



Monitoreo de calidad de agua y predicción de coliformes fecales en playas de Montevideo mediante algoritmos de aprendizaje automático

Water quality prediction using machine learning algorithms in recreational beaches from Montevideo, Uruguay

Previsão da qualidade da água usando modelos de aprendizado de máquina em praias recreativas de Montevideú, Uruguai

Ángel Segura¹, * <https://orcid.org/0000-0002-1989-8899>

Lía Sampognaro¹ <https://orcid.org/0000-0002-7718-9820>

Guzmán López¹ <https://orcid.org/0000-0002-1343-492X>

Carolina Crisci¹ <https://orcid.org/0000-0002-3089-8048>

Mathías Bourel² <https://orcid.org/0000-0002-7472-7179>

Victoria Vidal¹ <https://orcid.org/0000-0002-8623-7804>

Karina Eirin¹ <https://orcid.org/0000-0002-6588-4738>

Claudia Piccini³ <https://orcid.org/0000-0002-2762-1953>

Carla Kruk⁴ <https://orcid.org/0000-0003-0760-1186>

Gonzalo Perera^{1,2} <https://orcid.org/0000-0002-7530-3503>

*Autor de contacto: amsegura@fcien.edu.uy

¹Modelización Estadística de Datos e Inteligencia Artificial (MEDIA), Centro Universitario Regional Este (CURE), Universidad de la República. Rocha, Uruguay.

²Instituto de Matemática y Estadística Prof. Rafael Laguardia, Facultad de Ingeniería, Universidad de la República. Montevideo, Uruguay.

³Instituto de Investigaciones Biológicas Clemente Estable. Ministerio de Educación y Cultura. Montevideo, Uruguay.

⁴Instituto de Ecología y Ciencias Ambientales (IECA), Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay.

Recepción: 4 Agosto 2020 **Aprobación:** 16 Junio 2021

Esta obra está bajo una Licencia Creative Commons Atribución NoComercial 4.0 Internacional.

Resumen: En este trabajo se construyeron modelos de predicción de coliformes fecales (CF) para su aplicación en la gestión de calidad de agua de playas recreativas de Montevideo. Se analizó la base de datos histórica del monitoreo de CF en playas realizado por el Laboratorio de Calidad Ambiental de la Intendencia de Montevideo (IM), y se generaron modelos basados en Aprendizaje Automático (AA) para predecir excesos ($CF > 2.000$). Los datos abarcan 10 años de monitoreo en 21 playas de la capital ($N=19359$, noviembre 2009 a septiembre 2019), y presentan un amplio rango de salinidad y turbidez, con marcadas diferencias entre playas. Los CF presentaron una distribución asimétrica (mínima=4, mediana=250, media=1.047 y máxima=1.280.000) con excesos a la normativa en todas las playas. Las variables registradas *in situ*, meteorológicas y oceanográficas, fueron utilizadas para entrenar modelos de AA. El mejor modelo fue un bosque aleatorio estratificado con un porcentaje de acierto para los excesos del 86%. La predicción de excesos mejoró un 60% respecto al criterio actual de cierre de playas las 24 horas posteriores a precipitaciones. La generación de datos de calidad por parte de la Intendencia de Montevideo, junto con estrategias de modelización inteligente, resultan un insumo relevante para la gestión de playas recreativas.

Palabras clave: bosques aleatorios, datos desbalanceados, contaminación, playas recreativas, salud humana.

Abstract: We constructed Machine Learning (ML) models to predict faecal coliforms (FC) for water quality aid management in recreational beaches. Historical data base generated by the Laboratorio de Calidad Ambiental de la Intendencia de Montevideo (IM) was analyzed and ML models were constructed to predict FC excess ($FC > 2.000$). Ten years of monitoring 21 recreational beaches ($N=19359$, November 2009 to September 2019) presented a wide range of salinity and turbidity variability among beaches. FC showed an asymmetric distribution (min=4, median=250, average=1.047 and max=1.280.000) with values exceeding the threshold in all beaches. *In situ* registered, meteorological and oceanographic variables were used to train ML models. A stratified Random forests showed the best performance in the evaluated metrics with an overall accuracy of 86% and 60% of improvement in true positive rates with respect to baseline. High quality data generated by governmental institution together with modelling strategies provided a relevant framework to aid in beach and public health management.

Keywords: random forests, unbalanced data, contamination, recreational beach, human health.

Resumo: Construimos modelos de aprendizaje de máquina (AM) para prever a qualidade da água a fim de auxiliar o gerenciamento em praias recreativas e a balneabilidade. A base de dados históricos gerada pelo Laboratório de Calidad Ambiental de la Intendencia de Montevideo foi analisada e modelos de AM foram construídos para prever o excesso de coliformes fecais ($CF > 2.000$). Dez anos de monitoramento de 21 praias de lazer ($N=19359$, novembro de 2009 a setembro de 2019) apresentaram uma ampla gama de variabilidade de salinidade e turbidez entre as praias. O CF mostrou uma distribuição assimétrica (min=4, mediana=250, média=1,047

e máx=1.280.000) com valores acima do limiar em todas as praias. Registradas *in situ*, variáveis meteorológicas e oceanográficas foram usadas para treinar modelos de AM. Uma floresta aleatória estratificada mostrou o melhor desempenho nas métricas avaliadas, com uma precisão geral de 86% e 60% de melhoria nas taxas positivas verdadeiras em relação à linha de base. Dados de alta qualidade gerados por instituições governamentais, juntamente com estratégias de modelagem, forneceram uma estrutura relevante para auxiliar na gestão de praias e da saúde pública.

Palavras-chave: floresta aleatória, dados não balanceados, contaminação, praia recreativa, saúde humana.

INTRODUCCIÓN

El turismo es una de las actividades con mayor crecimiento en las últimas décadas. En Uruguay, el turismo de sol y playa es un componente importante tanto a nivel social como económico (Gorfinkiel, 2006). Este uso requiere el control de la calidad de las playas para la prevención de riesgos en la salud de los bañistas. La contaminación por aguas cloacales, indicada por concentraciones elevadas de bacterias (coliformes fecales - CF), es un riesgo importante para la salud ambiental y humana (Sabino, et al., 2014). Este es un problema actual en Uruguay (Kruk, et al., 2018) y en la región (Conde, et al., 2002), sobre el cual es necesario desarrollar medidas de gestión eficientes. Una de las grandes limitaciones para anticipar eventos de contaminación es que la estimación de la concentración de CF mediante métodos estándar requiere de 18 a 24 horas de incubación, lo que impide una predicción en tiempo real de la calidad de un ecosistema. La generación de modelos estadísticos que vinculen características del medio de fácil registro con el nivel de contaminación permite anticiparse a estos eventos y, por lo tanto, cuidar la salud de los bañistas (Bedri, et al., 2016).

La contaminación de las playas por aguas de origen cloacal es particularmente frecuente en las zonas costeras (Lotze, et al., 2006; Mara, 2013), especialmente en aquellas sin saneamiento adecuado, donde está directamente asociada al número de habitantes (Kruk, et al., 2019). La contaminación puede llegar a la costa no solo a través de la descarga de efluentes sino también desde el subsuelo contaminado (Giampaoli y Spica, 2014). La recreación en aguas contaminadas incrementa la probabilidad de contraer enfermedades que incluyen una amplia variedad de infecciones gastrointestinales, respiratorias, neurológicas, de la piel, de los oídos, de los ojos y de las heridas (Sabino, et al., 2014). Los niños, las mujeres embarazadas, los ancianos y las personas con el sistema inmunitario debilitado o comprometido son más susceptibles a contraer estas enfermedades (Wade, et al., 2008; Heaney, et al., 2009). Dado que los microorganismos patógenos suelen estar en baja concentración, y su detección y cuantificación resulta laboriosa (Savichtcheva y Okabe, 2006), se monitorea la

calidad bacteriológica a través de la abundancia de microorganismos cultivables indicadores de contaminación fecal (por ejemplo, coliformes, enterococos), cuya concentración se asocia con la probabilidad de desarrollo de una enfermedad específica (Sabino, et al., 2014; WHO, 2018). Los protocolos de monitoreo y valores límites permitidos e incluidos en las normativas varían entre países, e incluso dentro de un mismo país, según el cuerpo de agua monitoreado. En Uruguay, los resultados de coliformes fecales se comparan con el criterio establecido en el Decreto 253/79 (Uruguay, 1979), modificativos posteriores, y la resolución del Ministerio de Vivienda Orden Territorial y Medio Ambiente del 25 de febrero de 2005 (Uruguay, 2005), clasificando a las playas en la Clase 3. De acuerdo con esta clasificación, no se deberá exceder el límite de 2.000 UFC/100 mL en ninguna de al menos 5 muestras, debiendo estar por debajo de 1.000 UFC/100 mL la media geométrica móvil.

La red de saneamiento de Montevideo en la zona Este está conectada al sistema de alcantarillado pluvial que, luego de un tratamiento de las aguas, descarga a 2 km de la costa, a la altura de Punta Brava, en Punta Carretas (Uruguay. Ministerio de Vivienda Ordenamiento Territorial y Medio Ambiente, 2020). Además, en ciertos puntos de la costa hay estructuras destinadas a disminuir el flujo de la red cuando las precipitaciones son intensas, las cuales durante dichos eventos descargan directamente sobre la costa (Figura 1; Intendencia de Montevideo, 2019). En la zona Oeste, se está proyectando una red de saneamiento: actualmente, las descargas de efluentes domésticos se dan a través de la napa o de los cursos de agua locales, que están a su vez influidos por las descargas pluviales.

