Limits to Flood Forecasting in the Colorado Front Range for Two Summer Convection Periods Using Radar Nowcasting and a Distributed Hydrologic Model

HERNAN A. MORENO
School of Sustainable Engineering and the Built Environment, and Decision Center for a Desert City, Global Institute of Sustainability, Arizona State University, Tempe, Arizona

ENRIQUE R. VIVONI
School of Sustainable Engineering and the Built Environment, and School of Earth and Space Exploration, Arizona State University, Tempe, Arizona

DAVID J. GOCHIS
National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 27 August 2012, in final form 8 March 2013)

ABSTRACT

Flood forecasting in mountain basins remains a challenge given the difficulty in accurately predicting rainfall and in representing hydrologic processes in complex terrain. This study identifies flood predictability patterns in mountain areas using quantitative precipitation forecasts for two summer events from radar nowcasting and a distributed hydrologic model. The authors focus on 11 mountain watersheds in the Colorado Front Range for two warm-season convective periods in 2004 and 2006. The effects of rainfall distribution, forecast lead time, and basin area on flood forecasting skill are quantified by means of regional verification of precipitation fields and analyses of the integrated and distributed basin responses. The authors postulate that rainfall and watershed characteristics are responsible for patterns that determine flood predictability at different catchment scales. Coupled simulations reveal that the largest decrease in precipitation forecast skill occurs between 15- and 45-min lead times that coincide with rapid development and movements of convective systems. Consistent with this, flood forecasting skill decreases with nowcasting lead time, but the functional relation depends on the interactions between watershed properties and rainfall characteristics. Across the majority of the basins, flood forecasting skill is reduced noticeably for nowcasting lead times greater than 30 min. The authors identified that intermediate basin areas \(\sim \) (2–20) km\(^2\) exhibit the largest flood forecast errors with the largest differences across nowcasting ensemble members. The typical size of summer convective storms is found to coincide well with these maximum errors, while basin properties dictate the shape of the scale dependency of flood predictability for different lead times.

1. Introduction

Flood predictability in mountain watersheds is challenging because of our limited capacity to accurately forecast precipitation in time and space, the short response time of watersheds, and the inherent uncertainties present in hydrologic modeling. Nevertheless, the use of quantitative precipitation forecasts (QPFs) in hydrologic models of these settings can potentially improve streamflow predictions, as in other regions (Pessoa et al. 1993; Warner et al. 2000; Collier and Krzysztofowicz 2000; Berenguer et al. 2005; Vivoni et al. 2006; Chiang et al. 2007; Collier 2007). When QPFs are unavailable, the maximum lead time for flood warnings is approximately the basin response time, a value dependent on watershed characteristics and antecedent soil moisture. Nonetheless, the expected hydrologic gains in prediction time from QPFs are limited by the quality of the forecasted fields. Under warm-season convection, the short life span

Corresponding author address: Enrique R. Vivoni, School of Earth and Space Exploration, Arizona State University, ISTB4, Room 769, Tempe, AZ 85287-6004.
E-mail: vivoni@asu.edu

DOI: 10.1175/JHM-D-12-0129.1

© 2013 American Meteorological Society
and rapid evolution of these systems controls the accuracy of rainfall forecasts at different lead times (e.g., Ganguly and Bras 2003; Lin et al. 2005; Sharif et al. 2006). Thus, uncertainty about future rainfall distribution limits our capability for flood forecasting because of the sensitivity of runoff production to rapidly changing precipitation fields (Vivoni et al. 2006; Reed et al. 2007; Moreno et al. 2012).

For warm-season convective systems, radar nowcasts at short lead times (0–3 h) are often found to be the most skillful method for producing QPFs at high spatiotemporal resolutions (e.g., Collier 1991; Golding 2000; Ganguly and Bras 2003). The term radar nowcasting refers to a number of different algorithms that utilize sequences of rainfall fields to derive storm motion vectors applied to subsequent imagery (e.g., Dixon and Wiener 1993; Ganguly and Bras 2003; Bowler et al. 2004; Li and Lai 2004; Vivoni et al. 2006; Van Horne et al. 2006; Mass 2012). Several techniques are able to compute storm growth, movement, and dissipation, while providing quantitative measures of precipitation amounts. The availability of weather radar networks has expanded the applications of nowcasting techniques, primarily for regions where the errors from quantitative precipitation estimates (QPEs) are well understood (Berenguer et al. 2005; Sharif et al. 2006). In mountainous areas, however, challenges remain in the derivation of radar-based QPEs (e.g., Yates et al. 2001; Borga 2002; Hossain et al. 2004; Verbunt et al. 2007; Anagnostou et al. 2010; Moreno et al. 2012) and thus in the use of radar nowcasting techniques for predicting the timing, location, and magnitude of precipitation as input to hydrologic models. Uncertainties inherent in radar nowcasting QPFs are a consequence of the difficulty to forecast rainfall fields for extended periods given that extrapolation functions lose their correlation structures at large lead times (e.g., Sharif et al. 2006; Vivoni et al. 2007b).

Distributed hydrologic models are designed to continually ingest high-quality rainfall estimates and forecasts, allowing for real-time flood forecasting using information about future rainfall (Garrote and Bras 1995; Vivoni et al. 2006; Collier 2007). The distributed nature of these types of hydrologic models permits exploring the spatial properties of the basin response relative to the spatiotemporal evolution of precipitation forcing. For example, the streamflow properties can be assessed as a function of watershed area to understand the scale dependence of the flood forecast skill (e.g., Vivoni et al. 2007a,b). In addition, distributed modeling offers an opportunity to quantify the propagation of rainfall errors into the spatial hydrologic response and how these interact with basin properties. As a result of available spatial data on topography, soil, and land cover properties, differential basin responses to meteorological forcing can be assessed over a range of conditions in a region (Germann et al. 2009; Mascaro et al. 2010b; Schröter et al. 2011).

Previous studies have explored the limits to flood predictability through the use of distributed hydrologic models. For example, Berenguer et al. (2005), Vivoni et al. (2006), and Sharif et al. (2006) used different rainfall-runoff models to evaluate radar nowcasting techniques in various settings. The studies coincided in finding a decrease in the flood forecasting skill with rainfall forecast lead time, in accordance with theoretical models (Lin et al. 2005). On the other hand, while flood scaling theory is advanced in hydrology (e.g., Ogden and Dawdy 2003; Gupta 2004), few attempts have been made at analyzing the scale dependence of flood forecasting skill by inspecting results at a range of internal watershed sites. For example, Benoit et al. (2000) quantified hydrologic errors of radar nowcasts at 23 nested sites for a major flood event, while Vivoni et al. (2006) investigated the flood forecast skill from radar nowcasts at 15 internal sites during two separate flood events. Sangati et al. (2009) and Sangati and Borga (2009) found that the aggregation scales of rainfall and soil properties significantly affect the estimation of peak flows in mountain basins. All studies coincided in observing a reduction in the total forecasting error with increasing basin area as a result of the integration of different hydrologic processes in the watershed. Despite these prior efforts, the spatial and temporal limits to flood predictability in mountain catchments experiencing summer convection are currently unknown.

