

Modelling the spatial population structure and distribution of the queen conch, *Aliger gigas*, on the Pedro Bank, Jamaica

Ricardo A. Morris, Alvaro Hernández-Flores, Alfonso Cuevas-Jimenez

Universidad Marista de Mérida, Periférico Norte Tablaje 13941, Carretera Mérida-Progreso,
C.P. 97300 Mérida, Yucatan, Mexico.

(RAM) E-mail: ranthonymorris@gmail.com. ORCID-iD: <https://orcid.org/0000-0003-4759-5366>

(AH-F) (Corresponding author) E-mail: ahernandez@marista.edu.mx. ORCID-iD: <https://orcid.org/0000-0003-1900-9868>

(AC-J) E-mail: acuevas@marista.edu.mx. ORCID-iD: <https://orcid.org/0000-0002-8230-5021>

Summary: The estimation of reliable indices of abundance for sedentary stocks requires the incorporation of the underlying spatial population structure, including issues arising from the sampling design and zero inflation. We applied seven spatial interpolation techniques [ordinary kriging (OK), kriging with external drift (KED), a negative binomial generalized additive model (NBGM), NBGM plus OK (NBGM+OK), a general additive mixed model (GAMM), GAMM plus OK (GAMM+OK) and a zero-inflated negative binomial model (ZINB)] to three survey datasets to estimate biomass for the gastropod *Aliger gigas* on the Pedro Bank Jamaica. The models were evaluated using 10-fold cross-validation diagnostics criteria for choosing the best model. We also compared the best model estimations against two common design methods to assess the consequences of ignoring the spatial structure of the species distribution. GAMM and ZINB were overall the best models but were strongly affected by the sampling design, sample size, the coefficient of variation of the sample and the quality of the available covariates used to model the distribution (geographic location, depth and habitat). More reliable abundance indices can help to improve stock assessments and the development of spatial management using an ecosystem approach.

Keywords: spatial analysis; sedentary species; zero-inflation; species distribution models.

Modelación espacial de la estructura y distribución de la población de caracol rosado (*Aliger gigas*) en Pedro Bank, Jamaica

Resumen: Las estimaciones de índices de abundancia relativa para evaluar poblaciones de especies sedentarias requieren tener en cuenta su estructura espacial, el diseño del muestreo y el alto número de ceros registrados a la hora del muestreo. Para obtener un índice confiable de abundancia para el gasterópodo *Aliger gigas* en Pedro Bank, Jamaica, se aplicaron siete técnicas de interpolación espacial a tres conjuntos de datos: kriging ordinario (OK), kriging con desviación externa (KED), modelo aditivo generalizado binomial negativo (NBGM), NBGM más OK (NBGM+OK), modelo mixto aditivo general (GAMM), GAMM más OK (GAMM+OK) y modelo binomial negativo con ceros inflados (ZINB). La selección de los mejores modelos espaciales se basó en el criterio de validación cruzada con 10 iteraciones; asimismo, se aplicaron métodos de evaluación comúnmente usados, para destacar la importancia de tener en cuenta la estructura espacial de la distribución de la especie. Los mejores modelos fueron GAMM y ZINB, los cuales fueron fuertemente influenciados por el diseño de muestreo, el tamaño de muestra, el coeficiente de variación y la calidad de las covariables empleadas en la modelación (ubicación geográfica, profundidad y hábitat). Los índices de abundancia más confiables pueden contribuir a mejorar las evaluaciones y desarrollar el manejo espacial con enfoque de ecosistema.

Palabras clave: análisis espacial; especies sedentarias; distribución con ceros inflados; modelos de distribución de especies.

Citation/Como citar este artículo: Morris R.A., Hernández-Flores A., Cuevas-Jimenez A. 2022. Modelling the spatial population structure and distribution of the queen conch, *Aliger gigas*, on the Pedro Bank, Jamaica. Sci. Mar. 86(3): e040. <https://doi.org/10.3989/scimar.05269.040>

Editor: M. Gaspar.

Received: February 2, 2022. **Accepted:** June 21, 2022. **Published:** August 23, 2022.

Copyright: © 2022 CSIC. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License.

INTRODUCTION

Sedentary species aggregate over space to carry out biological and social functions resulting in a highly complex heterogeneous distribution (Stoner and Appeldoorn 2021). This spatial structure of populations becomes even more complex when these species are subject to exploitation. The estimation of precise indices of abundance critical for stock assessment must therefore consider this spatial structure (Rivoirard et al. 2000). Previous authors have addressed the spatial structure of sedentary species using methods based on assumptions suitable for determining data sets (Surette et al. 2007, Drexler and Ainsworth 2013, Lyashevskaya et al. 2016). Three of the most commonly applied spatial methods are geostatistical models (Matheron 1963), regression models (Lambert 1992, Hall 2000) and hybrid models (Hengl et al. 2007). The multi-model spatial analysis approach is often applied and has proven robust for modelling spatial distribution and improving the estimation of indices of abundance (Li and Heap 2011). This is because specific species distribution characteristics can be modelled and evaluated using various criteria to determine the most reliable one for a particular dataset (Rufino et al. 2021).

Geostatistical models are a group of spatial analysis and interpolation methods based on the theory of the regionalized variable (Matheron 1963), among which the most common is kriging (Kriging 1951). These have been applied to estimate the abundance and biomass of various sedentary fisheries species, including the snow crab *Chionoecetes opilio* (Surette et al. 2007), the sea cucumber *Isostichopus badionotus* (Hernández-Flores et al. 2015) and the red octopus *Octopus maya* (Avendaño et al. 2019). Generalized additive models (GAM) are regression methods that can model non-linear response-explanatory relationships without a priori assumptions about the probability distribution of the data (Hastie and Tibshirani 1990). These methods have also been applied in fisheries spatial analysis assuming Poisson probability distribution (Lyashevskaya et al. 2016), but for sedentary species the negative binomial probability may be more appropriate (Anderson and Seijo 2010). Zero-inflated models are another set of regression techniques with the advantage of directly modelling the over-dispersion and excessive zeros or zero inflation common to environmental data (Potts and Elith 2006). These techniques have been used to estimate abundance indices in fisheries stock assessment (Potts and Rose 2018). Each method can maintain the original spatial structure of sample datasets without the need to transform or remove outliers (Hastie and Tibshirani 1990, Martin et al. 2005, Webster and Oliver 2007). Transformation and removal of outliers of environmental data may risk the loss of valuable ecological information about the species (Arab et al. 2008), leading to model miscalibrations and incorrect interpretations (Martin et al. 2005).

In this study, we conducted a comprehensive comparative spatial analysis using seven models: ordinary kriging (OK), kriging with external drift (KED), negative binomial GAM (NBGM), a general additive

mixed model (GAMM), a zero-inflated negative binomial model (ZINB) and two hybrid regression kriging models, NBGM+OK and GAMM+OK. These particular models were chosen because they can deal with highly skewed and zero-inflated data. They were applied to three surveys of the gastropod queen conch, *Aliger gigas* (Linnaeus 1758), conducted in 2007, 2015 and 2018 on the Pedro Bank off the southern coast of Jamaica.

The queen conch is a highly exploited species on the Pedro Bank and throughout its wider Caribbean range, particularly over the last forty years (Prada et al. 2017). However, most queen conch assessments apply design-based methods (Cochran 1977) which do not consider the heterogeneous spatial structure of the species to determine abundance indices (Ehrhardt and Valle-Esquivel 2008, Baker et al. 2016). The fundamental issue with the design methods is that its assumptions of spatial homogeneity, perfect mixing of ages and sizes, and instantaneous redistribution after changes in abundance are invalid for sedentary species (Anderson and Seijo 2010). The failure of the design methods to account for spatial heterogeneity increases the risk of bias and estimation errors.