La Intendencia de Montevideo (IM) es la institución encargada de velar por la calidad de agua de las playas capitalinas. Cuenta con un sistema de monitoreo histórico que comprende 23 playas y varios puntos a lo largo de toda la costa de Montevideo. Las playas son muestreadas con una frecuencia de hasta 4 veces por semana en temporada estival. La sugerencia de cerrar playas por excesos se realiza en base a los indicadores de CF de los días previos, o en forma genérica luego de la ocurrencia de precipitaciones en las 24 o 36 horas previas (Intendencia de Montevideo, 2019). Estas decisiones se basan en información empírica o en la literatura proveniente de otros países, y no existe un modelo estadístico específico que permita aportar predicciones concretas.

Globalmente, se utilizan diversos tipos de algoritmos o criterios para predecir la contaminación mediada por CF en playas. Evitar el baño luego de copiosas lluvias es el criterio más ampliamente utilizado (Bouchalová, et al., 2013; Eregno, et al., 2016; Zhang, et al., 2015), aunque se registran variaciones según el país y entre playas de una misma zona. Por tanto, mejorar la capacidad de predicción con respecto a la clausura por lluvias es relevante para proteger la salud de los

bañistas, así como evitar el cierre innecesario de playas. Los modelos más utilizados son las regresiones lineales múltiples (Thoe y Lee, 2014; Shively, et al., 2016; Searcy, et al., 2018), que asumen un modelo determinado de relación lineal entre los CF y un grupo de variables ambientales de fácil registro. También se han incorporado otras técnicas de modelización como redes neuronales (He y He, 2008; Zhang, et al., 2015), modelos lineales mixtos (Jones, et al., 2013), modelos de combinación de árboles, regresiones lineales parciales y logísticas (Brooks, et al., 2013; 2016). La agencia de protección ambiental de Estados Unidos (EPA, por sus siglas en inglés) provee un programa disponible en línea para estimar los CF mediante regresión múltiple, regresión parcial y combinación de árboles mediante modelos boosting (United States Environmental Protection Agency, 2019). El desafío de predecir estos fenómenos radica en su fuerte comportamiento no lineal que requiere emplear técnicas y métodos con la flexibilidad suficiente (Bourel y Segura, 2018). El desarrollo de algoritmos flexibles y capaces de aprender de los datos ha evolucionado en las últimas décadas junto con el poder de cómputo, componiendo una serie de técnicas denominadas de "Aprendizaje automático" (Machine Learning) que han demostrado una mejor performance en este tipo de problemas (Avila, et al., 2018; Park, et al., 2018; Shively, et al., 2016). Una de las características más sobresalientes de estas aproximaciones es que, en general, se basan en algoritmos y no en un modelo estocástico predeterminado, pueden lidiar con interacciones de alto orden, y están desarrolladas para "aprender" de los datos que se disponen y generar mejores predicciones ante la generación de nuevos datos (Bourel y Segura, 2018; Crisci, et al., 2012). Las redes neurales, los bosques aleatorios (random forest; Breiman, 2001) y las máquinas de vectores soporte (Vapnik, 1998) han sido ampliamente utilizados en problemas de predicción de calidad de agua (He y He, 2008; Park, et al., 2018; Parkhurst, et al., 2005; Zhang, et al., 2015). Estos métodos han sido aplicados con éxito para predecir floraciones de cianobacterias tóxicas en Uruguay (Crisci, et al., 2017; Kruk, et al., 2017, 2015; Martínez de la Escalera, et al., 2017; Segura, et al., 2017) y floraciones marinas (Bourel, et al., 2017), pero no en predicción de coliformes fecales en playas recreativas. Además, el fuerte desbalance (90/10) en la relación entre casos negativos (que no exceden el límite de contaminación establecido en la normativa) y positivos (que exceden la normativa) requiere el uso de nuevas técnicas con las que entrenar los algoritmos.

En el presente trabajo se analizó la serie histórica de monitoreo de playas de Montevideo y se aplicó una variedad de modelos predictivos basados en técnicas de aprendizaje automático para: i) describir las principales tendencias y patrones espacio-temporales en la ocurrencia e incidencia de CF en las distintas playas de Montevideo y ii) desarrollar y evaluar la capacidad predictiva de los modelos de predicción de estos eventos. En base a estos resultados, se pretende aportar al desarrollo de un sistema que transforme la información

generada en un producto fácilmente aplicable a la gestión del baño en playas. Se realiza además una introducción básica a los modelos utilizados y al modo de evaluación de la calidad de éstos.

MATERIALES Y MÉTODOS

Plan de monitoreo de calidad de agua de la IM

La serie de datos analizada proviene del programa de monitoreo de agua de playas y costas del departamento de Montevideo, realizado por la Intendencia de Montevideo (Departamento de Desarrollo Ambiental, Servicio de Evaluación de Calidad y Control Ambiental), con el objetivo de controlar la calidad del agua de las playas y prevenir riesgos en la salud de los bañistas (para acceder a la base de datos, se debe solicitar al Servicio de Evaluación de Calidad y Control Ambiental de la IM). Los datos de las posiciones de los muestreos se encuentran en el Informe anual de evaluación de calidad del agua de playas y costas de Montevideo para el período abril 2016 - marzo 2017. Durante el monitoreo, las principales variables medidas *in situ* fueron: temperatura del agua, salinidad, turbidez, conductividad, oxígeno disuelto (O₂), y presencia de espuma cianobacteriana. Además, se tomaron muestras para su procesamiento en laboratorio, donde se estimó la concentración de clorofila *a*, feopigmentos, microcistinas totales, nitrógeno total, fósforo total, presencia de espuma y concentración de coliformes termotolerantes fecales. La concentración de coliformes fue estimada siguiendo los procedimientos estandarizados de operación del Laboratorio Ambiental de DINAMA (Uruguay. Ministerio de Ambiente, Dirección Nacional de Medio Ambiente, 2017), que se basa en la técnica de filtración por membrana de APHA (American Public Health Association, American Water Works Association y Water Environment Federation, 2012) en el Departamento de Desarrollo Ambiental, Servicio de Evaluación de Calidad y Control Ambiental de la Intendencia de Montevideo.

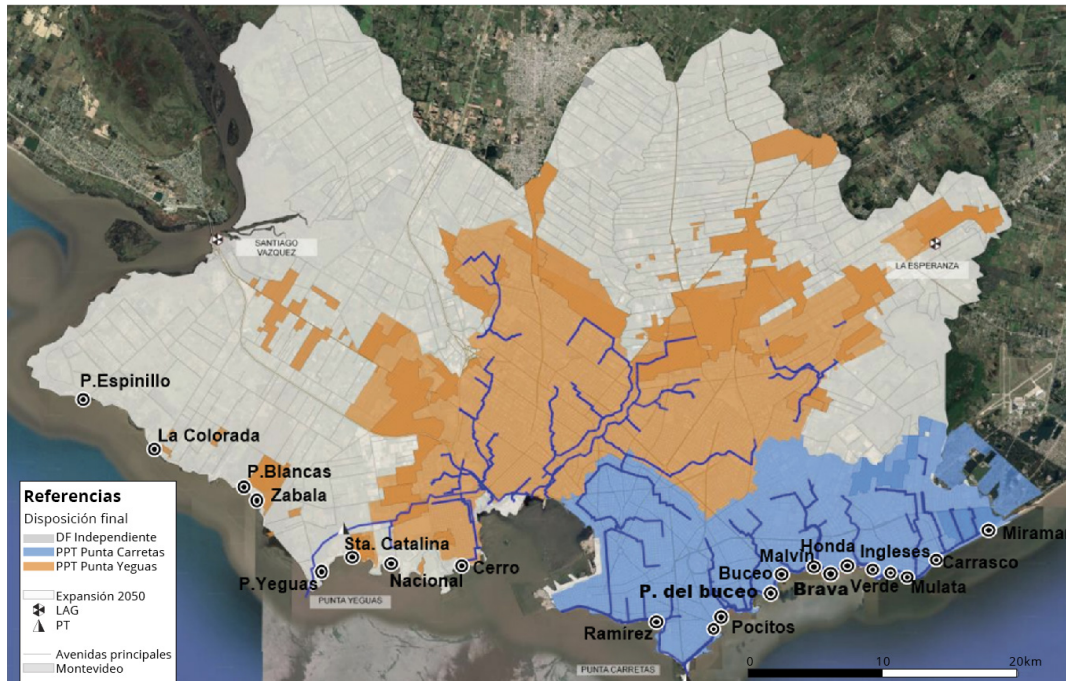


Figura 1. Mapa del departamento de Montevideo donde se muestran los tipos de saneamiento por zonas: Descarga final independiente (gris), Planta de pretratamiento Punta Carretas, tipo unitario (azul), y planta de pretratamiento Punta Yeguas, tipo separativo (naranja). Además, se indica la ubicación de los puntos de muestreo de playas (puntos negros). Modificado de Plan Nacional de Saneamiento (Uruguay. Ministerio de Vivienda Ordenamiento Territorial y Medio Ambiente, 2020).

Exploración de datos *in situ*

Se compilaron e integraron las matrices biológicas y ambientales derivadas de los monitoreos de calidad de agua de la IM y las posiciones espaciales de los sitios de muestreo, y se realizaron controles de calidad para depurar inconsistencias (por ejemplo, datos de temperatura menores a 0 y mayores a 40 °C y de salinidad menores a 0 y mayores a 35). La distribución histórica (Número de datos=19359; desde 2009 a 2019) de las principales variables como temperatura, salinidad, turbidez y CF (logaritmo en base 10), en cada una de las 21 playas monitoreadas de Oeste a Este, fue resumida temporal y espacialmente mediante gráficos descriptivos. La correlación de las variables ambientales entre playas fue evaluada mediante el coeficiente de correlación de Spearman (S).