This study seeks to quantify flood predictability using the triangulated irregular network (TIN)-based Real-time Integrated Basin Simulator (tRIBS; Ivanov et al. 2004a; Vivoni et al. 2007a) as a tool to generate flood predictions using radar nowcasting QPFs. With these coupled simulation tools, we quantify the relation of flood forecasting skill with lead time in a set of mountain basins that span several orders of magnitude in catchment scale. We pose the following question: Is the predictability of floods more limited at certain catchment scales by the interaction of rainfall errors with basin characteristics? If so, then differences in runoff production resulting from varying hydrologic processes at different scales may help determine flood predictability. We conduct our work in a set of headwater basins in the Colorado Front Range (CFR) because of its physiographic complexities and recurrent warm-season convective storms and their associated flood hazards. Given the limited length of the high-quality radar rainfall record in this region, we selected two warm-season storm events that resulted in simultaneous streamflow responses across the headwater...
basins. As a result, these storms lead to rapid and measurable runoff pulses that are deemed to occur frequently during the summer season, but may not constitute floods of record. We analyze the skill of ensemble precipitation forecasts relative to observed rainfall fields derived from a calibrated radar product for two summer convection periods (hereafter called storm events). Subsequently, we investigate the distributed flood forecasting skill and its dependence with lead time and catchment scale for the ensemble rainfall forecasts. In addition, we investigate how precipitation errors are transmitted to streamflow uncertainty at internal watershed sites as a flood wave progresses downstream. For these events and locations, we find characteristic patterns in flood predictability governed by the varying watershed characteristics through an analysis of the scale dependence of flood forecast errors. Finally, we discuss the limits of flood forecasting with radar nowcasting in mountain environments of the Colorado Front Range.

2. Methods

a. Study region and watershed characteristics

The Colorado Front Range in north-central Colorado, United States, was selected for its availability of hydrometeorological information and historical potential for floods during the summer season (e.g., Petersen et al. 1999; Ashley and Ashley 2008). Regional data include high-resolution (subhourly to hourly) information from 11 stream gauges, 7 rain gauges and meteorological stations, and three weather radars (KCYS, KFTG, and KPUX) and their 150-km coverage umbrellas. The associated geographic coverage of QPF products and the three weather radars are shown.
CFR (Jarret and Tomlinson 2000). Eleven headwater basins distributed on the east-facing slope of the CFR, northwest of the Denver urban corridor, were selected to quantify the flood forecasting skill obtained from radar nowcasting. Table 1 summarizes the major characteristics of the selected basins with drainage areas ranging between 37.2 and 359.5 km². The watersheds have considerable relief with mean elevations from 2287 to 3455 m. Given that some portions of their terrain are located at high altitude (above 2750 m), six of the basins [North Fork Big Thompson River (NFORK), Big Thompson River (BTHOM), North Saint Vrain Creek (NVRAIN), Middle Saint Vrain Creek (MVRAIN), South Saint Vrain Creek (SVRAIN), and Middle Boulder Creek (MBOUL)] are likely to be directly influenced by remnant snowmelt processes during the summer season. The contribution of snowmelt runoff to outlet hydrographs is clearly reduced during the months of July, August, and September. By that time, only small basin areas are snow covered. As a result, snowmelt contributions to soil moisture and the groundwater table position mostly influence baseflow in those catchments. Mean slopes vary between 28% and 40%, with high standard deviations induced by the presence of vast areas of steep bedrock and the sudden changes in terrain features. Sharp slopes, narrow valleys, and predominant dendritic patterns in the channel networks often lead to rapid runoff responses and short concentration (or response) times (Table 1). Figure 2 presents the spatial distribution of elevation, soil, and vegetation types and stream channel networks in the watersheds. Overall, the watersheds are characterized by a heterogeneous mixture of soil, vegetation, and exposed bedrock conditions. Dominant soils across the watersheds are sandy loam, loam, and bedrock, while vegetation is characterized by the prevalence of alpine regions and upper montane, subalpine forests, followed by lower montane grassland and shrublands.

b. Quantitative precipitation estimates and event characteristics

High-resolution QPEs were derived from level-II data from the NEXRAD radars at Denver, Colorado (KFTG); Pueblo, Colorado (KPUX); and Cheyenne, Wyoming (KCYS), which underwent a certain degree of signal processing to contain three-dimensional volume scans of radar reflectivity on a polar coordinate grid. The data were projected onto a 1-km common grid and constant altitude plan position indicator (CAPPI) levels. Subsequently, a composite reflectivity product was created that selected the maximum reflectivity return from each CAPPI from altitudes of 3-km to 5-km above ground level. This band of CAPPIs used in the development of the composite product lies above the terrain of the CFR, which minimizes, though not completely eliminates, the potential for false ground return echoes (i.e., clutter). The composite reflectivity data from the three radars were mosaicked together into a single product where the maximum return from each individual composite product was selected. In this way, we preserved as many heavy precipitation signals as possible prior to hail thresholding. Once the mosaicked composite reflectivity product was created, a hail threshold of 53 dBZ was applied and a minimum reflectivity return of 10 dBZ was used to eliminate weak echoes unrelated to precipitation. No additional vertical profile of reflectivity (VPR) correction was applied in this study.

A power law of the form \( Z = 700R^{1.3} \) was selected to convert reflectivity (Z) to 5-min, 1-km resolution rainfall rates (R) following an optimization procedure that minimized errors with collocated pixels at seven local
rain gauges (Moreno et al. 2013). To achieve computational savings, the QPEs were time aggregated to 15-min, 1-km rainfall depths for the radar nowcasting QPF bounding box region shown in Fig. 1. The impact of differences in the temporal scale of the precipitation forcing between 5 min and 15 min is not expected to influence our findings to a large degree given the times of concentration in the watersheds (Table 1). Two periods...

<table>
<thead>
<tr>
<th>NVRAIN</th>
<th>MVRAIN</th>
<th>SVRAIN</th>
<th>MBOUL</th>
<th>COAL</th>
<th>RALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>STALENCO</td>
<td>MIDSTECO</td>
<td>SSVWARCO</td>
<td>BOCMIDCO</td>
<td>COCREPCO</td>
<td>RALCRKCO</td>
</tr>
<tr>
<td>88.5</td>
<td>49.9</td>
<td>35.1</td>
<td>95.5</td>
<td>37.2</td>
<td>117.3</td>
</tr>
<tr>
<td>15.7</td>
<td>15.5</td>
<td>12.6</td>
<td>19.2</td>
<td>14.8</td>
<td>25.8</td>
</tr>
<tr>
<td>59.7</td>
<td>53.9</td>
<td>42.3</td>
<td>48.4</td>
<td>49.0</td>
<td>32.1</td>
</tr>
<tr>
<td>3302</td>
<td>3245</td>
<td>3455</td>
<td>3154</td>
<td>2509</td>
<td>2517</td>
</tr>
<tr>
<td>2527/4344</td>
<td>2635/4225</td>
<td>2858/4087</td>
<td>2495/4075</td>
<td>1993/3183</td>
<td>1847/3204</td>
</tr>
<tr>
<td>327</td>
<td>4029</td>
<td>344</td>
<td>337</td>
<td>187</td>
<td>387</td>
</tr>
<tr>
<td>39.4</td>
<td>40.3</td>
<td>30.0</td>
<td>34.3</td>
<td>34.2</td>
<td>29.2</td>
</tr>
<tr>
<td>29.4</td>
<td>30.7</td>
<td>26.7</td>
<td>23.4</td>
<td>20.8</td>
<td>17.9</td>
</tr>
<tr>
<td>Sandy loam (53.6)</td>
<td>Loam (38)</td>
<td>Sandy loam (44.1)</td>
<td>Sandy loam (45.2)</td>
<td>Sandy loam (39.9)</td>
<td>Loam (45.5)</td>
</tr>
<tr>
<td>Bedrock (30.4)</td>
<td>Sandy loam (36.7)</td>
<td>Loam (30.6)</td>
<td>Loam (41.9)</td>
<td>Loamy sand (34.2)</td>
<td>Sandy loam (34.4)</td>
</tr>
<tr>
<td>Loam (12.6)</td>
<td>Bedrock (25.3)</td>
<td>Bedrock (25.2)</td>
<td>Bedrock (12.4)</td>
<td>Bedrock (14.5)</td>
<td>Loamy sand (13.0)</td>
</tr>
<tr>
<td>Forest (48.4)</td>
<td>Forest (40.8)</td>
<td>Forest (43.1)</td>
<td>Forest (51.8)</td>
<td>Forest (80.8)</td>
<td>Forest (77.8)</td>
</tr>
<tr>
<td>Grassland (26.1)</td>
<td>Grassland (27.3)</td>
<td>Grassland (22)</td>
<td>Shrubland (28.3)</td>
<td>Grassland (9.3)</td>
<td>Grassland (14.6)</td>
</tr>
<tr>
<td>Shrubland (14.0)</td>
<td>Shrubland (21.3)</td>
<td>Shrubland (19.4)</td>
<td>Grassland (9.7)</td>
<td>Shrubland (8.8)</td>
<td>Shrubland (7.1)</td>
</tr>
<tr>
<td>1.63</td>
<td>1.68</td>
<td>1.58</td>
<td>2.07</td>
<td>1.69</td>
<td>3.04</td>
</tr>
</tbody>
</table>

