HYDROCLIMATIC CONDITIONS ASSOCIATED WITH RAPID FLOODING IN WATERSHEDS OF THE NEW YORK CITY WATER SUPPLY SYSTEM

A Thesis

by

NATALIE GRACE TEALE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Steven M. Quiring
Committee Members,	Oliver W. Frauenfeld
	Anthony T. Cahill
Head of Department,	David M. Cairns

August 2015

Major Subject: Geography

ABSTRACT

The New York City Water Supply System, partially sourced in the Catskill Mountains, NY, is an interconnected series of reservoirs, managed lakes, and aqueducts that serve a population of over 9 million residents in the New York City metropolitan area. This water supply system is unfiltered, as constructing and operating a water filtration facility to support this volume of water would be prohibitively expensive. Because of the unfiltered status, the water supply system is particularly vulnerable to changes in water quality. One such change may occur with flash flooding, which rapidly increases the turbidity of the water and decreases its quality beyond what is allowed in the water supply system.

This thesis seeks to understand flash flooding in these watersheds. In Objective 1, USGS 15-minute discharge data from two similar and adjacent watersheds were analyzed for flash floods from 1987-2014. A total of 32 flash flood peaks were detected in the two watersheds over 23 days. These flash flood events were related to hydrometeorological conditions. In Objective 2, unique synoptic-scale patterns ("synoptic types") were developed with NCEP/NCAR 500 mb geopotential height analyses data. When these 17 types were associated with flash flood events, a statistically significant relationship between types of a southwesterly flow was observed. This pressure gradient occurred significantly more frequently than normal in the days preceding flash floods. Another type, associated with an approaching ridge, was observed significantly less than normal in the days preceding flash flood events. In

Objective 3, a relationship between basic hydrometeorological variables and flash flood peak discharge was sought. Regression analyses showed that antecedent soil moisture and precipitation intensity over the 24 hours preceding the flash flood peak explained 42% of the variance in the peak discharge.

The novelty of this research largely lies with the synthesis of both synoptic scales and local scales in regard to understanding flash flood occurrence and magnitude in the Catskill Mountains, NY. The conclusions of this thesis provide information that could be utilized by watershed managers to improve understanding of flash floods in this water supply system. Direction and suggestions for future work are also provided.

DEDICATION

I would like to dedicate this thesis to my family. They are the reason why I continue to study, ask questions, and search for answers.

First, I dedicate this thesis to my sister, who has always blazed the trail in scholarly development and self-improvement for me. As playmates, teammates, college roommates, and best friends, her support and shining example have always inspired me to push myself further and faster.

I also dedicate this thesis to my mother, who will always be my strongest ally. Her ceaseless support is the wind in my sails, and the reason why I carry on. I can fearlessly test my limits and spread my wings, knowing that she is always in my corner.

Most importantly, I would like to dedicate this thesis to my father, who first taught me to ask questions, and then showed me that there is a sound and scientific method for achieving every answer. His confidence in me is my greatest source of power. Across distance and disciplines, I will always be his sputnik.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to Dr. Steven Quiring for his exceptional mentoring during all stages of the development of this thesis. This includes is but certainly not limited to guiding the research design, providing a funding situation that allowed me to focus on this research, and undertaking several rounds of diligent review of my writing.

In addition to my advisor, I would like to thank the other members of my thesis committee, Drs. Oliver Frauenfeld and Tony Cahill. Their support, questions, and conversations improved the research design and further fueled my curiosity in the field.

I owe sincere thanks to both Trent Ford and Matthew Samollow for all of their support and suggestions, but especially for their help with the coding in this project. I have learned more about coding in the past year with the help of these bright minds and excellent teachers, than I have learned in my entire education so far.

Finally, I would like to express my gratitude to the Climate Science Lab at the Texas A&M University Department of Geography. This research group has been incredibly supportive by providing informal feedback, offering new suggestions, delivering comic relief, and always supplying encouragement to keep working hard.

TABLE OF CONTENTS

Page

ABSTRACT ii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES xii

1. BACKGROUND AND INTRODUCTION 1

1.1. Background of the New York City Water Supply System 1
1.1.1. Changing regional climate 2
1.2. Objectives 4
1.2.1. Objective 1: Identify flash flood events in select watersheds of the New York City Water Supply System 5
1.2.2. Objective 2: Relate specific synoptic-scale circulation patterns to flash flooding in the study site. 5
1.2.3. Objective 3: Quantify the influence of local hydrometeorological drivers on flash flood peak discharge. 5

2. LITERATURE REVIEW 6

2.1. Introduction 6
2.2. Synoptic-scale and regional-scale hydroclimatic patterns 7
2.3. Extreme precipitation events 11
2.3.1. Flash floods 17
2.4. Background of methods used 19
2.4.1. Flash flood events 19
2.4.2. Synoptic typing 24

3. DATA AND METHODS 31

3.1. Study site 31
3.1.1 Neversink River near Claryville, NY (USGS Gauge 01435000) 32
3.1.2 Esopus Creek at Allaben, NY (USGS Gauge 01362200) 33
3.2. Discerning rapid floods in hydrologic record (Objective 1) 34

3.3. Associating rapid floods to synoptic-scale atmospheric conditions (Objective 2)....37
3.3.1. Spatial Synoptic Typer Tools (STT)....37
3.4. Regression analysis for local variables (Objective 3)....39

4. OBJECTIVE 1 RESULTS AND DISCUSSION....46

4.1. Flash floods detected....46
4.1.1. Description of flash flood days....48
4.2. Objective 1 discussion....114

5. OBJECTIVE 2 RESULTS AND DISCUSSION....118

5.1. Typical synoptic patterns....118
5.1.1. Type descriptions....121
5.2. Association of flash floods with synoptic types....134
5.3. Objective 2 discussion....139

6. OBJECTIVE 3 RESULTS AND DISCUSSION....143

6.1. Simple linear regression results....144
6.2. Stepwise linear regression results....147
6.3. Objective 3 discussion....151

7. CONCLUSIONS....157

7.1 Limitations and future work....159

REFERENCES....162

APPENDIX....171

LIST OF FIGURES

Page

Figure 1. Climograph for the Northeastern United States, with monthly data acquired from National Oceanic and Atmospheric Administration (NOAA) Northeast Climate Division for all available years (1895-2015). The Northeast Climate Division ranges from Maine through Maryland, USA. 8

Figure 2. Map of study area and watersheds. The two study watersheds are shaded in light and dark green. USGS gauge stations are indicated by squares. 32

Figure 3. Schematic diagram of the algorithm used to identify flash floods 36

Figure 4. Variable combinations in multiple linear regression analyses.....44

Figure 5. Monthly distribution of flash floods 47

Figure 6. Hourly distribution of flash flood peaks 48

Figure 7. Surface weather map for 1993-1-5 (NOAA Central Library Data Imaging Project). 52

Figure 8. Hydrograph for Neversink River, 1993-01-01 through 1993-01-05. These 15-minute discharge data are from USGS gauge Neversink River near Claryville, NY. 53

Figure 9. As in Figure 7, but for 1996-11-28. 54

Figure 10. As in Figure 8, but for 1993-11-24 through 1993-11-28 55

Figure 11. As in Figure 7, but for 1995-11-12. 57

Figure 12. Hydrograph for Esopus Creek, 1995-11-08 through 1995-11-12. These 15-minute discharge data are from USGS gauge Esopus Creek at Allaben, NY. 58

Figure 13. As in Figure 8, but for 1995-11-12 59

Figure 14. As in Figure 7, but for 1996-01-27 60

Figure 15. As in Figure 12, but for 1996-01-23 through 1996-01-27 61

Figure 16. As in Figure 7, but for 1996-11-09 62

Figure 17. As in Figure 12, but for 1996-11-05 through 1996-11-09 63

Figure 18. NOAA WPC Weather Forecast map issued in the morning for 2003-9-23 64

Figure 19. As in Figure 8, but for 2003-09-19 through 2003-09-23 65

Figure 20. NOAA WPC Weather Forecast map issued for 2003-10-29 67

Figure 21. As in Figure 12, but for 2003-10-25 through 2003-10-29 68

Figure 22. As in Figure 8, but for 2003-10-25 through 2003-10-29 69

Figure 23. NOAA WPC Weather Forecast map issued for 2004-07-23 70

Figure 24. As in Figure 8, but for 2004-07-19 through 2004-07-23 71

Figure 25. NOAA WPC Weather Forecast map issued for 2004-11-28 72

Figure 26. As in Figure 8, but for 2004-11-24 through 2004-11-28 73

Figure 27. As in Figure 12, but for 2004-11-24 through 2004-11-28 74

Figure 28. NOAA WPC Weather Forecast map issued for 2005-01-14 75

Figure 29. As in Figure 12, but for 2005-01-10 through 2005-01-14 76

Figure 30. NOAA WPC Weather Forecast map issued for 2005-10-08 78

Figure 31. As in Figure 8, but for 2005-10-04 through 2005-10-08 79

Figure 32. NOAA WPC Weather Forecast map issued for 2005-11-30 80

Figure 33. As in Figure 8, but for 2005-11-26 through 2005-11-30 81

Figure 34. As in Figure 12, but for 2005-11-26 through 2005-11-30 82

Figure 35. NOAA WPC Weather Forecast map issued for 2006-07-18 83

Figure 36. As in Figure 8, but for 2006-01-14 through 2006-01-18 84

Figure 37. NOAA WPC Weather Forecast map issued for 2006-06-28 85

Figure 38. As in Figure 12, but for 2006-06-24 through 2006-06-28 86
Figure 39. As in Figure 8, but for 2006-06-24 through 2006-06-28 87
Figure 40. NOAA WPC Weather Forecast map issued for 2007-10-27 88
Figure 41. As in Figure 8, but for 2007-10-23 through 2007-10-27 89
Figure 42. NOAA WPC Weather Forecast map issued for 2008-10-25 91
Figure 43. As in Figure 8, but for 2008-10-21 through 2008-10-25 92
Figure 44. NOAA WPC Weather Forecast map issued for 2008-12-12 94
Figure 45. As in Figure 8, but for 2008-12-08 through 2008-12-12 95
Figure 46. As in Figure 12, but for 2008-12-08 through 2008-12-12 96
Figure 47. NOAA WPC Weather Forecast map issued for 2010-01-25 97
Figure 48. As in Figure 12, but for 2010-01-21 through 2010-01-25 98
Figure 49. NOAA WPC Weather Forecast map issued for 2010-10-01 100
Figure 50. As in Figure 12, but for 2010-09-27 through 2010-10-01 101
Figure 51. NOAA WPC Weather Forecast map issued for 2010-12-01 102
Figure 52. As in Figure 12, but for 2010-11-27 through 2010-12-01 103
Figure 53. NOAA WPC Weather Forecast map issued for 2011-08-28 105
Figure 54. As in Figure 12, but for 2011-08-24 through 2011-08-28 106
Figure 55. As in Figure 8, but for 2011-08-24 through 2011-08-28 107
Figure 56. NOAA WPC Weather Forecast map issued for 2011-09-07 108
Figure 57. As in Figure 8, but for 2011-09-03 through 2011-09-07 109
Figure 58. As in Figure 12, but for 2011-09-03 through 2011-09-07 110
Figure 59. NOAA WPC Weather Forecast map issued for 2010-12-18 112

Figure 60. As in Figure 12, but for 2012-09-14 through 2012-09-18 113

Figure 61. As in Figure 8, but for 2012-09-14 through 2012-09-18 114

Figure 62. Scree plot for principal component analysis 118

Figure 63. Variance of (A) within-type scores and (B) between-type scores 120

Figure 64. 500 mb geopotential height composites of the 17 synoptic types from 1987-2013 121

Figure 65. Synoptic type frequencies by month based on 500 mb geopotential height for years 1987-2013. Types are separated into subplots by those peaking in (A) summer, (B) winter, or (C) transitional months. 123

Figure 66. Type occurrences before flood events on (A) the flood peak days and (B) the 1-5 days preceding the flood peak. 136

Figure 67. Bootstrapping analysis of synoptic type frequency 138

Figure 68. Observed vs. Predicted discharge (4 predictors) 146

Figure 69. Observed vs. Predicted discharge (2 predictors) 150

Figure 70. Normal probability plot of residuals 151

LIST OF TABLES

Page

Table 1. Description of predictor variables 40

Table 2. Flood and weather conditions for each flood event in (A) Neversink River watershed and (B) Esopus Creek watershed. Discharge data are from respective USGS streamflow gauges. Meteorological data are from the GHCND station at Slide Mountain, NY 50

Table 3. Precipitation (precip.) characteristics for each synoptic type. Daily precipitation data are from GHCND:USC00301521 Claryville, NY from 1987 through 2013 124

Table 4. Correlation matrix for all predictors 143

Table 5. Final correlation matrix for predictors 144

Table 6. Best simple linear regression table for predicting flash flood peak discharge 145

Table 7. Summary statistics of predictors in stepwise regression 147

Table 8. Stepwise linear regression table 148

1. BACKGROUND AND INTRODUCTION

As the magnitude and frequency of extreme precipitation in the northeastern United States increases, the impact of changing climate on the water supply systems of the northeast becomes critical (Douglas and Fairbank 2011; Frei et al. 2002). While trends in precipitation and streamflow have been studied in the region and recommend preparedness for larger precipitation events in water supply systems, little research has been carried out on the potential for and impacts of rapid flooding in the region. This matter is of critical importance in the New York City water supply system (NYCWSS), an unfiltered water supply with headwaters in the Catskill Mountains serving over nine million residents in eastern New York and the New York City metropolitan area (NYC-DEP 2014). This system is especially sensitive to flooding events, as the turbidity associated with flooding threatens the quality of the unfiltered water supply (Ralston et al. 2013). The uncertainties associated with the changing hydroclimatology of the area coupled with the high impact of this water supply system make a study of this kind essential.

1.1. Background of the New York City Water Supply System

Construction of the NYCWSS began in the 1842, with increasing demands for potable water by growing population in the Manhattan area (NYC-DEP 2014). The water supply system expanded through land accumulation in upstate New York, thereby linking the area of highest population density in the region to an area with the lowest.

The most recent developments of the system were completed in 1964 resulting in 19 reservoirs and three managed lakes in the Catskill Mountains of New York State (NYC-DEP 2014). The present capacity of the water supply system is approximately 580 billion gallons (NYC-DEP 2014). The NYCWSS is the only unfiltered water supply of its magnitude in the United States, after meeting a 150-point standard set by the U. S. Environmental Protection Agency with the Safe Drinking Water Act (U.S.E.P.A. 2014). The Filtration Avoidance Determination is not permanent, but rather requires periodic renewal (U.S.E.P.A. 2014). The high cost of building and maintaining filtration facilities for the magnitude of the NYCWSS requires that the NYC-DEP retain the Filtration Avoidance Determination through careful management of the watersheds and water supply system.

Flooding in the NYCWSS is a cause for concern, as high streamflows are often associated with high turbidity (Ralston et al. 2013; Samal et al. 2013), a condition not accepted under the Filtration Avoidance Determination. Therefore, increases in extreme streamflow may cause disruption of the aging water supply system (Burns et al. 2007; Degaetano and Castellano 2013). A thorough understanding of the hydroclimatic events that may lead to these infrastructure instabilities is necessary to create appropriate water management responses.

1.1.1. Changing regional climate

Several hydroclimatological studies analyzing historical trends and future projections have been previously conducted in the northeastern United States. Some of

these studies have focused on the effect of changing climate on water supplies, including the NYCWSS (e.g. Frei et al. 2002, Burns et al. 2007, Matonse et al. 2012, Pradhanang et al. 2013). While the literature unanimously suggests a general increase in temperature, literature has emerged that offers contradictory findings about the effect of the warming on the regional water supply. For example, Frei et al. (2002) suggests that increases in evapotranspiration caused by warmer temperatures are enough to substantially decrease water availability in the region. However, Burns et al. (2007) reports that the projected increase in precipitation is greater than that of evapotranspiration, thereby increasing the regional water supply. Furthermore, Seager et al. (2012) suggests that the variability in precipitation regimes observed in the last fifty years is due to variability on smaller spatial and temporal scales, and therefore water budget predictions cannot be made based on historical observations.

While some studies indicate that the regional climate will become wetter or drier in the future, the majority of studies in the literature predict a change in seasonal variability. Several studies agree that warmer temperatures will increase winter rainfall, which will increase winter streamflows (e.g. Matonse et al. 2011; Matonse et al. 2012). Furthermore, the same studies show that warmer temperatures will lead to an earlier spring snowmelt. The earlier snowmelt reduces the volume of water available for runoff later in the spring. The shift in snowmelt not only leads to a decrease in spring runoff, but also lengthens the amount of time the water is subjected to evaporation in reservoirs in the water supply system. This pattern ultimately leads to increased winter streamflow and decreased water availability in the summer (Moore et al. 1997; Matonse and Frei

2013; Pranhanag et al. 2013). The uncertainty associated with the timing and magnitude of these changes requires further analysis to improve our understanding of the hydroclimatology of this high-impact water supply system.

The turbidity associated with flooding is not allowed in the unfiltered water supply system. Rapid flooding events elevate turbidity, and the rapid onset of these flooding events leaves little time for waters supply system management to react to the altered streamflow. Studies show that the frequency and magnitude of extreme precipitation has increased (e.g. Douglas and Fairbank 2011; Degaetano and Castellano 2013), which increases the potential for rapid flooding. A more thorough understanding of rapid flooding in the region has implications for water quality. While flash floods are relatively rare events, awareness of the specific conditions leads to the preparedness that is essential for minimizing damage (Doswell et al. 1996). This knowledge must then be applied to climate projections so the NYCWSS can be managed to improve awareness and preparedness for these events.

1.2. Objectives

This thesis will improve understanding of past and future rapid flooding in the Catskill Mountains region. There are three primary objectives:

1.2.1. Objective 1: Identify flash flood events in select watersheds of the New York City Water Supply System

This objective provides the flash flood events to be used in this study. These flash flood events will be discerned from the hydrologic record, and then related to the atmospheric and environmental conditions.

1.2.2. Objective 2: Relate specific synoptic-scale circulation patterns to flash flooding in the study site.

This objective synthesizes methods from both hydrology and climatology to identify the rapid flooding events within the watershed and associate them with atmospheric circulations. This objective is separated into two parts: identifying the rapid flooding events, and associating these events with atmospheric circulation patterns.

1.2.3. Objective 3: Quantify the influence of local hydrometeorological drivers on flash flood peak discharge.

The purpose of this objective is to combine the synoptic-scale conditions associated with flash flooding with local conditions contributing to the maximum discharge during flash floods. This objective will define the hydrometeorological conditions within each watershed and examine how these conditions vary between flash flood events. A statistical model will be developed to predict the peak flash flood discharge based on local conditions.

2. LITERATURE REVIEW

2.1. Introduction

The northeastern United States (the Northeast) is known for its snowy winters, warm summers, and year-round moderate precipitation. Located in the midlatitudes and on the eastern portion of North America, the regional climatology allows the glacially-sculpted topography to support a wide range of ecoregions, including agriculturally-dominated areas, forest, and urban centers. These patterns largely reflect the geomorphology of the region. The regional distinctions in geology and land cover in the northeastern United States are described in Omernick (1987). The northeastern United States is also known for variations in population density. Population density in the Northeast region ranges from less than 1 to more than 1000 people per square kilometer (Hurlburt and Cohen 2014). The high degree of ecological and population variation in the northeastern United States highlights the importance of understanding the climatology of this region.

This chapter documents the patterns of atmospheric flow affecting the region, specifically focusing on advection of air masses. Understanding how these patterns influence the climate of the northeastern United States is important because the region experiences some effects of continentality from its location on the eastern side of the North American continent, and is also influenced by the midlatitude oceanic-atmospheric forcings from the Atlantic Ocean. Through these mechanisms, this region at times experiences both continental and maritime conditions.

The reliance on abundant water resources in the northeastern United States is affected by a dynamic climate in many ways. Dry conditions causing low streamflow are associated with degraded water quality, potentially affecting the large urban populations that are dependent on these water resources (Hayhoe et al. 2006). Additionally, the water temperature directly affects the thermoelectric productivity in the northeastern United States. Low streamflow during the summer months is associated with warmer water and this limits the cooling potential of thermoelectric operations, as well as prevents the return of water to the stream after serving as a coolant (van Vliet et al. 2012).

2.2. Synoptic-scale and regional-scale hydroclimatic patterns

The climatology of the northeastern United States is characterized by seasonal temperatures, with a distinct warm season, cold winter, and gradual transitions between. Summer temperatures average near 21°C and temperatures approaching 30°C are not unusual. The highest temperatures typically occur in June, July, and August. Winter temperatures shift to a minimum in the months of December, January, and February, with an average of approximately -4°C, though temperatures as low as -12°C are not uncommon. The transition between summer maximum and winter minimum temperatures is not abrupt. The seasonal variation in temperature in the northeastern United States is apparent in Figure 1. The monthly data used to assemble Figure 1 was acquired from National Oceanic and Atmospheric Administration (NOAA) Northeast Climate Division for all available years (1895-2015). The Northeast Climate Division ranges from Maine through Maryland, USA.

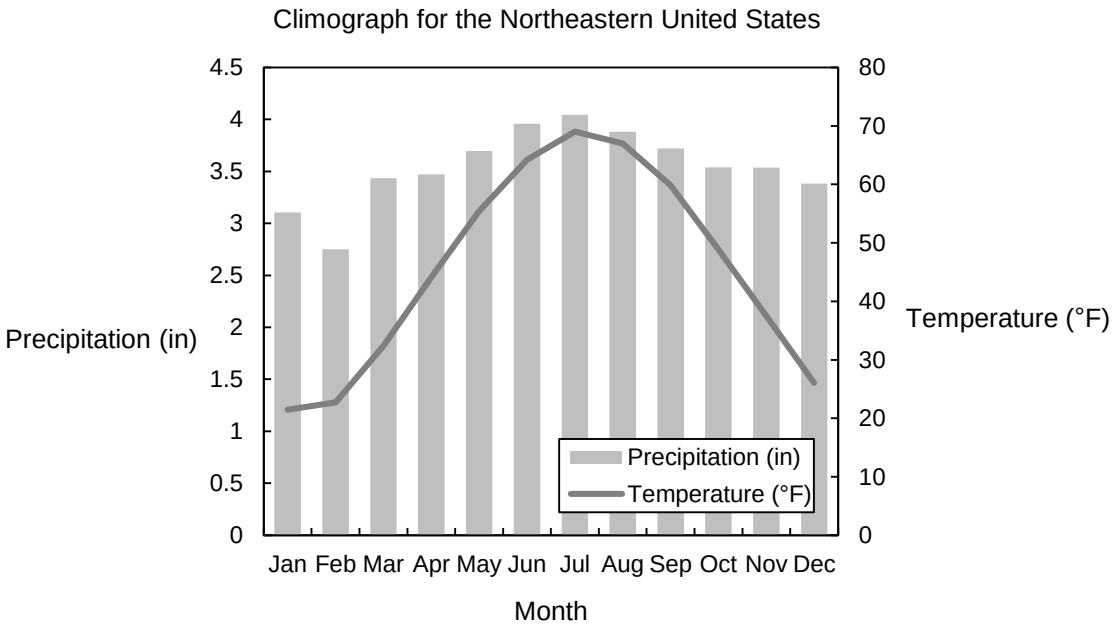

Figure 1. Climograph for the Northeastern United States, with monthly data acquired from National Oceanic and Atmospheric Administration (NOAA) Northeast Climate Division for all available years (1895-2015). The Northeast Climate Division ranges from Maine through Maryland, USA.

Precipitation is consistent in the region, with an absence of marked wet and dry seasons (Figure 1). In an analysis of intra-event precipitation variability of storms exceeding 3 hours throughout the United States, Brommer et al. (2013) analyzed intra-event precipitation variability throughout the United States and found that the mean duration of precipitation events in the Northeast is longer than that in other regions of the United States. Among the storms lasting at least three consecutive hours, the mean duration in the Northeast ranged from 5.5 to 5.75 hours, with increasing duration toward the New England coastline and Maine. In comparison, the mean duration of the longer (>3 hours) storms in Florida, the western Great Plains, the Southwest, and the Rocky

Mountains was less than 5 hours (Brommer et al. 2013). Brommer et al. (2013) also showed that the skewness of the distribution of precipitation within storms is relatively low in the Northeast, ranging from 2-3, meaning that precipitation intensity is relatively constant throughout the storm. In contrast, the skewness of precipitation in the Gulf States exceeds 10, indicating that precipitation intensity is greater earlier in the storm. This abrupt beginning in the Gulf States is associated with convective precipitation events (Brommer et al. 2013). Conversely, the low skewness observed in the Northeast is likely associated with stratiform events, which have a less abrupt onset (Brommer et al. 2013). This distribution may also be associated with the influence of extratropical cyclones (Brommer et al. 2013). The kurtosis of intra-event precipitation is closer to a normal Gaussian distribution east of the Rocky Mountains, though it is negative throughout the United States. In the Northeast, the kurtosis of precipitation in longer-duration events ranges from -0.85 to -0.90, while the kurtosis ranges from -1.00 to -1.05 over the Southwest and the Rocky Mountains (Brommer et al. 2013). Therefore, precipitation events exceeding 3 hours in the Northeast tend to have a relatively long duration, with precipitation distribution closer to a Gaussian distribution than elsewhere in the United States. Brommer et al. (2013) relates these characteristics to cold-core stratiform precipitation events occurring in synoptic-scale frontal systems. Several drivers affect the nature of the moisture entering the northeastern United States region and the manner in which it precipitates, resulting in variability of the precipitation between events in the region.

The topography of North America has been shown to influence large-scale atmospheric dynamics. Manabe and Broccoli (1990) demonstrated how topography (e.g. the Rocky Mountains) influences the synoptic climatology of North America using a comparison of general circulation models (GCMs) with and without topography. The authors showed that when the Rocky Mountains are removed, North America experiences more moisture transport and greater westerly flow, while the presence Rocky Mountains causes ridging and moisture loss over western North America. Manabe and Broccoli (1990) describe how this topography is a driving land-atmosphere interaction influencing the climatology of the northeastern United States.

A trough is frequently present over the eastern United States (the "East Coast Trough" (Bradbury et al. 2002a)) due in part to the presence of the Rocky Mountains. The East Coast Trough has been shown to influence climate and climate variability in the northeastern region. The strength and position of this trough at 500 hPa over the Atlantic coast was explored with regard to the winter climate variability of the northeastern United States in Bradbury et al. (2002b). A trough axis index, a measure of longitudinal position was significantly correlated with winter climate in the Northeast. Bradbury et al. (2002b) shows that the westward-shifted trough was associated with greater than average monthly winter precipitation, and an eastward shift was associated with lower than average monthly precipitation over inland New England. Similarly, the trough intensity index is strongly correlated with winter temperatures in the region (Bradbury et al. 2002a). The trough intensity influences the direction of atmospheric flow in the northeastern United States.

While Bradbury et al. (2002b) showed correlations between trough intensity index with winter temperatures in the northeast, Hubeny et al. (2011) found that the direction of flow is also related to precipitation in the northeastern United States. They found that the dominant air mass in the Northeast is either dry continental or transporting southerly moisture depending on a shift in flow from zonal to meridional flow. The evidence of these patterns in the sediment record indicates that these are strong, regular patterns of atmospheric circulation (Hubeny et al. 2011).

2.3. Extreme precipitation events

A shift from drought to a relatively wet period in the northeastern United States was observed in the 1970s. This transition was attributed to a shift in midlatitude circulation from a northerly and descending flow over the northeastern United States, which was described in Leathers et al. (2000) as contributing to the prior drought conditions, to a southerly and ascending flow (Seager et al. 2012). This shift is supported by other studies (e.g. Thibeault and Seth 2014). Low-level southwesterly flows from the Gulf of Mexico have been significantly correlated with relatively wet summers in the northeastern United States (Thibeault and Seth 2014). Thibeault and Seth (2014) indicate that this adjustment in atmospheric circulation patterns, linked to a westward shift in the Atlantic Subtropical High, caused conditions to switch from dry to wet in the northeastern United States. A period of unusually wet summers has been observed in recent years (Arndt et al. 2010).

The wet summers associated with southwesterly flow are also related to increased convergence over the Great Lakes and western portion of the northeastern United States. Wet summers in the region are associated with negative geopotential height anomalies over western New York and the Great Lakes and positive geopotential height anomalies off the Atlantic coast. These features occur in tandem with the patterns of convergence and divergence from the Great Lakes to the western Atlantic in Thibeault and Seth (2014). Wet summers in the northeastern United States are accompanied by increased convective activity and increased cyclone and trough activity (Thibeault and Seth 2014).

While precipitation in the northeastern United States is typically characterized by uniformly-distributed events of longer duration associated with frontal systems, especially during the summer months (Brommer et al. 2013), localized convective activity also influences the precipitation record in the region. Increased cyclone and trough activity is associated with increased convective activity in the Northeast and therefore the variability in summer mean atmospheric flow is related to distributions of convective storms (Murray and Colle 2011).

Johns (1994) discussed the association of severe weather outbreaks with an approaching surface pressure trough, which is related to the northwest flow that has been associated with tornadic events in the northeastern United States. However, Johns (1984) noted that a number of factors, particularly wind shear, must be considered when using synoptic-scale patterns to forecast severe weather. Similarly, in a climatology of convective precipitation in New Jersey, Croft and Schulman (1989) also found that a

combination of factors were important in producing convective activity in the region. The factors associated with convective activity shifted between mesoscale and synoptic-scale depending on the season. Murray and Colle (2011) found that the composite of active convective days in the Northeast was characterized by a mean trough over the Great Lakes, tying localized convection to synoptic-scale systems. Hurlbut and Cohen (2014) support other studies in showing that warm low-level temperatures under cooler mid-level temperatures, coupled with a midlevel trough, are also factors that are associated with severe thunderstorm environments in the northeastern United States.

The spatial distribution of severe weather related to convective storms has been well documented in the northeastern United States. Falconer (1984) used radar data to define the climatology of thunderstorm days in New York. The maximum number of thunderstorm days was located in southwestern New York along the New Jersey border with nearly 45 thunderstorm days per year, with decreasing numbers of thunderstorm days northward to 10 thunderstorm days per year on the shores of Lake Ontario. When Croft and Schulman (1989) separated thunderstorms from severe thunderstorms, the spatial variability further increased, with the highest frequencies of thunderstorms over southeastern Pennsylvania, but the greatest convection intensities occurring over southwestern New Jersey. Furthermore, adding to the spatial variability of convective storms in the region, Li and Colle (2014) studied warm-season convective environments from 1979 through 2010 and found that there is a decreasing trend in convective activity in western portions of the northeastern United States.

Wasula et al. (2002) examined the spatial distribution of severe weather across eastern New York and western New England to determine if major topographic features of the region such as the Adirondack, Catskill, Green, and Berkshire Mountains and the orientation of the Mohawk and Hudson River valleys influence convective initiation and severe weather. They determined that the terrain has a funneling effect on storm movement. Wasula et al. (2002) also demonstrated that storms forced through a topographic corridor near the Berkshires during days with a northwest flow produced lots of lightning. This agrees with the findings of Johns (1984). Overall, more lightning strikes occurred during events of southwest flow, but severe weather was suspected to be preferentially reported in the Catskills and Berkshires on days of northwest flow (Wasula et al. 2002). Topographic funneling of severe weather has also been studied with regard to tornadoes in the northeastern United States. One example of this situation occurred in May 1998 when two strong upper-level jets and one strong low-level jet introduced warm, moist air, which destabilized the atmospheric conditions. These unstable conditions were channeled through topography in a way that resulted in 32 tornadoes in eastern New York State (LaPenta et al. 2005).

Another situation associated with strong convection in the northeastern United States is an elevated mixed layer, related to the advection of air masses from the southwestern United States. The elevated mixed layer is a condition in which high convective available potential energy is maintained or increased due to an inversion, which acts as a lid in inhibiting convective release (Banacos and Ekster 2010). This vertical profile normally breaks down over the central United States. However, it is

possible that this profile may reach the northeastern United States through entrainment in the mid-tropospheric flow and horizontal advection with vertical downward motion. When the capping inversion deteriorates, strong convective events may ensue (Banacos and Ekster 2010). While the percentage of the convective events in the northeastern United States that are associated with the elevated mixed layer is small, Banacos and Ekster (2010) demonstrated the importance of this convective storm environment because of the unusual severity of these storms. Hurlbut and Cohen (2014) reported similar findings, suggesting that the presence of a mixed layer with inhibited convection and high convective available potential energy are significant in determining the magnitude of a storm event.

Tropical cyclones, tropical storms, and the remnants of tropical systems are associated with many of the most extreme precipitation events in the northeastern United States. In a study examining the transition of Atlantic tropical cyclones into extratropical cyclones, Hart and Evans (2001) showed that major cities in the northeastern United States typically receive 5 cm of precipitation from tropical cyclones every 3 to 6 years, and receive 10 cm of precipitation from tropical cyclones at a longer return interval of 10 to 20 years. Most of this precipitation falls in less than 48 hours (Hart and Evans 2001). In a spatiotemporal review of hurricane landings along the Atlantic coastline of the United States, Keim et al. (2007) found that the return period for tropical storms and hurricanes from New Jersey to Maine ranged from 6 to 12 years. When considering only hurricanes, the return intervals range from 21 to 105 years (Keim et al. 2007). While no hurricanes larger than a Category 3 have made landfall on this section of coast in the 105

years analyzed, it should be noted that discrepancies in the reported category of the New England Hurricane of 1938 caused that event to be reported as a Category 3 at landfall in some publications (Keim et al. 2007). The flooding caused by heavy precipitation from Hurricane Irene and Tropical Storm Lee in 2011 brought renewed attention to the hydroclimatic influence of tropical storms in the northeastern United States (Matonse and Frei 2013). However, Collins et al. (2014) demonstrated that the frequency of flooding from tropical cyclones is small compared to the influence of other flooding mechanisms.

Low-pressure systems are frequent causes of floods in the northeastern United States. Collins et al. (2014) provides a thorough analysis of these mechanisms by separating closed low-pressure systems by source area and tracks. Collins et al. (2014) differentiated the extratropical cyclone tracks through the region and was able to determine which of the storm tracks were more frequently associated with flooding in the Northeast. This study showed that pressure minima moving across the Great Lakes were most frequently associated with flooding the northeastern United States and Canada. The study also revealed that coastal lows (i.e. nor'easters) produce larger floods than lows passing over the Great Lakes, although they are less frequent.

Most floods occur in the early spring; however, the largest floods occurred with the passage of late summer and early fall hurricanes (Collins et al. 2014). In a study of the spatiotemporal variability of floods in New England, Magilligan and Graber (1996) show that the importance of floods occurring in the fall is inversely proportional to both the distance from the coast and latitude.

2.3.1. Flash floods

Flash floods are the aggregate product of both atmospheric and surface conditions. Both the Broxton et al. (2014) reference to this combination as a "'synergy' between hydrological and meteorological factors" (p. 399) and the Doswell et al. (1996) classification of a flash flood as "the concatenation of a meteorological event with a particular hydrological situation" (p. 560) illustrate the importance of understanding both the meteorological conditions and the surface hydrology in forecasting such events. The combination of rapid evolution of precipitating systems and complexity of surface hydrology is one reason for the difficulty in forecasting flash flood events (Kelsch 2001).

Flash floods usually occur in relatively small catchments. The flash floods analyzed in Kelsch (2001) occurred in drainage basins ranging from 10 km^2 to 125 km^2. Small, mountainous watersheds most frequently experience flash floods, as the small catchment length and steep slopes tend to produce flashy streamflow responses to precipitation. In addition to the geomorphic characteristics, the terrain of mountainous watersheds may force increased precipitation intensity and duration. Kelsch (2001) explains that the forced lifting of a convective cell over an orographic barrier may cause cooling that increases the precipitation rate. Kelsch (2001) also explains that topographic barriers may force convective cells in one system to precipitate heavily over the same localized areas as the system passes over mountainous terrain. This process, called regeneration by Kelsch (2001), in essence increases the duration of high rainfall rate in small areas, which increases the potential for flash flooding. Zhang et al. (2009) also

noticed this process associated with an extreme precipitation and flash flooding event in the New York-New Jersey area in October 2006.

Kelsch (2001) found that precipitation intensity is an important influence in flash floods. This characteristic presents as a leptokurtic distribution when plotting precipitation. Brommer et al. (2013) showed that the kurtosis of precipitation in events exceeding 3 hours in the Northeast was more platykurtic in the warm-season months than in the cool-season months.

Maddox et al. (1979) examined 151 flash flood events in the contiguous United States from 1973-1977 to derive information that could be used to improve flash flood forecasting, and found that the flash floods could be separated into four classes: those associated with synoptic-scale events, frontal events, or mesohigh events (associated with convective events), or western events, which occurred west of the eastern Wyoming border. Four common characteristics were found among the flash flood events: heavy precipitation from convective storms, high surface dewpoint temperatures, high moisture content through the troposphere, and weak vertical wind shear (Maddox et al. 1979). Maddox et al. (1979) further found that the formation or passage of multiple convective storms over the same area was associated with flash flooding, as was the location of a large-scale ridge at the 500 mb pressure height.

Maddox et al. (1979) found that the most flash floods in the contiguous United States occur in July, with 25% of the floods analyzed occurring in that month. The only flash flood included in Maddox et al. (1979) in the study area of this thesis occurred in the summer (June, July, or August).

2.4. Background of methods used

2.4.1. Flash flood events

The definition of flash flooding varies by situation and region. Kobiyama and Goerl (2007) list 16 different definitions of flash flooding, each from a different source. While the concept is similar in all the definitions, some definitions require that a flash flood occur within 6 hours of the causative precipitation or dam break (e.g. National Disaster Education Coalition; WMO 1994), while others allow the peak flow to occur 12 hours after the causative event (e.g. Georgakakos 1986). This widely varying definition of flash flooding may cause certain events to be included in some studies while excluded in others.

Maddox et al. (1979) found that most flash floods were related to heavy rain produced by convective storms, high surface dewpoint temperatures, large moisture contents through the troposphere, and weak to moderate vertical wind depth. In this analysis, Maddox et al. (1979) found that flash floods could be classified into distinct groups based on the causative meteorological conditions. Three of the four groups are relevant to flooding in the northeastern United States.

Doswell et al. (1996) show that the potential for storms to produce precipitation required for flash floods can be predicted using basic variables such as rain rate and duration. However, Doswell et al. (1996) show that the many additional factors influencing those basic ingredients make forecasting difficult. For example, precipitation

efficiency should be considered when predicting precipitation amounts, as not all of the water vapor brought into a cloud will fall as precipitation. However, precipitation efficiency is difficult to calculate, as evaporation from the cloud influences the amount of potential precipitation. Doswell et al. (1996) further show that evaporation from a cloud may decrease the amount of water available for precipitation, but evaporation may also cause cooling that strengthens downdrafts in convective cells and increases precipitation. This, among many others factors, shows that calculating each variable in the chain of influences on precipitation amount, rate, and duration is too complex to use in flash flood forecasting when time is of the essence.

There have been many advances in flash flood forecasting, as reviewed in Hapuarachchi et al. (2011). For example, the use of remotely sensed data from radar and satellites alone or in combination with observations provides quantitative precipitation estimates at a high spatial and temporal resolution. While this advance is useful, particularly in poorly gauged or ungauged areas, the accuracy and resolution of these data sources is variable and there are consequently rarely used in in applied hydrological operations (Hapuarachchi et al. 2011). Because flash floods typically occur in small, mountainous watersheds, radar precipitation data may not be available. Kelsch (2001) states that even where radar coverage is sufficient for meteorological analyses, the geomorphic data may not be at a high enough resolution to model the hydrologic processes and responses. Satellite precipitation data may be useful in these locations, however, as reviewed in Hapuarachchi et al. (2011), the resolution and accuracy may not

be suitable for this analysis. Therefore, this study adopts the philosophy of Doswell et al. (1996) in using only basic variables that are easily accessible.

The data used for analyzing flash floods vary among studies, largely due to this recurring lack of data in the small catchments were flash floods typically occur. Spatially representative data are required to analyze both the meteorological and surface conditions, and must also be at a relatively high temporal resolution. While extreme precipitation is widely accepted as a contributing factor to flash flooding, other meteorological conditions also influence the production of flash flood streamflows and should be considered in studies of such conditions. However, the availability of additional relevant data at sufficient spatial and temporal resolution is a limiting factor of study and flash flood prediction.

To develop a 5-year climatology of flash flooding in the United States, Maddox et al. (1979) used flash flood events reported in the NOAA publication *Storm Data*. While this method allowed 151 flash floods and corresponding extreme precipitation events to be included, the reliance on reports having no criteria limits the depth to which the events can be analyzed. For example, Maddox et al. (1979) noted that many of the reports did not include a time of observation, or the duration or amount of precipitation. This highlights the need, as mentioned in Maddox et al. (1979), for an improvement in flash flood reporting such as uniform inclusion of certain criteria necessary for comprehensive analyses.

Reed et al. (2007) used estimated times to peak to select flash flood basins that were used validate a distributed hydrologic model and threshold frequency-based

method in order to improve flash flood forecasts. In this method, the difference between the time of maximum rainfall and corresponding peak discharge was calculated. If that time was less than 6 hours for several events, the basin was considered to be a flash flood basin and was included in the study. This method of determining flash floods requires a reliable timeseries of precipitation and streamflow data, but may be an improvement than relying on incomplete reports as in Maddox et al. (1979). However, because this method for selecting flash flood events and basins is relatively labor intensive, and may only be appropriate at the regional scale or smaller. Further, because the method only requires that some times-to-peak to be fewer than 6 hours in order to be considered a flash flood basin, instead of instituting a threshold quantity, this method may still incorporate subjectivity as in Maddox et al. (1979).

Lumia et al. (2006) divided basins in New York State into 6 regions to formulate multiple regression equations for flood peak at different recurrence intervals. In this method, the watersheds of the New York City Water Supply System west of the Hudson River were separated into two separate regions. Stepwise regression algorithms were used to determine the optimal regression equation. The regression equations were compared and ranked based on multiple criteria, including standard error of the estimate, statistical significance of the explanatory variables, and coefficient of determination (r^2). Multicollinearity was assessed by correlation between predictor variables and the variance inflation factor. Lumia et al. (2006) found that different predictor variables were found to be useful in the two regions across which the New York City Water Supply System watersheds are spread. Given that these watersheds are adjacent, further

study of only these watersheds or the Catskill Mountains may reveal more about the flood characteristics in this region.

Roland and Stuckey (2008) used a multiple regression method to predict flood peaks in ungauged catchments in Pennsylvania, USA. Streamflow data from 322 stations and 24 basin characteristics were used in the development of the least-square regression equations. The predictor variables represented climatic, hydrologic, geologic, and physiographic conditions in the watersheds. The final regression equations included drainage area, mean elevation, percent carbonate bedrock, percent urban area, and percent storage as predictors of flood flows in 4 regions in Pennsylvania, with specific equations for selected recurrence intervals in each region.

In de Castro et al. (2013), a regression method was used to predict flash flood risk in the Phillipines in order to reduce damages. The authors found that forecasted water level and velocity were suitable predictors to be used in a regression model and calculated slope values for both of these variables. The predicted values, classified into stratifications based on their sign and magnitude, indicate the predicted risk of flash flood. The intended purpose of de Castro et al. (2013) is rapid communication of flash flood risk to the public using realtime streamflow data upstream. While this method is interesting and useful, the watersheds of the New York City Water Supply System are likely too small to allow for predictions once the flash flood conditions have already developed upstream.

2.4.2. Synoptic typing

Synoptic typing is a method used for identifying modes of variability in atmospheric variables and classifying instances of the patterns into typical synoptic patterns. Several methods for classifying climate data exist, each with different sets of advantages and disadvantages. Yarnal (1993) describes the methods as being either subjective or objective, or similarly, manual or automated. Further detail on methods of classification are readily available, such as in Yarnal (1993). Aiming for an objective as possible classification scheme, this study utilizes an eigenvector-based principal component analysis (PCA) to identify the main modes of variability, or the underlying structure, of pressure data. This method does require some semi-subjective decisions, but the subjectivity is limited compared to manual methods. Additionally, as an automated method, PCA is reproducible and can be performed quickly, making it the optimal classification system for this study.

The PCA a multi-step process with several decisions required of the researcher. The process begins with first selecting and preparing the data to be suitable for an eigenvector classification, as outlined in Yarnal (1993). The type of matrix used to calculate the eigenvectors and eigenvalues must be decided. This decision depends on the purpose of the analysis, as the correlation matrix and covariance matrix each have their limitations. Correlation matrices depict spatial patterns better than the covariance matrix, making it preferable in map classifications (Yarnal 1993). Covariance matrices better represents spatial deviations in intensity but may also center patterns on areas of

high variance (Yarnal 1993). Therefore, the correlation matrix is preferable to for map pattern classifications, such as those used in this study.

Before proceeding with the eigenvector analysis, the researcher must also determine whether or not to rotate the resulting eigenvectors, called components in PCA analysis. This decision depends on the purpose of the analysis. In some situations, the results derived from an unrotated analysis are necessary (Yarnal 1993). Richman (1986) provides a thorough investigation on the advantages and disadvantages of rotating the principal components produced in a PCA. Unrotated principal components may be considered advantageous as they are orthogonal and their patterns are less sensitive to the number of components retained (Richman 1986). These advantages make unrotated PCA appropriate for data reduction. However, limitations and disadvantages of using unrotated principal components also exist, and are discussed at length in Buell (1979), Richman (1986), Legates (1991), and Richman (1993). Rotated eigenvectors or components may be easier to interpret (Yarnal 1993), thus making rotation preferable in applied purposes such as map-pattern classifications.

When undertaking PCA, the researcher must make some subjective decisions. First, the researcher must decide how many principal components to retain. Establishing a minimum amount of variance explained by each component prior to the analysis creates a threshold after which components will be disregarded. However, adherence to this may also depend on a visual analysis of the scree plot. The break between the components that explain most of the variance and should be retained, and the

components with very little of the variance and should be disregarded, may be visually apparent in the scree plot.

If the results of the PCA are to be used in identifying types (e.g., map-pattern classification), then a cluster analysis must be performed to draw distinctions between different groups. When determining the number of types to include, both the between-type and within-type variance must be considered. High between-type variance indicates that types are greatly different from each other. Therefore, between-type variance should be maximized. High within-type variance suggests that there are large differences among maps of each particular type, whereas low within-type variance indicates conformity of the maps in each type. Therefore, within-type variance should be minimized. The number of types selected should optimize these two, maximizing the between-type variance and minimizing the within-type variance.

Kalkstein et al. (1987) tested three clustering methods after performing PCA using a variety of meteorological elements during winter months at Mobile, Alabama, USA. They showed that the type of clustering method selected greatly influences the number and characteristics of the clusters produced. Therefore, it is important to utilize an appropriate clustering methodology to correctly delineate typical synoptic patterns (Kalkstein et al. 1987).

K-means clustering has been widely used in synoptic typing studies (e.g. Kidson 1994; Galambosi et al. 1996). The method seeks to form groups with the least inner Euclidean distance, but also maintain a reasonable number of groups (Galambosi et al. 1996). As stated in Galambosi et al. (1996), grouping *M* data into *K* clusters using the K-

means method begins by assigning the first *K* elements of *M* to groups of one element each, and assigning the following *M-K* elements to one of the *K* centroids. The centroid of each cluster is then recomputed in the next iteration. The process of assigning elements to clusters and then recomputing the cluster centroid is recomputed until the cluster centroids become stable. This method is preferable to other clustering methods because it is less computationally intensive and the relative frequency of types better reflects extreme events than other methods (Galambosi et al. 1996). Galambosi et al. (1996) refer to the PCA with K-means clustering process as one method, with the components identified in the PCA used as a starting point for generating the clusters. However, Cuell and Bonsal (2009) point out that current computational abilities allow a K-means cluster analysis to be performed alone, without an antecedent PCA which may introduce limitations associated with da ta reduction. Despite the issues raised in Cuell and Bonsal (2009), most studies involving synoptic typing continue to use PCA and clustering algorithms together, as in Jiang et al. (2012), Erfani and Chouinard (2012), and Bettolli and Penalba (2014).

