

# **VII JORNADAS DE ESTADÍSTICA APLICADA**

Libro de Resúmenes  
La Paloma-Rocha 2023

# VII JORNADAS DE ESTADÍSTICA APLICADA

20 Y 21 DE OCTUBRE DE 2023  
LA PALOMA, ROCHA



UNIVERSIDAD  
DE LA REPÚBLICA  
URUGUAY

# Contenidos

<b>HOMENAJE A GONZALO PEREZ IRIBARREN</b>	<b>4</b>
Una manera de recordar a Gonzalo Pérez Iribarren. . . . .	4
<b>CONFERENCIAS PLENARIAS</b>	<b>11</b>
Optimización en procesos estocásticos. Ernesto Mordecki . . . . .	11
Comer y pescar: ecología trófica de lobas y leonas marinas y su interacción con pesquerías. Valentina Franco-Trecu . . . . .	11
Medidas de calidad para fusión de imágenes. Ana Georgina Flesia . . . . .	12
Deep learning for semantic segmentation; some medical applications. Badih Ghattas . . . . .	12
<b>PONENCIAS</b>	<b>13</b>
Regresión en Variedades. Alejandro Cholaquidis <sup>1*</sup> . . . . .	13
¿Qué es la inferencia causal? ¿Cómo podemos incorporarla al desarrollo de modelos de aprendizaje supervisado? Gabriel Illanes <sup>1</sup> . . . . .	14
Tres resultados para implementar la teoría asintótica extremal en el caso de muestras grandes de datos fuertemente dependientes. Gonzalo Perera <sup>1</sup> , Carolina Crisci <sup>1</sup> , Lía Sampognaro <sup>1</sup> . . . . .	15
Estimación univariada de la densidad adaptando el método root-unroot con una función núcleo. Mathias Bourel <sup>1</sup> y Jairo Cugliari <sup>2</sup> . . . . .	16
Métodos algebraicos para el estudio de tiempos y localizaciones de salida en difusiones estocásticas. Didier Henrion <sup>1</sup> , Mauricio Junca <sup>2</sup> , Mauricio Velasco <sup>3*</sup> . . . . .	17
Estimación de las direcciones de simetría de una distribución y aplicaciones. Alejandro Cholaquidis <sup>1</sup> , Juan Cuesta-Albertos <sup>2</sup> , Ricardo Fraiman <sup>1</sup> , Manuel Hernandez-Banadik <sup>3*</sup> . . . . .	18
Modelos de estimación de variables de calidad de agua en el Río Negro a partir de sensores remotos. Juan Manuel Barreneche <sup>1*</sup> , Bruno Guigou <sup>1</sup> , Federico Gallego <sup>1,2</sup> , Andrea Barbieri <sup>1,2</sup> , Virginia Fernández <sup>1,2</sup> . . . . .	19
Prospección del uso de imágenes de drones para el monitoreo detallado del pastizal nativo. Amparo Quiñones <sup>1*</sup> , Andrés Hirigoyen <sup>1</sup> , Jean Savian <sup>1</sup> . . . . .	21
Análisis preliminares de datos obtenidos con escáner láser terrestre para caracterización de la estructura de un bosque nativo ribereño en gradientes ambientales. Andrea Cardoso <sup>1*</sup> , Carolina Crisci <sup>1</sup> , Nestor Mazzeo <sup>2</sup> , Gonzalo Perera <sup>1</sup> , Santiago Verón <sup>3</sup> . . . . .	23
Una propuesta de modelización para predecir la distribución del Caracol Negro ( <i>Pachycymbiola brasiliana</i> ) como insumo para una pesquería artesanal sustentable. Ignacio Santillán <sup>1</sup> , Ximena Lagos <sup>2</sup> , Fabrizio Scarabino <sup>3</sup> , Angel M. Segura <sup>1</sup> . . . . .	25
Aplicando Machine Learning para predecir la condición reproductiva de peces. Andrés Flores <sup>1</sup> , Rodrigo Wiff <sup>2,3</sup> , Carl R. Donovan <sup>4,5</sup> , Patricio Gálvez <sup>6</sup> . . . . .	27
Revisión de métodos estadísticos y métricas para predecir contaminación fecal en playas recreativas. Victoria Vidal <sup>1*</sup> , Lia Sampognaro <sup>1</sup> , Angel Segura <sup>1</sup> , Fernanda de León <sup>1</sup> , Carla Kruk <sup>1</sup> , Gonzalo Perera <sup>1</sup> , Carolina Crisci <sup>1</sup> . . . . .	28
Comparación de algoritmos de inteligencia artificial para la predicción de concentración de partículas finas en diferentes usos del suelo a partir de variables satelitales. Rodríguez Nuñez, Martín <sup>1,2,3*</sup> , Mónica Balzarini <sup>2,3</sup> . . . . .	30
Detección de cambios en mamografías y evolución de calcificaciones mamarias. Gabriel Wolfmann <sup>1</sup> , Ma. Lucrecia Ballarino <sup>2</sup> , Valeria S. Rulloni <sup>3*</sup> . . . . .	31
Consistency of Nadaraya Watson estimator in non parametric regression model driven by fractional Brownian motion and sampled at random times. Tania Roa <sup>1*</sup> , Meryem Slaoui <sup>1</sup> , Soledad Torres <sup>2</sup> . . . . .	32
Modelización de la madurez sexual de la raya <i>Sympterygia acuta</i> de un Área Protegida de la costa Atlántica de Uruguay: Cerro Verde e Islas de la Coronilla. Ines Pereyra <sup>1*</sup> , Cristina Oddone <sup>2</sup> , Angel Segura <sup>3</sup> . . . . .	33
Una aplicación del Análisis de Sensibilidad al Aprendizaje Automático. Leonardo Moreno <sup>1*</sup> . . . . .	35
Predicción genómica mediante aprendizaje profundo. Maria Ines Faiello . . . . .	36

