Reproductive ecology studies for sustainable fisheries management in Iberoamerica.
M. Bezerra Figueiredo, R. Domínguez-Petit, G. Macchi, R. Lessa, M.A. Perera, G. Plaza and S. Garrido (Eds)

Scientia Marina 89 (4)
December 2025, e117, Barcelona (Spain)
ISSN-L: 0214-8358
https://doi.org/10.3989/scimar.05636.117

Classification models based on the gonadosomatic index to determine gonadal maturity stages: a case study in the Peruvian anchovy *Engraulis ringens*

Elmer Quispe-Salazar [1], Javier Sánchez [2], Ángel Perea [3]

[1] Área Funcional de Dinámica Poblacional y Evaluación de Recursos Pelágicos, Dirección General de Investigaciones de Recursos Pelágicos, Instituto del Mar del Perú, Callao, Perú.
(EQ-S) (Corresponding author) E-mail: oquispe@imarpe.gob.pe. ORCID iD: https://orcid.org/0000-0001-9229-6379
[2] Laboratorio de Biología Reproductiva, Dirección General de Investigaciones de Recursos Pelágicos, Instituto del Mar del Perú, Callao, Perú.
(JS) E-mail: jsanchez@imarpe.gob.pe. ORCID iD: https://orcid.org/0000-0003-3207-9406
[3] Laboratorio Costero de Chimbote, Instituto del Mar del Perú, Chimbote, Perú.
(APM) E-mail: aperea@imarpe.gob.pe. ORCID iD: https://orcid.org/0000-0001-6615-8802

Summary: Accurate classification of gonadal maturity stages is essential to ensure reliable estimates of reproductive indicators. In Peruvian anchovy (*Engraulis ringens*), two gonadal maturity scales have been used: the empirical scale proposed by Einarsson et al. (1966) and the histologically validated scale developed by Buitrón et al. (2015). The adoption of this scale underscores the need to harmonize or retrospectively adjust historical maturity records, which is essential for analysing long-term changes in anchovy reproductive dynamics. In this study, we evaluate the efficacy of the gonadosomatic index (GSI) in predicting stages and groups of gonadal maturity. Data from 115 research surveys conducted between 1984 and 2022 were analysed. Classification models using different algorithms were applied, with the random forest algorithm demonstrating the highest accuracy and robustness, achieving a mean accuracy of >90% and a logarithmic loss of 0.13. GSI and total length were identified as the most influential predictors of gonadal maturity. The classification of historical data (1984–2014) using the M-5 model (which includes GSI, total length, latitude, distance from the coast and condition factor) demonstrated high concordance with the reproductive patterns observed in recent periods (2015-2022), supporting its predictive reliability. The results indicate that the combined use of GSI and machine learning-based classification models is a valuable tool for identifying gonadal maturity stages from historical data and harmonizing them with the current maturity scale, facilitating the interpretation of variations in the reproductive behaviour of the resource over time.

Keywords: gonadal maturity; gonadosomatic index; Peruvian anchovy; classification models; random forest; machine learning.

Modelos de clasificación basados en el índice gonadosomático para determinar las fases de madurez gonadal: un caso de estudio en la anchoveta peruana *Engraulis ringens*

Resumen: La clasificación precisa de los estadios de madurez gonadal es esencial para calcular con fiabilidad los indicadores reproductivos. En la anchoveta peruana (*Engraulis ringens*), se han utilizado dos escalas de madurez gonadal: la escala empírica propuesta por Einarsson et al. (1966) y la escala validada histológicamente desarrollada por Buitrón et al. (2015). La adopción de esta escala pone de relieve la necesidad de armonizar o ajustar retroactivamente los datos históricos, lo cual es esencial para analizar los cambios en el comportamiento reproductivo de la anchoveta a lo largo del tiempo. En este estudio evaluamos la eficacia del índice gonadosomático (IGS) para predecir estadios y grupos de madurez gonadal. Se analizaron datos de 115 estudios de investigación realizados entre 1984 y 2022. Se aplicaron modelos de clasificación que utilizan diferentes algoritmos, destacando el algoritmo de bosque aleatorio como el más preciso y robusto, con una precisión media >90 % y una pérdida logarítmica de 0.13. El IGS y la longitud total se identificaron como los predictores más influyentes en la clasificación de la madurez gonadal. La clasificación de datos históricos (1984–2014) mediante el modelo M-5 (que incluye IGS, longitud total, latitud, distancia a la costa y factor de condición) demostró una alta concordancia con los patrones reproductivos observados en períodos recientes (2015–2022), respaldando su confiabilidad predictiva. Los resultados indican que el uso combinado de IGS y los modelos de clasificación basados en aprendizaje automático es una herramienta útil para identificar etapas de madurez gonadal a partir de datos históricos y armonizarlas con la escala de madurez actual, facilitando la interpretación de las variaciones en el comportamiento reproductivo del recurso a lo largo del tiempo.

Palabras clave: madurez gonadal; índice gonadosomático; anchoveta peruana; modelos de clasificación; random forest; aprendizaje automático.

Citation/Como citar este artículo: Quispe-Salazar E., Sánchez J., Perea A. 2025. Classification models based on the gonadosomatic index to determine gonadal maturity stages: a case study in the Peruvian anchovy *Engraulis ringens*. In: M. Bezerra Figueiredo et al. (Eds), Reproductive ecology studies for sustainable fisheries management in Iberoamerica. Sci. Mar. 89(4): e117. https://doi.org/10.3989/scimar.05636.117

Editor: R. Domínguez-Petit, M. Perera Garcia, M. Bezerra Figueiredo.

Received: April 11, 2025. **Accepted:** December 01, 2025. **Published:** January 12, 2025.

INTRODUCTION

In commercially harvested fish populations, gonadal maturation is a key aspect because it defines the fraction of the population contributing to reproduction and, consequently, determines recruitment potential and population resilience to fishing pressure and environmental variability (Lowerre-Barbieri et al. 2011, Kjesbu 2016). Thus, the correct classification of gonadal maturity stages (GMS) is essential for the calculation of the species' reproductive indicators, which in turn allows inferences to be made on spawning stock biomass (SSB), size, age at maturity and the species' spawning period (Kell et al. 2016, Domínguez-Petit et al. 2017, Wright and Rowe 2019).

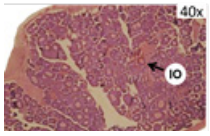
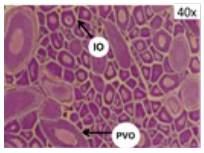
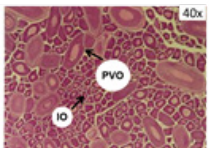
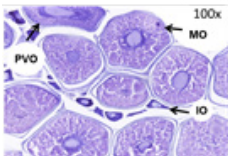
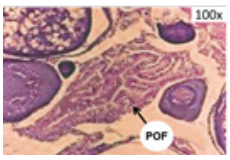
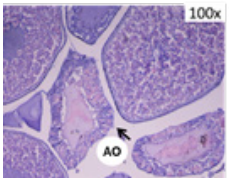
The Peruvian anchovy (*Engraulis ringens* Jenyns, 1842) is a small pelagic species that plays a key role in the food web of the Humboldt Current ecosystem (Taylor et al. 2008, Tam et al. 2010). It also sustains one of the largest single-species fisheries in the world (FAO 2024). Gonadal maturity in this species has been evaluated since the beginning of the fishery through macroscopic observations and over the last three decades through histological analyses of the gonads. The classification of maturity stages in this species has been based on scales that have evolved over time, using the empirical scale of Einarsson et al. (1966) until 2015, designed for a practical and operational classification of gonadal maturity in landings and research surveys (Tsukayama and Alvarez 1980, Santander and Flores 1983, Santander et al. 1984). However, like other macroscopic methodologies, this approach shows a high degree of subjectivity, especially if not validated (Bromley and Casey 2001, Hunter and Macewicz 2003, McPherson et al. 2011), which can lead to systematic errors in the classification of gonadal maturity. Such inaccuracies are particularly critical in the Peruvian anchovy fishery, as they can lead to the overestimation or underestimation of SSB and, consequently, affect population assessments and management strategies, such as reproductive closures (Fréon et al. 2008, Perea et al. 2011).

In this framework, efforts have been made to standardize cataloguing criteria through international workshops focused on the quality of reproductive data used in fish assessments and egg production studies (Kjesbu et al. 2003, Wyanski and Brown-Peterson 2010, Saborido-Rey 2014, ICES 2018). Among the main conclusions of these workshops, incorporating histological analysis was recommended to reduce uncertainty in the classification of GMS (Kjesbu et al. 2003). Following these recommendations, a new macroscopic gonadal maturity scale was implemented for the Peruvian anchovy, based on criteria such as colour, turgidity and vascularization of the gonads. This scale was validated through histological analysis, which characterizes the development of oocytes and spermatocytes, thereby improving precision and standardization into six stages for both sexes (Buitrón et al. 2015). A detailed summary of the characteristics that define each GMS in the Einarsson et al. (1966) and Buitrón et al. (2015) scales is presented in Table 1.

After the implementation of the new maturity scale in 2015 and recognition of the importance of studies based on historical time series for the estimation of maturity and fecundity (Kjesbu et al. 2003, Lowerre-Barbieri et al. 2011), a new challenge emerged: the need to adapt historical data to the current scale to ensure the interpretation of changes in seasonality, reproductive trends of the anchovy, and its response to environmental variability and biological changes (Mori et al. 2011, Cuba et al. 2020, Cuba et al. 2023).

The gonadosomatic index (GSI), also known as the maturity coefficient, represents the ratio of gonad weight to total or eviscerated body weight and has been widely used to describe reproductive behaviour in various species (DeVlaming et al. 1982, West 1990). It is known that the GSI increases during gonadal development (Potoschi et al. 1999). In partial spawners, such as anchovy, it has proved to be a helpful indicator for evaluating variations in reproductive activity throughout the annual cycle (Lowerre-Barbieri et al. 2011). Furthermore, the GSI has been used as an indirect indicator of gonadal maturity in situations where histological analysis is not feasible due to time or resource limitations (Skjæraasen et al. 2012, Flores et al. 2019), and as a proxy to improve the accuracy of macroscopic cataloguing and the interpretation of reproductive indicators (Vitale et al. 2006, McPherson et al. 2011, Flores et al. 2015).

Table 1. Gonadal maturity scales for female Peruvian anchovy (*Engraulis ringens*), describing gonadal maturity stages (GMS) based on macroscopic descriptions (Einarsson et al. 1966) and combined microscopic/macroscopic criteria (Buitrón et al. 2015). Stages in bold indicate gonadal maturity groups (GMG). AO, atretic oocytes; HO, hydrated oocytes; IO, immature oocytes; MO, mature oocytes; POF, postovulatory follicles; PVO, previtellogenic oocytes; VO, vitellogenic oocytes.