Fig. 2. Spatial distributions of (a) soil types and (b) vegetation classes in the CFR. Elevation contours and basin stream networks for the selected watersheds are also shown. The 30-m soil texture and vegetation type maps were obtained from the Soil Survey Geographic Database (Wang and Melesse 2006), and U.S. Geological Survey National Land Cover Databases (Homer et al. 2004).
with warm-season convective precipitation in the summers of 2004 (17–22 August) and 2006 (6–14 July) were selected for conducting simulations using the radar nowcasting and distributed hydrologic modeling tools. These storm periods were chosen because of 1) the availability of high-resolution radar measurements across the CFR, 2) the simultaneous presence of observed streamflows across most of the watersheds, 3) the development and propagation of intense convective cells in different areas leading to basin responses, and 4) the relatively low contribution of snowmelt to the precipitation-triggered flood events in most of the basins. The selection of the two storm periods is intended to capture simultaneous streamflow events across the basins. Nevertheless, not all basins exhibited significant responses that would be characterized as floods because of the large rainfall variability in the mountain setting.

The spatial distribution of cumulative rainfall depth for the two storm periods is shown in Fig. 3. The simulation windows were defined in a manner that the observed precipitation and hydrograph responses are fully captured across all watersheds. The first period, henceforth called “Storm 2004,” started at 0900 LST 17 August 2004 and consisted of several showers during three consecutive days over different areas that triggered streamflow responses extending for nearly 125 h in the largest basins. A series of thunderstorms moved from west to east and caused heavy precipitation during the afternoon hours (1300–1800 LST), while scattered convection was also observed in the lower elevation zones of the northernmost basins independently of the main storm cores. Most of the heavy rainfall was concentrated in the northern basins, although a streamflow response was observed in all basins. The second storm period began at 2200 LST 6 July 2006 (“Storm 2006”) and consisted of three main cores of convection during different days, primarily in the afternoons with some rain extending into the evening. A prevalent storm motion was observed from the south. Storm sequences generated flood responses in most of the study basins extending up to 160 h in some cases and were more evident in the southern region. Figure 4 shows estimates of the typical size of the convective storms for the two events. These were obtained by identifying areas or storm cores with rainfall depths falling between the 80th and 100th percentiles of the total basin precipitation ($P_T$). Note that storm cores with high precipitation amounts occupy small fractions of each basin area, ranging between 2 and 20 km$^2$.

c. Quantitative precipitation forecasts and radar nowcasting mode

The National Center for Atmospheric Research (NCAR) Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN; Dixon and Wiener 1993)
algorithm was used to generate short-term radar nowcasting QPFs over the CFR. Several studies have demonstrated its value for analysis of thunderstorm characteristics and real-time nowcasting applications (e.g., Pierce et al. 2004; Dance et al. 2010; Roberts et al. 2012). The algorithm allows for real-time automated identification, tracking, and short-term forecasting of thunderstorms based on volume-scan weather radar data. An optimization scheme was employed to match observed storms at one time instance with those at a following time, with geometric operations to deal with mergers and splits. The short-term forecasts of both position and size are based on a weighted linear fit to the storm track history data. This methodology provides the framework necessary to identify storms in radar data and to track them as physical entities (Dixon and Wiener 1993; Joe et al. 2004). Because of the number of parameters in TITAN for controlling forecast properties, we generated a set of nowcasting ensembles consisting of 27 members per forecast lead time. Ensemble radar nowcasting QPFs were produced at a fine resolution (1 km, 15 min) for lead times between 15 and 180 min (15, 30, 45, 60, 90, 120, 150, and 180 min). We varied the following TITAN model parameters within feasible ranges to generate each ensemble for each lead time: 1) minimum storm size (10, 20, and 30 km²), 2) tracking forecast weight rates (0.1, 0.25, and 0.5), and 3) reflectivity dual thresholds (5, 25, and 45 dB) using 20 dB as a static value. The minimum storm sizes are the smallest contiguous regions exceeding a reflectivity threshold, the tracking forecast weights rates distribute forecasting weights to past time steps according to tracking extrapolation, and the dual thresholds help identifying modified reflectivities to track mergers and splits of different cells. Further details on these parameters can be found in Dixon and Wiener (1993) as well as recent versions of the online documentation for TITAN.

TITAN was set to run in an extended-lead forecast mode (Dixon and Wiener 1993; Vivoni et al. 2006) to generate QPFs using available radar observations. In this mode, we eliminate the assumption of no future rainfall by providing nowcasting fields at a single rainfall lead time ($T_L$) over a flood forecasting window defined between the start of observed precipitation $t_i$ and $t_i + T_F$, the flood forecast end time. As shown in Fig. 5, the forecasting time ($T_F$) is discretized into $T_F/\Delta t$ time steps, where $\Delta t$ represents the time step at which forecasts are issued. As an example, Fig. 5 shows $T_F = 2T_L$ for clarity. Normally, many $T_L$ intervals are contained within $T_F$. Thus, a forecast starting at the time $t_i$ for a lead time ($T_L$) uses the $T_L/\Delta t$ most recent historical data (QPEs) to extrapolate the reflectivity field continuously for $\Delta t$ steps until reaching $T_F$ ($\Delta t$ is 15 min here). Subsequently, reflectivity values are converted to rainfall depths using $Z = 700R^{1.3}$ (Moreno et al. 2013). Rainfall forecasts (QPFs) of the same lead time ($T_L$) are assembled in a continuous manner separated by $\Delta t$ intervals. This ensures that each available QPE is extrapolated into a QPF with the same skill specified by an identical lead time (Vivoni et al. 2006). By increasing $T_L$, the time displacement between the QPEs and resulting QPFs is enlarged. Precipitation forecasts that change with lead time are expected to influence flood forecast skill at basin outlets and at internal watershed sites.

