

Predicción de calidad de agua para diversos usos mediante técnicas de Machine Learning



JOSEPE

C. Crisci , A. Segura, E. Rocca, MEDIA - CURE



FACULTAD DE INGENIERÍA



UNIVERSIDAD DE LA REPÚBLICA URUGUAY



SARAS² South American Institute for Resilience and Sustainability Studies



UNIDAD DE GESTIÓN DESCONCENTRADA MALDONADO



UBA Universidad de Buenos Aires



IFEVA

Predicción de calidad de agua para diversos usos mediante técnicas de Machine Learning



Crisci

JOSEPE

C. Crisci , A. Segura, E. Rocca, MEDIA - CURE



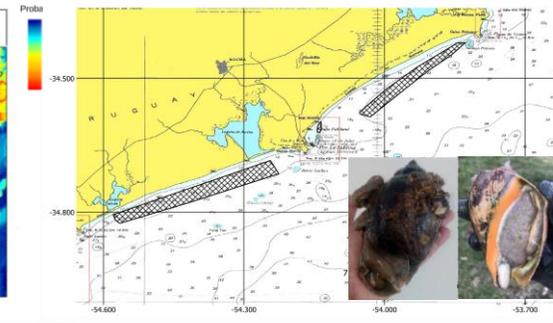
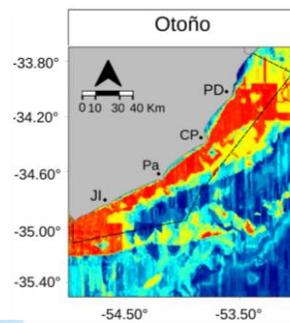
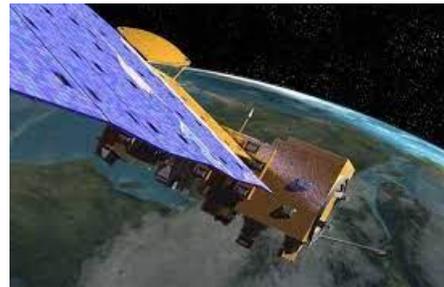
AGENCIA NACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN



ESPACIO
INTERDISCIPLINARIO
UDELAR

Algunas líneas de trabajo del Departamento de Modelización Estadística de Datos e Inteligencia Artificial (MEDIA)

- Aplicación de modelos predictivos mediante técnicas de ML
- Desarrollo de nuevas herramientas de ML
- Estadística de eventos extremos





Presentaremos:

- Predicción de calidad de agua para diversos usos mediante técnicas de Machine Learning (ML)
- Nueva algoritmo de clasificación para mejorar la predicción de clases raras: soto Forest

Algunos trabajos del MEDIA en estas líneas



Ecological Indicators
Volume 85, February 2018, Pages 1012-1021



Ecological Modelling
Volume 360, 24 September 2017, Pages 80-93



Water Research
Volume 202, 1 September 2021, 117450



Ecological Informatics
Volume 46, July 2018, Pages 97-102



Multiclass classification methods in ecology

M. Bourel ^{a,b,c} ✉, A.M. Segura ^c ✉



Ecological Informatics
Volume 42, November 2017, Pages 46-54



Multi-model approach to predict phytoplankton biomass and composition dynamics in a eutrophic shallow lake governed by extreme meteorological events

Carolina Crisci ^a ✉, Rafael Terra ^b ✉, Juan Pablo Pacheco ^c ✉, Badih Ghattas ^a ✉, Mario Bidegain ^e ✉, Guillermo Goyenola ^c ✉, Juan José Lagomarsino ^f ✉, Gustavo Méndez ^f ✉, Néstor Mazzeo ^c ✉

Machine learning methods for imbalanced data set for prediction of faecal contamination in beach waters

Mathias Bourel ^{a,b} ✉, Angel M. Segura ^b, Carolina Crisci ^b, Guzmán López ^b, Lia Sampognaro ^b, Victoria Vidal ^b, Carla Kruk ^{b,c,d}, Claudia Piccini ^{b,c}, Gonzalo Perera ^b

The first artificial intelligence algorithm for identification of bat species in Uruguay

G. Botto Nuñez ^{a,b,c,d} ✉, G. Lemus ^b, M. Muñoz Wolf ^a, A.L. Rodales ^b, E.M. González ^b, C. Crisci ^d

Consensus methods based on machine learning techniques for marine phytoplankton presence-absence prediction

M. Bourel ^{a,b} ✉, C. Crisci ^c ✉, A. Martínez ^d

Environmental Science and Policy 154 (2024) 103699

Freshwater Biology

ORIGINAL ARTICLE

Classification of Reynolds phytoplankton functional groups using individual traits and machine learning techniques

Carla Kruk ✉, Melina Devercelli, Vera L. M. Huszar, Esnedi Hernández, Guadalupe Beamud, Mónica Diaz Lúcia H. S. Silva, Angel M. Segura



ISSN 1688-6593 INNOTEC 2021, No. 22 (e555)
REVISTA DEL LABORATORIO TECNOLÓGICO DEL URUGUAY

<https://doi.org/10.26461/22.07>

Global Change Biology

RESEARCH ARTICLE

Rise of toxic cyanobacterial blooms is promoted by agricultural intensification in the basin of a large subtropical river of South America

Carla Kruk ✉, Angel Segura, Gervasio Piñeiro, Pablo Baldassini, Laura Pérez-Becoña, Felipe García-Rodríguez, Gonzalo Perera, Claudia Piccini



