UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

DECADAL DEVELOPMENT OF **CREST** HYDROLOGICAL MODEL FAMILY: REVIEW, IMPROVEMENTS, APPLICATIONS, AND OUTLOOK

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ZHI LI

Norman, Oklahoma

2022

DECADAL DEVELOPMENT OF **CREST** HYDROLOGICAL MODEL FAMILY:

REVIEW, IMPROVEMENTS, APPLICATIONS, AND OUTLOOK

A DISSERTATION APPROVED FOR THE SCHOOL OF CIVIL ENGIEERING AND ENVIRONMENTAL SCIENCE

BY THE COMMITTEE CONSISTING OF

Dr. Yang Hong, Chair

Dr. Randall Kolar

Dr. Pierre Kirstetter

Dr. Amy McGovern

Dr. Tiantian Yang

Dr. Yixin Wen

Dedication

This dissertation is dedicated to my family.

To my father, Yong Li, who has instilled me the grit, grind, and toughness.

To my mother, Yingbo Luo, who sacrificed to bring me up and support me in her continued belief.

To my grandparents, aunts, uncles, brothers and sisters, who have given their unselfish love to me. They always remind me of being a caring and giving person in life.

Acknowledgement

Growing up in a big family, I was influenced by how people care, help, and support each other. It comes naturally to me that one cannot go far without getting help from others. I am sincerely grateful for all the advices, help, and support from my committee members, colleagues, friends, and family. I wish to express my gratitude for their kind support along the way.

First, I am deeply indebted to my advisor and chair of my committee, Dr. Yang Hong, for his support both professionally and mentally. He has always been devoted and committed to working with students with his experience and vision, which manifests the true meaning of education. His encouragements and compliments have motivated me to challenge and break through myself. I would like to acknowledge the funding support from the Graduate College Hoving Fellowship and the HWS program.

To my committee members, Dr. Randall Kolar, Dr. Pierre Kirstetter, Dr. Amy McGovern, Dr. Tiantian Yang, and Dr. Yixin Wen, who have given their time to guide me and provide constructive criticism and feedback. They have set a life example for me to purse.

To Dr. Jonathan Gourley who has been a co-author for most of my publications. His professionals reflected in his suggestions and comments have made those publications possible. I appreciate him editing, polishing, and proofreading each of my publications.

To my colleagues, Dr. Mengye Chen, and Dr. Shang Gao, who are indispensable during my Ph.D. I am honored to have them with their practical advices, moral support, and challenges to help me grow.

To the HyDROS (Hydrometeorology and Remote Sensing Laboratory) group members, who have the belief in me. I am deeply inspired by their honor, courage, and commitment.

Lastly, I would like to extend my sincere thanks to my friends and family members who have been continuously supportive in this journey. The emotional support they brought has encouraged me to overcome every challenge in life and work.

Table of Contents

List of Figures

List of Tables

Abstract

A hydrological model is an indispensable tool in Earth system science and engineering operations to understand, predict, and manage water resources on Earth. The Coupled Routing and Excess Storage (CREST) model, released in 2011, is one such to simulate distributed hydrologic states and fluxes at variable scales. Over the last decade, CREST model has been actively under development and applied by different sectors to tackle water-related problems worldwide. This dissertation is dedicated to expanding the capacity of CREST model from three main fronts: (1) hydrologic data, (2) model development, and (3) applications. To start, the decadal development and applications of CREST model family were reviewed to lay the foundation for my contribution (Chapter 1). First, uncertainties in hydrologic input data were evaluated comprehensively for three state-of-the-science precipitation datasets derived from in-situ instruments, ground weather radar, and satellites during extreme events (Chapter 2); then a 120-year CONUS-wide flood database was compiled into a unified format as a validation source for models and hydroclimatic research (Chapter 3). From the model development front, a Hydrologic&Hydraulic (H&H) framework was developed to empower flood inundation mapping capacity for CREST (Chapter 4); furthermore, the re-infiltration, an important yet often ignored hydrologic process during the flooding period, was incorporated to improve the more realistic rainfall-runoff modeling representation (Chapter 5). To further improve the model efficiency, a vector-based CREST model was developed that can achieve 10x speedup for a continental-scale simulation, as well as improved model accuracy (Chapter 6). Finally on the model application, the high-resolution CREST model was applied in quantifying future US floods in a warmer climate: flood flashiness is becoming 7.9% higher for the continent (Chapter 7); and extreme rainfall and floods are becoming more frequent, widespread, and less seasonal (Chapter 8). The final Chapter 9 summarizes the contributions to the CREST model family development, outlooks, and general remarks for advancing our understanding of hydrologic science and engineering.

Graphical Abstract:

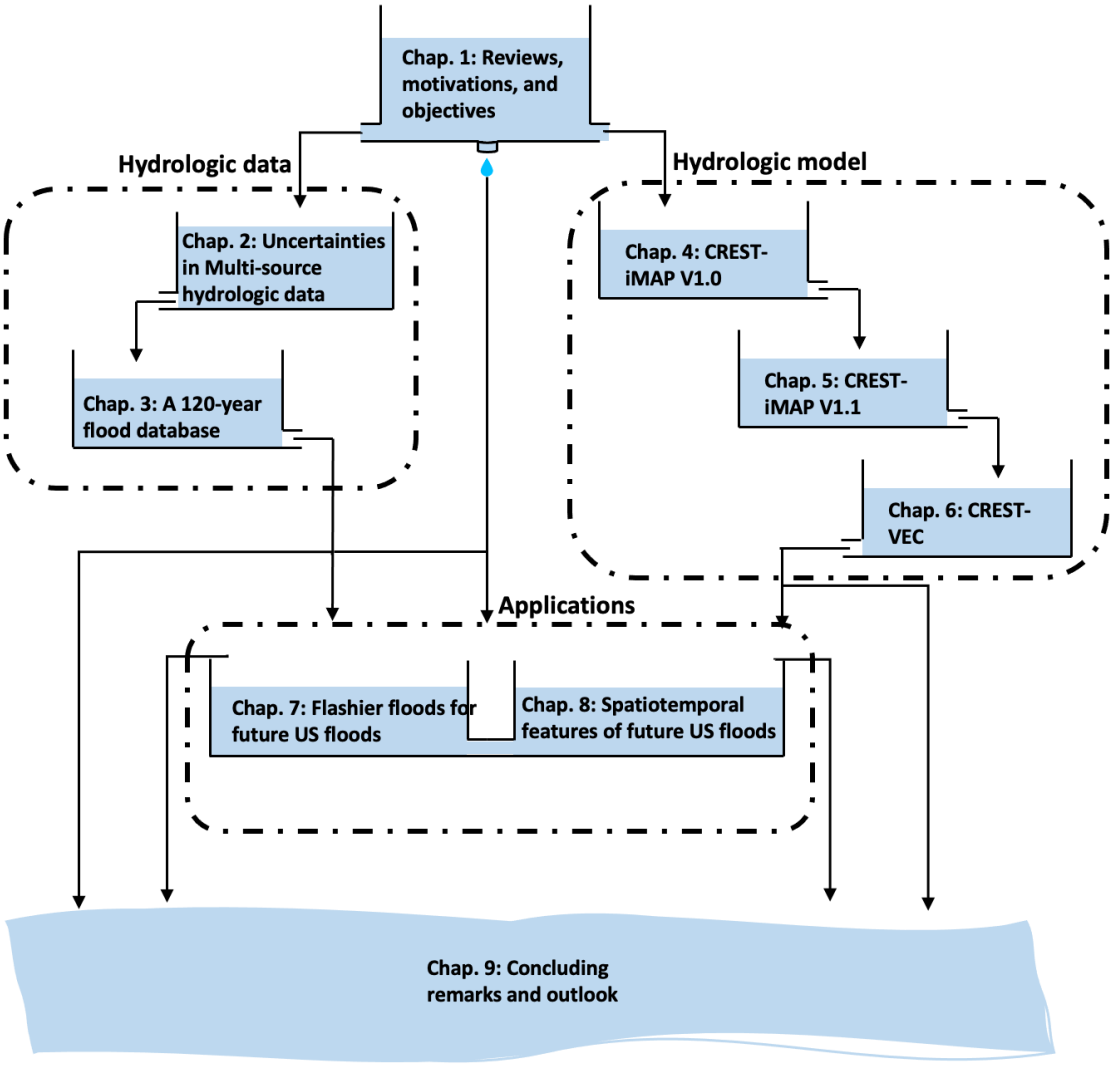


Figure 1. Flowchart of dissertation chapters.

Chapter 1:

Reviews, motivations, and objectives

1 Reviews, motivations, and objectives

Floods are one of the most devastating and deadliest natural hazards. In the United States, for instance, eight of the ten costliest weather disasters were floods between 1980 and 2019 (see https://www.ncdc.noaa.gov/billions/events). There is a median value of 81 flood fatalities per year from 1959 to 2005 (Ashley and Ashley, 2008), and almost ten percent of the flash floods have resulted in agricultural and economic losses beyond $100,000 (U.S. dollars) per event (Gourley et al., 2017). Several historical tropical cyclones and hurricanes (e.g., Hurricane Katrina, Hurricane Harvey, Tropical Cyclone Imelda, etc.) resulted in tremendous economic losses caused by coastal or inland flooding. The most damaging hurricane Katrina affected nine states and resulted monetary loss of $168.8 billion dollars.

In a pressing need to predict floods, a wealth of modelling approaches has been proposed since 1850s by an Irish engineer Mulvany T. J. who was considered as the founding father of hydrologic models (Mulvany, 1850). Darcy (1856) later laid the foundation for groundwater hydrology via his brilliant experiments relating flow speed through different soil medium and pressures. But only after the World War II, with revolutionary computers, hydrologic models to simulate the entire water cycle become feasible. One of the early pioneering works is the Stanford Watershed Model (Crawford and Linsley, 1966) which is the first distributed hydrologic model applied to large catchment. Nowadays, hydrologic models have essentially become one component of Global Climate Model (GCM) or Land Surface Model (LSM) to resolve large-scale water cycle.

The CREST model, namely Coupled Routing and Excess STorage, developed at the University of Oklahoma and National Aeronautics and Space Administration (NASA), is a distributed hydrologic model aiming to but not limited to predicting riverine floods in a fast implementation. The debut of CREST model was ten years ago in 2011. While celebrating its 10th anniversary, we in the following section review the development of CREST model and its applications to water-related issues (Figure 1.1). As of date (07/01/2022), CREST model has been widely accepted by the community, and 87 publications have used CREST model family for a range of applications, receiving 3524 citations in total. Figure 1.2 highlights the time series of number of publications and citations for CREST models, indicating a prevailing trend in the recent years (peaking in 2017). In the following paragraphs, we follow the chronologic order and focus

on the mainstem which is the main package to reveal the relative changes to the source. At last, we summarize the limitations and outlooks for the CREST model family.

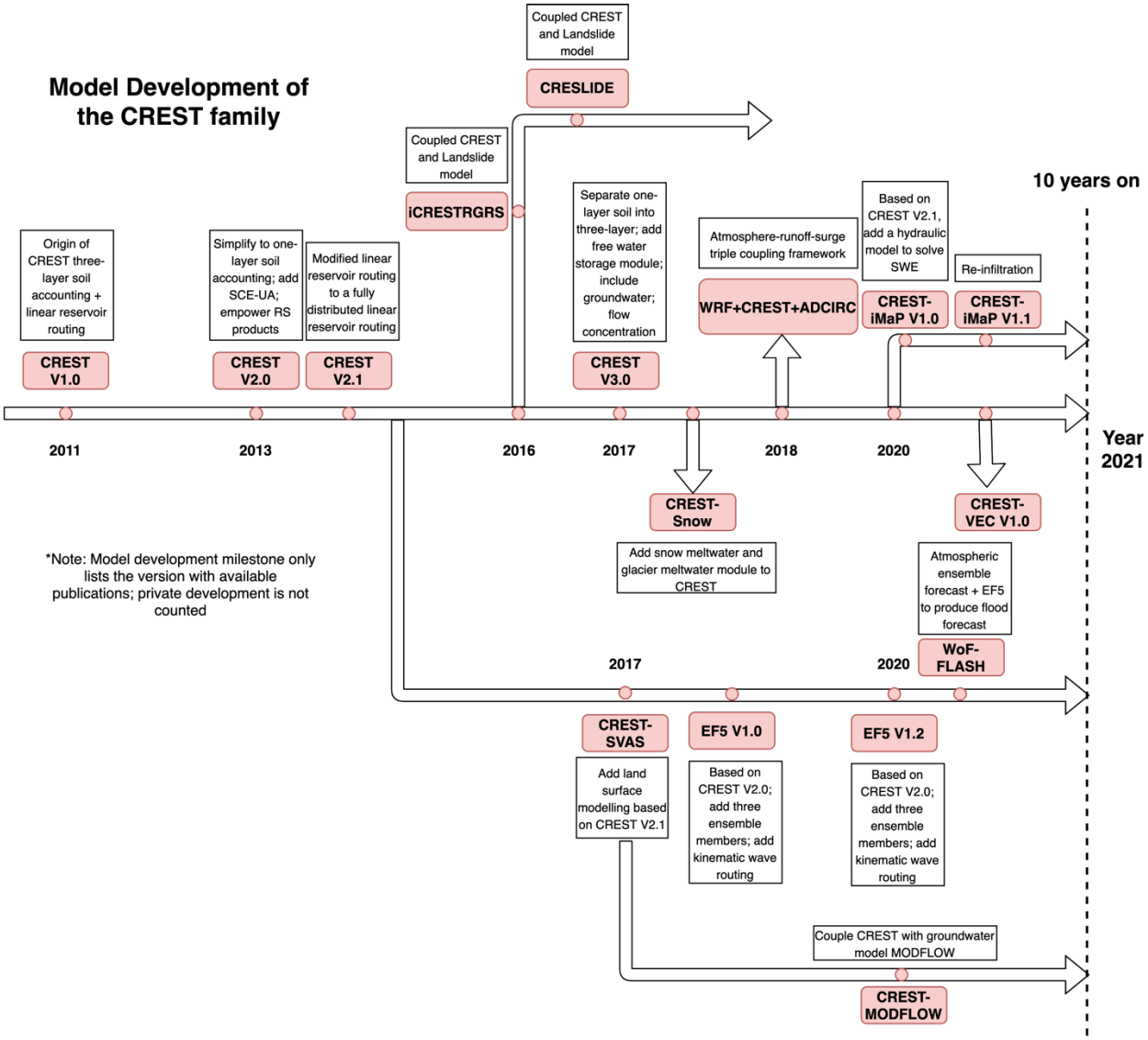


Figure 1.1 Evolution of CREST model

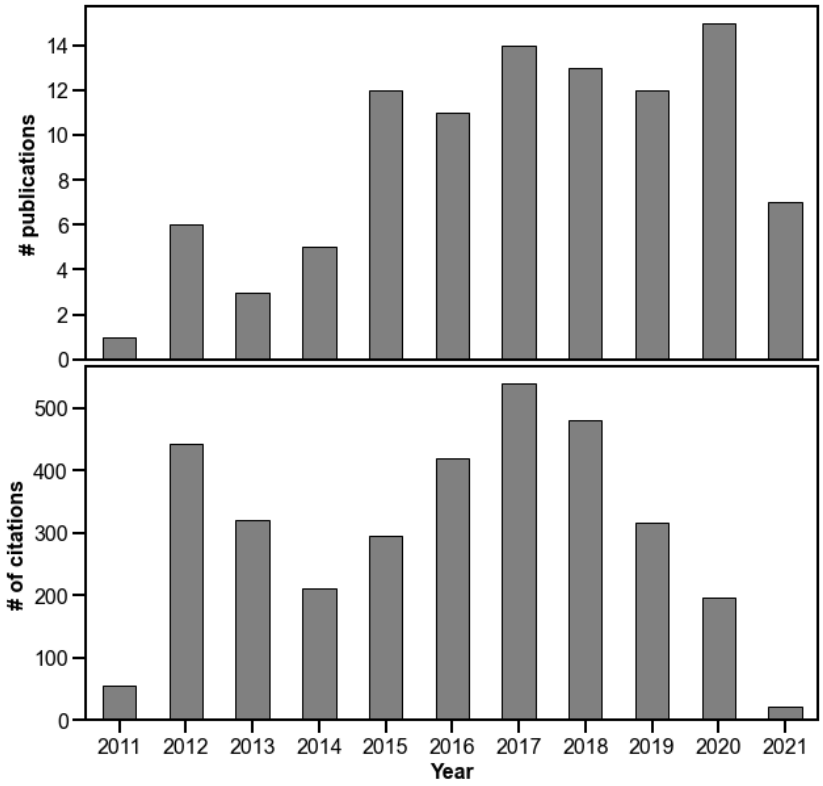


Figure 1.2 Time series of number of publications and citations.

1.1 CREST model mainstem evolution

There are three iterations of CREST mainstems in 2011 (CREST V1.0), 2013 (CREST V2.0 and CREST V2.1), and 2017 (CREST V3.0). The building blocks for the CREST model are conceptual bucket model at each grid cell to represent the spill-fill nature of water storage and movement. Figure 1.3 depicts the evolution of CREST model structure in these iterations.

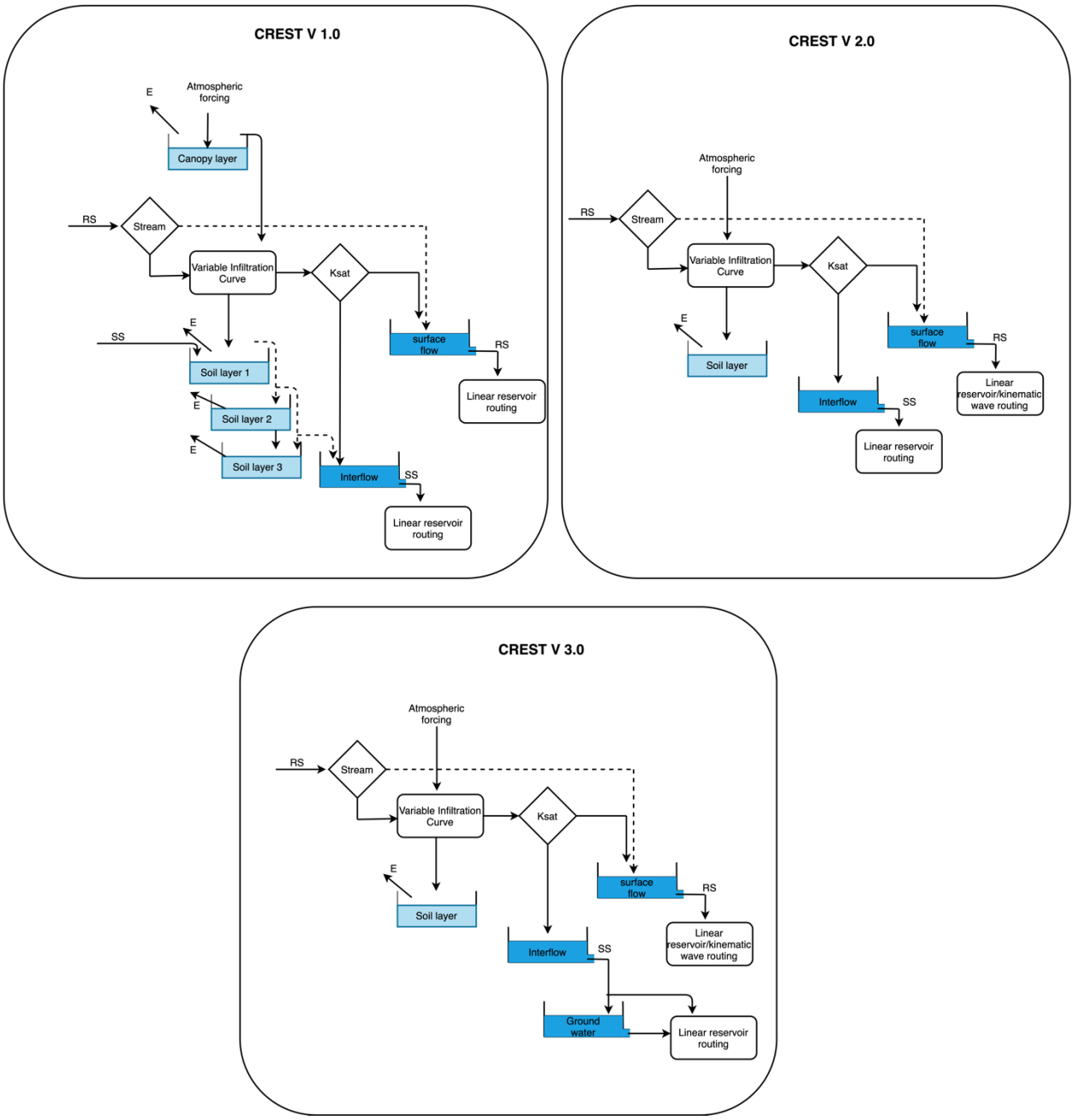


Figure 1.3 Model structure of three versions of CREST model.

1.1.1 CREST Version 1.x

CREST V1 features three important characteristics: 1) distributed rainfall-runoff generation process and cell-to-cell routing, 2) coupled routing and runoff generation mechanisms, and 3) sub-grid variability of soil moisture capacity (Wang et al., 2011). Back to the time when remotely sensed precipitation was prime (US radar-derived precipitation product and TRMM

product), CREST model features the seamless integration of distributed rainfall and runoff process, making it a standout out of various hydrologic models. The resultant atmospheric forcing, deficit of rainfall and evaporation, encounters canopy layer whose capacity (*CIC*) is parameterized as a linear function of Leaf Area Index (*LAI*), vegetation coverage (*d*), and coefficient of land cover (k_c) (Dickinson, 1989). Then the remanent water P_{soil}, if exceeding *CIC*, infiltrates into soil layers along with incoming interflow (SS) from upstream cells.

$$CIC = k_c \times d \times LAI \tag{1.1}$$

$$P_{soil} = P - (CIC - CI), \tag{1.2}$$

Where CI is the intercepted water in canopy.

The Variable Infiltration Curve (VIC) describes water partition in soils, which is a classic way founded in the Xinanjiang model and adopted by the University of Washington VIC model (Zhao, 1992; Liang et al., 1996). In CREST V1, three soil layers are incorporated to characterize shallow (0-0.5 meters), deep (0.5-2 meters), and very deep (>2 meters) soils. The infiltration process is mathematically represented in the following equations.

$$i = i_m[1 - (1 - A)^{1/b_i}] \tag{1.3}$$

$$i_m = W_m(1 + b_i) \tag{1.4}$$

$$W_m = W_{m1} + W_{m2} + W_{m3}, \tag{1.5}$$

Where i is the point infiltration capacity while i_m is the maximum infiltration capacity. A is the fractional area of the cell and b_i is the exponent of the curve. W_m is the maximum soil water capacity which is the sum of water capacity at three soil layers: W_{m1}, W_{m2}, W_{m3}. The infiltration water (*I*) is the deficit of maximum soil water capacity (W_m) and soil water state (*W*) if available soil water P_{soil} plus point infiltration capacity (i) is larger than maximum infiltration capacity (i_m). Otherwise, it follows the exponent function shown below.

$$I = (W_m - W) + W_m[1 - \frac{i+P_{soil}}{i_m}]^{1+b_i} \tag{1.6}$$

The runoff generation in CREST V1 is primarily based upon excess saturation runoff in which interflow is produced from the ratio of excess rain (*R*) and soil water (P_{soil}) and soil hydraulic conductivity (*K*). The overland runoff (RS) is the deficit of excess rain (R) and interflow (SS).

$$R = P_{soil} - I \tag{1.7}$$

$$SS = K\frac{R}{P_{soil}} \tag{1.8}$$

$$RS = R - SS \tag{1.9}$$

The potential evapotranspiration (PET) is required as a direct input to CREST V1, which is the potential amount of water can be evaporated, determined by atmospheric condition such as temperature, wind, and radiation, etc. The actual evaporation is calculated by distributing PET in canopy and soil layers. The rule of redistribution is laid on successive depletion (from canopy to soil) and water depth in each bucket. First, it judges whether canopy layer water is enough for PET. If yes, ET is coming all from canopy water. Otherwise, soil water contributes to the remaining part. ET in the shallow soil layer (first layer) is similar to the canopy layer based on depletion rule, as they both are open to air. If the first soil layer and canopy layer jointly cannot fulfill the PET amount, the second and/or third soil layer start to contribute based on exponent separation until it is completely depleted.

$$E_p = E_r\sqrt{\frac{W}{W_m}}, \tag{1.10}$$

Where E_p is evaporation in the second or third soil layer, and E_r is the deficit of PET and canopy layer plus first soil layer evaporation.

The routing scheme in original CREST model is linear reservoirs applied to surface flow and subsurface flow with different parameterization. The linear reservoir scheme accounts for sub-grid-scale routing, meaning that the grid spacing of DEM is immune to routing results. It makes CREST model scalable, especially suitable for global applications (Wu et al., 2012). The surface flow reservoir (SF) at time t+1 receives upstream flow (Qs) at time t and excess surface runoff (RS) and produces outflow (O_{SF}) for downstream routing. All the while, the interflow reservoir (SI) receives subsurface flow (SS) and deep soil spill flow (ES) if any. Eq. 1.11-1.12 represent these two processes in first-order Ordinary Differential Equation (ODE) which advances in time. For in-channel routing, we calculate the time of concentration at jth pixel (T^j) with Eq. 1.12, given the distance grid (l^j) and slope grid (S^j) as input and runoff velocity coefficient K_X as parameter. From there, we can solve how many grids runoff at jth pixel can advance with given time step. This is a Lagrangian approach that often applied in fluid mechanics. Finally, the discharge (Q) equals the sum of surface runoff and subsurface runoff multiplied by drainage area (A) and divided by time step (Δt).

$$SF_{t+1} = SF_t + RS_t + Q_s \times t \quad (1.11)$$

$$SI_{t+1} = SI_t + SS_t + ES_t \quad (1.12)$$

$$T^j = \frac{l^j}{K_X\sqrt{S^j}} \quad (1.13)$$

$$Q = (SF + SI) \times A/\Delta t \quad (1.14)$$

Some minor updates have been released such as CREST V1.6 which becomes scalable for both basin-scale (Khan et al., 2011) and global-scale applications (Wu et al., 2012). The input parameters are derived from field survey data, land cover maps, and vegetation coverage. Notably, the CREST V1.x considers lumped parameter set (Table 1) which is averaged over the study domain.

Table 1.1 Inputs and parameters used in CREST V1.x

Symbols	Description	Source	Unit
P	Gridded rainfall data	Remote sensing, weather/climate model, and gauges	mm/timestep
PET	Gridded potential evaporation data	Remote sensing, weather/climate model, and gauges	mm/timestep
LAI	Leaf Area Index	Remote sensing	m^2/m^2
DEM	Digital Elevation Model	Remote sensing/survey	M
FDIR	Flow direction	Derive from DEM	N/A
FAC	Flow Accumulation	Derive from DEM	Cells or km^2
S	Slope	Derive from DEM	Degree
l	Distance between cells	Derive from DEM	M
W_{m1}	Maximum soil water capacity at soil layer 1	Soil survey	mm
W_{m2}	Maximum soil water capacity at soil layer 2	Soil survey	mm

W_{m3}	Maximum soil water capacity at soil layer 3	Soil survey	mm
b_1	Exponent parameter of the VIC model at soil layer 1	Soil survey	N/A
b_2	Exponent parameter of the VIC model at soil layer 2	Soil survey	N/A
b_3	Exponent parameter of the VIC model at soil layer 3	Soil survey	N/A
Ksat	Mean saturated hydraulic conductivity	Soil survey	mm/hr
d	Vegetation coverage	Remote sensing	N/A
coeM	The overland runoff velocity coefficient	N/A	N/A
expM	The overland flow speed exponent	N/A	N/A
coeR	The multiplier used to convert overland flow to channel flow speed	N/A	N/A
coeS	The multiplier used to convert overland flow speed to interflow speed	N/A	N/A
K_S	Surface runoff velocity coefficient	N/A	m/s
K_I	Subsurface runoff velocity coefficient	N/A	m/s

1.1.2 CREST Version 2.x

Two years after the debut of the first version, Xue et al. (2013) introduced a next generation of CREST model – CREST V2.0, empowering more advanced features. The paradigm has been shifted thereafter towards fast implementation and dedicated to operational flood systems. Thus,

several major changes have been made to the previous generation including: (1) activating distributed parameters over regions where remotely sensed data is available, (2) the replacement of three soil layers to one bulk soil layer, (3) including impervious area ratio parameter to emulate fast runoff generation, (4) including a rainfall multiplier parameter to overcome rainfall bias, (5) automatic calibration using the SCE-UA algorithm (Duan et al., 1992), (6) parallel computing, (7) modular design framework.

Precipitation product, satellite product in particular, is biased because of systematic error due to instruments, infrequent sampling, and algorithms, especially in the early stage of development (Xue et al., 2013; Tang et al., 2016). To account for the systematic bias, Xue et al. (2013) introduced this RainFact parameter applied to remote sensing precipitation (Eq. 1.15). As such, it is a prerequisite to conduct a pre-analysis for precipitation product in use to derive this parameter.

$$P_{rain} = RainFact \times P, \tag{1.15}$$

where P_{rain} is the corrected rainfall rates prior to reach ground.

CREST 1.x generates surface runoff based upon saturation excess runoff, which is not readily suited for urban environment where pavements and built ups imped rainfall infiltration. On the other hand, flooding is mostly devastating and disastrous in populated urban regions. Urbanization has proved to sharpen hydrograph – increase magnitude and reduce flood-rising time (Smith et al., 2002; Yang et al., 2011; Zhang et al., 2018). It is of particular importance to consider such factor in urban environment. In doing so, Xue et al. (2013) splits P_{rain} into fast surface runoff (FS) and P_{soil} in Eq. 1.12-1.13, where IM is the impervious area ratio (%). Higher IM leads to greater amounts of fast runoff.

$$FS = IM \times P_{rain} \tag{1.16}$$

$$P_{soil} = P_{rain} - FS \tag{1.17}$$

The one bulk soil layer in lieu of three layers is attempting to represent soil within five meters from the surface (Figure 1.3). The rationale behind is threefold: (1) ease data preparing process; (2) reduce the number of model parameters; (3) speed up model implementation. Previous parameterization in VIC model (Eq. 1.3-1.6) has changed to Eq. 1.11, thereby reducing numbers of parameters in VIC model from 7 to 3.

$$i = i_{max} \times [1 - (1 - W/W_m)^{1/B}] \tag{1.18}$$

After CREST V2.0, Shen et al., (2017) published CREST V2.1, dedicated to improving the existing routing schemes. The proposed Fully Distributed Linear Reservoir Routing (FDLRR) scheme replaces the Quasi-Distributed Linear Reservoir Routing (QDLRR) in CREST V2.0. Specifically, the newer version addresses the underestimation of in-channel flow and discontinuous flow after storms because QDLRR only considers donor-to-receiver relationship while ignoring water along the routing nodes between donor node and receiver node. As pointed out in Shen et al. (2017), the QDLRR does not account for water continuity along river reaches, resulting in a substantial bias. Figure 1.4 depicts the difference between the two, where the newer version redistributes water along its pathway and thus conserves mass. The Eq. 1.14 for discharge at jth pixel has been modified to Eq. 1.19, where SF_{via} and SI_{via} are the surface runoff passing through node j.

$$Q^j = \frac{(SF_{out}+SI_{out})A_g+\sum SF_{via}A_g+\sum SI_{via}A_g}{\Delta t} \tag{1.19}$$

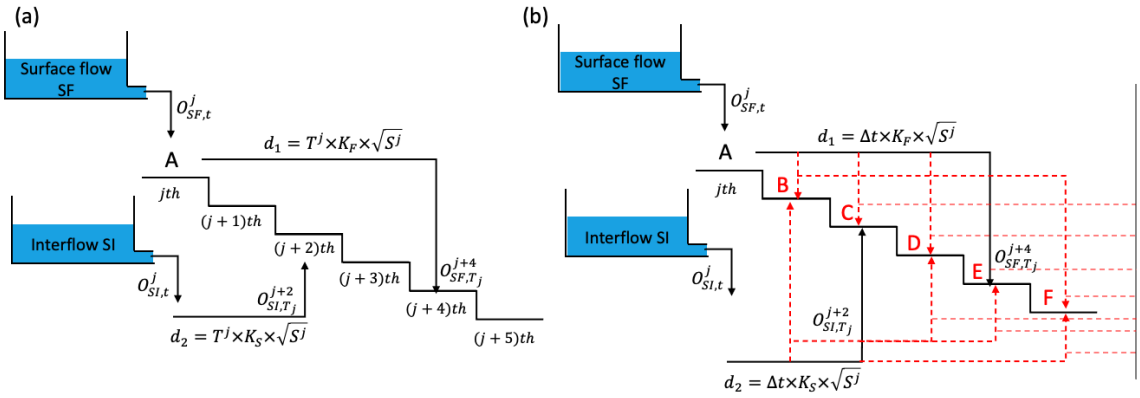


Figure 1.4 Illustration of routing scheme in CREST V1.x/2.0 (a) and V2.1 (b) with differences highlighted in red arrows.

1.1.3 CREST Version 3.x

Although there are several publications indicating development of CREST V3.0 and applications, they did not fully make codes available for open access (Kan et al., 2017; Li et al., 2018). In this study, we rearrange the versions of CREST to only consider public available models. The intention is to promote open-source models and community efforts to improve model performance. CREST V3.0 has adopted a conceptual groundwater model to further improve model performance, which makes it not only a flood-centric tool but also suitable for predicting water scarcity (Gao et al., 2021).

The conceptual bucket model inherits the classic fill-spill strategy, which is the current implementation in the US National Water Model (NWM) V2.1. A schematic plot is shown in Fig. 1.5. The ground water bucket receives recharge from upper layer soil or vadose zone and evaporates water in the meantime. The spill scheme generates lateral groundwater flow when the groundwater level (Z_{gw}) reaches its maximum level ($Z_{gw,max}$) in Eq. 1.20. The opening at the bottom of the bucket generates slow and continuous flow with Eq. 1.21, where parameter *GWC* is a multiplier and *GWE* is an exponent factor. Compared to CREST V2.x, we added three parameters $Z_{gw,max}$, *GWC*, *GWE*, which inferred from aquifer depth or groundwater table data if any.

$$Q_{gw,exc} = recharge + Z_{gw} - Z_{gw,max} \tag{1.20}$$

$$Q_{gw,exp} = GWC\left(\frac{Z_{gw}}{Z_{gw,max}} - 1\right)^{GWE} \tag{1.21}$$

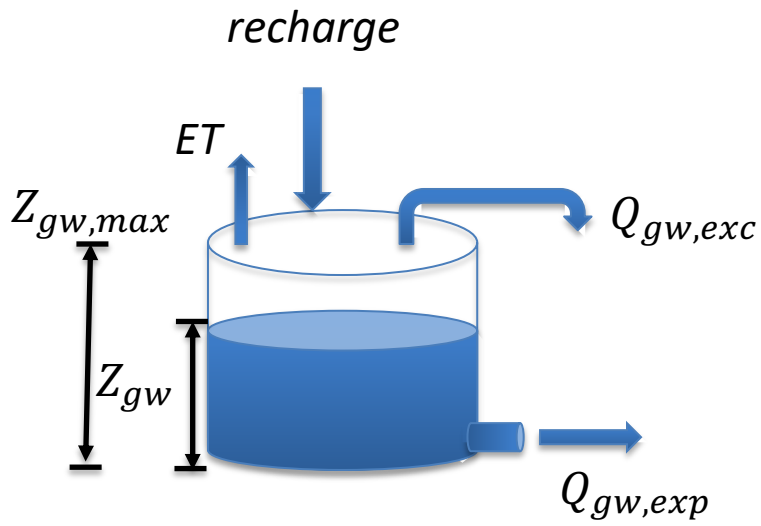


Figure 1.5 A conceptual groundwater model used in CREST V3.0

1.2 Ensemble framework for flash flood forecast (EF5)

Apart from CREST mainstem, an important framework (EF5) has been developed since 2016 as a joint effort by the University of Oklahoma and NOAA National Severe Storms Laboratory (NSSL) to acknowledge model structure uncertainties on top of parameter uncertainties. Since then, the EF5 has become operational across National Weather Service (NWS) in the US for flash flood forecasting by local Weather Forecast Offices (Clark et al., 2017; Gourley et al., 2017; Vergara et al., 2016). All the codes are written in C++ and work across common platforms

including Windows, Linux, and Mac OS X (Flamig et al., 2020). The current version EF5 V1.3 supports nine different models to adapt for a broader working environment, ranging from snowmelt model, to water balance model, and to inundation model as shown in Figure 1.6.

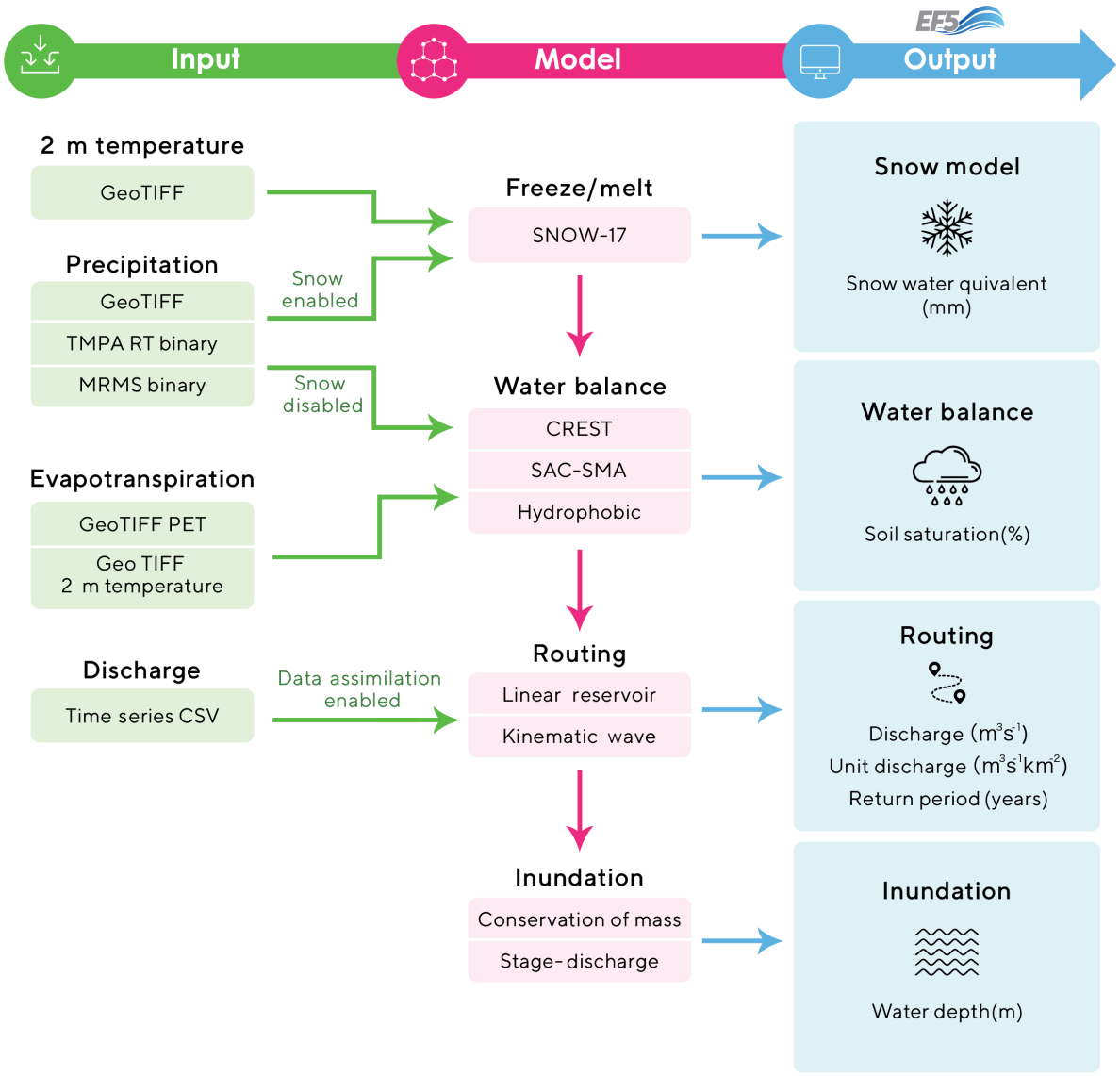


Figure 1.6 Schematic view of the EF5 framework. Adapted from Flamig et al. (2020) under copyright Creative Common Attribution 4.0 License held by the Copernicus Publications.

Previous CREST model lacks a snow module which becomes essential when simulating streamflow in snow-dominant regions such as the Rockies in the US. The EF5 framework adopts a Snow Accumulation and Ablation model (Snow-17) to parameterize snow process (Anderson, 1976, 2006). The Snow-17 model takes air temperature and precipitation as inputs to calculate

energy and water exchange. The accumulation in snow cover happens when precipitation falls, and temperature below freezing threshold. The energy exchange occurs between the snow-air interface to determine melting amount. Last, outflow from snow cover is computed in company with rain-on-snow.

The water balance models in EF5 framework include – CREST V2.1 (Xue et al., 2013), CREST V3.0, SAC-SMA (Koren et al., 2004), and hydrophobic model to generate ensemble predictions. The SAC-SMA is a classic hydrologic model which is still used by the US National Weather Service, but it has been modified to a distributed run in EF5. Readers are referred to Koren et al. (2004) and Yilmaz et al. (2008) for detailed description of SAC-SMA model. The hydrophobic model is the simplest model that requires no land surface parameters, treating surface as completely impervious. The essence of hydrophobic model is to not only provide a "worst case" scenario, but also reflect bias in forcing data.

In EF5, routing options have been enriched, from pure linear reservoir routing in CREST to kinematic wave routing. Compared to linear reservoir routing, kinematic wave is a more physically-based model, which solves Saint-Venant equation based on mass conservation and momentum conservation in Eq. 1.22-1.23.

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} = 0 \tag{1.22}$$

$$\frac{1}{A}\frac{\partial Q}{\partial t} + \frac{1}{A}\frac{\partial}{\partial x}\frac{Q^2}{A} + g\frac{\partial y}{\partial x} - gS_0 + gS_f = 0, \tag{1.23}$$

where Q is the volumetric flow, A is the cross-sectional area, x is longitudinal distance, y is the water depth, S_0 is the bottom slope, and S_f is the friction slope. The terms in the left hand of Eq. 1.23 represent local acceleration, convective acceleration, pressure force, gravity force, and friction force, respectively. While instead of solving full dynamic wave model, kinematic wave model ignores the acceleration terms and forcing terms to seek for faster computational speed. Because it is assumed that flow is uniform and steady, Eq. 1.23 is finally simplified to Eq. 1.24 where the flow is dependent on cross sectional area alone with parameters α and β. Substituting to Eq. 1.22, we arrive at the final form (Eq. 1.25) for kinematic wave model. As an implicit method, direct solution for eq.25 does not exist. In EF5, we used an iterative method to solve flow Q by substituting it repeatedly until it converges within a threshold.

$$Q = \alpha A^{\beta} \tag{1.24}$$

$$\frac{\partial Q}{\partial x} + \alpha\beta Q^{\beta-1}\frac{\partial Q}{\partial t} = inflow \quad (1.25)$$

To enhance flood hazard forecast capacity, two simple inundation mapping schemes were incorporated into the EF5 framework – mass-conserving and stage-discharge model. The mass-conservation model utilizes the flow output and calculates the total river volume at a given time step for basins of interest. Then, the exact amount of water is redistributed by pouring it into the basin based on the DEM data. The stage-discharge model is dependent upon a rating curve which relates discharge to river stage at river channel pixel by an exponential function. The river stage deducts the elevation in DEM to result in height which is compared to the Height Above Nearest Drainage map derived from the DEM data to have the inundation map. The later method is also adopted by the operational National Water Model V2.1 for generating flood inundation at large scales.

Aside from model developments, EF5 integrates several utilities to facilitate flood modeling. First, model calibration is a central component for hydrologic models, especially for lacking a-priori knowledge about the study area. The Differential Evolution Adaptive Metropolis (DREAM) algorithm by Vrugt et al. (2008) is adopted in lieu of SCE-UA algorithm used in CREST. The DREAM algorithm is an adaptive Markov Chain Monte Carlo (MCMC) algorithm that runs several chains in parallel to search for optimal parameter space. Second, a data assimilation module is enabled by direct insertion of observed streamflow at a gauge location, serving as a boundary condition for hydrologic simulations. Third, EF5 supports direct output of flood event period from simulated streamflow by fitting Log-Pearson type III distribution which is the most commonly used flood frequency distribution in the US.

1.3 Coupling with other models

1.3.1 Weather forecast model

Floods, flash floods in particular, are largely driven by atmospheric conditions such as synoptic and convective weather types (Merz et al., 2021; Ning et al., 2020). As such, coupling weather forecast model with hydrologic model holds great promise in predicting flood risks and protecting human lives and assets. A community model - Weather Research Forecast (WRF) - has been widely applied for short-term meteorological simulations. The WRF is configurable with multiple physics, dynamics, and parameterization to be applied under varying atmospheric

conditions. It has been served as the backbone for several US operational weather forecasts such as the High-Resolution Rapid Refresh (HRRR) and Rapid Refresh.

Blanton et al. (2018) for the first time coupled WRF with CREST to generate ensemble predictions of streamflow during Hurricane Isabel. The imperfect knowledge of model physics and parameterization schemes yields diverging results in precipitation and land surface components. In this study, they perturbed the model settings based on different initial conditions and physical schemes to output series of streamflow values. It is found the 6-day forecast generally capture the streamflow trends and peaks.

The NOAA Warn-on-Forecast (WoF)-FLASH program initiated an integrated framework to couple numerical weather forecasts with EF5 framework to operate probabilistic short-term flash flood prediction over the contiguous US in real time (Yussouf et al., 2020). The experimental WoF uses 36 ensemble members generated by the HRRR ensemble (HRRRE) at every hour and re-initializes system runs every 15 mins to assimilate readily available data such as the MRMS (Multi-Radar Multi-Sensor) radar reflectivity, radial velocity, cloud water path, and other meteorological observations (Jones et al., 2016). These probabilistic flash flood forecast products embrace the uncertain nature of weather forecasts and enable earlier detection of damaging flash floods in a systematic way.

1.3.2 Snowmelt model

CREST model is weak in snow and glacier melting and accumulations, limiting its application in high-mountain regions such as the Rockies in the US and the third pole – Tibetan Plateau. Chen et al. (2017) introduced the snow-resolving version of the CREST (CREST-Snow) that readily fills the gap. In the first stage, the total precipitation is separated into solid and liquid phases from a temperature criterion. Snow or glacier accumulates with an increase in solid precipitation. All the while, land surface receives snow meltwater and glacier ice meltwater in addition to rainfall-runoff process.

As hydrologic process becomes complex, more inputs and parameters are assigned therein. CREST-snow requires 19 parameters, nine of which are for snowmelt and glacier melt process. In addition to precipitation data, CREST-snow necessitates remotely sensed air temperature for determining precipitation phase and snow/glacier melts. It can be envisioned that model calibration is becoming challenging with many newly added parameters. Chen et al. (2017) proposed a two-

stage calibration strategy: the snow parameters are calibrated in the first stage and then others in the second stage. The rationale behind is that snowmelt process is relatively independent to rainfall-runoff process, and more importantly, there are intermediate observational states available for model calibration. For instance, the Snow Cover Area (SCA) and Snow Water Equivalent (SWE) can be retrieved from MODIS (Moderate Resolution Imaging Spectroradiometer) because of high reflectance of snow. In doing so, calibration burden is much alleviated.

1.3.3 Land surface model

One of the limitations for the CREST model, as compared to other land surface models such as Noah and Noah-MP models, is the poor representation of land surface process, especially for snow- and forest-covered regions where runoff generation mechanisms are more complex. Motivated by these challenges, Shen & Anagnostou (2017b) introduced a framework to solve the full cycle of Soil-Vegetation-Atmosphere-Snow process, named CREST-SVAS. The CREST-SVAS generates runoff by solving both water balance and energy balance in a more complex form. Correspondingly, more model inputs are required such as radiation, temperature, wind, etc. In fact, CREST-SVAS pushed the envelope that makes CREST model adaptable in different environment.

1.3.4 Groundwater model

The simple conceptual groundwater (GW) module in the CREST is subject to uncertainties in parameterizations, boundary conditions, and hydrologic stresses. In contrast, physical GW models offer more realistic simulation by solving physical governing equations. The MODFLOW (MODular 3-D finite-difference) model with NWT solver is promising in GW model communities, especially for its open interface to be coupled with other models. Khadim et al. (2020) for the first time coupled CREST model with MODFLOW and applied to the Blue Nile Basin in Ethiopia. The CREST model provides boundary conditions – recharge rate after infiltration and streamflow to force the MODFLOW. They also manifest the deviation of GW depth simulated from physical model and global conceptual model in the same region.

1.3.5 Landslide model

The CREST model has been extended throughout the ten years to couple with other models for predicting water-related natural hazards. He et al. (2016) published a coupled hydrological-geotechnical framework for landslide prediction, called CRESTLIDE (Coupled Routing Excess

Storage and Slope-Infiltration-Distributed Equilibrium). They couple CREST model with the SLIDE model which computes the slope stability as a factor of safety. The relatively more accurate subsurface hydrologic variables from CREST were feed to the SLIDE model in an integrated way. CRESTSLIDE assumes shallow depth landslide and simplifies the infinite-slope equation.

Unlike the simplified landslide model mentioned above, Zhang et al. (2016) coupled CREST model with a landslide model – Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability (TRIGRS) which uses the analytical Richard's equation to solve infinite-slope equation, termed as iCRESTRIGRS. Similar to the SLIDE model, TRIGRS only provides a simple representation of soil infiltration and runoff generation, placing CREST model upstream of TRIGRS can better predict landslide, as both infiltration rates and runoff are nontrivial for landslide models. Though via a loose coupling, the integrated system is seamlessly executed in a distributed manner at every time step. Specifically, this system takes distributed runoff and infiltration rates from CREST model and passes to TRIGRS to output pore-pressure and factor of safety for each grid cell.

1.4 Applications and outreach

Over the past ten years, CREST model family has been used worldwide for a range of purposes. Figure 1.7 depicts the distribution of CREST model applications (from published journal articles) by country. Apparently, US and China have used the CREST model the most, reaching 27 and 28, respectively. They also have similar focuses with regard to modeling purposes, with flood simulation ranked as the top concern, and followed by evaluating hydrologic utilities of satellite precipitation products. In fact, these applications are on par with the original objectives of developing CREST model. Notably, in China, due to strict policies with data sharing, streamflow data is hard to acquire. In this context, researchers have been devoted to developing strategies combining remote sensing data and frameworks to predict ungauged basins. These strategies made CREST model calibratable and verifiable in every corner of the world. There are also some instances using CREST model in developing countries especially in Africa, owing to the capacity building project in corporation with the NASA. In the following sections, we break down this topic into specific applications to illustrate the capacity and popularity of the CREST model family.

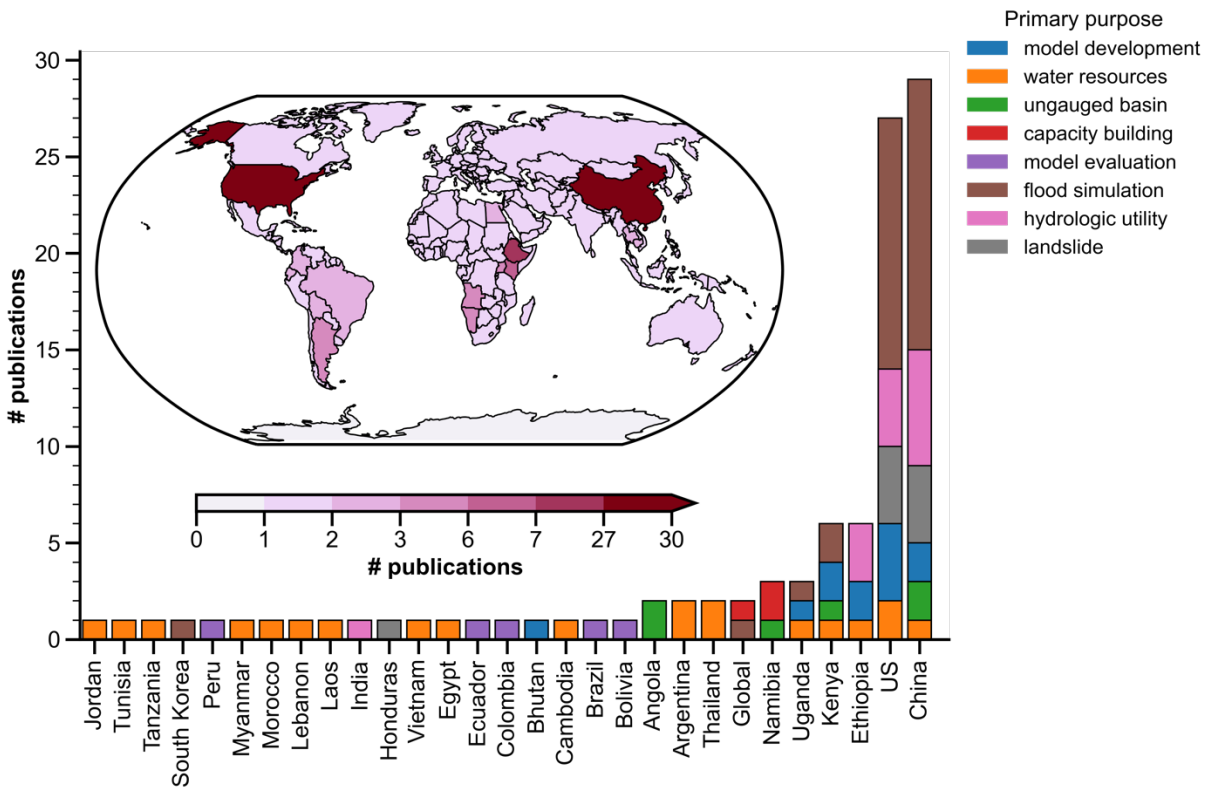


Figure 1.7 Map of CREST model applications in the world, grouped by primary purposes.

1.4.1 Evaluating hydrologic utility of the remote sensing precipitation products

As CREST model is dedicated to injecting remote sensing precipitation dataset, one of its famous applications is to assess and compare different precipitation products. In tradition, evaluation of precipitation products is done by comparing them to more trustworthy rain gauges. However, such approach is effective only at gauge locations which only cover a football-court size compared to global land surface (Kidd et al., 2017). Alternatively, the hydrologic utility of satellite or radar remote sensing offers another angle inspecting the errors propagating from precipitation to streamflow. Because streamflow is a basin-integrated result, some trivial changes (including location, magnitude, and timing) among precipitation products can be accumulated and magnified in the simulated streamflow.

Meng et al. (2014) provided some insights into the applicability of Tropical Rainfall Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) product, concluding that the daily rainfall rates by TMPA does not have much hydrologic utility but monthly rates agree quite well. Tang et al. (2016) assessed the hydrologic utility of after-launch Global Precipitation Measure

(GPM) mission, compared to its processor TRMM using CREST model in China. He experimented two sets of model parameters: (1) Static parameter set calibrated from rain gauges and (2) dynamic parameter set that is calibrated for each GPM and TRMM dataset at their optima. The hydrologic simulation results present a clear outperformance of GPM IMERG to TRMM-based products, which is not discernable if only by precipitation evaluations. Other evaluation studies can be found in Tang et al. (2015), Lakew et al. (2017), Li et al. (2017), Jiang et al. (2017), Ma et al. (2018), Sun et al., (2018), Yuan et al. (2018), Ma et al. (2018), Ma et al. (2019), Chen et al. (2020).

1.4.2 Flood simulations and forecasting

Of all the applications of CREST model, flood (flash flood) simulation and prediction are regarded as the primary utility in the following reasons. First, CREST model considers full scope of runoff generation schemes, including fast surface runoff by impervious area, saturation-excess runoff, and infiltration excess-runoff. Second, the rapid production by CREST delivers timely flood information to stakeholders, which makes it operational in several frameworks. Last, CREST model is scalable, approved to work in spatial scales ranging from 10 meters to 1000 km, which targets local flooding or pluvial flooding caused by intense rainfall and large-scale flooding or fluvial flooding.

The operational framework FLASH (flash.ou.edu) demonstrates the capacity of CREST/EF5 for continental flood simulation, driven by the highest available resolution observational radar precipitation product – MRMS at 1km (Gerard et al., 2021; Gourley et al., 2017). Funded by the NASA SEVIR project, CREST/EF5 has been deployed in African countries such as Namibia, Ethiopia, Kenya, Uganda, and surrounding areas to aid local decision making (Clark et al., 2017; Macharia et al., 2010; Yami et al., 2021). Wu et al. (2012) prototyped the global flood monitoring framework GFMS using TRMM-era precipitation product as input and CREST as hydrologic model to run at 3-hourly and 0.125-degree. Taking advantage of advanced weather forecast datasets in the US, there are experiments and frameworks using Quantitative Precipitation Forecast (QPF) and Probabilistic QPF (PQPF) data to alert local residents (Martinaitis et al., 2017; Yussouf et al., 2021). Zhang et al. (2015) evaluated the flood detectability of Global Hydrological Prediction System (GHPS) forced by Global Forecast System (GFS) and CREST model.

Besides these operational services worldwide, there are individual efforts to validate the efficacy of CREST simulated flood events. Khan et al., (2010) evaluated the inundation mapping

scheme in CREST and compared to the Earth observing satellite – Advance Microwave Scanning Radiometer for the Earth Observing System (AMSR-E), providing insights in using remote sensing data to validate hydrologic simulation in ungauged basins. Similarly, Gao et al. (2017) used the satellite-based precipitation products and CREST model to develop flood frequency analysis in ungauged basins.

1.4.3 Water resources management

Water resources management is a central component for human society. Groundwater, soil moisture, surface water, snow, and ice are five main components for available water resources on Earth. The dynamic changes of the five states have been of considerable interest in the recent years as (1) a warmer climate accelerates terrestrial water cycle (Huntington, 2006) and (2) more regions are under water stress due to climate change and anthropogenic influences (Rodell et al., 2018). Hydrologic models such as CREST are well equipped to simulate the dynamic changes of terrestrial water storage over a long time for water resources management.

Khan et al. (2011) first applied the CREST model to inspect the hydroclimatology of Lake Victoria in Africa. From there, they not only proved that CREST simulated states agree well with observations, but also analyzed the hydrologic behavior in such a region. Habib et al. (2012) found the reduced streamflow and groundwater by CREST simulation cannot sustain societal needs in the MENA region – Morocco, Tunisia, Egypt, Lebanon, and Jordan. Rossi & Ares (2015, 2016) applied the CREST-IRRIGATION model to improve agricultural irrigation efficiency in Argentina. Similarly, the soil moisture simulated by CREST can be used to inform crop yield (Yang et al., 2021). Li et al. (2019) used the streamflow by CREST to examine the benefits of water sharing for transboundary Lancang-Mekong River. Besides, there are applications to simulate terrestrial water components (Gathecha, 2015; Lazin et al., 2020; Ren et al., 2015; Rossi & Ares, 2016; Shen & Anagnostou, 2017).

1.4.4 Hydrologic prediction in ungauged basins

The Prediction in Ungauged Basins (PUB) is a long-standing challenge for the hydrologic community (Hrachowitz et al., 2013). Unfortunately, there is a decline in hydrometric stations over the recent years due to insufficient funding, inadequate institutional frameworks, criticisms of operating a network, and other factors such as wars or hazards (Mishra & Coulibaly, 2009). Under

these pressures, developing new prediction methods that embrace remote sensing data is imperative. For hydrologic simulations, model calibration and validation are necessary, although we hope models to be calibration-free if all physical processes are correctly represented.

The CREST model is born to integrate remote sensing products and ready to be calibrated or evaluated by them. Back in 2012, Khan et al. (2012) provided implications in PUB by evaluating CREST model against AMSR-E derived inundation maps. Later, Zhang et al. (2014) extracted streamflow signal from TRMM and AMSR-E to calibrate CREST model. Chen et al. (2017) calibrated CREST model against SWE (first stage) and GRACE (Gravity Recovery and Climate Experiment; second stage) to ungauged basins. Han et al. (2020) used the CREST model in accompany with satellite altimetry data to infer reservoir operation curves, filling data voids in ungauged basins or by reluctant data-sharing policy. It is worth mentioning that, Huang et al. (2020) proposed a framework CREST-RS which relies no stream gauges in calibration and validation stages. They tested two schemes: one was water level from satellite altimetry and the other one was river width plus rating curve parameters to emulate SWOT-like (the Surface Water and Ocean Topography) data. Although not been experimented yet, the remote sensing soil moisture product such as the SMAP (Soil Moisture Active Passive) can be used to calibrate and evaluate model in ungauged basins as well.

1.4.5 Capacity building and outreach

As mentioned above, CREST model developments and applications are funded by the NASA SEVIER program to help local government agencies build up their skills for model simulation and decision-making. It was in 2011 when CREST development team delivered its first training course to the East Africa node of NASA and the U.S. Agency for International Development's (USAID) SERVIR project, covering Kenya, Tanzania, Uganda, and surrounding areas. In the consecutive year, with the release of CREST V2.0, the development team spawned a new training course, aiming to help local specialists upgrade CREST model. Ever since, training courses and hands-on workshops were offered annually to local specialists, although some others attended from everywhere else (Clark et al., 2017, Yami et al., 2021). The participants from private sectors, government agencies, environmental scientists, and academics were able to pick up modeling skills owing to the concise model structure, minimum parameter preparation, well-structured model interface, adaptive cross platforms, and low-requiring computational resources.

Beyond hands-on and in-person workshops, our scientists made tutorial video publicly accessible online from basic trainings such as model description and simulation to more advanced trainings such as flood frequency analysis and inundation mapping. The upfront web interface (ef5.ou.edu) hosts a body of learning materials, videos, and model codes. We strongly support open-source model codes and request community efforts to extend model capacities. Upon marking the tenth-year anniversary of CREST model family, we operate a brand-new website (https://crest-family.readthedocs.io/en/latest/) where a collection of models codes (from CREST V1.x, V2.x, and V3.0), training materials, applications, and publications are made fully online.

1.5 Motivations and Objectives

The review of CREST model family exposes some opportunities for further improving the model. **First**, there lacks a systematic evaluation of remotely sensed precipitation products used by CREST model. Given the flood generating mechanisms, precipitation is still the main trigger of floods in most regions, especially for flash floods. Acquiring an accurate and high-resolution precipitation dataset is critical for streamflow simulation and flood prediction, as many studies demonstrated the error propagation (Meng et al., 2014; Tang et al., 2016). Meanwhile, the lack of flood data over the U.S. even the globe has struggled modelers in model validation, although there are proxy data available. It is worthwhile to synthesize and compile all available flood catalogue readily for flood models. **Second**, although the capacity of CREST model for water resources management has been strengthened over the years, flood prediction still serves the backbone of the CREST model. As compared to model developments for a wide range of applications, flood-related model tools have been relatively stagnant, which certainly has rooms to improve. The FLASH system, for instance, has been running in operation since 2017 at 1km grid spacing, but lately some frameworks have increased their resolutions to 250 meters (NWM). Upgrading model to a higher resolution is definitely worth in terms of flash flood prediction that occurs at street level or neighborhood level. Besides, to keep pace with increasing computational power, we have the ability to push the envelope by incorporating more advanced technology and models. **Last**, from the application perspective, future (flash) floods under a changing climate have been hindered by the lack of high-resolution climate forcing and hydrologic models. Given the improved data and model outlined in previous sections, we propose an integrated system to investigate how future US floods are becoming with respect to flashiness and spatiotemporal characteristics. The intention of

this dissertation is to re-cruise CREST model to its original driveway – flood prediction and research.

In light of these pressing needs and opportunities, the overarching goal of this dissertation is to improve CREST model physics and efficiency with incorporated advanced precipitation data to produce street-level flood estimates. My **contributions** in this dissertation are arranged in three sections: (1) **Hydrologic data for flood research**, (2) **Developments of multi-purpose flood simulation models across scales**, and (3) **Applications**. Each section contains several chapters for specific topics. Within **Section 1**, **Chapter 2** discusses deficiencies and uncertainties in high-resolution precipitation data to provide insights for evaluating their hydrologic utilities. In **Chapter 3**, we introduce a publicly available flood database across 120 years in the US, dedicated for verification of flood simulations, hydroclimatology, and disaster risk assessments. Within **Section 2**, we underline the development of hydrologic-hydraulic (H&H) models (**Chapter 4** and **5**) and vector-enabled routing (**Chapter 6**). **Section 3** mainly focuses on applications, centered in future floods under a warmer climate. For instance, **Chapter 7** reveals changes in future flash floods over the contiguous United States, and **Chapter 8** depicts their future spatiotemporal changes. Finally, **Chapter 9**, summarize all abovementioned works and envisions outlook for a future flourish of the CREST model family.

1.6 Reference

Anderson, E. A. (1976). A point energy and mass balance model of a snow cover. NOAA Technical Report, NWS 19, 1976.

Anderson, E. A. (2006). Snow accumulation and ablation model–SNOW-17. US National Weather Service, Silver Spring, MD, 61.

Blanton, B., Dresback, K., Colle, B., Kolar, R., Vergara, H., Hong, Y., Leonardo, N., Davidson, R., Nozick, L., & Wachtendorf, T. (2020). An Integrated Scenario Ensemble-Based Framework for Hurricane Evacuation Modeling: Part 2—Hazard Modeling. Risk analysis, 40(1), 117-133.

Chen, X., Long, D., Hong, Y., Zeng, C., & Yan, D. (2017). Improved modeling of snow and glacier melting by a progressive two-stage calibration strategy with GRACE and multisource data: How snow and glacier meltwater contributes to the runoff of the Upper Brahmaputra River basin?. Water Resources Research, 53(3), 2431-2466.

Chen, M., Nabih, S., Brauer, N. S., Gao, S., Gourley, J. J., Hong, Z., Kolar, R., & Hong, Y. (2020). Can remote sensing technologies capture the extreme precipitation event and its cascading hydrological response? A case study of Hurricane Harvey using EF5 modeling framework. Remote Sensing, 12(3), 445.

Clark, R. A., Flamig, Z. L., Vergara, H., Hong, Y., Gourley, J. J., Mandl, D. J., Frye, S., Handy, M., & Patterson, M. (2017). Hydrological modeling and capacity building in the Republic of Namibia. Bulletin of the American Meteorological Society, 98(8), 1697-1715.

Crawford NC, Linsley RK (1966) Digital simulation in hydrology: Stanford watershed simulation-IV. Technical Report 39, Stanford University, Palo Alto

David, C. H., Maidment, D. R., Niu, G. Y., Yang, Z. L., Habets, F., & Eijkhout, V. (2011). River network routing on the NHDPlus dataset. Journal of Hydrometeorology, 12(5), 913-934.

Dickinson, R. E. 1989. A regional climate model for the western united states. Climate Change, 15(1): 383–422.

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. Water resources research, 28(4), 1015-1031.

Gao, Z., Long, D., Tang, G., Zeng, C., Huang, J., & Hong, Y. (2017). Assessing the potential of satellite-based precipitation estimates for flood frequency analysis in ungauged or poorly gauged tributaries of China's Yangtze River basin. Journal of hydrology, 550, 478-496.

Gao, S., Chen, M., Li, Z., Cook, S., Allen, D., Neeson, T., Yang, T., Yami, T., & Hong, Y. (2021). Mapping Dynamic Non-Perennial Stream Networks Using High-Resolution Distributed Hydrologic Simulation: A Case Study in the Upper Blue River Basin. Journal of Hydrology, 126522.

Gathecha, H. M. (2015). Reconstruction of streamflow into Lake Naivasha using crest model and remote sensed rainfall and evapotranspiration (Master's thesis, University of Twente).

Gerard, A., Martinaitis, S. M., Gourley, J. J., Howard, K. W., & Zhang, J. (2021). An overview of the performance and operational applications of the MRMS and FLASH systems in recent

significant urban flash flood events. Bulletin of the American Meteorological Society, 1-29.

Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P. E., Clark, R. A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Earlingis, J. M., Hong, Y., & Howard, K. W. (2017). The FLASH project: Improving the tools for flash flood monitoring and prediction across the United States. Bulletin of the American Meteorological Society, 98(2), 361-372.

Habib, S., Kfouri, C., & Peters, M. (2012, July). Water information system platforms addressing critical societal needs in the MENA region. In 2012 IEEE International Geoscience and Remote Sensing Symposium (pp. 2767-2770). IEEE.

Han, Z., Long, D., Huang, Q., Li, X., Zhao, F., & Wang, J. (2020). Improving reservoir outflow estimation for ungauged basins using satellite observations and a hydrological model. Water Resources Research, 56(9), e2020WR027590.

He, X., Hong, Y., Vergara, H., Zhang, K., Kirstetter, P. E., Gourley, J. J., Zhang, Y., Qiao, G., & Liu, C. (2016). Development of a coupled hydrological-geotechnical framework for rainfall-induced landslides prediction. Journal of Hydrology, 543, 395-405.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., ... & Cudennec, C. (2013). A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrological sciences journal, 58(6), 1198-1255.

Huang, Q., Long, D., Du, M., Han, Z., & Han, P. (2020). Daily continuous river discharge estimation for ungauged basins using a hydrologic model calibrated by satellite altimetry: Implications for the SWOT mission. Water Resources Research, 56(7), e2020WR027309.

Huntington, T. G. (2006). Evidence for intensification of the global water cycle: Review and synthesis. Journal of Hydrology, 319(1-4), 83-95.

Jiang, S., Zhang, Z., Huang, Y., Chen, X., & Chen, S. (2017). Evaluating the TRMM Multisatellite Precipitation Analysis for extreme precipitation and streamflow in Ganjiang River basin, China. Advances in Meteorology, 2017.

Jones, T. A., Knopfmeier, K., Wheatley, D., Creager, G., Minnis, P., & Palikonda, R. (2016). Storm-scale data assimilation and ensemble forecasting with the NSSL experimental Warn-on-Forecast system. Part II: Combined radar and satellite data experiments. Weather and Forecasting, 31(1), 297-327.

Kan, G., Tang, G., Yang, Y., Hong, Y., Li, J., Ding, L., He, X., Liang, K., He, L., Li, Z., Hu, Y., & Cui, Y. (2017). An improved coupled routing and excess storage (CREST) distributed hydrological model and its verification in Ganjiang River Basin, China. Water, 9(11), 904.

Khadim, F. K., Dokou, Z., Lazin, R., Moges, S., Bagtzoglou, A. C., & Anagnostou, E. (2020). Groundwater modeling in data scarce aquifers: the case of Gilgel-Abay, Upper Blue Nile, Ethiopia. Journal of Hydrology, 590, 125214.

Khan, S. I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., Brakenridge, G. R., Policelli, F., Habib, S., & Irwin, D. (2010). Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria basin: Implications for hydrologic prediction in ungauged basins. IEEE Transactions on Geoscience and Remote Sensing, 49(1), 85-95.

Khan, S. I., Adhikari, P., Hong, Y., Vergara, H., F Adler, R., Policelli, F., Irwin, D., Korme, T., & Okello, L. (2011). Hydroclimatology of Lake Victoria region using hydrologic model and satellite remote sensing data. Hydrology and Earth System Sciences, 15(1), 107-117.

Khan, S. I., Hong, Y., Vergara, H. J., Gourley, J. J., Brakenridge, G. R., De Groeve, T., Flamig, Z., Policelli, F., & Yong, B. (2012). Microwave satellite data for hydrologic modeling in ungauged basins. IEEE Geoscience and Remote Sensing Letters, 9(4), 663-667.

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges?. Bulletin of the American Meteorological Society, 98(1), 69-78.

Koren, V., Reed, S., Smith, M., Zhang, Z., & Seo, D. J. (2004). Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. Journal of Hydrology, 291(3-4), 297-318.

Lakew, H. B., Moges, S. A., & Asfaw, D. H. (2017). Hydrological evaluation of satellite and reanalysis precipitation products in the Upper Blue Nile Basin: A case study of Gilgel Abbay. Hydrology, 4(3), 39.

Lazin, R., Shen, X., Koukoula, M., & Anagnostou, E. (2020). Evaluation of the hyper-resolution model-derived water cycle components over the upper Blue Nile Basin. Journal of Hydrology, 590, 125231.

Li, N., Tang, G., Zhao, P., Hong, Y., Gou, Y., & Yang, K. (2017). Statistical assessment and hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang River basin. Atmospheric research, 183, 212-223.

Li, Z., Yang, Y., Kan, G., & Hong, Y. (2018). Study on the applicability of the Hargreaves potential evapotranspiration estimation method in CREST distributed hydrological model (version 3.0) applications. Water, 10(12), 1882.

Li, D., Zhao, J., & Govindaraju, R. S. (2019). Water benefits sharing under transboundary cooperation in the Lancang-Mekong River Basin. Journal of Hydrology, 577, 123989.

Liang, X., Lettenmaier, D. P. and Wood, E. F. 1996. One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. J. Geophys. Res., 101(D16): 21,403–21,422.

Ma, Q., Xiong, L., Liu, D., Xu, C. Y., & Guo, S. (2018). Evaluating the temporal dynamics of uncertainty contribution from satellite precipitation input in rainfall-runoff modeling using the variance decomposition method. Remote Sensing, 10(12), 1876.

Ma, Q., Xiong, L., Xia, J., Xiong, B., Yang, H., & Xu, C. Y. (2019). A censored shifted mixture distribution mapping method to correct the bias of daily IMERG satellite precipitation estimates. Remote Sensing, 11(11), 1345.

Ma, Z., Tan, X., Yang, Y., Chen, X., Kan, G., Ji, X., Lu, H., Long, J., Cui, Y., & Hong, Y. (2018). The first comparisons of IMERG and the downscaled results based on IMERG in hydrological utility over the Ganjiang river Basin. Water, 10(10), 1392.

Macharia, D., Korme, T., Policelli, F., Irwin, D., Adler, B., & Hong, Y. (2010, October). SERVIR-Africa: Developing an integrated platform for floods disaster management in Africa. In 8th

International Conference African Association of Remote Sensing of the Environment (AARSE) (No. M10-1035).

Martinaitis, S. M., Gourley, J. J., Flamig, Z. L., Argyle, E. M., Clark III, R. A., Arthur, A., Smith, B. R., Erlingis, J. M., Perfater, S., & Albright, B. (2017). The HMT multi-radar multi-sensor hydro experiment. Bulletin of the American Meteorological Society, 98(2), 347-359.

Meng, J., Li, L., Hao, Z., Wang, J., & Shao, Q. (2014). Suitability of TRMM satellite rainfall in driving a distributed hydrological model in the source region of Yellow River. Journal of Hydrology, 509, 320-332.

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U., & Macdonald, E. (2021). Causes, impacts and patterns of disastrous river floods. Nature Reviews Earth & Environment, 1-18.

Mishra, A. K., & Coulibaly, P. (2009). Developments in hydrometric network design: A review. Reviews of Geophysics, 47(2).

Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., Viger, R., Markstorm, S. L., Hay, L. E., Woods, R., Arnold, J. R., & Brekke, L. D. (2016). mizuRoute version 1: a river network routing tool for a continental domain water resources applications. Geoscientific Model Development, 9(6), 2223-2238.

Mulvany TJ (1850) On the use of self-registering rain and flood gauges. In: Proceedings of the Institute Civil Engineers 4(2): 1–8, Dublin, Ireland.

Ning, L., Zhan, C., Luo, Y., Wang, Y., & Liu, L. (2019). A review of fully coupled atmosphere-hydrology simulations. Journal of Geographical Sciences, 29(3), 465-479.

Ren, G., Zhan, Y., Ren, Y., Chen, Y., Wang, T., Liu, Y., & Sun, X. (2015). Spatial and temporal patterns of precipitation variability over mainland China: I: Climatology. Advances in water science, 26(3), 299-310.

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., & Lo, M. H. (2018). Emerging trends in global freshwater availability. Nature, 557(7707), 651-659.

Rossi, M. J., & Ares, J. O. (2015). Efficiency improvement in linear-move sprinkler systems through moderate runoff–runon control. Irrigation science, 33(3), 205-219.

Rossi, M. J., & Ares, J. O. (2016). Overland flow from plant patches: Coupled effects of preferential infiltration, surface roughness and depression storage at the semiarid Patagonian Monte. Journal of Hydrology, 533, 603-614.

Shen, X., Hong, Y., Zhang, K., & Hao, Z. (2017a). Refining a distributed linear reservoir routing method to improve performance of the CREST model. Journal of hydrologic engineering, 22(3), 04016061.

Shen, X., & Anagnostou, E. N. (2017b). A framework to improve hyper-resolution hydrological simulation in snow-affected regions. Journal of hydrology, 552, 1-12.

Smith, J. A., Baeck, M. L., Morrison, J. E., Sturdevant-Rees, P., Turner-Gillespie, D. F., & Bates, P. D. (2002). The regional hydrology of extreme floods in an urbanizing drainage basin. Journal of Hydrometeorology, 3(3), 267-282.

Sun, W., Ma, J., Yang, G., & Li, W. (2018). Statistical and hydrological evaluations of multi-satellite precipitation products over Fujiang river basin in humid southeast China. Remote Sensing, 10(12), 1898.

Tang, G., Li, Z., Xue X., Hu, Q., Yong, B., & Hong, Y. (2015). A study of substitutability of TRMM remote sensing precipitation for gauge-based observation in Ganjiang River basin. Advances in Water Science, 26(3), 340-346.

Tang, G., Zeng, Z., Long, D., Guo, X., Yong, B., Zhang, W., & Hong, Y. (2016). Statistical and hydrological comparisons between TRMM and GPM level-3 products over a midlatitude basin: Is day-1 IMERG a good successor for TMPA 3B42V7? Journal of Hydrometeorology, 17(1), 121-137.

Vergara, H., Kirstetter, P. E., Gourley, J. J., Flamig, Z. L., Hong, Y., Arthur, A., & Kolar, R. (2016). Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the Conterminous United States. Journal of Hydrology, 541, 421-433.

Vrugt, J. A., Ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences and Numerical Simulation, 10(3), 273-290.

Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., Alder, R. F., Policelli, F., Habib, S., Irwin, D., Limaye, A. S., Korme, T., & Okello, L. (2011). The coupled routing and excess storage (CREST) distributed hydrological model. Hydrological sciences journal, 56(1), 84-98.

Wu, H., Adler, R. F., Hong, Y., Tian, Y., & Policelli, F. (2012). Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. Journal of Hydrometeorology, 13(4), 1268-1284.

Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., Dorji, C., & Chen, S. (2013). Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins?. Journal of Hydrology, 499, 91-99.

Yami, T. L., Gao, S., Chen, M., Li, Z., Vergara, H., Clark, R. R., Labriola, L., Gourley, J., Wara, C., Mandl, D., & Hong, Y. (2021). CREST/EF5 capacity building to enhance resilience to hydrodynamic disasters in emerging regions. African Journal of Environmental Science and Technology, 15(6), 230-242.

Yang, G., Bowling, L. C., Cherkauer, K. A., & Pijanowski, B. C. (2011). The impact of urban development on hydrologic regime from catchment to basin scales. Landscape and Urban Planning, 103(2), 237-247.

Yang, M., Wang, G., Lazin, R., Shen, X., & Anagnostou, E. (2021). Impact of planting time soil moisture on cereal crop yield in the Upper Blue Nile Basin: A novel insight towards agricultural water management. Agricultural Water Management, 243, 106430.

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. Water Resources Research, 44(9).

Yuan, F., Wang, B., Shi, C., Cui, W., Zhao, C., Liu, Y., Ren, L., Zhang, L., Zhu, Y., Chen, T., Jiang, S., & Yang, X. (2018). Evaluation of hydrological utility of IMERG Final run V05 and TMPA 3B42V7 satellite precipitation products in the Yellow River source region, China. Journal of Hydrology, 567, 696-711.

Yussouf, N., Wilson, K. A., Martinaitis, S. M., Vergara, H., Heinselman, P. L., & Gourley, J. J. (2020). The coupling of NSSL warn-on-forecast and FLASH systems for probabilistic flash flood prediction. Journal of Hydrometeorology, 21(1), 123-141.

Zhao, R. J. 1992. The Xianjiang model applied in China. J. Hydrol., 135(3): 371–381.

Zhang, Y., Hong, Y., Gourley, J. J., Wang, X., Brakenridge, G. R., De Groeve, T., & Vergara, H. (2014). Impact of assimilating spaceborne microwave signals for improving hydrological prediction in ungauged basins. Remote sensing of the terrestrial water cycle, 206, 439.

Zhang, Y., Hong, Y., Wang, X., Gourley, J. J., Xue, X., Saharia, M., Ni, G., Wang, G., Huang, Y., Chen, S., & Tang, G. (2015). Hydrometeorological analysis and remote sensing of extremes: Was the July 2012 Beijing flood event detectable and predictable by global satellite observing and global weather modeling systems? Journal of Hydrometeorology, 16(1), 381-395.

Zhang, K., Xue, X., Hong, Y., Gourley, J. J., Lu, N., Wan, Z., Hong, Z., & Wooten, R. (2016). iCRESTRIGRS: a coupled modeling system for cascading flood–landslide disaster forecasting. Hydrology and Earth System Sciences, 20(12), 5035-5048.

Zhang, W., Villarini, G., Vecchi, G. A., & Smith, J. A. (2018). Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. Nature, 563(7731), 384-388.

Chapter 2

Cross-examination of Similarity, Difference and Deficiency of Gauge, Radar, and Satellite precipitation Measuring Uncertainties for Extreme Events using Conventional Metrics and Multiplicative Triple Collocation

Publication

Li, Z., Chen, M., Gao, S., Hong, Z., Tang, G., Wen, Y., Gourley, J.J., Hong, Y. (2020). Cross-Examination of Similarity, Difference and Deficiency of Gauge, Radar and Satellite Precipitation Measuring Uncertainties for Extreme Events Using Conventional Metrics and Multiplicative Triple Collocation. Remote Sensing, 12, 1258. https://doi.org/10.3390/rs12081258

2 Chapter 2. Cross-examination of Similarity, Difference and Deficiency of Gauge, Radar, and Satellite precipitation Measuring Uncertainties for Extreme Events using Conventional Metrics and Multiplicative Triple Collocation

Abstract

Quantifying uncertainties of precipitation estimation, especially in extreme events, could benefit early warning of water-related hazards like flash floods and landslides. Rain gauges, weather radars, and satellites are three mainstream data sources used in measuring precipitation but have their own inherent advantages and deficiencies. With a focus on extremes, the overarching goal of this study is to cross-examine the similarities and differences of three state-of-the-art independent products (Muti-Radar Muti-Sensor Quantitative Precipitation Estimates, MRMS; National Center for Environmental Prediction gridded gauge-only hourly precipitation product, NCEP; Integrated Multi-satellitE Retrievals for GPM, IMERG), with both traditional metrics and the Multiplicative Triple Collection (MTC) method during Hurricane Harvey and multiple Tropical Cyclones. The results reveal that: (a) the consistency of cross-examination results against traditional metrics approves the applicability of MTC in extreme events; (b) the consistency of cross-events of MTC evaluation results also suggests its robustness across individual storms; (c) all products demonstrate their capacity of capturing the spatial and temporal variability of the storm structures while also magnifying respective inherent deficiencies; (d) NCEP and IMERG likely underestimate while MRMS overestimates the storm total accumulation, especially for the 500-year return Hurricane Harvey; (e) both NCEP and IMERG underestimate extreme rainrates (>= 90 mm/h) likely due to device insensitivity or saturation while MRMS maintains robust across the rainrate range; (g) all three show inherent deficiencies in capturing the storm core of Harvey possibly due to device malfunctions with the NCEP gauges, relative low spatiotemporal resolution of IMERG, and the unusual “hot” MRMS radar signals. Given the unknown ground reference assumption of MTC, this study suggests that MRMS has the best overall performance. The similarities, differences, advantages, and deficiencies revealed in this study could guide the users for emergency response and motivate the community not only to improve the respective sensor/algorithm but also innovate multidata merging methods for one best possible product, specifically suitable for extreme storm events.

