

RESEARCH ARTICLE

Probabilistic rainfall generator for tropical cyclones affecting Louisiana

Gabriele Villarini¹  | Wei Zhang^{1,2} | Paul Miller³  | David R. Johnson⁴ |
Lauren E. Grimley^{5,6} | Hugh J. Roberts⁵

¹IIHR—Hydroscience & Engineering, University of Iowa, Iowa City, Iowa, USA

²Department of Plants, Soils and Climate, Utah State University, Logan, Utah, USA

³Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, Louisiana, USA

⁴School of Industrial Engineering and Department of Political Science, Purdue University, West Lafayette, Indiana, USA

⁵The Water Institute of the Gulf, Baton Rouge, Louisiana, USA

⁶Department of Geological Sciences, University of North Carolina, Chapel Hill, North Carolina, USA

Correspondence

Gabriele Villarini, IIHR—Hydroscience & Engineering, The University of Iowa, 107C C. Maxwell Stanley Hydraulics Laboratory, Iowa City, IA 52242, USA.
Email: gabriele-villarini@uiowa.edu

Funding information

Louisiana Watershed Initiative, Grant/Award Number: N/A

Abstract

This study focuses on the development of a probabilistic rainfall generator for tropical cyclones (TCs) affecting Louisiana. We consider 12 storms making landfall along the Louisiana coast during 2002–2017 and generate ensembles of high-resolution (~5 km and 20 min) TC-rainfall fields for each storm. We develop a data-driven multiplicative model, relating observed rainfall to the rainfall obtained from a parametric TC rainfall model (Interagency Performance Evaluation Task Force Rainfall Analysis [IPET]) through the product of a deterministic and a stochastic component; the former accounts for rain-dependent biases, while the latter for the stochastic nature of the rainfall processes. As a preliminary step, we describe the overall bias of the IPET model as a function of total TC rainfall within the state and maximum wind speed at landfall. We then estimate the rain-dependent bias using a cubic spline. Finally, we characterize the random errors in terms of their probability distribution and spatial correlation. We show that the marginal distribution of the logarithm of the random errors can be described by a mixture of four Gaussian distributions, and its spatial correlation is estimated based on the nonparametric Kendall's τ . We then present a methodology to generate ensembles of random fields with the specified statistical properties. Here, the generation of probabilistic rainfall comes from the statistical modelling of the uncertainties between IPET rainfall and observations. While these results are valid for Louisiana and the IPET model, the methodology can be generalized to other parametric rainfall models and regions, and it represents a viable tool to improve our quantification of the risk associated with TC rainfall.

KEYWORDS

IPET, Louisiana, rainfall, rainfall generator, statistical modelling, tropical cyclones

1 | INTRODUCTION

Rainfall associated with landfalling tropical cyclones (TCs) represents a key ingredient for flooding and landslides (e.g., Bucknam *et al.*, 2001; Fuhrmann *et al.*, 2008; Villarini *et al.*, 2014a; Aryal *et al.*, 2018), leading to major societal and economic repercussions (e.g., Rappaport, 2014; Czajkowski *et al.*, 2017). These impacts are felt especially along the coast, where an increased concentration of population and wealth has led to increased vulnerabilities to these storms (e.g., Klotzbach *et al.*, 2018). Images of the extreme TC rainfall and flooding paint a tragic reality, and highlight the severity of impacts on communities (e.g., Miniussi *et al.*, 2020).

Louisiana is an area of the country significantly affected by these storms. Hurricane Katrina (2005) had an indelible influence on Louisiana's appreciation of TC impacts. Though Hurricane Katrina illustrated severe TC wind and storm surge impacts in the state and along the Gulf coast, recent years have witnessed additional severe TC events for which flood risk has been due in large part to TC-related rainfall. During the historic 2020 North Atlantic hurricane season, Tropical Storm Cristobal and Hurricanes Laura, Delta, and Zeta generated maximum storm total precipitation values of 16, 26.4, 25.9, and 18.7 cm over Louisiana, respectively. Additionally, Hurricane Sally, which made landfall as a Category-2 hurricane near Gulf Shores (Alabama) produced a staggering 76.2 cm of rain along the Alabama coastline. The year prior, Hurricane Barry (2019) approached Louisiana accompanied by grim forecasts of extreme precipitation over heavily populated southeastern Louisiana, though fortunately only 10–20 cm were ultimately recorded in the New Orleans and Baton Rouge metro areas (Cangialosi *et al.*, 2019). Consequently, Louisiana state authorities are increasingly interested in the superimposed, nonlinear effects of coincident TC storm surge and hydrological flooding, termed transition zone flooding (Bilskie and Hagen, 2018). While the methodology for simulating peak surge inundation associated with a prescribed TC path, size, and intensity is well established (e.g., Luettich and Westerink, 2004; Resio, 2007; Nadal-Caraballo *et al.*, 2015) the best approach for pairing the hypothetical TC with a representative precipitation field is less clear.

Several different parametric approaches have been proposed and developed to characterize the rainfall associated with TCs. The primary advantage of these rainfall models lies in their ability to produce a smooth, temporally varying precipitation surface using basic TC diagnostic variables (e.g., central pressure deficit, radius of maximum winds [RMW]) either directly recorded or easily derived from historical datasets such as HURDAT2 (Landsea and Franklin, 2013), Extended HURDAT data (Demuth *et al.*, 2006), and the International Best Track

Archive for Climate Stewardship (IBTrACS; Knapp *et al.*, 2010). For instance, the Rainfall Climatology and Persistence (R-CLIPER) model (Marks and DeMaria, 2003; Tuleya *et al.*, 2007) only requires the position of the storm and the corresponding maximum wind speed; however, R-CLIPER primitively assumes that the rainfall fields are azimuthally symmetric, ignoring the effects of vertical wind shear on TC precipitation.

Improving on this limitation, Lonfat *et al.* (2007) introduced the Parametric Hurricane Rainfall Model (PHRaM) which utilized R-CLIPER as the initial TC rainfall field, but then spatially redistributes the R-CLIPER estimate via two additional terms that account for the asymmetrical effects of shear and topography. Lastly, the Interagency Performance Evaluation Task Force Rainfall Analysis (IPET, 2006), based on Lonfat *et al.* (2004) and Chen *et al.* (2006), adopted elements of both aforementioned schemes, whereby a simple parametric model produces a symmetric precipitation surface; the asymmetrical effect of vertical wind shear is then crudely incorporated by multiplying all rainfall rates to the right of the direction of motion by 1.5 \times . Lastly, the PDF Precipitation-Climatology and Persistence (P-CLIPER) resembles R-CLIPER, but improves upon it by adding the frequency parameter, f , ranging from -90 to 90% which can be perturbed to reflect a deviation from the mean TC rainfall intensity (Geoghegan *et al.*, 2018). Brackins and Kalyanapu (2020) provided a detailed description, extensive review, and thorough evaluation of the R-CLIPER, IPET, and PHRaM models as well as seven configurations of P-CLIPER (i.e., f values of -90 , -60 , -30 , 0 , $+30$, $+60$, and $+90$), ultimately concluding that IPET tends to work the best in terms of storm total rainfall above 75 mm. However, Brackins and Kalyanapu (2020) also warned that the performance of these parametric rainfall models is currently insufficient for flood studies.

Another group of models is based on the characterization of the physical processes at play. For instance, Langousis and Veneziano (2009) developed an extension of the formulation by Smith (1968). Their model included a physically based component to describe the relationship between mean rainfall and some of the storms' characteristics (e.g., RMW, translation velocity), and a statistical part to describe the rainfall variability. Based on its configuration and the fact that it does not account for landfall effects, this model is best suited for open-water. Some of the shortcomings in Langousis and Veneziano (2009) were addressed by Lu *et al.* (2018), in which the authors developed a physics-based model capable of representing TC rainfall over the oceans and land (see also Zhu *et al.*, 2013).

All these approaches provide valuable contributions towards improving our characterization of this hazard. Even though these methodologies have different levels of complexity and sophistication, one thing that they

generally have in common is an attempt to develop the “best” or “most likely” representation of the rainfall field associated with these storms and are deterministic in nature. The development of probabilistic models of rainfall associated with these storms has received much less attention in the literature. The generation of probabilistic rainfall generally comes from the uncertainties in the TC track and its characteristics (e.g., Jing and Lin, 2020; Ko *et al.*, 2020; Yussouf *et al.*, 2020), with only few exceptions focused on the probabilistic modelling of TC rainfall (Langousis and Veneziano, 2009). For instance, Langousis and Veneziano (2009) proposed an approach in which the mean rainfall field associated with these storms is obtained using a physically based model, with the stochastic component that describes the rainfall variability at both large (i.e., among storms) and small (e.g., rainbands, local convection) scales.