# HOMENAJE A GONZALO PEREZ IRIBARREN

La Organización de la VII Jornadas de Estadística Aplicada de La Paloma, resolvió dedicar la última sesión de las mismas a homenajear a **Gonzalo Pérez Iribarren**, a 25 años de su desaparición física.



Se encomendó realizar una semblanza del homenajeado al Director del Departamento MEDIA, Gonzalo Perera, quien fue su alumno y compañero de trabajo. El siguiente texto es la referida semblanza.

## Una manera de recordar a Gonzalo Pérez Iribarren.

Gonzalo Perera, Departamento MEDIA, CURE, gperera@cure.edu.uy VII Jornadas de Estadística Aplicada de La Paloma, 21 de octubre del 2023.

Antes de entrar en materia, tres puntualizaciones sobre éste recuerdo de Gonzalo Pérez Iribarren.

En primer lugar, cualquier referencia a hechos o escenas en los que uno haya participado de alguna manera, son simplemente un recurso narrativo. El centro y protagonista aquí es Gonzalo Pérez Iribarren, pero los relatos impersonales, sin contexto, sin la descripción del impacto sobre otras personas, son, como el bronce, fríos y deshumanizadores, todo lo contrario a lo que queremos hacer.

En segundo lugar, aunque esta exposición fue escrita con cierto cuidado, no la leeré, pues prefiero la imperfección del relato que sigue una pauta predeterminada pero que deja espacio a la espontaneidad, nuevamente para que el relato cobre la mayor vitalidad y calidez, como lo merece largamente la persona que recordamos.