d. Distributed hydrologic modeling and numerical experiments

1) MODEL OVERVIEW

The tRIBS distributed hydrologic model was developed for flood forecasting using precipitation inputs at high spatial and temporal resolutions and has been tested in different mountainous regions (e.g., Vivoni et al. 2007b, 2009; Rinehart et al. 2008; Moreno et al. 2012). The tRIBS uses Voronoi polygons, derived from a TIN, to represent basin characteristics with a reduced number of nodes relative to the original data (Vivoni et al. 2004). Surface–subsurface moisture dynamics at each
computational node are resolved by tracking infiltration fronts, water table fluctuations, and lateral redistribution in the hillslope and channel system. Surface runoff during storm events is produced by infiltration excess, saturation excess, perched return flow, and groundwater exfiltration mechanisms, while flood routing is performed through hydrologic overland flow and hydraulic channel routing (e.g., Ivanov et al. 2004a). Water losses to the atmosphere occur through soil evaporation, plant transpiration, and evaporation of intercepted water. As a physically based model, the tRIBS is able to ingest spatially varying terrain, soil, and vegetation properties, as well as spatiotemporal meteorological forcing, to reproduce hydrologic process evolution at scales ranging from hillslopes to large river basins. The tRIBS can utilize QPFs from radar nowcasting to generate streamflow forecasts at the basin outlet and at interior or nested sites.

2) MODEL PARAMETERS AND INITIALIZATION

The distributed model requires parameters describing the surface, subsurface, vegetation, and channel characteristics that control the hydrologic response to storm and interstorm periods. A list of model parameters is provided in Ivanov et al. (2004a) and Moreno et al. (2012). In addition, the model requires a spatially distributed initial condition that characterizes the soil moisture and groundwater states. In some watersheds, initial conditions should account for remnant snowmelt processes that contribute to summer season baseflows. These are particularly critical for flood forecasting as the effect of initialization is not dissipated in short simulation periods. The lack of field measurements on antecedent moisture states as well as the lack of long-term (multiple year) simulations that could serve as a database of initial conditions restricted our approach to the following. An assumption of hydrostatic equilibrium allows inferring soil moisture profiles from the depth to the groundwater table (Ivanov et al. 2004a,b). This can be derived using a number of approaches. In this study, a long-term drainage experiment was conducted in each watershed following the procedure outlined by Vivoni et al. (2007a). Drainage experiments start with fully saturated basins that are allowed to drain for a long period (10 years) without weather or rainfall forcing, leading to hydrographs that are uniquely controlled by soil, channel network, and geomorphic characteristics of individual watersheds. As a result, the simulated instantaneous outlet discharges \( Q_b \) are related to model-based estimates of the spatial mean depth to groundwater \( N_{wt} \) through rating curves relating those variables (Vivoni et al. 2008). The multiple groundwater depth maps associated with specific outlet discharges allowed selecting a set of feasible scenarios (10 per basin) for \( N_{wt} \) corresponding to percentiles of the exceedence probability of the observed discharge at each stream gauge. The use of exceedence probabilities of the observed discharges offers a set of realistic streamflow values that are uniquely related to spatially distributed groundwater depths.

3) MODEL CALIBRATION AND TESTING STRATEGY

Hydrologic processes occurring in mountain catchments merit a careful analysis of model parameters and initial conditions at the storm event scale. The approach in this study first evaluated the relative importance of individual model parameters and initial conditions during one-at-a-time (OAT) analysis in several watersheds (Moreno et al. 2012). Results indicated that outlet streamflow responses were principally controlled by a limited set of parameters including the initial conditions (Table 2). We found the initial depth to groundwater \( \mu(N_{wt}) \) played an important role because of the relatively shallow aquifer (Birkeland et al. 2003) and the presence of snow processes in several basins. Parameters, other than those listed in Table 2, were assigned to...
Table 2. Calibrated parameters and initial conditions in each study watershed during the storm events of 17–22 August 2004. Calibration scores (RMSE and NS) are also shown. Parameters are saturated hydraulic conductivity \(K_s\) and its decay exponent with depth \(f\), soil moisture at saturation \(\theta_s\), pore-size distribution index \(\lambda_0\), air entry bubbling pressure \(N_{aw}\), ratio of horizontal to vertical conductivities \(A\), depth to bedrock \(B\), hillslope velocity exponent \(r\), and basin-averaged depth to groundwater \(\mu(N_{aw})\).

