



Universidade de Vigo

Traballo Fin de Mestrado

Rasgos de vida das especies demersais en Augas Ibéricas Atlánticas

Davinia Lojo Amoedo

Mestrado en Técnicas Estatísticas

Curso 2020-2021

Proposta de Traballo Fin de Mestrado

<p>Título en galego: Rasgos de vida das especies demersais en Augas Ibéricas Atlánticas</p>
<p>Título en español: Rasgos de vida de las especies demersales en Aguas Ibéricas Atlánticas</p>
<p>English title: Study of life-traits of demersal species in Iberian Atlantic Waters</p>
<p>Modalidade: Modalidade B</p>
<p>Autora: Davinia Lojo Amoedo, Universidade de Santiago de Compostela</p>
<p>Director: Javier Roca Pardiñas, Universidade de Vigo</p>
<p>Titor/as: Santiago Cerviño, M.Grazia Peninno, Marta Cousido-Rocha, Instituto Español de Oceanografía.</p>
<p>Breve resumo do traballo:</p> <p>Neste traballo analízase a variabilidade anual dos parámetros biolóxicos talla de primeira madurez sexual e factor de condición relativo da merluza europea en augas Ibéricas Atlánticas, para o período de estudo comprendido entre 1980 e 2019 con respecto a variables ambientais e biolóxicas empregando modelos GAM para avaliar os seus posibles efectos.</p>
<p>Recomendacións:</p> <ul style="list-style-type: none"> ▪ Interese en aplicacións prácticas. ▪ Dominio de R. ▪ Ter cursado a materia de Regresión Xeneralizada e Modelos Mixtos, ou ter coñecemento dos GLM e GAM
<p>Outras observacións:</p>

Don Javier Roca Pardiñas, da Universidade de Vigo, don Santiago Cerviño, dona M.Grazia Peninno, e dona Marta Cousido-Rochado Instituto Español de Oceanografía, informan que o Traballo de Fin de Mestrado titulado

Rasgos de vida das especies demersais en Augas Ibéricas Atlánticas

foi realizado baixo a súa dirección por doña Davinia Lojo para o Mestrado en Técnicas Estatísticas. Estimando que o traballo está rematado, dan a súa conformidade para a súa presentación e defensa ante un tribunal.

En Vigo, a 17 de Xuño de 2021.

O director:

O titor :

Don Javier Roca Pardiñas

Don Santiago Cerviño

A titora:

A titora:

Dona Marta Cousido-Rocha

Dona M.Grazia Peninno

A autora:

Davinia Lojo

Dona Davinia Lojo Amoedo

Agradecementos

Antes de proceder cos agradecementos e culminar así cun período de aprendizaxe intenso, declaro abertamente que cando rematei as materias que se cursan no Máster de Estatística, afrontei con algo de reticencia as prácticas do Mestrado no Instituto Español de Oceanografía. Subliño con algo de reticencia por adentrarme nun mundo totalmente alleo a min no campo académico pola miña formación como politóloga. A día de hoxe, miro cara atrás e, se ben foi todo un reto, tamén foi todo un acerto. Dende que comecei as prácticas, adoptei como miña unha mostra de 33094 merluzas europeas, participei en dous congresos e escribín, con axuda de máns amigas, o que algún día será o meu plan de tese o cal, abriume as portas ao Centro Tecnolóxico do Mar (CETMAR) no cal traballo actualmente.

Dito isto, agradezo ao IEO de Vigo pola oportunidade primeiro e, pola acollida despois, en especial a Santiago Cerviño, a M.Grazia Pennino e a Marta Cousido-Rocha pola súa profesionalidade e a súa valiosa aportación técnica. Ademáis, pese a que estas prácticas desenvolvéronse de xeito telemático, a distancia física non impediu a social a cal se consolidou mediante reunións semanais a través das cales se me facilitaron todas as ferramentas para completar o meu TFM satisfactoriamente.

Agradezo tamén a Javier Roca Pardiñas como director deste proxecto.

A Celestino García Arias, titor do meu TFG e referente no campo académico, pois sen a súa labor como docente seguramente non estaría escribindo este TFM e, por ende, este apartado.

Á familia que non se escolle os cales ensináronme que sen esforzo non hai nada e, por último, á familia que se escolle, en especial a Alex Almón porque e cito textualmente “se podes debuxalo, podes facelo”.

Índice xeral

Resumo	XI
Prefacio	XIII
1. Introducción e motivación do problema	1
1.1. Instituto Español de Oceanografía	1
1.2. Exposición do problema	2
1.3. Obxectivos	2
1.4. Métodos de resolución	2
2. Material e métodos	3
2.1. Área e Material de estudo	3
2.2. Metodoloxía	5
2.2.1. Estimación dos rasgos de vida	5
2.2.2. Modelos de regresión	6
2.2.3. Modelos Box-Jenkins	13
2.2.4. Aplicación das series temporais nos GAM	17
3. Análise de resultados	23
3.1. Análise descriptivo: Unha primeira aproximación	23
3.1.1. Talla da merluza europea	23
3.1.2. Peso vivo da merluza europea	26
3.2. Rasgos de vida	28
3.2.1. Talla de primeira madurez sexual	28
3.2.2. Factor de condición	31
3.3. Modelos de regresión	33
3.3.1. Talla de primeira madurez sexual	34
3.3.2. Factor de condición relativo	41
4. Conclusións	45
Apéndice	49
1. Análise datos	49
2. Contraste das hipóteses básicas sobre os residuos dos modelos	51
2.1. Contraste das hipóteses básicas sobre os residuos da talla de primeira madurez sexual para as merluzas macho	52
2.2. Contraste das hipóteses básicas sobre os residuos do modelo da talla de primeira madurez sexual para as merluzas femia	53
2.3. Contraste das hipóteses básicas sobre os residuos do modelo para factor de condición da poboación da merluza	54
2.4. Contraste das hipóteses básicas sobre os residuos do modelo para factor de condición das merluzas macho	55
2.5. Contraste das hipóteses básicas sobre os residuos do modelo para factor de condición das merluzas femia	56

Resumo

Resumo en galego

Comprender as dinámicas subxacentes dunha especie é esencial para a xestión e a avaliación dos recursos mariños. Así, neste traballo analízanse os parámetros biolóxicos talla de primeira madurez sexual, entendida como a talla á cal o 50% dos individuos acadan a madurez sexual (L_{50}), e factor de condición relativo (K_n), entendido como un indicador da condición corporal da merluza europea en augas Ibéricas Atlánticas, para o período de estudo comprendido entre 1980 e 2019. Ademais, analízase a variabilidade interanual destes rasgos de vida con respecto a variables ambientais como son a Oscilación Multidecadal do Atlántico, a Oscilación do Atlántico Norte e a Temperatura Superficial do Mar e biolóxicas como son a biomasa, a biomasa reproductora e a talla utilizando modelos aditivos xeneralizados para avaliar os seus posibles efectos. Os datos biolóxicos proporciónanse polo Instituto Español de Oceanografía de Vigo mentres que as variables ambientais extraense de bases de datos de libre acceso. Os resultados deste estudo amosan por un lado, unha tendencia temporal á diminución da talla de primeira madurez sexual da merluza europea e, polo outro, a variación do factor de condición relativo depende principalmente da talla. Estes resultados poden ser clave para entender o estado do stock da merluza europea nas augas Ibéricas Atlánticas.

English abstract

Understanding the underlying dynamics of a species is essential for the management and assessment of marine resources. Thus, in this study, we analyze the variability of size at first maturity (i.e., length at which the 50% of individuals reach sexual maturity, L_{50}) and the relative condition factor (i.e., an indicator of the body condition, K_n), of the European hake in the Atlantic Iberian waters for the study period 1980-2019. In addition, the interannual variability of these life traits is assessed with respect to environmental variables, (i.e., Atlantic Multidecadal Oscillation, the North Atlantic Oscillation and Sea Surface Temperature), and biological variables (i.e., individuals biomass, spawning biomass and size) using generalized additive models. The datasets of the European hake were provided by the Spanish Institute of Oceanography of Vigo while the environmental variables were extracted from freely accessible databases. The results of this study show, on the one hand, a temporal trend to decrease in the size at first maturity of European hake and, on the other hand, a variation of the relative condition factor depends mainly on the size. These results could be the key to understand the European hake stock status in Iberian Atlantic Waters.

Prefacio

A merluza europea (*Merluccius merluccius*) é unha das especies comerciais máis importantes no Atlántico Nororiental e é avaliada polo Consello Internacional para a Exploración do Mar (ICES, polas súas siglas en inglés “International Council for the Exploration of the Sea”) en dúas unidades: as poboacións do norte e do sur. A pesar do plan de recuperación aplicado en 2006 (CE, 2166/2005) e o plan de xestión plurianual para as augas occidentais (UE 2019), a taxa de mortalidade por pesca desta especie demersal, segue sendo superior á mortalidade por pesca que proporciona o máximo rendemento sostible (F_{msy}), aunque a biomasa da poboación reproductora está en condicións biolóxicas seguras (ICES, 2018)[17].

A información sobre os rasgos de vida das especies explotadas comercialmente tórnase esencial para a avaliación dos recursos mariños. Sen embargo, normalmente os análises dos rasgos de vida non teñen en conta as diferenzas intrínsecas que poidan ter os individuos capturados en diferentes anos. Neste senso, este traballo de fin de mestrado ten como finalidade, estudar a talla de primeira madurez sexual (a lonxitude á cal o 50 % da poboación presenta gónadas maduras) e o factor de condición relativo (indicador da condición corporal do animal) para proporcionar unha mellor comprensión destes rasgos de vida e determinar se a variabilidade temporal destes rasgos de vida, vese afectada significativamente por variables ambientais e biolóxicas.

A variabilidade da talla de primeira madurez sexual e o factor de condición relativo para cada sexo e a combinación de ambos ao longo do período de estudo (1980-2019), analízase mediante modelos aditivos xeneralizados (GAMs) introducidos por Hastie e Tibshirani en 1990 [16] considerando como variables predictoras, factores ambientais como a Oscilación Multidecadal do Atlántico (AMO, polas súas siglas en inglés “Atlantic Multidecadal Oscillation”, a Oscilación Atlántico Norte, (NAO do inglés “North Atlantic Oscillation”) e a Temperatura superficial do mar (SST do inglés “Sea Surface Temperature”) e factores biolóxicos como a biomasa entendida como a abundancia do stock en unidades de peso e a biomasa reproductora para a que se emprega o peso total dos individuos sexualmente maduros. Cada modelo axustouse utilizando a distribución gaussiana, como base de suavización os cubic regression splines e como método de estimación, a máxima verosimilitude restrinxida (REML, do inglés, “Restricted Maximum Likelihood”) validando nos residuos os supostos de estacionariedade, independencia, normalidade e homocedasticidade. Ao estar abordando series temporais, é posible que os residuos amosen dependencia temporal, neste caso, axustaranse a través da metodoloxía Box-Jenkins mediante un proceso ARMA (p, q).

A construción dos modelos realízase empregando o procedemento “backward” o cal consiste en partir dun modelo que inclúa todas as variables explicativas non correlacionadas entre sí, e logo extraer secuencialmente as menos influíntes segundo a proba t de Student. Ademáis, límitase o número de nodos para controlar o grado de suavidade das funcións dos modelos GAM evitando o sobreaxuste en aras de obter un modelo robusto e fiable.

Este traballo estrutúrase en catro capítulos. No primeiro, introdúcese o tema obxecto de estudo e plantéxanse os obxectivos do traballo. No segundo capítulo abórdanse os aspectos xerais e teóricos dos modelos estatísticos empregados neste traballo e, por último, preséntanse os resultados e as conclusións do traballo.

Capítulo 1

Introdución e motivación do problema

Neste primeiro capítulo, preséntase ao Instituto Español de Oceanografía (IEO), institución coa que se levou a cabo o estudo dos rasgos de vida da merluza europea nas augas Ibéricas Atlánticas para, seguidamente, abordar os motivos de interese do mesmo. Posteriormente, preséntanse os obxetivos e, por último, abórdanse os motivos polos cales se seleccionan cada unha das metodoloxías que se empregan para acadar os obxetivos.

1.1. Instituto Español de Oceanografía

O presente estudo realízase baixo a coordinación do **Instituto Español de Oceanografía** (IEO), un organismo público de investigación (OPI), dependente do Ministerio de Ciencia e Innovación, adicado á investigación en ciencias do mar, especialmente no relacionado co coñecemento científico dos océanos, a sustentabilidade dos recursos pesqueiros e o medio ambiente mariño.

Entre as funcións do IEO cómpre salientar:

- Investigación científica en oceanografía e ciencias do mar e o estudio multidisciplinar do mar.
- Asesoramento á Administración Xeral do Estado no concerninte a política pesquera e mariña.
- Promoción da cooperación en investigación mariña a escala rexional, nacional e internacional.
- Formación investigadores e difusión dos coñecementos oceanográficos.

O IEO representa a España na maioría dos foros científicos e tecnolóxicos internacionais relacionados co mar e os seus recursos. Conta cunha sede central en Madrid e nove centros oceanográficos costeiros: A Coruña, Baleares, Cádiz, Canarias, Gijón, Málaga, Murcia, Santander e Vigo; cinco plantas de experimentación de cultivos mariños; doce estacións mareográficas e unha estación receptora de imaxes de satélite. A súa flota oceanográfica, de máis dunha vintena de embarcacións, conta con cinco buques oceanográficos importantes, entre os que destacan Ramón Margalef e Ángeles Alvariño.

O estudo enmárcase dentro do proxecto IMPRESS (polas súas siglas en inglés “Improving scientific advice to fishery management for resources of interest for Spain in Atlantic waters”) (RTI2018-099868-B-I00)¹ cuxo obxectivo radica en mellorar a calidade do asesoramento científico para a xestión pesqueira dos recursos de interese para as frotas españolas no Atlántico Noroeste.

¹para máis información <https://impressproject.github.io/PROJECTIMPRESS/>

1.2. Exposición do problema

A merluza europea é un recurso de gran importancia comercial en Augas Ibéricas Atlánticas. A pesares do plan de recuperación implementado en 2006 e o Plan de Xestión plurianual das augas occidentais, a taxa de mortalidade por pesca desta especie segue estando por encima da que proporciona o rendemento máximo sostible (F_{msy}). Nembargantes, a biomasa reproductora está por encima dos niveis precautorios.

A información sobre os rasgos de vida é esencial no estudo da bioloxía das especies de interés comercial para deseñar modelos de avaliación eficientes, sen embargo, o coñecemento incompleto ou erróneo de aspectos claves da súa bioloxía ao non considerar a estrutura espacial ou a variabilidade temporal introduce un alto grado de incertidume na avaliación na que se basean tanto as medidas de xestión como os plans de recuperación.

1.3. Obxectivos

En liña co anteriormente dito, o estudo dos rasgos de vida dunha especie é esencial para comprender os procesos biolóxicos subxacentes á dinámica dunha poboación para unha avaliación e xestión eficientes dos recursos mariños. Isto motivou á investigación científica da merluza europea nas Augas Ibéricas Atlánticas, concretamente nas divisións 8c e 9a do Consello Internacional para a Exploración do Mar (ICES, polas súas siglas en inglés “International Council for the Exploration of the Sea”), cun dobre obxectivo:

- Estudar os parámetros biolóxicos da talla de primeira madurez sexual (definida esta como a lonxitude á cal o 50 % dos individuos son maduros sexualmente) e o factor de condición relativo (como indicador da condición corporal) da merluza europea.
- Analizar a variabilidade temporal destes rasgos de vida con respecto a factores ambientais e biolóxicos.

1.4. Métodos de resolución

O principal obxectivo deste estudo, tal e como se ven dicindo, é analizar a variabilidade temporal dos rasgos de vida da merluza europea. O gran número de momentos en que se realizan as medicións, fai que considerar a variable temporal como unha variable factor non sexa aconsellable e, axustar unha tendencia lineal global a todos os datos, tería pouco sentido xa que é posible que existan relacións moito máis complexas entre as covariables e as variables resposta do modelo. Unha alternativa é modelar a relación entre a variable resposta e a covariable temporal empregando funcións suaves basadas en “splines” [24] sen forzar a que a relación tome algunha forma particular. Por este motivo de entre os diferentes modelos de regresión dispoñibles na literatura, axústanse os modelos GAM (Sección 2.2.2.3). Non obstante, ao estar abordando series temporais, é posible que os residuos estén correlados. Neste caso, procederase ao modelado de dita dependencia a través da metodoloxía Box-Jenkins [12], unha técnica clásica de estudo de series temporais, a cal se fundamenta na consideración exclusiva dos valores pasados da serie temporal para explicar a súa evolución. Por último, cómpre sinalar que ademáis da metodoloxía GAM e Box-Jenkins, neste estudo, teñen cabida modelos máis sinxelos como os modelos lineais e os modelos lineais xeneralizados (LM e GLM) pois son empregados para a estimación das que serán as variables resposta no modelo GAM, o factor de condición e a talla de primeira madurez sexual respectivamente.

Capítulo 2

Material e métodos

2.1. Área e Material de estudo

Os datos biolóxicos e pesqueiros proporciónanse polo Instituto Español de Oceanografía (IEO) e estos proveñen de mostraxes mensuais das frotas comerciais que operan en Aguas Ibéricas Atlánticas, concretamente nas divisións 8c e 9a (que se amosan na figura 2.1) do ICES, para o período comprendido entre 1980 e 2019.

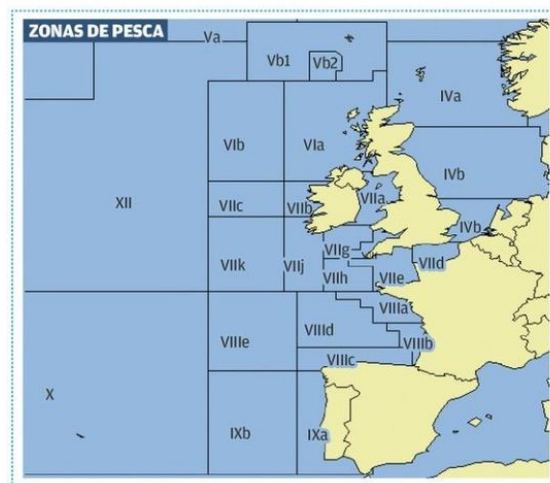


Figura 2.1: Mapa das subzonas e divisións do ICES, Atlántico nororiental extraído da nova de “La Voz de Galicia” [3].

Mostrexáronse un total de 33094 exemplares de merluza, aos cales se lles determinou medidas como a talla¹ (cm), o peso vivo² (gr), peso eviscerado³ (gr), sexo (macho, femia e descoñecido) e estado de madurez (inmaduros e maduros).

Neste senso cómpre salientar que se descoñece en que división (8c o 9a) se capturan 15070 das 33090 merluzas mostrexadas. Ademais, o 85 % dos individuos captúranse na división 9a fronte ao 15 % que se capturan na 8c, o que apunta que a mostra non é proporcional segundo zona. Ademais, non se dispón de datos en determinados meses para cada ano ao longo do período de estudo o que dificulta a análise con frecuencia mensual (ver Cadro 1 da Sección .1 do Apéndice)

Tendo como criterio a manexabilidade da información, nalgúñas variables agrupáronse categorías de forma homoxénea. A agregación das categorías, dada a perda de información que

¹Lonxitude

²Peso total do individuo

³Peso total do individuo sen vísceras

supón, únicamente se empregou tendo en conta as limitacións dos datos pola propia natureza dos mesmos. Deste xeito, procederáse a realizar a análise considerando como espacio as dúas áreas conxuntamente e, como unidade temporal o ano.

Cos datos rexistrados, estimáronse ragos de vida como son a talla de primeira madurez sexual e o factor de condición relativo para cada sexo e para a combinación de ambos para cada ano dende 1980 ata 2019. Cómpre destacar que a falta de información sobre algúns rasgos biolóxicos impediron estimar a talla de primeira madurez sexual e o factor de condición relativo para algúns anos. Estes valores faltantes cubríronse mediante medias móbiles [10].

Enténdese como talla de madurez sexual a lonxitude á cal o 50 % da poboación presenta gónadas maduras listas para a o desove, por esa razón tamén se coñece como L_{50} . Non se debe confundir coa talla ou longituxe á cal un individuo pode atoparse maduro, xa que ésta regularmente é moi inferior á L_{50} .

O factor de condición é a saúde física dunha poboación de peixes baseada na gordura ou gordura relativa do peixe. Dito doutro xeito, mídese a condición dun peixe para determinar se pesa máis ou menos do que cabería esperar en función da súa lonxitude, baseándose na suposición de que os animais máis pesados dunha determinada lonxitude exhiben mellores condicións físicas.

Ademáis o IEO tamén proporcionou para este traballo os datos da biomasa entendida como a abundancia do stock en unidades de peso e a biomasa reproductora para a que se emprega o peso total dos individuos sexualmente maduros da merluza europea para o período de estudo abordado.

Así mesmo, os datos das variables ambientais, como a Oscilación Atántico Norte (NAO, polas súas siglas en inglés “North Atlantic Oscillation”), a Oscilación Multidecadal do Atlántico (AMO, polas súas en inglés “Atlantic Multidecadal Oscillation”) e a Temperatura da Superficie do Mar (SST, do inglés “Sea Surface Temperature”), obtivéronse de bases de datos de libre acceso. Éstas extraírense da páxina web correspondente ao Laboratorio Oceanográfico e Meteorolóxico do Atlántico.

A Oscilación do Atlántico Norte, é un patrón de variabilidade atmosférica que ten un impacto significativo nas condicións oceánicas. Máis concretamente, descríbese como unha fluctuación a gran escala na presión atmosférica a nivel do mar entre o sistema de altas presións subtropical localizado perto das Azores no Océano Atlántico e o sistema de baixas presións subpolar cercano a Islandia [20]. Esta obténse da páxina <https://crudata.uea.ac.uk/cru/data/nao/> para o período comprendido entre (1980 – 2019) sen suavizar.

A Oscilación Multidecadal do Atlántico trátase dun fenómeno climático natural que incide na variación das temperaturas superficiais do mar no Atlántico Norte, con fases frías e cálidas que poden durar de 20 a 40 anos. Esta oscilación preséntase na maior parte do océano comprendida entre Groenlandia e o Ecuador e incluso pode afectar o Pacífico norte. Esta obténse da páxina <https://www.psl.noaa.gov/data/timeseries/AMO/>, para o período (1980 – 2019) sen suavizar.