Finalmente, quizás lo más importante: el sentido mismo de este recuerdo. Aprendimos de nuestros mayores y de lo que hemos podido conocer por el mundo que las tradiciones académicas fuertes y sanas son un gran activo para construir comunidades académicas fuertes y sanas. Y las tradiciones no se construyen sino hay legado, sino hay memoria permanente, que va pasando de generación en generación. Si hay una tal memoria, puede que esa tradición no tenga interpretación única: puede que dos colegas que se reconocen como frutos del mismo árbol, discrepen o tomen distintos caminos. Eso es propio al no-determinismo y libre albedrío de la especie humana, pero convengamos que, aún ese escenario de un legado que se ramifica en diversas interpretaciones, es francamente preferible a la ausencia de legado, donde todo tema parece presentarse por primera vez, donde se improvisa y cada tanto se redescubre la rueda o las disyuntivas académicas o éticas más básicas. Para decirlo alegóricamente, es preferible tener dos bibliotecas de donde aprender de la experiencia y el saber acumulado, que no tener ninguna biblioteca. Por lo tanto, aquí se trata, por un lado, de rendir homenaje a una persona, un académico, un universitario, que hizo aportes extraordinarios, tales como, para mencionar por ahora solamente uno y el más ligado a estas Jornadas, ser el impulsor de que haya hoy en Uruguay quienes centran su esfuerzo en el desarrollo de la Estadística Aplicada, desde distintos abordajes disciplinares e interdisciplinarios. Pero también, al mismo tiempo, es un intento de cultivar ese legado, mediante un ejercicio de memoria, de compartir recuerdos con las nuevas generaciones, obviamente desde un relato tan sesgado como cualquier otro, pero que invita a sentirse parte e intentar re-expresar en nuestras épocas algo que empezó mucho antes de nuestro tiempo y que, ojalá, nos sobreviva muy largamente.

Cualquier relato lineal sobre la vida de Gonzalo Pérez Iribarren debe empezar diciendo que nació en Carmelo el 26 de junio de 1936. Pero escapando muy rápida y drásticamente de la linealidad narrativa, señalemos que ese dato está lejos de ser una anécdota, ya que Carmelo era un dato muy relevante de la personalidad

de Gonzalo Pérez Iribarren. Lo era en su manera de hablar y reírse fuerte, mucho más de paisano que de académico (y ciertamente para nada de cortesano). En un viaje que hicimos a su pueblo natal, me indicaba árboles en los que se trepaba de gurí y se sentía protagonista de historias de piratas, o me contaba de sus zambullidas en el río.



Figure 1: A falta de imágenes de las zambullidas de infancia en el río...

Uno ha visto en su vida que quienes venimos de pueblitos del interior o de barrios humildes de Montevideo, de alguna manera u otra, llevamos ese sello permanentemente, No con localismos baratos, ni orgullos vanos, sino porque no hay forma de amar a la especie humana que no empiece por el más cercano, ni apego por el único territorio común a todos los humanos (que es nuestro planeta), que no empiece por atesorar imágenes, sonidos, olores y sabores del lugar donde uno nació. En Gonzalo era muy fácil reconocer ese sello, que había resistido estudios varios y pasajes por países diversos, era parte muy nítida de su ser.

Su pueblo de origen también lo marcaría de forma trágica, marcando una inflexión en su vida. Su hermano siete años mayor Ricardo Pérez Iribarren fue sin duda una de las grandes personalidades de la historia de la Facultad de Ingeniería, decisivo en el desarrollo de la Ingeniería Eléctrica en el país. Como mero reflejo de ello, mencionemos que Ricardo fue designado Profesor grado 5 en noviembre de 1961, con dedicación total en noviembre de 1962, y el 23 de abril de 1970 la Facultad le rindió homenaje nombrando al salón 002 “Aula de Electrotécnica Profesor: Ricardo Pérez Iribarren” A raíz de un accidente automovilístico en ruta 1, entre su pueblo natal y Montevideo, Ricardo falleció el 28 de febrero de 1968 y a su lado iba, y resultó seriamente lesionado, Gonzalo.