<table>
<thead>
<tr>
<th>Basin</th>
<th>Soil type</th>
<th>(K_s) (mm h(^{-1}))</th>
<th>(f) (mm(^{-1}))</th>
<th>(\theta_s) (-)</th>
<th>(\lambda_0) (-)</th>
<th>(\psi_0) (mm)</th>
<th>(A) (-)</th>
<th>(R) (-)</th>
<th>(B) (m)</th>
<th>(\mu(N_{aw})) (m)</th>
<th>RMSE</th>
<th>NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUCK</td>
<td>Sandy loam</td>
<td>7.96</td>
<td>0.033 36</td>
<td>0.32</td>
<td>0.81</td>
<td>-244.22</td>
<td>223.00</td>
<td>0.45</td>
<td>1.26</td>
<td>1.17</td>
<td>0.46</td>
<td>0.61</td>
</tr>
<tr>
<td>Loam</td>
<td>30.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFORK</td>
<td>Sandy loam</td>
<td>35.83</td>
<td>0.013 09</td>
<td>0.36</td>
<td>1.38</td>
<td>-27.03</td>
<td>274.70</td>
<td>0.42</td>
<td>1.42</td>
<td>0.39</td>
<td>1.44</td>
<td>-0.81</td>
</tr>
<tr>
<td>Bedrock</td>
<td>1.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-580.70</td>
<td>1.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTHOM</td>
<td>Sandy loam</td>
<td>30.69</td>
<td>0.005 42</td>
<td>0.58</td>
<td>1.18</td>
<td>-187.17</td>
<td>234.63</td>
<td>0.36</td>
<td>0.96</td>
<td>0.59</td>
<td>1.44</td>
<td>-2.42</td>
</tr>
<tr>
<td>Bedrock</td>
<td>3.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-209.35</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FISH</td>
<td>Sandy loam</td>
<td>0.41</td>
<td>0.009 27</td>
<td>0.6</td>
<td>1.59</td>
<td>-92.01</td>
<td>649.07</td>
<td>0.43</td>
<td>1.02</td>
<td>1.092</td>
<td>0.09</td>
<td>0.29</td>
</tr>
<tr>
<td>Bedrock</td>
<td>3.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-489.46</td>
<td>1.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTHOM</td>
<td>Sandy loam</td>
<td>4.78</td>
<td>0.042 26</td>
<td>0.57</td>
<td>1.42</td>
<td>-172.46</td>
<td>683.01</td>
<td>0.43</td>
<td>1.92</td>
<td>0.39</td>
<td>0.19</td>
<td>0.83</td>
</tr>
<tr>
<td>Loam</td>
<td>7.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-838.89</td>
<td>110.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVRAIN</td>
<td>Sandy loam</td>
<td>11.15</td>
<td>0.008 78</td>
<td>0.42</td>
<td>0.53</td>
<td>-29.68</td>
<td>887.97</td>
<td>0.45</td>
<td>2.90</td>
<td>0.18</td>
<td>0.12</td>
<td>0.75</td>
</tr>
<tr>
<td>Bedrock</td>
<td>0.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-70.87</td>
<td>1.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MVRAIN</td>
<td>Loam</td>
<td>13.09</td>
<td>0.006 99</td>
<td>0.58</td>
<td>1.01</td>
<td>-179.58</td>
<td>752.14</td>
<td>0.45</td>
<td>0.92</td>
<td>0.26</td>
<td>0.24</td>
<td>-0.26</td>
</tr>
<tr>
<td>Sandy loam</td>
<td>31.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-342.81</td>
<td>275.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SVRAIN</td>
<td>Sandy loam</td>
<td>10.55</td>
<td>0.001 70</td>
<td>0.34</td>
<td>1.62</td>
<td>-343.20</td>
<td>743.76</td>
<td>0.45</td>
<td>0.67</td>
<td>0.49</td>
<td>0.15</td>
<td>0.63</td>
</tr>
<tr>
<td>Loam</td>
<td>20.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-444.64</td>
<td>122.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MBOUL</td>
<td>Sandy loam</td>
<td>35.51</td>
<td>0.040 68</td>
<td>0.42</td>
<td>1.20</td>
<td>-75.36</td>
<td>394.13</td>
<td>0.37</td>
<td>1.79</td>
<td>0.35</td>
<td>0.22</td>
<td>0.38</td>
</tr>
<tr>
<td>Loam</td>
<td>6.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-20.81</td>
<td>647.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COAL</td>
<td>Sandy loam</td>
<td>22.28</td>
<td>0.023 50</td>
<td>0.38</td>
<td>1.66</td>
<td>-21.19</td>
<td>691.23</td>
<td>0.37</td>
<td>0.26</td>
<td>0.39</td>
<td>0.37</td>
<td>0.01</td>
</tr>
<tr>
<td>Loamy sand</td>
<td>84.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-160.97</td>
<td>791.19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RALS</td>
<td>Sandy loam</td>
<td>24.22</td>
<td>0.003 42</td>
<td>0.39</td>
<td>0.82</td>
<td>-680.26</td>
<td>80.94</td>
<td>0.43</td>
<td>2.31</td>
<td>0.84</td>
<td>0.23</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Reference values from the literature (e.g., Chow 1959; Bear 1972; Rutter and Morton 1977; Rawls et al. 1982; Shuttleworth 1988; Birkeland et al. 2003; Ivanov et al. 2004b; Mitchell et al. 2004; Todd and Mays 2005). The Shuffled Complex Evolution (SCE) algorithm (Duan et al. 1993) was then used to automatically find values for selected parameters and initial conditions within feasible ranges of variation reported in prior studies. Storm 2004 was selected to perform the calibration through objective functions that minimized the root-mean-squared error (RMSE) between the observed and simulated streamflow at each basin outlet over the defined period. Through the selection of Storm 2004 as a calibration event, the distributed model parameters are tailored for flood forecasting purposes under summer convection events. A 3-day spin-up period was allowed prior to the evaluation of QPFs during Storm 2004 to reduce the influence of the initial states on the evaluation of the different forecasts. Table 2 summarizes the values for the calibrated parameters for the two major soil types in each basin, along with the RMSE and Nash–Sutcliffe efficiency (NS) scores, relative to the observed streamflow. Calibrated parameter values differ among basins because of their terrain, soil and vegetation characteristics, and initial conditions and fall within realistic ranges. Unavoidably, parameters provide degrees of freedom to compensate for model uncertainties in a manner that differs from basin to basin. In addition to comparisons with streamflow data, flood forecasting skill was assessed with respect to the simulated hydrographs resulting from QPE forcings whose rainfall estimation errors have been minimized as compared with local rain gauge estimates.

Figure 6 presents the observed hydrographs and simulations derived from the calibration exercise for the 11 watersheds, spanning a range of sizes and differing amounts of snowmelt influence. The top 10% of the parameter sets obtained through the SCE procedure for each basin are represented by the gray envelopes. The model is able to reproduce the distinct hydrologic patterns resulting from the combination of rainfall forcing, basin properties, and initial conditions in each catchment. For example, compare the longer response times and extended recessions in larger basins with the faster time to peak discharge in smaller catchments. Furthermore, note the important role of wet initial conditions that amplify the total discharge and delay recession times in basins with an appreciable summer snowmelt signal (e.g., NVRAIN, MBOUL). These results indicate that the model is able to capture the different responses fairly well, with most of the watersheds attaining positive values of NS (Table 2) and RMSE ranging from 0.09 to 1.44 m\(^3\) s\(^{-1}\), depending on the properties of individual watersheds. The largest errors in streamflow are found in BTHOM and NFORK, where the model has difficulty in
Fig. 6. Observed (black lines) and simulated hydrographs (gray lines) resulting from the model calibration in all basins at 15-min intervals. The mean areal precipitation (MAP) in each basin is derived from the 15-min, 1-km level-II product. Uncertainty in parameter values and initial conditions is represented by the light gray envelopes.
replicating high base flows from snowmelt processes not depicted in this study. The remaining discrepancies can be explained by model structural uncertainties and measurement errors in precipitation and streamflow. Flood forecasts for Storm 2006 constitute an independent verification exercise to test the robustness of the calibrated parameters across the watersheds. During these experiments, no parameters are calibrated and only the initial condition is adjusted for the different year.

4) NOWCASTING EXPERIMENTS

Rainfall and flood forecasts generated by the TITAN and tRIBS models in the extended-lead forecasting mode accounted for 216 model runs (eight lead times and 27 ensemble members) per storm period in each basin, for a total of 4752 forecasts. The duration of each forecasting period \( T_F \) was 125 and 170 h, while the start \((t_i)\) and end \((t_i + T_F)\) times were 0900 LST 17 August to 1400 LST 22 August 2004 and 2200 LST 6 July to 0000 LST 14 July 2006, respectively. A hydrologic restart mode, at hours 75 and 30 for Storms 2004 and 2006, respectively, was used in a manner that initial conditions were preserved at the onset of each ensemble forecast and storm events were fully contained in the flood forecasting period \( T_f \). The hydrologic restart mode saves the entire model states in binary-formatted files at the end of a simulation period for use as the initial states directly read into the subsequent simulation. QPE forcing was provided to the model prior to the restarting times in each basin, guaranteeing equal initial conditions for each ensemble member in the flood forecasting period. Gains in computational efficiency for the ensemble simulations were achieved through the use of parallel computations that assign interior subbasins to different computer processors in a high-performance computing cluster (Vivoni et al. 2011). We assigned a relative low number of processors to each basin (eight in total) for the simulations. For the largest watershed in this study, the tRIBS issued flood forecasts at a rate of 2.5 forecasting hours per minute of “wall clock” time. However, on average, the model issued flood forecasts for the next 24 h in 1 min of wall clock time. Thus, the parallel performance suggests that the model can be used in operational forecasting environments.

