



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE GRADO

*Análisis de datos hidrológicos, meteorológicos y de calidad del
agua en la predicción de episodios de cianotoxinas*

Grado en Ingeniería Mecánica

ALUMNO: Santiago Montenegro Portela

DIRECTORES: Rosa Devesa Rey
Andrés Suárez García

CURSO ACADÉMICO: 2023-2024

Universida_{de}Vigo



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE GRADO

*Análisis de datos hidrológicos, meteorológicos y de calidad del
agua en la predicción de episodios de cianotoxinas*

Grado en Ingeniería Mecánica
Intensificación en Tecnología Naval
Cuerpo General

Universidade de Vigo

RESUMEN

Las cianobacterias destacan por su capacidad de fotosíntesis y simplicidad estructural. Sin embargo, en cuerpos de agua templados, como embalses y lagos, se pueden observar proliferaciones de cianobacterias que plantean graves preocupaciones ambientales y de salud pública. A pesar de investigaciones extensas, predecir la formación de estas floraciones sigue siendo un desafío debido a la interacción compleja de múltiples variables y causas aún por definir. Este trabajo se enfoca en la integración de tres métodos de análisis de datos: z-score, tercera derivada y descomposición por tendencia, estacionalidad y aleatoriedad, con el propósito de anticipar la aparición de estos episodios de cianotoxinas. Para llevar a cabo esta investigación, se han recolectado datos provenientes de las diversas fuentes pertinentes. Tanto el uso individual de estos métodos como la integración de estos tres métodos permite analizar los resultados obtenidos y proporciona una perspectiva más completa sobre la dinámica de los cuerpos de agua. Este estudio resalta el valor de combinar diferentes metodologías de análisis de datos para mejorar la capacidad de prever y afrontar los riesgos asociados con los episodios de cianotoxinas. Cobra importancia el validar los resultados obtenidos mediante estos métodos con las floraciones confirmadas por prensa.

PALABRAS CLAVE

Series temporales, Anomalía, Análisis tendencias.

AGRADECIMIENTOS

Quisiera expresar mi profundo agradecimiento a todas las personas que han contribuido de alguna manera a la realización de este Trabajo de Fin de Grado.

Merecen reconocimiento especial mi familia y amigos, quienes siempre estuvieron a mi lado brindándome su comprensión y ánimo en cada paso de este camino académico. Especialmente a D^a Martina Funes Lastra, sin su apoyo incondicional no habría sido posible llegar hasta aquí. También a D. Mikel Oyarzabal y D. Mikel Merino, que me han dado las fuerzas e ilusión necesarias para llegar hasta aquí.

De la misma manera agradecer a la promoción 424-154, a los compañeros con los que tanto tiempo he compartido. Gracias por hacer el día a día más llevadero.

¡Muchas gracias a todos!

CONTENIDO

Contenido	4
Índice de Figuras	6
Índice de Ecuaciones	7
Índice de Tablas.....	8
1. Introducción y objetivos.....	11
1.1 Introducción	11
1.2 Motivación	14
1.3 Objetivos	14
1.3.1 Recopilar y procesar datos ambientales relevantes para la investigación.....	14
1.3.2 Aplicar técnicas de análisis de datos para identificar relaciones entre variables.....	15
1.3.3 Evaluar la eficacia de los métodos analíticos corroborándolos con la prensa	15
1.4 Rasgo diferencial.....	15
1.5 Esquema del trabajo	15
2. Estado del arte	17
2.1 Series temporales	17
2.2 Series temporales hidrológicas.....	18
2.3 Anomalías en series temporales	19
2.4 Problema ambiental.....	20
2.5 El estudio de las cianotoxinas	22
2.6 Clasificación y tipos de cianotoxinas.....	23
2.7 Impacto en la salud humana y ambiental	25
2.8 Métodos de Detección y Monitoreo (Redes de Control de Calidad de las Aguas).....	26
3. Desarrollo del TFG	29
3.1 Obtención de series temporales.....	29
3.2 Preprocesado de datos.....	29
3.3 Detección de anomalías Z-scores.....	30
3.4 Detección de anomalías método de la tercera derivada	32
3.5 Detección de anomalías por descomposición en TEA.....	33
3.6 Aprendizaje por conjuntos	35
3.7 Validación mediante noticias de prensa.....	36
4. Resultados	39

4.1	Fechas prensa	39
4.2	Influencia meteorológica.....	40
4.3	Análisis Z-score	40
4.4	Análisis método de la tercera derivada	43
4.5	Análisis descomposición en TEA	45
4.6	Análisis resultados por métodos	46
4.6.1	Resultados por métodos para el río Limia	47
4.6.2	Resultados por métodos para el río Miño	47
4.7	Resultados aprendizaje por conjuntos.....	49
4.7.1	Resultados por conjuntos para el río Limia	49
4.7.2	Resultados por conjuntos para el río Miño	50
4.7.3	Resultados por conjuntos para el río Forcadas.....	51
4.7.4	Resultados por conjuntos para el río Tea	53
4.8	Validación de resultados	53
5	Conclusiones y líneas futuras.....	55
5.1	Conclusiones	55
5.2	Líneas futuras	56
6	Bibliografía	57
Anexo I: Implicaciones Sociales, Económicas y Ambientales		61
6.1	Sociales	61
6.2	Económicas	61
6.3	Ambientales	61
Anexo II: Reflexiones Éticas y Sociales		63

ÍNDICE DE FIGURAS

Figura 1-1. Agua contaminada con cianobacterias [2]	11
Figura 1-2. Distintos tipos de cianotoxinas [3]	12
Figura 1-3. Investigaciones científicas en relación con las cianobacterias [5].....	13
Figura 1-4. Embalse de As Forcadas [6]	13
Figura 1-5. Esquema ciclo de tratamiento de datos [7]	16
Figura 2-1. Ejemplo de una serie temporal en finanzas [9].....	17
Figura 2-2. Cálculo Hidrometeorológico de Aportaciones y Crecidas del Centro de Estudios Hidrográficos del CEDEX	18
Figura 2-3. Ejemplo de una anomalía en una serie temporal	19
Figura 2-4. Control de cianobacterias en el lago San Roque, Argentina [13].....	20
Figura 2-5. Agua afectada por la eutrofización [14]	21
Figura 2-6. Río Limia durante una floración de cianotoxinas [15]	21
Figura 2-7. Uso de análisis de datos hidrológicos para el caudal [16]	22
Figura 2-8. Toma de muestras de agua contaminada por cianotoxinas [17]	23
Figura 2-9. Microcistinas [19].....	24
Figura 2-10. <i>Gymnodinium catenatum</i> , productora de saxitoxinas [20].....	24
Figura 2-11. Cilindrospermopsina [21]	25
Figura 2-12. Perro consumiendo agua de un cuerpo de agua infectado de cianobacterias [23].....	26
Figura 2-13. Fauna acuática afectada por una floración de cianotoxinas [5]	26
Figura 2-14. División de la gestión de las cuencas hidrográficas en Galicia	27
Figura 3-1. Representación de la utilidad del preprocesado de datos	30
Figura 3-2. Distribución normal [29]	31
Figura 3-3. Estacionalidad en una serie temporal [35].....	34
Figura 3-4. Noticia de “La Voz de Galicia” acerca de las cianobacterias [36]	36
Figura 4-1. Situación geográfica del embalse de As Conchas y de la zona estudiada	49
Figura 4-2. Distancia entre la floración y el cuerpo de agua estudiado en el río Miño.....	50
Figura 4-3. Situación geográfica embalse de As Forcadas.....	52

ÍNDICE DE ECUACIONES

Ecuación Z-score (3-1).....	31
Ecuación de la 3ª derivada (3-2).....	32
Ecuación cálculo del CMA (3-3).....	34
Ecuación TPV (3-4).....	37

ÍNDICE DE TABLAS

Tabla 3-1. Conjunto de parámetros obtenidos de cada río	29
Tabla 3-2. Extracto de resultados 3ª derivada semanas 4 a 9 de la conductividad para el río Miño	33
Tabla 3-3. Extracto de amonio en el río Limia, con el valor anómalo en negrita	33
Tabla 3-4. Ejemplo de los datos que se obtienen por el método de estacionalidad para un cuarto de año	34
Tabla 3-5. Resultados finales para el método de la estacionalidad	35
Tabla 3-6. Explicación de la validación cruzada	36
Tabla 3-7. Matriz de confusión para falsos positivos/negativos	37
Tabla 4-1. Fechas de floraciones obtenidas por prensa de cada uno de los ríos estudiados	39
Tabla 4-2. Datos meteorológicos para las fechas de floraciones.	40
Tabla 4-3. Extracto de valores de Z de las primeras 10 semanas recogidas para el embalse del río Limia	41
Tabla 4-4. Extracto de los resultados en el que está subrayado el valor anómalo	41
Tabla 4-5. Resultados anómalos obtenidos por Z-score para el río Limia	41
Tabla 4-6. Resultados anómalos obtenidos por Z-score para el río Miño	42
Tabla 4-7. Resultados anómalos obtenidos por Z-score para el río Tea	42
Tabla 4-8. Resultados anómalos obtenidos por Z-score para el río Forcadas	43
Tabla 4-9. Extracto de valores de la 1ª, 2ª y 3ª derivada de las primeras 10 semanas muestreadas del amonio en el río Limia	43
Tabla 4-10. Valores anómalos obtenidos por el método de la 3ª derivada para el río Limia	44
Tabla 4-11. Resultados anómalos obtenidos por la 3ª derivada para el río Miño	44
Tabla 4-12. Resultados anómalos obtenidos por la 3ª derivada para el río Tea	44
Tabla 4-13. Resultados anómalos obtenidos por la 3ª derivada para el río Forcadas	45
Tabla 4-14. Extracto de valores divididos por cuartos para el río Tea	45
Tabla 4-15. Extracto de valores de CMA correspondientes a los factores del río Tea	46
Tabla 4-16. Extracto de la relación de cada valor respecto a su correspondiente CMA	46
Tabla 4-17. Resultados finales para el método de estacionalidad	46
Tabla 4-18. Comparación resultados Z-score con fechas documentadas para el río Limia	47
Tabla 4-19. Comparación resultados 3ª derivada con fechas documentadas para el río Limia	47
Tabla 4-20. Resultados por tendencia y estacionalidad del río Limia	47
Tabla 4-21. Comparación resultados Z-score con fechas documentadas para el río Miño	48
Tabla 4-22. Comparación resultados 3ª derivada con fechas documentadas para el río Miño	48
Tabla 4-23. Resultados por tendencia y estacionalidad del río Miño	49
Tabla 4-24. Resultados río Limia. Resaltados los elementos coincidentes	50

Tabla 4-25. Resultados por tendencia y estacionalidad del río Limia	50
Tabla 4-26. Resumen verificación cruzada resultados río Limia	50
Tabla 4-27. Resultados del río Miño. En negrita los elementos coincidentes	51
Tabla 4-28. Resultados por tendencia y estacionalidad del río Miño	51
Tabla 4-29. Resumen verificación cruzada resultados río Miño	51
Tabla 4-30. Tabla de resultados del río Forcadas	52
Tabla 4-31. Resultados por tendencia y estacionalidad del río Forcadas	52
Tabla 4-32. Resumen verificación cruzada resultados río Forcadas	52
Tabla 4-33. Resultados del río Tea	53
Tabla 4-34. Resultados por tendencia y estacionalidad del río Tea	53
Tabla 4-35. Comparación resultados río Forcadas	54
Tabla 4-36. Comparación resultados río Miño	54
Tabla 4-37. Comparación resultados río Limia	54

1. INTRODUCCIÓN Y OBJETIVOS

1.1 Introducción

En diversas regiones del mundo, la proliferación de cianobacterias y la consecuente presencia de cianotoxinas en embalses y fuentes de agua han emergido como un desafío ambiental de magnitudes preocupantes. Este fenómeno, lejos de constituir un hecho aislado, se manifiesta de manera periódica, generando una serie de consecuencias que van más allá de los límites naturales del ecosistema acuático.

Las cianobacterias son microorganismos primitivos que, a pesar de su simplicidad estructural, desempeñan un papel vital en el equilibrio ecológico de los ambientes acuáticos. En la Figura 1-1 se puede observar la apariencia que le pueden dar al cuerpo de agua y de ahí viene que se les llame también “algas verde azules”. Su capacidad para realizar la fotosíntesis las sitúa en la base de la cadena alimenticia acuática, siendo fuente de alimento para diversos organismos acuáticos. No obstante, la aparente carencia de ofensividad de estas cianobacterias se ve eclipsada en ocasiones por un fenómeno potencialmente perjudicial: la producción de cianotoxinas. Estas sustancias tóxicas, generadas durante ciertas condiciones propicias, pueden convertir las pacíficas aguas en fuentes de riesgo para la salud humana y la biodiversidad acuática. [1]



Figura 1-1. Agua contaminada con cianobacterias [2]

Las cianotoxinas son compuestos químicos que pueden tener efectos perjudiciales en diferentes organismos. Su presencia en el agua puede ser el resultado de una proliferación masiva de cianobacterias, conocidas como floraciones algales, que tiñen el agua con tonos verdosos. No todas las clases de cianotoxinas son iguales, las hay que son cianotoxinas hepatotóxicas que pueden afectar el hígado, mientras que otras actúan como neurotoxinas, amenazando el sistema nervioso. Las saxitoxinas, pueden causar parálisis respiratoria. Estas cianotoxinas pueden impactar adversamente a los organismos acuáticos, desencadenando consecuencias ecológicas de largo alcance.

La clave para entender este fenómeno radica en reconocer la compleja interacción entre las cianobacterias, las condiciones ambientales y la producción de cianotoxinas. Por ello es importante también saber que hay distintos tipos de cianotoxinas, cómo se muestra en la Figura 1-2, cada una con sus efectos y con su composición y aspecto. Factores como el aumento de la temperatura del agua, la presencia excesiva de nutrientes y patrones climáticos específicos pueden desencadenar estas floraciones algales y, por ende, la potencial liberación de cianotoxinas. La gestión efectiva de este problema requiere un enfoque que abarque desde la monitorización de cuerpos de agua hasta medidas preventivas y correctivas. La detección temprana de floraciones algales y la evaluación de la presencia de cianotoxinas son cruciales para tomar decisiones informadas y proteger la calidad del agua.

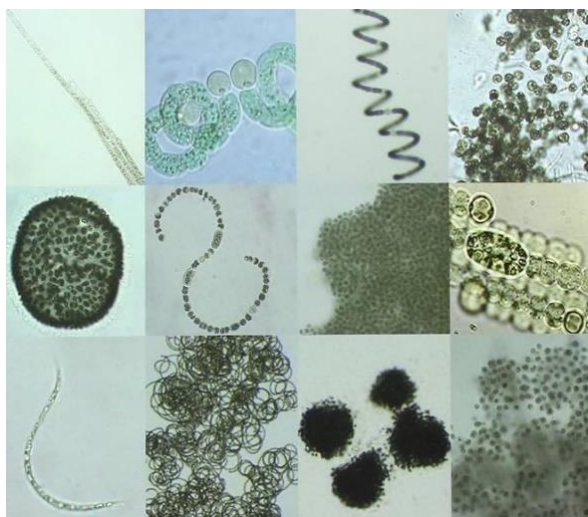


Figura 1-2. Distintos tipos de cianotoxinas [3]

El verdadero desafío reside en las cianotoxinas. En respuesta a esta problemática, las autoridades locales se ven obligadas, en ocasiones periódicas, a tomar medidas drásticas, como la interrupción de la red de saneamiento en zonas afectadas. La necesidad de cortar la red de saneamiento se torna imperativa cuando las concentraciones de cianotoxinas alcanzan niveles críticos, sobrepasando los límites aceptados para el consumo humano y poniendo en riesgo la integridad de la fauna y flora acuáticas. Esta interrupción temporal acarrea evidentes inconvenientes logísticos y económicos, afectando a comunidades locales que dependen del suministro constante de agua potable.

La complejidad de este problema radica en su naturaleza recurrente, ya que las condiciones ambientales propicias para la proliferación de cianobacterias pueden repetirse en ciclos estacionales o debido a cambios climáticos a largo plazo. La gestión efectiva de este fenómeno requiere no solo respuestas reactivas para abordar crisis puntuales, sino también estrategias a largo plazo para mitigar las condiciones que propician la proliferación de cianobacterias. [4]

En este contexto, la investigación científica y la implementación de tecnologías innovadoras son fundamentales para comprender y abordar este desafío ambiental de manera integral. La conciencia pública sobre la importancia de la conservación del agua y la prevención de la contaminación por cianotoxinas también desempeña un papel crucial en la búsqueda de soluciones sostenibles y a largo plazo. Por todo esto las cianobacterias, a pesar de su papel esencial en los ecosistemas acuáticos, plantean desafíos significativos.

En la Figura 1-3 se observa cómo se toman muestras de agua infectada por este fenómeno para su posterior análisis e investigación en un laboratorio, de este estudio se obtienen distintos datos. En base a esto, se busca mediante el análisis de estos datos ser capaz de observar una relación de parámetros que son los que provocan estas condiciones propicias que se den estas apariciones cianobacterias, y de esta manera ser capaces de predecir con antelación y efectividad cuando se van a dar para salvaguardar los cuerpos de agua y garantizar la seguridad de quienes dependen de ellos y de la misma manera para poder

actuar previamente para evitar males mayores para estas comunidades afectadas que una y otra vez ven su red de saneamiento interrumpida debido a este fenómeno.



Figura 1-3. Investigaciones científicas en relación con las cianobacterias [5]

Los datos han sido proporcionados por distintos organismos entre los que están el Ministerio de Transición Ecológica y el Reto Demográfico y la Xunta de Galicia mediante “Augas de Galicia”. En Galicia la responsabilidad de la recopilación de datos de los distintos embalses se encuentra en manos de dos administraciones distintas y cada una de estas administraciones tiene distintas cuencas hidrográficas a cargo de ellas. Esto significa que los documentos proporcionados carecen de cierta uniformidad a la hora de presentar dichos datos, que puede tener un impacto en los análisis que se hacen. Esto añade un punto de dificultad a la hora de la comparación y monitorización de los distintos embalses.

