Journal of Hydrology 541 (2016) 387–400

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

The impacts of climatological adjustment of quantitative precipitation estimates on the accuracy of flash flood detection

Yu Zhang[a,*], Sean Reed[d], Jonathan J. Gourley[c], Brian Cosgrove[a], David Kitzmiller[a], Dong-Jun Seo[b], Robert Cifelli[e]

[a] National Water Center, NOAA National Weather Service, Silver Spring, MD, United States
[b] University of Texas at Arlington, Arlington, TX, United States
[c] NOAA National Severe Storm Laboratory, Norman, OK, United States
[d] Mid-Atlantic River Forecast Center, NOAA National Weather Service, State College, PA, United States
[e] NOAA Earth System Research Laboratory, Boulder, CO, United States

ARTICLE INFO

Article history:
Available online 18 December 2015

Keywords:
Flash flood
Precipitation
Hydrologic model
Detection

SUMMARY

The multisensor Quantitative Precipitation Estimates (MQPEs) created by the US National Weather Service (NWS) are subject to a non-stationary bias. This paper quantifies the impacts of climatological adjustment of MQPEs alone, as well as the compound impacts of adjustment and model calibration, on the accuracy of simulated flood peak magnitude and that in detecting flood events. Our investigation is based on 19 watersheds in the mid-Atlantic region of US, which are grouped into small ($<500\ \text{km}^2$) and large ($>500\ \text{km}^2$) watersheds. NWS archival MQPEs over 1997–2013 for this region are adjusted to match concurrent gauge-based monthly precipitation accumulations. Then raw and adjusted MQPEs serve as inputs to the NWS distributed hydrologic model-threshold frequency framework (DHM-TF). Two experiments via DHM-TF are performed. The first one examines the impacts of adjustment alone through uncalibrated model simulations, whereas the second one focuses on the compound effects of adjustment and calibration on the detection of flood events. Uncalibrated model simulations show broad underestimation of flood peaks for small watersheds and overestimation those for large watersheds. Prior to calibration, adjustment alone tends to reduce the magnitude of simulated flood peaks for small and large basins alike, with 95% of all watersheds experienced decline over 2004–2013. A consequence is that a majority of small watersheds experience no improvement, or deterioration in bias (0% of basins experiencing improvement). By contrast, most (73%) of larger ones exhibit improved bias. Outcomes of the detection experiment show that the role of adjustment is not diminished by calibration for small watersheds, with only 25% of which exhibiting reduced bias after adjustment with calibrated parameters. Furthermore, it is shown that calibration is relatively effective in reducing false alarms (e.g., false alarm rate is down from 0.28 to 0.19 after calibration for small watersheds with calibrated parameters); but its impacts on detection rate are mixed. As an example, the detection rate of 2-Y events in fact declines for small watersheds after calibration is performed (from 0.4 to 0.28, and from 0.28 to 0.19 with raw and adjusted MQPE, respectively). These mixed outcomes underscore the complex interplays between errors in MQPEs, conditional bias in the reference gauge-based analysis, and structural deficiencies of the hydrologic model.

1. Introduction

Accurate detection and prediction of flash floods are of great importance to reducing flood-related life losses and property damages, and yet these are also among the most challenging aspects of hydrologic prediction due to the short response nature of the flooding events (Sene, 2012). Since the advent of weather radar, near real-time radar-based and radar-gauge blended quantitative precipitation estimates (QPEs) have been routinely used for flash flood monitoring and prediction in the world (Cosgrove et al., 2012; Sene, 2012; Berne and Krajewski, 2013). In the United States, most of the warnings are issued based on coupling of high resolution QPEs and Quantitative Precipitation Forecast with Flash Flood Guidance (Gourley et al., 2012), while an emerging paradigm of distributed Model-Threshold Frequency (DHM-TF; Reed et al.,

---

[*] Corresponding author.
E-mail address: yu.zhang@noaa.gov (Y. Zhang).

http://dx.doi.org/10.1016/j.jhydrol.2015.12.017
0022-1694/Published by Elsevier B.V.

Acronyms

ARI	Averaged Recurrence Interval	NWS	National Weather Service
CSI	Critical Success Index	POD	Probability of Detection
DHM-TF	Distributed Hydrologic Model – Threshold Frequency	PPS	Precipitation Processing System
FAR	False Alarm Ratio	PRISM	Parameter-elevation Regressions on Independent Slopes Model
FFD	Flood Frequency Distribution	QPE	Quantitative Precipitation Estimate
GPM	Global Precipitation Measurement	QPF	Quantitative Precipitation Forecast
LP3	Log Pearson type III	RDHM	Research Distributed Hydrologic Model
MAP	Mean Areal Precipitation	TE	Truncation Error
MPE	Multisensor Precipitation Estimator		
MQPE	Multisensor Quantitative Precipitation Estimate		

2007) has been gradually adopted. DHM-TF is based on a grid-based, distributed hydrologic model, and is therefore able to account for upstream inflow in calculating flood risk; it relies on historical streamflow simulations to define the thresholds for flooding and flood intensity levels, and thereby circumvents the difficulty in empirically establishing these thresholds at smaller reaches with no, or limited flow records. DHM-TF has been shown by Gourley et al. (2012) and Cosgrove et al. (2012) to outperform FFG in a number of experimental settings.

