Mixture Distributions and the Hydroclimatology of Extreme Rainfall and Flooding in the Eastern United States

JAMES A. SMITH, GABRIELE VILLARINI, AND MARY LYNN BAECK

Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

(Manuscript received 25 November 2009, in final form 20 August 2010)

ABSTRACT

Flooding in the eastern United States reflects a mixture of flood-generating mechanisms, with landfalling tropical cyclones and extratropical systems playing central roles. The authors examine the climatology of heavy rainfall and flood magnitudes for the eastern United States through analyses of long-duration records of flood peaks and maximum daily rainfall series. Spatial heterogeneities in flood peak distributions due to orographic precipitation mechanisms in mountainous terrain, coastal circulations near land–ocean boundaries, and urbanization impacts on regional climate are central elements of flood peak distributions. Lagrangian analyses of rainfall distribution and storm evolution are presented for flood events in the eastern United States and used to motivate new directions for stochastic modeling of rainfall. Tropical cyclones are an important element of the upper tail of flood peak distributions throughout the eastern United States, but their relative importance varies widely, and abruptly, in space over the region. Nonstationarities and long-term persistence of flood peak and rainfall distributions are examined from the perspective of the impacts of human-induced climate change on flood-generating mechanisms. Analyses of flood frequency for the eastern United States, which are based on observations from a dense network of U.S. Geological Survey (USGS) stream gauging stations, provide insights into emerging problems in flood science.

1. Introduction

Flooding in the eastern United States reflects a mixture of flood-generating mechanisms, with landfalling tropical cyclones and extratropical systems playing central roles (Hirschboeck 1987, 1988; Waylen 1991; Smith et al. 2010; Villarini and Smith 2010). We examine the climatology of heavy rainfall and flood magnitudes for the eastern United States through analyses of long-duration records of flood peaks and annual maximum daily rainfall series (section 2). Lagrangian analyses of rainfall distribution and storm evolution are presented for flood events in the eastern United States (section 3) and used to motivate new directions for stochastic modeling of rainfall and characterizing spatial extremes of flood magnitudes (section 4).

Spatial heterogeneities in flood peak distributions due to orographic precipitation mechanisms in mountainous terrain, coastal circulations near land–ocean boundaries, and urbanization impacts on regional climate are central features of flood peak distributions for the eastern United States, and for many other settings around the world. Orographic mechanisms in the eastern United States associated with heavy rainfall from tropical cyclones, winter–spring extratropical systems, and warm-season thunderstorms are diverse and in many settings poorly understood (e.g., Barros and Kuligowski 1998; Nykanen 2008; Smith et al. 1996; Pontrelli et al. 1999).

Flood frequency techniques have focused on analyses for single stations, with lesser attention given to multivariate procedures. Increasingly, there is a need for a broader examination of spatial structure of flood extremes. There are difficulties in extending methods based on univariate and multivariate extreme value theory to a broader spatial setting. Availability of stream gauging networks from the U.S. Geological Survey (USGS) with records of discharge time series covering multiple decades (see section 3 for examples) provides an important data resource for examining spatial structure of flood extremes. Closely linked with the characterization of spatial extremes of flood magnitude is the examination of scaling properties of flood peaks, which are of
interest—both from the flood science perspective (e.g., Gupta et al. 1994; Gupta and Dawdy 1995; Gupta et al. 1996; Robinson and Sivapalan 1997) and from the perspective of regional flood frequency estimation (e.g., Hosking and Wallis 1997).

Nonstationarities and long-term persistence of flood peak and rainfall distributions are examined from the perspective of the impacts of human-induced climate change on flood-generating mechanisms. Annual flood peak time series are stationary, provided that their distribution is invariant to translations of time (Brillinger 1975), implying that the time series is free of trends, shifts, or periodicity. Villarini et al. (2009) performed an analysis of 50 USGS stations with a record of at least 100 yr within the continental United States and found that change points were a more common source of non-stationarity than time trends (see also Changnon and Kunkel 1995; Lins and Slack 1999; Groisman et al. 2004; McCabe and Wolock 2002; Small et al. 2006; Juckem et al. 2008; Villarini et al. 2011; Villarini and Smith 2010). Flood records in the United States will play an important role in detecting changing frequency of floods associated with human-induced climate change, because of the exceptional observing record that has been developed and maintained by the USGS (USGS 1998; Blanchard 2007).