We aimed firstly to assess the performance of alternative spatial models in estimating biomass given the structure of the data. Secondly, we determined the implications of not accounting for the spatial population structure by comparing the spatial model estimations with two design methods, one with a stratified and the other with a non-stratified sampling design. The performance of the spatial interpolation models was evaluated using a ten-fold cross-validation residual analysis and visual examination of the resultant species distribution maps. The results of the analyses are discussed in terms of the main data issues presented by each survey data set, including sample design, zero-inflation and over-dispersion of variances, and how they affect the reliability of the estimations. The implications for spatial management of the species in light of its ecology are also discussed. All modelling procedures and the output thereof, including maps, were carried out and produced using packages available in the R program language (R Core Team 2021).

METHOD

Study site

The Pedro Bank is a shallow offshore bank located in the west-central Caribbean approximately 80 km south of the main island of Jamaica (Fig. 1). The bank has a total area of 8040 km² with maximum dimensions of 180 km west to east and 79 km north to south down to the 30 m depth contour (Morris 2016). The average depth is 24 m and the bottom consists predominantly of a coralline sand substrate which makes up approximately two-thirds of the bank. Infused throughout this primary substrate are patch coral reefs, coral pavement (weathered coral reefs), various species of macroalgae and gorgonians (Baldwin 2015). This habitat complex accommodates a marine ecosystem that supports im-

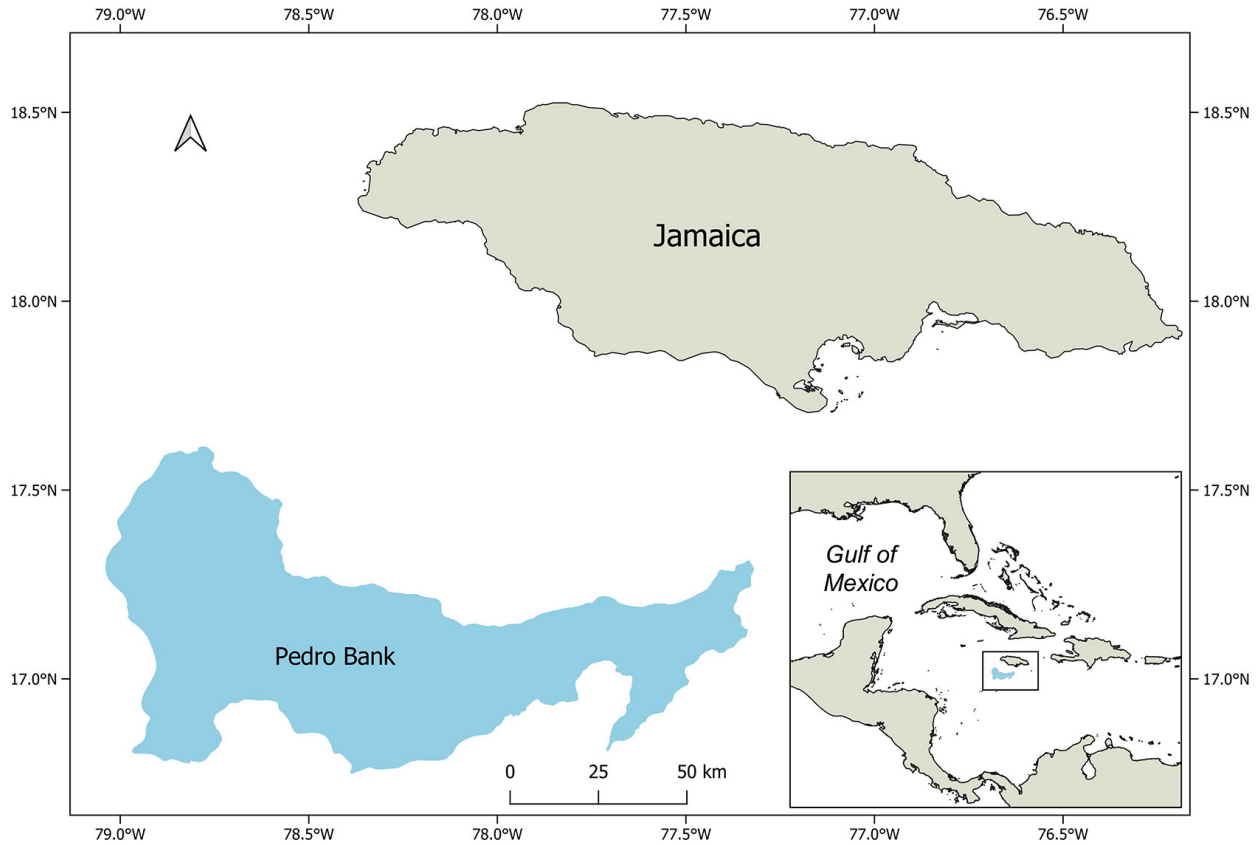


Fig. 1. – Location of the Pedro Bank, Jamaica.

portant fisheries for various finfish, spiny lobster (*Panulirus argus*) and queen conch (Baldwin 2015). The central and southern areas of the bank are known to contain the highest densities of both juvenile and adult conch (Morris 2016).

Data

Distribution data were compiled from underwater visual surveys conducted by the National Fisheries Authority of Jamaica (NFA) for the years 2007, 2015 and 2018. The NFA has conducted such surveys at three to five-year intervals since 1994, but only after 2007 did sampling target the entire extent of the bank. The years 2007, 2015 and 2018 were chosen to contrast sample design, sample sizes and spatial population structures such as the size and location of high-density patches. The 2007 survey consisted of 60 sites clustered in the centre of the study area, while the 2015 and 2018 surveys included 80 and 81 sites, respectively, systematically distributed down to the 30-metre contour (Fig. 2). Sample sites were defined as one to four belt transects with an area of 220 to 2250 m² spaced one metre from the end of one transect to the start of the following one. The number and area of transects were dependent on ocean conditions and diver safety considerations at the time of sampling. Data recorded within each transect included geographic location, depth, transect area,

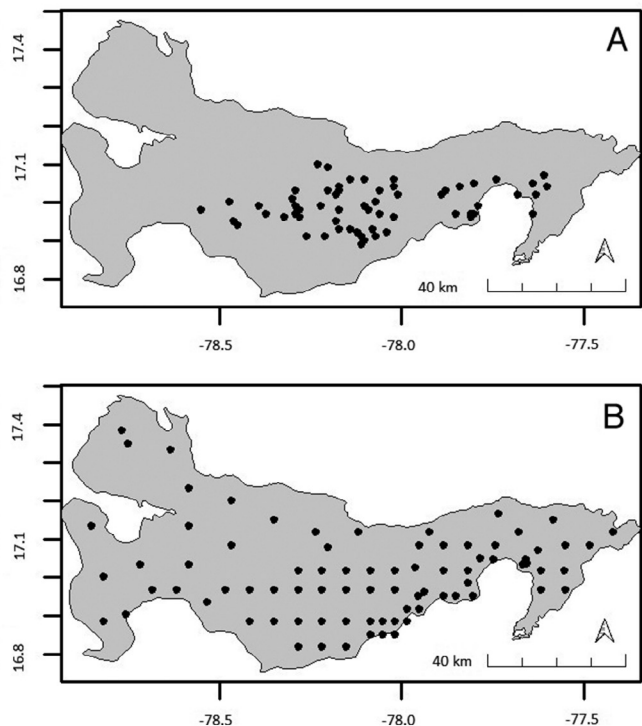


Fig. 2. – Sampling designs for the 2007 (A) and the 2015 and 2018 surveys (B).

habitat type and numbers of adult and juvenile conch. Juveniles were defined as individuals of shell length lower than 20 cm and without a folded shell aperture or lip, while adults were defined as having a shell length of 20 cm or greater and having a thickened or folded lip (Appeldoorn 1988). In recognition of the different data assumptions of the spatial models, calculated densities were used for the geostatistical models OK and KED, whereas count data were used for the other models.

