AQUATIC ECOLOGY

REVISTA DE
Biología Tropical

Satellite and historical data, and statistical modeling to predict potential fishing zones for dolphinfish, *Coryphaena hippurus* (Perciformes: Coryphaenidae) in Colombian Pacific

Adriana Martínez Arias[1]; https://orcid.org/0000-0002-2457-4047
Luis Octavio González Salcedo[1]; https://orcid.org/0000-0003-2460-6106
Iván Felipe Benavides Martínez[2]; https://orcid.org/0000-0002-1139-3909
John Josephraj Selvaraj[1]; https://orcid.org/0000-0002-9195-4883

1. Universidad Nacional de Colombia – sede – Palmira, Facultad de Ingeniería y Administración, Recursos Hidrobiológicos ; GEAL & GIMMA. Carrera 32 # 12 - 00, Palmira, 763533, Colombia; amartinezar@unal.edu.co, logonzalezsa@unal.edu.co, jojselvaraj@unal.edu.co
2. Universidad Nacional de Colombia – sede – Tumaco, Instituto de Estudios del Pacífico, San Andrés de Tumaco, 52835 Colombia; ibenavidesm@unal.edu.co

ABSTRACT

Introduction: The prediction of potential fishing areas is considered one of the most immediate and practical approaches in fisheries and is an essential technique for decision-making in managing fishery resources. It helps fishermen reduce their fuel costs and the uncertainty of their fish catches; this technique allows to contribute to national and international food security. In this study, we build different combinations of predictive statistical models such as Generalized Linear Models and Generalized Additive Models.

Objective: To predict the spatial distribution of PFZs of the dolphinfish (*Coryphaena hippurus* L.) in the Colombian Pacific Ocean.

Methods: We built different combinations of Generalized Linear Models and Generalized Additive Models to predict the Catch Per Unit Effort of *C. hippurus* captured from 2002 to 2015 as a function of sea surface temperature, chlorophyll-*a* concentration, sea level anomaly, and bathymetry.

Results: A Generalized Additive Model with Gaussian error distribution obtained the best performance for predicting PFZs for *C. hipurus*. Model validation was performed by calculating the Root Mean Square Error through a cross-validation approach. The R2 of this model was 50 %, which was considered suitable for the type of data used. January and March were the months with the highest Catch per Unit Effort values, while November and December showed the lower values.

Conclusion: The predicted PFZs of *C. hippurus* with Generalized Additive Models satisfactorily with the results of previous research, suggesting that our model can be explored as a tool for the assessment, decision making, and sustainable use of this species in the Colombian Pacific Ocean.

Key words: potential fishing zones; General Additive Models (GAM); Geographic Information System (GIS); prediction models.

Pelagic fishes such as the common dolphinfish (*Coryphaena hippurus* L) are highly migratory in tropical and subtropical waters (Zúñiga-Flores, 2009). However, similarly to other species, their mobility is restricted by oceanographic variability, which determines their survival, growth, and development (Bellido et al., 2008; Blaxter & Hunter, 1982).

Generally, changes in environmental conditions trigger fish migration, which can last for several months across big areas, generating good conditions for fishing (Bellido et al., 2008).

Several studies have indicated that the environmental variables in oceans influence the distribution of pelagic fish species such as the dolphinfish (Hsu et al., 2021). The physiology of this species restricts populations to occur in areas where the sea surface temperature (SST) varies between 21.5 °C and 27.5 °C, and chlorophyll-a (Chl-*a*) between 0.01 mg/m^3 and 9.5 mg/m^3 (Guzmán et al., 2010). It has been demonstrated that SST is an accurate predictor of habitat distribution, because it holds a close relationship with species richness and plays a vital role in the physiology of pelagic fish. On the other hand, Chl–*a* is recognized as an important oceanographic parameter as it determines primary production and biological productivity in oceans. Several findings suggest that SST and Chl–*a* are key factors influencing patterns of spatial distribution and abundance variability of pelagic fish like tuna, swordfish, and the common dolphinfish. However, for this last species, the spatial distribution of potential fishing zones (PFZ) in the tropical region (particularly in the Colombian Pacific Ocean), and their controlling factors are not clear enough (Zainuddin et al., 2013).

A PFZ is a prediction of fish aggregation in specific areas across the ocean, which is made based on the relationship of fish biology and the environmental context. PFZs are generally made using statistical parameters of SST and Chl-*a* (Natteshan & Kumar, 2016). Different researchers have applied and discussed a wide range of techniques for the prediction of PFZs, including remote sensing of a variety of oceanographic variables, geographic information systems (GIS) and statistical approaches such as Generalized Linear Models (GLM), Generalized Additive Models (GAM) and Classification and Regression Trees (CART). These approaches ultimately find the functional links between the environmental information and fish abundance data to produce a PFZ prediction (Marín-Enríquez et al., 2018).

Research published by Salvadeo et al. (2020) and Marín-Enríquez et al. (2018) show the application of these techniques to identify the distribution of dolphinfish along Southern California and the Mexican coast Herrera-Montiel et al. (2019) adopted this approach to identify and evaluate the PFZs of seerfish (*Scomberomorus sierra*) in the CPO. The identification of PFZs has several challenges associated with the requirements of habitat modeling, because, beyond the relationship between fish abundance and their environment, other factors such as management, the spatiotemporal variation of population sizes, the existence of regulatory zones for fishing and the damage to marine ecosystems have also an influence on fish distribution (Selvaraj et al., 2009).

Colombia is a country where fisheries contribute significantly to the economic growth and food security (FAO, 2021), hence the application of habitat models to determine patterns of abundance, spatial distribution, and the relationship of these patterns with the characteristics of the habitat and the fishing areas, is of crucial importance (Salazar et al., 2021; Selvaraj et al., 2009). Furthermore, considering the rapid changes of fish populations due to climate change in the oceans, it is highly important to determine their PFZs, specially for commercial fishes such as *C. hippurus*. So far, there is no yet available information on this issue.

Consequently, this research aimed at building, running, and comparing different models to predict the PFZs of *C. hippurus* in the CPO, considering the influence of determinant environment conditions for the species suitability.