Analyses of flood frequency for the eastern United States provide insights into emerging problems in flood science due to the range of flood mechanisms in the region and the extraordinary density of long-term stream gauging records maintained by the USGS. Flood records in the United States reflect the hydrologic response associated with a broad range of hydrometeorological settings, including storms associated with tropical cyclones, organized thunderstorm systems, and winter–spring extratropical systems (Miller 1990).

2. Mixture distributions—Flood hydroclimatology

In this section, we examine the climatology of floods and heavy rainfall, based on analyses of observations from long-duration records of flood peaks and maximum daily rainfall for the eastern United States. The region has an exceptionally dense network of stream gauging stations maintained by the USGS (e.g., Villarini and Smith 2010), including 572 stream gauging stations with at least 75 yr of record. We also utilize annual maximum daily rainfall from 38 stations in the central portion of the eastern United States with record lengths of more than 75 yr.

Flood peak occurrences in the eastern United States can be viewed in terms of mixtures of generating mechanisms associated with landfalling tropical cyclones and extratropical systems (Fig. 1). There is striking spatial
heterogeneity in the occurrence of tropical cyclone flood peaks (Fig. 1a), associated with mountainous terrain, land–ocean boundaries, and the north-to-south-varying climatology of extratropical transition (Hart and Evans 2001). Tropical cyclone flood peaks are defined as annual flood peaks for which a tropical cyclone passes within 500 km of the stream gauge within a 2-week window of the annual peak centered on the day of the peak. The Hurricane Database (HURDAT; Jarvinen et al. 1984; Neumann et al. 1993) of tropical cyclone tracks is used for determining tropical cyclone flood peaks [see Villarini and Smith (2010) for additional details]. There is an abrupt transition in tropical cyclone flood peaks across the Appalachian Mountain region, with a local maximum of more than 15% east of the Appalachians in the region extending from South Carolina through Maryland.

We also find large variability in counts of tropical cyclone floods per year. In Fig. 2, we show the fraction of annual flood peaks from the long-term USGS stream gauging stations that were produced by tropical cyclones. Analyses are based on the 572 stream gauging stations for the eastern United States (the number of stations available in a given year varies, so we present the analyses as the percentage of stations active for the given year). In 2004, tropical cyclones accounted for more than 70% of annual flood peaks (Fig. 2). During 14 yr, there were no annual flood peaks produced by tropical cyclones. Clustering of Atlantic basin tropical cyclone counts (see Villarini et al. 2010) is likely an important element of interannual variability in tropical cyclone floods for the eastern United States. Long-term trends in flood peaks associated with increasing frequency of Atlantic basin tropical cyclones is a major concern for flood hazards in the eastern United States (e.g., Landsea et al. 2006; Holland and Webster 2007; Knutson et al. 2007; Emanuel et al. 2008; Vecchi et al. 2008). There is not, however, significant evidence of increasing flood peaks from tropical cyclones (Fig. 2). The percent of the 572 USGS stations with annual flood peaks produced by tropical cyclones exhibits large interannual variability, but there is no significant long-term trend.

Winter–spring extratropical systems account for a large fraction of eastern U.S. flood peaks, especially in the northeastern and southeastern United States (Fig. 1b). We use the frequency of March–April flood peaks as a surrogate for winter–spring extratropical flood peaks. A combination of snowmelt and rain on snow determines the local maximum in flood peak occurrence in the northeastern United States (Fig. 1b), where March–April peaks account for up to 60% of annual peaks. Organized thunderstorm systems embedded in winter–spring extratropical systems, often associated with severe weather, are important flood agents in the southeastern United States (Fig. 1b). March–April peaks account for more than 50% of annual flood peaks in south Georgia, north Florida, and southeastern Alabama (Fig. 1b).