Estos organismos han proporcionado datos variados pudiendo así analizar distintos aspectos de diferentes muestras de agua alrededor de Galicia. De los ríos de los cuales se han recibido datos han sido el río Tea, el río Miño, el río Limia y del embalse de As Forcadas, como se observa en la Figura 1-4. Los datos proporcionados han sido a lo largo de los últimos 5 años completos (2018-2023, ambos inclusive) teniendo datos de todos los días de estos años, y entre estos datos se incluyen: nivel del río (m), caudal del río (m^3/s), precipitación acumulada (mm), temperatura ambiente ($^{\circ}\text{C}$), temperatura del agua ($^{\circ}\text{C}$), pH del agua (u.pH), conductividad del agua ($\mu\text{S}/\text{cm}$), cantidad de amonio por litro de agua (mg/L), oxígeno por litro de agua (mg/L), turbidez del agua (NTU), materia orgánica en el agua (%), cantidad de fosfatos por litro de agua (mg/L) y cantidad de nitratos por litro de agua (mg/L).



Figura 1-4. Embalse de As Forcadas [6]

1.2 Motivación

La motivación para emprender este trabajo de investigación surge de la creciente preocupación por los episodios de cianotoxinas en cuerpos de agua dulce, los cuales representan una amenaza significativa para la salud pública y el ecosistema acuático.

Los episodios de proliferación de cianobacterias y la subsiguiente liberación de cianotoxinas son fenómenos que han ido en aumento en todo el mundo debido a factores como el cambio climático, la eutrofización de cuerpos de agua y la contaminación por nutrientes. Estas toxinas pueden tener efectos adversos graves que van desde irritaciones cutáneas hasta enfermedades hepáticas y neurológicas, e incluso la muerte en casos extremos. Además, la presencia de cianotoxinas en fuentes de agua potable puede suponer un desafío para los sistemas de tratamiento convencionales, lo que plantea serias preocupaciones para la seguridad hídrica.

En este contexto, el análisis de datos emerge como una herramienta fundamental para comprender y predecir la ocurrencia de episodios de cianotoxinas. La complejidad de los factores ambientales y biológicos involucrados en la proliferación de cianobacterias demanda enfoques interdisciplinarios que integren datos de diversas fuentes, como información meteorológica, características del cuerpo de agua, niveles de nutrientes y la dinámica de las comunidades de microorganismos acuáticos.

La aplicación de técnicas de análisis de datos permite identificar patrones y relaciones entre variables que pueden ser utilizados para desarrollar modelos predictivos precisos. Estos modelos tienen el potencial de alertar a las autoridades y gestores de recursos hídricos sobre la probabilidad de aparición de episodios de cianotoxinas, lo que facilitaría la implementación de medidas preventivas y de mitigación.

Es decir, la motivación detrás de este trabajo radica en la necesidad urgente de desarrollar herramientas predictivas para mitigar los riesgos asociados con la proliferación de cianotoxinas. Es necesaria una combinación de conocimientos en ciencias ambientales y habilidades en análisis de datos para abordar este desafío y generar resultados significativos que puedan contribuir a la protección de la salud humana y la conservación del medio ambiente. Sería positivo que los resultados de este trabajo puedan servir como base para futuras investigaciones y acciones de gestión orientadas a prevenir y controlar los episodios de cianotoxinas, promoviendo así el bienestar de las comunidades y la sostenibilidad de los ecosistemas acuáticos.

1.3 Objetivos

Los objetivos de este trabajo de fin de grado están diseñados para abordar el desafío de predecir y mitigar los episodios de cianotoxinas en cuerpos de agua dulce, mediante el uso de técnicas de análisis de datos. Estos objetivos se fundamentan en la necesidad de desarrollar herramientas predictivas precisas que permitan una gestión óptima de los recursos hídricos, con el fin de proteger la sociedad y preservar la calidad del medio ambiente acuático. En base a esto se establecen los siguientes objetivos para este proyecto:

1.3.1 Recopilar y procesar datos ambientales relevantes para la investigación

El primer objetivo consiste en recopilar y procesar datos de diversas fuentes, incluidas estaciones meteorológicas, sensores remotos, muestreos de campo y bases de datos públicas, que contengan información relevante sobre los factores ambientales y biológicos asociados con los episodios de cianotoxinas. Esto puede implicar la integración de datos de múltiples plataformas y la realización de análisis espaciales y temporales para obtener conjuntos de datos completos y coherentes que puedan ser utilizados para el desarrollo de modelos predictivos.

1.3.2 Aplicar técnicas de análisis de datos para identificar relaciones entre variables

El segundo objetivo se centra en la aplicación de técnicas de análisis de datos, la modelización estadística y la minería de datos, para identificar patrones y relaciones entre variables en los conjuntos de datos recopilados. Estos modelos se basarán en variables ambientales y biológicas seleccionadas mediante análisis de importancia y validados mediante técnicas de validación cruzada y evaluación de su desempeño.

1.3.3 Evaluar la eficacia de los métodos analíticos corroborándolos con la prensa

El tercer y último objetivo consiste en evaluar la eficacia y la aplicabilidad de los métodos de análisis desarrollados en la práctica, mediante la comparación de las predicciones del modelo con datos observados de episodios de cianotoxinas en cuerpos de agua dulce en la prensa. Esto es esencial, la prensa es la fuente al alcance que proporciona fechas para los episodios de cianotoxinas en los embalses.

En resumen, los objetivos de este trabajo están orientados a desarrollar herramientas predictivas que puedan permitir una gestión proactiva y efectiva de los episodios de cianotoxinas en cuerpos de agua dulce.

1.4 Rasgo diferencial

Una de las características de este trabajo de fin de grado es la integración de disciplinas matemáticas de análisis de datos y de abordar el desafío de predecir episodios de cianotoxinas en cuerpos de agua dulce para la gestión sostenible de los recursos hídricos. Este estudio busca ofrecer una perspectiva holística y avanzada para comprender y mitigar los riesgos asociados con la proliferación de cianobacterias.

Un rasgo diferencial sería la integración de distintos métodos de análisis de datos. La parte más importante de este trabajo es el uso de distintos métodos para analizar los datos en busca de resultados anómalos que puedan ser significativos en la aparición de cianobacterias y que puedan tener un impacto directo en estas floraciones de tal manera que sabiendo que cuando se dan resultados como esos anómalos, es posible que vaya a darse un episodio de cianobacterias en ese cuerpo de agua. Esta validación cruzada entre distintos métodos aporta una fiabilidad de mayor grado a los resultados obtenidos ya que se descartan resultados que no concuerden entre los métodos.

Otro rasgo diferencial de este trabajo es su contribución a la gestión sostenible de los recursos hídricos. Al proporcionar herramientas en las que se trate de ver una manera de predecir estos episodios que servirían para abordar los episodios de cianotoxinas, este estudio busca promover prácticas de gestión más sostenibles que protejan la salud pública y preserven la calidad del medio ambiente acuático a largo plazo.

1.5 Esquema del trabajo

En este estudio, se ha seguido un esquema detallado. Este enfoque se divide en cuatro etapas principales: recopilación de datos, tratamiento de los datos, análisis de los datos y conclusiones.

Recopilación de Datos: La primera fase del trabajo implicó recopilar una amplia gama de datos relacionados con la presencia de cianotoxinas en los embalses y otros cuerpos de agua en las áreas de estudio. Esto incluyó datos históricos de monitoreo, registros meteorológicos, información sobre la calidad del agua y datos biológicos sobre la dinámica de las comunidades de cianobacterias. Para ello, se ha hecho uso de fuentes de datos gubernamentales, estudios científicos previos y bases de datos especializadas.

Tratamiento de los Datos: una vez recopilados los datos, procedimos a su tratamiento. Esto incluyó la limpieza de datos, así como la estandarización de formatos y unidades de medida. Ordenándolos todos de la misma manera de tal forma que se pudiese trabajar con todos por con los mismos métodos.

Análisis de los Datos: con los datos tratados, se pasa a la fase de análisis, que consiste en la aplicación una variedad de técnicas estadísticas para identificar patrones y relaciones entre variables. Utilizamos métodos de análisis descriptivo para caracterizar la distribución y variabilidad de las concentraciones de cianotoxinas en los embalses, así como técnicas de modelización predictiva para predecir la ocurrencia de episodios de cianotoxinas en función de variables ambientales y biológicas.

Validación: un apartado clave fue la validación de los datos puestos frente a los episodios de cianotoxinas ya publicados en distintas notas de prensa de tal forma que se confirma el nivel de veracidad y el nivel de fiabilidad de relación directa que aporta el resultado anómalo de cada uno de los componentes respecto a las floraciones de cianobacterias. Es decir, si mediante el análisis de datos se obtiene una fecha en la que se da un resultado anómalo, se valida un contraste con noticias de prensa para ver si se menciona algo al respecto o si ha habido una floración de cianotoxinas en ese periodo de tiempo, así validando una posible relación entre el resultado anómalo y la fecha documentada por la prensa.

Conclusiones: finalmente, poniendo la vista en los resultados del análisis de datos, se pueden extraer conclusiones sobre la dinámica de las cianotoxinas en los cuerpos de agua de estudio. Se han identificado factores clave que influyen en la proliferación de cianobacterias y la producción de cianotoxinas, así como patrones temporales y espaciales en la ocurrencia de episodios de contaminación. Estas conclusiones permiten proponer recomendaciones para la gestión y mitigación de los riesgos asociados con las cianotoxinas, así como identificar áreas de investigación futura para seguir avanzando en este campo.

Todas estas fases comentadas se ven representadas visualmente en la Figura 1-5, desde recopilación, tratamiento y análisis hasta la validación de los datos y las conclusiones a extraer del estudio.



Figura 1-5. Esquema ciclo de tratamiento de datos [7]

2. ESTADO DEL ARTE

2.1 Series temporales

Las series temporales son un componente esencial del análisis de datos en el ámbito de las matemáticas. Consisten en secuencias ordenadas de observaciones que se recopilan a lo largo del tiempo. Estas series se presentan en una amplia gama de contextos y disciplinas, desde la economía y las finanzas hasta la meteorología, la ingeniería y la medicina entre otras. Su importancia radica en la capacidad de proporcionar una comprensión de cómo evolucionan los fenómenos en el tiempo y cómo pueden ser modelados y utilizados para hacer predicciones futuras. [8]

Una de las aplicaciones principales de las series temporales en matemáticas es el análisis de tendencias y patrones. Al observar la evolución de una variable a lo largo del tiempo, se puede identificar si existe una tendencia ascendente, descendente o estacionaria en los datos. Esta información es crucial en diversos campos, en economía y finanzas, por ejemplo, donde se pueden estudiar los precios de los activos financieros, hasta la climatología, donde se pueden analizar las temperaturas medias a lo largo de los años. Un claro ejemplo de esto es la Figura 2-1, en la que se muestra la recopilación de datos del precio de un activo financiero respecto al tiempo, se observa la evolución de la variable en el periodo estudiado y a raíz de esta evolución se extraen conclusiones.



Figura 2-1. Ejemplo de una serie temporal en finanzas [9]

Además de analizar tendencias, las series temporales también se utilizan para prever valores futuros. Mediante el uso de técnicas como modelos autorregresivos, suavizado exponencial y redes neuronales, entre otros, se pueden hacer pronósticos sobre cómo evolucionarán los datos en el futuro. Estas

predicciones son esenciales para la planificación estratégica y la toma de decisiones en una amplia variedad de campos.

Otra aplicación importante de las series temporales en matemáticas es el modelado y la simulación de sistemas dinámicos. Al representar las relaciones entre diferentes variables en forma de series temporales, se puede estudiar cómo interactúan estas variables y cómo cambian con el tiempo. Además, el modelado de series temporales puede utilizarse para optimizar el rendimiento de sistemas en tiempo real, como la gestión de redes eléctricas o la planificación de rutas de transporte.

Las series temporales desempeñan un papel fundamental en el análisis de datos en el mundo de las matemáticas. Permiten a los investigadores y profesionales estudiar cómo evolucionan los fenómenos a lo largo del tiempo, identificar tendencias y patrones, y hacer predicciones sobre el futuro. Con el avance de la tecnología y el aumento en la disponibilidad de datos temporales, las series temporales se han convertido en una herramienta indispensable para comprender y modelar una amplia variedad de procesos y fenómenos en el mundo moderno. Su aplicación práctica es vasta y su importancia continúa creciendo a medida que se avanza hacia la era de la información y la ciencia de datos.

2.2 Series temporales hidrológicas

Las series temporales hidrológicas son simplemente registros de datos relacionados con el agua a lo largo del tiempo, como el caudal de los ríos, la precipitación y los niveles de agua en lagos y embalses. Estos datos son esenciales para comprender cómo funciona el ciclo del agua en diferentes regiones y cómo está cambiando con el tiempo. El propio Centro de Estudios Hidrográficos del CEDEX (perteneciente al Ministerio para la Transición Ecológica y el Reto Demográfico) basa su trabajo en cálculos sobre este tipo de series.

Uno de los aspectos más importantes de las series temporales hidrológicas es su capacidad para proporcionar información sobre los patrones y tendencias en el comportamiento del agua. Para poder crear este tipo de series temporales se depende de varios organismos, principalmente públicos, dependientes de organismos gubernamentales como el que se muestra en la Figura 2-2. Por ejemplo, al analizar las series temporales de caudal de un río, se puede identificar si hay cambios a largo plazo en el flujo de agua debido al cambio climático, la deforestación u otras influencias humanas.



Figura 2-2. Cálculo Hidrometeorológico de Aportaciones y Crecidas del Centro de Estudios Hidrográficos del CEDEX

Además de identificar tendencias a largo plazo, las series temporales hidrológicas también son útiles para prever eventos extremos, como inundaciones y sequías. Al estudiar cómo cambian las condiciones hidrológicas a lo largo del tiempo, se pueden desarrollar modelos predictivos que permitan anticipar y mitigar los impactos de estos eventos.

Las series temporales hidrológicas también son importantes para la gestión del agua y la planificación de recursos hídricos. Por ejemplo, los gestores de embalses pueden utilizar datos históricos

de niveles de agua para tomar decisiones informadas sobre cuándo liberar o retener agua para satisfacer la demanda de agua potable, riego y generación de energía.

Además de su importancia práctica, las series temporales hidrológicas también son valiosas para la investigación científica. Se pueden utilizar estos datos para estudiar una amplia gama de fenómenos hidrológicos, desde el ciclo del agua en ecosistemas naturales hasta la respuesta de los sistemas fluviales a eventos extremos como huracanes y tormentas.

Las series temporales hidrológicas son una herramienta invaluable para comprender y gestionar los recursos hídricos. Proporcionan información vital sobre cómo está cambiando el ciclo del agua con el tiempo, y cómo estos cambios pueden afectar a las comunidades humanas y los ecosistemas naturales. Al analizar y utilizar de manera efectiva los datos de series temporales hidrológicas, se puede trabajar hacia una gestión más sostenible y resiliente del agua en todo el mundo [10].

2.3 Anomalías en series temporales

La exploración de anomalías en series temporales hidrológicas se erige como un pilar esencial en la gestión eficiente de recursos hídricos y la salvaguarda de sistemas de agua sostenibles. En un contexto donde las demandas de agua crecen y los fenómenos climáticos extremos se vuelven moneda corriente, la identificación y comprensión temprana de irregularidades en el comportamiento hidrológico se vuelven críticas para garantizar la seguridad hídrica de las comunidades.

La detección de anomalías en series temporales hidrológicas no se limita simplemente a la eficiente distribución del agua, sino que abarca la identificación y mitigación de irregularidades que podrían afectar la disponibilidad y calidad del recurso. Aunque la automatización y la inteligencia artificial son cruciales, la necesidad de enfoques analíticos sólidos se vuelve aún más evidente en la gestión de recursos hídricos a gran escala.

La gestión integral del agua implica la adaptación de metodologías destinadas a detectar irregularidades en consumos urbanos a contextos hidrológicos más amplios. Para detectar estas anomalías se deben observar los datos a lo largo del tiempo y, fijarse especialmente en aquellos datos que disten del resto, como se observa en la Figura 2-3, que muestra en el eje y los datos recogidos de la cantidad de amonio en el agua y el eje x se corresponde a los días. Se observa que hay una clara anomalía en estos datos. Irregularidades en patrones de flujo, niveles de ríos o precipitaciones pueden indicar problemas en la infraestructura, cambios climáticos o eventos extremos. La aplicación de enfoques automatizados para identificar patrones anómalos se revela esencial en este sentido, ya que los grandes volúmenes de datos hidrológicos demandan análisis eficientes.



Figura 2-3. Ejemplo de una anomalía en una serie temporal

El desafío radica no solo en detectar pérdidas técnicas, sino también en comprender patrones no técnicos que pueden afectar la hidrología. La eficiencia de la identificación de consumos inusuales se ve maximizada con la automatización y flujos de trabajo inteligentes, reduciendo el tiempo necesario para identificar anomalías y permitiendo respuestas más rápidas a cambios significativos en el comportamiento del agua.

La diversidad de enfoques algorítmicos, destacada en investigaciones previas, proporciona una visión más completa de las posibles anomalías. En el ámbito hidrológico, donde múltiples factores influyen en la dinámica del agua, la aplicación de algoritmos diversos podría revelar patrones anómalos desde perspectivas múltiples.

La exploración de anomalías en series temporales hidrológicas requiere un enfoque integral que combine la automatización, la inteligencia artificial y metodologías analíticas sólidas. Si bien la investigación previa ha sentado las bases, la adaptación y expansión de estas metodologías para abordar desafíos específicos del ciclo hidrológico y eventos climáticos podría conducir a enfoques más robustos en la detección de anomalías hidrológicas.

2.4 Problema ambiental

La proliferación de cianobacterias y la producción de cianotoxinas en cuerpos de agua dulce representan un problema ambiental de creciente preocupación a nivel mundial. En este subapartado del estado del arte, se explora en detalle la naturaleza y las implicaciones de este problema, así como los avances recientes en la investigación sobre la predicción de episodios de cianotoxinas mediante el análisis de datos.

