

MODELING RETURN PERIODS OF TROPICAL CYCLONES
IN THE VICINITY OF HAWAI'I

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I AT MĀNOA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ATMOSPHERIC SCIENCES

October 2021

By

Xinping Zhang

Thesis Committee:

Pao-Shin Chu, Chairperson
Jennifer D. Small Griswold
Giuseppe Torri

Keywords: Return period, Statistical model, Tropical cyclones, Poisson regression

ACKNOWLEDGMENTS

I would like to thank all of you helping me in writing and revising my thesis. Firstly, professor Pao-Shin Chu is the deepest gratitude during my thesis working. He helped me in every step of finishing the thesis. He always gives me inspiration, encouragement, and confidence. Without his help, my thesis would not have been in its present form. Secondly, I would like to thank my thesis committee members, professor Jennifer D. Small Griswold and professor Giuseppe Torri. They give me valuable comments in revising thesis. Finally, I would like to thank my family for their love and support. I also want to thank my friends who have given me help during my difficult time.

ABSTRACT

The tropical cyclone is one of the most destructive natural disasters in Hawai'i. Tropical cyclones have caused a great number of damages on human lives and economies. The return periods of tropical cyclones in the vicinity of Hawai'i can give the important climate information to the government, insurance companies, engineers, and publics. This thesis developed a statistical model to estimate the return period of tropical cyclones in the vicinity. It contains the Poisson regression, the Gaussian-kernel density estimate, the extreme value distribution, and statistical resampling methods. Comparing the results in this model with the previous research based on the dataset from 1970 to 1995, high intensity of tropical cyclones' return periods becomes much shorter indicating a higher risk from tropical cyclone induced winds in Hawai'i.

Table of Contents

CHAPTER 1. INTRODUCTION	1
CHAPTER 2. DATA	3
2.1 Tropical cyclone data.....	3
2.2 Climate indices	5
2.3 Reanalysis	6
CHAPTER 3. STATISITCAL METHODs	6
3.1 Poisson process.....	6
3.2 Poisson regression	7
3.2.1 Poisson regression with climate indices as covariate factors.....	7
3.2.2 Poisson regression with SOI and circulation factors as covariate factors	8
3.2.3 Regression result.....	12
3.3 Model regression parameters	16
3.4 Kernel density estimate.....	16
3.5 Extreme value distributions	18
3.5.1 Generalized Pareto distribution	18
3.5.2 Generalized Extreme Value distribution.....	20
3.6 A Statistical model.....	21
CHAPTER 4. RESULT AND UNCERTAINTY ANALYSIS	21
4.1 Simulation results	21
4.2 Uncertainty analysis	23
CHAPTER 5. SUMMARY	23
References	25

CHAPTER 1. INTRODUCTION

Tropical cyclones (TCs) can cause major disasters for human lives and economies. For the Hawaiian Islands, a few TCs have struck or were very close to the islands and some of them resulted in great damage. In the past, 39 TCs were found in the vicinity of Hawai'i from 1966 to 2019 (Fig. 1a). Here we only consider TCs reaching at least tropical storm intensity (≥ 34 knots or named storms) in their lifespans. The definition of the vicinity of Hawai'i is the circular area with a radius of 250 n mi from Honolulu (Fig. 1b). TC season of Hawai'i mainly runs from July to October, with 95% of the total TCs occurring in these four months (Fig. 1c).

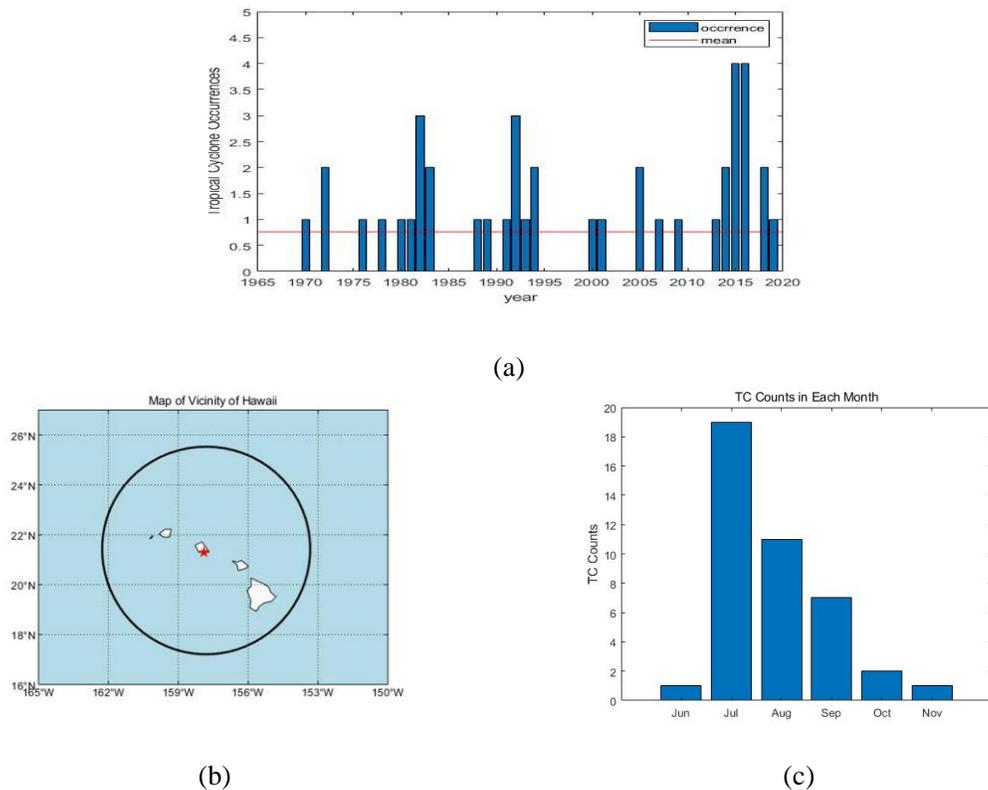


Fig.1 (a) Time series of TCs passing through the vicinity of Hawai'i. The blue rectangular represents the real number of TCs in each year from 1966 to 2019. The red line shows average the number of TC per year. (b) A cycle indicates the vicinity of Hawai'i. The red star denotes the location of Honolulu. (c) The seasonality of TCs in the vicinity of Hawai'i from June to November.

Hurricane Iniki (September, 6-13 September 1992) is the most destructive hurricane directly striking Hawai'i which caused over a hundred casualties and the total economic damages about \$3.19 billion in 1992. The damage brought by Iniki was widely spread around Kauai and western Oahu. Over 50 percent of the electric and telephone lines on Kauai were destroyed (National

Weather Service, 1993). Hurricane Lane (August 15-29 August 2018) is another intense storm that was very close to Oahu. The maximum wind speeds reached 140 knots and reduced to 105 knots when the storm passed through the cycle. It is the wettest TC in record in Hawai'i, simultaneously leading to flooding and landslides with heavy rainfall on the Island of Hawai'i (Nugent et. al, 2020).

Return period is one information for TC risks which is important for a variety of fields. To reduce losses by TC striking, engineers and managers involved in the coastal building programs need to know the TC risk information. Insurance companies can adjust premiums based on TC risk information. Governments can also make the natural disaster policy using this information.

In general, return period is defined as

$$R(x) = \frac{1}{\omega(1 - F(x))}, \quad (1.1)$$

where ω means the average sampling frequency, and $F(x)$ is the cumulative distribution function (CDF) evaluated at x . $R(x)$ is explained as the average waiting time between occurrences of events that reach the specific magnitude x (e.g., 80 knots) or over this value (Wilks, 2011). For TCs, return period of a specific wind speed TC is one of the TC risk information which is defined as the average waiting time of this fixed intensity TC that would recur once in how many years (Chu and Wang, 2008).

Previous research has developed many methods on estimating the return periods of TCs. The Gumbel distribution and the Weibull distribution were frequently applied in the early estimate methods (Rupp and Lander, 1996, Jagger et al., 2001). Jagger and Elsner (2006) used the Generalized Pareto Distribution (GPD) to estimate the return period of the maximum wind speeds associated with TCs near the U. S. coasts conditional on the North Atlantic Oscillation, El Niño, Atlantic Multidecadal Oscillation, and global temperature separately. Based on synthetic storm events, an empirical approach was developed by Emanuel and Jagger (2010) to estimate hurricanes return periods. For Hawai'i, Chu and Wang (1998) estimated the return periods of relative intensity and maximum wind speeds associated with TCs using the lognormal and Gumbel distributions and the Monte Carlo simulation.

The motivation of this these is to update the return periods of TCs in the vicinity of Hawai'i (Chu and Wang, 1998) using a modern statistical model advocated by Parisi and Lund (2008). In this research, the return period of Hawaii's TCs is estimated using the historical data introduced

in Chapter 2 and the statistical model described in Chapter 3. The results and analysis procedure are detailed in Chapter 4. Finally, the summary is given in Chapter 5.