Extratropical systems during the fall, which can produce comparable rainfall to winter–spring systems, are much less important as flood agents in the eastern United States (Fig. 1c) because of the climatological minimum in antecedent soil moisture. We use October–November peaks (that are not tropical cyclones) as a surrogate for fall extratropical flood peaks. October–November flood peaks, that are not tropical cyclone events, account for less than 15% of annual flood peaks throughout the eastern United States. Maxima in occurrence of October–November flood peaks are concentrated along the Appalachian region, suggesting that orographic precipitation mechanisms play a significant role in these flood events.

We analyzed peaks-over-threshold (POT) flood peaks from three “large” drainage basins in the central portion of the eastern United States with stream gauging records of more than 100 years: the Susquehanna [drainage area of 25 498 km$^2$, or 9960 miles$^2$ (mi$^2$); USGS ID 01536500], the Potomac [24 706 km$^2$ (9651 mi$^2$); USGS ID 01638500], and the James [5307 km$^2$ (2073 mi$^2$); USGS ID 02019500]. For these analyses we selected flood peaks from mean daily discharge data to provide two events per year, on average (Fig. 3). For the three stations, there is a pronounced seasonal maximum in flood occurrence during the March–April period, with the most pronounced seasonal peak in the most northern of the three basins, the Susquehanna. There is a second fall maximum, reflecting the contributions of both tropical and extratropical systems. Counts per year of flood peaks in the Potomac have an index of dispersion (variance divided by the mean of

![Fig. 2. Percentage of annual flood peaks in the eastern United States produced by tropical cyclone floods.](image-url)
annual counts) greater than 1 (Fig. 3), suggesting that clustering may play an important role in the occurrence process of flood events.

The flood of record for the Potomac River, as for many large drainages in the eastern United States and the Ohio River basin (see Miller 1990), occurred in March 1936. Miller (1990) presents the March 1936 flood as the prototype for winter–spring extratropical systems that produce extreme floods in large drainage basins of the eastern United States. Tropical cyclone floods are much more prominently represented in the upper tail of flood peak distributions than in the central and lower portions (Villarini and Smith 2010). For the Potomac record, 50% of the top 10 floods are tropical cyclone events, but less than 10% of other floods are tropical cyclone events. The March 1936 flood and a number of the tropical cyclone events are major flood events in all three basins, reflecting the potential for extreme flooding over large areas. The probability that a flood peak in the Potomac is also a flood peak in the James is 0.57, and the joint probability for Potomac and Susquehanna floods is 0.44. The lowest joint probability in flooding of 0.34 is between the James and Susquehanna, the two basins separated by the largest distance.

Mixtures of flood-generating mechanisms (Fig. 1) and orographic precipitation mechanisms result in large
spatial heterogeneity of flood magnitudes over the eastern United States. In Fig. 4, we show the ratio of at-site 100-yr flood estimates to regional estimates of the 10-yr flood. At-site estimates of the 100-yr flood are based on the generalized extreme value (GEV) distribution, parameterized as

$$F(x|\mu_i, \sigma_i, \xi_i) = \exp \left\{ -\left[ 1 + \xi_i \left( \frac{x - \mu_i}{\sigma_i} \right) \right]^{-1/\xi_i} \right\}, \quad (1)$$

where $\mu_i \in [-\infty, +\infty]$ is the location parameter of station $i$, $\sigma_i \in (0, +\infty]$ is the scale parameter, and $\xi_i \in [-\infty, +\infty]$ is the shape parameter. We developed at-site estimates of the 100-yr flood peak for stations with 75-yr records and without trends and abrupt changes in mean and variance (see Villarini and Smith 2010). The regional estimate of the 10-yr flood peak is based on scaling relationships,

