Regional Flood Frequency Analysis Using Physics-based Hydrologic Modeling

Ricardo Mantilla[1], Gabriel Perez[2], Nicolas Velasquez[3], Daniel Benjamin Wright[4], and Guo Yu[4]

[1]IIHR - Hydroscience & Engineering
[2]Vanderbilt University
[3]The university of iowa
[4]University of Wisconsin-Madison

November 26, 2022

Abstract

We investigate the validity of implicit assumptions in regional flood frequency analysis (RFFA) using Monte Carlo-style simulations of three distributed hydrological models forced with rainfall events generated using stochastic storm transposition. We test the long-standing assumption that for a set of sites within a region, *physical homogeneity* — defined in terms of the variability of meteorological inputs, the physics of runoff generation, and runoff routing — implies *statistical homogeneity* of peak flows— defined in terms of the existence of a common underlying statistical distribution with parameters that can be inferred using information from neighboring sites. Our modeling results do not support this assumption, with potentially important implications for RFFA methodologies and for the very definitions of homogeneity. We show that statistically homogeneous rainfall does not translate into predictable peak flow distribution parameters across drainage scales. Specifically, we show that changes in the coefficient of variation and skewness of peak flows cannot be inferred from upstream area alone, making popular regionalization techniques such as the index-flood method and quantile regression inadequate approximations for flood frequency estimation. Our findings are consistent across the three hydrological model formulations, lending confidence that our conclusions are not an artifact of epistemological model decisions. Finally, we argue that our methodology can serve as a framework to test new proposed empirical RFFA methods, and that it opens the door to a unified physics-informed framework for prediction of flood frequencies in ungauged basins embedded in gauged regions.

Regional Flood Frequency Analysis

Using Physics-based Hydrologic Modeling

[1]Ricardo Mantilla, [2]Gabriel Perez, [3]Nicolas Velasquez, [4]Daniel B. Wright, and [4]Guo Yu

[1]Iowa Flood Center, Department of Civil and Environmental Engineering, The University of Iowa, Iowa City, Iowa 52242, USA

[2]Department of Civil and Environmental Engineering, Vanderbilt University, Nashville, Tennessee 37235, USA.

[3] Iowa Flood Center, The University of Iowa, Iowa City, Iowa 52242, USA

[4]Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

Key Points

1. We test the validity of implicit assumptions in empirical regional flood frequency analysis using three hydrological models and stochastic storm transposition

2. We show that a link between physical homogeneity in hydrological processes and homogeneity of statistical distributions in peak flows cannot be assumed

3. Our work provides a framework to establish a physics-based regional flood frequency analysis

Submitted to

Water Resources Research

Corresponding author: Ricardo Mantilla, e-mail: ricardo-mantilla@uiowa.edu

Abstract: We investigate the validity of implicit assumptions in regional flood frequency analysis (RFFA) using Monte Carlo-style simulations of three distributed hydrological models forced with rainfall events generated using stochastic storm transposition. We test the long-standing assumption that for a set of sites within a region, *physical homogeneity* — defined in terms of the variability of meteorological inputs, the physics of runoff generation, and runoff routing — implies *statistical homogeneity* of peak flows— defined in terms of the existence of a common underlying statistical distribution with parameters that can be inferred using information from neighboring sites. Our modeling results do not support this assumption, with potentially important implications for RFFA methodologies and for the very definitions of homogeneity. We show that statistically homogeneous rainfall does not translate into predictable peak flow distribution parameters across drainage scales. Specifically, we show that changes in the coefficient of variation and skewness of peak flows cannot be inferred from upstream area alone, making popular regionalization techniques such as the index-flood method and quantile regression inadequate approximations for flood frequency estimation. Our findings are consistent across the three hydrological model formulations, lending confidence that our conclusions are not an artifact of epistemological model decisions. Finally, we argue that our methodology can serve as a framework to test new proposed empirical RFFA methods, and that it opens the door to a unified physics-informed framework for prediction of flood frequencies in ungauged basins embedded in gauged regions.

Keywords: Regional flood frequency analysis (RFFA), peak flow scaling, stochastic storm transposition, physical homogeneity, statistical homogeneity, hydrologic modeling, Bulletin 14D.

Plain Language Summary

So-called “regionalization” techniques are useful for estimating design peak river flows such as the “100-year flood” at locations that lack flow measurements. Although these strategies are widely used in flood protection, several associated assumptions are unproven because of limitations in data availability in terms of spatial sparsity and record lengths. This study tests some of the long-standing assumptions in regional flood frequency analysis (RFFA) using hydrologic simulations that include not only a large number of observed rainfall scenarios, but also by the support of three different hydrologic model formulations. Our findings indicate that several of the assumptions used in RFFA are incorrect, highlighting the need to consider new strategies that benefit from physics-informed hydrologic process modeling.

1. Introduction

The systematic power-law scaling of riverine flood peak flows (Q) with respect to drainage area (A), first identified by Fuller, (1914), has since become a well-established feature in hydrological research and practice. This scaling describes the regional change of characteristic flows (e.g. mean annual floods, bank-full flows, 100-year floods) across spatial scales as $Q = \alpha A^{\theta}$, with α and θ known as the scaling intercept and scaling exponents, respectively. This type of scaling arises in scale-invariant systems with self-similarity properties (V. K. Gupta et al., 2007) and seems to reflect a fundamental symmetry of nature (Schroeder, 2012). This result gave rise to empirical frameworks for Regional Flood Frequency Analysis (RFFA) to predict peak flow distributions at ungagged locations. RFFA, in its many forms, is the most widely accepted technique for estimation of annual flood quantiles (or also known as peak flow quantiles) in the United States and elsewhere, with results being used for infrastructure design, land use planning, and insurance.

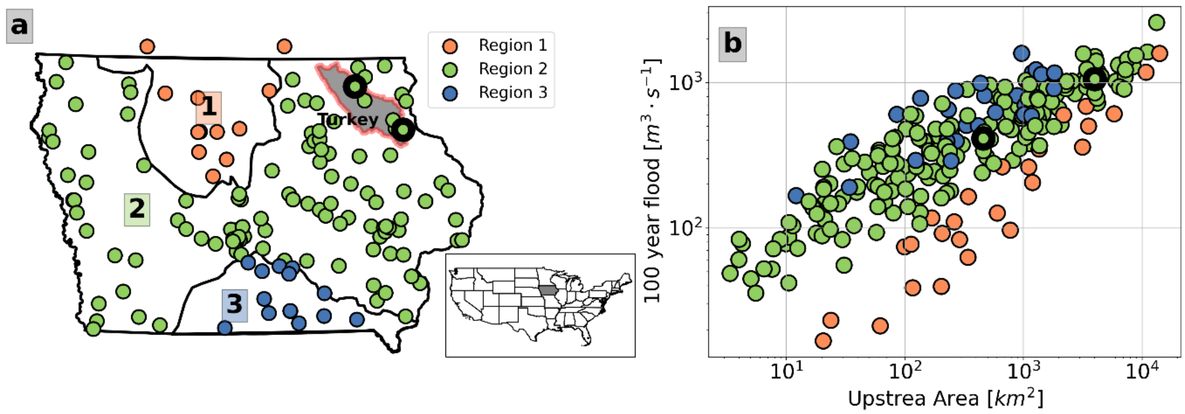


Figure 1. a) Location of long-term streamflow gauges in the state of Iowa, USA, color coded according to three homogeneous regions defined by Eash, (2001), and b) estimated 100-year flood peak for each gauge location. Both panels also show the locations of two gauges in the Turkey river watershed (dark outlined circles).

RFFA techniques generally employ both sparse data sets and unproven assumptions that seem to hold in idealized individual rainfall-runoff events (Furey et al., 2016; Perez et al., 2019b) and for highly idealized river network structures (e.g. Gupta et al., 1996; Menabde & Sivapalan, 2001). In Iowa, for example, 307,714 km of rivers are gauged at only 140 unevenly distributed locations (Figure 1a), which tend to coincide with the most densely populated municipalities. This leaves the large majority of the state's waterways—and riverine communities—ungauged and dependent on empirical RFFA techniques to determine flood protection strategies. The power law scaling pattern in Figure 1b between the estimated 100-year flood using Bulletin 14B methodologies is evident, but it's also remarkable that for a typical ungauged watershed the smallest estimate for the 100-year flood is 6 times smaller than the largest estimate. Using a regression line through the cloud of points to estimate the 100-year flood for a generic 100 km^2 basin would certainly underestimate flood levels half of the time and overestimate the other half.

While the steps and statistical inferential tools within particular empirical RFFA methodologies vary from country to country (England et al., 2019; Kjeldsen, 2011; Meigh et al., 1997), they all rely on a central assumption: the peak flow distribution at an ungauged site can be inferred from observations at nearby locations that are hydrologically similar. The statistical inference process also relies on an underlying set of assumptions that remain largely untested including independence of peak flow data, stationarity, statistical homogeneity, scaling of distribution parameters, and the appropriateness of statistical estimators. In Canada, the United Kingdom, and the European Union, the preferred RFFA methodology is the index flood method, or simple-scaling as defined by Gupta et al., (1994; see Section 2.1), and the preferred underlying distribution of peak flows is the generalized extreme value (GEV) distribution (Reed & Robson, 2008; Zhang & Stadnyk, 2020). These methodologies are detailed in national reports such as the Review of Applied European Flood Frequency Analysis Methods (Kjeldsen, 2011). In the United States, the US Geological Survey (USGS) recommends the use of the Log-Pearson Type III (LP3) distribution, and the more flexible regional quantile regression method (IACWD, 1982), which has been described as exhibiting multi-scaling (Gupta et al., 1994; see Section 2.1). More sophisticated geostatistical methods have been proposed but are note widely used (e.g. Ouarda, (2013); Salinas et al., (2013)).

We argue in this paper that three modern developments in theoretical and computational hydrology have opened the door to investigating the validity of the empirical formulas employed in RFFA in a systematic fashion. First is the identification of power laws that connect the peak flows from extreme rainfall-runoff events in nested catchments as a function of drainage area (Ayalew et al., 2014; Gupta, 2004; Gupta et al., 2010; Ogden & Dawdy, 2003; Perez, et al 2019a; Perez, et al 2019b). Second is the development of simplified but realistic distributed hydrological models, along with efficient algorithms to integrate mass balance and water transport differential equations, that can predict, simultaneously, peak flows resulting from storm events at multiple locations. The third is the deployment of weather radars that observe precipitation fields at high spatial and temporal resolution, which in turn can inform stochastic methods—such as the Stochastic Storm Transposition (SST) framework used in this study—to generate many realizations of extreme storms.

Based on these three developments, recent work by (Perez, et al 2019b) suggested that even in a medium-sized catchment (~4,000 km^2) there is a complex spatial pattern in the third moment of the peak flow distribution (skewness) that cannot be explained by upstream area alone. This finding implies that most of the common RFFA models (e.g. the so-called Index Flood Method and regional quantile regressions) contain significant epistemic errors, thus casting doubt on several RFFA assumptions. Although these results highlighted the need to consider new strategies to better capture peak flow distributions across scales, Perez, et al (2019b) was limited to a single hydrologic model formulation and thus contained inherent epistemic errors in the representation of hydrologic processes; it has been argued that this can give rise to misleading conclusions (Beven, 2018).

Our aim in this paper is not only to generalize the findings of Perez et al., (2019b) by limiting these epistemic errors and to correct several more minor methodological shortcomings, but also to further investigate the implicit and explicit assumptions of existing empirical RFFA methods. In particular, we seek to test the long-standing assumption that for a set of sites within a region, *physical homogeneity* — defined in terms of the spatial variability of meteorological inputs, and the physics of runoff generation and runoff routing — implies *statistical homogeneity* of peak flows— defined in terms of the existence of a common underlying statistical distribution, and the ability to infer local distribution parameters (e.g. distribution moments). To this end we implemented three distributed hydrological models in a medium size basin (~4,000 km^2). The three models are all able to simulate streamflow dynamics at multiple internal sub-catchments simultaneously but vary in the level of complexity of their physical processes, spatial resolution, and spatial variability of model parameters. The three models are forced with a very large number of extreme storms created using SST to provide an approximation of the population of realistic peak flows, which we analyze statistically (We henceforth refer to this very large sample as a population). The use of SST helps us address issues related to sample size, while the use of the three hydrological models helps reduce biases caused by epistemic errors. Three overarching questions drive this analysis: (*i*) What is the link between physical homogeneity and statistical homogeneity among locations within a watershed or region?, (*ii*) what are appropriate simplifications for the scaling of the statistical moments of peak flow distributions across scales?, (*iii*) can the peak flow distribution parameters at an ungauged site be inferred from measurements at neighboring sites?

In Section 2 (*Background*), we briefly review the most relevant research that led to the assumptions that we are testing. In Section 3 (*Study Area)*, we describe the study watershed and the parameters used by each model. In Section 4 (*Methodology)* we describe our experimental setup including the three hydrological models, the SST methodology, and the methods that we use to analyze model outputs. In Section 5 (*Results)*, we analyze the rainfall, runoff production, and peak flow distributions generated by our numerical experiment. In Section 6 (*Discussion)*, we detail how our results inform the assumptions we are testing and the potential implications. Finally, in Section 7 (*Conclusions*) we summarize our major findings and suggest a road map to replace the current empirical RFFA framework by one informed by decades of accumulated hydrological knowledge of the physical characteristics and processes involved in the rainfall-runoff and routing transformations responsible for riverine flooding.

2. Background

Dawdy et al., (2012) details the origins of RFFA equations in the United States, including the adoption of both power laws that relate flood quantiles to explanatory variables and the LP3 distribution for estimating flood quantiles at gauged sites. Their review reveals that the steps involved in RFFA are largely ad hoc but justified from the available empirical observations of available flood data. They also explain RFFA's reliance on the existence of statistical homogeneity (defined in Sec. 2.1) of flood data. They emphasized the need of a more physically-based framework to connect event scale observations of peak flow scaling to the scaling of annual flood quantiles. Their rationale is that unlike at the event scale—where the connection between the physical phenomena that give rise to peak flows along a river network can be revealed—flood quantiles and the physical phenomena that control them have proven more elusive due to issues of statistical estimators, sample sizes, and other procedural issues. From our point of view, a well-developed physically-based framework of RFFA should provide estimates of uncertainty introduced by epistemic errors in the description of physical processes, sample size, nonstationarity, and techniques for including all available hydrological data in a region.

2.1. Statistical Homogeneity

At the core of RFFA lies the definition of statistical homogeneity of peak flow data and under what circumstances it exists. The simplest example of statistical homogeneity can be given in the context of the index flood method (IFM). In it, the underlying assumption about the distribution of peak flows is that the ratio of a peak flow quantile to an index flood (e.g. mean or median annual flood, bank full flood) is constant over a homogeneous geographic region (Kinnison & Colby, 1945). This feature is known as simple scaling of statistical distributions. In this context, a set of catchments is said to be homogeneous if the coefficients of variation $CV = \sigma/\mu$ of their peak flow distributions are the same and if higher order moments such as skewness are constant spatially. Gupta et al. (1994) developed a more general framework for the analysis of peak flow data, called multi-scaling, that did not require a constant coefficient of variation. Instead, they defined statistical homogeneity more generally: “A homogeneous region has, for all gaging locations within the region, flood peaks that follow probability distributions which are rescaled versions of one another, based only on drainage area and nothing else.” Gupta et al. (1994) specifically showed that under the assumption that the skewness is constant in a region, regressions of flood quantiles with different exceedance probabilities p would scale with respect to drainage area with exponents that depend on the quantile value, that is $Q_p = \alpha_p A^{\theta_p}$, and that CV would systematically increase for small catchments up to an area threshold and then decrease beyond it, consistent with USGS flood data. The results by Gupta et al. (1994) provided a theoretical footing for the methodologies in Bulletin 17 (WRC, 1977) and Bulletin 17B (IACWD, 1982) of using similarity in regional skewness to define homogeneous regions that continue to be used today in the most recent release of the Bulletin 14C (England et al. 2019).

