Changes in the Discharge Characteristics of the Hudson River

Elen Kenny Teo  
Montclair State University

Follow this and additional works at: https://digitalcommons.montclair.edu/etd
Part of the Earth Sciences Commons, and the Environmental Sciences Commons

Recommended Citation
https://digitalcommons.montclair.edu/etd/642

This Thesis is brought to you for free and open access by Montclair State University Digital Commons. It has been accepted for inclusion in Theses, Dissertations and Culminating Projects by an authorized administrator of Montclair State University Digital Commons. For more information, please contact digitalcommons@montclair.edu.
Abstract

The Hudson River watershed has undergone significant urbanization in some areas and reforestation in others during the past century. The watershed has also recently experienced extreme flood events, including from Hurricane Irene, which caused 49 deaths and an estimated $15.8 billion in damage nationally. However, it is unclear how much changes in land use and/or climate have caused shifts in river discharge and the contributions of runoff and baseflow to the river. Determining the existence and magnitude of these shifts are important for managing water resources. This study analyzes changes in daily flows and maximum annual discharge events at multiple recurrence intervals in the context of urbanization and climate change at 13 US Geological Survey river gages in the Hudson River watershed. Smaller, more frequent floods (i.e., the 2, 5, and 10-year floods) are increasing in magnitude at 9 of 13 sites and the larger, while infrequent floods (i.e., 50 and 100-year floods) are decreasing in magnitude at 10 of 13 sites. Increases in population density are correlated with increases in 2-year flood magnitudes at all 5 sites in the Hydroclimatic Data Network of gages with limited human influence. Baseflow is increasing at many sites, often in conjunction with increasing population density, and there are significant seasonal variations in these changes. This study documents the changes in discharge over time and suggests that hydrologists and water managers should consider factors that are significantly correlated with discharge changes (such as local land use) and should not assume that flood magnitude distributions are stationary.
Changes in the Discharge Characteristics of the Hudson River

by

Elen Kenny Teo

A Master's Thesis Submitted to the Faculty of

Montclair State University

In Partial Fulfillment of the Requirements

For the Degree of

Master of Science

May 2014
Acknowledgements

I sincerely appreciate the help that so many members of the wonderful Montclair community have generously provided. Dr. Josh Galster has been a fantastic and enthusiastic mentor, both with guiding me through the research for this thesis and in exploring the hydrology field more broadly. His support has been invaluable in helping me to do good science and learn to manage complicated projects involving large data sets. I would like to thank Dr. Duke Ophori and Dr. Clement Alo for their feedback in writing this thesis and for their exciting courses that taught me so much about hydrology and climate.

I would like to thank Dr. Stephanie Brachfeld and Dr. Matthew Gorring for their guidance in coming to Montclair, obtaining funding to support my education, and the great experience of working with Antarctic sediments in Dr. Brachfeld’s lab.

I am grateful to my parents for the unique inspiration they have provided and to my friends and family for their love and support.
# Table of Contents

1. **Introduction** .......................................................................................................... 1  
2. **Background** ........................................................................................................... 1  
   2.1 Hudson River and Region ............................................................................... 1  
   2.2 Development and Urbanization ...................................................................... 3  
   2.3 Climate Change ................................................................................................ 5  
   2.4 Related Discharge Trends ................................................................................. 7  
   2.5 Stakeholders ..................................................................................................... 8  
   2.6 Summary ........................................................................................................ 11  
3. **Methods** ............................................................................................................... 12  
   3.1 Data Selection ................................................................................................. 12  
   3.2 HEC-SSP and Flood Magnitudes .................................................................... 13  
   3.3 Runoff and Baseflow from WHAT ................................................................. 14  
   3.4 GIS .................................................................................................................. 15  
   3.5 Population ....................................................................................................... 16  
   3.6 Precipitation .................................................................................................... 16  
   3.7 Fixed Effects Model ........................................................................................ 17  
4. **Results** .................................................................................................................. 19  
   4.1 Calculations Performed ................................................................................... 19  
   4.2 Precipitation and Population Density Trends .................................................. 19  
   4.3 Changes in Discharge Magnitudes ................................................................... 20  
   4.4 Runoff and Baseflow ...................................................................................... 24  
5. **Discussion** ............................................................................................................ 26  
   5.1 Runoff and Baseflow Trends .......................................................................... 26  
   5.2 Flood Trends ................................................................................................... 28  
   5.3 Limitations ...................................................................................................... 33  
6. **Conclusion** ........................................................................................................... 34  
7. **References** ........................................................................................................... 35  
8. **Tables and Figures** ............................................................................................. 43
1. Introduction

Global populations are increasing and causing significant changes to the surface of the land. These changes in land cover influence local and regional hydrology. Global climate change is also affecting precipitation magnitude, intensity, and seasonality. These combined factors increase the risk of flooding and drought and present a major challenge to cities seeking to mitigate and adapt to changes. The United States is hydrologically divided around the 100th meridian into the wetter east, where flooding kills more than one hundred people and causes millions of dollars of property damage each year, and arid west, where persistent drought threatens crops and raises food prices. The Hudson River watershed is located in the east and is home to New York City, where a population of over 8 million people live at the mouth of a flood-prone river. The watershed itself is highly variable in terms of population density and land cover. These unique qualities make it well suited for an analysis of hydrologic change in the context of urban development and climate variability.

2. Background

2.1 Hudson River and Region

The Hudson River watershed is 34,700 square kilometers and is located primarily in the state of New York with small areas extending into Vermont, Massachusetts, Connecticut, and New Jersey (Figure 1: Hudson Watershed Map). The main stem of the river is 507 kilometers long with headwaters near Newcomb, NY and discharges into the Atlantic Ocean near New York City. The main tributary of the Hudson River is the
Mohawk River, which joins from the west around the midpoint of the river. A large dam was built in Troy, NY by the federal government in 1823 and limits the impact of tides above the dam (U.S. Army, 1919). Below the dam is a tidal estuary that supports fish populations such as American shad and Atlantic salmon (Walburg and Nichols, 1960; Kahnle and Hattala, 2010).

The Hudson River flows south through the Great Valley lowlands and into the Highlands Province (USGS). The bedrock of the Great Valley is mostly folded Paleozoic shales and carbonate rocks and the Highlands Province is underlain by gneiss, schist, and marble (USGS). Both regions are covered by glacial and sedimentary deposits (USGS). The river’s watershed is bounded to the northwest by the Adirondack Mountains which are composed of anorthosite, quartzites, and marble, part of the 1,100 mya Grenville Orogeny (Landing, 2014). To the west are the Taconic Mountains and to the southwest are the Catskill and Shawangunk Mountains (Figure 2: New York Geologic Map and Figure 3: Hudson Watershed Geologic Map). All of the Hudson River watershed was covered by glaciers during the Pleistocene (Landing, 2014). Although the Hudson River existed prior to the Pleistocene, the glaciers modified the terrain significantly with glaciers moving large volumes of material and meltwater eroding layers of rock and sediment (Landing, 2014).

The Hudson River is a source of water for many uses and also serves as a site for recreational activities including boating and fishing (New York State Department of Environmental Conservation, 2014). In 2011, Hurricane Irene caused extensive flooding along the eastern coast of the US with damage estimates as high as $15.8 billion (Avila and Cangialosi, 2011). In New York City, below the dam, the influence of tides presented
a complicating factor. With large magnitudes of river discharge flowing down towards the bay and seawater levels rising, five New York City hospitals were evacuated; public transportation was closed in New York, New Jersey, Maryland and Pennsylvania; and a state of emergency was declared in nine states and Washington, D.C. (CNN, 2011). In light of these events, a more comprehensive analysis of maximum discharge events along the Hudson River over time will help to gage risk and to plan flood responses in the future, while examining baseflow conditions will aid management of water resources for uses such as recreation, drinking water, and fishing.

2.2 Development and Urbanization

While some regional qualities such as bedrock lithology, local relief, and slope generally remain constant over human time scales, changes in land use can affect stream flow by altering runoff and baseflow processes (Zhang and Schilling, 2006). Urbanization proceeds as small settlements of people in rural areas expand and the surrounding forests are converted to agriculture. These agricultural areas have different hydrologic characteristics than the forests which they replace and may require the withdrawal of water from nearby rivers or lakes to provide irrigation. Over time, settlements grow to accommodate larger populations. More buildings are constructed and roads are paved to provide transportation. Increasing land values increase conversion of agricultural fields and animal pastures to urban land uses (Clonts, 1970).