$$Q_i = a (A_i/A_0)^b, \quad (2)$$

where $Q_i$ is the sample 10-yr flood peak for station $i$, $A_i$ is the drainage area for station $i$, and $A_0$ is the reference drainage area (taken to be 1000 km$^2$). The estimated exponent for 10-yr flood quantiles in the eastern U.S. stations is 0.74 [see Villarini and Smith (2010) for additional details]. Maximum values of the normalized flood peaks are concentrated along the Appalachian region of the eastern United States, where 100-yr flood peaks are typically more than 4 times larger than the regional 10-yr flood peaks.

We performed frequency analyses of 38 annual maximum rainfall series with more than 75 yr of record for
the same region used for flood frequency results in Fig. 1. As with flood frequency analyses, we first assess the stationarity of rainfall series by testing for the presence of abrupt changes and long-term trends. Change points in mean and variance were examined by means of the Pettitt test (Pettitt 1979; see also Villarini et al. 2009; Villarini and Smith 2010). The presence of slowly varying trends was tested by means of Mann–Kendall and Spearman tests (e.g., Helsel and Hirsch 1993). For the annual maximum rainfall series, we found that the stationarity assumption was violated only in two stations (change point in variance). No statistically significant increasing or decreasing trends were detected. Annual maximum rainfall series were also examined for long-term persistence based on estimates of the aggregated variance estimator of the Hurst exponent (Montanari et al. 1999). For 12 of the stations, we obtained a value of the Hurst coefficient larger than 0.5, suggesting the potential for long-term persistence. The aggregated variance estimator, however, is affected by large sampling uncertainties. We computed the uncertainties associated with this estimator by means of bootstrap (Efron and Tibshirani 1993), testing the null hypothesis that the Hurst exponent was equal to 0.5. Out of the 12 stations with \( H > 0.5 \), only 1 station had an estimated Hurst coefficient that was significantly greater than 0.5. These results are also consistent with analyses of long-term persistence in annual peak flood records (Villarini et al. 2009).

For rain gauge stations in the central section of the eastern United States (with records of at least 75 yr), estimated GEV parameters of annual maximum daily rainfall series exhibit large spatial heterogeneity (Fig. 5). Location and scale parameters are largest in the eastern portion of the domain, with minimum values west of the Appalachian Mountains. As with flood peak distributions, the shape parameter of annual maximum rainfall series is positive, suggesting heavy tails for extreme rainfall. Morrison and Smith (2002) and Villarini and Smith (2010) show that mixtures of different flood-generating mechanisms can lead to elevated values of the shape parameter. Geographic variation in the estimated shape parameters is not as pronounced as for location and scale parameters, but there is a tendency for larger shape parameters in high-elevation portions of the region.
Warm-season thunderstorms are a prominent element of the occurrence of annual maximum daily rainfall (Fig. 3c), but they are not significant contributors to flooding in large basins of the eastern United States (Fig. 3). Some of the largest rainfall accumulations in the world at time intervals less than 6 h have been produced by orographic thunderstorm systems in the central Appalachians embedded within warm-season extratropical systems (Miller 1990). Warm-season thunderstorm systems are also important flood agents in urban watersheds along the urban megalopolis of the eastern United States (Ntelekos et al. 2007; Smith et al. 2002; Shepherd et al. 2002).