Environmental layers

We collated broad-scale spatial environmental layers for geographic location (latitude and longitude), habitat type and depth (bathymetry) of the bank to be used as covariates in the interpolation models. Geographic location was selected as a proxy for interactions between queen conch location and ecological processes (Gutierrez et al. 2008). Juveniles and adults aggregate in specific habitats for feeding, reproduction and predator avoidance, thus displaying heterogeneous distribution with respect to habitat (Stoner and Ray-Culp 2000). The habitat layer used was a generalized substrate categorization created at 250 m² resolution and adjusted to 100 m² from remote sensing and field measurements as part of the Pedro Bank Marine Spatial Planning Project (Baldwin 2015). Because the queen conch also displays stratification with depth (Morris 2016), we extracted depth readings down to the 30 m contour, coinciding with the maximum depth of the surveys, from the General Bathymetric Chart of the Oceans Compilation Group grid bathymetric database (GEBCO 2020).

Interpolation surface

We created a 100 m² gridded interpolation surface down to the 30 m depth contour consisting of 606000 pixels for a total area of 6060 km². Because the sampling area for 2007 was smaller, the grid was adjusted to 204000 pixels or approximately 2040 km² to avoid extrapolation for that dataset. Each environmental layer was converted to the same 100 m² resolution so that the interpolation outputs were comparable. For the interpolation process, each pixel was considered a discrete unit, so the interpolated value in each pixel was multiplied by the ratio of the total area of one pixel (1 km²) and the mean site area for each survey year to compensate for the difference between the unit used for the interpolation surface (pixel) and the area of the sampled sites (1 m²). Total biomass within a pixel was calculated by transforming the estimated abundances using queen conch weight conversion factors corresponding to the 50% clean processing level (shell, viscera and opercula removed) for adults (Aspra et al. 2009) and 36.69 g per individual for juveniles (Appeldoorn and Rodriguez 1994).

Design method biomass estimation

We calculated two design-based (Cochran 1977) biomass estimations for comparison with the spatial in-

terpolation methods: the first was a non-stratified method and the second one was stratified. The non-stratified method consisted of an estimate based on the product of the mean abundance and the total area and the weight conversion factors for adults and juveniles. Design method estimations can be significantly improved by stratification (Chang et al. 2017), so we introduced stratification based on depth (0–10, 11–20 and 21–30 m). The second design biomass was then calculated similar to the first method but for each stratum and then summed to obtain total biomass.

Geostatistical methods

Ordinary kriging and KED were applied to the datasets using functions and procedures in the R *gstat* package (Pebesma 2004). Geostatistical models use the variogram to estimate the spatial autocorrelation and covariance structures of the sample data, which are then used to determine kriging weights (Webster and Oliver 2007).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Z(x_i) - Z(x_i + h))^2$$

where $\gamma(h)$ is the semivariance at the distance interval h , $Z(x_i)$, is the number of conches at sample site x_i and $N(h)$ is the number of observation pairs separated by distance h .

Ordinary kriging (OK) assumes a stationary local mean around each sample and uses an unbiased linear combination of the surrounding weighted values produced by the variogram to estimate values at unsampled sites (Webster and Oliver 2007).

$$Z^{OK}(x_o) = \sum_{i=1}^n \lambda_{iOK} Z(x_i)$$

where Z^{OK} is the estimated density of conches at the unsampled point of interest, x_o , Z , is the observed value at site x_i , λ_{iOK} , is the weight and n is the value of samples within the search neighbourhood used for the estimation. We defined the local neighbourhood by designating a maximum distance over which semivariance values are calculated and the minimum and maximum number of observation points to be used in the kriging prediction. Defining a local neighbourhood in this way reduces bias from outliers and extreme values of highly skewed data (Surette et al. 2007).

KED assumes that the local trend or drift is a linear function of a secondary variable (Rivoirard et al. 2000), which in this case was the depth covariate.

$$E[Z(x)] = m(x) = a f(x) + b$$

where $f(x)$ is the drift function of the depth variable and a and b are linear coefficients. The KED estimation $Z^{KED}(x_o)$ is achieved by imposing two conditions: one where $\sum_{i=1}^n \lambda_{iKED} = 1$ as with OK, and one where $\sum_{i=1}^n \lambda_{iKED} f(x_i) = f(x_o)$ is the secondary variable at each sample site x_i and $f(x_o)$ is the globally known secondary variable at points x_o .

$$Z^{KED}(x_o) = \sum_{i=1}^n \lambda_{iKED} Z(x_i)$$

Generalized additive model and generalized additive mixed model

A spatial version of GAM (Hastie and Tibshirani 1990) assuming a negative binomial probability distribution (NBGM) was applied using the R `mgcv` package (Wood 2011), while the mixed model GAM or GAMM was applied using the R `nlme` package (Pinheiro et al. 2021) with procedures from Zuur et al. (2009).

$$Y_i = a + s(x_{i1}) + s(x_{i2}) + \dots + s(x_{ip}) + \varepsilon_i$$

where Y_i is the number of conches at site i , a is the intercept parameter, $s(x_{i1...p})$ are smoothing functions of the covariates and ε_i is the error term for the unexplained information. The starting model for each dataset was specified to include all available covariates (geographic location, depth and habitat). The best set of covariates was selected through a backward selection process using the Akaike information criterion (AIC).

$$\text{abundance}_i = \alpha + s(\text{depth}_i) + s(\text{latitude}_i) + s(\text{longitude}_i) + \text{factor}(\text{habitat}_i) + \varepsilon_i$$

The GAMMs included optimal variance and correlation structures to model the random error term ε_i , heterogeneity of the variance and residual correlation in the data. Optimal variance and correlation structures were selected for the GAMM by minimizing the AIC among competing structures (Zuur et al. 2009).

Regression kriging

Two hybrid regression kriging (RK) models (Hengl et al. 2007), GAM+OK and GAMM+OK, were used. The RK prediction Z^{RK} at the target location x_o was reached by summing the trend estimated by the regression procedure $m(x_o)$ and the kriged residuals using OK

$$Z^{RK}(x_o) = m(x_o) + \varepsilon(x_o)$$

Negative binomial zero-inflated model

The spatial negative binomial zero-inflated model (ZINB) directly models the two sources of excessive zeros which can cause bias in spatially sampled data: sampling error and ecological processes (Martin et al. 2005). The procedure included modelling the non-zero part of the data assumed to be negative binomially distributed and the probability processes that generate zeros. The selection of covariates for both processes was done using backward selection among the set of explanatory variables (geographic location, depth and habitat) by minimizing the AIC. The analysis was

implemented using the `zeroinfl` function of the R `pscl` package (Zeileis et al. 2008).

Model evaluation

Statistical evaluation and comparison of spatial model performance were done using ten-fold cross-validation. Two diagnostic statistics were calculated, the mean absolute error (MAE), a measure of prediction bias, and the root mean squared error (RMSE), which reflects prediction precision or prediction error (Willmott 1982). A smaller MAE or RMSE value represents a better performing model.

$$MAE = \frac{1}{n} \sum_{i=1}^n |p_i - o_i|$$

where n is the number of samples, p_i is the predicted value at a location and o_i is the observed value.

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

RESULTS

Structural data trends from the surveys included a high frequency of sites where zeros were recorded relative to sites with positive counts (Fig. 3). The proportion of zeros ranged from 14% to 36% and 24% to 48% for adult and juvenile datasets, respectively (Table 1). The data are also characterized by relatively high coefficients of variation (CoV) and degrees of right-skewness. These statistics are indicative of the patchy distribution, over-dispersion of variances, and zero inflation expected of typical ecological data (Martin et al. 2005). The 2015 sample had the least amount of skewness and CoV, while the 2007 data had the highest values for both.