CHAPTER 2. DATA

2.1 Tropical cyclone data

The revised Northeast and North Central Pacific hurricane database (HURDAT2) from the National Hurricane Center (NHC) and the Central Pacific Hurricane Center (CPHC) are used in the study. The data cover an area from the west coast of North and Central America westward to 180 °W over the North Pacific Ocean, including Hawai’i. Considering the accuracy of the data, only the dataset from satellite observation from 1966 is used.

Table 1 lists the TCs passing through the defined circle (Fig. 1b) during Hawai’i TC season from 1966 to 2019. Historically, seven TCs, Fico (9 July to 28 July 1978), Iniki (6 September to 13 September 1992), Emilia (16 July to 25 July 1994), Ana (13 October to 26 October 2014), Ignacio (23 August to 8 September 2015), Lester (24 August to 8 September 2016), and Lane (15 August to 29 August 2018), reached the hurricane intensity (wind speeds over 64 knots and denotes in the red color). Note that the magnitude of these wind speeds is the maximum value detected in the defined circle, and the maximum wind speeds of each TC could be larger outside the circle but is reduced when passing through the defined area.

Table 1. List of Hawai’i Tropical Cyclones

name	lifetime	Max Wind speeds (knots)
Maggie	8/20-8/27/1970	45
Diana	8/11-8/20/1972	60
Fernanda	8/20-9/1/1972	45
Gwen	8/5-8/18/1976	25
Fico	7/9-7/28/1978	100
Kay	9/16-9/30/1980	30
Jova	9/14-9/21/1981	35
Daniel	7/7-7/22/1982	25

Table 1. (continued)

Gilma	7/26-8/2/1982	30
Gil	7/23-8/5/1983	40
Raymond	10/8-10/20/1983	30
Gilma	7/28-8/3/1988	20
Dalilia	7/11-7/21/1989	55
Fefa	7/29-8/8/1991	25
Georgette	7/14-7/26/1992	30
Orlene	9/2-9/14/1992	25
Iniki	9/6-9/13/1992	125
Eugene	7/15-7/25/1993	30
Emilia	7/16-7/25/1994	85
No name	8/9-8/14/1994	30
Daniel	7/23-8/5/2000	60
Jova	9/12-9/25/2005	20
Kenneth	9/14-9/30/2005	25
Flossie	8/8-8/16/2007	50
Felicia	8/3-8/11/2009	30
Flossie	7/25-7/30/2013	35
Iselle	7/30-8/10/2014	60
Ana	10/13-10/26/2014	75
Ela	7/7-7/12/2015	20
Guillermo	7/27-8/8/2015	35
Ignacio	8/23-9/8/2015	65
Jimena	8/25-9/10/2015	30
Celia	7/6-7/21/2016	40

Table 1. (continued)

Darby	7/11-7/26/2016	40
Howard	7/31-8/7/2016	25
Lester	8/24-9/8/2016	70
Lane	8/15-8/29/2018	105
Olivia	9/1-9/14/2018	45
Flossie	7/28-8/7/2019	30

2.2 Climate indices

There are six climate indices running from 1966 to 2019 from National Oceanic and Atmospheric Administration (NOAA) in this study:

- (1) The Pacific Meridional Model (PMM) describes the meridional variability in the tropical-subtropical eastern Pacific Ocean, defined by the maximum covariance analysis to the sea surface temperature (SST) and the zonal and meridional components of the 10m wind field over the time period 1950-2005 using the NCEP/NCAR Reanalysis datasets.
- (2) The North Atlantic Oscillation (NAO) is a teleconnection pattern in the North Atlantic, Europe, and North America. It reflects fluctuations in the surface pressure between the Icelandic low and the Azore high, and the NAO index is usually defined through changes in atmospheric pressures between the center of these two weather systems.
- (3) The Southern Oscillation Index (SOI) represents the difference in average air pressure between Tahiti and Darwin. The term El Niño Southern Oscillation (ENSO) describes the ocean-atmosphere interaction and the relationship between oceanic El Niño and the atmospheric Southern Oscillation phenomenon.
- (4) The Pacific Decadal Oscillation (PDO) is the leading principal component of monthly SST anomalies in the North Pacific (20 °N to 70 °N). The PDO signals wax and wane approximately every 20 to 30 years from one polarity to another.
- (5) The Quasi-Biennial Oscillation (QBO) index is obtained from the zonal average of the 30hPa zonal wind at the equator.

(6) The Atlantic Multidecadal Oscillation (AMO) is defined from the SST patterns in the North Atlantic after removing linear trend in the data. The linear trend is intended to represent the externally forced, greenhouse gas induced global warming trend.

For each monthly climate indices, we use the concurrent period with Hawai'i TC season, which means the average value of July to October climate indices.

2.3 Reanalysis data

Monthly mean horizontal wind, SST, and relative humidity (RH) from the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis 1 dataset are used in this study. The resolution of grid is 2.5° longitude \times 2.5° latitude. The preprocessing method of the NCEP/NCAR dataset is the same as other climate indices.

CHAPTER 3. STATISITCAL METHODs

In this study, a statistical model developed by Parisi and Lund (2008) is applied and modified to fit for the Hawai'i TC records. It contains three parts:

- (1) A Poisson regression method is used to estimate the number of TCs in a given year.
- (2) A Kernel density estimate is used to model the arrival dates of TCs in that year.
- (3) The extreme value distribution is used to model the maximum wind speeds of each TC.

3.1 Poisson process

The Poisson distribution is a common statistical method to describe the occurrences of discrete events and the counts are non-negative integer values. The Poisson distribution could be understood as a limiting case of the binomial distribution with trials approaching infinity and the probability of the occurrence of the event of success on any one trial approaches zero (i.e., rare event).

The Poisson process is a proper method to describe the TC frequency. TCs' characters are satisfied to be a Poisson process because: 1) TC occurrences are discrete events. 2) TC occurrences are independent with each other. 3) TC events are rare enough that the probability of more than one TCs striking simultaneously in a limited region is very small. The probability density function (PDF) of y TCs occurring in one year is:

$$P(y) = \frac{\lambda^y \exp(-\lambda)}{y!}, \quad y = 1, 2, 3, \dots, \quad (3.1)$$

where λ is the Poisson parameter or the Poisson mean value, representing the average TC occurrence per year, and y is the TC counts in the specific year.

3.2 Poisson regression

3.2.1 Poisson regression with climate indices as covariate factors

TC occurrence is inhomogeneous in time. The time-homogeneous Poisson process, which means the average TC occurrence rate is fixed, as shown in Eq. (3.1), is suboptimal. An improvement can be made by varying the annual Poisson mean value, λ . A Poisson regression method is involved to model the annual Poisson parameter. Specifically, the Poisson mean rate varies with time and the concurrent climate indices (average value of each index during the period of JASO) are considered as the covariates.

We first calculate the Pearson correlation coefficient between each climate index and TC frequency to find proper regression parameters. Figures 2a, 2b, 2c show that there is no obvious correlation between TC frequency and QBO, NAO, and AMO. The PDO and PMM are positively correlated with TC frequency (Figs. 2d, 2e), while the SOI is negatively correlated with TC frequency (Fig. 2f). To test whether these correlation coefficients are significant, the Z-statistic is used. The Z-statistic is built by Fisher Z transformation which is an approach to test the significance of correlation coefficient. A Z-statistic follows the Gaussian distribution with mean value $\mu = 0$ and variance $\sigma = (t - 3)^{-\frac{1}{2}}$, where t is the number of sample size and the Z-statistic is

$$Z = \frac{1}{2} \ln \left(\frac{1 + R}{1 - R} \right). \quad (3.2)$$

where R is the Pearson correlation coefficient.

For $t=54$ (years), the critical value of the 95% confidence level is ± 0.27 . Only the SOI is significantly correlated with the TC occurrence. In El Niño years, the SOI is large and negative and there is a relatively large probability for TCs striking Hawai'i (Chu and Wang, 1997). In contrast, when SOI is large and positive, referring to as the La Niña year, the probability of TC striking Hawai'i is relatively low.

Taking the average values of JASO SOI as the regression parameter, the Poisson mean value is estimated by the Poisson regression method, as

$$\lambda_i = \exp^{\beta_0 + SOI_i \beta_1}. \quad (3.3)$$

All data points are fitted to a Poisson regression. Referring to the t-ratio and the p-value shown in Table 2, β_0 and β_1 are statistically significant at the 95% confidence level in this Poisson regression fitting. Thus, the SOI is first considered as a significant regression parameter in this model. The other five climate indices are not taken as regression parameters.