Clustering of heavy rainfall events was examined through Poisson analyses of annual counts of heavy rainfall events. A heavy rainfall event was defined as a daily accumulation greater than 25 mm. For the Poisson distribution, the dispersion parameter—that is, the ratio of the variance to the mean in annual counts—takes the value 1. For the rain gauge stations illustrated in Fig. 5, all but five of the stations have dispersion parameters for 25-mm rainfall occurrences greater than 1. There is a general increase in 25-mm days from west to east, mirroring the location and scale parameters for the annual maximum rainfall distributions (figure not shown). Clustering of heavy rainfall events can be linked to interannual variation in climate variables. Poisson regression analyses for the 25-mm rainfall counts were performed using North Atlantic Oscillation (NAO; Hurrell 1995; Hurrell and Van Loon 1997), Southern Oscillation index (SOI; Trenberth 1984) and Atlantic multidecadal oscillation (AMO; Enfield et al. 2001) as climate indices. The number of stations that show significant dependence of heavy rainfall counts on climate indices are 11 for AMO, 8 for SOI, and 7 for NAO. These analyses suggest that slowly varying climate processes modulate the occurrence of heavy rainfall and flooding in the eastern United States.

3. Mixture distributions—Storm event analyses

In this section, we examine mixtures of flood-generating mechanisms through analyses of rainfall and flooding from tropical and extratropical storm systems. Analyses focus
on a tropical cyclone in September 2008 and a winter–spring extratropical system in April 2007. We illustrate the spatial structure and temporal evolution of rainfall and associated flooding.

Hurricane Hanna made landfall in South Carolina at 0600 UTC 6 September 2008 as a tropical storm. The storm moved rapidly up the East Coast, producing heavy rainfall and flooding along the major urban centers. Storm total rainfall accumulations from Hanna were concentrated left of center of the track of the storm (Fig. 6). Rainfall analyses are based on composite radar rainfall estimates [Next Generation Weather Radar (NEXRAD) stage IV products; see Baldwin and Mitchell 1998; Lin and Mitchell 2005]. Maximum rainfall accumulations exceeded 250 mm, with the largest accumulations (and flooding) concentrated near landfall and in the Virginia suburbs of Washington, D.C. The left-of-center rainfall distribution was associated with the extratropical transition of the storm (see Hart and Evans 2001; Colle 2003; Atallah and Bosart 2003).

Despite the pronounced left-of-center concentration of rainfall, there was large temporal variation in rainfall structure following landfall (Fig. 7). The distribution of rainfall, relative to the center of circulation, is illustrated in Fig. 7. Analyses are presented for four quadrants: the “front left,” that is, left of track and ahead of track; front right, that is, right of track and ahead of track; rear left; and rear right. Analyses are presented for four periods: 0600 UTC 6 September, which is close to landfall, and for three subsequent 6-h periods. For the final period, the center of the storm is in northern New Jersey. Rainfall is systematically concentrated in the forward sector of the system. At 0600 and 1200 UTC, there are...
pronounced range-dependent maxima associated with eyewall and rainband rainfall. After 1200 UTC, there is a more uniform distribution of elevated rainfall in the forward sectors, principally left of center, with heavy rainfall extending to 300 km from the center of circulation. The left-of-center concentration of heavy rainfall (Fig. 6) is generally bounded on the western margin by elevated terrain of the Appalachians.

Rainfall-rate dynamics at short time intervals and point spatial scales play an important role in runoff production. Disdrometer measurements of raindrop size distributions in Princeton, New Jersey [see Smith et al. (2009) for instrumentation description], at 1-min intervals illustrate the temporal variability in rainfall rate associated with storm structures represented in Fig. 7. More than 100 mm of rainfall at the Princeton site was concentrated during three periods of rainband rainfall, followed by a trailing region of lower rain rates during the passage of the center of circulation over the instrument location (Fig. 8a).

Hanna produced annual flood peaks at USGS stream gauging stations from South Carolina through Connecticut (Fig. 9). Analyses in Fig. 9 are presented as the ratio of observed peak discharge for the event, based on unit values discharge data, to the regional 10-yr flood magnitude (see Fig. 4 for comparison). To highlight regional differences, the results are presented in log scale. The largest flood peaks were located in the Washington, D.C., metropolitan region and were principally linked to flooding in urban and suburban watersheds. Peak values of the log ratio exceed 0.8, reflecting flood peaks that are approximately 7 times larger than the 10-yr regional flood magnitude for the station.

