



A saturated stochastic simulator: synthetic US Gulf coast tropical cyclone precipitation fields

Jennifer Nakamura¹ · Upmanu Lall² · Yochanan Kushnir¹ · Patrick A. Harr³

Received: 25 July 2022 / Accepted: 27 September 2023
© The Author(s) 2023

Abstract

The space–time fields of rainfall during a hurricane and tropical storm (TC) landfall are critical for coastal flood risk preparedness, assessment, and mitigation. We present an approach for the stochastic simulation of rainfall fields that leverages observed, high-resolution spatial fields of historical landfalling TCs rainfall that is derived from multiple instrumental and remote sensing sources, and key variables recorded for historical TCs. Spatial realizations of rainfall at each time step are simulated conditional on the variables representing the ambient conditions. We use 6 hourly precipitation fields of tropical cyclones from 1983 to 2019 that made landfall on the Gulf coast of the US, starting from 24 h before landfall until the end of the track. A conditional K -nearest neighbor method is used to generate the simulations. The TC attributes used for conditioning are the pre-season large-scale climate indices, the storm maximum wind speed, minimum central pressure, the latitude and speed of movement of the storm center, and the proportion of storm area over land or ocean. Simulation of rainfall for three hurricanes that are kept out of the sample: Katrina [2005], Rita [2005], and Harvey [2017] are used to evaluate the method. The utility of coupling the approach to a hurricane track simulator applied for a full season is demonstrated by an out-of-sample simulation of the 2020 season.

Keywords N. Atlantic tropical cyclone · Precipitation · Stochastic model · Gulf states

1 Introduction

The 2020 North Atlantic hurricane season was record-breaking with 30 named tropical cyclones (TC) and 11 mainland US landfalls, of which eight were on the Gulf coast (NHC 2020). More named TC mainland landfalls occurred in 2020 than in any season on record. One of the Gulf coast landing TCs, Sally, moved slowly over Alabama resulting in 20–30 inches of rain in both coastal and inland areas, causing flash and river flooding (NWS

✉ Jennifer Nakamura
jennie@ldeo.columbia.edu

¹ Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

² Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA

³ Jupiter Intelligence, San Mateo, CA, USA

2020). In 2021, Hurricane Ida made landfall in Louisiana as a category 4 storm bringing flooding to the Gulf coast and the eastern seaboard as it exited to the North Atlantic (NHC 2021). These Gulf landfalls brought attention to the frequency and intensity of TC precipitation in this region during the North Atlantic hurricane season. An updated report of US casualties from North Atlantic TCs found that rainfall-induced freshwater floods accounted for a little more than a quarter of all fatalities (Rappaport 2014).

Stochastic simulation of hurricane tracks and precipitation is useful for hydrologic design and risk analysis. In prior work (Nakamura et al. 2015, 2021), we presented a hurricane track simulator (HITS) based on a K -nearest neighbor (k -NN), a semi-Markov process that allows a simulation of realistic stochastic hurricane tracks including several of their attributes (e.g., center latitude and longitude, maximum wind speed, and minimum pressure), using the NOAA National Hurricane Center HURDAT tropical storm (TC) track dataset for the North Atlantic. The model has exhibited good performance in simulating TC tracks and attributes as well as landfall statistics on the Gulf coast. While not designed for the prediction of evolving TCs, it has done well in out-of-sample probabilistic track predictions for several TC (Nakamura et al. 2015, Figs. 10 and 11). Recently, the simulator was extended to consider the stochastic simulation of TC tracks for an upcoming season by conditioning the track algorithm on early-season large-scale sea surface temperature (SST) conditions (Nakamura et al. 2021). Here, we seek to extend HITS to also simulate the space–time precipitation fields surrounding landfall. The rainfall field simulation uses a novel k -NN algorithm that conditions historical precipitation fields on the tropical cyclone attributes of maximum wind speed, minimum pressure, center latitude, speed of movement, the portion over land or ocean, and pre-season climate indices.

The background of the relevant literature and the rationale for selecting the conditional variables is presented in the next section. The data used, consisting of HURDAT2, PERSIANN-CCS-CDR precipitation, land fraction quadrants, and pre-season climate indices are described in Sect. 3. Methods are discussed in Sect. 4. Section 5 presents the design for the applications and the associated results. A summary and discussion of our current and forthcoming work with tropical cyclones conclude the paper.

2 Background

Deterministic process-based models have been used for TC precipitation modeling in the case of downscaling future climate simulations (Emanuel et al. 2008), for risk assessment (Langousis and Veneziano 2009a), for hindcast analysis (Emanuel 2017), and to address rainfall mechanisms as in Lu et al. (2018). Marks (2003) used satellite data to update the Tropical Cyclone Rainfall Climatology and Persistence (R-CLIPER) model. This model produces a symmetric (around the TC center) rain distribution that decays linearly in time after landfall. Parametric Hurricane Rainfall Model (PHRaM) updated R-CLIPER to include effects from shear and topography (Lonfat et al. 2007). Precipitation-Climatology and Persistence (P-CLIPER) updated R-CLIPER to include a spatial distribution as a function of distance from the center in the three intensity categories of (1) tropical storms, (2) category 1–2 TC, and (3) category 3–5 TC (Geoghegan et al. 2018). Zhu et al. (2013) modeled seasonal TC precipitation in Texas with linear regressions on three or fewer predictors. Many averaging areas, temporal scales, and meteorological variables were tested. They found ENSO, maximum potential wind velocity, and vorticity as the most important predictors of seasonal precipitation, precipitation days, and percentage of

tropical precipitation. Kleiber et al. (2020) recently presented a statistical TC space–time precipitation generator for Texas Gulf coast landfalls. The model was trained on patterns of precipitation from seven hurricanes, accumulated in 1-h increments from the Weather Research and Forecasting (WRF) model, giving both a high spatial and temporal resolution. Predictor variables common to North Atlantic hurricane track datasets were derived from the WRF model including an environmental pressure deficit at TC center, position of cyclone center and speed, radius of maximum winds, and distance to the Texas coastline. Precipitation was simulated in a Lagrangian framework, with the TC center as the center of the domain, by dividing it into mean and residual patterns on polar coordinates. Means were computed by empirical orthogonal functions and residuals by a random decision forest algorithm. Details of TC precipitation patterns were simulated by the method including asymmetry, changes in precipitation as the cyclone moves over land, incorporation of drier air, and total precipitation amounts.

Brackins and Kalyanapu (2020) compared the performance of several parametric models for TC rainfall and found that the IPET (2006) model performed best. The IPET model considers two zones with rainfall that is parameterized based on the TCs central pressure deficit, with constant rainfall in the central core, and an exponential decay rate for the rainfall in the outer zone. Villarini et al. (2022) followed this up with an approach to simulate precipitation fields for TCs that entail a bias correction step, followed by a scaling argument for spatial precipitation and the use of a parametric Gaussian copula to model the spatial dependence of a residual field from the basic IPET model relative to the observations. An application to fitting the data for 12 TCs with landfalls in Louisiana is presented.

A TC precipitation model that uses surface-level TC winds has been defined by Lu et al. (2018) and employed by Feldman et al. (2019) to define a TC precipitation climatology over eastern North America and by Xi et al. (2020) to assess risk due to TC-induced precipitation. The TC-induced precipitation model partitions the vertical motion at the top of the boundary layer into contributions from terrain, friction, stretching, baroclinic effects, and radiation. A vertical vapor flux is computed using the net vertical velocity and the saturation specific humidity, then modified by a precipitation efficiency factor to define precipitation rate.

At this point, it is useful to synthesize some key observations from the body of the literature as to variables that may potentially influence the precipitation pattern for TCs. First consider, maximum wind speed and minimum pressure are correlated measures of hurricane intensity. Touma et al. (2019) studied the link between TC intensity and precipitation over land. They found that TCs that were hurricane strength and then weakened to a tropical storm over land produced the most precipitation. The argument was that the attainment of a higher intensity led to a larger cyclone size that covered more area and hence more total precipitation.

Now, the radius of maximum winds is a measure of cyclone size and impacts precipitation amount through the size of the rain bands. In a study of four groups of TCs sized from smallest to largest that impacted North Carolina, the two largest groups produced the most rainfall (Konrad and Perry 2010). The smallest group only contained tropical storms or depressions; none were hurricane strength. At constant precipitation intensity, a slower speed of movement of the TC center will result in a higher overall storm-related precipitation amount than for a faster-moving cyclone (Kossin 2018; Chan 2019; Lai et al. 2020).

Precipitation fields from landfalling TCs are typically asymmetrical and have a high degree of variability. Corbosiera and Molinari (2003) found asymmetries due to TC motion (strongest in the right front core and on the right of motion in outer bands), linked to the vertical wind shear, beta effect, and the tilt with the height of the cyclone. Lu et al. (2018)

found frictional convergence to be the dominating factor in the rainfall pattern in two East Coast hurricanes. TC position affects the amount of surface friction. TCs overwater experience less friction. As the cyclone precipitation bands move over land friction increases. The fraction of land vs. sea that a TC covers (land–sea fraction) can be a proxy for the combined effect of the center position, relative land position, and landfall impact angle.

The cyclone center latitude is a factor in the transition of a TC to an extratropical system as the TC moves out of the tropics. The process of extratropical transition creates large precipitation asymmetries (Atallah and Bosart 2003) due to increasing storm translation speed and vertical wind shear due to interactions between the TC and the midlatitude flow into which the TC is moving. Interaction with extratropical systems and topography can also affect TC movement, increase upward vertical motion, and precipitation amount (Chan et al. 2019).

Based on this discussion, we identified the following attributes as potential predictors of the precipitation field associated with landfalling TCs in the Gulf region: the TC maximum wind speed, minimum central pressure, the latitude and speed of movement of the storm center, and proportion of storm area over land or ocean. Attributes of the larger-scale climate system, through SST indices, were also considered as potential predictors of precipitation. The larger-scale climate directs TC development, movement, and intensification, as well as ambient conditions in the region at the time of landfall. Thus, they can be considered latent variables that influence the basin scale and regional conditions attendant to a TC. To use them predictively, SST indices for different months that precede the June–November N. Atlantic hurricane season were considered.

