

# Ciguatera poisoning trends in Florida using a predictive hybrid model

Daniel Breininger<sup>a</sup>, Christopher Ryzowicz<sup>b</sup>, Motti Goldberger<sup>c</sup>, Michael Splitt<sup>d</sup>,  
Robert van Woesik<sup>e</sup>, Nezamoddin N. Kachouie<sup>a,f,\*</sup>

<sup>a</sup> Department of Mathematics and Systems Engineering, Florida Institute of Technology, Melbourne, FL, USA

<sup>b</sup> Florida State University, Tallahassee, FL, USA

<sup>c</sup> Yale University, New Haven, CT, USA

<sup>d</sup> College of Aeronautics, Florida Institute of Technology, Melbourne, FL, USA

<sup>e</sup> Department of Ocean Engineering and Marine Sciences, Florida Institute of Technology, Melbourne, FL, USA

<sup>f</sup> Department of Electrical Engineering and Computer Science, Florida Institute of Technology, Melbourne, FL, USA

## ARTICLE INFO

### Keywords:

Ciguatera poisoning  
Negative binomial count model  
Wavelet cross coherence  
Climate change  
Zero-inflated count model

## ABSTRACT

The aim of this study is to identify an optimal predictive model for ciguatera poisoning and to determine which variables and time lags best explain the number of cases reported in Florida between 2006 and 2020. Ciguatera poisoning is a debilitating condition caused by consuming coral reef fish contaminated with ciguatoxins, which originate from toxin-producing dinoflagellates and biomagnify through the food chain. In severe cases, the illness can be fatal. Global climate change is expected to increase both the incidence of ciguatera poisoning and its geographic range, extending from tropical and subtropical reefs into temperate regions. This makes understanding and predicting its dynamics particularly urgent, as millions of people worldwide depend on reef fish as a dietary staple. To address this need, we developed an integrated approach combining wavelet cross-coherence analysis with a count modeling framework. Candidate predictors included cumulative monthly landings of Amberjack, Red Snapper, Red Grouper, and Scamp Grouper; the number of tropical cyclones; maximum ocean temperatures; precipitation; season; and Florida's human population. The optimal model identified was a Zero-Inflated Negative Binomial model. Results showed positive associations between ciguatera cases and (i) maximum ocean temperatures, (ii) storm frequency, (iii) fish landings, and (iv) human population size, alongside a negative relationship with precipitation. By establishing a robust predictive framework, this study advances understanding of the environmental and anthropogenic drivers of ciguatera poisoning. The findings provide a foundation for forecasting outbreaks and offer actionable insights to fisheries and public health agencies aiming to reduce risks for Florida residents and tourists.

## 1. Introduction

Ciguatera poisoning is a common fish-borne illness annually affecting 50,000–200,000 people worldwide (Friedman et al., 2008; Gingold et al., 2014). Ciguatera poisoning occurs when humans ingest fish that contain high concentrations of ciguatoxin that are produced by some species of dinoflagellate microalgae, such as *Gambierdiscus polysensilis* and *Gambierdiscus excrucians*, commonly found on coral reefs (Bagnis et al., 1980; Friedman et al., 2017; Ansdell, 2019). Herbivorous reef fish that consume these microalgae accumulate the associated ciguatoxin in their tissues, leading to biomagnification of the toxins through the food chain. Over 400 species of fish, including Grouper, Amberjack, Snapper, Surgefish, and Barracuda, have been previously

recorded with ciguatoxins in their tissue (Tester et al., 2010). The ingestion of carnivorous fish with high concentrations of ciguatoxins is dangerous to humans (Randall, 1958), causing symptoms ranging from gastrointestinal irritation, cardiovascular failure, paralysis, and, in severe cases, can lead to death (Bagnis et al., 1980; Lewis, 2006; Palafox and Buenconsejo-Lum, 2001).

The most vulnerable human populations to ciguatera poisoning rely on wild-caught fish found on tropical and subtropical coral reefs. Among the societal problems associated with ciguatera poisoning outbreaks are increased health costs, a reduced labor market, reduced food sources, loss in fish sales, and declining tourism (Lewis, 1992). Since there are no cheap or reliable tests for ciguatera poisoning, health facilities are left reporting symptoms, administering intravenous therapy, and

\* Corresponding author at: Dept. of Mathematics and Systems Engineering, Florida Institute of Technology, Melbourne, FL, USA.

E-mail address: [nezamoddin@fit.edu](mailto:nezamoddin@fit.edu) (N.N. Kachouie).

<https://doi.org/10.1016/j.rsma.2025.104514>

Received 11 March 2025; Received in revised form 19 September 2025; Accepted 22 September 2025

Available online 25 September 2025

2352-4855/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

prescribing dietary changes (Gingold et al., 2014). The lack of concrete test results in poor observational data collection on the number of cases, making this illness difficult to study.

Previous studies that sought to predict ciguatera poisoning have primarily been conducted in the Pacific Ocean. Laboratory experiments show that optimal dinoflagellate microalgal growth occurs at temperatures around 29 °C, partially explaining why ciguatera poisoning is most prevalent in tropical regions with high ocean temperatures (Tester et al., 2010). Elevated temperatures also lead to coral bleaching and can lead to coral mortality, resulting in open spaces for dinoflagellate expansion (Randall, 1958; Rongo and van Woesik, 2011). Similarly, cyclones and hurricanes have been linked to ciguatera poisoning, as these disturbances open up space for dinoflagellate growth and expansion on coral reefs (Rongo and van Woesik, 2011, 2013). Ciguatera poisoning is predicted to become more prevalent with rising ocean temperatures and intensifying storms, however, previous studies in southern Florida had varying results regarding the impact of storms on ciguatera poisoning cases (Gingold et al., 2014; Baker et al., 2008; Chan, 2016; Barrett, 2014). The varying results across studies may be in part related to geographical differences in fish species composition and differences in temporal lags associated with predictive variables. In this study, we developed and evaluated an optimal predictive model for ciguatera poisoning in Florida to identify the key environmental, ecological, and demographic variables, as well as their potential time-lagged effects that best explain the number of reported cases between 2006 and 2020.

## 2. Methods

This study followed a multi-step process to identify the most effective predictive model for ciguatera poisoning in Florida between 2006 and 2020. The workflow began with data collection from multiple publicly available sources, followed by preprocessing and transformation of variables to ensure comparability across time. Reported ciguatera cases

were then analyzed alongside candidate predictors, including fish landings, ocean temperature, precipitation, tropical cyclone frequency, seasonal indicators, and human population. Exploratory analyses were conducted to evaluate temporal patterns and potential lag effects, after which statistical modeling was performed using count-based approaches. Alternative models were compared to determine the best fit, with the Zero-Inflated Negative Binomial framework ultimately selected as the most appropriate for handling the over dispersed and zero-inflated nature of the case data. This process allowed for a systematic assessment of both environmental and demographic drivers of ciguatera incidence. The schematic of the proposed methods, from data collection and preprocessing to modeling ciguatera fish poisoning cases in Florida from 2006 to 2020, is depicted in Fig. 1. Each process shown in the flow diagram is discussed in detail in this section.

