watershed Delineation

Three hundred ninety-one (391) watersheds at Fort Riley with average areas of 1 km^2 were delineated in this study. The small area of watershed captured homogeneous training intensities and landcover change/damage (Figure 4-8). Most watersheds were located in one of the training areas or in similar training intensities, improving the use of watersheds as replicates in a paired watershed assessment.

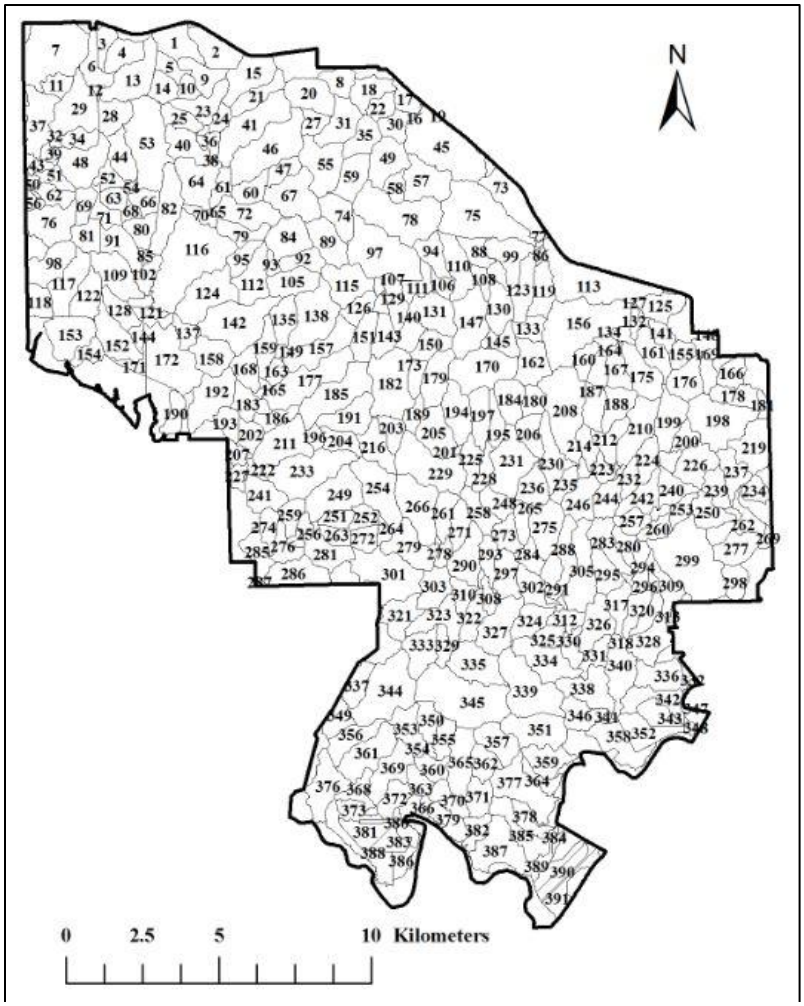


Figure 4-8: Delineated watersheds using hydrology tool in ArcGIS for paired watershed analysis

4.4.2 Topographic Parameters Selection

Figure 4-9 shows histograms of eight parameters that represent the distribution of data. Most parameters were not normally distributed and skewed. Martinez and Martinez (2008) advised researchers to be aware of distances that could be affected by magnitude and scale of variables used in clustering. Although curvature was distributed around mean value of zero, most values for percent flat area were greater than 40%.

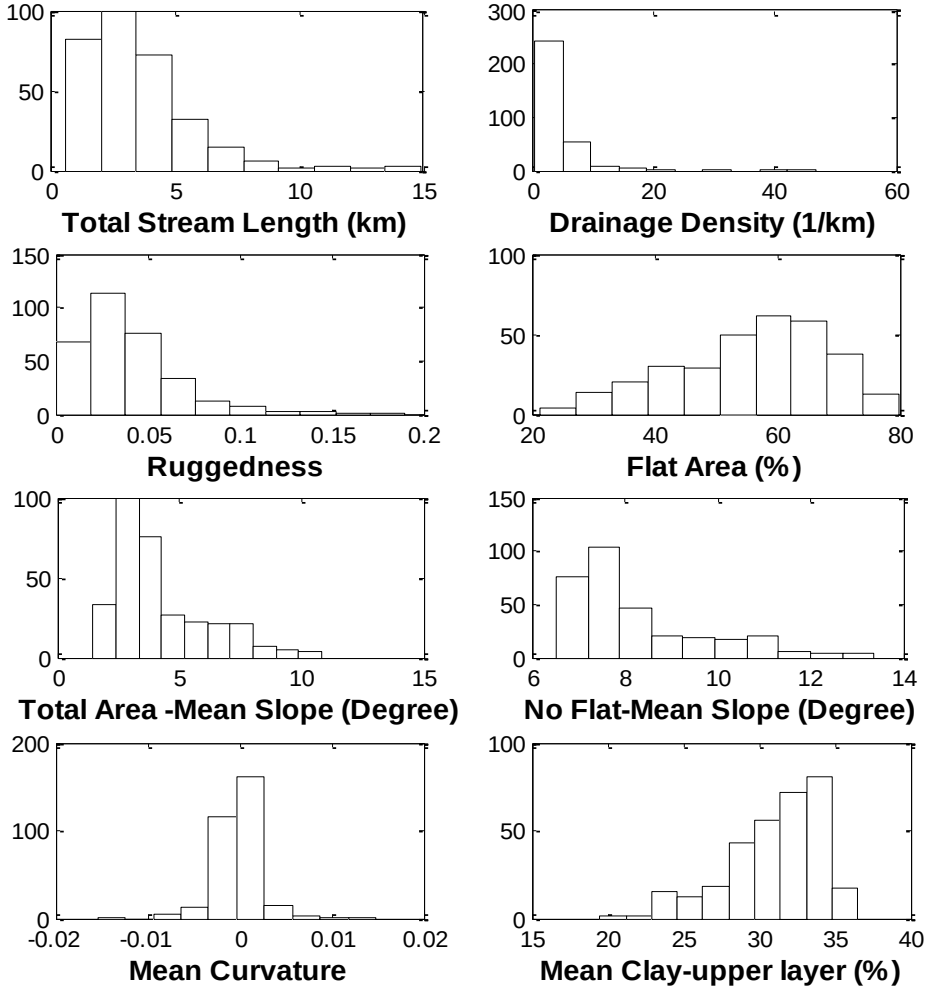


Figure 4-9: Histogram of eight variables used to select the paired watershed

Figure 4-10 shows boxplots of selected eight parameters that display summary values of minimum, maximum, median, and quantiles. Based on the figure, the parameters were not normally distributed and had a wide range of values with extreme values. The box plot also highlights outliers in each variable.

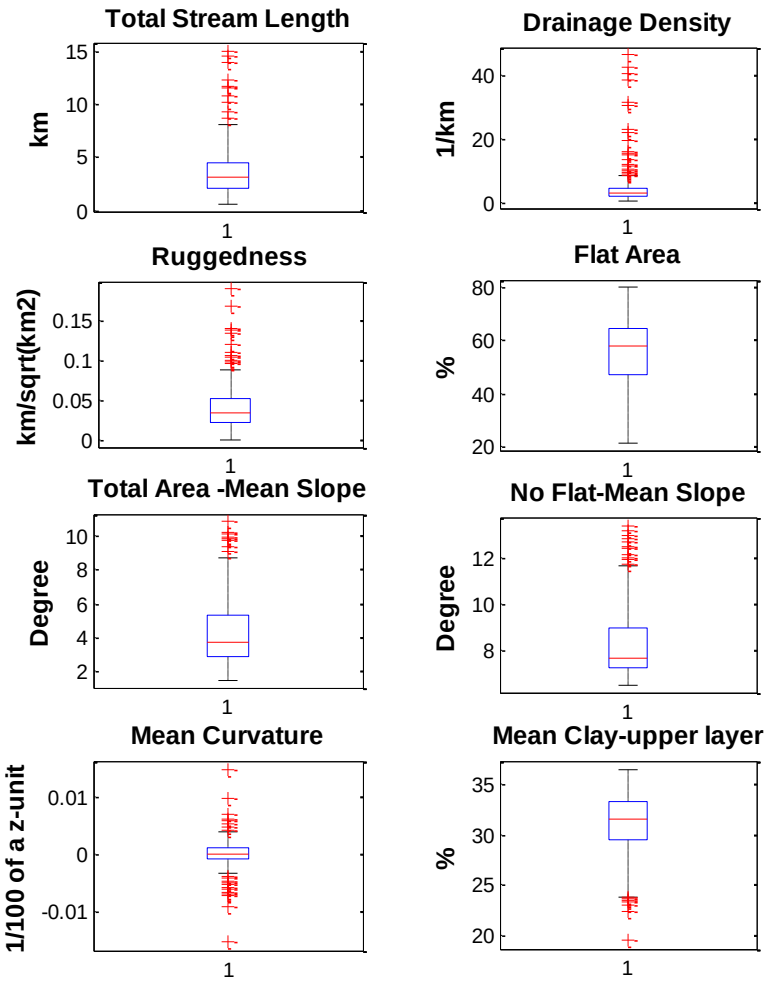


Figure 4-10: Box and whisker plots of watershed variables used in analysis

Table 4-3 summarizes statistics of raw and z-scored data of eight variables used in this analysis. The statistics show that variability was reduced compared to the raw data so that each variable equal role in statistical paired watershed selection. Z-score allows each variable to be represented when determining paired watersheds.

Table 4-3: Summary statistics of raw and z-scored values of eight variables used to select paired watershed

	Raw Data					Z-scored Data				
Variable	Min.	Max.	Mean	Std. Dev.	Range	Min.	Max.	Mean	Std. Dev.	Range
Total Length (km)	0.61	14.89	3.62	2.33	14.28	-1.29	4.84	0.00	1.00	6.13
Drainage Density (1/km)	0.50	46.67	4.66	5.72	46.17	-0.73	7.34	0.00	1.00	8.07
Ruggedness (km/sqrt(km2))	0.00	0.19	0.04	0.03	0.19	-1.43	5.43	0.00	1.00	6.86
Percent Flat Area (%)	21.41	79.89	55.68	12.38	58.48	-2.77	1.95	0.00	1.00	4.72
SlopeTotal Mean (Degree)	1.46	10.83	4.23	1.91	9.37	-1.45	3.45	0.00	1.00	4.90
Slope No Flat Mean (Degree)	6.52	13.37	8.28	1.47	6.85	-1.20	3.47	0.00	1.00	4.67
Mean Curvature (1/100 z units)	-0.02	0.01	0.00	0.00	0.03	-6.29	6.09	0.00	1.00	12.39
Soil Clay (%)	19.50	36.50	31.11	3.06	17.00	-3.79	1.76	0.00	1.00	5.56

The pattern of observations must be preserved when standardizing parameters. Figure 4-11 shows plots of each parameter before and after z-scoring of the observations. As shown, the z-scores did not change the patterns of observation; rather the scores standardized the dataset to alleviate domination of one or more parameter/s in the clustering process. The z-score values mimic data distribution of the raw data and improve the ability to compare variables.

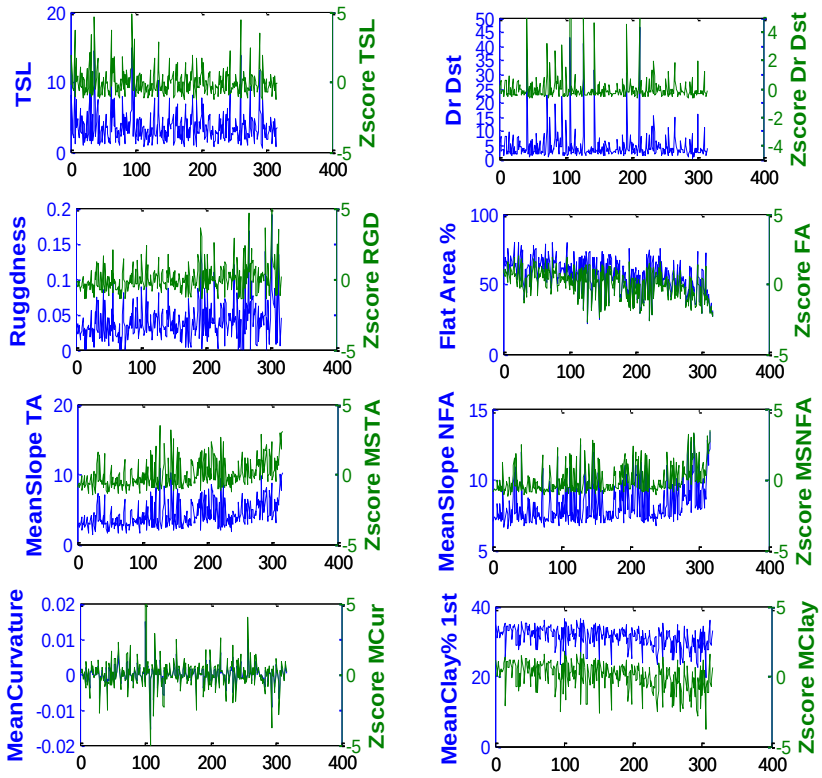


Figure 4-11: Plot of eight variables used to select the paired watershed

Figure 4-12 shows raw and z-scored data differences on two variables used for paired watershed selection analysis (total stream length and percent clay in the upper soil layer). Total stream length has an average value of 3.62 km (most values are below 5), but the percent clay average is 31.11% (most values are over 25 %). Therefore, when using the raw data, percent clay could have more to determine distances of k-means and agglomerative hierarchical statistics. The z-scored plots show that the two variables were within a similar range.

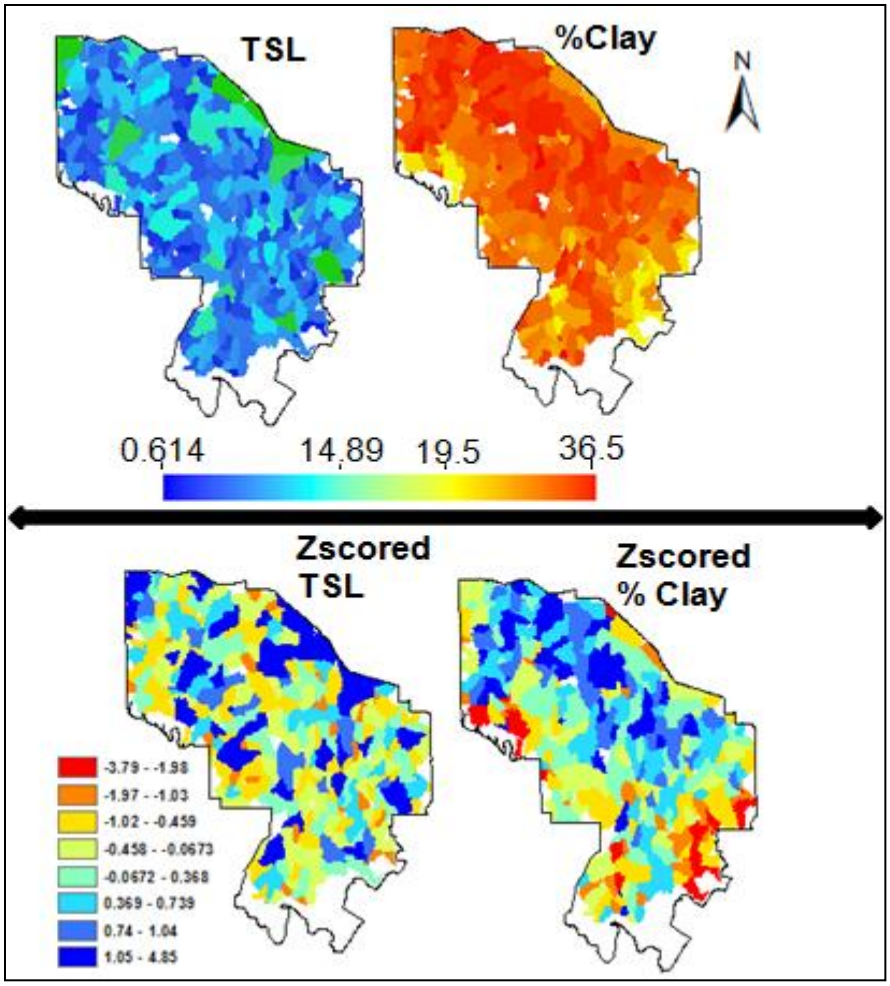


Figure 4-12: Data range comparison of raw data (upper) and z-scored standardized data (below) of total stream length and percent clay in upper soil layer

The correlation matrix technique explores whether or not dependencies are present in the dataset. Identifying nondependent variables is essential for successful clustering. Figure 4-13 shows the correlation matrix containing Pearson's correlation coefficients that helped visually identify the correlation of parameters. The correlation coefficients showed no strong positive or negative correlations between respective variables except total area slope and percent flat area. However, neither of the two variables showed identical correlation to the remaining variables, so all eight variables were included in the analysis.

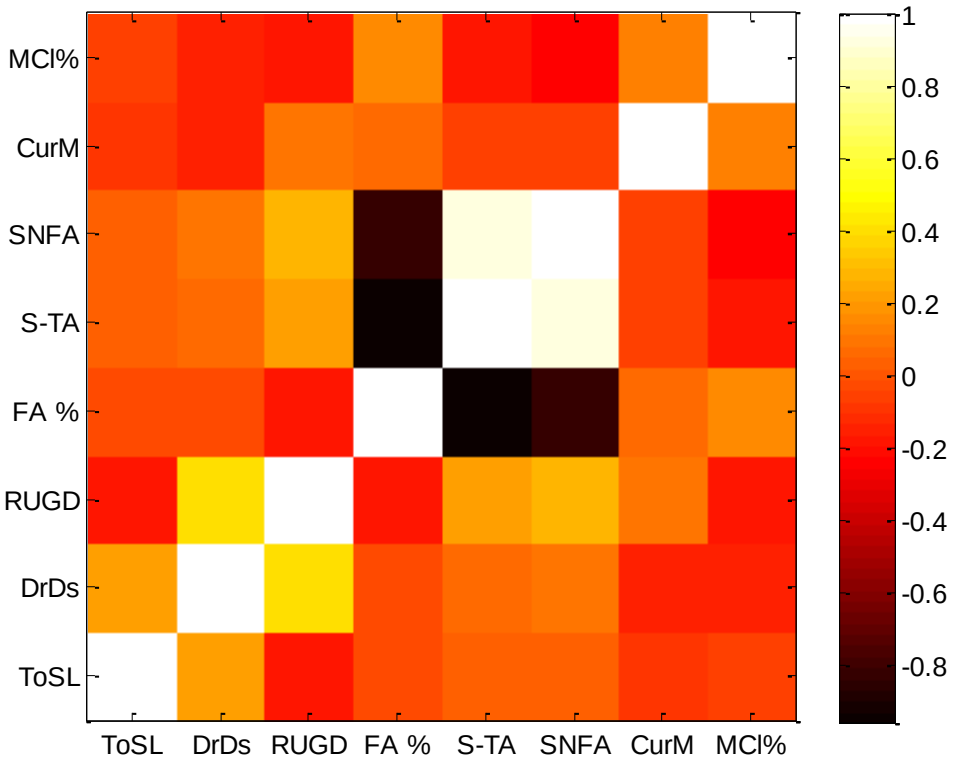


Figure 4-13: Correlation matrix containing Pearson's correlation coefficients for each pair of variables

Note: Light colors represent strong positive linear correlations; dark colors designate negative correlations. The range of values in orange suggests no correlation. (ToSL = total stream length, DrDs = drainage density, RUGD = ruggedness, FA% = percentage of flat area, S-TA = total area slope, SNFA = non-flat area slope, CurM = mean curvature, and MCl% = percentage of clay in the upper soil layer).

Variance-covariance matrix of the parameters showed similar results to Table 4-4. The diagonal of the table, which is bold and italicized, shows variances of the regression coefficients of each parameter, whereas the off-diagonal values are the covariance between the respective combined parameters. Mean curvature and ruggedness showed less variability, while percent flat area and drainage density had the most variability. Covariance values showed dependence between parameters, but the dependence was only in certain variables. Therefore, all parameters were included in the cluster analysis.

Table 4-4: Variance-covariance values of the data matrix of eight variables used in clustering analysis

Covariance MATRIX	Total Stream Length	Drainage Density	RUGN	% Flat Area	Mean Slope	Mean Slope No-Flat Area	Mean Curvature	Mean Clay 1st layer %
Total Stream Length (km)	***5.426***	2.915	-0.012	-1.104	0.130	0.170	-0.001	-0.455
Drainage Density (1/km)	2.915	***32.738***	0.065	-2.718	0.646	0.909	-0.002	-2.762
Ruggedness	-0.012	0.065	***0.001***	-0.063	0.012	0.011	0.000	-0.014
% Flat Area	-1.104	-2.718	-0.063	***153.335***	-22.667	-14.743	0.002	6.269
Mean Slope	0.130	0.646	0.012	-22.667	***3.658***	2.624	0.000	-0.971
Mean Slope No-Flat Area	0.170	0.909	0.011	-14.743	2.624	***2.156***	0.000	-1.007
Mean Curvature	-0.001	-0.002	0.000	0.002	0.000	0.000	***0.000***	0.001
Mean Clay 1st layer %	-0.455	-2.762	-0.014	6.269	-0.971	-1.007	0.001	***9.364***

4.4.3 Clustered Watersheds

4.4.3.1 K-means Clustering

The optimum number of groups was determined based on the Mojena- upper tail rule and silhouette statistics. As shown in Figure 4-14, many estimates were possible for the number of clusters. However, a distinct change occurred in raw fusion levels at 6, producing an elbow effect. Elbows at 9 and 11 provided other options of clusters; however, 6 clusters is supported by silhouette statistics (see below).

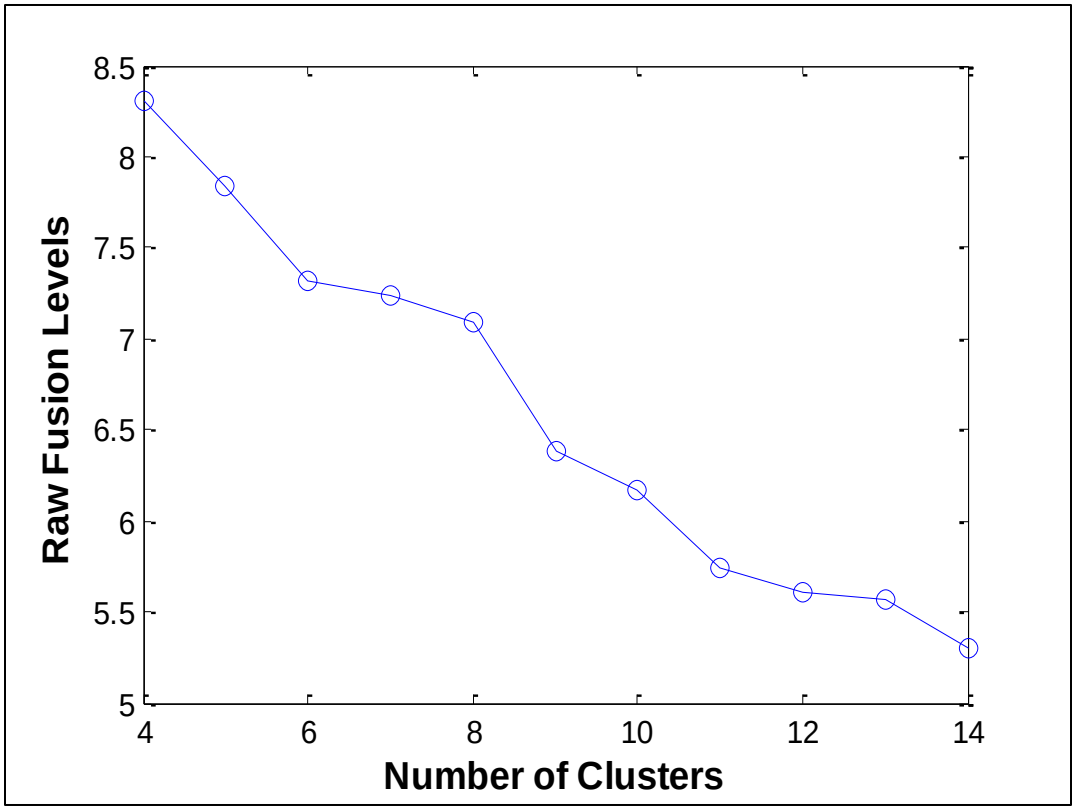


Figure 4-14: Raw fusion levels from complete data linkage to select the number of clusters in K-means clustering

Note: A distinct elbow effect is shown at 6 clusters; other possible elbows are 9 and 11.

In addition to the Mojena plot based on raw fusion levels, silhouette statistics can also provide the optimum number of clusters (Kaufman and Rousseeuw, 1990; Martinez and Martinez, 2008; MathWorks, 2012). Possible silhouette values range from -1 to 1, with greater silhouette values suggesting a better cluster (Figure 4-15). Silhouette statistics for this study suggested that 6 was the optimum number of clusters, as supported by the Mojena plot above. The range of mean silhouette value for various clusters ranged from 0.26 to 0.42. The overall average silhouette value for six clusters was 0.42, which was highest compared to other cluster numbers. Silhouette values vary little with repeated calculations, especially for cluster numbers greater than 10, indicating that an increased number of clusters would not improve the relationship. Therefore, the data were grouped into 6 clusters using k-means clustering.

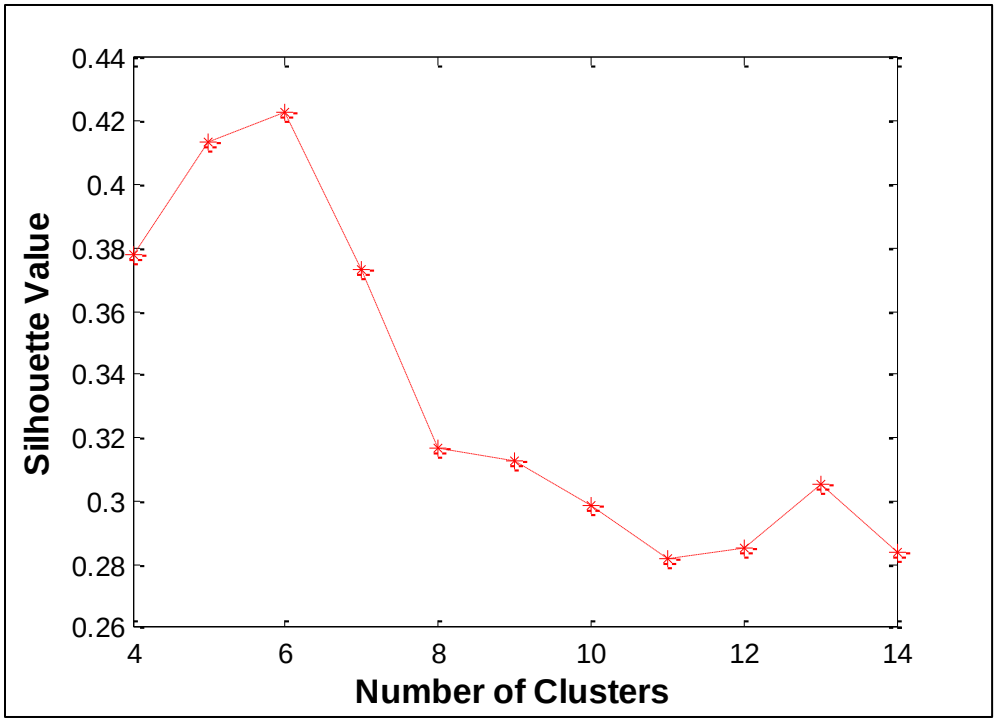


Figure 4-15: Silhouette values plot for 4–14 clusters using the eight watershed variables

A total of 316 watersheds were grouped into six distinct clusters. The number of watersheds in each cluster ranged from 29 to 69 (Table 4-5). Spatial distribution of the six clusters obtained from k-means clustering at Fort Riley is provided in Figure 4-16. A watershed's cluster can be identified by the watershed color.

