USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

UNIVERSIDADE DA CORUÑA

Universidade Vigo

Trabajo Fin de Máster

The Impact of EU Landing Obligation on Spanish coastal Trawl Fleet Fishing Strategies

Santiago Arce Pilo

Máster en Técnicas Estadísticas

Curso 2024-2025

Propuesta de Trabajo Fin de Máster

Título en galego: O impacto da Obriga de Desembarque da UE nas estratexias de pesca da frota de arrastre costeira española
Título en español: El impacto de la Obligación de Desembarque de la UE en las estrategias de pesca de la flota de arrastre costera española
English title: The Impact of EU Landing Obligation on Spanish coastal Trawl Fleet Fishing Strategies
Modalidad: Modalidad B
Autor/a: Santiago Arce Pilo, Universidad de Vigo
Director/a: Javier Roca Pardiñas, Universidad de Vigo
Tutor/a: José Castro Pampillón, ; Adriana Nogueira Gassent,
Breve resumen del trabajo: El reglamento de la UE que prohíbe arrojar el descarte al mar y obliga a su desembarque en puerto, recogido bajo una nueva categoría de captura no permitida para consumo humano (BMS), se aplica en su totalidad desde el año 2019. El análisis histórico del descarte pesquero, a partir de datos científicos de los programas de muestreo a bordo, evidencian una reducción del descarte desde 2019 en relación al período anterior. No obstante, esta reducción no se corresponde con el volumen de los nuevos desembarques de la categoría BMS, lo que podría deberse a los cambios realizados por los pescadores en su estrategia de pesca para reducir captura no deseada y evitar ocupar las bodegas de sus barcos con biomasa de escaso valor económico. En esta línea de investigación proponemos aplicar diferentes métodos de modelado multivariante sobre datos científicos del programa de muestreo a bordo con el objetivo de identificar aquellas variables significativas que puedan explicar los cambios que ha estado aplicando la flota española para reducir sus capturas no deseadas.

Don Javier Roca Pardiñas, catedrático de la Universidade de Vigo de la Universidad de Vigo, don José Castro Pampillón, Científico Titular de Instituto Español de Oceanografía (IEO-CSIC) , y doña Adriana Nogueira Gassent, Investigadora M3 de Instituto Español de Oceanografía (IEO-CSIC) , informan que el Trabajo Fin de Máster titulado

The Impact of EU Landing Obligation on Spanish coastal Trawl Fleet Fishing Strategies

fue realizado bajo su dirección por don Santiago Arce Pilo para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal. Además, Don Javier Roca Pardiñasy don Santiago Arce Pilo

√ sí ☐ no

autorizan a la publicación de la memoria en el repositorio de acceso público asociado al Máster en Técnicas Estadísticas.

En Vigo, a 21 de julio de 2025.

El/la director/a:
Don/doña Javier Roca Pardiñas

ROCA PARDIÑAS JAVIER - 77592181W
Firmado digitalmente por ROCA PARDIÑAS JAVIER - 77592181W
Fecha: 2025.07.21 14:17:03 +02'00'

El/la director/a:
Don/doña Director/a 2

El/la tutor/a:
Don/doña José Castro Pampillón

Firmado por CASTRO PAMPILLON JOSE ANTONIO - DNI

El/la tutor/a:
Don/doña Adriana Nogueira Gassent

NOGUEIRA GASSENT ADRIANA - DNI 14307098V
Firmado digitalmente por NOGUEIRA GASSENT ADRIANA - DNI 14307098V
Nombre de reconocimiento (DN): c=ES, o=CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS, ou=CERTIFICADO ELECTRONICO DE EMPLEADO PUBLICO, ou=CENTRO OCEANOGRAFICO DE VIGO, ou=14307098V, serialNumber=IDCES-14307098V, sn=NOGUEIRA GASSENT, givenName=ADRIANA, cn=NOGUEIRA GASSENT ADRIANA - DNI 14307098V
Fecha: 2025.07.21 12:53:10 +02'00'

El/la autor/a:
Don/doña Santiago Arce Pilo

Declaración responsable. Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, Disposición 2978 del BOE nm. 48 de 2022), **el/la autor/a declara** que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas,...)
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración,...sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.

Agradecimientos

En primer lugar, me gustaría dar las gracias a mis tutores, Adriana y José, por todo el apoyo y la atención que me han brindado a lo largo de este proceso. Os estoy profundamente agradecido. También a Hortensia, por invitarme a tantos cafés y hacerme pasar un buen rato cada vez que fui a las oficinas del IEO.

Quiero expresar igualmente mi agradecimiento a mi familia, y en especial a mi madre, por enseñarme el valor de la constancia y la resiliencia. A Sara, por estar a mi lado en todo momento y brindarme un apoyo incondicional.

Finalmente, quiero agradecer a mis amigos de KV por estar siempre presentes; a Olimpio y a Joel, por ayudarme a creer en mí y motivarme; y a Kike, por acogerme en Santiago y darme un poco de paz.

Gracias a todos por motivarme a ser mejor y a seguir buscando mi lugar en el mundo.

Índice general

Resumen **XI**

1. Introduction. **1**

2. Statistical Framework. **5**
- 2.1. Features of zero-inflated proportion data . . . 5
- 2.2. Generalized additive models (GAMs) . . . 5
  - 2.2.1. Fundamentals of GAMs . . . 5
  - 2.2.2. Specification of Effects in GAMs . . . 6
- 2.3. Evaluation criteria and validation . . . 7
- 2.4. Modeling approaches for semicontinuous data . . . 9
  - 2.4.1. Hurdle Model or Two part. . . . 9
  - 2.4.2. Division by period . . . 11
  - 2.4.3. Alternative models . . . 12

3. Materials and Methods. **15**
- 3.1. Study Area . . . 15
- 3.2. Sampling design and data source . . . 16
- 3.3. Variables and preprocessing . . . 18
- 3.4. Methods . . . 19
  - 3.4.1. Model selection and validation . . . 21

4. Results **23**
- 4.1. General Drivers of discarding . . . 23
  - 4.1.1. Probability of discard . . . 23
  - 4.1.2. Proportion discard. . . . 27
- 4.2. Effect of the Landing Obligation . . . 31
  - 4.2.1. Probability of discarding . . . 31
  - 4.2.2. Proportion discarded . . . 35

5. Discussion. **39**
- 5.1. General discard drivers . . . 39
- 5.2. Impact of the Landing Obligation . . . 42
- 5.3. Study limitations and future work . . . 43

6. Conclusions **45**

A. Apendix **47**

Bibliography **51**

Resumen

Resumen en español

La Obligación de Desembarque, introducida por la Unión Europea en 2013 y plenamente vigente desde 2019 para todas las especies sujetas a regulación, prohíbe arrojar al mar las capturas no deseadas, exigiendo su desembarco en puerto bajo una categoría específica que excluye su comercialización para consumo humano. Esta medida pretende incentivar una pesca más selectiva y sostenible. No obstante, su implementación ha suscitado interrogantes sobre su eficacia real, los cambios operativos inducidos y los factores que determinan el descarte.

Este Trabajo Fin de Máster analiza el impacto de esta normativa en las estrategias de pesca de la flota de arrastre costera española, con especial atención a la merluza europea (Merluccius merluccius), una especie de elevada importancia comercial y ecológica. El análisis se basa en datos del programa de muestreo científico a bordo, recogidos entre 2009 y 2023 en aguas del Cantábrico y Noroeste.

Se aplicó un enfoque estadístico en dos partes, mediante modelos aditivos generalizados. La primera parte modela la probabilidad de que se produzca un descarte y la segunda, la proporción descartada cuando este ocurre. Para identificar posibles cambios derivados de la normativa, se incluyeron términos diferenciados por periodo (antes y después de su implementación).

El estudio analiza de forma explícita los factores que influyen en el descarte, destacando la profundidad, la velocidad del buque, la distancia a la costa, la estacionalidad y el reclutamiento de juveniles. Los resultados muestran que, tras la entrada en vigor de la normativa, la proporción de capturas descartadas se ha reducido significativamente, mientras que la probabilidad de que ocurra un descarte no ha variado de forma apreciable. Esto sugiere que los pescadores han modificado sus prácticas para minimizar el volumen de captura no deseada, probablemente adoptando estrategias más selectivas.

Este trabajo aporta evidencia empírica sobre la respuesta adaptativa del sector pesquero ante un marco regulador complejo, y subraya la utilidad del modelado estadístico avanzado para entender cómo los factores ecológicos, técnicos y espaciales condicionan las decisiones de descarte. Los resultados contribuyen a mejorar el diseño y la evaluación de políticas pesqueras más eficaces y sostenibles.

English abstract

The Landing Obligation, introduced by the European Union in 2013 and fully implemented in 2019 for all regulated species, prohibits the discarding of unwanted catches at sea and requires their landing in port under a specific category that excludes them from human consumption. The regulation aims to encourage more selective and sustainable fishing practices. However, its implementation has raised concerns about its actual effectiveness, the operational changes it has triggered, and the underlying drivers of discarding behavior.

This Master's Thesis examines the impact of this regulation on the fishing strategies of the Spanish coastal trawl fleet, with a focus on European hake (Merluccius merluccius), a species of high commercial and ecological value. The analysis uses scientific at-sea observer data collected between 2009 and 2023 in the Cantabrian and Northwestern waters.

A two-part statistical modeling approach was applied, using generalized additive models to analyze separately (1) the probability that discarding occurs and (2) the proportion discarded when it does. Period-specific effects were included to evaluate changes before and after the regulation's implementation.

The study explicitly investigates the factors that influence discarding, identifying mean depth, vessel speed, distance to shore, seasonal variation, and juvenile recruitment as key drivers. The results show that while the probability of discarding has remained largely unchanged after the implementation of the regulation, the proportion of catch discarded during discard events has significantly decreased. This pattern suggests that fishers have adapted their strategies to reduce the volume of unwanted catch, likely through more selective fishing practices.

This research provides empirical evidence of the adaptive responses of the fishing sector to a complex regulatory framework. It also highlights the value of advanced statistical modeling in understanding how ecological, technical, and spatial factors influence discarding behavior. The findings offer useful insights for improving the design and evaluation of fisheries management measures aimed at reducing unwanted catch and enhancing sustainability.

Capítulo 1

Introduction.

Marine fisheries provide essential protein and economic benefits to billions of people globally, yet they simultaneously exert significant pressure on marine ecosystems through extraction activities that can alter population dynamics and community structure. Historically, marine resources were perceived as inexhaustible, leading to patterns of unlimited exploitation, particularly in international waters. However, technological advances in fishing techniques and species location methods have demonstrated that these resources are finite and can become depleted without proper management. This paradigm shift has motivated the creation of regulatory frameworks and management systems to govern fishing activities, protect marine ecosystems, and ensure the long-term sustainability of fisheries.Decades of intensive exploitation of fisheries, particularly on continental shelves, have resulted in steadily declining catches and the collapse of numerous fish and crustacean stocks (Pauly et al., 2003)

Among the various impacts of fishing activities, discarding represents one of the most problematic aspects of modern fisheries, generating both ecological and economic inefficiencies that undermine sustainable resource management. Discards, defined as "the total live weight of undersized, unsaleable or otherwise undesirable whole fish discarded at the time of capture or shortly afterwards" (FAO, 2020), represent a significant concern in fisheries management. Discarding arises from multiple factors: regulations prohibiting the landing of fish below minimum size, lack of commercial value, vessel hold space limitations, exhausted quotas for specific species, and prohibited species catches. Beyond direct mortality impacts, discarding practices fundamentally alter marine food webs and trophic chains, creating cascading effects throughout the ecosystem. When large quantities of dead or dying fish are released back into the sea, they can alter the dynamics, modify benthic habitats, and disrupt natural predator-prey relationships. Discarding practices directly increase mortality in fish populations particularly among juveniles of commercial species and disturb the ecological equilibrium of marine ecosystems (Borges, 2015), often in ways that are difficult to quantify in stock assessments due to insufficient data collection and monitoring. This practice generates a waste of biomass that could contribute to future recruitment and affects scientists' ability to accurately assess population status and ecosystem health.

To address this issue, the European Union (EU) established the Data Collection Regulation (DCR; EU, 2000) to outline principles for collecting scientific and fisheries data alignment with the Common Fisheries Policy (CFP). Data harmonization became crucial for managing fisheries that operate across national boundaries and for establishing conservation measures based on solid scientific evidence. This program played a fundamental role in harmonizing fishing data across EU Member States. In 2009, the DCR was replaced by the Data Collection Framework (DCF; EU, 2008), which introduced metier-based sampling—where a metier refers to a specific fishing operation characterized by a particular combination of fishing gear, target species, fishing area, and vessel type—instead of stock-based sampling, allowing ecosystem objectives to be achieved through more comprehensive monitoring. This approach recognizes that fishing operations are complex and that discard impacts vary considerably depending on the type of fishery, season, and location.

The European Union regulates the exploitation of its fishery resources through Total Allowable Catches (TACs), which are subsequently distributed among Member States as national quotas. This quota-based management system relies on accurate quantification of total catches, including both landed and discarded portions. Discards significantly complicate this management approach because TACs based solely on landed catches ignore a substantial fraction of actual fishing mortality, potentially leading to overexploitation of stocks and compromising the effectiveness of conservation measures.

Despite these improvements in data collection, the basic problem of extracting high levels of biomass only to return a significant portion into the sea persisted. The fundamental issue lies in the fact that fishers often capture species or sizes of fish that they cannot or do not wish to retain, whether due to legal restrictions, lack of market demand, or operational limitations. Therefore, the revised CFP (EU, 2013) established Article 15 termed the Landing Obligation, LO,which mandates that all individuals of species subject to management measures must be kept on board, landed, and counted against quotas. The Below Minimum Size (BMS) category was established to accommodate fish that would previously have been discarded—specifically, catches below the minimum conservation reference size that must now be landed but cannot be sold for direct human consumption. These fish, although having minimal commercial value, must be accounted for in fishing quotas and properly documented, which fundamentally changeswhile the economic and operational dynamics of fishing operations. The regulation's primary goal was to eliminate discards by encouraging fishers to adopt more selective fishing practices and avoid unwanted catches of managed stocks. Implementation of the LO began progressively from January 2015 until its full application to all European fisheries for stocks with management measures in January 2019.

However, the practical implementation of the LO has faced significant challenges, including technical limitations in handling and storing unmarketable fish, socioeconomic conflicts related to reduced vessel profitability, and operational difficulties in adapting fishing practices to avoid unwanted catches. The quality of discard data recorded in fishing logbooks remains limited, making scientific estimation essential for accurate stock assessments and policy evaluation.

Recent work has demonstrated that the LO has significantly reduced discards for species subject to management measures in Spanish fleets (Davie et al,2025). However, this reduction raises questions about whether discards have truly been eliminated or simply transferred to other recording categories. However, this reduction does not correspond with the volume of new Below Minimum Size (BMS) landings or the new catch category established by the L0 to record discards that must now be landed in port (personal communication). Additionally, informal interviews with fishing vessel captains suggest they have adopted changes in fishing strategy to reduce unwanted catch and avoid filling vessel holds with biomass of little economic value.

Despite these observations, a significant knowledge gap remains regarding the specific operational, spatial, and temporal factors that have driven these changes in fishing strategy. Understanding these changes is fundamental for evaluating the real effectiveness of the regulation and for developing more effective management strategies. Most studies have primarily focused on predicting discard patterns based on historical observations (Celic et al., 2018; Pennino et al., 2020), evaluating economic impacts (Villasante et al., 2019), or assessing compliance (STECF, 2024), but few have examined how fishing strategies have actually changed in response to the LO regulation using statistical modeling approaches.

The Northwestern Spanish waters (ICES Division 8c and 9aN) support important demersal and pelagic fisheries, with the Spanish non-Basque bottom trawl fleet being a significant component (Lema et al., 2006). These waters represent an ideal case study due to their fishing importance, the diversity of species caught, and the historical presence of high discard levels. The "baca" bottom otter trawl specifically targets demersal species such as European hake (Merluccius merluccius), megrim (Lepidorhombus spp), and anglerfishes (Lophius spp), historically showning substantial discard levels. ICES (International Council for the Exploration of the Sea) provides the scientific framework for fisheries management in the North Atlantic, dividing marine areas into statistical divisions for stock assessment and management purposes. This study focuses on quantifying and analyzing the discards of European hake, a species of particular commercial and ecological importance in these waters.

The objectives of the study are:

- To identify the most appropriate statistical approach for modeling the discard process.
- To analyze the factors influencing discarding using statistical models appropriate for the erratic and complex nature of this type of data.
- To assess the effect of a management measure originally designed to reduce discarding.

By employing robust statistical modeling techniques on a comprehensive data set spanning from 2009 to 2023, this study will provide valuable insights into how fishing fleets adapt their operational strategies in response to new management regulations, These findings would inform future policies aimed at reducing discards and improving the sustainability of European fisheries.

Capítulo 2

Statistical Framework.

2.1. Features of zero-inflated proportion data

Proportion data with an abundance of zeros represents a common analytical challenge in ecological and environmental research This type of data consist of continuous values constrained between 0 and 1, characterized by a significant accumulation of observations at exactly zero (Warton, 2005; Fletcher & Fortin, 2018). Unlike classic zero-inflated models that assume two distinct processes, this data situation indicates a single process generating both zeros and non-zero values, with zeros simply occurring at high frequency.

This structure is common in ecological studies where proportions represent coverage, relative abundance, or detection rates of organisms in environments where they are frequently absent or below detection limits. The key distinction here is that zeros are not generated by a separate mechanism but rather represent one extreme of the same underlying continuous process that generates all observations (Pirotta et al., 2011).

In statistical terms this type of distribution is positively skewed bounded on the left and the zero, with a probability mass concentrated at this boundary point. If Y represents the proportion variable where $Y \in [0, 1]$ and π the probability of observing a zero. The distribution can be considered as a continuous distribution with a significant probability mass at zero:

$$P(Y = y) = \begin{cases} 1 - \pi, & y = 0, \\ \pi\, f(y), & 0 < y \leq 1, \end{cases} \tag{2.0}$$

where $f(y)$ denotes the *continuous density* of Y conditional on $Y > 0$, satisfying $\int_0^1 f(y)\, dy = 1$.