### 2.1. Study site description

This study was conducted across the entire state of Florida, with data collected statewide for all predictor variables. Reported cases of ciguatera poisoning were obtained from the publicly available

### 2.2. Reported Ciguatera cases

Diseases Frequency Report maintained by the Florida Department of Health (Florida Department of Health, 2021). All documented incidents between 1992 and 2020 were initially included in the analysis. Data was obtained for all 67 counties in the state of Florida. It should be noted that some counties did not report any ciguatera poisoning cases using the specified search filters and thus did not appear in the data. The location of exposure, type of contaminated fish, and other specifications about the illnesses were not described in the report. For example, a fish containing ciguatoxin could have been eaten in Broward County and then reported from a hospital in Brevard County, where the patient received

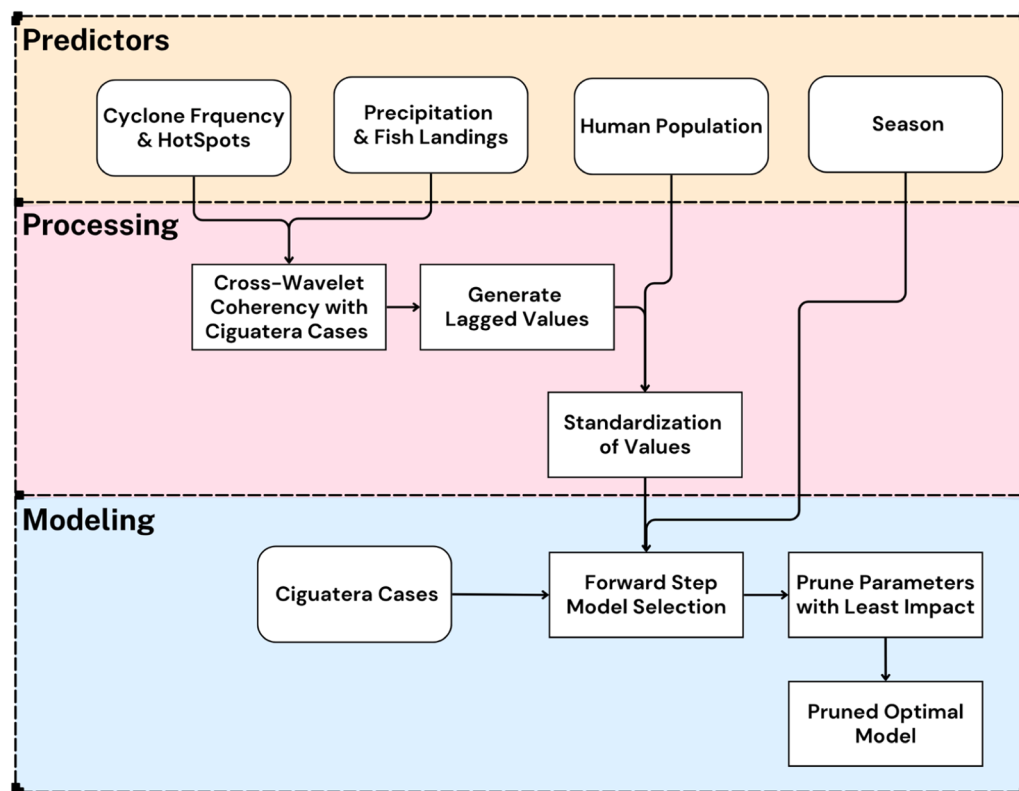


Fig. 1. Flow diagram of the methods to predict ciguatera poisoning cases in Florida from 2006 to 2020, starting with collecting predictors, processing the data, and modeling.

medical care. To account for the lack of spatial metadata, the total number of ciguatera poisoning cases in the entire state of Florida for each month was created by summing the counts of each county. This summation resulted in a data set comprising the total number of reported Florida ciguatera poisoning cases, sorted by month, beginning in 1992 and concluding in 2020.

Before 2006, it was typical for a month to have fewer than five cases. The longest duration, which each month had fewer than five cases, occurred between 2002 and 2006. After 2006, reported outbreaks increased, with a maximum of 20 cases monthly. Also, months with zero ciguatera poisoning cases became less frequent. With the significant variation in ciguatera poisoning over time, and with research suggesting inconsistencies and calling for a change in reporting (Begier, 2006), the most recent data from the years 2006–2020 were used in this study. The observed number of ciguatera poisoning cases from 2006 to 2020, are shown in Fig. 2a.

### 2.3. Temperature and precipitation data

Rising ocean temperatures have been previously associated with the expansion of ciguatera-containing dinoflagellates (Barrett, 2014). This expansion is associated with increasing growth rates and damage to the coral reef habitat (Steven R. Kibler, 2015). Coral mortality is caused by anomalously high ocean temperatures and the duration of those anomalous temperatures. Previously, scientists have used the Coral Bleaching HotSpot Product from NOAA as a measure of coral stress from temperature (NOAA Coral Reef Watch, 2020). This Hotspot metric collected is the difference in degrees Celsius between the highest

summertime mean temperature at a given site and the highest temperature occurring on a particular day at that site. In this study, the HotSpot values were gathered from two virtual stations in southern Florida and the Florida Keys. The process for matching the resolution of HotSpot with monthly ciguatera poisoning cases involved taking the maximum temperature value occurring in a month across either virtual site. NOAA similarly offers the Degree Heating Week (DHW) data product that considers lagged HotSpot values through time, although this was not used in the study because it has been specifically designed for corals, and the optimal lag for this study could differ.

Previous studies have hypothesized a positive association between ciguatera poisoning cases and precipitation. Increased precipitation results in runoff and thus an increase in nutrients or harmful substances that cause coral reef degradation, opening up new habitats for the proliferation of dinoflagellates with ciguatoxins (Ansdell, 2019). In this study, we used the monthly precipitation data of Miami-Dade County from the NOAA National Centers for Environmental Information (NOAA National Centers for Environmental Information, 2023).

### 2.4. Number of tropical cyclones

Previous studies in the Pacific Ocean have shown positive relationships between ciguatera poisoning and major tropical cyclones (Rongo and van Woesik, 2013). Tropical cyclones that may predict ciguatera poisoning cases were either tropical storms or hurricanes that crossed through southern Florida. Southern Florida was considered the relevant impact zone because of the presence of coral reefs, which the ciguatera-containing dinoflagellates inhabit. Tropical cyclone data were acquired from the National Hurricane Center's (National Hurricane Center, 2022) yearly reports, which contained the track of each major storm between 1992 and 2020. Previous studies have found that the storm's wind speed, the time of sustained winds, and the physical attributes of the coral reef itself determine how much destruction a storm causes to a reef (Matan Yuval, 2023; Nystrom, 2001). The data collected contained 6-hour snapshots of wind intensities, and thus, the wind speeds could not be accurately integrated through time or space. To minimize the possibility of false conclusions, the monthly frequency of storms was considered (Fig. 2-Top and -Middle).

### 2.5. Population

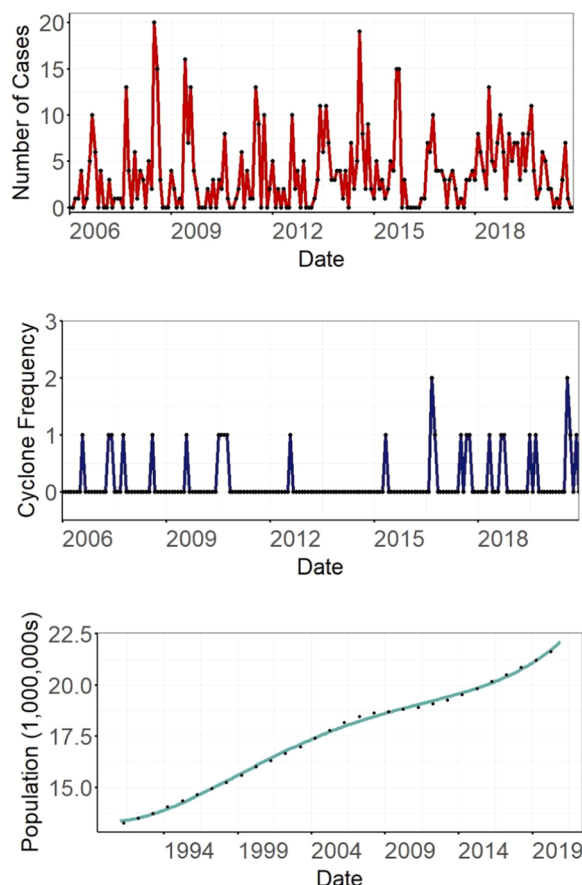
The US Census Bureau only provides Florida population data at 10-year intervals. To refine the data resolution required for the study, the yearly population estimates from 1992 to 2020 were obtained by the Bureau of Economic and Business Research at the University of Florida (University of Florida Bureau of Economic and Business Research, n.d.). To match the monthly ciguatera poisoning resolution with the population values, the data were interpolated using a fourth-degree polynomial fit to the data (Fig. 2-Bottom). The fit had a clear increasing trend, with the most rapid increase occurring in the most recent years (2020 and 2021).