Here we take a different point of view: rather than trying to develop a model with smaller and smaller uncertainties, our aim is to embrace these errors and characterize them explicitly. When incorporating predicted rainfall into design of protection systems, development of flood risk maps, or early-warning systems, it is critical to acknowledge and attempt to quantify uncertainty. As noted by Ben-Haim (2012), the performance of a system under “most likely” conditions is a “distinct and independent attribute” from the sensitivity of its performance to deviations from those conditions. Risk scholars working in complex systems (including flood risk management) emphasize the need to design systems that are robust to uncertainty, arguing for satisfying performance over a wide range of scenarios rather than optimal performance in the most likely scenario (Lempert *et al.*, 2003; Keller *et al.*, 2008; Cox, 2012; Aven, 2013; Johnson and Geldner, 2019). In the context of managing flood risk, research has examined the impact of uncertainty in topography, roughness coefficients, and flow data on inundation maps (Jung and Merwade, 2012), but less attention has been given to uncertainty in rainfall.

Part of the assumption behind our approach is that there is an inherent stochastic component associated with these storms and their associated rainfall, with small perturbations to the environmental conditions in which the storms developed that would have had potentially large impacts in terms of their track, strength, and rainfall. Therefore, the goal of this study is to develop an ensemble generator of rainfall associated with synthetic (i.e., hypothetical) or observed TCs making landfall along the Louisiana coast, that can be used to better quantify the risk associated with these storms through efforts such as the Louisiana Watershed Initiative (<https://www.watershed.la.gov/>) or Louisiana’s Coastal Master Plan (Fischbach *et al.*, 2017).

The paper is structured as follows. Section 2 presents the data and the methodology, followed by a description

of the results in Section 3. Finally, Section 4 summarizes the main points of the study and concludes it.

2 | DATA AND METHODOLOGY

The observed rainfall data are obtained from the Stage IV Quantitative Precipitation Estimates (QPE) products over the continental United State (CONUS) (<http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/>) released by the National Centers for Environmental Prediction (NCEP), and available since 2002. We use Stage IV data as the reference because of its high spatial and temporal resolution (i.e., ~4 km and hourly) and because it showed a good performance with respect to rain gage data (e.g., Villarini *et al.*, 2011; Luitel *et al.*, 2018). We will consider the Stage IV rainfall as the true rainfall.

In terms of the TC data, we use the Extended HURDAT data (Demuth *et al.*, 2006; http://rammb.cira.colostate.edu/research/tropical_cyclones/tc_extended_best_track_dataset/). It is available since 1988 and includes the storm name, month, day, time, year, latitude (deg N), longitude (deg W), maximum wind speed (kt), minimum central pressure (hPa), RMW speed (nm), eye diameter (nm), pressure of the outer closed isobar (hPa), radius of the outer closed isobar (nm), and radii (nm) of 34 kt, 50 kt, and 64 kt wind to the northeast, southeast, southwest, and northwest of the storm centre, respectively.

Here we consider storms post 2002 corresponding to the availability of Stage IV QPE, and focus on all the 12 storms that made landfall along the Louisiana coast (Figure 1): Tropical Storm Bertha (2002), Hurricane Isidore (2002), Hurricane Lili (2002), Tropical Storm Bill (2003), Tropical Storm Matthew (2004), Hurricane Cindy (2005), Hurricane Katrina (2005), Hurricane Rita (2005), Tropical Storm Bonnie (2010), Tropical Storm Lee (2011), Hurricane Isaac (2012), and Tropical Storm Cindy (2017). Detailed information about these storms is provided by the National Hurricane Center (<https://www.nhc.noaa.gov/data/#tcr>). The only storm we have not included is Hurricane Ivan (2004) because of its recurving track.

The first step in our approach is to generate storm total rainfall via a parametric TC rainfall model. Among the existing parametric models, we have selected to use the IPET model. Recently, Brackins and Kalyanapu (2020) examined the performance of four different parametric TC rainfall models, and found that IPET exhibited the most skill in reproducing high total precipitation. Here we provide just a brief overview of this model, and point the interested reader to the IPET (2006). The IPET rainfall model is a piecewise function that generates a single precipitation rate for all areas within the RMW (units: nm), and a second, exponentially decaying precipitation rate for areas beyond the RMW. Both regions of precipitation are

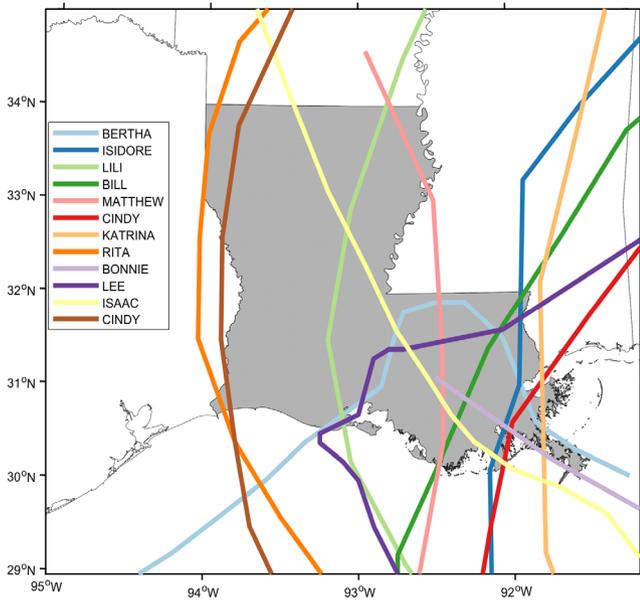


FIGURE 1 Track of the 12 tropical cyclones that made landfall over Louisiana between 2002 and 2017 and that are considered in this study [Colour figure can be viewed at wileyonlinelibrary.com]

parameterized using the TC's central pressure deficit (ΔP ; units: hPa) in the fashion shown below:

$$R_{\text{IPET}}(r) = \begin{cases} 1.14 + 0.12\Delta P, & r \leq \text{RMW} \\ (1.14 + 0.12\Delta P)e^{-0.3\left(\frac{r-\text{RMW}}{\text{RMW}}\right)}, & r > \text{RMW} \end{cases} \quad (1)$$

where $R_{\text{IPET}}(r)$ is the mean rainfall intensity ($\text{mm}\cdot\text{hr}^{-1}$) generated by IPET, and r is the radial distance (km) from the TC centre. To account for azimuthal asymmetry, IPET crudely applies a $1.5\times$ exaggeration factor to the modelled rainfall on the right-hand side of the storm azimuth. The diagnostic variables shown in the expression above were retrieved from the Extended HURDAT dataset, with all inputs linearly interpolated to 20-min resolution from HURDAT's 6-hourly entries during the lifetime of the storm to yield a smoother sequence of R_{IPET} fields. Each R_{IPET} surface was multiplied by 0.333 hr (i.e., $20 \text{ min}/60 \text{ min}\cdot\text{hr}^{-1}$) to produce a 20-min rainfall accumulation and all 20-min totals were summed to produce a single storm-total TC precipitation field estimate over Louisiana (storm total R_{IPET} termed R_m). The IPET rainfall footprint decays exponentially beyond the RMW (Equation (1)); here we calculate IPET rainfall for every storm over a fixed area from 10°N to 40°N and 100°W to 70°W . All IPET estimates were generated at 0.05 decimal degrees ($\sim 5 \text{ km}$) spatial resolution and validated against the Stage IV QPE data, which were regrided to the same spatial domain and resolution as the IPET rainfalls.

Our TC rainfall model takes the lead from Ciach *et al.* (2007) and the further developments in Villarini

et al. (2014b). We summarize the main steps involved in the estimation of the model components in Figure 2. Given the IPET rainfall (Figure 2a), we start by correcting for the bias in a storm's total rainfall (i.e., not varying from pixel to pixel or as a function of rainfall magnitude), which allows us to quantify and correct for the overall over- or under-estimation of the TC rainfall by the IPET model (Figure 2b). Given the dependence of the IPET rainfall on storm characteristics (e.g., RMW speed, TC's central pressure deficit), it is reasonable to assume that the bias may vary from storm to storm:

$$B_i = \frac{R_{o,i}}{R_{m,i}}, \quad (2)$$

where $i = 1, 2, \dots, 12$ (12 storms considered here), $R_{o,i}$ and $R_{m,i}$ are the storm total rainfall during the storm i derived from Stage IV observations and the IPET model, respectively, aggregated over the nonzero pixels. A bias value larger (smaller) than 1 indicates that the IPET model under-estimates (over-estimates) rainfall with respect to the Stage IV data.