As mentioned before, this type of data structure presents several analytical challenges. First, the large number of zeros creates a highly skewed distribution that violates the assumptions required by many standard statistical approaches. Second, the boundary constraint at zero, combined with the empirical concentration of values near zero, leads to heteroscedasticity, since variance tends to depend on the mean (Chen & Cheng, 2005). Finally, the presence of exact zeros alongside continuous positive values means that few standard continuous distributions can adequately model the entire range of data; even when they can, forcing everything into a single distribution may obscure important aspects of the underlying processes.

2.2. Generalized additive models (GAMs)

2.2.1. Fundamentals of GAMs

Generalized additive models (GAMs) are an extension of Generalized Linear Models (GLM) by allowing non-linear relationships between the response variable and the predictors.The way non linearity is introduce is by the use of smooth functions (Hastie & Tibshirani, 1990; Wood, 2017). This type of models are suitable for ecological data where in many cases the relationship between the response variable and the predictors is non-linear.

Considering that Y is the response variable, $g(\cdot)$ is the (single) link function, β_0 is the intercept, $s_j(x_j)$ are smooth functions of the continuous predictors x_j, and β_k are parametric coefficients for the categorical predictors z_k, the general formulation of a GAM is:

$$g\big(\mathbb{E}[Y \mid \mathbf{X}]\big) \;=\; \beta_0 + \sum_{j=1}^{p} s_j(x_j) + \sum_{k=1}^{q} \beta_k \, z_k, \tag{2.2}$$

where $\mathbf{X} = (x_1, \ldots, x_p)^\top$ denotes the vector of continuous predictors.

GAMs retain the distributional assumptions of GLMs, this allows to specify different error distributions (e.g., binomial, Poisson, gamma, beta) which are appropriate for different types of response variables.The link function $g(.)$ plays a crucial role in GAMs by connecting the linear predictor to the expected value of the response variable. They transform the predicted values to ensure they remain within appropriate bounds for the response variable while allowing the model to operate on an unbounded scale. The choice of link function substantially influences both model fit and interpretation of results.Mathematically a link function $g(.)$ relates the mean fo the response variable $\mu = \mathbb{E}(Y \mid X)$ to the linear predictor $\eta = \beta_0 + \sum_{j=1}^{p} s_j(x_j) + \sum_{k=1}^{q} \beta_k z_k$ by $\quad \mu = g^{-1}(\eta)$ where $g^{-1}(\cdot)$ is the inverse of the link function. Some common link functions are the logit link, used for binomial data, the log link ,for count and continuous positive data or the Identity link which is usually used for normality distributed data.

The estimation of GAMs involves penalised likelihood maximisation to find the optimal balance between model-fit accuracy and smoothness:

$$\big(\hat{\boldsymbol{\beta}}, \hat{s}_1, \ldots, \hat{s}_p\big) \;=\; \underset{\boldsymbol{\beta}, \mathbf{s}}{\arg\max} \Big\{ L(\boldsymbol{\beta}, \mathbf{s}) - \sum_{j=1}^{p} \lambda_j \int \big[s_j''(x) \big]^2 \, dx \Big\}, \tag{2.3}$$

where $L(\boldsymbol{\beta}, \mathbf{s})$ is the log-likelihood and λ_j are smoothing parameters that control the trade-off between model fit and smoothness. These parameters are typically selected via generalised cross-validation (GCV) or restricted maximum likelihood (REML).

In the context of this analysis, GAMs can provide provide a powerful framework for capturing the potentially complex non-linear effects of environmental, spatial, temporal, and operational factors on discard patterns without imposing restrictive parametric assumptions.

2.2.2. Specification of Effects in GAMs

The flexibility of Generalized Additive Models (GAMs) stems from their ability to incorporate various types of smooth functions, each designed to capture specific patterns in ecological data. These functions allow researchers to model complex relationships without imposing restrictive parametric forms (Wood, 2017).

The univariate smooths are the effect of a single continuous predictor on the response variable. These type of functions are designed to captured the non-linear relationships without making assumptions over the functional form(eg,linear or cuadratic).Univariate smooths are typically implemented using spline-based aproaches like cubic regression splines, thin plate regression splines or P-splines (Ruppert et al., 2003) .For example ,variables such as vessel speed or mean depth can benefit from this type of smoother and it can reveal complex relationships between them.This can been seen in where discard rate peaks at intermediate speeds where a greater diversity of species is captured, in the other hand more extreme speeds result in more selective fishing.

Tensor product interactions model how the joint effect of two continuous variables might differ from their individual effect. Tensor products are constructed from the products of basis functions for the marginal smooths, allowing for different levels of smoothness in each dimension. This is particularly important in ecological data where variables often operate on different scales and require different degrees of smoothness. Unlike isotropic smoothers like thin plate splines, tensor products can appropriately handle variables measured in different units (meter vs. degrees) or with different scales of variability (Márquez et al., 2022).s,these types of functions are essential for capturing complex ecological interactions (Wood, 2017).

Cyclic smooths are essential for temporal circular variables such as hour of day or month of year. These functions ensure continuity and smoothness at the boundaries, preventing artificial jumps between, for example, December and January, or between 23:00 and 00:00 hours (Simpson, 2018). Cyclic smooths are implemented by imposing constraints on the basis functions to ensure that the function value and its derivatives match at the endpoints of the range. In ecological studies, cyclic smooths are crucial for modeling seasonal effects, day patterns, and other cyclical environmental processes that influence species distributions and behaviors (Wood, 2017).

Spatial effects are typically modeled using bivariate smooths of geographic coordinates. Several approaches exist for modeling spatial effects, including thin plate splines, soap film smooths for complex boundaries, and Gaussian process models (Miller et al., 2019). These spatial smooths capture spatial autocorrelation and heterogeneity in ecological processes that cannot be explained by measured environmental predictors. They can reveal underlying spatial patterns in species distributions, ecosystem functions, or environmental responses that reflect unmeasured variables or historical legacies (Borcard et al., 2004).

Random effects account for hierarchical or grouped structure in the data. Random effects in GAMs are implemented as penalized ridge functions and can be viewed as smooths with a specific basis and penalty (Wood, 2017). They account for the non-independence of observations within groups and allow the estimation of both group-specific deviations and population-level effects. In ecological studies, random effects are commonly used to account for repeated measurements, spatial or temporal blocking structures, and variation among sampling units such as plots, transects, or individuals (Zuur et al., 2014).

In GAMs just like in many other regression techniques,it is often important to allow for differences in how predictors influence the response across distinct groups, conditions, or time periods. While traditional regression models address this using interaction terms, Generalized Additive Models (GAMs) provide a more flexible approach: group-specific smooths.This modeling strategy estimates a separate smooth function for each level of a categorical variable, such as treatment group, spatial region, or time period. For example, if there is the suspicion that a variable like temperature affects the outcome differently before and after a policy change.

2.3. Evaluation criteria and validation

The validation of Generalized Additive Models (GAMs) requires a structured methodological framework to ensure both statistical validity and ecological interpretability of results. Given the semiparametric nature of GAMs and the typical complexity of ecological data, multiple validation procedures are employed to evaluate different aspects of model adequacy (Zuur et al., 2009; Wood, 2017).

The selection of appropriate smooth terms constitutes a fundamental step in GAM validation. This selection should be based on the ecological understanding of the system being studied, the scales at which processes operate, and the structure of the available data. Model selection techniques such as AIC, BIC, or cross-validation can help determine which smooth terms to include, while visualization of the estimated smooth functions provides insights into the underlying ecological processes (Zuur, 2012).

For each smooth term included in the models, the basis dimension (k) should be evaluated using the k-index proposed by Wood (2017):

$$\text{k}-\text{index} = \frac{\hat{\rho}}{\sqrt{2/edf}} \tag{2.4}$$

where $\hat{\rho}$ represents the estimated first-order correlation of the residuals when ordered by the predictor variable, and edf denotes the effective degrees of freedom.

The statistic $\hat{\rho}$ in Equation (2.4) is the empirical estimate of first-order autocorrelation among the residuals associated with a given smooth term. It plays two complementary roles in GAM validation:

Diagnostic for residual independence $\hat{\rho}$ value close to zero implies that, after accounting for the fitted smooth, the residuals are approximately independent with respect to the ordering of the covariate. Pronounced positive or negative values indicate unmodelled structure—either remaining non-linearity (positive) or over-flexible fits (negative). This diagnostic guides refinement of the basis dimension and, if necessary, the inclusion of interaction or random-effect terms.

Scaling factor in the k-index.By normalising $\hat{\rho}$ with $\sqrt{2/edf}$, the k-index provides a scale-free measure that simultaneously accounts for the complexity of the smooth (via edf) and the strength of autocorrelation. A rule-of-thumb threshold of 1 is routinely adopted (Wood, 2017):

- If $k - index < 1$, the chosen basis dimension is adequate.
- If $k - index > 1$, increase k to allow additional wiggliness and re-fit the model.

Like multicolinearity in linear regression models there is a analog for non parametric models called concurvirty.The concurvity is used to identify potential redundancy among predictors.For a GAM model this metric can be computed by using the concurvity() function in the mgcv package, following the approach recommended by Simpson (2018). Concurvity values range from 0 (no concurvity) to 1 (perfect concurvity), with values exceeding 0.8 typically indicating problematic levels that may compromise model stability and parameter estimation (Wood, 2017). In ecological analyses, it is common to observe some concurvity between naturally correlated environmental predictors, such as depth and distance to coast (Borcard et al., 2004). On hand standard diagnosis plots mus be examined to asses model adequacy in regression models,following the recommendations of Zuur et al. (2009) and Wood(2017). Quantile-Quantile (Q-Q) plots are used to evaluate whether the residuals follow a the expected distributions.On the other hand graphics of residuals versus fitted values plots should be inspected for systematic patterns that might indicate incorrect specification over the model structure.

When GAMs are employed for binary classification problems, additional validation procedures specific to classification performance must be implemented alongside traditional regression diagnostics. The evaluation of classification GAMs requires assessment of the model's discriminatory ability and predictive accuracy through specialized metrics that quantify different aspects of classification performance (Hosmer et al., 2013; Kuhn & Johnson, 2013).

The Receiver Operating Characteristic (ROC) curve represents the primary tool for evaluating binary classification performance, plotting sensitivity (true positive rate) against 1-specificity (false positive rate) across all possible threshold values. The Area Under the Curve (AUC) provides a single summary measure of discriminatory performance, with values ranging from 0.5 (random classification) to 1.0 (perfect discrimination). Models with AUC values between 0.7-0.8 are considered acceptable, 0.8-0.9 indicate good performance, and values above 0.9 represent excellent discriminatory ability (Swets, 1988). Cross-validation techniques become particularly important for classification GAMs to ensure robust performance estimates and prevent overfitting. K-fold cross-validation should be employed to partition the dataset into training and validation subsets, with the model fitted on training data and performance evaluated on held-out validation data. This procedure should be repeated multiple times to obtain stable estimates of classification metrics and their confidence intervals (Hastie et al., 2009).

Confusion matrices provide detailed insight into classification errors by tabulating predicted versus observed classifications. From the confusion matrix, multiple performance metrics can be derived including sensitivity (proportion of true positives correctly identified), specificity (proportion of true negatives correctly identified), positive predictive value, and negative predictive value. The overall accuracy, while intuitive, can be misleading in cases of class imbalance and should be interpreted alongside other metrics (Fawcett, 2006).

For ecological classification problems with imbalanced datasets, where one outcome class substantially outnumbers the other, additional considerations become necessary. The precision-recall curve may provide more informative assessment than ROC curves in such scenarios, as it focuses on the performance regarding the minority class. The F1-score, representing the harmonic mean of precision and recall, offers a balanced measure that accounts for both false positives and false negatives (Davis & Goadrich, 2006).

Calibration assessment should also be conducted to evaluate whether predicted probabilities accurately reflect the true likelihood of outcomes. Calibration plots comparing predicted probabilities to observed frequencies across probability bins can reveal systematic over- or under-estimation of risk. The Hosmer-Lemeshow goodness-of-fit test provides a formal statistical assessment of calibration quality, though visual inspection of calibration plots often provides more practical insights (Steyerberg, 2009).

To make a comparison between the models fitted a criteria must be follow to ensure a robust selection.One of the most populars informative criterion are the Akaike Information Criterion (AIC) (Burnham and Anderson,2002) or the Bayesian Information Criterion (BIC) (Schwarz,1978) , both of these can be computed to quantify the balance between model fit and complexity:

$$\begin{aligned} \text{AIC} &= -2\ln(L) + 2p \\ \text{BIC} &= -2\ln(L) + p\ln(n) \end{aligned} \tag{2.5}$$

where L is the model likelihood, p is the effective number of parameters, and n is the sample size. Models with lower AIC and BIC values are considered to provide a better balance between fit and parsimony (Zuur et al., 2014). Additionally, the proportion of deviance explained should be calculated to assess the explanatory power of each model analogously to R^2 in linear regression.

As said before model evaluation can follow a systematic backward selection strategy, with this procedure the starting model must be comprehensive incorporating all potentially relevant predictors and subsequently removing terms based on the statistical significance and practical importance (Miller et al., 2019).Each simplification should be evaluated by examining its impact on AIC, BIC, and explained deviance, ensuring that model parsimony is achieved without substantial loss of explanatory power.

The validation procedures collectively should support the adequacy of final models for capturing patterns in ecological data while satisfying necessary statistical assumptions. This comprehensive approach to model validation aligns with best practices in ecological modeling (Zuur and Ieno, 2016) and provides confidence in the robustness of subsequent inferences about factors influencing the ecological patterns under study.

2.4. Modeling approaches for semicontinuous data

Multiple statistical frameworks have been developed to model semicontinuous data with excess zeros. This section presents the approaches most relevant to fisheries discard analysis, focusing first on the two-part hurdle model implemented in this study, followed by alternative approaches.

2.4.1. Hurdle Model or Two part.

The hurdle model represents a two-part modeling strategy that divides the data generating process into two components: a binary process that determining the presence or absence of the response variable (zero vs non zero values) and a continous process modeling the positive values but conditional to the response variable being non zero (Zuur et al., 2009; Miller et al., 2019).This approach, first formally proposed by Cragg (1971) for econometric application ,recognizes that the mechanisms that generates the zeros differs from the one influencing the magnitude of non-zero observations, a characteristic particularly relevant to discard data (Richards, 2008).

In fisheries science, hurdle models have gained significant traction, with Pennington (1983) being among the first to apply a delta approach to fisheries catch data, and more recently Maunder and Punt

(2004) demonstrating their utility in stock assessment contexts. These models have proven particularly valuable for analyzing discard data, as demonstrated by Cosandey-Godin et al. (2015) who employed hurdle models to investigate bycatch patterns in pelagic longline fisheries.

The hurdle model can be formulated

This formulation effectively "hurdles" over the zeros, treating them as a separate process (Pedersen et al., 2019).

A very important advantage of the hurdle approach is its flexibility when accommodating different types of covariates for each component, this allows ecological variables to influence the occurrence versus the magnitude of the discards(Zuur et al., 2014).For fisheries discard data, where zero observations might represent trips where species-specific discarding is effectively avoided, while non-zero observations reflect variable discard rates based on other factors, this flexibility is particularly valuable.

Binomial component

The binary component of the hurdle model predicts the probability of observing a non-zero value using a binomial GAM with a logit link function (Wood, 2017):

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \sum_{j=1}^{J} s_j(x_j) \tag{2.7}$$

where π is the probability of observing discards (i.e., a non-zero value), α is the intercept, and $s_j(\cdot)$ are smooth functions of the continuous predictors x_j.

In this case the choice of a logit link function is mainly because constrains predictions into the $(0, 1)$ interval,ensuring valid probability estimates (Simpson,2018).

The binomial component explains two important questions: the first one is what factors influence whether discards occur at all, and second, under what tyoe of conditions are zero discards more likely to happend.In fisheries contexts, this component can help explain the circumstances under which fishers can completely avoid discarding certain types of species under which fishers can completely avoid discarding certain species, this has led to management applications like in Borcard et al (2004).

Beta component

Once the presence or absence of discards is modeled, a second stage GAM is fitted for the continuous data using a beta distribution with a logit link function.Contrary to zero inflated models (these will be explain in section 3.4.3.2) that attempt to model zeros as part of the response distribution, the hurdle approach separates completely the modeling of zeros in the binomial component from the modeling of positive values in the beta component (Ferrari and Cribari-Neto, 2004). This clean separation is particularly appropriate for discard data where zeros often represent a distinct ecological process from positive values.

$$\text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \sum_{k=1}^{K} s_k(x_k), \quad \text{for } Y > 0. \tag{2.8}$$

where μ is the expected value of the discard proportion conditional on discards being present, β_0 is the intercept, and $s_k(\cdot)$ are smooth functions of the continuous predictors x_k. The beta distribution is applied exclusively to the subset of data where positive discard values are observed, completely excluding zeros from this part of the model.

The choice of predictors for the beta component may differ from those in the binomial component, reflecting potentially different processes driving discard occurrence versus discard rate. For example, certain spatial or seasonal factors might influence whether any discarding occurs, while vessel characteristics or catch composition factors might more strongly influence the magnitude of discarding when it does occur (Zuur et al., 2014).

An important practical consideration when implementing the beta component is handling values exactly equal to 1, as the beta distribution is defined on the open interval (0,1). A common approach is to apply a minor transformation to these boundary values:

$$y' = \begin{cases} \frac{y\,(n-1)+0{,}5}{n}, & y = 1, \\ y, & y \neq 1 \end{cases} \tag{2.9}$$

This transformation shrinks values of 1 slightly toward 0.5, preserving the relative ordering while making the values compatible with the beta distribution (Zuur, 2012). Since zeros are handled entirely by the binomial component, no transformation is needed for zero values in the beta component.

