Universidad de la República

Facultad de Ciencias

Programa de Posgrado en Ciencias Ambientales

Implementación de modelos predictivos como sistema de alerta temprana para la gestión de contaminación fecal en playas de uso recreativo de Montevideo

Autora

Victoria Vidal

(vvidalmadalena@gmail.com)

Orientador

Angel Segura

Tribunal

Dr. Pablo Muniz, Dr. Julio Gómez y Dra. Beatriz Yannicelli

2024

Tesis de Maestría en Ciencias Ambientales

RESUMEN

La contaminación por coliformes fecales (CCF) es un problema frecuente que afecta la calidad de las playas y por lo tanto el uso recreativo de dichos espacios. El desarrollo de modelos predictivos de CCF surge como un complemento de los sistemas de monitoreo tradicionales y su implementación en Sistemas de Alerta Temprana (SAT) es recomendada para la gestión de la calidad de playas recreativas. El objetivo general de esta tesis fue generar aportes para la implementación de un SAT en las playas monitoreadas por la Intendencia de Montevideo (IM). Para ello se realizó: i) una revisión de artículos académicos (desde 2000 al 2024) con información acerca de la estrategia de modelización de CCF y su desempeño. Además, se evaluó la implementación de modelos predictivos en SAT. ii) se construyeron modelos predictivos de Aprendizaje Automático (AA) para las playas de Montevideo, incluyendo variables satelitales y iii) se construyó una aplicación web en donde se implementó el modelo con mejor desempeño como una prueba de concepto. En la revisión de artículos científicos se registraron 67 artículos de predicción de CCF, donde la mayoría se desarrolló para sistemas continentales (67%) respecto a sistemas marinos y estuarinos (33%). Los modelos de Regresión Lineal Múltiple fueron los más utilizados en número. Sin embargo, los modelos de mejor desempeño fueron los de AA, incluyendo las Redes Neuronales Artificiales y modelos basados en árboles de decisión (CART y Random Forest: RF). Se identificó la necesidad de implementar técnicas que permitan lidiar con el desbalance de los datos para mejorar la capacidad predictiva de los modelos y la escasa implementación modelos predictivos en SAT. Se construyeron RF para las playas de Montevideo que presentaron buena capacidad predictiva, mejorando las métricas de la línea de base actual y similar a los modelos construidos previamente. Las variables satelitales fueron importantes para la predicción de CCF. La implementación en la interfaz web permite acceder a la base de datos histórica de la IM y a las predicciones de los modelos en mapas interactivos, lo que significa un gran avance para su implementación. La implementación de modelos predictivos en el sistema de gestión de playas de la IM representaría el primer sistema de SAT para Latinoamérica y entre los pocos a nivel mundial que incluya modelos de AA.

Palabras clave: Contaminación fecal, Playas recreativas, Modelos Predictivos, Aprendizaje Automatizado, Información Satelital, Sistemas de Alerta Temprana.

AGRADECIMIENTOS

A Angel por haberme impulsado a hacer esta tesis desde un principio, por su buena disposición, por estar presente, por su pragmatismo en los momentos de más confusión y por su paciencia.

Gracias a los miembros del tribunal por haberse tomado el tiempo de leer la tesis en detalle y hacer comentarios que me ayudaron a pensar más en profundidad el tema de la tesis y todo el proceso de implementación de los modelos.

Gracias a la Unidad de Calidad de Agua de la Intendencia de Montevideo por haber disponibilizado los datos que hicieron esta tesis posible y por su apertura en los intercambios sobre la dinámica de las playas y la implementación de los modelos.

Gracias a la ANII por haber financiado esta maestría.

Al CURE, Sede Rocha y la Universidad de la República por el espacio de trabajo.

Gracias a los compañeros del MEDIA por haber leído los disparates que escribo, haber aportado siempre con palabras de aliento y criticado constructivamente siempre con respeto.

Gracias a mis amigos que hicieron que este difícil momento de mi vida haya sido más ameno.

Gracias a mis padres por haber co-financiado esta tesis junto con la ANII. Por haber creído en mi cuando no tenía laburo, cuando no tenía mucha perspectiva de lo que iba a hacer en el futuro y ni donde, ni cómo. Mi motivación (además de criticar el capitalismo) siempre será verlos felices y orgullosos a ustedes.

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

test. N es el número de modelos incluidos en cada tipo de análisis. La caja representa el rango intercuartil, la barra horizontal la mediana y los bigotes se extienden 1.5 del rango intercuartil.

Figura 3.1 Mapa del departamento de Montevideo. En recuadros negros se indica la grilla espacial seleccionada para las variables satelitales (a modo de ejemplo se ilustra la grilla seleccionada para la variable Kd 490). En puntos rojos numerados se muestran las 20 playas utilizadas para el análisis de datos y la modelización: **Al Oeste**: 1- Punta Espinillo, 2- La Colorada, 3- Pajas Blancas, 4- Zabala, 5- Punta Yeguas, 6- Santa Catalina, 7- Nacional, 8- Cerro. **Al Este**: 9- Ramírez, 10- Pocitos, 11- Puerto del Buceo, 12- Buceo, 13-Malvín, 14- Brava, 15- Honda, 16- ingleses, 17- Verde, 18- Mulata, 19- Carrasco, 20- Miramar.

Figura 3.2 Funcionamiento del algoritmo de Bosque Aleatorio o Random Forest (RF). Se generan N muestras boostrap (una técnica estadística para obtener muestras de una población donde una observación se puede considerar en más de una muestra) del set de datos. Para cada muestra se genera un árbol y en cada separación de los nodos se utiliza sólo una porción *m* de las *p* variables predictoras Las salidas de todos los árboles se combinan en una salida final Y que se obtiene mediante alguna regla (generalmente el promedio, en RF de regresión y, conteo de votos, en RF de clasificación).

Figura 3.3 Diagrama con las partes fundamentales de una interfaz web construida con el paquete Shiny R. La interfaz se construye con dos partes: un código (Ui) que diseña lo que el usuario puede observar de la aplicación, y otro código (Server) que tiene las funciones reactivas que permiten que la aplicación sea interactiva.

Figura 3.4 Dinámica temporal del logaritmo de coliformes fecales (UFC/100ml) en el periodo de estudio (del 15 de noviembre de 2009 al 3 de marzo de 2023) de playas representativas de las diferentes zonas de la Montevideo. A) Playa Pocitos, B) Playa Ramirez y C) Playa del Cerro. Se seleccionaron tres estas playas debido a que presentaron mayor cantidad de datos que el resto. En una línea negra punteada se muestra la Media Móvil del logaritmo de CF. En una línea roja horizontal se muestra el umbral de CF permitido por la normativa.

Figura 3.5 Gráficos de series temporales de A) temperatura superficial del agua para la zona Este (SST, °C), y B) turbidez satelital (Kd490) para el periodo de estudio (del 15 de noviembre de 2009 al 3 de marzo de 2023), en puntos rojos se muestra la turbidez para la zona Oeste (Turbidez_w) y en puntos azules la turbidez para la zona Este (Turbidez_E). La línea punteada roja muestra la media móvil de la turbidez del Oeste (Mm_w) y la línea punteada azul muestra la media móvil de la turbidez del Este (Mm_E).

Figura 3.6 Gráficos de la variación estacional de la turbidez satelital (Kd490) para la zona Este (izquierda) y Oeste (derecha) para el periodo de estudio.

Figura 3.7 Se muestran las correlaciones entre las variables satelitales A) temperatura superficial del agua, B) Kd490 (el modelo de regresión segmentado se muestra con una línea roja) y C) velocidad del viento (WS, por su acrónimo en inglés) entre zona Este y Oeste.

Figura 3.8 Relación entre las variables satelitales y las variables *in situ* tanto para la zona Este a la izquierda y Oeste a la derecha. A) correlación entre la temperatura in situ y la temperatura superficial del agua satelital (SST) B) correlación entre la turbidez in situ y la turbidez satelital (Kd490) y C) correlación entre la velocidad del viento de estaciones meteorológicas y la velocidad del viento satelital (WS, por su acrónimo en inglés).

Figura 3.9 Importancia de variables del RF según la disminución del índice de Gini para el Modelo 1, Modelo 2 y Modelo 3. Las variables fueron la Playa que corresponde al sitio de muestreo, Valor de CF de un muestreo anterior (Lag_cf1), Turbidez *in situ*, Temperatura *in situ*

(Temp. *in situ*), Valor de CF de dos muestreos anteriores (Lag_cf2), Valor de CF de tres muestreos anteriores (Lag_cf2), Kd490 de zona Este (Kd490_e), Kd490 de zona Oeste (Kd490_w), Temperatura Superficial del Agua (SST), Salinidad, Precipitaciones acumuladas en 24hs (Pp24hs), Precipitaciones acumuladas en 48hs (Pp48hs), Precipitaciones de INIA (Pp_inia).

Figura 3.10 Boxplot de observados versus predichos para los tres mejores modelos testeados en este trabajo. En el eje de las "x" están las categorías a predecir "Excede" o "No Excede" y en el eje de las "y" los valores de contaminación fecal (UCF/100ml) del set de datos con los cuales se testeó el modelo en escala logarítmica. En una línea roja horizontal se muestra el umbral de CCF permitido por la normativa.

Figura 3.11 Aspecto general de la interfaz web creada con Shiny R. En la imagen se muestran las primeras dos pestañas de 1) presentación de la aplicación, 2) Instrucciones de uso para la aplicación.

Figura 3.12 Aspecto general de la pestaña en donde pueden visualizarse los niveles de coliformes fecales (UFC/1000ml, la leyenda en colores muestra en un gradiente de amarillo claro a rojo oscuro los valores más altos) en un mapa para la fecha seleccionada en cada playa monitoreada por la Intendencia de Montevideo en interfaz web creada con Shiny R. Los datos de coliformes fecales observados en los mapas provienen de la base de datos histórica de los monitoreos de calidad de playas de la IM.

Figura 3.13 Aspecto general de la pestaña en donde se muestra un gráfico puede visualizarse la evolución temporal de la variable seleccionada en la ventana temporal seleccionada para cada playa monitoreada por la Intendencia de Montevideo.

Figura 3.14 Aspecto general de la pestaña en donde se muestra una tabla en la que pude filtrarse cada columna de modo de acceder rápidamente a un día seleccionado, una playa, o un valor de coliformes en particular para cada playa monitoreada por la Intendencia de Montevideo.

Figura 3.15 Aspecto general de la pestaña en donde se muestran predicciones en mapas en la interfaz web creada con Shiny R.

LISTA DE ABREVIATURAS

AA: Aprendizaje automático

ANN: Redes Neuronales Artificiales

BN: Redes Bayesianas

BGM: Modelo de decisión Bagging

BLR: Regresión Logística Binaria

BDT: Árbol de decisión Boosting

CART: Árboles de regresión o clasificación

CT: Coliformes totales

CF: Coliformes fecales

CCF: Contaminación por coliformes fecales

ENT: Enterococos

FIB: Bacterias Indicadoras Fecales

GBM: Modelo Boosted Generalizado

IM: Intendencia de Montevideo

LME: Efectos lineales Mixtos

MLR: Regresión Lineal Múltiple

MLogR: Regresión Logística Multinomial

RF: Random Forest

OMS: Organización Mundial de la Salud

OLS: Regresión de Mínimos Cuadrados Ordinarios

PLS: Regresión de Mínimos Cuadrados Parciales

PNS: Plan Nacional de Saneamiento

USEPA: United States Environmental Protection Agency

SAT: Sistemas de Alerta Temprana

SECCA: Servicio de Evaluación de la Calidad y Control Ambiental

SVM: Máquinas de Vectores de Soporte

WA: Análisis Wavelet

CAPÍTULO 1

1.1 INTRODUCCIÓN GENERAL

El crecimiento demográfico de las zonas costeras, junto con un tratamiento inadecuado de las aguas residuales, han impactado negativamente la calidad del agua, aumentando la cantidad de contaminantes (Bae et al., 2010). La contaminación por coliformes fecales (CCF) es uno de los principales y más frecuentes problemas ambientales en las zonas costeras (Mallin et al., 2000). El origen y la concentración de la CCF varía ampliamente y se vierte a los cuerpos de agua por deposición directa, contaminación difusa, descargas de efluentes contaminados o por la llegada de contaminación proveniente de la napa freática (Shanks et al., 2006; Kang et al., 2010). Estos fenómenos afectan la calidad de las playas y por lo tanto el uso recreativo de dichos espacios en todo el mundo (He y He, 2008; Molina et al., 2014). Este problema, es particularmente frecuente en zonas sin saneamiento adecuado, sobre el cual es necesario desarrollar medidas de gestión que reduzcan los riesgos sanitarios (Organización Mundial de la Salud, 2003; 2021).

El uso recreativo de playas con niveles elevados de CCF supone un riesgo para la salud humana, ya que aumenta el riesgo de contraer enfermedades de distinta índole (tifoidea, cólera, diarreas, gastroenterocolitis, entre otras) (Shuval, 2003; Wade et al., 2006; Hlavsa et al., 2011; Sabino et al., 2014). La concentración de bacterias indicadoras fecales (FIB por su sigla en inglés), incluidos los coliformes totales (CT), coliformes fecales (o *Escherichia coli*) (CF), enterococos (ENT), los géneros Klebsiella, Enterobacter, y Citrobacter, son indicadores de riesgos potenciales para la salud que se utilizan para identificar niveles orientativos y clasificar a las playas en diferentes niveles de riesgo, siguiendo la recomendación de la Organización Mundial de la Salud (OMS). Los monitoreos microbiológicos de la calidad del agua permiten seguir la evolución del estado sanitario de las playas para desarrollar medidas de gestión de riesgos a la salud humana (Agencia de Protección Ambiental de EE.UU., 2010).

En Uruguay la CCF es un problema que se registra en las playas, principalmente en aquellas localidades costeras sin acceso al saneamiento, aunque también se registra en zonas con acceso a una red de saneamiento tradicional (Soumastre, 2016; Kruk et al., 2018; de León, 2019; Segura et al., 2021; Echeverriborda et al.,2022). Uruguay cuenta con un Plan Nacional de Saneamiento (PNS), en el cual se presentan los distintos tipos de saneamiento que existen en el país, los distintos porcentajes de cobertura y las

proyecciones de cobertura de saneamiento para las distintas localidades del país (PNS, 2020). En el caso de Montevideo, este departamento cuenta con la mayor cobertura de saneamiento centralizado (85%). Es un sistema compuesto de redes colectivas, de carácter dinámico, con tratamiento y disposición final de las aguas residuales centralizadas. Montevideo cuenta con dos tipos de saneamiento principales: el más antiguo, de tipo unitario, representa el 60 % de la cobertura de la red y el restante es separativo, más reciente, que continúa en expansión (Figura 1.1). En el saneamiento unitario las aguas residuales domésticas e industriales y las aguas pluviales se conducen por la misma red de colectores (PNS, 2020). La zona abarcada por este tipo de saneamiento es la zona centro-sur (Figura 1.1). El resto del departamento, fundamentalmente los barrios del Oeste de la ciudad capitalina no cuentan con saneamiento por red de cañerías, utilizando fosas sépticas para la deposición de las aguas negras que es removido por la barométrica (PNS, 2020). Este sistema diferenciado de saneamiento repercute en las dinámicas de CCF que ocurren en la zona costera del Este y del Oeste del departamento. Cuando se registran precipitaciones, en la zona Este ocurren eventos de CCF asociados al sistema de red unitario, mientras que en el Oeste se dan aportes difusos asociados a las cañadas que llegan a la costa (PNS, 2020).

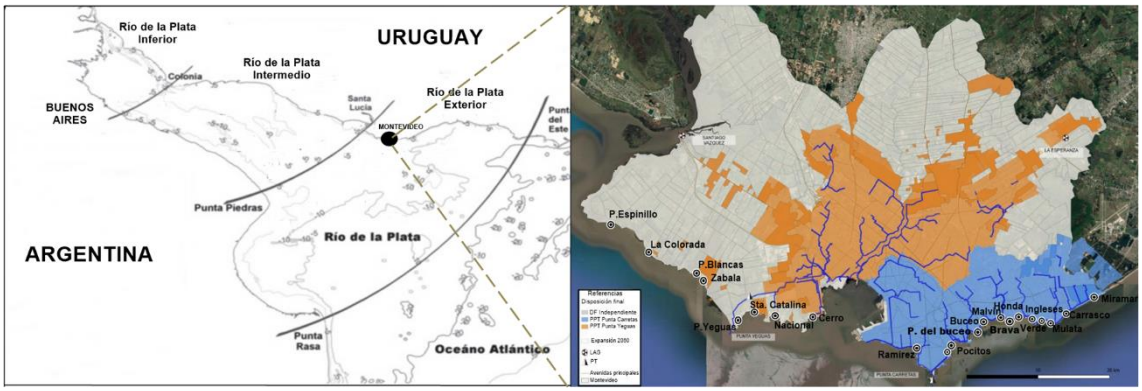

Figura 1.1. Área de estudio. A la izquierda ubicación de Montevideo en el Río de la Plata (se muestran tres regiones: Interior, Intermedio, y Exterior). A la derecha playas monitoreadas por la Intendencia de Montevideo con sus respectivos nombres y los tipos de saneamiento del departamento de Montevideo (en azul la zona correspondiente al colector de Punta Carretas, en naranja la zona correspondiente al colector de Punta Yeguas, en gris se muestra la zona que no está conectada a la red de saneamiento). Imagen modificada del Plan Nacional de Saneamiento, 2020.

La Intendencia de Montevideo (IM), a través del Departamento de Desarrollo Ambiental y del Servicio de Evaluación de la Calidad y Control Ambiental (SECCA) monitorea la calidad de las playas. Este monitoreo se realiza en 21 playas del departamento (desde playa Punta Espinillo a playa Miramar, Figura 1.1), de 2 a 3 veces

por semana en la temporada estival, desde el 15 de noviembre al 31 de marzo de cada año. En cada playa, se toman medidas *in situ* de variables físico-químicas del agua (temperatura, salinidad, turbidez, conductividad y pH) y muestras de agua para la cuantificación en el laboratorio de clorofila-a, nutrientes (fósforo y nitrógeno totales) y CF. El nivel de CF se conoce luego de la incubación entre 24 y 48hs y los resultados obtenidos son divulgados en la página oficial de la IM de forma semanal durante la temporada estival (IM, 2023).

Montevideo cuenta con una adecuada cobertura espacial y temporal del monitoreo de la calidad del agua de las playas, pero no cuenta con un sistema de implementación de modelos predictivos integrados a su sistema de gestión. La implementación de modelos predictivos se utiliza en diferentes partes del mundo (Francy et al., 2020) y es recomendada por organismos internacionales como complemento de los monitoreos de calidad de playas (OMS, 2018). A medida que aumenta la capacidad de cómputo y las herramientas computacionales, algunos programas de monitoreo incluyen diversas técnicas de modelización. Los modelos tienen la capacidad de sistematizar la información y pueden ayudar a anticipar las condiciones de CCF de las playas y así prevenir situaciones sanitarias riesgosas (Francy et al., 2020; Heasley et al., 2021). Esta implementación requiere de la comunicación entre las instituciones gestoras de la calidad de las playas y los desarrolladores de los modelos. En algunos casos las predicciones de los modelos son utilizadas por los gestores de la calidad de playas correspondientes de cada localidad/ciudad/región para tomar una decisión respecto a la habilitación o no de las playas. En otros casos, las predicciones se distribuyen a los usuarios de las playas a través de interfaces web o aplicaciones de celular (Thoe et al., 2014, 2018; Dada et al., 2019).

En Uruguay, el desarrollo de modelos predictivos con algoritmos de Aprendizaje Automático (AA) es un campo en crecimiento en distintas áreas del conocimiento (Borba et al., 2021; Garabedian et al., 2021; Velardez y Dima, 2022; Ramirez, 2022). Específicamente, en el campo de las ciencias naturales, se ha comenzado a implementar técnicas de AA para abordar diversas temáticas de biología, ecología y ciencias ambientales (Crisci et al., 2017; Kruk et al., 2017; Segura et al., 2017; Botto et al., 2018; Bourel y Segura, 2018; Serrón et al., 2020; Bourel et al., 2021; Cal, 2022). En el caso de la modelización predictiva de la CCF, se han desarrollado una diversidad de modelos de

AA para las playas de Montevideo (Segura et al., 2021; Bourel et al., 2021). Estos modelos obtuvieron una buena capacidad predictiva respecto a modelos construidos en el mundo (Segura et al., 2021). Sin embargo, aún no se han implementado en el sistema de gestión de calidad de playas para dicho departamento.

Objetivo General

El objetivo general de esta tesis es generar aportes para la implementación de un sistema de alerta temprana de calidad medida por coliformes fecales en las playas de uso recreativo de Montevideo.

Objetivos específicos

1. Realizar una revisión bibliográfica de los modelos estadísticos y sus métricas utilizadas para la predicción de la calidad de las playas de uso recreativo medida por CCF a escala mundial y evaluar la implementación de modelos estadísticos en la gestión y como SAT de contaminación a nivel mundial y en la región.

2. Incorporar variables satelitales a los modelos de predicción generados hasta el momento para las playas de Montevideo y evaluar su desempeño.

3. Desarrollar una interfaz web para implementar de forma accesible los modelos de predicción de coliformes fecales desarrollados para las playas de Montevideo.

En el capítulo 2 se presenta una revisión de los modelos predictivos de CCF desarrollados hasta el 2024 en la literatura a nivel mundial. Se detallaron los tipos de modelos, las variables que utilizaron para desarrollarse, la estrategia de modelización y su desempeño. Además, se evaluó la implementación de estos modelos en sistemas de gestión de calidad de playas. Mediante esta revisión se pudieron identificar vacíos de conocimiento y realizar recomendaciones o líneas de avance. Se realizaron recomendaciones respecto a las estrategias de modelización y de implementación de los modelos en sistemas de gestión.

En el capítulo 3 se presentan los resultados de nuevos modelos para predecir CCF que incluyeron variables satelitales (temperatura superficial del agua y turbidez como predictoras) para las playas de Montevideo. Se comparó el desempeño entre modelos construidos con variables remotas (precipitaciones y velocidad del viento tomadas de

estaciones meteorológicas y satelitales) y modelos construidos con variables tomadas *in situ* (temperatura del agua, salinidad, turbidez, valores de muestreos pasados de coliformes fecales). Luego, se presenta la implementación de los modelos con mejor desempeño en una interfaz web creada específicamente para las playas de Montevideo.

.

CAPÍTULO 2

REVISIÓN SISTEMÁTICA DE MODELOS Y MÉTRICAS PARA PREDECIR CONTAMINACION FECAL

2.1 INTRODUCCIÓN

Los programas de monitoreo de la calidad de playas son herramientas fundamentales diseñadas para prevenir riesgos a la salud en el corto plazo (semanas) y supervisar la evolución de la CCF ante diferentes medidas de gestión (OMS, 2018). Esta información es utilizada para la generación de alertas, la habilitación de los diferentes usos de las playas y para evitar la exposición directa de los seres humanos a concentraciones elevadas de CCF (Thoe et al., 2014; Shively et al., 2016). Los monitoreos se basan en su mayoría en el análisis de muestras en el laboratorio, lo que implica la colecta de muestras de agua y del cultivo de bacterias. Este procedimiento insume de 24 a 48 horas para obtener resultados (APHA, 1989), lo cual dificulta la realización de pruebas de monitoreo diarias e impide tomar decisiones de gestión rápidas con respecto a la calidad de la playa. Además, la frecuencia de colecta de muestras, en distintas regiones mundialmente, es relativamente baja (ej. Una o dos veces por semana) lo que implica que las decisiones de manejo y las notificaciones al público estén basadas, en su mayoría, en condiciones previas (Kim y Grant, 2004; Francy et al., 2020).

Los modelos predictivos han surgido recientemente como un complemento posible para superar estos problemas y su implementación es recomendada para la gestión de la calidad de playas recreativas por algunas agencias nacionales e internacionales (USEPA, 2012; OMS, 2018). Los modelos predictivos combinan variables ambientales (precipitaciones, irradiación solar, entre otras) y de calidad del agua en un modelo para estimar la concentración de FIB o la probabilidad de que se supere un umbral determinado (Bedri et al., 2016). La construcción de modelos es un paso que se debe incluir en un sistema de alerta temprana que debe ir acompañado del monitoreo *in situ* y de la transferencia de información a gestores y usuarios de las playas (Figura 2.1).

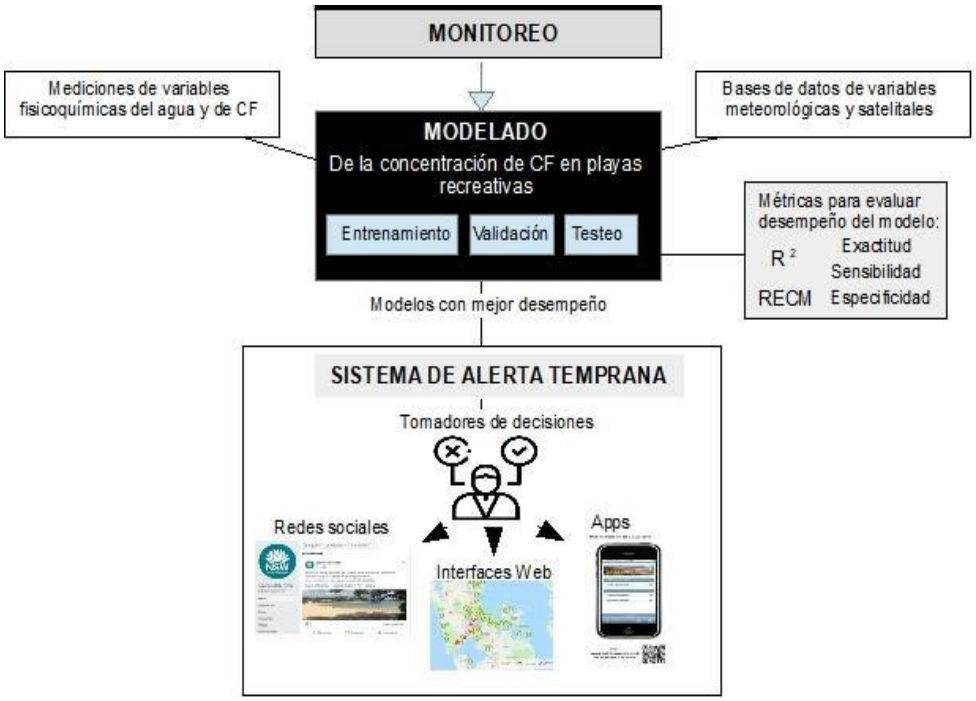

Figura 2.1 Diagrama de un Sistemas de Alerta Temprana (SAT) de contaminación por coliformes fecales en playas recreativas, con los diferentes pasos desde el monitoreo, la modelización, y la transferencia de información a los encargados de la gestión y los usuarios de las playas. La información generada en los sistemas de monitoreo de la calidad de las playas incluyendo la concentración de FIB y diferentes variables físico-químicas (temperatura, salinidad, turbidez, etc.) en conjunto con bases de datos de condiciones climáticas (dirección e intensidad del viento, nivel de marea, temperatura del aire, etc.) se utilizan para el desarrollo de modelos estadísticos predictivos. Los modelos se entrenan usando una fracción del conjunto de datos (conjunto de entrenamiento), y su desempeño se evalúa usando un conjunto de datos de validación y la fracción restante de los datos (conjunto de testeo) mediante métricas como: el coeficiente de determinación (R^2), la Raíz del Error Cuadrático Medio (RECM), la Precisión, Sensibilidad y Especificidad. Los modelos que presentan mejores métricas son seleccionados para formar parte de sistemas de alerta temprana o "Nowcasting" según las necesidades de cada entidad gestora y las características de cada playa. Los sistemas de alerta temprana suelen incluir interfaces web, aplicaciones móviles y redes sociales en las que se muestran las predicciones del modelo.

Se han empleado una amplia gama de técnicas de modelización predictiva de CCF en playas recreativas. Estas técnicas se pueden clasificar en *i)* modelos mecanicistas (físicos/hidrológicos), *ii)* enfoques estadísticos basados en datos (empíricos) o *iii)* una combinación de ambos. Revisiones previas han abordado la diversidad de estrategias de modelización y predicción centrándose en modelos de regresiones lineales y modelos mecanicistas en aguas superficiales en todo el mundo (de Brauwere et al., 2014; Heasley et al., 2021). En el trabajo de De Brauwere et al., (2014), los autores abordan una descripción exhaustiva de los modelos mecanicistas en los que las concentraciones de FIB se predicen basándose en la dinámica de las fuentes, los sumideros y los procesos internos que influyen en los CCF en el medio acuático.