### 2.6. Season

The season when the ciguatera poisoning case was reported was also used as a potential predictor of ciguatera poisoning cases. Winter was considered the months from December through February, spring from March through May, summer from June through August, and fall from September through November.

### 2.7. Reef fish landings

Commercial fish landings in Florida were collected from the Florida Fish and Wildlife Conservation Commission (FWC) (Florida Fish and Wildlife Conservation Commission, 2021). The online data tool provided access to data on dozens of fish species in Florida. The monthly



**Fig. 2.** Top) Monthly time series from 2006 to 2020 for the reported number of ciguatera fish poisoning cases in Florida; Middle) the number of tropical cyclones in southern Florida; Bottom) The estimated total Florida population using a fourth-degree polynomial on the yearly estimates.

landings, or total weight in pounds, of commercially caught Red Snapper (*Lutjanus campechanus*), Red Grouper (*Epinephelus morio*), Scamp Grouper (*Mycteroperca phenax*), and Greater Amberjack (*Seriola dumerili*) were collected as potential predictors of ciguatera poisoning. These fishes were selected because of their large landing sizes, their variability in behavior, and their tendency to contain ciguatoxins. The landings for each fish over time are provided in Fig. 3.

## 2.8. Wavelet coherence analysis

Excluding human population density and season, each predictor is expected to have an instantaneous impact as well as a lagged impact on the number of ciguatera poisoning cases. This is expected because the predictors could be related to multiple steps of the biomagnification process, and thus, each monthly predictor value could be related to ciguatera poisoning cases over multiple months. With the impossibility of collecting all necessary data to attempt the isolation of each step in the biomagnification process, an approximation of these relationships is necessary to create a relevant lag. Wavelet coherence analysis was performed using predictors and the reported number of ciguatera poisoning cases to find a generalization of the relations across multiple scales over time.

Wavelet transform is a localized time-frequency (or space-frequency) analysis to decompose a time series and determine the frequency content of narrower local time intervals (Torrence and Compo, 1998). Wavelet analysis is performed by taking a small wave called a wavelet (such as the Morlet wavelet) and comparing it with short pieces of the original time series and assigning a similarity coefficient. This process was repeated at different scales by stretching/compressing and shifting the time series.

To examine the relationships between ciguatera poisoning cases and environmental predictors, cross-wavelet analysis was applied to identify the intervals of significance between the lagged predictors and the number of reported ciguatera poisoning cases in the time-frequency domain.

The wavelet-coherence analysis was then performed to identify the coherence of the cross-wavelet transform in the time-frequency domain to obtain localized correlation coefficients between the two-time series (Grinsted et al., 2004). In this study, wavelet analysis was performed using the “wtc” function from the “biwavelet” package in R (Gouhier

et al., 2021). The heat map of the coefficient of determination ( $R^2$ ) is shown in Fig. 4, where regions of significance between the predictors and the number of ciguatera poisoning cases were outlined.

The optimal lag for each predictor is determined using the wavelet coherence plot by identifying the region in the frequency domain that consistently provides high  $R^2$  values throughout the study from 2006 to 2020. The selected temporal region for each predictor represents the consistently high correlation between ciguatera poisoning and the predictor's frequency. For example, in Fig. 4a, there was a clear outline shown by pixels marked in red, for the most frequent lag of [0–3 months] over the study period (2006–2020). It suggests a likely association between red snapper landings and reported ciguatera poisoning cases within three months of the fish landing. Notice that a climate predictor could impact the observed ciguatera poisoning cases at different scales (different time lags) of a predictor. Hence, whenever it was suggested by the coherence plot, the predictor was represented using two separate lagged parameters in the model.

## 2.9. Count models

A Generalized Linear Model (GLM) is defined by:

$$\eta(E[Y|X]) = X\beta \quad (1)$$

where  $\eta$  is the link function,  $E[Y|X]$  is an  $n \times 1$  vector of the expected response,  $X$  is an  $n \times (K+1)$  matrix of predictors,  $\beta$  is  $(K+1) \times 1$  vector of coefficients,  $n$  is the number of observations, and  $K$  is the number of predictors in the model. If  $\eta$  is an identity link function, GLM will be simplified to a simple/multiple linear regression model. The response in the multiple linear regression model can assume continuous values. Because the ciguatera poisoning case only assumes non-negative integers, a multiple regression model is not relevant for this data. Each observation regarding ciguatera poisoning cases is a non-negative integer with a Bernoulli distribution representing reported illness or no illness. Hence, a count model is considered for predicting the expected number of ciguatera poisoning cases. A count model assumes that the observed response has a discrete distribution, such as a Poisson distribution.

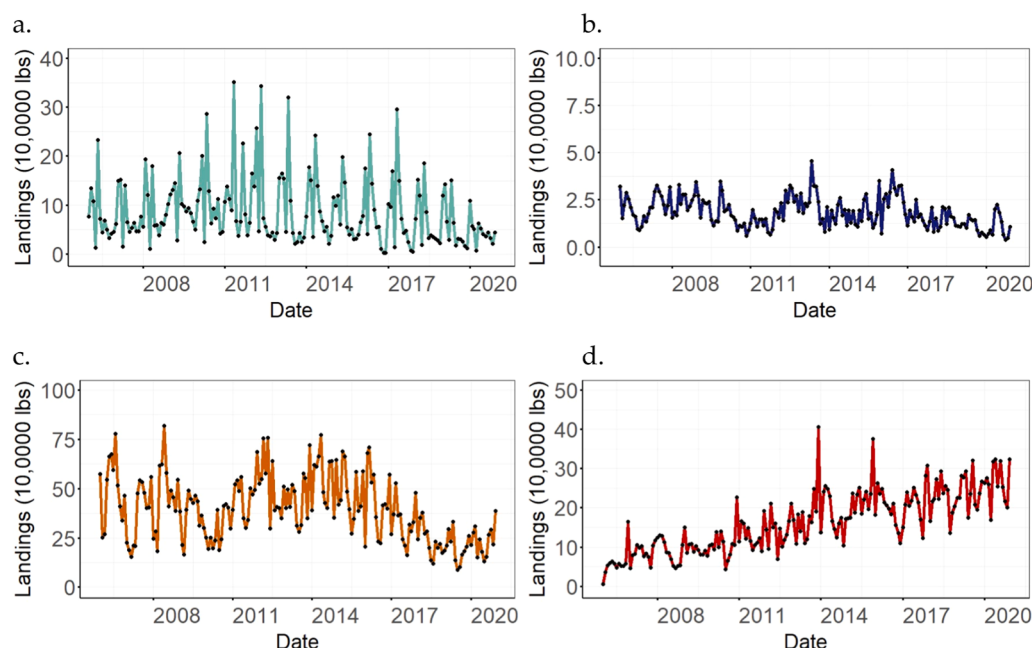


Fig. 3. Time series of monthly commercial fish landings in Florida from 2006 to 2020. a) Amberjack; b) Scamp Grouper; c) Red Grouper; d) Red Snapper.

