A remote sensing-based tool for assessing rainfall-driven hazards

Daniel B. Wright a, *, Ricardo Mantilla b, Christa D. Peters-Lidard c

a University of Wisconsin-Madison, Madison, WI, United States
b University of Iowa, Iowa City, IA, United States
c NASA Goddard Space Flight Center, Greenbelt, MD, United States

ABSTRACT

RainyDay is a Python-based platform that couples rainfall remote sensing data with Stochastic Storm Transposition (SST) for modeling rainfall-driven hazards such as floods and landslides. SST effectively lengthens the extreme rainfall record through temporal resampling and spatial transposition of observed storms from the surrounding region to create many extreme rainfall scenarios. Intensity-Duration-Frequency (IDF) curves are often used for hazard modeling but require long records to describe the distribution of rainfall depth and duration and do not provide information regarding rainfall space-time structure, limiting their usefulness to small scales. In contrast, RainyDay can be used for many hazard applications with 1–2 decades of data, and output rainfall scenarios incorporate detailed space-time structure from remote sensing. Thanks to global satellite coverage, RainyDay can be used in inaccessible areas and developing countries lacking ground measurements, though results are impacted by remote sensing errors. RainyDay can be useful for hazard modeling under nonstationary conditions.

© 2016 Elsevier Ltd. All rights reserved.

Software availability

Name of Software: RainyDay Rainfall Hazard Modeling System
Developer: Daniel B. Wright
Contact: Daniel B. Wright; Address: Room 1269C Engineering Hall, 1415 Engineering Drive, Madison, WI 53706, USA; Email: danielb.wright@wisc.edu
Year first available: 2015
Required hardware and software: RainyDay requires Python 2.7 or newer (not tested with Python 3.0 or higher) with Numpy and Scipy. The Netcdf4 and GDAL APIs are also required. RainyDay will run on Macintosh, Linux, and Windows machines
Cost: Free. RainyDay is currently available by request. Open-source release under version 3.0 of the GNU General Public License (http://www.gnu.org/licenses/gpl-3.0.en.html) is planned

1. Introduction

Rainfall-driven hazards such as floods and landslides are the most common natural disasters worldwide, and amongst the most devastating. A growing number of computational hazard models are available to transform extreme rainfall inputs into hazard predictions, including distributed hydrologic models for the movement of water into and through river systems (e.g., Smith et al., 2004); hillslope stability and run-out models for landslide initiation and subsequent motion (e.g., Brenning, 2005; Preisig and Zimmermann, 2010; respectively); and hydraulic models for flood wave propagation in channels and floodplains (e.g., Horritt and Bates, 2002). These models have seen significant advances in recent decades, and have become key components in probabilistic hazard and risk assessment in fields such as natural catastrophe risk insurance, infrastructure design, and land-use planning. The hazard predictions produced by these models tend to be highly sensitive to the amount, timing, and spatial distribution of rainfall inputs. Unfortunately, progress on developing realistic rainfall inputs for probabilistic hazard and risk assessment has been relatively limited. This paper introduces RainyDay, a Python-based platform that addresses this shortcoming by coupling rainfall remote sensing data from satellites or other sources with a technique for temporal resampling and spatial transposition known as Stochastic Storm Transposition (SST) to generate highly realistic probabilistic rainfall scenarios.

Rainfall inputs for long-term hazard and risk assessment require a probabilistic description of three interrelated components:
duration, intensity, and space-time structure. Efforts to jointly model these components are usually referred to as rainfall frequency analysis, a simple term that belies the complexity of the physical phenomena and analytical methods involved. The probability structure of the first two components, rainfall duration and intensity, has been a focus of research and application for decades (see U.S. Weather Bureau, 1958 and Yarnell, 1935 for early examples). These two components are strongly linked and together they determine the probability distribution of rainfall volume (or depth) at a point or over an area. The third component, space-time structure, describes the spatial and temporal variability of rainfall and is determined by storm size, velocity, and temporal evolution of spatial rainfall coverage. Space-time structure can thus be understood as describing the “when” and “where” of extreme rainfall, whereas intensity and duration describe “how much.”

Rainfall space-time structure can be an important hazard determinant. For example, a rainstorm that is short-lived and small in spatial extent may pose a significant flash flood threat in a narrow mountain valley or urban area, but may not represent a hazard on a larger river system. Conversely, a month-long rainy period could lead to flooding on a major river due to the gradual accumulation of water in soils, river channels, and reservoirs, but may never feature a short-lived “burst” of rainfall sufficiently intense to cause flash flooding at smaller scales. Similarly, a storm that covers a large area or passes over a series of valleys could lead to more widespread landslide or debris flow occurrences than a smaller or stationary storm. Rainfall space-time structure and its importance as a hazard trigger, therefore, must be understood within the context of the particular geography and scale in question. Due to its complexity, rainfall space-time structure has traditionally been less well understood than intensity and duration, and its representation in hazard modeling has been less sophisticated.

The probability distribution of rainfall depth or intensity for a given duration is usually derived from rain gages and distilled into Intensity-Duration-Frequency (IDF) curves, such as those provided by the National Oceanic and Atmospheric Administration’s (NOAA) Atlas 14 (Bonnin et al., 2004). Records spanning many decades are generally needed to define the extreme tail of such distributions. The challenge of measuring extreme rainfall over long time periods and over large areas using rain gages has hindered IDF estimation in many developed countries, while the lack of data in poor countries and in inaccessible terrain means that IDF estimation using such methods is virtually impossible in many locations. Furthermore, measurements of rainfall space-time structure at a high level of detail using dense networks of rain gages are nonexistent outside of a handful of wealthy cities and research-oriented efforts. “Regionalization,”—the pooling of hazard information over a larger area in order to inform analyses at particular locations (see, e.g. Alexander, 1963 for an early discussion of rainfall regionalization and Stedinger et al., 1993 for a review)—has helped with IDF estimation in areas where rain gage densities are moderate or high. These techniques offer little help, however, in parts of the world where rain gages are few or nonexistent, and do not offer a framework for incorporating rainfall space-time properties into hazard estimation. Even where long rainfall records do exist, nonstationarity due to climate change may mean that earlier portions of the record are no longer representative of current or future IDF properties.

Several techniques, which generally fall under the term of design storm methods, are used in long-term hazard estimation to link IDF properties to space-time structure for probabilistic flood hazard assessment (commonly referred to as flood frequency analysis). Design storm methods include linking rainfall duration to rainfall intensity via a measure of flood response time, such as the time of concentration (e.g. McCuen, 1998), deriving estimates of area-averaged rainfall from point-scale rainfall estimates using area reduction factors (ARFs; U.S. Weather Bureau, 1958), and using dimensionless temporal disaggregation such as the family of U.S. Soil Conservation Service 24-h rainfall distributions (e.g. McCuen, 1998). Each is highly empirical, laden with assumptions (see Wright et al., 2014a; Wright et al., 2014b; Wright et al., 2013), valid only in certain contexts, and often misunderstood or misused (K. Potter, personal communication, May 6, 2015).

SST explicitly links IDF to rainfall space-time properties, providing certain advantages over design storm methods. Similar to other regionalization techniques, SST aims to effectively “lengthen” the period of record by using nearby observations, albeit using a fundamentally different approach involving temporal resampling and spatial transposition of rainstorms drawn from a catalog of observed rainfall events from the surrounding region. The inclusion of nearby storms at least partially addresses the difficulty of accurately estimating rainfall hazards using short records. SST can be used to estimate rainfall IDF properties and also to facilitate modeling of interactions of rainfall space-time structure with geographic features (such as hillslopes and river networks) at the appropriate spatial and temporal scales. It accomplishes this by generating large numbers of extreme rainfall “scenarios,” each of which has realistic rainfall structure based directly on observations.

Alexander (1963), Foufoula-Georgiou (1989), and Fontaine and Potter (1989) describe the general SST framework, while Wilson and Foufoula-Georgiou (1990) use the method for rainfall frequency analysis and Gupta (1972) and Branchini et al. (1996) use it for flood frequency analysis. In those days, however, the method was of limited practical use due to the lack of detailed rainfall datasets with large areal coverage. Those studies also did not focus explicitly on the aspects of SST related to rainfall space-time structure nor its implications for hazard modeling.