Comparison of spatial interpolation models

The performance of the spatial models reflected their underlying interpolation assumptions and the structural characteristics of the survey datasets. There was no clear correspondence between bias measured by the MAE and the prediction precision based on the RMSE. In other words, a model returning the lowest MAE did not necessarily return the lowest RMSE (Table 2). Only for the 2015 adult conch estimations did the model with the lowest MAE and RMSE relate to the same model. The best-performing models in 2007 and 2018 were generally those that included GAM or GAMM in full or in part through regression kriging (NBGM, NBGM+OK, GAMM and GAMM+OK). For 2015 the best models also included OK and ZINB. Specifically for the juvenile data, the best models were ZINB (MAE=0.13) and GAMM+OK (RMSE=14.46) in 2007, GAMM (MAE=0.06) and ZINB (RMSE=15.17) in 2015, and ZINB (MAE=0.27)

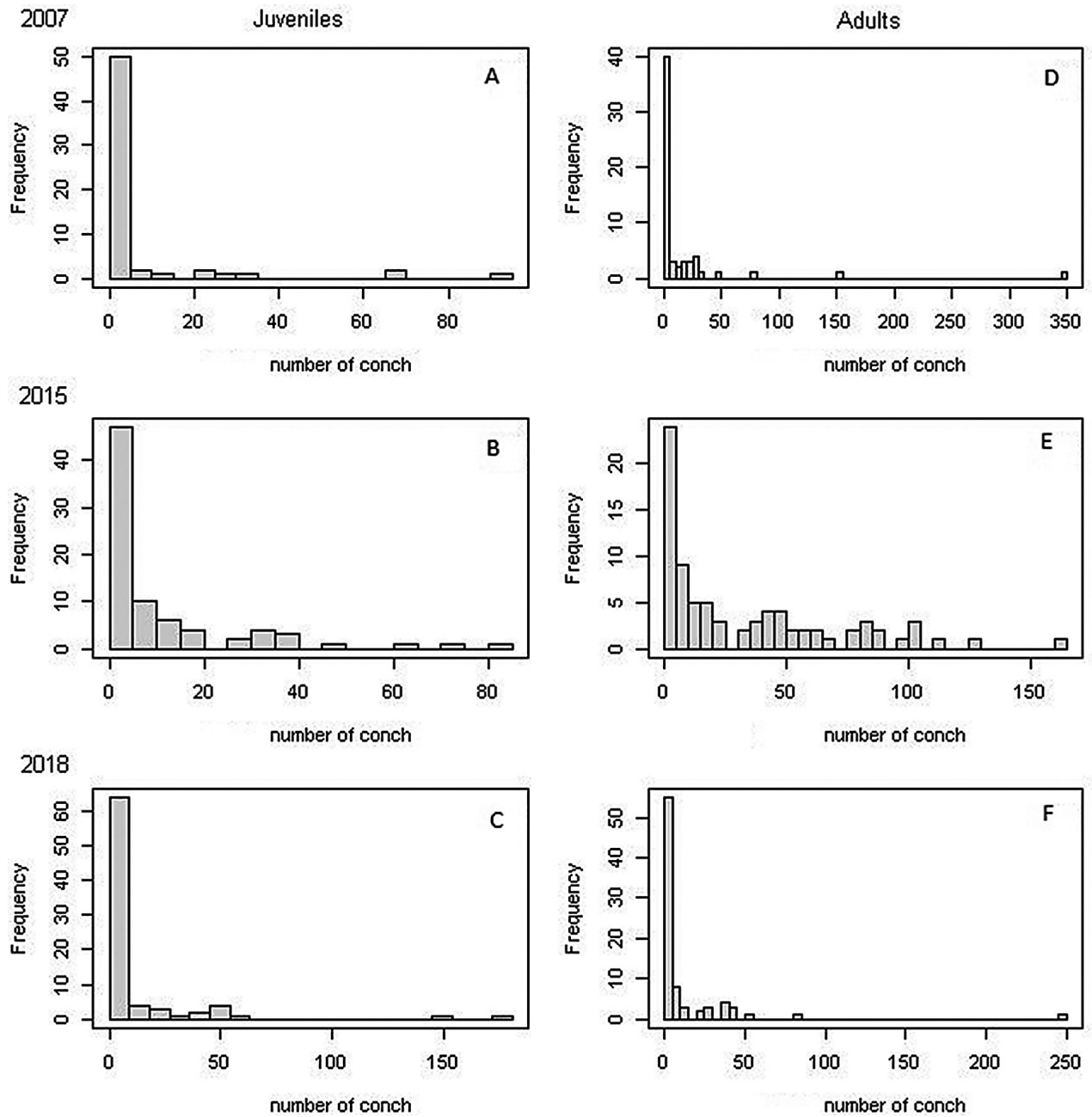


Fig. 3. – Frequency distribution of juvenile and adult queen conch for the 2007, 2015 and 2018 surveys.

Table 1. – Descriptive statistics of queen conch survey data for the years 2007, 2015 and 2018.

	2007		2015		2018	
	Adult	Juvenile	Adult	Juvenile	Adult	Juvenile
Sample size	60	60	80	80	81	81
Minimum	0	0	0	0	0	0
Maximum	350	95	165	85	250	181
Mean	17	7	34	11	12	12
Median	1	1	17	3	2	2
Variance	2458	315	1387	290	958	795
Standard deviation	49.6	17.8	37.3	17.0	31.0	28.2
Coefficient of variation	3.0	2.6	1.1	1.6	2.7	2.4
Kurtosis	32.1	11.7	0.9	5.9	41.0	19.9
Skewness	5.4	3.4	1.2	2.4	5.9	4.2
Percentage of zeros	33	48	14	24	36	36

Table 2. – Juvenile and adult queen conch biomass estimation from alternative spatial interpolation models^a and corresponding ten-fold cross-validation diagnostics^b. Bold: estimated biomass selected for adults and juveniles.

Year	Method	Juvenile			Adult		
		Biomass (t)	MAE	RMSE	Biomass (t)	MAE	RMSE
2007							
	OK	306	0.34	17.70	2171	0.57	56.68
	KED	347	0.31	18.32	2484	2.78	61.38
	NBGAM	63	0.45	15.89	1173	0.22†	20.52
	NBGAM+OK	47	0.92	15.70	1104	0.72	20.40
	GAMM	242	0.25	14.87	1783	0.57	20.00*
	GAMM+OK	320	1.27	14.46*	3535	88.08	208.72
	ZINB	143	0.13†	14.47	2645	5.02	48.77
2015							
	OK	2060	0.58	16.82	16775	0.04†	33.60*
	KED	1856	0.09	16.64	14928	0.20	35.83
	NBGAM	1801	0.61	17.69	18184	0.81	38.38
	NBGAM+OK	1783	0.16	17.64	17135	0.44	38.37
	GAMM	2536	0.06†	16.94	23515	0.06	38.11
	GAMM+OK	2173	1.06	16.04	18751	0.19	38.12
	ZINB	1447	0.29	15.17*	19811	0.25	36.40
2018							
	OK	2194	0.61	27.00	4258	0.47	33.38
	KED	1857	0.52	29.53	4158	0.37	34.24
	NBGAM	1416	1.47	23.98	5606	0.15†	27.52
	NBGAM+OK	1277	1.90	23.96	5372	0.31	27.47
	GAMM	1848	1.56	23.50*	5584	1.82	22.59*
	GAMM+OK	1693	4.54	31.76	7615	2.36	23.87
	ZINB	1883	0.27†	25.53	7394	0.17	23.00

^a Interpolation model: OK = ordinary kriging; KED = kriging with external drift; NBGAM = negative binomial generalized additive model; NBGAM+OK = NBGAM plus OK; GAMM = general additive mixed model; GAMM+OK = GAMM plus OK; ZINB = zero-inflated negative binomial model.

^b Model evaluation and comparative statistics: MAE = mean absolute error; RMSE = root mean square error.

† #Least biased model.