3 Data

The predictors considered were obtained from the revised best-track dataset of Atlantic (HURDAT2), the Worldbath world elevation, and Hadley SSTA datasets. Several track-associated precipitation datasets were considered. Station gauge data and radar data were missing data areas as the track approached from the ocean. Satellite data have much better coverage spatially and temporally but typically underrepresent tropical precipitation. The underestimation of TC precipitation using satellite products such as the NASA Tropical Rain Measurement Mission (TRMM) was a problem in R-CLIPPER. Averaged over five TCs, R-CLIPPER predicted less than half of the peak gage-estimated TC-total precipitation (Marks 2003). The PERSIANN CCS-CDR dataset employs a neural net analysis of cloud clusters and a climate data record to capture higher intensity TC precipitation well. Ashouri et al. (2015) found that the PERSIANN-CDR had good agreement with stage IV radar data during Hurricane Katrina in 2005.

3.1 HURDAT2

The Atlantic hurricane database 2 (HURDAT2) is an updated version of the best-track HURDAT of TCs, subtropical cyclones, and tropical depression tracks in the North Atlantic Ocean, Gulf of Mexico, and Caribbean Sea (Landsea et al. 2015) available at <https://oasishub.co/dataset/hurdat-2-atlantic-hurricane-database>. It includes cyclone number, name, year, month, day, hour, and minute (synoptic times of 00Z, 06Z, 12Z, and 18Z and non-synoptic times), center latitude and longitude, landfall indicator, minimum center pressure, and maximum wind speed for the period studied. Non-synoptic

times were removed for a dataset of observed track locations every 6 h for the 84 tracks that occurred during the precipitation dataset time period 1983–2019.

3.2 PERSIANN-CCS-CDR precipitation

The Center for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine (UCI), created the ‘Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks’ (PERSIANN) using neural networks on geostationary satellite brightness temperatures and is available at <https://chrsdata.eng.uci.edu> (Nguyen 2019). The PERSIANN-CCS-CDR is a combination of the PERSIANN-CCS, which includes the cloud clustering algorithm, and the PERSIANN-CDR, which includes the climate data record. The PERSIANN-CCS improves on the PERSIANN by using cloud patches that have distinctive properties rather than directly fitting pixel brightness to rainfall rate (Hong et al. 2004). PERSIANN-CDR incorporates the Global Precipitation Climatology Project (GPCP) monthly product for consistency (Ashouri et al. 2015). The use of satellite products allows better spatial coverage than radar products off the coastline. The native resolution of 0.04 (1/25) degrees was downgraded to 0.25 (1/4) degrees to reduce the number of pixels. The 6-h temporal resolution was selected to match the HURDAT2.

3.3 Land fraction quadrants

Boxes of size 16 by 16 degrees and centered on the TC were segmented into four (8 by 8 degree) quadrants for calculation of the fraction of land/sea in each. World bathymetry from the ETOPO5 5 by 5-min Navy database (Data Announcement 88-MGG-02 1988) available at <http://iridl.ldeo.columbia.edu/SOURCES/WORLDBATH/bath/> was gridded to match the 0.25-degree gridded PERSIANN-CCS-CDR with elevations 0 m and above as marked as land (1) and below 0 m as ocean (0) and then averaged over the quadrant. Land fraction quadrants are a measure of both dynamic and thermodynamic forces on a cyclone through surface friction and the area of available heat and moisture from the ocean.

3.4 Preseason climate indices

The start of the traditional June–November N. Atlantic hurricane season has been trending earlier (Kossin 2008). A precipitation model can account for early TCs in May by using April variables. The preseason climate index of NINO3.4 (5S–5N, 170W–120W) represents the state of ENSO. Positive (warm) ENSO inhibits overall N. Atlantic TC development (Gray 1984). The main development region (MDR) (10N–20N, 80W–20W) represents the genesis of TCs from easterly waves that lead to most major hurricanes, with higher SST increasing TC activity (Goldenberg et al. 2001). April NINO3.4 and MDR were computed from the Met Office Hadley HadSST.4.0.1.0 SSTA is available at a 5°-degree resolution from 1850 to current from <https://www.metoffice.gov.uk/hadobs/hadsst4/>. These indices serve the dual purpose of linking ENSO and development region to landfall precipitation and also link climatologically similar seasons.

4 Methods

Section 4.1 presents the method for the K -nearest neighbor (k -NN) precipitation model that creates simulated precipitation field distribution for every point along a tropical cyclone track, both observed tracks and simulated tracks based on track-based predictors. Section 4.2 outlines a method to evaluate the k -NN simulated precipitation field distribution using a mean squared error threshold evaluation. Once a method has been established for evaluation, track-based predictors can be tested for fitness in the model in Sect. 4.3. In Sect. 4.4, an overall detection for irregularities in the simulated precipitation field is tested by Benford's law. Section 4.5 outlines the method for testcase of linking the k -NN precipitation model to a N. Atlantic track simulator for a sample year.

4.1 K -nearest neighbor (k -NN) precipitation

A k -NN approach for the full precipitation field centered on the TC at each time step is used in the precipitation model. We first identify the k -NN of the predictor variables (described in Sect. 2 and below) using Euclidean distance applied to a scaled version of each of the variables. The scaling is from 0 to 1, using the minimum and the maximum of each observation. This has the effect of equally weighting each predictor since the scale is now the same. It also enables the exploration of interpretable weights for each variable. However, for the applications reported here, uniform weights were used as no physical reason was known for applying non-uniform weights. The storm precipitation field is then simulated by randomly drawing from many fields that correspond to these k -NN, using a probability kernel associated with the k -NN.

Consider a TC precipitation field, $P(x, y, t)$, or $\mathbf{P}(t)$, defined over p^*q grid cells at every time step of a TC track. We seek to simulate from the conditional probability density $f(\mathbf{P}(t)|\mathbf{v}(t))$, where \mathbf{v} is a set of potential predictors of the field. The p^*q frame for precipitation is defined such that it is centered at on the TC, i.e., the track position, and extends symmetrically around that location. As members of \mathbf{v} , consider $q_1(t)$ to $q_4(t)$, $w(t)$, $p(t)$, $\text{lat}(t)$, NINO3.4(t), MDR(t), and $s(t)$, which refer to the land fraction in each of the four quadrants of the frame, the max wind speed, the central pressure of the TC, the center latitude, pre-season NINO3.4, pre-season MDR, and the TC translation speed at time t .

The dimension, $n_p^*n_q$ of the precipitation field is typically large (4225 pixels), and any given frame may be dominated by 0 precipitation values, while also containing bands or hot spots of very high precipitation. These aspects pose challenges for the application of a traditional multivariate probability density function (including a copula representation) that accounts for the heterogeneous spatial dependence structure of precipitation in the frame as well as the zero values and the skewed distributions and the high dimension. These aspects motivated the choice of a non-parametric approach. Conditional simulations or bootstraps using an implicit k -NN conditional density estimation process have been shown to give good results for meteorological variables (Lall et al. 1996; Lall and Sharma 1996; Rajagopalan and Lall 1999; Gangopadhyay et al. 2009), and we explored this approach here.

The k -NN conditional bootstrap introduced in Lall and Sharma (1996) has the following structure. Consider simulating from the conditional distribution $f(y|x=x^*)$ by finding the k -nearest neighbors of x^* in x , and then drawing the y corresponding to these neighbors using a discrete kernel or probability function. Lall and Sharma (1996) suggested the kernel,

$$K(j) = \frac{1/j}{\sum_{i=1}^k 1/j} \tag{1}$$

where j is the index of the neighbor (1 is nearest), and k is the number of neighbors considered. They suggest a nominal choice of $k = \sqrt{n}$, as a default choice, where n is the sample size of the data on x . In our application, n , which is the number of precipitation frames (fr) over the 1983–2019 period without any missing pixels, is 1439, and so k was selected as 37. There is only modest sensitivity to the choice of k , and the kernel adapts in shape to the local density of x observations around x^* . The k -NN algorithm is sensitive to the choice of a distance metric for the selection of neighbors, and only the Euclidean distance was used here.

In Eq. (1), the numerator corresponds to the weight, while the denominator represents the summation of all weights, serving the purpose of normalizing the kernel function. Consider the following, when $j=1$, the associated weight (numerator) is 1; for $j=2$, the weight is 0.5; extending further, for $j=3$, the weight is 1/3, and so forth, finishing at $j=37$ with a weight of 1/37. These individual weights undergo normalization by division with the aggregate sum of all 37 weights, ensuring that the resultant values are scaled appropriately.

A nominal choice is to select neighbors using a Euclidean distance metric d_E ,

$$d_E(x, x^*) = \sqrt{(x - x^*)^2} \tag{2}$$

However, if the components of x are correlated, the Mahalanobis distance d_M (Mahalanobis 2018, reprinted from 1936),

$$d_M(x, x^*) = \sqrt{(x - x^*)^T S_x^{-1} (x - x^*)^T} \tag{3}$$

is considered useful where S_x is the covariance matrix between x and x^* .

Karlsson and Yakowitz (1987) consider a general distance version d_W

$$d_W(x, x^*) = \sqrt{(x - x^*)^T W (x - x^*)} \tag{4}$$

where W is a diagonal matrix of weights that can be selected by optimization to yield the best $f(y|x)$.

Instead of d_M or d_W , but with a similar motivation to address covariation across predictors, and relevance to the response, we explored a canonical correlation analysis of the total and maximum precipitation in the frame versus the predictors and then used just the two canonical covariates for the predictors that were maximally correlated with the two precipitation indicators. However, the results from this approach were not superior to those from a direct application of the k -NN resampling using d_E in Eq. (2).