2.1 Introduction

Extreme rainfall events associated with flash floods (Smith et al., 2011; Villarini and Smith 2010), landslides (Kirschbaum et al., 2012; Hong et al., 2007), debris flow (Dong et al., 2010) often lead to tremendous damages to properties and fatalities. According to Cerveny et al. (2007) and Mazzoglio et al. (2019), extreme rainfall events are becoming more frequent. It is thus an increasing concern for emergency agencies, researchers and publics especially residents in coastal cities (Gao et al., 2009). Tropical cyclones carry a considerable amount of water vapor from oceans inland and are the predominant cause of deaths in the United States and disruption to transportation, utilities, communications, and agriculture (Elsberry 2002; Knight and Davis 2007; Gao et al., 2009; Dare et al., 2012). Hurricane Harvey, starting as tropical storm, became a category four hurricane while making landfall in Texas causing devastating urban flooding and fatalities in August 2017. Based on the survey by Emanuel (2017), Hurricane Harvey had reached the heaviest rainfall record in the history of the United States and caused at least 70 casualties, and economic loss beyond 150 billion US dollars (Emanuel, 2017; Omranian et al., 2018; Chen et al., 2019). In Fall 2019, another tropical cyclone Imelda made landfall in Texas, with recorded 1096 mm of rainfall that occurred in Jefferson County, which is ranked as the fourth-highest rainfall record in the history. Two additional tropical cyclones Bill and Cindy influenced southeast Texas and brought in copious amounts of rainfall in 2015 and 2017, respectively.

Traditionally, rainfall is estimated by gauges as a direct measurement but only at a point scale (Sarachi et al., 2015). Even though gauge data are often treated as the "ground truth" for rainfall measurement, they are still not impeccable due to splash-out during heavy rainfall, lack of sensitivity to light rain rates, under-catching by wind drift, and evaporation (Luyckx & Berlamont, 2001; Molini et al., 2005; Dai et al., 2018). Especially in heavy rain events, multiple studies have demonstrated that the error caused by these inherent factors is not trivial (Luyckx & Berlamont, 2001; Molini et al., 2005). Luyckx (2001) investigated the disadvantages of tipping bucket rain gauges during extreme weather conditions and found there was underestimation of rainfall volumes due to loss of water during the tipping action from the comparison of 24 well calibrated gauges. Molini et al. (2015) quantified the bias as an underestimation from 60% to 100% for the 1h design rainfall and return periods from 20 to 200 years. The wind from tropical cyclones would substantially affect the performance of rain gauges as well, with the relative bias ranging from 5% to 80% (Medlin et al.,2007; Pollock et al., 2010). Besides the systematic error, when interpolating

point samples, errors caused by spatial interpolation accounts for 50% to 80% of the total difference depending upon the gauge quality and density (Ciach and Krajewski 1999; Dai et al., 2018). Stampoulis and Anagnostou (2012) discovered that the convective nature of rainfall can also increase gauge-interpolation uncertainties. All this may result in problems when one applies gauge-based corrections for other products e.g., radar QPE, satellite precipitation estimation in extreme events.

In the past three decades, the emerging radar technology has been applied in meteorology to estimate rainfall by emitting and receiving electromagnetic signals. The most prominent advantage of radar over rain gauges is that radar provides a more refined spatiotemporal scale and a larger area coverage (He et al., 2013). However, since it is an indirect measurement of rain rate, radar technology itself produces and propagates errors in its end products. The uncertainties of radar-based precipitation estimation can be categorized as incorrect calibration, sampling representativeness, non-weather echoes, and errors in Z-R relations (Gourley et. al., 2007; Medlin et al., 2007; Zhang et. al., 2008; Ryzhkov et al., 2014; Kirstetter et al., 2015; Cao et al., 2018; Dai et al., 2018). These inherent systematic errors are challenging to mitigate by improving radar technology alone. Thus, many researchers chose to blend radar with rain gauges and satellite data to improve the performance e.g., Kriging with External Drift (KED; Jewell & Gaussiat, 2015; Cecinati et al., 2018), Mean Field Bias Correction (MFB; Yoo et al., 2014). A few studies investigated the bias of radar rainfall product in excessive rainfall events (Medlin et al., 2007; Kidd et al., 2011; Chen et al., 2013; Cao et al., 2018). Kidd et al. (2011) compared ground-radar Quantitative Precipitation Estimation (QPE) with gauge data in Germany, and the results indicated radar overestimation in convective rainfall regimes. Medlin et al. (2007) evaluated National Weather Service (NWS) Weather Surveillance Radar-1988 Doppler (WSR-88D) during Hurricane Danny. They concluded that both radar and rain gauge seriously underestimated rainfall. Cao et al. (2018) evaluated the performance of S-band dual-polarized radar in Hurricane Irma in which the radar showed nearly 50% of underestimation. Gao et al. (2018) evaluated the performance of Multi-Radar Multi-Sensor (MRMS) QPE during hurricane Harvey. MRMS QPE underestimated total accumulated rainfall by a small factor and overestimated very light precipitation.

Another commonly used data source is remote sensing satellite, attributing to the advantages of broad spatial coverage and an ability to scan downward over complex terrain (Mei et al., 2014; Sarachi et al., 2015). Satellite-based precipitation products utilize the information provided by

visible-infrared (VIS-IR) channels from geostationary (GEO) satellites, passive microwave (PMW) sensors, and spaceborne radars from low orbiting (LEO) satellites. Just like weather radar QPE, satellite rainfall retrieval is also an indirect measurement of rainfall. Many previous works (Mei et al., 2014; Omranian et al., 2018; Huang et al., 2019) stated that satellite data would underestimate the magnitude of high rainfall rates. Hong et al. (2006) stated that satellite performance degrades with the increase of rain rate but has higher relative bias in the light rain. Chen et al. (2013) compared four satellite precipitation products (SPPs) and ground-based radar with gauge references for Typhoon Morakot, and they found that SPPs underestimate extreme rainfall. Omranian et al. (2018) evaluated Integrated Multi-satellitE Retrievals for GPM (IMERG) Version 5 final gauge-adjusted products with the National Weather Service/National Centers for Environmental Prediction (NWS/NCEP) River Forecast Center (RFC) Stage-IV Quantitative Precipitation Estimates (QPEs) as the reference in Hurricane Harvey. They revealed that IMERG is able to detect the spatial variability of the rainfall field but suffered in heavy rainrate regions.

Given the context that these independent data sources come with its inherent deficiencies, researchers performed stochastic approaches to analyze these uncertainties (Tian & Peters-Lidard, 2010; Sarachi et al., 2015). Triple Collocation (TC) has been proven to be a powerful statistical tool to estimate uncertainties within each of three independent products (Stoffelen, 1998; Caires, 2003; Zwieback et al., 2012; Roebeling et al., 2012; Ratheesh et al., 2013; Gentemann, 2014; McColl et al., 2014; Alemohammad et al., 2015; Massari et al., 2017; Li et al., 2018). TC was firstly applied to evaluate ocean surface wind variability by inputting different wind products (Stoffelen, 1998). After that, it has been extended to determine errors of sea surface temperature (Gentemann, 2014), sea surface salinity (Ratheesh et al., 2013), and wave height (Caires, 2003). Roebeling et al. (2012) were the first to apply TC in hydrometeorology to accommodate remote sensing, weather radar and rain gauges in Europe. Massari et al. (2017) compared the performance of five satellite precipitation products over the U.S. and applied the TC method to evaluate correlation coefficients worldwide. Alemohammad et al. (2015) introduced the Multiplicative Triple Collocation method (MTC), suggesting its appropriateness in rainfall error estimation and proposed a way to decompose the error term in order to investigate the violation of assumptions. Li et al. (2018) used TC method to perform uncertainty analysis for five SPPs, reanalysis data, and gridded gauge data over ungauged regions in Tibetan Plateau in China after validating TC with

traditional statistics. Tang et al. (2020) first evaluated the accuracy of snow observations for satellite precipitation products and reanalysis products.

To our knowledge, the TC and particular MTC methods have not been applied for evaluating gauge-radar-satellite precipitation products, exclusively for extreme events. Thus, the overarching goals of this study are to: (a) evaluate the applicability of MTC method in uncertainty assessment under extreme events; and (b) investigate the similarity, difference and deficiencies of each product along with its inherent deficiency in extreme events. More specific objectives of this study include:

1) Evaluate the applicability of MTC in extreme events with the cross-examination of three products using traditional metrics.

2) Further examine the stability of MTC method's performance in multiple extreme events.

3) Understand gauge rainfall product uncertainties in extreme events, which are often associated with splash-out, wind undercatch as well as interpolation uncertainties.

4) Understand satellite QPE uncertainties in extreme events, which are associated with signal that are indirectly tied to surface precipitation and poor spatiotemporal resolutions.

5) Understand radar QPE uncertainties in extreme events, which are associated with incorrect Z-R formulations, non-weather signals, inadequate sampling.

This article is organized into four sections: Section 2 introduces the study domain and briefly reviews the three state-of-the-art datasets being used in this study; Section 3 describes in detail the formulas to derive Root Mean Squared Error (RMSE) and Correlation Coefficient (CC) from TC method; Section 4 analyzes the results from a broad overview and further examines the specific event (Hurricane Harvey); and Section 5 provides the conclusions and future works.

Traditionally, rainfall is estimated by gauges as a direct measurement but only at a point scale. Even though gauge data are often treated as the "ground truth" for rainfall measurement, they are not impeccable due to splash-out during heavy rainfall, lack of sensitivity to light rain rates, under-catching by wind drift, and evaporation (Dai et al., 2018; Molini et al., 2005). Especially in heavy rainfall events, multiple studies have demonstrated that errors caused by these inherent factors are not trivial. In addition to gauge rainfall products, radar Quantitative Precipitation Estimation (QPE) has been advantageous to provide a more refined spatiotemporal resolution and a large area coverage. However, radar-based products bear uncertainties that can be categorized as incorrect calibration, sampling representativeness, non-weather echoes, and errors

in Z-R relations (Gourley et al., 2007; Kirstetter et al., 2015; Zhang et al., 2010). Another commonly used data source is remote sensing satellite, attributing to the advantages of broad spatial coverage and an ability to scan downward over complex terrain (Kidd et al., 2017; Skofronick-Kackson, et al., 2017). Just like weather radar QPE, satellite rainfall retrieval is also an indirect measurement of rainfall, which suffers from ambiguous surface emissivity, infrequent sampling, and coarse spatial resolution.

In this study, we quantify the inherent uncertainties of three mainstreaming precipitation products via a Multiplicative Triple Collocation (MTC) method for four extreme precipitation events. The overarching goal is to (a) evaluate the applicability of MTC in uncertainty assessment under extreme events, and (b) investigate the similarities, differences, and deficiencies of each product along with its inherent deficiency in extreme events. More specific objectives of this study include: (1) evaluate the applicability of MTC in extreme events with cross-examination of three products using traditional metrics, (2) furfure examine the stability of MTC method's performance in multiple extreme events, (3) understand gauge rainfall product uncertainties in extreme events, which are often associated with splash-out, wind undercatch, as well as interpolation uncertainties, (4) understand satellite precipitation products uncertainties in extreme events, which are associated with signals that are indirectly tied to surface precipitation and poor spatiotemporal resolutions, and (5) understand radar precipitation product uncertainties in extreme events, which are associated with incorrect Z-R formulations, non-weather signals, and inadequate samplings.

2.2 Study area and datasets

2.2.1 Study domain

The area of interest is in the Gulf Coast of the U.S., where it endures several events recently and historically as shown in Figure 2.1. It is one of the most frequently impacted areas by hurricanes, such as Hurricane Dennis (2005), Isaac (2012), Harvey (2017) and tropical storms such as Bill (2015), Cindy (2017), Imelda (2019) also affected this area in the past five years. Given that this study focuses on GPM era, Figure 2.1 illustrates the storm tracks and accumulative precipitation amount of Bill, Cindy and Imelda as well as Harvey, all after year of 2015. Due to the unprecedented nature of Hurricane Harvey, the three other tropical storms were selected for analysis to improve the generality of the study findings. All three tropical storms were accumulated in total to match the Hurricane Harvey, which gave the analogous data sample sizes and

accumulative rainfall amounts for two comparable study groups: Harvey and Other. The impacted area of the aforementioned events contains the states of Texas, Oklahoma, Louisiana, Arkansas, Tennessee, Mississippi, and Alabama, which almost accounts for 10 percent of the conterminous United States (CONUS). These events had similar moving patterns in that they approached inland from the Gulf of Mexico and then bending towards the northwest after making landfall, except for Imelda which dissipated shortly after making the landfall. Table 2.1 lists the details of four individual events, including their durations and amount of rainfall. Harvey was observed to contiguously produce rainfall across seven days. The duration of other three tropical cyclones (Bill, Cindy, Imelda) were short as 3 days, 2 days, and 4 days, respectively.

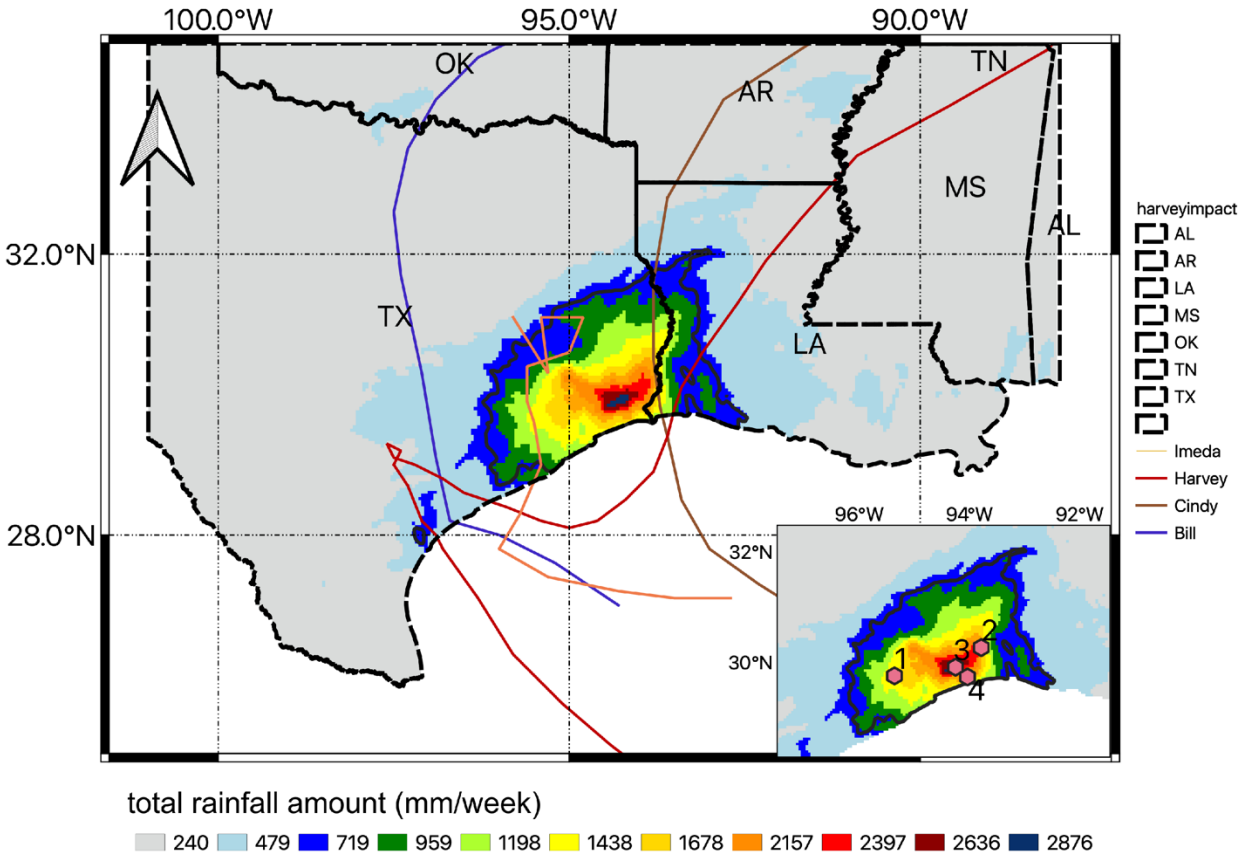


Figure 2.1 Accumulative rainfall for four events combined is measured by Multi-Radar Multi-Sensor (MRMS) radar data. Four cyclone tracks are illustrated with lines. Red dots (numbers from 1 to 4) in the bottom right panel are selected representative pixels in the Harvey core region.

Table 2.1 Events overview.

Hurricane/Storm	Start Date	End Date	Duration	Maximum Rainfall Amount
Harvey	25 August 2017	31 August 2017	7 days	1625 mm
Bill	16 June 2015	18 June 2015	3 days	496 mm

Cindy	22 June 2017	23 June 2017	2 days	233 mm
Imelda	18 September 2019	21 September 2019	4 days	1126 mm

2.2.2 Dataset description

The NCEP gridded gauge-only hourly precipitation product (hereafter denoted as NCEP) is an operational product (Lin, 2011) covering the CONUS and parts of Puerto Rico. It is automatically derived from approximately 3000 operational hourly rain gauge observations across 48 states to produce a 4km/hour rainfall field, using the interpolation method from (Seo 1998), which introduced Double Optimal estimation (DO) and Single Optimal estimation (SO) to gain a conditional expectation of the rainfall estimation. This technique accounts for fractional coverage of rainfall due to sparse gauge networks. Gourley et al. (2009) performed inter-comparisons of NCEP, NCEP Stage IV radar QPE and PERSIANN-CSS SPP, and demonstrated that NCEP delivers the best performance at longer time scale e.g., seasonal, daily. However, it encounters underperformance at shorter time scale i.e., 1 hour, especially for the convective scale storms. NCEP data were obtained from National Center for Atmospheric Research/Earth Observing Laboratory (NCAR/EOL): https://data.eol.ucar.edu/dataset/21.088.

MRMS radar-only product (hereafter MRMS) has around 180 integrated operational radars, including 146 S-band and 30 C-band radars, creating seamless 3D radar mosaic across the CONUS and Southern Canada at 1km/2min resolution (Zhang et al., 2016). It is selected because of the strict quality control, involving filtering out non-hydrometeor signals, corrections for anomalous propagation, beam blockage, vertical profile reflectivity (VPR) correction, and adaptive Z-R relations (Zhang et al., 2016). Despite the adoption of these rigorous quality control steps, it still suffers from uncertainties of common issues mentioned before. For our study, 1km/2min radar-based QPE were retrieved and aggregated by averaging to 4km/hourly, in order to be compatible with NCEP. Historical MRMS radar-only QPE was downloaded at http://mtarchive.geol.iastate.edu/.

IMERG satellite precipitation final product V06 (hereafter IMERG; Huffman, 2019) is an integrated SPP from its core satellite GPM Core Observatory (GPM CO), microwave constellations, Infrared, and additional constellations, aiming at providing global coverage of rainfall field (90N-S from V06 onward) beyond its predecessor Tropical Rainfall Measurement

Mission (TRMM). GPM CO has additional channels of dual-frequency precipitation radar (DPR) and GPM Microwave Imager (GMI), which can detect very light precipitation and falling snow (Skofronick-Jackson et al., 2017). It produces three stages: early run, late run, and final run with 4 hours latency, 12 hours latency, and 3.5 months latency respectively to accommodate different purposes. The early run provides near real-time brief observations with inter-calibrated satellite products primarily for operational forecasting, and the late run adds up the late coming high-quality PMW data and climatological calibration to serve for agricultural purposes. The final run compares the late run product with the Global Precipitation Climatology Project (GPCP) and adjust the factor to compensate for under/over-estimation (Huffman et al., 2019). To account for the independence, we selected the final run without calibration in this study. In order to perform pixel-wise analysis, we further downscaled IMERG data using a nearest neighbor approach and then accumulated it up to hourly. Current IMERG final run product V06 can be accessed at https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGHH_06/summary?keywords=IMERG.

2.3 Methodology

2.3.1 TC evaluation

2.3.1.1 Assumptions

A few assumptions are needed to comply for the TC method: (1) the three datasets should be independent, for instance, they are derived from different instruments; (2) The errors of the datasets should be independent or unrelated which is referred to as zero cross-correlation. (3) The expectation of error is treated as zero known as the unbiasedness, which is often described in geostatistical analysis e.g., kriging. Yilmaz and Crow (2014) conducted experiments on the TC errors due to the relevance of three products, and the results revealed that the more independent they are, the less TC induced error. It is essential to consider the relevance of the inputs in order to make the TC method more reliable (Li et al., 2018). The three precipitation products selected, i.e. gauge-only, radar-only, and satellites-only meet the above criteria well because they are derived from different and independent instruments.

2.3.1.2 Expressions

The basics of TC method is to treat three independent datasets as equally important, and thus no bias is produced in between them. Since no ground truth values are assumed, TC method then uses a linear combination of three products and affine transformed error model to derive RMSE and CC (Zwieback et al., 2012).

$$R_i = a_i + b_i \mathrm{G} + \varepsilon_i, \tag{1}$$

where R_i indicates each of the independent source data, G is the "relative truth", a_i, b_i correspond to the weights and biases to adjust, and ε_i represents error for each product. Tian et al. (2013) then proposed a way to transform the additive error model to multiplicative by logarithmic transformation, and it is proved to be more appropriate in rainfall error estimation (Tian et al., 2013; Alemohammad et al., 2015; Li et al., 2018). Hence, the error model can be reformed as:

$$\mathrm{R_i} = \alpha_\mathrm{i} \mathrm{G}^{\beta_\mathrm{i}} \epsilon_\mathrm{i} \tag{2}$$

From that, we can derive rain rate and error model by transforming back into linear combination so that it fits into TC method.

$$\mathrm{r_i} = \mathrm{a_i} + \mathrm{b_i g} + \varepsilon_\mathrm{i}, \tag{3}$$

where r_i is the logarithmic form of rain rate R_i, $a_i = \ln \alpha_i$ demonstrates the multiplicative error, $\epsilon_i = \ln \varepsilon_i$ indicates the residual error, and $b_i = \beta_i$ as the deformation error.

From linear equation (3), we are able to derive RMSE in the following set of equations based on the covariance of triples (McColl et al., 2014):

$$\begin{cases} {\sigma_{r1}}^2 = Cov(r_1, r_1) - \dfrac{Cov(r_1, r_2)Cov(r_1, r_3)}{Cov(r_2, r_3)} \\ {\sigma_{r2}}^2 = Cov(r_2, r_2) - \dfrac{Cov(r_1, r_2)Cov(r_2, r_3)}{Cov(r_1, r_3)} \\ {\sigma_{r3}}^2 = Cov(r_3, r_3) - \dfrac{Cov(r_1, r_3)Cov(r_2, r_3)}{Cov(r_1, r_2)} \end{cases} \tag{4}$$

Because the model is transformed to be in additive form, these parameters along with error should also be in logarithmic form. In the analysis of Alemohammad et al. (2015), they transformed error to linear scale by Taylor series expansion as first order approximation.

$$\sigma_{r_i}{}^2 = \left(\frac{\sigma_{R_i}}{\mu_{R_i}}\right)^2 \tag{5}$$

$$\sigma_{R_i} = \mu_{R_i}\sigma_{r_i} \tag{6}$$

σ_{R_i}, σ_{r_i}, μ_{R_i} here represent the RMSE in linear form, logarithmic form, and the mean of field. In doing so, the error field could be identified in linear scale, meaning the same unit of mm/hour as rain rate.

(McColl et al., 2014) introduced a way to evaluate correlation coefficient (CC) from manipulating covariance matrices in which CC is expressed below as a set of equations.

$$\begin{cases} CC_1{}^2 = \dfrac{Cov(r_1, r_2)Cov(r_1, r_3)}{Cov(r_1, r_1)Cov(r_2, r_3)} \\ CC_2{}^2 = \dfrac{Cov(r_1, r_2)Cov(r_2, r_3)}{Cov(r_2, r_2)Cov(r_1, r_3)} \\ CC_3{}^2 = \dfrac{Cov(r_1, r_3)Cov(r_2, r_3)}{Cov(r_3, r_3)Cov(r_1, r_2)} \end{cases} \tag{7}$$

To clarify the TC calculated RMSE with traditional evaluation, thereafter we only use RMSE that refers to TC results and RMSD for traditional evaluation. Since both RMSE and CC are derived from covariance between the three products, the TC statistics reveal the relative error, which is treated as uncertainties. Therefore, the less uncertain product or the best performance is the one that has the lowest RMSE and highest CC. Likewise, the most uncertain is associated with the highest RMSE and lowest CC. In the following content, the unit of RMSE is in logarithmic mm/hour and CC is unitless.

2.3.1.3 Data preparation

The rainfall data retrieved from each source have different spatial extends and resolutions. To make them comparable, IMERG and MRMS are downscaled with nearest neighbors and aggregated to the same resolution as NCEP, with 4km in space and 1-hour in time. To fit in MTC,

the rainfall data should be logarithmic transformed. Previous studies demonstrated that the treatment of zero values can be done either by simply removing zero values (Alemohammad et al., 2015; Massari et al., 2017) or replacing with near-zero values (Roebeling et al., 2012; Li et al., 2018). However, the displace with near-zero values can retain more sample sizes, especially in event analysis (Roebelling et al., 2012). Based upon their experiences, the NAN values are then removed and zeros are replaced with 10^{-3} in this study. To obtain more robust results and filter out noise, the bootstrapping strategy is utilized with 500 trials for evaluation at each pixel and stored the mean values of RMSE and CC.

2.3.2 Conventional statistical metrics

A list of evaluation metrics is summarized in Table 2.2. As there is assumed no absolute ground truth in the extreme events, the "error" is termed as "difference" between two products e.g., Root Mean Squared Difference (RMSD). They are all computed in a domain at each pixel (4km by 4km) for each pair. For categorical indices between two products, the Probability of Detection (POD), False Alarm Rate (FAR), Critical Success Index (CSI) are chosen as indicators to investigate the sensitivity of rainfall detection. The reasons that NCEP is not regarded as reference are: (1) NCEP is not well quality controlled; (2) It involves uncertainties from interpolation in space; (3) NCEP likely encounters device problems in extreme rainfall events. It thus exposes more uncertainties, and is justified in our study analysis. The "reference" data in the denominator is chosen by the less uncertain product calculated with regard to TC method. The threshold for computing these categorical indices is set to 0.1 mm/hour (Tang et al., 2020).

Table 2.2 A list of computational metrics used in this study.

Metrics		**Equation**	**Best Value**	**Conditional Values**
Continuous Indices	Correlation coefficient	$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}}$	1	$CC_{cond,p}$
	RMS difference (RMSD)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-R_i)}$	0	$RMSD_{cond,p}$
Categorical Indices	POD	$\frac{\text{Hits}}{Hits+\text{Misses}}$	1	$POD_{cond,p}$
	FAR	$\frac{\text{False Alarms}}{Hits+\text{False alarms}}$	0	$FAR_{cond,p}$

CSI	$\frac{\text{Hits}}{Hits\ +\ \text{Misses} + \text{False alarms}}$	1	$CSI_{cond,p}$

Note: For categorical indices, hits is defined by both evaluated value and reference value larger than the threshold. The false alarm is an evaluated value larger than the threshold but a reference less than it. Misses are evaluated values less than the threshold while the reference is larger than it. Conditional metrics are defined by given source data larger than a threshold, i.e., p percentile of the daily rainfall for extreme events.

Since it is also insightful to compute these metrics in a conditional way during extreme events (Sukovich et al., 2014), hence they are conditioned to regions within the core of Harvey to validate the consistency between unconditioned results and conditioned (i.e. RMSD, $RMSD_{cond,p}$). This condition is based upon what the percentile of rainfall rate the product exceeds with similar formula in Table 2. To take the extreme values into account, three percentiles are selected (i.e., 50 percentiles, 75 percentiles, and 95 percentiles) with an analogy to previous work (Sukovich et al., 2014). Therefore, at point-level, only rain rates beyond those conditions are used to compute the differences in Harvey core.

2.3.3 Hierarchical evaluation

Three hierarchical levels of evaluation are performed: In level one, all events are combined (denoted as **All**) and further events breakdown as **Harvey** event only and **Other** non-Harvey events for cross-examination; In level two, extreme Harvey event is exclusively taken into account to test the consistency of TC results with conventional metrics; And last, analysis is conducted specifically within the storm core of Hurricane Harvey to examine uncertainty in such ultra-extreme region, with the core delineated with 400mm rainfall isoline. In core region, the conditional metrics described above are also calculated and several representative pixels are selected to investigate their temporal variations. These pixels from 1 to 4 as shown in Figure 1 are: (1) Houston metropolitan area where NCEP is highly uncertain; (2) the maximum amount of rainfall observed by IMERG; (3) the maximum amount of rainfall observed by MRMS; and (4) the highly uncertain place for MRMS.

2.4 Results

2.4.1 Cross-events comparison

2.4.1.1 Conventional inter-comparison

Figure 2.2 depicts the accumulative rainfall maps, and Table 2.3 lists the inter-comparison results with conventional metrics for three cases (i.e., All, Harvey, and Other) from the three datasets (i.e., NCEP, MRMS, and IMERG). The rainfall spatial patterns observed by each product are similar, as their spatial correlations are all above 0.8 cross cases in Table 2.3. MRMS/IMERG has the highest spatial correlation for All (0.91) and Harvey event (0.94) while similar to NCEP/MRMS in Other non-Harvey events (both 0.87). Despite these similarities, neither NCEP nor IMERG adequately capture the high spatial variability inside the core regions, due to their relatively coarse spatial resolutions. Especially in the Harvey case, one can observe the "patchiness" pattern for NCEP, which is further investigated in Section 3.3. From the rainfall distribution histogram inserted in Figure 2.2, the non-Harvey events have higher densities of rainfall amount in the light rain range (<200 mm), but Harvey produces more in the heavy rain range (>800 mm).

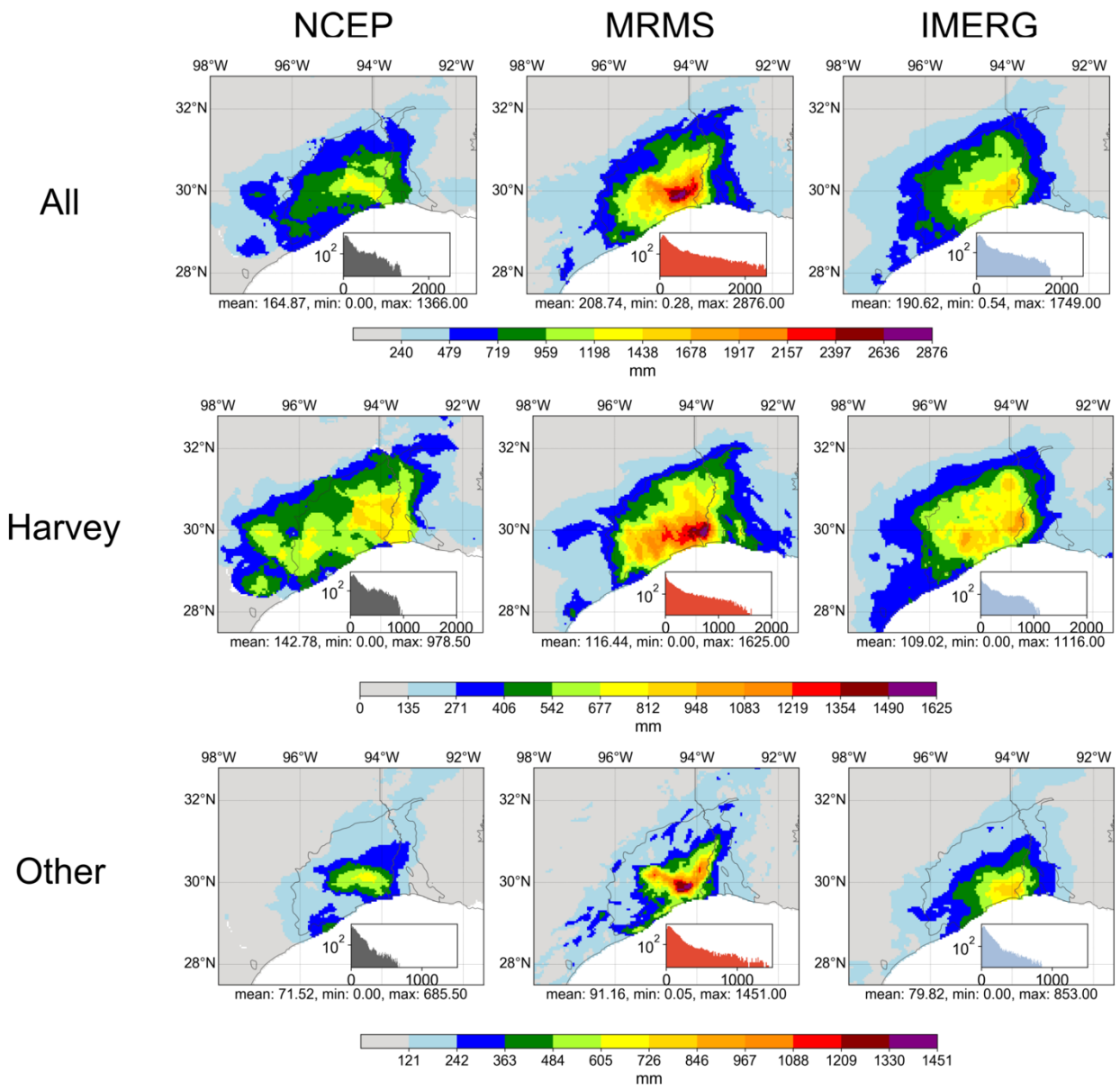


Figure 2.2 Accumulative rainfall observed in three cases. Inside each axes, the inset corresponds to the histogram of accumulative rainfall.

Table 2.3 lists the inter-comparison results with conventional metrics. In terms of total accumulation, MRMS always observes the highest accumulative rainfall amount, followed by IMERG and NCEP. Both spatial and temporal correlations demonstrate the highest similarity between the two remote sensing technology-derived IMERG and MRMS in all cases. For categorical metrics (i.e., POD, FAR, CSI), IMERG and MRMS also show the best agreement in general, suggesting NCEP possibly suffers from gauge device malfunction in these extreme events.

Table 2.3 Overall conventional statistics for three cases (i.e., All, Harvey, and Other) and pairs of comparison.

Metrics		All	Harvey	Other
Max. total rain (mm)	NCEP	1366	979	686
	MRMS	2876	1625	1451
	IMERG	1749	1116	853
RMSD (mm/h)	NCEP/IMERG	**2.51**	1.44	**1.87**
	NCEP/MRMS	3.08	1.55	2.54
	IMERG/MRMS	2.71	**1.42**	2.20
Temporal CC	NCEP/IMERG	0.52	0.48	0.48
	NCEP/MRMS	0.49	0.54	0.40
	IMERG/MRMS	**0.62**	**0.56**	**0.60**
Spatial CC	NCEP/IMERG	0.85	0.90	0.85
	NCEP/MRMS	0.80	0.87	**0.87**
	IMERG/MRMS	**0.91**	**0.94**	**0.87**
POD	NCEP/IMERG	0.64	0.52	0.56
	NCEP/MRMS	**0.72**	0.58	0.62
	IMERG/MRMS	0.70	**0.69**	**0.62**
FAR	NCEP/IMERG	0.28	0.47	0.39
	NCEP/MRMS	0.29	0.39	0.42
	IMERG/MRMS	**0.19**	**0.41**	**0.27**
CSI	NCEP/IMERG	0.52	0.44	0.41
	NCEP/MRMS	0.57	**0.52**	0.43
	IMERG/MRMS	**0.60**	0.51	**0.51**

2.4.1.2 MTC comparison

Because no absolute ground truth is assumed, MTC instead evaluates the relative deviation from "truth" among three independent products. The first overview of TC analyzed results for All, Harvey, and Other are illustrated in Figure 2.3. Both CC (relative agreement) and RMSE (relative deviation) indicate MRMS (CC = 0.80; RMSE = 2.77 mm/h) is the overall best product (closest to truth) across all events, followed by IMERG (CC = 0.72; RMSE = 4.35 mm/h) and then NCEP (CC = 0.66; RMSE = 5.16 mm/h). Therefore, the similarity and difference of NCEP and IMERG relative to the MRMS are mostly notable in the storm core Houston area, especially during the Harvey extreme event. For example, NCEP differs from MRMS more in the Houston metropolitan areas, as marked in black circles in Figure 3b. This result is on a par with the investigation of Chen et al. (2020) in the Harvey event that MRMS has the highest CC (0.91) value and correspondingly lowest RMSE (5.75 mm/h) among NCEP gauge-only products and IMERG V06A in Hurricane Harvey. IMERG, on the other hand, exhibits better agreements with MRMS in the rainfall core regions than NCEP. Notably, the deviation from the most uncertain NCEP to the best overall MRMS is much more magnificent in the Harvey case, which can also be quantified as the range

of the median value of CC (0.14, **0.27**, and 0.09) and RMSE (1.79, **2.01**, 1.52 mm/h) for All, **Harvey**, and Other, respectively. This reveals that MTC can capture the high variability in this 500-year return Harvey extreme event. In summary, the above analysis results from both the MTC method and the conventional metrics suggest the consistence and robustness of TC in extreme events.

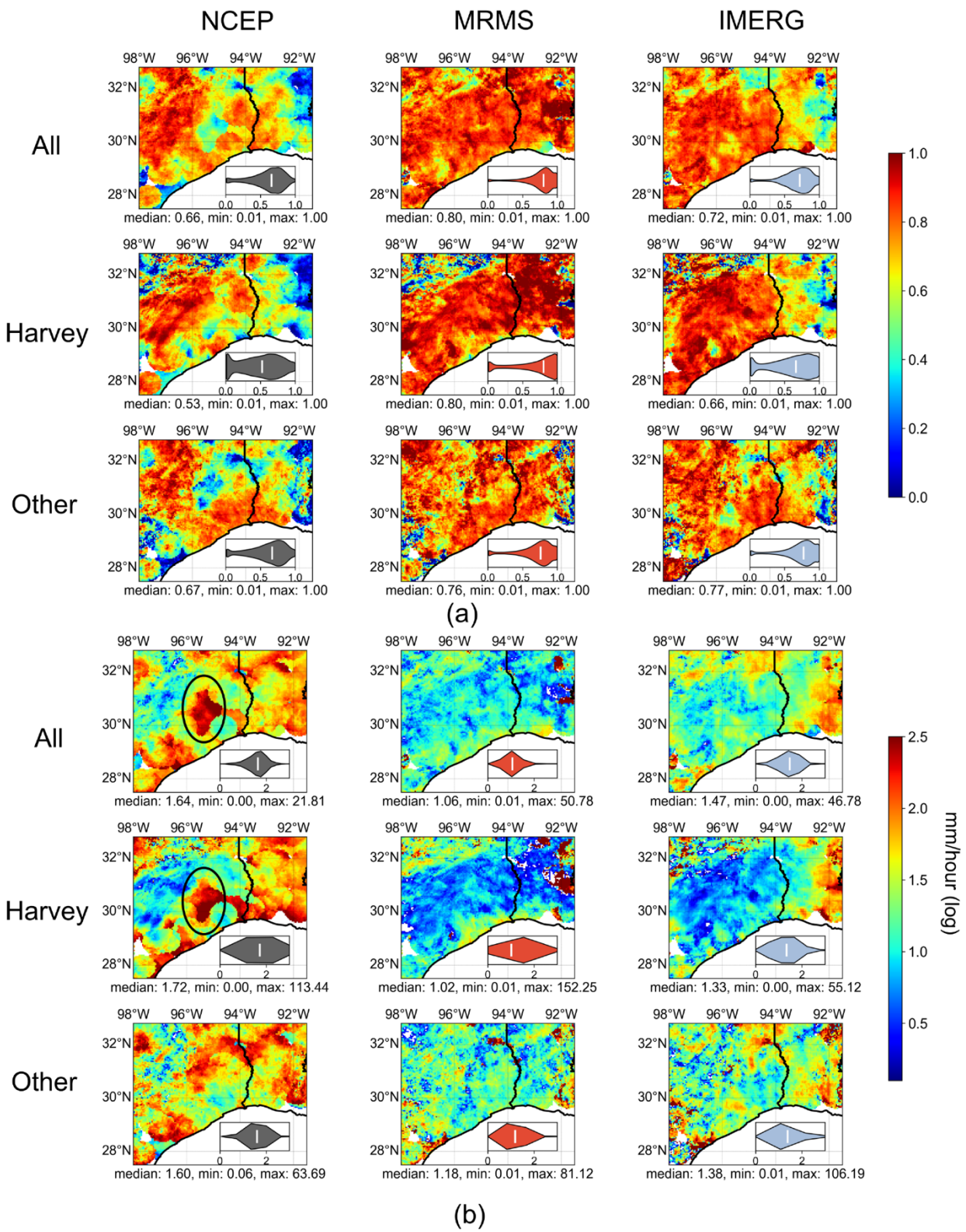


Figure 2.3 Spatial plot of Multiplicative Triple Collocation (MTC) results (a) Correlation Coefficient (CC); (b) Root Mean Squared Error (RMSE) for three cases (concatenated all events, Harvey-only, and non-Harvey) in each row, and each column represents each product. The small

panel inside each axis is the violin plot for the metric, and the white bar is where the median value is located. Two marked circles emphasize where the National Centers for Environmental Prediction product is highly uncertain.

2.4.2 Hurricane Harvey Analysis

2.4.2.1 Conventional inter-comparison and MTC results

The excessive rainfall extreme and high diverging behaviors of the TC results of Hurricane Harvey in the previous results motivate us to further investigate this event at the pixel level, as shown in Figure 2.4. First, the conventional CC indicates the MRMS/IMERG pair has the highest correlation values (0.69) with the least range than other pairs; and MTC calculated CC is also ranked MRMS (0.86), IMERG (0.79), and NCEP (0.65) from high to low. Second, even though conventional RMSD for each pair is similar (2.23 mm/h for NCEP/MRMS and NCEP/IMERG, 2.16 mm/h for MRMS/IMERG), MRMS/IMRG (NECP/MRMS) has the smallest (largest) range; similarly, MTC ranks MRMS to have the least (2.25 mm/h), followed by IMERG (2.60 mm/h) and NCEP (4.80 mm/h) based on RMSE. A previous study by Omranian et al. (2018) revealed the average POD, FAR, and CSI score to be 0.9, 0.3, and 0.7, respectively, for the IMERG Final product against radar data in the Harvey case. This result is consistent with our study (POD = 0.92; FAR = 0.25; CSI = 0.73). Overall, IMERG/MRMS pair has the highest agreement among categorical indices. Again, the results at point-scale corroborate the consistence and robustness of MTC in this extreme Harvey event.

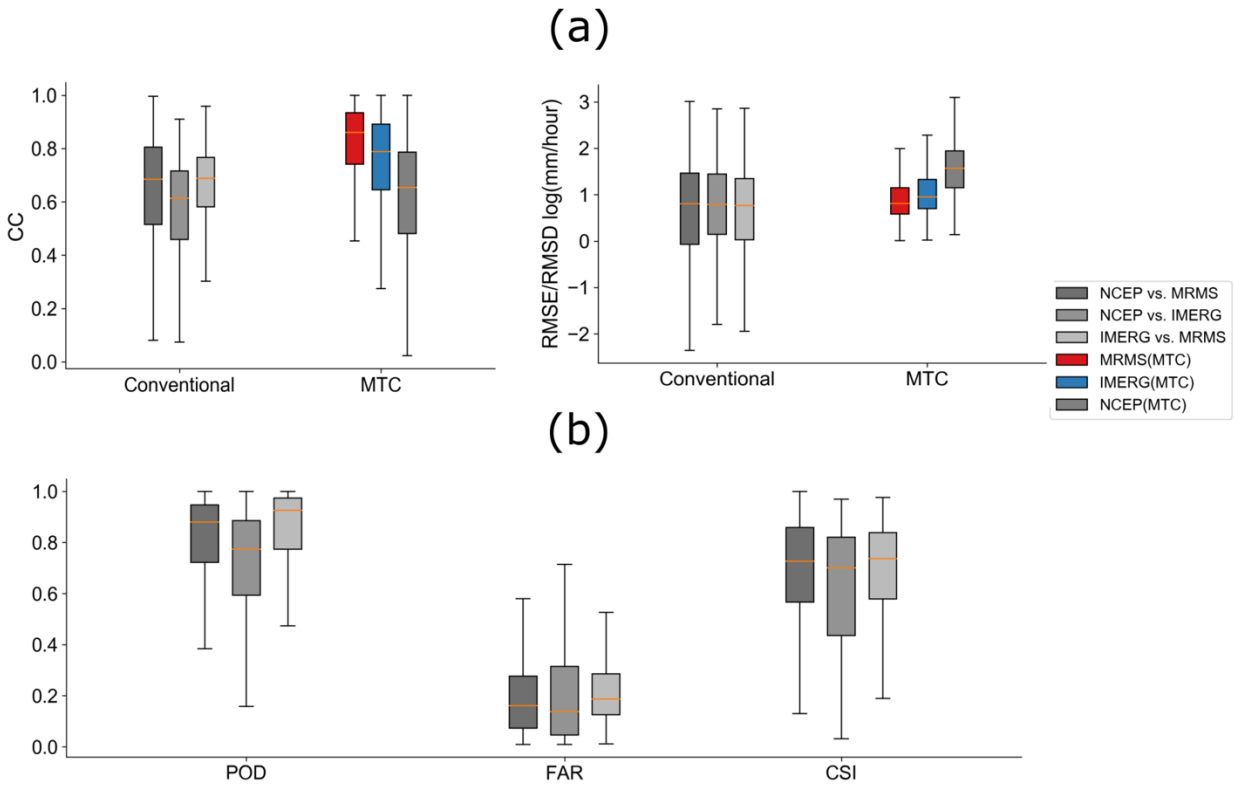


Figure 2.4 Boxplot for conventional metrics and MTC results in Hurricane Harvey. (a) Continuous indices and MTC results; (b) categorical indices.

2.4.2.2 Further exploration of MTC results

Since both cross-events results and exclusive Harvey case show the applicability of MTC in extreme events in previous sections, MTC is further applied to investigate the performance of each product with respect to accumulative rainfall ranges at 50 mm intervals, as shown in Figure 2.5. Overall, MRMS performs consistently well, with higher CC cross the range and lower RMSEs (only slightly higher than IMERG at a higher range 1100 mm). Interestingly, IMERG shows higher uncertainties in the light rain (i.e., below 150 mm), but with improved performance gradually toward a higher range. This is similar to reported literature (Guo et al., 2016; Sharifi et al., 2016; Sungmin et al., 2017). Specifically, Omranian et al. [12] also concluded that the IMERG final product generally has better performance with higher precipitation rates compared to lower rates in the case of hurricane Harvey.

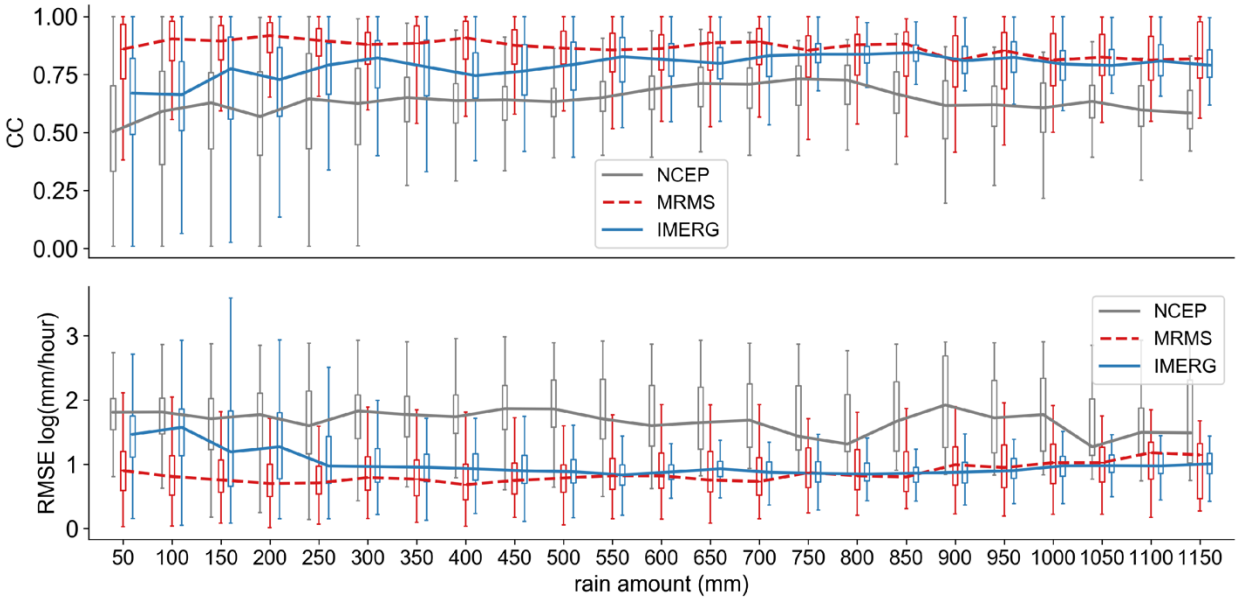


Figure 2.5 Boxplot of MTC measured CC, RMSE (from top down) in Hurricane Harvey at each accumulative rain bins for a 50 mm interval based on the MRMS data. The lines connected the median valus of each box for the corresponding product.

2.4.3 Storm core of Harvey event

Figure 2.6 depicts the MTC results of CCs and RMSEs for whole area, storm core, and non-core of Harvey with the core delineation of 400 mm rain isoline. First, the results (i.e., RMSEs and CCs) are generally better inside the core region because the rainy samples inside the core are more than non-cores and thus strengthen the robustness of MTC. Second, the performance for each product remains consistent: MRMS > IMERG > NCEP. Therefore, recall previous results, MTC

gives consistent rankings of these products in three different levels: Cross-events, Harvey event, and Harvey storm core.

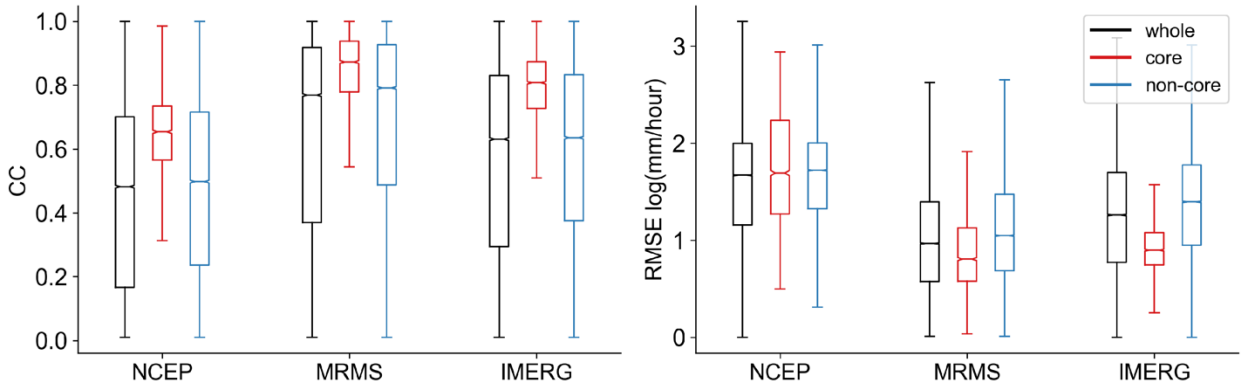


Figure 2.6 Boxplot of MTC measured CCs and RMSEs for three products grouped by whole, core, and non-core regions in Hurricane Harvey. Extreme values, i.e., outliers (outside of 1st and 3rd quartile + interquartile), are ignored to visualize the difference. The notch represents the median value for the samples in the region.

By looking at the core of Harvey alone, the accumulative rainfall distribution in Figure 2.7 reveals the systematic difference in this 500-year return extreme event. NCEP has the highest density within the low range of 400 to 600 mm, contributing almost 60% of the occurrence. In this same low range, IMERG with MRMS illustrates more similarities. In the middle range (i.e., 650 to 900 mm), on the contrary, NCEP and MRMS achieve better agreements and IMERG seems to overdetect. Unlike NCEP and IMERG, which drop off after 1100 mm, MRMS continuously observes rainfall in the higher end. Hence, both NCEP and IMERG underestimate high-end rainfall and overestimate at the lower/middle range.

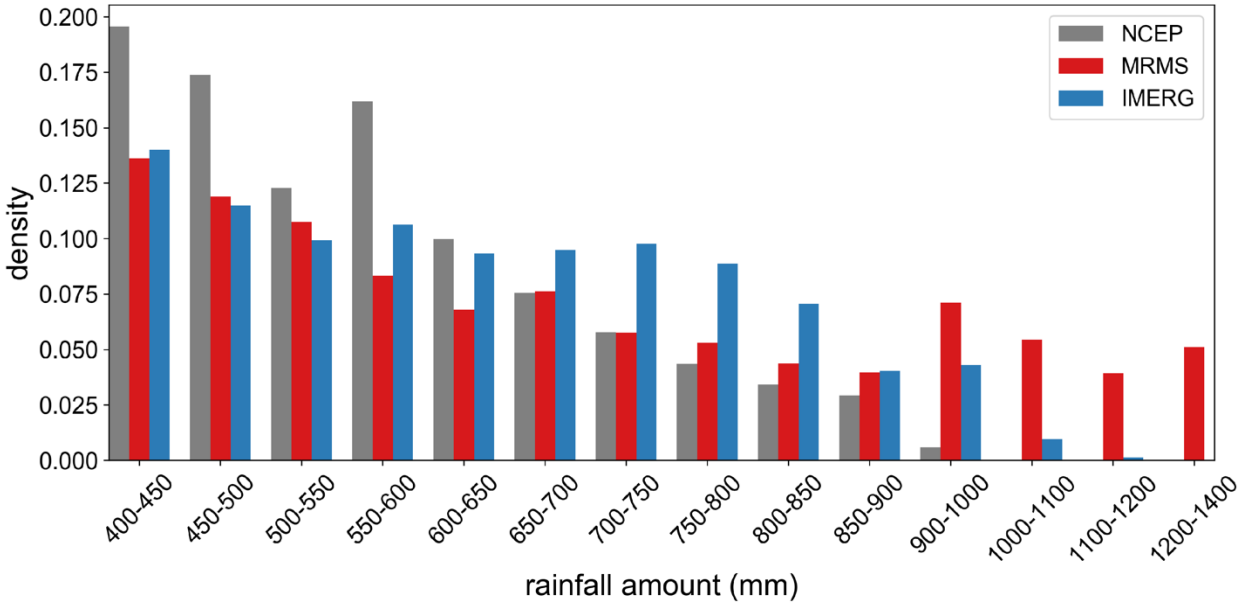


Figure 2.7 Histogram of accumulative rainfall binned at 50 mm interval from 400 (core) to 1400 mm in Hurricane Harvey. Vertical axis indicates the density.

Figure 2.8 further examines their performance conditioned at the medium to high end accumulative rainfall at three different percentiles (50th, 75th, and 95th). Figure 2.8a shows the conventional metrics with consistent results to MTC results across the three percentiles. Moreover, the differences of categorical indices in Figure 2.8b indicate IMERG/MRMS performs the best among the three pairs. In general, these results (conditional high rainfall) again approve the applicability of MTC in (ultra) extreme events.

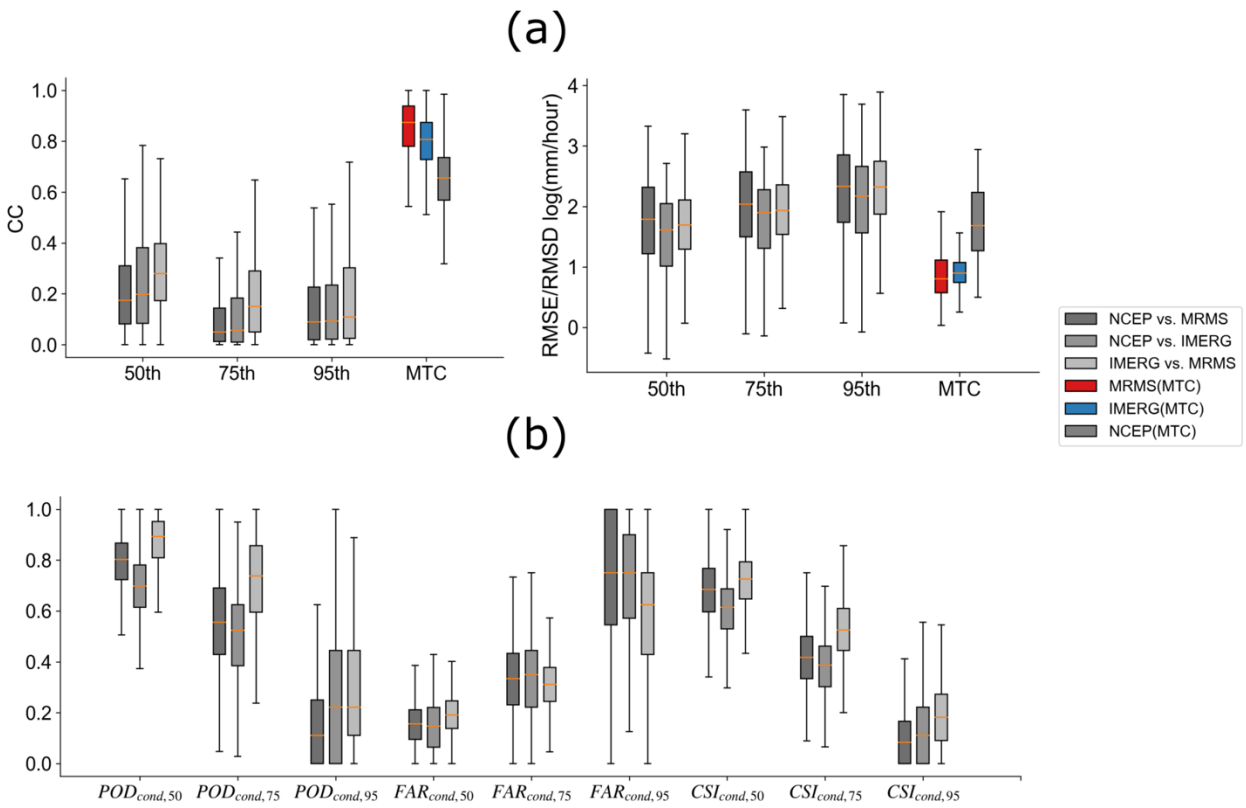


Figure 2.8 Boxplot of metrics conditioned at 50, 75, and 95th percentiles in the core region: (a) continuous indices; (b) categorical indices.

To look into event temporal variability, the time series of storm core average and some representative pixels of extreme are shown in Figure 2.9. Overall, the areal average time series in the uppermost panel shows they all indeed captured the temporal variability of the storm event. However, there are variations in pixel (local) details for each product. Sites 1, 3, and 4 refer to the place where NCEP have very high uncertainty (MTC calculated RMSE beyond 8 mm/h). The gray windows in the three corresponding time series plots indicate when NCEP data either showed zero value (site 1, 3) or stopped recording any data (site 4), while the other two remote sensing data sources captured intense rainfall. Site 2 is selected as IMERG measured the maximum amount of rainfall in this event. The horizontal blue line marks the upper maximum rain rate of 60 mm/h that GMI can observe due to its sensor sensitivity (Skofronick-Jackson et al., 2017). In other words, the IMERG product has an upper bound due to its sensor sensitivity. The red-shaded windows in sites 1, 3, and 4 indicate when MRMS rises up instantaneously while NCEP and IMERG show some degree of agreements at lower rain rates. Such spike anomalies could be caused either by

non-weather echoes, or more likely high-sensitivities in the Z-R relation to hails mixed with rainstorms (Gourley et al., 2007; Zhang et al., 2010).

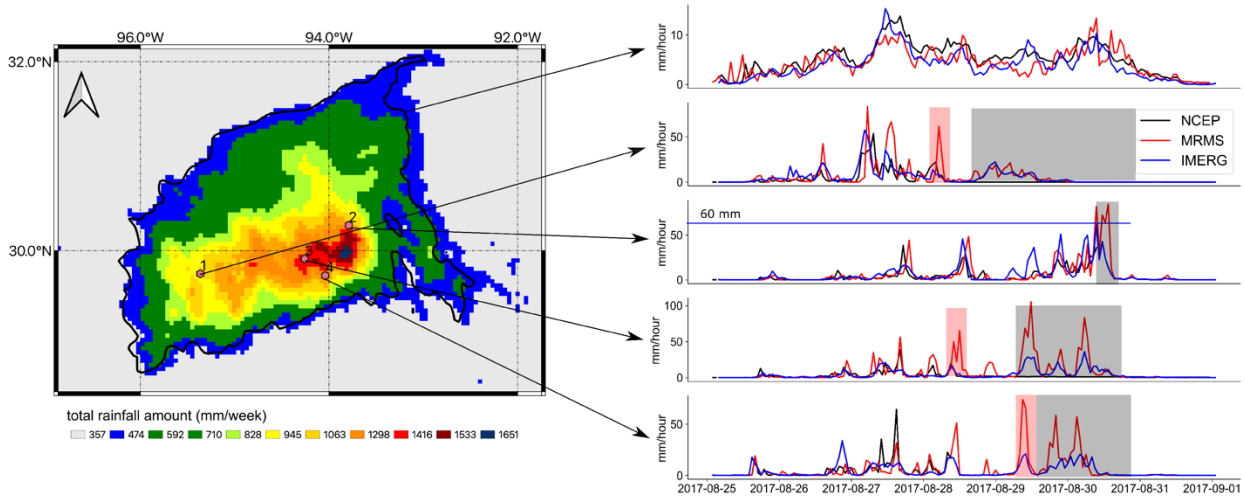


Figure 2.9 Core separated plot with selected pixel-wise time series. The black thick line in the left panel is the 400 mm accumulated rainfall contour line to separate the Harvey core regions. The windows in the right panel highlight special characteristics, with black corresponding to the NCEP observation, red for MRMS, and blue for IMERG.

2.5 Discussion

Throughout this study, three state-of-the-art precipitation products are intercompared with the conventional and MTC method. It is worth noting that during extreme events, particularly for 500-year return Harvey event, the ground-based observations (i.e., gauge and radar) are possibly contaminated due to saturation/malfunction and hot signals. Observations from space (i.e., satellite) suffer from relatively coarse spatiotemporal resolution and is thus unrepresentative of the storm core. Previous studies described in detail the impact of spatiotemporal resolution to the accuracy of rainfall products (Hong et al., 2006; Omranian et al., 2018). With the limitation on identifying "ground truth", this study adopts the MTC method and complements with the conventional method to provide the evaluations of the three products in extreme events. These agreements between MTC and conventional metrics corroborate these research findings and suggest the potentials of the MTC method in such circumstances. Beyond that, it is opportunistic to utilize the results of MTC and further develop innovative data merging/assimilation methods to integrate for one best possible combined product, especially suitable for extreme events.

2.6 Conclusions

In order to better understand uncertainties of precipitation estimation in extreme events, this study utilizes both traditional metrics and also the MTC method to cross-examine the similarities and differences of three independent rainfall datasets, i.e., MRMS, IMERG, and NCEP in Houston Storm/Hurricane events during the GPM era.

First, the applicability of the MTC method to extreme events is cross-evaluated with the conventional metrics and the results reveal that: (a) the consistency of cross-examination results against traditional metrics (both conditioned and non-conditioned continuous/categorical metrics) approves the applicability of MTC in multiple extreme events; (b) the consistency of cross-event evaluation results also indicates the robustness of the MTC method across such individual extreme events.

Second, both traditional metrics and especially MTC results conclude that: (c) all three independent products demonstrate their capacity of capturing the spatial and temporal variability of the storm structures while also magnify respective inherent deficiencies; (d) NCEP and IMERG likely underestimate, while MRMS overestimates the storm total accumulation, especially for the 500-year return Hurricane Harvey; (e) both NCEP and IMERG underestimate extreme rain rates (>= 90 mm/h), while MRMS maintains robust performance across the rain storm range; (g) all three products show respective inherent deficiencies in capturing the storm core of Harvey. For example, the NCEP discontinuity in time (zero or no records) is likely due to in-situ tipping buck gauge saturation or device malfunctions during the Harvey extreme; the relative coarse spatiotemporal resolution of IMERG under-capture the extremes; and while MRMS experiences "hot" radar signals possibly from non-weather echoes or hail contamination during such extremes.

Given the unknown ground reference assumption of MTC, this study suggests that MRMS has the best overall performance, followed by IMERG and NCEP. The similarities, differences, advantages, and deficiencies revealed in this study could guide the users, such as operational agencies, to carefully select products for emergency planning and response during extreme storms, and also hopefully motivate the research community to improve respective gauge network designs and sensor/algorithms. Furthermore, this study also calls for developing innovative multi-platform multi-sensor data merging/assimilation methods to integrate one best possible combined product, especially suitable for extreme storm events.

2.7 Reference

Alemohammad, S. H., McColl, K. A., Konings, A. G., Entekhabi, D., Stoffelen, A. (2015). Characterization of precipitation product errors across the United States using multiplicative triple collocation. Hydrology and Earth System Science, 19, 3489–3503.

Caires, S. (2003). Validation of ocean wind and wave data using triple collocation. Journal of Geophysical Research, 108.

Cao, Q., Knight, M., & Qi, Y. (2018) Dual-pol radar measurements of Hurricane Irma and comparison of radar QPE to rain gauge data. In Proceedings of the 2018 IEEE Radar Conference (RadarConf18), Oklahoma City, OK, USA, 23–27 April 2018; pp. 496–501.

Cecinati, F., Moreno-Ródenas, A., Rico-Ramirez, M., ten Veldhuis, M.-C., & Langeveld, J. (2018). Considering Rain Gauge Uncertainty Using Kriging for Uncertain Data, Atmosphere, 9, 446.

Cerveny, R. S., Lawrimore, J., Edwards, R., & Landsea, C. (2007). Extreme Weather Records. Bulletin of the American Meteorological Society, 88, 853–860.

Chen, S., Hong, Y., Cao, Q., Kirstetter, P. E., Gourley, J. J., Qi, Y., & Wang, J. (2013). Performance evaluation of radar and satellite rainfalls for Typhoon Morakot over Taiwan: Are remote-sensing products ready for gauge denial scenario of extreme events? Journal of Hydrology, 506, 4–13.

Chen, M., Nabih, S., Brauer, N. S., Gao, S., Gourley, J. J., Hong, Z., Kolar, R. L., & Hong, Y. (2020). Can Remote Sensing Technologies Capture the Extreme Precipitation Event and Its Cascading Hydrological Response? A Case Study of Hurricane Harvey Using EF5 Modeling Framework. Remote Sensing, 445.

Ciach, G. J., & Krajewski, W. F. (1999). On the estimation of radar rainfall error variance. Advances in Water Resources, 22, 585–595.

Dai, Q., Yang, Q., Zhang, J., & Zhang, S. (2018). Impact of Gauge Representative Error on a Radar Rainfall Uncertainty Model. Journal of Applied Meteorology and Climatology, 57, 2769–2787.

Dare, R. A., Davidson, N. E., & McBride, J. L. (2012). Tropical Cyclone Contribution to Rainfall over Australia. Monthly Weather Review, 140, 3606–3619.

Dong, M., Chen, L., Li, Y., & Lu, C. (2010). Rainfall Reinforcement Associated with Landfalling Tropical Cyclones. Journal of Atmospheric Science, 67, 3541–3558.

Emanuel, K. Assessing the present and future probability of Hurricane Harvey's rainfall. Proceedings of the National Academy Science of the United States of America, 48, 12681–12684.

Gao, S., Meng, Z., Zhang, F., & Bosart, L. F. (2006). Observational Analysis of Heavy Rainfall Mechanisms Associated with Severe Tropical Storm Bilis (2006) after Its Landfall. Monthly Weather Review, 137, 1881–1897.

Gao, S., Zhang, J., Li, D., Jiang, H., & Fang, N. Z. (2018). Evaluation of Multi-Radar Multi-Sensor (MRMS) and Stage IV Gauge-adjusted Quantitative Precipitation Estimate (QPE) During Hurricane Harvey. AGU Fall Meeting Abstract, NH42A-07.

Gentemann, C. L. (2014). Three way validation of MODIS and AMSR-E sea surface temperatures. Journal of Geophysical Research, 119, 2583–2598.

Gourley, J. J., Tabary, P., & Parent du Chatelet, J. (2007). A Fuzzy Logic Algorithm for the Separation of Precipitating from Nonprecipitating Echoes Using Polarimetric Radar Observations. Journal of Atmospheric and Oceanic Technology, 24, 1439–1451.

Gourley, J. J., Hong, Y., Flamig, Z. L., Li, L., & Wang, J. (2009). Intercomparison of Rainfall Estimates from Radar, Satellite, Gauge, and Combinations for a Season of Record Rainfall. Journal of Applied Meteorology and Climatology, 49, 437–452.

Guo, H., Chen, S., Bao, A., Behrangi, A., Hong, Y., Ndayisaba, F., & Stepanian, P.M. (2016). Early assessment of Integrated Multi-satellite Retrievals for Global Precipitation Measurement over China. Atmospheric Research, 121–133.

Hong, Y., Hsu, K.-l., Moradkhani, H., & Sorooshian, S. (2006). Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. Water Resources Research, 42.

Hong, Y., Adler, R.F., Negri, A., & Huffman, G. (2007). Flood and landslide applications of near real-time satellite rainfall products. Natural Hazards, 43, 285–294.

Huang, C., Hu, J., Chen, S., Zhang, A., Liang, Z., Tong, X., & Zhang, Z. (2019). How Well Can IMERG Products Capture Typhoon Extreme Precipitation Events over Southern China? Remote Sensing, 11, 70.

Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2019). GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06. Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC). Available online: 10.5067/GPM/IMERG/3B-HH/06 (accessed on 11 November 2019).

Jewell, S. A., & Gaussiat, N. (2015). An assessment of kriging-based rain-gauge-radar merging techniques. Quarterly Journal of Royal Meteorological Society, 141, 2300–2313.

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., & sHsu, K. L. (2011). Braithwaite, D. Intercomparison of High-Resolution Precipitation Products over Northwest Europe. Journal of Hydrometeorology, 13, 67–83.

Kirschbaum, D., Adler, R., Adler, D., Peters-Lidard, C., & Huffman, G. H. (2010). Global Distribution of Extreme Precipitation and High-Impact Landslides in 2010 Relative to Previous Years. Journal of Hydrometeorology, 13, 1536–1551.

Mazzoglio, P., Laio, F., Balbo, S., Boccardo, P., & Disabato, F. (2019). Improving an Extreme Rainfall Detection System with GPM IMERG data. Remote Sensing, 11, 677.

Knight, D. B., & Davis, R. E. (2007). Climatology of tropical cyclone rainfall in the Southeastern United States. Physical Geography, 28, 126–147.

Kirstetter, P.-E., Gourley, J. J., Hong, Y., Zhang, J., Moazamigoodarzi, S., Langston, C., & Arthur, A. (2015). Probabilistic precipitation rate estimates with ground-based radar networks. Water Resources. Research, 51, 1422–1442.

Li, C., Tang, G., & Hong, Y. (2017). Cross-evaluation of ground-based, multi-satellite and reanalysis precipitation products: Applicability of the Triple Collocation method across Mainland China. Journal of Hydrology, 562, 71–83.

Luyckx, G., & Berlamont, J. (2001). Simplified Method to Correct Rainfall Measurements from Tipping Bucket Rain Gauges. In Proceedings of the Conference: Specialty Symposium on Urban Drainage Modeling at the World Water and Environmental Resources Congress, Orlando, FL, USA, 20–24 May 2001; pp. 767–776.

Massari, C., Crow, W., & Brocca, L. (2017). An assessment of the performance of global rainfall estimates without ground-based observations. Hydrology and Earth System Science, 21, 4347–4361.

McColl, K. A., Vogelzang, J., Konings, A.G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. Geophysical Research Letters, 41, 6229–6236.

Medlin, J. M., & Kimball, S. K., Blackwell, K. G. (2007). Radar and Rain Gauge Analysis of the Extreme Rainfall during Hurricane Danny's (1997) Landfall. Monthly Weather Review, 135, 1869–1888.

Mei, Y., Anagnostou, E. N., Nikolopoulos, E. I., & Borga, M. (2014). Error Analysis of Satellite Precipitation Products in Mountainous Basins. Journal of Hydrometeorology, 15, 1778–1793.

Molini, A., Lanza, L. G., & La Barbera, P. (2005). The impact of tipping-bucket rain gauge measurement errors on design rainfall for urban-scale applications. Hydrological Processes, 19, 1073–1088.

Omranian, E., & Sharif, H. (2018). Evaluation of the Global Precipitation Measurement (GPM) Satellite Rainfall Products over the Lower Colorado River Basin, Texas. Journal of American Water Resources Association, 54, 882–898.

Omranian, E., Sharif, H., & Tavakoly, A. (2018). How Well Can Global Precipitation Measurement (GPM) Capture Hurricanes? Case Study: Hurricane Harvey. Remote Sensing, 2018, 10, 1150.

Ratheesh, S., Mankad, B., Basu, S., Kumar, R., & Sharma, R. (2013). Assessment of Satellite-Derived Sea Surface Salinity in the Indian Ocean. IEEE Geoscience and Remote Sensing Letters, 10, 428–431.

Roebeling, R. A., Wolters, E. L. A., Meirink, J. F., & Leijnse, H. (2012). Triple Collocation of Summer Precipitation Retrievals from SEVIRI over Europe with Gridded Rain Gauge and Weather Radar Data. Journal of Hydrometeorology, 13, 1552–1566.

Ryzhkov, A., Diederich, M., Zhang, P., & Simmer, C. (2014). Potential Utilization of Specific Attenuation for Rainfall Estimation, Mitigation of Partial Beam Blockage, and Radar Networking. Journal of Atmospheric and Oceanic Technology, 31, 599–619.

Sarachi, S., Hsu, K., & Sorooshian, S. (2015). A Statistical Model for the Uncertainty Analysis of Satellite Precipitation Products. Journal Hydrometeorology, 2015, 16, 2101–2117.

Seo, D.-J. (1998). Real-time estimation of rainfall fields using rain gage data under fractional coverage conditions. Journal of Hydrology, 208, 25–36.

Sharifi, E., Steinacker, R., & Saghafian, B. (2016). Assessment of GPM-IMERG and Other Precipitation Products against Gauge Data under Different Topographic and Climatic Conditions in Iran: Preliminary Results. Remote Sensing, 8, 135.

Smith, J. A., Villarini, G., & Baeck, M. L. (2012). Mixture distributions and the hydroclimatology of extreme rainfall and flooding in the eastern United States. Journal of Hydrometeorology, 13, 588–603.

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., & Wilheit, T. (2017). The Global Precipitation Measurement (Gpm) Mission for Science and Society. Bulletin of American Meteorological Society, 98, 1679–1695.

Stampoulis, D., & Anagnostou, E. N. (2012). Evaluation of Global Satellite Rainfall Products over Continental Europe. Journal of Hydrometeorology, 13, 588–603.

Stoffelen, A. (1998). Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. Journal of Geophysical Research, 103, 7755–7766.

Sukovich, E. M., Ralph, F. M., Barthold, F. E., Reynolds, D. W., Novak, D. R. (2014). Extreme Quantitative Precipitation Forecast Performance at the Weather Prediction Center from 2001 to 2011. Weather and Forecasting, 29, 894–911.

Sungmin, O., Foelsche, U., Kirchengast, G., Fuchsberger, J., Tan, J., & Petersen, W. A. (2017). Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. Hydrology and Earth System Science, 21, 6559–6572.

Tang, G., Clark, M. P., Papalexiou, S. M., Ma, Z., & Hong, Y. (2020). Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. Remote Sensing of Environment, 240, 111697.

Tian, Y., & Peters-Lidard, C. D. (2010). A global map of uncertainties in satellite-based precipitation measurements. Geophysical. Research Letters, 37, L34407.

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., & Wu, H. (2013). Modeling errors in daily precipitation measurements: Additive or multiplicative? Geophysical Research Letters, 40, 2060–2065.

Villarini, G., & Smith, J. A. (2010). Flood peak distributions for the eastern United States. Water Resources Research., 46.

Yilmaz, M. T., & Crow, W. T. (2014). Evaluation of Assumptions in Soil Moisture Triple Collocation Analysis. Journal of Hydrometeorology, 15, 1293–1302.

Yoo, C., Park, C., Yoon, J., & Kim, J. (2014). Interpretation of mean-field bias correction of radar rain rate using the concept of linear regression. Hydrological Processes, 28, 5081–5092.

Zwieback, S., Scipal, K., Dorigo, W., & Wagner, W. (2012). Structural and statistical properties of the collocation technique for error characterization. Nonlinear Processes in Geophysics, 19, 69–80.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., & Kitzmiller, D. (2016). Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. Bulletin of American Meteorological Society, 97, 621–638.

Zhang, J., & Qi, Y. (2010). A Real-Time Algorithm for the Correction of Bright band Effects in Radar-Derived QPE. Journal of Hydrometeorology, 11, 1157–1171.

Chapter 3

A multi-source 120-year flood database with a unified common format and public access

Publication

Li, Z., Chen, M., Gao, S., Gourley, J. J., Yang, T., Shen, X., Kolar, R., & Hong, Y. (2021). A multi-source 120-year US flood database with a unified common format and public access, Earth System Science Data, 13, 3755–3766, https://doi.org/10.5194/essd-13-3755-2021.

3 Chapter 3. A multi-source 120-year flood database with a unified common format and public access

Abstract

Despite several flood databases available in the United States, there is a benefit to combine and reconcile these diverse data sources into a comprehensive flood database with a unified common format and easy public access in order to facilitate flood-related research and applications. Typically, floods are reported by specialists or media according to their socioeconomic impacts. Recently, data-driven analysis can reconstruct flood events based on in situ and/or remote-sensing data. Lately, with the increasing engagement of citizen scientists, there is the potential to enhance flood reporting in near-real time. The central objective of this study is to integrate information from seven popular multi-sourced flood databases into a comprehensive flood database in the United States, made readily available to the public in a common data format. Natural language processing, geocoding, and harmonizing processing steps are undertaken to facilitate such development. In total, there are 698 507 flood records in the United States from 1900 to the present, which highlights the longest and most comprehensive recording of flooding across the country. The database features event locations, durations, date/times, socioeconomic impacts (e.g., fatalities and economic damages), and geographic information (e.g., elevation, slope, contributing area, and land cover types retrieved from ancillary data for given flood locations). Finally, this study utilizes the flood database to analyze flood seasonality within major basins and socioeconomic impacts over time. It is anticipated that thus far the most comprehensive yet unified database can support a variety of flood-related research, such as a validation resource for hydrologic or hydraulic simulations, hydroclimatic studies concerning spatiotemporal patterns of floods, and flood susceptibility analysis for vulnerable geophysical locations. The dataset is publicly available with the following DOI: https://doi.org/10.5281/zenodo.4547036 (Li, 2020).

3.1 Introduction

Floods are one of the most common and costliest natural hazards globally (World Health Organization). In fact, around 74 % of natural hazards between 2001 and 2018 were water-related, among which floods were the most devastating. In the United States, eight of the 10 costliest weather disasters (in billions of USD) were floods between 1980 and 2019 (https://www.ncdc.noaa.gov/billions/events, last access: 10 February 2021), and almost 10 % of the flash floods have resulted in agricultural and economic losses beyond USD 100 000 (US dollars) per event (Gourley et al., 2017). The flood-producing storms and hurricanes frequently strike the coastal regions with devastating socioeconomic impacts, among which the most damaging Hurricane Katrina affected nine states and resulted in monetary losses of over USD 168 billion. Moreover, under the influences of climate change, the increasingly intensified hydrologic cycle and sea level rise pose more threats to coastal areas (Alfieri et al., 2016; Tabari, 2020). IPCC AR5 (2014) has reported that the frequency and intensity of floods in the United States are changing, which challenges current water-related infrastructure and water management principles. In light of flood risks, a compilation of a comprehensive flood database can provide insights into both national and regional flood characteristics.

Many published works have hitherto been limited to developed countries such as European countries and the United States. Developing countries either restrict data sharing or lack the resources to collect and assemble flood events. With respect to the available period, not many works continuously offer up-to-date flood data accessible to the public or research communities (e.g., Fiorillo et al., 2018; He et al., 2018; Luu et al., 2019; Petrucci et al., 2019; Shi, 2003). However, it is noteworthy that there are means of collecting flood information. Conventionally, flood reports are produced by local specialists with limited and sometimes delayed information (e.g., Filrilo et al., 2018; He et al., 2018; Luu et al., 2019; Petrucci et al., 2019). Later on, media outlets (e.g., newspaper) start to participate in timely flood reporting, but typically on the high-impact floods (e.g., Hilker et al., 2009; Shi, 2013; Smith et al., 2012; Vos et al., 2010). Insurance companies collectively offer valuable information on flood damages and people affected from a financial perspective (Swiss Re, 2010). Until recently, the increasing engagement of social scientists has greatly supported near-real-time flood reporting with web or mobile applications (Chen et al., 2016; de Bruijn et al., 2019), although these reports are often confined to populated urban areas. In addition to human-led reporting, stream gage and opportunistic sensors (e.g.,

surveillance cameras, ground radars, and satellites) can also augment flood monitoring in real time (Hall et al., 2015; Shen et al., 2019).