Because we want to develop a generator that is applicable beyond the 12 storms considered here, we need to estimate the bias as a function of different storm characteristics:

$$B'_i = f(x_1, x_2, \dots, x_n), \quad (3)$$

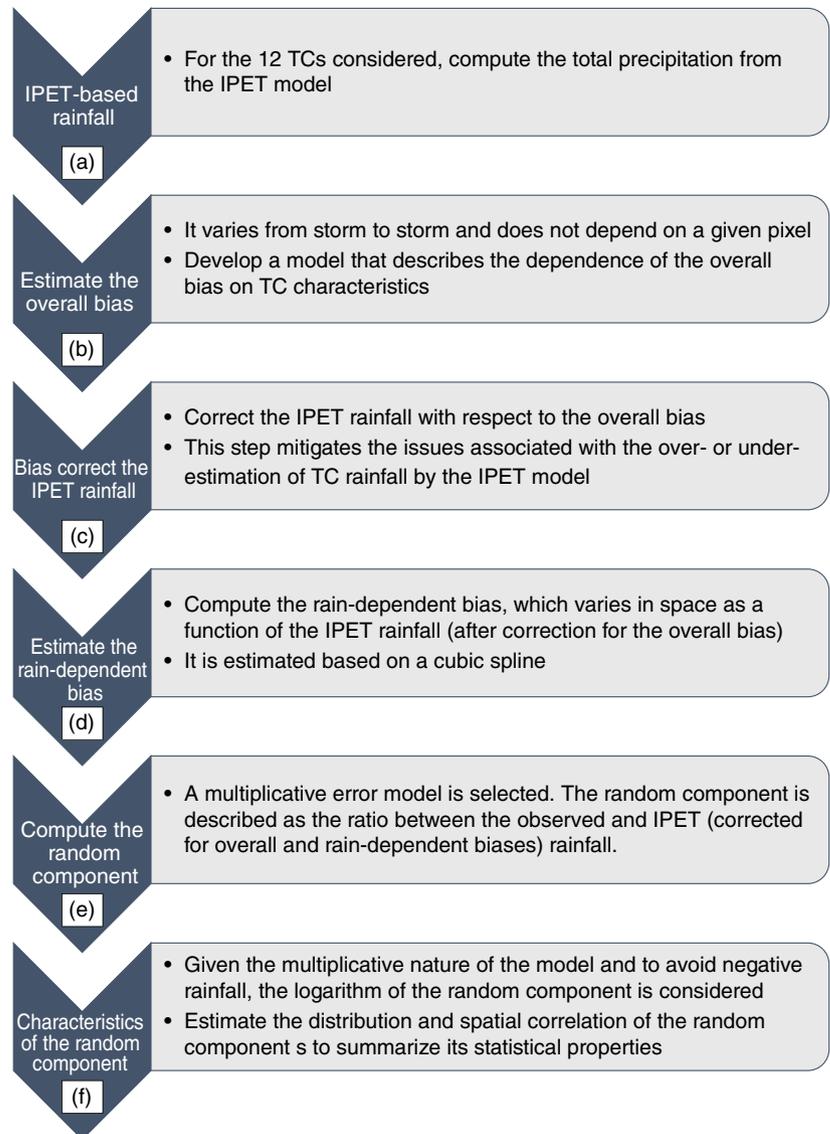
where x_k ($k = 1, 2, \dots, n$) represents the k th storm characteristics, and $f(\cdot)$ a function that maps storm characteristics x_k to B'_i (dependent upon the data). Here we will consider maximum sustained wind, minimum sea level pressure and translation velocity at landfall, and storm total rainfall across the area in Louisiana within a 200-km buffer of a storm's track as potential factors. We select the best model with respect to the adjusted coefficient of determination (i.e., \bar{R}^2 value). Notice the superscript symbol for the bias in Equation (3) compared to Equation (2): this indicates that the bias value is based on the fitted function, not one estimated from the 12 TCs.

Therefore, if we define $R_{m,i}(l)$ to be the modelled rainfall at location (i.e., pixel) l for storm i , we first adjust the location-specific rainfall totals by the fitted overall bias factor B'_i (Figure 2c) from Equation (3):

$$R'_{m,i}(l) = B'_i \cdot R_{m,i}(l). \quad (4)$$

Note that the effect of this step is to correct the storm total rainfall by applying the same correction factor at every location. However, this does not mean that the adjusted value from this step will match observed totals at each location. Thus, we next turn to estimating the residual spatial errors.

FIGURE 2 Flow chart summarizing the main tasks involved in the estimation of the properties of the TC-rainfall generator. IPET, Interagency Performance Evaluation Task Force; TC, tropical cyclone [Colour figure can be viewed at wileyonlinelibrary.com]



Similar to Ciach *et al.* (2007) and Villarini *et al.* (2014b), we consider a multiplicative error model, in which the true rainfall R_o at location l can be described as:

$$R_o(l) = h(R'_m(l)) \cdot \varepsilon(l), \quad (5)$$

where R'_m is the precipitation from the IPET model after correcting for overall bias (see Equation (4)). Notice that we have dropped the subscript i as we will be pooling all the storms together. The function $h(R'_m(l))$ is the deterministic component, while $\varepsilon(l)$ is the random component; the former accounts for the rain-dependent bias, while the latter for the residual fluctuations. Even though the relationship between true and IPET-based rainfall could be described by a number of other formulations, here we have selected a multiplicative error model based on the results by Tian *et al.* (2013), Villarini and Krajewski (2010), and

Villarini *et al.* (2014b). In the development of the model, we will focus on an area of 200 km around the TC track and intersecting Louisiana.

The next step is the estimation of the deterministic component $h(\cdot)$ to capture the relationship between observed and modelled rainfall (Figure 2d). This systematic component varies from pixel to pixel and allows capturing the potential dependence of the bias on modelled rainfall. There are different ways in which it can be estimated (e.g., Villarini *et al.*, 2008). Here we will use a cubic spline, with its degrees of freedom estimated by means of leave-one-out cross-validation.

Once we have estimated the deterministic part, we can turn our attention to the random component (Figure 2e):

$$\varepsilon(l) = \frac{R_o(l)}{h(R'_m(l))}. \quad (6)$$

To fully characterize the random component, we need to estimate its probability density function and spatial correlation. Given the multiplicative nature of the error model, to ensure that we do not get any negative rainfall values we will focus on the modelling of (Figure 2f):

$$\varepsilon'(l) = \log(\varepsilon(l)) = \log\left(\frac{R_o(l)}{h(R'_m(l))}\right). \quad (7)$$

Building on Serinaldi and Kilsby (2012) and similar to Villarini *et al.* (2014b), we focus on mixtures of n Gaussian distributions, which we can write as:

$$f(\varepsilon'(l)) = \sum_{j=1}^n w_j p_j(\varepsilon'(l)), \quad (8)$$

where n represents the total number of components for the mixture of distributions, p_j the j th Gaussian probability density function, while the weights w_j ($j = 1, \dots, n$) need to satisfy the conditions $\sum_{j=1}^n w_j = 1$ and $0 < w_j < 1 \forall j$.

We use the Schwarz Bayesian Criterion (SBC; Schwarz, 1978) to select the number of components n . We will compute the SBC value for an increasing number of components and select the mixture with the lowest SBC. We evaluate the goodness of fit of our mixture of distributions by comparing empirical and fitted cumulative density and survival functions.

The next step is the estimation of the spatial correlation of the random component to reflect the expectation that values at one pixel are more closely related to values at nearby pixels rather than distant ones. Because of the non-Gaussian nature of the error distribution, we cannot use Pearson's correlation coefficient; we instead employ Kendall's τ because it does not require distributional assumptions (e.g., Serinaldi, 2008), and use the weighted average of the spatial correlations for each storm (the weights are based on the sample size used to compute the value of Kendall's τ for a given distance between two pixels).