We note that the algorithm as presented is fully automatic, in that the kernel, the distance metric, and the number of k -NN are specified a priori and are not tuned using a goodness of fit metric. Figure 1 summarizes the simulation strategy. The historical dataset of predictors is used to find the k -NN of the current predictor vector using the algorithm fast k -NN from R version 1.4.1106. Then, 1000 precipitation fields that correspond to this 37 k -NN (j) using the probability kernel (K) associated with the k -NN are drawn for each time step for each proposed TC as in Eq. (1). The choice of 1000 was motivated by several considerations: It aligns with the order of magnitude of the original dataset, it was large enough for adequate representation of the population distribution, and it facilitated a straightforward calculation of percentiles. The current predictor vector is then updated

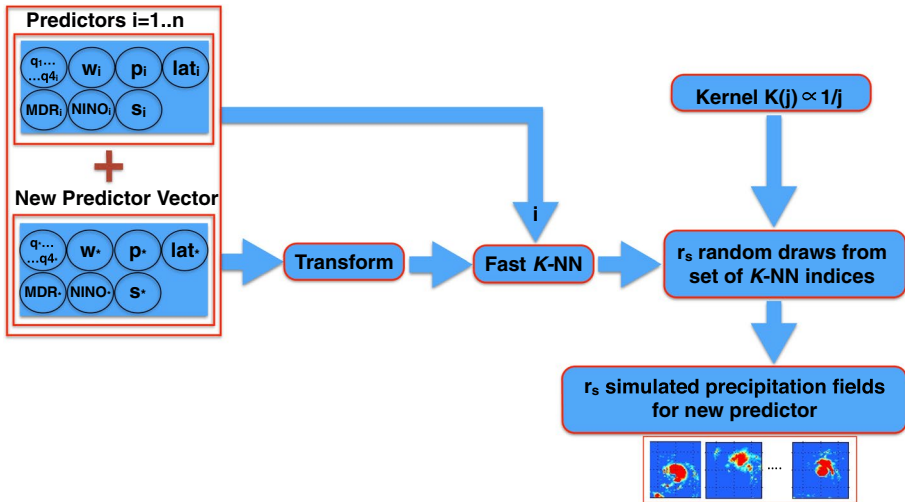


Fig. 1 Flowchart illustrating the method of creating simulated precipitation fields. First, we combine the predictors over the historical precipitation field (n number of observations) with the new predictor vector. Then, the combined matrix is normalized and transformed by rescaling to equally weight the predictors. We identify the K -nearest neighbors (Fast k -NN) of the new predictor variables, using the indices of the historical precipitation field (i). The new precipitation field is then simulated by drawing precipitation fields that correspond to these k -NN, using a probability kernel associated with the rank of the k -NN where j is the number of neighbors

as one moves forward in time steps. Thus, if hurricane tracks have been simulated by any method (e.g., HITS), at each time step, the current predictor vector is known from that simulation and is used to identify the k most similar situations in the entire database of Gulf landfalling TCs, and these are resampled based on their degree of similarity to the current predictor vector to provide a stochastic simulation of the potential precipitation field at each time step as the landfall progresses. The stochastic simulation yields 1000 precipitation fields for each discrete time step.

4.2 MSE threshold evaluation

Following Casati et al. (2004), who assessed forecasts using the mean squared error (MSE) skill score of binary images, three thresholds of precipitation were selected to create binary images. Threshold binary images normalize the discontinuous, noisy, heavily right-hand skewed with large outliers, and large zero populations typical of precipitation data. MSE threshold evaluation was chosen because of these reasons. The discontinuities from the abrupt spatial changes in precipitation data can make it difficult to apply scoring methods that assume continuous data. Precipitation data are noisy, skill scores are meant to capture meaningful signal, and excessive noise can obscure this signal. Precipitation data are heavily right-hand skewed with large outliers, some skill scores rely on assumptions of normality or symmetry. Precipitation data have large zero populations which can complicate the calculation of percentiles.

To determine the threshold values, the maximum precipitation observed for each pixel across all 1439 frames was computed. The minimum precipitation was 3.7 mm/6 h, while

the maximum reached 478 mm/6 h. The median precipitation value was 108 mm/6 h, and the mean was 129 mm/6 h. The threshold values were selected based on these computed values. In results using all Gulf landfalling TCs during 1983–2019 ‘all,’ thresholds were set at low: 3.5 mm/6 h to 50 mm/6 h, mid: 50–100 mm/6 h, and high: over 100 mm/6 h.

For in-sample and out-of-sample individual TC results, recognizing that the peak rainfall can vary significantly by storm, and hence, the thresholds may need to vary by duration and by storm, for the long duration of Hurricane Harvey, we were able to use the same thresholds as the ‘all’ TC. For Katrina, low: 1–25 mm/6 h, mid: 25–50 mm/6 h, and high: over 50 mm/6 h, and for Rita, low: 1–35 mm/6 h, mid: 35–70 mm/6 h, and high: over 70 mm/6 h.

Binary values (1—yes and 0—no) were assessed for each TC observed frame pixel and its associated model 1000 draws of neighbors (Eq. 1) for each category using the low, mid, and high precipitation thresholds above. Also, 1000 random frames were drawn as a control. Figure 2 shows an example of an observed precipitation frame 200 (F200), the first TC of 1994, Tropical Storm Alberto precipitation accumulated from 1800 to 2400Z July 1, 1994 in (a). (b) shows the pixels that fall into the low category marked in blue. (c) is

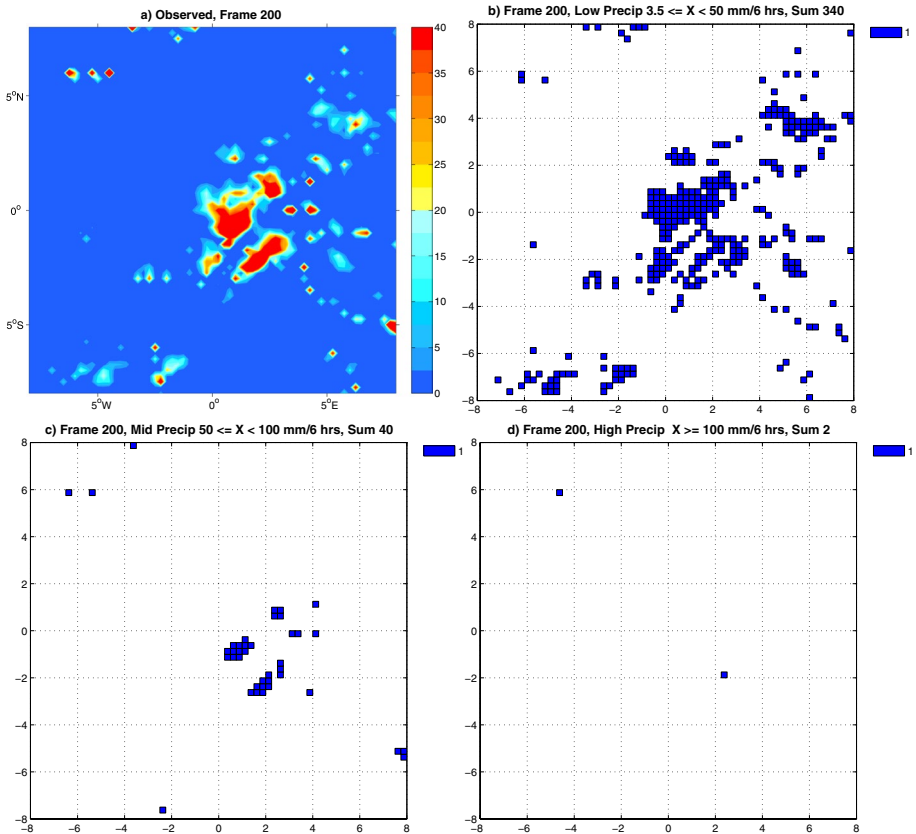


Fig. 2 a Example precipitation frame of the first TC of 1994, Tropical Storm Alberto precipitation accumulated from 1800–2400Z July 1, 1994. Binary (1 yes and 0 no) pixels in three categories. **b** Low defined as greater than or equal to 3.5 and less than 50 mm per 6 h, **c** middle defined as greater than or equal to 50 and less than 100 mm per 6 h, and **d** high defined as greater than or equal to 100 mm per 6 h

the middle category, and finally, (d) is the high category. Looking at frequency only, these binary values are averaged over the whole frame then subtracted from the model results averaged over the frame, and the result squared to create the mean MSE as in

$$MSE_f = \frac{1}{fr} \sum_1^{fr} \left(\left(\sum_1^n p - \sum_1^n m \right)^2 \right) \tag{5}$$

where MSE_f is 1000 mean squared errors computed for frequency only for each of the three thresholds of low, mid, and high, fr is the number of frames, n is the number of pixels in the frame, p contains the pixel observed precipitation binary counts, and m contains the pixel model precipitation binary counts.