Los modelos mecanicistas requieren la formulación de relaciones entre variables explicativas (por ejemplo, temperatura del agua, salinidad, turbidez, precipitaciones, niveles de descarga de ríos y arroyos, tasas de flujo y de corriente de agua) y la concentración de FIB, que son ajustados para la modelización de la CCF en ríos, estuarios y playas costeras (Bolker, 2007). Estos modelos se concentran en las causas, procesos y mecanismos que explican la presencia de CCF, utilizando funciones, y ecuaciones basadas en teorías explicativas (Bolker, 2007). En este grupo de modelos están incluidos modelos con representación de la hidrodinámica en una, dos o tres dimensiones (ej. el modelo MOHID, y modelos hidro-ecológicos SENEQUE, MIKE) (Palazon et al., 2017; Kayode y Kumarasamy, 2018). Otros modelos, como los modelos de decaimiento de patógenos o los modelos cuyo objetivo es identificar y modelar las fuentes de FIB en ambientes acuáticos, requieren la formulación de ecuaciones de descomposición de patógenos, incluida la concentración inicial de patógenos y la tasa de descomposición, que depende de factores como la luz solar, la temperatura y la salinidad, entre otros (ej. Bacterial Water Quality Model, Environmental Fluid Dynamics Code, Soil and Water assessment tool: SWAT) (Bradford et al., 2013). Estos modelos han sido ampliamente desarrollados, aunque, al basarse en diferentes supuestos, conceptos y propósitos, se recomienda su estandarización (Wang et al., 2013).

Los métodos estadísticos basados en datos para predecir FIB o el exceso de un determinado umbral de FIB (variable de respuesta) se basan en ajustar un modelo a los datos utilizando co-variables (variables explicativas) y los modelos más comunes incluyen Regresiones Lineales Múltiples (MLR por su acrónimo en inglés), Regresiones de Mínimos Cuadrados Parciales u Ordinarios (PLS, OLS por su acrónimo en inglés), entre otros (Heasley et al., 2021). Dichos modelos han resultado ser útiles debido a su facilidad de implementación e interpretación. Sin embargo, rara vez se cumplen sus supuestos (de Brauwere et al., 2014).

En las últimas dos décadas, los algoritmos de Aprendizaje Automático (AA) han surgido como una herramienta versátil capaz de modelar respuestas no lineales entre variables y de mejorar la precisión de la predicción de CCF (Park et al., 2018; Shively et al., 2016; Segura et al., 2021). Entre las ventajas de estas técnicas se encuentra que pueden manejar relaciones entre variables altamente no lineales, valores atípicos, y tienen una alta flexibilidad por su capacidad de aprender de nuevos datos. El objetivo principal de las técnicas de AA supervisado es estimar una función *f* capaz de predecir una variable *Y*

basado en un set de variables explicativas o *X*. La variable a predecir usualmente se nombra como *output* y las variables explicativas o de entrada se nombran como *input* (James et al., 2013). Si la variable a predecir se trata de una variable continua o cuantitativa entonces se trata de un problema de regresión (ej. Predecir la concentración de CCF). Si la variable a predecir *Y* es una variable categórica, es decir, no es numérica (ej. una categoría de calidad del agua: playa abierta o cerrada) el problema es de clasificación. Las Redes Neuronales Artificiales (ANN por su acrónimo en inglés), las técnicas basadas en árboles de decisión (CART y Random Forest por su acrónimo en inglés) y las Máquinas de Vectores de Soporte (SVM por su acrónimo en inglés) se encuentran entre los algoritmos más utilizados (He y He, 2008; Park et al., 2018; Parkhurst et al., 2005; Zhang et al., 2015).

Dada su flexibilidad, una desventaja de los algoritmos de AA es el riesgo de sobreajuste a los datos de entrenamiento. Para superar este problema, deben utilizarse técnicas específicas de validación del desempeño de los modelos que usualmente implica dividir el conjunto de datos en dos o tres subconjuntos de datos (entrenamiento, validación y testeo). El conjunto de datos de entrenamiento, se utiliza para construir el modelo y ajustar los hiperparámetros. El conjunto de datos de validación sirve para evaluar los mejores modelos. Finalmente, en el conjunto de datos de testeo se evalúa el desempeño del modelo en datos nunca antes vistos durante la etapa de evaluación del modelo (Cawley y Talbot, 2010; James et al., 2013). En el contexto de la modelización predictiva de CCF, uno de los desafíos para las técnicas de AA, es lidiar con conjuntos de datos desbalanceados (Han et al., 2005; He y He, 2008; Xu et al., 2020). Esto es debido a que los eventos de alto riesgo para la salud son relativamente raros en comparación con situaciones de bajo riesgo, por lo que se requieren algoritmos y funciones específicos para predecir con precisión una fracción menor de los días en los que la calidad del agua excede los umbrales permitidos por las normativas de cada país (Bourel et al., 2021).

Para evaluar el desempeño de un modelo y cuantificar en qué medida el valor de respuesta previsto para una observación determinada se acerca al valor de respuesta real para esa observación se utilizan diversas métricas (James et al., 2013). Para los modelos de regresión, que predicen una variable de respuesta continua *Y*, se busca minimizar el error cuadrático medio (MSE por su acrónimo en inglés) y la raíz del error cuadrático medio (RMSE por su acrónimo en inglés). El RMSE es la raíz del promedio de la diferencia entre los valores observados y los predichos al cuadrado. Para los métodos de

clasificación que predicen la clase o categoría para una observación determinada (por ejemplo, una categoría de calidad del agua: playa abierta o cerrada), con frecuencia se utilizan la precisión, la sensibilidad y la especificidad de la clasificación. Estas métricas se estiman dependiendo de cuántas observaciones fueron predichas correcta o incorrectamente por el modelo (James et al., 2013). Se prefiere y selecciona el modelo con mejores métricas sobre otros (en este caso, estas medidas intentan maximizarse). La comparación entre métricas de los modelos, proporciona una forma objetiva de seleccionar modelos. Si bien estos criterios están bien instaurados en la literatura de modelización de AA, su aplicación y uso para predecir CCF no ha sido evaluada.

Los modelos seleccionados en base a su desempeño predictivo pueden implementarse en Sistemas de Alerta Temprana (SAT) o en sistemas de "Nowcasting". Este último término en inglés es el que se utiliza para nombrar a una herramienta o sistema que utiliza modelos predictivos para proporcionar predicciones (en este caso de CCF) en tiempo real o casi tiempo real ya sea a los gestores de una playa o a los usuarios (Francy et al., 2020; Heasley et al., 2021). El SAT implica, además de la implementación de los modelos predictivos, del desarrollo de campañas de concientización a la población y en particular a los usuarios de las playas, la traducción del lenguaje de la modelización al lenguaje de la comunicación (por ejemplo un código de colores: rojo significa alerta, naranja peligro, verde fuera de peligro) y de la comunicación de las predicciones en interfaces web amigables o aplicaciones al celular que puedan ser fácilmente utilizadas por los usuarios de las playas (Figura 2.1) (Francy et al., 2020). La implementación de modelos en SAT se observa mayoritariamente en países que cuentan con sistemas de monitoreo de largo plazo en conjunto con el desarrollo de modelos predictivos (Francy et al., 2020). A pesar de la cantidad de modelos desarrollados para la predicción de CCF, su implementación en SAT de calidad en playas recreativas es aún escaso (Francy et al., 2020).

En esta revisión, primero se sistematizó la información sobre las estrategias de modelización utilizadas para predecir FIB en aguas recreativas naturales en las últimas dos décadas (entre 2000 y 2024) y se describió con detalle la diversidad de estrategias de predicción de AA y la evaluación de sus desempeños. Luego, se exploraron las variables utilizadas como variables de entrada (inputs) y los criterios y métricas utilizados para evaluar el desempeño y la selección de los modelos. Finalmente, se presentaron y describieron la variedad de SAT que han sido implementadas por instituciones gestoras

de la calidad de playas de cada país/región. En conjunto, se espera poder aportar a la implementación de los modelos predictivos en sistemas de gestión de calidad de playas, identificando los desafíos que se encuentran para desarrollar los mismos, las técnicas necesarias para mejorar el desempeño de los modelos e identificar rutas de avance en la modelización predictiva de CF.

2.2 MÉTODOS

Revisión de literatura y criterio de búsqueda

El período de búsqueda incluyó artículos publicados desde febrero de 2000 hasta febrero de 2024 utilizando el motor de búsqueda de acceso abierto Google Académico (título y resumen), y plataformas de acceso gratuito a los artículos (ej. Portal TIMBO. http://www.timbo.org.uy/) disponibles en Uruguay como un acceso universal a la literatura científica en línea. Las búsquedas se realizaron utilizando las siguientes combinaciones de palabras clave: “monitoreo de la calidad de playas”, “playas recreativas” y “modelado de bacterias fecales/indicadoras fecales”, “*Escherichia coli*” y “coliformes totales/fecales”. Para todos los artículos relevantes, se siguieron los enlaces a "artículos relacionados" y "citados por" para identificar artículos de interés. Los dominios del estudio se restringieron a las aguas superficiales naturales, tal como las define la OMS como “cualquier área costera, estuarinas o de agua dulce donde un número significativo de usuarios realizan cualquier tipo de uso recreativo del agua” (OMS, 2003). Se consideraron artículos internacionales publicados en revistas revisadas por pares que incluían información de cualquier tipo de modelo cuyo objetivo fuera la predicción, ya sea que emplearan modelos mecanicistas o modelos basados en enfoques estadísticos basados en datos (empíricos). Dentro del alcance de esta revisión, en una segunda selección de artículos, seleccionamos exclusivamente aquellos que presentaban enfoques estadísticos basados en datos (empíricos), ya sea que su objetivo fuera predecir concentraciones de FIB (regresión) o predecir la excedencia de un determinado umbral de FIB (clasificación).

Para evaluar cuantos de los modelos desarrollados en la literatura eran implementados en SAT. Buscamos interfaces web en línea que mostraran predicciones

de modelos estadísticos para realizar notificaciones públicas sobre la calidad de playas recreativas. Incorporamos las palabras claves (en español e inglés) seleccionadas: "pronóstico de la calidad de agua en playas", "programas de calidad de agua en playas recreativas" y "Nowcasting para predecir la calidad del agua de las playas". Los criterios de selección se centraron en programas con modelos predictivos operativos en sistemas de gestión, y que brindaran predicciones de CCF. Se seleccionaron las páginas web que contaban con interfaces específicamente creadas para informar al público sobre la calidad de las playas utilizando modelos predictivos. También se tuvieron en cuenta las aplicaciones para celulares.

Análisis de datos

Se analizaron un total de 96 artículos destinados a predecir CCF. Dentro de estos, 16 artículos incluyeron estrategias de modelación hidrodinámica e hidro-ecológicos y 13 cuyo objetivo era estimar la tasa de decaimiento de FIB. Estos artículos no se incluyeron en el siguiente análisis, ya que no estaban dentro del alcance de este capítulo.

Para los 67 artículos restantes se registró la información sobre el país, el área de estudio, el tipo de ecosistema (de agua dulce, estuarinos o costero-marino), datos sobre el sistema de monitoreo incluyendo: frecuencia, estación del año y la estructura de los datos de muestreo (por ejemplo, tamaño del conjunto de datos, número de casos en los que el FIB excede ciertos valores guía). También se registró el enfoque de modelación, es decir, que tipo de modelos eran utilizados para la predicción (ej. modelos no supervisados o supervisados, modelos de regresión o clasificación) y las características de las variables de entrada y salida.

Los modelos estadísticos de los artículos seleccionados se agruparon siguiendo a Huang et al., (2021) en dos tipos principales de estrategias de modelado: aprendizaje supervisado y aprendizaje no supervisado. Huang et al., (2021) asignaron a los modelos de aprendizaje profundo (Deep learning, en inglés), que refieren básicamente a diferentes tipos de redes neurales, en un tercer grupo separado, que en este trabajo se clasificaron dentro de la estrategia de modelos supervisados principales. Dentro del aprendizaje supervisado, se diferencian métodos de regresión y métodos de clasificación. Los métodos de regresión típicamente incluyeron a la Regresión Lineal Múltiple y la Regresión Polinómica y de Mínimos cuadrados parciales. Los métodos de clasificación,

incluyen algoritmos como el clasificador de Bayes naif "Naive Bayesian Classifier" y la Regresión Logística binaria o multinomial. Las técnicas de AA (que se pueden utilizar tanto para regresión como para clasificación) incluyen, entre otros, las Máquinas de Vectores de Soporte (SVM), modelos basados en árboles de decisión (CART y Random Forest), técnicas de Boosting (Boosting Tree, AdaBoost) y Redes Neuronales Artificiales (ANN: NIO, NAR, PNN). Finalmente, el aprendizaje no supervisado incluye, entre otros, Análisis de Componentes Principales (PCA), análisis de agrupamiento (Clustering) y Redes Bayesianas.

El desempeño de los modelos estadísticos se evaluó para métodos de regresión mediante el RMSE y el coeficiente de determinación (R^2) calculado en el conjunto de datos de testeo (es decir, datos independientes utilizados para ajustar el modelo). Estas métricas fueron registradas y resumidas en una tabla por tipo de estrategia de modelado. En estudios con respuestas categóricas, donde se intenta predecir el exceso de un determinado valor guía o nivel de alerta, se registró la tasa de falsos positivos (FP), falsos negativos (FN), verdaderos positivos (VP) y verdaderos negativos (VN). La precisión es (VP+VN) /(VP+FP+FN+VN), la sensibilidad o tasa de verdaderos positivos: VP/VP + FN y la especificidad o tasa de verdaderos negativos: VN/VN + FP (Tabla 2.1) (James et al., 2013).

Tabla 2.1 Matriz de confusión teórica con los posibles resultados de la predicción de dos clases. En las columnas se agregan los valores observados en la realidad y en las filas los valores predichos.

		Observado	
		Excede	No Excede
Predicho	Excede	Verdadero Positivo (VP)	Falso Positivo (FP)
	No Excede	Falso Negativo (FN)	Verdadero Negativo (VN)

En algunos casos, como parte de la estrategia de modelización se utiliza una regla de decisión para interpretar el resultado de un modelo de regresión. Las predicciones de un modelo de regresión se clasifican *a posteriori* en función de un umbral regulatorio (que se define en base a la legislación de cada estado o país, por ejemplo, 235 UFC/100 ml, en los estados de Wisconsin y Ohio en los Estados Unidos y 2000 UFC/100ml en Uruguay) para separar las predicciones en excede y no excede el umbral. En estos casos,

el modelo ajustado es un modelo de regresión, pero la predicción se categoriza en dos categorías *a posteriori*. En estos casos, se registraron métricas de clasificación como la FP, FN, VP, VN, sensibilidad, especificidad y precisión.

En este capítulo, se resumen los resultados del desempeño de los modelos (métricas) para modelos desarrollados al menos cuatro veces. Registramos las métricas de los modelos evaluados en el conjunto de datos de test o de validación (conjunto de datos de validación o prueba) para evaluar el desempeño del modelo. Aquellos artículos que evaluaron el ajuste en los mismos datos utilizados para entrenar en el modelo no se incluyeron en el análisis de desempeño pues esta forma de evaluación tiene alta probabilidad de sobreajuste (James et al., 2013). Además, se evaluaron los métodos de selección de variables predictoras (inputs) y se identificaron y enumeraron las variables predictoras más importantes incluidas en los modelos finales. También se llevó a cabo una búsqueda de métodos para abordar el desbalance de clases del conjunto de datos.

Para evaluar la aplicabilidad de los modelos predictivos creados en el marco académico, primero se evaluó si había un vínculo entre estos y algún desarrollo de SAT. Luego buscamos interfaces operativas de implementación de modelos predictivos, y se registró el país, la institución o empresa encargada de la gestión de la calidad de las playas, el tipo de modelo (por ejemplo, estadístico o hidrodinámico), el tipo de FIB utilizado como referencia (por ejemplo, EC, CF o ENT) y el dominio de estudio (agua dulce, estuarinos, costero-marino). Se registró toda la información relevante sobre el modelo operativo implementado en un SAT.

2.3 RESULTADOS

Artículos de modelación

Los 67 artículos que cumplieron con nuestros criterios se desarrollaron en 12 países de 4 continentes (Figura 2.2, Tabla 1 del Anexo). La mayoría de los estudios (66%) se concentraron en Estados Unidos, aunque en el período de 2015 a 2024 la diversidad de países aumentó (Figura 2.2).

De las 28 técnicas de modelado registradas, 27 correspondieron a aprendizaje supervisado y 1 empleó aprendizaje no supervisado (Tabla 2 en el Anexo). Las MLR, las ANNs y otras técnicas de AA (árboles de decisión) se encuentran entre los modelos mayormente desarrollados (Figura 2.3). Hacia el final del período analizado, el número de artículos, el número total de modelos y la diversidad de estrategias de modelos aumentaron, 20%, 65% y 50% respectivamente (Figuras 2.2 y 2.3).

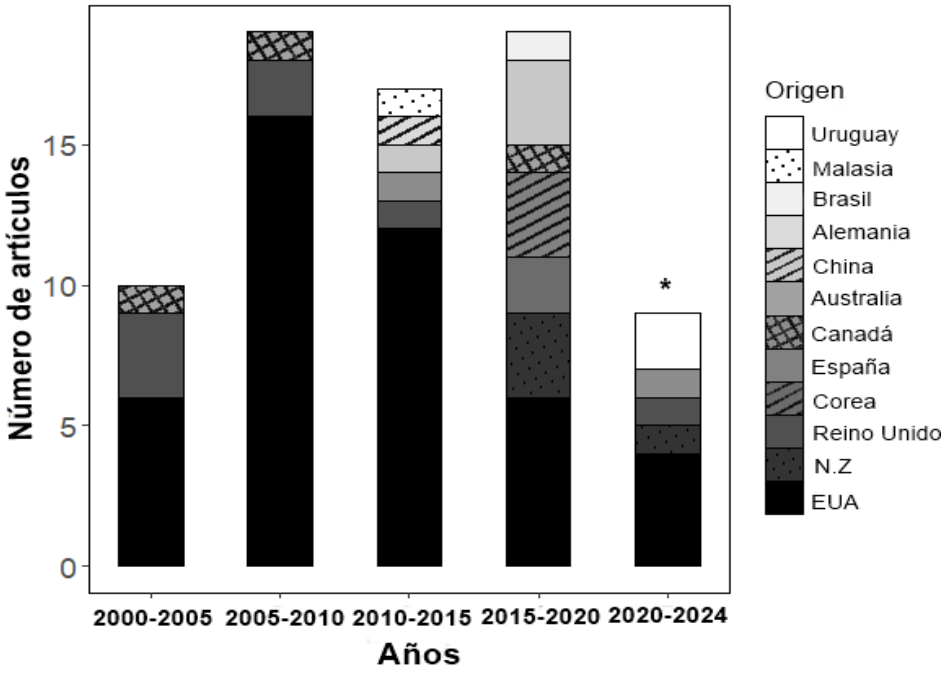

Figura 2.2 Origen de los artículos académicos registrados en el periodo de estudio analizado (2000 al 2024). Con un asterisco se indica que el último periodo analizado abarca del 2000 a febrero de 2024 (3 años), en vez de 5 años como el resto de los periodos analizados.

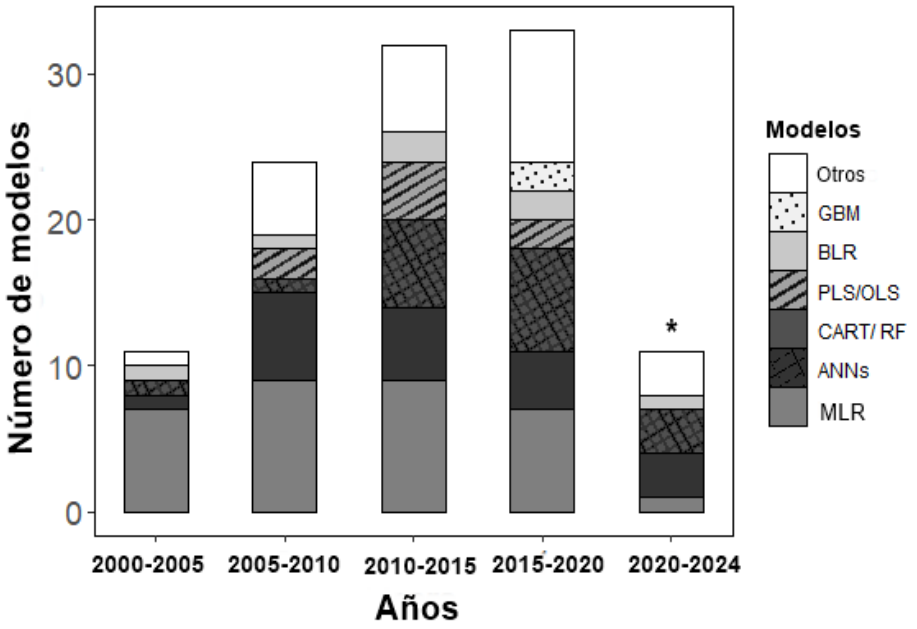

Figura 2.3 Tipos de modelos registrados durante el periodo de estudio analizado (2000-2024). Se registraron Regresiones Lineales Múltiples (MLR), Redes Neuronales Artificiales (ANNs), árboles de decisión (CART y Random Forest (RF)), así como también Regresión Polinómica y de Mínimos cuadrados parciales (PLS/OLS), Regresión Logística Binaria (BLR), Modelo Boosted generalizado (GBM). Los modelos que fueron desarrollados menos de 4 veces en la literatura académica fueron incluidos como "Otros". Con un asterisco se indica que el último periodo analizado abarca del 2000 a febrero de 2024 (3 años), en vez de 5 años como el resto de los periodos analizados.

Los modelos analizados fueron desarrollados mayoritariamente para ambientes de agua dulce (lagos, lagunas, ríos y arroyos, 67%), mientras que los sistemas costeros-marinos, y estuarinos, alcanzaron un porcentaje menor (33%). La extensión de las series temporales de los programas de monitoreo varió entre menos de un año y 24 años, con un promedio de 5 años. Se ajustaron modelos predictivos para una playa en 18 casos y para múltiples playas en 49 casos. El número promedio de puntos de muestreo (sitios de muestro dentro de una playa) por playa y por tipo de FIB (por ejemplo, EC, CF, ENT, TC) fue 322 (desvío estándar (ds) =273). En general, la mayoría de los programas de monitoreos llevaron a cabo muestreos con una frecuencia semanal. La frecuencia máxima de muestreo fue diaria en verano con un promedio de frecuencia de muestreo de 2,3 días (ds= 1,6) por semana. La mayoría de los estudios (75%) sólo monitorearon la calidad de las playas durante los meses de verano y la temporada de baño, mientras que sólo en cuatro casos se dispuso de datos durante todo el año. El período promedio de temporada de muestreo fue de 6 meses (ds= 3,3) al año. La base de datos más grande constaba de 20908 observaciones de 6 años correspondientes a 25 playas y tres tipos de FIB (CF, TC,

ENT), con algunas playas muestreadas diariamente y otras semanalmente (Thoe et al., 2015).

Métricas de desempeño de los modelos

Se registraron métricas de desempeño de modelos predictivos de un total de 40 artículos (Tabla 2.2). Los artículos restantes (N= 27) no dividieron el conjunto de datos en entrenamiento y evaluación y no presentaron resultados del modelo evaluados con un conjunto de datos independiente. La mayoría de los artículos (N=14) implementaron una combinación de ambas estrategias (primero ajustar un método de regresión y luego categorizar el resultado), mientras que un total de 13 artículos utilizaron solo métodos de regresión. Siete artículos implementaron ambos tipos de estrategias de modelo (regresión y clasificación por separado). Seis artículos utilizaron únicamente métodos de clasificación. Los métodos de regresión transformaron principalmente la concentración de FIB a logaritmo o utilizaron la media geométrica de FIB como variable a predecir.

Los métodos de clasificación utilizaron dos categorías como variable de respuesta en 9 artículos y dos, tres y cuatro categorías en 3 artículos. Un estudio evaluó la predicción de cinco categorías. Considerando todos los modelos de clasificación binaria, hubo un desbalance entre las clases, donde la clase minoritaria (exceso) estuvo representada en promedio en el 15% (ds= 8,3) de los casos con un mínimo del 3%. Para los modelos con más de dos clases, se registró un desbalance entre las clases en todos los artículos, en donde la clase minoritaria representó en promedio el 12,5% (ds= 4,5) de los casos, con un mínimo del 6%. En 7 de los 40 artículos se registró la aplicación de métodos que abordan el desbalance de clases mediante diferentes algoritmos: el algoritmo de muestreo sintético adaptativo (ADAsyn) para generar casos de FIB sintéticos por encima del umbral (Xu et al., 2020), el método de agrupamiento de K-medias (Choi y Seo, 2018) en el proceso de muestreo de conjuntos de datos de prueba y entrenamiento (al muestrear una determinada porción de cada clase, los conjuntos de datos de entrenamiento y prueba pueden mostrar distribuciones similares), y por último, manipular el umbral de probabilidad de exceso para lograr la sensibilidad y especificidad deseadas (Searcy et al., 2018).

Tres artículos generaron modelos en el software Virtual Beach, un software específico desarrollado por la Agencia de Protección Ambiental de los Estados Unidos

(por su acrónimo en inglés: USEPA) para construir modelos estadísticos de niveles de indicadores de patógenos en playas recreativas basados en MLR (Cyterski et al., 2013).

Variables input

El conjunto de variables *inputs* se agruparon en las siguientes categorías: *i)* mediciones *in situ* de variables físico-químicas y biológicas de la calidad del agua (ej. temperatura del agua, turbidez, conductividad/salinidad, oxígeno disuelto, pH, nitrógeno total, fósforo y carbono orgánico, sólidos en suspensión y clorofila-a); *ii)* variables meteorológicas (ej. duración e intensidad de las precipitaciones, radiación solar, nubosidad, temperatura y presión del aire, humedad, velocidad y dirección del viento); *iii)* variables hidrodinámicas (ej. altura y período de las olas, nivel y amplitud de las mareas, caudales de ríos y arroyos y nivel de descarga); *iv)* concentración reciente e histórica de la CCF; *v)* variables registradas a partir de satélite (ej. Temperatura superficial del agua, turbidez) *vi)* otros (ej. presencia de manto de algas, número de aves y nadadores, densidad de población, nivel de urbanización, día del año) (Figura 2.4; Tabla 3 en Anexo).

Las condiciones de precipitaciones previas (acumuladas, 24, 48 y 72 hs) y la temperatura del agua fueron los inputs más frecuentes y relevantes elegidos tanto en los ambientes costeros-marinos como de agua dulce. La velocidad y dirección del viento, la altura de las olas, la altura de la marea y la radiación solar fueron relevantes para modelar FIB en dominios costeros-marinos, mientras que la velocidad, el volumen, la intensidad del flujo de agua y la turbidez fueron relevantes en el caso de ambientes de agua dulce (Figura 2.4; Tabla 3 en Anexo).

Las estrategias para la selección de variables y la definición de la importancia de las variables inputs variaron entre artículos y estrategias de modelado. La mayoría de los estudios se basaron en el método de regresión de selección de variables paso a paso o "Stepwise selection" por su nombre en inglés. Este método se basa en incluir una variable independiente (input) a la vez en el modelo de regresión si es estadísticamente significativa o incluyendo todas las variables independientes potenciales en el modelo y eliminando aquellas que no son estadísticamente significativas. Se realizaron análisis de correlación (por ejemplo, correlación de Pearson) entre las covariables y el factor de

inflación de la varianza (VIF por sus siglas en inglés) fue el método más utilizado para abordar el problema de la multicolinealidad (6 artículos). En el caso de las técnicas de AA, los artículos modelados con CART incluyeron el índice de Gini para medir la impureza de los nodos y en el caso de Random Forest, este algoritmo incluyó métodos de selección de variables basados en la disminución media de la exactitud (MDA).

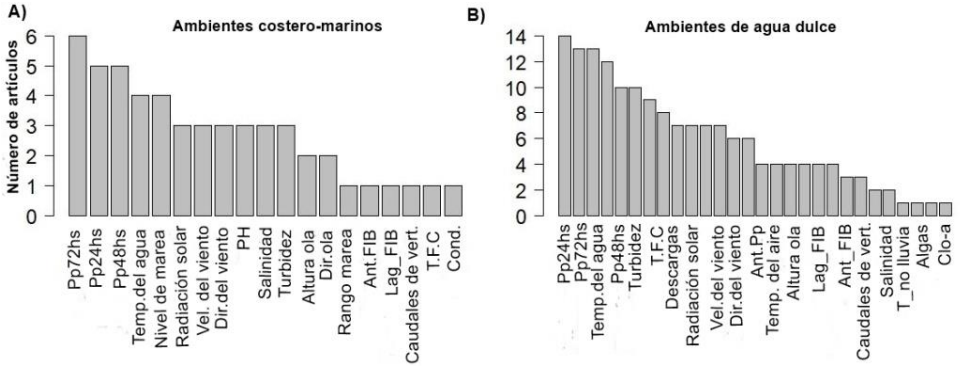

Figura 2.4 Frecuencia de las variables de ingreso (input) utilizadas para la modelización en A) ambientes costero-marinos y B) ambientes de agua dulce (ríos, arroyos, lagos, lagunas). Precipitaciones acumuladas de 24,48 y 72 horas previas (Pp24,48,72hs), Temperatura del agua (Temp. del agua), Nivel de marea, Radiación solar, Velocidad del viento (Vel. Del viento), Dirección del viento (Dir. del viento), pH, Salinidad, Turbidez, Altura de la ola, Dirección de la Ola, Rango de marea, Concentraciones de CCF de días previos (Lag_FIB), Caudales de las vertientes, Tasa de flujo del caudal (T.F.C), Conductividad (cond.), Descarga del río (descarga), Antecedentes de precipitaciones (Ant_Pp), Antecedentes de concentraciones de CCF (Ant_FIB), Tiempo desde la última precipitación registrada (T_no lluvia), Algas y Clorofila-a (Clo-a).