To train the regression model, a leave one out cross validation (LOOCV) is used to test the regression result. Cross validation simulates prediction by repeating the entire fitting procedure on data subsets, and then examining the predictions made for the data portions left out of each subset. The LOOCV is to leave only one data out during the fitting procedure and to make the sample size of each subset as n-1. By repeating the procedure, the results are n different prediction equations. The n predicted Poisson parameters are estimated separately by these n equations. Figure 3 shows the time series of the observed TC counts and the series of estimated Poisson parameters using the LOOCV, suggesting that the Poisson regression model does not catch the annual TC frequency accurately. The Pearson correlation coefficient between the observed and predicted TC counts is only 0.28. The Poisson regression (3.3) with only the SOI as regression parameter is not good enough to represent Hawai'i TCs on interannual time scales.

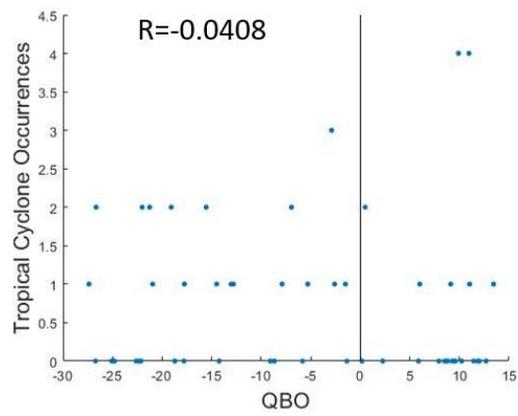
Table 2. The fitting result of Poisson regression using SOI as the parameter

	Estimate	SE	t-statistic	p-value
β_0	-0.42	0.18	-2.38	0.02
β_1	0.13	0.06	2.16	0.03

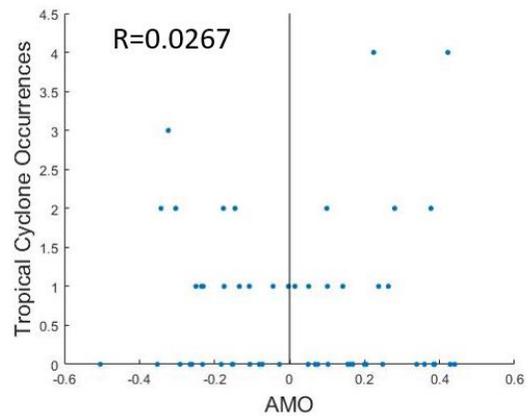
3.2.2 Poisson regression with SOI and circulation factors as covariate factors

To improve the regression model, other regression parameters should be introduced. Adding circulation factors to the regression parameters with SOI might be a scheme to improve the model prediction skill. The possible factors are 850hPa relative vorticity (ξ), 700hPa relative humidity (RH), vertical wind shear (VWS), and sea surface temperature (SST) (Chu, 2002). The 850hPa relative vorticity is computed as

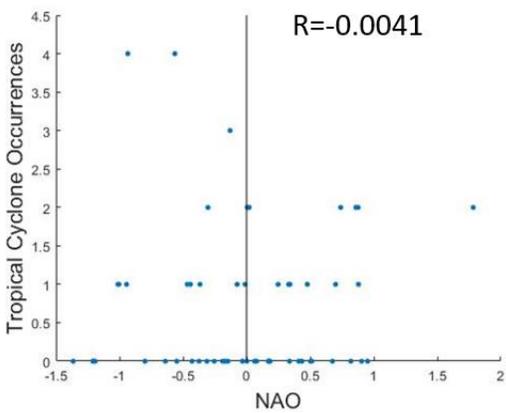
$$\xi = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} + \frac{u}{a} \text{tg}(\varphi). \quad (3.4)$$



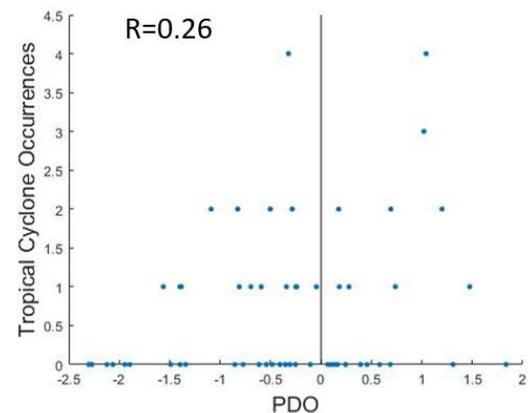
(a)



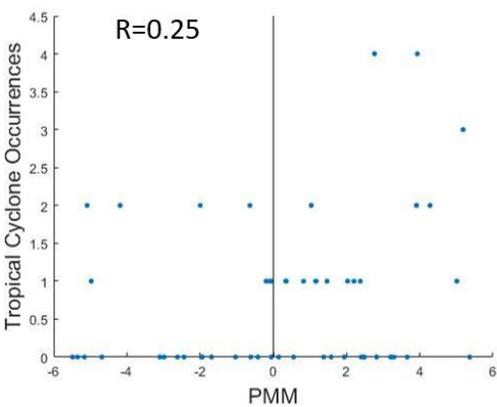
(b)



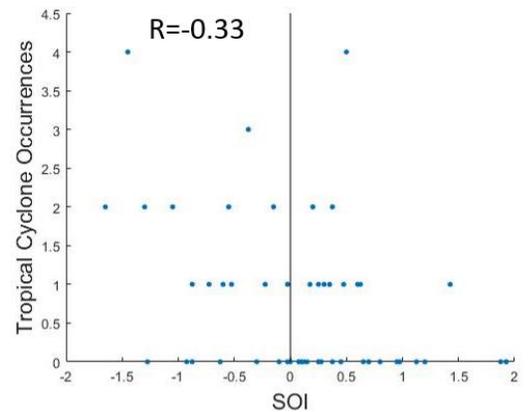
(c)



(d)



(e)



(f)

Fig. 2 The scatter plot of each climate indices an TC occurrence.

(a) QBO, (b) AMO, (c) NAO, (d) PDO, (e) PMM, (f) SOI.

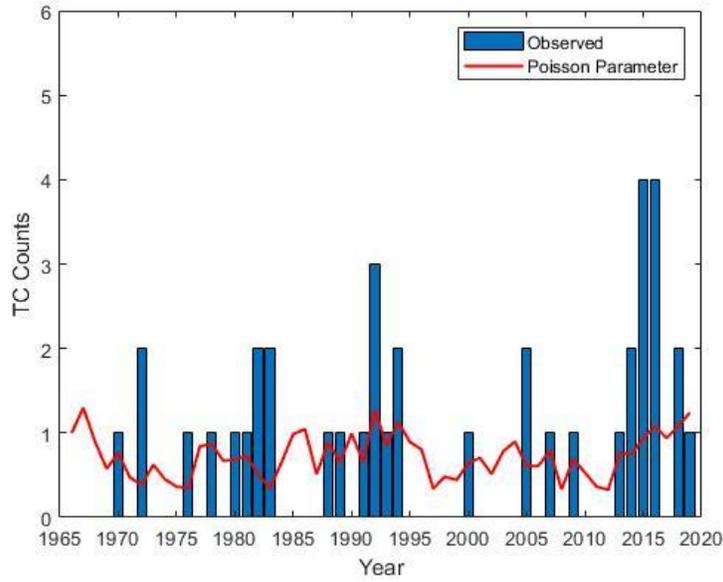


Fig. 3 The time series of historical TC numbers (blue rectangle) and hindcasts (red line).

To find out the critical region of each factor, we will compute the Pearson correlation coefficients between TC counts and each of the circulation factors during JASO over the central North Pacific. Over the southeast of the Hawaiian Islands, an area shows large positive correlation coefficients (0.59) between TC counts and 850hPa relative vorticities (Fig. 4). Considering the resolution of the dataset and the significance of the correlation coefficients, the average value of the relative vorticity in the red box is taken to form the critical region (Fig. 4a). The scatter plot in Figure 4b independently confirms the strong and positive relationship between 850hPa relative vorticity and TC counts. When low-level vorticity is large and positive (i.e., cyclonic vorticity), the circulation is more likely to strengthen the spin-up process, and therefore favor TC development. On the opposite, the TC counts would be lower when the low-level vorticity is largely negative.

There are two steps to do before adding 850hPa vorticity into the regression model. The first one is to clarify whether ξ is independent with the SOI. The Pearson correlation coefficient between ξ and SOI is about zero. The scatter plot of these two factors also shows that they are independent with each other (Fig. 5b). The second one is to ensure the computed relative vorticity is reliable. A climatology map of 850hPa relative vorticity and the wind field during Hawai'i hurricane season is drawn as a reference (Fig. 5a). In the midlatitude, the large anticyclone

circulation is shown, and the subtropical high is spread over the eastern and central North Pacific so the computed values are deemed reliable.

The 850 hPa relative vorticity is introduced in this regression model based on the above tests.