These binary values are also aggregated into larger boxes (5 by 5 pixels) to capture spatial coherence rather than individual pixel agreement for spatial and frequency. Figure 3 shows an example of the aggregation of the observed frame binary values. In space and

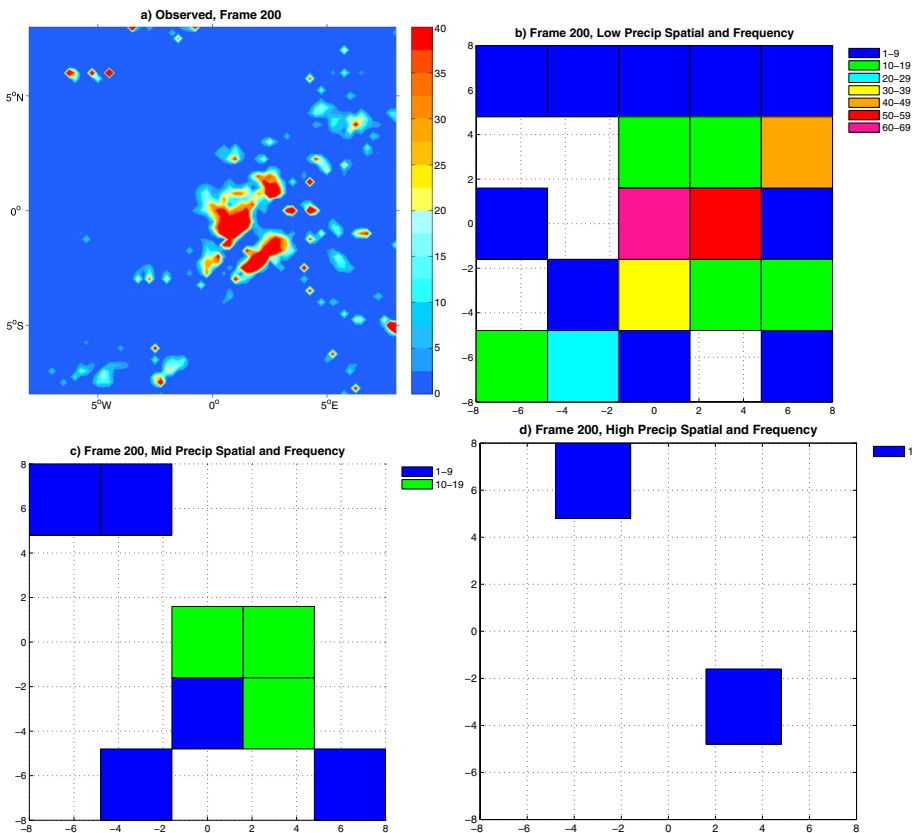


Fig. 3 a Example precipitation frame in the upper left of the first TC of 1994, Tropical Storm Alberto precipitation accumulated from 1800 to 2400Z July 1, 1994. Aggregated binary pixels in the three categories. b Low defined as greater than or equal to 3.5 and less than 50 mm per 6 h, c middle defined as greater than or equal to 50 and less than 100 mm per 6 h, and d high defined as greater than or equal to 100 mm per 6 h. Values are counts of binary values in each of the larger boxes

frequency, the 25 aggregated boxes are subtracted from the aggregated boxes of each model frame, squared, and then averaged over the frame to create the mean MSE as

$$MSE_{sf} = \frac{1}{fr} \sum_1^{fr} \frac{1}{a} \sum_1^a (p_a - m_a)^2 \tag{6}$$

where MSE_{sf} is 1000 mean squared errors computed for spatial and frequency for each of the three thresholds of low, mid, and high, fr is the number of frames, a is the number of aggregated boxes in the frame, p_a contains the observed precipitation binary counts in the aggregated boxes, and m_a contains the model precipitation binary counts in the aggregated boxes.

Mean MSE values are computed over all frames simulated, broken into the six components of low, mid, and high binary, and aggregated binary values shown in Figs. 2 and 3. As a benchmark of performance, we compare the median values of MSEs across all frames or a complete TC using the k -NN simulations vs. the MSEs from an equal number of simulations of randomly selected precipitation frames centered at the TC.

4.3 Predictor selection sensitivity

Predictors available in the HURDAT2 include those recorded and those calculated from the recorded variables. External to the HURDAT2, the April NINO3.4 and MDR computed from Hadley SSTA were also tested. Each predictor is used successively in the k -NN precipitation model and the MSE threshold evaluation. The MSE in each of the three threshold categories (low, mid, and high precipitation) and two spatial scales (MSE_f and MSE_{sf} of Eqs. 5 and 6) is tested against the values for the previous best model using the Kolmogorov–Smirnov (KS) test. The two-sample KS test is a nonparametric test to determine if the underlying distribution is statistically significantly different from the previous test. If the median MSE is lower for the new case and at least one of the three categories indicates that the underlying distribution of the 1000 MSE values is different from the previous test, the predictor is included. If the median MSE is higher or none of the categories shows a significant difference, then the predictor is discarded. Each predictor is added until all are tested.

4.4 Benford’s law

The observed and simulated TC precipitation of the Gulf landfalls 1983–2019 was tested to see if they followed Benford’s law. The precipitation is heavily positively skewed with a long asymmetrical tail of extremely high values. This is also typical of the probability distribution of many natural systems that follow Benford’s law. The law indicates the chance of the leading number 1 of about one-third and decreasing in probability as the first number increases, with 9 occurring to less than 5%,

$$P(\text{leading digit} = d) = \log_{10} ((d + 1)/d), \text{ where } d = 1, 2, 3, \dots 9 \tag{7}$$

(Fewster 2009). Benford’s law was used to investigate tropical cyclone length homogeneity in Joannes-Boyar et al. (2015), fraud detection in financial matters such as Enron’s accounting scandal, and Greece’s economic reports to European authorities (Goodman 2016). Here, we use it as a check to see if the simulations have the same property as the observations. The process of resampling introduces randomness, as it allows for the same data point to be sampled multiple times while excluding others. This variability can result

in differences between resampled and observed distributions. Even if the original dataset follows Benford's law, it is not a guarantee that the simulated will.

4.5 Linking k -NN precipitation and C^3 HITS for sample 2020 year

The sample year of 2020 was run with (cluster climate conditioned) C^3 HITS following Nakamura et al. (2021). An individual simulation was picked that had approximately the same total TC count as observed in 2020 (31 TC—30 named TC and one subtropical TC). C^3 HITS creates a simulated season by starting a specified number of tracks in each N. Atlantic track cluster based on a Poisson regression of early-season indices of NINO3.4 and MDR.

5 Results and discussion

5.1 Predictor selection sensitivity

Using the MSE thresholding evaluation of Sect. 4.2 and the KS tests explained in Sect. 4.3, 13 predictors tested for the k -NN model included in order: four land fraction quadrants, maximum wind speed, minimum pressure, the latitude of the center position, the longitude of the center position, April MDR, April NINO3.4, X component of TC motion, Y component of TC motion, and direct TC translation speed. All had a lower error and statistically significant differences in MSE median except center longitude, X component of TC motion, and Y component of TC motion. These three non-significant or higher error predictors were removed. Center latitude was only significant for the lower threshold, but it also had a lower error for that category, so it was retained. After testing, 10 predictors remained the four land fraction quadrants, maximum wind speed, minimum pressure, center latitude, April MDR and NINO3.4, and TC movement speed.

An example of the precipitation neighbors based on the k -NN model of 10 predictors is shown in Fig. 4. The upper left of Fig. 4 is the observed frame (Obs) which is frame number 200 (F200 as in Fig. 2). N1–N37 are the 37 closest matches to the predictors with 1 being the closest with increasing distance to 37. N1 is the next step on the current track and matches well, but other neighbors have various patterns of precipitation some closer to observations and some diffuse as in N17 or N31. In Eq. (1), the kernel is biased toward selecting frames that are closer to N1 than to N37. Therefore, the 1000 frames that are chosen are expected to be more closely associated with N1.

5.2 Benford's law

The underlying probability distribution given in Eq. (5) was plotted for all observed and modeled precipitation data. Both the observed (blue solid line, Fig. 5) and the k -NN model simulated median precipitation (red circles, Fig. 5) follow Benford's law and match indicating that the underlying natural probability distribution is matched by the simulated distribution.

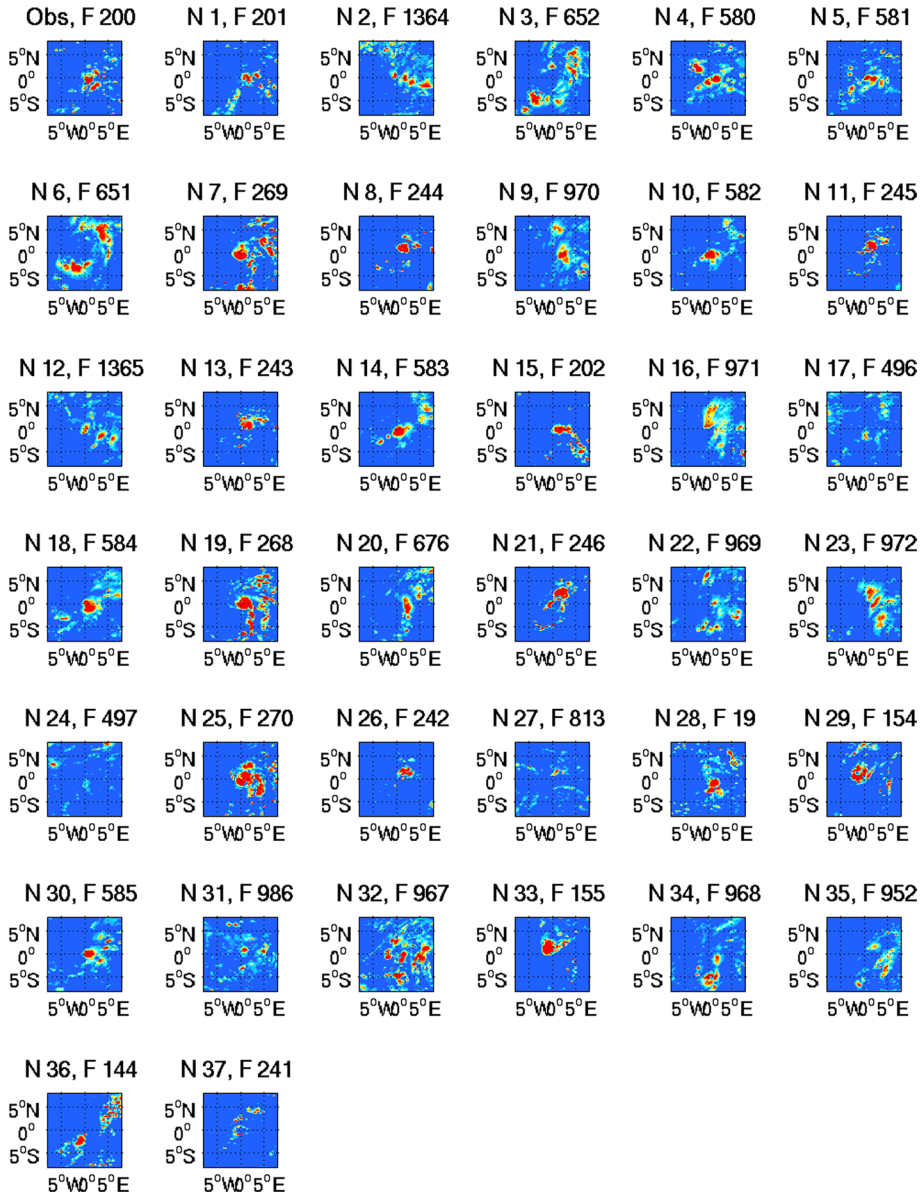


Fig. 4 Observed PERSIANN-CCS-CDR precipitation frame 200 (F 200), Tropical Storm Alberto precipitation accumulated from 1800 to 2400Z July 1, 1994, and the 37 closest neighbors based on the model predictors titled N1–N37 and the associated historical frame number