Despite long-established flood records (reports), there are few studies attempting to merge multi-source flood databases, especially considering the increasing number and diversity of flood databases available. The motivations of a merged dataset are primarily twofold. First and foremost, we are still under-utilizing all sorts of flood information that can be used for model validation and flood risk analysis (Scotti et al., 2020). Second, each individual dataset has its own limitations, and thereby no single database holistically describes flooding in a given region (Gourley et al., 2013). For instance, flood reports by government agencies or the media are skewed towards high-impact events, whereas local community-level, low-end floods are oftentimes ignored. In light of these motivations, efforts should be undertaken to collectively merge all possible sources to provide off-the-shelf data support to complement flood-related research. Gourley et al. (2013) assembled a georeferenced US database from three primary sources: (1) discharge observations from the U.S. Geological Survey (USGS), (2) flood reports by the National Weather Service from 2006 to 2013, and (3) witness reports from the public. Amponsah et al. (2018) merged a high-resolution flash flood database in Europe with a set of spatial data, rainfall data, and discharge data from 1991 to 2015. Petrucci et al. (2019) collaboratively harmonized five regional flood databases from 1980 to 2015 in the Mediterranean region to investigate the causes of deaths in flood events. These merged datasets are relatively short in time and not complete. In this study, we introduce a comprehensive United States Flood Database – USFD, which compiles seven individual databases and converts them into a common data format. Sources to compile this database include (1) reports from news media, (2) reconstructed flood events from gage and satellite instruments, and (3) crowdsourcing data queried from the web and mobile applications. As a result, a 120-year flood database in the United States is assembled, unified, and published for public access, as well as an interactive web interface for immediate use. This dataset includes diverse flood subtypes, including riverine flooding, coastal flooding, flash flooding, etc., and features the longest and most comprehensive recording of flooding across the country. It is anticipated that this database can support a variety of flood-related research, such as a validation source for hydrologic/hydraulic simulation, climatic studies concerning spatiotemporal patterns of floods given this long-term and US-wide coverage, and flood risk analysis for vulnerable geophysical locations. Primary

assessments on flood occurrences across the Unites States, flood seasonality within major basins, and socioeconomic impacts across time are carried out to share insights into US floods.

This article is structured as follows. Section 2 details seven individual databases and ancillary datasets used to form our database. Section 3 describes methods to retrieve (query), clean, and unify these datasets in a processing pipeline. Lastly, Sect. 4 serves as a pre-assessment on floods in the United States over the past 120 years, spatially aggregated by geopolitical boundaries and for major US river basins.

3.2 Database components

3.2.1 Individual database

In this section, we detail seven individual databases, which are the NOAA National Weather Service (NWS) storm reports, Emergency Events Database (EM-DAT), Dartmouth Flood Observatory database (DFO), the University of Connecticut Flood Events Database (FEDB), cyber-infrastructure flood database (CyberFlood), meteorological Phenomena identification near the ground data (mPing), and Global Flood Monitoring (GFM). Each candidate of this compiled database has to satisfy the following criteria: (1) published by trustworthy organizations, (2) has been used in at least one traceable high-impact publication, and (3) contains useful information for flood-related research.

3.2.1.1 National Weather Service storm reports

The NOAA NWS routinely publishes post-event reports of floods from trained spotters, local authorities, and emergency management officials. This dataset is arguably the most exhaustive meteorology-driven reporting in the United States. The descriptors can be mainly categorized into the geophysical location (e.g., begin and end location), time period (e.g., begin and end time), causes (e.g., heavy rain), impacts (e.g., fatalities and damages), and narratives (see technical documentation for details: https://www.nws.noaa.gov/directives/sym/pd01016005curr.pdf, last access: 20 December 2020). Limitations of this database for flood events are summarized in Gourley et al. (2013), including (1) imprecise event location, (2) times related to meteorological events, (3) relying on in-person witness accounts, and (4) limited information about the site exposure to

antecedent condition. We retrieve all flood records from 1950 to the present, which totals 144 313 reports.

3.2.1.2 Emergency Events Database (EM-DAT)

The EM-DAT database is produced and maintained by the Centre for Research on the Epidemiology of Disasters (CRED) in Belgium, which contains all types of global natural disasters in the world from 1900 to the present. These recorded events should meet one of the following criteria: (1) > 10 people dead, (2) > 100 people affected, (3) declaration of a state of emergency, or (4) a call for international assistance. Therefore, regional floods or street-level floods are not included in this database. The sources of information stem from government agencies, non-government organizations, insurance companies, research institutes, and press agencies. The EM-DAT provides information including geographic location, time entry, fatalities, and economic damages. The geographic location is uncertain to studies considering precise flood locations such as inundation mappings, and the economic damages are obtained from insurance claims. All the flood-related data entries are collected via public access at https://www.emdat.be/ (last access: 20 December 2020). Due to its stringent reporting criteria, there are only 189 events recorded in the United States.

3.2.1.3 Dartmouth Flood Observatory (DFO)

The DFO data, regarded as one of the most popular flood databases in the world, collects flood events from news, government agencies, and stream gauges and remote-sensing instruments from 1985 to the present (Brakenridge, 2020). Different from other databases, the DFO collectively retrieves spatial flood information from satellite remote-sensing products, such as the Moderate Resolution Image Spectroradiometer (MODIS), Sentinel-1, and Landsat. Flood extent is accordingly provided as shapefiles for easy integration into Geographic Information Systems software. The tabular data include geophysical location, date/time, fatalities, affected area, displaced people, flood severity, and primary causes. However, events without significant river flooding are not included in this database, and they are subject to uncertainties from satellite-derived flood extent such as water-like echoes in urban areas and limitations due to cloudiness. A total of 469 events have been retrieved from its tabular data in the United States.

3.2.1.4 University of Connecticut Flood Events Database (FEDB)

Taking advantage of nation-wide flow records at 6301 stations operated by the U.S. Geological Survey (USGS) and radar rainfall measurements, a comprehensive flood database is reconstructed from 2002 to 2013, using the characteristic point method (Shen et al., 2017). At each gauge site, flood events are identified by baseflow separation and filtered with non-significant peaks (i.e., less than 95th percentile). Additionally, flood-producing rainfall events are traced within a certain time window to portray an event. The FEDB provides shapefiles of stream gauges with a series of flood event attributes (e.g., flow peak, flow period, rainfall event period, base flow, rainfall–runoff coefficient, and spatial moments-based characteristics). The original dataset is retrieved from https://ucwater.engr.uconn.edu/fedb (last access: 25 December 2020). Limitations of this database are that the reconstructed events may not necessarily lead to damages, which may undermine its role in flood-impact-related research, and the flood events must occur in USGS-gauged basins. In major flooding events (e.g., Hurricane Harvey), some stream gauges may be flushed away, so no event is recorded. Over 542 000 events have been reconstructed in the United States, making this flood database the biggest contributor to the combined database.

3.2.1.5 CyberFlood

CyberFlood is a crowdsourced flood database by collecting event reports in a web application developed at the University of Oklahoma (Wan et al., 2014). It is regarded as one of the first integrated systems that collect, organize, visualize, and manage a flood database globally. However, the flood records may be falsely reported due to lack of cross-reference scrutiny. We queried the latest results of CyberFlood, which contains flood events, geographic locations, date/time, country code, causes, and fatalities. The latest version of CyberFlood has 203 flood records from 1998 to 2008. To facilitate data unification, we convert all the code-based descriptors to strings (i.e., country and causes) with key matching methods.

3.2.1.6 Meteorological Phenomena Identification Near the Ground (mPING)

The mPing app is a crowdsourcing, weather-reporting software jointly developed by NOAA National Severe Storms Laboratory (NSSL) and the University of Oklahoma (Elmore et al., 2014). Members of the public who downloaded this app based on their GPS-enabled smartphones can report the weather event at their locations. Time, geophysical coordinates, and

standard event types (e.g., flood events classified into four severity levels, tornado, precipitation type, wind) are provided. The four flood severity levels are based on the flash flood severity index (FFSI) proposed by Schroeder et al. (2016). Like any other crowdsourcing datasets, the major limitation lies in the data validation, as some events are improperly misreported or even hacked with data injection. Chen et al. (2013) compared these reports to ground radar observations with respect to precipitation types, and a satisfying correspondence is found between the two. mPing data provide REST API for research-purpose uses, and we queried flood-related events from 2013 to the present with 5000 flood events counted.

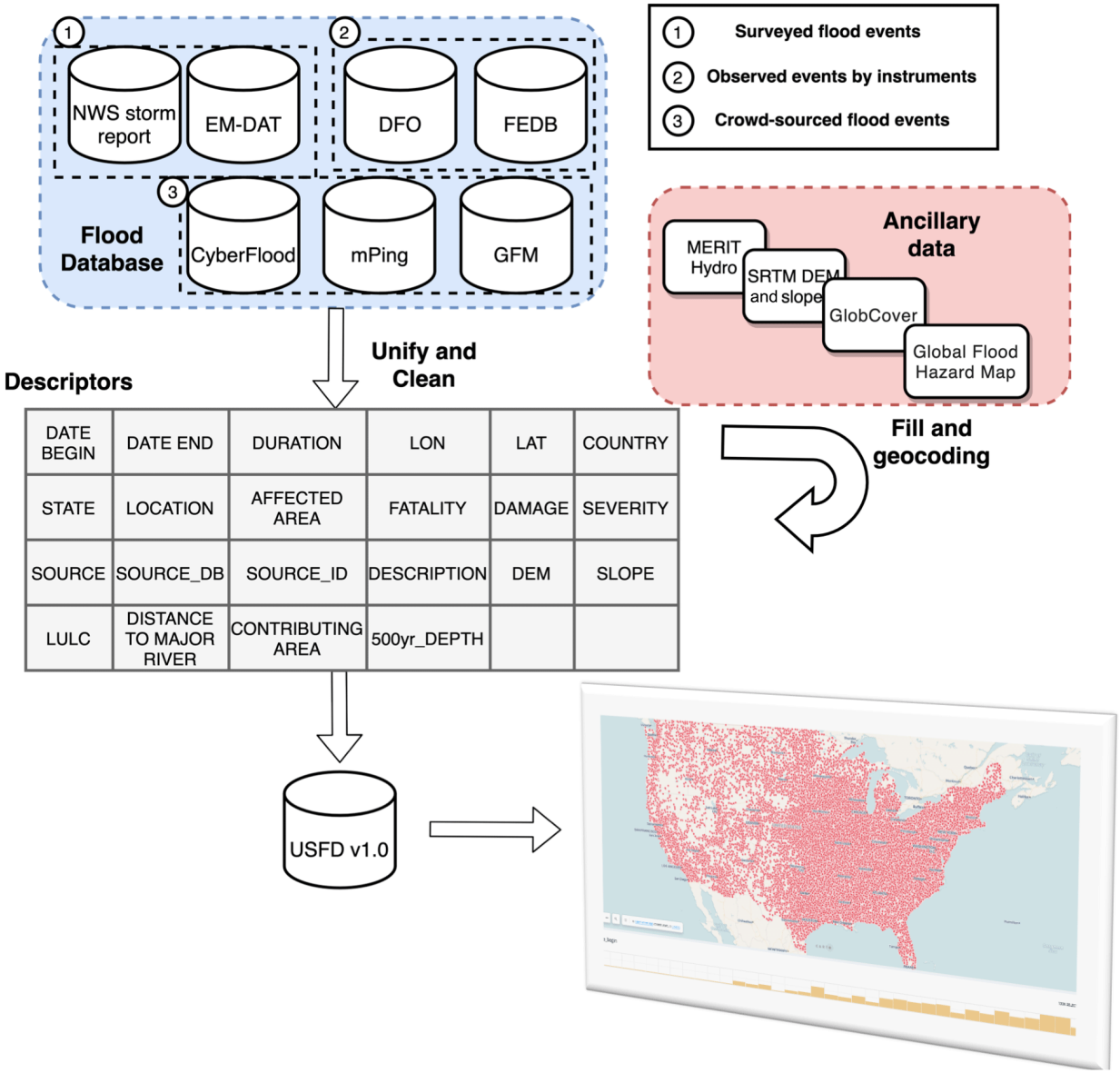
3.2.1.7 Global flood monitoring (GFM)

The GFM data are produced and managed by de Bruijn et al. (2019), with over 88 million Twitter tweets over the globe since 2014. Contents tied to flood observations are filtered with the natural language processing (NLP) tool BERT, which extracts time of observation and toponyms (in token) and assigns reports to the database attributes after a quality assessment. It is found in the study of Bruijn et al. (2019) that around 90 % of the events are correctly detected when compared to another disaster database. Table attributes include event_id, location_id, location_ID_url, country_ID, country_ISO3, and the time of detection. Due to privacy issues, all the locations are archived in tokens, which requires further decoding. Uncertainties related to this dataset are its geographic locations and technical algorithms to filter events. For studies requiring precise flood locations, this dataset may not be a good candidate. Data are publicly accessible at https://www.globalfloodmonitor.org/download (last access: 22 December 2020). Given the latest database, we retrieved 6315 flood events in the United States and subsequently processed them as described in Sect. 3.

3.2.2 Ancillary dataset

Since one purpose of this database is for flood susceptibility analysis, contributing factors to flooding are also incorporated for a given location. Land use–land cover (LULC), digital elevation model (DEM), slope, distance to a major river, drainage area, and 500-year flood depth are factored into the data attributes. The LULC value is retrieved from the Copernicus Global Land Service (CGLS) at 100 m resolution, covering urban, cultivated land, forest, vegetation, wetland, water, and ice. The topographic inputs (i.e., DEM and slope) are acquired from the NASA Shuttle

Radar Topography Mission (SRTM) at 90 m spatial resolution, and hydrography datasets (i.e., river networks, drainage area) are acquired from MERIT Hydro at the same resolution (Yamazaki et al., 2019). The 500-year flood depth is downloaded from the Joint Research Centre Data Catalogue at https://data.jrc.ec.europa.eu/collection/floods (last access: 27 December 2020) at 1 km spatial resolution. All the extensive computations (i.e., sampling) are processed using the Google Earth Engine platform (Gorelick et al., 2017).



DATE BEGIN	DATE END	DURATION	LON	LAT	COUNTRY
STATE	LOCATION	AFFECTED AREA	FATALITY	DAMAGE	SEVERITY
SOURCE	SOURCE_DB	SOURCE_ID	DESCRIPTION	DEM	SLOPE
LULC	DISTANCE TO MAJOR RIVER	CONTRIBUTING AREA	500yr_DEPTH		

Figure 3.1 Flowchart of dataset processing.

Table 3.1 Description in the USFD database.

Headers	Unit/format	preprocessing	Description	Sample
DATE_BEGIN	yyyymmdd (hh:mm:ss)	Pattern extraction, converting	Event beginning date–time (UTC)	2.02E+11
DATE_END	yyyymmdd (hh:mm:ss)		Event end date–time (UTC)	
DURATION	Days	End date–begin date	Duration of an event	10
LON	Decimal degree	Check validity; remove	Longitude of an event	(−120,60)
LAT	Decimal degree	unreasonable (long,lat)	Latitude of an event	
COUNTRY	String	Mapping country code to full name	Country name	The United States of America
STATE	String	Replace or fill names with (long, lat)	State name	Oklahoma
LOCATION	String	None	Location of an event	Bryan
AREA	Square kilometers	Calculate affected areas from reported storm range	Event-affected area	1000
FATALITY	Int	Check data type	Number of fatalities	43
DAMAGE	USD	Sum up sub-category costs; check data type; convert to dollars	Economic damages (direct)	107
SEVERITY	NA	None	Severity of an event(according to DartmouthFlood Observatory data)	1,1.5,2
SOURCE	String	None	Collecting sources	Newspaper
SOURCE_DB	String	None	The original recorded database	NOAA storm report
SOURCE_ID	String	None	The original event ID insource database	102300
CAUSE	String	None	Causes of a flood event	Heavy rain
DESCRIPTION	String	None	Event narratives	River overflowing/bankfull
DEM	Meters	Retrieved from Shuttle Radar Topography Mission	Elevation	120
SLOPE	Degree	Derived from DEM	Slope	10
LULC	Class	Retrieved from Copernicus global land cover 2019	Land use–land cover classification	Urban
DISTANCE_RIVER	Kilometers	Distance computed in Google Earth Engine	Distance from event location to nearest major river	3.5
CONT_AREA	Square kilometers	Retrieved from MERIT Hydro	Contributing area	1.35
500yr_DEPTH	Meters	Retrieved from 500-year flooddepth	500-year flood depth	1.23

3.3 Processing methods

Figure 3.1 displays the processing flowchart, including pre-processing, merging, and unifying all seven individual databases comprising the USFD v1.0. There are 22 descriptors in the database, including the start and end time (UTC), duration (days), longitude and latitude (decimal degrees), toponyms (country, state, and location), flood impacts (i.e., affected area, severity, damage, and fatality), sources (i.e., source database, source ID, collecting sources), event description, and environmental variables as mentioned in Sect. 2.2. Detailed header descriptions and pre-processing steps are summarized in Table 3.1. Two intrinsic factors describing flood events are date–time and locations. The date–time information varies in different databases. Some early reports do not record the precise date–time of an event onset, using only year or year and month. For clarity, we format them in a concatenated string with maximum available information. The date 23 June 2015 00:00:00 is recoded as "20150623000000", and June 2015 is inserted as "201506". The date–time of all records are converted to UTC.

It is challenging to completely retrieve the location of events from some databases. The NWS storm report has some missing entries in geographic coordinates, but it has detailed narratives. To compensate, we use the NLP toolkit provided by the spaCy package, which contains pre-trained models for English multi-tasks. spaCy firstly tokenizes the event narratives and subsequently parses and tags each word with respective entities. Then, we can geocode locations into geographic coordinates via calling the Google Map API. The GFM also does not contain precise geophysical locations to protect user privacy. Therefore, the geographic coordinates are inferred by first converting location tokens into administrative locations (e.g., cities and villages) via the GeoNames API and then geocoding them into geographic coordinates. The state names of a merged dataset are firstly validated with geophysical locations from an inverse geocoding. If they do not match, a new name from GeoNames is assigned to replace the original one. Meanwhile, the empty fields are also filled during this process. We also processed supplementary flood information such as affected areas and damages from the original database. The affected area for each event is calculated by assuming a circular area whose radius is approximated by the recorded range if available. For economic damages, we sum up all available sub-category damages (e.g., agriculture, property, and structure) to give a holistic view. For those single databases that do not provide information or information that cannot be inferred from the specific header, we uniformly treat them as not-a-number (NAN) values. As a result, we merged 698 507 total flood records in

the United States from 1900 to the present. In the DOI link, a merged database USFD, along with seven individual databases, is provided in either comma-separated format or Excel format for general readability. Additionally, an interactive web interface is built and hosted at http://hydro.ou.edu/research/us-flood-database/ (last access: 10 July 2021), where users can do immediate analysis online and download the datasets.

3.4 Pre-assessment

3.4.1 Nationwide distributions

Figure 3.2a shows the nationwide distribution of flood events at state levels. Because the total event numbers might be skewed by replicated events in different databases, we standardize the event counts of each state by the maximum number of events, including repetitive events from different sources, out of all states to reveal the relative composition. It is noteworthy that we do not intend to discard repetitive events because they are reported with different attributes and uncertainties in different candidate databases; it is up to users to select the one that fits into their scopes. North Carolina, Texas, Missouri, and Pennsylvania are the top-listed regions with over 50 % of the total population, among which North Carolina experiences the most cases. From the meteorological perspective, North Carolina is prone to flooding due to a mixture of a flood-generating mechanisms, with landfalling tropical cyclones and extratropical systems being the primary large-scale drivers, in conjunction with warm-season thunderstorms. In the meantime, the combination of snowmelt and rain-on-snow contributes to flood peak occurring on the lee side of the Appalachian Mountains (Smith et al., 2011). The most devastating flood, estimated as a 500-year flood, was caused by Hurricane Floyd and led to 35 fatalities. Texas, similar to North Carolina, is affected by tropical cyclones and hurricanes, which produce compound inland and coastal flooding (Li et al., 2020). Anthropogenic effects such as urbanization and regulation, apart from meteorological effects, are equally critical for inland flooding. Blum et al. (2020) in a recent study noted that these listed regions experienced increased urbanization from 1974 to 2012, resulting in an average 3.3 % increased flood magnitude by changes in impervious cover alone.

Figure 3.2b lists detailed event numbers for each state and the composition of each individual flood database. The FEDB contributes a major portion of the unified database because of the data length of flow records, and states with higher gauge densities undoubtedly yield more event numbers than gauge-sparse regions. The data nonuniformity underlies a major limitation of

this specific dataset. Regions with more exposure to observational sources (e.g., densely populated and gauge-dense areas) likely have more recorded events. However, it is expected that by including more observations from remote-sensing sources, this gap can be potentially compensated for. Following the composition, NWS storm reports comprise the second largest number of events because of the long data length (i.e., 70 years). Other databases, such as EM-DAT, though the longest available length, only record very-high-impact events, and the crowdsourcing databases are limited by their short lengths.

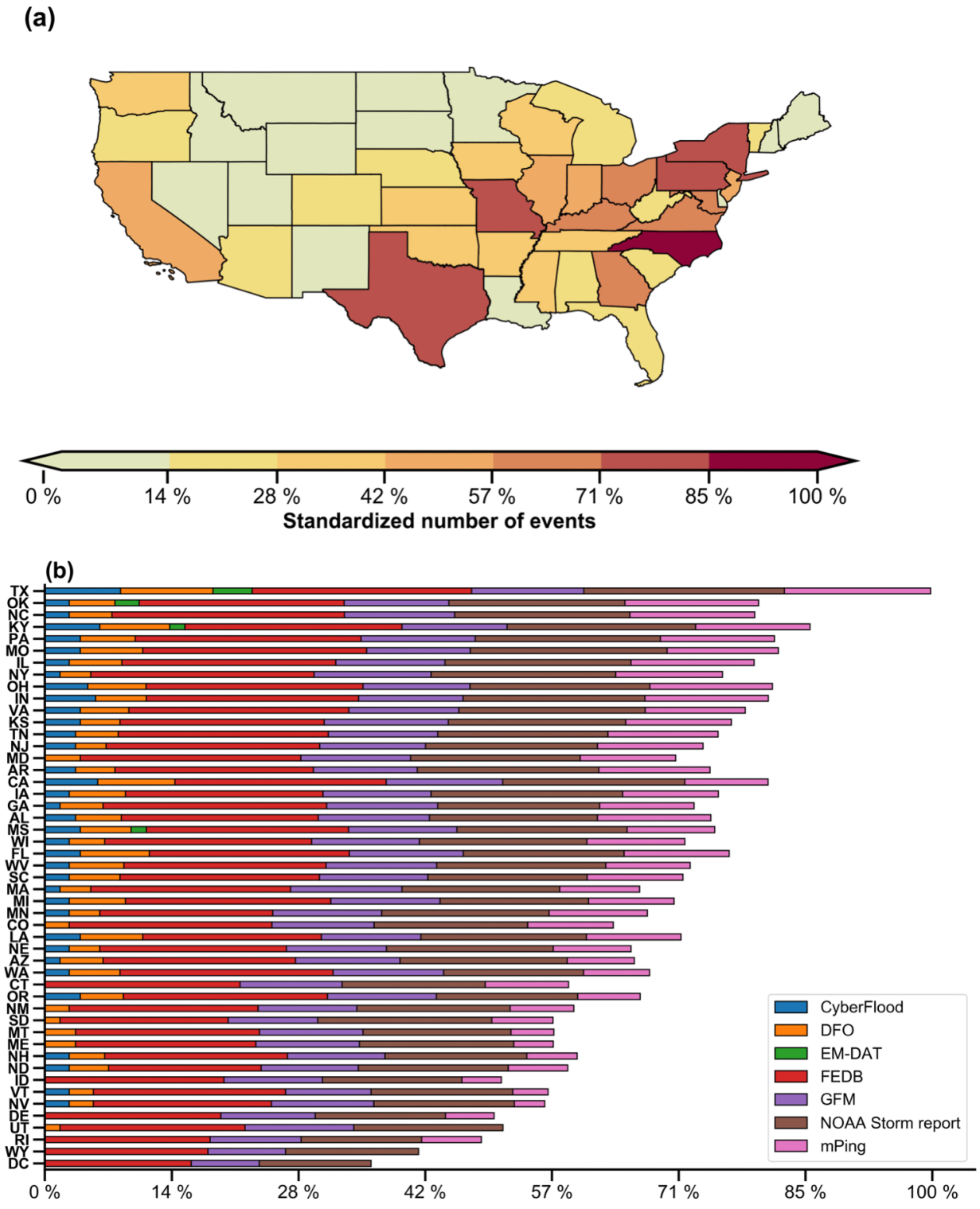


Figure 3.2 (a) Map of min-max standardized flood occurrences in the United States; (b) fractions of logarithmic event numbers of each candidate database to logarithmic total event numbers.

3.4.2 Flood seasonality in major water basins

Flood variability is highly associated with seasonal atmospheric pathways of moisture delivery and basin attributes (Dickinson et al., 2019). In this regard, we segregate the nation-wide events into major basins and months. The hydrologic unit code (HUC) four-digit basins, as shown in Fig. 3.3, are obtained from the national hydrography dataset. Figure 3.3a depicts months with the highest number of events. Flooding generally happens between January and June over the majority of the US basins, similar to that of other studies (e.g., Brunner et al., 2020; Dickinson et al., 2019; Villarini, 2016). The basins are clustered into several regions according to local hydroclimatologies. On the west coast (e.g., western Washington, Oregon, and California), flood events are dominant in winter months because of atmospheric rivers (ARs) as a main driving factor, which are a carrier of water vapor from the tropics (Ralph et al., 2006). Moving to the east, floods in the Rockies (i.e., Upper Colorado and Great basins) are characterized by spring snowmelt in snow-fed rivers, whereas in the Desert Southwest (e.g., Lower Colorado and Rio Grande regions), floods likely occur in late summer, which is ascribed to the North American monsoon and North Pacific tropical cyclones. Closer to the Gulf of Mexico, flooding events during late spring and summer are due to severe thunderstorm activity and mesoscale convective systems. The lower Mississippi, Ohio, and Tennessee river basins experience their biggest floods during the spring from extratropical cyclones (Lavers and Villarini, 2013). The lower Florida Peninsula features high numbers of summer flood events, which are tied to North Atlantic tropical cyclones (Villarini et al., 2014). In the northeastern United States, tropical cyclones, winter–spring extratropical cyclones, and warm-season thunderstorms are the primary flood agents, yet winter–spring extratropical cyclones account for larger fractions (Smith et al., 2011; Villarini, 2016). Figure 3.3b displays the number of flood events within each basin, grouped by months. The Mid-Atlantic region (HUC2-02) takes seven places out of the top 20 HUC4 basins listed in Fig. 3.3b, with the Delaware River basin near the coast (HUC2-0204) being the highest one. In terms of flood seasonality, it is relatively evenly spanned across seasons and months for these listed basins, which suggests its susceptibility to widespread floods. This symmetric feature around the Appalachian Mountains across seasons is also highlighted in Villarini (2016), in which they suggest flow regulations play an essential role in weakening the seasonal cycle. In a recent study by Brunner et al. (2020), these basins are identified as severe or moderate widespread flooding in space, and our results indicate these regions also have widespread flooding in time (month).

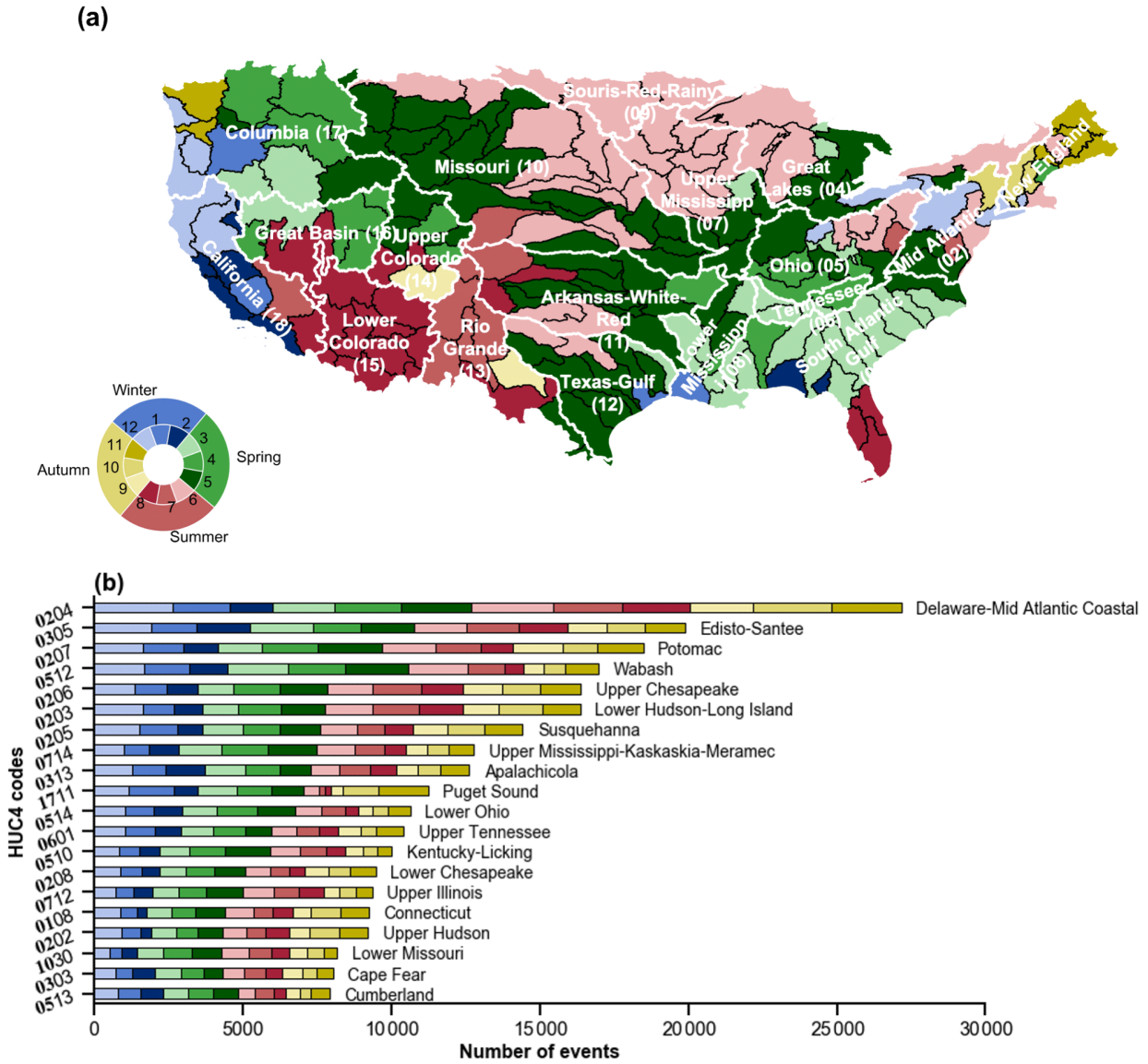


Figure 3.3 (a) Map of months with largest flood occurrences in major US watersheds (Hydrologic Unit Code 4). (b) Flood occurrences of the top 20 HUC4 basins (HUC2 codes in bold), grouped by months.

Figure 3.4 depicts the flood seasonality for two separate periods, 2000–2010 (panel a) and 2010–2020 (panel b), which investigates potential shifts in flood timing for US river basins. On the west coast (California and Columbia basin), there is a shift from early winter flooding to late winter or early spring flooding, especially near the northern coast. This probably relates to snowmelt occurring in early spring, in conjunction with the rain-on-snow effect. The Great Plains feature an earlier maximum flood frequency, transitioning from early summer to late spring, which relates to enhanced and earlier timing of thunderstorm activities due to spring warming. The south

Atlantic Coast shows a delayed maximum flood frequency from winter to spring. The lower Florida Peninsula, however, does not present a clear monthly shift, which is still controlled by tropical cyclones.

3.4.3 Flood impact assessment

In the USFD, flood impacts are based on affected areas, economic damages, and fatalities. Since affected areas are relatively subjective, they are not analyzed in this study. All the economic damages (US dollars) are adjusted for inflation with GDP deflectors obtained from the World Bank. To avoid repeated counts of damages or fatalities due to repetitive events, we herein calculate the mean values per event. Figure 3.5 depicts the fatalities and damages by year. Although events continuously span from 1900 to the present, impact assessments were not provided in the earlier years (before 1980). The 10-year running mean represents the long-term trend. Both fatalities and damages begin at high rates in the early years, possibly due to the immature understanding of floods and lack of flood protection measures. The 1964 flood event which happened in the Pacific Northwest and northern California during the Christmas holiday, also known as the "Thousand Year Flood", caused hundreds of millions of dollars in damages, and over 10 people lost their lives. Since 1990, however, with the improved measures in flood prediction, management, and protection, fatality rates have started to decrease except for some highlighted major events. Yet, in recent years, damages have shown a slight upward trend, which is tied to frequent floods caused by intensified active hurricane events and anthropogenic effects. For instance, the 2011 and 2012 Atlantic hurricane seasons are deemed the third and fourth most active hurricane season on record. The 2017 hurricane season featuring Harvey, Irma, and Maria was the costliest hurricane season on record, which is reflected in flood-related damages. Land surface changes such as urbanization continue to develop a conducive environment for urban flooding.

State-specific damages across time shown in Fig. 3.6 reveal the trends in flooding hotspots. For Texas and Louisiana, consistent upward trends are present because of intensified extreme events in the Gulf Coast. The slopes of annual flood damage curves are the greatest among the identified hotspots, manifesting potential flood risks. North Carolina has experienced an increase in damages in the past 10 years, accompanied by major events during the 2016 and 2018 Atlantic hurricane seasons. Florida, similar to North Carolina, has encountered active hurricane seasons

and accompanying damages. In summary, these flood hotspots with increased flood damages should raise awareness from policy-makers and the public.

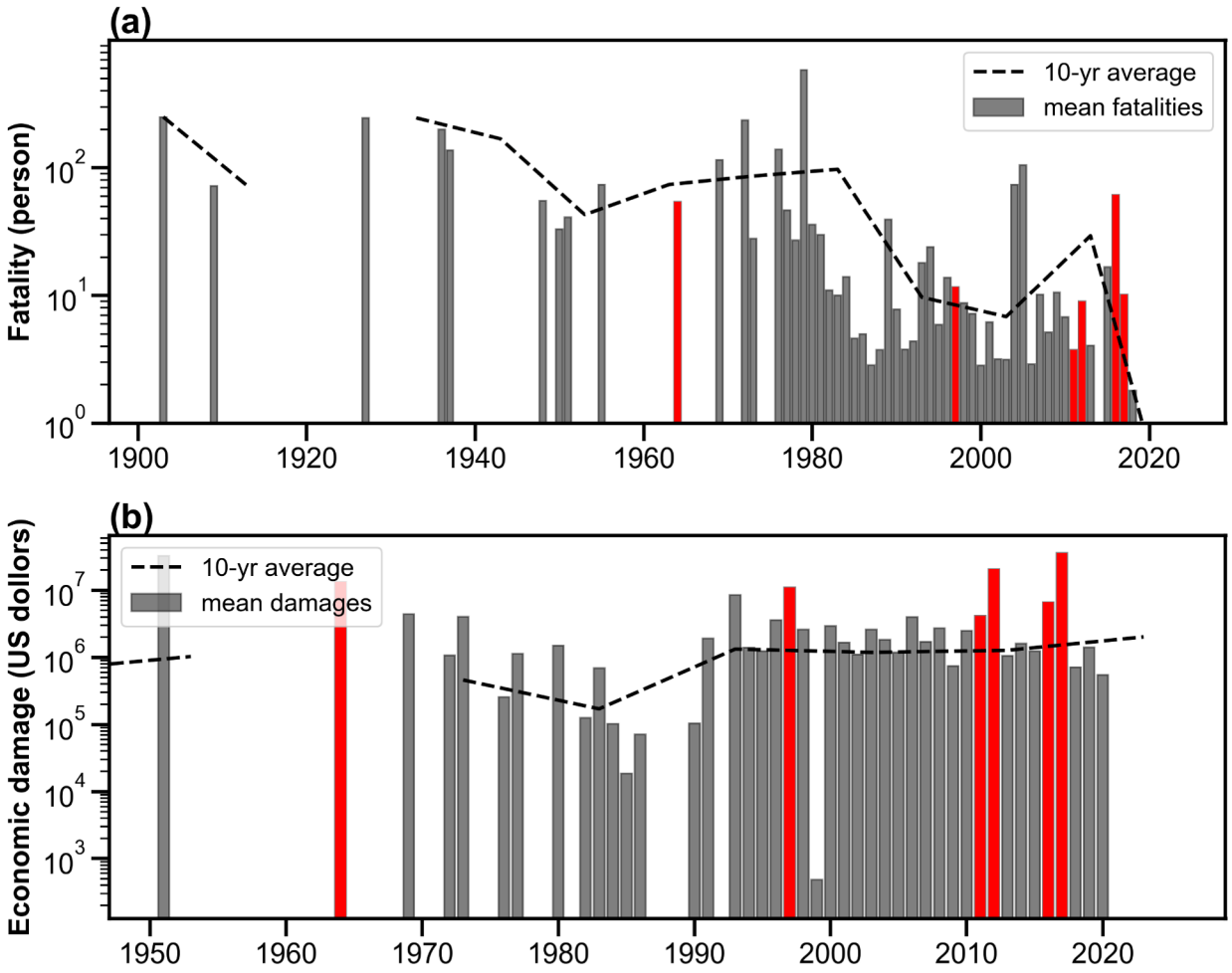


Figure 3.4 Time series of flood impacts: (a) mean fatalities per event and (b) mean economic damages per event (US dollars based on 2020). The highlighted bars represent 1964, 1997, 2011, 2012, 2016, and 2017, which are the active hurricane seasons on record.

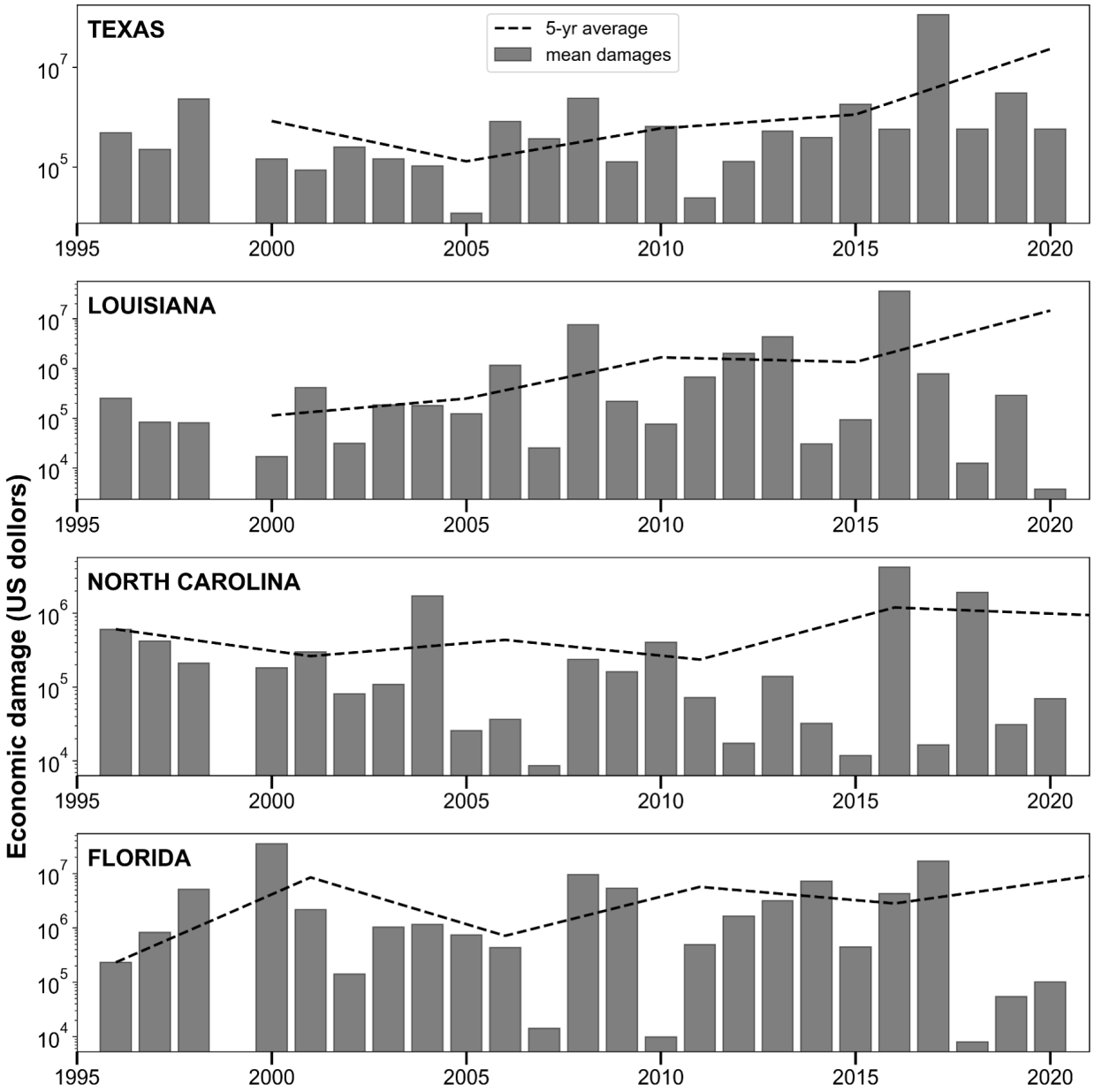


Figure 3.5 Time series of regional economic damages at 2020 values.

3.5 Conclusions and outlook

This work presents a merged United States Flood Database (USFD) that features the longest and most comprehensive recording of flooding across the country. The merged database, integrated from multiple sources, can overcome limitations inherent to the individual databases and thus maximize benefits. It is expected that this database can support a variety of flood-related research, such as a validation resource for hydrologic and hydraulic model simulations,

hydroclimatic studies concerning spatiotemporal patterns of floods, and flood susceptibility analysis for vulnerable geophysical locations.

We showcase three analyses based on the developed flood database. For flood occurrences across the United States, Texas, Pennsylvania, and Missouri are highlighted with great exposure to floods in total amount, which could raise awareness from policy-makers and the public. Flood seasonality in major river basins generally follows the large-scale synoptic weather patterns. In addition, delayed timings of maximum flood frequency are observed in the west coast and Atlantic River basin, possibly due to earlier snowmelt than in prior decades that now contributes to spring floods. Floods in the Great Plains, on the contrary, feature an earlier month of maximum flood frequency, which is possibly tied to intensified thunderstorm activities because of earlier spring warming. Lastly, flood impacts are assessed in terms of economic damages and fatalities, and we found a slight increasing trend in damages in recent years. Especially in Texas and Louisiana, a consistent increase in damages is evident, which relates to intensified storm activity and expanding urban zones. Under a warming climate, storms are projected to occur more frequently in the future, which challenges current water infrastructures and water management principles (IPCC, 2014).

Notwithstanding, there are some limitations associated with the current version of USFD. First, the individual databases disproportionally make up the merged one. FEDB, taken from streamflow records over a long history, consists of the majority of the flood events. The emerging crowdsourced databases are expected to play a significant role with the increasing engagement of citizen scientists. Space-based observations could markedly bridge the gaps between well-observed urban areas and gauged basins to gauge-sparse areas in rural zones. For instance, complete use of the MODIS imagery on board Terra and Aqua satellites, in association with Landsat or synthetic aperture radar data, can reconstruct global flood events at daily resolution. In addition to flood extent, flood depth can also be approximated through the use of high-resolution DEM data. Floods, reported by insurance companies, offer another angle to not only record events, but relate flood hazards to societal impacts comprehensively. In the future, we hope to incorporate such a dataset to enrich our database at a global scale. Second, uncertainties exist in each candidate database. The web-based crowdsourcing dataset may be less reliable because of lack of stringent data scrutiny, especially for studies that require precise flood locations. Developing new guidance for citizen scientists is a remedy. The instrument uncertainties are subject to confined locations such as stream gauges and technical algorithms to retrieve flood information such as remotely

sensed observations. Algorithm developers should take into account how to quantify uncertainties in addition to the end product. The flood reports from government agencies are relatively less uncertain. Therefore, it is highly recommended that users select the candidate databases that fit their research scope. We also encourage each flood database, during its original development, to produce uncertainty measures quantitatively. In a future work, we will consider ways of incorporating uncertainty measures from individual flood databases and harmonize them into one for delivering comprehensive flood information. Third, the data-processing framework needs to be automated in real time. We plan to migrate processing codes to the cloud, so that they can query records from the child databases and update the parent database on a regular basis.

3.6 Reference

Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., De Roo, A., & Salamon P. (2016). Global projections of river flood risk in a warmer world, Earth's Future, 5, 171–182, https://doi.org/10.1002/2016EF000485.

Amponsah, W., Ayral, P.-A., Boudevillain, B., Bouvier, C., Braud, I., Brunet, P., Delrieu, G., Didon-Lescot, J.-F., Gaume, E., Lebouc, L., Marchi, L., Marra, F., Morin, E., Nord, G., Payrastre, O., Zoccatelli, D., & Borga, M. (2018). Integrated high-resolution dataset of high-intensity European and Mediterranean flash floods, Earth System Science Data, 10, 1783–1794, https://doi.org/10.5194/essd-10-1783-2018.

Brakenridge, G. R. (2020). Global active archive of large flood events, Dartmouth Flood Observatory, University of Colorado, available at: http://floodobservatory.colorado.edu/Archives/index.html, last access: 28 December 2020.

Brunner, M. I., Papalexiou, S., Clark, M. P., & Gilleland, E. (2020). How probable is widespread flooding in the United States?, Water Resources Research, 56, e2020WR028096, https://doi.org/10.1029/2020WR028096.

Chen, S., Gourley, J. J., Hong, Y., Cao, Q., Carr, N., Kirstetter, P. E., Zhang, J., & Flamig, Z. (2016). Using citizen science reports to evaluate estimates of surface precipitation type, Bulletin of American Meteorological Society, 97, 187–193, https://doi.org/10.1175/BAMS-D-13-00247.1.

de Bruijn, J. A., de Moel, H., Jongman, B., de Ruiter, M. C., Wagemaker, J., & Aerts, J. C. J. H. (2019). A global database of historic and real-time flood events based on social media, Scientific Data, 6, 311, https://doi.org/10.1038/s41597-019-0326-9.

Dickinson, J. E., Harden, T. M., & McCabe, G. J. (2019). Seasonality of climatic drivers of flood variability in the conterminous United States, Scientific Reports, 9, 15321, https://doi.org/10.1038/s41598-019-51722-8.

Elmore, K. L., Flamig, Z. L., Lakshmanan, V., Kaney, B. T., Farmer, V., Reeves, H. D., & Rothfusz, L. P. (2014). MPING: Crowd-Sourcing Weather Reports for Research, Bulletin of American Meteorological Society, 95, 1335–1342, https://doi.org/10.1175/BAMS-D-13-00014.1.

Fiorillo, E., Crisci, A., Issa, H., Maracchi, G., Morabito, M., & Tarchiani, V. (2018). Recent Changes of Floods and Related Impacts in Niger Based on the ANADIA Niger Flood Database, Climate, 6, 59, https://doi.org/10.3390/cli6030059.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R., (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, 202, 18–27, https://doi.org/10.1016/j.rse.2017.06.031.

Gourley, J. J., Hong, Y., Flamig, Z. L., Arthur, A., Clark, R., Calianno, M., Ruin, I., Ortel, T., Wieczorek, M. E., Kirstetter, P. E., Clark, E., & Krajewski, W. F. (2013). A unified Flash flood database across the United States, Bulletin of American Meteorological Society, 94, 799–805, https://doi.org/10.1175/BAMS-D-12-00198.1.

Hall, J., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Burlando, P., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Gaál, L., Gorbachova, L., Gül, A., Hannaford, J., Kiss, A., Kjeldsen, T., Kohnová, S., Koskela, J. J., Macdonald, N., Mavrova-Guirguinova, M., Ledvinka, O., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Osuch, M., Parajka, J., Perdigão, R. A. P., Radevski, I., Renard, B., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., & Blöschl, G. (2015). A European Flood Database: facilitating comprehensive flood research beyond administrative boundaries, Proceedings of IAHS, 370, 89–95, https://doi.org/10.5194/piahs-370-89-2015.

He, B., Huang, X., Ma, M., Chang, Q., Tu, Y., Li, Q., Zhang, K., & Hong, Y. (2018). Analysis of flash flood disaster characteristics in China from 2011 to 2015, Natural Hazards, 90, 407–420, https://doi.org/10.1007/s11069-017-3052-7.

Hilker, N., Badoux, A., & Hegg, C. (2009). The Swiss flood and landslide damage database 1972–2007, Natural Hazards and Earth System Science, 9, 913–925, https://doi.org/10.5194/nhess-9-913-2009.

IPCC: Climate Change 2014: Synthesis Report, Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Core Writing Team, Pachauri, R. K., & Meyer, L. A., IPCC, Geneva, Switzerland, 151 pp., 2014.

Li, Z.: United States Flood Database (Version v1.0), Zenodo, https://doi.org/10.5281/zenodo.4547036, 2020.

Li, Z.: chrimerss/USFD (Version 1.0) [code], Zenodo, https://doi.org/10.5281/zenodo.5090020, 2021.

Li, Z., Chen, M., Gao, S., Hong, Z., Tang, G., Wen, Y., Gourley, J. J., & Hong, Y. (2020). Cross-Examination of Similarity, Difference and Deficiency of Gauge, Radar and Satellite Precipitation Measuring Uncertainties for Extreme Events Using Conventional Metrics and Multiplicative Triple Collocation, Remote Sensing, 12, 1258, https://doi.org/10.3390/rs12081258.

Mallakpour, I. & Villarini, G. (2015). The changing nature of flooding across the central United States, Nature Climate Change, 5, 250–254, https://doi.org/10.1038/nclimate2516.

Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R., & White, A. B. (2006). Flooding on California's Russian River: role of atmospheric rivers, Geophysical Research Letters, 33, L13801, https://doi.org/10.1029/2006GL026689.

Petrucci, O., Papagiannaki, K., Aceto, L., Boissier, L., Kotroni, V., Grimalt, M., Llasat, M. C., Llasat-Botija, M., Rossello, J., Pasqua, A. A., & Vinet, F. (2019). MEFF: The database of MEditerranean Flood Fatalities (1980 to 2015), Journal of Flood Risk Management, 12, e12461, https://doi.org/10.1111/jfr3.12461.

Schroeder, A., Gourley, J., Hardy, J., Henderson, J., Parhi, P., Rahmani, V., Reed, K., Schumacher, R., Smith, B., & Taraldsen, M. (2016). The development of a flash flood severity index, Journal of Hydrology, 541, 523–532, https://doi.org/10.1016/j.jhydrol.2016.04.005.

Shen, X., Mei, Y., & Anagnostou, E. N. (2017). A comprehensive database of flood events in the contiguous United States from 2002 to 2013, Bulletin of American Meteorological Society, 98, 1493–1502, https://doi.org/10.1175/BAMS-D-16-0125.1.

Shen, X., Anagnostou, E. N., Allen, G. H., Brakenridge, R., & Kettner, A. J. (2019). Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar, Remote Sens. Environ., 221, 302–315, https://doi.org/10.1016/j.rse.2018.11.008.

Shi, P. J. (2003). Atlas of natural disaster system of China, Science Press, Beijing.

Smith, A. E. R., Bates, P. D., & Hayes, C. (2012). Evaluation of a coastal flood inundation model using hard and soft data, Environmental Modelling and Software, 30, 35–46, https://doi.org/10.1016/j.envsoft.2011.11.008.

Smith, J. A., Villarini, G., & Baeck, M. L. (2011). Mixture Distributions and the Hydroclimatology of Extreme Rainfall and Flooding in the Eastern United States, Journal of Hydrometeorology, 12, 294–309, https://doi.org/10.1175/2010JHM1242.1.

Storm Events Database: https://www.ncdc.noaa.gov/stormevents/, last access: 30 October 2020.

Swiss Re sigma (2009). Natural catastrophes and man-made disasters in 2009, Swiss Reinsurance Company Ltd, Zurich, Switzerland.

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability, Scientific Reports, 10, 13768, https://doi.org/10.1038/s41598-020-70816-2.

Villarini, G. (2016). On the seasonality of flooding across the continental United States, Advances in Water Resources, 87, 80–91, https://doi.org/10.1016/j.adv.watres.2015.11.009.

Villarini, G., Goska, R., Smith, J. A., & Vecchi, G. A. (2014). North Atlantic Tropical Cyclones and U.S. Flooding, Bulletin of American Meteorological Society, 95, 1381–1388, https://doi.org/10.1175/BAMS-D-13-00060.1.

Wan, Z., Hong, Y., Khan, S., Gourley, J., Flamig, Z., Kirschbaum, D., & Tang, G. (2014). A cloud-based global flood disaster community cyber-infrastructure: Development and demonstration, Environmental Modelling and Software, 58, 86–94, https://doi.org/10.1016/j.envsoft.2014.04.007.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A high-resolution global hydrography map based on latest topography dataset, Water Resources Research, 55, 5053–5073, https://doi.org/10.1029/2019WR024873.

Chapter 4

CREST -iMAP V1.0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction

Publication

Li, Z., Chen, M., Gao, S., Luo, X., Gourley, J. J., Kirstetter, P., Yang, T., Kolar, R., McGovern, A., Wen, Y., Hong, Y. (2021). CREST-iMAP v1. 0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction. Environmental Modelling & Software, 141, 105051.

4 Chapter 4. CREST -iMAP V1.0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction

Abstract

Under the impacts of hydrologic extremes, there is a growing need for integrated frameworks for flood inundation mapping. In this study, we introduce a framework: the Coupled Routing and Excess Storage for inundation MApping and Prediction (CREST-iMAP). It utilizes a highly parallelized and fully integrated hydrologic-hydraulic model and aims to improve flood prediction. To highlight the advantages, a synthetic rainfall event and the 500-year Hurricane Harvey event were investigated using the CREST-iMAP, compared with a hydrologic model, a hydraulic model, and a simplified hydrologic-hydraulic model. The results indicate the CREST-iMAP achieves good performances in both hydrologic simulation and floodplain inundation mapping. Moreover, the antecedent soil moisture is the most sensitive parameter to model accuracy, followed by the land surface characteristics; the infiltration process to a flood forecasting model is significant. Overall, the CREST-iMAP delivers accurate and timely flood information, making it one of the latest developments of an integrated hydrologic-hydraulic expert system.

4.1 Introduction

A hydrologic model normally comprises a water balance module and a river routing module. The water balance module transforms rainfall into surface runoff via the hydrologic processes in the vertical direction. Laterally, these runoffs are routed to an outlet point by the routing module. However, the routing scheme (e.g., unit hydrograph methods, linear reservoir, and 1D kinematic wave model) is often oversimplified for most hydrologic models, which is perceived as a weakness for flood forecasts (Felder et al., 2017; Kim et al., 2012; Nguyen et al., 2016; Pontes et al., 2017; Wood et al., 2011). Regarding the pure hydraulic model, the ignored or crudely represented infiltration process likely leads to overestimations in flood simulation (Ni et al., 2020). Therefore, several studies have hitherto attempted to couple a hydrologic model with a hydraulic (hydrodynamic) model, herein denoted as the H&H model (Felder et al., 2017; Montanari et al., 2009; Phu et al., 2016; Pontes et al., 2017; Sampson et al., 2015; Wu et al., 2014; Zischg et al., 2018). The coupled models enable streamflow generation, flood extent delineation, and peak flood depth estimation all at once and have been applied at local, regional, continental, and global scales. The real-time Global Flood Monitoring System (GFMS), jointly developed by the National Aeronautics and Space Administration (NASA) and the University of Maryland, sets an example that couples the Variable Infiltration Capacity (VIC) model and Dominant River Tracing-Based Runoff-Routing (DRTR) model at 12-km spatial resolution (Wu et al. 2014). Nguyen et al. (2016) coupled the Hydrology Laboratory Research Distributed Hydrologic Model (HL-RDHM) with the BreZo model and showcased results in the Baron Fork, Oklahoma. Felder et al. (2017) integrated a semi-distributed hydrologic model PREVAH and a hydraulic model BASEMENT to examine the joint advantages in peak flow simulation specifically.

Despite a compelling need for developing H&H models, a few challenges still remain for a coupled system. First, different means of integrating the two models are critical for operational flood monitoring. A majority of the previous studies loosely couple the two (i.e., offline couple), in which hydrologic simulation in the upstream basin is a prerequisite to provide boundary conditions before the hydraulic model runs (Montanari et al., 2009; Nguyen et al., 2016; Felder et al., 2017). In this regard, the offline coupling compromises its timely data delivery for emergency responders. Additionally, the offline coupling entails calibrating two models separately, making it time-consuming, although the hydraulic model typically has fewer parameters to tune. Second, the

design of computational meshes that are scalable and suitable for all applications is challenging. Structural meshes (quadrilateral cells) at fixed resolution are so far a common approach in hydraulic models (e.g., TUFLOW Classic 2D, Flowroute-I, DIVAST, SOBEK Suite, CaMa-Flood, and LISFLOOD-FP) because they resemble the grids as the topographic data and thus exhibit less structural errors (Kim et al., 2014). The structured mesh, however, could lead to numerous computational elements, limiting the utility of the coupled modelling system for large-scale implementation (Saksena et al., 2020; Teng et al., 2017). The recent development of unstructured meshes (triangular meshes or blended triangular and quadrilateral meshes such as ANUGA, DELFT-3D, and MIKE FM) enhances the flexibility of the model, to have fine resolution in complex watersheds to reflect subtle changes in topography or coarse resolution in areas of less interest (or less heterogeneity) to save computational time (Ivanov et al., 2004; Saksena et al., 2020). The flexibility in this regard helps to make the model scalable and efficient while not sacrificing too much accuracy. Third, there is a necessity for a fully integrated system that not only couples the H&H model, but also includes extended features such as uncertainty quantification, mesh generation, sensitivity analysis, and calibration.

In the presence of these challenges and limitations, we developed a new end-to-end H&H modeling framework: the Coupled Routing and Excess Storage inundation MApping and Prediction (CREST-iMAP). Such a framework contains a flexible mesh generator that efficiently reduces structural errors, a H&H model that stems from the widely used the CREST model (CREST; Wang et al., 2011) for the water-balance part and the ANUGA-Hydro for the hydraulic part (Nielsen et al., 2005), a suite of sensitivity and uncertainty analysis tools, and global calibration schemes. The physical representation of hydrologic processes and routing in conjunction with flexible and efficient mesh grids makes this framework one of the latest developments of an H&H expert system. In this paper, the main purpose is to release and introduce the first version of CREST-iMAP that is suited for operational flood forecasts. For the analysis, we examine the similarity and difference of CREST-iMAP with a pure hydraulic model, a hydrologic model, and a hydraulic model combined with a simple infiltration scheme. This article is structured as follows. The second section details each component within this framework, along with a methodology describing the designed experiment. The third section presents the results from the synthetic experiment and showcases a real flood simulation against observations from in-situ instruments and remote sensing products for the 500-year Hurricane Harvey event.

4.2 Methods

4.2.1 CREST-iMAP framework

Fig. 4.1 shows an overview of the designed CREST-iMAP framework. The model engine consists of two components: the CREST hydrologic model that vertically transforms rainfall into surface runoff and a 2D hydraulic model that routes surface runoff spatially driven by topographic inputs. For the model setup, it is required to first design the mesh and prepare a-priori distributed parameters. Detailed information regarding mesh and parameter generation are provided in Section 2.3 Hydraulic components, 2.4 Flexible mesh generator, respectively. For model calibrations, we provided two approaches, one conventional optimization approach that minimizes the desired objective function through a set of advanced algorithms; and the other one, termed surrogate modeling, to mitigate extensive or infeasible calibration efforts by learning the response surface. Furthermore, we also include a sensitivity analysis that can investigate how the parameters interact and influence the results.

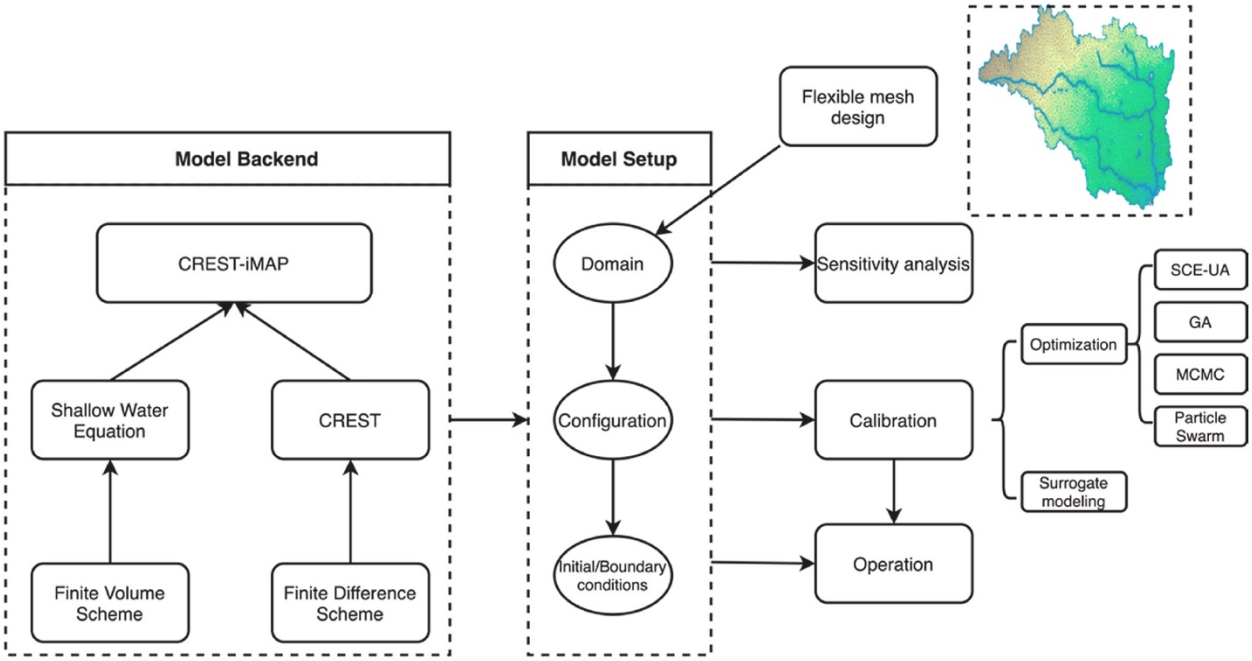


Figure 4.1 The proposed modeling framework for CREST-iMAP v1.0.

4.2.2 Hydrologic processes

The natural hydrologic process is depicted in Fig. 4.2a, in which both natural landscapes and urban areas are taken into account for flood monitoring. It is often recognized that a hydrologic model has the advantage of mimicking natural hydrologic processes while being limited in

representing water routing in urban regions due to complicated hydraulic structures and controls. On the contrary, a hydraulic model could be applied for urban flood simulation while ignoring the hydrologic processes of infiltration, canopy interception, etc. By coupling these two models, as in CREST-iMAP, we preserve the advantage of modeling hydrologic process and extends its capacity for flood prediction and inundation in urban areas. It is noteworthy that traditional hydrologic models often omit local ponds and only simulate the streamflow in the natural waterway. In this coupled system, it is possible to also account for ponding that could impact the timing and magnitude of localized flooding.

The detailed hydrologic processes represented in CREST-iMAP is shown in Fig. 4.2b, which originates from CREST V2.1 (Flamig et al., 2020; Xue et al., 2013). The CREST model is a distributed hydrologic model that is jointly developed by the University of Oklahoma (https://hydro.ou.edu) and the NASA SERVIR project team (Wang et al., 2011). It has been implemented as a flood detection toolkit across the globe (Wu et al., 2012) and for operational flash flood forecasting in the United States and territories (Gourley et al., 2017) in near-real-time. The atmospheric forcing data, acquired from the gridded products, are firstly partitioned to surface runoff and subsurface flow according to the surface characteristics and prior model states. The subsurface flow is further separated via a variable infiltration curve (Zhao, 1995). Thereafter, multiple linear reservoirs are used to represent sub-grid routing for overland flow and interflow. The detailed relationships and expressions of theses interactions can are discussed in Xue et al. (2013). In flash flood forecasting, model efficiency can be maintained by routing surface flow through a hydraulic model while ignoring the less impactful fluxes such as the interflow term. As an intermediate product, users can export the fluxes such as surface flow and subsurface flow, and the soil moisture state as shown in Fig. 4.2c.

The proposed CREST-iMAP framework has seven distributed or lumped parameters as described in Table 4.1. Amongst them, the saturated hydraulic conductivity (Ksat), mean soil water capacity (WM), exponent of the variable infiltration curve (B), impervious area ratio (IM), and the ratio of the PET to actual evapotranspiration (KE) can be estimated as a-priori parameter set using the land cover types and soil texture data based on a look-up table (Chow et al., 1988). The initial soil moisture (SM0) is an antecedent model state, which can be approximated from an external data source such as Soil Moisture Active Passive (SMAP) or interpolated soil moisture

values from ground observations or via a long warm-up period. It is the only lumped parameter that scales the WM from 0 to 1.

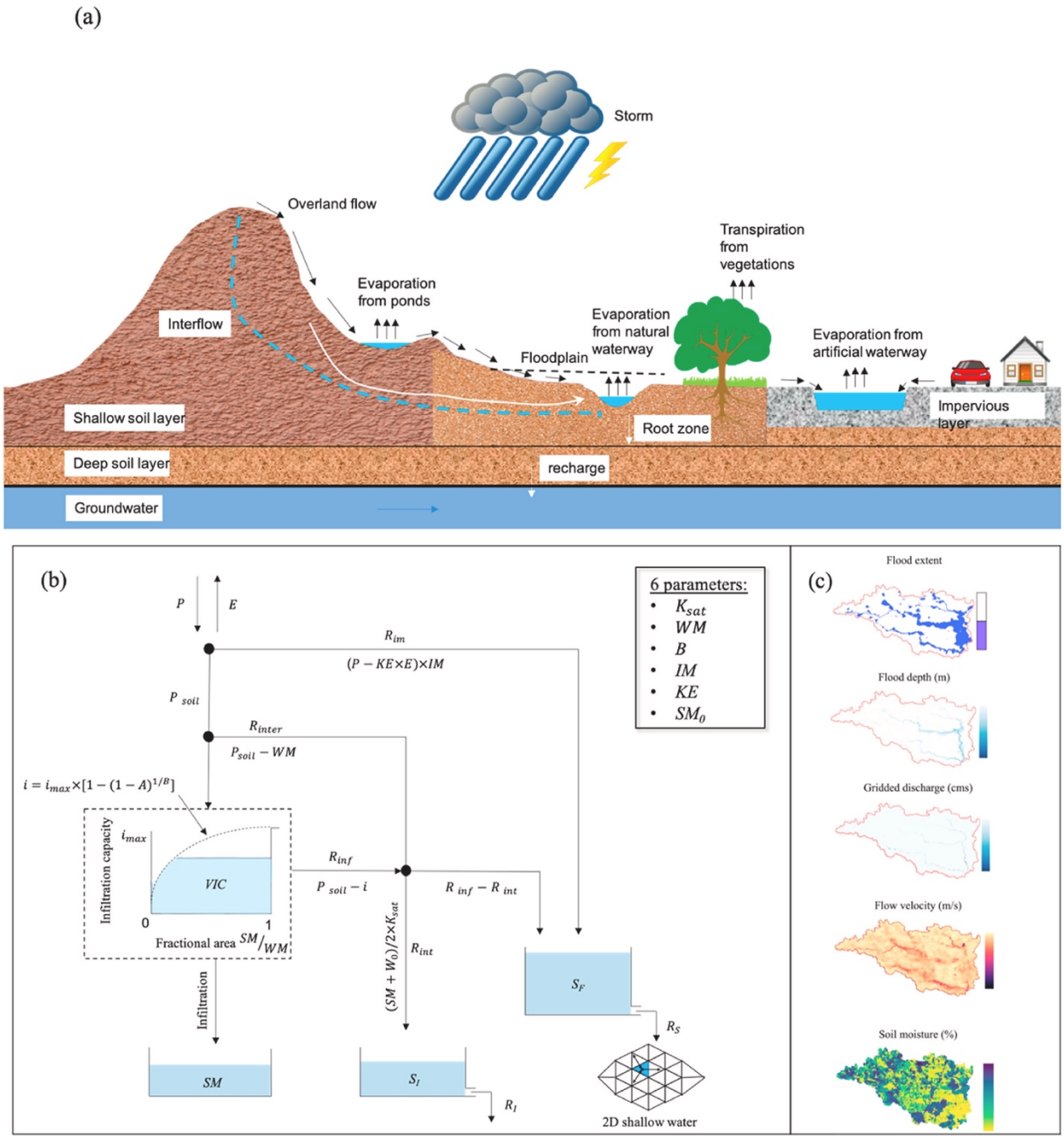


Figure 4.2 Schematic view of (a) natural hydrologic processes, (b) how CREST-iMAP simulates them, and (c) the output variables.

Table 4.1 Parameters required in CREST-iMAP framework.

Parameters	description	Range

Ksat	Soil saturated hydraulic conductivity (mm/d)	0–2827.2
WM	Mean soil water capacity (mm)	80–200
B	The exponent of the variable infiltration curve	0.05–1.5
IM	Impervious area ratio (%)	0–100
KE	The ratio of the PET to actual evapotranspiration	0.1–1.5
SM_0	Initial soil moisture	0–1
Manning's n	The coefficient for the use of manning's equation in channel flow	0–1

4.2.3 Hydraulic components

The hydraulic model is derived from ANUGA-Hydro V2.1 (Nielsen et al., 2005), which is an open-source software that is suitable for predicting the consequences of hydrological disasters such as riverine flooding, storm surges, and tsunamis. This model has been applied to a wide range of 2D flood simulations (Griffin et al., 2015; Guerra et al., 2014; Issermann and Chang, 2020). It solves the 2D shallow water equation (depth-averaged Navier-Stokes equation) that is derived from mass balance and momentum balance equations by ignoring the vertical acceleration term. The 2D Finite Volume Method (FVM) is used as the solver to the mathematical representations. In the CREST-iMAP framework, the surface runoff output from the hydrologic model is set as the input for this hydraulic model. For the end results, users can choose to export the flood extent, gridded flood depth, flow velocity, and volumetric flow rate. Notably, the gridded volumetric flow rate output requires a set of cross-sections as input that can be derived from the System for Automated Geoscientific Analyses (SAGA) package – Cross Profiles (Conrad et al., 2015).

4.2.4 Flexible mesh generator

There is a tradeoff of structuring the mesh that could minimize the structural error from topographic inputs and also retaining the computational efficiency. Traditionally, the unstructured meshes are designed according to the user-defined maximum and minimum areas, minimum angle of the triangular area as illustrated in Fig. 4.3a. Some advanced generators can refine the areas of interest in the study domain with nested cells, such as the one shown in Fig. 4.3b. However, the landscape heterogeneity is mostly ignored but could substantially increase the computational burden if the mesh is not optimized in conjunction with the mesh quality (Marsh et al., 2018). Experiments have demonstrated that the computational effort is significantly reduced by 90% by simulating complex watersheds using an adaptive mesh to represent terrain (Marsh et al., 2018; Ivanov et al., 2004). Therefore, a quality-guaranteed and flexible mesh design is essential to meet these challenges. Here we identify the flexible mesh by being able to satisfy specific user

demands, which can accomplish a) dense meshes in the regions that have large heterogeneity of input features (e.g., elevation, soil property, and flow accumulation), (b) coarse meshes in the regions that are of less interest or homogeneous, and (c) dividing the meshes based on user specified constraints (e.g., river networks). In this framework, we adopted the novel mesh generator scheme proposed by Marsh et al. (2018) and integrated it into our modeling framework.

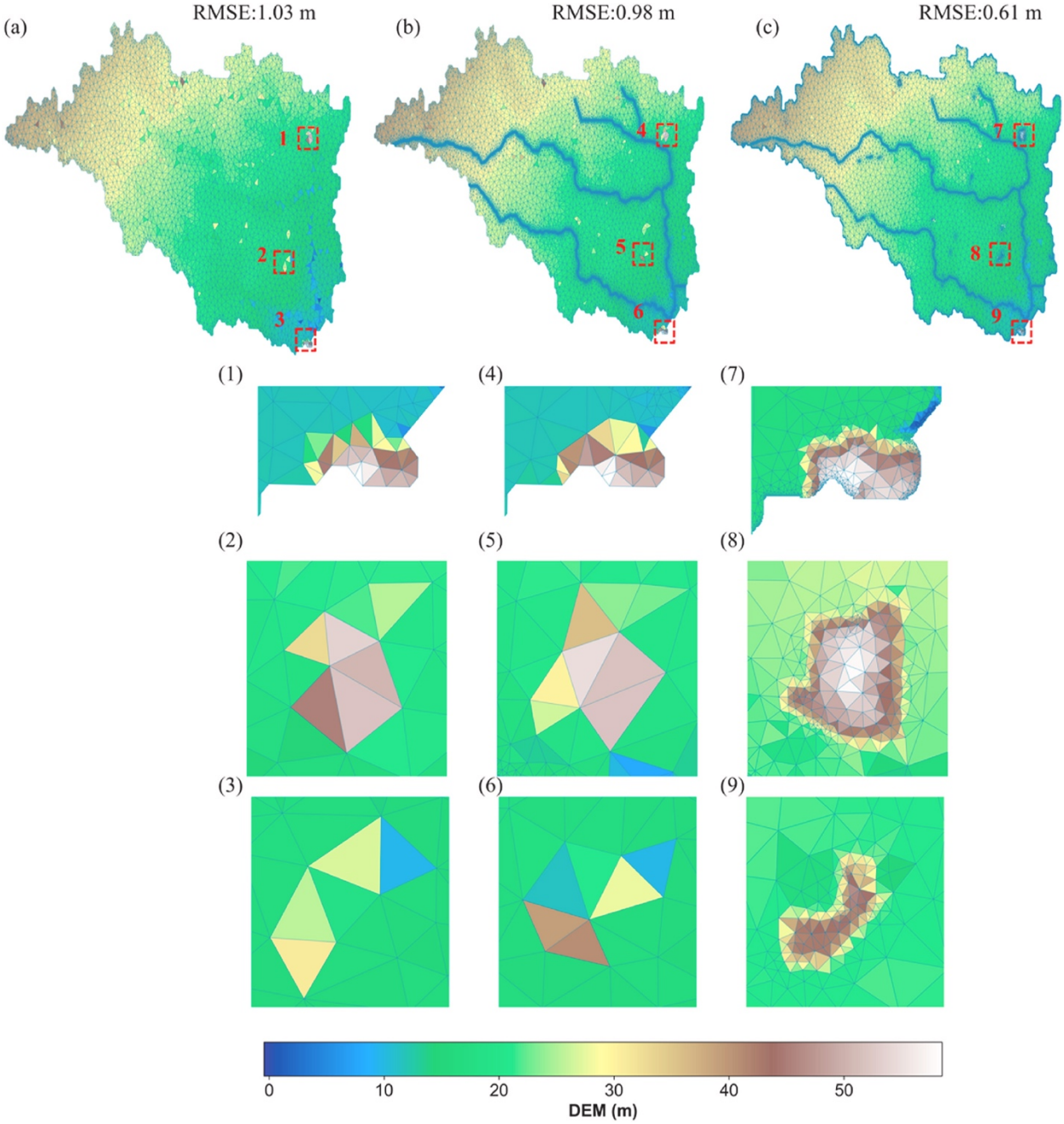


Figure 4.3 Examples of generated meshes: (a) traditional uniform mesh based on area of interest; (b) refined river networks in addition to uniform mesh; (c) flexible mesh refined by topographic

variation. (1)–(9) illustrate the samples encompassed by dashed rectangular regions (in red) in each mesh design.

Fig. 4.3 displays the (a) traditional uniformly spaced meshes, (b) refined meshes for the inner river network, and (c) flexible meshes that are based on the variation of topographic inputs. Intuitively, the traditional way cannot reveal the local abrupt changes of terrain (e.g., concave and convex) while the flexible design enables more meshes to reflect local variations and less so in homogeneous surfaces. The Root Mean Square Error (RMSE) quantitatively demonstrates that the flexible mesh could best reconstruct the structural input (DEM) among the three with the lowest RMSE (0.61 m). The refined river network improves slightly compared to the uniform mesh. In response to flow simulations, it is widely reported in the literature that the adaptive meshes could significantly reduce the computational demand and preserve the localized flow characteristics (Ivanov et al., 2004; Marsh et al., 2018; Saksena et al., 2020; Teng et al., 2017).

4.2.5 Sensitivity analysis and calibration

Parameters are inherent and uncertain in most environmental models (Beven and Binley, 1992; Beven and Freer, 2001; Duan et al., 2006; Gupta et al., 1998; Teng et al., 2017). Parameter sensitivity and uncertainty have profound implications for hydrologic models. Sensitivity analysis is advantageous in environmental models, including uncertainty assessment, model calibration, and evaluation. An exhaustive review on the sensitivity analysis that is specific to environmental modeling can be found in Pianosi et al. (2016). The intentions here of incorporating sensitivity analysis in our framework are two-fold. First, the sensitivity result is a pre-examination against our prior hypotheses of the basin characteristics. For instance, soil property-related parameters are theoretically less sensitive than friction terms in a highly urbanized area. Second, only the sensitive parameters will be calibrated in the automatic calibration scheme for the sake of efficiency (Gan et al., 2018). We herein include two of the most widely applied algorithms: One-At-a-Time (OAT) and All-At-a-Time (AAT). The OAT simply measures the partial derivative of a parameter at the nominal value, and it is referred to as the local sensitivity method, while the AAT measures global sensitivity by aggregating individual sensitivities. The specific AAT method used in this framework is the well-established Morris method (Morris, 1991).

To quantify the model parameters efficiently necessitates a global search calibration algorithm. The well-known Shuffle Complex Evolution (SCE-UA) by Duan et al. (1992) is

implemented in this framework. Additionally, a set of evolutionary algorithms such as the Genetic Algorithm (GA), Markov Chain Monte Carlo (MCMC), and Particle Swarm are also included for different user demands. The calibration is conducted by comparing simulated river stage and discharge with stream gauges. But in the future, we believe the calibration can be extended with promising remotely sensed data such as flood extent/stage derived from Synthetic Aperture Radar (SAR; Montanari et al., 2009) and crowdsourcing data (Chen et al., 2016).

4.2.6 Study region and data acquisition

In the following experiment, we apply our modeling framework to the Greens Bayou watershed in the Houston metropolitan area, illustrated in Fig. 4.4. This area endures frequent tropical cyclone and hurricane activity, which was most evident with Hurricane Harvey in 2017 where 955 mm of rainfall impacted the region over five days. Heavy rainfall from tropical systems combined with expansive urban development, results in numerous fatalities and vast economic losses. The sub-basin in Fig. 4.4c is upstream of Buffalo Bayou in Houston. The total drainage area is 457.9 km2, and the average elevation (slope) is 23.65 m (1.46%). Due to the flat slope, the simplified 1D-routing scheme (i.e., kinematic wave) in the hydrologic model is problematic, which complicates streamflow generation (Getirana and Paiva, 2013; Vergara et al., 2016). Fig. 4.4d depicts the 15 different land surface types (from the National Land Cover Database), of which the majority (89.7%) is developed land, especially in the southwest. This areal impervious ratio is 33.8% on average. Forests are situated on the eastern part of this region, creating a natural buffer to storm rainfall. There are five United States Geological Survey (USGS) stream gauges located at three main tributaries (Garners Bayou, Greens Bayou, and Halls Bayou from north to south). Because the upstream basin of the gauge 08075900 is a highly developed urban area, the flow simulation is complicated by urban controls and stores on the hydraulic simulation. Therefore, in the following flow analysis, this gauge is excluded.

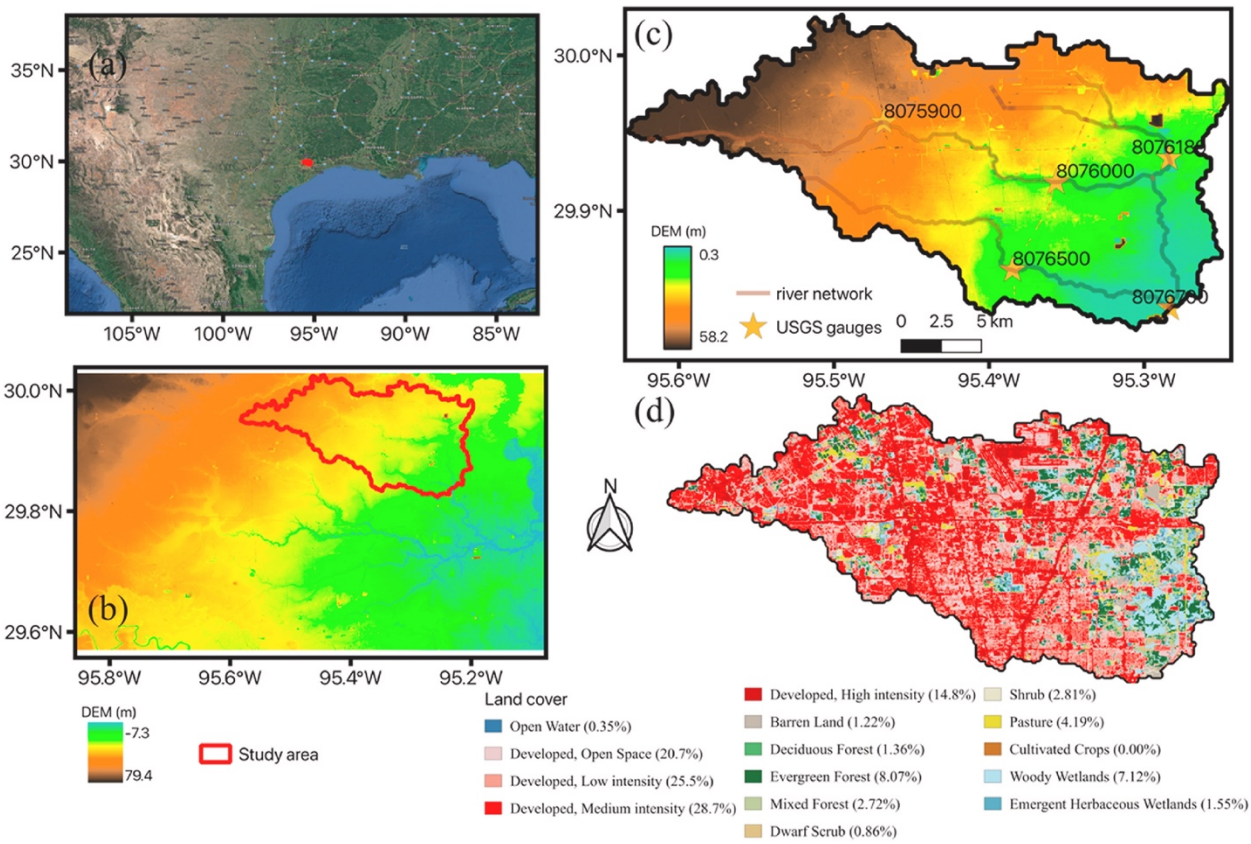


Figure 4.4 Maps of the (a) study area, (b) surrounding terrains, and (c) detailed information including river networks and USGS gauge sites (c). The land cover is shown in (d) with the respective compositions in percentage (%).