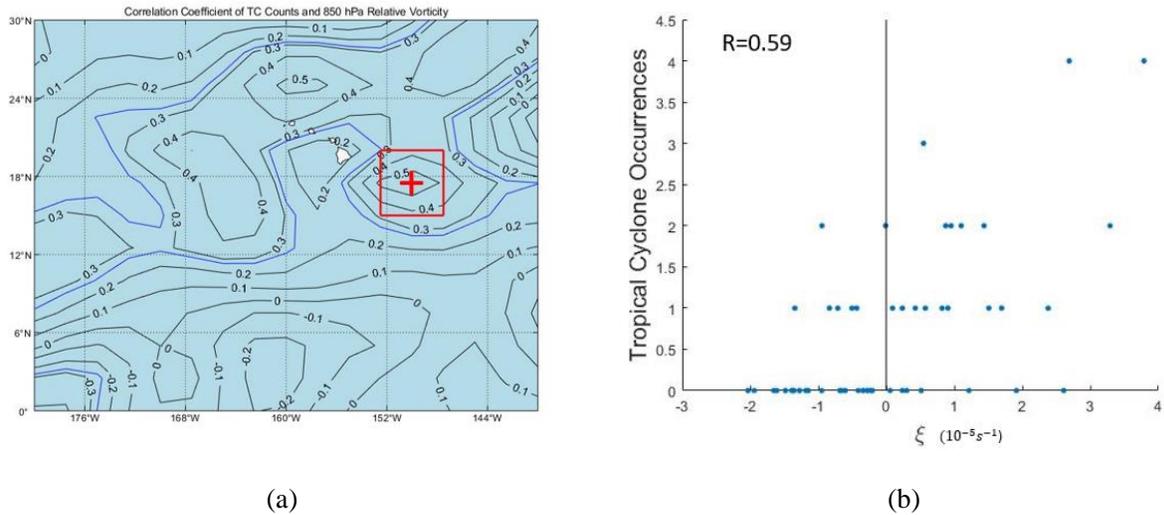


Fig. 4 (a) The correlation map of low-level vorticity and TC numbers. The blue line shows the boundary of significant correlation. Red box shows the domain area. (b) The scatter plot of TC numbers and low-level vorticity.

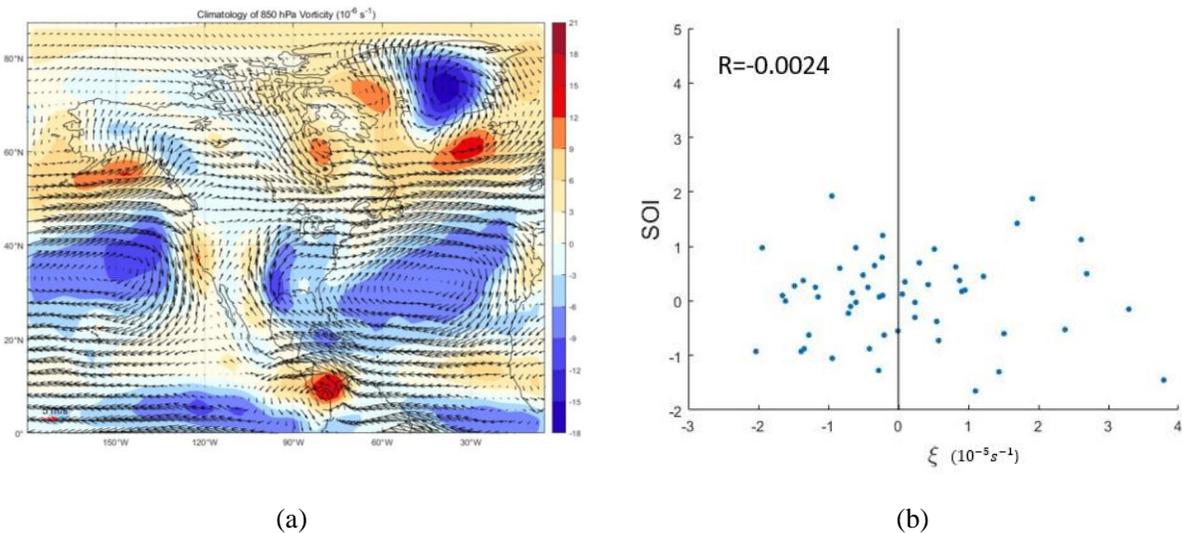


Fig. 5 (a) The climatology map of low-level vorticity. The blue shows the negative value indicating the cyclone circulation and the red area is on the opposite. (b) The scatter plot of SOI and low-level vorticity.

The second circulation factor is the vertical wind shear (VWS) calculated as

$$VWS = \sqrt{(u_{200} - u_{850})^2 + (v_{200} - v_{850})^2}, \quad (3.5)$$

where, u and v are the standard horizontal wind speed components and the subscripts in (3.5) represent the pressure levels. In this correlation map (Fig.6a), to the south of Hawai'i, largely negative correlation is shown, and to the north of Hawai'i, there is an area showing largely positive correlation. Large VWS tends to destroy the development of TCs because of the ventilation effect. The negative correlated area is preferred to be the critical region shown in the red box. The average value in this area is negatively correlated (-0.46) with TC counts (Fig. 6b).

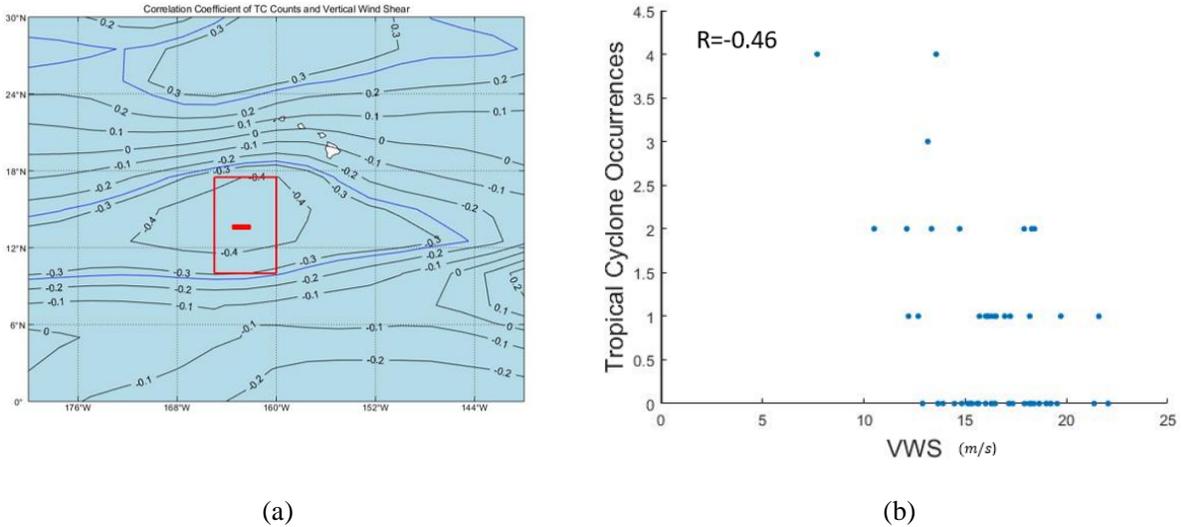


Fig. 6 The same as Fig. 4 but for vertical wind shear.

To avoid overfitting, the independence of VWS and other regression parameters is tested. VWS is positively correlated with SOI ($R = 0.33$) and negatively correlated with 850hPa relative vorticity ($R = -0.54$) shown in Fig. 7. The VWS is not introduced because of its dependency with the SOI and 850hPa relative vorticity.

The relative humidity (RH) is in a similar condition as the VWS. A large and positive correlation coefficient is found over Hawai'i, and the red box is the critical region (Fig. 8a). The scatter plot also confirms this relationship (Fig. 8b). However, RH is not independent from the SOI and 850hPa vorticity (Fig. 9). For this reason, the RH is not added into the existing two-parameter regression model.

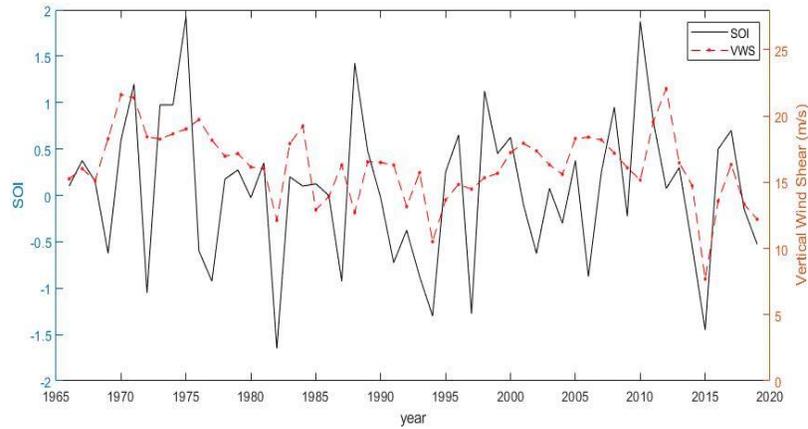
The sea surface temperature (SST) does not show the large correlation with TC counts (Fig. 10) and the SST is not considered as the regression parameter.

