Trabajo Fin de Máster

Modelización de los cierres por toxinas DSP en polígonos de cultivo del mejillón y su relación con las variables ambientales

Adrián Rodríguez López

Máster en Técnicas Estadísticas

Curso 2025-2026

Propuesta de Trabajo Fin de Máster

Título en galego: Modelización dos peches por toxinas DSP en polígonos de cultivo do mexillón e a súa relación coas variables ambientais
Título en español: Modelización de los cierres por toxinas DSP en polígonos de cultivo del mejillón y su relación con las variables ambientales
English title: Modeling DSP Toxin-Related Harvesting Closures in Mussel Farming Areas and Their Relationship with Environmental Variables
Modalidad: Modalidad B
Autor/a: Adrián Rodríguez López, Universidad de A Coruña
Director/a: Javier Roca Pardiñas, Universidad de Vigo
Tutor/a: Xosé Antonio Padín Álvarez , Instituto de Investigacións Mariñas (CSIC); Alexandre Alonso Fernández, Instituto de Investigacións Mariñas (CSIC)
Breve resumen del trabajo: Este trabajo analiza los factores ambientales asociados a los cierres por toxinas DSP en los cultivos de mejillón de Galicia mediante modelos aditivos generalizados (GAM). Además, desarrolla modelos predictivos capaces de anticipar la evolución de estos episodios utilizando variables oceanográficas, climáticas y biológicas.
Recomendaciones:
Otras observaciones:

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

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

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

RODRIGUEZ LOPEZ ADRIAN - 35623182T

Firmado digitalmente por RODRIGUEZ LOPEZ ADRIAN - 35623182T
Fecha: 2026.07.17 22:03:25 +02'00'

77592181W JAVIER ROCA (R: S1500132D)

Firmado digitalmente por 77592181W JAVIER ROCA (R: S1500132D)
Fecha: 2026.07.20 17:10:10 +02'00'

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas e instituciones que han contribuido, de una u otra forma, a la realización de este Trabajo Fin de Máster.

En primer lugar, deseo agradecer al Instituto de Investigaciones Marinas (IIM-CSIC) por acogerme en sus instalaciones y brindarme la oportunidad de desarrollar este trabajo en un entorno de investigación de primer nivel. De manera muy especial, quiero expresar mi agradecimiento a Toni, por su constante disponibilidad, por facilitarme los datos empleados en esta investigación y por compartir conmigo ideas, conocimientos y sugerencias que fueron fundamentales para orientar el tratamiento de la información y el desarrollo del trabajo. Su acompañamiento y apoyo durante todo el proceso han sido esenciales para la realización de este proyecto. Asimismo, quiero agradecer a Álex su ayuda y sus valiosas aportaciones a lo largo del desarrollo del trabajo.

Quiero agradecer también a mi tutor académico, Javier Roca, por su orientación, disponibilidad y apoyo durante todo el proceso. En particular, deseo destacar su ayuda en el desarrollo de los modelos, proporcionándome el script utilizado para la selección de variables y guiándome en su adaptación e interpretación, aspectos clave para la metodología empleada en este trabajo.

Por último, quiero agradecer a mi familia, a mis amigos y a todas aquellas personas que han estado a mi lado durante esta etapa. Su apoyo, confianza, paciencia y ánimo han sido fundamentales para afrontar este reto y hacerlo posible.

Índice general

Resumen **XI**

1. Introducción **1**

2. Material y métodos **5**
2.1. Contexto oceanográfico del área de estudio . . . 5
2.2. Sistema de cultivo del mejillón . . . 6
2.3. Base de datos . . . 9
2.3.1. Variable respuesta . . . 9
2.3.2. Variables explicativas . . . 9

3. Análisis exploratorio **15**

4. Modelización de los cierres por toxinas DSP **21**
4.1. Introducción a los Modelos Aditivos Generalizados (GAM) . . . 21
4.2. Selección de variables . . . 22
4.3. Modelado Explicativo . . . 24
4.3.1. Modelo 1 - BD Completa . . . 24
4.3.2. Modelo 2 - BD Completa, sin toxinas . . . 25
4.3.3. Modelo 3 - BD sin toxinas ni lags DSP . . . 26
4.3.4. Modelo 4 - Manchas solares . . . 27
4.3.5. Modelo 5 - Variables ambientales . . . 28
4.3.6. Modelo 6 - Temperatura, Producción Primaria y Clorofila . . . 29
4.3.7. Conclusiones sobre las variables ambientales . . . 30
4.4. Predicción . . . 31

5. Conclusiones y trabajo futuro **35**
5.0.1. Conclusiones . . . 35
5.0.2. Limitaciones . . . 35
5.0.3. Líneas futuras . . . 36

Bibliografía **37**

Resumen

Resumen en español

Este trabajo analiza los principales factores ambientales asociados a los episodios de cierre de los polígonos de cultivo de mejillón en Galicia provocados por la presencia de toxinas diarreicas (DSP), así como su capacidad predictiva a medio plazo. Para ello se integró una base de datos con información histórica de cierres, variables oceanográficas, meteorológicas, climáticas, hidrológicas y de actividad solar, correspondientes a las principales zonas de producción de mejillón de las rías gallegas. Mediante modelos aditivos generalizados (GAM) se evaluó la influencia de los distintos predictores y se exploraron las relaciones no lineales entre las variables ambientales y la duración anual de los cierres.

Los resultados muestran que la persistencia temporal de los episodios de cierre y las variables relacionadas con el afloramiento costero constituyen los principales factores explicativos de la variabilidad observada. Asimismo, variables como la temperatura superficial del mar, los aportes fluviales, los índices climáticos de gran escala y la actividad solar presentaron una contribución significativa, aunque de menor magnitud. Los mejores modelos alcanzaron una elevada capacidad explicativa, con valores superiores al 90 % cuando se incorporaron variables relacionadas con las toxinas y cercanos al 80 % utilizando únicamente variables ambientales. Además, los modelos predictivos mostraron un mejor rendimiento para horizontes temporales de dos a cinco años, evidenciando el potencial de estas metodologías como herramienta de apoyo para la gestión y planificación del sector mejillonero gallego.

English abstract

This study investigates the main environmental factors associated with shellfish harvesting closures caused by diarrhetic shellfish poisoning (DSP) toxins in the mussel farming areas of Galicia (NW Spain), as well as their medium-term predictive potential. A historical database was compiled, integrating information on harvesting closures together with oceanographic, meteorological, climatic, hydrological, and solar activity variables from the main Galician mussel production areas. Generalized Additive Models (GAMs) were used to assess the influence of the different predictors and to characterize the nonlinear relationships between environmental variables and the annual duration of harvesting closures.

The results indicate that the temporal persistence of closure events and coastal upwelling-related variables are the most influential factors explaining the observed variability. Sea surface temperature, river discharge, large-scale climate indices, and solar activity also showed significant effects, although their contribution was comparatively smaller. The best-performing models achieved a high explanatory power, exceeding 90 % when toxin-related variables were included and around 80 % when only environmental predictors were considered. Furthermore, the predictive models performed better over forecasting horizons of two to five years than for one-year forecasts, highlighting the potential of these approaches as decision-support tools for improving the management and planning of the Galician mussel farming sector.

Capítulo 1

Introducción

Figura 1.1: Bateas de mejillones

Las floraciones de algas nocivas (Harmful Algal Blooms, HABs) constituyen uno de los principales desafíos ambientales para los ecosistemas costeros y la acuicultura a escala mundial. Aunque la proliferación de microalgas forma parte de la dinámica natural de los ecosistemas marinos, determinadas especies pueden alcanzar concentraciones elevadas o producir potentes toxinas que afectan a la fauna marina, comprometen la seguridad alimentaria y generan importantes pérdidas económicas asociadas al cierre de zonas de producción de moluscos bivalvos. Durante las últimas décadas se ha documentado un aumento en la frecuencia, duración y extensión geográfica de estos episodios en numerosas regiones del mundo, favorecido tanto por la mejora de los programas de observación como por cambios ambientales relacionados con el calentamiento oceánico, la modificación de los patrones de circulación, la eutrofización y el incremento de los fenómenos climáticos extremos [11, 3, 22].

Entre las diferentes tipologías de HABs, las asociadas a especies del género Dinophysis poseen una especial relevancia en las costas europeas debido a la producción de ácido okadaico y dinofisistoxinas, compuestos lipofílicos responsables de intoxicación diarreica por moluscos (Diarrhetic Shellfish Poisoning, DSP). Estas toxinas son acumuladas por moluscos filtradores durante su alimentación sin producirles efectos fisiológicos aparentes, aunque pueden alcanzar concentraciones superiores a los límites establecidos por la legislación sanitaria para el consumo humano. Cuando esto ocurre, las autoridades competentes decretan el cierre temporal de las zonas de producción hasta que los niveles de toxinas descienden por debajo de los umbrales legales, una medida imprescindible para garantizar la

seguridad alimentaria, pero con importantes repercusiones económicas para el sector productor [18, 6].

Galicia constituye una zona afectada por una elevada persistencia e incidencia de cierres de la extracción de mejillón por la acumulación de toxinas lipofílicas en este recurso debido a las frecuentes proliferaciones de dinoflagelados del género Dinophysis. La comunidad autónoma concentra prácticamente la totalidad de la producción española de mejillón (Mytilus galloprovincialis) y se sitúa entre las principales regiones productoras a escala internacional. Según los datos más recientes de la FAO y del Anuario de Acuicultura de Galicia, la producción gallega representa alrededor del 97 % de la producción nacional y cerca de la mitad de la producción de la Unión Europea, con una producción media próxima a las 230.000 toneladas anuales y un valor económico de varios cientos de millones de euros. Más allá de su dimensión productiva, la mitilicultura constituye una actividad estratégica para la economía del litoral gallego, sustentando miles de empleos directos e indirectos y una extensa industria de transformación, comercialización y exportación [7, 5, 26].

El éxito de la mitilicultura gallega está estrechamente ligado a las características oceanográficas de las Rías Baixas. Estas rías forman parte del sistema de afloramiento del margen ibérico noroccidental, el límite septentrional del gran sistema de afloramiento del Atlántico Nordeste. Durante la primavera y el verano, los vientos dominantes del norte y nordeste inducen el transporte de Ekman hacia mar abierto y favorecen el ascenso de aguas oceánicas frías y ricas en nutrientes sobre la plataforma continental. La entrada de estas aguas por el fondo de las rías intensifica la circulación estuarina positiva y sostiene elevadas tasas de producción primaria, situando a las Rías Baixas entre los ecosistemas costeros más productivos del Atlántico europeo [9, 8, 2].

Paradójicamente, las mismas condiciones que favorecen la elevada producción fitoplanctónica también convierten a las rías gallegas en un entorno especialmente favorable para el desarrollo recurrente de HABs. En Galicia, las toxinas DSP representan la principal causa de cierre temporal de los polígonos de cultivo de mejillón y constituyen el episodio tóxico más frecuente registrado por los programas oficiales de vigilancia. La recurrencia de estas proliferaciones responde a la interacción de procesos físicos, químicos y biológicos que actúan a distintas escalas temporales y espaciales, incluyendo la intensidad del afloramiento costero, la estratificación de la columna de agua, los aportes continentales, la disponibilidad de nutrientes, la circulación residual de las rías y la dinámica poblacional de Dinophysis. Esta naturaleza multifactorial explica que, pese a más de cuatro décadas de investigación, la predicción de estos episodios continúe siendo uno de los principales retos de la oceanografía costera y de la gestión de la acuicultura [16, 20, 21, 17].

Durante los últimos años, el incremento de la disponibilidad de observaciones oceanográficas y meteorológicas de alta resolución, junto con el desarrollo de nuevas técnicas de análisis estadístico y aprendizaje automático, ha abierto nuevas posibilidades para estudiar este tipo de fenómenos complejos. A diferencia de los enfoques tradicionales basados en relaciones bivariantes entre una única variable ambiental y la presencia de microalgas tóxicas, las técnicas modernas de modelización permiten integrar simultáneamente un elevado número de variables explicativas, explorar relaciones no lineales e identificar interacciones entre procesos que difícilmente pueden detectarse mediante aproximaciones clásicas. No obstante, la capacidad real de estas metodologías para describir la ocurrencia de episodios DSP en sistemas altamente dinámicos como las rías gallegas continúa siendo poco conocida y constituye un área de investigación de creciente interés [4, 22, 13].