Flooding from Hanna was most significant in small urban watersheds (especially in the Washington, D.C., metropolitan region; Fig. 6). Scaling analyses of flood peaks from Hanna (Fig. 10) are based on instantaneous maxima from all USGS stream gauging stations in the region. A threshold of 1 m$^3$ s$^{-1}$ km$^{-2}$ was imposed; that is, we do not include stations for which the peak discharge (m$^3$ s$^{-1}$) divided by drainage area (km$^2$) is less than 1. The most extreme peaks are for basin scales in the 1–100-km$^2$ range.

An extratropical system produced heavy rainfall and flooding along the entire eastern United States during the period 14–16 April 2007 (Figs. 11 and 12). Heavy rainfall was associated with a rapidly intensifying cyclone. The surface low tracked from the lower Mississippi Valley at 1200 UTC 14 April to the New York City metropolitan region at 1200 UTC 16 April, with minimum sea level pressure decreasing from 1003 hPa at 0000 UTC 15 April to 987 hPa at 0000 UTC 16 April (Fig. 11).
The storm produced heavy rainfall and flooding over much of the eastern United States, with the heaviest rainfall concentrated in the Delaware–New Jersey–New York corridor (Fig. 11). Maximum rainfall accumulations exceeded 200 mm and were concentrated during a 24-h period. Rainfall in the southeastern United States was principally in the form of prefrontal squall lines (Fig. 11), which produced streaks of rainfall accumulations exceeding 100 mm. The region in the northeastern United States, with rainfall accumulations exceeding 200 mm, was associated with rapid cyclogenesis and strong transport of moisture ahead of the frontal boundary (Fig. 12).

The storm produced more than 125 mm of rainfall over the Princeton, New Jersey, disdrometer location and a rainfall-rate distribution (Fig. 13a) that contrasts sharply with the September 2008 storm (Fig. 8). Peak 1-min rainfall rates for the April 2007 storm were less than 40 mm h\(^{-1}\) and extreme rainfall accumulations were associated with the extended duration of elevated rainfall rates. In addition to the rainfall-rate differences, raindrop size distributions contrast sharply between the two heavy rain events (Figs. 13b and 13c; Figs. 8b and 8c). The April 2007 storm produced much larger drop arrival rates and the September 2008 storm exhibited much larger drop diameters, especially for the periods of peak rainfall rates. In addition to the contrasts between storms, there are also pronounced changes in raindrop size distributions within the two storm events [see Chapon et al. (2008) for related analyses].

The April 2007 storm produced annual flood peaks from Georgia to New York. The most extreme flooding (Fig. 14) was concentrated in the New Jersey–New York corridor with the largest storm total accumulations. Flood peaks in this area were more than 6 times larger than the regional 10-yr flood magnitude. An area of elevated flood peaks in West Virginia associated with smaller storm total rainfall accumulations reflects the combined contributions of snowmelt and runoff from storm event rainfall. Scaling properties of flood peaks for the April 2007 storm (Fig. 10) exhibited a more uniform spread around the log–log scaling line than for flood peaks from Hurricane Hanna.

4. Discussion and conclusions

In this section, we synthesize results from previous sections and discuss directions for stochastic modeling of rainfall that are linked to analyses of the spatial and temporal structure of flooding. Principal conclusions from the analyses of the previous sections are as follows:

1) Flood peak distributions in the eastern United States reflect mixtures of flood-generating mechanisms associated with tropical cyclones and extratropical systems. There is large spatial heterogeneity in the occurrence of tropical cyclone floods, winter–spring extratropical floods, and fall extratropical floods. The upper tail of flood peak distributions in much of the eastern United States reflects a greater contribution from tropical cyclone floods. The apparent thickness of tails of flood peaks and annual maximum rainfall is linked to mixtures of flood-generating mechanisms.