PCA and clustering have been widely used to relate synoptic-scale atmospheric patterns and to surface conditions, particularly in conjunction with synoptic typing. Kidson (1994) used an eigenvector-based methodology and K-means cluster analysis for a synoptic classification in New Zealand. Kidson (1994) used 9 years of daily 5° resolution pressure data for a 30° latitude-by-25° longitude extent centered over New Zealand to discern 13 typical synoptic patterns. Because the temporal extent of the data

was relatively short, data were not separated by season. Thus, the relative frequency of the types reveal seasonality of the patterns.

While also using a PCA and K-means cluster analysis to construct atmospheric circulation patterns in the southwestern United States, Galambosi et al. (1996), unlike Kidson (1994), separated the pressure data by season and performed the PCA and cluster method separately for each of 4 seasons. Using a 35 points of daily pressure data in a 30° latitude-by- 40° longitude diamond-shaped grid over the western United States, Galambosi et al. (1996) found that the success in modeling precipitation based on the synoptic types varied by season, with best results in winter types and improvement needed in summer types.

Romero et al. (1999) used a PCA and K-means cluster analysis methodology to identify 19 synoptic atmospheric patterns influencing significant rainfall events in Mediterranean Spain. Larger classification regions were tested in Romero et al. (1999), but associations with rainfall events were poor due to influence from circulation features distant from the study focus, thus indicating the importance of using an appropriately-sized area of data when relating typical synoptic patterns to surface observations.

PCA and cluster analysis in synoptic typing has been used in numerous regions across the globe. To identify which of the patterns were related to wind erosion events measured at 3 locations in Scania, Sweden, Ekstrom et al. (2002) found success in using PCA and cluster analysis to discern 14 separate pressure patterns for a 50° longitude-by-35° latitude area centered over southern Sweden. Alonso-Pérez et al. (2011) used complementary PCA and K-means clustering to classify 5 years of monthly pressure

anomalies into 3 types, which were then used in identifying circulation patterns enabling African dust to intrude into the marine boundary layer of the subtropical Eastern North Atlantic Region. When comparing the results of a self-organizing map procedure to a PCA and a K-means clustering analysis to classify typical synoptic patterns over eastern Australia, Jiang et al. (2012) found similar circulation patterns represented in both methods. To evaluate the ability of GCMs to represent daily rainfall in the Pampas plain, Argentina, Bettolli and Penalba (2014) used a PCA and K-means cluster analysis methodology for daily sea level pressure over a 47.5° longitude-by-45° latitude extent over southern South America. Raziei et al. (2012) used 12 synoptic types formed via PCA and K-means cluster analysis to determine the typical circulation patterns associated with local and regional precipitation in Iran.

Several classifications of typical synoptic patterns exist for the study region, broadly defined as eastern North America. Kalkstein and Corrigan (1986) used a PCA and Ward's clustering methodology with 5 years of daily weather maps over the United States and, with a focus on maintaining homogenous air masses in each type, found that 10 typical synoptic patterns were inclusive of all daily maps. They then applied the synoptic types to pollution concentrations to determine frequency of elevated sulfur dioxide concentrations in Delaware. Yarnal (1993), who heralded Kalkstein and Corrigan (1986) as advancing synoptic climatology by demonstrating the method and application of synoptic typing, also performed an example of a PCA and cluster analysis. Using 10 years of surface observations from Pittsburgh, Yarnal (1993) filtered the

seasonality of the data by removing 13-day running means from the variables, which left the synoptic conditions present but removed the longer-term conditions from the data.

3. DATA AND METHODS

3.1. Study site

Two streamflow gauges in the southern extent of the NYCWSS were selected for this study based on the size of the drainage area: the drainage area had to be large enough to indicate a regional pattern, but would be small enough to have a clear signal undiluted by other processes. The USGS gauge stations Neversink River near Claryville, NY (01435000) and Esopus Creek at Allaben, NY (01362200) were selected due to similarities in land cover and drainage area (Figure 2). In Figure 2, the two watersheds used in this study are shaded in green and the USGS streamflow gauges are indicated by squares. Pink shading indicates a village or census-designated place. The drainage area at the Neversink River near Claryville, NY and Esopus Creek at Allaben, NY gauge is 66.6 mi^2 and 63.7 mi^2, respectively. These drainage areas are larger than those discussed in Kelsch (2001), where drainage areas of less than 38.6 mi^2 were considered a commonality of flash flood events. However, the history of flash floods in the selected watersheds suggests that while smaller watersheds may produce more flash floods as in Kelsch (2001), watersheds of the selected size are suitable for this study.

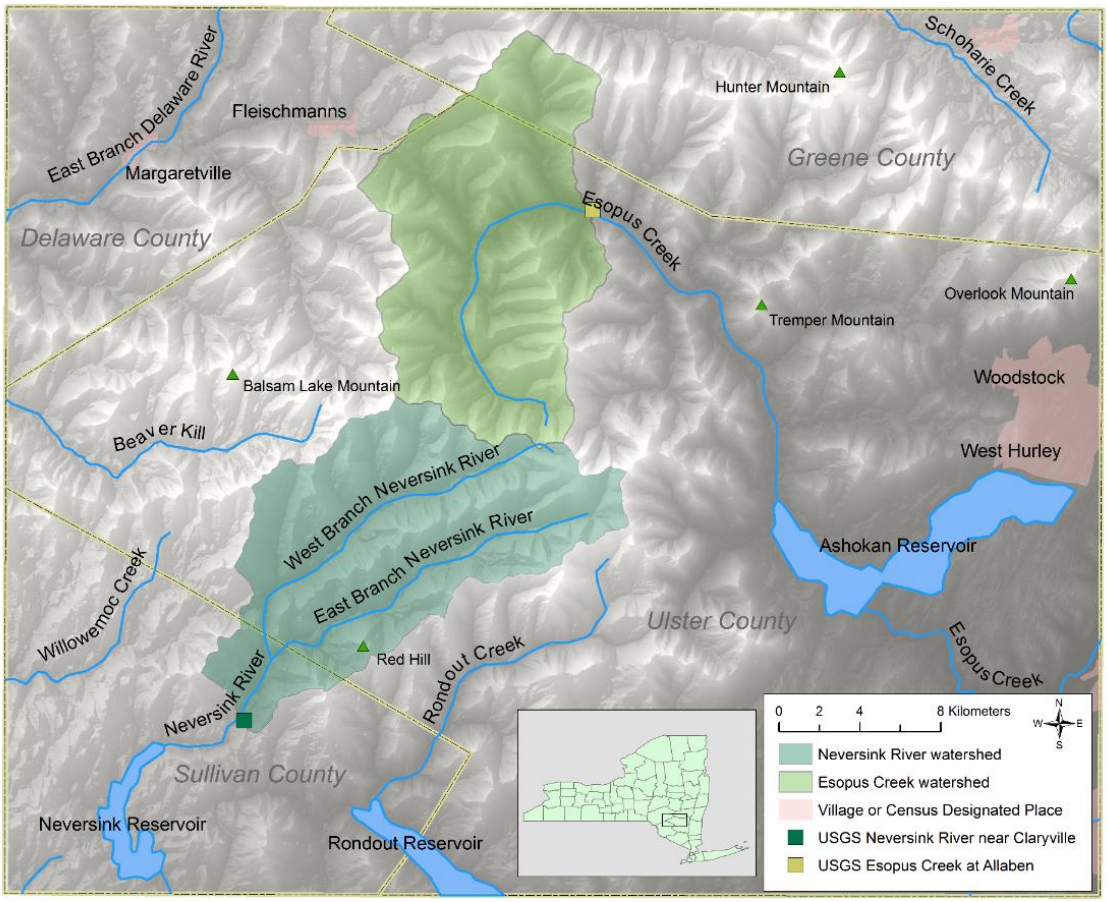

Figure 2. Map of study area and watersheds. The two study watersheds are shaded in light and dark green. USGS gauge stations are indicated by squares.

3.1.1 Neversink River near Claryville, NY (USGS Gauge 01435000)

The USGS Gauge 01435000 Neversink River near Claryville, NY is located at 41°53'24" N, 74°35'24" W, two miles downstream of the confluence of the East and West Branches of the Neversink River and 2.2 miles southwest of Claryville, NY. The direction of flow in this watershed is northeast to southwest, with most frequent slope aspects in this watershed being southeast and northwest, accounting for 18.4% and 15.1% of slopes respectively. The highest elevation in the watershed is 4186 ft ASL and

the elevation of the gauge is 1522.3 ft ASL, giving a relief of 2663.7 ft in the watershed. The period of record extends through November 1937; however, the data used in this study begin in 1987. The mean depth and width of the Neversink River at this gauge are 4.2 ft and 102.2 ft, respectively, with a cross sectional area of 426.5 ft^2. Bankfull discharge is 4182 cfs, with a return interval of 1.3 years. The Rosgen stream classification at this gauge station is C3 (Miller and Davis 2003), which is represents a low gradient, meandering river with riffle-pool bed morphology accompanied by well-defined floodplains (Rosgen 1994). The drainage area of this gauge is 66.6 mi^2.

3.1.2 Esopus Creek at Allaben, NY (USGS Gauge 01362200)

The USGS Gauge 01362200 Esopus Creek at Allaben, NY is located at 42°07'01" N, 74°22'49" W near Allaben, NY, 200 ft downstream from the tributary Fox Hollow Creek. The direction of flow in this watershed is south to north, curving to flow toward the east near the gauging station. The slope aspects are generally east-northeast and south-southwest, though accounting for only 50.6% of the slopes in the watershed. The highest point in this watershed is 3733.6 ft ASL and the gauge is located at 1000.7 ft ASL, making the relief of this watershed 2732.9 ft. The period of record extends through October 1963, though the discharge data used in this study begin in 1988. The mean depth and width are 4.3 and 80.5 ft, respectively. The cross-sectional area at this gauge is 342.9 ft^2, with a bankfull discharge of 2772 cfs having a return interval of 1.65 years. The Rosgen stream classification for this river site is F3 (Miller and Davis 2003), which represents the gentle gradient at this site, the high width-depth ratio, lateral instability,

and a riffle-pool bed morphology (Rosgen 1994). The drainage area of this gauge is 63.7 mi^2.

Discharge data for these sites were obtained from the USGS for the entire available record where 15-minute records exist. Discharge data to at this temporal frequency is available from 1987 through 2014 at the Neversink River near Claryville, NY gauge, and from 1988 through 2014 at the Esopus Creek at Allaben, NY gauge. Gaps in these data may be attributed to ice. The quality of the discharge data are assumed to be suitable for this study.

3.2. Discerning rapid floods in hydrologic record (Objective 1)

A flash flood-detecting algorithm was developed and run over the hydrologic record for both sites to empirically detect rapid floods in the record. This algorithm was constructed as a decision tree:

1. The script isolated 24 hours of discharge data, and found the maximum discharge during that day. If the peak did not exceed bankfull discharge, then the peak was not qualified as a flood and the algorithm shifted to the next 24 hour segment. If this condition was met, the algorithm moved to the next criterion.
2. If the peak exceeded bankfull discharge, then the minimum discharge value in the 6 hours preceding the flood peak was identified. In the same step of the decision tree, the average discharge for two days from 6 hours prior to the peak was calculated. This value, with 500 cfs added for minor fluctuations, is

considered baseflow in this study. This is a suitable estimation for baseflow as it is relative to the surrounding conditions and does not rely on other estimations of seasonality.

3. If the minimum value in the 6 hours preceding the peak was greater than baseflow, then while this event may have been a flood, this peak did not meet the criteria for a flash flood. In this case, the algorithm disregarded any variables and moved to begin the process again by shifting to the next 24 hour segment of record. If the minimum value in the 6 hours preceding the peak was less than or equal to baseflow, then this peak was retained as a flash flood for this study.

This process ensures that floods rising from baseflow to a maximum exceeding bankfull within six hours are detected, with respect to the preexisting stream conditions and regardless of season. Flash floods detected in this method were manually inspected to ensure that rapid increases in discharge were not attributed to missing data. This decision tree is illustrated in Figure 3.

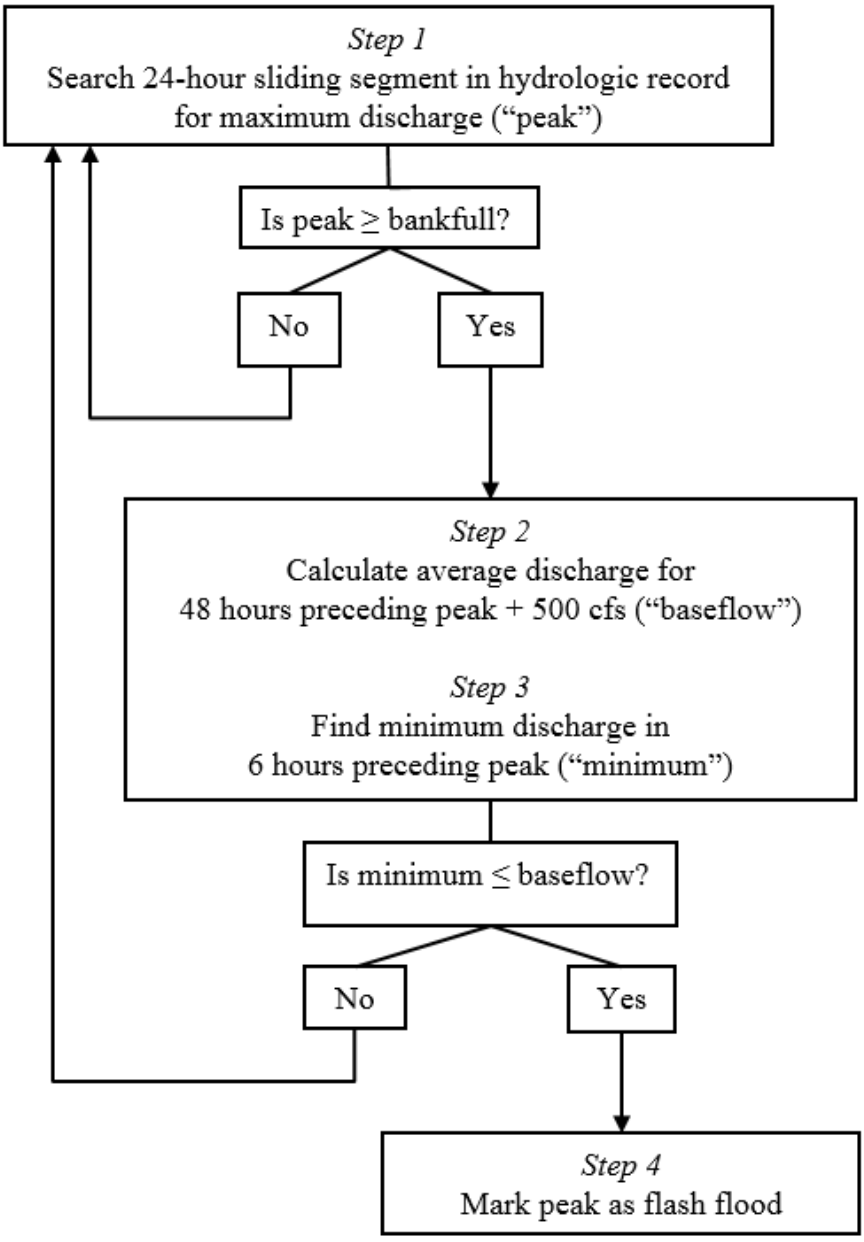

Figure 3. Schematic diagram of the algorithm used to identify flash floods

Weather forecast maps from the Weather Prediction Center will be used to describe each flood day individually. Data from the United States Global Historical

Climate Network (GHCND) will also be used to supplement these individual flood day descriptions. The GHCND site USC00307799 at Slide Mountain, NY at 2649.9 ft ASL was selected for its location between the two gauge sites as well as its representation of the higher elevations in the watersheds. Daily data are available for the study period with few gaps.

3.3. Associating rapid floods to synoptic-scale atmospheric conditions (Objective 2)

The application of synoptic typing in this study will be to determine which large-scale atmospheric patterns are associated with observed flash flood events. Implementing this method links environmental conditions to the simultaneous or preceding atmospheric conditions, which may influence the development or onset of the flash flood conditions. Synoptic typing can be executed in a variety of ways, as listed in detail in Yarnal (1993). The typing for this analysis was performed using an existing program that has been accepted in similar environment-to-circulation studies (e.g. Smith 2012).

3.3.1. Spatial Synoptic Typer Tools (STT)

The Spatial Synoptic Typer Tools (Spatial STT) is a program written in Interactive Data Language (IDL) to performing a supervised map-pattern classification of gridded atmospheric data, particularly National Center for Environmental Protection / National Center for Atmospheric Projections gridded geopotential height (Blair et al. 2011). The classification is performed through principal component analysis (PCA).

PCA is an eigenvector-based classification technique which first determines the underlying patterns in the data, to which a clustering algorithm can be applied to distinguish a user-defined number of map types (Yarnal 1993).

When directed to input climate data, STT executes a loop conducting a PCA with a range of user-defined numbers of principal components and output classes, and a K-means cluster analysis. The program prompts several options, including whether a correlation or covariance matrix should be used to calculate the eigenvectors, and whether or not to perform an orthogonal rotation on the components. The program provides a scree plot and between-type variance and among-type variance plots for the range of user-defined number of principal components to retain and clusters to form, from which the user can discern how many principal components to retain and how many classes to use.

National Center for Environmental Prediction / National Center for Atmospheric Research (NCEP/NCAR) 500 mb geopotential height over the study region were the input data from which synoptic weather types were classified. Because the majority of floods peaked during night and early morning hours local time (to be discussed in Chapter IV), the classification analysis was performed using data from 6z. The data were bound by 277.5°-290° E and 37.5°-35° N to form a six-by-six grid of data, with the study watersheds slightly in the eastern portion of the grid. As in Kidson (1994), data were not stratified by season; thus, the relative frequencies of the types are anticipated to reflect the mid-latitude seasonality of the study area.

A bootstrapping methodology was used to determine if the days preceding flash flood events were a statistically significant departure from normal conditions. First, the 5 days preceding each of the flood days were compiled into a subset. The synoptic type for each of these 115 pre-flood days was recorded and the frequency of each type in this pre-flood subset was calculated. This calculation of type frequencies was then performed for 10,000 random samples of 115 days throughout the study period. The frequency of each type for all of the bootstrapped random samples were compiled into a histogram to illustrate the normal frequency of occurrence of each type. The 2.5^{th} and 97.5^{th} percentiles were calculated for each of the types and plotted onto the histogram. If the number of occurrences of a type in the pre-flood subset was outside these limits, then that type was considered to show significant deviation from normal occurrences in the days preceding a flood at the $\alpha = 0.05$ level.

3.4. Regression analysis for local variables (Objective 3)

Linear regression was performed to determine if local hydrometeorological conditions could be used to predict the peak discharge of a flash flood event. If a relationship is found, discharge should approximately equal an intercept or starting value (β_0) with the influence ($\beta_1, \beta_2, \ldots, \beta_n$) of each predictor ($x_1, x_2, \ldots, x_n$) with a random error ($\varepsilon$).

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \varepsilon$$

Equation 1

Antecedent bankfull, antecedent soil moisture, 24-hour precipitation amount, 24-hour precipitation duration, 24-hour precipitation intensity, and maximum precipitation were used as predictors. A description of each of these predictors follows Table 1 and in the subsequent paragraphs. In Table 1, the variables removed from analysis due to multicollinearity are italicized.

Table 1. Description of predictor variables

Predictor variable	Definition, units
Antecedent percent bankfull	Percent of bankfull discharge 4 days preceding flood peak, %
Soil moisture	Average root zone soil moisture for day preceding flood peak, kg/m^2
24-hour precipitation total	*Sum of hourly precipitation for 24 hours preceding flood peak, mm*
Precipitation duration	Number of hours with precipitation recorded for 24 hours preceding flood peak, hours
24-hour precipitation intensity	$\frac{24-hour\ precipitation\ total}{Duration}$, mm/hr
Maximum precipitation intensity	*Maximum precipitation recorded in one hour for the 24 hours preceding flood peak, mm/hr*

Hourly root zone soil moisture data (hereafter simply "soil moisture") were obtained from the North American Land Data Assimilation (NLDAS)-2 Noah Model at .125 degree resolution for the area bounded by -74.307° E, -74.653° E, 42.131° N,

41.88° N for the day preceding each flood event. The average of these hourly data for each day provided an estimate of soil moisture for the day preceding each flood event.

Hourly precipitation data were obtained from the National Climate Data Center at the cooperative station Claryville, NY (COOP: 301521) from 1987 through the last available records in 2011. This station is located at 41.9134°, -74.5723° and is 503.8 m above sea level. Because not every flash flood day was represented in this dataset, data from cooperative station at Mongaup Valley 4 SSW, NY US (COOP: 305435) were substituted. Data for this station, located at 41.5705°, -74.7933° and 379.8 m above sea level, were available through 2012. A two-sample F-test for variances showed that the variances in 24-hour precipitation totals between the Claryville precipitation data (mean = 47.0 mm (1.9 in), standard deviation = 25.3 mm (0.93 in)) and Mongaup Valley precipitation data (mean = 68.3 (2.5 in), standard deviation = 29.0 mm (1.1 in) were equal ($p < 0.05$). A two-tailed paired-samples t-test was then performed on precipitation data for days included in both the Claryville and Mongaup Valley to determine if significant differences between the two stations exist on flood days. There was no significant difference in the Claryville precipitation data and the Mongaup Valley precipitation data distributions ($p < 0.05$). These preliminary analyses support the substitution of Mongaup Valley precipitation data as appropriate.

Hourly precipitation from the Claryville and Mongaup Valley stations together were available for 26 days on which flash flood peaks occurred in the Neversink River and Esopus Creek watersheds occurred. Both stations provided hourly precipitation data for 13 flash flood days. Given the closer proximity of the Claryville station to the study

watersheds, data from this station were preferred over data from the Mongaup Valley station on days when data were available from both stations. Overall, hourly precipitation data were available for 26 of the flash floods, leaving 6 flash floods without hourly precipitation data and therefore being excluded from these analyses.

To determine if the precipitation data from the Claryville and Mongaup Valley stations were significantly different from each other, an F-test for variance and a two-tailed paired-samples t-test were performed. Because temporal differences were expected in the hourly precipitation due to their distance, the hourly precipitation was summed for the 24 hours preceding the flood peak in each watershed to provide 24-hour total precipitation in these preliminary analyses.

A correlation matrix was used to determine if any of the standardized predictor variables had a greater influence on other predictor variables than the model. Both watersheds and both precipitation datasets were included to maintain the highest possible number of observations. A high correlation between two predictors suggests collinearity and consequent inflation of the influence that that factor has on the model. This problem is accounted for by removing one of the variables with high correlation from the model. A threshold of absolute correlations greater than 0.5 resulted in the removal of one of the variables in this study. Removing the variables with dependence on other variables also improves the model by removing extraneous variables that lead to overfitting the model.

Multiple combinations of these variables were assembled, and a linear regression was performed for each combination. The 24 combinations are illustrated in Figure 4. First, the combinations were separated by including either the Claryville and Mongaup

Valley precipitation stations together or just the Claryville station. This differentiation was included to determine if the Mongaup Valley station data had an effect on the predictive ability of the model. The second difference in combinations was whether to include both watersheds together, or each watershed separately. While including both watersheds incorporated the highest number of flash flood events, separating by watershed was done to determine if using the same precipitation and soil moisture data to predict two different flood peak discharges had an influence on the predictive success of the model. These differences divided the combinations into 6 groups, indicated in Figure 4. The third difference in combinations was to include the peak discharge (cfs) or the log of the peak discharge. This step was included as the discharge appeared to have a more logarithmic trend, which would not be fit well by linear regression. The final difference in combinations was to evaluate the use predictor variables in their original units versus standardized variables. This step was included as the range of magnitudes of the predictor variables was large. A z-score transformation was performed to standardize the data. Predictor variables were retained if they were significant at the $p = 0.05$ level for all combinations. Including these 4 different combinations of the predictor and discharge data in each of the 6 groups resulted in 24 separate linear regressions.

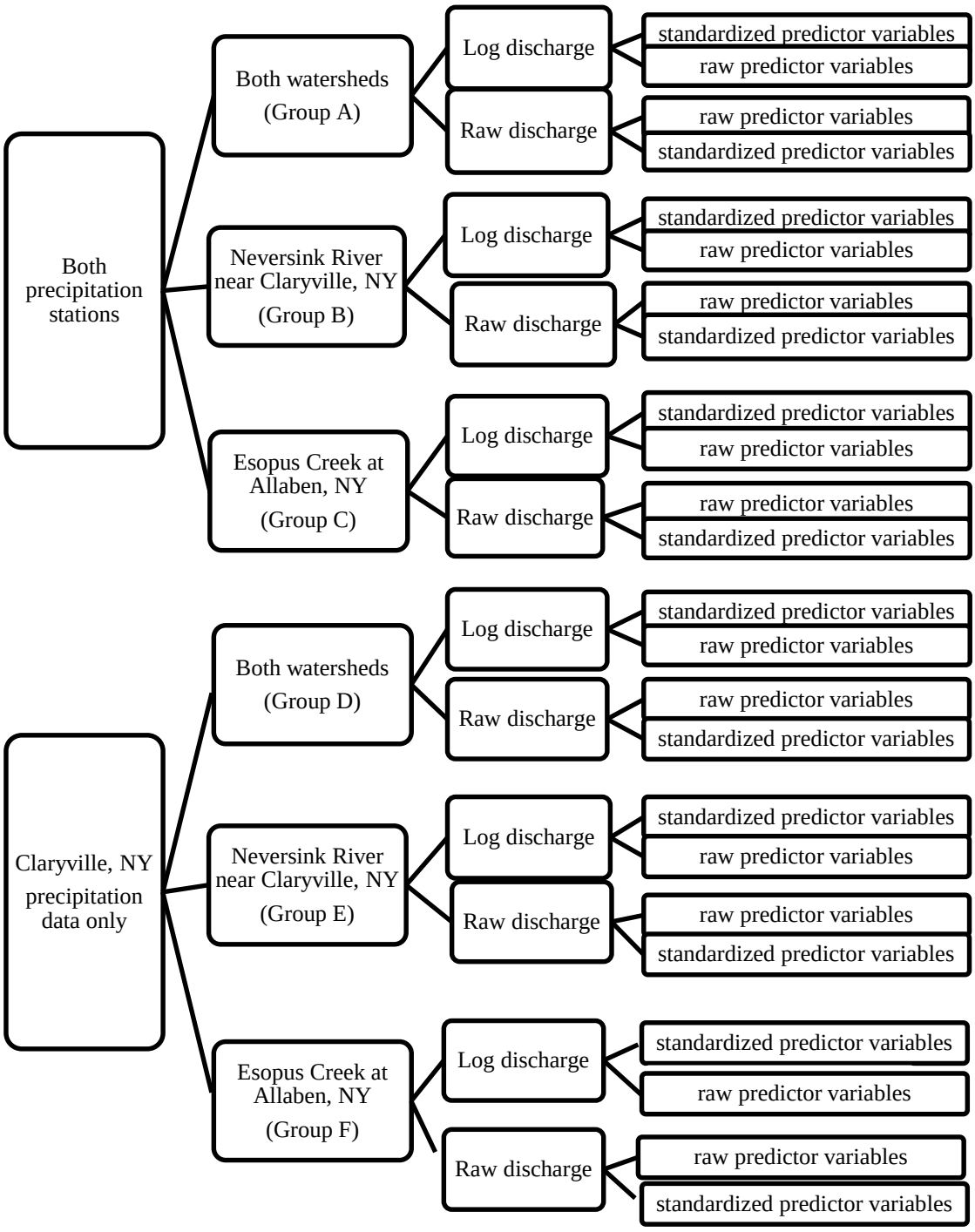

Figure 4. Variable combinations in multiple linear regression analyses

Following the methodology of Lumia et al. (2006) and Roland and Stuckey (2008), stepwise linear regression was performed with the standardized predictors for both watersheds using precipitation data from both precipitation stations. Predictors were manually added and removed until an optimal model was produced. Predictors were removed in rank order of their *p*-value, then added back in until a balance between the adjusted r^r penalty for including many variables, value of the F-statistic, and minimizing the overall *p*-value was achieved. The primary limitation of this approach is that the number of predictor variables is too high as compared to the number of observations and so there is a tendency to overfit the model.

4. OBJECTIVE 1 RESULTS AND DISCUSSION

4.1. Flash floods detected

The goal of Objective 1 was to detect flash flood within the hydrologic record and associate those peaks with meteorological and streamflow conditions. The hydrologic record was analyzed for flash floods from 01-April-1987 at Neversink River near Claryville, NY and 01-December-1988 at Esopus Creek at Allaben, NY through 29-September-2014 at both sites. If multiple flood peaks were detected within 48 hours, they were assumed to be forced by the same hydroclimatological conditions and only the first peak was retained for analysis, as distinguishing streamflow responses for synoptic-scale conditions at this timescale would be impractical.

Seventeen flash floods were detected at the Neversink River near Claryville site, with discharges ranging from slightly over bankfull at 4360 cfs to five times greater than bankfull at 20900 cfs. Fifteen flash floods were detected at the Esopus Creek at Allaben site, similarly with the discharges ranging from slightly over bankfull at 2940 cfs to more than ten times bankfull at 29300 cfs. Nine of the flash flood events were experienced at both watersheds. Overall, 32 total flash flood peaks in 23 total flash flood days over the study period were detected in the two watersheds.

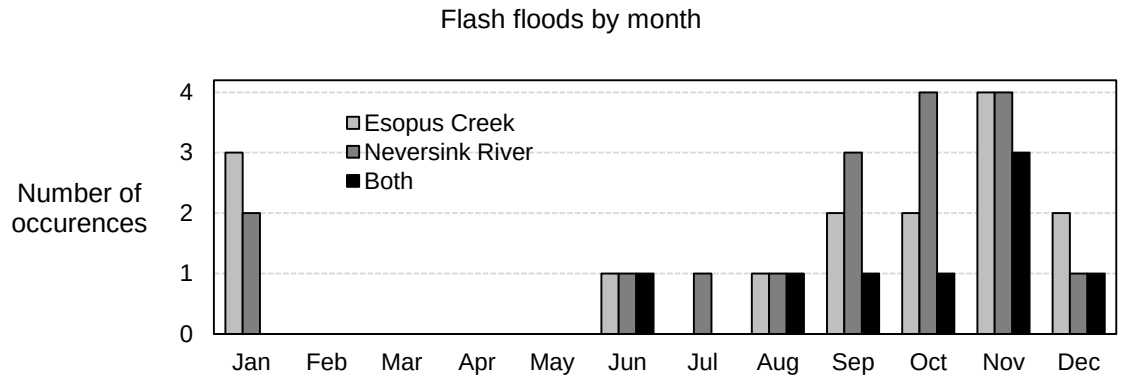

Figure 5. Monthly distribution of flash floods

Figure 5 shows the monthly distribution of flash floods recorded at USGS gauging stations at Esopus Creek at Allaben, NY and Neversink River near Claryville, NY from 1987-2014. The shared count indicates flash floods occurring simultaneously in both watersheds. The majority of flash floods occurred in the fall and winter months, with 14 of the 16 floods at Esopus Creek at Allaben, NY and 13 of the 17 floods at Neversink River near Claryville, NY occurring from September through January (Figure 5). No flash floods were detected at either site from February through May. Figure 6 shows the hourly distribution of flash flood peaks in local time during the study period. The majority of floods were recorded during local morning and daytime hours, with relatively few events peaking between 17:00 and 4:00 local time (Figure 6).

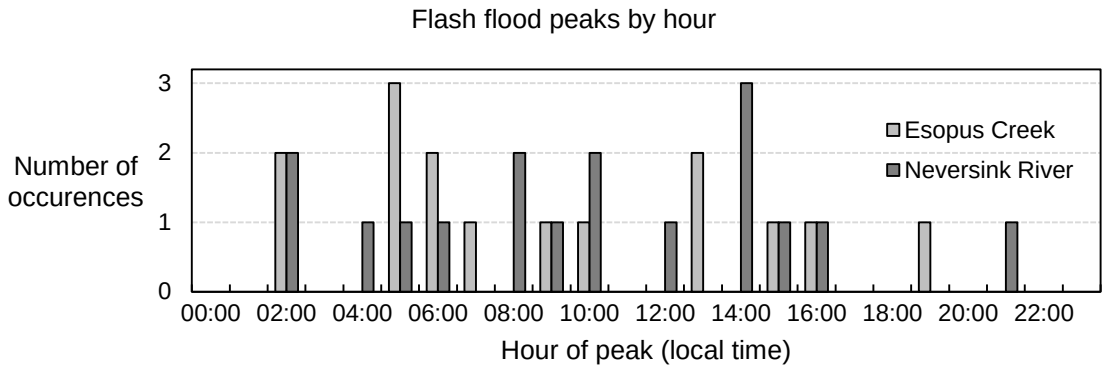

Figure 6. Hourly distribution of flash flood peaks

Several of these days correspond with tropical storms or tropical storm remnants passing over the area, including Tropical Storm Tammy, Tropical Storm Lee, and Tropical Storm Irene. Others correspond with notable floods caused by rapid snowmelt, heavy precipitation, or a combination of factors.

4.1.1. Description of flash flood days

A brief description of each flash flood event follows in this section. A weather map is provided for each flash flood day. The maps for days prior to 2003 are from the National Oceanic and Atmospheric Administration (NOAA) Daily Weather Map series. For days during and after 2003, a NOAA Weather Prediction Center (WPC) surface weather or forecast map for the morning each flash flood day is given. The map is followed by a summary of the daily precipitation, maximum and minimum temperatures, and the streamflow as a percent of bankfull discharge. The meteorological data are from the GHCN station at Slide Mountain, NY.

A hydrograph of the 4 days preceding and including the flash flood peak is also given. The blue line shows the river discharge in cfs. The red line shows the bankfull discharge in cfs at the gauge. Each of the hydrographs shows a rapid increase in streamflow, rising from relatively normal conditions to exceeding bankfull discharge in fewer than 6 hours. The streamflow data are from the respective USGS gauge stations, as indicated by the figure title. The percent of bankfull discharge 4 days before each flood peak is also given as a reference for baseflow, in addition to the visual comparison available through the hydrograph figures. The meteorological and stramflow variables can be found in Table 2-A for the events occurring in the Neversink River basin from 1987-2013 (n=17). Similarly, these basic meteorological and streamflow variables corresponding with flood events at the USGS streamflow gauge Esopus Creek at Allaben, NY are found in Table 2-A (n=15).

Table 2. Flood and weather conditions for each flood event in (A) Neversink River watershed and (B) Esopus Creek watershed. Discharge data are from respective USGS streamflow gauges. Meteorological data are from the GHCND station at Slide Mountain, NY.

A Flood and weather conditions for each event in the Neversink River watershed (n=17)

Event	Date	Flood peak discharge (cfs)	Percent of bankfull at flood peak	Percent of bankfull 4 days prior	Precipitation (in)	Snow depth (in)	Snow fall (in)	Maximum temperature (°F)	Minimum temperature (°F)
1	1993- 01- 05	5020	120.0%	22.9%	1.3	8.0	0	50.0	33.1
2	1993- 11- 28	7760	185.6%	3.4%	3.0	0	0	53.1	33.1
3	1995- 11- 12	8030	192.0%	7.7%	3.7	2.0	2.5	55.0	21.0
6	2003- 9- 23	4930	117.9%	7.4%	2.6	0	0	64.0	52.0
7	2003- 10- 29	4570	109.3%	3.5%	1.3	0	0	48.9	34.0
8	2004- 07- 23	6720	160.7%	4.3%	0.4	0	0	75.9	64.0
9	2004- 11- 28	4860	116.2%	2.5%	1.3	0	0	45.0	21.2
11	2005- 10- 08	4360	104.3%	0.7%	4.6	0	0	64.0	61.0
12	2005- 11- 30	6310	150.9%	3.8%	2.9	*Missing*	0	55.9	39.0
13	2006- 01- 18	6600	157.8%	33.2%	1.3	0	*Trace*	53.1	19.0
14	2006- 06- 28	11300	270.2%	2.8%	2.5	0	0	69.1	63.0
15	2007- 10- 27	4580	109.5%	2.7%	1.0	0	0	57.9	43.0
16	2008- 10- 25	4440	106.2%	1.3%	0.1	0	0	51.1	33.1
17	2008- 12- 12	6340	151.6%	3.2%	2.3	0	0	36.0	28.0
21	2011- 08- 28	20900	499.8%	3.6%	8.6	*Missing*	*Missing*	*Missing*	*Missing*
22	2011- 09- 07	5620	134.4%	6.1%	2.8	*Missing*	*Missing*	57.0	52.0
23	2012- 09- 18	17800	425.6%	1.1%	0.7	*Missing*	*Missing*	68.0	46.9

Table 2. (continued)

B Flood and weather conditions for each event in the Esopus Creek watershed (n=15)

Event	Date	Flood peak discharge (cfs)	Percent of bankfull at flood peak	Percent of bankfull 4 days prior	Precipitation (in)	Snow depth (in)	Snow fall (in)	Maximum temperature (°F)	Minimum temperature (°F)
3	1995-11-12	3430	123.7%	5.6%	3.7	2.0	1.5	55.0	21.0
4	1996-01-27	3540	127.7%	21.2%	2.2	9.0	*trace*	44.1	12.9
5	1996-11-09	3530	127.3%	3.9%	3.9	0.0	0.0	60.1	45.0
7	2003-10-29	2940	106.1%	4.4%	1.3	0.0	0.0	48.9	34.0
9	2004-11-28	2940	106.1%	2.3%	1.3	0.0	0.0	45.0	21.2
10	2005-01-14	2970	107.1%	5.8%	1.6	2.0	2.0	55.0	28.0
12	2005-11-30	3020	108.9%	5.5%	2.9	*Missing*	0.0	55.9	39.0
14	2006-06-28	4420	159.5%	4.0%	2.5	0.0	0.0	69.1	63.0
17	2008-12-12	2970	107.1%	4.1%	2.3	0.0	*trace*	36.0	28.0
18	2010-01-25	5020	181.1%	2.0%	1.7	11.0	0.0	48.9	24.1
19	2010-10-01	10100	364.4%	0.4%	1.4	0.0	0.0	68.0	51.1
20	2010-12-01	4090	147.5%	5.9%	1.4	0.0	0.0	51.1	37.0
21	2011-08-28	29300	1057%	4.4%	8.6	*Missing*	*Missing*	*Missing*	*Missing*
22	2011-09-07	5730	206.7%	11.4%	2.8	*Missing*	*Missing*	57.0	52.0
23	2012-09-18	6120	220.8%	0.7%	0.7	*Missing*	*Missing*	68.0	46.9

Event 1: 1993-1-5

The surface weather map 1993-1-5, the date of the first flood, shows a closed low-pressure minima over northern New York and northern New England (Figure 7). A cold front accompanied this low pressure, extending over the Appalachian Mountains and advancing on the study region in the Catskill Mountains of the northern Appalachians.

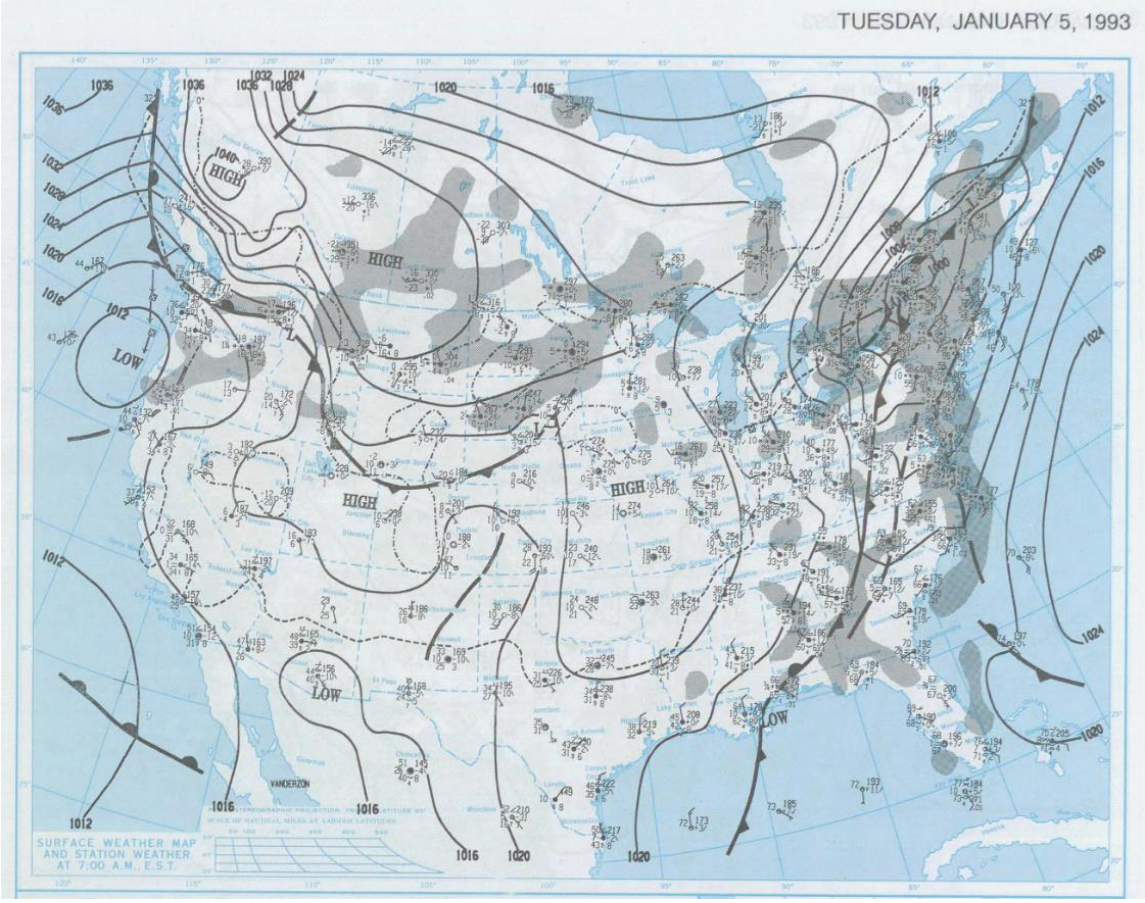

Figure 7. Surface weather map for 1993-1-5 (NOAA Central Library Data Imaging Project, hereafter NOAA CLDIP).

The Neversink River watershed experienced a flash flood during this event, while the Esopus Creek watershed did not. The discharge of the Neversink River at the gauge was 22.9% of bankfull 4 days prior to the flash flood, then increased to 120% of bankfull at the peak of the flash flood (Figure 8, Table 2). On the day of the flash flood, 1.3 in (33.8 mm) of precipitation fell on 8 in (203 mm) of snow, with temperatures between 33.8°F (6°C) and 32.1°F (10°C) (Table 2).

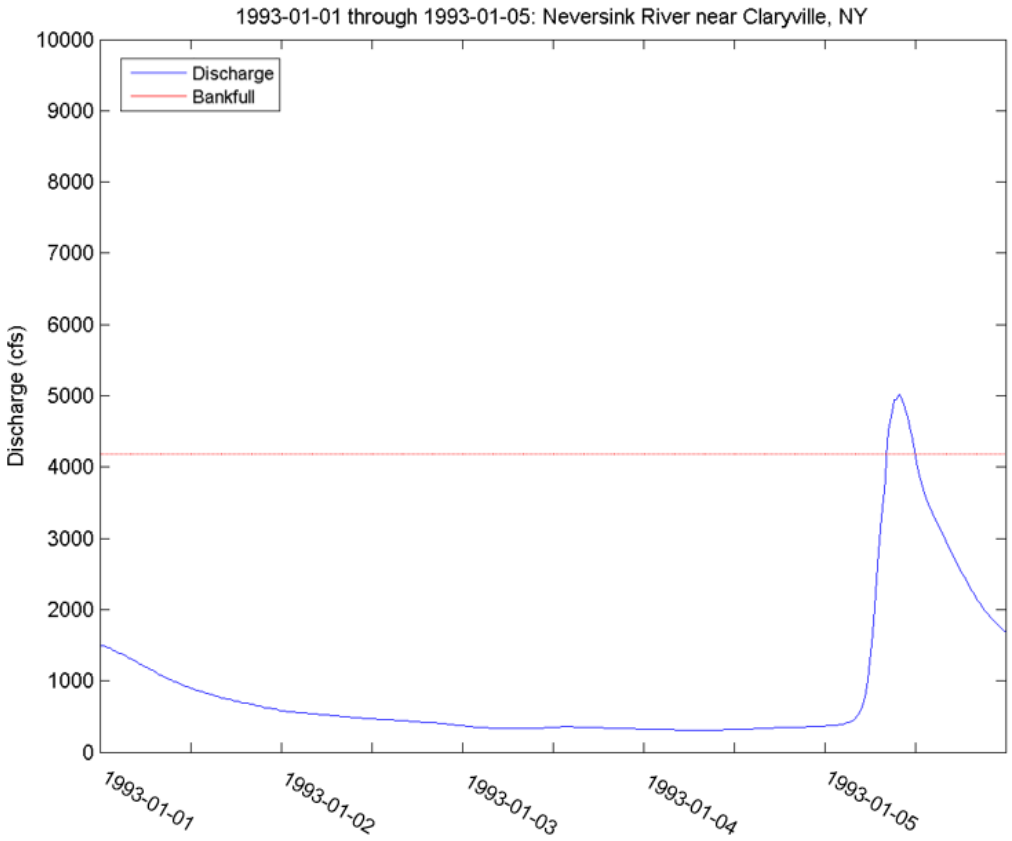

Figure 8. Hydrograph for Neversink River, 1993-01-01 through 1993-01-05. These 15-minute discharge data are from USGS gauge Neversink River near Claryville, NY.

Event 2: 1993-11-28

The weather map for 1993-11-28 shows a closed low-pressure centered over New York (Figure 9). This pressure minima moved northward into the study region along the Appalachian Mountains. A cold front existed south of the pressure minima, advancing on a warm front off the coast of Maryland and Virginia. While not clear in Figure 9, these fronts likely formed an occluded front in Pennsylvania and New York.

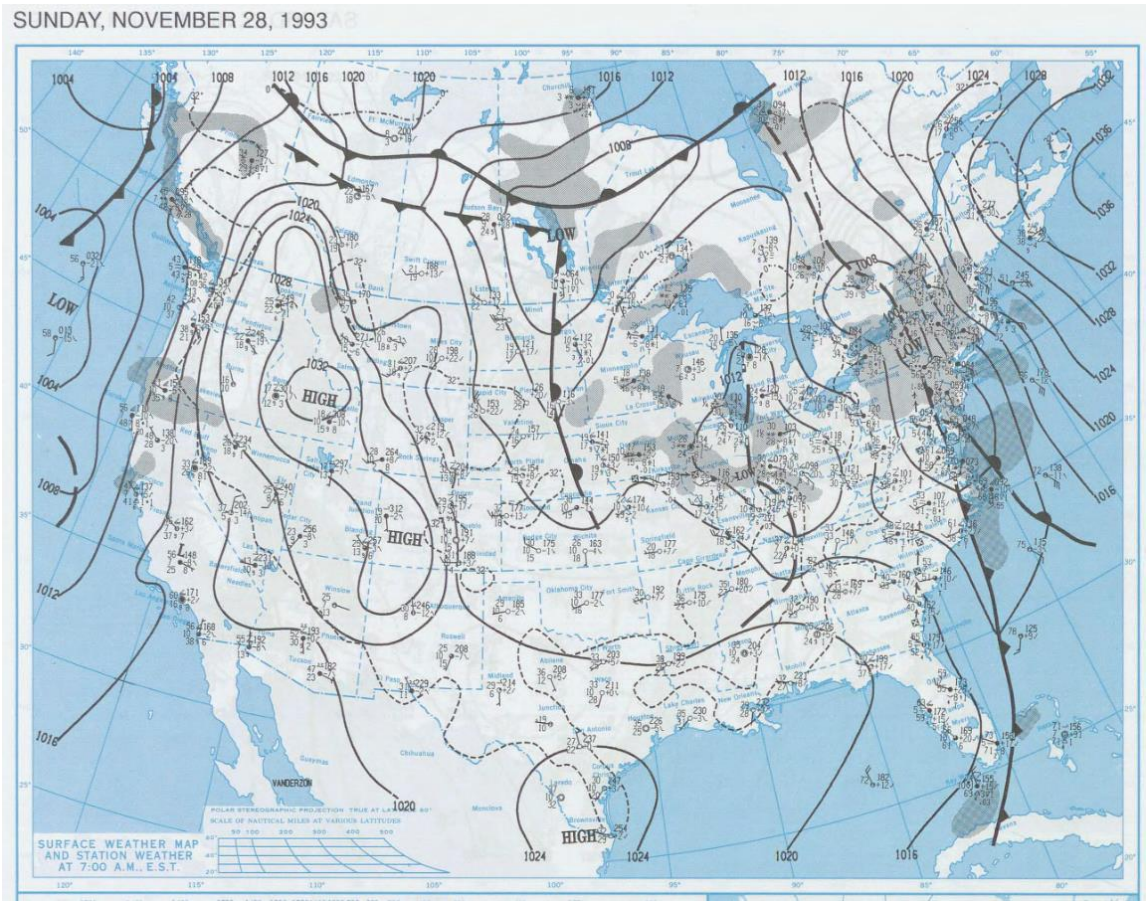

Figure 9. As in Figure 7, but for 1993-11-28 (NOAA CLDIP).