De este modo, el objetivo de este proyecto será desarrollar y evaluar modelos predictivos capaces de estimar el porcentaje de cierres de los polígonos de producción de mejillones debido a la presencia de toxinas DSP, y establecer relaciones con los factores que los condicionan. Para ello, se elaborará una base de datos mediante la integración de variables meteorológicas, oceanográficas y ambientales obtenidas de diferentes fuentes de información. Asimismo, se calcularán variables derivadas adicionales con el fin de aportar información complementaria que pueda mejorar el rendimiento del modelo.

Posteriormente, tras elegir el modelo, se llevará a cabo un proceso de selección de variables mediante un algoritmo de selección de características (feature selection), con el objetivo de identificar el conjunto de predictores que aporte una mayor capacidad explicativa y predictiva al modelo. Una vez definido el modelo final, se evaluará su rendimiento mediante las métricas correspondientes y se analizará su capacidad predictiva. Finalmente, el modelo se utilizará para realizar predicciones de los cierres de polígonos debido a las toxinas DSP en años posteriores.

Capítulo 2

Material y métodos

2.1. Contexto oceanográfico del área de estudio

El estudio se desarrolló en las principales zonas de cultivo de mejillón de Galicia, localizadas entre los 42 y 43º N de la costa noroeste de la Península Ibérica (Figura2.2). La producción se concentra en las rías de Vigo, Pontevedra, Arousa y Muros-Noia, pertenecientes a las Rías Baixas, y en la ría de Ares-Betanzos, situada en la costa norte gallega. En conjunto, estas cinco rías albergan más de 3.300 bateas distribuidas en polígonos de cultivo gestionados por la administración autonómica, constituyendo la principal región productora de mejillón de Europa.

Desde el punto de vista oceanográfico, las rías gallegas son valles fluviales inundados que funcionan como estuarios parcialmente mezclados. Su circulación está dominada por el denominado patrón estuarino positivo, caracterizado por una salida superficial de aguas menos salinas procedentes del drenaje continental y una entrada en profundidad de aguas oceánicas más densas. Esta circulación favorece el intercambio continuo entre la plataforma continental y el interior de las rías, condicionando la distribución de nutrientes, fitoplancton y partículas en suspensión [9, 19].

La intensidad de esta circulación está fuertemente controlada por el régimen estacional de vientos asociado a la posición del anticiclón de las Azores. Durante la primavera y el verano predominan los vientos de componente norte, que inducen transporte de Ekman hacia mar abierto y favorecen episodios de afloramiento costero. Como consecuencia, aguas frías, ricas en nutrientes y relativamente pobres en oxígeno ascienden sobre la plataforma e penetran por el fondo de las rías, intensificando la circulación estuarina y estimulando el desarrollo de floraciones fitoplanctónicas. Por el contrario, durante el otoño y el invierno predominan los vientos del sur y suroeste, generando condiciones de hundimiento (downwelling), una disminución del aporte oceánico de nutrientes y una mayor retención de las aguas superficiales en el interior de las rías [1, 8].

El sistema de afloramiento del margen ibérico noroccidental constituye el extremo septentrional del gran sistema de afloramiento del Atlántico Nordeste, uno de los cuatro principales sistemas de afloramiento costero del planeta. La coincidencia temporal entre la disponibilidad de nutrientes aportados por el afloramiento y la elevada radiación solar durante la primavera y el verano permite alcanzar tasas excepcionalmente altas de producción primaria, situando a las Rías Baixas entre los ecosistemas costeros más productivos de Europa. Esta elevada disponibilidad de alimento explica las altas tasas de crecimiento del mejillón gallego, pero también favorece el desarrollo recurrente de determinadas especies fitoplanctónicas potencialmente tóxicas.

Las proliferaciones de Dinophysis muestran una estrecha relación con esta variabilidad hidrodinámica. Numerosos estudios han demostrado que la alternancia entre episodios de afloramiento y relajación,

el grado de estratificación de la columna de agua, la circulación residual de las rías y la posición de los frentes hidrográficos condicionan el transporte, la acumulación y el desarrollo de estas poblaciones. En consecuencia, la evolución de los episodios DSP responde a la interacción de procesos que actúan desde escalas sinópticas, controladas por el viento, hasta escalas estacionales e interanuales relacionadas con la variabilidad climática del Atlántico Nordeste. Esta complejidad convierte a las rías gallegas en un entorno idóneo para evaluar el potencial de modelos que integren simultáneamente información meteorológica, oceanográfica y ambiental.

Las principales productores de las toxinas DSP en la costa gallega son los dinoflagelados Dinophysis acuminata y Dinophysis acuta.

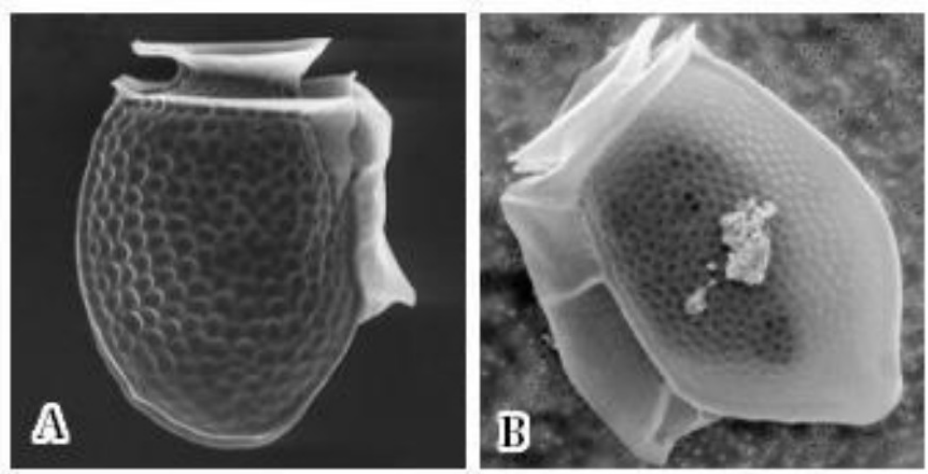


Figura 2.1: Dinophysis acuminata (A) y Dinophysis acuta (B). Fuente: [10].

2.2. Sistema de cultivo del mejillón

La mitilicultura gallega se desarrolla casi exclusivamente mediante un sistema extensivo de cultivo en suspensión sobre plataformas flotantes denominadas bateas, una tecnología introducida en la década de 1940 y que ha experimentado escasas modificaciones desde su implantación. Actualmente, el cultivo se distribuye en más de 3.300 bateas agrupadas en polígonos de producción localizados principalmente en las Rías Baixas y, en menor medida, en la ría de Ares-Betanzos (Figura 2.2). A pesar de que el número de bateas se ha mantenido prácticamente estable desde principios de la década de 1970, las mejoras en el manejo del cultivo y la elevada productividad natural de las rías han permitido consolidar a Galicia como la principal región productora de mejillón de Europa y una de las más importantes del mundo [15, 26].

Área de cultivo	**Nº bateas**	**Producción (t)**	**Producción/batea (t)**	**Bateas/titular**	**Titulares >4 bateas**
Ría de Ares–Sada	103	7 150	69.41	14.71	4
Ría de Muros–Noia	118	8 225	69.70	4.07	8
Ría de Arousa Norte	699	174 000	75.91	1.10	2
Ría de Arousa Sur	1 593	–	–	1.19	7
Ría de Pontevedra	346	26 125	75.51	2.42	10
Ría de Vigo	478	34 500	72.18	1.58	8
Galicia	**3 337**	**250 000**	**74.92**	**1.35**	**42**

Cuadro 2.1: Distribución de bateas por zona de cultivo y distribución de la propiedad de las concesiones en 1997.

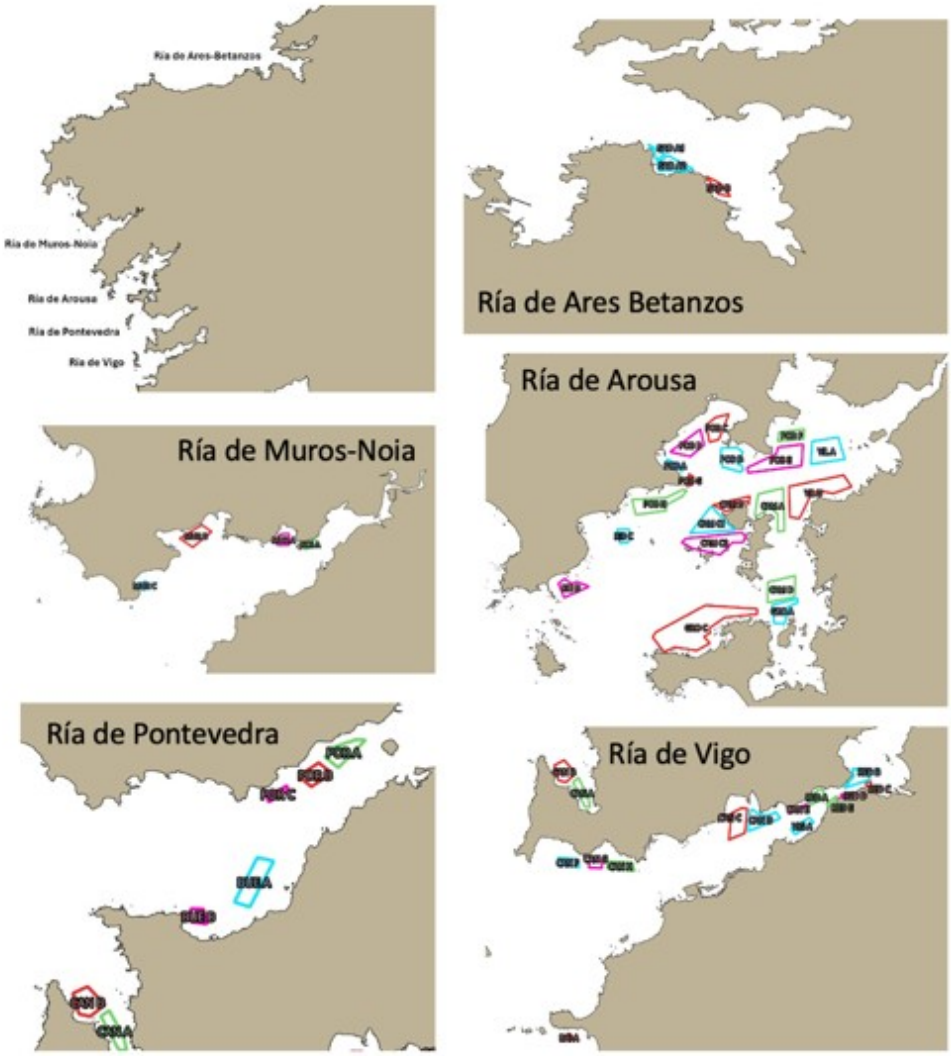


Figura 2.2: Distribución geográfica de las principales rías gallegas y localización de los polígonos de cultivo de mejillón incluidos en el estudio.

Cada batea consiste en una plataforma flotante de madera, de aproximadamente 500 m^2, fondeada al fondo mediante un sistema de anclaje del que se suspenden hasta 500 cuerdas de cultivo con una longitud máxima cercana a los 12 m. Sobre estas cuerdas se fija la semilla de mejillón (Mytilus galloprovincialis), obtenida tradicionalmente de afloramientos rocosos del litoral o mediante colectores instalados en las propias bateas. Durante los primeros días, los juveniles permanecen sujetos mediante una malla biodegradable que posteriormente desaparece, quedando fijados de forma permanente gracias a los filamentos del biso.