Once we estimate the components of the random error, the final step is the development of a generator that can reproduce the estimated characteristics. Here we will follow an approach similar to the one detailed in Villarini *et al.* (2014b) and build on the concept of copulas (e.g., Genest and Favre, 2007; Salvadori and De Michele, 2007; Nelsen, 2010). We adopt a meta-Gaussian copula because of its correspondence to a meta-Gaussian-random field in high dimensions. Therefore, our steps include performing a normal quantile transformation to obtain marginals that are close to the Gaussian distribution. In terms of spatial correlation, we transform the

Kendall's τ to the corresponding Pearson's correlation coefficient ρ as (e.g., Fang *et al.*, 2002):

$$\tau = \frac{2}{\pi} \sin^{-1} \rho. \quad (9)$$

Using Equation (9), we obtain the Pearson's correlation coefficient ρ that we need as input for our generator. Here we use a power law model to describe the spatial dependence (e.g., Chilès and Delfiner, 2012):

$$C(d) = \exp\left[-\left(\frac{d}{a}\right)^\alpha\right], \quad (10)$$

where d is the distance between pixels, and a and α are two parameters that we need to estimate.

Therefore, we will need to generate Gaussian-distributed random fields with the spatial structure described by Equation (10). Once we use the inverse of the normal quantile transformation, we will have spatially correlated error fields, with the marginals that are described by a mixture of n Gaussian distributions, and the spatial correlation based on the estimated Kendall's τ .

It is worth highlighting that, while our focus here is on the IPET rainfall model, other deterministic TC rainfall models could have been used; in that case, the parameters of our statistical model would need to be re-estimated given that they are specific to a given TC rainfall model.

All the calculations are performed in R (R Core Team, 2019) using the freely available `gamlss` (Rigby and Stasinopoulos, 2005; Stasinopoulos and Rigby, 2007) and `RandomFields` (Schlather *et al.*, 2015) packages.

3 | RESULTS

The first step in the development of the generator is the calculation of the overall bias (Table 1). When we pool all the storms together, the total overall bias is equal to 0.63, which means that the IPET model tends to overestimate TC rainfall with respect to the Stage IV observations. There is, however, variability in the values of B among the different storms (Table 1), ranging from 0.20 (Bertha, 2002) to 1.98 (Lili, 2002). In the spirit of Equation (3), we have examined the dependence of B on different storm characteristics, including the minimum pressure, translation velocity and maximum wind speed at landfall, and modelled rainfall. After examining different combinations of predictors and functional relationships, we have found that the model with the largest adjusted coefficient of determination is:

TABLE 1 Summary of the overall bias values for the individual storms (first 12 rows) and after pulling all the storms together (last row) based on the original model outputs (second column), after applying the bias correction in Equation (11) using all the storms (third column) or through leave-one-out cross validation (last column)

Storm name (year)	Original bias value	After bias correction	After bias correction (cross-validation)
Bertha (2002)	0.20	0.52	0.40
Isidore (2002)	0.41	0.75	0.72
Lili (2002)	1.98	1.91	2.15
Bill (2003)	0.91	1.12	1.14
Matthew (2004)	1.41	2.03	2.35
Cindy (2005)	1.46	1.11	1.16
Katrina (2005)	0.59	0.44	0.26
Rita (2005)	1.43	1.36	1.54
Bonnie (2010)	0.62	0.65	0.46
Lee (2011)	0.52	1.57	1.97
Isaac (2012)	0.45	0.92	0.90
Cindy (2017)	0.45	0.92	0.91
Total	0.63	1.09	1.03

$$\log_{10}(B') = 1.43 + 0.005 \cdot x_1 - 0.353 \cdot \log_{10}(x_2), \quad (11)$$

where x_1 is the maximum wind speed at landfall (in knots) and x_2 is the total rainfall (in mm) within 200 km of the storm track and intersecting Louisiana.

Overall, this simple relationship can generally capture the observed values (the correlation coefficient is equal to 0.7), with the advantage that it provides a functional form to correct the IPET model: for any observed or synthetic storm, once we use IPET to determine the modelled rainfall across Louisiana and around the storm track and the maximum wind speed at landfall, we can obtain a value of B' based on Equation (11).

The results in Table 1 highlight the performance of this first layer of bias correction. Here, we correct the precipitation associated with each storm based on the functional form of B' in Equation (11). In general, there is a tendency for the bias values to move towards 1 (i.e., overall unbiased estimate), with an overall bias of 1.09 when we consider all the storms together (compared to 0.63 based on the original model simulations). This is also visualized in Figure 3, where the scatterplots prior to this step (left panel) exhibited considerable variability, with the traces for the individual storms that are visible. After bias correction (right panel), the situation improves significantly, with points that are now much more tightly grouped and closer to the 1:1 line. We have also computed the root-mean-square error (RMSE) between

observations and IPET model before and after bias-correction: prior to the bias correction, the RMSE was equal to 91 mm, while it decreases to 59 mm after correcting for the overall bias, highlighting the improvements associated with the bias correction.

After the estimation of the overall bias, we can focus on the rain-dependent bias. This task is performed using a cubic spline, shown in Figure 3 (right panel). Overall, the rain-dependent bias tends to be around the 1:1 line up to about 200 mm. As the modelled rainfall increases further, there is a switch towards over-estimation, especially for the larger values (i.e., above 250 mm).

We evaluate the robustness of these results through leave-one-out cross validation: we remove one storm at the time and use the information from the remaining 11 storms to estimate the overall and rain-dependent biases; we then evaluate how well our model predicts the independent storm. As shown in Table 1, the results for the overall bias are generally comparable to what we computed when using all the storms. When we plot the observed rainfall against the modelled rainfall after bias correction and leave-one-out cross validation (Figure 4, left panel), the results are largely consistent with what obtained pooling all the storms together (Figure 4, right panel). The largest difference is for the trace of Hurricane Katrina (2005), which is outside of the cloud of points against the 1:1 line in the left panel of Figure 4. This is because the maximum wind speed at landfall associated with this storm is the largest among all the TCs

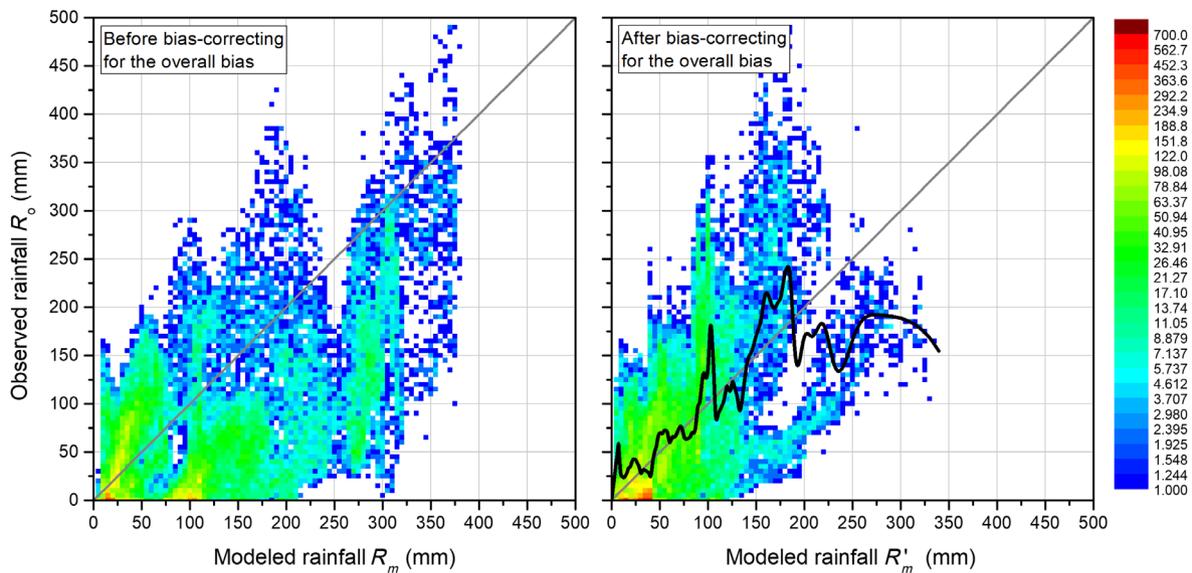


FIGURE 3 Two-dimensional histogram between observed and modelled precipitation before (R_m ; left panel) and after (R'_m ; right panel) correcting for the overall bias at each pixel l using the parametric model in Equation (11). The black line in the right panel represents the rain-dependent bias estimated using cubic splines, with the degrees of freedom optimized using leave-one-out cross validation [Colour figure can be viewed at wileyonlinelibrary.com]