As some regions of New York have urbanized, other areas have become reforested as settlement patterns change and agricultural use declines (New York State Dept. of Environmental Conservation). In New York, most early urban development occurred along the rivers because of the availability of hydropower for milling, textiles,
logging, and other industrial activities (Figure 4: Land Use Map). As urbanization occurs, buildings and pavements reduce the permeability of the land surface (Leopold, 1968; Sharp et al., 2003). Precipitation that falls onto these low permeability surfaces will infiltrate more slowly and more overland flow (runoff) will be generated. Urbanization can affect river discharge by altering the infiltration of precipitation through the land surface and into the local groundwater (Simmons and Reynolds, 1982; Bronstert et al., 2002). Rivers in heavily urbanized areas generally produce hydrographs with a flood peak that is higher and reached more quickly than in forested watersheds (Anderson and Woessner, 2002). Subsurface transport of water in sewers can also impact the runoff-baseflow system via water pumping and leakage from pipes which may have a net effect of increasing the amount of recharge to groundwater and potentially baseflow (Lerner, 2002; Sharp et al., 2003).

The variability of changes in the Hudson watershed presents a unique opportunity to examine the impact of a wide variety of land uses over a relatively small area. This study focuses on changes in river discharges magnitude model predictions and runoff and baseflow components by using population density as an estimate of land use. Fixed effects models with panel data are used to analyze how historical flow magnitudes have changed over time along with population and precipitation.

This study uses population density as a proxy for increased proportions of impervious surfaces and urbanization more generally. This is a useful method because it does not require the use of aerial photographs. Using aerial photographs can provide more precise information about historical land use, but records are too incomplete for hydrologic analysis. Stankowski (1972) studied the relationship between population
density and land use across New Jersey and found that population density can serve as a good proxy for land use. Although there is significant variation between the amount of impervious surfaces across multiple land use types, (for example, 12-40% impervious for single-family residential areas and 80-100% in commercial zones) Stankowski found a strong relationship between population density and the percent of impervious land cover. The relationship found was non-linear which suggests that logarithmic methods may be useful in the data analysis for future studies.

Population density can serve as a useful metric for other human impacts on the environment as well. Greater populations require more water for personal use, and larger economies have associated commercial and industrial activity, which also require water. Certain non-consumptive uses of water keep the resource within the watershed and examples include drinking, bathing, and lawn irrigation. Consumptive uses cause the water to be removed from the watershed and include evaporation from reservoirs, water used in energy production, and agricultural irrigation of crops intended for export. Water for both use types may come from the Hudson River after treatment or may be provided by groundwater pumping (United States Environmental Protection Agency, 2013). As the population increases, these withdrawals will generally increase. Leaky pipes lose 20-25% of the water they transport, which becomes a source of groundwater recharge (Lerner, 1986). This movement of water from surface bodies to groundwater can have significant effects on the hydrologic system (Ellis and Revitt 2002; Wolf et al. 2004).

2.3 Climate Change

Temperature changes have been documented over the past century at both global and local scales. Globally, temperatures have increased an average of 0.2 degrees C per
decade since the 1980s and 0.61 degrees C between 1861 and 2000 (Hansen et al. 2006; Folland et al. 2001). In New York State specifically, the average annual temperature has increased by approximately 1.5 degrees C between 1895 and 2013 (NOAA, 2014). The urban heat island effect an additional increase in temperature that occurs in populated areas. The effect is difficult to detect in areas with populations under 10,000 and is stronger in areas of greater population density (Karl et al., 1988). Urbanization primarily increases the daily minima, with an average increase of 0.13 degrees and a smaller effect on daily maxima, with an average increase of 0.01 degrees C (Karl et al., 1988).

By raising the minimum daily temperature, these increases may cause snow to melt earlier in the year and influence stream flow. Changes in the seasonality of flooding have been identified using directional statistics in New England (Magilligan and Graber, 2005). This earlier arrival of spring has also been documented in changes to the first leaf date and first flower date in some woody perennials (Wolf et al., 2004). Additionally, a demonstrated connection exists between air temperatures and soil temperatures (Parton and Logan, 1981) and between soil temperatures and infiltration rates (Jaynes, 1990). The effect of increasing average temperatures on infiltration rates may influence the runoff and baseflow pathways followed by precipitation as it moves through the watershed and to the river.

As temperatures increase, the water-holding capacity of the atmosphere also increases (Trenberth, 2011). This results in rainy areas such as the Hudson River watershed receiving more rain. Over the past century, average precipitation levels in the eastern United States have risen, particularly as a result of increasing intensity of precipitation events (IPCC, 2007). In the northeastern part of the country, precipitation is
increasing at a rate faster than the national average, with a 12% increase in the magnitude of the most extreme 1-day precipitation event over the period 1910-1995 compared to a 7% increase nationally and a faster than national average increase in metrics of extreme intensity events, median 1-day events, and moderate intensity events (Karl and Knight, 1998). This increase in precipitation is expected to continue as global temperatures rise with projected climate change, with a predicted 6-7% increase in precipitation per degree C (Wentz, 2007 and Lambert, 2008). When combined with the effects of urbanization on river hydrograph peak and timing in the country’s most densely populated region, the stage is set for flood events becoming increasingly common as time passes.

### 2.4 Related Discharge Trends

Recent research finds that some changes in regional discharges near the Hudson watershed are correlated with changes in climate. Collins (2009) studied the relationship between flood magnitudes and the North Atlantic Oscillation in New England. The North Atlantic Oscillation is a weather phenomenon involving changes in the distribution of atmospheric mass between the Arctic and the Atlantic (Hurrell et al., 2003). When there is a large pressure difference between these areas, the North Atlantic Oscillation is considered to be positive and causes mild winters in both Europe and eastern North America (Hurrell et al., 2003). Because of these warm winters, increased flood magnitude is possible due to snow melt runoff or precipitation as rain instead of snow (Hurrell et al., 2003; Collins, 2009). Collins performed a nonparametric Mann-Kendall test on the long term flood records of 28 New England streams. He found that the flood magnitude has increased since 1970 and that there is a correlation between the North Atlantic Oscillation and the flood magnitude. Since the 1970s, the sign of the oscillation has been mainly
positive and this may be a contributing factor of increased flood magnitudes in New England during the past few decades (Collins, 2009).

Villarini and Smith (2010) studied eastern US rivers for trends in extreme discharge events. They found that there were no significant linear trends but they did detect change points in the time series of 29% of the gages by using the nonparametric Pettit test (Villarini and Smith, 2010). Methods such as the Pettit test are useful in hydrology because hydrologic data often lacks the normal distribution, independence, and non-seasonality necessary to meet the assumptions for most statistical tests (Kundzewicz and Robson, 2004). The Pettit test consists of a rank-based method combined with the Mann-Whitney statistic to determine whether or not the samples come from the same population (Reeves et al., 2007). Villarini and Smith’s research used a data set of the maximum instantaneous annual flood peaks for the prior 75 years at 572 sites across the eastern United States. In New York State, they detected change points in the hydrograph records primarily around the year 1970 with some sites having earlier change points. Many sites within the Hudson River watershed did not have a detected change point.

2.5 Stakeholders

Accurate discharge information is important to a large number of stakeholders from diverse areas such as emergency planning, real estate, and nature enthusiasts. The Federal Emergency Management Agency uses discharge data (along with other data) to calculate flood insurance rates and designate hazard zones (FEMA, 2013). Hazard Zone A consists of all areas that would be affected by the 100-year flood. This area is determined based on the magnitude of the 100-year flood and the topography of the region (FEMA, 2013). All properties that have received federal disaster assistance are
required to purchase flood insurance, with more stringent requirements for properties located within the Hazard Zone, but the federal government has wavered on whether or not the flood insurance rates should reflect the true risk (FEMA, 2012; Grimm, 2013). Reliable discharge data is also necessary for constructing and maintaining bridges and dams. Changes in discharge characteristics over time will have impacts on future development in the Hudson region.

Understanding the local effects of urbanization on rivers in the Hudson River watershed is important for protecting the quality of the environment. Millions of people live in the watershed and rely on the river’s water for drinking and household use. Agriculture and hydroelectric dams also make significant use of the river and it also serves as a popular space for outdoor recreation. Many of these uses require that the water be of a certain quality but some urban processes threaten this quality and may make the river a less desirable destination.

Increases in impervious surfaces are an important indicator of environmental quality (Arnold and Gibbons, 1996) and increase the amount of runoff produced by precipitation. As this runoff moves through the urban environment, it collects pollutants such as nutrients, pesticides, and metals (Paul and Meyer, 2001). Decreases in stream quality are present even at relatively low levels of urbanization between 10-15% impervious surfaces, although a single factor cannot fully explain the complexity of the urban processes (Brabec et al., 2002). These impervious surfaces prevent the polluted runoff from being filtered through the soil where there is a possibility of ecosystem remediation (Arnold and Gibbons, 1996) and transport it directly to the river where it
causes increases in bacteria levels and algal blooms and decreases the biodiversity of macrophytes, invertebrates, and possibly fish (Paul and Meyer, 2001).