3. Results and discussion

a. Regional evaluation of quantitative precipitation forecasts

The spatiotemporal properties of radar nowcasting QPFs over the CFR are assessed using two grid-to-grid verification methods. These consider ensemble members for each lead time in categorical and quantitative analyses that help elucidate the regional properties of the QPFs with respect to radar QPEs. The first approach uses a probabilistic analysis in terms of the probability of detection (POD), false alarm rate (FAR), and critical success index (CSI) from contingency tables for distinct forecast thresholds (Ganguly and Bras 2003; Wilks 2006; Gochis et al. 2009). Figure 7 illustrates the different evaluation metrics for the storm periods in 2004 and 2006. A threshold of zero indicates forecast success or failure based on whether rainfall is observed. For subsequent thresholds, successes are only achieved if forecasted pixel precipitation at 15-min intervals is greater or equal to the specified value. Both storm periods exhibit similar results, with lower forecast skill (low POD, high FAR, low CSI) as lead time increases at all threshold values. Nowcasting skill deteriorates at a faster rate for short lead times (e.g., between 15 and 45 min) as compared to the performance change for longer \( T_F \) (e.g., from 120 to 180 min). This occurs since the nowcasting algorithm has difficulty in reproducing the evolution of rapidly developing rainfall cells. As a result, predictive hits decrease and false alarms dramatically increase for short lead times. The POD of rainfall occurrence (when the threshold equals zero) is 0.6 in 2004 and 0.5 in 2006, on average for all lead times. However, a decrease in the forecast skill scores occurs for larger events. The magnitude of this decrease depends on lead time, but 5 mm appears to be the value after which no further decreases in skill are observed. Complementary skill scores (e.g., bias, Clayton’s skill score, equitable threat score, Heidke skill score, and Pierce skill score) confirmed the behaviors summarized above. The underlying cause for the rainfall forecast skill decreases with threshold value is related to the difficulty in accurately predicting the evolution of localized, high-magnitude precipitation events that are comparatively infrequent in the region.

The second approach is tailored to quantify the reduction in rainfall forecasting skill with lead time using the RMSE of \( P_R \), correlation coefficient (CC) and mean ensemble difference (DIFF) between the QPF members and corresponding QPEs, defined as

\[
\text{DIFF} = \frac{1}{n} \sum_{i=1}^{n} (\overline{QPF_{i,j}} - \overline{QPE_{i,j}}) \quad \forall_i = \{1, 2, 3, \ldots, 27\},
\]

where overbars represent spatial means over the entire CFR domain, \( i \) is the forecast time, and \( j \) is the ensemble member. DIFF can be interpreted as the average difference representing under- or overestimation of precipitation over the region. Figure 8 presents these metrics as a function of lead time for the two storm periods in the form of boxplots that capture the ensemble distributions.
A similar pattern in the variation of each metric with lead time is observed in the two storm periods, though differing magnitudes are present. With higher lead time, an asymptotic increase in forecast $P_R$ and an asymptotic reduction in CC are found. In general, the spread among ensemble members is larger for smaller lead times for the $P_R$ and CC metrics. This is not the case for DIFF, where under- or overestimations can average out to small standard deviations at small lead times. Rainfall forecasting skill decreases rapidly at small lead times and becomes negligible at 30- to 45-min lead time, the limit at which forecasts have reduced utility. The time at which predictions no longer worsen appears to be 150 min. Positive values of DIFF and their increase with lead time indicates that radar nowcasts tend to overestimate precipitation, leading to more false alarms, especially for large lead times.

b. Lead time dependence of flood forecasting skill

Rainfall errors should be reflected in flood forecasting skill across individual watersheds. Basin characteristics, specified by the model domain, parameters, and initial conditions, however, are anticipated to play a role in the hydrologic response so that rainfall errors are not transmitted identically to streamflow forecasts. Figure 9 presents the flood forecasting skill at four selected watershed and storm pairs through the RMSE of streamflow ($Q_R$), NS, and mean ensemble difference (DIFF$_Q$), evaluated at basin outlets. DIFF$_Q$ is defined as

$$\text{DIFF}_Q = \frac{1}{n} \sum_{i=1}^{n} (QSF_{it} - QSE_{it}) \quad \forall i = \{1, 2, 3, \ldots, 27\},$$

where QSF and QSE are the instantaneous forecasted and estimated outlet streamflows at time $t$ for the $i^{th}$ ensemble member. These watershed and storm pairs exhibit representative behaviors for other basins, whose patterns in $Q_R$, NS, and DIFF$_Q$ will be discussed next. As a general rule, the flood forecasting skill decreases with lead time across the metrics, although an asymptotic behavior is not necessarily observed for all basins. This results from the variability in streamflow response for QPFs with different errors because of the basin effects on flood timing and magnitude. The quantities $Q_R$ and NS illustrate similar patterns within the same watershed and storm period with respect to the shape and relative spread of the probability distributions, but a slightly different behavior is presented by DIFF$_Q$, with larger ensemble spreads at small lead times. Consistent with prior analyses, DIFF$_Q$ has positive values in all cases, likely as a result of regional rainfall overestimation. Both $Q_R$ and
NS indicate that flood forecasting skill is no better than the mean value over the period (NS < 0) for lead times greater than 30 min. This is consistent with the QPF skill dependence on lead time, indicating the critical role of nowcasting errors on flood forecast skill. Two exceptions are SVRAIN and MBOUL in Storm 2006 that present NS below zero after the 60-min lead time as a result of the snowmelt influence on streamflow.

Three different patterns are observed for the variation of streamflow QR, NS, and DIFFQ with lead time. First, the ensemble mean of QR and DIFFQ grows (declines for NS) asymptotically and interquartile range increases with lead time as observed for the Little Thompson River (LTHOM) and NFORK in Storm 2004 and is replicated in six of the 11 studied cases. This behavior occurs when rainfall predictability exerts a clear influence on flood forecasting skill, thus preserving similar functional relations with lead time. Increases in ensemble dispersion are due to variations in streamflow responses induced when precipitation events exceed hydrologic thresholds, such as infiltration capacity. Second, in three of the studied cases, a similar overall pattern is observed, but after a particular lead time (120 min for NVRAIN in Storm 2006), the interquartile range decreases slightly, possibly because of a reduction in rainfall ensemble spread at individual watersheds. Third, in the remaining cases, a similar pattern to the second pattern is observed, except that both the ensemble mean and interquartile ranges decrease after a certain lead time (60 min in SVRAIN in Storm 2006). This case occurs for small catchments under low rainfall or snowmelt-dominated basins where increases in lead time do not necessarily translate into streamflow error. Complementary boxplot diagrams of Pearson and Spearman rank correlations and bias confirm the reduction in flood forecast skill with lead time and the generalized overestimation of predicted streamflow values (not shown).

In summary, radar nowcasting QPF errors play a significant role on the functional relations between flood forecasting skill and lead time that translate into their limited utility beyond 30 min, except in basins where snowmelt is a major driver. However, watershed initial conditions and properties induce different ensemble responses that shape the functional relations for long lead times. In the next section, the flood forecast skill is assessed as a function of catchment area for each storm event to identify potential spatial limits to predictability within the study basins.

d. Scale dependence of flood forecasting skill

The impacts of warm-season rainfall variability in the CFR region are explored through distributed measures of runoff and internal channel discharges to elucidate the potential relation between flood predictability and...
basin scale. Figure 10 shows examples of the spatial distribution of total precipitation and runoff from the QPE forcing along with the mean ensemble differences from the QPF forcing at two lead times (60 and 180 min) for Storm 2004 in LTHOM. Note that the location of storm cores in the north-central part of the basin might favor increased runoff, but the maximum runoff amounts do not necessarily overlap. This indicates that basin properties (e.g., terrain slope, soil hydraulic conductivity, and initial soil wetness) play a critical role in the basin susceptibility to flooding. The mean ensemble differences in rainfall and runoff are primarily positive in the basin, indicating a general overestimation of precipitation and runoff amounts by the QPFs and the flood forecasts derived from these. As expected, larger positive differences in rainfall and runoff occur for larger lead times (180 min versus 60 min). More interestingly, the changes in the spatial distribution of forecasted precipitation (Figs. 10c,e) with lead time are more dramatic than in runoff (Figs. 10d,f), as watershed characteristics tend to dampen the rainfall forecast errors. Thus, while the spatial distribution and magnitude of QPFs show changes with lead time, the expected differences in basin response are mostly reflected in runoff magnitudes, while spatial patterns remain fairly constant in response to static basin properties.