The Digital Elevation Model (DEM) was obtained from the USGS Elevation Products (3DEP) in the USGS at 1/3 arc-second (roughly 10 m) resolution. The original DEM underwent a series of pre-processing steps, including noise removal with the speckle filter, depression fill, and burn-in of river channels. For the atmospheric forcing, the Multi-Radar Multi-Sensor (MRMS) data at 2-min and 1-km resolution are used for rainfall inputs (Zhang et al., 2016); and the potential evapotranspiration rate is obtained from USGS FEWS data port (https://earlywarning.usgs.gov/fews) at daily and 1° spatial resolution (Allen et al., 1998). These forcing data are specific for the Hurricane Harvey case study from 2017-08-25 to 2017-08-31. In addition, the distributed model parameters are acquired from the CREST model configured as part of the Flooded Locations and Simulated Hydrographs (FLASH) project, where the existing parameters are provided and calibrated for delivering operational US-wide discharge (Gourley et al., 2017) (Fig. 4.5). The surveyed soil type compositions are downloaded from United States Department of Agriculture SSURGO dataset at https://websoilsurvey.nrcs.usda.gov.

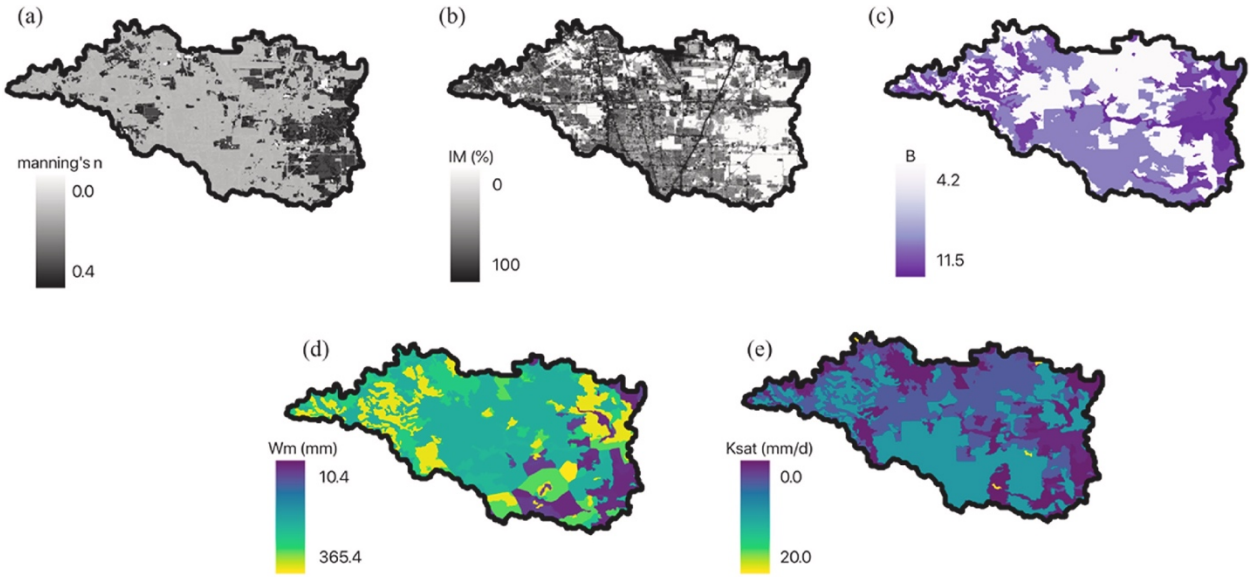


Figure 4.5 The distributed parameters in the CREST-iMAP model: (a) manning's n coefficient, (b) impervious area ratio, (c) the exponent of the variable infiltration curve, (d) mean soil water capacity, and (e) soil saturated hydraulic conductivity.

In order to validate the model-simulated flood extent with observations, two remote-sensing products are correspondingly retrieved as follows. The Radar-Produced Inundation Diary (RAPID) incorporates the Sentinel-1 SAR data to produce high-resolution (10 m) flood extent (Shen et al., 2019). The Dartmouth Flood Observatory flood extent (herein denoted as DFO) was produced by merging multiple data sources, including the Sentinel-1, Moderate Resolution Imaging Spectroradiometer (MODIS), RadarSat 2, and the Constellation of Small Satellites for Mediterranean basin Observation (COSMO SkyMed). These images are resampled to the same resolution as our simulation results. In addition to inundated areas, the USGS High-Water Marks (HWMs), collected during a post-event ground survey of water marks, are utilized in this study to comprehensively compare to the simulated peak flood depths. It is noteworthy that these observations, though used as reference, are not error-free.

4.2.7 Experiments and computational metrics

To assess the uncertainty of the model engines (i.e., CREST-iMAP, non-coupled hydraulic, and non-coupled hydrologic models) and model parameters, we implement a synthetic rainfall study as well as a real case study. Additionally, in order to reflect the advances in infiltration process simulated by the hydrologic model, the hydraulic component plus a simple infiltration model added to it are compared to one another. The infiltration model for this study is adopted

from an empirical method described by Huggins and Monke (1968) in Eq. 4.1. Therefore, for the model comparison, CREST-iMAP, pure hydraulic component (herein referred to as hydraulic), hydrologic component (herein referred to as CREST), and a hydraulic plus simple infiltration (denoted as hydraulic + infiltration) are taken into consideration.

$$f = K_{sat} + 5K_{sat}((\theta_{sat} - \theta_i)L - F)/L\theta_{sat}{}^{0.65} \quad (4.1)$$

where f represents the infiltration rate (mm h^{-1}), and F is the accumulated infiltration. Ksat is the saturated hydraulic conductivity (mm h^{-1}), approximated from the look-up table. θ_i, θ_{sat} are the initial and saturated moisture content.

The parameter sensitivity is assessed using the Morris method (Morris, 1991), which evaluates the global sensitivity AAT. The corresponding computational metrics are summarized in Table 4.2. The Nash-Sutcliff Efficiency (NSE) is a standardized indicator to evaluate the correspondence and magnitude of the simulated and observed streamflow; Relative Bias (RB) indicates the difference of the flow volume in a systematic manner; Time-to-peak Error in hours (h) and Peak-flow Error in cubic meter per second (cms) indicate the timing and magnitude difference of the peak flow. When comparing flood extent in a binary form, a contingency table is constructed to compute the percentage of True Positives (TP), False Negatives (FN), False Positives (FP), and True Negatives (TN).

Name	Expression	Range	Best value
NSE	$1 - \frac{\sum_{n=1}^{N}(O_n - S_n)^2}{\sum_{n=1}^{N}(O_n - \bar{O})^2}$	(-inf,1)	1
Relative Bias	$\frac{\sum_{n=1}^{N} S_n - \sum_{n=1}^{N} O_n}{\sum_{n=1}^{N} O_n}$	(-inf, +inf)	0
Time-to-peak Error (h)	$t(O_{max}) - t(S_{max})$	(-inf, +inf)	0
Peak-flow Error (cms)	$O_{max} - S_{max}$	(-inf, +inf)	0
True Positives (%)	$\frac{\sum_{n=1}^{N}(O_n = 1) \cap (S_n = 1)}{N} \times 100\%$	(0,100)	100
False Positives (%)	$\frac{\sum_{n=1}^{N}(O_n = 0) \cap (S_n = 1)}{N} \times 100\%$	(0,100)	0
False Negatives (%)	$\frac{\sum_{n=1}^{N}(O_n = 1) \cap (S_n = 0)}{N} \times 100\%$	(0,100)	0
True Negatives (%)	$\frac{\sum_{n=1}^{N}(O_n = 0) \cap (S_n = 0)}{N} \times 100\%$	(0,100)	100

Note: the symbol O represents the observations, and S is the estimates. N indicates the length of data.

In terms of the real case, we simulated the Hurricane Harvey event from 2017-08-25 to 2017-09-01. Since model parameters have been well-calibrated previously, we simply split this simulation period to a fine-tuning period (2017-08-25 to 2017-08-26) and a verification period

(2017-08-27 to 2017-09-01). The purpose of model fine-tuning is to refine the soil moisture states and water levels in the streams. Some sensitive parameters discovered in the sensitivity analysis are tuned to better represent the basin attributes. We applied a scaling factor in the calibration stage to perturb the distributed parameters, similar to the approach used in the CREST model (Flamig et al., 2020; Xue et al., 2013). The objective function for the fine-tuning period is based on the NSE.

4.3 Results

4.3.1 Synthetic experiment

In this section, we examine the performance of the four model engines for a synthetic rainfall event according to the following three aspects. First, the sensitivity of the model parameters of CREST-iMAP is investigated, which is primarily linked to basin characteristics. Second, the hydrologic comparison provides insights on the difference of the hydrologic infiltration process and soil moisture dynamics. Lastly, the flood simulation cross-compares the flood extent and flood depth generated by different model engines. In this experiment, the synthetic rainfall is uniformly distributed in this study area at a rate of 100 mm h−1 for a 2-h duration. Twenty-four-hour simulations from the four model configurations are thereafter generated. The sensitivity is conducted via the Morris method, and statistical metrics are computed at the basin outlet by perturbing seven model parameters for CREST-iMAP. The standard deviation and means of the gradient are calculated to reflect the change of parameter value per perturbation.

The simulation time of each model engine is as follows. CREST-iMAP costs 1.42 h in total using 36 CPU cores implemented on Intel Xeon @ 3.00 GHz. Similarly, hydraulic and hydraulic + infiltration require 1.58 h and 1.42 h, respectively. Notably, this total simulation time is an accumulation of time run per step, which is explicitly derived to satisfy the known Courant-Fredrichs-Levy (CFL) condition in eq. 4.2.

$$\Delta t \leq \frac{u}{C\Delta x} \tag{4.2}$$

where u is the magnitude of flow velocity (m s−1), and Δx is the radius of the inscribed circle of a cell (m). C is the Courant number that is typically fixed to 1. Therefore, the time per step is determined by the flow velocity and the shape of the unstructured grid. The pure hydraulic model, which solely transforms all rainfall to overland flow, is anticipated to have a higher flow rate and thus more time spent per step compared to CREST-iMAP and the scenario of

hydraulic + infiltration configuration. In contrast, the CREST costs 1.72 h on a single CPU core since the routing scheme is not fully parallelized.

4.3.1.1 Model parameter sensitivity

The sensitivity of model parameters in this study is investigated to not only examine the basin characteristics, but to identify a set of parameters to calibrate the model. Fig. 4.6 depicts the global sensitivity of each parameter from the means and standard deviations of the gradient, benchmarked using the a-priori parameter set. By default, the initial soil moisture (*SM0*) is set to 0. Across all the metrics, *SM0* stands out to be the foremost sensitive parameter to the model results. It is anticipated that antecedent soil moisture is a determining factor to flooding, and previous studies also exemplified this point using a range of urbanized basins (Smith et al., 2002; Yang et al., 2011). In addition, the manning's coefficient *n* is the second most sensitive parameter that governs overland flow. For the hydraulic model, it is the only parameter to tune. The parameter *IM*, as the third most sensitive parameter, has important implications for timing of the peak flows. Broadly speaking, a higher fraction of impervious ratio (urbanization) leads to flashier hydrographs. Other parameters related to soil properties and evaporation are less significant in this study, as the majority of the land cover is well-developed. Therefore, out of the seven parameters in CREST-iMAP, only the most sensitive ones (i.e., *SM0*, *IM*, *n*) are selected to fine-tune the model for the real case using the SCE-UA algorithm. Meantime, all model configurations are optimized with the same technique in the fine-tuning period.

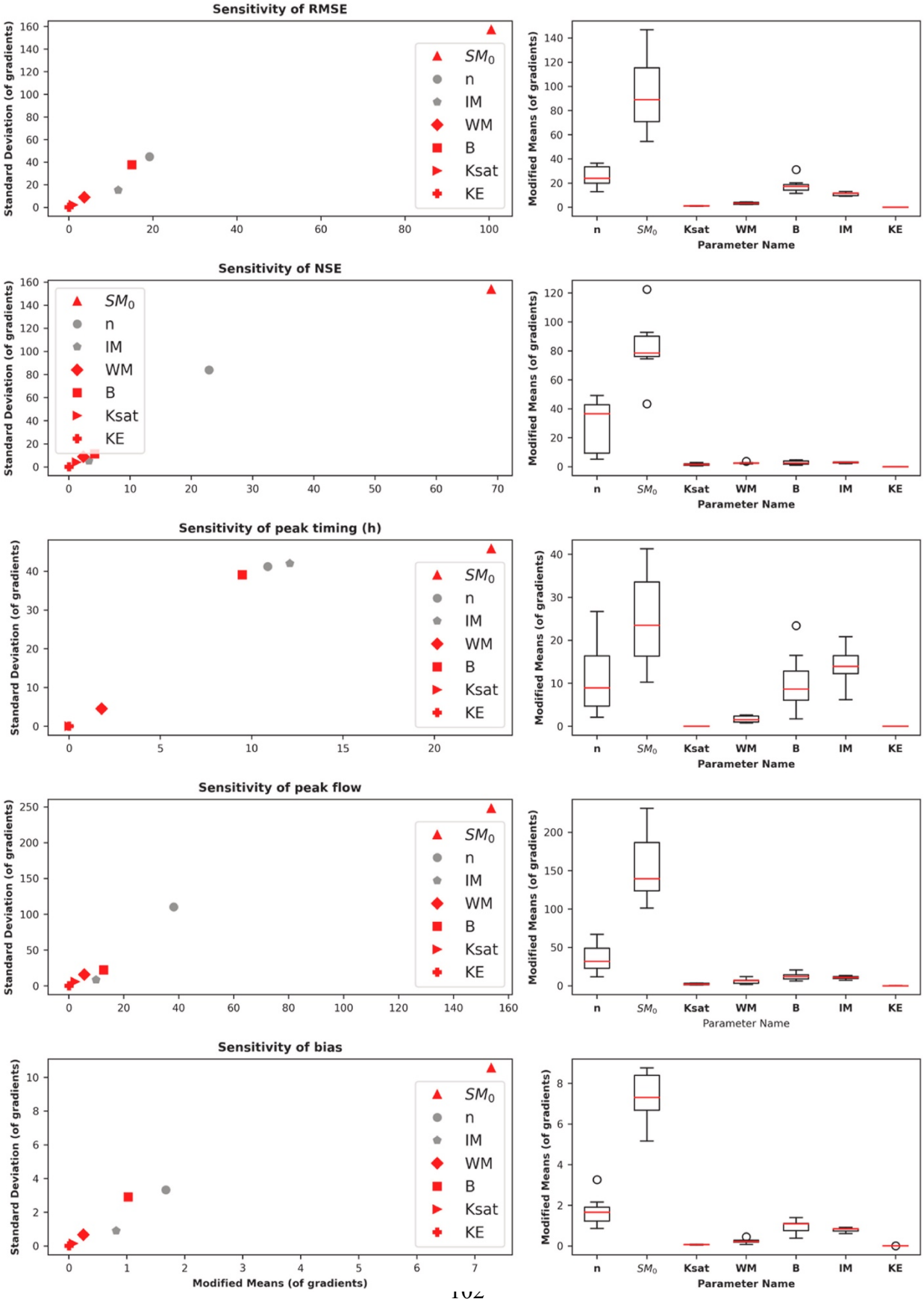

Sensitivity of RMSE
Standard Deviation (of gradients)
Modified Means (of gradients)
Parameter Name
Sensitivity of NSE
Sensitivity of peak timing (h)
Sensitivity of peak flow
Sensitivity of bias
Modified Means (of gradients)
SM_0
n
IM
WM
B
Ksat
KE

Figure 4.6 Model sensitivity using the Morris method with respect to the metrics summarized in Table 4.2.

4.3.1.2 Infiltration and soil moisture

The distributed intermediate variables have expanded the data sources for model evaluation, especially when considering the availability of remotely-sensed data in addition to more conventional use of basin-integrated streamflow. Furthermore, these intermediate variables can be used in the process of data assimilation to improve the model prediction skill. In this section, the intermediate soil moisture, infiltration rate, and overland flow rate at a specific timeframe are compared to one another.

The overland flow rate and soil moisture at 1 h after the synthetic rainfall has been introduced are shown in Fig. 4.7. It is visually obvious that CREST-iMAP shares a spatial similarity with CREST in contrast to the other two configurations in terms of the overland flow rate and the soil moisture content. Moreover, the basin-averaged time series of infiltration rate of the two configurations exponentially decay with time. The soil moisture change in time is also comparable, peaking at the same time (~2 h). However, the difference of percent of saturated soil cells with time is discernible in Fig. 7h. The CREST model contains a higher fraction of saturated grid cells than CREST-iMAP after 12 h. This difference can be attributed to the structural difference caused by different representation of mesh grids, as CREST-iMAP uses the triangular meshes that is at a pseudo-10-m resolution while CREST uses regular 10-m grid cells. In addition, due to different settings of interflow, the soil water depletion could essentially differ. Under the scenario of hydraulic only (Fig. 4.7a), the model treats every drop of rainfall as overland flow to route but omits the soil infiltration capacity, and therefore the overflow rate saturates at 100 mm h^{-1} homogeneously across the region. Under the scenario of hydraulic with simple infiltration (Fig. 4.7b), the soil saturates across the majority of the land surface because water-soil interaction is not properly represented due to the simplification (i.e., infiltration rate is not a function of soil moisture state). But at least, this simple infiltration captures saturation in the developed land surfaces, leaving the areas with high soil water capacity as unsaturated. From the basin-average infiltration rate shown in Fig. 4.7g, the simple infiltration performs similarly with CREST-iMAP and CREST, except for a higher infiltration rate at the beginning and earlier saturation. In summary, both CREST-iMAP and CREST exhibit more hydrologically reasonable

spatial variation of the soil moisture and infiltration process, as opposed to the simulation results under the hydraulic only and hydraulic with infiltration scenarios.

4.3.1.3 Flood simulation

The end products, including river stage, discharge, flood extent, and flood depth, are inter-compared. CREST uses a 1D channel routing scheme and does not produce flood inundation maps, so it is excluded in this analysis. Figures 4.8a and b depict the basin-integrated river stage and discharge at the outlet point, respectively. CREST-iMAP, with advanced infiltration process, shares similar performance with the hydraulic + infiltration. However, the hydraulic + infiltration configuration generates slightly greater flow volume in the rising limb and earlier peak time because its soil state reaches saturation earlier than CREST-iMAP. The hydraulic simulation, due to zero infiltration, generates a higher river stage and volumetric flow rate compared to the other two. Regarding the flood extent comparisons, CREST-iMAP is considered the benchmark (or observation). Despite the hydraulic-only model capturing higher percent of TP (21.7%) than the hydraulic + infiltration (19.9%), it raises markedly more FP (13.4% versus 1.0%) because of the systematically higher overland flow. The same reason leads to the regionally positive difference of the maximum water depth between the hydraulic and CREST-iMAP in Fig. 4.8e; however, most of the grid cells in Fig. 4.8f show negative differences between the hydraulic + infiltration and CREST-iMAP.

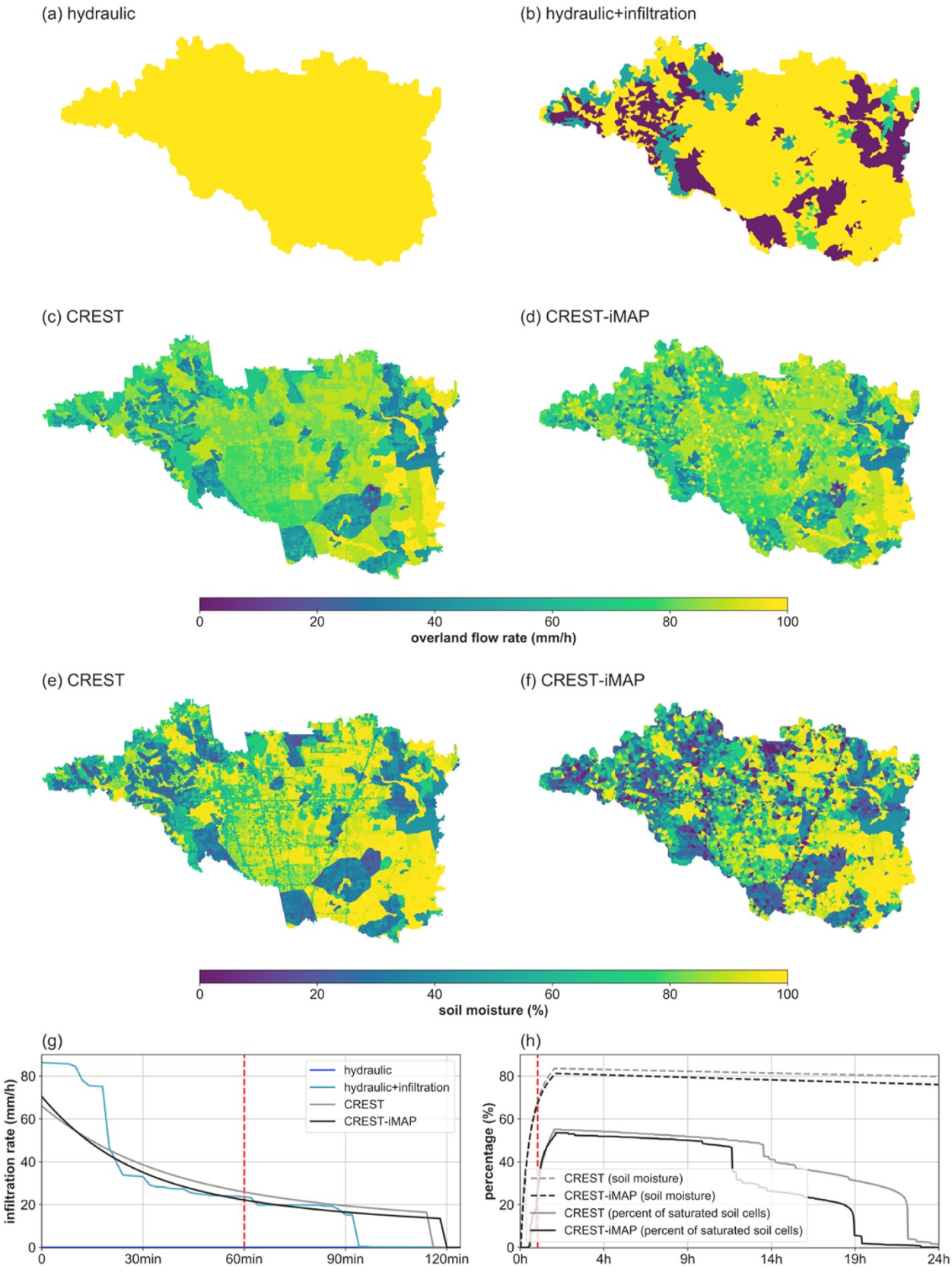

(a) hydraulic
(b) hydraulic+infiltration
(c) CREST
(d) CREST-iMAP
0
20
40
60
80
100
overland flow rate (mm/h)
(e) CREST
(f) CREST-iMAP
0
20
40
60
80
100
soil moisture (%)
(g)
infiltration rate (mm/h)
hydraulic
hydraulic+infiltration
CREST
CREST-iMAP
0
30min
60min
90min
120min
(h)
percentage (%)
CREST (soil moisture)
CREST-iMAP (soil moisture)
CREST (percent of saturated soil cells)
CREST-iMAP (percent of saturated soil cells)
0h
4h
9h
14h
19h
24h

Figure 4.7 Maps of the overland flow rate for (a) hydraulic, (b) hydraulic + infiltration, (c) CREST, and (d) CREST-iMAP, and maps of soil moisture (%) for (e) CREST-iMAP and (f) CREST. Time series of (g) the basin averaged infiltration rate and (h) basin averaged soil moisture, where the dashed red line in the panels corresponds to the selected timeframe for the above maps.

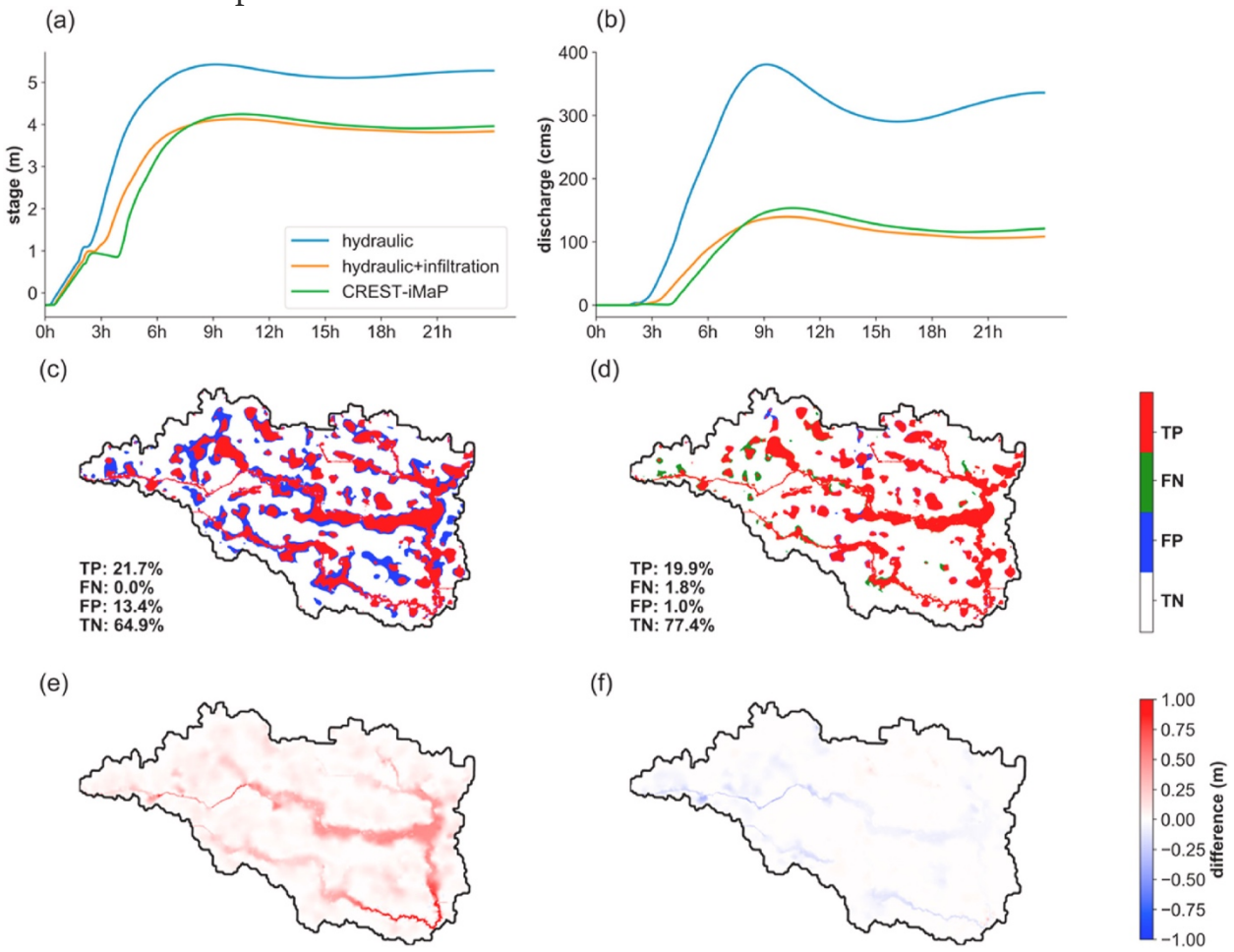


Figure 4.8 Differences of (a) river stage and (b) discharge time series at outlet. The binary counts of flood extent from the (c) hydraulic and (d) hydraulic + infiltration, compared to CREST-iMAP. The difference of maximum water depth from (e) the hydraulic and (f) hydraulic + infiltration, compared to CREST-iMAP.

4.3.2 Case study: Hurricane Harvey

The above study based on a synthetic rainfall event revealed the inter-comparison of different model engines and the sign of the results were expected. There is a need to verify the models with respect to observations in a real-world case study. To this end, we simulate an extreme event – the 500-year Hurricane Harvey, that has been a well-studied event for a modeling purposes (Chen et al., 2020; Noh et al., 2019; Saksena et al., 2019; Wing et al., 2019). Hurricane Harvey, starting as a tropical storm, became a category four hurricane while making landfall in Texas,

causing devastating urban flooding in August 2017. The regional accumulative rainfall set a record (1625 mm) since the 1880s for its seven-day lifespan (Li et al., 2020). The torrential rainfall has resulted in unprecedented pluvial and fluvial combined flooding, thereby leading to at least 70 fatalities and economic damages beyond 125 billion (US dollars). The basin-average rainfall rate time series is illustrated in Fig. 4.9, with markedly intense rates (beyond 60 mm h−1) occurring during 08/27/2017 to 08/28/2017. Four individual model engines at their optimal performance (fine-tuned) are intercompared and also verified against the observations for this event.

4.3.2.1 Integrated streamflow

The streamflow at collocated USGS sites are retrieved for four model engines: the CREST-iMAP, CREST, hydraulic, and hydraulic + infiltration as shown in Fig. 4.9. Table 3 correspondingly summarizes the validation metrics described in Table 4.2. Notably, the outlet gauge (USGS 08076700) only records reliable streamflow data until 08/28/2017, probably due to damage of the in-situ instrument. Compared to the hydrologic models (i.e., CREST), the (coupled) hydraulic models (i.e., CREST-iMAP, hydraulic, and hydraulic + infiltration) overall better capture the streamflow as indicated visually and with higher NSE values. This observation relates to two limitations of the CREST model. First, the routing scheme (i.e., kinematic wave) encounters issues in flat terrain as previous studies stated (Getirana and Paiva, 2013; Vergara et al., 2016). Second, the hydrologic model mandates sufficient calibration data to match the observation while the tuning period is not adequate. Conversely, it points to the main advantage of this coupled model, which requires less calibration efforts to produce reasonable results. Amongst all hydraulic-involved models, CREST-iMAP outperforms the other two for all the gauge stations with the highest NSE values, the lowest RMSE and peak flow errors. On the other hand, the hydraulic and hydraulic + infiltration reach comparable results. The hydraulic-only configuration unsurprisingly overpredicts the streamflow as indicated by the RB due to ignored infiltration process at five gauges; however, the hydraulic + infiltration, due to crudely represented infiltration process, exhibits less flow volume than CREST-iMAP. In summary, the infiltration process is essential and sensitive for the generation of basin-integrated streamflow, and the hydrologically sound infiltration process improves the streamflow prediction significantly.

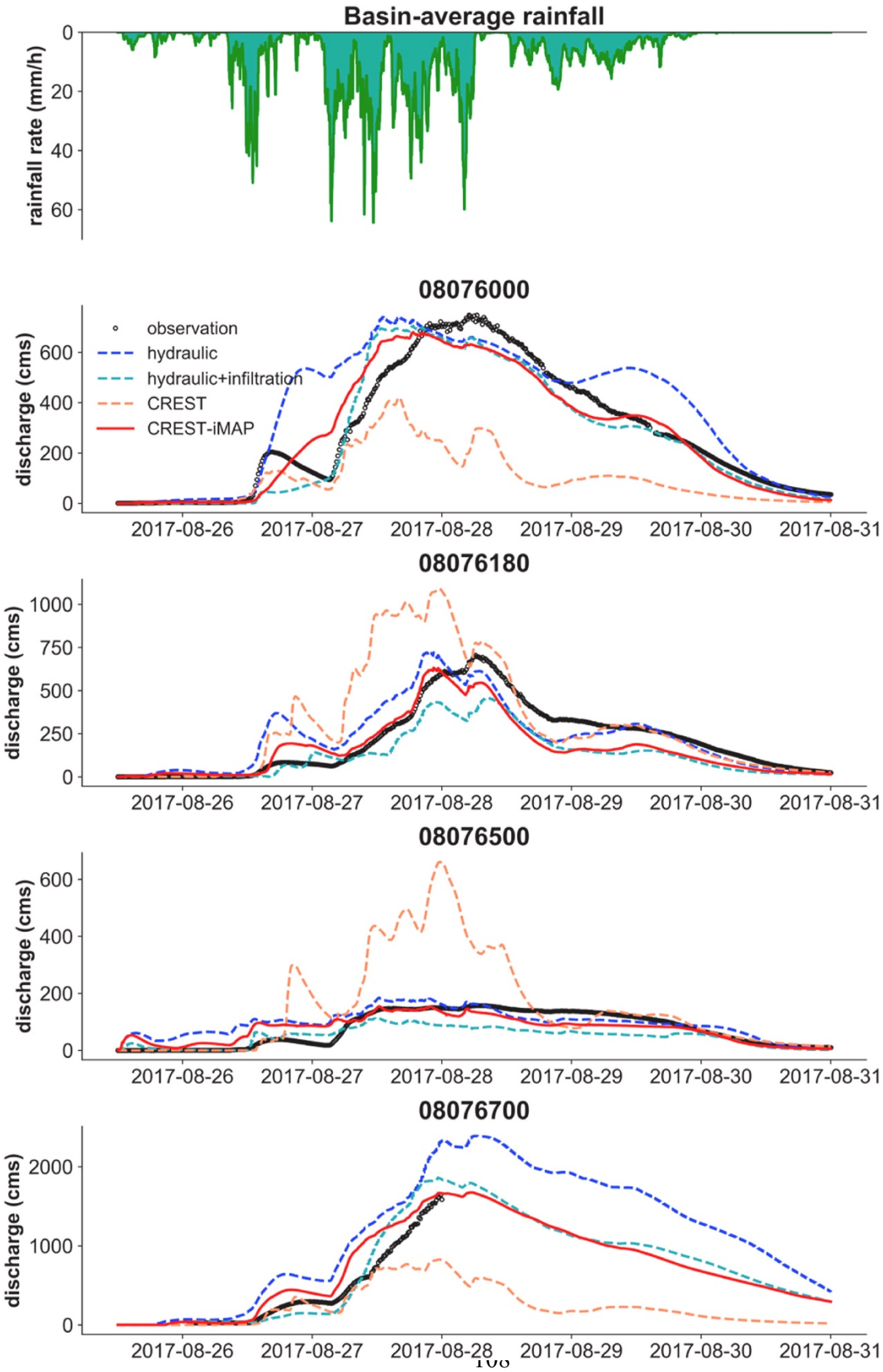

Basin-average rainfall
rainfall rate (mm/h)
0
20
40
60
08076000
discharge (cms)
observation
hydraulic
hydraulic+infiltration
CREST
CREST-iMAP
0
200
400
600
2017-08-26
2017-08-27
2017-08-28
2017-08-29
2017-08-30
2017-08-31
08076180
discharge (cms)
0
250
500
750
1000
08076500
discharge (cms)
0
200
400
600
08076700
discharge (cms)
0
1000
2000

Figure 4.9 The basin-average rainfall and corresponding streamflow simulations for the CREST-iMAP, CREST, hydraulic, and hydraulic + infiltration at collocated USGS sites.

Table 4.2 Validation results for the CREST-iMAP, CREST, hydraulic, and hydraulic + infiltration at collocated USGS sites. The boldfaced cells represent the best model engine (metrics) according to the overall performance.

Site	Model	NSE	RMSE (cms)	Peak flow error (cms)	Time-to-peak error (hours)	RB
08076000	hydraulic	0.63	149	43.2	10.3	0.28
	hydraulic + infiltration	0.90	76.6	**7.72**	13.0	−0.08
	CREST	−0.09	255	326.7	12.9	−0.64
	CREST-iMAP	**0.91**	**73.1**	67.4	**10.3**	**−0.03**
08076180	hydraulic	0.73	103	−20.1	8.75	**0.10**
	hydraulic + infiltration	0.66	115	249	**−2.00**	−0.39
	CREST	−0.39	234	−386.5	6.78	0.46
	CREST-iMAP	**0.77**	**94.6**	**72.2**	7.78	−0.20
08076500	hydraulic	0.55	39.9	−27.0	−5.87	0.27
	hydraulic + infiltration	0.45	43.9	45.8	**−3.00**	−0.35
	CREST	−5.55	152	−503.8	6.93	1.03
	CREST-iMAP	**0.71**	**32.2**	**2.59**	−5.90	**−0.00**
08076700	hydraulic	0.40	370.7	−221.4	0.48	0.68
	hydraulic + infiltration	0.72	254.3	−690.3	−0.25	**0.17**
	CREST	0.68	270.9	811.8	0.32	−0.32
	CREST-iMAP	**0.81**	**207.2**	**−35.2**	**0.18**	0.33

Note: The positive (negative) sign for time-to-peak error indicates an early (late) arrival of flood peaks.

4.3.2.2 Flood extent

The flood extent from two remotely sensed products (i.e., RAPID and DFO) and simulated by CREST-iMAP are overlaid at the same time as shown in Fig. 4.10a and b. The percentages of binary counts with respect to TP, FP, FN, and TN are shown in Fig. 10c and d, referenced by the RAPID and DFO flood extent products. First, there is certainly a visual agreement between the remote-sensing product and CREST-iMAP simulated results, especially near the confluence of the Greens Bayou and Garners Bayou. In contrast, the DFO and model results are overlaid with higher percentages of TP (4.60% on average), as opposed to RAPID (2.00%). In this case, the merged product DFO may better measure the inundation area rather than the single-sourced product RAPID. However, the higher fraction of FN (2.07% vs. 0.83%) in DFO might be subject to the non-distinguishable backscattering signals from water surfaces and urban structures. Instead, RAPID undergoes rigorous preprocessing steps (such as morphologic removal, water classification, and machine-learning corrections) to eliminate noise. Second, our simulations

tend to predict more inundated areas than the remote-sensing observations in the floodplain. Especially downstream, such as the deep forested area near the basin outlet, the remote-sensing products did not capture the flooded areas. This could be explained by the inherent issues for the remote sensing measurements (i.e., SAR and passive microwave), as the vegetated areas and submerged wetlands enhance the dihedral scattering (Martinis and Rieke, 2015). In that regard, it is likely that they miss flood events in the deep forested area. Some related studies such as Saksena et al. (2020) also generate similar flood extent as CREST-iMAP, ensuring the rationality of the result.

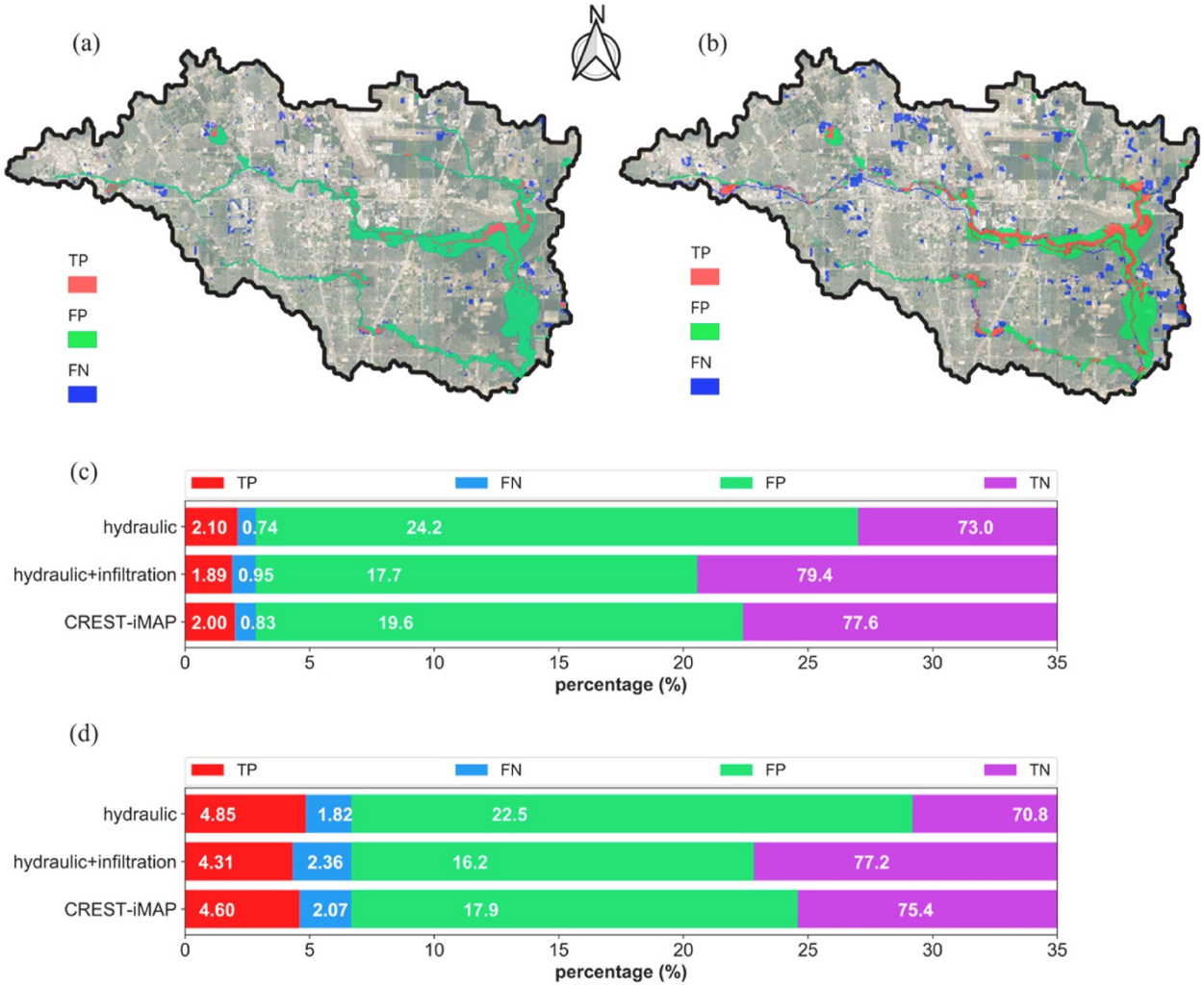


Figure 4.10 2D flood extent results: (a) CREST-iMAP simulated extent compared to RAPID algorithm derived extent; (b) CREST-iMAP simulated extent compared to DFO derived extent; (c) The percentage of hits, misses, false alarms, and correct negatives, comparing CREST-iMAP with RAPID; (d) The percentage of hits, misses, false alarms, and correct negatives, comparing CREST-iMAP to DFO.

For the model engines, CREST-iMAP achieves a comparable percentage of TP with the hydraulic, referenced by **RAPID** and DFO. However, the FP are adequately reduced (**19.6**% and 24.2% for the hydraulic vs. **17.9**% and 22.5% for CREST-iMAP). Compared to hydraulic + infiltration, CREST-iMAP shows more than 0.1% higher percentage of TP and similar FP. Because the hydraulic-only configuration obviously generates more inundated grid cells than models including infiltration processes, it is challenging to justify the performance solely based on TP. Meanwhile, remote-sensing products suffer inherent errors which may bias our judgement. Therefore, it is only reliable to conclude the flood extent by CREST-iMAP is comparable with the hydraulic model, which proves its applicability in flood prediction.

4.3.2.3 Flood depth

The simulated peak flood depth is compared with HWMs collected from 23 sites shown in Fig. 4.11a. First, all the HWMs are located within the simulated flood extent. Second, for the Greens Bayou, the CREST-iMAP exhibits higher depths than the HWMs while lower depths for the Halls Bayou in the south. Nevertheless, according to the Harris County local flood report (https://www.hcfcd.org/Portals/62/Harvey/immediate-flood-report-final-hurricane-harvey-2017.pdf) during Hurricane Harvey, the Greens Bayou gauge broke the flood record (500-year return), but the Halls Bayou only experienced a 10–50 year return flood. Our simulations align well with this frequency analysis. It is worth mentioning the uncertainties when comparing HWMs to model simulations. First, there is a large deviation between surveyed elevation and DEM used for modeling, generally ranging from 1 to 2 m depending on the location. Second, the qualities of HMWs vary, resulting in differences of ±0.5 m. Third, uncertainties between point values (i.e., HWMs) and area values (i.e., simulations) are apparent. The locations of HWMs could be off tens of meters because of truncation errors. Despite of this fact, the maximum difference of 2.6 m is reasonable and acceptable.

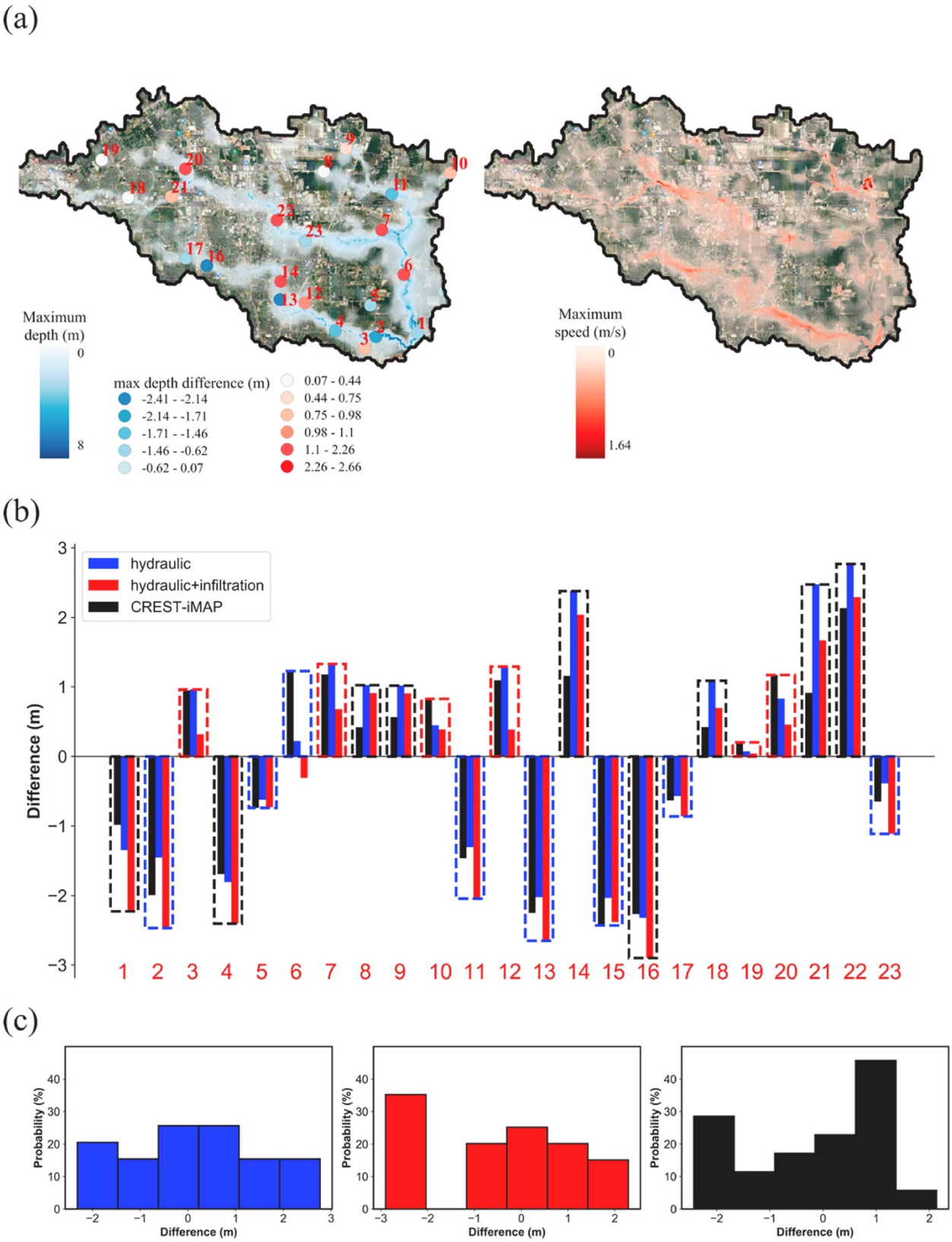


Figure 4.11 Simulated maximum flood depth and flow speed throughout Harvey event. The High-Water Marks are color coded by the difference between the observed maximum depth and

simulated maximum depth. (b) The bar plot of the difference of High-Water Marks and three model engines. The color of the dashed line represents the best engine out of the three. (c) the difference distributions between simulation and High-Water Marks.

With respect to the model engines, Fig. 4.11b reveals the peak depth difference. Of all 23 HWMs, 9 (8) sites show that the CREST-iMAP (hydraulic) outperforms with the lowest absolute difference; the remaining 6 sites have better relative simulations from the hydraulic + infiltration model. Also, the distributions of peak flow differences in Fig. 11c indicate the performance in a broad view. For CREST-iMAP, the errors between estimates and observations vary from −2.4 to 2.1 m, with 74% of them situated within (−1.5, 1.5) meters, which is comparable with the results of 79% from Saksena et al. (2020). However, their results experience overestimations of 3 m. For the hydraulic, the errors vary from −2.3 to 2.8 m, with 69% of them situated within (−1.5, 1.5) meters. Lastly, for the hydraulic + infiltration, the errors vary from −2.9 to 2.3 m, with 57% of them located within (−1.5, 1.5) meters. It is obvious that CREST-iMAP has the narrowest range of the differences between simulations and observations, and the highest percentage of differences within an acceptable error range among all model configurations. The CREST-iMAP exhibits the best performance due to the joint advantages of hydrologic and hydraulic simulations.

4.4 Conclusions and future work

This study introduces a new flood forecasting framework, which is comprised of a mesh generator, an H&H model, and a set of uncertainty/sensitivity analysis tools and calibration schemes. To explore the advances and sensitivities, a synthetic rainfall study and the 500-year Hurricane Harvey event are simulated to systematically evaluate the different model configurations (CREST, hydraulic-only, hydraulic + infiltration, and CREST-iMAP). Here are the main findings for CREST-iMAP:

1) For the model efficiency, it is highly parallelized and fully integrated, promoting its applicability for operational flood prediction and inundation mapping;

2) The flexible mesh generator minimizes structural errors compared to traditional mesh designs;

3) For the parameter sensitivity, antecedent soil moisture is the foremost sensitive parameter for these case studies, followed by land surface characteristics (manning's n and impervious area ratio); the other parameters related to soil moisture are less sensitive, which is attributed to the characteristics of study basin (89.7% developed land);

4) For the hydrologic physical process representation, the CREST-iMAP and CREST reasonably capture the spatial and temporal dynamics of the soil moisture and infiltration rates, while the hydraulic and hydraulic + infiltration misrepresent them, which leads to the different performance in the end products;

5) The infiltration process is central to a flood mapping model, as it greatly influences the basin-integrated discharge, flood extent, and peak flood depth. A real case study of Hurricane Harvey also indicates that CREST-iMAP systematically outperforms the hydraulic simulation without infiltration and the hydraulic + infiltration with crudely represented infiltration process.

Further assessment of CREST-iMAP is necessary to improve this framework. First, current work does not include surveyed channel geometries, which is acceptable with respect to model performance in hydraulic simulations. We expect to see the added values by incorporating refined channel cross sections and terrains. Second, the mesh generator needs to be systematically evaluated with respect to different basin attributes, and the strategies of designing an optimized mesh should be explored. Lastly, an impact-based forecasting module could complete this modeling chain to specifically focus on the potential damages for public awareness, instead of the natural hazard characteristics alone (e.g., intensity, extent, discharge, and frequency).

4.5 Reference

Abbott, M. B., Cunge, J. C., O'Connell, P. E., & Rasmussen J. (1986). An introduction to the European Hydrological System – Systeme Hydrologique Européen 'SHE' 1: History and philosophy of a physically based distributed modelling system, Journal of Hydrology, 87, 45-59.

Allen, R. G., Pereira, L., Raes, D., & Smith, M., 1998. Crop Evapotranspiration, Food and Agriculture Organization of the United Nations, Rome, Italy. FAO publication 56. ISBN 92-5-104219-5. 290p.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large-area hydrologic modeling and assessment: Part I. Model development. Journal of American Water Resources Association, 34 (1), 73-89.

Ashley, S. T., & Ashley, W. S. (2008). Flood Fatalities in the United States. Journal of Applied Meteorology and Climatology, 47, 805–818. https://doi.org/10.1175/2007JAMC1611.1.

Beven, K., & Kirkby, M. J. (1979). A physically based variable contributing area model of basin hydrology, Hydrological Sciences Bulletin, 24(1), 43-69.

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction, Hydrological Processes, 6(3), 279-298. https://doi.org/10.1002/hyp.3360060305.

Beven, K., & Freer, J. (2001). Equifinality, data assimulation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, Journal of Hydrology, 249, 11-29. https://doi.org/10.1016/S0022-1694(01)00421-8.

Chen, M., Nabih, S., Brauer, N. S., Gao, S., Gourley, J. J., Hong, Z., Kolar R. L., & Hong, Y. (2020). Can Remote Sensing Technologies Capture the Extreme Precipitation Event and Its Cascading Hydrological Response? A Case Study of Hurricane Harvey Using EF5 Modeling Framework. Remote Sensing, 12, 445. https://doi.org/10.3390/rs12030445

Chen, S., Gourley, J. J., Hong, Y., Cao, Q., Carr, N., Kirstetter, P., Zhang, J., & Flamig, Z. (2016). Using Citizen Science Reports to Evaluate Estimates of Surface Precipitation Type. Bulletin of American Meteorological Society, 97, 187–193. https://doi.org/10.1175/BAMS-D-13-00247.1

Chow, V.T., Maidment, D.R., Mays L.W. (1988). Applied Hydrology: McGraw-Hill Series in Water Resources and Environmental Engineering, McGraw-Hill, Inc., New York.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geoscientific Model Development, 8, 1991-2007. https://doi.org/10.5194/gmd-8-1991-2015

Duan Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resources Research, 28(4), 1015-1031.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H.V., Gusev, Y.M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O.N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., & Wood, E.F. (2006). Model Parameter Estimation

Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, Journal of Hydrology, 320, 3-17. https://doi.org/10.1016/j.jhydrol.2005.07.031

Felder, G., Zischg, A., & Weingartner, R. (2017). The effect of coupling hydrologic and hydrodynamic models on probable maximum flood estimation, Journal of Hydrology, 550, 157-165. https://doi.org/10.1016/j.jhydrol.2017.04.052

Gan, Y., Liang, X., Duan, Q., Ye, A., Di, Z., Hong, Y., & Li J. (2018). A systematic assessment and reduction of parametric uncertainties for a distributed hydrologic model, Journal of Hydrology, 564, 697-711. https://doi.org/10.1016/j.hydrol.2018.07.055

Getirana, A.C.V., & Paiva R.C.D. (2013). Mapping large-scale river flow hydraulics in the Amazon basin, Water Resources Research, 49, 2437-2445. https://doi.org/10.1002/wrcr.20212

Gourley, J. J., Flamig, Z.L., Vergara, H., Kirstetter, P., Clark, R.A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., & Yang, H. (2017). The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. Bulletin of American Meteorological Society, 98, 361–372. https://doi.org/10.1175/BAMS-D-15-00247.1

Griffin, J., Latief, H., Kongko, W., Harig, S., Horspool, N., Hanung, R., Rojali, A., Maher, N., Fuchs, A., Hossen, J., Upi, S., Dewanto, S.E., Rakowsky, N., & Cummins, P. (2015). An evaluation of onshore digital elevation models for modeling tsunami inundation zones, Frontiers in Earth Science, 3(32). https://doi.org/10.3389/feart.2015.00032

Guerra, M., Cienfuegos, R., Escauriaza, C., & Marche, F. (2014). Modeling Rapid Flood Propagation Over Natural Terrains Using a Well-Balanced Scheme, Journal of Hydraulic Engineering, 140(7). https://doi.org/10.1061/%28ASCE%29HY.1943-7900.0000881

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resources Research, 34(4), 751– 763. https://doi.org/10.1029/97WR03495

Huggins, L. F., & Monke, E. J. (1968). A Mathematical Model for Simulating the Hydrologic Response of a Watershed, Water Resources Research, 4(3), 529-539. https://doi.org/10.1029/WR004i003p00529

Issermann, M., & Chang, F. (2020). Uncertainty Analysis of Spatiotemporal Models with Point Estimate Methods (PEMs) – The Case of the ANUGA Hydrodynamic Model, 12(1), 229. https://doi.org/10.3390/w12010229

Ivanov, V.Y., Vivoni, E.R., Bras, R.L., & Entekhabi, D. (2004). Preserving high-resolution surface and rainfall data in operational-scale basin hydrology: a fully-distributed physically-based approach, Journal of Hydrology, 298, 80-111. https://doi.org/10.1016/j.hrol.2004.03.041

Kim, J., Warnock, A., Ivanov, V. Y., & Katopodes, N. D. (2012). Coupled modeling of hydrologic and hydrodynamic processes including overland and channel flow, Advances in Water Resources, 37, 104-126.

Kim, B., Sanders, B.F., Schubert, J.E., & Famiglietti, J.S. (2014). Mesh type tradeoffs in 2D hydrodynamic modeling of flooding with a Godunov-based flow solver, Advances in Water Resources, 68, 62-61. https://doi.org/10.1016/j.advwatres.2014.02.013

Li, Z., Chen, M., Gao, S., Hong, Z., Tang, G., Wen, Y., Gourley J.J., & Hong, Y. (2020). Cross-Examination of Similarity, Difference and Deficiency of Gauge, Radar and Satellite Precipitation Measuring Uncertainties for Extreme Events Using Conventional Metrics and Multiplicative Triple Collocation. Remote Sensing, 12, 1258. https://doi.org/10.3390/rs12081258

Martinis, S., & Rieke, C. (2015). Backscatter analysis using multi-temporal and multi-frequency SAR data in the context of flood mapping at River Saale, Germany, Remote Sensing, 7, 7732-7752. https://doi.org/10.3390/rs70607732

Marsh, C.B., Spiteri, R.J., Pomeroy, J.W., & Wheater H.S. (2018). Multi-objective unstructured triangular mesh generation for use in hydrological and land surface models, Computational Geosciences, 119, 49-67, https://doi.org/10.1016/j.cageo.2018.06.009

Montanari, M., Hostache, R., Matgen, P., Schumann, G., Pfister, L., & Hoffmann, L. (2009). Calibration and sequential updating of a coupled hydrologic-hydraulic model using remote sensing-derived water stages. Hydrology and Earth System Science, 13(3), 367-380. https://doi.org/10.5194/hess-13-367-2009

Morris, M. (1991). Factorial sampling plans for prelininary computational experiments, Technometrics, 33(2), 161-174.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I – A discussion of principles, Journal of Hydrology, 10(3), 282-290. https://doi.org/10.1016/0022-1694(70)90255-6

Nguyen, P., Thorstensen, A., Sorooshian, S., Hsu, K., AghaKouchak, A., Sanders, B., Koren, V., Cui, Z., & Smith, M. (2016). Journal of Hydrology, 541, 401-420. https://doi.org/10.1016/j.jhydrol.2015.10.047

Ni, Y., Cao, Z., Liu, Q., & Liu Q. (2020). A 2D hydrodynamic model for shallow water flows with significant infiltration losses. Hydrological Processes, 34. https://doi.org/10.1002/hyp.13722

Nielsen, O., Roberts, S., Gray, D., McPherson, A., & Hitchman, A. (2005). Hydrodynamic modelling of coastal inundation, MODSIM 2005 International Congress on Modelling and Simulation, 518-523.

Noh, S. J. Lee, J. H., Lee S., & Seo, D. J. (2019). Retrospective Dynamic Inundation Mapping of Hurricane Harvey Flooding in the Houston Metropolitan Area Using High-Resolution Modeling and High-Performance Computing. Water, 11, 597. https://doi.org/10.3390/w11030597

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow, Environmental Modelling and Software, 79, 214-232. https://doi.org/10.1016/j.envsoft.2016.02.008

Pontes, P. R. M., Fan, F. M., Fleischmann, A. S., de Paiva, R. C. D., Buarque, D. C., Siqueira, V. A., Jardim, P. F., Sorribas, M. V., & Collischonn, W. (2017). MGB-IPH model for hydrological and hydraulic simulation of large floodplain rivers systems coupled with open source GIS, Environmental Modelling and Software, 94, 1-20, https://doi.org/10.1016/j.envsoft.2017.03.029

Saksena, S., Dey, S., Merwade, V., & Singhofen, P. J. (2020). A computationally efficient and physically based approach for urban flood modeling using a flexible spatiotemporal structure. Water Resources Research, 56, e2019WR025769. https://doi.org/10.1029/2019WR025769

Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., & Freer, J. E. (2015). A high-resolution global flood hazard model, Water Resources Research, 51, 7358–7381. https://doi.org/10.1002/2015WR016954

Savenije, H. H. G., 2010. HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)", Hydrology and Earth System Science, 14, 2681–2692. https://doi.org/10.5194/hess-14-2681-2010.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F. W., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environmental Modelling and Software, 90, 201-216. https://doi.org/10.1016/j.envsoft.2017.01.006

Vergara, H., Kirstetter, P., Gourley, J. J., Flamig, Z. L., Hong, Y., Arthur, A., & Kolar, R. (2016). Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the Conterminous United States, Journal of Hydrology, 541, 421-433. https://doi.org/10.1016/j.jhydrol.2016.06.011

Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., Adler, R. F., Policelli, F. S., Habib, S., Irwn, D., Limaye, A. S., Korme, T., & Okello, L. (2011). The coupled routing and excess storage (CREST) distributed hydrological model, Hydrological Sciences Journal, 56 (1), 84-98

Wing, O. E. J., Sampson, C. C., Bates, P. D., Qionn, N., Smith, A. M., & Neal, J. C. (2019). A flood innundation forecast of Hurricane Harvey using a continental-scale 2D hydrodynamic model, Journal of Hydrology, 4. https://doi.org/10.1016/j.hydra.2019.100039

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H.,. Bierkens, M. F. P, Blyth, E., de Roo, A., Doll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giese, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., & Whitehead, P. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring, Water Resources Research, 47, W05301, https://doi.org/10.1029/2010WR010090

Wu, H., Adler, R.F., Hong, Y., Tian, Y., & Policelli, F. (2012). Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model. Journal of Hydrometeorology, 13, 1268–1284. https://doi.org/10.1175/JHM-D-11-087.1

Wu, H., Adler, R.F., Tian, Y., Huffman, G.J., Li, H., & Wang, J. (2014). Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model, Water Resources Research, 50, 2693.2717, https://doi.org/10.1002/2013WR014710

Shen, X., Anagnostou, E. N., Allen, G. H., Brakenridge G. R., & Kettener, A. J. (2019). Near-real-time non-obstructed flood inundation mapping using synthetic aperture radar, Remote Sensing of Environment, 221, 302-315. https://doi.org/10.1016/j.rse.2018.11.008

Singh, V. P., ed. (1995). Computer Models of Watershed Hydrology, Highlands Ranch, CO: Water Resources Publications

Sood, A., & Smakhtin, V. (2014). Global hydrological models: a review, Hydrological Sciences Journal, 60(4), 549-565. https://doi.org/10.1080/02626667.2014.950580

Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G.J., Khan, S.I., Doriji, C., & Chen, S. (2013). Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation priducts 3B42V7 ready for use in ungauged basins? Journal of Hydrology, 499, 91-99, https://doi.org/10.1016/j.hydrol.2013.06.042

Yang, G., Bowling, L. C., Cherkauer, K. A., & Pijanowski, B. C. (2011). The impact of urban development on hydrologic regime from catchment to basin scales, Landscape and Urban Planning, 103(2), 237-249. https://doi.org/10.1016/j.landurbplan.2011.08.003

Zhao, R. J. (1995), The xinanjiang model. Computer models of watershed hydrology, 215-232

Zhu, Z., Oberg, N., Morales, V. M., Quijano, J. C., Landry, B. J., & Garcia, M. H. (2016). Integrated urban hydrologic and hydraulic modelling in Chicago, Illinois, Environmental Modelling and Software, 77, 63-77, https://doi.org/10.1016/j.envsoft.2015.11.014

Zischg, A. P., Felder, G., Mosimann, M., Rothlisberger, V., Weingatner, R. (2018). Extending coupled hydrological-hydraulic model chains with a surrogate model for the estimation of flood losses, Environmental Modelling and Software, 108, 174-185, https://doi.org/10.1016/j.envsoft.2018.08.009

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., & Kitzmiller, D. (2016). Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. Bulletin of American Meteorological Society, 97(4), 621-638. https://doi.org/10.1175/bams-d-14-00174.1

Chapter 5

CREST-iMAP V1.1: A re-infiltration-enabled hydrologic-hydraulic modeling framework for better flood prediction

Publication

Li, Z., Chen, M., Gao, S., Wen, Y., Gourley, J. J., Yang, T., & Kolar, Hong, Y. (2022). Can re-infiltration process be ignored for flood inundation mapping and prediction during extreme storms? A case study in Texas Gulf Coast region. Environmental Modelling & Software, 155, 105450.

5 CREST-iMAP V1.1: A re-infiltration-enabled hydrologic-hydraulic modeling framework for better flood prediction

Abstract

Coupled Hydrologic & Hydraulic (H&H) models have been widely applied for flood simulations, yet the modern H&H models suffer from one-way and weak coupling and particularly disregarded run-on infiltration, which could compromise the model accuracy. In this study, we assess the H&H model performance with and without re-infiltration process in extreme flooding events. Results highlight that the re-infiltration process should not be disregarded even in extreme flood simulations. Saturated hydraulic conductivity and antecedent soil moisture are found to be the prime contributors to such differences. For the Hurricane Harvey event, the model performance is verified against stream gauges and high water marks, from which the re-infiltration scheme increases the Nash Sutcliffe Efficiency score by 140% on average and reduces maximum depth differences by 17%. Meanwhile, the recent update of the CREST-iMAP model Version 1.1, which incorporates two-way coupling and re-infiltration scheme, is released for public access.

5.1 Introduction

Flooding, as a costly natural hazard, has been increasingly threatening human lives and economies (Gourley et al., 2017; Hirabayashi et al., 2013; Li et al., 2021a). In the United States, most billion-dollar natural hazards are tied to either local or regional flooding, making it the major cost to human society. Unfortunately, under a warmer climate with anthropogenic pressure, flood crises are likely to continue expanding, as the flood frequency accelerates and flood magnitude rises (Bates et al., 2021; Hirabayashi et al., 2013; Li et al., 2022a; Tabari, 2020; Triet et al., 2020; Swain et al., 2020; Viero et al., 2019). To combat flood risks, researchers have been developing hydrologic/hydraulic models to deliver accurate and timely flood information for local communities and decision-makers (Gourley et al., 2017). In the United States, two pronounced flood forecasting systems – the NWM (National Water Model) (Cohen et al., 2018; Viterbo et al., 2017) and FLASH (Flooded Locations And Simulated Hydrographs Project) (Gourley et al., 2017; Yussouf et al., 2020) – are capable of both simulating real-time floods and forecasting floods in a short range.

These large-scale flood monitoring systems, although claimed to offer inundation maps and predictions, weaken their hydrodynamic simulation due to computational constraints. For instance, the NWM adopts the Height Above Nearest Drainage (HAND) method to produce flood inundation maps along the river channels by mapping discharge to stage via rating curves (Johnson et al., 2019). This conceptual method, however, overlooks the physics of floodwater propagation because of no flow dynamics being represented (Wing et al., 2017). Moreover, it cannot simulate the pluvial flood, which is a local effect caused by intense rainfall rates and does not normally occur along river channels (Bates et al., 2021). More recently, some emerging hydrodynamic models have been successfully deployed and evaluated at continental or global scales (Bates et al., 2021; Grimaldi et al., 2019; Sampson et al., 2015; Wing et al., 2017; Yamazaki et al., 2011). These models simplify the full Shallow Water Equation (SWE) to speed up the flood simulation. Nevertheless, they normally do not represent the hydrologic process well, especially for the infiltration process, which is proven to be critical in flood simulations (Li et al., 2021b; Ni et al., 2020). As such, a coupled physically-based hydrologic & hydraulic (H&H) model appears to be a better choice, which takes the complementary advantages for accurate flood modeling (Dullo et al., 2021; Felder et al., 2017; Kim et al., 2012; Nguyen et al., 2016; Pontes et al., 2017; Sebastian et al., 2021). Readers are referred to Teng et al. (2017) and Grimaldi et al. (2019) for a detailed

review of coupled models. Most of such models, however, adopt one-way and weak coupling, meaning that there is no interplay between the hydrologic component and hydraulic component (Bravo et al., 2013). They normally produce surface runoff outputs first to drive the hydraulic model. The recent development of the Coupled Routing and Excess STorage inundation MApping and Prediction (CREST-iMAP) version 1.0 also uses this one-way coupling strategy (Chen et al., 2021; Li et al., 2021b).

Two-way coupling for the H&H models has not hitherto been well-recognized. The accumulated surface water (hydraulic feature), along with excess surface runoff during a flood event in principle would alter infiltration rates, whereby both the flood magnitudes and timings could differ. Therefore, we activate the surface water infiltration along its way to downslope, which is called run-on infiltration or re-infiltration in short (Smith & Hebbert, 1979; Nahar et al., 2004; Zhang et al., 2020). Nahar et al. (2004) defined this re-infiltration as the infiltration of surface water that, as it moves downslope, encounters areas where moisture deficit has not yet been satisfied. It is often ignored in rainfall-runoff studies, while it can be significant when the random nature of infiltration properties is taken into account (Corradini et al., 1998; Nahar et al., 2004). Smith and Hebbert (1979) simulated the run-on process with varying saturated hydraulic conductivity and rainfall rates, and they reported that the effect of the run-on process is to decrease the ponding time dramatically. Corradini et al. (2002) compared models with and without re-infiltration, and they suggested that re-infiltration greatly reduces surface flow and alters both rising and recession limbs of the hydrograph. Nahar et al. (2004) emphasized the influence of re-infiltration in hillslope hydrograph using the Green-Ampt model with a 1D kinematic wave surface routing. A recent study by Zhang et al. (2020) takes it one step further, in which they applied the community model WRF-Hydro (Weather Research Forecasting model-Hydrological modeling system) to explore the influence of rainfall rates, topography, soil types on the re-infiltration process. However, none of these studies have considered the implication of re-infiltration to hydrodynamic studies, where the overland flow is driven by the 2D Shallow Water Equation (SWE) instead of 1D routing. To do so, we can obtain a more realistic view of how re-infiltration plays a role in flood simulations.

During extreme flood events, the infiltration process is often disregarded because the infiltration rates are relatively low compared to excess rainfall rates. Yet, some studies claim that the infiltration process is critical to determine flood wave propagation, such as arrival and

dissipation, especially in flat plain or regions with highly permeable soil media (Corradini et al., 2002; Mahapartra, et al., 2020; Nahar et al., 2004; Li et al., 2021b; Woolhiser et al., 1996). A hydrodynamic model without infiltration is likely to overestimate flood depth (Kim et al., 2012; Li et al., 2021b; Ni et al., 2020). Nevertheless, it still remains unclear, or at least not as clear as infiltration, whether re-infiltration is essential for H&H models in extreme flood events. In other words, whether it is worth encapsulating such a scheme in modern flood simulation frameworks. To our knowledge, few studies have attempted to answer this question under the context of extreme flood events. Moreover, further questions can arise as to what the determining factors are during such a process and how it interacts with both flood magnitude and dynamics. **In light of these questions, the objectives of this study are to explore 1) the effectiveness and importance of the re-infiltration scheme to an H&H model, 2) the contributing factors to the differences between with and without re-infiltration, and 3) whether and to what extent the re-infiltration process can help improve flood inundation mapping and prediction of extreme events.** We first test its effectiveness and importance on a 100-year design extreme rainfall event during a sensitivity test and then apply it to a real case study – Hurricane Harvey – to validate the efficacy. It is anticipated to provide insightful information for model developers and researchers to understand the importance of the re-infiltration process to flood modeling. In this study, we also release our latest development of CREST-iMAP V1.1, which features a two-way coupling and re-infiltration scheme on top of the previous version (Chen et al., 2021; Li et al., 2021b).

The rest of this paper is structured as follows. Section 2 introduces the study area and necessary datasets for the model setup, followed by experimental designs. Section 3 presents the results from the sensitivity test and the Hurricane Harvey event. Section 4 discusses limitations of this study as well as recommendations for input data and future model development. At last, Section 4 concludes the main findings of this study.

5.2 Methods

5.2.1 Forcing data

Precipitation is the major driver of local or regional flooding, and it is thus central to acquire an accurate and high-resolution dataset. In the U.S., the Multi-Radar Multi-Sensor (MRMS) precipitation product, developed at the National Severe Storms Laboratory (NSSL), provides 2-min and 1-km rainfall field, making it suitable for flash flood forecasting (Yussouf et

al., 2016). It integrates ~180 WSR-88D operational radars, creating a seamless radar mosaic across the CONUS and southern Canada. Recent studies (e.g., Chen et al., 2020; Li et al., 2020) verified the efficacy of MRMS data when compared to gauge-based and satellite-based products during the Hurricane Harvey event. The advantage of using radar rainfall is obvious for flood inundation modeling, as conventional rain gauges cannot readily represent the spatially variable rainfall fields. The MRMS data was downloaded at https://mtarchive.geol.iastate.edu/.

Besides precipitation data, potential evapotranspiration (PET) is a common input into a hydrologic system when the surface radiation is not resolved in the modeling process. In this real-case study, we obtain the PET data from the USGS FEWS data port (https://earlywarning.usgs.gov/fews) at daily temporal and 1° spatial resolution (Allen et al., 1998).

5.2.2 Environmental data

The modeling system requires inputs from the terrain, Land Use Land Cover (LULC), and soil type and depth. Among these variables, terrain data arguably plays the utmost important role in hydraulic simulation (Dullo et al., 2021; Schumann & Bates, 2018). There has been a thorough investigation of terrain data affecting flood inundation modeling since the early development of hydrologic/hydraulic models (Kenward et al., 2000; Sanders, 2007). Lately, with the increasing interest in deploying macro-scale flood inundation simulations, global terrain data assessment has been again brought up (Mohanty et al., 2020; Sampson et al., 2015). Generally, three types of data are favored and available in the U.S.: 1) airborne light ranging and detection (LiDAR) that resolves terrain with a high degree of vertical accuracy (0.05–0.2 m) and comes with a high spatial resolution but limited areal coverage, 2) spaceborne radar interferometry (IfSAR, e.g., Shuttle Radar Terrain Mission) that provides global coverage but poor vertical accuracy (~10 m) and spatial resolution (~90 m), and 3) a mixed product such as the National Elevation Dataset (NED) from the USGS that merges LiDAR surveys and the USGS quadrangle maps, whose accuracy (~5-7 m) and resolution (~5/10/30 m) sit in between the former two products. A general consensus from these studies is that LiDAR data is the most favorable DEM owing to improved vertical accuracy in flood modeling (Mohanty et al., 2020; Sanders, 2007; Schumann & Bates, 2018) but they have to be accompanied by surveyed channel profile since channel depth is not reflected. IfSAR, however, degrades its quality because of poor vegetation penetration and speckle noise

while the NED smooths some artifacts. The NED 10 m data accurately represents the river channel morphology than high-resolution LiDAR data that cannot penetrate water surface. Therefore, in this study, we select the 10 m DEM data from the NED dataset in the study area. To confirm the river channel bathymetry, 13 surveyed river geometries from the Harris Country Flood Control are curated and compared to NED 10 m, shown in Table 5.1. The average difference is found to be small (~0.55 m).

The LULC and impervious area data are acquired from the National Land Cover Database (NLCD) at 30 m resolution to derive a-priori parameter sets. The soil type dataset is retrieved from the United States Department of Agriculture.

Table 5.1 Digital Elevation Model and surveyed channel bottom elevation comparison

id	lat	lon	gage (feet)	gage (meter)	NED_10m	Difference (m)
1650	29.965025	-95.271954	52.82	16.099536	16.53	-0.430464
1630	29.933648	-95.233574	31.61	9.634728	10.51	-0.875272
1680	29.861698	-95.334883	39.21	11.951208	12.04	-0.088792
1690	29.892682	-95.396717	59.72	18.202656	18.7	-0.497344
1675	29.849306	-95.282843	18	5.4864	5.5	-0.0136
1620	29.837012	-95.233773	-2.12	-0.646176	0.29	-0.936176
1685	29.849798	-95.229007	0.51	0.155448	1.63	-1.474552
1640	29.91791	-95.306565	35.39	10.786872	11.96	-1.173128
1600	29.891907	-95.237623	17.03	5.190744	5.75	-0.559256
1670	29.948972	-95.51941	99.2	30.23616	30.33	-0.09384
1160	29.973483	-95.598483	101.14	30.827472	31.14	-0.312528
1660	29.956166	-95.416142	64	19.5072	20.06	-0.5528
1655	29.972595	-95.435011	76.91	23.442168	23.63	-0.187832

5.2.3 Study area

Greens Bayou Basin, located in the north of the Houston metropolitan region, is one of the areas that are susceptible to regional flooding because, firstly, landfalling tropical cyclones and hurricanes bring torrential rainfall within a short period; secondly, the urban development in the recent years have altered the local ecosystem (e.g., replacement of soil with built-up structures). The basin is relatively flat (~1.5%), with an average elevation of around 23.65 meters, and the total drainage area is 457.9 km^2. Three main streams flow across this region. Reinhardt Bayou (drainage

area: 86.3 km^2) flows from north to south, met with Greens Bayou to form the longest river in this area. Halls Bayou (drainage area: 225.1 km^2), the second-longest river, meets Greens Bayou at the basin outlet (Figure 5.1a). The five USGS stream gauges, situated at each mainstream, monitor instantaneous streamflow at a 15-min time interval. Nearly 90% of the area is well-developed, especially in the western portion; forests and wetlands are present downstream, close to the basin outlet (Figure 5.1b). The soil types are dominated by a mixture of sand, clay, and loam (Figure 5.1c). The typical runoff generation mechanism in this region is infiltration excess runoff when extreme rain rates surpass soil infiltration capacity, indicated by relatively low hydraulic conductivity values (Buchanan et al., 2018). Meanwhile, the correlation between rainfall and streamflow is above 0.6, pointing to a flashy hydrograph (Berghuijs et al., 2016).

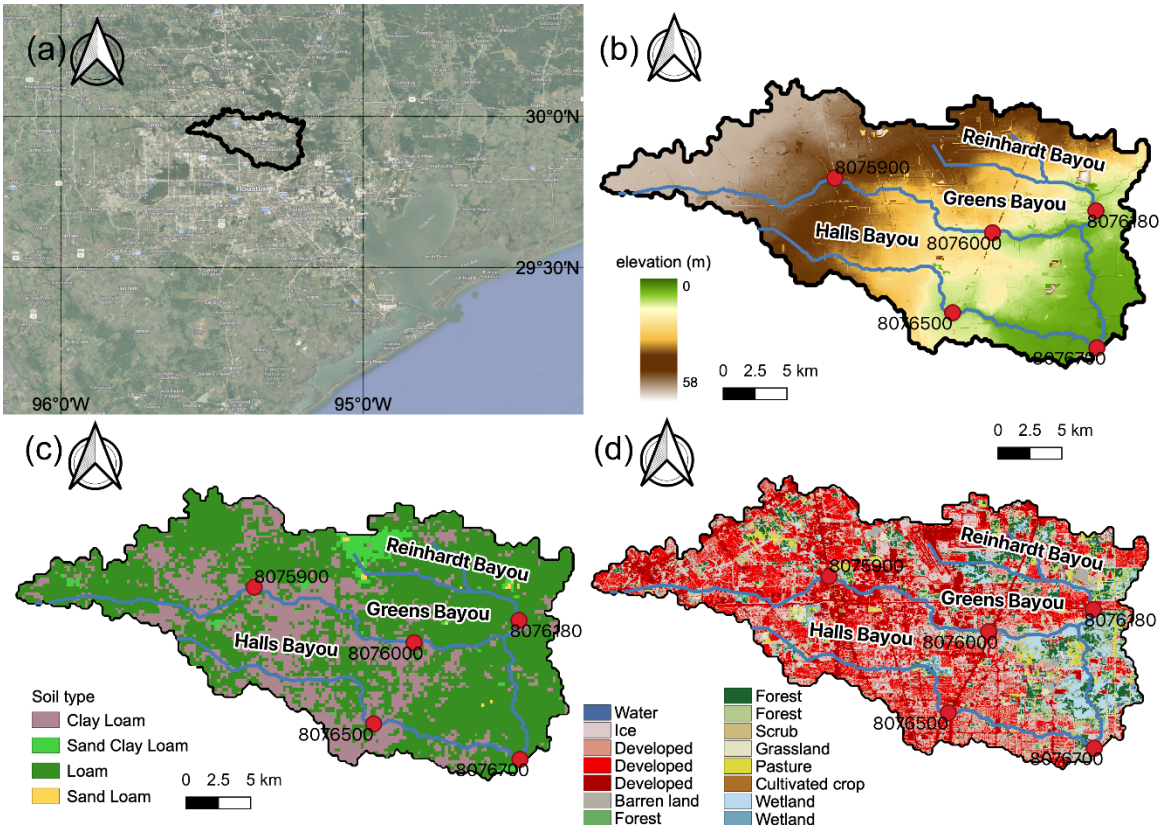


Figure 5.1 Maps of the (a) location of the study region, (b) digital elevation model, (c) soil type, and (d) land use land cover.

During the 500-year Hurricane Harvey event, this region is largely inundated due to record-breaking 1600 mm rainfall over a one-week storm lifespan (Chen et al., 2020; Li et al., 2020). According to the Harris Country flood report, both Greens Bayou and Halls Bayou experienced a

500-year water level downstream and 50-year to 100-year in between upstream. Greens Bayou broke previous water level records in 2002 and observed flooding occurred along the entire channel.

5.2.4 CREST-iMAP model

Hydrologic modeling is thus far a common approach to deliver timely flood information for the sake of scalability and efficiency (Gourley et al., 2017). Yet, conventional hydrologic models bear large uncertainties in such developed regions, which is mainly due to 1) simplified representation of terrain (Dullo et al., 2021) and 2) one-dimensional routing that raises issues in flat regions (Flamig et al., 2020; Getirana & Paiva, 2013; Li et al., 2021b). On the other hand, hydraulic models do not excel in representing hydrologic processes. In light of these issues, the newly developed Coupled Routing and Excess STorage inundation MApping and Prediction (CREST-iMAP) model is used to investigate the importance of the re-infiltration scheme in flood inundation models. The CREST-iMAP integrates CREST V2.1 for the hydrologic part that simulates vertical water distribution by land surface and ANUGA V2.1 for the hydraulic routing that distributes spatial water over terrain by solving 2D shallow water equation. Its performance has been evaluated in this region against the non-coupled hydrologic models and other popular coupled models – WRF-Hydro+HAND and LISFLOOD FP (Chen et al., 2021; Li et al., 2021b). CREST-iMAP achieves similar performance with LISFLOOD-FP, if not better, and generally outperform WRF-Hydro+HAND. However, the previous version of CREST-iMAP V1.0 does not include the re-infiltration scheme, meaning that surface running water is not allowed to re-enter the soil. Here, we release the CREST-iMAP V1.1, an upgrade version, which considers two-way coupling via exchanging surface water between the hydraulic and hydrologic module and re-infiltration. Two different schemes are illustrated schematically in Figure 5.2, where the left panel represents the re-infiltration scheme, and the right does not. The CREST-iMAP V 1.0 and V1.1 are openly accessible from https://github.com/chrimerss/CREST-iMAP.