No era frecuente que Gonzalo hablara de esa circunstancia trágica o de otros dolores que le tocó vivir. Pero, para marcar los matices de su personalidad, a veces, en alguna conversación “mano a mano,” tocaba el tema, y al hacerlo su voz y su mirada cambiaban, se reposaban, se atenuaban, en lo que a uno le parecía un claro indicio de un dolor tan asentado por el paso del tiempo, como profundo y siempre presente. Gonzalo falleció el 27 de agosto de 1998, habiendo superado muy largamente todo pronóstico realizado cuando se le diagnosticó su enfermedad. Cinco años después, yo era director del IMERL y entre dicho instituto, el CMAT y el IESTA, los tres lugares donde Gonzalo dictó clase desde el reinicio de la vida democrática, se organizó el Primer Encuentro Nacional de Jóvenes Matemáticos y Estadísticos, en honor a Gonzalo Pérez Iribarren. El acto de cierre se hizo en el salón 002, que lleva el nombre de su hermano y obviamente, no fue ninguna casualidad. Pero mucho menos casualidad aún fue que para homenajearlo se hiciera un evento académico centrado en los jóvenes colegas. Muchos y muchas que por ese entonces hacían sus primeras armas, hoy ya han realizado una trayectoria sumamente destacada y ciertamente eso mucho alegraría a Gonzalo. Por ello, promover las nuevas generaciones siempre es la mejor manera de transmitir su legado. Volveremos sobre este punto hacia el final, pero bien vale enfatizarlo desde ya.



Gonzalo era portador de una sensibilidad muy fina, que a veces sorprendía a quien presuponía que la fuerte voz paisana y la sutileza no podían habitar el mismo ser. Gonzalo pintaba con talento, escribía con talento. Cristiano, católico apostólico romano, estudió en Roma y estuvo a punto de ordenarse sacerdote.



Figure 2: Gonzalo estudiando en Roma

No tomó los hábitos y decidió formar una gran familia, junto a Beatriz de Sierra, pero su luna de miel fue ir a vivir entre los cañeros de Artigas en 1968, legando el único libro que yo conozca que detalla como se sentía, como se vivía, estando en ese lugar y momento..



Figure 3: El casamiento de Gonzalo Pérez Iribarren y Beatriz de Sierra y la pareja años despues.

En la primera mitad de los 90 formamos dupla para ir a dictar un curso de Series de Tiempo para los ingenieros de Despacho de Carga de UTE, en Melilla. Un auto de la UdelaR nos llevaba y traía, y en una parte del curso iba él, en otra iba yo, y en el camino pasábamos por los accesos del noroeste de Montevideo y por la entrada de los asentamientos, que por ese entonces se multiplicaban rápidamente. Cerca del puente de Carlos María Ramírez, había una esquina particularmente paradójal. Volviendo de Melilla a Montevideo, uno veía a su izquierda el flanco de un gran edificio que había sido pintado íntegramente con una publicidad de un conocido refresco, invitando a disfrutar los placeres de la vida. Mientras, a la derecha se amontonaba el rancherío de lacerante precariedad, que marcaba el ingreso a un asentamiento por ese entonces recién instalación en el Norte del Cerro. Esa publicidad, en frente a una tan clara expresión de la exclusión social,

parecía una broma de muy mal gusto, pero uno la miraba una y otra vez, y seguía adelante. Cerca de fin de año, y con motivo de la Navidad, Gonzalo me regaló una tarjeta. Pero no una comercial, sino una hecha por él. En ella había dibujado con tinta negra y sobre fondo blanco, de manera perfecta, aquella esquina, con su edificio publicitando los placeres de la vida y sus tan precarios ranchitos. Gonzalo había escrito, abajo de su dibujo; “Si volviera a nacer hoy...dónde nacería?” Gonzalo veía los Evangelios en los que basaba su vida, en tiempo presente, y siempre, desde la sensibilidad.

Esa sensibilidad no le impedía tener un carácter fuerte y posiciones firmes, que incluso podían a veces parecer tercas. Cuando alguna vez me tocó ser objeto de alguna crítica de su parte, no fue ni elíptico ni complaciente, y aunque no me gustara lo que oía, era imposible no apreciar su frontalidad y honestidad. Cuando ante ciertos hechos de la comunidad académica, tuvo discrepancias que para él no eran salvables, renunció a alguna posición, en una situación en que, seguramente, la inmensa mayoría de las personas hubieran sido más pragmáticas.