MATERIALS AND METHODS

Study area: The CPO belongs to the Eastern Pacific Ocean, which is located along the West coast of Colombia. It has a coast length of 1 300 km and borders to the North with the isthmus of Panama and to the South with Ecuador (Fig. 1). The oceanographic conditions of CPO are subject to the seasonal changes in the position of the Inter-Tropical Convergence Zone
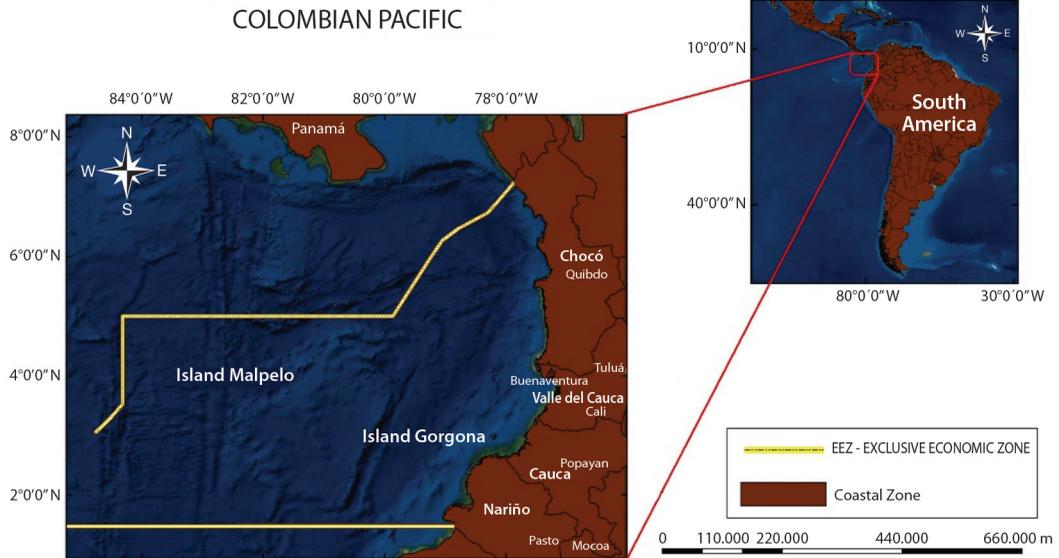
Fig. 1. Location of the study area showing its Exclusive Economic Zone (EEZ).

(ITCZ) and to the strong inter-annual changes due to El Niño South Oscillation (ENSO), both irregular fluctuations that involve changes in the general circulation of winds, insolation and hydrography across the Tropical Eastern Pacific (Rodríguez-Rubio et al., 2003). ENSO implies an unstable interaction between the SST and the atmospheric pressure, causing strong variability in the winds, precipitation and the depth of the thermocline, affecting the biological productivity and the reproduction and survival of marine species (Beltrán, 2011; Fiedler, 2002) (Fig. 1).

Data sources and structure: Catch data of *C. hippurus* was provided by the fishery Sepulveda Rodgers LTDA and contains the historical catch data of semi-industrial fishing boats in the study area during the species's season (November to March), from years 2000 to 2010, 2012 and 2015. These catches were performed with longlines of 1 100 circular hooks. The number of captured fish divided by the 1 100 total hooks is reported here as the Catch per Unit of Effort (Selvaraj, 2010), which is a proportion ranging from 0 to 1. Each CPUE value results from a particular fishing set, from

fishing trips that last between 2 and 10 days at the sea. Each particular CPUE value represents the catch efficiency, and indirectly, the fish abundance during each of the 150 total fishing that occurred during the studied period. All the CPUE values were positive, implying no zero catches. Additionally, each CPUE is linked to a geographical position where each fish trip was performed along the 800 km of the CPO.

Oceanographic variables: Satellite data of SST and Chl-*a* was obtained from the MODIS sensor (*Moderate Resolution Imaging Spectroradiometer*) at the website http://www.oceancolor.gsfc.nasa.gov/. We downloaded daily L3 level data (4 km) that matched the catch dates for the period 2002-2015. Data of Sea Level Anomalies (SLA) was obtained from the AVISO project (*Archiving Validation and Interpretation of Satellite Oceanographic data*), which is supported by the Centre National D'Etudes Spatiales (CNES) in France (Piedrahíta et al., 2013) (http://www.las.aviso.oceanobs.com/las/getUI.do). SLA data was combined to available information from satellites ENVISAT, ERS1/2, GFO, JASON1, JASON 2 and TOPEX/POSEIDON, with a

579

TABLE 1
Data sets used to build the predictive statistical models and their sources

Data sets	Source	Period	Spatial Reference	Resolution	
				Spatial	Temporal
C. hippurus captures	Records from Sepúlveda Rodger Ltda	2000-2010 2012 and 2015	GCS_WGS_1984	NA	Daily
Chl-a	MODIS sensor	2002-2015	GCS_WGS_1984	4 km	Daily
TSM	MODIS sensor	2002-2015	GCS_WGS_1984	4 km	Daily
SLA	ENVISAT, ERS1/2, GFO, JASON1, JASON 2 y TOPEX/POSEIDON	2002-2015	ITRF2008	4 km	Daily
Bathymetry	ETOPE 1	2002-2015	GCS_WGS_1984	1 km	NA

spatial resolution of 0.25 decimal degrees. The bathymetric information was acquired from the website: https://www.ngdc.noaa.gov/mgg/global/ with a spatial resolution of 1 minute of arc (1 km). This resolution was modified to 4 km by using a bilinear resampling algorithm in ArcGIS version 10.5. This oceanographic data set was processed together with the fishing data in order to build the models for PFZ's prediction of *C. hippurus* in the CPO (Table 1).

Statistical modeling: The prediction of the spatial distribution of PFZ's of *C. hippurus* was performed using parametric and semi-parametric statistical models such as GLMs and GAMs respectively. GLMs were used as they allow the use of specific probability distributions appropriate for normal errors (Gaussian), asymmetric errors (Gamma) and errors coming from proportion data (Beta); plus, link functions to account for nonlinear relationships

between predictors and responses (Myers et al., 2012). GAMs were used because besides specifying specific probability distributions and link functions, they offer the flexibility of fitting nonlinear relationships via smoothed functions (spline or kernel) combined with parametric functions. Both models perform parameter estimation via Maximum Likelihood Estimation (MLE), (Wood, 2017).