Contents lists available at ScienceDirect

Environmental Science and Policy

journal homepage: www.elsevier.com/locate/envsci



Inter and transdisciplinarity strategies for evaluating and improving water quality monitoring systems: Uruguay as a study case

Néstor Mazzeo ^{a,b}, Ana Lía Ciganda ^{a,c}, Camila Fernández Nion ^d, Francisco J. Peñas ^e, Alexia María González-Ferreras ^e, Carolina Crisci ^f, Cristina Zurbriggen ^{b,g}, Daniel Pérez ^{b,h}, José Barquin ^e, Ismael Díaz ^{d,*}

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



Data from: Cyanobacterial blooms in subtropical riverine and estuarine ecosystems of South America

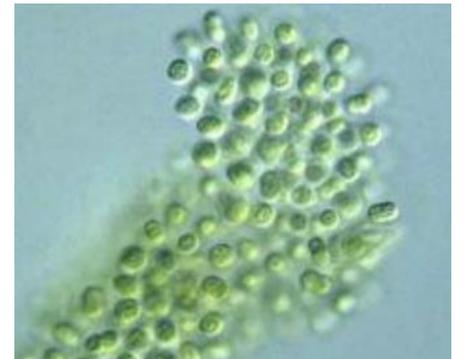
Sampognaro, Lia ¹ ✉; Segura, Angel M. ¹ ✉; Piccini, Claudia ² ✉; Kruk, Carla ¹ ✉

Author affiliations ▼

Published Jul 21, 2024 on Dryad. <https://doi.org/10.5061/dryad.9w0vt4bpb>

Predicción de calidad de agua mediante técnicas de ML

- ¿Para qué predecimos?
- ¿Qué predecimos?
- ¿Qué datos consideramos? (gran trabajo de recopilación de datos)
- Ajuste y resultados de los modelos
- Cómo acoplamos los resultados de los modelos a ámbitos de la gestión

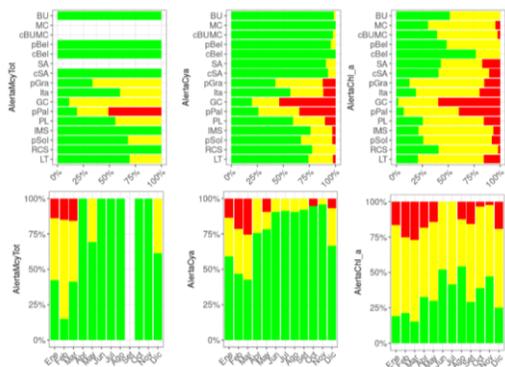


Predicción de calidad de agua mediante técnicas de ML

- **Para qué predecimos:**
- Anticiparnos a condiciones no deseadas de calidad de agua destinada a usos recreativos para prevenir a la población a través de las insituciones competentes (mejorar el escenario de “predicciones para mañana”)
- Brindar a plantas potabilizadores insumos para la toma de decisiones en tiempo real
- Aportar a sistemas de alerta temprana

Predicción de calidad de agua mediante técnicas de ML

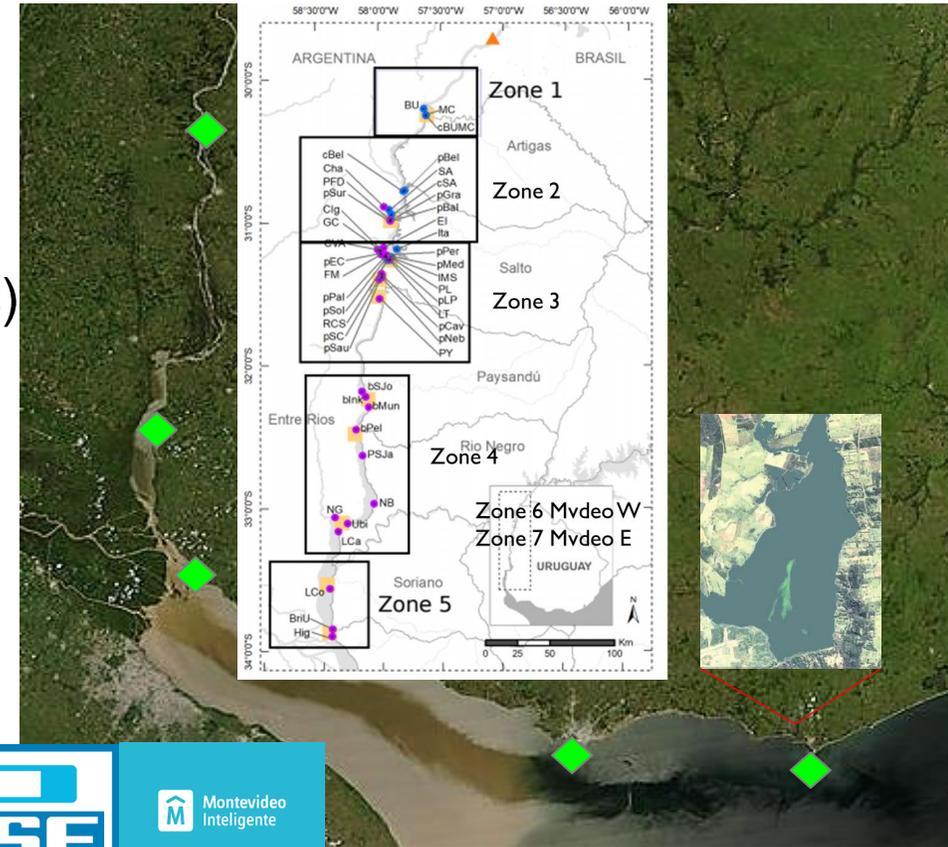
- **Qué predecimos:** variables de respuesta de los modelos
- Variables que representan algún tipo de afectación de la calidad del agua ya sea para consumo humano o recreación
- Clorofila-a (proxy de biomasa algal, toda la comunidad del fitoplancton –ug/L-)
- Cianobacterias (en cel/ml o biovolumen -mm³/L-)
- Microcistina (ug/L)
- Presencia/ausencia de cierta especie o grupo funcional
- Inspección visual