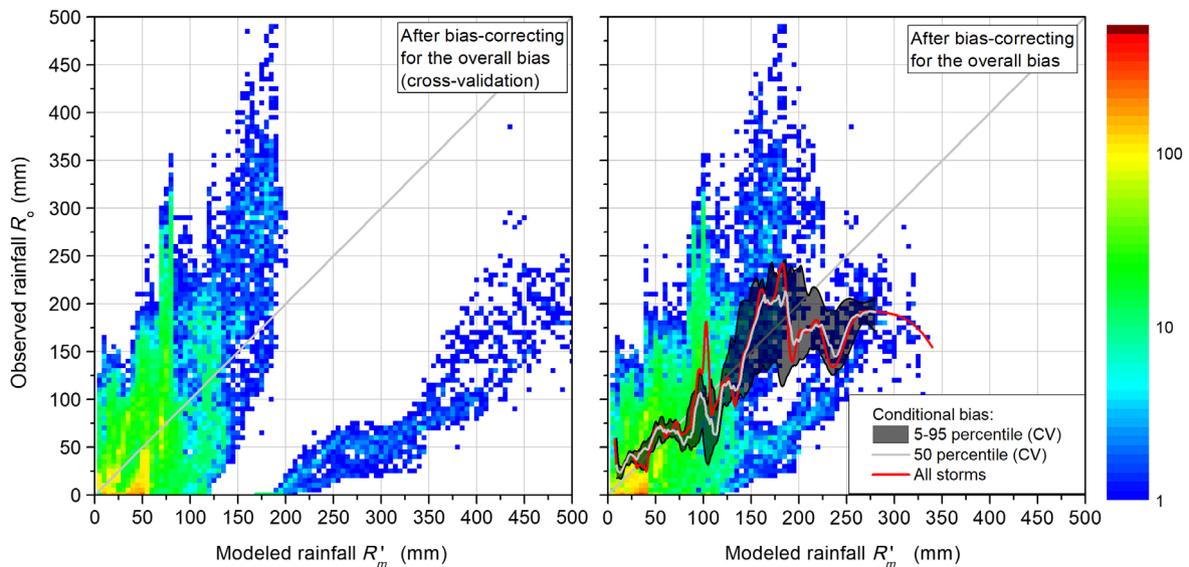


FIGURE 4 Two-dimensional histogram between observed and modelled precipitation after correcting for the overall bias at each pixel l based on cross validation (left panel) or after pooling all the storms together (right panel; same as the right panel in Figure 3). The red line in the right panel represents the rain-dependent bias estimated pooling all the storms and using cubic splines, with the degrees of freedom optimized using leave-one-out cross validation; the light grey line (black area) represents the median (the area between the 5th and the 95th percentile) of all the rain-dependent bias functions based on leave-one-out cross validation [Colour figure can be viewed at wileyonlinelibrary.com]

considered here, leading to a coefficient of 0.009 for x_1 in Equation (11), much larger than for any other storm; this, in turn, leads to an overall bias value of 0.26 compared to 0.44 if all the TCs were considered (Table 1). The results of the cross-validation for the rain-dependent bias suggest

that, while there is some variability, the overall pattern is consistent with respect to what we estimated using all the storms (Figure 4, right panel).

Figure 5 shows the results for the 12 storms after correcting for overall and rain-dependent biases. Depending

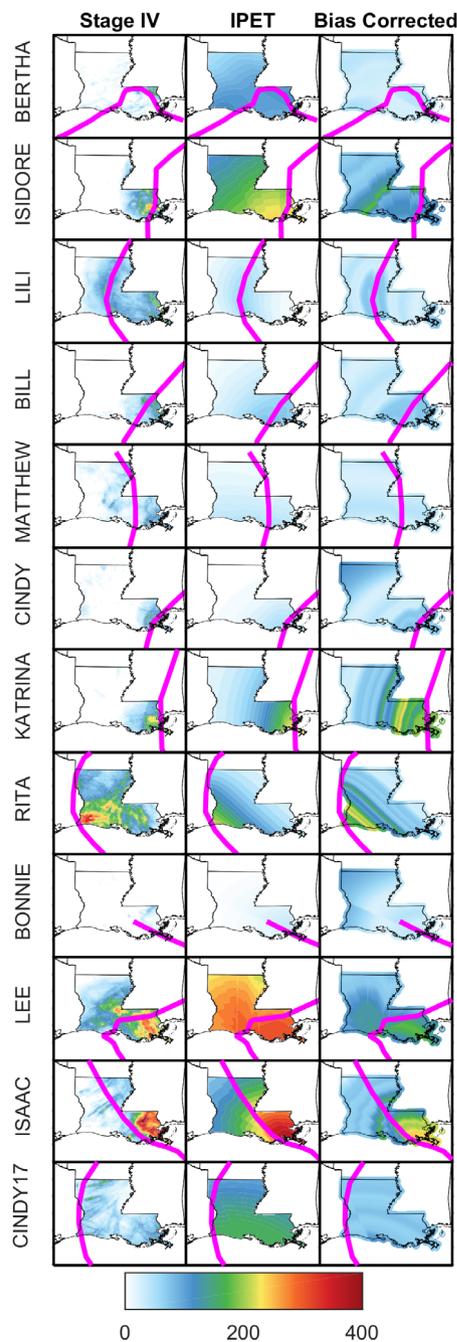


FIGURE 5 Storm total rainfall (unit: mm) associated with the 12 storms considered here (one per row). The results in the left panels are based on Stage IV data, while those in the middle and right panels are based on the Interagency Performance Evaluation Task Force (IPET) model before and after bias correction [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

on the storm, the bias correction either increases or decreases precipitation totals, with results that are now closer to the observations. One element that persists is the tendency for these fields to be smooth. This is largely due to the IPET model and the smooth fields that it (as well as other aforementioned parametric TC

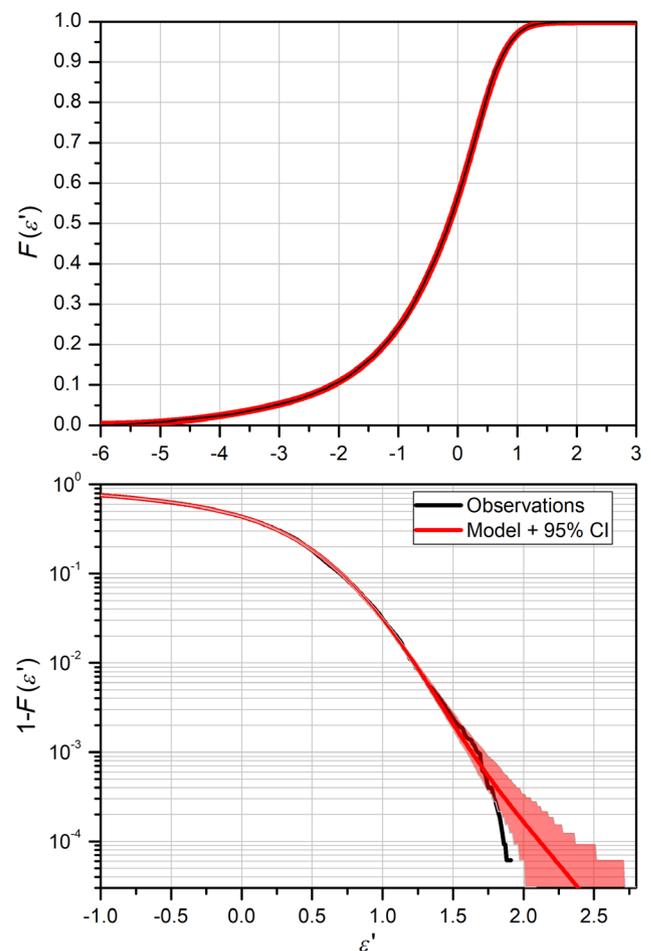


FIGURE 6 Comparison of the empirical (black line) and fitted (red line; based on the mixture of four Gaussian distributions) cumulative distribution (top panel) and survival (bottom panel) function. The light red region represents the 95% confidence intervals representing the sampling uncertainties [Colour figure can be viewed at [wileyonlinelibrary.com](#)]

precipitation models) generates, and the random component will address that issue by capturing the variability around the deterministic function (Figure 3, right panel).