Evaluación de los modelos

El desempeño de los modelos se evaluó principalmente mediante error de validación cruzada (18 artículos). En 10 artículos, los modelos utilizaron una única partición aleatoria de datos, entrenaron modelos con una fracción de los datos y evaluaron los resultados de los datos restantes que no se utilizaron para el entrenamiento del modelo. En 12 artículos se seleccionó una serie temporal de las observaciones para entrenar el modelo y testearon el modelo con la serie temporal restante.

Métricas utilizadas para evaluar los modelos de regresión

En 23 de 40 artículos la bondad de ajuste o el desempeño del modelo se evaluó mediante el coeficiente de determinación R^2, entre valores ajustados y observados que varió ampliamente entre 0,04 y 0,95. Para los modelos de ANN, el R^2 fue en promedio mayor que MLR y OLS ($p < 0.01$) (Figura 2.5). El RMSE (y=log (FIB)) se utilizó como el principal determinante del desempeño de los modelos en 27 de 40 artículos, y varió ampliamente para los diferentes tipos de modelo (MLR =0,15 a 1,9 y ANN =0,08 a 8,2), y fue menor en PLS=0,43 a 0,48. En el caso de los árboles de regresión (CART y RF), el RMSE varió entre 0,3 y 0,7 con una media de 0,4 (Tabla 2.2). En cuanto a los modelos continuos en el cual la salida fue categorizada *a posteriori*, la sensibilidad fue menor que la precisión y especificidad para todas las estrategias de modelización (Tabla 2.2).

Métricas utilizadas para evaluar los modelos de clasificación

Las métricas utilizadas en los métodos de clasificación implicaron principalmente la precisión, sensibilidad y especificidad (Tabla 2.1). En general, se hizo hincapié en minimizar los falsos negativos (es decir, la sensibilidad) para proteger la salud pública. Se estimó la precisión, la sensibilidad y la especificidad en 29 de 40 artículos. El valor medio y el rango de estas métricas se resumen por tipo de modelo en detalle en la Tabla 2.2. Las ANNs, CART y RF presentaron un desempeño similar, considerando las tres métricas (Tabla 2.2).

Tabla 2.2. Valor medio y rango (entre paréntesis) de las principales métricas de evaluación de los modelos de regresión y clasificación recopilados en la literatura científica para predecir CCF. N es el número de modelos desarrollados para cada tipo de modelo. Lista de modelos: Redes Bayesianas (BN), árboles de clasificación y regresión (CART y RF), Regresión Logística Multinomial (MLogR), Regresión Lineal Múltiple (MLR), Regresión Logística Binaria (BLR), Regresiones de Mínimos Cuadrados Parciales u Ordinarios (PLS, OLS), Redes Neuronales Artificiales (ANN), Modelo Boosted Generalizado (GBM), Modelo de decisión Bagging (BGM), Árbol de decisión Boosting (BDT), Efectos lineales Mixtos (LME), regresión de vectores de soporte (SVR), máquina de vectores de soporte (SVM) y Análisis Wavelet (WA). La columna Referencias, indica el artículo científico donde se implementó cada modelo.

Tipos de modelos	**Modelo**	**Acrónimo**	**N**	**R^2**	**RMSE (y=log (FIB))**	**RMSE (y= FIB)**	**Precisión**	**Sen**	**Esp**	**Referencias**
Modelos de Regresión	Regresiones Lineales Múltiples	MLR	9	0.61 (0.21-0.7)	0.31 (0.15-0.4)					1,2
	Regresiones Lineales Múltiples en Virtual Beach	MLR-VB	5	0.59 (0.45-0.83)						3,4
	Redes Neuronales Artificiales (ANN, NIO, NAR, WA-NAR)	ANNs (ANN, NIO, NAR, WA-NAR)	38	0.79 (0.50-0.95)	2.78 (0.09-8.27)	1007.2 (55.4-3319.0)				6,7,8,9,10,11,13
	Regresión de vectores de soporte	SVR	4	0.55 (0.3-0.9)		760 (48-2190)				10
	Árboles de decision	CART, RF	6	0.28 (0.25-0.31)						14,38
Modelos de Regresión (predicción categorizada *a posteriori*)	Regresiones Lineales Múltiples	MLR	62	0.4 (0.1-0.8)	1.3 (0.4-1.9)		85 (58-100)	37 (0-100)	90 (65-100)	14, 15,16,17,18,19,20,21, 22,23,24,25,36
	Regresiones de Mínimos Cuadrados Parciales	PLS	6		0.4 (0.43-0.48)		86 (80-93)	46 (28-92)	93 (78-100)	24,26

	Regresiones de Mínimos Cuadrados Ordinarios	OLS	4			465,3 (454-486)	87 (79-96)	65 (38-92)	90 (77-98)	12,26
	Redes Neuronales Artificiales (ANN, NIO, NAR, WA-NAR)	ANNs (ANN, NIO, NAR, WA-NAR)	34	0.73 (0.04-0.95)	1.52 (0.08-8.2)	1147,8 (213-3319)	89 (70-100)	56 (0-100)	95 (78-100)	5, 7, 27, 8, 9, 28, 11,29, 37
	Árboles de decision	CART, RF	9		0.4 (0.3-0.7)		79 (75-87)	52 (27-80)	93 (82-100)	14, 30, 31
	Modelo boosted generalizado	GBM	17				76 (61-89)	51 (24-85)	79 (61-93)	15
	Modelo de persistencia	PM	22				78 (64-92)	24 (0-42)	86 (75-96)	15,17
Modelos de Clasificación	Redes Neuronales Artificiales (ANN, NIO, NAR, WA-NAR)	ANNs (ANN, NIO, NAR, WA-NAR)	11				90 (85-939	66 (54-79)	95 (90-100)	5, 32, 33,39
	Árboles de decision	CART, RF	14				86 (73-93)	57 (24-86)	91 (71-98)	24, 25, 30, 34, 40
	Regresión Logística Binaria	BLR	9				78 (41-94)	35 (0-86)	90 (69-100)	20, 24, 35

1- Herrig et al., (2015), 2- de Souza et al., (2018), 3- Frick et al., (2008), 4- Zhang et al., (2015), 5- Chandramouil et al., (2007), 6- Choi and Bae (2018), 7- García-Alba et al., (2019), 8- Lin et al., (2003), 9- Lin et al., (2008), 10- Park et al., (2018), 11- Zhang et al., (2012), 12-Mas and Ahlfeld (2007) 13- Bae et al., (2010), 14- Harai and Porto., (2016), 15-Brooks et al., (2016), 16- Eleria and Vogel (2005), 17- Francy et al., (2013), 18- Gonzalez et al., (2012), 19- Heberger et al., (2008), 20- Searcy et al., (2018), 21- Shively et al., (2016), 22- Dada and Hamilton (2016), 23,24 and 25- Thoe et al., (2012, 2014, 2016), 26-Brooks et al., (2013), 27- He and He (2008), 28- Gazzaz et al., (2012), 29- Zhang et al., (2018), 30-Avila et al., (2018), 31- Jones et al., (2013), 32- Jin and Englande (2006), 33- Xu et al., (2020), 34- Choi and Seo (2018), 35- Rossi et al., (2020), 36- Francy et al., (2020), 37- Kashefipour et al., (2005), 38- Bae et al., (2010), 39- Laureano-Rosario et al., (2019), 40- Segura et al., (2021).

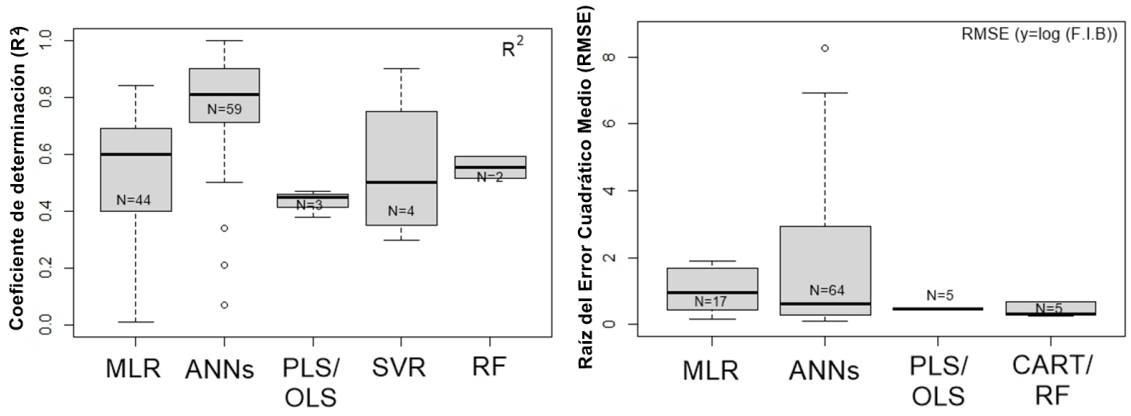

Figura 2.5 Distribución del coeficiente de determinación (R^2) y de la raíz del error cuadrático medio (RMSE), para las Regresiones Lineales Múltiples (MLR), las Redes Neuronales Artificiales (ANNs), Regresión Polinómica y de Mínimos cuadrados parciales (PLS/OLS), regresión de vectores de soporte (SVR), árboles de decisión (CART/RF) evaluados en muestra test. N es el número de modelos incluidos en cada tipo de análisis. La caja representa el rango intercuartil, la barra horizontal la mediana y los bigotes se extienden 1.5 del rango intercuartil.

Implementación operativa de los modelos en SAT

Se registraron 11 programas que implementaron SAT y Nowcasting de calidad del agua en 6 países (Estados Unidos, Escocia, Dinamarca, Australia, Nueva Zelanda y China, Tabla 2.3). Las predicciones de los modelos se presentaron en páginas web junto con datos meteorológicos (por ejemplo, viento, mareas, temperatura, etc.) e información detallada de la playa (playa accesible para sillas de ruedas, transporte público, etc.). Siete de ellos fueron implementados por instituciones gubernamentales encargadas de la gestión de la calidad ambiental (por ejemplo, la Agencia de Protección Ambiental de los Estados Unidos (USEPA), la Agencia Escocesa de Protección Ambiental (SEPA), la Agencia Danesa de Protección Ambiental (DEPA), la Autoridad de Protección Ambiental de Nueva Gales del Sur (NSWEPA) y el Servicio Regional de Salud Pública de Auckland (ARPHS)). Tres se desarrollaron en cooperación entre instituciones gubernamentales y universidades como la Universidad de California (UC) y la Universidad de Stanford (SU). Dos se crearon conjuntamente con organizaciones no gubernamentales (Salvavidas de la Región Norte (SLSNR)) o por iniciativas privadas (por ejemplo, DHI) (Tabla 2.3).

Tabla 2.3 Programas de monitoreo de calidad de playas que implementaron modelos predictivos en Sistemas de Alerta Temprana. Se detalla el país, el programa institucional de calidad de playas, su enlace a la página web. Se presenta el umbral definido como la CCF por la legislación de cada país.

País	Estado/ Ciudad	Institución	Nombre del Programa	Link a la página web	Tipo de Modelo	FIB	Umbral	Referencias
Estados Unidos	Ohio, Pensilvania y Nueva York	USGS[1], USEPA[16]	Great Lakes Nowcast	https://ny.water.usgs.gov/maps/now	MLR	EC	235 EC/100 ml	1
Estados Unidos	California	UC[2], US3	Beach report card	https://beachreportcard.org/	MLR, CT, BLR	EC; ENT	235 EC/100 ml	2
Estados Unidos	Georgia	USGS, USEPA[16][1]	BacteriALERT	https://www2.usgs.gov/water/southatlantic/ga/bacteria/	MLR	EC	235 EC/100 ml	-
Estados Unidos	Carolina del Norte y del Sur, Florida	SECOORA[4], USC[5], UM[6]	How's The Beach?	https://howsthebeach.org/	MLR, BLR	ENT	235 EC/100 ml	-
Estados Unidos	Filadelfia	PWD[7]	PhillyRiverCast	https://www.phillyrivercast.org/	MLR	EC:CF	235 EC/100 ml	3
Estados Unidos	Wisconsin	WDNR[8]	-	https://dnr.wisconsin.gov/topic/Beaches	Árboles de decisión	EC	235 EC/100 ml	4,5
Escocia, Reino Unido	-	SEPA[9]	-	https://www2.sepa.org.uk/bathingwatershttps://www2.sepa.org.uk/bathingwatersthingwaters	MIKE 3 and ECOLab	EC; ENT	500 CFU/100 ml	6
Dinamarca	Copenaghen	DHI[10], DEPA [11]	Badevandsudsigten	http://newkbh.badevand.dk/#om-badevandsprognoser		EC; ENT	500 CFU/100 ml	7
Australia	NSW	NSWEPA [12]	Beachwatch	https://www.environment.nsw.gov.au/" \t "_tophttps://www.environment.nsw.gov.au/" \t "_topau	GLM, MLR	EC; ENT	500 CFU/100 ml	8

Nueva Zelanda	Auckland	ARPHS[13] SLSNR[14]	Safeswim	https://safeswim.org.nz	MLR 'black box'-'white box'	ENT	280 MPN/ 100 ml	9
China	Hong Kong	HKU[15]	Waterman	http://www.waterman.hku.hk	MLR	EC	610 EC/100 ml	10

1-Fancy et al., (2020) 2- Thoe et al., (2014), 3- Maimone et al., (2007), 4- Mednick (2012), 5- Cyterski et al., 2012, 6- Stidson et al., (2012), 7-Mark y Erichsen (2021), 8- Hose et al., (2005), 9- Dada et al., (2019), 10- Thoe et al., (2018).

1-Servicio Geológico de los Estados Unidos (USGS), 2- Universidad de California (UC), 3- Universidad de Stanford (US), 4- Asociación Regional de Observación de los Océanos Costeros del Sudeste (SECOORA), 5- Universidad de Carolina del Sur (USC), 6- Universidad de Maryland (UM) 7- Departamento de Agua de Philadelphia 8- Departamento de Recursos Naturales de Wisconsin, 9- Agencia Escocesa de Protección Ambiental (SEPA), 10- DHI, 11-Agencia Danesa de Protección Ambiental Agencia de Protección Ambiental (DEPA), 12- Autoridad de Protección Ambiental de Nueva Gales del Sur (NSWEPA), 13-Servicio de Salud Pública Regional de Auckland (ARPHS), 14- Salvavidas de la Región Norte (SLSNR), 15- Universidad de Hong Kong (HKU), 16-Agencia de Protección Ambiental de los Estados Unidos (USEPA).

La mayoría de los programas detallados en la Tabla 2.3 utilizaron algoritmos estadísticos para obtener predicciones de CCF (por ejemplo, MLR, GBM, árboles de decisión, etc.), mientras que DHI desarrolló un modelo hidrodinámico (MIKE 3 y Ecolab) (Tabla 2.3). Los modelos estadísticos revisados utilizaron *Escherichia coli* (EC) y Enterococos (ENT) como FIB, y los umbrales permitidos de CCF para designar si una playa es de buena calidad o no varió según la legislación vigente en cada país.

2.4 DISCUSIÓN

Esta revisión evidenció que en la última década se registró un aumento del número de países que utilizaron modelos predictivos de CCF y también un aumento en la diversidad de técnicas de modelización. Se identificaron tres cuestiones críticas que requieren mayor atención, como el preprocesamiento de datos, la forma de evaluación de la capacidad predictiva de los modelos y una lenta transferencia de información sobre cómo utilizar los modelos a los gestores de las playas. Los resultados evidencian que las técnicas de aprendizaje automático mejoran la capacidad de predicción de la CCF, pero

esto requiere una estrategia de selección de variables y la aplicación de métodos para tratar el desbalance en los conjuntos de datos, siendo este último un problema importante para la modelización de la CCF (Bourel et al., 2021). En esta revisión se ampliaron los resultados de revisiones anteriores que analizaron diferentes tipos de estrategias de modelado predictivo (de Brauwere et al., 2014) y de modelos predictivos en ambientes de agua dulce (Heasly et al., 2021). La integración de los sistemas de monitoreo con la modelización y la implementación de SAT es necesaria para mejorar la difusión de la información acerca de las predicciones de CCF y los excesos a las normativas, a los responsables de la toma de decisiones y a los usuarios de las playas para proteger su salud.

En los últimos veinte años se han desarrollado con relativo éxito diversos tipos de modelos predictivos de CCF en playas recreativas alrededor del mundo. Se desarrolló una amplia gama de modelos para los lagos de Estados Unidos, lo que explica la sobrerrepresentación de los ambientes continentales respecto a los costero-marinos en este trabajo. Hubo una diversificación de estrategias de modelado, desde Regresiones Lineales Múltiples que dominaron a principios del 2000 (de Brauwere et al., 2014; Heasley et al., 2021) hasta técnicas de AA como Redes Neuronales Artificiales y árboles de decisión (CART y Random Forest) (Ávila et al., 2018; Choi y Seo, 2018; Thoe et al., 2016). Estos últimos han mostrado una mejor capacidad predictiva para la CCF similar a lo encontrado para problemas complejos en diferentes campos del conocimiento (Voyant et al., 2017; Liu et al., 2018; Aguilera-Venegas et al., 2023). La modelización Bayesiana y el Análisis Wavelet también se han hecho más populares en la última década, pero con una aplicación limitada para la predicción de CCF (Ávila et al., 2018; Seis et al., 2018; Zang et al., 2018). Esta tendencia, de aumento de implementación de modelos de AA, estuvo acompañada del aumento de la potencia de los computadores y la utilización de software estadísticos gratuitos como R o Python.

El desempeño de los modelos se evaluó utilizando diferentes métricas, principalmente el R^2 y el RMSE para los métodos de regresión, y las métricas derivadas de la matriz de confusión (precisión, sensibilidad, especificidad) para los métodos de clasificación. Los resultados de esta revisión revelan que los modelos de AA (ANNs, y basados en árboles de decisión) se desempeñan mejor en términos de precisión que las MLR, como también lo indicaban trabajos anteriores (Thoe et al., 2012, 2014; Zhang et al., 2018). En cuanto a los métodos de clasificación, la sensibilidad fue menor que la

especificidad para todos los tipos de modelos, y tanto las ANNs como los árboles de decisión presentaron una especificidad mayor que el resto de modelos. La sensibilidad de un modelo representa la capacidad de predecir un exceso a los umbrales permitidos de CCF y, por lo tanto, es fundamental para proteger a los usuarios de condiciones riesgosas de contaminación. La especificidad, en cambio, refleja la capacidad de predecir los días sin excesos, los cuales, en general, suelen ser la mayoría de las observaciones. Existe un compromiso entre la sensibilidad y la especificidad que debe manejarse con cuidado, ya que un aumento de las falsas alarmas (es decir, el cierre de la playa cuando es seguro) puede generar un descrédito del modelo y del SAT en general por parte de las y los usuarias/os. La mayoría de los estudios de modelización presentan datos desbalanceados que requieren técnicas de balanceo para mejorar su desempeño (Bourel et al., 2021). Algunas de estas técnicas incluyen el tratamiento de la ponderación positiva de la clase minoritaria o la manipulación de los conjuntos de datos de entrenamiento para equilibrar la proporción de clases mediante la replicación de casos de la clase rara (ej. upsampling), la eliminación aleatoria de casos para la clase mayoritaria (ej. downsampling) o la generación sintética de casos en la clase rara (ej. SMOTE). Estas técnicas han sido evaluadas y han demostrado su utilidad para superar el problema del desbalance, pero su aplicación en la modelización de CCF sigue siendo escasa (Bourel et al., 2021; Watthaisong et al., 2024).

La modelización predictiva de la CCF en ambientes dinámicos como ríos, arroyos o playas costeras, presenta grandes desafíos. Históricamente se ha utilizado como estándar a nivel mundial el "Modelo de Persistencia" o "Modelo naif", que consiste en asumir que tanto las concentraciones de FIB como las condiciones ambientales actuales son las mismas que las del día anterior. Las estrategias de modelado predictivo mejoran las métricas de predicción con respecto a los modelos de persistencia (Searcy et al., 2018; Francy et al., 2020; Avila et al., 2018) y, por ende, aumentan la capacidad para anticiparse a eventos de CCF. La concentración de FIB en las playas puede ser muy variable, incluso en 24 o 48 horas, que es el tiempo necesario para la incubación de bacterias en el laboratorio. Los resultados actuales sugieren que los modelos de persistencia no son aconsejables como método para predecir la concentración de FIB en playas recreativas (Whitman et al., 2004; Whitman y Nevers, 2008). Sin embargo, la inclusión de concentraciones pasadas de FIB en los modelos predictivos parece ser una variable importante que ayuda a mejorar el desempeño de los modelos (Thoe et al., 2012).

El desempeño de un modelo predictivo de CCF depende también de la selección de las variables de entrada (Francy et al., 2020). Las mediciones *in situ* de la calidad físico- química y biológicas del agua (por ejemplo, la temperatura del agua) y las variables meteorológicas (por ejemplo, las condiciones previas de precipitación (acumulada, 24, 48 y 72 hs)) fueron las variables más importantes para explicar la presencia de CCF, ya sea en dominios de agua dulce o costeros. Esto se explica por el hecho de que los sistemas de saneamiento son en su mayoría unitarios, es decir, las aguas residuales y los pluviales se encuentran en los mismos sistemas de drenaje y, por lo tanto, los episodios de precipitaciones dirigen las aguas residuales a las zonas costeras. Esto apunta a la ineficacia de las plantas de tratamiento y a la necesidad de mejorar y replantear los sistemas de saneamiento tradicional. La densidad de población y el nivel de urbanización, por otra parte, casi no se tuvieron en cuenta en el proceso de construcción de los modelos, lo que podría explicarse porque en el marco temporal de un programa de monitoreo no se producen cambios sustanciales en esos atributos demográficos. Sin embargo, podrían ser variables importantes a incluir en un futuro.

La implementación de modelos predictivos en los sistemas de gestión de playas recreativas es altamente recomendada por instituciones de salud pública, pero los SAT aún son escasos en comparación con la variedad de modelos desarrollados en la literatura científica. Se encontró una clara falta de modelos predictivos operacionales para Centro y Sudamérica, aunque, recientemente, se registran avances hacia la implementación de modelos en estas áreas del mundo (de Souza et al., 2018; Segura et al., 2021). La implementación de un protocolo completo de alerta temprana, que incluya componentes de asesoramiento como apps, carteles, etc., reduce los riesgos sanitarios de la exposición humana a la CCF en playas recreativas (Searcy et al., 2018). Estas implementaciones operativas de modelos predictivos requieren la automatización de algunos procesos, como la recopilación de datos para generar predicciones, la actualización continua de las bases de datos, el acceso rápido a nuevos datos y las interfaces fáciles de usar para que los responsables de la gestión de las playas implementen las predicciones de los modelos en la toma de decisiones. La adición de datos remotos, por ejemplo, datos de satélite (por ejemplo, temperatura de la superficie del mar, turbidez, etc.) mejoran los modelos de AA (Stampoulis et al., 2020; Shirmard et al., 2022) y pueden proporcionar datos frecuentes para combinarlos con programas de seguimiento *in situ* (Laureano-Rosario et al., 2019). Sin embargo, esta estrategia rara vez se ha aplicado para predecir la CCF. En este sentido,

la instrumentación específica de boyas hidrográficas de calidad del agua, estaciones meteorológicas, telemetría y programas informáticos son útiles para mejorar el desempeño, la rapidez y la facilidad de implementación de los modelos (Francy et al., 2009; Thoe y Lee, 2014).

2.5 CONCLUSIONES

En general, en las dos últimas décadas se han hecho grandes avances en las estrategias de modelización predictiva de CCF. Sin embargo, se puede perfeccionar la capacidad predictiva mediante la utilización de técnicas que permitan tratar el desbalance entre clases. En todo el mundo se utiliza una amplia gama de estrategias de modelización, pero sólo unas pocas están implantadas en SAT. Los SAT deberían ampliarse a otras zonas geográficas y, fundamentalmente, a las zonas costero-marinas y estuarinas. La integración debería incluir tres ejes principales interconectados: *i)* el monitoreo *in situ* de la calidad del agua, *ii)* la generación de modelos predictivos precisos y *iii)* un SAT con una interfaz web interactiva de fácil acceso y uso. La implementación de los modelos en los SAT, en complemento con monitoreos a largo plazo, son críticos para prevenir el contacto de los bañistas con aguas contaminadas.

CAPÍTULO 3

IMPLEMENTACIÓN DE MODELOS PREDICTIVOS EN LA GESTIÓN DE LAS PLAYAS DE USO RECREATIVO EN MONTEVIDEO

3.1 INTRODUCCIÓN

En el contexto de la predicción de contaminación coliformes fecales (CCF) en playas, la implementación de modelos predictivos en los sistemas de gestión de calidad de playas es una medida recomendada por la OMS (OMS, 2018; 2021). Los modelos que históricamente han sido mayormente desarrollados en la literatura para la predicción de CCF han sido las Regresiones Lineales Múltiples (Brauwere et al., 2014; Heasley et al., 2021) pero conforme avanza la capacidad de cómputo y el desarrollo de más y mejores técnicas de modelización predictiva, se comenzaron a implementar distintos modelos de Aprendizaje Automático (AA). Los modelos de AA son ampliamente utilizados actualmente en distintos campos del conocimiento y específicamente en las ciencias ambientales (Zhong et al., 2022). Son poderosas técnicas estadísticas que dan cuenta de complejidad de los datos y han demostrado buena capacidad predictiva en playas recreativas (Zhong et al., 2022; Tahmasebi et al., 2020). Se han desarrollado una gran diversidad de modelos de AA y de estrategias de modelización para mejorar la capacidad predictiva de los modelos de contaminación. Las técnicas de AA supervisado son las más comunes (Capitulo 2) y buscan predecir una variable de salida o de respuesta *Y* (relacionada a la CCF) en función de variables de entrada o predictoras *X*. Estos modelos pueden ser de clasificación, cuando su objetivo es predecir una variable categórica con al menos dos categorías (por ejemplo, playa cerrada o abierta, o excede o no excede un umbral determinado de CF) o de más de dos categorías (verde, naranja, rojo), o de regresión, si se busca predecir un valor en específico de CCF (ej. concentración de CF). Entre los más utilizados y con mejor desempeño para la predicción de CCF se encuentran las Redes Neuronales Artificiales y los basados en arboles de decisión (CART, RF) (Heasley et al., 2021; Capitulo 2 de esta tesis).

Una gran parte de la literatura dirigida al desarrollo de modelos predictivos de CCF ha sido desarrollada para ambientes de agua dulce (lagos y lagunas) y para ambientes costero-marinos. La modelización predictiva de CCF en zonas estuarinas está poco representada en la literatura a nivel mundial (Heasley et al., 2021; Capítulo 2 de esta tesis). Esto se debe a que la modelización en este tipo de ambientes presenta desafíos debido a que estos ecosistemas son altamente dinámicos, en donde confluyen las aguas de fuentes continentales con aguas marinas, que, a su vez, son influenciados por el viento, las mareas, las precipitaciones, entre otros (González et al., 2012; García-Alva., 2019).

Las variables input o predictoras que se utilizan para la modelación predictiva de CCF en ambientes estuarinos generalmente se basan en medidas *in situ* (Jin y Englade, 2006; Lin et al., 2008). Típicamente la temperatura del agua, del aire, la turbidez y la salinidad. El incremento de la salinidad tiene en general una relación negativa con la CCF, es decir, a mayor salinidad menor concentración de CCF (González et al., 2012). La turbidez, por otra parte, suele relacionarse de forma directa y positiva con la CCF en este tipo de ambientes (García-Álava et al., 2019). El nivel de las precipitaciones, por otra parte, tiene una fuerte influencia en las dinámicas de la CCF (Hose et al., 2005). En general se utilizan los datos de las precipitaciones acumuladas de los días previos, ya sea 24, 48, 72hs previas. Las precipitaciones tienen influencia en la concentración de la CCF ya sea por elevar el nivel de descargas de los afluentes de agua dulce (ríos, arroyos y cañadas) que llegan a las playas y dispersan la CCF de diversas fuentes difusas (González et al., 2012; García-Álava et al., 2019), o porque en algunos casos, los sistemas de saneamiento, al ser unitarios (las mismas cañerías que desagotan los pluviales también conducen las aguas negras de los hogares), descargan CCF en días de precipitaciones elevadas (Segura et al., 2021).

Variables remotas como las que se derivan de productos satelitales en cambio, aún no son implementadas con frecuencia en la construcción de los modelos predictivos de CCF, a pesar de que han empezado a utilizarse conforme aumentan la cantidad de bases de datos y tipos de información que se puede obtener de los mismos (Laureano-Rosario et al., 2019). Dentro de las variables utilizadas para la predicción de CCF, la temperatura superficial del agua (SST por su sigla en inglés) es una de las de más fácil acceso a nivel global gracias a bases de datos de acceso libre provistas por la Administración Nacional de Aeronáutica y el Espacio de Estados Unidos (NASA) o la Oficina Nacional de Administración Oceánica y Atmosférica (NOAA, por su sigla en inglés). Además, por la naturaleza de los sensores, este producto satelital permite obtener datos donde otros tipos de sensores tienen dificultad, por ejemplo, por la presencia de nubosidad (Huang et al., 2017). Otras variables que pueden obtenerse de bases de datos de productos satelitales, y que tienen influencia en la presencia de CCF, son algunos *proxis* de la turbidez (ej. Kd490), la irradiación solar o irradiancia y aquellos que se relacionan con el color del agua como la Clorofila-a que, generalmente, se relaciona con floraciones de fitoplancton. La implementación de variables satelitales en el ámbito del desarrollo de modelos

predictivos de CCF, sin embargo, ha sido escasamente implementada hasta el momento (Laureano-Rosario et al., 2019).