* Most precise model.#

and GAMM (RMSE=23.50) in 2018. For the adult data, the best models were NBGAM (MAE=0.57) and GAMM (RMSE=20.00) in 2007, OK (MAE=0.23, RMSE=33.60) in 2015, and NBGAM (MAE=0.15) and GAMM (RMSE=22.59) in 2018.

Adult biomass estimated from spatial models varied approximately in a range equal to the minimum estimate for each year, with a CoV of 0.40 in 2007, 0.15 in 2015 and 0.24 in 2018. For the juvenile biomass, the CoV was 0.60 in 2007, and 0.18 in 2015 and 2018. This indicates that there were no great differences between the results of the models in terms of biomass estimation.

Comparative analysis of spatial interpolation and design methods

The design-based estimations tended to be close to the biomass estimated by spatial models (Tables 2 and 3), but a major limitation is that the simplicity of

this biomass estimation is not enough to develop ecosystem-based management (Garcia et al. 2003). For adults, the percentage differences between the design methods and spatial interpolation biomass estimations ranged from -5% to -14% and +7% to +43% for the stratified and non-stratified methods, respectively. For juveniles, the differences ranged from +33% to -43% and +6% to +51%, respectively.

Species distribution analysis

The spatial distribution maps of the model predictions (Figs 4 and 5) highlighted the concentration of adult and juvenile conch biomass along the southern edges of the bank, as has been reported by Morris (2016). This area is characteristically shallower and contains a more complex habitat structure, including coral reefs, seagrass, gorgonians and macroalgae, as opposed to predominantly sand and algae in the other areas. However, each model differed in its degree of

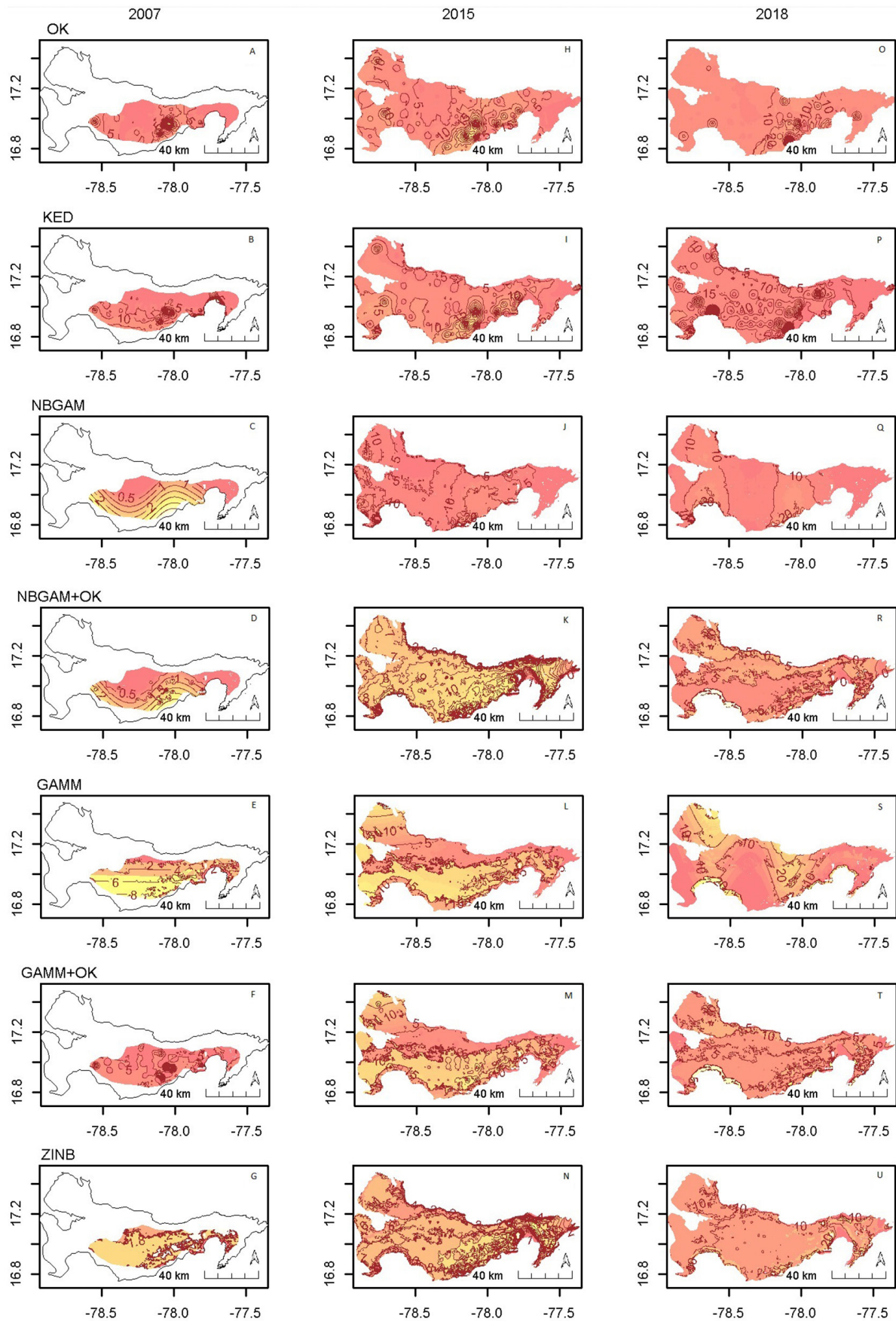


Fig. 4. – Spatial interpolation maps of juvenile conch biomass on the Pedro Bank for the years 2007, 2015 and 2018 produced from seven models: OK, KED, NBGAM, NBGAM+OK, GAMM, GAMM+OK and ZINB.