The recent advent of satellite-based remote sensing provides a relatively low-cost means of measuring extreme rainfall over large parts of the globe at moderately high spatial and temporal resolution (30 min–3 h, 4—25 km), while ground-based weather radar offers higher-resolution estimates (5—60 min, typically 1—4 km) over smaller regions. While the accuracy of rainfall remote sensing can be poor (particularly for satellite-based estimates, e.g. Mehran and AghaKouchak, 2014; and in mountainous regions, e.g. Nikolopoulos et al., 2013; Stampoulis et al., 2013), such data nonetheless offer unprecedented depictions of rainfall over large areas, offering opportunities for hazards research and practice at various scales, ranging from forecasting and post-event analysis to long-term hazard assessment.

In the context of SST, the ongoing accumulation of remote sensing data to lengths of 10—20 years or more “unlocks” many of the as-yet unrealized opportunities offered by SST. Wright et al. (2013) demonstrated the coupling of SST with a 10-year high-resolution radar rainfall dataset for IDF estimation, and the method was extended to flood frequency analysis for a small urban watershed using a distributed hydrologic model in Wright et al. (2014b). These two papers, along with Wright et al. (2014a) show that commonly-used design storm practices (ARFs, dimensionless time distributions) have serious shortcomings in representing the multiscale space-time structure of extreme rainfall and critical interactions with of this structure with watershed and river network features. Wright et al. (2014b) also show that when SST is coupled with rainfall remote sensing data and a distributed hydrologic model, it can reproduce the role that this structure plays in determining multi-scale flood response. The RainyDay software described in this paper was developed to facilitate the use of SST in conjunction with rainfall remote sensing data.

Though SST was developed in the context of flood hazard estimation, it may prove useful for rainfall-triggered landslides and other mass movements, subject to the oftentimes poor accuracy of
remote sensing data in steep terrain as well as other limitations that will be discussed subsequently. Rainfall space-time structure governs the temporal distribution of rainfall volume onto individual hillslopes, as well as the number of hillslopes subject to rainfall. In addition, steep landslide-prone terrain often has poorer rain gauge coverage than lowland areas due to limited accessibility, suggesting that remote sensing rainfall estimates are potentially useful in such regions, particularly if improvements in accuracy can be realized (e.g. Shige et al., 2013).

Section 2 provides a description of the SST methodology used in RainyDay. Section 3 discusses the specific implementation of SST in RainyDay and some of the software's important features. Section 4 provides sample results from RainyDay and sensitivity analyses using different input rainfall datasets for rainfall and flood frequency analysis in order to illustrate its capabilities and some of its limitations, including for flood frequency analysis in nonstationary conditions. Section 5 includes discussion and concluding remarks.

2. The SST methodology

In this section, we provide a step-by-step methodology for SST-based rainfall frequency analysis for a user-defined geographic “area of interest,” of arbitrary shape. A high-level description of software features is left to Section 3, but it merits mention that in RainyDay, A can be a single remote sensing pixel, a rectangular area containing multiple pixels, or a contiguous area defined by a user-supplied polygon in shapefile format.

The following five steps describe the SST methodology, as implemented in RainyDay:

1. Identify a geographic transposition domain A’ that encompasses the area of interest A. One could confine A’ to regions with homogeneous extreme rainfall properties, (e.g. flat areas far from large water bodies and topographic features). However, such homogeneity would likely be difficult to rigorously determine in practice and regardless, such strict interpretation is likely to be overly limiting. RainyDay offers several diagnostic aids, discussed in Section 3.3, that help the user to understand rainfall heterogeneity over the region A’ and to improve the performance of the SST procedure in cases where rainfall heterogeneities do exist. Additional issues related to the selection of A’ are explored in Section 4.3.

2. Identify the largest m temporally non-overlapping storms in A’ from an n-year rainfall remote sensing dataset, in terms of rainfall accumulation of duration t and with the same size, shape, and orientation of A. For example, the principal axis of the Turkey River watershed in northeastern Iowa in the central United States is oriented roughly northwest-southeast and has an area of 4400 km². In this case, the m storms are those associated with the m highest t-hour rainfall accumulations over an area of 4400 km² with the same size, shape, and orientation as the Turkey River watershed. We refer to this set of storms henceforth as a “storm catalog,” with the same geographic extent as A’ and the same spatial and temporal resolution as the input rainfall data. We refer to the m storms in the storm catalog henceforth as “parent storms.” In RainyDay, the user can specify whether to exclude certain months (such as wintertime) from the storm catalog. Previous studies have shown that there can be low bias introduced in high-exceedance probability (i.e. frequent, low-intensity) events if m is small (e.g. Foufoula-Georgiou, 1989; Franchini et al., 1996; Wilson and Foufoula-Georgiou, 1990; see Wright et al., 2013 for a discussion). The sensitivity of SST results to the choice of m and A’ is explored in detail in Section 4.3, but m = 10n generally minimizes the low bias for frequent events, and would likely be a good starting point for new analyses. Low exceedance probability (i.e. rare) events are less sensitive to the choice of m (see Section 4.3).

In RainyDay, duration t is a user-defined input, and if t is neither very short nor very long relative to the time scale of hazard response in A, subsequent hazard modeling results will be relatively insensitive to the chosen value. In this respect, the duration t in SST differs conceptually from design storm methods, in which hazard response is intrinsically sensitive to the user-specified duration, and this feature is indeed one of the chief advantages of SST over design storm methods for multi-scale flood hazard estimation (see Wright et al., 2014b for analysis and discussion). In the case of SST-based flood frequency analysis, t should be at least as long as the watershed time of concentration and preferably somewhat longer.

3. Randomly generate an integer k, which represents a “number of storms per year.” In previous SST literature, the assumption was made that k follows a Poisson distribution with a rate parameter λ storms per year. The m parent storms are selected such that an average of λ = m/n storms per year are included in the storm catalog. For example, if m = 100 storms selected from a ten-year remote sensing dataset, then λ = 100/10 = 10.0 storms per year. RainyDay will generate k using either the Poisson distribution or an empirical distribution, discussed in Section 3.3. If the Poisson distribution is selected, RainyDay will automatically calculate λ based on user-specified m and the length of the input dataset.

4. Randomly select k parent storms from the storm catalog. For each selected parent storm, transpose all rainfall fields associated with that storm by an east-west distance Δx and a north-south distance Δy, where Δx and Δy are drawn from the distributions D(x) and D(y) which are bounded by the east-west and north-south extents of A’, respectively. The motion and structure of the parent storm is unaltered during transposition and only the location is changed. The distributions D(x) and D(y) were taken to be uniform in Wright et al. (2013, 2014b), but RainyDay offers additional options, described in Section 3.3. We illustrate this step schematically in Fig. 1. For each of the k transposed storms, compute the resulting t-hour rainfall accumulation averaged over A.

Step 4 can be understood as temporal resampling and spatial transposition of observed storm events within a probabilistic framework to synthesize one year of heavy rainfall events over A’ and, by extension, over A. RainyDay and previous SST efforts retain the largest (in terms of rainfall intensity) of the k events for subsequent steps and discard the k-1 remaining events, though in principle these events could be retained. The single retained storm can be understood as a “synthetic” annual rainfall maximum, analogous to those annual rainfall maxima that are extracted from rain gauge records for rainfall frequency analysis. It should be noted that these rainfall events do not form a continuous series, meaning that neither inter-storm periods nor the temporal sequencing of the k storms are considered.

5. Repeat steps 3 and 4 a user-specified Tmax number of times, in order to create Tmax years of t-hour synthetic annual rainfall maxima for A. RainyDay then assigns each annual maxima a rank i according to its rainfall intensity relative to all others. Each of these ranked maxima can then be assigned an annual exceedance probability p*e where p*e = i/Tmax. Exceedance probability p*e is the probability in a given year that an event of equal or greater intensity will occur. The “return period” or “recurrence interval” Ti, commonly used in hazard analysis, is simply Ti = 1/p*e. It is possible to directly infer exceedance probabilities of 1.0 ≥ p*e ≥ 10^-3 (recurrence intervals of 1 ≤ Ti ≤ 10^3). Each of these rainfall events can then serve as one datum of an empirical IDF estimate or as a rainfall scenario for hazard modeling.
To ensure accessibility for users inexperienced with Python, all the necessary Python modules are supported within recent versions of the Anaconda Python distribution from Continuum Analytics (https://store.continuum.io/cshop/anaconda). The user must install NetCDF4 libraries and any requisite dependencies. If the user wishes to use shapefile functionality, necessary for defining A to be a shape other than a rectangle or a single rainfall pixel, the GDAL library (http://www.gdal.org) and any necessary dependencies must also be installed.