Einarsson et al. (1966)		Buitrón et al. (2015)			
Stages	Macroscopic description	Stages	Microscopic/macroscopic description		
I – Virginal immaturity	Thin, pale yellow tubular ovaries without visible oocytes.	**O – Virginal** **GMG: Virgin**	Germ cells, IO, PVO. Thin, transparent, white or pale yellow; no vascularization or visible oocytes.		
II - Early maturing or recovering	Elongated, tubular, yellow-orange ovaries with increased vascularization	I – Resting GMG: Immature	IO, PVO, VO, MO (may or may not be present), remnants of POF, high presence of AO. Flaccid, wine-red colour, with yellow spots from atretic oocytes.		
III - Maturing	Enlarged, cylindrical, bright orange ovaries with visible oocytes.	II-Maturing GMG: Immature	IO, PVO, VO. Medium turgidity, orange-red, slight vascularization; oocytes visible to the naked eye.		
IV - Advanced maturing or hydrated	Large, globular, bright orange ovaries with translucent, easily detached oocytes.	**III – Mature** **GMG: Mature-active**	VO, MO (mature oocytes with completed vitellogenesis). Turgid, pale orange or reddish, strong vascularization, visible oocytes.		
V - Spawning	Reduced ovaries, reddish, with residual mature oocytes.	**IV - Spawning** **GMG: Mature-active**	VO, MO, HO, and/or postovulatory follicles (POF). Large, vascularized, hydrated (crystalline) or blood-tinged flaccid ovaries.		
VI - Spawned	Flaccid, reduced, orange-violet ovaries	V- Recovery GMG: Mature-inactive	HO remnants, POF, AO. Flaccid, vascularized, signs of resorption and atresia.		

This study uses GSI to harmonize the gonadal maturity scale historically employed (Einarsson et al. 1966) with the scale currently used (Buitrón et al. 2015), thereby enabling the retrospective analysis and integration of information on the reproductive condition of the Peruvian anchovy. The objective of this study is to evaluate the accuracy of classifying GMS. To this end, we developed classification models based on GSI and other explanatory variables, intending to classify historical GMS observations.

MATERIALS AND METHODS

Study area and sampling

In Peru, anchovy is divided into two population subunits: the northern-central stock and the southern stock. This study focuses on the northern-central stock, which spans the latitudinal range between 4.0°S and 16.0°S and extends longitudinally up to 150 nautical miles offshore. The data come from biological sampling conducted during pelagic resource research surveys carried out by the Instituto del Mar del Perú (IMARPE), which are performed at a minimum frequency of twice a year.

The samples were collected between 1984 and 2022, encompassing a total of 115 surveys, 4649 hauls, 203024 specimens collected, and 191564 gonads catalogued (Fig. 1A). For this study, the sampling effort is divided into two periods that correspond to the macroscopic maturity scales used to catalogue gonads. The first period comprises 89 surveys conducted between 1984 and 2014 (3709 hauls and 129271 specimens; 64% of the sampling effort; Fig. 1B), in which gonadal maturity was catalogued using the empirical macroscopic scale of Einarsson et al. (1966). The second period includes 26 surveys conducted between 2015 and 2022 (940 hauls and 62293 specimens; 31% of the sampling effort; Fig. 1C), during which gonadal maturity was catalogued using the histologically validated scale of Buitrón et al. (2015). The remaining 5% of the sampling effort corresponds to specimens for which gonadal maturity could not be catalogued or data were unavailable. In this study, data from the second period (2015–2022), catalogued according to the histologically validated scale of Buitrón et al. (2015), were used to train and validate the classification models. Data from the first period (1984–2014), originally catalogued with the empirical macroscopic scale of Einarsson et al. (1966), were then used to classify historical records using the selected model and to assess concordance between observed and model-predicted maturity categories.

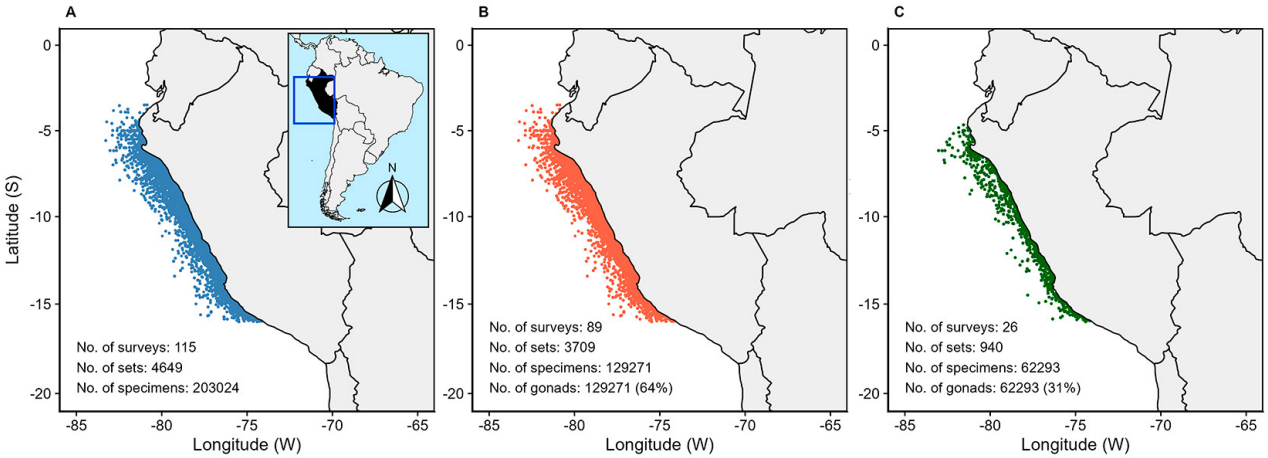


Fig. 1. – Spatial distribution of Peruvian anchovy biological sampling hauls conducted along the Peruvian coast: (A) overall sampling coverage for 1984–2022; (B) hauls in which gonadal maturity was catalogued according to the macroscopic scale of Einarsson et al. (1966); and (C) hauls in which gonadal maturity was catalogued according to the histologically validated scale of Buitrón et al. (2015).

Biological and explanatory variables

The classification of GMS was based on two reference scales. The empirical scale of Einarsson et al. (1966), used until 2015, depended on macroscopic features such as gonad colour, transparency, vascularization and overall size. Although practical for field use, this scale lacked histological validation and thus introduced subjectivity. In 2015, Buitrón et al. (2015) validated a new scale through histological analysis of oocytes and spermatocytes, offering standardized and more accurate criteria for six maturity stages in both sexes. A summary of both scales is shown in Table 1. In this study, the probability of an individual being classified into one of the six GMS categories, as defined by the Buitrón et al. (2015) scale, served as the response variable. This scale was applied as the benchmark for model development and was used to retroactively reclassify historical observations that had originally been recorded with the Einarsson et al. (1966) scale, ensuring temporal comparability across datasets. Covariates related to the biology and condition of the individuals, as well as spatial, temporal and environmental variables, were also considered. In total, 12 explanatory variables of Peruvian anchovy were analysed to classify GMS from 2015 to 2022 (Table 2).

Biological and biometric variables were recorded for each fishing haul during research surveys conducted following standardized protocols (Castillo et al. 2009, Buitrón et al. 2011). Each sample was classified regarding the spawning period (PRD), identifying the main period between August and October (PRD-1) (Jordán 1980, Peña et al. 1989) and a secondary peak from February to March (PRD-2) (Saetersdal and Valdivia 1964). The GSI was calculated as the percentage ratio of gonad weight (GW) to eviscerated body weight (EW), calculated as GSI=(GW/EW)*100. The condition factor (CF), an indicator of physiological condition in fish (Heincke 1908), was derived from total weight (TW), gonad weight, and total length (TL) as $CF=((TW-GW)/TL^3)*100$.

Regional environmental variability associated with El Niño–Southern Oscillation (ENSO) events was assessed based on the Coastal El Niño Index (ICEN) developed by Takahashi et al. (2014). For each biological observation, the following data were recorded: (1) the sampling month, (2) the corresponding ICEN value, and (3) the ENSO phase category, classified as cold phase (La Niña, ICEN ≤–0.7°C), neutral phase (ICEN between –0.7°C and +0.5°C), and warm phase (El Niño, ICEN ≥+0.7°C).

Table 2. – Description of the explanatory variables analysed in the study, including their type and value range for the period 2015-2022.

Category	Variable	Name	Type	Range
Reproductive	GSI	Gonadosomatic index (%)	Numerical (continuous)	0.10–11.80
	GW	Gonad weight (g)	Numerical (continuous)	0.01–4.30
	GL	Gonad length (cm)	Numerical (continuous)	0.01–4.00
Condition	SX	Sex	Categorical (nominal)	male, female
	CF	Condition factor (g/cm^3)	Numerical (continuous)	0.46–0.81
Biometry	TL	Total length (cm)	Numeric (discrete)	7.50–19.00
	TW	Total weight (g)	Numerical (continuous)	1.30–79.80
	EW	Eviscerated weight (g)	Numerical (continuous)	1.00–74.10
Spatial	LAT	Latitude (°S)	Numerical (continuous)	4.00–16.00
	DC	Distance to coast (nm)	Numerical (continuous)	0.42–110.96
Temporal	MNH	Sampling month	Categorical (nominal)	February to November
	PRD	Spawning period	Categorical (nominal)	PRD-0, PRD-1, PRD-2
Environmental	ICEN	ENSO phase	Categorical (ordinal)	El Niño, Neutral, La Niña

Exploratory data analysis

Data processing and validation included the detection and exclusion of outliers using the Z-score technique (Quinn and Keough 2002, Jamshidi et al. 2022). This procedure was applied to the variables GW, EW, TW, GSI, and CF, by TL class and GMS, using a threshold of ±2.7 standard deviations. Subsequently, an exploratory data analysis was conducted to characterize the distribution and patterns of the main explanatory variables in relation to the GMS. The exploratory analysis was conducted using box plots and proportional bar plots to visualize trends, differences and associations across maturity stages. As part of this analysis, a correlation matrix was constructed using Spearman's rank correlation coefficient to evaluate the strength and direction of monotonic relationships among the variables. This analysis allowed for the identification of redundancies and potential multicollinearity issues ($|\rho| \geq 0.75$), which could compromise the performance and interpretability of the classification models (Supplementary Fig. S1). Based on this evaluation, a subset of variables with low intercorrelation was selected, ensuring greater independence among predictors and reducing the risk of overfitting in the models.