Tampoco esa sensibilidad le impedía tener actitudes de genuina picardía para ejercer la docencia, que tanto amaba. Mi primer trabajo estadístico aplicado fue hacia fines del año 88. Gonzalo me decía que estaba haciendo muchos estudios teóricos en Probabilidad y Estadística, u otras ramas de la Matemática, y que eso estaba muy bien, pero que le parecía que era parte importante de la formación “embarrarse” con los datos, y ver cómo se aplicaba la teoría en la cruel realidad. Fue así que, obviamente en papel, me entregó un montón de datos sobre el crecimiento de colonias de hongos ante distintos tratamientos, que había que comparar entre sí, clasificar, seleccionar los que tuvieran mayor efectividad, etc. Obviamente, tras entender lo que había que hacer desde el punto de vista sustantivo, le manifesté no tener la menor idea de cómo aplicar lo que había aprendido sobre ese mar de números. Me dijo; “Tocayo, esto te va a ayudar, lo que precisás está en el capítulo siete” y me entregó “The Analysis of Variance” de Scheffé, un gran libro pero de difícil digestión y que para llegar al capítulo siete sin rendirse y con alguna chance de entender algo, requería cierta perseverancia. Gonzalo sabía perfectamente la dificultad que me estaba planteando y por eso lo hizo de forma casi provocadora, una suerte de “a ver si te da el cuero,” El trabajo sobre problemas concretos nunca me abandonó desde ese entonces. Por si fuera poco, en el 90, Gonzalo nos invitó a Ernesto Mordecki, al inolvidable Ariel Roche, a José Diaz Milessi y a mí a trabajar sobre los coliformes fecales en las playas de Montevideo, en tiempos de la construcción del emisario subacuático. No puedo evitar acordarme una y otra vez de esa experiencia, cuando hoy, en nuestro Departamento MEDIA trabajamos y hemos trabajado, con otras herramientas y objetivos, sobre los colifecales y otros contaminantes. Más de tres décadas después sigo tratando de ver si realmente entendí el capítulo siete de Scheffé, trabajando en equipos donde siempre he tenido el placer de trabajar con bellas personas y además, con colegas mucho más capaces que yo. Todo ello gracias a aquellos consejos e invitaciones de Gonzalo Pères Iribarren, o como le decía todo el mundo, “el flaco.”

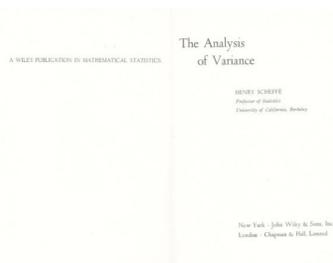


Figure 4: Una picardía de Gonzalo

Gonzalo entró como Ayudante grado 1 con 20 horas semanales en el entonces IME en 1969, y luego pasó a Asistente grado 2 con 30 horas semanales de 1971 hasta que en 1973 la dictadura interviene la Universidad y arrasa (en particular) al IME. Tuvo que irse del país, estuvo en Buenos Aires donde fue padre por primera vez, se hizo luego maracucho estudiando y trabajando en la Universidad del Zulia, en Venezuela, donde culminó su maestría bajo la orientación de Enrique Cabaña quien por ese entonces residía en Caracas y por quien siempre expresó una enorme admiración intelectual.