La implementación de modelos predictivos para las playas de Montevideo registra importantes esfuerzos. En el marco de un trabajo conjunto entre la Unidad de Calidad de Agua del SECCA de la IM y el Departamento de Modelización Estadística de Datos e Inteligencia Artificial (CURE, Rocha) se llevó a cabo un proyecto que evaluó la capacidad predictiva de distintos tipos de modelos y mejoró en un 60% la capacidad predictiva respecto al criterio actual de cierre de playas las 24 horas posteriores a precipitaciones (Segura et al., 2021). Se entrenaron varios algoritmos para la predicción de CF en playas capitalinas que incluyeron: Análisis Lineal Discriminante (LDA), Máquinas de Vectores Soporte (SVM), Modelos Lineales Generalizados (GLM), árboles de clasificación y regresión (CART), y métodos de combinación de árboles: Adaboost y bosques aleatorios (RF). Dentro de éstos, los algoritmos de RF fueron los de mejor capacidad predictiva (Bourel et al., 2021). En la etapa de construcción de modelos se identificaron desafíos y líneas de avance en el desarrollo de los modelos y en la implementación de los mismos en el sistema de gestión de la IM. Las principales oportunidades que se identificaron fue la de incorporar variables remotas para contar con mayor frecuencia y cobertura espacial de datos. La incorporación de variables tomadas a partir de datos satelitales permite una frecuencia diaria y según el producto y con una amplia cobertura espacial y resolución (Jutz y Milagro-Pérez, 2018; Martinis et al., 2021). Algunas variables importantes para la modelización predictiva de CCF como la temperatura del agua o la turbidez, son variables que disponen los productos satelitales y podrían implementarse en modelos para las playas capitalinas. Por otra parte, se identificaron algunos desafíos, como, por ejemplo, la estructura desbalanceada de los datos (el porcentaje de días en los que se excede el umbral es mucho menor (aprox. 8%) que el porcentaje de días en los que no se excede el umbral (92%). Esto genera una dificultad durante el entrenamiento de los modelos, que debe ser abordada para generar predicciones más precisas (Bourel et al., 2021).

El estuario del Río de la Plata, es uno de los estuarios más grandes del mundo y recibe los aportes de los Ríos Uruguay y Paraná, que conforman la segunda cuenca hídrica más grande de Latinoamérica (Guerrero y Piola, 1997). Este estuario puede dividirse en tres regiones: una región interna, una región media y otra externa (Figura 1.1). Las dos

últimas pueden ser distinguidas por un límite transversal en donde se encuentran las aguas turbias de la descarga del Río de la Plata, con el agua más clara y salina proveniente del océano Atlántico. Allí se observa un cambio repentino en la batimetría, y los frentes de salinidad y turbidez típicamente se localizan en esta región, normalmente llamado Frente de Turbidez del Río de la Plata (Parker y López-Laborde, 1989; Framiñan y Brown, 1996; Piola et al., 2000; Maciel et al., 2018). Las dinámicas espacio-temporales de este frente repercuten en las características físico-químicas de las aguas que ocurren en la zona costera de Montevideo y por lo tanto en las dinámicas (tiempo de residencia, mortandad, decaimiento, etc.) de la CCF (IM, 2023).

Montevideo, localizada en la porción media de la cuenca del estuario del Rio de la Plata, presenta una gran complejidad del sistema de playas debido a la rica dinámica entre las aguas marinas y fluviales que repercuten en los gradientes de salinidad y turbidez entre playas (Nagy et al., 1997; Brugnoli et al., 2018; Segura et al., 2021). La zona costera de Montevideo se encuentra limitada entre los Ríos Santa Lucia hacia el Oeste y al Este por el Arroyo Carrasco. Además, tiene aportes de arroyos como el Miguelete, el Pantanoso y el Seco, y de pequeñas cañadas que desembocan en las playas (Figura 1.1). Además, se suma la descarga de efluentes cloacales proveniente de la red capitalina producidos por un millón y medio de personas (Álvarez, 2019; Artelia et al., 2019). El 60% del sistema de saneamiento de Montevideo es unitario donde las aguas residuales se conducen por las mismas cañerías que el agua de las precipitaciones que en eventos de precipitaciones, al saturar la red, el agua desagota por caños aliviaderos directamente en las playas (Plan Nacional de Saneamiento, 2019). La zona Oeste de Montevideo no cuenta con conexión unitaria por lo que se generan aportes difusos de las cañadas y las napas a las playas, que generan eventos de CCF (IM, 2023) (Figura 1.1).

La Intendencia de Montevideo (IM) a través de la Unidad de Calidad de Agua del SECCA cuenta con un sistema de monitoreo que comprende 21 playas a lo largo de toda la costa del departamento (Figura 1.1; IM, 2023). Dichos monitoreos se realizan con una frecuencia de dos veces por semana en la temporada no estival (abril a noviembre) y de dos a cuatro veces por semana en la temporada estival (noviembre a marzo). El monitoreo registra variables fisicoquímicas del agua como la temperatura, la salinidad, la turbidez, pH, y conductividad y se toman muestras de agua de las playas para realizar estudios bacteriológicos y estimar la concentración de CF.

A largo plazo, por ejemplo, en informes anuales de calidad de playas de la Unidad de Calidad de Agua de la IM, las decisiones respecto a habilitar o no una playa se realizan con un criterio en conjunto con el Dirección Nacional de Control y Evaluación Ambiental (DINACEA, 2021) elaborados en la Guía para definir la aptitud y la categorización de las playas. Esta guía establece que la Media Geométrica deberá ser calculadas a partir de cinco muestras consecutivas, tomadas dentro de un período de tiempo de 45 días (MG5) y no deben exceder las 1000 UFC/1000ml. Este procedimiento implica que la habilitación o no de una playa responda a criterios históricos. Algunas playas, por ejemplo, Puerto del Buceo y playa Miramar se encuentran permanentemente inhabilitadas para baño debido a que antecedentes históricos indican que no presentan condiciones homogéneas durante la temporada, pudiendo aparecer eventualmente valores puntuales muy superiores a los límites que indica la reglamentación vigente. En dicho informe anual se publican Medias Geométricas y las Medias Geométricas Móviles (MG5) de toda la temporada (no estival y estival por separado) para cada playa (IM, 2023). Los días que se tienen en cuenta para realizar estos cálculos son aquellos considerados "representativos" (es decir, "aquellos en los que no se registraron vertimientos ocasionados por precipitaciones siempre que las mismas hubieran ocurrido dentro de las 24 horas anteriores al muestreo", IM, 2023). Debido a que en Montevideo el saneamiento es unitario cuando se registran precipitaciones acumuladas altas se registran eventos de CCF. La IM, por lo tanto, recomienda no bañarse en las 24hs posteriores a las precipitaciones y son días considerados no representativos (IM, 2023).

A mediano plazo, los resultados de los monitoreos son publicados semanalmente (en la temporada estival) en la página oficial de la IM través del sitio web institucional: https://montevideo.gub.uy/areas-tematicas/ambiente/evaluacion-de-la-calidad-del-agua-en-lasplayas/informe-semanal-de-calidad-del-agua-de-las-playas-de-montevideo. Estos resultados se basan en los resultados de los monitoreos de la semana previa. La concentración de CF que presentan las playas se compara respecto a la normativa del Decreto 253/79 que clasifica los cuerpos de agua según sus usos y que asigna la Clase 2b a aquellos cuerpos de agua de "recreación por contacto directo" (Decreto 253/79). Este decreto indica para esa categoría un umbral de CF de 2000 UFC/100ml. Las playas que son habilitadas para el baño se indican en una tabla en color verde. Las playas que presentan irregularidades en sus valores se colocan en un nivel de "alerta" en color naranja con la advertencia "momentáneamente no se recomienda hacer uso para

recreación". Por último, en color rojo se muestran playas inhabilitadas para el baño permanentemente por presentar concentraciones de CF sistemáticamente por encima del umbral permitido (Puerto del Buceo y Miramar) (IM, 2023).

Para alertas en periodos más cortos (diarios), durante la temporada estival, en Montevideo existe la posibilidad de la colocación de la Bandera Sanitaria por parte del Servicio de Guardavidas cuando ocurren eventos que este servicio considere que puedan implicar riesgo sanitario para la población (IM, 2023). Si bien la IM cuenta con un sistema de calidad de playas de monitoreo frecuente y que presenta las playas habilitadas y no habilitadas para baño de forma semanal en su página web, aún no cuenta con un sistema dinámico sitio-específico. Los resultados se muestran de forma estática en tablas con un código de colores en un archivo PDF, indicando cuáles playas están habilitadas y cuáles no según los registros previos.

Por lo tanto, los objetivos de este capítulo son:

1. Incorporar variables satelitales en los modelos de predicción de CCF para playas de Montevideo y evaluar su desempeño.

2. Desarrollar una interfaz web para implementar de forma accesible los modelos de predicción de coliformes fecales desarrollados para las playas de Montevideo.

3.2 MÉTODOS

Base de datos

Se utilizó una base de datos provista por la Unidad de Calidad de Agua del SECCA de la IM que abarcó una ventana temporal desde el 15 de noviembre de 2009 hasta el 03 de marzo de 2023. Las playas seleccionadas para la modelación predictiva fueron 20 en total. Dichas playas son las que incluyen para su monitoreo durante la temporada estival (Tabla 3.1). La frecuencia de registro de los datos fue de dos a cuatro veces por semana según los monitoreos realizados por la IM.

La base de datos incluyó información descriptiva correspondiente a los muestreos: fecha y hora del muestreo, ubicación, muestreo (Oblicuo, No Oblicuo), tipo de muestreo

(Representativo, No Representativo), nombre de la playa, código de la playa, coordenadas del sitio de muestreo en cada playa, el periodo de muestreo (estival, no estival), presencia de espuma de cianobacterias (si/no), información de variables ambientales medidas *in situ* como la temperatura del agua (°C), la salinidad, la conductividad, la turbidez (NTU), pH, oxígeno disuelto (ppm) y variables cuantificadas *a posteriori* en el laboratorio como los nutrientes Nitrógeno Total (mg/l) y Fósforo total (mg/l), la Clorofila-a (µg/l), microcistinas totales (µg/l), feopigmentos, y la concentración de coliformes fecales (UFC/100ml) (IM, 2023). La concentración de coliformes es estimada siguiendo los procedimientos estandarizados de operación del Laboratorio Ambiental de DINAMA (Ministerio de Ambiente, Dirección Nacional de Medio Ambiente, 2017), que se basa en la técnica de filtración por membrana de APHA (APHA, 1989; IM, 2023).

Por otra parte, se compilaron bases de datos de variables meteorológicas como la precipitación (mm), la temperatura del aire (°C), la humedad relativa, la velocidad y la dirección del viento (km/h). Estos datos fueron tomados de las páginas web del Instituto Nacional de Investigación Agropecuaria (INIA), y de Meteomanz (http://meteomanz.com/). A partir de la variable de precipitación se calculó la precipitación acumulada de 1 a 7 días para todas las playas.

Se extrajeron variables derivadas de productos satelitales utilizadas para la modelización predictiva de CCF. La temperatura superficial del agua, la turbidez satelital (Kd490) e intensidad del viento fueron extraídas de la base de datos de la Agencia Nacional de Administración Oceánica y Atmosférica de Estados Unidos (NOAA) (https://coastwatch.pfeg.noaa.gov/erddap/search/). En esta plataforma se accedió a los datos georeferenciados diarios de temperatura superficial del agua, utilizando el producto Multi-scale Ultra-high Resolution (MUR) Sea Surface Temperature (SST por su acrónimo en inglés). Se seleccionó una base de datos diaria, con una grilla de 0.01° (≈1km), que incluyó datos desde el 2002 hasta el presente. Por otra parte, se utilizó el coeficiente de atenuación difusa a 490 nm (Kd490) como *proxy* de la turbidez (Science Quality, Global 4km, Level 3, 2012-present, Daily). Esta variable es tomada por el sensor Moderate Resolution Imaging Spectroradiometer (MODIS), a bordo del satélite Terra, e indica a que tasa se atenúa la luz a una longitud de onda de 490 nanómetros con la profundidad y es tomada como aproximación de turbidez del agua. La grilla utilizada fue de resolución 0.04°. Finalmente, se incluyó la velocidad del viento extraída del proyecto

“Radiómetro avanzado de muy alta resolución, Pathfinder (AVHRR, por sus siglas en inglés). Esta base de datos proviene de la NOAA y contiene datos, tanto de SST como de velocidad del viento, desde 1981 hasta el presente con una grilla de resolución 0.0417°. Estas tres variables fueron extraídas de forma separada para las playas del Este entre las coordenadas de latitud -34.89375 y -34.96875 y longitud -56.15625 y -56.006249. Y para las playas del Oeste en latitud -34.8187499 y -34.96875 y longitud -56.493749 y -56.15625 (Figura 3.1). La ventana temporal seleccionada para todas las variables fue la misma que la disponible en la base de datos de variables *in situ*, desde el 15 de noviembre de 2009 hasta el hasta el 03 de marzo de 2023. Con excepción del indicador Kd490, cuya base de datos estaba disponible desde el 2012.

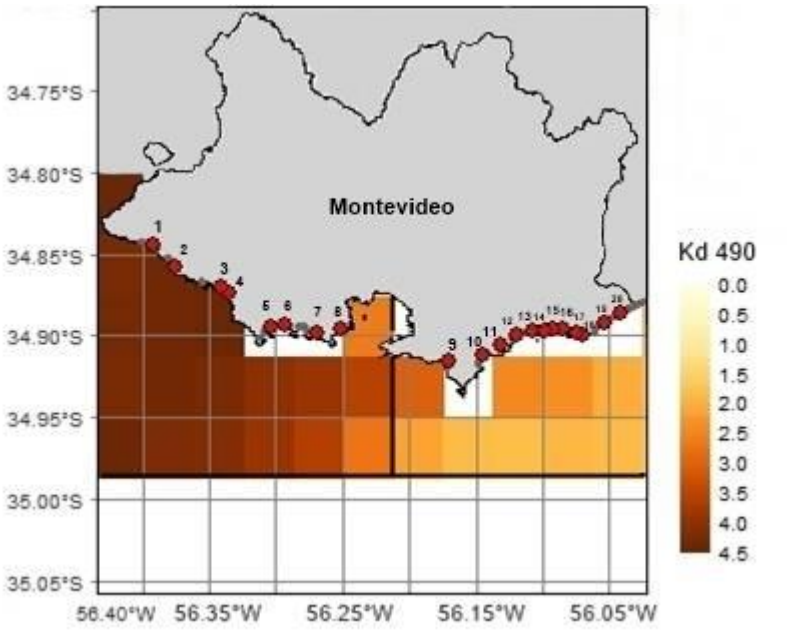

Figura 3.1 Mapa del departamento de Montevideo. En recuadros negros se indica la grilla espacial seleccionada para las variables satelitales (a modo de ejemplo se ilustra la grilla seleccionada para la variable Kd490). En puntos rojos numerados se muestran las 20 playas utilizadas para el análisis de datos y la modelización: **Al Oeste**: 1- Punta Espinillo, 2- La Colorada, 3- Pajas Blancas, 4- Zabala, 5- Punta Yeguas, 6- Santa Catalina, 7- Nacional, 8- Cerro. **Al Este**: 9- Ramírez, 10- Pocitos, 11- Puerto del Buceo, 12- Buceo, 13-Malvín, 14- Brava, 15- Honda, 16- ingleses, 17- Verde, 18- Mulata, 19- Carrasco, 20- Miramar.

Análisis de Datos

Se realizaron análisis descriptivos del comportamiento de las playas en base a algunas variables ambientales (temperatura del agua, salinidad, turbidez) en cada playa del Este y Oeste, se determinó el porcentaje de excesos de CF al umbral permitido de cada playa, y el porcentaje de excesos sobre los muestreos totales. Se construyeron gráficos de Boxplot de algunas variables *in situ* como temperatura del agua (°C), salinidad, turbidez para cada playa. Y gráficos de serie temporal de CF para cada playa, y de las variables satelitales analizadas SST y el Kd490. Se registraron datos descriptivos como la media de CF, el máximo, mínimo, desvío estándar de CF para cada playa. Se quitaron los datos faltantes (NAs).

Para evaluar si existían diferencias significativas respecto a las variables satelitales SST, Kd490 y velocidad del viento tanto, entre playas del Este y el Oeste y las variables *in situ* y las satelitales se realizaron test de correlaciones entre las playas localizadas al Este y Oeste del departamento de Montevideo. Para esto se seleccionaron tres playas consideradas como representativas para el Este y para el Oeste. Estas playas fueron seleccionadas por el número de datos disponible por playa (aquellas con las que se contaba con mayor cantidad de datos). Las playas seleccionadas como representativas para el Este fueron Pocitos, Carrasco y Malvín y para las playas del Oeste fueron Cerro, Santa Catalina y Pajas Blancas. También se evaluó la correlación entre las variables tomadas *in situ* y las satelitales tanto para las playas del Este como del Oeste. Las correlaciones evaluadas fueron entre la temperatura del agua *in situ* y la SST, entre la turbidez *in situ* y el Kd490 y entre la velocidad del viento tomada por las centrales meteorológicas y la velocidad del viento estimada por satélite. Estas correlaciones fueron realizadas utilizando la función “cor.test” del paquete “Stats” de RStudio (R Core Team, 2023). Se utilizaron correlaciones de Pearson y estas correlaciones se evaluaron según el p-valor y coeficiente de correlación (r).

Estrategia de modelación y evaluación de desempeño

Para evaluar el desempeño predictivo de nuevos modelos incorporando variables satelitales se tomó como línea de base los trabajos de Segura et al., (2021) y Bourel et al., (2021) en los cuales se analizaron diversas estrategias de modelización y se presentaron los modelos de mejor capacidad predictiva. En dichos trabajos, además, se presentaron

las variables de mayor relevancia para la predicción de CF en las playas de Montevideo. Los modelos de mejor desempeño fueron árboles de clasificación Random Forest y las variables inputs de mayor relevancia predictiva fueron la playa, la concentración de CF ("Lags") de 1, 2 y hasta 3 muestreos previos, la turbidez, la salinidad, y la temperatura del agua tomada *in situ* y las precipitaciones de 48 y 24 hs previas a los muestreos.

Los Random Forest (RF) son árboles de decisión y están dentro de las técnicas de AA supervisado. La idea básica detrás de los árboles de decisión consta en estratificar o segmentar el espacio de los predictores en un número determinado de subregiones. Para hacer una predicción para una determinada observación, normalmente se utiliza la media o la moda de las observaciones de entrenamiento en la región a la que pertenece. Dado que el conjunto de reglas de división utilizadas para segmentar el espacio predictor se puede resumir en un árbol, estos tipos de enfoques se conocen como métodos de árboles de decisión (Cutler et al., 2007; James et al., 2013). Es un procedimiento para predecir los valores de variables categóricas (árboles de clasificación) o continuas (árboles de regresión) en base a variables predictoras (continuas y/o categóricas). También son útiles para explorar o entender la importancia de variables y su interrelación dada su salida grafica de fácil interpretación (Breiman et al., 1984; 2001).

Los modelos de bosque aleatorio o RF por su nombre en inglés, son un tipo de algoritmo que combina las predicciones de varios árboles de decisión obtenidos de muestras boostrap del set de datos (una técnica estadística para obtener muestras de una población donde una observación se puede considerar en más de una muestra) y en donde se sortea en cada rama un número bajo de variables explicativas. Este algoritmo creado a principios de los 2000 por Breiman, (2001) se basa en un consenso de la predicción de N arboles generados a partir de N muestras boostrap (Figura 3.2). La predicción final es, por lo tanto, un promedio, o un consenso de la predicción de N árboles. Las observaciones no estimadas en los árboles (también conocidas como "out of the bag") se utilizan para validar el modelo. Las salidas de todos los árboles se combinan en una salida final Y que se obtiene mediante alguna regla (generalmente el promedio, en RF de regresión y, conteo de votos, en RF de clasificación). Este tipo de estrategia de combinación de árboles ha demostrado que incrementa significativamente la precisión de las predicciones (James et al., 2014). Estos modelos han sido implementados para la predicción de CCF en playas

recreativas con buena capacidad predictiva (Ávila et al., 2018; Choi y Seo., 2018; Segura et al., 2021)

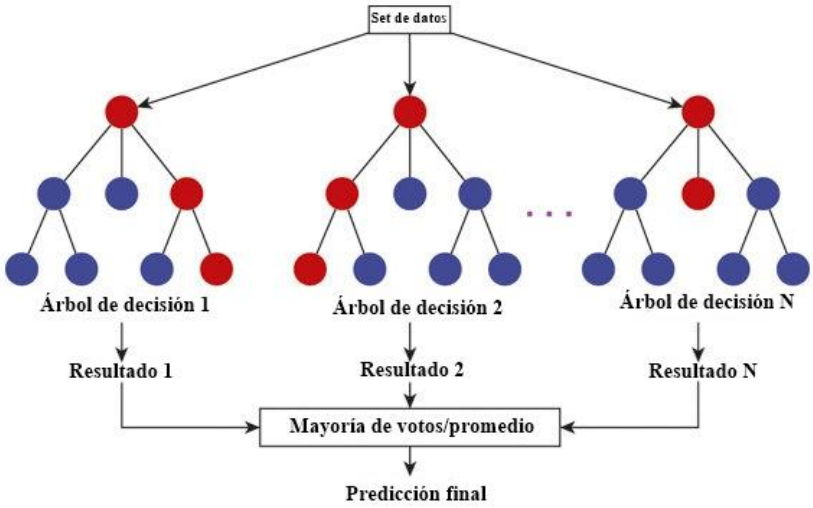

Figura 3.2 Funcionamiento del algoritmo de Bosque Aleatorio o Random Forest (RF). Se generan N muestras boostrap (una técnica estadística para obtener muestras de una población donde una observación se puede considerar en más de una muestra) del set de datos. Para cada muestra se genera un árbol y en cada separación de los nodos se utiliza sólo una porción *m* de las *p* variables predictoras Las salidas de todos los árboles se combinan en una salida final Y que se obtiene mediante alguna regla (generalmente el promedio, en RF de regresión y, conteo de votos, en RF de clasificación).

En este trabajo se compararon modelos Random Forest con variables tomadas *in situ* y modelos con variables remotas, se consideró la siguiente estrategia de modelización: se construyeron tres tipos de modelos; modelos que incluyeron variables *in situ*, meteorológicas y satelitales (Modelo 1), modelos que incluyeron variables meteorológicas y satelitales (Modelo 2) y modelos que incluyeron únicamente variables satelitales (Modelo 3). En los Modelos 1 y 2 igualmente se conservó la variable de los valores de CF de muestreos pasados ("lags"). Todos los modelos fueron desarrollados en RStudio utilizando la función "randomForest" del paquete con el mismo nombre (R Core Team, 2023). Las variables medidas *in situ* utilizadas fueron la turbidez, la temperatura, la salinidad y los valores de los muestreos previos de CF (Lags de CF) (Bourel et al., 2021; Segura et al., 2021).

Luego se evaluó si existían diferencias entre las métricas de desempeño de estas tres estrategias de modelado. Las métricas evaluadas para la comparación fueron la precisión total del modelo, la sensibilidad y la especificidad. Estas métricas son las

calculadas a partir de la matriz de confusión (Ver capítulo 2). Las métricas que se quieren maximizar serán los aciertos de los modelos (vardaderos positivos y verdaderos negativos). Y las métricas que se quieren minimizar son las clasificaciones erróneas del modelo (falsos positivos y falsos negativos). Sin embargo, en el contexto de la gestión de usos de playas recreativas los distintos errores del modelo tienen diferentes consecuencias. En el caso de los falsos positivos, implican que el modelo predice un exceso al umbral permitido cuando en verdad no lo era. Y en el caso de los falsos negativos el modelo predice que no hay un exceso al umbral permitido cuando sí lo había. En el primer caso, podría implicar inhabilitar una playa sin razón (una falsa alarma) y en el segundo caso podría implicar exponer a los usuarios a CCF (un riesgo para la salud).

Para todos los tipos de modelos primero se manipuló la base de datos para imputar datos faltantes (NAs) de las variables input con la función "RfImput" del paquete "random Forest" de RStudio. Dicha función utiliza la matriz de proximidad de Random Forest para la imputación de los NA. Para los predictores continuos, el valor imputado es el promedio ponderado de las observaciones no faltantes, donde las ponderaciones son las proximidades. Para los predictores categóricos, el valor imputado es la categoría con la mayor proximidad promedio. La variable "Y" a predecir fue la CCF en dos categorías: Excede o No excede el umbral permitido por el decreto 253/79 de 2000 UCF/100ml.

Para las tres estrategias de modelización (Modelo1, 2 y 3) se aplicaron a su vez, tres técnicas de modelización. Las técnicas utilizadas fueron: la utilización de Random Forest estratificado (al que se nombró como Rf_St), la implementación de una técnica de generación artificial de datos de la clase minoritaria (SMOTE) (al que se nombró como Rf_SMOTE) y la modificación de un parámetro (Cutoff) dentro de RF que permite darle mayor prioridad a la clasificación de la clase minoritaria (al que se nombró como Rf_cutoff).

La técnica de SMOTE, por su acrónimo en inglés, refiere a la Synthetic Generation of Cases, genera casos de la clase minoritaria artificialmente, y un muestreo reducido para la case mayoritaria, de modo de balancear las clases y entrenar el modelo con clases balanceadas para mejorar su capacidad predictiva. Se utilizó la función "SMOTE" del paquete "DMwR" en RStudio, que consiste de algunos parámetros como "perc.over" y "perc.under" para aumentar casos de la clase minoritaria o disminuir casos de la clase mayoritaria. Para aumentar la cantidad de casos de la clase minoritaria se seleccionan

casos dicha clase de forma aleatoria, construidos por la interpolación de valores de las variables explicativas de los casos seleccionados y sus k-vecinos (Chawla et al., 2002). Cabe destacar que esta técnica se realiza de forma previa al entrenamiento del modelo.

El Cutoff, por otra parte, es un parámetro dentro de la función Random Forest, que permite modificar la probabilidad de que una predicción sea asignada a cada clase. En el caso de problemas de clasificación de dos clases, esta probabilidad es, por defecto, 0,5 para cada clase. Este parámetro puede modificarse para una categoría se asigne con mayor probabilidad, por ejemplo, para datos con clases desbalanceadas. La clase "ganadora" para una observación es aquella que tiene la relación máxima de proporción de votos sobre el Cutoff. El valor predeterminado es 1/k, donde k es el número de clases (es decir, gana la mayoría de votos). En este trabajo se evaluaron distintos Cutoff hasta seleccionar el Cutoff que maximizaba el desempeño de predicción de la clase minoritaria en los modelos.

Random Forest estratificado consiste en utilizar muestras boostrap estratificadas para entrenar el modelo. Lo que permite definir cuantas veces como mínimo se muestrea una clase en cada muestra boostrap. Asegurando así, que la clase minoritaria esté representada en cada muestra boostrap en la que se entrena el modelo (Bourel et al., 2021).

El modelo con mejor desempeño de las tres técnicas previamente descritas fue seleccionado y testeado con datos “unseen” es decir, previamente no vistos por el modelo El set de datos utilizado para entrenar los modelos fueron las observaciones desde el 2009 al 2020, se validaron los modelos con las observaciones del 2020 al 2021. Se seleccionó el mejor modelo y éste se testeó con datos del 2021 al 2023.

Los modelos de mejor desempeño desarrollados en este trabajo fueron, a su vez, comparados con modelos construidos en trabajos previos por Segura et al., (2021) y con la línea de base. La línea de base es la correspondiente al criterio que utiliza la IM actualmente, tomando en cuenta precipitación en las 24hs previas como criterio para considerar un exceso de CCF.

La importancia de las variables input fue evaluada mediante la visualización que otorga el paquete “randomForest” que permite mediante la función “VarImpPlot” graficar

la importancia en orden decreciente según el decrecimiento del índice de Gini. La disminución media del coeficiente de Gini es una medida de cómo cada variable contribuye a la homogeneidad de los nodos y las hojas en el RF resultante. Cuanto mayor sea el valor de la exactitud de disminución media de Gini, mayor será la importancia de la variable en el modelo.

Para evaluar diferencias significativas entre las observaciones que fueron clasificadas en las dos categorías ("Excede y "No Excede") por los modelos se llevaron a cabo test de Mann-Whitney. Este es un test no paramétrico que permite comparar dos muestras independientes y se toma en cuenta la significancia (p) para rechazar ($p < 0.05$) o no ($p > 0.05$) la hipótesis nula de igualdad dos muestras. Este test realizó mediante la función "wilcox.test".