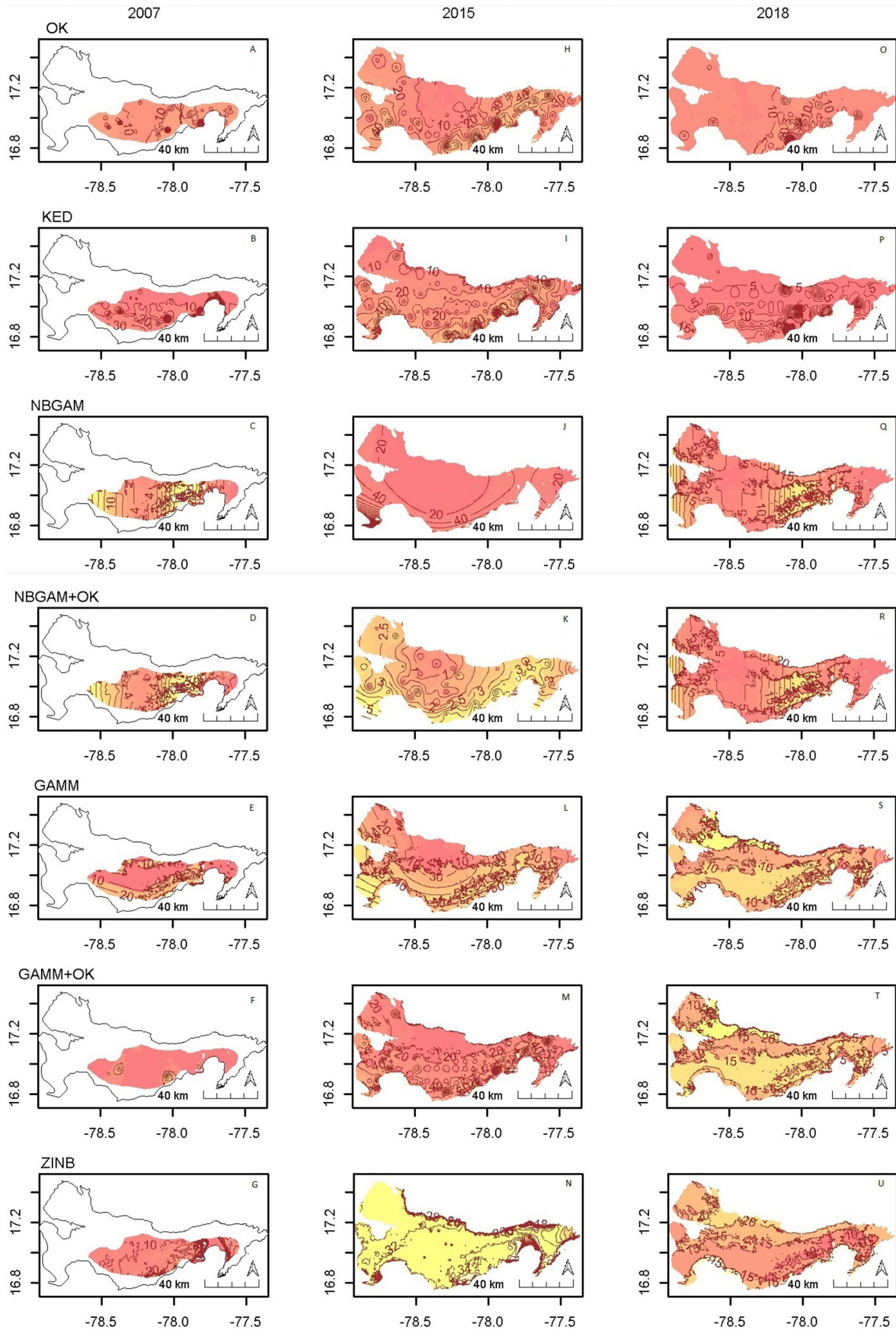


Fig. 5. – Spatial interpolation maps of adult conch biomass on the Pedro Bank for the years 2007, 2015 and 2018 produced from seven models: OK, KED, NBGAM, NBGAM+OK, GAMM, GAMM+OK and ZINB.

local versus global smoothing and identification of the local and broader-scale spatial structures. For instance, the OK, KED and RK models better identified local high-density patches while the NBGM, GAMM, and ZINB models identified larger-scale distributional trends largely reflecting the environmental covariates. High-density patches of juveniles were generally fewer and smaller in terms of abundance in comparison with adults. They were also farther away from the southern edges of the bank than the adults, suggesting a slight distributional difference. The spatial distribution maps therefore validated the performance of the interpolation model by identifying conch aggregations related to the environmental covariates.

DISCUSSION

Given the socio-economic importance of sedentary fisheries and their high vulnerability to overexploitation because of their slow-moving habit, abundance indices must be well estimated and understood in order to conserve and manage stocks. The set of spatial analysis techniques now available has allowed for more reliable estimations of abundance indices and spatial population structure while reducing bias and increasing estimation accuracy to reduce the risk of errors, particularly overestimations. Errors in estimations can be substantial, as our results have shown, if the spatial component of the species distribution is not suitably accounted for. Failing this, there is an increased risk of fishery collapse or a decrease in the efficiency of the management and utilization of the resource as a result of poor decision-making (Anderson and Seijo 2010). We have demonstrated the utility of the multi-model spatial analysis approach to estimate a common index of abundance (biomass) from highly skewed zero-inflated datasets by contrasting selected spatial interpolation techniques and then contrasting these with design methods which do not explicitly consider spatial structuring in the population.

Spatial model performance

Modelling the spatial component of exploited sedentary stocks may be critical to precise biomass esti-

mations, maintaining reproductively viable population structures and avoiding the pitfalls of recruitment overfishing. In this study, the models that assumed a priori probability distribution or used covariates to model the spatial structure (NBGM, NBGM+OK, GAMM, GAMM+OK and ZINB) generally performed better than those that did not, such as OK. This contrast clearly shows the importance of closely considering the spatial component of the abundance index being estimated related to the species biology and exploitation. Queen conch, like many sedentary species, aggregate in specific habitats to carry out biological activities and are subject to high levels of exploitation which constantly alter their population structure (Stoner and Appeldoorn 2021). While the weight of evidence has long favoured model-based approaches over design-based ones, the set of models is often generic in their calibration or is otherwise limited by the availability of covariates (Li and Heap 2011, Chang et al. 2017). In our case, careful knowledge of the species biology, level of exploitation and the study area were required to select and specify the models that were used.

From our results, the design methods returned biomass estimations that were sometimes very close to the best interpolation models (Table 3). Despite this, the presence of unit stock assumptions and failure to account for the population spatial structure renders design methods inadequate for sedentary species (Anderson and Seijo 2010), so they should not be applied to queen conch abundance index estimation. In other cases, the design methods returned estimations with large differences from the best interpolation models. From an ecosystem management standpoint, the risk of estimation errors is particularly problematic for density-dependent sedentary species such as the queen conch, in which allee effects can be enhanced by fishing leading to negative population growth and stock collapse (Stoner and Ray-Culp 2000). Chang et al. (2017) suggested that predictions from design methods may be comparable to those from statistical models if the study area is accurately stratified. Accurate stratification can, however, be difficult or subjective and may contain large bias due to the unmodelled spatial autocorrelation from the spatial structure (Segurado et

Table 3. – Comparison of biomass estimations from the best* spatial interpolation model and from the stratified and non-stratified design methods.

Survey	Spatial* (t)	Biomass estimation method and difference (%)			
		Stratified (t)	% diff.	Non-stratified (t)	% diff.
Adults					
2007	1783	1691	-5	2555	+43
2015	16775	15934	-5	21144	+26
2018	5584	4784	-14	5981	+7
Juveniles					
2007	320	183	-43	339	+6
2015	1447	1919	+33	2182	+51
2018	1848	1911	+3	1980	+7

*model returning the lowest RMSE.

al. 2006). In our case, we stratified according to ocean depth, which generally improved on the non-stratified estimate in terms of the relative difference from the best interpolation model. Other stratification criteria could have also been considered, such as habitat type or age class, but these were either impractical to apply or the data were not available.

In this study, the best models were generally the GAMM and the ZINB based on the two evaluation statistics, MAE and RMSE. It was, however, clear that model performances were strongly influenced by the structure of each dataset, including its sampling design, sample size, coefficient of variation and the quality of the available covariates used. As a result, the best model was not consistent for either juveniles or adults across the three surveys. This is consistent with theory and studies that suggest that spatial interpolation model performance is often data- and case-specific (Isaaks and Srivastava 1989, Rufino et al. 2021) and is reflected in the level of bias and prediction accuracy (Li and Heap 2011). For the 2015 adults, for example, OK was the best model for both statistics but did not feature among the best for either of the two other years. Geostatistical methods, particularly OK, perform relatively poorly with clustered samples and high CoV (Webster and Oliver 2007). The clustered sampling of the 2007 data and the relatively high CoV for both the 2007 and 2015 datasets could explain the model performances in these years.

The inclusion of globally inputted environmental layers as covariates served to improve model performance but also likely added bias to the analysis. The inclusion of depth, habitat, and geographic location layers in the spatial analysis was justified based on the biology and ecology of the species, but these layers were also included because they were easy to obtain. The resolution of these layers, some of which were created for different purposes, did not necessarily reflect the scale at which ecological processes that affect queen conch distribution occur. As a result, undue bias may have been introduced, possibly affecting local autocorrelation and resulting in the false regression significance of some variables (Segurado et al. 2006). The effect of these biases is reflected in the output of the species distribution maps, on which some distributions appear to closely follow the trend of one or more of the covariates, so they poorly captured the local-scale spatial population structures. In addition, the significance of the covariates differed among the various models, as would be expected for models with different assumptions and specifications (Potts and Elith 2006). Therefore, model performance, and by extension the reliability of the estimated abundance index, may have been affected.