Tras aproximadamente cinco o seis meses de crecimiento se realiza el desdoble, una operación en la que los individuos se redistribuyen sobre nuevas cuerdas para reducir la competencia por espacio y alimento y favorecer un crecimiento homogéneo. A partir de ese momento los mejillones permanecen en

suspensión hasta alcanzar la talla comercial, proceso que, dependiendo de las condiciones ambientales y de la disponibilidad de alimento, suele completarse entre los 9 y los 18 meses. Las elevadas concentraciones de fitoplancton asociadas al afloramiento costero permiten alcanzar densidades de cultivo excepcionalmente altas y tasas de crecimiento entre las mayores descritas para esta especie.

El carácter filtrador del mejillón convierte a esta especie en un excelente indicador de la calidad del medio marino, pero también favorece la acumulación de contaminantes y biotoxinas presentes en el agua. Durante los episodios de proliferación de especies tóxicas del género Dinophysis, los mejillones incorporan las células durante la filtración y acumulan ácido okadaico y dinofisistoxinas en sus tejidos. La concentración de estas toxinas depende tanto de la intensidad y duración de la proliferación como de la tasa de filtración, la temperatura, la disponibilidad de alimento y los procesos fisiológicos de depuración del propio organismo.

Con objeto de garantizar la seguridad alimentaria, Galicia dispone de uno de los programas de monitorización de biotoxinas marinas más extensos y continuados de Europa, basado en el muestreo periódico de fitoplancton y moluscos en todos los polígonos de producción. Este programa está gestionado por el Instituto Tecnolóxico para o Control do Medio Mariño de Galicia [14]. Cuando la concentración de toxinas supera los límites establecidos por la legislación europea, la extracción y comercialización de moluscos queda temporalmente suspendida hasta que los niveles vuelven a situarse por debajo del umbral reglamentario. Estos cierres constituyen la principal consecuencia económica de los episodios DSP y representan la variable de respuesta analizada en el presente trabajo para evaluar su relación con las condiciones meteorológicas, oceanográficas y ambientales.

2.3. Base de datos

Se ha construido una base de datos anual a partir de la integración de distintas fuentes de información. Los datos abarcan el período comprendido entre 1991 y 2025, con algunos datos adicionales de años anteriores. En general, los datos anuales son comunes para todos los polígonos, excepto que se indique en el apartado considerado.

2.3.1. Variable respuesta

- **DSP**: Número de días de cierre anual por presencia de toxinas lipofílicas (DSP)

2.3.2. Variables explicativas

Además de la variable respuesta, la base de datos incluye 90 variables explicativas. A continuación se presentan divididas por categorías.

Anomalías

Se incluyen anomalías de ciertos grupos de variables explicativas. Se indicarán con **(*)**. Hay 2 diferentes que se detallan a continuación.

- **Anom_XXX_Anual**: Para la variable indicada, se calculan las medias mensuales de toda la serie temporal por polígono, y se le substraen a los valores de ese año. Finalmente, esas diferencias mensuales se promedian para obtener valores anuales.

- **Anom_XXX_Abrnov**: Se calcula de forma análoga a la anual, pero solamente teniendo en cuenta el periodo de cierres altos **Abril-Noviembre**.

Clasificación

- **Poligono**: Nombre del polígono de cultivo.
- **Pol_id**: Identificador numérico asignado a cada polígono de producción (Aparecen ordenados de norte a sur).
- **Ano**: Año correspondiente a la observación.
- **Ría**: Ría a la que pertenece el polígono de cultivo.

Biotoxinas

Nota: Estas variables son **únicas** para cada polígono.

- **ASP**: Número de días de cierres anuales por presencia de toxinas amnésicas (ASP).
- **PSP**: Número de días de cierres anuales por presencia de toxinas paralizantes (PSP).
- **AO_Medio**: Concentración media anual de ácido okadaico.
- **AO_Max**: Concentración máxima anual de ácido okadaico.
- **AO_Min**: Concentración mínima anual de ácido okadaico.

Oceanográficas (*)

Nota: Estas variables son **únicas** para cada polígono.

- **Temp_Diarias**: Temperatura media diaria anual de la superficie del mar.
- **Prod_Primaria**: Producción primaria media anual del fitoplancton.
- **Clorofila**: Concentración media anual de clorofila.

Ambientales (*, excepto Sunspot)

- **WS_RRAA**: Velocidad media anual del viento en las rías altas.
- **WS_RRBB**: Velocidad media anual del viento en las rías baixas.
- **UI_RRAA**: Índice medio anual de afloramiento en las rías altas.
- **UI_RRBB**: Índice medio anual de afloramiento en las rías baixas.
- **Sunspot_Mean**: Número medio anual de manchas solares.
- **Sunspot_SD**: Desviación típica anual del número de manchas solares.

Ríos (*)

- **Miño**: Caudal anual del río Miño
- **Rio_Estuario**: Caudal del río principal (o conjunto de ríos) asociado a la ría donde se localiza cada polígono. Se muestran en la tabla 2.2.

Ría	**Río(s) considerado(s)**
Ares-Betanzos	Eume + Mandeo
Muros-Noia	Tambre
Arousa	Ulla + Umia
Pontevedra	Lérez
Vigo	Verdugo

Cuadro 2.2: Ríos considerados para el cálculo del caudal asociado a cada ría.

Afloramiento

Se calcularon una serie de variables a partir del índice de afloramiento del periodo de estudio. Cada una de ellas se descompone en dos, correspondiéndose con las rías altas y bajas (RRAA y RRBB). Se muestran a continuación:

- **AF_Media**: Media del índice de afloramiento anual.
- **AF_Mediana**: Mediana del índice de afloramiento anual.
- **AF_Dias_Positivos**: Número de días con afloramiento anual positivo.
- **AF_Dias_Negativos**: Número de días con afloramiento anual negativo.
- **AF_Suma_Total**: Sumatorio del índice de afloramiento anual.
- **AF_Suma_Positivos**: Sumatorio de todos los valores positivos del índice de afloramiento anual.
- **AF_Suma_Negativos**: Sumatorio de todos los valores negativos del índice de afloramiento anual.
- **AF_Eventos_Negativos**: Número de eventos persistentes negativos anuales (definidos como períodos de tres días o más con valores consecutivos positivos o negativos).
- **AF_Duracion_Media_Neg**: Duración media anual de los eventos persistentes negativos.
- **AF_Eventos_Positivos**: Número de eventos persistentes positivos anuales (definidos como períodos de tres días o más con valores consecutivos positivos o negativos).
- **AF_Duracion_Media_Pos**: Duración media anual de los eventos persistentes positivos.
- **AF_Intermitencia**: Intermitencia del afloramiento, definida como el número de cambios de signo del índice a lo largo del año.
- **AF_Dias_Pos_250**: Número de días con un índice de afloramiento intenso superior a 250 m^3 s^{-1} kg^{-1}.
- **AF_Dias_Neg_250**: Número de días con un índice de afloramiento intenso inferior a -250 m^3 s^{-1} kg^{-1}.
- **AF_Ratio_250**: Cociente entre el número de días con afloramiento intenso positivo y el número de días con afloramiento intenso negativo para el umbral de ± 250 m^3 s^{-1} kg^{-1}.

- **AF_Dias_Pos_750**: Número de días con un índice de afloramiento intenso superior a 750 m^3 s^{-1} kg^{-1}.

- **AF_Dias_Neg_750**: Número de días con un índice de afloramiento intenso inferior a -750 m^3 s^{-1} kg^{-1}.

- **AF_Ratio_750**: Cociente entre el número de días con afloramiento intenso positivo y el número de días con afloramiento intenso negativo para el umbral de ± 750 m^3 s^{-1} kg^{-1}.

Índices climáticos (*)

- **NAO**: Índice de la Oscilación del Atlántico Norte (*North Atlantic Oscillation*), calculado a partir de la diferencia de presión atmosférica entre las Azores e Islandia. Este índice controla la intensidad y trayectoria de los vientos del oeste y de las borrascas en el Atlántico Norte.

- **wNAO**: Valor medio invernal del índice NAO (*winter North Atlantic Oscillation*), calculado habitualmente para el periodo comprendido entre diciembre y marzo, cuando su influencia sobre la circulación atmosférica y oceánica es máxima.

- **EA**: Índice del patrón del Atlántico Este (*East Atlantic Pattern*), segundo modo principal de variabilidad atmosférica del Atlántico Norte. Modula la temperatura, la precipitación y los patrones de viento sobre Europa y el Atlántico.

- **AMO**: Índice de la Oscilación Multidecenal del Atlántico (*Atlantic Multidecadal Oscillation*), que describe las variaciones de la temperatura superficial del Atlántico Norte a escala multidecenal y ejerce una influencia sobre el clima, la circulación oceánica y la productividad marina.

Retardos

Se crearon retardos tanto de la variable de interés **DSP** para incorporar la tendencia temporal de la serie, como de las manchas solares **Sunspot Mean** debido a su posible relación observada con la variable de interés durante el análisis exploratorio (Ver Figura 3.5).

- **DSP_lag1**: Número de días de cierre por toxinas DSP del año anterior.

- **DSP_lag2**: Número de días de cierre por toxinas DSP correspondiente a dos años antes.

- **DSP_lag3**: Número de días de cierre por toxinas DSP correspondiente a tres años antes.

- **DSP_lag12**: Número de días de cierre por toxinas DSP correspondiente a doce años antes.

- **Sunspot_lag1**: Número medio de manchas solares correspondiente al año anterior.

- **Sunspot_lag2**: Número medio de manchas solares correspondiente a dos años antes.

- **Sunspot_lag3**: Número medio de manchas solares correspondiente a tres años antes.

- **Sunspot_lag4**: Número medio de manchas solares correspondiente a cuatro años antes.

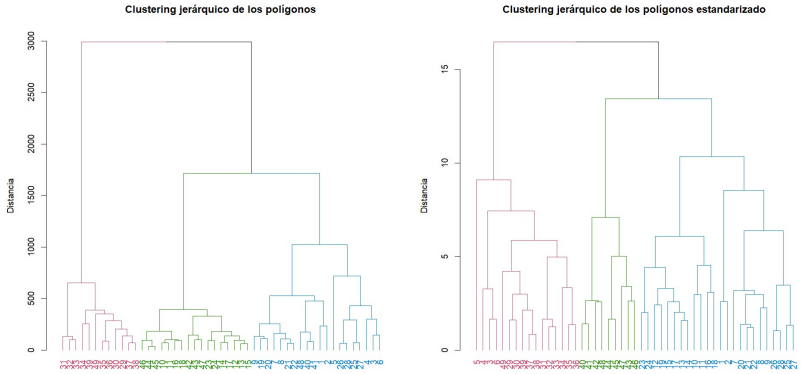
- **Sunspot_lag5**: Número medio de manchas solares correspondiente a cinco años antes.

- **Sunspot_lag6**: Número medio de manchas solares correspondiente a seis años antes.

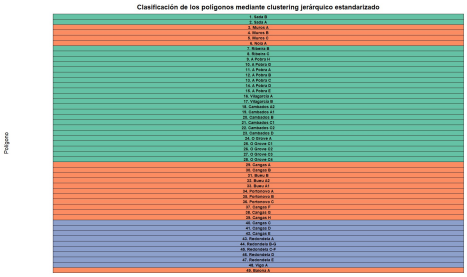
Derivadas

- **grupo**: Grupo asignado a cada polígono mediante el análisis de conglomerados jerárquico.
- **grupo_std**: Grupo asignado a cada polígono mediante el análisis de conglomerados jerárquico tras la estandarización de las variables.

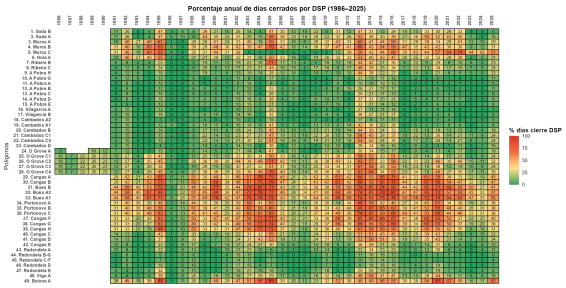
El objetivo de estas variables, consistió en dividir la serie en una serie de grupos, distinguiendo cierres bajos, medios y altos. Para ello, se empleó el clustering jerárquico, ya que son pocos datos y se puede interpretar bien gracias al dendograma. Mostramos a continuación, los dendogramas, las tablas con los clústeres ordenados y el heatmap para comparar los resultados.