Todos los análisis se realizaron en el programa estadístico R utilizando la interfaz de RStudio en su versión 4.3.0 (R Core Team, 2023)

Construcción de la interfaz web

Para facilitar la interpretación de las predicciones de los modelos y la implementación de los modelos predictivos construidos en el sistema de gestión de calidad de playas de la IM se realizó una interfaz web con el paquete "Shiny" de R Studio (R Core Team, 2023). Dicho paquete permite la construcción de interfaces web interactivas directamente desde R Studio. La creación de la aplicación requiere de dos componentes fundamentales de un mismo código: un parte del código en donde se diseña la interfaz que ve el usuario (interfaz **Ui**) y contiene los inputs (pestañas, botones, distintas secciones de una aplicación) y outputs (los textos, gráficos, imágenes, tablas o mapas que se muestren en la aplicación), y otra parte del código llamada "**Servidor o Server"**, por sus nombre en inglés, que recibe los inputs del Ui y con ellos genera outputs reactivos (Figura 3.3).

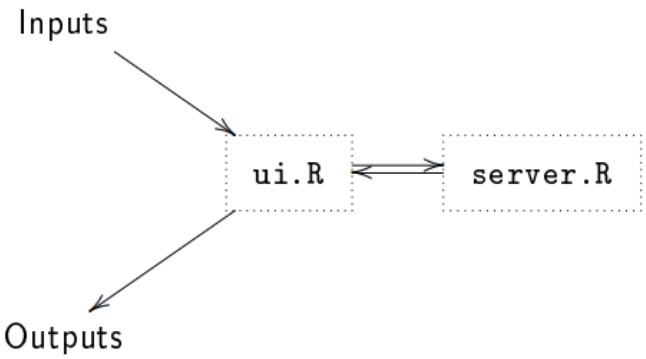

Figura 3.3 Diagrama con las partes fundamentales de una interfaz web construida con el paquete Shiny R. La interfaz se construye con dos partes: un código (Ui) que diseña lo que el usuario puede observar de la aplicación, y otro código (Server) que tiene las funciones reactivas que permiten que la aplicación sea interactiva.

Para la construcción de la interfaz web se utilizó un diseño de página tipo "Dashboard page" que permite la visualización y compartimentación de la información en diferentes pestañas utilizando el paquete "shinydashboard" de R Studio. Dentro de la Ui, una Dashboard page tiene tres partes fundamentales: un "header", es decir un título principal, una "sidebar" que consiste en la barra izquierda a donde se muestran las diferentes pestañas y el "body" que es el cuerpo de la aplicación que es el contenido que se muestra cuando se selecciona cada pestaña.

Para la visualización de los datos históricos de la IM y para la visualización de las predicciones de los modelos se utilizó una combinación de mapas, gráficos y tablas. Para implementar mapas interactivos, y darles reactividad se utilizó el paquete "Leaflet" y la función "renderLeaflet" que permite construir mapas a partir de la fuente de código abierto JavaScript, en donde se le adicionó una capa con las playas de Montevideo. Esta función también permite la visualización de datos en los mapas al asignar un variable ya sea cuantitativa (concentración de CCF) o cualitativa (clase predicha por el modelo). Por otra parte, se utilizó la función "dygraphs" del paquete "dygraph" para crear un gráfico interactivo. Con la función "DT" del paquete "DT" se construyó una tabla interactiva para visualización de los datos históricos de calidad de playas de la IM.

3.3 RESULTADOS

Descriptivos

La CCF varió temporalmente y espacialmente entre playas y entre años a lo largo de la ventana temporal analizada (Figura 3.4 y Table 3.1). En la Tabla 3.1 se describe el número total de días muestreados desde el 2009 al 2023 para cada playa, así como el valor mínimo, máximo, medio y desvío de CF (Log10). La proporción de excesos puntuales a la normativa (CF > 2.000) global fue de 7%. Las playas con mayor porcentaje de excesos a la normativa fueron Puerto del Buceo (24,6%), Santa Catalina (18,2%), Cerro (16,6%), Ramírez (13,3%) y Miramar (12,0%) y las playas con menor porcentaje de excesos fueron Punta Espinillo (0,3%), Punta Yeguas (0,7%), y playa Nacional (0,9%).

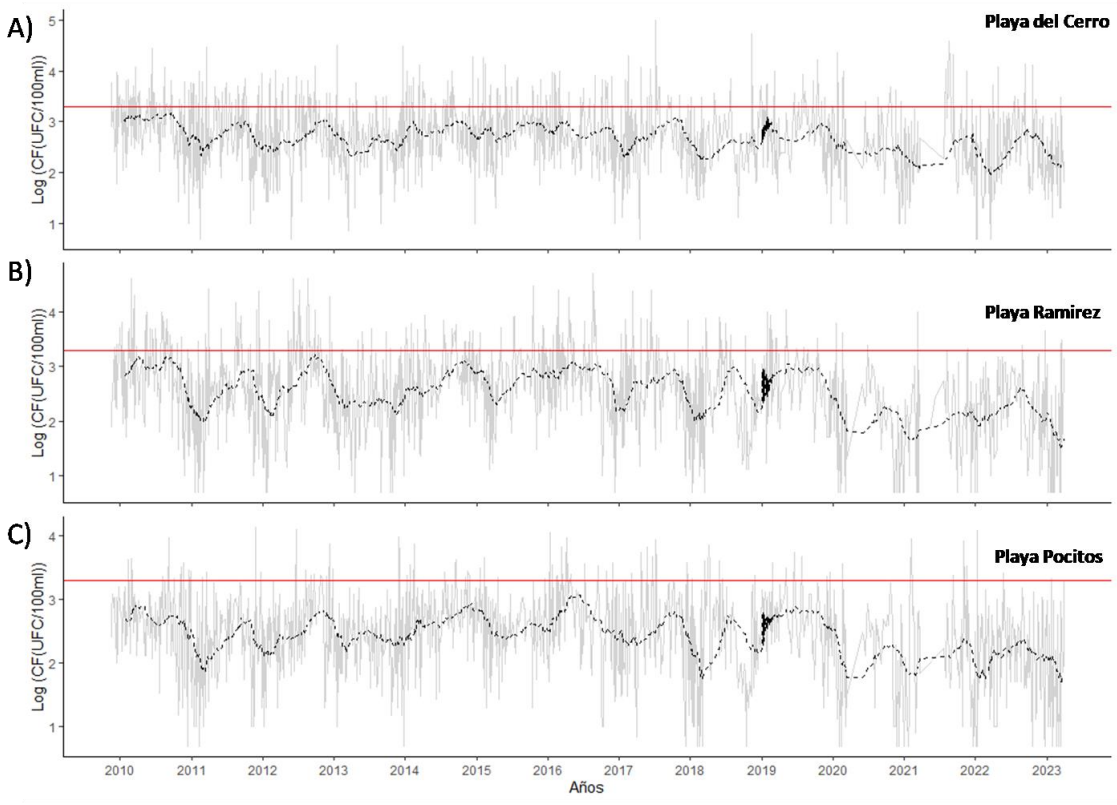

Figura 3.4 Dinámica temporal del logaritmo de coliformes fecales (UFC/100ml) en el periodo de estudio (del 15 de noviembre de 2009 al 3 de marzo de 2023) de playas representativas de las diferentes zonas de la Montevideo. A) Playa Pocitos, B) Playa Ramirez y C) Playa del Cerro. Se seleccionaron tres estas playas debido a que presentaron mayor cantidad de datos que el resto. En una línea negra punteada se muestra la Media Móvil del logaritmo de CF. En una línea roja horizontal se muestra el umbral de CF permitido por la normativa.

Tabla 3.1. Resumen estadístico del monitoreo de la contaminación por coliformes fecales en las playas de Montevideo por parte del Servicio de Evaluación de la Calidad y Control Ambiental de la Intendencia de Montevideo. Nombre de las playas de Montevideo y sus códigos ordenadas de Oeste a Este. Se muestra la zona a la que fue asignada cada playa (Oeste: W y Este: E), número de veces que esa playa fue muestreada (N), el mínimo (Min CF), máximo (Max CF), promedio (Prom CF) y desvío estándar (DS) del logaritmo en base 10 de la concentración de contaminación fecal para cada playa. Además, se muestra el porcentaje (%) de excesos a la normativa permitida (3.3 (Log CFU/100ml)) para cada playa (% Excesos).

Zona	Playa	Código	N	Max CF	Prom CF	DS	% Excesos
W	Punta Espinillo	PE	589	3.77	1.43	0.61	0.3
W	La Colorada	LC	1316	4.00	1.70	0.68	1.2
W	Pajas Blancas	PB	1469	4.67	1.88	0.70	1.9
W	Zabala	Z	748	4.83	1.61	0.73	1.7
W	Punta Yeguas	PY	662	3.69	1.85	0.60	0.7
W	**Santa Catalina**	**SC**	**1474**	**4.99**	**2.68**	**0.74**	**18.2**
W	Nacional	PN	811	4.25	2.02	0.62	0.9
W	**Cerro**	**PA**	**1530**	**4.99**	**2.68**	**0.64**	**16.6**
E	**Ramirez**	**RAM**	**1541**	**4.70**	**2.54**	**0.75**	**13.3**
E	Pocitos en Barreiro	POC	1535	4.12	2.44	0.63	6.5
E	Pocitos en Avenida Brasil	POCB	1090	3.91	2.46	0.58	5.8
E	Buceo	BUC	1456	5.18	2.47	0.67	8.1
E	**Puerto del Buceo**	**PPB**	**727**	**6.17**	**2.73**	**0.86**	**24.6**
E	Malvín	MAL	1514	5.38	2.32	0.73	6.2
E	Brava	BR	873	4.56	2.04	0.70	2.0
E	Honda	H	1417	4.70	2.14	0.69	2.1
E	Ingleses	ING	1486	4.89	2.27	0.65	3.9
E	Verde	VDE	1436	4.20	2.15	0.71	3.7
E	Mulata	MTA	730	4.49	2.02	0.68	1.9
E	Carrasco	CAR	1500	4.55	2.42	0.66	8.2
E	**Miramar**	**MIR**	**1398**	**4.69**	**2.53**	**0.67**	**12.0**

Respecto a las variables *in situ* analizadas para el periodo de estudio (2009-2023), se registraron comportamientos similares a los previamente registrados por Segura et al., (2021). Se registraron variaciones entre playas del Este y Oeste respecto a la salinidad, con mayor dispersión en los datos en las playas del Oeste con un rango de 0.10 a 34 y promedio de 6.24. Para las playas del Este el rango fue de 0.10 a 36 con un promedio de 12.09. Respecto a la turbidez se registró mayor turbidez en las playas del Oeste (1.10 mínima, 953 máxima y promedio de 41.33 NTU) y mayor dispersión de los datos en las playas del Este (1.00 mínima, 825 máxima y promedio de 33.22 NTU). La temperatura del agua (°C) fue similar entre playas con un comportamiento estacional esperable con mínimas en invierno y máximas en verano (6.80 y 30.30°C para el Oeste y 8.40 y 32.10 °C para el Este).

Respecto al comportamiento de las variables registradas a partir de satélites, la temperatura superficial del agua (SST por su acrónimo en inglés) mostró una oscilación estacional esperada para esta variable, con máximos en verano (28.4°C) y mínimos en invierno (9.0 °C) (Figura 3.5 A). La turbidez satelital (Kd490) registró máximos en invierno tanto para la zona Este (6.55) como para en el Oeste (6.55) y mínimos en verano (0.84 y 0.90 para Este y Oeste respectivamente), aunque la zona Este, durante el verano y el otoño presentó mayor número de valores extremos de Kd490 (Figura 3.5 B y Figura 3.6).

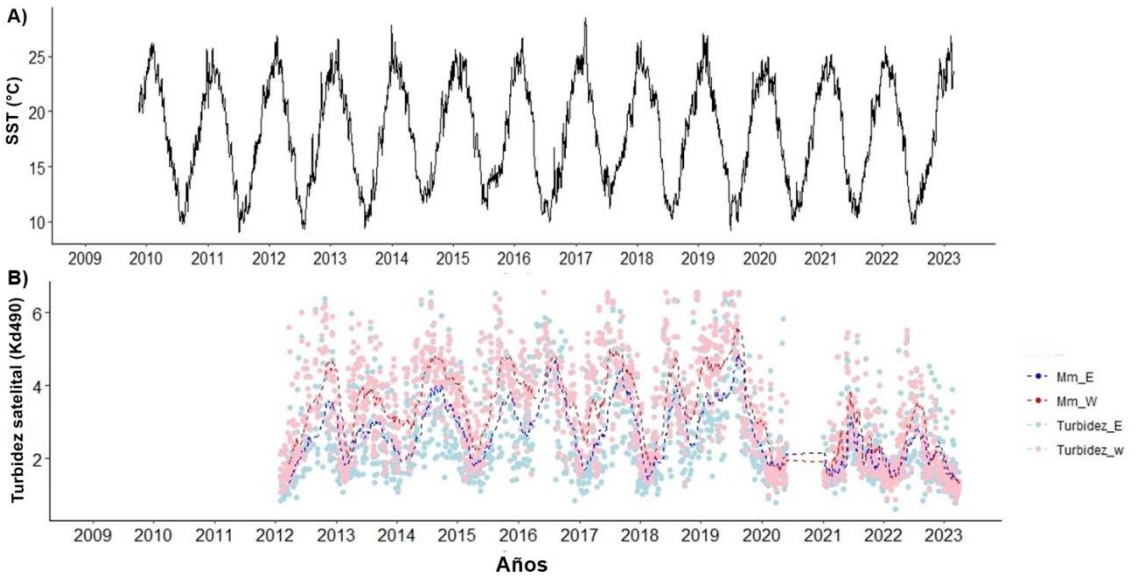

Figura 3.5 Gráficos de series temporales de A) temperatura superficial del agua para la zona Este (SST, °C), y B) turbidez satelital (Kd490) para el periodo de estudio (del 15 de noviembre de 2009 al 3 de marzo de 2023), en puntos rojos se muestra la turbidez para la zona Oeste (Turbidez_w) y en puntos azules la turbidez para la zona Este (Turbidez_E). La línea punteada roja muestra la media móvil de la turbidez del Oeste (Mm_w) y la línea punteada azul muestra la media móvil de la turbidez del Este (Mm_E).

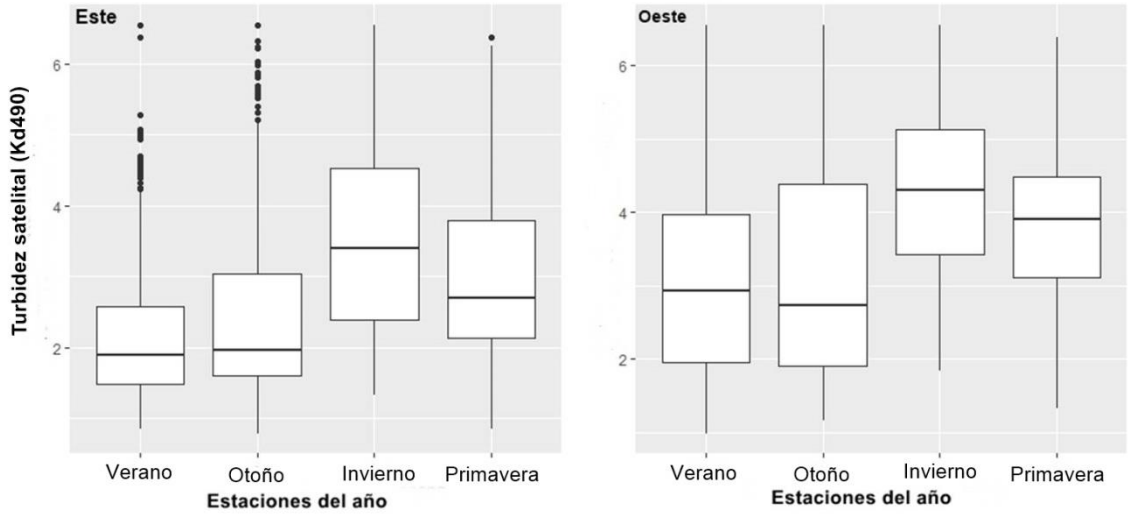

Figura 3.6 Gráficos de la variación estacional de la turbidez satelital (Kd490) para la zona Este (izquierda) y Oeste (derecha) para el periodo de estudio.

Correlaciones entre las zonas Este y Oeste para las variables satelitales

Tanto la temperatura superficial del agua (SST) y la velocidad del viento (km/h) presentaron una correlación significativa y positiva entre la zona Este y Oeste ambos con

un r=0.99 y un p < 0.01 (Figura 3.7 A y C). Mientras que el Kd490 presentó una correlación positiva entre Este y Oeste, pero con menor intensidad que las anteriores, con un r= 0.78 y p < 0.01 (Figura 3.7 B).

El ajuste de un modelo de regresión segmentado entre el Kd490 entre las zonas del Este y el Oeste arrojó un punto de quiebre en 2.9 (ds=0.16) con un R^2 ajustado de 0.62. A su vez, las correlaciones entre Kd490 menor a 3 entre Este y Oeste fue de 0.45 (p < 0.01) y para valores mayores a 3 fue de 0.69 (p < 0.01). Es decir, la correlación entre zonas Este y Oeste respecto al Kd490 es mayor cuando esta toma valores mayores a 3 (Figura 3.7 B).

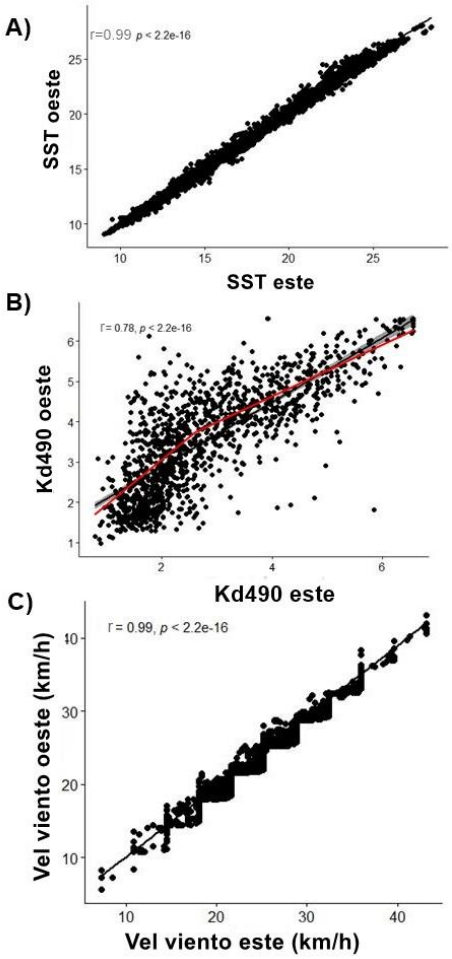

Figura 3.7 Se muestran las correlaciones entre las variables satelitales A) temperatura superficial del agua, B) Kd490 (el modelo de regresión segmentado se muestra con una línea roja) y C) velocidad del viento (WS, por su acrónimo en inglés) entre zona Este y Oeste.

Correlaciones entre variables in situ y satelitales

Las correlaciones entre las variables medidas *in situ* y las variables satelitales mostraron que para la temperatura (°C) *in situ* y la SST presentaron una correlación de $r=0.99$ y $p < 0.01$ para la zona Este y $r=0.97$ y $p < 0.01$ para la zona Oeste (Figura 3.8 A). La turbidez (NTU) *in situ* y el Kd490 presentaron una correlación de $r=0.43$ y $p < 0.01$ para la zona Este y $r=0.59$ y $p < 0.01$ para la zona Oeste (Figura 3.8 B). La velocidad del viento (km/h) *in situ* y la velocidad del viento satelital presentaron una correlación de $r=0.20$ y $p < 0.01$ para la zona Este y $r=0.21$ y $p < 0.01$ para la zona Oeste (Figura 3.8 C). Como la correlación fue significativa pero baja esta variable no fue utilizada *a posteriori* para construir los modelos predictivos.

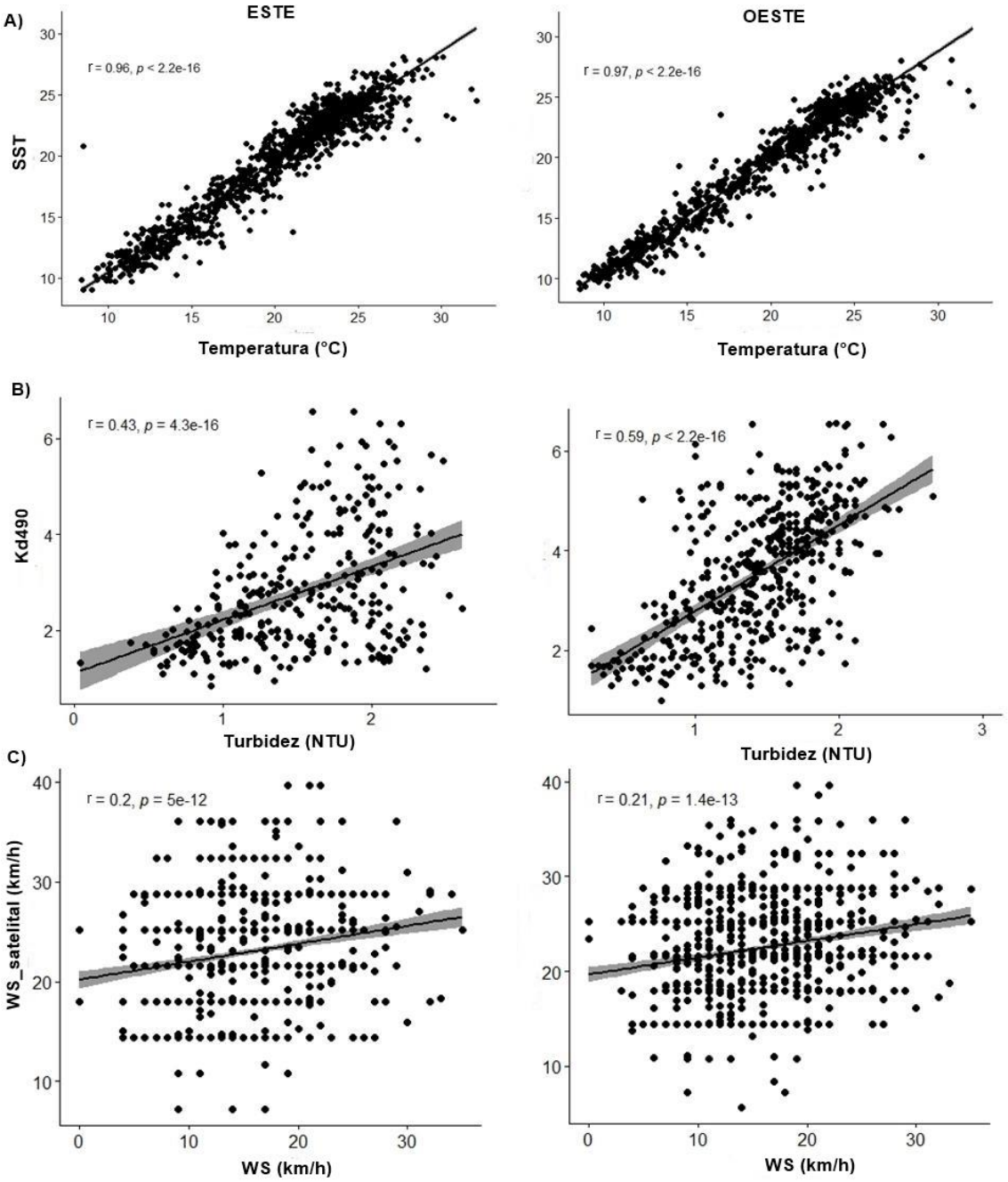

Figura 3.8 Relación entre las variables satelitales y las variables *in situ* tanto para la zona Este a la izquierda y Oeste a la derecha. A) correlación entre la temperatura *in situ* y la temperatura superficial del agua satelital (SST) B) correlación entre la turbidez *in situ* y la turbidez satelital (Kd490) y C) correlación entre la velocidad del viento de estaciones meteorológicas y la velocidad del viento satelital (WS, por su acrónimo en inglés).

Desempeño de modelos predictivos

Se entrenaron un total de 30 modelos para las 20 playas monitoreadas por el departamento de Calidad Ambiental de la Intendencia de Montevideo. Los resultados de

los 3 modelos con mejores métricas dentro de las tres técnicas evaluadas fueron los RF con Cutoff (0.09-0.6) de 500 árboles y el número máximo de 3 variables utilizadas en cada nodo. El error OOB del modelo fue de 15.6 % durante el entrenamiento. Con esta técnica se evaluaron los tres tipos de modelos (Modelo 1, 2 y 3). El Modelo 1 obtuvo una Precisión de 63%, una sensibilidad del 74% y una especificidad del 63%. El Modelo 2 obtuvo una Precisión de 81%, una sensibilidad del 54% y una especificidad del 81%. El Modelo 3 obtuvo una Precisión de 90%, una sensibilidad del 43% y una especificidad del 91% (Tabla 3.2).

Tabla 3.2 Comparación de métricas (Tasa de Aciertos, Sensibilidad, y Especificidad) entre modelos construidos en este trabajo (Random Forest 1, 2 y 3), la línea de base que se basa en predecir un exceso en caso de registrarse precipitaciones las 24hs previas (Pp24hs) y los construidos por Segura et al., 2021 y Bourel et al., 2021 (Random Forest estratificados: Rf_st, Maquinas de soporte de vectores: SVM y Adaboost). La columna Datos refiere a la base de datos utilizada para el entrenamiento de los modelos, es Originales cuando no se manipula la base de datos, Imputados, cuando se imputaron datos faltantes con la función RfImpute y SMOTE cuando se utilizó esta técnica para el balance de clases.

Trabajo	**Modelo**	**Datos**	**Precisión (Tasa de Aciertos, %)**	**Sensibilidad (TPR, %)**	**Especificidad (TNR, %)**
Este trabajo					
Modelo 1 (in situ+ meteo+ satelitales)	Rf_cutoff	Imputados	63	74	63
Modelo 2 (lags + meteo + satelitales)	Rf_cutoff	Imputados	81	54	81
Modelo 3 (lags + satelitales)	Rf_cutoff	Imputados	90	43	91
Línea de base	Pp24hs	Originales	82	40	85
Segura et al. 2021	Rf st	Originales	84	63	85
Bourel et al.2021	SVM	Upsampling	84	58	86
	SVM	SMOTE	84	48	86
	Adaboost	SMOTE	83	64	84

Las variables input o de entrada más importantes para el Modelo 1 fueron la Playa (variable categórica que representa el sitio de muestreo), y los valores de CF de un muestreo previo ("Lag_cf1"), la turbidez y la temperatura *in situ*. Las variables más importantes para el Modelo 2 fueron la también la Playa y el "Lag" de la CF de muestreo previo (Lag_cf1). A esto le siguió el "Lag" de la CF de dos muestreos previos (Lag_cf2) y la SST. Por último, las variables más importantes del Modelo 3 fueron nuevamente el "Lag" de la CF de muestreo previo (Lag_cf1), seguido de la turbidez satelital (Kd490) y la temperatura del agua satelital (SST) (Figura 3.9).

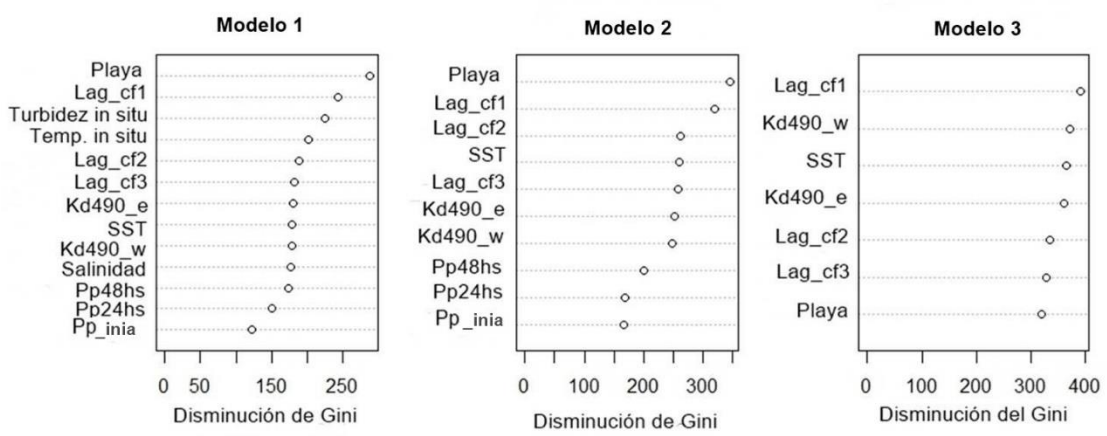

Figura 3.9 Importancia de variables del RF según la disminución del índice de Gini para el Modelo 1, Modelo 2 y Modelo 3. Las variables fueron la Playa que corresponde al sitio de muestreo, Valor de CF de un muestreo anterior (Lag_cf1), Turbidez *in situ*, Temperatura *in situ* (Temp. *in situ*), Valor de CF de dos muestreos anteriores (Lag_cf2), Valor de CF de tres muestreos anteriores (Lag_cf2), Kd490 de zona Este (Kd490_e), Kd490 de zona Oeste (Kd490_w), Temperatura Superficial del Agua (SST), Salinidad, Precipitaciones acumuladas en 24hs (Pp24hs), Precipitaciones acumuladas en 48hs (Pp48hs), Precipitaciones de INIA (Pp_inia).

La distribución de la concentración de CF mostró diferenciación cuando los modelos (Modelo 1, 2 y 3) clasificaron como excesos y no excesos (Figura 3.10). El test Mann-Whitney arrojó diferencias significativas entre los valores de CF (log (CF)) asignados a las categorías "Excede" y "No Excede" ($p < 0,01$) en los tres casos.