GAM:

$$g(\mu_i) = \alpha_0 + S_1(x_{1i}) + S_2(x_{2i}) + ...S_n(x_{ni}) + \varepsilon \text{ (1)}$$

Where g is a link function, μ_i is the expected value of the response variable (CPUE), α_0 is a model constant for every oceanographic predictor, S_n is a smoothing function for each predictor and ε is a random error (Wood, 2017) with a specific probability distribution (Table 2).

TABLE 2
Results of the modeling and validation procedures showing the type of model, probability distribution of errors, link-functions and the metrics used for model selection and validation

	Modeling						Validation
Model	Probability distribution	Link-function	Df	AICc	Adjusted R²	D	RMSE
GAM	Gaussian	Log	14	-525.5	0.50	54.2 %	0.037
GAM	Gamma	Log	10	-526.2	0.38	34.9 %	0.040
GAM	Beta	Logit	10	-522.4	0.39	41.1 %	0.040
GLM	Beta	Logit	4	-525.4	0.28	25.2 %	0.044
GLM	Gaussian	Log	6	-494.2	0.27	30.1 %	0.045
GLM	Gamma	Log	6	-519.3	0.17	27.4 %	0.052

Df = degrees of freedom.

GLM:

$$(g)Yi = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_k X_{ki} + \mu_i \quad i1,2,...,T \quad (2)$$

Where is a link function, Yi is the ith observation of CPUE, X_{1i}, X_{2i}, X_{ki} are the ith observations of the k oceanographic predictors, μ_i is the ith term of the random error with a specific probability distribution and β_1, β_2 β_k are the effect parameters for each predictor (Espinoza-Morriberón, 2010).

These models were built with all the oceanographic variables as predictors using a random data subset with 70 % of the dataset (N = 105). The remaining 30 % (N = 45) was kept for model validation (see next section). The Akaike Information Criterion (AIC) (Akaike, 1973; García et al., 2014), the coefficient of determination (R^2) and the explained deviance (*D*) were used as the criteria to select the best model among all candidate models that were built. To identify differences in CPUE between months of the year, we used a Kruskal-Wallis test plus a-posteriori pairwise multiple comparisons (Zar, 2010), where P values lower than 0.05 were considered as statistically significant.

To identify spatial and temporal autocorrelation patterns in the CPUE data that would risk the reliability of the model predictions, we performed Moran tests (Legendre, 1993) and Partial Autocorrelation Functions (PACF) (Tsitsika et al., 2007) respectively on the residuals of the best model selected after model validation. Moran test showed no significant spatial autocorrelation between residual values and the matrix of spatial distances (P = 0.69) between catch points, and the PACF showed only a weak autocorrelation in lag 1 (R = 0.22). For Gaussian GAM and GLM models, normal distribution and homoscedasticity were checked by performing QQ-plots and residuals vs fitted plots. All these assumptions were correctly fulfilled to ensure model validation and applicability for PFZs prediction.

Model validation: We used a cross-validation approach for the predictions of CPUE (Pérez-Planells et al., 2015). This approach

was performed with a random subset containing 30 % of the data that was not involved in modeling. We used the Root Mean Square Error (RMSE) as a metric to evaluate prediction accuracy validate and model performance. This metric quantifies the quality of predictions through the difference between the observed and predicted values (Martínez, 2016). RMSE is calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (p_i - a_i)^2}{n}} \quad (4)$$

Where is the actual observed value, is the predicted value and is the number of observations. All the statistical analyses were performed in R-studio (ver 1.1.442) with packages base, MuMin, ggplot2, mgcv, ape, metrics, betareg, boot and raster.

RESULTS

Oceanographic variables: SST varied between 25.13 and 29.06 ºC, being higher in January and lower in March. Higher values of CPUE occurred in temperatures between 25.5 and 27.5 ºC. Above 27.5 °C and below 25.5 °C, CPUE decreases. Chl–*a* varied between 0.15 and 2.2 mg/m^3, being higher in December and February. Higher values of CPUE occurred in places where Chl–*a* varied between 0.5 and 2.2 mg/m^3, Below and above these values respectively, CPUE decreased. SLA varied between -21.18 and 8.9 cm. Higher CPUE values occurred between -10 and 0.1 cm. Below and above these values respectively, CPUE decreased. Bathymetry varied between -24 and -400 m, with no optimal ranks for CPUE.

Model selection and validation: According to the values shown in Table 2, the Gaussian GAM, Beta GAM, and Gamma GAM were the best candidate models. Despite Gamma GAM had the lowest AICc (slightly lower than the Gaussian GAM), the R^2 and *D* values were considerably higher for the Gaussian GAM. Similarly, the RMSE values for the cross-validation
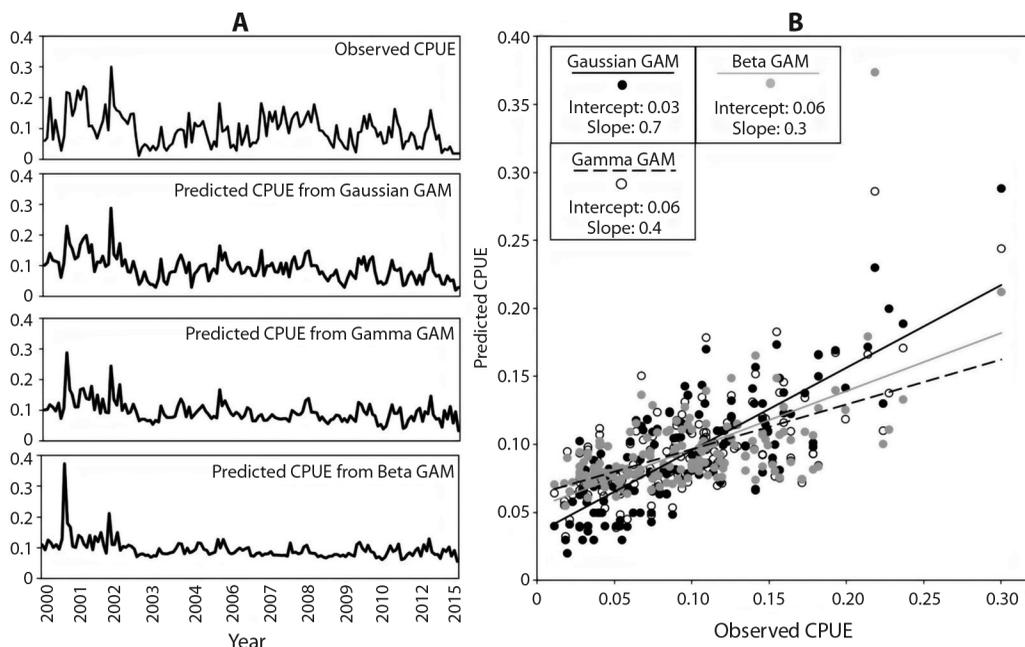
Fig. 2. A. Total observed and predicted CPUE values over capture years for the whole study area according to the three best candidate models: Gaussian GAM, Gamma GAM, and Beta GAM. Predictions are shown in the original CPUE scale after exponential transformation of fitted values. **B.** Relationship between observed and predicted CPUE values according to the three best candidate models. Lines represent a linear regression between the observed and predicted CPUE data for every model. Intercept and slope parameters are shown in the upper boxes for every model.