Wastewater treatment facilities release effluent into the Hudson and a reliable understanding of discharge is important for it to be released safely. Although this water is treated, it is not possible to completely remove all the impurities from the wastewater and this altered water chemistry has significant impacts on stream ecology (Paul and Meyer, 2001). Combined sewer overflows (CSOs) are systems in which stormwater and untreated sewage are combined and released directly into rivers during storm events when the amount of water to be treated is significantly more than the treatment capacity. Changes in discharge may necessitate a reevaluation of the handling of these polluted waters because of the changing ratio of fresh water to unclean water. For example, if discharge is decreasing in some areas of the river, releases of CSOs or wastewater effluent may raise the level of nutrients such as phosphorous and nitrogen or other pollutants in the river above an acceptable level (Paul and Meyer, 2001). Knowledge of any increases in the proportion of discharge contributed via runoff is also important for water quality management because urban runoff is frequently contaminated with a variety of pollutants such as salt and heavy metals which can easily be introduced into the river (Forman and Alexander, 1998). The challenges are expected to increase as urbanization and climate change proceed (Semadeni-Davies et al., 2008). By observing the local variations in river discharge it will be possible to create management plans which are tailored to the local circumstances rather than general conditions in the state.

Large magnitude discharges serve many important ecological functions for rivers and riparian habitats. The large flows fill the bank of the channel, removing debris and
transporting sediments. Bankfull discharge refers to a magnitude of discharge which will fill the river channel. This flood is particularly important for transporting suspended sediments and altering the shape of the river channel (Simon et al, 2004; Wu et al, 2008; Andrews, 1979) When the river flows over onto the banks, the infiltration of water and the disturbance to vegetation can provide specialized riparian habitat (Nilsson and Svedmark, 2002).

Flows of a certain rate at different times of year are necessary to sustain healthy aquatic communities. Migratory fish such as American shad (*Alosa sapidissima*) are important to the Hudson ecosystem and their harvest has long supported the local economy and served as an abundant source of food (Kahnle and Hattala, 2010). The value that the region places on healthy fish from healthy rivers has been demonstrated in recent decades in the form of the Lambertville Shad Festival (Hinrichson, 1996). The shad population is now under threat and harvests from New York fisheries declined from 2,200,546 pounds in 1896 to 472,261 pounds in 1960 (Walburg and Nichols, 1960). Shad populations have continued their decline in recent years and their harvest is now banned in the Hudson River (Kahnle and Hattala, 2010). Although commercial over harvesting is believed to be the primary cause of the decline, river discharge analysis will provide insight into the changing environment of these economically and culturally important river fish.

**2.6 Summary**

Determining whether and how flood magnitudes and the relative proportions of runoff and baseflow to a river’s average flow are impacted by urbanization and reforestation is important to hydrology and water resource management. This study
builds upon previous research by examining the relationship between population density, precipitation, predicted flood magnitudes, and the magnitude and percentage discharge that is supplied by runoff and baseflow. Fixed effects models using panel data is constructed for 13 gage sites with a series of probabilistic flood magnitudes calculated by HEC-SSP, historical monthly runoff and baseflow data from the hydrograph analysis tool WHAT, and census population records that have been normalized for spatial and temporal precipitation variability. This method quantifies the historical trends in discharges and determines how land use and population density are correlated with discharge in the Hudson River watershed. The study seeks to characterize and quantify changes in discharge events at multiple probability levels (e.g., 100-year flood, 2-year flood).

3. Methods

3.1 Data Selection

Discharge data are from the US Geological Survey’s (USGS) river gage records. The records are continuous and of varying lengths, containing data on both maximum annual flow and monthly discharges. The earliest record begins in 1869 and the most recent ends in 2011. Only gages with a minimum record of 50 years and no multi-year gaps were selected. These requirements for the data allows for the construction of 30-year moving window plots with 20 or more data points per plot to observe changes over time.

13 gages from the Hudson River watershed were selected (Table 1: Table of Gages). Multiple sites are along the Hudson’s major tributary, the Mohawk River. Included in the 13 sites are five sites in the Hydroclimatic Data Network. This is a set of gages that meet criteria for minimal human influence on flows and allows for a
comparison between urbanized and less urbanized drainage basins and to investigate potential climate changes while minimizing the signals from human activity (Lins and Slack, 1999; Regonda et al., 2005).

### 3.2 HEC-SSP and Flood Magnitudes

The computer program HEC-SSP (Hydrologic Engineering Center – Statistical Software Package version 2.0) was used to calculate the computed magnitudes of the 2, 5, 10, 20, 50, and 100-year floods using the method described in the US Geological Survey’s Bulletin 17B (U.S. Geological Survey, 1982). This bulletin uses the Log-Pearson Type III distribution to calculate the computed magnitudes of these predicted floods. This is a widely accepted method for flood calculation and is used by groups such as the Federal Emergency Management Agency and the Army Corps of Engineers (U.S. Army Corps of Engineers, 2008; FEMA, 2012). This method is non-parametric, meaning that it does not require the data to fit a known distribution and is acceptable for hydrologic use (U.S. Geological Survey, 1982). The 100-year flood refers to the magnitude of discharge that will occur once every 100 years on average, or has a 1% chance of occurring annually. The 2-year flood is the magnitude of flow that has a 50% annual chance of occurring.

The flood magnitudes were computed using a 30 year moving window and the results were recorded in Microsoft Excel. This size window was chosen so that changes over longer time scales would be discernable, and that smaller fluctuations would not be taken into account. This length of time is greater than the 25 year minimum suggested by Konrad and Booth (2002) and Brandes (2005). Linear regression for the graphs of the calculated flood magnitude in cubic feet per second (cfs) over time was performed to
obtain the accompanying statistics of the slope (m) and the significance level (p) at each of the 13 gages. The coefficient value for each flood year at each site was then divided by the accompanying average magnitude of flow over the entire gage record to find the average annual percent change in predicted flood magnitude to allow for comparison between watersheds of different size (Figure 5: Rome Example).

3.3 Runoff and Baseflow from WHAT

Separation of runoff and baseflow components of flow is useful to observe how the sources of water entering a stream changes over time. Baseflow provides a substantial contribution to streamflow in many rivers (Anderson and Woessner, 2002). Quantifying runoff and baseflow parameters allows for a variety of hydrologic analysis including aquifer thickness (Dewandel et al., 2003), recharge (Wittenberg and Sivapalan, 1999), and nitrate loading (Doležal and Kvítek, 2004, Schilling and Zhang, 2004, Lim et al., 2005).

WHAT is a web-based hydrograph analysis tool that separates total flows into runoff and baseflow components and is an improvement over graphical hydrograph analysis because of its ability to handle long time periods and consistently differentiate baseflows while considering flow duration and estimating baseflow more consistently than manual separations (Lim et al., 2005). WHAT incorporates two digital filters, BFLOW and Eckhardt (Lim et al., 2005). Two of the filtering equations are

\[ q_t = \alpha \times q_{t-1} + ((1 + \alpha) / 2) \times (Q_t - Q_{t-1}) \]

where \( q \) is the filtered direct runoff (amount of water that the filter shows is a component of the runoff), \( \alpha \) is the filter parameter, and \( Q \) is the total streamflow (amount of all water
in the stream) and

\[ b_t = \frac{(1 - BF_{\text{max}}) \times \alpha + b_{t-1} + (1 - \alpha) \times BF_{\text{max}} \times Q_t}{(1 - \alpha \times BF_{\text{max}})} \]

where \( b \) is the filtered base flow (amount of water that is part of the steady flow caused by water seeping into and out from the channel according to the filter) and \( BF_{\text{max}} \) is the maximum value of the long term ratio of base flow to total streamflow (Lim et al., 2005).

Monthly flow records were analyzed and the results were downloaded from the WHAT website for each study site. Most sites had WHAT data that matched the flood peak data in record length. Runoff and baseflow were calculated as percentages of the total flow at each site for each month and annual sum. Graphs were constructed to compare changes in runoff and baseflow percentage of flow over time. Regression analyses were performed to determine trends and significance levels. Volumetric changes in runoff, baseflow, and total flow were also studied.