A temperatura da superficie do mar obténse da páxina <https://www.ncdc.noaa.gov/oisst>. Cómpre salientar que se trata dun conxunto de datos globais a unha resolución diaria dispoñibles a partir do 1 de setembro de 1981, polo que para obter a temperatura superficial do mar con frecuencia anual para as áreas obxecto de estudo neste traballo, séguense os pasos descritos no seguinte tutorial https://cran.rstudio.com/web/packages/heatwaveR/vignettes/OISST_preparation.html. Ademáis, tendo en conta que os datos están dispoñibles a partir de setembro de 1982, extraírense os mesmos a partir do ano 1982.

2.2. Metodoloxía

Esta Sección xira en torno á revisión das técnicas estatísticas empregadas para acadar os obxectivos mencionados na Sección 1.3. Deste xeito, esta Sección estrutúrase en catro partes: na primeira, abórdase a metodoloxía para a estimación dos rasgos de vida para a merluza europea, a talla de primeira madurez sexual e o factor de condición relativo. A segunda Sección, leva ao lector por un percorrido dos distintos modelos de regresión para ofrecer un mellor entendemento dos distintos modelos aplicados neste traballo. Tendo en conta que se abordan series temporais, na Sección 2.2.3, preséntase a metodoloxía Box Jenkins. Tras isto, expónse brevemente un apartado que permite ao lector unha guía para a aplicación práctica 2.2.4.

2.2.1. Estimación dos rasgos de vida

Neste apartado abórdase a estimación dos rasgos de vida que serán, tal e como se dixo anteriormente, as variables dependentes deste estudo, concretamente a talla de primeira madurez sexual e o factor de condición relativo.

2.2.1.1. Talla de primeira madurez sexual

A talla de primeira madurez sexual estímase para cada ano entre 1980-2019 e para cada sexo, e tamén para a combinación de ambos, empregando un modelo loxístico da seguinte forma:

$$P_L = \frac{1}{(1 + \exp^{-(B_0 + B_1 L)})} \quad (2.1)$$

onde P_L é a proporción de exemplares maduros á lonxitude (L) e B_0 e B_1 , son os parámetros que representan a posición e a pendente da curva, respectivamente. Estes parámetros estímase por máxima verosimilitude asumindo unha distribución binomial (maduro/inmaduro).

A lonxitude de madurez queda definida cando a función loxística acada o 50% da probabilidade de observar individuos maduros a esa clase de lonxitude (L_{50}) e obtéñese como a razón entre os parámetros B_0 e B_1

$$L_{50} = \frac{B_0}{B_1}.$$

Emprégase o paquete `SizeMat` [15] de R software.

2.2.1.2. Factor de condición relativo

A talla e o peso son medidas biométricas dunha especie que permiten determinar o factor de condición relativo, é dicir, se un individuo pesa máis ou menos do que cabería esperar en función da súa lonxitude.

Se ben para estimar este parámetro emprégase normalmente o peso eviscerado do individuo, que non está afectado pola variabilidade no peso causado pola dieta, neste traballo estímase o factor de condición co peso total. Esta decisión tómase a tenor de evitar a menor perda de información posible, derivada dunha análise exploratoria previa dos datos.

En consonancia co anteriormente dito, calcúlase o factor de condición relativo para cada individuo de acordo á seguinte fórmula:

$$K_n = \frac{P_v}{P_e} \quad (2.2)$$

onde P_v é o peso vivo total rexistrado e P_e é o peso esperado para a lonxitude total rexistrada [5].

O peso vivo e a lonxitude total foron relacionados mediante a ecuación

$$P_v = aL^b \quad (2.3)$$

empregando o método de mínimos cadrados tras linealizar a curva aplicando o logaritmo as dúas variables. Os valores do coeficiente a e o potencial b (coeficiente angular), empregáronse para calcular o peso esperado (P_e) para ambos sexos combinados e para cada sexo, segundo a lonxitude total, mediante a ecuación 2.3.

2.2.2. Modelos de regresión

Esta Sección levará ao lector por un percorrido dende os modelos de regresión máis simples ata outros máis complexos facendo fincapé nas principais vantaxes e desvantaxes para explicar adecuadamente o comportamento da talla de primeira madurez sexual e o factor de condición relativo da merluza ao longo do período de estudo (1980-2019) para cada sexo e para a combinación de ambos.

Seguindo as palabras de Hastie, Trevor, Tibshirani, Robert and Friedman, Jerome [18]

“Partiremos dos modelos máis simples para poder chegar a comprender o funcionamento dos modelos máis complexos”.

Non obstante, comezarase pola exposición do modelo de regresión máis simple, o Modelo Lineal (2.2.2.1), o cal pese a súa rixidez goza dunha ampla aplicación histórica debido a súa sinxela formulación ata os Modelos Aditivos Xeneralizados (2.2.2.3), (GAM, do inglés “Generalized Additive Models”) que proporcionan unha maior flexibilidade, pasando polo Modelo de regresión Lineal Xeneralizado (2.2.2.2), coñecido como GLM (do inglés, “Generalized Linear Models”).

2.2.2.1. Modelo de regresión lineal

O plantexamento de métodos estatísticos nos que se trata de explicar o comportamento dunha ou varias variables resposta, a través dun conxunto de variables explicativas, require a elección dun modelo que describa a estrutura da relación entre as variables.

Xeralmente o modelo máis empregado é de tipo lineal no que se modeliza a variable ou variables obxectivo, ou algunha característica delas, mediante unha combinación lineal das variables explicativas.

O modelo lineal clásico consiste en expresar a esperanza condicionada da variable resposta Y como a combinación lineal das variables explicativas $\mathbf{X} = (X_1, X_2, \dots, X_p)$

$$E[Y|\mathbf{X}] = \mu(\mathbf{X}) \quad (2.4)$$

$$\mu(\mathbf{X}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (2.5)$$

onde $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ son parámetros descoñecidos. A relación existente entre a variable resposta e as covariables consideradas segue a seguinte formulación $Y = \mu(\mathbf{X}) + \epsilon$ onde ϵ é un erro aleatorio i.i.d (independente e idénticamente distribuído) $N(0, \sigma^2)$.

A pesar de ser un modelo amplamente empregado en diferentes contextos, o uso correcto destes modelos está moi limitado pois só se pode empregar cando a variable resposta é normal. Isto fai que en moitos contextos prácticos estes modelos proporcionen resultados erróneos debido a súa aplicación a datos que non verifican a asunción de normalidade. Por este motivo, aparecen os Modelos Lineais Xeneralizados (GLM, do inglés “General linear model”) que son unha extensión dos modelos de regresión clásicos.

2.2.2.2. Modelo de regresión lineal Xeneralizado

Nos modelos de regresión lineal a variable resposta segue unha distribución normal e as co-variables teñen un efecto lineal sobre a variable resposta. O modelo lineal xeneralizado é moito máis versátil toda vez que non se impón o comportamento normal na distribución. Neste senso, os modelos lineais xeneralizados empréganse para modelizar variables respostas binarias, de recuento, etc.

Cando a variable resposta non sexa gaussiana, non vai ser posible unha conexión directa entre o valor esperado e o predictor lineal, pois o dominio de $E[Y|\mathbf{X}]$ non é a recta real.

Lémbrese ao lector que nun modelo de regresión lineal $E[Y|\mathbf{X}] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p = \mu(\mathbf{X})$. Polo que nun modelo lineal xeneralizado precísase realizar unha transformación, h para expresar o dominio do valor esperado correctamente, é dicir:

$$\mu(\mathbf{X}) = h(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p),$$

o cal pódese rescribir como

$$h^{-1}(\mu(\mathbf{X})) = g(\mu(\mathbf{X})) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p,$$

sendo g a función link ou de enlace, e $\beta_0, \beta_1, \dots, \beta_p$ os parámetros descoñecidos. Así mesmo, nos GLM realízase a suposición de que a variable resposta Y segue unha distribución pertencente á familia de distribucións exponenciais. Neste senso, nestes modelos deberase escoller a función link dependendo da distribución da variable resposta. Cómpre salientar que se a variable resposta segue unha distribución Normal a función link é a identidade, polo tanto, estarase ante un modelo de regresión lineal.

No Cadro 2.1, amósanse as funcións links máis empregadas segundo o tipo de distribución da variable resposta.

Familia	Función link por defecto	Rango de Y	$Var(Y X_1 = x_1, \dots, X_k = x_k)$	Outras funcións link posibles
Gaussiana	Identidad	$(-\infty, +\infty)$	ϕ	Logarítmica
Binomial	logit	$\frac{0,1,\dots,N}{N}$	$\frac{\mu(1-\mu)}{N}$	Logarítmica, probit, cloglog, loglog
Poisson	log	0, 1, 2...	μ	Identidade, raíz cadrada
Gamma	Inversa	$(0, \infty)$	$\phi\mu^2$	Identidade, logarítmica

Cadro 2.1: Función link por defecto, rango da resposta, función da varianza condicionada para varias distribucións da familia exponencial. ϕ é o parametro de dispersión ou escala. Se non se amosa vale 1. $\mu =$ é a media condicionada de y dados os valores das variables predictoras. Na familia binomial N é o número de ensaios.

Estimación dos parámetros do modelo

O axuste dun modelo lineal xeneralizado determínase polo vector $\hat{\beta}$ de estimacións dos parámetros. Para estimar os β 's usarase o método de máxima verosimilitude, no cal se trata de obter os coeficientes que maximizan o logaritmo da función de verosimilitude.

Agora ben, á hora de estimar os coeficientes mediante o método de máxima verosimilitude, atópase o inconveniente de que as ecuacións resultantes do problema de maximización formulado na aplicación deste método non son lineais. A non linearidade das ecuacións imposibilita a resolución exacta das mesmas sendo necesario aplicar algún método iterativo para a súa resolución aproximada. Estes métodos consisten en considerar unha solución inicial $\hat{\beta}^0$ de β e ir dando iterativamente estimacións $\hat{\beta}^1, \hat{\beta}^2, \hat{\beta}^3, \dots$ que deben converxer a un valor $\hat{\beta}$ denominado estimador de máxima verosimilitude de β .

Un dos métodos máis empregados é o método de mínimos cadrados ponderados IRLS (do inglés, "Iteratively Reweighted Least Squares") tamén denominado método das marcas de Fisher (pode verse en [24], Subsección 2.1.2 onde se trata este algoritmo no caso de máis dunha variable resposta). Este método é unha variante do método de Newton-Raphson (para máis información pode consultarse [26], Sección 3.1)

Deseguido enuméranse os pasos do algoritmo:

- **Inicio.** Calcular os coeficientes iniciais $\hat{\beta}_0^0 = h^{-1}(\bar{Y}) = g(\bar{Y}), \hat{\beta}_1^0 = \dots = \hat{\beta}_p^0 = 0$
- **Paso 1.** Obter a variable dependente linealizada, para $j = 1, \dots, n$

$$Z_j = \eta_j(\hat{\beta}^0) + g'(\hat{\mu}_{0j})(Y_j - \hat{\mu}_{0j})$$

e os pesos

$$W_j = \frac{1}{g'(\hat{\mu}_{0j})^2 V(\hat{\mu}_{0j})}$$

sendo

$$\eta_j(\hat{\beta}^0) = \hat{\beta}_0^0 + \hat{\beta}_1^0 X_{j1} + \dots + \hat{\beta}_p^0 X_{jp}$$

e

$$\hat{\mu}_{0j} = g^{-1}(\eta_j(\hat{\beta}^0))$$

e V , unha función de $\hat{\mu}_{0j}$ cuxa definición pode atoparse en [24] Sección 2.1.1, páxina 63. Nótese que no caso de que a distribución sexa gaussiana o seu valor é 1.

- **Paso 2.** Obter os coeficientes actualizados mediante un Modelo Lineal ponderado

$$\hat{\beta} = (\hat{\beta}_0, \dots, \hat{\beta}_p)^t = (X^t W X)^{-1} X^t W Z$$

con

$$X = \begin{pmatrix} 1 & X_{11} & \dots & X_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{pmatrix} \text{ e } W = \text{diag}(W_1, \dots, W_n)$$

- **Paso 3.** Repetir o **Paso 1** e **Paso 2** substituíndo as estimacións iniciais, $\hat{\beta}^0$ polas estimacións $\hat{\beta}$ obtidas no **Paso 2**, ata acadar a converxencia. Un criterio de converxencia que se pode empregar é para o algoritmo cando se verifique que

$$\frac{\|\hat{\beta} - \beta^0\|}{\|\beta^0\|} \leq \varepsilon,$$

sendo $\|\cdot\|$ a norma euclídea.

Convén sinalar que o estimador de máxima verosimilitude minimiza a *deviance*, un concepto de gran importancia para avaliar a calidade do axuste e que se define como segue:

Definición 2.2.2.2.1. *O estatístico deviance obtense mediante:*

$$D = 2\{l(\hat{\beta}_{max}) - l(\hat{\beta})\}\phi = \sum_{j=1}^p 2w_j\{y_j(\tilde{\theta}_j - \hat{\theta}_j) - b(\tilde{\theta}_j) + b(\hat{\theta}_j)\}$$

onde $l(\hat{\beta}_{max})$ indica a verosimilitude maximizada do modelo saturado, tantos parámetros como datos, $l(\hat{\beta})$ é o valor máis alto que a verosimilitude podería ter dados os datos. Con $\tilde{\theta}_j$ e $\hat{\theta}_j$, as estimacións dos parámetros canónicos para os modelos saturado e o modelo de interés respectivamente e ϕ é o parámetro escala.

En definitiva, a deviance é unha medida que avalía en qué medida a verosimilitude do modelo saturado, tantos parámetros como datos, supera á verosimilitude do modelo proposto. Se se realizou un bo axuste debería de obterse unha deviance pequena.

Hipóteses e diagnose do modelo

As **hipóteses** básicas do modelo GLM son:

- **Linealidade:** $g(E[Y|\mathbf{X}]) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$.
- **Distribución:** A distribución condicional da Y será dunha familia de tipo exponencial con esperanza condicional $E[Y|\mathbf{X}] = \mu(\mathbf{X})$.
- **Independencia:** Os residuos do modelo deben ser independentes.

Diagnose do modelo

Nos modelos lineais xeneralizados necesitanse residuos xeneralizados que se poidan aplicar a todas as distribucións que poidan remplacear á normal. Deste xeito, a análise dos residuos nos GLM realízanse mediante os *residuos de Pearson* e os *residuos da deviance*.

Definición 2.2.2.2.2. *Residuos de Pearson:*

$$\hat{\epsilon}_j^p = \frac{y_j - \hat{\mu}_j}{\sqrt{V(\hat{\mu}_j)}}$$

Definición 2.2.2.2.3. *Residuos Deviance*

$$\hat{\epsilon}_j^d = \text{signo}(y_j - \hat{\mu}_j)\sqrt{d_j}$$

onde $d_j = 2[y_j \log(y_j/\hat{y}_j) - (y_j - \hat{y}_j)]$ son os sumandos da deviance [2.2.2.2.1](#)

Para máis información, pódese consultar [\[24\]](#), Sección 2.1.9

Os residuos de Pearson no GLM deben ter media cero e varianza ϕ . Noutros termos, non deben presentar tendencia, nin en media nin en varianza, cando se representan fronte aos valores axustados ou fronte a algunha das covariables do modelo. Debido a que a distribución dos *residuos de Pearson* soe ser asimétrica, na práctica empréganse os *residuos deviance*.

Cráterios de selección e bondade de axuste

Pode ocorrer que haxa varios modelos válidos M_1, \dots, M_m . Para escoller un deles calcúlase a porcentaxe de deviance explicada, a cal ven dada por:

$$D_{\text{explicada}} = \frac{\text{Null deviance} - \text{Res deviance}}{\text{Null deviance}} \times 100$$

onde a “Null deviance” é a “deviance” do modelo que non depende de ningunha variable explicativa, é dicir, o modelo que só inclúe o intercepto para explicar o comportamento da variable resposta e a “Res deviance” é a diferenza entre a “deviance” do modelo que non depende de ningunha variable explicativa, “Null Deviance”, menos a “deviance” do modelo que inclúe as variables explicativas consideradas. Nótese que “Null deviance” - “Res deviance” = “deviance” do modelo que engade variables explicativas.

Este concepto é análogo ao coeficiente de determinación axustado o cal se denota por R^2 . Este coeficiente toma valores entre cero e un e canto máis cerca esté de un, máis próximo estarán os datos á recta axustada (para máis información sobre o R^2 consultar [4], Sección 1.4).

2.2.2.3. Modelos de regresión aditiva xeneralizada

No apartado anterior viuse que o Modelo Lineal Xeneralizado asume que a influencia das variables explicativas sobre a variable resposta é lineal, é dicir:

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

Pode ocorrer que o efecto das variables explicativas sobre a dependente teña forma descoñecida e, nese caso, a estrutura do modelo pode representarse da seguinte forma:

$$\eta = \beta_0 + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) \quad (2.6)$$

onde $f_j(X_j)$ é o efecto parcial suave descoñecido de X_j no predictor. Para evitar problemas de identificación do modelo debemos engadir á formulación do mesmo a condición $E[f_j(X_j)] = 0$. Este novo predictor permite engadir no modelo todo tipo de efectos e relacións non lineais entre as variables.

Estimación

Segundo Hastie, Trevor and Tibshirani, Robert [16]:

“ *A estimación de β e f_1, \dots, f_p lógrase replazando a regresión lineal ponderada na regresión da variable dependente estimada mediante un algoritmo apropiado para axustar un modelo aditivo ponderado*”

Polo tanto, os GAM axustaranse mediante “a maximización da probabilidade penalizada” a cal na práctica conséguese a través de mínimos cadrados penalizados iterativos exposto na estimación do modelo GLM. A modificacións deste método para unha correcta adecuación á estimación dun modelo GAM realízase no **Paso 2** onde se axusta un Modelo Aditivo ponderado en lugar dun Modelo Lineal Ponderado.

Sexa $(X_1, Y_1), \dots, (X_n, Y_n)$ unha mostra de observacións. A estimación do modelo GAM pódese obter mediante a seguinte adaptación do algoritmo iterativo “Fisher scoring”

- **Inicio.** Calcular as estimacións iniciais $\hat{\beta}_0^0 = h^{-1}(\bar{Y}) = g(\bar{Y}) = \bar{Y}$, $\hat{f}_1^0 = \dots = \hat{f}_p^0 = 0$
- **Paso 1.** Calcular, para $j = 1, \dots, n$, as respostas linealizadas

$$Z_j = \hat{\eta}_j^0 + g'(\hat{\mu}_{0j})(Y_j - \hat{\mu}_{0j})$$

e os pesos

$$W_j = \frac{1}{g'(\hat{\mu}_{0j})^2 V(\hat{\mu}_{0j})}$$

sendo

$$\hat{\eta}_j^0 = \hat{\beta}_0^0 + \hat{f}_1^0(X_{j1}) + \dots + \hat{f}_p^0(X_{jp})$$

e

$$\hat{\mu}_{0j} = g^{-1}(\eta_j(\hat{\beta}^0))$$

e V , unha función de $\hat{\mu}_{0j}$ cuxa definición pode atoparse en [24] Sección 2.1.1, páxina 63. Nótese que no caso de que a distribución sexa gaussiana o seu valor é 1.

- **Paso 2.** Axustar un Modelo Aditivo de Z sobre X , ponderado por W , obtendo para $j = 1, \dots, n$ as actualizacións.

$$\hat{f}_1(X_{j1}), \dots, \hat{f}_p(X_{jp})$$

e

$$\hat{\eta}_j = \hat{\beta}_0 + \hat{f}_1(X_{j1}) + \dots + \hat{f}_p(X_{jp}).$$

- **Paso 3.** Repetir o **Paso 1** e o **Paso 2** substituíndo $\hat{\eta}_j^0$ por $\hat{\eta}_j$ ata acadar a converxencia. Por exemplo, cando se verifique $|\hat{\eta}_j - \hat{\eta}_j^0| \leq \varepsilon$.

O procedemento de estimación do modelo aditivo no **Paso 2** depende do número e da natureza das funcións parciais engadidas no modelo. Cando as funcións parciais son paramétricas, $f_j(X_j) = \alpha_j X_j$, o modelo de regresión aditivo resólvese usando mínimos cadrados ponderados pois o modelo GAM é un GLM. Cando ditas funcións non son paramétricas, precísase axustar un Modelo Aditivo ponderado para conseguir as actualizacións \hat{f}_j .

Para a estimación do Modelo Aditivo ponderado presente no **Paso 2** existen diferentes técnicas, este traballo centrase nunha delas que é a metodoloxía Regresión Spline Penalizada. Esta metodoloxía parte da idea de representar cada unha das funcións $f_j(x_j)$ mediante unha combinación lineal dunha base de funcións, concretamente dunha base de B-splines. Non obstante, hai outros sistemas base que dependendo da natureza dos datos mostrais poden ser máis adecuadas como por exemplo, as bases Fourier para datos periódicos.

$$f_j(x_j) = \sum_{k=1}^{K_j} b_{jk}(x_j) \beta_{jk} \quad (2.7)$$

onde $b_{jk}(\cdot)$, $k = 1, \dots, K_j$ son as funcións da base de B-splines considerada, polo cal son funcións coñecidas, K_j é o número de nodos considerado, e $\beta_{jk}(\cdot)$, $k = 1, \dots, K_j$ son parámetros descoñecidos.

Con esta representación, cada función f_j pasa a ter unha forma paramétrica dependendo unicamente dos parámetros β_{jk} , $k = 1, \dots, K_j$. Así, o modelo aditivo do **Paso 2** pode verse como un modelo lineal.

$$\eta_j = \sum_{k=1}^{K_1} b_{1k}(x_{1j}) \beta_{1k} + \dots + \sum_{k=1}^{K_p} b_{pk}(x_{pj}) \beta_{pk} = X_j \beta. \quad (2.8)$$

sendo

$$X_j = (b_{11}(x_{1j}) \quad \dots \quad b_{1K_1}(x_{1j}) \quad \dots \quad b_{p1}(x_{pj}) \quad \dots \quad b_{pK_p}(x_{pj}))$$

e

$$\beta = (\beta_{11} \quad \dots \quad \beta_{1K_1} \quad \dots \quad \beta_{p1} \quad \dots \quad \beta_{pK_p})^t.$$

Con todo o anteriormente dito, pódese considerar empregar as técnicas de estimación do GLM para estimar o modelo. Non obstante, isto presenta o inconveniente de que previamente á aplicación das técnicas expostas para os GLM habería que determinar o número e a posición dos nodos para cada unha das covariables.