2.4.2. Division by period

The division of fisheries data into distinct temporal periods represents a critical consideration in modeling discard patterns, as ecological processes, management regulations, and fishing practices often exhibit substantial temporal heterogeneity (Pedersen et al., 2019). This temporal stratification approach acknowledges that the drivers of discarding behavior may fundamentally differ across distinct time periods, necessitating separate model parametrizations or even different model structures.

In fisheries research, temporal divisions typically follow management-relevant units such as quarters, seasons, or regulatory periods (Miller et al., 2019). The optimal temporal resolution depends on both the ecological processes under investigation and the management context. For example, seasonal divisions may better capture changes in species composition and recruitment patterns, while regulatory periods more effectively model shifts in fisher behavior in response to management interventions (Zuur et al., 2014).

When implementing temporal stratification within a GAM framework, several approaches can be employed. The simplest method involves fitting separate models to data from different periods, which allows for complete flexibility in model specification across periods but may reduce statistical power and increase computational demands (Wood, 2017). A more integrated approach incorporates period as a categorical variable interacting with smooth terms, allowing the shape of functional relationships to vary across periods while retaining the full dataset structure (Simpson, 2018). This can be implemented using the “by” variable functionality in mgcv.

Another approach uses varying coefficient models, where the effect of key predictors is allowed to vary smoothly with time, capturing gradual changes in relationships rather than abrupt transitions between periods (Zuur, 2012). This is particularly useful when the temporal changes in discarding behavior are expected to be continuous rather than discrete.

The selection between these approaches should be guided by both ecological understanding and model diagnostics. Comparing model performance metrics (AIC, BIC, explained deviance) across different temporal stratification schemes provides objective criteria for determining the optimal temporal resolution (Richards, 2008). Additionally, examining the estimated smooth functions for each period can reveal important insights into how the drivers of discarding behavior change over time.

The division by period approach carries important implications for interpretation and management. Identifying period-specific relationships between fishing practices and discard rates can inform temporally-targeted management interventions, such as seasonal closures, periodic gear restrictions, or time-varying landing obligations (Borcard et al., 2004). Furthermore, understanding temporal heterogeneity in discard drivers enhances the ability to forecast future discard patterns under changing environmental or management conditions.

However, researchers must also consider the trade-offs involved in temporal stratification. Excessive division can lead to overfitting and reduced statistical power, particularly when data are limited for certain periods. Conversely, insufficient temporal resolution may obscure important patterns and lead to misspecification of management recommendations. The optimal balance depends on the specific research questions, data availability, and management objectives (Augustin et al., 2013; Hartig, 2020).

2.4.3. Alternative models

While the hurdle model approach represents a flexible framework for addressing semicontinuous data with excess zeros, alternative modeling approaches may be more appropriate in certain contexts. This section explores two prominent alternatives: Tweedie models and zero-inflated beta models, discussing their statistical properties, advantages, limitations, and applications in fisheries discard analysis.

Tweedie models

Tweedie distributions belong to the exponential family and are characterized by a variance function of the form $Var(Y) = \phi\,\mu^p$, where μ is the mean, ϕ is the dispersion parameter, and p is the power parameter (Wood, 2017). For $1 < p < 2$, Tweedie distributions can model semicontinuous data with exact zeros and continuous positive values within a single model framework, avoiding the need for separate model components (Pedersen et al., 2019).

The probability density function of the Tweedie distribution cannot generally be expressed in closed form, but can be approximated numerically. Within the GAM framework, Tweedie models can be implemented using the mgcv package with the family specification 'tw(p)', where p is either fixed a priori or estimated from the data (Wood, 2017).

A key advantage of Tweedie models in fisheries discard analysis is their ability to model the discard process as a unified phenomenon rather than as two separate processes. This approach may be more appropriate when the zero and positive components share common drivers or when the distinction between the occurrence and magnitude processes is not ecologically meaningful (Miller et al., 2019,Browne 2024). Additionally, Tweedie models avoid potential issues with combining predictions from two separate models, as required in hurdle approaches (Zuur et al., 2009).

However, Tweedie models also present significant challenges. The parameter p controlling the distribution shape must be either specified a priori or estimated from the data, introducing an additional source of uncertainty. Furthermore, interpretation of model coefficients is less intuitive than in hurdle models, as effects simultaneously influence both the probability of zeros and the magnitude of positive values (Zuur and Ieno, 2016).

In discard analysis applications, Tweedie models have demonstrated utility particularly when the research question focuses on total discard amounts rather than the separate processes governing discard occurrence and magnitude. For example, when modeling total discard rates across multiple species, the exact zero mechanism may be less important than accurately capturing the overall pattern (Richards, 2008).

The appropriate value of the power parameter p in fisheries applications typically falls between 1.2 and 1.7, with values closer to 1 indicating a distribution dominated by zeros and values closer to 2 indicating a distribution increasingly resembling a gamma distribution (Augustin et al., 2013). Cross-validation approaches can help identify the optimal power parameter for a given dataset, although simulation studies suggest that model performance is often robust to moderate misspecification of this parameter (Hartig, 2020).

Zero-inflated beta models

For proportional data with values constrained to the interval [0,1], such as discard rates expressed as proportions of total catch, zero-inflated beta (ZIB) models offer an alternative approach to handling excess zeros. ZIB models combine a beta distribution for continuous proportions with a point mass at zero, modeling the data-generating process as a mixture distribution (Simpson, 2018). Unlike the hurdle model approach, which conditions the continuous component on the binary outcome, ZIB models treat zeros as arising from either the beta distribution or a separate zero-generating process. This distinction reflects a different conceptualization of the zero-generating mechanism, potentially more appropriate when zeros can arise from multiple ecological processes (Borcard et al., 2004).

The continuous component of the ZIB model follows a beta distribution, defined on the open interval $(0, 1)$. This component governs the behavior of the positive, non-zero proportions. Its probability density function is given by:

$$f_\beta(y;\mu,\phi) = \frac{y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}}{B(\mu\phi,(1-\mu)\phi)}, \qquad 0 < y < 1,$$

where: ($\mu \in (0,1)$) is the conditional mean of the distribution, ($\phi > 0$) is the precision parameter, and ($B(a,b)$) denotes the beta function.

This specification implies that the variance of the response, conditional on being non-zero, is:

$$\mathrm{Var}(Y \mid Y > 0) = \frac{\mu(1-\mu)}{1+\phi}.$$

In the GAM framework, this model requires estimating separate linear predictors for π, μ, and potentially ϕ (Wood, 2017). ZIB models offer several advantages in fisheries discard analysis. They explicitly model the probability of structural zeros (e.g., no discards due to perfect selectivity) separately from sampling zeros that might arise from the beta process. This distinction may be particularly relevant in fisheries where different mechanisms can lead to zero discards, such as regulatory compliance, market conditions, or gear selectivity (Miller et al., 2019). However, implementing ZIB models involves greater computational complexity than hurdle models or Tweedie models. The mixture distribution structure requires specialized estimation approaches, and convergence issues can arise with sparse data or complex predictor relationships. Additionally, model diagnostics for ZIB models are less developed than for more common distributional families (Zuur et al., 2014). In discard analysis applications, ZIB models have shown particular utility when the research focus includes understanding the different processes that can generate zero observations.

For example, when studying the implementation of selective fishing technologies, ZIB models can help distinguish between zeros arising from successful selectivity and those from other operational or environmental factors (Pedersen et al., 2019). Model selection between hurdle, Tweedie, and ZIB approaches should be guided by both statistical fit criteria and ecological understanding of the discard-generating process. Comparative evaluation using information criteria such as AIC or BIC, coupled with residual diagnostics, can help identify the most appropriate modeling framework for a given dataset and research question (Zuur and Ieno, 2016).

Capítulo 3

Materials and Methods.

3.1. Study Area

The study area corresponds to the Cantabrian-Northwest fishing grounds (Figure 3.1), which primirly encompass the outer continental shelf and upper slope, at depths ranging from 100 to 600 meters. This area is characterized by muddy and sandy bottoms with some rocky areas that provide important habitats for demersal and benthic species, particularly European hake.Stock and fisheries are assessed by the ICES.

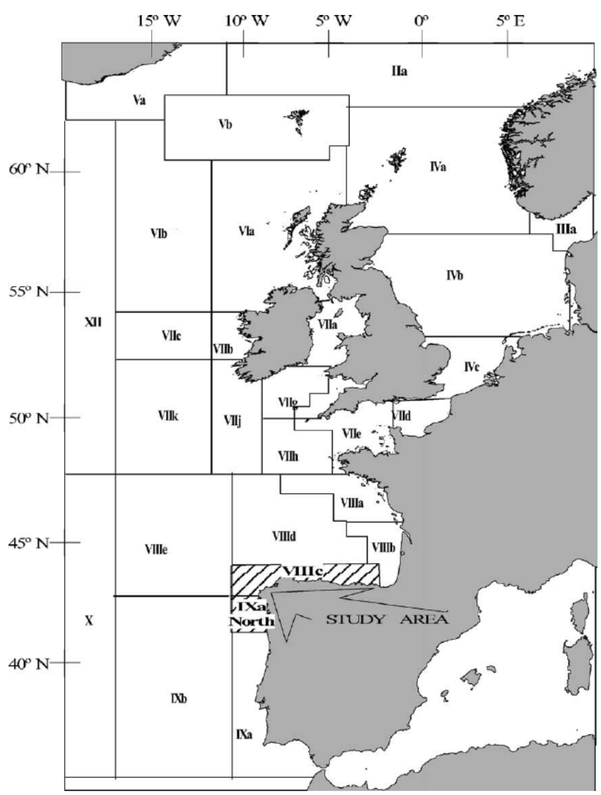

Figura 3.1: West Europe. Dashed divisions (ICES 8c and 9a) are the location of the study area.

3.2. Sampling design and data source

The data analyzed in this study were obtained through the Spanish scientific at-sea sampling program (3.2), coordinated by the Spanish Institute of Oceanography (IEO-CSIC) in collaboration with the Spanish Ministry of Fisheries. This program systematically collects information on fishing activity, including operational characteristics (such as speed or fishing duration), environmental and spatial conditions (such as depth or fishing location), biological data (e.g., species caught), and vessel-level information.

Figura 3.2: Trawling vessel with nets deployed in operation and observer on board.

The dataset covers the period from January 2009 to December 2023. The selected period also coincides with the formal implementation of the European Union's Data Collection Framework (DCF), which standardized the classification of fishing activities into so-called metiers—essentially fishing activity types defined by gear used, target species, and fishing area.

Sampling followed a multi-stage design common in observational fisheries data. At each level, vessels were first selected, then fishing trips (defined as a complete outing at sea), and finally individual fishing operations (hauls). For each haul (Figure: 3.3) , the catches were classified into retained (landed) and discarded fractions, and samples were collected accordingly.

Until 2015, the selection of vessels was largely based on voluntary participation. Since 2016, a probabilistic design has been introduced, where vessels are randomly selected from the official fleet registry, and the first subsequent trip is sampled. If the selected vessel refuses to carry an observer onboard, this is recorded and accounted for as a refusal.

This study focuses on a single metier, identified as "baca" gear, which refers to bottom trawl operations a mixture of demersal species,such as hake, monfish or megrim. The "baca" trawl (Figure 3.4) is a type of fishing gear consisting of a large, funnel-shaped net that is dragged along the seabed.

It is widely used in demersal fisheries due to its efficiency in capturing species that live near or on the seafloor, such as European hake. However, it is also known for its relatively high discard rates, since the gear is non-selective and captures a wide range of species and sizes, many of which are not commercially valuable or subject to fishing restrictions. For this analysis, the focus will be specifically on European hake (Merluccius merluccius) discards.

The selected time period and area aim to capture seasonal and long-term changes in fishing behavior and discarding practices, with special attention to the influence of the LO policy implemented under the European CFP, which progressively banned discards at sea.

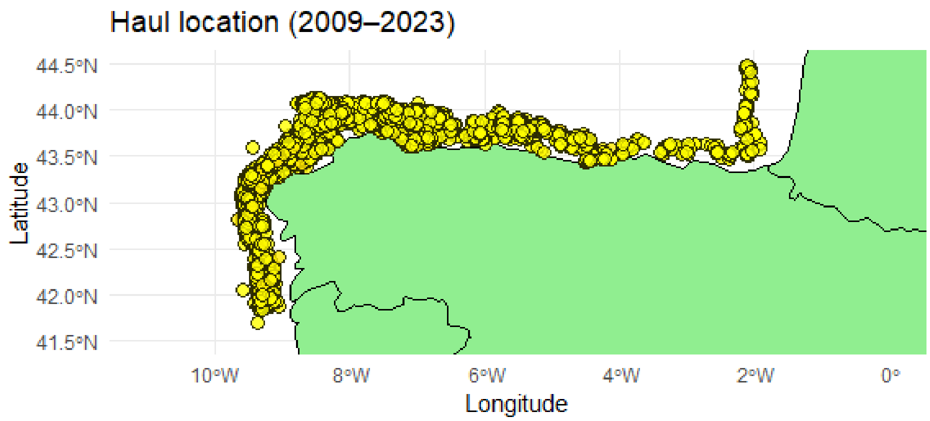

Figura 3.3: Spatial distribution of the sampled hauls, collected by Observers-at-sea for the period 2009-2023 in the Cantabrian-Northwest fishing ground, ICES 8c and 9a.

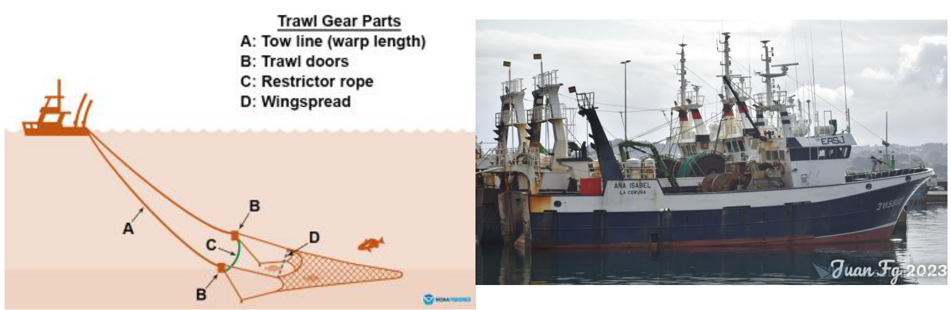

Figura 3.4: Trawling vessel gear composition and observer on board.

3.3. Variables and preprocessing

The initial dataset comprised 2811 hauls and 90 variables, which underwent a systematic cleaning and validation process prior to analysis, following the recommendations by Zuur and Leno (2016). Before modeling, a comprehensive data cleaning and consistency verification process was performed to ensure the internal coherence and analytical reliability of the data. This included validation of fishing effort records, spatiotemporal consistency, and discard estimates. Only after confirming the completeness and integrity of the cleaned dataset was the variable selection process undertaken.

This selection followed a two-stage process aimed at enhancing analytical relevance and model performance:

In the first stage, variables considered theoretically irrelevant or redundant were excluded based on ecological, operational, or statistical reasoning. These included:

- Descriptive or administrative fields unrelated to discard processes, such as species code and species name.
- Redundant vessel characteristics, such as horsepower or gross tonnage, already encapsulated in the vessel identifier.
- Variables with little or no variability or theoretical significance, such as “Net size” or “Name in English”.

The goal in this stage was to retain only variables with clear relevance to ecological patterns or fishing activity.The variables that were removed in this stage can be seen in the table

In the second stage, a set of variables that were initially deemed relevant were removed after exploratory analysis due to:

- A high proportion of missing values (e.g., wind speed, vertical net opening).
- Internal inconsistencies or lack of statistical contribution in preliminary models.
- Low variability despite theoretical relevance, like in the case of net size where over 90 % of the data

  had the same value.

These decisions are documented in table A.1, which outlines both the excluded variables and the rationale behind their removal.

Finally, new variables were constructed from existing fields or external sources to enrich the dataset and improve interpretability. These included:

- **Recruitment indices of European hake** (Merluccius merluccius) provided by the outputs of ICES stock assessments (ICES, 2024).
- **Distance to coast**, computed using spatial coordinates of each haul.
- **Discard rate**, calculated by dividing the total weight discarded and total weight catch.
- **Discard presence**, calculated by computing 0 if the discard rate is 0 and 1 if is positive.
- **Period**, a categorical variable where the data is divided into two time periods, before (2009-2015) and after (2016-2023) LO implementation on hake.

As shown in table A.2, the final set of variables selected included environmental, operational, and vessel-specific factors.

The main response variable in this study is the discard rate, which is defined as the proportion of the weight discarded over the total weight captured in each fishing operation:

$$\text{Discard rate} = \frac{\text{Weight discarded (kg)}}{\text{Total weight caught (kg)}} \tag{3.1}$$

This metric is bounded within the interval $[0, 1]$ and provides a direct measure of discard intensity, consistent with methodological approaches used in previous fisheries research (Catchpole et al., 2005; Depestele et al., 2011,Berg et al .,2022). Given its proportional nature, discard rate represents a standardized measure that facilitates comparison across different fishing operations, vessel types, and temporal periods, making it particularly useful for evaluating potential effects of management interventions such as the Landing Obligation regulation.

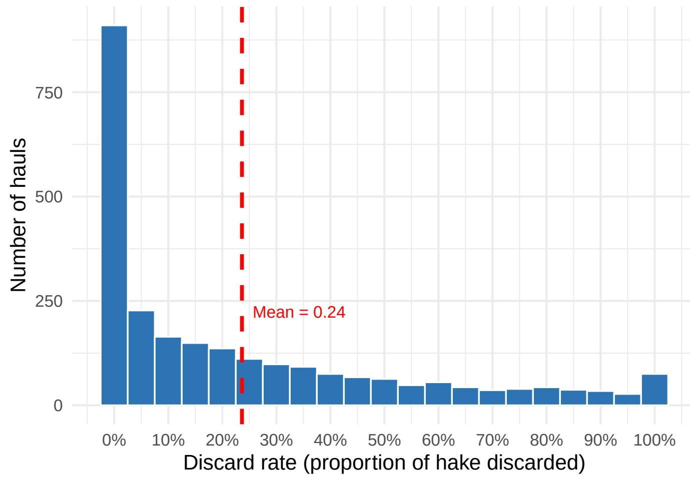

Figura 3.5: Distribution of hake discard rate (n = 2,569 hauls).

