

Predicción de floraciones de Cianobacterias en el Río Uruguay: modelos de clasificación con datos des-balanceados

Ignacio Alcántara^{1,2}; Claudia Piccini¹; Angel Segura³; Madeleine Renom⁴; Carla Kruk⁵

1. Departamento de Microbiología, Instituto de Investigaciones Biológicas Clemente Estable, MEC. nalcann@gmail.com
2. Departamento de Bioestadística, Facultad de Veterinaria, Universidad de la República.
3. Maren
4. Departamento de Ciencias de la Atmósfera, Instituto de Física, Facultad de Ciencias, Universidad de la República.
5. Limnología-IECA, Facultad de Ciencias y Ecología Funcional de Sistemas Acuáticos, CURE-Rocha, Universidad de la República.

Las floraciones de cianobacterias (FC) ocurren con mayor frecuencia durante periodos estivales, en sistemas con altas concentraciones de nutrientes y columna de agua estable. Las variables meteorológicas e hidrológicas son fundamentales ya que operan en conjunto sobre la estructura química y térmica del agua derivando en efectos sobre las FC. En el Río Uruguay las FC son frecuentes, por esta razón la Comisión Administradora del Río Uruguay monitorea 37 playas y emite comunicados sobre la aptitud de baño. Los comunicados clasifican al riesgo de exposición en tres niveles de peligrosidad, siendo el nivel rojo el que implica riesgo para la salud. El objetivo de este trabajo es construir un modelo de Random Forest (RF) para predecir la probabilidad de ocurrencia de estos niveles. Para ello utilizamos información de variables hidrológicas, meteorológicas, limnológicas y de la distribución y frecuencia de FC en el RU (periodo 2011–2015). Los niveles no se distribuyeron de manera equitativa en la base de datos. El porcentaje de cada nivel sobre el total fue: verde=84%, amarillo=6% y rojo=10%. Dado que los RF son afectados por un desabalance en el número de casos por categoría se implementó una aproximación basada en muestreo para construir modelos con distintas proporciones entre la clase mayoritaria y la clase de interés. El desempeño del clasificador se evaluó mediante matrices de confusión generadas por muestra de prueba. Se seleccionó el modelo cuya proporción de rojos fue doble respecto al resto. Las métricas derivadas de la matriz de confusión fueron: sensibilidad= 80%, especificidad= 84% y precisión= 35%. Como perspectivas se plantea evaluar el desempeño de RF modificando los límites de decisión y ponderar a la clase minoritaria. Generar modelos de predicción de FC permite anticipar cambios que puedan propiciar crecimiento y acumulación de estos organismos y seleccionar medidas de gestión/mitigación adecuadas.