3.4 GIS

GIS (Geographic Information System) maps of the watershed were produced using ArcGIS v. 10.2 (ESRI, Redlands, California). Digital elevation models (DEM) of the region was downloaded from the National Map Viewer. The fill, flow direction, and flow accumulation tools in the ArcGIS hydrology toolkit were used to identify flow paths. Subwatersheds were delineated by adding the study gage locations as points in shapefiles and calculating their contributing watersheds. A layer containing the NY state county boundaries was added to the GIS model. The select features and attributes tools were used to display a table of which counties were partially or fully contained within a watershed. The contributing percentage of each county to the area of the entire watershed was calculated and recorded using the formula

\[ p_{\text{county}} = \frac{a_{\text{county}}}{a_{\text{watershed}}} \]

where \( p_{\text{county}} \) is the
proportional coefficient of the contributing county, $a_{\text{county}}$ is the area of the county within the watershed, and $a_{\text{watershed}}$ is the area of the whole watershed.

### 3.5 Population

Population density is used as an indicator of land use change. Historical population data was downloaded from the New York State Department of Economic Development. The data source includes the decadal census population of each New York State county from the 1700s until 1990 (NY Dept. of Economic Development, 2013). The records for 1850-1990 were collected and added to a spreadsheet. Data from the 2010 census was used to extend the population data to the end of the flood record period. The values for the years in between the census years were estimated using linear interpolation.

The census data was integrated with the GIS watershed-county data to provide an estimate of the historical spatial distribution of people into each subwatershed. Although people are not distributed evenly throughout a county, this gives a general estimate of the level of urbanization within the area. If a subwatershed spans multiple counties, then a weighted average was used with the formula $d_{\text{watershed}} = \sum p_{\text{county}}d_{\text{county}}$ where $d_{\text{watershed}}$ is the weighted average of the population density across the watershed, $p_{\text{county}}$ is the proportional coefficient from section 2.4, and $d_{\text{county}}$ is the population density of each county in people per square kilometer. This was repeated for each year in the gage record at each site.

### 3.6 Precipitation

Historical precipitation data was downloaded from National Oceanic and Atmospheric Administration (NOAA) National Climatic Data. The NOAA data provides
average precipitation records at the state level and at the climate division scale. Climate divisions originated from measurements taken by the U.S. Army and were further developed under the U.S. Department of Agriculture, with boundaries are often defined by watersheds or crop type (Guttman and Quayle, 1995). The records at the climate division scale are the unweighted mean of all recording sites within that climate division (Guttman and Quayle, 1995). As climate divisions do not correspond exactly with actual climate regions, this is a limitation of the precipitation data. However, it provides an enhanced resolution for precipitation across the state as compared to the whole-state averages. The use of local data has been shown to produce significantly different results compared to the use of state-wide averages in other regions (Russo and Fisher., 2013). It is important to have this enhanced resolution so that precipitation events which occur in one part of the watershed do not incorrectly influence the results at gage locations where the precipitation event did not take place (Figure 6: New York Climate Divisions).

Because the spatial and temporal variability in rainfall can be significant, precipitation records were used to normalize the flood magnitude data prior to its use in certain runs of the fixed effects model. The final form of the flood magnitude is in cubic meters per second and is analyzed in relation to precipitation in the form of meters of annual rainfall per square meter of the watershed. The precipitation value used is the average rainfall over each 30 year window.

**3.7 Fixed Effects Model**

Fixed effects models using panel data and dummy variables were employed with regression analysis. The panel data is constructed with columns containing various metrics of flood magnitude, population density, and a grid of linking cells. The sites’
columns contain dummy variables that link the discharge data for to each individual watershed. The dummy variables serve to estimate the y-intercept for each subset (Clonts, 1970; Murray, 2006). The fixed effects model provides an estimate of the effect of variables such as population density on measurements of runoff, baseflow, and total flow across multiple sub watersheds over time.

A fixed effects model assumes that the effect of the independent variable (such as population density) is constant (i.e., fixed) over time. It allows the y-intercept to vary between watersheds to account for variations in inter-watershed conditions (slope, bedrock lithology, relief, soil type, etc..) that are not measured in this study. The model seeks to find the best fitting slope that will apply to the full data set. However, each subset of data (each set of gage data is considered a subset) is allowed to have its overall best-fit line to have a y-intercept which varies from gage to gage. This is a valuable method because it considers multiple subsets of data at the same time to detect a signal which may be obscured by the other factors. The method can also produce results with higher significance levels than performing regression on each subset alone (Allison, 2005). The result of the FEM provides an estimate of the effect of the independent on the dependent variable across all the data.

Linear regression was also performed at each individual gage site. This allows for the results of each individual site to be compared to the overall fixed effects model results. This can provide useful information about the gages that do not follow the major trend and suggests avenues for future research. In this study, individual site analysis frequently provided clearer information than the fixed effects model.
4. Results

4.1 Calculations Performed

The calculations discussed in this paper include: a) precipitation and population density regression against time, b) HEC-SSP Q regression against year using moving windows c) Q/A vs present population density, d) Q/PA vs 30-year average population density, e) Q/A vs 30 year average precipitation, f) volumetric analysis of annual runoff and baseflow in comparison to total annual discharge g) runoff percentage as calculated by WHAT vs year for both annual and monthly flows, h) runoff percentage vs population density. Q refers to the flow magnitude and is measured in cubic meters per second. A refers to the area of the watershed in square meters. P is the precipitation in inches. Population density is measured in people per square kilometer. Calculations were performed using both the population density in the final year of the 30-year moving window period (present) or the average during the 30 year period. Similar results were found using both measurements and one example of each is presented.

4.2 Precipitation and Population Density Trends

(a) Linear regression of precipitation against year was performed and detected an increase in annual precipitation at all sites. The amount of the increase varies between 0.23 and 0.93 inches per decade. P values are below 0.05 for 11 sites and below 0.06 for all sites (Figure 7: Annual Change in Precipitation). The population density at the study sites covers a large range, from 0.88 to 145.28 people per square kilometer (Figure 8: Population Density Range).
4.3 Changes in Discharge Magnitudes

(b) Statistically significant trends in magnitude of flooding at a variety of recurrence intervals were detected at sites across the study region. The 2-year and 100-year flood are particularly important for flood planning and water management. Regression of the HEC-SSP predicted flood magnitude against time (year) shows that the size of the 2-year flood is increasing significantly at 8 of 13 sites and the 100-year flood is decreasing significantly in 9 of 13 sites. The increase in the predicted 2-year flood magnitude is between 0.10 - 1.08% annually. Of the 3 sites where the 2-year flood magnitude is decreasing, all are non-HCDN gages with decreases between 0.19 - 0.58% annually. The HCDN sites most clearly display this trend of larger 2-year floods and smaller 100-year floods, with three of the five showing a pattern of both significantly increasing 2-year flood magnitudes and decreasing 100-year flood magnitudes. A fourth gage displays increasing magnitudes for all exceedance levels but the result is not significant for the 100-year flood. The fifth HCDN gage shows decreasing magnitudes for all exceedance levels except the 2-year flood which is not a statistically significant result.

Among the non-HCDN sites, the 2-year flood is increasing at four sites and decreasing at three sites. The 100-year flood is increasing at two sites and decreasing at five sites. Two gages follow the trend described above, two show an increase for all flood years, and two show a decrease for all flood years. One site is significantly decreasing for the 100, 50, 20, 10, and 5 year flood and results are not significant for the 2-year flood. The last site is decreasing significantly for the 2, 5, and 10 year recurrence interval (Table 2: HEC HCDN Regression Table, Table 3: HEC Non-HCDN Regression Table, Figure 9: [Image Reference])
HEC Regression HCDN Scatter Plot, Figure 10: HEC Regression Non-HCDN Scatter Plot.

(c) Significant trends were also identified when these flood magnitude predictions were considered in comparison to population density variation instead of time. In this case, the dependent variable flow metric studied was the predicted magnitude \( (Q, \text{in m}^3/\text{s}) \) divided by the area \( (A, \text{m}^2) \) of the watershed flowing to the point, for a value measured in m/s at multiple exceedance probability levels (2, 5, 10, 20, 50 and 100-year). The independent variable is population density, measured in people per square kilometer. The result of this analysis is a measurement of how land use changes as represented by a proxy of increases or decreases in population density (people/km\(^2\)) affect the contribution to the flow from each unit area of the land surface. The units of the result are in m/s per person/km\(^2\), or m\(^3\)/person-s \(E+06\). This calculation is the same as one of discharge magnitude \( (Q) \) vs population.