On the day of the flood peak, 76.2 mm of precipitation fell, with temperatures between 33.1°F (6°C) and 53.1°F (11.7°C) (Table 2). A flash flood on this day was only recorded in the Neversink River watershed, where discharge increased from only 3.4% of bankfull four days prior to the flood, to 185.6% of bankfull discharge at the flood peak (Figure 10, Table 2-A).

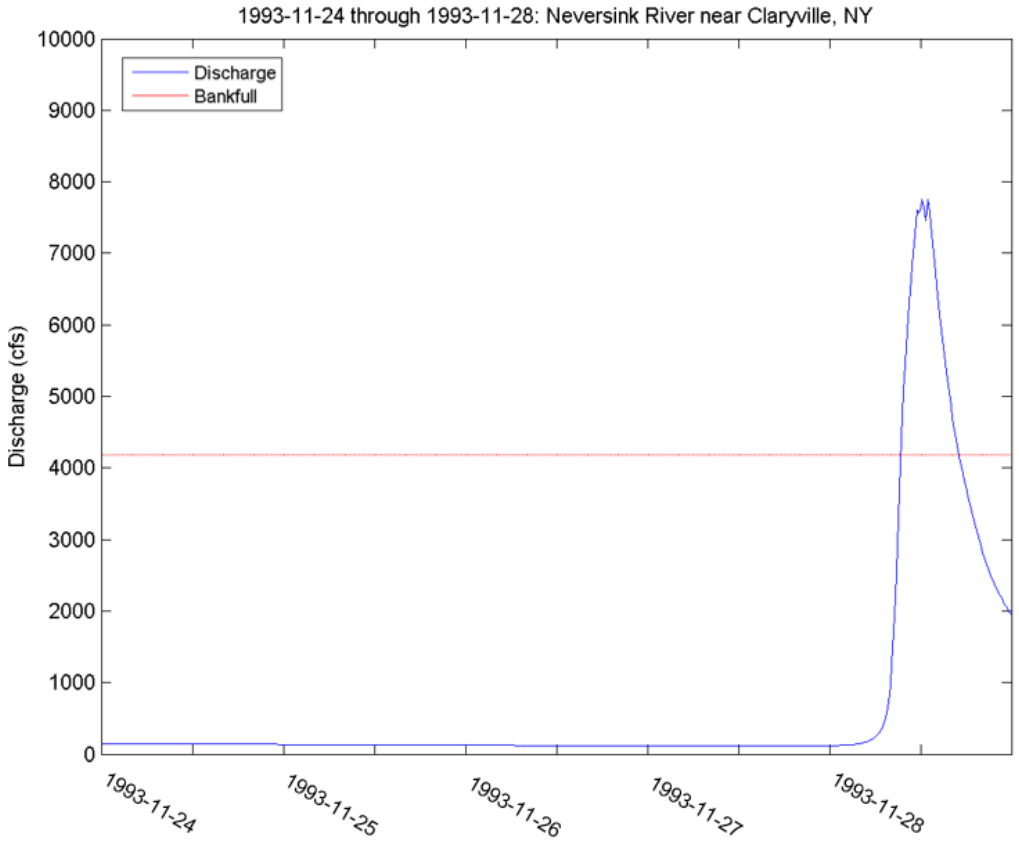

Figure 10. As in Figure 8, but for 1993-11-24 through 1993-11-28

Event 3: 1995-11-12

The weather map for 1995-11-12 shows an elongated closed low-pressure system moving over the Great Lakes region in the days preceding the flash flood event (Figure 11). The cold front accompanying the low pressure extended through Quebec and New England, angling offshore in Massachusetts. Areas of high pressure existed behind the cold front nearly uniformly, with high pressure areas ranging from the southeastern United States through central Canada. Other fronts exist in the western half of the visible continent, though lacking the pressure gradient seen surrounding the cold front located over the study region. Ninety-three mm (3.7 in) of precipitation was recorded on this day, including 2.0 in (38 mm) of snow which fell on 2.5 in (51 mm) of accumulated snow (Table 2-A, Table 2-B). The maximum temperature on 1995-11-12 was 55.0°F (12.8°C) and the minimum temperature was below freezing at 21°F (-6.1°C) (Table 2-A, Table 2-B).

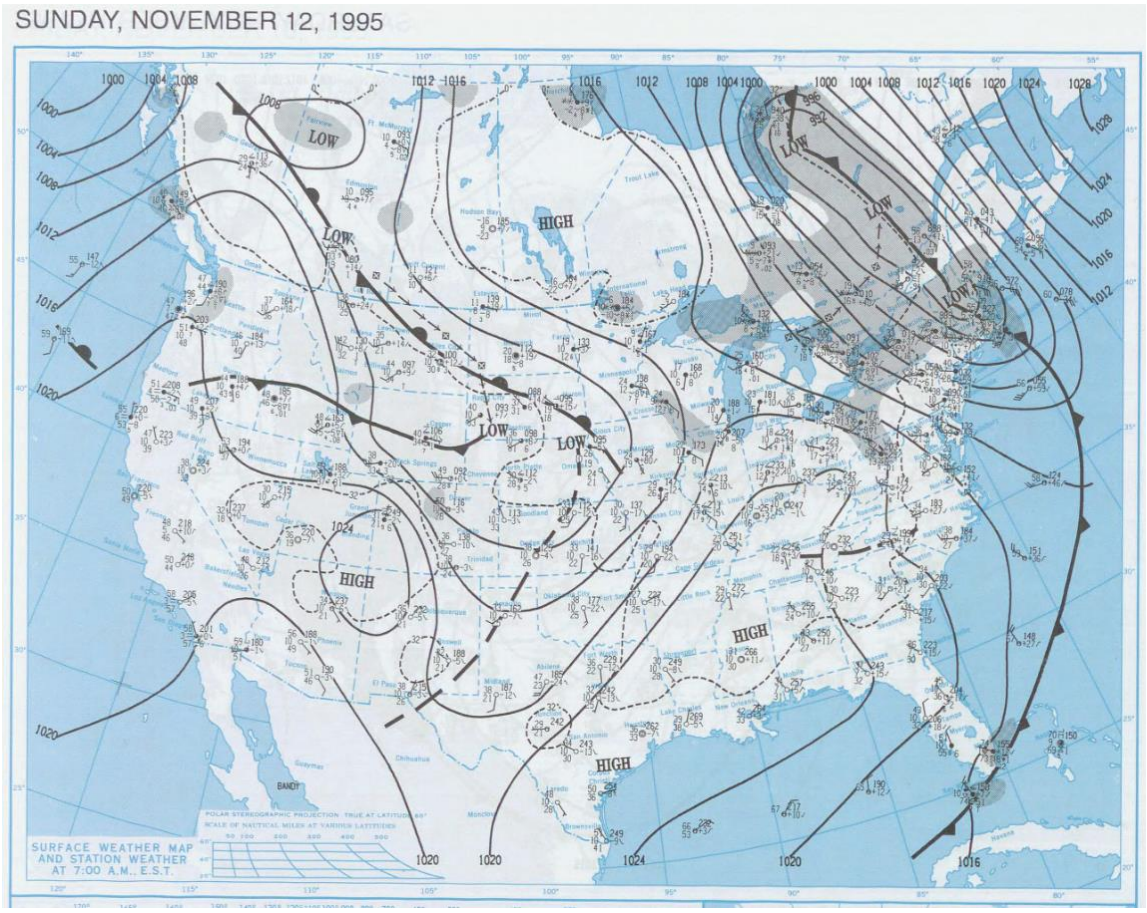

Figure 11. As in Figure 7, but for 1995-11-12 (NOAA CLDIP).

On this day, flash floods were recorded in both the Esopus Creek and Neversink River watersheds. The discharge at the Esopus Creek watershed was 5.6% of bankfull four days before the flood peak, and rose to 123.7% of bankfull at the peak of the flash flood (Table 2-A, Figure 12).

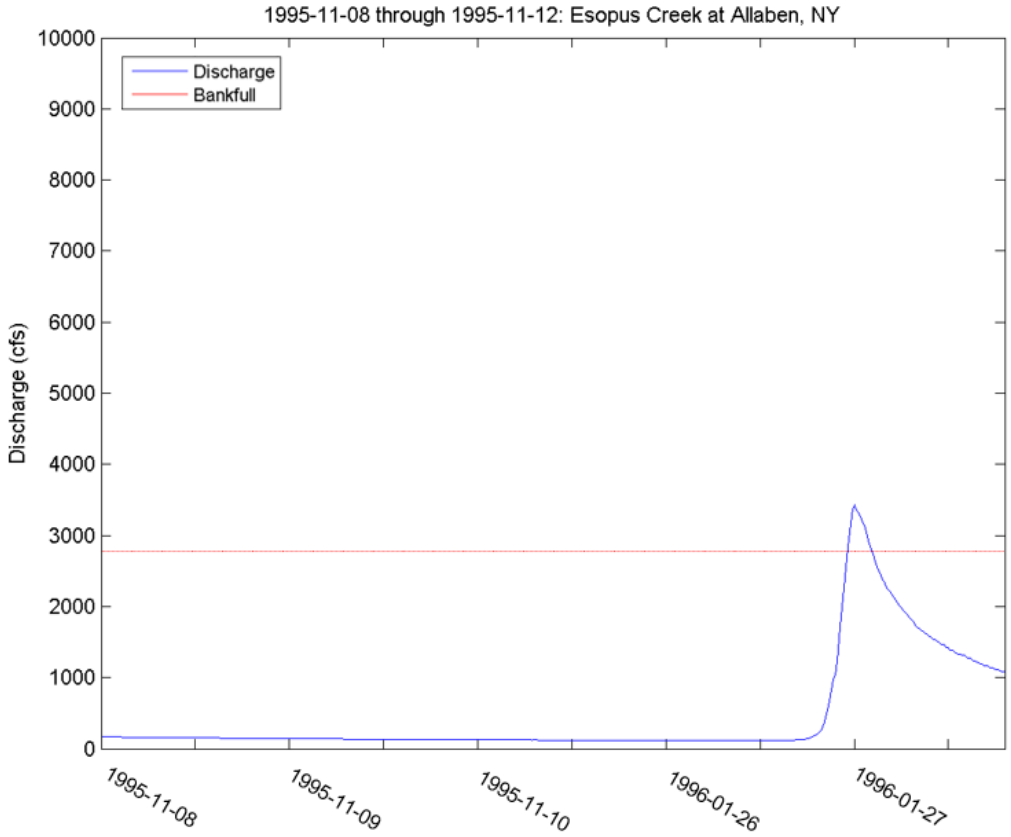

Figure 12. Hydrograph for Esopus Creek, 1995-11-08 through 1995-11-12. These 15-minute discharge data are from USGS gauge Esopus Creek at Allaben, NY.

The discharge at the Neversink River was slightly higher 4 days before the flood, at 7.7% of bankfull, and rose to 192% of bankfull during the flash flood peak (Figure 13).

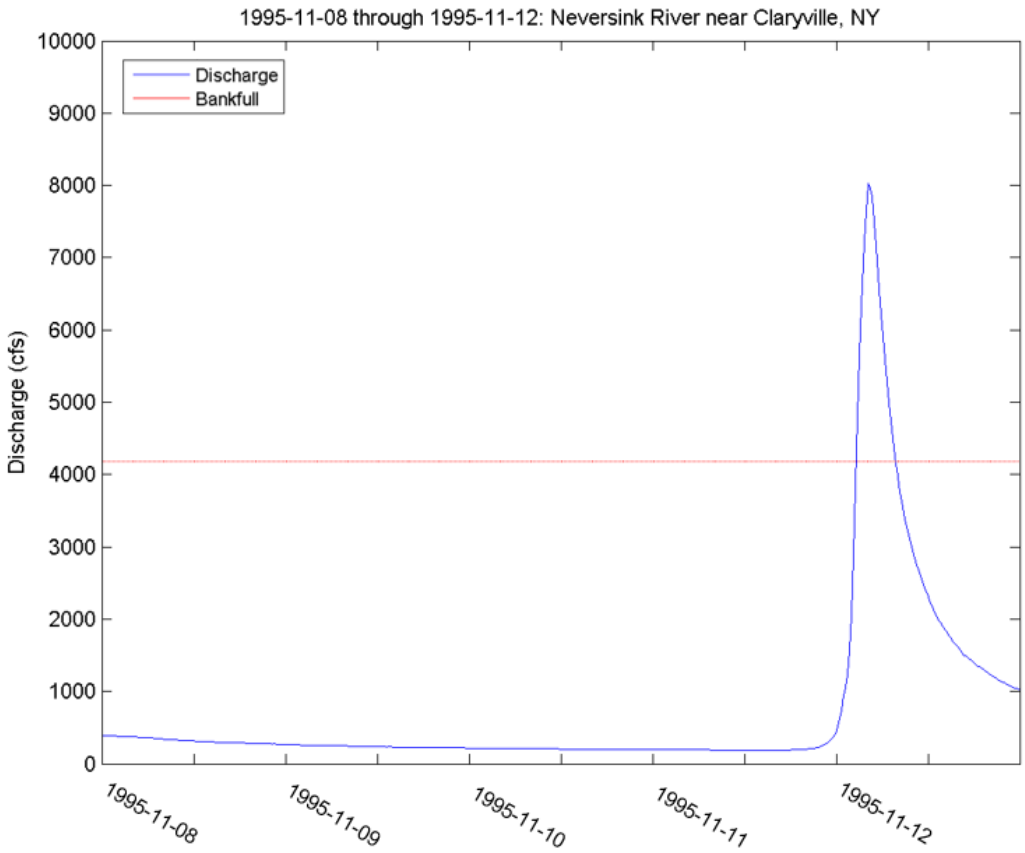

Figure 13. As in Figure 8, but for 1995-11-12

Event 4: 1996-1-27

The flash flooding event of 1996-1-27 occurred with the passage of a strong mid-latitude cyclone through the region centered over the Michigan and Ontario, directly over Lakes Superior, Michigan, and Huron (Figure 14). This pressure minima passed over Iowa, Wisconsin, and northern Lake Michigan in the days preceding the flash flooding event. Several fronts affecting the study region are present. The latitudinal warm front extending from the pressure minima ranged through Ontario and Quebec,

moving northward. The longitudinal cold front extending from the pressure minima bent southward, covering lower Ontario, Ohio, and West Virginia. Opposing remnants of the warm front were positioned to encompass the high elevations of Virginia and West Virginia, dissipating into Pennsylvania and the southern tier of New York. Wind in the study area was from the southeast. These conditions are typically associated with moisture advection from the Atlantic.

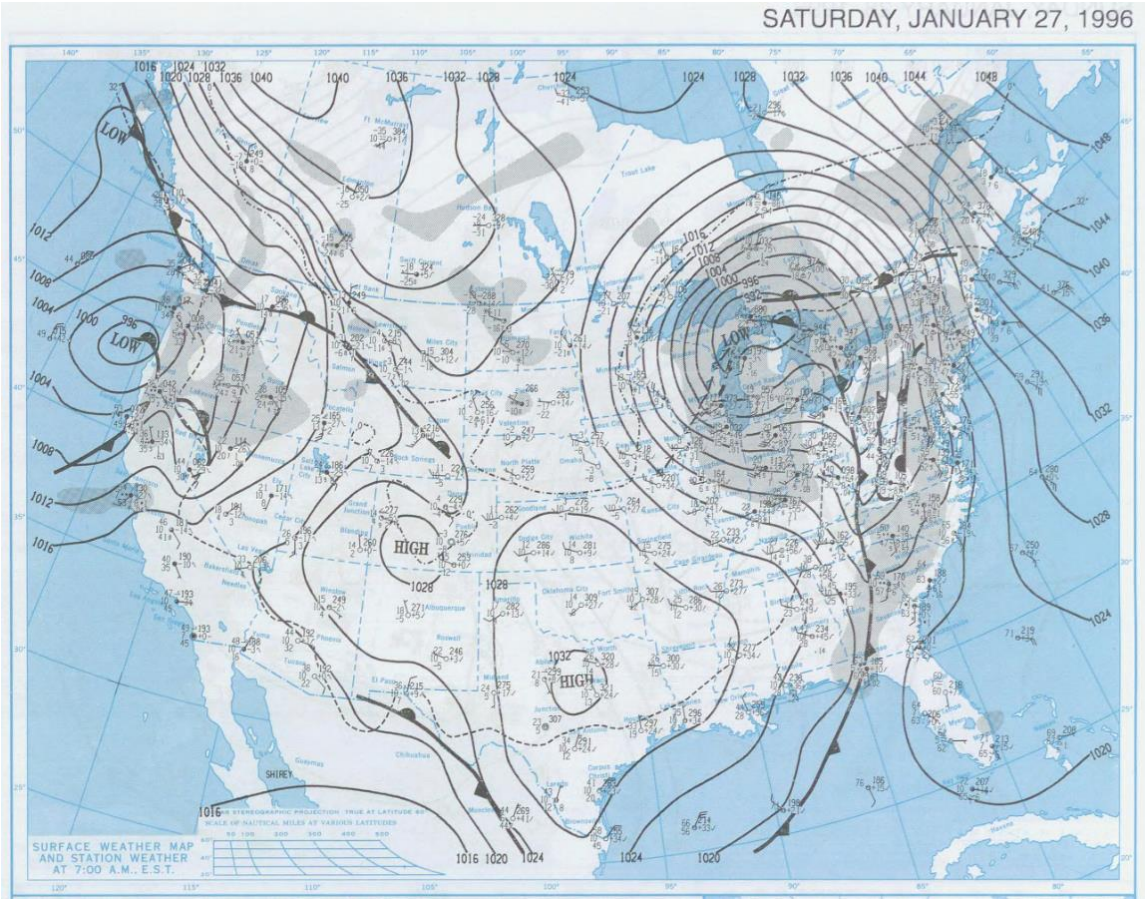

Figure 14. As in Figure 7, but for 1996-01-27 (NOAA CLDIP).

About 2.2 in (549 mm) of precipitation was recorded on the day of the flood peak, falling on 9 in (229 mm) of accumulated snow (Table 2-B). While the maximum temperature was 44.1°F (6.7°C) and the minimum temperature was 12.9° (-10.6°C), no snowfall was recorded on this day (Table 2-B). The Esopus Creek watershed flooded in less than 6 hours on this day, increasing from 21.2% of bankfull four days before the flood to 127.7% of bankfull at the peak of the flash flood (Table 2-B, Figure 15).

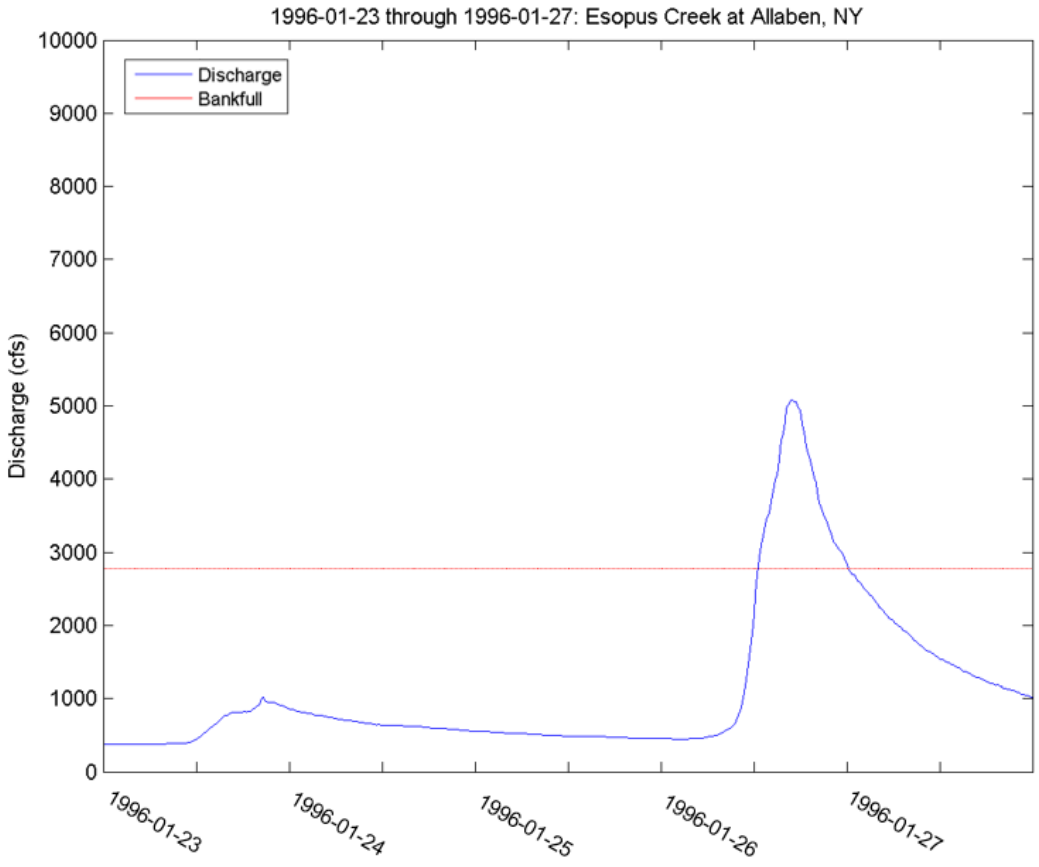

Figure 15. As in Figure 12, but for 1996-01-23 through 1996-01-27

Event 5: 1996-11-9

The weather map for 1996-11-9 shows a closed low pressure minima centered over Montreal, Quebec, Canada with a stationary front extending northward through Quebec into Newfoundland and a cold front extending southward through New York veering offshore in New Jersey (Figure 16). Surface winds were southerly on this day in the study watersheds.

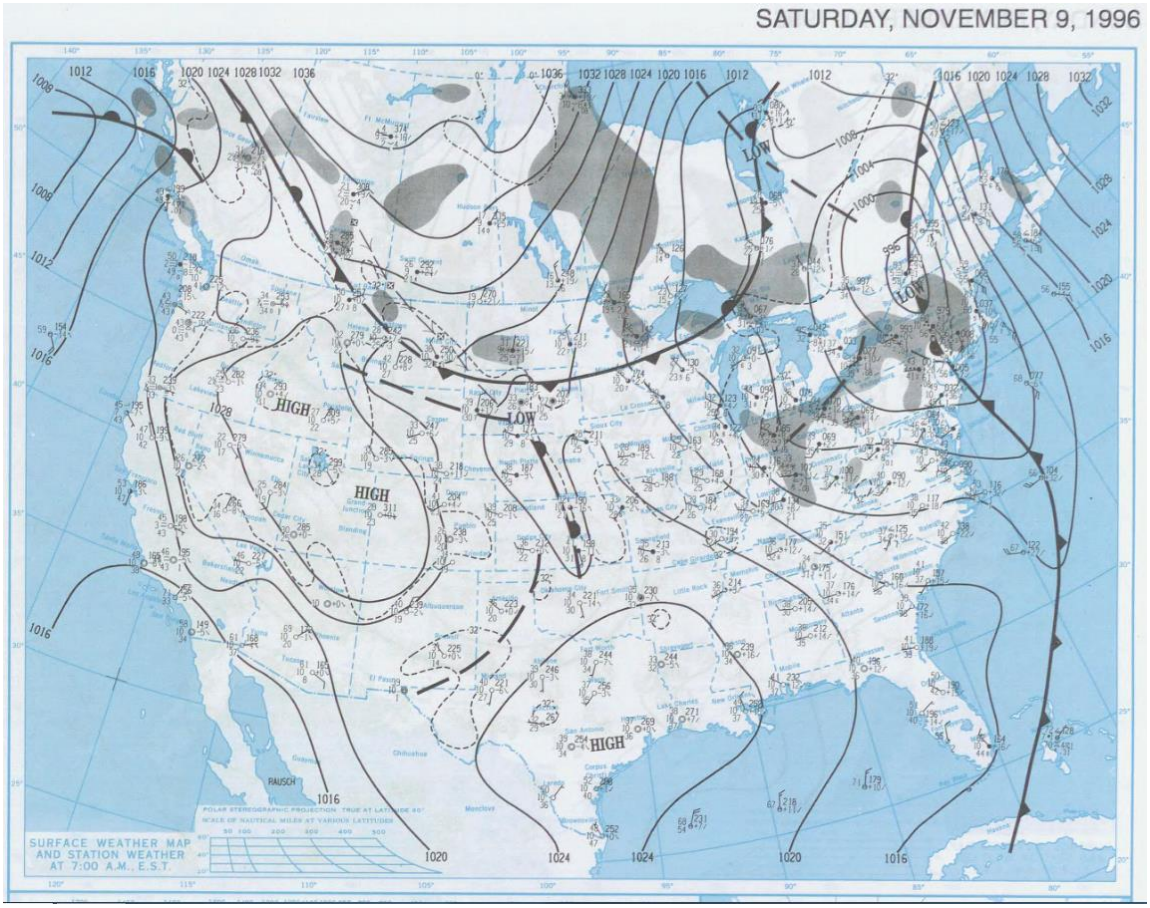

Figure 16. As in Figure 7, but for 1996-11-09 (NOAA CLDIP).

This is one of the wettest events. This low pressure minima resulted in 98 mm of precipitation in the watershed, with temperatures ranging between 45.0°F (7.2°C) to 60.1°F (15.6°C)(Table 3). The Neversink River watershed did not produce a discharge that qualified as a flash flood during this day. The Esopus Creek watershed did show a flash flood on this day, rising from only 3.9% of bankfull four days before the flood to 127.3% of bankfull during the flash flood peak (Figure 17).

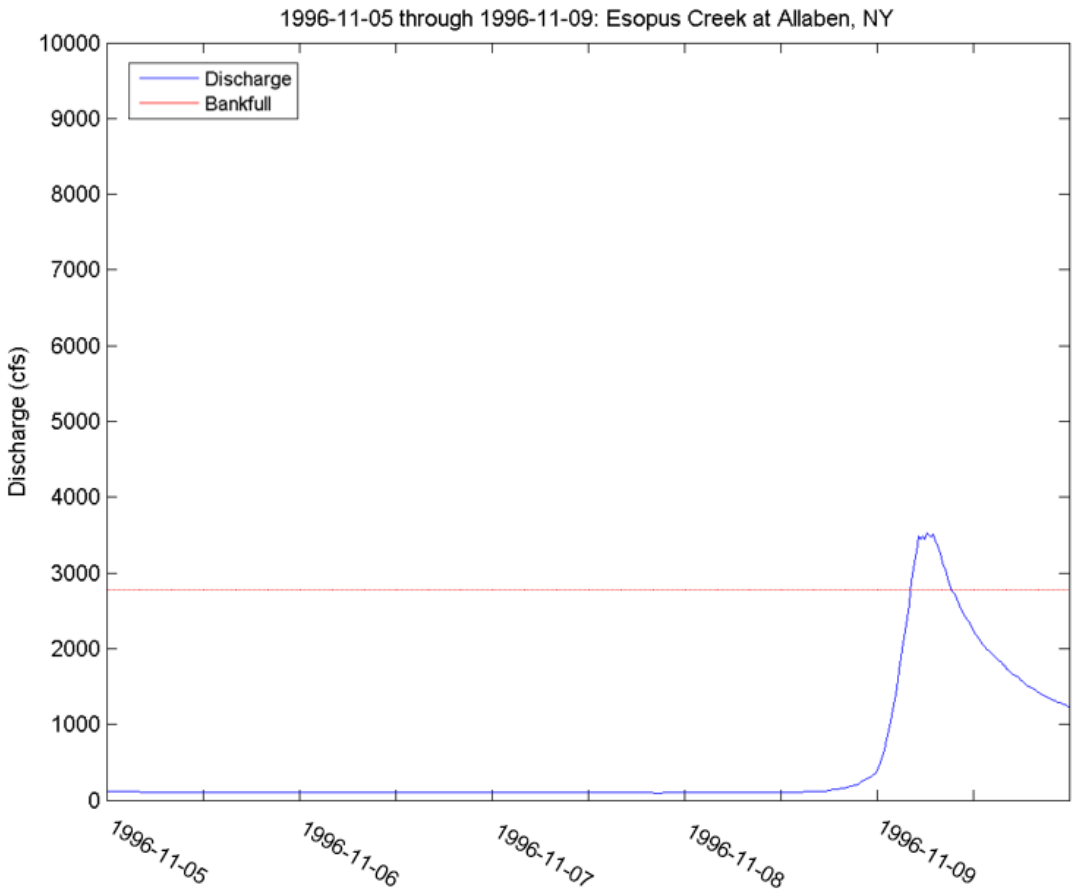

Figure 17. As in Figure 12, but for 1996-11-05 through 1996-11-09

Event 6: 2003-9-23

The flash flooding event occurring on 2003-9-23 was accompanied by a mid-latitude cyclone centered over New England, with the cold front extending southward along the eastern seaboard (Figure 18). The map shows precipitation and thunderstorms in the region, including the study watersheds, which are commonly associated with the frontal activity present with the mid-latitude cyclone.

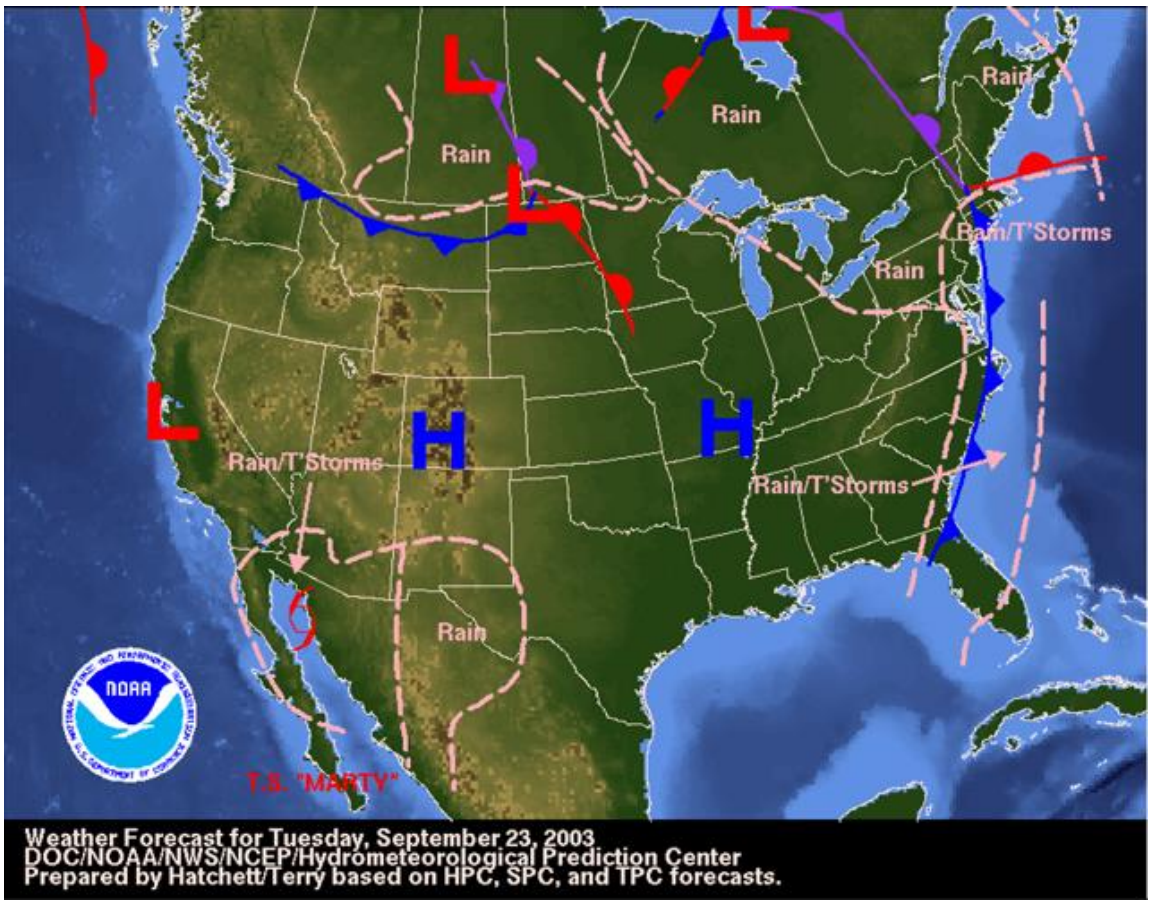

Figure 18. NOAA WPC Weather Forecast map issued in the morning for 2003-9-23 (NOAA National Weather Service / Weather Prediction Center, hereafter NOAA NWS/ WPC).

With temperatures between 52°F (11.1°C) and 64°F (17.8°C), 2.6 in (66.3 mm) of precipitation fell on this day (Table 2). A flash flood was recorded in the Neversink River watershed on this day, but not in the Esopus Creek watershed. The discharge of the Neversink River at the gauge increased from 7.4% of bankfull four days before the flood to 117.9% during the flood peak (Table 2-A, Figure 19).

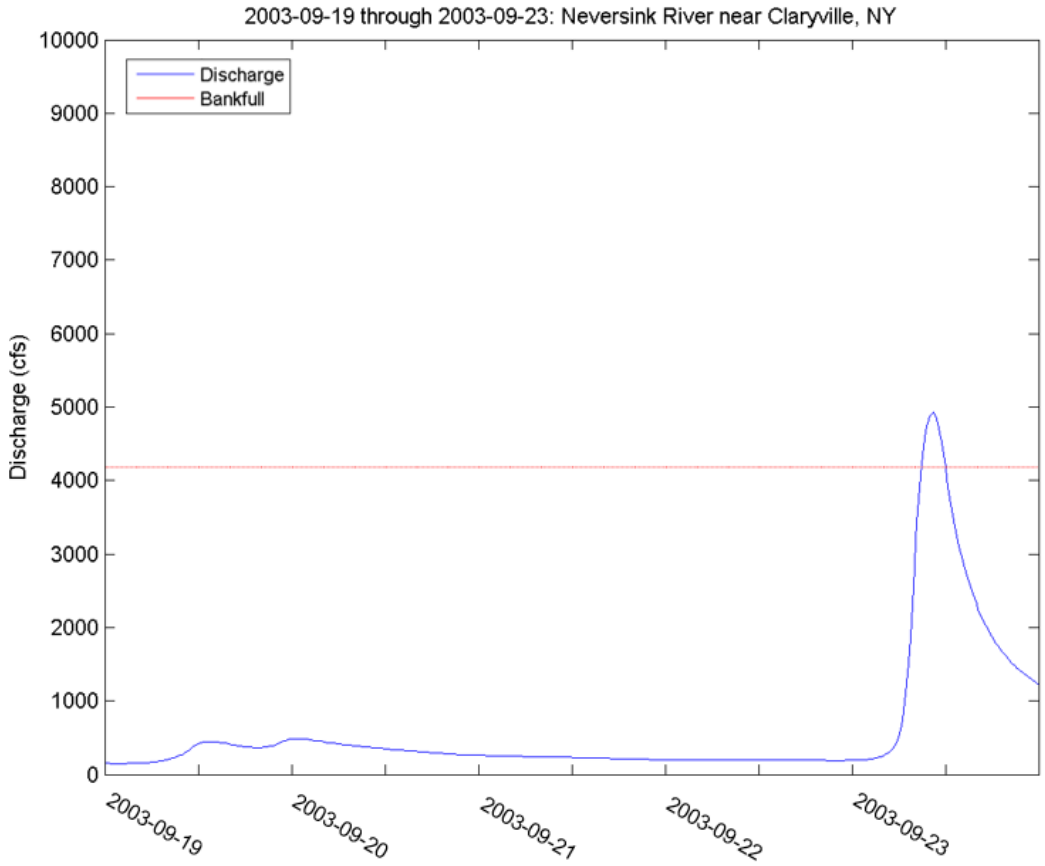

Figure 19. As in Figure 8, but for 2003-09-19 through 2003-09-23

Event 7: 2003-10-29

The weather forecast map for 2003-10-29 shows a mid-latitude cyclone centered over Maine and New Hampshire, with the warm front extending through Maine and New Brunswick and the cold front extending southward through Massachusetts and bending out over the Atlantic coastline south of Cape Cod (Figure 20). A second cold front passed over the Great Lakes and approached the study region from the west. This placed the study watersheds directly between the two cold fronts. The secondary front was indicated as being stationary over the Great Lakes region. The forecast map shows possible flash flooding in the low-pressure center over Maine and New England, though this does not encompass the watersheds in which flash floods were detected for this study. Two cold fronts also exist in the western United States, the southernmost of which is likely indicative of another mid-latitude cyclone. The forecast map shows that most of the United States was under areas of low pressure, with only the southeastern United States between the two cold fronts as experiencing high pressure.

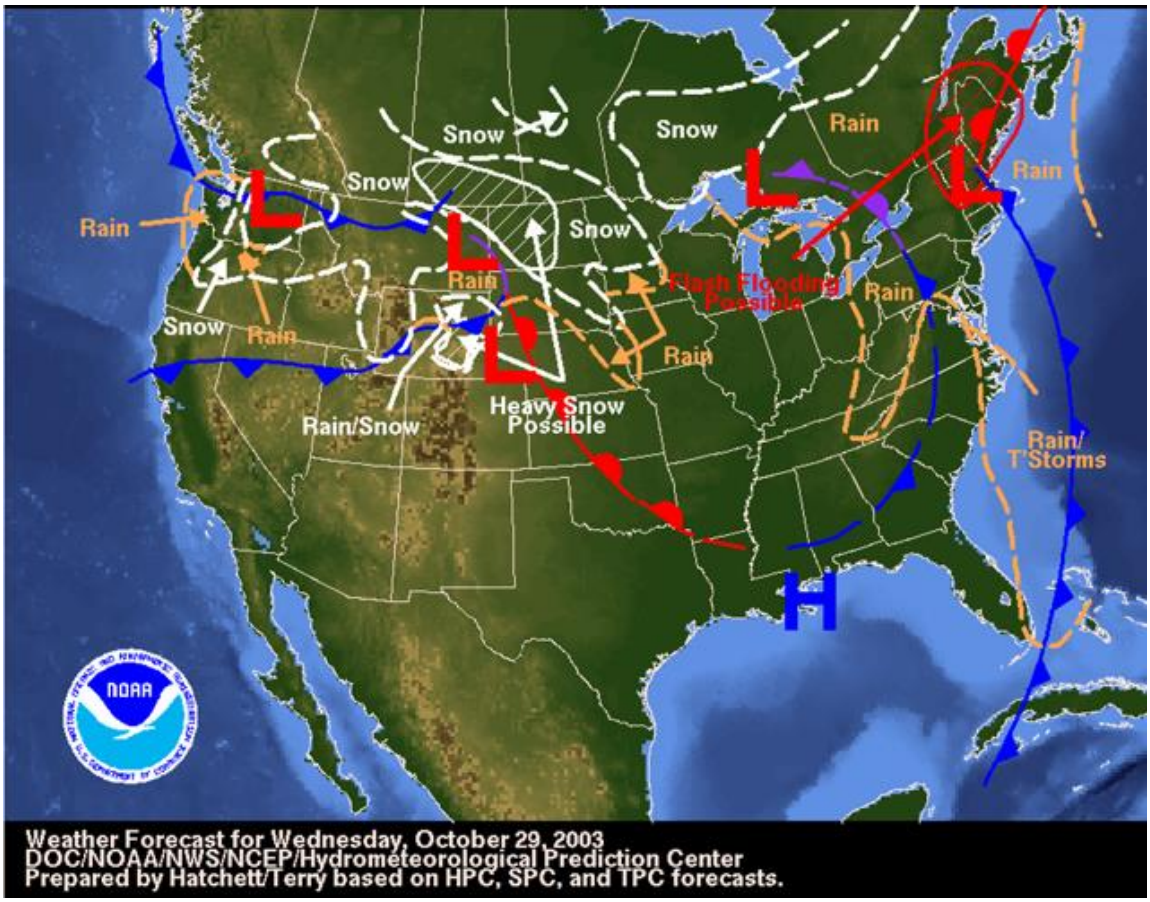

Figure 20. NOAA WPC Weather Forecast map issued for 2003-10-29 (NOAA NWS/ WPC)

On the day of the flood peaks, 1.3 in (33 mm) of precipitation fell, with temperatures between 34.0°F (1.1°C) and 48.9° F (9.4°C)(Table 2-A, Table 2-B). Flash floods were detected in both watersheds on this day. The discharge at the Esopus Creek gauge increased from 4.4% of bankfull to 106.1% during the flood peak (Table 2-B, Figure 21).

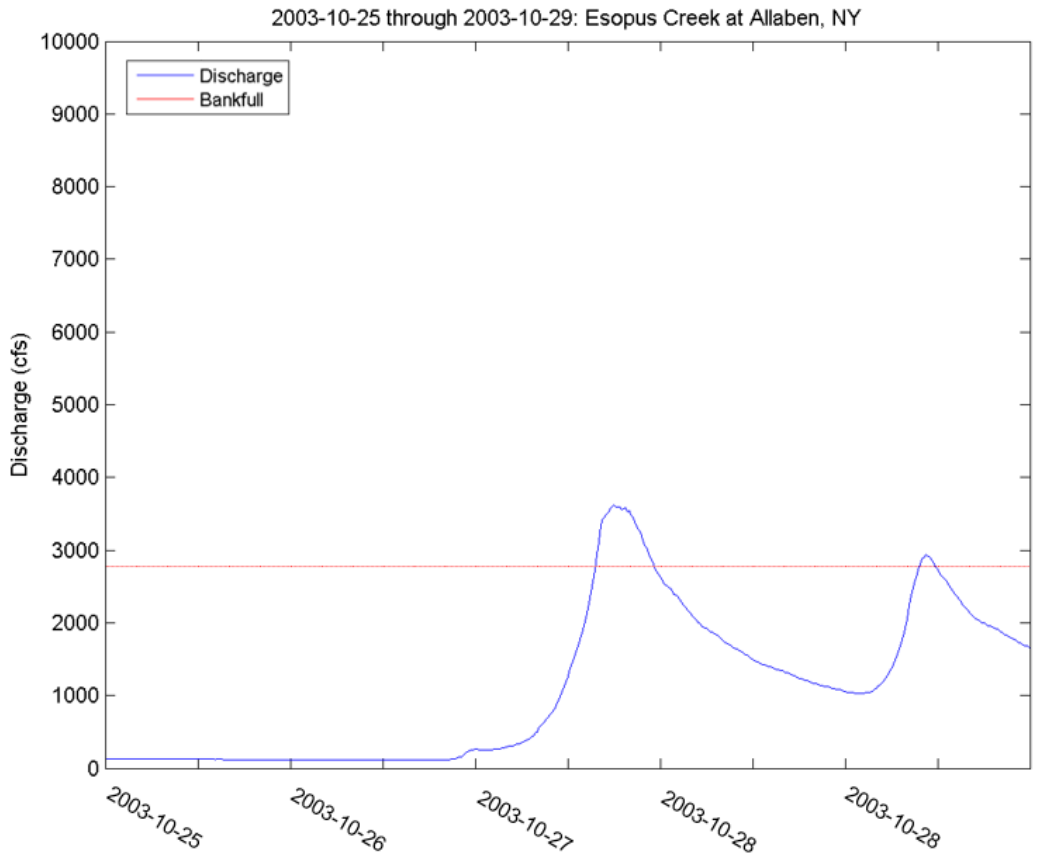

Figure 21. As in Figure 12, but for 2003-10-25 through 2003-10-29

The discharge at Neversink River gauge increased from 3.5% of bankfull to 109.3% of bankfull at the peak flood discharge (Table 2-A, Figure 22).

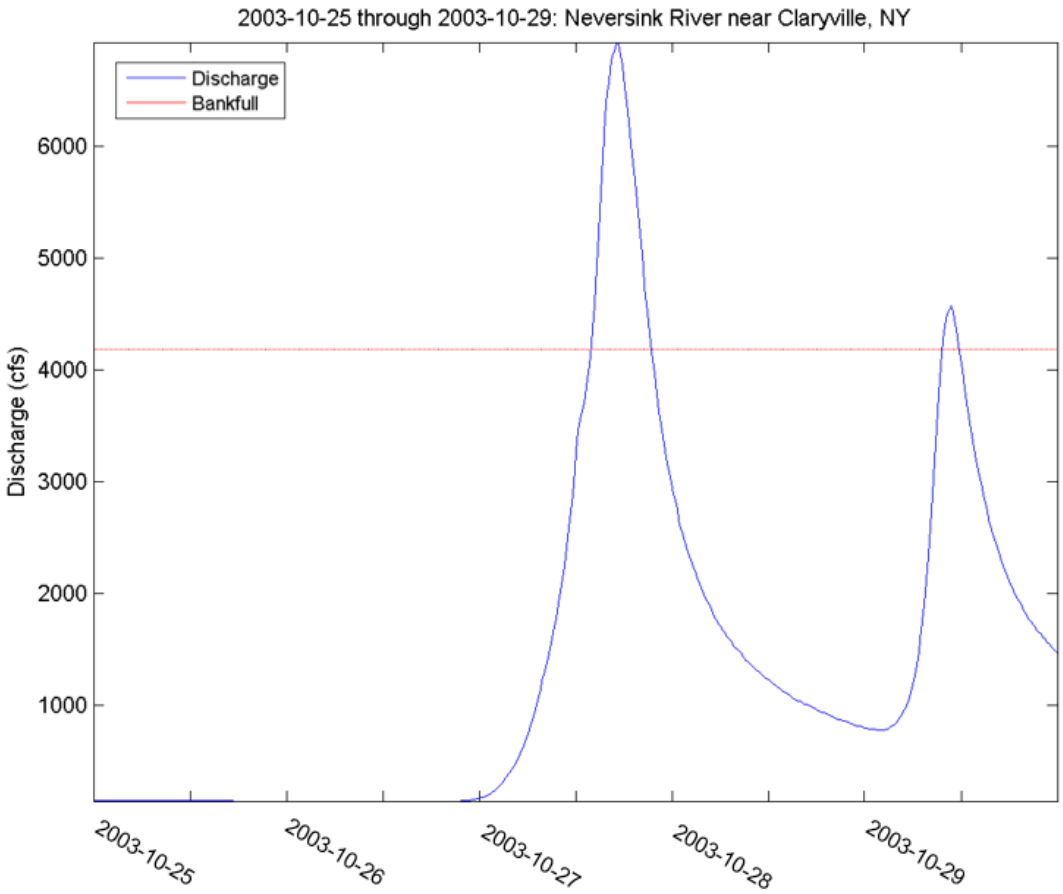

Figure 22. As in Figure 8, but for 2003-10-25 through 2003-10-29

Event 8: 2004-7-23

The flash flood event of 2004-7-23 occurred with the approach and passage of a cold front over the study region (Figure 23). The cold front extended from Quebec through Tennessee, passing directly over New York and approaching the study watersheds. The forecast map suggests rain and thunderstorms along the frontal boundary. Temperatures ranged from 64°F (11.1°C) to 75.9°F (17.8°C) during the flood peak day, with 0.4 in (11.4 mm) of precipitation (Table 2-A).