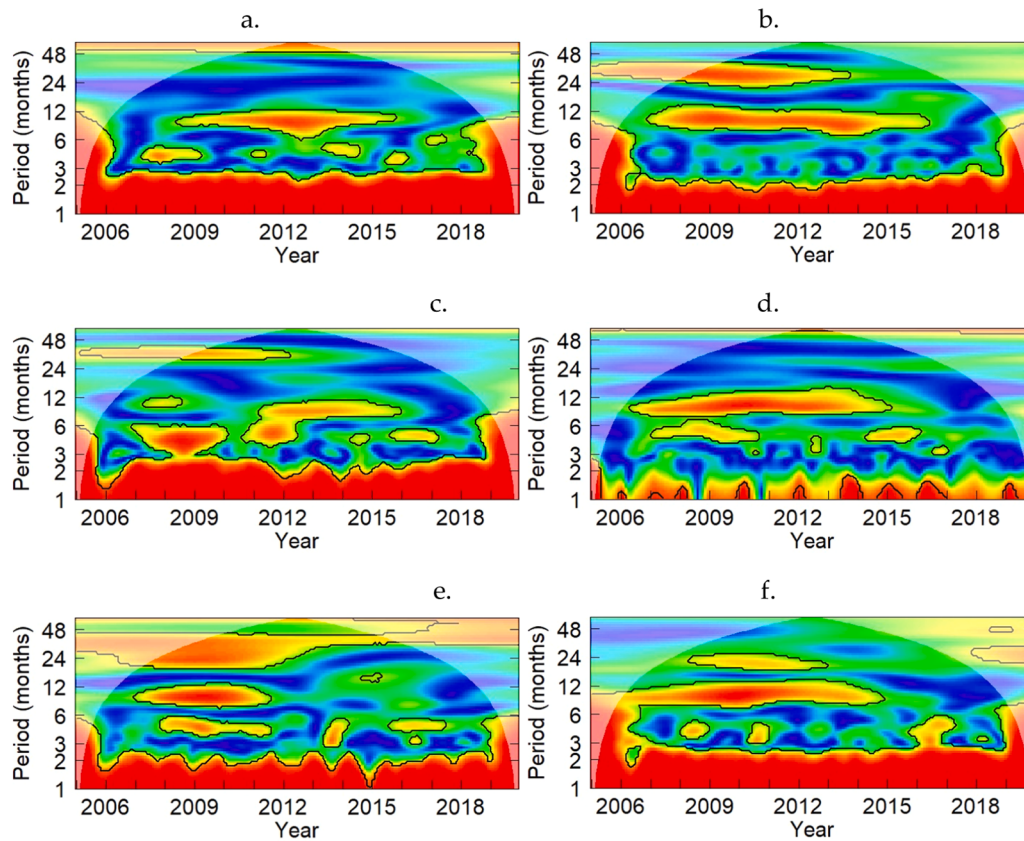


Fig. 4. Wavelet cross-coherence plots for the number of reported ciguatera poisoning cases in Florida (from 2006 to 2020) with each predictor: a. Red Snapper landings; b. Amberjack landings; c. Red Grouper landings; d. HotSpot; e. Frequency of tropical cyclones; f. Monthly precipitation in southern Florida.

### 2.9.1. Poisson model

The probability mass function (PMF) for a Poisson random variable is:

$$P(Y = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (2)$$

with

$$E(Y) = \text{Var}(Y) = \lambda \quad (3)$$

where  $\lambda$  and  $k$  are the average and the possible number of ciguatera poisoning cases in a given time period. Poisson regression is a GLM with a log-link function and with a response variable that follows a Poisson distribution. Thus, the expected value of the response is continuous and in the space of all real numbers. Modeling count data using Poisson regression has the following form:

$$Y_i \sim \text{Pois}(\lambda_i), \quad \lambda_i = \exp(X_i^T \beta) \quad (4)$$

where  $X_i$  is the  $i^{\text{th}}$  predictor vector associated with response  $Y_i$ . The expected value and variance are as follows:

$$E[Y_i | X_i] = \text{Var}[Y_i | X_i] = \exp(X_i^T \beta) \quad (5)$$

This model was not relevant because the equal dispersion assumption was not being met, given that ciguatera poisoning was over-dispersed, with a mean of 2.272 and a variance of 11.935.

### 2.9.2. Negative binomial model

To represent over-dispersed count data, a negative binomial (NB) distribution can be considered (Number Crunching Statistical Systems Software, 2021a):

$$\varepsilon_i \sim \text{Gamma}(r, 1/r), \quad f(\varepsilon_i) = \frac{r^r}{\Gamma(r)} \varepsilon_i^{r-1} e^{-r\varepsilon_i} \quad (6)$$

with

$$E[\varepsilon_i] = 1, \text{Var}[\varepsilon_i] = \frac{1}{r} \quad (7)$$

A NB probability density function (Number Crunching Statistical Systems Software, 2021b) is:

$$f(Y = y_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1}) y_i!} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_i} \right)^{\alpha^{-1}} \left( \frac{\mu_i}{\alpha^{-1} + \mu_i} \right)^{y_i} \quad (8)$$

where  $\alpha = \frac{1}{r}$  is the dispersion parameter, and mean and variance are:

$$\begin{aligned} E[Y_i | X_i] &= \exp(X_i^T \beta), \text{Var}[Y_i | X_i] \\ &= E[Y_i | X_i] + \alpha(E[Y_i | X_i])^2 \end{aligned} \quad (9)$$

In this way, the NB distribution is modeled using a Gamma distribution as a mixture of Poisson random variables, where  $\mu$  is the average number of events that occur during a specific time period, depending on  $i$ .

### 2.9.3. Zero-inflation model

A zero-inflated negative binomial model (ZINB) is a negative binomial model that can account for excess zeros. Excess zeros are data points that are believed to be zero by a process independent of the count process (UCLA Statistical Methods and Data Analytics, n.d.). The zero-inflated negative binomial model tests two sets of hypotheses:

1. An unobserved ciguatera poisoning case is an excess or a true zero.
2. The observed ciguatera poisoning case is from a negative binomial distribution.

Because many ciguatera poisoning cases are not reported or misdiagnosed as food poisoning, excess zeros will be considered for modeling ciguatera poisoning cases using the aforementioned predictors by a ZINB probability distribution (Number Crunching Statistical Systems Software, 2021b):

$$P(Y_i = j) = \begin{cases} \pi_i + (1 - \pi_i) f(Y_i = 0_i) & j = 0 \\ (1 - \pi_i) f(Y_i) & j > 0 \end{cases} \quad (10)$$

with  $\pi$  being the probability that a zero is an excess zero and  $f(Y_i)$  being a NB distribution given by Eq. (8). Thus, ZINB regression combines two models (Eq. 10) to yield a single probability of a given number of successes for each data point. Given that previous literature has estimated that a significant number of cases in Florida have been unreported (Friedman et al., 2008) and that the ciguatera poisoning cases were over dispersed a ZINB model was determined to be most relevant.

### 2.10. Model selection

After an admissible count model was determined, the predictors that can optimally predict ciguatera poisoning cases must be identified. Moreover, interactions with biological relevance between predictors will be recognized. Therefore, the interactions between each predictor with population, season, HotSpot, number of storms, and precipitation were considered. These interaction terms were included in the model to separate a predictor's trend that resulted from a relationship with another predictor. For example, including the interaction between season and red snapper landings into the model would be expected to differentiate the impact of a change in red snapper landings on the number of ciguatera poisoning cases depending on the season. The interaction could also remove noise from the red snapper landings coefficient that resulted from consistent seasonal changes. Possible seasonal variation is likely from fishing restrictions and randomness in the fish population. Thus, each predictor and the interactions mentioned above generate the parameters of the full model. Before the optimal grouping of parameters was found, each parameter was standardized for an even comparison of coefficients. Then, a goodness of fit for the full model was tested to initiate a reference point in the model selection process.