A preliminary distributional analysis(Figure 3.5) revealed that the empirical distribution of discard rates exhibits two distinct statistical features:

- A substantial point mass at zero (34 % of observations), representing fishing operations with no discarding.
- A continuous component on the interval (0,1) for operations where discarding occurred.

In datasets of discards data„ zeros are typically structural rather than structuralzeros differ fundamentally from left-censoring or detection-limit situations, where zeros arise because observations fall below a threshold. Modelling both the binary decision to discard and the continuous proportion discarded with a single continuous distribution violates key statistical assumptions and risks biased inference, so a two-part (hurdle) framework that separates these processes is more appropriate. Traditional single-equation approaches for proportion data—such as logit or arcsine transformations—handle true zeros poorly and require arbitrary constants, introducing additional bias and obscuring interpretation.

3.4. Methods

1) Two-stage modelling approach

To identify the factors that drive discarding behaviour—and given the two-stage nature of the process revealed by the data—a two-part (hurdle) modelling framework was implemented,both were described in section 2.4:

- Binomial model: Predicts the probability of discard occurrence (presence/absence) described in section 2.4.1.
- Beta regression model: Models the proportion of discards when discarding occurs, accommodating the bounded nature of proportion data (section 2.4.1).

For the binomial model the analysis began with a full model specification that includes all relevant coviariates:

$$\begin{aligned}\text{logit}(p) = \beta_0 &+ \beta_1 \cdot \text{Light} \\ &+ s_1(\text{Haul\_Duration}) + s_2(\text{Speed}) + s_3(\text{MeanDepth}) \\ &+ s_4(\text{Recruitment}) + s_5(\text{DistanceCoast}) \\ &+ s_6(\text{Month}) + s_7(\text{Hour}) + s_8(\text{Longitude}, \text{Latitude}) \\ &+ b_1(\text{Vessel}) + b_2(\text{Home\_Port})\end{aligned} \tag{4.1}$$

where p represents the probability of discarding occurrence,β_0,β_1 are parametric coefficients, and $s_1, ..., s_8$ represent smooth functions estimated using penalized regression splines and b_1, b_2 random effects of categorical variables.

In the case of the model for the proportion discard the full model is:

$$\begin{aligned}\text{logit}(\mu) = \beta_0 &+ \beta_1 \cdot \text{Light} \\ &+ s_1(\text{Haul\_Duration}) + s_2(\text{Speed}) + s_3(\text{MeanDepth}) \\ &+ s_4(\text{Recruitment}) + s_5(\text{DistanceCoast}) \\ &+ s_6(\text{Month}) + s_7(\text{Hour}) + s_8(\text{Longitude}, \text{Latitude}) \\ &+ b_1(\text{Vessel}) + b_2(\text{Home\_Port})\end{aligned} \tag{4.2}$$

where μ represents the expected proportion of discards conditional on discarding occurring,β_0,β_1 are parametric coefficients, and $s_1, ..., s_8$ represent smooth functions estimated using penalized regression splines and b_1, b_2 random effects of categorical variables.

From an operational perspective, this two-stage approach reflects the actual sequence of choices made by fishers. The initial decision—whether to retain all catch or discard some portion—is fundamentally categorical and likely influenced by factors such as the presence of undersized individuals, market conditions, and quota limitations. The subsequent decision of how much to discard is a continuous choice influenced by catch characteristics, vessel capabilities, and economic considerations.

The discarding decisions may be influenced by different factors at each stage, or by the same factors operating through different relationships. For example, the decision to discard might be primarily driven by regulatory constraints, while the amount discarded could be more strongly influenced by the size distribution of the catch or storage constraints.

This modeling approach has a well-established tradition in fisheries science for analyzing data with similar structures (Stefánsson, 1996; Maunder and Punt, 2004). Studies comparing various modeling approaches for similar data have consistently found that two-part models provide better predictive accuracy and more meaningful interpretations than alternative approaches (Lecomte et al., 2013; García-de-Leaniz et al., 2021).

Both components of the model were implemented using Generalized Additive Models (GAMs) to capture the potentially complex non-linear relationships between operational, environmental, and spatio-temporal factors and discard patterns. The models incorporated vessel-specific random effects to account for unobserved heterogeneity at the vessel level.

2) Adding period-specific smooth terms to each GAM

A key objective of this analysis is to identify changes in discard behavior before and after the implementation of the LO. To refine the analysis, we augmented each GAM—both the binomial and beta components—with period-specific smooth terms with the by = period syntax. The categorical variable period splits the time-series into two phases: Pre-LO (2009-2015) vs Post-LO (2016-2023) The year 2016 is set as the breakpoint because that is the year in which the LO was fully implemented for European hake (*Merluccius merluccius*), although the regulation became fully applicable to all species under the discard plan in 2019. This approach allowed the estimation of separate non-linear relationships for each temporal period, enabling the detection of changes in how discard patterns relate to covariates across time.

This specification enables the separate estimation of the covariates effects for each period, making it possible to evaluate whether the relationships changed after the implementation of the LO.

Period-by-period smooths offer a robust and interpretable way to visualize and quantify these changes (Pedersen et al., 2019), especially when interactions between temporal context and ecological or operational processes are suspected.

The model for changes in probability before and after LO is:

$$\begin{aligned} logit(p) &= \beta_0 + \beta_1 \cdot Period + s_1(HaulDuration \mid Period) \\ &+ s_2(Speed \mid Period) \\ &+ s_3(MeanDepth \mid Period) + s_4(Recruitment \mid Period) \\ &+ s_5(Month \mid Period) + s_6(Longitude, Latitude \mid Period) \\ &+ b_1(Vessel) + b_2(HomePort) \end{aligned} \tag{4.3}$$

Where p represents the probability of discarding occurrence,β_0 and β_1 are parametric coefficients,$s1, ..., s6$ represent smooth functions estimated using penalized regression splines, | indicates by-factor interaction with period and b_1, b_2 the random effect of categorical variables.

The model for changes in proportion before and after LO is:

$$\begin{aligned} logit(\mu) &= \beta_0 + \beta_1 \cdot Period + s_1(HaulDuration \mid Period) \\ &+ s_2(Speed \mid Period) \\ &+ s_3(MeanDepth \mid Period) + s_4(Recruitment \mid Period) \\ &+ s_5(Month \mid Period) + s_6(Longitude, Latitude \mid Period) \\ &+ b_1(Vessel) + b_2(HomePort) \end{aligned} \tag{4.3}$$

Where μ represents the probability of discarding occurrence,β_0 and β_1 are parametric coefficients,$s1, ..., s6$ represent smooth functions estimatndefineded using penalized regression splines, | indicates by-factor interaction with period and b_1, b_2 the random effect of categorical variables.

This model specification includes separate smooth functions for each continuous predictor conditional on period (before/after LO), allowing for different functional forms in each period. The implementation follows the factor-smooth interaction approach described by Wood (2017), which enables statistical comparison of smooth effects between groups or periods in our case , as recommended by Augustin et al. (2013) for spatio-temporal modeling of fisheries data.

To formally evaluate whether the estimated smooth effects of covariates differed significantly between periods, we computed difference smooths using the `difference_smooths()` function from the `gratia` package. This procedure calculates the pointwise difference between period-specific smooth terms (pre-LO minus post-LO), accompanied by simultaneous 95 % confidence intervals. A covariate was considered to exhibit a significant temporal change in its effect when the corresponding confidence band excluded zero across a meaningful portion of its domain.

3.4.1. Model selection and validation

Binomial component

Model selection for the binomial GAM began with a fully saturated specification—twelve candidate predictors comprising three categorical factors entered linearly and nine continuous covariates entered as smooth functions—and proceeded by backward elimination. Terms lacking statistical support ($p > 0.05$) were removed unless ecological reasoning justified their retention; smooths whose effective degrees of freedom approached one were re-specified as linear effects; and basis dimensions were increased whenever the k-index indicated undersmoothing. Competing models were ranked with Akaike and Bayesian information criteria, while deviance explained and adjusted R^2 were tracked to ensure that parsimony was not achieved at the expense of explanatory power. The ultimately retained structure contains one linear effect (Haul Duration), six univariate smooths (Speed, Mean Depth, Recruitment, Distance-to-Coast, Month, Hour), a bivariate spatial smooth (Longitude–Latitude), and random intercepts for Vessel and Home Port; the complete list of retained terms is given in Table A.3 (see Results). Model adequacy was finally checked with residual QQ-plots, residual-versus-fitted diagnostics, concurvity measures, and—given the binary response—ROC curves and their associated Area-Under-the-Curve statistic to summarise discriminatory performance.

Beta component

Model selection for the beta-regression GAM followed a similar strategy to the one applied for the binomial model. Starting with a full model specification, we included all potentially informative variables and progressively simplified the structure using a stepwise approach. Non-significant terms ($p > 0.05$) were excluded unless supported by ecological relevance. Smooth terms with estimated effective degrees of freedom (EDF) close to one were reparametrized as linear effects, and the basis dimension was increased whenever the k-index indicated undersmoothing.

The retained structure includes three linear effects (Haul Duration, Distance to Coast, Intercept) and four smooth terms (Recruitment, Month, Hour, Vessel). A summary of their estimates, standard errors, EDF values, and associated significance tests is presented in TableA.4. All retained terms were statistically significant at the 0.05 level, and the smooth terms—especially Recruitment and Vessel—exhibited high EDF values and highly significant chi-squared statistics, indicating substantial non-linear structure.

Final model validation included diagnostic checks of residuals, EDF values, and concurvity diagnostics to ensure robust estimation. As the beta regression is sensitive to boundary values, we also verified that response values were appropriately constrained to the (0, 1) interval, ensuring model assumptions were met.

Capítulo 4

Results

4.1. General Drivers of discarding

4.1.1. Probability of discard

The comparison of three candidate binomial GAM fits for predicting the presence or absence of hake discards is presented in Table 4.1.The initial specification (Fit 1) contained twelve predictors—three categorical factors entered linearly and nine continuous covariates entered as smooth functions. In Fit 2 the non-significant factors *Light* and *Home Port* were removed, improving parsimony without appreciable loss of deviance explained. Fit 3 introduced two further adjustments: (i) *Haul Duration*, whose estimated edf was 1, was re-specified as a linear effect, and (ii) basis dimensions were increased for smooths whose k-index indicated undersmoothing.

These changes reduced the AIC from 2413 (Fit 1) and 2413 (Fit 2) to 2393 (Fit 3); the > 20-unit drop provides strong support for Fit 3 over the alternatives (Burnham and Anderson, 2002). BIC echoed this result while rewarding the lower effective degrees of freedom. Adjusted R^2 and deviance explained declined by only 0.02 percentage points relative to the most complex model, confirming that the simplifications achieved parsimony at negligible cost in explanatory power. Consequently, Fit 3 was retained for all subsequent inference and visualization.

The final fit combines one linear effect, six smooth effects and one random effect.The smooth terms include key continuous covariates (speed,mean depth,recruitment, distance to coast),seasonal and daily variation (month and hour), spatial position and vessel specific random effects.Together these components provide a comprehensive and interpretable representation of the drivers of discarding behavior.

The detailed term-level diagnostics for the retained specification—including significance levels and effective degrees of freedom—are reported in table 4.2.

Figure 4.1 shows the estimated smooth functions for the significant continuous terms in the final binomial GAM model.The effect of speed (Figure A) reveals a distinct non-linear pattern. At lower velocities, up to approximately 2.75 knots, the log-odds of discard are negative, indicating a reduced probability of discard in this range. Between 2.75 and 3.2 knots, the log-odds increase, suggesting a higher likelihood of discards. Beyond 3.2 knots, the effect diminishes, and the probability of discard decreases once more. The smooth function for mean depth (Figure B) shows a broad and gradual U-shaped trend. The estimated effect starts at a positive value at lower depths, then decreases smoothly until it reaches a minimum around 500 meters. From this point onward, the curve begins a slight upward trajectory, eventually flattening out at greater depths. The shape is symmetric and smooth, with relatively narrow confidence intervals across most of the depth range, though they begin to widen slightly at the deepest values. The smooth function for recruitment (Figure C) displays a wave-like pattern. The estimated effect begins near zero, rises to a local maximum around 200,000 units, then descends to a local minimum around 280,000. From there, it increases again, reaching a second peak

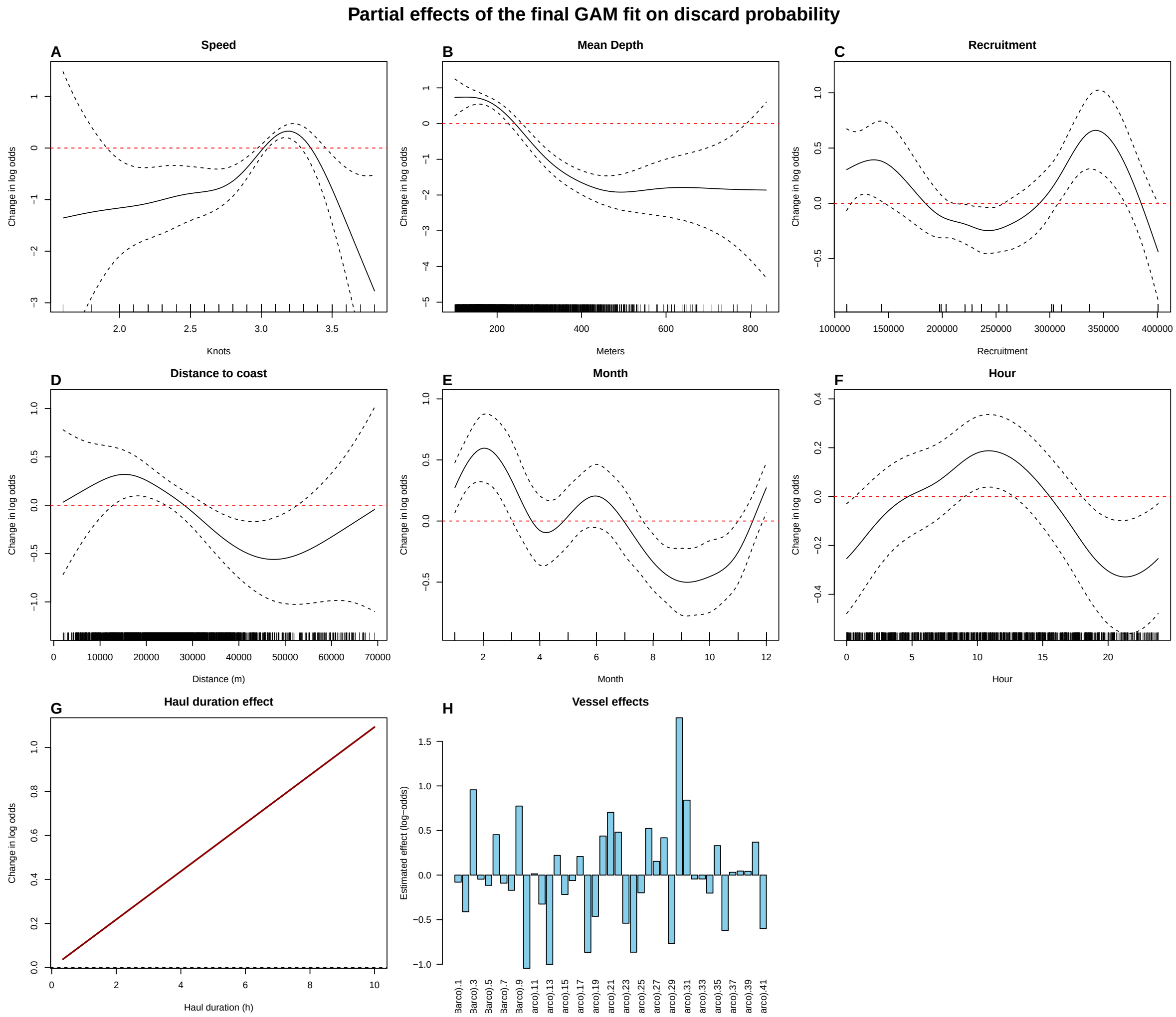

Figura 4.1: Effects of predictors over the discard probability of the Spanish bottom otter trawl flee operating in Spanish North-western waters during the 2009-2023 period.(Binomial model).A) Speed effect, B) Mean depth effect,C) Recruitment effect,D) Distance to the coast effect,E) Month effect,F) Hour effect,G) Haul duration effect and H) Vessel effect .

Cuadro 4.1: Statistical comparison of three BINOMIAL GAM models

Metric	Fit.1	Fit.2	Fit.3
AIC	2413.304	2412.856	**2392.762**
BIC	2899.532	2889.307	**2815.030**
Degrees of freedom	77.88	76.12	**72.36**
R-squared adjusted	0.328	0.327	**0.326**
Deviance explained	28.8 %	28.7 %	**28.6 %**
REML	1246.7	1247.2	**1248.6**
Sample size	2529	2529	**2529**

Cuadro 4.2: Summary of fixed and smooth terms in the Fit 3 GAM (binomial family, logit link)

Effect.Type	Term	Estimate	Std.Error	EDF	Chi.sq	P.value
Linear effects	Intercept	0.691	0.247	-	-	0.0051
Linear effects	Haul duration	0.109	0.055	-	-	0.0487
Smooth terms	Speed	-	-	4.665	59.05	p<2e-16
Smooth terms	Mean depth	-	-	4.474	181.75	p<2e-16
Smooth terms	Recruitment	-	-	5.238	43.49	p<2e-16
Smooth terms	Distance to coast	-	-	3.356	13.63	0.00105
Smooth terms	Month	-	-	5.371	48.84	p<2e-16
Smooth terms	Hour	-	-	2.621	10.94	0.00399
Smooth terms	Position	-	-	13.176	226.21	0.00429
Smooth terms	Vessel	-	-	24.492	92.70	p<2e-16

near 350,000 before sharply dropping toward the right end of the range. The confidence intervals are relatively symmetric and show moderate widening at both extremes. The smooth function for distance to coast (Figure D) exhibits a complex, undulating shape. It begins with a relatively flat segment at short distances, followed by a sharp rise around 10,000–20,000 meters. The curve then dips again near 40,000 meters, rises to a secondary peak near 60,000, and drops sharply afterward. The pattern is non-linear with multiple inflection points. Confidence intervals remain relatively narrow throughout. The smooth for month (Figure E) shows a cyclic, sinusoidal pattern. The curve rises steadily from January, peaking around May, and then declines to a trough near October. It begins to rise again toward the end of the year. The periodic nature is consistent with the use of a cyclic cubic spline.