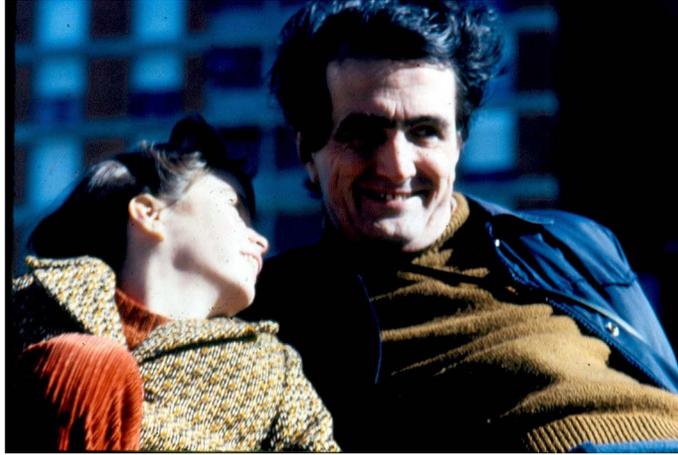


Figure 5: Gonzalo con su hija Leticia

De esa etapa en Maracaibo, muy marcada en su vida desde todo punto de vista y de la que le brotaban una multitud de anécdotas, surgió además una publicación conjunta con Heber Nieto, a la que recordaba con mucho cariño (Una ley de arco seno discreto, de 1984).

Volvió al país aún en dictadura, daba clases particulares en su pueblo natal y venía a Montevideo para participar en las primeras reuniones donde se pensaba lo que terminaría siendo el PEDECIBA y todo el renacimiento de la vida académica en el país. En 1984 asume un cargo de Profesor Agregado, grado 4 con 40 horas semanales del IME, asumiendo además posteriormente la dirección del instituto en un momento en que todo estaba por hacerse y vaya que hizo unas cuantas cosas.



Figure 6: Gonzalo Director del IME, con Mariel Etchemendy, secretaria.

Por ejemplo, se acercó y promovió a un joven Ayudante, llamado Ernesto Mordecki. El nivel de la trayectoria académica de mi amigo Ernesto, posiblemente baste por sí sola para valorar cuánto el flaco cambió y para bien en la vida de personas y de instituciones. Fue Profesor también en el CMAT y dictó clases en la Licenciatura en Estadística, en el IESTA. Fue además profesor y adscrito en Enseñanza Secundaria. Estudió problemas de Estadística no paramétrica, hizo dos publicaciones al respecto en co-autoría con Ricardo Fraiman (Conservative confidence bands for nonparametric regression and Nonparametric regression estimation in models with weak error's structure, ambos de 1991), luego retomó trabajos de Wald aplicándolos a propiedades de la estimación por máxima verosimilitud en modelos paramétricos mal planteados (On the likelihood ratio and the Kullback-Leibler distance, en 1992. Trabajó junto a Carlos Mechoso de manera pionera sobre los impactos hidrológicos de El Niño y La Niña ( Streamflow in southeastern South America and the southern oscillation en 1992 y, también junto a José Luis Genta, A recent increasing trend in the streamflow of rivers in southeastern South America en 1998) y en 1999 se publica un muy lindo trabajo de modelización hidrológica ( Cadenas de Markov gobernando algunos procesos aplicables a los ríos: Aplicaciones estadísticas a algunos ríos de la región ).



Este último trabajo fue leído por un entonces joven estudiante, Jorge Graneri, y no me cabe duda que fue un insumo relevante para su tesis de Maestría en Ingeniería Matemática (Análisis de datos hidrológicos y procesos de memoria larga), aprobada en el 2014. Ese aprendizaje juvenil de Jorge se enmarcó en los trabajos que realizaba el Laboratorio de Probabilidad y Estadística del CMAT, que Gonzalo fundó y dirigió y desde donde se estudiaron diversos problemas, como los aportes a las represas de UTE y otros requerimientos de la empresa eléctrica nacional, al mismo tiempo que se formaba Jorge, que yo intentaba convencerme de que había digerido el capítulo 7 de Scheffé, o realizaban parte fermentales de su formación destacados colegas como Héctor Cancela (ex Decano de FING, Presidente de ADUR, etc.) o María Isabel Cañette (hoy Principal Mathematician and Statistician at StataCorp LLC). El que en 1999 hayamos relanzado el LPE, pero entonces desde el IMERL (el IME adquirió el nombre de su fundador, Rafael Laguardia, en 1987) no fue ninguna coincidencia, si había que resignarse a no poder escuchar la sonora carcajada del flaco, era absolutamente impensable no continuar trabajando sistemáticamente desde la Universidad para muchas instituciones de la sociedad que requerían estudios estadísticos de mediano y largo plazo y que al mismo tiempo brindaban la oportunidad de formarse a nuevas generaciones de jóvenes colegas. Según palabras de María Simón, entonces Decana de FING, en el homenaje del 2003, sobre Gonzalo: “Mantuvo siempre el interés de hacer matemática vinculada a problemas de la tecnología y de la naturaleza, encontrando en esa vinculación, fuente de investigación matemática de calidad y acercándose a las otras áreas con la voluntad e inteligencia necesaria para aprehender y formular los problemas de la realidad en términos rigurosos. En ese sentido es recordado con gran aprecio por muchos colegas ingenieros.”