Bases de datos meteorológicas

Se extrajeron series temporales de valores de variables meteorológicas relevantes de bases de datos digitales del INIA (Instituto Nacional de

Investigación Agropecuaria, s.d.) y de Meteomanz.com (Meteomanz.com, s.d.). Las variables relevadas fueron: precipitación, temperatura del aire, humedad relativa, nubosidad e intensidad y dirección del viento. A partir de la variable precipitación (Pp) se calculó la precipitación acumulada en 1, 2, 3 y hasta 7 días.

Modelización

En este trabajo aplicaremos técnicas de aprendizaje automático supervisadas (Hastie, et al., 2009). Más precisamente se quiere buscar modelar o predecir una variable de salida o de respuesta Y en función de variables de entrada o explicativas X . En este caso, la variable de respuesta Y es una variable binaria que toma dos valores: 0 o 1. Si la playa excede (E), en este caso Y vale 1, o no excede (NE), en este caso Y vale 0, el umbral máximo definido en la normativa. Si el registro en una playa es mayor a 2000 UFC/100 mL, definimos el estado de la playa como E; mientras que si se registra una concentración menor a 2000 UFC/100 mL, definimos el estado de la playa como NE. De esta forma, tenemos un problema que se conoce como de clasificación. En este contexto, entrenaremos un modelo que nos permitirá predecir si una playa excede (E) o no excede la normativa (NE) sin necesidad de obtener el valor de CF en el laboratorio. En el caso de clasificación desbalanceada, es decir, cuando la relación entre casos de cada clase no mantiene una proporción equitativa, es necesario además introducir técnicas de preparación de los datos que disminuyan el desbalance y mejoren la eficiencia de los algoritmos sobre la clase minoritaria, que es en general la que presenta mayor dificultad en predecir adecuadamente. En nuestro contexto, esta clase minoritaria es E y tiene gran relevancia en términos sanitarios. El tratamiento extensivo de estas técnicas está fuera del objetivo de este trabajo, pero quien tenga interés se puede referir a los trabajos de Chawla y otros (2002) y Kuhn y Johnson (2016).

Algoritmos

Para clasificar el estado de las playas se emplearon 9 algoritmos en combinación con un tratamiento específico de los datos para manejar el desbalance. Los algoritmos utilizados fueron: análisis lineal discriminante (LDA), máquinas de vectores soporte (SVM), modelos lineales generalizados (GLM), árboles de clasificación y regresión (CART), y dos métodos de combinación de árboles: Adaboost y bosques aleatorios. Sobre este último consideramos tres variantes: con punto de corte variable (RFcf), RF con muestreo estratificado (RFst) y los RF con pesos diferenciales para las clases (Rfcw). Una introducción conceptual y formal a estos modelos se puede encontrar en Bourel y Segura, 2018. Las variantes de RF han sido desarrolladas por su creador (Breiman, 2001), y programadas y utilizadas en numerosas aplicaciones en ecología y calidad de agua (Cutler, et al., 2007; Kruk, et al., 2015; Segura, et al., 2017). Todos los

modelos y análisis fueron ejecutados en el software estadístico de libre acceso R (R Core Team, 2020).

Además, se consideró también el modelo propuesto por la EPA-EEUU para calidad bacteriológica del agua "Virtual Beach", que está disponible en línea de forma gratuita. Este programa implementa una regresión lineal múltiple y un algoritmo de combinación de árboles (Xgboost) (Cyterski, et al., 2014; Zepp, et al., 2010).

En todos los casos, el conjunto de datos L ingresados al modelo tuvieron cuatro posibles pretratamientos, i) el mismo conjunto L de datos originales, ii) un conjunto donde a L se le agrega réplicas de los datos de la clase minoritaria para balancear artificialmente el conjunto (Upsampling) (Kuhn y Johnson, 2016), iii) un conjunto obtenido a partir de L en el que la clase mayoritaria fue reducida mediante selección aleatoria de casos al número de la clase minoritaria (Downsampling) (Kuhn y Johnson, 2016), y iv) un conjunto donde se le agrega a L, de manera sintética, observaciones de la clase minoritaria (Chawla, et al., 2002).

Estrategia de modelización

Para construir y evaluar los modelos, se utilizó una estrategia de partición de los datos en la cual: i) se entrenaron los algoritmos empleando los datos hasta el mes de septiembre de 2017 (no inclusive) y ii) se evaluó la capacidad predictiva con los datos comprendidos entre septiembre de 2017 y septiembre de 2019. Esta partición permite evaluar predicciones reales e incluye un gran número de casos para el entrenamiento y la evaluación, por lo que no conlleva a sesgos creados por la misma. De esta forma se evalúa sobre datos "frescos" (que no fueron utilizados para entrenar los métodos) la adecuación de los modelos (Crisci, et al., 2012; Hastie, et al., 2009).

Evaluación de modelos

Para definir una playa con riesgo de contaminación fecal se tomó como línea de base el registro de precipitaciones en las 24 horas previas. Definimos que la playa presenta un exceso (E) si en las 24 previas se registraron precipitaciones positivas. Utilizaremos este ejemplo para definir la matriz de confusión, e introducir algunas métricas que serán útiles para evaluar el desempeño predictivo del modelo. La matriz de confusión es una tabla, en cuyas celdas se registran los aciertos y los errores de utilizar un determinado criterio de clasificación respecto a lo observado (Tabla 1). De la intersección de los valores observados y predichos se desprende que puede haber dos tipos de aciertos, predecir E cuando la observación es E, los verdaderos positivos (VP), y predecir

NE cuando la observación es NE, los verdaderos negativos (VN). Estos aciertos se encuentran repertoriados en la diagonal. En un clasificador ideal, todos los casos deberían estar en la diagonal. Entre los dos tipos de errores es fácil identificar que no tienen la misma importancia, pues predecir que no hay exceso cuando sí lo hubo (falso negativo, FN) implica que los bañistas estuvieron expuestos a condiciones con potenciales efectos sobre su salud. Mientras que predecir un exceso cuando no lo hubo (falso positivo, FP) implica que se cierra la playa sin un motivo aparente.

Tabla 1. Matriz de confusión teórica. En las columnas se agregan los valores observados en la realidad y en las filas los valores predichos.

		Observado	
		Excede (E)	No Excede (NE)
Predicción de un modelo	Excede (E)	Verdadero Positivo (VP)	Falso positivo (FP) (Cierra la playa sin motivo)
	No Excede (NE)	Falso Negativo (FN) (Peligro para la salud)	Verdadero Negativo (VN)

Una métrica inmediata que se puede calcular a partir de esta tabla es el porcentaje total de aciertos (Acc):

$$\text{Acc \%} = 100 \frac{\text{VP} + \text{VN}}{\text{VP} + \text{FP} + \text{FN} + \text{VN}} \quad (1)$$

Sin embargo, cuando las clases están desbalanceadas, como en el caso de los excesos de CF, es necesario considerar otras métricas. En efecto, si tenemos una muy baja proporción de datos de la clase minoritaria E, por ejemplo 5%, tener un modelo que acierta en un 95% podría ser un modelo que no acierta en ninguna de las observaciones de la clase minoritaria, a pesar de que su desempeño global es muy bueno pero que expone a los bañistas en ese 5% restante de los casos a riesgos para la salud. Una métrica que pondera los aciertos de la clase minoritaria es la Tasa de Verdaderos Positivos (TVP), que divide los aciertos en la clase E sobre el total de Excesos observados:

$$\text{TVP \%} = 100 \frac{\text{VP}}{\text{VP} + \text{FN}} \quad (2)$$

TVP es una métrica importante de maximizar, junto con la Tasa de Verdaderos Negativos (TVN):

$$\text{TVN \%} = 100 \frac{\text{VN}}{\text{VN} + \text{FP}} \quad (3)$$

El mejor modelo será aquel que maximice ambas métricas, pero -como se verá- en general existe un compromiso entre ambas, y en cierto punto, cuando una aumenta, la otra tiende a disminuir.

A los efectos de ilustrar las métricas consideradas anteriormente analizaremos en la Tabla 2 un caso hipotético.

Tabla 2. Ejemplo hipotético de días, estado de las playas observado (Excede -E- y No Excede -NE-) y predicción de la línea de base (lluvias en las 24hs previas, Pp24) y de un modelo predictivo hipotético que llamaremos AA.

Día	1	2	3	4	5	6	7	8	9	10
Observado	E	NE	NE	NE	NE	NE	NE	E	NE	NE
Predicción Pp24	E	E	NE	NE	NE	NE	E	NE	NE	NE
Predicción del modelo AA	E	NE	NE	NE	NE	NE	NE	E	E	NE

La matriz de confusión resultante de aplicar el criterio de Pp24 es la que aparece en la Tabla 3.

Tabla 3. Matriz de confusión basada en la línea de base (Pp24) de la tabla 2.

		Observado	
		Excede (E)	No Excede (NE)
Predicción	Excede (E)	1	2
	No Excede (NE)	1	6

En este caso, observamos que el clasificador acertó a un exceso y en 6 ocasiones que no hubo exceso, pero se cerró la playa en 2 ocasiones donde no era necesario, mientras que en una ocasión no se predijo un exceso cuando lo hubo. El modelo de AA en el caso del ejemplo superó en todas las métricas a la línea de base (Pp24), y por lo tanto es más adecuado para modelar este problema (Tabla 4). Los modelos fueron generados para todo el conjunto de playas y evaluados de forma global y calculando las métricas (Acc, TVP, TVN) para cada playa.