Based on Figure 5, one way to think of the random component is by adding multiplicative “noise” to the results in the right column. To accomplish this task, we need to estimate its distribution and spatial correlation. As mentioned in Section 2, we model ε' , which is the logarithm of the random component. One of the advantages of modelling the logarithm of the random component, rather than the random component itself, is that we enforce the errors in the linear space to be strictly positive, avoiding issues related to potentially negative values due to the multiplicative nature of our model. Based on the results in Figure S1, a mixture of four Gaussian distributions is selected based on SBC (see Table S1 for a summary of each of the components). As shown in Figure 6,

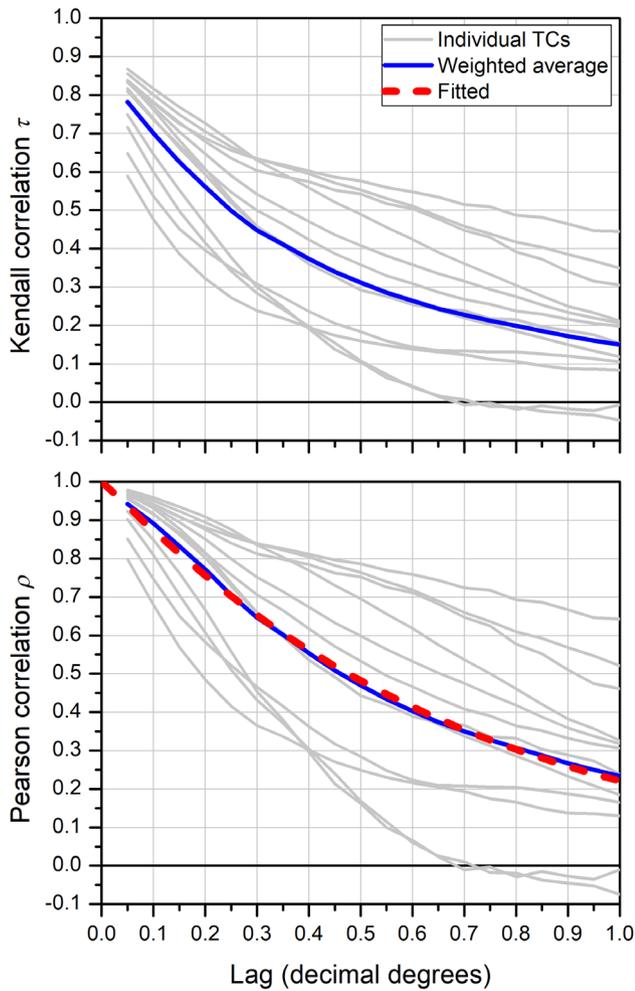


FIGURE 7 Plot of the spatial correlation of ε' based on Kendall's τ (top panel) and the corresponding results based on Pearson's ρ (bottom panel; based on Equation (9)). The light grey lines are the results for each of the 12 tropical cyclones (TCs), while the blue lines represent their weighted averages (with the weights based on the sample size for each lag). The red dashed line represents the fitted spatial correlation function based on Equation (10) [Colour figure can be viewed at wileyonlinelibrary.com]

the fitted distribution can reproduce well the empirical results, providing additional supporting evidence of the goodness of our model.

The results for the spatial correlation of ε' are shown in Figure 7. Because of its nonparametric nature, we start with Kendall's τ (Figure 7, top panel), and then convert it to the corresponding Pearson's ρ based on Equation (9). While there is some variability in the spatial dependence among the different TCs, the weighted average captures well the overall patterns. Here we fit the results based on Pearson's ρ with the power law model in Equation (10). We estimate a value of a equal to 0.68 decimal degrees and α equal to 1.05, which suggests that it decorrelates (using $1/\exp$ (~ 0.37) as reference) at a distance of 0.68

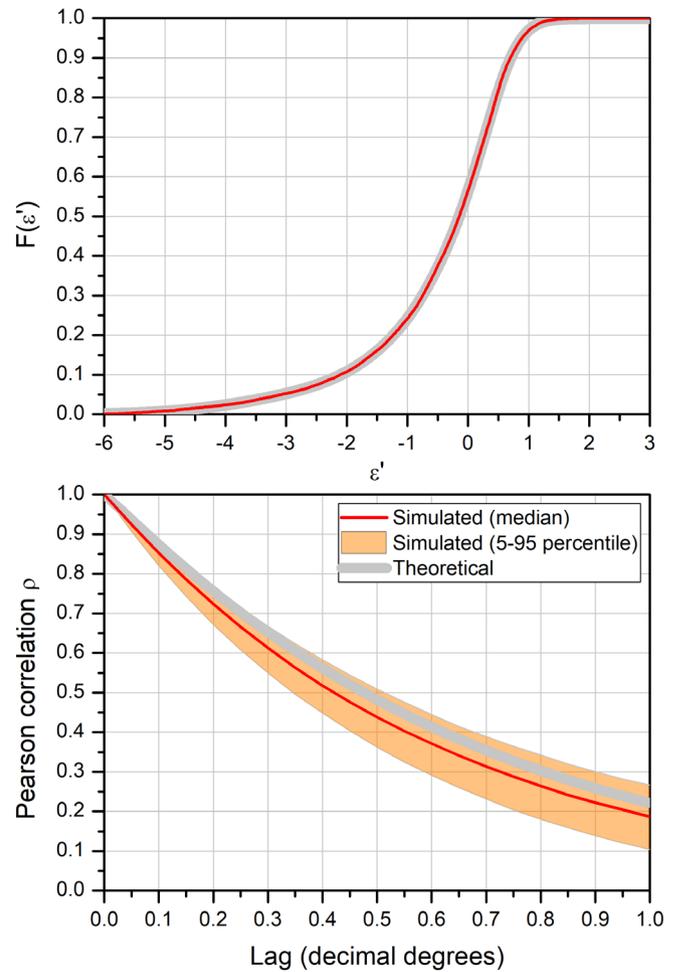


FIGURE 8 Plot of the input (grey line) and simulated (red line; the orange region represents the area between the 5th and the 95th percentiles) cumulative distribution function (top panel) and spatial correlation (bottom panel). In the top panel, the confidence intervals are too small to be visible [Colour figure can be viewed at wileyonlinelibrary.com]

decimal degrees; moreover, the shape coefficient α controls the behaviour of the correlation at small distances, with a value of 1.05 indicating that correlation tends to be close to the corresponding exponential function. Overall, the fitted model can reproduce well the empirical results.

At this point, we have estimated the distribution and spatial dependence structure of ε' , which represent the inputs for the ensemble generator. Based on the methodology described in Section 2, we can generate Gaussian random fields (which we can then back-transform to mixtures of Gaussian distributions using the inverse of the normal quantile transformation; Figure 8, top panel), with the prescribed spatial correlation (Figure 8, bottom panel). Therefore, we have developed a generator that is able to correct the IPET outputs for overall and rain-dependent biases and perturb them with noise that is

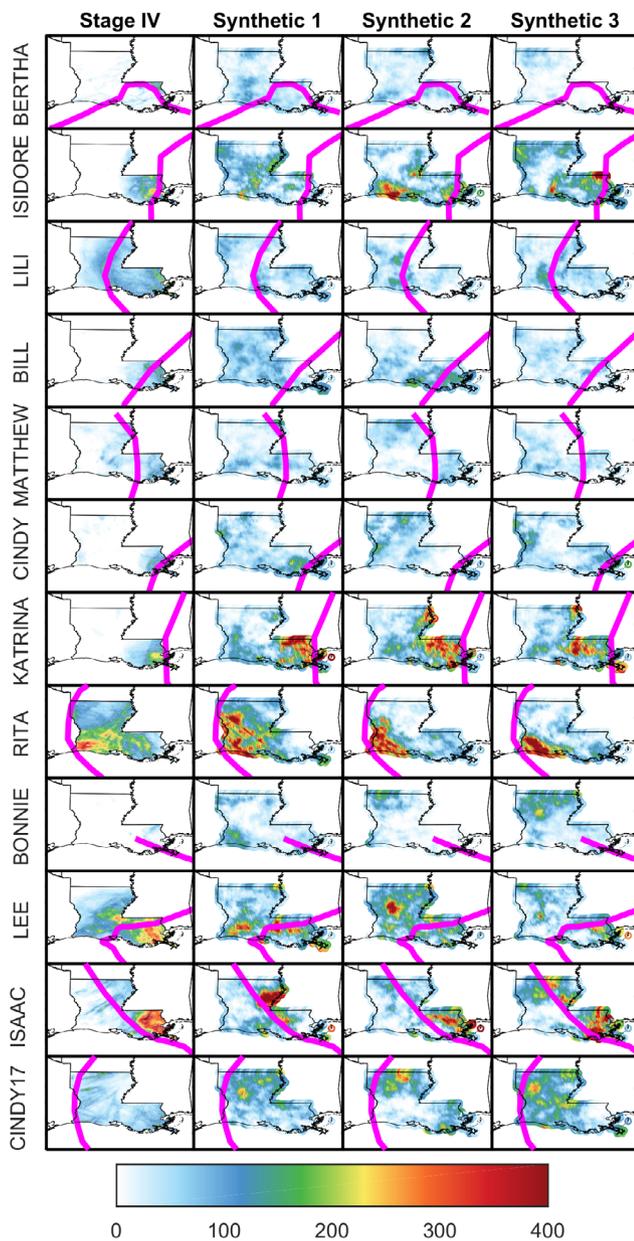


FIGURE 9 Examples of the synthetic storm total rainfall (unit: mm) associated with the 12 storms considered here (one per row). The left row represents the observed rainfall (Stage IV), while the remaining three columns have examples of synthetic realizations based on the ensemble generator [Colour figure can be viewed at wileyonlinelibrary.com]