3.2.3 Regression result

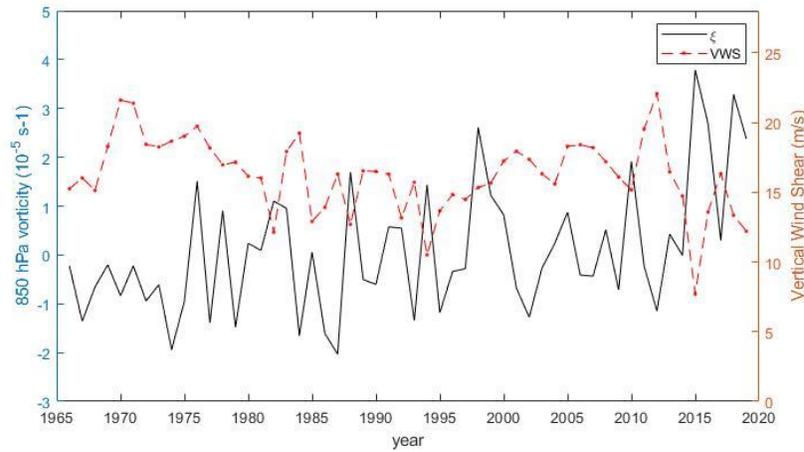
Only the 850hPa relative vorticity (ξ) is added to the initial Poisson regression model, and the Poisson regression parameters include both the SOI and ξ . The new regression function is

$$\lambda_i = \exp^{\beta_0 + \beta_1 SOI_i + \beta_2 \xi_i}, \quad (3.6)$$

Using 54-year data to fit this regression, the result shows that both SOI and ξ are statistically significant (Table 3).

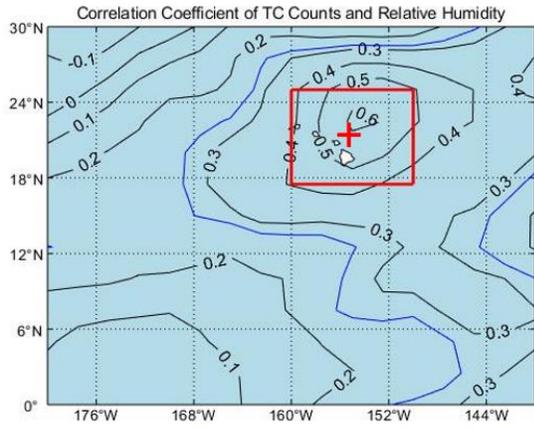


(a)

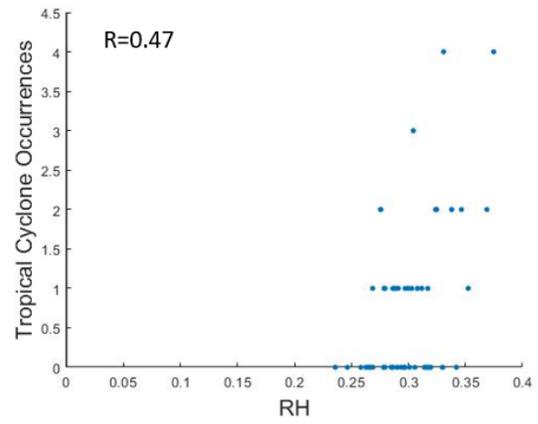


(b)

Fig. 7 The correlation of vertical wind shear with (a) SOI, (b) low-level vorticity.

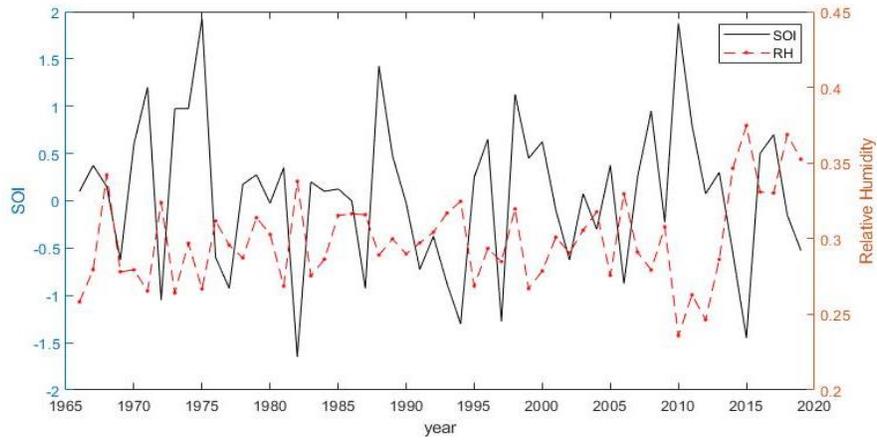


(a)

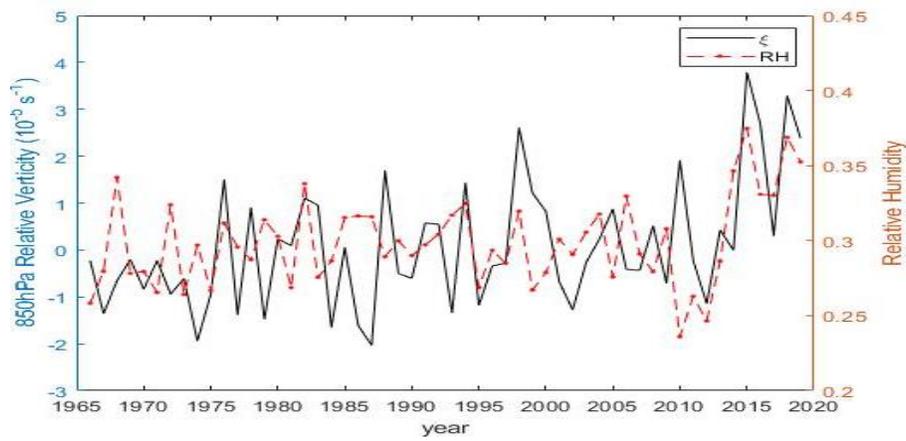


(b)

Fig. 8 The same as Fig. 4 but for relative humidity.



(a)



(b)

Fig. 9 The same as Fig. 7 but for relative humidity.

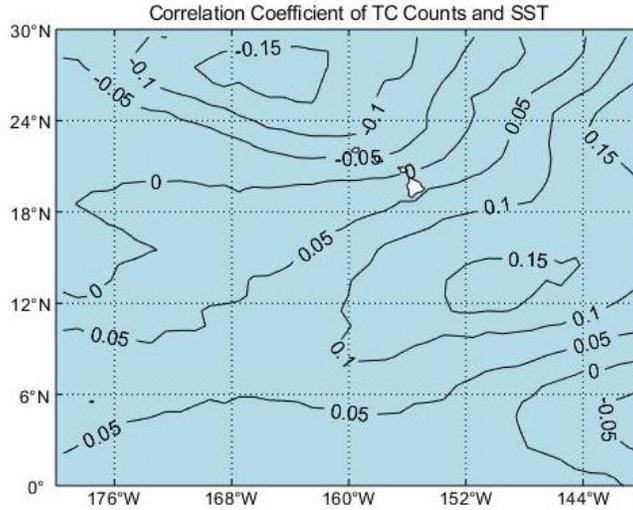


Fig. 10 The same as Fig. 4a but for SST.

The LOOCV is used to train this model. The training result suggests that the new regression model is improved by adding ξ into the Poisson regression model with only SOI (Fig. 11). Most peaks are captured by this regression model. The estimated Poisson mean value is highly correlated with the observed TC numbers and the Pearson correlation coefficient is raised from 0.28 (Fig. 3) to 0.66.

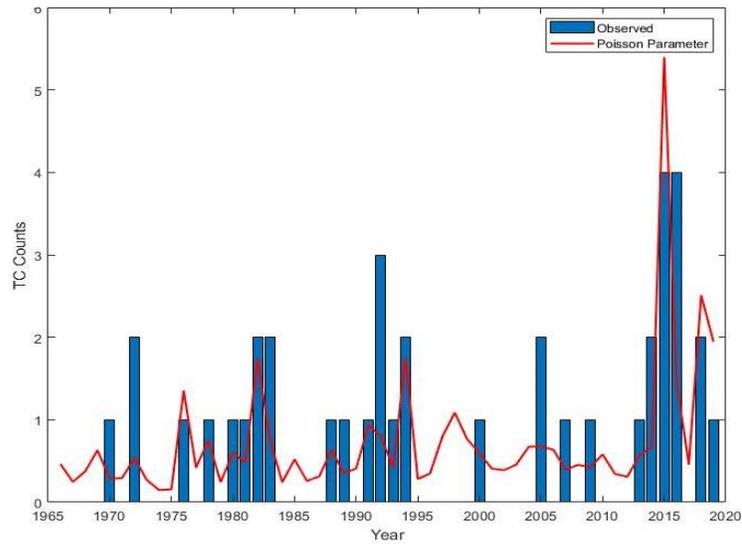


Fig. 11 The time series of historical TC numbers (blue rectangle) and hindcasts (red line).