procedure showed that the predictions resulting from Gaussian GAM were the best fitted to the observed CPUE data. When plotting CPUE predictions over the years (Fig. 2A), Gaussian GAM was able to capture higher time variability than other models. This was particularly conspicuous over periods of abrupt oscillations in fish captures like those between 2001-2003, 2006-2009 and 2010-2015. When comparing the CPUE predictions of every model to the observed CPUE (Fig. 2B) and fitting simple linear regressions on the data, we found again a higher performance for Gaussian GAM, which had a higher slope (0.7) and lower intercept (0.03) in comparison to Gamma GAM and Beta GAM.

Despite all these criteria in favor of Gaussian GAM, it is clear that this model still underestimates CPUE and its R^2 cannot be considered high. However, due to the high variability of these kinds of data both in the spatial and temporal dimensions and to the multiplicity of

factors not able to be included in our models, we considered that the performance of Gaussian GAM was enough as a first approach to the prediction of PFZ's for *C. hippurus* in the CPO. The results of the Gaussian GAM are shown in Table 3. For this model, the intercept as well as the effect of Chl-*a,* SST, and SLA were statistically significant, yet based on the F-values, Chl-*a* had the highest effect on the behavior of CPUE.

Prediction of PFZs: Using the CPUE predictions of the Gaussian GAM, we generated a series of maps according to the monthly seasonality of *C. hippurus* (November-March) for all the capture years (2002-2015) (Fig. 3). These maps show the probability of aggregation of *C. hippurus*, and consequently, reflect the occurrence of PFZs across the CPO. The spatial distribution of PFZs during November expanded uniformly from coordinates 1°00' N & 79°00' W to 8°00' N & 83°00' W along the Pacific

TABLE 3

Parametric coefficients and approximate significance smooth terms resulted from Gaussian GAM model

	Parametric coefficients			
	Estimate	**SE**	**t-value**	**P-value**
Intercept	0.08	0.04	-59.45	< 0.001
	Approximate significance of smooth terms			
	EDF	**RDF**	**F-value**	**P-value**
Chl-*a*	2.84	3.47	17.56	< 0.001
SST	4.86	5.60	2.08	0.05
SLA	3.50	4.28	3.62	0.006
Bathymetry	1.00	1.00	0.01	0.91

EDF = Effective degrees of freedom; SE = Standard error; RDF = Reference degrees of freedom.

coast of Colombia. Meanwhile, the PFZ's distribution in the oceanic area was not uniform in comparison to the coast. However, important fish aggregation was observed between coordinates 6°00' N & 80°00' W (coastal area) and 4°00' N & 82°00' W (oceanic area), especially to the South (border with Ecuador) in the Nariño region, with CPUE values around 0.24. In December, the PFZs are mainly grouped in front of the Nariño region, similarly to the pattern of November, with CPUE values ranging from 0.17 to 0.19 between coordinates 6°00' N & 77°05' W to 1°00' N & 79°05' W near to the coast. Two oceanic conglomerations were observed at coordinates 3°00' N & 81°02' W to 2°00'N & 82°00' W and 2°08' N & 82°03' W to 1°05' N & 82°03' W approximately. Mean CPUE values of December are lower in comparison to November, with maximum values up to 0.234.

For January, PFZs show a persistent occurrence in front of the Colombian coast from coordinates 6°00' N & 77°05' W to 8°00' N & 78°5'W to 1°00' N & 83°1' W, with mean CPUE values around 0.24. This zone starts on the Panama coast and extends southwards to the Chocó and Valle del Cauca regions in the Colombian coast, with a mean CPUE of 0.24, which gradually decreases towards the Cauca and Nariño regions. CPUE values in January were significantly higher than in December

and November for all the study years. During February there were two conspicuous PZFs, one between coordinates 1°17' N & 79°00' W to 6°53' N & 77°42' W and the other at the oceanic area of the Panama basin from coordinates 7°18'5" N & 79°46' W to 3°20°N & 80°83' W. Mean CPUE values in these zones are around 0.18.

Finally, the spatial distribution of PFZs for March showed higher CPUE values. Locations of PFZs persisted near the coast at coordinates 6°00' N & 77°5' W to 1°00' N & 79°05' W and at an oceanic area with coordinates 3°00' N & 81°02' W to 2°00' N & 82°00' W, 2°08' N & 82°03' W and 1°05' N & 82°03' W approximately. Generally, the PFZs were spatially persistent over the months and years. Nevertheless, the PFZs with larger sizes occurred during January, February, and March, with mean CPUE values around 0.25. The difference of CPUE values between all pairs of months were highly significant according to Kruskal-Wallis tests (P < 0.001 for every a-posteriori pairwise multiple comparison).