3. SST internal variability

In RainyDay, the user specifies N, the number of $T_{\text{max}}$-year long “ensemble members” to be generated. This enables examination of “internal variability,” i.e. how much variation in rainfall intensity is possible for a given $p_e$ for a given input rainfall dataset and set of user-defined parameters. For example, if the user specifies $T_{\text{max}} = 10^3$ and $N = 100$, then there will be 100 intensity estimates for each $p_e$ between 1.0 and $10^{-3}$. RainyDay will automatically generate text file and graphics files containing the results of this rainfall frequency analysis, including the rainfall mean, minimum, and maximum (or, optionally, a quantile interval) for each $p_e$, computed from the $N$ ensemble members.

If the scenarios generated by RainyDay are fed through a hazard model, then the ensemble spread will propagate through to generate ensemble hazard estimates. A useful and interesting feature of SST and RainyDay that is not examined in this paper, but is discussed at length in Wright et al. (2014b), is that the exceedance probability of rainfall and of subsequent hazards can be decoupled using SST, particularly if some realistic scheme is used to account for the initial conditions in A (such as soil moisture or baseflow). Consider the example where $N = 1$ and $10^3$ rainfall scenarios ($T_{\text{max}} = 10^3$) are created as input to a distributed flood hydrologic model. One of these rainfall scenarios has $p_e = 0.01$ (in terms of watershed-average t-hour rainfall depth over an area A). Even if initial conditions are kept constant across all $T_{\text{max}}$ simulations, the $p_e$ of the peak discharge or volume predicted by the model for this particular scenario need not be equal to 0.01, since the space-time structure of the rainfall scenario and its interactions with watershed and river network features can dampen or magnify the flood severity. If variability in initial conditions within the hazard model are considered, this dampening or magnification can be even greater. This property of SST contrasts with design storm methods, which typically assume a 1:1 relationship between the exceedance probability of rainfall and of subsequent hazards.

3. RainyDay software

3.1. Overview of software

We wrote RainyDay to render SST more accessible and to streamline the code for speed and ease-of-use using Python. The majority of subroutines utilize the Scipy (Jones et al., 2011) and Numpy packages (Walt et al., 2011). Fig. 2 shows a schematic of workflow in RainyDay.

While the ranking of rainfall events described in Step 5 of the SST methodology in Section 2 is based on rainfall intensity averaged over A, RainyDay will create NetCDF4 files (http://www.unidata.ucar.edu/software/netcdf) that contain the transposed rainfall scenarios with full depictions of rainfall space-time structure at the native spatial and temporal resolution of the input. This is an important feature because space-time structure, and not just average rainfall intensity over area A and duration $t$, is important in determining hazard response. For example, one rainfall scenario may produce a more severe flood response than another scenario, even if it has a lower overall average rainfall intensity over A and $t$, due to interactions with watershed features (see Section 3.2 of this paper for discussion and Wright et al., 2014b for analysis).

We will provide the RainyDay source code, examples, and user documentation upon request, and intend to release it under version 3 of the GNU General Public License (http://www.gnu.org/copyleft/gpl.html) once we have completed sufficient testing and documentation. The code is currently not parallelized. Computational time is determined mainly by the size of the input dataset (record length n, input resolution, and geographic size of A and A’), while other factors, such as m, t, $T_{\text{max}}$ and N can impact runtime. Computational speed, even without parallelization, is not prohibitive on a modern desktop or laptop computer (several seconds to several hours for typical configurations and input datasets).
Ensemble spread is shown throughout Section 4 to illustrate various aspects of SST-based rainfall and flood frequency analysis. If the user is only interested in examining internal variability of SST-based rainfall IDF, then the number of ensemble members can be large (e.g. $N \geq 100$). If the user wishes to perform hazard simulations, however, $N$ should be selected with consideration of the computational and storage costs associated with large numbers of simulations, which can be substantial depending on the particular hazard model. To help manage the number of simulations required, the user can specify a rainfall return period threshold, below which output scenarios will not be created. For example, if the user specifies a 5-year threshold, no rainfall scenarios with a rainfall...
depth less than the 5-year return period depth will be written, which reduces the number of hazard simulations by 80% for a given value of $N$ while still retaining the most extreme scenarios.

### 3.3. Rainfall heterogeneity and non-uniform spatial transposition

A common criticism of SST is that its validity is restricted to regions with homogenous extreme rainfall properties. As previously mentioned, depending on how rigidly this criterion is enforced, the method would be limited to small, flat regions far from topographic features, water bodies, etc. It is unclear how homogeneity would be determined, particularly given the paucity of rainfall data in most regions. Instead, steps can be taken to use SST in more varied geophysical settings. Regardless of the setting, the selection of $A'$ requires an understanding of regional rainfall patterns and of the intrinsic assumptions of SST. Though more work is needed to understand the geographic limits of the applicability of RainyDay in complex terrain, the work of England et al. (2014) provides an example of SST usage in complex terrain.

RainyDay provides several tools to help understand the issue of rainfall heterogeneity, and, to some extent, to mitigate it. First, RainyDay produces a map showing the location of the rainfall centroids for all storms in the storm catalog, overlaid on a smoothed field of the spatial probability of storm occurrence within $A'$. This spatial probability of occurrence map is generated by applying a two-dimensional Gaussian kernel smoother to the $(x,y)$ locations of the rainfall centroids for all the storms in the storm catalog. This smoothed field is then normalized such that the sum of all grid cells in $A$ equals 1.0, thus creating a two-dimensional probability density function (PDF) of storm occurrence. A second plot shows these rainfall centroids overlaid with the average rainfall per storm across $A'$. These diagnostic plots assist in understanding regional variations in storm occurrences and rainfall over $A'$. Examples of these diagnostic plots for a region $A'$ encompassing most of the state of Iowa in the central United States are shown in Fig. 3. The top panel suggests that storms are somewhat more frequent in the southernmost third or so of the transposition domain (top panel), along with slightly elevated activity in the northeast quadrant. The bottom panel shows somewhat higher average storm rainfall in these two areas. Caution should be taken when drawing firm conclusions from these diagnostic plots, however, since rainfall heterogeneities evident in both storm occurrences and average storm rainfall may be the result of spatial biases in rainfall remote sensing estimates or of randomness in the climate system over the relatively short remote sensing record, rather than from “true” heterogeneity in the underlying rainfall hydroclimate.

Additional optional diagnostic outputs include static and animated rainfall maps for each storm in the storm catalog (not shown). These storm rainfall maps are useful for diagnosing “bad data,” particularly in rainfall datasets that use ground-based weather radar contaminated by radar beam blockage and other artifacts. Anomalous storm periods must be identified by the user (i.e. no automatic data quality checking is provided), but such periods can be excluded from subsequent analyses.

The two-dimensional PDF of spatial storm probability of storm occurrence can optionally be used as the basis for non-uniform spatial transposition (providing the $D_0(x)$ and $D_0(y)$ described in Step 4 and Fig. 1 in Section 2) so that the spatial distribution of storm occurrences will be preserved between the input data and output rainfall scenarios and IDF estimates. Section 4.3 examines the impact of this optional feature on results for the Iowa study region, along with potential implications.

It is important to note that this approach only addresses the spatial heterogeneity of storm occurrences, not of spatial variations in the climatology of rainfall intensity (due to topography or other factors). For example, if $A'$ contains a flat plain and an adjacent mountain range, the probability of storm occurrence will vary across $A'$. This variation will be captured in the two-dimensional PDF of spatial storm probability and, using the optional non-uniform spatial transposition scheme, will be reflected in RainyDay outputs. In this example, however, rainfall intensity from these storms will also vary according to the underlying topography. The current transposition scheme in RainyDay cannot explicitly account for this variation, which is likely to be a serious constraint in some regions.