The predictors of the model were first chosen through a forward selection method by optimizing a goodness of fit metric using the Akaike Information Criterion (AIC). In turn, predictors are added one by one to the model to find the optimal AIC. The forward selection process was repeated until there were no parameters left to add to improve the model. Then, through a backward selection method, the model was pruned by removing the parameters, one parameter at a time, that resulted in the least change in AIC. The final model was then validated using varying time series splits with the data prior to the split for training and the remaining for testing.

## 3. Results

The optimal lag for each predictor, excluding scamp grouper landings, was found using the wavelet cross-coherence plots (Fig. 4). Table 1 describes the relationships, from the cross-coherence plots, between the reported number of ciguatera poisoning cases and each parameter. Fig. 4a to c show the wavelet cross-coherence plot for the reported ciguatera poisoning cases with the amount of Red Snapper, Amberjack, and Red Grouper landings, respectively. These plots suggested that a lag of 3 months was appropriate as the optimal lag for Red Snapper and Red Grouper, whereas there were high R-squared values at 0–2 months across all years of the study, suggesting a 2-month lag for Amberjack landings.

The wavelet cross-coherence plot between the monthly maximum HotSpot value and the reported ciguatera poisoning cases is provided in Fig. 4d. There were high R-squared values between 0 and 2 months

**Table 1**

Identified optimal lag for each predictor in the model. The lags were determined from the wavelet coherence plots between the reported number of ciguatera poisoning cases in Florida and the predictors.

| Predictor                                             | Lag                              |
|-------------------------------------------------------|----------------------------------|
| Season                                                | Current month                    |
| Population                                            | Current month                    |
| Amberjack ( <i>Seriola dumerili</i> ) Landings        | Current to the Previous 2 Months |
| Red Snapper ( <i>Lutjanus campechanus</i> ) Landings  | Current to the Previous 3 Months |
| Red Grouper ( <i>Epinephelus morio</i> ) Landings     | Current to the Previous 3 Months |
| Scamp Grouper ( <i>Mycteroperca phenax</i> ) Landings | Current to the Previous 3 Months |
| Short-term Tropical Cyclone Frequency                 | Current to the Previous 3 Months |
| Long-term Tropical Cyclone Frequency                  | Previous 32–46 Months            |
| Short-Term HotSpot                                    | Current and Previous Month       |
| Long-Term HotSpot                                     | Previous 8–15 Months             |
| Short-Term Precipitation                              | Current to the Previous 3 Months |

across approximately half of the time period in this study, denoting an optimal short-term lag of the current plus previous month for the HotSpot parameter. A long-term lag for HotSpot was also selected using the previous 8–15 months for the HotSpot parameter and had a correlation of 0.522 with the short-term parameter. The wavelet cross-coherence plot between the number of tropical cyclones and the reported ciguatera poisoning cases is provided in Fig. 4e. An optimal short-term lag of 0–3 months was selected for the number of tropical cyclones because that lag was significant across all years. Similarly, the long-term lag of the previous 32–46 months was considered for tropical cyclones. The correlation between the short-term and long-term lagged tropical cyclone parameters was 0.117. The wavelet cross-coherence plot between monthly precipitation in southern Florida and the reported ciguatera poisoning cases is provided in Fig. 4f, and a short-term lag of the current to the previous 3 months was considered.

The selected parameters for each optimal model are provided in Table 2. The first model was the full negative binomial (NB) model, with 47 parameters and performing stepwise selection. The optimal negative binomial model had an AIC of 802 (Table 2). The second model was a zero-inflated negative binomial model (ZINB), implemented to consider the potential excessive zeros caused by unreported ciguatera poisoning cases. The stepwise selection method yielded an optimal model with the lowest AIC (than the NB model) of 784.099, with six parameters in the count model and eight parameters in the zero model. By pruning the insignificant parameters from the optimal ZINB model (that was identified through the forward selection method), the third optimal (pruned) model was obtained, with an AIC of 788.625.

The model summary for the pruned optimal ZINB model is given in Table 3. The model summary provides the coefficients and significance values for each predictor in the count and zero models. The stepwise and pruned model had all the count coefficients being significant. In the count portions of the models, each model had the same parameters with the same signs of the coefficients, except the pruned model had an insignificant interaction between the Florida population and the short-term HotSpot value removed. The zero portions of the optimal ZINB models did not differ in parameters or the signs of their coefficients. In the pruned model, a one standardized unit increase would result in a multiplicative change in the expected number of ciguatera poisoning cases by the exponent of the coefficient. For the zero portion of the model, a one standardized unit increase would result in a multiplicative

**Table 2**

The goodness of fit and the number of model parameters for each optimal model to predict the number of ciguatera poisoning cases in Florida.

| Model | Name | Zero Model (# Param.) | Count Model (# |
|-------|------|-----------------------|----------------|
|-------|------|-----------------------|----------------|

**Table 3**

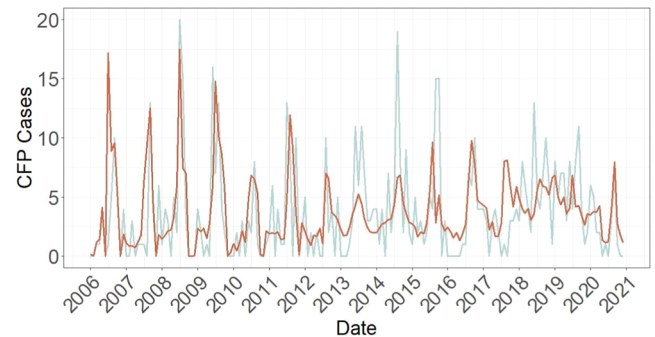
Model summary of the pruned optimal zero-inflated negative binomial model to predict the number of ciguatera poisoning cases in Florida.

| COUNT MODEL COEFFICIENTS         |             |            |         | ZERO-INFLATION MODEL COEFFICIENTS         |             |            |         |
|----------------------------------|-------------|------------|---------|-------------------------------------------|-------------|------------|---------|
| NEGBIN WITH LOG LINK             |             |            |         | BINOMIAL WITH LOGIT LINK                  |             |            |         |
| Parameter                        | Coefficient | Std. Error | p-Value | Parameter                                 | Coefficient | Std. Error | p-Value |
| Intercept                        | 1.218       | 0.071      | < 0.001 | Intercept                                 | -22.362     | 11.832     | 0.059   |
| HotSpot <sub>s</sub>             | 0.432       | 0.068      | < 0.001 | Red Snapper                               | -21.497     | 11.222     | 0.055   |
| Storm <sub>L</sub> : Red Grouper | 0.192       | 0.070      | 0.006   | Amberjack                                 | 4.293       | 2.705      | 0.113   |
| Population: Precipitation        | -0.388      | 0.098      | < 0.001 | Red Snapper: HotSpot <sub>s</sub>         | 15.412      | 7.898      | 0.051   |
| Storm <sub>L</sub> : Amberjack   | 0.308       | 0.093      | 0.001   | Population: Storm <sub>L</sub>            | 10.862      | 5.858      | 0.064   |
| Red Snapper: Storm <sub>s</sub>  | 0.196       | 0.104      | 0.060   | Red Snapper: Spring                       | 10.173      | 6.551      | 0.120   |
| log( $\theta$ )                  | 0.868       | 0.202      | < 0.001 | Red Snapper: Summer                       | -8.171      | 4.616      | 0.077   |
|                                  |             |            |         | Red Snapper: Winter                       | 8.766       | 5.972      | 0.142   |
|                                  |             |            |         | HotSpot <sub>s</sub> : Storm <sub>L</sub> | -8.887      | 5.747      | 0.122   |

change in the odds of a zero being an excess zero by the exponent of the coefficient. Thus, an increase in a positive coefficient relates to an increase in the odds that an observed zero is an excess zero, and the opposite for a negative coefficient predictor. Fig. 5 provides a graphical representation of the pruned model and is shown in Table 3.