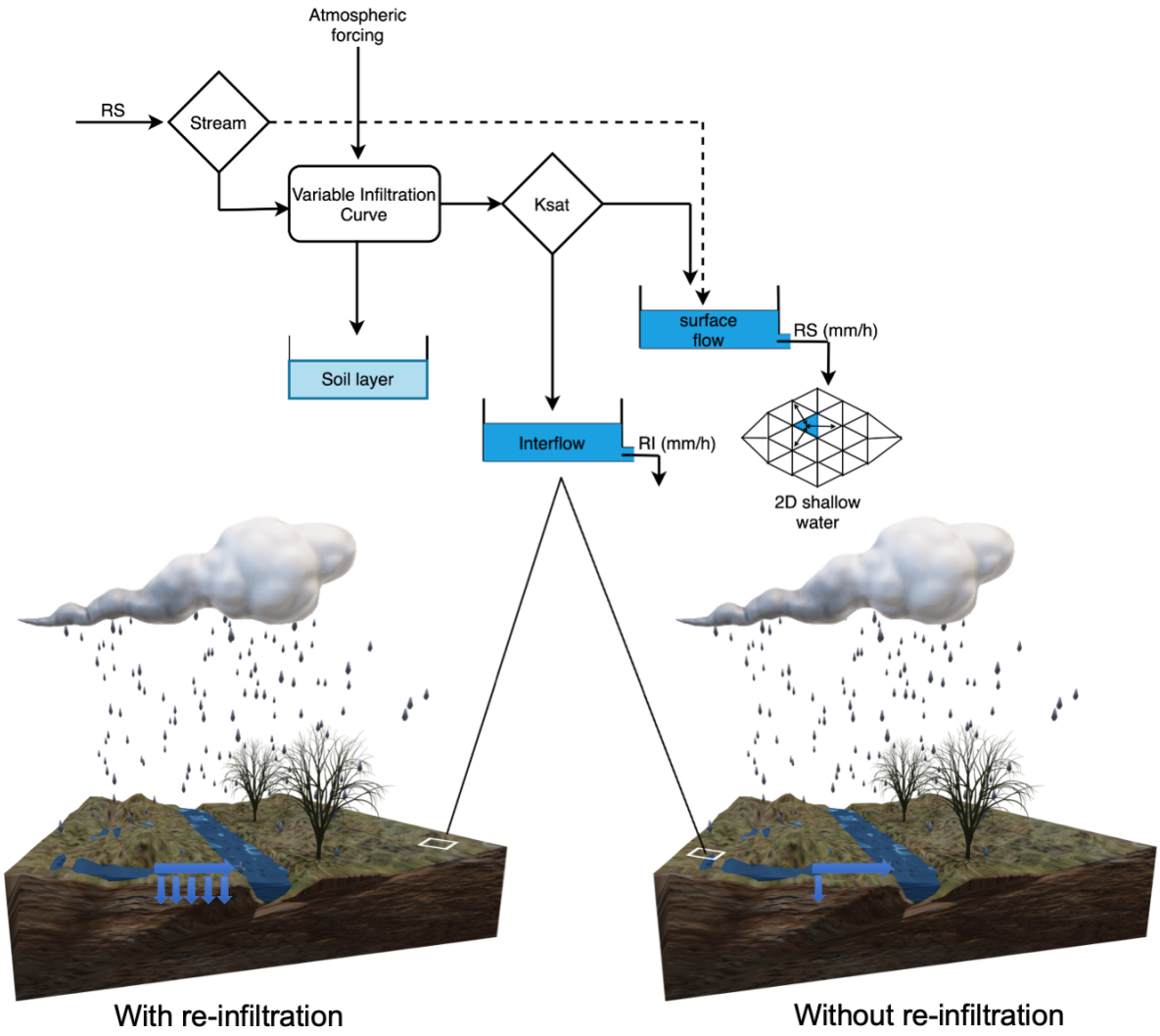


Figure 5.2 Schematic illustration of the re-infiltration scheme.

CREST-iMAP inherits the previous version of the CREST model, which simulates saturation excess runoff as the primary runoff generation process (Wang et al., 2011; Xue et al., 2013; Flamig et al., 2020; Li et al., 2022). The schematic model structure is depicted in Figure 5.2. The study area is discretized in variable triangular meshes which allow higher density in river channels to resolve high-resolution river flow. Each modeling unit receives excess rainfall (rainfall minus evaporation) from forcing data. Then surface water is divided into overland flow and soil water according to the impervious area ratio through linear weighting. Overland flow is generated once soil water exceeds its holding capacity; otherwise, soil water is separated into the remaining amount and interflow based on the Variable Infiltration Curve (VIC) concept, as shown in Eq.5.1. The VIC model is a widely recognized infiltration model that has been applied in several classic

hydrologic models (Liang et al., 1994; Zhao, 1995). Overland flow, combined with the impervious area and saturation excess flow, is eventually fed into the 2D shallow water equation solver – the Finite Volume Scheme. It solves water depth and momentum distributed at each grid cell and propagates across boundaries. The outputs of the model include water depth, velocity, discharge, and soil moisture at a desired time step. The flexibility of the unstructured mesh in CREST-iMAP allows dense meshes in regions that reflect high terrain variability (e.g., river channel) and sparse meshes in other regions (e.g., flood plain). This study simulates the extreme flood events at 10-m resolution using the embedded unstructured mesh generator.

$$i = i_{max} \times [1 - (1 - A)^{\frac{1}{B}}], \tag{5.1}$$

Where i is the infiltration rates, i_{max} is the maximum infiltration capacity, A is the fractional area of the curve, and B is the exponent of the VIC curve.

There are five hydrologic parameters and one hydraulic parameter for the CREST-iMAP, which are listed in Table 5.2 along with parameter ranges. It is noteworthy that all these parameters are spatially distributed to account for the spatial heterogeneity of land cover and soil types. The mean soil saturated hydraulic conductivity, *Ksat* from 0 to 20 mm/d, indicates the soil infiltration capability. Higher *Ksat* values imply higher infiltration rates if soils are not saturated while reaching plateau for the saturated soils. The mean soil water capacity, *WM* from 10.4 to 365.4 mm, measures the total water content the soil can hold with lower value representing the impermeable soils. The exponent of the Variable Infiltration Curve (VIC), *B*, determines soil water partitioned to saturation excess runoff or interflow, with a higher *B* value corresponding to higher infiltration rates. *KE* is the ratio of the potential evapotranspiration to actual evapotranspiration, similar to the concept of pan coefficient. These soil-related a-priori parameters can be approximated from a look-up table at an individual grid cell basis (Chow et al., 1988). There are also CONUS-wide optimized parameter sets that are configured for operational flood monitoring systems (Flamig et al., 2020). The impervious area ratio, *IM* from 0% to 100%, is obtained directly from the NLCD dataset; the manning's n coefficient is derived from the LULC via a look-up table. Both parameters determine water conveyance capacity, meaning that higher values relate to faster and larger flood peaks. The hydrologic parameters are configured at their optima based on previous study – a CONUS-wise calibrated parameters (Flamig et al., 2020), but for the hydraulic parameter – manning coefficient, we manually adjusted it in a preceding event to ensure generating timely and accurate possible

flood peaks. Specifically, we use 2017-08-20 to 2017-08-25 to calibrate the manning coefficient, as well as in-channel water stage.

Table 5.2 Parameters required in CREST-iMAP framework.

Parameters	description	Range
Ksat	Soil saturated hydraulic conductivity (mm/d)	0–2827.2
WM	Mean soil water capacity (mm)	80–200
B	The exponent of the variable infiltration curve	0.05–1.5
IM	Impervious area ratio (%)	0–100
KE	The ratio of the PET to actual evapotranspiration	0.1–1.5
SM_0	Initial soil moisture	0–1
Manning's n	The coefficient for the use of manning's equation in channel flow	0–1

5.2.5 Experiment

5.2.5.1 Synthetic experiment

The importance of re-infiltration in principle is governed by (1) soil properties, (2) soil water saturation, and (3) excess rainfall rates. To quantify the relative importance and generalize our results, we decide to conduct a sensitivity test in this study area to mimic different environment while preserving other variables. The sensitivity analysis addresses the following hypotheses: 1) discernable differences exist when switching on and off re-infiltration scheme, 2) re-infiltration alters flood inundation magnitude and dynamics, 3) differences are amplified when increasing soil infiltration rates and drying antecedent soil saturation, and 4) differences increase with more frequent rainfall storms. Of five hydrologic parameters, we select two soil parameters (i.e., *Ksat* and *B*) that have a direct interaction with infiltration rates. Increase in *Ksat* and *B* promote re-infiltration amount. Additionally, the antecedent soil moisture (SM_0) and roughness parameter (n), proven to be critical for flood generation (Li et al., 2021b; Yang et al., 2011), is another term to change infiltration dynamics. We applied a multiplier to each parameter of interest, ranging from 0.0 to 2.0 with 0.1 spacing except for SM_0 that only ranges from 0.0 (completely dry) to 1.0 (fully saturated) with 0.1 spacing.

For the forcing data in this experiment, we consider a 100-year extreme in the study area by looking up the local Intensity-Duration-Frequency table. This determined rainfall rates are uniform across 2 hours without spatial heterogeneity to eliminate the impact of rainfall spatial structure because we solely consider the impact by soils. We run the model for 24 hours for each parameter, totaling 50 runs.

5.2.5.2 Real case – Hurricane Harvey

Hurricane Harvey is one of the most destructive extreme weather events happened in this study area with substantial damaging winds and urban flooding. The storm was stalled over the Houston region for one week with continuous falling of extreme rains to develop pluvial and fluvial flooding, compounded by costal surges. According to the precipitation estimates by gauges and radars, 1,539 mm maximum rainfall was observed, and most locations in the study area recorded at least 760 mm rainfall, making it the wettest tropical cyclone on record. As a result, almost 25-30 percent of Harris Country was submerged during this event, leading to at least $125 billion economic damage, the second largest natural disasters in US history. Many stream gauges malfunctioned (e.g., being flushed) during high flows. Owing to the socioeconomic impact, a variety of flood simulations were conducted in this region (Chen et al., 2021, 2022; Dullo et al., 2021; Li et al., 2021b; Sebastian et al., 2021). The simulation in our study is conducted from 2017-08-26 to 2017-09-01, during which we did not vary model parameters between scenarios with and without re-infiltration. The parameter values are optimized from a previous study (Li et al., 2021b). The initial soil moisture states are obtained from the operational FLASH project (flash.ou.edu/new).

5.2.6 Computational metrics and results interpretation

A set of computational metrics are selected for this study. The binary assessment comparing the scenarios with and without re-infiltration is considered with Positive Positives (PP), Positive Negatives (PN), and Negative Positives (NP). The first notion indicates whether the model results with re-infiltration detects floods, while the second for model results without re-infiltration scheme. The rationale behind this is that flood extent observations, e.g., witness reports, watermark, satellite-derived flood extent, and insurance claims, are still uncertain without ground truth (Bates, 2004; Chen et al., 2021, 2022). For flood magnitude, the depth, area, and volume are calculated as a basin-integrated ratio. For flood dynamics, we inspect the initial inundation timings and total inundation duration that are often factored in flood risk assessments (Merz et al., 2010). The first six metrics listed in Table 3 are calculated at the maximum flood depth across the simulation period. In the real case study, we verify the performance of two schemes against stream gauge measurements, which is so far the most conventional and trustworthy source. During the verification, the Nash-Sutcliffe Efficiency (NSE) and Correlation Coefficient (CC) are the primary

evaluation scores with each indicating the best value of 1. The detailed formulas for calculating these variables are listed in Table 3, as well as their ranges.

Table 5.3 Computational metrics used in this study.

Metrics	**Formula**	**Range**
Positive Positives (PP)	$\frac{\sum_{n=1}^{N}(B_n = 1) \cap (S_n = 1)}{N} \times 100\%$	(0, 100) %
Negative Positives (NP)	$\frac{\sum_{n=1}^{N}(B_n = 1) \cap (S_n = 0)}{N} \times 100\%$	(0, 100) %
Positive Negatives (PN)	$\frac{\sum_{n=1}^{N}(B_n = 0) \cap (S_n = 1)}{N} \times 100\%$	(0, 100) %
Flood area ratio (R_F)	$R_F = \frac{Area_s}{Area_b}$	(0, 1)
Mean water depth ratio (R_H)	$R_H = \frac{\sum_{i=1}^{n} H_{s,i}}{\sum_{i=1}^{n} H_{b,i}}$	(0, 1)
Surface water volume ratio (R_V)	$R_V = \frac{\sum_{i=1}^{n} H_{s,i} \times Area_s}{\sum_{i=1}^{n} H_{b,i} \times Area_b}$	(0, 1)
Initial inundation time differences (T_{init})	$T_{init} = T_s - T_b$	(-T, +T)
Inundation duration differences (D)	$D = D_s - D_b$	(-T, +T)
Nash- Sutcliffe Efficiency coefficient (NSE)	$1 - \frac{\sum_{n=1}^{N}(S_n - O_n)^2}{\sum_{n=1}^{N}(S_n - \bar{O})^2}$	(-inf, 1)
Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{N}\sum_{n=1}^{N}(S_n - O_n)^2}$	(0, inf)
Correlation Coefficient (CC)	$\frac{\sum_{i=1}^{n}(o_i - \bar{o})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{n}(s_i - \bar{s})^2}\sqrt{\sum_{i=1}^{n}(o_i - \bar{o})^2}}$	(0, 1)

Note: subscript s represents the simulation by turning on the re-infiltration scheme, and subscript b indicates the benchmark that turns off the re-infiltration scheme. T is the total simulation time and O denotes the observed river stage.

The RMSE can be further decomposed to reveal the systematic error and random error (Tang et al., 2020). First, we assume an additive error model by fitting a linear regression to our simulated stage to determine regression coefficients *a* and *b*. We assign the new variable as *F*. Then the residual is calculated by the difference of observed river stage *O* and fitted river stage *F*.

$$F = a \times S + b \tag{5.2}$$

$$RMSE_S = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(S - F)^2} \tag{5.3}$$

$$RMSE_R = \sqrt[2]{\frac{1}{n}\sum_{i=1}^{n}(F-O)^2} \quad (5.4)$$

, where *S* is denoted as the simulated river stage and *O* is the observed.

In the results section, we present it in two parts: sensitivity analysis and real case study. The former includes basin-wise difference in an integral to reveal general differences regarding parameters (Section 5.3.1.1) and storm intensity (Section 5.3.1.2). In the real case study, we focus on the efficacy of re-infiltration scheme by comparing to river stage observations (Section 5.3.2.1), and the High Water Marks surveyed in the aftermath of the event (Section 5.3.2.2). In Section 5.3.2.3, we investigate the importance of re-infiltration scheme in real case study by cross-comparing it to the synthetic results.

5.3 Results

5.3.1 Sensitivity analysis

5.3.1.1 Parameter sensitivity

5.3.1.1.1 Basin-average statistics

The overall flood-related differences between scenarios with and without re-infiltration are shown in Figure 5.3, calculated as basin-integral change by averaging each metric over the whole grid cells that are wet (water depth larger than 0.01 m). First, the differences are discernible comparing the two, especially for the surface water volume ratio (RV), which varies from 0.7 to 1. It suggests the surface water with re-infiltration scheme could only account for 70% of the condition without re-infiltration. Previous studies agree that the re-infiltration results in a substantial reduction of river flow discharge, which can be translated to lower flood depth ratio (RH) (Nahar et al., 2004; Woolhiser et al., 1996). For different conditions, the antecedent soil moisture, as expected, exhibits the largest impact on flood inundation dynamics when comparing the two scenarios. Lower initial soil moisture leads to greater differences in flood depth ratio (RH), area ratio (RF), volume ratio (RV), and dynamics. For instance, when the initial soil is completely dry, the average flood depth ratio (RH), area ratio (RF), and volume ratio (RV) ratios of the re-infiltration scheme are 85%, 85%, 67%, respectively.

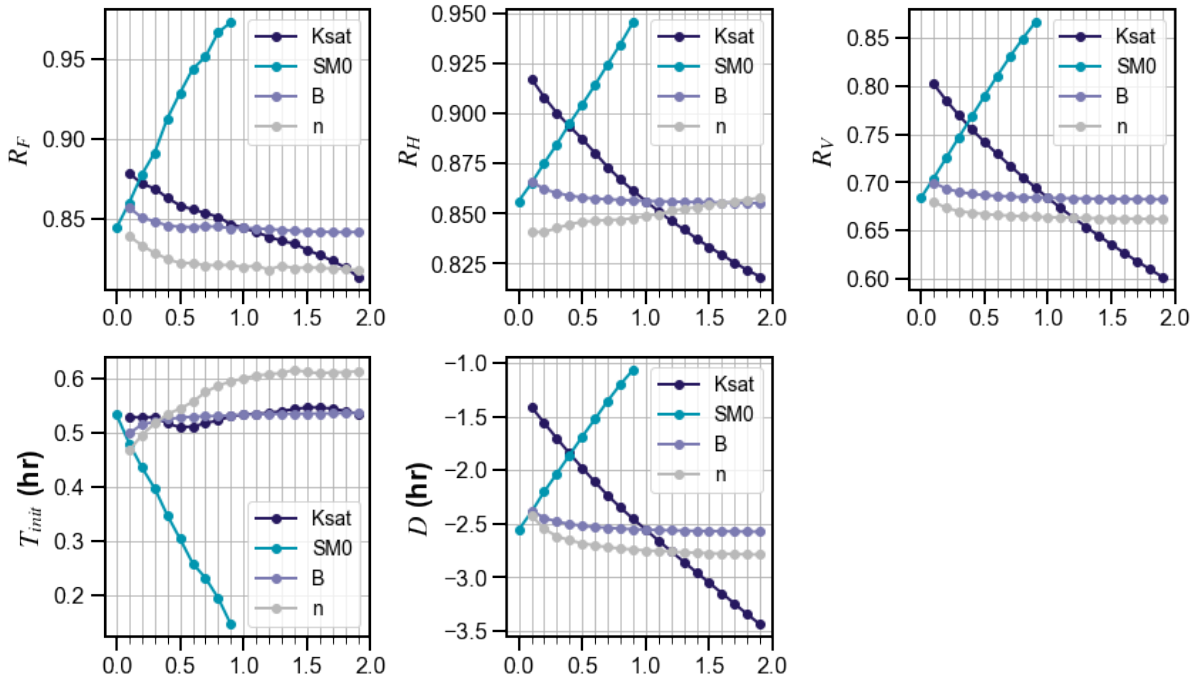


Figure 5.3 Plots of parameter sensitivity with metrics indicated in Table 3. R_F: Flood area ratio; R_H: Flood depth ratio; R_V: Water volume ratio; T_{init}: Flood timing differences; D: Flood duration differences. The no-difference point should be located at (1.0,1.0) for R_F, R_H, R_V and (1.0,0.0) for T_{init} and D.

The initial inundation timing for re-infiltration delays around 0.5 hours (T_{init}=*0.5*), and the total inundation duration is 2.5 hours shorter than the scenario without re-infiltration (D=2.5). The total inundation duration is an important factor for flood risk management (Merz et al., 2010; Triet et al., 2020). As soil gradually approaches saturation, the differences diminish. Saturated hydraulic conductivity, *Ksat*, ranked as the second most sensitive parameter during the test, exponentially reduces flood depth/area/volume by 10%/7%/20% when its multiplier increasing from 0.0 to 2.0. Furthermore, the differences of inundation duration (D) range from 1.5 hours to 3.5 hours, making *Ksat* the most influential parameter; however, the initial inundation timing is relatively insensitive to it, as opposed to initial soil saturation condition. This is due to the fact that *Ksat* only changes infiltration flux along the way while exerting less impact on the initial inundation timings. For higher surface roughness (n), Flood area ratio (R_F) and volume ratio (R_V) decrease. It is expected because as water flows slowly, more water is accumulated above the surface, leaving higher potential to infiltrate. One exception is for the flood depth ratio (R_H), which increases with roughness. It means differences in water depth are shrinking between two schemes for higher roughness. This implies that with increasing roughness, flood inundation calculated with re-

infiltration will result in more concentrated (higher local water depth) yet less widespread (lower inundation areas) flooding. The infiltration parameter *B*, however, has the least impact on the flood inundation dynamics among the three. These measures exhibit the greatest changes at small *B* multipliers (0.1-0.3) and then level out irrespective of increasing *B* multipliers. A plateau is reached because of the constrain of the maximum infiltration capacity. In summary, this sensitivity analysis tests our three main hypotheses, indicating the non-negligible differences between the two schemes and how the soil type and condition influence the results.

5.3.1.1.2 Spatiotemporal relationships

To explore the spatiotemporal differences of scenarios with and without re-infiltration, we set three parameters to their default values, namely an initially dry soil condition and normal soil infiltration rates (i.e., a-priori setting). Figure 5.4 shows the difference in flood extent for the two schemes. Despite a considerable amount of grid cells showing positive agreements (i.e., both detect floods; PP=16.0%), there are still 1.7% of the grid cells issuing NPs, amounting to 15.6k grid cells (~1.56 km^2) in this model configuration. Specifically, those NPs accumulate around upstream floodplains while the downstream such as areas near the basin outlet does not present discernable differences, as the flood depth there due to accumulation is not sensitive to inundation thresholds for flooded cells. Moreover, Figure 5.5 portrays the spatial distribution of the differences with respect to maximum depth, initial inundation timings, and total inundation duration. Figure 5.5a depicts the major differences that are situated in floodplains and river channels where surface water is accumulated via routing, and the maximum depth difference is up to 3 meters in the river channel, especially downstream of Halls Bayou. However, the initial inundation timings and durations are scattered sparsely over the study area, with a majority of the grid cells showing earlier and longer inundations for the case without re-infiltration scenario. Therefore, it is likely that the flood is over-predicted by models without the re-infiltration scenario.

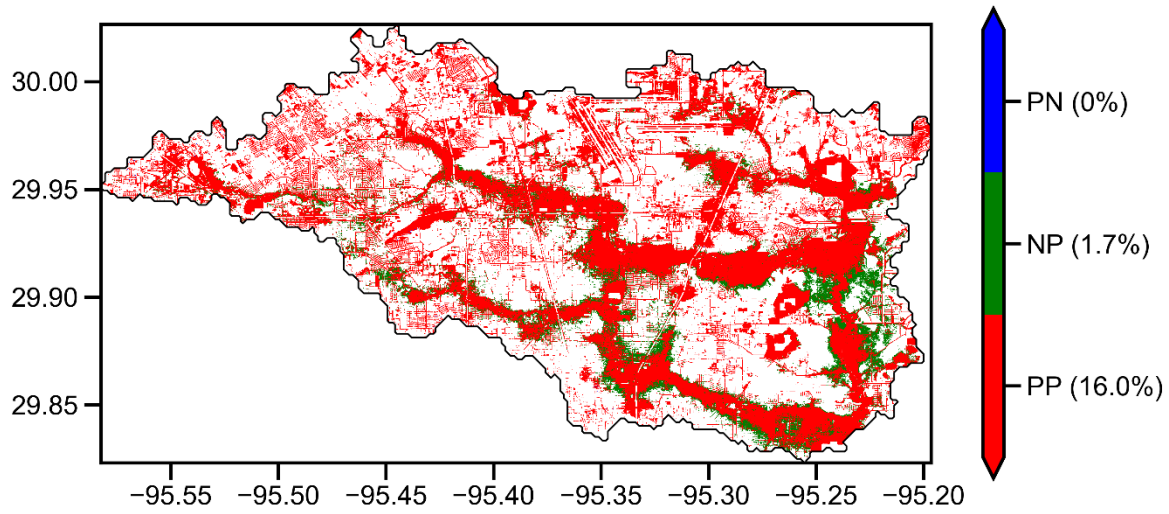


Figure 5.4 Map of binary flood detection comparison. PP: Positive Positives; PN: Positive Negatives; NP: Negative Positives.

Meanwhile, we notice that there are some samples in the opposite distribution, indicating delayed and/or shorter inundation time (Figure 5.5). Arguably, this could be some local effects when the soil reaches earlier saturation in the re-infiltration scenario, thereby leading to earlier flooding. A supporting material is found in Figure 5.5d, in which the basin-average soil moistures of two schemes are compared. Notably, evapotranspiration is not considered in this ideal test, so the soil moisture does not deplete with time. During the storm lifetime, soil moisture surges from completely dry to 85% saturation for the scenario without re-infiltration and to 95% saturation for scenario with re-infiltration. Early saturation reduces infiltration rates later on and thus has pronounced effects on local flooding when surface water is not routed timely. Figure 5e presents the evolution of surface water volume by integrating surface water depth along with grid cells. Although both scenarios concurrently reach the maximum surface water volume, their recession limbs show considerable differences. The re-infiltration scenario apparently has a steep exponential decay, as both still water and running water infiltrates into the soil; for the scenario without re-infiltration, in contrast, there is a mild decay and even levels out at the end of the simulation. The difference between the two increases with time, as shown in the shaded area, up to 0.4×10^8 m^3 volume difference, which equates to almost half of the total surface water volume.

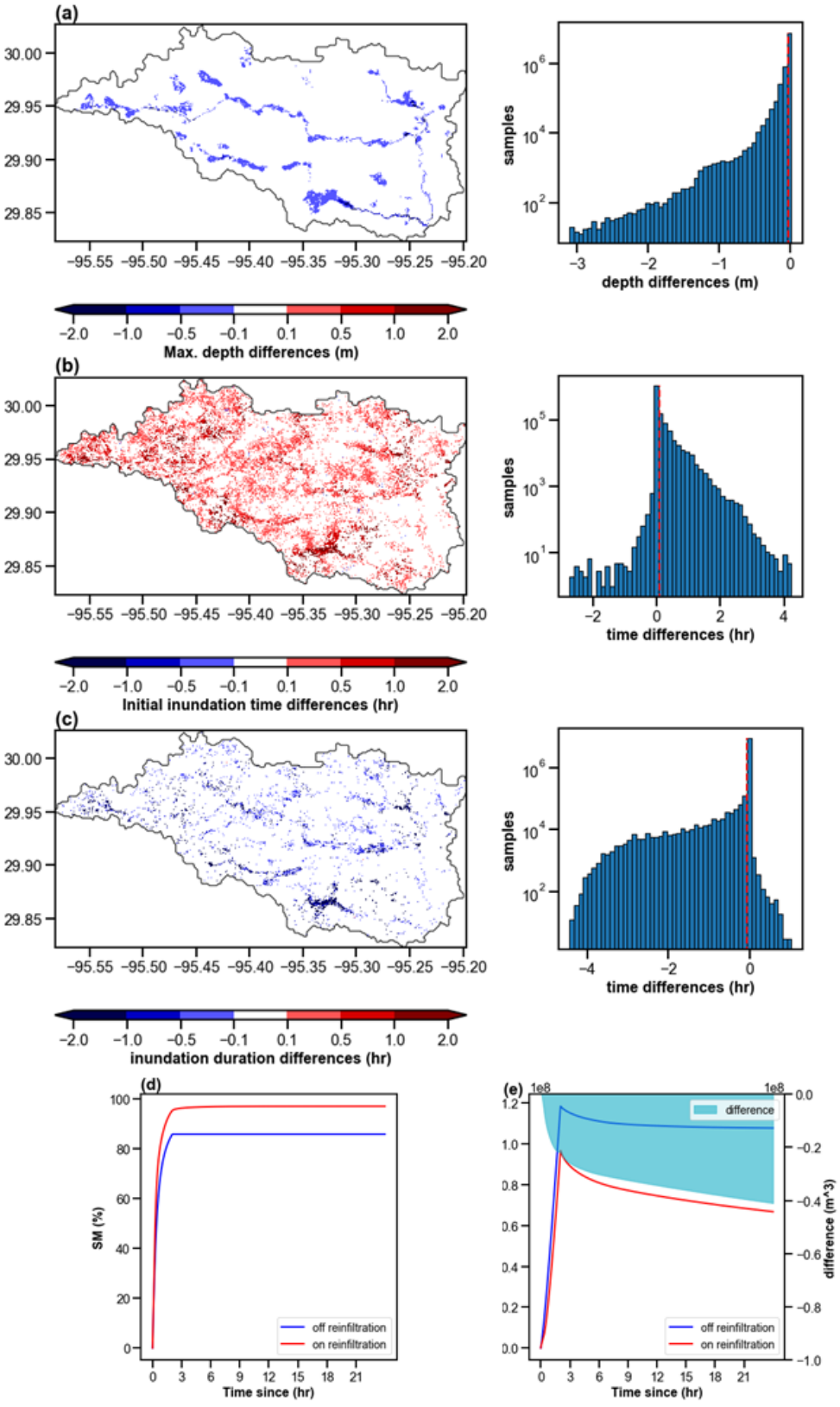

(a)
30.00
29.95
29.90
29.85
−95.55 −95.50 −95.45 −95.40 −95.35 −95.30 −95.25 −95.20
samples
depth differences (m)
−2.0 −1.0 −0.5 −0.1 0.1 0.5 1.0 2.0
Max. depth differences (m)
(b)
time differences (hr)
Initial inundation time differences (hr)
(c)
inundation duration differences (hr)
(d)
SM (%)
off reinfiltration
on reinfiltration
Time since (hr)
(e)
1e8
difference
difference (m^3)

Figure 5.5 Spatial distribution of differences of (a) maximum depth, (b) initial inundation timings, and (c) inundation durations along with respective sample distributions (red line represents the mean value of the distribution). Temporal evolution of (d) soil moisture (%) and (e) surface water volume. The difference of surface water volume in (e) is plotted in the shaded area.

In summary, re-infiltration scheme indeed influences flood magnitude and timings via surface water-soil interaction, and it possibly reduces flood magnitude and delays (shortens) flood timing (duration). Flood magnitude differences are pronounced downstream or in depressions, while flood timings are scattered. Such results are markedly tied to soil condition (wet or dry) and soil characteristics (infiltration capacity).

5.3.1.2 Sensitivity to storm frequency

For less intense storms yet more frequent, the impact of re-infiltration is expected to be larger than more intense storms, as rainfall rates are less likely to exceed infiltration capacity. We looked up 1-year, 2-year, 5-year, 10-year, 25-year, 50-year, and 100-year event rainfall from the NOAA Atlas 14 IDF curve and simulated those events collectively with default parameters. As shown in Figure 5.6, three flood magnitude-related metrics (i.e., R_F, R_H, and R_V) verify our speculation. For the 1-year storm specifically, flood areas with re-infiltration only account for 30% of that without re-infiltration. Likewise, flood depth and surface water volumes are reduced by 50% and 75%, respectively. However, there is no monotonic trend for flood timing. Floods are delayed for all storms, but the delayed time increases with return periods prior to peaking at 0.83 hours for 10-year storm, after which it decreases. The flood duration differences again decrease with return periods.

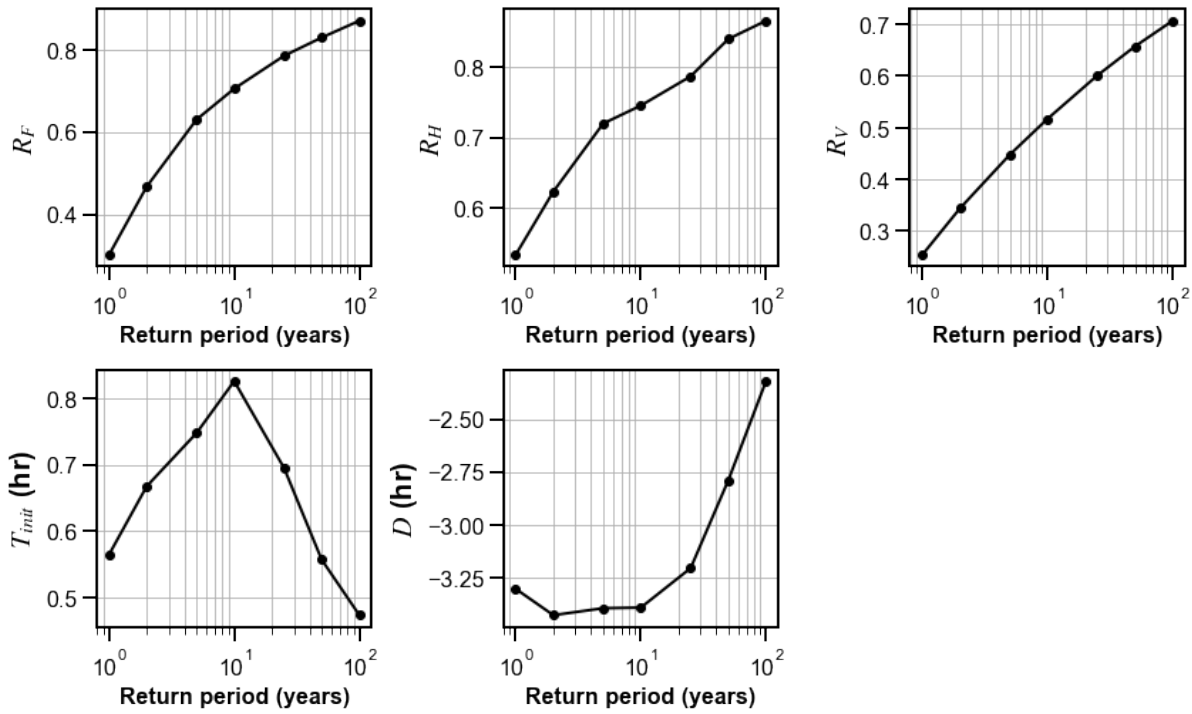


Figure 5.6 Similar to Figure 4, but for storms at different frequencies.

5.3.2 Case study: Hurricane Harvey

The previous sensitivity test indicates that the re-infiltration scheme is not only physically sound, but it exerts considerable influences on model simulations. Although the CREST-iMAP tested under theoretical scenarios, it is relevant to compare one another in a real case study against observations during Hurricane Harvey.

5.3.2.1 Verification against stream gauges

Gauged water heights from five USGS stream gauges within the model domain are retrieved during model simulations at the 15-min interval. Surface water levels from two simulation schemes are extracted at collocated stream gauge locations. It is worth mentioning that the terrain elevation imposes great uncertainties when comparing model simulations to observations, as the sub-grid variation cannot be resolved in the current settings. Despite the resolution mismatch, these gauge readings are still the most widely used source to verify the model performance. Table 5.4 shows the respective performance for with and without re-infiltration scenarios with respect to observations. The re-infiltration scheme greatly improves the NSE scores (+139.9%) and CC (+7.24%) while reducing RMSE (-18.2%). Especially for the gauge 08075900,

there is more than a 400% increase in NSE score, jumping from 0.12 to 0.69. By breaking down the RMSE into systematic error $RMSE_S$ and random error $RMSE_R$, we see the reduced errors are largely attributed to systematic error (-31.2%), relative to the random error (-13.1%). Thus, the systematic bias is much alleviated by considering the re-infiltration scheme.

The reason for such a performance leap comes from better characterization of its flow recession limbs, as shown in Figure 5.7. Both schemes are capable of simulating the peak water height values without delays, but water in the scenario without re-infiltration falls mildly in the recession stage, resulting in much higher water level than the observations. On the other hand, flow for the re-infiltration scenario follows the gauge readings closely, especially after the first peak (from 2017-08-26 to 2017-08-27). Apart from this best-performing gauge, the re-infiltration scheme improves capturing falling water across all the gauge stations, thereby leading to significant performance gains. Consistently, previous studies also highlighted that the re-infiltration markedly reduces recession limbs in the hydrograph (Nahar et al., 2004).

To be noted, water heights during flood recession period are both overestimated by two schemes, pointing to a systematic error in our CREST-iMAP framework. First, the model is calibrated to capture the peaking water height while ignoring the recession period. Second, surface runoff generated by the water balance model has been found to overestimate (Li et al., 2022b). Third, missing model physics such as subsurface exfiltration to channels and manmade structures complicate results interpretation.

Table 5.4 Model performance at stream gauge locations. The bolded values are the better ones from the off and on re-infiltration comparison. "Off" represents scenario without re-infiltration and "On" represents with re-infiltration scenario. NSE: Nash-Sutcliffe Coefficient; RMSE: Root Mean Squared Error; CC: Correlation Coefficient.

Metrics\Gauges		**08075900**	**08076000**	**08076180**	**08076500**	**08076700**
NSE	Off	0.12	0.20	0.47	0.34	0.921
	On	0.69	0.44	0.52	0.66	0.919
	Improvement (%)	+475	+120	+10.6	+94.1	+0.2
RMSE (m)	Off	1.89	2.24	1.85	1.61	1.14
	On	1.13	1.89	1.76	1.15	1.12
	Improvement (%)	-40.2	-15.6	-4.86	-28.6	-1.8

$RMSE_S$ (m)	Off	1.64	1.39	0.44	1.02	0.96
	On	1.00	1.02	0.26	0.59	0.89
	Improvement (%)	-39.0	-26.6	-40.9	-42.2	-7.3
$RMSE_R$ (m)	Off	0.94	1.77	1.80	1.25	0.61
	On	0.53	1.58	1.74	0.99	0.69
	Improvement (%)	-43.6	-10.7	-3.33	-20.8	+13.1
CC	Off	0.88	0.71	0.70	0.77	0.99
	On	0.96	0.78	0.73	0.87	0.99
	Improvement (%)	+9.09	+9.86	+4.29	+13.0	0

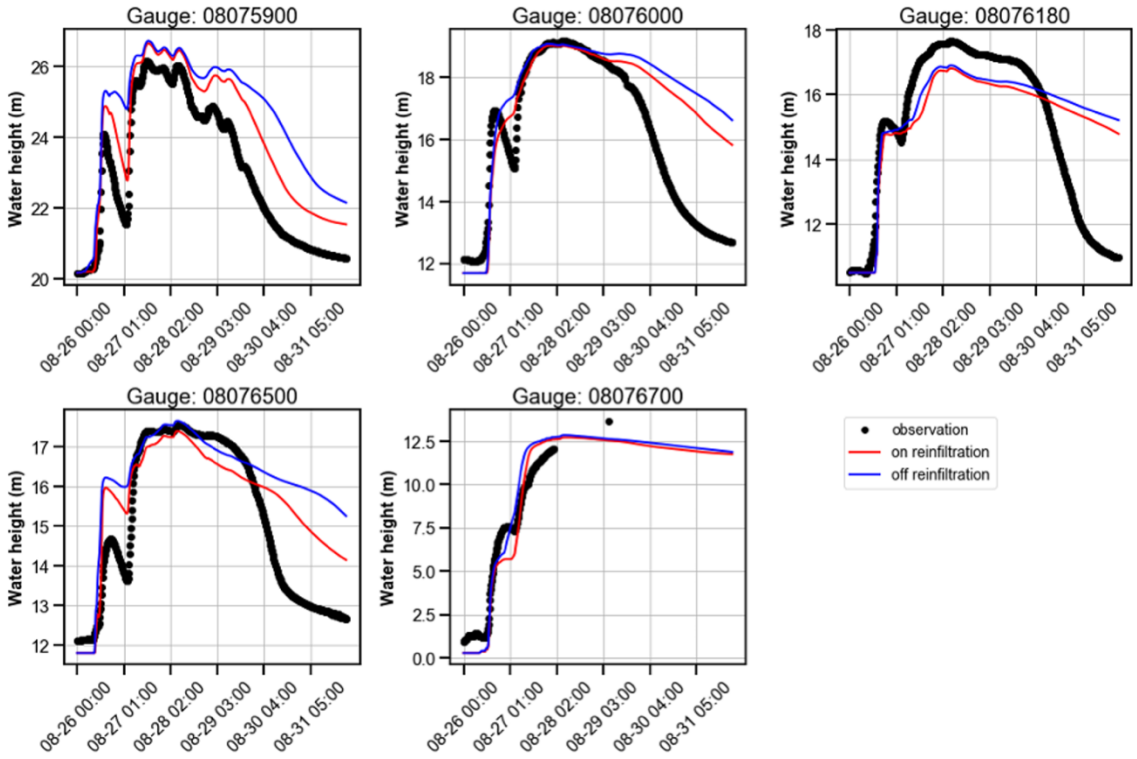


Figure 5.7 Simulated and observed time series of surface water level at five USGS stream gauges.

5.3.2.2 Verification against High Water Marks

Because a direct assessment for flood inundation is not feasible, some watermarks or stains in the aftermath of a flood event can be used as proxy data for model evaluation although with great uncertainties. The USGS team routinely publishes their surveyed High Water Marks

(HWMs) after some major flood events, which can subsequently be used for model evaluations (Chen et al., 2021; Li et al., 2021b; Sebastian et al., 2021; Wing et al., 2017). Figure 5.8 shows the cell-wise maximum flood depth of the two schemes compared to the HWMs. Both schemes present better performance upstream of Halls Bayou, with a difference smaller than 0.5 meters. However, the model over-predicts water depth in Greens Bayou up to 1.5 meters. This is consistent with the over-prediction of in-channel water level, as shown in Figure 5.7. The distribution of the differences is shown in Figure 5.8c, pointing to a generally better performance of the re-infiltration scenario than without it, as the absolute mean depth difference of the re-infiltration (0.51 m) is 17.2% smaller than that of the scenario without re-infiltration (0.60 m). It is worth noting that HWMs themselves come with uncertainties that are due to the data quality, and errors could be up to 0.2 meters (Koeig et al., 2016). For instance, tranquil water represents a smooth trend that has small uncertainties. There are also spurious errors that are related to human mistakes or values being rounded off. This is particularly true for the recorded geographical coordinates which requires more floating points to pin down the location exactly. In Figures 5.8d and e, the two HWMs are marked with an absolute difference greater than 1 m but only several pixels away (i.e., tens of meters) from their true values. Despite this, the re-infiltration greatly alleviates the over-prediction of the previous model.

5.3.2.3 Intercomparisons of flood magnitude and dynamics during Hurricane Harvey

The intercomparison of flood magnitude and dynamics helps to understand the effects of re-infiltration in a real 500-year event. Figure 5.9, similar to Figure 5.5, depicts the basin-integrated differences. For flood dynamics, the initial inundation timing (T_{init}) and total inundation duration (D) could vary from -2 (delayed) to 4 (earlier) hours and 0 to 15 hours, respectively. For the temporal evolution of the Harvey event, it is featured by two subsequent events. The first event from 2017-08-26 to 2017-08-27 saturates the soils immediately, during which the large differences of surface water volume and soil moisture are present between the two schemes. The soil moisture content for the re-infiltration scheme is about 10% more than without the re-infiltration scheme; the surface water volume, however, is 40% less. The second event does not produce a large difference because of the saturated soils over the domain (Figure 5.9d). As indicated by the sensitivity analysis, this effect is highly dependent on soil condition, soil types, and rainfall characteristics. The extreme rainfall from Harvey leaves less room for water to infiltrate, compared

to other less intense events. It is therefore expected to have more pronounced improvements for less intense rainfall or other regions with high soil infiltration capacity.

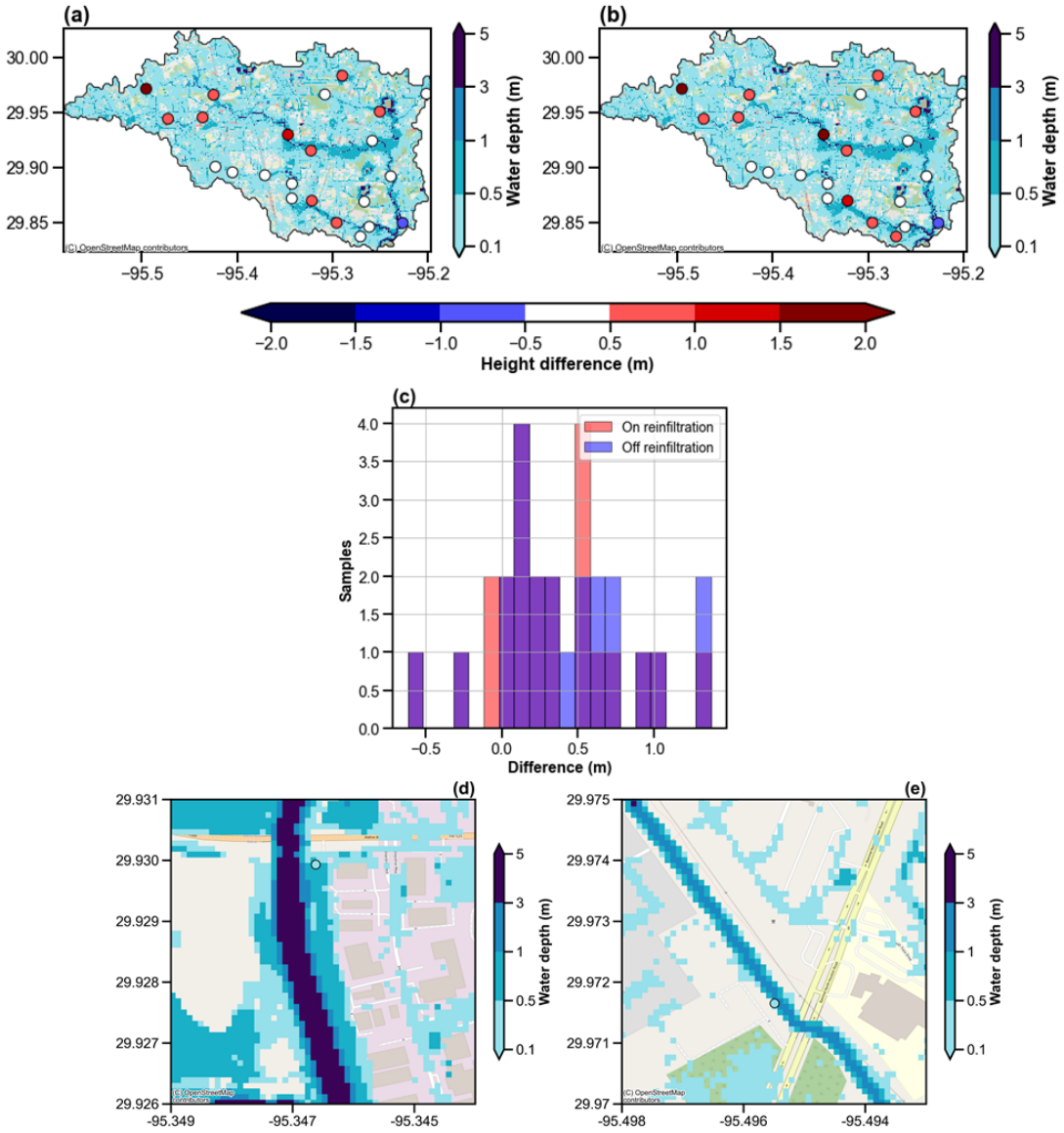


Figure 5.8 Maps of maximum surface water depth for (a) on re-infiltration and (b) off re-infiltration with differences against High Water Marks (HWMs). (c) histogram of water depth difference. Maps of two USGS high water marks (d) and (e), with the difference larger than one meter between the simulation and recorded.

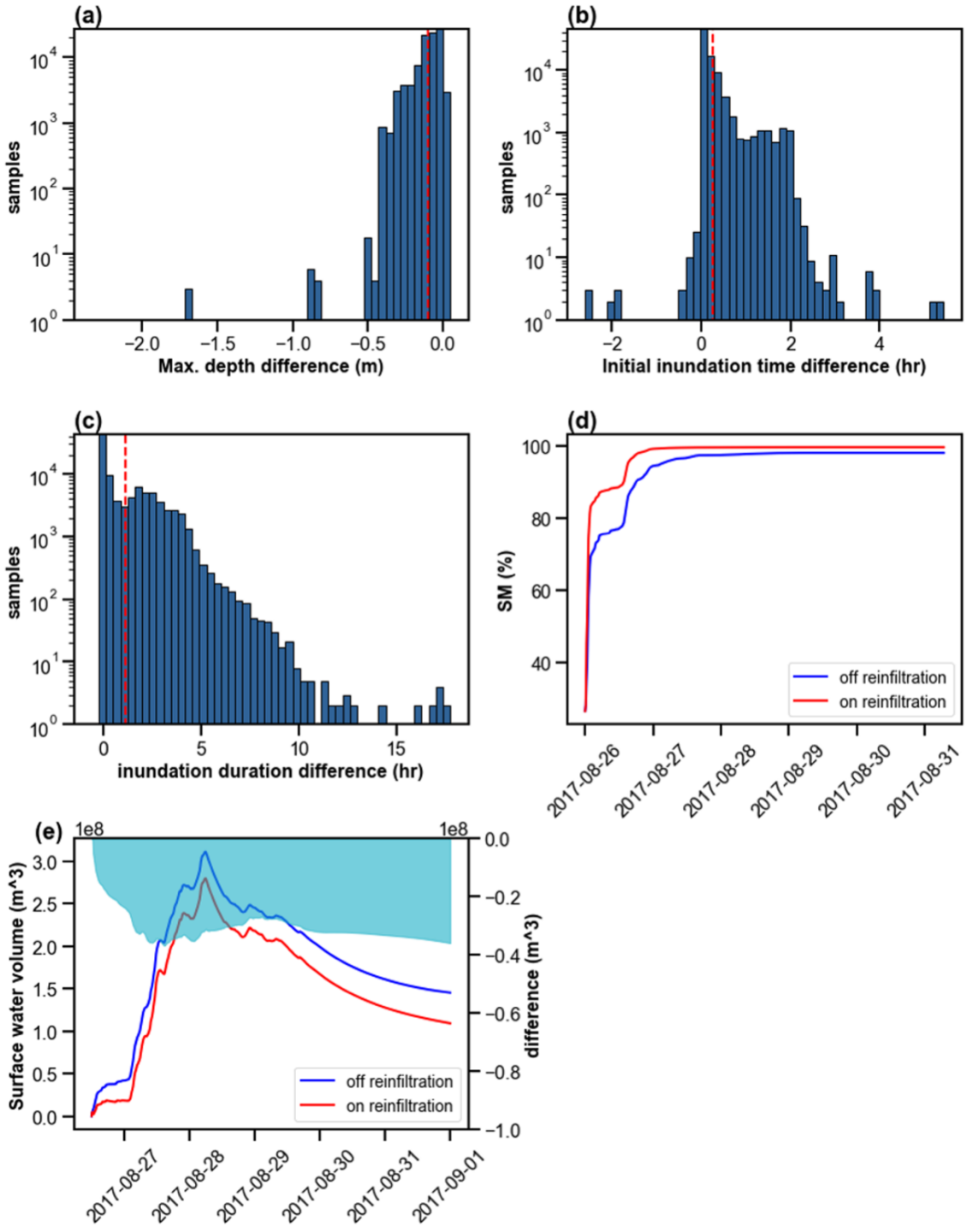


Figure 5.9 Basin aggregated distribution of (a) maximum depth differences, (b) initial inundation timing differences, and (c) inundation duration differences averaged over the simulation period. Time series of basin-average (a) soil moisture and (b) surface water volume. The vertical dashed line indicates the mean value of all samples.

5.4 Discussion

In this study, only local variations of four parameters – initial soil moisture, manning's roughness, hydraulic conductivity, and the exponent of the VIC model – are tested independently. However, the interactions among these parameters are not explored herein. Global sensitivity analysis, such as the Morris method used in the previous study (Li et al., 2021b), can measure the variation of each parameter relative to other parameters, so it provides a clearer picture of the parameter interactions. Needless to say, initial soil saturation state is the dominant controller for the differences between the simulations with and without re-infiltration process. When the soils are fully saturated, the with and without re-infiltration scenarios are almost identical if other parameters are the same. Combined with our previous study that underlies the importance of infiltration and initial soil moisture for flood inundation modeling, we highly recommend taking into consideration the initial soil moisture state, as it has not been well-recognized in the hydraulic model community. This can be achieved via three ways: 1) warm up the model for a relatively long period prior to the simulation period (Chen et al., 2020); 2) parameterize the initial soil moisture and calibrate it, similar to the way we treat initial in-channel water depth (Xue et al., 2013); 3) approximate it using observations or other model simulations, like what has been done in the real case study in Section 5.3.2 (Flamig et al., 2020). The first approach is ideal because it eliminates uncertainties in parameterization (such as equifinality) or error propagation from observations/simulations to models; it is, however, the most computationally expensive approach for hydraulic modeling compared to the other two. Approach two and three are more pragmatic, while both inherit uncertainties or errors. We prefer the third approach if the data source is found to be trustworthy. For instance, in our case study, we used the simulated soil moisture product from the operational CREST/EF5 model which shares the same land surface processes as the CREST-iMAP.

The results relating to different rainfall events are considered in this study (i.e., 1-year, 2-year, 5-year, 10-year, 25-year, 50-year, and 100-year rainfall). As for lower storm intensity, the differences between the two schemes enlarge, as rainfall rates are less likely to exceed the infiltration capacity, and soils are not saturated. Other environmental factors pertaining to different topography and physiography are also likely to interact and change the results. For instance, an increase in slope will leave less room for surface water to re-infiltrate, which explains why re-

infiltration compromises its importance in hillslope hydrology (Corradini et al., 2002; Zhang et al., 2020).

5.5 Conclusions

This study focuses on the influence of the re-infiltration process for 100-year and 500-year flood events, which has so far not been well-recognized by the hydrologic/hydraulic modeling community. The sensitivity experiment and a 500-year Hurricane Harvey example both highlight the discernable differences between the with and without re-infiltration scheme. The major conclusions are summarized as follows:

1. In the 100-year design rainfall event, re-infiltration is found to make discernible differences with less flood extent (~1.56 km2), depth (~3 m), and dynamics (~4-hour delayed flooding and ~4-hour shorter inundation duration), compared to without re-infiltration. The differences are increasing with more frequent storms. The 500-year Hurricane Harvey event shows a magnified difference in inundation duration up to 15 hours because of the longer event duration. However, the flood depth difference is less in the Harvey event due to the rapid saturation of the soils.

2. The hydraulic conductivity and antecedent soil condition from the designed sensitivity test are found to be the prime contributors to the difference between with and without re-infiltration, and comparatively, the antecedent soil moisture condition is the most sensitive among the four tested factors.

3. For the Harvey event, the differences are verified with stream gauge observations. On average, a 139.9% increase in NSE scores is found for re-infiltration with respect to without it. The improvements are mostly tied to better characterization of the recession limb after peak flow while the peak flows are well-captured by both. The proxy data – USGS High Water Marks – also indicate better performance with the inclusion of the re-infiltration scheme, as the re-infiltration scheme presents a 17.2% less flood depth difference than the case without the re-infiltration. The differences are further expected to enlarge for less intensive events and regions with a higher percentage of permeable soil media.

This study aims to raise attention to the important re-infiltration process in coupled H&H flood modeling to provide more accurate flood information, e.g., depth and timings. For future work, we will continue improving the current CREST-iMAP model framework by incorporating flood mitigation measures such as levees and dams into the system. Also, it is critical to couple

with the NWP model to advance flood prediction lead time, which ensures more time for residents at risk to evacuate.

5.6 Reference

Allen, R.G., Pereira, L., Raes, D., & Smith, M. (1998). Crop Evapotranspiration, Food and Agriculture Organization of the United Nations, Rome, Italy. FAO publication 56. ISBN 92-5-104219-5. 290p.

Bravo, J.M., Allasia, D., Paz, A.R., & Collischonn, W. (2012). Coupled Hydrologic-Hydraulic Modeling of the Upper Paraguay River Basin, Journal of Hydrological Engineering*, 17*, 635-646. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000494

Bates, P. (2004). Remote sensing and flood inundation modelling. Hydrological Processes, 18, 2593-2597.

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., Savage, J., Olcese, G., Neal, J., Schumann, G., Giustarini, L., Coxon, G., Porter, J. R., Amodeo, M. F., Chu, Z., Lewis-Gruss, S., Freeman, N. B., Houser, T., Delgado, M., Hamidi, A., Bolliger, I., McCusker, K. E., Emanuel, K., Ferreira, C. M., Khalid, A., Haigh, I. D., Couasnon, A., Kopp, R. E., Hsiang, S., & Krajewski W. F. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. Water Resources Research, 57, e2020WR028673. https://doi.org/10.1029/2020WR028673

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States, Geophysical Research Letters, *43*, pp.4382– 4390, doi:10.1002/2016GL068070.

Buchanan, B, Auerbach, DA, Knighton, J, Evensen, D., Fuka, D.R., Easton,Z., Wieczorek, M., Archibald, J. A., McWilliams, B., & Walter, T. (2018). Estimating dominant runoff modes across the conterminous United States. Hydrological Processes. *32*, 3881–3890. https://doi.org/10.1002/hyp.13296

Chen, M., Nabih, S., Brauer, N. S., Gao, S., Gourley, J. J., Hong, Z., Kolar R. L., & Hong, Y. (2020). Can Remote Sensing Technologies Capture the Extreme Precipitation Event and Its Cascading Hydrological Response? A Case Study of Hurricane Harvey Using EF5 Modeling Framework. Remote Sensing, 12, 445. https://doi.org/10.3390/rs12030445

Chen, M., Li, Z., Gao, S., Luo, X., Wing, O. E. J., Shen, X., Gourley, J. J., Kolar, R. L., & Hong, Y. (2021). A Comprehensive Flood Inundation Mapping for Hurricane Harvey Using an Integrated Hydrological and Hydraulic Model, Journal of Hydrometeorology, 22(7), 1713-1726.

Chen, M., Li, Z. & Gao, S. (2022). Multisensor Remote Sensing and the Multidimensional Modeling of Extreme Flood Events. In Remote Sensing of Water-Related Hazards (eds K. Zhang, Y. Hong and A. AghaKouchak). https://doi.org/10.1002/9781119159131.ch5

Chow, V.T., Maidment, D.R., & Mays L.W. (1988). *Applied Hydrology: McGraw-Hill Series in* Water Resources and Environmental Engineering, McGraw-Hill, Inc., New York.

Cohen, S., Praskievicz, S., & Maidment, D.R. (2018). Featured Collection Introduction: National Water Model. Journal of the American Water Resources Association*, 54*, 4 767– 769. https://doi.org/10.1111/1752-1688.12664.

Corradini, C., Morbidelli, R., & Melone, F. (1998). On the interaction between infiltration and Hortonian runoff, Journal of Hydrology, 204, pp.52-67.

Corradini, C., Govindaraju, R.S., & Morbidelli, R. (2002). Simplified modelling of areal average infiltration at the hillslope scale. Hydrological Processes, 16, 1757-1770.

Dullo, T. T., Ganguade, S., Morales-Hernandez, M., Sharif, M. B., Kao, S.C., Kalyanapu, A. J., Ghafoor, S., & Evans, K. J. (2021). Simulation of Hurricane Harvey flood event through coupled hydrologic-hydraulic models: challenges and next steps. Journal of Flood Risk Management, 14, 1-26. https://doi.org/10.1111/jfr3.12716

Felder, G., Zischg, A., & Weingartner, R. (2017). The effect of coupling hydrologic and hydrodynamic models on probable maximum flood estimation, Journal of Hydrology, 550, 157-165. https://doi.org/10.1016/j.jhydrol.2017.04.052

Flamig, Z. L., Vergara, H., & Gourley, J. J. (2020). The Ensemble Framework For Flash Flood Forecasting (EF5) v1.2: description and case study. Geoscientific Model Development, 13, 4943-4958. https://doi.org/10.5194/gmd-13-4943-2020

Freeze, R.A., 1974. Streamflow generation. Review of Geophysics. 12, 627-647. https://doi.org/10.1029/RG012i004p00627

Gourley, J. J., Flamig, Z.L., Vergara, H., Kirstetter, P., Clark, R.A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., & Yang, H. (2017). The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. Bulletin of American Meteorological Society, 98, 361–372. https://doi.org/10.1175/BAMS-D-15-00247.1

Getirana, A. C. V., & Paiva R. C. D. (2013). Mapping large-scale river flow hydraulics in the Amazon basin, Water Resources Research, 49, 2437-2445. https://doi.org/10.1002/wrcr.20212

Grimaldi, S., Schumann, G. J.-P., Shokri, A., Walker, J. P., & Pauwels, V. R. N. (2019). Challenges, opportunities and pitfalls for global coupled hydrologic-hydraulic modeling of floods. Water Resources Research, 55, 5277– 5300. https://doi.org/10.1029/2018WR024289

Hirabayashi, Y., Mahendran, R., Koirala, S. Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S. (2013). Global flood risk under climate change. Nature Climate Change, 3, 816–821 (2013). https://doi.org/10.1038/nclimate1911

Johnson, J.M., Munasinghe, D., Eyelade, D., & Cohen, S. (2019). An integrated evaluation of the National Water Model (NWM)-Height Above Nearest Drainage (HAND) flood mapping methodology, Natural Hazards Earth System Sciences, 19, 2405-2420. https://doi.org/10.5194/nhess-19-2405-2019

Kenward, T., Lettenmaier, D.P., Wood, E.F., & Fielding, E. (2000). Effects of Digital Elevation Model Accuracy on Hydrologic Predictions. Remote Sensing of Environment, 74, 3, 432-444. https://doi.org/10.1016/S0034-4257(00)00136-X.

Kim, J., Warnock, A., Ivanov, V.Y., & Katopodes, N.D., 2012. Coupled modeling of hydrologic and hydrodynamic processes including overland and channel flow. Advances in Water Resources, 37, 104-126. https://doi.org/10.1016/j.advwatres.2011.11.009

Koenig, T.A., Bruce, J.L., O'Connor, J.E., McGee, B.D., Holmes, R.R., Jr., Hollins, Ryan, Forbes, B. T., Kohn, M. S., Schellekens, M. F., Martin, Z. W., & Peppler, M.C., 2016, Identifying and preserving high-water mark data: U.S. Geological Survey Techniques and Methods, book 3, chap. A24, p.47, http://dx.doi.org/10.3133/tm3A24.

Li, Z., Chen, M., Gao, S., Hong, Z., Tang, G., Wen, Y., Gourley J.J., & Hong, Y. (2020). Cross-Examination of Similarity, Difference and Deficiency of Gauge, Radar and Satellite Precipitation Measuring Uncertainties for Extreme Events Using Conventional Metrics and Multiplicative Triple Collocation. Remote Sensing, 12, 1258. https://doi.org/10.3390/rs12081258

Li, Z., Chen, M., Gao, S., Gourley, J. J., Yang, T., Shen, X., Kolar, R., & Hong, Y. (2021a). A multi-source 120-year US flood database with a unified common format and public access, Earth System Science Data, 13, 3755–3766, https://doi.org/10.5194/essd-13-3755-2021.

Li, Z., Chen, M., Gao, S., Luo, X., Gourley, J. J., Kirstetter, P., Yang, T., Kolar, R., McGovern, A., Wen, Y., Hong, Y. (2021b). CREST-iMAP v1. 0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction. Environmental Modelling & Software, 141, 105051.

Li, Z., Gao, S., Chen, M., Gourley, J.J., Liu, C., Prein, A., & Hong, Y. (2022a). The conterminous United States are projected to become more prone to flash floods in a high-end emissions scenario. Commun Earth Environ 3, 86. https://doi.org/10.1038/s43247-022-00409-6

Li, Z., Gao, S., Chen, M., Gourley, J., Mizukami, N., & Hong, Y. (2022b). CREST-VEC: A framework towards more accurate and realistic flood simulation across scales, Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2022-61, in review.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research and Atmosphere, 99, 14415-14428. https://doi.org/10.1029/94JD00483.

Mahapatra, S., Jha, M.K., Biswal, S., & Senapati, D. (2020). Assessing Variability of Infiltration Characteristics and Reliability of Infiltration Models in a Tropical Sub-humid Region of India. Scientific Reports, 10, 1515. https://doi.org/10.1038/s41598-020-58333-8

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article "Assessment of economic flood damage", Natural Hazards and Earth System Sciences, 10, pp.1697–1724, https://doi.org/10.5194/nhess-10-1697-2010, 2010.

Mohanty, P. M., Nithya, S., Nair, A. S., Indu, J., Ghosh, S., Bhatt, C. M., Rao, G. S., Karmakar, S., & Karmakar, S. (2020). Sensitivity of various topographic data in flood management: Implications on inundation mapping over large data-scarce regions. Journal of Hydrology, 590, 125523, https://doi.org/10.1016/j.jhydrol.2020.125523.

Nahar, N., Govindaraju, R.S., Corradini, C., & Morbidelli, R., 2004. Role of run-on for describing field-scale infiltration and overland flow over spatially variable soils, Journal of Hydrology, 286, pp.36-52, 10.1016/j.jhydrol.2003.09.011.

Nguyen, P., Thorstensen, A., Sorooshian, S., Hsu, K., AghaKouchak, A., Sanders, B., …, & Smith, M., 2016. A high resolution coupled hydrologic-hydraulic model (HiResFlood-UCI) for flash flood modeling. Journal of Hydrology, 541, pp.401-420. https://doi.org/10.1016/j.jhydrol.2015.10.047

Ni, Y., Cao, Z., Liu, Q., & Liu Q., 2020. A 2D hydrodynamic model for shallow water flows with significant infiltration losses. Hydrological Processes, 34. https://doi.org/10.1002/hyp.13722

Pontes, P.R.M., Fan, F.M., Fleischmann, A.S., de Paiva, R.C.D., Buarque, D.C., Siqueira, V.A., …, & Collischonn, W., 2017. MGB-IPH model for hydrological and hydraulic simulation of large floodplain rivers systems coupled with open source GIS, Environmental Modelling and Software, 94, pp.1-20, https://doi.org/10.1016/j.envsoft.2017.03.029

Saksena, S., Dey, S., Merwade, V., & Singhofen, P.J., 2020. A computationally efficient and physically based approach for urban flood modeling using a flexible spatiotemporal structure. Water Resources Research, 56, p.e2019WR025769. https://doi.org/10.1029/2019WR025769

Sampson, C.C., Smith, A.M., Bates, P., Neal, J.C., Alfieri, L., & Freer, J.E., 2015. A high-resolution global flood hazard model. Water Resources Research, 51, pp.7358-7381, 10.1002/2015WR016954.

Sanders, B., 2007. Evaluation of on-line DEMs for flood inundation modelling, Advances in Water Resources, 30, pp.1831-1843, 10.1016/j.advwatres.2007.02.005.

Schumann, G.J.P. & Bates P.D. (2018) The Need for a High-Accuracy, Open-Access Global DEM. Frontier Earth Science, 6, 225. doi: 10.3389/feart.2018.00225

Sebastian, A., Bader, D.J., Nederhiff, C.M., Leijnse, T.W.B., Bicker, J.D., & Aarninkhof, S.G., 2021. Hindcast of pluvial, fluvial, and coastal flood damage in Houston, Texas during Hurricane Harvey (2017) using SFINCS. Natural Hazards. https://doi.org/10.1007/s11069-021-04922-3

Smith, R. E., and Hebbert, R. H. B. (1979), A Monte Carlo Analysis of the hydrologic effects of spatial variability of infiltration, Water Resources Research, 15, 2, pp.419– 429, https://doi.org/10.1029/WR015i002p00419.

Swain, D. L., Wing, O. E. J., Bates, P. D., Done, J. M., Johnson, K. A., & Cameron, D. R., 2020. Increased flood exposure due to climate change and population growth in the United States. Earth's Future, 8, p.e2020EF001778. https://doi.org/10.1029/2020EF001778

Tabari, H., 2020. Climate change impact on flood and extreme precipitation increases with water availability. Scientific Reports, 10, 13768. https://doi.org/10.1038/s41598-020-70816-2

Tang, G., Clark, M.P., Papalexiou, S.M., Ma, Z., & Hong, Y., 2020. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with

nine satellite and reanalysis datasets. Remote Sensing and Environment, 240, 111697. https://doi.org/10.1016/j.rse.2020.111697

Teng, J., Jakeman, A.J., Vaze, J., Croke, B.F.W., Dutta, D., & Kim, S., 2017. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis, Environmental Modeling and Software, 90, pp.201-216. https://doi.org/10.1016/j.envsoft.2017.01.006

Triet, N.V.K., Dung, N.V., Hoang, L.P., Duy, N.L., Tran, D.D., Anh, T.T., …, & Apel, H., 2020, Future projections of flood dynamics in the Vietnamese Mekong Delta, Science of The Total Environment, 742, 140596, https://doi.org/10.1016/j.scitotenv.2020.140596.

Viero, D.P., Rober G., Matticchio, B., Define, A., & Tarolli, P., 2019. Floods, landscape modifications and population dynamics in anthropogenic coastal lowlands: The Polesine (northern Italy) case study. Science of The Total Environment, 651, pp.1435-1450. https://doi.org/10.1016/j.scitotenv.2018.09.121

Viterbo, F., Mahoney, K., Read, L., Salas, F., Bates, B., Elliott, J., …, & Cifelli, R., 2020. A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland, Flood, Journal of Hydrometeorology, 21, 3, pp.475-499. https://doi.org/10.1175/JHM-D-19-0125.1

Wing, O. E. J., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., & Erickson, T. A. , 2017. Validation of a 30 m resolution flood hazard model of the conterminous United States, Water Resources Research, 53, pp.7968– 7986, doi:10.1002/2017WR020917.

Woolhiser, D.A., Smith, R.E., & Giraldez, J.V., 1996. Effects of Spatial Variability of Saturated Hydraulic Conductivity on Hortonian Overland Flood, Water Resources Research, 32, pp.671-678, https://doi.org/10.1029/95WR03108

Xue, X., Hong, Y., Limaye, A.S., Gourley, J.J., Huffman, G.J., Khan, S.I., Doriji, C., & Chen, S., 2013. Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? Journal of Hydrology, 499, pp.91-99. https://doi.org/10.1016/j.hydrol.2013.06.042.

Yamazaki, D., Kanae, S., Kim, H., & Oki, T., 2011. A physically based description of floodplain inundation dynamics in a global river routing model, Water Resources Research, 47, W04501, https://doi.org/10.1029/2010WR009726.

Yang, G., Bowling, L.C., Cherkauer, K.A., & Pijanowski, B.C., 2011. The impact of urban development on hydrologic regime from catchment to basin scales, Landscape and Urban Planning, 103, 2, pp.237-249. https://doi.org/10.1016/j.landurbplan.2011.08.003

Yussouf, N., Wilson, K. A., Martinaitis, S. M., Vergara, H., Heinselman, P. L., & Gourley, J. J., 2020. The Coupling of NSSL Warn-on-Forecast and FLASH Systems for Probabilistic Flash Flood Prediction, Journal of Hydrometeorology, 21, 1, pp.123-141. https://doi.org/10.1175/JHM-D-19-0131.1

Zhao, R. J., 1995. The xinanjiang model. Computer models of watershed hydrology, pp.215-232

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., & Kitzmiller, D., 2016. Multi-Radar Multi-Sensor (MRMS) Quantitative Precipitation Estimation: Initial Operating Capabilities. Bulletin of American Meteorological Society, 97, 4, pp.621-638. https://doi.org/10.1175/bams-d-14-00174.1

Zhang J., Lin, P., Gao, S., & Fang, Z., 2020. Understanding the re-infiltration process to simulating streamflow in North Central Texas using the WRF-hydro modeling system, Journal of Hydrology, 587, 124902, https://doi.org/10.1016/j.jhydrol.2020.124902.

Chapter 6

CREST-VEC: Vector-based hydrologic routing framework for large-scale flood simulation

Publication

Li, Z., Gao, S., Chen, M., Gourley, J., Mizukami, N., and Hong, Y. (2022). CREST-VEC: A framework towards more accurate and realistic flood simulation across scales, Geoscientific. Model development, Discuss. [preprint], https://doi.org/10.5194/gmd-2022-61, in review, 2022.

6 CREST-VEC: Vector-based hydrologic routing framework for large-scale flood simulation

Abstract

Large-scale (i.e., continental and global) hydrologic simulation is an appealing yet challenging topic for the hydrologic community. First and foremost, model efficiency and scalability (flexibility in resolution and discretization) have to be prioritized. Then, sufficient model accuracy and precision are required to provide useful information for water resource applications. Towards this goal, we craft two objectives for improving US current operational hydrological model: (1) vectorized routing and (2) improved hydrological processes. This study presents a hydrologic modeling framework CREST-VEC that combines a gridded water balance model and a newly developed vector-based routing scheme. First, in contrast to a conventional fully gridded model, this framework can significantly reduce the computational cost of river routing by at least ten times, based on experiments at regional (0.07 sec/step vs. 0.002 sec/step) and continental scales (0.35 sec/step vs. 7.2 sec/step). This provides adequate time efficiency for generating operational ensemble streamflow forecasts and even probabilistic estimates across scales. Second, the performance using the new vector-based routing is improved, with the median-aggregated NSE (Nash-Sutcliffe Efficiency) score increasing from -0.06 to 0.18 over the CONUS. Third, with the lake module incorporated, the NSE score is further improved by 56.2%, and the systematic bias is reduced by 17%. Lastly, over 20% of the false alarms on two-year floods in the US can be mitigated with the lake module enabled, at the expense of only missing 2.3% more events. This study demonstrated the advantages of the proposed hydrological modeling framework that could provide a solid basis for continental and global scale water modeling at fine resolution. Furthermore, the use of ensemble forecasts can be incorporated into this framework; and thus, optimized streamflow prediction with quantified uncertainty information can be achieved in an operational fashion for stakeholders and decision-makers.

6.1 Introduction

Flooding all over the world has affected millions of people, especially those who reside in floodplains (Tellman et al., 2021). In the US, flooding, as the primary cause of billion-dollar weather disasters, costs $3.9 billion monetary losses and 15 deaths per year over the past four decades according to the NOAA National Centers for Environmental Information U.S. Billion-Dollar Weather and Climate Disasters (2021). In light of frequent flooding in the US, several public agencies have been operating real-time flood forecasting systems, such as the NOAA NSSL (National Severe Storms Laboratory) FLASH project (https://flash.ou.edu) and NOAA Office of Water Prediction (OWP). However, flood warnings are still either missed or unverified due to uncertainties ranging from precipitation forcing, hydrologic model structure, model parameterization, and/or hydrologic routing. As revealed by Martinaitis et al. (2017), 12.8% of flash floods in the US go completely unwarned per year, let alone falsely warned. Apart from pursuing accurate weather forecasts, improving hydrologic simulation is the key to issuing flood warnings properly.

Flow routing in hydrology is the lateral transport of water on the land surface, subsurface, and in waterways (namely hydrologic compartments). It is an inseparable component in hydrologic simulation to redistribute and exchange water between compartments and is also relatively time-consuming. In a lumped hydrologic model (watershed as an integrated unit), routing can be simplified to convolution in time, such as the Unit Hydrograph (UH) or referred to Impulse Response Function (IRF) (Chow, 1988). However, variable velocities over the land surface and in waterways are difficult to be physically considered. Parameterization is a pragmatic way, but too many parameters could lead to equifinality (Beven, 2006). In addition, only outlet streamflow can be simulated in a lumped model. A semi-distributed model was then born to resolve flow pathways using Digital Elevation Models (Quinn et al., 1991). Owing to ever-increasing computing power, gridded hydrologic models with spatially distributed routing become feasible over large domains (Shaad, 2018). Terrain (or hillslope) routing and river channel routing at grid scales can be explicitly represented in model settings with distributed solvers such as linear reservoirs (Liston et al., 1994; Wang et al., 2011; Shen et al., 2017), kinematic wave model (Vergara et al., 2016), and diffusive wave model (Lighthill & Whitham, 1955; Ponce et al., 1978; de Almeida & Bates, 2013). More recently, vector-based routing has attracted more attention instead of raster-based routing for large-scale (i.e., continental and global extent) simulation. In theory, vector-based

routing and raster-based routing differ in defining unit catchments and river networks. For instance, a raster-based routing model discretizes both catchments and river networks on Cartesian coordinates, while a vector-based routing model builds upon the irregular shape of unit catchments (i.e., polygon) and river networks (i.e., polyline).

The pioneering experiment of vector-based routing can be dated back to the early 2000s, in which river network models were incorporated in emerging Geographic Information System (GIS) software (Wang et al., 2000). With the burgeoning availability of global-scale hydrography datasets (e.g., HydroSHEDS and NHDPlus), vector-based routing models have been gaining considerable interest in recent years (David et al., 2011; Lehner & Grill, 2013; Mizukami et al., 2016; Paiva et al., 2011; Yamazaki et al., 2011). Among those developments, three frameworks have become popular and stand out in the hydrologic model community. First, David et al. (2011) introduced the RAPID routing framework that is based on the Muskingum method. The RAPID has been coupled with the National Water Model operated by the NOAA OWP (Office of Water Prediction) (Lin et al., 2018) and the Global Flood Awareness System (GLoFAS) developed by the ECWMF (European Center for Median-Range Weather Forecasts). Second, Yamazaki et al. (2011) developed the CaMa-Flood framework which generates flood inundation at a large scale by solving the 1D diffusive equation and spilling water over floodplains. Third, the recent development of the mizuRoute framework by Mizukami et al. (2016) offers terrain routing and multiple channel routing schemes (e.g., IRF and kinematic wave), making it more physically based compared to RAPID which ignores terrain routing. The mizuRoute has been used together with the hydrologic framework SUMMA (Structure for Unifying Multiple Model Alternatives) (Knoben et al., 2021) and is planned to be implemented in the CESM (Community Earth System Model). These vector-based routing models overcome several challenges for large-scale hydrologic simulations faced by raster-based routing models. First, higher model resolution in raster-based models comes at the expense of higher computational cost, which prohibits global hydrological simulations at tens or hundreds of meters. However, the vector-based routing model is much more scalable and computationally efficient, irrespective of increasing resolution. Second, river networks can be more realistically represented in a vector form. In conventional hydrologic models, the river network in a raster form has to be delineated based on a DEM as a preprocessing step. River networks generated in such a way do not always align well with natural river centerlines. For studies investigating hydrologic connectivity in particular, river grid cells in a raster form can

easily become discontinuous without considering river topology. Alternatively, river networks in popular hydrography data are digitalized based on satellite optical imagery and manual inspection (Lin et al., 2021). Another weakness of raster-based routing stems from the traditional D8 flow strategy which means water in the central grid can only be permitted to flow through one of its neighboring grid cells (Tarboton, 1997). On the contrary, vector-based routing offers a more flexible approach from vector representation of river networks.

To date, modern vector-based routing models such as RAPID and mizuRoute have neglected the subsurface routing, which is either assumed to be minimum (Mizukami et al., 2020) or treated the way same as surface routing (Lin et al., 2019; Yang et al., 2021). However, subsurface routing is an important hydrologic process and dominates over regions that have intermittent flow behaviors (Freeze, 1972). For flood simulation, ignoring subsurface routing could underestimate the peak flow and miscalculate the flood timing, both of which directly affect decision-making processes. An equally important research thrust is the representation of lakes and reservoirs in vector formats, since they markedly alter flow response not only at a local scale, but also downstream rivers. One of the functions of lakes and reservoirs in the US is for flood control, so simulation without incorporating such a process is likely to result in falsely issued flood warnings.

In light of the advantages of vector-based routing, this study introduces a coupled modeling framework CREST-VEC (Coupled Routing and Excess STorage with VECtor routing), which strives to facilitate real-time flood forecasting across scales. This framework seamlessly integrates the current operational flash flood forecast model structure – CREST model and the vector-based routing framework – mizuRoute. We utilize a case study to demonstrate the advantages of this coupled framework and to investigate some updates we made to improve the existing routing scheme. Four questions are posed in this regional case study: (1) What are the performance gains for CREST-VEC compared to the CREST model? (2) Does the included subsurface routing improve model performance? (3) Can a simple natural lake simulation improve model performance in a downstream urban area? and (4) How does CREST-VEC model adopt to flood warnings? In the second part, we apply this framework to the continental US for a comprehensive evaluation. We ask one additional question: How many floods are falsely alarmed without considering reservoir operations? It is anticipated that findings from this work could motivate the

future development of large-scale hydrologic models and raise awareness on whether and how much flood forecasts by model simulations should be trusted without proper representation of lakes.

6.2 Data and methods

6.2.1 Hydrography data

In this study, we use the vectorized river network and Hydrologic Response Unit (HRU) dataset derived from the high-accuracy Multi-Error-Removed-Improved-Terrain (MERIT) Hydro hydrography dataset (Yamazaki et al., 2017, 2019). The flowlines were created from the 90-meter DEM data (MERIT DEM), covering the full global land surface (60°S-90°N). A minimum channelization threshold of 25 km^2 (upstream area) was applied to restrict river channel grid cells in the MERIT Hydro dataset. The HRUs were processed along with flowlines by the TauDEM software and trimmed with the HydroBASINS level-II boundaries. Detailed processing of the hydrography data is listed in Lin et al. (2019). This set of hydrography data has been validated against 30-year Landsat imagery (Lin et al., 2021) and empowered the global reconstruction of historical streamflow (Lin et al., 2019; Yang et al., 2021). Over the CONUS, we have obtained 341,921 river reaches and the same amount of unit catchments for the routing component.

Lakes and reservoirs in the U.S. play a significant role in regulating streamflow (Tavakoly et al., 2021). Major river basins (e.g., Mississippi and Columbia River Basins) are highly regulated, as shown in Fig.6.1a and 6.1b, results obtained from Lehner et al. (2011). The HydroLAKES dataset provides a global catalog of lake polygons and pour points that can be easily integrated into hydrologic models (Messager et al., 2016). Over one million natural lakes and constructed reservoirs were identified globally, with a minimum surface area larger than 10 ha. Over the U.S., there are 96,874 lakes recorded in the HydroLAKES data, of which 94,865 are natural lakes without human intervention, and 1,992 (17) lakes are reservoirs (regulated lakes), as shown in Fig.6.1c. Of regulated lakes or reservoirs, 20.0% are primarily used for irrigation, 19.9% for hydroelectricity, 17.6% for water supply, 17.2% for flood control, 14.1% for recreation, 1.9% for navigation, 0.7 for fisheries, and 8.6% for others (Fig.1d). The total lake volume, estimated from the lake bathymetry, is a required field in our modeling framework to approximate outflow.

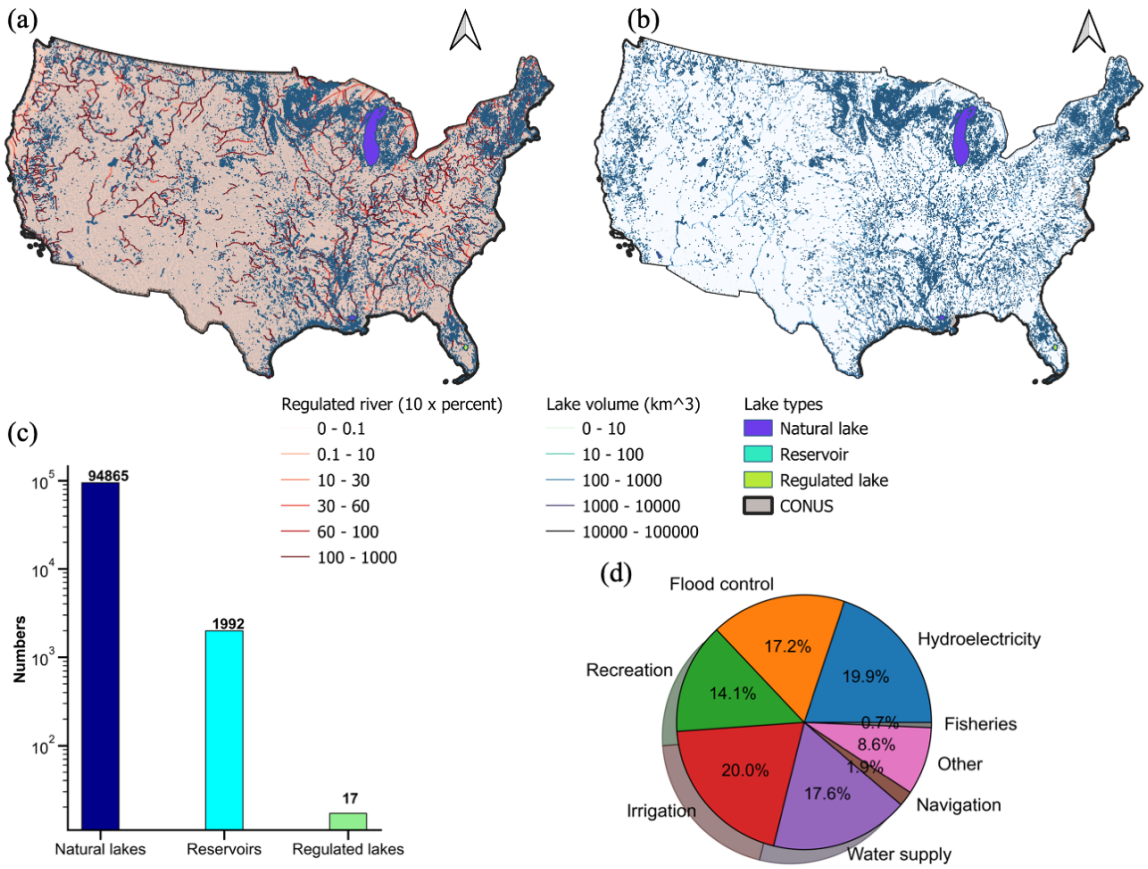


Figure 6.1 Maps of (a) percent of regulated river and (b) regulated lake volume; (c) bar plot of lake classifications; (d) pie plot of US regulated lake or reservoir function purposes.