While the assumptions of simple or multi-scaling have been adopted to describe peak flow variability over a region of interest, the reality is that these assumptions cannot be rejected by observations, mainly because of large sampling errors—that is, severely limited sample sizes—that obscure the true variability of the peak flow distributions over a region. (Perez et al., 2019b) showed this via minimization of sampling errors through the generation of a large number of realistic rainfall-runoff events. Their work demonstrated that skewness exhibits a complex spatial structure that cannot be explained solely by drainage area, in contrast with the definition of simple scaling and multi-scaling. Our study builds on that approach.

2.2. Physical Homogeneity

A parallel definition—*physical homogeneity*—is also prevalent in RFFA literature. It refers to similarity in the geophysical characteristics of catchments and the meteorological drivers such as rainfall and evapotranspiration. For, example, the extent of Region 1 in Iowa (Fig. 1) coincides with the Des Moines Lobe, a landform resulting from the last glacial episode. Similar examples are easy to find in other states' RFFA studies. A connection between physical homogeneity and statistical homogeneity is widely assumed and has been spelled out in the empirical region of influence (ROI) RFFA framework introduced by Burn (1990) and tested in Zrinji & Burn (1994) and Burn (1997). Burn (1997) used a dissimilarity measure introduced by (Webster & Burrough, (1972) and tested if the peak flow distributions of the catchments could be considered statistically homogeneous using the *H* metric introduced by Hosking & Wallis, (1993). The methodology requires "the choice of a threshold value that functions as a cut-off point for a dissimilarity measure. All sites that have a dissimilarity measure with the catchment of interest that is greater than the threshold value are excluded from the region of influence for that particular catchment" (Burn, 1997). Catchments that pass the test are expected to be statistically homogeneous and are considered appropriate for pooling data.

ROI methods have been refined to better characterize catchment dissimilarity and the definition of statistical homogeneity using L-moments. Hall & Minns, (1999) used non-linear clustering methods to define homogeneous regions using a variety of geomorphological descriptors. Viglione et al., (2007) investigated statistical homogeneity measures and showed that L-moments based tests are more powerful when the samples are slightly skewed, while rank tests have better performance in cases of high skewness. They also show that many of the indices used to determine statistical homogeneity lack the power to discriminate between homogeneous and heterogeneous regions. Recently, Ilorme & Griffis, (2013) introduced more sophisticated measures for physical homogeneity and statistical similarity *H*, while Basu & Srinivas, (2014) used non-linear kernels to divide the attribute space into groups that are not necessarily continuous in space. The attributes include physiographic, land use/land cover, drainage, climate and flood seasonality descriptors of the watersheds contributing to flow at the sites, and geographical location attributes. Underlying all of this work is the assumption that physical homogeneity implies statistical homogeneity. Even more recently (Zhang et al., 2020) proposed "a novel region revision procedure to complement the well-known region of influence and L-Moments techniques that automates the identification of homogeneous regions across continental domains" that could be used in the Canadian statistical flood estimation guideline under the FloodNet project (www.nsercfloodnet.ca).

3. Study Area and Data Sources

We select the Turkey River watershed (4,385 km^2), located in northeast Iowa in the midwestern United States, as a case study to address our objectives. Turkey River is located in Iowa's Region 2 according to the USGS RFFA report by Eash (2001) (Figure 2). This watershed is the same one used by Perez et al. (2019b) and it also has been widely used in hydrologic modeling frameworks to understand the control of different hydrologic process in the peak flow response. Examples include the evaluation of the importance of the level of storm spatial and temporal detail (Zhu et al., 2018) and climate-driven shifts in the seasonality of snowmelt and soil moisture (Yu et al., 2019) in peak flow distributions.

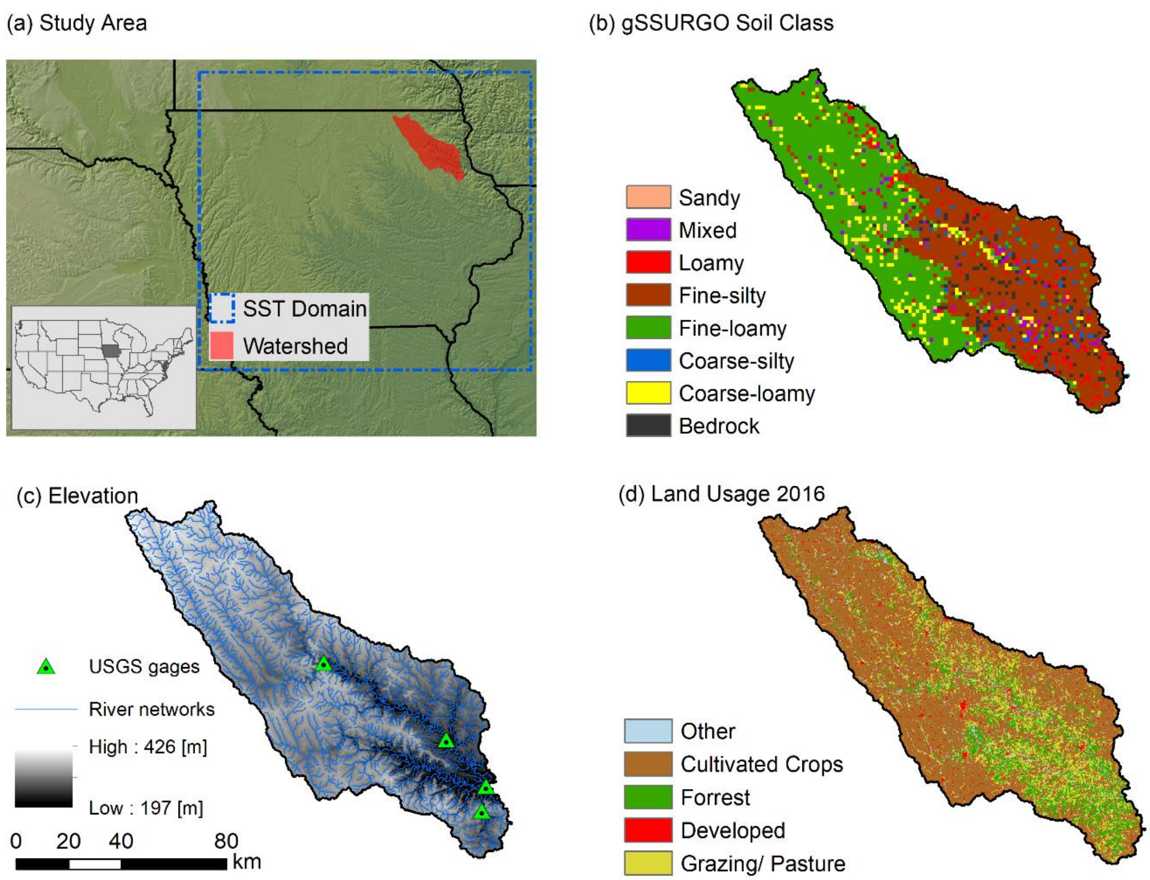


Figure 2. The Turkey River basin and the corresponding variability of soil types and landcover in the region.

4. Methodology

Our experimental design seeks to generalize the findings of Perez et al. (2019b) by addressing four RFFA-relevant questions about the distributions of peak flows along a river network:

1. Can the moments of the peak flow distribution for an arbitrary location in the river network be inferred using data from other sites within the network?
2. Are power-law equations sufficient to describe the scaling of peak flow distribution moments with respect to drainage area?

3. Is the coefficient of variation ($CV = \sigma/\mu$) or the skewness (γ) of peak flows constant across drainage scales and throughout a region under conditions of physical homogeneity in meteorology, runoff generation, and runoff transport mechanisms?
4. Are the IFM (simple scaling) or the quantile regression (multi-scaling) frameworks appropriate for describing flood data in a purportedly physically homogenous region?

To this end we used three different hydrological models to simulate peak flows across the river network associated to an equal number of storms derived using the SST framework described below. Our experiment is designed to ensure that rainfall distributions are homogeneous across the river network. To accomplish this, our experiment creates an unprecedented data set of 50,000 spatiotemporally realistic rainstorms, comprising 10,000 synthetic years that assume that 5 storms occur every year, which are used to simulate annual peak flows at many locations in the river network using each hydrological model. This is a five-fold increase over the number of storms and flood events considered in Perez et al. (2019b), and it was selected to ensure homogeneity in rainfall inputs and to minimize sampling uncertainty in order to approximate the population of the peak flows.

4.1. Stochastic Storm Transposition (SST) Framework

To address our research questions, we created 50,000 rainfall scenarios comprising 10,000 synthetic years. The needs of our study impose two basic requirements: *i*) individual rainstorms must have realistic spatiotemporal structure within rainstorms to ensure realistic multiscale depiction of flood events across the river network, and *ii*) each portion of the watershed must “see” approximately identical rainfall distributions so that sampling error associated with rainfall inputs is minimal. Creating such an event set is nontrivial—existing high-resolution precipitation datasets generally span only several decades and thus suffer from high sampling error (Wright, 2018), while most stochastic rainfall generation techniques either lack spatiotemporal detail or struggle to reproduce observed extreme rainfall rates (Furrer & Katz, 2008; Wright et al., 2020).

To address these issues, we couple the open-source SST software RainyDay (Wright et al., 2017) with 17 years (2002-2018) of Stage IV gage-corrected radar-based precipitation (Du, 2011; Lin, n.d.) to generate 50,000 storm scenarios for Turkey River. SST generates these scenarios via temporal resampling and spatial transposition of observed precipitation events from the surrounding transposition domain. Wright 2020 reviewed the development and applications of SST since its inception. SST, as implemented in RainyDay, ensures that the aforementioned first requirement (e.g. realistic intrastorm spatiotemporal structure), to the extent that the Stage IV observations reflect the true precipitation structure (resolution-related effects remain, e.g. Zhu et al., (2018). The second requirement is satisfied through the following steps:

1. The transposition domain (Figure 2a) refers to a region over which observed extreme precipitation should be statistically similar. While we do not formally verify this regional rainfall homogeneity here, it's existence or lack thereof will have no influence on the results of this study due to other aspects of the SST methodology. The transposition domain used here spanned 94.25-89.25°W and 40.5-45.5°N.

2. RainyDay used April-November Stage IV precipitation to create a "storm catalog" consisting of the 320 most intense precipitation events within the domain. "Most intense" is defined with respect to 72-h duration rainfall accumulations over areas the size and shape of the Turkey River watershed. This same duration, as well as similar a time period and storm catalog size, has been used in previous SST-based FFA studies in Turkey River (Perez et al., 2019b; Yu et al., 2019, 2020).

3. RainyDay randomly selected k storms from the catalog to represent a synthetic year. The number of storm arrivals k was calculated by RainyDay using the Poisson distribution with rate parameter λ, defined as the ratio of total number of events in the storm catalog to the number of years in the record (i.e., $\lambda = 320/17 = 18.82$ storms yr^{-1}).

4. RainyDay transposed the selected k storms randomly (with uniform probability) within the transposition domain and computed the precipitation field over the watershed, thus creating a synthetic year of precipitation scenarios for the study watershed. These RainyDay-based precipitation events have the identical spatial and temporal resolution and structure as the input Stage IV data (hourly, ~5 km).

RainyDay repeated steps 3-4 to create 10,000 synthetic years, each comprised of k precipitation events in Turkey River. Within each synthetic year, the largest five of the k 72-h rainfall events with respect to the size and shape of the watershed were retained for hydrologic simulations. Testing showed that five events per synthetic year (i.e. 50,000 storms in total) was sufficient to generate the simulated annual flood peaks across the entire range of subbasin scales represented (i.e. up to 4,385 km^2). This constitutes a methodological improvement over Perez et al (2019b), which retained only the largest single event per year, for a total of 10,000 events. Follow-up testing revealed that using only one event per year could lead to spatial biases in the resulting flood simulations, since the storms that produce annual flood peaks on small headwaters tributaries tend to have different characteristics from those that produce annual peaks along the river's mainstem.

4.2. Description of Three Hydrological Models

4.2.1. HLM

The Hillslope-Link Model (HLM) is a distributed hillslope-scale rainfall-runoff model that partitions the landscape into individual control volumes following the landscape decomposition into hillslope areas outlined in Mantilla & Gupta, (2005). The model is parsimonious, using ordinary differential equations to describe transport between adjacent control volumes (Krajewski et al., 2017; Mantilla, 2007; Mantilla et al., 2006). All flows in HLM are downstream, all the closure relations are unique storage-flow relations, and there are no feedback loops of surface and subsurface flows. In addition, the parameters describing vertical and horizontal flows in hillslopes are assumed to be the same throughout the watershed. HLM is the simplest and more spatially homogenous of the three models used.

4.2.2. WRF-Hydro

The physics-based distributed Weather Research and Forecasting Hydrological modeling system (Gochis et al., 2020) calculates energy and moisture fluxes over square grid cells using the Noah-MP land surface model (Niu et al., 2011). Noah-MP is coupled to modules for surface and subsurface routing over the terrain and through river channels. While WRF-Hydro is highly flexible, we use the same configuration as in National Water Model (NOA, 2016) which is used by National Oceanic and Atmospheric Administration (NOAA) for operational flood forecasting. One exception is that snow processes are not used in this study to render WRF-Hydro results more comparable to the other two models, which lack snow process representations. The model was calibrated manually with respect to long-term water balances, infiltration processes, and flood hydrographs. See (Yu et al., 2020) for more information on this WRF-Hydro configuration and calibration procedures. From here on, we will use the abbreviation “WRF” to refer to WRF-Hydro in figures and tables.

4.2.3. Watershed Modeling Framework (WMF)

The Watershed Modeling Framework (WMF) is a Python-Fortran programable hydrological modeling and analysis tool (Velásquez et al., 2020; Velez et al., 2018). The hydrological model of WMF is a modification of the hydrological TETIS model (Francés et al., 2007; Velez, 2001). WMF represents the hydrological processes using grid elements. Each element has five storages representing the runoff, canopy, subsurface, aquifer, and channel processes. The water moves vertically across storages and horizontally from an upstream element to a downstream element. The vertical movement is mainly governed by the evapotranspiration rate, the infiltration rate, and the percolation rate. The downstream movement uses the rill approximation to represent the runoff (Foster et al., 1984) and a non-linear equation to represent the subsurface (Kubota & Sivapalan., 1995). The channel routing follows the GKW (Geomorphological Kinematic Wave) approach developed by Velez (2001). We use the information of POLARIS (Chaney et al., 2016) to distribute the parameters of the soil at a resampled resolution of 150m.