Classification models

All GMS were represented in the dataset, but with an asymmetry in their frequencies. To avoid bias in the training of classification models, stratified oversampling was applied using the Synthetic Minority Over-sampling Technique (SMOTE, Chawla et al. 2002), generating synthetic observations for minority stages until 10000 cases were reached (the average frequency per maturity stage). Methodologically, SMOTE interpolates between nearest neighbours, which ensures that the synthetic data remain within the multivariate space of the empirical observations (Lemaître et al. 2017). This procedure prevented information loss, improved stability in model training, and avoided the introduction of artificial patterns, particularly in relation to environmental covariates.

Eight models were adjusted using a forward selection process to identify the most relevant combinations of covariates for the classification of GMS:

$$P(GMS = k \mid x) = f(x)$$

where GMS is the gonadal maturity stage; k denotes one of GMS categories; x is the vector of predictor variables (GSI, TL, ...); and f(x) is a function that estimates the probability of an individual belonging to category k given the covariate vector x.

Six classification algorithms were used to evaluate the models (Breiman et al. 1984, Breiman 2001, Venables and Ripley 2002): linear discriminant analysis, quadratic discriminant analysis, multinomial logistic regression, recursive partitioning decision trees, random forests (RFO), and neural networks with skip connections, through the R-project packages *MASS*, *nnet*, *rpart* and *randomForest* (Liaw and Wiener 2002, Venables and Ripley 2002, Therneau et al. 2013).

Considering that reproductive activity is defined as the number of females with mature, hydrated oocytes and/or post-ovulatory follicles relative to the total number of females (Buitrón et al. 2011), grouping GMS into broader categories was deemed appropriate. Based on this reproductive criterion, GMS were grouped into four gonadal maturity groups (GMGs): virginal (Stage O), immature (Stages I–II), mature-active (stages III–IV), and mature-inactive (stage V) (Table 1) (Flores et al. 2019, Flores et al. 2024). This grouping serves as a functional differentiation of the GMS, allowing sensitivity analysis to evaluate the performance and robustness of the classification models.

Performance metrics

To evaluate predictive performance, the dataset was split into training (80%) and test (20%) sets using stratified random sampling. Stratification ensured that all GMS were proportionally represented in both subsets, thereby avoiding bias and maintaining balanced stage distribution. The performance of each model was evaluated using the confusion matrix from the test dataset. Four performance metrics were considered. First, accuracy (ACC) represents the proportion of correct predictions over the total number of observations. Second, Cohen's kappa coefficient (κ) measures the degree of agreement between predictions and observations after adjusting for chance (Cohen 1960). Third, the F1 score was calculated as the harmonic mean between precision (the proportion of true positives among predicted cases) and recall (the proportion of true positives among actual cases) (Sokolova and Lapalme 2009). Finally, log loss, also known as logarithmic loss, quantifies the discrepancy between predicted probabilities and observed categories, penalizing incorrect predictions made with high confidence (Bishop 2006). These metrics are described as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\kappa = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}$$

$$F1 = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity}; Precision = \frac{TP + FP}{TP}; Sensitivity = \frac{TP + FN}{TP}$$

$$LLS = \frac{-1}{N} \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} log(p_{ij})$$

where *TP* is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives, *FN* is the number of false negatives, *N* is the number of observations, *M* is the number of classes, y_{ij} is 1 if observation *i* belongs to class *j*, and 0 otherwise, and p_{ij} is the predicted probability that observation *i* belongs to class *j*.

For the RFO, we evaluated the importance of the covariates using multi-way importance (Paluszynska et al. 2022). This approach incorporates three related metrics. First, we consider the number of trees in which the predictor served as the root—indicating the first and most informative split. Next, we assess the average depth in the trees where the predictor first splits, as shallower splits are deemed more informative. Finally, we evaluate the number of times the predictor is split across the forest; splits that occur more frequently suggest a greater contribution. These metrics together provided a comprehensive view of variable importance.

Concordance between observed data and model predictions

Historical samples (1984–2014) were reclassified into GMS and GMG using the selected classification model, and concordance was evaluated against observed data from the 2015–2022 period (gonadal maturity catalogued according to the Buitrón et al. 2015 scale) within the reproductive season to ensure biological comparability. Concordance of patterns according to GSI and TL was quantified using Kendall's τ, with statistical significance assessed through two-sided *p*-values obtained by bootstrap resampling (B=2000). Absolute concordance with the 1:1 line was estimated with Lin's concordance correlation coefficient (CCC) and its 95% confidence intervals, while systematic changes and limits of agreement were assessed with Bland–Altman analysis on paired class medians to estimate the mean change (observed–predicted). Finally, the distributions of GSI and TL were compared between periods for each maturity class using the two-sample Kolmogorov-Smirnov test, applying the Holm–Bonferroni adjustment to correct for type I error.

RESULTS

Gonadal maturity distribution and covariate selection

During the period 2015–2022, 62293 gonads were classified, distributed among the gonadal maturation stages O (18.33%), I (6.34%), II (10.80%), III (46.93%), IV (16.72%) and V (0.88%). Based on the exploratory analysis and the correlation matrix of the continuous variables (Supplementary Fig. S1), eight covariates were selected for constructing the classification models. These included five continuous variables—GSI, TL, CF, latitude (LAT) and distance to the coast (DC)—and three categorical variables: sex, spawning period and ENSO category. All covariates were incorporated into the models as explanatory variables (Fig. 2).

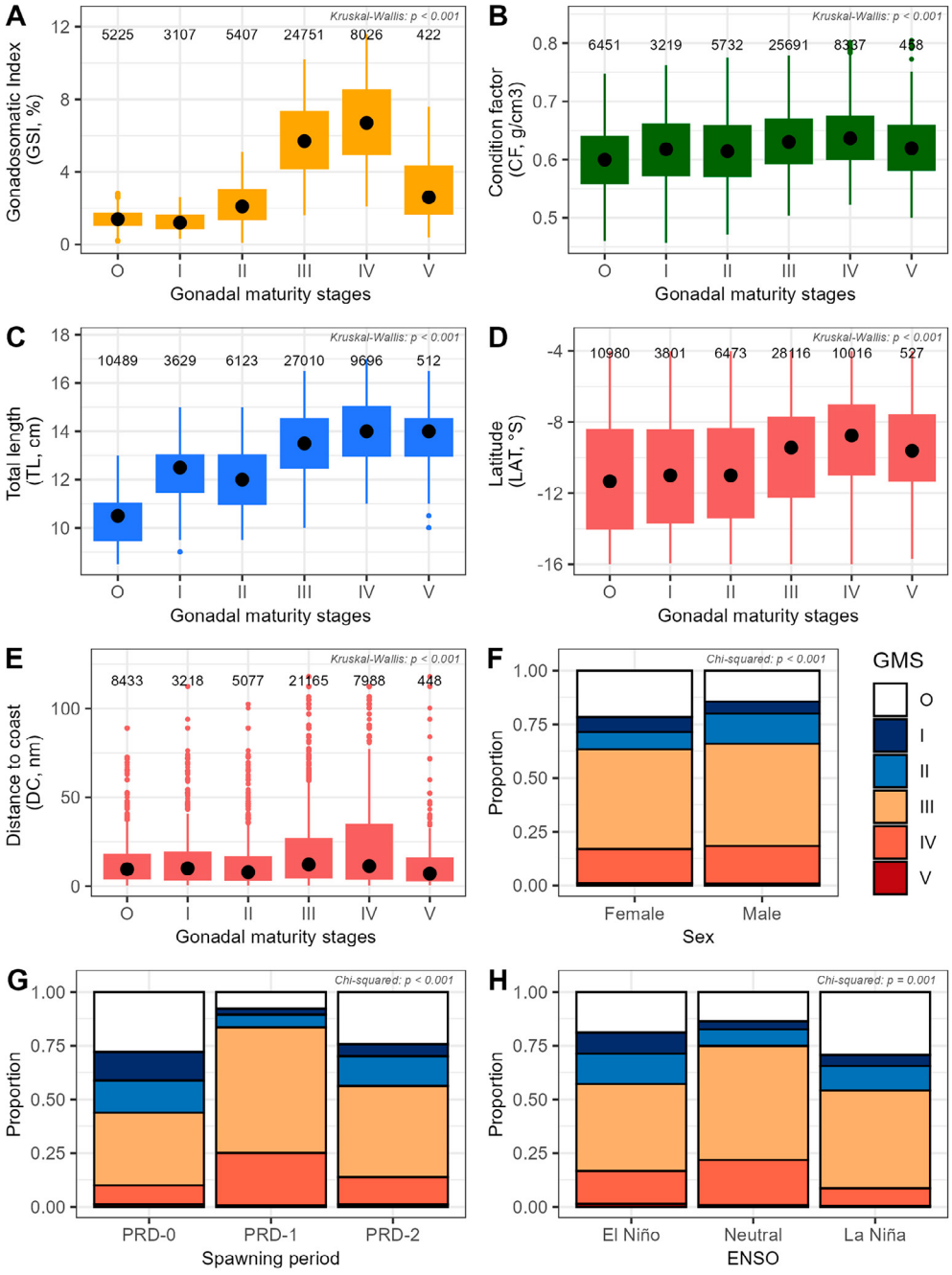


Fig. 2. – Relationship between biological, spatial and environmental variables and gonadal maturity stages in Peruvian anchovy. (A) Gonadosomatic index, (B) condition factor, (C) total length, (D) latitude, (E) distance to coast, (F) sex, (G) spawning period, and (H) ENSO phase (La Niña, Neutral, El Niño). Black points represent the mean. The values above each box indicate the number of individuals per gonadal maturity stage. Statistical tests are indicated at the top of each panel: Kruskal–Wallis test for continuous variables (A–E) and chi-square test for categorical variables (F–H).

Covariate patterns by gonadal maturity stage

GSI showed a progressive increase with the advancement of maturity, reaching the highest values in stages III (μ=5.80) and IV (μ=6.78), followed by a decrease in stage V (μ= 3.07), and these differences among maturity stages were statistically significant according to the Kruskal–Wallis test ($p<0.001$) (Fig. 2A). CF did not exhibit a clear progressive pattern across the stages of gonadal maturity; however, Kruskal–Wallis tests revealed significant differences between several stages ($p<0.001$; Fig. 2B). TL scaled positively with gonadal development. Immature stages (O–II) showed smaller mean sizes (10.46–12.43 cm), whereas advanced maturity stages (III–V) showed larger values (13.30–13.96 cm). Overall, mature individuals were larger than immature ones (Kruskal–Wallis, $p<0.001$; Fig. 2C). LAT showed a northward shift in the more advanced maturity stages, with early stages O–II occupying more southerly latitudes (μ=–11.18 to –10.88), whereas advanced stages III–V occurred farther north (μ=–9.83 to –9.14) (Kruskal–Wallis, $p<0.001$; Fig. 2D). DC increased from immature stages O–II (μ=12.92–13.85) to advanced stages III–IV (μ=19.82–22.15), before decreasing again in stage V (μ=13.58). This indicates that mature individuals (III–IV) were, on average, found farther offshore (Kruskal–Wallis, $p<0.001$; Fig. 2E). However, in general, the best conditions for maturation and spawning were observed in nearshore areas (<50 nm), with only a few cases of mature individuals recorded as far as 110 nm from the coast (Fig. 2E).