It should be noted that the model predicts the average number of ciguatera poisoning cases in comparison with the observed ciguatera poisoning cases. As such, the model demonstrated a fairly accurate fit with the number of reported ciguatera poisoning cases (Fig. 6). While some predicted zeros did match the reported zeros (i.e., true zeros), some predicted non-zero ciguatera poisoning cases did not match the reported zeros (i.e., excess zeros). The model under-predicted the peaks, particularly after 2011, perhaps due to predicting average ciguatera poisoning. The model had a slight increase in the predicted number of cases over time, but with a decline starting in 2020. On average, both the predicted and the observed ciguatera poisoning cases peaked in the summer months. We also note that the model was poor at predicting cases of ciguatera poisoning between 2013 and 2016, during a period of few hurricanes.

To have a closer look and visualize the goodness of the model, the rootogram of predicted and true counts are shown in Fig. 7 for different frequencies of ciguatera poisoning cases. The figure displays the number of ciguatera poisoning cases on the x-axis vs. the frequency (number of months) of the number of ciguatera poisoning cases. Predicted (expected) ciguatera poisoning counts by the model are shown by the thick red line; observed ciguatera poisoning counts are shown as gray bars, which are hanging from the red line of expected counts on the x-axis, such as the count bin, 0 ciguatera poisoning count, 1 ciguatera poisoning count, 2 ciguatera poisoning count, and so on. The Y-axis is the square root of the observed and expected ciguatera poisoning counts, where the square root transformation allows for departures from expectations to be seen even at small frequencies. Overall, predicted expected ciguatera poisoning counts follow the same trends as observed ciguatera poisoning counts. The model under-predicted the frequency of 4 ciguatera

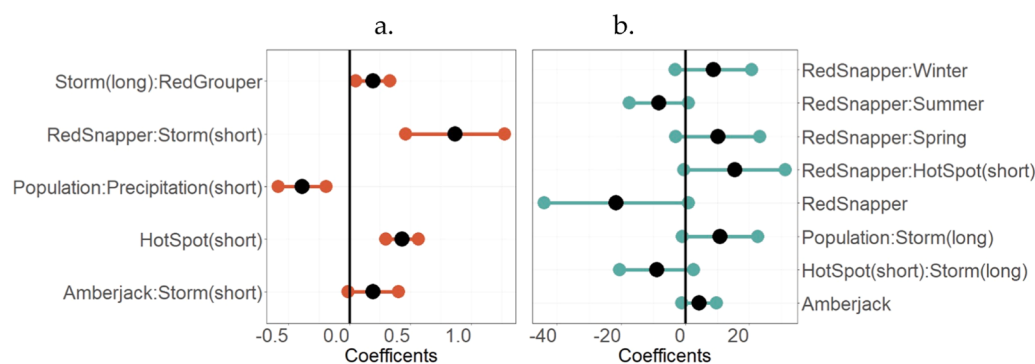


**Fig. 6.** Predicted number of ciguatera poisoning cases (red) using the pruned zero-inflated negative binomial model superimposed on the observed number of ciguatera poisoning cases (green) in Florida from 2006 to 2020.

poisoning cases, slightly underpredicted the frequency of zero, 7, 10, 12, and 15 ciguatera poisoning cases, and slightly overpredicted the frequency of 1, 2, and 3 ciguatera poisoning cases. However, we should point out that the predicted values by the model are the expected or average number of ciguatera poisoning cases, while the observed values are the true number of ciguatera poisoning cases that have been reported. Therefore, the predicted number of ciguatera poisoning cases by the model is relatively accurate.

To evaluate the predictive capabilities of the pruned optimal model, classic cross-validation techniques were not relevant and were not performed to avoid data bleeding of the lagged values, which could bias the predictions.

Rather, a time-progressive time series split validation was implemented. The validation was performed using three time series with train-test splits with the first 65, 60, and 55 % of the data for training, and the remaining portion, i.e., 35, 40, and 45 %, used for testing the models. These three time series splits were selected to have enough data



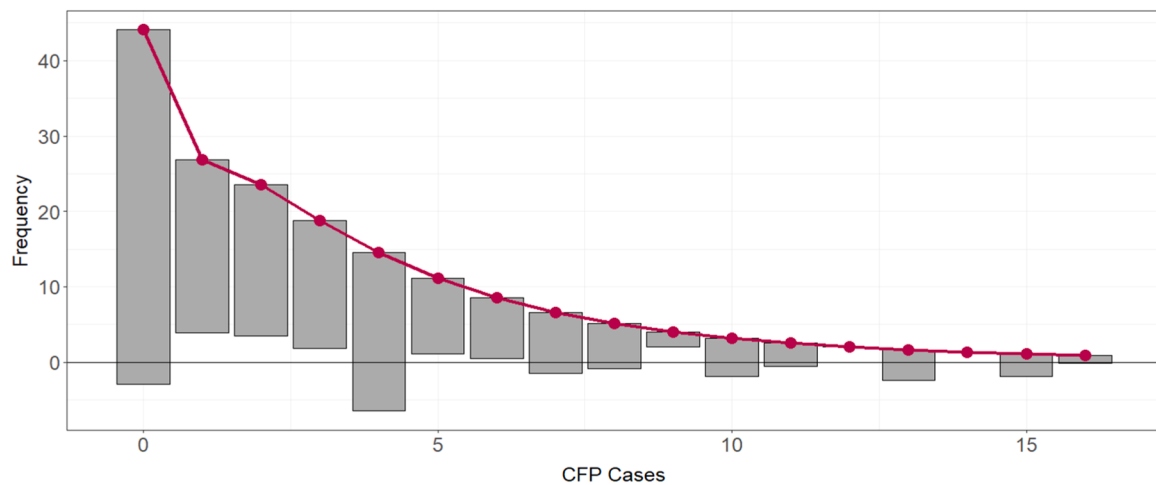


Fig. 7. Rootogram of the predicted (expected) ciguatera poisoning cases (solid red line) vs. observed ciguatera poisoning cases (gray bars).

for both training and testing. The testing model's raw predictions and the number of cases over time were plotted in Fig. 8. This process was repeated to observe general trends using loess smoothing at a span of 0.25 on the predictions and cases. The mean squared error (MSE) of the raw model predictions with the true number of ciguatera poisoning cases was calculated and superimposed onto each plot. The two predictions using 65 and 60 % of the data for training had a smaller MSE than the predictions of the model using fewer data for training. While the predictions using the models with 60 and 65 % training only differ by about 1.18 MSE, the 60 % trained model predictions had a better fit, following the dominant trends of the number of ciguatera poisoning cases in Florida, for its testing time range.

#### 4. Discussion

The awareness of Ciguatera outbreaks in Florida has increased over the past decades, but inaccurate reporting has limited our understanding of the dynamics and severity of these outbreaks. There have been few studies on the number of cases of Ciguatera in Florida and the potential environmental influences that can cause outbreaks. In this study, the number of reported Ciguatera cases in Florida was modeled as a response to several predictors, including the HotSpot index, precipitation, number of tropical storms, the human population density of Florida, season, and the commercial landings of four common fish species.