Table 1. Comparison of main attributes for the three hydrological models used in our study.

	Equations Class	**Spatial resolution**	**Rainfall-Runoff**	**Subsurface**	**River Network Transport**
HLM	ODE	Hillslopes ~ 0.018 km^2	Top-layer	Linear Reservoir	Non-linear reservoirs
WRF-Hydro	ODE-PDE	Square Cells 1 km^2 (land column); ~ 0.063 km^2 (overland flow)	Noah-MP	Vertical Richard’s Equation (upper); 2D-Darcy’s Flow (lower)	Muskingum-Cunge
WMF	DE	Square Cells ~ 0.022 km^2	Phi-index + exfiltration	Kubota & Sivpalan (1995)	Kinematic Wave

4.3. Process-based framework for flood frequency analysis

Continuous simulations were conducted with all three models for the period 2002-2018 to (*i*) determine their ability to simulate streamflow variability at five gauged locations in Turkey River, and (*ii*) to create a daily-scale “database” of initial conditions to be used to initialize the models for each of the 50,000 flood events. These continuous simulations, as well as the SST framework described in Section 4.1, were forced using Stage IV precipitation and other meteorological forcings (which varied by model) from NLDAS-2 (Mitchell et al., 2004).

The process-based flood frequency simulation framework is the same as described in Yu et al. (2019) and Yu et al. (2020); it is only briefly described here. The initial condition for each event was randomly selected amongst daily model snapshots. The selection of the date corresponding to the initial condition was preserved for the three models (i.e. all three models start each event with similar initial conditions). The goal of this process was to have all the models on a similar footing at the beginning of each rainfall-runoff event to minimize the effects of differences in initial conditions in our analysis.

4.3 Evaluation of physical and statistical homogeneity

4.3.1. Meteorological homogeneity and realism

Meteorological homogeneity, or more specifically, homogeneity in the probability distributions of extreme rainfall across a range of durations, is a primary focus of regional rainfall frequency analysis (Svensson & Jones, 2010), a branch of research and practice that shares a common lineage and methods with RFFA. Meteorological homogeneity is also a prerequisite for physical homogeneity in runoff production. The SST methodology (Sec. 4.1) was deliberately conducted to ensure that the populations of rainfall events that are experienced by each computational unit (e.g. grid cell or hillslope) in the hydrologic models are effectively identical. Furthermore, by transposing observed storms, SST preserves the spatial and temporal structure of observed rainfall, meaning that homogeneity will be consistent across scales. It should be noted that the relevance of existing metrics and tests for evaluating rainfall homogeneity (e.g. Hosking & Wallis, 1993) is questionable, since such tests were developed under the presumption of limited sample sizes, as opposed to the very large samples used in this study.

In addition to homogeneity, the population of rainfall events used in our hydrological simulations must be relatively realistic in terms of frequency and magnitude. Otherwise, conclusions arising from the simulations would have limited relevance to real-world conditions. This realism is evaluated by creating rainfall intensity-duration-frequency (IDF) curves at the pixel scale, following Wright et al. (2017). These SST-based IDF curves are then compared against published rain gage-based IDF estimates from NOAA's Atlas 14 project (Bonnin et al., 2006).

4.3.2. Physical homogeneity of runoff generation

Physical homogeneity of runoff generation is typically approached from the point of view of the spatial variability of physical properties associated to runoff production (e.g. hydraulic conductivities, terrain roughness, land use, land cover, and soil types) and their representation through relevant model parameters. This can be misleading, however, because it is difficult to determine what is an acceptable range of variability in parameter values that can influence the behavior of nonlinear equations that can lead to large differences in runoff production in the models. Instead of examining homogeneity in the context of model parameters, therefore, we will examin it in the context of the event-based runoff coefficients *RC*, defined by the ratio between total event rainfall and total surface runoff during the event $RC = \sum q_s / \sum P$. The runoff coefficient will vary from event to event and amongst model control volumes (e.g. hillslope to hillslope or grid cell to grid cell). While it will depend primarily on the total amount of rainfall and on antecedent soil moisture conditions, other factors play a role including the rainfall variability during the event, differences in slope, infiltration capacities, soil depth, and porosity.

Recently Jadidoleslam et al. (2019) introduced the deficit-based rainfall metric $P_\theta = \sum P / \left(D_s(\phi - \theta_o)\right)$ to categorize runoff coefficients observed in catchments in Iowa, where $\sum P$ is the total event precipitation, ϕ is the porosity, and θ_o is the antecedent soil moisture, and D_s is the integration depth of soil moisture values. They found that the index collapsed the variability of runoff coefficients significantly, providing predictability for the runoff coefficient with spearman correlation coefficients of 0.6. This metric is ideal to contrast runoff production from event to event and from one control volume to another, but it also gives us an objective comparison between our three different models. Note that in the HLM, parameters related to the rainfall-runoff transformation are constant in space, though a dependency on hillslope length introduces spatial variability. In WRF-Hydro, the control volumes are square grid cells, but there is significant variability in parameter values specified from land use and soil maps. In WMF, the two sources of variability are combined, because the control volumes are irregular in size and have variable parameters

4.3.3. Statistical Homogeneity of Peak Flow Distributions

After establishing the range of homogeneity (or lack thereof) in runoff processes, we can establish the level of statistical homogeneity amongst peak flow distributions for locations along the river network. Peak flows are a scale-dependent random variable, which makes establishing homogeneity more complicated than rainfall and runoff. Here, we use the spatial variability of the coefficient of variation and skewness to determine if the distributions can be rescaled using this information alone. As in the case of rainfall, the 50,000 events are treated as 10,000 synthetic years and all peak flow statistics are calculated using the most extreme 10,000 events for each channel link in the river network.

5. Results

5.1. Model Performance Evaluation

The three models were set up for Turkey River. Parameters in WRF-Hydro and WMF were determined using data for the period 2002 to 2018 via calibration, while HLM was parameterized using values from the operational IFC model for the Iowa domain (Quintero et al., 2016). We determine the ability of each of the models to capture peak flows for a total of 782 rainfall runoff events. Simulated peak flows were compared against observations, as were observed and simulated hydrographs at five gauged locations within the watershed (Figure 3). While bias and random error are evident in all three models, no single model is clearly superior and all three models exhibit reasonable performance at capturing both peak flows and other streamflow dynamics.

It is hard to determine if any model is superior. To evaluate the models, we compute the peak flows correlation coefficient (R^2), and the streamflow KGE (Kling Gupta Efficiency) index (H. V. Gupta et al., 2009) and Pbias (volume bias). R^2 represents the simulated peak flow accuracy. The KGE summarizes the correlation, the mean bias, and the deviation bias to determine the performance, and the Pbias represents the total percentage at the simulated volume. WMF shows the higher peak flow correlation coefficient with an R^2 equals to 0.87 (Figure 3c). Followed by HLM with an R^2 equal to 0.85 and WRF with an R^2 of 0.72. According to Figure 3 d to h, there is no model with a dominant KGE or Pbias. Regardless of the observed performances, we accept the three models for the purpose of the current work.

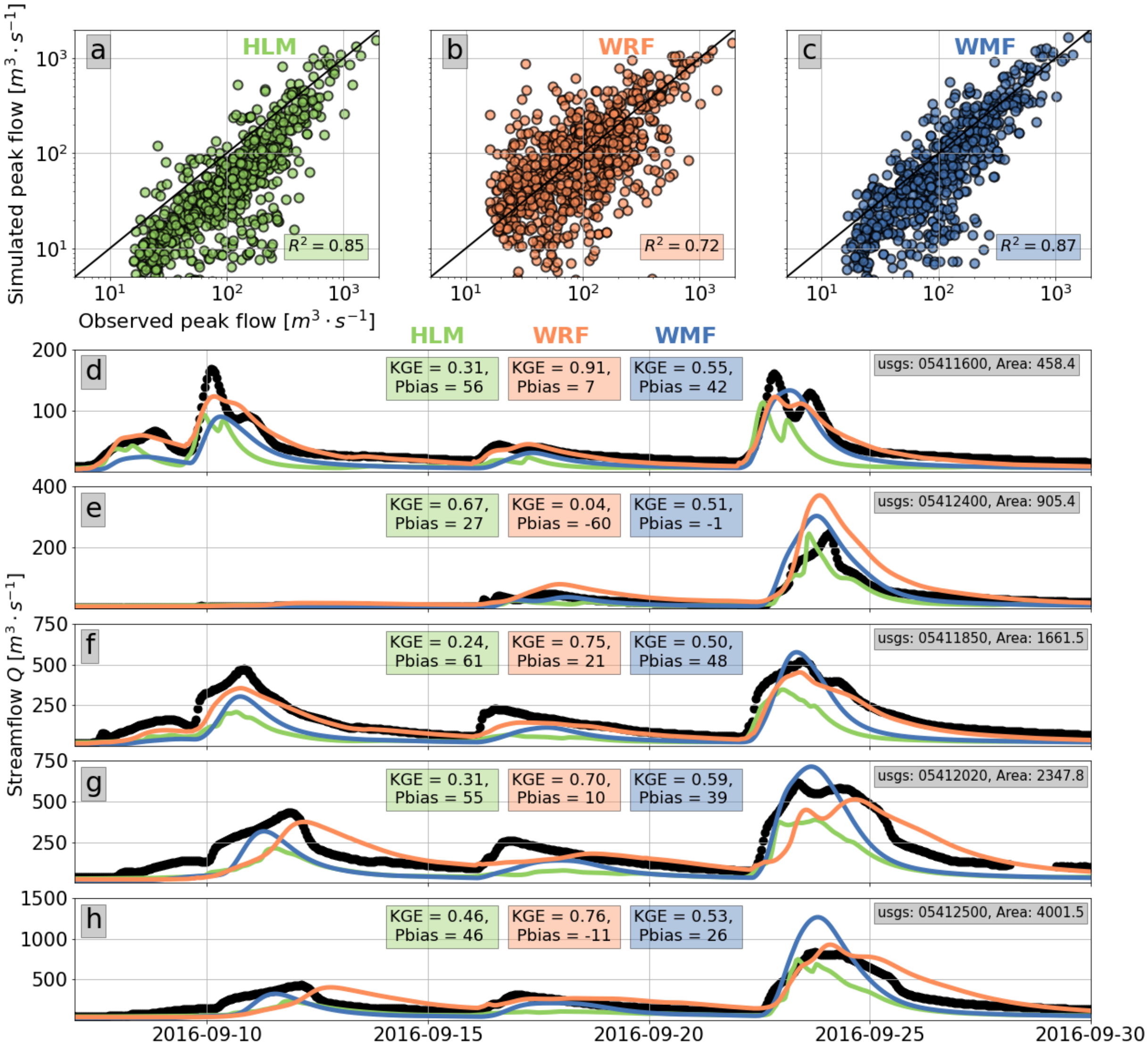


Figure 3. A comparison of model performance to simulate peak flows and hydrographs. Panels a,b, and c show the simulated vs the observed peakflows for the HLM, WRF, and WMF respectively. Panels d to h show the observed streamflow (black) and the simulated streamflow for the HLM (green), WRF (orange), and WMF (blue) models at five USGS gauges.

5.1 Realism of Rainfall Intensity Duration Frequency (IDF) Curves

Rainfall intensity-duration-frequency (IDF) curves at the pixel scale, following Wright et al. (2017) are calculated for the Turkey river domain. In the Figure 4, we compare the SST-based IDF curves against published rain gage-based IDF estimates from NOAA's Atlas 14 project (Bonnin et al., 2006).

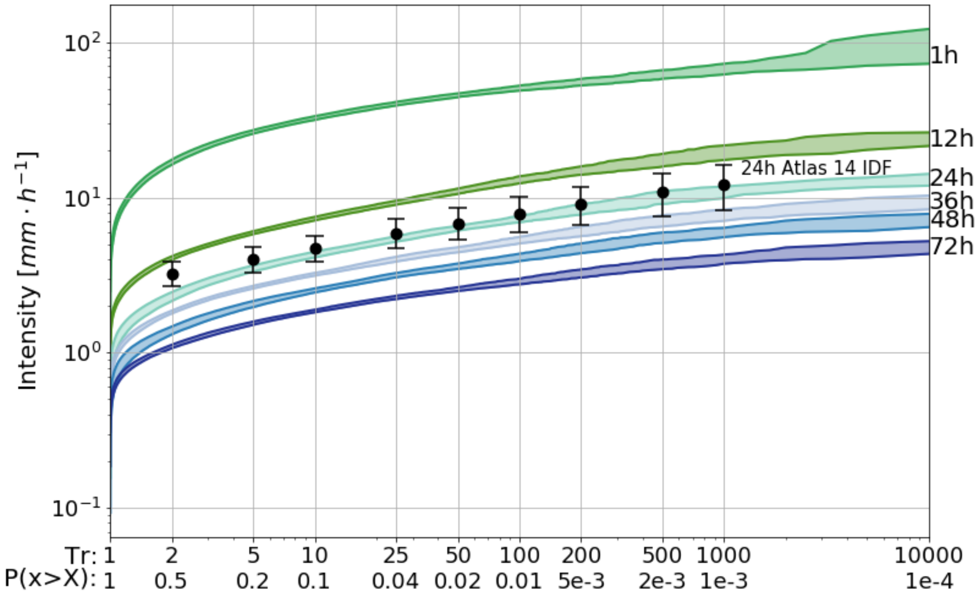


Figure 4. SST Intensity Duration Frequency (IDF) curves for accumulations ranging between 1h and 72h. And, Atlas-14 IDF curves for accumulations of 24h.

5.2 Physical Homogeneity of Runoff Generation

Violin plots show the variability in simulated flood event runoff coefficients of individual control volumes (i.e. hillslopes or grid cells) as a function of P_θ (Figure 5). The variability of RC vs. P_θ is analogous to the concept of curve number, as it represents the expected runoff coefficient for a given precipitation total and the antecedent soil moisture, and it represents the prediction of each one of the models for the statistics of runoff regeneration in the region. In the case of the HLM model (green violins) we observe the largest values of runoff coefficient as the model structure assumes that the majority of quick runoff entering the channels is surface runoff. The extent of the variability of RCs for different values of P_θ is small, as expected, because all the hillslopes in the domain share the same values. In the case of RCs predicted by WRF-Hydro we see the smallest values of runoff coefficient, which indicates that a significant fraction of the quick runoff into the channels is provided by subsurface flows. Finally, the WMF model predicts an intermedia prediction for surface runoff, but it exabits the largest variability or RCs. These results reveal that even in landscapes that appear to have very similar land cover, land use, and soil types, such as those found in the Midwest, significant spatial and temporal heterogeneity can be introduced by the physical processes that control runoff production.