### 3.4. Empirical temporal resampling

As mentioned in Step 3 of the SST procedure described in Section 2, previous SST work has employed the assumption that the annual number of storm counts follows a Poisson distribution, which in turn serves as the basis for the temporal resampling of storms (i.e. for generating the number of storms per year $k$ that will
be spatially transposed). RainyDay supports Poisson-based resampling, but also allows the use of an empirical distribution. This distribution is derived from the number of storms that enter into the storm catalog from each calendar year in the rainfall input dataset. Then, during the temporal resampling step, \( k \) is obtained by randomly selecting one of these values. This feature may be useful in regions where storm occurrences exhibit strong clustering (i.e., where there is strong evidence for more storms in some years and fewer in other years for persistent climatological reasons; e.g., Villarini et al., 2013). Section 4.3 examines the impact of this choice on SST results. Other discrete probability distributions, such as the two-parameter negative binomial (Pascal) distribution, can also be used to model clustered storm occurrences. RainyDay does not currently use such distributions, however, since short (typically 10–20 year) remote sensing records may yield poor parameter estimates stemming from the limited number of statistical degrees of freedom.

3.5. “Spin-up” of initial conditions

A key issue in the modeling of rainfall driven hazards is to adequately represent initial conditions. In many flood and landslide modeling efforts, the most critical of these is antecedent soil moisture, while other states such as snowpack, river baseflow, and water table position may also be relevant. Many hydrologic models allow for the specification of such initial conditions, and thus many design storm-based hazard modeling efforts rely on an assumed soil moisture state, such as an average or fully saturated condition. Such approaches have previously been used with SST (Wright et al., 2014b), and could be combined with the rainfall scenarios generated via RainyDay. This approach has the downside, however, that the true variability antecedent soil moisture is not captured in hazard predictions. This is particularly important in regions in which heavy rainfall does not necessarily occur in the same season as high soil moisture conditions. A second approach that can capture this variability would be to derive a distribution of antecedent soil moisture from previous long-term (ideally continuous multidecadal) model simulations. Since there can be substantial variation in how soil moisture is represented in different hazard models, the same model should be used for these long-term simulations and for the hazard scenario modeling. RainyDay offers an alternative option, however, in which initial soil moisture can be “spun up” within the hazard model to represent seasonally realistic initial conditions without the need for long-term simulations.

The spin-up procedure is described for a single rainfall scenario. The month of occurrence of the rainfall scenario is identified based on the “parent storm” that created it. Then RainyDay identifies the set of \( X \)-day periods (where \( X \) is a user-defined spin-up period) preceding all parent storms that occur within a user-defined number of months from the date of occurrence of the parent storm. One of the \( X \)-day periods is randomly selected and prepended to the rainfall scenario. This scheme helps to ensure that spin-up conditions are reasonable for the given season. It also helps ensure that spin-up conditions have realistic temporal correlations when pre-pended to the rainfall scenario (for example, if there is a historical tendency for several days of moderate rain prior to heavy storms but several days of heavy rain prior to the main storm doesn’t have historical precedent, these conditions will be properly represented). It is important to note, however, that the 10 to 20-year records typical of rainfall remote sensing records may not capture the full variability of “true” initial conditions.

This pre-pending procedure creates rainfall scenario output files that are of duration \( X + t \). The modeler can then assign an average initial soil moisture condition to initialize each model run, and use the rainfall scenario as input. Soil moisture within the model will then evolve over the spin-up period based on the rainfall (or lack thereof), evapotranspiration, and other modeled processes. This spin-up procedure has storage and computational costs since it can substantially increase the size of the rainfall scenario files generated by RainyDay and increase the length of each hazard simulation. The importance of these limitations depends on the size of \( A \), the resolution of the input rainfall dataset, and the computational burden of the hazard model. In Section 4.2, for example, \( X = 6 \) days and \( t = 4 \) days. This spin-up period is likely sufficient to spin up moisture in the upper soil layers, but not to fully establish baseflow or deeper groundwater flow. The modeler should evaluate the tradeoffs between longer \( X \) and the associated storage and computational costs.

3.6. Parametric rainfall intensity

Instead of relying on the rainfall intensity derived from a remote sensing input dataset, a user might prefer to use a parametric distribution to impose rainfall depths on the rainfall output scenarios. RainyDay supports this option. The user can supply a \( t \)-hour rainfall depth distribution. This distribution is then applied to the output rainfall scenarios via a normalization procedure that assumes that the supplied distribution corresponds to the annual maximum \( t \)-hour rainfall intensity for a single rainfall grid cell. Rainfall space-time structure is still derived from the remote sensing data. It should be noted, however, that when the resolution of the input remote sensing dataset is coarse relative to the spatial coverage of the rainfall measurement device upon which the parametric distribution is based (for example, the 16–625 km\(^2\) footprint of many satellite rainfall datasets relative to the 0.1 m\(^2\) sampling area of a single rain gage), this approach may be problematic. This procedure is also problematic in regions where such parametric rainfall distributions might be the synthesis of “mixture distributions” of distinct storm types in which rainfall intensity is intrinsically linked to rainfall space-time structure (e.g. Smith et al., 2011), since RainyDay does not distinguish between different storm types. Currently only the three-parameter generalized extreme value distribution (Walshaw, 2013) is supported, though it would be straightforward to add additional options.

4. Rainfall and flood case studies

4.1. Rainfall IDF

We used RainyDay to generate IDF results for durations from 3 to 96 h and \( p_r \) ranging from 0.5 to \( 10^{-3} \) for single rainfall grid cells in the vicinity of Iowa City, Iowa (Fig. 4) using rainfall data from Stage IV (Lin and Mitchell, 2003) and version 7.0 of the Tropical Rainfall Measurement Mission Multi-Satellite Precipitation Analysis (TMPA; Huffman et al., 2010). Stage IV is available through the National Weather Service (NWS) National Center for Environmental Prediction and provides hourly, 4 km resolution rainfall estimates by merging data from the NWS Next-Generation Radar network (NEXRAD; Crum and Alberty, 1993) with rain gages and, in some instances, satellite rainfall estimates. Stage IV has been extensively used in studies of extreme rainfall and flooding. All Stage IV analyses in this paper use data from 2002 to 2014. TMPA merges passive microwave, active radar, and infrared observations from multiple satellites to create a near-global (\( \pm 50^\circ \) latitude) rainfall dataset with 3-hourly, 0.25\(^\circ\) (approximately 25 km) resolution. Unless otherwise noted, TMPA analyses this study use the final “research version” of TMPA from 1998 to 2014, which includes a monthly rain gage-based bias correction. For the results in Fig. 4, and most subsequent analyses in this study, \( A \) is the rectangular area shown in Fig. 3. A is set to a single rainfall pixel and each run...
Fig. 4. Comparison of IDF curves from Atlas 14 and RainyDay using the Stage IV and TMPA rainfall datasets for 3-, 6-, 12-, 24-, 48-, and 96-h durations. Shaded areas for RainyDay estimates denote the ensemble spread. Bars on the NOAA Atlas 14 IDF estimates denote the 90% confidence intervals. Key RainyDay parameters: $m = 150$ storms, $A' = [40^\circ$ to $44^\circ$ N, $90^\circ$ to $96^\circ$ W]. $A$ is a single rainfall pixel (approximately 16 km$^2$ for Stage IV, 625 km$^2$ for TMPA), $N = 100$, $T_{\text{max}} = 1000$. Spatially-uniform transposition and Poisson-based temporal resampling are selected. Stage IV period of record is 2002–2014, TMPA period of record is 1998–2014. Analyses are restricted to April–November period.
increased temporal aggregation. Thus, while not de-
and spatial mismatch effects, both of which diminish with
convergence between Stage IV-based RainyDay IDFs and Atlas 14
Habib et al., 2009 for evidence of conditional biases in TMPA). The
using both datasets could potentially be explained by conditional
more frequent events and the underestimation for more rare events
overestimation of rainfall depth from TMPA (relative to Stage IV) for
Arti-
not. Finally, it includes the 60-min, approximately 4 km version of
gage-based bias correction scheme, and CMORPH Raw, which does
Joyce et al., 2004): CMORPH Corrected, which uses a daily rain-
2014. It also includes two versions of the 30-min resolution, 8 km
based bias correction, and TMPA-RT, which is produced in near-
procedure itself.
In order to highlight both the potential for IDF estimation and
and probabilistic hazard assessment in data-sparse regions using
RainyDay with satellite remote sensing and some of the associated
challenges, we compare 24-h IDF curves generated using RainyDay
for various satellite rainfall datasets for the vicinity of Iowa City
(Fig. 5). This comparison includes two versions of TMPA: the
aforementioned final version which includes monthly rain gage-
based bias correction, and TMPA-RT, which is produced in near-
time, does not feature bias correction, and runs from 2000 to
2014. It also includes two versions of the 30-min resolution, 8 km
Climate Prediction Center (CPC) Morphing Technique (CMORPH;
Joyce et al., 2004): CMORPH Corrected, which uses a daily rain-
gage-based bias correction scheme, and CMORPH Raw, which does
not. Finally, it includes the 60-min, approximately 4 km version of
Precipitation Estimation from Remotely Sensed Information Using
Artificial Neural Networks Global Cloud Classification System
(PERSIANN-GCCS; Sorooshian et al., 2000), which does not use
gage-based bias correction. The results in the top panel of Fig. 5
show relatively good agreement between point-scale NOAA Atlas
14 IDFs and single-pixel RainyDay-based IDFs for bias-corrected
TMPA and PERSIANN-GCCS, particularly considering the spatial
sampling mismatch between the remote sensing data and Atlas 14
mentioned previously, while results based on CMORPH Corrected
show systematic underestimation.