Confidence intervals remain narrow and symmetric across the entire time span. The smooth function for hour (Figure F) has a bell-shaped structure. The effect increases gradually from midnight, peaks between 8:00 and 10:00, and then decreases smoothly toward the evening. A slight rise is visible in the final hours of the day. The function is smooth and symmetric, with tight confidence bands in the central part of the range and moderate widening at the extremes. Panel G displays a linear effect of haul duration. The plot shows a perfectly straight, ascending line with no curvature, as expected from modeling the term as a linear predictor. There is no shaded confidence band due to the parametric nature of the term. Panel H presents the estimated random effects for individual vessels. Each bar represents a vessel-specific deviation from the overall intercept on the log-odds scale. The bars are centered around zero, with varying magnitudes and directions.

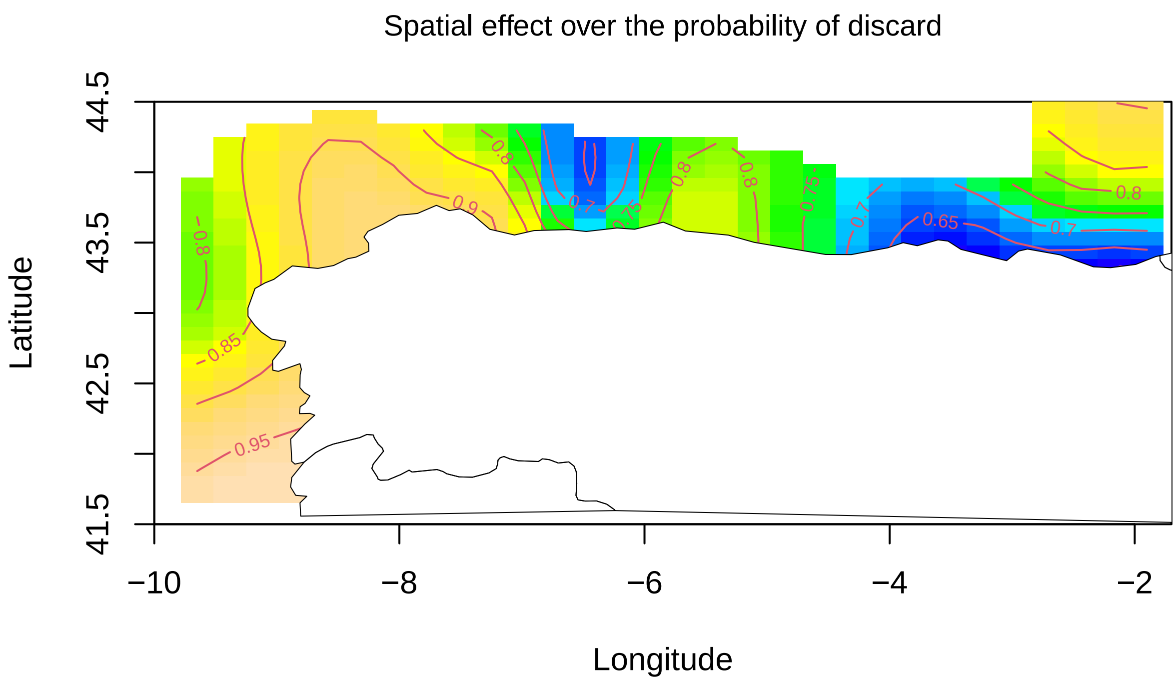

Figura 4.2: Effects of spatial effect over the discard probability of the Spanish bottom otter trawl flee operating in Spanish North-western waters during the 2009-2023 period.Values shown are predicted probabilities (response scale: 0 = 0 %, 1 = 100 %)

The spatial effect (Figure 4.2) reveals considerable geographic heterogeneity across the study area, with lower predicted probabilities (blue areas) concentrated in the eastern coastal regions and greater probabilities (yellow-green areas) in the central and western offshore areas. The contour lines indicate smooth transitions between different probability zones.

Validation

As noted in Section 3.3, each GAM must undergo a rigorous evaluation to determine whether its specification is appropriate.

The ROC curve analysis (Figure 4.3) demonstrates that the Binomial GAM model exhibits strong discriminatory performance for predicting the outcome variable. With an Area Under the Curve (AUC) of 0.84, the model shows good predictive ability, as this value falls well above the threshold of 0.5 that would indicate random classification and approaches the ideal value of 1.0.

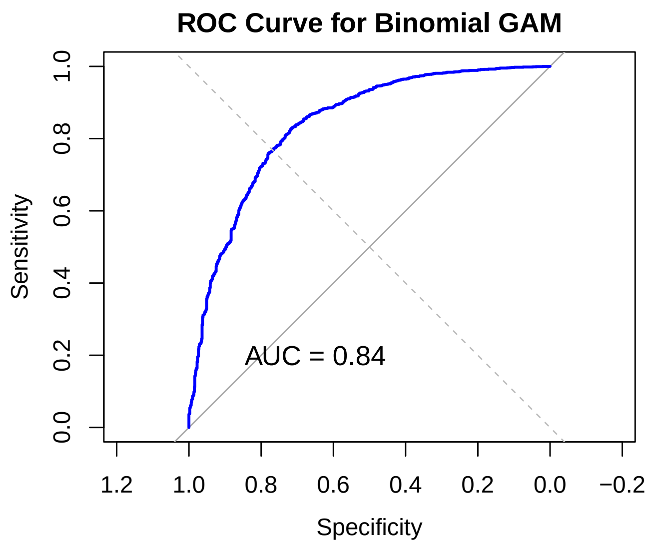

Figura 4.3: ROC curve for the Binomial GAM model.

The model was validated using a data set containing 2,529 observations, with 1,733 cases (positive outcomes, desc_pres = 1) and 796 controls (negative outcomes, desc_pres = 0). This represents approximately a 2.2:1 ratio of cases to controls, indicating a moderately imbalanced data set that favors positive outcomes. The ROC curve's pronounced bow above the diagonal reference line (which represents random chance) confirms that the model's predicted probabilities effectively distinguish between the two outcome groups across various threshold settings.

The AUC value of 0.84 suggests that there is approximately an 84 % probability that the model will correctly rank a randomly selected positive case higher than a randomly selected negative case in terms of predicted probability. This level of discrimination indicates that the GAM successfully captures the underlying relationships between the predictor variables and the binary outcome, making it a reliable tool for classification purposes. The smooth, well-defined curve without erratic fluctuations also suggests model stability and robust performance across different sensitivity and specificity trade-offs.

4.1.2. Proportion discard.

Table 4.3 presents the results for the three candidate beta models. Once again, the third fit delivers the strongest performance: it reports the lowest AIC and BIC, reflecting the most favorable balance between goodness-of-fit and model parsimony. Although this model employs a larger number of effective degrees of freedom, the gain in explanatory power justifies the added complexity

Table 4.4 summaries the term-level diagnostics for the three beta-regression fits. The refinement protocol mirrored that used for the binomial analysis. We began with a fully saturated model in which every candidate predictor entered as a smooth term. Non-significant effects were then pruned, smooths whose effective degrees of freedom approximated one were recast as linear terms, and basis dimensions were increased whenever the k-diagnostic signaled under-fitting. This iterative procedure once again converged on a third model that delivers the best trade-off between explanatory power and parsimony.

Cuadro 4.3: Statistical comparison of three BETA-regression GAM models

Metric	**Fit.1**	**Fit.2**	**Fit.3**
AIC	-929.5123	-942.2149	**-1036.2174**
BIC	-572.9420	-612.3799	**-671.1948**
Degrees of freedom	65.81726	60.88234	**68.54863**
R-sq adjusted	0.203	0.204	**0.259**
Deviance explained	26.6 %	26.8 %	**32.4 %**
REML	-428.95	-429.5	**-452.32**
Sample size	1665	1665	**1665**

Cuadro 4.4: Summary of fixed and smooth terms in the Fit 3 GAM (Beta regression, logit link)

Effect.Type	Term	Estimate	Std.Error	EDF	Chi.sq	P.value
Linear effects	Intercept	-0.29900	0.13420	-	-	0.02591
Linear effects	Haul duration	-0.06637	0.02849	-	-	0.01985
Linear effects	Distance to coast	-0.00000653	0.00000241	-	-	0.00681
Smooth terms	Recruitment	-	-	11.697	423.93	<2e-16
Smooth terms	Month	-	-	4.505	24.39	0.00504
Smooth terms	Hour	-	-	3.266	32.36	<2e-16
Smooth terms	Vessel	-	-	28.156	176.11	<2e-16

The final model fitted can be seen in Table 4.4 were the variables selected are the haul duration, distance to the coast,recruitment, month of the year, hour of the day and the vessel.The first two act as a linear effect and the rest are smooth effects.The over significance of the variables is high, meaning that they explained some part of the data.

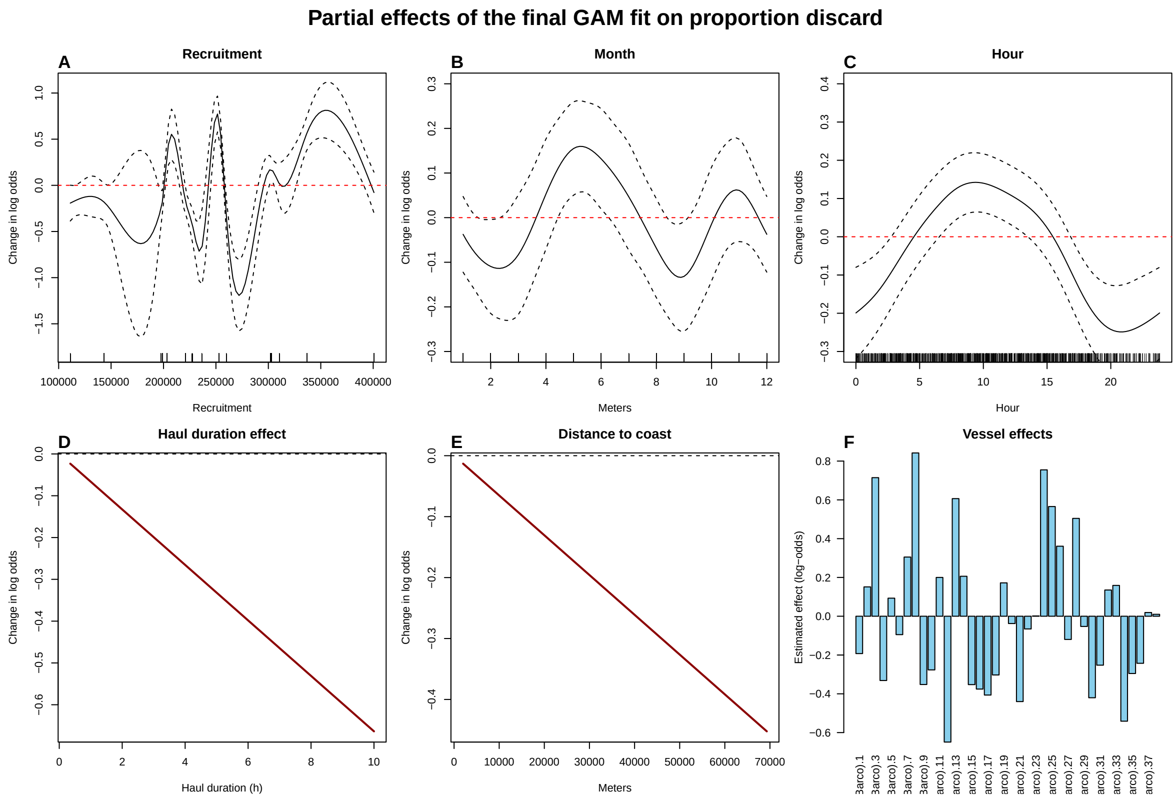

Figura 4.4: Effects of predictors on discard proportion (Beta GAM model) of the Spanish bottom otter trawl flee operating in Spanish North-western waters during the 2009-2023 period.(Binomial model).A) Recruitment effect, B) Month effect, C) Hour effect ,D) Haul Duration effect ,E) Distance to coast effect ,F) Vessel effect

Figure 4.4 shows the estimated smooth functions for the significant continuous terms,the slope of linear terms and the difference between vessels in the final beta GAM model for discard proportion.

Panel A depicts the smooth effect of the recruitment index. The solid line shows the estimated change in log-odds across the observed range of recruitment, while the dashed lines mark the approximate 95 % confidence envelope; individual observations are indicated by tick marks along the rug at the plot base.

Panel B presents the smooth for month. The response surface exhibits two broad peaks and a mid-range trough, with confidence limits widening at the distribution tails.One of the peaks happens to be on spring and the other one by the end of the year.

Panel C illustrates the cyclic smoother for hour of day. The curve rises towards mid-morning, flattens around midday and declines in the late afternoon; the density of tick marks demonstrates that the data cover the entire diel cycle.

Panel D shows the linear effect of haul duration (h). The fitted line is strictly decreasing, and the narrow confidence band (largely overlapping the estimate) indicates precise estimation across the sampled range.

Panel E displays the linear relationship with distance to coast (m). A nearly one-to-one negative slope is evident, again with a very tight confidence band.

Panel F summarises the random-effect estimates for individual vessels. Bars represent deviations from the overall intercept on the log-odds scale; positive bars identify vessels with higher estimated discard proportions, whereas negative bars indicate the opposite. The variation is visibly asymmetric, ranging from approximately -0.6 to $+0.8$ log-odds.

Validation

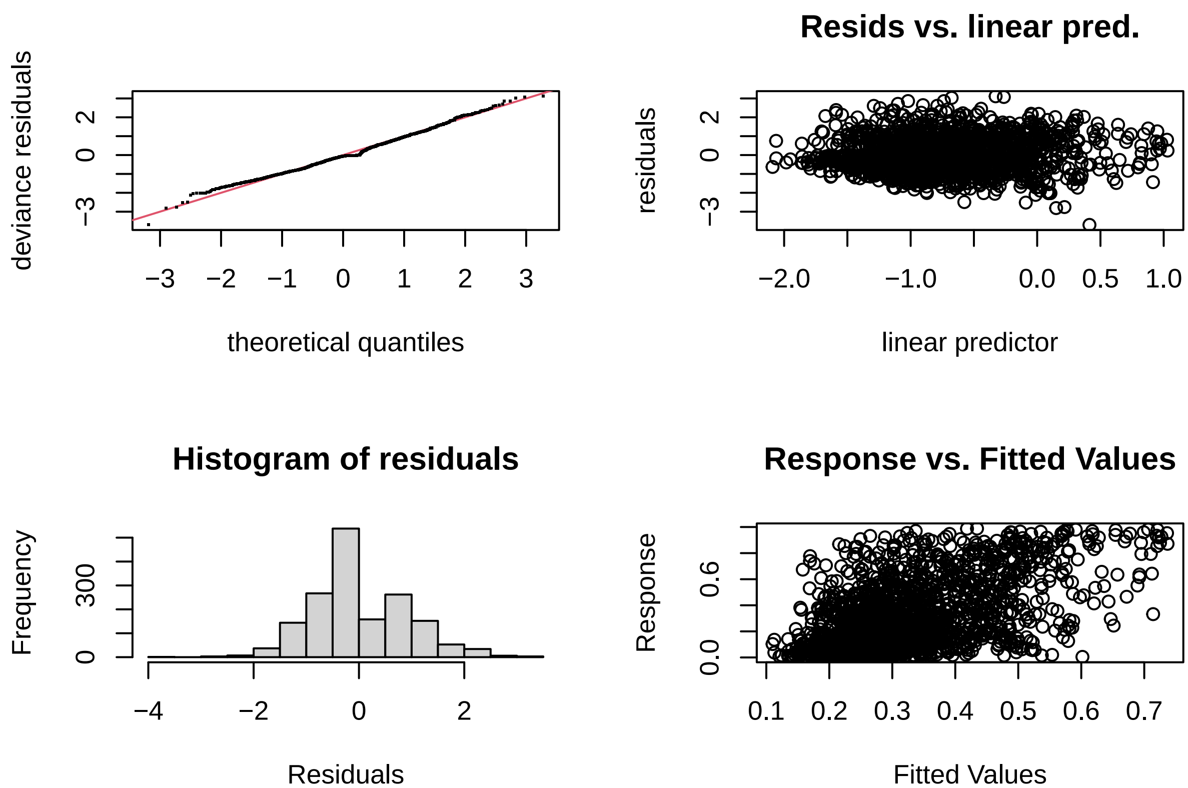

Figura 4.5: Validation plots for Beta GAM. Top left : Q-Q plot, Top right: Residuals vs linear predictors,Bottom left: Histogram of the residuals, Bottom right: Response against fitted values.

Figure 4.5 presents the diagnostic plots from gam.check() for the final beta GAM model validation. The four-panel diagnostic display shows overall satisfactory model performance with some areas for consideration.

The Q-Q plot (upper left) shows deviance residuals against theoretical quantiles, with points following the diagonal reference line reasonably well in the central portion but displaying some deviation at both extremes, particularly noticeable departures in the upper tail where points fall above the reference line, suggesting potential outliers or slight distributional misspecification.

The residuals versus linear predictor plot (upper right) displays deviance residuals scattered around zero across the linear predictor range from -2.0 to 1.0, showing relatively random scatter without strong systematic patterns, though some increased variability appears at intermediate predictor values around -1.0 to -0.5.