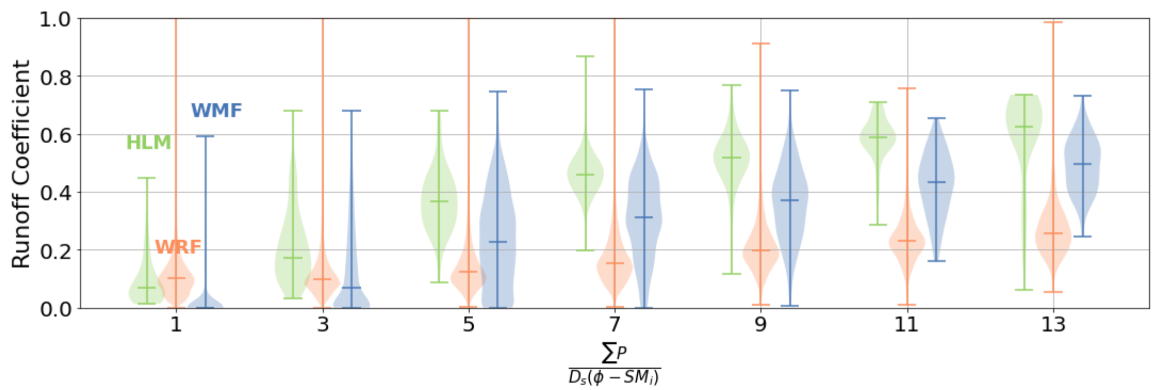


Figure 5. Runoff ratios calculated using three hydrological models as a function of deficit-based precipitation for 50,000 simulated rainfall runoff events.

5.2. Statistical Homogeneity of Peak Flow Distributions

Scaling plots show the mean μ, standard deviation σ, coefficient of variation $CV = \sigma/\mu$, skewness γ, and kurtosis κ with respect to drainage area for the three models' simulated peak flow distributions using the 10,000 peak flow samples at each location (Figure 6). Recall from Sec. 2.1 that the homogeneity requirement for the index flood method (i.e. simple scaling) is for CV to be constant across scales and for every location in the region. For the three models this assumption is clearly violated for the population of peak flows (Figure 6g-i). It could be argued that the variations are within acceptable margins of error for any of the given empirical RFFA techniques; however, as is shown by Perez et al (2019b), an incorrect representation of the skewness in the estimation of the index flood method will add systematic bias to resulting peak flow estimates that changes across scales. In addition, the index flood method requires that the index flood be predicted using a power law regression equation with respect to drainage area. Residuals of power law fits (supplemental Fig. S1) show the existence of one or more scale breaks between 100 and 1000 km^2, depending on the model.

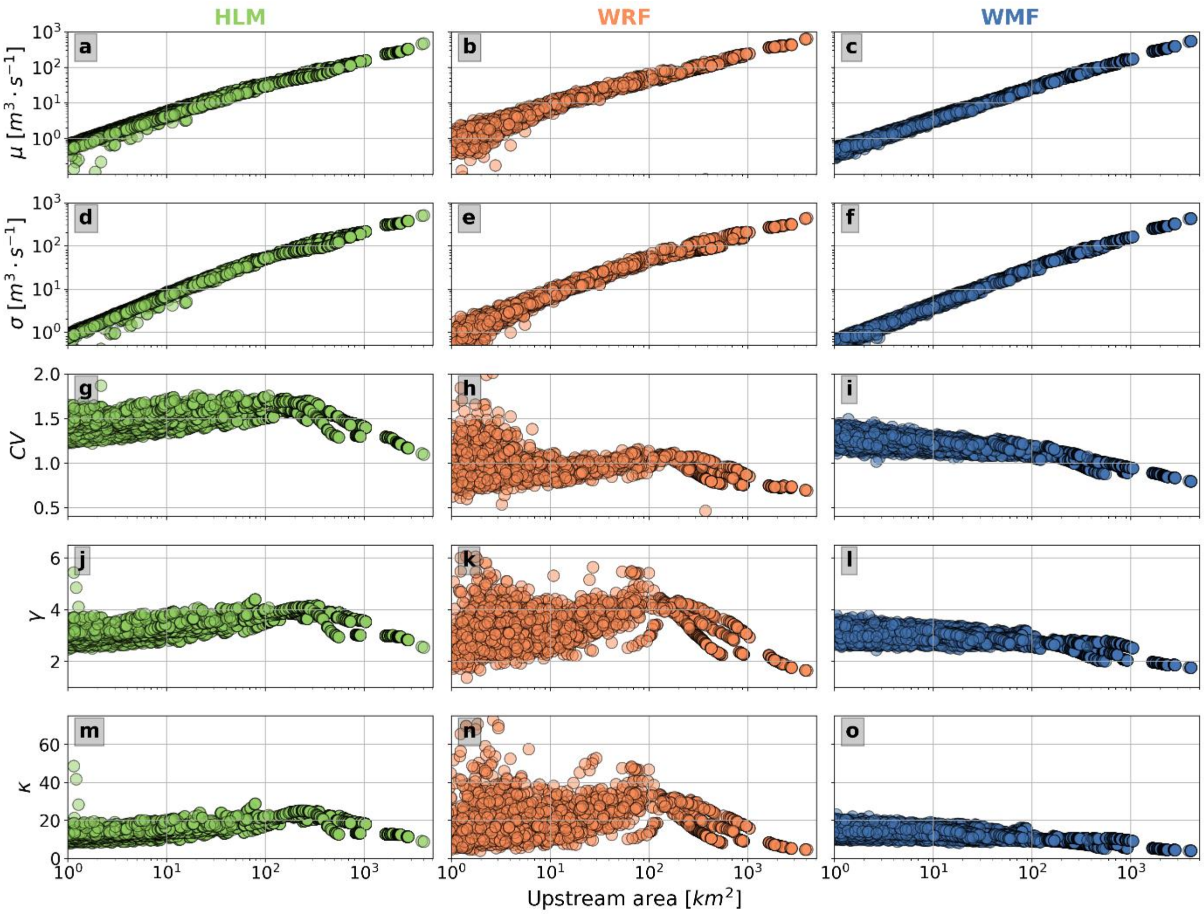


Figure 6. Scaling plots for the mean μ, standard deviation σ, coefficient of variation $CV = \sigma/\mu$, skewness γ, and kurtosis κ of peak flow distributions with respect to upstream drainage area A, from the three hydrological models considered in this study.

The multi-scaling framework (Sec. 2.1) relaxes the requirement that CV is constant across scales, but it requires that all distribution moments scale as a power-law of drainage area. While this appears reasonable for the mean and standard deviation, all three models show significant variability in skewness that is not explained by area alone (Figure 6j-l).

The spatial distribution of skewness in the drainage network offers an additional view into the limits of the concept of a regional skew value and its existence (Figure 7). The maps of skewness for the three models reveal systematic variations of skew across the drainage network, which are undoubtedly related at least in part to the interaction of storm patterns with network geometry, with parameter variability related to land use and soil properties likely playing roles in WRF-Hydro and WMF results. The map generated by WRF-Hydro exhibits the most unstructured pattern of skewness in space and the widest range of variation of its value.

It is important to note that the spatial variability of skewness shown in Figure 7 is substantial, on the order of magnitude of the skewness values across Iowa in the most recent USGS statewide flood frequency study (Eash et al., 2013), and as large as those found for the entire Midwest region of the United States in Bulletin 17B (IACWD 1982) (see supplemental Fig. S2).

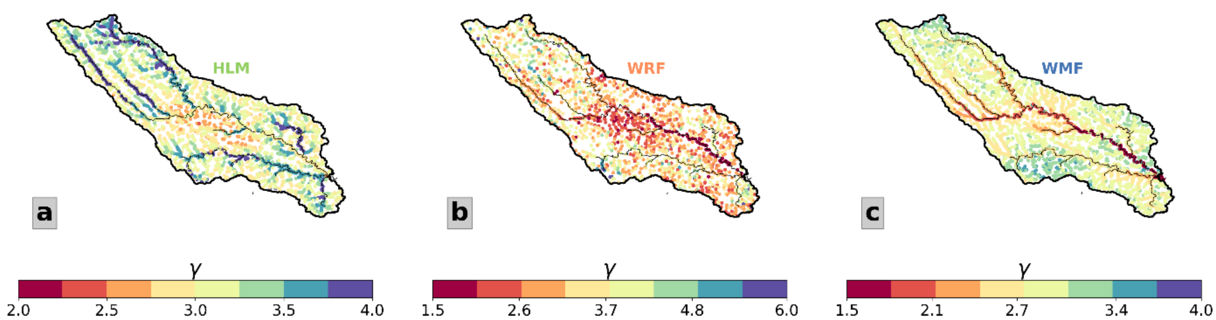


Figure 7. Spatial distribution of skewness γ of the peak flow distribution with respect to upstream drainage area A from the three hydrological models considered in this study.

5.3 Physics-based Predictions in the Context of Observed Flood Data

Each of the three models provides a deterministic representation of hydrologic response to a given rainfall input, and therefore a unique prediction of the peak flow distribution for every computational node in the river network. These distributions and particular quantiles from them can be compared against observations. In Figure 8, we focus on simulated 100-year peak flows and compare them against USGS estimated 100-year events (Eash et al. 2001) at gauged locations in Iowa's Region 2 (see Figure 1). In addition to the population estimate using 10,000 annual flood peaks, Figure 8 also shows the values that would be estimated if we used smaller samples of 100, and 1,000 annual peaks. Variability in these estimates provides a depiction of plausible deviations from true population values due to sample size.

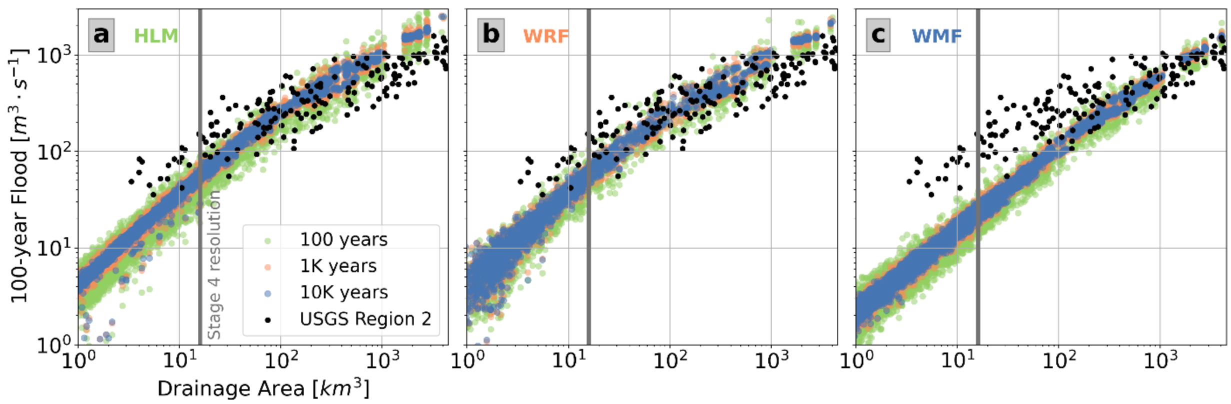


Figure 8. Predictions of the 100-year peak low in Turkey River. Colored dots represent simulated peak flows quantiles, with different colors highlighting how uncertainty varies with sample size. Black dots show the 100-year peak flow estimates (Eash et al. 2001) using observations in Region 2 in Iowa (see Fig.1 for map of RFFA regions). Blue vertical lines denote the 16 km^2 spatial resolution of Stage IV precipitation used in this study. Predictions below that scale are questionable, and rainfall resolution effects may explain the apparent bias relative to observations at small drainage scales. The effect of resolution scale is less relevant for larger scales (e.g. Cunha et al., 2012; Mandapaka, P. V. et al., 2009).

The simulated results in Figure 8 show predictions that emerge from the assumptions that are built into model equations and parameterizations as well as the rainfall fields generated through SST. This result is in high contrast to the purely empirical RFFA approach of fitting power law regressions and using these regressions to estimate values at ungauged sites. The physically based estimated frequencies are calculated for each channel in the river network, and they include the available information about the hydrology of the region that is encoded by each one of the hydrologic models.

The magnitude of variability of the 100-year flood estimate as a function of sample size is a consequence of the rate of convergence of the moments of the peak flow distribution. In Figure 9 we illustrate this issue by showing the results of a bootstrap analysis of estimation for the mean, standard deviation, skewness and kurtosis, as a function of sample size at three different locations in the basin that drain different watershed areas.

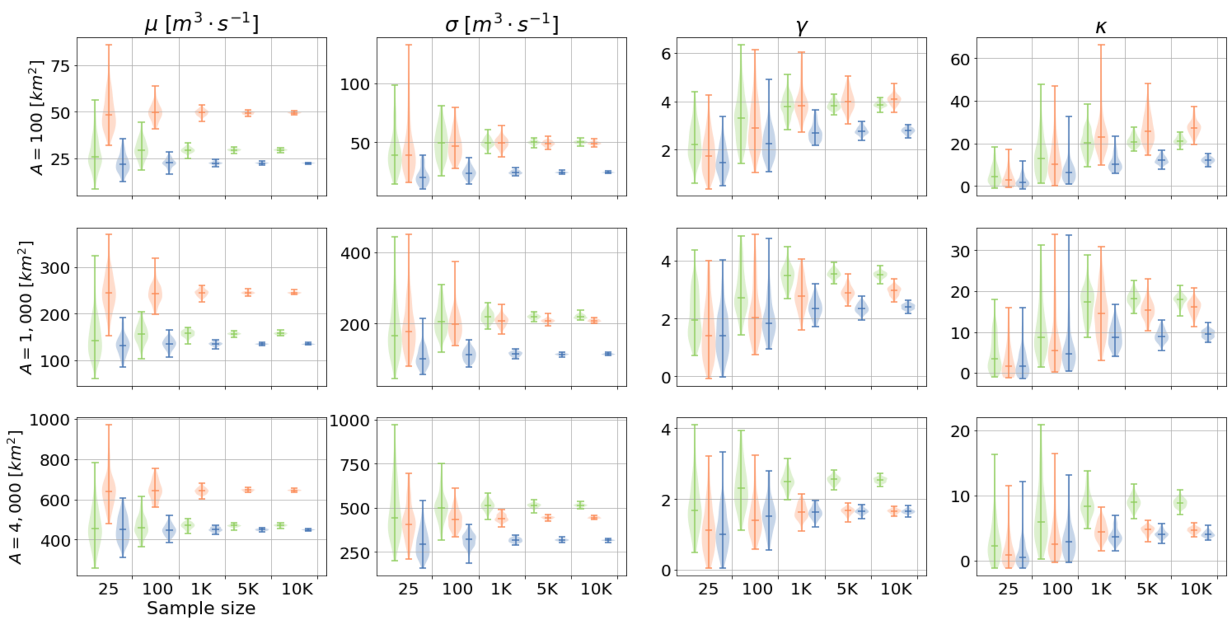


Figure 9. Estimated peak flow distribution moments (mean μ, standard deviation, σ, skewness, γ, and kurtosis, κ) at three different catchment scales for varying sample sizes.

Figure 9 shows that peak flow distribution moments can be accurately estimated for sample sizes larger than 1,000. In the case of small sample sizes, the estimation error can be larger than 50% for all the moments. In the case of the mean and standard deviation, the small sample estimates seem to be unbiased, however, for the skewness and kurtosis, it can be seen that the estimates using small sample sizes are biased toward low values. The rates of convergence appear to be scale independent. The need for a sample size of 1,000 peak flows to estimate population moments offers a glaring contrast with the length of datasets that are typically used in RFFA that typically range between 30 to 100 years.