In the stepwise and the optimal pruned model, there was a strong positive relationship between ciguatera poisoning cases and the short-term HotSpot values. The interaction between the short-term HotSpot

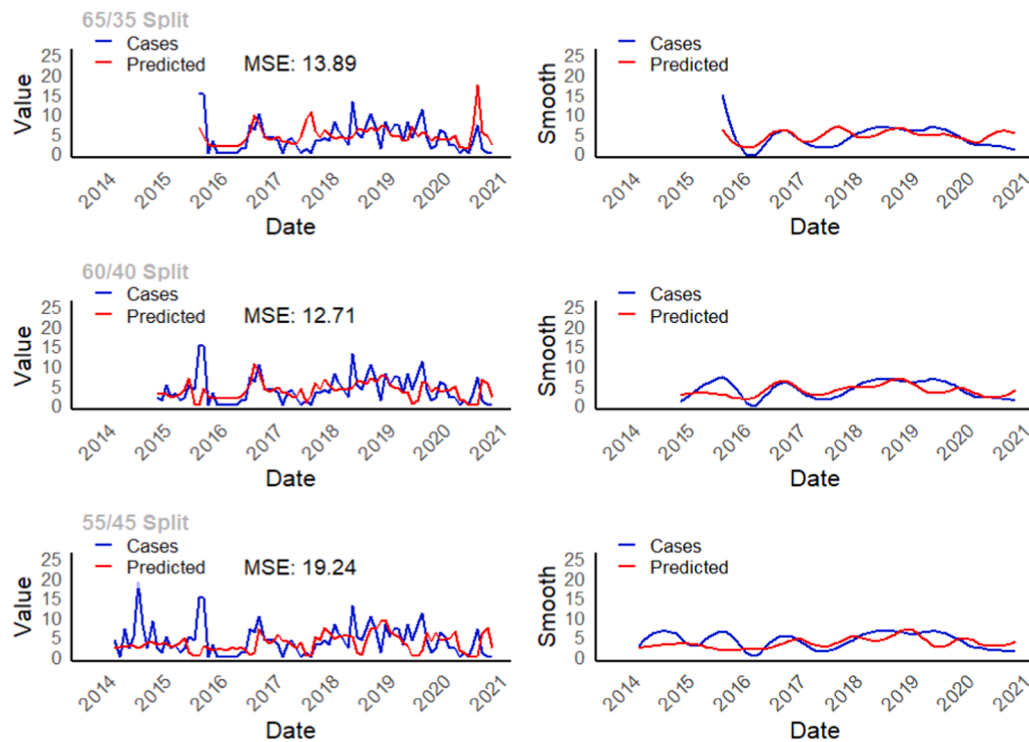


Fig. 8. Predicted number of ciguatera poisoning cases (red) from the testing sets using the last 35 %, 40 %, and 45 % of the data from top to bottom, respectively. Each iteration used the pruned zero-inflated negative binomial model features trained on the remaining portions of the split time series. The predictions are superimposed on the observed number of ciguatera poisoning cases (blue) in Florida from 2006 to 2020.

values with the human population density was present in the stepwise model, although it was insignificant and was removed to create the optimal pruned model. These results suggest that, given all other parameters remain constant, high short-term HotSpot values are associated with an increase in the expected number of ciguatera poisoning cases. The long-term effect of HotSpot values, from 8 to 15 months, did not play a major role in the models.

The results of studies in the Pacific Ocean, accompanying some studies in Florida, have previously shown a strong relationship between large storms and ciguatera poisoning, with large storms clearing reefs and thereby creating habitat for ciguatoxin-containing dinoflagellates (Gingold et al., 2014; Baker et al., 2008; Rongo and van Woesik, 2013; Chan, 2016; Barrett, 2014). For the optimal model of the study, the interaction between the long-term impact from the number of tropical cyclones and the landings of two fish species, Red Grouper and Amberjack, had a strong positive relationship with the expected number of ciguatera poisoning cases. Also, the interaction between the short-term impact of the number of tropical cyclones with Red Snapper was present in the count model, with a positive relationship. Therefore, these fish can be used as a proxy for the number of ciguatoxin-infected fish that are consumed. The interaction between the long-term impact of the number of tropical storms and the human population of Florida showed a positive coefficient and a p-value of 0.064 in the zero-inflated model. This interaction suggests that an increase in storm frequency will likely reduce the chance of having zero ciguatera poisoning cases. Unlike the number of tropical storms, the interaction between the population of Florida and precipitation had the only negative and significant coefficient in predicting the number of ciguatera poisoning cases. Thus, when the number of people in Florida remained constant, an increase in precipitation in Southern Florida correlates to a decrease in the number of ciguatera poisoning cases.

The landings of three of the fish species, Red Snapper, Red Grouper, and Amberjack, had significant coefficients in the optimal models. It is important to note that Scamp Grouper was not a significant parameter in either model, likely due to its consistently low landings, and therefore, few infected fish were consumed. Even though Amberjack had a slightly decreasing trend over time, and can be partially pelagic, it still had a larger positive coefficient in the count model than the other fish species. This effect could be because Amberjack had a shorter lag, or that Amberjack can reach larger sizes than Red Grouper or Red Snapper and can therefore accumulate higher levels of toxins. Interestingly, Red Snapper appeared in most of the parameters for the zero-inflated model and also had the largest regions of correlation to ciguatera poisoning cases in the cross-coherence plot, possibly because the trend for the landings of Red Snapper had the least variability in comparison to Amberjack and Red Grouper. Individually, Red Snapper had the most extreme coefficient in the zero-inflated model. The result suggests that if a month has no reported cases, when the Red Snapper landings are high, the zero is more likely to be a true zero than an excess zero. The interaction between Red Snapper and short-term HotSpot had the largest positive coefficient, with a p-value of 0.051, suggesting that having zero ciguatera poisoning cases is unlikely with high anomalous ocean temperatures.

The optimal pruned model was validated using three time series splits using the first 65, 60, and 55 percent of the data for training and the remaining for testing. Using 60 % of the time series provided the lowest MSE, although comparable to when using 65 % of the data for training, while using 55 % of the data for training had an increase in error likely because of lacking enough data to train on. The testing results lacked the power to make accurate monthly predictions in comparison to the model fit on the whole data but could predict the long-term trends of ciguatera poisoning cases. This suggests the optimal pruned model is not just overfit to the data but rather the predictors have a relationship to the overall trend of ciguatera cases in Florida.

The validation was performed using three time series with train-test splits with the first 65, 60, and 55 % of the data for training, and the

remaining portion, i.e., 35, 40, and 45 %, used for testing the models. These three time series splits were selected to have enough data for both training and testing. The testing model's raw predictions and the number of cases over time were plotted in Fig. 8. This process was repeated to observe general trends using loess smoothing at a span of 0.25 on the predictions and cases. The mean squared error (MSE) of the raw model predictions with the true number of ciguatera poisoning cases was calculated and superimposed onto each plot. The two predictions using 65 and 60 % of the data for training had a smaller MSE than the predictions of the model using fewer data for training. While the predictions using the models with 60 and 65 % training only differ by about 1.18 MSE, the 60 % trained model predictions had a better fit, following the dominant trends of the number of ciguatera poisoning cases in Florida, for its testing time range. More specifically, for 65/35 split, the model captures many of the fluctuations in case counts, particularly in periods of moderate incidence. Due to smoothing of extreme variations, some peak outbreaks are underestimated, while some troughs are overestimated. For 60/40 split, yields the lowest MSE with a balanced performance. The model tracks the temporal dynamics of case counts reasonably well, with improved alignment to both peaks and troughs compared to the 65/35 split. The smooth predictions show that the general trends are predicted with fair accuracy. For 55/45 split with a larger proportion of test data, the model does not track the fluctuations yielding a higher MSE. Model predictions using this data split are particularly smoother than observed data, leading to poor alignment with outbreak peaks and increased error.

## 5. Conclusions

Although the proposed model varied in capturing the different aspects of the data, the model appeared to predict the observed ciguatera poisoning cases reasonably accurately and underestimated or overestimated the number of ciguatera poisoning cases depending on the year. In 2006, both models poorly predicted cases of ciguatera poisoning. This is potentially because 2006 was the first year of the study when reporting was likely to be more sporadic than in other years. This study has some other limitations, including the lack of monthly reporting per county, which hopefully can be resolved for future studies.