The histogram of residuals (lower left) reveals an approximately normal distribution of deviance residuals centered around zero with a slight right skew. Most residuals fall within the -1 to +1 range, with a small number of extreme values extending to approximately ± 3, consistent with the Q-Q plot findings. The response versus fitted values plot (lower right) shows observed discard proportions against model predictions, with fitted values spanning 0.1 to 0.7 and observed responses covering the full 0 to 1 range. The scatter pattern appears appropriate for proportion data, showing reasonable correspondence between fitted and observed values without obvious systematic bias, though some boundary effects are visible as expected for beta regression models.

The GAM equivalent of multicollinearity in linear models (concurvity) was evaluated to identify potential redundancy among predictors. The pairwise and worst-case concurvity were calculated using the concurvity() function in the mgcv package as recommended by Simpson (2018).The results are shown in table 4.5

Cuadro 4.5: Selected concurvity values between model predictors

Predictor_pairs	Concurvity_value
Speed - Depth	0.45
Depth - Distance to coast	0.68
Speed - Month	0.31
All terms (worst case)	0.72

The maximum observed concurvity value was 0.72, below the problematic threshold of 0.80 suggested by Wood (2017). This indicates that while some correlation exists between predictors, it is not severe enough to compromise model stability or interpretability. The moderate concurvity observed between depth and distance to coast (0.68) is expected due to the continental shelf bathymetry in the study area (Sánchez and Serrano, 2003).

4.2. Effect of the Landing Obligation

4.2.1. Probability of discarding

Cuadro 4.6: Summary of GAM with period interactions (binomial component)

Component	Value
Model type	Binomial GAM with logit link
Response variable	Presence/absence of discards
Sample size	2529
Deviance explained (%)	33.9 %
R-squared (adj)	0.371
REML score	-0.081643

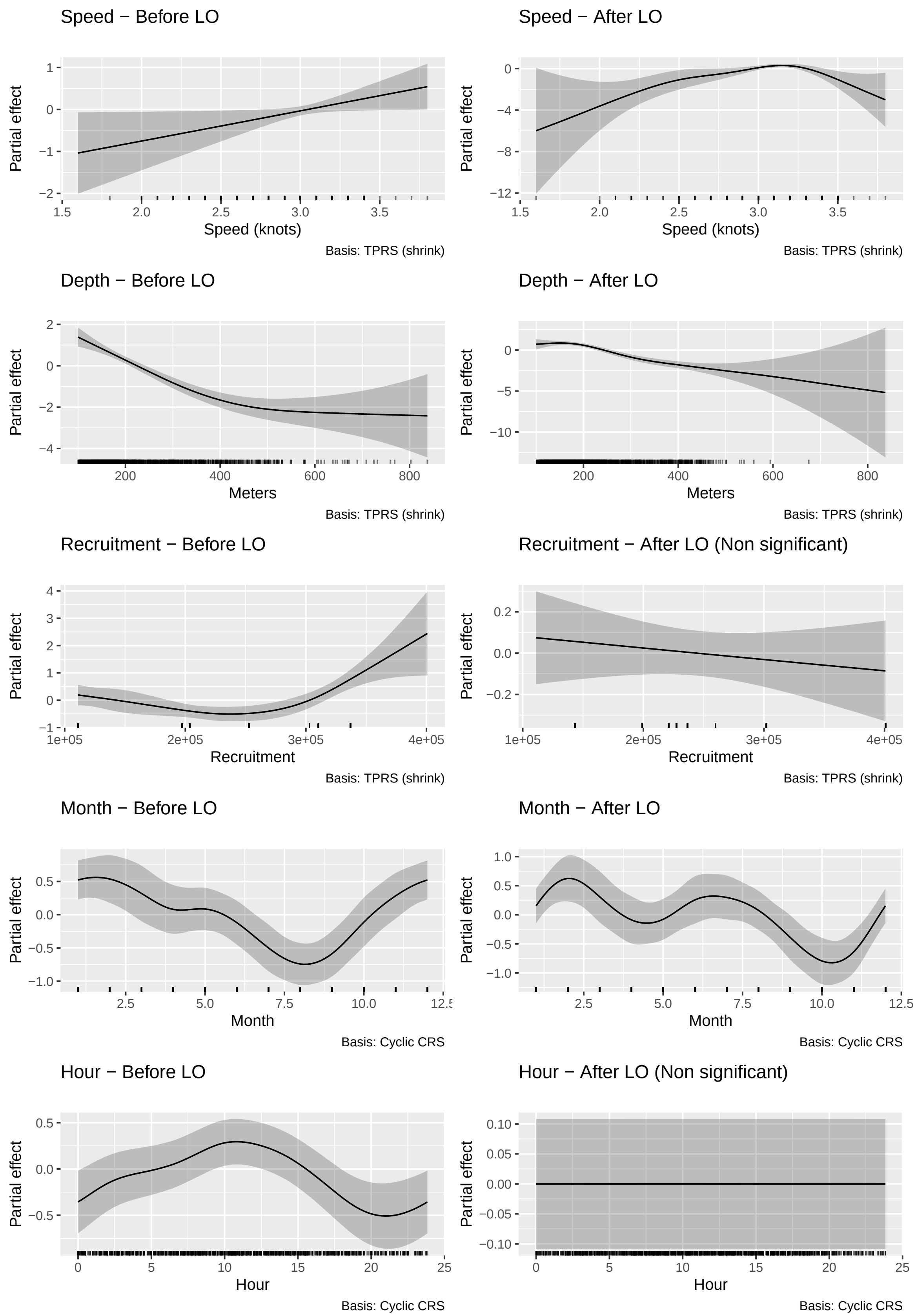

Figura 4.6: Effect of variables over the discard probability pre-LO and post-LO (Bionomial GAM using period interaction

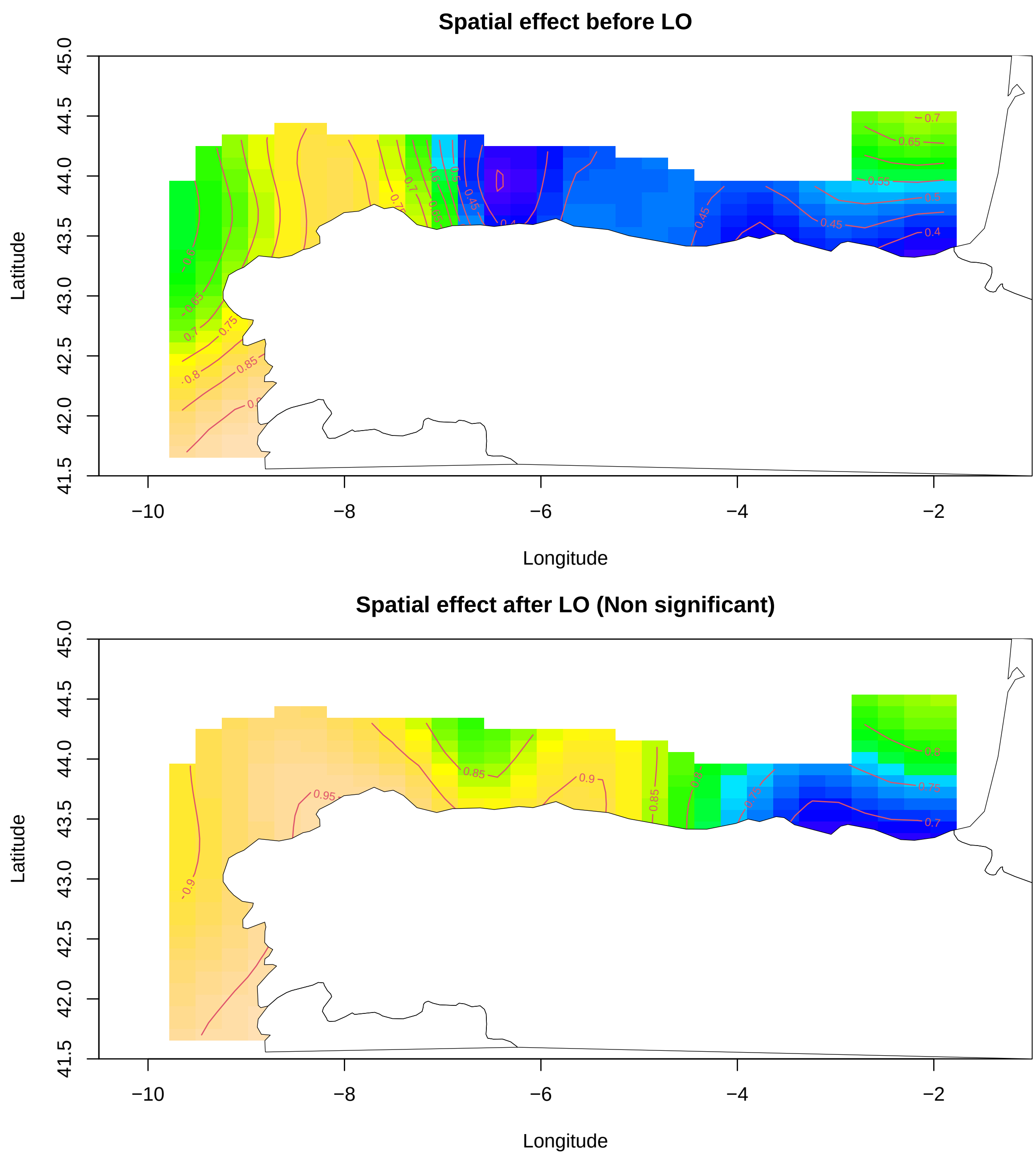

Figura 4.7: Effects of spatial effect over the discard probability before and after LO of the Spanish bottom otter trawl flee operating in Spanish North-western waters during the 2009-2023 period.

Cuadro 4.7: Smooth term significance and parametric coefficients (binomial component)

	Before LO		After LO	
Component	Edf	p-value	Edf	p-value
Intercept (Before LO)	0.6651	0.985		
Period effect (After LO)			0.4065	0.728
Haul duration	0.4115	0.149	0.00014	0.221
Distance to coast	3.340	0.103	4.894	0.0168
Speed	3.363	0.017	3.809	0.000962
Mean depth	2.504	<2e-16	4.703	<2e-16
Recruitment	5.601	2.8e-05	0.0934	0.237
Month	3.225	<2e-16	5.271	<2e-16
Hour	6.354	5.09e-05	0.00001	0.819
Spatial	11.98	3.53e-06	12.92	0.00465
Vessel effect	35.32	0.0087	35.32	0.008725
Base port	1.664	0.511	1.664	0.511
Landing port	12.00	0.0329	12.00	0.0329

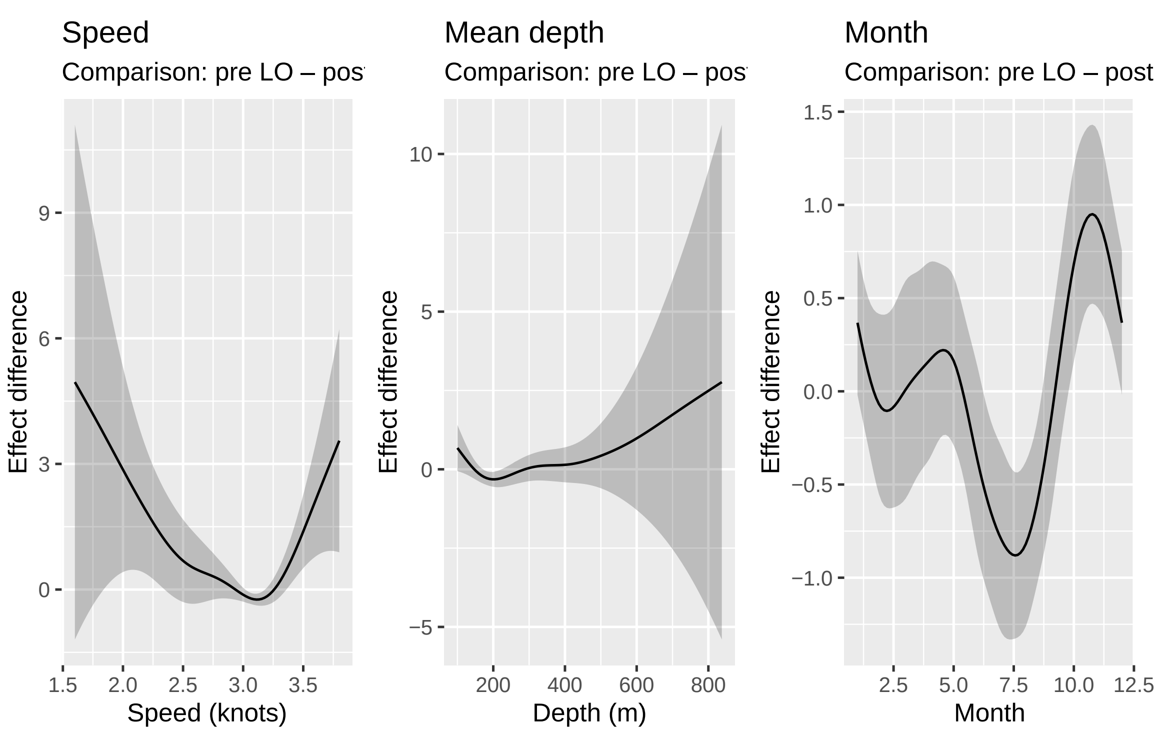

Figura 4.8: Difference in variable effects after and before LO for the probability of discard

Binomial models with period interactions (Tables 4.7, 4.6) showed no significant main effect of period when other covariates were held constant ( ).

The smooth term was highly significant in both periods (pre-LO post-LO). Pre-LO, discard probability increased approximately linearly with speed. Post-LO, the relationship was hump-shaped, peaking at ~3 kn and declining thereafter. The difference smooth was positive at speeds < 2.2 kn and > 3.3 kn and near zero at intermediate speeds.

Depth had a strong, non-linear effect in both periods. Discard probability decreased with depth in both cases, with a steeper decline pre-LO. The difference smooth was centred on zero across the full depth range.

Recruitment was significant only in the pre-LO period; discard probability rose sharply at high recruitment values pre-LO, while the post-LO smooth was flat. The difference smooth was positive only in the upper tail.

A marked seasonal pattern occurred in both periods. Minimum discard probabilities were in July pre-LO and in October post-LO. The difference smooth was positive in mid-winter and late autumn and negative from early to mid-summer.

The diel smooth was significant only in the pre-LO period . Discard probability peaked around midday and dipped in the evening pre-LO; post-LO, the smooth was flat. The difference smooth was positive from dawn to early afternoon and negative in the evening.

Spatial structure was strong pre-LO and marginal post-LO. Figure 4.7 shows a pre-LO hotspot (> 0.9 probability) on the western outer shelf that weakened slightly post-LO, eastward expansion of medium-probability areas, and enlargement of low-probability areas (< 0.4) along the mid-shelf, producing a smoother west–east gradient.

4.2.2. Proportion discarded

Cuadro 4.8: Summary of GAM with period interactions (Beta component)

Component	Value
Model type	Beta GAM with logit link
Response variable	Proportion discarded
Sample size	1665
Deviance explained (%)	26.6 %
R-squared (adj)	0.217
REML score	-448.32

The results of the beta-regression model with period interactions (Table 4.9) and overall performance (Table 4.8) summarize the main factors associated with variations in discard rates—rather than discard presence/absence—across the two periods before and after the implementation of the Landing Obligation (LO).The model has been fitted with 1665 observations and it explains 26.6 % of the deviance and has an adjusted r-squared of 0.217 Controlling for speed, depth, haul duration, and other covariates through period-specific non-linear smooth functions, the post-LO period shows 53 % lower odds of discarding (OR 0.47, $p < 0.001$), which translates into a drop in the modeled mean discard proportion from 0.44 to 0.27.In contrast with the binomial model, the main effect of period was statistically significant ($p < 0.001$). To analyze how the effect of the variables has changed we relay on figure4.9 and figure 4.10 who respectably show the fitted smooth functions for each period

Cuadro 4.9: Smooth term significance and period effect coefficients

	Before LO		After LO	
Component	Edf	P-value	Edf	P-value
Intercept (Before LO)(Coefficient)	-0.2372	0.0233		
Period effect (After LO)(Coefficient)			-0.7485	2e-11
Haul duration	0.41	0.149	0.0001	0.221
Distance to coast	3.34	0.103	4.89	0.0168
Speed	3.36	0.017	3.81	0.00096
Mean depth	2.50	<2e-16	4.70	<2e-16
Recruitment	5.60	2.8e-05	0.093	0.237
Month	3.23	<2e-16	5.27	<2e-16
Hour	6.35	5.1e-05	0.00001	0.819
Spatial	11.98	3.5e-06	12.92	0.0047
Vessel effect	35.32	0.0087	12.00	0.0329

and variable and the corresponding difference smooths (after minus before). The haul duration showed a significant effect only in the post-LO period (p = 0.026), while being non-significant before, this is consistent with the smooth function where a shallow U-shape before the LO can be seen, with discard proportions lowest at intermediate tow lengths.After the LO, the relationship becomes nearly linear and downward-sloping but since the confidence intervals include zero the variable becomes non significant.The difference smooth confirms this effect where it is negative for short hauls and positive beyond four hours. In the case of the distance to the coast in only showed significance after the LO.For this variable the pre-LO effect is mostly flat, while the post-LO curve displays a pronounced U-shape: discard proportion declines until around 20 km from the coast and rises again offshore. The difference smooth becomes increasingly positive beyond 15 km. The recruitment index of hake was strongly significant in both periods, though the effect was notably stronger before the LO (Chi-sq = 70.8 vs 23.2).The effect in the periods diverge in shape: the pre-LO curve shows a steep rise at very high recruitment values, while the post-LO curve is flatter.This is corroborated by the difference smooth, which is positive only at the upper tail. Month of the year displays no clear pattern before the LO due to only be significant for the post-LO period,yet a pronounced seasonal trend emerges afterwards, with discard proportions peaking in late spring and early summer and falling during late autumn.When looking a the smooth differences it shows that discard proportions were higher before the LO in winter and late autumn, but higher after the LO from spring through summer. Hour its significant for both periods and it shows and interesting pattern where the maximum discard proportion happens during the midday and then the minimum is at the evening. Comparing both smooths structure persists after the LO but with reduced amplitude, this is also highlighted by the smooth difference where there is an attenuation being positive from dawn to early afternoon and negative after mid-afternoon Finally the spatial smooth was only significant in the pre-LO period, suggesting that discard rates were more spatially structured before the regulation, while the effect weakened afterwards.