Nivel de Vigilancia	Nivel de Alerta 1	Nivel de Alerta 2
<p><10 Clorofila a (↔g/L) o</p> <p><5.000 Cianobacterias (cel/ml) o</p> <p>Ausencia de cúmulo* o</p> <p><2 Microcistina (↔g/L)</p>	<p>10-50 Clorofila a (↔g/L) o</p> <p>5.000-50.000 Cianobacterias (cel/ml) o Ausencia de cúmulo* o</p> <p>2-10 Microcistina (↔g/L)</p>	<p>>50 Clorofila a (↔g/L) o</p> <p>>50.000 Cianobacterias (cel/ml) o</p> <p>Presencia de cúmulo² o</p> <p>>10 Microcistina (↔g/L)</p>

Predicción de calidad de agua mediante técnicas de ML

- **Qué datos consideramos**
- 80 sitios RU, RdelP, Laguna del Sauce (altísima heterogeneidad)
- 25866 casos
- Más de 80 variables
 - ✓ Verdad acuática (proyectos, instituciones varias)
 - ✓ Meteorológicas, hidrológicas, climatológicas (instituciones y servidores varios)
 - ✓ Satelitales (MODIS GEE, intensidad bandas, índices)

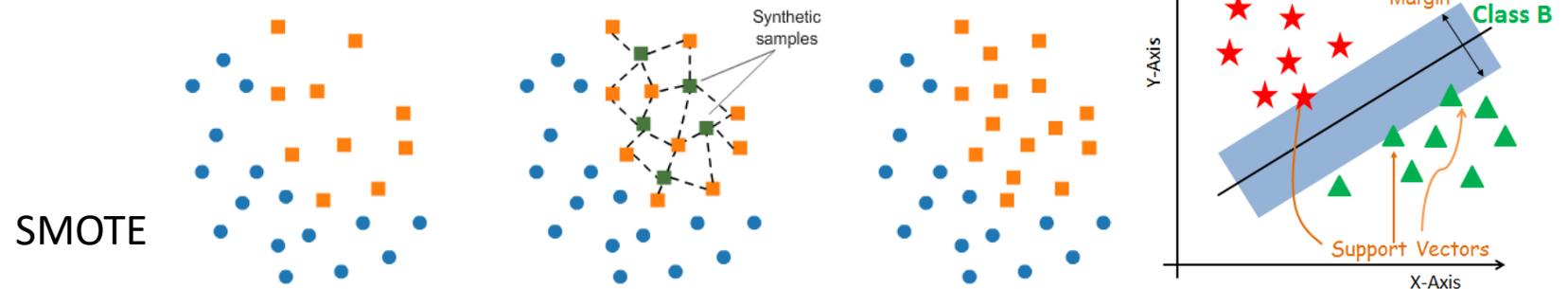
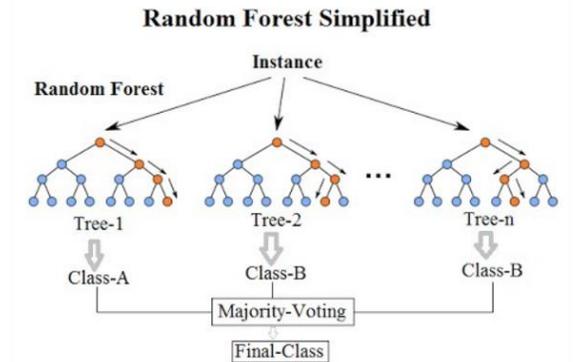


Predicción de calidad de agua mediante técnicas de ML

- Ajuste y algunos resultados de los modelos

MODELOS

- Random Forests (Breiman, 2001) + SMOTE en clasificación (Chawla et al., 2002)
- Selección de variables con VSURF
- SVM (Cortes & Vapnik, 1995) + SMOTE
- soto Forest (Segura et al., in prep)
- Desafíos: Datos desbalanceados/Datos faltantes



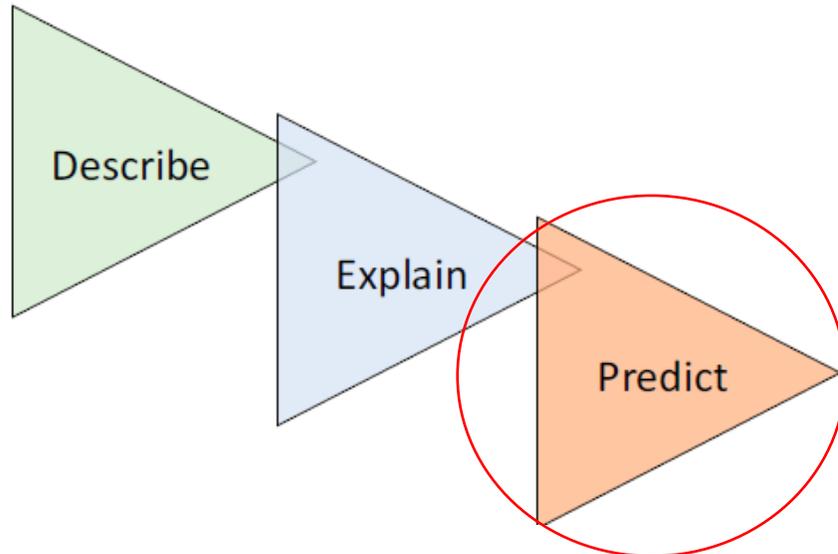
Predicción de calidad de agua mediante técnicas de ML