In conclusion, the optimal zero-inflated negative binomial model suggested the strongest predictors of the number of reported ciguatera poisoning in Florida were short-term HotSpot values, precipitation, the number of storms, and Amberjack landings. These models provide a basis for predicting Ciguatera fish poisoning, a relatively unknown and potentially upcoming health and economic issue in Florida. One of the main challenges of the study was the lack of a complete geographical data set for both response and predictor variables. Obtaining more complete and accurate data sets on reported cases, such as what type of fish resulted in the poisoning, or the number of fish caught for each fish species in each county, would help improve the predictive models. The number of tropical cyclones had a positive relationship with Ciguatera outbreaks, and future work should be performed to further investigate this impact. This research agrees with studies in the Pacific Ocean that show the increase in intensity and frequency of severe storms increases the likelihood of Ciguatera outbreaks (Rongo and van Woesik, 2013). With hurricanes in Florida becoming more severe, and with an increase in anomalous ocean temperatures, increasing the intensity and frequency of hurricanes, further research should investigate the implications of global climate change on ciguatera poisoning in Florida. The data collection for ciguatera poisoning cases can be improved by implementing protocols for more accurate diagnosis and reporting. In addition, an improved database on ciguatera poisoning cases will help to reduce the risk of ciguatera poisoning, potentially protecting millions of Florida residents and tourists annually.

In this work, we studied the Ciguatera poisoning trends and proposed a model for prediction of Ciguatera poisoning counts. This is a particularly challenging task due to the interplay of multiple and often

nonlinear factors such as environmental changes (e.g., algal blooms), dietary exposure pathways, reporting variability, and population movement. These inherent complexities mean that perfect alignment between predicted and observed cases is unlikely. The proposed model yields moderate but meaningful trend detection performance. While the MSE values are not exceptionally low by standard forecasting benchmarks, they are decent and promising in this context. Importantly, the model captures the seasonal and cyclical trends in Ciguatera case incidence which is critical for public health preparedness. The model performs better at predicting broad temporal dynamics rather than exact case counts which are evident in the smooth graphs where predicted trends closely follow the observed cycles, even if individual peaks differ in magnitude. Such performance is valuable, specifically where forecasting whether the trend is likely to rise or fall may be far more actionable than predicting exact numbers. It was revealed that there is a trade-off between train/test split ratios. The 60/40 split demonstrates the best overall performance with lowest MSE and the best alignment with observed dynamics. Too little training data in 55/45 split reduces predictive power, while too much training data in 65/35 split may limit generalization. This balance highlights the importance of carefully selecting data partitioning strategies for prediction of rare events. In summary, although the MSE values suggest there is room for improvement, the model's ability to capture temporal fluctuations and broad risk dynamics of Ciguatera poisoning should be viewed as a success given the inherent difficulty of modeling this condition. From a public health standpoint, even moderately accurate forecasts can guide early warning systems, targeted awareness campaigns, and proactive clinical preparedness in regions prone to outbreaks. This establishes a valuable baseline model that can be iteratively refined through incorporation of environmental predictors such as sea surface temperature, harmful algal bloom, dietary patterns, and socio-economic factors, which could further reduce predictive error and enhance reliability.

### Author statement

We, the authors, have addressed all reviewers' comments and accordingly revised the article. We thoroughly proofread the manuscript and marked the modifications and additions in green.

We are grateful for your consideration of our revised manuscript and the feedback provided.

### CRediT authorship contribution statement

**Daniel Breininger:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Nezamoddin N. Kachouie:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Robert van Woesik:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Michael Splitt:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Motti Goldberger:** Writing – original draft, Visualization, Investigation, Formal analysis, Data curation. **Christopher Ryzowicz:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This work was done as part of the REU (Research Experiences for Undergraduates) program in Statistical Models with Applications to Geoscience, awarded by NSF (Grant # 1950768).

### Data availability

Publicly available data will be shared.

### References

- Ansdell, V., 2019. Seafood poisoning. In: Keystone, J.S., Kozarsky, P.E., Connor, B.A., Nothdurft, H.D., Mendelson, M., Leder, K. (Eds.), *Travel medicine*, 4th. Elsevier, pp. 449–456. <https://doi.org/10.1016/B978-0-323-54696-6.00049-5>.
- Bagnis, R., Chanteau, S., Chungue, E., Hurtel, J.M., Yasumoto, T., Inoue, A., 1980. Origins of ciguatera fish poisoning: a new dinoflagellate, *Gambierdiscus toxicus* adachi and fukuyo, definitively involved as a causal agent, 1 *Toxicon* 18 (2), 199–208, 1.
- Baker, A.C., Glynn, P.W., Riegl, B., 2008. Climate change and coral reef bleaching: an ecological assessment of long-term impacts, recovery trends and future outlook. *Estuar. Coast. Shelf Sci.* 80 (4), 435–471.
- Barrett, J., 2014. Under the weather with ciguatera fish poisoning: climate variables associated with increases in suspected cases. *Environ. Health Perspect.* 122.
- Chan, T.Y., 2016. Characteristic features and contributory factors in fatal ciguatera fish poisoning—implications for prevention and public education. *NCBI* 94 (4), 704–709, 4.
- Florida Department of Health. Reportable Diseases Frequency Report. Accessed June 2021 [Online]. URL <https://www.flhealthcharts.gov/ChartsReports/rdPage.aspx?rdReport=FrequencyMerlin.Frequency&FirstTime=True>.
- Florida Fish and Wildlife Conservation Commission. Commercial Fisheries Landings Summaries. Accessed June 2021 [Online]. URL <https://app.myfwc.com/FWRI/P/FDM/ReportCreator.aspx>.
- Friedman, M.A., Fernandez, M., Backer, L.C., Dickey, R.W., Bernstein, J., Schrank, K., Kibler, S., Stephan, W., Gribble, M.O., Bienfang, P., Bowen, R.E., Degrasse, S., Flores Quintana, H., Loeffler, C.R., Weisman, R., Blythe, D., Berdalet, E., Ayyar, R., Clarkson-Townsend, D., Swajian, K., Benner, R., Brewer, T., Fleming, L.E., Litaker, R. W., 2017. An updated review of ciguatera fish poisoning: Clinical, epidemiological, environmental, and public health management. *Marine Drugs* 15 (3), 72. <https://doi.org/10.3390/md15030072>.
- Friedman, M.A., Fleming, L.E., Fernandez, M., Bienfang, P., Schrank, K., Dickey, R., Bottein, M.-Y., Backer, L., Ayyar, R., Weisman, R., et al., 2008. Ciguatera fish poisoning: treatment, prevention and management. *Mar. Drugs* 6 (3), 456–479.
- Gingold, D.B., Strickland, M.J., Hess, J.J., 2014. Ciguatera fish poisoning and climate change: analysis of national poison center data in the United States, 2001–2011. *Environ. Health Perspect.* 122 (6), 580–586.
- Gouhier, T

- Rongo, T., van Woesik, R., 2013. The effects of natural disturbances, reef state, and herbivorous fish densities on ciguatera poisoning in rarotonga, Southern Cook Islands. *Toxicon* 64, 87–95.
- Tester, P.A., Feldman, R.L., Nau, A.W., Kibler, S.R., Wayne Litaker, R., 2010. Ciguatera fish poisoning and sea surface temperatures in the Caribbean sea and the west indies. *Toxicon* 56 (5), 698–710.
- Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. *Bull. Am. Meteor. Soc.* 79, 61–78.
- UCLA Statistical Methods and Data Analytics. (n.d.). Zero-inflated negative binomial regression. Retrieved June 29, 2021, from <https://stats.oarc.ucla.edu/r/dae/zinb/>.
- University of Florida Bureau of Economic and Business Research. (n.d.). Florida estimates of population. Retrieved June 2021, from <https://www.bebr.ufl.edu/population/population-data-archive/>.