Tabla 4. Resumen de las métricas de evaluación de la calidad de predicción al aplicar la línea de base (Pp24) y el modelo de inteligencia artificial (AA) para predecir el ejemplo hipotético del estado de la playa.

Modelo	Acc	TVP	TVN
Pp24	0.7	0.5	0.75
Modelo AA	0.9	1	0.875

RESULTADOS

Las variables ambientales periódicamente monitoreadas *in situ* en las 21 playas analizadas desde noviembre de 2009 hasta septiembre de 2019 tuvieron comportamientos estacionales. La temperatura mostró oscilaciones entre 7.1 y 32.2 °C, y un comportamiento estacional similar para todas las playas (Figura 2). La turbidez (rango 0-1000) y la salinidad (rango 0-35) presentaron fluctuaciones estacionales menos periódicas que la temperatura, con una mayor variabilidad entre playas (Figura 2). Fue notoria la diferencia en la salinidad entre las playas del Este y del Oeste (Figura 3).

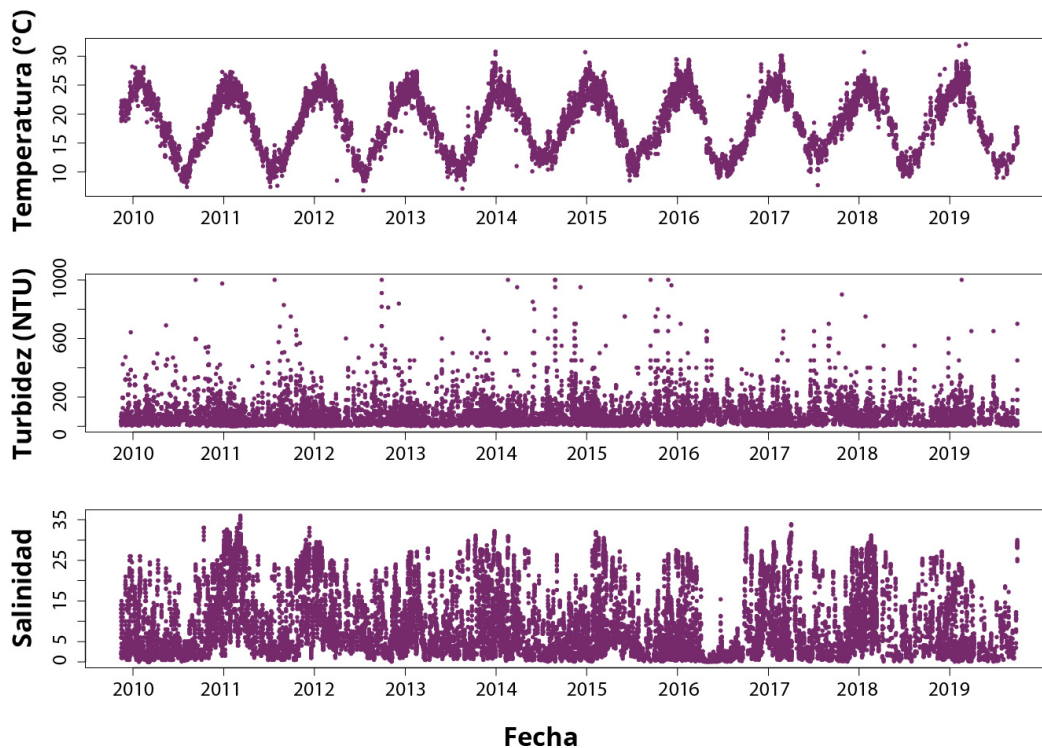


Figura 2. Dinámica temporal de las variables ambientales más relevantes registradas *in situ* en el plan de monitoreo de la IM.

Se cuantificó la abundancia de coliformes fecales (CF; UFC/100 mL) en 19359 casos (Tabla S1). Todas las playas presentaron casos con excesos puntuales (CF > 2000 UFC/100 mL) en el transcurso del monitoreo.

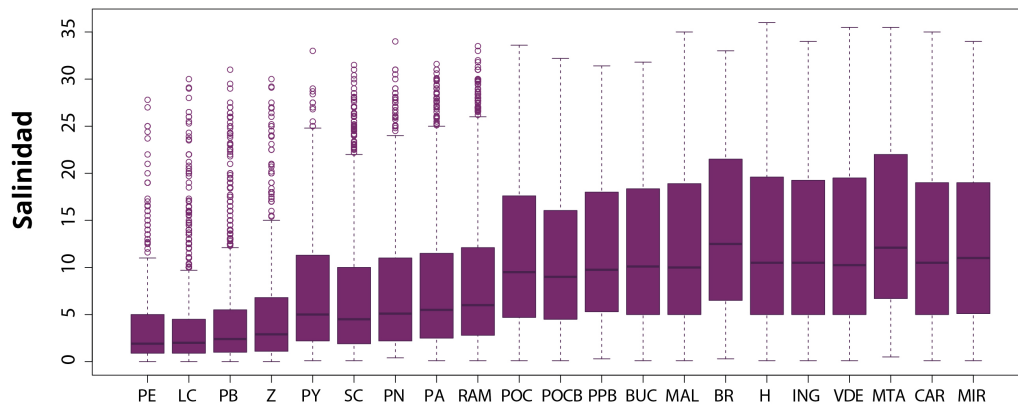


Figura 3. Salinidad por playa ordenadas de Oeste a Este. Nótese el incremento de salinidad de valores menores a 5 en el Oeste a valores de 10 en la zona Este (luego de RAM). La línea negra horizontal en cada caja indica la mediana de la distribución, y las cajas los cuantiles 0.25 (borde inferior) y 0.75 (borde superior) de la distribución.

Los CF tuvieron una distribución asimétrica (mínima=4, mediana=250, media=1.047 y máxima=1.280.000; Figura 4) y la proporción de excesos puntuales (CF > 2.000) global fue cercana a 8% con diferencias entre playas (Figura 4). El número de excesos fue variable entre playas (entre 2 y 275), al igual que la proporción de excesos sobre las muestras realizadas, que se ubicó entre 0.45 y 29% (Tabla S1). Las playas con mayores proporciones de excesos fueron Puerto del Buceo (PPB=29%), Santa Catalina (SC=24%), Playa del Cerro (PA=19%), Ramírez (RAM=16%) y Miramar (MIR=13%). Las playas con menor número de excesos se encontraron en general al Oeste, y fueron Punta Espinillo (PE=0.45%), Playa Nacional (PN=0.9%), Punta Yeguas (PY=1.2%) y La Colorada (LC=1.6%).

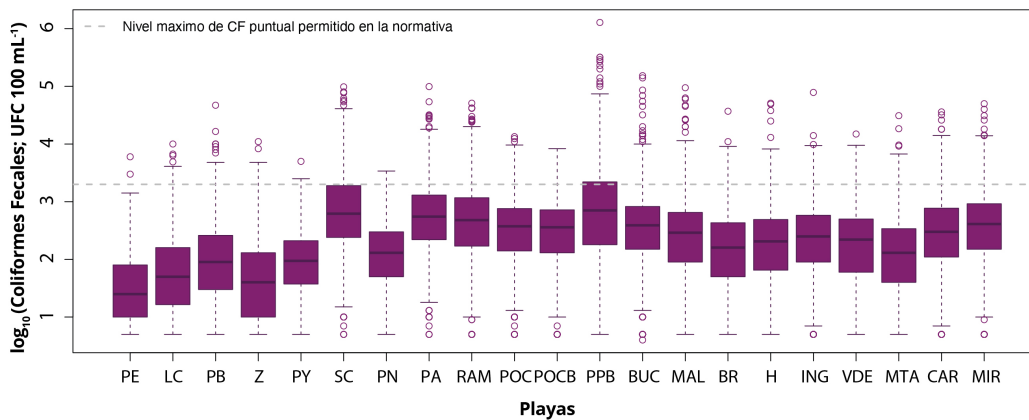


Figura 4. Distribución de los coliformes fecales (logaritmo en base 10) en cada playa en toda la serie de datos de 2009 a 2019. La línea punteada gris representa el valor máximo permitido por la normativa para excesos puntuales en playas recreativas (CF < 2000 UFC/100 mL). La línea negra horizontal en cada caja indica la mediana de la distribución, y las cajas los cuantiles 0.25 (borde inferior) y 0.75 (borde superior) de la distribución. Las playas se encuentran ordenadas de Oeste a Este.

Modelación de la dinámica de coliformes fecales

La línea de base, tomando en cuenta la precipitación en las 24 horas previas como criterio para considerar un exceso, presentó un porcentaje total de aciertos del 82%, una tasa de verdaderos positivos (TVP) del 40% y una tasa de verdaderos negativos (TVN) de 85%.

Los modelos que superaron la performance de la línea de base en todas las métricas fueron el bosque aleatorio estratificado (RFst), la máquina de vectores soporte (SVM) con datos balanceados por SMOTE y upsampling (Tabla 5). La combinación de árboles con Adaboost en los datos balanceados con SMOTE no superó la línea de base en el total de aciertos o en la tasa de verdaderos negativos, pero estuvo muy cercana y con la mejor performance en los verdaderos positivos (Tabla 5). El modelo de bosque aleatorio estratificado mostró un desempeño superior, mejorando a sus competidores directos en todas las métricas y a la línea de base (Tabla 5).