DISCUSSION

The GIS approach used in this study was successful to capture enough variability in the oceanographic variables, which helped to understand the spatiotemporal dynamics of the PFZs of C. *hippurus*. This is in accordance with the results described by Devis et al. (2002), where the variation in Intertropical Convergence Zone (ITCZ) triggers changes in SST and Chl-*a*. This phenomenon creates a gradual seasonality in climate and oceanic conditions, allowing the presence of different temperatures along the surface water. Concerning the spatial distribution of SST, our results resemble the findings of Rodríguez-Rubio (2003), who mentioned that this variable evolves seasonally between January and March, when surface waters of low temperature (25-26 °C) extend from the Gulf of Panama to the center of the basin in a South-to-Southwest orientation. From April to June the temperature increases gradually eastwards and by July-September, the cold-water plume
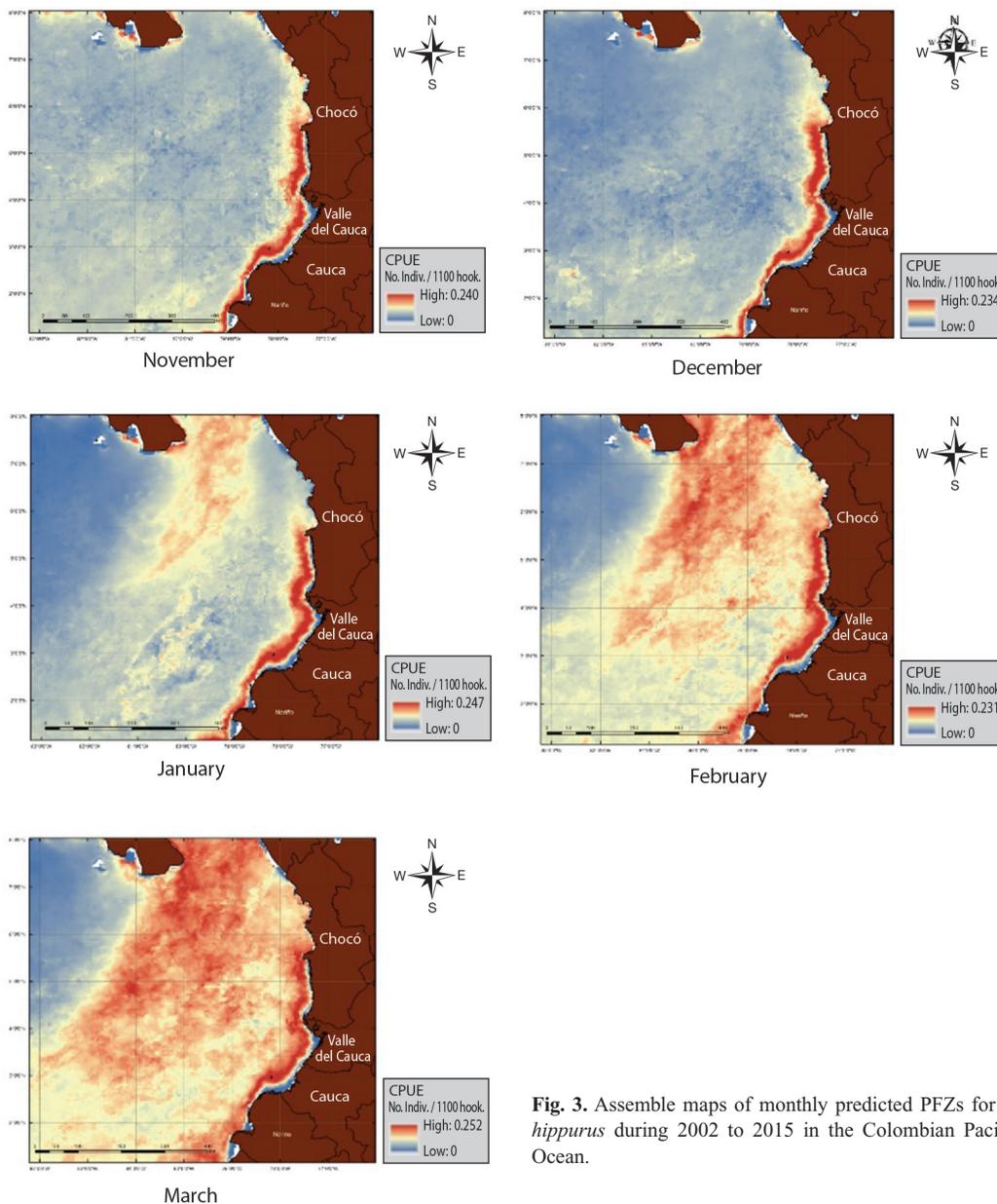
Fig. 3. Assemble maps of monthly predicted PFZs for *C. hippurus* during 2002 to 2015 in the Colombian Pacific Ocean.

disappears altogether with 3 °C increase to the North and 1 °C decrease to the South. By the end of the year (October-December), the spatial pattern of temperature stays, though with a ~1 °C decrease in across the CPO.

On the other hand, the results obtained with the GAM model with Gaussian distribution are in accordance with the findings made by Bigelow et al. (1999), Rodríguez-Marín et al. (2003), Zagaglia et al. (2004), Zainuddin et al. (2013), and Setiawati et al. (2015). These authors indicated that this modeling approach is highly efficient to predict PFZs of pelagic species in the Pacific and Atlantic oceans. This efficiency relies on the fact that GAMs offer high flexibility regarding the type of data

and the link between predictors and responses, allowing non-linear and non-monotonous relationships to make predictions in the response variable (França & Cabral, 2015; Guisan et al., 2002; Yee & Mitchell, 1991). Maunder and Punt (2004), compared the performance of GAMs, GLMs and GLMMs (Generalized Linear Mixed Model) to predict the spatial distribution of valuable fish species such as tuna and sharks, remarking the benefits of GAMs.

Though we consider that the Gaussian GAM model used in this study was successful as a first approach to predict the spatial distribution and PFZs of *C. hippurus* in the CPO, our results suggest that the precision of the model still needs to be increased for better results. When observing the model predictions over the capture years in (Fig. 3), it is clear that for many capture months the model underestimates or overestimates the observed CPUE, with maximum difference values up to 9.6 %. We suggest that these underperformances can be solved by including other biological variables not considered in this study, such as physiologic traits, life-history traits, genetic structuring, predation, and migration rates of *C. hippurus* populations. When including some of these variables in the model, a higher CPUE variability could be captured, hence increasing the chance of making more accurate predictions. This would help the developing of more powerful frameworks for studies that consider fish management and decision making under climate change scenarios.