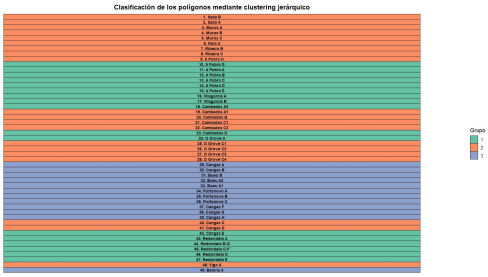
(a) Dendogramas de grupo y grupo_std



(b) Grupo_std ordenada geográficamente



(c) Heatmap cierres anuales DSP



(d) Grupo_std ordenada geográficamente

Figura 2.3: Resultados clustering jerárquico

Vemos que ambos clústeres son capaces de distinguir 3 grupos de cierres bajos, medios y altos. En concreto, vemos que la versión estandarizada agrupa los polígonos en bloques grandes contiguos.

Capítulo 3

Análisis exploratorio

Como paso previo al desarrollo de los modelos explicativos y predictivos, se realizó un análisis exploratorio con el objetivo de caracterizar la distribución de los episodios de cierre por presencia de toxinas DSP en los polígonos de cultivo de mejillón de Galicia.

En primer lugar, la Figura 3.1 muestra la distribución anual del porcentaje de días de cierre por toxinas DSP para cada polígono de cultivo durante el periodo (1986-2025).

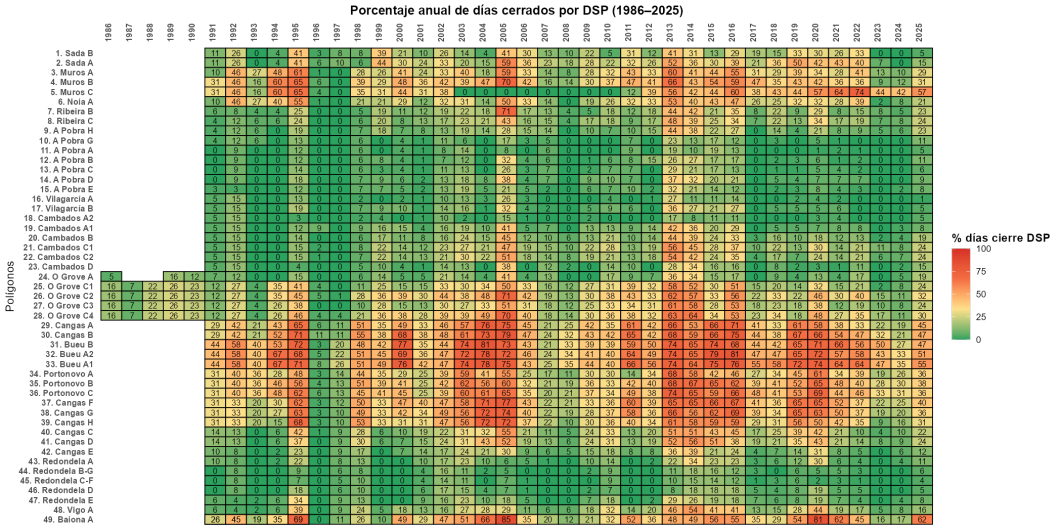
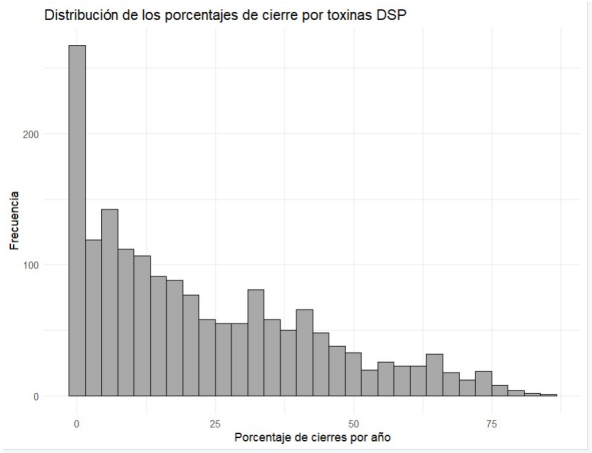


Figura 3.1: Porcentaje de cierres anuales por toxinas DSP

En primer lugar, destaca la naturaleza espacio-temporal de los datos, permitiendo distinguirse zonas con cierres altos y bajos unidas de forma continua. Aquí hay que tener en cuenta que los polígonos están ordenados siguiendo la costa gallega de norte a sur, empezando en Sada y acabando en Baiona, en la parte inferior de la ría de Vigo. Además, vemos comportamientos similares en muchos polígonos dentro de las rías: En **Sada (1-2)**, **Muros-Noia (3-6)**, **Arousa (7-24)**, **Pontevedra (29-39)** y **Vigo (40-48)**. Para terminar, las zonas rojas parecen indicar una cierta periodicidad de las etapas de cierres altos, aunque tampoco se da de forma precisa.

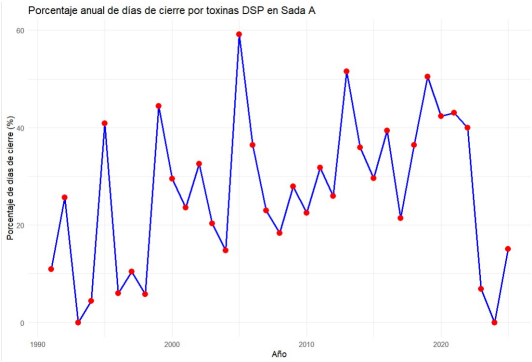
Para complementar el análisis de la distribución de los cierres, se presenta a continuación un histograma, además de la evolución temporal del porcentaje anual de días de cierre en cuatro polígonos representativos. La selección incluye un polígono con baja incidencia de cierres, otro con incidencia intermedia y otro con elevada incidencia, con el objetivo de ilustrar los distintos patrones observados

en la serie histórica. Además, se incorpora el caso de **Baiona A**, considerado de forma independiente debido a su comportamiento diferenciado respecto al resto de polígonos.

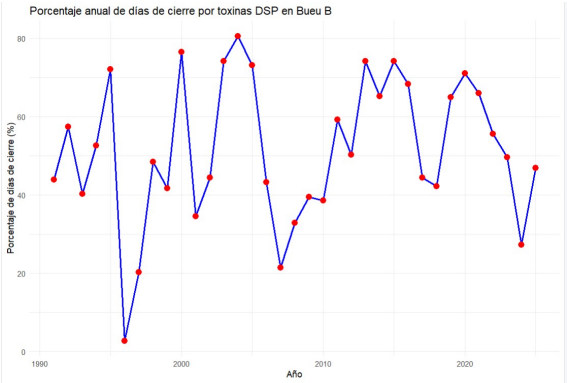


(a) Distribución de los porcentajes anuales de cierre por DSP.

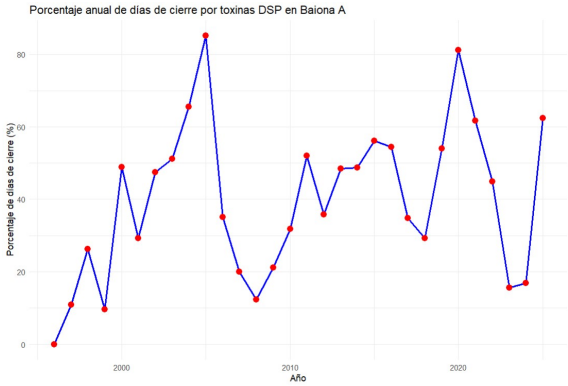
(b) Polígono con baja incidencia.



(c) Polígono con incidencia media.



(d) Polígono con alta incidencia.



(e) Caso particular de Baiona A.

Figura 3.2: Distribución de los porcentajes anuales de cierre y evolución temporal en cuatro polígonos representativos.

Relacionado con los cierres anuales, también es interesante analizar cómo se distribuyen los cierres a lo largo del año, es decir, en que meses se producen. Lo podemos ver en la siguiente figura:

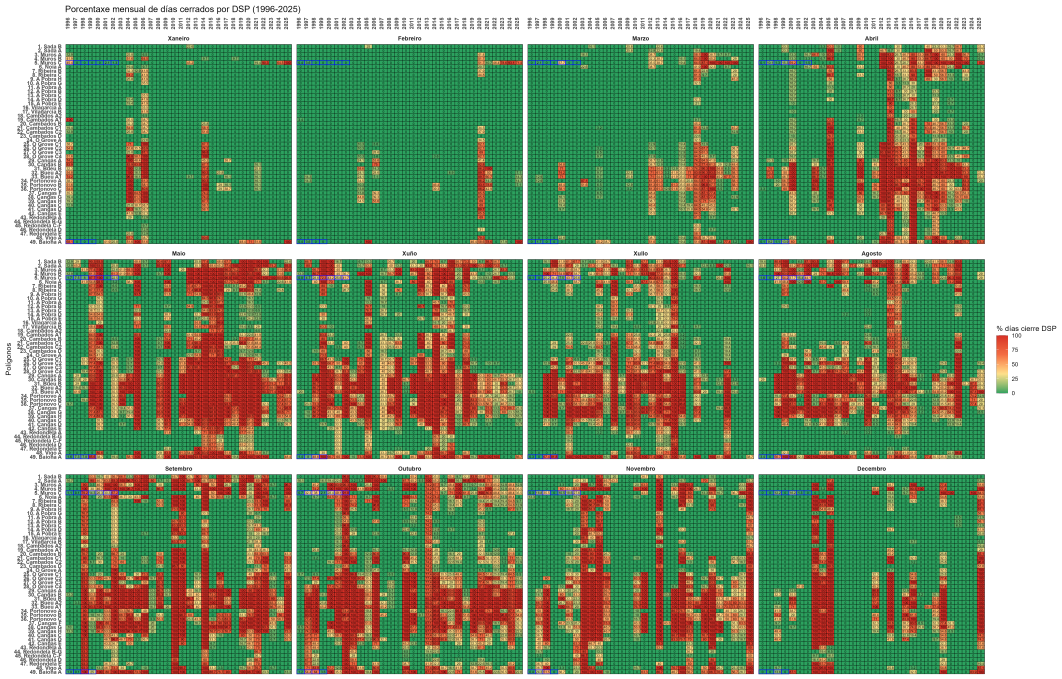


Figura 3.3: Porcentaje de cierres mensuales por toxinas DSP

Analizar esta figura en profundidad se escapa de nuestro interés ya que trabajaremos con datos anuales. Aún así, sí que ve que los episodios de cierre se concentran en meses determinados a lo largo del año. Asimismo, a partir de ahora consideraremos el periodo de **Abril-Noviembre** como el periodo de cierres altos. Este dato lo emplearemos para calcular anomalías de determinadas variables durante ese periodo, ya que disponemos de los datos diarios y/o mensuales y podemos aprovecharnos de ello.

Continuamos el análisis mostrando la matriz de correlaciones de las variables. Dado que tenemos una gran cantidad de ellas y es complicado distinguirlas cuando tenemos tantas, se seleccionarán las más relevantes.