Un xeito de evitar o problema que se expón anteriormente é seleccionar un número de nodos suficientemente alto e, controlar o grado de suavización incorporando unha penalización no proceso de estimación dos coeficientes do modelo. Consecuentemente, os coeficientes β estímense minimizando a suma dos residuos cadrados penalizados

$$\sum_{j=1}^n W_j (Z_j - X_j \beta)^2 + \lambda_1 \int (f''(X_1))^2 dx_1 + \dots + \lambda_p \int (f''(X_p))^2 dx_p, \quad (2.9)$$

A variabilidade de cada función suave contrólase penalizando o grado de curvatura da función mediante a súa segunda derivada. Os valores λ_j denomínanse parámetros de suavizado, valores de $\lambda_j \rightarrow \infty$ conduce a unha estimación de liña recta para f , mentres que $\lambda = 0$ da lugar a unha estimación de regresión spline non-penalizada. Polo tanto un valor demasiado pequeno ou demasiado grande do parámetro de suavizado implica que a estimación da función suave \hat{f}_j non está próxima á verdadeira función f_j . Así, faise latente a importancia de determinar o valor adecuado para cada unha das penalizacións á hora de obter bos estimadores para o modelo.

O inconveniente de determinar o número e a posición dos nodos para cada covariable transfórtese agora no problema de estimación do parámetro de suavización. Existen diferentes criterios e métodos de estimación do parámetro de suavizado, sen embargo, hai dúas formas principais de elixir λ automaticamente durante o axuste do modelo. A primeira forma é elixir λ de tal xeito que minimize o erro de predicción do modelo, o cal pódese acadar mediante Validación Cruzada (CV) e a Validación Cruzada Xeneralizada (GCV). O segundo enfoque é tratar a suavidade como un efecto aleatorio, no que λ é agora un parámetro de varianza que se estima empregando a máxima verosimilitude ou a máxima verosimilitude restrinxida.

Neste traballo, empregárase REML proposto por Patterson e Thompson [11] para a selección da suavidade. A estimación por máxima verosimilitude restrinxida trata de reducir o sesgo que teñen os estimadores de máxima verosimilitude das compoñentes da varianza ao non ter en conta os grados de liberdade que se perden ao estimar a media. Este procedemento consiste en factorizar a verosimilitude completa en dúas partes independentes, unha das cales non contén a media, asumindo que ao usar esta parte da verosimilitude non se perde información con respecto a usar a verosimilitud completa. En realidade, a verosimilitude restrinxida, corresponde coa verosimilitude asociada a unha combinación lineal das observacións, cuxa media é nula e cumpre as condicións de ser un factor independente do outro co que se produce a verosimilitude completa e, non supoñer a perda de información con respecto ao uso dos datos orixinais. Para máis información pódese consultar [24] Sección 6.2.5.

Diagnose

Por último, destacar que ao igual que o GLM visto no apartado 2.2.2.2 o modelo será válido se os residuos son normais, homocedásticos e independentes. Se algunha destas hipóteses non se cumpre, deberíase axustar un novo modelo. Sen embargo, ao tratarse de series temporais é posible que a hipótese de independencia non se cumpra. Dito isto, axustárase un proceso $ARMA(p, q)$ sobre os mesmos. Deste xeito, conseguírase modelizar a dependencia temporal existente entre a variable resposta e as covariables. A continuación na 2.2.3 preséntanse os modelos Box-Jenkins para proporcionar ao lector un mellor entendemento dos fundamentos de dita metodoloxía de cara a realizar unha correcta aplicación da mesma nos GAM presentados neste apartado coa finalidade de acadar os obxetivos plantexados.

2.2.3. Modelos Box-Jenkins

No presente apartado considérase un conxunto de variables aleatorias Z_t onde t toma valores nun conxunto C , sendo este un conxunto ordenado segundo os instantes temporais (días, meses, anos...). Entón, para cada valor t do conxunto C terase unha variable aleatoria, Z_t , e o conxunto de valores observados en cada instante formarán unha serie temporal (z_1, z_2, \dots, z_t) .

Este apartado estrutúrase en torno ao proceso de construción dun modelo válido baixo a metodoloxía Box Jenkins. Deste xeito, nas Seccións 2.2.3.1 e 2.2.3.2 amósanse os modelos ARMA e ARIMA que permiten a modelización de múltiples series temporais. Na Sección 2.2.3.3 preséntanse as ferramentas necesarias para identificar os ordes dos modelos ARMA e ARIMA. Na Sección 2.2.3.4 preséntase de forma sucinta distintos métodos para a estimación dos parámetros así como distintos criterios para tomar unha decisión sobre o modelo que finalmente se empregara ante a dúbida de seleccionar un modelo entre varios e, para finalizar, preséntanse as técnicas necesarias para a comprobación das hipóteses realizadas sobre o modelo, é dicir, expomos o procedemento de diagnose do modelo.

2.2.3.1. Modelos para procesos estacionarios

Unha serie temporal é estacionaria se as súas propiedades estatísticas permanecen constantes ao longo do tempo. Se unha serie é estacionaria (en media y varianza) cúmplase que:

$$\bullet \mu_t = \mu = cte \quad \bullet \sigma_t = \sigma = cte \quad \bullet \gamma(t, t_k) = E\{(z_t - \mu)(z_{t_k} - \mu)\} = \gamma_k$$

Os procesos ARMA (p, q) son froito da combinación dun proceso autorregresivo de orden (p) , AR (p) os cales explican o valor actual da variable resposta, como unha función lineal de p valores anteriores ao actual e un proceso de medias móbiles de orden q , MA (q) , cuxo valor actual depende das q últimas innovacións. Entón, unha serie Z_t segue un proceso ARMA (p, q) se:

Definición 2.2.3.1.1.

$$Z_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t + \theta_1 a_{t-2} + \dots + \theta_q a_{t-q} \quad (2.10)$$

onde $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ son constantes.

2.2.3.2. Modelos para procesos non estacionarios

O modelo ARMA que se expuxo con anterioridade pertence a unha familia moi flexible de procesos estacionarios, xa que calquera proceso estacionario ou ben é lineal ou ben transfórmase para que o sexa (“Descomposición de Wold”, pode verse en Peña (2010) [22], Sección 2.3). Cómpre sinalar que non abundan series reais xeradas por procesos estacionarios, pois soen presentar compoñentes determinísticas como:

- **Heterocedastidade:** A variabilidade da serie non é constante no tempo senon que aumenta ou diminúe ao facelo no nivel da serie. É dicir:

$$\sigma_t = cf(\mu_t)$$

onde c é unha constante e f unha función da media. Se consideramos como función $f(\mu_t) = \mu_t^{1-\lambda}$, situación moi usual na práctica, unha familia de transformacións moi útiles é a familia de transformacións Box-Cox introducida por George E.P. Box e David Cox no ano 1964 [6].

- **Tendencia:** O nivel da serie non é estable no tempo.

Sexa o proceso $Z_t = \beta_0 + \beta_1 t + V_t$, un proceso non estacionario pola presenza de tendencia determinista. Para eliminar a tendencia aplícaselle á serie d diferenciacións regulares. Se despois de diferenciar regularmente a serie non é estacionaria vólvese a diferenciar a serie ata que esta sexa estacionaria. En xeral, diferencíase a serie como moito 3 veces, $d \geq 3$. Unha vez que a serie é estacionaria pódese modelizar mediante un ARMA (p, q) . A este proceso denomínaselle, ARIMA (p, d, q) .

Un proceso ARIMA (p, d, q) é aquel proceso que, despois de aplicarlle d diferenzas regulares, convértese nun proceso ARMA (p, q) . É dicir:

$$\{Z_t\}_{t \in R} \text{ é ARIMA } (p, d, q) \iff (1 - B)^d Z_t \text{ é ARMA } (p, q)$$

onde B correspóndese co operador retardo nido por $BZ_t = Z_{t-1}$

Os procesos ARIMA que acabamos definir permiten captar non estacionariedades provocadas pola presenza de tendencia pero non capturan non estacionariedades provocadas pola presenza de compoñente estacional. Por este motivo seguidamente procedemos á ampliación da clase de procesos ARIMA presentada co obxectivo de que a nova clase capture tanto non estacionariedades provocadas pola presenza de tendencia como de compoñente estacional.

- **Compoñente estacional:** A serie presenta patróns repetitivos.

Os modelos ARMA expostos ata o momento só modelizan dependencia entre observacións consecutivas acontecidas no pasado inmediato (dependencia regular), polo cal é necesario presentar os modelos ARMA estacionais os cales permiten modelizar dependencia entre observacións acontecidas en instantes separados por múltiplos do período estacional s , é dicir, dependencia estacional.

Sexa o proceso $Z_t = S_t + V_t$, onde $\{S_t\}_{t \in R}$ non é estacionario pero $\{V_t\}_{t \in R}$ sí, e $S_t = S_{t-s}$ correspóndese cunha tendencia estacional determinista. Entón o proceso diferenciado $Z_t - Z_{t-s} = V_t - V_{t-s}$ é estacionario. Polo tanto, se unha serie presenta dependencia estacional, ao aplicarlle D diferenzas estacionais convértese nun proceso ARMA (p, q) . É dicir:

$$\{Z_t\}_{t \in R} \text{ é ARIMA } (P, D, Q) \iff (1 - B^s)^D Z_t \text{ é ARMA } (P, Q)_s$$

con operador retardo estacional $B^s Z_t = Z_{t-s}$

- **Tendencia e compoñente estacional:** A serie presenta tendencia e patróns repetitivos.

Se as series se viran afectadas pola tendencia como pola dependencia estacional, o proceso que nos permite modelar tal situación corresponderíase cun proceso ARIMA estacional multiplicativo ARIMA $(p, d, q) \times (P, D, Q)_s$.

O proceso ARIMA estacional multiplicativo é posiblemente o máis empregado na modelización de series temporais univariantes pola súa flexibilidade para axustar múltiples series temporais.

2.2.3.3. Identificación

No apartado anterior amósanse dous tipos de procesos estocásticos claramente diferenciados, os procesos estacionarios dentro dos cales destacamos a familia dos procesos ARMA, e os procesos non estacionarios onde vimos a familia dos procesos ARIMA. O seguinte obxectivo é dotar ao lector dos coñecementos precisos para identificar a partires dunha serie real un destes procesos como posible xerador da mesma.

Para iso será necesario definir os conceptos de: media mostral, función de autocovarianzas da mostra, función de autocorrelacións simples da mostra e función de autocorrelacións parciais da mostra.

Definición 2.2.3.3.1. *Sexa $\{Z_t\} = (z_1, z_2, z_3, \dots, z_T)$ unha serie temporal, con $t = 1, \dots, T$. Defínese a media mostral como:*

$$\bar{z} = \frac{1}{T} \sum_{t=1}^T z_t$$

Definición 2.2.3.3.2. *Sexa $\{Z_t\} = (z_1, z_2, z_3, \dots, z_T)$ unha serie temporal con $t = 1, \dots, T$. Defínese a función de autocovarianzas mostrais entre observacións separadas por k instantes temporais como:*

$$\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (z_t - \bar{z})(z_{t-k} - \bar{z})$$

con $\hat{\gamma}_k = \hat{\gamma}_{-k}$, para $k = 0, 1, \dots, T - 1$

Definición 2.2.3.3.3. *Sexa $\{Z_t\} = (z_1, z_2, z_3, \dots, z_T)$ unha serie temporal, con $t = 1, \dots, T$. A función de autocorrelacións simples mostrais (FAS) asigna a cada retardo k o valor:*

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

Definición 2.2.3.3.4. *Sexa $\{Z_t\} = (z_1, z_2, z_3, \dots, z_T)$ unha serie temporal con $t = 1, \dots, T$. A función de autocorrelacións parciais da mostra (FAP) asigna a cada retardo k o valor:*

$$\hat{\alpha}_k = \hat{\alpha}_{kk}$$

onde $\hat{\alpha}_{kk}$ correspóndese co estimador mínimo cuadrático de α_{kk} na regresión $z_t = \alpha_{k0} + \alpha_{k1}z_{t-1} + \dots + \alpha_{kk}z_{t-k} + \varepsilon$.

Unha vez presentados os conceptos previos, os pasos a seguir para a identificación das ordes do proceso xerador da serie correspondente con:

1. Se a serie presenta heterocedasticidade, eliminala mediante a transformación Box-Cox.
2. Se a serie (pode que transformada no Paso 1) presenta tendencia, eliminala a través da diferenciación regular.
3. Se a serie (pode que transformada nos Pasos 1 e 2) presenta estacionalidade, eliminala mediante a diferenciación estacional.
4. A través das funcións de autocorrelación simple e parcial da serie temporal, identificar un modelo ARMA seguindo a Táboa 2.2.

Proceso	FAS $\rho(k)$	FAP $\sigma(k)$
$AR(p)$	Moitos coeficientes non nulos	Anúlase para os retardos maiores que
$MA(q)$	Anúlase para os retardos maiores que	Moitos coeficientes non nulos
$ARMA(p, q)$	Moitos coeficientes non nulos	Moitos coeficientes non nulos*
$AR(P)_s$	Moitos coeficientes non nulos*	Anúlase para os retardos maiores que $(Q)_s$
$MA(Q)_s$	Anúlase para os retardos maiores que Q_s	Moitos retardos non nulos*
$ARMA(P, Q)_s$	Moitos coeficientes non nulos*	Moitos coeficientes non nulos*

Cadro 2.2: Comportamentos das FAS e as FAP mostrais para a identificación de procesos. Con * indícase que os coeficientes nos múltiplos de P (modelo AR), de Q (modelo MA) e o de p (parte AR do modelo ARMA) o q (parte MA do modelo ARMA), son nulos

2.2.3.4. Selección do modelo

Unha vez tense o modelo posible xerador da serie, estímase os parámetros do mesmo. Dado que non é obxecto de estudo neste traballo estimar os parámetros dun modelo Box-Jenkins, remítese ao lector a [12], Sección 10.4 por se fose de interese.

O habitual é que se atope un conxunto de modelos M_1, \dots, M_m válidos. Neste senso, sería desexabel seleccionar o mellor modelo que explique a serie observada. Dito isto, defínese k como a cantidade de coeficientes dun modelo ARMA (p,q) , é dicir, $k = p + q + 1$ o $k = p + q$ se o ARMA ten ou non constante, respectivamente e, por φ_{k+1} o vector formado polas estimacións de máxima verosimilitude de k coeficientes e de σ_a^2 .

No caso de ter varios modelos, proponse seleccionar aquel modelo que minimize o Criterio de Información de Akaike [?] (AIC, do inglés “Akaike Information Criterion”):

$$AIC = -2\log(L(\hat{\varphi}(p, q))) + 2k.$$

No caso de que a diferenza dos AICs sexa menor de dúas unidades, seleccionárase o modelo que teña un menor número de parámetros ou o que sexa máis doado de explicar. Para máis información sobre o criterio citado anteriormente pódese consultar [12], Sección 10.7.

2.2.3.5. Diagnose do modelo

Tras axustar o modelo ARMA comprobarase se o modelo é válido. Para que o modelo sexa válido, a diagnose do modelo Box Jenkins require verificar que as innovacións, a_t , sexan ruído branco, é dicir os erros deben ter media cero, varianza constante, falta de correlación para calquera retardo e seguir unha distribución normal. Se ben a hipótese de normalidade non invalidaría o modelo, baixo normalidade, a incorrelación equivale a independencia, polo que non estaríamos deixando información por modelizar. Se o modelo que se axusta non cumpre ningunha das hipóteses mencionadas, descartárase este modelo como posible xerador da serie temporal e terase que voltar a axustar un modelo ata que sexa válido.

2.2.3.6. Valores atípicos

Con moita frecuencia ocorren nas series reais feitos puntuais que descoñecemos. As observacións afectadas por estas intervencións poden presentar unha estrutura distinta das demais e aparecer

como valores atípicos, é dicir, datos que aparentemente non foron xerados igual que os demais. O correcto tratamento de valores atípicos é moi importante pois estes valores poden sesgar a estimación dos parámetros.

Consideramos dous tipos de valores atípicos, aditivos e innovativos. Diremos que a serie temporal presenta un atípico aditivo (AO) no instante h se o valor da serie temporal nese instante foi xerado de maneira distinta ao resto. O modelo que segue unha serie observada, z_t , afectada por un atípico aditivo no instante h é.

$$Z_t = \begin{cases} Y_t & t \neq h \\ Y_t + \omega_a & t = h \end{cases}$$

onde y_t é a serie non contaminada por atípicos que se supón que segue un modelo ARIMA, ou que indicamos como $Y_t = \psi(B)a_t$. Entón, o modelo que segue a serie observada, é:

$$Z_t = \omega_A I_t^h + \psi(B)a_t,$$

Onde $I_t^{(h)} = 0$ para $t \neq h$ e $I_h^{(h)} = 1$.

Por outra banda, a serie temporal presenta un atípico innovativo (IO) no instante h cando o valor da innovación nese punto está afectado por unha cantidade descoñecida debido a un suceso imprevisto. O modelo que segue unha serie temporal que presenta un atípico innovativo de magnitude w_I no instante h é.

$$Z_t = \psi(B)(\omega_I I_t^{(h)} + a_t)$$

Na práctica, a natureza e a posición dos valores atípicos que poden aparecer na serie temporal é descoñecida. Por este motivo precísase un procedemento para identificar e clasificar estes valores. O modo de proceder consiste en:

1. Detectar o momento de aparición do dato atípico.
2. Identificar o tipo de dato atípico.
3. Estimar a magnitude do mesmo.

Os detalles necesarios para levar a cabo cada unha das etapas enunciadas pódense atopar en Peña (2010) [12], Sección 13.6 .

2.2.4. Aplicación das series temporais nos GAM

Ao estar empregando series temporais, pode pasar que os residuos do modelo GAM estén correlados. Neste caso proporase un modelo ARMA e axustaríanse conxuntamente os parámetros da regresión e do ARMA. Isto lévase a cabo empregando a función `gamm` do paquete `mgcv` de Wood [25]. A continuación ilústrase un exemplo para proporcionar ao lector unha mellor comprensión desta función. Para elo, consideraránse como variables unha única variable explicativa e unha variable resposta, as cales se amosan na Figura 2.2.

Axústase un GAM, utilizando como base suavizadora a cubic regression spline e como método de estimación dos parámetros o REML. Unha vez axustado o modelo e tras realizar os contraste de significación dos parámetros do modelo, os cales resultan significativos a un nivel de significación do 5% , procederase a realizar o contraste das hipóteses básicas do modelo:

- Contraste de media 0 de Student's t-Test sobre os residuos.
- Contraste de incorrelación de Ljung-Box para os 30 primeiros retardos temporais.
- Contraste de normalidade de Jarque bera e Shapiro-Wilk.

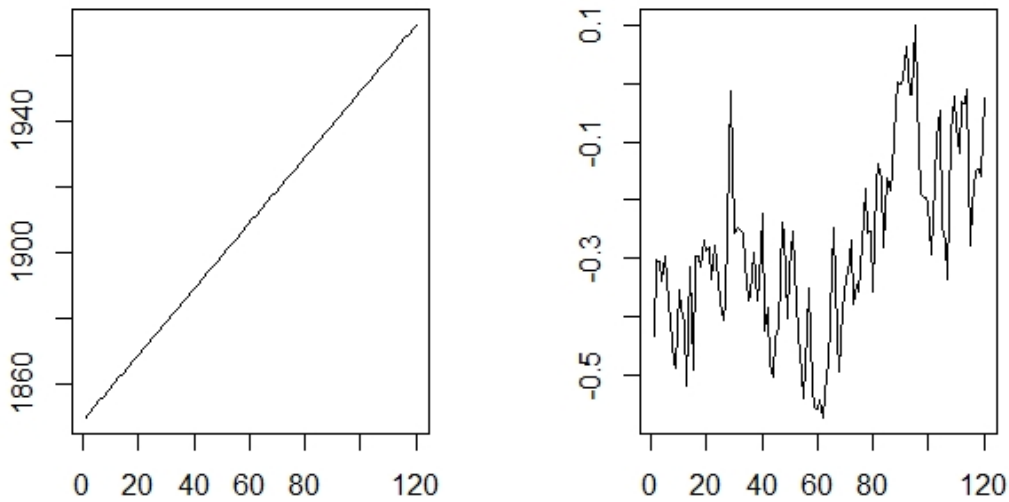


Figura 2.2: Series temporais: variable explicativa (esquerda) e variable resposta (dereita)

- Contraste de estacionariedade de Dickey-Fuller mediante a función `adf.test()` do paquete `TSA` [9] de R. Cómpre salientar que a hipótese nula é que os datos non son estacionarios fronte á hipótese alternativa de que sí o son.

Considerando como nivel de significación $\alpha = 0.05$ os residuos seguen unha distribución normal tanto para o test de Jarque Bera como o de Shapiro-Wilk que se realizaron mediante as funcións `jarque.bera.test` e `shapiro.test`. Ademais os residuos teñen media cero e son estacionarios (ver Cadro 2.3). En canto á hipótese de independencia contrastase mediante o test de Ljung-Box para cada os primeiros 30 retardos da serie. No gráfico 2.3 represéntase os p -valores do estatístico de contraste de Ljung-Box para distintos retardos temporais, nela obsérvase que a partir do quinto retardo a hipótese de incorrelación deixa de cumprirse. Non obstante, os residuos do modelo están correlados xa que baixo normalidade a independencia equivale a incorrelación. A estrutura de correlación dos residuos do modelo obsérvase no gráfico 2.4

Contraste	Hipóteses	P -valor
Jarque-Bera	Normalidade	0.16
Shapiro-Wilk	Normalidade	0.10
Dickey-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 2.3: Contrastos sobre os residuos do modelo

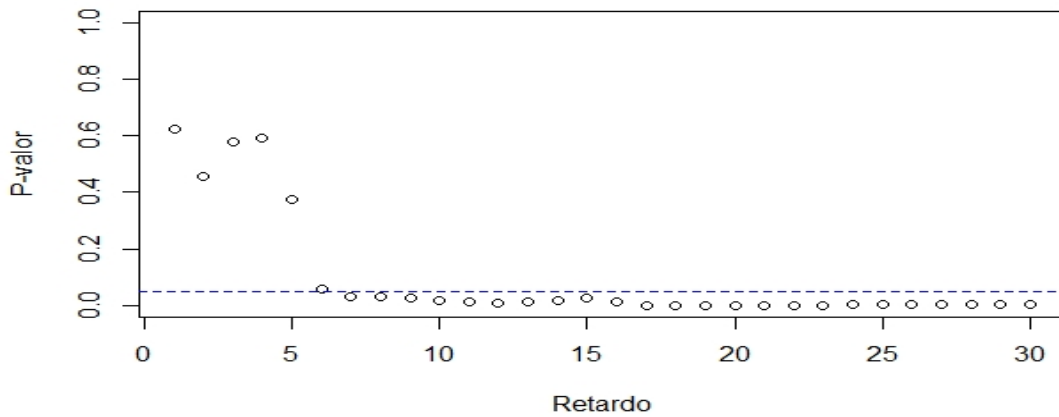


Figura 2.3: P -valores do contraste de Ljung-Box para os diferentes retardos dos residuos do modelo.