In the future, estimations could be further improved by obtaining geospatial layers of variables that reflect the resolution and scale of the biological processes of the species. Incorrect model calibration, unmodelled interactions and missing variables are also important sources of bias in spatial modelling (Segurado et al. 2006, Yu et al. 2013). Given the level of density-dependence and high exploitation of the queen conch (Stoner

and Ray-Culp 2000), covariates explaining the impact of density-dependent mechanisms (such as the allee effect) and fishing on the species distribution may be important missing variables that could have improved the models. The systematic targeting of high-density patches of sedentary species is a key factor in modifying their spatial distribution directly by removing individuals and indirectly by affecting their habitat (Stoner et al. 2018). The inclusion of these two covariates would be prudent for future studies of this species and similar exploited sedentary species using regression-type spatial models.

Fisheries management implications

Despite the improved access and availability of spatial interpolation techniques, there are fisheries such as that of the queen conch in which design methods of abundance indices estimation are widely used. Queen conch is a CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) Appendix II listed species, a conservation status which requires careful management if the fishery is to be sustained (Prada et al. 2017). Improving abundance index estimations and understanding the spatial structuring of juvenile and adult aggregations is therefore an important part of this management process and can be used to develop integrated ecosystem management.

An ecosystem approach to fisheries (Garcia et al. 2003) implies explicit inclusion of the spatial aspect of species management, including small and larger-scale spatial trends caused by biological, ecological and environmental factors. The position and extent of the aggregations realized by the distribution map isolines are indicative of habitat preferences of juveniles and adult spawning stocks that are critical for management (Fig. 6). The changes in size and structure of these aggregations over time are indicative of the distributional response of conch biomass to environmental changes and the level of exploitation. High-density adult aggregations from our models can also serve as a proxy indicator of fisher behaviour since fishers are likely to target these areas. The spatial structures produced from our analysis are therefore a means for prioritizing areas for spatial planning measures such as zonation and targeted research, for example, to maintain viable spawning stock biomass above a reference point. This type of focused spatial management can help to understand critical habitat requirements and maintain population structure and genetic connectivity between local and regional populations (Kitson et al. 2018).

CONCLUSION

The estimation of reliable indices of abundance for sedentary stocks such as queen conch requires that their underlying spatial population structure be accounted to include issues arising from the sampling design, excessive zeros and over-dispersion of variances. The multi-model comparative approach applied here to estimate queen conch biomass was shown to be robust but was dependent on the model specification

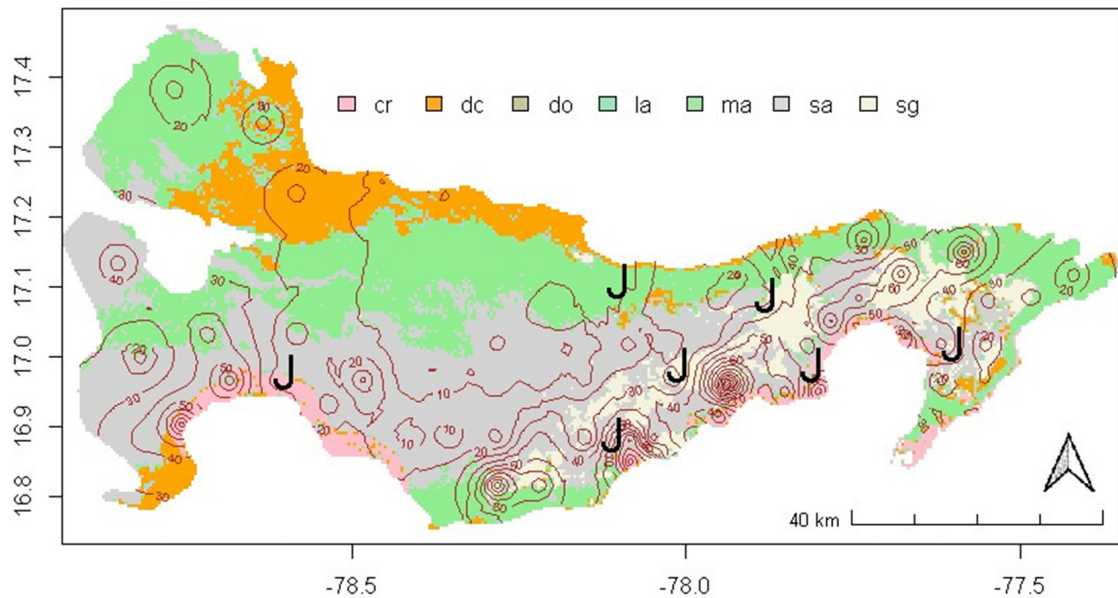


Fig. 6. – Habitat map of the Pedro Bank based on Baldwin (2015) overlaid with predicted adult abundance distribution 2015 (isolines) and the relative location of high-density juvenile patches (“J”). Abundance isolines units are in number of conches per pixel. cr, coral reef; dc, deep coral; do, deep ocean; la, land; ma, macroalgae hardground; sa, sand and sediment; sg, seagrass.

to the specific case and dataset. The traditional use of design biomass estimation should be discontinued for queen conch, as there is the risk of stock collapse owing to poor management decisions based on estimation errors. Our results indicated differences in design method estimation as high as +51% of the biomass estimated by the best spatial interpolation model. Of the seven interpolation models, GAMM and ZINB were among the best-performing models for both juveniles and adults. Aside from the inherent assumptions of the interpolation models, model performance was largely influenced by the sampling design, the sample size, the coefficient of variation of the sample and the quality of the available set of covariates. Model performance could be further improved by including covariates that better represent the scale and resolution of the processes affecting the species distribution, including the effect of fishing. It is therefore clear that to ensure the reliability of abundance index estimation in an ecosystem management context, strong knowledge of the biological, environmental and exploitation-related factors must be understood and incorporated to address specific species and datasets.

ACKNOWLEDGEMENTS

This study was supported by data from the National Fisheries Authority of Jamaica (NFA), the National Environment and Planning Agency (NEPA), and The Nature Conservancy (TNC). Special thanks are due to Stephen Smikle and the survey staff from the National Fisheries Authority of Jamaica. The first author thanks the National Council of Science and Technology, Mexico (CONACYT) for the scholarship at the Universidad Marista de Mérida for a PhD degree. The authors would like to acknowledge the useful feedback and

comments on the manuscript provided by two anonymous reviewers.

REFERENCES

- Anderson L., Seijo J. 2010. Bioeconomics of Fisheries Management. Ames: Wiley-Blackwell. Oxford, UK. 305 pp.
- Appeldoorn R. 1988. Age determination, growth, mortality and age of the first reproduction in adult queen conch, *Strombus gigas* L., off Puerto Rico. Fish. Res. 6: 363-378.
[https://doi.org/10.1016/0165-7836\(88\)90005-7](https://doi.org/10.1016/0165-7836(88)90005-7)
- Appeldoorn R., Rodriguez B. 1994. Queen conch, *Strombus gigas*, biology, fisheries and mariculture. Latinamerican Malacological Congress. Fundacion Cientifica Los Roques, Caracas, 356 pp.
- Aspra B., Barnutty R., Mateo J., et al. 2009. Conversion factors for processed queen conch to nominal weight. FAO Fisheries and Aquaculture Circular No. 1042, Rome, 97 pp.
- Arab A., Wildhaber M., Wikle C., Gent C. 2008. Zero-inflated modeling of fish catch per unit area resulting from multiple gears: Application to channel catfish and shovelnose sturgeon in the Missouri River. North. Am. J. Fish. Manage. 28: 1044-1058.
<https://doi.org/10.1577/M06-250.1>
- Baker N., Appeldoorn R., Torres-Saavedra P. 2016. Fishery-independent surveys of the queen conch stock in Western Puerto Rico, with an assessment of historical trends and management effectiveness. Mar. Coast. Fish. 8: 567-579.
<https://doi.org/10.1080/19425120.2016.1223232>
- Baldwin K. 2015. Marine spatial planning for the Pedro Bank, Jamaica. Final Report. For the Nature Conservancy and NEPA, Government of Jamaica.
- Chang J., Shank B., Hart D. 2017. A comparison of methods to estimate abundance and biomass from belt transect surveys: Population estimation from belt transect surveys. Limnol. Oceanogr. 15: 480-494.
<https://doi.org/10.1002/lom3.10174>
- Drexler M., Ainsworth C. 2013. Generalized Additive Models Used to Predict Species Abundance in the Gulf of Mexico: An Ecosystem Modeling Tool. PLoS ONE, 8: 5.
<https://doi.org/10.1371/journal.pone.0064458>
- Ehrhardt N., Valle-Esquivel M. 2008. Conch (*Strombus gigas*) stock assessment manual. San Juan (PR): Caribbean Fisheries Management Council. 128 pp.
- Garcia S., Zerbi A., Aliaume C., et al. 2003. The ecosystem approach to fisheries. Issues, terminology, principles, institutional founda-