La proliferación de cianobacterias o floración de algas nocivas, se ha vuelto más frecuente y severo en los últimos años debido a una serie de factores antropogénicos y ambientales. Bajo ciertas condiciones, como altas concentraciones de nutrientes y temperaturas cálidas, las poblaciones de cianobacterias pueden experimentar un crecimiento explosivo, formando densas capas en la superficie del agua conocidas como “bloom” o floración. [11]

Como ya se ha comentado con anterioridad, las cianotoxinas pueden causar una amplia gama de efectos negativos, que van desde irritaciones cutáneas y gastrointestinales hasta daños en el hígado, el sistema nervioso y los órganos internos, e incluso la muerte en casos extremos. Por ello al trabajar en el control de estas, los trabajadores deberán tomar siempre las precauciones oportunas y deberán estar en todo momento correctamente pertrechados, como se muestra en la Figura 2-4. La presencia de cianotoxinas en fuentes de agua potable representa un riesgo significativo, ya que los sistemas de tratamiento convencionales pueden no ser efectivos para eliminar completamente estas sustancias nocivas. [12]



Figura 2-4. Control de cianobacterias en el lago San Roque, Argentina [13]

Varios factores contribuyen al aumento de las floraciones de cianobacterias y la producción de cianotoxinas en cuerpos de agua dulce. Entre ellos, la eutrofización puede ser uno de los principales impulsores de este problema. La eutrofización se refiere al enriquecimiento de nutrientes, como nitrógeno y fósforo, en los cuerpos de agua, generalmente como resultado de la vertiente de fertilizantes agrícolas, aguas residuales urbanas y otras fuentes de contaminación. Estos nutrientes proporcionan el alimento necesario para el crecimiento rápido de las cianobacterias, promoviendo así la formación de floraciones y la producción de cianotoxinas. En la Figura 2-5 se observa este fenómeno.



Figura 2-5. Agua afectada por la eutrofización [14]

El cambio climático también está exacerbando este problema al alterar las condiciones ambientales que pueden favorecer la proliferación de cianobacterias. El aumento de las temperaturas del agua y los patrones de precipitación irregulares pueden crear condiciones más propicias para el crecimiento de cianobacterias, mientras que los cambios en los patrones de circulación del agua pueden contribuir a la persistencia de las floraciones y la propagación de cianotoxinas a través de los sistemas acuáticos.

Además, la degradación del hábitat natural y la modificación de los cursos de agua, como la construcción de represas y la canalización de ríos, pueden alterar los flujos de nutrientes y la dinámica de las comunidades acuáticas, creando condiciones más favorables para la proliferación de cianobacterias. La urbanización y el desarrollo costero también pueden aumentar la carga de contaminantes en los cuerpos de agua, exacerbando aún más el problema de la eutrofización y la proliferación de cianobacterias, en la Figura 2-6 se observa un episodio en uno de los ríos estudiados.



Figura 2-6. Río Limia durante una floración de cianotoxinas [15]

En los últimos años, ha habido avances significativos en la investigación sobre la predicción de episodios de cianotoxinas mediante el análisis de datos. Los enfoques tradicionales de monitoreo y muestreo han sido complementados y en algunos casos reemplazados por métodos basados en el análisis de grandes volúmenes de datos ambientales y biológicos. Estos enfoques aprovechan la capacidad de las tecnologías modernas, como sensores remotos, estaciones meteorológicas automatizadas y sistemas de

información geográfica, para recopilar datos a gran escala sobre factores ambientales y biológicos relevantes para la proliferación de cianobacterias.

El análisis de datos permite identificar patrones y tendencias en los datos recopilados, lo que puede ser utilizado para desarrollar modelos predictivos de la ocurrencia de episodios de cianotoxinas. Estos modelos pueden integrar múltiples variables ambientales y biológicas, como la temperatura del agua, la concentración de nutrientes, la velocidad y dirección del viento, la radiación solar y la dinámica de las comunidades de cianobacterias, para predecir la probabilidad de formación de floraciones y la producción de cianotoxinas en un cuerpo de agua dado.

Además, el análisis de datos también puede ayudar a identificar factores de riesgo y condiciones predisponentes para la proliferación de cianobacterias, lo que puede ser útil para la gestión proactiva de los recursos hídricos. Por ejemplo, la identificación de áreas propensas a la eutrofización o la identificación de patrones climáticos que favorecen la proliferación de cianobacterias puede ayudar a orientar la implementación de medidas preventivas, como la reducción de la carga de nutrientes en los cuerpos de agua o la implementación de prácticas de gestión de tierras que minimicen la escorrentía de contaminantes.

Por lo tanto, el análisis de datos ofrece nuevas oportunidades para mejorar la comprensión y la predicción de episodios de cianotoxinas en cuerpos de agua dulce. Al integrar datos de múltiples fuentes y aplicar técnicas avanzadas de análisis, los investigadores pueden desarrollar modelos predictivos precisos que ayuden frente a la creciente amenaza de las floraciones de cianobacterias y la producción de cianotoxinas. Mediante el análisis de datos hidrológicos en la Figura 2-7 se puede observar la concentración de caudal de un río y se puede dividir en zonas de mayor o menos caudal según las conclusiones que se extraen.



Figura 2-7. Uso de análisis de datos hidrológicos para el caudal [16]

2.5 El estudio de las cianotoxinas

El estudio de las cianotoxinas en ecosistemas acuáticos ha experimentado un crecimiento significativo en las últimas décadas, impulsado por la creciente preocupación por la calidad del agua y sus implicaciones. Las cianotoxinas, producidas por cianobacterias durante floraciones algales, han emergido como un tema crítico de investigación debido a sus efectos perjudiciales y su prevalencia en diversos cuerpos de agua en todo el mundo.

Este crecimiento en la investigación refleja una creciente conciencia social sobre la importancia crucial de mantener nuestros ecosistemas acuáticos. El fenómeno de las floraciones algales, impulsado por la proliferación de cianobacterias, ha despertado una atención urgente hacia los riesgos asociados con las cianotoxinas. Estos compuestos tóxicos plantean desafíos significativos para la estabilidad y el equilibrio de los ecosistemas acuáticos.

La investigación en cianotoxinas (Figura 2-8) no se limita a observar estos fenómenos naturales, sino que se ha convertido en un esfuerzo vital para comprender los mecanismos subyacentes de su

proliferación y los impactos que generan. Este campo de estudio, multidisciplinario y dinámico, explora la compleja interacción entre factores climáticos, la disponibilidad de nutrientes, las relaciones biológicas y las consecuencias para el entorno. Además de abordar los problemas actuales, la investigación en cianotoxinas busca anticipar y prepararse para desafíos futuros, particularmente en el contexto del cambio climático y otras alteraciones ambientales.



Figura 2-8. Toma de muestras de agua contaminada por cianotoxinas [17]

La comprensión profunda de las cianotoxinas y su influencia en los ecosistemas acuáticos es fundamental para la gestión sostenible de nuestros recursos hídricos. La investigación no solo se centra en identificar y medir estas toxinas, sino también en evaluar sus efectos a largo plazo y en este caso, en predecir las apariciones de estos episodios para poder mitigar sus consecuencias o incluso tratar de evitar que se produzcan estos “bloomings” con algún tipo de contramedidas. A medida que se avanza en el tiempo, la preservación de nuestros recursos acuáticos cobra aún más importancia, la investigación en cianotoxinas no solo proporciona respuestas a los desafíos actuales, sino que también arroja luz sobre cómo se puede proteger y prevenir amenazas futuras.[12]

2.6 Clasificación y tipos de cianotoxinas

La clasificación de las cianotoxinas y su diversidad de tipos constituyen un amplio campo de estudio que se ha desarrollado a lo largo de décadas, impulsado por la necesidad de comprender los riesgos asociados con estas sustancias tóxicas en los ecosistemas acuáticos. Cada grupo exhibe propiedades tóxicas únicas, afectando diferentes sistemas biológicos, desde daño hepático hasta efectos neurotóxicos, lo que enfatiza la necesidad de una comprensión detallada de cada cianotoxina. El estudio de la taxonomía de las cianobacterias se erige como un pilar crucial en el entendimiento de la relación entre la diversidad de estos microorganismos y la síntesis de cianotoxinas.

A través de investigaciones meticulosas, se han desentrañado las complejidades de las diferentes cepas de cianobacterias, identificando sus características distintivas y explorando cómo estas peculiaridades se relacionan con su capacidad para producir cianotoxinas específicas. Esta exploración taxonómica ha llevado a avances significativos, permitiendo a los investigadores clasificar las diversas especies de cianobacterias. A medida que se descubren nuevas cepas y se perfeccionan las técnicas de análisis, se logra una comprensión más profunda de cómo la taxonomía de las cianobacterias influye en la dinámica de las poblaciones y, por ende, en la posible producción de cianotoxinas. [18]

Entre las cianotoxinas más estudiadas se encuentran las microcistinas, hepatotóxicas por naturaleza, que pueden causar daño hepático y están vinculadas a problemas de salud a largo plazo. Estas toxinas son producidas principalmente por cianobacterias del género *Microcystis* y pueden contaminar cuerpos de agua dulce. Su aspecto físico es el de la Figura 2-9.

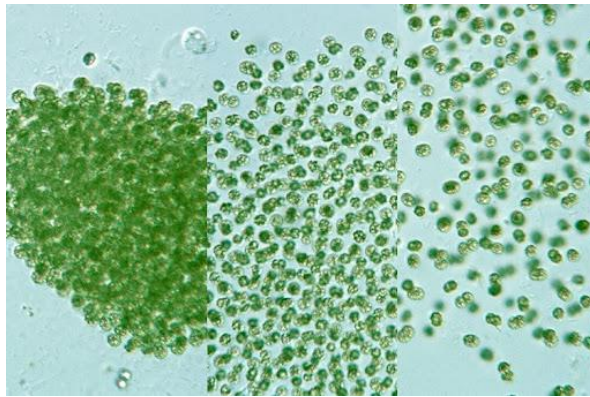


Figura 2-9. Microcistinas [19]

Otro grupo de cianotoxinas destacado son las anatoxinas, conocidas por ser neurotóxicas. La anatoxina-a, en particular, puede actuar rápidamente y causar síntomas como convulsiones y parálisis. El impacto de estas toxinas se extiende a los organismos acuáticos y a quienes dependen del agua para su consumo.

Las saxitoxinas, clasificadas como neurotoxinas, son comúnmente asociadas con las "mareas rojas" en ambientes marinos. Estas cianotoxinas pueden provocar parálisis respiratoria y, en casos extremos, llevar a la muerte. Su presencia en fuentes de agua dulce también plantea preocupaciones significativas, especialmente en áreas propensas a floraciones algales, y son producidas por un dinoflagelado marino que se muestra en la Figura 2-10.



Figura 2-10. *Gymnodinium catenatum*, productora de saxitoxinas [20]

Las nodularinas, hepatotóxicas como las microcistinas, representan otra categoría de cianotoxinas con potencial daño hepático. Su estudio se ha centrado en comprender cómo estas toxinas actúan, contribuyendo así a estrategias de gestión más efectivas.

Cilindropermopsinas, hepatotóxicas y vinculadas a problemas hepáticos, también son objeto de atención en la clasificación de cianotoxinas. La preocupación por su presencia en cuerpos de agua dulce destaca la necesidad de una comprensión más profunda de sus efectos y la implementación de medidas preventivas. Se muestran en la Figura 2-11.

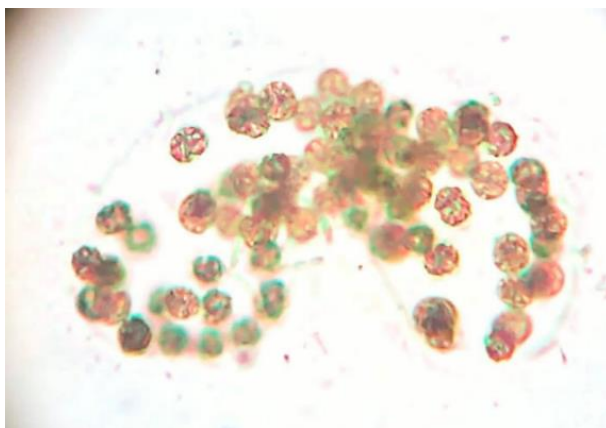


Figura 2-11. Cilindrospermopsina [21]

Este panorama diverso de cianotoxinas subraya la importancia de la investigación continua en la clasificación y caracterización de estas sustancias. A medida que se identifican nuevas cepas de cianobacterias y se profundiza en la comprensión de las interacciones entre factores ambientales y la producción de cianotoxinas, se abre el camino para estrategias más eficaces de monitoreo y gestión. El desafío persiste en anticipar y abordar las complejidades de estos compuestos tóxicos, salvaguardando así la integridad de los ecosistemas acuáticos y la seguridad de las comunidades que dependen de estas fuentes de agua. [22]

2.7 Impacto en la salud humana y ambiental

Como ya se ha comentado con anterioridad, la amenaza que representan las cianotoxinas en fuentes de agua potable trasciende los límites de la preocupación, impactando de manera directa la salud humana y generando efectos perjudiciales en los delicados equilibrios de los ecosistemas acuáticos. La exposición a estas toxinas plantea riesgos considerables que van más allá de simples inquietudes, manifestando consecuencias tanto agudas como crónicas que resaltan la urgencia de abordar este problema de manera integral.

En términos de salud humana, la exposición a cianotoxinas puede dar lugar a efectos adversos inmediatos y a largo plazo. Las consecuencias agudas pueden incluir impactos en el sistema nervioso, manifestándose a través de síntomas como mareos, náuseas y debilidad muscular. En casos más graves, la exposición a concentraciones elevadas de cianotoxinas puede conducir a efectos neurotóxicos severos, resultando en convulsiones, parálisis e incluso la muerte en situaciones extremas. Además de las repercusiones inmediatas, la exposición crónica a cianotoxinas ha sido asociada con problemas de salud a largo plazo. La afectación del sistema hepático se destaca como una preocupación particular, con la posibilidad de daño progresivo al hígado y el desarrollo de enfermedades hepáticas crónicas. Estos riesgos enfatizan la necesidad de medidas preventivas y una gestión eficaz para evitar la contaminación de fuentes de agua potable con cianotoxinas. Su influencia puede ir más allá de la fauna acuática y del consumo humano, ya que diversas especies pueden consumir agua de estas características, como en la Figura 2-12.



Figura 2-12. Perro consumiendo agua de un cuerpo de agua infectado de cianobacterias [23]

El impacto ambiental de las cianotoxinas se extiende más allá de la esfera humana, afectando directamente a la fauna acuática y alterando los ecosistemas acuáticos, como se puede observar claramente en la Figura 2-13. La biodiversidad se ve comprometida, ya que las cianotoxinas pueden tener efectos perjudiciales en organismos acuáticos, desde invertebrados hasta peces. Las floraciones algales masivas pueden dar lugar a zonas muertas, agotando los niveles de oxígeno y afectando negativamente a la vida acuática en estas áreas. [24]



Figura 2-13. Fauna acuática afectada por una floración de cianotoxinas [5]

La gestión efectiva de las cianotoxinas es crucial para preservar la integridad y la funcionalidad de los ecosistemas acuáticos. Estrategias de monitoreo y prevención, combinadas con prácticas de gestión sostenible, son esenciales para abordar este desafío complejo y proteger tanto a las comunidades humanas como a los frágiles equilibrios naturales que dependen del agua para su supervivencia.

2.8 Métodos de Detección y Monitoreo (Redes de Control de Calidad de las Aguas)

La detección temprana y precisa de cianotoxinas es fundamental para la gestión eficiente de los riesgos asociados. Los métodos analíticos, han evolucionado para mejorar la sensibilidad y especificidad en la identificación de diferentes cianotoxinas. La implementación de sistemas de monitoreo en tiempo real también ha avanzado, proporcionando datos cruciales para la toma de decisiones oportunas.

En el contexto de Galicia, la gestión de las cuencas hidrográficas se divide en dos ámbitos distintos, cada uno bajo la responsabilidad de diferentes organismos gubernamentales. En primer lugar, tenemos la cuenca Galicia-costas, que abarca las zonas costeras de la región, y su gestión está a cargo de la Xunta de Galicia, el gobierno autonómico de la comunidad. Por otro lado, encontramos las aguas interiores, que comprenden los ríos y cuerpos de agua situados tierra adentro, y su gestión corresponde al Ministerio de Transición Ecológica y el Reto Demográfico (MITECO) [25]. Esta división se muestra en la Figura 2-14.

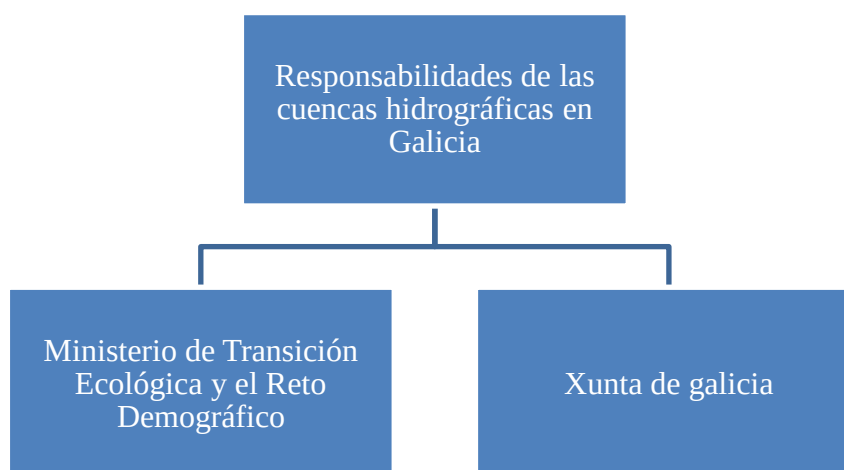


Figura 2-14. División de la gestión de las cuencas hidrográficas en Galicia

Esta división de responsabilidades puede generar ciertas problemáticas en cuanto a la recopilación y análisis de datos relacionados con la calidad del agua y otros aspectos hidrológicos. Por un lado, al estar a cargo de diferentes entidades, la coordinación y colaboración entre los organismos puede no ser óptima, lo que podría dificultar la compartición de información y la implementación de estrategias de gestión integrada de los recursos hídricos. Esto puede resultar en lagunas en la recopilación de datos y en la falta de una visión holística de la situación de las cuencas hidrográficas.

Además, las diferencias en los protocolos y métodos de recopilación de datos entre la Xunta de Galicia [26] y el Ministerio pueden introducir sesgos o inconsistencias en los conjuntos de datos disponibles, lo que podría afectar la calidad y fiabilidad de los análisis realizados. Por ejemplo, podrían existir discrepancias en los parámetros medidos, las frecuencias de muestreo o los métodos de análisis utilizados, lo que dificultaría la comparación de datos entre las diferentes cuencas hidrográficas y la identificación de tendencias regionales o patrones comunes.