The middle panel of Fig. 5 shows how RainyDay can be used to
examine the effect of rain gage-based bias correction on satellite-
based IDF estimates. In the case of CMORPH, the Raw version
overestimates rainfall intensity at all $p_e$, while results for the Cor-
corrected version shows that the daily-scale bias correction scheme
leads to systematic underestimation. The TMPA-RT also over-
estimates at all $p_e$, though not as severely as CMORPH Raw, while
the monthly bias correction scheme used in the final version of
TMPA appears to offer superior performance to the daily-scale
routine used by CMORPH Corrected. It is not immediately clear
why this is the case, but relevant considerations include the effect of
rainfall detection errors on bias correction (Tian et al., 2007) and
the challenge of correcting for conditional biases at short time
scales (Wright et al., 2014c). The apparent strong performance of
the monthly bias correction is encouraging in the context of Inte-
grated Multi-satellite Retrievals for GPM (IMERG), a state-of-the-
art rainfall dataset that combines various elements from TMPA,
CMORPH, and PERSIANN, including TMPA’s monthly bias correction
(Huffman et al., 2014). The IMERG dataset is not analyzed in this
study since the full retrospective dataset is not yet available.

The bottom panel of Fig. 5 shows results similar to those in the
top panel, but with $A$ set to a $0.5$ by $0.5$ (approximately 2500 km$^2$)
box centered on Iowa City. The results demonstrate that RainyDay
can easily generate spatially aggregated rainfall IDF curves. This is
not achievable using standard gage-based IDF curves without the use of ARFs, which, as previously mentioned, have been shown to have limitations. We omit gage-based IDF curves from the bottom panel of Fig. 5 for this reason.

The results shown in Figs. 4 and 5 have implications for using RainyDay for IDF and hazard estimation in data-sparse regions using satellite remote sensing. First, there can be substantial differences in extreme rainfall estimates between satellite rainfall datasets, and these differences will propagate through to IDF estimates (and to probabilistic hazard estimates, as will be shown in Section 4.2). Furthermore, while comparison with gage-based IDFs (when available) can be used to understand these differences, spatial sampling mismatches complicate comparisons. Findings may not be transferable across regions since the performance of satellite rainfall retrievals varies with region and latitude (e.g. Ebert et al., 2007) and because the quality of the gage-based bias correction schemes that some of satellite datasets employ will vary regionally with the density of rain gage observations that are available.

4.2. Flood frequency analysis

In this section, we present flood peak frequency analyses for the 4400 km² Turkey River watershed in northeastern Iowa using rainfall scenarios from RainyDay as inputs to the Iowa Flood Center (IFC) Model, a calibration-free distributed hydrologic modeling framework designed primarily for multi-scale flood research and applications (see Cunha et al., 2012; Demir and Krajewski, 2013; Mantilla and Gupta, 2005; Moser et al., 2015; Small et al., 2013). Moser et al. (2015) provides a detailed model description and Cunha et al. (2012) performed validation for flood events in Iowa, showing that performance of the IFC Model is generally comparable to that of the more heavily-calibrated operational SAC-SMA flood forecast model (Burnash, 1995). This study aims only to demonstrate basic features of RainyDay for flood hazard analysis and so does not provide detailed discussion of the IFC Model or comparisons with other available platforms. For a discussion of the value of calibration-free, distributed hydrologic models for multi-scale flood modeling, the reader is directed to Wright et al. (2014b) and, in particular, Cunha et al. (2012). The full multi-scale hazard estimation capabilities of SST and RainyDay can, in principle, be harnessed using any distributed hydrologic or mass wasting model, while some of the capabilities can be achieved through the use of lumped models.

A limited set of model hydrograph validation is provided in Fig. 6 for the 2008 and 2014 April–July periods, during which major flooding occurred throughout Iowa (see Smith et al., 2013 for a detailed examination of the hydrometeorology of the 2008 floods). The model is run both with Stage IV and gage-corrected TMPA. Comparisons with U.S. Geological Survey (USGS) stream gage observations are provided at four locations, with upstream drainage areas ranging from 900 to 4000 km². All hydrographs are normal observations are provided at four locations, with upstream drainage areas ranging from 900 to 4000 km². All hydrographs are normal.
the USGS StreamStats system. The first is developed using standardized methods described in Bulletin 17B (Interagency Advisory Committee on Water Data, 1982) using the log-Pearson Type III distribution (henceforth referred to as the LP3 distribution) with a regionalized skew coefficient. The second is based on regional regression equations that consider drainage basin area and shape as well as soil properties. Eash et al. (2013) report 121 years of data for Turkey River at Garber (4002 km²), near Eldorado (1660 km²), and above French Hollow (2338 km²), while 63 years are reported for Volga River at Littleport (901 km²) and 45 years for Turkey River at Spillville (458 km²). It should be noted that these record lengths refer to “historic record length” described in Section V.B.10 of Bulletin 17B and do not correspond to length of the USGS annual maxima streamflow timeseries available on the USGS National Water Information System (http://nwis.waterdata.usgs.gov/nwis), which are much shorter. All available USGS streamflow observations for the five sites are also shown, where \( p_e \) is estimated using the Cunnane plotting position (Cunnane, 1978; \( p_e = \frac{i - 0.4}{X + 0.2} \), where \( i \) is the rank of the observation and \( X \) is the number of observations). Other common plotting position formulae produce similar results (not shown) and do not alter subsequent findings.

For all five locations shown in Fig. 8, the SST-based peak discharge estimates using TMPA are higher than those using Stage IV for \( p_e < 0.01 \), generally converging toward the Stage IV estimates as \( p_e \) decreases, and in some cases yielding lower estimates for \( p_e \) less than about 0.005. This is consistent with the rainfall IDF results from RainyDay shown in Fig. 4 and are suggestive of conditional biases in the TMPA dataset. This in indeed confirmed in Fig. 9, which shows watershed-specific IDF curves for the entire Turkey River watershed from RainyDay using TMPA and Stage IV. The USGS streamflow observations shown in Fig. 8 agree reasonably well with the Stage IV-based estimates for \( p_e > 0.5 \), with the exception of the smallest subwatershed, Turkey River at Spillville, where Stage IV produces low peak estimates. For \( p_e < 0.5 \), there is a lack of consistency. For example, Turkey River at Garber shows higher estimates from Stage IV than the streamflow observations, while the reverse is true for Turkey River at French Hollow and near Eldorado. Deviations from the USGS observations do not show a systematic scale dependency.

Both RainyDay-based frequency analyses and the USGS streamflow observations are generally higher than the USGS frequency analyses for \( p_e \) less than about 0.2. One exception is the set of USGS observations for Turkey River at Spillville, which is lower than both the RainyDay estimates and the regional regression but generally consistent with the Bulletin 17B analysis. The regional regression results for Turkey River at Spillville are greater than the USGS regionalized LP3 estimates, while the reverse is true for the four larger subwatersheds. Interestingly, some of the USGS observations fall outside of the 90% confidence intervals of the LP3 analyses for Turkey River near Eldorado, Volga River at Littleport, and Turkey River at Garber. In the case of the latter station, the five most intense floods are near or above the upper 95% confidence bound, a finding that is explored in more detail in the following paragraphs.