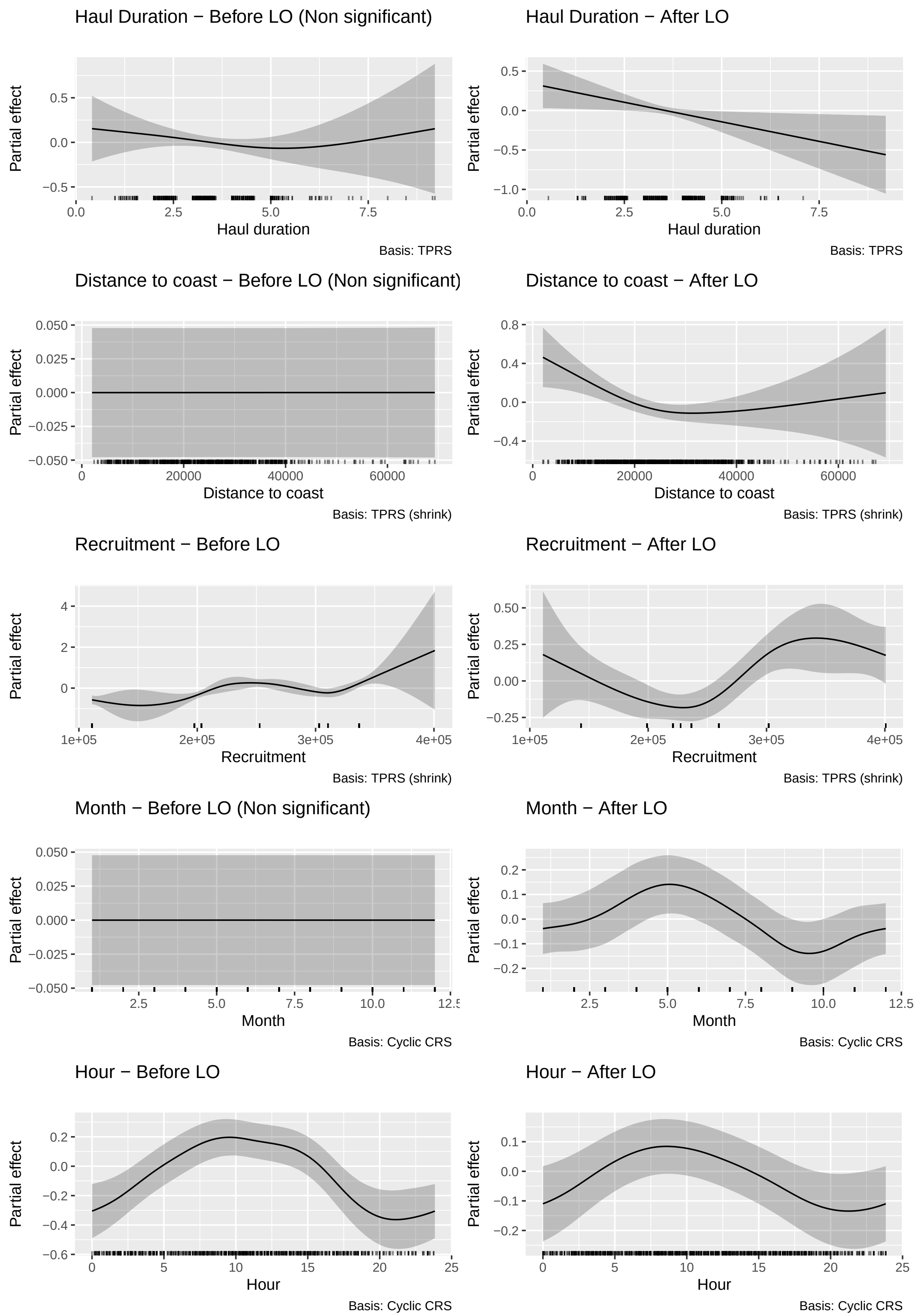

Figura 4.9: Effect of variables over the proportion discarded pre-LO and post-LO (Beta GAM using period interaction)

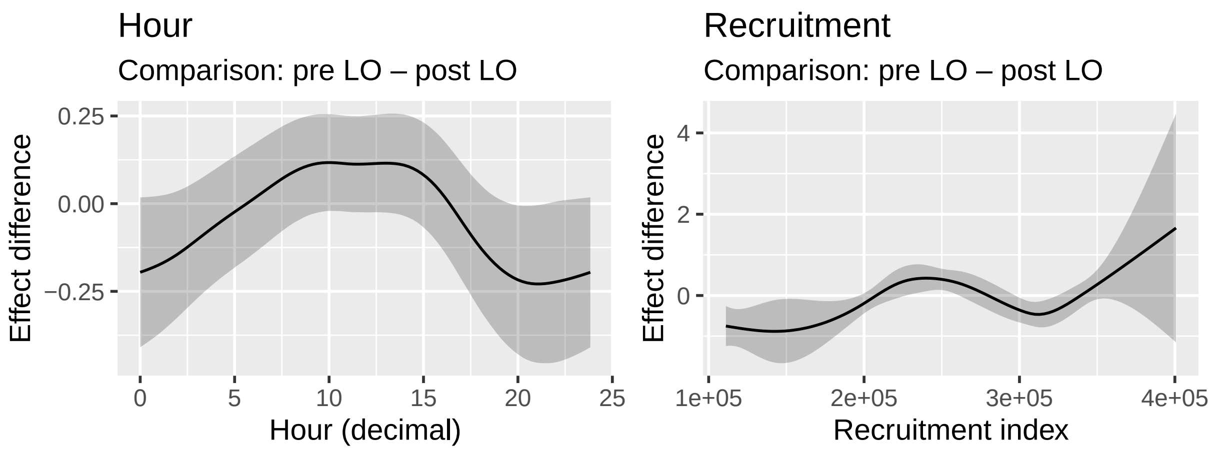

Figura 4.10: Difference in variables effect after and before LO for the proportion discard

Capítulo 5

Discussion.

The two-part modeling approach successfully decomposed discarding behavior into its constituent components, revealing fundamentally different variable structures for each process. The binomial component modeling the probability of discarding and the beta component modeling the proportion discarded demonstrated distinct covariate relationships, validating the methodological decision to treat these as separate processes rather than applying a single unified model.

The model structure revealed that the set of significant predictors differed substantially between the two components. Variables that strongly influenced the probability of discarding showed weak or non-significant effects on the proportion discarded, and vice versa. This differential variable importance across model components confirms that the underlying data-generating processes are distinct and would be poorly represented by a single modeling framework.

The beta-binomial structure effectively captured the overdispersion present in the discard data, with the two-part decomposition providing superior model fit compared to alternative single-component approaches. The separate modeling of zeros (non-discard events) and continuous proportions (when discarding occurs) allowed for more precise parameter estimation and better accommodation of the data's distributional characteristics.

This modeling architecture demonstrates that discarding behavior cannot be adequately characterized by a single process but requires recognition of its dual nature: the occurrence of discarding events and the intensity of discarding when it occurs. The successful separation of these components through the two-part model provides a more nuanced understanding of the mechanisms underlying discarding decisions and their responses to regulatory and operational factors.

5.1. General discard drivers

The factors driving the discarding process exhibit distinct differences between variables affecting discard probability and those influencing discard proportion. These differences manifest in both the significance of variables across models and their respective effects, justifying the use of a two-part modeling approach where combining both processes in a single model would dilute their individual effects. Although neither model achieves exceptional overall performance, they effectively illuminate the complex phenomenon of discarding behavior.

Discard probability demonstrates a more intricate relationship with environmental variables and hake population dynamics, particularly where these factors influence juvenile concentrations susceptible to discarding. This pattern suggests that discard probability is primarily driven by ecological factors determining the spatial and temporal distribution of vulnerable life stages within fishing grounds. Conversely, discard probability shows less sensitivity to technical variables related to fishing strategies, indicating that the likelihood of encountering discardable hake depends more on natural population patterns and environmental conditions than on operational fishing decisions. This distinction highlights the predominantly ecological nature of discard probability, where environmental drivers and recruitment dynamics play the dominant role in determining where and when discarding events are most likely to occur.

In contrast, the proportion discarded model demonstrates a more parsimonious structure, relying on a smaller subset of variables compared to the probability model. Specifically, the proportion model is influenced by four environmental variables plus two technical variables, suggesting that once a discarding event occurs, the magnitude of that discard is determined by a more limited set of factors. This reduced variable complexity indicates that the quantity of fish discarded during a discarding event follows different drivers than the likelihood of discarding occurring initially. While discard probability appears sensitive to a broader range of environmental and population dynamics, the actual proportion discarded seems governed by more specific operational and environmental constraints that directly influence sorting and retention decisions made during fishing operations.

Environmental variables

Environmental variables (mean depth, month, hour, and spatial position) significantly influence hake discarding patterns, with their effects being more pronounced on discard probability than on discard proportion. While all four variables significantly affect probability, only two variables significantly influence the proportion model, highlighting the differential sensitivity of these two response metrics to environmental conditions.

The most striking pattern emerges from the interplay between depth, spatial distribution, and hake recruitment. Discard probability decreases markedly with increasing depth, a trend that directly reflects the bathymetric habitat preferences and life cycle of European hake. Juveniles concentrate along the northern Iberian continental shelf, particularly in Cantabrian Sea and Galician waters at depths between 120-200 meters, where persistent recruitment areas have been documented. This shallow-water aggregation makes juveniles highly susceptible to trawl gear, thereby elevating discard probability in these zones.

The spatial analysis reveals that areas with the highest discard probability are concentrated over the Rías Baixas and the Artabro Gulf, regions that share a common characteristic: relatively shallow depths due to their location on extended continental shelf platforms. These areas function as critical juvenile nursery grounds, creating a convergence of ecological and operational factors that intensify discarding practices. As hake mature and gradually migrate to deeper offshore waters, the discard probability correspondingly decreases, explaining the inverse relationship between depth and discarding observed in the data (Izquierdo et al., 2021).

This spatial-bathymetric pattern reflects the broader ecological context of the Galician shelf system. The Galician continental shelf is notably wider and shallower compared to the eastern shelf, making it a particularly important recruitment area for European hake. Consequently, the combination of shallow depths, high juvenile density, and intensive fishing pressure creates distinct hotspots of discarding activity that are both spatially and temporally predictable based on the species' life history and habitat requirements.

The temporal variables (month and hour) reveal distinct patterns in hake discarding that reflect both seasonal and daily behavioral cycles of the species. Monthly variations demonstrate a clear seasonal component in discard risk, with peaks and troughs associated with temporal changes in recruitment dynamics, fishing effort, and species behavior. This seasonal variability in abundance and catchability is characteristic of hake fisheries, largely driven by spawning cycles and post-recruitment dispersal patterns (Alvarez et al., 2004).

The seasonal patterns differ between the two model components: discard probability peaks during winter, while discard proportion reaches its maximum in spring. Notably, the probability model shows greater complexity with a bimodal distribution, exhibiting peaks in both winter and spring/summer periods. These temporal peaks coincide with periods of higher juvenile concentrations, suggesting that seasonal recruitment pulses and subsequent dispersal patterns directly influence discarding intensity throughout the year.

Daily temporal patterns show consistent effects across both probability and proportion models, with peak discarding occurring during morning hours followed by a progressive decline through afternoon and night. This diurnal pattern likely reflects the diel vertical migration behavior of hake, particularly juveniles, which typically rise in the water column at night and descend during daylight hours (Dominguez-Petit et al., 2008). Consequently, morning trawling operations encounter higher concentrations of hake in the benthic zone, increasing their vulnerability to bottom trawl gear and elevating discard rates during these hours.

Technical variables

Technical variables—haul duration, towing speed, distance from the coast, and vessel code—are operational levers that skippers can adjust, and they encapsulate each vessel's particular fishing style. Among them, towing speed stands out: the discard-probability curve peaks precisely at the fleet's mean speed of ~2.9 knots. This is the speed band that maximizes catch efficiency for fast-swimming demersal species such as European hake. When vessels tow markedly slower or faster than this optimum, catches tend to fall, and with them the incentive (and opportunity) to discard. A physiological explanation underpins the pattern: juvenile hake can sustain only about 1.5 kn while adults manage ~2 kn, and both age classes reach their short burst limit at roughly 4 kn (Wardle, 1980). Operating near 2.9 kn therefore coincides with the speed at which gear efficiency is highest but the fish are still catchable, resulting in the highest probability of discard.When combined with long hauls the discard increases a lot when it comes to the probability The positive linear effect of haul duration on discard probability suggests that longer trawling times result in a higher likelihood of discarding. This relationship likely arises from a cumulative catch process, where more individuals—including undersized ones—are captured over time. Longer tows increase gear saturation and potential mortality of non-target individuals due to fatigue or damage in the codend (Catchpole et al., 2005). On the other hand the effect of the haul duration has the contrary effect over the proportion discard.

Distance from shore and mean depth are tightly coupled in the study area: the continental shelf narrows so quickly that every extra nautical mile offshore places the gear in appreciably deeper water. Both variables therefore display the same qualitative pattern—discard probability falls as depth (and thus distance) increases—yet only distance from shore remains significant in the model for the proportion discarded. This divergence likely arises because distance is a more comprehensive descriptor: it captures the depth gradient while also subsuming operational factors such as steaming time, fuel cost, and hold capacity that affect how much of the catch is retained. In addition, the distance–discard relationship is almost linear, whereas depth shows a curved response, making distance the cleaner and more powerful predictor when modelling proportions.

Population

Recruitment emerges as a fundamental factor for both models, exhibiting significant interactive effects with nearly all variables considered in the analysis. The general pattern shows that higher recruitment values are consistently associated with increased discard probability and greater proportions of discarded catch, reflecting the direct relationship between juvenile abundance and discarding intensity.

The influence of recruitment is closely intertwined with spatial and bathymetric factors, particularly depth, distance to coast, and spatial position. The previously discussed areas, such as the Rías Baixas and Artabro Gulf, function as established nursery grounds where elevated juvenile concentrations create predictable hotspots of discarding activity. These recruitment-environment interactions demonstrate that discarding patterns are not merely a function of fishing operations but are fundamentally shaped by the spatial distribution of vulnerable life stages within the ecosystem.

This recruitment-discard relationship has been extensively documented in European hake fisheries (ICES, 2021), confirming that understanding juvenile distribution patterns is crucial for predicting and managing discarding practices. The strong interactive effects between recruitment and environmental variables underscore the importance of considering ecological processes when developing discard reduction strategies.

5.2. Impact of the Landing Obligation

General patterns

The results from the models incorporating period-specific effects reveal a key insight: while the probability of discarding has remained broadly similar before and after the implementation of the Landing Obligation (LO), the proportion of the catch that is discarded has significantly decreased in the post-LO period. This suggests that although discards still occur, their relative intensity has been reduced—likely as a result of changes in onboard practices and compliance with the regulation. The combination of model significance tests and difference plots provides a robust framework for interpreting these patterns. This approach allows us not only to identify which variables were influential in shaping discarding behaviour within each period, but also to determine whether and how the effects of those same variables changed after the policy was introduced. The comparison of smooth terms across periods reveals how certain covariates not only remain relevant but also shift in influence after the implementation of the Landing Obligation (LO), suggesting changes in operational strategies or ecological conditions.

Environmental variables

Environmental variables demonstrate contrasting effects between the two model components, with their influence being more pronounced on discard probability than on discard proportion, where effects are more limited. The depth effect reveals particularly interesting changes between periods: at equivalent depths, there is now a lower probability of discard occurring in the post-LO period. This pattern is especially noteworthy given that mean fishing depths performed between periods remain remarkably similar, suggesting that operational strategies have evolved independent of bathymetric targeting changes.

The reduced significance of spatial position variables in the post-LO period—where spatial effects are only significant during the pre-LO period—indicates that regulatory changes have diminished the importance of area-specific fishing strategies in determining discard probability. This suggests that the observed changes in depth effects are not driven by shifts in spatial fishing patterns but rather by modifications in operational behavior within the same fishing grounds.

Temporal variables reveal significant strategic adaptations following regulation implementation. Monthly effects show a notable shift in seasonal discarding patterns: while winter months were associated with higher discard probability in the pre-LO period with summer representing the lowest discard period, the post-LO period exhibits increased probability during both summer and winter, with the lowest point now occurring in autumn. This temporal shift toward year-end may reflect vessels approaching their discard quotas and consequently adopting more selective fishing practices to comply with regulations,this makes this variable have not only a environmental component but a technical as well.

This regulatory influence is further evidenced in the proportion discard model, where monthly effects were non-significant in the pre-LO period but now demonstrate a clear annual evolution throughout the post-LO period. This transformation reinforces the importance of temporal factors in the new regulatory framework, likely reflecting vessels' need to manage their remaining quota allocations as the fishing year progresses.

Technical variables

Technical variables show less defined effects across regulatory periods, suggesting their reduced importance in explaining period-specific differences in discarding patterns. The most notable interactions include speed in the probability model, and both duration and distance in the post-LO proportion model.

The speed effect reveals that for equivalent trawling speeds, there is lower discard probability in the post-LO period, despite average speeds remaining nearly identical between periods. For haul duration, the effect indicates higher discard probability during shorter hauls in the post-LO period. These patterns may reflect a spatial relocation of fishing effort away from coastal areas where juveniles are more abundant, either as a direct regulatory response or as a by-product of changes in target species preferences, fishing zone selection, or fleet dynamics.