Table 4-5: Number of watersheds in each cluster obtained by k-means technique

Custer	Watersheds
1	51
2	29
3	36
4	65
5	67
6	68
sum	316

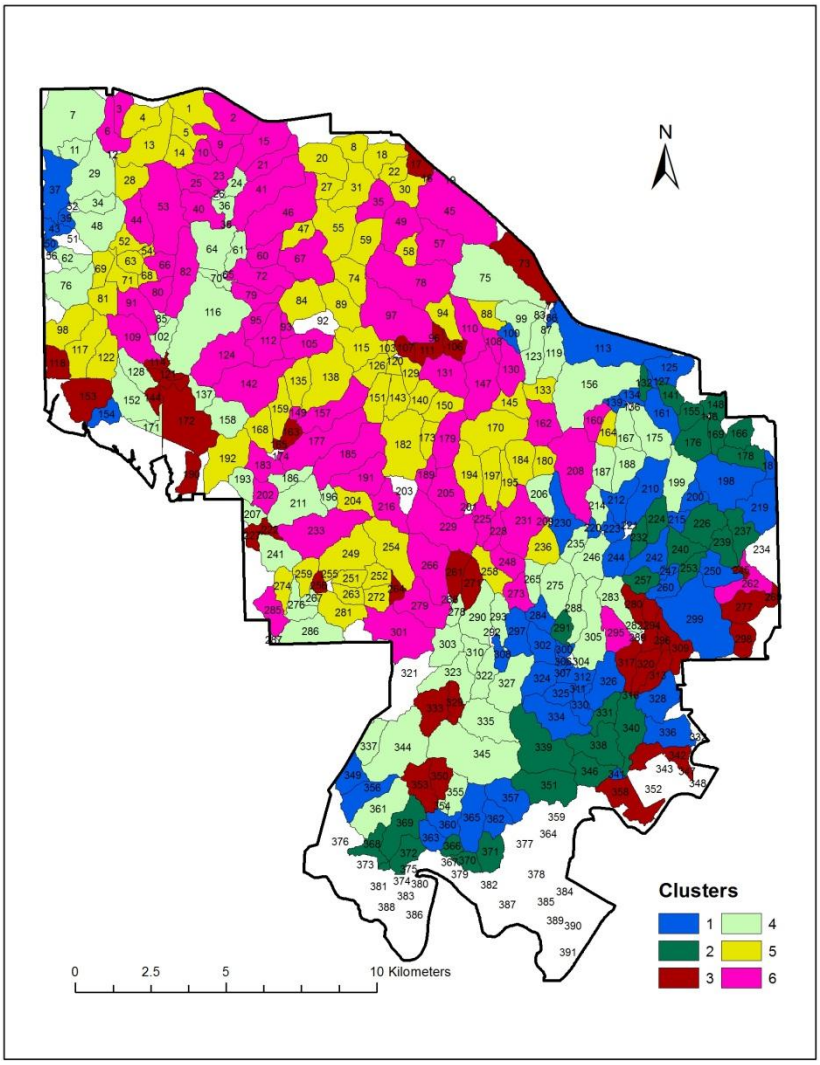


Figure 4-16: Map of watersheds at Fort Riley with specified clusters obtained by k-means clustering

Scatter plots in Figure 4-17 show possible combinations of eight parameters used in paired watershed selection of the six clusters. These scatter plots show that each parameter contributed to the formation of clusters even though visual identification of the contribution of individual parameters in cluster computation with many variables is difficult.

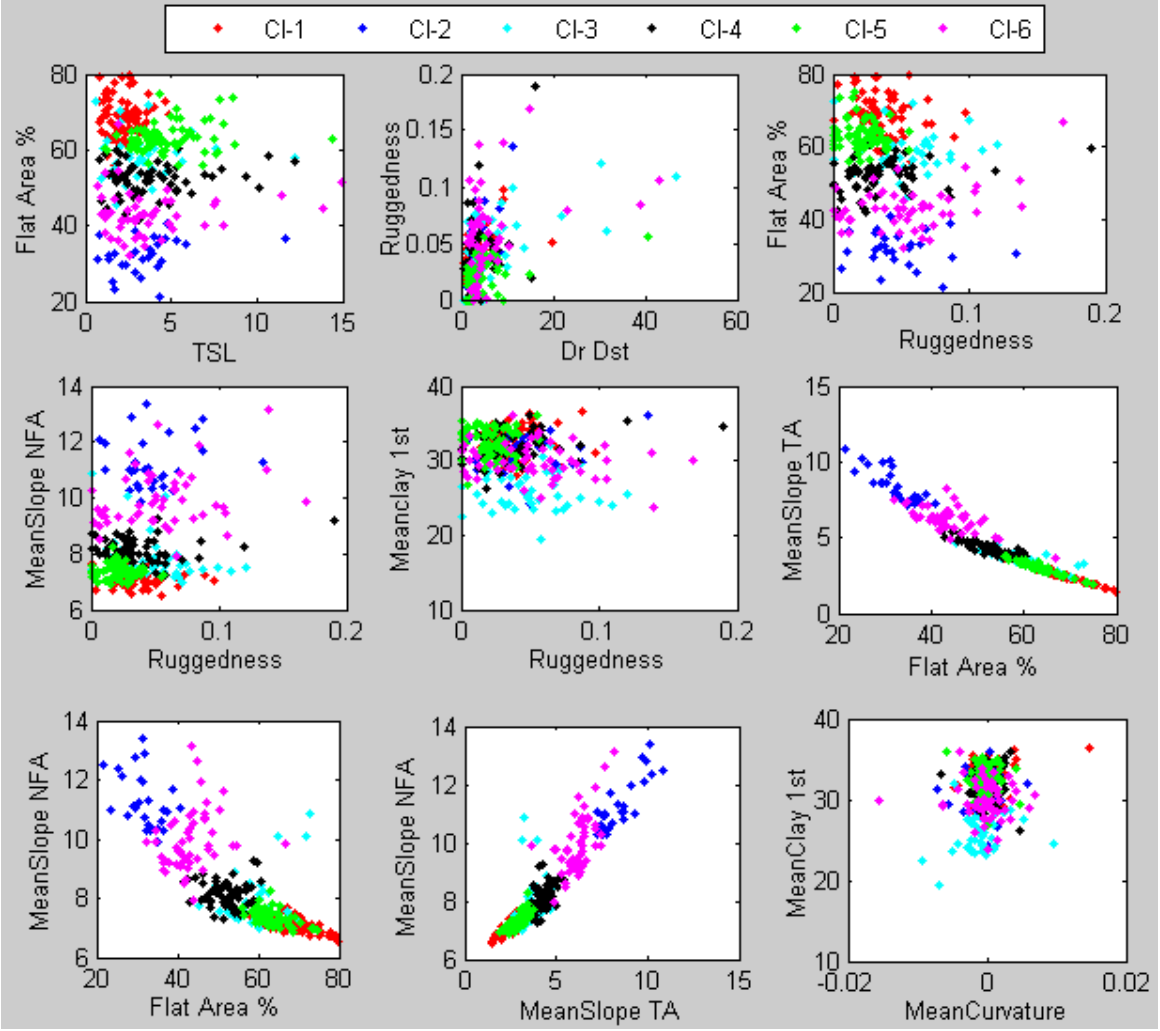


Figure 4-17: Map of clusters based on two possible parameter combinations of Fort Riley watersheds in analysis

4.4.3.2 Agglomerative Hierarchical Technique

The agglomerative hierarchical method was used for each cluster of watersheds obtained by k-means clustering in order to pair watersheds. Cophenetic correlation coefficients, the standard measure of how well a hierarchical cluster fits the data, were greater than 0.80 for six of the clustered groups. This proved valid hierarchical clustering (Table 4-6), which measured the linear correlation coefficient between cophenetic distances from the hierarchical tree of watersheds and the original distance used to construct the tree based on the centroid of each cluster of watersheds (MathWorks, 2012). Based on obtained cophenetic correlation coefficients

(Table 4-6), cluster 6 was selected as the best paired watersheds composed group. Cluster 6 watersheds are also clearly shown in terms variable contributions in Figure 4-17. Watersheds were selected from this cluster for runoff analysis.

Table 4-6: Cophenetic correlation coefficient of agglomerative hierarchical clustering of each cluster

Watersheds	Cophenetic Correlation Coefficient
Cluster 1	0.9044
Cluster 2	0.8404
Cluster 3	0.8703
Cluster 4	0.8159
Cluster 5	0.9168
Cluster 6	0.9341

The dendrogram for cluster 2 watersheds in Figure 4-18 indicates the hierarchical relationship of individual watersheds. The key for watershed ID and the dendrograms of the other five clusters are attached in Appendix B. The dendrogram also indicates possible small clusters within the dataset. The y-axis shows matched watersheds in hierarchy based on clustering distance. Based on the hierarchy of the dendrogram, possible paired watersheds can be identified. For example, denodrogram ID 9 and 16 (watershed ID 224 and 257) are similar watersheds (Appendix for matching dendrogram and watershed IDs). The height of the hierarchy represents the distance linkage computed between watersheds. High hierarchy indicates less similarity between watersheds. Appendix B shows the remaining dendrogram of the agglomerative analysis and the matched table of IDs in denedogram and watersheds.

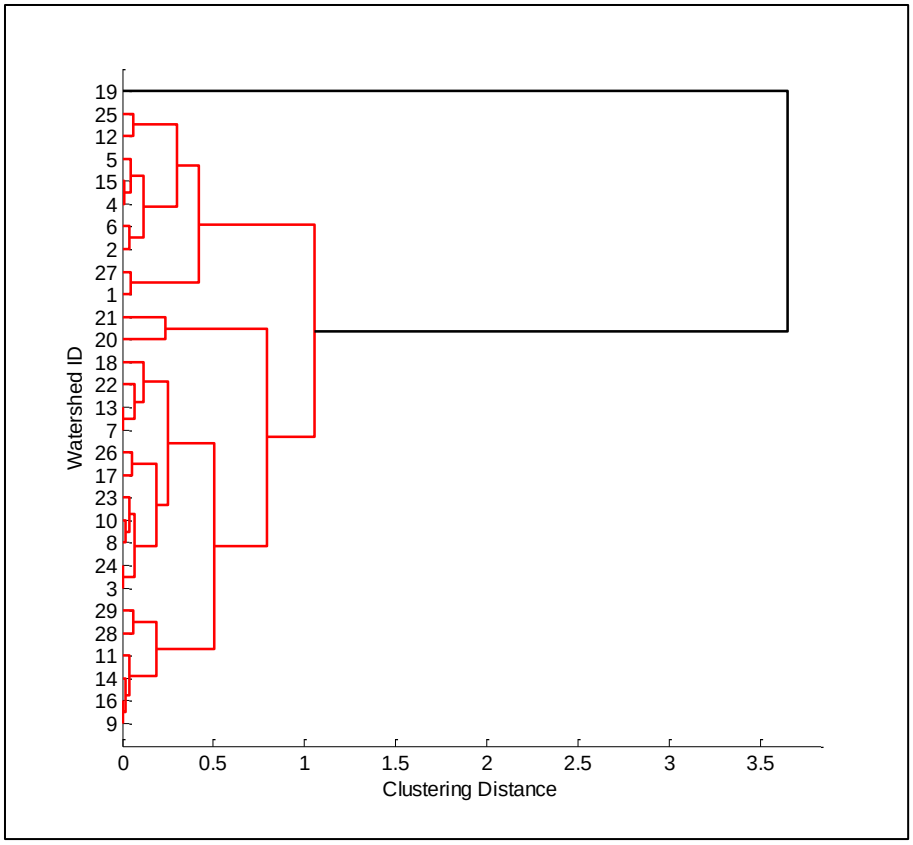


Figure 4-18: Hierarchical matches of watersheds in one cluster

4.4.4 Application of CN_{NDVI} for Runoff Estimation

Initially, CN_{NDVI} modeled runoff data was assumed to be compared to actual data gathered by ISCO water samplers installed at five small watersheds outlets. However, no significant runoff was recorded during the study period. In fact, one ISCO in the west area was reinstalled in a different cross section because no runoff was recorded due to very small contributing area for the watershed at the particular cross section. The assumption was made that the very small sizes of watersheds, low sensitivities of ISCO samplers, and low rainfall events caused insufficient runoff. Due to the low number of observations, all watershed data were combined for analysis. ISCO-recorded runoff at Fort Riley from watersheds versus runoff based

on CN_{NDVI} is shown Figure 4-19; the graph also shows how it varies with SCS-CN modeled runoff.

Figure 4-20 shows that most CN_{NDVI} runoff was higher than ISCO-recorded values, especially low flows. However, CN_{NDVI} model runoff was similar to standard SCS-CN runoff, although the coefficient of determination was higher in SCS-CN due to similar land cover in both conditions since the watersheds were very small. Low-resolution NDVI was unable to accurately capture spatial variation in land cover. Watersheds were very small and had few NDVI grids because NDVI was low resolution (231 m). So it is expected that if very high spatial resolution NDVI acquired a better result might be seen. Nevertheless, the CN_{NDVI} accounted for 72% of observed runoff even though only one value had significant influence. The ISCOs had different issues during the collection period.

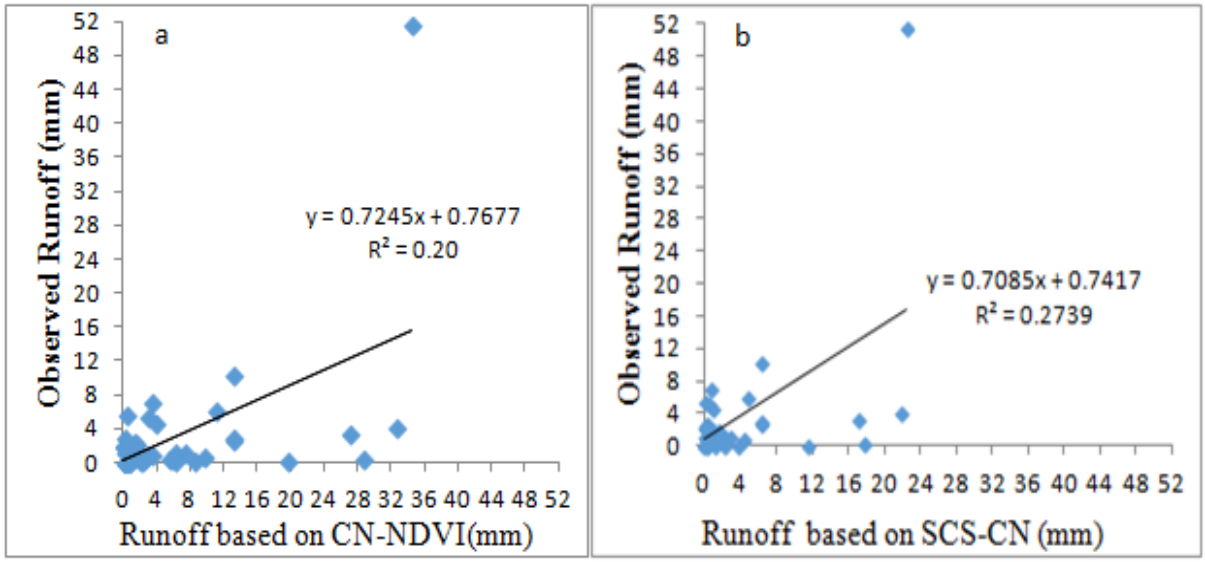


Figure 4-19: Runoff comparison of ISCO recorded at small watersheds at Fort Riley: a) CN_{NDVI} and b) SCS-CN

In addition to uncertainties in runoff, rainfall measurements and the inability to accurately predict low flows may also be related to the initial abstraction used in CN_{NDVI} development. The original SCS I_a/S ratio value of 0.2 may not able to represent the actual

condition, thereby requiring a higher value, especially when significant military impacts are present, including removal of vegetation, soil compaction, and gullies. These effects may need to be included in initial abstraction.

Temporal pattern in CN_{NDVI} over 16-day NDVI acquisition periods in 2010 is given in Figure 4-20. The maps show that CN_{NDVI} captured variability of the rainfall-runoff relationship by capturing spatiotemporal variability in CN_{NDVI}.

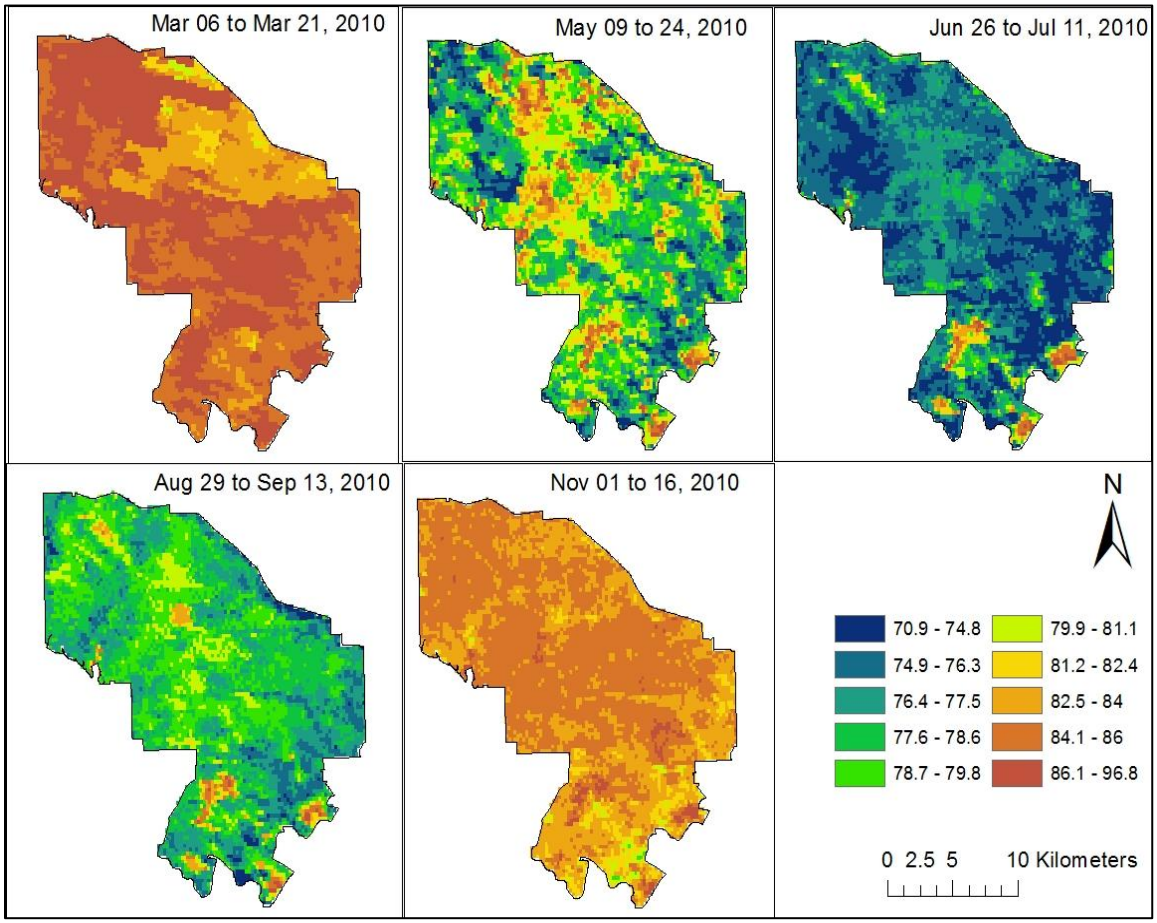


Figure 4-20: NDVI-CN based on the 16-day NDVI period in 2010

The pixel reliability rank and VI quality QA assessment was conducted for NDVI imagery used in runoff of analysis and spatiotemporal mapping of CN_{NDVI} (Figure 4-21). As shown in the figure, a majority of March 6–21, 2010, pixel reliability ranking imagery was under

marginal data (useful but look other QA), and the upper part of Fort Riley May 9–24, 2010, imagery pixel reliability maps were partly marginal data and cloudy.

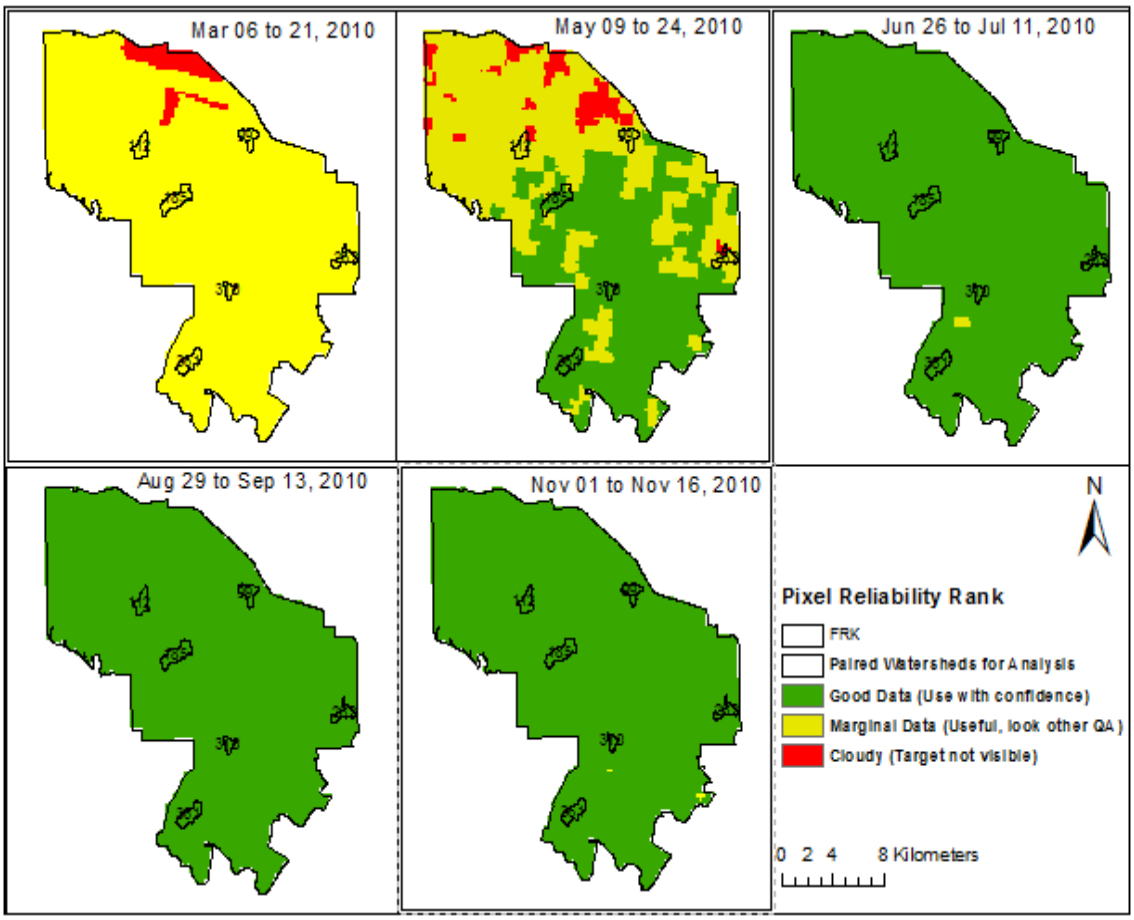


Figure 4-21: Pixel reliability ranking of MODIS VIs (MOD13Q1) at Fort Riley

Furthermore, VI quality QA analysis showed that March 6–21 and May 9-24, 2010, imagery had multiple values with and without clouds, resulting in decreased VI quality or poor VI quality (Figure 4-21). However, VI quality QA results showed that a majority of the area had NDVI products without intermediate and mixed clouds. In addition, both images satisfactorily showed the different growing period (March 6-crop emerging or no crop period; May 09-developed crop).

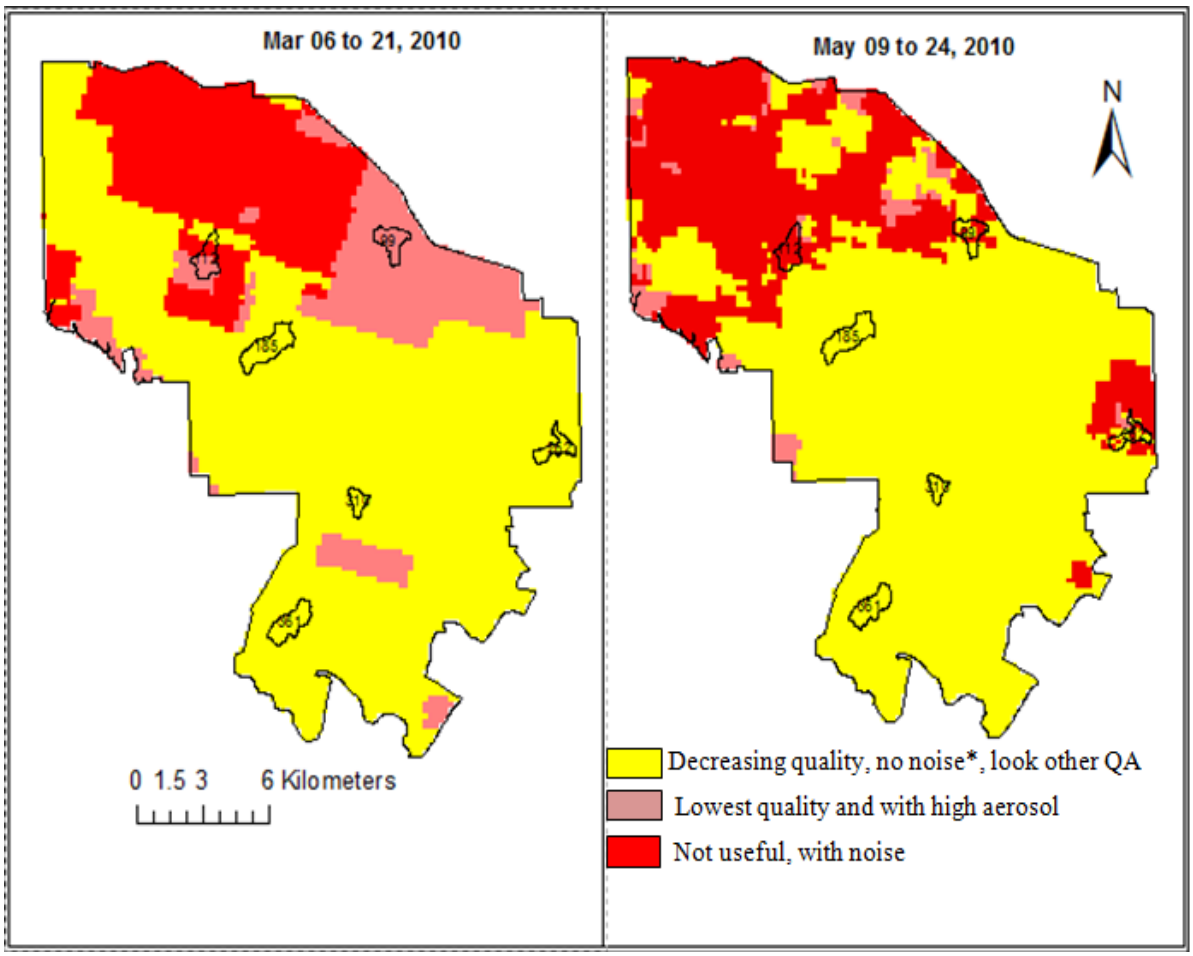


Figure 4-22: VI quality of MODIS (MOD13Q1) at Fort Riley watersheds

Note: noise * is aerosols, adjacent and mixed clouds, possible snow/ice, and shadow. VI quality summary description is included in Appendix A.