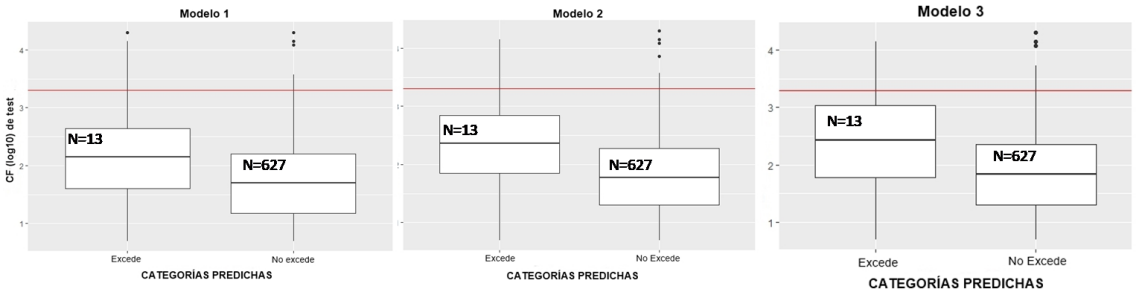

Figura 3.10 Boxplot de observados versus predichos para los tres mejores modelos testeados en este trabajo. En el eje de las "x" están las categorías a predecir "Excede" o "No Excede" y en el eje de las "y" los valores de contaminación fecal (UCF/100ml) del set de datos con los cuales se testeó el modelo en escala logarítmica. En una línea roja horizontal se muestra el umbral de CF permitido por la normativa.

Implementación de la Interfaz Web

La interfaz construida fue una Dashboard Page (un tipo de diseño predeterminado de aplicación que provee el paquete Shiny de R) que permite la generación de un menú desplegable en donde pueden abrirse diferentes pestañas. Dicha interfaz web cuenta con un menú principal de pestañas en donde, primero, se introduce al funcionamiento de la página con una pestaña de "Bienvenida" y otra de "Instrucciones de uso" (Figura 3.11).

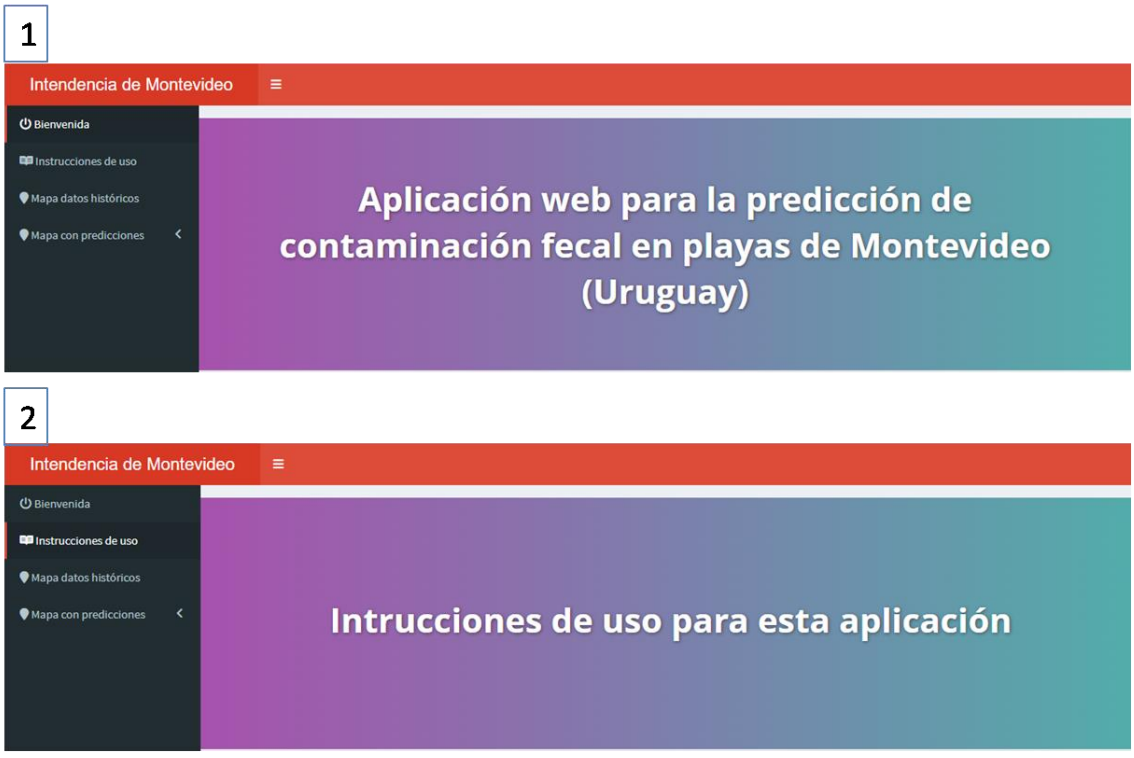

Figura 3.11 Aspecto general de la interfaz web creada con Shiny R. En la imagen se muestran las primeras dos pestañas de 1) presentación de la aplicación, 2) Instrucciones de uso para la aplicación.

En otra pestaña puede visualizarse la base de datos histórica de la IM en gráficas, tablas y mapas (Figura 3.12, 3.13 y 3.14). Esto permite comparar fácilmente valores de CF entre playas y acceder de forma rápida a valores históricos, tanto de CF como de cualquier variable monitoreada en las playas. El conjunto de visualizaciones diseñado permitirá identificar fácilmente patrones y eventos extremos de CCF, o, por ejemplo, de temperatura, salinidad, o turbidez en la base de datos. En el mapa se puede visualizar los valores de CF (UFC/100ml) para una fecha determinada para cada playa. Estos valores son representados en una escala del blanco al rojo oscuro para representar en orden creciente desde el 0 al máximo registrado de CF para esa fecha (Figura 3.12). Por otra parte, inmediatamente abajo del mapa puede visualizarse un gráfico interactivo en donde se muestran, para un rango de fechas seleccionadas, valores de los valores de las variables ambientales del monitoreo (salinidad, temperatura, turbidez) (Figura 3.13). Luego, se incluyó una tabla en donde se puede filtrar por fecha y por playa para ver, por ejemplo, en donde están los valores máximos registrados de CF o de variables ambientales y en qué momento (Figura 3.14). Los datos que se pueden visualizar en esta pestaña de la interfaz son los correspondientes a los monitoreos históricos de calidad de playas de la IM.

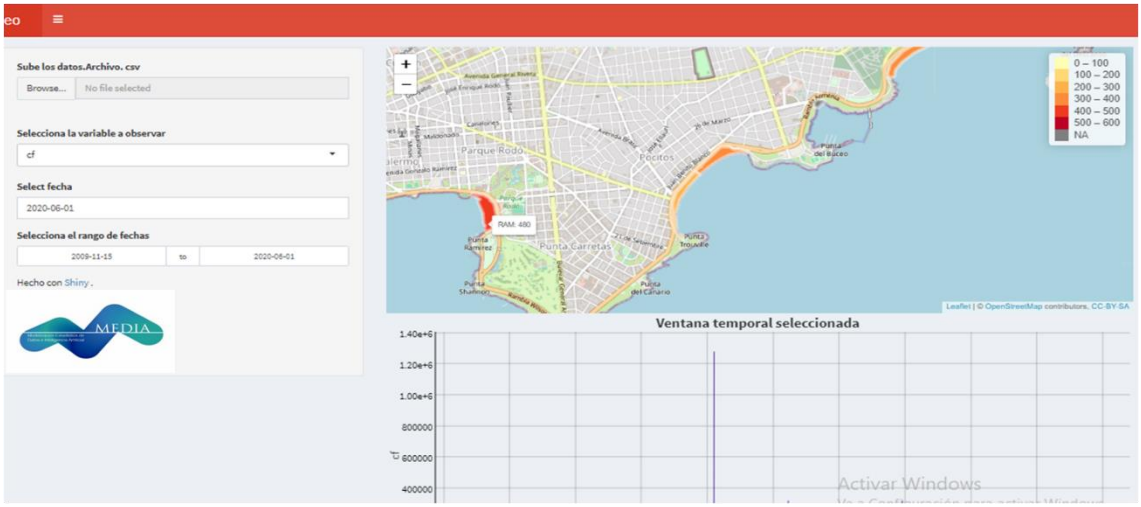

Figura 3.12 Aspecto general de la pestaña en donde pueden visualizarse los niveles de coliformes fecales (UFC/1000ml, la leyenda en colores muestra en un gradiente de amarillo claro a rojo oscuro los valores más altos) en un mapa para la fecha seleccionada en cada playa monitoreada

por la Intendencia de Montevideo en interfaz web creada con Shiny R. Los datos de coliformes fecales observados en los mapas provienen de la base de datos histórica de los monitoreos de calidad de playas de la IM.

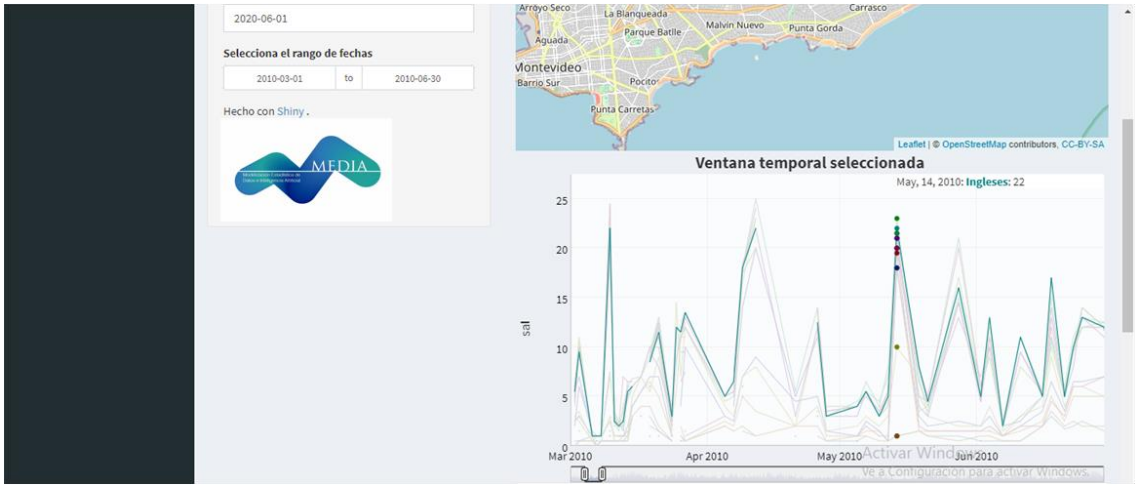

Figura 3.13 Aspecto general de la pestaña en donde se muestra un gráfico puede visualizarse la evolución temporal de la variable seleccionada en la ventana temporal seleccionada para cada playa monitoreada por la Intendencia de Montevideo.

Figura 3.14 Aspecto general de la pestaña en donde se muestra una tabla en la que puede filtrarse cada columna de modo de acceder rápidamente a un día seleccionado, una playa, o un valor de coliformes en particular para cada playa monitoreada por la Intendencia de Montevideo.

En las siguientes pestañas puede visualizarse las predicciones para el mismo día del mejor modelo de clasificación seleccionado y estas predicciones son visualizadas en mapas de fácil interpretación (Figura 3.15). En estos mapas se muestra por un lado la categoría predicha por el modelo “Excede” en rojo y “No Excede” en verde para cada

playa y en otro mapa las probabilidades de excedencia del umbral permitido por la legislación uruguaya para cada playa (2000 UFC/100ml). Esta probabilidad se presenta en un código de colores del blanco al rojo (de 0 a 1) (Figura 3.15).

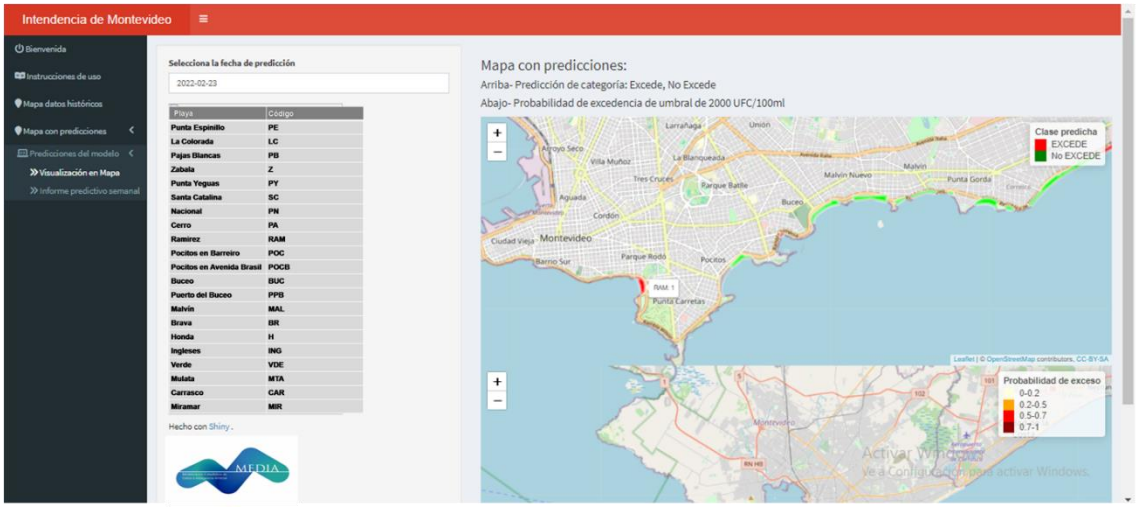

Figura 3.15 Aspecto general de la pestaña en donde se muestran predicciones de los modelos en mapas en la interfaz web creada con Shiny R

Esta aplicación fue presentada a los integrantes de la Unidad de Calidad de Agua del SECCA de la IM. La valoración de la aplicación fue positiva. En donde los gestores indicaron que era de utilidad poder acceder a los datos históricos de forma rápida e intuitiva y que la visualización de las predicciones de los modelos en mapas simplifica su interpretación.

3.4 DISCUSIÓN

El presente capítulo aporta al proceso de implementación de modelos predictivos en el sistema de gestión de calidad de playas de la Intendencia de Montevideo. Se construyeron modelos predictivos de Aprendizaje Automático para las playas de Montevideo incorporando variables satelitales y su desempeño fue similar a los mejores modelos logrados para la zona y con capacidad predictiva con niveles de los que presentaron mejor performance a nivel mundial. Se implementó el modelo con mejor desempeño en una interfaz web que permite visualizar las predicciones de los modelos de forma interactiva. Esta implementación significa un avance significativo en la operativización de los modelos en el sistema de gestión de calidad de playas de

Montevideo y es un camino a explorar para facilitar la gestión de sistemas acuáticos del país.

Desarrollo y desempeño de los modelos

Las fluctuaciones temporales y espaciales de las variables satelitales analizadas en este trabajo en la zona costera de Montevideo están alineadas con las observaciones sobre la gran dinámica del estuario (Fossati y Piedra-Cueva, 2013; Santoro et al., 2017). La turbidez fue diferente para las zonas Este y Oeste del departamento, reafirmando lo que ya había sido reportado por trabajos previos utilizando mediciones *in situ* (Naggy et al., 1997; Muniz et al., 2019; Segura et al., 2021). El indicador de turbidez satelital Kd490 es una variable que se registra como importante para la predicción de CCF para las playas de Montevideo. La temperatura superficial del agua (SST), presentó una fuerte correlación con los valores medidos *in situ* y aportó a los modelos de predicción. Sin embargo, la velocidad del viento presentó baja correlación con la velocidad del viento tomada por estaciones meteorológicas, así como un alto número de datos faltantes, por lo que su estimación satelital no aportó a la construcción de los modelos predictivos. Desde el punto de vista de la modelización, los datos satelitales presentan algunas ventajas: los conjuntos de datos obtenidos por satélites suelen estar disponibles para dominios espaciales más extensos que las recogidas *in situ*, por lo general, con una mayor frecuencia temporal (diaria) (Kim et al., 2017). Si bien, estudios recientes muestran que hay un aumento en la implementación de datos extraídos de productos satelitales para predecir distintos atributos de la calidad de agua (Hassan y Woo, 2021), la utilización de información satelital para la construcción de los modelos predictivos específicamente en el contexto de la modelización del CCF es aún escaso (Laureano-Ramos et al., 2019). En este trabajo se evaluó y demostró que las variables satelitales Kd490 y SST son de relevancia para la predicción de CCF. Nuevos productos satelitales, como provenientes de los satélites Sentinel 2, por ejemplo, podrían ser implementadas también ya que cuentan con mayor precisión y resolución (Jutz y Milagro-Perez, 2018; Martinis et al., 2021). Estas tecnologías ya han sido evaluadas en otros ambientes acuáticos del Uruguay, por ejemplo, para la estimación de la turbidez en Río de la Plata y en el embalse de Palmar, Río Negro (Maciel et al., 2018; Ministerio de Ambiente, 2022; Zabaleta et al., 2022) y podrían utilizarse en conjunto con los modelos de AA para la predicción de CCF en las playas de Montevideo.

Por otra parte, una de las desventajas que suele presentar la utilización de variables de origen satelital es la cantidad de datos faltantes que presentan (entre el 53 y 95% en este estudio, dependiendo del producto) (Gerber et al., 2018; Liu y Wang, 2018). En este trabajo se utilizó la técnica de imputación de datos para completar los datos faltantes con la función "RfImpute" dentro del paquete "RandomForest" de RStudio (Ishioka, 2012) que permitió entrenar los modelos con más observaciones. Se han implementado diversos algoritmos de AA para superar el problema de los datos faltantes en el caso de la estimación de la clorofila-a, por ejemplo, para reconstruir valores en los días nublados (Park et al., 2020; Zhang y Zhou, 2023; Catipovic et al., 2023). La imputación de datos mediante RfImpute es ampliamente usado para lidiar con datos faltantes en el contexto de los modelos AA (Tabo et al., 2022; Campbell et al., 2020; Kaur Dhaliwal et al., 2022), pero poco implementada para datos satelitales en predicción de calidad de agua.

En este trabajo, las variables tomadas *in situ* tuvieron una mayor relevancia en todos los casos al comparar modelos utilizando variables *in situ* y variables remotas (tanto meteorológicas como satelitales). La variable categórica Playa fue la más importante junto con los valores de los muestreos previos de CCF ("Lags") similar a lo recogido por investigaciones previas (Bourel et al., 2021; Segura et al., 2021; Jones et al. 2013; Gazzaz et al., 2012). Esto apunta a la importancia de mantener los monitoreos y el registro de variables *in situ* para la predicción de CCF en complemento con las variables remotas (tanto meteorológicas como satelitales). Las dinámicas de cada playa, su orientación, circulaciones internas, o tal vez caños aliviaderos que desembocan en las playas, perecen definir la concentración de CCF de un modo relevante. Algunas playas, debido a estas características, presentan mayores proporciones de excesos de CCF a la normativa y eso puede verse reflejado en que la variable playa sea la que más importante para predecir CCF en Montevideo.

Las precipitaciones acumuladas aparecen como variables relevantes para la predicción de CCF en Montevideo. Esta es una variable clave en la modelización de CCF en otros sistemas, particularmente las precipitaciones acumuladas de 24 y 48hs previas (Sokolova et al., 2021; Tselemponis et al., 2023) y es de especial relevancia en ambientes estuarinos (Hose et al., 2005; González et al., 2012). Las precipitaciones dispersan y hacen llegar a las zonas costeras la CCF de fuentes difusas, ya sea hogares, industrias, etc. Las precipitaciones podrían tener un rol importante en explicar la presencia de CCF

en Montevideo por dos razones. La primera es que el sistema unitario de Montevideo, indefectiblemente afecta a los niveles de CCF cuando se registran precipitaciones. La segunda es que, si bien el Plan de Saneamiento de Montevideo cubre una buena parte de la ciudad, aún no cubre partes del departamento que podrían estar aportando CCF a arroyos y cañadas que desembocan a la zona costera capitalina (Plan Nacional de Saneamiento, 2020).

El desempeño de modelos estadísticos desarrollados para las playas de Montevideo, tanto en trabajos previos como en este estudio, presentan buenas métricas en comparación con el desempeño de otros modelos desarrollados a nivel global (Heasley et al., 2021; Francy et al., 2020). Al comparar el desempeño de los modelos construidos en este trabajo con la línea de base utilizada por la IM hasta el momento se observa que los modelos de este trabajo tienen sensibilidades por encima de las de la línea de base. A su vez, el desempeño global de los modelos es similar a los construidos en trabajos previos (Bourel et al., 2021). La sensibilidad de un modelo de clasificación representa que porcentaje de acierto tiene el modelo para predecir eventos de excesos a los umbrales de CCF previstos por las normativas de cada país o región. En general, se observa un compromiso entre métricas, al aumentar la sensibilidad, disminuye la precisión y la especificad, y viceversa. La sensibilidad tiende a ser la métrica de calificación más baja en todos los modelos, tanto en los modelos desarrollados en este trabajo como en los trabajos anteriores. Las bases de datos de CCF, presentan un desbalance asociado al bajo porcentaje de días en los que se registran excesos que representa un desafío para la modelización que ya ha sido documentado en trabajos previos en la zona (Bourel et al., 2021) y nivel mundial (Zhang et al., 2023; Jang et al., 2023, Nafsin y Li, 2023; Watthaisong et al., 2024). En el caso de Uruguay, la legislación vigente prevé un umbral permitido de 2000 UCF/100ml, este umbral, es alto al compararlo con la legislación permitida de otras regiones del mundo (Stidson et al., 2012; Fancy et al., 2020; Mark y Erichsen, 2007; Dada et al., 2019; Thoe et al., 2018). En este trabajo el porcentaje de días en los que se registraron excesos a la normativa fue 7% de los días muestreados totales, lo que genera una dificultad inherente de alcanzar sensibilidades mayores. El desarrollo de otras estrategias de modelización, como la combinación de modelos de AA (Metalearning, por su nombre en inglés) que capturen la complejidad el set de datos podrían ser implementadas para avanzar sobre este desafío (Marmion et al., 2009, Bourel et al., 2017).

Implementación de los modelos en el sistema de gestión de playas recreativas de Montevideo

La implementación de modelos en SAT para la gestión de CCF es aún escasa. En Latinoamérica se registran avances en el desarrollo de modelos predictivos de CCF, pero aún las predicciones de los modelos no se muestran en interfaces web interactivas (de Souza et al., 2018; Segura et al., 2021). La implementación de modelos predictivos en el sistema de gestión de calidad de playas de la IM podría ser pionero entonces en la región. Los modelos pueden ser implementados en diferentes etapas en la toma de decisiones, ya sea para complementar los resultados de los monitoreos y contribuir a la toma de decisión de los gestores respecto al cierre o habilitación de una playa, o para mostrar los resultados de las predicciones a los usuarios de las playas de forma directa (Francy et al., 2020). Si, bien en este trabajo la interfaz web desarrollada será en principio utilizada por los gestores (en este caso los funcionarios del Laboratorio de Calidad Ambiental de la IM), podría ser adaptada para informar a los y las usuarios/as de las playas los niveles de alerta para cada playa según las predicciones de los modelos. Este paso, implicaría generar una campaña de información en la página web con la que ya cuenta la IM, informando sobre cómo funcionan los modelos y la información que proveen. Este tipo de implementación contribuye a la democratización sobre los niveles de contaminación en las playas y a evitar el contacto con eventos de CCF. La mayor eficiencia en la reducción de los riesgos sanitarios se consigue cuando las predicciones de los modelos se utilizan en complemento con otras medidas de gestión, ya sean los monitoreos, o (en el caso de la IM), la utilización de banderas sanitarias (Francy et al., 2020; OMS, 2021). Sin embargo, esta difusión de predicciones a los usuarios también presenta desafíos. Por ejemplo, el nivel de aciertos que presenta el modelo o de errores pude generar un gran nivel de confianza o de descreimiento de los usuarios hacia los mismos. Es por esta razón que, generalmente, los modelos pasan una etapa de prueba en donde son utilizados por los gestores de las playas recreativas para validar su funcionamiento. Luego, si las predicciones son mostradas a los usuarios, cuanto mayor sea la información que acompañe a los modelos, acerca de qué información se está mostrando, qué implican los coloreas asociados a las clases (ej. excede o no excede el umbral permitido por la legislación) y que medidas de prevención deben tomarse por parte de los usuarios, más exitosa será la implementación (Francy et al., 2020).

La implementación de la información satelital en la construcción de los modelos, podría utilizarse para avanzar hacia un sistema de predicción en tiempo casi real o "Nowcasting" que es a lo que se tiende en algunos lugares del mundo (Francy et al., 2020; Pras y Mamane, 2023). Sin embargo, dentro de los desafíos de la utilización de este tipo de información actualmente es la necesidad de la formación de personal con entrenamiento para obtener y analizar este tipo de información que automatice los procesos, que además tengan la capacidad de transferir la información satelital desde la obtención hasta la implementación en el desarrollo de los modelos de AA (Schollaert Uz et al., 2019). Esto apunta al hecho de que para que un sistema SAT sea implementado de forma exitosa precisa de la generación de capacidades en distintas áreas del conocimiento (quienes manejan la información satelital, quienes manejan herramientas de modelización y quienes toman las decisiones sobre la gestión de las playas) y de una comunicación interinstitucional fluida, para la disponibilización de los datos, ya sea de los monitoreos o de las variables meteorológicas y satelitales para la construcción y actualización de los modelos. Luego, para la comunicación entre quienes modelan y quienes deben interpretar las predicciones de los modelos para la gestión de los usos de las playas recreativas (Francy et al., 2020).

Las campañas de difusión y comunicación son de central importancia para la implementación exitosa de las SAT (Thoe et al., 2014; Dada et al., 2019). La información debe ser clara y concisa; por ejemplo, para la difusión de las predicciones de los modelos, se utiliza un código de colores que puede ir del verde al rojo, en donde se indica el nivel de alerta según los niveles de CCF predichos por los modelos. Además, se explicita con el mayor detalle posible que significa cada categoría y las medidas para evitar riesgos a la salud. Sería deseable explicar cómo funcionan los modelos utilizados, y los niveles de error de predicción tal como se realizan en interfaces web a nivel mundial (Francy et al., 2013). En el caso de la gestión de las playas de Montevideo, podría implementarse cartelería informativa, combinada con códigos QR para acceder fácilmente a las predicciones de los modelos. También podría utilizarse la bandera sanitaria en casos de CCF, con comunicación directa a los guardavidas de cada playa. En la página web de la IM en la que actualmente se presentan las predicciones podrían implementarse los mapas en los avisos a los usuarios, en conjunto con información que incluyan códigos de colores simples y claros acerca de los niveles de alerta, de modo de aumentar la capacidad de

interpretación que brindan los modelos para hacer más democrática la llegada de la información sobre la CCF.

En resumen, en este capítulo se generaron nuevos modelos que incorporaron variables de productos satelitales a modelos de AA para las playas recreativas de Montevideo con buena capacidad predictivas. Las variables satelitales fueron de importancia para la predicción, en conjunto con variables *in situ.* Estos modelos fueron implementados en una interfaz web interactiva que puede ser utilizada por los gestores de las playas y que fue bien valorada por los mismos para su utilización.

4. CONCLUSIONES

- El presente trabajo identificó vacíos de información y formas de abordar problemas respecto a la modelización de la CCF en playas recreativas. Se detectó una baja implementación de técnicas para lidiar con el desbalance del set de datos en los modelos desarrollados para predecir CCF en playas recreativas. La mayor parte de los modelos estadísticos fueron desarrollados para ambientes de agua dulce y costero-marinos, mientras que los sistemas estuarinos fueron los menos representados. Por otra parte, se detectó que existe una baja proporción de modelos estadísticos que sean implementados en SAT. La mayor parte de los modelos estadísticos implementados ocurren en sitios con monitoreos a largo plazo.

- Respecto a la modelización predictiva para las playas recreativas de Montevideo, los modelos construidos con variables satelitales (Temperatura Superficial del Agua: SST, y el indicador de turbidez del agua Kd490) tienen un buen desempeño respecto a otros modelos alrededor del mundo y respecto a los modelos previamente generados. Por lo que se presenta una buena oportunidad para implementarlos en el sistema de gestión de calidad de playas de la Intendencia de Montevideo.

- Se generó un prototipo de interfaz web desarrollada con el paquete Shiny de RStudio. Esta interfaz permite a los gestores de la IM la rápida visualización de los eventos extremos tanto de las variables físico-químicas del agua, como de eventos de CCF, con una rápida visualización a través de gráficos, tablas, mapas y acceder fácilmente a las predicciones de CCF para todas las playas monitoreadas.

5. PERSPECTIVAS

Esta tesis deja planteadas algunas rutas a seguir respecto a la modelización predictiva de la contaminación fecal en playas recreativas. Entre ellas la necesidad de implementación de técnicas y algoritmos para tratar los desafíos que implica la predicción de contaminación fecal en datos desbalanceados. Implementar técnicas para lidiar con datos desbalanceados y realizar análisis que permitan observar la estructura de los datos es clave. También es importante realizar una evaluación honesta (dividir el set de datos para entrenar modelos y testearlos con datos independientes no vistos previamente por el

modelo) de los modelos para poder comparar los desempeños entre técnicas u algoritmos. En cuanto a la operativización de los modelos, conforme avanza la tecnología y la disponibilización en tiempo real de datos remotos como las precipitaciones o datos satelitales, se recomienda utilizar datos con mayor cobertura temporal para poder generar predicciones de corto plazo (para el día siguiente o el mismo día en la tarde). En futuros estudios podría implementarse la utilización de información de dirección y velocidad del viento de anemómetros localizados *in situ*, representativos de las distintas regiones/playas lo que permitiría capturar dinámicas locales en las playas.