consistent with what we estimated with respect to the observations. Given that the random component was not dependent on the rainfall values, we are able to generate a large ensemble of random error fields offline, which we can then use to perturb the bias-corrected IPET fields. We show the storm total rainfall based on observations and three synthetic examples for each of the 12 storms in Figure 9. The simulated fields tend to place the heavier precipitation close to the track, consistent with observations.

However, there is also rainfall that is further away from the track that, even though it did not materialize during these events, could have happened. These fields appear realistic compared to the observations, even though there are differences in terms of the areas that receive some of the largest rainfall amounts. This is because we are not performing conditional simulations (i.e., we are not forcing to have certain values at certain locations); hence our results suggest that there could have been other locations within the state that could have experienced heavy precipitation and potentially flooding other than those that effectively experienced the largest impacts.

4 | CONCLUSIONS

We have proposed and developed an ensemble generator for rainfall associated with landfalling TCs. Our focus was on the state of Louisiana, and 12 storms that made landfall along its coastline between 2002 and 2017. It is a multiplicative model that includes two components, a deterministic and a stochastic one. The first step corrected the storm total rainfall associated with the IPET parametric model for the overall bias, which we found to depend on the total amount of modelled rainfall and maximum wind speed at landfall. Then, we estimated the rain-dependent bias (i.e., a bias function of the rainfall value at a given pixel) using a cubic spline. To fully characterize the random component, we estimated its probability distribution and spatial correlation. Due to the multiplicative nature of our model, we focused on the logarithm of the multiplicative errors, which we found could be described by a mixture of four Gaussian distributions. We used Kendall's τ to estimate its correlation structure; we then transformed it into Pearson's ρ and parametrized using a power law correlation function. After having estimated the model's components, we used them as inputs for our generator, which we showed was able to reproduce these characteristics well.

We used leave-one-out cross validation to evaluate the robustness of our model. While the performance of the model was generally preserved when considering independent storms, we highlighted the role of a major hurricane like Hurricane Katrina in impacting the results. Based on the insights of these analyses, it is important to include a variety of storms in the estimation of the model's parameters to avoid extrapolating the results too far from the training set.

Given the stochastic nature of these events, future work could examine a wide range of possible rainfall impacts for these historical events, providing a more comprehensive view of risk associated with these storms (i.e., not just what happened but also what could have happened). Beyond re-analysis of historical events, our

method could be used to better incorporate uncertainty in rainfall distributions into estimates of compound flood risk from rainfall, riverine overflows, and/or storm surge.

Moreover, there are several aspects worth discussing as far as possible future developments and extensions. Here we have focused on the IPET model to provide the initial quantification of the TC rainfall. However, given the data-driven nature of our approach, other TC-rainfall simulators could be used; while our approach would be applicable, the parameters and components of the model would need to be estimated based on the different TC-rainfall model.

While here we have focused on Louisiana, our modeling framework is not limited to this area, but could be expanded to other coastal regions affected by landfalling TCs; if very large domains were of interest, then a preliminary step would be to identify sub-regions to capture regional variability in precipitation and TC characteristics. Furthermore, the application of this generator is not just in a hindcast mode, but it could also be used in conjunction with the forecasting of TCs: based on the storm track and characteristics, once we generate the TC rainfall based on the IPET model, we can first bias-correct it, and then generate ensembles of possible TC rainfall fields; these data could then be used as input to hydrologic models for flood forecasting. The same line of thinking could be applied to global climate models and superimpose our synthetic rainfall fields on the tracks simulated by the global climate models.

ACKNOWLEDGEMENTS

This research was supported by the Louisiana Watershed Initiative; for more information visit their website <https://www.watershed.la.gov/>. Comments and suggestions by Alex McCorquodale (The Water Institute of the Gulf) and three anonymous reviewers are gratefully acknowledged.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The Stage IV Quantitative Precipitation Estimates (QPE) products over the continental United State (CONUS) are available from <http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/stage4/>. The Extended HURDAT data are available from http://rammb.cira.colostate.edu/research/tropical_cyclones/tc_extended_best_track_dataset/.

ORCID

Gabriele Villarini  <https://orcid.org/0000-0001-9566-2370>

Paul Miller  <https://orcid.org/0000-0002-0512-8295>

REFERENCES

- Aryal, Y.N., Villarini, G., Zhang, W. and Vecchi, G.A. (2018) Long term changes in flooding and heavy rainfall associated with North Atlantic tropical cyclones: roles of the North Atlantic Oscillation and El Niño-Southern Oscillation. *Journal of Hydrology*, 559, 698–710.
- Aven, T. (2013) On how to deal with deep uncertainties in a risk assessment and management context. *Risk Analysis*, 33, 2082–2091.
- Ben-Haim, Y. (2012) Doing our best: optimization and the management of risk. *Risk Analysis*, 32, 1326–1332.
- Bilskie, M.V. and Hagen, S.C. (2018) Defining flood zone transitions in low-gradient coastal regions. *Geophysical Research Letters*, 45, 2761–2770.
- Brackins, J.T. and Kalyanapu, A.J. (2020) Evaluation of parametric precipitation models in reproducing tropical cyclone rainfall patterns. *Journal of Hydrology*, 580, 124255.
- Bucknam, R.C., Coe, J.A., Chavarria, M.M., Godt, J.W., Tarr, A.C., Bradley, L.-A., Rafferty, S.A., Hancock, D., Dart, R.L. and Johnson, M.L. (2001) *Landslides triggered by Hurricane Mitch in Guatemala – inventory and discussion*. Open-File Report.
- Cangialosi, J.P., Hagen, A.B. and Berg, R. (2019) Hurricane Barry (AL022019). *National Hurricane Center Tropical Cyclone Report*.
- Chen, S.Y.S., Knaff, J.A. and Marks, F.D. (2006) Effects of vertical wind shear and storm motion on tropical cyclone rainfall asymmetries deduced from TRMM. *Monthly Weather Review*, 134, 3190–3208.
- Chilès, J.-P. and Delfiner, P. (2012) *Geostatistics: Modeling Spatial Uncertainty*, 2nd edition. Hoboken, New Jersey: Wiley, p. 734
- Ciach, G.J., Krajewski, W.F. and Villarini, G. (2007) Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with NEXRAD data. *Journal of Hydrometeorology*, 8, 1325–1347.
- Cox, L.A., Jr. (2012) Confronting deep uncertainties in risk analysis. *Risk Analysis*, 32, 1607–1629.
- Czajkowski, J., Villarini, G., Montgomery, M., Michel-Kerjan, E. and Goska, R. (2017) Assessing current and future freshwater flood risk from North Atlantic tropical cyclones via insurance claims. *Scientific Reports*, 7, 41609.
- Demuth, J.L., DeMaria, M. and Knaff, J.A. (2006) Improvement of advanced microwave sounding unit tropical cyclone intensity and size estimation algorithms. *Journal of Applied Meteorology and Climatology*, 45, 1573–1581.
- Fang, H.B., Fang, K.T. and Kotz, S. (2002) The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis*, 82, 1–16.
- Fischbach, J., Johnson, D.R., Kuhn, K., Pollard, M., Stelzner, C., Costello, R., Molina, E. and Sanchez, R. (2017) *2017 Coastal Master Plan Modeling: Attachment C3-25 – Storm Surge and Risk Assessment*. Baton Rouge, LA: LA Coastal Protection and Restoration Authority.
- Fuhrmann, C.M., Konrad, C.E. and Band, L.E. (2008) Climatological perspectives on the rainfall characteristics associated with landslides in western North Carolina. *Physical Geography*, 29, 289–305.
- Genest, C. and Favre, A.-C. (2007) Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, 12, 347–368.