The map shows that CN_{NDVI} decreased from March to July and increased during the remaining periods. A majority of land surface had higher CN_{NDVI} values during the non-growing season due to low NDVI values, potentially leading to lower interception and infiltration and higher runoff potential. During growing season, increases in vegetation cover resulted in decreased CN_{NDVI} (May, June, and August), potentially representing higher ET and infiltration and thus lower runoff.

CN_{NDVI} also showed the effect of training intensity (Figure 4-23 Figure 4-24). Figure 4-24 shows the error bars of low, medium, and high training intensity categories at Fort Riley and

the Konza Prairie for the 12 years averaged for each time period. As shown in the figure, low military intensity closely matched the Konza Prairie, especially during peak growing period. High military intensity and resulting high CN showed more impact on vegetation during peak season. Wide error bars in higher intensity also showed potential variation of runoff through various CN_{NDVI} for this area.

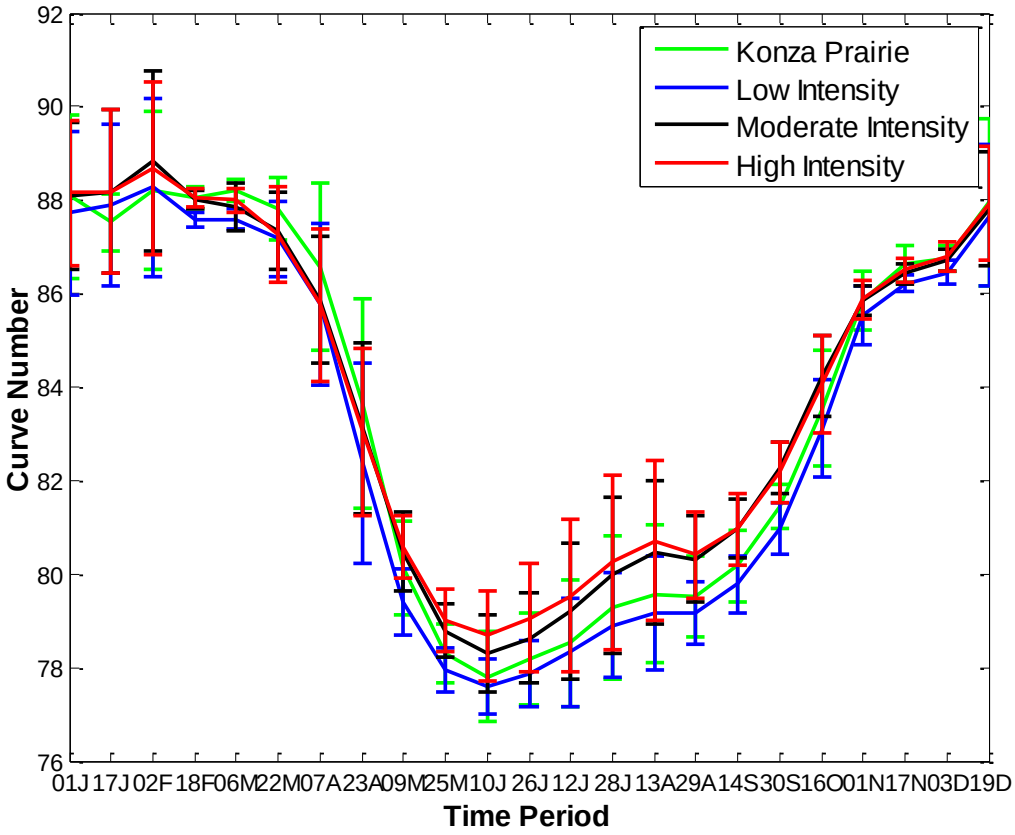


Figure 4-23: Error bar of CN based on 23 NDVI acquisition time periods for low, medium, and high training intensities at Fort Riley

The spatial (different training intensities) and temporal (each NDVI acquisition period) variability of CN_{NDVI} showed significant differences (p = 0.001, Table 4-7). The temporal variability pattern among the four areas (three intensities at Fort Riley and Konza) was similar, as shown in Figure 4-23. Spatial variability was not as wide as temporal variability, although the training intensity and Konza CN were lower in general. Various factors caused the low variability, including that the CN was averaged for relatively wider areas, all areas were in

similar land use/cover category (primarily grasslands), and NDVI coarse spatial resolution may not capture accurate training intensity effects. However, the runoff potential from a possible rainfall event; 1 or 2 CN differences could cause significant increase in the runoff potential. The pairwise comparison of various intensities was conducted and low intensity CN showed significant difference (p-value < 0.0001) from moderate and high intensity CNs; however, no statistically significant difference was evident between the moderate and high intensity areas.

Table 4-7: ANOVA statistical output of CN based on training intensity and NDVI acquisition period

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	25	17289.80685	691.59227	450.99	<.0001
Error	1078	1653.09509	1.53348		
Corrected Total	1103	18942.90194			

R-Square	Coeff Var	Root MSE	CN Mean
0.912733	1.511634	1.238339	81.92055

Source	DF	Anova SS	Mean Square	F Value	Pr > F
Period	22	17189.01833	781.31901	509.51	<.0001
Intensity	3	100.78852	33.59617	21.91	<.0001

Detailed boxplots of CN_{NDVI} for each 16-day NDVI period for high, moderate, and low intensity areas at Fort Riley (Figure 4-24) show that CN_NDVI variability during the growing season was lower compared to moderate and high intensity areas, especially during growing season. A wider range of CN_{NDVI} for moderate and high intensities during July and August relatively than low intensity areas was attributed to changes in land cover due to military maneuvers.

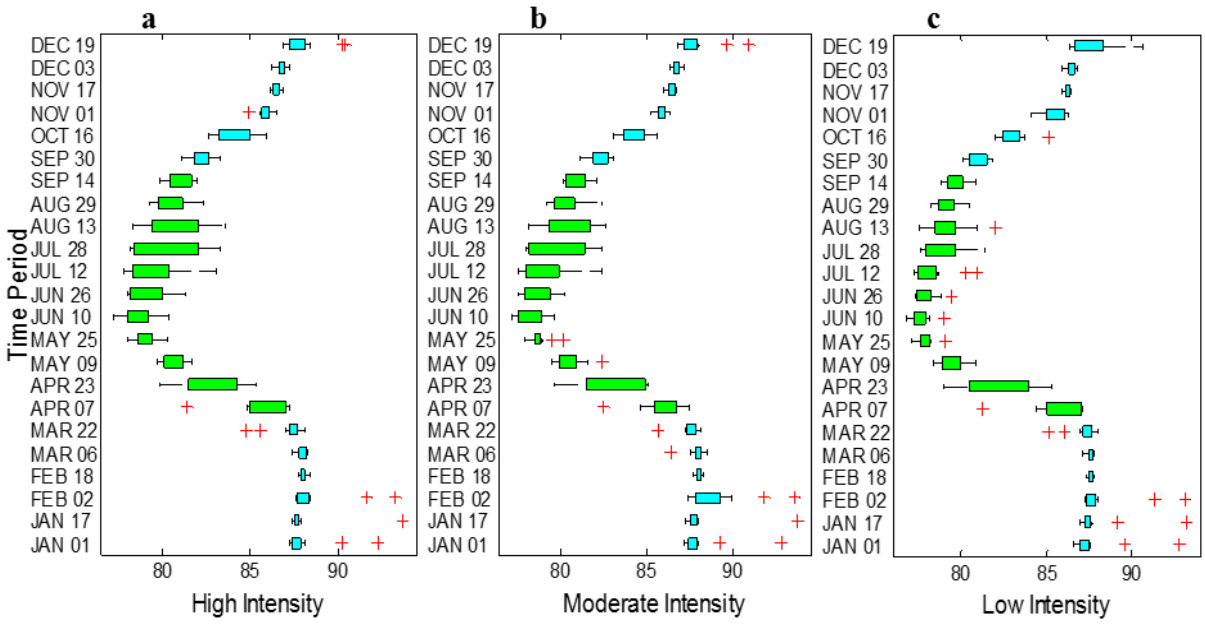


Figure 4-24: Box plots of CN_{NDVI} for each 16-day NDVI period for three types of intensities

In order to compare the effects of training intensities on runoff generation, the percent runoff changes of moderate and high training intensity watersheds in relative-to-low training intensities were estimated as described in methods section. The percent runoff increases from low to moderate intensity and high intensity for selected watersheds were assessed (Figure 4-25). The percent increase was calculated by taking low intensity runoff as a benchmark so that the amount of percent increase was computed for moderate and high intensity. In general, for both watershed combinations, the percent runoff increase from low intensity to high intensity was higher than from low to moderate intensity. Because a uniform rainfall event was used, this result showed that runoff increased with increasing intensity, potentially creating significant erosion potential. The effect was higher during the growing period, clearly demonstrating the influence of vegetation on runoff. High intensity maneuvers cause higher loss of vegetation, resulting in lower interception and infiltration by creating a micro environment in root zone vegetation that increases infiltration and leads higher runoff. Although VI quality QA maps showed some cloud contamination (Figure 4-22), the study watersheds in other NDVI periods in

analysis were under good data (use with confidence) classification using pixel reliability ranking imageries (Figure 4-21).

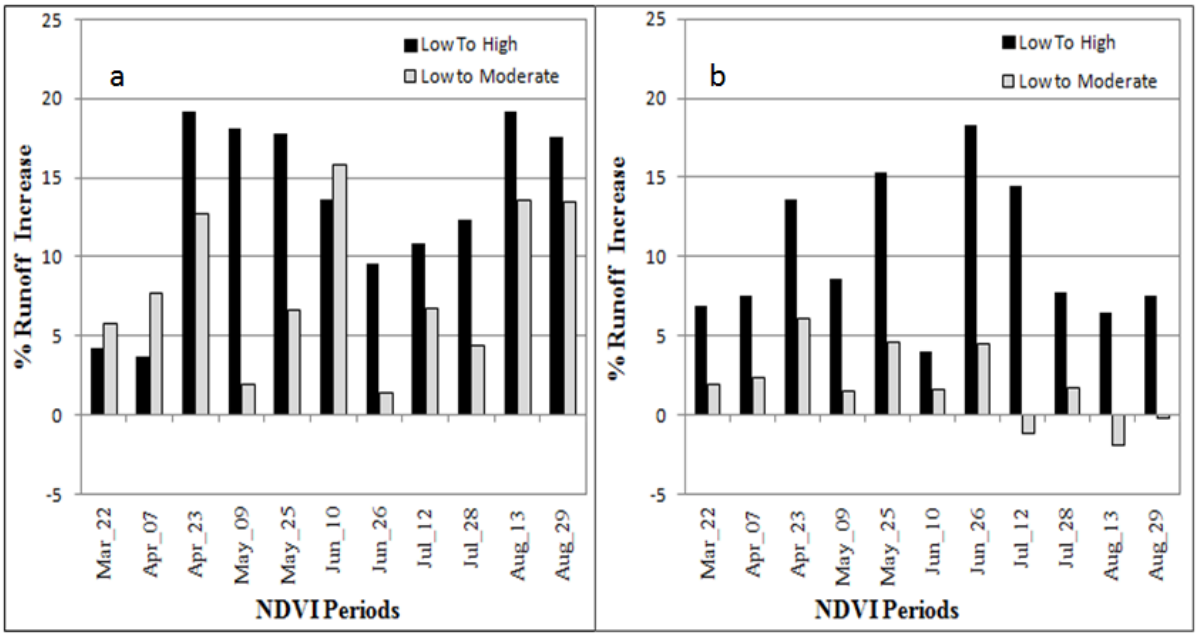


Figure 4-25: Percent increase in runoff for a) watersheds 112, 185, and 262; and b) watersheds 99, 310, and 361

4.5 Conclusion

Paired watershed study has been widely practiced in water resources research to study water quality and quantity and conservation impact. However, no objective criteria or method exists to select paired watersheds. A combination of K-means and hierarchical-agglomerative clustering techniques were investigated in this study to identify hydrologically homogeneous clusters of watersheds at Fort Riley, Kansas. Eight hydrologically important topographic parameters, including one soil parameter, were used in classification. Those parameters were transformed using standardized z-score to improve generalizability of the results and eliminate the domination of a certain variable. K-means clustering techniques grouped 316 watersheds into six groups, and the agglomerative hierarchical technique classified those into small groups and into paired watersheds. Six paired watersheds identified through this process were used to apply

previously developed CN_{NDVI} and to study the impact of military maneuvers on runoff generation at smaller scale. These results showed that the paired watershed selection technique used in this study effectively identified land cover changes due to natural and/or anthropogenic disturbances. Although the study pertained to military maneuvers, the same parameter selection method and classification techniques could be applied for any landscape. However, additional caution is recommended when selecting parameters since it is challenging and can be subjective. The following chapter describes how the developed model could perform with relatively diverse land use/cover set up and with relatively larger watersheds.

References

Almeida, J. S. and C. A. Prieto (2014). Automated unsupervised classification of the sloan digital sky survey stellar spectra using k-means clustering, *Astrophys*. J. 763:50.

Althoff, D. P., Althoff, P. S., Lambrecht, N. D., Gipson, P. S., Pontius, J. S., and Woodford, P. B. (2007). Soil properties and perceived disturbance of grasslands subjected to mechanized military training: Evaluation of an index. *Land Degradation Development* 18(3): 269–288. doi:10.1002/ldr.773

Althoff, P., M. Kirkham, T. Todd, S. Thien, and P. Gipson (2009 a). Influence of Abrams M1A1 Main Battle Tank Disturbance on Tallgrass Prairie Plant Community Structure, *Rangeland Ecology Management* 62:480–490.

Althoff, P., T. Todd, S. Thien, and M. Callaham (2009 b). Response of soil microbial and invertebrate communities to tracked vehicle disturbance in tallgrass prairie, *Applied Soil Ecology* 43:122–130.

Althoff P. S. and S. J. Thien (2005). Impact of M1A1 main battle tank disturbance on soil quality, invertebrates, and vegetation characteristics. *Journal of Terramechanics* 42:159–176.

Andréassian, V. (2004). Waters and forests: From historical controversy to scientific debate. *Journal of Hydrology* 291(1): 1–27.

Bhat, S., J. H. Jacobs, H. Kirk, and J. Prenger (2005). Relationships between stream water chemistry and military land use in forested watersheds in Fort Benning, Georgia. *Ecol. Ind*. 6:458–466.

Bloschl, G. and Sivapalan, M. (1995). Scale issues in hydrological modelling—a review. ¨ *Hydrological Processes* 9(3–4): 251–290.

Bunger, C. (2013). City Commission Agenda Memo. The Wildcat Creek Floodplain Management Plan (FMP) is a summary document of the discussions and efforts of the Wildcat Creek Watershed Working Group (WCWWG).

Burn, D. H. and Boorman, D. B. (1993). Estimation of hydrological parameters at ungauged catchments. *Journal of Hydrology* 143:429–454.

Cao, J., Z. Wu, J. Wu, and W. Liu (2013). Towards information-theoretic K-means clustering for image indexing, *Signal Process* 93:2026–2037.

Cavadias, G. S., Ouarda, T., Bobee, B., Girard, C. (2001). A canonical correlation approach to the determination of homogeneous regions for regional flood estimation of ungauged basins. *Hydrological Sciences Journal* 46(4): 499–512.

Chang, M. (2006). Forest Hydrology: An Introduction to Water and Forests, 2nd ed. CRC Press, Bacon Raton, FL.

Christopher, O., Idown A. O., and Olugbenga A. S. (2010). Hydrological Analysis of Onitsha North East Drainage Basin using Geoinformatic Techniques. *World Applied Science Journal* 11:1297–1302.

Clausen, J. C. and J. Spooner (1993). Paired watershed study design. U.S. Enviromental Protection Agency.

Cuell, C. and B. Bonsal (2009). An assessment of climatological synoptic typing by principal component analysis and K-means clustering, *Theoretical and applied climatology* 98:361–373.

DeFries, R. and N. Eshleman (2004). Land-use change and hydrologic processes: a major focus for the future, *Hydrol. Process*. 18:2183–2186.

DeFries, R., J. Foley, and G. Asner (2004). Land-use choices: balancing human needs and ecosystem function, *Frontiers in Ecology and the Environment* 2:249-257.

Di Prinzio, M., Castellarin, A., and Toth, E. (2011). Data-driven catchment classification: application to the pub problem. *Hydrology and Earth System Sciences* 15:1921–1935.

Ford, C. R., Stephanie S. L.,Wayne T. S., and M. V. James (2011). Can forest management be used to sustain water-based ecosystem services in the face of climate change? *Ecol. Appl*. 21:2049–67.

Foster, J., P. Ayers, A. Lomardi-Przybylowicz, and K. Simmons (2006). Initial Effects of Light Armored Vehicle use on Grassland Vegetation at Fort Lewis, Washington. *Journal of Environmental Management* 81.4:315–22.

Galluccio, L., Michel, O., Comon, P., and A. O. Hero (2012). Graph based K-means clustering, *Signal Process* 92:1970–1984.

Gaston, K. J., T. M Blackburn, and K. K. Goldewijk (2003). Habitat conversion and global avian biodiversity loss, *Proceedings - Royal Society.Biological sciences* 270:1293–300.

Genereux, D., M. Jordan, and D. Carbonell (2005). A paired-watershed budget study to quantify inter-basin groundwater flow in a lowland rain forest, Costa Rica, *Water Resour. Res.* 41.

Grantham, W., E. Redente, C. Bagley, and M. Paschke (2001). Tracked Vehicle Impacts to Vegetation Structure and Soil Erodibility, *J. Range Manage*. 54:711–716.

Haugen, L., P. Ayers, and A. Anderson. (2003). Vehicle Movement Patterns and Vegetative Impacts during Military Training Exercises. *Journal of Terramechanics* 40.2:83–95.

Horton, R. E. (1932). Drainage basin characteristics, *Trans. Amer. Geophys. Union*.13:350361.

Howard, H. R., Wang, G., Singer, S., and A. B. Anderson (2011). Modeling and prediction of land condition for Fort Riley Military Installation. A specialty Conference of the American Society of Agricultural and Biological Engineers held in conjunction with the annual meeting of the Association of Environmental & Engineering Geologists.

Jetz, W., D. S. Wilcove, and A. P. Dobson (2007). Projected impacts of climate and land-use change on the global diversity of birds., *PLoS Biology* 5:e157–1219.

Johnson, S., G. Wang, H. Howard, and A. Anderson. (2010). Identification of superfluous roads in terms of sustainable military land carrying capacity and environment. *Journal of Terramechanical* 48(2): 97–104.

Kahya, E., Kalaycý, S., and Piechota, T. C. (2008). Streamflow Regionalization: Case Study of Turkey. *Journal of Hydrologic Engineering*. 13(4): 205–214.

Kaufman, L. and P. J. Rousseeuw (1990). Finding groups in data - An introduction to cluster-analysis, New York: John Wiley & Sons.

King, K. W., P. C. Smiley, B. J. Baker, and N. R. Fausey (2008). Validation of paired watersheds for assessing conservation practices in the Upper Big Walnut Creek watershed, Ohio, *J. Soil Water Conserv*., 63, 380–395.

Kun, L., P. Ayers, H. Howard, and A. Anderson (2009). Influence of turning radius on wheeled military vehicle induced rut formation. *Journal of Terramechanics.* 46-2: 49–55.

Lu, Y. and Y. Wan (2013). PHA: A fast potential-based hierarchical agglomerative clustering method, *Pattern Recognition* 46:1227–1239.

MathWorks (2012). Statistics Toolbox: User's Guide (R2012a). Retrieved January 14, 2013, from http://www.mathworks.com/help/pdf_doc/stats/stats.pdf, 869–905.

Martinez, W. L. and A. R. Martinez (2008). *Computational Statistics Handbook with MATLAB*, (2nd ed).New York: Chapman and Hall.

Martinez, W. L. and A. R. Martinez (2004). *Exploratory Data Analysis with MATLAB: Computer Science and Data Analysis Series*. New York: Chapman and Hall.

Milchunas, D., K. Schulz, and R. Shaw (2000). RESEARCH: Plant Community Structure in Relation to Long-Term Disturbance by Mechanized Military Maneuvers in a Semiarid Region, *Environ. Manage*. 25:525–539.

Mojena, R. (1977). Hierarchical grouping methods and stopping rules: an evaluation, *The Computer Journal* 20:359–363.

"NASA LP DAAC" (2013). MODIS Land Products Quality Assurance Tutorial: Part 2, How to interpret and use MODIS QA information in the Vegetation Indices product suite https://lpdaac.usgs.gov/sites/default/files/public/modis/docs/MODIS_LP_QA_Tutorial-2.pdf

Nathan, R. J. and T. A. McMahon. (1990). Evaluation of Automated Techniques for Baseflow and Recession Analysis. *Water Resources Research*. 26(7): 1465–1473.

Natural Resources Conservation Service (NRCS) United States Department of Agriculture (USDA) (2007). Kansas Annual Precipitation. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_032018.pdf

Nobert, J. and J. Jeremiah (2012). Hydrological Response of Watershed Systems to Land Use/Cover Change. A Case of Wami River Basin, *The Open Hydrology Journal* 6:78–87.

Olaya, V. (2009). Basic land-surface parameters. In Hengl T, and H.I. Reuter (eds), Geomorphometry: concepts, software, applications. *Developments in soil science, Elsevier*. 33:141–170.

Perkins, D., N. Haws, J. Jawitz, B. Das, and P. Rao (2007). Soil hydraulic properties as ecological indicators in forested watersheds impacted by mechanized military training, *Ecological Indicators* 7-3:589–597.

Pockrandt, B. R. (2014). A multi-year comparison of vegetation phenology between military training lands and native tallgrass prairie using TIMESAT and moderate-resolution satellite imagery. Thesis. Master of Arts. Kansas State University.

Prokopy, L., Z. A. Goecmen, J. Gao, S. Allred, and J. Bonnell (2011). Incorporating Social Context Variables Into Paired Watershed Designs to Test Nonpoint Source Program Effectiveness1, *J. Am. Water Resour. Assoc*. 47:196–202.

Prosser, C., K. Sedivec, and W. Barker (2000). Tracked Vehicle Effects on Vegetation and Soil Characteristics, *J. Range Manage*. 53:666–670.

Quist, M. C., P. A Fay, C. S. Guy, A. K. Knapp, and B. N. Rubenstein (2003). Military training effects on terrestrial and aquatic communities on a grassland and military installation. *Ecological Applications* 13(2): 432–442.

Raper, R. L. (2005). Agricultural traffic impacts on soil, *J. Terramech*. 42:259–280.

Razavi Zadegan, S. M., Mirzaie, M., and Sadoughi, F. (2013). Ranked k-medoids: A fast and accurate rank-based partitioning algorithm for clustering large datasets. *Knowledge-Based Systems* 39:133–143.

Ricker, M. C., B. K. Odhiambo, and J. M. Church (2008). Spatial analysis of soil erosion and sediment fluxes: a paired watershed study of two Rappahannock River tributaries, Stafford County, Virginia, *Environ. Manage.* 41:766–778.

Riley, S. J., S. D. DeGloria, and R. Elliot (1999). A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences* 5(1–4): 23–27.

Rao, A. R and V. V. Srinivas (2006). Regionalization of Watersheds by Fuzzy Cluster Analysis. *Journal of Hydrology* 318:57–79.

Rose, S. and Peters, N. E. (2001). Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. *Hydrological Processes* 15:1441–1457.

Sala, O. E., E. Berlow, J. Bloomfield, F. S. Chapin, J. J. Armesto, R. Dirzo, E. Huber-Sanwald, L. F. Huenneke, R. B. Jackson, A. Kinzig, R. Leemans, D. M. Lodge, H. A. Mooney, M. Oesterheld, N. L. Poff, M. T. Sykes, B. H. Walker, M. Walker, and D. H. Wall (2000). Global biodiversity scenarios for the year 2100. *Science* 287:1770–4.

Schilling, K. E. (2002). Chemical transport from paired agricultural and restored prairie watersheds. *Journal of Environmental Quality* 31:1184–1193.

Schilling, K. E. and Spooner, J. (2006). Effects of watershed-scale land use change on stream nitrate concentrations. *Journal of Environmental Quality* 35:2132–45.

Singh, W., Hjorleifsson, E., and G. Stefansson (2011). Robustness of fish assemblages derived from three hierarchical agglomerative clustering algorithms performed on Icelandic groundfish survey data. *ICES Journal of Marine Science* 68:189–200.

Ssegane, H., D. M. Amatya, G. M. Chescheir, R. W. Skaggs, E. W. Tollner, and J. E. Nettles (2012). Consistency of hydrologic relationships of a paired watershed approach. *Am. J. Clim. Change* 2:147–164.

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. *Trans. Amer. Geophys. Unio*n. 38:913–920.

Strahler, A. N. (1964). Quantitative geomorphology of drainage basins and channel networks. *Handbook of Applied Hydrology*, New York, Mc.Graw Hill Book Company.

Storck, P., Bowling, L., Wetherbee, P., and Lettenmaier, D. (1998). Application of a GIS-based distributed hydrology model for prediction of forest harvest effects on peak streamflow in the Pacific Northwest. *Hydrological Processes* 12:889–904.

Thurow, T. L., S. D. Warren, and D. H. Carlson (1993). Tracked vehicle traffic effects on the hydrologic characteristics of Central Texas rangeland. *Transactions of the American Society of Agricultural Engineers* 36:1645–1650.

USDA Soil Conservation Service in cooperation with Kansas Agricultural Experiment Station (1975). Soil survey of Riley and part of Geary County, Kansas.

Varshavsky, R., Horn, D., and M. Linial (2008). Global considerations in hierarchical clustering reveal meaningful patterns in data. *PLoS ONE* 3:2247.

Veum, K. S., K. W. Goyne, P. P. Motavalli, and R. P. Udawatta (2009). Runoff and dissolved organic carbon loss from a paired-watershed study of three adjacent agricultural Watersheds, Agric., *Ecosyst. Environ*. 130:115–122.

Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo (1997). Human domination of Earth's ecosystems. *Science* 277:494 and 278:21–21.