Among HCDN gage locations, increases in population density are well-correlated with increases in the predicted magnitude of the two year flood. Four of these gages have a significant positive slope varying between 2.94E-09 and 1.37E-08. The non HCDN sites show a more mixed picture, which is consistent with the heterogeneity of urban and urbanizing environments. Five of these gages showed a negative correlation between the predicted magnitude and the population density, with slopes between -3.16E-09 and -1.1E-07. The two sites where the correlation is positive have slope values of 1.09E-08 and 2.12E-09.

The 100-year flood analysis tells a different story at the HCDN sites. For this test, the correlation between population density and predicted flood magnitude was found to
be negative at four sites, all statistically significant, with values ranging between -1.81E-08 to -6.56E-08. The fifth HCDN result showed a non-significant correlation. The non-HCDN sites showed similar results to the 2-year flood analysis. The correlation was negative and significant at four sites (values -8.39E-09 to -4.76E-07) and positive and significant at two sites values (4.39E-09 and 1.81E-08).

Among HCDN sites, the correlation between discharge magnitude and population density is generally positive at the high-frequency recurrence intervals (2 or 5 year) and negative for the 10, 20, 50, and 100 year recurrence intervals. The exception to this is the Cohoes site, where the correlation is positive for all recurrence intervals. For the non-HCDN sites, the correlation is negative at all recurrence intervals for five of the sites. It is positive for all recurrence intervals at one site, and the remaining two sites have mixed results. Detailed results of these calculations are presented in the appendix (Figure 11: Q/A vs Population Density: 2, 5, 10, 20, 50, and 100 year flood).

The fixed effects model for Q/A vs population density shows that there is a general correlation between an increase in population density and a decrease the predicted magnitude of the 5, 10, 20, 50 and 100 year flood. Each person in the watershed corresponds to a decrease of 1.907E-02 m³/s in the magnitude of the 100-year flood, 1.245E-02 m³/s for the 50-year flood, 6.298E-03 m³/s for the 20-year flood, 3.132E-03 m³/s for the 10-year flood, and a 7.769E-04 m³/s decrease per person for the 5-year flood. The general correlation of each additional person on the 2-year flood is an increase of 7.769E-04 m³/s. P values are very good for the fixed effects model, with all below 0.001. R-squared values are high, all above 0.96.
(d) When Q/A is normalized for the average precipitation within each watershed, the study variable is Q/PA, with units of m²/second-person. When regressed against the 30 year average population density, the results are similar in character to the previous section, with a negative correlation for the 5, 10, 20, 50 and 100 year recurrence interval flood and a positive relationship for the 2-year flood. The 30 year average of population density is used here to allow for a lag time in the effects of population density changes. Analysis at individual gage locations shows again that there is significant heterogeneity between sites. Although 9 sites show a statistically significant negative correlation between Q100/PA and the 30 year population density, the variation in coefficients is two orders of magnitude, from -1.72E-09 to -4.63E-07. Only one to three sites show a positive relationship for this flood metric at the 50, 20, 10, and 5 year recurrence intervals. For the 2-year flood, five sites have a positive relationship and two have a negative relationship. 3 of the positive relationships and neither of the negative relationships are HCDN gages.

(e) Linear regression of Q/A against a 30-year moving average of precipitation find that increases in average precipitation are significantly correlated with increases in the predicted magnitude of the 2-year flood. According to the fixed effects model, each additional meter of precipitation increases the flow generated by a square meter of land surface by 3.48E-10 m³/s. Individual regression calculations at each gage site show a much larger effect. The slope of the regression is positive and varies between 1.41E-07 and 1.43E-06 at the 9 sites with significant results. The slope is negative and statistically significant at one site with a value of -2.39E-07. Results are similar for the 5-year flood, with the FEM estimating a general coefficient of 6.020E-10.
Results for the 10, 20, and 50 year floods show a transition to a more mixed relationship between Q/A and precipitation as the recurrence interval increases. The fixed effects model describes a generally positive relationship, with coefficients of 8.085E-10 (10-year), 1.035E-09 (20-year), and 1.374E-09 (50-year). For these exceedance levels, results are split between positive and negative slope coefficients.

At the 100-year exceedance level, the fixed effects model estimates that a 1 meter increase in average annual precipitation will increase the flow generated by each square meter of the watershed by 1.667E-09 m³/s. The slope coefficient of the gage-level individual regressions is positive and statistically significant at 4 sites, with values from 2.05E-07 to 8.96E-07. Six of the other sites show a statistically significant negative correlation, with slope values between -4.61E-07 to -4.19E-06 (Figure 12: Q/A vs Precipitation: 2, 5, 10, 20, 50, and 100 year flood).

4.4 Runoff and Baseflow

(f) Figure 13: Volumetric Changes in Discharge shows how the magnitude of total annual discharge, runoff, and baseflow have been changing over the study period. Total annual discharge is increasing at eight sites and decreasing at one site. Runoff volume is increasing at six sites, four of which are in the hydroclimatic data network, and decreasing at two non-HCDN sites. Baseflow is increasing at six sites, again with four in the HCDN.

(g) At the HCDN sites, linear regression did not find statistically significant results for the overall yearly percentage of discharge that is contributed by runoff over time, as p values ranged from 0.11 to 0.65. At non-HCDN sites, percent runoff has been
increasing at two of eight sites and decreasing at five. The increases are 0.019% and 0.047% annually and the decreases range from -0.03 to -0.23% (Figure 14: WHAT Regression). The fixed effects model of this shows a coefficient less than 0.001 for this relationship. This is likely caused by the variability in data. All sites except Prattsville have at least one month where a statistically significant change was detected in percent runoff contribution over time. The majority of the data points show an increase in percent baseflow but there is a lot of variability in the magnitude and timing of the changes (Figure 15: Monthly Percent Runoff and Figure 16: HCDN Monthly Percent Runoff). Because Baseflow + Runoff = Total Flow, the changes in percent baseflow are equal in magnitude and opposite in sign to the changes in percent runoff, and vice versa.

(h) Some significant results were found when percent runoff and percent baseflow change was compared with population density instead of time. The largest change is at the Indian Lake non-HCDN site, where the slope of the regression shows that a 1 person per square kilometer increase in population density corresponds to a 63.47% decrease in percent runoff contribution to the river (p>0.05). This is more likely an effect of dam construction and discharge management. The other sites have much more modest relationships. There is a similarly negative relationship between percent runoff contribution and population density at four of the other non-HCDN sites. The range for these relationships is -0.34% to -8.16%, with four significant. At the HCDN sites, only the Prattsville and Cohoes results are significant. The coefficient for Prattsville relates to an increase of +0.17% per person in each square kilometer. Cohoes has a negative correlation of -0.13% (Figures 17a and 17b. Runoff and Baseflow Percent vs. Population Density).
5. Discussion

5.1 Runoff and Baseflow Trends

The positive relationship between population density and percent baseflow at six of the seven sites with significant results suggests that urbanization may alter watershed hydrology in ways other than increasing impervious surfaces and causing greater runoff. The network of distribution pipes, sewers, and septic fields under cities and suburbs has been described as an urban karst (Sharp et al., 2003; Garcia, 2007). Leakage of water from this network can be significant (Burns et al., 2005; Heisig, 2000; and Lerner, 2002). The results indicate that some watersheds are affected by urban development in a way that causes urban groundwater recharge to outweigh the reductions in infiltration caused by impervious surfaces, increasing the amount of baseflow.

Although the percentage of flow that is contributed to the river by runoff is decreasing, the actual magnitude of runoff is increasing at many sites, which is consistent with current literature about the effects of urbanization on surface permeability and runoff (Arnold and Gibbons, 1996; Bronstert et al., 2002). This is possible because total discharge is also increasing. When a larger portion of the increase is baseflow, the percent baseflow increases and the percent runoff decreases, even though the volume or magnitude of both is increasing (Figure 13: Volumetric Changes in Discharge). This means that urban development can increase both runoff and baseflow, and similar results have been identified by Brandes et al. (2005) who found that urbanization does not reduce baseflow and Garcia-Fresca’s (2007) analysis of Austin, TX, where groundwater recharge has doubled since before the area urbanized even though direct recharge from precipitation has decreased because of impervious surfaces. Sharp et al. (2003) discusses
the effects of urbanization on increasing groundwater recharge. Significant water leakage from pipes has also been described (Lerner, 1986).

It is possible that the baseflow effect is influenced by the regional geology. These results are in line with Lopes (2013), who found that urbanization is generally associated with increased baseflow in the 14 sites studied in the Appalachian physiographic province and may be accounted for by leaky water supply lines and interbasin transport of water. One of these sites was in the Hudson River watershed. An opposite trend was found in the nearby Coastal Plain province (Lopes 2013). This study expands the analysis of sites in and near the Appalachian province to include an additional 12 New York gages.