The proportion of the mature group (stages III, IV and V) was comparable between females (51.5%) and males (48.5%). However, differences were observed in stages O, I, and II, with a higher frequency of males in stage II (males 60%, females 40%) (chi-square test, $p<0.001$; Fig. 2F). This could be attributed to the faster rate of spermatogenesis compared with oogenesis. During the spawning periods PRD-1 and PRD-2, a higher proportion of mature-active and mature-inactive individuals was observed (III=51.81%, IV=60.95% in PRD-1; III=30.48%, IV=25.91% in PRD-2) than during PRD-0, where these stages were less represented (III=17.70%, IV=13.14%), consistent with a significant association between GMS and the spawning period (chi-square test, $p<0.001$; Fig. 2G). Regarding ENSO conditions, under neutral conditions, a higher frequency of mature specimens was recorded (III=50.24%, IV=55.88%). In contrast, during El Niño events, the proportion was moderate (III=33.46%, IV=35.82%), and markedly lower under La Niña (III=16.30%, IV=8.30%) (Fig. 2H), the distribution of GMS differed significantly among ENSO phases (chi-square test, $p=0.001$; Fig. 2H).

Detailed Kruskal–Wallis statistics and pairwise comparisons for all continuous variables, together with chi-square test results showing significant associations between GMS and the categorical variables are provided in Supplementary Tables S1 and S2.

Model formulation and performance

Classification models were structured as probability functions of GMS and GMG, systematically incorporating covariates based on biological relevance. Model progression followed the following structure: gonadometric (M-1), body metrics (M-2,3), spatial distribution (M-4,5), reproductive biology (M-6,7) and environmental interaction (M-8). The eight models are described below:

M-1: $P(GMS = k) = f(GSI)$
M-2: $P(GMS = k) = f(GSI, TL)$
M-3: $P(GMS = k) = f(GSI, TL, CF)$
M-4: $P(GMS = k) = f(GSI, TL, CF, LAT)$
M-5: $P(GMS = k) = f(GSI, TL, CF, LAT, DC)$
M-6: $P(GMS = k) = f(GSI, TL, CF, LAT, DC, SX)$
M-7: $P(GMS = k) = f(GSI, TL, CF, LAT, DC, SX, PRD)$
M-8: $P(GMS = k) = f(GSI, TL, CF, LAT, DC, SX, PRD, ENSO)$

The classification of the groups and GMS showed a similar pattern in model performance, with a progressive increase as more covariates were included (Fig. 3). In terms of accuracy, RFO stood out for its superior performance, achieving ACC values above 0.90 (Fig. 3A, E). The high agreement between predicted and observed values was also supported by κ values above 0.80 for RFO (Fig. 3C, D). For the F1-score—an indicator of the balance between precision and sensitivity—a similar pattern was observed as in the accuracy and concordance values. RFO achieved the highest values, reaching an F1-score of 0.96 (Fig. 3E, F). The results showed that RFO outperformed the other models and maintained greater stability, with a minimum log loss value of 0.13 (Fig. 3G, H). Together, these findings established RFO as the most robust and consistent algorithm for classifying both GMS and GMG. Sequential model assessments showed that M-5 achieved the greatest marginal improvement in predictive performance. This was particularly evident when DC was included alongside GSI, TL, CF and LAT. For GMS, transitioning from M-4 to M-5 resulted in an increase in accuracy by 3.1%, kappa by 3.7%, and F1 by 3.1%, while also decreasing log loss by 12.9%. For GMG, the gains were 1.5% in accuracy, 2.0% in kappa, 1.5% in F1 score, and a 7.5% decrease in log loss. Beyond this point, additional models (M-6 to M-8) gave negligible improvements, indicating a performance

plateau. Thus, M-5 was selected as the optimal model, balancing predictive power, consistency and parsimony, and aligning with our ecological interpretation (Fig. 3).

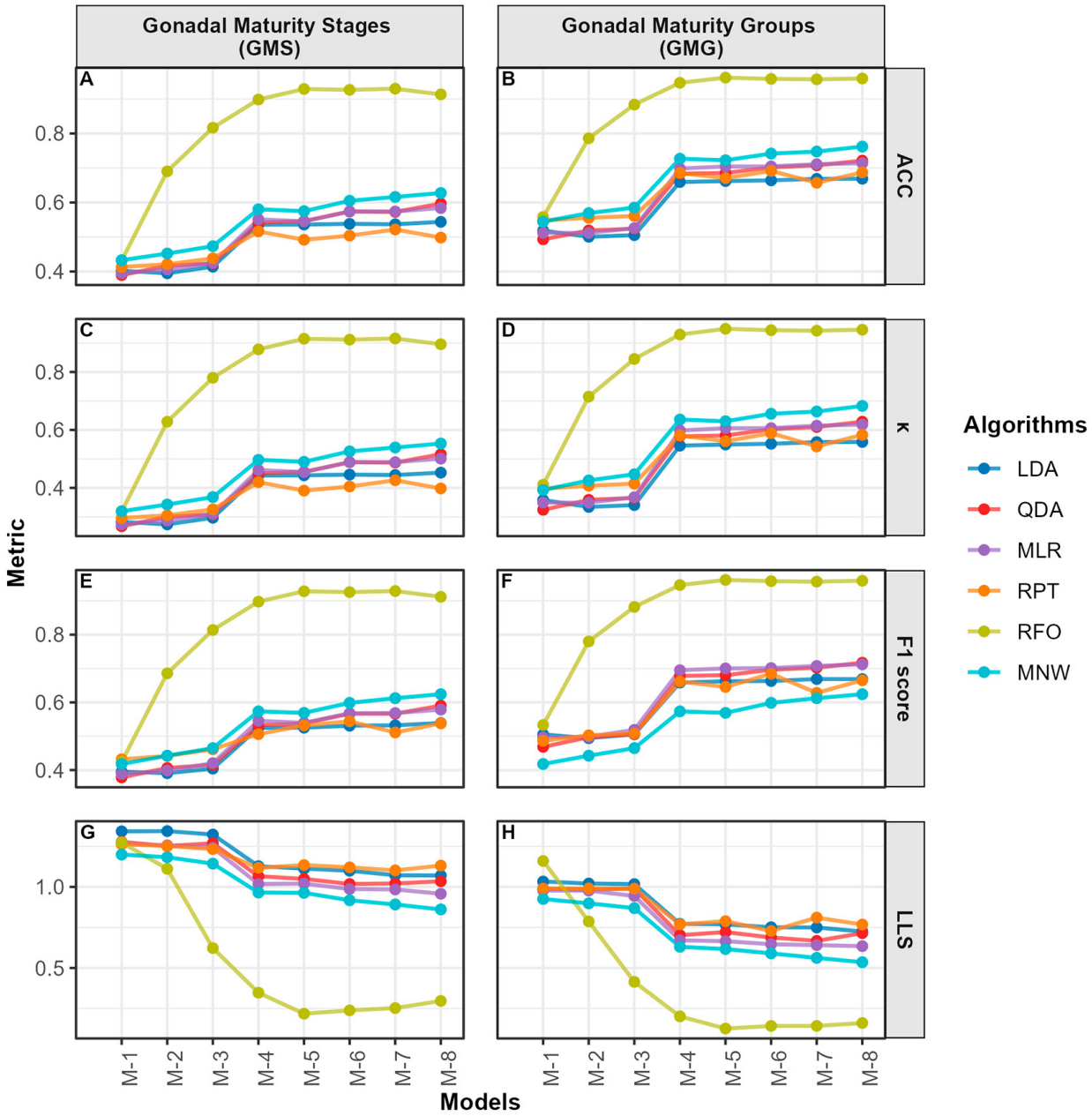


Fig. 3. – Performance and sensitivity of the models for different classification algorithms. Accuracy (ACC), Kappa coefficient (κ), F1 score and log-loss (LLS) are shown in rows A–H for the classification of gonadal maturity stages (left column) and gonadal maturity groups (right column).

Importance of predictor variables

In the assessment of the relative contribution of each covariate in classifying gonadal maturity at M-5, GSI and TL stood out as the most influential predictors, showing high frequencies as root nodes (205 and 140 times), lower average depths (0.84 and 1.14), and a large number of appearances in the trees (>50000) (Fig. 4). Although the predictive contribution of TL was lower than that of GSI, it was consistently selected as a root node and remains biologically relevant because larger individuals are more likely to reach advanced maturity stages. This was especially evident in the classification of GMS stage V, where GSI decreased after peaking at stage IV, while TL continued to

increase. LAT showed secondary importance, with 95 appearances as a root node and a minimum average depth of 1.56, and the remaining covariates contributed to a lesser degree, supporting the discrimination of maturity stages (Fig. 4).

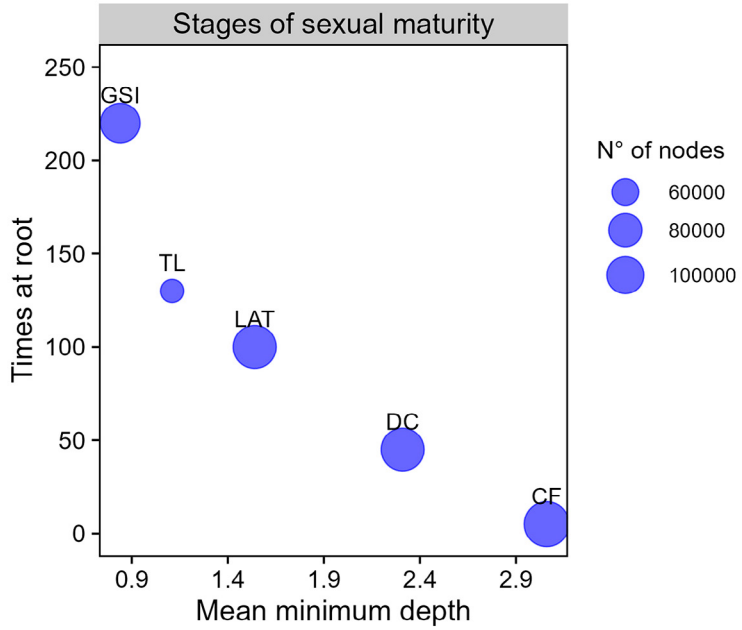


Fig. 4. – Multidirectional importance of the predictor variables in the M-5 model using random forest for the classification of gonadal maturity stages and groups in Peruvian anchovy. Three measures evaluate the importance of each predictor: a) times in a root, which corresponds to the number of trees in which the predictor was used as the root node; b) mean minimum depth, which refers to the depth or level of complexity in the tree structure (a lower value indicates a simpler and more accurate tree); and c) number of nodes, indicating how many times each predictor was included in the forest (larger circles represent greater representation).