Werth, D. and Avissar, R. (2002). The local and global effects of Amazon deforestation. *Journal of Geophysical Research –Atmospheres* 107(D20)808. doi: 10-1029/2002J8000717

Wilson, J. (2012), Digital terrain modeling, *Geomorphology* 137:107–121.

Yu, Y., Yoon, S. O., Poulogiannis, G., Yang, Q., Ma, X. M., and J. Villén (2011). Quantitative phosphoproteomic analysis identifies the adaptor protein Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. *Science* 332:1322–1326.

Chapter 5 - Application of NDVI-based Curve Number Model to Quantify Rainfall-Runoff Relationship in Diverse Land Use/Cover Areas

Abstract

The CN_{NDVI}, developed using back-calculated CN and NDVI, was applied to study the performance of CN_{NDVI} in larger and more diverse land use/cover watersheds. Daymet precipitation and CN_{NDVI} were used to estimate runoff, which was compared with direct runoff extracted from stream flow gauging stations in the study watersheds, Chapman (grassland dominated) and Upper Delaware (agriculture dominated). Low, moderate, and high flow years were selected based on the stream flow observed at USGS runoff gauging points in the last decade. The overall flow of selected years and moderate flow performed well compared to low and high flow years in both watersheds. In the Chapman watershed, for every 1.057 increase in observed flow, the predicted flow increased by 1 for overall study years that means the model under-predicted by about 6 %; and the intercept showed less error in the model which is 0.23 mm. And for moderate flow year, model under-predicted flow by about 30 % with every 1.3 increase in CN_{NDVI} flow, the observed flow increased by 1. The CN_{NDVI} flow over-predicted by 38% for the overall flow year in Upper Delaware. The moderate flow over predicted by the model by 22% and the intercept in all cases of Delaware showed less error in the model. And the moderate year was the best performing relatively as that of grassland dominated watershed. In all seasons of Upper Delaware and three seasons in Chapman (spring, summer, and autumn), molded runoff values were closely matched with observed values statistically. However in the case of Chapman Creek, the grassland dominated watershed, the model flow was significantly

different than the observed flow during winter period. In both watersheds, the low flow year runoff comparison was showed the worst performance statistically. Daymet precipitation and CN_{NDVI} could provide an opportunity for timely surface runoff water planning and management.

5.1 Introduction

Surface runoff, an important component of hydrological processes, is widely studied in water resources for planning and management practices. Surface runoff studies have become crucial as worldwide land use modifications and the importance of drought and flood predictions evolve. The rainfall-runoff process is complex, dynamic, and nonlinear (Fan et al., 2013; Song et al., 2011), controlled by many interrelated physical factors. As land use/cover and soil physicochemical properties become increasingly diverse, the complexity of rainfall-runoff relationship increases. Therefore, predicting the amount and rate of runoff becomes increasingly difficult and time-consuming. The CN, also referred to as the SCS-CN method, is one of the most commonly used empirical methods to estimate surface runoff from rainfall events. It represents combined hydrologic effects of land use/cover, HSG, agricultural land treatment class, and hydrologic condition. The CN method has been widely used for watershed modeling and has been incorporated in numerous hydrological models such as Soil and Water Assessment Tool (SWAT), Hydrologic Engineering Center-Hydrologic Modeling System (HEC-HMS), Erosion Productivity Impact Calculator (EPIC), and Agricultural Non-Point Source Pollution Model (AGNPS) (Kousari et al., 2010; Hawkins et al., 2002; Woodward et al., 2002). The wide applicability of CN shows how crucial the method is in water resources for planning and management practices. Even though, it is widely applied and accepted as a method for runoff estimation; studies have indicated that it should be evaluated and adapted to local and regional conditions (D'Asaro et al., 2014; Epps et al., 2013; Soulis et al., 2009; Yuan, 2014).

Hawkins (2014) stated that some watersheds perform differently from SCS-CN runoff response patterns, leading to significant differences between the model and reality. Several other studies also mentioned that the estimated runoff was inaccurate in semiarid watersheds in southeastern Arizona due to higher retention capacity of the soil (Baltas et al., 2007; Hjelmfelt, 1980; Soulis et al., 2009; Yuan, 2014). Garen and Moore (2005) stated that the use of CN is beset with a number of problems, issues, and misinterpretations that weaken its usage to estimate accurate representation of the amount, paths, and source areas upon which erosion and water quality predictions depend. Many of these unresolved issues in the empirical CN based rainfall-runoff relationship are due to the fact that the method is very simplified though the relationship is so complex and dependents on multiple watershed related factors such as land use/cover and soil related factors, mainly soil moisture at the time of rainfall event.

The CN is generally determined using SCS-CN lookup tables that primarily contain land use/cover and land surface characteristics that affect the rainfall-runoff relationship. An area-weighted average CN for the entire watershed/subwatershed is often calculated since a watershed is usually a combination of land-use/cover and soil conditions. However, the CN approach does not reflect temporal variability of hydrological conditions and it cannot reflect the effects of seasonal or dynamic land-use changes, which are essential for reflecting seasonality of hydrological conditions. Many researchers have attempted to include spatial variability of land in CN calculation. Hong and Adler (2008) used land cover, soils, and antecedent moisture conditions to develop a global SCS-CN runoff map. Canters et al. (2006) derived a catchment scale CN using impervious surface, vegetation, bare soil, and water/shade information. Reistetter and Russell (2011) proposed a CN calculation method by integrating percentages of impervious surface, tree canopy density, and pervious surface. However, none of these methods used

temporal/ seasonal variability of land cover of CN or combined spatial and temporal variability of land cover.

The major reason for this gap in CN research is limited spatiotemporal data. Because hydrological data are typically point/field measurements, satellite remote sensing data and GIS techniques are viable alternatives or supplements because it is available over vast ungauged regions. Based on the considerations of seasonal or dynamic land-use changes of landscape and characteristics of medium-resolution satellite imagery, a previous study developed CN_{NDVI} model using remotely sensed data. The CN_{NDVI} model was developed to address the seasonality of land use/cover effects and capture the temporal variability of hydrological conditions. This study attempted to address CN_{NDVI} model's applicability for different land uses and watershed sizes. The goal of this study was to evaluate the CN_{NDVI} model for estimating runoff, using Daymet precipitation data, for large watershed with more complex land use patterns. The specific objectives were to assess the application of Daymet precipitation and CN based on NDVI to estimate runoff in relatively larger watersheds; and to evaluate runoff in grassland and agriculture dominated land use/cover watersheds using USGS outlets.

5.2 Study Area

Two watersheds with different land use/cover and hydrological conditions in northeast Kansas were used for study: Chapman and Upper Delaware watersheds. The Chapman watershed (HUC 10: 1026000806) was dominated by grassland, while the Upper Delaware (HUC10: 1027010301 [Muddy Creek] and HUC10: 1027010302 [Grasshopper]) was dominated by agricultural land use/cover. NCDC rainfall gauge points (two in each watershed) were located in Chapman and the Upper Delaware (Figure 5-1). Daymet precipitation points (Thornton et al., 1997) enabled high spatial representation (1 km x 1 km) of precipitation. Figure 5-1 shows the

study watersheds, available NCDC rainfall gauges, Daymet precipitation points used to generate runoff in this study, and USGS streamflow outlet locations. Watersheds were chosen based on dominant land use/cover, HSG, and stream gauge locations. Land cover data was obtained from the 2011 USGS National Land Cover Dataset (NLCD) at a 1:24,000 scale (30 m resolution). Soil data were derived from the USDA-NRCS SSURGO database. Hydrological characteristics of study watersheds are listed in Table 5-1.

Table 5-1: Watershed hydrological characteristics

	Chapman Creek Watershed	Upper Delaware Watershed
Area (km2)	776.2	1130
Elevation range (m)	338–488	282–422
Average annual precipitation (mm)	762–813	864–965
Long-term average flow (m3/s)	2.78	7.07
Dominant HSG	C (49.8%) and B (36.3%)	D (48.2%) and B (32%)
Dominant land use/cover	Forest/shrub land (56.7%) and Agriculture (33.5%)	Agriculture (74.9%)

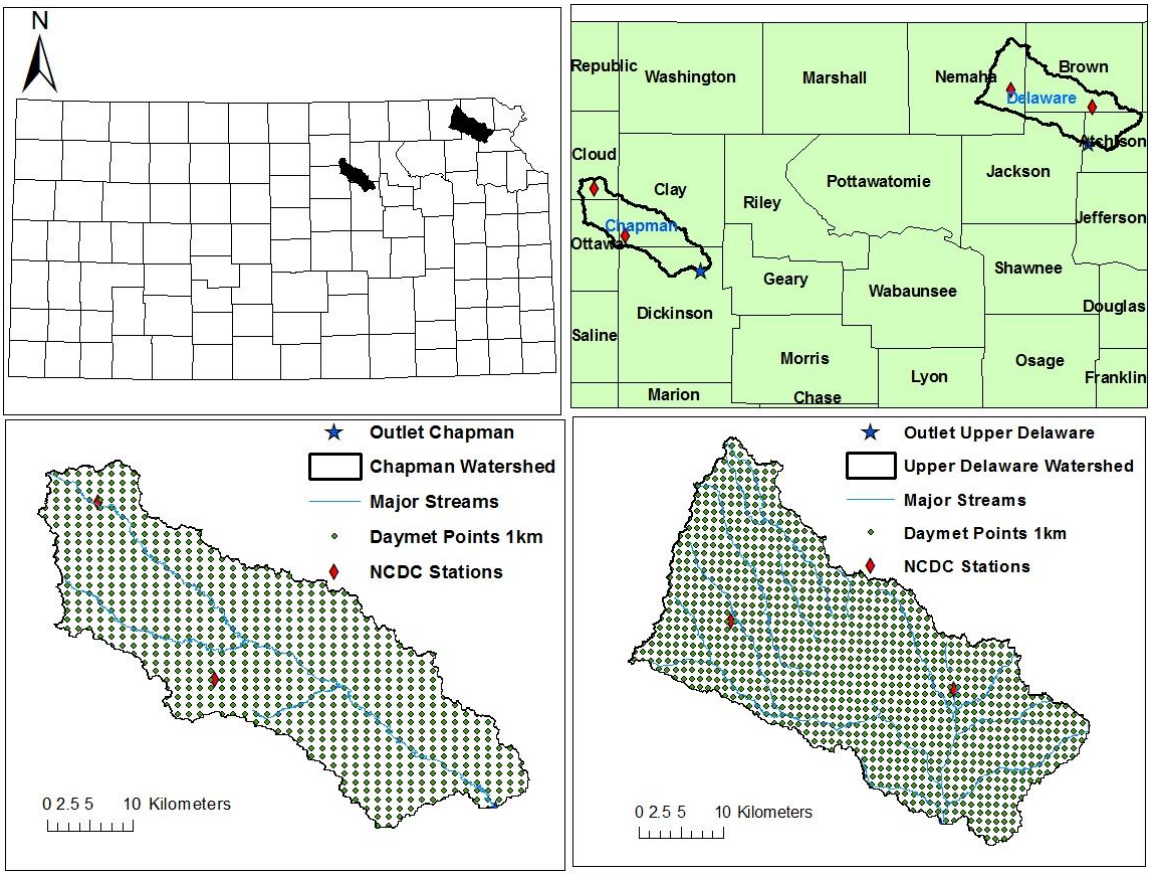


Figure 5-1: Chapman Creek and Upper Delaware watersheds in Kansas, Dayment points, NCDC stations, and USGS outlets

5.2.1 Chapman Creek Watershed

The Chapman watershed is located in northeast Kansas and has a 776.2 km^2 drainage area. The Chapman Creek watershed was delineated with the USGS station (06878000 Chapman Creek near Chapman, Kansas) as the outlet (Figure 5-1). Therefore, the current study area covered approximately 91% of 10-digit hydrological unit code (HUC 10-1026000806) area. The average annual precipitation of the watershed region ranged from 762 mm to 813 mm (NRCS, 2007), with a long-term average annual stream flow of 2.78 m^3/s.

The elevation of Chapman watershed ranges from 338 m to 488 m, with average 4.15% land slope (Figure 5-2). Soils in the Chapman Creek watershed are mainly Class B and C HSG, accounting for 36.3% and 49.8%, respectively, and categorized as moderate runoff potential

(Figure 5-3). Based on NLCD (2011), land cover (Figure 5-3) in the watershed is approximately 57% grassland; the remainder is cultivated crop (33.5%) and developed area (3.4%) (Table 5-1 and Figure 5-3).

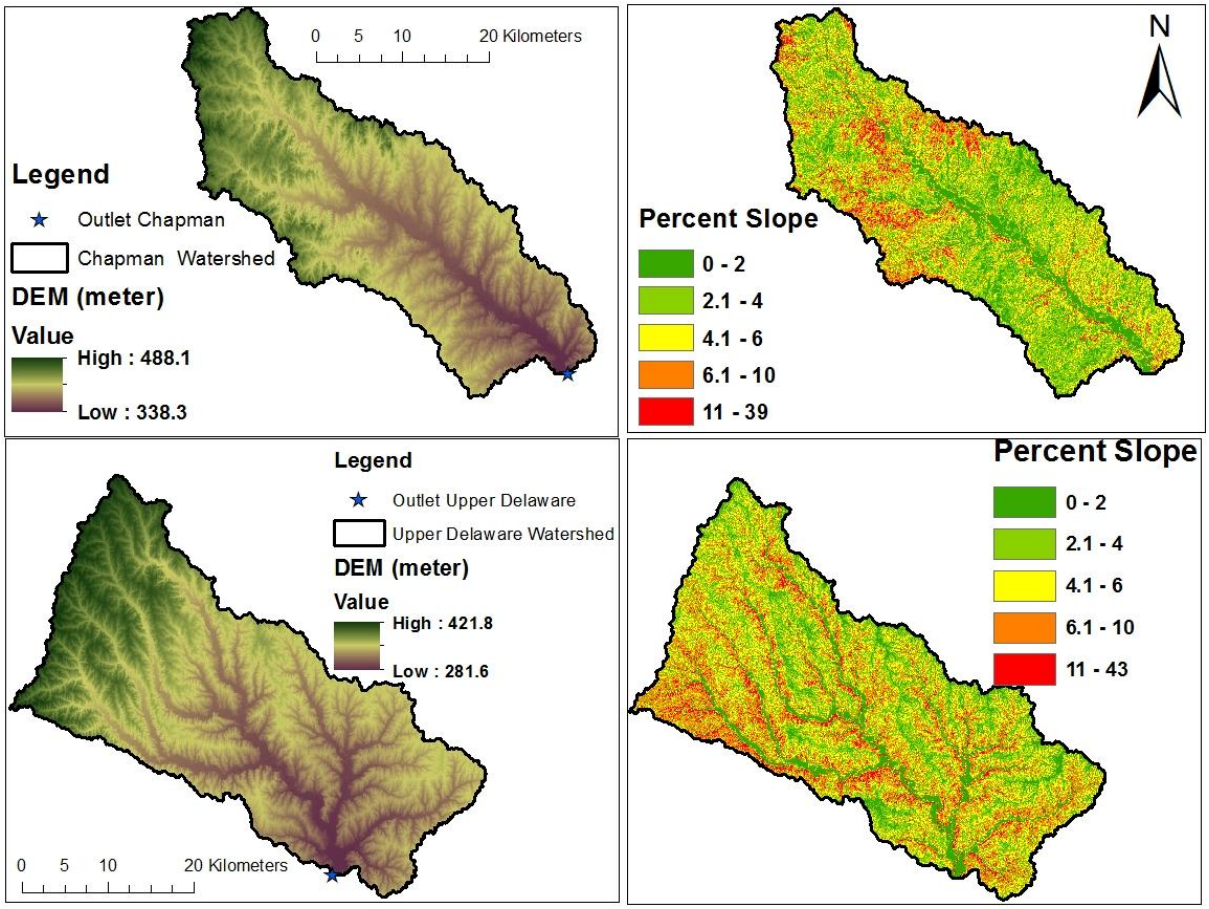


Figure 5-2: DEM and percent slope of Chapman Creek (upper) and Upper Delaware (lower) watersheds

5.2.2 Upper Delaware Watershed

The Upper Delaware watershed has a drainage area (Figure 5-1) approximately 1130 km^2, covering two HUC 10s (Upper Delaware [HUC10s- 1027010301 Muddy Creek and 1027010302 Grasshopper]). The elevation of the watershed ranges from 282 m to 422 m, with an average 4.73% land slope. Soils in the Upper Delaware watershed are mainly Class D HSG (48.2%), with high runoff potential, followed by Class B HSG (32%), which has moderately low

runoff potential (Figure 5-3). The Upper Delaware watershed was considered for this study due to its high percentage of agricultural land cover. Based on NLCD (2011), land cover in the watershed is approximately 75% cultivated land (cultivated crops and pasture/hay), and the remainder is grassland (10%), forest (8%), and developed area (5%) (Table 5-1 and Figure 5-3). Hay, dryland corn, and soybean are the major cultivated crops in the watershed. The average annual precipitation of the watershed region ranges from 864 mm to 965 mm (NRCS, 2007). USGS station (06890100) Delaware River near Muscotah is located in the outlet of watershed (Figure 5-1). The Upper Delaware had 7.07 m^3/s average annual total long-term flow (1971–2014). The Upper Delaware watershed faces many challenges associated with high runoff, such as sedimentation, nutrient management, fecal coliform bacteria contamination, pesticide contamination, household hazardous wastes, water wells, and point source pollution (Bosworth, 2007).

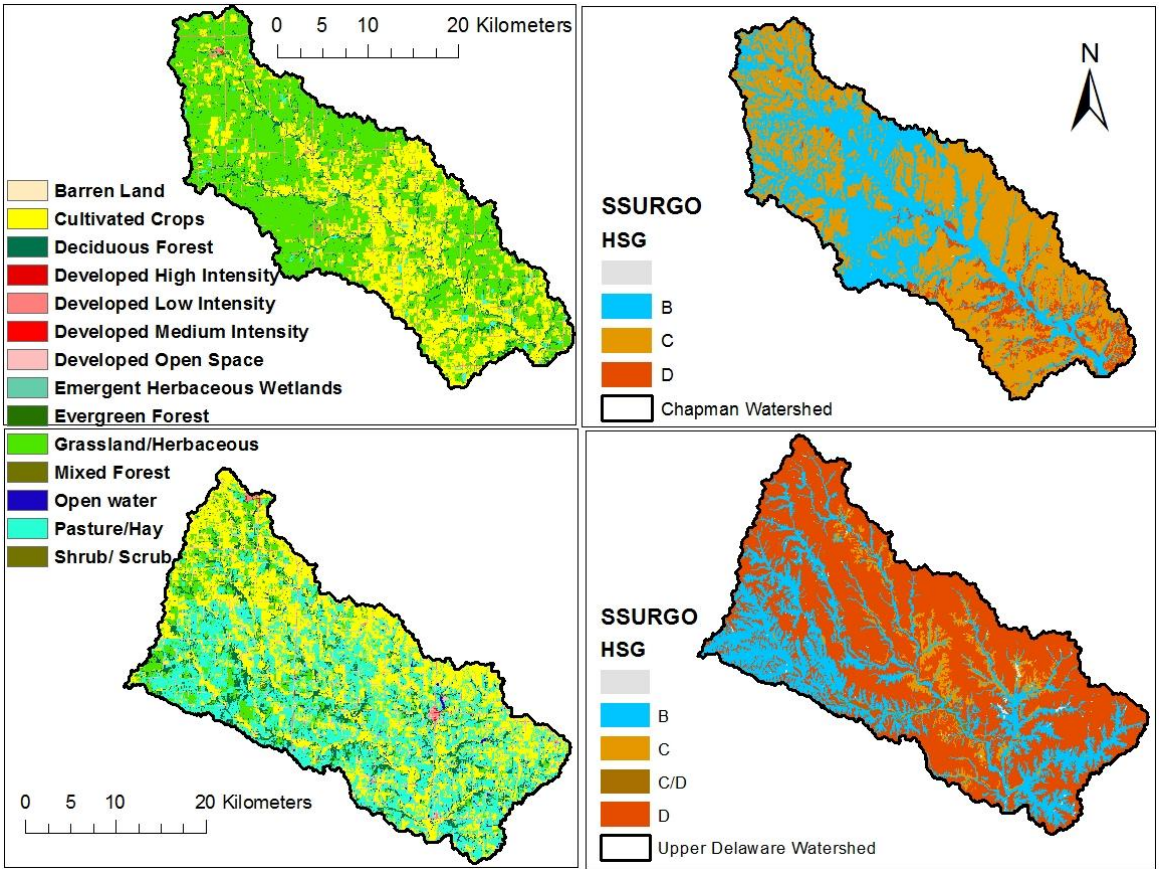


Figure 5-3: Land use/cover based on NLCD and hydrologic soil group based on SSURGO of Chapman Creek (top) and Upper Delaware watershed (bottom)

5.2.3 Study Period

Three flow periods (high flow, moderate flow, and low flow) from 2000 to 2014 were analyzed. Any annual average flow less than or equal to 75% of long-term average flow (45 years) was considered low flow. Long-term average flow plus/minus 25% was considered moderate flow, and flow greater than 125% of long-term average was considered high flow. Based on this flow evaluation, flows from the year 2003 were considered low flow, flows from the year 2007 were moderate flow, and flows from the year 2010 were high flow. Those years were selected based on availability of NDVI for analysis. Figure 5-4 shows the 45-year period long-term stream flow average for Chapman and Upper Delaware watersheds (1970–2014). Figure 5-5 shows stream flow from 2000 to 2014, which is the time period NDVI is

available. Figure 5-5 also shows selected stream flow as high, moderate, and low flow for this study. The criteria used to select low, moderate, and high flow is described in Table 5-5.

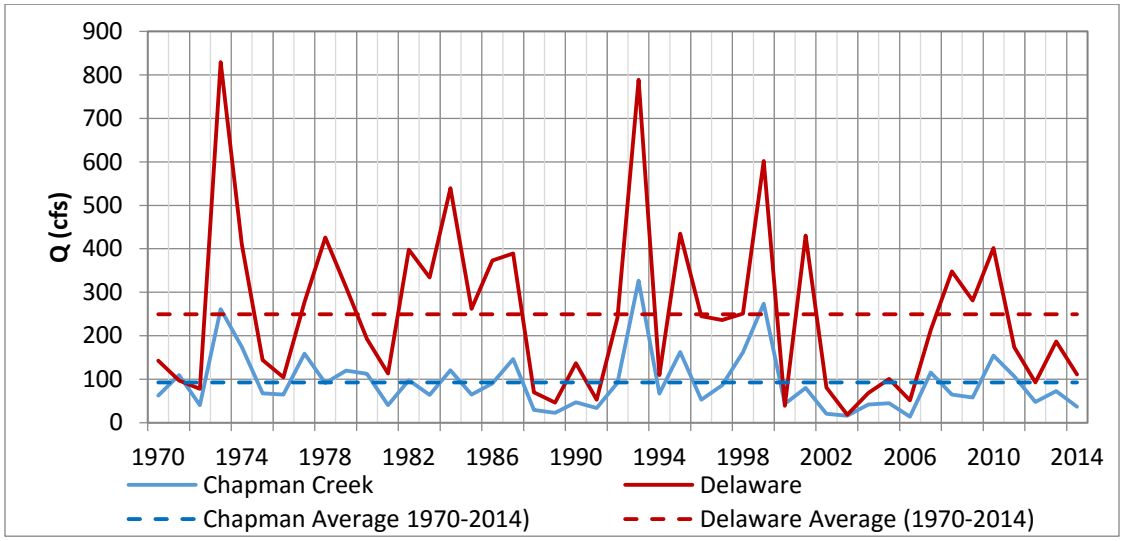


Figure 5-4: Annual average long term flow for Chapman Creek and Upper Delaware watersheds

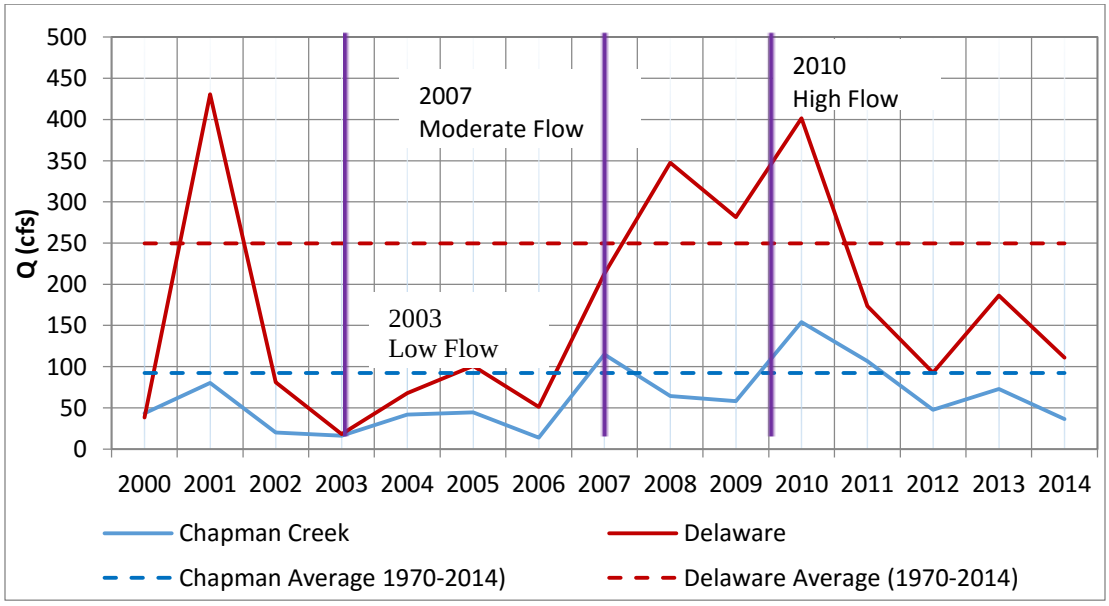


Figure 5-5: Annual stream flow for Chapman Creek and Upper Delaware watersheds

Table 5-2: Study period flow categorization and criteria

Study Period Flow Categorization	Criteria
Low	≤75% long-term average flow†
Moderate	75%–125% long-term average flow†
High	≥125% long-term average flow†

† Long-term flows for Chapman Creek and Upper Delaware are 92.3 cfs and 249. 53 cfs, respectively.

5.3 Materials and Methods

5.3.1 General Method

The following steps were used to estimate runoff from rainfall events in this study.

Step 1: NDVI was extracted from multi-parameter MODIS-NDVI 16-day composite grid data (MOD13Q1) and processed into a scale of 0–255 from the MODIS raw data scale (-2000–10000).

Step 2: CN_{NDVI} from NDVI was calculated based on the model developed in Chapter 3.

Step 3: Daymet precipitation was downloaded and processed based on 1 km latitude and longitude interval.

Step 4: Daymet precipitation was interpolated for each study area using the Inverse Distance Weighting (IDW) method.

Step 5: Runoff was calculated based on SCS-CN formula using the processed CN_{NDVI} and precipitation input.

Step 6: Stream flow at gauging points of USGS National Streamflow Information Program (NSIP) was downloaded to compare to runoff estimated based on CN_{NDVI} and SCS CN.

Step 7: Baseflow was separated from stream flow to estimate direct runoff for comparison using the Baseflow Filter Program developed by Arnold and Allen (1999).