5.3 Hurricanes Harvey, Rita, and Katrina

The precipitation patterns of Hurricanes Harvey (Fig. 6a), Rita (Fig. 6b), and Katrina (Fig. 6c) summed from 24 h before landfall until the end of the track in mm/6 h. Both Harvey and Rita were under low vertical wind shear, which is normally the case for Gulf

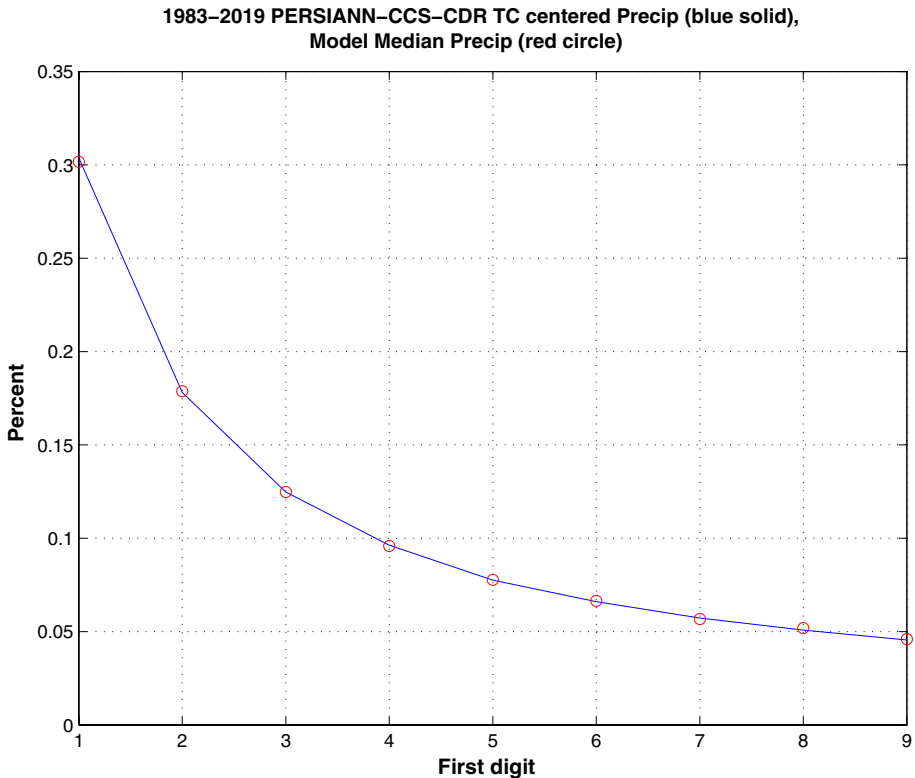


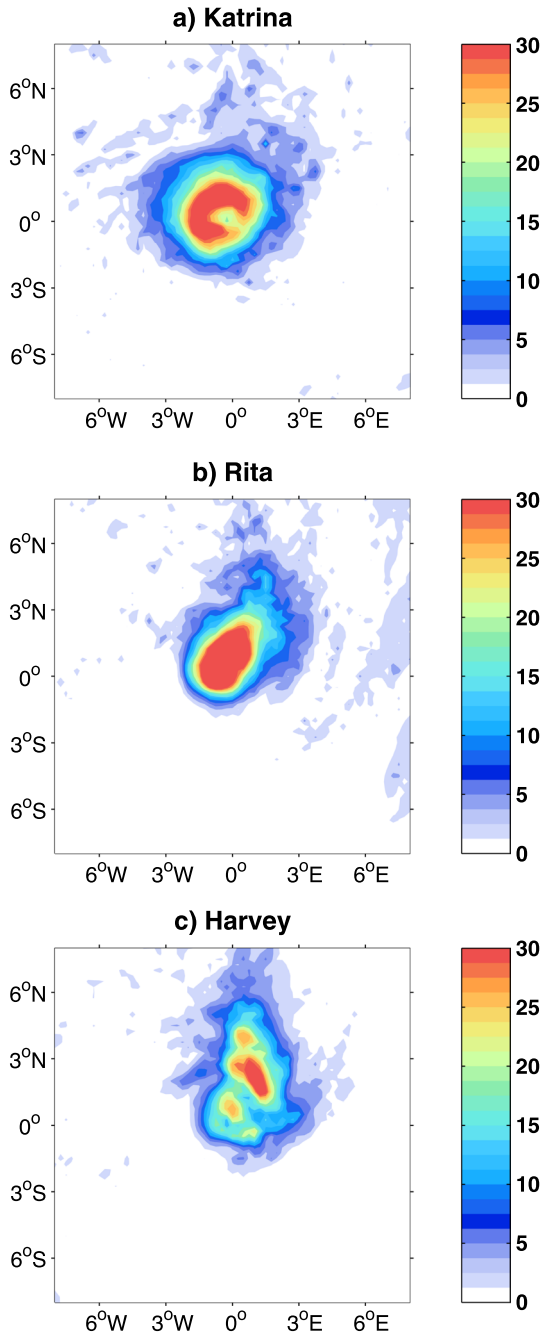
Fig. 5 Benford's law (first digit vs. percentage) applied to 1983–2019 PERSIANN-CCS-CDR tropical cyclone-centered precipitation as a solid blue line and model median precipitation as red circles

landfalls. Consequently, the strong and extensive precipitation was in the northeastern quadrant (Fig. 6b and c). Katrina interacted with a mid-latitude trough causing increasing vertical wind shear, with most of the precipitation falling in the northwestern quadrant in the onshore flow. Ashouri et al. (2015) showed the PERSIANN-CDR captures a wide view of the precipitation and hurricane landfall during severe weather when radar sites go down as the Lake Charles radar site did in Southwest Louisiana during Katrina. Figure 6 is produced with the updated PERSIANN-CDR-CCS.

After masking the observed average of 1 mm/6 h over the track frames, the percentile of the observed value relative to the 1000 k -NN model ensemble is shown in Fig. 7. Dark blues indicate the model overestimated the precipitation, light blues, greens, and yellows indicate that the precipitation was well represented, and oranges and reds indicate that the precipitation was underestimated. The orange values on the fringe of hurricanes are low values shown in Fig. 6. In model validation, the median of the model simulation can be used as a reference point for comparative analysis. While exact correspondence with observations is not expected, a degree of proximity is anticipated, suggesting that the model is capable of reasonably simulating the underlying processes.

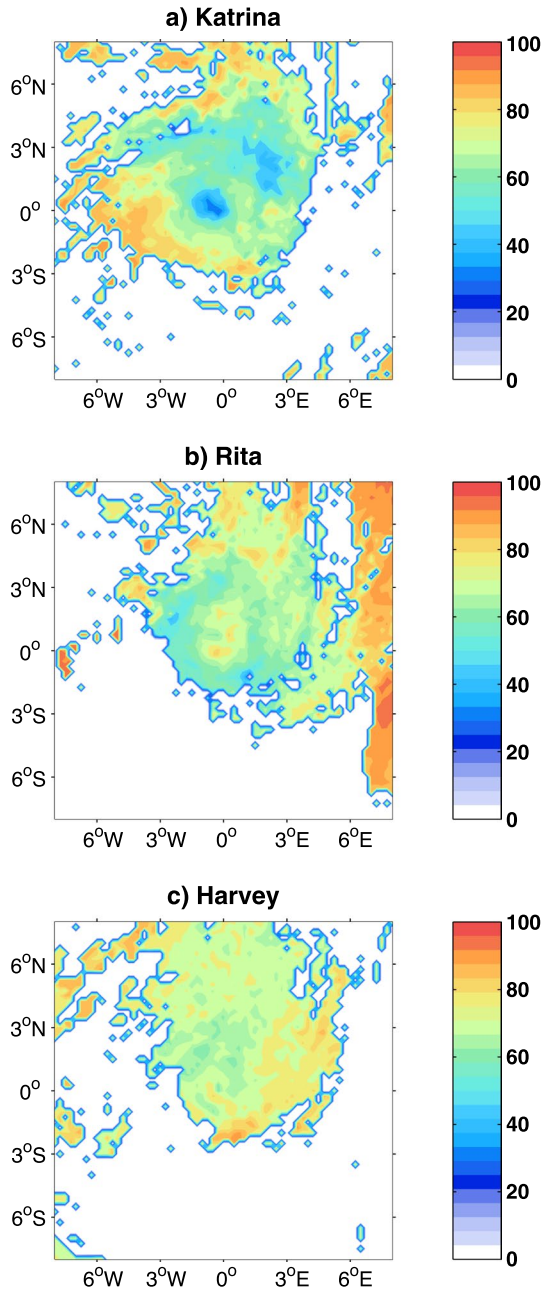
Using the MSE threshold evaluation of Sect. 4.2, Fig. 8 shows the results for the k -NN model precipitation (blue) and a random draw of precipitation frames (red) over low a), mid b), and high c) categorized precipitation amounts. This view shows the distribution

Fig. 6 Six-h average tropical cyclone-centered precipitation over a track within 24 h of landfall to the end of the track in mm/6 h



of MSE errors. Frequency only computes the MSE after averaging the binary image in total (Eq. 5). In and out of sample performance is shown for individual tracks in categories for the three Gulf coast landfalling hurricanes Katrina (K), Rita (R), and Harvey (H). In sample (In), the remainder of the track is included as possible neighbors and out