En cuanto a la implementación en el sistema de gestión de Montevideo, se recomienda la implantación escalonada de los modelos seleccionados. La interfaz web generada en esta tesis podría ser utilizada por los gestores (Unidad de Calidad de Agua de la IM). Este es un paso necesario ya que permite que éstos accedan a los datos históricos y a las predicciones de los modelos de forma fácil, accesible e interactiva. Idealmente se recomienda, para siguientes fases de la implementación, que se genere una interfaz web que pueda ser utilizada por los usuarios de las playas. Esta implementación implicaría el desarrollo de una campaña de difusión y comunicación que acompañen las predicciones de los modelos, en conjunto con la mayor información posible acerca de las playas, cómo funcionan los modelos y que tipo de información brindan. Estos componentes de comunicación son fundamentales para que las herramientas de implementación de los modelos sean utilizadas exitosamente por los usuarios de las playas.

BIBLIOGRAFÍA

Aguilera-Venegas, G., López-Molina, A., Rojo-Martínez, G., Galán-García, J.L., 2023. Comparing and tuning machine learning algorithms to predict type 2 diabetes mellitus. Journal of Computational and Applied Mathematics 427, 115115. https://doi.org/10.1016/j.cam.2023.115115

Álvarez, L. 2019. Programa de Saneamiento Urbano de la Ciudad de Montevideo PSU IV. UR-L1005, 1819/OC-UR. Evaluación final. Montevideo: Intendencia de Montevideo.

APHA, 1989. Standard methods for the examination of water and wastewater. American Public Health Association.

Artelia, Halcrow, Rhama y CSI Ingenieros. 2019. Plan Director de Saneamiento y Drenaje Urbano de Montevideo. Resumen ejecutivo. Montevideo: Intendencia de Montevideo.

Avila, R., Horn, B., Moriarty, E., Hodson, R., Moltchanova, E., 2018. Evaluating statistical model performance in water quality prediction. Journal of Environmental Management 206, 910–919. https://doi.org/10.1016/j.jenvman.2017.11.049

Bae, H.-K., Olson, B.H., Hsu, K.-L., Sorooshian, S., 2010. Classification and regression tree (CART) analysis for indicator bacterial concentration prediction for a Californian coastal area. Water Science and Technology 61, 545–553. https://doi.org/10.2166/wst.2010.842

Bedri, Z., Corkery, A., O'Sullivan, J.J., Deering, L.A., Demeter, K., Meijer, W.G., O'Hare, G., Masterson, B., 2016. Evaluating a microbial water quality prediction model for beach management under the revised EU Bathing Water Directive. Journal of Environmental Management 167, 49–58. https://doi.org/10.1016/j.jenvman.2015.10.046

Bolker, B.M., 2008. Ecological models and data in R. Princeton university press, Princeton.

Borba, C., Maldonado, S., Otero, S., 2021. Desarrollo de receptores bioinspirados con aplicaciones de aprendizaje automático para electrorrecepci on en entornos acuáticos. Universidad de la República, Montevideo, Uruguay.

Botto Nuñez, G., Lemus, G., Muñoz Wolf, M., Rodales, A.L., González, E.M., Crisci, C., 2018. The first artificial intelligence algorithm for identification of bat species in Uruguay. Ecological Informatics 46, 97–102. https://doi.org/10.1016/j.ecoinf.2018.05.005

Bourel, M., Crisci, C., Martínez, A., 2017. Consensus methods based on machine learning techniques for marine phytoplankton presence–absence prediction. Ecological Informatics 42, 46–54. https://doi.org/10.1016/j.ecoinf.2017.09.004

Bourel, M. y Segura, A.M., 2018. Multiclass classification methods in ecology. Ecological Indicators 85, 1012–1021. https://doi.org/10.1016/j.ecolind.2017.11.031

Bourel, M., Segura, A.M., Crisci, C., López, G., Sampognaro, L., Vidal, V., Kruk, C., Piccini, C., Perera, G., 2021. Machine learning methods for imbalanced data set for prediction of

faecal contamination in beach waters. Water Research 202, 117450. https://doi.org/10.1016/j.watres.2021.117450

Bradford, S.A., Morales, V.L., Zhang, W., Harvey, R.W., Packman, A.I., Mohanram, A., Welty, C., 2013. Transport and Fate of Microbial Pathogens in Agricultural Settings. Critical Reviews in Environmental Science and Technology 43, 775–893. https://doi.org/10.1080/10643389.2012.710449

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification And Regression Trees, 1st ed. Routledge. https://doi.org/10.1201/9781315139470

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brooks, W., Corsi, S., Fienen, M., Carvin, R., 2016. Predicting recreational water quality advisories: A comparison of statistical methods. Environmental Modelling & Software 76, 81–94. https://doi.org/10.1016/j.envsoft.2015.10.012

Brooks, W.R., Fienen, M.N., Corsi, S.R., 2013. Partial least squares for efficient models of fecal indicator bacteria on Great Lakes beaches. Journal of Environmental Management 114, 470–475. https://doi.org/10.1016/j.jenvman.2012.09.033

Brugnoli, E., Verocai, J., Muniz, P., y García-Rodríguez, F. 2018. Weather, hydrological and oceanographic conditions of the northern coast of the Río de la Plata Estuary during ENSO 2009–2010. Estuary. InTech, London, United Kingdom, 19-38. 10.5772/intechopen.71808

Cal, A., 2022. Metodología automática para mapeo y seguimiento de la condición de cultivos agrícolas durante la zafra a partir de imágenes satelitales y aprendizaje automático en Uruguay. Universidad de Buenos Aires, Buenos Aires.

Campbell, M.J., Dennison, P.E., Tune, J.W., Kannenberg, S.A., Kerr, K.L., Codding, B.F., Anderegg, W.R.L., 2020. A multi-sensor, multi-scale approach to mapping tree mortality in woodland ecosystems. Remote Sensing of Environment 245, 111853. https://doi.org/10.1016/j.rse.2020.111853

Ćatipović, L., Matić, F., Kalinić, H., 2023. Reconstruction Methods in Oceanographic Satellite Data Observation—A Survey. JMSE 11, 340. https://doi.org/10.3390/jmse11020340

Cawley, G.C. y Talbot, N.L.C., 2010. On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation 29.

Chandramouli, V., Brion, G., Neelakantan, T.R., Lingireddy, S., 2007. Backfilling missing microbial concentrations in a riverine database using artificial neural networks. Water Research 41, 217–227. https://doi.org/10.1016/j.watres.2006.08.022

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. jair 16, 321–357. https://doi.org/10.1613/jair.953

Choi, S.-W., Bae, H.-K., 2018. Daily prediction of total coliform concentrations using artificial neural networks. KSCE J Civ Eng 22, 467–474. https://doi.org/10.1007/s12205-017-0739-y

Choi, S.Y., Seo, I.W., 2018. Prediction of fecal coliform using logistic regression and tree-based classification models in the North Han River, South Korea. Journal of Hydro-environment Research 21, 96–108. https://doi.org/10.1016/j.jher.2018.09.002

Christensen, V.G., Rasmussen, P.P., Ziegler, A.C., 2002. Real-time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams. Water Science and Technology 45, 205–219. https://doi.org/10.2166/wst.2002.0240

Crisci, C., Ghattas, B., Perera, G., 2012. A review of supervised machine learning algorithms and their applications to ecological data. Ecological Modelling 240, 113–122. https://doi.org/10.1016/j.ecolmodel.2012.03.001

Crisci, C., Goyenola, G., Terra, R., Lagomarsino, J.J., Pacheco, J.P., Díaz, I., González-Madina, L., Levrini, P., Méndez, G., Bidegain, M., Ghattas, B., Mazzeo, N., 2017. Dinámica ecosistémica y calidad de agua: estrategias de monitoreo para la gestión de servicios asociados a Laguna del Sauce (Maldonado, Uruguay). INNOTEC. https://doi.org/10.26461/13.05

Crowther, J., Kay, D., Wyer, M.D., 2001. Relationships between microbial water quality and environmental conditions in coastal recreational waters: the fylde coast, UK. Water Research 35, 4029–4038. https://doi.org/10.1016/S0043-1354(01)00123-3

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. RANDOM FORESTS FOR CLASSIFICATION IN ECOLOGY. Ecology 88, 2783–2792. https://doi.org/10.1890/07-0539.1

Cyterski, M., Brooks, W., Galvin, M., Wolfe, K., Carvin, R., Fienen, M., Corsi, S., 2013. Virtual Beach 3.0.7: User's Guide 83.

Cyterski, M., Zhang, S., White, E., Molina, M., Wolfe, K., Parmar, R., Zepp, R., 2012. Temporal Synchronization Analysis for Improving Regression Modeling of Fecal Indicator Bacteria Levels. Water Air Soil Pollut 223, 4841–4851. https://doi.org/10.1007/s11270-012-1240-3

Dada, A.C., Hamilton, D.P., 2016. Predictive Models for Determination of E. coli Concentrations at Inland Recreational Beaches. Water Air Soil Pollut 227, 347. https://doi.org/10.1007/s11270-016-3033-6

Dada, C.A., 2019. Seeing is Predicting: Water Clarity-Based Nowcast Models for E. coli Prediction in Surface Water. GJHS 11, 140. https://doi.org/10.5539/gjhs.v11n3p140

David, M.M. y Haggard, B.E., 2011. Development of Regression-Based Models to Predict Fecal Bacteria Numbers at Select Sites within the Illinois River Watershed, Arkansas and

Oklahoma, USA. Water Air Soil Pollut 215, 525–547. https://doi.org/10.1007/s11270-010-0497-7

de Brauwere, A., Ouattara, N.K., Servais, P., 2014. Modeling Fecal Indicator Bacteria Concentrations in Natural Surface Waters: A Review. Critical Reviews in Environmental Science and Technology 44, 2380–2453. https://doi.org/10.1080/10643389.2013.829978

de León, F., 2019. Modelización de la calidad del agua basada en coliformes fecales en playas de La Paloma, Rocha,Uruguay como insumo para la gestión. Centro Universitario de Regional del Este, Universidad de la República, Rocha.

de Souza, R.V., Campos, C.J.A., Garbossa, L.H.P., Seiffert, W.Q., 2018. Developing, cross-validating and applying regression models to predict the concentrations of faecal indicator organisms in coastal waters under different environmental scenarios. Science of The Total Environment 630, 20–31. https://doi.org/10.1016/j.scitotenv.2018.02.139

DINACEA, 2021. Guía para definir la aptitud y la categorización de las playas. Dirección Nacional de Control y Evaluación Ambiental. Ministerio de Ambiente.

Echeverriborda, G., Mesa, F., Chalar, G., Kruk, C., Piccini, C., 2022. Experiencia de aplicación de microorganismos efectivos nativos (MEN) para el tratamiento de aguas residuales. INNOTEC 24. https://doi.org/10.26461/24.06

Eleria, A., y Vogel, R.M., 2005. Predicting fecal coliform bacteria levels in the Charles River, Massachusetts, usa. J Am Water Resources Assoc 41, 1195–1209. https://doi.org/10.1111/j.1752-1688.2005.tb03794.x

EPA, 2010. Predictive Tools for Beach Notification Volume I: Review and Technical Protocol 71.

Framiñan, M.B., Brown, O.B., 1996. Study of the Río de la Plata turbidity front, Part 1: spatial and temporal distribution. Continental Shelf Research 16, 1259–1282. https://doi.org/10.1016/0278-4343(95)00071-2

Fossati, M., Piedra-Cueva, I., 2013. A 3D hydrodynamic numerical model of the Río de la Plata and Montevideo's coastal zone. Applied Mathematical Modelling 37, 1310–1332. https://doi.org/10.1016/j.apm.2012.04.010

Francy, D.S., 2009. Use of predictive models and rapid methods to nowcast bacteria levels at coastal beaches. Aquatic Ecosystem Health & Management 12, 177–182. https://doi.org/10.1080/14634980902905767

Francy, D.S., Stelzer, E.A., Duris, J.W., Brady, A.M.G., Harrison, J.H., Johnson, H.E., Ware, M.W., 2013. Predictive Models for Escherichia coli Concentrations at Inland Lake Beaches and Relationship of Model Variables to Pathogen Detection. Appl Environ Microbiol 79, 1676–1688. https://doi.org/10.1128/AEM.02995-12

Francy, D.S., Brady, A.M.G., Cicale, J.R., Dalby, H.D., Stelzer, E.A., 2020. Nowcasting methods for determining microbiological water quality at recreational beaches and drinking-water source waters. Journal of Microbiological Methods 175, 105970. https://doi.org/10.1016/j.mimet.2020.105970

Frick, W.E., Ge, Z., Zepp, R.G., 2008. Nowcasting and Forecasting Concentrations of Biological Contaminants at Beaches: A Feasibility and Case Study. Environ. Sci. Technol. 42, 4818–4824. https://doi.org/10.1021/es703185p

Garabedian, S., Porteiro, R., Pena, P., 2021. Redes neuronales artificiales para la predicción de flujos de carga aplicadas al sistema de trasmisión de Uruguay. TM. https://doi.org/10.18845/tm.v34i7.6040

Gerber, F., De Jong, R., Schaepman, M.E., Schaepman-Strub, G., Furrer, R., 2018. Predicting Missing Values in Spatio-Temporal Remote Sensing Data. IEEE Trans. Geosci. Remote Sensing 56, 2841–2853. https://doi.org/10.1109/TGRS.2017.2785240

García-Alba, J., Bárcena, J.F., Ugarteburu, C., García, A., 2019. Artificial neural networks as emulators of process-based models to analyse bathing water quality in estuaries. Water Research 150, 283–295. https://doi.org/10.1016/j.watres.2018.11.063

Gazzaz, N.M., Yusoff, M.K., Aris, A.Z., Juahir, H., Ramli, M.F., 2012. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Marine Pollution Bulletin 64, 2409–2420. https://doi.org/10.1016/j.marpolbul.2012.08.005

Ge, Z., Frick, W.E., 2009. Time-Frequency Analysis of Beach Bacteria Variations and its Implication for Recreational Water Quality Modeling. Environ. Sci. Technol. 43, 1128–1133. https://doi.org/10.1021/es8024116

Ge, Z., Frick, W.E., 2007. Some statistical issues related to multiple linear regression modeling of beach bacteria concentrations. Environmental Research 103, 358–364. https://doi.org/10.1016/j.envres.2006.11.006

Gonzalez, R.A., Conn, K.E., Crosswell, J.R., Noble, R.T., 2012. Application of empirical predictive modeling using conventional and alternative fecal indicator bacteria in eastern North Carolina waters. Water Research 46, 5871–5882. https://doi.org/10.1016/j.watres.2012.07.050

Guerrero, R., Piola, A., 1997. Masas de agua en la plataforma continental. 1 107–118.

Han, H., Wang, W.-Y., Mao, B.-H., 2005. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning, in: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (Eds.), Advances in Intelligent Computing, Lecture Notes in Computer Science. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 878–887. https://doi.org/10.1007/11538059_91

Hassan, N., Woo, C.S., 2021. Machine Learning Application in Water Quality Using Satellite Data. IOP Conf. Ser.: Earth Environ. Sci. 842, 012018. https://doi.org/10.1088/1755-1315/842/1/012018

He, L.-M. (Lee), He, Z.-L., 2008. Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern California, USA. Water Research 42, 2563–2573. https://doi.org/10.1016/j.watres.2008.01.002

Heasley, C., Sanchez, J.J., Tustin, J., Young, I., 2021. Systematic review of predictive models of microbial water quality at freshwater recreational beaches. PLoS ONE 16, e0256785. https://doi.org/10.1371/journal.pone.0256785

Heberger, M.G., Durant, J.L., Oriel, K.A., Kirshen, P.H., Minardi, L., 2008. Combining Real-Time Bacteria Models and Uncertainty Analysis for Establishing Health Advisories for Recreational Waters. J. Water Resour. Plann. Manage. 134, 73–82. https://doi.org/10.1061/(ASCE)0733-9496(2008)134:1(73)

Hellweger, F.L., 2007. Ensemble modeling of E. coli in the Charles River, Boston, Massachusetts, USA. Water Science and Technology 56, 39–46. https://doi.org/10.2166/wst.2007.588

Herrig, I.M., Böer, S.I., Brennholt, N., Manz, W., 2015. Development of multiple linear regression models as predictive tools for fecal indicator concentrations in a stretch of the lower Lahn River, Germany. Water Research 85, 148–157. https://doi.org/10.1016/j.watres.2015.08.006

Hirai, F.M., Porto, M.F.D.A., 2016. O desenvolvimento de ferramentas de predição de balneabilidade baseadas em níveis de precipitação: estudo de caso da praia de Cachoeira das Emas (SP). Eng. Sanit. Ambient. 21, 797–806. https://doi.org/10.1590/s1413-41522016131249

Hose, G.C., Murray, B.R., Gordon, G., McCullough, F.E., Pulver, N., 2005. Spatial and rainfall related patterns of bacterial contamination in Sydney Harbour estuary. Journal of Water and Health 3, 349–358. https://doi.org/10.2166/wh.2005.060

Hou, D., Rabinovici, S.J.M., Boehm, A.B., 2006. Enterococci Predictions from Partial Least Squares Regression Models in Conjunction with a Single-Sample Standard Improve the Efficacy of Beach Management Advisories. Environ. Sci. Technol. 40, 1737–1743. https://doi.org/10.1021/es0515250

Huang, B., Thorne, P.W., Banzon, V.F., Boyer, T., Chepurin, G., Lawrimore, J.H., Menne, M.J., Smith, T.M., Vose, R.S., Zhang, H.-M., 2017. NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5. https://doi.org/10.7289/V5T72FNM

Huang, R., Ma, C., Ma, J., Huangfu, X., He, Q., 2021. Machine learning in natural and engineered water systems. Water Research 205, 117666. https://doi.org/10.1016/j.watres.2021.117666

Intendencia de Montevideo, 2023. PROGRAMA DE MONITOREO DE AGUA DE PLAYAS Y COSTA DEL DEPARTAMENTO DE MONTEVIDEO. INFORME ANUAL. Servicio de Evaluación de la Calidad y Control Ambiental.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, Springer Texts in Statistics. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4614-7138-7

Jang, J., Abbas, A., Kim, H., Rhee, C., Shin, S.G., Chun, J.A., Baek, S., Cho, K.H., 2023. Prediction and interpretation of pathogenic bacteria occurrence at a recreational beach using data-driven algorithms. Ecological Informatics 78, 102370. https://doi.org/10.1016/j.ecoinf.2023.102370

Jin, G., Englande, A.J., 2006. Prediction of swimmability in a brackish water body. Management of Environmental Quality: An International Journal 17, 197–208. https://doi.org/10.1108/14777830610650500

Jones, R.M., Liu, L., Dorevitch, S., 2013. Hydrometeorological variables predict fecal indicator bacteria densities in freshwater: data-driven methods for variable selection. Environ Monit Assess 185, 2355–2366. https://doi.org/10.1007/s10661-012-2716-8

Jutz, S., Milagro-Pérez, M.P., 2018. Copernicus Program, in: Comprehensive Remote Sensing. Elsevier, pp. 150–191. https://doi.org/10.1016/B978-0-12-409548-9.10317-3

Kang, J.-H., Lee, S.W., Cho, K.H., Ki, S.J., Cha, S.M., Kim, J.H., 2010. Linking land-use type and stream water quality using spatial data of fecal indicator bacteria and heavy metals in the Yeongsan river basin. Water Research 44, 4143–4157. https://doi.org/10.1016/j.watres.2010.05.009

Kashefipour, S.M., Lin, B., Falconer, R.A., 2005. Neural networks for predicting seawater bacterial levels. Proceedings of the Institution of Civil Engineers - Water Management 158, 111–118. https://doi.org/10.1680/wama.2005.158.3.111

Kaur Dhaliwal, J., Panday, D., Saha, D., Lee, J., Jagadamma, S., Schaeffer, S., Mengistu, A., 2022. Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. Computers and Electronics in Agriculture 199, 107107. https://doi.org/10.1016/j.compag.2022.107107

Kayode, O., Kumarasamy, M., 2018. Assessment of Some Existing Water Quality Models. Nature Environment and Pollution Technology 17, 10.

Kim, H.-C., Son, S., Kim, Y.H., Khim, J.S., Nam, J., Chang, W.K., Lee, J.-H., Lee, C.-H., Ryu, J., 2017. Remote sensing and water quality indicators in the Korean West coast: Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. Marine Pollution Bulletin 121, 425–434. https://doi.org/10.1016/j.marpolbul.2017.05.026

Kim, J.H., Grant, S.B., 2004. Public Mis-Notification of Coastal Water Quality: A Probabilistic Evaluation of Posting Errors at Huntington Beach, California. Environ. Sci. Technol. 38, 2497–2504. https://doi.org/10.1021/es034382v

Kruk, C., Devercelli, M., Huszar, V.L.M., Hernández, E., Beamud, G., Diaz, M., Silva, L.H.S., Segura, A.M., 2017. Classification of Reynolds phytoplankton functional groups using individual traits and machine learning techniques. Freshwater Biology 62, 1681–1692. https://doi.org/10.1111/fwb.12968

Kruk, C., Dobroyan, M., González, L., Segura, A.M., Balado, I., Trabal, N., León, F.D., Martínez, G., Piccini, C., Chalar, G., Verrastro, N., 2018. CALIDAD DE AGUA Y SALUD ECOSISTÉMICA EN PLAYAS RECREATIVAS DE LA PALOMA, ROCHA. Trama 9, 11.

Laureano-Rosario, A.E., Duncan, A.P., Symonds, E.M., Savic, D.A., Muller-Karger, F.E., 2019. Predicting culturable enterococci exceedances at Escambron Beach, San Juan, Puerto Rico using satellite remote sensing and artificial neural networks. Journal of Water and Health 17, 137–148. https://doi.org/10.2166/wh.2018.128

Liu, Z., Peng, C., Work, T., Candau, J.-N., DesRochers, A., Kneeshaw, D., 2018. Application of machine-learning methods in forest ecology: recent progress and future challenges. Environ. Rev. 26, 339–350. https://doi.org/10.1139/er-2018-0034

Liu, X., Wang, M., 2018. Gap Filling of Missing Data for VIIRS Global Ocean Color Products Using the DINEOF Method. IEEE Trans. Geosci. Remote Sensing 56, 4464–4476. https://doi.org/10.1109/TGRS.2018.2820423

Maciel, F., Santoro, P., Cueva, I., Pedocchi, F., 2018. Validación sinóptica de un modelo hidrodinámico del río de la plata mediante teledetección del frente de turbidez. XXVIII Congreso latinoamericano de hidráulica, Buenos Aires.

Maciel, F., Ponce de León, L., Pedocchi, F., 2019. Teledetección de sólidos suspendidos y clorofila-a en aguas costeras turbias:avances para una estimación confiable. VI Simposio sobre métodos experimentales en hidráulica, Paysandú.

Maimone, M., Crockett, C.S., Cesanek, W.E., 2007. PhillyRiverCast: A Real-Time Bacteria Forecasting Model and Web Application for the Schuylkill River. J. Water Resour. Plann. Manage. 133, 542–549. https://doi.org/10.1061/(ASCE)0733-9496(2007)133:6(542)

Mallin, M.A., Williams, K.E., Esham, E.C., Lowe, R.P., 2000. EFFECT OF HUMAN DEVELOPMENT ON BACTERIOLOGICAL WATER QUALITY IN COASTAL WATERSHEDS. Ecological Applications 10, 1047–1056. https://doi.org/10.1890/1051-0761(2000)010[1047:EOHDOB]2.0.CO;2

Mark, O., Erichsen, A.C., 2007. Towards Implementation of the new EU Bathing Water Directive : Case studies: Copenhagen and Århus, Denmark.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions 15, 59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

Martinis, S., Wieland, M., Rättich, M., 2021. An Automatic System for Near-Real Time Flood Extent and Duration Mapping Based on Multi-Sensor Satellite Data, in: Earth Observation for Flood Applications. Elsevier, pp. 7–37. https://doi.org/10.1016/B978-0-12-819412-6.00002-X

Mas, D.M.L., Ahlfeld, D.P., 2007. Comparing artificial neural networks and regression models for predicting faecal coliform concentrations. Hydrological Sciences Journal 52, 713–731. https://doi.org/10.1623/hysj.52.4.713

Mednick, A.C., Services, W.B. of S., 2012. Building Operational "Nowcast" Models for Predicting Water Quality at Five Lake Michigan Beaches. Bureau of Science Services, Wisconsin Department of Natural Resources.

Ministerio de Ambiente, 2022. Evaluación de modelos de turbidez con imágenes Sentinel 2 en el embalse de Palmar, Serie Técnica de la División Información Ambiental. Dirección Nacional de Calidad y Evaluación Ambiental. Minsterio de Ambiente.

Molina, M., Hunter, S., Cyterski, M., Peed, L.A., Kelty, C.A., Sivaganesan, M., Mooney, T., Prieto, L., Shanks, O.C., 2014. Factors affecting the presence of human-associated and fecal indicator real-time quantitative PCR genetic markers in urban-impacted recreational beaches. Water Research 64, 196–208. https://doi.org/10.1016/j.watres.2014.06.036

Muniz, P., Venturini, N., Brugnoli, E., Gutiérrez, J.M., Acuña, A., 2019. Río de la Plata: Uruguay, in: World Seas: An Environmental Evaluation. Elsevier, pp. 703–724. https://doi.org/10.1016/B978-0-12-805068-2.00036-X

Nafsin, N., Li, J., 2023. Prediction of total organic carbon and *E. coli* in rivers within the Milwaukee River basin using machine learning methods. Environ. Sci.: Adv. 2, 278–293. https://doi.org/10.1039/D2VA00285J

Nagy, G., Martinez, C., Caffera, M., Pedrosa, G., Forbes, E., Perdomo, A., Laborde, J., 1997. The Hydrological and Climatic setting of the Río de la Plata, in: The Río de La Plata. An Environmental Overview. ECOPLATA Projectbackgroung Report, Dalhousie Univ, Halifax, Nova Scotia, pp. 17–68.

Nevers, M.B., Whitman, R.L., 2011. Efficacy of monitoring and empirical predictive modeling at improving public health protection at Chicago beaches. Water Research 45, 1659–1668. https://doi.org/10.1016/j.watres.2010.12.010

Nevers, M.B., Whitman, R.L., 2008. Coastal Strategies to Predict Escherichia coli Concentrations for Beaches along a 35 km Stretch of Southern Lake Michigan. Environ. Sci. Technol. 42, 4454–4460. https://doi.org/10.1021/es703038c

Nevers, M.B., Whitman, R.L., 2005. Nowcast modeling of Escherichia coli concentrations at multiple urban beaches of southern Lake Michigan. Water Research 39, 5250–5260. https://doi.org/10.1016/j.watres.2005.10.012

Nevers, M.B., Whitman, R.L., Frick, W.E., Ge, Z., 2007. Interaction and Influence of Two Creeks on *Escherichia coli* Concentrations of Nearby Beaches: Exploration of Predictability and Mechanisms. J. Environ. Qual. 36, 1338–1345. https://doi.org/10.2134/jeq2007.0025

Olyphant, G.A., 2005. Statistical basis for predicting the need for bacterially induced beach closures: Emergence of a paradigm? Water Research 39, 4953–4960. https://doi.org/10.1016/j.watres.2005.09.031

Olyphant, G.A., Whitman, R.L., 2004. Elements of a Predictive Model for Determining Beach Closures on a Real Time Basis: The Case of 63rd Street Beach Chicago. Environ Monit Assess 98, 175–190. https://doi.org/10.1023/B:EMAS.0000038185.79137.b9

Organización Mundial de la Salud (Ed.), 2003. Guidelines for safe recreational water environments. World Health Organization, Geneva, Switzerland.

Organización Mundial de la Salud, 2018. WHO Recommendations on Scientific, Analytical & Epidemiological Developments Relevant to the Parameters for Bathing Water Quality in the Bathing Water Directive (2006/7/EC).

Organización Mundial de la Salud, 2021. Guidelines on recreational water quality. Volume 1, Coastal and fresh waters, 2021. Geneva, Switzerland.

Palazón, A., López, I., Aragonés, L., Villacampa, Y., Navarro-González, F.J., 2017. Modelling of Escherichia coli concentrations in bathing water at microtidal coasts. Science of The Total Environment 593–594, 173–181. https://doi.org/10.1016/j.scitotenv.2017.03.161

Park, J., Kim, H.-C., Bae, D., Jo, Y.-H., 2020. Data Reconstruction for Remotely Sensed Chlorophyll-a Concentration in the Ross Sea Using Ensemble-Based Machine Learning. Remote Sensing 12, 1898. https://doi.org/10.3390/rs12111898

Park, Y., Kim, M., Pachepsky, Y., Choi, S., Cho, J., Jeon, J., Cho, K.H., 2018. Development of a Nowcasting System Using Machine Learning Approaches to Predict Fecal Contamination Levels at Recreational Beaches in Korea. J. environ. qual. 47, 1094–1102. https://doi.org/10.2134/jeq2017.11.0425

Parker, G., y López-Laborde, J. 1989. Morfología y variaciones morfológicas del lecho del Río de la Plata. In SHIN-SOHMA (Divs Geología Marina)

Parkhurst, D.F., Brenner, K.P., Dufour, A.P., Wymer, L.J., 2005. Indicator bacteria at five swimming beaches—analysis using random forests. Water Research 39, 1354–1360. https://doi.org/10.1016/j.watres.2005.01.001

Piola, A., Campos, E., Moller, O., Charo, M., Martinez, C., 2000. Subtropical Shelf Front off eastern South America. Journal of geophysical research 105, 6565–6578.