- Ajuste y algunos resultados de los modelos

MODELOS DATOS IN SITU

MODELOS DATOS
SATELITALES

MODELOS DATOS IN SITU +
SATELITALES

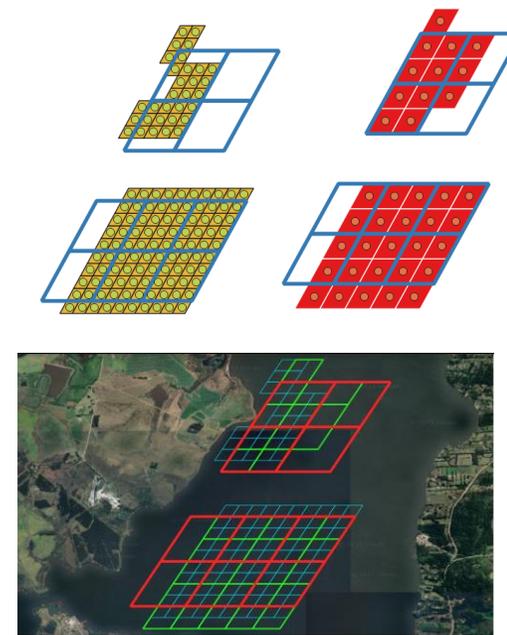
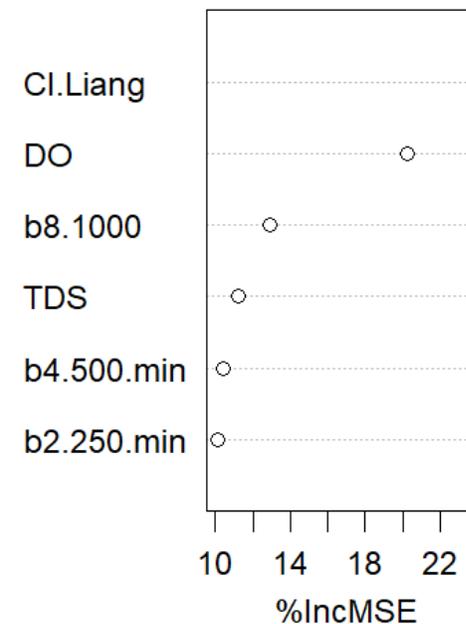


Interés de transferencia de resultados
al ámbito de la gestión

Predicción de calidad de agua mediante técnicas de ML

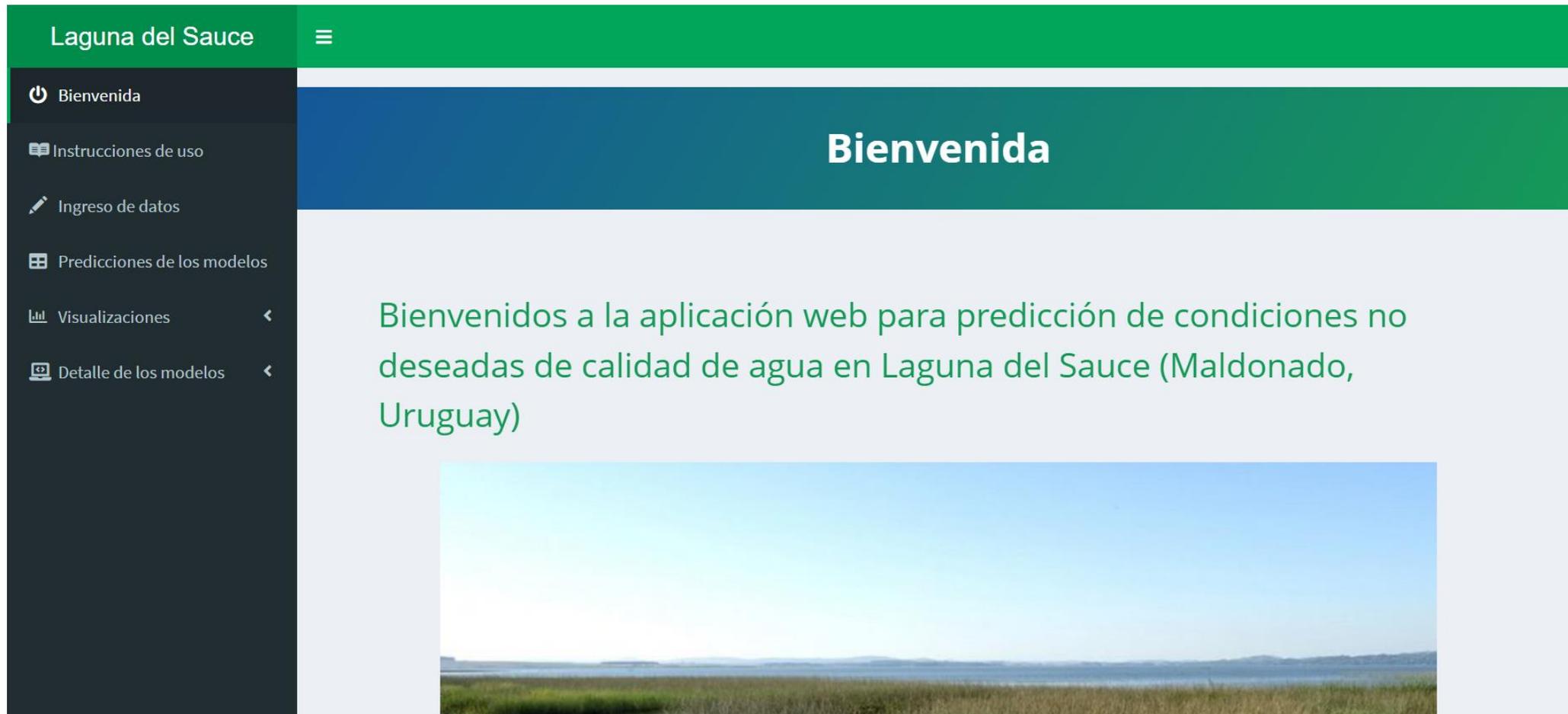
- **Resultados de modelo que integra datos in situ y satelitales (L. del Sauce)**
- RF, VSURF para selección de variables (Robin et al., 2015)
- Datos in situ + satelitales (reescalamiento espacial para considerar píxeles de diferente tamaño)
- Variable de respuesta: Biovolumen de cianobacterias