Fig. 7 Percentile of the observed sum of tropical cyclone-centered precipitation over a track to 1000 modeled precipitation runs summed over a track within 24 h of landfall to the end of the track. Masked below average 1 mm/6 h per 6 h frame. Hurricanes Katrina and Rita (2005) contain 14 frames and Hurricane Harvey (2017) 35



of sample (Out), the remainder of the track is not included. Individual frames are always calculated out of sample. The k - NN precipitation model is not calibrated for the precipitation field; therefore, cross-validation is not required. However, the frame predicted is not used, which provides cross-validation. The random draw (red) from the full population of frames is the same for in or out of sample. Random selects the same number of

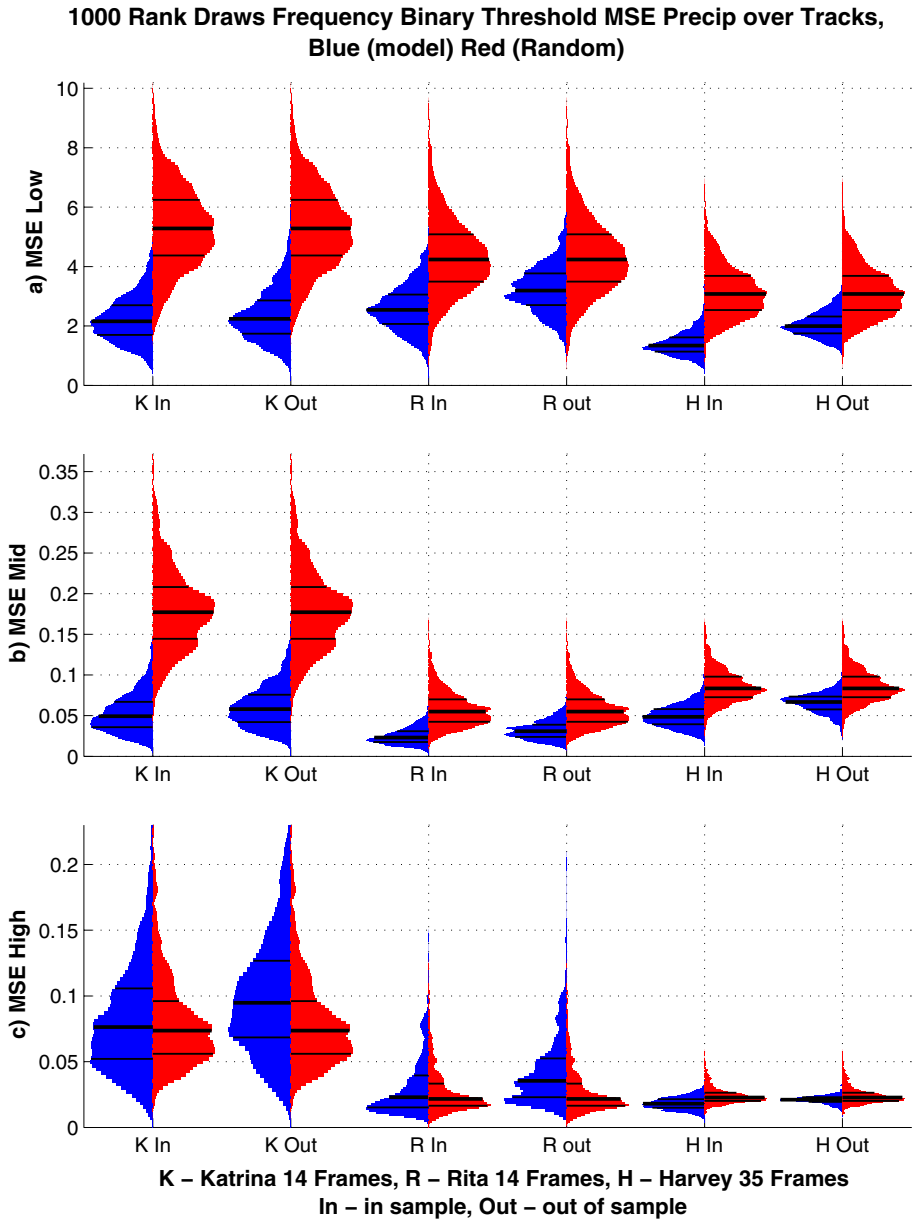


Fig. 8 Frequency binary threshold mean squared error precipitation over tracks within 24 h of landfall to the end of the track. Left blue for model results and right red for random results. Median as a dark black line, 25th and 75th percentile as thin black line

frames as the *k*-NN model: 14 for Katrina, 14 for Rita, and 35 for Harvey and computes the error from the observed frames. The *k*-NN (blue) shows a slightly lower MSE for in-sample tests. Harvey has the lowest out of sample error followed by Rita and Katrina. Harvey also has the most frames and was the slowest moving of the three hurricanes.

Differences between the three hurricanes are largest for the highest precipitation category (c), the largest category.

Figure 9 shows similar to Fig. 8, but for spatial and frequency distributions. Spatial and frequency distributions are evaluated by binning the binary data into boxes and comparing those to the observed (Eq. 6). Harvey and Rita have smaller errors than Katrina for the

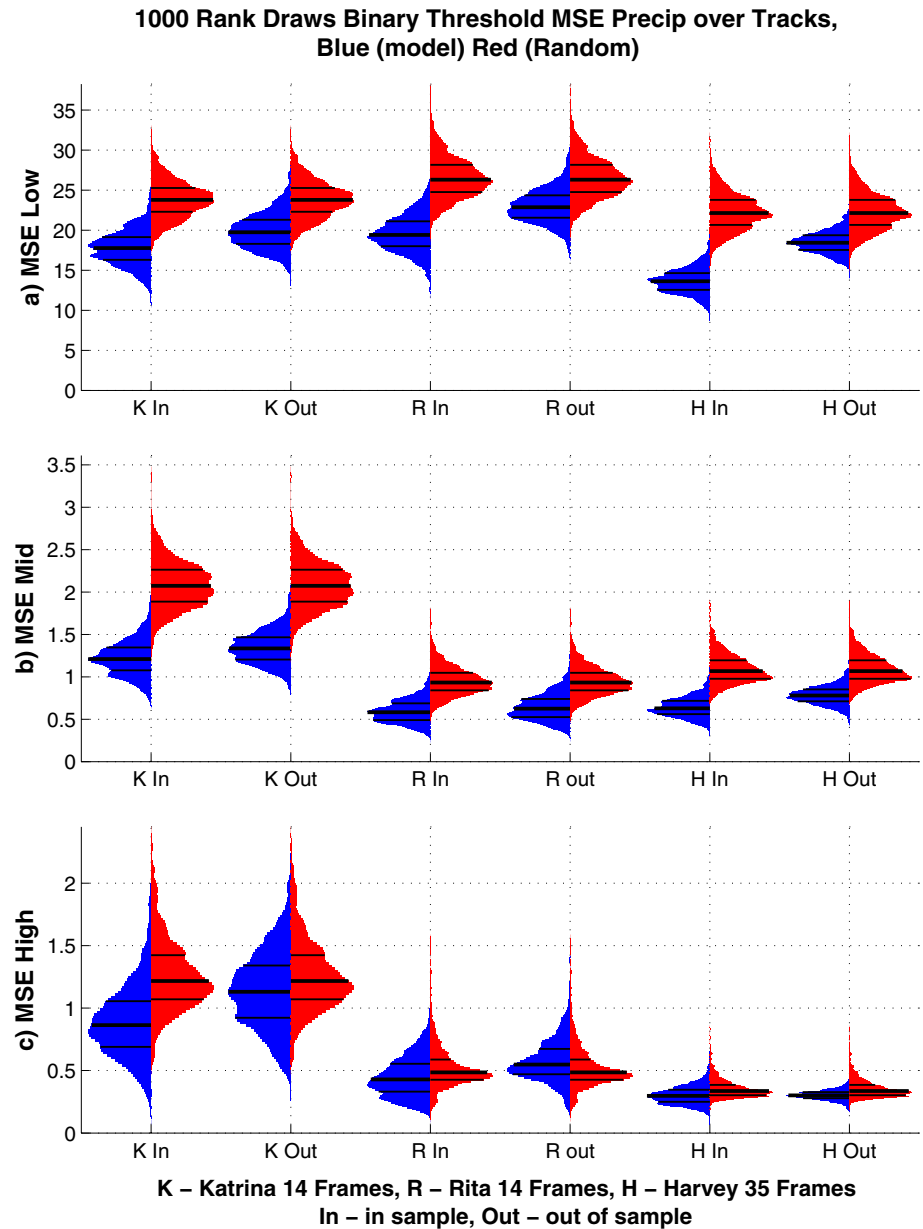


Fig. 9 As in Fig. 8, but for frequency and spatial distribution of mean squared error precipitation

mid and high categories. The center of Katrina is overestimated (Fig. 7a) leading to higher errors in Figs. 8 and 9. Katrina has larger errors in the k -NN model due to the high vertical wind shear leading to a rainfall pattern heaviest in the northwestern quadrant (Fig. 6a). Sharma and Varma (2022) found in high environmental shear, the eye is shifted at the cloud top level toward the heaviest precipitation quadrant. Harvey and Rita have more average error patterns representative of the full set of Gulf coast landfalling TCs that primarily make landfall under a low vertical wind shear environment.

5.4 Gulf coast landing tropical cyclones 1983–2019

Using the MSE threshold evaluation of Sect. 4.2, Fig. 10 shows the results for all Gulf coasting landing tropical cyclones during the PERSIANN-CDR-CCS period of 1983–2019. Blue is the k -NN model precipitation, and red is a random draw of precipitation frames over low a and b), mid c and d), and high e and f) categorized precipitation amounts. The largest median skill separation between model and random appears in the low category spatial and frequency distributions (b) and the lowest in the high category frequency only distributions (e). Overall, the high precipitation areas (e and f) have a lower MSE than the low precipitation areas (a and b). This is a promising result as the high precipitation category contributes more to the overall TC amount than the low precipitation category. However, there are many more low precipitation pixels than the high.