2) The occurrence processes for flooding and heavy rainfall in the eastern United States exhibit large annual variability in counts, but they do not exhibit significant evidence of trends associated with changing climate. Clustering of heavy rainfall and flood events is likely an important element of the occurrence process. Slowly varying climate processes that are linked to properties of tropical cyclone and extratropical cyclone occurrences may play an important role in controlling clustering properties of flood and heavy rainfall occurrence.

3) Spatial correlation in flood occurrences is an important feature of long-duration flood records, as illustrated through analyses of peaks-over-threshold data.
from the Susquehanna, Potomac, and James River basins. Analyses of flood peaks from the September 2008 tropical system and April 2007 extratropical system highlight the striking spatial correlation in flood peaks and the contrasting spatial structure of flood magnitudes between extratropical and tropical systems. Scale-dependent flood response and spatial heterogeneities in rainfall associated with orographic precipitation mechanisms are key features of the spatial structure of flood peaks for a storm.

4) USGS discharge data for the eastern United States provide an exceptional resource for characterizing spatial extremes of flood magnitudes. Analyses of the April 2007 and September 2008 flood events are based on flood peak data from more than 2000 stream gauges. Composite flood peak data from a large sample of flood events can provide the observational resources for data-driven analyses of spatial flood extremes.

Characterization of rainfall and flood extremes has centered on univariate and (low dimensional) multivariate analyses based on extreme value theory (e.g., Davison and Smith 1993). Extensions to spatial extremes is an important goal of flood hazard characterization for both applications (floodplain mapping over large regions and flood insurance, for example) and for advances in flood science. Data-driven approaches can be based on combinations of properties of flood extremes based on analyses of peak observations from a large number of stream gauges (Figs. 9 and 14) and properties of rainfall extremes, as formulated below.

The point process framework introduced by Le Cam (1961) has provided stimulus for rainfall modeling (see Waymire et al. 1984; Rodriguez-Iturbe et al. 1987; Foufoula-Georgiou 1997) in a Lagrangian framework. These stochastic models of rainfall are centered on data-sets that are “event oriented,” as opposed to centered on time series observations from rain gauge stations. The structure of stochastic models of rainfall begins with counts of storm events, $N^i$—that is, the number of storms during year $i$—and represents the rainfall-rate distribution in terms of covariate processes, $(T^i_j, U^i_j, X^j(t), Z^j(t,x,z), R^j(t,x); j = 1, \ldots, N^i)$, where $T^i_j$ is the initiation time of the $j$th storm during year $i$; $U^i_j$ is an indicator for storm
type (tropical versus extratropical, for example); $X_j^t(t)$ is the location at time $t$ of the $j$th storm during year $I$; $Z_j^i(t,x,z)$ is a vector of ancillary state variables at time $t$, spatial location $x$, and elevation $z$ for the $j$th storm; and $R_j^i(t,x)$ is the rainfall rate at time $t$ and spatial location $x$ from the $j$th storm. Two basic notions that we derive from analyses of flood mixtures in the eastern United States are that differing mechanisms associated with tropical and extratropical systems should be distinguished and that rainfall rate should be examined relative to the center of the storm environment, which is naturally represented by the center of circulation for tropical and extratropical systems.

An important element of stochastic models of rainfall is the development of statistical procedures for parameter estimation and inference. The formulation above presumes that observations can be obtained for the ancillary variables representing storm initiation, storm type, and storm position. The HURDAT provides precisely these observations for tropical cyclones, and comparable analysis products have been developed for extratropical cyclones (Jarvinen et al. 1984; Neumann et al. 1993). The atmospheric state variables represented by $Z_j^i(t,x,z)$ can be derived from reanalysis fields (Kistler et al. 2001), which are state estimates of atmospheric fields based on surface and upper-air observations, constrained by conservation equations of the atmospheric modeling system. Reanalysis fields, such as the NCEP–NCAR product, have been derived for long time periods; the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) products cover the period from 1948 to the present. An important problem is determining how to condense reanalysis fields to data structures that are both informative on rainfall distribution and of manageable size for analysis. For extratropical systems, reanalysis fields can characterize the spatial structure of moisture transport and vertical motion fields (see Fig. 12) and provide insights into the clustering of storm occurrences (e.g., Mailier et al. 2006; Vitolo et al. 2009). For tropical systems, variables characterizing extratropical transition (see Atallah and Bosart 2003; Hart and Evans 2001) can play an important role in conditioning for rainfall distribution properties, such as the left-of-center versus right-of-center structure of tropical cyclone rainfall in the eastern United States (see Fig. 7). Developing formal procedures for statistical inference and parameter estimation is especially challenging for stochastic models of rainfall, even for simple formulations in the Le Cam framework (see Smith and Karr 1990).