Step 8: Direct runoff from USGS and runoff based on CN_{NDVI} and SCS-CN were analyzed to assess model performance.

5.3.2 Curve Number based on NDVI

A CN_{NDVI} model based on NDVI (Equation 1, 2) calculated in Chapter 3 was used to estimate runoff. The CN_{NDVI} model was developed using four small northeastern Kansas grassland watersheds with an average area of 1 km^2, located on the Konza Prairie Long-Term Ecological Research (LTER) site Twelve years (2001-2012) of daily precipitation and runoff were used in CN_{NDVI} model development. The rain gauges in the watersheds operated from April 1 to October 31. The annual average precipitation at the Konza was 835 mm. The watersheds were dominated with hydrologic soil group C soils. Detail description on CN_{NDVI} development can be seen in the chapter 3.

$$NDVI = 0.0208 * NDVI_{org} + 42 \qquad (5\text{-}1)$$

where *NDVI* is processed NDVI for CN model (value 0–255) and $NDVI_{org}$ is original NDVI downloaded from MODIS website (16-bit signed integer ranging -2000–10000).

$$CN_{NDVI} = -0.108 * NDVI + 98 \qquad (5\text{-}2)$$

where *NDVI-CN* is curve number based on NDVI and *NDVI* is NDVI with value 0–255. CN model development is detailed in Chapter 3 of this dissertation.

5.3.3 Daymet Precipitation

Daymet (Thornton et al., 1997) provides gridded estimates of weather data at 1 km × 1 km spatial resolution and daily time step. It contains estimations of daily weather parameters: day length (s/day), precipitation (mm/day), shortwave radiation (W/m^2), snow water equivalent (kg/m^2), maximum air temperature (^oC), minimum air temperature (^oC), and water vapor pressure (Pa). Daymet estimates are based on daily meteorological observations (observations of maximum temperature, minimum temperature, and precipitation from NCDC daily ground-based meteorological stations, and the SNOwpack and TELemetry (SNOTEL) dataset managed and distributed by the NRCS. DEM (1-km or 30 arc second DEM) and land value "Mask" (information of land and water areas on earth) were used as a secondary data. According to Oak Ridge National laboratory (ORNL) website, Daymet data has been available from 1980 through the latest full calendar year for the United States, Mexico, and Canada (south of 52 degrees North) as station density allows. Daily precipitation includes all forms converted to water-equivalent. Daymet data has been broadly applied to fields such as hydrology, terrestrial vegetation growth models, carbon cycle science, and regional-to-large-scale climate change analysis (Thornton et al., 2014). For this study, Daymet data was accessed through the ORNL Distributed Active Archive Center (DAAC). For the study area, the latitude and longitude based on 1 km × 1 km spatial resolution was processed by creating point shape files within each watershed; Chapman watershed had 782 points, and Upper Delaware had 1148 points (Figure 5-1). Precipitation was extracted and processed using MATLAB from all points within each watershed.

5.3.4 Surface Runoff Estimation

Stream flow at the outlets of the two watersheds, Chapman Creek and the Upper Delaware River (Figure 5-1), was downloaded from USGS National Streamflow Information Program (NSIP) website, and direct runoff was extracted and used to compare to runoff estimated based on the CN_{NDVI}. These stations were selected because they have only natural or unregulated flow (Rasmussen and Perry, 2001) with no diversions and dams or other significant regulations within the watersheds. (The CN method does not account for reservoirs or other diversions.) The assumption was made that minimal diversions or ponds were contained within both watersheds.

5.3.4.1 Baseflow Separation

Baseflow is the component of streamflow that comes from groundwater or return flow. Kansas baseflow displays significant spatial variability as a consequence of variation in climatic conditions, land use, topography, and hydrogeological heterogeneity (Sophocleous and Wilson, 2000). In this study baseflow was separated using the automated recursive digital filter program (or baseflow filter program) (Nathan and McMahon, 1990; Arnold et al, 1995). The filter was run over the streamflow three times (forward/Pass 1, backward/Pass 2, and forward/Pass 3), separating low frequency baseflow from high frequency quick streamflow. From Pass 1 to 3 the percentage of baseflow contribution decreased (Nathan and McMahon, 1990; Santhi et al., 2008). Fraction of baseflow (BF) contribution to total streamflow (average of Pass 1 and 2), a baseflow recession constant (alpha factor), and baseflow days of the two watersheds were calculated during baseflow analysis (Table 5-3). Recession constant (alpha factor) is a direct index of groundwater flow responses to changes in recharge. Low flow response values for both

watersheds represented slow response to recharge. Baseflow days represented the required number of days for baseflow recession to decline through one log cycle (Arnold et al., 1995)

Table 5-3: Calculated baseflow as a fraction of total streamflow, alpha factor and baseflow

USGS Station	Pass 1 Baseflow Fraction	Pass 2 Baseflow Fraction	Average Baseflow Fraction	Alpha Factor (1/days)	Baseflow (days)
Chapman	0.44	0.32	0.38	0.0605	54
Delaware	0.40	0.25	0.325	0.0640	36

Average of Pass 1 and 2 baseflow fraction was used to separate baseflow from streamflow. Thereafter, runoff volume was converted to depth of runoff based on each watershed area.

5.3.4.2 Potential Runoff-Contributing Areas

Runoff is the result of two processes that produce overland flow. The first process is infiltration-excess overland flow, which occurs when precipitation intensity exceeds the rate of infiltration. The second process is saturation-excess overland flow, which occurs when precipitation falls on temporarily or permanently saturated land surface areas. Saturated areas are produced during storms and disappear during dry periods. Based on Juracek (1999), high runoff potential areas are characterized by high antecedent soil moisture and high rainfall intensity. Therefore, identification of potential runoff-contributing areas could be helpful to estimate relatively accurate response areas for a single rainfall event.

In order to estimate infiltration excess and saturation-excess overland flow that contributed to direct runoff, potential runoff contributing areas must be estimated. Juracek (1999) estimated potential runoff-contributing areas using topographic and soil information based on the topographic wetness index (TWI) derived from flow accumulation and slope. In this study

potential runoff-contributing areas were estimated to 63% of the watershed for Chapman and 89% of the Upper Delaware based on 100 m resolution potential runoff of grid (Juracek, 1999), which was downloaded from Kansas Data Access and Support Center for analysis.

5.3.5 Model Evaluation

Efficiency of the runoff estimation model was assessed using various model evaluation statistical techniques. Modeled and observed runoff was evaluated using regression graphs slope and *y*-intercept (Pineiro et al., 2008) and three quantitative statistics: Pearson's correlation coefficient (*r*), Nash-Sutcliffe efficiency (NSE) (Equation 5-3), and percent bias (PBIAS) (Equation 5-4). These graphics and statistics are commonly used in model evaluation (Moriasi et al., 2007; Pineiro et al., 2008; Shaw and Walter, 2008; Patil et al., 2008). Pineiro et al. (2008) evaluated models that compared observed and predicted data, suggesting that Observed (y-axis or ordinate) versus Predicted (x-axis or abscissa) should be used to estimate the slope and intercept accurately with less error in the model.

$$\text{NSE} = 1-\left[\frac{\sum_{i=1}^{n}(Y_i^{Obs}-Y^{sim})^2}{\sum_{i=1}^{n}(Y_i^{Obs}-Y^{mean})^2}\right] \qquad (5\text{-}3)$$

where Y_i^{Obs} is the i^{th} observation (i.e., measured runoff) value evaluated, Y^{sim} is the ith simulated (i.e., predicted runoff) value evaluated, Y^{mean} is the mean of observed data, and *n* is the total number of observations.

$$\text{PBIAS} = \left[\frac{\sum_{i=1}^{n}(Y_i^{Obs}-Y_i^{sim})*(100)}{\sum_{i=1}^{n}(Y_i^{Obs})}\right] \qquad (5\text{-}4)$$

The theoretical NSE value range between $-\infty$ and 1.0 (1 inclusive), with NSE =1 being the optimal value (Moriasi et al., 2007). Moriasi et al. (2007) stated that ranges of NSE values between 0.0 and 1.0 are generally acceptable levels of performance. However, the authors suggested that values ≤0.0 indicate that mean observed values are better predictors than

simulated values that indicate unacceptable performance of the model. Percent bias measures the average tendency of predicted values to be larger or smaller than the observed values (Moriasi et al., 2007; Gupta et al., 1999), where the optimal value is zero, positive values indicate overestimation bias, and negative values indicate underestimation bias.

In order to estimate model efficiency on seasonal variability, runoff was grouped into four seasons in this study. Runoff event values from March through May were grouped for spring, June through August for summer, September through November for autumn, and December through February for winter. A two-sample t-test (assuming unequal variance) tested whether or not any significant difference was present between observed and CN_{NDVI} model predictions.

5.4 Results and Discussion

5.4.1 Runoff Estimation

5.4.1.1 Runoff Estimation for Grassland Dominated Watershed

CN_{NDVI} model more accurately predicted runoff than the existing standard (literature) CN method. Figure 5-7 shows higher r^2 for all observed runoff versus CN_{NDVI}-predicted runoff, except for the low-flow period (2003) during which r^2 was slightly higher for observed runoff versus standard CN runoff. The slope and intercept of each comparison for overall study years (all three years), low flow (2003), moderate flow (2007), and high flow (2010) year data is shown in Figure 5-7. The slope and intercept (Figure 5-6) of the linear regression fit indicates how well the predicted data matched the observed data; the slope indicates the relative relationship between predicted and observed, and the intercept indicates the presence of a lag or lead between model predictions and measured data (Moriasi et al., 2007).

Based on Figures, the overall study period (Figure 5-6) and the moderate flow year (Figure 5-8) performed relatively well compared to low and high flow years in the standard CN-based runoff and CN_{NDVI}-based runoff estimation. In general (for the overall period, moderate, and high flow years), CN_{NDVI} performed better than the standard CN based on slope, r^2, and NSE and PBIAS in Table 5-3. However, the r^2 was better for standard CN flow. The slope expresses the consistent relationship of the two variables in comparison. The slope of the overall study period for CN_{NDVI} prediction was 1.06, indicating that for every 1 increase in CN_{NDVI} runoff, the observed flow increased by 1.06. Therefore, the CN_{NDVI} model underpredicted by approximately 6%. For standard CN in the overall period, for 1 increase in standard CN flow, the observed flow increased by 0.52, demonstrating that the standard CN flow overpredicted by 47%. The intercept showed less error in the CN_{NDVI} runoff model (0.34 mm) than the literature CN (0.72 mm). In the moderate flow year, although the CN_{NDVI} and standard CN flow overpredicted, the CN_{NDVI} model was better than the standard CN. However, low-flow year predictions were the worst statistically. Figures 5-7 and 5-9 show for low-flow year (2003) and high-flow years (2010) and performed poorly compared to overall and moderate flow years.

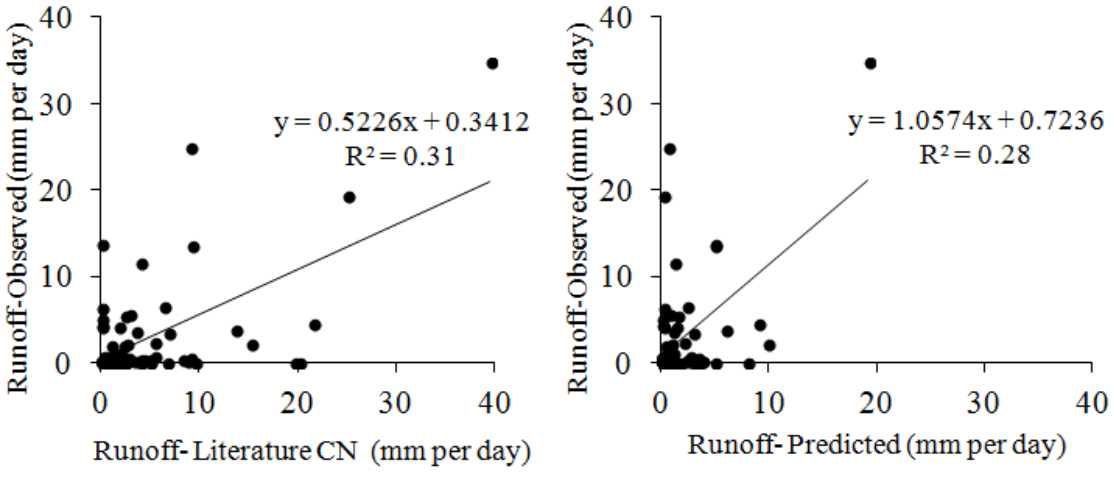


Figure 5-6: Comparison of observed runoff vs. standard CN runoff (left) and observed runoff vs. CNNDVI predicted runoff (right) at Chapman watershed outlet during overall flow years (2003, 2007, and 2010)

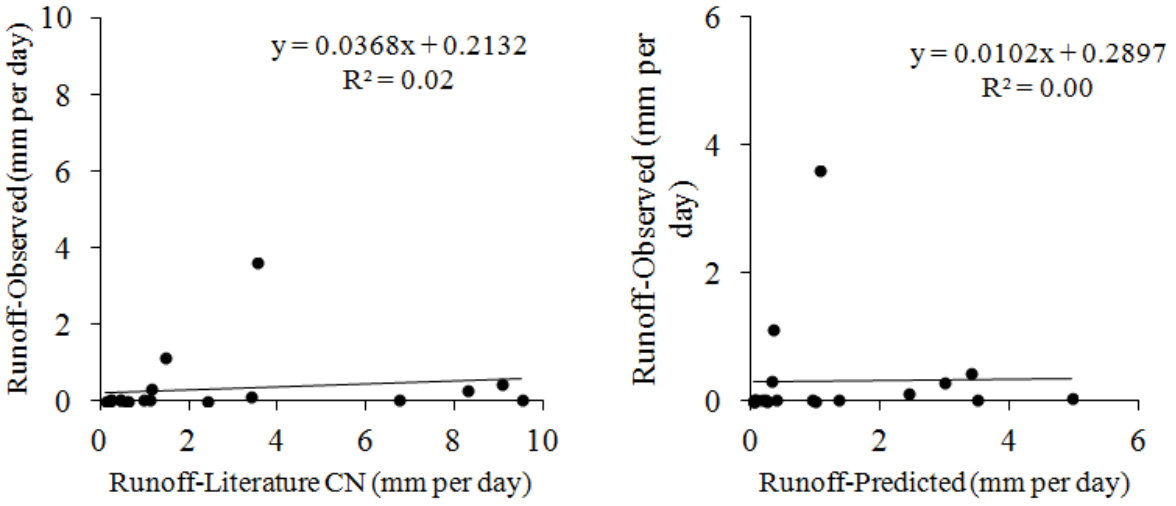


Figure 5-7: Comparison of observed runoff vs. standard CN runoff (left) and observed runoff vs. CNNDVI predicted runoff (right) at Chapman watershed outlet during low flow year (2003)

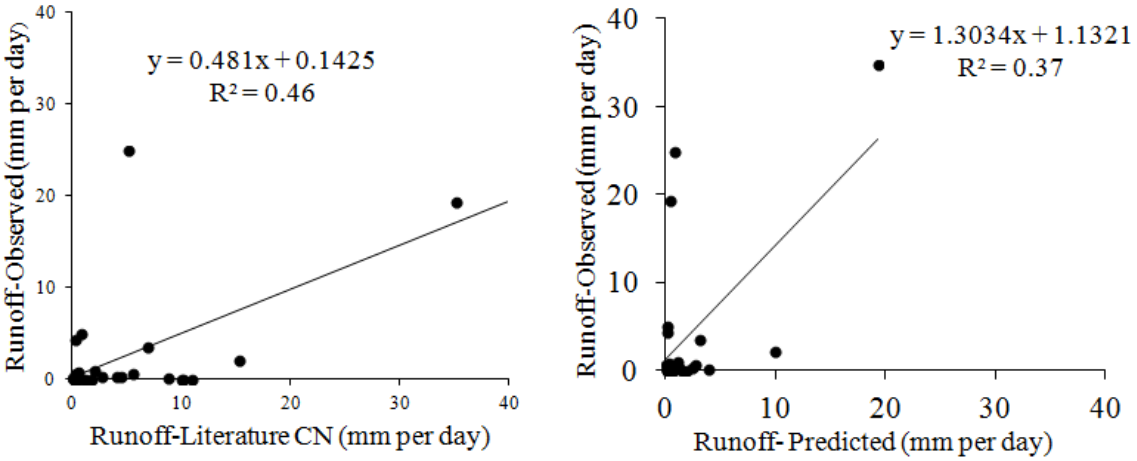


Figure 5-8: Comparison of observed runoff vs. standard CN runoff (left) and observed runoff vs. CNNDVI predicted runoff (right) at Chapman watershed outlet during moderate flow year (2007)

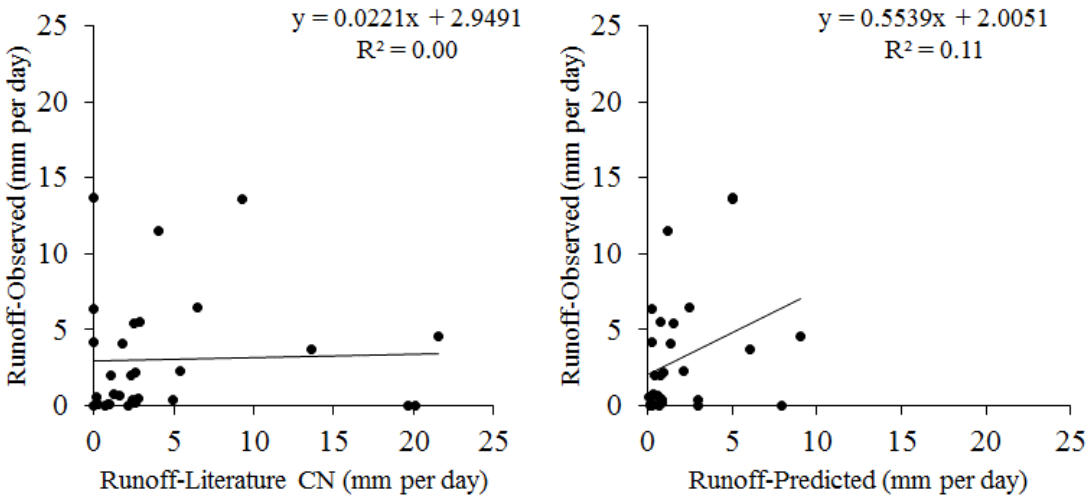


Figure 5-9: Comparison of observed runoff vs. standard CN runoff (left) and observed runoff vs. CNNDVI predicted runoff (right) at Chapman watershed outlet during high flow year (2010)

The same pattern was observed in quantitative statistics measures (Table 5-4). In all conditions, CN_{NDVI} model predictions showed better results than the SCS-CN prediction. For CN_{NDVI} predictions, all quantitative statistics were relatively good throughout the entire study period and for the moderate flow year. NSE values for those periods were positive, PBIAS was comparatively lower, and the correlation coefficient (*r*) was above 0.60.

Table 5-4: Model performance statistics of Chapman watershed

Statistics	Overall Study Period		Yr 2003		Yr 2007		Yr 2010	
	Observed runoff vs. std. CN runoff	Observed runoff vs. CN_{NDVI} runoff	Observed runoff vs. std. CN runoff	Observed runoff vs. CN_{NDVI} runoff	Observed runoff vs. std. CN runoff	Observed runoff vs. CN_{NDVI} runoff	Observed runoff vs. std. CN runoff	Observed runoff vs. CN_{NDVI} runoff
NSE	-0.02	0.26	-21.64	-4.31	-0.28	0.63	-2.44	-0.05
Correlation coefficient (*r*)	0.56	0.53	0.14	0.02	0.68	0.61	0.03	0.33
PBIAS (%)	-65	33.05	-693.26	-274.82	48.01	-41.54	-46.67	38.22

In summary, the moderate flow period runoff, which was approximately 25% greater than long-term average flow (Table 5-2), performed well than the low flow, which was 17% of long-term average flow, and high flow, which was 167% of long-term average flow, runoffs. Results showed that the CN_{NDVI} model accurately predicted moderate flow runoff but not low and high flows or extreme flow runoff of grassland. However, the CN_{NDVI} model predicted runoff better than the standard SCS-CN model (Table 5-4) as that of the overall year period and moderate flow year.

Inability to accurately predict low-flow period runoff may be related to uncertainties in runoff, rainfall measurements, NDVI, initial abstraction, and/or base flow extraction. CN_{NDVI} developed in this study (Chapter 3) used the original SCS I_a/S ratio value of 0.2 as the initial abstraction. In general, results showed that the CN_{NDVI} overpredicted runoff during low runoff period (Figure 5-6b, Table 5-4). Therefore, a higher initial abstraction value may improve runoff prediction and improve other measurement techniques, such as rainfall and runoff, especially

true when rainfall events are sparse and scattered and initial abstraction and infiltration play big role in runoff generation. However, initial abstraction may not able to represent the actual condition. Altering the initial abstraction throughout the season with changing CN may improve the results (Jacobs and Srinivasan, 2005).

Baseflow was separated using the automated recursive digital filter program (or baseflow filter program) (Nathan and McMahon, 1990; Arnold et al., 1995) in this study. Stewart (2015) asserted that extra caution should be used when streamflow is far below low flow and during high flows. During a low-flow period, streamflow is dominated by baseflow, but during a high-flow period, streamflow is composed of comparable amounts of quick- flow and baseflow. Stewart (2015) stated that various mechanisms must be used in baseflow during those periods. Baseflow values calculated during low- and high-flow periods in this study may not sufficiently address these runoff conditions.

Results also showed that higher uncertainties were present during low-flow conditions as compared to regular flow measurements due to the difficulty of measuring low-flow discharge as a result of surface-groundwater exchange and the presence of ice and vegetation (Hamilton, 2008; Shrestha et al., 2013; Sinnathamby, 2014). There may be also a potential source of error in the rainfall data input. MODIS NDVI also contains uncertainties: Light reflectance from the soil surface can influence NDVI values. The effect of soil is expected to be higher in grasslands, which tend to have higher cover of bare ground and exposed rock, especially during drought or low-flow periods (Huete and Jackson, 1988). The good performance demonstrated throughout the entire study period could be explained by the increase in the number of events compared to the individual flow period analysis.

5.4.1.2 Runoff Estimation for Agricultural Dominated Watershed

Results similar to the grassland dominated area were observed in the Upper Delaware, which is an agriculture dominated watershed (Figure 5-10–5-13 and Table 5-5). Both r^2 and statistics considered in this study (NSE, *r*, and PBIAS) were better for CN_{NDVI} runoff predictions compared to literature-based CN runoff predictions. Overall, CN_{NDVI} model performance was relatively better for the grassland dominated watershed than the agricultural dominated watershed (Table 5-4 and 5-5, NSE, and *r*). However, PBIAS (%) for the Upper Delaware was better than Chapman Creek for most of the condition due to the NDVI uncertainty for grassland (Heute and Jackson, 1988). Moderate flow period performed very well also performed very well for grassland. CN_{NDVI} moderate flow demonstrated an increase of observed flow by 0.78 for the predicted flow increase by 1 mm, and the intercept showed less error in the model (i.e., 0.84 mm). For the overall years of study, for every 0.62 increase in observed flow, the predicted flow increased by 1. The moderate year was the best performing relatively.

Based on all results, the CN_{NDVI} model overpredicted runoff for grassland and agriculture land areas during low-flow periods. However, the CN_{NDVI} model performed better than the SCS-CN-based predictions, even at low flows. Similar reasons explained for Chapman Creek low flow period could explain the overpredicted runoff in Upper Delaware runoff during runoff. The improved results for the combined study period were due to the dramatic increase in the number of events, as with Chapman Creek.