6.2.2 Forcing data

Forcing data is required as model inputs to drive the hydrologic model. Hourly precipitation rates are obtained from the Multi-Radar Multi-Sensor (MRMS) data, operated at the NOAA NSSL (Zhang et al., 2016). The MRMS is a state-of-the-art radar-gauge merged product, providing instantaneous rates at a 1-km spatial resolution over the CONUS and parts of southern Canada and northern Mexico. We used the one-hour accumulated and gauge-corrected precipitation product in this study for streamflow simulation. The performance and hydrologic utility of MRMS data has been corroborated in previous studies (Li et al., 2020, 2021). The daily temperature from the PRISM (Parameter-elevation Relationships on Independent Slopes Model) is used to simulate snow accumulation and melt (PRISM Climate Group, 2014). The PRISM team routinely collects meteorological data from meteorological stations over the U.S. and interpolates them into 4-km gridded data based on the elevation dependence (Daly et al., 2008). The potential

evapotranspiration (PET) data is obtained from the USGS FEWS data port (https://earlywarning.usgs.gov/fews) at daily and 1° spatial resolution (Allen et al., 1998). Forcing data at different spatial resolutions is re-gridded to a 1-km model resolution. All of these data are collected from the simulation period in complete calendar days from 2015 to 2019.

6.2.3 CREST model

As jointly developed by the University of Oklahoma (OU) and NASA, the CREST model has been released for a decade (Wang et al., 2011). It is a distributed hydrologic model whose primary purposes are (1) flood simulation and forecasting, (2) evaluating the hydrologic utility of satellite precipitation datasets, and (3) water resources management (Xue et al., 2013; Tang et al., 2016; Gourley et al., 2017; Gao et al., 2021; Li et al., 2021; Chen et al., 2022). Owing to its relatively simple structure and computationally efficient simulation, the CREST model has been promoted by the NOAA NSSL for real-time flash flood forecasting over the continental U.S. and its territories (Gourley et al., 2017; Flamig et al., 2020). As shown in Fig. 6.2, the effective rain (deficit of rainfall rates and evaporation rates) reaches the land surface and is partitioned into fast runoff from urban impervious area ratio and infiltration into the soils. A Variable Infiltration Curve (VIC) model is incorporated to determine the infiltration rate (Liang et al., 1994). Surface runoff is generated when infiltration rates become higher than the maximum infiltration capacity. In the meantime, slow-flowing interflow is produced while soil water content is depleted. In the CREST model, flow routing is handled in two ways. Terrain routing and in-channel river routing are done by the kinematic wave model which simplifies the Saint-Venant equation by ignoring the acceleration and forcing terms (Vergara et al., 2017). The interflow is routed by a conceptual linear reservoir with parameterized velocity (Shen et al., 2017). We refer to the CREST model hereafter as a standalone package that couples the water balance model with gridded terrain and channel routing. The original code is written in C++.

To account for snowmelt, we coupled the original CREST model with the Snow-17 model, which is part of the National Weather Service River Forecast System in the U.S (Franz et al., 2008). The Snow-17 model is a conceptual snowmelt scheme that simulates snow accumulation and ablation based on temperature and precipitation as inputs (Anderson, 2006). Although the physics behind it is not as comprehensive as the energy balance model, Snow-17 is advantageous for

having less required input data and performing "at least as good as" energy-based models (Ohmura, 2001).

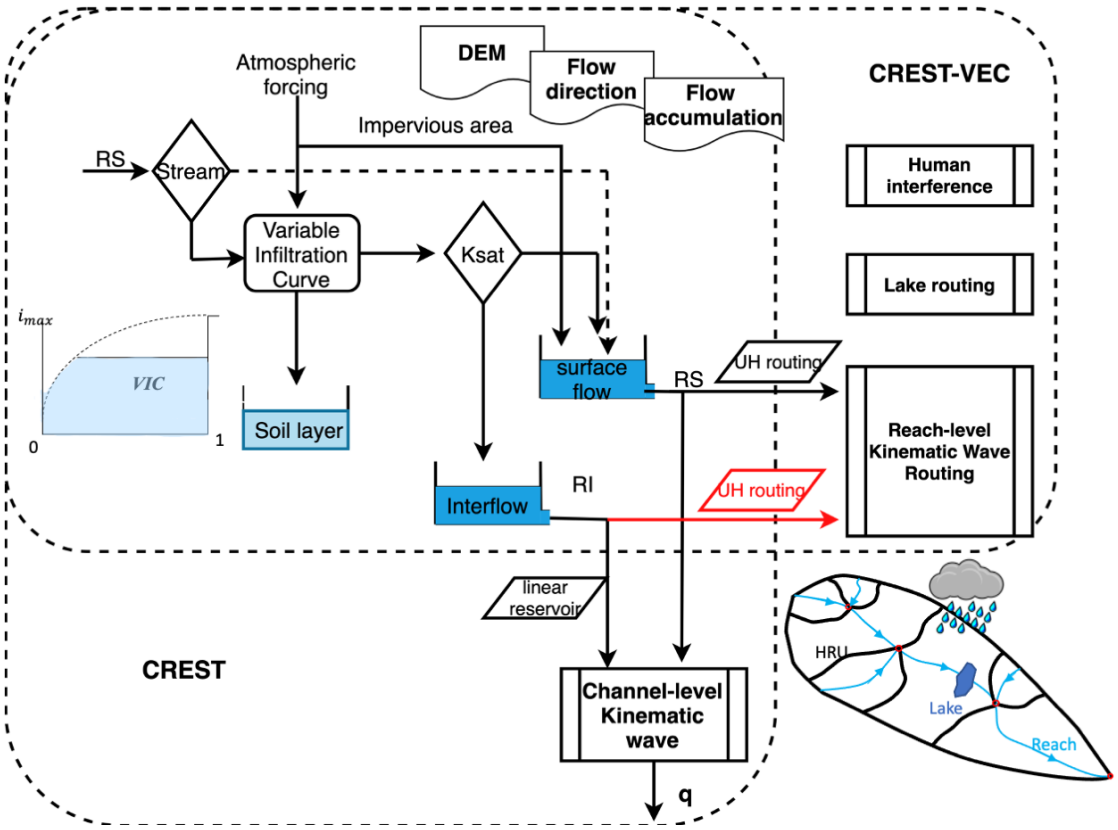


Figure 6.2 Schematic view of the CREST-VEC framework. The red arrow highlights the newly added subsurface routing option to the original mizuRoute framework.

6.2.4 mizuRoute

The mizuRoute river routing model, developed at the NCAR (National Center for Atmospheric Research), is a vector-based routing framework that incorporates both terrain and channel routing for large-domain river routing applications (Mizukami et al., 2016, 2021). For the terrain routing, the IRF or UH is used with parameters associated with gamma distribution to adjust the shape and scale. For the channel routing, user-defined options are IRF, kinematic wave with Lagrangian solution, and kinematic wave with Euler solution. A recent version of mizuRoute (Version 2.0.1) includes two lake routing schemes (Gharari et al., 2022; Vanderkelen et al., 2022) – one based on Döll et al. (2003) with a simple level-pool equation for natural lakes and the other more complicated one based on Hanasaki et al. (2006) which includes reservoir operation rules. These

two schemes have been applied to the other global hydrologic models (e.g., WaterGAP, VIC, and CWatM) to account for regulated streamflow. The original code is written in Fortran.

The current version of mizuRoute does not explicitly account for subsurface runoff routing over terrain, which is critical in the Great Plains and regions where streams are intermittent across a year (Salas et al., 2017). In this study, we enable an option to turn on or off subsurface routing as defined in the model configuration file. Similar to surface runoff routing, the subsurface flow is routed using the IRF scheme but with a much slower velocity and reduced magnitude. We use a two-parameter Gamma distribution function to materialize the IRF method as shown in Eq. 6.1.

$$y(t) = \frac{1}{\Gamma(a)\theta^a} t^{a-1} e^{-\frac{t}{\theta}} \quad (6.1)$$

Where t is the time variable, a is a shape parameter, and θ is a time-scale parameter. Both a and θ determine the flood peaking time and flashiness. After calculating instantaneous rates based on the gamma function, we use convolution to compute flow rates Q at time t. $R(t-s)$ is the (sub)surface runoff at the time $(t-s)$, and s is an increment of time from 0 to $tmax$ (also denoted as the time window). The default values of a and θ for hillslope surface routing are set to 2.5 and 8000. For subsurface flow routing, the a and θ are 10 and 86400, respectively.

$$Q(t) = \int_0^{tmax} y(t) \times R(t-s) ds \quad (6.2)$$

6.2.5 CREST-VEC

The framework, CREST-VEC, and the difference compared to its precedent CREST model are shown in Fig. 6.2. The main difference comes from the routing process, where the original CREST model routes surface flow and interflow via a kinematic wave routing model and a conceptual linear reservoir model in a gridded manner. However, the CREST-VEC model requires area-averaged time series of surface and subsurface flow at each river reach to be separately routed downstream. The gridded outputs from the CREST model (i.e., surface runoff and subsurface runoff) are extracted and averaged over each unit catchments or HRU using the newly developed Python package EASYMORE (EArth SYstem Modeling REmapper), publicly available from https://github.com/ShervanGharari/EASYMORE.

The framework is loosely coupled with two models written in different programming languages. A bash file calls three executables after model compilation subsequently (CREST-EASYMORE-mizuRoute). The input files for this model chain include forcing data (gridded precipitation,

potential evaporation, and temperature), topography data (gridded digital elevation model, flow direction, flow accumulation, river network topology, and hydrologic response unit), and configuration files. The topography data can be accessed from the HydroSHEDS website which consists of grid-based and vector-based topography data.

We use the IRF scheme in this study for both terrain routing and channel routing in this study and activate the lake model with the Döll et al. (2003) lake model. The parameters for lake parameters such as the outflow coefficient *a* and exponent *b* of Eq.6.3, are based on suggested values in Döll et al. (2003) and Gharari et al. (2022). For lakes that have monitored storage provided by the US Geological Survey (USGS), we directly insert storage time series into the model. As reservoir operation is not considered in this study, we exclude observed streamflow that is regulated by reservoirs and regulated lakes, as shown in Fig. 6.1c. So, only results from natural lakes, which account for 98% of US lakes or reservoirs, are considered valid for statistical comparison. To initialize model states, especially for initial lake volumes, we warm up the CREST-VEC model from 1948 to 2014 using the GLDAS forcing (Global Land Data Assimilation System) at a daily time step.

$$Q_{out} = a \times S_f \times (S_f / S_{f,max})^b, \quad (6.3)$$

where *a* and *b* are the outflow coefficient (1/day) and exponent, respectively; S_f is the actual lake storage (m^3); $S_{f,max}$ is the maximum lake storage (m^3).

6.3 Results

6.3.1 Case study: Houston region

As mentioned in the objectives of this study, we first conduct a case study analysis to assess the relative contributions of subsurface flow routing and lake routing to streamflow simulation based on the CREST-VEC framework. The original CREST model is used as a benchmark. We chose the Houston region (Fig. 3a) because there are two large natural lakes - Lake Barker and Lake Addicks that impact hydrologic simulations (Fig. 1a). For the CREST model with gridded routing, we calibrate the model using the DREAM (Differentiable Evolution Adaptive Metropolis) optimizer (Vrugt et a., 2009) from 2016-06-01 to 2017-06-01 at an hourly time step and performed evaluation from 2017-06-01 to 2020-01-01. The NSCE is used as the objective function for calibration, and the model is warmed up for one year from 2016-06-01 to 2017-06-01. We run the CREST model at three spatial resolutions: 1 km, 250 m, and 90 m. To be comparable with CREST-

VEC simulations, whose hydrography data is built upon a 90-meter resolution DEM, we only use CREST model results at 90 meters for statistical comparisons and use the results at 1 km and 250 meters to assess computational efficiency. The evaluation metrics shown in Fig. 3c are based on the evaluation period. The river flows from 22 stream gauges are curated from the USGS.

6.3.1.1 Model speedup

Figure 6.3b shows the computational cost (elapsed time at seconds per step) for a series of model configurations for the routing process. All the tests were run on a single core Intel i7-6700K CPU (4.00 GHz). The grid-based CREST model costs 0.01, 0.08, and 0.12 seconds per step at 1-km, 250-m, and 90-m resolutions, respectively. However, the CREST-VEC model can reduce this to approximately 0.002 seconds per step, regardless of grid resolutions from forcing data. There is little difference among the three scenarios (i.e., CREST-VEC, CREST-VEC+subq: CREST-VEC plus subsurface routing, and CREST-VEC+subq+lake: CREST-VEC with subsurface routing and lake routing). Relatively speaking, CREST-VEC can speed up the current operational CREST model at 1-km by 10x, let alone at finer resolutions.

6.3.1.2 Performance improvement

Regarding model skills, the CREST model and CREST-VEC achieve similar median NSE (Fig. 6.3c) based on observations from 22 stream gauges, even though the CREST model takes advantage of automatic calibration. CREST-VEC and CREST-VEC+subq overestimate flows downstream of two natural lakes, resulting in poor scores. But after incorporating lake routing schemes, the CREST-VEC+subq+lake model achieves not only better median scores but also less spread (quantified by the interquartile range). Notably, both CREST-VEC+subq and CREST-VEC+subq+lake have positive NSE values and smaller uncertainty ranges, primarily owing to included subsurface routing. The time series in Fig. 6.4 highlights the model performance at three stream gauges affected by upstream lakes. The CREST-VEC overestimates streamflow by a considerable amount (i.e., three times higher than observation in Hurricane Harvey), resulting in low NSE scores: 0.11, 0.16, and 0.18, respectively. With lake routing considered in the CREST-VEC+subq+lake, the simulated streamflow aligns well with observations, achieving NSE scores of 0.61, 0.65, and 0.64, respectively. Although the CREST model captures streamflow magnitude after calibration with the NSE scores – 0.37, 0.52, and 0.54, the peak timing is at least one-day

delayed for Hurricane Harvey. In summary, the advantages for the general CREST-VEC framework against the gridded CREST model are threefold: (1) improve computational efficiency by at least ten times, (2) improve overall model skill, (3) reduce uncertainty ranges.

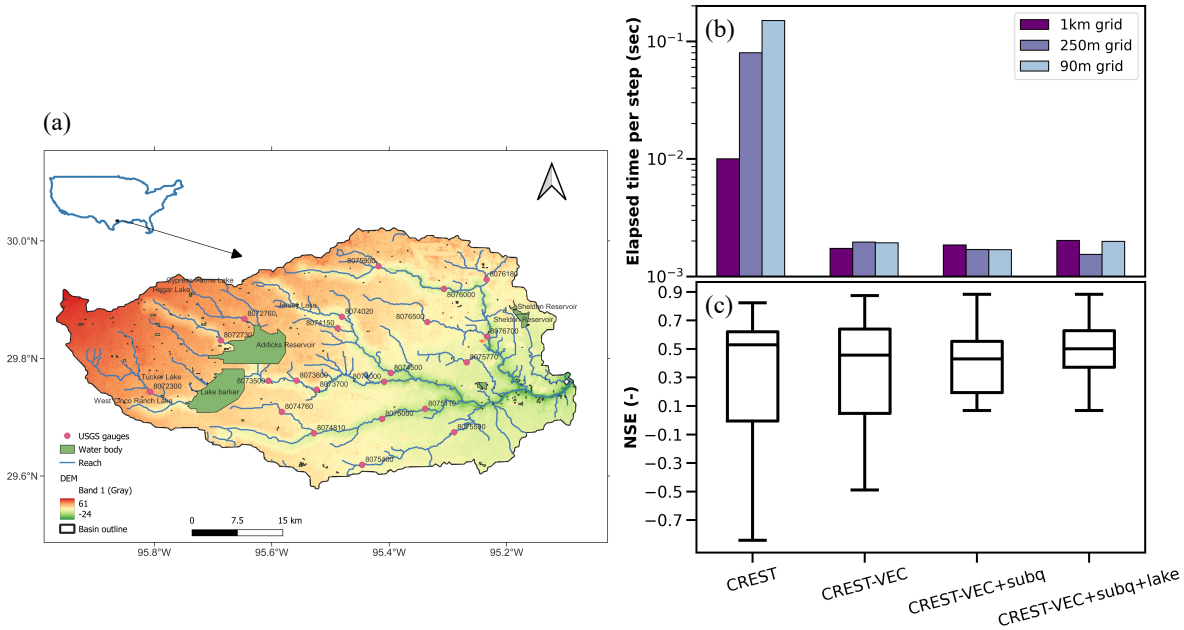


Figure 6.3 (a) Map of the study area (Houston region) showing river networks and water bodies; (b) Computation time per step for CREST at three resolutions and CREST-VEC model at four configurations on the x-axis; (c) Nash-Sutcliffe efficiency values for CREST and CREST-VEC model.

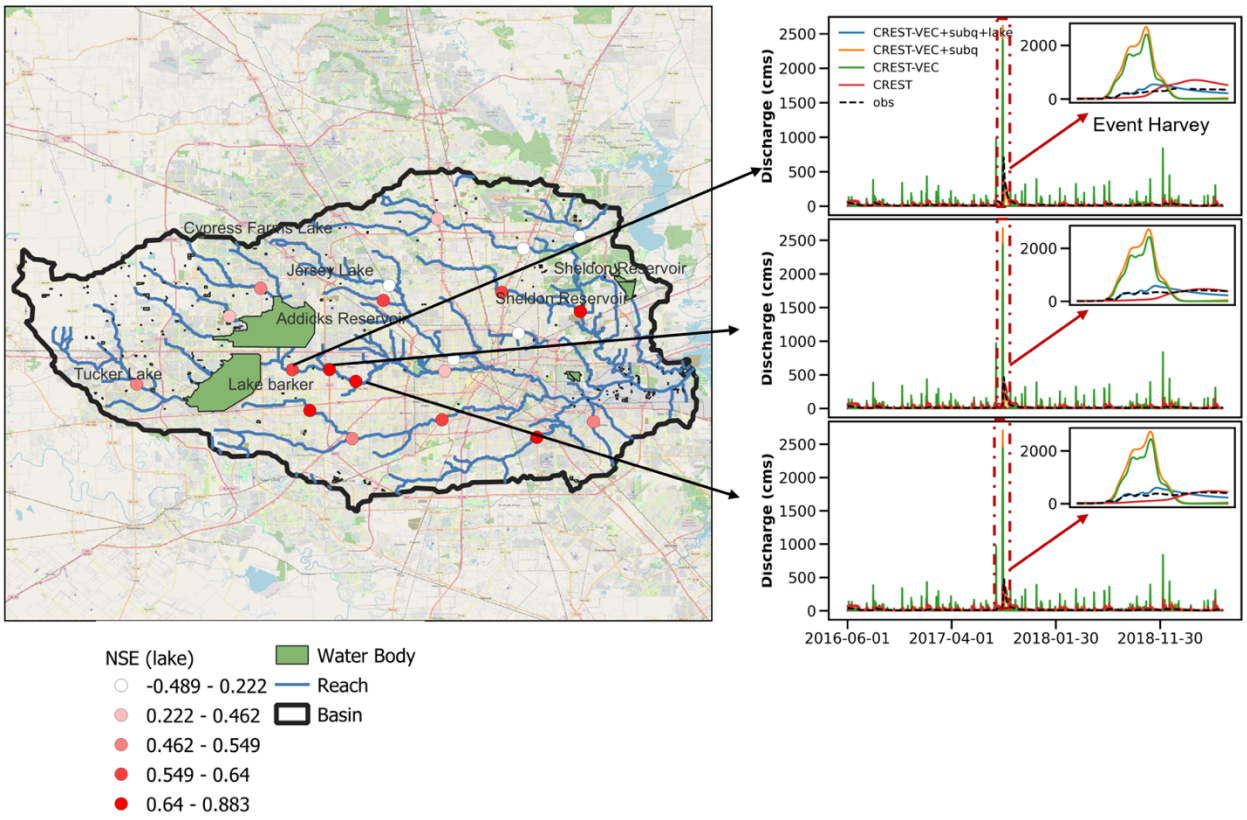


Figure 6.4 Performance of models downstream of two lakes. The Nash-Sutcliffe Efficiency coefficients are obtained from the CREST-VEC model with lake routing and subsurface routing. Three plots of time series of stream gauges (from upstream to downstream: 08073500, 08073600, 08074000) are pointed aside by the map, and the Hurricane Harvey event is highlighted in red box and insets.

6.3.2 CONUS simulation

Moving towards continental-scale hydrologic simulation, the CREST-VEC model excels at reducing computational costs, leaving room for quantifying uncertainties from forcing, model structure, and parameters in real-time. The following question is whether and how much the new lake routing improves a continental simulation. To answer this question, we simulate CREST-VEC with and without lake routing over the CONUS from 2017-06-01 to 2020-01-01 at an hourly time step. Notably, subsurface routing is activated for both models with and without lake routing, so we expect the difference in results to be primarily due to lake simulation. Streamflow data from 5,350 stream gauges in the same period are collected and used for model verification. For this case, the CREST-VEC model parameters are based on the pre-configured CONUS-wide parameters, the same as the ones used in Flamig et al. (2020).

6.3.3 Model speedup

Table 6.1 lists model performance with respect to total computational costs and evaluation scores of streamflow simulation. CREST-VEC certainly improves streamflow simulation not only via a higher resolution (from 1-km to 90-m) but with faster computational speed (149.2 hours to 29.9 hours in total; 7.2 sec/step to 0.37 sec/step for routing step only). Considering all preprocessing steps altogether, CREST-VEC model is still at least four times faster than the original framework. To be noted, a considerable amount of time is spent on mapping gridded runoff data to a vector form (>50% of the time). Future attention should be drawn on how to optimize the efficiency while preserving certain degrees of accuracy for this process.

Table 6.1 Statistical comparison of model performance over the continental U.S. Bolded numbers indicate the best metrics among the three model configurations. The computational speed is calculated as an average speed over a whole simulation period.

Metrics	**Gridded CREST (Flamig et al., 2020)**	**CREST-VEC (w/o lake)**	**CREST-VEC (w/ lake)**
Simulation resolution	1 km	**90 m**	**90 m**
Total computational cost (hours)	149.2	**29.9**	32.96
Computational Speed for routing (sec/step)	7.2	**0.35**	0.37
Max NSE	0.71	**0.87**	**0.87**
Median NSE	-0.06	0.12	**0.18**
% gauges NSE>0	41.8 %	50.6 %	**56.2 %**
Max CC	**1.0**	0.96	0.96
Median CC	0.40	**0.67**	**0.67**
Median bias	**9%**	27%	17%

6.3.4 Performance improvement

The median NSE score has increased from -0.06 (gridded) to 0.12 (no lake) and 0.18 (lake). The fraction of gauges with positive NSE scores has been improved from 41.8% (gridded CREST) to 50.6% (CREST-VEC without lake) and to 56.2% (CREST-VEC with lake). However, the CREST-VEC results are more biased than gridded CREST, partly due to the systematic overestimation of streamflow by the IRF routing scheme in the CREST-VEC. The difference

would be primarily attributed to the different routing processes, as CREST permits leakage in the interflow reservoir, thereby leading to lower positive bias. Results with lake simulation have reduced BIAS from 27% to 17%, as part of the water is being held in the lake. The CC (Correlation Coefficient), however, does not vary much between scenarios with and without lake simulation, as shown in Fig. 6.5. One of the reasons is that the CREST-VEC model does not simulate regulated lakes or reservoirs which have strong control of streamflow time shifts. Notably, the IQRs (interquartile ranges) of NSE and BIAS for lake simulation are lower than without lakes, meaning that this method particularly boosts scores at gauge locations that had poor performance previously. Figure 6.6 depicts the spatial map of model skill (with lake) and its difference between scenarios with and without lake simulation. CREST-VEC with lake module in regions like the West Coast and Upper Mississippi River Basin have relatively good performance (NSE>0.4), yet over the Great Plains and East Coast, the model bias is high (BIAS>1), yielding low NSE scores. Similar issues are found in the literature with other models (Clark et al., 2008; Newman et al., 2015; Mizukami et al., 2017; Salas et al., 2017; Lin et al., 2019; Konben et al., 2020; Knoben et al., 2020; Yang et al., 2021; Tijerina et al., 2021). Taking the Great Plains as an example (highlighted box in Fig. 6.6c), the model physics of CREST-VEC does not correctly represent the real hydrologic processes by two means. First, the surface runoff (before routing) simulated by CREST-VEC is biased. We compare the annual surface runoff by CREST-VEC to the public community dataset GRFR (Global Reach-level Flood Reanalysis) in Fig. 6.7. The runoff in GRFR is simulated by the VIC model and undergoes stringent bias correction against observations via the discrete quantile mapping technique (Yang et al., 2021; Lin et al., 2019). There is a 116.3% higher surface runoff by the CREST-VEC than theirs, partly explaining the high BIAS and low NSE scores in such region. We suspect the singular bulk soil layer represented in the CREST model yields such systematic differences. Second, the missing representation of playas, small and rain-fed lakes that are prominent in the Great Plains, leads to falsely produced runoff (Hay et al., 2016; Solvik et al., 2021). However, even when accounting for multiple hydrologic model structures, performance in this region is still ranked as one of the poorest (Clark et al., 2008; Knoben et al., 2020). For example, Knoben et al. (2020) analyzed 36 hydrologic models over the U.S., in which the maximum KGE scores out of those models are lower than 0.5 over the Great Plains.

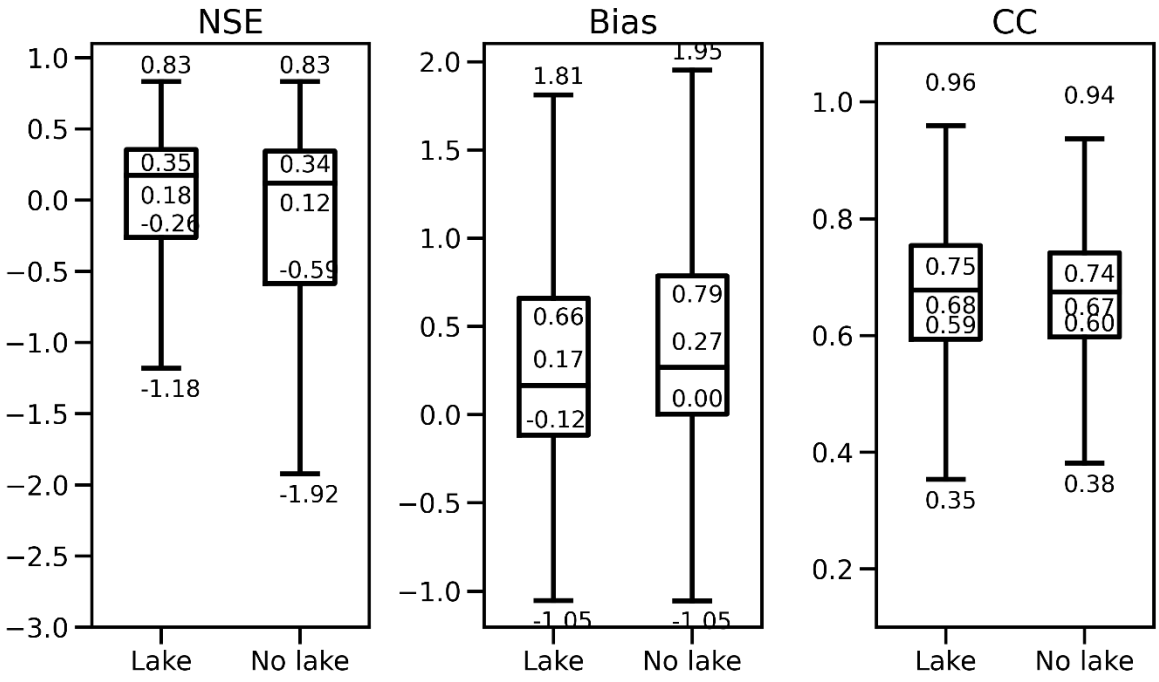


Figure 6.5 Boxplot of model performance comparing results with lake routing and without it.

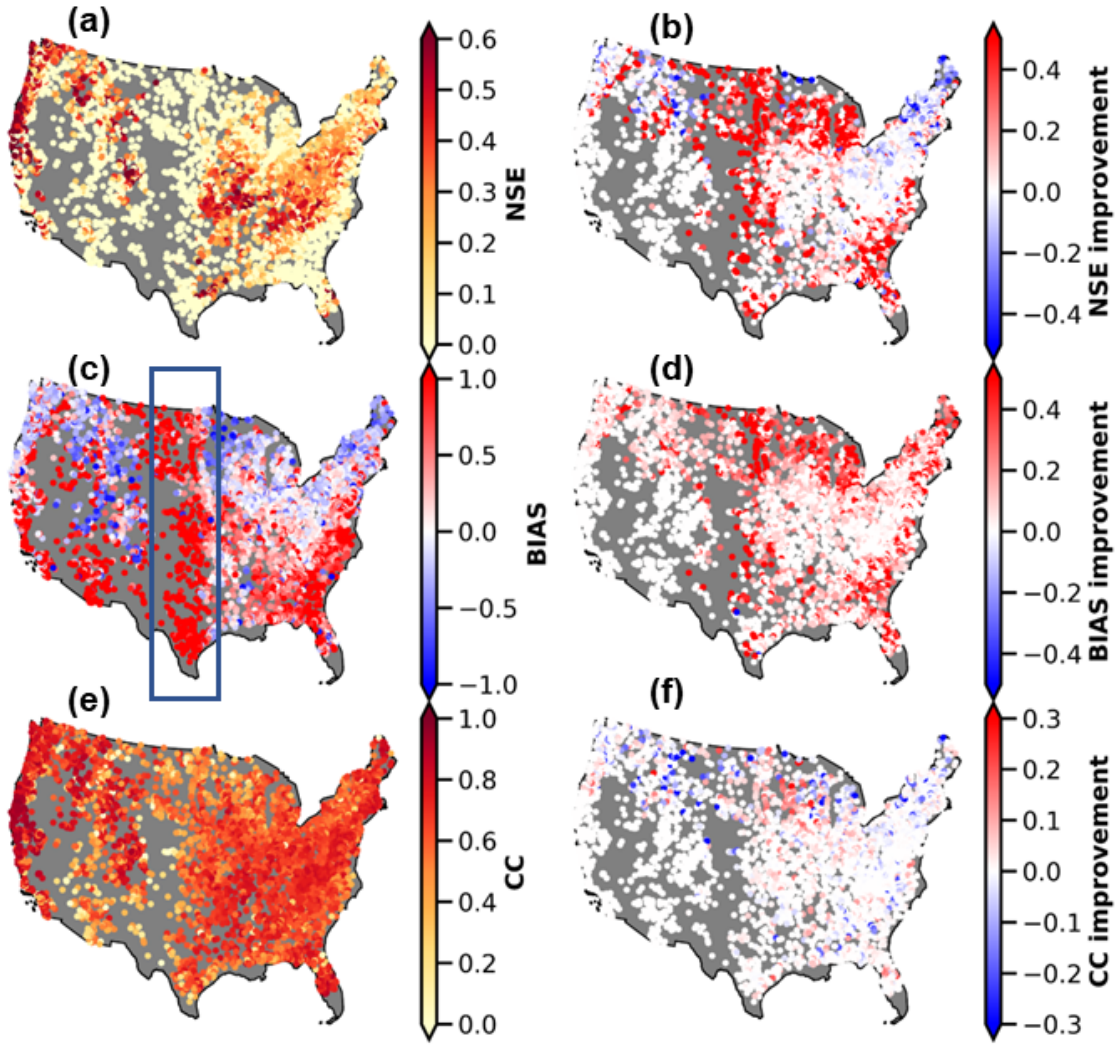


Figure 6.6 Spatial map of model performance with the lake (left column) and the difference between with and without lake simulation (right column). (a): NSE scores; (b) NSE differences (results with lake minus results without lake); (c) BIAS; (d) BIAS difference; (c) Correlation Coefficient (CC); (f) CC difference. The blue box in (c) highlights the region where high positive BIAS is present.

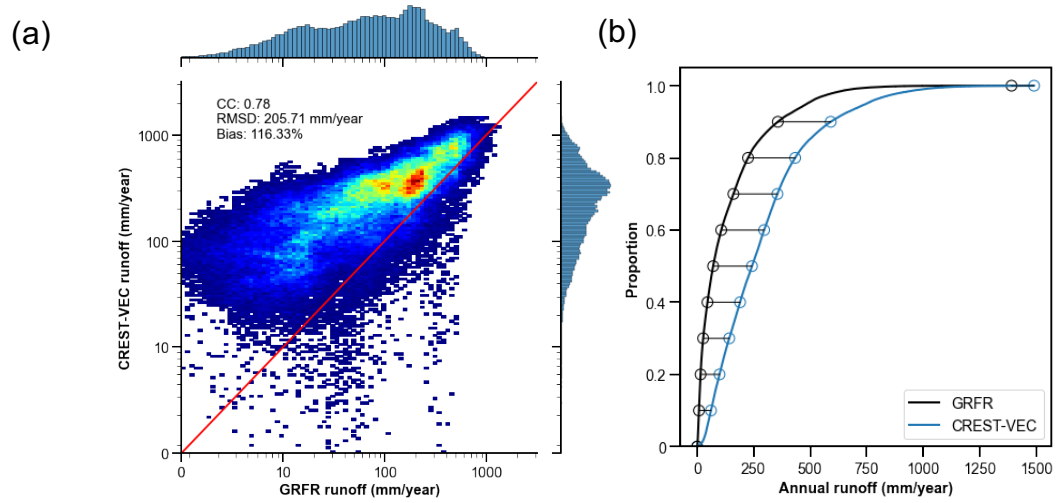


Figure 6.7 (a) density plot of CREST-VEC simulated annual surface runoff against GRFR (Global Reach-level Flood Reanalysis) in the Great Plains; (b) cumulative density function (CDF) of CREST-VEC and GRFR simulated annual runoff.

6.3.5 How likely are floods falsely detected?

In this section, we shift gears to explore how likely US floods are falsely alarmed if no lake simulations are included. We selected 283 gauges that are downstream of natural lakes (Fig. 6.8), with most of them located in the middle and eastern US. The hourly time series of streamflow of those gauges are compared against advised flood thresholds (2-year flooding) provided by the US Geological Survey. They fit a log-Pearson III type distribution to the annual maxima streamflow from long-term records and extract values with given flood frequency. Following a similar approach as in Yang et al. (2021), consecutive yet independent events have to be two days apart from one another. From there, we calculated the Probability of Detection (POD), False Alarm Ratio (FAR), and Critical Success Index (CSI) based on the contingency table.

As expected, median FAR is reduced from 0.63 (without lake simulation) to 0.50 (with lake), resulting in a slightly higher CSI of 0.36 than that of 0.31 for no lake simulation (Fig. 6.8a). Additionally, previous research reported that simulation results with the lake module mitigate the seasonal variability of the river discharge (Tokuda et al., 2021). The decrease in FAR values implies five instances: (1) decrease in false alarms while hits remain the same; (2) increase in hits while false alarms remain the same; (3) decrease in false alarms while increase in hits; (4) decrease in both false alarms and hits; and (5) increase in both false alarms and hits. We find that however POD values decrease from 0.87 without lakes to 0.85 with lakes, from which we can infer that both hits and false alarms are decreasing, but false alarms decrease at a higher rate. That is a fact

of reducing simulated flood peak, which results in fewer hits in flood forecasts but meanwhile less falsely alarmed floods. As most studies focus on flood detection, they inevitably arrive at more falsely detected floods. Too many false alarms could make people disregard the warnings, despite a real threat, causing the "cry wolf" effect.

Maps in Fig. 6.8b display the distributions of flood detectability with lake simulation and its improvements compared to results without lake simulation. High POD and FAR values co-exist in the Great Plains, where the model simulates considerably higher streamflow values than observations. Moderate FAR values are found near the Florida panhandle and parts of Georgia. Lower FAR values are found in the Midwest and West Coast. Compared to results without the lake, FAR values are reduced reasonably over the East Coast, Midwest, Gulf Coast, and West Coast, although POD values remain relatively unchanged or even decreased.

Five local cases are shown in Fig. 6.9, which depicts the river topology and time series of hourly streamflow. One can infer that these lakes are not heavily regulated from recorded streamflow time series, therefore showing the effectiveness of our model. In Fig. 6.9a, the simulated streamflow without lakes is heavily overestimated, peaking at 1200 cms in the year 2017, whereas the actual flow rate is around 400 cms. The scenario with lake simulation, however, produces a magnitude much closer to the observation. Due to decreased systematic bias, the lake scenario boosts the NSE score from -0.2 to 0.5. There is also an 8% less chance of issuing false alarms than the model without lake simulation. Figure 6.9b shows a case where FAR is reduced from 0.70 to 0.17, a reduction rate of 75.7%. The flood detectability, i.e., CSI, is greatly improved from 0.29 to 0.57. Figure 6.9c exemplifies a case with all improved metrics (i.e., NSE, POD, FAR, and CSI). All these three cases in Figs. 6.9a-c are located along the St. Johns River, in which we expect a systematic improvement along this river after incorporating the lake simulation. Figure 6.9d displays more common cases where a reduction of FAR comes at the expense of reducing POD (i.e., flood detection), almost at the same pace. Figure 6.9e shows that although the model with lakes produces better baseflow, it underestimates flood peaks, resulting in lower NSE values (0.3) than results without lakes (0.4). It implies that parameters governing the lake outflow need to be improved.

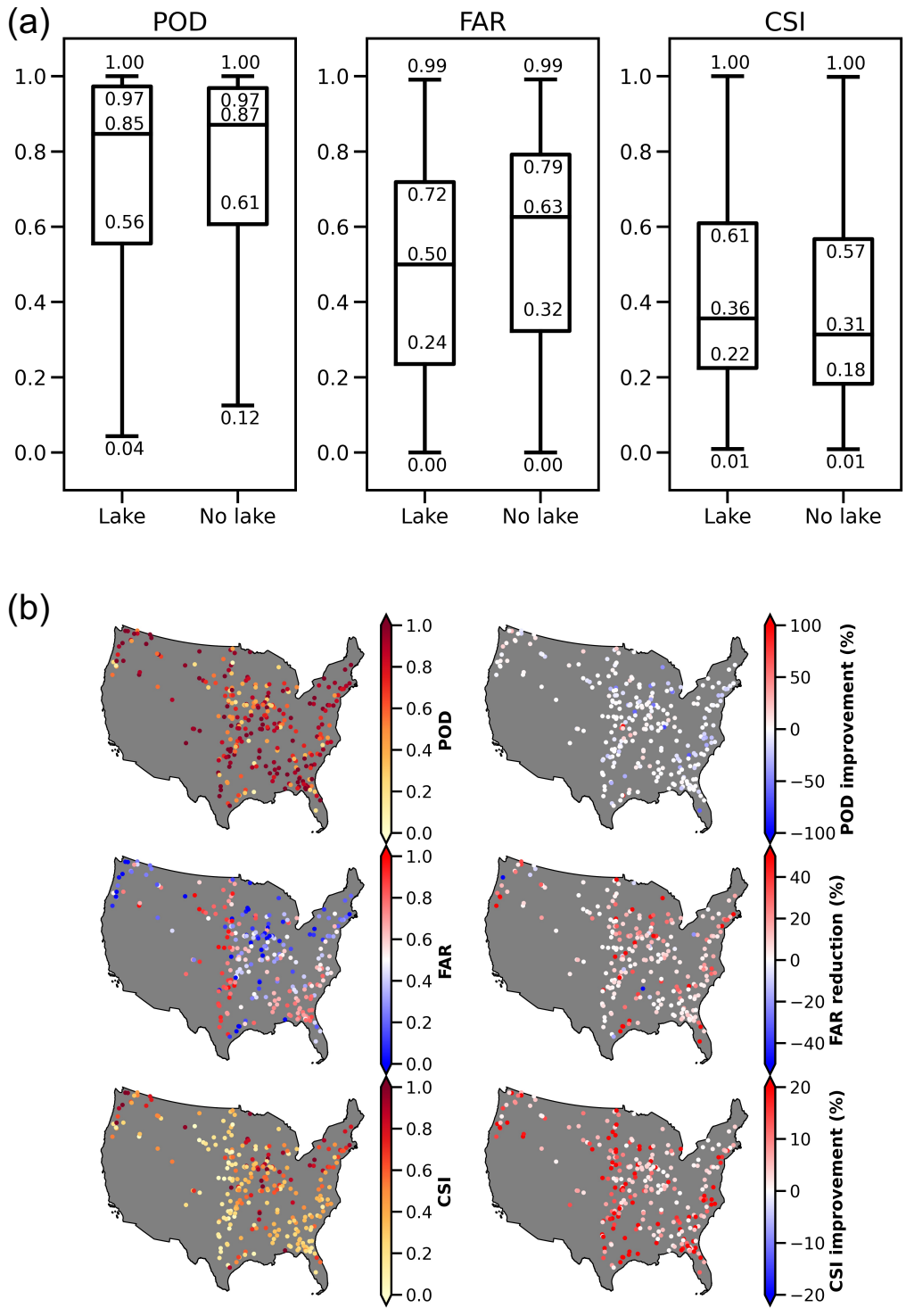


Figure 6.8 Flood detection performance comparing lake and no lake simulation. (a) Similar to Figure 5, but for flood detectability; (b) Similar to Figure 6, but for flood detectability.

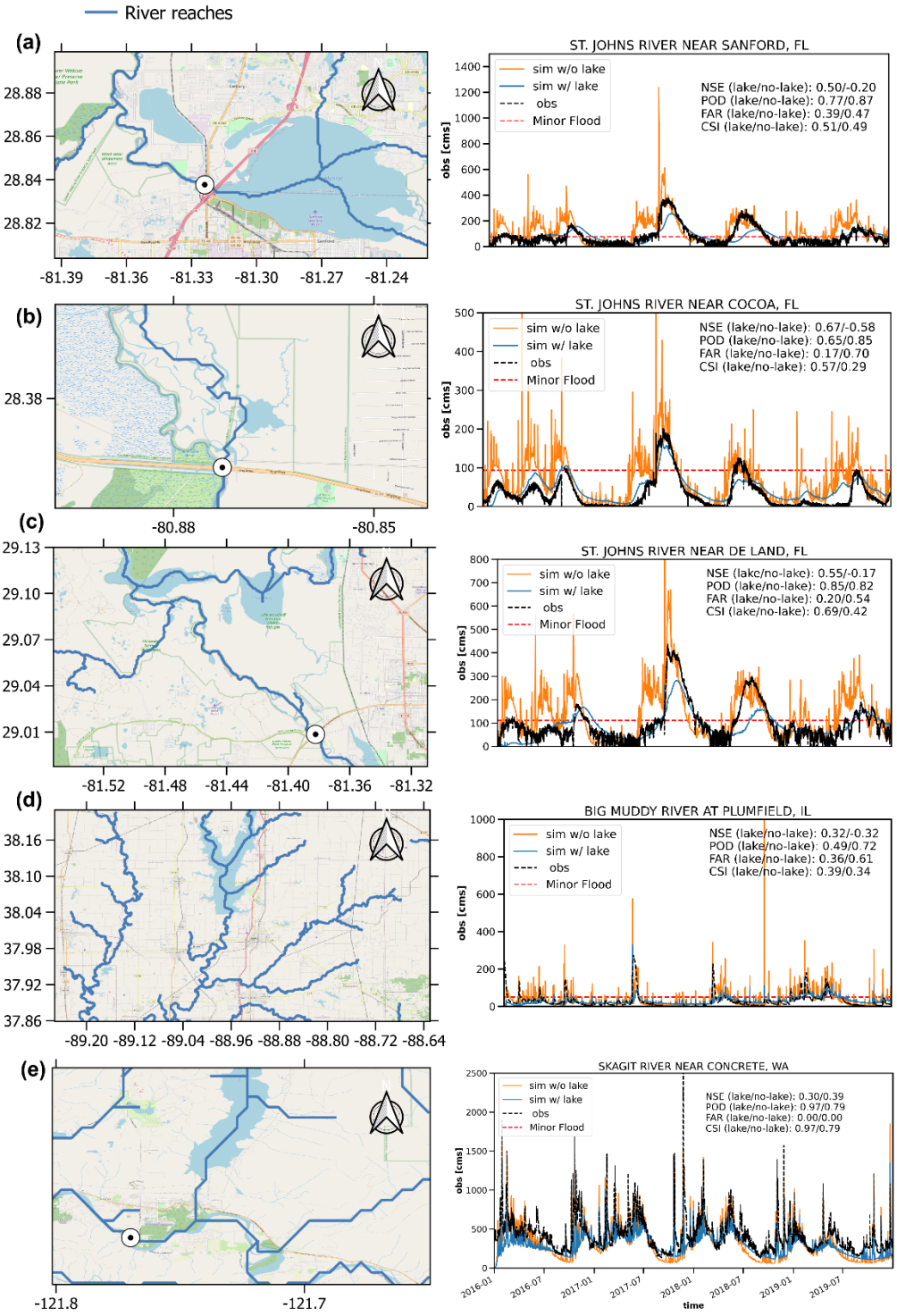


Figure 6.9 Five case examples of streamflow time series at gauges downstream of lakes: (a) St. Johns River near Sanford, FL; (b) St. Johns River near Cocoa, FL; (c) St. Johns River near De

Land, FL; (d) Big Muddy River at Plumfield, IL; (e) Mississippi River at Clinton, IA. Images courtesy of Google Map.

6.4 Discussion

6.4.1 Vector vs. Raster routing

In this study, we compare the advantages of vector-based routing with respect to conventional raster-based routing in two aspects: (1) model efficiency and (2) model accuracy. Overall, the vector-based routing shows great promise, as it speeds up the routing process by at least ten times, compared to grid-based routing, for both the regional simulation (0.07 sec/step vs. 0.002 sec/step) and the CONUS simulation (0.35 sec/step vs. 7.2 sec/step). In terms of results against observations, the CONUS-wide performance is improved regarding NSE values. However, the variable river reach lengths (from hundreds of meters to tens of kilometers) in large-scale simulation pose challenges for estimating routing parameters such as the time and shape parameters in a unit hydrograph. Second, most land surface models are still grid-based, making a type mismatch (grid-based land surface model vs. vector-based routing model) (Lehner & Grill, 2013). To integrate the two, we need a processing step by mapping surface and subsurface runoff onto representative HRUs. Different aggregation strategies are present and subject to the primary purpose of interest. At present, there is an ongoing effort to seamlessly integrate these two processes together (Gharari et al., 2020). However, it is yet to be efficient and draws further attention to improving this mapping scheme. Third, the many-to-one river network is established but not for one-to-many, meaning river bifurcation is challenging to represent and tackle (Yamazaki et al., 2014).

Raster-based routing comes at the resolution of the input DEM data, albeit at a slower computational speed. Having matured over the years, most raster-based routing models are seamlessly integrated with water balance models so that model can be set up at minimum effort by a modeler. Mostly derived from hydrodynamic models, the concept of “raster” can be extended to “grid” because of the emerging unstructured grids such as triangles, curvilinear, hexagons, etc. These flexible grid types mimic the real flow directions and reduce computational costs. However, they are yet to be accepted/applied by the hydrologic model community.

As one objective of this study, we want to examine the potential improvement from the with-lake configuration on streamflow simulation over a wide range of hydrometrical and

geographical settings in the CONUS, rather than provide some optimal model setup and parameterization at the CONUS scale, which we believe is way beyond our scope and several steps forward from the current CREST-VEC or any existing CONUS models. As far as what qualifies 'an adequate base simulation', there may be some room for debate but should be some bottom-line principles: first, one should be clearly aware of the sources of uncertainties, including forcing, model structure, parameterization, streamflow observation as the reference, etc. Optimization, though effective in improving the model performance, compensates for uncertainties from the other sources simply via adjusting model parameters. This has been acceptable for operational purposes but is not appropriate for this study where a modification of the model structure is introduced. Instead, we use an a-priori parameter set that was developed based on remote sensing datasets and also evaluated at the CONUS scale (Vergara et al., 2016). The physical base of these a-priori parameters set a solid foundation for examining the new with-lake configuration, thus should not be compromised via parameter tunning.

The CREST-VEC model, by no means, represent all physical hydrological processes. Instead, it is a conceptual flood forecast model that aims to deliver timely flood information to stakeholders, decision-makers, and broader users. We admit that some processes such as vadose zone modeling, snow melting, hillslope routing, in-channel river routing, and reservoir operations are simplified, and some processes such as vegetation and groundwater modeling are missing from the current version. For the lake module, we expect to include more sophisticated multi-layer decision processes instead of a level-pool process. Lake evaporation is another important factor to be considered for improved water balance. Since it is a compromise between model complexity and efficiency, we hope to continuously push the envelope on this front to optimize the real-time flood forecast system.

6.4.2 Room for improving large-scale hydrologic simulation

Large-scale hydrologic simulation is still a long-standing challenge for the hydrologic community, especially with debates on developing a "one-model-fits-all" structure or a "malleable" structure (Burek et al., 2020; Clark et al., 2015a; Fenicia et al., 2011; Savenije, 2008). The CREST model, in our study, systematically overestimates surface runoff over the Great Plains and Southeast, a result of some misrepresented or missing processes, yet excels in flash flood simulation. Diverse hydrologic model structures, on the other hand, hope to overcome individual

limitations and offer joint benefits (Horton et al., 2021). We, therefore, promote the "malleable" model structure from the efficiency point of view - a flexible structure disables redundant hydrologic processes. Then, the central question becomes: How do we adapt the model to variable catchment processes? In such a context, intercomparisons and discussions of different hydrologic models in varying catchment processes become particularly valuable (Clark et al., 2015a; Knoben et al., 2020; Tijerina et al., 2021). Notably, simply relying on the NSE or KGE score to assess the model performance can be misguiding (Clark et al., 2021).

Hydrologic calibration is powerful in boosting model accuracy, yet large-scale models oftentimes suffer from the complexity that impedes credible model calibration. River routing schemes and their parameters can affect streamflow simulations, especially at fine time scales such as sub-daily (Mizukami et al., 2021). Our current study used an IRF scheme in which the impulse response function is derived from a diffusive wave equation (see Lohman et al., 1996; Mizukami et al., 2016) and includes two parameters – diffusivity and celerity. These parameters need to be exposed to calibration in addition to the hydrologic model parameters. Furthermore, to fully understand the routing model's impact on streamflow simulations, it is necessary to consider other routing schemes including a diffusive wave model as well as a kinematic wave model, which may be suited for flood forecasting.

Lastly, the computational costs for large-scale simulation can be optimized from accelerated hardware (multi-core CPUs and GPUs) once codes are parallelized and scalable. Advances in Reduced Order Modeling (ROM), a surrogate model which develops a parsimonious solution to replace the computationally intensive part, hold promise to reduce costs (Clark et al., 2015b). For instance, to integrate reservoir simulation into the CREST-VEC system, we can build an offline ML model which is promising in mimicking human decisions (Yang et al., 2021) and plug it into the system.

6.4.3 Towards improved flood forecasting with lake routing

Flood forecasts are difficult because of their rarity, and their hits and misses are typically low while false alarms are high (Bartholmes et al., 2009; Cloke & Pappenberger, 2009). Results in this study demonstrate a dilemma in which the model with a lake module reduces false alarms but at the cost of more missed flood events compared to the one without a lake module. Although the combined metric CSI has a certain degree of improvement, this leaves a question – should we

reduce a large number of false alarms at the expense of missing a small number of real events? Before discussing this point, we acknowledge that the current lake routing process is simple and imperfect, and improvement in this process possibly leads to an optimal situation where both false alarms and misses can be improved. However, in most situations, tradeoffs exist in hydrologic predictions. A good strategy in our case would be running both simulations with and without the lake module concurrently and making the "without lake" results the worst-case scenario. Since the CREST-VEC model has the advantage of efficiency, running two scenarios is totally feasible. A decision-maker can be trained to assess the situation – results from two scenarios disagree – from the perspective of flood severity and consequences.

6.5 Conclusion

This study compares a conventional raster-based routing scheme with the emerging vector-based routing approach in hydrologic models for regional case and continental simulations. From the continental run, we demonstrate the improvement in streamflow simulation after incorporating the lake storage and release module. Last but not least, flood-related false alarms can be greatly reduced by including the lake module. The following points summarize the primary findings of the study:

1. Vector-based routing can accelerate continental-scale river routing by up to ten times, compared to a grid-based routing, for both a regional case (0.07 sec/step) and a continental case (0.002 sec/step). This leaves adequate room for generating ensemble predictions with variable forcing, parameters, and/or model structures. Furthermore, it improves streamflow simulation from -0.06 to 0.18, according to the aggregated median NSE values.

2. A newly developed lake model increases the NSE score by 56.2% and reduces systematic BIAS by 17% for the continental simulation.

3. Flood false alarm ratios can be mitigated by 20.6% after enabling the lake module at the expense of missing 2.3% more floods on a continental scale.

We recommend the use of ensemble simulations stemming from different model structures to overcome and adapt to varying catchment processes. Optimized streamflow prediction with quantified uncertainty information can be achieved in an operational manner for stakeholders and decision-makers. Future studies can fully investigate the limitation and uncertainty of different forcing, parameters, and/or model structures to catchment signatures such as climatology, dominant hydrologic processes, lithology, etc. Vector-based routing, in such a context, can enable

fair comparison by excluding the effect of different routing schemes while focusing on discrepancies in water balance models alone. For future work, we hope to have the best possible model-simulated streamflow product in the US, fused with multi-model structures and observations. Another direction is to improve current lake and reservoir outflow simulation with a hybrid model – process-based and ML-based.

6.6 Reference

Allen, R.G., Pereira, L., Raes, D., & Smith, M. (1998). Crop Evapotranspiration, Food and Agriculture Organization of the United Nations, Rome, Italy. FAO publication 56. ISBN 92-5-104219-5. 290p.

Anderson, E. A. (2006). Snow accumulation and ablation model–SNOW-17. US National Weather Service, Silver Spring, MD, 61.

Bartholmes, J., Thielen, J., Ramos, M., & Gentilini, S. (2009). The European flood alert system EFAS – Part 2: statistical skill assessment of probabilistic and deterministic operational forecasts. Hydrology and Earth System Science, 13 (2) 141–153.

Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1-4), 11-29, 2001.

Beven, K. (2006). A manifesto for the equifinality thesis, Journal of Hydrology, 320, 18-36, https://doi.org/10.1016/j.hydrol.2006.07.007.

Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F., & Wada, Y. (2020). Development of the Community Water Model (CWatM v1.04) – a high-resolution hydrological model for global and regional assessment of integrated water resources management, Geoscientific Model Development, 13, 3267–3298, https://doi.org/10.5194/gmd-13-3267-2020.

Chen, M., Li, Z. & Gao, S. (2022). Multisensor Remote Sensing and the Multidimensional Modeling of Extreme Flood Events. In Remote Sensing of Water-Related Hazards (eds K. Zhang, Y. Hong and A. AghaKouchak). https://doi.org/10.1002/9781119159131.ch5.

Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). Applied Hydrology: McGraw-Hill Series in Water Resources and Environmental Engineering, McGraw-Hill, Inc., New York.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., & Hay, L. E. (2008). Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resources Research, 44, W00B02, https://doi.org/10.1029/2007WR006735.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski D., Rupp, D. E., Woods, R. A., Freeer, J. E., Gutmann, E. D., Wood, A. W., Gochis, D. J., Rasmussen, R. M., Tarboton, D. G., Mahat, V., Flerchinger G. N., & Marks, D. G. (2015a). A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resources Research, 51, 2498 – 2514, https://doi.org/10.1002/2015WR017198.

Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., Mackay, D. S., Maxwell, R. M., Shen, C., Swenson, S. C., & Zeng, X. (2015b) Improving the representation of hydrologic processes in Earth System Models, Water Resources Research, 51, 5929– 5956, https://doi.org/10.1002/2015WR017096.

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Shervan, G., Freer J. E., Whitfield, P. H., Shook, K. R., & Papalexiou, S. M. (2021) The abuse of popular performance metrics in hydrologic modeling. Water Resources Research, 57, e2020WR029001. https://doi.org/10.1029/2020WR029001.

Cloke, E. & Pappenberger, F. (2009). Ensemble flood forecasting: A review. Journal of Hydrology, 375, 613-626. https://doi.org/10.1016/j.jhydrol.2009.06.005.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. A. (2008). Physiographically-sensitive mapping of temperature and precipitation across the conterminous United States. International Journal of Climatology, 28, 2031-2064.

David, C. H., Maidment, D. R., Niu, G., Yang, Z., Habets, F., & Eijkhout, V. (2011) River Network Routing on the NHDPlus Dataset, Journal of Hydrometeorology, 12(5), 913-934.

de Almeida, G. A. M., & Bates, P. (2013). Applicability of the local inertial approximation of the shallow water equations to flood modeling, Water Resources Research, 49, 4833– 4844, https://doi.org/10.1002/wrcr.20366.

Döll, P., Kaspar, F., & Lehner, B. (2003). A global hydrological model for deriving water availability indicators: model tuning and validation. Journal of Hydrology, 270, 105-134. https://doi.org/10.1016/S0022-1694(02)00283-4.

Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resources Research, 47, W11510, https://doi.org/10.1029/2010WR010174.

Flamig, Z. L., Vergara, H., & Gourley, J. J. (2020). The Ensemble Framework For Flash Flood Forecasting (EF5) v1.2: description and case study, Geoscientific Model Development, 13, 4943–4958, https://doi.org/10.5194/gmd-13-4943-2020.

Franz, K. J., Hogue, T. S., & Sorooshian, S. (2008). Operational snow modeling: Addressing the challenges of an energy balance model for National Weather Service forecasts. Journal of Hydrology, 360(1-4), 48-66.

Freeze, R. A. (1972). Role of subsurface flow in generating surface runoff: 2. Upstream source areas, Water Resources Research, 8, 1272– 1283, doi:10.1029/WR008i005p01272.

Gao, S., Chen, M., Li, Z., Cook, S., Allen, D., Neeson, T., Yang, T., Yami, T., & Hong, Y. (2021) Mapping dynamic non-perennial stream networks using high-resolution distributed hydrologic simulation: A case study in the upper blue river basin. Journal of Hydrology, 600, 126522.

Getirana, A. C. V., Boone, A., Yamazaki, D., Decharme, B., Papa, F., & Mognard, N. (2012). The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon Basin, Journal of Hydrometeorology, 13(6), 1641-1665.

Gharari, S., Clark, M. P., Mizukami, N., Knoben, W. J. M., Wong, J. S., & Pietroniro, A. (2020). Flexible vector-based spatial configurations in land models, Hydrology and Earth System Science, 24, 5953–5971, https://doi.org/10.5194/hess-24-5953-2020.

Gharari, S., Vanderkelen, I., Tefs, A., Mizukami, N., Stadnyk, T. A., Lawrence, D., & Clark, M. P. (2022). A Flexible Multi-Scale Framework to Simulate Lakes and Reservoirs in Earth System Models, Earth and Space Science Open Archive.

Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P., Clark, R. A., III, Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., Hong, Y., & Howard, K. W. (2017). The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. Bulletin of American Meteorological Society, 98, 361–372. https://doi.org/10.1175/BAMS-D-15-00247.1.

Hanasaki, N, Kanae, S., & Oki, T. (2006). A reservoir operation scheme for global river routing models, Journal of Hydrology, 327, 22-41.

Hay, L, Norton, P, Viger, R, Markstrom, S, Regan, RS, & Vanderhoof, M. (2018). Modelling surface-water depression storage in a Prairie Pothole Region. Hydrological Processes, 32, 462–479. https://doi.org/10.1002/hyp.11416.

Horton, P., Schaefli, B., & Kauzlaric, M. (2021). Why do we have so many different hydrological models? A review based on the case of Switzerland. Wiley Interdisciplinary Review: Water, e1574. https://doi.org/10.1002/wat2.1574.

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A brief analysis of conceptual model structure uncertainty using 36 models and 559 catchments. Water Resources Research, 56, e2019WR025975. https://doi.org/10.1029/2019WR025975.

Knoben, W. J. M., Clark, M., Bates, J., Bennet, A., Gharari, S., Marsh, C., Nijssen, B., Pietroniro, A., Spiteri, R., Tarboton, D. J., & Wood, A. J. (2021). Community Workflows to Advance Reproducibility in Hydrologic Modeling: Separating model-agnostic and model-specific configuration steps in applications of large-domain hydrologic models, Earth and Space Science Open Archive, https://doi.org/10.1002/essoar.10509195.1.

Lehner, B., Liermann, C.R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., Döll, P., Endejan, M., Frenken, K., Magome, J., Nilsson, C., Robertson, J.C., Rödel, R., Sindorf, N. & Wisser, D. (2011). High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. Frontiers in Ecology and Environment, 9, 494-502. https://doi.org/10.1890/100125.

Lehner, B. & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. Hydrological Processes, 27: 2171-2186. https://doi.org/10.1002/hyp.9740.

Li, H., Wigmosta, M. S., Wu, H., Huang, M., Ke, Y., Coleman, A. M., & Leung, L. R. (2013). A Physically Based Runoff Routing Model for Land Surface and Earth System Models, Journal of Hydrometeorology, 14(3), 808-828, 2013.

Li, Z., Chen, M., Gao, S., Hong, Z., Tang, G., Wen, Y., Gourley, J. J., & Hong, Y. (2020). Cross-examination of similarity, difference and deficiency of gauge, radar and satellite precipitation measuring uncertainties for extreme events using conventional metrics and multiplicative triple collocation. Remote Sensing, 12(8), 1258.

Li, Z., Chen, M., Gao, S., Luo, X., Gourley, J. J., Kirstetter, P., Yang, T., Kolar, R., McGovern, A., Wen, Y., Rao, B., Yami, T., & Hong, Y. (2021). CREST-iMAP v1. 0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction. Environmental Modelling and Software, 141, 105051.

Li, Z. (2022). CREST-VEC: A framework towards more accurate and realistic flood simulation across scales (v1.0). Zenodo. https://doi.org/10.5281/zenodo.6305817.

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models, Journal of Geophysical Research, 99(D7), 14415– 14428, https://doi.org/10.1029/94JD00483.

Lohman, D., Nolte-holube, R. & Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes. Tellus A: Dynamic Meteorology and Oceanography, 48: 708-721. https://doi.org/10.1034/j.1600-0870.1996.t01-3-00009.x.

Lighthill M. J. & Whitham G. B. (1955). On kinematic waves I. Flood movement in long rivers. Proceedings of Royal Society of London, A229281–316. http://doi.org/10.1098/rspa.1955.0088.

Lin, P., Rajib, M. A., Yang, Z., Somos-Valenzuela, M., Merwade, V., Maidment, D. R., Wang, Y., & Chen, L. (2018). Spatiotemporal Evaluation of Simulated Evapotranspiration and Streamflo over Texas Using the WRF-Hydro-RAPID Modeling Framework. Journal of American Water Resources Association, 54(1), 40-54. https://doi.org/10.1111/1752-1688.12585, 2018.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. G., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., & Wood, E. F. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. Water Resources Research, 55, 6499– 6516. https://doi.org/10.1029/2019WR025287.

Lin, P., Pan, M., Wood, E.F. Yamazaki, D., & Allen, G. H. (2021). A new vector-based global river network dataset accounting for variable drainage density. Scientific Data, 8, 28, https://doi.org/10.1038/s41597-021-00819-9.

Liston, G. E., Sud, Y. C. & Wood, E. F. (1994). Evaluating GCM land surface hydrology parameterizations by computing river discharges using a runoff routing model. Journal of Applied Meteorology, 33: 394–405.

Martinaitis, S. M., Gourley, J. J., Flamig, Z. L., Argyle, E. M., Clark, R. A., III, Arthur, A., Smith, B. R., Erlingis, J. M., Perfater, S., & Albright, B. (2017). The HMT Multi-Radar Multi-Sensor Hydro Experiment, Bulletin of American Meteorological Society, 98(2), 347-359.

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nature Communications, 13603. https://doi.org/10.1038/ncomms13603.

Messager, M.L., Lehner, B., Cockburn, C. Lamouroux, N., Pella, H., Snelder, T., Tockner K., Trautmann, T., Watt, C. & Datry, T. (2021). Global prevalence of non-perennial rivers and streams. Nature 594, 391–397, https://doi.org/10.1038/s41586-021-03565-5.

Mizukami, N., Clark, M. P., Sampson, K., Nijssen, B., Mao, Y., McMillan, H., Viger, R. J., Markstrom, S. L., Hay, L. E., Woods, R., Arnold, J. R., and Brekke, L. D. (2016). mizuRoute version 1: a river network routing tool for a continental domain water resources applications, Geoscientific Model Development, 9, 2223–2238, https://doi.org/10.5194/gmd-9-2223-2016.

Mizukami, N., Clark, M. P., Newman, A. J., Wood, A. W., Gutmann, E. D., Nijssen, B., Rakovec, O., & Samaniego, L. (2017). Towards seamless large-domain parameter estimation for hydrologic models, Water Resources Research, 53, 8020– 8040, https://doi.org/10.1002/2017WR020401.

Mizukami, N., Clark, M. P., Gharari, S., Kluzek, E., Pan, M., Lin, P., Beck, H. E., & Yamazaki, D. (2021). A vector-based river routing model for Earth System Models: Parallelization and global applications. Journal of Advabces in Modeling Earth Systems, 13, e2020MS002434, https://doi.org/10.1029/2020MS002434.

Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L., Koutsoyiannis, D., Cudennec, C., Toth, E., Grimaldi, S., Blöschl, G., Sivapalan, M., Beven, K., Gupta, H., Hipsey, M., Schaefli, B., Arheimer, B., Boegh, E., Schymanski, S.J., Di Baldassarre, G., Yu, B., Hubert, P., Huang, Y., Schumann, A., Post, D., Srinivasan, V., Harman, C., Thompson, S., Rogger, M., Viglione, A., McMillan, H., Characklis, G., Pang, Z., & Belyaev, V. (2013). "Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022. Hydrological Sciences Journal, 58 (6) 1256–1275.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., & Duan, Q. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, https://doi.org/10.5194/hess-19-209-2015.

NOAA National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters. https://www.ncei.noaa.gov/access/billions/, DOI: 10.25921/stkw-7w73, 2022.

Ohmura, A. (2001). Physical basis for the temperature-based melt-index method. Journal of Applied Meteorology, 40, pp. 753-761.

Paiva, R. C. D., Collischonn, W., & Tucci, C. E. M. (2011). Large scale hydrologic and hydrodynamic modeling using limited data and a GIS based approach. Journal of Hydrology, 406 (3), 170–181. https://doi/org/10.1016/j.jhydrol.2011.06.007.

Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook, Water Resources Research., 51, 7090– 7129, https://doi.org/10.1002/2015WR017780.

Ponce, V. M., Li, R., & Simons, D. B. (1978). Applicability of kinematic and diffusion models, Journal of the Hydraulic Division, 104(HY3), 353– 360.

PRISM Climate Group, Oregon State University, https://prism.oregonstate.edu, data created 4 Feb 2014, accessed 16 Sep 2020.

Quinn, P., Beven, K., Chevallier, P. & Planchon, O. (1991). The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. Hydrological Processes, 5: 59-79. https://doi.org/10.1002/hyp.3360050106.

Rameshwaran, P., Bell, V. A., Brown, M. J., Davies, H. N., Kay, A. L., Rudd, A. C., & Sefton, C. (2022). Use of abstraction and discharge data to improve the performance of a national-scale hydrological model. Water Resources Research, 58, e2021WR029787, https://doi.org/10.1029/2021WR029787

Salas, F. R., Somos-Valenzuela, Marcelo A., Dugger, A., Maidment, D. R., Gochis, D. J., David, C. H., Yu, W., Ding, D., Clark, E. P., & Noman, N. (2018). Towards Real-Time Continental Scale Streamflow Simulation in Continuous and Discrete Space. Journal of American Water Resources Association, 54, 7- 27, https://doi.org/10.1111/1752-1688.12586

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, W05523, https://doi.org/10.1029/2008WR007327.

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., Eisner, S., Müller Schmied, H., Sutanudjaja, E. H., Warrach-Sagi, K., & Attinger, S. (2017). Toward seamless hydrologic predictions across spatial scales, Hydrology and Earth System Science, 21, 4323–4346, https://doi.org/10.5194/hess-21-4323-2017.

Savenije, H. H. (2009). HESS Opinions" The art of hydrology". Hydrology and Earth System Science, 13(2), 157-161.

Shaad, K. (2018). Evolution of river-routing schemes in macro-scale models and their potentials for watershed management, Hydrological Sciences Journal, 63, 1062-1077. https://doi.org/10.1080/02626667.2018.1473871

Shen, X., Hong, Y., Zhang, K., & Hao, Z. (2017). Refining a distributed linear reservoir routing method to improve performance of the CREST model. Journal of Hydrological Engineering, 22(3), 04016061.

Solvik, K., Bartuszevige, A. M., Bogaerts, M., & Joseph, M. B. (2021). Predicting playa inundation using a long short-term memory neural network. Water Resources Research, 57, e2020WR029009. https://doi.org/10.1029/2020WR029009.

Tang, G., Zeng, Z., Long, D., Guo, X., Yong, B., Zhang, W., & Hong, Y. (2016). Statistical and Hydrological Comparisons between TRMM and GPM Level-3 Products over a Midlatitude Basin: Is Day-1 IMERG a Good Successor for TMPA 3B42V7?, Journal of Hydrometeorology, 17(1), 121-137.

Tarboton, D. G. (1997). A new method for the determination of flow directions and upslope areas in grid digital elevation models, Water Resources Research, 33(2), 309– 319, https:/doi.org/10.1029/96WR03137.

Tavakoly, A. A., Gutenson, J. L., Lewis, J. W., Follum, M. L., Rajib, A., LaHatte, W. C., & Hamilton, C. O. (2021). Direct integration of numerous dams and reservoirs outflow in continental

scale hydrologic modeling. Water Resources Research, 57, e2020WR029544. https://doi.org/10.1029/2020WR029544.

Tellman, B., Sullivan, J.A., Kuhn, C. Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A. & Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. Nature 596, 80–86, https://doi.org/10.1038/s41586-021-03695-w.

Tijerina, D., Condon, L., FitzGerald, K., Dugger, A., O'Neill, M. M., Sampson, K., Gochis, D., & Maxwell, R. (2021). Continental hydrologic intercomparison project, phase 1: A large-scale hydrologic model comparison over the continental United States. Water Resources Research, 57, e2020WR028931. https://doi.org/10.1029/2020WR028931, 2021.

Tokuda, D., Kim, H., Yamazaki, D., & Oki, T. (2021). Development of a coupled simulation framework representing the lake and river continuum of mass and energy (TCHOIR v1.0), Geoscientific Model Development, 14, 5669–5693, https://doi.org/10.5194/gmd-14-5669-2021.

Vanderkelen, I., Gharari, S., Mizukami, N., Clark, M. P., Lawrence, D. M., Swenson, S., Pokhrel, Y., Hanasaki, N., van Griensven, A., & Thiery, W. (2022). Evaluating a reservoir parametrisation in the vector-based global routing model mizuRoute (v2.0.1) for Earth System Model coupling, Geosci. Model Dev. Discuss. [preprint], https://doi.org/10.5194/gmd-2022-16, in review.

Vergara, H., Kirstetter, P., Gourley, J.J., Flamig, Z.L., Hong, Y., Arthur, A., & Kolar, R. (2016). Estimating a-priori kinematic wave model parameters based on regionalization for flash flood forecasting in the Conterminous United States, Journal of Hydrology, 541, 421-433. https://doi.org/10.1016/j.jhydrol.2016.06.011, 2016.

Vrugt, J. A., ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling, International Journal of Nonlinear Science and Numerical Simulation, 10, 273–290, https://doi.org/10.1515/ijnsns.2009.10.3.273, 2009

Wang, X., White-Hull, C., Dyer, S., & Yang, Y.: GIS-ROUT: A River Model for Watershed Planning. Environmental Planning B: Planning and Design, 27(2), 231–246, https://doi.org/10.1068/b2624, 2000

Wang, J., Hong, Y., Li, L., Gourley, J.J., Khan, S.I., Yilmaz, K.K., Adler, R.F., Policelli, F.S., Habib, S., Irwn, D., Limaye, A.S., Korme, T. & Okello L. (2011). The coupled routing and excess storage (CREST) distributed hydrological model, Hydrological Sciences Journal, 56 (1), 84-98.

Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., Doriji, C., & Chen, S. (2013). Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? Journal of Hydrology, 499, 91-99. https://doi.org/10.1016/j.hydrol.2013.06.042.

Yamazaki D, Kanae S, Kim H, & Oki T. (2011). A physically based description of floodplain inundation dynamics in a global river routing model. Water Resources Research, 47: 1– 21. https://doi.org/10.1029/2010WR009726.

Yamazaki, D., Sato, T., Kanae, S., Hirabayashi, Y., & Bates, P. D. (2014). Regional flood dynamics in a bifurcating mega delta simulated in a global river model, Geophysical Research. Letters, 41, 3127– 3135, https://doi.org/10.1002/2014GL059744.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., & Bates, P. D. (2017). A high-accuracy map of global terrain elevations, Geophysical Research Letters, 44, 5844–5853. https://doi.org/10.1002/2017GL072874.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M. (2019). MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset, Water Resources Research, 55, 5053–5073, https://doi.org/10.1029/2019WR024873.

Yang, Y., Pan, M., Lin, P., Beck, H. E., Zeng, Z., Yamazaki, D., David, C. H., Lu, H., Yang, K., Hong, Y., & Wood, E. F. (2021). Global Reach-Level 3-Hourly River Flood Reanalysis (1980–2019), Bulletin of American Meteorological Society, 102(11), E2086-E2105.

Yang, T., Zhang, L., Kim, T., Hong, Y., Zhang, D., & Peng, Q. (2021). A large-scale comparison of Artificial Intelligence and Data Mining (AI&DM) techniques in simulating reservoir releases over the Upper Colorado Region, Journal of Hydrology, 602, 126723.

Chapter 7

The conterminous United States are projected to become more prone to flash floods in a high-emissions scenario

Publication

Li, Z., Gao, S., Chen, M. Gourley, J.J., Liu, C., Prein, A., & Hong, Y. (2022). The conterminous United States are projected to become more prone to flash floods in a high-end emissions scenario. Communications Earth & Environment, 3, 86 (2022). https://doi.org/10.1038/s43247-022-00409-6

7 The conterminous United States are projected to become more prone to flash floods in a high-emissions scenario

Abstract

Flash floods are largely driven by high rainfall rates in convective storms that are projected to increase in frequency and intensity in a warmer climate in the future. However, quantifying the changes in future flood flashiness is challenging due to the lack of high-resolution climate simulations. Here we use outputs from a continental convective-permitting numerical weather model at 4-km and hourly resolution and force a numerical hydrologic model at a continental scale to depict such change. As results indicate, US floods are becoming 7.9% flashier by the end of the century assuming a high-emissions scenario. The Southwest (+10.5%) has the greatest increase in flashiness among historical flash flood hot spots, and the central US (+8.6%) is emerging as a new flash flood hot spot. Additionally, future flash flood-prone frontiers are advancing northwards. This study calls on implementing climate-resilient mitigation measures for emerging flash flood hot spots.

7.1 Introduction

Floods in the United States are the most devastating water-related natural hazards (Gourley et al., 2013; Hirabayashi et al., 2013; Khajehei et al., 2020; Li et al., 2021; Merz et al., 2021; Saharia et al., 2017). According to the National Weather Service storm reports, floods have cost more than 159 billion US dollars and 2000 fatalities from 1996 to 2020 (last access: July 27, 2021), let alone immeasurable damage to ecosystems. Among all types of floods, flash floods are one of the most devastating subsets, accounting for nearly half of those economic losses (Ahamadalipour & Moradkhani, 2019; Gourley et al., 2017; Saharia et al., 2017). The central question for the hydrologic science community, practitioners, and more urgently the weather service agencies becomes: How do future floods evolve under a changing climate (Fowler et al., 2021)?

Flood risks are largely tied to increased precipitation extremes induced by a warmer climate in addition to more direct anthropogenic effects (Swain et al., 2020; Tabari, 2020). Although the Clausius-Clapeyron equation infers a theoretical 7% increase in atmospheric water holding capacity per degree Celsius warming (Allen & Ingram, 2002), extreme precipitation might be increasing even more due to changes in storm structure (Prein et al., 2017), storm dynamics (Westra et al., 2014), and large-scale weather patterns (Folwer et al., 2021). Global Climate Models (GCMs), which are mainly designed for simulating large-scale climate variables, poorly represent mesoscale weather systems because of their coarse resolutions (>10 km) and insufficient parameterization schemes (Zhang et al., 2020; Prein et al., 2015). GCM-based climate change-related flood studies are limited in assessing changes in fine-scale flood dynamics that are particularly important for flash flooding (Hirabayashi et al., 2013; Swain et al., 2020; Bates et al., 2020). In contrast, convection-permitting models that operate at kilometer-scale grid spacing (Prein et al., 2015; Clark et al., 2016) start to resolve convective processes resulting in largely improved simulations of sub-daily extreme precipitation rates (Prein et al., 2015; Clark et al., 2016; Wehner et al., 2016; Younis et al., 2008). Convection-permitting models, therefore, offer a valuable alternative for improved understanding of climate change impacts on flash flooding (Younis et al., 2008).

In this study, hourly data from a 13-year retrospective convection-permitting climate simulation with 4-km grid spacing and a 13-year future climate counterpart are used to investigate underlying future changes in flash floods. The retrospective simulation (also denoted as control run; CTL) downscales the ERA-Interim reanalysis (Dee et al., 2011) over the period from October

2000 to September 2013. The future simulation uses the Pseudo Global Warming approach (PGW) (Schär et al., 1996; Rasmussen et al., 2014) by adding climate change perturbations to the ERA-Interim boundary conditions and applies to the same period (2000-2013). Those perturbations are derived from an ensemble mean of 19 CMIP5 (Coupled Model Intercomparison Project 5) GCMs under a high emission scenario (Representative Concentration Pathway: RCP8.5) during the period 2071 - 2100, compared to the reference period 1976 - 2005. The PGW simulation focuses on assessing more deterministic thermodynamic climate change impacts and does not allow studying the impacts of more uncertain changes in large-scale weather patterns (Liu et al., 2017). It is noteworthy that the historical synoptic scale precipitation events are reproduced both in CTL and PGW owing to the use of spectral nudging (Liu et al., 2017).

Several studies have verified the retrospective simulation (CTL) against precipitation climatology[21], the simulation of hurricanes (Gutmann et al., 2018), atmospheric rivers (Dougherty et al., 2020), snowmelt (Musselman et al., 2018), and extreme precipitation (Prein et al., 2017; Dougherty et al., 2020; Bukovsky, 2011). In addition, we verified the CTL run with respect to the most commonly used precipitation dataset, National Centers for Environment Prediction Stage IV at four frequency-duration levels (i.e., 50-year 6-hour, 50-year 24-hour, 100-year 6-hour, and 100-year 24-hour events). For example, a 50-year term refers to frequency while 6-hour refers to precipitation accumulation intervals. We aggregate our results to 17 “homogeneous” climate regions (the Bukovsky regions (Bukovsky, 2011); Supplementary Figure 7.1) to interpret regional climate differences on future flood changes.

Future flood risks pertaining to flash floods have not been thoroughly investigated under climate change scenarios, although the close relationship between rainfall rates and flood flashiness is well established by hydrometeorologists as a synoptic or mesoscale phenomenon (Maddox et al., 1979). Flood resilience, termed as the ability to mitigate the socioeconomic impacts of floods, has been brought up frequently in the wake of climate change (McClymont et al., 2020). Taking advantage of the benefits of convection-permitting models, we quantify the impacts of climate change on future flood flashiness changes over the conterminous US (CONUS), which conveys information such as flood hazards, exposure, vulnerability, and impacts in the future (Merz et al., 2021). In doing so, we use the CTL and PGW simulations to force a hydrologic model – the Ensemble Framework For Flash Flood Forecasting (EF5) which has been used operationally in the U.S. National Weather Service for flash flood forecasting across the CONUS

and territories since 2017 (Gourley et al., 2017; Flamig et al., 2020; Li et al., 2021; Xue et al., 2013). For the analysis, the rainfall-flood event isolation and association are essentially needed to ensure the dependence between streamflow, precipitation, and antecedent soil moisture (Wasko et al., 2019). Based on simulation outputs from 11 full calendar years (2001-2012) at an hourly time step, we extract flood events (defined by two-year streamflow values determined by CTL) for both CTL and PGW runs. For each flood event, we identify the corresponding rainfall event using the Characteristic Points Method (CPM) (Shen et al., 2017).

This study hopes to provide quantitative assessments on changes in future flood-producing storms and flood flashiness, including geographical shifts in flash flood hotspots. It can be served as a basis for adapting nationwide flash flood planning strategies and calls on implementing climate-resilient mitigation measures for emerging flash flood hotspots.

7.2 Data and Methods

7.2.1 Datasets

To evaluate the performance of retrospective simulations, we curated historical extreme rainfall events archived by the Colorado State University Real-Time Weather Data (http://schumacher.atmos.colostate.edu/precip_monitor/50y24h/event_images/year_list.php). The extreme precipitation thresholds are determined by NOAA Atlas 14, which is a gridded product that is derived by over a hundred-year rainfall gauge dataset and fitted with an assumed distribution. The hourly precipitation data is based on the National Centers for Environment Prediction Stage IV radar-gauge merged product from the University Cooperation for Atmospheric Research/National Center for Atmospheric Research Earth Observing Laboratory. Because Stage IV hourly data availability on the West Coast is generally under 60%, the evaluation extent is confined to the Midwest and the eastern U.S. There is a systematic bias (>10%; underestimation) for precipitation between Stage IV and CTL, partly due to imperfect physical schemes, boundary feedbacks for regional climate model (RCM) simulations, and climate variability (Ehret et al., 2012).