- Geoghegan, K.M., Fitzpatrick, P., Kolar, R.L. and Dresback, K.M. (2018) Evaluation of a synthetic rainfall model, P-CLIPER, for use in coastal flood modeling. *Natural Hazards*, 92, 699–726.
- Interagency Performance Evaluation Task Force (IPET). (2006) *Performance evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System draft final report of the Interagency Performance Evaluation Task Force volume VIII – engineering and operational risk and reliability analysis*.
- Jing, R. and Lin, N. (2020) An environment-dependent probabilistic tropical cyclone model. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001975.
- Johnson, D.R. and Geldner, N.B. (2019) Contemporary decision methods for agricultural, environmental, and resource management and policy. *Annual Review of Resource Economics*, 11, 19–41.
- Jung, Y. and Merwade, V. (2012) Uncertainty quantification in flood inundation mapping using generalized likelihood uncertainty estimate and sensitivity analysis. *Journal of Hydrologic Engineering*, 17, 507–520.
- Keller, K., Yohe, G. and Schlesinger, M. (2008) Managing the risks of climate thresholds: uncertainties and information needs. *Climatic Change*, 91, 5–10.
- Klotzbach, P.J., Bowen, S.G., Pielke, R. and Bell, M. (2018) Continental US hurricane landfall frequency and associated damage – observations and future risks. *Bulletin of the American Meteorological Society*, 99, 1359–1376.
- Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J. and Neumann, C.J. (2010) The International Best Track Archive for Climate Stewardship (IBTrACS) unifying tropical cyclone data. *Bulletin of the American Meteorological Society*, 91, 363–376.
- Ko, M.C., Marks, F.D., Alaka, G.J. and Gopalakrishnan, S.G. (2020) Evaluation of Hurricane Harvey (2017) rainfall in deterministic and probabilistic HWRF forecasts. *Atmosphere-Basel*, 11, 1–17.
- Landsea, C.W. and Franklin, J.L. (2013) Atlantic hurricane database uncertainty and presentation of a new database format. *Monthly Weather Review*, 141, 3576–3592.
- Langousis, A. and Veneziano, D. (2009) Long-term rainfall risk from tropical cyclones in coastal areas. *Water Resources Research*, 45, W11430. <https://doi.org/10.1029/2008WR007624>
- Lempert, R.J., Popper, S.W. and Bankes, S.C. (2003) *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*. Santa Monica, California: RAND Corporation.
- Lonfat, M., Marks, F.D. and Chen, S.Y.S. (2004) Precipitation distribution in tropical cyclones using the tropical rainfall measuring Mission (TRMM) microwave imager: a global perspective. *Monthly Weather Review*, 132, 1645–1660.
- Lonfat, M., Rogers, R., Marchok, T. and Marks, F.D., Jr. (2007) A parametric model for predicting hurricane rainfall. *Monthly Weather Review*, 135, 3086–3097.
- Lu, P., Lin, N., Emanuel, K., Chavas, D. and Smith, J. (2018) Assessing hurricane rainfall mechanisms using a physics-based model: Hurricanes Isabel (2003) and Irene (2011). *Journal of the Atmospheric Sciences*, 75, 2337–2358.
- Luettich, R. and Westerink, J. (2004) *Formulation and Numerical Implementation of the 2D/3D ADCIRC Finite Element Model Version 44.XX*.
- Luitel, B., Villarini, G. and Vecchi, G.A. (2018) Verification of the skill of numerical weather prediction models in forecasting rainfall from US landfalling tropical cyclones. *Journal of Hydrology*, 556, 1026–1037.
- Marks, F.D. and DeMaria, M. (2003) *Development of a Tropical Cyclone Rainfall Climatology and Persistence (R-CLIPER) Model*. NOAA/OAR/AOML/Hurricane Research Division.
- Miniussi, A., Villarini, G. and Marani, M. (2020) Analyses through the metastatistical extreme value distribution identify contributions of tropical cyclones to rainfall extremes in the eastern United States. *Geophysical Research Letters*, 47, e2020GL087238.
- Nadal-Caraballo, N.C., Melby, J.A., Gonzalez, V.M. and Cox, A.T. (2015) *Coastal Storm Hazards from Virginia to Maine*. US Army Corps of Engineers – Engineer Research and Development Center, p. 221.
- Nelsen, R.B. (2010) *An Introduction to Copulas*. New York, NY: Springer Publishing Company Incorporated.
- R Core Team. (2019) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rappaport, E.N. (2014) Fatalities in the United States from Atlantic tropical cyclones. *Bulletin of the American Meteorological Society*, 95, 341–346.
- Resio, D.T. (2007) *White Paper on Estimating Hurricane Inundation Probabilities Coastal and Hydraulics Laboratory*. U.S. Engineer Research and Development Center (U.S.), p. 124.
- Rigby, R.A. and Stasinopoulos, D.M. (2005) Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54, 507–544.
- Salvadori, G. and De Michele, C. (2007) On the use of copulas in hydrology: theory and practice. *Journal of Hydrologic Engineering*, 12, 369–380.
- Schlather, M., Malinowski, A., Menck, P.J., Oesting, M. and Strokorb, K. (2015) Analysis, simulation and prediction of multivariate random fields with package random fields. *Journal of Statistical Software*, 63, 1–25.
- Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, 6, 461–464.
- Serinaldi, F. (2008) Analysis of inter-gauge dependence by Kendall's tau(K), upper tail dependence coefficient, and 2-copulas with application to rainfall fields. *Stochastic Environmental Research and Risk Assessment*, 22, 671–688.
- Serinaldi, F. and Kilsby, C.G. (2012) A modular class of multisite monthly rainfall generators for water resource management and impact studies. *Journal of Hydrology*, 464, 528–540.
- Smith, R.K. (1968) The surface boundary layer of a hurricane. *Tellus*, 20, 473–484.
- Stasinopoulos, D.M. and Rigby, R.A. (2007) Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, 23, 1–46.
- Tian, Y., Huffman, G.J., Adler, R.F., Tang, L., Sapiano, M., Maggioni, V. and Wu, H. (2013) Modeling errors in daily precipitation measurements: additive or multiplicative? *Geophysical Research Letters*, 40, 2060–2065.
- Tuleya, R.E., DeMaria, M. and Kuligowski, R.J. (2007) Evaluation of GFDL and simple statistical model rainfall forecasts for US landfalling tropical storms. *Weather and Forecasting*, 22, 56–70.

- Villarini, G., Goska, R., Smith, J.A. and Vecchi, G.A. (2014a) North Atlantic tropical cyclones and U.S. flooding. *Bulletin of the American Meteorological Society*, 95, 1381–1388.
- Villarini, G. and Krajewski, W.F. (2010) Sensitivity studies of the models of radar-rainfall uncertainties. *Journal of Applied Meteorology and Climatology*, 49, 288–309.
- Villarini, G., Seo, B.-C., Serinaldi, F. and Krajewski, W.F. (2014b) Spatial and temporal modeling of radar rainfall uncertainties. *Atmospheric Research*, 135, 91–101.
- Villarini, G., Serinaldi, F. and Krajewski, W.F. (2008) Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches. *Advances in Water Resources*, 31, 1674–1686.
- Villarini, G., Smith, J.A., Baeck, M.L., Marchok, T. and Vecchi, G. A. (2011) Characterization of rainfall distribution and flooding associated with US landfalling tropical cyclones: analyses of Hurricanes Frances, Ivan, and Jeanne (2004). *Journal of Geophysical Research-Atmospheres*, 116, D23116. <https://doi.org/10.1029/2011JD016175>
- Yussouf, N., Jones, T.A. and Skinner, P.S. (2020) Probabilistic high-impact rainfall forecasts from landfalling tropical cyclones

using Warn-on-Forecast system. *Quarterly Journal of the Royal Meteorological Society*, 146, 2050–2065.

- Zhu, L., Quiring, S.M. and Emanuel, K.A. (2013) Estimating tropical cyclone precipitation risk in Texas. *Geophysical Research Letters*, 40, 6225–6230.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Villarini, G., Zhang, W., Miller, P., Johnson, D. R., Grimley, L. E., & Roberts, H. J. (2021). Probabilistic rainfall generator for tropical cyclones affecting Louisiana. *International Journal of Climatology*, 1–14. <https://doi.org/10.1002/joc.7335>