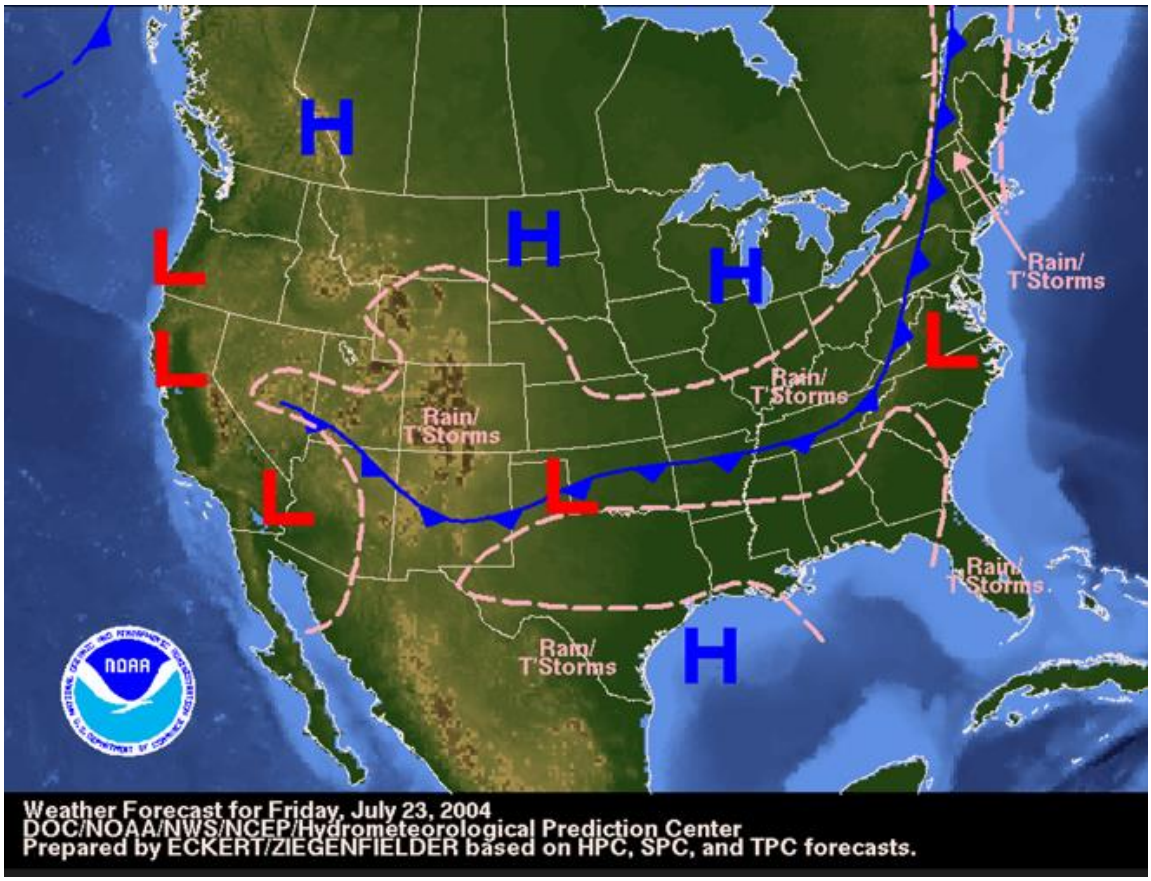

Figure 23. NOAA WPC Weather Forecast map issued for 2004-07-23 (NOAA NWS/ WPC).

The flash flood detected on this day occurred in the Neversink River watershed, where discharge rose from 4.3% of bankfull discharge to 160.7% at the flood peak (Table 2-A, Figure 24).

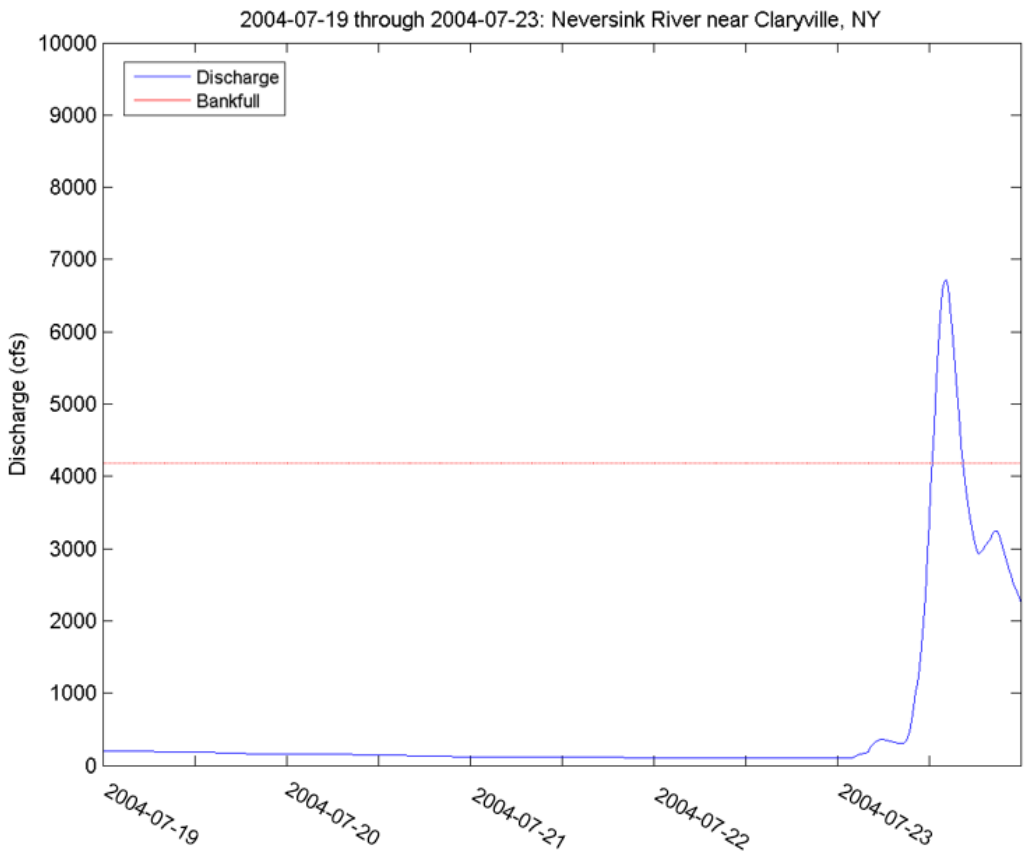

Figure 24. As in Figure 8, but for 2004-07-19 through 2004-07-23

Event 9: 2004-11-28

The weather forecast map for 2004-11-28 shows two distinct low-pressure systems acting on the study region (Figure 25). The cold front curving out of the low-pressure minima in Ontario is moving eastward across New York, potentially forcing precipitation in the study watersheds. A second pressure minima exists off the coast of Delaware and New Jersey. Precipitation associated with warm fronts typically occurs ahead of the frontal boundary on the surface, possibly causing precipitation in the study

region or interacting with the cold front from the more northerly pressure minima. The forecast map shows rain and thunderstorms in the study watersheds. With temperatures ranging from 21.2°F (-6°C) to 45.0°F (7.2°C), 1.32 in (33.5 mm) of precipitation was recorded (Table 2-A, Table 2-B).

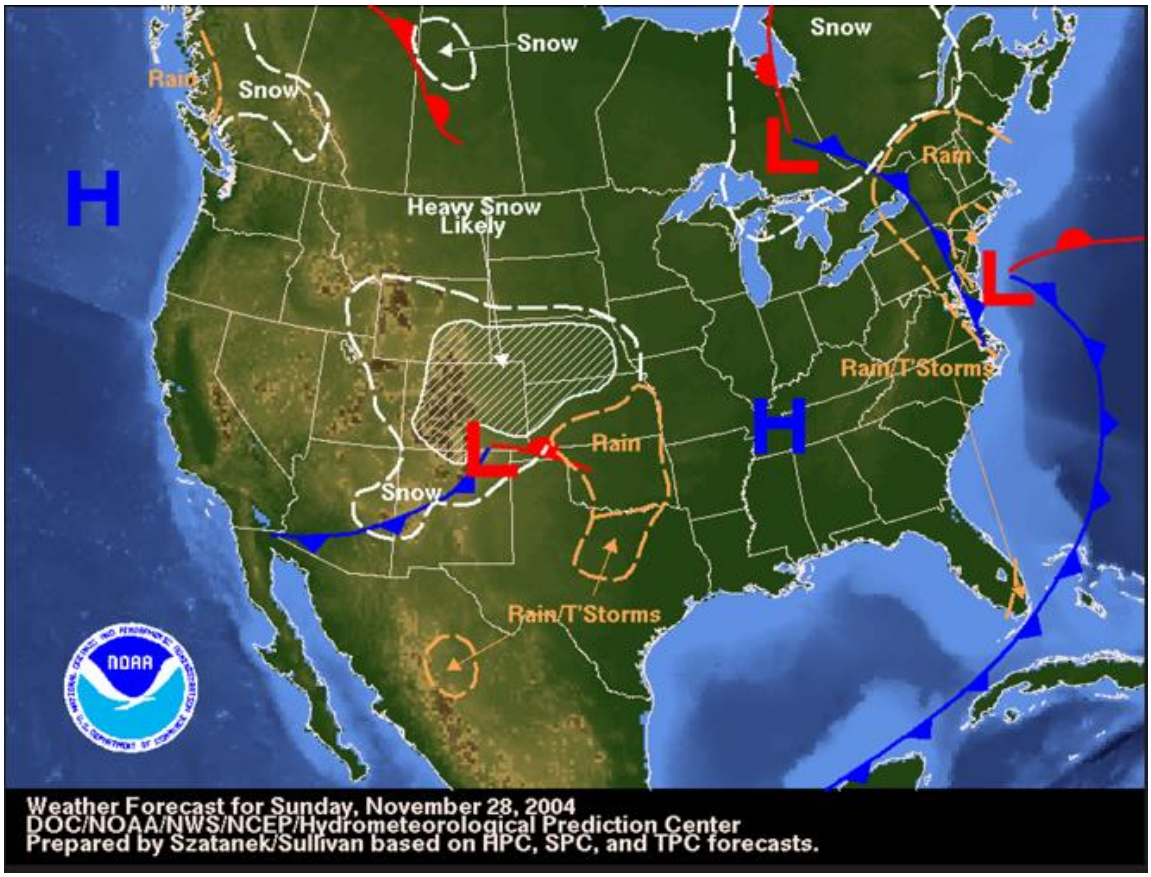

Figure 25. NOAA WPC Weather Forecast map issued for 2004-11-28 (NOAA NWS/ WPC).

Flash floods were detected in both watersheds, with the discharge rising from 2.5% of the bankfull discharge 4 days before the flood day to 116.2% of bankfull at the flood peak at the Neversink River near Claryville, NY gauge (Figure 26).

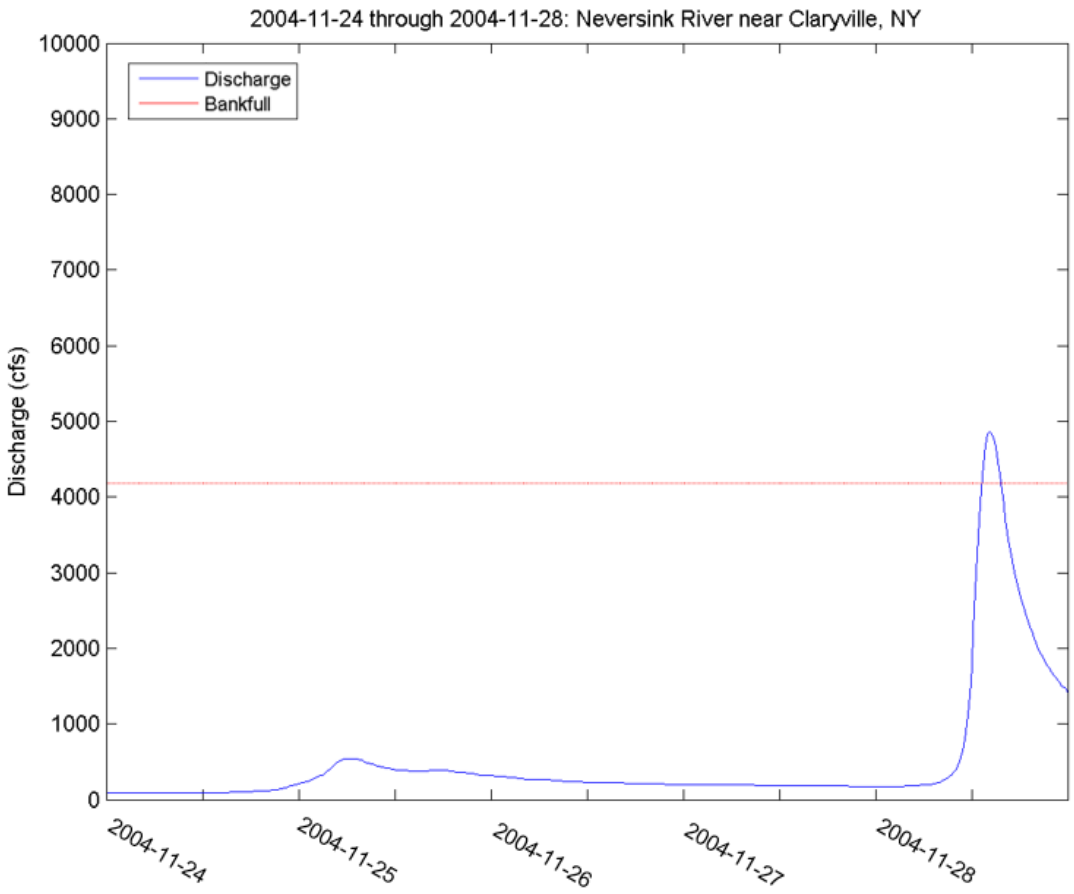

Figure 26. As in Figure 8, but for 2004-11-24 through 2004-11-28

In the Esopus Creek watershed, discharge increased from 2.3% of bankfull four days before the flood day to 106.1% of bankfull at the flood peak (Figure 27).

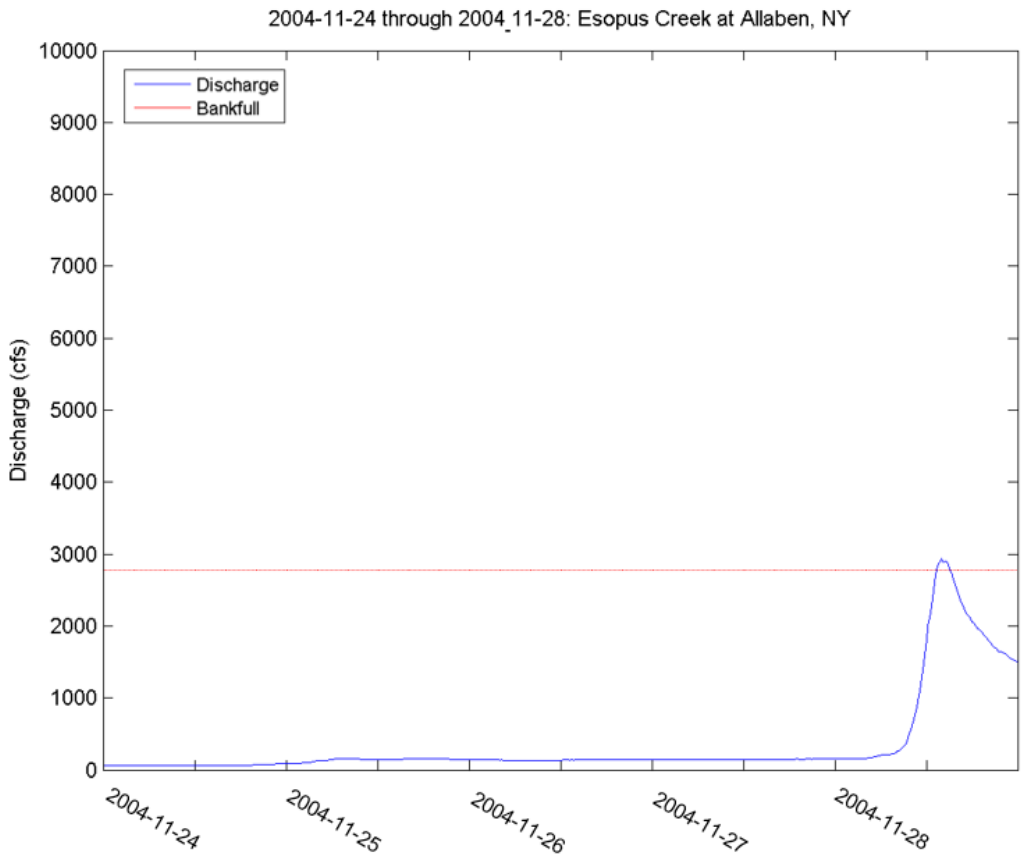

Figure 27. As in Figure 12, but for 2004-11-24 through 2004-11-28

Event 10: 2005-1-14

The flash flooding event on 2005-1-14 occurred after the passage of one cold front, and slightly in the advance of a second cold front (Figure 28). The first cold front is shown to be in the Atlantic just past the coastline. The second cold front is shown moving eastward, located along the Quebec-New Brunswick border and extending into northern New England. The forecast map suggests snow and heavy rain in the study region, with flash flooding possible in central Pennsylvania and the southern tier of New

York and heavy rain over the study area. Snow was also forecasted in the study region, though restricted to the areas typically influenced by lake-effect snow. Fifty-one mm of snow fell on the study day, contributing to the total precipitation of 1.6 in (41.7 mm). The high temperature was 55.0°F (12.8°C), and the low was 28.0°F (-2.2°C) (Table 2-B).

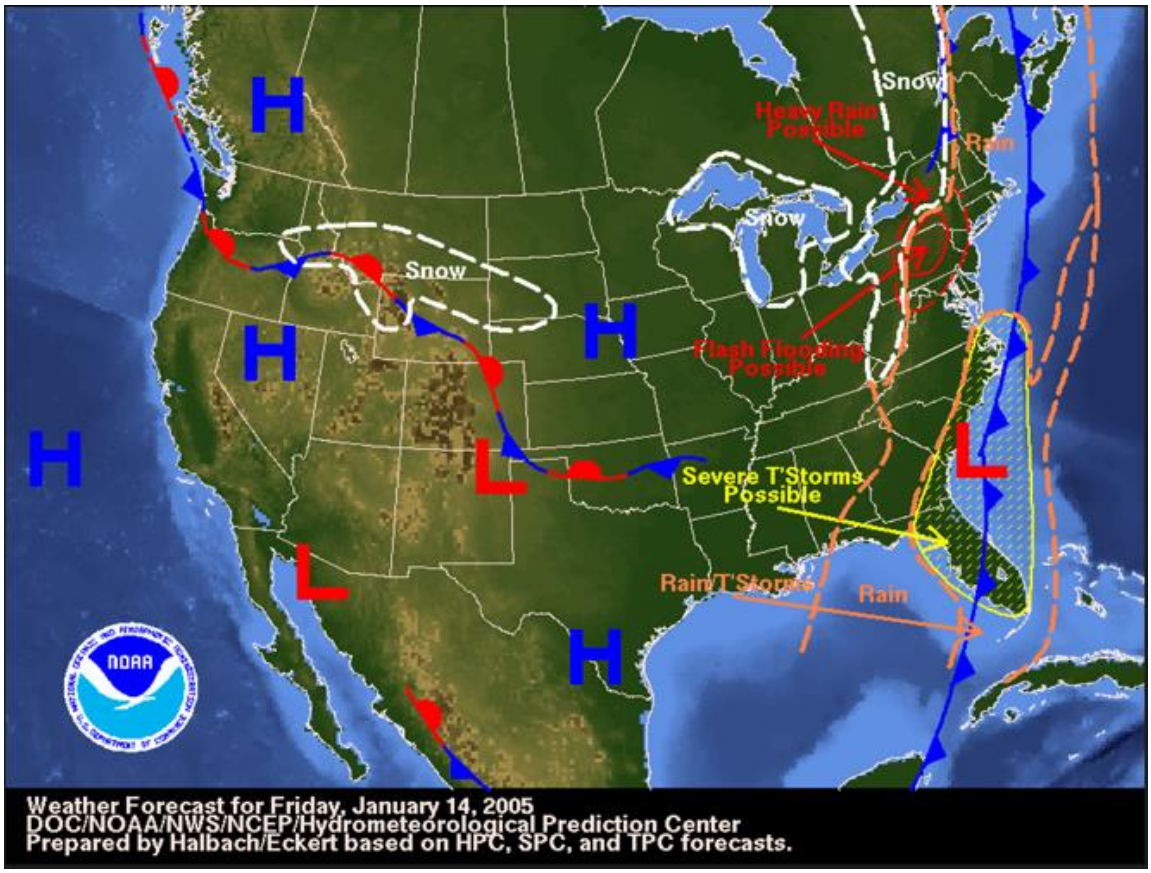

Figure 28. NOAA WPC Weather Forecast map issued for 2005-01-14 (NOAA NWS/ WPC).

Flash flooding was only recorded in the Esopus Creek watershed, with discharge rising to 107.1% of bankfull from 5.8% of bankfull four days earlier (Figure 29, Table 3).

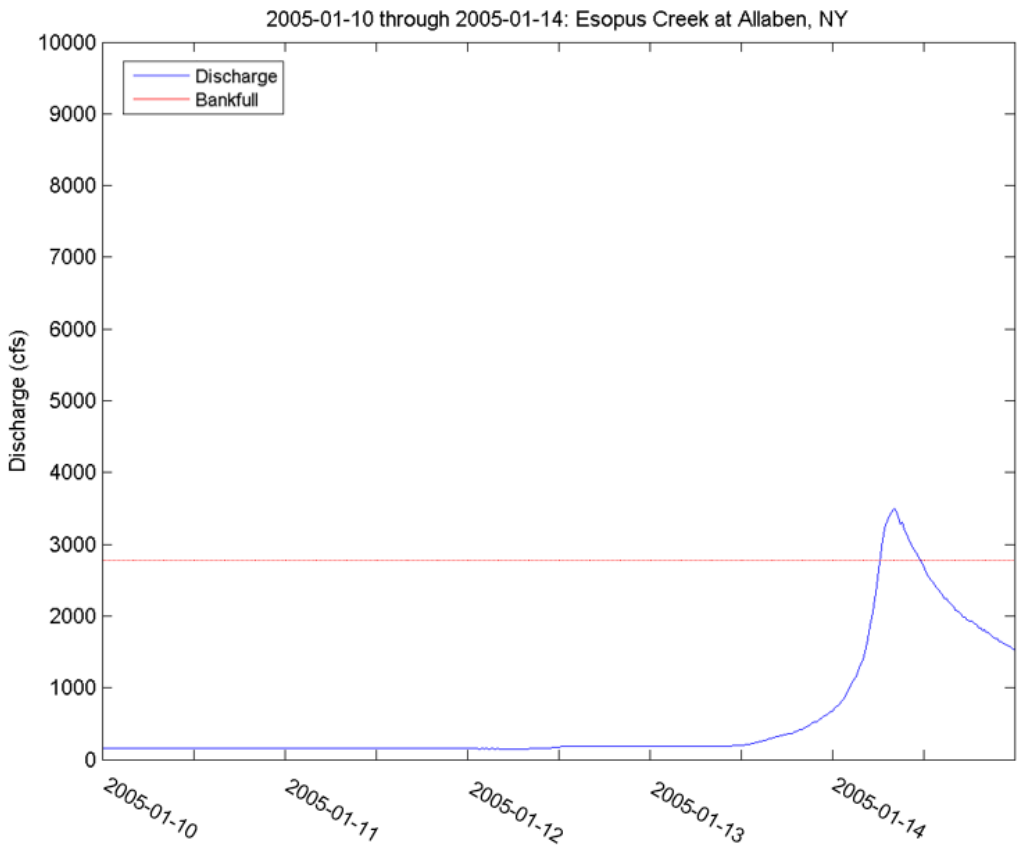

Figure 29. As in Figure 12, but for 2005-01-10 through 2005-01-14

Event 11: 2005-10-8

The weather forecast map for 2005-10-8 shows a stationary front extending from Maine through a low-pressure minima in South Carolina and Georgia (Figure 30). Rain was forecasted directly over the study region on the western side of the front, with flash flooding forecasted along the frontal boundary and in the study watersheds. While flash flooding was only recorded at the Neversink River near Claryville, NY gauge, the precipitation associated with this event was among the highest of all the detected flash flood days with 4.6 in (117.3 mm) (Table 2-A). Temperature fluctuations were slight, ranging only from 61° F (16.1°C) to 64° F (17.8°C) (Table 2-A). The forecast map shows many other low and high pressure systems. Many of these systems appear opposing in the southwestern states. Snow is forecasted in some elevations in the Rocky Mountains of Montana, Idaho, and Wyoming. The Midwestern and central states are relatively free of frontal activity on this day.

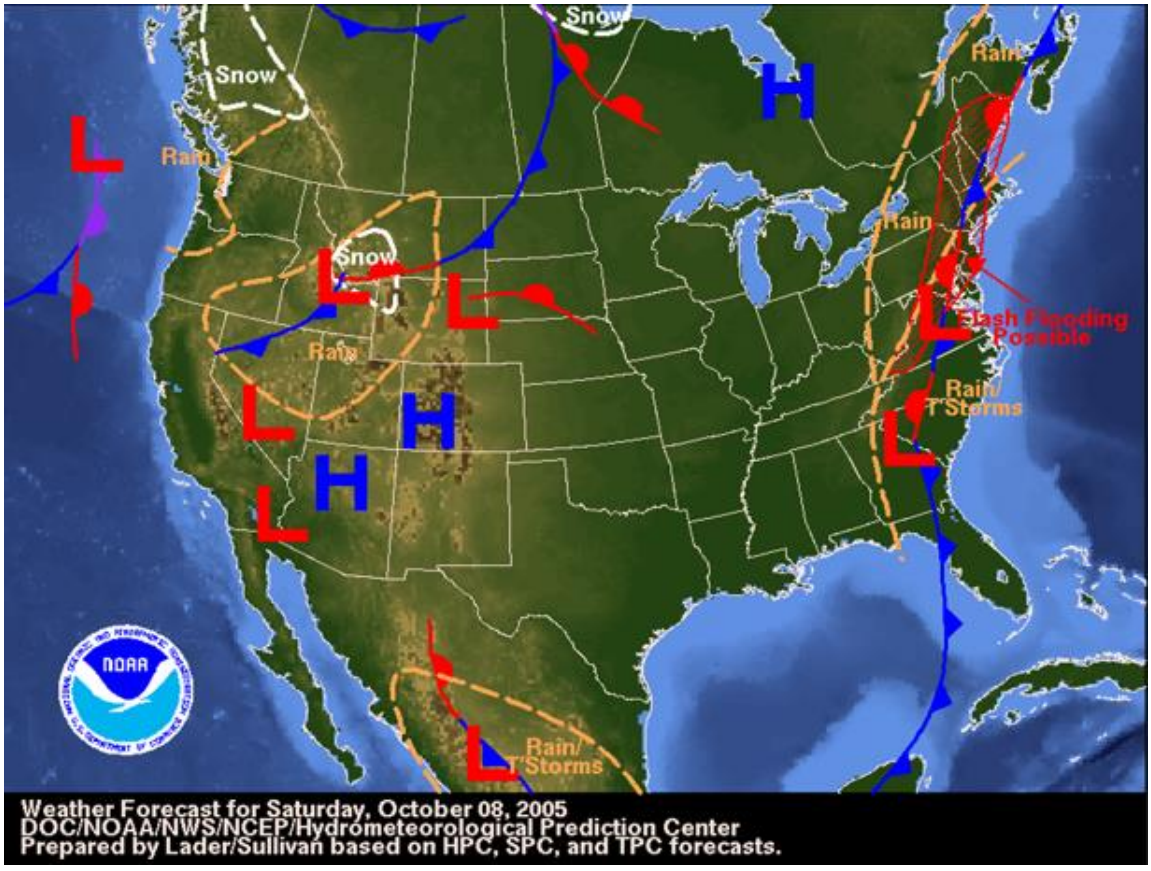

Figure 30. NOAA WPC Weather Forecast map issued for 2005-10-08 (NOAA NWS/ WPC).

Though the precipitation was high, the flash flood peak was only 104.3% above bankfull (Figure 31, Table 2). However, four days prior to the flash flood, the discharge at the gauge was only 0.7% of bankfull discharge, suggesting dry antecedent conditions (Table 2). These conditions could also explain why flash flooding was not observed in the Esopus Creek at Allaben, NY gauge, at which the bankfull discharge is only 66% of the Neversink River gauge discharge.

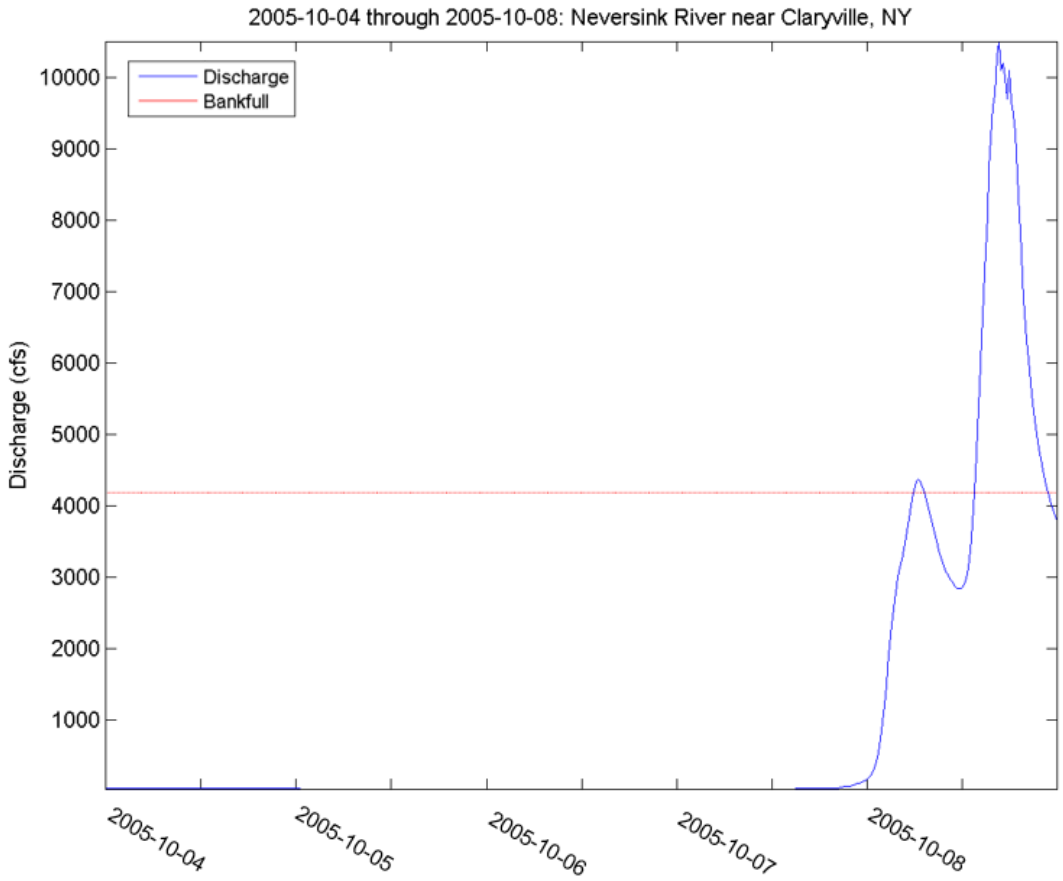

Figure 31. As in Figure 8, but for 2005-10-04 through 2005-10-08

Event 12: 2005-11-30

The flash flooding event on 2005-11-30 occurred after the passage of a cold front in the region a low-pressure area (Figure 32). This cold front extended from Quebec through New England and off of the Atlantic Coast south of Connecticut, moving to the east. The weather forecast map suggests flash flooding in Maine and New Hampshire, and rain and thunderstorms from the study watersheds through the majority of New England. A second cold front approached from Ontario, stretching over the Great Lakes.

Snow was forecasted in the lee of each of the Great Lakes, though not encompassing the study watersheds in southwestern New York. Temperatures ranged from 39° F (3.9°C) to 55.9°F (13.3°C), with 73.7 mm of precipitation (Table 2-A, Table 2-B).

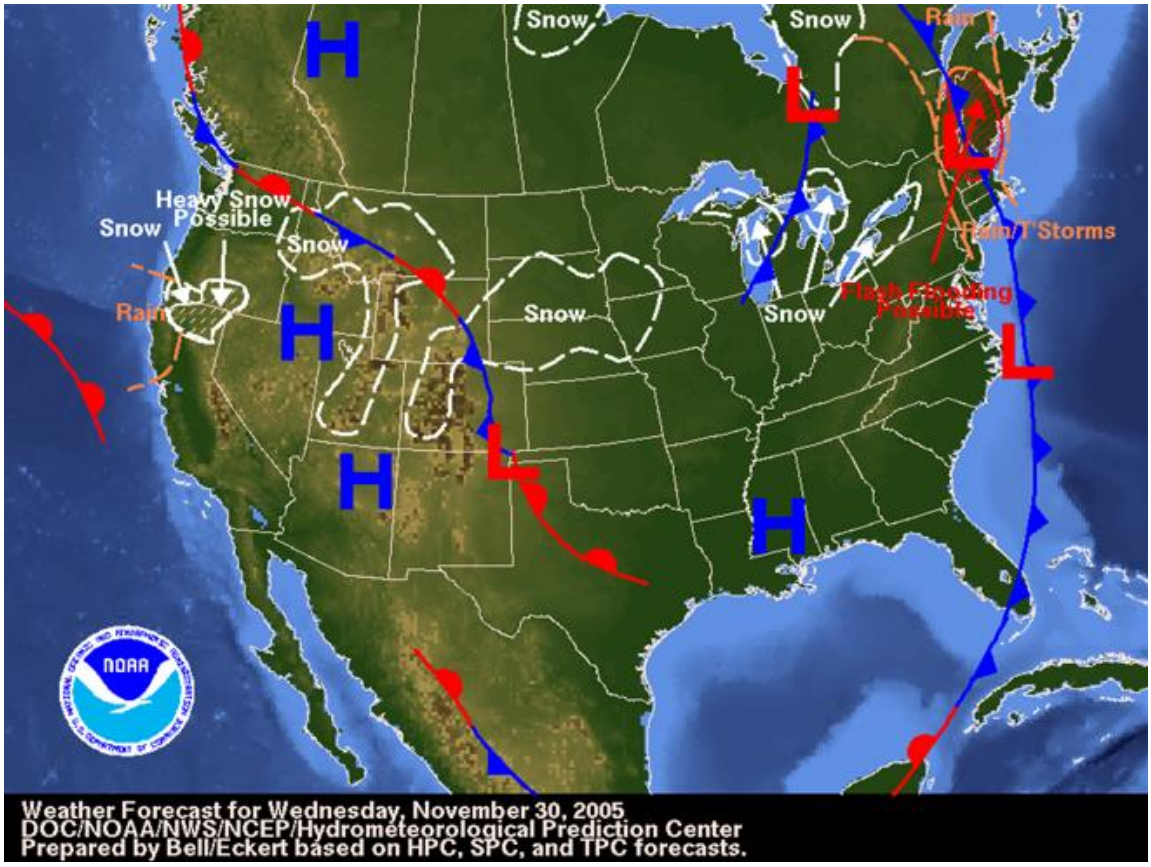

Figure 32. NOAA WPC Weather Forecast map issued for 2005-11-30 (NOAA NWS/ WPC).

Flash floods were detected at the gauges in both watersheds, though more pronounced in the Neversink River watershed. At the Neversink River gauge, discharge

rose from 3.8% of bankfull four days prior to the flood to 150.9% at the flood peak (Figure 33, Table 2-A).

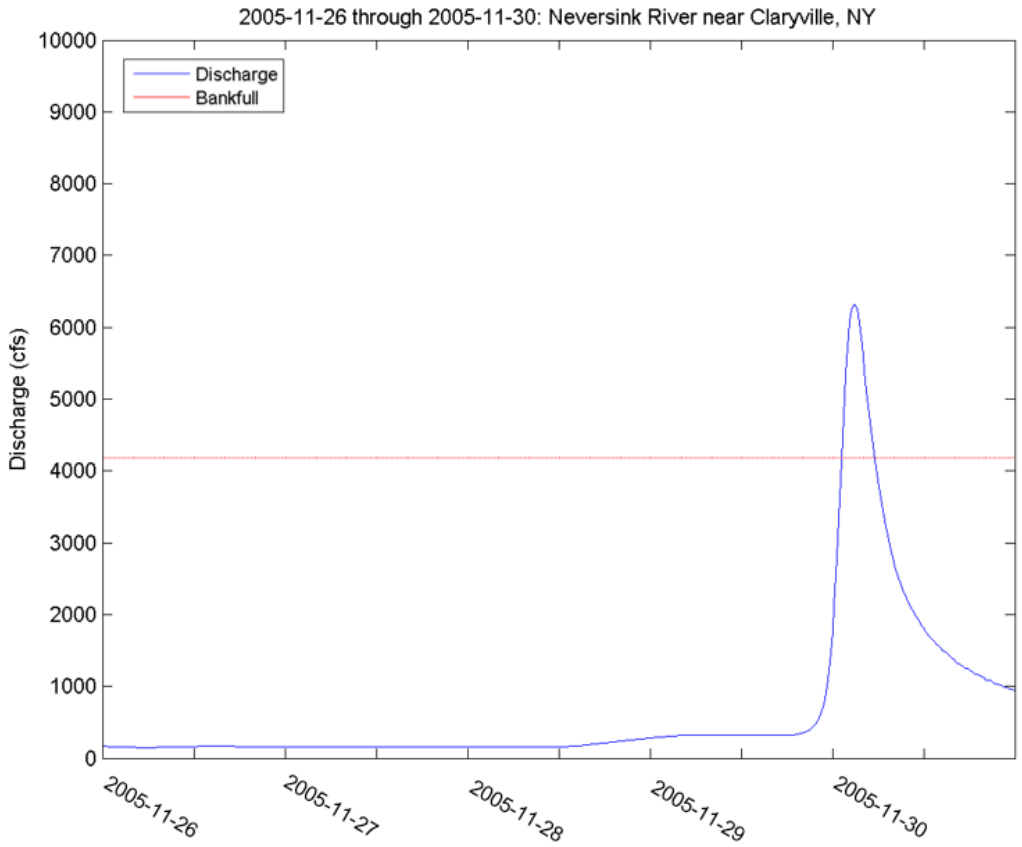

Figure 33. As in Figure 8, but for 2005-11-26 through 2005-11-30

The peak discharge for this event at the Esopus Creek gauge was only 108.9% of bankfull, rising from 5.5% of bankfull four days earlier (Figure 34, Table 2-B).

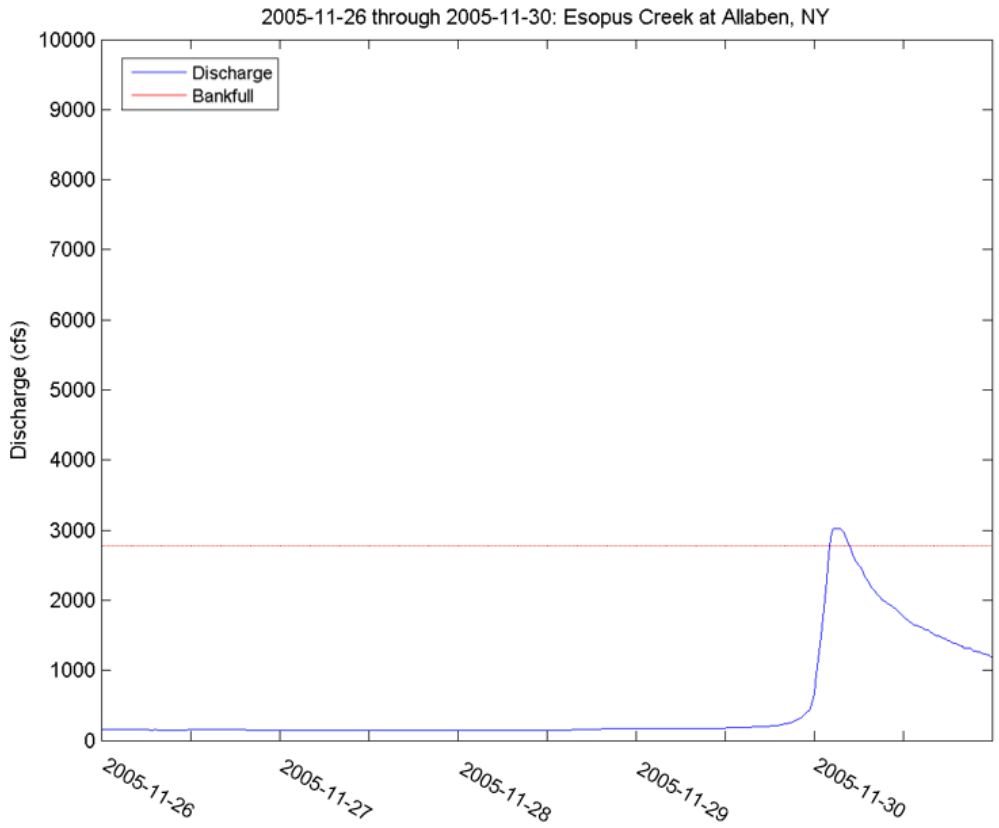

Figure 34. As in Figure 12, but for 2005-11-26 through 2005-11-30

Event 13: 2006-1-18

The weather forecast map for 2006-1-18 shows the passage of a mid-latitude cyclone passing over New England (Figure 35). The cold front of this system was located over Massachusetts, extending into the Atlantic off of Rhode Island. The warm front extended from the southeastern corner of New Hampshire into the Atlantic. The occluded front extended from New Hampshire and Vermont into Quebec. The precipitation ahead of the warm and occluded fronts was forecasted to possibly fall as freezing rain, while the forecast map suggests that the precipitation directly behind the

cold front would fall as rain, with thunderstorms over the study watersheds. The maximum temperature on this day was 53.1°F (11.7°C) and the minimum was 19.0°F (-7.2°C), with trace snow and 1.3 in (33.5 mm) of precipitation (Table 2-A).

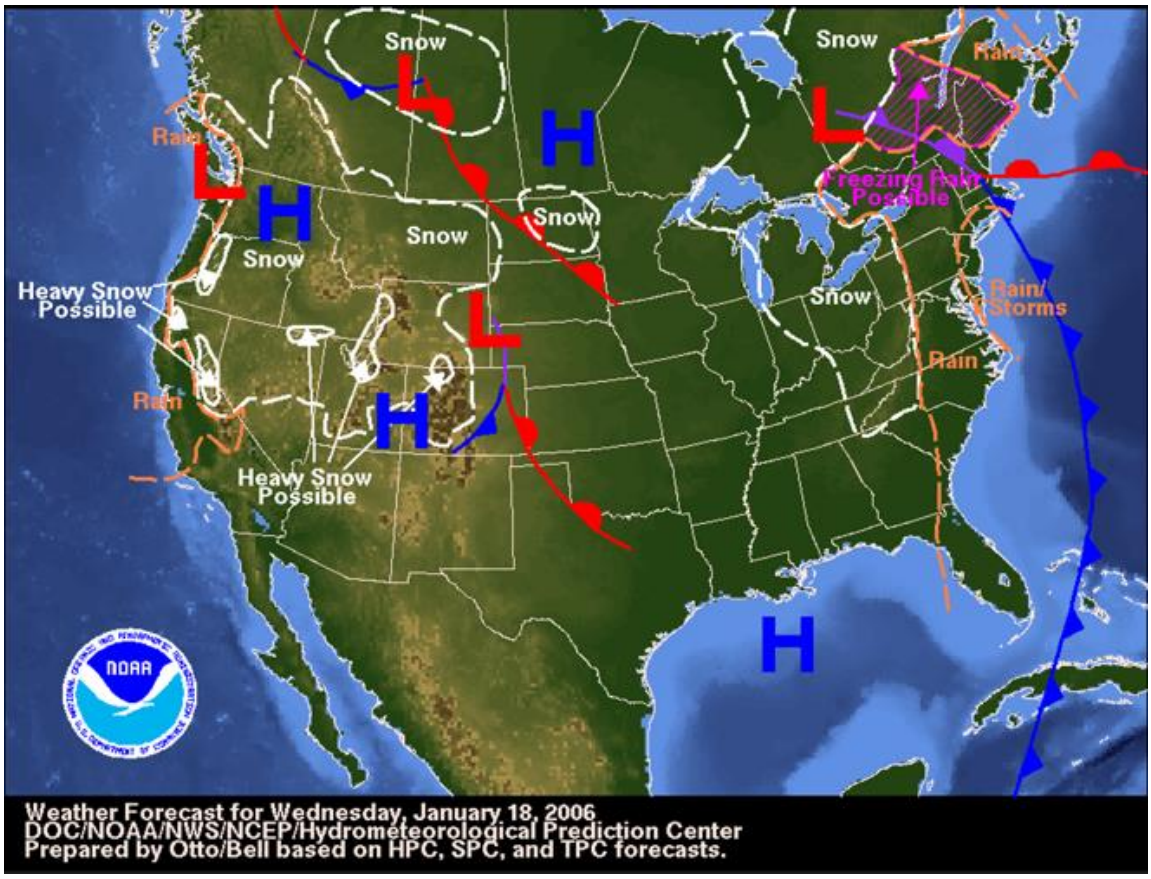

Figure 35. NOAA WPC Weather Forecast map issued for 2006-07-18 (NOAA NWS/ WPC).

The flash flood was detected in the Neversink River watershed only, rising to 157.8% of bankfull discharge from 33.2% bankfull four days earlier (Figure 36, Table 2-A).

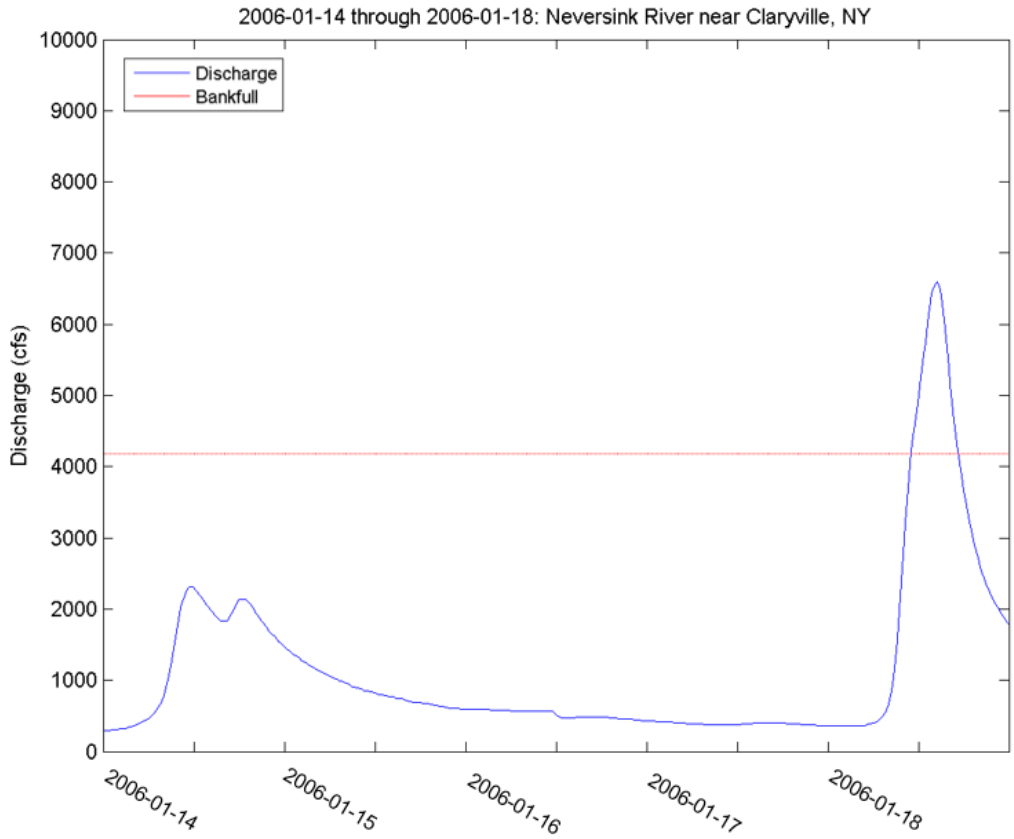

Figure 36. As in Figure 8, but for 2006-01-14 through 2006-01-18

Event 14: 2006-6-28

The flash flood event on 2006-6-28 was forecasted in the Weather Prediction Center forecast map (Figure 37). A stationary front curves southward from a low-pressure center over southern Quebec, along the eastern United States through Florida. While the map suggests the possibility of severe thunderstorms associated with this frontal boundary, the possibility of flash flooding was limited to eastern New York, eastern Pennsylvania, New Jersey, and the New England states excluding Maine. A total

of 2.5 in (64 mm) of precipitation was recorded on this day. The maximum temperature was 69.1°F (20.6°C), and the minimum was 63.0°F (17.2°C) (Table 2-A, Table 2-B).