The Global Streamflow Characteristics Dataset, which provides extreme streamflow values at daily resolution, is used to verify our hydrologic simulations (Beck, 2016). This dataset is a machine-learning-aided product, training catchment streamflow characteristics based on a set of climate and physiographic features with historical data from 9169 USGS stream gauges (Beck et

al., 2015). The produced results have been assessed against five global hydrologic models for further independent evaluation. A negative systematic bias (>8%) for surface runoff persists as propagated from the aforementioned precipitation bias, but other contributing factors within the hydrologic model simulations (i.e., calibrated parameters based on different forcing datasets, uncertainties in hydrologic model structure) slightly compensate for the precipitation bias.

7.2.2 Climate simulation

Two 4-km and hourly datasets from climate simulations are analyzed and used as forcing for hydrologic models. Such simulations are created with the community Weather Research and Forecast (WRF) model Version 3.4.1 (see Liu et al. (2017) for reference), encompassing the whole CONUS and portions of Canada and Mexico. The retrospective simulation (CTL) from 1 October 2000 to 30 September 2013 downscales ERA-Interim reanalysis data (Dee et al., 2011), with spectral nudging being applied. The future climate simulation uses the Pseudo Global Warming approach (PGW) (Schär et al., 1996, Rasmussen et al., 2014) by perturbing the climatic boundary conditions (horizontal wind, geopotential heights, temperature, specific humidity, sea surface temperature, soil temperature, sea level pressure, and sea ice). An ensemble mean of climate signals from 19 CMIP5 models for the period 2070 to 2100, relative to the reference period 1976 to 2005, is added to ERA-Interim boundary conditions under RCP8.5. In principle, the essence of the PGW run is to assess thermodynamic climate change signals in contrast to changes in large-scale weather patterns such as shifts of the large-scale storm tracks and North Atlantic jet stream (Cook et al., 2103). The benefits of PGW simulations are that systematic GCM errors are mostly excluded in the downscaled simulation, and the effects from climate internal variability on climate change signals can be thus ignored. Previous model comparisons between PGW run and CMIP5 reveal similar climate change patterns over the CONUS (Liu et al., 2017).

7.2.3 Hydrologic models

In this study, the simulated retrospective and future precipitation events are taken separately as forcing to drive a hydrologic model. We use the widely recognized flash flood forecast model – the Ensemble Framework For Flash Flood Forecasting (EF5), developed jointly at the University of Oklahoma and NOAA National Severe Storms Laboratory (Flamig et al. 2020). The EF5 model has been in operation for real-time flash flood forecast across the CONUS

and territories at 1-km spatial resolution and updates every 10 min (Gourley et al., 2017). Forecasters in the U.S. National Weather Services (NWS) utilize this product to issue flash flood warnings. The model is developed across operating systems (Linux, macOS, and Windows) and is publicly available at https://github.com/HyDROSLab/EF5.

In this study, we choose the subset hydrologic model - Coupled Routing and Excess STorage (CREST) V2.1 (Flamig et al., 2020) and routing model – kinematic wave within the EF5 framework to simulate streamflow at 1km spatial resolution and hourly temporal resolution from 2001 to 2011. The nearest neighbor method is used to downscale precipitation data at 4km to be consistent with topographic resolution at 1km. The distributed parameters for the CREST and the kinematic wave model remain the same as the operational runs. To separate errors from hydrologic models and forcing input, we similarly run EF5 using Stage IV as precipitation inputs for the same period (2001-2011). From the previous model evaluation, this model is generally suited for flash floods or heavy-rainfall-induced floods while not modeling snowmelt events or rain-on-snow events (Flamig et al., 2020).

7.2.4 Hydrograph separation and rainfall-flood event association

Prior to comparisons, the time series of outputs from hydrologic models (i.e., streamflow and basin-average rainfall rates) are extracted at each HUC8 basin over the CONUS. Then, flood events are identified according to the following two criteria. First, peak streamflow should exceed the threshold for a two-year flood, typically considered as a bankfull condition (Bates et al., 2020). The two-year flood is determined by fitting annual maximum streamflow into a log-Pearson Type III distribution, which is a conventional method used by the U.S. Geological Survey, and extracting values at two-year exceedance. Second, to specifically focus on flashier floods, we set a threshold for the flood rising slope, peak streamflow divided by rising time. This threshold is computed as a function of upstream flow accumulation values and is empirically suggested in Chow et al. (1988), meaning that large basins allow more time to be considered as flash floods than small basins. For rainfall-flood event association, we use the Characteristic Point Method, which has been applied to compile a comprehensive flood database over the CONUS (Shen et al., 2017). The Characteristic Point Method firstly separates storm event flow and base flow with the filtered revised constant k method, which is a hybrid method of the revised constant k and recursive digital filter. Secondly, event identification is achieved through linking separated flow events and rainfall

events within a search window. For a detailed description and application of the method, the reader is referred to Mei & Anagnostou (2015). Some event signatures are exported to describe the matched rainfall and flood events, such as rainfall duration (D_r), peak rainfall rate (R_p), rainfall volume (V_r), flood duration (D_f), peak flow rate (Q_f), flood volume (V_f), rainfall-flood lag time (Lag). On top of these signatures, we calculate the flood flashiness index per event with respect to flood rising time and peak flow rate as shown in Eq. 7.1 (Saharia et al., 2017).

$$Flashiness_{s,i,j} = \frac{Qp_{s,i,j} - Qb_{s,i,j}}{A_i T_{s,i,j}} \quad (7.1)$$

where the subscripts s, i, j indicate each scenario (CTL or PGW), HUC8 basin, and flood event; Q_p and Q_b refer to the peak flow and base flow, and A and T are the drainage area and flood rising time. The baseflow is calculated following the Characteristic Point Method, which uses filtered revised constant k method – a combination of revised constant k and recursive digital filter – to separate storm flow and baseflow. To scale the flashiness index values between 0 and 1, we fit a collection of flashiness values with an empirical cumulative distribution function (ecdf). Then the basin-level flashiness index is calculated as the median value of all event-based flashiness values within each basin. Results of flashiness index calculated using our model outputs are compared to the study of Saharia et al. (2017) at U.S. Geological Survey stream gauge locations (Supplementary Figure 7.5).

7.3 Results

The results shown in the Supplementary Figure 7.2 corroborate the satisfactory performance of CTL, as compared to Stage IV, with the correlation coefficient above 0.6. The simulated runoff by EF5 is compared with the community dataset - Global Streamflow Characteristic Dataset (Beck, 2016) at three percentiles (Quantiles 90, Quantiles 95, and Quantiles 99) (Supplementary Figures 7.3 and 7.4), which verifies the efficacy of our model results. The rainfall-flood event association and calculation of the flashiness index (definition detailed in Method) and other characteristics are shown in Fig. 7.1a. The frequency change in Fig. 7.1b highlights that the US basins (e.g., the Southwest, Great Plains, and Prairie) - characterized by an intermittent streamflow regime, which is dominated by weak seasonality and local floods due to episodic precipitation events related to thunderstorms or fronts (Brunner et al., 2020) - will likely experience more floods in the PGW simulation. Small-to-medium size US rivers have a diverse

response to warmer climates as their flood frequency changes range from -50 % to 350 % (Fig. 7.1c). Large river basins, however, tend to be more stable with reduced uncertainty ranges. With that being said, small basins have more closely linked precipitation and streamflow, while floods in large basins are more likely to be modulated by antecedent basin wetness (Ivancic & Shaw, 2015).

Despite a general trend of increasing extreme precipitation at the end of this century in the PGW run (Dougherty & Rasmussen et al., 2020), changes in the flood-producing event precipitation differ across climate divisions (Fig. 7.2). Significant positive percent changes (PC) in extreme precipitation are found across the Rockies (d and h) (PC_{mean}=34.5%; $PC_{99\%}$=28.5%) and Appalachia-Atlantic regions (p, n, and q) (PC_{mean}=24.7%; $PC_{99\%}$=21.9%). Flood events in these regions are typically related to spring snowmelt and rain-on-snow events (Li et al., 2021; Dougherty et al., 2020; Villarini, 2016; Kunkel et al., 2013). The flood-producing storms in the future, however, are likely associated with rainfall excess that is favored in a warming climate rather than snowfall (Dougherty & Rasmussen, 2020). For these regions, future flood frequencies are nearly doubled (Fig. 7.1b).

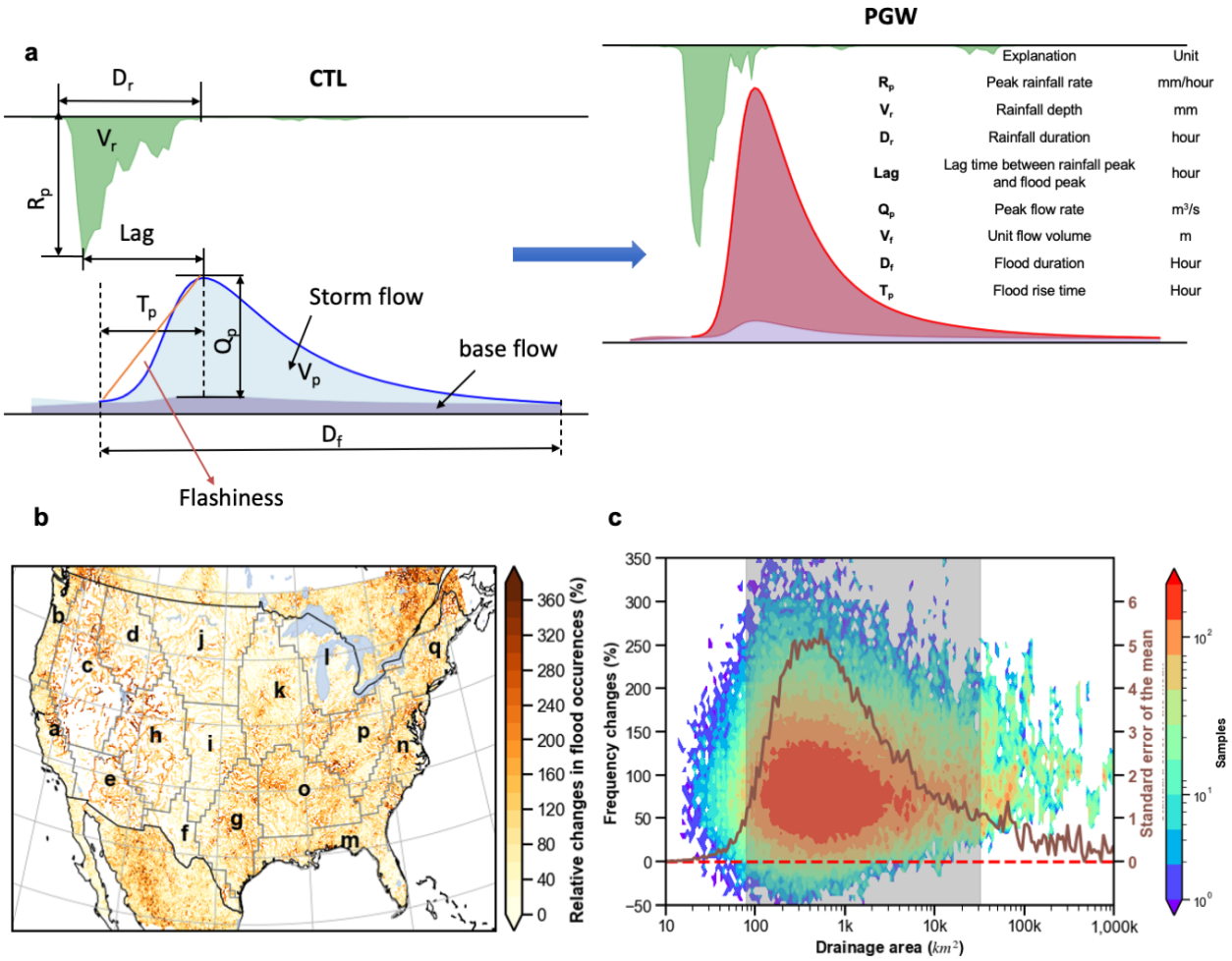


Figure 7.1 Schematic view of rainfall-runoff event association at US HUC8 catchment scale. (a) illustration of calculated flood characteristics; (b) the percentage change of flood occurrences comparing the future (PGW) and retrospective analysis (CTL) at 1-km spatial resolution; (c) conditional plot of frequency changes against drainage area (shaded contour plot) and standard error of the mean in dark red line. Maps and figures are produced using the Python package Matplotlib and Cartopy.

Water vapor availability in the atmosphere in the central and eastern US plays an essential role in contributing to extreme precipitation changes (Tabari, 2020; Westra et al., 2014; Kunkel et al., 2013). Additionally, extreme precipitation events are typically caused by specific storm types, e.g., (extra)tropical cyclones (Gutmann et al., 2018; Kunkel et al., 2013), mesoscale convective systems (Prein et al., 2017; Diffenbaugh et al., 2013), and the North American Monsoon (NAM) (Kunkel et al., 2013). The 26.6% increase of extreme precipitation in the eastern US agrees well

within previous estimates of a 25%-40% increase in mesoscale convective event rainfall (Prein et al., 2017), which translates to 104% more frequent floods in the future (Fig. 7.2).

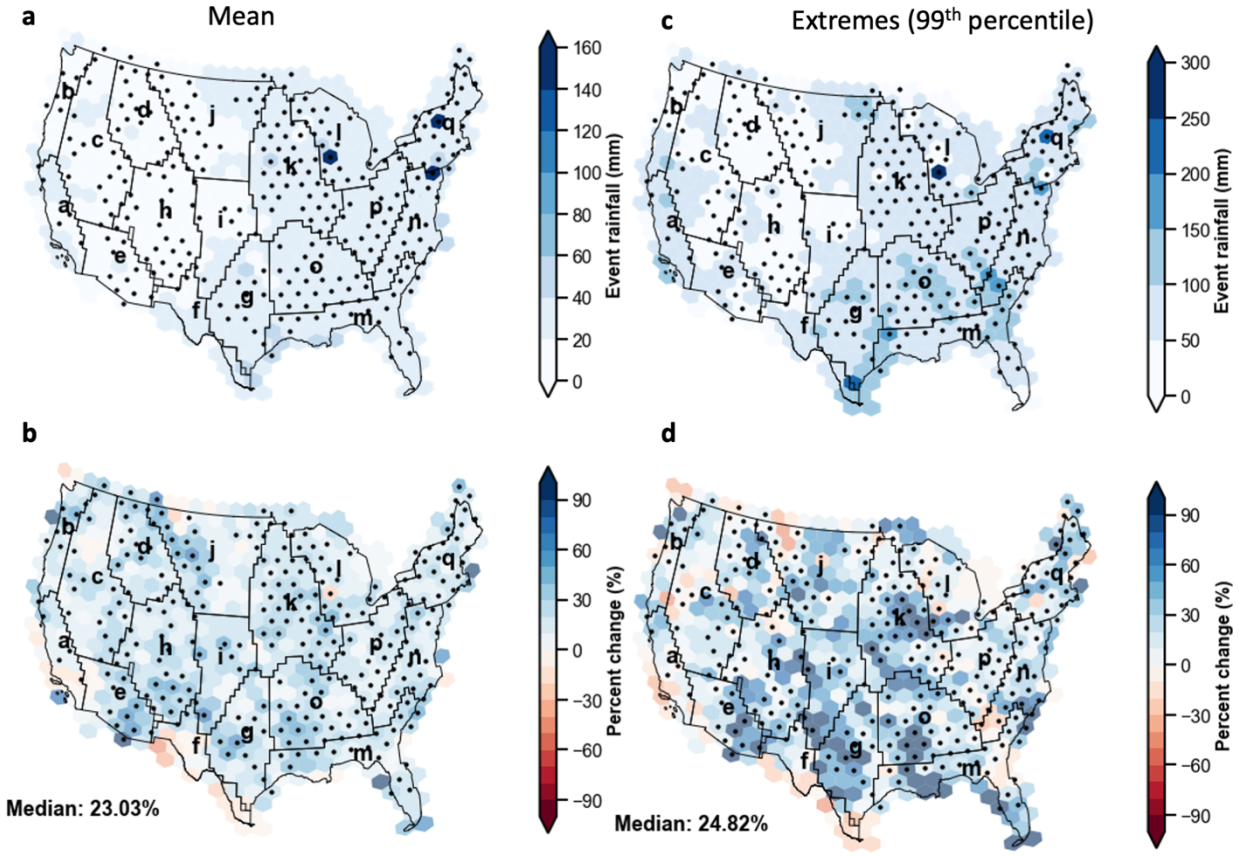


Figure 7.2 Future increase in two-year flood-producing mean and 99th percentile event rainfall (mm) at 2.5-deg spatial resolution, grouped by Bukovsky climate divisions.

(a) Retrospective mean event rainfall, (b) Percent changes (pseudo global warming minus retrospective simulation), (c) Retrospective extreme event rainfall (99th percentile), and (d) Percent changes (pseudo global warming minus retrospective simulation) in extreme event rainfall. The dots highlight significant changes with Kolmogorov-Smirnov two-side test for p-values less than 0.05. Maps are produced using the Python package Matplotlib and Cartopy.

The largest change in flash-flood-producing precipitation occurs in the Southwest (e) (PC_{mean}=+40.4%; $PC_{99\%}$=51.1%), where the main forcing agent for extreme precipitation is the NAM. The warm, moist air emanating from the Gulf of California fuels the storms and supports the "it never rains, but it pours" pattern in the arid climates, thereby leading to destructive flash floods (Khajehei et al., 2020; Saharia et al., 2017; Tabari, 2020). Although CMIP5 ensemble members project a weakening effect of NAM (Cook & Seager, 2013), its extreme precipitation

rates are still likely to increase (Prein et al., 2017). Correspondingly, future floods are becoming 123% more frequent there, representing the largest increase than other regions.

On the other hand, flood-producing extreme precipitation in the Pacific Southwest (a) and Northwest (b), Mezquital (f) regions is relatively insensitive to climate change, with only a few grid cells showing significant changes. Flood frequencies there increase by 67.1%, half of the probability of the aforementioned sensitive regions. Extreme precipitation is even projected to largely decrease in Mezquital (f) for both mean and extreme conditions (Figs. 7.2b and 7.2d), where the thermodynamic changes are likely limited by atmospheric moisture availability.

Consistent decreases of rainfall and flood event durations (D_r and D_f) are found across the CONUS climate divisions (Wasko et al., 2021), measured by the fraction of events with negative changes (Fig. 7.3b, c), among which the Southern Plains (g; 55.8% and 56.7% of the total events), Deep South (o; 54.1% and 59.9% of the total events) and Southeast (m; 56.6% and 59.1% of the total events) have above-normal negative change events (52.1% and 52.4%) in the future. The shortened duration of rainfall and flooding could be partly explained by intense yet faster-moving storms in these regions (Prein et al., 2017). In contrast, the majority of events with positive changes of peak rainfall rates and flow rates (R_p and Q_p) are located in the western US (Figs. 7.3b and 3c), i.e., Pacific Southwest (a; 79.6% and 75.9% of the total events), Pacific Northwest (b; 83.2% and 80.6% of the total events), Great Basins (c; 79.1% and 80.2% of the total events), and Northern Rockies (d; 80.8% and 80.2% of the total events). Similarly, these regions also see positive changes in rainfall and flood volumes (V_r and V_f, a – 67.2% and 68.1%; b – 72.6% and 72.3%; c – 70.8% and 72.8%; d – 69.9% and 71.6%). The future 7.5% decrease of lag time between rainfall and flood events is largely associated with antecedent soil moisture conditions and/or precipitation intensities, as the wetter soils and/or heavier precipitation rates accelerate flood-rise time (Merz et al., 2021; Li et al., 2021). Specifically, humid regions with higher soil moisture content (e.g., Mid-Atlantic and Southeast) in the eastern US have a higher fraction of events (65.1%) with reduced lag time than arid regions (e.g., Southwest and Mezquital; 60.3%), pointing to flashier hydrographs. Combined with factors such as peak flow rates (R_f) and time lag (lag), the flashiness index (flashiness) in the future is increased by 7.9% over the CONUS.

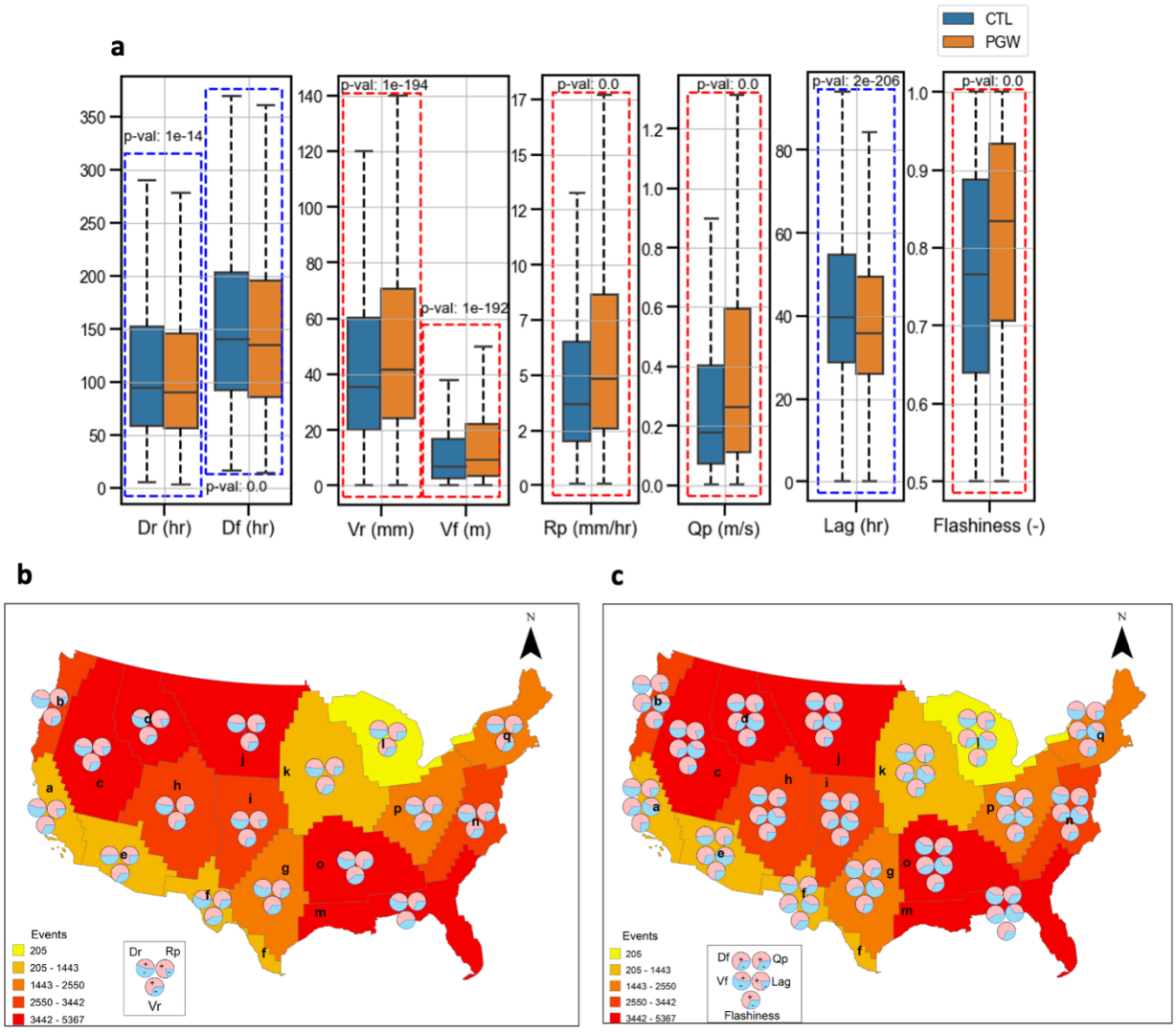


Figure 7.3 Future rainfall/flood durations and rainfall-flood lag time are decreasing while rainfall/flood peak values, volumes, and flashiness index are increasing.

(a) Boxplot of CONUS-wide statistics of rainfall-flood characteristics where blue dotted line indicates a future decrease while red dotted line indicates a future increase; (b) maps of the fraction of increase (red) or decrease (blue) for rainfall characteristics (Dr – duration, Rp – peak rates, and Vr - volume) at Bukovsky climate divisions; (c) map of the fraction of increase (red) or decrease (blue) samples for flow characteristics (Df - duration, Vf - volume, Qp – peak rates, Lag – rainfall-flood-peak lag time, and Flashiness) at Bukovsky climate divisions. Shaded colors in (b) and (c) indicate the sample sizes.

The blue (red) dashed box indicates a significant decrease (increase) of each feature (p-value less than 0.05). The significance test is conducted by Kolmogorov-Smirnov test. Maps are produced using the ArcGIS.

The present-day flashiness index distribution at the catchment scale (HUC8, Fig. 7.4a) is closely related to topography, climatology, river networks, and event characteristics (e.g., antecedent soil moisture, peak rainfall, and duration). The West Coast (1), Arizona (2), Rockies (3), Flash Flood Alley (4), and Appalachia (5) emerge as historical flash-flood hot spots (flashiness index > 0.8). We found similar clusters to the study of Sahari et al. (2017), in spite of a different approach. At the end of this century, the flashiness values are projected to increase across 87.4% of the HUC8 basins in the CONUS (Figure 7.4b). Outside of the NAM-impacted region of Arizona, other flash-flood hot spots have a moderate increase of flashiness indices (4.1%). For these basins, the scaling rates, percent of flashiness index increase per warming temperature degree Celsius (averaged in each climate division), are close to zero (Fig. 7.4c). We suspect that the flashiness increase to a large extent is constrained by the physical limits of the basin geomorphology and land cover. For Arizona, however, the increase of flood-producing storms (Fig. 7.2) contributes to an increase in flashiness indices (+10.5%). In addition, the central US (e.g., Great Plains – j, i, g and Prairie - k), which used to be less prone to flash floods, has on average an 8.5% increase in flashiness indices, posing a potential threat to future floodwater management. In particular, the Prairie and Deep South will transition to hot spots in the future (Fig. 7.4b). We identify these regions (e.g., Southwest (+10.5%), Prairie (9.5%), Great Plains (+9.0%)) as climate sensitive regions to flash floods. Flood risk management including hard measures and soft measures is challenged by unawareness to potential flash flood risks. Interestingly, positive flashiness changes are advancing northward (regions d, j, k, and l).

To link the changes in extreme precipitation and flashiness, Fig. 7.4d depicts the relationships among three variables – peak rainfall rates (R_p), rainfall depth (V_r), and rainfall duration (D_r). As expected, rainfall duration is weakly tied to the flashiness index because it is not the determining factor to either flood rising time or peak flow rates. Yet, both peak rainfall rates and rainfall volumes are positively correlated with flashiness. These two variables share a very similar rate of change (Fig. 7.4d) – around a 100% increase of extreme rainfall leads to a 10% increase of flashiness indices. When extreme rainfall change exceeds more than 120%, the flashiness index reaches a plateau. The plateau represents the physical limit of maximal channel conveyance which is jointly determined by catchments' geographical properties, including channel morphology, slope, land surface roughness, and soil properties (Berghuijs et al., 2016; Villarini, 2016).

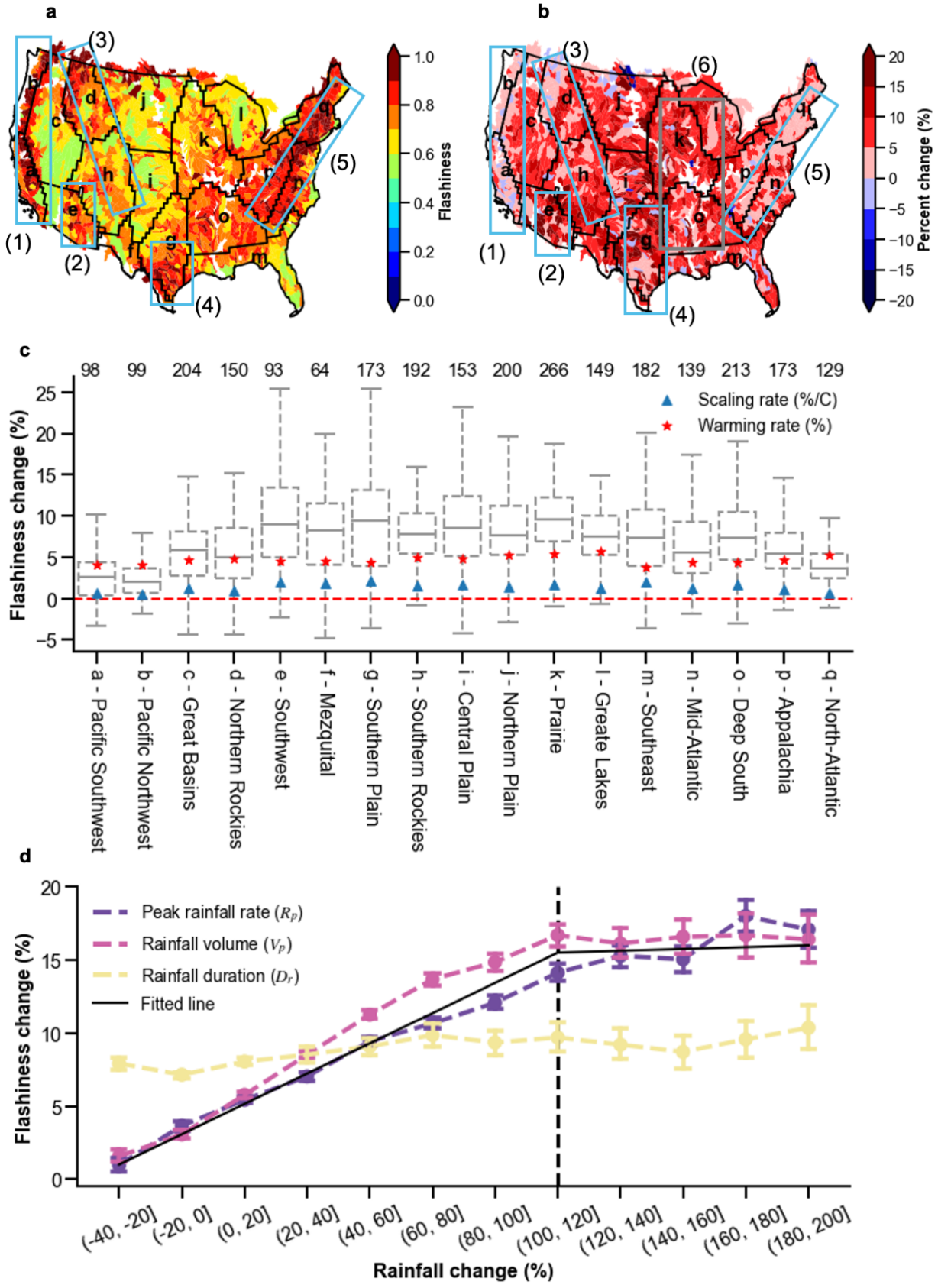

a
b
Flashiness
Percent change (%)
c
Flashiness change (%)
Scaling rate (%/C)
Warming rate (%)
98 99 204 150 93 64 173 192 153 200 266 149 182 139 213 173 129
a - Pacific Southwest
b - Pacific Northwest
c - Great Basins
d - Northern Rockies
e - Southwest
f - Mezquital
g - Southern Plain
h - Southern Rockies
i - Central Plain
j - Northern Plain
k - Prairie
l - Greate Lakes
m - Southeast
n - Mid-Atlantic
o - Deep South
p - Appalachia
q - North-Atlantic
d
Peak rainfall rate (R_p)
Rainfall volume (V_p)
Rainfall duration (D_r)
Fitted line
Flashiness change (%)
Rainfall change (%)

Figure 7.4 Future flashiness indices will increase by 7.9%, while Southwest being affected the most (10.5%).

(a) present-day flashiness indices driven by the retrospective analysis (blue box encompass historical flash flood hot spots identified from Sahari et al.7); (b) future flashiness changes of the difference of PGW run and CTL run (blue box encompasses historical flash flood hot spots, and grey box shows the emerged future flood hot spot); (c) changes of flashiness indices for all climate divisions, accompanied by warming rate and scaling rate; (d) plot regarding binned rainfall changes (including rates – purple line, volumes – red line, and duration – yellow line) on the x axis versus flashiness indices changes on the y axis. Maps and figures are produced using the Python package Matplotlib and Cartopy.

This study conveys some similar conclusions found in other studies. For instance, decreasing trends in the duration of extreme rainfall and floods has been widely reported using either historical observations or climate simulations (Prein et al., 2017; Wasko et al., 2021). High confidence can be obtained for increases in extreme precipitation amounts across the continental U.S., but there is variability due to model uncertainties and changes in regional weather systems. However, disagreements in streamflow responses to these precipitation changes are found across studies. For instance, a range of studies indicates decreasing trends in low-end flood frequencies based on historical observations, in which they deemphasize the dependence of floods on precipitation but rather emphasize the importance of antecedent catchment states (Sharma et al., 2018; Wasko et al., 2019). Even within climate model simulations, results can vary considerably over a given region, provided uncertainties in climate models, hydrologic model structures, and/or parameterizations (Hirabayashi et al., 2013; Tabari, 2020; Clark et al., 2016). Our results differ from some other simulations regarding flood magnitudes because firstly we use the worst-case emissions scenario – RCP 8.5 - that dramatically warms and invigorates the atmosphere. Second, unlike both dynamic and thermodynamic evolution of weather patterns in GCMs, our PGW scheme only permits investigation of the less uncertain component of thermodynamic changes to the atmosphere. Third, the hourly and 4-km resolution model simulations resolve convective-scale weather phenomena that are only parameterized in GCMs. Precipitation simulations are more realistic, conditioned on the thermodynamic changes associated with RCP 8.5, and thus can be applied in a hydrologic modeling context. To be noted, we only investigated the rainfall indices based on rates, volume, time-to-peak, and duration. Other proper indices such as ratios of peak rainfall rates and volumes could be meaningful to interpret the dynamic evolution of rainfall storms.

This work by no means intends to deliver an exhaustive depiction of future floods. The PGW approach is based on a high emissions scenario (RCP 8.5), which may or may not be realized in the future. Furthermore, anthropogenic impacts such as Land Use Land Cover type changes and river regulations are not considered in the modeling settings. One of the reasons is the uncertain prediction of such processes, even though some standards are being developed as a community effort[51]. The Land Use Land Cover changes especially for urban areas could even worsen the future situation by reshaping and magnifying hydrologic responses – less infiltration and more surface runoff[52]. It is also reported that urbanization exacerbates extreme rainfall due to increased aerosols and altered circulations from urban heat island. Therefore, our results could serve as a basis or benchmark if not the worst-case scenario.

In summary, there is a pressing need to understand how future flood flashiness responds to climate change for early mitigation and adaptation measures. Here we present results that translate rainfall-flood event changes from convection-permitting climate simulations to changes in flood flashiness. Four main findings presented in this study include:

1. A CONUS-wide 7.9% increase in future flood flashiness means that future floods will onset more rapidly with higher peak runoff resulting in even shorter opportunities for early warning.

2. Regions outside of the North American Monsoon have moderate increases in flashiness (+4.1%) in a warmer climate. However, floods in the Southwest become much flashier (+10.5%).

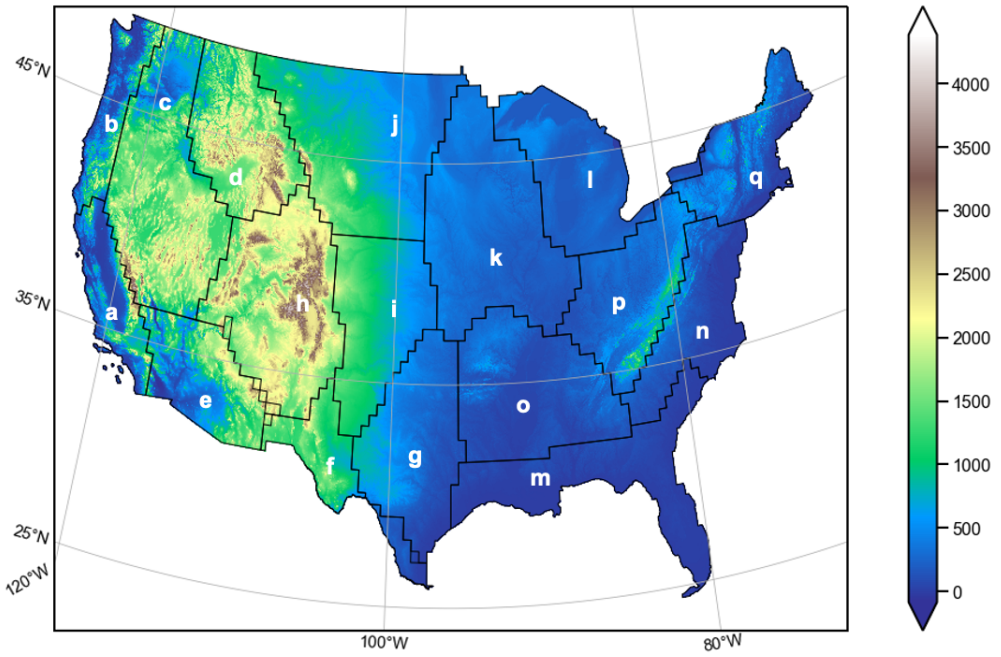
3. The central US have an on average an 8.5% increase in flood flashiness. Particularly, basins in the Deep South and Prairie will transition into new flash flood hot spots in a warmer, future climate.

4. Future flash flood risks are advancing northwards (Northern Rockies, Northern Plains, and Prairie), which poses challenges to local flood resilience measures.

This study shows a pressing need to adapt to climate change and mitigate future flood risks because of an overall increase of flashiness over the CONUS, resulting in less response time for local communities to react. Reassessment of current floodwater management and planning for future flood risk is necessary for judging current standards[53]. More critically, emerging flash flood hot spot regions will be facing unprecedented challenges because of the historically unpreparedness of flood impacts in conjunction with aging infrastructure and outdated flood risk measures (Wright et al., 2019). Potential future studies can examine the seasonal cycles and spatial

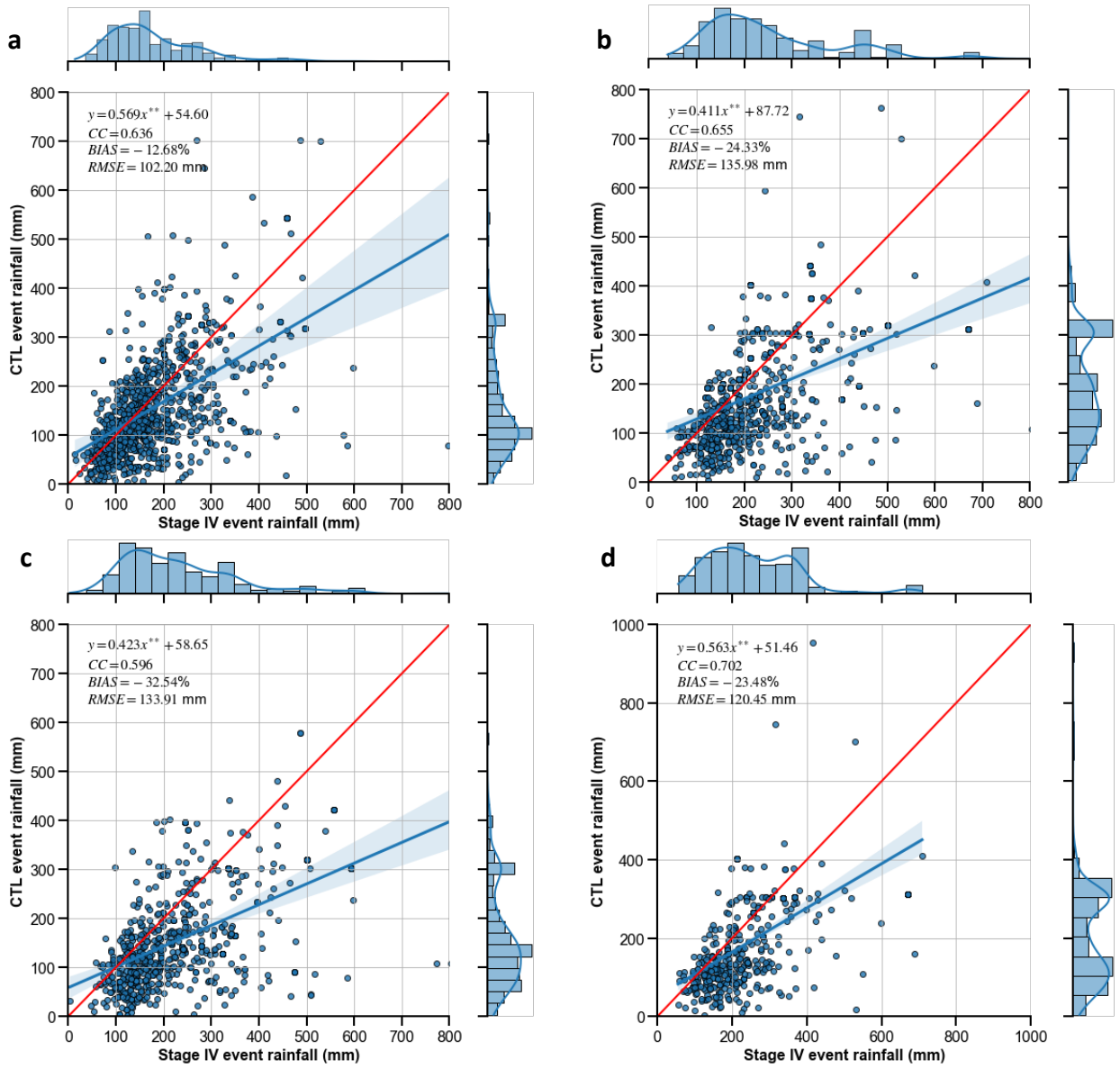
extent of both storms and floods to reveal the spatiotemporal correlation between the two. In addition, it is critical to consider more anthropogenic influences on floods such as river controls and increased urbanization (Blum et al., 2020; Wyżga et al., 1997).

7.4 Supplementary Figures

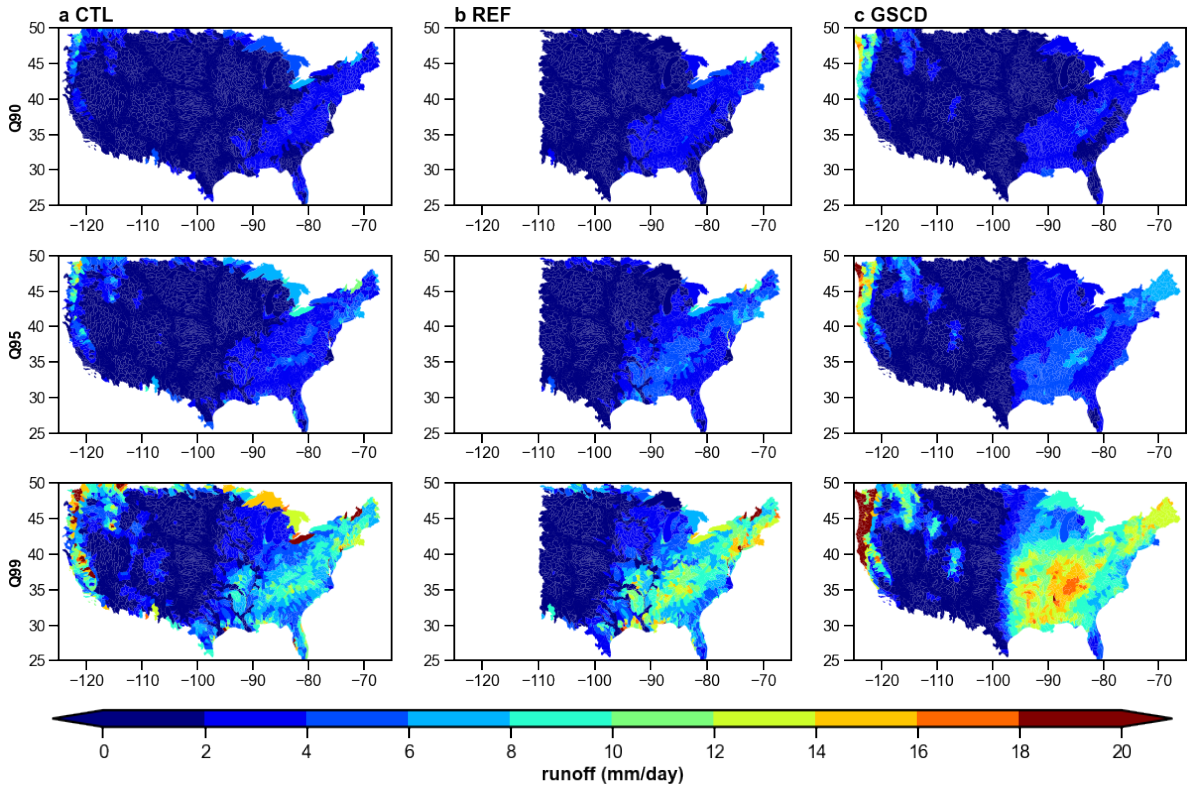


a: Pacific Southwest	d: Northern Rockies	g: Southern Plains	j: Northern Plains	m: Southeast	p: Appalachia
b: Pacific Northwest	e: Southwest	h: Southern Rockies	k: Prairie	n: Mid-Atlantic	q: North Atlantic
c: Great Basin	f: Mezquital	i: Central Plains	l: Great Lakes	o: Deep south	

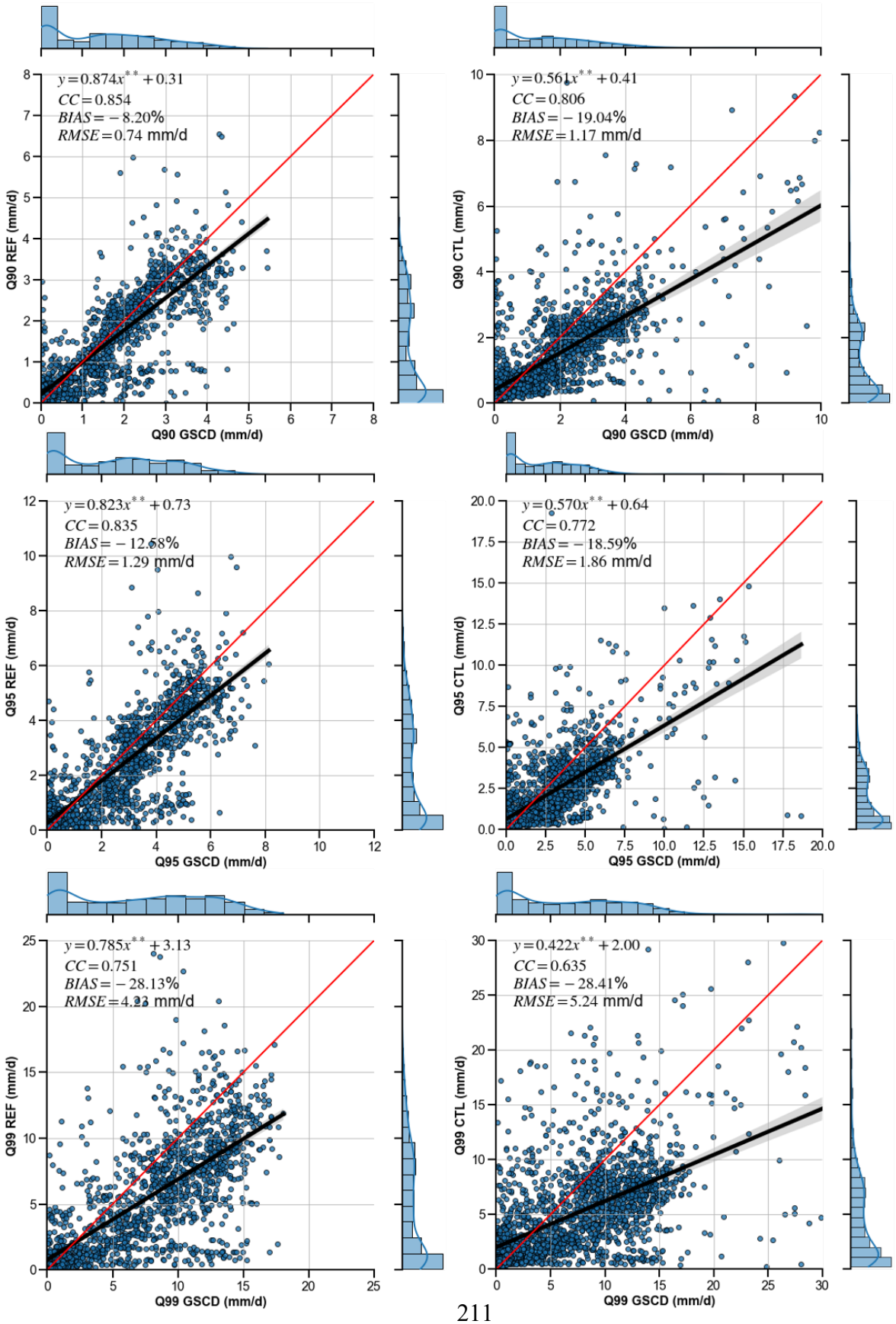
Supplementary Figure 7.1. Study area with Bukovsky climate divisions. Figures are produced using the Python package Matplotlib and Cartopy.



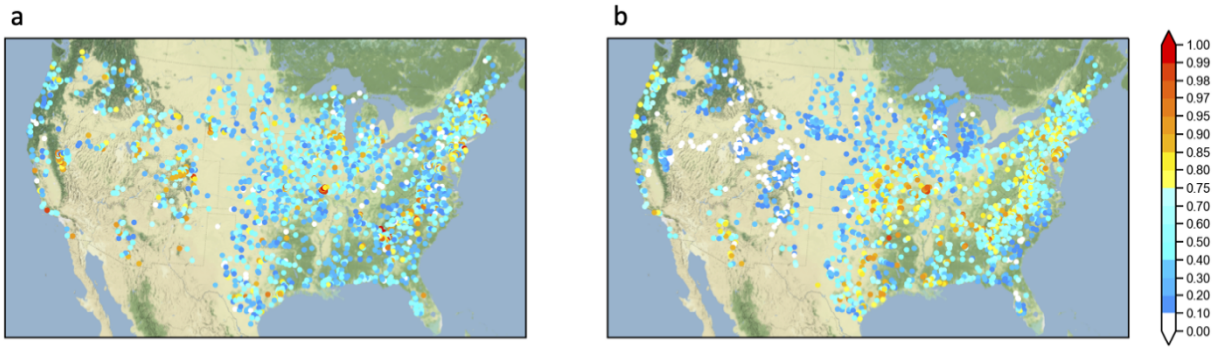
Supplementary Figure 7.2. Scatter plots of retrospective analysis (CTL) against Stage IV data. (a) 50-year 6-hour event; (b) 50-year 24-hour event; (c) 100-year 6-hour event; (d) 100-year 24-hour event. Figures are produced using the Python package Matplotlib.



Supplementary Figure 7.3. Daily runoff at three extreme quantiles (Q90, Q95, and Q99). (a) results forced by retrospective runs (CTL); (b) results forced by Stage IV (REF); (c) the Global Streamflow Characteristics Dataset (GSCD) (Beck et al., 2016). Figures are produced using the Python package Matplotlib.


y = 0.874x** + 0.31
CC = 0.854
BIAS = − 8.20%
RMSE = 0.74 mm/d
Q90 REF (mm/d)
Q90 GSCD (mm/d)
y = 0.561x** + 0.41
CC = 0.806
BIAS = − 19.04%
RMSE = 1.17 mm/d
Q90 CTL (mm/d)
Q90 GSCD (mm/d)
y = 0.823x** + 0.73
CC = 0.835
BIAS = − 12.68%
RMSE = 1.29 mm/d
Q95 REF (mm/d)
Q95 GSCD (mm/d)
y = 0.570x** + 0.64
CC = 0.772
BIAS = − 18.59%
RMSE = 1.86 mm/d
Q95 CTL (mm/d)
Q95 GSCD (mm/d)
y = 0.785x** + 3.13
CC = 0.751
BIAS = − 28.13%
RMSE = 4.23 mm/d
Q99 REF (mm/d)
Q99 GSCD (mm/d)
y = 0.422x** + 2.00
CC = 0.635
BIAS = − 28.41%
RMSE = 5.24 mm/d
Q99 CTL (mm/d)
Q99 GSCD (mm/d)

Supplementary Figure 7.4. Inter-comparisons of daily runoff for retrospective forcing (CTL), Stage IV forcing (REF), and Global Streamflow Characteristics Dataset (GSCD) (Beck et al., 2016). Figures are produced using the Python package Matplotlib.



Supplementary Figure 7.5. Maps of flashiness at gauge locations: (a) Saharia's method with flood action stage; (b) flood isolation method used in this paper. Maps are produced using Python package Matplotlib and Cartopy. Basemap: Stamen Terrain.

7.5 Reference

Ahmadalipour, A., & Moradkhani, H. (2019). A data-driven analysis of flash flood hazard, fatalities, and damages over the CONUS during 1996–2017. Journal of Hydrology, 578, 124106.

Allen, M. R. & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. Nature, 419, 224–232 (2002).

Barnes, E. A., & Polvani, L. (2013). Response of the midlatitude jets, and of their variability, to increased greenhouse gases in the CMIP5 models. Journal of Climate, 26, 7117–7135.

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., Savage, J., Olcese, G., Neal, J., Schumann, G., Giustarini, L., Coxon, G., Porter, J. R., Amodeo, M. F., Chu, Z., Lewis-Gruss, S., Freeman, N. B., Houser, T., Delgado, M., Hamidi, A., Bolliger, I., McCusker, K. E., Emanuel, K., Ferreira, C. M., Khalid, A., Haigh, I. D., Couasnon, A., Kopp, R. E., Hsiang, S., & Krajewski W. F. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. Water Resources Research, 57, p.e2020WR028673. https://doi.org/10.1029/2020WR028673

Beck, H. E., de Roo, A., & van Dijk, A. I. J. M. (2015). Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments, Journal of Hydrometeorology, 16, 1478-1501 (2015).

Beck, H. (2016). A global map of mean annual runoff based on discharge observations from large catchments. Zenodo. http://doi.org/10.5281/zenodo.44782.

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States, Geophysical Research Letters, 43, 4382– 4390.

Blum, A. G., Ferraro, P. J., Archfield, S. A., & Ryberg, K. R. (2020). Causal effect of impervious cover on annual flood magnitude for the United States. Geophysical Research Letters, 47, e2019GL086480.

Brunner, M. I., Melsen, L. A., Newman, A. J., Wood, A. W., & Clark, M. P. (2020). Future streamflow regime changes in the United States: assessment using functional classification, Hydrology and Earth System Science, 24, 3951–3966.

Bukovsky, M. S. (2011) Masks for the Bukovsky regionalization of North America, Regional Integrated Sciences Collective, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO. Downloaded 2021-07-05. http://www.narccap.ucar.edu/contrib/bukovsky/

Bukovsky, M. S., Gao, J., Mearns, L. O., & O'Neill, B. C. (2021). SSP-based land-use change scenarios: A critical uncertainty in future regional climate change projections. Earth's Future, 9, e2020EF001782.

Chow, V. T., Maidment, D. R., & Mays, L. W. (1988). *Applied hydrology* (McGraw-Hill, New York, 1988).

Clark, P., Roberts, N., Lean, H., Ballard, S.P. & Charlton-Perez, C. (2016). Convection-permitting models: a step-change in rainfall forecasting. Meteorological Applications **23**, 165-181.

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A. W., Fowler, H. J., Prudhomme, C., Arnold, J. R., Brekke, L. D. (2016). Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. Current Climate Change Reports, 2, 55–64.

Cook, B. I., & Seager, R. (2013). The response of the North American Monsoon to increased greenhouse gas forcing, Journal of Geophysical Research: Atmosphere, 118, 1690– 1699.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Blasamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J., Park, B., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J., & Vitart, F., (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society, 137, 553-597.

Diffenbaugh, N. S., Scherer, M., & Trapp, R. J. (2013). Robust increases in severe thunderstorm environments in response to greenhouse forcing, Proceedings of the National Academy of Sciences of the United States of America, 110, 16361-16366.

Dougherty, E., Sherman, E., & Rasmussen, K. L. (2020). Future Changes in the Hydrologic Cycle Associated with Flood-Producing Storms in California, Journal of Hydrometeorology, 21, 2607-2621.

Dougherty, E., & Rasmussen, K. L. (2020). Changes in Future Flash Flood–Producing Storms in the United States, Journal of Hydrometeorology, 21, 2221-2236.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., & Liebert, J. (2012). HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrology and Earth System Science, 16, 3391–3404 (2012).

Flamig, Z. L., Vergara, H. & Gourley, J. J. (2020). The Ensemble Framework For Flash Flood Forecasting (EF5) v1.2: description and case study. Geoscientific Model Development, 13, 4943-4958.

Flamig, Z. (2020). HyDROSLab/EF5-US-Parameters: EF5 parameters for USA (v1.0.0). Zenodo.

https://doi.org/10.5281/zenodo.4009759.

Fowler H. J., Wasko, C, & Prein, A. F. (2021). Intensification of short-duration rainfall extremes and implications for flood risk: current state of the art and future directions. Philosophical Transactions of the Royal Society, A379: 20190541.

Fowler, H. J. Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H., Guerreiro, S., Haerter, J. O., Kendon, E. J., Lewis, E., Schaer, C., Sharma, A., Villarini, G., Wasko, C., & Zhang, X. (2021). Anthropogenic intensification of short-duration rainfall extremes. Nature Reviews Earth and Environment, 2, 107-122.

Gourley, J. J., Hong, Y., Flamig, Z. L., Arthur, A., Clark, R., Calianno, M., Ruin, I., Ortel, T., Wieczorek, M. E., Kirstetter, P. E., Clark, E., & Krajewski, W. F. (2013). A unified Flash flood database across the United States, Bulletin of American Meteorological Society, 94, 799–805, https://doi.org/10.1175/BAMS-D-12-00198.1.

Gourley, J. J., Flamig, Z. L., Vergara, H., Kirstetter, P., Clark, R. A., III, Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J. M., Hong, Y., & Howard, K. W. (2017). The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. Bulletin of American Meteorological Society, 98, 361–372. https://doi.org/10.1175/BAMS-D-15-00247.1.

Gutmann, E. D., Rasmussen, R. M., Liu, C., Ikeda, K., Bruyere, C. L., Done, J. M., Garre, L., Friis-Hansen, P., & Veldore, V. (2018). Changes in Hurricanes from a 13-Yr Convection-Permitting Pseudo–Global Warming Simulation, Journal of Climate, 31, 3643-3657 (2018).

Hirabayashi, Y., Mahendran, R., Koirala, S. Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S. (2013) Global flood risk under climate change. Nature Climate Change, 3, 816–821 (2013). https://doi.org/10.1038/nclimate1911

Ivancic, T. J., & Shaw, S. B. (2015). Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. Climatic Change, 133, 681–693.

Khajehei, S. Ahmadalipour, A., Shao, W., Moradkhani, H. (2020). A Place-based Assessment of Flash Flood Hazard and Vulnerability in the Contiguous United States. Scientific Reports, 10, 448.

Kunkel, K. E., Karl, T. R., Brooks, H., Kossin, J., Lawrimore, J. H., Arndt, D., Bosart, L., Changnon, D., Cutter, S. L., Doesken, N., Emanuel, K., Groisman, P. Y., Katz, R. W., Knutson, T., O'Brien, J., Paciorek, C. J., Peterson, T. C., Redmond, K., Robinson, D., Trapp, J., Vose, R., Weaver, S., Wehner, M., Wolter, K., & Wuebbles, D. (2013). Monitoring and Understanding Trends in Extreme Storms: State of Knowledge, Bulletin of the American Meteorological Society, 94(4), 499-514.

Li, Z., Chen, M., Gao, S., Gourley, J. J., Yang, T., Shen, X., Kolar, R., & Hong, Y. (2021). A multi-source 120-year US flood database with a unified common format and public access, Earth System Science Data, 13, 3755–3766, https://doi.org/10.5194/essd-13-3755-2021.

Li, Z., Chen, M., Gao, S., Luo, X., Gourley, J. J., Kirstetter, P., Yang, T., Kolar, R., McGovern, A., Wen, Y., Hong, Y. (2021). CREST-iMAP v1. 0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction. Environmental Modelling & Software, 141, 105051.

Li, Z. (2022). Flashier Floods over the Conterminous US under a Changing Climate. figshare. Dataset. https://doi.org/10.6084/m9.figshare.19127186.v1.

Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., Chen, F., Chen, L., Clark, M., Dai, A., Dudhia, J., Eidhammer, T., Gochis, D., Gutmann, E., Kurkute, S., Li, Y., Thompson, G., & Yates, D. (2017). Continental-scale convection-permitting modeling of the current and future climate of North America. Climate Dynamics, 49, 71–95.

Lopez-Cantu, T. & Samaras, C. (2018). Temporal and spatial evaluation of stormwater engineering standards reveals risks and priorities across the United States. Environmental Research Letters, 13, 074006.

Maddox, R. A., Chappell, C. F., & Hoxit, L. R. (1979). Synoptic and Meso-scale Aspects of Flash Flood Events Bulletin of American Meteorological Society, 60, 115-123 (1979).

McClymont, K., Morrison, D., Beevers, L., & Carmen, E. (2020). Flood resilience: a systematic review, Journal of Environmental Planning and Management, 63, 1151-1176.

Mei, Y., & Anagnostou, E. N. (2015). A hydrograph separation method based on information from rainfall and runoff records. Journal of Hydrology, 523, 636–649.

Merz, B., Blöschl, G., Vorogushyn, S. Dottori, F., Aerts, J. C. J.H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U. & Macdonald, E. (2021). Causes, impacts and patterns of disastrous river floods. Nature Review Earth & Environment, 592–609. https://doi.org/10.1038/s43017-021-00195-3

Musselman, K. N., Flavio, L., Ikeda, K., Clark, M., Prein, A. P., Liu, C., Barlarge, M., Rasmussen, R. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. Nature Climate Change, 8, 808–812.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., Tolle, M., Gutjahr, O., Feser, F., Brisson, E., Kollet, S., Schmidli, J., van Lipzig, N. P. M., & Leung, R. (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. Review of Geophysics, 53, 323-361 (2015).

Prein, A., Rasmussen, R. M., Ikeda, K., Liu, C., Clark, M. P., Holland, G. J. (2017). The future intensification of hourly precipitation extremes. Nature. Climate Change, 7, 48–52.

Prein, A.F. Liu, C., Ikeda, K., Trier, S. B., Rasmussen, R. M., Holland, G. J. & Clark, M. (2017). Increased rainfall volume from future convective storms in the US. Nature Climate Change, 7, 880–884. https://doi.org/10.1038/s41558-017-0007-7

Rasmussen, R., Ikeda, K., Liu, C., Gochis, D., Clark, M. P., Dai, A., Gutmann, E., Dudhia, J., Chen, F., Barlage, M., Yates, D., & Zhang, G. (2014). Climate change impacts on the water balance of the Colorado headwaters: High-resolution regional climate model simulations. Journal of Hydrometeorology, 15, 1091-1116.

Rasmussen, R., & Liu., C. (2017). High Resolution WRF Simulations of the Current and Future Climate of North America. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6V40SXP. Accessed 09/01/2020 (2017).

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L. A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L.,

Doelman, J. C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., & Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environmental Change, 42, 153-168.

Saharia, M., Kirstetter, P., Vergara, H., Gourley, J. J., Hong, Y., & Giroud, M. (2017). Mapping Flash Flood Severity in the United States, Journal of Hydrometeorology, 18(2), 397-411.

Schär, C., Frei, C., Lüthi, D., & Davies, H. C. (1996). Surrogate climate-change scenarios for regional climate models, Geophysical Research Letters, 23, 669-672.

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If precipitation extremes are increasing, why aren't floods? Water Resources Research, 54(11), 8545-8551.

Shen, X., Mei, Y., & Anagnostou, E. N. A. (2017). comprehensive database of flood events in the contiguous United States from 2002 to 2013, Bulletin of American Meteorological Society, 98, 1493-1502.

Swain, D. L., Wing, O. E. J., Bates, P., Done, J. M., Johnson, K. A., & Cameron, D. R. (2020). Increased flood exposure due to climate change and population growth in the United States. Earth's Future, 8, e2020EF001778.

Tabari, H. (2020). Climate change impact on flood and extreme precipitation increases with water availability. Scientific Reports, 10, 13768.

Villarini, G. (2016). On the seasonality of flooding across the continental United States, Advances in Water Resources, 87**,** 80-91.

Wasko, C., Sharma, A., & Lettenmaier, D. P. (2019). Increases in temperature do not translate to increased flooding. Nature communications, 10(1), 1-3 (2019).

Wasko, C., Nathan, R., Stein, L., & O'Shea, D. (2021). Evidence of shorter more extreme rainfalls and increased flood variability under climate change. Journal of Hydrology, 603, 126994.

Wehner, M. F., Arnold, J. R., Knutson, T., Kunkel, K. E., & LeGrande, A. (2017). N. *Climate Science Special Report: Fourth National Climate Assessment Volume I Ch. 8* (U.S. Global Change Re-search Program, Washington, 2017).

Westra, S. Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., & Roberts, N. M. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall, Review of Geophysics, 52, 522– 555 (2014).

Wright, D. B., Bosma, C. D., & Lopez-Cantu, T. (2019). U.S. Hydrologic Design Standards Insufficient Due to Large Increases in Frequency of Rainfall Extremes. Geophysical Research Letters*,* 46, 8144–8153.

Wyżga, B. (1997). Methods for studying the response of flood flows to channel change, Journal of Hydrology, 198, 271-288.

Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., Doriji, C., & Chen, S. (2013). Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? Journal of Hydrology, 499, 91-99. https://doi.org/10.1016/j.hydrol.2013.06.042.

Younis, J., Anquetin, S., & Thielen, J. (2008). The benefit of high-resolution operational weather forecasts for flash flood warning, Hydrology and Earth System Science, 12, 1039–1051 (2008).

Zhang, B., Wang, S., & Wang, Y. (2020). Probabilistic projections of multidimensional flood risks at a convection-permitting scale. Water Resources. Research, 56, e2020WR028582 (2020).

Zhang, W., Villarini, G., Vecchi, G. A. & Smith, J. (2018). Urbanization exacerbated the rainfall and flooding caused by hurricane Harvey in Houston. Nature, 563, 384–388 (2018). https://doi.org/10.1038/s41586-018-0676-z

Chapter 8

Spatiotemporal characteristics of US floods: Current status and forecast under a future warmer climate

Publication

Li, Z., Gao, S., Chen, M., Gourley, J. J., & Hong, Y. (2022). Spatiotemporal characteristics of US floods: Current status and forecast under a future warmer climate. Earth's Future, 10, e2022EF002700. https://doi.org/10.1029/2022EF002700

8 Spatiotemporal US flood characteristics under a current and warmer climate

Abstract

Floods in the US exhibit strong spatiotemporal variability, mainly controlled by precipitation types and catchment attributes. In a future warmer climate, it remains largely unclear how such features change, relative to the current climate. Global climate models at coarse resolution cannot resolve convective-scale storms which is one of the major weather systems causing devastating floods in the US. Alternatively, coupled, high-resolution and continental-scale climate and flood simulations can advance our understanding. In this study, we couple a 4-km convection-permitting climate simulation model with a 1-km hydrologic model to simulate the spatiotemporal variability of rainfall and flooding over the contiguous US. In particular, changes in rainfall and flood frequency, spatial scale, and seasonality are explored in major climate divisions. We found: (1) an overall increase in flood frequency (+101.7%) and spatial extent (+44.9%), mainly attributed to more extreme rainfall and variability in the future; (2) weakening rainfall and flood seasonality, resulting in more random and unpredictable events throughout the year; (3) earlier flooding season onsets in the West and snow-dominated regions while delayed onsets in the East driven by drier antecedent soil moisture; (4) correlation between extreme rainfall and flood onsets is becoming stronger in the West yet weaker in the East in the future. Findings in this study can potentially serve as a basis for future flood exposure and risk assessments, as well as more scientific understanding of changing flood-generating mechanisms across the CONUS.

8.1 Introduction

Population and assets exposed to floods have been ever-increasing over the globe (Hirabayashi et al., 2013; Jongman et al., 2012; Tellman et al., 2021). Changes in land use have shifted so the United States is the leading country with the highest exposed asset values (Jongman et al., 2012; Rajib et al., 2021). Continuing urbanization, aging infrastructures, and intensified storms can exacerbate the current situation.

Future flood changes are often quantified by frequency (Hirabayashi et al., 2013; Li et al., 2022; Swan et al., 2020;), magnitude (Bates et al., 2020; Brunner et al., 2021), and spatiotemporal characteristics (Alfieri et al., 2020; Tellman et al., 2021) with greater attention on the former two. Flood spatiotemporal features, spatial scale and seasonality in particular, are especially central to anticipating flood exposure and risk management (Blöschl et al., 2017; Jongman et al., 2012; Tellman et al., 2021). Moreover, they are recognized as attributes to generalize flood-generating mechanisms (Brunner et al., 2020; Li et al., 2021a; Merz et al., 2021; Smith et al., 2011; Stein et al., 2021; Villarini, 2016). A body of studies has successfully described US flood spatial dependence and seasonality based on historical observations (Brunner et al., 2020; Li et al., 2021a; Stein et al., 2021; Villarini, 2016). Villarini (2016) completed a comprehensive study on US streamflow seasonality, in which he did not observe strong temporal shifts in flood seasonality, but seasonality strength is weakening due to human interference. Berghuijs et al. (2019) proposed the term “flood synchrony scale” to characterize the flood extent of which at least half of the river gauges co-experience floods. Brunner et al. (2020) applied this concept in US river gauges and identified 15 regions with similar flood behaviors. Nonetheless, the aforementioned studies have provided little understanding outside gauge locations that are sparsely sampled over the US To fill the gap, continental-scale hydrologic simulations at each river reach are essential. Moreover, continuous simulation in lieu of event-based simulation is critical as the antecedent catchment condition can thus be well represented (Wasko et al., 2019).

However, future changes in spatiotemporal flood characteristics are more equivocal, although some generic features such as earlier snowmelt season are generally expected. Primary challenges hindering additional insights are: (1) the lack of high-resolution climate simulations underpinning fine-scale flood simulations, especially for flood extent, (2) uncertain climate simulations that often require large ensemble predictions to replace the traditional one-model-one-vote approach (Clark et al., 2016; Giuntoli et al., 2021), and (3) complex flood-generating

mechanisms (Ashish et al., 2018; Wasko et al., 2019). For (1) and (2), these challenges persist to date, but workarounds have been proposed. For instance, downscaling Global Climate Models (GCM) has been explored over the last two decades to enable regional climate projections. Generally, two downscaling approaches have become mainstream: statistical downscaling and dynamic downscaling (the predictor-predict approach). Non-stationarity in atmospheric components has been the obstacle for statistical downscaling and is yet to be resolved by the community. Likewise, future flood characteristics are unlikely to remain stationary since the atmosphere is the main driver for floods. In contrast, dynamic downscaling uses the outputs from the GCMs and then drives a regional climate model, yielding more realistic simulations. Lately, dynamic downscaling has been prevailing and holding great promise owing to the increased computational power and physics parameterizations (Clark et al., 2016; Liu et al., 2017; Prein et al., 2017). The recent CONUS-1 dataset (Liu et al., 2017), produced at hourly and 4-km resolution, has become one of the highest resolution continental climate simulations, which has empowered a wide range of regional climate studies, including extreme precipitation (Prein et al., 2017), snowmelt (Musselman et al., 2018), flash flood-producing storms (Dougherty & Rasmussen, 2020), flood frequency analysis (Yu et al., 2020), etc. By applying the Pseudo Global Warming (PGW) approach (Schär et al., 1996), this dataset circumvents the more uncertain dynamic changes yet focuses on more deterministic thermodynamic changes. As such, it is suited for continental flood simulations requiring high resolution and broad spatial coverage.

The focus on underlying physical processes recognizes that floods are not only dependent on extreme rainfall, for which there is high confidence in increases over North America in the future (Douville et al., 2021), but with modulations by the land surface processes (Merz et al., 2020). Undoubtedly, increasing rainfall extremes will directly amplify flood magnitude in a controlled environment (meaning other variables remain unchanged). However, there is widespread evidence that antecedent catchment conditions (soil moisture and groundwater level) are becoming drier as temperature rises in conjunction with more partitioning of snowfall to rain and earlier snowmelt (Ashish et al., 2018; Ivancic & Shaw, 2015; Wasko et al., 2021). These conditions counteract increases in extreme rainfall magnitudes to modulate the flood-generating process. While there are so many conceptual discussions in the literature, there are still ongoing works perceiving extreme precipitation as a proxy for high streamflow (Ivancic & Shaw, 2015).

Moreover, many details are lacking about which processes dominate and for which times of the year and climate divisions.

Motivated by the lack of continental-scale studies of reach-level flood characteristics beyond gauge locations and elusive future changes of such, in this study we seek to address the following objectives: (1) assess the relative influences of atmospheric and land surface drivers to flood spatiotemporal characteristics by comparing present-day conditions to future ones, (2) quantify the spatial changes in rainfall and flood extents, and (3) estimate temporal shifts and changes of the seasonality of rainfall and flooding in the CONUS. We achieve these goals via simulating US reach-level streamflow using forcing data from the outputs of a continental convection-permitting climate model. Methodologies are then developed to quantify rainfall/flood spatial scales after spatiotemporal clustering. To our best knowledge, we, for the first time, analyze the spatiotemporal changes in continental convective-scale rainfall and kilometer-scale floods based on a proposed framework that works for distributed models. It is anticipated that this study provides a practical assessment of US extreme flood characteristics, and more importantly, on their future changes so that local communities can be well prepared to mitigate their impacts.

8.2 Data and methods

8.2.1 US climate divisions

Statistical results throughout this study are aggregated to US climate regions to provide a more holistic view. The popular Koppen-Geiger climate classification, used widely in climatological studies, however bears inhomogeneous divisions in the US For instance, the Southwest, representing an arid climate, is divided into four sub-classes, while the vast southeastern US is represented by a single class, as shown in Fig. S1a. In contrast, the Bukovsky climate division in Fig. S8.1b equally divides the US into 17 classes based on temperature and precipitation (Bukovsky, 2011). This climate regionalization yields subdivisions with similar areas and is thus preferred in this study as well as in other climate studies (Prein et al., 2017). A list of data used in this study can be found in Supplementary Table 8.1.

8.2.2 Convective-permitting climate simulation

In this study, we use the outputs from a 13-year CONUS climate simulation at hourly time step and 4-km grid spacing to force a continuous hydrologic simulation. The retrospective

simulation (control run; CTL) from 2000 to 2013 downscales ERA-Interim reanalysis data, with spectral nudging applied (Liu et al., 2017). In parallel, a future climate counterpart is simulated by applying the Pseudo Global Warming approach (PGW) which perturbs climatic boundary conditions (e.g., wind, temperature, humidity, and sea surface temperature) from an ensemble mean of 19 CMIP5 models under the Representative Concentration Pathway (RCP) 8.5. Climatic changes are evaluated by subtracting each variable in the simulation period 2071 – 2100 from the reference period 1976 - 2005. This approach highlights relative changes and minimizes the absolute uncertainties. The essence of the PGW run is to assess more deterministic thermodynamic climate change impacts while not allowing the impacts of more uncertain dynamic changes. Because of spectral nudging, we can then conduct event-to-event analysis of rainfall and floods based on the current and future climate.

8.2.3 Hydrologic simulation

The Coupled Routing Excess STorage (CREST) V2.1 is used in this study to simulate hydrologic responses to climate change. The CREST model is a distributed hydrologic model that couples Snow-17 for snowmelt, Variable Infiltration Curve (VIC) for soil infiltration, and kinematic wave solution for overland and channel routing. A conceptual linear reservoir routing is used for subsurface flows. Prior studies have successfully applied the CREST model for operational flash flood forecasts in the US (Gourley et al., 2017), detecting global floods and landslides (Wu et al., 2012; Zhang et al., 2016), mapping flood inundation extents and depths (Chen et al., 2021; Li et al., 2021b), and simulating the water cycle (Li et al., 2018). Building upon a previous legacy that operates the CREST model for real-time flash flood forecasting (Flamig et al., 2020; Gourley et al., 2017; Vergara et al., 2017), we use the a-priori distributed parameters over the CONUS. Since potential evapotranspiration (PET) in the future will increase given global warming, we calculated future daily PET based on the Thornthwaite equation that depends only on daily temperature from climate simulations (Thornthwaite, 1948). Other variables such as hourly precipitation and temperature are direct inputs to the CREST model. We simulated hourly streamflow at 1-km grid spacing from 2001 to 2013, with the first year used to warm up the model states.

8.2.4 Rainfall/Flood spatial scales

Rainfall spatial scale is measured in this study by firstly considering the contour lines of daily rainfall enclosed by a certain extreme threshold, referred to as the areal extent (Fig. 8.1). In practice, we plot contour lines for daily rainfall greater than 2-year average recurrence interval thresholds for the retrospective simulation. The use of 2-year rainfall as thresholds can be found in the literature such as Lamjiri et al. (2017). A gamma distribution is fitted to the annual maximum rainfall values, and a certain value exceeding the 2-year return interval is determined from the distribution. The goodness-of-fit test (Anderson-Darling test) regarding extreme value distribution is shown in Fig. S8.2, where 92% of the samples pass the significance test at a significant level of 0.01. then, we sample vertices along the contour line. The area can be approximated numerically using Green's theorem shown in eq.1:

$$A = \iint dA = \frac{1}{2}\oint_C xdy - ydx, \tag{8.1}$$

where C is a closed curve that bounds regions with extreme rainfall, and x, y refer to the coordinate of the vertices sampled along the contour line. Lastly, the rainfall spatial scale is computed by measuring the distance between the centroid of the enclosed area and the farthest point on curve C. For the example given in Fig. 1, the 2-year rainfall threshold in such a region is 28.6 mm/day, and the area of an enclosed curve is calculated as 3,211 km^2 using Eq. 8.1. The spatial scale of this event is 260.2 km.

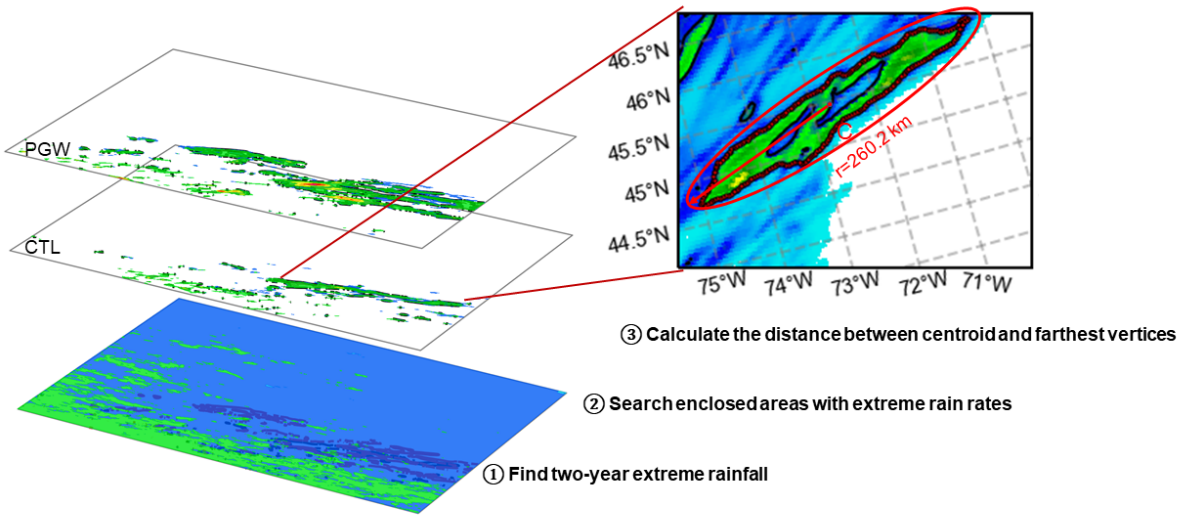


Figure 8.1 An illustration of rainfall spatial scale calculation, where symbol C represents the enclosed curve. CTL: historical simulation; PGW: future simulation.

River flooding in nature, unlike continuous rainfall fields, occurs along elevation-dependent river networks. Therefore, a different strategy needs to be developed to account for river network topology. Flood spatial scale, by definition, is the maximum spatial extent in which multiple gauges or reaches co-experience flooding synchronously. Berghuijs et al. (2019) proposed the measure of flood synchrony scale by computing the maximum radius of gauges within which at least 50% of them are flooded. Similar approaches have been adopted in the literature (Brunner et al., 2020; Kemter et al., 2020). However, their works are limited to gauge sites that are sparsely distributed, and there lacks a detailed representation of river topology. As a consequence, the flood synchrony scale may be overestimated by disconnected river networks. In contrast, distributed hydrologic simulations can overcome this limitation by incorporating pixel-wise sampling with a river network embedded in the model setting. A challenge notwithstanding for distributed models in flood synchrony studies is that there is no prior standard to calculate flood spatial scale.

In this study, we propose a method that fills this gap and can be applied in any distributed streamflow simulation. First, the flood threshold is computed based on an N-year event. Here "N" can be replaced with any number by considering the length of data and application. In this study, given the total 10-year simulation, we select a 2-year streamflow value as the threshold, which is approximated as a bankful condition for river channels (He & Wilkerson, 2011). Similar to 2-year rainfall, a log-Pearson type III distribution is fitted to the annual maxima streamflow values, which is a conventional way to calculate return intervals by the US Geological Survey (Veilleux et al., 2014). The goodness-of-fit test (Anderson-Darling test) of the distribution is shown in Fig. S8.3, where 91.6% of the samples pass the significance test at a significant level of 0.01. With the flood thresholds available at all grid cells, flood occurrences at each pixel can be estimated by simply counting the events that exceed this threshold. Second, the flood occurrences can be rearranged in a matrix ($m \times n$) form with binary values, termed the co-occurrence matrix. Here "m" represents the number of days or hours, depending on the data frequency, and "n" represents the number of pixels in output streamflow simulation. The co-occurrence matrix is used to select synchronous flood events and cluster pixels in proximity. Third, at each timestamp, we cluster flooded pixels with an unsupervised machine learning algorithm, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al. 1996). The DBSCAN is a density-based clustering non-parametric algorithm that groups high-density points in space and marks isolated points as outliers. In doing so, we can remove isolated flood pixels while focusing on densely

flooded regions. Lastly, we find the centroid geophysical location within each cluster and compute the maximum distance to enclosed edge nodes. A schematic flowchart of these processing steps is shown in Fig. 8.2. Similar approach can be found in Brunner et al. (2019).