Increases in population often are accompanied by increases in impermeable surfaces (Stankowski, 1972). Infiltration rates decrease as the proportion of impermeable surfaces increases, causing more runoff, shifting discharges higher and decreasing lag times (Anderson and Woessner, 2002). The effect of urbanization is particularly visible at the non-HCDN sites, where seven of eight sites had significant results (Figure 14: WHAT Regression). Most of these sites show an increase in baseflow over period of record, suggesting that unintentional urban groundwater recharge may outweigh decreases in infiltration.

The HCDN gages allow an insight into flows in areas which are unaffected by dams or major anthropogenic activity such as withdrawal of water from the river (Slack and Landwehr, 1988). Although HCDN sites generally lack large urban centers, the HCDN watersheds around the Hudson do contain significant residential populations which have in some cases doubled over the past century (U.S. Census Bureau, 2013).
Some of the changes at HCDN sites are visible when monthly analysis is used, comparing January to January, February to February, and so on. Annual results for these gages were not statistically significant. This suggests that runoff and baseflow changes may be influenced by seasonal factors (Figure 16: HCDN Monthly Runoff).

These increases in population are strongly linked with changes in land use and can be used as a proxy for these changes (Stankowski, 1972). The increase in the magnitude of 2-year flood flow generated by a unit increase in population density for each unit area of the watershed (Q/A) and also normalized for precipitation (Q/PA) supports a relationship between increases in population and increases in the magnitude of flooding contributed by baseflow. The 2-year flood describes a discharge event which happens frequently and for which HEC-SSP can make reliable predictions. The relative frequency means that it will be dominantly influenced by local variables such as the vegetation and average precipitation (Castro and Jackson, 2001). A proportion of this average precipitation goes into the soils and provides the moisture that makes up the baseflow in the river between rain events. Urban development of water distribution systems leak water and supplement the baseflow. The increased baseflow contributes to the increasing annual discharges and the size of the 2-year flood.

5.2 Flood Trends

The climate of the northeastern United States is predicted to become warmer and wetter over the coming century (Hansen et al. 2006; Folland et al. 2001; Karl and Knight, 1998). Combined with the results of this study regarding precipitation and urbanization, it is reasonable to expect that the magnitude of frequent flood events (i.e., 2-year) will continue increase in the future.
Warming climates are associated with reductions in snowpack (Mote et al., 2003) and it is possible that decreasing snowpack has contributed to some of the increases found in baseflow. Water has a better opportunity to seep slowly into the ground and recharge the groundwater if it falls as rain instead of snow in the winter or if winter temperatures frequently rise to allow the snowpack to melt (Barnett et al., 2005; Null et al., 2010). The warmer temperatures can also thaw the frozen ground, allowing for greater infiltration rates and contrasts with colder winters where snowpack builds up and melts in the spring, producing seasonal floods.

The effect of these warmer winters on stream flow has been discussed in the recent hydrologic literature. Collins (2009) identified a correlation between the North Atlantic Oscillation and a step increase in flows in New England around 1970. This is similar to the results of this study, which also identifies an increase in some measurements of stream flow and the presence of step changes and also uses the HEC-SSP method for flood frequency analysis. Because the NAO’s effects become weaker as one moves inland, it is geographically constrained. The method presented in this paper is a more broadly applicable way to measuring changes in stream flow and may be better suited to New York State, particularly in the western areas of the state around the Mohawk River. The step changes identified in this paper may be a result of the same factors as in Collins (2009) but may also be a result of individual watershed events such as dam construction or the inclusion and exclusion of large floods as the HEC-SSP moving window is shifted. This paper joins many others in providing evidence to challenge the Bulletin 17B assumption of stationary climate (Karl and Knight, 1998; Madsen and Figdor, 2007; IPCC 2007; Collins, 2009; Villarini and Smith, 2010).
The results of this study indicate that changes in precipitation may account for some of the changes in discharge. Increases in average precipitation are generally correlated with an increase in the amount of flow generated by each unit of the land surface, and the effects can be seen in the 2-year flood analysis. However, the magnitude of the relationship described here is very small and it is likely that other factors such as urbanization are influencing the flows to a much greater extent than the change in average annual precipitation (Figures 18a and 18b. Population Density Comparison). It is possible that considering precipitation as a 30 year average does not capture enough information about the variability of timing and intensity of precipitation events. The variations in monthly runoff-contributed percentage of flows suggest that there are multiple factors influencing the discharge changes.

Outlier events are expected to become even more extreme as climate variability increases (Groisman et al., 1999; Coumou and Rahmstorf, 2012). The results indicate that reductions in the magnitude of the 100-year flood are possibly the result of a number of factors, including increases in population density, increases in precipitation at the majority of sites both within and outside of the Hydroclimatic Data Network, or other unstudied variables. Decreases in the magnitude of the 100-year flood (Q100) over time were identified at five non-HCDN sites and four HCDN sites. The decrease also exists at the same sites with the exception of Gooley when Q100 is normalized for average precipitation and watershed area. It is possible that these decreases may be an effect of flood mitigation infrastructure but could also be attributable to precipitation timing, water storage capacity, length of the hydrologic record, or other factors.
Using 30 years of data to calculate the magnitude of the 100-year flood limits

HEC-SSP to extrapolate to large flood magnitudes from the available data. Depending on
the timing of extreme precipitation events, this 30-year window may not capture enough
information about the variability of the flood regime and over- or underestimate the flood
magnitude. An expanding window might reduce this source of error but would
necessitate the inclusion of data from NAO-positive and NAO-negative years.

Considering the presence of change points in many hydrologic records (Villarini and
Smith, 2010) it may be more sensible to use the moving windows so that only recent data
is used for each point.

Changes in the water storage capacity of soil and the dependent processes of
evapotranspiration and photosynthesis have been identified as an effect of climate change
(Porporato et al., 2004; Seneviratne et al., 2010). It is possible that these changes affect
the amount of water that can be stored in the soil during a precipitation event. Flood
magnitudes will change if more water can be stored in the environment. Increases in the
amount of forested land in the Hudson watershed has likely changed the permeability of
the soil and increased the amount of water that can be stored. Similarly, flood retention
basins and riverine wetlands can desynchronize the timing of flows and reduce the size of
extreme discharge events.

A broad analysis of change points and flood peak distributions was performed by
Villarini and Smith (2010). They determined that change points are much more common
than linear trends, but did not consider different flood recurrence intervals. In contrast,
this paper’s analysis of the 2-year and 5-year flood detects statistically significant linear
trends in the predicted flood magnitude at 84% (11/13) of gages. Stewart’s Bridge can be
considered a 12th gage if data from prior to the 1930 reservoir construction is excluded. In some cases, the length of the record prior to a change point causes the overall regression line to be the opposite sign of the most recent trend. This is generally a problem in the 100 year flood regressions because of the infrequent timing of extreme floods and record length, although it is also present at Stewart’s Bridge for the 2-year flood because of the reservoir construction.

The results show the presence of step changes or change points of varying size at almost all sites. Some step changes may be the result of climate shifts or anthropogenic activity but can also be an artifact of the moving window method (Collins, 2009, Magilligan and Nislow, 2005). A clear example of a step change is at the Stewart’s Bridge gage, where the construction of a reservoir caused a very large reduction in flows (Figure 19. Stewart’s Bridge). This reduction is reflected in the graph of the predicted magnitude of the 100 year flood over time as a sharp decrease in Q. Sites such as Hadley show more mild effects, with two step decreases in predicted flow and one step increase in 2011 when Hurricane Irene greatly increased the predicted magnitude for low frequency floods (Figure 20. Hadley).

In 2011, Hurricane Irene caused very large discharge events across the northeastern United States and caused predicted flood magnitudes to increase significantly. Because this occurred in the last year of the discharge record, the impact of the Hurricane Irene flood is minimal in some of the results. Future analysis of low frequency discharge events will include this high value as a more significant component (Figure 21. Hurricane Irene).
5.3 Limitations

The use of county-level population and precipitation data is a limitation of this study. Most areas of the Hudson watershed are forested with densely populated cities. Without high-resolution population and climate records, population density and precipitation was considered to be homogenous across each county. Actual land use change generally spreads out from each urban center to include large agricultural areas and actual precipitation is influenced by topography. The approximations of population density in this study are unable to consider these localized factors. More detailed information about historical and current land use patterns will allow an analysis that weights the proportions of urban, agricultural, and forested land to determine which zones are contributing the most to changes in discharge. The analysis performed in this paper does not consider the lack of independence between data points in the hydrologic record. Certain environmental variables such as soil moisture and available subsurface storage capacity will affect the ability of water to move though the system and affect the river discharge.