- tions, implementation and outlook. FAO Fisheries Technical Paper No. 443, Rome, 71 pp.
- Gridded Global Bathymetry Data (GEBCO). 2020. British Oceanographic Data Centre, Liverpool, United Kingdom. <https://doi.org/10.5285/a29c5465-b138-234d-e053-6c86abc040b9>
- Gutierrez N., Martinez, A. Defeo, O. 2008. Identifying environmental constraints at the edge of a species' range: Scallop *Psychrochlamys patagonica* in the SW Atlantic Ocean. *Mar. Ecol. Prog. Ser.* 353: 147-156. <https://doi.org/10.3354/meps07184>
- Hall D. 2000. Zero-inflated Poisson binomial regression with random effects: a case study. *Biometrics*, 56: 1030-1039. <https://doi.org/10.1111/j.0006-341X.2000.01030.x>
- Hastie T., Tibshirani R. 1990. *Generalized Additive Models*. Chapman and Hall, Washington D.C., 352 pp.
- Hengl T., Heuvelink G., Rossiter D. 2007. About regression-kriging: From equations to case studies. *Comput. Geosci.* 33: 1301-1315. <https://doi.org/10.1016/j.cageo.2007.05.001>
- Isaaks E., Srivastava, R. 1989. *An introduction to applied geostatistics*. Oxford University Press, New York, 592 pp.
- Kitson-Walters K., Candy A., Truelove N., Roye M., Webber M., Aiken K., Box, S. 2018. Fine-scale population structure of *Lobatus gigas* in Jamaica's exclusive economic zone considering hydrodynamic influences. *Fish. Res.* 199: 53-62. <https://doi.org/10.1016/j.fishres.2017.11.010>
- Krige D. 1951. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, J. South. Afr. Inst. Min. Metall.* 52: 119-139.
- Lambert D. 1992. Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics* 34: 1-14. <https://doi.org/10.2307/1269547>
- Li J., Heap A. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecol. Inform.* 6: 228-241. <https://doi.org/10.1016/j.ecoinf.2010.12.003>
- Lyashevskaya O., Brus D., van der Meer L. 2016. Mapping species abundance by a spatial zero-inflated Poisson model: a case study in the Wadden Sea, the Netherlands. *Ecol. Evol.* 6: 532-543. <https://doi.org/10.1002/ece3.1880>
- Martin T., Wintle B., Rhodes J., et al. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecol. Lett.* 8: 1235-1246. <https://doi.org/10.1111/j.1461-0248.2005.00826.x>
- Matheron G. 1963. Principles of geostatistics. *Econ. Geol.* 58: 1246-1266. <https://doi.org/10.2113/gsecongeo.58.8.1246>
- Morris R. 2016. Distribution of Queen conch (*Strombus gigas*) on the Pedro Bank, Jamaica: descriptive and predictive distribution models. MS thesis, University of Iceland, 67 pp.
- Pebesma E. 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30: 683-691. <https://doi.org/10.1016/j.cageo.2004.03.012>
- Pinheiro J., Bates D., DebRoy S., et al. 2021. nlme: Linear and Non-linear Mixed Effects Models. R package version 3.1-153, 3.1-153.
- Potts J., Elith J. 2006. Comparing species abundance models. *Ecol. Model.* 199: 153-163. <https://doi.org/10.1016/j.ecolmodel.2006.05.025>
- Potts S., Rose K. 2018. Evaluation of GLM and GAM for estimating population indices from fishery independent surveys. *Fish. Res.* 208: 167-178. <https://doi.org/10.1016/j.fishres.2018.07.016>
- Prada M., Appeldoorn R., Van Eijs S., Pérez M. 2017. *Conch Fisheries Management and Conservation Plan*. FAO Fisheries and Aquaculture Technical Paper T610, Rome, 72 pp.
- R Core Team. 2021. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rivoirard J., Simmonds J., Foote K., Fernandes P., Bez N. 2000. *Geostatistics for Estimating Fish Abundance*. Blackwell Science Ltd., Oxford, 216 pp. <https://doi.org/10.1002/9780470757123>
- Rufino M., Albouy C., Brind'Amour A. 2021. Which spatial interpolators I should use? A case study applying to marine species. *Ecol. Model.* 449: 109501. <https://doi.org/10.1016/j.ecolmodel.2021.109501>
- Segurado P., Araujo M., Kunin W. 2006. Consequences of spatial autocorrelation for niche-based models. *J. Appl. Ecol.* 43: 433-444. <https://doi.org/10.1111/j.1365-2664.2006.01162.x>
- Stoner A., Appeldoorn R. 2021. Synthesis of Research on the Reproductive Biology of Queen Conch (*Aliger gigas*): Toward the Goals of Sustainable Fisheries and Species Conservation. *Rev. Fish. Sci. Aquac.* 1-45. <https://doi.org/10.1080/23308249.2021.1968789>
- Stoner A., Ray-Culp M. 2000. Evidence for Allee effects in an over-harvested marine gastropod: density-dependent mating and egg production. *Mar. Ecol. Prog. Ser.* 202: 297-302. <https://doi.org/10.3354/meps202297>
- Stoner A., Davis M., Kough A. 2018. Relationships between fishing pressure and stock structure in queen conch (*Lobatus gigas*) populations: synthesis of long-term surveys and evidence for overfishing in The Bahamas. *Rev. Fish. Sci. Aquac.* 26: 51-71. <https://doi.org/10.1080/23308249.2018.1480008>
- Surette T., Marcotte D., Wade E. 2007. Predicting snow crab (*Chionoecetes opilio*) abundance using kriging with external drift with depth as a covariate. *Can Tech Rep Fish Aquat Sci.* 2763: 1488-5379
- Webster R., Oliver M. 2007. *Geostatistics for Environmental Scientists*, 2nd Edn. John Wiley and Sons, Ltd. Chichester, 336 pp. <https://doi.org/10.1002/9780470517277>
- Willmott C. 1982. Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc.* 63: 1309-1313. [https://doi.org/10.1175/1520-0477\(1982\)063<1309:SCOTE0>2.0.CO;2](https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTE0>2.0.CO;2)
- Wood S. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc. Series. B. Stat. Methodol.* 73: 3-36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>
- Yu H., Jiao Y., Carstensen L. 2013. Performance comparison between spatial interpolation and GLM/GAM in estimating relative abundance indices through a simulation study. *Fish. Res.* 147: 186-195. <https://doi.org/10.1016/j.fishres.2013.06.002>
- Zeileis A., Kleiber C., Jackman S. 2008. Regression Models for Count Data in R. *J. Stat. Softw.* 27: 1-25. <https://doi.org/10.18637/jss.v027.i08>
- Zuur A., Ieno E., Walker N., Saveliev A., Smith, G. 2009. *Mixed Effects Models and Extensions in Ecology With R*. Springer Science+Business Media LLC., New York, 574 pp. <https://doi.org/10.1007/978-0-387-87458-6>