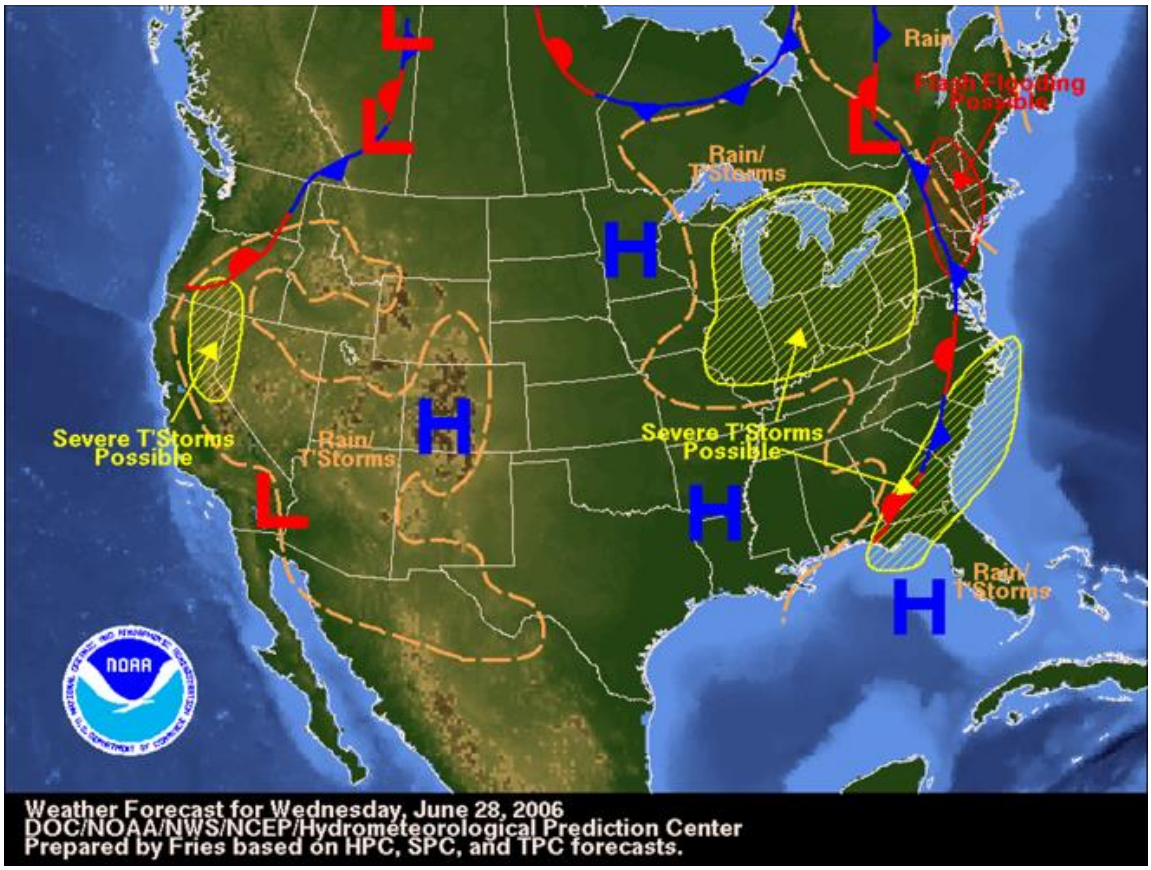

Figure 37. NOAA WPC Weather Forecast map issued for 2006-06-28 (NOAA NWS/ WPC).

Flash flooding was detected in both watersheds, as suggested in the forecast map. Discharge at the Esopus Creek at Allaben, NY gauge increased from 4% of bankfull four days before the flood day, to 159.5% of the bankfull discharge at the flood peak (Figure 38, Table 2-B).

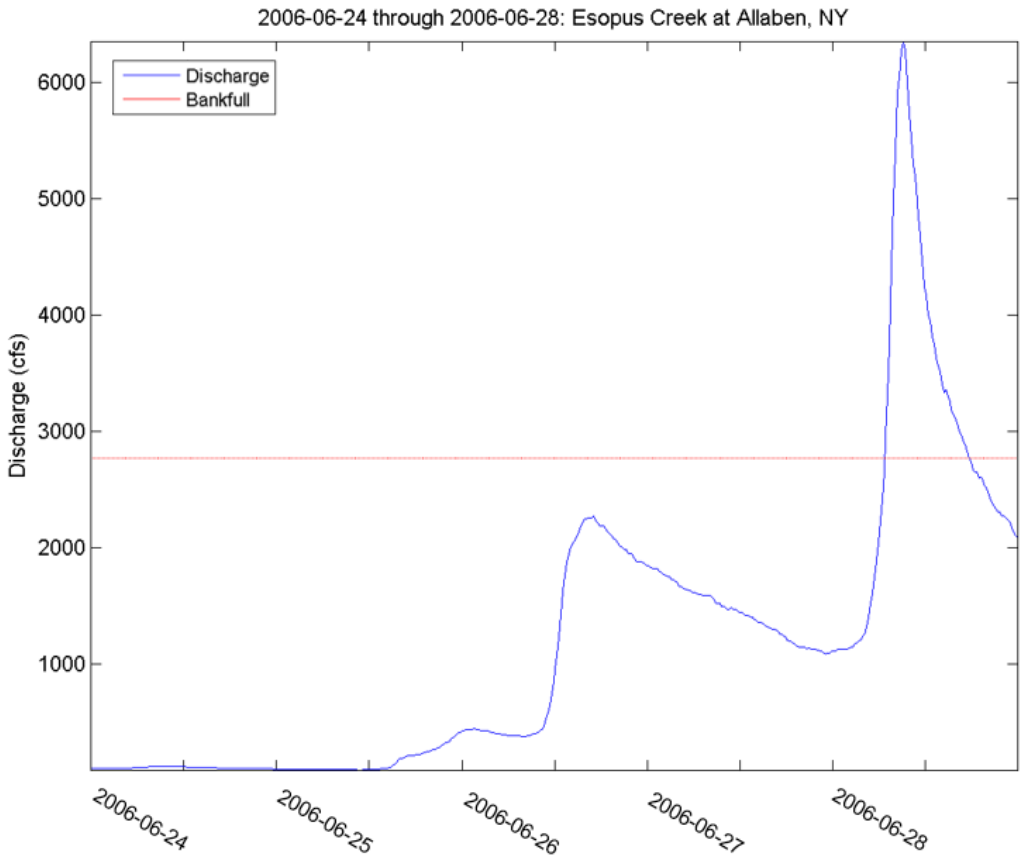

Figure 38. As in Figure 12, but for 2006-06-24 through 2006-06-28

The discharge before the flood day was slightly lower at the Neversink River near Claryville, NY gauge, at only 2.8% of bankfull discharge, but increased to 270.2% of bankfull discharge during the flood peak (Figure 39, Table 2-A).

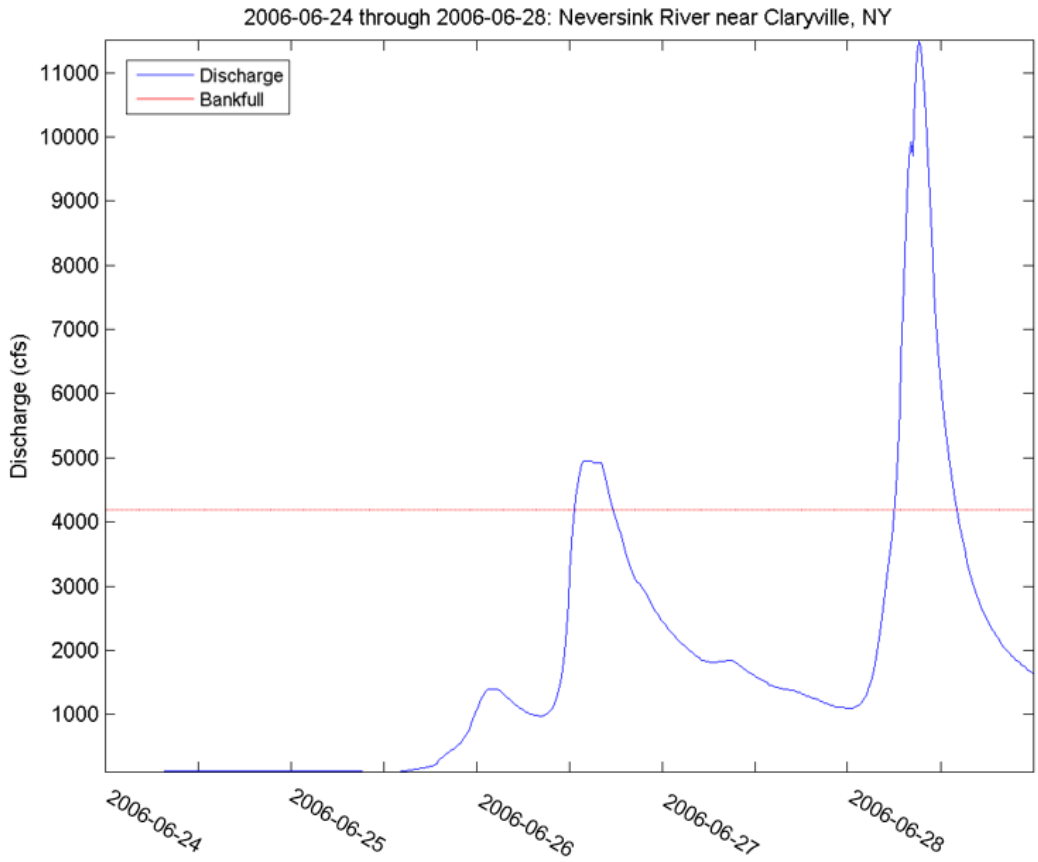

Figure 39. As in Figure 8, but for 2006-06-24 through 2006-06-28

Event 15: 2007-10-27

The weather forecast map for 2007-10-27 shows the passage of a mid-latitude cyclone passing over the northeastern United States, with the low-pressure center over eastern Ontario and the cold, warm, and occluded fronts extending out of Vermont (Figure 40). The cold front extends southward through New England and the Atlantic. The warm front, moving northward, extends through New Hampshire and off the Gulf of

Maine. The occluded front extends through Vermont and Quebec. Rain was forecasted west of the system, behind the cold front and in the study watersheds. In the study watersheds, 1.0 in (25.4 mm) of precipitation fell. Temperatures ranged from 43°F (6.1°C) to 57.9°F (14.4°C) (Table 2-A).

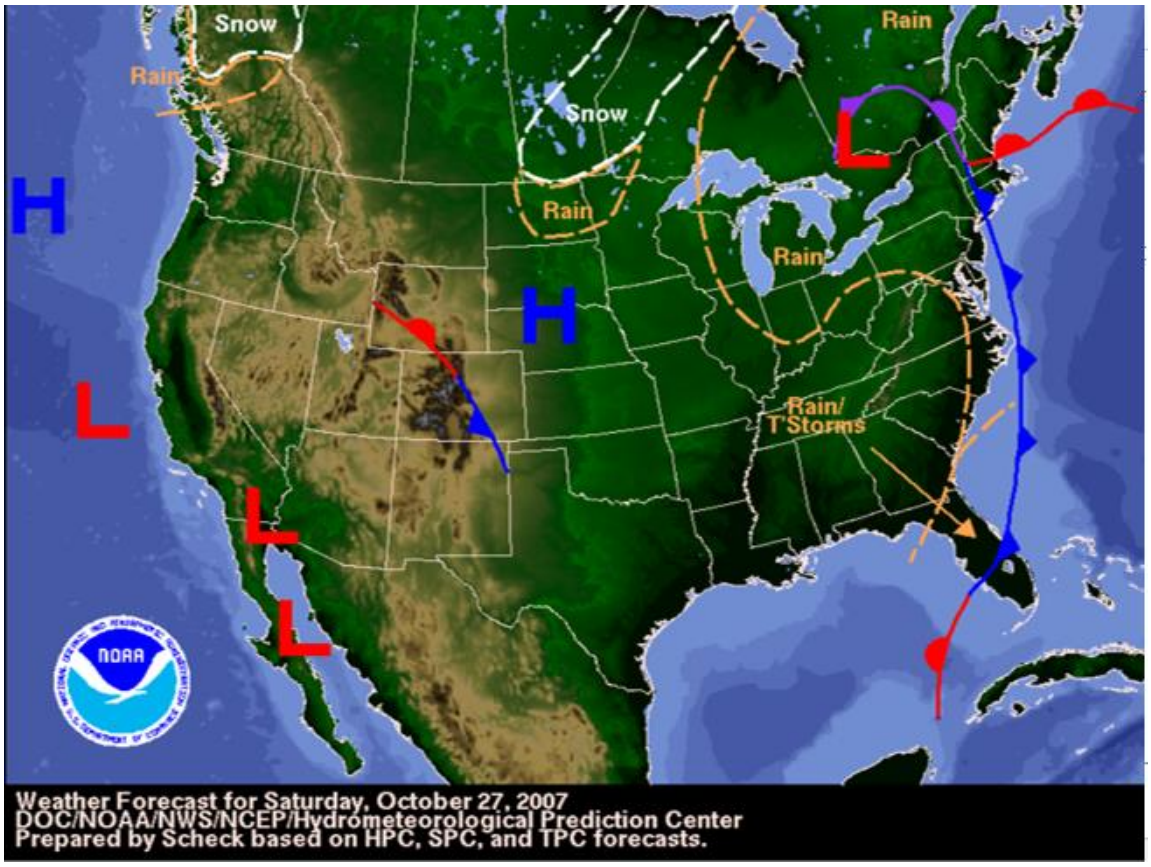

Figure 40. NOAA WPC Weather Forecast map issued for 2007-10-27 (NOAA NWS/ WPC).

The discharge at the Neversink River gauge increased from 2.7% of bankfull discharge to 109.5% of bankfull at the flood peak (Figure 41, Table 2-A). Flash flooding was not detected at the Esopus Creek at Allaben, NY gauge on this day.

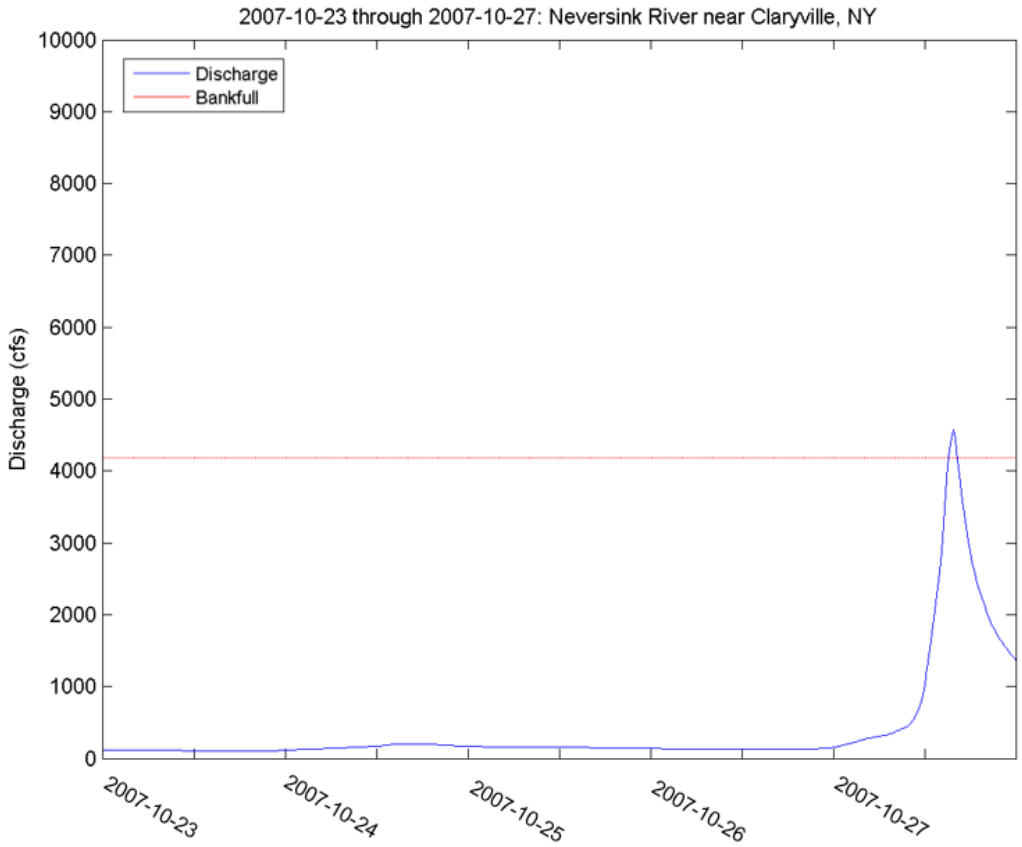

Figure 41. As in Figure 8, but for 2007-10-23 through 2007-10-27

Event 16: 2008-10-25

The weather forecast map for 2008-10-25 was issued at 3:59 EDT, while the flash flood on that day did not occur at the Neversink River near Claryville, NY watershed gauge until 21:30 EDT. The forecast map shows a low-pressure center over

Virginia and North Carolina, with an occluded front in both of the states and the cold front approaching the warm front off the coast of South Carolina. Rain and thunderstorms were widely forecasted over the northeastern quadrant of the United States and the majority of southern Canada (Figure 42). Only 0.1 in (1.8 mm) of precipitation was recorded on this day at this station on the day of the flood peak. However, 3.0 in (75.4 mm) was recorded on the next day at this station (not shown). This shows that this precipitation event may not have tracked over this gauge at the same time as other parts of the catchment. This inconsistency also demonstrates the need for more frequent temporal and spatially representative precipitation data.

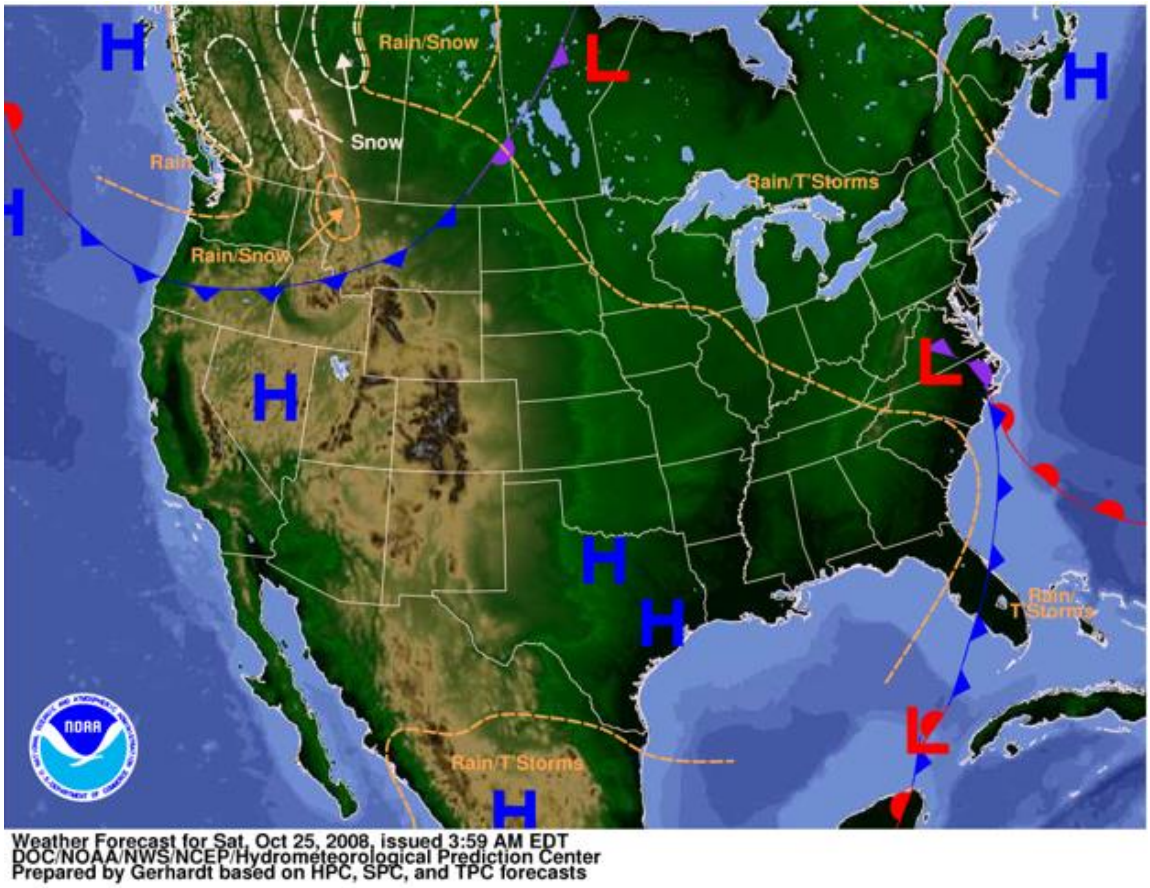

Figure 42. NOAA WPC Weather Forecast map issued for 2008-10-25 (NOAA NWS/ WPC).

The discharge at the Neversink River near Claryville, NY gauge rose from 1.3% of bankfull discharge 4 days before the flood peak, to 106.2% of bankfull at the flood peak (Figure 43, Table 2-A).

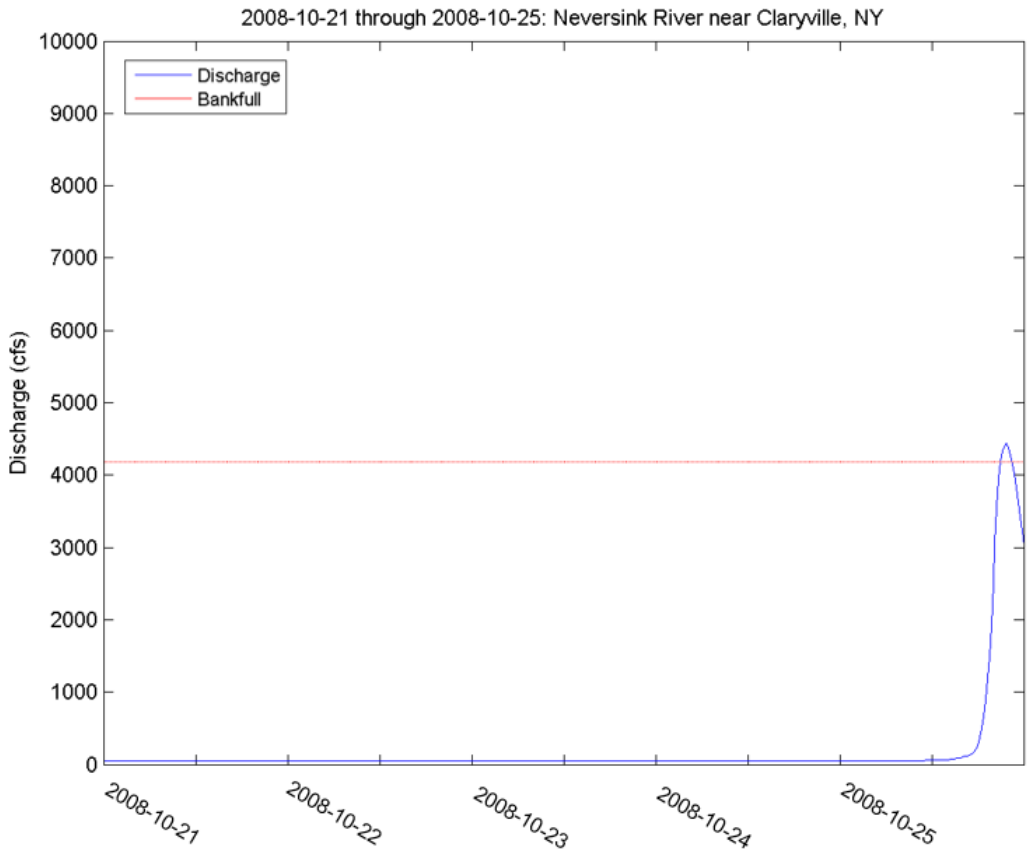

Figure 43. As in Figure 8, but for 2008-10-21 through 2008-10-25

Event 17: 2008-12-12

The flash flood event of 2008-12-12 was forecasted in the Weather Prediction Center forecast map for the day (Figure 44). The map shows a mid-latitude cyclone exiting eastern Maine, with the cold front extending southward off the coast. The weather forecast map suggests rain turning to snow in the study watersheds, with flash flooding possible south of the study area in Long Island. The map shows freezing rain

possible northeast of the study watersheds in New Hampshire and Maine. A second cold front approached the study area, though it is indicated as a weakening cold front in the stages of frontolysis, or the disintegration of the front. On the synoptic scale, the map indicates many other features of interest. Heavy snow is indicated as possible in the high elevations in the Rocky Mountains as far south as Arizona, and also on the west coast through the Canadian Rockies. Snow was also forecasted over the Great Lakes area, over the study region, and southward along the high elevations of the Appalachian and Smokey Mountains. High pressure was indicated over the middle latitudes of the United States, both over the high elevations of the Rocky Mountains and into the Great Plains. The maximum temperature in the watershed was 36.0° F (2.2°C), and the minimum temperature was 28.0°F (-2.2°C) (Table 2-A, Table 2-B). Along with trace snowfall, 2.3 in (59.4 mm) of precipitation was recorded (Table 2-A, Table 2-B).

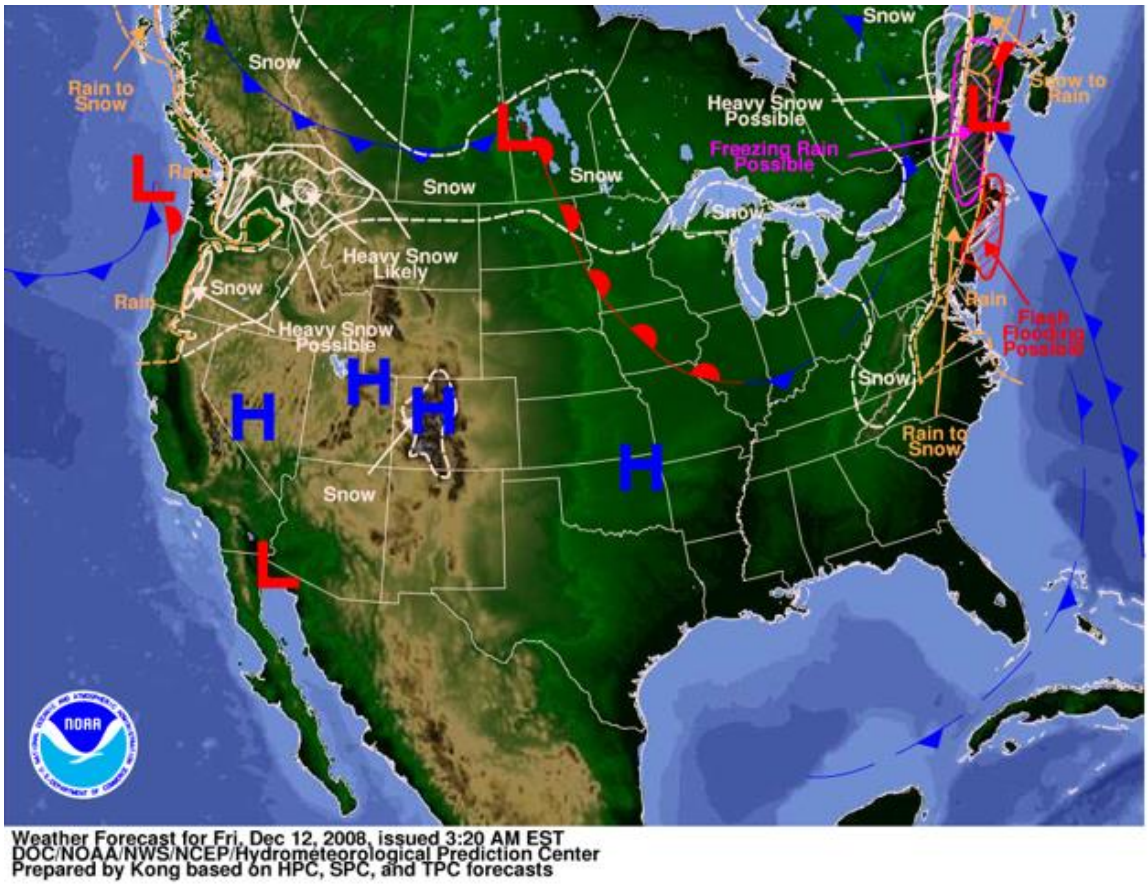

Figure 44. NOAA WPC Weather Forecast map issued for 2008-12-12 (NOAA NWS/ WPC).

Flash floods were detected in both study watersheds. In the Neversink River watershed, discharge increased from 3.2% of bankfull 4 days before the flood peak, to 151.6% of bankfull discharge during the flood peak (Figure 45, Table 2-A).

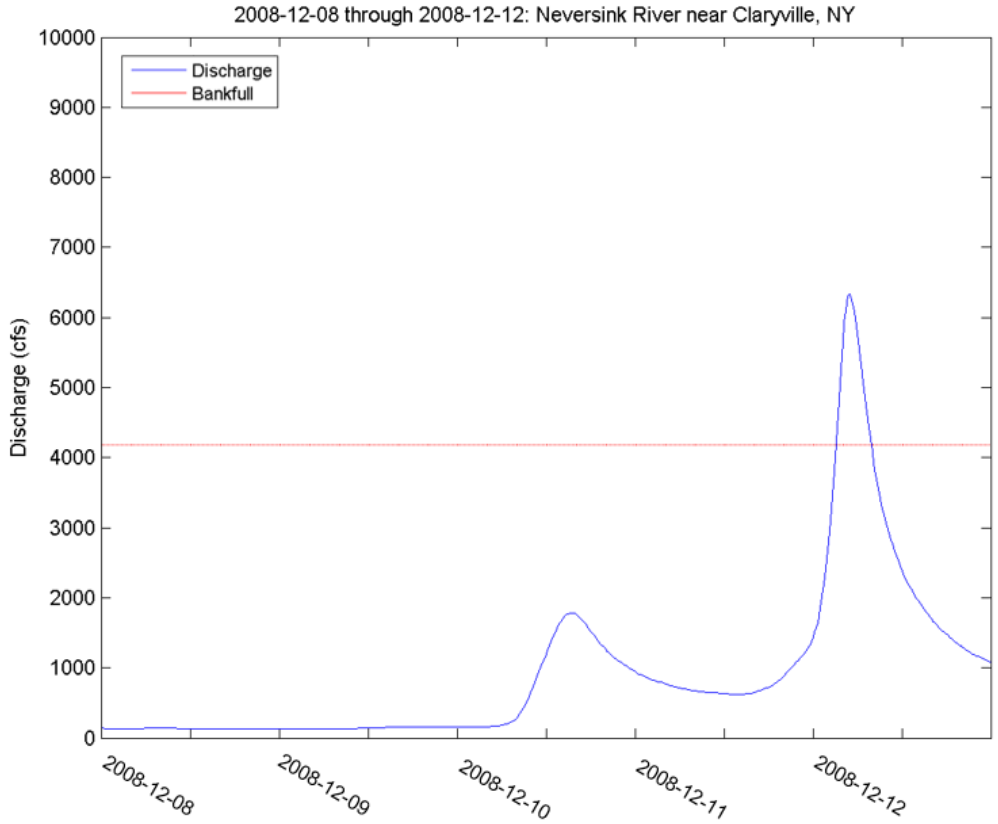

Figure 45. As in Figure 8, but for 2008-12-08 through 2008-12-12

The increased discharge was less pronounced in the Esopus Creek watershed, where discharge was 4.1% of bankfull discharge 4 days preceding the flood peak and increased to 107.1% of bankfull at the flood peak (Figure 46, Table 2-B).

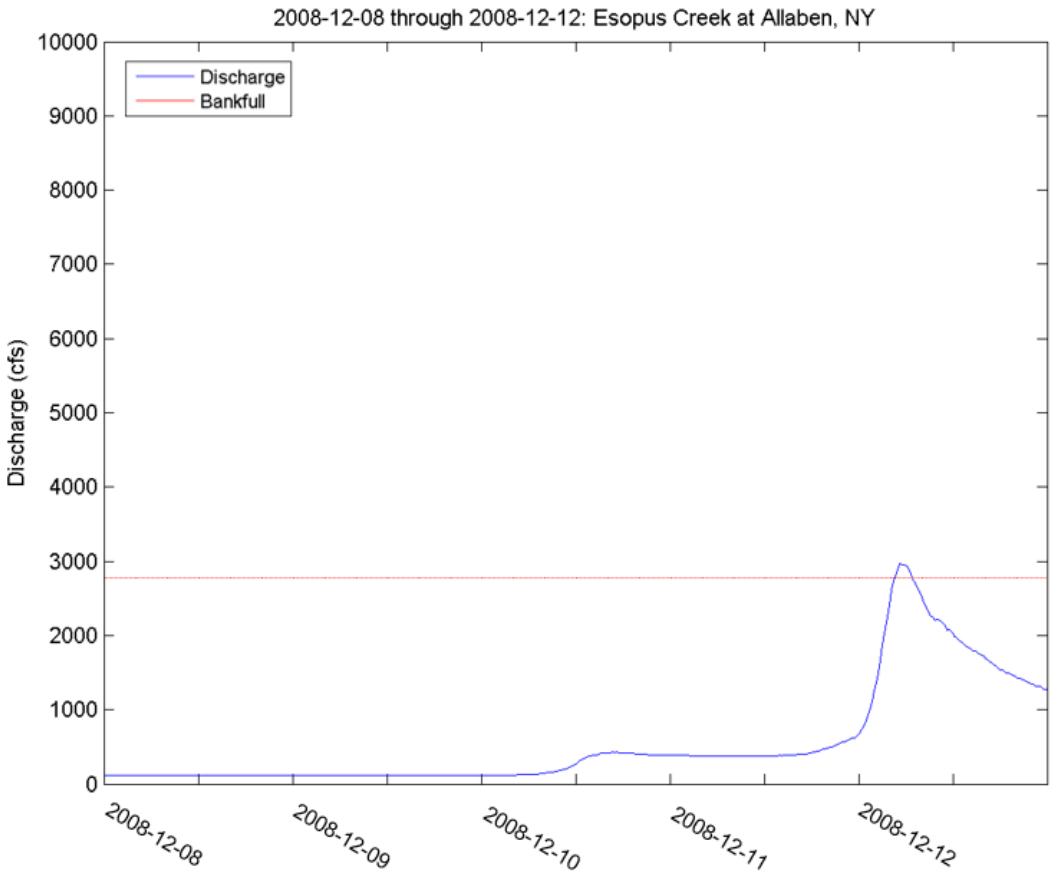

Figure 46. As in Figure 12, but for 2008-12-08 through 2008-12-12

Event 18: 2010-1-25

The weather forecast map for 2010-1-25 shows a low-pressure system over the Quebec-Ontario border, with the occluded, warm, and cold fronts extending out of the area directly over the study watersheds (Figure 47). A second cold front approached the study watersheds and the mid-Atlantic coastline. Rain and thunderstorms in the study region were forecasted, as well as widespread possible flash flooding from North

Carolina through Maine. Snow was forecasted behind the second cold front, though precipitation was forecasted over the study area. The 1.7 in (43.4 mm) of precipitation fell on a depth of 11.0 in (279 mm) accumulated snow, though none of the precipitation was recorded as snow (Table 2-B). The maximum temperature on this day was 48.9°F (9.4°C), and the minimum temperature was 24.1°F (-4°C) (Table 2-B).

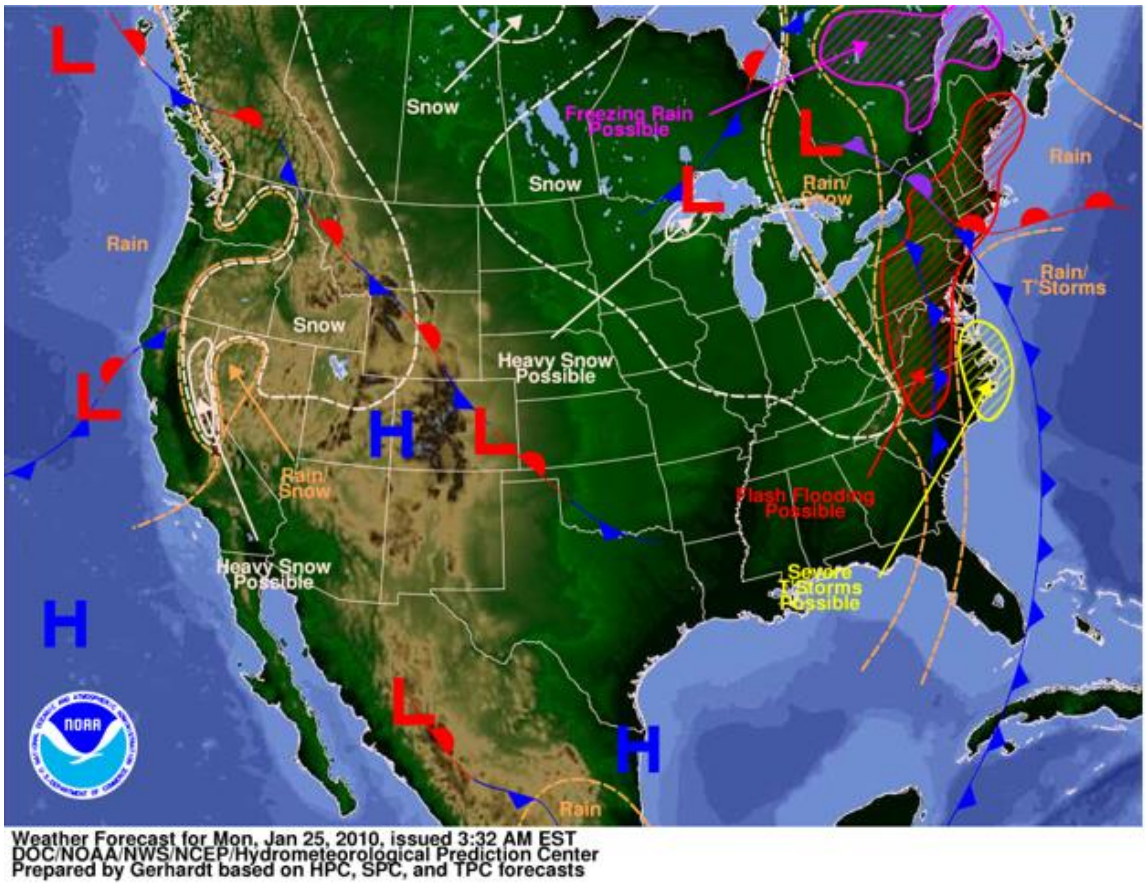

Figure 47. NOAA WPC Weather Forecast map issued for 2010-01-25 (NOAA NWS/ WPC).

Flash flooding was detected only at the Esopus Creek at Allaben, NY gauge, with a peak flood discharge of 181.1% of bankfull discharge rising from just 2% of bankfull discharge four days earlier (Figure 48, Table 2-B).

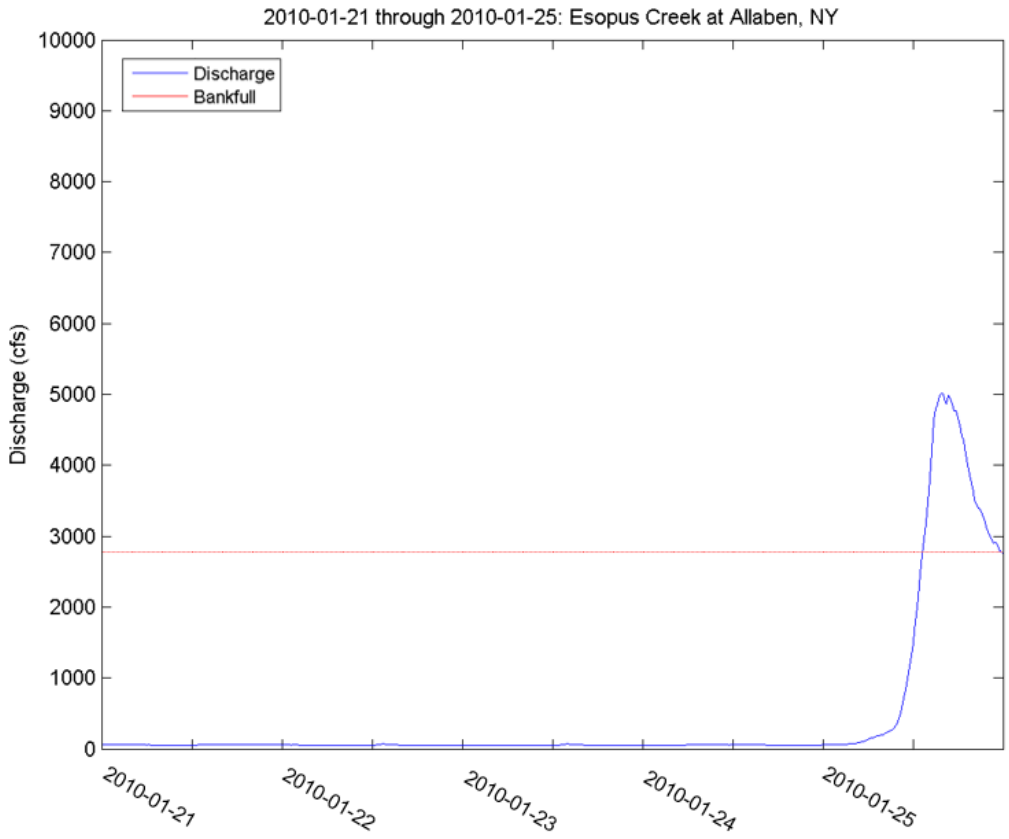

Figure 48. As in Figure 12, but for 2010-01-21 through 2010-01-25

Event 19: 2010-10-1

The flash flood event on 2010-10-1 was forecasted in the weather forecast map. The map shows a low-pressure system over the northeastern United States, with cold fronts extending northward and southward of the pressure minima (Figure 49). Flash flooding was forecasted from Pennsylvania to Maine behind the northern cold front, with rain forecasted more generally from the northeastern United States down through North Carolina. The forecast map also shows a small cold front with rain forecasted over the Great Lakes region. This is in contrast with the rest of the map view, which shows the United States, Mexico, and Southern Canada to generally be experiencing high pressure systems. In the study area, the maximum temperature was 68.0°F (10.6°C), and the minimum was 51.1°F (10.6° C) (Table 2-B). This was one of the largest precipitation events for the detected flash floods, as 1.4 in (35.1 mm) of precipitation was recorded falling on this day (Table 2-B).

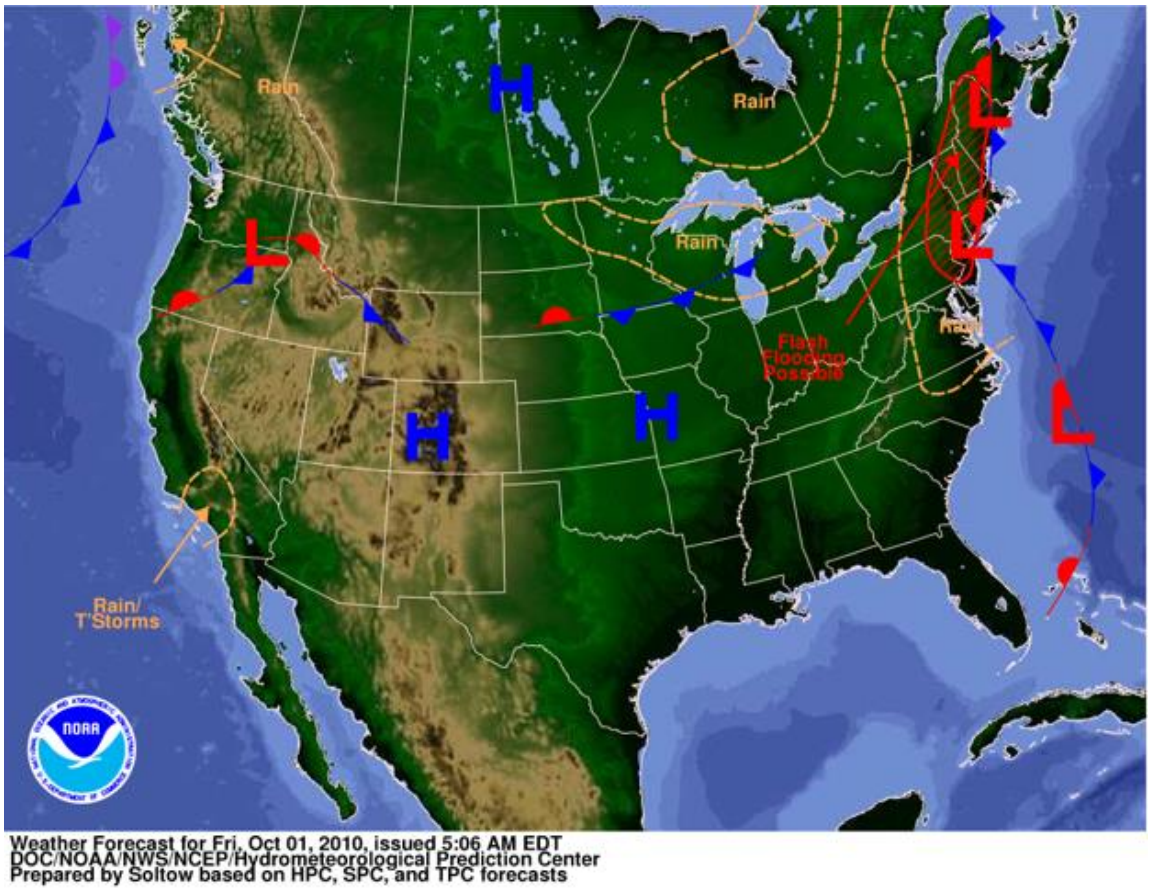

Figure 49. NOAA WPC Weather Forecast map issued for 2010-10-01 (NOAA NWS/ WPC).

This flood had the second highest flash flood peak at the Esopus Creek gauge, at 364.4% of bankfull discharge, rising from just 0.4% of bankfull discharge four days earlier (Figure 50, Table 3). Despite the high precipitation, no flash flooding was recorded in the Neversink River watershed.