It should be noted that with the exception of Turkey River at Garber, the differences between the RainyDay-based frequency
analyses are roughly similar in magnitude to the differences between the two USGS approaches. This, along with the underestimation shown by USGS frequency analyses relative to the USGS peak discharge observations at several sites, suggests that the RainyDay-based frequency analyses should not be dismissed out of hand as being too high for low $p_e$. In fact, as the next example shows, there is observational evidence that supports the validity of the RainyDay-based results in light of possible nonstationarity in flooding. It should be noted that discharge-based frequency analyses, even in stationary situations with long records, are not necessarily superior to hydrologic modeling methods. Analyses by Smith et al. (2013) suggest that peak discharge measurement errors may be substantial for a recent major flood event in Iowa. The propagation of discharge measurement errors through frequency analysis is poorly understood (e.g., Petersen-Øverleir and Reitan, 2009; Petersen- Øverleir, 2004; Potter and Walker, 1985). Rogger et al. (2012) reported significant differences between two commonly-used flood frequency analysis approaches for ten small alpine watersheds in Austria, one based on a stream gage-based statistical method and the other on design storm methods combined with a hydrologic model. The latter method produced higher discharge values than the former, and the authors discuss possible explanations and deficiencies in both approaches, concluding that hydrologic modeling using rainfall inputs can produce superior results in certain situations.

Of the five USGS stream gage locations shown in Fig. 8, only the gage at Garber, Iowa has a long (82-year), unbroken annual peak discharge record. We use this record to better understand the discrepancies between the RainyDay-based results and the USGS frequency analyses from Eash et al. (2013), and in particular to contrast the methods in the context of potential nonstationarity in flood processes. The top panel of Fig. 10 shows the same results as Fig. 8 for Turkey River at Garber, except that the USGS observations have been divided into two groups: one for all peaks occurring from 1933 to 1989, and the second for all peaks occurring from 1990 to 2014. The plotting position-based $p_e$ is recalculated for each group. The 1933–1989 group shows higher discharges than either RainyDay Stage IV or USGS discharges for $p_e > 0.5$, and lower discharges for $p_e$ less than about 0.2. The 1990–2014 group, meanwhile, matches closely with the RainyDay-based frequency analyses with Stage IV.

Taken together, this suggests a regime shift toward more extreme flooding since 1990 accompanied by a reduction in the magnitude of more average floods. Evidence of this regime shift can be seen in the annual peak time series in the bottom panel of Fig. 10. We fit a nonparametric linear regression to the 1933–2014 time series using the Theil-Sen estimator (Sen, 1968) and a statistically significant ($p$-value < 0.05) downward trend was found. In contrast, using ordinary least squares, an insignificant upward trend is found over the same period. Thus when the influence of the most extreme values is minimized through nonparametric methods, there is a tendency toward smaller flood peaks over time that is not evident with parametric methods, which are more sensitive to the recent extremes.

Fig. 10 shows that the period of apparent elevated flood activity is well captured by RainyDay, while the preceding period is not, presumably because the IFC model reflects recent land use changes and because the input rainfall data are relatively recent. In general, whether or not this constitutes a strength or limitation of RainyDay depends on the underlying causation of nonstationary flood activity. If nonstationarity results from a climate-driven secular trend in extreme rainfall, then the results from RainyDay using relatively short and recent rainfall remote sensing records should be understood as more “up-to-date” estimates of flood frequency compared to approaches, such as the USGS analyses, that use longer stream gage or rain gage records. The same is true if there is a secular trend in flooding due to urbanization or other land-use changes, so long as these changes are properly incorporated into the hydrologic
Fig. 8. Peak discharge analyses using RainyDay with Stage IV and TMPA rainfall remote sensing data and the IFC Model, compared against USGS stream gage-based analyses for five subwatersheds of the Turkey River in northeastern Iowa. Shaded areas for RainyDay estimates denote the ensemble spread. Bars on the USGS Bulletin 17B estimates denote the 90% confidence intervals. Confidence intervals are not available for the USGS regional regression. Key RainyDay parameters: $m = 150$ storms, $A' = [40°$ to $44°$ N, $90°$ to $96°$ W], $A$ is the watershed upstream of the USGS streamgage at Garber, IA, $N = 10$, $T_{max} = 500$, $t = 96$ h. Spatially-uniform transposition and Poisson-based temporal resampling are selected. Stage IV period of record is 2002–2014, TMPA period of record is 1998–2014. RainyDay Analyses are restricted to April–November period.
model. In the case of Iowa, flooding has been shown to be affected by land-use change (Villarini and Strong, 2014) and by climate change (Mallakpour and Villarini, 2015). If, on the other hand, flood or rainfall nonstationarity has a periodic structure due to a slowly-varying climate mode, then the results from SST may adequately reflect the true flood frequency only for the phase of the mode that overlaps with the remote sensing record. It should also be recognized that a period of higher or lower flood activity at a particular location could result from pure randomness (i.e. in absence of both secular and periodic trends). SST should be relatively robust to this possibility through the sampling storms from a larger region.

4.3. SST sensitivity to record length and user-defined parameters

In this section, we examine the sensitivity of SST to the length of the input dataset and to the different user-defined parameters and options introduced in Sections 2 and 3. Specific topics that are examined include the optional non-uniform spatial transposition (Section 3.3), empirically-based temporal resampling (Section 3.4) and the size of the transposition domain $A'$. In all cases, the specific results pertain to the Iowa study area and may not be entirely generalizable to other locations. The intention is to demonstrate some important concepts and pitfalls associated with RainyDay, and provide a possible framework for assessing performance in different locations and applications.

A common critique of coupling SST with rainfall remote sensing datasets is that such data records are relatively short (approximately 10–20 years at time of writing) and thus may not contain sufficient numbers of extreme events at the regional scale to leverage “space-for-time substitution” to accurately recreate the properties of rare rainfall events. To examine this critique, we turn to a longer dataset: CPC-Unified, a daily rain gage-based gridded rainfall dataset that has a spatial resolution of 0.25° over the conterminous United States (Chen et al., 2008; Xie et al., 2007). Though the spatial and temporal resolution of CPC-Unified is generally insufficient for fine-scale flood modeling, its long record (1948 to present) makes it ideal for evaluating the sensitivity of SST-based IDF estimates to record length. We examined several stationarity measures over the transposition domain $A'$ (which, as in Section 4.1, roughly encompasses the state of Iowa), including the average number of storm counts per year and the mean, median, and standard deviation of storm rainfall depth. None of these measures revealed significant temporal trends (results not shown). This may contradict the apparent flood nonstationarity in the Turkey River watershed discussed in Section 4.2, or may point to land-use change as the predominant source of non-stationarity in Turkey River, but rigorous examination is beyond the scope of this

---

**Fig. 9.** IDF analyses for Turkey River using RainyDay with Stage IV and TMPA rainfall remote sensing data. Shaded areas for RainyDay estimates denote the ensemble spread. Key RainyDay parameters: $m = 150$ storms, $A' = [40°$ to $44° N, 90°$ to $96° W], A$ is the 4400 km² watershed upstream of the confluence with the Mississippi River. $N = 100$, $T_{max} = 500$, $t = 96$ h, and spatially-uniform transposition and Poisson-based temporal resampling are selected. Stage IV period of record is 2002–2014, TMPA period of record is 1998–2014. Analyses are restricted to April–November period.