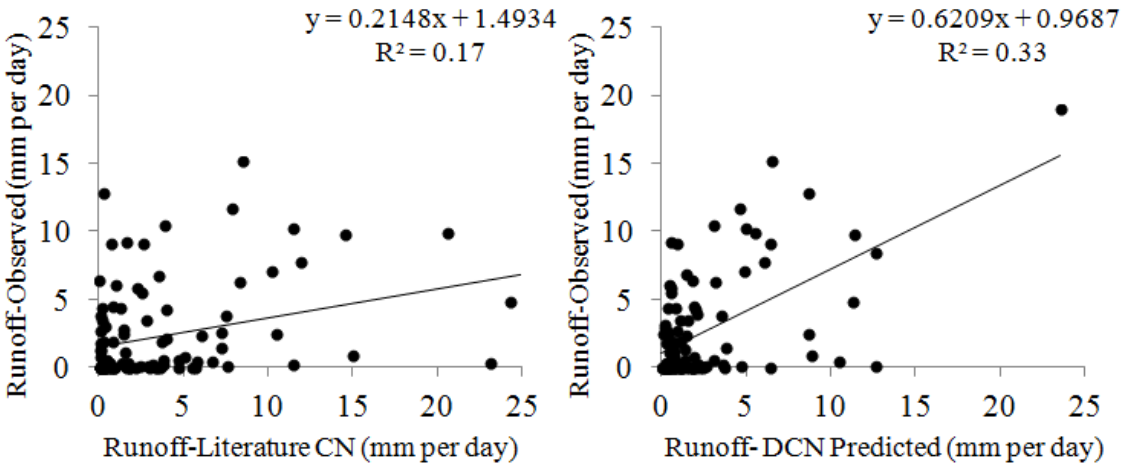


Figure 5-10: Comparison of observed runoff vs. literature CN runoff (left) and observed runoff vs. CNNDVI predicted runoff (right) at Upper Delaware watershed outlet during overall study years (2003, 2007, and 2010)

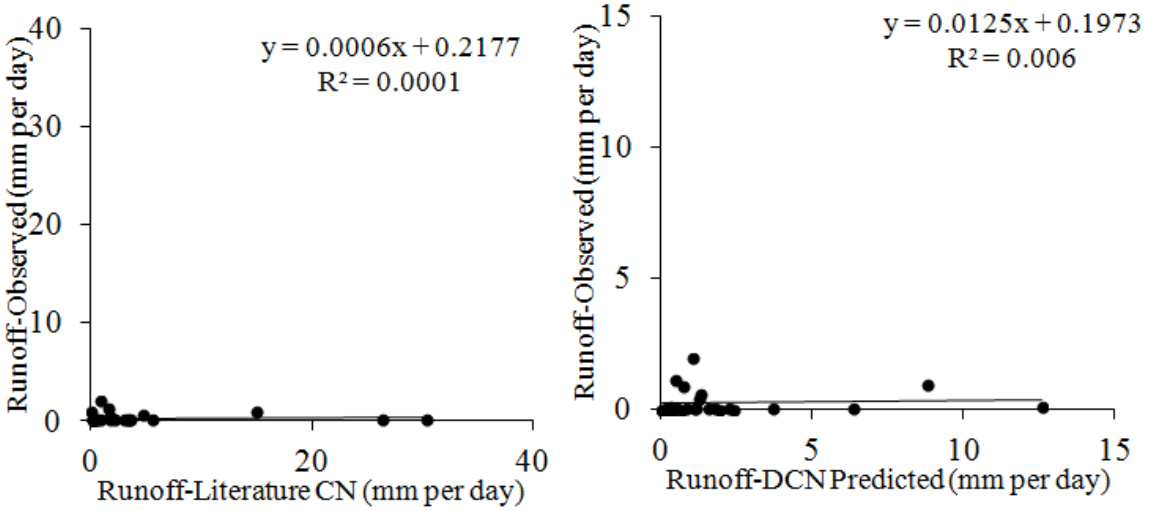


Figure 5-11: Comparison of observed runoff vs. literature CN runoff (left) and observed runoff vs. CNNDVI predicted runoff (right) at Upper Delaware watershed outlet during low flow year (2003)

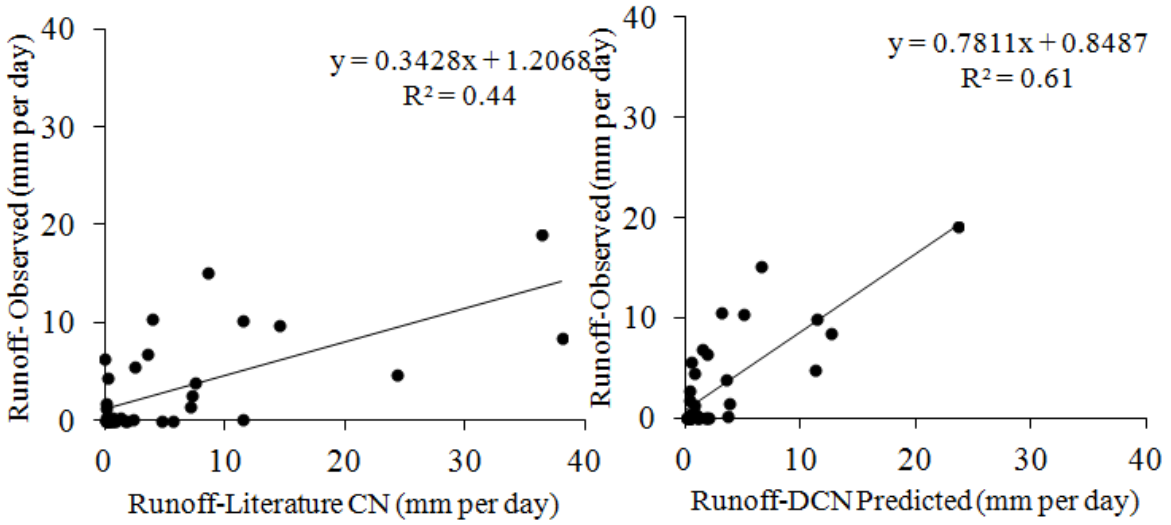


Figure 5-12: Comparison of observed runoff vs. literature CN runoff (left) and observed runoff vs. CNNDVI predicted runoff (right) at Upper Delaware watershed outlet during moderate flow year (2007)

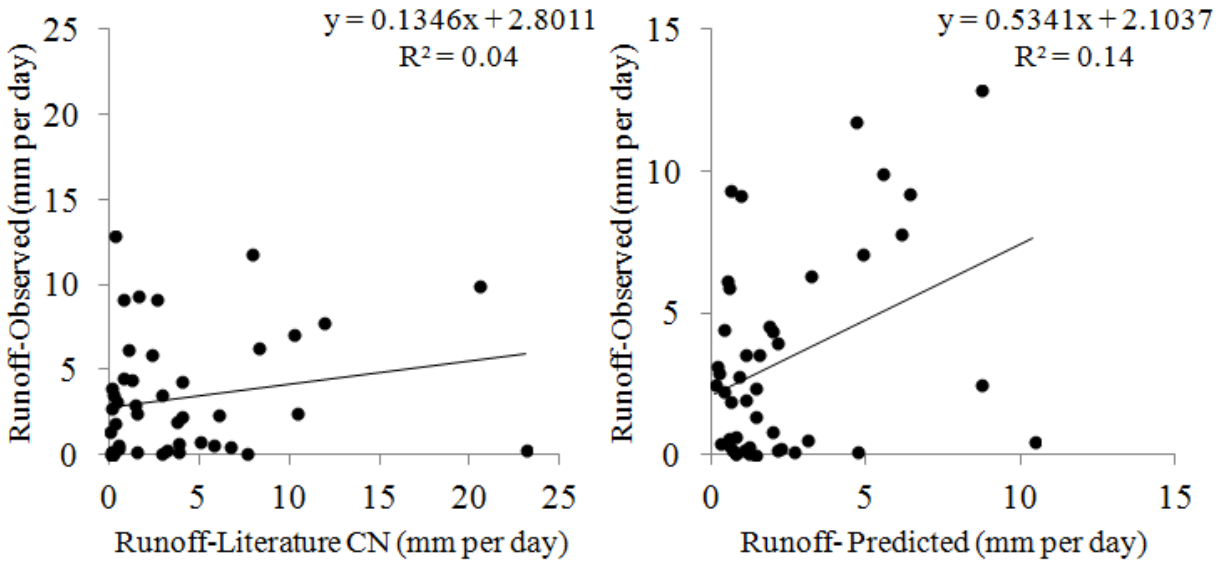


Figure 5-13: Comparison of observed runoff vs. literature CN runoff (left) and observed runoff vs. CNNDVI predicted runoff (right) at Upper Delaware watershed outlet during high flow year (2010)

Table 5-5: Model performance statistics of Upper Delaware watershed

Statistics	Study Period		Yr 2003		Yr 2007		Yr 2010	
	Observed runoff vs. std. CN runoff	Observed runoff vs. CN_{NDVI} runoff	Observed runoff vs. std. CN runoff	Observed runoff vs. CN_{NDVI} runoff	Observed runoff vs. std. CN runoff	Observed runoff vs. CN_{NDVI} runoff	Observed runoff vs. std. CN runoff	Observed runoff vs. CN_{NDVI} runoff
NSE	-2.36	0.20	-327.80	-50.44	-1.40	0.56	-1.51	-0.06
Correlation Coefficient, r	0.41	0.57	0.01	0.07	0.66	0.77	0.19	0.38
PBIAS (%)	-74.69	4.22	-241.71	-271.71	-72.48	8.76	-12.45	32.08

5.4.2 Estimation of Seasonal Variability

In general, the CN_{NDVI} model underpredicted runoff for spring, summer, and autumn and overpredicted runoff for winter in grassland. Literature CN-based runoff was higher all the times. The seasonal change in runoff was captured more in CN_{NDVI} than the standard CN-based runoff. For Chapman Creek, the grassland dominated watershed, the model flow was significantly different than the observed flow during the winter period. In all other seasons, however, modeled runoff values were closely matched to observed values (Figure 5-14).

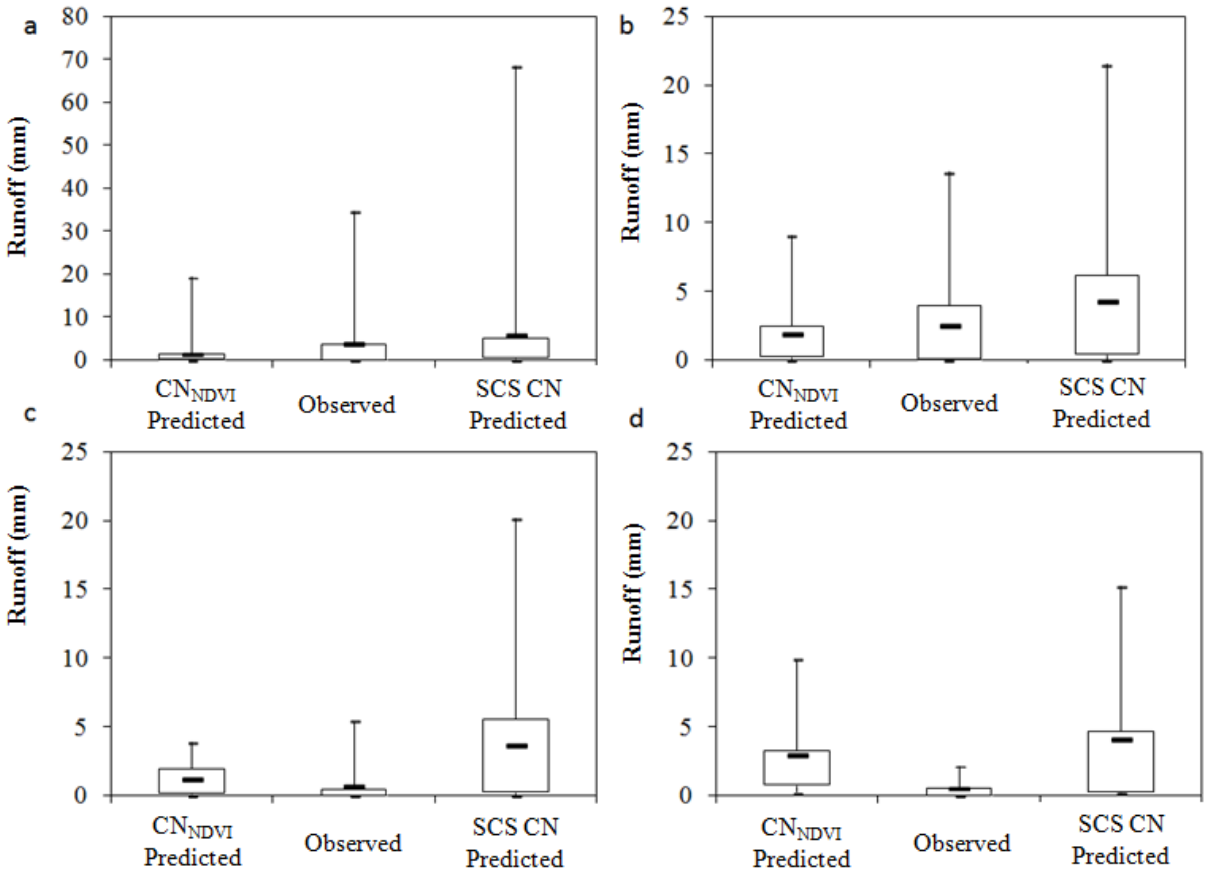


Figure 5-14: Comparison of CNNDVI predicted, observed, and literature CN-based runoff variability for a) spring, b) summer, c) autumn, and d) winter for grassland dominant (Chapman Creek) watershed

In the Upper Delaware watershed, cropping seasons were identified as August to October for hay, April to October for corn, and May to October for soybean according to K-State Extension documents. Therefore, any type of crop was present in the Upper Delaware watershed from April to October, including spring, summer, and part of autumn. NDVI-captured variation in CN values on a seasonal basis may result observed overall model performance on runoff prediction (Price, 1998; Van Mullem et al., 2002). Figure 5-15 shows seasonal flow of observed, CN_{NDVI}, and standard CN boxplots.

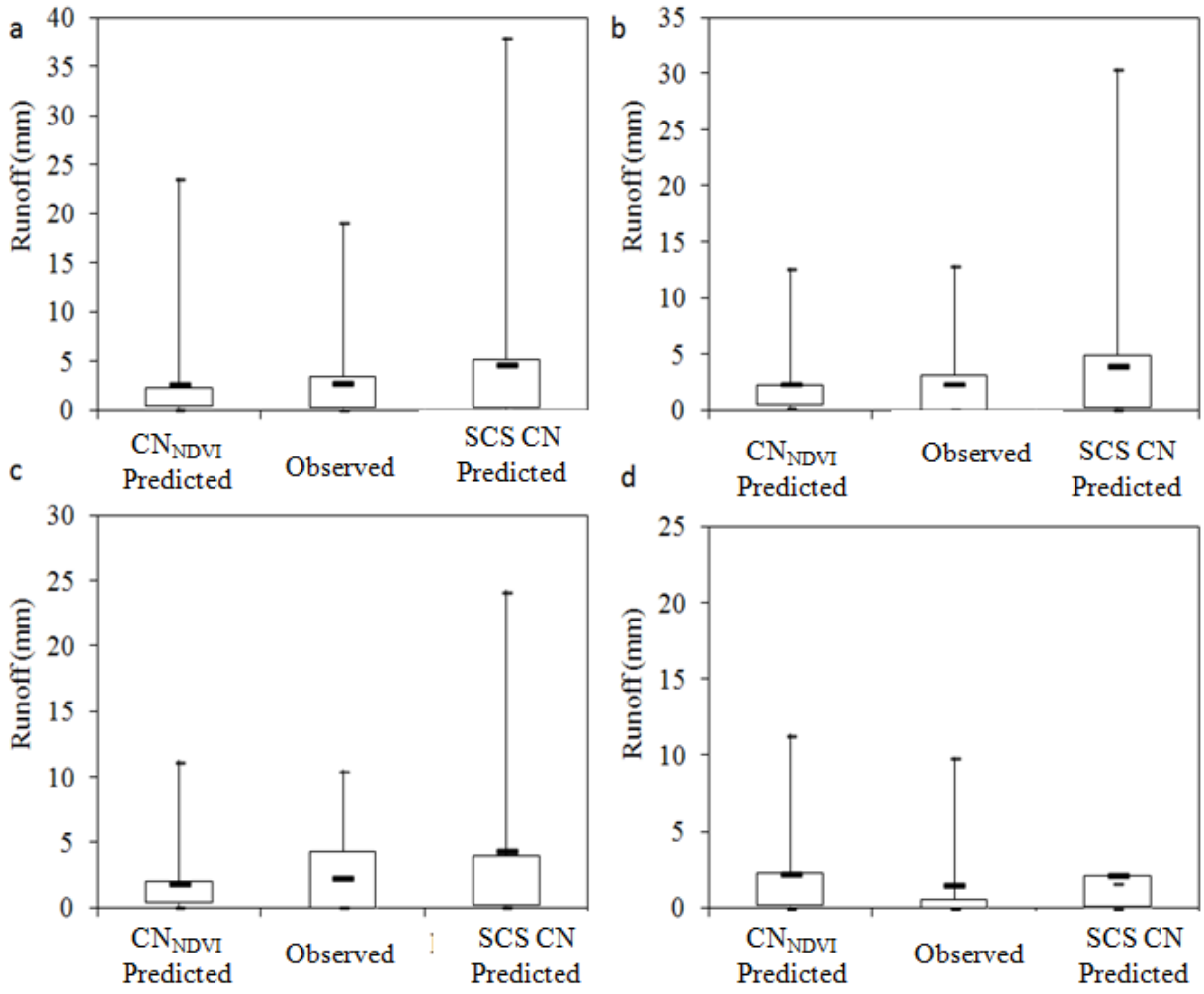


Figure 5-15: CNNDVI predicted, observed, and literature CN-based runoff variability for a) spring, b) summer, c) autumn, and d) winter for agricultural land dominant (Upper Delaware) watershed

Figure 5-16 shows spatiotemporal maps of three selected NDVI periods in 2007 for Chapman and Upper Delaware watersheds. The maps show that the CN_{NDVI} increased from March to September. However, the lowest CN_{NDVI} (higher NDVI) values were observed at different periods. The lowest CN_{NDVI} for Chapman was observed during the end of May and the beginning of June; for the Upper Delaware watershed, the lowest CN_{NDVI} was in September, potentially due to the fact that Upper Delaware is agriculture dominated and the beginning of September is the peak greenness period of the area. The increased variability observed March

22–April 6 at both watersheds CN_{NDVI} could be explained by the spatial variability of slope and land use/cover. Significant higher CN_{NDVI} values were observed along riparian vegetation from August 29 to September 13 in Chapman Creek due to increased soil moisture.

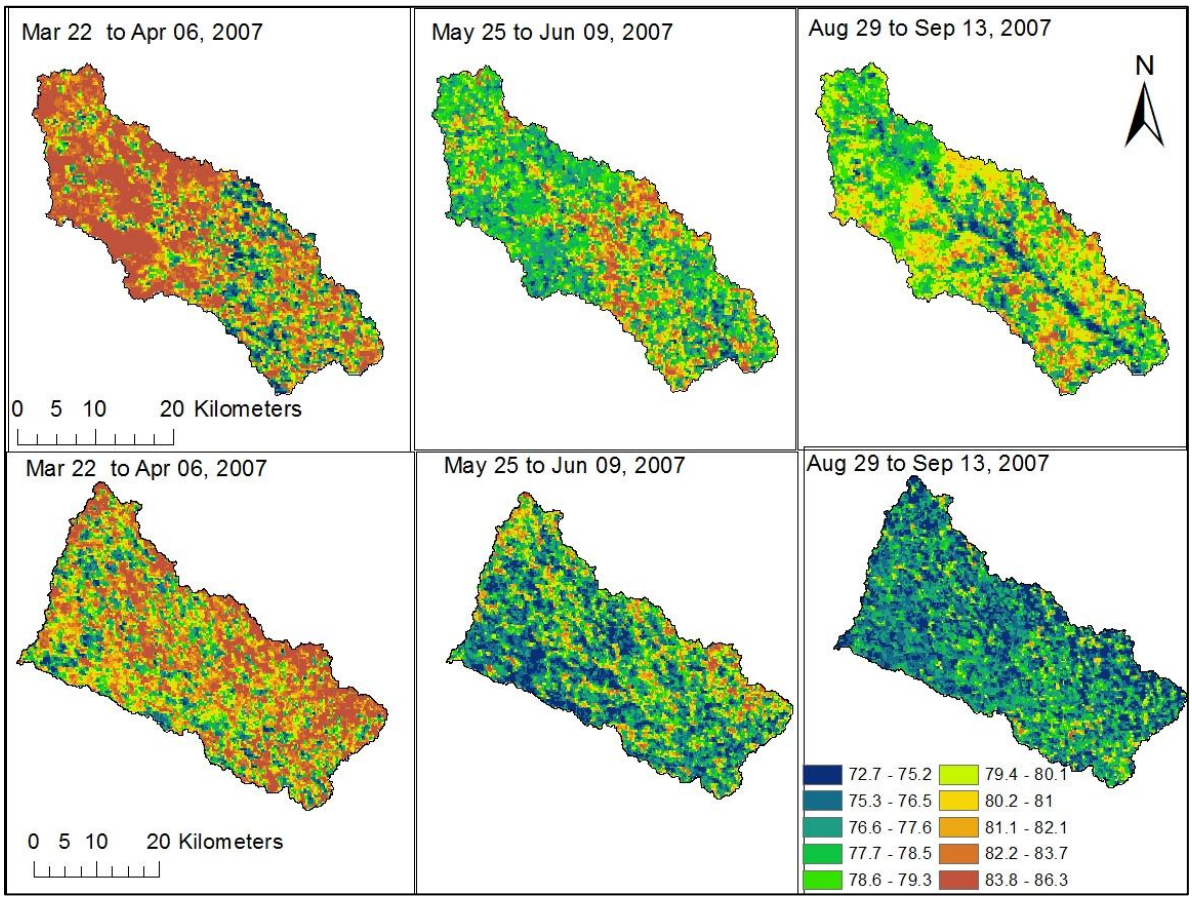


Figure 5-16: CN based on 16-day NDVI periods in 2007: Chapman (upper) and Upper Delaware (lower)

VI quality analysis of these six periods (or images) revealed that NDVI used in this analysis were in good condition (pixel reliability = 0) and no clouds contamination was present. VI quality analysis utilized pixel reliability ranking imagery for the maps in Figure 5-16. Therefore, March 22, 2007, Chapman Creek watershed variability of upper and lower parts of the watershed was due to land use changes. The upper part of the watershed is grassland, and the lower part is cropland.

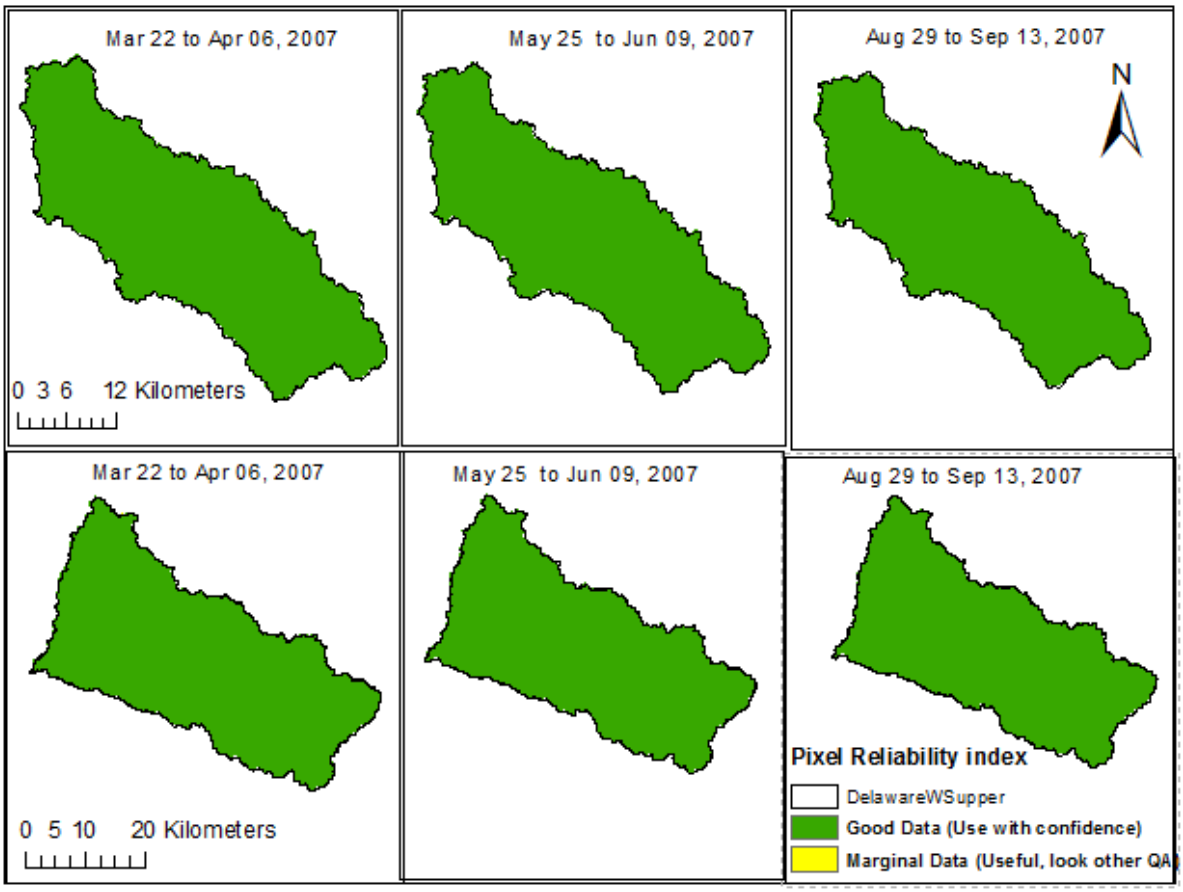


Figure 5-17: Pixel reliability ranking of MODIS VIs (MOD13Q1) at Chapman Creek watershed (upper) and Upper Delaware watershed (lower)

5.5 Conclusion

Although the SCS-CN method is widely applied, modification is required. The CN_{NDVI} model developed in Chapter 3 was applied to study the performance of CN_{NDVI} in larger and more diverse land use/cover watersheds. Results showed that the CN_{NDVI} model predicted moderate-flow runoff well than the low flow and high flow runoffs at larger scales for grassland and agricultural watersheds. Inability to accurately predict low-flow period runoff may relate to initial abstraction, baseflow extraction, or uncertainties in runoff, rainfall measurements, and NDVI. However, the CN_{NDVI} model more accurately predicted runoff than the standard SCS-CN model. The spatiotemporal pattern observed in CN_{NDVI} revealed that CN_{NDVI} captured

spatiotemporal variability of the rainfall-runoff relationship, thereby improving runoff modeling at larger scale. Use of Daymet and CN_{NDVI}, as shown in this study, could provide opportunity for new applications of a modified CN model with improved runoff estimation results. This method could be used to estimate runoff in a real-time fashion if an online model is developed based on this analysis. The method could also be automated and results could be used in water resource planning and management. Longer study years may improve the analysis.

References

Arnold, J. G. and P. M. Allen (1999). Automated methods for estimating baseflow and ground water recharge from streamflow records. *Journal of the American Water Resources Association* 35(2): 411–424.

Bosworth, M. (2007). Delaware River Watershed Restoration and Protection Strategy (WRAPS). http://www.kswraps.org/files/attachments/delaware_plansummary.pdf

Baltas, E. A., Dervos, N. A., and M. A. Mimikou (2007). Determination of the SCS initial abstraction ratio in an experimental watershed in Greece. *Hydrology and Earth System Sciences* 11:1825–1829.

Canters, F., Chormanski, J., Van de Voorde, T., and O. Batelaan (2006). Effects of Different Methods for Estimating Impervious Surface cover on Runoff Estimation at Catchment Level. In Proceedings of 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, Lisbon, Portugal, July 5–7, 2006: 557–566.

D'Asaro, F., G. Grillone, and R. H. Hawkins (2014). Curve Number: Empirical evaluation and comparison with Curve Number handbook tables in Sicily. *Journal of Hydrologic Engineering* doi: 10.1061/(ASCE)HE.1943-5584.0000997

Epps, T. H., Hitchcock, D. R., Jayakaran, A. D., Loflin, D. R., Williams, T. M., and D. M. Amatya (2013). Curve number derivation for watersheds draining two headwater streams in lower coastal plain South Carolina, USA. *Journal of the American Water Resources Association* 49:1284–1295.

Fan, F., Deng, Y., Hu, X., and Weng, Q. (2013). Estimating Composite Curve Number Using an Improved SCS-CN Method with Remotely Sensed Variables in Guangzhou, China. *Remote Sens.* 5:1425–1438.

Garen, D. C. and D. S. Moore (2005). Curve number hydrology in water quality modeling: uses, abuses, and future directions. *Journal of the American Water Resources Association*. 41:377–88.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrologic Eng.* 4(2): 135–143.

Hjelmfelt, A. T., Jr. (1980). An empirical investigation of curve number technique. *Journal of the Hydraulics Division* 106–9:1471–1476.

Hong, Y. and R. F. Adler (2008). Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *International Journal of Remote Sensing* 29:471–477.

Kousari, M. R., H. Malekinezhad, H. Ahani, and M. A. A. Zarch (2010). Sensitivity analysis and impact quantification of the main factors affecting peak discharge in the SCS curve number method: An analysis of Iranian watersheds. *Quaternary International* 226:66–74.

Hawkins, R. H. (2014). Curve Number Method: Time to Think Anew? *Journal of Hydrologic Engineering* 19(6): 1059–1059.

Hawkins, R. H., R. Jiang, D. E.Woodward, A. T. Hjelmfelt, and J. E.Van Mullem (2002). Runoff curve number method: Examination of the initial abstraction ratio. U.S. Geological Survey Advisory Committee on Water Information Second Federal Interagency Hydrologic Modeling Conference. July 28–August 1, (2002), Las Vegas, Nevada.

Hong, Y. and R. F. Adler (2008). Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *International Journal of Remote Sensing* 29:471–477.

Huete, A. R. and Jackson, R. D. (1988). Soil and atmosphere influences on the spectra of partial canopies. *Remote Sensing of the Environment* 25:89–105.

Jacobs, J. H. and Srinivasan, R. (2005). Effects of Curve Number Modification on Runoff Estimation Using WSR-88D Rainfall Data in Texas Watersheds. *Journal of Soil and Water Conservation* 60:274–278.