5.5 Linking k -NN precipitation and C^3 HITS for sample 2020 year

Individual sample C^3 HITS run of 2020 using the observed number of 2020 TCs as explained in Sect. 4.5. The N. Atlantic Gulf landfalling TCs are recorded 24 h before landfall to the track end in 6-h increments. The observed 2020 had eight Gulf landfalls, but our simulation had 6. The six tracks have 124 6-h frames. The median of the 1000 rank draws is computed and summed over the 124 frames giving the total precipitation of Gulf coast landfalling TCs over the simulated 2020 season (Fig. 11). Many C^3 HITS can be run with many median or maximum k -NN simulated precipitation seasons resulting in theoretical bounds on precipitation amounts from Gulf landfalling hurricanes from a particular season for risk analysis.

6 Summary and future work

The precipitation of Gulf coast landfalls was simulated over the PERSIANN-CDR-CCS data period including in-sample and out-of-sample examinations of Hurricanes Katrina, Rita, and Harvey with those showing more skill during standard low shear Gulf coast climate conditions. Historically statistical TC precipitation models of CLIPPER origin have significantly underestimated peak precipitation (Marks 2003), but using the PERSIANN-CDR-CCS data as a base for this new precipitation model, Fig. 7 shows an underestimation of the low precipitation (Fig. 6) fringe edges but not in the peak areas.

Benefits of this k -NN precipitation model include the addition of correlated predictors does not cause a problem as it does in a regression model, the precipitation frames are selected only by predictors so no cross-verification is needed, and the inclusion of the satellite-based neural network and climate data record enhanced precipitation product PERSIANN-CDR-CCS results in better-defined precipitation fields than previous

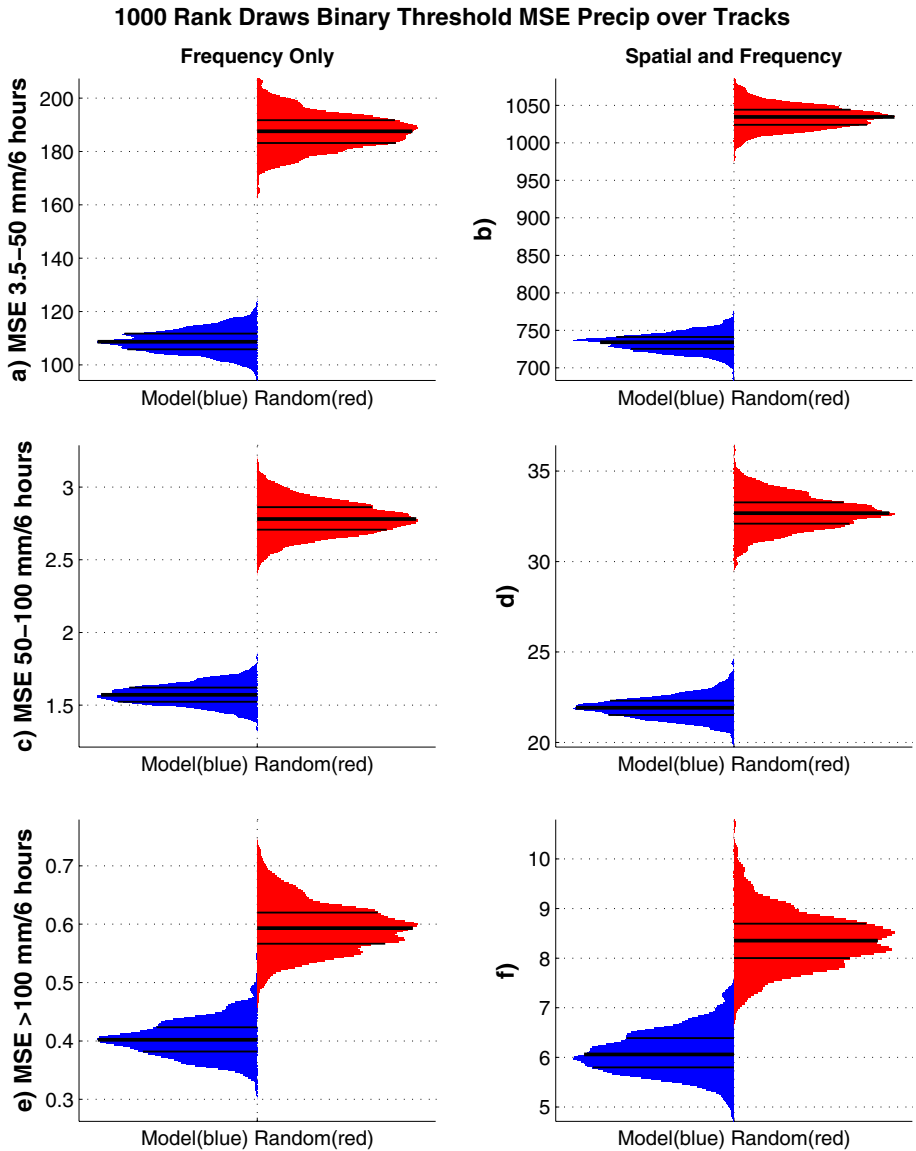


Fig. 10 Frequency only and frequency and spatial binary threshold mean squared error precipitation over all tracks within 24 h of landfall to the end of the tracks. Left blue for model results and right red for random results. Median as a dark black line, 25th and 75th percentile as thin black line

satellite or radar-based statistical-based TC precipitation models. Drawbacks include a relatively short database of TCs from 1983 to 2019 that had 84 Gulf landfalling tracks. This is sufficient for modeling purposes, but more tracks would increase the possible choices of precipitation data frame neighbors to select. With time, more tracks can be added. Also, a decision was made not to include east coast landfalling tracks due to synoptic differences in shear with more encountering mid-latitude systems. Shear changes

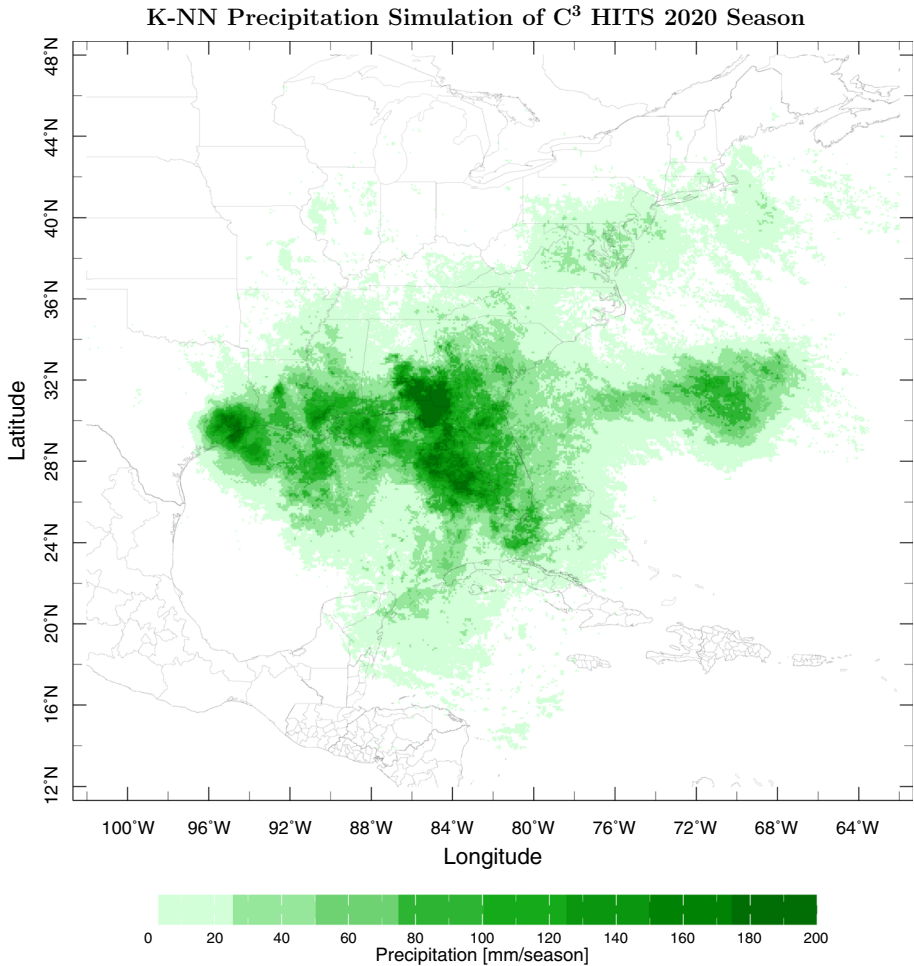


Fig. 11 Median of k -NN model precipitation of an example C³ HITS 2020 N. Atlantic hurricane season

the precipitation amount and distribution, as illustrated in Fig. 6a in relation to the precipitation from Hurricane Katrina.

Future work includes fully linking the precipitation model to both the C³ HITS statistical model of N. Atlantic hurricane tracks and tracks derived using model future scenarios and exploring east coast landfalls. The benefit of linkage to the statistical model of hurricane tracks is that full distributions of TC precipitation can be used in a risk analysis that can then be used with rainfall-runoff models and topographic and land surface models to assess the potential runoff characteristics. Furthermore, the data can be combined with a surge model for full risk analysis of TC-driven flood hazards. Across all k -NN precipitation model ensembles, potential hot spots of very high precipitation corresponding to flat land areas could be highlighted for potential protection. A financial risk analysis of flooding is beneficial for the insurance industry and government entities.

Funding This work was funded by Jupiter Intelligence under Grant Number: JI CU18-0639.