Classification performance and stage overlap

In the assessment of the classification within GMG and GMS with the confusion matrices obtained for the M-5 model, these reflected a robust performance in classifying GMS (Fig. 5A). The matrix based on maturity stages showed high accuracy levels in stages O, I, II and V, with correct classification rates higher than 92%. However, stages III and IV showed a greater degree of overlap, particularly with 14.4% of stage III observations classified as stage IV, and 10.2% of stage IV observations classified as stage III. On the other hand, the classification in GMG (Fig. 5B) yielded better results, with correct classification rates exceeding 91.0% for the maturity groups. Overall, the simplification in GMG reduced ambiguity between adjacent classes and supported its predictive reliability of the model.

Concordance between observed data and model predictions

Observed patterns for 2015–2022 (Buitrón et al. 2015 scale) and predicted patterns for 1984–2014 (M-5 random forest classification) showed strong concordance at GMS level for both GSI (τ=0.87, p=0.016; Fig. 6A) and TL (τ=0.867, p=0.017; Fig. 6B). These findings confirm that GSI increases consistently from stage O to IV and that individuals with more advanced gonadal development tend to be larger. Lin's CCC supported these findings (GSI-GMS=0.97 [0.83, 0.99]; TL-GMS=0.81 [0.71, 0.94]), and KS tests indicated that GSI distributions did not differ significantly between periods across GMS (Fig. 6A), whereas TL distributions showed significant differences in most GMS (Fig. 6B). At GMG level, rank concordance was also high (GSI: τ=0.959, p=0.043; TL: τ=0.96, p=0.044; Fig. 6C, D). CCC values confirmed strong agreement (GSI-GMG=0.97 [0.72, 1.00]; TL-GMG=0.85 [0.70, 0.95]). KS tests further indicated that GSI distributions did not differ significantly between periods across GMG (Fig. 6C), whereas TL distributions showed significant differences in most groups (Fig. 6D). Based on this, Bland–Altman analyses revealed no systematic change for GSI at GMS level (change=0.35; Fig. 6E) but a consistent negative change for TL (≈−0.98; Fig. 6F), while at GMG level there was no systematic change for GSI (0.36; Fig. 6G), and TL showed a persistent negative change (≈−0.99; Fig. 6H). Overall, these results demonstrate that the M-5 model is robust in reproducing historical reproductive patterns, particularly the structure of GSI and TL, while also revealing a decline in total length in the more recent period.

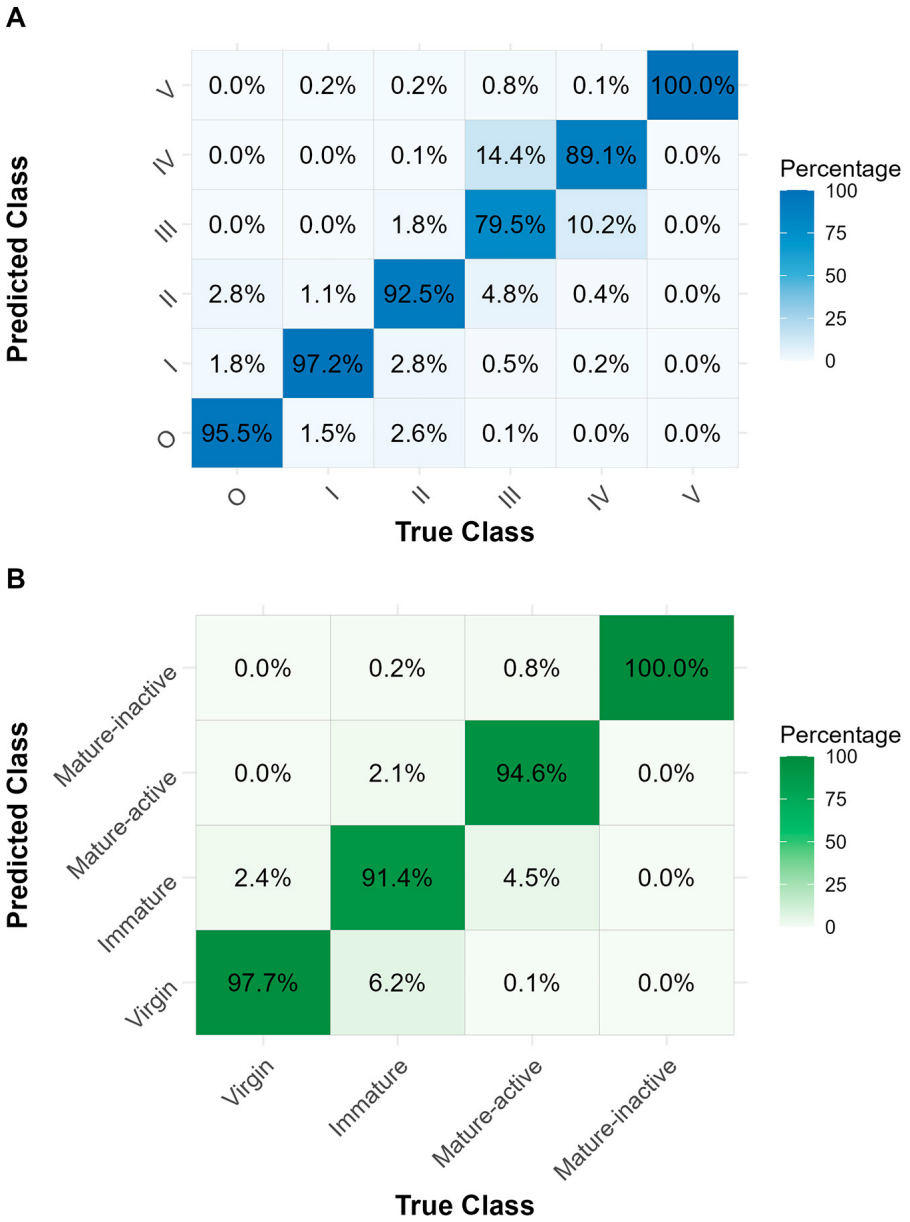


Fig. 5. – Confusion matrices of the classification models using random forests algorithm for (A) gonadal maturity stages (O, I, II, III, IV and V) and (B) gonadal maturity groups (virginal, immature, mature-active and mature-inactive) in Peruvian anchovy. Predictions correspond to the final selected model (M-5).

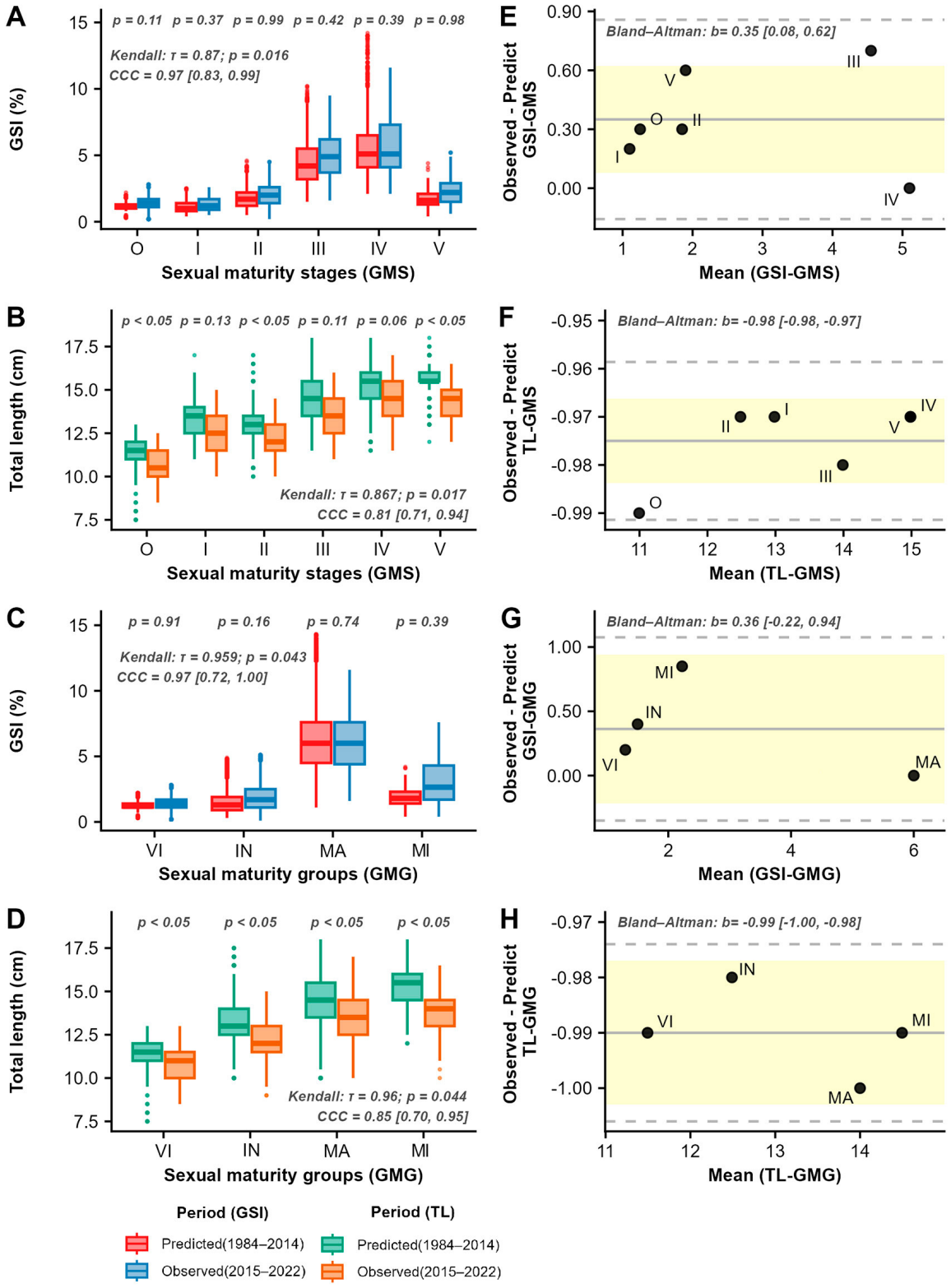


Fig. 6. – Concordance between observed data (2015–2022) and model predictions (1984–2014) of reproductive patterns of Peruvian anchovy. Comparisons of gonadosomatic index (GSI) and total length by gonadal maturity stages (GMS: A, B) and gonadal maturity groups (GMG: C, D). VI, virginal; IN, immature; MA, mature-active; MI, mature-inactive. Observed values correspond to empirical data collected in the period 2015–2022 (Buitrón et al. 2015 scale), while predicted values correspond to classifications for 1984–2014 obtained from the M-5 random forest model. Boxplots (A-D) display medians and interquartile ranges, annotated with Kolmogorov–Smirnov test results for distributional differences. The pattern concordance between periods was quantified with Kendall's τ (± p), absolute agreement with Lin's concordance correlation coefficient (CCC, 95% CI), and systematic change with Bland–Altman analyses (E-H).