**Fig. 10.** Top panel—four peak discharge analyses for the location of the USGS stream gage at Garber, IA: RainyDay with Stage IV and TMPA rainfall and USGS frequency analyses using regional regression relationships and Bulletin 17B methods. Shaded areas for RainyDay estimates denote the ensemble spread. Bars for the Bulletin 17B-based analysis denote the 90% confidence intervals. Confidence intervals are not available for the USGS regional regression. Bottom panel—annual peak discharge time series for 1932–2014 for the Garber gage. Linear trend lines in the bottom panel use non-parametric Thiel-Sen regression (Sen, 1968) and ordinary least squares (OLS). Key RainyDay parameters: $m = 150$ storms, $A' = [40°$ to $44° N, 90°$ to $96° W], A$ is the watershed upstream of the USGS stream gage at Garber, IA, $N = 10$, $T_{max} = 500$, $t = 96$ h. Spatially-uniform transposition and Poisson-based temporal resampling are selected. Stage IV period of record is 2002–2014, TMPA period of record is 1998–2014. RainyDay Analyses are restricted to April–November period.
paper.

We use a bootstrapping approach to examine variability in IDF estimates derived from the CPC-Unified data using RainyDay and how this variability evolves as the length of the record increases. All IDF estimates in this section are for 1-day rainfall over averaged over a 0.5° by 0.5° box. We generate n-year long input rainfall datasets by randomly selecting n years of CPC-Unified data without replacement from the 1948–2014 period. Each of these datasets is then used as the basis for a single run of RainyDay with 100 ensemble members and with \( m = 10n \) (leading to \( k = 10 \) storms per year). We repeat this procedure to create 25 datasets for each value of \( n = 10, 20, 30, 40, 50 \) years.

Greater variability is evident in the ensemble mean and spread of the IDF estimates using 10 years of CPC-Unified data than using 20 years, while change in variability is generally small between runs using 20 years and 30 years of data (Fig. 11). We also examined the variability of relative deviations in the ensemble IDF means, minima, and maxima from RainyDay between the n-year runs and IDFs based on the full 67-year dataset (Fig. 12). The boxplots show that most deviations in the n-year IDF ensemble means, minima, and maxima are less than 10% and that the vast majority are less than 20% for any given \( n \). For most \( n \), there are substantial reductions in deviation when the records increase in length from \( n = 10 \) to \( n = 20 \) years. The reductions in deviation are less when the record length increases beyond 20 years.

Unless the intensity of the rainfall inputs is perturbed stochastically, SST-based frequency analyses have an upper bound that corresponds to the most intense rainstorm in the storm catalog transposed in such a way that rainfall over A is maximized. The lack of positive deviations in the ensemble maxima at \( p_e = 10^{-3} \) (middle panel of Fig. 12; also in certain realizations shown in Fig. 11) show where the SST procedure “encounters” this upper limit.

While the results in this section are by no means exhaustive and the conclusions are specific to the Iowa study region and could vary in different physiographic regions, they nonetheless suggest that concerns over the use of relatively short rainfall remote sensing records with SST may be overstated and that such datasets, many of which are approaching 20 years in length, should provide relatively robust estimates that will improve as these datasets continue to grow in length. This emphasizes the fact that rainfall events that would be considered rare from the perspective of a single location or watershed can occur relatively frequently from a regional perspective. This is qualitatively consistent with the findings of Troutman and Karlinger (2003), who estimate that a flood with \( p_e > 10^{-2} \) occurs on average every 4.5 years at least one of the 193 USGS stream gage sites in their Puget Sound study region.

A potentially important issue related to short data records in SST, previously mentioned in Section 3.3, can arise if, instead of assuming that the probability of storm occurrence is uniform across the transposition domain, non-uniform spatial transposition is used instead (such as the approach used in Wilson and Foufoula-Georgiou, 1990 or the optional scheme in RainyDay described in Section 3.3). Using the bootstrapping approach with the CPC-Unified dataset described above, visual inspection of storm probability-of-occurrence maps such as the one shown in Fig. 3 reveal that there can be substantial variations in the spatial distribution of historical storms when rainfall records are short (results not shown). These variations tend to diminish as the length of record increases, as do their impacts on IDF estimates. More variation is evident in the median IDFs from independent runs of RainyDay, for example, using non-uniform transposition than using uniform transposition when \( n = 10 \) years (Fig. 13, left panels). When using non-uniform transposition, variability diminishes when \( n = 20 \) years and a systematic increase in rainfall intensity for \( p_e > 0.02 \), relative to the uniform transposition case, emerges (Fig. 13, right panels). Given these results, we recommend that the assumption of uniform transposition be used in the absence of strong physically-based reasoning and observational support for non-uniform transposition. It is possible, however, that this explains the IDF underestimation by RainyDay with Stage IV for high \( p_e \) relative to Atlas 14 shown in Fig. 4, where uniform spatial transposition was used.

As mentioned previously, RainyDay supports either the Poisson-based resampling that has traditionally been used with SST, or an empirical scheme described in Section 3.4. There do not appear to be substantial systematic differences between the results from RainyDay using these two schemes with 10-year records (Fig. 14, left panels), but like Fig. 13, when 20-year records are used, there is a tendency toward higher rainfall estimates for \( p_e > 0.02 \). Results may differ in other regions where temporal clustering of storms is very strong or where rainstorms are very infrequent. It is recommended that the modeler assess clustering using an independent long-term rainfall data source if available, in addition to assessing sensitivity to this option in RainyDay. As with the spatial transposition schemes, the choice of temporal resampling scheme does.
not appear to have a substantial impact on low $p_e$ estimates.

We also examine the sensitivity of RainyDay results to the size of $A'$ (Fig. 15). To do so, we run RainyDay for various square domains ranging from $1^\circ$ by $1^\circ$ up to $10^\circ$ by $10^\circ$, while holding $A$ fixed at a $0.5^\circ$ by $0.5^\circ$ box. Then the evolution of rainfall intensity is examined for a range of $p_e$ as a function of $A$. This is repeated for a several different record lengths and for two values of $\lambda$. Interestingly, while there is a general tendency for intensity estimates to stabilize as $A'$ grows, the behavior is not asymptotic (though roughly so for $n = 68$ years). The high exceedance probability estimates ($p_e = 0.5$) tend to be stable over a large range of $A'$ and then decrease for very large values, due to the tendency for synthetic years to be created in which no storm is transposed directly over $A$. This is the root of potential low biases mentioned in Step 2 of the SST procedure described in Section 2. However, Fig. 15 demonstrates that this tendency for a decrease in intensity estimates for large $A'$ extends to smaller $p_e$ values as well, and that there is a critical value of $A'$ at which the estimated intensity is roughly maximized. This critical value appears to vary more by the particular period of record than by the length of record. For example, the 20-year record from 1976 to 1995 yielded a critical value of $A'$ that is lower than the critical value from 20-year record from 1996 to 2015. This points to the fact that the existence and number of major storms within $A'$ during the record period is very important (Wright et al., 2014b reached the same conclusion through different means).

These results also indicate that increasing $m$ (thus increasing $\lambda$) can mitigate the reduction in estimated intensity for values of $A'$ larger than the critical value. This result suggests that, if the modeler is interested in hazard estimation across a range of $p_e$, she should choose a relatively large $m$. A diagnostic framework within the RainyDay software to identify this critical value of $A'$ for a given value of $m$ (or vice versa) for different $p_e$ would be useful but does not currently exist.

5. Discussion and conclusions

We introduce RainyDay, a Python-based platform that couples rainfall remote sensing data with a technique known as Stochastic Storm Transposition (SST) that effectively “lengthens” the extreme rainfall record through temporal resampling and spatial transposition of observed rainstorms. It produces probabilistic extreme rainfall scenarios that include realistic estimates of rainfall duration, intensity, and space-time structure that can be used for probabilistic flood and landslide hazard and risk assessment at a wide range of scales.

The SST technique implemented in RainyDay has two important features that distinguish it from IDF and design storm methods for describing the relationships between the intensity, duration, and structure of extreme rainfall. First, it leverages the detailed picture of rainfall space-time structure offered by ground-based radar or satellite-based sensors. This structure can play an important role in landslides and floods because the variability in the concentration and intermittency of extreme rainfall in space and time can lead to a complex and diverse spectrum of hazard response. This structure is difficult to measure using rain gages due to the high gage densities and sampling rates required, and so rain gage-based methods for analysis of rainfall-driven hazards, such as IDF relations and design storm methods, typically neglect this higher-order variability. The reader is directed to Wright et al. (2014b) for a deeper examination of this feature of SST in the context of urban flood
The second important feature of RainyDay is that, because of the near-global coverage of satellite rainfall datasets, it is possible to generate realistic representations of extreme rainfall in remote or poorly-instrumented regions where rain gage or stream gage records are lacking. Such regions are common even in wealthy nations and are ubiquitous in developing countries, many of which are characterized by rapidly-growing exposure to rainfall-driven hazards due to urbanization and climate change. The authors are not aware of other approaches that offer the ability to generate realistic rainfall inputs for probabilistic hazard modeling nearly anywhere on the globe with minimal computational effort.