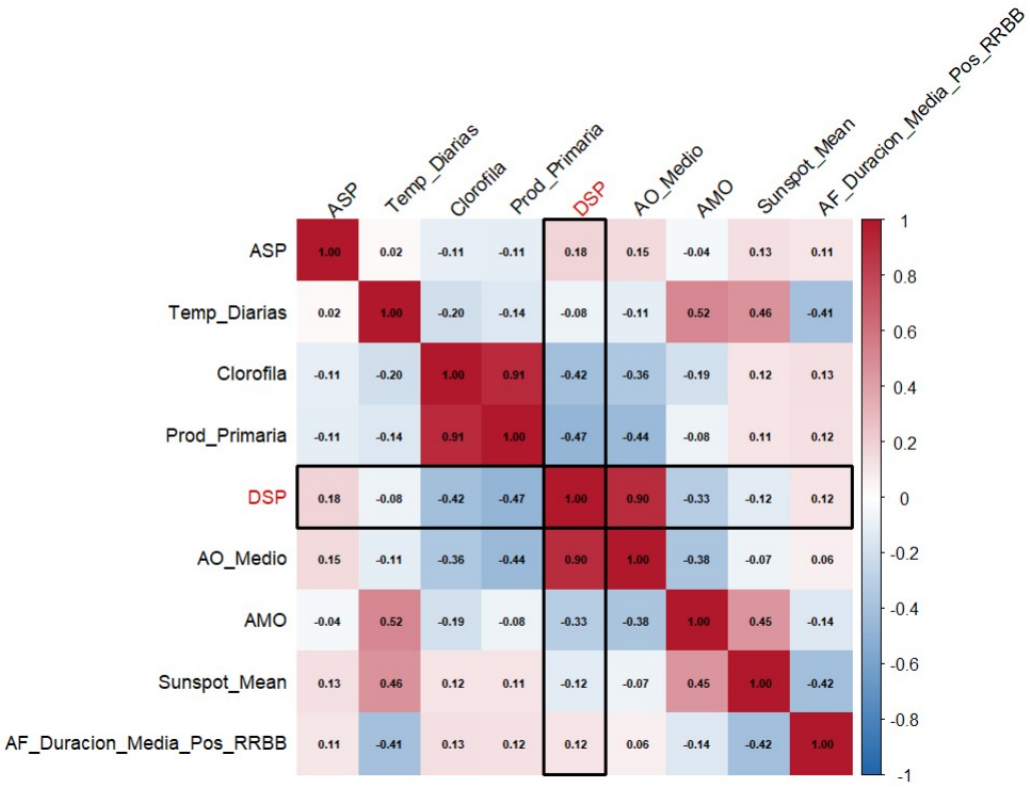
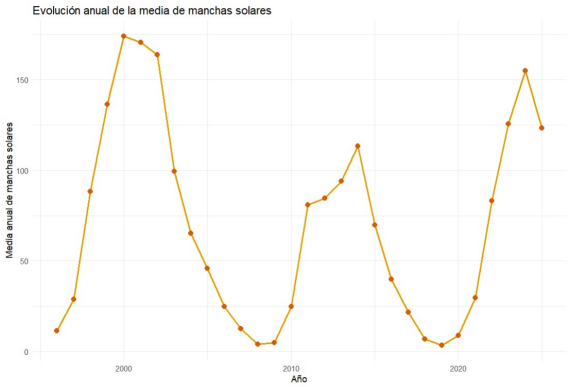


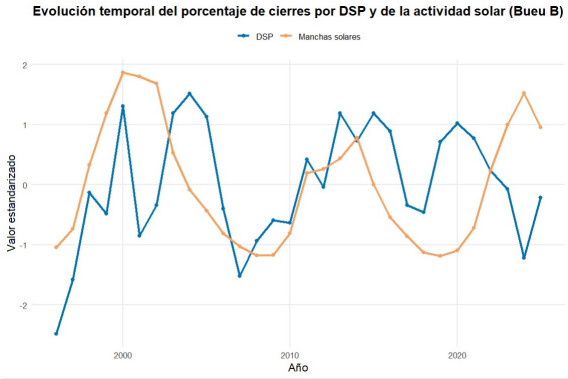
Figura 3.4: Matriz de correlación de DSP con variables

Destaca principalmente la correlación del ácido okadaico medio (**AO_medio**) con la variable de interés **DSP** con un valor de **0.9**. Esto se debe a que el **AO** es el principal causante de las intoxicaciones por **DSP** en humanos. También se muestran otras correlaciones, aunque no tan significativas como por ejemplo **Clorofila** o **Producción Primaria** con valores superiores a **0.4** (negativos en este caso). Además, también se muestra una alta correlación entre determinadas variables ambientales, como la **Clorofila** y la **Producción Primaria**.

Para terminar con el análisis, mostramos otro resultado interesante obtenido. Se trata de las manchas solares, que siguen una tendencia cíclica de unos 11 años. Al comparar su evolución con la de los cierres de los polígonos con valores altos, parece que intenta replicar su tendencia (en cierto modo). Mostramos este fenómeno en la Figura 3.5



(a) Evolución de las manchas solares



(b) Comparación de las manchas solares con los cierres de Bueu B

Figura 3.5: Evolución y comparación de las manchas solares con los cierres (1996-2025)

Capítulo 4

Modelización de los cierres por toxinas DSP

4.1. Introducción a los Modelos Aditivos Generalizados (GAM)

Los Modelos Aditivos Generalizados (Generalized Additive Models, GAM) constituyen una extensión de los Modelos Lineales Generalizados (GLM) que permite describir relaciones no lineales entre una variable respuesta y un conjunto de variables explicativas mediante funciones suavizadas estimadas a partir de los propios datos [12, 25].

A diferencia de los modelos lineales tradicionales, donde cada predictor contribuye mediante un coeficiente constante, los GAM sustituyen estas relaciones por funciones suaves capaces de adaptarse a comportamientos más complejos. De esta forma, el modelo no impone una relación lineal entre las variables, sino que permite que la forma de dicha relación venga determinada por la información contenida en los datos.

De forma general, un modelo GAM puede expresarse como:

$$g(\mu_i) = \beta_0 + \sum_{j=1}^{p} f_j(x_{ij}) \tag{4.1}$$

donde g representa la función de enlace, μ_i es el valor esperado de la variable respuesta para la observación i, β_0 corresponde al intercepto del modelo, x_{ij} representa el valor de la variable explicativa j para la observación i, y f_j son las funciones suavizadas que describen la relación entre cada predictor y la variable respuesta [25]. En este trabajo se empleó una distribución gaussiana con función de enlace identidad, de modo que $g(\mu_i) = \mu_i$.

La variable respuesta corresponde al número anual de días de cierre de los polígonos de cultivo de mejillón debido a la presencia de toxinas DSP. Como variables explicativas se consideran distintos indicadores ambientales, oceanográficos y biológicos potencialmente relacionados con la aparición de estos episodios de cierre.

En este trabajo se empleó una formulación específica implementada mediante el paquete `mgcv` [24, 25], en la que gran parte de los predictores se introducen mediante términos del tipo

$$s(\text{Pol_id},\ \text{by} = X)$$

donde X representa una variable explicativa.

Esta formulación permite modelar un efecto no lineal diferente para cada valor de la variable explicativa, utilizando la posición del polígono como eje sobre el que se construye la función suavizada. De este modo, el efecto espacial asociado a cada predictor puede variar entre polígonos, permitiendo capturar diferencias en la respuesta de cada zona de cultivo frente a una misma variable ambiental.

La utilización de modelos GAM resulta especialmente adecuada para este estudio debido a la complejidad de las relaciones existentes entre las condiciones ambientales y la aparición de episodios de cierre por toxinas DSP. Su flexibilidad permite representar relaciones no lineales entre la variable respuesta y los distintos predictores sin necesidad de imponer una forma funcional determinada. De este modo, estos modelos permiten integrar la información ambiental con la estructura espacial de los polígonos de cultivo y la dependencia temporal presente en la serie histórica, proporcionando así una representación más precisa del fenómeno estudiado[25].

4.2. Selección de variables

Debido al gran número de variables explicativas, se implementó un procedimiento automático de selección de variables basado en el Criterio de Información Bayesiano **(BIC)**. Esto se realizó a partir de un script proporcionado por el tutor del trabajo, Javier Roca, adaptado posteriormente a las necesidades específicas de este estudio.

El procedimiento sigue una estrategia de **selección progresiva con optimización iterativa**. En primer lugar, se evalúan todas las variables de forma individual y se selecciona aquella que produce el menor valor del BIC. A continuación, el modelo se amplía incorporando una nueva variable, escogida entre todas las candidatas disponibles, de forma que la combinación resultante minimice nuevamente dicho criterio.

Una vez fijado el número de variables del modelo, la selección no se considera definitiva. El algoritmo revisa sucesivamente cada una de las variables ya incluidas y estudia si puede ser sustituida por cualquier otra variable disponible que permita mejorar el ajuste. Este proceso de sustitución se repite para todas las posiciones del modelo, de manera que las variables seleccionadas en etapas anteriores no permanecen necesariamente en el modelo final, sino que pueden ser reemplazadas si existe otra combinación con un menor valor del **BIC**.

Durante el ajuste del modelo, las variables continuas con más de 20 valores únicos se incorporaron mediante funciones de suavizado de la forma $s(\texttt{Pol_id}, \texttt{by} = X_i)$, donde X_i representa la variable explicativa correspondiente. Por el contrario, las variables categóricas y las variables continuas con un número reducido de valores distintos se incorporaron como efectos lineales. Además, el identificador del polígono (**Pol_id**) se incluyó como efecto fijo en todos los modelos evaluados. (Esto se hace para ver si hay un efecto lineal conforme avanzamos por la línea de costa. En caso de no mejorar el BIC, se elimina este término)

Posteriormente, se incrementa en una unidad el número de variables del modelo y se repite nuevamente el procedimiento. En este trabajo se evaluaron modelos con entre dos y ocho variables explicativas, almacenándose para cada tamaño del modelo la combinación de variables seleccionada y el **valor mínimo del BIC obtenido**.

Finalmente, la evolución del BIC en función del número de variables permitió comparar la mejora obtenida al aumentar la complejidad del modelo. Así, decidimos quedarnos aquellas variables que **reduzcan el BIC en 2 unidades o más con respecto a la iteración anterior**.

Se muestra a continuación el código R del script empleado:

```r
BIC_Javi <- function(var0, datos, suavizar = TRUE){
  ff <- NULL
  for(i in var0){
    ff2 <- i
    if(suavizar &&
       !is.factor(datos[, i]) &&
       length(unique(datos[, i])) > 20){
      ff2 <- paste0("s(Pol_id, by=", i, ")")
    }
    ff <- c(ff, ff2)
  }
  ff <- c("Pol_id", ff)
  ff <- paste(ff, collapse = "+")
  ff <- as.formula(
    paste(names(datos)[1], "~", ff)
  )
  modelo <- gam(ff, data = datos)
  list(
    modelo = modelo,
    res = BIC(modelo)
    )
}

Buscaj <- function(var0, datos, j = 2, suavizar = TRUE){
  var <- names(datos)[-1]
  var.todas <- var
  var <- setdiff(var, var0[-j])
  nvar <- length(var)
  error <- numeric(nvar)
  dentro <- matrix(
    NA,
    nrow = nvar,
    ncol = length(var.todas)
    )

  for(k in 1:nvar){
    var0[j] <- var[k]
    error[k] <- BIC_Javi(
      var0,
      datos,
      suavizar
    )$res
    dentro[k, ] <- var.todas %in% var0
  }

  historico <- data.frame(
    p = apply(dentro, 1, sum),
    error,
    dentro  )
  var0[j] <- var[which.min(error)]
  list(
    historico = historico,
    var0 = var0,
    error0 = min(error)
    )
}
```

```
var0 <- NULL
historico <- NULL
A <- Buscaj(var0, datos, j = 1, suavizar = TRUE)
var0 <- A$var0
VAR <- list()
ERROR <- numeric()
ERROR[1] <- A$error0
VAR[[1]] <- var0

for(k in 2:8){

  var0 <- A$var0

  for(l in 1:k){

    A <- Buscaj(
      var0,
      datos,
      j = l,
      suavizar = TRUE
    )

    var0 <- A$var0
  }
  VAR[[k]] <- var0
  ERROR[k] <- A$error0
}
```

4.3. Modelado Explicativo

El objetivo de esta fase no consiste en obtener un único modelo con la mayor capacidad predictiva, sino en identificar y caracterizar la influencia de las distintas variables ambientales sobre el número anual de días de cierre por presencia de toxinas DSP. Para ello, se ajustó una **serie de modelos GAM**, y se analizó la contribución de los predictores seleccionados.