Our model detected Chl-*a* as the most efficient to predict CPUE, agreeing with Sund et al. (1981), who considers it as the most restrictive oceanic variable for the horizontal and vertical distribution of pelagic species. Furthermore, it has been a successful proxy to estimate the available food in pelagic ecosystems, since this pigment is present in most of the phytoplankton species and forms the basis of trophic chains (Martínez-Rincón, 2012; Sartimbul et al., 2010). Loukos et al. (2003) suggested specific large-scale changes on the fish habitat in the equatorial Pacific by means of SST effects, which could explain the distribution of

available habitats for pelagic fishes on higher latitudes due to the increase of temperature by global warming (Cheung et al., 2009; Mugo et al., 2010). Despite SLA and Bathymetry showed a lower effect on the CPUE, they were included in the model because they are related to the spatial distribution and movement of pelagic species (Maravelias, 1999). SLA is related to mesoscale cyclonic movements that occur at the beginning of the year (Rodríguez-Rubio, 2003), bringing oceanic upwelling, low temperatures, and high chlorophyll concentrations. According to Sabarros et al. (2009), it is highly probable that these mesoscale processes are related to the presence of PFZs of pelagic fishes.

The predicted PFZs for *C. hippurus* and their spatial distribution was similar to the findings drawn by Lasso and Zapata (1999), who mentioned that the higher fish abundances occur during February and March. Furthermore, this season could be associated with phenomena that trigger fish migration, since it coincides with the lowest SST and higher Chl–*a* concentrations in comparison to the rest of the year. Giraldo et al. (2010) and Guzmán et al. (2010) demonstrated that the aggregation of *C. hippurus* are linked to mesoscale oceanic processes, such as the oceanic thermal fronts. Lasso and Zapata (1999) suggested that the migratory path of *C. hippurus* could be associated with surface ocean currents in the eastern Pacific, as a response to the changes driven by ENSO, which cause fish populations to migrate southwards in search for warmer waters (Caetano-Nunes, 2013; Lasso & Zapata, 1999).

The predicted PFZs shown by our model from February and March agree with the reports of the highest catch rates in Colombia (Lasso & Zapata, 1999) an also, with the isotherm changes from January to April, and the salinity increase in the Pacific Ocean that brings an upsurge of nutrients in front of the Panama Bay (Valencia et al., 2013). Generally, these results show that the distribution patterns of PFZs for *C. hippurus* are associated with changes in SST, which influence fish

aggregation, growth, reproduction, and survival (Boyce et al., 2008; Martínez-Rincón, 2012).

As shown in this study, the identification of PFZs of fish species emerges from the combination of remote sensing, geographic information systems, historical registers of in situ data and statistical modeling tools, which can be used as a robust package to improve the management and decision making of fish resources. As stated by Selvaraj et al. (2009), the maps resulting from this multidisciplinary approach can be used for conservation o extraction purposes. However, to achieve these purposes, we still need a standardization of data sharing, data depuration and analysis and training of the entities and people in charge of its administration and distribution.

The development of these prediction techniques has created new alternatives that allow understanding the behavior and distribution of fish populations. However, the variability of the prediction performances among different models results in contrasting predictions. This implies that a predictive model must be selected very carefully, especially if it is intended for practical applications (Zhang et al., 2017). The advantage of using GAMs relies to their robustness to outliers and its ability to capture complex interactions between predictors. These advances in the model's predictive power allow a more reliable understanding of potential fishing areas, especially in those areas at the local level (Venables & Dichmont, 2004).

ACKNOWLEDGMENTS

RESUMEN

Datos satelitales e históricos, y modelado estadístico, para predecir zonas de pesca potencial del pez dorado, *Coryphaena hippurus* (Perciformes: Coryphaenidae) en el Pacífico colombiano

Introducción: La predicción de zonas potenciales de pesca se considera uno de los enfoques más inmediatos y efectivos en las pesquerías, es una técnica importante para la toma de decisiones en el manejo de los recursos pesqueros. Ayuda a los pescadores a reducir su costo de combustible y también a disminuir la incertidumbre de sus capturas, esta técnica permite contribuir a la seguridad alimentaria nacional e internacional. En este estudio, se construyeron diferentes combinaciones de modelos estadísticos predictivos como modelos lineales generalizados y modelos aditivos generalizados.

Objetivo: predecir la distribución espacial de las zonas potenciales de pesca del pez dorado (*Coryphaena hippurus* L.) en el Pacífico colombiano.

Métodos: La variable de respuesta se expresó en escala de captura por unidad de esfuerzo, es decir, el número de individuos de *C. hippurus* capturados por un número total de anzuelos disponibles entre 2002 y 2015. Temperatura de la superficie del mar, concentración de clorofila, anomalía del nivel del mar y batimetría, se utilizaron como variables explicativas para los meses de estacionalidad de *C. hippurus* (noviembre - marzo).

Resultados: El modelo con mejor rendimiento para la predicción de zonas potenciales de pesca fue un modelo aditivo generalizado con distribución de error gaussiana y función de enlace de registro, que se seleccionó en función del criterio de información de Akaike, el R^2 y la desviación explicada. La validación del modelo se realizó calculando el error cuadrático medio a través de un enfoque de validación cruzada. El ajuste de este modelo fue del 50 %, lo que puede considerarse adecuado para el tipo de datos utilizados. Enero y marzo fueron los meses con mayor captura por unidad de esfuerzo y noviembre-diciembre los meses con menor.

Conclusión: Las zonas potenciales de pesca previstas coincidieron satisfactoriamente con investigaciones anteriores, lo que sugiere que nuestro modelo es una herramienta poderosa para la evaluación, toma de decisiones y uso sostenible de los recursos pesqueros de *C. hippurus* en el Pacífico colombiano.

Palabras clave: zona potencial de pesca; Modelo Aditivo Generalizado (GAM); sistema de información geográfica; modelos de predictivos.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov, & F. Csaki (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Akademiai Kiado.