Tabla 5. Métricas para evaluación de los mejores modelos seleccionados para la predicción de excesos puntuales. Los mejores modelos fueron los bosques aleatorios estratificados (RFst) utilizando los datos originales (Originales), las máquinas de vectores soporte (SVM) y Adaboost ambos entrenados sobre una base de datos modificada (Upsampling o SMOTE). Se presenta el porcentaje total de aciertos (Acc%), la tasa de verdaderos positivos (TVP%) y la tasa de verdaderos negativos (TVN%). El tratamiento de los datos y el tipo de modelo

utilizado se compara con la línea de base (Pp24). Se presenta además el caso en el cual ningún dato excede el umbral máximo definido en la normativa (Nunca Excede) y el modelo propuesto por la agencia de protección ambiental de Estados Unidos (EPA).

Datos	Modelo	Acc%	TVP%	TVN%
Originales	RFst	84	63	85
Upsampling	SVM	84	58	86
SMOTE	SVM	84	48	86
SMOTE	Adaboost	83	64	84
Originales	Pp24	82	40	85
Originales	Nunca Excede	93	0	0
Originales	EPA-RMCP	89	3	1

Las variables más importantes obtenidas a partir de RFst, según el criterio de importancia de variables definido para RF, fueron la playa, la concentración de los coliformes en esa playa en el muestreo previo (lag_cf), la salinidad y la turbidez registradas *in situ* y las precipitaciones acumuladas en 24 horas (Figura 5). En las playas con mayor número de excesos (PA, SC, RAM), el modelo presentó una capacidad de predicción de los casos riesgosos para la salud (TVP) de 86, 82 y 76% respectivamente (Figura 6).

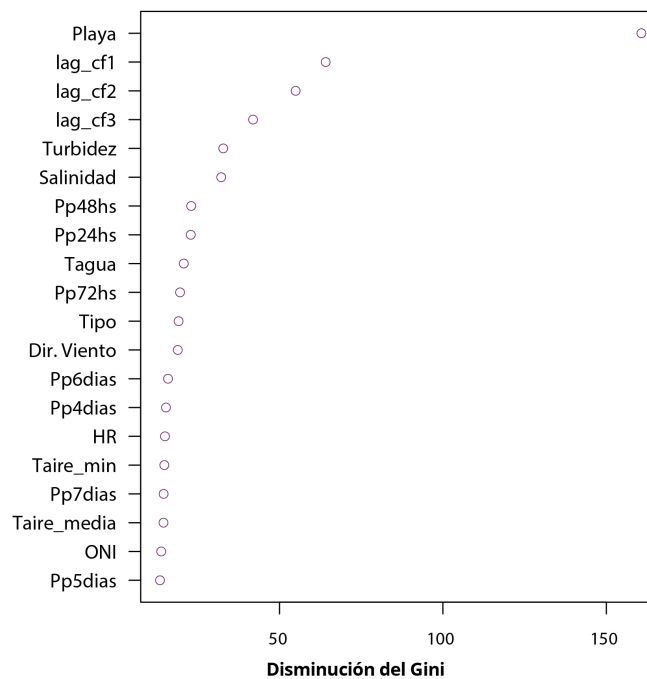


Figura 5. Importancia de las principales variables medida como la disminución en el índice de Gini, ordenadas de forma decreciente para el modelo de bosques aleatorios estratificados (RFst). Por ejemplo, los lag de coliformes fecales (lag_cf) son más relevantes que el Índice Oceánico del Niño (ONI). No se muestran las variables con importancia relativa menor.

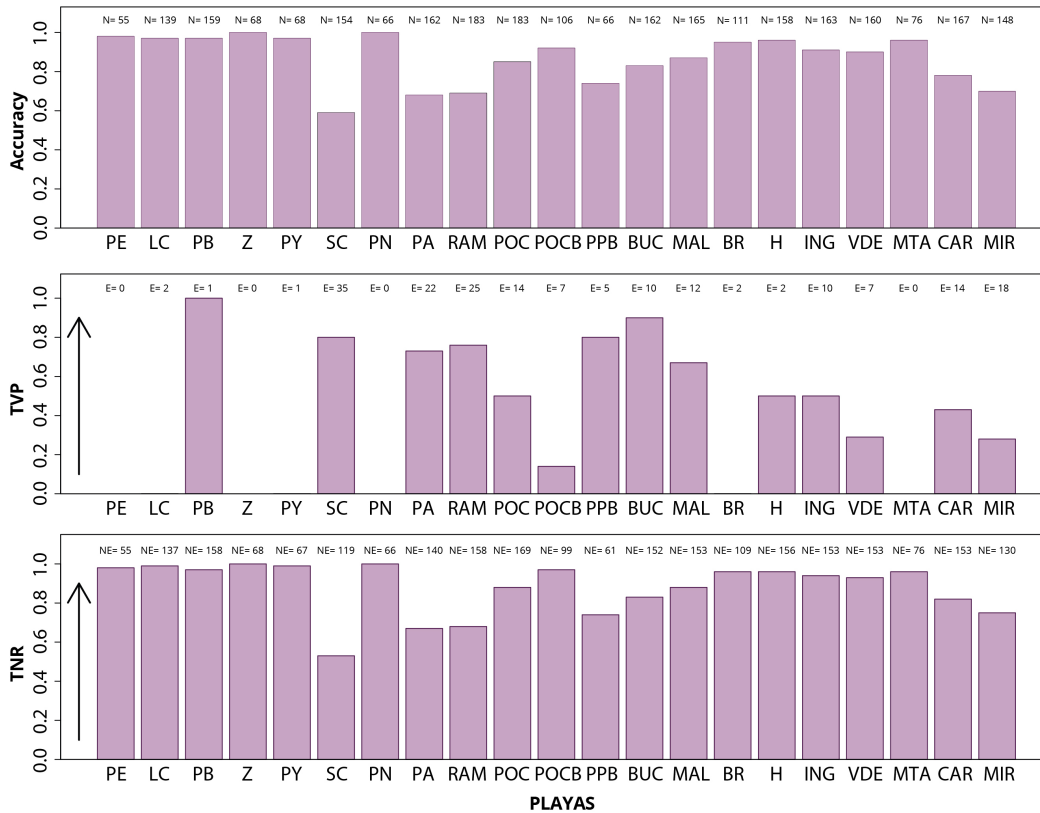


Figura 6. Métricas por playa ordenadas de Oeste a Este para el RFst, que fue considerado el mejor modelo. Se muestra además el número de casos totales (N), el número de excesos por playa (E) y el número de veces que la playa presentó valores dentro de la reglamentación (NE).

DISCUSIÓN

En este trabajo se emplearon técnicas de aprendizaje automático para modelar una extensa serie de datos espacio-temporales con dinámicas complejas. Dichos modelos mejoraron la predicción de eventos donde se excedió la normativa de contaminación fecal en playas recreativas con respecto a la línea de base establecida en función de las precipitaciones en las 24 hs previas. Se describieron además los patrones espacio-temporales de variabilidad ambiental en las playas y el gradiente entre las zonas Oeste y Este de Montevideo. Este tipo de modelización aporta herramientas para minimizar el impacto de la contaminación en la salud humana, advirtiendo a las

autoridades con antelación la probabilidad de exceso de coliformes fecales en el agua. Los modelos utilizados son flexibles, proveen resultados en tiempo real, y tienen la capacidad de “aprender” asimilando nueva información, por lo cual pueden ser incorporados en un esquema de gestión y monitoreo de ciudades inteligentes (Searcy, et al., 2018). Sin embargo, es crucial mejorar el tratamiento de las aguas que se vierten en el estuario del Río de la Plata, pues se registraron valores de concentración de coliformes fecales que exceden la normativa en todas las playas, y en algunos casos con valores órdenes de magnitud por encima de la normativa.

Este trabajo presenta un resumen de la información colectada en más de 10 años de muestreo en las playas de Montevideo. Se evidencian las diferencias en la dinámica entre las Playas al Oeste y al Este de Montevideo, tomando como sitio de corte Punta Brava. Los cambios estacionales en las variables hidrográficas reflejan la dinámica de la zona media del estuario (Calliari, et al., 2005; Simionato, et al., 2010), marcando la transición que genera el desplazamiento estacional del frente de turbidez y de salinidad. Es importante destacar que las mediciones son realizadas en una zona somera (prof = 1m), por lo tanto, existen efectos locales que pueden generar variabilidad entre playas (descargas locales, radiación solar, etcétera). Pero en su conjunto, los datos reflejan la dinámica oceanográfica de esta zona relevante del estuario.

La concentración de coliformes fecales fue variable, pero en todas las playas se registraron valores que excedieron el máximo estipulado en la normativa para valores puntuales (2000 UFC/100 mL). Las playas del extremo Oeste, desde Punta Espinillo hasta Punta Yeguas, presentaron en general valores moderados, con una media inferior al promedio general. Santa Catalina y la Playa del Cerro presentaron mayores valores y un alto porcentaje de excesos. La playa Ramírez se sitúa en una posición geográfica intermedia con características diferenciales de orientación y forma de la playa. Al ser una bahía cerrada, propicia la retención de la contaminación que recibe, y se observaron valores de coliformes fecales elevados y un gran porcentaje de excesos. Las playas de la región Este, desde Pocitos a Miramar, presentaron en promedio valores de coliformes fecales intermedios y similares. La playa del Puerto del Buceo representa una excepción con valores muy superiores de coliformes fecales, está localizada dentro de un recinto portuario deportivo y actualmente permanece como no habilitada para baños.