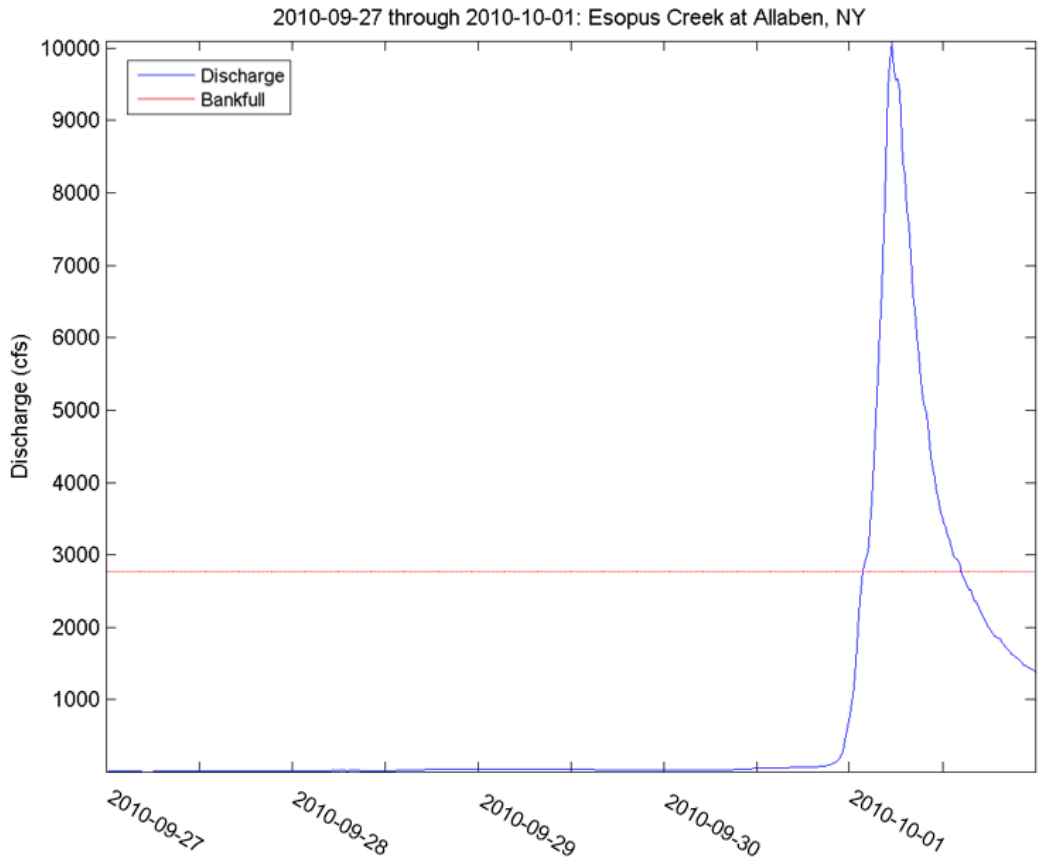

Figure 50. As in Figure 12, but for 2010-09-27 through 2010-10-01

Event 20: 2010-12-1

The flash flooding event on 2010-12-1 was also forecasted. The map shows a mid-latitude cyclone passing over the northeastern United States, centered over Pennsylvania and New Jersey (Figure 51). The occluded front extended to the northwest through Quebec. The warm front extended into the Atlantic off of New Jersey. Two cold fronts accompanied this system. The first cold front approached the warm front off the coast. The second cold front extended from New York, through the study watersheds,

curving through Maryland and ending in South Carolina. The map suggests severe thunderstorms possibly between the two cold fronts, and flash flooding possible in New York and Pennsylvania, including the study watersheds. The map indicated rain throughout the majority of the eastern United States and Canada associated with the mid-latitude cyclone, and snow in the typical lake-effect snow areas east of the Great Lakes. Temperatures ranged from 37.0° F (2.8°C) to 57.0°F (10.6°C), with 1.4 in (35.1 mm) of precipitation (Table 2-B).

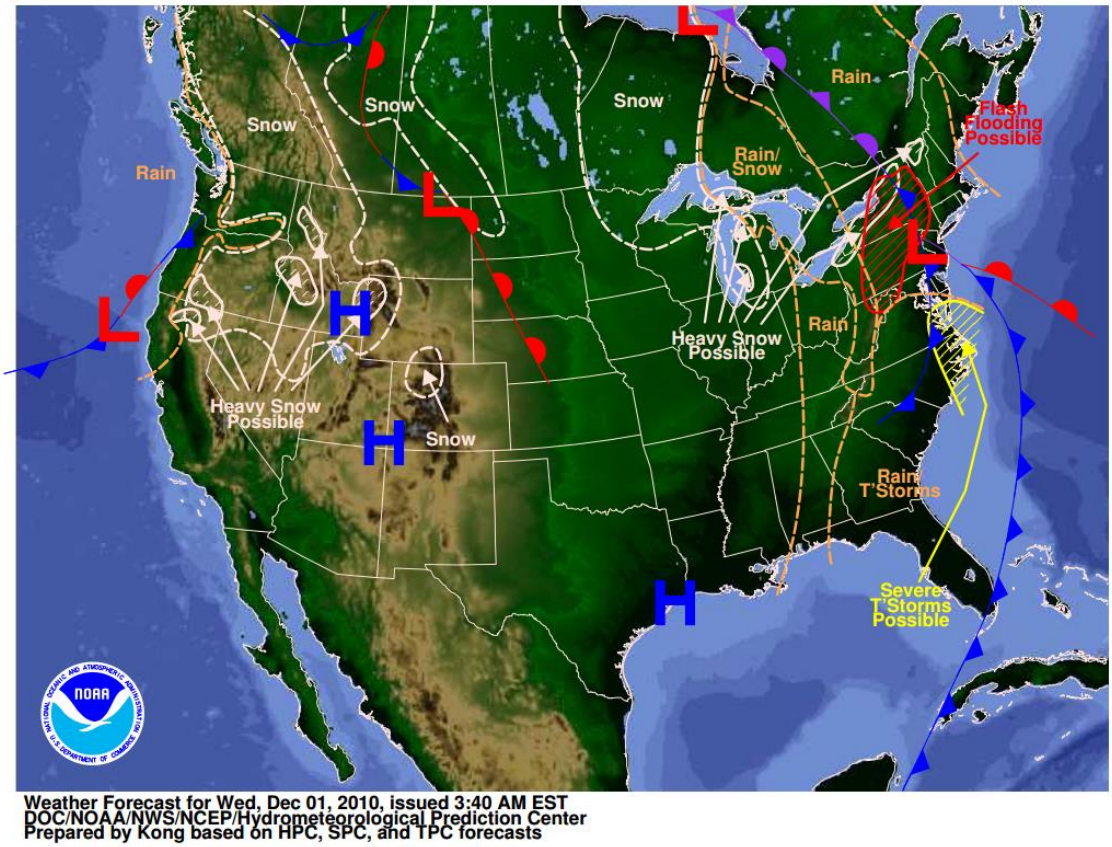

Figure 51. NOAA WPC Weather Forecast map issued for 2010-12-01 (NOAA NWS/ WPC).

The discharge at the Esopus Creek at Allaben, NY gauge increased from 5.9% of bankfull four days before the flash flood peak, to 147.5% of bankfull during the peak discharge (Figure 52, Table 2-B). No flash flooding was detected in the Neversink River watershed.

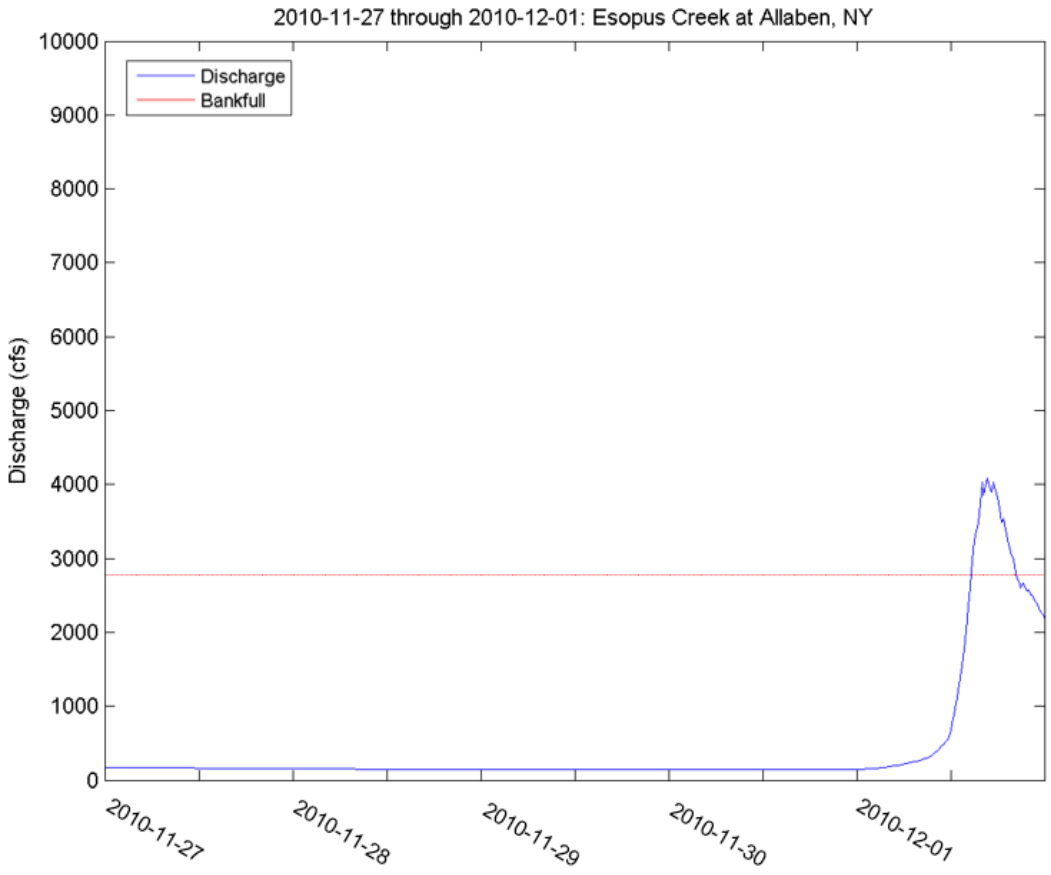

Figure 52. As in Figure 12, but for 2010-11-27 through 2010-12-01

Event 21: 2011-8-28

The weather forecast map for 2011-8-28, issued at 4:11 EDT shows Hurricane Irene approaching New Jersey (Figure 53). A cold front approached the hurricane from the north, suggesting that the fronts would interact directly over the study watersheds. The map indicated rain and flash flooding in much of the Northeast and New England, and severe thunderstorms north of the hurricane. The map also shows the hurricane moving inland and opposing an approaching cold front in the Great Lakes area. This front is the second front shown on the map, with one cold front trailing the other. Overall, precipitation is forecasted over the entire United States, with the exception of the Great Lakes region under a high pressure system. This was the largest precipitation event associated with flash flooding in this study, with 8.6 in (218.9 mm) of precipitation falling on the day of the largest flash floods in both watersheds (Table 2-A, Table 2-B).

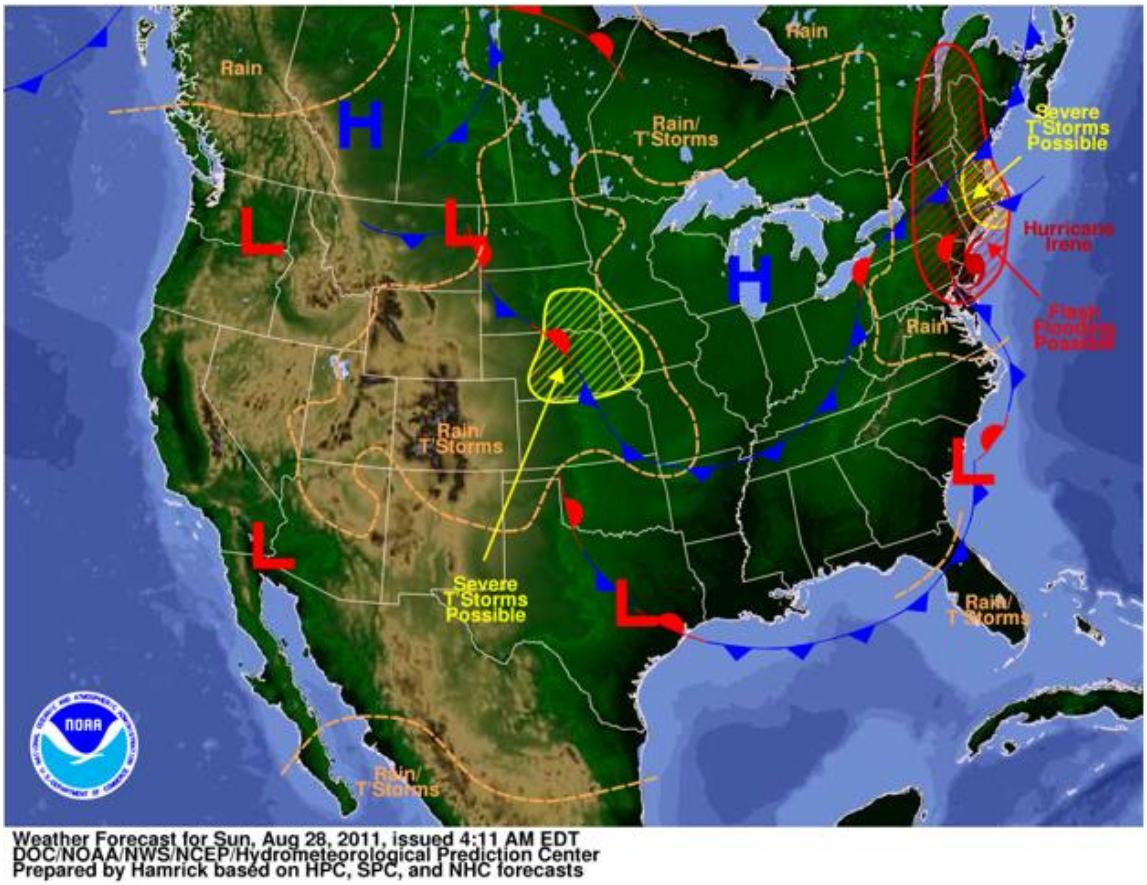

Figure 53. NOAA WPC Weather Forecast map issued for 2011-08-28 (NOAA NWS/ WPC).

Discharge in the Esopus Creek watershed increased from 4.4% of bankfull four days before the flood to 1057% of the bankfull during the flood peak (Figure 54, Table 2-B).

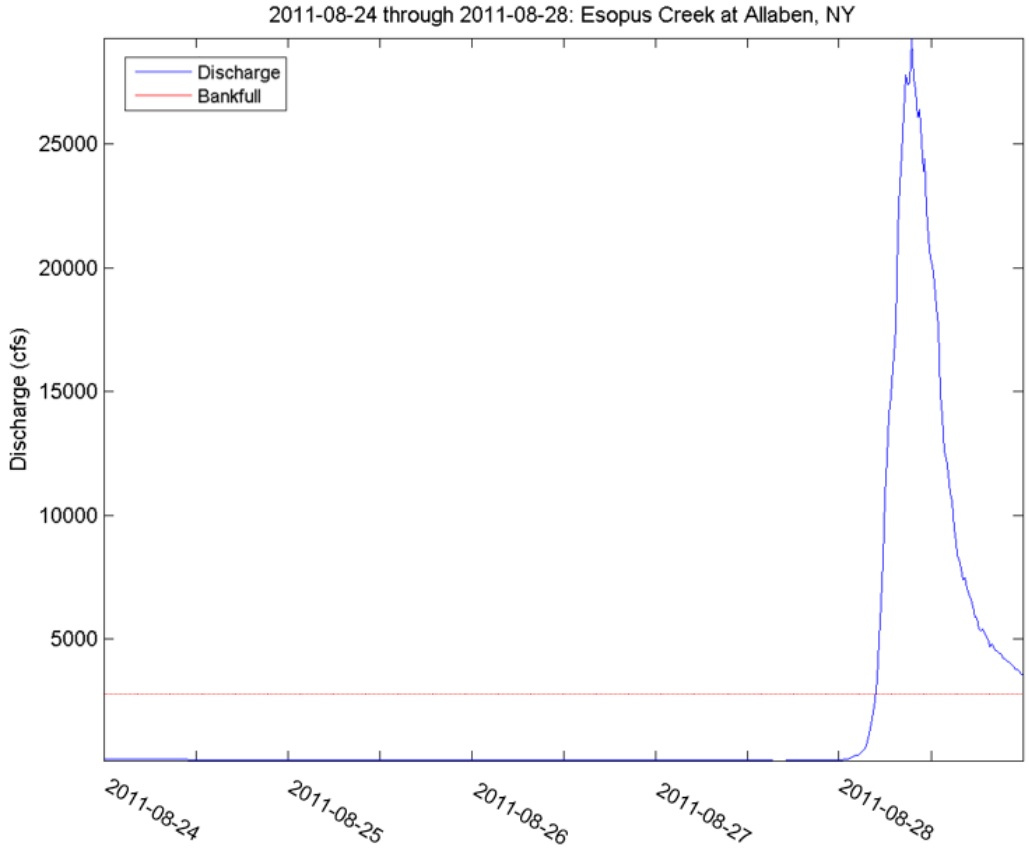

Figure 54. As in Figure 12, but for 2011-08-24 through 2011-08-28

Discharge in the Neversink River watershed rose increased from 3.6% of bankfull four days before the flood to 499.8% of the bankfull discharge during the flood peak (Figure 55, Table 2-A).

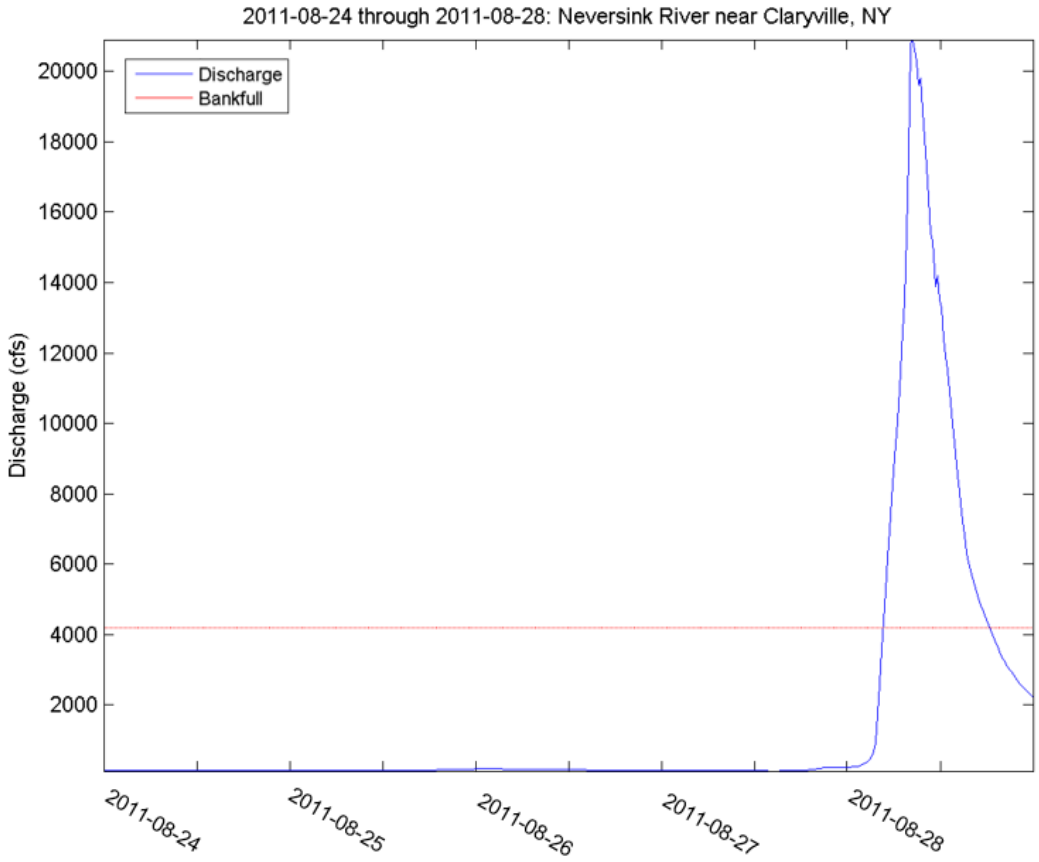

Figure 55. As in Figure 8, but for 2011-08-24 through 2011-08-28

Event 22: 2011-9-7

The flash flooding event on 2011-9-7 was also forecasted in the weather forecast map. The map shows a mid-latitude cyclone in the eastern United States, with the warm front extending from Maryland through New Jersey (Figure 55). Rain and possible flash flooding were forecasted ahead of the warm front in New York, Pennsylvania, and New Jersey, with severe thunderstorms over the Chesapeake Bay where the distance between

the warm and cold fronts is diminishing into the low pressure center. The maximum temperature on this day in the study area was 57.0°F (13.9°C), and the minimum was 52.0°F (11.1°C) (Table 2-A, Table 2-B). A total of 2.8 in (71.9 mm) of precipitation fell, associated with flash flooding in both watersheds (Table 2-A, Table 2-B).

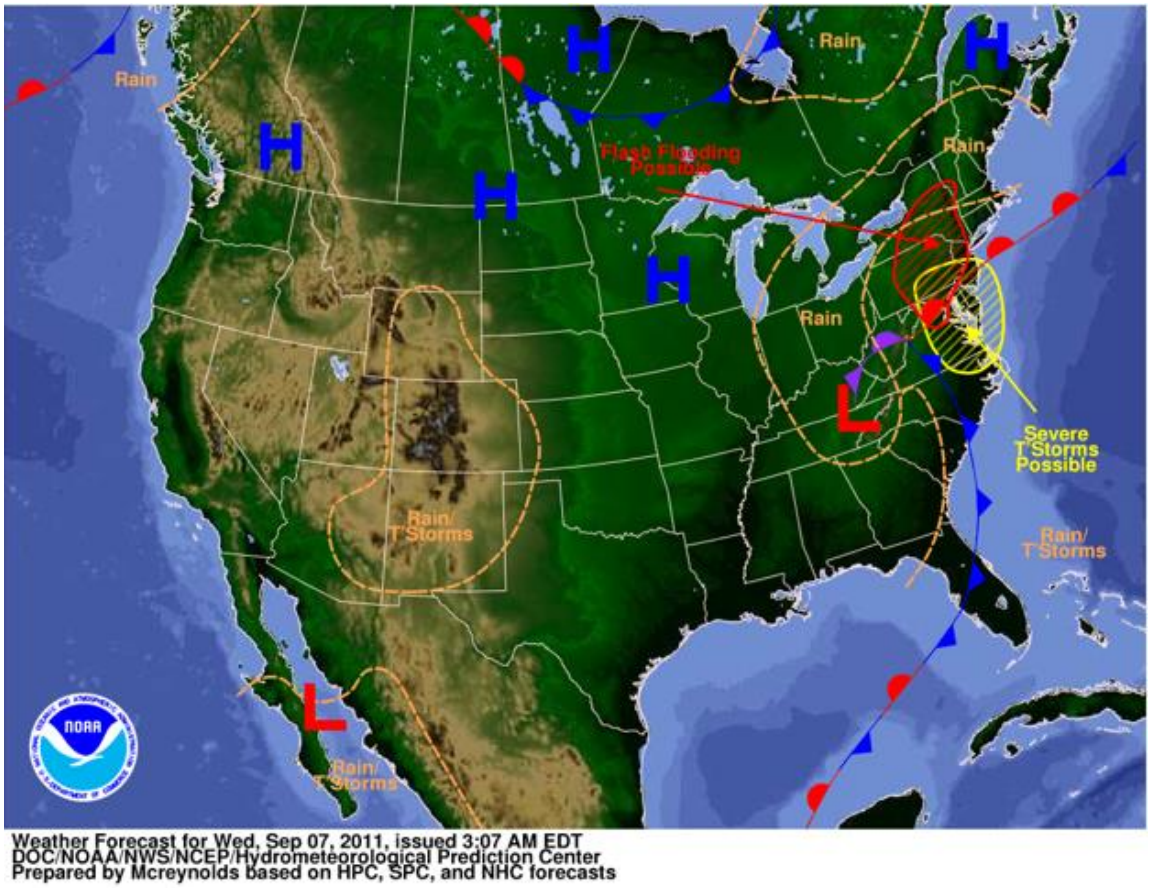

Figure 56. NOAA WPC Weather Forecast map issued for 2011-09-07 (NOAA NWS/ WPC).

At the Neversink River near Claryville, NY gauge, a discharge of 6.1% of bankfull was recorded four days before the flood peak, increasing to 134.4% of bankfull at the flood peak (Figure 57, Table 2-A).

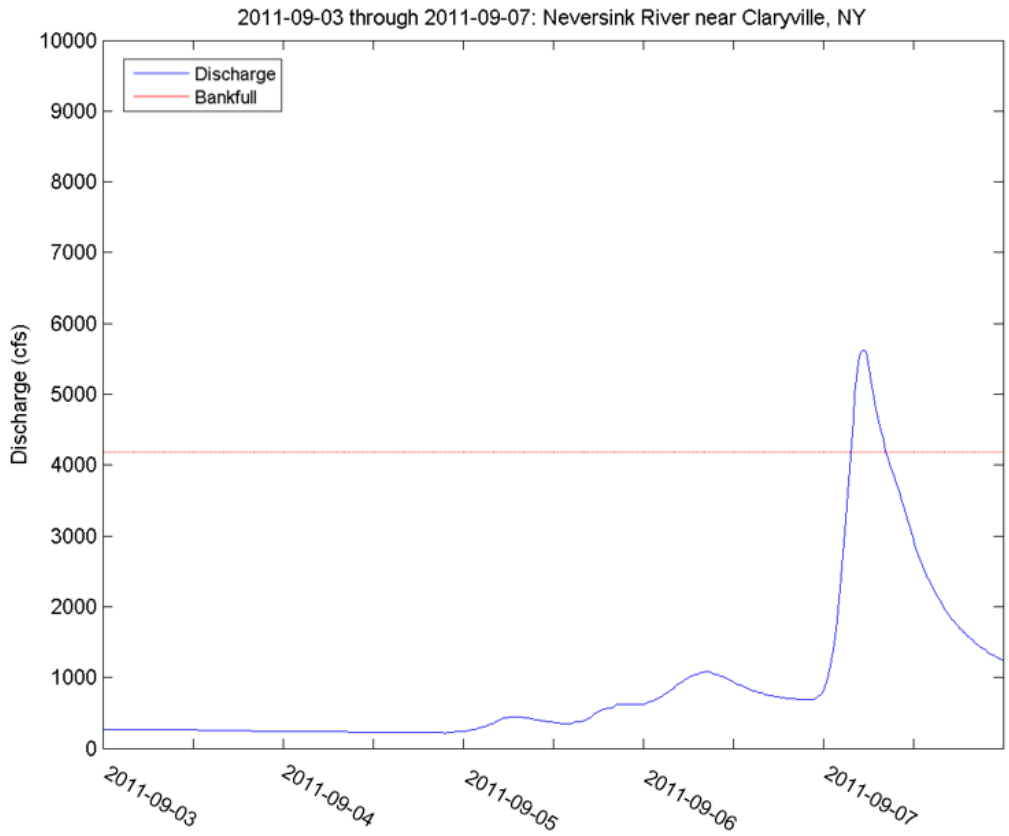

Figure 57. As in Figure 8, but for 2011-09-03 through 2011-09-07

The Esopus Creek at Allaben, NY gauge showed an increase from 11.4% of bankfull four days before the flood peak to 206.7% of bankfull at the flood peak (Figure 58, Table 2-B).

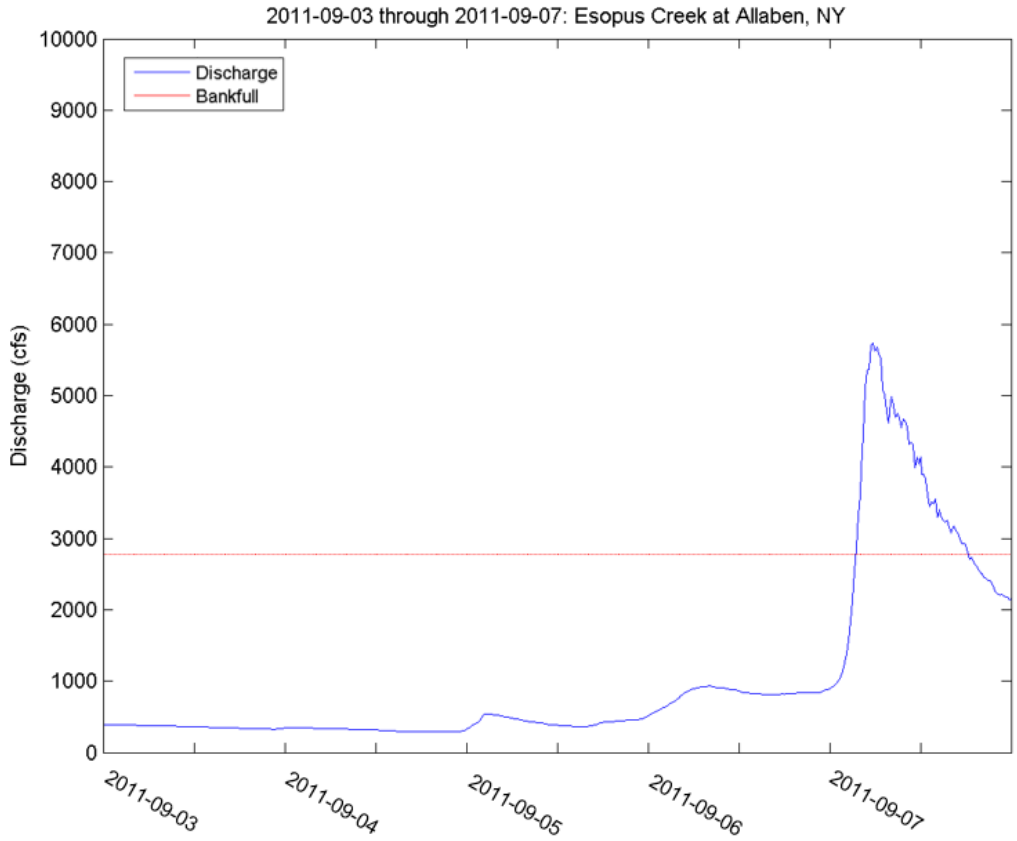

Figure 58. As in Figure 12, but for 2011-09-03 through 2011-09-07

Event 23: 2012-9-18

The last flash flood detected, on 2012-9-18, was forecasted in the weather forecast map. The map shows a large low pressure system moving eastward, extending from Alabama through Quebec, with the front changing from a cold front in the southern extent to a stationary from northward of Ohio (Figure 59). Severe thunderstorms were forecasted from New Hampshire through South Carolina. The map indicated possible flash flooding from Maine through Tennessee and North Carolina along the Appalachian Mountains east of the low pressure, with precipitation over the eastern third of the United States. The areas of possible flash flooding are partially overlapped with forecasts for possible severe thunderstorms as well, adding to the potential for heavy precipitation. Tropical disturbance Kristy is depicted south of California, further alluding to the general atmospheric instability at the synoptic scale on this day. With the maximum temperature of 68°F (20°C) and a minimum of 46.9°F (8.3°C), 0.7 in (17.8 mm) of precipitation was recorded on this day (Table 2-A, Table 2-B).

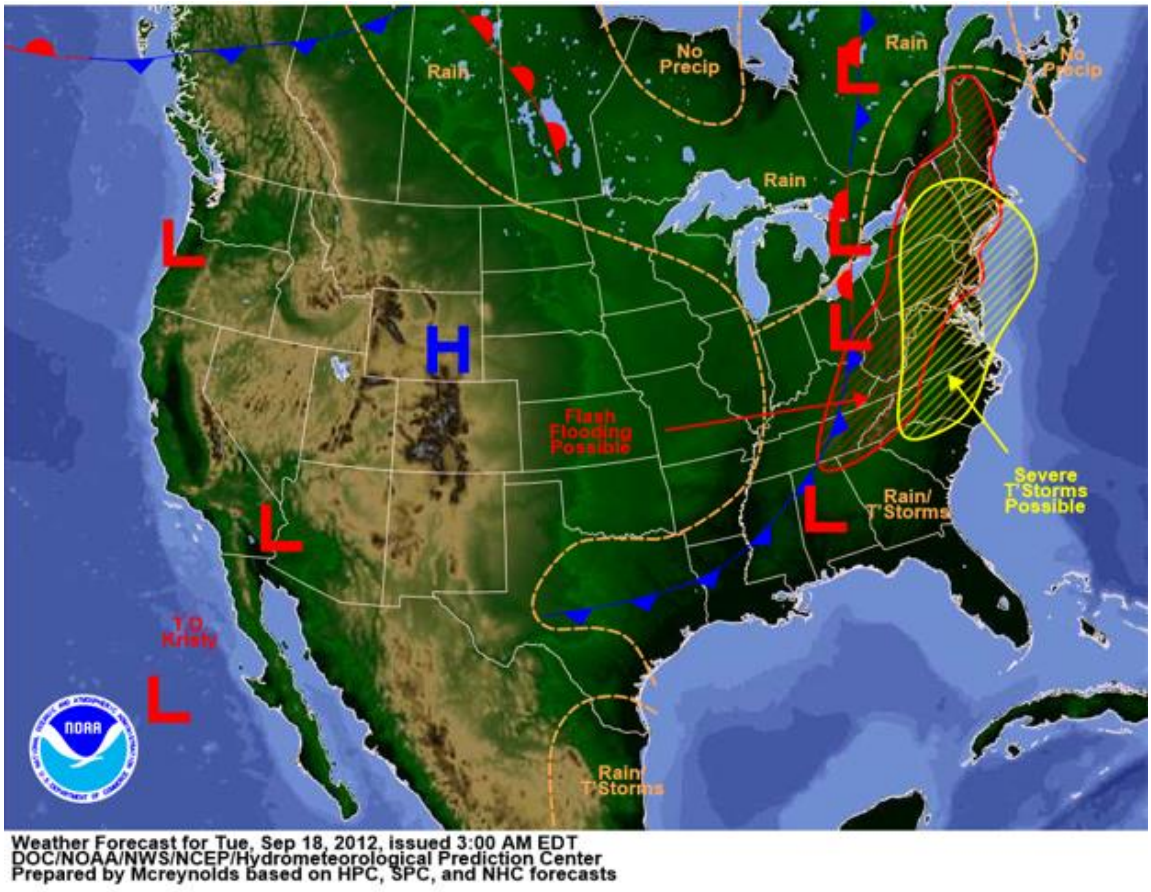

Figure 59. NOAA WPC Weather Forecast map issued for 2010-12-18 (NOAA NWS/ WPC).

Flash floods were detected in both watersheds. The discharge at the Esopus Creek watershed rose from 0.7% of bankfull discharge 4 days before the flood to 220.8% during the flood peak (Figure 60, Table 2-B).

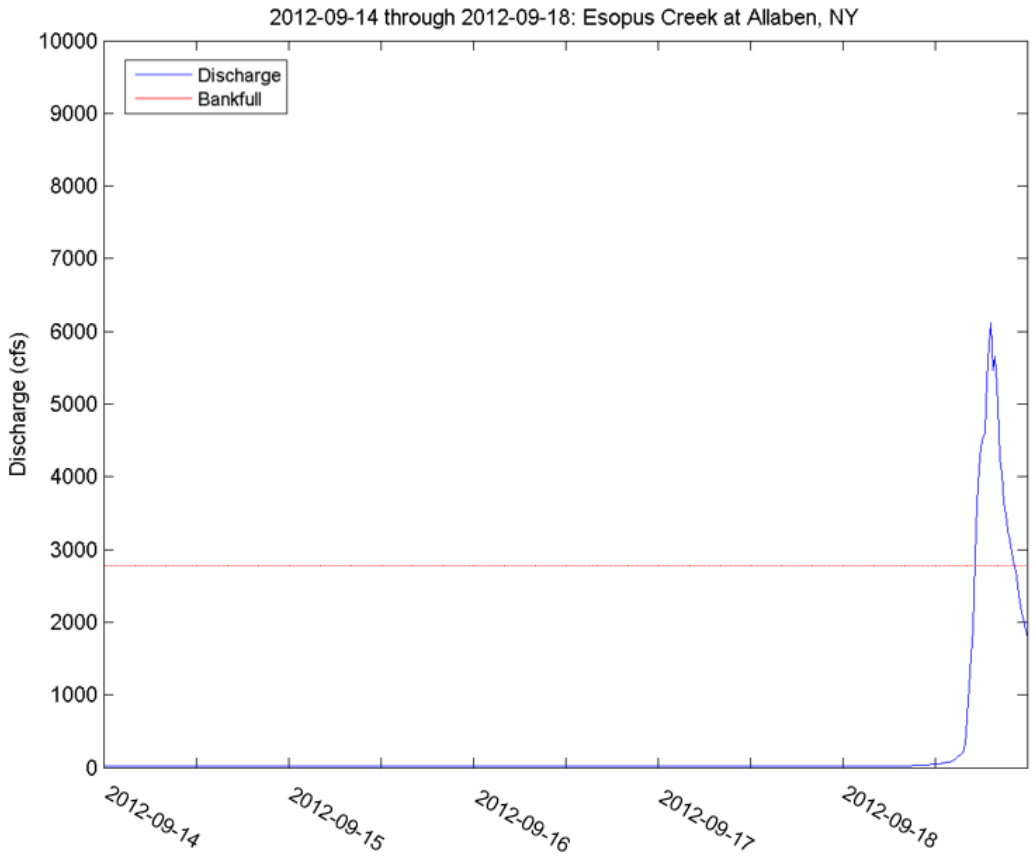

Figure 60. As in Figure 12, but for 2012-09-14 through 2012-09-18

Discharge at the Neversink River increased from 1.1% four days before the flood to 425.6% of bankfull discharge at the peak of the flood (Figure 61, Table 2-A).

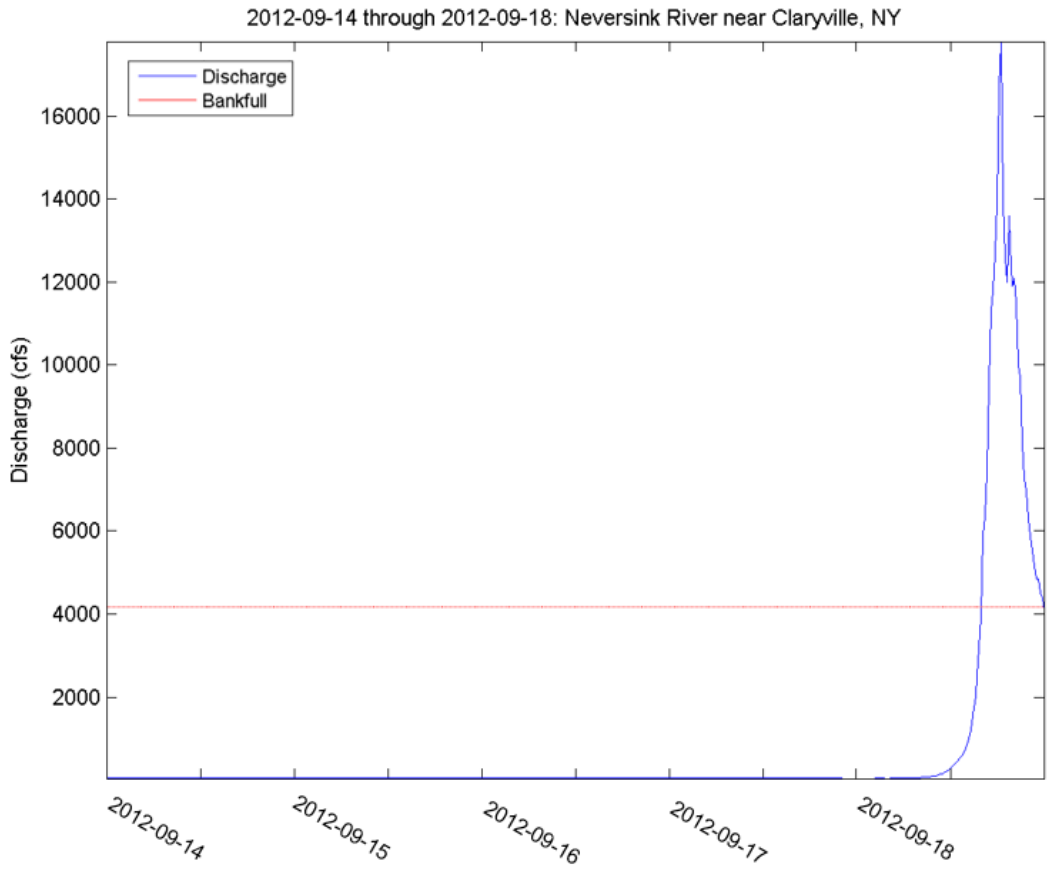

Figure 61. As in Figure 8, but for 2012-09-14 through 2012-09-18

4.2. Objective 1 discussion

The grouping of flash floods in fall and winter months, accompanied by the relative absence of flash floods in the spring and summer months, suggests that these events are associated with mid-latitude cyclones passing through the region, though frontal boundaries and surface conditions were not analyzed. This distribution differs from that in Maddox et al. (1979), who found that 25% of 151 flash flood events in the

contiguous United States from 1973-1977 occurred in July, and 86% of the events occurred in the warm season between April and September. However, difference in study area, methodology and flash flood detection may account for some of this variance. Some of the summer flash flood events occurred within two weeks of the passage of tropical storm remnants, from which the soil moisture was still likely near saturation. Rain-on-snow events were likely triggers of some of the winter flash floods, particularly Events 1, 3, 4, 10, and 18 where snow was present on the ground on the day of the flash flood peak.

The monthly distribution of flash floods associated with synoptic events in the current study differs from the findings in Maddox et al. (1979). While Maddox et al. (1979) showed a bimodal monthly distribution of flash flood-producing synoptic events peaking in the spring and fall months, the current study contrasts by finding no flash flood events in the spring months. A variety of differences could explain this discrepancy. These studies used different data and methods and their spatial and temporal extents were also different. The events included in Maddox et al. (1979) occurred throughout the contiguous United States, with most of their flash flood-producing synoptic events located throughout the Midwest, South, and southern Great Plains, with only one occurring in the same region as this study. This could be interpreted as there being no evidence of the bimodal distribution of this type of flash flood being observed in the northeastern United States, thereby discounting the lack of flash floods in the spring months in the current study. Further supporting this interpretation is the observance that the only flash flood associated with a synoptic event

in Maddox et al. (1979) occurred in the summer months (June, July, or August), while the bimodal distribution suggests higher probability of this type of event occurring in the spring or fall. While sample sizes are too small in either study to support this interpretation, this occurrence suggests that the bimodal distribution is true for the contiguous United States and not the New York watersheds that were the focus of this study. Several factors are responsible for the differences in flash flood occurrence in this region as compared to the rest of the United States; these include evapotranspiration amount and rate, differences in moisture influx, and thermodynamic patterns.

The three largest flash floods in the Neversink River watershed were among the six largest floods at the Esopus Creek watershed. However, while the second greatest flash flood precipitation event produced the second highest flash flood discharge in the Esopus Creek watershed, a flash flood was not recorded in the Neversink River watershed on this day. This likely indicates different streamflow regimes, which is expected given the differences in channel, bankfull discharge, and orientation of the watersheds. Another possible explanation for this could be differences in antecedent conditions not detected in the daily meteorological data. Further, the differences in flash flood peak discharge rank could be attributed to small storms only producing flash floods in one of the watersheds, due to small size, a storm track that misses one of the two watersheds, or a combination of both factors. Evidence of these small systems may not be detected in the basic daily meteorological data found in Tables 2 and 3.

Contrary to expectations, no flash floods occurred in the typical period for snowmelt in the region in February through April. This unexpected finding suggests that

while concerns have been raised regarding the shift in snowmelt to earlier in the spring, these changes do not appear to be frequently responsible for flash flooding in the region. Seasonal snowmelt was likely not detected as a flash flood because of a slower onset, with the rise to peak discharge exceeding the 6-hour threshold defined by this analysis. That is, snowmelt may indeed flooding, but it does not usually cause flash flooding in these watersheds.

5. OBJECTIVE 2 RESULTS AND DISCUSSION

5.1. Typical synoptic patterns

The input data for the synoptic typing were NCEP/NCAR 500 mb geopotential height reanalysis data from 1987-2013. The percent variance explained by each additional principal component declines rapidly after 3 components, with the first three components explaining 97% of the variance and each following component explaining very little variance (Figure 62). Therefore, three principal components were used in this analysis.

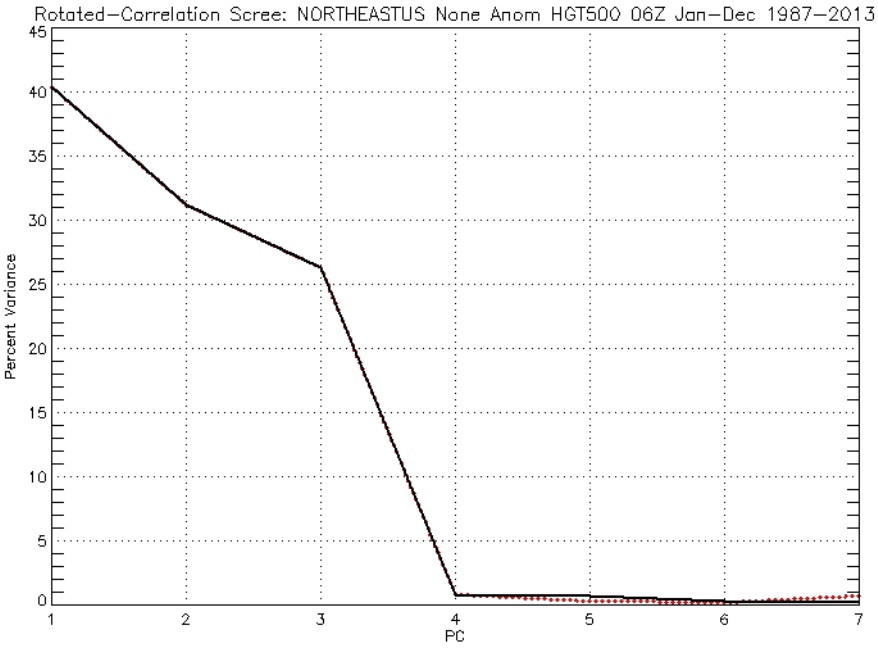

Figure 62. Scree plot for principal component analysis

Type variance plots were generated in the Spatial STT for 10 to 20 types. Visual assessment of the within-type variance plot shows that the range of variance is smallest and lowest for thirteen and seventeen types (Figure 63-A). However, a clustering of low within-type variances is apparent with seventeen types and is absent with thirteen types. Additionally, the variance between types is higher with seventeen types than with thirteen types (Figure 63-B), indicating that 17 typical patterns better classify the regional climatology.

The sequences of weather types occurring during and before flash floods in the Neversink River at and Esopus Creek watersheds were examined using both the thirteen-type and seventeen-type schemes. Finding that each of the types occurring on flood days had a matching type in the other classification scheme, detecting circulation patterns associated with flash flooding was not greatly affected by the classification scheme chosen. Therefore, the statistical preference for a seventeen-type classification scheme indicated by the variance plots led to the classification being selected and used for all analyses.

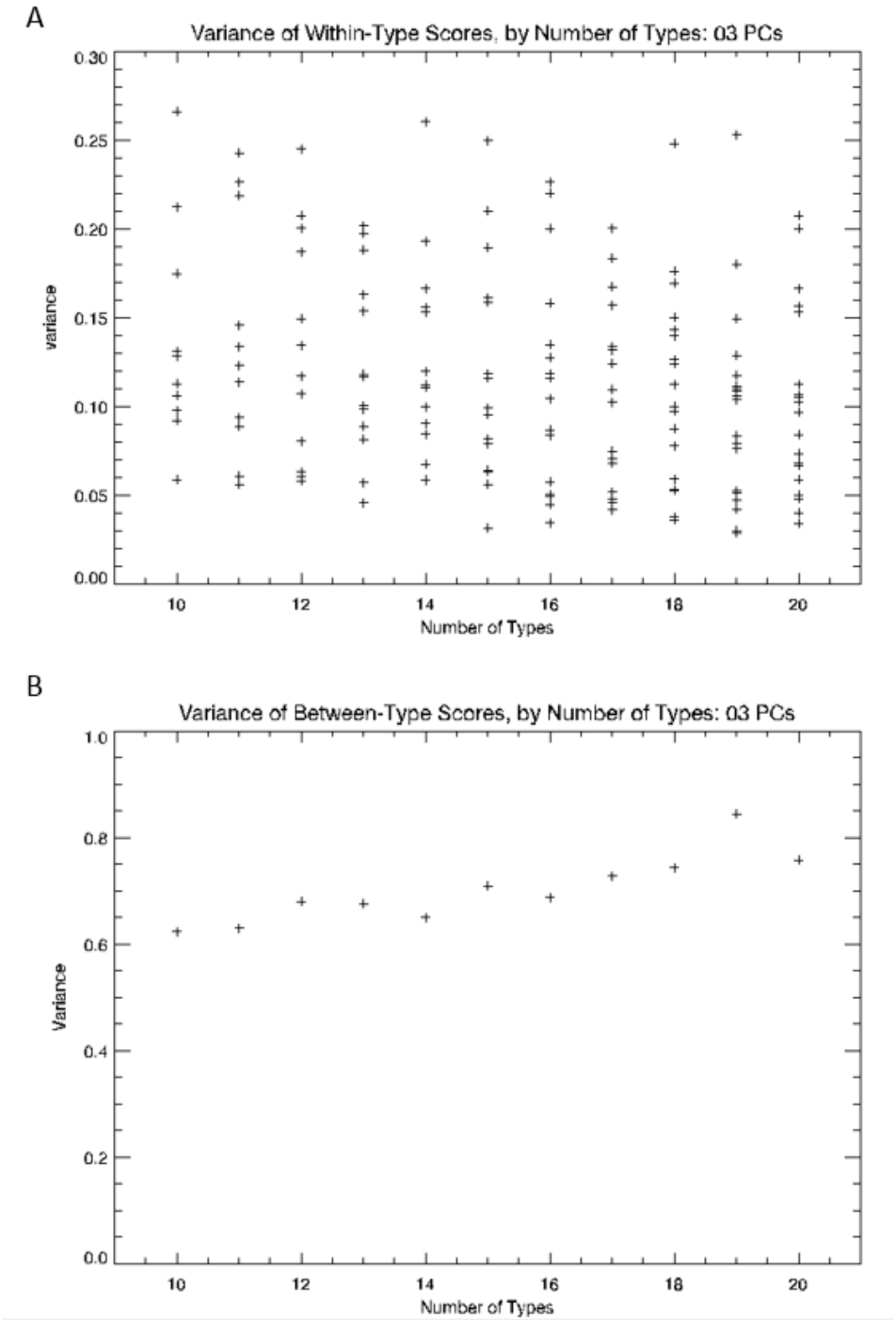

Figure 63. Variance of (A) within-type scores and (B) between-type scores

5.1.1. Type descriptions

Types are numbered according to their frequency, with ascending type numbers indicating descending total frequency for the study period. Composites of each of the 17 typical synoptic patterns generated in Spatial STT for 1987-2013 are shown in Figure 64.