Bellido, J. M., Brown, A. M., Valavanis, V. D., Giráldez, A., Pierce, G. J., Iglesias, M., & Palialexis, A. (2008). Identifying essential fish habitat for small pelagic species in spanish Mediterranean waters. *Hydrobiologia*, *612*(1), 171–184.

Beltrán, F. M. C. (2011). *Evaluación de las pesquerías de los bancos de pesca artesanal en el Pacífico sur colombiano, Colombia* (Tesis de pregrado). Universidad Nacional de Colombia, Colombia.

Blaxter, J. H. S., & Hunter, J. R. (1982). *The biology of the clupeoid fishes.* Academic Press.

Bigelow, K., Boggs, C., & He, X. (1999). Environmental effects on swordfish and blue shark catch rates in the US North Pacific longline fishery. *Fisheries Oceanography*, *8*(3), 178–198.

Boyce, D., Tittensor, D., & Worm, B. (2008). Effects of temperature on global patterns of tuna and billfish richness. *Marine Ecology Progress Series*, *355*(5), 267–276.

Caetano-Nunes, B. R. (2013). *Caracterización genética y ambiental del dorado (*Coryphaena hippurus L.*) en el Pacífico colombiano* (Tesis doctoral). Universidad Nacional de Colombia, Colombia.

Cheung, W. W., Lam, V. W., Sarmiento, J., Kearney, K., Watson, R., & Pauly, D. (2009). Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries, 10*(3), 235–251.

Devis, A., García, I., Málikov, I., & Villegas, N. (2002). *Compilación oceanográfica de la cuenca pacífica colombiana*. Centro Control Contaminación del Pacífico, Tumaco, Colombia.

Espinoza-Morriberón, D. (2010). *Estandarización de la CPUE de la flota industrial de cerco del stock norte-centro de anchoveta peruana (*Engraulis ringens Jenyns 1842*) entre 1996 y el 2008* (Tesis de pregrado). Universidad Nacional Mayor de San Marcos, Perú.

FAO. (2021). *Visión general del sector acuícola nacional-Colombia. National Aquaculture Sector Overview fact sheets* (Sitio web). https://www.fao.org/fishery/en/countrysector/co/es

Fiedler, P. C. (2002). Environmental change in the eastern tropical Pacific Ocean: Review of ENSO and decadal variability. *Marine Ecology Progress Series*, *244*(2002), 265–283.

França, S., & Cabral, H. N. (2015). Predicting fish species richness in estuaries: Which modelling technique to use? *Environmental Modelling & Software*, *66*(2015), 17–26.

García, M. D. C., Castellana, N., Rapelli, C., Koegel, L., & Catalano, M. (2014). Criterios de información y predictivos para la selección de un modelo lineal mixto. *SaberE*s, *6*(2), 61–76.

Giraldo, A., Payán, L., & Chirima, H. (2010). *Variabilidad térmica en el ambiente pelágico de la Isla Gorgona entre marzo 2009-febrero 2010* (Presentación oral). XIV Semanario Nacional de Ciencia y Tecnologías del Mar, Calí, Colombia.

Guisan, A., Edwards, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, *157*(2), 89–100.

Guzmán, A. I., Selvaraj, J. J., & Martínez, A. (2010). *Variación de la temperatura superficial del mar, clorofila "a" y su relación con capturas de dorado (*Coryphaena hippurus*) en la costa pacífica colombiana* (Presentación oral). III Congreso Brasileiro de Oceanografía, Fortaleza, Brasil.

Herrera-Montiel, S. A., Coronado-Franco, K. V., & Selvaraj, J. J. (2019). Predicted Changes in the Potential Distribution of Seerfish (*Scomberomorus sierra*) under multiple climate change scenarios in the Colombian Pacific Ocean. *Ecological Informatics, 53*(2019), 100985.

Hsu, T. Y., Chang, Y., Lee, M. A., Wu, R. F., & Hsiao, S. C. (2021). Predicting skipjack tuna fishing grounds in the western and central Pacific Ocean based on high-spatial- temporal-resolution satellite data. *Remote Sensing*, *13*(5), 1–17.

Lasso, J., & Zapata, L. (1999). Fisheries and biology of *Coryphaena hippurus* (Pisces: Coryphaenidae) in the Pacific coast of Colombia and Panama. *Scientia Marina*, *63*(3-4), 387–399.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology*, *74*(6), 1659–1673.

Loukos, H., Monfray, P., Bopp, L., & Lehodey, P. (2003). Potential changes in skipjack tuna (*Katsuwonus pelamis*) habitat from a global warming scenario: modeling approach and preliminary results. *Fisheries Oceanography, 12*(4-5), 474–482.

Maravelias, C. D. (1999). Habitat selection and clustering of a pelagic fish: effects of topography and bathymetry on species dynamics. *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(3), 437–450.

Marín-Enríquez, E., Seoane, J., & Muhlia-Melo, A. (2018). Environmental modeling of occurrence of dolphinfish (*Coryphaena* spp.) in the Pacific Ocean off Mexico reveals seasonality in abundance, hot spots

587

and migration patterns. *Fisheries Oceanography, 27*(1), 28–40.

Martínez, D. N. (2016). *Predicción y Análisis de los Retrasos en los Vuelos* (Tesis de pregrado). Universidad Autónoma de Barcelona, España.

Martínez-Rincón, R. O. (2012). *Efecto de la variabilidad ambiental en la distribución de las capturas incidentales de pelágicos mayores en el Océano Pacífico Oriental* (Tesis doctoral). Instituto Politécnico Nacional, México.

Maunder, M. N., & Punt, A. E. (2004). Standardizing catch and effort data: a review of recent approaches. *Fisheries Research, 70*(2), 141–159.

Mugo, R., Saitoh, S. I., Nihira, A., & Kuroyama, T. (2010). Habitat characteristics of skipjack tuna (*Katsuwonus pelamis*) in the Western North Pacific. *Fisheries Oceanography, 19*(5), 382–396.

Myers, R. H., Montgomery, D. C., Vining, G. G., & Robinson, T. J. (2012). *Generalized linear models: with applications in engineering and the sciences* (2nd Ed.)*.* John Wiley & Sons Inc.

Natteshan, N. V. S., & Kumar, N. S. (2016). A survey of various methods of potential fishing zone identification. *International Journal of Control Theory and Applications, 9*(40), 697–703.