DISCUSSION

GSI as a robust indicator of gonadal maturity

The use of GSI has been applied to other species of the genus *Engraulis*, allowing the description of their reproductive behaviour under different space-time scales (Yoneda et al. 2013, Djurovic et al. 2018, Basilone et al. 2020, Rahmani et al. 2020). The choice of GSI as the primary descriptor of reproductive behaviour is supported by several factors. (1) It is independent from body size, allowing comparisons of individuals of different sizes without biases associated with growth (DeVlaming et al. 1982). (2) It meets the isometric condition in the slope of the log-log relationship between gonadal weight and eviscerated weight by maturity stages (Somarakis et al. 2004). (3) It shows homogeneity of the slopes between different maturity stages (Erickson et al. 1985). In the case of anchovy, GSI increased with TL from stages O to III. Still, this trend plateaued at advanced maturity (stage IV), where larger individuals did not show higher GSI values, indicating that GSI values are independent of body size (Fig. 2A, C). Although the assumptions of isometry and homogeneity were not directly evaluated in this study, GSI showed the highest discriminative power among the variables analysed, followed by TL (Fig. 4). This discriminative capacity is further illustrated by the bivariate distribution of TL and GSI by sex (Fig. 7), which shows an increase in GSI with TL from the early stages up to stage III, followed by stabilization at stage IV. This pattern is evident in both sexes and reinforces the interpretation that GSI, together with TL, provides a consistent descriptor of reproductive development across all maturity stages, becoming even more apparent when GMS are aggregated into GMG. Moreover, the combined use of GSI and TL helped reduce subjectivity and improve retrospective classification in the historical series from 1984 to 2014. These findings are consistent with previous studies, which have shown that GSI remains a reliable indicator of gonadal maturity (Flores et al. 2019), particularly when integrated with classification models based on machine learning (Flores et al. 2024).

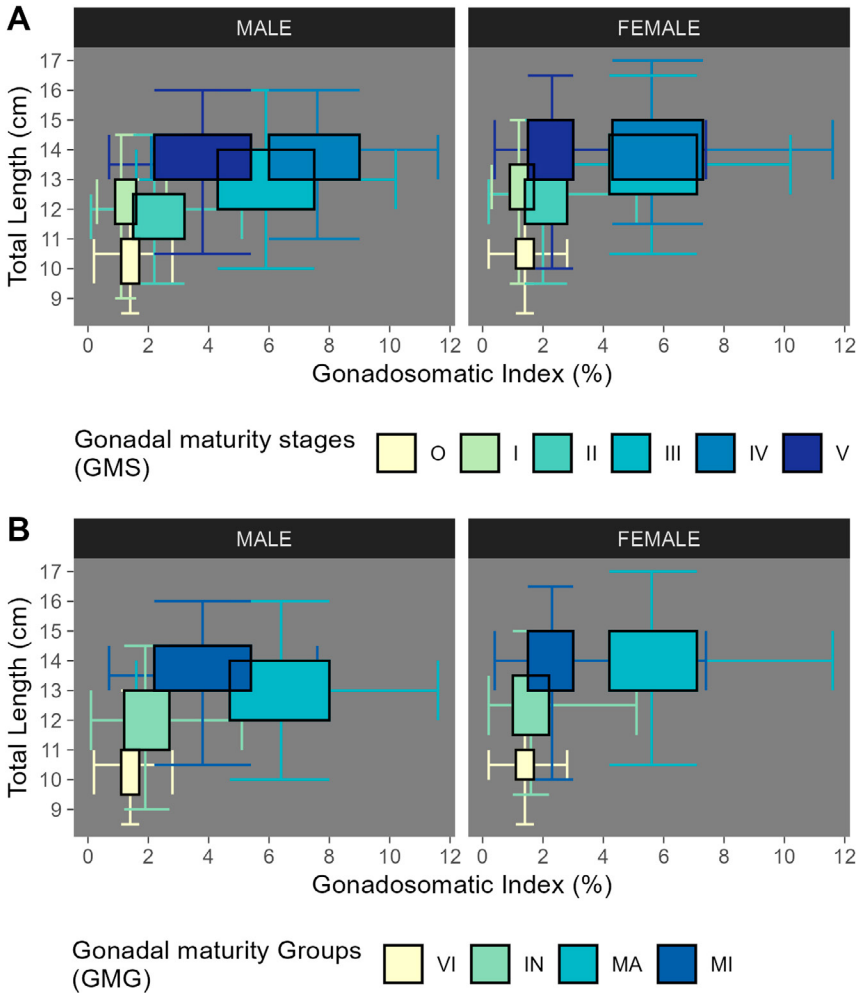


Fig. 7. – Bivariate distribution of total length and gonadosomatic index of Peruvian anchovy by sex, shown according to (A) gonadal maturity stages and (B) gonadal maturity groups. VI, virginal; IN, immature; MA, mature-active; MI, mature-inactive.

Machine learning advances in fish reproductive ecology

Multiple classification algorithms have been used to study life history traits in fish. Traditional models such as linear discriminant analysis and multinomial logistic regression have proven useful in identifying patterns related to maturation (Engelhard et al. 2003), the relationship between age and length (Gerritsen et al. 2006), early-stage development (Ibaibarriaga et al. 2007) and recruitment (Fukuda et al. 2013). However, these techniques have limitations when the relationships between variables are nonlinear or when there is high collinearity among predictors, which reduces their accuracy and generalization ability (Breiman et al. 1984). In contrast, machine learning algorithms such as random forests (RFO) and neural networks have shown improvements in predicting age (Benzer et al. 2022), reproductive behaviour (Brownscombe et al. 2020, Brown-Peterson et al. 2021, Flores et al. 2024) and recruitment (Fernandes et al. 2010, Fernandes et al. 2015, Okamura et al. 2024), proving more effective in situations with complex nonlinear relationships and high data dimensionality (Breiman 2001, Venables and Ripley 2002).

Limitations of RFO in classifying gonadal maturity stages

In this study, the RFO algorithm successfully classified GMS in anchovy, using a histologically validated scale as a reference (Buitrón et al. 2015). However, as with all supervised classification models, certain limitations were identified that should be considered:

1) The class imbalance in the data, with less frequent maturity stages, which can affect the ability to correctly classify these categories. To address this, stratified oversampling was applied (Chawla et al. 2002, Lemaître et al. 2017), improving the representation of minority classes and the model's stability.
2) The quality of the training data and its generalization capacity in the model's performance. In this study, the data collected from research surveys, a scientific platform that ensured data reliability through standardized sampling protocols (Castillo et al. 2009, Buitrón et al. 2011, 2015), were used. Additionally, the Z-score was applied to detect and exclude outliers (Quinn and Keough 2002, Jamshidi et al. 2022).
3) The low interpretability of the model, considered a *black-box model*, makes it difficult to understand the relationships between variables (Breiman 2001). However, it allows for the evaluation of the importance of predictor variables (Paluszynska et al. 2022), facilitating the identification of variables with the highest discriminatory power in the classification of GMS.

Limitations in the macroscopic discrimination of gonadal maturity stages

The lower accuracy of the M-5 model (Fig. 5A) in classifying stages III (mature) and IV (spawning) is attributed to the subjectivity in macroscopic evaluation by observers, based on visual characteristics such as coloration, volume and vascularization of the gonads, which introduces inter-observer variability (Vitale et al. 2006, McPherson et al. 2011), along with the overlap in GSI between these stages. This difficulty is further accentuated by the gradual histological transition between the two stages: in females, stage III is characterized by the coexistence of oocytes in different stages of development, whereas in stage IV, the presence of hydrated oocytes and/or post-ovulatory follicles is typical; in males, stage III shows seminiferous tubules filled with spermatocytes and sperm, whereas in stage IV, seminiferous tubules exhibit empty zones with fewer sperm (Buitrón et al. 2015). These transitions between stages in both females and males complicate macroscopic discrimination in advanced stages of gonadal maturity. However, when stages III and IV were grouped under the functional category of mature-active (developing and reproductively capable), the accuracy model improved, overcoming the macroscopic limitations between stages (Fig. 5B). In this regard, although a validated macroscopic scale exists, it is recommended to complement it with microscopic staging to achieve a more precise cataloguing and finer discrimination between stages (Kjesbu et al. 2003, Wyanski and Brown-Peterson 2010). We consider that the choice between using stages or grouping them into functional categories represents a trade-off between the level of biological resolution and the statistical stability of the model, where stages provide detail, while groups offer greater robustness and simplicity for operational applications

Environmental and biological drivers of reproductive variability

The advantages of the random forest model over macroscopic analysis are evident, as it incorporates multiple variables, including both biological (GSI, TL and CF) and spatial (LAT and DC) factors. Although variables such as sex, reproductive period, and environmental conditions were not selected in the final model (M-5), their influence was indirectly reflected through changes in GSI and TL. This ability to reflect changes in reproductive behaviour in GSI and TL has been documented in several studies, indicating that larger individuals tend to maintain prolonged spawning activity throughout the year (Perea et al. 2011), while smaller individuals mature and spawn for the first time during reproductive periods PRD-1 and PRD-2 (Saetersdal and Valdivia 1964, Jordán 1980, Peña et al. 1989). Moreover, events such as El Niño and La Niña impact GSI (Perea et al. 2015, Cuba et al. 2020, 2023), reflecting

alterations in the reproductive behaviour of anchovies, including shifts in the reproductive cycle, decreased fertility and reduced spawning frequency (Buitrón and Perea 2000). To reduce the variability introduced by the reproductive period and sex, it is recommended that the operational application of the model focus on females and during the main reproductive periods, considering the environmental scenario, as suggested in previous studies (Flores et al. 2015, Flores et al. 2019). Additionally, analysis should prioritize females, as their reproductive indicators, such as GSI, spawning fraction and atresia index, support management decisions for temporary closures during the reproductive cycle of Peruvian anchovy (Perea et al. 2011).

Temporal consistency in GSI-TL patterns

Finally, the strong concordance observed between GSI and TL patterns across both periods supports the robustness of the M-5 model. This finding suggests that the biological relationship between GMS, gonadal development and body growth in anchovy has remained consistent (Fig. 6), despite possible decadal variations (Mori et al. 2011) or changes in gonadal maturity scales (Einarsson et al. 1966, Buitrón et al. 2015). Nevertheless, although correlations were high, it is noteworthy that, in the most recent period, the average TL values in both GMS and GMG were lower than those in 1984–2014. This pattern may reflect changes in population structure or environmental conditions that influence reproductive biology and somatic growth in anchovy, and it should be taken into account in future studies.