Declarations

Competing interests The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ashouri H, Hsu KL, Sorooshian S, Braithwaite DK, Knapp KR, Cecil LD et al (2015) PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull Am Meteor Soc* 96(1):69–83. <https://doi.org/10.1175/BAMS-D-13-00068.1>
- Atallah EH, Bosart LF (2003) The extratropical transition and precipitation distribution of Hurricane Floyd (1999). *Mon Weather Rev* 131(6):1063–1081. [https://doi.org/10.1175/1520-0493\(2003\)131%3C1063:TETAPD%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131%3C1063:TETAPD%3E2.0.CO;2)
- Brackins JT, Kalyanapu AJ (2020) Evaluation of parametric precipitation models in reproducing tropical cyclone rainfall patterns. *J Hydrol* 580:124255. <https://doi.org/10.1016/j.jhydrol.2019.124255>
- Casati B, Ross G, Stephenson DB (2004) A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorolog Appl: J Forecast Pract Appl Train Tech Model* 11(2):141–154. <https://doi.org/10.1017/S1350482704001239>
- Chan KT (2019) Are global tropical cyclones moving slower in a warming climate? *Environ Res Lett* 14(10):104015. <https://doi.org/10.1088/1748-9326/ab4031>
- Chan KT, Chan JC, Wong WK (2019) Rainfall asymmetries of landfalling tropical cyclones along the South China coast. *Meteorol Appl* 26(2):213–220. <https://doi.org/10.1002/met.1754>
- Corbosiero KL, Molinari J (2003) The relationship between storm motion, vertical wind shear, and convective asymmetries in tropical cyclones. *J Atmos Sci* 60(2):366–376. [https://doi.org/10.1175/1520-0469\(2003\)060%3C0366:TRBSMV%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(2003)060%3C0366:TRBSMV%3E2.0.CO;2)
- Data Announcement 88-MGG-02 (1988) Digital relief of the Surface of the Earth. NOAA, National Geophysical Data Center, Boulder, Colorado
- Emanuel KA, Sundararajan R, Williams J (2008) Hurricanes and global warming: results from downscaling IPCC AR4 simulations. *Bull Am Meteor Soc* 89:347–367. <https://doi.org/10.1175/BAMS-89-3-347>
- Emanuel KA (2017) Assessing the present and future probability of Hurricane Harvey's rainfall. *Proc Natl Acad Sci* 114(48):12681–12684. <https://doi.org/10.1073/pnas.1716222114>
- Feldman M, Emanuel K, Zhu L, Lohmann U (2019) Estimation of Atlantic tropical cyclone rainfall frequency in the United States. *J Appl Meteor and Climate* 58:1853–1866. <https://doi.org/10.1175/JAMC-D-19-0011.1>
- Fewster RM (2009) A simple explanation of Benford's law. *Am Stat* 63(1):26–32. <https://doi.org/10.1198/tast.2009.0005>
- Gangopadhyay S, Harding BL, Rajagopalan B, Lukas JJ, Fulp TJ (2009) A nonparametric approach for paleohydrologic reconstruction of annual streamflow ensembles. *Water Resour Res* 45(6). <https://doi.org/10.1029/2008WR007201>
- Geoghegan KM, Fitzpatrick P, Kolar RL, Dresback KM (2018) Evaluation of a synthetic rainfall model, P-CLIPER, for use in coastal flood modeling. *Nat Hazards* 92(2):699–726. <https://doi.org/10.1007/s11069-018-3220-4>
- Goldenberg SB, Landsea CW, Mestas-Núñez AM, Gray WM (2001) The recent increase in Atlantic hurricane activity: causes and implications. *Science* 293(5529):474–479. <https://doi.org/10.1126/science.1060040>
- Goodman W (2016) The promises and pitfalls of Benford's law. *Significance* 13(3):38–41. <https://doi.org/10.1111/j.1740-9713.2016.00919.x>

- Gray WM (1984) Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb quasi-biennial oscillation influences. *Monthly Weather Rev* 112(9):1649–1668. [https://doi.org/10.1175/1520-0493\(1984\)112%3C1649:ASHFPI%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1984)112%3C1649:ASHFPI%3E2.0.CO;2)
- Hong Y, Hsu KL, Sorooshian S, Gao X (2004) Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J Appl Meteorol* 43(12):1834–1853. <https://doi.org/10.1175/JAM2173.1>
- 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
- Joannes-Boyau R, Bodin T, Scheffers A, Sambridge M, May SM (2015) Using Benford’s law to investigate Natural Hazard dataset homogeneity. *Sci Rep* 5:12046. <https://doi.org/10.1038/srep12046>
- Karlsson M, Yakowitz S (1987) Nearest-neighbor methods for nonparametric rainfall-runoff forecasting. *Water Resour Res* 23(7):1300–1308. <https://doi.org/10.1029/WR023i007p01300>
- Kleiber W, Sain S, Madaus L, Harr P (2020) Stochastic tropical cyclone precipitation field generation. <https://doi.org/10.48550/arXiv.2011.09918>
- Konrad CE, Perry LB (2010) Relationships between tropical cyclones and heavy rainfall in the Carolina region of the USA. *Int J Climatol: J R Meteorol Soc* 30(4):522–534. <https://doi.org/10.1002/joc.1894>
- Kossin JP (2008) Is the North Atlantic hurricane season getting longer? *Geophys Res Lett* 35(23). <https://doi.org/10.1029/2008GL036012>
- Kossin JP (2018) A global slowdown of tropical-cyclone translation speed. *Nature* 558(7708):104–107. <https://doi.org/10.1038/s41586-018-0158-3>
- Lall U, Rajagopalan B, Tarboton DG (1996) A nonparametric wet/dry spell model for resampling daily precipitation. *Water Resour Res* 32(9):2803–2823. <https://doi.org/10.1029/96WR00565>
- Lall U, Sharma A (1996) A nearest neighbor bootstrap for time series resampling. *Water Resour Res* 32(3):679–693. <https://doi.org/10.1029/95WR02966>
- Lai Y, Li J, Gu X, Chen YD, Kong D, Gan TY et al (2020) Greater flood risks in response to slowdown of tropical cyclones over the coast of China. *Proc Natl Acad Sci* 117(26):14751–14755. <https://doi.org/10.1073/pnas.1918987117>
- Landsea C, Franklin J, Beven J (2015) The revised Atlantic hurricane database (HURDAT2). <https://www.nhc.noaa.gov/data/hurdat/hurdat2-format-atlantic.pdf> Accessed 25 Apr 2021
- Langousis A, Veneziano D (2009) Theoretical model of rainfall in tropical cyclones for the assessment of long-term risk. *Journal of Geophysical Research: Atmospheres* 114(D2). <https://doi.org/10.1029/2008JD010080>
- Lonfat M, Rogers R, Marchok T, Marks FD Jr (2007) A parametric model for predicting hurricane rainfall. *Mon Weather Rev* 135(9):3086–3097. <https://doi.org/10.1175/MWR3433.1>
- Lu P, Lin N, Emanuel K, Chavas D, Smith J (2018) Assessing hurricane rainfall mechanisms using a physics-based model: Hurricanes Isabel (2003) and Irene (2011). *J Atmos Sci* 75(7):2337–2358. <https://doi.org/10.1175/JAS-D-17-0264.1>
- Mahalanobis PC (2018) On the generalized distance in statistics. *Sankhyā: Indian J Stat Ser A* (2008-) 80: S1–S7
- Marks Jr FD (2003) Development of a Tropical Cyclone Rainfall Climatology and Persistence (R-CLIPER) Model. Final Report. https://origin.www.nhc.noaa.gov/jht/final_rep/R-CLIPER_final.pdf Accessed 23 Sep 2021
- Nakamura J, Kushnir Y, Lall U, Harr PA, McCreery K (2021) Early season hurricane risk assessment: climate-conditioned HITS simulation of north atlantic tropical storm tracks. *J Appl Meteor and Climatol* 60:559–575. <https://doi.org/10.1175/JAMC-D-20-0237.1>
- Nakamura J, Lall U, Kushnir Y, Rajagopalan B (2015) HITS: hurricane intensity and track simulator with North Atlantic ocean applications for risk assessment. *J Appl Meteor Climatol* 54:1620–1636. <https://doi.org/10.1175/JAMC-D-14-0141.1>
- National Hurricane Center (NHC) (2020) Tropical cyclone advisory archive. <https://www.nhc.noaa.gov/archive/2020/> Accessed 14 Feb 2021
- National Hurricane Center (NHC) (2021) Tropical cyclone advisory archive. <https://www.nhc.noaa.gov/archive/2021/> Accessed 1 Sept 2021
- National Weather Service (NWS), Mobile/Pensacola Office (2020) Hurricane Sally September 15–16, 2020. https://www.weather.gov/mob/sally_rainfall. Accessed 27 Feb 2021
- Nguyen P, Shearer EJ, Tran H, Ombadi M, Hayatbini N, Palacios T et al (2019) The CHRS Data Portal, an easily accessible public repository for PERSIANN global satellite precipitation data. *Sci Data* 6(1):1–10. <https://doi.org/10.1038/sdata.2018.296>

- Rajagopalan B, Lall U (1999) A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resour Res* 35(10):3089–3101. <https://doi.org/10.1029/1999WR900028>
- Rappaport EN (2014) Fatalities in the United States from Atlantic tropical cyclones: new data and interpretation. *Bull Am Meteor Soc* 95(3):341–346. <https://doi.org/10.1175/BAMS-D-12-00074.1>
- Sharma N, Varma AK (2022) Impact of vertical wind shear in modulating tropical cyclones eye and rainfall structure. *Nat Hazards* 112:2083–2100. <https://doi.org/10.1007/s11069-022-05257-3>
- Touma D, Stevenson S, Camargo SJ, Horton DE, Diffenbaugh NS (2019) Variations in the intensity and spatial extent of tropical cyclone precipitation. *Geophys Res Lett* 46(23):13992–14002. <https://doi.org/10.1029/2019GL083452>
- Villarini G, Zhang W, Miller P, Johnson DR, Grimley LE, Roberts HJ (2022) Probabilistic rainfall generator for tropical cyclones affecting Louisiana. *Int J Climatol* 42(3):1789–1802. <https://doi.org/10.1002/joc.7335>
- Xi D, Lin N, Smith J (2020) Evaluation of a physics-based tropical cyclone rainfall model for risk assessment. *J. Hydrometeorology* 21:2197–2217. <https://doi.org/10.1175/JHM-D-0035.1>
- Zhu L, Frauenfeld OW, Quiring SM (2013) Seasonal tropical cyclone precipitation in Texas: a statistical modeling approach based on a 60-year climatology. *J Geophys Res: Atmos* 118(16):8842–8856. <https://doi.org/10.1002/jgrd.50663>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.