- ✓ **Var explicada: 77% (OOB)**
- ✓ **Pseudo varianza explicada sobre muestra test: 58%**
- ✓ La varianza explicada aumenta respecto al modelo que solo considera variables in situ, importancia de índices



Predicción de calidad de agua mediante técnicas de ML

- Cómo acoplamos los resultados de los modelos a ámbitos de la gestión



The screenshot shows a web application interface. On the left is a dark sidebar with a menu containing: 'Bienvenida', 'Instrucciones de uso', 'Ingreso de datos', 'Predicciones de los modelos', 'Visualizaciones', and 'Detalle de los modelos'. The main content area has a green header with 'Laguna del Sauce' and a hamburger menu icon. Below the header is a blue banner with the word 'Bienvenida' in white. The main body contains a green text message: 'Bienvenidos a la aplicación web para predicción de condiciones no deseadas de calidad de agua en Laguna del Sauce (Maldonado, Uruguay)'. Below the text is a landscape photograph of a wetland area with water and grass under a clear sky.



Predicción de calidad de agua mediante técnicas de ML

- Cómo acoplamos los resultados de los modelos a ámbitos de la gestión

Laguna del Sauce

☰







🔌 Bienvenida

📖 Instrucciones de uso

✍ Ingreso de datos

🏠 Predicciones de los modelos

📊 Visualizaciones <

➤ Series de tiempo

➤ Histograma

➤ Gráficas bivariadas

🔍 Detalle de los modelos <

Mediciones del día

Fecha

Fecha:

Turbidez (NTU)

Color (UPC)

Nitratos (ppm NO₃)

Amonio (ppm NH₄)

Conductividad (uS.cm)

Nivel (m)

Clorofila (mg/m³)

pH

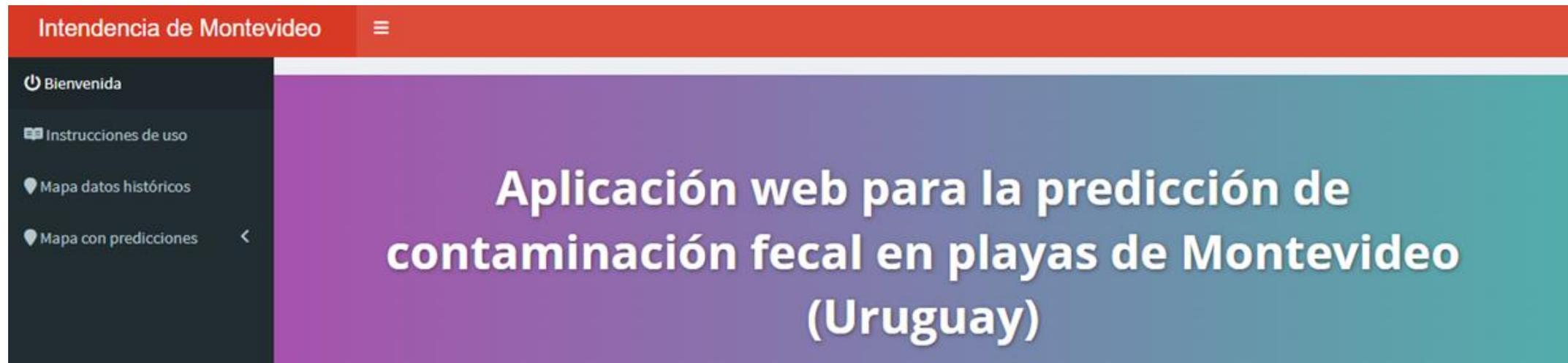
Oxidabilidad (ppm O₂)

Alcalinidad (p

	Accuracy test	C1 test	C2 test	C3 test
Chl-a	0.82	0.89	0.37	0.89
Biovol cianos	0.78	0.85	0.5	0.86
	Accuracy test	TPR test	TNR test	
GIII	0.84	0.81	0.85	
GVII	0.87	0.78	0.88	
<i>M. aeruginosa</i>	0.93	0.92	0.94	

Predicción de calidad de agua mediante técnicas de ML

- **Cómo acoplamos los resultados de los modelos a ámbitos de la gestión**
- Otro ejemplo: sistema de alerta temprana para coliformes fecales en playas de Montevideo