Juracek, K. E. (1999). Estimation of potential runoff-contributing areas in the Kansas-lower Republican River Basin, Kansas: U.S. Geological Survey Water Resources Investigations Report 99–4089.

Rasmussen, T. J. and C. A. Perry (2001). Estimation of Peak Streamflows for Unregulated Rural Streams in Kansas. Water-Resources Investigations Report 00–4079. U.S. Department of the Interior and U.S. Geological Survey.

Reistetter, J. A. and M. Russell (2011). High-resolution land cover datasets, composite curve numbers, and storm water retention in the Tampa Bay, Florida region. *Applied Geography* 31:740–747.

Santhi, C., N. Kannan, J. G. Arnold, and M. Di Luzio (2008). Spatial calibration and temporal validation of flow for regional-scale hydrologic modeling. *Journal of the American Water Resources Association* 44(4): 829–846.

Shaw, S. B., Makhlouf, R., Walter, M. T., and J. Y. Parlange (2008). Experimental testing of a stochastic sediment transport model. *Journal of Hydrology* 348:425–430.

Sophocleous, M. A. and B. B. Wilson (2000). Surface water in Kansas and its interaction with groundwater: accessed January 03, 2016, at URL http://www.kgs.ukans.edu/HighPlains/atlas/atswqn.htm

Moriasi, D., J. Arnold, M. Van Liew, R. Bingner, R. Harmel, and T. Veith (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASAE* 50(3): 885–900.

Nathan, R. and McMahon, T. (1990). Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology* 121(1): 217–238.

Natural Resources Conservation Service (NRCS) (2007). Kansas Annual Precipitation. United States Department of Agriculture. KS-08-22. Salina, KS: Resource Conservation Staff. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_032018.pdf (last accessed November 25, 2015)

Pineiro, G., Perelman, S., Guerschman, J. P., and Paruelo, J. M. (2008). How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling* 2 1 6:316–322.

Sinnathamby, S. (2014). Modeling tools for eco-hydrological characterization. Dissertation, Doctor of Philosophy, Kansas State University.

Song, X. M., Zhan, C. S., Kong, F. Z., and J. Xia (2011). Advances in the study of uncertainty quantification of large-scale hydrological modeling system. *Journal of Geographical Sciences* 21(5): 801–819.

Stewart, M. K. (2015). Promising new baseflow separation and recession analysis methods applied to streamflow at Glendhu Catchment, New Zealand. *Hydrol. Earth Syst. Sci.* 19:2587–2603.

Soulis, K., J. Valiantzas, N. Dercas, and P. Londra (2009). Analysis of the runoff generation mechanism for the investigation of the SCS-CN method applicability to a partial area experimental watershed. *Hydrol Earth Syst Sci*. 6:373–400.

Thornton, P. E., M. M. Thornton, B. W. Mayer, N. Wilhelmi, Y. Wei, R. Devarakonda, and R. B. Cook (2014). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2. ORNL DAAC, Oak Ridge, Tennessee, USA. Accessed Month DD, YYYY. Time period: YYYY-MM-DD to YYYY-MM-DD. Spatial range: N=DD.DD, S=DD.DD, E=DDD.DD, W=DDD.DD. http://dx.doi.org/10.3334/ORNLDAAC/1219

Thornton P. E., Running S. W., and White M. A. (1997). Generating surfaces of daily meteorology variables over large regions of complex terrain. *Journal of Hydrology* 190:214–251.

Woodward, D. E., R. H. Hawkins, A. T. Hjelmfelt, J. A. VanMullem, and Q. D. Quan (2002). Curve number method: Origins, applications, and limitations. U.S. Geological Survey Advisory Committee on Water Information - Second Federal Interagency Hydrologic Modeling Conference. July 28–August 1, (2002), Las Vegas, Nevada.

Yuan, Y., W. Nie, S. C. McCutcheon, and E. V. Taguas (2014). Initial abstraction and curve numbers for semiarid watersheds in Southeastern Arizona. *Hydrological Processes* 28(3): 774–783.

Chapter 6 - Summary and Conclusion

Land surface and its transformation plays a fundamental role in modulating the atmospheric, geomorphic, hydrologic, and ecological processes on or near the earth's surface so that understanding of the land surface could provide valuable information on natural process functions (Wilson, 2012). Quantifying hydrological processes and accounting for their spatiotemporal variability due to changes in climate, occurrences of variable extreme events and landscape disturbances on hydrological processes are challenging tasks that require tremendous parameterizations of different processes and input. Understanding and predicting the rainfall-runoff relationship is crucial for water resources planning and management; and to understand processes of erosion, sediment transport, and contaminant loading. However, the rainfall-runoff process is a complex, dynamic and nonlinear process (Fan et al., 2015; Song et al., 2011) to quantify in any universal way. How much runoff generated from a certain rainfall event is dependent on the hydrologic condition of an area in addition to the amount, intensity and duration of the rainfall. Reliable predictions of amount and rate of runoff from land surface is becoming a difficult and time consuming process as its complexity increases.

The overall goal of the study was to develop a mechanism to improve surface water hydrologic modeling through incorporating spatiotemporal variability of rainfall-runoff relationship; and devise statistical method for hydrologically homogenized selection of paired watersheds. The first study tried to overcome the inadequate representation of spatiotemporal variability of the rainfall-runoff because of static nature curve number.

The SCS-CN method is a popular method to estimate runoff events; is determined based on land use/cover, hydrologic soil groups, agricultural treatments, and hydrologic conditions of the area. Even though the SCS-CN method is widely applied and incorporated in numerous

hydrologic models (Kousari et al., 2010; Hawkins et al., 2008; Woodward et al., 2002); the SCS-CN method does not adequately reflect the detailed spatiotemporal variability of the rainfall-runoff relationship because of limited spatial and temporal data availability. The study used Moderate Resolution Imaging Spectroradiometer – Normalized Difference Vegetation Index (MODIS-NDVI) and back-calculated CN using regression analysis to develop CN_{NDVI}. This study have used NDVI to develop CN_{NDVI} to better estimate or predict the runoff from rainfall events and address the spatiotemporal variability of rainfall-runoff relationship based on a 12 years observed precipitation and runoff. The pixel reliability ranking and VI quality pixel bit numbers were used to monitor quality of all NDVI used. In addition, the developed CN_{NDVI} have applied a similar land cover smaller watershed, but very different management and disturbance processes and in larger watersheds with different land use/cover. Result showed that derived models behave satisfactorily during calibration and validation period irrespective of the fact that the rainfall event-runoff relationship is so complex. It was also noticed that the calibrated model has improved the prediction of runoff compared to the existing standard CN/static method.

The developed CN_{NDVI} has predicted statistically better runoff from rainfall events compared to the standard CN based runoff. The coefficient of correlation (r) which measures the strength of the linear relationship between the observed flow and estimated flow supported the analysis. The analysis showed the r value of 0.40 for the standard CN based flow; 0.48 of regression model flow; and 0.85 for calibrated model flow in a relationship with the observed (measured) flow, respectively. The CN_{NDVI} showed a vivid change during the beginning of the growing season (late March to April) with the lowest CN_{NDVI} around June 10. The CN_{NDVI} standard deviation was greatest during the spring (April 23), which was expected based on the varying spring weather and changes of the landscape frost; and wider during the growing season

in general where vegetation health varies based on various weather conditions (e.g. available soil moisture). The variability of CN_{NDVI} in the growing season could give a chance to calculate relatively accurate runoff from rainfall events since most rainfall occurs from during the season.

The study has also provided insights on the suitability of developed CN_{NDVI} to estimate runoff in relatively disturbed grassland with similar size watersheds, and with larger and diverse watersheds of grassland and agriculture dominated areas. The results show that, CN_{NDVI} model was able to estimate runoff for different land use watersheds especially in moderate runoff condition especially at larger scale. However, CN_{NDVI} model was not able to predict extremely low and higher flow period runoffs accurately. Inability to accurately predict low flow period runoff might relate to initial abstraction or baseflow extraction or uncertainties in runoff, rainfall measurements and NDVI. CN_{NDVI} model also was able to capture seasonal variability in runoff of both watersheds. These results suggest that the CN_{NDVI} and Daymet data could be useful in capturing the spatiotemporal variability of runoff.

In general, the model is able to produce better CN_{NDVI} than the popularly used SCS-CN values to predict runoff for a rainfall event. The method used in this study can be adjusted (based on the local data) and used in any watersheds and climatic conditions to develop CN_{NDVI} method. Improved rainfall and runoff measurement techniques for various soil and climatic conditions and issues related to the improvement of satellite data (especially higher resolution NDVI to capture the land use/cover variability on the ground) could improve regression model development.

Predicting CN using any variable is more complicated than implied by a linear regression, however in this study; NDVI was indicated to be a better predictor than the existing standard CN method. Modeling CN based runoff using satellite data can be improved as the

resolution of the satellite data is higher. In addition, the advancement of remote sensing might provide other hydrologic variables to incorporate in the model such as soil moisture and evapotranspiration. Improved CN_{NDVI} method could be automated with online website to estimate runoff from rainfall events for better water resources planning and management.

In addition to development of CN_{NDVI}, the study conducted a paired watershed selection that could be used to study land use/cover change impacts in hydrology. Eight hydrologically important topographic variables including one soil variable (total stream length, drainage density, ruggedness, total mean slope, no flat area slope, percent flat area, curvature, and percent clay in the upper layer) and combined K-means and hierarchical-agglomerative clustering techniques used in this study give a promising method of paired watershed selection as long as appropriate caution is done in selecting major variables affecting the hydrologic processes on the study watersheds.

6.1 Recommendations

- Improved rainfall and runoff measurement techniques, especially during low rainfall periods, and improved satellite data (such as higher resolution NDVI to capture land use/cover variability on the ground) could improve regression model development. Low-flow measurements have higher uncertainty than regular flow due to higher variability because of ground-surfacewater (or base flow) interactions and the presence of ice and vegetation.
- Inclusions of various distinct land uses and long set of precipitation and runoff data from a variety of climatic conditions would improve model performance.
- This analysis contained limitations due to the complex nature of the rainfall-runoff relationship. The model attempted to incorporate the effects of various factors into one parameter, NDVI, but possible errors in precipitation and runoff estimations could have been

present on low values since the data was in the low range. Those uncertainties may be associated with measurement technique precision and environmental conditions, such as wind effects. In addition, the coarse spatial scale of NDVI could add errors and uncertainty.

- Using different initial abstractions for various seasons when developing CN_{NDVI} could improve runoff prediction. Testing runoff estimation with different initial abstractions in different maneuver intensities may also be valuable.
- This study intended to use only readily available MODIS NDVI to predict CN, but the advancement of remote sensing may provide other hydrologic variables to incorporate into the model, such as soil moisture and ET.
- Stream flow characteristics, if available, could be a better parameter in the paired watershed selection process.

References

Fan, F. M., Collischonn, W., Jiménez, K. Q., Sorribas, M., Buarque, D., and Siqueira, V. (2015). Flood forecasting on the Tocantins River using ensemble rainfall forecasts and real-time satellite rainfall estimates. *Journal of Flood Risk Management*. DOI: 10.1111/jfr3.12177

Hawkins, R. H., Ward, T. J., Woodward, D. E., and Van Mullem, J. A. (2008). Curve number hydrology: State of the art. ASCE Publication.

Kousari, M. R., Malekinezhad, H., Ahani, H., and Zarch, M. A. A. (2010). Sensitivity analysis and impact quantification of the main factors affecting peak discharge in the SCS curve number method: An analysis of Iranian watersheds. *Quaternary International* 226(1): 66–74.

Song, X., Kong, F., Zhan, C., and Han, J. (2011). Hybrid optimization rainfall-runoff simulation based on xinanjiang model and artificial neural network. *Journal of Hydrologic Engineering* 17(9): 1033–1041.

Wilson, J. P. (2012). Digital terrain modeling. *Geomorphology* 137(1): 107–121.

Woodward, D. E., R. H. Hawkins, A. T. Hjelmfelt, J. A. VanMullem, and Q. D. Quan (2002). Curve number method: Origins, applications, and limitations. U.S. Geological Survey Advisory Committee on Water Information - Second Federal Interagency Hydrologic Modeling Conference. July 28–August 1, (2002), Las Vegas, Nevada.

Appendix A- Chapter 3 Additional Information

Figure A-1: Fit diagnostics for NDVI back-calculated CN

Number of Observations Read	173
Number of Observations Used	173

Root MSE	11.62292	**R-Square**	0.1445
Dependent Mean	77.86529	**Adj R-Sq**	0.1395
Coeff Var	14.92695		

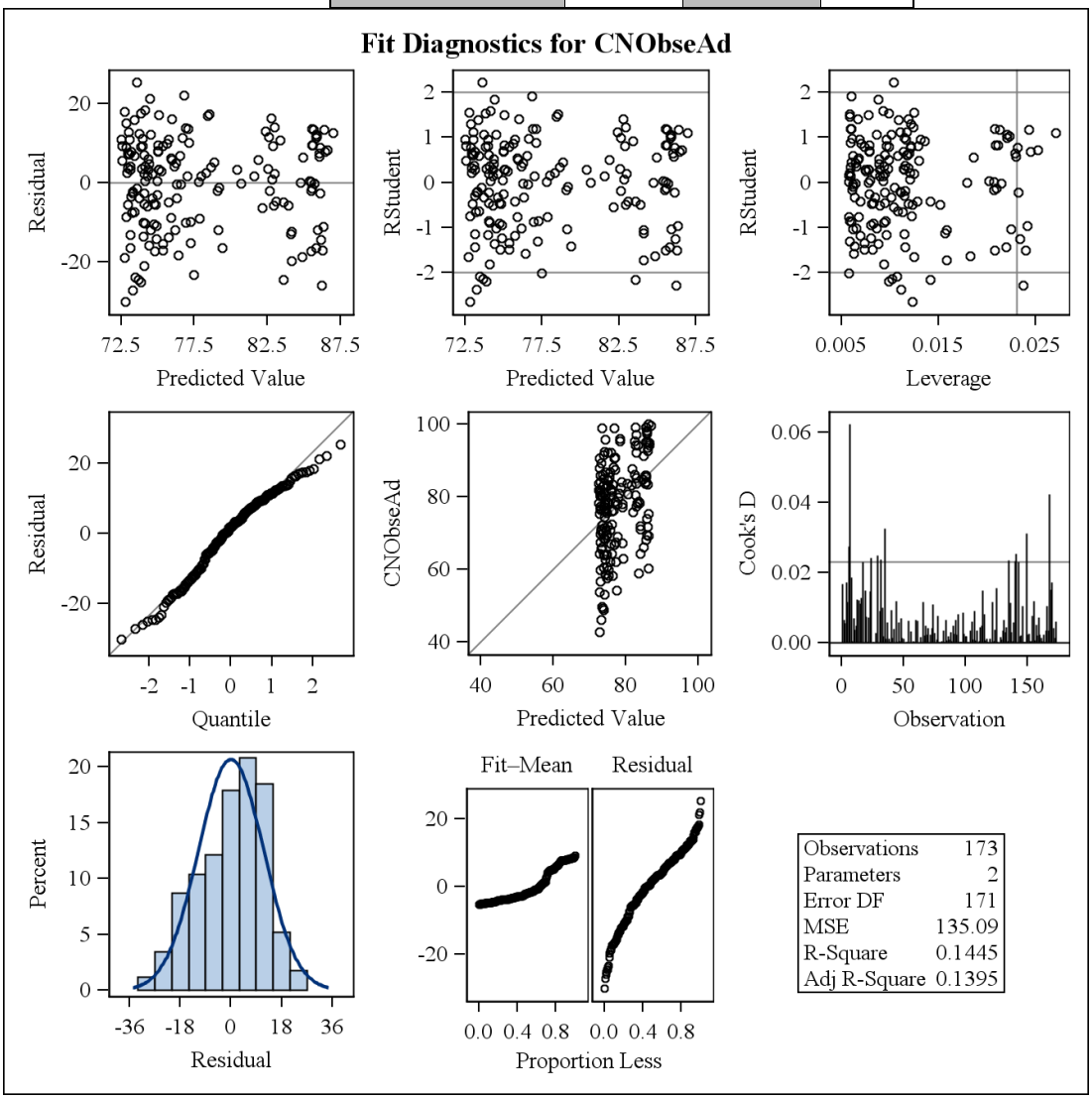


Figure A-2: Residuals

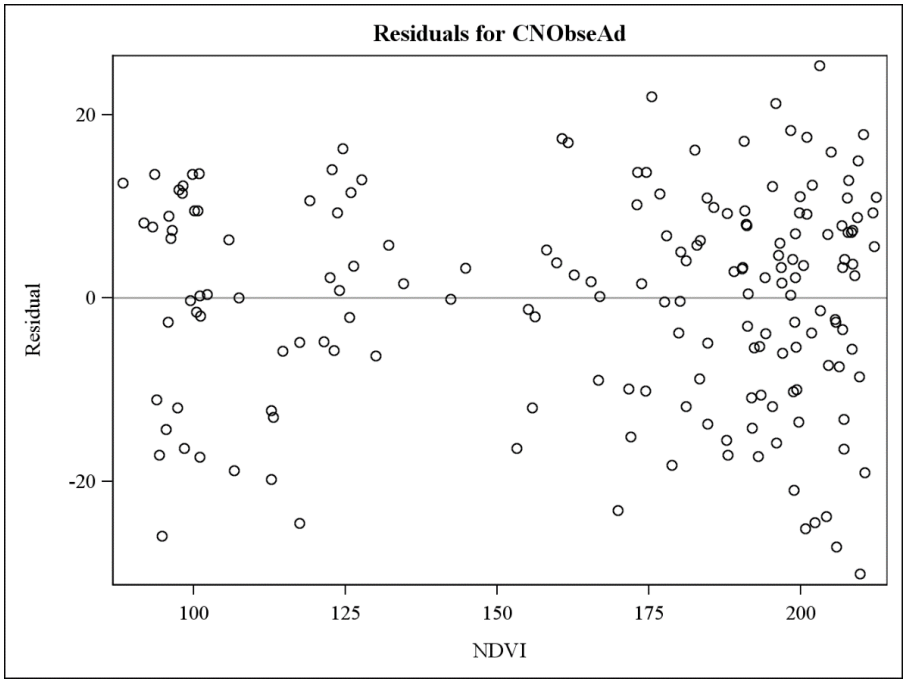


Figure A-3: NDVI versus observed CN

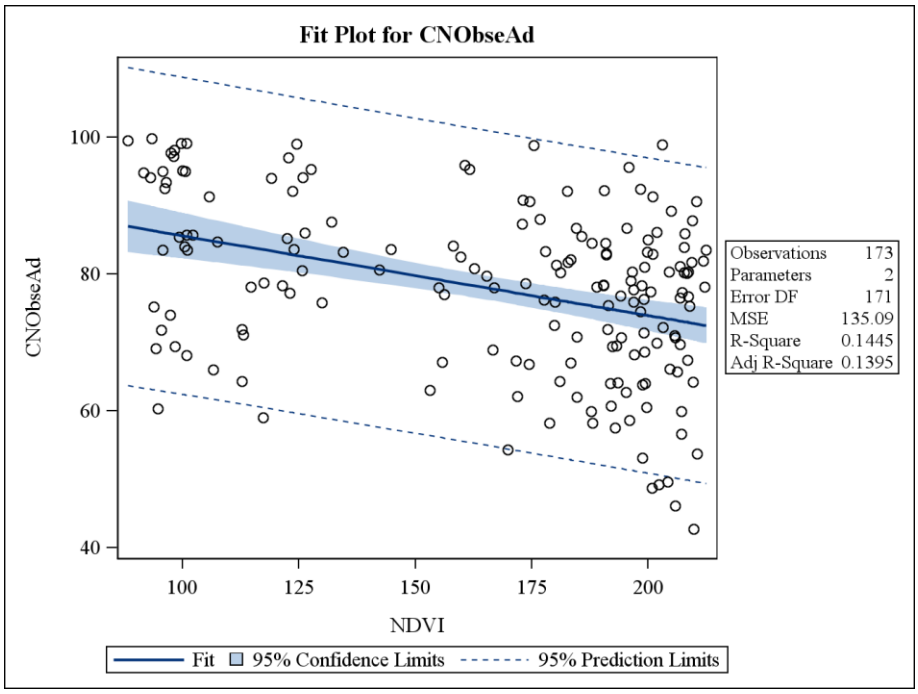


Figure A-4: Regression residuals

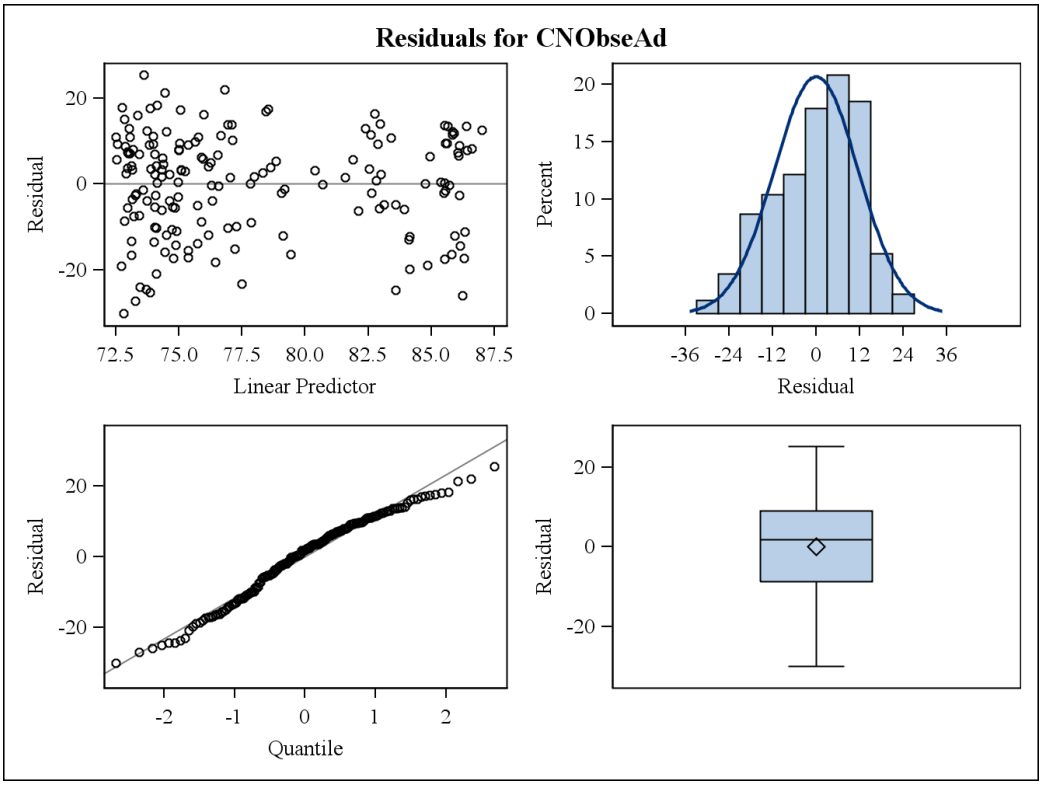


Table A-1: Pixel realization rank values and their interpretation in Konza Prairie watersheds (Adopted from NASA LP DAAC, 2013)

Pixel Reliability Rank	Summary QA	Description
-1	Fill/No data	Not processed
0	Good data	Use with confidence
1	Marginal data	Useful but look at other QA information
2	Snow/Ice	Target covered with snow/ice
3	Cloudy	Target not visible, covered with cloud
4	Estimated	Based on MODIS historic time-series. All products are gap-filled, indicating whether or not the value was interpolated from long-term averages.