En definitiva, o modelo estimado non é válido sendo necesario a búsqueda dun modelo $ARMA(p, q)$ para modelizar a estrutura de dependencia temporal presente nos residuos do modelo. Para determinar que tipo de modelo de series de tempo podería ser adecuado para os residuos, lévase a cabo unha análise dos gráficos dos residuos mediante as funcións de autocorrelación simple e parcial tal e como se menciona na Sección 2.2.3.3 e, mediante a función `auto.arima()` do paquete `forecast` de Hyndman et al. (2019). Observando os gráficos da fas e fap dos residuos do modelo GAM, Figura 2.4, pódese suxerir que os residuos seguen un modelo $AR(1)$.

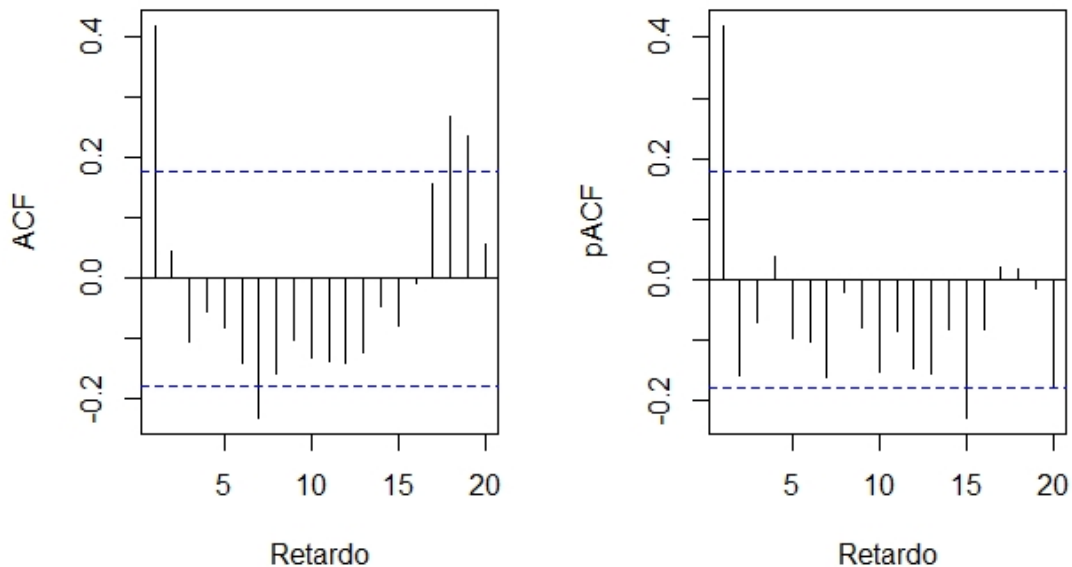


Figura 2.4: Gráficos de correlacións simples e parciais da serie temporal.

Para a construción do modelo empregárase a función `gamm`, mencionada anteriormente, indicándolle que os residuos seguen un $AR(1)$, o modelo sería:

```
mod = gamm(y ~ s(x, bs = "cr"), method = "REML", correlation = corARMA(form=~x ,p = 1, q = 0))
```

Os contrastes dos residuos do modelo final pódense observar no Cadro 2.4 e na gráfica 2.5. Se se toma como nivel de significación $\alpha = 0.05$, obsérvase que os residuos son ruído branco pois son independentes, estacionarios, teñen media cero e seguen unha distribución normal.

Contraste	Hipóteses	<i>P</i> -valor
Jarque-Bera	Normalidade	0.67
Shapiro-Wilk	Normalidade	0.61
Dickey-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 2.4: Contrastos sobre os residuos do modelo.

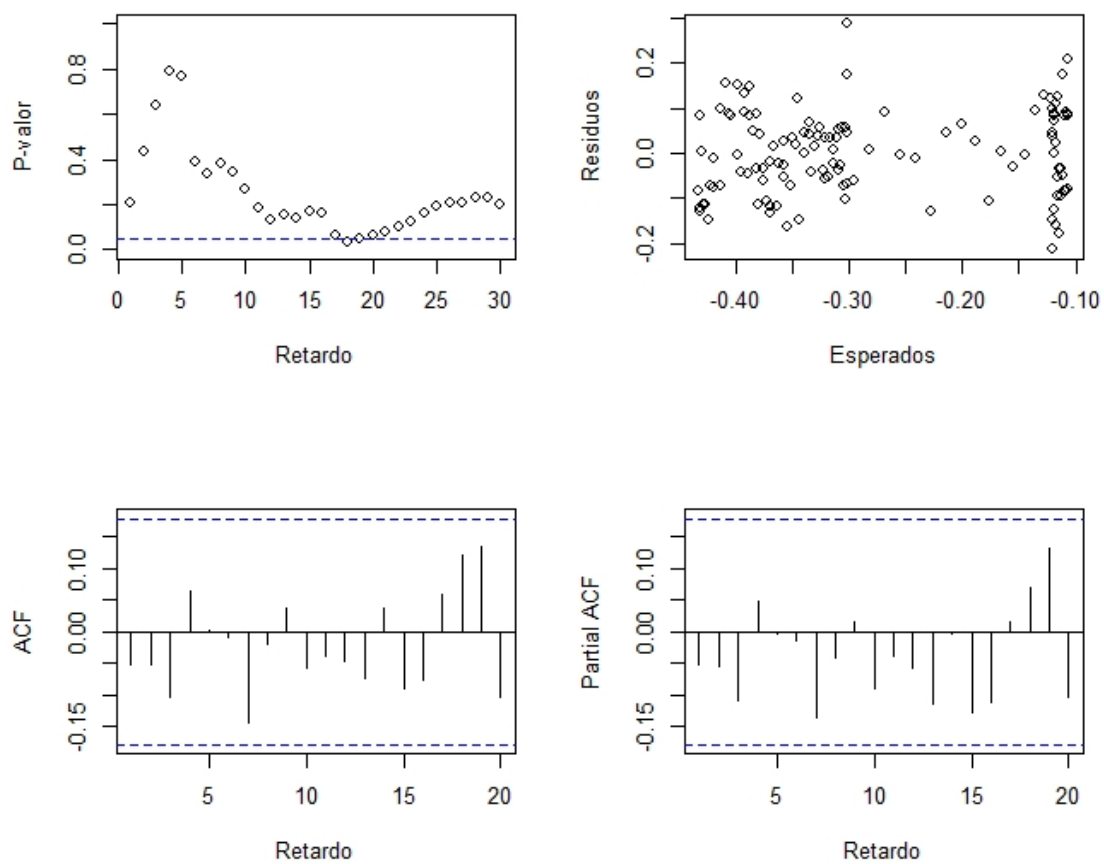


Figura 2.5: Gráficos de contrastes das hipóteses básicas do modelo sobre os residuos. A gráfica superior esquerda, correspóndese cos p -valores do contraste de independencia Ljung-Box e a superior á dereita ven a sinalar a homocedasticidade dos residuos do modelo. As gráficas inferiores representan as correlacións simples e parciais dos residuos.

Capítulo 3

Análise de resultados

Este capítulo xira en torno a presentar os resultados en liña a dar resposta aos obxectivos que guían este estudo. Neste senso, este capítulo consta de tres partes, a primeira na que se describe a talla e o peso vivo da merluza europea en aras de proporcionar unha mellor comprensión destas variables biométricas, as cales empréganse posteriormente para estimar os rasgos de vida, a talla de primeira madurez sexual e o factor de condición relativo, variables resposta deste estudo que se abordan na segunda parte deste capítulo. Finalmente, preséntanse os modelos que explican a variabilidade anual destes rasgos de vida con respecto a variables ambientais e biolóxicas para ambos sexos combinados e para para cada sexo individualmente.

3.1. Análise descriptivo: Unha primeira aproximación

No presente apartado preséntase a análise da talla (cm) e do peso vivo (gr) da merluza europea, variables coas que se estiman os rasgos de vida que definen a dinámica poboacional do recurso.

3.1.1. Talla da merluza europea

A análise de tallas da merluza europea amosa que en promedio a merluza capturada nas áreas 8c e 9a ten unha lonxitude de 39.62 cm, nun rango de tallas de 8 a 91 cm (Cadro 3.1). De xeito máis específico, observouse que as femias amosaron tallas superiores con respecto aos machos (Figura 3.1). O promedio de tallas dos machos foi de 35.46, cun rango de tallas de 8 e 66 cm. O promedio das femias 42.9, cun rango de tallas de 11 a 91 cm.

Medida	Xeneral	Macho	Femia
Media	39.62	35.46	42.9
Mínimo	8	8	11
Máximo	91	66	91
N	27629	12203	15489

Cadro 3.1: Promedio, mínimo, máximo e N das tallas (cm) da merluza europea.

Para determinar se as tallas da merluza varían significativamente segundo sexo, aplicouse a proba de Mann-Whitney empregando a función `wilcox.test` do paquete `stats` de R [13]. Os resultados da proba indican que sí existen diferenzas significativas entre as tallas de machos e femias da merluza nas áreas de muestreo sinaladas durante os anos de prospección realizados, sendo as femias máis grandes ($W = 127505657$, $p\text{-valor} < 0.05$).

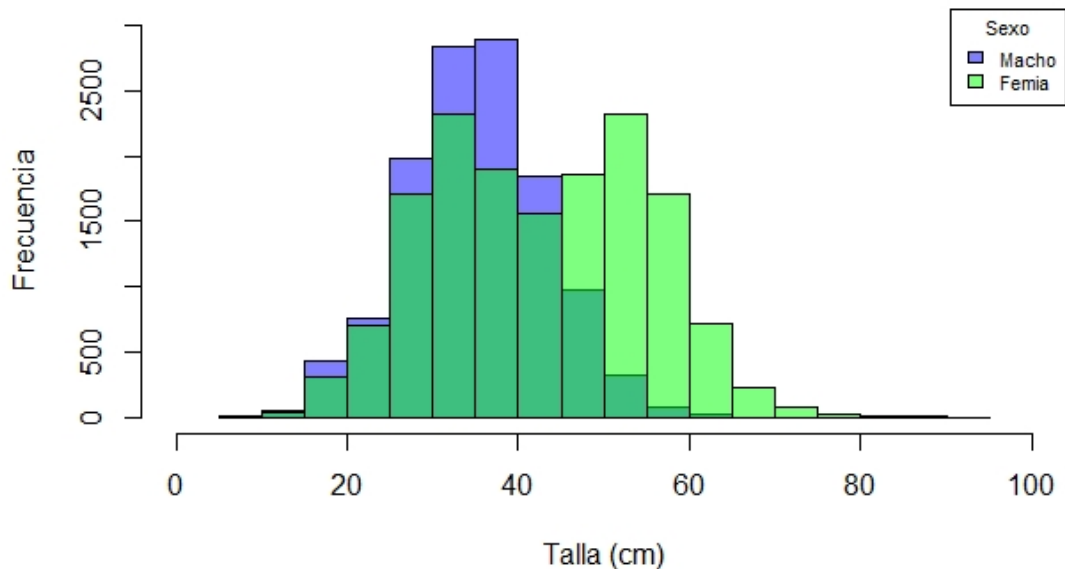


Figura 3.1: Histograma de frecuencias de tallas por sexos da merluza europea.

A análise anual da talla dos datos mostrexados de merluza (ambos sexos combinados) apunta a unha tendencia crecente a partir do ano 1993 acadando o seu máximo no ano 2018 cun promedio de 49.56 cm (ver Figura 3.2). Desagregando esta información segundo sexo, observouse a mesma estrutura en relación á variación anual da talla para a poboación de merluza (ver Figura 3.3), unha tendencia crecente a partir do ano 1993 acadando as femias o seu máximo no ano 2015 cun promedio de 54.63 cm e os machos no ano 2014 cun valor promedio de 43.46 cm. O feito de que se observe unha tendencia crecente da talla a partir do ano 1993 débese a que, a partir deste ano, máis do 50 % dos individuos mostrexados son maduros (Figura 1 na Sección 1 do Apéndice).

Para determinar se as tallas das merluzas varían significativamente entre os anos, aplicouse unha proba de Kruskal-Wallis a través da función `kruskal.test` do paquete `stats`, empregando todas as tallas da poboación (ambos sexos combinados) e posteriormente separados segundo sexos. Os resultados destas probas indican que existen diferenzas significativas entre as tallas de machos e femias da merluza nas augas Ibéricas Atlánticas para o período de estudo analizado ($[X_{39}]^2 = 3823.2, p\text{-valor} < 0.05$; $[Machos]X_{39}^2 = 1385.5, p\text{-valor} < 0.05$; $[Femias]X_{39}^2 = 2326, p\text{-valor} < 0.05$).

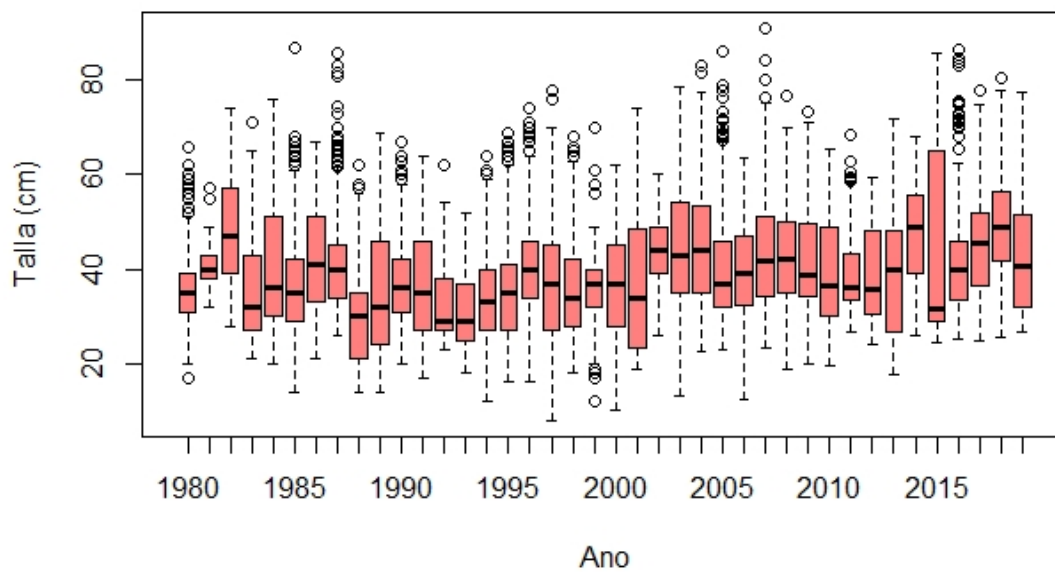


Figura 3.2: Variación anual da talla das mostras de merluza ao longo do período de estudo.

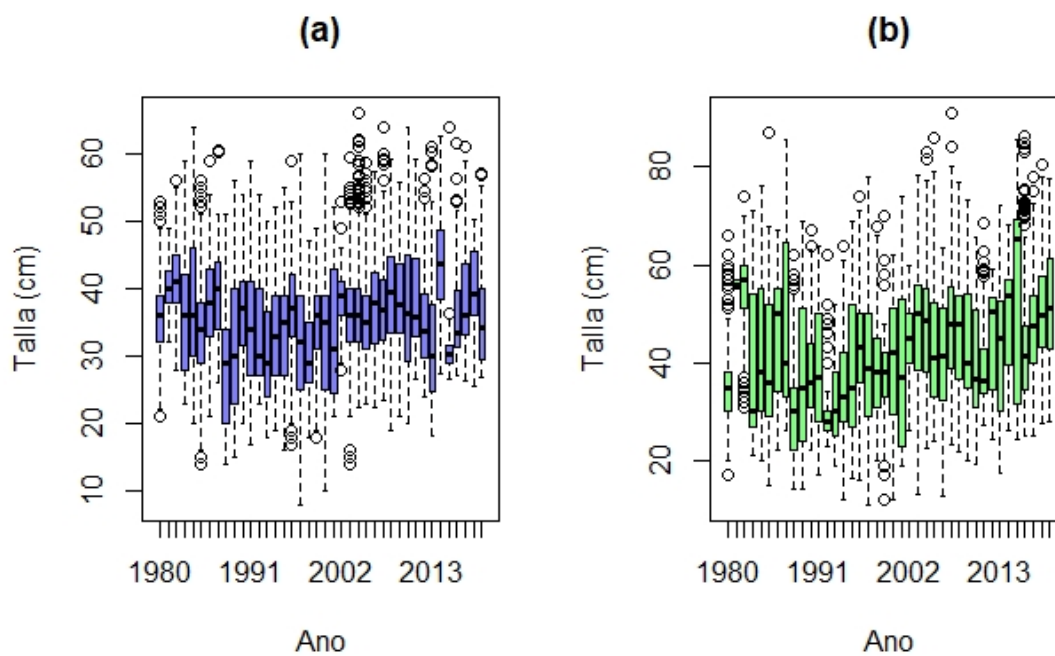


Figura 3.3: Variación anual da talla da merluza europea ao longo do período de estudo para (a) machos e (b) femias.

3.1.2. Peso vivo da merluza europea

Antes de proceder á presentación da análise do peso vivo para a merluza europea lémbrese ao lector que o peso vivo é o peso total do individuo. Neste senso, non é baladí que as merluzas femia pesen máis que os machos pois acadan tallas maiores e as gónadas das femias son máis pesadas en comparación cos individuos macho. Tendo en conta isto, non ten cabida realizar unha análise comparada do peso vivo segundo sexo.

O peso vivo promedio observado durante o período de estudo é de 572.7 gr cun rango de 15.5 a 7120 gr. Para as merluzas femia o promedio do peso vivo é de 739.0449 gr nun rango de 15.5 a 7120 gr, mentras que o peso promedio dos machos é de 366.6190 gr nun rango de 17.7 a 217.0 gr (ver Cadro 3.2).

Medida	Xeneral	Macho	Femia
Media	572.7	366.6190	739.0449
Mínimo	15.5	17.7	15.5
Máximo	7120	2170	7120
N	24415	10906	13509

Cadro 3.2: Promedio, mínimo, máximo e N dos pesos (gr) da merluza europea.

A análise anual do peso vivo das mostras de merluza apunta a unha tendencia crecente a partir do ano 1994. Cómpre destacar que no ano 1998 o peso vivo acada o seu máximo cun valor promedio de 1865.70 gr (ver Figura 3.6), isto débese a que neste ano só se dispón de información de merluzas femia clasificadas como maduras cun tamaño de mostra de 15. Así mesmo, non escapa ao lector o vacío temporal dende o ano 1999 ata o 2001 que se amosa na Figura 3.6 pola falta de información desta variable biométrica nestes anos.

Desagregando esta información segundo sexo (ver Figura 3.5), se ben as merluzas femia seguen a mesma estrutura que a poboación de merluza, unha tendencia ao aumento do peso vivo a partir do ano 1994, o peso vivo das merluzas macho fluctúa arredor da media sen amosar ningunha tendencia ao longo do período de estudo. Doutra banda, ao igual que no total, para as femias non se dispón de información do peso vivo dende o ano 1999 ata o 2001 e, para os machos, dende o 1998 ata o 2001. A non dispoñibilidade de información do peso vivo para ambos sexos combinados e para cada sexo nos anos sinalados previamente, impediu estimar o factor de condición relativo (Sección 3.2.2) para a poboación de merluza e para as merluzas femia dende o 1999 ata o 2001 e, para os machos dende o 1998 ata o 2001.

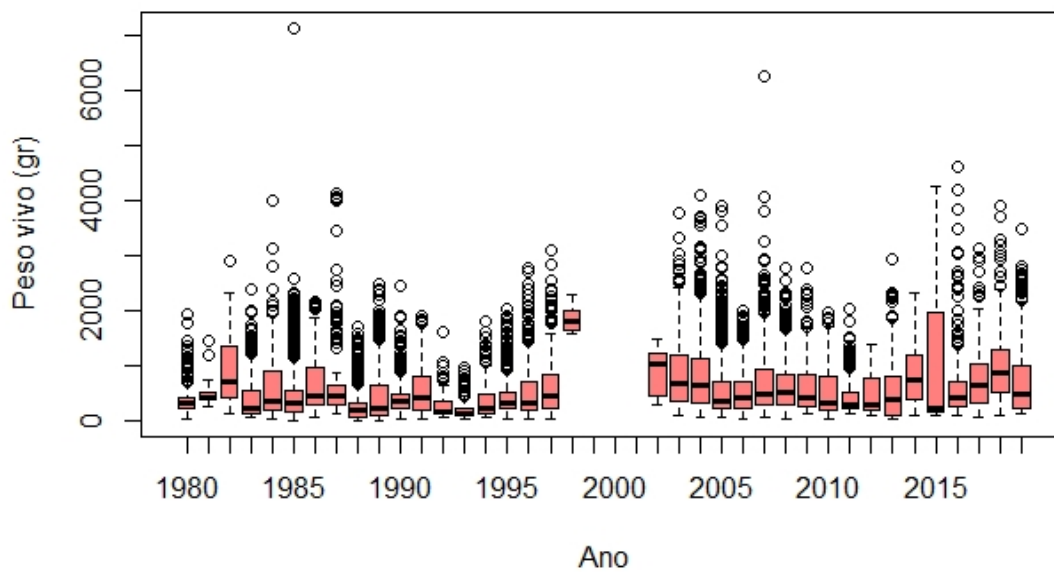


Figura 3.4: Variación anual do peso vivo para a poboación de merluza.