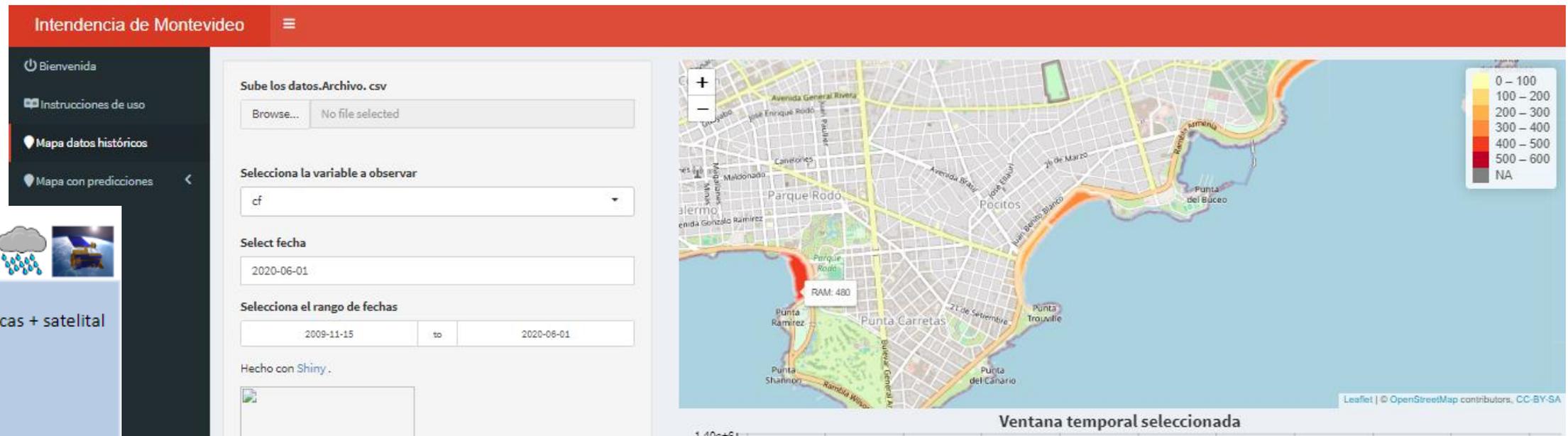


Tesis de Maestría en Ciencias Ambientales de Victoria Vidal



Predicción de calidad de agua mediante técnicas de ML

- **Cómo acoplamos los resultados de los modelos a ámbitos de la gestion**
- Otro ejemplo: sistema de alerta temprana para coliformes fecales en playas de Montevideo



Modelo 2
Meteorológicas + satelital

Lags CF

Meteo= Precipitaciones
y velocidad del viento

Satelital = SST y Kd490



Nueva herramienta de clasificación para mejorar la predicción de clases raras: soto Forest

Problema, similar al que presentamos recién, pero la variable de respuesta en este caso es categórica donde una de las clases se encuentra mucho menos representada que la otra (e.g. presencia floración).



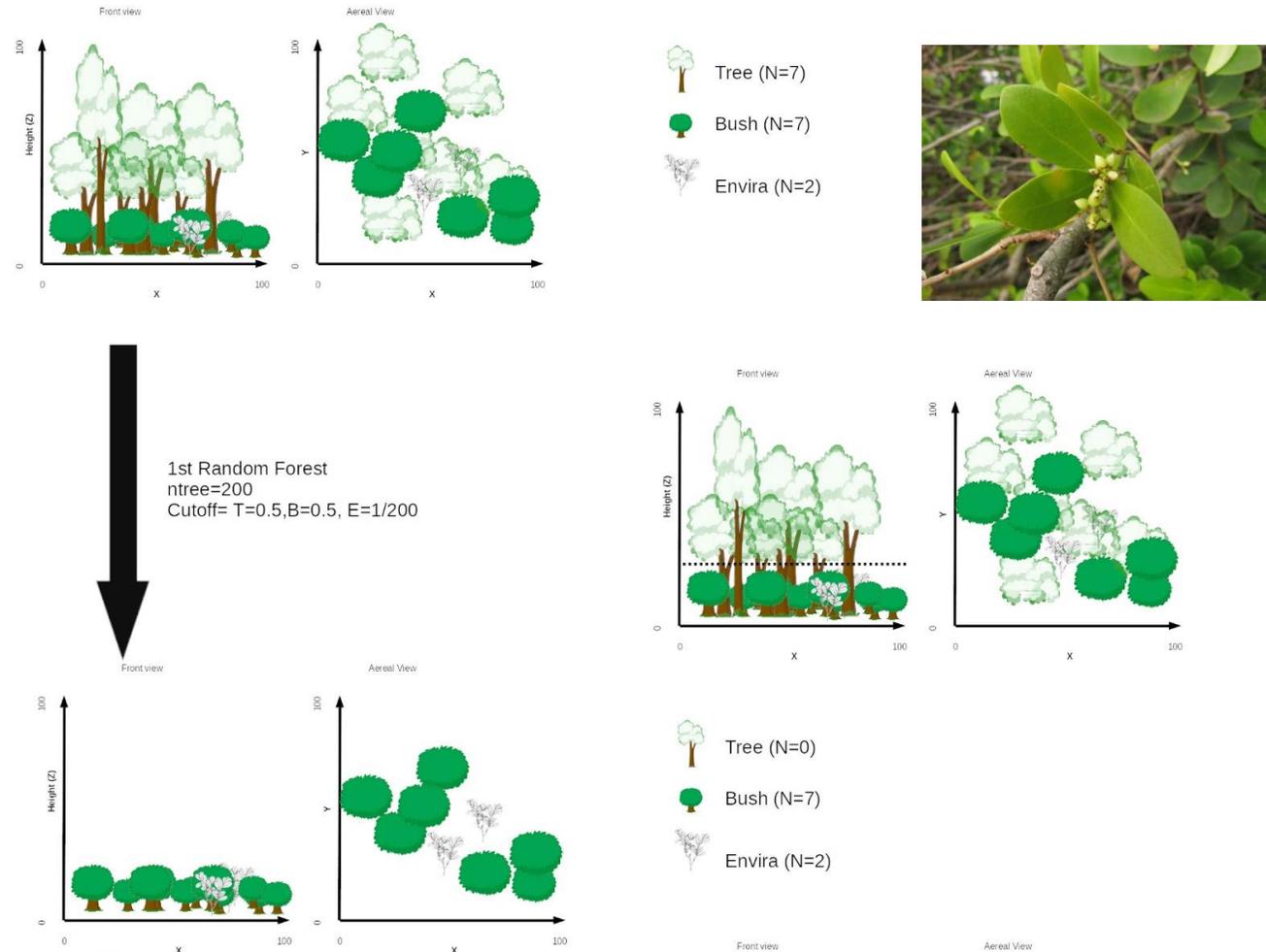
Nueva herramienta de clasificación para mejorar la predicción de clases raras: soto Forest