4.3.1. Modelo 1 - BD Completa

Las variables del primer modelo se han empleado seleccionando toda la base de datos. La principal limitación es que al emplear el ácido okadaico, sólo están disponibles los datos desde 2014. Aún así, sigue siendo el modelo que presenta un mejor rendimiento.

A continuación, se muestra la estructura del modelo (Que solo aparecerá en este caso para evitar redundancia, ya que todas son similares), y posteriormente la tabla incluyendo los efectos suavizados.

```
DSP ~ s(Pol_id, by = AO_Medio) +
  s(Pol_id, by = DSP_lag1) +
  s(Pol_id, by = AF_Dias_Pos_250_RRBB) +
  s(Pol_id, by = AF_Mediana_RRBB) +
  s(Pol_id, by = AF_Suma_Positivos_RRBB) +
  s(Pol_id, by = ASP) +
  s(Pol_id, by = AF_Ratio_750_RRBB) +
  s(Pol_id, by = AO_Max)
```

Variable	**edf**	**F**	**p-valor**	$\boldsymbol{\Delta R^2}$
AO_Medio	6.08	14.49	$<$ **0,001**	0.017
DSP_lag1	2.00	140.13	$<$ **0,001**	0.039
AF_Dias_Pos_250_RRBB	2.00	82.33	$<$ **0,001**	0.025
AF_Mediana_RRBB	7.29	35.23	$<$ **0,001**	0.048
AF_Suma_Positivos_RRBB	3.37	27.40	$<$ **0,001**	0.014
ASP	3.34	19.18	$<$ **0,001**	0.008
AF_Ratio_750_RRBB	9.00	14.56	$<$ **0,001**	0.008
AO_Max	3.43	7.98	$<$ **0,001**	0.006

Cuadro 4.1: Importancia de los efectos suavizados incluidos en el primer modelo GAM.

El modelo presentó un ajuste muy elevado, con un $\boldsymbol{R^2 = 0{,}917}$. Los grados de libertad efectivos (**edf**) revelaron que algunas variables presentaban relaciones marcadamente no lineales con la respuesta. En particular, el ácido okadaico medio (**edf=6.08**), la mediana del afloramiento (**edf=7.29**) y el ratio de afloramiento (**edf=9.00**) mostraron patrones complejos, mientras que otras variables, como **DSP_lag1** o **AF_Dias_Pos_250_RRBB**, presentaron relaciones considerablemente más simples.

Todos los términos suavizados incluidos en el modelo fueron altamente significativos ($\boldsymbol{p < 0{,}001}$), con valores altos del estadístico **F**. Entre ellos, destacó especialmente (**DSP_lag1**), con un valor de **140.13**.

La contribución relativa de cada predictor al modelo se evaluó mediante la reducción del $\boldsymbol{R^2}$ ajustado $\boldsymbol{(\Delta R^2)}$ al eliminar individualmente cada término del modelo completo. La mediana de la intensidad del afloramiento (**AF_Mediana_RRBB**), fue el predictor de mayor importancia, provocando una disminución de **0.048** unidades en el $\boldsymbol{R^2}$ ajustado tras su exclusión. También destacó **DSP_lag1** con $\boldsymbol{\Delta R^2 = 0{,}039)}$ y el número de días con afloramiento positivo (**AF_Dias_Pos_250_RRBB**, $\boldsymbol{\Delta R^2 = 0{,}025}$), lo que pone de manifiesto la elevada capacidad explicativa de las variables relacionadas con el afloramiento. El ácido okadaico medio (**AO_Medio**) mostró una contribución intermedia ($(\boldsymbol{\Delta R^2 = 0{,}017})$), mientras que la suma de valores positivos del afloramiento (**AF_Suma_Positivos_RRBB**), el índice(**ASP**), el ratio de afloramiento superior a 750 (**AF_Ratio_750_RRBB**) y el ácido okadaico máximo (**AO_Max**) presentaron aportaciones más modestas, con valores inferiores a **0.01**.

4.3.2. Modelo 2 - BD Completa, sin toxinas

En este segundo modelo se excluirán el ácido okadaico y el resto de toxinas para poder detectar mejor las variables predominantes. De este modo, también podremos trabajar con una serie temporal más amplia.

Variable	**edf**	**F**	**p-valor**	$\boldsymbol{\Delta R^2_{aj}}$
DSP_lag12	2.00	11.60	$<$ **0,001**	0.006
DSP_lag1	9.52	72.57	$<$ **0,001**	0.122
AF_Dias_Pos_250_RRAA	2.00	153.03	$<$ **0,001**	0.055
Sunspot_lag6	7.00	9.51	$<$ **0,001**	0.013
AF_Duracion_Media_Neg_RRBB	9.76	57.37	$<$ **0,001**	0.074
AF_Dias_Positivos_RRAA	2.00	108.42	$<$ **0,001**	0.039
AF_Suma_Negativos_RRAA	2.42	22.11	$<$ **0,001**	0.010
AF_Duracion_Media_Neg_RRAA	2.00	45.95	$<$ **0,001**	0.018

Cuadro 4.2: Importancia de los efectos suavizados incluidos en el modelo 2.

El modelo basado en retardos temporales y variables relacionadas con el afloramiento presentó un elevado ajuste, alcanzando un $\boldsymbol{R^2}$ ajustado de **0.809** y explicando el **81.5 %** de la desviación observada. Todos los efectos suavizados incluidos en el modelo resultaron altamente significativos ($\boldsymbol{p < 0{,}001}$).

El análisis de la reducción del $\boldsymbol{R^2}$ ajustado al eliminar cada predictor mostró que el primer retardo de la variable respuesta (**DSP_lag1**) constituye el factor de mayor importancia dentro del modelo, provocando una disminución de **0.122** unidades en el $\boldsymbol{R^2}$ ajustado. También destacaron la duración media de episodios negativos de afloramiento en la RRBB (**AF_Duracion_Media_Neg_RRBB**, $\boldsymbol{\Delta R^2 = 0{,}074}$) y el número de días de afloramiento positivo en la RRAA (**AF_Dias_Pos_250_RRAA**, $\boldsymbol{\Delta R^2 = 0{,}055}$), evidenciando de nuevo la fuerte influencia tanto de la persistencia temporal de los cierres como del afloramiento sobre la aparición de episodios de cierres por DSP.

Por el contrario, variables como el retardo de las manchas solares a 6 años (**Sunspot_lag6**) o la suma de episodios negativos de afloramiento en las rías altas(**AF_Suma_Negativos_RRAA**), produjeron reducciones muy pequeñas del $\boldsymbol{R^2}$ ajustado.

4.3.3. Modelo 3 - BD sin toxinas ni lags DSP

Ajustamos ahora un modelo similar al anterior, pero eliminando los lags del DSP, ya que el **lag1** era la variable que reducía más la variabilidad en el modelo anterior.

Variable	$\boldsymbol{\beta}$	**p-valor**
Intercepto	-155.45	$<$ **0,001**
Grupo 2	34.64	$<$ **0,001**
Grupo 3	84.96	$<$ **0,001**

Cuadro 4.3: Efectos lineales del grupo en el modelo 3

Variable	**edf**	**F**	**p-valor**	$\boldsymbol{\Delta R^2_{aj}}$
Sunspot_lag6	7.55	17.33	$<$ **0,001**	0.029
AF_Duracion_Media_Pos_RRBB	10.00	70.69	$<$ **0,001**	0.137
Sunspot_lag3	4.30	92.81	$<$ **0,001**	0.097
EA	8.54	32.46	$<$ **0,001**	0.061
Anom_Miño_Anual	3.90	54.65	$<$ **0,001**	0.051
NAO	2.00	70.24	$<$ **0,001**	0.029
wNAO_DJFM	5.93	18.86	$<$ **0,001**	0.026

Cuadro 4.4: Importancia de los efectos suavizados incluidos en el modelo 3.

El modelo basado en agrupaciones de polígonos, índices climáticos y variables de afloramiento alcanzó un $\boldsymbol{R^2}$ ajustado de **0.770**. A diferencia de los modelos anteriores, vemos que la variable **grupo** resultó altamente significativa ($\boldsymbol{p < 0{,}001}$), distinguiendo zonas de cierres bajos, medios y altos (Intercepto, Grupo 2 y Grupo 3 respectivamente).

Todos los efectos suavizados incluidos en el modelo fueron altamente significativos ($\boldsymbol{p < 0{,}001}$). No obstante, su contribución al ajuste fue diferente. La variable con mayor influencia fue la duración media de los episodios positivos de afloramiento en las rías baixas (**AF_Duracion_Media_Pos_RRBB**), cuya eliminación redujo el $\boldsymbol{R^2}$ ajustado en **0.137** unidades. También destacaron la actividad solar con un retardo de tres años (**Sunspot_lag3**, $\boldsymbol{\Delta R^2 = 0{,}097}$), el patrón atmosférico del Atlántico Este (**EA**, $\boldsymbol{\Delta R^2 = 0{,}061}$) y la anomalía anual del caudal del río Miño (**Anom_Miño_Anual**, $\boldsymbol{\Delta R^2 = 0{,}051}$). Por otro lado, las manchas solares con un retardo de seis años (**Sunspot_lag6**), el índice **NAO** y la **NAO** invernal (**wNAO_DJFM**) mostraron reducciones del $\boldsymbol{R^2}$ ajustado inferiores a **0.03**.

4.3.4. Modelo 4 - Manchas solares

En este apartado creamos un modelo basándonos exclusivamente en las manchas solares y en índices ambientales.

Variable	$\boldsymbol{\beta}$	**p-valor**
Intercepto	91.78	$<$ **0,001**
Pol_id	1.23	0.028

Cuadro 4.5: Efectos lineales incluidos en el modelo basado en manchas solares

Variable	**edf**	**F**	**p-valor**	$\boldsymbol{\Delta R^2_{aj}}$
Sunspot_lag6	7.25	13.76	$<$ **0,001**	0.032
Sunspot_SD	10.00	12.49	$<$ **0,001**	0.026
Sunspot_lag2	2.69	19.05	$<$ **0,001**	0.019
Sunspot_Mean	9.73	12.80	$<$ **0,001**	0.016
Sunspot_lag1	2.00	13.48	$<$ **0,001**	0.008
AMO	9.76	17.72	$<$ **0,001**	0.044

Cuadro 4.6: Importancia de los efectos suavizados incluidos en el modelo 4.

El modelo basado únicamente en variables relacionadas con la actividad solar y el índice oceánico AMO obtuvo un $\boldsymbol{R^2}$ ajustado de **0.624**, explicando el **63.8 %** de la desviación observada. El identificador del polígono resultó estadísticamente significativo ($\boldsymbol{p < 0{,}05}$), indicando una ligera tendencia creciente de los cierres de los polígonos al avanzar hacia el sur.

Todos los efectos suavizados incluidos en el modelo fueron altamente significativos ($\boldsymbol{p < 0{,}001}$). El índice oceánico **AMO** fue el predictor que mayor contribución realizó al ajuste del modelo, ya que su eliminación redujo el $\boldsymbol{R^2}$ ajustado en **0.044** unidades. Entre las variables relacionadas con la actividad solar destacó el retardo de seis años (**Sunspot_lag6**), con una reducción de **0.032** unidades, seguido de la desviación estándar anual de las manchas solares (**Sunspot_SD**, $\boldsymbol{\Delta R^2 = 0{,}026}$). Por el contrario, la media anual de manchas solares (**Sunspot_Mean**) y, especialmente, el retardo de un año (**Sunspot_lag1**) mostraron una contribución mucho más reducida al ajuste del modelo, con disminuciones del $\boldsymbol{R^2}$ ajustado de únicamente **0.016** y **0.008** unidades, respectivamente. Aún así, se ve que en conjunto el modelo esta determinado en gran medida por las manchas solares.