Table A-2: VI quality Bit-No, parameter, Bit-word and their interpretation in Konza Prairie watersheds (Adopted from NASA LP DAAC, 2013)

Bit-No	Parameter	Bit-Word	Interpretation
0-1	VI Quality	00	VI produced with good quality
		01	VI produced but check other QA
		10	Pixel produced but most probably cloudy
		11	Pixel not produced due to reasons other than clouds
2-5	VI usefulness	0000	Highest quality
		0001	Low quality
		0010,0001,0010,0100, 1000,1001, 1010	Decreasing quality
		1100	Lowest quality
		1101	Quality so low that it is not useful
		1110	L1B data faulty
		1111	Not useful for any other reason
6-7	Aerosol quantity	01	Low
		10	Intermediate
8	Adjacent cloud detected	0	No
9	Atmospheric BRDF Correction	0	No
10	Mixed clouds	0	No
11-13	Land-water mask	001	Land (nothing else but land)
14	Possible snow/ice	0	No
15	possible shadow	0	No

Table A-3: MODIS NDVI quality analysis based on pixel reliability and VI bit quality

	NDVI Period	**Pixel Reliability**		**VI bit Quality**			
		0 Good Data (% Area)	**1 Marginal Data (% Area)**	**2116* (% Area)**	**2120** (% Area)**	**2185*** (% Area)**	**2189**** (% Area)**
1	MOD13Q1_20010306_20010321	100	0	100	0	0	0
2	MOD13Q1_20010322_20010406	100	0	100	0	0	0
3	MOD13Q1_20010407_20010422	100	0	100	0	0	0
4	MOD13Q1_20010423_20010509	99	1	99	0	1	0

5	MOD13Q1_20010509_20010524	100	0	100	0	0	0
6	MOD13Q1_20010525_20010609	100	0	100	0	0	0
7	MOD13Q1_20010610_20010625	100	0	0	0	100	0
8	MOD13Q1_20010626_20010711	100	0	100	0	0	0
9	MOD13Q1_20010712_20010727	100	0	100	0	0	0
10	MOD13Q1_20010728_20010812	100	0	100	0	0	0
11	MOD13Q1_20020423_20020509	100	0	100	0	0	0
12	MOD13Q1_20020509_20020524	100	0	100	0	0	0
13	MOD13Q1_20020525_20020609	100	0	100	0	0	0
14	MOD13Q1_20020610_20020625	100	0	0	0	100	0
15	MOD13Q1_20030407_20030422	100	0	90	10	0	0
16	MOD13Q1_20030423_20030509	0	100	0	100	0	0
17	MOD13Q1_20030509_20030524	100	0	100	0	0	0
18	MOD13Q1_20030525_20030609	100	0	100	0	0	0
19	MOD13Q1_20030610_20030625	96	4	96	4	0	0
20	MOD13Q1_20030626_20030711	100	0	100	0	0	0
21	MOD13Q1_20030712_20030727	100	0	100	0	0	0
22	MOD13Q1_20030728_20030812	100	0	100	0	0	0
23	MOD13Q1_20040407_20040422	100	0	100	0	0	0
24	MOD13Q1_20040423_20040509	100	0	100	0	0	0
25	MOD13Q1_20040509_20040524	20	80	20	80	0	0
26	MOD13Q1_20040525_20040609	100	0	100	0	0	0
27	MOD13Q1_20040610_20040625	100	0	100	0	0	0
28	MOD13Q1_20040626_20040711	100	0	100	0	0	0
29	MOD13Q1_20040712_20040727	100	0	100	0	0	0
30	MOD13Q1_20040728_20040812	100	0	100	0	0	0
31	MOD13Q1_20040813_20040828	100	0	100	0	0	0
32	MOD13Q1_20040829_20040913	100	0	97	3	0	0
33	MOD13Q1_20050322_20050406	100	0	100	0	0	0
34	MOD13Q1_20050407_20050422	100	0	100	0	0	0
35	MOD13Q1_20050423_20050509	100	0	100	0	0	0
36	MOD13Q1_20050509_20050524	100	0	98	2	0	0
37	MOD13Q1_20050525_20050609	100	0	100	0	0	0
38	MOD13Q1_20050610_20050625	100	0	100	0	0	0
39	MOD13Q1_20050626_20050711	100	0	100	0	0	0
40	MOD13Q1_20050712_20050727	100	0	100	0	0	0
41	MOD13Q1_20050728_20050812	100	0	100	0	0	0
42	MOD13Q1_20050813_20050828	100	0	100	0	0	0
43	MOD13Q1_20050829_20050913	100	0	100	0	0	0
44	MOD13Q1_20050914_20050929	100	0	100	0	0	0
45	MOD13Q1_20050930_20051015	100	0	99	1	0	0

46	MOD13Q1_20051016_20051031	100	0	99	1	0	0
47	MOD13Q1_20060322_20060406	100	0	100	0	0	0
48	MOD13Q1_20060407_20060422	100	0	100	0	0	0
49	MOD13Q1_20060423_20060509	25	75	25	0	75	0
50	MOD13Q1_20060509_20060524	100	0	90	10	0	0
51	MOD13Q1_20060525_20060609	100	0	100	0	0	0
52	MOD13Q1_20060610_20060625	100	0	100	0	0	0
53	MOD13Q1_20060626_20060711	100	0	100	0	0	0
54	MOD13Q1_20060712_20060727	100	0	100	0	0	0
55	MOD13Q1_20060728_20060812	100	0	100	0	0	0
56	MOD13Q1_20060813_20060828	100	0	100	0	0	0
57	MOD13Q1_20060829_20060913	100	0	100	0	0	0
58	MOD13Q1_20060914_20060929	100	0	100	0	0	0
59	MOD13Q1_20060930_20061015	100	0	100	0	0	0
60	MOD13Q1_20070407_20070422	100	0	100	0	0	0
61	MOD13Q1_20070423_20070509	100	0	100	0	0	0
62	MOD13Q1_20070509_20070524	100	0	100	0	0	0
63	MOD13Q1_20070525_20070609	100	0	100	0	0	0
64	MOD13Q1_20070610_20070625	0	100	100	0	0	0
65	MOD13Q1_20070626_20070711	99	1	99	0	1	0
66	MOD13Q1_20080407_20080422	100	0	100	0	0	0
67	MOD13Q1_20080423_20080509	100	0	100	0	0	0
68	MOD13Q1_20080509_20080524	100	0	100	0	0	0
69	MOD13Q1_20080525_20080609	100	0	100	0	0	0
70	MOD13Q1_20080610_20080625	100	0	100	0	0	0
71	MOD13Q1_20080626_20080711	100	0	100	0	0	0
72	MOD13Q1_20080712_20080727	100	0	100	0	0	0
73	MOD13Q1_20080728_20080812	100	0	100	0	0	0
74	MOD13Q1_20080813_20080828	100	0	85	15	0	0
75	MOD13Q1_20080829_20080913	0	100	0	100	0	0
76	MOD13Q1_20080914_20080929	100	0	100	0	0	0
77	MOD13Q1_20090322_20090406	100	0	0	100	0	0
78	MOD13Q1_20090407_20090422	100	0	100	0	0	0
79	MOD13Q1_20090423_20090509	100	0	100	0	0	0
80	MOD13Q1_20090509_20090524	100	0	100	0	0	0
81	MOD13Q1_20090525_20090609	100	0	100	0	0	0
82	MOD13Q1_20090610_20090625	100	0	100	0	0	0
83	MOD13Q1_20090626_20090711	100	0	100	0	0	0
84	MOD13Q1_20090712_20090727	100	0	100	0	0	0
85	MOD13Q1_20090728_20090812	100	0	100	0	0	0
86	MOD13Q1_20090813_20090828	100	0	100	0	0	0

87	MOD13Q1_20090829_20090913	100	0	100	0	0	0
88	MOD13Q1_20090914_20090929	100	0	100	0	0	0
89	MOD13Q1_20090930_20091015	100	0	100	0	0	0
90	MOD13Q1_20091016_20091031	100	0	100	0	0	0
91	MOD13Q1_20091101_20091116	100	0	100	0	0	0
92	MOD13Q1_20100322_20100406	100	0	100	0	0	0
93	MOD13Q1_20100407_20100422	100	0	100	0	0	0
94	MOD13Q1_20100423_20100509	100	0	100	0	0	0
95	MOD13Q1_20100509_20100524	75	25	80	0	20	0
96	MOD13Q1_20100525_20100609	100	0	100	0	0	0
97	MOD13Q1_20100610_20100625	100	0	100	0	0	0
98	MOD13Q1_20100626_20100711	100	0	100	0	0	0
99	MOD13Q1_20100712_20100727	100	0	99	1	0	0
100	MOD13Q1_20100728_20100812	100	0	100	0	0	0
101	MOD13Q1_20110306_20110321	100	0	100	0	0	0
102	MOD13Q1_20110322_20110406	100	0	100	0	0	0
103	MOD13Q1_20110407_20110422	100	0	100	0	0	0
104	MOD13Q1_20110423_20110509	100	0	100	0	0	0
105	MOD13Q1_20110509_20110524	100	0	100	0	0	0
106	MOD13Q1_20110525_20110609	100	0	100	0	0	0
107	MOD13Q1_20110610_20110625	100	0	100	0	0	0
108	MOD13Q1_20110626_20110711	100	0	100	0	0	0
109	MOD13Q1_20110712_20110727	100	0	100	0	0	0
110	MOD13Q1_20110728_20110812	100	0	100	0	0	0
111	MOD13Q1_20110813_20110828	100	0	100	0	0	0
112	MOD13Q1_20110829_20110913	100	0	100	0	0	0
113	MOD13Q1_20111219_20111231	100	0	0	100	0	0
114	MOD13Q1_20120101_20120116	100	0	3	97	0	0
115	MOD13Q1_20120117_20120201	100	0	100	0	0	0
116	MOD13Q1_20120202_20120217	100	0	100	0	0	0
117	MOD13Q1_20120218_20120305	100	0	99	1	0	0
118	MOD13Q1_20120306_20120321	100	0	100	0	0	0
119	MOD13Q1_20120322_20120406	100	0	100	0	0	0
120	MOD13Q1_20120407_20120422	100	0	75	25	0	0
121	MOD13Q1_20120423_20120509	100	0	100	0	0	0
122	MOD13Q1_20120509_20120524	100	0	100	0	0	0
123	MOD13Q1_20120525_20120609	100	0	99	1	0	0
124	MOD13Q1_20120610_20120625	100	0	100	0	0	0

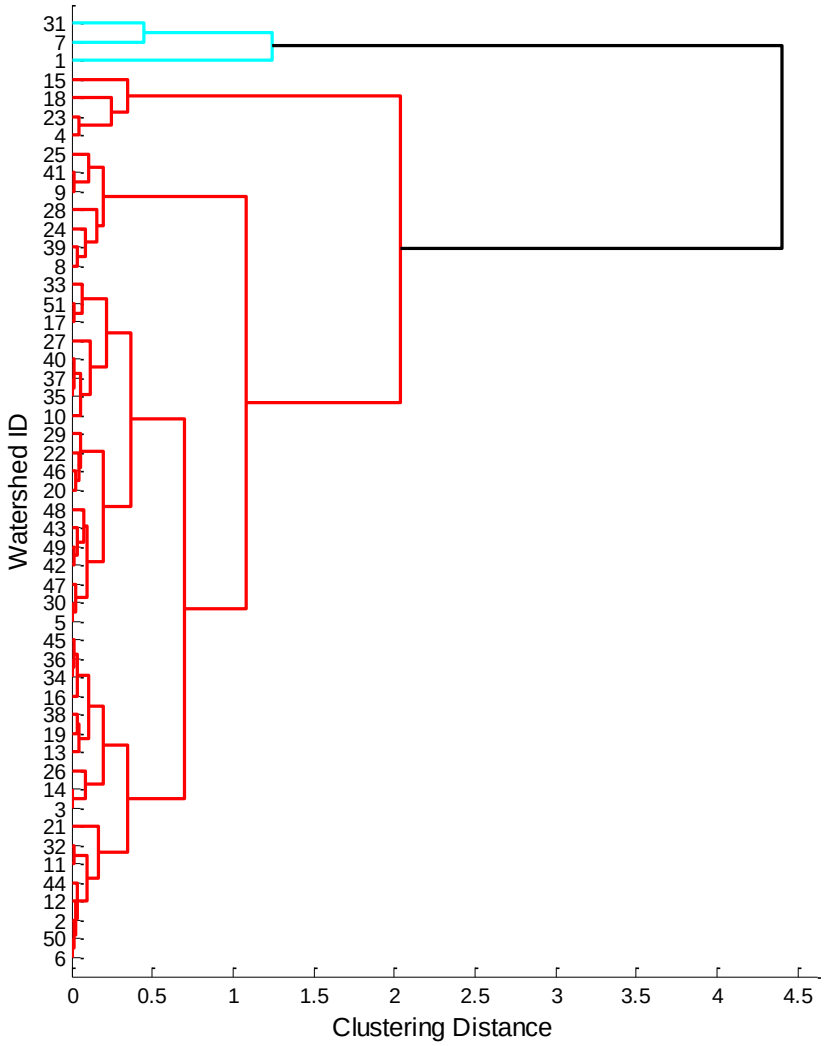
*2116 = 0|0|001|0|0|0|01|0001|00 VI produced with good quality, low quality VI usefulness, low aerosol quantity, no adjacent cloud detected, no atmospheric BRDF correction, no mixed clouds, only land, no possible snow/ice, and possible shadow

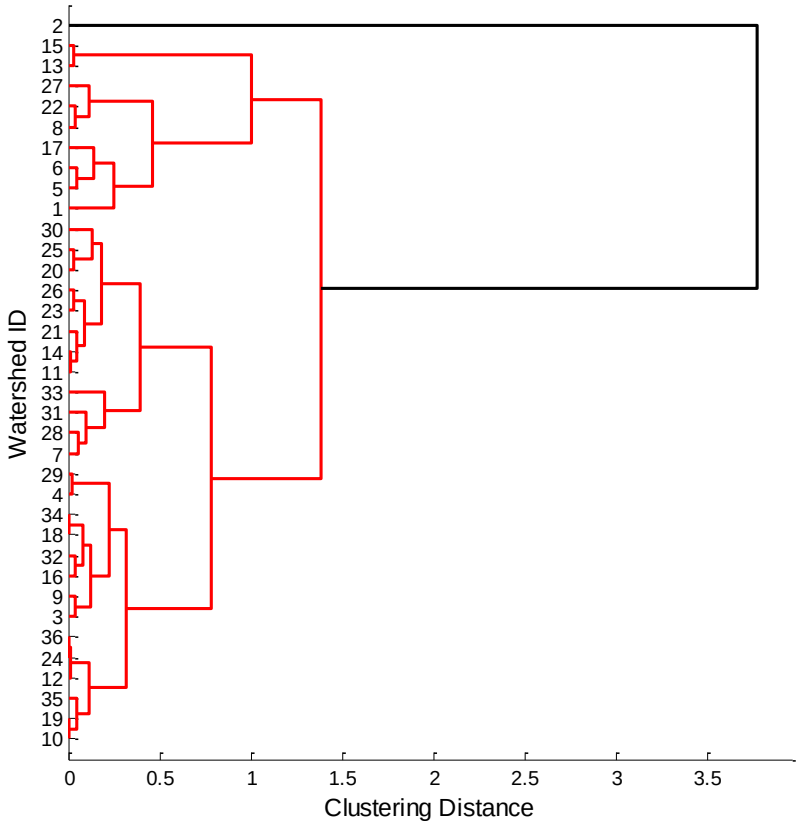
**2120 = 0|0|001|0|0|0|01|0010|00 VI produced with good quality, decreasing quality VI usefulness, low aerosol quantity, no adjacent cloud detected, no atmospheric BRDF correction, no mixed clouds, only land, no possible snow/ice, and possible shadow

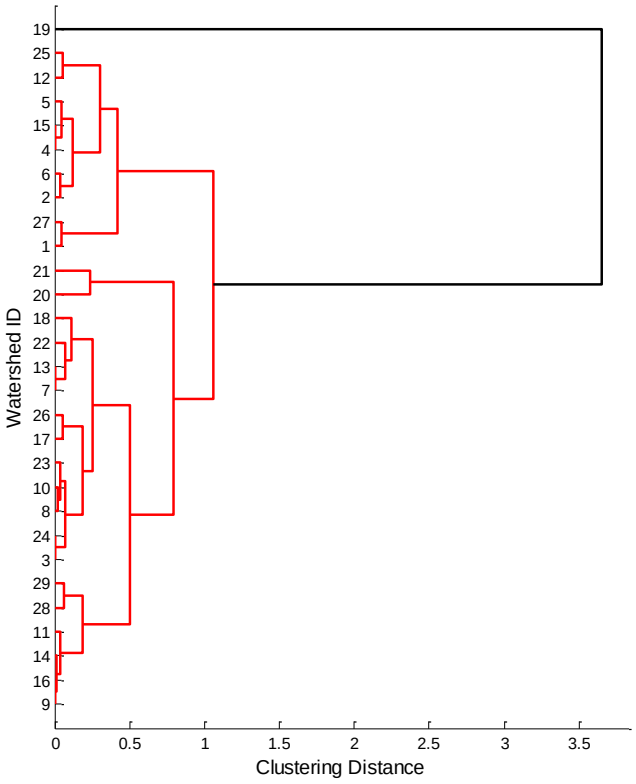
***2185 = 0|0|001|0|0|0|10|0010|01 VI produced but check other QA, decreasing quality VI usefulness, intermediate aerosol quantity, no adjacent cloud detected, no atmospheric BRDF correction, no mixed clouds, only land, no possible snow/ice, and possible shadow

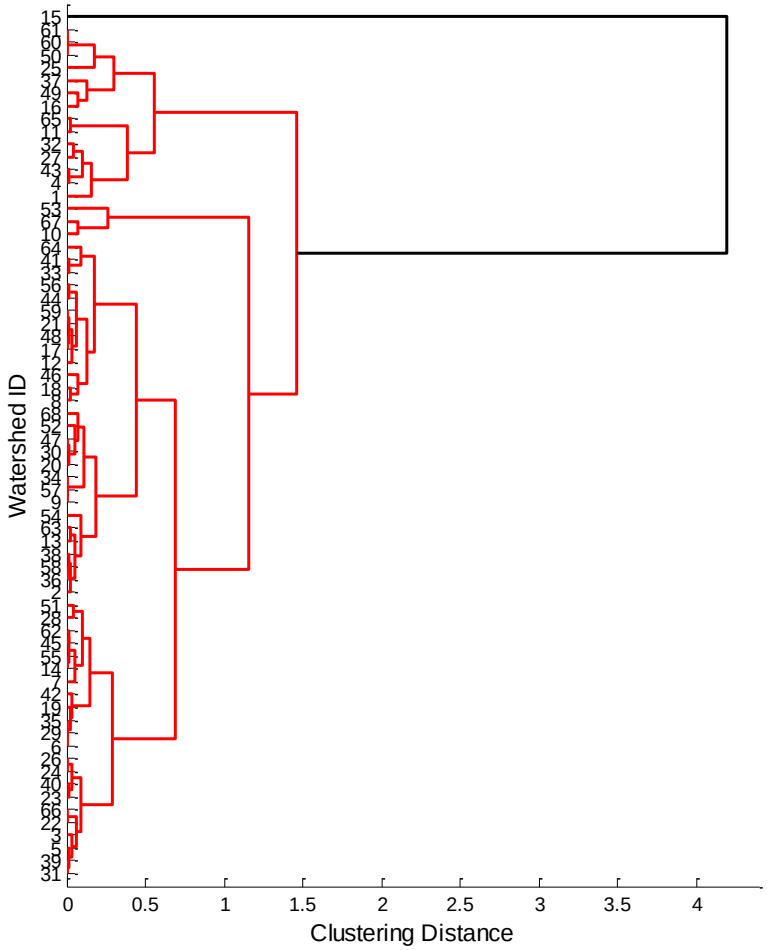
****2189 = 0|0|001|0|0|0|10|0011|01 VI produced but check other QA, no grouping found for VI usefulness, intermediate aerosol quantity, no adjacent cloud detected, no atmospheric BRDF correction, no mixed clouds, only land, no possible snow/ice, and possible shadow

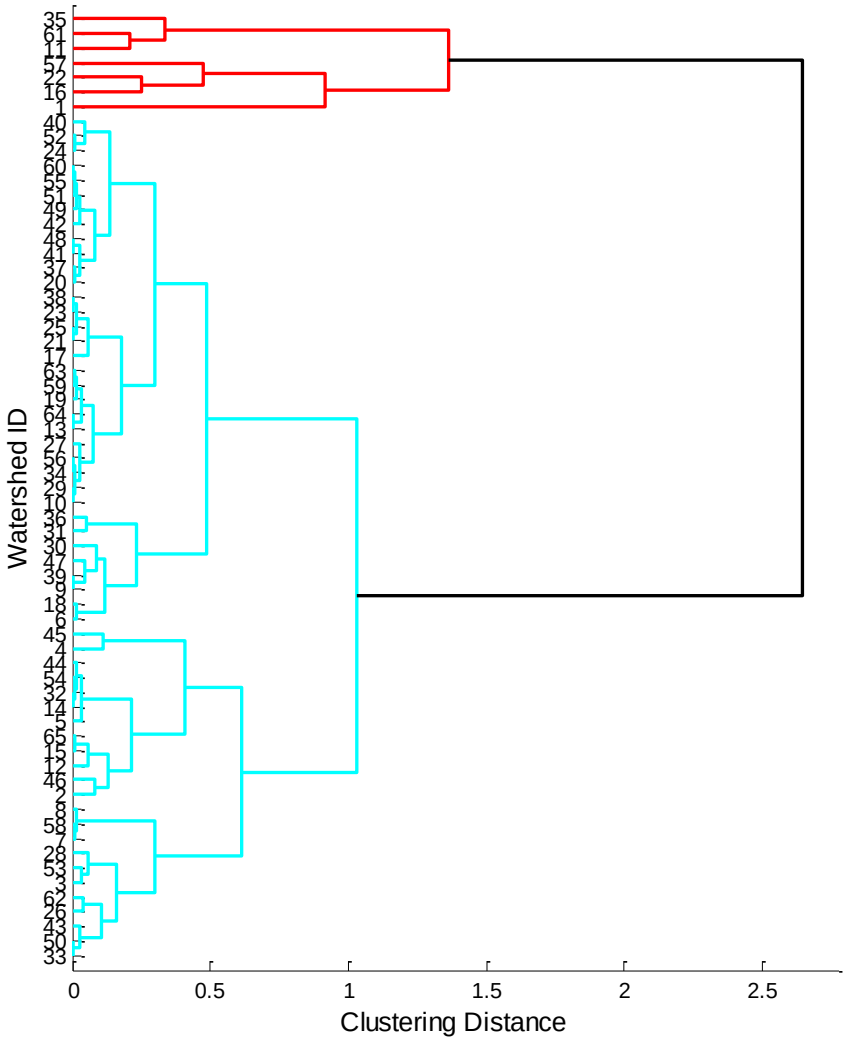
Appendix B- Chapter 4 Additional Information

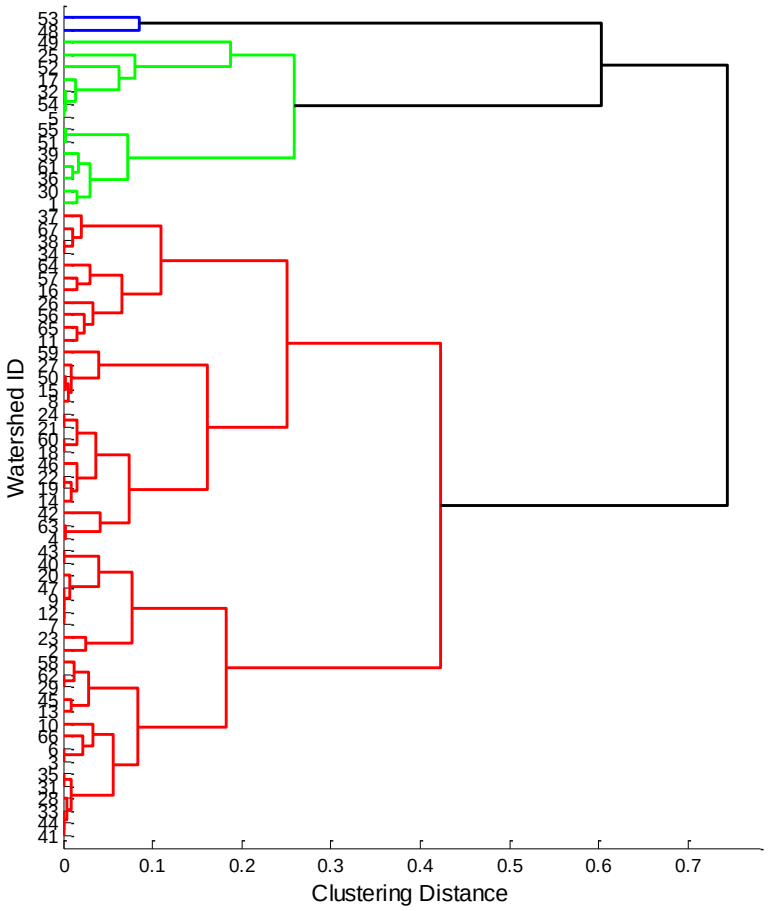



Watershed ID
2
15
13
27
22
8
17
6
5
1
30
25
20
26
23
21
14
11
33
31
28
7
29
4
34
18
32
16
9
3
36
24
12
35
19
10
0
0.5
1
1.5
2
2.5
3
3.5
Clustering Distance


Watershed ID
19
25
12
5
15
4
6
2
27
1
21
20
18
22
13
7
26
17
23
10
8
24
3
29
28
11
14
16
9
0
0.5
1
1.5
2
2.5
3
3.5
Clustering Distance


Watershed ID
Clustering Distance
0
0.5
1
1.5
2
2.5
3
3.5
4


Watershed ID
0
0.5
1
1.5
2
2.5
Clustering Distance


Watershed ID
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Clustering Distance

Cluster	Dendrogram ID	Watershed ID
1	1	37
1	2	39
1	3	43
1	4	50
1	5	86
1	6	100
1	7	113
1	8	125
1	9	127
1	10	134
1	11	139
1	12	154
1	13	161
1	14	181
1	15	198
1	16	200
1	17	210
1	18	212
1	19	215
1	20	219
1	21	220
1	22	223
1	23	230
1	24	242
1	25	244
1	26	247
1	27	250
1	28	260
1	29	284
1	30	297
1	31	299
1	32	300
1	33	302
1	34	306
1	35	308
1	36	312
1	37	324
1	38	325
1	39	326
1	40	328
1	41	330
1	42	334
1	43	336
1	44	341
1	45	349
1	46	356
1	47	357
1	48	360
1	49	362
1	50	363
1	51	365

Cluster	Dendrogram ID	Watershed ID
2	1	132
2	2	141
2	3	148
2	4	155
2	5	166
2	6	169
2	7	176
2	8	178
2	9	224
2	10	226
2	11	232
2	12	237
2	13	239
2	14	240
2	15	253
2	16	257
2	17	291
2	18	331
2	19	338
2	20	339
2	21	340
2	22	346
2	23	351
2	24	366
2	25	368
2	26	369
2	27	370
2	28	371
2	29	372

Cluster	Dendrogram ID	Watershed ID
3	1	17
3	2	73
3	3	106
3	4	107
3	5	111
3	6	114
3	7	118
3	8	121
3	9	144
3	10	153
3	11	163
3	12	165
3	13	172
3	14	190
3	15	222
3	16	227
3	17	245
3	18	261
3	19	264
3	20	269
3	21	271
3	22	277
3	23	280
3	24	294
3	25	296
3	26	298
3	27	309
3	28	313
3	29	317
3	30	320
3	31	329
3	32	333
3	33	342
3	34	350
3	35	353
3	36	358

Cluster	Dendrogram ID	Watershed ID
4	1	7
4	2	11
4	3	24
4	4	29
4	5	34
4	6	36
4	7	48
4	8	61
4	9	62
4	10	70
4	11	75
4	12	76
4	13	85
4	14	99
4	15	102
4	16	116
4	17	119
4	18	123
4	19	128
4	20	137
4	21	152
4	22	156
4	23	158
4	24	167
4	25	171
4	26	175
4	27	186
4	28	187
4	29	188
4	30	193
4	31	196
4	32	199
4	33	206
4	34	207
4	35	211
4	36	214
4	37	225
4	38	235
4	39	241
4	40	246
4	41	256
4	42	265
4	43	267
4	44	273
4	45	275
4	46	276
4	47	278
4	48	283
4	49	286
4	50	288
4	51	290
4	52	293
4	53	303
4	54	305
4	55	310
4	56	322
4	57	323
4	58	327
4	59	335
4	60	337
4	61	344
4	62	345
4	63	354
4	64	355
4	65	361

Cluster	Dendrogram ID	Watershed ID
5	1	1
5	2	4
5	3	5
5	4	8
5	5	13
5	6	14
5	7	20
5	8	22
5	9	27
5	10	28
5	11	30
5	12	31
5	13	47
5	14	52
5	15	54
5	16	55
5	17	58
5	18	59
5	19	63
5	20	68
5	21	69
5	22	71
5	23	74
5	24	81
5	25	84
5	26	88
5	27	89
5	28	94
5	29	98
5	30	103
5	31	115
5	32	117
5	33	120
5	34	122
5	35	126
5	36	129
5	37	133
5	38	135
5	39	138
5	40	140
5	41	143
5	42	145
5	43	150
5	44	151
5	45	159
5	46	164
5	47	168
5	48	170
5	49	173
5	50	180
5	51	182
5	52	184
5	53	192
5	54	194
5	55	197
5	56	204
5	57	236
5	58	249
5	59	251
5	60	252
5	61	254
5	62	258
5	63	259
5	64	263
5	65	272
5	66	274
5	67	281

Cluster	Dendrogram ID	Watershed ID
6	1	2
6	2	3
6	3	6
6	4	9
6	5	10
6	6	15
6	7	18
6	8	21
6	9	23
6	10	25
6	11	35
6	12	40
6	13	41
6	14	44
6	15	45
6	16	46
6	17	49
6	18	53
6	19	57
6	20	60
6	21	64
6	22	66
6	23	67
6	24	72
6	25	78
6	26	79
6	27	80
6	28	82
6	29	91
6	30	93
6	31	95
6	32	97
6	33	105
6	34	108
6	35	109
6	36	110
6	37	112
6	38	124
6	39	130
6	40	131
6	41	142
6	42	147
6	43	149
6	44	157
6	45	160
6	46	162
6	47	177
6	48	179
6	49	183
6	50	185
6	51	189
6	52	191
6	53	195
6	54	202
6	55	205
6	56	208
6	57	216
6	58	228
6	59	229
6	60	231
6	61	233
6	62	248
6	63	262
6	64	266
6	65	279
6	66	285
6	67	295
6	68	301