- ✓ Se realiza una secuencia de M bosques con diferentes cutoff para las clases desde un mínimo de cutoff- green= $1/2$, red= $2/ntree$
- ✓ Si hay en el bosque 1 árbol que clasifica como la clase de interés- en este caso floración- red, entonces el bosque predice floración
- ✓ El cutoff cambia linealmente en cada paso hasta 0.5-0.5 en el último paso (M) que es el RF por defecto (cutoff= $1/k$)
- ✓ El único hiperparámetro que se agrega es Numero de Capas del Bosque (M), que es el número de random forests secuenciales que se realizan y se define a priori

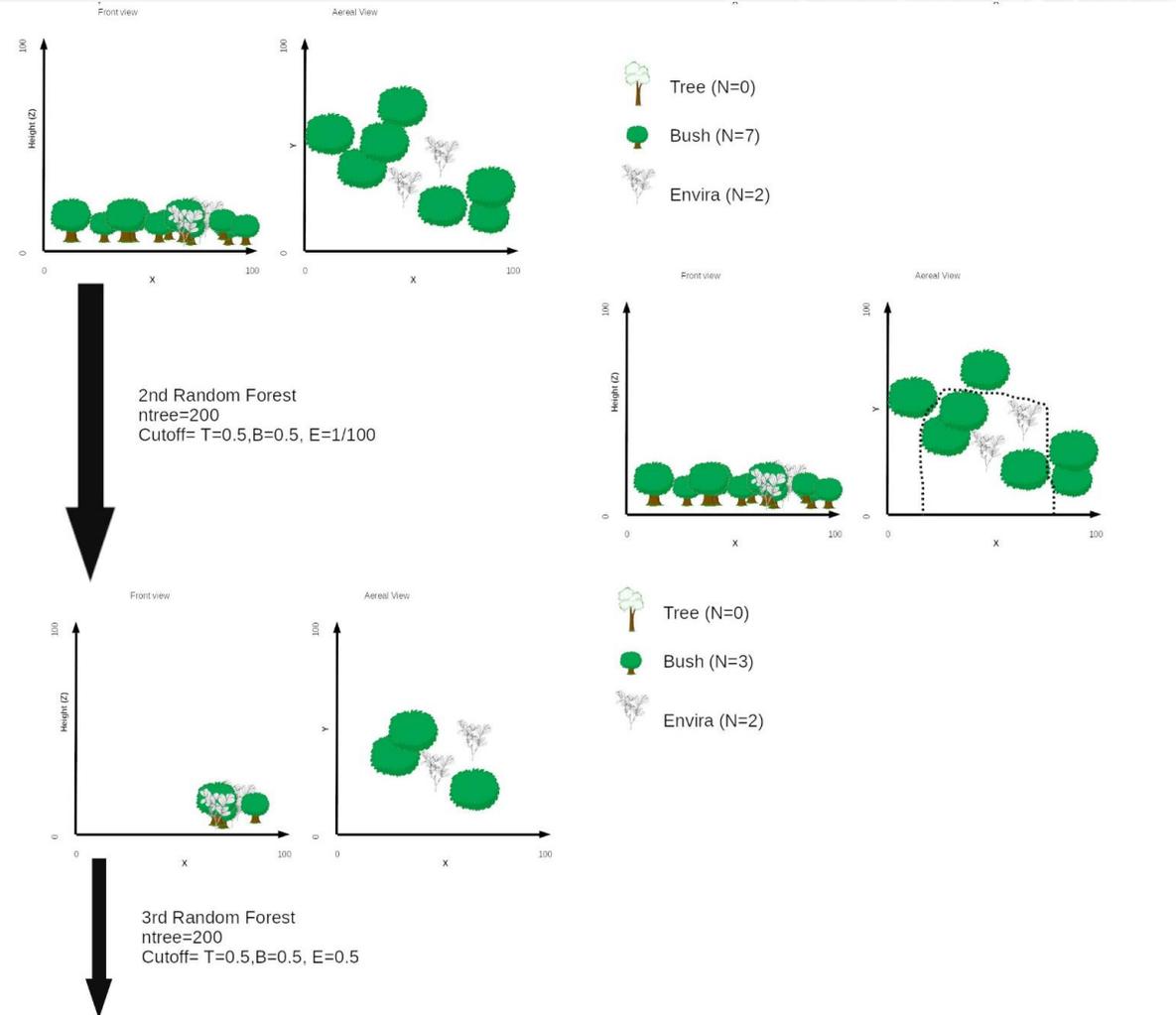
Undergrowth forest: going deep into the forest to improve prediction of rare classes

Segura A.M., Crisci, C., Bourel, M. Sampognaro, L., Kruk, C & Perera G.