The construction of dams at sites like Indian Lake is another limitation. The dam at Indian Lake maintains the water level of the Indian Lake, and heavily influences the discharge. This type of human management makes it more difficult to determine how indirect human activity influences the river using historical data from these sites. However, the vast majority of rivers in the United States have dams. Intense human influence is now the normal condition for most rivers so analyzing changes in dammed rivers will provide useful information about the rivers that are of the greatest importance to human society. The HCDN sites serve as an imperfect control for comparison.
6. Conclusion

This study has examined the changes in flood magnitude distribution and runoff-baseflow in the context of urbanization and climate change. The results show that the flood magnitude probability distributions have changed over the past century and that runoff-baseflow processes and contributions to discharge have also changed. This challenges the Bulletin 17B assumption of stationary climate that allows hydrologists to use historical data to predict future flood magnitudes. The changes vary significantly between sites. Urbanization affects the environment by decreasing the permeability of the land surface and increasing the amount of water that leaks out of water distribution infrastructure and recharges the groundwater. Other factors such as precipitation timing and intensity also influence river discharge. It is necessary to consider how these factors have changed in the past and are expected to change in the future and to include the relationships between these factors and hydrologic processes in order to make more reliable predictions of river discharge.
7. References


   <http://dx.doi.org/10.1016/j.jhydrol.2004.03.010>.


   <http://dx.doi.org/10.1016/j.geomorph.2003.07.003>


Figure 1. Hudson Watershed Map Map of the watershed of the Hudson River including portions in New York, Vermont, Connecticut, and New Jersey. From New York State Department of Environmental Conservation.
Figure 2. New York Geologic Map Map of the geologic formations in New York State. The Hudson River watershed (Figure 1) covers most of the state except the western portion, and is predominantly underlain by folded shales, carbonate rocks, gneiss, schist, and marble. From Isachsen et al., 1990, New York State Museum.
Figure 3. Hudson Watershed Geologic Map Cross-sectional view of the geologic layers in the lowlands of the Hudson River Valley. These formations mark the geologic boundaries of the Hudson River watershed and show how the land developed over millions of years. From New York State Department of Environmental Conservation.
New York State Land Use Quad Color

Red = Urban (Residential/Commercial/Industrial/Highways) - 10%
Yellow = Agriculture (Corn/Hay/Alfalfa/Other) - 24%
Green = Forest or Wild Lands (Forests/Barren) - 62%
Blue = Water - 4%

**Figure 4. Land Use Map** GIS map of land use types in New York State. From Arthur, 2011.
<table>
<thead>
<tr>
<th>Gage Name</th>
<th>USGS ID Number</th>
<th>Years of Record</th>
<th>HCDN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gooley</td>
<td>1314000</td>
<td>1917 – 1968</td>
<td>No</td>
</tr>
<tr>
<td>North Creek</td>
<td>1315500</td>
<td>1908 – 2011</td>
<td>No</td>
</tr>
<tr>
<td>Hadley</td>
<td>1318500</td>
<td>1913 – 2011</td>
<td>Yes</td>
</tr>
<tr>
<td>Prattsville</td>
<td>1350000</td>
<td>1909 – 2011</td>
<td>Yes</td>
</tr>
<tr>
<td>Mechanicville</td>
<td>1335500</td>
<td>1869 – 1956</td>
<td>No</td>
</tr>
<tr>
<td>Green Island</td>
<td>1358000</td>
<td>1936 – 2011</td>
<td>No</td>
</tr>
<tr>
<td>Hope</td>
<td>1321000</td>
<td>1912 – 2011</td>
<td>Yes</td>
</tr>
<tr>
<td>Stewart's Bridge</td>
<td>1325000</td>
<td>1908 – 2011</td>
<td>No</td>
</tr>
<tr>
<td>Rome</td>
<td>1336000</td>
<td>1928 – 2011</td>
<td>No</td>
</tr>
<tr>
<td>Indian Lake</td>
<td>1315000</td>
<td>1912 – 2011</td>
<td>No</td>
</tr>
<tr>
<td>Wappinger's Falls</td>
<td>1372500</td>
<td>1929 – 2011</td>
<td>Yes</td>
</tr>
<tr>
<td>Little Falls</td>
<td>1347000</td>
<td>1901 – 2011</td>
<td>No</td>
</tr>
<tr>
<td>Cohoes</td>
<td>1357500</td>
<td>1913 – 2011</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 1. Table of Gages  List of 13 gages used including site names, USGS gage numbers, years of record and inclusion in the Hydroclimatic Data Network. A total of 13 gages were used for this study, including 5 that are HCDN gages and 8 that are not. The longest record is 110 years and the shortest record is 69 years. The median record is 98 years long. All sites are in New York State.
Figure 5. Rome Example An example of the moving window flood magnitude calculations. Each data point is generated using HEC-SSP on a 30-year section of the historical flood data. The program provides the calculated magnitude for the 2, 5, 10, 20, 50, and 100 year recurrence interval floods. Linear regression is used to generate a trend line and find the average annual change. Trend lines for the 2-year and 100-year are shown on the Rome graph.
Figure 6. New York Climate Divisions Map showing the climate divisions of New York State. From Green, 1925.
Figure 7. Annual Change in Precipitation The average annual change in precipitation at each gage location using data from 1895 to 2010. The chart shows the coefficient of the regression, which is equivalent to the average annual increase in precipitation at each site. All sites showed an increase in precipitation at significant levels better than $p < 0.05$. 
Figure 8. Population Density Range This graph shows the range of population densities at each study site in people per square kilometers. Urbanization intensity is highly variable in the Hudson River watershed, making this a particularly interesting area to study to see the diversity of effects.
<table>
<thead>
<tr>
<th>SITE</th>
<th>FLOOD</th>
<th>P VALUE</th>
<th>% M/AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohoes</td>
<td>100 year</td>
<td>8.01E-02</td>
<td>0.06%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>3.13E-04</td>
<td>0.11%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>1.69E-08</td>
<td>0.16%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>9.65E-11</td>
<td>0.19%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>1.82E-12</td>
<td>0.21%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>1.24E-12</td>
<td>0.20%</td>
</tr>
<tr>
<td>Hadley</td>
<td>100 year</td>
<td>8.45E-19</td>
<td>-0.54%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>7.85E-16</td>
<td>-0.38%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>8.54E-09</td>
<td>-0.19%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>5.24E-02</td>
<td>-0.05%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>6.91E-04</td>
<td>0.08%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>8.55E-13</td>
<td>0.19%</td>
</tr>
<tr>
<td>Hope</td>
<td>100 year</td>
<td>4.43E-06</td>
<td>-0.23%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>1.35E-04</td>
<td>-0.16%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>2.32E-02</td>
<td>-0.08%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>6.19E-01</td>
<td>-0.01%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>5.84E-02</td>
<td>0.04%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>5.88E-07</td>
<td>0.10%</td>
</tr>
<tr>
<td>Prattsville</td>
<td>100 year</td>
<td>3.31E-04</td>
<td>-0.31%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>7.73E-04</td>
<td>-0.25%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>1.44E-02</td>
<td>-0.14%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>4.34E-01</td>
<td>-0.04%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>2.02E-02</td>
<td>0.10%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>1.17E-14</td>
<td>0.40%</td>
</tr>
<tr>
<td>Wappinger's Falls</td>
<td>100 year</td>
<td>9.21E-17</td>
<td>-1.91%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>1.90E-16</td>
<td>-1.58%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>2.67E-15</td>
<td>-1.13%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>4.22E-13</td>
<td>-0.79%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>2.65E-08</td>
<td>-0.44%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>2.66E-01</td>
<td>0.06%</td>
</tr>
</tbody>
</table>