4.3.5. Modelo 5 - Variables ambientales

Creamos en este apartado un nuevo modelo centrándonos en una selección de variables ambientales. Los resultados se muestran a continuación. (Tabla 4.7)

Variable	**edf**	**F**	**p-valor**	ΔR^2
Temp_Diarias	9.96	192.08	**< 0,001**	0.522
Anom_WS_RRBB_Abrnov	3.15	56.54	**< 0,001**	0.057
Anom_Miño_Abrnov	2.00	39.29	**< 0,001**	0.021
Anom_Temp_Abrnov	2.55	37.09	**< 0,001**	0.030
Anom_UI_RRBB_Abrnov	2.00	16.36	**< 0,001**	0.008
Anom_WS_RRAA_Abrnov	2.00	70.82	**< 0,001**	0.039
Anom_WS_RRAA_Anual	6.07	2.58	0.0111	0.002

Cuadro 4.7: Importancia de los efectos suavizados incluidos en el modelo de variables ambientales.

En este caso se obtuvo un $\boldsymbol{R^2}$ de **0.618**. La temperatura superficial media (**Temp_Diarias**) constituye el predictor más relevante, ya que su eliminación reduce el $\boldsymbol{R^2}$ ajustado en más de **0.5**. Este resultado confirma que la temperatura es el principal factor ambiental asociado a la aparición de episodios de cierre por toxinas DSP.

Entre el resto de variables destacan las anomalías de la velocidad del viento durante el periodo **abril-noviembre**, tanto para RRBB como para RRAA, así como la anomalía de temperatura en dicho periodo. Todas ellas presentan valores elevados del estadístico $\boldsymbol{F}$ y reducciones apreciables del $\boldsymbol{R^2}$.

Las anomalías del caudal del río Miño también contribuyen de forma moderada al ajuste del modelo, mientras que la anomalía anual de la velocidad del viento en las rías altas apenas modifica el $\boldsymbol{R^2}$ ajustado, a pesar de resultar estadísticamente significativa.

4.3.6. Modelo 6 - Temperatura, Producción Primaria y Clorofila

A diferencia de la mayor parte de las variables analizadas, cuyos valores son comunes para todos los polígonos en un mismo instante temporal, la **temperatura**, la **clorofila** y la **producción primaria** presentan una variación espacial entre polígonos. Esta característica permite evaluar no solo su importancia estadística, sino también cómo cambia su influencia sobre la aparición de episodios de cierre por toxinas DSP a lo largo de la costa gallega.

Por este motivo, planteamos este último modelo incluyéndolas.

El modelo alcanzó un $\boldsymbol{R^2}$ ajustado de **0.59**, explicando el **59.5 %** de la variabilidad observada en el número anual de días de cierre por toxinas DSP.

Variable	**F**	**p-valor**	ΔR^2_{aj}
Temp_Diarias	72.10	$<$ **0,001**	0.020
Clorofila	46.78	$<$ **0,001**	0.017
Anom_Temp_Abrnov	31.61	$<$ **0,001**	0.018
Anom_Clorofila_Anual	16.00	$<$ **0,001**	0.000
Prod_Primaria	12.48	$<$ **0,001**	0.007
Anom_Clorofila_Abrnov	6.49	0.0016	0.003

Cuadro 4.8: Importancia de los efectos suavizados incluidos en el modelo 6.

Todas las variables ambientales locales incluidas en el modelo resultaron estadísticamente significativas ($p < 0{,}01$). La temperatura superficial del mar (**Temp_Diarias**) fue la variable de mayor relevancia, reduciendo el R^2 ajustado en **0.020** unidades. Fue seguida por la anomalía de temperatura durante el periodo abril-noviembre (**Anom_Temp_Abrnov**, $\Delta R^2 = 0{,}018$) y de la concentración de clorofila (**Clorofila**, $\Delta R^2 = 0{,}017$).

Aún así, podemos concluir que las tres variables ambientales son relevantes para los **cierres DSP**, ya que **Prod_Primaria**, aunque no tenga un ΔR^2 tan grande, sí que presenta una gran correlación con la **Clorofila**, por lo que ya recoge parte de su efecto.

4.3.7. Conclusiones sobre las variables ambientales

El análisis conjunto de los diferentes modelos GAM permitió identificar qué variables ambientales presentan una relación más consistente con el número anual de días de cierre por toxinas DSP.

- **Retardos de DSP**: Es uno de los grupos de variables con mayor capacidad explicativa. En particular, el retardo de un año (**DSP_lag1**).
- **Variables de afloramiento**: Son, junto con los retardos de DSP, los predictores más relevantes del modelo. Destacan especialmente la duración media de los episodios de afloramiento negativo y el número de días con afloramiento positivo. **(AF_Duracion_Media_Neg_RRBB y AF_Dias_Pos_250_RRAA)**
- **Temperatura superficial del mar**: La temperatura media anual y su anomalía durante el periodo abril-noviembre mostraron una elevada importancia, siendo las variables ambientales locales con mayor contribución al ajuste de los modelos.
- **Productividad biológica**: Tanto la concentración de clorofila como la producción primaria presentaron efectos significativos sobre la variable respuesta, aunque menores a la temperatura. No obstante, debido a la elevada correlación existente entre ambas, parte de la información aportada es redundante.
- **Caudal del río Miño**: Las anomalías del caudal mostraron una influencia moderada sobre los cierres, sugiriendo que los aportes fluviales pueden modificar las condiciones ambientales favorables para el desarrollo del fitoplancton tóxico pero en menor medida.
- **Índices climáticos**: Los índices de gran escala (**NAO**, **wNAO**, **EA** y **AMO**) resultaron estadísticamente significativos en distintos modelos. Entre ellos, el índice **AMO** fue el que mostró una mayor contribución al ajuste, mientras que el resto presentaron efectos más moderados.

- **Actividad solar**: Las variables relacionadas con las manchas solares fueron significativas, siendo capaces de definir un modelo prácticamente por si mismas. Sin embargo, su capacidad explicativa fue inferior a la observada para otras variables.

- **Grupo**: Esta variable permitió capturar diferencias estructurales entre zonas de cultivo en el **Modelo 3**, aunque de forma ligera ($\boldsymbol{\Delta R^2 = 0{,}026}$).

4.4. Predicción

En este apartado se desarrollan modelos predictivos con el objetivo de estimar el número anual de días de cierre por presencia de toxinas DSP en los polígonos de cultivo de mejillón de las Rías Baixas. Tras comparar sus $\boldsymbol{R^2}$ y su capacidad predictiva sobre datos no utilizados durante el entrenamiento, los modelos seleccionados se emplearán para obtener predicciones correspondientes al periodo **2026–2030**.

Dado que no poseemos información de las variables para los próximos años, se desarrolló un **modelo específico para cada horizonte comprendido entre uno y cinco años**. Cada uno de ellos fue construido utilizando las variables disponibles para el desfase temporal correspondiente, permitiendo evaluar cómo evoluciona la capacidad predictiva conforme aumenta el horizonte de predicción.

A continuación, se muestra la tabla comparativa de los modelos para cada horizonte. Para el conjunto de test se empleó el periodo **(2021-2025)**.

Horizonte	$\boldsymbol{R^2}$ **(test)**	**RMSE (test)**	$\boldsymbol{R^2}$ **(serie completa)**
1 año	0.2255	69.82	0.782
2 años	0.7086	39.81	0.796
3 años	0.6403	39.32	0.803
4 años	0.6590	40.07	0.835
5 años	0.6846	37.16	0.843

Cuadro 4.9: Rendimiento de los modelos desarrollados para cada horizonte temporal.

Los resultados obtenidos muestran que la capacidad predictiva de los modelos GAM varía en función del horizonte temporal considerado (Tabla 4.9). El modelo correspondiente al horizonte de un año presentó un rendimiento predictivo notablemente inferior ($\boldsymbol{R^2 = 0{,}23}$), acompañado del mayor error de predicción (**RMSE = 69.82**), a pesar de alcanzar un elevado ajuste sobre la serie completa ($\boldsymbol{R^2 = 0{,}78}$). Esta diferencia sugiere una limitada capacidad de generalización para este horizonte temporal, indicando que la variabilidad anual de los episodios de cierre está influenciada por procesos de elevada incertidumbre que dificultan su predicción.

Por el contrario, los modelos correspondientes a horizontes temporales de dos a cinco años mostraron un comportamiento considerablemente más estable, alcanzando valores de $\boldsymbol{R^2}$ sobre el conjunto de **test** comprendidos entre **0.64** y **0.71**, con errores de predicción inferiores a **40 días** en todos los casos. Aunque el ajuste sobre la serie completa aumentó ligeramente conforme se incrementó el horizonte temporal ($\boldsymbol{R^2 = 0{,}80\text{-}0{,}84}$), las diferencias entre estos modelos fueron relativamente reducidas, lo que indica una capacidad predictiva similar para horizontes de medio plazo.

A continuación se muestran las predicciones (en días de cierre):

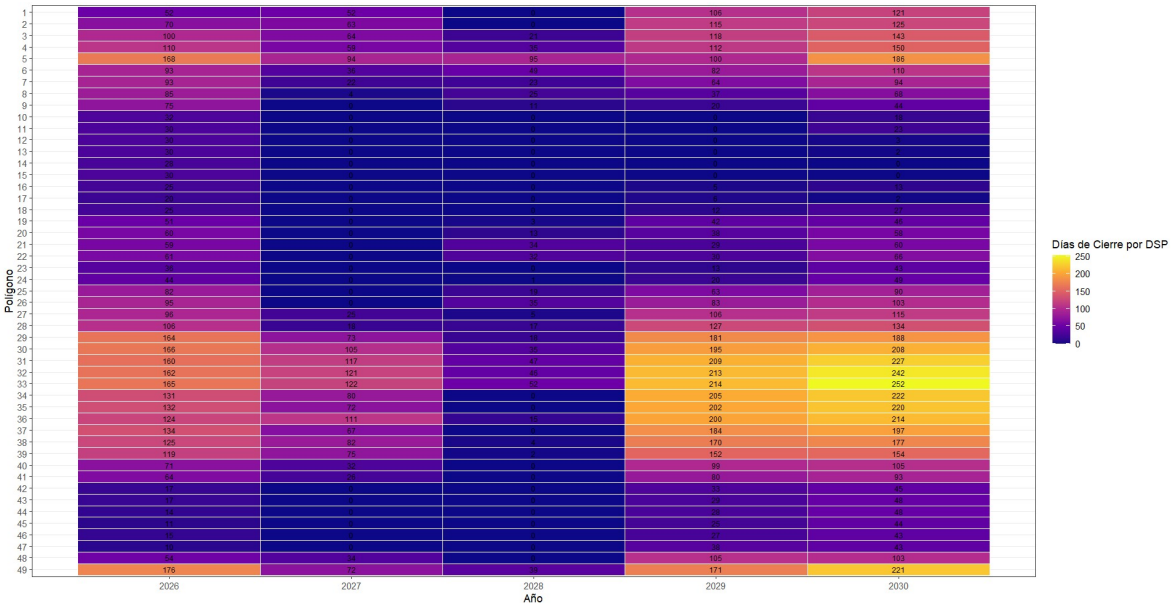


Figura 4.1: Predicciones de los cierres de los polígonos por DSP para los próximos 5 años

Se obtienen unos resultados bastante curiosos. Para empezar, se obtiene una predicción bastante razonable para el **horizonte 1**, aquel que mostraba un peor rendimiento en el conjunto de test. En el resto de casos, también se obtienen estimaciones razonables, exceptuando el **horizonte número 3**, donde son relativamente bajas. Para terminar, las últimas predicciones parecen indicar que se acerca un nuevo periodo de cierres altos, algo que ya parecía razonable al ver los últimos años del heatmap inicial.

A continuación se muestra el heatmap del periodo completo de cierres DSP tras añadir las predicciones, para visualizar mejor estos resultados.