CONCLUSIONS

The Peruvian anchovy is one of the most important fisheries globally (FAO 2024) and plays a key role in the trophic structure of the Humboldt Current ecosystem (Taylor et al. 2008, Tam et al. 2010). Therefore, understanding its reproductive biology is crucial for the conservation and sustainable management of the resource. Our results demonstrate that GSI is the primary predictor of maturity and, together with TL, enables a reliable classification of GMS in this species. In this context, classification algorithms—and particularly the random forest applied here—proved to be highly effective. The superiority of random forest was evident not only in its high predictive accuracy but also in its ability to handle the biological, temporal and spatial complexity associated with gonadal maturity.

The application of predictive models for the historical classification of maturity stages represents a significant advancement in standardizing and maintaining the continuity of the time series of reproductive indicators in anchovies. However, it is essential to consider the spawning process, as it depends on still unknown variables that require further study. Despite this limitation, this approach not only facilitates the coherent integration of historical data with the current gonadal maturity scale but also provides opportunities for retrospective analysis of anchovy reproductive behaviour. As a result, a more accurate and reliable interpretation of the reproductive patterns of this species is achieved, along with a deeper understanding of its long-term trends.

Against this background, the implementation of GSI, combined with the random forest algorithm, is as a complementary tool for macroscopic cataloguing in commercial fishing monitoring programmes and research surveys, evaluating its effectiveness through histological analysis, the primary reference for assessing gonadal maturation in fish. This approach not only enhances the ability to integrate multiple variables but also opens the possibility of exploring various machine learning algorithms. It is suggested that its applicability be explored in other pelagic species of interest, with the aim of optimizing the historical classification of GMS, standardizing it to their validated scale, and ultimately enhancing the understanding of the reproductive behaviour of these species.

SUPPLEMENTARY MATERIAL

The following supplementary material is available through the online version of this article:

Fig. S1. – Spearman's correlation matrix between variables for the classification of Peruvian anchovy gonadal maturity stages. The number of observations used for each variable is indicated in parentheses. The variables were ordered using hierarchical clustering. CF, condition factor; DC, distance to coast; EW, eviscerated weight; GL, gonad length; GSI, gonadosomatic index; GW, gonad weight; LAT, latitude; TL, total length; TW, total weight.

Table S1. – Summary of Kruskal–Wallis statistics and Dunn's post hoc pairwise comparisons (with Benjamini–Hochberg adjusted *p*-values) for continuous variables evaluated across different gonadal maturity stages. Only significant pairwise differences are reported. CF, condition factor; DC, distance to coast; GSI, gonadosomatic index; LAT, latitude; TL, total length.

Table S2. – Summary of chi-square statistics and post hoc evaluation using standardized residuals (|residual| >2) for categorical variables associated with gonadal maturity stages.

DATA AVAILABILITY

The biological data used in this study are part of the historical research database of the Instituto del Mar del Perú (IMARPE) and are available upon request with institutional authorization.

ACKNOWLEDGEMENTS

We sincerely thank the anonymous reviewers for their valuable comments and suggestions, which helped improve earlier versions of this manuscript.

This work is dedicated to the entire pelagic resources biology group at the Instituto del Mar del Perú (IMARPE), whose effort and commitment in the collection, recording and cataloguing of data aboard research surveys form a fundamental pillar for the conducting of scientific research and the sustainable management of the main pelagic fisheries.

DECLARATION OF COMPETING INTERESTS

The authors of this article declare that they have no financial, professional or personal conflicts of interest that could have inappropriately influenced this work.

FUNDING SOURCES

This research was funded by institutional resources allocated to the scientific research surveys targeting the main marine resources, conducted by the Instituto del Mar del Perú (IMARPE).

AUTHORSHIP CONTRIBUTION STATEMENT

Elmer Quispe-Salazar: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft, Writing – review and editing. **Javier Sánchez:** Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – review and editing. **Ángel Perea:** Formal analysis, Investigation, Methodology, Project administration, Writing – review and editing.

REFERENCES

Bishop C.M., Nasrabadi N.M. 2006. Pattern Recognition and Machine Learning. Springer, New York, 738 pp.

Basilone G., Ferreri R., Barra M., et al. 2020. Spawning ecology of the European anchovy (*Engraulis encrasicolus*) in the Strait of Sicily: Linking variations of zooplankton prey, fish density, growth, and reproduction in an upwelling system. Prog. Oceanogr. 184: 102330. https://doi.org/10.1016/j.pocean.2020.102330

Breiman L., Friedman J., Olshen R.A., et al. 1984. Classification and Regression Trees. Chapman and Hall/CRC, New York, 368 pp. https://doi.org/10.1201/9781315139470

Breiman L. 2001. Random Forests. Machine Learning 45(1): 5-32. https://doi.org/10.1023/A:1010933404324

Bromley P.J., Casey J. 2001. An attempt to ascertain the spawning fraction of female North Sea whiting based on visual staging of maturity. In: Kjesbu O.S., Hunter J.R., Witthames P.R. Report of the working group on Modern approaches to assess maturity and fecundity of warm- and cold-water fish and squids. Fish. Board Norw., Inst. Mar. Res., Bergen, pp. 25-28.

Brown-Peterson N.J., Leaf R.T., Leontiou A.J. 2021. Importance of depth and artificial structure as predictors of female red snapper reproductive parameters: Red Snapper Reproduction. Trans. Am. Fish. Soc. 150(1): 115-129. https://doi.org/10.1002/tafs.10277

Brownscombe J.W., Griffin L.P., Morley D., et al. 2020. Application of machine learning algorithms to identify cryptic reproductive habitats using diverse information sources. Oecologia 194(1): 283-298. https://doi.org/10.1007/s00442-020-04753-2

Buitrón B., Perea A. 2000. Aspectos reproductivos de la anchoveta peruana durante el periodo 1992–2000. Bol. Inst. Mar Perú 19: 45-53.

Buitrón B., Perea A., Mori J., et al. 2011. Protocolo para estudios sobre el proceso reproductivo de peces pelágicos y demersales. Inf. Inst. Mar Perú 38(4): 373-383. https://hdl.handle.net/20.500.12958/2172

Buitrón B., Perea A., Mori J., et al. 2015. Escala de madurez gonadal de anchoveta peruana *Engraulis ringens* (Jenyns, 1842). Bol. Inst. Mar Perú 30(1-2): 10-19. https://hdl.handle.net/20.500.12958/2936

Castillo R., Peraltilla S., Aliaga A., et al. 2009. Protocolo técnico para la evaluación acústica de las áreas de distribución y abundancia de recursos pelágicos en el mar peruano. Inf. Inst. Mar Perú 36(1-2): 1-23. https://hdl.handle.net/20.500.12958/2001

Chawla N.V., Bowyer K.W., Hal L.O., et al. 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16: 321-357. https://doi.org/10.1613/jair.953

Cohen J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1): 37-46. https://doi.org/10.1177/001316446002000104

Cuba A., Sánchez J., Mori J., et al. 2020. Anomalías de los índices reproductivos, fracción desovante e índice gonadosomático de anchoveta peruana *Engraulis ringens (*Jenyns, 1842) del stock norte-centro del Perú en relación a El Niño Costero 2017. The Biologist 17(2): 253-261. https://doi.org/10.24039/rtb2019172371

Cuba A., Sánchez J., Mori J. 2023. Effect in the reproduction of north-center Peruvian anchovy *Engraulis ringens* (Jenyns, 1842) stock from Peru during La Niña Event 2017-2018 categorized by reproductive index anomalies. The Biologist 21(2): 141-146. https://doi.org/10.24039/rtb20232121606

DeVlaming V., Grossman G., Chapman F. 1982. On the use of the gonosomatic index. Comp. Biochem. Physiol. A Physiol. 73(1): 31-39. https://doi.org/10.1016/0300-9629(82)90088-3

Djurovic M., Joksimović A., Pešić A., et al. 2018. Reproductive pattern of the anchovy, *Engraulis encrasicolus* (Linnaeus, 1758), in the Boka Kotorska Bay (Montenegro, southern Adriatic Sea). Acta Adriat. 59(2): 173-184. https://doi.org/10.32582/aa.59.2.2

Domínguez-Petit R., Anastasopoulou A., Cubillos L., et al. 2017. Maturity. Action COST FRESH (F0601). NAFO Working Group on Reproductive Potential.

Einarsson H., Flores F., Miñano J. 1966. El ciclo de madurez de la anchoveta peruana (*Engraulis ringens J.*). In: I Seminario Latinoamericano sobre el Océano Pacífico Oriental. Univ. Nacional Mayor, San Marcos, Lima, pp. 128-135.

Engelhard G.H., Dieckmann U., Godø O. R. 2003. Age at maturation predicted from routine scale measurements in Norwegian spring-spawning herring (*Clupea harengus*) using discriminant and neural network analyses. ICES J. Mar. Sci. 60(2): 304-313. https://doi.org/10.1016/S1054-3139(03)00017-1