Los mecanismos que explican esta diferencia geográfica en el comportamiento de las concentraciones de coliformes fecales están dados por los tipos y el esquema de la red de saneamiento metropolitano, la densidad poblacional en las microcuencas, y la circulación dentro y entre playas (Uruguay. Ministerio de Vivienda Ordenamiento Territorial y Medio Ambiente, 2020). Las

playas del Oeste reciben aportes de fuentes de agua contaminadas de forma directa, pues esa región no posee un esquema de saneamiento centralizado, que está actualmente en desarrollo (Uruguay. Ministerio de Vivienda Ordenamiento Territorial y Medio Ambiente, 2020). Las playas del Este reciben en mayor medida el efecto del transporte del agua contaminada proveniente del colector subacuático y los aportes puntuales generados por la descarga de los aliviaderos (Figura 1). Estos hechos generan una predictibilidad diferencial para las playas sujetas a diferentes sistemas de tratamiento. En la zona Oeste, existe mayor variabilidad en los aportes, que son puntuales y dispersos en el territorio y, por lo tanto, más difíciles de predecir. En la zona Este, se observa una menor variabilidad entre playas, tanto en las condiciones ambientales como en la concentración de los coliformes. El efecto de la descarga centralizada de todos los aportes de la cuenca por el colector, luego del tratamiento, promedia los aportes de las microcuencas (barrios) y el transporte de la circulación oceanográfica a mesoescala (decenas de km) costera termina definiendo el efecto en las playas. Esta variabilidad de mecanismos fue bien capturada por el abordaje de modelización utilizado de AA. En este esquema de modelización es posible agregar nuevas variables, como las descargas de los aliviaderos, o la altura de marea, que podrían mejorar la capacidad predictiva, pues son variables que afectan directamente la carga de contaminación que reciben los cuerpos de agua y que no fueron incorporadas en la presente formulación.

La estrategia de modelización presentada permite avanzar en la capacidad predictiva de los excesos de coliformes fecales. Se alcanzó una mejora del 60% en la tasa de verdaderos positivos (TVP) respecto a la línea de base basada en precipitación. En playas con situaciones de contaminación comprometida (PA, SC, RAM), el modelo presentó un incremento importante en la capacidad de predicción de excesos, aumentando la TVP respecto a la línea de base en un factor de 4 en Playa Nacional y Santa Catalina, y un factor de 2 en Ramírez. De los algoritmos evaluados y las estrategias de tratamiento de los datos se desprende que los bosques aleatorios con muestreo estratificado (RFst) fueron los que presentaron mejor desempeño y superaron en todas las métricas los pronósticos generados por la línea de base. Las máquinas de vectores soporte y Adaboost fueron competitivos cuando las clases fueron balanceadas artificialmente mediante upsampling y SMOTE. Esto es esperable debido a que los algoritmos son "voraces" en el sentido de maximizar el porcentaje de aciertos, pero no consideran el acierto en las clases particulares (Breiman, 2001; Segura, et al., 2017). Esta voracidad es problemática cuando la clase minoritaria tiene una importancia superlativa, pues implica no exponer a los bañistas a niveles excesivos de contaminación. El bosque aleatorio con muestreo estratificado aumentó la capacidad de predicción en todas las métricas y en particular en la clase de Excesos, mejorando con respecto a las técnicas de remuestreo (por ejemplo, downsampling) o de incremento artificial de la clase

minoritaria (por ejemplo, SMOTE). Este resultado es relevante, y sugiere que en bases de datos donde la clase mayoritaria presenta casos poco informativos, la estratificación que realiza el algoritmo de RFst permite utilizar toda la información disponible, mientras que los algoritmos de downsampling o SMOTE no permiten capturar toda la riqueza de la información, pues los datos se remuestran con técnicas rígidas, sin aprovechar todo el conjunto de la información. El balanceo de datos para obtener mejores predicciones cuando los problemas son desbalanceados es un área en constante desarrollo y crecimiento (Kuhn y Johnson, 2016).

El algoritmo RFst presenta gran eficiencia computacional, pues el entrenamiento de un RFst con aproximadamente 19000 casos y clase minoritaria de 8% tiene un tiempo de ejecución en un computador portátil con i7 y 16Gb de memoria RAM de unos 5 segundos. La predicción luego de entrenado el modelo con 3000 nuevos casos es instantánea. En este sentido, es más eficiente que los algoritmos que requieren ajustes secuenciales, como Adaboost o aquellos que requieren un tratamiento de datos previo mediante, por ejemplo, upsampling o SMOTE, pues incrementan artificialmente el número de datos y por lo tanto el costo computacional. La eficiencia computacional del RFst permite que el algoritmo sea ejecutado sin la necesidad de grandes plataformas de cómputo. Esto es importante en un contexto de aplicación del modelo (por ej. en otras intendencias departamentales) o en oficinas descentralizadas, pues permite su manejo aún en condiciones de escasa capacidad computacional operativa.

Los algoritmos programados en el software "Virtual Beach" de la agencia de protección ambiental de USA (EPA-EEUU) tuvieron una baja performance, y en el caso del Boosting no fue posible ejecutarlos pues el modelo no convergió dadas las probabilidades iniciales de las clases. Estos resultados resaltan la importancia del desarrollo de algoritmos y modelos que contemplen los aspectos específicos de las bases de datos locales, y que la aplicación de modelos producidos en otros contextos, en general, generan soluciones que no son satisfactorias y no son adecuadas a la realidad local. Esto es aún más crítico en el área ambiental, donde se requiere un equipo de profesionales que posea una visión holística del tema y pueda interactuar entre saberes para desarrollar soluciones integrales. En Uruguay, actualmente existen profesionales altamente capacitados en estas áreas, y a partir de la descentralización de la Universidad de la República en particular, se han desarrollado equipos con una visión interdisciplinaria del ambiente que permiten aventurar un salto en la generación de conocimiento y desarrollo de estas técnicas y sus aplicaciones.

En resumen, el presente trabajo aportó a sintetizar información relevante sobre el estado de las playas. Los algoritmos de aprendizaje automático,

entrenados con una excelente base de datos histórica generada por la IM, permiten predecir eventos de contaminación y representan un insumo importante para la gestión de las playas y el cuidado de la salud humana. Incluir estos algoritmos en un esquema de ciudades inteligentes es inmediato. Asimismo, la inteligencia artificial o el conjunto de métodos para hacer "ciudades inteligentes" requiere que los objetivos de gestión estén integrados y reconozcan el valor de una buena calidad ambiental como forma de disminuir las desigualdades en el acceso a los bienes naturales.

AGRADECIMIENTOS

Al Laboratorio de Calidad Ambiental de la Intendencia de Montevideo, y en especial a Jimena Risso y Bruno D'Alessandro por su apoyo en todo el proyecto y la generación de la base de datos. Se agradece a la ANII por el apoyo al proyecto titulado "Predicción de coliformes fecales en playas capitalinas de interés turístico", (código FSDA_1_2017_1_142164), del Fondo Sectorial de Análisis de Datos de la ANII, cuyos responsables fueron AMS y GP.

REFERENCIAS

- American Public Health Association, American Water Works Association y Water Environment Federation, 2012. *Standard methods for the examination of water and wastewater*. 22a ed. Washington: APHA. Standard Method. 9222 E, Approved 2015.
- Avila, R., Horn, B., Moriarty, E., Hodson, R. y Moltachanova E., 2018. Evaluating statistical model performance in water quality prediction. En: *Journal of Environmental Management*, 206, pp.910–919. DOI: <https://doi.org/10.1016/j.jenvman.2017.11.049>
- Bedri, Z., Corkery, A., O'Sullivan, J.J., Deering, L.A., Demeter, K., Meijer, W.G., O'Hare, G. y Masterson, B., 2016. Evaluating a microbial water quality prediction model for beach management under the revised EU Bathing Water Directive. En: *Journal of Environmental Management*, 167, pp.49–58. DOI: [10.1016/j.jenvman.2015.10.046](https://doi.org/10.1016/j.jenvman.2015.10.046)
- Bouchalová, M., Wennberg, A. y Tryland, I., 2013. Impact of rainfall on bathing water quality—a case study of Fiskevollbukta, Inner Oslofjord, Norway. En: *Vann*, 4, pp.491–498.
- Bourel, M., Crisci, C. y Martínez, A., 2017. Consensus methods based on machine learning techniques for marine phytoplankton presence–absence prediction. En: *Ecological Informatics*, 42, pp.46–54. DOI: [10.1016/j.ecoinf.2017.09.004](https://doi.org/10.1016/j.ecoinf.2017.09.004)

- Bourel, M. y Segura, A.M., 2018. Multiclass classification methods in ecology. En: *Ecological Indicators*, 85, pp.1012–1021. DOI: [10.1016/j.ecolind.2017.11.031](https://doi.org/10.1016/j.ecolind.2017.11.031)
- Breiman, L., 2001. Random forests. En: *Machine Learning*, 45(1), pp.5–32.
- Brooks, W.R., Fienen, M.N. y Corsi, S.R., 2013. Partial least squares for efficient models of fecal indicator bacteria on Great Lakes beaches. En: *Journal of Environmental Management*, 114, pp.470–475. DOI: [10.1016/j.jenvman.2012.09.033](https://doi.org/10.1016/j.jenvman.2012.09.033)
- Brooks, W., Corsi, S., Fienen, M. y Carvin, R., 2016. Predicting recreational water quality advisories: a comparison of statistical methods. En: *Environ. Model. Softw.*, 76, pp.81–94. DOI: <https://doi.org/10.1016/j.envsoft.2015.10.012>
- Calliari, D., Gómez, M. y Gómez, N., 2005. Biomass and composition of the phytoplankton in the Río de la Plata estuary: large scale distribution and relationship with environmental variables during a Spring cruise. En: *Continental Shelf Research*, 25(2), pp.197–210. DOI: [10.1016/j.csr.2004.09.009](https://doi.org/10.1016/j.csr.2004.09.009)
- Chawla, N.V., Bowyer, K.W., Hall, L.O. y Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. En: *Journal of Artificial Intelligence Research*, 16, pp.321–357. DOI: [10.1613/jair.953](https://doi.org/10.1613/jair.953)
- Conde, D., Arocena, R. y Rodríguez-Gallego, L., 2002. Recursos acuáticos superficiales de Uruguay: ambientes algunas problemáticas y desafíos para la gestión. En: *AMBIOS*, III(10), pp.5-9 y IV(11), pp.32-33.
- Crisci, C., Ghattas, B. y Perera, G., 2012. A review of supervised machine learning algorithms and their applications to ecological data. En: *Ecological Modelling*, 240, pp.113–122. DOI: <https://doi.org/10.1016/j.ecolmodel.2012.03.001>
- Crisci, C., Terra R., Pacheco, J.P., Ghattas, B., Bidegain, M., Goyenola, G., Lagomarsino, J.J., Méndez, G. y Mazzeo, M. 2017. Multi-model approach to predict phytoplankton biomass and composition dynamics in a eutrophic shallow lake. En: *Ecological Modelling*, 360, pp.80-93. DOI: <https://doi.org/10.1016/j.ecolmodel.2017.06.017>
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J. y Lawler, J.J., 2007. Random forests for classification in ecology. En: *Ecology*, 88(11), pp.2783–2792. DOI: [10.1890/07-0539.1](https://doi.org/10.1890/07-0539.1)