Table 3. The fitting result of Poisson regression using SOI and low-level vorticity as the parameters

	Estimate	SE	t-statistic	p-value
β_0	-0.63	0.20	-3.17	0.0015
β_1	-0.42	0.20	-2.17	0.03
β_2	0.45	0.10	4.47	7.85e-06

3.3 Model regression parameters

After we have selected the parameters, it is necessary to know the theoretical distribution of these parameters so they can be properly represented in the statistical model (to be explained in Chapter 3.6). For the Gaussian distribution, its PDF is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (3.7)$$

where μ is the mean value of each series and σ is the standard deviation. Here for the SOI, $\mu = 0.06$, and $\sigma = 0.79$, and for vorticity, $\mu = 0.09$ and $\sigma = 1.33$. The goodness of fit is shown through a comparison of the data histogram and the theoretical Gaussian distribution (Fig. 12a) and the quantile-quantile diagram (Fig. 12b). Both the SOI and low-level vorticity follow the Gaussian PDF well. The two Q-Q plots show that most SOI and low-level vorticity points align along the perfect 45° lines. For low-level vorticity, when the value is extremely positive or negative, the theoretical Gaussian distribution is slightly less optimal.

3.4 Kernel density estimate

The kernel density estimate is a nonparametric estimator. It can be applied to the empirical frequency distribution of a data set. The advantage of a kernel density estimate is to avoid the subjective error of the histogram center choices. A kernel density smoothing is achieved by stacking kernel shapes. Given a value x , the probability of x is the summation of heights of all the kernel functions at this point. The PDF of kernel density is

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right), i = 1, 2, 3, \dots, n. \quad (3.8)$$

where n is the sample size and h is the smoothing parameter, expressed as

$$h = \frac{\min \{0.9s, \frac{2}{3}IQR\}}{n^{\frac{1}{5}}}. \quad (3.9)$$

where s is sample variance and IQR is the Interquartile Range, and it equals to the difference between the upper and lower quartiles. When h increase, the PDF becomes more smoothed. A Gaussian kernel function K is

$$K(t) = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{t^2}{2}\right], \quad -\infty < t < \infty \quad (3.10)$$

where, $t = \frac{x_0 - x_i}{h}$ when plug it into the kernel density PDF. x_i is the center of each kernel.

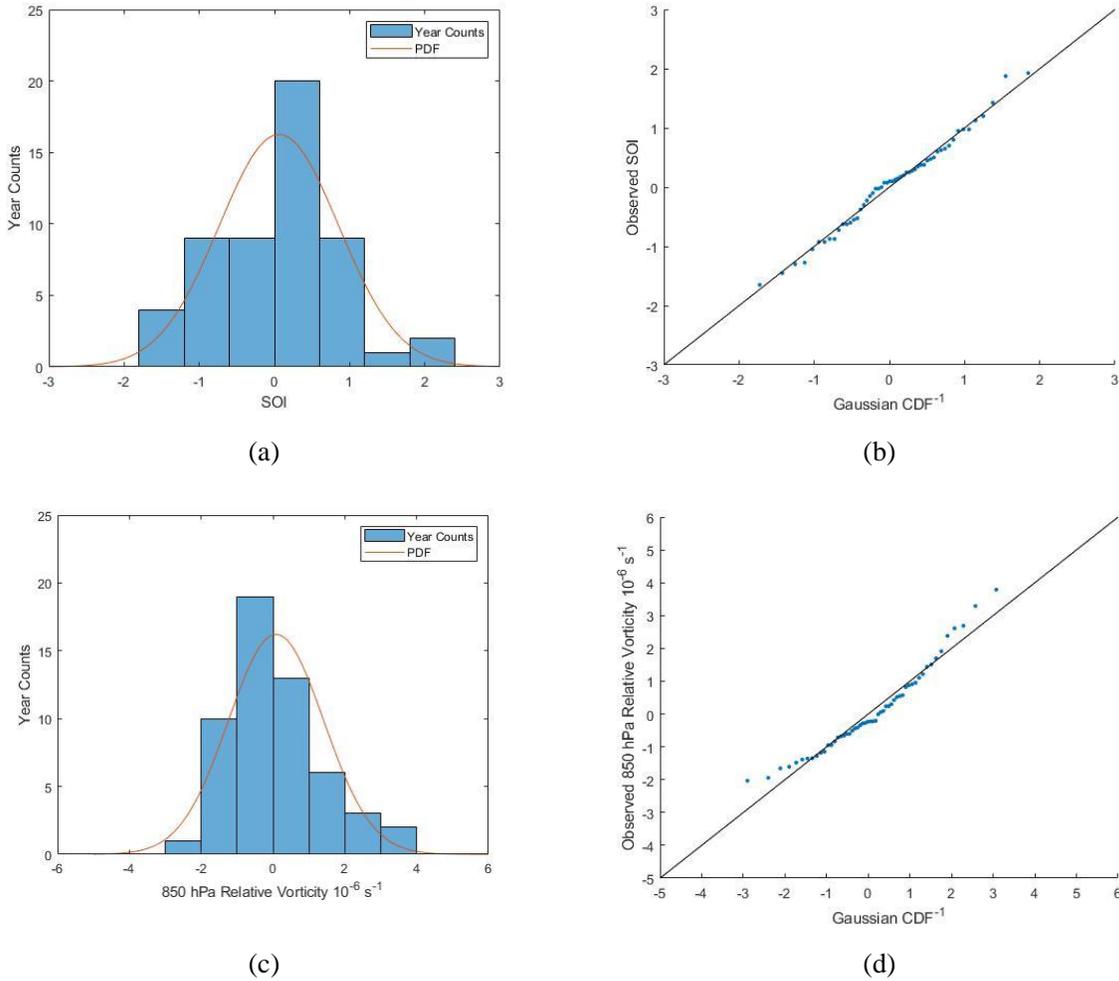


Fig. 12 The goodness of fit for Gaussian distribution. (a) and (c) The Probability density function (red line) plot with the observed data (blue rectangle) and (a) for SOI, (c) for low-level relative vorticity. (b) and (d) Quantile-Quantile (Q-Q) plot, x axis indicates the theoretical value and y axis represents the observed value. (b) is for SOI and (d) is for low-level vorticity.

A Poisson regression can model the number of TCs striking in a particular year conditional on climate factors. To compliment the Poisson regression model, a kernel density estimate is used to model these TCs striking dates in that year. The striking date is defined as the date of the TCs when the highest wind speed occurs in the defined circle in Fig. 1b. The PDF of kernel density estimate using TC data is

$$f(d) = \frac{1}{\frac{39}{h} \sum_{i=1}^{39} K\left(\frac{d - d_i}{h}\right)}, \quad 0 \leq d < 365, \quad (3.11)$$

where the smoothing parameter $h=13.72$. There are 39 TCs in the target region so $n=39$ (Table 1). The kernel density estimate of TC occurrences shows the peak value in around late July and early August, which agrees well with the observed TC peak season in July and August (Fig. 13).

3.5 Extreme value distributions

The next part of this study is to model the striking intensity of TCs. Here, the intensity refers to the maximum wind speed in the defined circle (Fig.1b). The extreme value distributions (EVDs) can be fitted for the TC intensities because the extreme wind intensities are positively skewed and nonnegative. There are two types of EVDs involved: the generalized Pareto distribution and the generalized extreme value distribution.

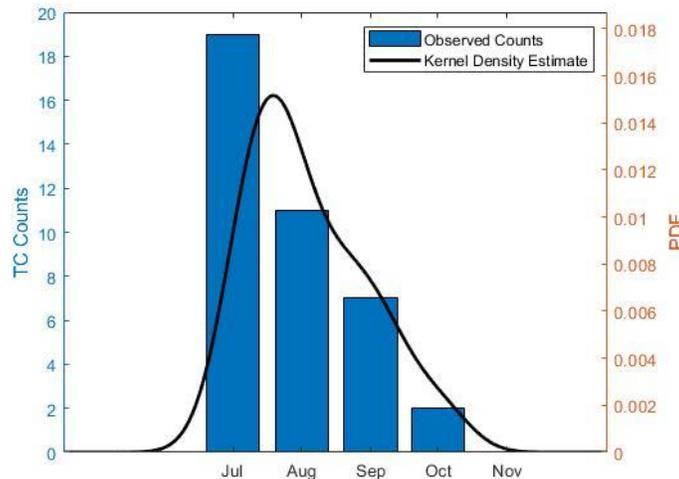


Fig. 13 The PDF of the kernel density estimate (black curve) plot with the observed TC numbers (blue rectangle).