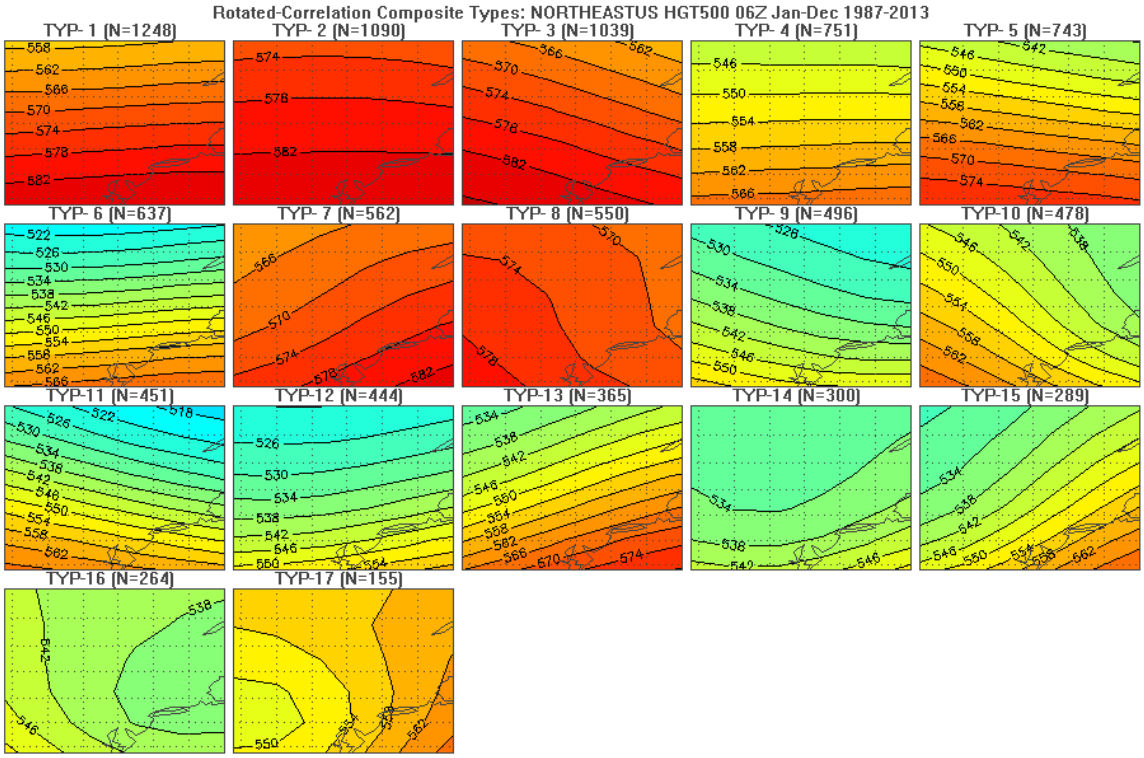

Figure 64. 500 mb geopotential height composites of the 17 synoptic types from 1987-2013.

Relative monthly frequencies of each type from 1987-2013 are shown in Figure 65. For viewing ease, subplot 65-A shows summer types, while subplot 65-B and 65-C show winter and transitional types, respectively. The types were divided into seasons based on frequency for the purpose of clarity and as such, statistical tests for differences were not warranted. Descriptions of each of the types follow. For clarity, month names in this section refer to all recurrences of that month throughout the study period (e.g. "January" indicates every month of January from 1987 through 2013). Describing the months as such allows for relative frequency description in the text to align with the axes in the type frequencies plot. The seasonality of each type is clear, with the increases in summer types occurring at the loss of winter types, and vice versa. Many of the summer and winter types occur with a fairly normal monthly distribution. The transitional types are all bimodal and are not necessarily symmetric in monthly distribution.

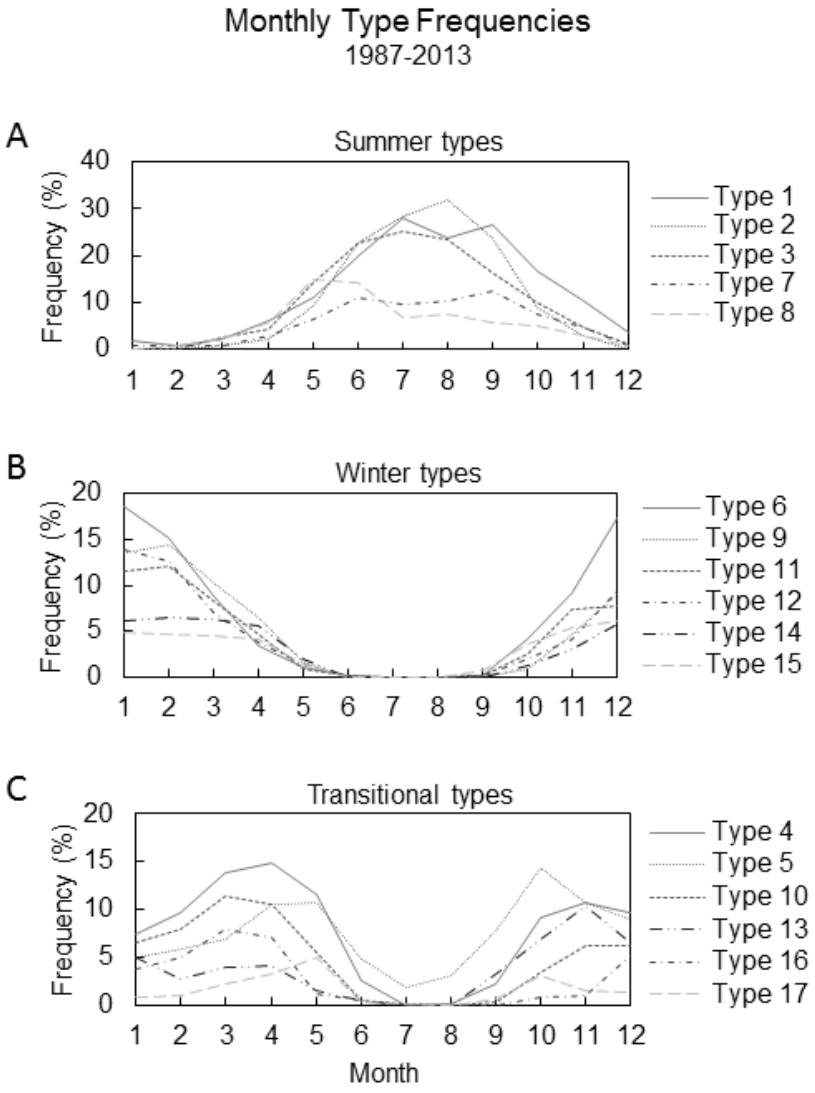

Figure 65. Synoptic type frequencies by month based on 500 mb geopotential height for years 1987-2013. Types are separated into subplots by those peaking in (A) summer, (B) winter, or (C) transitional months.

Table 3 shows the precipitation characteristics for each of the 17 synoptic types. Precipitation measurements were made at the GHCND station at Claryville, NY from 1987-201

Table 3. Precipitation (precip.) characteristics for each synoptic type. Daily precipitation data are from GHCND:USC00301521 Claryville, NY from 1987 through 2013

		Precipitation						Snowfall					
Type	Sample Size	Percent of days with precip.	Maximum precip. (in)	Minimum precip. (in)	Mean precip. (in)	Median precip. (in)	Standard Deviation	Percent of days with snowfall	Maximum precip. (in)	Minimum precip. (in)	Mean precip. (in)	Median precip. (in)	Standard Deviation
1	1248	44.6%	3.6	0.0	0.2	0.0	0.4	0.0%	0.1	0.0	0.0	0.0	0.0
2	1090	31.9%	8.6	0.0	0.2	0.0	0.5	0.0%	0.1	0.0	0.0	0.0	0.0
3	1039	26.3%	2.7	0.0	0.1	0.0	0.2	0.0%	0.1	0.0	0.0	0.0	0.0
4	751	43.0%	2.6	0.0	0.1	0.0	0.3	0.0%	1.2	0.0	0.0	0.0	0.1
5	743	35.7%	1.9	0.0	0.1	0.0	0.2	0.0%	0.7	0.0	0.0	0.0	0.0
6	637	46.2%	2.1	0.0	0.1	0.0	0.3	0.1%	0.9	0.0	0.0	0.0	0.1
7	562	66.0%	4.6	0.0	0.4	0.1	0.7	0.0%	0.6	0.0	0.0	0.0	0.0
8	550	29.9%	4.1	0.0	0.1	0.0	0.4	0.0%	0.3	0.0	0.0	0.0	0.0
9	496	40.8%	2.0	0.0	0.0	0.0	0.2	0.1%	1.9	0.0	0.0	0.0	0.1
10	478	33.7%	1.0	0.0	0.0	0.0	0.1	0.1%	0.6	0.0	0.0	0.0	0.1
11	451	40.8%	1.5	0.0	0.0	0.0	0.1	0.1%	0.9	0.0	0.0	0.0	0.1
12	444	58.2%	2.7	0.0	0.2	0.0	0.4	0.1%	1.2	0.0	0.1	0.0	0.2
13	365	73.0%	6.8	0.0	0.5	0.2	0.7	0.1%	0.8	0.0	0.0	0.0	0.1
14	300	65.1%	4.0	0.0	0.2	0.1	0.5	0.2%	2.2	0.0	0.2	0.0	0.3
15	289	88.1%	4.8	0.0	0.8	0.5	0.9	0.1%	2.1	0.0	0.1	0.0	0.3
16	264	57.7%	3.1	0.0	0.2	0.0	0.5	0.2%	2.5	0.0	0.1	0.0	0.3
17	155	84.6%	6.6	0.0	1.0	0.6	1.2	0.1%	1.2	0.0	0.1	0.0	0.3

Type 1

The composite map pattern for Type 1 (n=1248) exhibits zonal flow, with moderate to high geopotential heights ranging from 558 to 582 decameters over the classification area, also displaying a moderate pressure gradient (Figure 64). This pattern indicates no immediate approach of a frontal boundary or low (high) pressure system. Type 1 was more frequent in warm months, with peak frequencies in July and September accounting for 28.5% and 26.5%, respectively, of the type days in those months (Figure 65-A). In contrast, Type 1 occurrences made up less than 1% of days in the February. Precipitation occurred on 44.6% of Type 1 days, but fell as snow on less than 1% of days (Table 3).

Type 2

The Type 2 (n=1090) composite map and examples show a similar zonal flow to that in Type 1, but with higher heights, indicating warmer air temperatures (Figure 64). The gradient of 500 mb height is shallow, ranging from 574 to 582 decameters over the area classified. The frequency of Type 2 peaks in August, accounting for 32.7% of the days in that month (Figure 65-A). The slight majority of days are classified as Type 2 in July as well, occurring 28.6% of the days. As the warmest type with highest frequency in July and August, there were no instances of Type 2 occurring in February. Precipitation occurred on 31.9% of Type 2 days, falling as snow less than 0.1% of occurrences (Table 3).

Type 3

A departure from zonal flow is apparent in the Type 3 (n=1039) composite maps and examples, with a slight southeasterly shift in isobars suggesting either the lee of a trough or advance of a ridge (Figure 64). With moderately high 500 mb heights ranging from 562 to 582 decameters, the pressure gradient is greater than that of Type 2. Type 3 was the driest of the types, with precipitation falling on 26.3% of Type 3 days and snow falling on less than 0.1% of days (Table 3). Type 3 is most frequent in July, accounting for 25% of days in that month, but is never the single most frequent type in any month (Figure 65-A). Type 3 and Type 2 each account for 23% of days in June, which is as high as Type 3 ranks in most frequent type by month.

Type 4

Type 4 (n=1039) maps are characterized by zonal flow, with a similar north-south geopotential height gradient as in Type 1, but is notably cooler (Figure 64). This type is most frequent in transition seasons; while nonexistent in July and rare in August, it is the most frequent type observed in March, April, and November, accounting for 14%, 15%, and 11% of the days in those months, respectively (Figure 65-C). Precipitation was recorded on 43% of Type 4 days, with 10.4% of the type days experiencing snow (Table 3).

Type 5

Similar to Type 4, Type 5 (n=743) is also a transitional season type, peaking in October with 14% of days (Figure 65-C). Cyclical high frequencies of Type 5 are also observed in April and May, accounting for 10% and 11% of days, respectively. The Type 5 composite map shows predominantly zonal flow, with a slight northwest-southeast tilt in the geopotential height gradient (Figure 64). The gradient in this type is higher than the first four types, ranging from 542 decameters in the northeastern portion of the input region to 574 decameters in the southern extent. Precipitation occurred on 35.7% of Type 5 days though falling as snow on only 4.9% of the type days (Table 3).

Type 6

The Type 6 (n=637) composite map is characterized by the second highest geopotential height gradients of all the types, ranging from 522 decameters in the northern extend of the analysis region to 566 decameters in the south (Figure 64). The flow in this type is predominantly zonal. The composite type map shows a slight southwest-northeast tilt in the flow, but the shift is not strong enough to be considered anything but zonal. Type 6 is a winter type, as the most common type in December, January and February (17%, 19%, and 15% respectively), and was rare or nonexistent between May and October (Figure 65-B). The geopotential heights also support the statement that this is a winter type, with the second lowest height of 522 decameters in the study extent. Furthermore, while 46.2% of Type 6 days total experienced

precipitation, 19.7% of days experienced snow, supporting that this is a winter type (Table 3).

Type 7

The Type 7 (n=562) composite map is the first pattern to show notable departure from zonal flow, with a trough advancing over the central United States (Figure 64). The isobars suggest the advection of southerly air masses along the Atlantic coastline into the study region, which tend to have warm, moist characteristics. Type 7 is a warm type, peaking in June and September with 11% and 12% of the days in those months, respectively (Figure 65-A), with decreasing frequency in winter months. Frequencies of this type remain higher in July and August, accounting for just <11% of the days, and decreases in the winter months, though still observed. The geopotential height gradient is somewhat shallow in this type, ranging from 566 decameters in the northwest portion of the analyzed area to 582 decameters in the southeast. Precipitation occurred on 66% of Type 7 days, but only fell as snow on 1% of the days of this type (Table 3).

Type 8

The geopotential height gradient in the Type 8 (n=550) composite map is the shallowest of all the type composite maps, only ranging from 500 mb pressure height at 570 decameters in the northeast corner of the analyzed extent to 578 decameters in the southwest (Figure 64). Opposite of Type 7, the Type 8 composite shows the advance of a ridge into the study region, likely forcing the advection of cool, dry air masses from

central Canada into the region. These conditions are likely followed by warm, dry air masses as the ridge passes over the study area. Type 8 was one of the driest types, with precipitation falling on 29.9% of days and falling as snow less than 0.1% of the time (Table 3). While the previous transitional season types have shown dual peak frequencies in spring and fall months, Type 8 is most frequent in May and June, accounting for 15% and 14% of the days in those months respectively, but does not show peak frequency again in the fall months (Figure 65-A). There is a gradual decline in frequency of Type 8 from August, accounting for 7% of days, to January, in which Type 8 was not observed.

Type 9

The Type 9 (n=496) composite map shows a similar flow to that in Type 8, though less meridional and with a stronger geopotential height gradient (Figure 64). The geopotential heights in Type 9 range from 526 decameters in the northeast corner of the analyzed extent, to 554 decameters in the southwest corner. These lower heights suggest cooler temperatures. This notion is supported by the monthly frequencies of the type: Type 9 was not observed from June through September of the study period (Figure 65-B). Type 9 is most frequent in February, accounting for 14% of the days, and decreases rapidly in frequency until the summer months. Type 9 patterns begin to appear again in October, accounting for just 1% of the days in that month, and rapidly become more frequent through February. While snow was recorded on 25.4% of Type 9 days (Table 3), precipitation in general was observed on 40.8% of days of this type.

Type 10

Type 10 (n=478) is also not common in the summer months, accounting for less than 1% of days in each month from June through September, with no type observations in August (Figure 65-C). The frequency of Type 10 observations peaks after Type 9, peaking in March, where the type accounts for 11% of the days. While the frequency of Type 9 patterns is more evenly distributed around the February peak, the frequency of Type 10 observations increases gradually from September through the March peak frequency, before rapidly decreasing in frequency to less than 1% in June. Type 10, while similar to Type 9, is slightly warmer and more meridional than Type 9 (Figure 64). The 500 mb geopotential heights in the Type 10 composite map range from 538 decameters in the northeast portion of the analyzed extent to 566 in the southwest corner. This type is also indicative of an advancing ridge, with isobars suggesting the advection of cool, dry northerly air masses into the study region. Precipitation occurred on 33.7% of Type 10 days, with more than half of precipitation days (17.5%) recording snow (Table 3).

Type 11

The Type 11 (n=451) composite map has the highest 500 mb geopotential height gradient of all the types, ranging from 518 decameters in the northeast portion of the analyzed area, to 566 in the southwestern corner (Figure 64). The geopotential heights indicate the advection of cooler air masses from central Canada into the study region. 23.8% of Type 11 days experienced snowfall (Table 3), while precipitation in general

was recorded on 40.8% of days. Type 11 is most frequent in February, accounting for 12% of the days, and gradually deceases in frequency until July and August, in which no Type 11 patterns were observed for the study period (Figure 65-6). Type 11 patterns begin again in September, accounting for less than 1% of the days, and increases in frequency through January, accounting for 11% of days in that month.

Type 12

The composite map for Type 12 (n=444) indicates the advance of a trough into the study region (Figure 64). While the pressure gradient isn't as steep as observed in Type 11, the pressure heights are similar, ranging from 526 decameters in the northern section of the study region to 558 in the southernmost extent. Type 12 is similar to Type 11 in terms of frequency, peaking in January with 13% of days and decreasing to no observances in June, July, and August before rising to accounting for 9% of days in December (Figure 65-B). Type 12 is one of the snowiest types, with snowfall recorded on 37.5% of Type 12 days and precipitation in general on 58.2% of days of this type (Table 3).

Type 13

An advance of a trough is also present in the composite map of Type 13 (n=365), though at an earlier stage than Type 12 (Figure 64). With 500 mb geopotential heights ranging from 534 decameters in the northwest portion of the analyzed area, to 574 decameters in the southeast corner, the days in this type are warmer on average than

in Type 12. The positioning of the pressure gradient is likely indicative of increased frontal activity in the advance of the trough, suggesting precipitation in the study area. Type 13 was one of the wettest types, with precipitation recorded on 73% of the days of this type and snow recorded on 14.2% of days (Table 3). Type 13 is most frequent in November, accounting for 10% of the days. The frequency of Type 13 decreases to 3% of days in February, increases slightly to 4% of days in April, and decreases down to no observances again in July and August before increasing again through September and October (Figure 65-C).

Type 14

The composite for Type 14 (n=300) indicates a deep trough in the study region, with 534 decameters dipping through the northern part of the study region and pressures increasing only to only 545 decameter geopotential height in the southern edge of the analyzed area (Figure 64). Type 14 is a winter type, with no observations in the summer months of June through September (Figure 65-B). The most Type 14 patterns are observed from December through April, though only accounting for 6% of the days in each of those months. The lower heights associated with this trough indicate cooler temperatures, a claim which is supported by the monthly frequencies as well. Type 14 is the snowiest of types, with 49.3% of days of this type having snow recorded but does not exhibit the highest amount of precipitation in general (Table 3).

Type 15

The monthly frequency of Type 15 (n=289) days is similar to Type 14, with frequencies at 5% of days from January through March, no observations in June, July, and August, then increasing to accounting for 6% of days in December (Figure 65-B). The 500 mb geopotential height ranges from 530 decameters in the northwest to 566 in the southwestern corner of the area analyzed (Figure 64). The direction of the pressure gradient suggests a trough advancing into the region, with the advection of southerly air masses along the coast into the study region. The combined presence of the trough with the advection of southerly air masses suggests precipitation in the region. Type 15 was the wettest type, with 88.1% of Type 15 days having precipitation. A total of 28.3% of Type 15 days experienced snow (Table 3).

Type 16

The composite of Type 16 (n=264) patterns suggests the passage of a mid-latitude cyclone through the study region, with the pressure minima exiting the east side of the analyzed area over New England (Figure 64). The pressure gradient suggested by the composite map is shallow, ranging from 500 mb geopotential height at 538 decameters over the New England coast to 546 decameters over the southwestern corner of the analyzed region. Type 16 is infrequent, accounting for 8% of days in a March peak and not observed from July through September with gradual increases and decreases in frequency (Figure 65-C). Low-pressure systems along the coastline, such as that in the Type 16 composite, are often associated with precipitation in the region. Type

16 was one of the snowiest types, with 39.2% of days in this type producing snowfall (Table 3), but with 57.7% of Type 15 days having precipitation, it was not the most particularly wet.

Type 17

Type 17 (n=155), the most infrequent type, is similar to Type 16 in that it indicates a closed low pressure system entering the region, but the pressure minima is over the inland mid-Atlantic portion of the area analyzed (Figure 64). The type is generally warmer than Type 16, with 500 mb pressure heights ranging from 550 decameters at the center of the low to 562 decameters along the coastline. The frequency reflects this difference from Type 16 as well, with the peak frequency at only 5% of days in May (Figure 65-C). The frequency of Type 17 decreases through no observations in August, and increases to have a second peak frequency—though only accounting for 3% of days—in October. The frequency decreases after October and remains at accounting for 1% or fewer of days from December through February before increasing back to the peak frequency in May. Type 17 was the second wettest type, with 84.6% of Type 17 days having precipitation. 17.6% of the Type 17 days experienced snowfall (Table 3).

5.2. Association of flash floods with synoptic types

Type 7 was the most frequent atmospheric pattern associated with flood events, with 47.8% of flood events passing through at least one Type 7 in the five days preceding and including the flood peak. Type 15 was nearly as frequent, with 43.5% of

flood events passing through the type in the same preceding interval. Type 17, the least frequent of the synoptic types, did not precede any of the flood events.

Types 7, 13, and 15 were the most frequent types concurrent with flood peaks. 96% of floods passed through at least one Type 7, 13, or 15 in the 5 days preceding and including the flood peak. The only flood peak that was not associated with Type 7, 13, or 15 was associated with the passage of Tropical Storm Irene (Event 21).

Figure 66-A shows the number of times a type was observed on the day of a flash flood. Figure 66-B shows the number of times a type was observed on 1 through 5 days before the flood peak. Counts are shown above and within the bar for reference. The high counts of type observances in the days before the flash flood peak suggests that the flash flood conditions develop in a few days or less, as the low type numbers by design linked to high frequency and normal conditions. This figure shows that most of the floods peaked under Types 7, 13, and 15, with 10 of the 23 flood peaks peaking in Type 15 circulation pattern. No types peaking under a different type were preceded by a Type 15 event (not shown). This shows that when Type 15 is associated with a flash flood, the Type 15 pattern only occurs on the day of the flood, and is not observed in the days leading up to the flood. Eight flood peaks occurred in Type 13 circulation. Only one event peaking under a different type was preceded by a Type 13 pattern, which occurred 5 days before the flood peak (Figure 66).

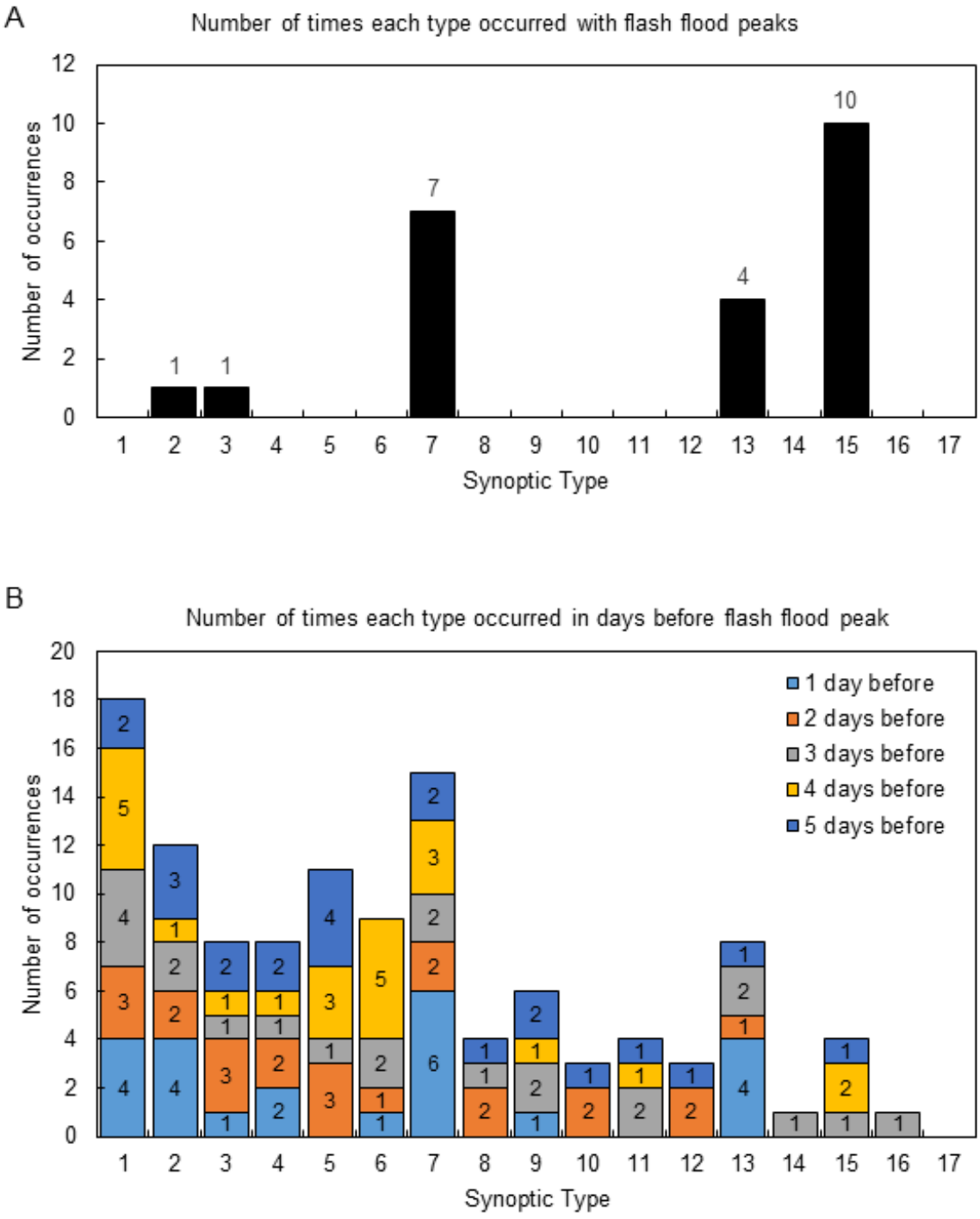

Figure 66. Type occurrences before flood events on (A) the flood peak days and (B) the 1-5 days preceding the flood peak.

A random-sampling bootstrapping was utilized analyze the pre-flash flood days for significant departure from normal type sequences, as outlined in Chapter 3. The results of this analysis are shown in Figure 67. Axes for all types are as shown for Type 1. The bars show the frequency of the number of times that each type was observed in a selection of random days (n=10000 random samples). The dashed lines indicate the 2.5% and 97.5% of the distribution. The red solid line shows the number of times each type occurred in the pre-flash flood subset. Statistically significant departure from normal conditions preceding flash flood days is deduced if the red solid line is outside of the dashed lines. Types 7, 13, and 15 were found to be significantly more frequent in the days preceding flash floods than normal, with the number of occurrences of these types before floods outside the 97.5th percentile of the normal distribution of type days. Type 10 was significantly less frequent in the days preceding flash floods than normal days. No other types showed significant deviation from normal before flood days.

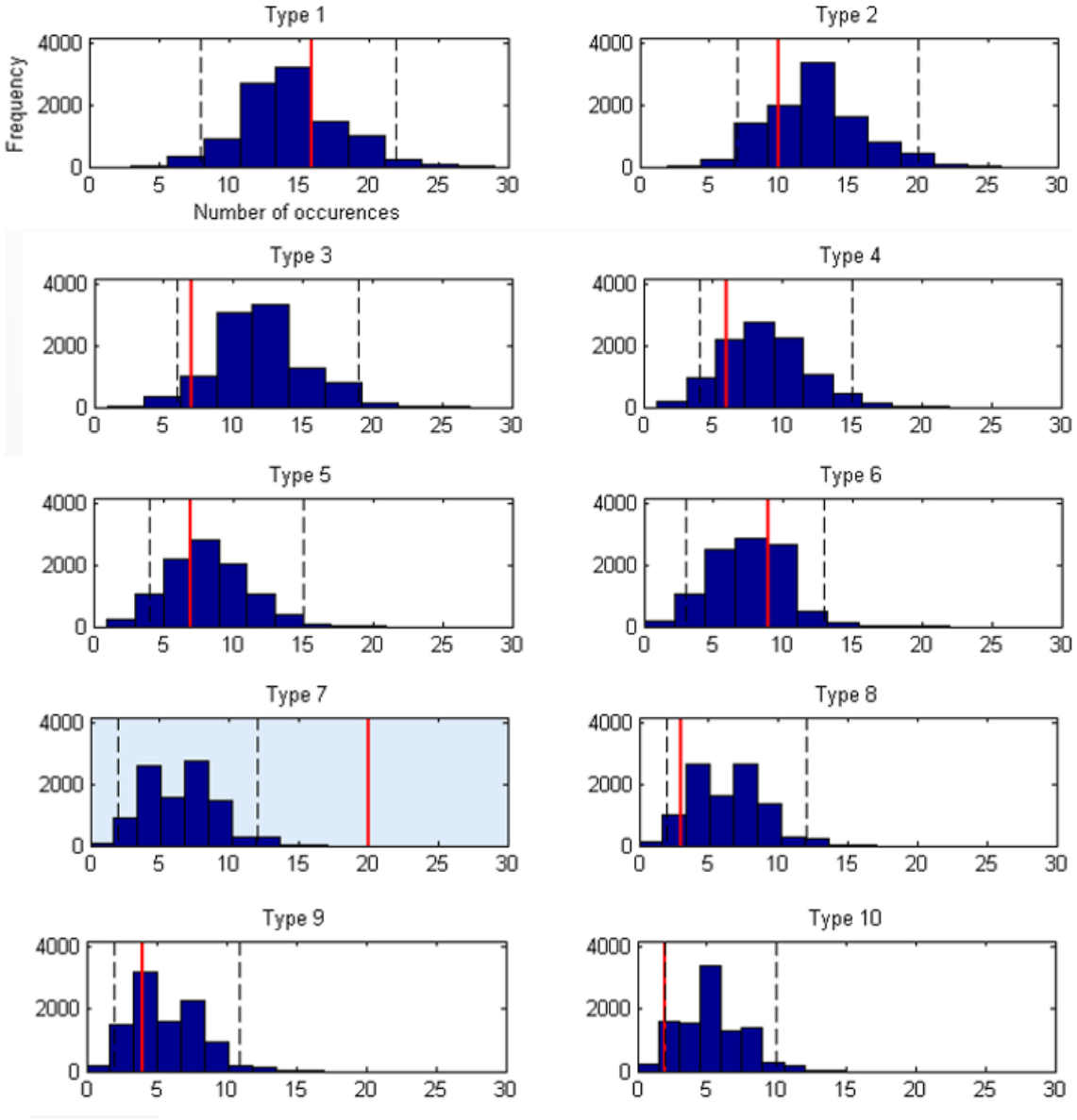

Figure 67. Bootstrapping analysis of synoptic type frequency

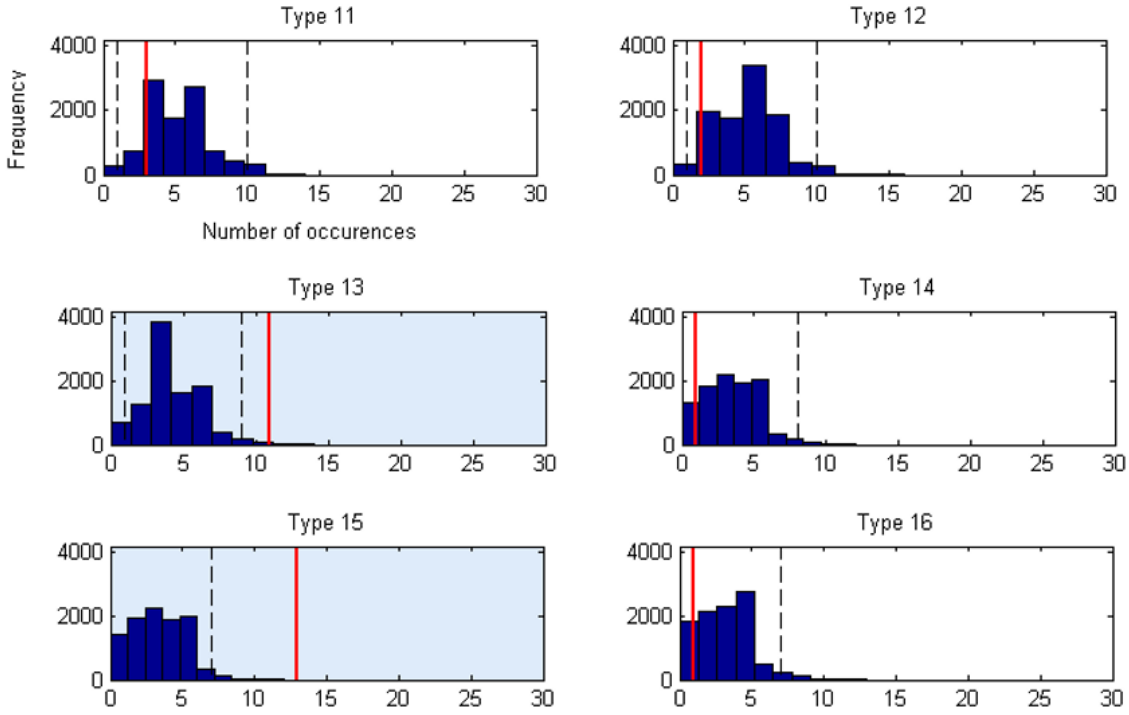

Figure 67. Continued. Axes for all types as shown for Type 11.

5.3. Objective 2 discussion

The prominence of Types 7, 13, and 15 before flash flood days is significant and worthy of further consideration. These three types are similar in they are the only types of all the typical patterns to show strong southwesterly flow at the 500 mb pressure height, suggesting the advance of a trough and frontal activity into the region. The advection of southerly air masses along the eastern seaboard, drawing warm, moist air masses into the region is connected to precipitation events. Furthermore, the direction of flow indicated in the 500 mb pressure height composites in Type 7, 13, and 15 is similar to that in flash flood-producing synoptic events described in Maddox et al. (1979).

Maddox et al. (1979) found that 20% of the flash floods over a 5 year period in the contiguous United States were associated with synoptic-scale events, usually an eastward-moving intense cyclone or trough. This pattern was observed in this study both through the weather maps of each event, and the composite of the flash flood-producing synoptic types. Maddox et al. (1979) further explain that these systems are often associated with severe thunderstorms and heavy rains, and that the slow movement of the systems may cause these conditions to persist for multiple days in some areas. Furthermore, Maddox et al. (1979) note that due to their larger expanse, these systems may produce flash floods in multiple watersheds. This was also seen in the present study, with nearly 40% of the flash flood days having flash floods in both the Neversink River near Claryville, NY and Esopus Creek at Allaben, NY watersheds. Therefore, it is clear that direction of flow at the 500 mb pressure level can still be strongly associated with flash flooding in the region. It should be noted, however, though most of the occurrences of these types and this directionality of flow are not associated with flash floods.

The significant departure absence of Type 10 days before flash flood days is interesting. This finding was unexpected and suggests that Type 10 days may act as a block of sorts to flash flood conditions. Type 10 is characterized by a trough and a northwesterly flow into the study area (Figure 64). Precipitation was recorded on approximately one third of Type 10 days, with approximately half of those precipitation days experiencing snow. It is possible while precipitation is occurring on one third of Type 10 days, the depth of precipitation is low.

While the results of the synoptic typing show preference for certain types and geopotential height gradients before flash flood events, linking synoptic-scale circulation patterns to watershed-scale flooding, more information could possibly be extracted if the discrepancies in physical and temporal scales were resolved. Given that the adjacent watersheds do not always experience flash flooding simultaneously, it could be suggested that differences in slope and aspect of the terrain between the basins could impact the trajectory and persistence of small storms in the watersheds, thus resulting in differences in timing of flash floods.

Kelsch (2001) reports increased precipitation efficiency when convective storms flow perpendicular to the topography, as lifting of the system is forced by the terrain and causes further cooling and precipitation. The effect of orography on extreme precipitation and urban flash flooding was observed in the New York-New Jersey region in October 2006, when a convective line from a cold front first precipitated heavily while passing over the Appalachian Mountains, and the lingering effects of the thunderstorms enabled development of a second phase of convective activity to precipitate over the same localized area again (Zhang et al. 2009). This event did not produce flash floods detected in this study, likely because the meteorological observations in Zhang et al. (2009) were located approximately 160 km south of the watersheds selected for the present study. In addition to modifying rainfall rates and duration, terrain has also been known to influence the trajectory of convective activity in the region: Wasula et al. (2002) determined that the terrain has a funneling effect on storm movement across eastern New York and western New England. Any effect of

terrain on storm systems will not be detected in the 500 mb pressure height reanalysis data used to classify the types, possibly providing explanation for the insignificant results and motivating a smaller-scale approach to this research question.

Collins et al. (2014) found that combining pressure both at the surface and aloft better classified storms northeastern United States and Canada than only using one level of pressure data, as was the case in this analysis. Furthermore, Collins et al. (2014) couples an environment-to-circulation approach with a circulation-to-environment subset to further classify storms. The procedure in this thesis utilizes circulation-to-environment analysis; that is, classifying atmospheric conditions and applying the classes to explain surface conditions or events (Yarnal 1993). By adding another component to the study, such an environmental classification subset, the results may be enriched and more detailed.

As found in other studies such as Collins et al. (2014), the use of synoptic typing at these scales has strengths in reducing large amounts of data into a desired number of classes. However, the nature of reducing the data in this way leads to limitations, including replacing unique patterns by composite maps which may not fully represent all factors influencing environmental conditions. While the purpose of synoptic typing in this study was to determine the atmospheric patterns associated with flash flooding, the typical synoptic patterns identified do not provide novel information on atypical flash flooding events.

6. OBJECTIVE 3 RESULTS AND DISCUSSION

The purpose of Objective 3 was to determine if the peak discharge of flash floods in the Neversink River and Esopus Creek watersheds could be estimated using basic, easily accessible measurements and forecasts. This objective was achieved through carrying out simple and stepwise linear regression analyses on a variety of transformed and untransformed variables.

A correlation matrix was assembled to examine the predictors for potential problems with multicollinearity (Table 4). In Table 4, a number of potential multicollinearity problems were apparent. These issues are highlighted in red for ease of viewing. The 24-hour precipitation correlated with 24-hour precipitation intensity (r= 0.77) and maximum precipitation intensity (r= 0.69). The maximum precipitation intensity correlated highly with 24-hour precipitation intensity (r= 0.72).

Table 4. Correlation matrix for all predictors

	Antecedent percent bankfull	*Antecedent soil moisture*	*24-hour precip. total*	*Precip. duration*	*24-hour precip. intensity*	*Maximum precip. intensity*
Predictor	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1.0	0.3	-0.5	-0.4	-0.3	-0.3
x_2		1.0	-0.1	-0.1	< 0.1	-0.1
x_3			1.0	0.4	0.8	0.7
x_4				1.0	-0.2	< 0.1
x_5					1.0	0.7
x_6						1.0

Precipitation abbreviated by 'precip.'

These potential problems were circumvented by removing the 24-hour precipitation total and the maximum intensity from the model. In removing these variables, the two correlations closest to the threshold—24-hour precipitation total with both antecedent percent bankfull (r= -0.50) and precipitation duration (r= 0.43)—were also removed in addition to those exceeding the r> 0.5 threshold. The four remaining predictors show little evidence of multicollinearity (Table 5). Note that the identifying predictor number (e.g. x_1, x_2, etc.) changed as variables were removed.

Table 5. Final correlation matrix for predictors

	Antecedent percent bankfull	*Antecedent soil moisture*	*Precipitation duration*	*24-hour precipitation intensity*
Predictor	x_1	x_2	x_3	x_4
x_1	1.0	0.3	-0.4	-0.3
x_2		1.0	-0.1	< 0.1
x_3			1.0	-0.2
x_4				1.0

6.1. Simple linear regression results

The regressions outlined in Figure 4 were performed in 6 groups, A-F. Each of the groups has 4 combinations of transformed and untransformed data. The only significant models were found in Group A, where all combinations of raw or standardized predictors with raw or log discharge were significant when using both precipitation datasets with both watersheds. The sample size was the largest in this group, with 26 flash flood events included in the analysis. Consequently, the F-statistics

were higher for all models in this group than in any other group, with the lowest model p-values. The F-statistics were slightly lower when using log discharge (F = 4.33) than when using the raw discharge data (F = 5.17). The 24-hour precipitation was the only significant predictor. Interestingly, the β coefficient for the antecedent percent bankfull was negative for this and several other groups. The best regression model in Group A is presented in Table 6, using both precipitation datasets and both watersheds, with standardized predictors and untransformed discharge. The regression tables for the other models in Group A with different combinations of transformed and untransformed data can be found in Appendix A along with the non-significant models from groups B-F.

Table 6. Best simple linear regression table for predicting flash flood peak discharge

Variable combination Group A, 2: Both datasets, both watersheds, standardized x, raw y				
	Estimated coefficients			
	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	7459.6	942.0	7.9	0
x1	-268.3	1224.6	-0.2	0.8
x2	1440.6	1016.7	1.4	0.2
x3	910.9	1152.8	0.8	0.4
x4	4201.5	1081.5	3.9	<0.01
Number of observations: 26, Error degrees of freedom: 21				
Root Mean Squared Error: 4800				
R-squared: 0.50, Adjusted R-Squared: 0.4				
F-statistic vs. constant model: 5.17, p-value = <0.01				

The regression equation when using standardized predictors and the untransformed discharge values is assembled in Equation 2.

$$y \sim 7459.6 - 268.3x_1 + 1440.6x_2 + 910.9x_3 + 4201.5x_4$$

Equation 2

Where *y* is the estimated discharge in cfs, x_1 is the antecedent percent bankfull, x_2 is the antecedent soil moisture, x_3 is the precipitation duration, and x_4 is the 24-hour precipitation intensity. This model explains 40% of the variance in peak discharge. The observed discharge and the predicted discharge from Equation 2 are shown in Figure 68. A 1:1 line is shown for reference. In this figure, outlying points are apparent and show that the model tends to underpredict high discharge events.

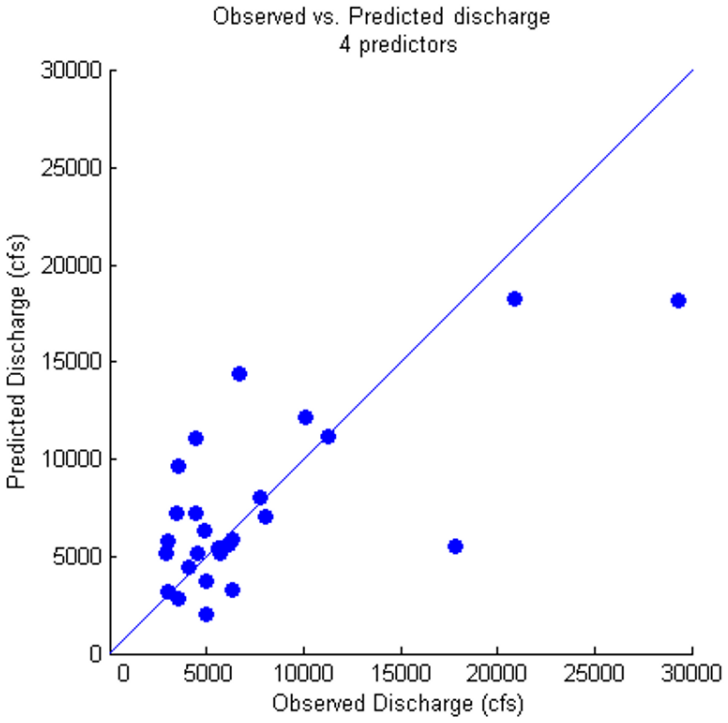

Figure 68. Observed vs. Predicted discharge (4 predictors)

Overall, the model results tended to be better (higher F-statistic, lower *p*-value, higher r^2) when using raw discharge instead of log discharge for all models. One point to note is that the β coefficient for antecedent bankfull is positive only when both precipitation datasets are used for the Esopus Creek watershed, and when the Claryville precipitation data only is used in the Neversink River watershed. However, none of the predictors or models were significant. All other predictors had positive β coefficients throughout the 4 models in each of the 6 groups. More interpretation of the non-significant models can be found in Appendix A.

6.2. Stepwise linear regression results

Summary statistics for the four predictors are show in Table 7. The stepwise linear regression was used to identify which predictor variables to retain in the model.

Table 7. Summary statistics of predictors in stepwise regression

Variable combination: Both datasets, both watersheds, standardized x, raw y

	Estimate (β)	Standard error	*t*-statistic	*p*-value
x_1	284.8	1187.0	0.2	0.8
x_2	1369.6	942.6	1.5	0.2
x_3	934.0	985.6	1.0	0.4
x_4	4246.6	985.6	4.3	< 0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 26

Predictor variables were added and then removed to find the optimal model fit. The best fit was found to include only antecedent soil moisture and 24-hour precipitation intensity

(Table 7). While the removal of precipitation duration from the model results in slight increase in RMSE, excluding this variable reduced the penalty for additional variables in the adjusted r^2. In addition to the decreased distance between the r^2 and adjusted r^2 and improvement of F- and *p*-values makes this step the optimal model (Table 8). A following step of including just the 24-hour precipitation intensity was also performed, though the removal of the soil moisture term caused an increase of RMSE and lowered the amount of variance explained.