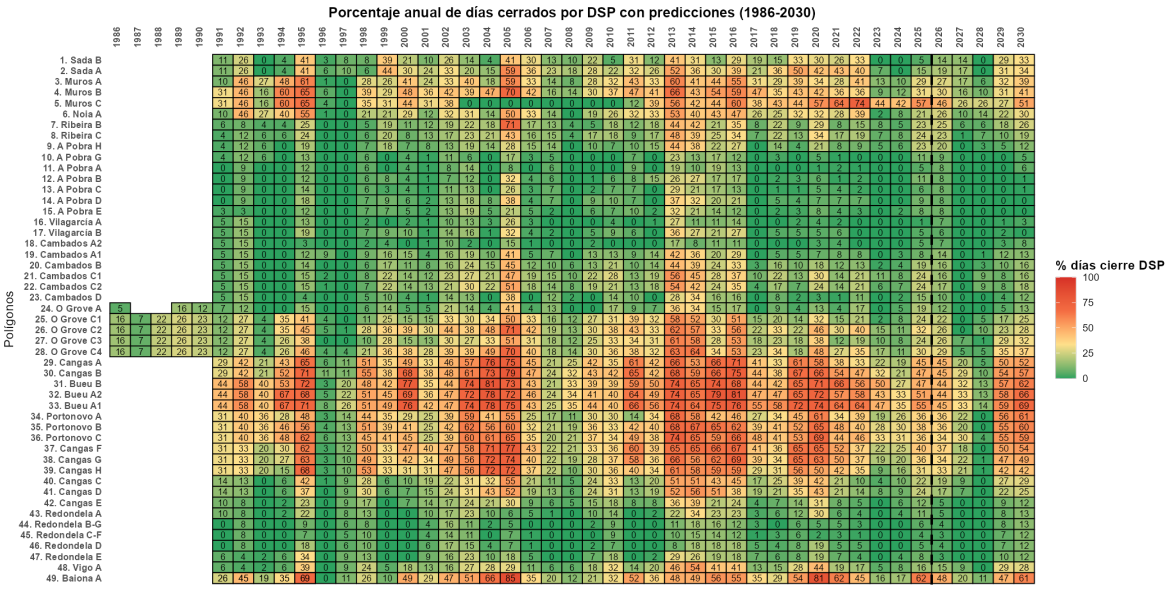


Figura 4.2: Heatmap de cierres DSP con predicciones

Sobre el heatmap las predicciones parecen razonables, indicando el inicio de un nuevo periodo de cierres altos durante la próxima década.

Capítulo 5

Conclusiones y trabajo futuro

El objetivo principal de este trabajo fue estudiar los factores asociados a los cierres por la presencia de toxinas DSP en los polígonos de cultivo de mejillón de Galicia, analizando la influencia de diferentes variables, junto a los componentes espacio-temporales de los datos mediante modelos aditivos generalizados. Además, se desarrollaron modelos predictivos para evaluar la evolución futura de los episodios de cierre.

5.0.1. Conclusiones

- **Capacidad explicativa de los modelos**: Los modelos aditivos generalizados (**GAM**) permitieron describir con precisión la variabilidad de los cierres por toxinas DSP, alcanzando valores de capacidad explicativa superiores al **90 %** cuando se incorporaron variables relacionadas con las toxinas y en torno al **80 %** utilizando únicamente variables ambientales.

- **Variables relacionadas con los cierres**: La persistencia temporal de los episodios de cierre y las variables relacionadas con el afloramiento constituyeron los principales factores explicativos de la variabilidad del DSP. Asimismo, las variables ambientales locales, los aportes fluviales, los índices climáticos y la actividad solar mostraron una contribución significativa, aunque de menor magnitud.

- **Capacidad predictiva**: Los modelos desarrollados mostraron capacidad para anticipar la evolución de los cierres por DSP con varios años de antelación, obteniendo mejores resultados para horizontes temporales de medio-largo plazo que para predicciones a corto plazo.

- **Aplicabilidad**: La metodología desarrollada constituye una herramienta útil para el estudio y la predicción de los episodios de cierres por DSP, pudiendo servir de apoyo a la gestión de las zonas de producción y a la planificación de las campañas de seguimiento.

- **Aplicabilidad**: La metodología desarrollada constituye una herramienta útil para anticipar periodos de elevada incidencia de cierres por toxinas DSP, proporcionando información que puede contribuir a la planificación temporal de la actividad en las zonas de producción.

5.0.2. Limitaciones

- **Disponibilidad y calidad de los datos**: Algunas de las variables empleadas presentaron una cobertura temporal limitada o valores ausentes, lo que redujo el número de observaciones disponibles para determinados modelos. En particular, la concentración de ácido okadaico, una de las variables con mayor capacidad explicativa, solo estuvo disponible a partir de 2014 y no para todos los polígonos de cultivo, restringiendo el periodo de estudio y la cobertura espacial de los modelos que la incorporaban.

- **Naturaleza de los datos**: Una parte importante de las variables ambientales y climáticas fue común a todos los polígonos para un mismo instante temporal, por lo que no reflejaron la variabilidad espacial específica de cada zona de producción, lo que pudo limitar la capacidad de los modelos para capturar diferencias locales entre polígonos.

- **Interpretación de los modelos**: Aunque los modelos aditivos generalizados permitieron capturar relaciones no lineales complejas entre las variables explicativas y los cierres por DSP, la interpretación de estos efectos resulta menos directa que en modelos lineales convencionales. En particular, la formulación empleada mediante funciones de suavizado dificulta cuantificar de forma sencilla el efecto individual de cada variable sobre la respuesta, por lo que los resultados deben interpretarse principalmente desde una perspectiva predictiva y descriptiva.

5.0.3. Líneas futuras

- **Mayor precisión temporal**: Desarrollar modelos basados en datos diarios o semanales permitiría representar con mayor precisión la dinámica de las floraciones tóxicas y de los episodios de cierre, capturando eventos y variaciones que no pueden identificarse utilizando agregaciones anuales.

- **Incorporación de nuevas variables**: Ampliar el conjunto de variables explicativas mediante la incorporación de información ambiental con mayor resolución espacial y temporal, así como de nuevos indicadores oceanográficos, meteorológicos y biológicos que puedan influir en la aparición de episodios de DSP.

- **Desarrollo de modelos espacio-temporales**: Explorar modelos que incorporen explícitamente la dependencia espacial entre polígonos, aprovechando su proximidad geográfica y las conexiones existentes entre las distintas rías.

- **Comparación con otros enfoques**: Evaluar el rendimiento de los modelos aditivos generalizados frente a otras técnicas de aprendizaje automático, como **Random Forest**, **XGBoost** o **redes neuronales**, para determinar si es posible mejorar la capacidad predictiva.

El presente trabajo ha permitido identificar los principales factores asociados a la aparición de episodios de cierre por toxinas DSP en las zonas de producción de mejillón de Galicia, demostrando el potencial de los modelos aditivos generalizados para describir y predecir este fenómeno. Los resultados obtenidos ponen de manifiesto que la integración de información ambiental, oceanográfica y climática constituye una estrategia prometedora para mejorar el conocimiento de la dinámica de las floraciones tóxicas y avanzar hacia el desarrollo de herramientas predictivas que contribuyan a una gestión más eficiente del sector mejillonero gallego.

Bibliografía

[1] Álvarez-Salgado XA, Figueiras FG, Pérez FF, Groom S, Nogueira E, Borges AV, Chou L, Castro CG, Moncoiffé G, Ríos AF, Miller AEJ, Frankignoulle M, Savidge G, Wollast R (2003) The Portugal Coastal Counter Current off NW Spain: New insights on its biogeochemical variability. Progress in Oceanography 56(2):281–321.

[2] Álvarez-Salgado XA, Labarta U, Fernández-Reiriz MJ, Figueiras FG, Rosón G, Piedracoba S, Filgueira R, Cabanas JM (2008) Renewal time and carrying capacity of the Rías Baixas for bivalve aquaculture. Journal of Marine Systems 72:278–292.

[3] Anderson DM, Cembella AD, Hallegraeff GM (2012) Progress in understanding harmful algal blooms: Paradigm shifts and new technologies for research, monitoring, and management. Annual Review of Marine Science 4:143–176.

[4] Anderson DM, Fensin E, Gobler CJ, Hoagland P, Kleindinst JL, et al. (2019) Marine harmful algal blooms (HABs) in the United States: History, current status and future trends. Harmful Algae 102107.

[5] APROMAR (2025) La acuicultura en España 2025. Asociación Empresarial de Acuicultura de España, Madrid.

[6] EFSA Panel on Contaminants in the Food Chain (CONTAM) (2023) Marine biotoxins in shellfish – Summary and scientific opinions of the EFSA CONTAM Panel. EFSA Journal.

[7] Food and Agriculture Organization of the United Nations (FAO) (2024) Fishery and Aquaculture Statistics. Global aquaculture production database. FAO, Rome.

[8] Figueiras FG, Labarta U, Fernández-Reiriz MJ (2002) Coastal upwelling, primary production and mussel growth in the Rías Baixas of Galicia. Hydrobiologia 484:121–131.

[9] Fraga F (1981) Upwelling off the Galician coast, northwest Spain. In: Richards FA (ed) Coastal Upwelling. Coastal and Estuarine Sciences, vol 1. American Geophysical Union, Washington, DC, pp 176–182.

[10] Gvozdenović Nikolić S, Mandić M, Drakulović D, Joksimović A (2015) The Shellfish Biotoxins. Agriculture & Forestry 61:135–142.

[11] Hallegraeff GM (1993) A review of harmful algal blooms and their apparent global increase. Phycologia 32(2):79–99.

[12] Hastie TJ, Tibshirani RJ (1990) Generalized Additive Models. Chapman & Hall, London.

[13] He X, Liu Y, Conklin A, Westrick J, Weavers LK, Dionysiou DD, Lenhart JJ, Mouser PJ, Szlag D, Walker HW (2021) Harmful algal blooms (HABs) and their environmental factors in fresh, marine and brackish waters: A review. Journal of Environmental Management 300:113718.

[14] Instituto Tecnolóxico para o Control do Medio Mariño de Galicia (INTECMAR) (2025) INTECMAR. https://www.intecmar.gal. Último acceso: julio de 2026.

[15] Labarta U, Fernández-Reiriz MJ (2019) The mussel industry in Galicia (NW Spain): Production, management and markets. Aquaculture International 27:337–354.

[16] Reguera B, Bravo I, Fraga S (1995) Bloom dynamics and physiology of Dinophysis species. In: Lassus P, Arzul G, Erard E, Gentien P, Marcaillou C (eds) Harmful Marine Algal Blooms. Lavoisier, Paris, pp 801–806.

[17] Reguera B, Velo-Suárez L, Raine R, Park MG (2012) Harmful Dinophysis species: A review. Harmful Algae 14:87–106.

[18] Reguera B, Riobó P, Rodríguez F, Díaz PA, Pizarro G, Paz B, Franco JM, Blanco J (2014) Dinophysis toxins: Causative organisms, distribution and fate in shellfish. Marine Drugs 12(1):394–461.

[19] Rosón G, Pérez FF, Álvarez-Salgado XA (1997) Water masses and water renewal in the Ría de Arousa (Galicia, NW Spain). Estuarine, Coastal and Shelf Science 45(2):213–223.

[20] Smayda TJ (2006) Harmful Algal Blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea. Limnology and Oceanography 51(1, part 2):728–745.

[21] Trainer VL, Kudela RM, Hunter MV, et al. (2010) The state of U.S. harmful algal blooms assessments, policy and legislation. Toxicon 55(5):1024–1034.

[22] Wells ML, Trainer VL, Smayda TJ, Karlson B, Trick CG, Kudela RM, Ishikawa A, Bernard S, Wulff A, Anderson DM, Cochlan WP (2020) Harmful algal blooms and climate change: Learning from the past and present to forecast the future. Harmful Algae 91:101591.

[23] Wood SN (2006) Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC, Boca Raton.

[24] Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B 73(1):3–36.

[25] Wood SN (2017) Generalized Additive Models: An Introduction with R, 2nd edn. Chapman & Hall/CRC, Boca Raton.

[26] Xunta de Galicia (2024) Anuario de Acuicultura de Galicia 2024. Consellería do Mar, Santiago de Compostela.