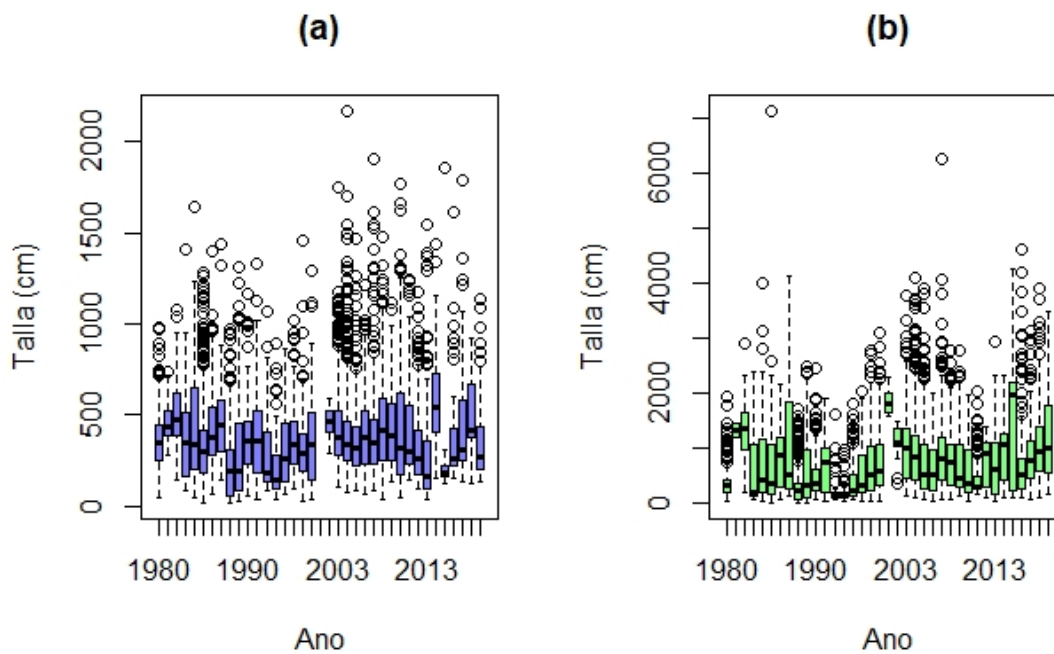


Figura 3.5: Variación anual do peso vivo das merluzas (a) macho e (b) femia.

3.2. Rasgos de vida

Esta Sección gravita en torno a presentar os rasgos de vida que se abordan neste traballo, a talla de primeira madurez sexual e o factor de condición relativo, respectivamente. Tal e como se mencionou anteriormente, os rasgos de vida permiten coñecer a dinámica poboacional dun recurso para deseñar modelos de avaliación eficientes. Neste senso, a talla de primeira madurez sexual é unha das informacións máis importantes que se necesitan coñecer dunha poboación, sendo clave na avaliación e xestión dos recursos explotados comercialmente ao empregarse como punto de referencia para establecer a talla mínima de captura. Así mesmo, o factor de condición relativo emprégase amplamente como índice indirecto do crecemento dunha poboación, asumindo un mellor estado fisiolóxico naqueles individuos con valores máis elevados.

3.2.1. Talla de primeira madurez sexual

A aplicación da función `gonad.mature` do paquete `sizeMat` á información da merluza europea no período 1980-2019, cun tamaño de mostra de 27151 exemplares, permite estimar unha talla de madurez sexual para a poboación de merluza de 37.8 cm de lonxitude. Así mesmo, desagregando esta información segundo sexo, a L_{50} das merluzas femia toma un valor superior ao correspondente nos individuos machos (44.6 e 32.2 cm, respectivamente).

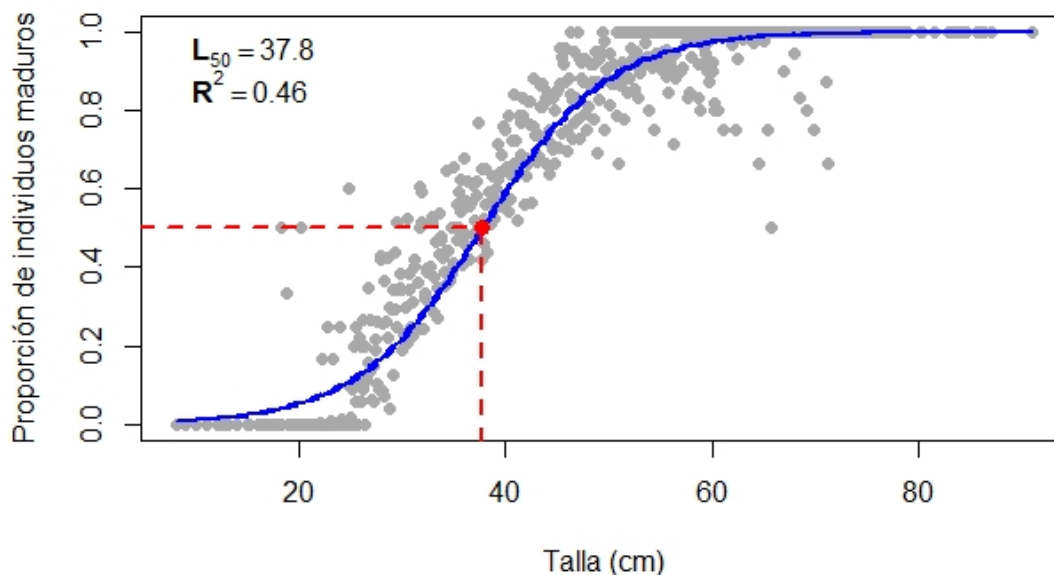


Figura 3.6: Talla de primeira madurez sexual para a poboación de merluza. Os puntos grises representan as observacións, a liña azul e o modelo axustado e a liña vermella representa a talla a cual o 50% da poboación está madura.

Para estudar se existen diferenzas significativas entre os valores da talla de primeira madurez sexual segundo sexo, estudamos se os parámetros da regresión loxística difiren significativamente. Para isto, emprégase a función `Anova` do paquete `car` [14], concluíndo que sí existen diferenzas significativas entre a talla de primeira madurez sexual da merluza segundo sexo ($X^2 = 63930$, p -valor < 0.05).

As estimacións anuais da talla de primeira madurez sexual da merluza europea (Figura 2), amosa que entre 1980 e 1992 os valores da L_{50} presentan variacións relativamente pequenas en

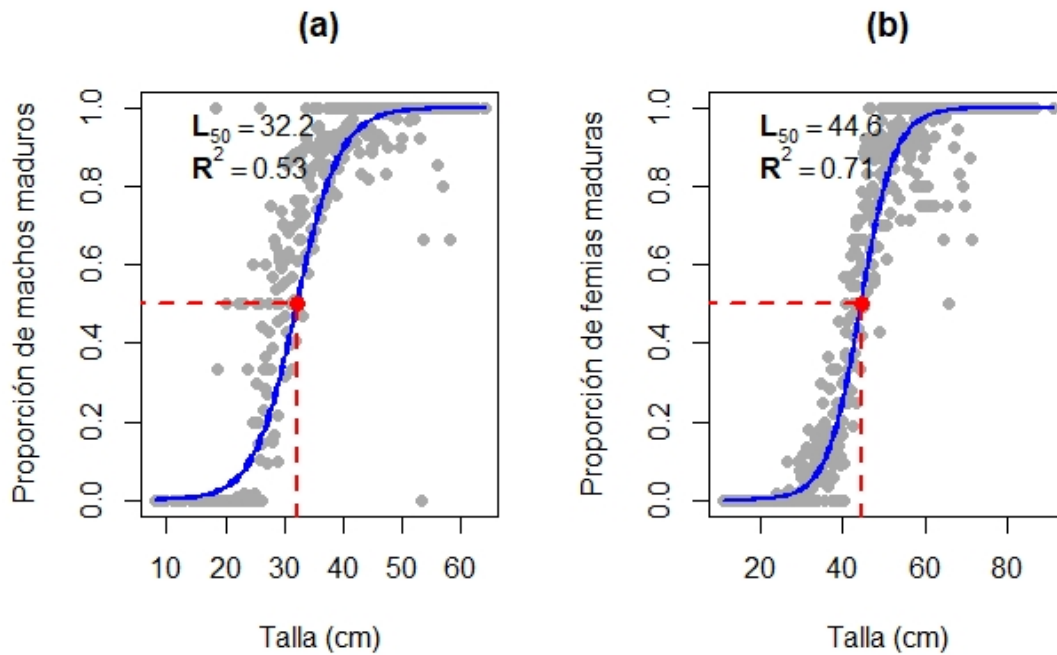


Figura 3.7: Talla de primeira madurez sexual para (a) machos e (b) femias da merluza europea para o período de estudo.. Os puntos grises representan as observacions, a linea azul e o modelo axustado e a linea vermella representa a talla a cual o 50% da poboación está madura.

torno á media sen amosar unha tendencia clara. A partir do ano 1992 prodúcese un incremento da talla de primeira madurez sexual que se mantén ata o ano 1998 acadando o seu máximo neste ano (49.7 cm). Non obstante, é a partir deste ano, cando se observa unha tendencia á diminución da talla de primeira madurez sexual da poboación de merluza acadando a talla mínima de primeira madurez de 22.3 cm no ano 2019.

Analizando a talla de primeira madurez sexual segundo sexo ao longo do período de estudo (ver Figura 2), ademais de notar, como se ven dicindo, que a L_{50} estimada para as femias é maior ca dos machos, obsérvase que a L_{50} para as femias segue unha tendencia similar á poboación de merluza, presentada anteriormente, mentres que os machos presentan unha tendencia á diminución da L_{50} para todo o período de estudo.

Así mesmo, cómpre salientar que a escasa proporción de exemplares de femias inmaduras na mostra para o 1982 e para os machos nos anos 2014 e 2015, impediou estimar un valor verosímil da L_{50} (ver Figura 2 da Sección .1 do Apéndice). En liña co obxectivo principal deste estudo, e en coherencia co que se ven dicindo ao longo deste capítulo, tendo en conta como criterio a menor perda de información posible, estes valores faltantes cubríronse mediante o procedemento de medias móbiles a través do paquete `imputeTS` empregando a función `na.ma`. A continuación, na Figura 3.9 amósase a talla de primeira madurez sexual ao longo do período de estudo para cada sexo resultado de imputar os valores faltantes das series temporais.

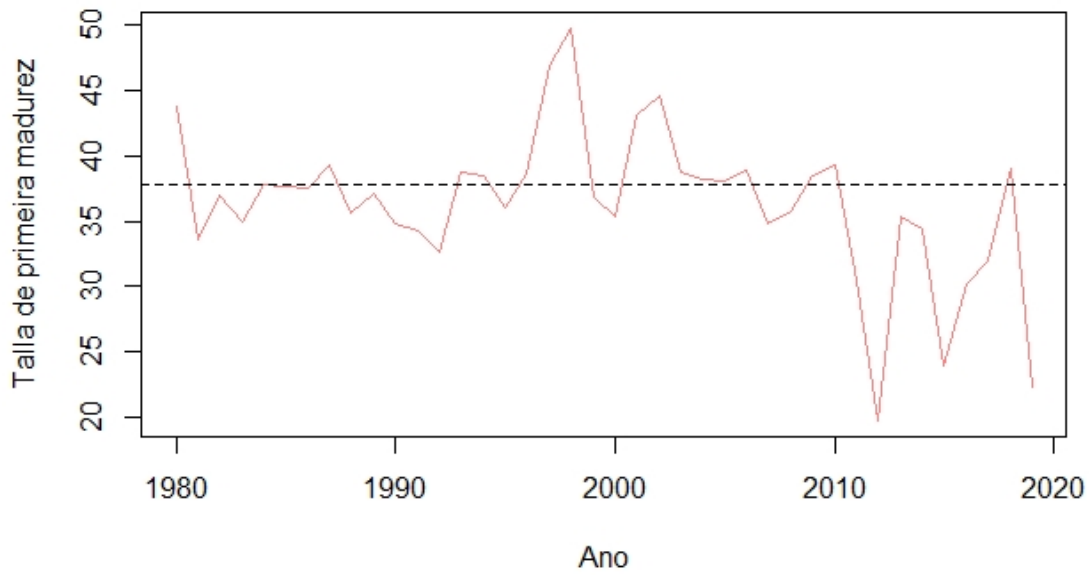


Figura 3.8: Variación anual da talla de primeira madurez sexual para a poboación da merluza.

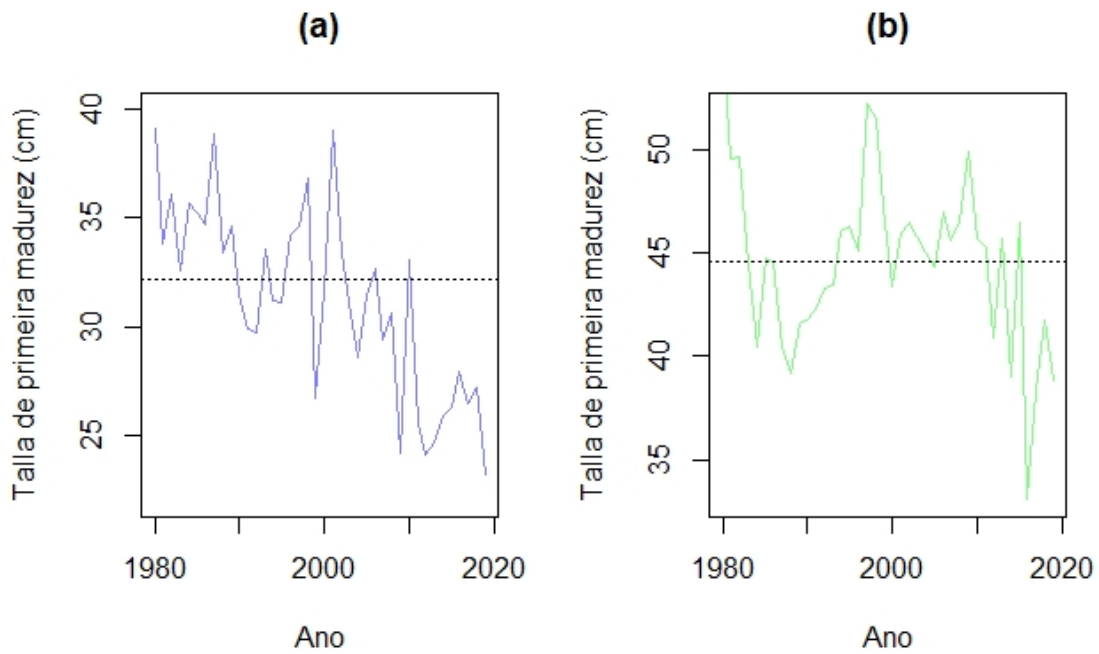


Figura 3.9: Variación anual da talla de primeira madurez sexual para as merluzas (a) macho e (b) femia ao longo do período de estudo despois do procedemento de imputación.

3.2.2. Factor de condición

Tal e como se expuxo na Sección 2.2.1.2, o factor de condición relativo (K_n) estímase co peso vivo en lugar do peso eviscerado pola propia limitación dos datos. En coherencia co dito e, tal e como se expresou ao principio da Sección 3.1.2, non procede realizar unha análise en termos comparativos segundo sexo do factor de condición relativo.

Os datos apuntan que a poboación de merluza ao longo do período de estudo ten un valor promedio de 0.85 nun rango de 0.024 e 8.38. Así mesmo, os valores do factor de condición relativo oscilan entre 0.053 e 8.38 para as femias e entre 0.060 e 2.853 para os machos cun promedio de 1.08 e 0.60, respectivamente.

A análise anual do factor de condición relativo para a poboación de merluza apunta a unha tendencia ao aumento a partir do ano 1994. Segundo sexo (ver Figura 4), o factor de condición relativo para as merluzas femia segue a mesma estrutura que para a poboación da merluza. Non obstante, o factor de condición relativo para as merluzas macho non presenta unha tendencia clara.

Tal e como se dixo na Sección 3.1.2, o non dispor de información para o peso vivo dende o 1999 ata o 2001 para a poboación de merluza e para as merluzas femia e, dende o 1998 ata o 2001 para os machos, impediu estimar o factor de condición relativo para a poboación de merluza e para cada sexo nos anos citados anteriormente o cal se observa nas Figuras 3 e 4 da Sección .1 do Apéndice. Estes valores faltantes cubríronse mediante o procedemento de medias móbiles. A continuación, nas Figuras 3.10 e 3.11 amósase o factor de condición relativo para a poboación de merluza e para cada sexo tras cubrir os valores faltantes das series temporais para este rasgo de vida.

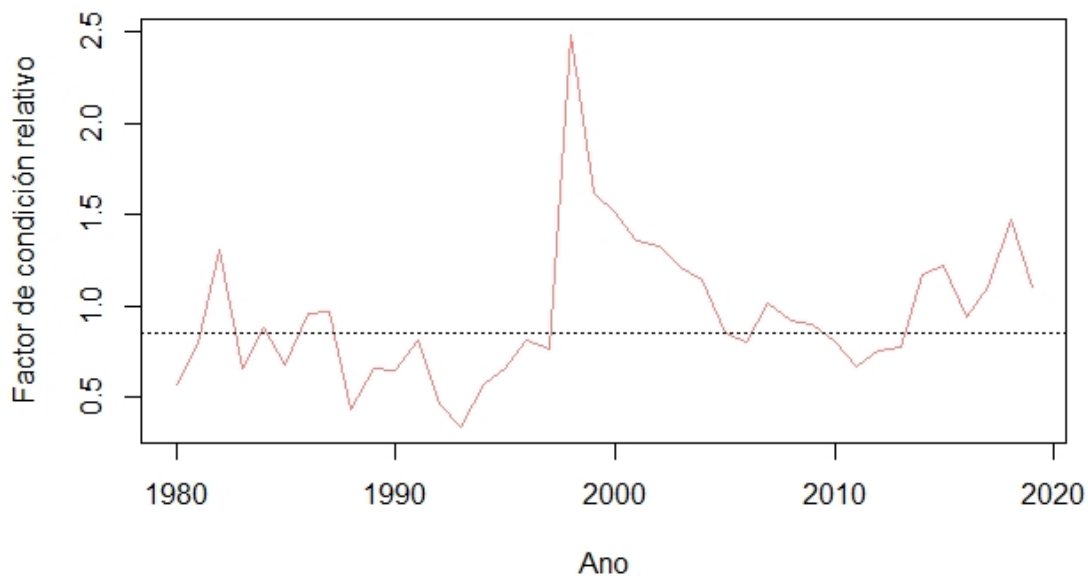


Figura 3.10: Variación anual do factor de condición relativo para a poboación da merluza ao longo do período de estudo tras cubrir os valores faltantes mediante medias móbiles.

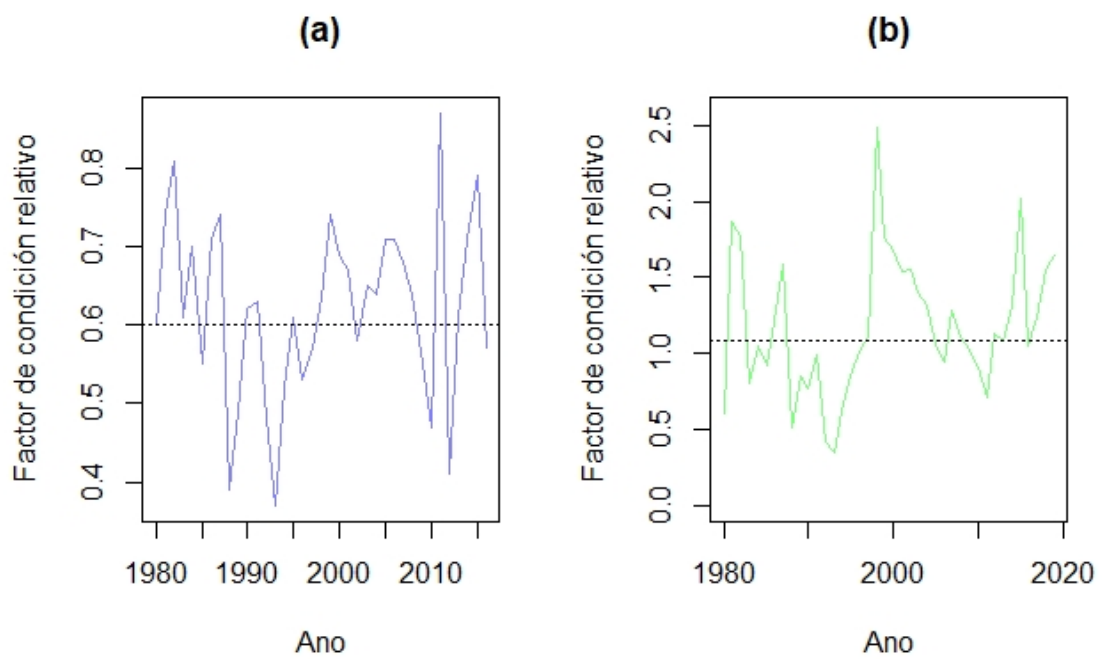


Figura 3.11: Variación anual do factor de condición relativo para as merluzas macho (a) e para as merluzas femia (b) ao longo do período de estudo tras cubrir os valores faltantes mediante medias móbiles.

3.3. Modelos de regresión

No presente apartado preténdese explicar as variacións anuais detectadas na talla de primeira madurez sexual e no factor de condición relativo para cada sexo e para ambos sexos combinados en función dos posibles efectos que as variables ambientais e biolóxicas exercen sobre ditos rasgos de vida. Nótese que en aras de explicar a variabilidade da talla de primeira madurez sexual e o factor de condición relativo, as variables explicativas que se teñen en conta en función do rasgo de vida a estudar, son diferentes. É dicir, para explicar a variabilidade da talla de primeira madurez sexual considéranse as variables ambientais AMO, NAO e a SST, e a biomasa como factor biolóxico. Non obstante, se ben para o factor de condición relativo as variables ambientais son as mesmas que para a talla de primeira madurez sexual, como factores biolóxicos téñense en conta a biomasa reproductora e a talla. Cómpre salientar, que a selección das variables explicativas consideradas para explicar a variabilidade anual de cada rasgo de vida non é baladí xa que responde ao coñecemento dos coordinadores deste proxecto en bioloxía pesqueira.

En orde a determinar a variabilidade dos rasgos de vida e tendo en conta os motivos expostos na Sección 1.4, aplícanse modelos GAM empregando a distribución gaussiana, como base suavizadora os splines de regresión cúbica e como método de estimación dos parámetros o REML. Ademais, coa finalidade de controlar o grado de suavidade das funcións límitase o número de nodos.