3.5.1 Generalized Pareto distribution (GPD)

In the original model developed by Parisi and Lund (2008), the generalized Pareto distribution (GPD) is fitted to estimate the U.S. landfalling hurricane intensities. The GPD is an extreme value analysis using the peaks-over-threshold (POT) sampling. The PDF of the GPD is

$$f(x) = \frac{1}{\sigma^*} \left[1 + \frac{\kappa(x - \mu)}{\sigma^*} \right]^{-\frac{1}{\kappa} - 1}, \quad (3.12)$$

and its CDF is

$$F(x) = 1 - \left[1 + \frac{\kappa(x - \mu)}{\sigma^*} \right]^{-\frac{1}{\kappa}}. \quad (3.13)$$

Here μ is the threshold for the POT sampling, which should be a relatively high value. κ the shape parameter, and σ^* the scale parameter. These parameters could be estimated using the method of maximum likelihood. The threshold in this case is the minimum of the maximum TC wind speed in the defined circle which is 20 knots.

The fitted result is shown in Fig. 14. The estimated three parameters are $\mu = 20, k = -0.47,$ and $\sigma = 63.90$. Both PDF and Quantile-Quantile (Q-Q) plot suggest that the GPD did not fit well for Hawai'i's TC intensities. Only low intensity and super high intensity TCs could be estimated well, and the rest of TCs' intensities are overestimated. The GPD model is not a good extreme value distribution method for Hawai'i TCs.

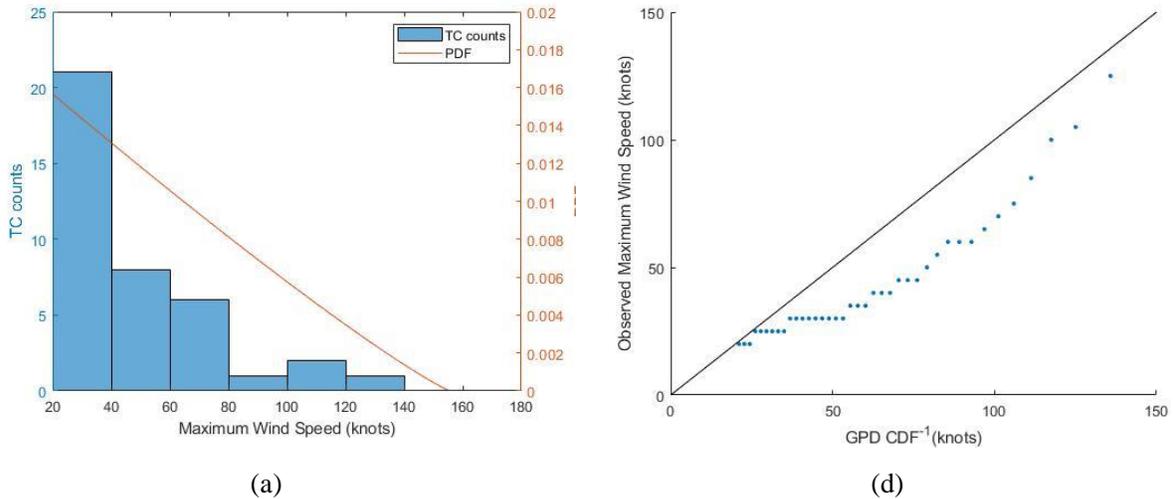


Fig. 14 The goodness of fit for the GPD. (a) The probability density function (red line) plot with the observed maximum wind speeds (blue rectangle). (b) The Quantile-Quantile (Q-Q) plot. The x axis indicates the theoretical value and y axis represents the corresponding observed value.

3.5.2 Generalized extreme value (GEV) distribution

Because the GPD is not fitted well for Hawai'i TC data, another extreme value distribution method should be explored. The generalized extreme value distribution (GEV) is applied to estimate the TC intensities. The GEV is the distribution with the characteristic of convergence at the largest of m values. If daily data in a year is used, then $m=365$ so that the highest value in a year is used. The PDF of the GEV is

$$f(x) = \frac{1}{\beta} \left[1 + \frac{\kappa(x - \mu)}{\sigma^*} \right]^{1 - \frac{1}{\kappa}} \exp \left\{ - \left[1 + \frac{\kappa(x - \mu)}{\sigma^*} \right]^{-\frac{1}{\kappa}} \right\}, \quad 1 + \frac{\kappa(x - \mu)}{\sigma^*} > 0. \quad (3.14)$$

The CDF of the GEV is

$$F(x) = \exp \left\{ - \left[1 + \frac{\kappa(x - \mu)}{\sigma^*} \right]^{-\frac{1}{\kappa}} \right\}, \quad (3.15)$$

Here μ is the location parameter, and the other two parameters have the same meaning as the GPD parameters.

The fitting result is shown in Fig. 15. The three parameters are $\mu = 31.23$, $\kappa = 0.47$, and $\sigma = 11.51$. Both PDF and Q-Q plot show that the GEV fitting is better than the GPD. Most estimate results are close to the observed data except the largest one. The GEV is taken to estimate the TC intensity in this study.

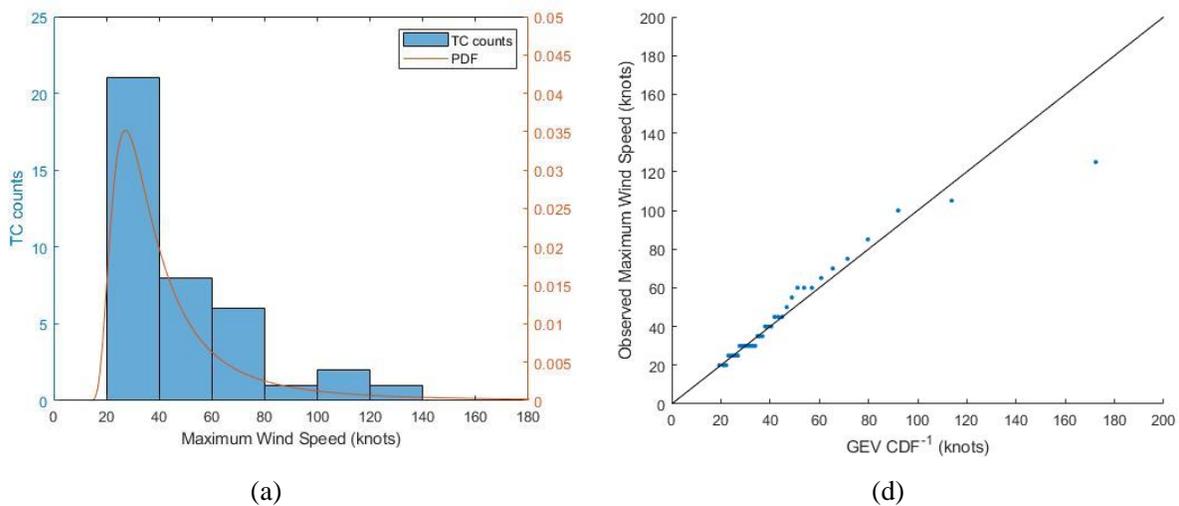


Fig. 15. Same as Fig. 14 but for the GEV.

3.6 A Statistical model

Utilizing the aforementioned methods, a statistical model is developed

. The return period of the TC with a striking wind speed of w knots in this model is the average waiting time from 1 January of a given year until a TC appears with a speed of w knots or above this magnitude. Each run of simulation generates a draw of a ‘level w ’ return period. By repeating a large number of simulations, the average value of these simulation results is calculated as the ‘level w ’ return period. The flow chart of this statistical model is shown in Fig. 16.

To run this model, we first set the number of year $t=1$ and generate the Poisson regression parameters SOI_t and ξ_t via function (3.7). Substitute these two parameters into (3.6) and simulate the number of TCs striking Hawai’i n for the given year t . The arrival date d_i ($i=1, 2, \dots, n$) of each TC in this year is simulated through (3.11) independently and the maximum wind speeds w_i ($i=1, 2, \dots, n$) of each TC in this year is modeled by (3.14) separately. If these wind speeds w_i do not reach the magnitude of ‘level w ’, go on to the next year ($t=t+1$) and continue the loop. Until the first TC maximum wind speeds reach the ‘level w ’ ($w_i \geq w$), the single simulated ‘level w ’ return period would be generated which is

$$R_w = (t - 1)years + (d_i)days. \quad (3.16)$$

Repeat this sequence over many independent simulations (10,000 times), the expected ‘level w ’ return period would be calculated through averaging these separate results.

CHAPTER 4. RESULT AND UNCERTAINTY ANALYSIS

4.1 Simulation results

The estimated return periods of various maximum wind speeds of TCs are listed in Table 4. Additionally, the previous estimated results by Chu and Wang in 1998 using data up to 1995 are also listed for comparison in the third column.