En resumen, la división de responsabilidades sobre las cuencas hidrográficas de Galicia entre la Xunta de Galicia y el Ministerio puede plantear desafíos significativos en términos de recopilación y análisis de datos. Para abordar estas problemáticas, es fundamental promover la colaboración y coordinación entre los diferentes organismos responsables, así como estandarizar los protocolos y métodos de recopilación de datos para garantizar la consistencia y fiabilidad de la información disponible. De esta manera, será posible obtener una visión más completa y precisa de la situación de los recursos hídricos en la región y tomar decisiones informadas para su gestión y protección. [27]

3. DESARROLLO DEL TFG

3.1 Obtención de series temporales

Este apartado se enfocará en la obtención de series temporales donde se han analizado datos de distintos embalses con el objetivo de predecir la presencia de cianotoxinas. Los embalses de los que se han solicitado datos son del río Miño, del río Limia, del río Forcadas y datos del río Tea. La obtención de series temporales es un paso fundamental en nuestro estudio, ya que permite entender la evolución de variables clave a lo largo del tiempo y establecer relaciones significativas entre estas variables y la presencia de cianotoxinas.

Para comenzar, se han recopilado datos históricos de diversos embalses, incluyendo mediciones de parámetros como la temperatura del agua, la cantidad de oxígeno, la turbidez, la cantidad de amonio y otros indicadores de la calidad del agua. Estos datos se han obtenido de las páginas de las agencias gubernamentales correspondientes y organismos de gestión de recursos hídricos ya mencionados, donde se deben pedir los datos personalmente y unas semanas más tarde los envían al correo electrónico que se haya indicado en la solicitud.

Los parámetros de la Tabla 3-1, son de los que se han recibido datos para cada uno de los ríos. Como se puede ver, estos datos han sido recopilados diariamente en el periodo del 2018 al 2023 a excepción del embalse de As Forcadas en el que se han obtenido a intervalos más irregulares, sin embargo, durante un periodo más largo de investigación ya que estos son desde el año 2010 al 2023.

Río	Parámetros	Periodo de tiempo
Miño	Conductividad, pH, amonio, turbidez, materia orgánica.	2018-2023
Tea	Conductividad, pH, oxígeno, turbidez, materia orgánica.	2018-2023
Limia	Conductividad, amonio, turbidez, fosfatos.	2018-2023
Forcadas	Fosfatos, nitratos, pH, temperatura.	2010-2023

Tabla 3-1. Conjunto de parámetros obtenidos de cada río

3.2 Preprocesado de datos

Una vez recopilados los datos, se ha procedido a su procesamiento y limpieza. Esto implica identificar y corregir posibles errores, que podrían afectar la calidad de las series temporales. Además, se ha realizado una exploración inicial de los datos para detectar tendencias, estacionalidades y otros

patrones que puedan influir en la presencia de cianotoxinas en los embalses. Esto se ve representado en la Figura 3-1, donde se aprecia la importancia de este paso al preparar y ordenar los datos de la manera adecuada.

Con los datos limpios y preparados, se procedió a la construcción de las series temporales. Esto implica organizar los datos y calcular estadísticas resumen, como promedios o medianas, para cada intervalo. También se ha considerado la posibilidad de utilizar técnicas de interpolación para completar lagunas en los datos y garantizar la continuidad de nuestras series temporales.

Además, se han llevado a cabo análisis de correlación entre las diferentes variables para identificar posibles relaciones entre estas y la presencia de cianotoxinas. Esto ha permitido determinar qué variables tienen un mayor impacto en la aparición de floraciones de cianobacterias y, por lo tanto, son las más relevantes para la investigación en curso.

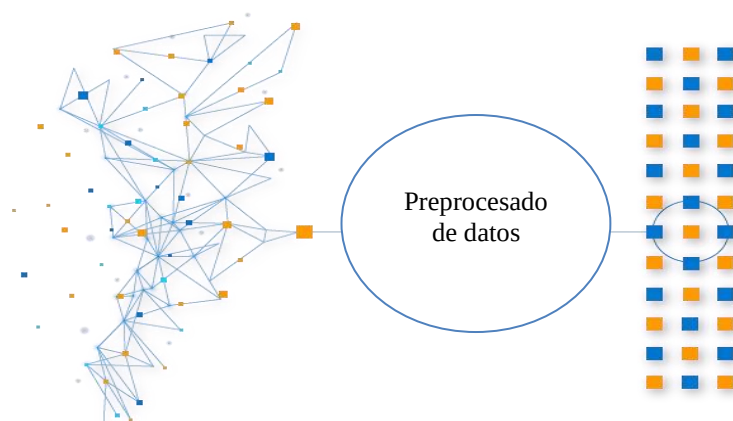


Figura 3-1. Representación de la utilidad del preprocesado de datos

La obtención de series temporales es un paso crucial en el presente estudio, ya que proporciona la base de datos necesaria para analizar y predecir la presencia de cianotoxinas en embalses de agua. A través de este proceso, se identifican patrones y relaciones significativas entre variables ambientales y la presencia de cianobacterias, lo que permitirá desarrollar modelos predictivos precisos y efectivos para mitigar los riesgos asociados con estas floraciones.

3.3 Detección de anomalías Z-scores

La investigación en el ámbito de la ingeniería ambiental es crucial para abordar desafíos complejos relacionados con la conservación de los recursos naturales y la protección del medio ambiente. La problemática de las cianobacterias, cada vez más relevante debido al cambio climático y la contaminación, requiere un enfoque multidisciplinar que combine conocimientos de biología, química y análisis de datos.

Lo primero como ya se ha mencionado fue la recopilación de datos diarios desde el 2018 al 2023 proporcionados por los órganos pertinentes, incluyendo mediciones de variables ambientales como la cantidad de oxígeno en el agua, la turbidez del agua, el pH del agua, la cantidad de amonio en el agua, entre otros. Sin embargo, le sigue el desafío de analizar y comprender esta gran cantidad de datos para extraer conclusiones significativas sobre la presencia de cianobacterias.

Es en este punto donde entra en juego el método de análisis z-score, una herramienta estadística poderosa que permite evaluar la posición relativa de un valor dentro de un conjunto de datos y detectar valores atípicos o extremos. Una de las ventajas del método de análisis z-score es su capacidad para proporcionar una medida objetiva y estandarizada de la importancia relativa de cada variable en el modelo. Esto facilita la interpretación de los resultados y la identificación de posibles áreas de enfoque para futuras investigaciones. Por ejemplo, al observar que la turbidez del agua tiene un z-score

significativamente más alto que otras variables, se puede inferir que este factor puede ser un predictor importante de la presencia de cianobacterias y, por lo tanto, merece una atención particular en futuros estudios y estrategias de gestión. [28]

El origen del método de detección de anomalías por z-score se remonta a los principios fundamentales de la estadística y la teoría de probabilidad. El concepto de z-score se basa en la idea de la distribución normal, también conocida como la distribución de Gauss o campana de Gauss. En una distribución normal, la mayoría de los valores se agrupan alrededor de la media, y la probabilidad de encontrar valores extremadamente altos o bajos disminuye a medida que la media se aleja. Esta representación se puede ver en la Figura 3-2. Su aplicabilidad en la presente investigación sobre la proliferación de cianobacterias ha sido fundamental.

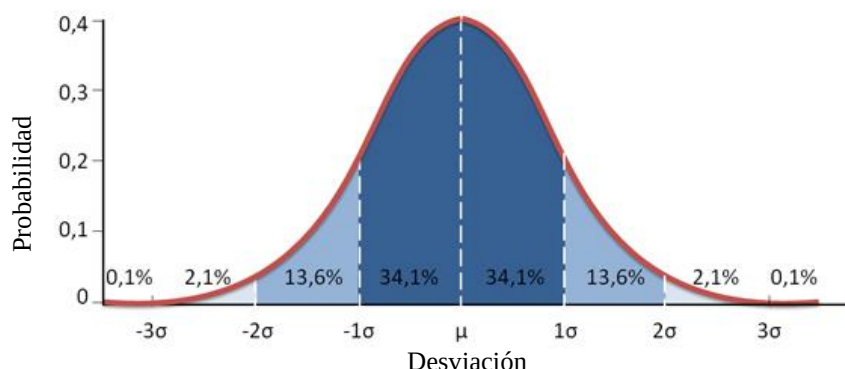


Figura 3-2. Distribución normal [29]

Para cada una de las variables ambientales recopiladas, se calcula el z-score correspondiente utilizando la fórmula estándar, que compara cada valor individual con la media y la desviación estándar del conjunto de datos. Este proceso permite identificar valores inusuales que podrían indicar condiciones ambientales excepcionales que favorecen el crecimiento de cianobacterias. Por ejemplo, al analizar la cantidad de oxígeno en el agua, los valores con z-scores negativos indicaron niveles de oxígeno por debajo de la media, lo que podría ser un indicador de condiciones anaeróbicas propicias para la proliferación de cianobacterias. En la ecuación, μ se refiere a la media de los datos y σ a su desviación típica. [30]

$$z = \frac{x - \mu}{\sigma} \quad \text{Ecuación Z-score (3-1)}$$

Lo esencial al utilizar este método es el hecho de que se considerarán valores extremos aquellos en los que el valor absoluto de la Z, al aplicar la Ecuación Z-score (3-1), sea mayor que 3. Esto se ha decidido en base a que en las bases de datos generales alrededor del 99,7% de los datos entran dentro de este umbral en el que $0 < Z < 3$. Por esto se interpreta que los valores que se salen de este umbral son resultados anómalos.

Al mismo tiempo, se ha explorado la relación entre las diferentes variables ambientales y la presencia de cianobacterias. Utilizando técnicas de correlación y regresión, se han podido identificar patrones significativos que sugieren una influencia directa de factores como la turbidez del agua y el pH del agua en la proliferación de cianobacterias.

Además de identificar relaciones entre variables individuales y la presencia de cianobacterias, el análisis z-score permitió evaluar la importancia relativa de cada variable en relación con otras variables en el modelo. Mediante la comparación de los z-scores de diferentes variables, se ha podido determinar qué variables tienen un mayor impacto en la presencia de cianobacterias y, por lo tanto, merecen una atención especial en futuros estudios y estrategias de gestión. [31]

3.4 Detección de anomalías método de la tercera derivada

Entre las herramientas utilizadas, se encuentra el método de la tercera derivada, el cual ha demostrado ser sumamente útil en el análisis de datos relacionados con la evolución temporal de los parámetros del agua. Esta herramienta matemática que analiza cómo cambian las tasas de cambio en función del tiempo o de una variable independiente, ofrece una perspectiva más profunda sobre la dinámica de los sistemas hidrológicos y la calidad del agua. Al aplicar este método al análisis de datos hidrológicos y de calidad del agua, se puede examinar detalladamente los cambios en la aceleración o desaceleración de los parámetros en cuestión a lo largo del tiempo.

El método de la tercera derivada se puede comprender fácilmente como una forma de examinar los cambios en la tasa de cambio de una serie de datos en función del tiempo o de otra variable independiente. Se puede tomar como ejemplo la variación en la concentración de ciertos componentes en el agua de los embalses a lo largo del tiempo. Si se calcula la primera derivada, se obtiene la tasa de cambio de esa concentración, es decir, cuánto cambia la concentración en cada intervalo de tiempo. Al calcular la segunda derivada, se puede ver cómo varía esa tasa de cambio, lo que proporciona información sobre si la concentración está aumentando o disminuyendo a un ritmo constante o cambiante. Al aplicar la tercera derivada lleva a lo que se conoce en inglés como "jerk", que representa la tasa de cambio de la aceleración con respecto al tiempo. Aunque pueda sonar peculiar, este concepto es relevante en física, dinámica de sistemas y otros aspectos de la estadística y el análisis de datos. Y es aquí donde entra cobra importancia la tercera derivada.[32]

La tercera derivada permite analizar cómo cambia la tasa de cambio de la concentración en cada momento específico. En esencia, proporciona una visión detallada de la aceleración o desaceleración de los cambios en la concentración del contaminante en el embalse. Este enfoque ayuda a identificar momentos críticos en los datos, tales como cambios bruscos en la evolución de la concentración, que pueden indicar eventos significativos, como vertidos contaminantes o condiciones ambientales extremas, como tormentas intensas.

Además, la tercera derivada permite distinguir entre variaciones estacionales y tendencias a largo plazo en los datos. Por ejemplo, puede revelar si los cambios observados son parte de un patrón estacional predecible o si indican una tendencia más preocupante que requiere atención inmediata. Esta capacidad de discernimiento es crucial para la gestión efectiva de los recursos hídricos y la protección de la calidad del agua. [33]

La aplicación del método de la tercera derivada en el análisis de datos hidrológicos proporciona una comprensión más profunda de la dinámica de los sistemas acuáticos y los factores que influyen en su calidad. Su capacidad para detectar cambios sutiles y significativos en los datos es fundamental para la toma de decisiones informadas en la gestión ambiental y la protección de los ecosistemas acuáticos.

Se trata de aplicar la ecuación de la aproximación de derivadas por diferencias finitas en los datos para sacar una tabla y poder extraer los datos que se buscan. Se sugiere una combinación lineal de valores en diferentes puntos. La Ecuación de la 3ª derivada (3-2) proporciona una forma de calcular la tercera derivada utilizando valores de la función de cuatro puntos equidistantes, siendo esta $h=1$. Esta forma es reminiscente de un enfoque conocido en análisis numérico para aproximar derivadas utilizando diferencias finitas.

$$f'''(x) \approx \frac{f(x+3h) - 3f(x+2h) + 3f(x+h) - f(x)}{h^3} \quad \text{Ecuación de la 3ª derivada (3-2)}$$

Se ha trabajado sobre una tabla como la del ejemplo de la Tabla 3-2 para todos los elementos estudiados, para todas las semanas estudiadas y para todos los ríos estudiados.

N.º semana	Conductividad	3ª derivada
4	61.756	6.723
5	63.549	3.870
6	63.874	-8.252
7	59.151	-3.579
8	58.801	9.419
9	57.884	-4.938

Tabla 3-2. Extracto de resultados 3ª derivada semanas 4 a 9 de la conductividad para el río Miño

En base a estos resultados obtenidos para la columna de la 3ª derivada, se puede deducir cuales son los resultados anómalos. En la siguiente Tabla 3-3 se observa subrayado el valor anómalo claro que dista en gran medida del resto de valores

N.º semana	Media amonio	3º derivada
35	135.232	0.944
36	140.445	-10.385
37	142.027	1.095
38	61.213	-78.761
39	142.321	244.312

Tabla 3-3. Extracto de amonio en el río Limia, con el valor anómalo en negrita

De esta manera se obtienen una serie de desviaciones respecto al resto de valores que merecen estudio y se han corroborado con los valores obtenidos por el resto de los métodos.

3.5 Detección de anomalías por descomposición en TEA

En el entorno del análisis de datos, existe una técnica llamada análisis de anomalías por descomposición en tendencia, estacionalidad y aleatoriedad. A pesar de su nombre complejo, este enfoque es bastante simple de comprender y resulta de gran utilidad en la detección de patrones inusuales en los datos.

La esencia de este método radica en dividir los datos en tres partes fundamentales: la tendencia, la estacionalidad y la aleatoriedad. La tendencia indica la dirección general en la que se mueven los datos a lo largo del tiempo, si están aumentando, disminuyendo o permanecen estables. Por ejemplo, si estamos examinando la concentración de ciertos contaminantes en el agua de un embalse a lo largo de varios meses, la tendencia podría indicar si esa concentración está aumentando gradualmente o manteniéndose constante.

La aleatoriedad representa las fluctuaciones impredecibles en los datos, como el ruido de fondo. Estas variaciones podrían deberse a factores como errores en la medición o eventos imprevistos. La aleatoriedad en series temporales es un concepto clave en el análisis de datos, ya que determina si los valores observados en una serie temporal están distribuidos de manera impredecible y sin un patrón discernible. [34]

La estacionalidad, por otro lado, muestra patrones que se repiten en momentos específicos. En el caso del agua de un embalse, podríamos observar si hay ciertas épocas del año en las que la concentración de ciertos contaminantes tiende a ser más alta o baja. Como se muestra en la Figura 3-3, en la que se utiliza un ejemplo claro de datos que dependen de la estación del año, la cantidad de viajes que se realizan. Donde se repiten las tendencias periódicamente dependiendo del mes del año en el que se estudian los datos. Esto podría estar influenciado por factores estacionales como las condiciones climáticas o el uso de pesticidas en la agricultura durante ciertas temporadas.

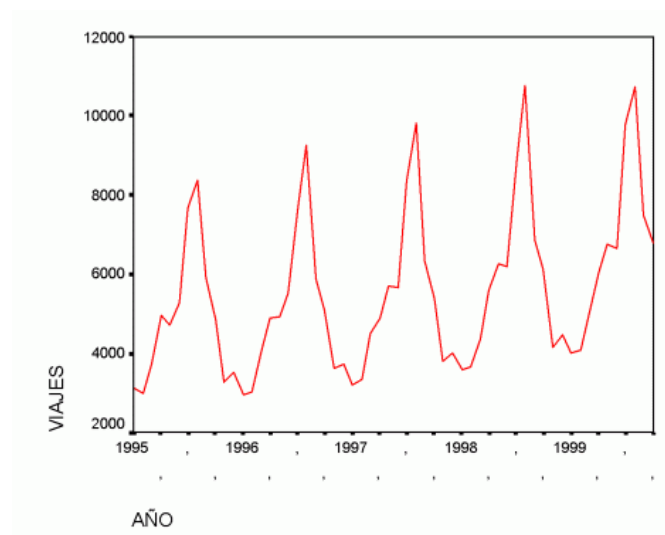


Figura 3-3. Estacionalidad en una serie temporal [35]

Esta división por estacionalidad tiene una cierta connotación con la que se interpreta que se estudian datos meteorológicos también, ya que al estudiarse datos divididos en estaciones se observa el comportamiento del agua en distintas estaciones, con todo lo que ello conlleva: variación en precipitaciones, temperatura media, etc. Esto le añade otra dimensión al análisis permitiendo centrarse en ciertos periodos anuales.