These results suggest that rainfall forecast errors are not the only driver of flood forecast skill across mountain
watersheds for the same lead time. To investigate this issue further, we selected channel locations corresponding to different contributing areas within and downstream of the major storm cores in 2004 and 2006. Two basin groups were created corresponding to the major storm locations for each period: 2004 [Buckhorn Creek (BUCK), NFORK, BTHOM, Fish Creek (FISH), and LTHOM] and 2006 [NVRAIN, MVRAIN, SVRAIN, MBOUL, Ralston Creek (RALS), and Coal Creek (COAL)]. Figure 11 presents the spatial scale dependence of RMSE in forecasted $P_R$ and $Q_R$ relative to the QPE and its derived flood forecast. The symbols represent
the ensemble mean, while the vertical bars depict the ensemble standard deviation. Three different basins (NFORK, LTHOM, and SRVAIN) and two lead times (60 and 180 min) were selected as representative examples. Results reveal that, although $PR$ and $QR$ increase with lead time in most basins, a clear pattern is not present between $PR$ and catchment area ($Ac$) that can explain the growing dependence of the ensemble mean $QR$ and its standard deviation (or ensemble spread) with $Ac$. Furthermore, no compensating or amplifying behaviors in the ensemble mean or spread are observed for $PR$ with basin area that supports the scale dependence of $QR$. The growing trends in $QR$ with $Ac$ are instead due to the dependence of the streamflow forecast errors on flood magnitudes, which naturally increase with basin area, as noted by Moreno et al. (2013). This evidence points to the need to integrate the spatial characteristics of rainfall forecasts and the corresponding patterns in runoff production, which are linked to watershed properties, to obtain a full picture of the flood predictability in space.

d. Scale dependence on ensemble properties of streamflow errors

Spatial differences in streamflow errors due to variations in rainfall and basin properties can be assessed through the specific error (SE), defined as (Moreno et al. 2013):

$$SE = \frac{QR}{Ac \cdot MAP}, \quad (3)$$

where $QR$ is the RMSE in forecasted streamflow at internal watershed sites characterized by an upstream area ($Ac$) and mean areal precipitation (MAP). Figure 12 presents SE as a function of $Ac$ at BTHOM during Storm 2004 for a lead time of 180 min as an example. The selection of this lead time enhances the visualization of the scale dependency, though similar patterns are observed at other lead times. Selected internal channel locations span a range of catchment areas and are nested along a downstream path from the storm cores to the outlet in each basin. As illustrated in this case, the scale dependence of SE reveals an interesting pattern with a bell-shaped variation with $Ac$. Note that MAP (gray circle size in Fig. 12) can have larger values at small and intermediate scales but decreases with basin area beyond approximately 100 km$^2$, as a result of the smoothing of precipitation fields when integrated over large areas. This example indicates the presence of reduced SE properties at small scales, subsequent increases in the mean and dispersion of SE at intermediate-sized basins, and posterior reductions in SE at the basin outlets. This pattern should be linked with the spatial distribution and rates of runoff production as dictated by the typical storm size and the underlying watershed properties.

To help interpret this result, Fig. 13 shows the variation of $SE/SE_{max}$ with $f/f_{max}$ at four selected watershed-storm pairs. The value $f$ is the runoff ratio defined as the runoff volume divided by the rainfall volume. The quantities $SE_{max}$ and $f_{max}$ are maximum values across all scales for the same lead time and watershed-storm pair. An interesting pattern is observed, which is in Fig. 11. Rainfall (right axes) and streamflow (left axes) error propagation with basin area for (left to right) three basins and two lead times [(top) 60 and (bottom) 180 min] during Storms 2004 and 2006. Symbols represent the ensemble mean, while the vertical bars are the ensemble standard deviation.
accordance with preliminary observations on runoff production areas and rainfall distribution. SE/SE_{max} has a proportional relation with $f/f_{max}$ that is replicated for all lead times. Thus, areas in the watershed with $f/f_{max}$ close to 1 with high runoff productions tend to exhibit higher errors and more limited flood predictability that is tied to precipitation forecast skill. These cases occur for intermediate scales where the full areal storm cover is superimposed on the basin areas. Conversely, low and intermediate $f/f_{max}$ values can be attributed to small and large basin areas, whose errors are smoothed by basin properties, the spatial aggregation occurring in the mean areal precipitation, and the more limited presence of runoff production zones as a fraction of the entire area.

To generalize the patterns across watersheds, Fig. 14 compiles the ensemble mean ($\mu$) and standard deviation ($\sigma$) of normalized SE for three different basin groups organized by similar areas for coincident storm periods. These results indicate that at small $A_c$ (0.1%–1% of total area) a low $\mu$ and $\sigma$ are present. This is due to a relatively large MAP that is unaffected by areal smoothing and by the small $Q_R$ as watershed characteristics mitigate the impact of QPF errors. At intermediate $A_c$ (up to 5% or 10% of the total area, depending of the basin), increased $\mu$ and $\sigma$ are observed as this scale range corresponds to the typical size of warm-season convective systems in the region (see Fig. 4) that lead to a high number of runoff-producing areas. Subbasins of this size present a higher and more variable $Q_R$ under heavy precipitation. Under these conditions, watershed characteristics, such as areas of low permeability and high slopes, trigger variable runoff and streamflow responses. At large $A_c$ (from 10% to total area), lower $\mu$ and $\sigma$ are caused primarily by a significant reduction in $Q_R$ due to the integration effects of the channel network as the flood wave propagates (Vivoni et al. 2006; Mascaro et al. 2010a), but also by an average areal reduction in runoff rates that also decreases the total uncertainty. As a result, the typical size and organization of warm-season convection as well as the runoff characteristics in the basin play a fundamental role on the scale dependence of specific errors in streamflow. We might expect that other type of rainfall systems or basins may have a different functional relation between these normalized quantities.