The distance effect in the post-LO proportion model suggests stronger spatial segregation of discarding risk following the policy implementation. This combination of technical variable effects points to subtle but important operational adaptations that are not fully captured by the environmental and spatial variables alone.

These findings indicate that underlying processes within technical operations remain partially unexplained by the current model structure, suggesting that additional operational or behavioral variables may be necessary to fully understand the mechanisms driving post-regulation changes in discarding practices.

Population

Recruitment patterns reveal fundamental shifts in fleet behavior following the Landing Obligation implementation. For discard probability, recruitment was a strong predictor in the pre-LO period, particularly at higher abundance levels, but substantially lost influence post-LO. This suggests operators have strategically adjusted their spatial or temporal fishing patterns to avoid peak juvenile concentrations, or improved on-board selectivity.

Regarding discard proportion, recruitment remained significant across both periods but with markedly different responses. The pre-LO period showed a stronger, nonlinear relationship where high recruitment years produced substantially larger discard fractions, indicating vessels frequently captured large quantities of juveniles in individual hauls. Post-LO, this relationship flattens considerably, suggesting fishing practices have adapted to reduce exposure to juvenile aggregations.

The most pronounced changes occur at high recruitment levels, where discard proportions decreased significantly after regulatory implementation (ICES, 2021). This indicates the fleet has developed more sophisticated avoidance strategies and enhanced selectivity measures, transitioning from opportunistic fishing to more selective practices that actively minimize encounters with vulnerable life stages.

5.3. Study limitations and future work

While the models developed in this study provide useful insights into the drivers of discarding behaviour before and after the implementation of the Landing Obligation (LO), they also present notable limitations that should be considered.

The overall explanatory power of the models remains moderate. In the case of discard probability (binomial GAM), the deviance explained reached approximately 28.6 %, and for the discard proportion (beta GAM), it was slightly higher at 34 %. These values indicate that a large portion of the variability in the response remains unaccounted for. This suggests that, although the selected covariates capture relevant effects, other important factors likely influence discarding behaviour and are not represented in the current framework.

This limitation partly comes from the nature of the data. Although onboard observer data offer a robust and direct view of fishing operations, they are restricted . Several potentially relevant variables—such as gear configuration, haul-by-haul catch handling, detailed biological variables, or fisher decision-making strategies—were either not recorded or not available at the necessary resolution. Environmental conditions (e.g., temperature gradients, prey distribution) were also excluded, despite their known influence on fish distribution and catchability.

Incorporating more variables or modelling complex interactions between predictors could potentially improve the models' fit. However, this would also increase the complexity and reduce the interpretability of the models, making it more difficult to derive clear conclusions or actionable recommendations. A key trade-off in modelling fisheries data lies between explanatory richness and practical clarity, especially when the goal is to support policy or management decisions.

In summary, while the models successfully highlight some consistent patterns and period-specific shifts in discard behaviour, they should not be viewed as fully deterministic representations of the system. Further research using richer datasets and integrating additional ecological and operational variables would be necessary to more completely explain the processes behind discards in this fishery.

Capítulo 6

Conclusions

- Although the models explain only a moderate proportion of deviance (26 % for discard probability and 29 % for discard proportion), they provide valuable insights into key drivers. The complexity of discard behaviour, involving biological, economic, regulatory, and operational dimensions, makes it difficult to capture fully through standard observer data. Including more explanatory variables (e.g., economic incentives, gear configurations, species composition) or interaction terms could enhance model accuracy but at the cost of reduced interpretability.

- The study successfully identified that different sets of variables influence the two key components of discarding: the probability that discards occur and the proportion discarded once they do. This distinction validates the use of a two-part modelling approach, as merging both processes into a single framework would obscure the distinct drivers behind each. Biological availability, operational effort, and fleet behavior all played different roles across these phase

- Using statistical models with interaction by period and spatial-temporal smoother, we found that the Landing Obligation (LO) had a measurable impact on the proportion of discards, which decreased significantly in the post-LO period. However, the overall probability of discarding did not change substantially, indicating that while discards still occur with similar frequency, their intensity has been reduced, possibly due to improved selectivity or better on-board practices.

- It can be observed that, prior to the regulatory change, only environmental and population-related variables were significant in the models. In contrast, technical variables became significant only during the post-Landing Obligation (LO) period. This pattern may indicate a shift in fishing strategies, whereby vessels began relying more on technical adjustments to reduce discards.

- While the models show some changes in how certain variables affect discarding—especially things like vessel speed, recruitment, and where fishing happens—most of these changes are quite small. In many cases, the overall pattern stays similar before and after the Landing Obligation, and only certain parts of the data show clear differences. This means that instead of a big shift in fishing behavior, the LO probably led to more subtle adjustments in how and where vessels operate.

Apéndice A

Apendix

Cuadro A.1: Variables excluded from the analysis, including their meaning and the reasons for their exclusion.

Variable	Description	Observations
C.ICES	ICES statistical fishing area where the haul was recorded.	Introduces redundant information when working with latitude and longitude.
Moon	Lunar phase at the time of the haul.	Many NA values, significantly reduces the dataset size.
Vertical and Horizontal Net Opening	Opening dimensions of the trawl net during operation.	Many NA values, significantly reduces the dataset size.
Surface temperature	Sea surface temperature during the haul.	Many NA values, significantly reduces the dataset size.
Heads of the net	Net component that helps maintain shape and opening.	Many NA values, significantly reduces the dataset size.
Wind Speed	Wind intensity measured during the fishing operation.	Many NA values, significantly reduces the dataset size.
Net size	Overall dimensions of the fishing net.	Very homogeneous, provides little information.
Course	Compass direction in which the vessel was moving.	Many NAs and not significant in the analyses.
English name	Common English name of the main species targeted or caught.	Not relevant for the analysis.
Ship direction	Direction of the vessel at the time of the haul.	Many NA values, significantly reduces the dataset size.
Cable	Cable length or configuration used in the trawl.	Very homogeneous, provides little information.
SobreCopo	Presence of additional mesh on top of the codend (selectivity device).	Theoretically illegal.
Sea State	Wave and sea conditions during the haul.	Many NA values, significantly reduces the dataset size.
Observer	Identifier of the onboard scientific observer.	Not filled correctly.
HP	Engine power of the vessel in horsepower.	Introduces redundant information when working with the variable Vessel.
Gross Registered Tonnes	Total registered tonnage of the vessel.	Introduces redundant information when working with the variable Vessel.
Lenght of the ship	Length of the vessel measured overall.	Introduces redundant information when working with the variable Vessel.

Cuadro A.2: Variables recorded for each fishing operation included in the analysis.

Variable	Unit	Description	Category
Discard rate	Proportion (0-1)	Primary response variable: proportion of hake discarded relative to total caught	Response
Discard presence/absence	Binary (0/1)	Binary response variable: absence (0) or presence (1) of hake discards	Response
Light conditions	Categorical (Night/Day)	Daylight conditions during fishing operation	Environmental
Vessel speed	knots	Average vessel speed during trawling	Effort
Mean depth	m	Average depth during trawling	Environmental
Haul duration	hours	Total duration of trawling operation	Effort
Recruitment	Number of juveniles per year	Number of juvenile fish that survive to a size or age at which they join the exploitable stock.	Catch
Distance to coast	m	Shortest distance to continental coastline	Environmental
Month	1-12	Month of the year (circular variable)	Temporal
Hour	0-23	Hour of the day when fishing operation started (circular)	Temporal
Latitude	Decimal degrees	Latitude of haul position (central point)	Spatial
Longitude	Decimal degrees	Longitude of haul position (central point)	Spatial
Haul position in the trip	Whole number	Position of the haul within the trip	Operational
Vessel code	Categorical	Unique vessel identifier	Vessel-specific
Home port	Categorical	Base port of the vessel	Operational
Landing Port	Categorical	Landing port for the catch	Operational

Cuadro A.3: Results of GAM models for discard presence/absence

Term	Fit.1.p	Fit.1.edf	Fit.2.p	Fit.2.edf	Fit.3.p	Fit.3.edf
Light	0.222	1	0.208	1	-	-
Haul duration (linear)	-	-	-	-	0.0487	1
Speed	p<2e-16	4.63	p<2e-16	4.63	p<2e-16	4.67
Mean depth	p<2e-16	4.53	p<2e-16	4.47	p<2e-16	4.47
Haul duration (smooth)	0.0568	0.73	0.0458	0.75	-	-
Recruitment	3.76e-5	5.18	3.77e-5	5.17	p<2e-16	5.24
Distance to coast	9.78e-4	3.18	0.00141	3.29	0.00105	3.36
Month	p<2e-16	5.29	p<2e-16	5.30	p<2e-16	5.37
Hour	0.0708	1.88	0.0835	1.78	0.00399	2.62
Position	p<2e-16	14.46	0.00274	13.02	0.00429	13.18
Vessel	7.07e-5	23.18	p<2e-16	23.01	p<2e-16	24.49
Home port	0.595	12.32	-	-	-	-

Cuadro A.4: Results of Beta–GAM models for discard rate

Term	Fit.1.P.value	Fit.1.EDF	Fit.2.P.value	Fit.2.EDF
Light	0.251	1	-	-
Haul duration (linear)	-	-	0.01985	1
Speed	0.276	0.196	-	-
Mean depth	0.178	2.35	-	-
Haul duration (smooth)	0.0121	0.857	-	-
Recruitment	<2e-16	1.205	<2e-16	11.697
Distance to coast	0.00225	2.188	0.00681	-
Month	0.0168	4.017	0.00504	4.505
Hour	0.00027	2.785	<2e-16	3.266
Position	0.0911	7.595	-	-
Vessel	<2e-16	30.787	<2e-16	28.156
Home port	0.4567	2.114	-	-

Bibliografía

[Álvarez(2004)] Álvarez, P., Fives, J., Motos, L., & Santos, M. B. (2004). Distribution and abundance of European hake larvae and post-larvae (Merluccius merluccius) in relation to hydrographic conditions in the Bay of Biscay. *Fisheries Research*, 70(1), 1–15.

[Augustin et al.(2013)] Augustin, N. H., Sauleau, E. A., & Wood, S. N. (2013). On quantile–quantile plots for generalized linear models. *Computational Statistics & Data Analysis*, 56, 2404–2409.

[Borcard et al.(2004)] Borcard, D., Legendre, P., Avois-Jacquet, C., & Tuomisto, H. (2004). Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, 85, 1826–1832.

[Borges(2015)] Borges, L. (2015). The evolution of a discard policy in Europe. *Fish and Fisheries*, 16, 534–540.

[Browne et al.(2024)] Browne, M., Calderwood, J., Brophy, D., & Minto, C. (2024). Single-species quotas drive discards by multi-species trawlers in the Celtic Seas ecoregion when their relative abundance fluctuates. *ICES Journal of Marine Science*, fsae122.

[Burnham & Anderson(2002)] Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd ed.). Springer-Verlag, New York.

[Catchpole et al.(2005)] Catchpole, T. L., Frid, C. L. J., & Gray, T. S. (2005). Discards in North Sea fisheries: causes, consequences and solutions. *Marine Policy*, 29, 421–430.

[Celić et al.(2018)] Celić, I., Libralato, S., Scarcella, G., Raicevich, S., Marčeta, B., & Solidoro, C. (2018). Ecological and economic effects of the landing obligation evaluated using a quantitative ecosystem approach: a Mediterranean case study. *ICES Journal of Marine Science*, 75, 1992–2003.

[Cosandey-Godin et al.(2015)] Cosandey-Godin, A., Krainski, E. T., Worm, B., & Flemming, J. M. (2015). Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. *Canadian Journal of Fisheries and Aquatic Sciences*, 72, 186–197.

[Cragg(1971)] Cragg, J. G. (1971). Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica*, 39, 829–844.

[Davie et al.(2025)] Davie, S., Wakeford, R. C., Whitley, C., van den Berg, P., Rimpler, A., Burke, A., et al. (2025). Study supporting the evaluation of the landing obligation – Common Fisheries Policy. Publications Office of the European Union. doi:10.2926/5282226.

[Depestele et al.(2011)] Depestele, J., Vandemaele, S., Vanhee, W., Polet, H., Torreele, E., Leirs, H., & Vincx, M. (2011). Quantifying causes of discard variability: an indispensable assistance to discard estimation and a paramount need for policy measures. *ICES Journal of Marine Science*, 68, 1719–1725.

[Domínguez-Petit et al.(2008)] Domínguez-Petit, R., Alonso-Fernández, A., & Saborido-Rey, F. (2008). Reproductive strategies of European hake (Merluccius merluccius L.) in the Galician Shelf (NW Spain). *Fisheries Research*, 90(1–3), 108–115.

[EU(2013)] EU (2013). Regulation (EU) No 1380/2013 of the European Parliament and of the Council of 11 December 2013 on the Common Fisheries Policy. *Official Journal of the European Union*, L354, 22–61.

[FAO(2020)] FAO (2020). *FAO Yearbook. Fishery and Aquaculture Statistics 2018.* FAO, Rome.

[Feekings et al.(2012)] Feekings, J., Bartolino, V., Madsen, N., & Catchpole, T. (2012). Fishery discards: factors affecting their variability within a demersal trawl fishery. *PLoS ONE*, 7, e36409.

[Ferrari & Cribari-Neto(2004)] Ferrari, S. L. P., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31, 799–815.

[Hartig(2020)] Hartig, F. (2020). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package v0.3-3.

[ICES(2021)] ICES (2021). Hake (Merluccius merluccius) in Division 8.c and Subarea 9 (Cantabrian Sea and Atlantic Iberian waters). ICES Advice 2021. https://doi.org/10.17895/ices.advice.7757

[Izquierdo et al.(2021a)] Izquierdo, F., Paradinas, I., Cerviño, S., Conesa, D., Alonso-Fernández, A., Velasco, F., Preciado, I., & Pennino, M. G. (2021). Spatio-Temporal Assessment of the European Hake (Merluccius merluccius) Recruits in the Northern Iberian Peninsula. *Frontiers in Marine Science*, 8, 614675.

[Izquierdo et al.(2021b)] Izquierdo, A., et al. (2021). Integrating dynamic habitat models in fisheries advice: the case of European hake in the Bay of Biscay and Iberian waters. *Frontiers in Marine Science*, 8, 676853.

[Lecomte et al.(2013)] Lecomte, J. B., Benoît, H. P., Ancelet, S., Etienne, M. P., Bel, L., & Parent, E. (2013). Compound Poisson–gamma vs. delta–gamma to handle zero-inflated continuous data under a variable sampling volume. *Methods in Ecology and Evolution*, 4, 1159–1166.

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

[Miller et al.(2019)] Miller, D. L., Brickman, D., Roth, J. D., Adams, P., Arroyo, K. L., Barton, C. J., et al. (2019). The use of generalized additive models in ecology. *Methods in Ecology and Evolution*, 10, 1517–1531.

[Pedersen et al.(2019)] Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: an introduction with mgcv. *PeerJ*, 7, e6876.

[Pennington(1983)] Pennington, M. (1983). Efficient estimators of abundance for fish and plankton surveys. *Biometrics*, 39, 281–286.

[Pennino et al.(2014)] Pennino, M. G., Muñoz, F., Conesa, D., López-Quílez, A., & Bellido, J. M. (2014). Bayesian spatio-temporal discard model in a demersal trawl fishery. *Journal of Sea Research*, 90, 44–53.

[Pennino et al.(2020)] Pennino, M. G., Bevilacqua, A. H., Torres, M. A., Bellido, J. M., Solé, J., Steenbeek, J., & Coll, M. (2020). Discard ban: a simulation-based approach combining hierarchical Bayesian and food-web spatial models. *Marine Policy*, 116, 103324.

[Plet-Hansen et al.(2020)] Plet-Hansen, K. S., Bastardie, F., & Ulrich, C. (2020). The value of haul-by-haul fish size information for monitoring discards and unwanted catches. *ICES Journal of Marine Science*, 77, 2729–2740.

[Richards(2008)] Richards, S. A. (2008). Dealing with over-dispersed count data in applied ecology. *Journal of Applied Ecology*, 45, 218–227.

[Rochet & Trenkel(2005)] Rochet, M. J., & Trenkel, V. M. (2005). Factors for the variability of discards: assumptions and field evidence. *Canadian Journal of Fisheries and Aquatic Sciences*, 62, 224–235.

[Sánchez & Serrano(2003)] Sánchez, F., & Serrano, A. (2003). Variability of groundfish communities of the Cantabrian Sea during the 1990s. *ICES Marine Science Symposia*, 219, 249–260.

[Sánchez et al.(2008)] Sánchez, F., Gil, J., & Preciado, I. (2008). Spatial distribution patterns and habitat preferences of juvenile European hake (Merluccius merluccius) in the Cantabrian Sea. *ICES Journal of Marine Science*, 65(9), 1400–1407.

[Schwarz(1978)] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.

[Simpson(2018)] Simpson, G. L. (2018). Modelling palaeoecological time series using generalized additive models. *Frontiers in Ecology and Evolution*, 6, 149.

[Villasante et al.(2019)] Villasante, S., Pita, P., Antelo, M., Neira, J. A., & de Santiago, R. P. (2019). Socio-economic impacts of the landing obligation of the European Union Common Fisheries Policy on Galician (NW Spain) small-scale fisheries. *Ocean & Coastal Management*, 170, 60–71.

[Wardle(1980)] Wardle, C. S. (1980). Effects of temperature on the maximum swimming speed of fish. *Marine Biology*, 56, 145–156.

[Wood(2017)] Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman & Hall/CRC, Boca Raton.

[Zuur(2012)] Zuur, A. F. (2012). *A Beginner's Guide to Generalized Additive Models with R*. Highland Statistics Ltd, Newburgh.

[Zuur & Ieno(2016)] Zuur, A. F., & Ieno, E. N. (2016). A protocol for conducting and presenting results of regression-type analyses. *Methods in Ecology and Evolution*, 7, 636–645.

[Zuur et al.(2009)] Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2009). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 3–14.

[Zuur et al.(2014)] Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., & Smith, G. M. (2014). *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York.