6 Discussion

The three hydrological models represent different approximations of the true hydrological cycle in the Turkey River study watershed. There is no doubt that the actual flow paths are more complex than those represented by any of the models. The models appear to capture key aspects of hillslope and channel processes in the basin (Figure 3), suggesting that the simulated variability of peak flows depict a plausible representation of reality. Furthermore, the three models provide estimates for the 100-year flood for a catchment in Region 2 of Iowa that are relatively consistent (i.e. within an order of magnitude) of estimates made using flood peak observations (Figure 8). We acknowledge that this simulated variability has its limits, however. While all three models offer a variety of optional process configurations and virtually infinite parameter spaces, only a single configuration and parameter set were used for each model in this study. Furthermore, the SST methodology's reliance on transposing observed storms limits the variability that can arise from precipitation inputs (Wright et al., 2020). Finally, we totally neglect snowmelt and rain-on-snow flooding, which can be relevant for the study region. Thus, it is reasonable to expect that real peak flows would exhibit more, not less, variability in space and time than the variability modeled here. While it is thus difficult to make definitive statements based on our results, but they serve as a useful starting point to discuss the magnitude of peak flow variability in space and across scales. In addition, some of the limitations of our study (e.g. parameter equifinality, snowmelt processes) are well within the capabilities of many contemporary distributed modeling frameworks.

Our findings are consistent with and help to generalize the results of Perez et al., (2019b). Namely, simulated peak flow quantiles exhibit a level of variability that cannot be simplified into a simple- or multi-scaling framework, indicating that the underlying assumptions in widely-used RFFA methods such as IFM and quantile regressions will introduce estimation biases. Simple scaling is rejected by the spatial variability of the coefficient of variation, while multi-scaling is rejected by the variability of skewness in space and across scales (Figure 7Figure 8, and S2). Furthermore, scaling of the mean annual flow and standard deviation with respect to drainage area does not collapse to a simple power law across scales (Figure 9). In fact, we observe a scaling break in all three models ranging from 100-1,000 km^2, possibly tied to a scale at which river routing and network topology effects start exerting control over peak flow variability. Further assessment of the errors introduced by assuming simple- or multi-scaling of peak flow data is outside the scope of our work and is instead left as an important future research task.

In general, the simulated peak flows reject the premise that data from neighboring sites provide information about the specific attributes of the peak flow distribution at an ungauged site. We found strong patterns of variability in skewness (Figure 7Figure 8) that would introduce biases when interpolating or extrapolating from one site to another. Furthermore, the degree of variability in skewness within the 4,385 km^2 study watershed is comparable to the variability of published estimates of flood peak skewness for Iowa (Eash et al. 2013) and for the entire midwestern US (IACWD, 1982) shown in supplemental Fig. S2.

The spatial structure of skewness predicted by the three models is remarkable in several ways. First, the differences in values estimated by the three models suggests that skewness is very sensitive to the underlying local physical processes controlling the magnitude of flood peaks. We attempted several analyses to identify the variables involved in explaining the variability of skewness, but no simple predictive equation has emerged (results not shown). It would thus appear that every aspect of the physics of peak flow generation matters when it comes to determining skewness. Second, the value of local skewness differ from the average value of skewness in the catchment across all scales, indicating that the distribution of peak flows is sensitive to both small-scale (e.g. infiltration dynamics; Figure 7) and large-scale (e.g. river network topology; see Mantilla (2010) for a theoretical context on the role of river networks) sources of variability in runoff production and transport. Third, variability in skewness implies that the notion of a single regional skew value, which is central to RFFA practice in the United States and also implicit in other empirical RFFA methodologies, is an incorrect assumption. It is difficult to pinpoint what aspects of each model lead to a particular value of skewness; variability arises not only from parametric differences but also dynamics that include non-linearities and threshold values and the spatial and temporal variability of the rainfall fields.

Another important insight from our simulation exercise is just how slowly the peak flow sample moments converge to the "population" moments (Figure 9). This implies that sample size dominates the uncertainty in estimates of flood quantiles, even using regionalization techniques, confirming Perez et al. (2019b). Our estimates of the 100-year flood using 100 synthetic years in Fig. 10 can easily double or halve the "true" population value, and it takes roughly 1,000 synthetic years of simulated data to reliably obtain estimates that are within 20% of the population values. The consistency in this finding across the three models suggests that this feature is a property of the peak flow random variable, which is greatly influenced by both initial states (e.g. soil moisture, storage in channels) and the spatiotemporal structure of rainfall (Zhu et al., 2018). These results should raise concerns about both data-driven studies that try to estimate nonstationary trends in peak flows over short timespans and studies that aim to project changes in flood regimes under different climate change scenarios.

A frequent criticism of the usage of hydrologic models, including physics-based ones, in flood frequency studies is the perception that model uncertainties are prohibitively high. As argued in Wright et al., (2020), such criticisms generally presume the primacy of data while glossing over the numerous and poorly-understood uncertainties in statistically-based RFFA. We agree wholeheartedly with the critical role of data, but argue that statistically-based RFFA makes poor use of the wealth of relevant datasets currently available—including distributed high-resolution rainfall, soils, land cover, and hydrography, not to mention the vast majority of streamflow observations that do not rise to the level of annual maxima. Physically-based approaches such as the one used here not only leverages these rich resources, but also our understanding (which is admittedly imperfect, and further simplified within models) of the movement of water through the landscape. This makes the estimates made through our process more robust and less susceptible to outliers in the data.

7 Conclusions

Three hydrological models and stochastic storm transposition (SST) were used to investigate the validity of implicit assumptions in the empirical methodology of regional flood frequency analysis (RFFA) for prediction of flood quantiles at ungauged locations. This combined physics-based framework allows us to estimate peak flow frequencies throughout the river network by calculating thousands of realistic dynamic flooding scenarios. These scenarios give us an unprecedent look into the nature of the "annual peak flow random variable," which is at the heart of the empirical RFFA. The framework allowed us to investigate the four questions, reprinted from Section 2.3: (*i*) what is the link between physical homogeneity and statistical homogeneity among locations within a watershed or region?, (*ii*) what are appropriate simplifications for the scaling of the statistical moments of peak flow distribution across scales?, (*iii*) can the peak flow distribution parameters at an ungauged site be inferred from measurements at neighboring sites?, and (*iv*) what are the necessary sample sizes to characterize peak flow distributions.

Our results do not support the link between *physical homogeneity*, defined as the spatial variability in the physics of runoff generation and runoff routing, and *statistical homogeneity*, defined as the existence of a common underlying statistical distribution of peak flows and the possibility of inferring the distribution parameters using regional observations of peak flows. However, our results cannot ultimately reject the link, instead they show that more general and restrictive definitions of *homogeneity* are needed. Specifically, our simulated peak flow results show that the local value of the third moment in the population distribution (skewness) cannot be inferred from regional values, and instead depends on very specific local conditions including but not limited to land use and soil distribution and their effect on runoff production, river network topology and geometry, and local hydraulic geometry. While not explicitly evaluated here, basin shape and orientation relative to prevailing storm motions and space-time structures likely to exert important controls on skewness; this was partially addressed by Perez et al. (2019b), but it remains an open problem.

The spatial variability of skewness predicted by the three hydrologic models indicates that simple-scaling (i.e. the index flood method, generally used in Europe and Canada) and multi-scaling (e.g. quantile regressions used in the United States) frameworks provide at best first-order approximations of peak flow variability at ungauged locations. Those scaling frameworks require that the moments of a peak flow distribution be inferred from basin area alone through power laws or other deterministic functions. Instead, our results suggest that well-validated distributed hydrological models may be needed to estimate local moments for peak flow distributions at ungauged locations, due to the complexity of rainfall-runoff transformations and river routing. In addition, the methodology presented here can be used to quantify the magnitude of sampling error involved in the inference of particular peak flow quantiles such as the 100-year flood.

A careful inspection of differences between the three models allow us to identify key features in the rainfall-runoff and runoff-transport processes that play important roles in determining the degree of spatial heterogeneity of peak flows distributions. In particular, in terms of sources of spatial variability of skewness, we can list four factors:

1. River Network Topology: the connectivity of side tributaries creates consistent spatial patterns in the variability of skewness along the river network. Although the three models used here employ somewhat different depictions of the river network (e.g. different drainage densities) in their formulations, they all show consistent patterns of variation along the major rivers in the watershed, although different amongst models, reflecting the complex aggregation and attenuation of flows that determine peak flow quantiles.
2. Routing: The routing schemes in our paper range from kinematic based ODE formulations to one dimensional PDEs, some of which include channel and terrain slope information. These different formulations give rise to different patterns of variability in skewness along a particular channel.
3. Infiltration Processes at the characteristic scale of the model: the three models use different discretization scales (e.g. grid structures) to describe physical processes. This translates into differences in sub-catchment scale responses and suggests that spatial variability in infiltration processes in the real world will leave a fingerprint on peak flow distributions.
4. Routing of Runoff into River Network Channels: akin to infiltration processes, the routing schemes for surface and subsurface runoff toward river channels differs among the three models. Those differences are reflected in the skewness maps, indicating that the routing of water over and through real-world hillslopes will have a signature in peak flow distributions across spatial scales within a catchment.

Which physical processes control the ultimate distribution of peak flows, and how they control them, remains to be revealed and quantified. The models used in this paper are gross simplifications of real-world physical processes, which are no doubt far more complicated. In addition, other flood generation mechanisms such as snowmelt would add a new dimension to the problem. There is little reason to believe, however, that more complex and realistic model physics would lead to smoother spatial fields of peak flow distribution moments. In addition, our storm resampling procedure in SST does not fully reflect the extent of the real rainfall spatial and temporal variability of a true population of storms. For example, a recent study shows that the orientation of the catchment in relation to the directionality of storms is a relevant control on flood quantiles (Perez et al., 2020). We have made an effort to provide all the data and models needed to replicate the results of this study in public repositories, however we are well aware that replicating a scientific study of this magnitude can be daunting (Hutton et al., 2016). However, we believe that our results are generic enough that they can be generated with any other high-resolution distributed hydrological model. The refinement and generalization of our ideas would require the engagement of other research groups interested in addressing these questions in other regions in the world.

Although SST is in its early stages and the physical models employed in this work have limited levels of sophistication, our results show the urgent need to transition from purely statistically-based empirical RFFA to approaches that are informed by the best available local characterization of the physical processes that determine peak flows.

The results presented in this paper lead us to argue that the tools and theoretical underpinnings to establish a physics-based RFFA framework are within reach. We invite the hydrological community and the relevant agencies to test our conclusions using their own hydrological models.

If our findings hold under further scrutiny, we urge reconsideration of the current path of empirical RFFA, which consists of periodic updating using recent observations, along with incremental refinement of statistical methods. Our results show that the variability of peak flows cannot be captured with the relatively short records in the data sets available today (generally less than 100 years). Instead, it appears that about 1,000 independent peak flows at multiple sites would be needed to approximate the true quantile values within acceptable margins of error (see Fig. 10). It seems difficult to believe that a new statistical technique, based on the same limited data, could overcome such a hurdle. On the other hand, the path anticipated by (Eagleson, 1972; V. K. Gupta et al., 2007; Klemeš, 1987) for a dynamical framework for prediction of flood quantiles based on our understanding of the physical processes that determine peak flow magnitude is within reach and it can provide a robust estimation framework for the spatial and temporal variability of peak flows frequencies. It is our hope that in the future, “Bulletin 17D” will mark a reinvention of flood frequency estimation that truly reflects the monumental advances in distributed hydrometeorological observations and modeling of the last several decades and that is supported by universally agreed scientific theory.

Acknowledgements: RM’s and NV’s contributions were supported by the Iowa Flood Center at the University of Iowa and by grants from the the Iowa Department of Transportation (Grant TR-699), and the Mid-American Transportation Center (MATC). GY’s and DW’s contributions were supported by the U.S. National Science Foundation Hydrologic Sciences Program (award number EAR-1749638). This University of Wisconsin-Madison (UW-Madison) team used computing resources and assistance from the UW-Madison Center for High Throughput Computing for WRF-Hydro simulations, and also thank WRF-Hydro team at the National Center for Atmospheric Research for their guidance. NLDAS-2 forcings are available at Goddard Earth Sciences Data and

Information Services Center (https://disc.gsfc.nasa.gov/datasets?keywords=NLDAS); Stage IV precipitation is available at https://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/. USGS streamflow data is available at https://maps.waterdata.usgs.gov/mapper/index.html. Our work builds on foundational work and on the exchange of ideas over the years with our colleagues Vijay K. Gupta, the late David Dawdy, Brent Troutman, Ed Waymire, and Witold Krajewski.

References

(IACWD), I. C. on W. D. (1982). *Guidelines for determining flood flow frequency: Bulletin 17B (revised and corrected)*. Hydrol. Subcomm.

Ayalew, T. B., Krajewski, W. F., & Mantilla, R. (2014). Connecting the power-law scaling structure of peak-discharges to spatially variable rainfall and catchment physical properties. *Advances in Water Resources*, *71*, 32–43. https://doi.org/10.1016/j.advwatres.2014.05.009

Basu, B., & Srinivas, V. V. (2014). Regional flood frequency analysis using kernel-based fuzzy clustering approach. *Water Resources Research*, *50*(4), 3295–3316. https://doi.org/https://doi.org/10.1002/2012WR012828

Beven, K. J. (2018). On hypothesis testing in hydrology: Why falsification of models is still a really good idea. *WIREs Water*, *5*(3), e1278. https://doi.org/https://doi.org/10.1002/wat2.1278

Bonnin, G. M., Martin, D., Lin, B., Parzybok, T., Yekta, M., & Riley, D. (2006). *NOAA Atlas 14 Precipitation-Frequency Atlas of the United States Volume 2 Version 3.0: Delaware, District of Columbia, Illinois, Indiana, Kentucky, Maryland, New Jersey, North Carolina, Ohio, Pennsylvania, South Carolina, Tennessee, Virginia, West Virgini*. *2*, 295. http://www.nws.noaa.gov/oh/hdsc/PF_documents/Atlas14_Volume2.pdf

Burn, D. H. (1990). Une évaluation de la méthode “d’influence régionale” pour l’analyse des fréquences de crues. *Hydrological Sciences Journal*, *35*(2), 149–165. https://doi.org/10.1080/02626669009492415

Burn, D. H. (1997). Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, *202*(1), 212–230. https://doi.org/https://doi.org/10.1016/S0022-1694(97)00068-1

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma*, *274*, 54–67. https://doi.org/https://doi.org/10.1016/j.geoderma.2016.03.025

Cunha, L. K., Mandapaka, P. V., Krajewski, W. F., Mantilla, R., & Bradley, A. a. (2012). Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model. *Water Resources Research*, *48*(10), n/a-n/a. https://doi.org/10.1029/2012WR012138

Dawdy, D. R., Griffis, V. W., & Gupta, V. K. (2012). Regional Flood-Frequency Analysis: How We Got Here and Where We Are Going. *Journal of Hydrologic Engineering*, *17*(9), 953–959. https://doi.org/10.1061/(asce)he.1943-5584.0000584

Du, J. (2011). *GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0. UCAR/NCAR - Earth Observing Laboratory*.

Eagleson, P. S. (1972). Dynamics of flood frequency. *Water Resources Research*, *8*(4), 878–898. https://doi.org/https://doi.org/10.1029/WR008i004p00878

Eash, D. A. (2001). Techniques for Estimating Flood-Frequency Discharges for Streams in Iowa. *United States Geological Survey, Water-Resources Investigations Report*, *4233*.