A dispoñibilidade de información con frecuencia anual para a SST e a biomasa reproductora a partir do ano 1982 e 1983 respectivamente, impediu abordar o estudo da variabilidade dos rasgos de vida para o período comprendido entre 1980 e 2019. Neste senso, a talla de primeira madurez sexual para cada sexo e para a combinación de ambos abrangue o período comprendido entre 1982 e 2019 e o factor de condición relativo, dende o ano 1983.

Antes de axustar os modelos, emprégase a función `tsclean` do paquete `forecast` para identificar e substituír os valores atípicos das series temporais (tanto das variables explicativas como das variables resposta) [21].

Previamente á construción dos modelos, cómpre estudar a posible correlación e colinealidade entre as variables explicativas. A correlación entre variables estúdase empregando o contraste de correlación de Pearson a través do paquete de R `corrplot` [23], namentras que a colinealidade analízase calculando os factores de variación-inflación xeneralizados (GVIF, do inglés “generalized variance-inflation factors”) que son os valores de VIF corrixidos polos grados de liberdade da variable predictora (para máis detalles véxase [7]). O GVIF de cada variable explicativa obténse mediante a función `corvif` en R. Nótese que se identifican como variables conflictivas aquelas cuxo GVIF sexa superior a 3 unidades. Neste estudo, soamente as variables AMO e ano presentan unha correlación alta de 0.71 e, un GVIF superior a 3 unidades para AMO. En liña co obxectivo principal deste estudo, estudar a variabilidade temporal dos rasgos de vida, decídese prescindir da variable AMO nos modelos de regresión.

A selección das variables explicativas engadidas no modelo final para explicar a variabilidade anual da talla de primeira madurez sexual e do factor de condición relativo, para cada sexo e para a combinación de ambos, realízase mediante o procedemento “stepwise backward”. Ademais, sobre os residuos dos modelos estúdanse as hipóteses básicas empregando os seguintes contrastes:

- Contraste de estacionariedade de Dickey-Fuller.
- Contraste de normalidade a través dos contrastes de Jarque-Bera e de Shapiro-Wilk
- Contraste de media 0 “Student’s t-Test”.
- Contraste de incorrelación de Ljung-Box para os 30 primeiros retardos temporais.
- Contraste de homocedasticidade mediante gráfico dos residuos fronte aos valores pronosticados o que confirmará a existencia ou non dun patrón.

Tal e como se menciona no Prefacio e na Sección 2.2.2.3, ao estar empregando series temporais é posible que os residuos dos modelos presenten unha estrutura de correlación temporal. Neste senso, procederáse ao axuste dun modelo ARMA (p, q) para os mesmos seguindo o procedemento que se amosa na Sección 2.2.4. Unha vez se valida a comprobación das hipóteses básicas do modelo que se propón, procédese ao axuste do modelo completo.

A continuación, sinálase de xeito breve os pasos a seguir para obter un modelo válido.

- **Paso 1** “Stepwise backward” coa finalidade de atopar un modelo onde as variables do mesmo resulten significativas a un nivel de significación do 5% e a influencia das mesmas, sobre a variable resposta, sexa coherente coa lóxica biolóxica.
- **Paso 2** Comprobación da condición de estacionariedade dos residuos e a hipótese de independencia dos mesmos. Unha vez comprobado que son estacionarios, de incumplirse a hipótese de independencia procédese ao axuste dun modelo ARMA (p, q) sobre os residuos.
- **Paso 3** Comprobación da significación das ordes do modelo proposto para os residuos e aplicación dos contraste das hipóteses básicas do mesmo.
- **Paso 4** Axuste do modelo completo mediante o modelo de regresión GAM e o proceso Box-Jenkins que modeliza a dependencia temporal dos residuos.

3.3.1. Talla de primeira madurez sexual

Para explicar o comportamento da talla de primeira madurez sexual da merluza, para cada sexo e para ambos sexos combinados para o período de estudo comprendido entre 1982-2019 considéranse como variables explicativas, os factores ambientais, SST, NAO e AMO, a biomasa como factor biolóxico e o ano para capturar a variabilidade anual.

3.3.1.1. Talla de primeira madurez sexual da poboación de merluza

Neste apartado abórdase a **variabilidade da talla de primeira madurez sexual para a poboación de merluza**. En tanto que o resto de modelos se realizan de xeito análogo, describiráse este modelo en detalle para, no resto dos modelos, engadir unicamente os resultados. Así, o modelo final que se obtén no paso 1 mediante o procedemento “stepwise backward” detalláse no Cadro 3.3. A continuación, procédese a realizar o contraste das hipóteses básicas do mesmo.

Coeficientes paramétricas				
	Estimación	Std Erro	t valor	Pr
β_0	36.22	0.75	48.61	2e-16
Significación das funcións suaves				
	Edf	Ref df	F	P-valor
$f(\text{ano})$	2.826	3.339	6.357	0.00129 **

Cadro 3.3: Estimación dos coeficientes do modelo.

Como se observa no Cadro 3.4 para un nivel de significación do 5% os residuos son normais baixo os criterios do test de normalidade de Jarque Bera e o de Shapiro-Wilk. Ademais os residuos teñen media 0 e son estacionarios. En canto á hipótese de homocedasticidade, no gráfico superior da dereita da Figura 3.12, compróbase que os residuos non seguen ningún patrón polo que os residuos son homocedásticos. Non obstante, en relación á hipótese de independencia, gráfico á esquerda da

Figura 3.12, representáanse os p -valores do test de Ljung-box para os 30 primeiros retardos da serie, os resultados levan a rexeitar a hipótese de independencia dos residuos.

En definitiva, conclúese que o modelo non é válido, pois os residuos están correlados xa que baixo normalidade a independencia equivale a incorrelación. Isto confirmase mediante os gráficos inferiores, fas e fap respectivamente, da Figura 3.12. Polo tanto, é preciso a búsqueda dun modelo ARMA (p, q) que modelice a estrutura de dependencia temporal presente nos residuos do modelo.

Contraste	Hipóteses	P -valor
Jarque-Bera	Normalidade	0.20
Shapiro-Wilk	Normalidade	0.25
Dickeye-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 3.4: Contrastes sobre os residuos do modelo para a talla de primeira madurez sexual para ambos sexos combinados.

A través da función `auto.arima` e dos gráficos da fas e da fap (gráficos inferiores da Figura 3.12), compróbase se existe algún proceso que modelice a estrutura temporal dos residuos do modelo. Entre os diferentes modelos propostos seleccionase, en base ao criterio AIC, o modelo ARMA(2,2).

Unha análise do axuste anterior en busca de valores atípicos localizou a presenza dun atípico aditivo na posición 31. A modelización deste atípico no modelo pasa por considerar que se ben algúns coeficientes do modelo antes eran significativamente distintos de cero, pasan a ser nulos quedando un modelo ARMA $(2, 2)$ co atípico mencionado (ver Cadro 3.5). Este modelo explica o 59.9% da deviance ou, o que é o mesmo, o 59.9% da variabilidade da talla de primeira madurez sexual da poboación de merluzas explícase pola variable temporal ano.

	Estimación	Std Erro	Z valor	Pr ($> z $)
ar2	-0.97189	0.04293	-22.6392	$< 2.2e - 16$
ma2	0.78540	0.15457	5.0811	$3.753e - 07$
AO.31	-10.23732	2.96178	-3.4565	0.0005473

Cadro 3.5: Estimación dos coeficientes do modelo ARMA (p, q) .

Finalmente, procédese novamente á validación do modelo. Tal e como pode verse no Cadro 3.6 e nos gráficos da Figura 3.13, confírmase que os residuos seguen unha distribución normal, son estacionarios e teñen media cero. Ademáis os residuos son incorrelados pois baixo normalidade, a hipótese de independencia equivale a incorrelación.

Trala obtención dun modelo válido, na Figura 3.14 repórtase o efecto estimado do ano sobre a talla de primeira madurez sexual para a poboación da merluza a través do cal se observa unha tendencia á diminución a partir do ano 2000 ata o final da serie.

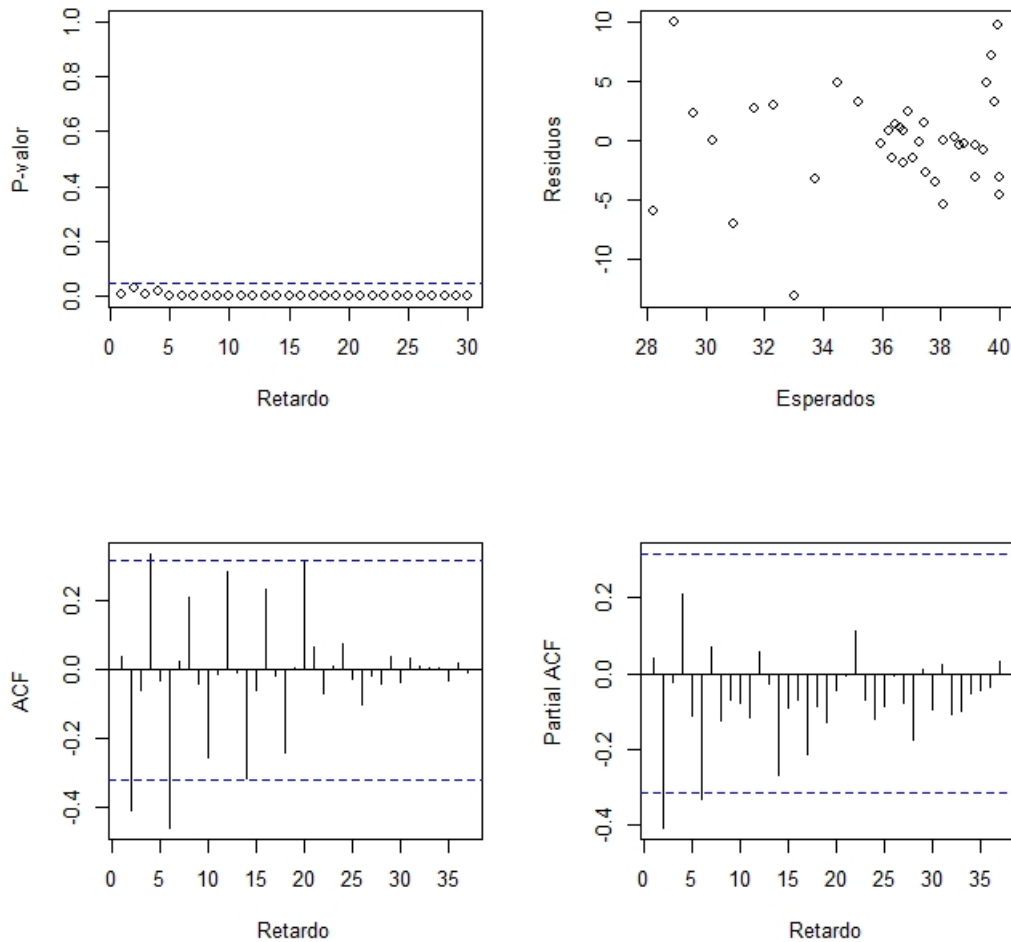


Figura 3.12: Contrastes gráficos sobre os residuos do modelo GAM para a talla de primeira madurez sexual para a poboación de merluza. A gráfica superior esquerda, correspóndese cos p -valores do contraste de independencia Ljung-Box e, a superior á dereita, ven a sinalar a homocedasticidade dos residuos do modelo. As gráficas inferiores representan as correlacións simples e parciais dos residuos.

Contraste	Hipóteses	P -valor
Jarque-Bera	Normalidade	0.20
Shapiro-Wilk	Normalidade	0.25
Dickeye-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 3.6: Contrastes das hipóteses básicas sobre os residuos do modelo da talla de primeira madurez sexual froito de modelizar a estrutura temporal dos residuos mediante un ARMA (2,2).

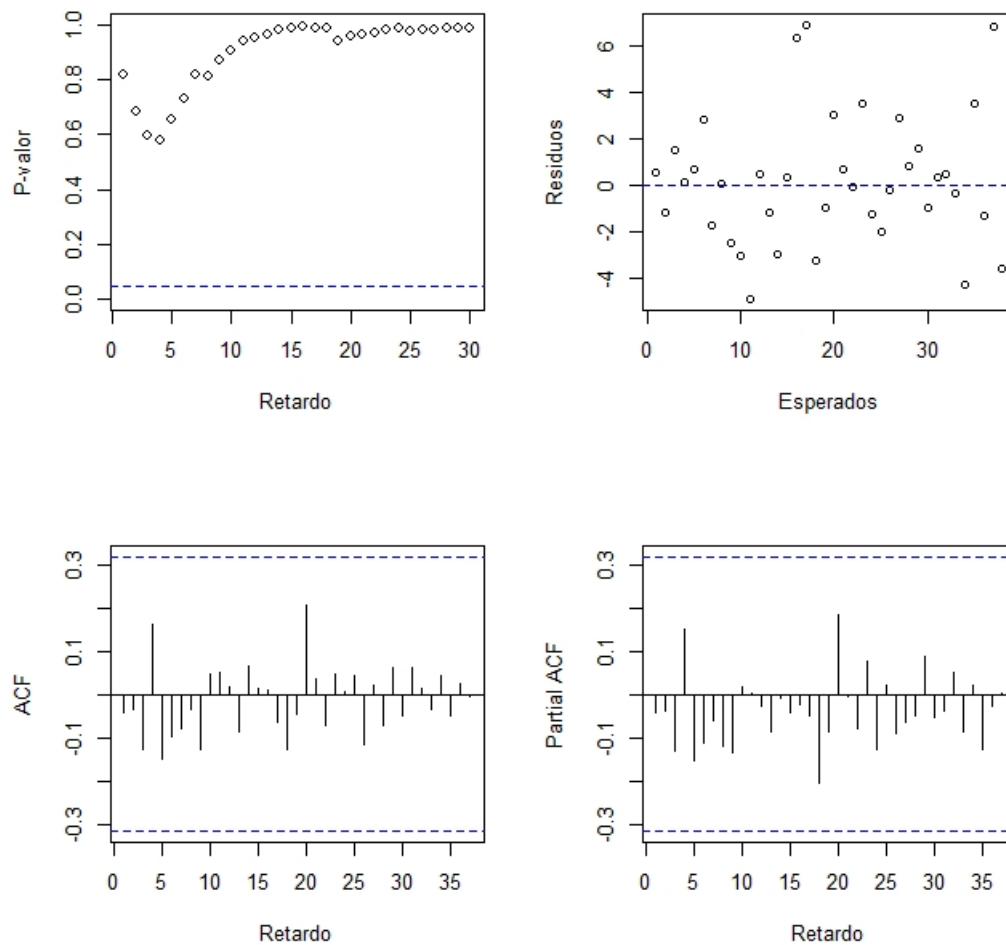


Figura 3.13: Contrastes gráficos sobre os residuos do modelo da talla de primeira madurez sexual froito de modelizar a estrutura temporal dos residuos mediante un ARMA (2, 2). A gráfica superior esquerda, correspóndese cos p -valores do contraste de independencia Ljung-Box e, a superior á dereita, ven a sinalar a homocedasticidade dos residuos do modelo. As gráficas inferiores representan as correlacións simples e parciais dos residuos.

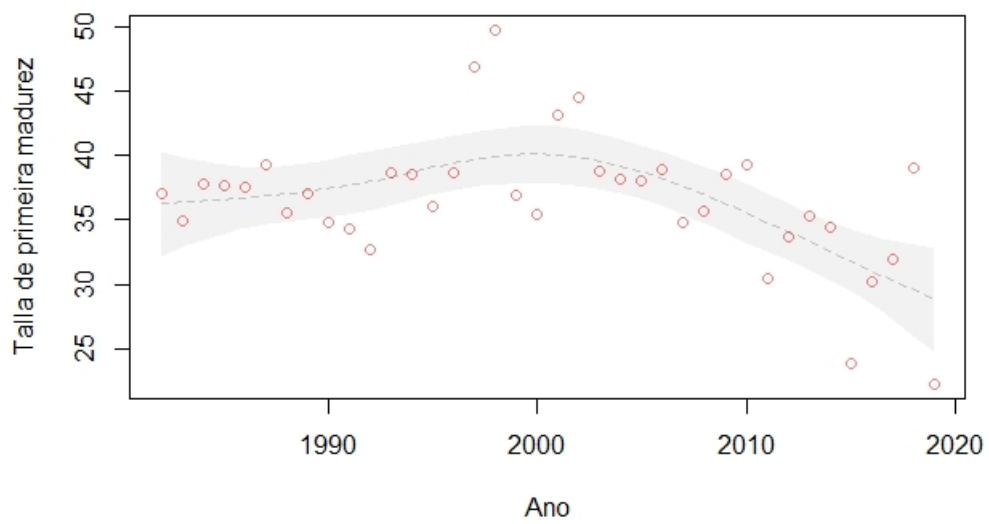


Figura 3.14: Efecto estimado do ano sobre a talla de primeira madurez sexual para a poboación total de merluza. Os círculos vermellos representan os valores anuais estimados e a sombra gris os intervalos de confianza do 95%.

3.3.1.2. Talla de primeira madurez sexual das merluzas macho

Neste apartado abórdase a variabilidade da **talla de primeira madurez sexual para as merluzas macho**. O modelo final pode verse no Cadro 3.7, no cal se observa que tanto o intercepto (β_0) como a función suave correspondente á variable temporal ano, $f(\text{ano})$, son significativas a un nivel de significación do 5%. Este modelo explica o 56.5% da deviance ou, o que é o mesmo, o 56.5% da variabilidade da talla de primeira madurez sexual das merluzas macho explícase pola variable temporal ano. Como se observa na Figura 3.15 a talla de primeira sexual madurez presenta unha tendencia á diminución ao longo da serie temporal. Finalmente, os residuos do modelo cumpren a hipótese de normalidade, son homocedásticos, estacionarios, teñen media cero e son independentes, xa que baixo normalidade, a incorrelación equivale a independencia (ver Cadro 2 e Figura 5 da Sección .2.1 do Apéndice).

Coeficientes paramétricas				
	Estimación	Std Erro	t valor	Pr ($> t $)
β_0	36.22	0.75	48.61	2e-16
Significación das funcións suaves				
	Edf	Ref df	F	P -valor
$f(\text{ano})$	2.826	3.339	6.357	0.00129 **

Cadro 3.7: Estimación dos coeficientes do modelo da talla de primeira madurez sexual das merluzas macho.

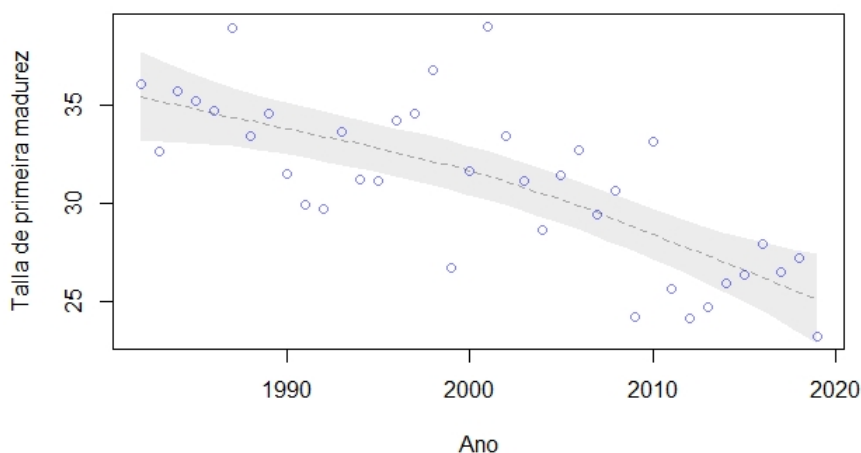


Figura 3.15: Efecto estimado do ano sobre a talla de primeira madurez sexual das merluzas macho. Os círculos azuis representan os valores anuais estimados e a sombra gris os intervalos de confianza do 95%.

3.3.1.3. Talla de primeira madurez sexual das merluzas femia

O modelo resultante para explicar a talla de primeira madurez sexual das merluzas femia amósase no Cadro 3.8, no cal se observa que as estimacións dos coeficientes do modelo, tanto o intercepto (β_0) como a función suave ano ($f(\text{ano})$), son significativas a un nivel de significación do 5%.

Coeficientes paramétricas				
	Estimación	Std. Erro	t valor	Pr ($> t $)
β_0	44.2158	0.4499	98.28	$< 2e - 16$ * **
Significación das funcións suaves				
	Edf	Ref df	F	P-valor
$f(\text{ano})$	4.949	5.832	5.925	0.00463 ***

Cadro 3.8: Estimación dos coeficientes do modelo da talla de primeira madurez sexual das merluzas femia.

Non obstante, o 54.6% da variabilidade da talla de primeira madurez sexual das merluzas femia explícase pola variable temporal ano. Ademais, como se observa na Figura 3.16 a talla de primeira madurez sexual das merluzas femia presenta unha tendencia á diminución dende o ano 2000 ata o final da serie temporal.

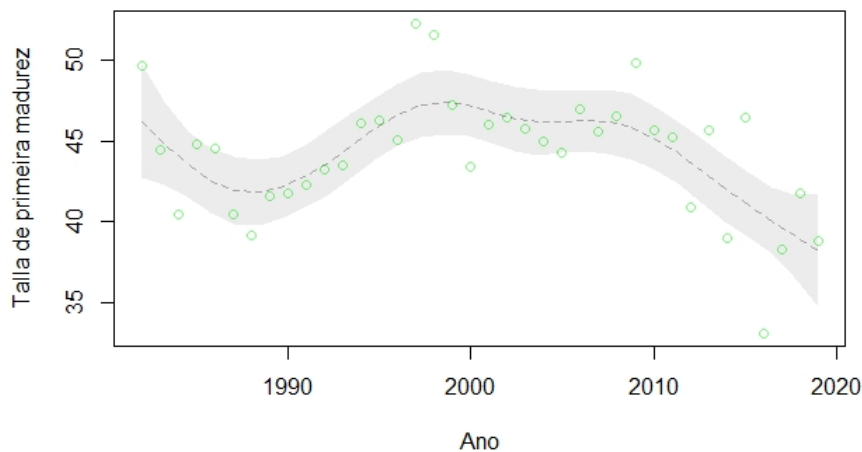


Figura 3.16: Efecto estimado do ano sobre a talla de primeira madurez sexual das merluzas femia. Os círculos verdes representan os valores anuais estimados e a sombra gris os intervalos de confianza do 95%.

Para finalizar, cómpre salientar que os residuos do modelo cumpren as hipóteses básicas en tanto que son estacionarios, teñen medio cero e son independentes. Ademais, son homocedásticos

e seguen unha distribución normal. Isto obsérvase no Cadro 3 e na Figura 6 da Sección .2.2 do Apéndice.