Erickson D.L., Hightower J.E., Grossman G.D. 1985. The relative gonadal index: an alternative index for quantification of reproductive condition. Comp. Biochem. Physiol. A 81: 117-120. https://doi.org/10.1016/0300-9629(85)90276-2
FAO. 2024. The State of World Fisheries and Aquaculture 2024 – Blue Transformation in action. Rome. https://doi.org/10.4060/cd0683en
Fernandes J.A., Irigoien X., Goikoetxea N., et al. 2010. Fish recruitment prediction, using robust supervised classification methods. Ecol. Modell. 221(2): 338-352. https://doi.org/10.1016/j.ecolmodel.2009.09.020
Fernandes J.A., Irigoien X., Lozano J.A., et al. 2015. Evaluating machine-learning techniques for recruitment forecasting of seven North East Atlantic fish species. Ecological Informatics 25: 35-42. https://doi.org/10.1016/j.ecoinf.2014.11.004
Flores A., Wiff R., Díaz E. 2015. Using the gonadosomatic index to estimate the maturity ogive: application to Chilean hake (*Merluccius gayi gayi*). ICES J. Mar. Sci. 72(2): 508-514. https://doi.org/10.1093/icesjms/fsu155
Flores A., Wiff R., Ganias K., et al. 2019. Accuracy of gonadosomatic index in maturity classification and estimation of maturity ogive. Fish. Res. 210: 50-62. https://doi.org/10.1016/j.fishres.2018.10.009
Flores A., Wiff R., Donovan C.R., et al. 2024. Applying machine learning to predict reproductive condition in fish. Ecological Informatics 80: 102481. https://doi.org/10.1016/j.ecoinf.2024.102481
Fréon P., Bouchon M., Mullon C., et al. 2008. Interdecadal variability of anchoveta abundance and overcapacity of the fishery in Peru. Prog. Oceanogr. 79(2-4): 401-412. https://doi.org/10.1016/j.pocean.2008.10.011
Fukuda N., Aoyama J., Miller M.J., et al. 2013. Discriminant analysis of age to enable rapid estimation of 0-age Japanese eels Anguilla japonica. Coast. Mar. Sci. 36: 19-25.
Gerritsen H.D., McGrath D., Lordan C. 2006. A simple method for comparing age–length keys reveals significant regional differences within a single stock of haddock (*Melanogrammus aeglefinus*). ICES J. Mar. Sci. 63(6): 1096-1100. https://doi.org/10.1016/j.icesjms.2006.04.008
Heincke F. 1908. Bericht über die Untersuchungen der Biologischen Anstalt auf Helgoland zur Naturgeschichte der Nutzfische. In: Deutsche Wissenschaftliche Kommission für Meeresforschung, Die Beteiligung Deutschlands an der Internationalen Meeresforschung 1908:4/5:67-155.
Hunter J.R., Macewicz B.J. 2003. Improving the accuracy and precision of reproductive information used in fisheries. In: Kjesbu O.S., Hunter J.R., Witthames P.R. (eds), Report of the working group on Modern approaches to assess maturity and fecundity of warm- and cold-water fish and squids. Fish. Board Norw., Inst. Mar. Res., Bergen, pp. 57-68.
Ibaibarriaga L., Bernal M., Motos L., et al. 2007. Characterization of stage-classified biological processes using multinomial models: a case study of anchovy (*Engraulis encrasicolus*) eggs in the Bay of Biscay. Can. J. Fish. Aquat. Sci. 64(3): 539-553. https://doi.org/10.1139/f07-030
ICES. 2018. Report of the Workshop for Advancing Sexual Maturity Staging in Fish (WKASMSF), 30 April - 4 May 2018, ICES Headquarters, Copenhagen, Denmark. ICES CM/EOSG: 38. 75 pp.
Jamshidi E.J., Yusup Y., Kayode J.S., et al. 2022. Detecting outliers in a univariate time series dataset using unsupervised combined statistical methods: A case study on surface water temperature. Ecological Informatics 69:101672. https://doi.org/10.1016/j.ecoinf.2022.101672
Jordán R. 1980. Biology of the anchoveta. In: Summary of the present knowledge. Proceedings of the Workshop on the Phenomenon known as "El Niño". Guayaquil, Ecuador, 4–12 December 1974, United Nations Educational, Scientific and Cultural Organization, Paris, pp. 249-276.
Kell L., Nash R., Dickey-Collas M., et al. 2016. Is spawning stock biomass a robust proxy for reproductive potential? Fish and Fisheries 17(3): 596-616. https://doi.org/10.1111/faf.12131
Kjesbu O.S., Hunter J.R., Witthames P.R. 2003. Report of the working group on modern approaches to assess maturity and fecundity of warm- and cold-water fish and squids, Fisken og. Havet 12: 117-124.
Kjesbu O. 2016. Applied fish reproductive biology: contribution of individual reproductive potential to recruitment and fisheries management. In: Jakobsen T., Fogarty M., Megrey B., et al. (eds), Fish reproductive biology: implications for assessment and management. John Wiley and Sons, pp. 293-332. https://doi.org/10.1002/9781118752739.ch8
Lemaître G., Nogueira F., Aridas C. K. 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. J. Mach. Learn. 18(17): 1-5.
Liaw A., Wiener M. 2002. Classification and Regression by Random Forest. R News 2: 18-22.
Lowerre-Barbieri S., Ganias K., Saborido-Rey F., et al. 2011. Reproductive timing in marine fishes: variability, temporal scales, and methods. Mar. Coast. Fish. 3(1): 71-91. https://doi.org/10.1080/19425120.2011.556932
McPherson L.R., Ganias K., Marshall C.T. 2011. Inaccuracies in routinely collected Atlantic herring (*Clupea harengus*) maturity data and correction using a gonadosomatic index model. J. Mar. Biol. Assoc. U.K. 91(7): 1477-1487. https://doi.org/10.1017/S002531541100018X
Mori J., Buitrón B., Perea A., et al. 2011. Interannual variability of the reproductive strategy of the Peruvian anchovy off northern-central Peru. Ciencias Marinas 37(4B): 513-525. https://doi.org/10.7773/cm.v37i4B.1872
Okamura H., Morita S., Kuroda H. 2024. Forecasting fish recruitment using machine learning methods: A case study of arabesque greenling. Fish. Res. 278: 107096. https://doi.org/10.1016/j.fishres.2024.107096
Paluszynska A., Biecek P., Jiang Y., 2022. Explaining and Visualizing Random Forests in Terms of Variable Importance. Available: https://CRAN.R-project.org/package
Peña N., Mendo J., Pellón J.. 1989. Sexual maturity of Peruvian anchoveta (Engraulis ringens), 1961–1987. In: Pauly D, Tsukayama I (eds), The Peruvian Anchoveta and its Upwelling Ecosystem: Three Decades of Change. ICLARM Studies and Reviews 15, pp. 132-142.
Perea Á., Peña C., Oliveros-Ramos R., et al. 2011. Producción potencial de huevos, reclutamiento y veda reproductiva de la anchoveta peruana (*Engraulis ringens*): Implicaciones en el manejo pesquero. Ciencias marinas 37(4b): 585-601. https://doi.org/10.7773/cm.v37i4B.1827
Perea Á., Buitrón B., Mori J., et al. 2015. Anomalías de los índices reproductivos de anchoveta *Engraulis ringens* en relación al ambiente. Boletín Trimestral Oceanográfico 1: 25-26.
Potoschi A., Renones O., Cannizzaro L. 1999. Sexual development, maturity and reproduction of dolphinfish (*Coryphaena hippurus*) in the western and central Mediterranean. Sci. Mar. 63(3-4); 367-372. https://doi.org/10.3989/scimar.1999.63n3-4367
Quinn G.P., Keough M.J. 2002. Experimental design and data analysis for biologists. Cambridge university press, 557 pp. https://doi.org/10.1017/CBO9780511806384
Rahmani A., Kacher M., Iguer-Ouada M. 2020. Oocytes number and surface area as microsopic and quantative parameters to evaluate reproductive activity in *Engraulis encrasicolus* L. (Clupeiformes: Engraulidae). Cah. Biol. Mar. 61: 271-277.
Saborido-Rey F., Basilone G., Belcari P., et al. 2014. Report of the Workshop for Maturity Staging Chairs (WKMATCH), 11–15 June 2012, Split, Croatia. ICES, 59 pp.
Saetersdal G., Valdivia J. 1964. Un estudio del crecimiento, tamaño y reclutamiento de la anchoveta (*Engraulis ringens J*) basado en datos de frecuencia de longitudes. Bol. Inst. Invest. Recurs. Mar 1(4): 85-136.
Santander H., Flores R. 1983. Los desoves y distribución larval de cuatro especies pelágicas y sus relaciones con las variaciones del ambiente marino frente al Perú. FAO Fisheries Report (FAO). No. 291(2): 835-870.
Santander H., Alheit J., Smith P.E. 1984. Estimación de la biomasa de la población desovante de anchoveta peruana *Engraulis ringens* en 1981. Por aplicación del "método de producción de huevos". Bol. Inst. Mar Perú 8: 209-250.
Skjæraasen J.E., Nash R.D.M., Korsbrekke K., et al. 2012. Frequent skipped spawning in the world's largest cod population. Proc. Natl. Acad. Sci. U.S.A. 109(23): 8995-8999. https://doi.org/10.1073/pnas.1200223109
Sokolova M., Lapalme G. 2009. A systematic analysis of performance measures for classification tasks. Information processing and management 45(4): 427-437. https://doi.org/10.1016/j.ipm.2009.03.002

Somarakis S., Ganias K., Tsepes G., et al. 2004. Ovarian allometry and the use of gonadosomatic index: a case of study in the Mediterranean sardine, *Sardina pilchardus*. Mar. Biol. 146: 181-189. https://doi.org/10.1007/s00227-004-1419-2

Takahashi K., Mosquera K., Reupo J. 2014. El Índice Costero El Niño (ICEN): historia y actualización. Boletín técnico: Generación de modelos climáticos para el pronóstico de la ocurrencia del Fenómeno El Niño. Instituto Geofísico del Perú 1(2): 8-9.

Tam J., Blaskovic V., Goya E., et al. 2010. Relación entre anchoveta y otros componentes del ecosistema. Bol. Inst. Mar Perú 25: 95-108.

Taylor M.H., Tam J., Blaskovic V., et al. 2008. Trophic modeling of the Northern Humboldt Current Ecosystem, Part II: Elucidating ecosystem dynamics from 1995 to 2004 with a focus on the impact of ENSO. Prog. Oceanogr. 79: 366-378. https://doi.org/10.1016/j.pocean.2008.10.008

Therneau T., Atkinson B., Ripley B. 2013. Rpart: Recursive Partitioning. R Package Version 4.1-3. http://CRAN.R-project.org/package=rpart

Tsukayama I., Alvarez M.A. 1980. Fluctuations in the spawning stock of anchoveta during the spring reproduction cycles from 1964-1978. In: Sharp G.D. (ed), Workshop on the Effects of Environmental Variation on the Survival of Larval Pelagic Fishes. IOC Workshop Report No. 28. IOC/UNESCO, París, pp. 233-253.

Venables W.N., Ripley B.D. 2002. Modern Applied Statistics with S. Springer, New York, 497 pp. https://doi.org/10.1007/978-0-387-21706-2

Vitale F., Svedäng H., Cardinale M. 2006. Histological analysis invalidates macroscopically determined maturity ogives of the Kattegat cod (*Gadus morhua*) and suggests new proxies for estimating maturity status of individual fish. ICES J. Mar. Sci. 63(3): 485-492. https://doi.org/10.1016/j.icesjms.2005.09.001

Wyanski D.M., Brown-Peterson N. 2010. 4th Workshop on Gonadal Histology of Fishes. Fish Reproduction and Fisheries (FRESH), COST Action FA0601.

West G. 1990. Methods of assessing ovarian development in fishes: a review. Mar. Freshw. Res. 41(2): 199-222. https://doi.org/10.1071/MF9900199

Wright, P., Rowe S. 2019. Reproduction and spawning. In: Rose G. (ed), Atlantic Cod: A Bio-Ecology. John Wiley and Sons, pp. 87-132. https://doi.org/10.1002/9781119460701.ch3

Yoneda M., Kitano H., Selvaraj S., et al. 2013. Dynamics of gonadosomatic index of fish with indeterminate fecundity between subsequent egg batches: application to Japanese anchovy *Engraulis japonicus* under captive conditions. Mar. Biol. 160: 2733-2741. https://doi.org/10.1007/s00227-013-2266-9