Despite the advantages that SST and RainyDay offer over other methods for assessing rainfall-driven hazards (e.g. design storms, discharge frequency analysis), a number of issues remain. Perhaps the biggest limitation to coupling SST with rainfall remote sensing is the uncertain accuracy of the input rainfall data. Significant efforts have been made to better understand and minimize the errors in remote sensing estimates of rainfall, both from satellites (e.g. Petty and Krajewski, 1996; Tian and Peters-Lidard, 2007; Tian et al., 2009) and from ground-based radar (e.g. Villarini and Krajewski, 2010). Such studies demonstrate that remote sensing estimates can vary significantly from reference observations in terms of rainfall intensity and differentiation between rainy and non-rainy areas, with important implications for hazard applications. In the case of satellite-based rainfall estimates, heterogeneities in the underlying land or water surfaces can be difficult to distinguish from variations in cloud and rainfall properties (e.g. Ferraro et al., 2013), while both ground-based radar and space-based sensors tend to suffer in mountainous areas due to dramatic variations in rainfall physical properties over short time and length scales. Furthermore, the spatial and temporal resolution of remote sensing estimates, particularly from satellites, can be too coarse for modeling at very small scales, especially in urban areas and fast-responding mountain or desert catchments where surface runoff generation from intense, short-duration rainfall on sub-hourly, sub-kilometer scales can be a key driver of hazards. The uncertainties associated with rainfall remote sensing data pose serious challenges for flood or landslide forecasting and monitoring, which require accurate rainfall estimates in real-time. These issues may be somewhat less critical in the SST framework or in long-term hazard assessment more generally, since the rainfall estimates need only have fidelity in the statistical sense. SST will be somewhat robust to random errors in rainfall data, as the underestimation of rainfall intensity from some storms in the storm catalog can be compensated by overestimation of rainfall intensity from others. In contrast, SST is not robust to systematic rainfall biases, as demonstrated in several examples in this paper. IMERG, NASA’s newest satellite multi-sensor dataset, will feature improved accuracy and relatively high resolution (0.1°, 30-min), thus addressing some of these issues once the full retrospective dataset becomes available.

In the case of flood hazard modeling using SST, a practical upper limit on the size of the area of interest A can arise. The sizes of A and A’ can be limited due to the challenges posed by transposition in the presence of complex terrain features. Furthermore, as A becomes
larger, the rainfall duration $t$ needed to properly model hazard response becomes longer. While RainyDay does not restrict the choice of $t$, practical limitations exist. In large watersheds, floods are usually the result of specific space-time arrangements of multiple distinct storm systems over the span of perhaps a week to several months, often linked to persistent large-scale atmospheric phenomena. One could specify a long $t$ (a month, for example) in RainyDay to “capture” multiple storm systems within a single storm catalog entry. Such long $t$, however, means there could only be relatively few entries in the storm catalog, given the limited record length of the input dataset. Such an approach would be constrained by the few space-time configurations of these storm systems that were observed, while many other non-observed configurations are hypothetically possible. A tradeoff thus emerges as $A$ (and thus $t$) increases relative to the area of the transposition domain $A'$. If $A$ is a large fraction of $A'$, then there is little opportunity to leverage the “space-for-time” substitution that is at the core of the SST approach. If the user instead decides to increase the size of $A'$, she must ensure that this transposition is performed in a realistic manner. This effectively precludes modeling of regions that approach continental scales. The maximum scale at which SST can be feasibly used is an open question with no simple answer. It should be noted that IDF and design storm methods face similar and perhaps even more acute limitations in terms of an upper area limit, though for different reasons (e.g. conceptual and practical shortcomings of point-based IDF, temporal rainfall distributions, and area reduction factors).

As mentioned in Section 4.3, a common critique of the methodology presented in this study is that the relatively short remote sensing records may not contain enough truly extreme rainfall events. Sensitivity to record length is not unique to SST; frequency estimates of rare hazards will be driven by the largest several events in the historical record, regardless of the chosen analysis technique. The results in Section 4.3 demonstrate that this concern may be somewhat exaggerated in the case of SST since very extreme rainfall events that are considered rare from a local viewpoint can occur much more frequently when viewed regionally. Like more commonly-used regionalization techniques, SST leverages this fact to improve hazard analysis. As the rainfall remote sensing record grows, the robustness of estimates produced by SST and RainyDay should increase as additional extreme storms are observed (and as their accuracy improves due to technological advances). Estimates of rainfall intensity will improve more per unit of additional observational period using SST than using point-based techniques due to SST’s regional nature, while new patterns of rainfall space-time structure will add to the realism of SST-based flood and landslide hazard estimates since a broader spectrum of hazard outcomes will be possible. RainyDay makes such updating simple, while IDF databases and design storm methods are generally updated through slow and costly procedures (Y. Zhang, personal communication, May 14, 2015).

As highlighted in Section 4.2, SST and RainyDay have important features in the context of nonstationary hazards. Extreme rainfall scenarios from RainyDay are generally based on more recent

Fig. 14. The effect of the temporal resampling scheme on daily rainfall IDF curves estimated using RainyDay with the CPC-Unified daily rainfall over Iowa, United States. Each panel shows the ensemble mean (solid lines) for ten independent runs of RainyDay. The shaded areas denote the maximum spread across the ten runs. The specific years that comprise the input dataset vary. Key RainyDay parameters: $m = 10n$ storms (where $n$ varies by specified record length), $A' = [40° to 44° N, 90° to 96° W]$, $A$ is a 0.5° by 0.5° box, $N = 100$, $T_{max} = 1000$, $t = 1$ day. Spatially uniform transposition is used. Analyses are restricted to April–November period.
Fig. 15. The effect of the size of the transposition domain $A'$ on daily rainfall IDF curves estimated using RainyDay with the CPC-Unified daily rainfall over Iowa, United States using a range of record lengths. Key RainyDay parameters: $m = 10$ storms (where $n$ varies by specified record length), $A'$ is a square of varying size, $A$ is a 0.5° by 0.5° box, $N = 100$, $T_{\text{max}} = 1000$, $t = 1$ day, spatially-uniform transposition and Poisson-based temporal resampling. Analyses are restricted to April–November period.
observations than existing rain gage or stream gage-based frequency analyses such as Atlas 14 IDF relations, which contain older records that may not be representative of the current state of the climate. In this respect, hazard analyses based on RainyDay can be understood as relatively current “snapshots” based on recent climate. The performance of RainyDay is very dependent on major storms having occurred one or more times within the transposition domain, however, meaning that spatial transposition is not a perfect remedy for short data records. Furthermore, if the rainfall remote sensing record deviates significantly from the true long-term properties of extreme rainfall over the region of interest due to random chance, decadal-scale climate variability, or systematic measurement bias, then caution must be taken when using RainyDay. It can be challenging in practice to diagnose such non-stationarities and biases due to a lack of long-term independent observational data, particularly in remote or underdeveloped regions. Meanwhile, as discussed in Wright et al. (2014b), combining SST (or other rainfall-based approaches, e.g. Cunha et al., 2011) with a distributed hazard model allows the analyst to incorporate changes in land use and land cover into non-stationary hazard estimates.

Acknowledgments

This work was made possible through the fellowship support of the NASA Postdoctoral Program, administered by Oak Ridge Associated Universities, Oak Ridge, Tennessee. We also acknowledge the support of the University of Wisconsin-Madison, the Wisconsin Alumni Research Foundation, and the Iowa Flood Center and the University of Iowa. We would also like to thank Scott Small, Chi Chi Choi, and Tibeju Ajayew at the University of Iowa for their support in configuring and troubleshootiing the IFC Model. Computing resources supporting the hydrologic modeling were provided by the NASA High-End Computing Program through the NASA Center for Climate Simulation at Goddard Space Flight Center. We would also like to thank the editor and the two anonymous reviewers whose constructive criticisms contributed greatly to the study.

References