- Cyterski, M., Brooks, W., Galvin, M., Wolfe, K., Carvin, R., Roddick, T., Fienen, M. and Corsi, S., 2014. *Virtual Beach 3.0.6: user's guide* [En línea]. [s.l.]: USEPA. [Consulta: 9 de junio de 2019]. Disponible en: https://www.epa.gov/sites/default/files/2016-03/documents/vb3_manual_3.0.6.pdf
- Eregno, F.E., Tryland, I., Tjomsland, T., Myrmel, M., Robertson, L. y Heistad, A., 2016. Quantitative microbial risk assessment combined with hydrodynamic modelling to estimate the public health risk associated with bathing after rainfall events. En: *The Science of the Total Environment*, 548–549, pp.270–279. DOI: [10.1016/j.scitotenv.2016.01.034](https://doi.org/10.1016/j.scitotenv.2016.01.034)
- Giampaoli, S. y Spica, V.R., 2014. Health and safety in recreational waters. En: *Bulletin of the World Health Organization*, 92(2), pp.79–79. DOI: [10.2471/BLT.13.126391](https://doi.org/10.2471/BLT.13.126391)
- Gorfinkiel, D., 2006. The economic valuation of coastal areas: the case of Uruguay. En: *Ocean Yearbook*, 20(1), pp.411–434. DOI: <https://doi.org/10.1163/22116001-90000115>
- Hastie, T.J., Tibshirani, R.J. y Friedman, J.H., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Nueva York: Springer. (Springer Series in Statistics).
- Heaney, C.D., Sams, E., Wing, S., Marshall, S., Brenner, K., Dufour, A.P. y Wade, T.J., 2009. Contact with beach sand among beachgoers and risk of illness. En: *American Journal of Epidemiology*, 170(2), pp.164–172. DOI: <https://doi.org/10.1093/aje/kwp152>
- He, L. y He, Z., 2008. Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in Southern California, USA. En: *Water Research*, 42, pp.2563–2573. DOI: [10.1016/j.watres.2008.01.002](https://doi.org/10.1016/j.watres.2008.01.002)
- Instituto Nacional de Investigación Agropecuaria, s.d. *Clima* [En línea]. Montevideo: INIA. [Consulta: 13 de mayo de 2021]. Disponible en: <http://www.inia.uy/gras/Clima/>
- Intendencia de Montevideo, 2019. *Programa de monitoreo de agua de playas y costa del departamento de Montevideo. Informe anual 2018-2019* [En línea]. Montevideo: Intendencia de Montevideo. [Consulta: 12 de abril de 2020]. Disponible en: https://montevideo.gub.uy/sites/default/files/biblioteca/informeannualcalidaddeaguadelacosta-2018-2019_0.pdf

- Jones, R.M., Liu, L. y Dorevitch, S., 2013. Hydrometeorological variables predict fecal indicator bacteria densities in freshwater: data-driven methods for variable selection. En: *Environmental Monitoring and Assessment*, 185(3), pp.2355–2366. DOI: [10.1007/s10661-012-2716-8](https://doi.org/10.1007/s10661-012-2716-8)
- Kruk, C., Dobroyan, M., Segura, A.M., Balado, I., Trabal, N., Piccini, C., Sampognaro, L., De Leon, F., Rodríguez, A., y Verrastro, N. 2019. Calidad de agua y su percepción en playas: La Paloma, Rocha [En línea]. En: AUGM. // *Congreso de Agua, Ambiente y Energía*. Montevideo, Uruguay (25-27 de setiembre de 2019). Montevideo: Uruguay. [Consulta: 13 de mayo de 2021]. Disponible en: https://www.fing.edu.uy/imfia/congresos/caae/assets/trabajos/37_Calidad_de_agua_y_su_percepci%C3%B3n_en_playas_La_Paloma_Rocha.pdf
- Kruk, C., Dobroyan, M., González, L., Segura, A.M., Balado, I., Trabal, N., De León, F., Martínez, G., Rodríguez, A., Piccini, C., Chalar, G. y Verrastro, N., 2018. Calidad de agua y salud ecosistémica en playas recreativas de la Paloma, Rocha [En línea]. En: *Revista Trama*, 9(9), pp.1-10. [Consulta: 13 de mayo de 2021]. Disponible en: <http://www.auas.org.uy/trama/index.php/Trama/article/view/179>
- Kruk, C., Piccini, C., Segura, A., Nogueira, L., Carballo, C., Martínez de la Escalera, G., Calliari, D., Ferrari, G., Simoens, M., Cea, J., Alcántara, I., Vico, P. y Miguez, D., 2015. Herramientas para el monitoreo y sistema de alerta de floraciones de cianobacterias nocivas: Río Uruguay y Río de la Plata. En: *INNOTEC*, (10), pp.23–39. DOI: <https://doi.org/10.26461/10.02>
- Kruk, C., Segura, A.M., Nogueira, L., Alcántara, I., Calliari, D., Martínez de la Escalera, G., Carballo, C., Cabrera, C., Sarthou, F., Scavone, P. y Piccini, C., 2017. A multilevel trait-based approach to the ecological performance of *Microcystis aeruginosa* complex from headwaters to the ocean. En: *Harmful Algae*, 70, pp.23–36. DOI: [10.1016/j.hal.2017.10.004](https://doi.org/10.1016/j.hal.2017.10.004)
- Kuhn, M. y Johnson, K., 2016. *Applied predictive modeling*. 5ta. imp. cor. Nueva York: Springer.
- Lotze, H.K., Lenihan, H.S., Bourque, B.J., Bradbury, R.H., Cooke, R.G., Kay, M.C., Kidwell, S.M., Kirby, M.X., Peterson, C.H. y Jackson, J.B.C., 2006. Depletion, degradation, and recovery potential of estuaries and coastal seas. En: *Science*, 312, pp.1806-1809. DOI: <https://doi.org/10.1126/science.1128035>
- Mara, D. 2013. *Domestic wastewater treatment in developing countries* [En línea]: Londres: Earthscan. [Consulta: 13 de mayo de 2021]. Disponible en: https://www.researchgate.net/publication/287291244_Domestic_Wastewater_Treatment_in_Developing_Countries#fullTextFileContent

- Martínez de la Escalera, G., Kruk, C., Segura, A.M., Nogueira, L., Alcántara, I. y Piccini, C., 2017. Dynamics of toxic genotypes of *Microcystis aeruginosa* complex (MAC) through a wide freshwater to marine environmental gradient. En: *Harmful Algae*, 62, pp.73–83. DOI: [10.1016/j.hal.2016.11.012](https://doi.org/10.1016/j.hal.2016.11.012)
- Meteomanz.com, s.d. *Meteomanz.com* [En línea]. [s.l.]: [s.n.]. [Consulta: 13 de mayo de 2021]. Disponible en: <http://meteomanz.com/>
- Park, Y., Kim, M., Pachepsky, Y., Choi, S.H., Cho J.G., Jeon, J. y Cho, K.H., 2018. Development of a nowcasting system using machine learning approaches to predict fecal contamination levels at recreational beaches in Korea. En: *Journal of Environment Quality*, 47(5), pp.1094-1102. DOI: [10.2134/jeq2017.11.0425](https://doi.org/10.2134/jeq2017.11.0425)
- Parkhurst, D.F., Brenner, K.P., Dufour, A.P. y Wymer, L.J., 2005. Indicator bacteria at five swimming beaches—analysis using random forests. En: *Water Research*, 39(7), pp.1354–1360. <https://doi.org/10.1016/j.watres.2005.01.001>
- R Core Team, 2020. R: A language and environment for statistical computing [En línea]. Viena: R Foundation for Statistical Computing. [Consulta: 30 de marzo de 2021]. Disponible en: <http://www.r-project.org/index.html>
- 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. y Brandão, J., 2014. Routine screening of harmful microorganisms in beach sands: implications to public health. En: *Science of The Total Environment*, 472, pp.1062–1069. DOI: [10.1016/j.scitotenv.2013.11.091](https://doi.org/10.1016/j.scitotenv.2013.11.091)
- Savichtcheva, O. y Okabe, S., 2006. Alternative indicators of fecal pollution: relations with pathogens and conventional indicators, current methodologies for direct pathogen monitoring and future application perspectives. En: *Water Research*, 40(13), pp.2463–2476. DOI: [10.1016/j.watres.2006.04.040](https://doi.org/10.1016/j.watres.2006.04.040)
- Searcy, R.T., Taggart, M., Gold, M. y Boehm, A.B., 2018. Implementation of an automated beach water quality nowcast system at ten California oceanic beaches. En: *Journal of Environmental Management*, 223, pp.633–643. DOI: [10.1016/j.jenvman.2018.06.058](https://doi.org/10.1016/j.jenvman.2018.06.058)
- Segura, A.M., Piccini, C., Nogueira, L., Alcántara, I., Calliari, D. y Kruk, C., 2017. Increased sampled volume improves *Microcystis aeruginosa* complex (MAC) colonies detection and prediction using Random Forests. En: *Ecological Indicators*, 79, pp.347–354. DOI: [10.1016/j.ecolind.2017.04.047](https://doi.org/10.1016/j.ecolind.2017.04.047)