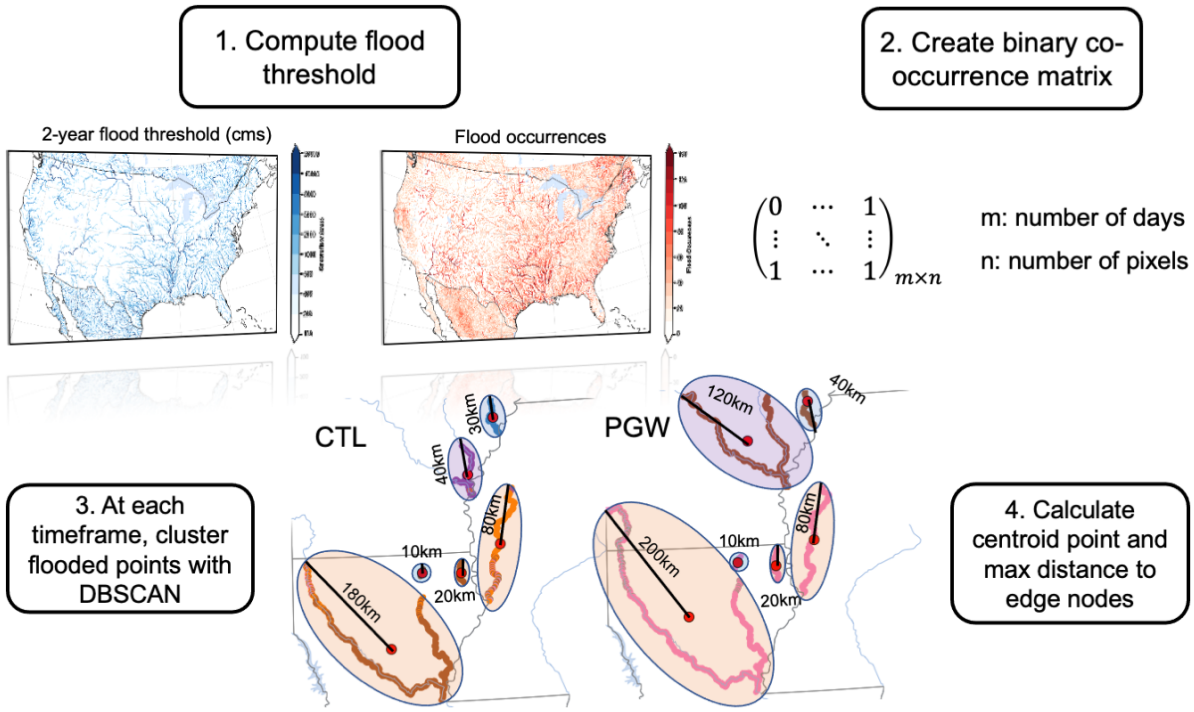


Figure 8.2 Schematic view of calculating flood spatial scales.

8.2.5 Flood seasonality measures

We aggregate the hourly streamflow data to daily to investigate flood seasonality over the CONUS. The circular statistics (i.e., circular mean, circular variance) are used for data such as flood calendar days on a polar coordinate system (Burn, 1997; Villarini, 2016). Circular mean and resultant length are the two primary measures to determine the average rainfall/flood days in Days of the Year (DOY) and strength of the seasonality, respectively. The strength of seasonality reflects how strong the seasonality is, with 1 indicating a strong seasonal cycle and 0 indicating no seasonal cycle. For brevity, we term the strength of seasonality as the seasonality index (SI) hereafter. Equations 8.2 and 8.3 express the calculation of circular mean and SI in a complex domain.

$$Mean = A\left(\sum_{j=1}^{n} exp\left(i \cdot \alpha_j\right)\right) \tag{8.2}$$

$$SI = \frac{1}{N} \left\| \sum_{j=1}^{n} exp\left(i \cdot \alpha_j\right) \right\| \quad (8.3)$$

where function A returns the angle of the complex; i is the imaginary unit; α_j is the jth day of the year with a range of (0,365). The $\|\cdot\|$ denotes the absolute value.

8.3 Results

8.3.1 Frequency and spatial changes

8.3.1.1 Extreme rainfall

Extreme rainfall appears to increase both in occurrences (+118.0%) and spatial scales (101.7%) across all US climate regions under future climate simulation (Fig. 8.3). Driven by different weather patterns, such changes vary by region. In the Pacific West Coast, the increases in both extreme rainfall occurrences and spatial scales are the least (+37.7%/58.4% for the change of occurrences/scale) among US climate zones. The main forcing agent there for extreme rainfall is atmospheric rivers (ARs) which are narrow and long plumes that originated from the tropics to replenish water resources and cause flooding along the West Coast. It is found that 78% - 100% of the storms in the Pacific Northwest (PNW) are associated with ARs, although such fraction reduces to 60% in the Pacific Southwest (PSW) (Lamjiri et al., 2017). In a warmer climate, the dominant thermodynamic change leads to a 35% increase in AR frequency (Hagos et al., 2016), which is comparable to our results. The Southwest, influenced by the North American Monsoon (NAM), is projected to experience 114.7% more events and 45.6% greater scales. The Great Plains have been mainly impacted by mesoscale convective systems (MCSs) and tropical cyclones, which are expected to increase precipitation extremes owing to strengthened convective updrafts and intensified cyclone circulation in a warmer climate (Prein et al., 2017; Nie et al., 2018). We see on average a 99.9% and 128.1% increase in frequency and spatial scale, indicating that the increase in spatial scale outpaces that in frequency. In the central US, where the MCSs contribute 30-70% of the total warm-season precipitation (Fritsch et al., 1986; Feng et al., 2016), extreme rainfall occurrences (167.4%) are the mostly increased among climate divisions. The East Coast, influenced by tropical and extratropical cyclones, has a 123.3% and 128.9% increase in frequency and spatial scale. The East Coast has the largest change in rainfall spatial scale, partly attributed to increased storm motion (Prein et al., 2017). In general, we observe an increasing rainfall spatial scale from the West Coast to the East Coast (Fig. 8.3b), implying a transition from convective-

scale to synoptic-scale storms in the future. Six individual storms in Fig. 8.4 visually manifest the increase in rainfall spatial scale in PGW. Two factors result in the increase in spatial scale: (1) single storm grows in length in PGW run (Figs. 8.4b and 8.4e); (2) individual storms in CTL are connected to become large storms in PGW (Figs. 8.4a, 8.4c, 8.4d, and 8.4f).

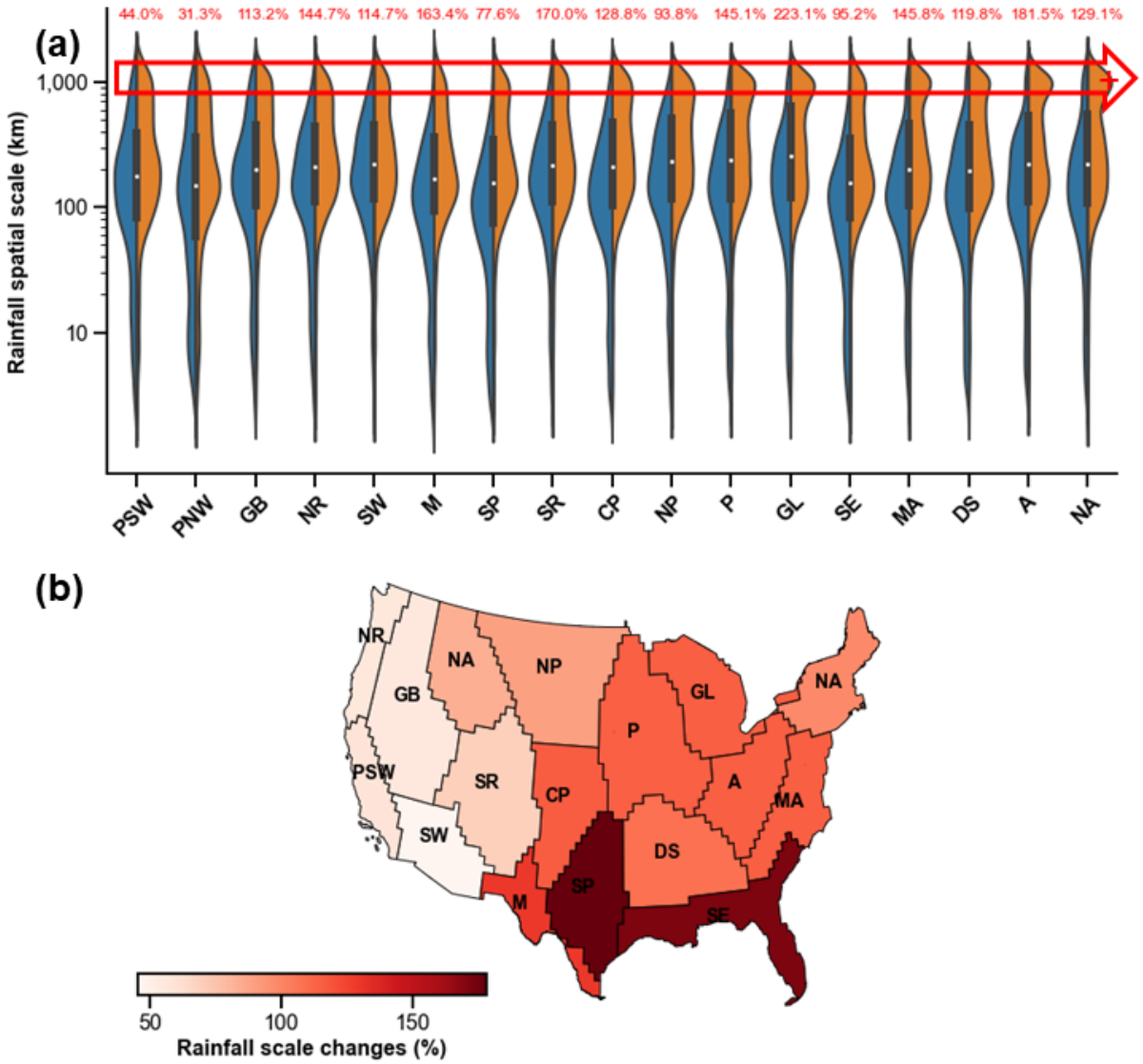


Figure 8.3 (a) Violin plot of current and future rainfall spatial scales grouped by the Bukovsky regions. Numbers in red text indicate relative changes (%) in extreme rainfall occurrences; (b) Map of rainfall spatial scale changes, averaged over the Bukovsky regions.

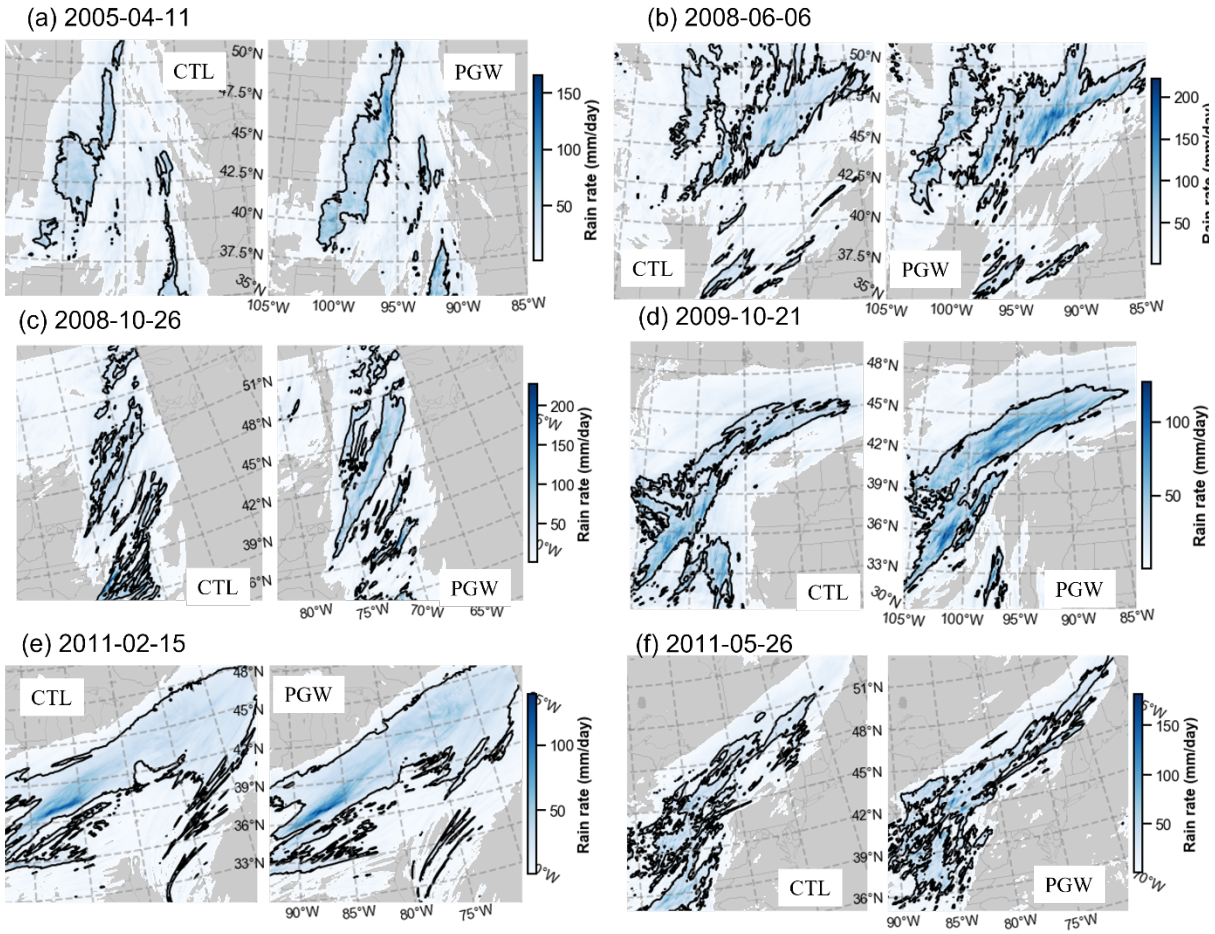


Figure 8.4. Maps of six storm events show the spatial scale difference between CTL and PGW. The black contour line encloses the area whose rainfall rates exceed the two-year rainfall threshold.

8.3.1.2 Floods

Roughly one-third of the extreme rains in the US translates to floods by comparing their occurrences in Figs. 8.3 and 8.5, which is close to the ratio of 36% estimated by Ivancic & Shaw (2015) and Ashish et al. (2018). The increase in extreme rainfall and snowmelt overcome the drier catchment states under anthropogenic warming, resulting in an increase in flood frequency and spatial scales across the US climate zones in Fig. 8.5.

Similar to the frequency changes of extreme rainfall, flood frequencies (scales) in Fig. 8.5 are increased by 55.2% (45.9%) in the Great Plains. However, the central US is likely to experience 74.2% more frequent floods and 33.6% greater flood extent, much less than the changes in the western US The East Coast also sees a reduced increase in flood spatial scale (35.8%) compared to the West Coast (70.6%). Relatively speaking, floods in the western US are becoming more

severe in the future than those in the central and eastern US, opposite to rainfall changes. Such a pattern corroborates the multidisciplinary nature of flooding rather than being a pure atmospheric process. From the perspective of flood-generating mechanisms, the western US features steeper terrain and shallower soil profiles that more typically yield surface runoff by the infiltration excess process (i.e., rainfall rates exceed maximum soil infiltration capacity) and snowmelt (Brunner et al., 2020; Stein et al., 2021). The infiltration excess process is less dependent on catchment wetness (antecedent soil moisture or groundwater level) but rather on rainfall intensity. Increased surface water availability from snowmelt and rain on snow in a warmer climate apparently increase flood risk in the western US (Musselman et al., 2018). On the contrary, the relatively flatter terrain and deeper soil profiles in the central and eastern US yield runoff generation by the saturation excess process (i.e., surface runoff is produced after the soil saturates from the bottom up), thereby leaving floods sensitive to antecedent soil moisture and groundwater levels. These two states, however, are more likely to decrease in a warmer climate because of increased atmospheric temperature and potential evapotranspiration (Ashish et al., 2018; Wasko & Sharma, 2017).

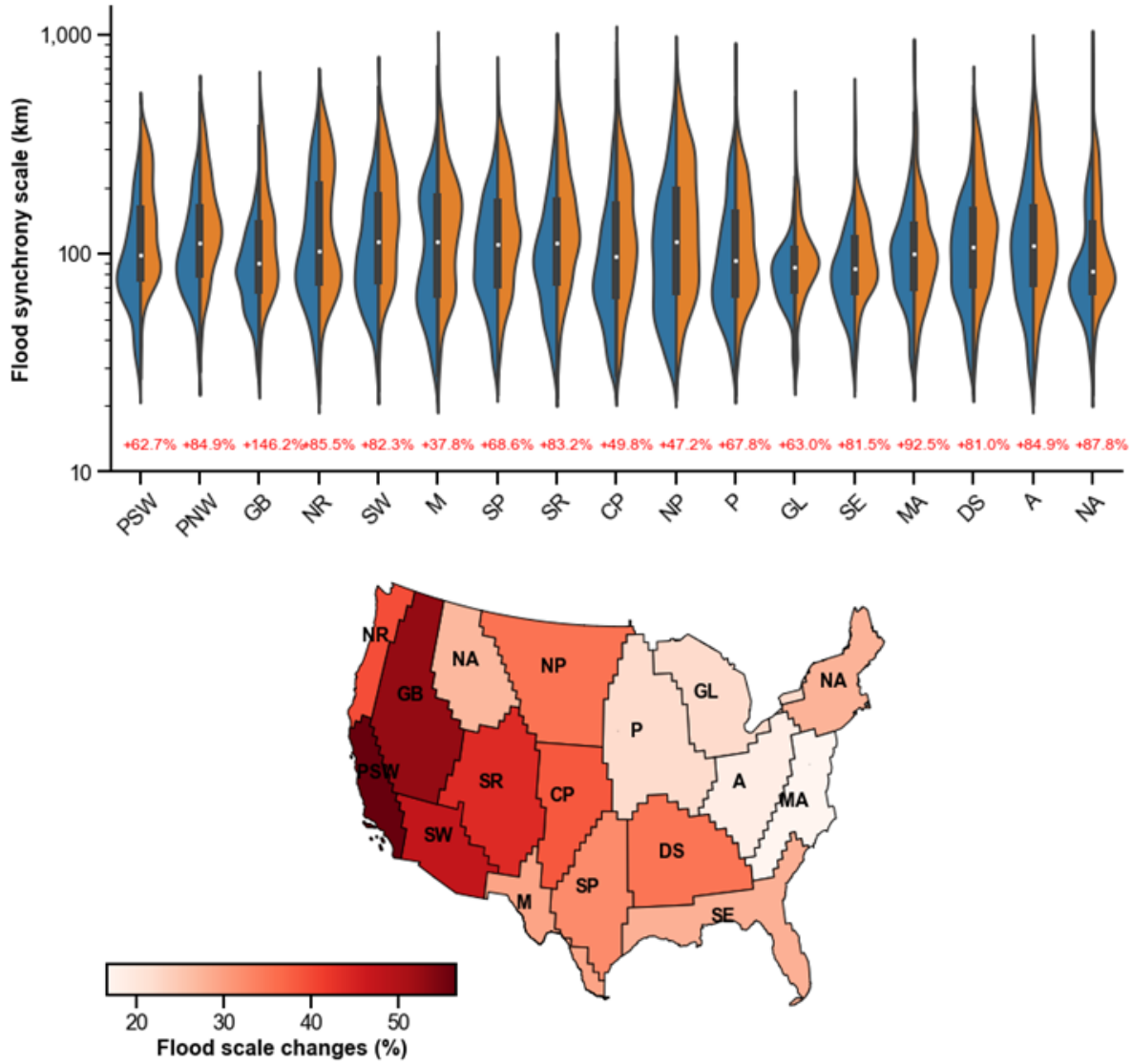


Figure 8.5 Similar to Figure 3, but for floods.

8.3.2 Seasonal changes

8.3.2.1 Western United States

Figures 8.6 and 8.7 depict the seasonality of extreme rainfall simulated by CTL and its changes by taking the *DOY* difference (PGW minus CTL). Figures 8.6a and 8.7a, showing the spatial distribution of the months at which extreme rainfall occurs, are on par with US extreme rainfall climatology. The Pacific West Coast, as discussed in Section 3.1, is dominantly influenced by ARs that occur in winter (mostly in January). Therefore, the flood season in such a region, as shown in Figs. 8.8a and 8.9a, follows closely with the extreme rainfall season. In fact, it is also the place

that has the strongest rainfall/flood seasonality, with the SI of both rainfall and floods above 0.8. Future changes of peak days and seasonality index are marginal for both rainfall (Figs. 8.7b and 8.7d) and floods (Figs. 8.9b and 8.9d), with a small IQR (Inter-Quartile Range). Moving towards the East, SI values for both rainfall (Fig. 8.7c) and floods (Fig. 8.9c) tend to decrease with increasing IQR, pointing to more complex inland weather systems. Future SI values in the east of the Pacific Coast indicate negative changes in the future, manifesting weakening seasonal cycles of rainfall (Fig. 8.7d) and flooding (Fig. 8.9d). The Great Basin is jointly impacted by ARs, troughs, and cutoff lows (Prein & Mearns, 2021), leading to a more diverse seasonality of extreme rainfall (Fig. 8.7a) and flooding (Fig. 8.9a), yet centered in wintertime. Over the Northern Rockies, extreme rainfall or snowfall spans the winter-spring-summer season. Due to the large fraction of snowmelt, flood seasons are centered in late spring and early summer. Future floods are becoming, on average, 22 days earlier because of earlier snowmelt caused by increasing temperatures, comparable to the 3-5 weeks results by Xu et al. (2021). The Southern Rockies are featured by spring-summer-fall extreme rainfall and floods occurring during the warm season from spring through fall. Notably, flood seasonality (0.77) over the Rockies is much stronger than rainfall seasonality (0.53), which is attributed to the contribution of snowmelt that typically occurs in late spring and early summer. However, flood seasonality in such a region in the future will decrease by 0.23, a result of less snowmelt contributing to flooding generation from a decreasing snowpack. As a consequence, one can infer that the reduced flood seasonality complicates predictability in downstream snow-fed streams. The Southwest region is equally influenced by summer-fall NAM and winter inland-penetrating Pacific storms, while over half of the flood events occur in fall. However, no significant temporal shifts are observed in the future simulation.

8.3.2.2 Central United States

Weather types in the central US are composed of spring-summer MCSs (Ashley et al., 2003; Prein et al., 2017), summer-fall tropical cyclones (Chalise et al., 2021; Li et al., 2021a), and spring extratropical cyclones (Barbero et al., 2019), among which extreme rainfall events in summer are the most prevalent (Fig. 8.7a). Flood events in late spring are common in the northern plains (Fig. 8.9a), mainly caused by the contribution from snowmelt and rain-on-snow events (Brunner et al., 2020; Villarini, 2016). Owing to earlier onsets of seasonal snowmelt (Fig. 8.7b; Li et al., 2021a), flood events in the Northern Plains, Great Lakes, and Prairie will trend, on average, approximately

one week earlier in the future (Fig. 8.9b). Similar to the Rockies, the seasonality of extreme rainfall and flood in the central US is weakening in the future (Figs. 8.7d and 8.9d), which makes flood prediction more challenging (Ledingham et al., 2019).

8.3.2.3 Eastern United States

Extreme rainfall in the eastern US is associated with cool-season extratropical cyclones and warm-season convective rainfall. The seasonality of rainfall on the East Coast reaches 0.6. Flood events, caused by convective rainfall and exacerbated by snowmelt in the Appalachians and New England, tend to span across seasons (Fig. 8.7a). Flood seasonality in the eastern US is 0.3 instead (Fig. 8.9b). As discussed in Section 3.1.2, flood generation depends more on catchment states in these regions, meaning rainfall exerts a less important role than that in the western US Due to drier antecedent states (e.g., soil moisture and groundwater level), flood events will be delayed for tens of days in a warmer climate (Fig. 8.9b), especially for the Mid-Atlantic (14 days) and Deep South (10 days).

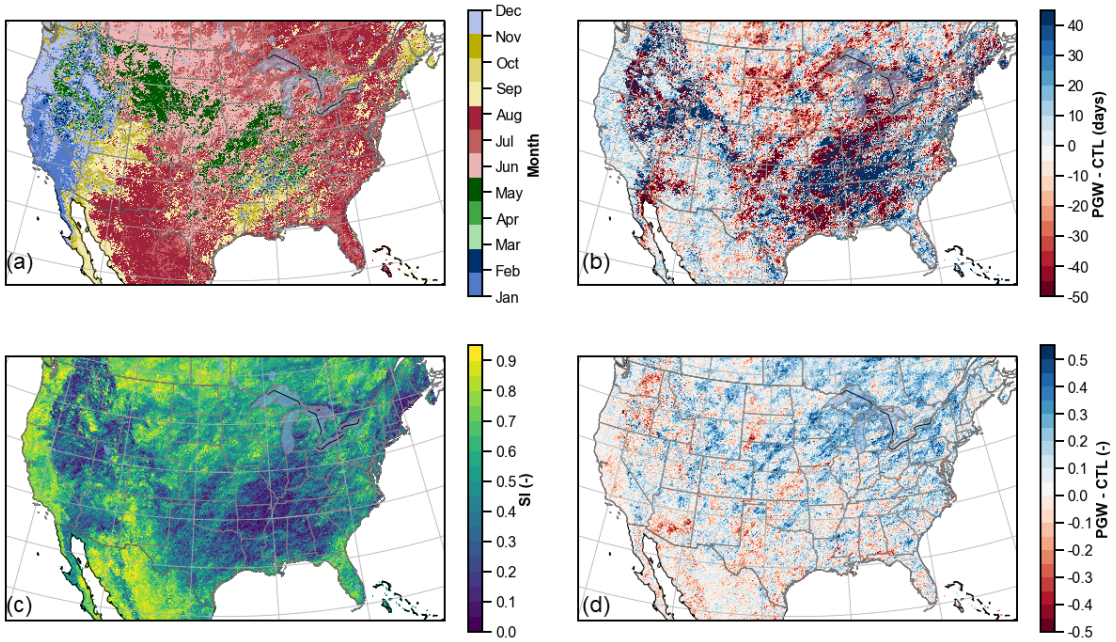


Figure 8.6 Maps of temporal changes of extreme rainfall: (a) retrospective rainfall seasonality (month); (b) difference of peak day between PGW and CTL; (c) rainfall seasonality index; (d) differences of seasonality index between PGW and CTL.

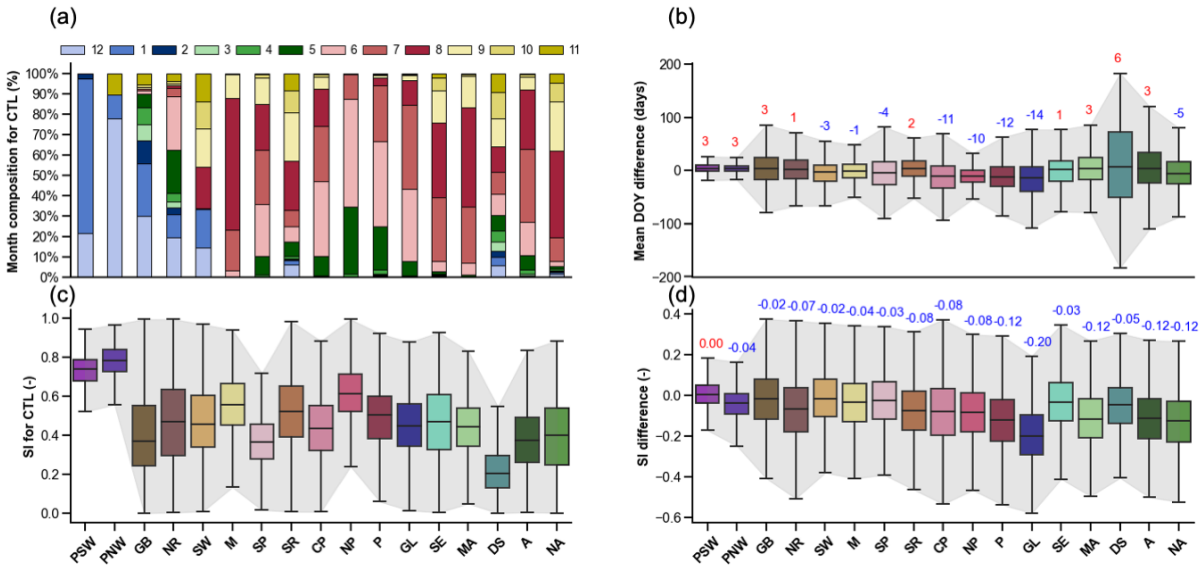


Figure 8.7 Retrospective rainfall seasonality and changes grouped by the Bukovsky regions. (a) month compositions in retrospective setting; (b) differences in Date Of Year (DOY) between PGW and CTL; (c) seasonality index (SI); (d) differences of SI between PGW and CTL. The Inter-Quartile Range (IQR) (defined as the upper and lower bound of the boxplot) is the shaded gray color. The numbers above the boxplot are the median values, which indicate increase (red) or decrease (blue) in seasonality.

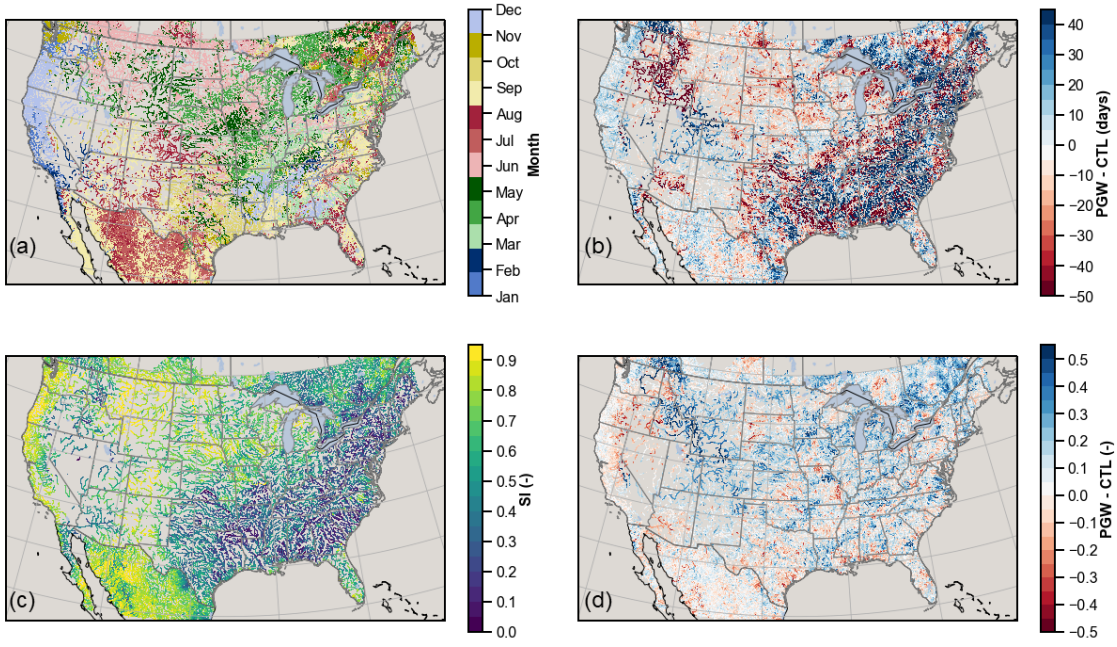


Figure 8.8 Similar to Figure 6, but for floods.

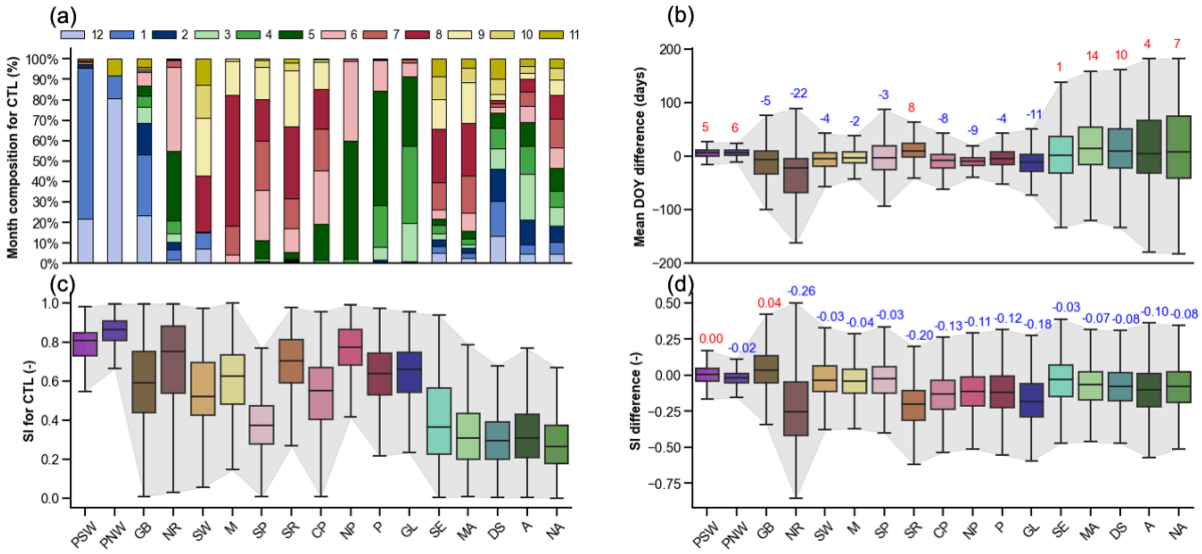


Figure 8.9 Similar to Figure 7, but for floods.

8.3.3 Temporal correlation between rainfall and floods

Despite decreasing rainfall and flood seasonality in the future, the circular correlation of event date between extreme rainfall and floods increases (Fig. 8.10). This can be explained by less accumulating snowpack in the US which delays rainfall-runoff transformation. Similarly, the correlation of seasonality between rainfall and floods increases from 0.63 (CTL) to 0.77 (PGW). The seasonal cycle of rainfall and flooding diminishes in the future, while rainfall and flooding events become more correlated. Both correlations can act as predictive tools to infer future flood characteristics.

Breaking down the correlations into months in Fig. 8.10e, we can diagnose the monthly correlation variations. We observe the correlation of rainfall-flood date in PGW is apparently higher in the future for the cool seasons (January, February, March, and April). Correlation is nearly zero for CTL simulation in these months because of the delayed rainfall-runoff process. However, earlier snowmelt in the future shortens that delay, thereby leading to a higher correlation (~0.1). On the other hand, the present-day correlations in October, November, and December are slightly higher than those in PGW. The correlation of seasonality in the future is consistently higher than present across months, while the correlations in February and March for both CTL and PGW are small (0.1).

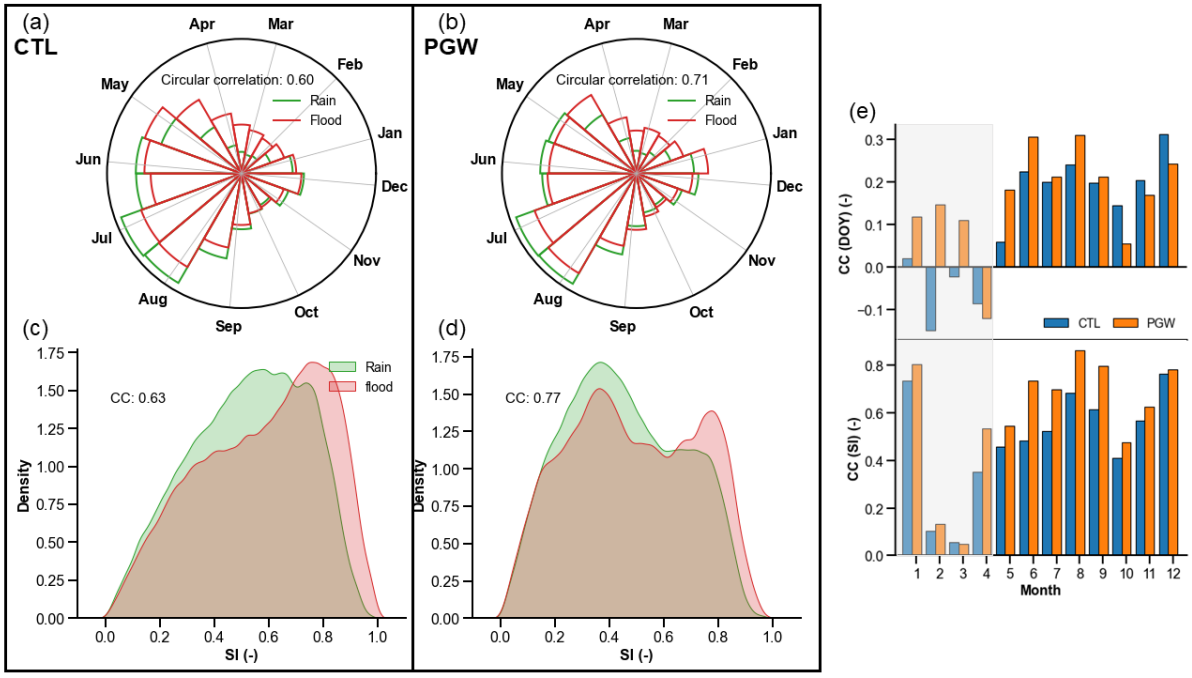


Figure 8.10 Temporal correction between rainfall and floods: (a) circular correlation in rainfall-flooding date in historical simulation; (b) circular correlation in rainfall-flooding date in future simulation; (c) histogram of seasonality index (SI) in historical simulation; (d) histogram of seasonality index in future simulation; (e) correlations of Date of Year (DOY) and seasonality index by month.

The spatial map showing a rainfall-flood correlation in Fig. 8.11 identifies regions with strong/weak correlations. The rainfall-flood date correlation is generally stronger in the western US (0.31) than in the eastern US (0.11). The Southwest and Mezquital, in particular, have CC values above 0.6 because of intense storms and low-infiltrating soils, resulting in flashier floods. This correlation even becomes 14.9% stronger in the future, possibly due to intensified storms fueled by enhanced temperature (energy) and atmospheric moisture (water availability). The Great Basin and Rockies will also have a 34.4% higher correlation in the future, which is related to earlier snowmelt that reduces rainfall-runoff lag time. The central and eastern US have slight increases in the correlation of DOY. The correlation of seasonality is increased across the CONUS, except for the Great Lakes, Prairie, and North-Atlantic.

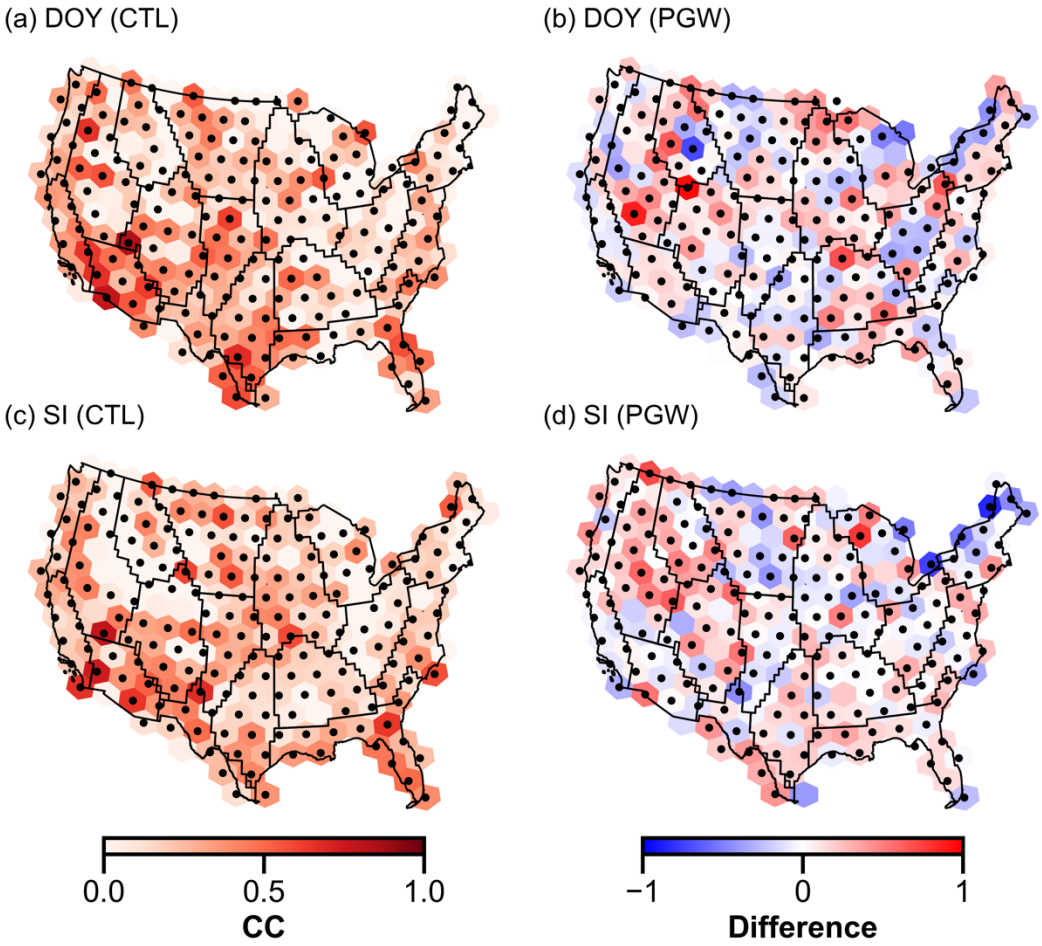


Figure 8.11 Map of correlation over the CONUS: (a) circular correlation between rainfall and flood dates in CTL simulation; (b) differences of circular correlation between rainfall and flood dates (PGW - CTL); (c) Spearman correlation between rainfall and flood seasonality in CTL simulation; (d) differences of spearman correlation between rainfall and flood seasonality (PGW-CTL). The stipples indicate a significant correlation (p-value less than 0.05) between rainfall and flood.

8.4 Discussion

8.4.1 Representativeness of flood synchrony scale by hydrologic models

This study introduces a new methodology that uses a high-resolution distributed hydrologic model to overcome limited coverages by stream gauges that are conventionally used to derive flood synchrony scale (Berghuijs et al., 2019; Brunner et al., 2020). The added values of model simulation against stream gauges are two-fold. First, in terms of spatial extent, the reach-scale representation of flood enabled by modeling is more informative than stream gauges at point-scale. In a simple illustration (Fig. S8.4), let us assume gauge-based approach indicates flood synchrony

at A, B, C, and D with near-perfect confidence, despite the inherent uncertainty sources from subjective parameterization and gauge measurement itself. Such confidence probably drops quickly as the target of estimation moves away from these gauge points. Put simply, one may argue the reaches in between these gauges have synchronous floods, but it would be hard for him/her to infer the extent outside these gauges. In comparison, the level of confidence at gauge locations and their contributing area is comparable. Moreover, we can now not only infer synchronous floods occurred between A-B, B-C, and C-D as a way of interpolation but also at extended reaches i.e., A-E, D-F, and B-G. In essence, the modeling approach can be seen as a way to physically interpolate/extrapolate information from gauges.

The second fold is whether model simulated flood synchrony is realistic, which necessitates careful validation. We gathered flow frequency data at more than 6,000 USGS gauges and compared simulated two-year flow with observed values (Fig. S8.5). The underestimation of two-year flow is obvious, with the bias being -1.1. Further, the validity of spatially synchronous flood is verified using a metric connectedness error (Fig. S8.6), which measures the differences between number of gauges by simulation and by observation co-experiencing floods. Overall, the median connectedness error is centered around zero, with slight underestimation (-16). Particularly for spring floods, the degree of underestimation is magnified (-121). The overall underestimation of both two-year flow and synchronous flood events point to an overall dry bias, as reported by Liu et al. (2017) and Li et al. (2022).

However, such validation is difficult to directly assess the flood synchrony scale by observations and simulations. Spatially coherent observation is more insightful. A collection of satellite observations such as optical sensors (e.g., MODIS, Landsat) and synthetical aperture radars (e.g., Sentinel-1) is promising to construct a flood synchrony dataset for model validation. In this study, we instead emphasize the relative changes due to climate change while did not conduct full-blown evaluation.

8.4.2 Sensitivity of the results to climate divisions and extreme thresholds

Results interpreted in this study are subject to choices of climate divisions and thresholds to determine extreme rainfall or floods. The Bukovsky climate zones, relative to the Koppen-Geiger climate zones, are assumed to be more homogenous. We expect the zonal mean statistics calculated in e.g., Figs. 8.3 and 8.5 can vary as if using different climate divisions. However, the general

message should be similar – western US will likely experience greater changes in floods yet less changes in extreme rainfall, as counter to the eastern US. Second, the extreme thresholds for rainfall and floods have great impacts on occurrences of those events. A detailed depiction of extreme rainfall and flooding events such as low-end and high-end extreme events is insightful to cover the full spectrum of frequency. This study establishes a methodology and paves a way for future research on such a scope.

8.4.3 Results intercomparison

Despite many existing climate studies that focus on flooding, few of them have fully considered convective-scale storms based on the GCM, which forms the foundational differences when making comparisons. For instance, previous studies indicate that some dry regions such as the Southwest will experience less frequent floods (Hirabayashi et al., 2013), However, we see an increase in flood frequency and scales (Fig. 8.5). This contrasting result is likely a reflection of different climate data at different scales. First, our high-resolution climate simulation can overcome the poor representation of extreme precipitation (especially convective-scale storms) in GCMs that are often used in large-scale flood simulation (Farnham et al., 2018; Kendon et al., 2012). Second, this dataset is primarily used for assessing thermodynamic changes in the atmosphere, which exerts great importance in invigorating storms in the future, while the dynamic changes included in other flood studies are not allowed. Some similarities coexist with differences. For instance, intensification of extreme rainfall and flood events with respect to their spatial scales is found in Yu et al. (2020). Earlier flood onsetting over the Rockies and later in the eastern US have been supported by other studies (Li et al., 2021a; Villarini, 2016; Xu et al., 2021).

8.4.4 Reconciling different rainfall-flood patterns in the western and eastern US

Floods are modulated not only by the atmosphere and climate mechanisms but also catchment states (e.g., antecedent soil moisture, groundwater level, land-use changes, snowmelt, and rain on snow) and river networks (Merz et al., 2020). In this study, we found different patterns for the changes in extreme rainfall and floods across the US. The western US is likely to experience greater (less) changes in flood (rainfall) frequency and spatial scales, as opposed to eastern US whose changes in rainfall (flood) frequency and spatial scales are greater (lesser). One plausible reason is the changes in antecedent soil moisture. To verify whether the future catchment state is

drying, especially in the eastern US, we extracted antecedent soil moisture, one day prior to the onset of flooding events, as shown in Fig. S8.7. Distinct spatial patterns appear in Fig. S8.7a – higher values in the eastern US and lower in the western US. In the future, there is a drying tendency of antecedent soil moisture conditions over a large portion of the US, especially in the Southern Plains. The eastern US generally experiences 0-5% drier soil moisture conditions in the future, in contrast to 5-10% wetter in the Pacific Coast and Great Basin (Fig. S8.7b). It explains why the future increase in rainfall extent is more closely tied to flood extent in the West than the East, as the dry antecedent soils modulate the intense rainfall.

8.5 Conclusion

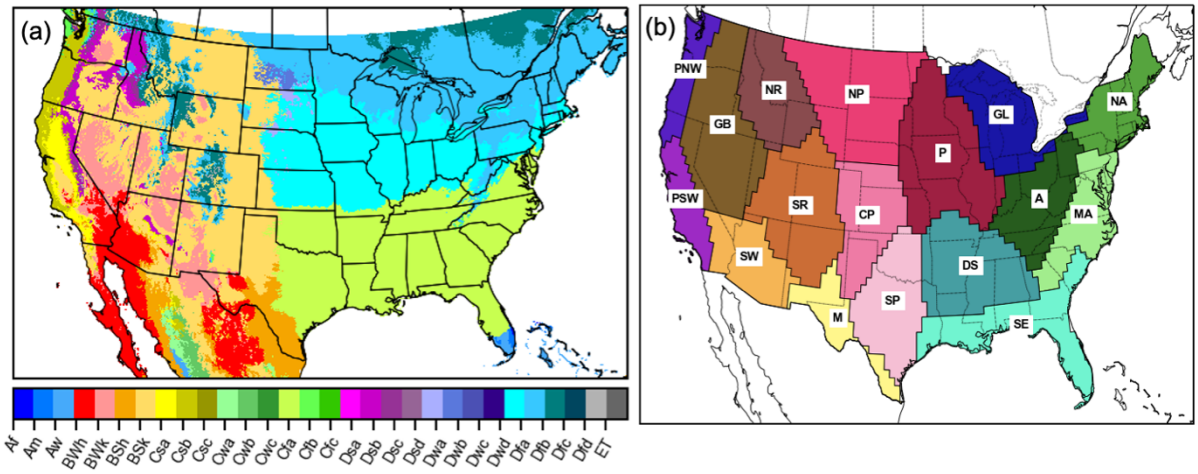
In this study, we explore the spatiotemporal patterns of US floods under the current climate and how they change under a warmer climate. A convection-permitting climate simulation enables high-resolution flood simulation (CREST model + Snow 17 model), which reveals a continental perspective of reach-level flood spatiotemporal characteristics. The following points summarize the primary findings of the study:

1. In the future climate scenario, the two-year rainfall occurrences and spatial scales increase by 118.0% and 101.7%, respectively. As a consequence, the two-year flood occurrences and spatial scales increase by 101.7% and 44.9% across the US.
2. Earlier and Faster snowmelt bring earlier flood season onset and shorten rainfall-to-flood timing, which results in 0.11 and 0.14 stronger correlation in date of occurrence and seasonality.
3. Regionally, floods in the western US become relatively more severe (+70.6% greater flood spatial scale) in the future than those in the east (+35.8% greater flood spatial scale), as snowmelt exacerbates flooding in the West, while the drier antecedent catchment soil moisture in the East partially offsets flooding.
4. The Pacific Coast, both currently and in the future, has the strongest seasonality for extreme rainfall (0.80) and floods (0.85) in the US However, seasonality decreases moving towards the East Coast (0.4 for rainfall and 0.3 for floods). The seasonal cycles of rainfall and floods become less pronounced in the East, partly caused by more evenly distributed extreme

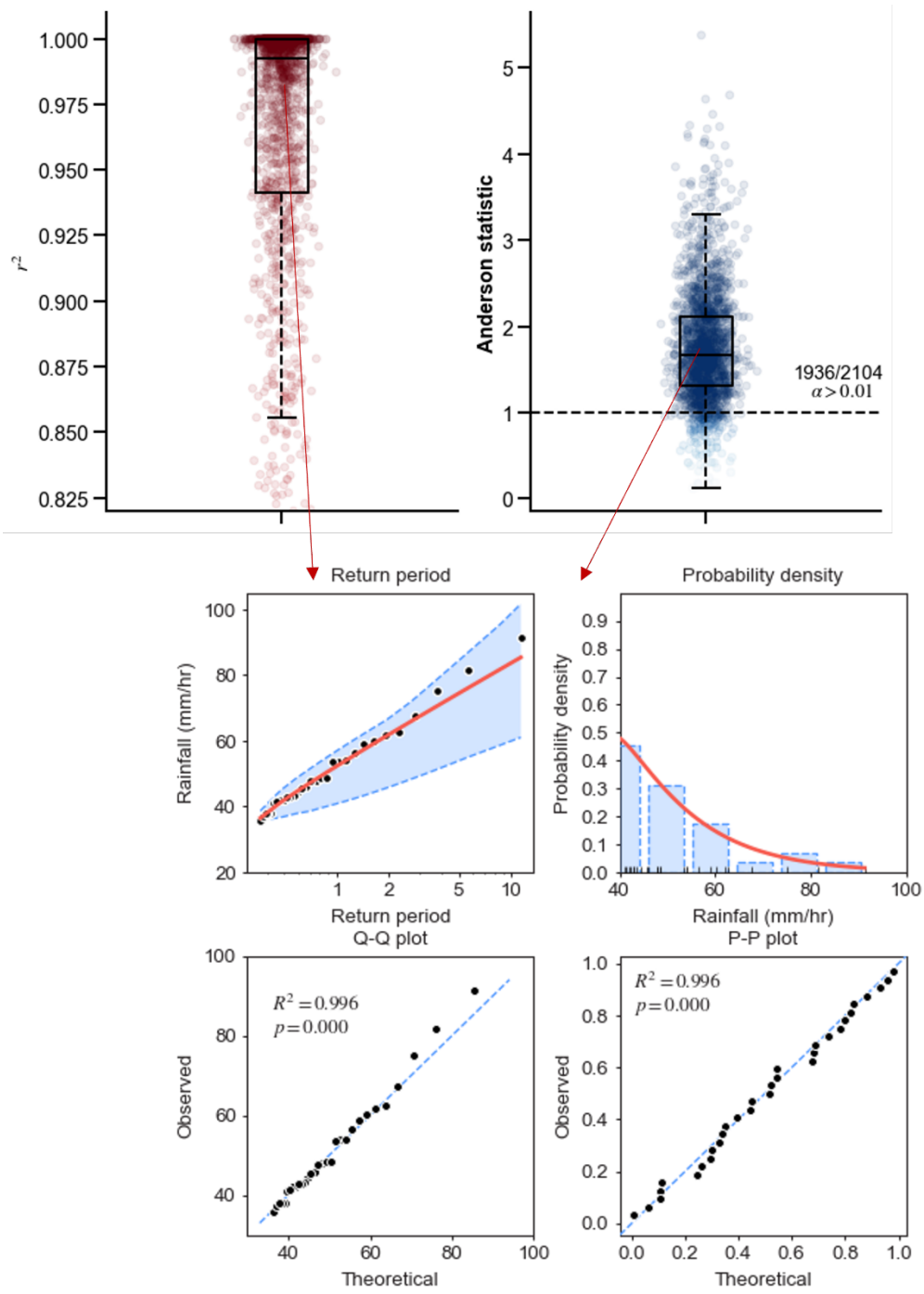
rainfall days, earlier snowmelt, and lengthened rainfall-runoff lag time in watersheds with drier soils in the future.

This study reveals a pressing need for the national weather and hydrology community to adapt to climate change because of not only intensified frequency in rainfall and floods but also more widespread rainfall and floods. Even worse, the weakening seasonal cycle of extreme rainfall and floods challenges flood predictability and preparedness. We recommend future works incorporating anthropogenic effects on floods in detail to present an economic assessment of future flood damages and resilience measures.

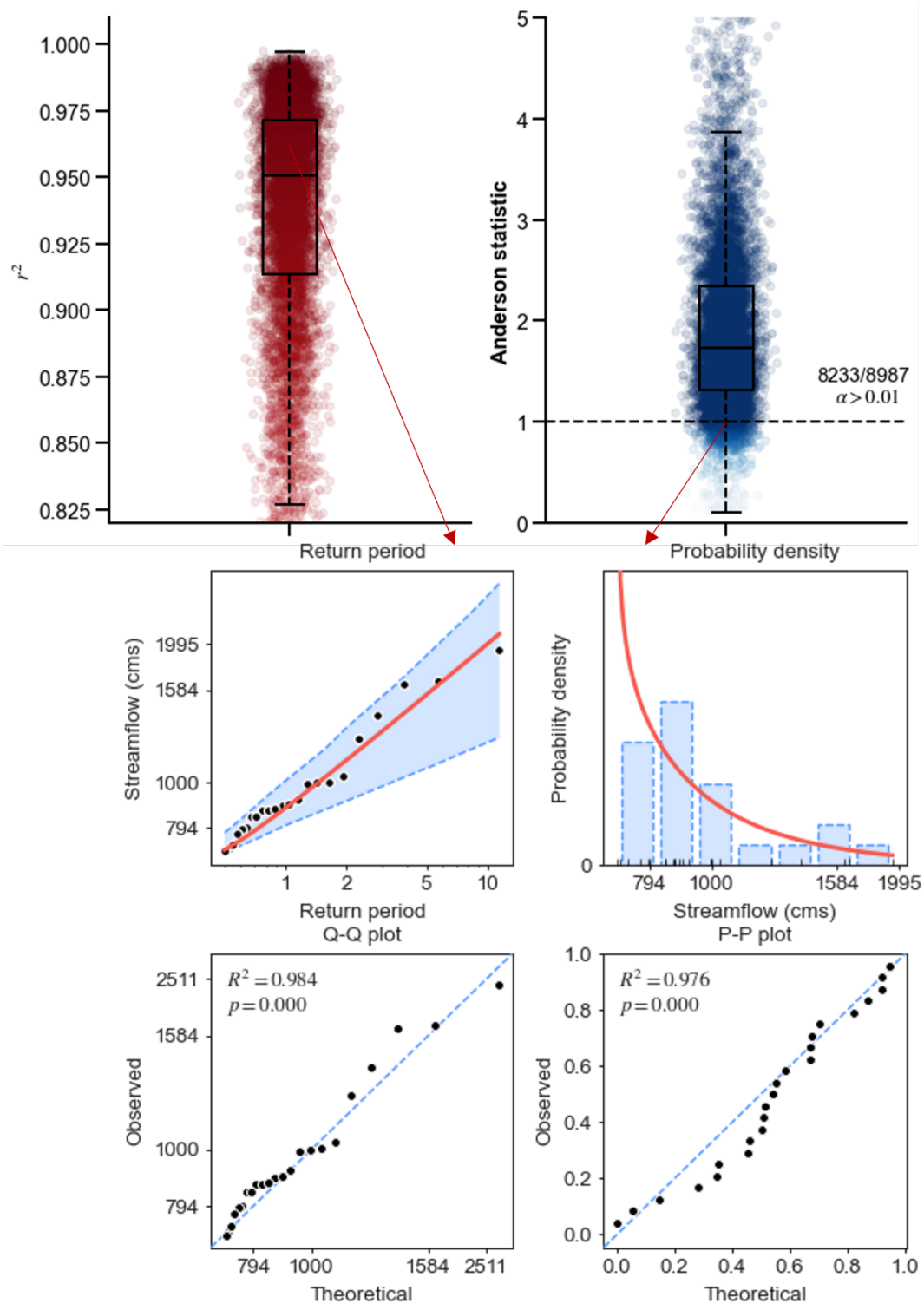
8.6 Supplementary information



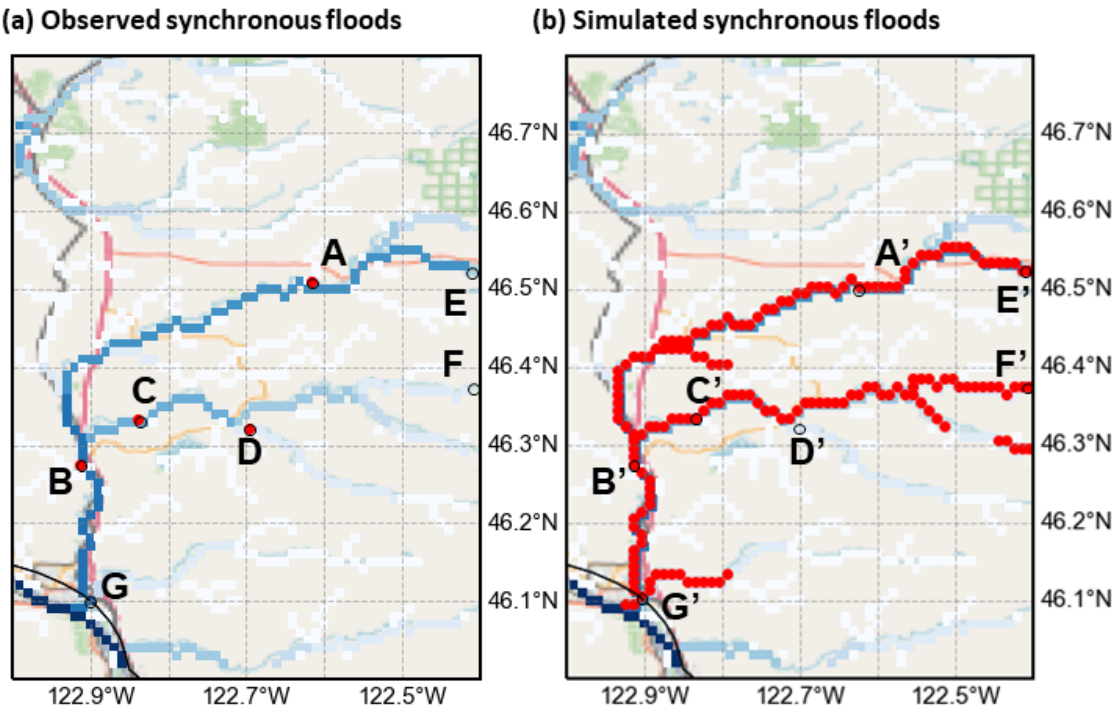
Supplementary Figure 8.1. Maps of Climate divisions based on (a) Koppen-Geiger and (b) Bukovsky. Symbols in (a) represent: Af – Tropical rainforest; Am – Tropical monsoon; Aw – Tropical savannah; BWh – Arid, desert, hot; BWk – Arid, desert, cold; BSh – Arid, steppe, hot; BSk – Arid, steppe, cold; Csa – Temperate, dry summer, hot summer; Csb – Temperate, dry summer, warm summer; Csc Temperate, dry summer, cold summer; Cwa – Temperate, dry winter, hot summer; Cwb – Temperate, dry winter, warm summer; Cwc – Temperate, dry winter, cold summer; Cfa – Temperate, no dry season, hot summer; Cfb – Temperate, no dry season, warm summer; Cfc – Temperate, no dry season, cold summer; Dsa – Cold, dry summer, hot summer; Dsb – Cold, dry summer, warm summer; Dsc – Cold, dry summer, cold summer; Dsd – Cold, dry summer, very cold winter; Dwa – Cold, dry winter, hot summer; Dwb – Cold, dry winter, warm summer; Dwc – Cold, dry winter, cold summer; Dwd – Cold, dry winter, very cold winter; Dfa – Cold, no dry season, hot summer; Dfb – Cold, no dry season, warm summer; Dfc – Cold, no dry season, cold summer; Dfd – Cold, no dry season, very cold winter; E.T. – Polar, tudra; E.F. – Polar frost. Symbols in (b) represent: PNW – Pacific Northwest; PSW – Pacific Southwest; G.B. – Great Basin; S.W. – Southwest; N.R. – Northern Rockies; S.R. – Southern Rockies; M – Meziquital; N.P. – Northern Plain; C.P. – Central Plain; S.P. – Southern Plain; P – Prairie; D.S. – Deep South; S.E. – Southeast; G.L. – Great Lakes; A – Appalachian; M.A. – Mid-Atlantic; N.A. – North-Atlantic.



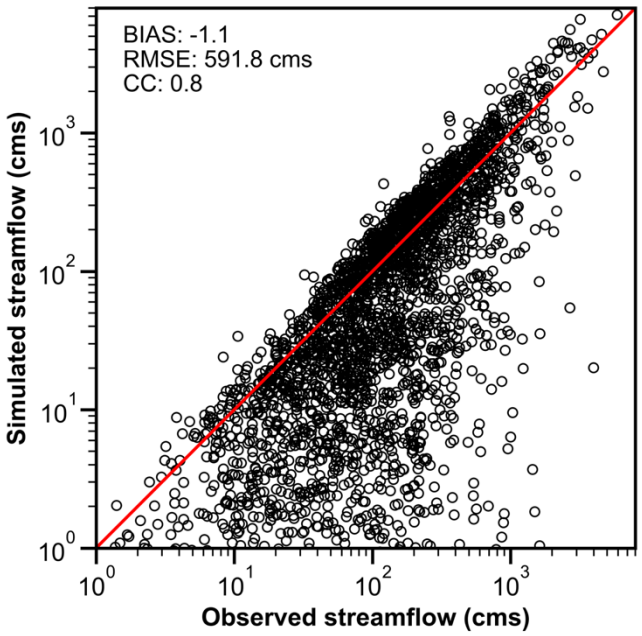
Supplementary Figure 8.2. Goodness-of-fit test for extreme rainfall (fitted by Gamma distribution) and tested with Anderson-Darling test.



Supplementary Figure 8.3. Goodness-of-fit test for extreme streamflow (fitted by Log-Pearson type III distribution) and tested with Anderson-Darling test.



Supplementary Figure 8.4. Illustration of advantages of simulating flood synchrony by distributed hydrologic models. The red dots are synchronous floods by (a) USGS stream gauges and (b) distributed hydrologic model. The flood event is taken from 2003-02-01, an atmospheric river event contributing to catastrophic flood event in California. Flooded points are overlaid with 1-km flow accumulation image used for hydrologic simulation.



Supplementary Figure 8.5. Scatter plot of simulated and observed two-year flow. BIAS: relative bias, calculated by summation of simulated streamflow minus summation of observed streamflow and divide by summation of observed. RMSE: Root Mean Square Error. CC: Correlation Coefficient.

8.7 Reference

Alfieri, L., Dottori, F., Salamon, P., Wu, H., & Feyen, L. (2020). Global modeling of seasonal mortality rates from river floods. Earth's Future, 8, e2020EF001541. https://doi.org/10.1029/2020EF001541

Ashley, W. S., Mote, T. L., Dixon, P. G., Trotter, S. L., Powell, E. J., Durkee, J. D., & Grundstein, A. J. (2003). Distribution of Mesoscale Convective Complex Rainfall in the United States, Monthly Weather Review, 131(12), 3003-3017.

Barbero, R., Abatzoglou, J.T. & Fowler, H.J. (2019). Contribution of large-scale midlatitude disturbances to hourly precipitation extremes in the United States. Climate Dynamics, 52, 197–208. https://doi.org/10.1007/s00382-018-4123-5

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., Savage, J., Olcese, G., Neal, J., Shumann, G., Giustarini, L.,Coxon, G., Porter, J. R., Amodeo, M. F., Chu, Z., Lewis-Gruss, S., Freeman, N. B., Houser, T., Delgado, M., Hamidi, A., Bolliger, I., McCusker, K. E., Emanuel, K., Ferreira, C. M., Khalid, A., Haigh, I. D., Couasnon, A., Kopp, R. E., Hsiang, S., & Krajewski, W. F., Combined modelling of US fluvial, pluvial and coastal flood hazard under current and future climates. Water Resources Research, 56, e2020WR028673 (2020).

Berghuijs, W. R., Allen, S. T., Harrigan, S., & Kirchner, J. W. (2019). Growing spatial scales of synchronous river flooding in Europe. Geophysical Research Letters, 46, 1423– 1428. https://doi.org/10.1029/2018GL081883

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Vorga, M., Canjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gul, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen T. R., Kohnova, S., Koskela, J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J., Sauquet, E., Sraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K. & Živković, N. (2017). Changing climate shifts timing of European floods. Science, 357(6351), 588-590. https://doi.org/10.1126/science.aan2506

Bukovsky, M.S., (2011). Masks for the Bukovsky regionalization of North America, Regional Integrated Sciences Collective, Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research, Boulder, CO. Downloaded 2021-07-05. http://www.narccap.ucar.edu/contrib/bukovsky/

Burn, D.H., 1997. Catchment similarity for regional flood frequency analysis using seasonality measures. Journal of Hydrology, 202, 212-230. https://doi.org/10.1016/S0022-1694(97)00068-1

Brunner, M. I., Gilleland, E., Wood, A., Swain, D. L., & Clark, M. (2020). Spatial dependence of floods shaped by spatiotemporal variations in meteorological and land-surface processes. Geophysical Research Letters, 47, e2020GL088000. https://doi.org/10.1029/2020GL088000

Brunner, M. I., Swain, D. L., Wood, R. R. Willkofer, F., Done, J. M., Gilleland, E. & Ludwig, R.. (2021) An extremeness threshold determines the regional response of floods to changes in rainfall

extremes. Communications Earth and Environment, 173. https://doi.org/10.1038/s43247-021-00248-x

Clark, M. P., Wilby, R. L., Gutmann, E. D., Vano, J. A., Gangopadhyay, S., Wood, A., Fowler, H. J., Prudhomme, C., Arnold, J. R. & Brekke, L. D. (2016). Characterizing Uncertainty of the Hydrologic Impacts of Climate Change. Current Climate Change Reports, 2, 55–64.

Dougherty, E., & Rasmussen, K. L. (2020). Changes in Future Flash Flood–Producing Storms in the United States, Journal of Hydrometeorology. 21, 2221-2236 (2020).

Farnham, D. J., Doss-Gollin, J., & Lall, U. (2018). Regional extreme precipitation events: Robust inference from credibly simulated GCM variables. Water Resources Research, 54, 3809– 3824. https://doi.org/10.1002/2017WR021318.

Feng, Z., Leung, L., Hagos, S. Houze, R. A., Burleyson, C. D., & Balaguru K., (2016). More frequent intense and long-lived storms dominate the springtime trend in central US rainfall. Nature Communications, 7, 13429. https://doi.org/10.1038/ncomms13429

Flamig, Z. (2020). HyDROSLab/EF5-US-Parameters: EF5 parameters for USA (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.4009759

Fritsch, J. M., Kane, R. J. & Chelius, C. R. (1986). The contribution of mesoscale convective weather systems to the warm-season precipitation in the United States. Journal of Applied Meteorology and Climatology, 25, 1333–1345.

Giuntoli, I., Prosdocimi, I., & Hannah, D. M. (2021). Going beyond the ensemble mean: Assessment of future floods from global multi-models. Water Resources Research, 57, e2020WR027897. https://doi.org/10.1029/2020WR027897

Gourley, J.J., Flamig, Z.L., Vergara, H., Kirstetter, P., Clark, R.A., Argyle, E., Arthur, A., Martinaitis, S., Terti, G., Erlingis, J.M., Hong, Y., & Howard, K.W., 2017. The FLASH Project: Improving the Tools for Flash Flood Monitoring and Prediction across the United States. Bulltin of the American Meteorlogical Society, 98, 361–372, https://doi.org/10.1175/BAMS-D-15-00247.1

Hagos, S. M., Leung, L. R., Yoon, J.-H., Lu, J., & Gao, Y. (2016), A projection of changes in landfalling atmospheric river frequency and extreme precipitation over western North America from the Large Ensemble CESM simulations, Geophysical Research Letters, 43, 1357– 1363, https://doi.org/10.1002/2015GL067392.

He, L. & Wilkerson, G.V. (2011), Improved Bankfull Channel Geometry Prediction Using Two-Year Return-Period Discharge. Journal of the American Water Resources Association, 47: 1298-1316. https://doi.org/10.1111/j.1752-1688.2011.00567.x

Hirabayashi, Y., Mahendran, R., Koirala, S. Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., Kanae, S. (2013) Global flood risk under climate change. Nature Climate Change, 3, 816–821 (2013). https://doi.org/10.1038/nclimate1911

Ivancic, T.J., & Shaw, S.B. (2015). Examining why trends in very heavy precipitation should not be mistaken for trends in very high river discharge. Climatic Change 133, 681–693. https://doi.org/10.1007/s10584-015-1476-1

Jongman, B., Ward, P., & Aerts, J. C.J.H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. Global Environmental Change, 22, 823-835.

Kemter, M., Merz, B., Marwan, N., Vorogushyn, S., & Blöschl, G. (2020). Joint trends in flood magnitudes and spatial extents across Europe. Geophysical Research Letters, 47, e2020GL087464. https://doi.org/10.1029/2020GL087464

Kendon, E. J., Roberts, N. M., Senior, C. A., & Roberts, M. J. (2012). Realism of rainfall in a very high-resolution regional climate model. Journal of Climate, 25(17), 5791– 5806. https://doi.org/10.1175/JCLI-D-11-00562.1

Lamjiri, M. A., Dettinger, M. D., Ralph, F. M., & Guan, B. (2017), Hourly storm characteristics along the US West Coast: Role of atmospheric rivers in extreme precipitation, Geophysical Research Letters, 44, 7020– 7028, https://doi.org/10.1002/2017GL074193.

Ledingham, J., Archer, D., Lewis, E., Fowler, H., Kilsby, C., 2019. Contrasting seasonality of storm rainfall and flood runoff in the UK and some implications for rainfall-runoff methods of flood estimation. Hydrology Research; 50 (5): 1309–1323. https://doi.org/10.2166/nh.2019.040

Li, Z., Yang, Y., Kan, G., & Hong, Y. (2018). Study on the Applicability of the Hargreaves Potential Evapotranspiration Estimation Method in CREST Distributed Hydrological Model (Version 3.0) Applications. Water, 10, 1882. https://doi.org/10.3390/w10121882

Li, Z., Chen, M., Gao, S., Gourley, J. J., Yang, T., Shen, X., Kolar, R., & Hong, Y., (2021a). A multi-source 120-year US flood database with a unified common format and public access, Earth System Science Data, https://doi.org/10.5194/essd-2021-36.

Li, Z., Chen, M., Gao, S., Luo, X., Gourley, J. J., Kirstetter, P., Yang, T., Kolar, R., McGovern, A., Wen, Y., Rao, B., Yami, T. & Hong, Y. (2021b). CREST-iMAP v1. 0: A fully coupled hydrologic-hydraulic modeling framework dedicated to flood inundation mapping and prediction. Environmental Modelling & Software, 141, 105051.

Li, Z., Gao, S., Chen, M. Gourley, J.J., Liu, C., Prein, A., Hong, Y. (2022). The conterminous United States are projected to become more prone to flash floods in a high-end emissions scenario. Commun Earth Environ 3, 86. https://doi.org/10.1038/s43247-022-00409-6

Merz, B., Blöschl, G., Vorogushyn, S. Dottori, F., Aerts, J. C. J.H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U. & Macdonald, E. (2021). Causes, impacts and patterns of disastrous river floods. Nature Review Earth & Environment, 592–609. https://doi.org/10.1038/s43017-021-00195-3

Musselman, K.N., Lehner, F., Ikeda, K., Clark, M., Prein, A., Liu, C., Barlage M. & Rasmussen R. (2018). Projected increases and shifts in rain-on-snow flood risk over western North America. Nature Climate Change, 8, 808–812. https://doi.org/10.1038/s41558-018-0236-4

Nie, J., Sobel, A.H., Shaevitz, D. A., & Wang, S. (2018). Dynamic amplification of extreme precipitation sensitivity, Preceedings of the National Academy of Sciences of the United States of America, 115, 9467-9472. https:/doi.org/10.1073/pnas.1800357115

Prein, A.F., Gobiet, A., Suklitsch, M. Truhetz, H., Awan, N. K., Keuler, K., & Georgievski, G. (2013). Added value of convection permitting seasonal simulations. Climate Dynamics, 41, 2655–2677. https://doi.org/10.1007/s00382-013-1744-6

Prein, A.F. Liu, C., Ikeda, K., Trier, S. B., Rasmussen, R. M., Holland, G. J. & Clark, M. (2017). Increased rainfall volume from future convective storms in the US Nature Climate Change, 7, 880–884. https://doi.org/10.1038/s41558-017-0007-7

Prein, A. F., & Mearns, L. O. (2021). US Extreme precipitation weather types increased in frequency during the 20th century. Journal of Geophysical Research: Atmospheres, 126, e2020JD034287. https://doi.org/10.1029/2020JD034287

Rajib, A., Zheng, Q., Golden, H. E., Wu, Q., Lane, C. R., Christensen, J. R., Morrison, R. R., Annis, A. & Nardi, F. (2021). The changing face of floodplains in the Mississippi River Basin detected by a 60-year land use change dataset. Scientific Data, 8, 271. https://doi.org/10.1038/s41597-021-01048-w

Douville, H., K. Raghavan, J. Renwick, R. P. Allan, P. A. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Y. Gan, J. Gergis, D. Jiang, A. Khan, W. Pokam Mba, D. Rosenfeld, J. Tierney, O. Zolina, 2021, Water Cycle Changes. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu and B. Zhou (eds.)]. Cambridge University Press. In Press.

Schär, C., Frei, C., Lüthi, D., & Davies, H. C. (1996). Surrogate climate-change scenarios for regional climate models, Geophysical Research Letters, 23, 669-672.

Smith J.A., Villarini, G., & Baeck, M. L. (2011). Mixture distributions and the climatology of extreme rainfall and flooding in the Eastern US, Journal of Hydrometeorology, 12 (2), 294-309.

Stein, L., Clark, M. P., Knoben, W. J. M., Pianosi, F., & Woods, R. A. (2021). How do climate and catchment attributes influence flood generating processes? A large-sample study for 671 catchments across the contiguous USA. Water Resources Research, 57, e2020WR028300. https://doi.org/10.1029/2020WR028300

Swain, D. L. Wing, O. E.J., Bates, P. D., Done, J. M., Johnson, K. A. & Cameron, D. R. (2020). Increased flood exposure due to climate change and population growth in the United States. Earth's Future, 8, e2020EF001778.

Tellman, B., Sullivan, J.A., Kuhn, C. Kettner, A. J., Doyle, C. S., Brakenridge, G. R., Erickson, T. A. & Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. Nature, 596, 80–86. https://doi.org/10.1038/s41586-021-03695-w

Thornthwaite, C. W. (1948) An approach toward a rational classification of climate. Geographical Review, 38, 55-94.

Titley, H. A., Cloke, H. L., Harrigan, S., Pappenberger, F., Prudhomme, C., Robbins, J. C., Stephens, E. M., & Zsótér, E. (2021). Key Factors Influencing the Severity of Fluvial Flood Hazard from Tropical Cyclones, Journal of Hydrometeorology, 22(7), 1801-1817.

Villarini, G. (2016). On the seasonality of flooding across the continental United States, Advances in Water Research, 87, 80-91. https://doi.org/10.1016/j.advwatres.2015.11.009.

Wasko, C., & Sharma, A. (2017). Global assessment of flood and storm extremes with increased temperatures. Scientific Reports, 7, 7945. https://doi.org/10.1038/s41598-017-08481-1

Wasko, C., Sharma, A. & Lettenmaier, D.P. (2019). Increases in temperature do not translate to increased flooding. Nature Communications, 10, 5676. https://doi.org/10.1038/s41467-019-13612-5

Wasko, C., Nathan, R., Stein, L., & O'Shea, D. (2021). Evidence of shorter more extreme rainfalls and increased flood variability under climate change. Journal of Hydrology, 603, 126994. https://doi.org/10.1016/j.jhydrol.2021.126994

Wu, H., Adler, R. F., Hong, Y., Tian, Y., & Policelli, F. (2012). Evaluation of Global Flood Detection Using Satellite-Based Rainfall and a Hydrologic Model, Journal of Hydrometeorology, 13(4), 1268-1284.

Xu, D., Ivanov, V. Y., Li, X., & Troy, T. J. (2021). Peak runoff timing is linked to global warming trajectories. Earth's Future, 9, e2021EF002083. https://doi.org/10.1029/2021EF002083

Yu, G., Wright, D. B., & Li, Z. (2020). The upper tail of precipitation in convection-permitting regional climate models and their utility in nonstationary rainfall and flood frequency analysis. Earth's Future, 8, e2020EF001613. https://doi.org/10.1029/2020EF001613

Zhang, K., Xue, X., Hong, Y., Gourley, J. J., Lu, N., Wan, Z., Hong, Z., & Wooten, R. (2016). iCRESTRIGRS: a coupled modeling system for cascading flood–landslide disaster forecasting, Hydrology and Earth System Science, 20, 5035–5048. https://doi.org/10.5194/hess-20-5035-2016, 2016.

9 Chapter 9. Concluding remarks and outlook

9.1 Summary and conclusion

This dissertation is dedicated to a decadal review, development, application, and outlook of CREST hydrological model family, centered on three fronts: (1) hydrologic data, (2) model development, and (3) applications. To start, the decadal development and applications of CREST model were reviewed to lay the foundation for the contribution of this dissertation. First, the uncertainties in essential hydrologic data were evaluated and quantified for three state-of-the-science precipitation data sources derived from in-situ instruments, ground weather radars, and satellites during extreme events (**Chapter 2**); then a 120-year CONUS-wide flood database was compiled into an unified format for public access (**Chapter 3**). From the model development front, an original Hydrologic&Hydraulic (H&H) – CREST-iMAP model was developed to empower flood inundation mapping for CREST (**Chapter 4**); and furthermore, the re-infiltration, an important yet often ignored hydrologic process during flooding period, was incorporated to improve the more realistic rainfall-runoff modeling representation (**Chapter 5**); and later, a vector-based version of CREST model – CREST-VEC was developed to achieve 10x speedup for a continental-scale simulation (**Chapter 6**), all of which enhance flood prediction capability that is a backbone for CREST model. At last, the high-resolution CREST model was applied in quantifying future US floods in a warmer climate: US flood flashiness (**Chapter 7**) and spatiotemporal characteristics (**Chapter 8**). Major conclusions from this dissertation are summarized below:

1. CREST model family has been a widely recognized and used hydrologic framework over the last decade, a cornerstone for remote sensing and hydrologic streamflow prediction, flood prediction and forecast, water resources management, and capacity building and educational outreach.
2. Modern gridded precipitation products bear large uncertainties in extreme rainfall events, which can propagate to streamflow prediction. Triple Collocation is proven to be a valid tool to quantify such uncertainties without reference.
3. The 120-year flood database offers unprecedented opportunities for a variety of flood-related research, such as validation source for model simulations, hydroclimatic studies, and flood risk analysis. Flood resulted damages have been increasing over the last 50 years, especially for coastal states like Texas and Louisiana.

4. CREST-iMAP model, a highly parallelized and integrated H&H model, demonstrates its applicability for regional operational flood prediction and inundation mapping. This coupled H&H model is superior to a standalone hydrologic or hydraulic model.
5. Re-infiltration is an important yet ignored hydrologic process in modern hydrodynamic models. CREST-iMAP model with re-infiltration generates better streamflow and flood depth, even for a 500-year Hurricane Harvey event than that without re-infiltration.
6. CREST-VEC, a vector version of CREST model, can accelerate continental-scale river routing by up to ten times, while improving streamflow simulation by 56.2%, compared to original CREST model. Importantly, flood false alarm ratios can be mitigated by 20.6% due to enabled lake routing.
7. A CONUS-wide 7.9% increase in future flood flashiness is projected, meaning future floods will onset more rapidly with higher peak runoff, resulting in even shorter times for early warning. In particular, floods in the Southwest (central US) become 10.5% (8.5%) flashier, both of which are above nationwide average.
8. Projected by a warmer climate, extreme rainfall and floods are becoming more frequent, widespread, and less seasonal in the US. Flood timing is shifted to earlier flooding in the West (because of snowmelt) while later in the East (because of drying soils).

CREST model has been an easy-to-implement and powerful hydrologic model in community and will continue to serve as flood forecasting and water resources management tools for the community. In the future, there are several fronts we can push the envelope to make it more accurate and efficient for large-scale hydrologic simulation. Below is a list of outlooks and challenges for future development.

9.2 Outlook and challenge

CREST model carries the mission of global high-resolution hydrologic simulation to inform stakeholders, decision-makers, and the public. Future development should be centered in this end. In practice, we eye on improving the model efficiency, accuracy, and precision as a multi-objective function.

First, a large-scale high-resolution hydrologic simulation demands fast model execution. CREST-VEC exemplifies the efforts we put on improving the routing efficiency, but there is still room to optimize the lag time between the water balance model and routing model. Coupled with

high resolution remote sensing dataset, it is opportunistic to operate CREST-VEC model at a scale of tens of meters.

Second, we hope to implement a large-scale model calibration to further enhance model accuracy, although it is a challenge for modern hydrologic models. There are promising tricks such as parameter regionalization and surrogate modeling that can reduce computational costs for calibrating a distributed model.

Third, we can adopt ensemble prediction to enhance model precision and reduce uncertainty. Ensemble members can be generated from perturbations from hydrologic data, parameters, and model structures to embrace uncertainties from these three sources.