Eash, D. A., Barnes, K. K., & Veilleux, A. G. (2013). *Methods for Estimating Annual Exceedance-Probability Discharges for Streams in Iowa , Based on Data through Water Year 2010 Scientific Investigations Report 2013 – 5086*.

England, J.F., Jr., Cohn,T.A., Faber,B.A., Stedinger, J.R., Thomas, W.O., Jr., Veilleux, A.G., Kiang, J.E., and Mason, R.R., J. (2019). Guidelines for Determining Flood Flow Frequency Bulletin 17C Book 4, Hydrologic Analysis and Interpretation. In *Book:* (Issue May). https://pubs.usgs.gov/tm/04/b05/tm4b5.pdf

Francés, F., Vélez, J. I., & Vélez, J. J. (2007). Split-parameter structure for the automatic calibration of distributed hydrological models. *Journal of Hydrology*, *332*(1–2), 226–240. https://doi.org/10.1016/j.jhydrol.2006.06.032

Fuller, W. E. (1914). Flood flows. *ASCE Transactions1*, *77*, 567–617.

Furey, P. R., Troutman, B. M., Gupta, V. K., & Krajewski, W. F. (2016). Connecting Event-Based Scaling of Flood Peaks to Regional Flood Frequency Relationships. *Journal of Hydrologic Engineering*, *21*(10), 4016037. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001411

Furrer, E. M., & Katz, R. W. (2008). Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, *44*(12). https://doi.org/https://doi.org/10.1029/2008WR007316

Gochis, D. J., Barlage, M., Cabell, R., Casali, M., Dugger, A., FitzGerald, K., McAllister, M., McCreight, J., RafieeiNasab, A., Read, L., Sampson, K., Yates, D., & Zhang, Y. (2020). The

WRF-Hydro modeling system technical description, (Version 5.1.1). *NCAR Technical Note*, 107. https://ral.ucar.edu/sites/default/files/public/projects/wrf_hydro/technical-description-user-guide/wrf-hydro-v5.1.1-technical-description.pdf

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, V. K. (2004). Emergence of statistical scaling in floods on channel networks from complex runoff dynamics. *Chaos, Solitons & Fractals*, *19*(2), 357–365. https://doi.org/https://doi.org/10.1016/S0960-0779(03)00048-1

Gupta, V. K., Castro, S. L., & Over, T. M. (1996). On scaling exponents of spatial peak flows from rainfall and river network geometry. *Journal of Hydrology*, *187*(1–2), 81–104. https://doi.org/10.1016/S0022-1694(96)03088-0

Gupta, V. K., Mantilla, R., Troutman, B. M., Dawdy, D., & Krajewski, W. F. (2010). Generalizing a nonlinear geophysical flood theory to medium-sized river networks. *Geophysical Research Letters*, *37*(11), 1–6. https://doi.org/10.1029/2009GL041540

Gupta, V. K., Mesa, O. J., & Dawdy, D. R. (1994). Multiscaling theory of flood peaks: Regional quantile analysis. *Water Resources Research*, *30*(12), 3405–3421. https://doi.org/10.1029/94WR01791

Gupta, V. K., Troutman, B. M., & Dawdy, D. R. (2007). Towards a Nonlinear Geophysical Theory of Floods in River Networks: An Overview of 20 Years of Progress. *Nonlinear Dynamics in Geosciences*, 121–151.

Hall, M. J., & Minns, A. W. (1999). The classification of hydrologically homogeneous regions. *Hydrological Sciences Journal*, *44*(5), 693–704. https://doi.org/10.1080/02626669909492268

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resource Research*, *29*(92).

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science? *Water Resources Research*, *52*(10), 7548–7555. https://doi.org/https://doi.org/10.1002/2016WR019285

Ilorme, F., & Griffis, V. W. (2013). A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *Journal of Hydrology*, *492*, 151–162. https://doi.org/10.1016/j.jhydrol.2013.03.045

Jadidoleslam, N., Mantilla, R., Krajewski, W. F., & Goska, R. (2019). Investigating the role of antecedent SMAP satellite soil moisture, radar rainfall and MODIS vegetation on runoff production in an agricultural region. *Journal of Hydrology*, *579*, 124210. https://doi.org/https://doi.org/10.1016/j.jhydrol.2019.124210

Kinnison, H. B., & Colby, B. R. (1945). Flood formulas based on drainage-basin characteristics. *ASCE Transactions*, *110*, 849–904.

Kjeldsen, T. (2011). COST Action ES0901: European procedures for flood frequency estimation (FloodFreq) [Keynote]. *Geophysical Research Abstracts*, *13*, 10327–10327.

Klemeš, V. (1987). Hydrological and Engineering Relevance of Flood Frequency Analysis. In V. P. Singh (Ed.), *Hydrologic Frequency Modeling*. Reider Publishing Company.

Krajewski, W. F., Ceynar, D., Demir, I., Goska, R., Kruger, A., Langel, C., Mantillla, R., Niemeier, J., Quintero, F., Seo, B. C., Smallll, S. J., Weber, L. J., & Young, N. C. (2017). Real-time flood forecasting and information system for the state of Iowa. *Bulletin of the American Meteorological Society*, *98*(3), 539–554. https://doi.org/10.1175/BAMS-D-15-00243.1

Kubota, J., & Sivapalan, M. (1995). *Towards a Catchment-Scale Model of Subsurface Small-Scale Process-Based Modelling and Runoff Generation Based on Synthesis of Field Studies*. *9*(November 1994), 541–554.

Lin, Y. (n.d.). *GCIP/EOP surface: Precipitation NCEP/EMC 4KM gridded data (GRIB) stage IV data. UCAR/NCAR-Earth Observing Laboratory.* Retrieved January 20, 2021, from https://doi.org/10.5065/D6PG1QDD

Mandapaka, P. V., Krajewski, W. F., Mantilla, R., & Gupta, V. K. (2009). Dissecting the effect of rainfall variability on the statistical structure of peak flows. *Advances in Water Resources*, *32*(10), 1508–1525.

Mantilla, R. (2007). Physical Basis of Statistical Scaling in Peak Flows and Stream Flow Hydrographs for Topologic and Spatially Embedded Random Self-similar Channel Networks. In *University of Colorado*.

Mantilla, R., & Gupta, V. K. (2005). A GIS numerical framework to study the process basis of scaling statistics in river networks. *IEEE Geoscience and Remote Sensing Letters*, *2*(4), 404–408. https://doi.org/10.1109/LGRS.2005.853571

Mantilla, R., Gupta, V. K., & J. Mesa, O. (2006). Role of coupled flow dynamics and real network structures on Hortonian scaling of peak flows. *Journal of Hydrology*, *322*(1), 155–167.

https://doi.org/https://doi.org/10.1016/j.jhydrol.2005.03.022

Meigh, J. R., Farquharson, F. A. K., & Sutcliffe, J. V. (1997). Comparaison à l'échelle mondiale entre les méthodes régionales d'estimation de crue et le climat. *Hydrological Sciences Journal*, *42*(2), 225–244. https://doi.org/10.1080/02626669709492022

Menabde, M., & Sivapalan, M. (2001). Linking space–time variability of river runoff and rainfall fields: a dynamic approach. *Advances in Water Resources*, *24*(9), 1001–1014. https://doi.org/https://doi.org/10.1016/S0309-1708(01)00038-0

Mitchell, K. E., Lohmann, D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B. A., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., … Bailey, A. A. (2004). The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. *Journal of Geophysical Research: Atmospheres*, *109*(D7). https://doi.org/https://doi.org/10.1029/2003JD003823

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, *116*(D12). https://doi.org/https://doi.org/10.1029/2010JD015139

NOA. (2016). *The national water model*. https://water.noaa.gov/about/nwm

Ogden, F. L., & Dawdy, D. R. (2003). Peak Discharge Scaling in Small Hortonian Watershed. *Journal of Hydrologic Engineering*, *8*(2), 64–73. https://doi.org/10.1061/(ASCE)1084-

0699(2003)8:2(64)

Ouarda, T. (2013). *Hydrological Frequency Analysis, Regional*. https://doi.org/10.1002/9780470057339.vnn043

Perez, G., Mantilla, R., Krajewski, W. F., & Quintero, F. (2019). Examining Observed Rainfall, Soil Moisture, and River Network Variabilities on Peak Flow Scaling of Rainfall-Runoff Events with Implications on Regionalization of Peak Flow Quantiles. *Water Resources Research*, *55*(12), 10707–10726. https://doi.org/https://doi.org/10.1029/2019WR026028

Perez, G., Mantilla, R., Krajewski, W. F., & Wright, D. B. (2019). Using Physically Based Synthetic Peak Flows to Assess Local and Regional Flood Frequency Analysis Methods. *Water Resources Research*, *55*(11), 8384–8403. https://doi.org/10.1029/2019WR024827

Quintero, F., Krajewski, W. F., Mantilla, R., Small, S., & Seo, B. C. (2016). A spatial-dynamical framework for evaluation of satellite rainfall products for flood prediction. *Journal of Hydrometeorology*, *17*(8), 2137–2154. https://doi.org/10.1175/JHM-D-15-0195.1

R. Foster, G., F. Huggins, L., & D. Meyer, L. (1984). A Laboratory Study of Rill Hydraulics: I. Velocity Relationships. *Transactions of the ASAE*, *27*(3), 790–796. https://doi.org/https://doi.org/10.13031/2013.32873

Reed, D., & Robson, A. (2008). *Statistical procedures for flood frequency estimation*.

Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins-Part 2: Flood and low flow studies. *Hydrology and Earth System Sciences*, *17*(7), 2637–2652. https://doi.org/10.5194/hess-17-2637-2013

Schroeder. (2012). *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. Dover Publications.

Svensson, C., & Jones, D. A. (2010). Review of rainfall frequency estimation methods. *Journal of Flood Risk Management*, *3*(4), 296–313. https://doi.org/https://doi.org/10.1111/j.1753-318X.2010.01079.x

Velásquez, N., Hoyos, C. D., Vélez, J. I., & Zapata, E. (2020). Reconstructing the 2015 Salgar flash flood using radar retrievals and a conceptual modeling framework in an ungauged basin. *Hydrology and Earth System Sciences*, *24*(3), 1367–1392. https://doi.org/10.5194/hess-24-1367-2020

Velez, J. I. (2001). *Desarrollo de un Modelo Hidrologico Conceptual y Distribuido Orientado a la Simu- lacion de Crecidas*. Universidad Politecnica de Valencia.

Velez, J. I., Velasquez, N., & Hoyos, C. D. (2018). Watershed modelling framework, a programable tool for hydrologic simulation and analysis. *American Geophysical Union Fall Meeting*. https://agu.confex.com/agu/fm18/meetingapp.cgi/Paper/467285

Viglione, A., Laio, F., & Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis. *Water Resources Research*, *43*(3), 1–10. https://doi.org/10.1029/2006WR005095

Webster, R., & Burrough, P. A. (1972). computer-based soil mapping of small areas from sample data. *Journal of Soil Science*, *23*(2), 222–234. https://doi.org/https://doi.org/10.1111/j.1365-2389.1972.tb01655.x

WRC. (1977). *Guidelines for determining flood flow frequency. Bulletin 17A of the Hydrology*

Subcommittee.

Wright, D. B. (2018). Rainfall Information for Global Flood Modeling. In *Global Flood Hazard* (pp. 17–42). American Geophysical Union (AGU). https://doi.org/https://doi.org/10.1002/9781119217886.ch2

Wright, D. B., Mantilla, R., & Peters-Lidard, C. D. (2017). A remote sensing-based tool for assessing rainfall-driven hazards. *Environmental Modelling & Software*, *90*, 34–54. https://doi.org/https://doi.org/10.1016/j.envsoft.2016.12.006

Wright, D. B., Yu, G., & England, J. F. (2020). Six decades of rainfall and flood frequency analysis using stochastic storm transposition: Review, progress, and prospects. *Journal of Hydrology*, *585*(February), 124816. https://doi.org/10.1016/j.jhydrol.2020.124816

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*(10), e2020EF001613. https://doi.org/https://doi.org/10.1029/2020EF001613

Yu, G., Wright, D. B., Zhu, Z., Smith, C., & Holman, K. D. (2019). Process-based flood frequency analysis in an agricultural watershed exhibiting nonstationary flood seasonality. *Hydrology and Earth System Sciences*, *23*(5), 2225–2243. https://doi.org/10.5194/hess-23-2225-2019

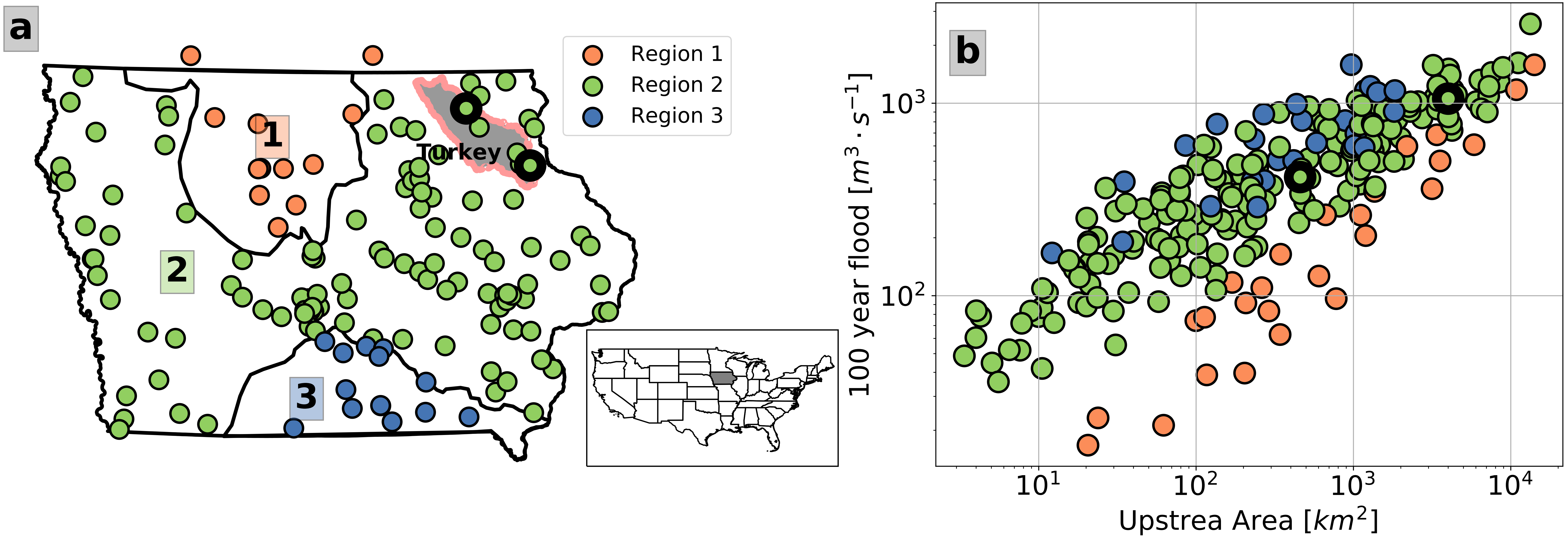
Zhang, Z., & Stadnyk, T. A. (2020). Investigation of attributes for identifying homogeneous flood regions for regional flood frequency analysis in Canada. *Water (Switzerland)*, *12*(9), 1–24. https://doi.org/10.3390/W12092570

Zhang, Z., Stadnyk, T. A., & Burn, D. H. (2020). Identification of a preferred statistical distribution for at-site flood frequency analysis in Canada. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *45*(1), 43–58. https://doi.org/10.1080/07011784.2019.1691942

Zhu, Z., Wright, D. B., & Yu, G. (2018). The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis. *Water Resources Research*, *54*(11), 8983–8998. https://doi.org/https://doi.org/10.1029/2018WR023550

Zrinji, Z., & Burn, D. H. (1994). Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology*, *153*(1–4), 1–21. https://doi.org/10.1016/0022-1694(94)90184-8

Figure 1.



a
Region 1
Region 2
Region 3
1
2
3
Turkey
b
100 year flood $[m^3 \cdot s^{-1}]$
Upstrea Area $[km^2]$
10^1
10^2
10^3
10^4

Figure 2.