Table 8. Stepwise linear regression table

Step	Predictors included	RMSE	r^2	Adjusted r^2	F	*p*
1		0	0	0	N/A	N/A
2	x_1	6183.6	0.1	0.0	1.2	0.3
3	x_1, x_2	6065.1	0.1	0.0	1.6	0.2
4	x_1, x_2, x_3,	6152.1	0.1	0.0	1.1	0.4
5	x_1, x_2, x_3, x_4	4803.1	0.5	0.4	5.2	<0.01
6	x_1, x_2, x_3	4698.0	0.5	0.4	7.2	<0.01
7*	x_2, x_4	4714.7	0.5	0.4	10.1	<0.01
8	x_4	4799.9	0.4	0.4	17.7	<0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 26

Overall, Step 7 of this model shows improvement over the comparable standard multiple linear regression model shown in Table 6. The adjusted r^2 in the stepwise regression shows slight improvement by explaining 42% of the variance in peak discharge, and the RMSE and F-statistic are also improved in this model as well (Table 8). Further, while the number of flash flood observations is too small to have high

statistical power in either model, the reduction of predictors included in the stepwise regression model reduces the potential for overfitting to the data. This regression equation, shown in Equation 3 where x_2 is the antecedent soil moisture and x_4 is 24-hour precipitation intensity, was considered an improvement over the standard multiple linear regression.

$$y \sim 7459.6 + 1369.6x_2 + 4246.6x_4$$

Equation 3

The predicted discharge values from this equation were plotted against the observed discharge values (Figure 69). A 1:1 line is shown for reference.

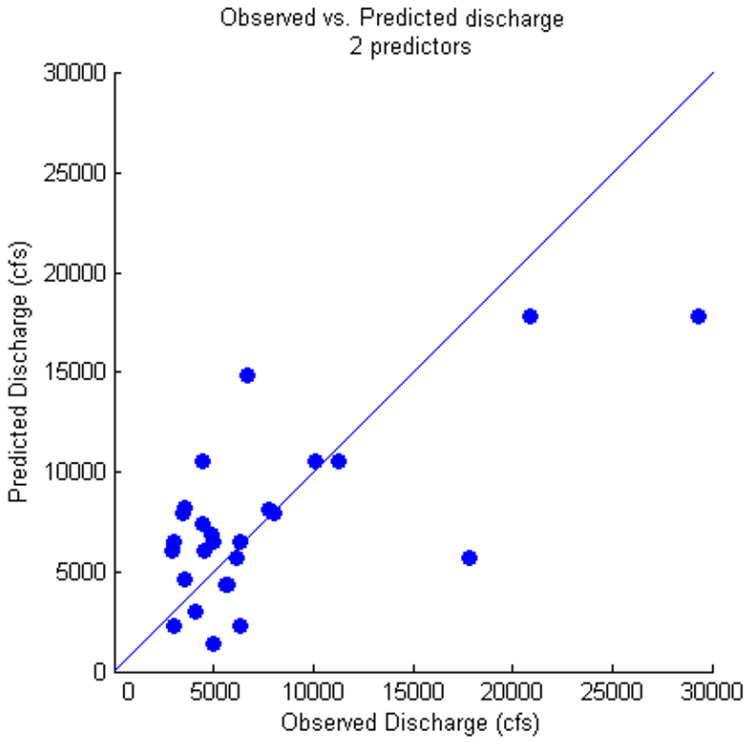

Figure 69. Observed vs. Predicted discharge (2 predictors)

Figure 70 shows the probability distribution of the residuals. The closer these points are to the dashed line, the closer the distribution is to normal. This figure shows that the residuals are distributed fairly normally, with some outliers at high values. This confirms that the model has difficulty accurately predicting high discharge values, as anticipated from Figure 69. The supposition that the points are outliers could be confirmed or rejected by utilizing Tukey's method or calculating Cook's distance. However, as flash floods are by nature extreme events and the sample size is already small, such efforts were purposefully omitted. With the exception of two outliers, Figure 70 shows that the residuals are fairly normally distributed.

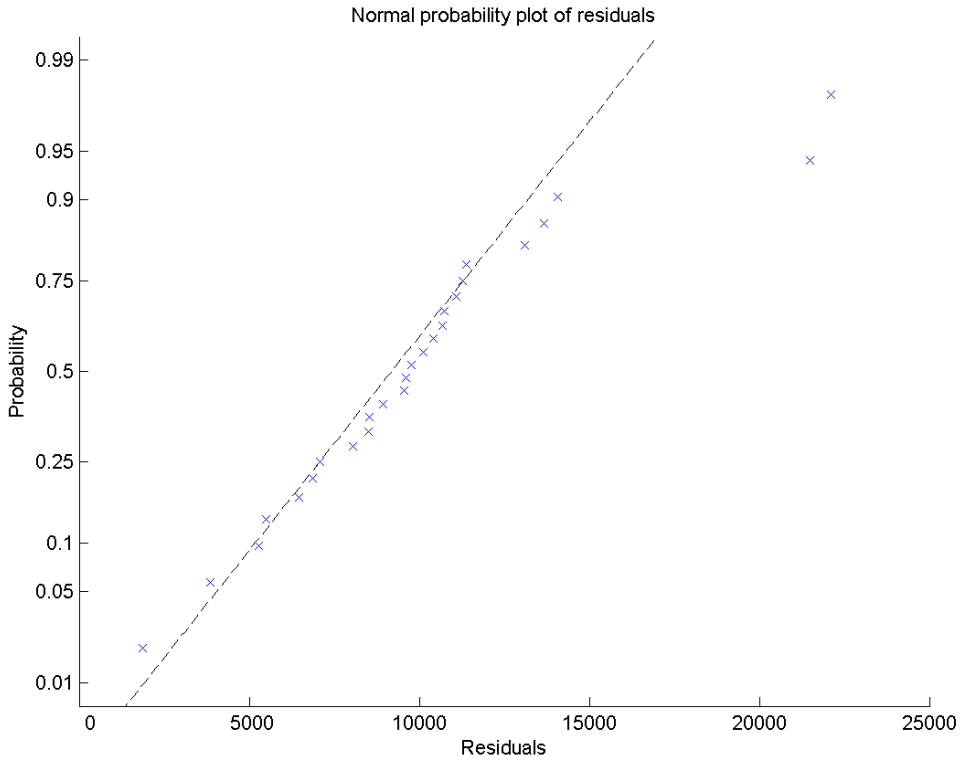

Figure 70. Normal probability plot of residuals

6.3. Objective 3 discussion

The simple linear regression model including antecedent percent of bankfull, antecedent soil moisture, precipitation duration, and 24-hour precipitation intensity explains 40% of the variance of peak discharge, while the stepwise regression model including only antecedent soil moisture and 24-hour precipitation intensity explains 42% of the same variance. Considering the number of environmental and meteorological factors not included in this study, these correlations are reasonably good.

The positive sign of the β coefficients of all predictor variables except antecedent percent bankfull in the simple regression model were logical. The increased antecedent soil moisture suggests that as soil becomes more saturated, less precipitation will be able to infiltrate into the soil. This may result in increased surface flow. This relationship was observed in all of the models. The process for soil moisture is likely closely linked to the precipitation duration. As duration of a precipitation event increases, the soil moisture increases, consequently decreasing the amount of infiltration and increasing the surface flow. The final predictor, the 24-hour precipitation intensity, is closely related to precipitation duration, as shown in Table 1. As the intensity of the precipitation increases, the likeliness of the precipitation rate to surpass the infiltration rate increases.

Contrary to expectations, the β coefficient of antecedent percent bankfull was negative in the best simple regression model (Table 6, Equation 2). This illogically suggests that lower streamflows will produce a higher flash flood peak. Because the flash flood detection algorithm utilized in this study does not require a defined increase in discharge other than rising from baseflow to flood peak in less than 6 hours, the logical relationship would be for the antecedent percent bankfull to be relatively high before the flash flood events as a smaller increase in streamflow would be required to meet the conditions. While this unanticipated negative relationship shown in the regression model could be explored further, it is likely indicating that this variable at it stands is not suitable for this analysis. This problem could be addressed by analyzing antecedent bankfull closer to the start of the rise in discharge, perhaps on the preceding day instead of 4 days before the flood peak. As mentioned in Chapter II, De Castro et al.

(2013) used hydrologic variables stream height and velocity upstream of the target area to predict flash flood risk instead of using streamflow data collected at same site as the expected flash flood peak. This method may also provide a means for improving the antecedent streamflow component of this model. Further, using upstream data may provide more lead time before the flash flood peak, which is critical in flash flood forecasting.

The stepwise linear regression expanded the results of the simple linear regression. The weakness of antecedent percent bankfull as a predictor was reiterated in the stepwise linear regression, where it was the first predictor to be removed in an effort to reduce variables and improve the adjusted r^2. Similarly, the precipitation duration was highlighted for removal in the stepwise linear regression. Though the adjusted r^2 remained nearly the same after the removal of this variable and the RMSE increased only slightly, the F-statistic increased greatly. The *p*-value and F-statistic were the best when only the 24-hour precipitation intensity was included in the model. However, this final step worsened the adjusted r^2 and RMSE, leading to the re-accepting of the antecedent soil moisture into the model.

While the antecedent soil moisture was not the most influential variable in any of the regression analyses performed, it is logical that such a variable should still be included in any analysis of flash flooding. Further data that could be used to support this idea include soil maps and infiltration rate throughout the watersheds. The addition of this data may also highlight the link between soil moisture and precipitation intensity on this short timescale.

Overall, the 24-hour precipitation intensity was the strongest predictor variable of flash flood discharge. This is logical, as the value of this variable is increased either through the increase of precipitation amount or the decrease of precipitation duration. Both of these contributing factors relate to the soil moisture in the watershed. Increased precipitation while holding duration constant leads to the increased likeliness that soil is at saturation, which will cause excess water from precipitation to flow at the surface and contribute to an elevated streamflow. Decreased precipitation duration while holding precipitation amount constant will increase the rate of precipitation. Increasing the rate of precipitation increases the likeliness that the rate of infiltration will be exceeded, which will also cause excess water from precipitation to flow at the surface and contribute to an elevated streamflow. Doswell et al. (1996) highlight the importance of sustained heavy precipitation in developing flash flood conditions. The 24-hour precipitation accounts for this variable by representing both the amount of accumulated precipitation and the precipitation duration in one variable.

As highlighted in Doswell et al. (1996), accounting for the large number of additional factors influencing even the few variables here could lead to a staggering number of potential outcomes with only a few predictor variables. It is important to understand each variable and its interactions with other variables when attempting to forecast flash floods. Adhering to any fixed formula such as the regression model suggested in this thesis without careful attention to the error term could lead to "false alarms" or failure to predict flash floods (Doswell et al. 1996). The importance of understanding each variable and the hydrologic response of the watershed in question is

therefore reiterated. By including only a very small number of predictor variables, this study follows the philosophy of Doswell et al. (1996).

The coefficient of determination in this study (r^2 =0.42) is smaller than that of similar studies. Roland and Stuckey (2008) presented an r^2 >0.7 by stratifying floods by recurrence interval, and calculating the coefficients for the predictor variables for each of these flood sizes, using a similar methodology to that in Lumia et al. (2006). However, several fundamental differences between Lumia et al. (2006) and of Roland and Stuckey (2008) and this thesis prevent the replication of this method. Difference in sample size was the largest deterrent from adopting this method: While Roland and Stuckey (2008) had 322 stations with at least 10 years of data, this thesis focused on only two watersheds. Dividing flood peaks into smaller groups based on recurrence interval would fatally reduce the robustness of results with this sample size.

Lumia et al. (2006) and Roland and Stuckey (2008) also used 4 and 1-4 basin characteristics respectively as predictors for each of these groups and recurrence intervals, while thesis found that using only 2 predictors produced the highest r^2 values. Further, the predictors retained in this thesis differ drastically from those in Roland and Stuckey (2008), who did not retain any hydrometeorological or streamflow in their regression equations. However, the purposes of the studies vary greatly. While this thesis attempts to predict flash flood peaks for individual events for these two watersheds only, Roland and Stuckey (2008) provide the framework for predicting peak flows at selected recurrence intervals in the state of Pennsylvania. Adopting certain predictor variables from the methodology of Roland and Stuckey (2008) may account for differing basin

characteristics and improve the model for use in both the Esopus Creek and Neversink River watersheds. Lumia et al. (2006) found different basin characteristics to affect flood peak discharge in the separate regions in which the Esopus Creek and Neversink River are located. While drainage area, basin lag factor, and mean annual runoff were predictors of flood peak discharge in both watersheds, basin storage was included for the region containing the Esopus Creek watershed while the seasonal maximum snow depth was significant for the region containing the Neversink River watershed. This further motivates the inclusion of basin characteristics in future flash flood peak regression modelling.

7. CONCLUSIONS

This study utilized hydrologic and climatic data from 1987-2014 to further understand the specific hydrometeorologic conditions association with flash flooding in two watersheds contributing to the New York City Water Supply System. Preparedness for flash floods is important in any area but is particularly important in watersheds of this unfiltered water supply system, as the turbidity associated with flash flooding is not allowed by the Filtration Avoidance Determination. The rapid onset of these conditions requires thorough knowledge of the conditions associated with flash flooding in these watersheds in order to minimize the reaction time in decision making.

This study contributes to the body of literature in hydroclimatology and flash flood forecasting by combining analyses at the synoptic and local scales. Further, this study defines the influence of multiple variables on flash flood peak discharge through regression equations specific to these watersheds.

From 1987-2014, 32 flash flood peaks have been measured in these two watersheds, occurring on 23 days. These flash floods were all recorded as rising from just over bankfull to many times bankfull in less than 6 hours. Most of these flash floods occur in fall and winter, and in the absence of snow.

The geopotential height patterns in the study area can be distinguished into 17 unique synoptic types. Synoptic typing showed that a southwesterly flow at the 500 mb geopotential height is important in the development of flash flood conditions. 22 of the 23 flash flood event days in this study were associated with southwesterly flow on the

peak day and/or in the five days preceding the peak discharge. The type most commonly observed on flash flood days was Type 15. Types 7 and 13 were also frequently associated with flash flood peaks and the preceding days. The synoptic types in the 5 days before and including flood peak occurred in a significant departure from normal conditions, as shown through a bootstrapping methodology. Types 7, 13, and 15 were observed more frequently than normal in the days preceding flash flood events, while Type 10 was observed less frequently than normal in the same pre-flood days.

Peak discharge of flash floods can be best predicted in the Esopus Creek and Neversink River watersheds using antecedent soil moisture and 24-hour precipitation intensity. Antecedent soil moisture accounts for modified storage capacity of the soil in the days preceding a flash flood peak and the possibility of increased runoff. The 24-hour precipitation intensity accounts for both the accumulated precipitation in the day leading up to the flash flood as well as the time over which the precipitation occurred. These predictors explain 42% of the variance in flash flood peak discharge.

While these results are interesting and useful, further work is required to solidify these findings and expand to other watersheds. A number of limitations in the data and sample size were identified and motivate further research, including improvements to the flash flood detection algorithm, synoptic typing, and the multiple regression component. These limitations are discussed in section 7.1. Overall, this research serves as a multi-scale starting point for further research on flash flooding in watersheds of the New York City Water Supply System.

7.1 Limitations and future work

Several limitations restrict this work and motivate further study. These limitations and future prospects are listed below.

1. This study was constricted by a small sample size and short study period. The small sample size was not unexpected given that flash flood events fortunately are relatively rare. The use of 15-minute discharge data allows flash floods to be defined empirically of relying on reports, which may or may not be able to be confirmed or checked for accuracy. However, equipment to record at this frequency was not installed until the late 1980s, leaving fewer than 20 years of data at the time of this study. However, expanding to other watersheds in attempts to detect more flash floods may introduce differences in LU/ LC and add more noise to the regression model. Reproducing parts of this study in the 2020's may include more flash flood events, but may also include more unaccounted error through changes in LU/ LC. Future researchers may have to face these decisions in order to obtain a sample size large enough for statistical robustness.
2. The typical synoptic patterns may be better distinguished if more variables were included in the process. Multi-level classification with NCEP/NCAR variables is being tested for use with STT (Blair et al. 2013). Performing the classification using both surface and aloft pressure heights may reveal more about the transport of air masses in the region. Likewise, performing the classification using air

temperature or wind direction may also provide information on the environment of the type days.

3. The analysis of storm tracks has been found useful in other studies (e.g. Collins et al. 2014), and may apply to this study as well. Computing the backward trajectory of regional air masses at different heights may provide more information on the characteristics of the atmosphere on flash flood and preceding days. Further, this endeavor may reveal how long an air mass tracks over a region, and any interactions with local topography. Exploratory analysis using the NOAA Air Resources Laboratory Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) Model showed southwesterly flow at 500 mb pressure height coupled with clockwise advection of air masses from the mid-Atlantic and New England coastline at the surface. The result of this investigation were not quantified.
4. Spatially representative data would likely improve results over single-station data. The inclusion of radar data at multiple hours preceding the flash flood days may illuminate important characteristics in storm tracks and stalling over the watershed. Further, the inclusion of LU/ LC could offer additional information on the streamflow regimes.
5. The variables in this study are limited to those at a daily resolution or better within a relatively small radius of the watersheds. Other variables such as wind direction and convective available potential energy could be included to better differentiate and define each precipitation event.

Future work should integrate results of synoptic-scale and local-scale analyses. This could be done in a number of ways, such as by including the synoptic type for the flash flood and preceding days in the multiple linear regression. While the multi-scale purpose of this study is achieved through a two-step process, future work fusing the synoptic-scale and local-scale analyses and results may be effective in progressing the state of the knowledge in flash flood forecasting.

REFERENCES

Alonso-Pérez, S., E. Cuevas, and X. Querol, 2011: Objective identification of synoptic meteorological patterns favouring African dust intrusions into the marine boundary layer of the subtropical eastern north Atlantic region. Meteorol. Atmos. Phys., 113, 109–124, doi:10.1007/s00703-011-0150-z.

Arndt, D. S., M. O. Baringer, and M. R. Johnson, 2010: State of the Climate in 2009. Bulletin of the American Meteorological Society, 91, 1-222.

Banacos, P. C., and M. L. Ekster, 2010: The Association of the Elevated Mixed Layer with Significant Severe Weather Events in the Northeastern United States*. Weather and Forecasting, 25, 1082-1102.

Bettolli, M. L., O. C. Penalba, and W. M. Vargas, 2010: Synoptic weather types in the south of South America and their relationship to daily rainfall in the core crop-producing region in Argentina. Atlantic, 60, 37–48.

Blair, D., R. Smith, R. Dahni, 2013. Spatial Synoptic Typer Tools version 4.0. Accessed via http:// stt.uwinnipeg.ca/STT/HOME.html.

Bradbury, J. A., B. D. Keim, and C. P. Wake, 2002a: U. S. East Coast Trough Indices at 500 hPa and New England Winter Cliamte Variability. Journal of Climate, 15, 3509-3517.

Bradbury, J. A., S. L. Dingman, and B. D. Keim, 2002b: New England drought and relations with large scale atmospheric circulation patterns. Journal of the American Water Resources Association, 38, 1287-1299.

Brommer, D. M., R. S. Cerveny, and R. C. J. Balling, 2013: An analysis of intra-event precipitation variability for the United States (1980-2010). Physical Geography, 34, 456-470.

Broxton, P., P. A. Troch, M. Schaffner, C. Unkrich, and D. Goodrich, 2014: An all-season flash flood forecasting system for real-time operations. Bull. Am. Meteorol. Soc., 95, 399–407, doi:10.1175/BAMS-D-12-00212.1.

Buell, C. E., 1979: On the physical interpretation of empirical orthogonal functions. 6th Conference on Probability and Statistics in the Atmospheric Sciences, American Meteorological Society, 112–117.

Burns, D. A., J. Klaus, and M. R. McHale, 2007: Recent climate trends and implications for water resources in the Catskill Mountain region, New York, USA. Journal of Hydrology, 336, 155-170.

Collins, M. J., J. P. Kirk, J. Pettit, T. Arthur, M. S. Mccown, T. C. Peterson, and N. Tiffany, 2014: Annual floods in New England (USA) and Atlantic Canada : synoptic climatology and generating mechanisms. Phys. Geogr., 1–25, doi:10.1080/02723646.2014.888510.

Croft, P., and M. Schulman, 1989: A Five-Year Radar Climatology of Convective Precipitation for New Jersey. Int. J. Climatol., 9, 581–600.

Cuell, C., and B. Bonsal, 2009: An assessment of climatological synoptic typing by principal component analysis and kmeans clustering. Theor. Appl. Climatol., 98, 361–373, doi:10.1007/s00704-009-0119-8.

De Castro, J. T., G. M. Salistre, Jr., Y. C. Byun and B. D. Gerardo, 2013: Flash Flood Prediction Model based on Multiple Regression Analysis for Decision Support System. Proceedings of the World Congress on Engineering and Computer Science Vol II. WCECS 2013, 23-25 October, 2013, San Francisco, USA.

Degaetano, A. T., and C. M. Castellano, 2013: Recent and future changes in extreme rainfall in the Catskills region of New York. Annals of the New York Academy of Sciences.

Doswell, C. A., H. E. Brooks, and R. A. Maddox, 1996: Flash Flood Forecasting: An Ingredients-Based Methodology. Weather Forecast., 11, 560–581, doi:10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2.

Douglas, E. M., and C. A. Fairbank, 2011: Is Precipitation in Northern New England Becoming More Extreme? Statistical Analysis of Extreme Rainfall in Massachusetts, New Hampshire, and Maine and Updated Estimates of the 100-Year Storm. Journal of Hydrologic Engineering, 16, 203-217.

Erfani, R., and L. Chouinard, 2012: Automated synoptic typing of freezing rain events for hazard analysis. Atmos. Res., 111, 58–70.

Falconer, P. D., 1984: A Radar-Based Climatology of Thunderstorm Days across New York State. J. Clim., 23, 1115–1120.

Frei, A., R. L. Armstrong, M. P. Clark, and M. C. Serreze, 2002: Catskill Mountain water resources: Vulnerability, hydroclimatology, and climate-change sensitivity. Annals of the Association of American Geographers, 92, 203-224.

Galambosi, A., L. Duckstein, and I. Bogardi, 1996: Evaluation and analysis of daily atmospheric circulation patterns of the 500 HPA pressure field over the southwestern USA. Atmos. Res., 40, 49–76, doi:10.1016/0169-8095(95)00025-9.

Hart, R. E., and J. L. Evans, 2001: A climatology of the extratropical transition of Atlantic tropical cyclones. J. Clim., 14, 546–564, doi:10.1175/1520-0442(2001)014<0546:ACOTET>2.0.CO;2.

Hayhoe, K., and Coauthors, 2006: Past and future changes in climate and hydrological indicators in the US Northeast. Climate Dynamics, 28, 381-407.

Hubeny, J. B., J. W. King, and M. Reddina, 2011: Northeast US precipitation variability and North American climate teleconnections interpreted from late Holocene varved sediments. Proceedings of the National Academy of Sciences, 108, 17895–17900.

Hurlbut, M. M., and A. E. Cohen, 2014: Environments of Northeast U.S. Severe Thunderstorm Events from 1999 to 2009. Weather and Forecasting, 29, 3-22.

Jiang, N., K. Cheung, K. Luo, P. J. Beggs, and W. Zhou, 2012: On two different objective procedures for classifying synoptic weather types over east Australia. Int. J. Climatol., 32, 1475–1494. 10.1002/joc.2373.

Johns, R., 1984: A Synoptic Climatology of Northwest-Flow Severe Weather Outbreaks. Part II. Meteorological Parameters and Synoptic Patterns. Mon. Weather Rev., 112, 449–464.

Kalkstein, L. S., and P. Corrigan, 1986: A Synoptic Climatological Approach For Geographical Analysis: Assessment of Sulfur Dioxide Concentrations. Ann. Assoc. Am. Geogr., 76, 381–395, doi:10.1111/j.1467-8306.1986.tb00126.x.

——, G. Tan, and J. A. Skindlov, 1987: An Evaluation of Three Clustering Procedures for Use in Synoptic Climatological Classification. J. Clim. Appl. Meteorol., 26, 717–730, doi:10.1175/1520-0450(1987)026<0717:AEOTCP>2.0.CO;2.

Keim, B. D., R. A. Muller, and G. W. Stone, 2007: Spatiotemporal patterns and return periods of tropical storm and hurricane strikes from Texas to Maine. J. Clim., 20, 3498–3509, doi:10.1175/JCLI4187.1.

Kelsch, M., 2001: Hydrometeorological Characteristics of Flash Floods. Coping With Flash Floods SE - 18, E. Gruntfest and J. Handmer, Eds., Vol. 77 of NATO Science Series, Springer Netherlands, 181–193 http://dx.doi.org/10.1007/978-94-010-0918-8_18.

Kidson, J. W., 1994: An Automated Procedure for the Identification of Synoptic Types Applied to the New Zealand Region. Int. J. Climatol., 14, 711–721, doi:10.1002/joc.3370140702.

Kobiyama, M., and R. F. Goerl, 2007: Quantitative method to distinguish flood and flash flood as disaster. Hydrol. Res. Lett., 1, 11–14.

LaPenta, K. D., L. F. Bosart, T. J. Galarneau, Jr., and M. J. Dickinson, 2005: A Multiscale Examination of the 31 May 1998 Mechanicville, New York, Tornado. Weather and Forecasting, 20, 494-513.

Leathers, D. J., A. J. Grundstein, and A. W. Ellis, 2000: Growing season moisture deficits across the northeastern United States. Climate Research, 14, 43-55.

Legates, D. R., 1991: The effect of domain shape on principal components analyses. Int. J. Climatol., 11, 135–146, doi:10.1002/joc.3370110203.

Li, H., and B. Colle, 2014: Multidecadal Changes in the Frequency and Ambient Conditions of Warm Season Convective Storms over the Northeastern United States. J. Clim., 27, 7285–7300, doi:10.1175/JCLI-D-13-00785.1.

Lumia, R., D. A. Freehafer, and M. J. Smith, 2006: Magnitude and frequency of floods in New York: U.S. Geological Survey Scientific Investigations Report.

Maddox, R. A., C. F. Chappell, and L. R. Hoxit, 1979: Synoptic and Meso-alpha Scale Aspects of Flash Flood Events. Bull. Am. Meteorol. Soc., 60, 115–123.

Magilligan, F. J., and B. E. Graber, 1996: Hydroclimatological and geomorphic controls on the timing and spatial variability of floods in New England, USA. J. Hydrol., 178, 159–180, doi:10.1016/0022-1694(95)02807-2.

Matonse, A. H., D. C. Pierson, A. Frei, M. S. Zion, E. M. Schneiderman, A. Anandhi, R. Mukundan, and S. M. Pradhanang, 2011: Effects of changes in snow pattern and the timing of runoff on NYC water supply system. Hydrol. Process., 25, 3278–3288, doi:10.1002/hyp.8121.

Matonse, A. H., and A. Frei, 2013: A Seasonal Shift in the Frequency of Extreme Hydrological Events in Southern New York State. Journal of Climate, 26, 9577-9593.

Matonse, A. H., D. C. Pierson, A. Frei, M. S. Zion, A. Anandhi, E. Schneiderman, and B. Wright, 2012: Investigating the impact of climate change on New York City's primary water supply. Climatic Change, 116, 437-456.

Miller, S. J. and D. Davis, 2003. Optimizing Catskill Mountain Regional Bankfull Discharge and Hydraulic Geometry Relationships. American Water Resources Association 2003 International Congress, 29 June- 2 July, 2003, New York, USA.

Moore, M. V., and Coauthors, 1997: Potential effects of climate change on freshwater ecosystems of the New England/ Mid-Atlantic region. Hydrological Processes, 11, 925-947.

Murray, J. C., and B. A. Colle, 2011: The Spatial and Temporal Variability of Convective Storms over the Northeast United States during the Warm Season. Monthly Weather Review, 139, 992-1012.

NOAA Central Library Data Imaging Project. Published by US Department of Commerce / NOAA Central Library. Silver Spring, Maryland, USA.

NOAA National Weather Service / Weather Prediction Center. Archive of the National Forecast Chart. Published by US Department of Commerce / National Oceanic and Atmospheric Administration / National Weather Service. Silver Spring, Maryland, USA.

NYC-DEP: New York City Department of Environmental Protection, 2015. Drinking Water. The City of New York. Accessed from http://www.nyc.gov/html/dep/html/drinking_water/index.shtml.

Pradhanang, S. M., and Coauthors, 2013: Streamflow Responses to Climate Change: Analysis of Hydrologic Indicators in a New York City Water Supply Watershed. JAWRA Journal of the American Water Resources Association, 49, 1308-1326.

Ralston, D. K., J. C. Warner, W. R. Geyer, and G. R. Wall, 2013: Sediment transport due to extreme events: The Hudson River estuary after tropical storms Irene and Lee. Geophysical Research Letters, 40, 5451-5455.

Reed, S., J. Schaake, and Z. Zhang, 2007: A distributed hydrologic model and threshold frequency-based method for flash flood forecasting at ungauged locations. J. Hydrol., 337, 402–420, doi:10.1016/j.jhydrol.2007.02.015.

Richman, M. B., 1986: Rotation of principal components. Int. J. Climatol., 6, 293–335, doi:10.1002/joc.3370060305. http:// dx.doi.org/10.1002/joc.337006030.

Roland, M. A. and M. H. Stuckey, 2008. Regression Equations for Estimating Flood Flows at Selected Recurrence Intervals for Ungaged Streams in Pennsylvania. U.S. Geological Survey Scientific Investigations Report 2008-5102, 57 p.

Rosgen, D. L., 1994: A classification of natural rivers. CATENA, 22, 169–199, doi:10.1016/0341-8162(94)90001-9.

Samal, N. R., A. H. Matonse, R. Mukundan, M. S. Zion, D. C. Pierson, R. K. Gelda, and E. M. Schneiderman, 2013: Modelling potential effects of climate change on winter turbidity loading in the Ashokan Reservoir, NY. Hydrological Processes, 27, 3061-3074.

Seager, R., N. Pederson, Y. Kushnir, J. Nakamura, and S. Jurburg, 2012: The 1960s Drought and the Subsequent Shift to a Wetter Climate in the Catskill Mountains Region of the New York City Watershed*. Journal of Climate, 25, 6721-6742.

Thibeault, J. M., and A. Seth, 2014: A Framework for Evaluating Model Credibility for Warm-Season Precipitation in Northeastern North America: A Case Study of CMIP5 Simulations and Projections. Journal of Climate, 27, 493-510.

United States Environmental Protection Agency, 2007: New York City Filtration Avoidance Determination.

van Vliet, M. T. H., J. R. Yearsley, F. Ludwig, S. Vögele, D. P. Lettenmaier, and P. Kabat, 2012: Vulnerability of US and European electricity supply to climate change. Nature Climate Change, 2, 676-681. doi:10.1038/nclimate1546.

Wasula, A. C., L. F. Bosart, and K. D. LaPenta, 2002: The Influence of Terrain on the Severe Weather Distribution across Interior Eastern New York and Western New England. Weather Forecast., 17, 1277–1289, doi:10.1175/1520-0434(2002)017<1277:TIOTOT>2.0.CO;2.

Yarnal, B., 1993: Synoptic Climatology in Environmental Analysis. Belhaven Press, London,

APPENDIX

MULTIPLE LINEAR REGRESSION TABLES: STATISTICAL OUTPUT

Linear regression model:
$y \sim 1 + x_1 + x_2 + x_3 + x_4$

Table A-1. Group A, variable combination 1

Variable combination Group A, 1: Both datasets, both watersheds, raw x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	-32283.00	17295.00	-1.87	0.08
x_1	-49.67	226.73	-0.22	0.83
x_2	38.54	27.20	1.42	0.17
x_3	293.79	371.80	0.79	0.44
x_4	1997.50	514.15	3.89	<0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 26, Error degrees of freedom: 21,
Root Mean Squared Error: 4800
R-squared: 0.50, Adjusted R-Squared: 0.4
F-statistic vs. constant model: 5.17, p-value = <0.01

Table A-2. Group A, variable combination 2

Variable combination Group A, 2: Both datasets, both watersheds, standardized x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	7459.60	941.96	7.92	<0.01
x_1	-268.26	1224.60	-0.22	0.83
x_2	1440.60	1016.70	1.42	0.17
x_3	910.94	1152.80	0.79	0.44
x_4	4201.50	1081.50	3.89	<0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 26, Error degrees of freedom: 21
Root Mean Squared Error: 4.8e+03
R-squared: 0.50, Adjusted R-Squared 0.4
F-statistic vs. constant model: 5.17, p-value = <0.01

Table A-3. Group A, variable combination 3

Variable combination Group A, 3: Both datasets, both watersheds, raw x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	2.38	0.76	3.15	<0.01
x_1	<0.01	0.01	-0.31	0.76
x_2	<0.01	<0.01	1.01	0.32
x_3	0.01	0.02	0.84	0.41
x_4	0.08	0.02	3.57	<0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 26, Error degrees of freedom: 21
Root Mean Squared Error: 0.21
R-squared: 0.45, Adjusted R-Squared 0.35
F-statistic vs. constant model: 4.33, p-value = 0.01

Table A-4. Group A, variable combination 4

Variable combination Group A, 4: Both datasets, both watersheds, standardized x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	3.78	0.04	91.87	<0.01
x_1	-0.02	0.05	-0.31	0.76
x_2	0.05	0.04	1.01	0.32
x_3	0.04	0.05	0.84	0.41
x_4	0.17	0.05	3.57	<0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 26, Error degrees of freedom: 21
Root Mean Squared Error: 0.21
R-squared: 0.45, Adjusted R-Squared 0.35
F-statistic vs. constant model: 4.33, p-value = 0.01

The sample size was reduced to 13 flash flood events when the floods in the Esopus Creek watershed were removed to isolate conditions in the Neversink River watershed in Group B. No predictors or models were significant in this group, and the F-statistics were considerably lower than when both watersheds were analyzed together. Similar to Group A, the β coefficient for the antecedent percent bankfull was negative. Overall, the results when using the log discharge (r^2 = 0.07, F = 1.22) were nearly identical to the raw discharge data (r^2 = 0.06, F = 1.2). Both models were not statistically significant.

Table A-5. Group B, variable combination 1

Variable combination Group B, 1: Both datasets, Neversink River only, raw x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	-17628.00	24600.00	-0.72	0.49
x_1	-59.56	391.01	-0.15	0.88
x_2	18.62	40.57	0.46	0.66
x_3	500.58	817.03	0.61	0.56
x_4	1437.60	809.82	1.78	0.11

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 5.05e+03
R-squared: 0.38, Adjusted R-Squared 0.06
F-statistic vs. constant model: 1.2, p-value = 0.38

Table A-6. Group B, variable combination 2

Variable combination Group B, 2: Both datasets, Neversink River only, standardized x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	8441.50	1401.10	6.03	<0.01
x_1	-334.04	2193.10	-0.15	0.88
x_2	728.07	1586.60	0.46	0.66
x_3	1349.60	2202.80	0.61	0.56
x_4	3025.60	1704.40	1.78	0.11

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 5.05e+03
R-squared: 0.38, Adjusted R-Squared 0.06
F-statistic vs. constant model: 1.2, p-value = 0.38

Table A-7. Group B, variable combination 3

Variable combination Group B, 3: Both datasets, Neversink River only, raw x, log y

	Estimated coefficients			
	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	2.77	1.02	2.72	0.03
x_1	<0.01	0.02	-0.16	0.88
x_2	<0.01	<0.01	0.46	0.66
x_3	0.02	0.03	0.69	0.51
x_4	0.06	0.03	1.75	0.12

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 0.209
R-squared: 0.38, Adjusted R-Squared 0.07
F-statistic vs. constant model: 1.22, p-value = 0.38

Table A-8. Group B, variable combination 4

Variable combination Group B, 4: Both datasets, Neversink River only, standardized x, log y

	Estimated coefficients			
	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	3.87	0.06	66.73	<0.01
x_1	-0.01	0.09	-0.16	0.88
x_2	0.03	0.07	0.46	0.66
x_3	0.06	0.09	0.69	0.51
x_4	0.12	0.07	1.75	0.12

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 0.21
R-squared: 0.38, Adjusted R-Squared 0.07
F-statistic vs. constant model: 1.22, p-value = 0.38

The sample size was 13 flash flood events when the Esopus Creek peaks were analyzed separately from those in the Neversink River watershed in Group C. While the 24-hour precipitation intensity was a statistically significant variable in all models in this group, none of the final regression models overall were statistically significant, with p = 0.06 when using raw discharge data and p = 0.07 when using the log discharge data. The r^2 and adjusted r^2 were larger in this watershed than in the Neversink River (r^2 = 0.64 and adjusted r^2 = 0.46 with raw discharge). Unlike in Group B, the models were slightly improved when using the raw discharge data instead of the log discharge data.

Table A-9. Group C, variable combination 1

Variable combination Group C, 1: Both datasets, Esopus Creek only, raw x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	-44803.00	29304.00	-1.53	0.16
x_1	83.47	354.75	0.24	0.82
x_2	49.23	45.99	1.07	0.32
x_3	363.23	508.54	0.71	0.50
x_4	2599.90	795.39	3.27	0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 5.23e+03
R-squared: 0.64, Adjusted R-Squared 0.46
F-statistic vs. constant model: 3.57, p-value = 0.06

Table A-10. Group C, variable combination 2

Variable combination Group C, 2: Both datasets, Esopus Creek only, standardized x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	6477.70	1451.90	4.46	<0.01
x_1	451.85	1920.40	0.24	0.82
x_2	1828.60	1708.50	1.07	0.32
x_3	1191.30	1667.90	0.71	0.50
x_4	5546.80	1697.00	3.27	0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 5.23e+03
R-squared: 0.64, Adjusted R-Squared 0.46
F-statistic vs. constant model: 3.57, p-value = 0.06

Table A-11. Group C, variable combination 3

Variable combination Group C, 3: Both datasets, Esopus Creek only, raw x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	1.91	1.19	1.60	0.15
x_1	<0.01	0.01	0.24	0.81
x_2	<0.01	<0.01	0.73	0.48
x_3	0.02	0.02	1.21	0.26
x_4	0.10	0.03	3.10	0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 0.21
R-squared: 0.61, Adjusted R-Squared 0.42
F-statistic vs. constant model: 3.13, p-value = 0.08

Table A-12. Group C, variable combination 4

Variable combination Group C, 4: Both datasets, Esopus Creek only, standardized x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	3.69	0.06	62.60	<0.01
x_1	0.02	0.08	0.24	0.81
x_2	0.05	0.07	0.73	0.48
x_3	0.08	0.07	1.21	0.26
x_4	0.21	0.07	3.10	0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 13, Error degrees of freedom: 8
Root Mean Squared Error: 0.21
R-squared: 0.61, Adjusted R-Squared 0.42
F-statistic vs. constant model: 3.13, p-value = 0.08

Opting to use only precipitation data from Claryville, NY instead of supplementing with data from the Mongaup Valley, NY station reduced the sample size to only 17 flash floods, when both watersheds are considered together in Group D. The β coefficient of antecedent percent bankfull was negative in this group, similar to Groups A and B. None of the models were statistically significant.

Table A-13. Group D, variable combination 1

Variable combination Group D, 1: Claryville, NY data only, both watersheds, raw x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	-11333.00	12889.00	-0.88	0.40
x_1	-82.08	150.66	-0.54	0.60
x_2	20.97	18.20	1.15	0.27
x_3	89.23	256.61	0.35	0.73
x_4	490.31	376.71	1.30	0.22

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 17, Error degrees of freedom: 12
Root Mean Squared Error: 2.31e+03
R-squared: 0.21, Adjusted R-Squared -0.05
F-statistic vs. constant model: 0.82, p-value = 0.54

Table A-14. Group D, variable combination 2

Variable combination Group D, 2: Claryville, NY data only, both watersheds, standardized x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	5251.20	561.36	9.35	<0.01
x_1	-364.99	670.01	-0.54	0.60
x_2	692.17	600.71	1.15	0.27
x_3	240.01	690.26	0.35	0.73
x_4	846.45	650.34	1.30	0.22

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 17, Error degrees of freedom: 12
Root Mean Squared Error: 2.31e+03
R-squared: 0.214, Adjusted R-Squared -0.0483
F-statistic vs. constant model: 0.816, p-value = 0.54

Table A-15. Group D, variable combination 3

Variable combination Group D, 3: Claryville, NY data only, both watersheds, raw x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	2.73	1.00	2.74	0.02
x_1	-0.01	0.01	-0.55	0.59
x_2	<0.01	<0.01	0.81	0.44
x_3	0.01	0.02	0.28	0.78
x_4	0.04	0.03	1.25	0.23

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 17, Error degrees of freedom: 12
Root Mean Squared Error: 0.179
R-squared: 0.184, Adjusted R-Squared -0.0877
F-statistic vs. constant model: 0.677, p-value = 0.62

Table A-16. Group D, variable combination 4

Variable combination Group D, 4: Claryville, NY data only, both watersheds, standardized x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	3.69	0.04	84.99	<0.01
x_1	-0.03	0.05	-0.55	0.59
x_2	0.04	0.05	0.81	0.44
x_3	0.01	0.05	0.28	0.78
x_4	0.06	0.05	1.25	0.23

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 17, Error degrees of freedom: 12
Root Mean Squared Error: 0.18
R-squared: 0.18, Adjusted R-Squared -0.09
F-statistic vs. constant model: 0.68, p-value = 0.62

Treating each watershed separately in Group E further reduces the number of flash floods to only 9. As a result, none of the models are statistically significant.

Table A-17. Group E, variable combination 1

Variable combination Group E, 1: Claryville, NY data only, Neversink River only, raw x, raw y

	Estimated coefficients			
	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	-29135.00	14405.00	-2.02	0.11
x_1	140.93	348.95	0.40	0.71
x_2	41.00	19.51	2.10	0.10
x_3	525.82	380.36	1.38	0.24
x_4	670.25	426.11	1.57	0.19

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 9, Error degrees of freedom: 4
Root Mean Squared Error: 1.89e+03
R-squared: 0.61, Adjusted R-Squared 0.23
F-statistic vs. constant model: 1.59, p-value = 0.33

Table A-18. Group E, variable combination 2

Variable combination Group E, 2: Claryville, NY data only, Neversink River only, standarized x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	6711.10	630.95	10.64	<0.01
x_1	295.38	731.37	0.40	0.71
x_2	1466.20	697.66	2.10	0.10
x_3	1118.90	809.35	1.38	0.24
x_4	1225.20	778.89	1.57	0.19

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 9, Error degrees of freedom: 4
Root Mean Squared Error: 1.89e+03
R-squared: 0.61, Adjusted R-Squared 0.23
F-statistic vs. constant model: 1.59, p-value = 0.33

Table A-19. Group E, variable combination 3

Variable combination Group E, 3: Claryville, NY data only, Neversink River only, raw x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	1.75	0.94	1.87	0.13
x_1	0.01	0.02	0.58	0.59
x_2	<0.01	<0.01	1.75	0.16
x_3	0.03	0.02	1.38	0.24
x_4	0.04	0.03	1.50	0.21

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 9, Error degrees of freedom: 4
Root Mean Squared Error: 0.12
R-squared: 0.56, Adjusted R-Squared 0.12
F-statistic vs. constant model: 1.27, p-value = 0.41

Table A-20. Group E, variable combination 4

Variable combination Group E, 4: Claryville, NY data only, Neversink River only, standardized x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	3.81	0.04	92.87	<0.01
x_1	0.03	0.05	0.58	0.59
x_2	0.08	0.05	1.75	0.16
x_3	0.07	0.05	1.38	0.24
x_4	0.08	0.05	1.50	0.21

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 9, Error degrees of freedom: 4
Root Mean Squared Error: 0.12
R-squared: 0.56, Adjusted R-Squared 0.12
F-statistic vs. constant model: 1.27, p-value = 0.41

The sample size of Group F was the smallest, with only 8 flash floods in the Esopus Creek watershed. None of the models were statistically significant.

Table A-21. Group F, variable combination 1

Variable combination Group F, 1: Claryville, NY data only, Esopus Creek only, raw x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	-1853.60	8816.50	-0.21	0.85
x_1	-7.92	75.00	-0.11	0.92
x_2	7.79	12.99	0.60	0.59
x_3	89.46	140.98	0.63	0.57
x_4	-73.80	268.73	-0.27	0.80

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 8, Error degrees of freedom: 3
Root Mean Squared Error: 1e+03
R-squared: 0.22, Adjusted R-Squared -0.81
F-statistic vs. constant model: 0.22, p-value = 0.91

Table A-22. Group F, variable combination 2

Variable combination Group F, 2: Claryville, NY data only, Esopus Creek only, standarized x, raw y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	3608.80	354.94	10.17	<0.01
x_1	-48.37	458.07	-0.11	0.92
x_2	240.72	401.44	0.60	0.59
x_3	294.54	464.13	0.63	0.57
x_4	-108.32	394.42	-0.27	0.80

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 8, Error degrees of freedom: 3
Root Mean Squared Error: 1e+03
R-squared: 0.22, Adjusted R-Squared -0.81
F-statistic vs. constant model: 0.21, p-value = 0.91

Table A-23. Group F, variable combination 3

Variable combination Group F, 3: Claryville, NY data only, Esopus Creek only, raw x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	2.89	1.00	2.88	0.06
x_1	<0.01	0.01	<0.01	1.00
x_2	<0.01	<0.01	0.60	0.59
x_3	0.01	0.02	0.68	0.54
x_4	<0.01	0.03	-0.16	0.89

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 8, Error degrees of freedom: 3
Root Mean Squared Error: 0.114
R-squared: 0.21, Adjusted R-Squared -0.84
F-statistic vs. constant model: 0.20, p-value = 0.92

Table A-24. Group F, variable combination 4

Variable combination Group F, 4: Claryville, NY data only, Esopus Creek only, standardized x, log y

Estimated coefficients

	Estimate	Standard Error	t-Statistic	p-value
(Intercept)	3.55	0.04	88.01	<0.01
x_1	<0.01	0.05	<0.01	1.00
x_2	0.03	0.05	0.60	0.59
x_3	0.04	0.05	0.68	0.54
x_4	-0.01	0.04	-0.16	0.89

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 8, Error degrees of freedom: 3
Root Mean Squared Error: 0.114
R-squared: 0.21, Adjusted R-Squared -0.84
F-statistic vs. constant model: 0.20, p-value = 0.92

Table A-25. Summary statistics of predictors in stepwise regression

Summary statistics of predictors in stepwise regression
Both datasets, both watersheds, standardized x, raw y

	Estimate (β)	Standard error	*t*-statistic	*p*-value
x_1	284.84	1187.04	0.240	0.813
x_2	1369.60	942.59	1.453	0.160
x_3	933.95	985.61	0.947	0.353
x_4	4246.57	985.61	4.309	0.0003

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 26

Table A-26. Results of each step of the stepwise linear regression. Step 7* was considered the optimal combination of predictor variables.

Stepwise Regression results

Step	Predictors included	RMSE	r^2	Adjusted r^2	F-statistic	*p*-value
1		0	0	0	N/A	N/A
2	x_1	6183.6	0.05	0.01	1.15	0.29
3	x_1, x_2	6065.13	0.12	0.04	1.57	0.23
4	x_1, x_2, x_3	6152.08	0.13	0.02	1.14	0.36
5	x_1, x_2, x_3, x_4	4803.09	0.50	0.40	5.17	0.01
6	x_2, x_3, x_4	4698.02	0.50	0.43	7.19	<0.01
7*	x_2, x_4	4714.69	0.47	0.42	10.13	<0.01
8	x_4	4799.93	0.43	0.40	17.74	<0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity
Number of observations: 26

Table A-27. Comparison of standard multiple linear regression and stepwise linear regression

Comparison of Standard and Stepwise Linear Regression Results

Model	Predictors included	RMSE	r^2	Adjusted r^2	F	p
Standard Multiple Linear Regression	x_1, x_2, x_3, x_4	4800.0	0.50	0.40	5.17	< 0.01
Stepwise Linear Regression	x_2, x_4	4714.69	0.47	0.42	10.13	< 0.01

x_1 = Antecedent percent bankfull, x_2 = Antecedent soil moisture
x_3 = Precipitation duration, x_4 = 24-hour precipitation intensity

The differences in significant predictor variables between groups is worthy of further consideration. Limitations in the input data could be a contributing factor to these differences. If so, these results would suggest that the hourly precipitation data from one or two stations cannot be used to predict flash flood peak discharge beyond some distance away from the precipitation gauge, or that the temporal component of the time-averaged variables such as soil moisture is inappropriate for the analysis. Steps in resolving these limitations, such as by using more spatially sensitive data such as radar or assessing the most appropriate lead time for antecedent conditions, would likely yield stronger results.