For small intensity storms, the new results are slightly different from the previous research. For example, the estimated return periods of ‘level 64 knots’ TCs is raised from 6.6 years to 9.9 years. This change results from the differences of models used and the data used. The GEV is used to generate the wind speeds in the current study while the lognormal and Gumbel distributions were applied in Chu and Wang (1998). Moreover, the data used in Chu and Wang’s study covered

the period from 1966 to 1995. Since 1995, many more TCs with stronger intensity were observed in the vicinity of Hawai'i. For example, Hurricane Lane in 2018 reached 105 knots. For storms over 100 knots (Table 4), the return periods are substantially shortened from Chu and Wang (1998). For example, the 'level 100 knots' return period is decreased from 42 years to 26.4 years and the 'level 125 knots' return period is reduced substantially from 202 years to 40.9 years.

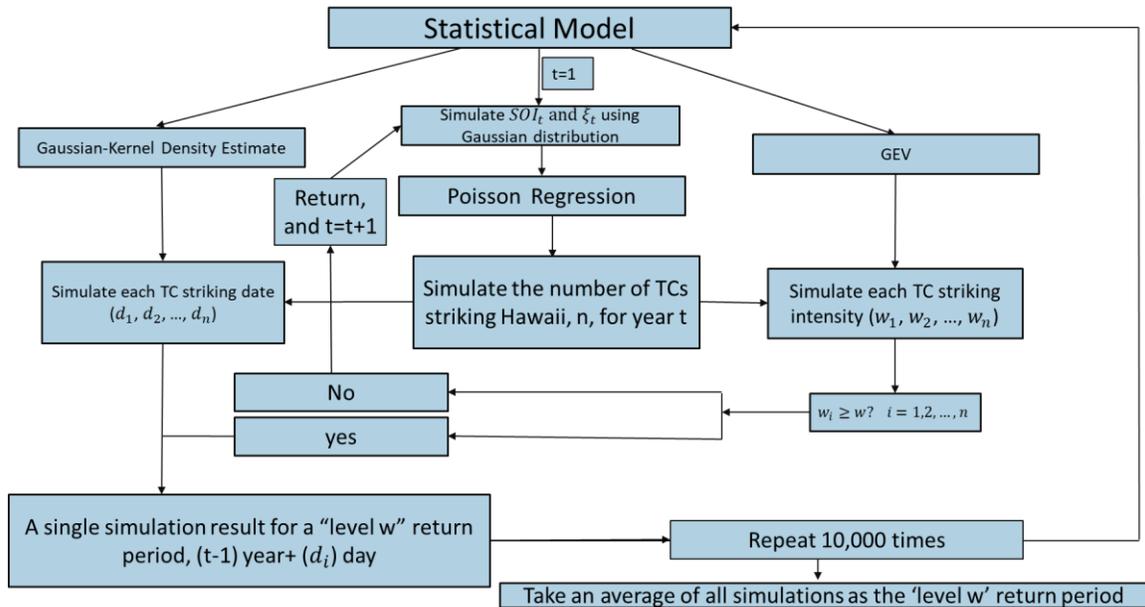


Fig. 16 The flow chart of the statistical model

Table. 4 Simulation result

Wind speed level (kt)	1966-2019 Return period (yr)	1970-95 (Chu and Wang, 1998)
34	3.0	3.2
50	6.0	4
64	9.9	6.6
80	16.2	13
100	26.4	42
110	31.7	81
125	40.9	202

4.2 Uncertainty analysis

The estimated results for each level TC in Table 4 are the exact value. This value may fluctuate from time to time. An uncertainty analysis is needed to calculate the confidence interval of the return periods. To establish a confidence interval of the deterministic value from our statistical model, a bootstrap method is advocated. The bootstrap is a resampling method by constructing the artificial data batches with replacements from the original data of size n with resampling data. By repeating a large number of S times, S bootstrap samples of size n can be constructed. For this research, the original data set is the observed maximum wind speeds of TCs with sample size ($n=39$). We construct the first sample that contain 39 draws from the original dataset and put them in the statistical model and run the flow explained in Chapter 3.6. The first resampling result is generated. Repeat the above steps for 10,000 times and we get ($S=10,000$) estimated results. The 95% confidence interval (Table 5) is the range of the central 9500 samples. For example, for the strongest intensity (125kt) TC, its return period is estimated as 40.9 years and the 95% confidence interval of that level lies between 30.1 and 55.6 years.

Table. 5 Confidence interval

Wind speed level (kt)	1966-2019 Return period (yr)	Confidence interval (yr)
34	3.0	1.9-3.6
50	6.0	3.4-8.0
64	9.9	6.5-14.1
80	16.2	13.2-22.8
100	26.4	20.4-35.6
110	31.7	28.4-40.7
125	40.9	30.1-55.6

CHAPTER 5. SUMMARY

A statistical model is developed to estimate the return period of Hawai'i TCs. It contains three parts, the Poisson regression, the kernel density estimate and the generalized extreme value

distribution. In Hawai'i's TC season, which is from July to October, two concurrent factors, the SOI and 850hPa relative vorticity, can approximately describe the annual TC frequency by using the Poisson regression (3.6). These two factors approximately follow the Gaussian distribution. Through the LOOCV, the correlation of the Poisson regression hindcasts and the observed TC numbers is 0.66 and the model catches most of peak values.

The kernel density estimate is introduced to get a higher resolution of TC arrival date on daily time scale. The GEV model is a reasonable fit except the most intense TC, while the GPD deviates from the data histograms. The GEV is a better model than the GPD in Hawai'i area to estimate the intensity of TCs.

The return period of high intensity TCs greater than 100 knots become much shorter than the previous estimate (Chu and Wang, 1998). The major hurricane with at least 125 knots wind speed in the past was considered as a 202-year event. Using more TC data, together with the modern statistical method, this event is shortened considerably to a return period of 40.9 years with the 95% confidence interval between 30.1 and 55.6 years. This result indicates that there is a higher risk for Hawai'i area to be struck by intense TCs. This is consistent with the available data, in which more TCs striking Hawai'i after 2013 (Fig. 1a).

One limitation of this study is that the model accuracy is highly dependent on the three distribution models. The correlation of Poisson regression and the historical data is 0.66 (not perfect) and the GEV slightly overestimates the most intense TCs. These issues could bias the model's accuracy. However, given the current scientific data and statistical methods, we believe that the statistical model developed in this study is perhaps the best tool one can come up for estimating return periods of hurricane wind intensities.

References

- Chu, P.-S. (2002). Large-Scale Circulation Features Associated with Decadal Variations of Tropical Cyclone Activity over the Central North Pacific. *Journal of climate*, 15(18), 2678–2689.
- Chu, P.-S. and Wang, J. (1997). Tropical Cyclone Occurrences in the Vicinity of Hawai'i: Are the Differences between El Niño and Non-El Niño Years Significant? *Journal of climate*, 10(10), 2683–2689.
- Chu, P.-S. and Wang, J. (1998). Modeling Return Periods of Tropical Cyclone Intensities in the Vicinity of Hawai'i. *Journal of Applied Meteorology*, 37(9), 951–960.
- Chu, P.-S. and Wu, P. (2008). Climatic Atlas of Tropical Cyclone Tracks over the Central North Pacific. University of Hawai'i, Department of Meteorology.
- Emanuel, K. and Jagger, T. (2010). On Estimating Hurricane Return Periods. *Journal of Applied Meteorology*, 49(5), 837–844.
- Jagger, T. and Elsner, J. B. (2006). Climatology Models for Extreme Hurricane Winds Near the United States. *Journal of climate*, 19(13), 3220–3236.
- Jagger, T., Elsner, J. B. and Niu, X. (2001). A dynamic probability model of hurricane winds in coastal counties of the United States. *Journal of Applied Meteorology*, 40, 853–863.
- Parisi, F. and Lund, R. (2008). Return Periods of Continental U.S. Hurricanes. *Journal of Climate*, 21(2), 403–410.
- Parisi F, and Lund R. (2000). Seasonality and Return Periods of Landfalling Atlantic Basin Hurricanes. *Australian & New Zealand Journal of Statistics*, 42(3), 271–282.

- Rupp, J. A. and Lander, M. A. (1996). A technique for estimating recurrence intervals of tropical cyclone-related high winds in the Tropics: Results for Guam. *Journal of Applied Meteorology*, 35, 627–637.
- Mayfield, M. and Rappaport, E.N. (1992). Eastern North Pacific hurricane season of 1991. *Monthly Weather Review*, 120, 2697-2708.
- Nugent, A.D., Longman R. J., Trauernicht, C., Lucas, M.P., Diaz, H.F., Giambelluca, T.W. (2020). Fire and Rain: The Legacy of Hurricane Lane in Hawai'i. *Bulletin of the American Meteorological Society*, 101(6), E954.
- United States. National Weather Service. (1993). Hurricane Iniki, September 6-13, 1992. Silver Spring, Md.: U.S. Dept. of Commerce, National Oceanic and Atmosphere Administration, National Weather Service.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press, 34-36, 157-166 pp.