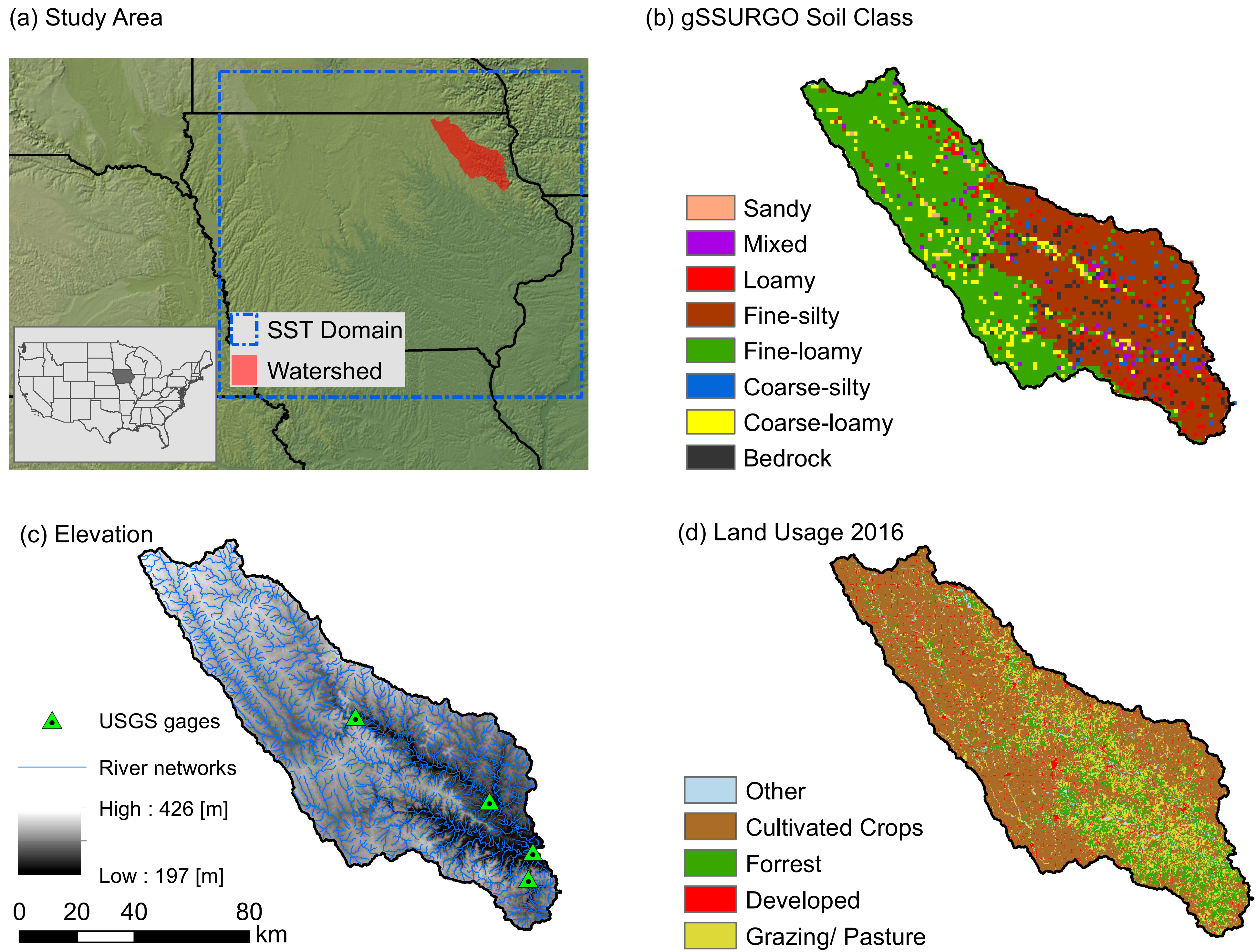
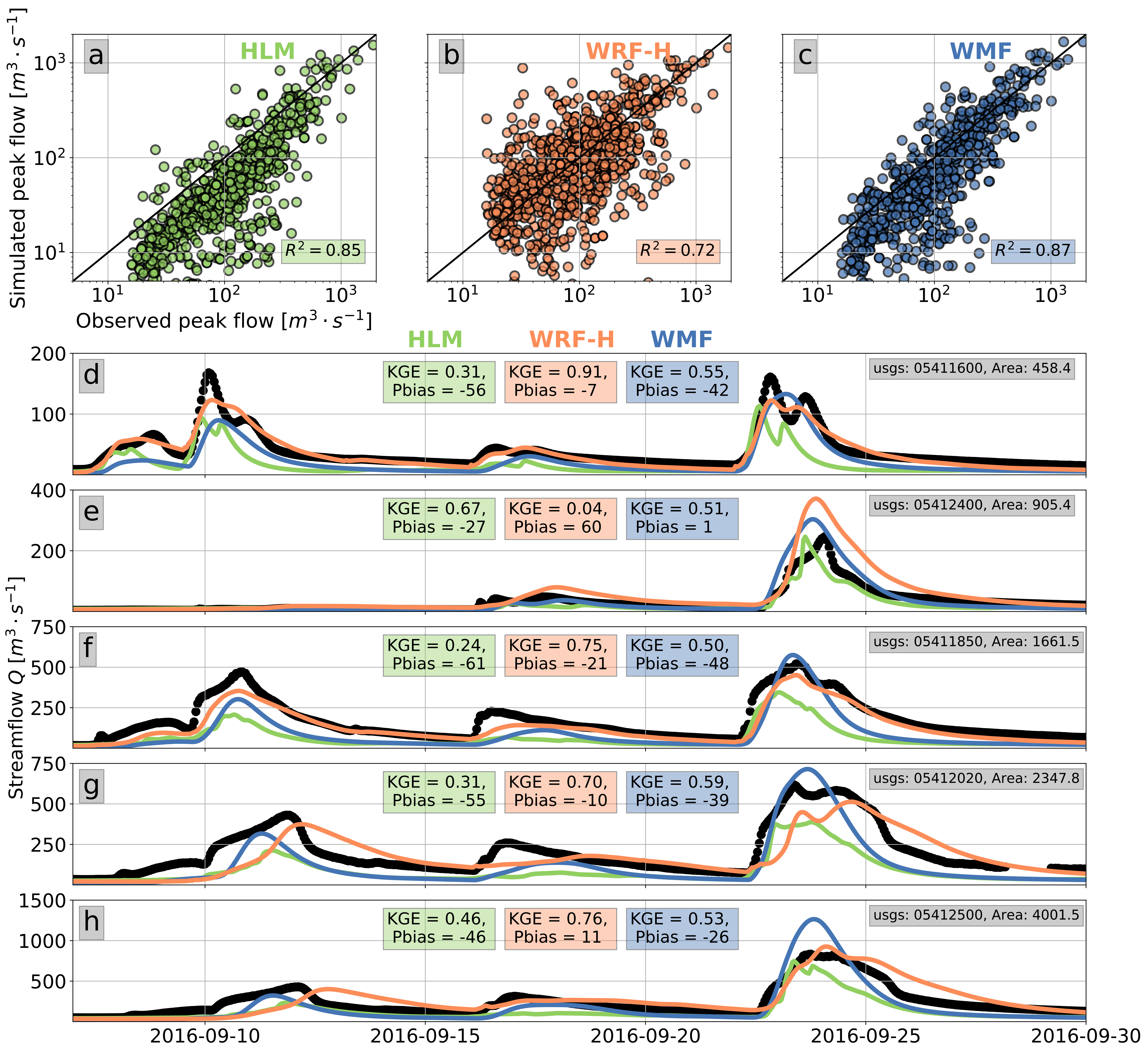
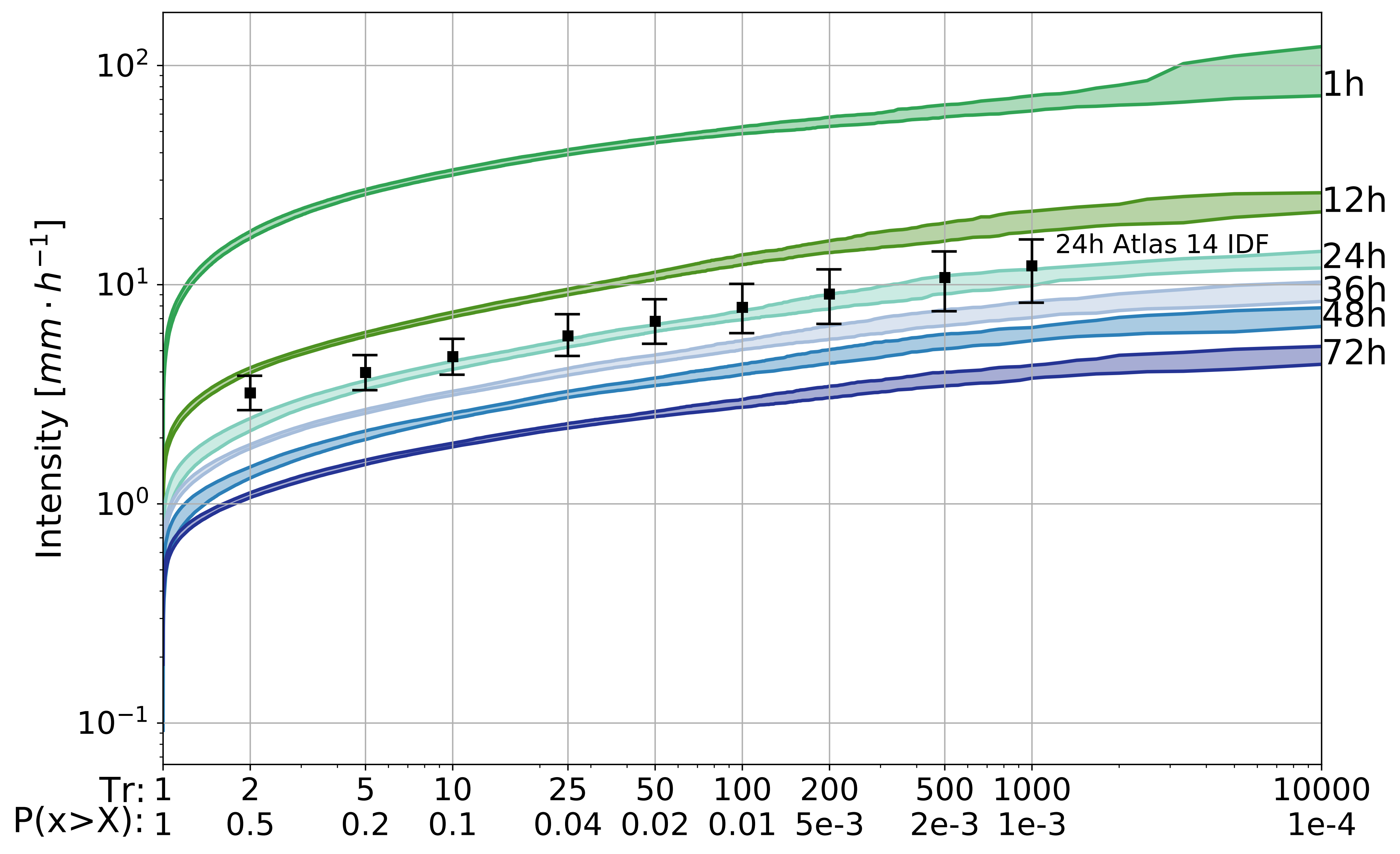

(a) Study Area
SST Domain
Watershed
(b) gSSURGO Soil Class
Sandy
Mixed
Loamy
Fine-silty
Fine-loamy
Coarse-silty
Coarse-loamy
Bedrock
(c) Elevation
USGS gages
River networks
High : 426 [m]
Low : 197 [m]
0
20
40
80
km
(d) Land Usage 2016
Other
Cultivated Crops
Forrest
Developed
Grazing/ Pasture

Figure 3.



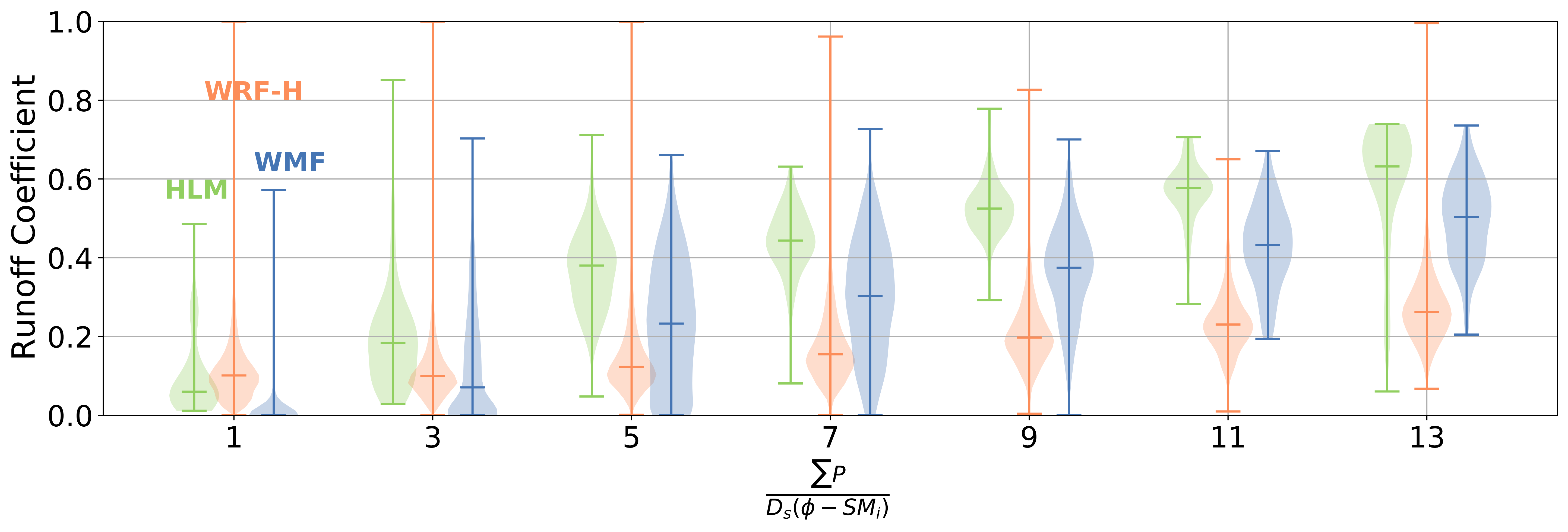
a
HLM
$R^2 = 0.85$
b
WRF-H
$R^2 = 0.72$
c
WMF
$R^2 = 0.87$
Simulated peak flow $[m^3 \cdot s^{-1}]$
Observed peak flow $[m^3 \cdot s^{-1}]$
HLM
WRF-H
WMF
d
KGE = 0.31, Pbias = -56
KGE = 0.91, Pbias = -7
KGE = 0.55, Pbias = -42
usgs: 05411600, Area: 458.4
e
KGE = 0.67, Pbias = -27
KGE = 0.04, Pbias = 60
KGE = 0.51, Pbias = 1
usgs: 05412400, Area: 905.4
f
KGE = 0.24, Pbias = -61
KGE = 0.75, Pbias = -21
KGE = 0.50, Pbias = -48
usgs: 05411850, Area: 1661.5
g
KGE = 0.31, Pbias = -55
KGE = 0.70, Pbias = -10
KGE = 0.59, Pbias = -39
usgs: 05412020, Area: 2347.8
h
KGE = 0.46, Pbias = -46
KGE = 0.76, Pbias = 11
KGE = 0.53, Pbias = -26
usgs: 05412500, Area: 4001.5
Streamflow Q $[m^3 \cdot s^{-1}]$
2016-09-10
2016-09-15
2016-09-20
2016-09-25
2016-09-30

Figure 4.