Este es el enfoque clave de este método y para potenciarlo, se han calculado las medias de todos los datos por años y estaciones (cuartos de año), agrupándolos de esta manera en unas series temporales distintas aportando una perspectiva distinta. Una vez hecha esta división, el siguiente paso es calcular el CMA, este es el “Centered Moving Average”, o lo que es lo mismo, el promedio móvil centrado. Este es un tipo de promedio que trata de evitar el retraso que tienen los promedios móviles habituales. Mientras que los promedios móviles simples solo miran hacia atrás en el tiempo, este también considera datos futuros, ofreciendo una visión más completa de cómo están cambiando los datos. Esto lo hace muy útil para encontrar tendencias y posibles puntos donde los datos podrían cambiar de dirección. Se calcula mediante la Ecuación cálculo del CMA (3-3) en la cual: [33]

$$S_3 = Y_1/8 + (Y_2 + Y_3 + Y_4)/4 + Y_5/8.$$

Ecuación cálculo del CMA (3-3)

En esta ecuación para hallar el valor se utilizan 5 valores distintos. Suponiendo que se quiere obtener el valor de CMA de la conductividad del primer cuarto de 2019. Se utilizarían los datos del tercer y último cuarto de 2018 y los datos del segundo y tercer cuarto de 2019, además del valor del primer cuarto de 2019, evidentemente. El más antiguo y el más moderno de los datos se dividen entre 8, el resto de los datos entre 4. Se hace un sumatorio de esto y de esta manera se obtiene el CMA de la conductividad para ese periodo del año. El siguiente paso es calcular una relación entre el valor medio de cada cuarto dividido entre su CMA correspondiente

Por lo que un resumen de los datos que hay en las tablas obtenidas sería el que se muestra en la Tabla 3-4:

Cuarto de año	Conductividad	CMA	Conductividad/CMA
1º	10,102	30,845	0.328

Tabla 3-4. Ejemplo de los datos que se obtienen por el método de estacionalidad para un cuarto de año

Es este último de los datos el que es interesante, se tomaría este resultado para cada uno de los años y se calcula la media de todos los años, así obteniendo la tendencia estacional de ese elemento en cada cuarto de año.

Como se ve en la Tabla 3-5, se han hecho las medias de los cuartos de todos los años y se han reunido en esta tabla, de esta manera se puede comparar el efecto que se da en estos factores según el cuarto de año en el que se encuentre. El método dictamina que aquellos valores que en esta última tabla son menores que 1, se denominan los valores “bust” en los que este factor no es propicio a sobresalir en ese periodo del año, serían los datos que se ven en un tono normal en la tabla. Por el otro lado, los valores que se ven en negrita son aquellos mayores que 1, los denominados “boom”. Los que señalan que esta es una época del año propicia para que los factores florezcan.

	Conductividad	Oxígeno	Turbidez	Materia Orgánica	pH
1^{er} cuarto	0.902	1.122	1.185	0.791	1.007
2^o cuarto	0.914	1.007	0.981	0.637	0.993
3^{er} cuarto	1.317	0.882	0.696	0.652	1.054
4^o cuarto	1.019	1.143	1.205	1.829	1.078

Tabla 3-5. Resultados finales para el método de la estacionalidad

Una vez se han descompuesto los datos en estos tres componentes, se comparan las observaciones reales con lo que se espera ver según la tendencia y la estacionalidad. Si se encuentran discrepancias significativas entre los valores reales y los valores esperados, eso podría indicar la presencia de una anomalía en los datos.

Por ejemplo, si se monitorea la presencia de cianobacterias en un embalse y aparece un aumento repentino en su concentración que no puede ser explicado por la tendencia o la estacionalidad esperada, eso podría indicar un brote inusual que requiere atención adicional.

3.6 Aprendizaje por conjuntos

El enfoque de Aprendizaje por Conjuntos en el análisis de datos se ha convertido en una herramienta invaluable para obtener una comprensión más completa y confiable de los fenómenos estudiados. En este sentido, la relevancia del trabajo de fin de grado radica en la utilización y comparación de tres métodos de análisis de datos diferentes: la Descomposición en TEA, el Z-Score y el Método de la Tercera Derivada.

El uso de estos tres métodos ofrece una perspectiva multifacética sobre los datos recopilados, permitiendo una evaluación exhaustiva de los mismos y una mayor confianza en los resultados obtenidos. En primer lugar, el Z-Score permite cuantificar la desviación de un punto de datos con respecto a la media de la distribución y expresar esta desviación en términos de desviaciones estándar. Esto ayuda a identificar puntos de datos atípicos que se encuentran significativamente por encima o por debajo de la media, lo que puede indicar la presencia de valores extremos o inusuales en los datos.

Por otro lado, el Método de la Tercera Derivada brinda información sobre cómo cambian las tasas de cambio de los datos en un momento dado, lo que permite identificar momentos críticos en los que la variación puede estar acelerándose o desacelerándose. Esta técnica es especialmente útil para detectar cambios abruptos en la tendencia de los datos o eventos que pueden tener un impacto significativo en el fenómeno estudiado. Finalmente, la Descomposición en TEA proporciona una visión detallada de la evolución de los datos a lo largo del tiempo, identificando patrones de tendencia, estacionales y fluctuaciones aleatorias. Esta técnica es especialmente útil para detectar anomalías o cambios inusuales en los datos, lo que puede proporcionar información valiosa sobre eventos significativos o problemas potenciales.

Al comparar los resultados obtenidos mediante estos tres métodos, se puede obtener una validación cruzada de los mismos, lo que aumenta la fiabilidad de los hallazgos y proporciona una imagen más

completa y precisa de los datos analizados. Esta se vería resumida en la Tabla 3-6. Esto contribuye a una toma de decisiones más informada y fundamentada en evidencia, lo que es fundamental en numerosos campos de aplicación, desde la investigación científica hasta la toma de decisiones empresariales.

Método Z-score	Método 3ª derivada	Conclusión
Resultado en la media	Resultado en la media	Resultado en la media
Resultado en la media	Resultado anómalo	Resultado en la media
Resultado anómalo	Resultado anómalo	Resultado anómalo

Tabla 3-6. Explicación de la validación cruzada

3.7 Validación mediante noticias de prensa

Son varios los periódicos locales (Figura 3-4) que se hacen eco de la aparición de cianobacterias y cianotoxinas cuando se dan estos episodios. Es evidente que es un tema digno de difusión ya que afecta directamente a la población que puede ver sus recursos esenciales limitados, por lo que la cobertura de estos hechos es esencial en dar a entender a la sociedad los problemas que pueden conllevar estas apariciones esporádicas y qué deben hacer para no verse afectados en demasía.

Para encontrar estos artículos de prensa que son esenciales para el trabajo se han utilizado los modificadores oportunos con el motor de búsqueda. La manera en la que se puede acotar para que los resultados se correspondan a los periodos temporales de los datos obtenidos es utilizando simultáneamente al final de la búsqueda la herramienta “after:yyyy/mm/dd” y “before:yyyy/mm/dd”. De esta manera, los resultados que esta búsqueda proporciona están acotados entre las dos fechas que se han introducido en dichos comandos. Este método ha sido de gran ayuda ya que ha permitido encontrar las noticias en el espacio temporal de interés.



Figura 3-4. Noticia de “La Voz de Galicia” acerca de las cianobacterias [36]

Además, la cobertura mediática puede ejercer una cierta influencia sobre las autoridades responsables y los organismos reguladores para que tomen medidas adecuadas para abordar el problema. Esto puede incluir la implementación de programas de monitoreo más rigurosos, la aplicación de estándares de calidad del agua más estrictos y la inversión en tecnologías de tratamiento de agua más

avanzadas. Sin la atención de los medios de comunicación, estos problemas podrían pasar desapercibidos o ignorarse, lo que podría tener consecuencias devastadoras.

Cabe señalar que cada vez más periódicos de mayor calado en el ámbito nacional han publicado noticias acerca de este tema. En periódicos a nivel nacional se habla de las cianobacterias, pero no se terminan de recalcar los efectos negativos sobre los recursos hidrológicos de ciertas localidades y la problemática que estas elevan. Haría falta una concienciación más responsable de estas bacterias y los efectos que pueden tener al consumirse productos procedentes de aguas infectadas.

De otra manera, los medios a nivel más local como “La Voz de Galicia” al notar su audiencia el impacto directo de estas floraciones, además de publicar cuando estos episodios de cianobacterias suceden para informar y alertar a la población de este suceso, publican artículos mucho más conscientes y de carácter informativo de lo que estos brotes pueden significar para la población por lo que tienen un enfoque distinto.

Una vez se han validado estos resultados y contrastado con la prensa se procede a calcular la tasa de positivos verdaderos. La importancia de estas tasas radica en su capacidad para evaluar con precisión la efectividad de un sistema de detección. Cuando se habla de positivos verdaderos, se habla de los casos en los que la prueba o el sistema identifica correctamente la presencia de algo que estamos buscando, como una enfermedad o un objeto de interés. Por otro lado, los negativos verdaderos son aquellos en los que la prueba o sistema descarta correctamente la ausencia de lo que estamos buscando.

Estas tasas son cruciales porque determinan cuán confiable es el resultado que obtenemos. Un alto porcentaje de positivos verdaderos significa que la prueba o sistema es efectivo para detectar lo que busca. De manera similar, un alto porcentaje de negativos verdaderos indica que el sistema es bueno para descartar lo que no estamos buscando. Por el contrario, altas tasas de falsos positivos y falsos negativos pueden llevar a conclusiones incorrectas y decisiones erróneas.

Después se observarán cuáles de estos resultados son falsos positivos, falsos negativos y positivos verdaderos siguiendo la metodología de la matriz de confusión (Tabla 3-7). [37]

	Positivo documentado	Negativo documentado
Positivo algoritmo	Positivo verdadero	Falso positivo
Negativo algoritmo	Falso negativo	Negativo verdadero

Tabla 3-7. Matriz de confusión para falsos positivos/negativos

Para calcular las tasas de positivos y negativos verdaderos se debe dividir el dato que se ha obtenido entre la suma de este dato y su opuesto. Es decir, como se muestra en la Ecuación TPV (3-4):

$$TPV = PV / (PV + FN) \quad \text{Ecuación TPV (3-4)}$$

En esta ecuación las siglas “TPV” significan “Tasa de Positivos Verdaderos”. Asimismo “PV” son los “Positivos Verdaderos” y “FN” hace referencia a los “Falsos Negativos”.

La tasa de positivos verdaderos es esencial porque proporciona una medida de confianza en la precisión de los métodos de detección y clasificación, esta tasa suele indicar la sensibilidad de los métodos.

4. RESULTADOS

4.1 Fechas prensa

Uno de los elementos esenciales de este trabajo es la investigación y el hallazgo de fechas de floraciones de cianotoxinas en los cuerpos de agua estudiados, resumidos en la Tabla 4-1.

RÍO	FECHA
Limia	30/09/2019 (tomado de T21 [38])
Limia	03/11/2019 (tomado de Faro de Vigo [39])
Limia	28/07/2021 (tomado de Faro de Vigo [40])
Limia	16/09/2022 (tomado de Ministerio de Sanidad [41])
Miño	03/11/2019 (tomado de Faro de Vigo [39])
Miño	26/10/2021 (tomado de La Voz de Galicia [42])
Forcadas	23/07/2015 (tomado de Diario de Ferrol [43])
Forcadas	28/08/2019 (tomado de La Voz de Galicia [44])
Forcadas	27/07/2020 (tomado de La Voz de Galicia [45])
Forcadas	21/05/2021 (tomado de Faro de Vigo [46])
Forcadas	06/09/2023 (tomado de La Voz de Galicia [47])

Tabla 4-1. Fechas de floraciones obtenidas por prensa de cada uno de los ríos estudiados

Se puede observar que las fuentes principales de información para dar con las fechas se compusieron por los medios locales “La Voz de Galicia” y “Faro de Vigo”. Estas fechas son esenciales para más tarde poder validar los resultados obtenidos mediante el análisis de datos y compararlos con las fechas reales en las que se dieron estos episodios.

Cabe mencionar que a pesar de que estas son las fechas que fueron publicadas en la prensa, es posible que haya habido floraciones de cianobacterias que no han sido publicadas y que, por lo tanto, no han quedado documentadas. Esto puede tener un cierto impacto en los resultados obtenidos por el modelo predictivo. Al no tener una fiabilidad total de que todos los casos han sido registrados se da cabida a la posibilidad de que una predicción esté en lo cierto y sí que se haya dado un episodio de cianobacterias, pero al no haber quedado documentado se da como una predicción errónea.

4.2 Influencia meteorológica

Una vez se conocen estas fechas que confirman las floraciones, primero se analiza la posible influencia que ha podido tener el factor climático sobre estas floraciones y hasta qué punto son relevantes los cambios meteorológicos en este asunto.

Mirando el historial en la web de Meteogalicia [48] se pueden obtener los datos de la Tabla 4-2. Se han tomado los datos de las estaciones de Ourense, A Capela y Xinzo de Limia. Se observan condiciones climáticas muy similares en todos los casos. Esto sucede debido a que todas las fechas documentadas son en la segunda mitad del año, entre los meses de julio y octubre mayormente, cuando se dan menos precipitaciones y hay unas temperaturas más altas y mayores niveles de humedad relativa. Aquellos en los que el apartado de precipitaciones aparece descrito como “muy débil” describe un nivel de precipitaciones menor a un promedio de 0,5 L/m² por hora.

Fechas	Temperatura media	Humedad relativa	Precipitaciones
30/09/2019 Limia	16°C	85%	NO
03/11/2019 Limia	14°C	96%	MUY DÉBIL
28/07/2021 Limia	17°C	80%	NO
16/09/2022 Limia	19°C	90%	MUY DÉBIL
03/11/2019 Miño	10°C	95%	NO
26/10/2021 Miño	17°C	75%	NO
23/07/2015 Forcadas	18°C	85%	NO
28/08/2019 Forcadas	19°C	89%	NO
27/07/2020 Forcadas	16°C	85%	MUY DÉBIL
21/05/2021 Forcadas	10°C	89%	MUY DÉBIL
06/09/2023 Forcadas	20°C	80%	NO

Tabla 4-2. Datos meteorológicos para las fechas de floraciones.

4.3 Análisis Z-score

La primera etapa del tratamiento de datos fue calcular las medias y estudiar el perfil de los datos para el intervalo solicitado. Los datos recibidos fueron diarios a lo largo de 5 años, por lo que para pasarlos a una manera más fácil de manipular se consideró la opción de hacer las medias semanales y tratarlos de tal manera. De este modo se dispone de más de 150 cifras representativas. El análisis empleado consiste en usar la ecuación ya mencionada con anterioridad. De estas cifras se calcula la media de todas ellas, y con la desviación típica de cada una se consigue un valor de Z como se indicó en la fórmula de la Ecuación Z-score (3-1). A continuación, en la Tabla 4-3 se muestra un ejemplo de un extracto de los datos para ver el formato y cómo quedan recopilados dichos datos.

Semana	Z-score amonio	Z-score conductividad	Z-score turbidez	Z-score fosfatos
1	0.412	0.792	-0.812	-0.879
2	0.644	1.318	-0.643	-0.876
3	0.595	1.994	-0.788	-0.874
4	0.488	2.011	-0.844	-0.872
5	0.582	2.239	-0.839	-0.871
6	-0.045	1.834	-0.813	-0.872
7	0.066	1.697	-0.821	-0.874
8	0.221	1.653	-0.814	-0.875
9	0.159	0.735	-0.756	-0.792
10	-0.089	0.423	0.972	-0.815

Tabla 4-3. Extracto de valores de Z de las primeras 10 semanas recogidas para el embalse del río Limia

Como se ve en la tabla superior, una vez obtuvimos los valores de Z para cada semana lo único que quedaba por hacer era identificar todos aquellos valores anómalos. Para esto la regla establecida ha sido que para que un valor sea definido como anómalo, el valor absoluto de su Z debe ser mayor que 3. Por lo que en la Tabla 4-4 se ve como el valor anómalo se destaca por ser mayor que 3. Así se puede ver claramente cuáles son los valores anómalos y se pueden reunir y tratar uno a uno.

Semana	Z-score amonio	Z-score conductividad	Z-score turbidez	Z-score fosfatos
151	-0.813	-0.762	-0.533	0.620
152	-0.826	-0.823	-0.385	-0.872
153	-0.822	-0.898	0.352	-0.816
154	-0.827	-1.281	-0.157	0.430
155	-0.829	-1.405	0.329	3.289

Tabla 4-4. Extracto de los resultados en el que está subrayado el valor anómalo

La mejor manera de afrontar esto consistió en reunir todos los resultados extremos para, de esta manera, poder compararlos con los resultados obtenidos por otros métodos. Por este motivo se creó otra tabla con los resultados anómalos obtenidos en este proceso, la Tabla 4-5. En esta tabla se resumen los resultados para el método de Z-score indicándose cuál fue la propiedad del agua que presentó la anomalía, la semana en la que se dio esta anomalía y, por último, el valor de Z relacionado a esta anomalía.

Resultados anómalos	Z-score
Turbidez semana 57	4.331
Turbidez semana 67	5.786
Turbidez semana 103	5.073
Fosfatos semana 155	3.289

Tabla 4-5. Resultados anómalos obtenidos por Z-score para el río Limia

Esta tabla servirá para ponerla junto a la tabla que obtendremos de los resultados de la tercera derivada y comparar resultados para ver si algunos de estos coinciden. Unidos a estos resultados del río Limia obtenemos resultados de Z-score para el Miño (Tabla 4-6), para el río Tea (Tabla 4-7) y para el Forcadás (Tabla 4-8).