Table 2. HEC HCDN Regression Table Results of linear regression of the HEC-SSP moving window over time for 6 recurrence intervals at the 5 HCDN sites. The chart includes the p-value and %M/AVG, which is the correlation coefficient (slope) divided by the average of the predicted flood magnitudes for all years. The slope is a value of the increase or decrease in discharge per year in cms. Dividing by the average normalizes the flows for multiple rivers by comparing them against the average discharge magnitude at the sites. This value shows a normalized magnitude of the annual percent change in calculated discharge. P values in green are >0.05. P values in red are <0.05. Values in the %M/AVG column are shown with a color gradient showing increases (blue) and decreases (red).
<table>
<thead>
<tr>
<th>SITE</th>
<th>FLOOD</th>
<th>P VALUE</th>
<th>% M/AVG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gooley</td>
<td>100 year</td>
<td>2.73E-03</td>
<td>-0.41%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>5.17E-04</td>
<td>-0.38%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>2.98E-05</td>
<td>-0.34%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>5.22E-06</td>
<td>-0.30%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>4.01E-06</td>
<td>-0.26%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>9.42E-05</td>
<td>-0.19%</td>
</tr>
<tr>
<td>Green Island</td>
<td>100 year</td>
<td>8.78E-04</td>
<td>0.27%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>1.67E-03</td>
<td>0.21%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>4.16E-03</td>
<td>0.15%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>5.78E-03</td>
<td>0.13%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>1.78E-03</td>
<td>0.13%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>1.09E-08</td>
<td>0.24%</td>
</tr>
<tr>
<td>Indian Lake</td>
<td>100 year</td>
<td>1.14E-15</td>
<td>-0.92%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>4.62E-17</td>
<td>-0.84%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>8.11E-19</td>
<td>-0.73%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>1.99E-19</td>
<td>-0.66%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>9.44E-19</td>
<td>-0.60%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>4.16E-17</td>
<td>-0.58%</td>
</tr>
<tr>
<td>Little Falls</td>
<td>100 year</td>
<td>5.74E-23</td>
<td>0.56%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>2.10E-25</td>
<td>0.50%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>1.28E-28</td>
<td>0.42%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>2.56E-27</td>
<td>0.36%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>2.66E-20</td>
<td>0.31%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>3.49E-11</td>
<td>0.27%</td>
</tr>
<tr>
<td>Mechanicville</td>
<td>100 year</td>
<td>5.91E-01</td>
<td>-0.15%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>4.11E-01</td>
<td>-0.20%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>1.66E-01</td>
<td>-0.25%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>3.81E-02</td>
<td>-0.30%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>1.41E-03</td>
<td>-0.34%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>3.28E-09</td>
<td>-0.41%</td>
</tr>
<tr>
<td>North Creek</td>
<td>100 year</td>
<td>4.23E-09</td>
<td>-0.22%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>9.39E-07</td>
<td>-0.16%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>8.42E-03</td>
<td>-0.07%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>8.12E-01</td>
<td>-0.01%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>2.96E-02</td>
<td>0.05%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>6.13E-06</td>
<td>0.11%</td>
</tr>
<tr>
<td>Rome</td>
<td>100 year</td>
<td>2.09E-08</td>
<td>-0.49%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>1.30E-03</td>
<td>-0.20%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>4.09E-04</td>
<td>0.17%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>8.55E-16</td>
<td>0.46%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>8.25E-25</td>
<td>0.74%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>1.86E-27</td>
<td>1.08%</td>
</tr>
<tr>
<td>Stewart's Bridge</td>
<td>100 year</td>
<td>1.31E-08</td>
<td>-1.25%</td>
</tr>
<tr>
<td></td>
<td>50 year</td>
<td>1.10E-07</td>
<td>-1.08%</td>
</tr>
<tr>
<td></td>
<td>20 year</td>
<td>3.13E-06</td>
<td>-0.84%</td>
</tr>
<tr>
<td></td>
<td>10 year</td>
<td>5.24E-05</td>
<td>-0.66%</td>
</tr>
<tr>
<td></td>
<td>5 year</td>
<td>1.02E-03</td>
<td>-0.46%</td>
</tr>
<tr>
<td></td>
<td>2 year</td>
<td>5.91E-02</td>
<td>-0.19%</td>
</tr>
</tbody>
</table>

Table 3. HEC Non-HCDN Regression Table  Results of linear regression of the HEC-SSP moving window over time for 6 recurrence intervals for the non-HCDN sites. See table 2 for a full explanation.
The change in calculated flood magnitudes of the HCDN gages normalized by the average flow magnitude for each recurrence interval. This is the slope of the regression line given in Figure 5 (Rome Example) divided by the average magnitude of the discharge over the whole time period. The slope is a value of the increase or decrease in discharge per year in cms. Dividing by the average normalizes the flows for multiple rivers by comparing them against the average discharge magnitude at the individual sites. This shows the change in calculated discharge magnitude every year as a percentage of the average expected discharge. Only statistically significant results are shown.

Figure 9. HEC Regression HCDN Scatter Plot
Figure 10. HEC Regression Non-HCDN Scatter Plot The change in calculated flood magnitudes of the non-HCDN gages normalized by the average flow magnitude for each recurrence interval. This is the slope of the regression line given in Figure 5 (Rome Example) divided by the average magnitude of the discharge over the whole time period. This shows the change in calculated discharge magnitude every year as a percentage of the average expected discharge.
Figure 11. Q/A vs Population Density: 2, 5, 10, 20, 50, and 100 year flood Regression results for Q/A against the population density. This shows how an increase or a decrease in the population density in the watershed flowing to a site is correlated with the calculated magnitude of flow that will be generated by each square kilometer of the watershed for six recurrence intervals. The large change at the Indian Lake site is discussed in section 4.3.
Figure 12. Q/A vs Precipitation: 2, 5, 10, 20, 50, and 100 year flood Regression results for Q per A per the average precipitation for the 30 years prior to the end of the moving window to demonstrate how an increase or a decrease in the precipitation in the watershed is correlated with the calculated magnitude of flow that will be generated by each square kilometer of the watershed for six recurrence intervals. There is a better correlation between average precipitation and smaller discharges rather than large ones because annual precipitation does not capture much information about the type of storms that cause 50 or 100-year flood events.
Figure 13. Volumetric Changes in Discharge This chart shows the statistically significant results of linear regression of the WHAT times series of total annual discharge, annual runoff, and annual baseflow. Only the volume or discharge magnitude is considered, not the percentage of flow. This shows that runoff may be increasing in magnitude while the percent of discharge contributed by runoff may be decreasing, as baseflow makes up the larger portion of the increase.
Figure 14. WHAT Regression Linear regression through the historical data of the amount of annual filtered baseflow as a percentage of total discharge over time. It shows the average annual increase or decrease in baseflow at all statistically significant sites. Results were not significant for any of the HCDN sites so they are not displayed.
Figure 15. Monthly Percent Runoff Linear regression through the month-to-month historical data of the amount of filtered baseflow as a percentage of total discharge over time. It shows the average annual increase or decrease in baseflow. For example, the point for Indian Lake at 1 (January) is the regression of the percentage of baseflow in January only of each year. The majority of statistically significant monthly changes are decreases in runoff but that there is a lot of heterogeneity both seasonally and between sites.
Figure 16. HCDN Monthly Percent Runoff This chart shows only the statistically significant results for the HCDN sites monthly runoff analysis using WHAT. See Figure 14 for a full explanation.
Figures 17a and 17b. Runoff and Baseflow Percent vs. Population Density. The coefficient of the relationship between runoff % and population density using linear regression. Most sites have a negative relationship between these variables, showing that increases in population are usually correlated with a decrease in the runoff percentage of the river. This could be caused by artificial groundwater recharge through subsurface...
pipes or flood mitigation infrastructure. Because Baseflow + Runoff = Total Flow, the changes in baseflow are equal in magnitude and opposite in sign to the changes in runoff.

**Figures 18a and 18b. Population Density Comparison** This chart was produced by multiplying the coefficient (m) of the regression from Tables 2 and 3 in m³/s by the number of years of the record. This produces a value that shows how the river discharge was calculated to change. The second chart shows how much of this expected change can be accounted for by the change in population density. To calculate the values for the second chart, the change in population was multiplied by the Q/A vs pop density coefficient for the corresponding site. The similarity between these charts shows that the
The majority of the change in discharge can be explained by the changes in population density and the land use changes that accompany human settlement.

**Figure 19. Stewart's Bridge** The Stewart's Bridge site shows the effect of dam construction on a river in an urbanizing area. The discharge decreased around 1960 when a dam was constructed in order to fill a reservoir. The discharge has been increasing since then but has not returned to its pre-dam levels. The effect of the dam construction on calculated flood magnitude is largest for the lower frequency floods (100-year recurrence interval) because the dam limits high flow events.
Figure 20. Hadley This site shows the step changes that are produced with the moving window method. Because the data range shifts, the calculations change when a large magnitude discharge event is dropped included from the analysis. An example of this can be seen here around 1967, 1978, and 2005 where the 100-year flood line decreases in a step-like manner as large flood events are dropped. At the end in 2011, there is a stepwise increase as the Hurricane Irene data point is added to the data range.
Figure 21. Hurricane Irene This site shows an example of the very high discharges associated with Hurricane Irene at many sites. This data point increased the calculated magnitude of the 100 year recurrence interval flood by the most and also increased calculated magnitudes for the 50, 20, 10, and 5 year recurrence interval floods to a lesser extent.