Plan Nacional de Saneamiento, 2020. Plan Nacional de Saneamiento. Ministerio de Ordenamiento Territorial y Medio Ambiente.

Pras, A., Mamane, H., 2023. Nowcasting of fecal coliform presence using an artificial neural network. Environmental Pollution 326, 121484. https://doi.org/10.1016/j.envpol.2023.121484

Ramírez, C.A., 2022. Aplicación de la metodología de Ciencia de Datos para analizar datos de facturación de energía eléctrica. Caso de estudio: Uruguay 2000-2022. Rev.Investig.sist.inform. 15, 127–138. https://doi.org/10.15381/risi.v15i1.23544

Rossi, A., Wolde, B.T., Lee, L.H., Wu, M., 2020. Prediction of recreational water safety using Escherichia coli as an indicator: case study of the Passaic and Pompton rivers, New Jersey. Science of The Total Environment 714, 136814. https://doi.org/10.1016/j.scitotenv.2020.136814

Sabino, R., Rodrigues, R., Costa, I., Carneiro, C., Cunha, M., Duarte, A., Faria, N., Ferreira, F.C., Gargaté, M.J., Júlio, C., Martins, M.L., Nevers, M.B., Oleastro, M., Solo-Gabriele, H., Veríssimo, C., Viegas, C., Whitman, R.L., Brandão, J., 2014. Routine screening of harmful microorganisms in beach sands: Implications to public health. Science of The Total Environment 472, 1062–1069. https://doi.org/10.1016/j.scitotenv.2013.11.091

Santoro, P., Fossati, M., Tassi, P., Huybrechts, N., Pham Van Bang, D., Piedra-Cueva, J.C.I., 2017. A coupled wave–current–sediment transport model for an estuarine system: Application to the Río de la Plata and Montevideo Bay. Applied Mathematical Modelling 52, 107–130. https://doi.org/10.1016/j.apm.2017.07.004

Schollaert Uz, S., Kim, G.E., Mannino, A., Werdell, P.J., Tzortziou, M., 2019. Developing a Community of Practice for Applied Uses of Future PACE Data to Address Marine Food Security Challenges. Front. Earth Sci. 7, 283. https://doi.org/10.3389/feart.2019.00283

Searcy, R.T., Taggart, M., Gold, M., Boehm, A.B., 2018. Implementation of an automated beach water quality nowcast system at ten California oceanic beaches. Journal of Environmental Management 223, 633–643. https://doi.org/10.1016/j.jenvman.2018.06.058

Segura, A.M., Piccini, C., Nogueira, L., Alcántara, I., Calliari, D., Kruk, C., 2017. Increased sampled volume improves Microcystis aeruginosa complex (MAC) colonies detection and prediction using Random Forests. Ecological Indicators 79, 347–354. https://doi.org/10.1016/j.ecolind.2017.04.047

Segura, A.M., Sampognaro, L., Lopez, G., Crisci, C., Bourel, M., Vidal, V., Eirin, K., Piccini, C., Kurk, C., Perera, G., 2021. Monitoreo de calidad de agua y predicción de coliformes

fecales en playas de Montevideo mediante algoritmos de aprendizaje automático. INNOTEC 22. https://doi.org/10.26461/22.07

Seis, W., Zamzow, M., Caradot, N., Rouault, P., 2018. On the implementation of reliable early warning systems at European bathing waters using multivariate Bayesian regression modelling. Water Research 143, 301–312. https://doi.org/10.1016/j.watres.2018.06.057

Serron, A., Coitiño, H., Segura, A., n.d. Atropellos de mamíferos en la Región Este de Uruguay y su relación con los atributos del paisaje. 2020 20, 139–157. https://doi.org/10.12461/20.05

Shanks, O.C., Nietch, C., Simonich, M., Younger, M., Reynolds, D., Field, K.G., 2006. Basin-Wide Analysis of the Dynamics of Fecal Contamination and Fecal Source Identification in Tillamook Bay, Oregon. Appl Environ Microbiol 72, 5537–5546. https://doi.org/10.1128/AEM.03059-05

Shirmard, H., Farahbakhsh, E., Müller, R.D., Chandra, R., 2022. A review of machine learning in processing remote sensing data for mineral exploration. Remote Sensing of Environment 268, 112750. https://doi.org/10.1016/j.rse.2021.112750

Shively, D.A., Nevers, M.B., Breitenbach, C., Phanikumar, M.S., Przybyla-Kelly, K., Spoljaric, A.M., Whitman, R.L., 2016. Prototypic automated continuous recreational water quality monitoring of nine Chicago beaches. Journal of Environmental Management 166, 285–293. https://doi.org/10.1016/j.jenvman.2015.10.011

Shuval, H., 2003. Estimating the global burden of thalassogenic diseases: human infectious diseases caused by wastewater pollution of the marine environment. Journal of Water and Health 1, 53–64. https://doi.org/10.2166/wh.2003.0007

Sokolova, E., Ivarsson, O., Lilliesträm, A., Speicher, N.K., Rydberg, H., Bondelind, M., 2022. Data-driven models for predicting microbial water quality in the drinking water source using E. coli monitoring and hydrometeorological data. Science of The Total Environment 802, 149798. https://doi.org/10.1016/j.scitotenv.2021.149798

Soumastre, M., 2016. Echeverriborda, G., Mesa, F., Chalar, G., Kruk, C., Piccini, C., 2022. Experiencia de aplicación de microorganismos efectivos nativos (MEN) para el tratamiento de aguas residuales. Instituto de Investigaciones Biológicas Clemente Estable, Montevideo, Uruguay.

Stampoulis, D., Damavandi, H.G., Future H2O, Office of Knowledge Enterprise Development, Arizona State University, Tempe, AZ 85281, USA, Boscovic, D., Center for Assured and Scalable Data Engineering, Arizona State University, Tempe, AZ 85281, USA, Sabo, J., Future H2O, Office of Knowledge Enterprise Development, Arizona State University, Tempe, AZ 85281, USA, 2020. Using Satellite Remote Sensing and Machine Learning Techniques Towards Precipitation Prediction and Vegetation Classification. J ENVIRON INFORM. https://doi.org/10.3808/jei.202000427

Stidson, R.T., Gray, C.A., McPhail, C.D., 2012. Development and use of modelling techniques for real-time bathing water quality predictions: Development and use of modelling techniques. Water and Environment Journal 26, 7–18. https://doi.org/10.1111/j.1747-6593.2011.00258.x

Tabo, Z., Neubauer, T.A., Tumwebaze, I., Stelbrink, B., Breuer, L., Hammoud, C., Albrecht, C., 2022. Factors Controlling the Distribution of Intermediate Host Snails of Schistosoma in Crater Lakes in Uganda: A Machine Learning Approach. Front. Environ. Sci. 10, 871735. https://doi.org/10.3389/fenvs.2022.871735

Tahmasebi, P., Kamrava, S., Bai, T., Sahimi, M., 2020. Machine learning in geo- and environmental sciences: From small to large scale. Advances in Water Resources 142, 103619. https://doi.org/10.1016/j.advwatres.2020.103619

Thoe, W., Choi, K.W., Lee, J.H., 2016. Predicting 'very poor' beach water quality gradings using classification tree. Journal of Water and Health 14, 97–108. https://doi.org/10.2166/wh.2015.094

Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M.L., Boehm, A.B., 2015. Sunny with a Chance of Gastroenteritis: Predicting Swimmer Risk at California Beaches. Environ. Sci. Technol. 49, 423–431. https://doi.org/10.1021/es504701j

Thoe, W., Gold, M., Griesbach, A., Grimmer, M., Taggart, M.L., Boehm, A.B., 2014. Predicting water quality at Santa Monica Beach: Evaluation of five different models for public notification of unsafe swimming conditions. Water Research 67, 105–117. https://doi.org/10.1016/j.watres.2014.09.001

Thoe, W., Lee, J.H.W., 2014. Daily Forecasting of Hong Kong Beach Water Quality by Multiple Linear Regression Models. J. Environ. Eng. 140, 04013007. https://doi.org/10.1061/(ASCE)EE.1943-7870.0000800

Thoe, W., Lee, O.H.K., Leung, K.F., Lee, T., Ashbolt, N.J., Yang, R.R., Chui, S.H.K., 2018. Twenty five years of beach monitoring in Hong Kong: A re-examination of the beach water quality classification scheme from a comparative and global perspective. Marine Pollution Bulletin 131, 793–803. https://doi.org/10.1016/j.marpolbul.2018.05.002

Thoe, W., Wong, S.H.C., Choi, K.W., Lee, J.H.W., 2012. Daily prediction of marine beach water quality in Hong Kong. Journal of Hydro-environment Research 6, 164–180. https://doi.org/10.1016/j.jher.2012.05.003

Tselemponis, Athanasios, Stefanis, C., Giorgi, E., Kalmpourtzi, A., Olmpasalis, I., Tselemponis, Antonios, Adam, M., Kontogiorgis, C., Dokas, I.M., Bezirtzoglou, E., Constantinidis, T.C., 2023. Coastal Water Quality Modelling Using E. coli, Meteorological Parameters and Machine Learning Algorithms. IJERPH 20, 6216. https://doi.org/10.3390/ijerph20136216

Tufail, M., Ormsbee, L., Teegavarapu, R., 2008. Artificial Intelligence-Based Inductive Models for Prediction and Classification of Fecal Coliform in Surface Waters. J. Environ. Eng. 134, 789–799. https://doi.org/10.1061/(ASCE)0733-9372(2008)134:9(789)

Uruguay, 2009. Uruguay. Decreto 253/979, de 09 de mayo de 2009. Diario Oficial, 31 de mayode 1979, p.1473.

USEPA, 2012. Recreational Water Quality Criteria. Office of Science and Technology, United States (U.S.) Environmental Protection Agency (EPA),.

Velardez, O., Dima, G., 2022. Desarrollo de una herramienta de aprendizaje automático (machine learning) para establecer relaciones entre ocupaciones y programas de capacitación en el Uruguay (Documentos de Proyectos (LC/TS.2022/2)). Comisión Económica para América Latina y el Caribe (CEPAL).

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review. Renewable Energy 105, 569–582. https://doi.org/10.1016/j.renene.2016.12.095

Wade, T.J., Calderon, R.L., Sams, E., Beach, M., Brenner, K.P., Williams, A.H., Dufour, A.P., 2006. Rapidly Measured Indicators of Recreational Water Quality Are Predictive of Swimming-Associated Gastrointestinal Illness. Environ Health Perspect 114, 24–28. https://doi.org/10.1289/ehp.8273

Wang, Q., Li, S., Jia, P., Qi, C., Ding, F., 2013. A Review of Surface Water Quality Models. The Scientific World Journal 2013, 1–7. https://doi.org/10.1155/2013/231768

Watthaisong, T., Sunat, K., Muangkote, N., 2024. Comparative Evaluation of Imbalanced Data Management Techniques for Solving Classification Problems on Imbalanced Datasets. Stat., optim. inf. comput. 12, 547–570. https://doi.org/10.19139/soic-2310-5070-1890

Whitman, R.L., Nevers, M.B., 2008. Summer *E. coli* Patterns and Responses along 23 Chicago Beaches. Environ. Sci. Technol. 42, 9217–9224. https://doi.org/10.1021/es8019758

Whitman, R.L., Nevers, M.B., Korinek, G.C., Byappanahalli, M.N., 2004. Solar and Temporal Effects on *Escherichia coli* Concentration at a Lake Michigan Swimming Beach. Appl Environ Microbiol 70, 4276–4285. https://doi.org/10.1128/AEM.70.7.4276-4285.2004

Xu, T., Coco, G., Neale, M., 2020. A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning. Water Research 177, 115788. https://doi.org/10.1016/j.watres.2020.115788

Zabaleta, B.V., Aubriot, L., Olano, H., Achkar, M., 2022. Satellite assessment of eutrophication hot spots and algal blooms in small and medium-sized productive reservoirs in Uruguay's main drinking water basin (preprint). In Review. https://doi.org/10.21203/rs.3.rs-1886972/v1

Zhang, Z., Deng, Z., Rusch, K.A., 2012. Development of predictive models for determining enterococci levels at Gulf Coast beaches. Water Research 46, 465–474. https://doi.org/10.1016/j.watres.2011.11.027

Zhang, Z., Deng, Z., Rusch, K.A., 2015. Modeling Fecal Coliform Bacteria Levels at Gulf Coast Beaches. Water Qual Expo Health 7, 255–263. https://doi.org/10.1007/s12403-014-0145-3

Zhang, J., Qiu, H., Li, X., Niu, J., Nevers, M.B., Hu, X., Phanikumar, M.S., 2018. Real-Time Nowcasting of Microbiological Water Quality at Recreational Beaches: A Wavelet and Artificial Neural Network-Based Hybrid Modeling Approach. Environ. Sci. Technol. 52, 8446–8455. https://doi.org/10.1021/acs.est.8b01022

Zhong, S., Zhang, K., Bagheri, M., Burken, J.G., Gu, A., Li, B., Ma, X., Marrone, B.L., Ren, Z.J., Schrier, J., Shi, W., Tan, H., Wang, T., Wang, X., Wong, B.M., Xiao, X., Yu, X., Zhu, J.-J., Zhang, H., 2021. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. Environ. Sci. Technol. acs.est.1c01339. https://doi.org/10.1021/acs.est.1c01339

Zhang, X., Zhou, M., 2023. A General Convolutional Neural Network to Reconstruct Remotely Sensed Chlorophyll-a Concentration. JMSE 11, 810. https://doi.org/10.3390/jmse11040810

Zhang, S., He, R., Wang, Q., Qu, Z., Wang, J., Wang, Y., Ren, H., 2023. Machine-Learning-Based Approach To Assessing Water Quality in a Specific Basin: The Case of Wujingang Basin. ACS EST Water acsestwater.3c00153. https://doi.org/10.1021/acsestwater.3c00153

ANEXO

Tabla 1. Lista de los artículos incluidos en la revisión desde el 2000 hasta el 2024. Se muestra el país de origen, el sitio de estudio, el tipo de ambiente (costero-marino, agua dulce o estuarino), la variable a predecir (Var_y; *Escherichia coli* (EC), coliformes fecales (CF), coliformes totales (CT), enterococos (ENT)) el método de modelización utilizado (Métodos de Regresión y Métodos de Clasificación), el tipo de aprendizaje (supervisado, no supervisado) y la referencia bibliográfica.

País	Sitio	Tipo de ambiente	Var_y	Métodos de Regresión	Métodos de Clasificación	Supervisado /No supervisado	Referencia
Nueva Zelanda	Río Oreti, Wallacetown	agua dulce	EC	DynReg; Naive model; MLR; CART y RF	Markov chain;CART y RF; MlogR; LDA, QDA; BN	Supervisado	Avila et al. (2018)
Estados Unidos	Playa Aliso, Orange County	costero-marinos	CF, CT, ENT	CART		Supervisado	Bae et al. (2010)
Estados Unidos	Lago Erie beaches (Huntington, Edgewater, Maumee, Upper lake park)	agua dulce	EC	PLS; OLS		Supervisado	Brooks et al. (2013)
Estados Unidos	Grandes lagos de Winsconsin	agua dulce	EC	MLR; GBM; PLS; Persistence	BLR	Supervisado	Brooks et al. (2016)
Estados Unidos	Río Kentuky	agua dulce	CF	ANN	ANN	Supervisado	Chandramoulli et al. (2007)
Estados Unidos	Playa Aliso, Orange County	costero-marinos	TC	ANN		Supervisado	Choi and Bae (2018)
Korea del Sur	Río Daesung-ri	agua dulce	CF		LRM; CART; BGM; RF	Supervisado	Choi and Seo (2018)
Estados Unidos	Río Kansas	agua dulce	CF	MLR		Supervisado	Christensen et al. (2002)
United Kingdom	Playas Fylde	costero-marinos	CF, CT	MLR		Supervisado	Crowther et al. (2001)
Estados Unidos	South Shore, Milwaukee	agua dulce	ENT	MLR		Supervisado	Cyterski et al. (2012)
Nueva Zelanda	Ríos de New Zeland	agua dulce				Supervisado	Dada (2019)
Nueva Zelanda	Río Rotorua beaches	agua dulce	EC	MLR-VB		Supervisado	Dada and Hamilton (2016)
Estados Unidos	Río Illinois	agua dulce	EC, CF	MLR		Supervisado	David and Haggard (2011)

Brasil	Bahía Norte y Sur, Santa Catarina	costero-marinos	FIO	MLR		Supervisado	de Souza et al. (2018)
Estados Unidos	Río Larz Anderson Bridge	agua dulce	CF	MLR	BLR	Supervisado	Eleria and Vogel (2005)
Estados Unidos	Lago Ohio	agua dulce	EC	MLR-VB		Supervisado	Francy et al. (2013)
Estados Unidos	Ríos de Ohio, Pensilvania and New York	agua dulce	EC	MLR		Supervisado	Francy et al. (2020)
Estados Unidos	Lago Huntington	agua dulce	EC	MLR	Static model	Supervisado	Frick et al. (2008)
España	Estuario Eo	estuarinos	EC	ANN; Procces-based-model		Supervisado	Garcia-Alba et al. (2019)
Malasia	Río Kinta	agua dulce	EC, TC	ANN		Supervisado	Gazzaz et al. (2012)
Estados Unidos	Lago Erie, Huntington, Ohio	agua dulce	EC	MLR		Supervisado	Ge and Frick (2007)
Estados Unidos	Lago Erie, Huntington, Ohio	agua dulce	EC	WA; MLR		Supervisado	Ge and Frick (2009)
Estados Unidos	Estuario Ware y Oyster Creek	estuarinos	EC, ENT	MLR		Supervisado	Gonzalez et al. (2012)
Estados Unidos	Playas San Elijo y Torrey Pines State Beaches	costero-marinos	EC, TC, ENT	ANN		Supervisado	He and He (2008)
Estados Unidos	Río Mystic	agua dulce	ENT	MLR		Supervisado	Heberger et al. (2008)
Estados Unidos	Playas de San Diego	costero-marinos	EC	MLR; HYDRO+MLR		Supervisado	Hellweger (2007)
Alemania	Río Lahn	agua dulce	EC, ENT	MLR		Supervisado	Herrig et al. (2015)
Brasil	Río Mogi-Guaçu, Pirassununga	agua dulce	CT	MLR		Supervisado	Hirai and Porto (2016)
Australia	Bahía de Sydney (Port Jackson)	estuarinos	CF, ENT	GLM; MLR		Supervisado	Hose et al (2005)
Estados Unidos	Playa Huntington State and Huntington City Beach	costero-marinos	ENT	PLS		Supervisado	Hou et al. (2006)
Estados Unidos	Lago Pontchartrain, playa Lincon	estuarinos	EC, CF, ENT		PNN; MLogR	Supervisado	Jin and Englande (2006)
Estados Unidos	Lagos Chicago Area Waterway System y Michigan	agua dulce	EC, ENT	CART; RF; LME		Supervisado	Jones et al. (2013)
Inglaterra	Playas Firth of Clyde, Scotland	costero-marinos	CF	ANN		Supervisado	Kashefipour et al. (2005)

Inglaterra	Playas de Firth of Clyde	costero-marinos	CF	ANN		Supervisado	Lin et al. (2003)
Inglaterra	Estuario Ribble	estuarinos	CF	ANN		Supervisado	Lin et al. (2008)
Estados Unidos	Río Schuylkill	agua dulce	EC, CF	MLR		Supervisado	Maimone et al. (2007)
Dinamarca	Bahía de Århus	costero-marinos	EC, ENT				Mark and Erichsen (2021)
Estados Unidos	Lago Gates Brook , Massachuset	agua dulce	CF	MLR; ANN	BLR	Supervisado	Mas and Ahlfeld (2007)
Estados Unidos	Lago Michigan, Wisconsin	agua dulce	EC	MLR		Supervisado	Mednick (2012)
Estados Unidos	Playas Porter and Lake Counties, Indiana	agua dulce	EC	MLR		Supervisado	Nevers and Whitman (2005)
Estados Unidos	Lago Michigan, Indiana	agua dulce	EC	MLR		Supervisado	Nevers and Whitman (2008)
Estados Unidos	Lago Michigan	agua dulce	EC	MLR		Supervisado	Nevers and Whitman (2011)
Estados Unidos	Playas Mount Baldy and Central Avenue	agua dulce	EC	MLR		Supervisado	Nevers et al. (2007)
Estados Unidos	Playas Illinois and Indiana	agua dulce	EC	MLR		Supervisado	Olyphant (2005)
Estados Unidos	Playa 63rd Street, Chicago	agua dulce	EC	MLR		Supervisado	Olyphant and Whitman (2004)
Corea	Playas Haeundae and Gwangalli, Busan city	costero-marinos	EC, ENT	ANN; SVR		Supervisado	Park et al. (2018)
Estados Unidos	Playas Wollaston, Imperial, Belle Isle Park y West, Miami	costero-marino/estuarinos /agua dulce	EC, ENT	RF		Supervisado	Parkhurst et al. (2005)
Estados Unidos	Playas Pompton and Passaic, Nueva Jersey	agua dulce	EC		BLR	Supervisado	Rossi et al. (2020)
Estados Unidos	Río Tangipahoa	agua dulce	CF	NPMR		Supervisado	Schulz and Childers (2011)
Estados Unidos	Playas de California	costero-marinos	CF, TC, ENT	MLR	BLR	Supervisado	Searcy et al. (2018)
Alemania	Río Havel	agua dulce	EC	BN		No Supervisado	Seis et al. (2018)
Estados Unidos	Playas de Illinois	agua dulce	EC	MLR; RF		Supervisado	Shively et al. (2016)

Inglaterra	Playas Troon, Irvine, Saltcoats/Ardrossan and Ettrick Bay	costero-marinos	EC	DynReg; Naive model; MLR; CART y RF		Supervisado	Stidson et al. (2011)
China	Playa Hong Kong	costero-marinos	EC	MLR; ANN		Supervisado	Thoe et al. (2012)
Estados Unidos	Playas de Santa Monica, Los Angeles	costero-marinos	EC, ENT	MLR; ANN; PLS	BLR; CART	Supervisado	Thoe et al. (2014)
Estados Unidos	Playas de California	costero-marinos	CF, TC, ENT	MLR; ANN; PLS	BLR; CART	Supervisado	Thoe et al. (2015)
China	Playa Hong Kong	costero-marinos	EC	MLR	CART	Supervisado	Thoe et al. (2016)
Estados Unidos	Río Kentucky	agua dulce	CF	MLR; ANN; GA		Supervisado	Tufail et al. (2008)
Estados Unidos	Playas de Chicago	costero-marinos	EC	MLR		Supervisado	Whitman and Nevers (2008)
Nueva Zelanda	Playas Milford, Narrow Neck, Judges Bay, Weymouth y Clarks	costero-marinos	FIB		ANN; KNN; SVM; BDT	Supervisado	Xu et al. (2020)
Estados Unidos	Playas del golfo, Louisiana	costero-marinos	ENT	ANN		Supervisado	Zhang et al. (2012)
Estados Unidos	Playa Holly, Louisiana	costero-marinos	CF	MLR-VB; ANN		Supervisado	Zhang et al. (2015)
Estados Unidos	Lago Michigan	agua dulce	EC	ANN (NARX; NIO; NAR); WA-NAR		Supervisado	Zhang et al. (2018)
Estados Unidos	Playa Escambron, San Juan	costero-marinos	ENT		ANN	Supervisado	Laureano-Rosario et al. (2019)
Uruguay	Playas de Montevideo	estuarinos	EC		RF	Supervisado	Segura et al. (2021)
Uruguay	Playas de Montevideo	estuarinos	EC		RF	Supervisado	Bourel et al. (2021)

Tabla 2. Lista de modelos predictivos de contaminación fecal registrados en la revisión bibliográfica desde 2000 al 2024. Se muestra el tipo de modelo (supervisado, no supervisado), el método de modelización (Regresión o Clasificación), los nombres de los modelos y sus acrónimos.

Tipo de modelo	**Método de Regresión/Clasificación**	**Modelo**	**Acrónimo**
Supervisado			
	Métodos de Regresión	Modelos Lineales Míxtos	LME
	Métodos de Regresión	Regresión multiplicativa no paramétrica	NPMR
	Métodos de Regresión	Modelos Lineales Generalizados	GLM
	Métodos de Regresión	Modelos Generalizados aditivos	GAMs
	Métodos de Regresión	Regresión dinámica	DynReg
	Métodos de Regresión	Regresiones Lineales Múltiples	MLR

	Métodos de Regresión	Regresiones de Mínimos Cuadrados Parciales	PLS
	Métodos de Regresión	Regresiones de Mínimos Cuadrados Ordinarios	OLS
	Métodos de Regresión	Modelo de persistencia	Persistence
	Ambos Regresión y Clasificación	Modelo generalizado boosted	GBM
	Ambos Regresión y Clasificación	Modelo de decisión Bagging	BGM
	Ambos Regresión y Clasificación	Árboles de decisión Boosting	BDT
	Ambos Regresión y Clasificación	Redes Neuronales Artificiales	ANN
	Ambos Regresión y Clasificación	Modelo Input-Output no-lineal	NIO
	Ambos Regresión y Clasificación	Modelo exógeno autorregresivo no lineal	NAR(X)
	Ambos Regresión y Clasificación	Redes Neuronales probabilísticas	PNN
	Ambos Regresión y Clasificación	Máquinas de vectores de soporte	SVM
	Ambos Regresión y Clasificación	Regresión de vectores de soporte	SVR
	Ambos Regresión y Clasificación	Vecinos k-nearest	KNN
	Ambos Regresión y Clasificación	Análisis lineal discriminante	LDA
	Ambos Regresión y Clasificación	Análisis cuadrático discriminante	QDA
	Ambos Regresión y Clasificación	Random Forest	RF
	Ambos Regresión y Clasificación	Árboles de Regresión/ Clasificación	CART
	Ambos Regresión y Clasificación	Análisis Wavelet	WA
	Ambos Regresión y Clasificación	Algoritmo genético de conjunto funcional fijo	FFSGA
	Métodos de Clasificación	Modelo Naive	Naive model
	Métodos de Clasificación	Regresión logística Binaria	BLR
	Métodos de Clasificación	Regresión Logística Múltiple	MlogR
	Métodos de Clasificación	Modelo Markov Chain	Markov Chain
No supervisado		Redes Bayesianas	BN

Tabla 3. Lista de las variables input registradas durante la revisión bibliográfica de modelos predictivos de contaminación fecal en ambientes costero-marinos y ambientes de agua dulce y sus acrónimos.

Ambientes costero-marinos		**Ambientes de agua dulce**	
Input	**Acrónimo**	**Input**	**Acrónimo**
Acumulación de precipitaciones en las 72 horas anteriores	Pp72hs	Acumulación de precipitaciones en las 24 horas anteriores	Pp24hs
Acumulación de precipitaciones en las 24 horas anteriores	Pp24hs	Acumulación de precipitaciones en las 72 horas anteriores	Pp72hs
Acumulación de precipitaciones en las 48 horas anteriores	Pp48hs	Temperatura del agua	Temp. del agua
Temperatura del agua	Temp. del agua	Turbidez	Turbidez
Nivel de marea	NM	Acumulación de precipitaciones en las 48 horas anteriores	Pp48hs
Radiación solar	Radiación solar	Radiación solar	Radiación solar
Velocidad del viento	Vel. Del viento	Velocidad del viento	Vel. Del viento
Dirección del viento	Dir. del viento	Descarga del río	Descarga
Salinidad	Salinidad	Dirección del viento	Dir. del viento
Turbidez	Turbidez	pH	pH
Altura de Ola	Altura de la ola	Oxígeno disuelto	OD
Oxígeno disuelto	OD	Caudales de las vertientes	Caudales de vertinetes
Rango de marea	RM	Antecedentes de niveles de precipitaciones	ANT rainfall
Antecedentes de niveles de FIB	ANT. FIB	Alto de marea	TH
Niveles previos de FIB	Lag_FIB	Temperatura del aire	AT
Caudales de las vertientes	Caudales de vertientes	Salinidad	Salinidad
Tasa de flujo del caudal	T.F.C	Altura de Ola	Altura de Ola
Conductividad	cond	Dirección del viento	Dir. del viento
Temperatura superficial del agua satelital	SST	Niveles previos de *E.coli*	Lag_FIB
Irradiancia	DNI	Amonio	NH4
Punto de rocío	DP	Antecedentes de niveles de FIB	Lag_FIB
		Conductividad	cond
		Tasa de flujo del caudal	T.F.C
		Presencia de aves	Aves
		Algas	Algal mat
		Clorofila-a	Clo-a
		Tiempo desde la última precipitación	T_no lluvia
		Número de usuarios de la playa	N° usuarios