Nueva herramienta de clasificación para mejorar la predicción de clases raras: soto Forest



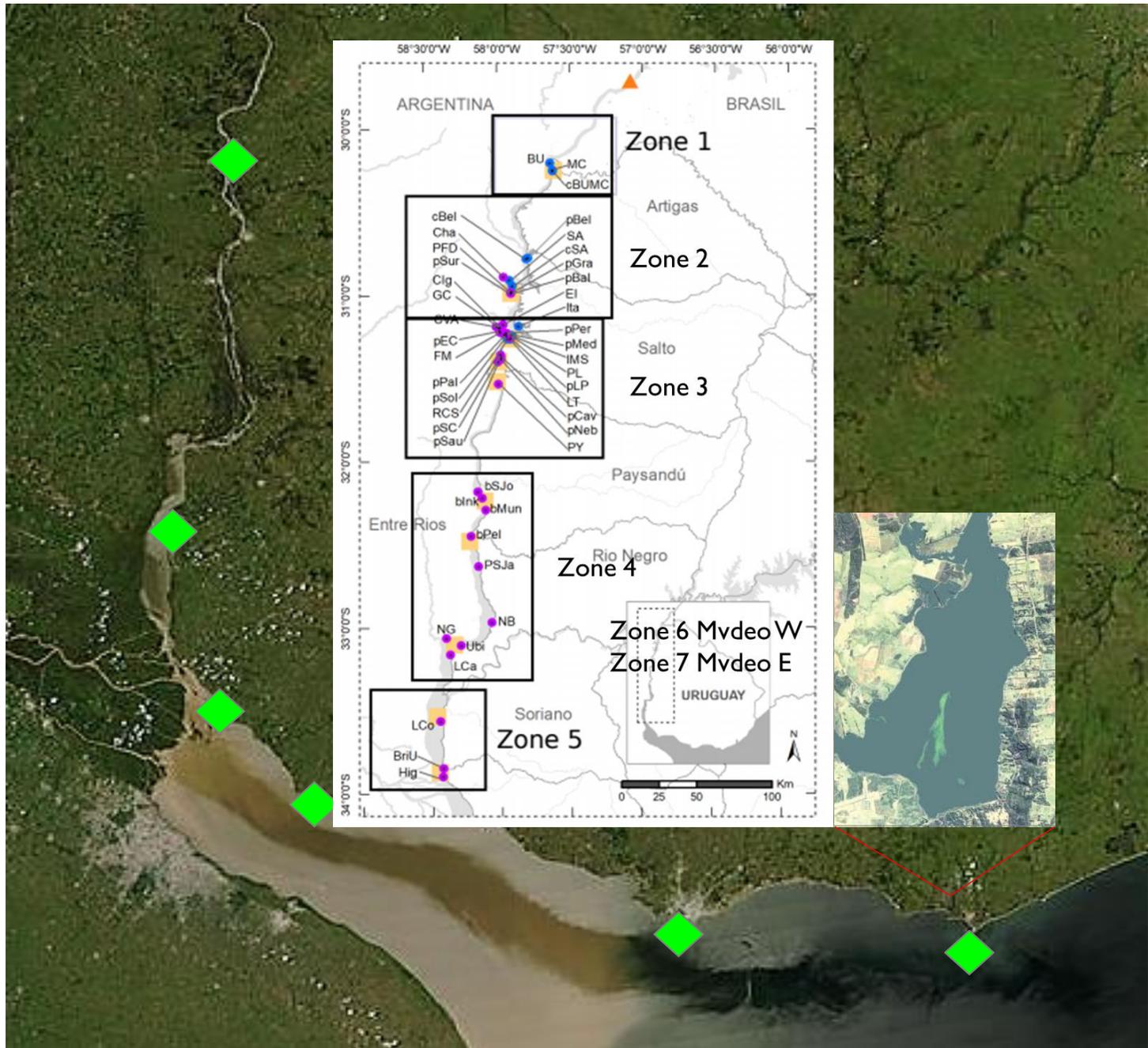
Nueva herramienta de clasificación para mejorar la predicción de clases raras: soto Forest



Nueva herramienta de clasificación para mejorar la predicción de clases raras: soto Forest

Primeros resultados

	Accuracy	Kappa	sensitivity- TPR	1-specificity- FPR	
	sotoRF	0,872	0,517	0,595	0,077
	RF	0,882	0,472	0,436	0,036
	RFstrat	0,876	0,497	0,535	0,063
	SVM	0,866	0,317	0,26	0,021
	sotoRFsatSite	0,846	0,419	0,515	0,093
	RFsatSite	0,863	0,389	0,382	0,048
	RFstratsatSite	0,853	0,427	0,491	0,078
	SVMsatSite	0,855	0,245	0,205	0,026
	sotoRFSatOnly	0,849	0,271	0,254	0,041



¡Gracias!



<https://media.cure.edu.uy/>