Resultados anómalos	Z-score
Amonio semana 21	6.054
Turbidez semana 104	3.083
Turbidez semana 155	3.318
Amonio semana 206	4.211
Amonio semana 262	4.074
Amonio semana 263	5.069
Amonio semana 264	6.742
Amonio semana 265	5.757
Turbidez semana 260	3.411
Turbidez semana 261	3.124
Turbidez semana 262	4.882
Turbidez semana 263	3.650
Turbidez semana 264	6.362
Turbidez semana 265	4.521
Mat. Orgánica semana 309	3.058
Mat. Orgánica semana 310	3.443
Mat. Orgánica semana 311	3.359
Mat. Orgánica semana 312	3.210
Mat. Orgánica semana 313	3.444

Tabla 4-6. Resultados anómalos obtenidos por Z-score para el río Miño

Resultados anómalos	Z-score
Turbidez semana 39	3.672
Conductividad semana 142	3.025
Turbidez semana 150	6.137
Turbidez semana 163	4.072
Turbidez semana 168	3.759
Turbidez semana 169	7.423
Mat. Orgánica semana 191	3.380
Mat. Orgánica semana 192	3.461
Mat. Orgánica semana 193	4.846
Turbidez semana 251	3.317
Mat. Orgánica semana 251	3.680
Turbidez semana 256	4.191
Mat. Orgánica semana 256	4.394
Turbidez semana 260	3.387
Turbidez semana 261	5.092

Tabla 4-7. Resultados anómalos obtenidos por Z-score para el río Tea

Resultados anómalos	Z-score
Fosfatos 08/03/2010	5.901
Fosfatos 28/02/2011	5.594
Temperatura 08/11/2011	-3.491
Fosfatos 18/05/2022	3.378

Tabla 4-8. Resultados anómalos obtenidos por Z-score para el río Forcadas

Como se puede ver en esta última tabla, los datos recibidos para el embalse de Forcadas no son diarios, sino que corresponden a días dispersos por lo que se toman estos datos de manera directa. Esto puede ser significativo para el río Forcadas ya que la irregularidad en los datos es un inconveniente que puede conllevar imprecisión en los resultados.

4.4 Análisis método de la tercera derivada

Para el método de la tercera derivada de nuevo trataremos los datos con su media semanal para que estos sean algo más manejables. Este método es sencillo, pero obliga a tratar cada propiedad del agua estudiada de forma separada, en el sentido en que no es una tabla conjunta en la que obtengo resultados del amonio, turbidez, fosfatos, etc. (como si se da en el análisis por Z-score). Para analizar mediante la tercera derivada debe haber una tabla para cada uno de estos en la que se calculan los valores de la primera, segunda y tercera derivada de los promedios semanales obtenidos (Tabla 4-9).

N.º semana	Media amonio	1º derivada	2º derivada	3º derivada
1	0.142			
2	0.161	0.025		
3	0.165	-0.006	-0.032	
4	0.152	-0.017	-0.005	0.029
5	0.167	0.014	0.025	0.025
6	0.099	-0.068	-0.070	-0.101
7	0.104	0.014	0.074	0.158
8	0.120	0.012	0.005	-0.076
9	0.112	-0.007	-0.023	-0.033
10	0.097	-0.024	-0.016	0.007

Tabla 4-9. Extracto de valores de la 1ª, 2ª y 3ª derivada de las primeras 10 semanas muestreadas del amonio en el río Limia

Como se puede observar en la Tabla 4-9, para cada valor (a partir del cuarto valor) representando cada semana se puede obtener una tercera derivada. Es esta 3ª derivada la que va a ser significativa en nuestros estudios ya que es aquí donde se van a ver representados los valores anómalos. Similarmente al método de Z-score es en esta columna de la 3ª derivada donde se debe buscar e identificar estos valores. Mirando la media de las cifras obtenidas para la 3ª derivada es simple designar un valor límite a partir del cual las cifras que lo sobrepasen serán consideradas anómalas.

Cuando un valor dista del resto de valores obtenidos para la tercera derivada, se destaca. Este proceso debe repetirse para todos los factores estudiados y para todas las semanas de datos recibidas. Así obteniendo los resultados anómalos para cada uno de estos factores y reuniéndolos todos para su posterior comparación con el resto de los métodos (Tabla 4-10).

Resultados anómalos	3ª derivada
Conductividad semana 39	244.314
Conductividad semana 104	106.062
Turbidez semana 58	-81.927
Turbidez semana 59	78.991
Turbidez semana 68	-115.084
Turbidez semana 69	112.988

Tabla 4-10. Valores anómalos obtenidos por el método de la 3ª derivada para el río Limia

En la Tabla 4-10, se observan reunidos los resultados anómalos para este método indicándose el factor al que corresponde esta anomalía y el número de semana en la que ocurre. A continuación, se muestran los resultados para los ríos Miño (Tabla 4-11), Tea (Tabla 4-12) y Forcadás (Tabla 4-13).

Resultados anómalos	3ª derivada
Conductividad semana 40	35.103
Conductividad semana 45	31.532
Conductividad semana 92	42.205
Conductividad semana 98	58.719
Conductividad semana 104	31.724
Conductividad semana 105	36.405
Conductividad semana 190	32.201
Conductividad semana 206	30.721
Conductividad semana 249	31.199
Conductividad semana 250	31.575
pH semana 207	1.602
pH semana 208	2.844
pH semana 209	1.357
Turbidez semana 156	33.312
Turbidez semana 265	31.431
Mat. Orgánica semana 267	6.816

Tabla 4-11. Resultados anómalos obtenidos por la 3ª derivada para el río Miño

Resultados anómalos	3ª derivada
Conductividad semana 43	26.064
Conductividad semana 46	30.292
Conductividad semana 95	29.427
Conductividad semana 110	24.865
Conductividad semana 139	29.593
Conductividad semana 144	25.835
Conductividad semana 198	34.385
Conductividad semana 252	44.538
Conductividad semana 304	30.789
Turbidez semana 152	24.551
Turbidez semana 165	24.880
Turbidez semana 170	24.585
Turbidez semana 171	39.889

Tabla 4-12. Resultados anómalos obtenidos por la 3ª derivada para el río Tea

Resultados anómalos	3ª derivada
Temperatura 22/11/2010	22.504
Temperatura 28/02/2011	28.319
Temperatura 27/02/2012	57.622
Temperatura 25/06/2012	35.746
Temperatura 28/07/2015	23.759
pH 15/06/2017	8.382
pH 05/09/2023	8.176

Tabla 4-13. Resultados anómalos obtenidos por la 3ª derivada para el río Forcadás

Como se puede observar el formato de los resultados del método de Z-score y del método de la tercera derivada son muy similares. Como ambos ofrecen una semana específica en la que se da la anomalía, el objetivo es que una vez que ya se han realizado estos dos métodos puedan compararse a primera vista y comprobar si estos resultados anómalos coinciden entre sí y como no puede ser de otra manera, comprobar que coinciden con las fechas de los episodios.

4.5 Análisis descomposición en TEA

Para este tipo de análisis se han dividido los datos recibidos por años y posteriormente por cuartos de año, es decir, se ha calculado la media de los datos de enero, febrero y marzo de un año y el valor obtenido representará el valor del primer cuarto de dicho año. Esto se ha hecho para cada uno de los factores y para cada una de las aguas que se ha estudiado. Se muestra en la Tabla 4-14 un extracto de estos valores para el río Tea a modo de ejemplo para poder ver su distribución.

Año	Cuarto	Conductividad	Oxígeno	Turbidez	Materia orgánica	pH
2018	1	10.103	10.102	1.612	0.375	6.076
	2	25.846	9.466	1.524	0.002	6.248
	3	42.451	7.922	1.594	0.0002	6.458
	4	35.254	8.949	1.857	0.485	6.269
2019	1	29.509	9.896	1.289	0.091	6.282
	2	25.813	9.650	3.314	0.084	6.325
	3	32.372	7.964	1.757	0.020	6.433
	4	25.740	9.654	2.142	0.719	5.935
2020	1	31.379	10.682	2.2021	0.493	6.352
	2	32.410	9.468	1.700	0.707	6.088
	3	48.167	7.740	1.033	1.064	6.360
	4	27.711	9.953	3.261	0.671	6.081

Tabla 4-14. Extracto de valores divididos por cuartos para el río Tea

Una vez se ha obtenido esta tabla, se calcula el CMA como se ha explicado con anterioridad obteniendo una tabla como la Tabla 4-15.

CMA COND	CMA OXIG	CMA TURB	CMA MAT ORG	CMA PH
-	-	-	-	-
-	-	-	-	-
30.842	9.089	1.618	0.167	6.293
33.263	9.081	1.793	0.134	6.329
32.005	9.112	2.038	0.151	6.338
29.551	9.201	2.092	0.190	6.292
28.598	9.395	2.249	0.289	6.256
29.658	9.472	2.150	0.410	6.233

Tabla 4-15. Extracto de valores de CMA correspondientes a los factores del río Tea

Inevitablemente, la fórmula empieza en el tercer valor de la tabla, por lo que los dos primeros valores no tienen un CMA correspondiente a ellos. Esta última tabla se encontraría inmediatamente a la derecha de la Tabla 4-14. El siguiente paso es calcular una relación entre el valor medio de cada cuarto dividido entre su CMA correspondiente, como aparece en la Tabla 4-16.

COND/CMA	OXIG/CMA	TURB/CMA	MAT ORG/CMA	PH/CMA
-	-	-	-	-
-	-	-	-	-
1.386	0.877	0.991	0.625	1.032
1.069	0.984	1.046	3.010	0.997
0.928	1.092	0.632	0.628	0.992
0.878	1.055	1.593	0.432	1.019
1.139	0.853	0.786	0.095	1.034
0.872	1.025	1.008	1.753	0.950

Tabla 4-16. Extracto de la relación de cada valor respecto a su correspondiente CMA

En esta tabla, las dos primeras filas están vacías intencionadamente ya que, al no poder calcularse el CMA de estos primeros cuartos, evidentemente no les corresponden cálculo de la relación respecto al CMA. Teniendo en cuenta que los datos a extraer van a ser por cuartos de año, se ha tratado de agrupar cada cuarto de año y hacer una media de estos (Tabla 4-17).

	Conductividad	pH	Amonio	Materia Orgánica	Turbidez
1er cuarto	0.93	1.02	1.27	1.49	1.45
2º cuarto	0.91	0.99	0.27	0.88	0.76
3er cuarto	1.03	1.00	0.73	0.76	0.56
4º cuarto	1.25	1.10	1.62	1.03	1.40

Tabla 4-17. Resultados finales para el método de estacionalidad

Como se ve en la Tabla 4-17, se pueden distinguir fácilmente los valores “boom” y “bust” comentados con anterioridad.

El análisis de anomalías por descomposición en TEA es una herramienta valiosa para detectar patrones anómalos en los datos relacionados con la calidad del agua y la presencia de cianobacterias. Ayuda a los investigadores a comprender mejor la evolución de estos fenómenos y a tomar medidas adecuadas en caso de que se detecten anomalías.

4.6 Análisis resultados por métodos

Tras obtener resultados para cada uno de los métodos, primero se comparan individualmente para tratar de valorar la eficacia individual de cada uno de estos en comparación con el método de validación cruzada que se utilizará posteriormente. Esto permite extraer conclusiones acerca del método y la

perspectiva apropiada para tratar de afrontar el problema y poder predecir las floraciones con mayor exactitud. Esto se va a realizar únicamente para los ríos Limia y Miño ya que son de los que se disponen datos regulares y diarios y al mismo tiempo hay fechas documentadas por la prensa. Para el río Forcadás como ya se ha comentado no se dispone de datos diarios, y para el Tea no hay fechas documentadas de floraciones de cianobacterias.

4.6.1 Resultados por métodos para el río Limia

Primeramente, esto se llevará a cabo para el río Limia. Se analizan los datos obtenidos para este río por el método de Z-score y se comparan con las fechas confirmadas por la prensa en la Tabla 4-18. En la tabla se observa que se obtienen 4 alertas arrojadas por el método de Z-score, y ninguna de estas coincide con las fechas reales documentadas por la prensa.

Resultados anómalos Z-score	Fechas documentadas
Turbidez semana 57 - 02/2019	30/09/2019
Turbidez semana 67 - 04/2019	03/11/2019
Turbidez semana 103 - 12/2019	28/07/2021
Fosfatos semana 155 - 12/2020	16/09/2022

Tabla 4-18. Comparación resultados Z-score con fechas documentadas para el río Limia

De la misma manera, se repite el proceso para el método de la tercera derivada para este río en la Tabla 4-19. Donde se concluye que para este método hay 4 fechas distintas con resultados inusuales y sin embargo de nuevo, ninguna de estas coincide con ninguna de las fechas documentadas por los periódicos locales.

Resultados anómalos 3ª derivada	Fechas documentadas
Conductividad semana 39 - 09/2018	30/09/2019
Conductividad semana 104 - 12/2019	03/11/2019
Turbidez semana 58 - 02/2019	28/07/2021
Turbidez semana 59 - 02/2019	16/09/2022
Turbidez semana 68 - 04/2019	
Turbidez semana 69 - 04/2019	

Tabla 4-19. Comparación resultados 3ª derivada con fechas documentadas para el río Limia

Por último, para este río, se extraen los resultados por el método de descomposición en estacionalidad y se analizan individualmente, comparándolos con las fechas verificadas en la Tabla 4-20. Para estas fechas, comenzando con la primera fecha que se corresponde al 30/09/2019, donde los elementos prósperos son la conductividad y los fosfatos, estos se repiten en las dos últimas fechas, las del 28/07/ 2021 y del 16/09/2022. Para la floración del 03/11/2019 los valores en auge son los del último cuarto de año que son la conductividad y el amonio.

	Amonio	Conductividad	Turbidez	Fosfatos
1º cuarto	0.80	0.68	1.49	0.14
2º cuarto	0.31	0.49	1.57	0.80
3º cuarto	0.60	1.11	0.76	1.86
4º cuarto	1.50	1.51	0.69	0.85

Tabla 4-20. Resultados por tendencia y estacionalidad del río Limia

4.6.2 Resultados por métodos para el río Miño

Siguiendo con este análisis por métodos, el siguiente río en ser analizado es el río Miño. Se comienza de nuevo observando los resultados presentados por Z-score en la Tabla 4-21. El método indica resultados atípicos en un total de 7 meses distintos para el periodo estudiado, de los cuáles ninguno coincide con las floraciones comprobadas.

Resultados anómalos Z-score	Fechas documentadas
Amonio semana 21 – 05/2018	03/11/2019
Turbidez semana 155 – 12/2020	
Amonio semana 206 - 12/2021	
Amonio semana 262 - 01/2023	
Amonio semana 263 - 01/2023	
Amonio semana 264 - 01/2023	
Amonio semana 265 - 01/2023	
Turbidez semana 260 – 12/2022	
Turbidez semana 262 – 01/2023	
Turbidez semana 263 – 01/2023	
Turbidez semana 264 - 01/2023	
Turbidez semana 265 - 01/2023	
Mat. Orgánica semana 310 - 12/2023	
Mat. Orgánica semana 311 - 12/2023	
Mat. Orgánica semana 313 - 12/2023	

Tabla 4-21. Comparación resultados Z-score con fechas documentadas para el río Miño

Continuando con el análisis de los resultados obtenidos por el método de la tercera derivada en la Tabla 4-22, de nuevo se obtienen una cantidad abundante de resultados aislados, un total de 12 meses distintos. Sin embargo, en este caso sí que se observa lo que se llama un positivo verdadero. Lo que significa que atendiendo únicamente a este método se habría conseguido predecir con éxito una de las dos floraciones documentadas en este río. De la misma manera, se habría obtenido una cantidad total de 11 falsos positivos.

Resultados anómalos 3ª derivada	Fechas documentadas
Conductividad semana 40 – 10/2018	03/11/2019
Conductividad semana 45 – 11/2018	26/10/2021
Conductividad semana 92 -10/2019	
Conductividad semana 98 – 11/2019	
Conductividad semana 104 – 12/2019	
Conductividad semana 105 – 01/2020	
Conductividad semana 190 – 08/2021	
Conductividad semana 206 – 12/2021	
Conductividad semana 249 – 10/2022	
Conductividad semana 250 – 10/2022	
pH semana 207 – 12/2021	
pH semana 208 – 12/2021	
pH semana 209 – 12/2021	
Turbidez semana 156 – 12/2020	
Turbidez semana 265 – 01/2023	
Mat. Orgánica semana 267 – 02/2023	

Tabla 4-22. Comparación resultados 3ª derivada con fechas documentadas para el río Miño

Por último, se realiza el análisis por estacionalidad únicamente para las fechas comprobadas (Tabla 4-23). Al encontrarse ambas fechas confirmadas en el último cuarto del año, se comprobaría que para todos los elementos estudiados esta es una época de floración.

	COND	PH	AMONIO	MAT ORG	TURB
1^{er} cuarto	0.93	1.02	1.27	1.49	1.45
2^o cuarto	0.91	0.99	0.27	0.88	0.76
3^{er} cuarto	1.03	1.00	0.73	0.76	0.56
4^o cuarto	1.25	1.10	1.62	1.03	1.40

Tabla 4-23. Resultados por tendencia y estacionalidad del río Miño

4.7 Resultados aprendizaje por conjuntos

Tras el análisis de cada método por separado, se concluye que el uso de los tres métodos simultáneamente es más beneficioso para el estudio ya que esta validación cruzada que se aplicará más adelante actúa como filtro para minimizar el número de resultados y aumentar la eficacia de los resultados que sí que pasan el filtro de los tres métodos. Por lo que se completa el análisis al comparar los resultados de los distintos métodos y determinando los valores anómalos en los que los métodos concuerdan y viendo qué relación pueden tener estas fechas obtenidas respecto a las fechas de las floraciones de cianobacterias y así poder concluir que ciertos parámetros tienen un impacto directo. O por lo contrario se concluye que estos factores no están directamente relacionados con estas apariciones.

4.7.1 Resultados por conjuntos para el río Limia

Comenzando por el río Limia, en este caso los positivos se han dado en el embalse de As Conchas, mientras que el cuerpo de agua estudiado es el del tramo del río a la altura de A Ponte Liñares (entre Forxa y O Ribeiro), siendo este punto 15 km previos al embalse mencionado. A pesar de la posible presencia de microclimas en esta zona, se considera que las condiciones meteorológicas en la zona estudiada y en el embalse donde se reportó la floración son similares al ser una distancia relativamente corta de 15 km. En la Figura 4-1 se observa la distancia entre ambos puntos.