3.3.2. Factor de condición relativo

En orde a determinar que factores explican a varibilidade do factor de condición relativo, para ambos sexos combinados e para cada sexo para o período comprendido entre 1983 e 2019, tivéronse en conta os factores ambientais AMO, NAO e SST e os factores biolóxicos, a talla e a biomasa reproductora.

3.3.2.1. Factor de condición relativo da poboación de merluza

O modelo obtido para explicar o **factor de condición relativo da poboación de merluza**, detállase no Cadro 3.9, no cal se observa que tanto o intercepto (β_0) como a función suave correspondente á variable biolóxica talla ($f(talla)$), son significativas a un nivel de significación do 5%. Este modelo explica o 95.1% da deviance, ou dito doutro xeito, o 95% da variabilidade do factor de condición relativo para a poboación de merluza explícase pola variable talla (cm) .

Coeficientes paramétricas				
	Estimación	Std Erro	<i>t</i> valor	Pr (> <i>t</i>)
β_0	0.92217	0.01154	79.91	< 2e - 16 ***
Significación das funcións suaves				
	Edf	Ref df	F	P-valor
$f(talla)$	2.344	2.707	2389	< 2e - 16 ***

Cadro 3.9: Estimación dos coeficientes do modelo do factor de condición da poboación de merluza.

Os residuos do modelo cumpren as hipóteses básicas en canto son independentes, estacionarios, homocedásticos, teñen media cero e seguen unha distribución normal (ver Cadro 7 e Figura 4 da Sección .2.3 do Apéndice). Como se observa no gráfico 3.18 o factor de condición relativo para a poboación de merluza aumenta a medida que a talla é máis grande.

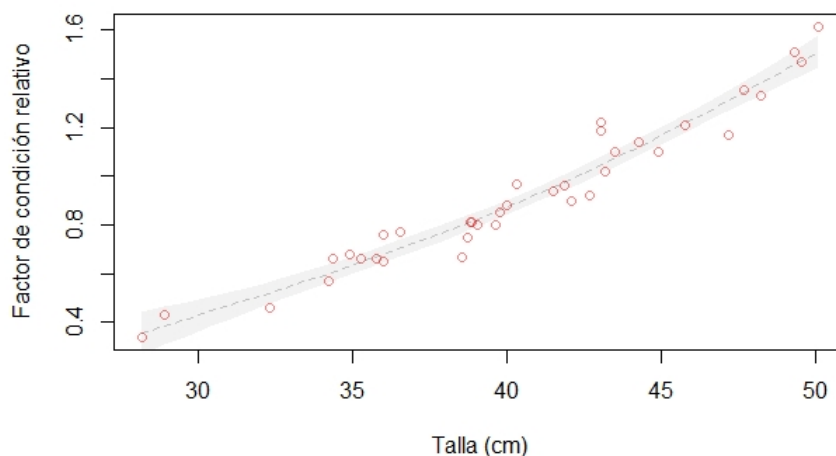


Figura 3.17: Efecto estimado da talla sobre o factor de condición relativo para a poboación de merluza.

3.3.2.2. Factor de condición relativo das merluzas macho

A variabilidade do **factor de condición relativo para as merluzas macho** explícase pola influencia da variable temporal ano e a talla. Este modelo amósase no Cadro 3.11 onde, a un nivel de significación do 5 %, todos os coeficientes do modelo son significativos. Este modelo explica o 96.7 % da deviancia. Por outra banda, os residuos do modelo son independentes, estacionarios, normais, teñen media cero e non seguen un patrón polo tanto, cumpren as hipóteses básicas (ver Cadro 5 e Figura 8 da Sección .2.4 do Apéndice). Na Figura 3.18, amósase o efecto estimado da parte non paramétrica do modelo sobre a variable resposta. Neste senso, cómpre salientar que o factor de condición relativo para as merluzas macho correlaciónase de xeito positivo coa talla, sen embargo, se ben a variable ano contribúe a explicar a variabilidade, a través da análise gráfica pódese salientar dúas cousas: por unha banda que o factor de condición relativo permanece constante ao longo do período de estudo e, pola outra, un mal axuste da función de suavizado aos valores observados o que leva a concluir que a variable temporal ano non explica a variabilidade do factor de condición.

Coeficientes paramétricas				
	Estimación	Std Erro	<i>t</i> valor	Pr ($> t $)
β_0	0.615152	0.003474	177	$< 2e - 16$ * **
Significación das funcións suaves				
	Edf	Ref df	F	<i>P</i> -valor
$f(\text{ano})$	1.945	2.324	4.017	0.0231*
$f(\text{talla})$	1.506	1.835	471.178	$< 2e - 16$ * **

Cadro 3.10: Estimación dos coeficientes do modelo para o factor de condición relativo para as merluzas macho.

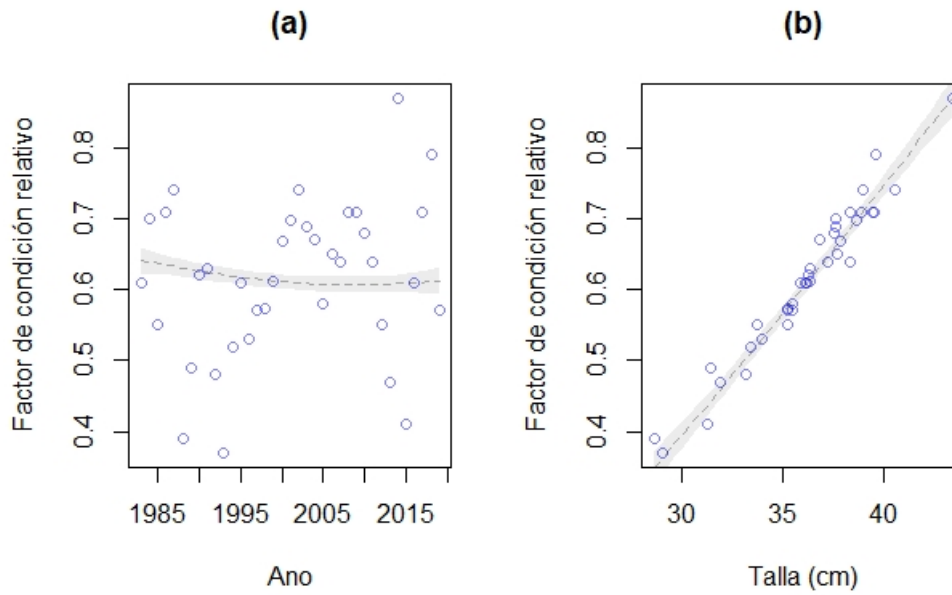


Figura 3.18: Efecto estimado do ano (a) e da talla (b) sobre o factor de condición relativo para as merluzas macho.

3.3.2.3. Factor de condición relativo das merluzas femia

Antes de proceder a presentar o modelo final que contribúe a explicar a variabilidade do **factor de condición relativo para as merluzas femias** que se amosa no Cadro 3.11, cómpre salientar que a pesar de suavizar o comportamento das covariables coa aplicación da función `tsclean`, mediante a análise das hipóteses básicas sobre os residuos do modelo, detéctase un atípico no gráfico dos residuos fronte aos valores esperados o que impide validar a hipótese de homocedasticidade. Mediante a aplicación da función `tso` do paquete `tsoutliers` [8], identifícase un atípico na covariable talla, concretamente no ano 1998. Consecuentemente, impútase este valor mediante medias móbiles.

Coeficientes paramétricas				
	Estimación	Std Erro	<i>t</i> valor	Pr (> <i>t</i>)
β_0	0.615152	0.003894	158	$< 2e - 16$ ***
Significación das funcións suaves				
	Edf	Ref df	F	<i>P</i> -valor
$f(\text{ano})$	1.773	2.137	7479	0.00188**
$f(\text{talla})$	2.809	2.970	588.072	$< 2e - 16$ ***

Cadro 3.11: Estimación dos coeficientes do modelo do factor de condición relativo para as merluzas femia.

Non obstante, o 98.5% da variabilidade do factor de condición das merluzas femia explícase

pola variable talla e polo ano. Sen embargo, observando os parámetros de suavizado na Figura 3.19 pódese concluir o mesmo que para o factor de condición relativo dos machos: por unha banda, o factor de condición relativo para as femias correlaciónase de xeito positivo coa talla e, pola outra, este rasgo de vida para as femias permanece constante ao longo do período de estudo. Así mesmo, a función de suavizado ano non se axusta correctamente aos valores observados da variable resposta ao longo do período de estudo abordado. Para rematar, o modelo cumpre as hipóteses básicas de independencia, estacionariedade, media cero, homocedasticidade e normalidade (ver Cadro 6 e Figura 9 na Sección .2.5 do Apéndice).

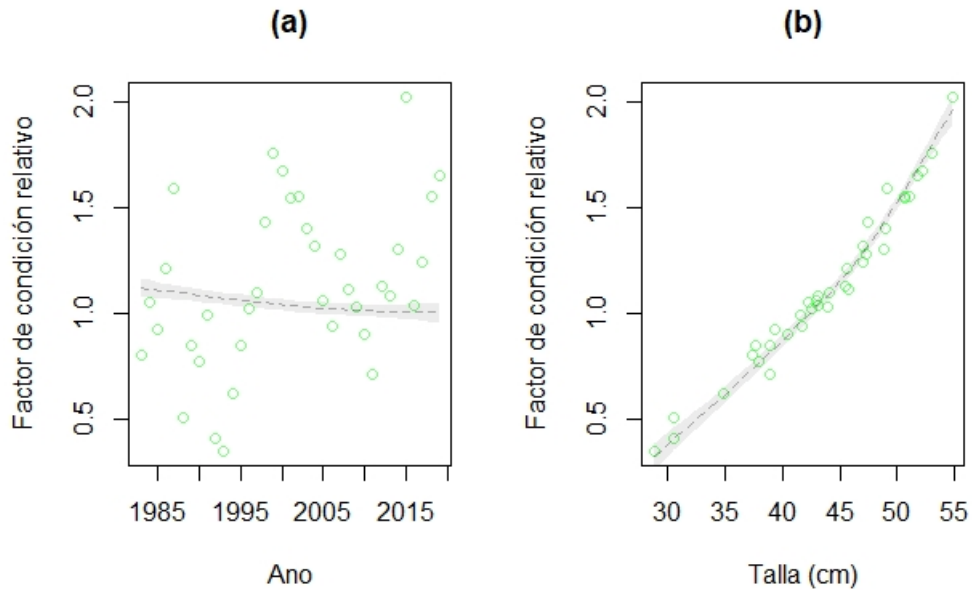


Figura 3.19: Efecto estimado do (a) ano e da (b) talla sobre o factor de condición relativo para as merluzas femia.

Capítulo 4

Conclusiones

Explicar a variabilidade anual dos rasgos de vida con respecto a factores ambientais e biolóxicos é o obxectivo principal exposto na Sección 1.3, ao que se pretende dar resposta neste estudo. Isto implicou a consecución de obxectivos secundarios como 1) proporcionar un mellor coñecemento sobre a estrutura das variables biométricas como é a talla e o peso vivo da merluza variables empregadas para 2) estimar os rasgos de vida talla de primeira madurez sexual e factor de condición relativo para ambos sexos e para a combinación de ambos, para cada ano ao longo do período de estudo. Así mesmo, tal e como se mencionou no prefacio e na Sección 1.4, en aras de acadar o obxectivo principal deste estudo aplícanse os modelos GAM. É por iso que, ao longo do capítulo de metodoloxía, lévase ao lector por un percorrido dende os modelos de regresión máis simples ata os modelos GAM. Ademais, ao tratarse de series temporais tamén se aborda a metodoloxía Box Jenkins concluíndo este apartado cunha guía de aplicación práctica na Sección 2.2.4 en aras de facilitarlle ao lector as ferramentas necesarias para proporcionarlle un mellor entendemento non só de cómo se acada o obxectivo anteriormente citado senón tamén no concernente a interpretación dos resultados.

Da análise descritiva conclúese que as merluzas femia presentan tallas superiores con respecto ás merluzas macho. Así mesmo, a variación da talla anual da poboación de merluza para cada sexo presenta unha tendencia ao aumento da talla a partir do ano 1993 o cal débese porque a maioría dos individuos mostrexados a partir deste ano son individuos maduros.

No que respecta ao peso vivo sen entrar nunha análise comparativa segundo sexo polo que se explicou na Sección 3.1.2, cómpre salientar unha tendencia ao aumento a partir do ano 1994 para a poboación de merluza e para as merluzas femia amosando un peso vivo moi elevado no 1998, o que se debe a que a mostra dese ano componse de 15 individuos femia clasificados como maduros. Doutra banda, o peso vivo das merluzas macho fluctúa arredor da media non observando ningunha tendencia clara ao longo do período de estudo.

A talla de primeira madurez sexual é superior para as femias en comparación cos machos. Así mesmo, a variabilidade anual amosa unha tendencia á diminución da talla de primeira madurez. Mentres que as merluzas femia seguen a mesma tendencia que a poboación total de merluza, unha tendencia á diminución dende o ano 2000 ata o final da serie temporal, as merluzas macho presentan unha tendencia á diminución para todo o período de estudo. Ademais, os resultados dos modelos GAM apuntan que a única variable que contribúe a explicar a variabilidade deste rasgo de vida para cada sexo e para ambos sexos combinados é o ano.

O factor de condición relativo segue a mesma estrutura que a observada para o peso vivo, unha tendencia ao aumento para a poboación de merluza e as merluzas femia a partir do ano 1994. Non obstante, o factor de condición relativo dos machos no amosa ningunha tendencia clara ao longo do período de estudo. Mediante os modelos GAM, conclúese que a variabilidade para este rasgo de vida, tanto para cada sexo como para ambos sexos combinados, explícase pola influencia da talla coa que se correlaciona de xeito positivo. Cómpre destacar que se ben os modelos finais para as merluzas femias e os machos a variable ano é significativa, os gráficos de suavizado amosan por

unha banda, que o factor de condición permanece constante ao longo de estudo e, pola outra, un mal axuste da función de suavizado aos datos o que permite concluir que, en efecto, é a talla a que contribúe a explicar a variabilidade do factor de condición.

Por último, cómpre salientar que non se atopa ningún efecto significativo das variables ambientais empregadas neste traballo. Cecaís debido ao feito de que a súa influencia xa está recollida na variabilidade temporal explicada polo ano.

Bibliografía

- [1] Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. URL <https://CRAN.R-project.org/doc/Rnews/>
- [2] Arnau Gras J (2001), *Diseños de series temporales: técnicas de análisis*. Edicions Universitat de Barcelona.
- [3] Bruselas, A., 2021. Bruselas propone rebajar un 36% la cuota de merluza en el Cantábrico e indigna al sector. [online] La Opinión de A Coruña. Available at: <https://www.laopinioncoruna.es/galicia/2016/10/28/bruselas-propone-rebajar-36-cuota-24391549.html>. [Accessed 20 June 2021].
- [4] Carmona Pontaque, F., 2005. Modelos lineales. Barcelona: Publicacions i Edicions Universitat de Barcelona.
- [5] Cren, E. (1951). The length-weight relationship and seasonal cycle in gonad weight and condition in the perch. *Journal of Animal Ecology*, 20, 201.
- [6] George E.P. Box and David R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 211-252, 1964.
- [7] Fox, J., Weisberg, S. (2018). An R companion to applied regression. Sage publications
- [8] Javier López-de-Lacalle (2019). tsoutliers: Detection of Outliers in Time Series. R package version 0.6-8. <https://CRAN.R-project.org/package=tsoutliers>
- [9] Kung-Sik Chan and Brian Ripley (2020). TSA: Time Series Analysis. R package version 1.3. <https://CRAN.R-project.org/package=TSA>
- [10] Moritz, Steffen Bartz-Beielstein, Thomas. (2017). imputeTS: Time Series Missing Value Imputation in R. The R Journal. 9. 10.32614/RJ-2017-009.
- [11] Patterson, H. D., and R. Thompson. 1971. Recovery of interblock information when block sizes are unequal. *Biometrika* 58: 545-554.
- [12] Peña, Daniel (2010). *Análisis de series temporales*. Alianza Editorial.
- [13] R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- [14] John Fox and Sanford Weisberg (2019). An R Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- [15] Josymar Torrejon-Magallanes (2020). sizeMat: Estimate Size at Sexual Maturity. R package version 1.1.2. <https://CRAN.R-project.org/package=sizeMat>
- [16] Hastie, Trevor and Tibshirani, Robert. (1990) *Generalized Additive Models*. Chapman and Hall, New York.

- [17] ICES. 2018. Report of the Working Group for the Bay of Biscay and the Iberian Waters Ecoregion (WGBIE), 3–10 May 2018 , ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:12. 642 pp
- [18] Hastie, Trevor and Tibshirani, Robert and Friedman, Jerome (2009). *The Elements of Statistical Learning. Springer Series in Statistics*. Springer New York, NY.
- [19] Reiss, P.T. and R.T. Ogden (2009) Smoothing parameter selection for a class of semiparametric linear models. *Journal of the Royal Statistical Society, Series B* 71, 505-524
- [20] Hughes, S.L.; Holliday, N.P.; Beszczynska-Möller, A. (Eds) (2008) ICES Report on Ocean Climate 2007. ICES Cooperative Research Report No. 291
- [21] Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O’Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeeen F (2020). Forecast: Forecasting functions for time series and linear models. R package version 8.13, ¡URL: <https://pkg.robjhyndman.com/forecast/>
- [22] Ucm.es. 2021. [online] Available at: <https://www.ucm.es/data/cont/docs/518-2013-11-11-JAM-IASST-Libro.pdf>. [Accessed 20 June 2021].
- [23] Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., Zemla, J. (2017). Package ‘corrplot’. *Statistician*, 56, 316-324.
- [24] Wood SN, (2006), *Generalized Additive Models: an introduction with R*. CRC Press Taylor Francis Group.
- [25] Wood S (2015) mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML Smoothness Estimation. R package version 1.8-9. <http://cran.r-project.org/package=mgcv>. Accedido 24 de noviembre de 2015.
- [26] Wood, Simon N. (2017). *Generalized Additive Models*. Chapman and Hall/CRC.

Apéndice

.1. Análise datos

	1	2	3	4	5	6	7	8	9	10	11	12
1980	0	0	0	0	177	115	333	0	97	0	0	99
1981	50	0	0	0	0	0	0	0	0	0	0	0
1982	81	43	152	61	31	30	33	0	0	0	0	0
1983	0	93	49	151	0	24	0	0	0	0	0	0
1984	22	142	302	232	0	0	48	22	0	119	122	0
1985	167	168	21	50	252	245	328	294	0	277	155	251
1986	63	181	45	0	0	0	0	103	0	93	0	0
1987	0	0	0	0	0	5	165	0	0	0	3	0
1988	0	76	0	0	0	462	240	231	91	0	0	0
1989	0	29	290	57	345	230	38	198	0	0	0	0
1990	0	98	130	43	14	66	0	142	0	34	58	0
1991	58	0	35	77	332	25	41	74	0	0	0	0
1992	0	26	134	0	0	0	0	0	0	0	0	0
1993	0	0	0	38	32	135	0	0	0	0	0	33
1994	0	83	60	0	143	36	0	0	12	20	117	35
1995	0	81	154	175	0	94	67	32	10	0	80	0
1996	0	5	227	8	83	152	71	199	201	16	41	18
1997	12	73	101	8	47	119	0	143	0	110	0	0
1998	0	0	101	0	162	0	23	0	5	0	0	0
1999	0	76	83	51	34	0	2	0	0	2	0	3
2000	31	43	7	50	0	40	22	0	76	0	55	0
2001	10	14	42	0	22	0	0	0	0	0	0	0
2002	0	0	0	0	74	177	133	0	35	40	27	12
2003	95	280	164	159	323	298	358	146	48	103	195	140
2004	272	162	284	164	200	281	80	165	140	63	318	64
2005	138	198	260	75	272	3	215	419	0	0	286	0
2006	248	116	80	348	0	33	224	0	0	0	0	52
2007	162	39	258	161	116	143	118	78	148	52	0	109
2008	152	225	25	246	65	75	84	157	0	0	76	153
2009	140	21	29	100	27	0	0	169	0	0	0	0
2010	147	106	66	112	48	0	151	0	41	0	11	25
2011	0	0	286	0	20	0	62	112	0	0	106	0
2012	44	71	121	0	0	0	0	0	0	0	0	0
2013	0	30	0	98	47	97	0	0	56	0	0	36
2014	0	51	0	23	110	0	0	0	0	0	0	18
2015	0	73	0	0	18	0	0	0	0	0	0	0
2016	0	76	95	37	79	49	0	0	0	0	0	0
2017	0	0	0	20	270	0	50	0	0	0	42	38
2018	0	65	0	0	72	15	0	0	0	0	43	0
2019	46	46	0	56	119	0	0	0	0	0	0	0

Cadro 1: Distribución mensual das capturas comerciais da merluza europea nas augas Ibéricas Atlánticas segundo ano ao longo do período de estudo.

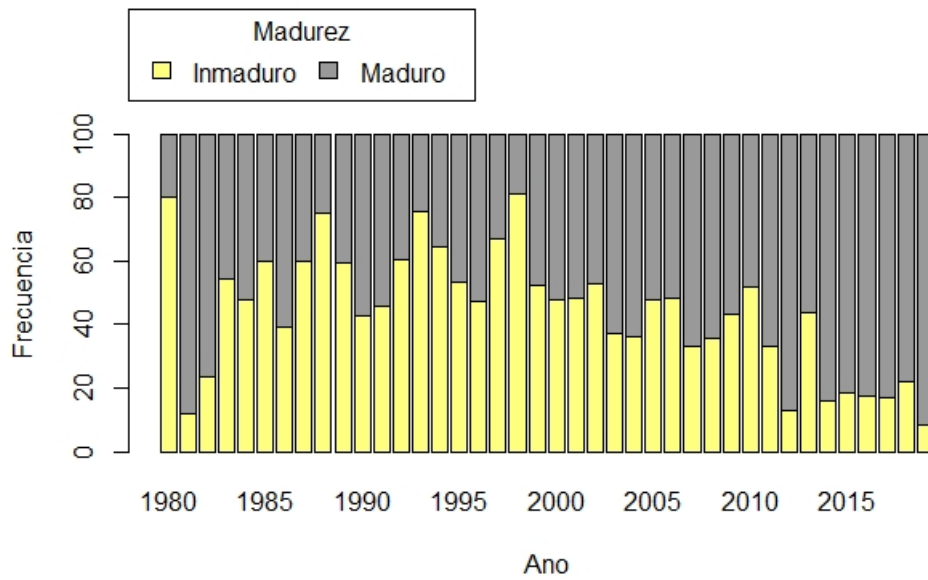


Figura 1: Porporción de individuos clasificados segundo estado de madurez ao longo do período de estudo.