Accurate measurements of rainfall in space and time are critical for progress in the development of new stochastic models of rainfall. Rain gauge networks will

FIG. 12. The 850-hPa specific humidity (g kg$^{-1}$) and wind field (m s$^{-1}$) based on simulations using the WRF model on a 9-km grid with NCEP–NCAR reanalysis fields used for initial and boundary conditions [see Ntelekos et al. (2008, 2009) and Zhang et al. (2009) for similar model implementations for the eastern United States].
continue to play an important role in rainfall analysis, but weather radar networks and satellite-based sensors will continue to assume increasing importance. Technical problems remain [see Krajewski et al. (2010) for a 30-yr retrospective on quantitative rainfall measurement from radar], but applications of existing measurement systems are producing steady advances in understanding space–time properties of rainfall. Accurate, high-resolution rainfall products that cover large areas (see Figs. 6 and 11) are required from composite radar products (see Krajewski et al. 2008, 2011) and satellite rainfall products (e.g., Huffman et al. 2007).

Spatial heterogeneity in rainfall associated with mountainous terrain is one of the key elements of rainfall and flooding for the eastern United States and for many regions around the world. Treatment of terrain impacts on rainfall should move beyond the inclusion of spatially varying parameters to direct treatment of orographic mechanisms. Similar issues arise for treatment of land–water boundaries, heterogeneities in vegetation, and urbanization.

Seasonal cycles of rainfall are linked to the occurrence and evolution of tropical and extratropical systems. Interannual variation of rainfall, as reflected in the tropical cyclone flood counts (Fig. 2), modulates the seasonal cycle and provides important links to slowly varying climate variables. Inferences concerning long-term trends in precipitation and associated flood regimes can be usefully formulated in terms of problems concerning changing frequencies and properties of tropical and extratropical cyclones. These problems are closely linked with the characterization of clustering of storm occurrences and

---

**Fig. 13.** (a) Rainfall rate, (b) drop arrival rate, and (c) mean diameter (bottom) time series on 15 Apr 2007, based on disdrometer measurements in Princeton, NJ.
long-term persistence in statistics of rainfall and flood extremes.

Dynamics of short-term surface rainfall rate play a central role in flood hydrology. As illustrated in Figs. 8 and 13, there are striking contrasts in dynamics of rainfall rate—both between storm systems and within an individual storm. Distributional properties or rainfall rate over the life cycle of a storm system are poorly understood and represent a major challenge for stochastic modeling of rainfall (e.g., Georgakakos and Krajewski 1996; Sempere-Torres et al. 1998; Krajewski et al. 2003; Uijlenhoet et al. 2003). The upper tail of rainfall-rate distributions is of particular importance for examining spatial extremes of flooding.

The eastern United States is an interesting study region because of the complex environment, diversity of physical processes, and exceptional data resources. It is an important setting for the study of rainfall and the associated flood processes.

Acknowledgments. This research was funded by the Willis Research Network, the NOAA Cooperative Institute for Climate Sciences, NASA, and the National Science Foundation (Grants BES-0607036, EF-0709538, and ITR-0427325). NEXRAD rainfall data provided by NCAR/EOL under sponsorship of the National Science Foundation (available online at http://data.eol.ucar.edu/).

REFERENCES