Note that since DHM-TF establishes the thresholds on the basis of flow simulations, it requires high-resolution, accurate *historical* QPEs in addition to real-time QPEs and reliable hydrologic model representations. Historical QPEs can be subject to a number of deficiencies. In the US, the widely used multisensor QPEs (MQPEs) based on blending radar and gauge observations are known to exhibit a time varying bias (Zhang et al., 2011a). This trending bias has clear implications for hydrologic prediction. Zhang et al. (2011a) demonstrated that the water balance based on uncalibrated runs of a distributed hydrologic model exhibits a conspicuous upward trend between 1998 and the early-mid 2000. Zhang et al. (2011a) further experimented with re-adjusting the MQPEs using monthly gauge-based precipitation analysis. Though the authors found that this adjustment greatly reduced the trending bias in simulated water balance, they also suggested that the adjustment may be detrimental to resolving the magnitude of rainfall and flood peaks.

Bias and inaccuracy of both real-time and climatological QPE products, and the associated impacts on flood and flash flood prediction have both been active research areas (Smith et al., 1996; Young et al., 1999; Young et al., 2000; Hardegree et al., 2008; National Research Council, 2005; Oudin et al., 2006; Kitzmiller et al., 2011; Looper et al., 2012), so is calibration of hydrologic model (Duan et al., 1993; Gupta et al., 1998; Winsemius et al., 2009; Westerberg et al., 2011; Singh and Bàrdossy, 2012). Yet, to date, few studies have addressed the linkage between climatological adjustments and the accuracy of flash flood detection and prediction, though a few did examine the impacts of uncertainties in forcings and parameters. Oudin et al. (2006), for example, illustrated that some of the impacts of random and systematic errors in precipitation can be compensated by model calibration. The authors, however, did not explore climatological adjustment as a means to suppress the random and systematic errors. Zhang et al. (2011a)'s analysis on climatological adjustment focused on simulated water balance rather than on detection of flash flood events, and the authors did not address the relative effects of model calibration and adjustment. Strauch et al. (2012) attempted to account for the uncertainty in precipitation and parameters simultaneously by calibrating the model against an ensemble of precipitation inputs. Looper et al. (2012) assessed the compound effects of adjustment and model calibration. Neither of the latter two studies, however, delve into the mechanistic causes of precipitation errors and bias, nor did they address the impacts of calibration and adjustment on flood detection per se. The present study is intended to fill this gap by investigating isolated and compound impacts of climatological adjustment, both prior to and after model calibration, on the detection of flash floods over 19 watersheds in the eastern US. In this work, a long-term radar-gauge MQPE data set is adjusted using monthly gauge-based analysis, and both the original and adjusted MQPEs serve as inputs for calibrating a distributed hydrologic model. The streamflow simulation series from model with *a priori* and calibrated parameters are then used as the basis of the detection experiment. The work also complements a body of literature attempting to disentangle the impacts of structural and input errors on uncertainty in model prediction (e.g., Renard et al., 2010; Sun and Bertrand-Krajewski, 2013) by examining the differential impacts of calibration in the presence of nonstationary rainfall bias.

The remainder of the paper is organized as follows. Section 2 describes the data and methods. Section 3 summarizes the observations. Section 4 discusses the results, and Section 5 summarizes the key conclusions.

2. Data and methodology

2.1. Study watersheds

Selected for this study are 19 watersheds located within the service area of Mid-Atlantic River Forecast Center (Fig. 1; Table 1), whose drainage areas range from 84 to 2116 km^2. These watersheds are divided into two groups: (a) *small* watersheds – those with drainage area below 500 km^2 and (b) *large* watersheds, with drainage area above 500 km^2. The threshold of 500 km^2 was chosen as it roughly divides the watersheds with short response time and therefore prone to flash floods from those of much longer response time: synthetic unit hydrographs generated using a distributed hydrologic model (to be described later) indicate that all except one (**WASHB**) small basins in the former group are associated with time to peak (T_p) less than 6 h, whereas only one in the latter group does. The large watersheds are included in the analysis, as short-fused floods can also take place with an opportune combination of the spatio-temporal configuration of storm systems and antecedent soil moisture conditions (Zhang and Smith, 2003).

For each basin, flood events were identified from the hourly time series collected by the United States Geological Survey (USGS) using the 2-Y Averaged Recurrence Interval (ARI) values as thresholds; the former of these is widely considered a rough indicator of the over-bank flow (Reed et al., 2007). In this study, these ARI values are established based on the annual maximum hourly peak discharge using the standard procedure outlined in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982; Reed