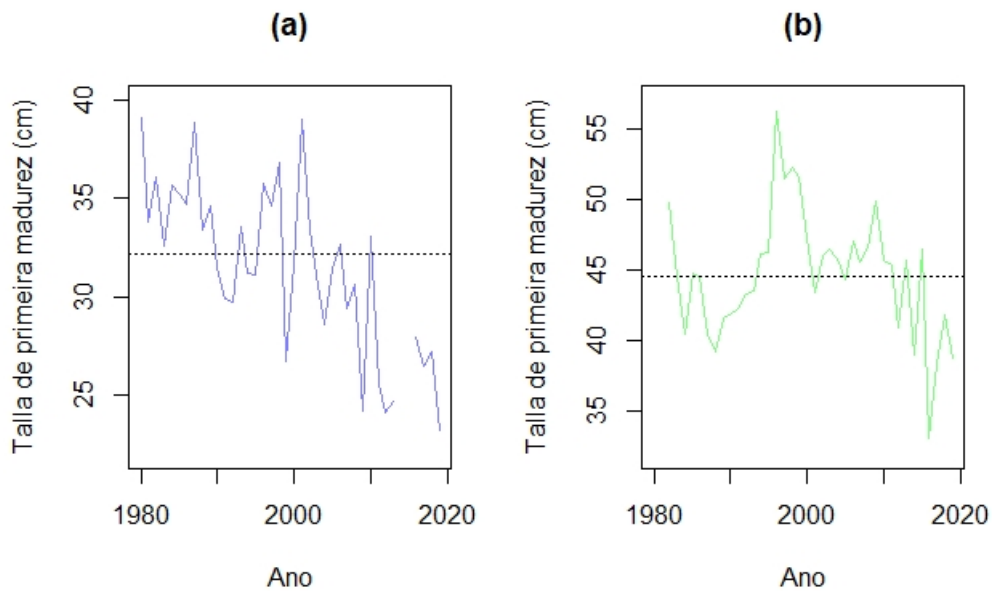


Figura 2: Variación anual da talla de primeira madurez sexual para as merluzas (a) macho e (b) femia ao longo do período de estudo antes de imputar os valores faltantes.

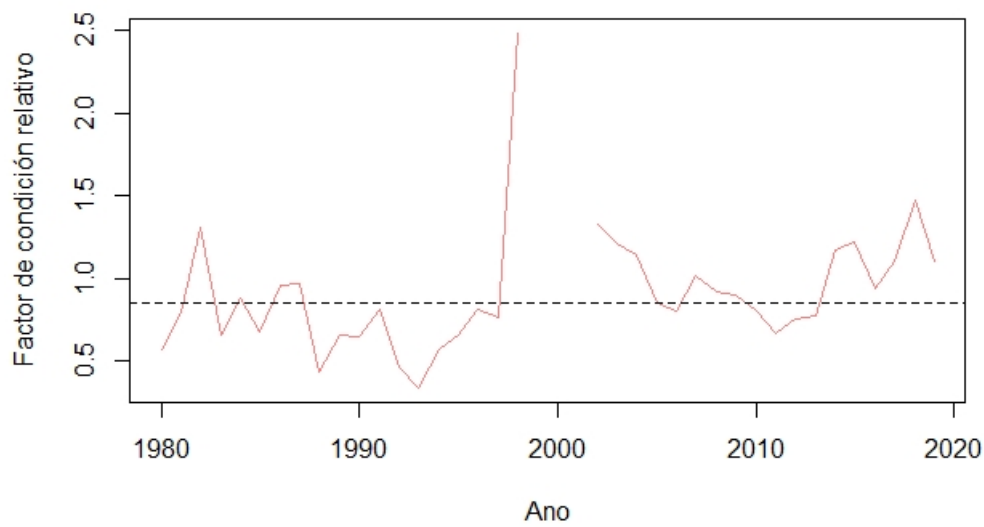


Figura 3: Variación anual do factor de condición relativo para a poboación de merluza antes de imputar os valores faltantes.

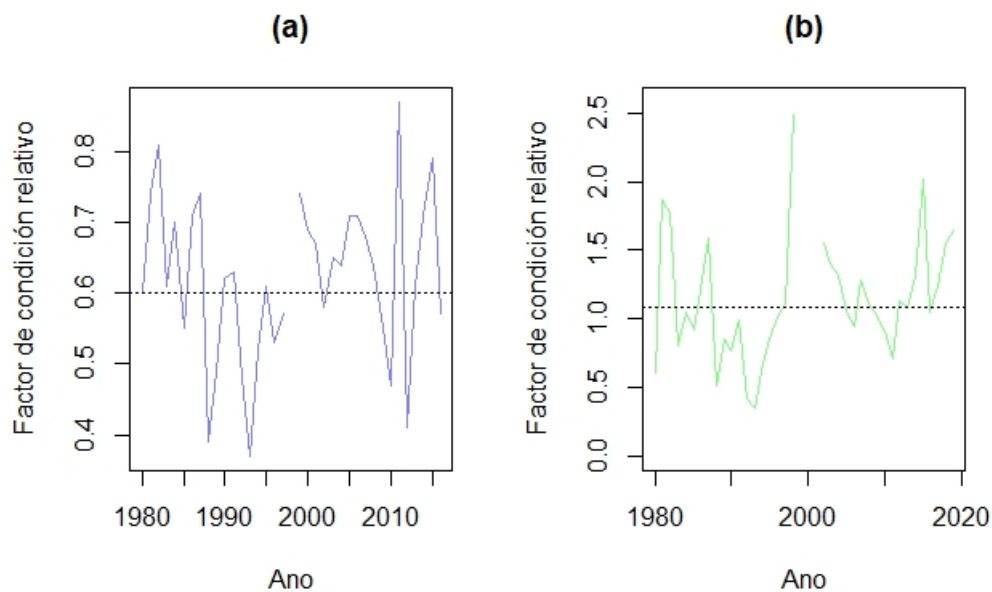


Figura 4: Variación anual do factor de condición relativo para as merluzas (a) macho e para as (b) femia antes de imputar os valores faltantes.

.2. Contraste das hipóteses básicas sobre os residuos dos modelos

No corpo do TFM abórdase, a efectos ilustrativos, os contrastes do modelo concernente á talla de primeira madurez sexual para a poboación da merluza. Neste apartado preséntase ao lector, os

contrastes das hipóteses básicas sobre os residuos do resto de modelos considerados neste traballo como son a talla de primeira madurez sexual para cada sexo e o factor de condición relativo para a poboación da merluza e para cada sexo, nesta orde.

.2.1. Contraste das hipóteses básicas sobre os residuos da talla de primeira madurez sexual para as merluzas macho

Contraste	Hipóteses	P -valor
Jarque-Bera	Normalidade	0.54
Shapiro-Wilk	Normalidade	0.55
Dickeye-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 2: Contrastes das hipóteses básicas sobre os residuos do modelo para a talla de primeira madurez sexual das merluzas macho.

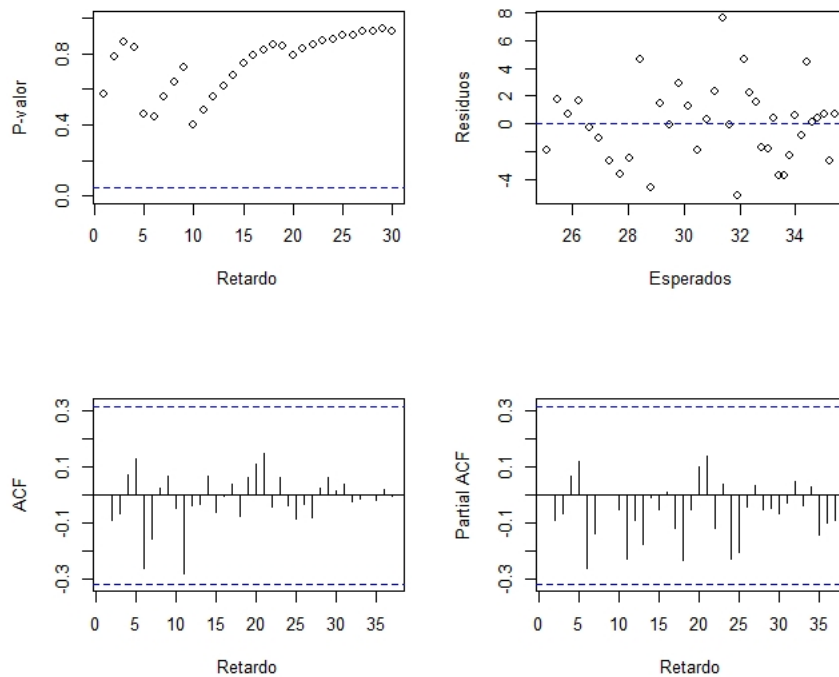


Figura 5: Contrastes gráficos sobre os residuos do modelo. A gráfica superior esquerda, correspóndese cos p -valores do contraste de independencia Ljung-Box e, a superior á dereita, ven a sinalar a homocedasticidade dos residuos do modelo. As gráficas inferiores representan as correlacións simples e parciais dos residuos.

.2.2. Contraste das hipóteses básicas sobre os residuos do modelo da talla de primeira madurez sexual para as merluzas femia

Contraste	Hipóteses	<i>P</i> -valor
Jarque-Bera	Normalidade	0.70
Shapiro-Wilk	Normalidade	0.28
Dickey-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 3: Contrastes das hipóteses básicas sobre os residuos do modelo para a talla de primeira madurez sexual das merluzas femia.

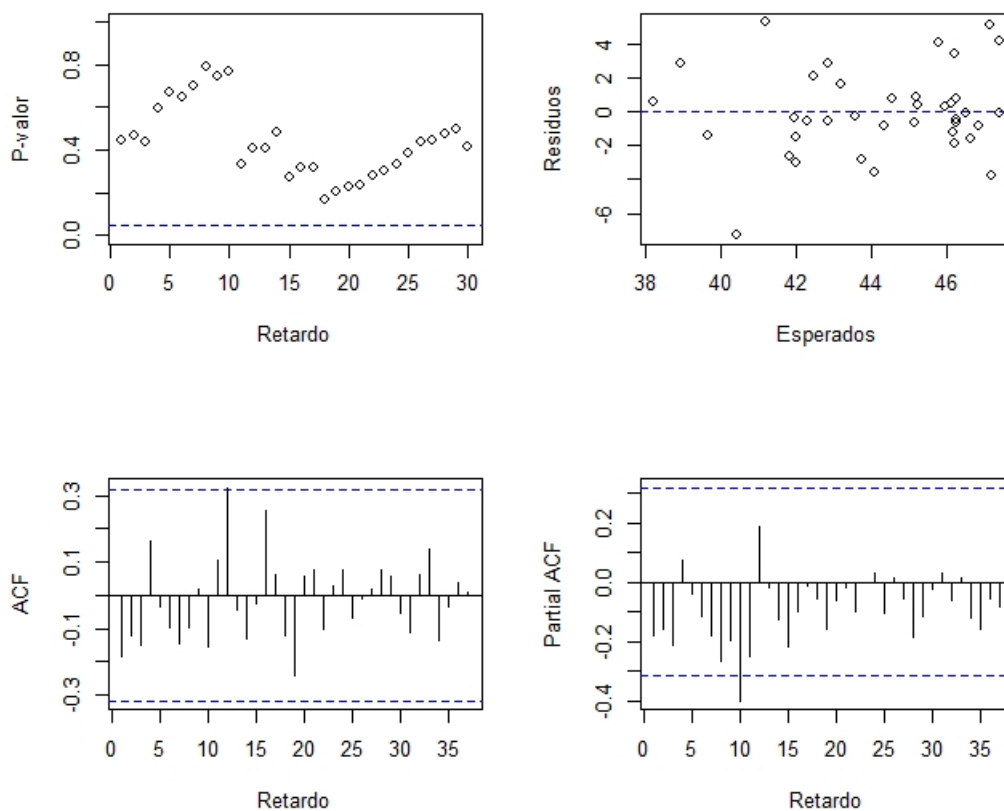


Figura 6: Contrastes gráficos sobre os residuos do modelo. A gráfica superior esquerda, correspóndese cos *p*-valores do contraste de independencia Ljung-Box e, a superior á dereita, ven a sinalar a homocedasticidade dos residuos do modelo. As gráficas inferiores representan as correlacións simples e parciais dos residuos.

.2.3. Contraste das hipóteses básicas sobre os residuos do modelo para factor de condición da poboación da merluza

Contraste	Hipóteses	<i>P</i> -valor
Jarque-Bera	Normalidade	0.60
Shapiro-Wilk	Normalidade	0.64
Dickey-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 4: Contrastes das hipóteses básicas sobre os residuos do modelo para o factor de condición relativo da poboación de merluza.

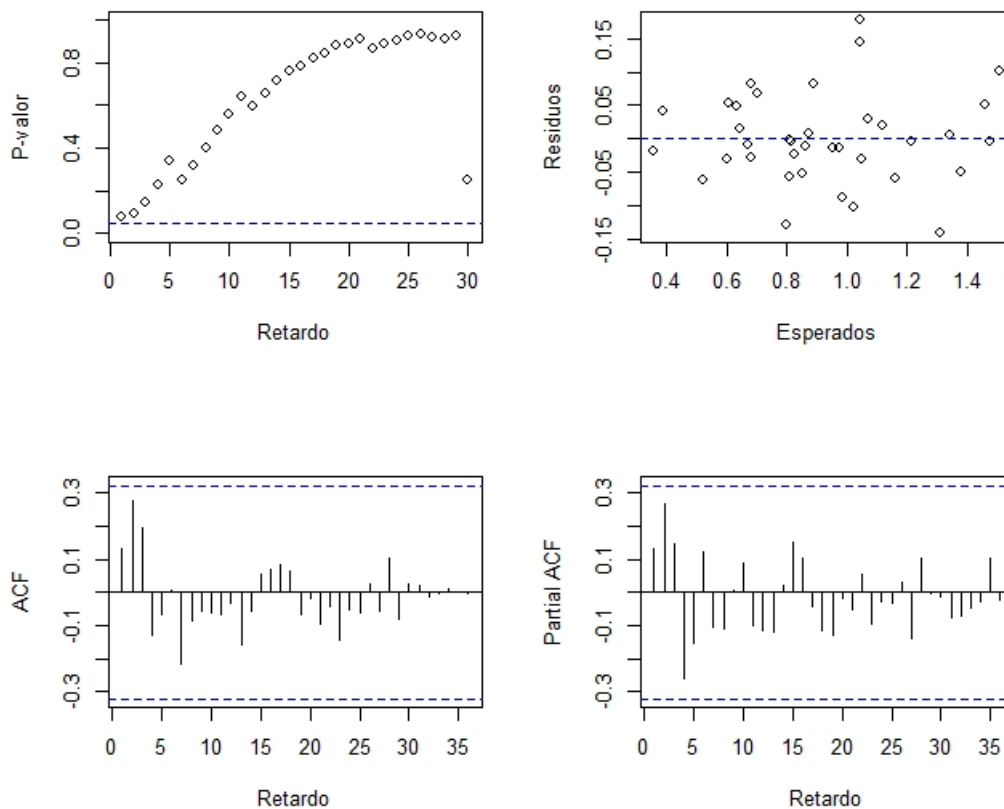


Figura 7: Contrastes gráficos sobre os residuos do modelo. A gráfica superior esquerda, correspóndese cos *p*-valores do contraste de independencia Ljung-Box e, a superior á dereita, ven a sinalar a homocedasticidade dos residuos do modelo. As gráficas inferiores representan as correlacións simples e parciais dos residuos.

.2.4. Contraste das hipóteses básicas sobre os residuos do modelo para factor de condición das merluzas macho

Contraste	Hipóteses	<i>P</i> -valor
Jarque-Bera	Normalidade	0.14
Shapiro-Wilk	Normalidade	0.0845
Dickey-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 5: Contrastes das hipóteses básicas sobre os residuos do modelo para o factor de condición relativo das merluzas macho.

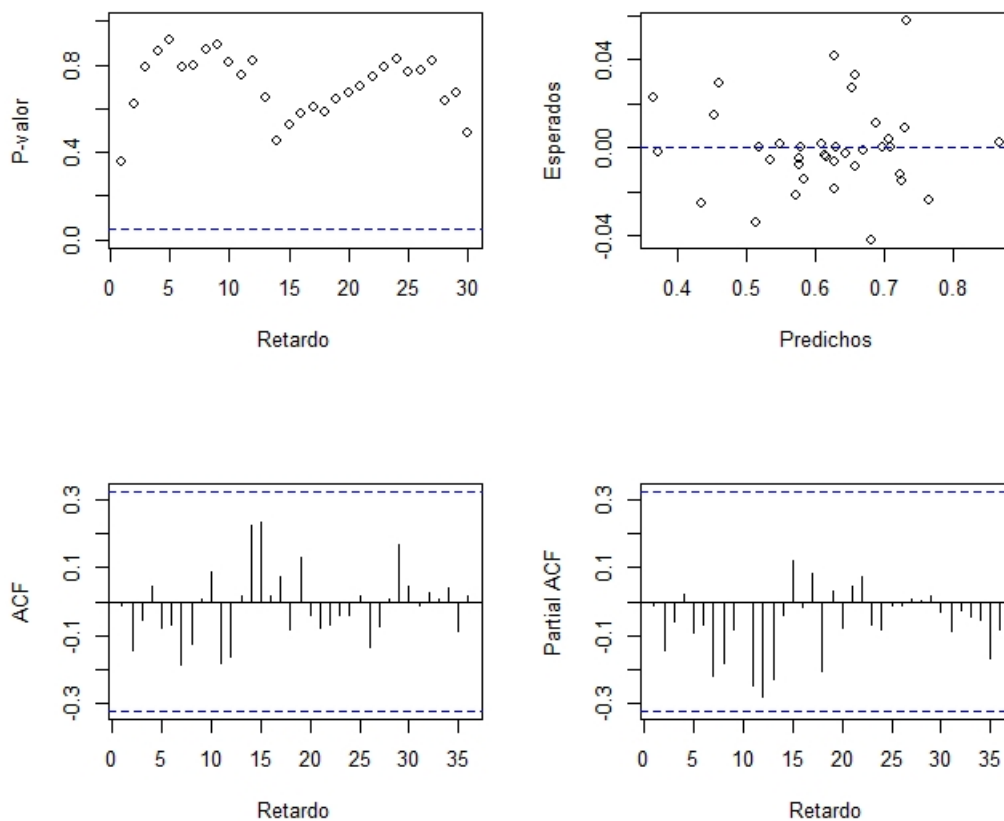


Figura 8: Contrastes gráficos sobre os residuos do modelo. A gráfica superior esquerda, correspóndese cos *p*-valores do contraste de independencia Ljung-Box e, a superior á dereita, ven a sinalar a homocedasticidade dos residuos do modelo. As gráficas inferiores representan as correlacións simples e parciais dos residuos.

.2.5. Contraste das hipóteses básicas sobre os residuos do modelo para factor de condición das merluzas femia

Contraste	Hipóteses	P -valor
Jarque-Bera	Normalidade	0.50
Shapiro-Wilk	Normalidade	0.47
Dickey-Fuller Test	Estacionariedade	0.01
Student's t-Test	Media cero	1

Cadro 6: Contrastes das hipóteses básicas sobre os residuos do modelo para o factor de condición relativo das merluzas femia.

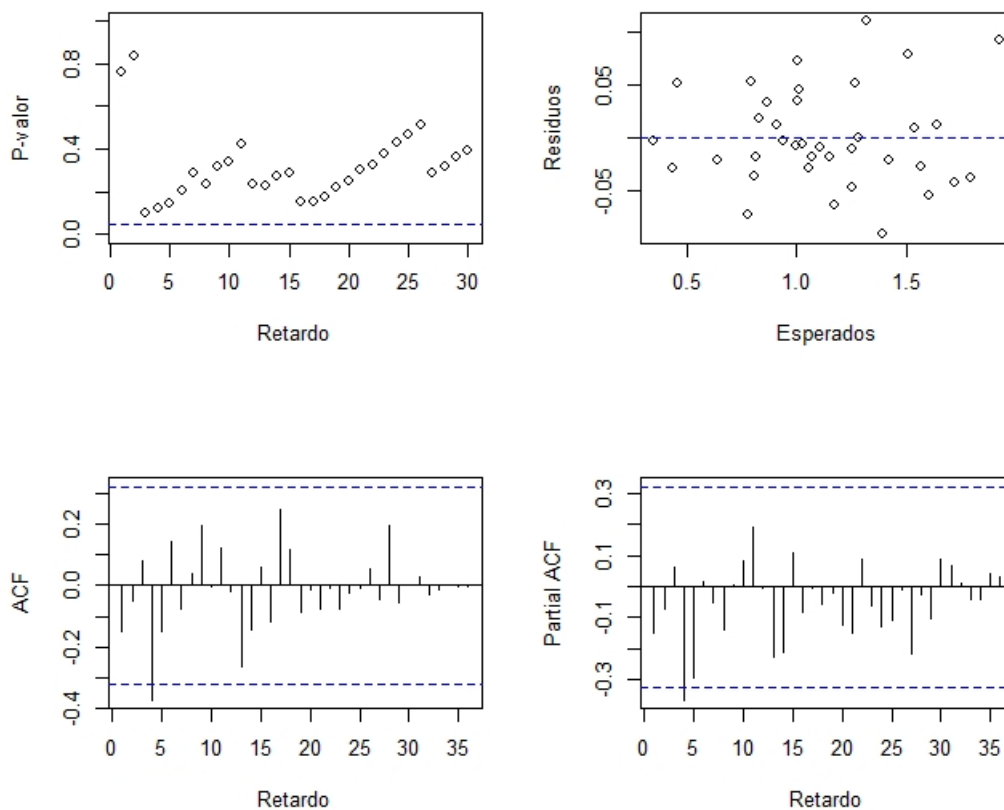


Figura 9: Contrastes gráficos sobre os residuos do modelo. A gráfica superior esquerda, correspóndese cos p -valores do contraste de independencia Ljung-Box e, a superior á dereita, ven a sinalar a homocedasticidade dos residuos do modelo. As gráficas inferiores representan as correlacións simples e parciais dos residuos.