- Shively, D.A., Nevers, M.B., Breitenbach, C., Phanikumar, M.S., Przybyla-Kelly, K., Spoljaric, A.M. y Whitman, R.L., 2016. Prototypic automated continuous recreational water quality monitoring of nine Chicago beaches. En: *Journal of Environmental Management*, 166, pp.285–293. DOI: [10.1016/j.jenvman.2015.10.011](https://doi.org/10.1016/j.jenvman.2015.10.011)
- Simionato, C.G., Clara Tejedor, M.L., Campetella, C., Guerrero, R. y Moreira, D., 2010. Patterns of sea surface temperature variability on seasonal to sub-annual scales at and offshore the Río de la Plata estuary. En: *Continental Shelf Research*, 30(19), pp.1983–1997. DOI: [10.1016/j.csr.2010.09.012](https://doi.org/10.1016/j.csr.2010.09.012)
- Thoe, W. y Lee, J.H.W., 2014. Daily forecasting of Hong Kong beach water quality by multiple linear regression models. En: *Journal of Environmental Engineering*, 140(2). DOI: [10.1061/\(ASCE\)EE.1943-7870.0000800](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000800)
- United States Environmental Protection Agency, 2019. *Virtual beach* [En línea]. [s.l.]. USEPA. [Consulta: 28 de junio de 2019]. Disponible en: <https://www.epa.gov/ceam/virtual-beach-vb>
- Uruguay. Decreto 253/979, de 09 de mayo de 2009. *Diario Oficial*, 31 de mayo de 1979, p.1473.
- Uruguay. Ministerio de Ambiente, Dirección Nacional de Medio Ambiente, 2017. Técnica de filtración por membrana 5053UY. En: Uruguay. Ministerio de Ambiente, Dirección Nacional de Medio Ambiente. *Manual de procedimientos analíticos para muestras ambientales* [En línea]. Montevideo: DINAMA. [Consulta: 12 de marzo de 2021]. Disponible en: <https://www.gub.uy/ministerio-ambiente/politicas-y-gestion/manual-procedimientos-analiticos-para-muestras-ambientales-tercera-edicion-2017>.
- Uruguay. Ministerio de Vivienda Ordenamiento Territorial y Medio Ambiente, 2020. *Plan nacional de saneamiento* [En línea]. Montevideo: MVOTMA. [Consulta: 30 de mayo de 2020]. Disponible en: <https://www.gub.uy/ministerio-ambiente/politicas-y-gestion/planes/plan-nacional-saneamiento>
- Uruguay. Resolución S/N del 25 de febrero de 2005. *Diario Oficial*, 2 de marzo de 2005, p.543.
- Vapnik, V., 1998. *Statistical learning theory*. Nueva York: John Wiley and Sons, Inc.

- Wade, T.J., Calderon, R.L., Brenner, K.P., Sams, E., Beach, M., Haugland, R. y Dufour, A.P., 2008. High sensitivity of children to swimming-associated gastrointestinal illness: results using a rapid assay of recreational water quality. En: *Epidemiology*, 19(3), pp.375-383. DOI: [10.1097/EDE.0b013e318169cc87](https://doi.org/10.1097/EDE.0b013e318169cc87)
- WHO, 2018. *WHO recommendations on scientific, analytical and epidemiological developments relevant to the parameters for bathing water quality in the Bathing Water Directive (2006/7/EC)*. [s.n.]: WHO.
- Zepp, R.G., Cyterski, M., Parmar, R., Wolfe, K., White, E.M. y Molina, M., 2010. *Predictive modeling at beaches. Volume II: predictive tools for beach notification*. Washington: USEPA.
- Zhang, Z., Deng, Z. y Rusch, K.A., 2015. Modeling fecal coliform bacteria levels at gulf coast beaches. En: *Water Quality, Exposure and Health*, 7(3), pp.255–263. DOI: <https://doi.org/10.1007/s12403-014-0145-3>

ANEXO

Tabla S1. Resumen de las principales características de las 21 playas capitalinas analizadas ordenadas de Oeste a Este. CF refiere a densidad de coliformes fecales (UFC/100 mL) y a los estadísticos mínimo (Mín), máximo (Máx), la mediana (Med) y el promedio aritmético (Prom). El número de casos (N) y el número de excesos (E; cuando CF>2000 UFC/100 mL). Las variables ambientales registradas en el agua temperatura (T), turbidez (Turb) y salinidad (S) junto con los estadísticos descriptivos (Ds= desvío estándar).

Playa*	CF Mín	CF Máx	CF Med	CF Prom	N	E	T Mín	T Máx	T Prom	T Ds	Turb. Mín	Turb. Máx	Turb. Prom	Turb. Ds	S Mín	S Máx.	S Prom.	S Ds
PE	5	6000	25	101	433	2	8.8	29.2	21.4	4.1	2.8	310.0	37.4	32.0	0.0	27.8	3.8	4.9
LC	5	10000	50	198	1059	18	8.8	29.2	19.5	4.9	2.2	650.0	43.3	47.2	0.0	30.0	3.6	4.6
PB	5	47000	90	339	1119	24	6.8	29.2	19.0	5.0	1.1	811.0	45.9	60.3	0.0	31.0	4.4	5.3
Z	5	11000	40	210	570	12	10.3	29.2	22.8	2.5	2.1	450.0	34.8	34.7	0.0	30.0	5.2	5.9
PY	5	5000	94	204	404	5	8.9	28.9	21.6	3.9	2.6	370.0	36.2	40.2	0.1	33.0	7.8	7.2
SC	5	98000	620	2317	1135	275	8.8	29.2	19.6	4.9	1.8	1000.0	44.0	61.8	0.1	31.5	7.1	7.2
PN	5	3400	130	260	645	6	8.8	29.2	21.4	4.2	2.0	303.0	34.5	34.3	0.4	34.0	7.9	7.5
PA	5	99000	550	1550	1169	224	7.1	30.3	19.7	5.0	1.7	1000.0	46.7	59.6	0.1	31.6	8.2	7.5
RAM	5	51000	480	1432	1179	190	8.4	32.1	19.8	5.1	1.6	800.0	46.0	65.8	0.1	33.5	8.8	7.9
POC	5	13400	375	715	1168	86	8.5	29.3	19.5	4.8	1.1	1000.0	86.8	89.2	0.1	33.6	11.5	8.1
POCB	5	8300	360	618	1003	63	8.8	29.2	19.4	4.9	2.8	684.5	86.7	85.7	0.1	32.2	10.9	7.9
PPB	5	1E+06	705	8390	550	159	10.3	29.2	22.8	2.6	2.8	600.0	81.0	75.8	0.3	31.4	11.7	7.8
BUC	4	2E+05	390	1347	1103	105	8.8	29.2	19.6	4.9	1.6	900.0	93.7	95.2	0.1	31.8	11.9	8.2
MAL	5	94550	290	990	1149	77	8.9	29.6	19.8	4.8	1.9	1000.0	98.0	117.6	0.1	35.0	12.1	8.4
BR	5	37000	160	431	633	15	10.3	29.2	22.9	2.5	2.8	600.0	77.7	71.8	0.3	33.0	14.1	8.6
H	5	51000	205	546	1082	27	8.8	29.2	19.7	4.9	2.6	750.0	84.4	96.7	0.1	36.0	12.4	8.5
ING	5	78400	250	601	1127	52	8.8	29.2	19.6	4.9	2.1	1000.0	91.8	118.7	0.1	34.0	12.3	8.4
VDE	5	14900	220	482	1086	47	8.8	29.2	19.6	4.9	2.8	684.5	85.5	85.0	0.1	35.5	12.3	8.5
MTA	5	31000	130	432	545	13	10.3	29.2	22.9	2.5	2.8	1000.0	79.9	82.0	0.5	35.5	14.1	8.7
CAR	5	36000	300	866	1137	102	9.1	30.1	20.0	4.8	2.3	1000.0	110.1	131.9	0.1	35.0	12.3	8.3
MIR	5	50000	410	1099	1063	137	8.8	29.2	19.6	4.9	2.8	600.0	86.1	83.5	0.1	34.0	12.4	8.3

*Nombres de las playas en orden de Oeste a Este (de arriba a abajo en la tabla): Punta Espinillo, La Colorada, Pajas Blancas, Zabala, Punta Yeguas, Santa Catalina, Nacional, Cerro, Ramírez, Pocitos en Barreiro, Pocitos en Av. Brasil, Puerto del Buceo, Buceo, Malvín, Brava, Honda, Ingleses, Verde, Mulata, Carrasco, Miramar. (ver informes con descripción detallada en el Programa de monitoreo de agua de playas y costa del departamento de Montevideo. Informe anual 2018-2019 (Intendencia de Montevideo, 2019).