Pérez-Planells, L., Delegido, J., Rivera-Caicedo, J. P., & Verrelst, J. (2015). Análisis de métodos de validación cruzada para la obtención robusta de parámetros biofísicos. *Revista Española de Teledetección*, *4*(2015), 55–65.

Piedrahíta, C., Selvaraj, J. J., & Guzmán, A. I. (2013). Detección de afloramientos de algas usando datos modis de fluorescencia en el pacífico colombiano. En L. M. Marina, & C. García (Eds.), *Investigación en ciencias del mar* (pp. 57–69). Aportes de la Universidad Nacional de Colombia.

Rodríguez-Marín, E., Arrizabalaga, H., Ortiz, M., Rodríguez-Cabello, C., Moreno, G., & Kell, L. T. (2003). Standardization of bluefin tuna, *Thunnus thynnus*, catch per unit effort in the baitboat fishery of the Bay of Biscay (Eastern Atlantic). *ICES Journal of Marine Science, 60*(6), 1216–1231.

Rodríguez-Rubio, E. (2003). *Variabilidad de la circulación de mesoescala en la cuenca oceánica del Pacífico colombiano* (Tesis doctoral). Universidad de Concepción, Chile.

Rodríguez-Rubio, E., Schneider, W., & Abarca del Río, R. (2003). On the seasonal circulation within the Panama Bight derived from satellite observations of wind, altimetry and sea surface temperature. *Geophysical Research Letters*, *30*(7), 631–634.

Sabarros, P. S., Ménard, F., Lévénez, J. J., Tew-Kai, E., & Ternon, J. F. (2009). Mesoscale eddies influence distribution and aggregation patterns of micronekton in the Mozambique Channel. *Marine Ecology Progress Series, 395*(2009), 101–107.

Salazar, J. E., Benavides, I. F., Cabrera, C. V. P., Guzmán, A. I., & Selvaraj, J. J. (2021). Generalized additive models with delayed effects and spatial autocorrelation patterns to improve the spatiotemporal prediction of the skipjack (*Katsuwonus pelamis*) distribution in the Colombian Pacific Ocean. *Regional Studies in Marine Science*, *45*(2021), 101829.

Salvadeo, C., Auliz-Ortiz, D. M., Petatán-Ramírez, D., Reyes-Bonilla, H., Ivanova-Bonchera, A., & Juárez-León, E. (2020). Potential poleward distribution shift of dolphinfish (*Coryphaena hippurus*) along the southern California current system. *Environmental Biology of Fishes*, *103*(8), 973–984.

Sartimbul, A., Nakata, H., Rohadi, E., Yusuf, B., & Kadarisman, H. (2010). Variations in chlorophyll-a concentration and the impact on *Sardinella lemuru* catches in Bali Strait, Indonesia. *Progress in Oceanography*, *87*(1-4), 168–174.

Selvaraj, J. J., Martínez, A., Guzmán, A. I. (2010). *Environmental preference of dolphinfish (*Coryphaena hippurus*) derived from remotely sensed data along the Pacific coast of Colombia* (Oral presentation). ICES Annual Science Conference, Nantes, France.

Selvaraj, J. J., Rajasekharan, M., & Guzmán, Á. I. (2009). Aplicaciones de los sistemas de información geográfica y sensores remotos al manejo de pesquerías marinas y desafíos para su desarrollo en Colombia. *Boletín de Investigaciones Marinas y Costeras, 38*(1), 105–120.

Setiawati, M. D., Sambah, A. B., Miura, F., Tanaka, T., & As-Syakur, A. R. (2015). Characterization of bigeye tuna habitat in the Southern waters off Java-Bali using remote sensing data. *Advances in Space Research, 55*(2), 732–746.

Sund, P. N., Blackburn, M., & Williams, F. (1981). Tunas and their environment in the Pacific Ocean: a review. *Oceanography and Marine Biology an Annual Review, 1981*(19), 443–512.

Tsitsika, E. V., Maravelias, C. D., & Haralabous, J. (2007). Modeling and forecasting pelagic fish production using univariate and multivariate ARIMA models. *Fisheries Science*, *73*(5), 979–988.

Valencia, B., Lavaniegos, B., Giraldo, A., & Rodríguez-Rubio, E. (2013). Temporal and spatial variation of hyperiid amphipod assemblages in response to hydrographic processes in the Panama Bight, eastern tropical Pacific. *Deep-Sea Research I, 73*(2013), 46–61.

Venables, W. N., & Dichmont, C. M. (2004). GLMs, GAMs and GLMMs: an overview of theory for applications

in fisheries research. *Fisheries Research, 70*(2-3), 319–337.

Wood, S. M. (2017). *Generalized additive models, an introduction with R* (2nd Ed.). Chapman and Hall/CRC Press.

Yee, T. W., & Mitchell, N. D. (1991). Generalized additive models in plant ecology. *Journal of Vegetation Science, 2*(5),587–602.

Zagaglia, C., Lorenzzetti, J., & Stech, J. (2004). Remote sensing data and longline catches of yellowfin tuna (*Thunnus albacares*) in the equatorial Atlantic. *Remote Sensing of Environment, 93*(1), 267–281.

Zainuddin, M., Nelwan, A., Farhum, S. A., Hajar, M. A. I., & Kurnia, M. (2013). Characterizing potential fishing zone of skipjack tuna during the southeast monsoon in the Bone Bay-Flores Sea using remotely sensed oceanographic data. *International Journal of Geosciences, 4*(1), 259–266.

Zar, J. H. (2010). *Biostatistical analysis* (5th Ed.). Prentice Hall.

Zhang, X., Saitoh, S. I., & Hirawake, T. (2017). Predicting potential fishing zones of Japanese common squid (*Todarodes pacificus*) using remotely sensed images in coastal waters of south-western Hokkaido, Japan. *International Journal of Remote Sensing, 38*(21), 6129–6146.

Zúñiga-Flores, M. S. (2009). *Dinámica poblacional del dorado (*Coryphaena hippurus*) en Baja California Sur, México: implicaciones para su manejo* (Tesis doctoral). Instituto Politécnico Nacional, México.