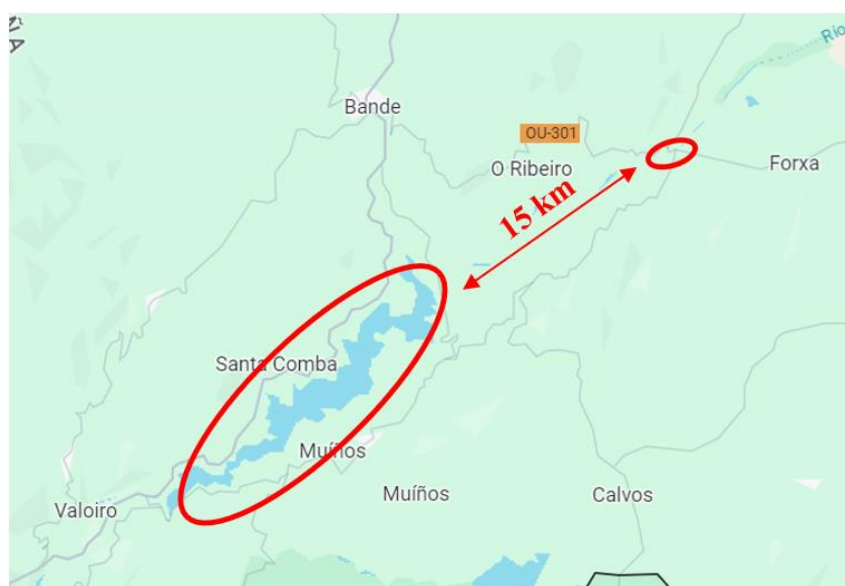


Figura 4-1. Situación geográfica del embalse de As Conchas y de la zona estudiada

Observando los resultados de los métodos se ve que en los dos métodos llaman la atención como resultados anómalos aquellos referidos a la turbidez. Más concretamente, en ambos casos por el método de Z-score se identifica un resultado anómalo y las dos semanas posteriores a esas se identifican como resultados anómalos por el método de la tercera derivada (Tabla 4-24). Esto sucede en tanto en el caso de la semana 57 como en el caso de la semana 67. Estas se ven corroboradas por el análisis por descomposición en tendencia, estacionalidad y aleatoriedad (Tabla 4-25), ya que corresponden al primer y al segundo cuarto del año donde, como se puede ver, para el parámetro de turbidez está en negrita para remarcarlo. Estas dos semanas corresponderían a la semana del 28 de enero del 2019 y a la semana del 08 de abril de 2019.

Resultados anómalos	Z-score	Resultados anómalos	3ª derivada
Turbidez semana 57	4.337	Conductividad semana 39	244.310
Turbidez semana 67	5.783	Conductividad semana 104	106.069
Turbidez semana 103	5.069	Turbidez semana 58	-81.923
Fosfatos semana 155	3.283	Turbidez semana 59	78.999
		Turbidez semana 68	-115.084
		Turbidez semana 69	112.987

Tabla 4-24. Resultados río Limia. Resaltados los elementos coincidentes

	Amonio	Conductividad	Turbidez	Fosfatos
1^{er} cuarto	0.80	0.68	1.49	0.14
2^o cuarto	0.31	0.49	1.57	0.80
3^{er} cuarto	0.60	1.11	0.76	1.86
4^o cuarto	1.50	1.51	0.69	0.85

Tabla 4-25. Resultados por tendencia y estacionalidad del río Limia

Retornando a la tabla en la que se muestran las fechas de floraciones confirmadas por la prensa, la Tabla 4-1, se pueden comparar en la Tabla 4-26 estas con los resultados obtenidos. Se puede observar la fecha de floración de cianobacterias en el río Limia más cercana es el 30 de septiembre de 2019, siendo esta más de cinco meses posterior al resultado anómalo de turbidez obtenido.

Z-score	3ª derivada	TEA	Verificación
Semana 57 (02/2019)	Semanas 58/59 (02/2019)	Auge	Alerta
Semana 67 (04/2019)	Semanas 68/69 (04/2019)	Auge	Alerta

Tabla 4-26. Resumen verificación cruzada resultados río Limia

4.7.2 Resultados por conjuntos para el río Miño

Continuando con el análisis por conjuntos para el río Miño, en esta ocasión se encuentra que hay cierta distancia entre donde se publica que ha sido la floración y donde se ha realizado el estudio ya que el estudio es a la altura de Ourense y la floración se da a la altura de Chantada. Se puede ver en la Figura 4-2 gráficamente esta separación. De nuevo al tratarse de una distancia de unos 50 km, se desestima que las condiciones meteorológicas entre ambos puntos sean muy diferentes por lo que a efectos de toma de muestras se consideran unas condiciones similares.



Figura 4-2. Distancia entre la floración y el cuerpo de agua estudiado en el río Miño

Fijando la atención en la Tabla 4-27, se vislumbran ciertas similitudes con el análisis realizado en el río Limia. A pesar de que aquí se han obtenido más resultados anómalos, coincidentes en ambos métodos solo se han encontrado en el parámetro de la turbidez, como ya sucedió en el estudio del río Limia. De la misma manera, el método de la tercera derivada y el de Z-score coinciden en dos fechas, la de la semana 155/156 y la de la semana 265. Estas fechas corresponden a la semana del 14 de diciembre del 2020 y a la del 23 de enero de 2023. En la tabla de resultados del método de tendencia y estacionalidad (Tabla 4-28) se ven respaldados estos resultados ya que corresponden al primer y último cuartos del año, que se encuentran subrayados para remarcarlos.

Resultados anómalos	Z-score	Resultados anómalos	3ª derivada
Amonio semana 21	6.053	pH semana 208	2.010
Amonio semana 206	4.217	Turbidez semana 156	33.925
Amonio semana 262	4.070	Turbidez semana 265	31.886
Amonio semana 263	5.065	Mat. Orgánica semana 267	6.832
Amonio semana 264	6.740		
Amonio semana 265	5.758		
Turbidez semana 260	3.486		
Turbidez semana 155	3.318		
Turbidez semana 262	4.859		
Turbidez semana 263	3.606		
Turbidez semana 264	6.360		
Turbidez semana 265	4.528		
Mat. Orgánica semana 310	3.446		
Mat. Orgánica semana 311	3.359		
Mat. Orgánica semana 313	3.446		

Tabla 4-27. Resultados del río Miño. En negrita los elementos coincidentes

	COND	PH	AMONIO	MAT ORG	TURB
1^{er} cuarto	0.93	1.02	1.27	1.49	1.45
2^o cuarto	0.91	0.99	0.27	0.88	0.76
3^{er} cuarto	1.03	1.00	0.73	0.76	0.56
4^o cuarto	1.25	1.10	1.62	1.03	1.40

Tabla 4-28. Resultados por tendencia y estacionalidad del río Miño

Haciendo referencia a la Tabla 4-1 se puede ver de nuevo una cierta distancia entre las fechas confirmadas por la prensa de floraciones de cianobacterias en el río Miño y las fechas obtenidas por medio del análisis de datos al compararlas en la Tabla 4-29. La distancia mínima entre estas dos agrupaciones de fechas es de diez meses, entre diciembre de 2020 y octubre de 2021. De nuevo los métodos arrojan unos resultados unos meses antes de la floración de cianobacterias.

Z-score	3ª derivada	TEA	Verificación
Semana 155 (12/2020)	Semana 156 (12/2020)	Auge	Alerta
Semana 265 (01/2023)	Semana 265 (01/2023)	Auge	Alerta

Tabla 4-29. Resumen verificación cruzada resultados río Miño

4.7.3 Resultados por conjuntos para el río Forcadás

El siguiente río para el que se combinan los tres métodos es el río Forcadás. En este caso todas las fechas obtenidas por prensa se dan en el mismo embalse de As Forcadás en el que se han estudiado los componentes del agua. Este se sitúa dónde se indica en la siguiente Figura 4-3.



Figura 4-3. Situación geográfica embalse de As Forcadas

En esta ocasión los métodos de la tercera derivada y de Z-score no coinciden en los resultados anómalos (Tabla 4-30). A pesar de esto, en las fechas del 28/07/2015 y 05/09/2023 tanto la temperatura como el pH dan resultados anómalos por el método de la tercera derivada y al mismo tiempo coincide con que están en auge por el método de descomposición en TEA (Tabla 4-31).

Resultados anómalos Z-score	Resultados anómalos 3ª derivada
Fosfatos 08/03/2010	Temperatura 27/02/2012
Fosfatos 28/02/2011	Temperatura 25/03/2013
Temperatura 08/11/2011	Temperatura 28/07/2015
Fosfatos 18/05/2022	pH 05/09/2023

Tabla 4-30. Tabla de resultados del río Forcadas

	Fosfatos	Nitratos	pH	Temperatura
1 ^{er} cuarto	0.96	2.12	0.91	0.69
2 ^o cuarto	0.92	1.21	1.02	1.13
3 ^{er} cuarto	1.08	0.19	1.11	1.34
4 ^o cuarto	1.03	0.43	0.99	0.95

Tabla 4-31. Resultados por tendencia y estacionalidad del río Forcadas

Al volver a la Tabla 4-1 de fechas confirmadas por la prensa es destacable que en la floración del 23 de julio de 2015 coincide con el resultado anómalo de la temperatura del 28 de ese mismo mes y la floración del 06 de septiembre de 2023 coincide con el resultado anómalo del pH del 05/09/2023, un día antes de dicha noticia. Ambos resultados han sido obtenidos por el método de la tercera derivada y estas fechas son del tercer y último cuarto del año y se ven en la tabla de tendencia y estacionalidad en negrita por lo que al compararlos teniendo en cuenta las irregularidades en la toma de datos de este río, son resultados concluyentes para que salte la alerta en el modelo predictivo (Tabla 4-32).

Z-score	3ª derivada	TEA	Verificación
-	28/07/2015	Auge	Alerta
-	05/09/2023	Auge	Alerta

Tabla 4-32. Resumen verificación cruzada resultados río Forcadas

4.7.4 Resultados por conjuntos para el río Tea

Por último, se analizan los datos del río Tea. En este análisis se observa que no hay ningunas fechas coincidentes entre los métodos como se observa en la Tabla 4-33. Volviendo a la Tabla 4-1, se observa que no hay constancia de floraciones de cianobacterias en el río Tea. Por lo que en este caso se puede decir que la predicción coincidió con la realidad.

Resultados anómalos	Z-score	Resultados anómalos	3ª derivada
Turbidez semana 39	3.675	Turbidez semana 152	24.556
Conductividad semana 142	3.024	Turbidez semana 165	24.888
Mat. Orgánica semana 191	3.383	Turbidez semana 170	24.585
Mat. Orgánica semana 192	3.465	Turbidez semana 171	39.881
Mat. Orgánica semana 193	4.842		
Turbidez semana 251	3.319		
Mat. Orgánica semana 251	3.682		
Turbidez semana 256	4.196		
Mat. Orgánica semana 256	4.390		
Turbidez semana 260	3.381		
Turbidez semana 261	5.094		

Tabla 4-33. Resultados del río Tea

Como se puede ver en la Tabla 4-34 evidentemente hay ciertas épocas en las que sí que se dan auges de ciertos parámetros siendo la época más destacable el último cuarto de año dónde todos los parámetros están en negrita, especialmente la materia orgánica. Sin embargo, en el análisis por Z-score y por la tercera derivada no terminan de coincidir por lo que la verificación cruzada descarta una alarma por proliferación para el río Tea.

	Conductividad	Oxígeno	Turbidez	Mat. Org.	pH
1 ^{er} cuarto	0.90	1.12	1.18	0.79	1.00
2 ^o cuarto	0.91	1.00	0.98	0.63	0.99
3 ^{er} cuarto	1.31	0.88	0.69	0.65	1.05
4 ^o cuarto	1.01	1.14	1.20	1.82	1.07

Tabla 4-34. Resultados por tendencia y estacionalidad del río Tea

4.8 Validación de resultados

Todos estos resultados obtenidos deben validarse y compararse. Gracias a estos tres métodos se han identificado unas posibles fechas en las cuales podría darse un episodio de proliferación de cianobacterias por lo que se deben contrastar con las fechas obtenidas por la prensa en la Tabla 4-1 y comprobar la validez de estos resultados. Comparando ambas tablas se obtiene que:

En el río Tea no hay floraciones publicadas. El análisis de datos tampoco indica ninguna alarma por proliferación de cianotoxinas. Estos resultados se complementan y se validan entre sí.

Para el río Forcadás se ha alarmado de floraciones en dos fechas, siendo estas dos fechas exitosas ya que se confirman con las fechas publicadas en prensa (Tabla 4-35). Sin embargo, de las tres floraciones anteriores en el periodo estudiado no se ha activado la alarma por lo que se consideran falsos negativos.

Fechas por análisis Forcadas	Fechas prensa Forcadas
28/07/2015	23/07/2015
05/09/2023	28/08/2019
	27/07/2020
	21/05/2021
	06/09/2023

Tabla 4-35. Comparación resultados río Forcadas

Para el río Miño se ha alarmado de episodios de cianobacterias en dos ocasiones, en ninguna de esas ocasiones hubo una floración reportada por la prensa. La prensa si reportó las floraciones de noviembre de 2019 y de octubre de 2021, de ninguna se detectó ningún resultado anómalo para alarmar de estos episodios (Tabla 4-36).

Fechas por análisis Miño	Fechas prensa Miño
12/2020	03/11/2019
01/2023	26/10/2021

Tabla 4-36. Comparación resultados río Miño

En el río Limia se dieron dos resultados anómalos relativamente seguidos ya que fueron en febrero y abril de 2019, no hay noticias en prensa acerca de proliferación de cianotoxinas en esa franja de tiempo. Sí las hubo más tarde ese año, curiosamente también seguidos (septiembre y noviembre 2019). En 2021 y 2022 hubo floraciones no alarmadas por el modelo predictivo (Tabla 4-37).

Fechas por análisis Limia	Fechas prensa Limia
02/2019	30/09/2019
04/2019	03/11/2019
	28/07/2021
	16/09/2022

Tabla 4-37. Comparación resultados río Limia

Validando con la prensa se obtiene que por análisis de los datos se ha alarmado en 6 ocasiones, dándose un episodio de cianotoxinas en 2 de estas ocasiones. De las 11 fechas confirmadas de floraciones en las aguas estudiadas se detectó por análisis de datos en estas 2 ocasiones. Por lo que se obtuvieron 4 falsos positivos y 9 falsos negativos. Curiosamente las dos veces que si se ha alarmado con éxito se han dado en el río Forcadas y ambas han sido detectadas por el método de la 3ª derivada conjunto con el análisis por descomposición en TEA. Es aquí donde entra en juego las tasas de positivos verdaderos. Donde se debe hacer uso de la ecuación mencionada en el apartado 3.7: la Ecuación TPV (3-4).

En el caso de los resultados obtenidos se obtiene una Tasa de Positivos Verdaderos de 2 positivos verdaderos dividido entre esos 2 positivos verdaderos sumado a los 9 falsos negativos, el resultado numérico es de 0,182, o lo que viene siendo lo mismo, de un 18,2%.

5 CONCLUSIONES Y LÍNEAS FUTURAS

5.1 Conclusiones

En el inicio de este trabajo se marcaron una serie de objetivos para este proyecto. El primero era recopilar y procesar datos ambientales relevantes para esta investigación. Tras contactar con los organismos responsables de las aguas estudiadas se han obtenido una serie de datos. Dichos datos se han procesado, organizado y proyectado para su posterior análisis. Se han adaptado a cantidades representativas de la muestra que se puedan analizar y se han representado de tal manera que su exposición es más gráfica y conveniente.

El segundo objetivo se refería al propio análisis de datos. Para efectuar este análisis de datos se han utilizado los tres métodos mencionados a lo largo de este documento. El análisis a través del método de z-score ha permitido identificar desviaciones significativas en los datos, señalando posibles anomalías en la turbidez y otras variables relevantes. Si bien estas desviaciones no siempre están directamente relacionadas con la presencia de cianotoxinas, sirven como indicadores tempranos de cambios en las condiciones del agua que podrían preceder a la floración de cianobacterias.

Por otro lado, el método de la tercera derivada ha sido útil para detectar puntos de inflexión y cambios abruptos en la serie temporal de datos. Este enfoque ha revelado momentos críticos en los que parámetros muestran fluctuaciones inesperadas, lo que podría estar asociado con factores ambientales que favorecen el crecimiento de cianobacterias. Por lo tanto, se puede obtener información valiosa sobre la dinámica de los cambios en la calidad del agua a lo largo del tiempo.

El último método, la descomposición por TEA ha permitido desglosar la serie temporal en componentes fundamentales, facilitando la identificación de patrones estacionales y tendencias a largo plazo. Este análisis ha revelado cambios estacionales que podrían estar relacionados con la actividad biológica en el cuerpo de agua, incluida la proliferación de cianobacterias durante ciertas épocas del año.

A través de la aplicación de estos métodos, se buscaba encontrar relaciones significativas entre variables ambientales y la presencia de cianobacterias, proporcionando una base sólida para futuras investigaciones y estrategias de gestión. El uso conjunto de estos tres métodos al mismo tiempo proporciona una verificación cruzada que ofrece una fiabilidad mayor en el análisis de estos datos.

El último objetivo está relacionado con la evaluación de la eficacia de estos métodos para predecir los episodios de cianotoxinas corroborando las fechas obtenidas con la prensa. Esto se da en el último apartado de “Resultados” donde se validan los resultados con las fechas confirmadas obtenidas por prensa, así comparando y observando la eficacia de los resultados.

En cuanto a los resultados se puede concluir que no han sido del todo fiables ya que tan solo 2 de 11 floraciones han sido detectadas con éxito y ha dado 4 falsos positivos de floración cuando no han

sido reportados. La Tasa de Positivos Verdaderos es de un 18,2%, no termina de ser un resultado favorable a la hora de evaluar la fiabilidad de un método, por lo que no se puede concluir que los resultados anómalos de estos parámetros para los cuerpos de agua estudiados ofrezcan una gran fiabilidad a la hora de predecir los episodios. Aunque se puede sacar como nota positiva a este estudio la observación de la influencia meteorológica sobre este fenómeno ya que todos los episodios de cianotoxinas publicados se dieron entre los meses de julio y octubre prácticamente.

De todas formas, cabe recordar que en este estudio se depende también de la prensa, como ya se ha comentado con anterioridad. Es posible que haya habido floraciones que no han quedado documentadas, de esta manera influenciando los resultados claramente.

Además, se puede considerar el uso de este método para la alarma de estos episodios de proliferaciones para poder al menos estar preparado para la toma de unas reacciones iniciales para que en caso de que la floración finalmente sí que se dé, se pueda actuar de una forma más eficiente que si esta floración apareciese de forma sorpresiva.