e. Residual errors from model structural and parametric uncertainty

While flood forecasting skill clearly decreases because of QPF errors, other sources of model uncertainties also affect the total forecast error with respect to observed streamflows. Figure 15 presents the ensemble mean RMSE of the flood forecasts with 1) the model simulations driven with QPEs (RMSE_{QPE}) and 2) the observed streamflows (RMSE_{obs}) at each basin outlet for two lead times (15 and 180 min). Clearly, the magnitude of both errors increases with lead time in all basins. As expected, RMSE_{obs} are typically greater than or equal to RMSE_{QPE}. Differences between RMSE_{obs} and RMSE_{QPE} can be considered as residual errors caused by model structural or parametric uncertainty. Residual errors tend to be
small for Storm 2004 used in the model calibration, but
grow substantially for Storm 2006, reaching the same
magnitude as RMSE_{OPE} in some basins. For the cali-
bration period in Storm 2004, flood forecasting errors are
primarily due to QPF uncertainty, as evidenced by the
reduction of residual errors and the overall increase in
RMSE_{OPE} between the 15- and 180-min lead times.
Meanwhile, the verification period in Storm 2006 exhibits
larger residual errors at the 15-min lead time, as com-
pared to the calibration period, because of the presence
of model structural and parameter uncertainty for this
event. Interestingly, residual errors are significantly re-
duced at a lead time of 180 min for Storm 2006 for most
basins at the expense of an increase in uncertainty in-
duced by QPF errors. As a result, we can conclude that
sources of uncertainty other than nowcasting errors can
worsen flood forecast skill at small lead times for verifi-
cation periods. Undoubtedly, these differences are less
notable at larger lead times since precipitation forecasting
errors increase. Nonetheless, in some basins, residual er-
rors continue to be the largest contributor to total flood
forecast errors during the verification exercise, suggesting
that a single-event model calibration introduces addi-
tional sources of uncertainty even after the initial condi-
tion has been adjusted.

4. Summary and conclusions

In this study, we investigated the propagation of ra-
dar nowcasting errors into distributed flood forecast
skill during two storm periods in 2004 and 2006 in the
Colorado Front Range. This region is known for its
propensity of summer convection that triggers significant
runoff events and occasional flooding. We used high-
resolution radar observations to produce nowcasts from
the NCAR TITAN algorithm for lead times ranging from
15 to 180 min. Using the tRIBS model, we quantified the
resulting flood forecast skill as a function of lead time and

![Figure 13](image-url)

Fig. 13. The SE/SE_{max} as a function of φ/φ_{max} for all lead times in four selected basin and storm pairs.
catchment area. The distributed model allowed depicting the spatial patterns in basin response that explain local differences in flood forecast skill introduced by the distributions of rainfall and watershed properties. For this purpose, we evaluated regional radar nowcasting QPF errors relative to the radar-based QPEs, quantified the dependence of flood forecasting skill on lead time at basin outlets, and identified the scale dependence of flood forecast errors at internal sites. An emphasis was placed on obtaining a detailed picture of the rainfall-runoff error propagation through normalized metrics that removed the effects of basin area and mean areal precipitation. We also quantified how the lead-time decay of radar nowcasting QPF skill affected the flood forecast skill as compared to parametric and model structural uncertainty through comparisons of the calibration and verification periods. The study results indicate the following.

Fig. 14. Relation between SE/SE\textsubscript{max} properties [ensemble mean (left) $\mu$ and ensemble standard deviation (right) $\sigma$] and the normalized catchment area ($A_c$, percent) in (top to bottom) three watershed groupings with similar sizes and behaviors.
1) Radar nowcasting skill decreases with lead time and rainfall magnitude across the CFR, with the most noticeable reduction in forecast skill occurring between 15- and 45-min lead times. For both storm periods, the radar nowcasts tend to overestimate precipitation values, increasing the number of false alarms, in particular for large forecast lead times.

2) Flood forecasting skill also decreases with lead time, but the functional forms follow a different pattern as a result of the interaction with watershed properties, in particular when rainfall intensities exceed hydrologic thresholds. For these studied cases, flood forecasting skill is not better than the forecasted mean for lead times greater than 30 min. Snowmelt-dominated basins have a more limited impact of rainfall uncertainties on the predicted discharges.

3) Watershed properties in conjunction with storm characteristics play a determinant role on the differential susceptibility to high runoff production and flooding. Rainfall-runoff maps show that, despite changes in the spatial distribution of QPFs with lead time, only variations in runoff magnitude are triggered. Analyses of precipitation and streamflow errors also indicate a low correspondence between those variables across different scales, whereas the scale dependence of streamflow errors is primarily due to increasing flood magnitudes.

4) A characteristic pattern was revealed in the scale dependence of specific error (SE) at different lead times. Basin areas coinciding with the typical size of convective storms experience the highest flood forecast errors with the largest differences among ensemble members. Thus, intermediate-sized basins have more limited flood predictability. Watershed properties dictate the shape of the scale dependence as they control rainfall error propagation downstream and modulate the ensemble dispersion across watersheds and lead times. Although MAP is removed from the

Fig. 15. Comparison of streamflow errors at basin outlets derived from flood forecasts relative to observed hydrographs (RMSE_{obs}, black bars) and simulations using QPEs (RMSE_{QPE}, gray bars) for lead times of (left) 15 and (right) 180 min for Storm (top) 2004 and (bottom) 2006.
analysis, precipitation patterns have a principal role in the differential runoff responses.

5) In comparison to rainfall forecast errors, the uncertainties related to model parameters and structural errors can reach similar orders of magnitude in particular for small lead times. At large lead times, QPF errors tend to reduce flood forecasting skill more significantly in most watersheds, though residual errors can remain important in some cases when model structural and parametric uncertainties amplify the disparities in forecasted discharges.

The use of the TITAN as a single nowcasting algorithm is intended to provide a set of reasonable, spatially distributed forecasts during warm-season convection. While a multimodel nowcasting approach is preferable, possibly with comparisons to simpler methods such as persistence (Vivoni et al. 2006), we explored how the model parameter uncertainty in TITAN propagated into the precipitation forecast skill. Additionally, the results of this study are based on the use of a distributed hydrologic model that was calibrated during a storm period in 2004, independently for each basin using a level-II 1-km, 15-min radar product (Moreno et al. 2013). Initial conditions were then adjusted for a verification period in 2006. While single-event model calibration is not ideal for operational settings, it offers the possibility to quantify the errors introduced by rainfall forecasts, independent of model structural and parametric uncertainty. Results are primarily shown relative to model simulations forced with QPEs that we consider as the ground truth. We demonstrated the benefits of using distributed hydrologic models to produce flood forecasts from radar nowcasting as these allow identifying spatial runoff errors and their scale dependence along the channel network that would have been difficult to address using lumped hydrologic approaches. Clearly, improving flood predictability from radar nowcasting analysis alone is difficult given the strong roles played by basin heterogeneities in topography, soils, and vegetation on the runoff response. We found that the interaction of radar nowcasting QPFs and watershed characteristics lead to a distinct pattern in flood predictability, with the greatest errors in intermediate-sized basins. High mean areal precipitation and watershed features tend to reduce the flood forecast uncertainties in small catchments, while channel routing and the areal aggregation of storm systems are responsible for reduced errors in large basins. This scale dependence illustrates the limits of flood predictability in mountain catchments of the Colorado Front Range under two summer convection periods. Additional studies on this dependence during other convective events and in different environments and precipitation regimes are needed to generalize the findings of this study. A possible avenue for exploring this issue systematically is through the coupling of precipitation downscaling approaches with distributed hydrologic models (e.g., Mascaro et al. 2010a,b).

Acknowledgments. This research was supported by the National Weather Service Office of Hydrologic Development (Grant NWS-NWSPO-2007-2000799). We thank the AmeriFLUX and Mesowest networks, the Colorado Division of Water Resources, the NOAA Center for Satellite Applications and Research, and the Center for Hydrometeorology and Remote Sensing at the University of California, Irvine, for data products. We also thank three reviewers whose excellent comments helped us to improve the content and clarity of the manuscript.

REFERENCES