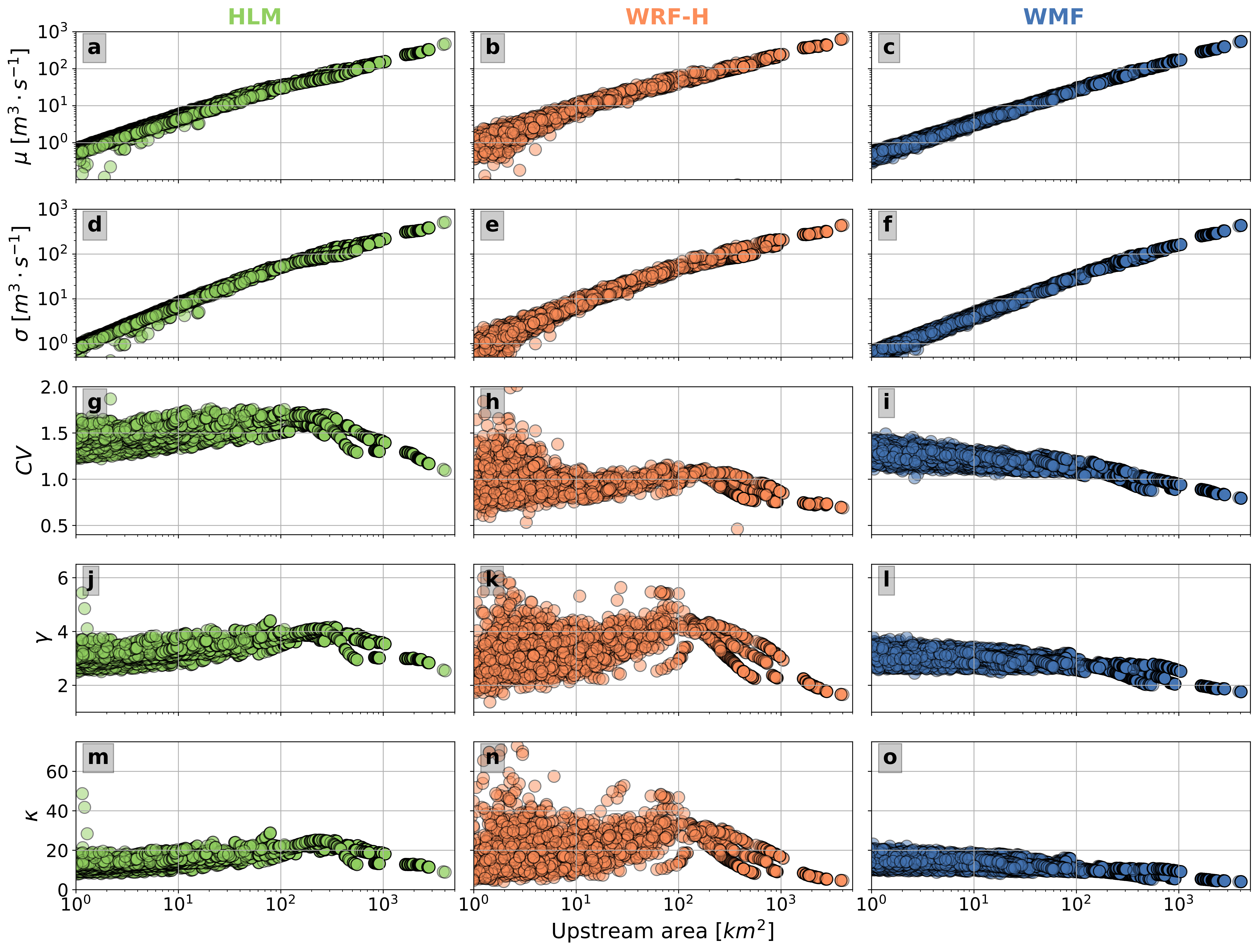
Intensity [$mm \cdot h^{-1}$]
10^2
10^1
10^0
10^{-1}
1h
12h
24h
36h
48h
72h
24h Atlas 14 IDF
Tr: 1 2 5 10 25 50 100 200 500 1000 10000
P(x>X): 1 0.5 0.2 0.1 0.04 0.02 0.01 5e-3 2e-3 1e-3 1e-4

Figure 5.



WRF-H
WMF
HLM
Runoff Coefficient
1.0
0.8
0.6
0.4
0.2
0.0
1
3
5
7
9
11
13
$\frac{\sum P}{D_s(\phi - SM_i)}$

Figure 6.



HLM
WRF-H
WMF
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
$\mu\ [m^3 \cdot s^{-1}]$
$\sigma\ [m^3 \cdot s^{-1}]$
CV
γ
κ
Upstream area $[km^2]$

Figure 7.

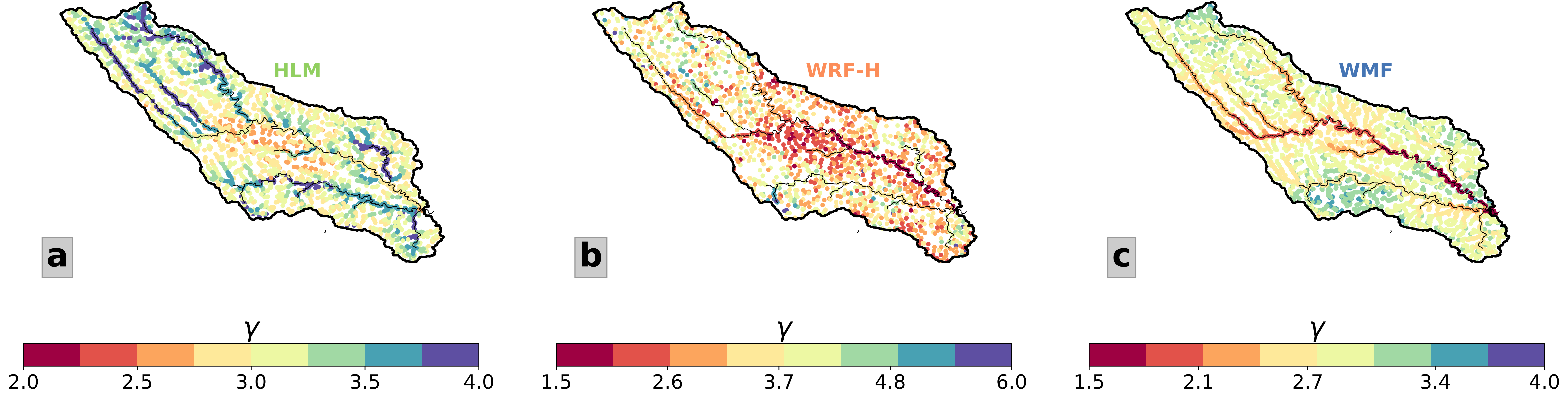
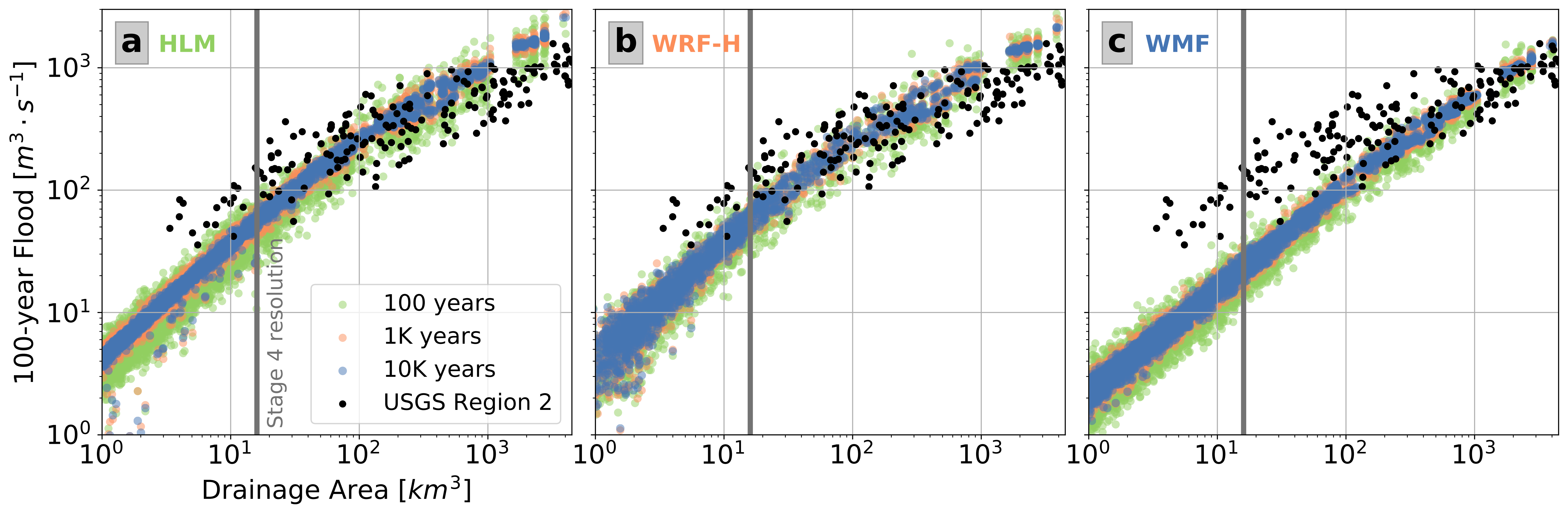
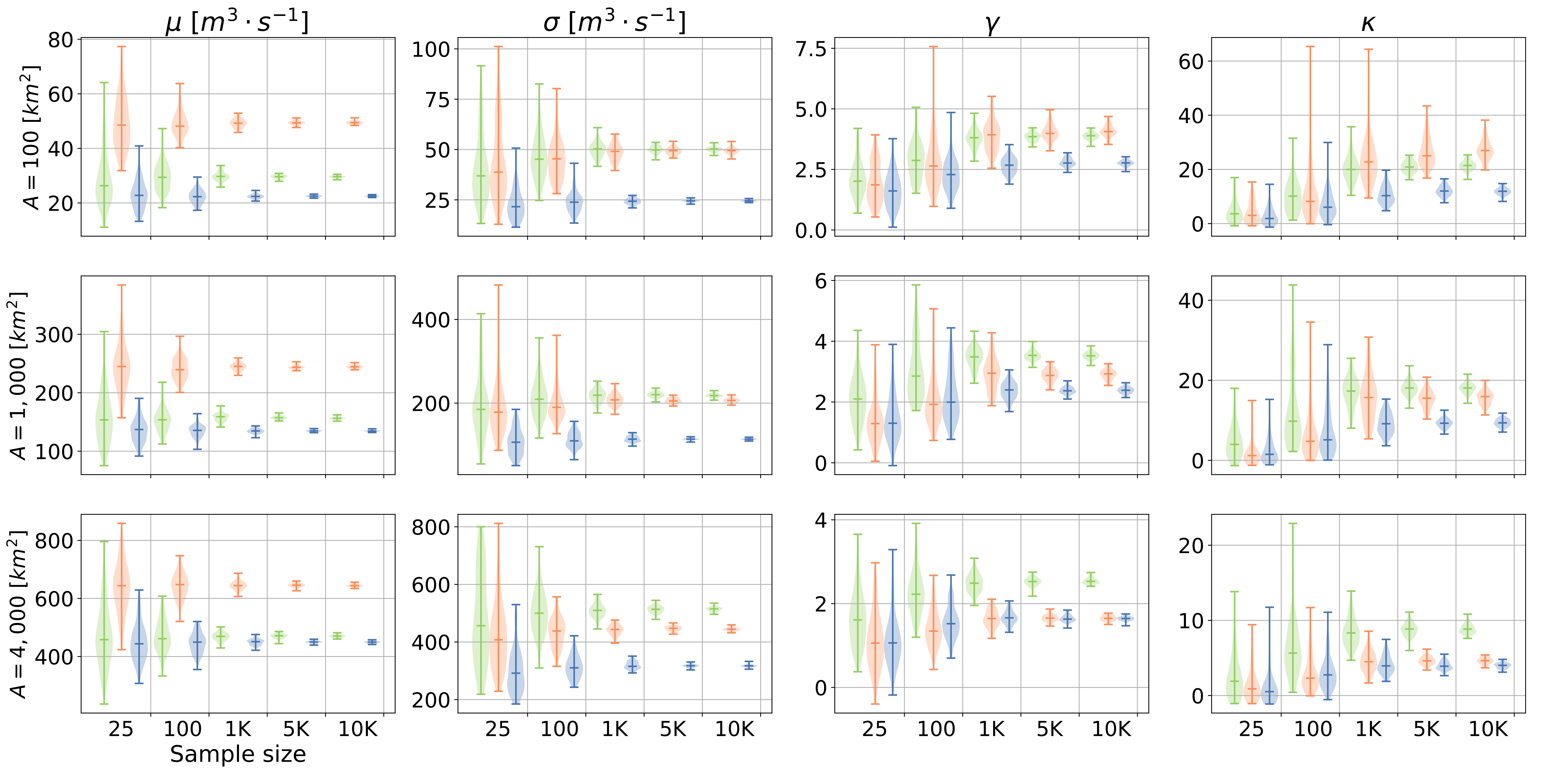

HLM
a
WRF-H
b
WMF
c
γ
2.0
2.5
3.0
3.5
4.0
γ
1.5
2.6
3.7
4.8
6.0
γ
1.5
2.1
2.7
3.4
4.0

Figure 8.



a
HLM
b
WRF-H
c
WMF
100-year Flood [$m^3 \cdot s^{-1}$]
Drainage Area [km^3]
Stage 4 resolution
100 years
1K years
10K years
USGS Region 2

Figure 9.



$\mu\ [m^3 \cdot s^{-1}]$
$\sigma\ [m^3 \cdot s^{-1}]$
γ
κ
$A = 100\ [km^2]$
$A = 1,000\ [km^2]$
$A = 4,000\ [km^2]$
25
100
1K
5K
10K
Sample size