5.2 Líneas futuras

Para avanzar en la comprensión de estos fenómenos, se sugiere una investigación más profunda que aborde cada método en detalle y explore su aplicabilidad en la predicción de episodios de cianotoxinas. Esto podría incluir la refinación de los algoritmos utilizados en cada método y la evaluación de su precisión y robustez en diferentes contextos hidrológicos.

Además, sería positivo examinar la interacción entre los diferentes parámetros analizados, como la temperatura del agua, la concentración de nutrientes y la velocidad del flujo, para comprender mejor los factores que influyen en la aparición de cianobacterias. Esto podría implicar la realización de experimentos controlados en entornos de laboratorio o el desarrollo de modelos computacionales que integren múltiples variables.

Por último, seguir con estas líneas de investigación no solo contribuirá al avance del conocimiento científico en el campo de la hidrología, sino que evidentemente también tendrá implicaciones prácticas importantes para la gestión y conservación de los recursos hídricos. La detección temprana y la predicción precisa de episodios de cianotoxinas son cruciales para proteger la salud humana y el medio ambiente, y este estudio proporciona una base para futuras investigaciones en esta área crítica.

6 BIBLIOGRAFÍA

- [1] W. Carmichael, «The Toxins of Cyanobacteria», *Sci. Am.*, vol. 270, pp. 78-86, feb. 1994, doi: 10.1038/scientificamerican0194-78.
- [2] D. Millet, «Cianobacterias: todas las respuestas», *La Provincia - Diario de Las Palmas*. Accedido: 30 de marzo de 2024. [En línea]. Disponible en: <https://www.laprovincia.es/sociedad/2017/08/14/cianobacterias-respuestas-9650264.html>
- [3] S. Cires Gómez, *Catálogo de cianobacterias planctónicas potenciales productoras de cianotoxinas*.
- [4] Lorena Rosso y Leda Giannuzzi, «Factores ambientales y antropogénicos que afectan la formación de floraciones de cianobacterias y cianotoxinas».
- [5] T. Gunther, «Floraciones de Cianobacterias: causas, riesgos y tratamiento», *LG Sonic*. Accedido: 30 de marzo de 2024. [En línea]. Disponible en: <https://www.lgsonic.com/es/cianobacterias/>
- [6] «El embalse de As Forcadas recupera agua y llega el 65% de ocupación», *Diario de Ferrol*. Accedido: 30 de marzo de 2024. [En línea]. Disponible en: <https://www.diariodeferrol.com/articulo/ferrol/embalse-as-forcadas-recupera-agua-llega-65-ocupacion-4064826>
- [7] Nodd3r, «¿Análisis de datos en tiempo real? Descubre su importancia y ventajas.» Accedido: 30 de marzo de 2024. [En línea]. Disponible en: <https://nodd3r.com/blog/analisis-de-datos-en-tiempo-real-descubre-su-importancia-y-ventajas>
- [8] K. Rojas-Jimenez, *Capítulo 8 Análisis de Series de Tiempo | Ciencia de Datos para Ciencias Naturales*. Accedido: 3 de julio de 2024. [En línea]. Disponible en: https://bookdown.org/keilor_rojas/CienciaDatos/an%C3%A1lisis-de-series-de-tiempo.html
- [9] «Tipos de series temporales - Definición, qué es y concepto | Economipedia». Accedido: 3 de abril de 2024. [En línea]. Disponible en: <https://economipedia.com/definiciones/tipos-de-series-temporales.html>
- [10] E. Cepeda Cuervo, J. A. Achcar, y M. G. Andrade, «Series de Tiempo hidrológicas y meteorológicas estacionales», *Earth Sci. Res. J.*, vol. 22, n.º 2, pp. 83-90, jun. 2018, doi: 10.15446/esrj.v22n2.65577.
- [11] C. de I. y D. T. del A. Universidad de Salamanca, «CAPÍTULO 7 ALGAS Y CIANOBACTERIAS EN AGUAS DULCES».
- [12] E. Lucena, «Aspectos sanitarios de las cianotoxinas», 2008.

- [13] gmaldonado, «Lago San Roque: Ambiente realiza tareas de remediación para frenar el ingreso de cianobacterias», Córdoba Ahora. Accedido: 30 de marzo de 2024. [En línea]. Disponible en: <https://cordobaahora.com/lago-san-roque-ambiente-realiza-tareas-de-remediacion-para-frenar-el-ingreso-de-cianobacterias/>
- [14] Aquanova, «¿Qué es el DBO y DQO? - Análisis de Agua», Aquanova. Accedido: 30 de marzo de 2024. [En línea]. Disponible en: <https://www.aquanova.es/2020/04/que-es-el-dbo-y-dqo/>
- [15] «La Manga gallega está en As Conchas, el embalse de A Limia intoxicado por la ganadería intensiva que promueve Coren | Público». Accedido: 3 de abril de 2024. [En línea]. Disponible en: <https://www.publico.es/sociedad/manga-gallega-as-conchas-embalse-limia-intoxicado-ganaderia-intensiva-promueve-coren.html>
- [16] «Estudios hidrológicos - HidrojING». Accedido: 3 de abril de 2024. [En línea]. Disponible en: <https://www.hidrojing.com/servicios/estudios-hidrológicos/>
- [17] E. Fernández, «Análisis de Cianobacterias y Cianotoxinas en Cuerpos de Agua», TSI Group - Tecnosoluciones Integrales. Accedido: 3 de abril de 2024. [En línea]. Disponible en: <https://tecnosolucionescr.net/blog/547-analisis-de-cianobacterias-y-cianotoxinas-en-cuerpos-de-agua>
- [18] «ESTADO DE LA TAXONOMÍA DE CIANOBACTERIAS; EL EJEMPLO DE STIGONEMATALES». M VERGARA MARTÍNEZ
- [19] Gloriselle Negrón Ríos, «Microcistinas».
- [20] «¿Qué es la saxitoxina que provocó la veda en el Golfo de California? - Ensenada.Net». Accedido: 3 de abril de 2024. [En línea]. Disponible en: <https://www.ensenada.net/noticias/nota.php?id=40819>
- [21] redaccion iAgua, «Colombia invierte en proyectos de diagnóstico de toxinas en aguas de riego de cultivo», iAgua. Accedido: 3 de abril de 2024. [En línea]. Disponible en: <https://www.iagua.es/noticias/universidad-nacional-colombia/colombia-invierte-proyectos-diagnostico-toxinas-aguas-riego>
- [22] M. F. Abdallah, W. H. R. Van Hassel, M. Andjelkovic, A. Wilmotte, y A. Rajkovic, «Cyanotoxins and Food Contamination in Developing Countries: Review of Their Types, Toxicity, Analysis, Occurrence and Mitigation Strategies», *Toxins*, vol. 13, n.º 11, Art. n.º 11, nov. 2021, doi: 10.3390/toxins13110786.
- [23] «Cianobacterias, el peligro verde que nos espera en el agua – Diario El Telégrafo». Accedido: 3 de abril de 2024. [En línea]. Disponible en: <https://www.eltelegrafo.com/2019/12/cianobacterias-el-peligro-verde-que-nos-espera-en-el-agua/>
- [24] E. A. Cantoral Uriza, A. D. Asencio Martínez, M. Aboal Sanjurjo, E. A. Cantoral Uriza, A. D. Asencio Martínez, y M. Aboal Sanjurjo, «Cianotoxinas: efectos ambientales y sanitarios. Medidas de prevención», *Hidrobiológica*, vol. 27, n.º 2, pp. 241-251, ago. 2017.
- [25] «GALICIA», Ministerio para la Transición Ecológica y el Reto Demográfico. Accedido: 25 de marzo de 2024. [En línea]. Disponible en: <https://www.miteco.gob.es/es/agua/temas/saneamiento-depuracion/actuaciones/galicia.html>
- [26] «Augas de Galicia adjudica por 1,5 millones de euros un sistema para optimizar el consumo energético y garantizar la calidad del agua vertida en las EDAR | Innovaugas». Accedido: 3 de abril de 2024. [En línea]. Disponible en: <https://innovaugas.xunta.gal/es/actualidad/augas-de-galicia-adjudica-por-15-millones-de-euros-un-sistema-para-optimizar-el-consumo>
- [27] «Portada - Augas de Galicia». Accedido: 25 de marzo de 2024. [En línea]. Disponible en: <https://augasdeg Galicia.xunta.gal/>
- [28] Ana Elizabeth Flores González, «Práctica 2: Mediciones y análisis estadístico de datos».
- [29] «Campana de Gauss | Blog MBA Cámara de Comercio de Málaga». Accedido: 1 de abril de 2024. [En línea]. Disponible en: <https://www.master-malaga.com/master-mba/campana-de-gauss/>
- [30] Pértegas Díaz S. y Pita Fernández S., «La distribución normal».
- [31] Natalia Ruiz Fuentes, «Una aproximación al análisis exploratorio de datos».
- [32] Carlos Eugenio Tapia Ayuga, «Aplicación de las terceras derivadas a procesos astrofotográficos».
- [33] Flores, J., «Data analysis: An introduction to applied statistics».

- [34] Ignacio Arnott Iglesias, «Análisis y detección de anomalías usando técnicas de deep learning: case de estudio en cotizaciones de Pfizer y Moderna».
- [35] «ÍNDICE DE VARIACIÓN ESTACIONAL». Accedido: 2 de julio de 2024. [En línea]. Disponible en: http://www.ub.edu/aplica_infor/spss/cap8-4.htm
- [36] «Vuelve al embalse de Vilasouto la contaminación por cianobacterias», La Voz de Galicia. Accedido: 3 de julio de 2024. [En línea]. Disponible en: https://www.lavozdegalicia.es/noticia/lemos/2022/08/30/vuelve-embalse-vilasouto-contaminacion-cianobacterias/0003_202208M30C1991.htm
- [37] C. S. Hong y T. G. Oh, «TPR-TNR plot for confusion matrix», *Commun. Stat. Appl. Methods*, vol. 28, n.º 2, pp. 161-169, mar. 2021, doi: 10.29220/CSAM.2021.28.2.161.
- [38] RedacciónT21, «El embalse Las Conchas de Orense produce toxinas de guerra química • Tendencias21», Tendencias21. Accedido: 1 de abril de 2024. [En línea]. Disponible en: https://tendencias21.levante-emv.com/el-embalse-las-conchas-de-orense-produce-toxinas-de-guerra-quimica_a45468.html
- [39] Ourense, «Un estudio revela que todos los embalses de la cuenca Miño-Sil presentan cianobacterias», Faro de Vigo. Accedido: 1 de abril de 2024. [En línea]. Disponible en: <https://www.farodevigo.es/ourense/2019/11/03/estudio-revela-embalses-cuenca-mino-15496115.html>
- [40] A.F, «Alertan de un alto nivel de cianobacterias en el embalse de As Conchas, en Bande», Faro de Vigo. Accedido: 1 de abril de 2024. [En línea]. Disponible en: <https://www.farodevigo.es/ourense/2021/07/28/alertan-alto-nivel-cianobacterias-embalse-55559033.html>
- [41] Ministerio de Sanidad, «Informe nacional calidad de las agua de baño. Informe anual 2022.»
- [42] «Aparecen manchas verdes en el agua del Miño entre Chantada y O Saviñao», La Voz de Galicia. Accedido: 1 de abril de 2024. [En línea]. Disponible en: https://www.lavozdegalicia.es/noticia/lemos/o-savinao/2021/10/27/aparecen-manchas-verdes-agua-mino-chantada-savinao/0003_202110M27C3991.htm
- [43] «El embalse de As Forcadas está en alerta por proliferación de cianobacterias desde el 20 de mayo», Diario de Ferrol. Accedido: 4 de abril de 2024. [En línea]. Disponible en: <https://www.diariodeferrol.com/texto-diario/mostrar/2397378/embalse-as-forcadas-esta-alerta-proliferacion-cianobacterias-desde-20-mayo>
- [44] «Un estudio sobre contaminación de embalses no incluye el Umia». Accedido: 1 de abril de 2024. [En línea]. Disponible en: https://www.lavozdegalicia.es/noticia/pontevedra/2019/08/28/estudio-sobre-contaminacion-embalses-incluye-umia/0003_201908P28C1992.htm
- [45] «La presa del Umia es una de las tres en alerta por cianobacterias», La Voz de Galicia. Accedido: 1 de abril de 2024. [En línea]. Disponible en: https://www.lavozdegalicia.es/noticia/pontevedra/2020/07/27/presa-umia-tres-alerta-cianobacterias/0003_202007P27C5991.htm
- [46] D. Domínguez, «Ganadería, agricultura y erosión amenazan la calidad del agua de doce embalses», Faro de Vigo. Accedido: 1 de abril de 2024. [En línea]. Disponible en: <https://www.farodevigo.es/galicia/2021/05/21/ganaderia-agricultura-erosion-amenazan-calidad-52068529.html>
- [47] «Vecinos de Valdoviño se quejan del olor pestilente del embalse de As Forcadas», La Voz de Galicia. Accedido: 4 de abril de 2024. [En línea]. Disponible en: <https://www.lavozdegalicia.es/noticia/ferrol/valdovino/2023/09/06/vecinos-valdovino-quejan-olor-pestilente-embalse-as-forcadas/00031694017506025691238.htm>

- [48] «Histórico de la red meteorológica - MeteoGalicia». Accedido: 3 de julio de 2024. [En línea]. Disponible en:
https://www.meteogalicia.gal/observacion/estacionshistorico/historico.action?idEst=19008&requiest_locale=es

ANEXO I: IMPLICACIONES SOCIALES, ECONÓMICAS Y AMBIENTALES

El estudio sobre episodios de cianotoxinas no solo tiene relevancia en el ámbito científico, sino que también tiene importantes implicaciones en la sociedad, la economía y el medio ambiente.

6.1 Sociales

Como ya se ha mencionado en repetidas ocasiones a lo largo de este trabajo, las cianotoxinas representan una amenaza evidente. La exposición a estas toxinas puede causar una variedad de efectos adversos, que van desde irritación en la piel hasta problemas más graves como daño hepático y neurotoxicidad. Esto plantea preocupaciones significativas para las comunidades que dependen del agua para consumo humano, recreación y actividades agrícolas.

Además, los episodios de proliferación de cianobacterias pueden afectar la calidad del agua potable, lo que puede resultar en costos adicionales para los sistemas de tratamiento de agua y para los consumidores que deben recurrir a alternativas más costosas para obtener agua segura. Además, es evidentemente un apartado importante que merece una gestión ejemplar por parte del gobierno y de las administraciones correspondientes.

6.2 Económicas

Los impactos económicos de los episodios de cianotoxinas pueden ser significativos. Por un lado, las empresas que dependen del agua para sus operaciones, como las industrias agrícola, pesquera y turística, pueden experimentar pérdidas debido a la contaminación del agua y la disminución de la demanda de productos relacionados con el agua.

Por otro lado, los costos asociados con el monitoreo, tratamiento y remedio de la proliferación de cianobacterias pueden ser elevados para los gobiernos y las agencias encargadas de la gestión de recursos hídricos. Estos costos adicionales pueden repercutir en los presupuestos públicos y en última instancia en la población. Por lo que parece claro que estos episodios tienen un impacto económico sobre la sociedad.

6.3 Ambientales

Uno de los impactos más claros de la proliferación de las cianobacterias, ya que los episodios pueden tener impactos negativos en los ecosistemas acuáticos. La proliferación de estas algas puede provocar la eutrofización de los cuerpos de agua, agotando el oxígeno y afectando la biodiversidad local. Además, las cianotoxinas, contaminando el agua afectan a los organismos acuáticos, incluidos peces, crustáceos y aves.

Estos impactos pueden tener efectos cascada en los ecosistemas terrestres y marinos que dependen de los recursos acuáticos. Por ejemplo, la disminución de la población de peces debido a la contaminación del agua puede afectar a las industrias pesqueras y a la seguridad alimentaria de las comunidades que dependen de ellas.

En conclusión, el análisis de datos en la predicción de episodios de cianotoxinas parece crucial no solo para comprender y prevenir los riesgos que conllevan, sino que también para abordar los desafíos sociales y económicos asociados con estas floraciones. Un enfoque que considere las implicaciones

sociales, económicas y ambientales es fundamental para desarrollar estrategias efectivas de gestión y mitigación de los impactos de la proliferación de cianobacterias.

ANEXO II: REFLEXIONES ÉTICAS Y SOCIALES

El análisis de para la predicción de episodios de cianotoxinas no solo ofrece un acercamiento científico a un problema ambiental de relevancia, sino que también invita a reflexionar sobre diversas consideraciones éticas y sociales que deben ser contempladas en esta área de investigación.

Desde un punto de vista ético, es esencial asegurar que la recopilación y el manejo de los datos se lleven a cabo de manera ética y transparente. Esto implica seguir rigurosamente los protocolos establecidos para obtener la información, respetando la confidencialidad y la privacidad de los datos cuando sea necesario. Además, se debe evitar cualquier forma de manipulación o distorsión de los datos que pueda comprometer la integridad del estudio.

En cuanto al uso de los resultados del análisis, es esencial que estos se utilicen de manera responsable y en beneficio del bien común. Los resultados obtenidos en esta investigación deben ser compartidos de manera abierta y accesible, asegurando que la información esté disponible para todos los interesados y que se tomen medidas adecuadas para abordar los problemas identificados.

Desde una perspectiva social, es importante considerar cómo este análisis puede afectar a las comunidades locales y a las personas que dependen del agua para su sustento y bienestar. Los episodios de cianotoxinas pueden tener consecuencias significativas en el sustento de las personas, especialmente en áreas donde la pesca y la agricultura son actividades económicas importantes. Por lo tanto, es crucial tener en cuenta el impacto social de estos eventos y trabajar en colaboración con las comunidades afectadas para encontrar soluciones efectivas.

Además, es necesario prestar especial atención a la publicación de estos análisis a la población. Una comunicación clara y comprensible es esencial para asegurar que las personas estén informadas sobre los riesgos asociados con los episodios de cianotoxinas y puedan tomar medidas para proteger su bienestar.

En resumen, el análisis de para la predicción de episodios de cianotoxinas plantea importantes desafíos éticos y sociales que deben ser afrontados con sensibilidad y responsabilidad. Al considerar estas reflexiones y trabajar en colaboración con todas las partes interesadas, se puede garantizar que el uso de los datos contribuya positivamente a la protección del medio ambiente y al bienestar de las todos los afectados.