Gonzalo fue gremialista, fue consejero, fue un universitario a carta cabal, fue un formador, un tendedor de puentes entre disciplinas, una persona extraordinariamente culta, poseedora de muy amplios e inusuales conocimientos que disimulaba con humildad y a los que sólo apelaba cuando eran necesarios, fue palabra de aliento, de estímulo, y también de consuelo. Lo conocí a principios de 1985 por razones personales. En enero de 1986, un domingo de mañana, me invitó a la casa porque me quería convencer que estudiara la Licenciatura en Matemática. Los hechos muestran que fue muy persuasivo. En lo personal no hay ningún texto que yo pudiera escribir que ni siquiera se aproximara a reflejar cuánto le debo y cuánto le agradezco a la vida haber tenido el privilegio de conocerlo, quererlo y recibir tantas enseñanzas e influencias, que espero haber logrado asimilar al menos en parte, incluyendo el capítulo siete de Scheffé.

Estas torpes palabras intentaron recordar y para muchos hacer conocido a Gonzalo Pérez Iribarren, en un tiempo bastante acotado, tarea singularmente difícil por la riqueza del personaje.

Pero para que ose trate solo de palabras, sino hacer lo que seguramente Gonzalo hubiera deseado, que es recordarlo haciendo lo que para él era lo más importante, como se decidió hacer en el 2003, me voy a permitir presentar en público una propuesta: que a partir de las próximas Jornadas, que serán las octavas, **una de las conferencias plenarias lleve el nombre de Gonzalo Pérez Iribarren** y que la persona a cargo sea la que resulte ganadora de un concurso, con reglas precisas a determinar, de trabajos de Estadística Aplicada realizados por jóvenes. Reconocer a una o un colega joven y darle el espacio para que pueda quizás hacer su primera conferencia en un evento, me permite sospechar que sería lo que Gonzalo preferiría frente a cualquier mar de palabras. Obviamente, esta propuesta no es exclusiva para el MEDIA, sino para todos los departamentos o institutos aquí representados y todo aquel que comparta el objetivo de desarrollar la Estadística Aplicada.

Ojalá pueda haber ayudado en algo para que algo del espíritu del flaco haya asomado y sea visto por todos

Uds como alguien respetable y respetado, admirable y admirado, recordable y recordado.  
Muchas gracias.

## CONFERENCIAS PLENARIAS

### **Optimización en procesos estocásticos. Ernesto Mordecki**

Centro de Matemática, Facultad de Ciencias, UdelaR.

Se presenta una matriz de problemas versus modelos en procesos estocásticos. De esa forma aparecen múltiples problemas de optimización de funcionales de procesos estocásticos, que van desde los problemas de barrera (determinación de la probabilidad de alcanzar un nivel dado para un proceso), el problema de parada óptima (elegir el mejor momento para reclamar un pago que depende de un proceso), los juegos de Dynkin (una variante de la situación anterior con dos jugadores) y los juegos de campo medio (donde el segundo jugador es la síntesis de un grupo grande de jugadores). Una vez presentada la matriz de problemas, se presenta una revisión comenzando con la solución de algunos que ya son clásicos, concluyendo con un enunciado de algunos problemas abiertos.

### **Comer y pescar: ecología trófica de lobas y leonas marinas y su interacción con pesquerías. Valentina Franco-Trecu**

Departamento de Ecología y Evolución, Facultad de Ciencias, UdelaR.

Durante mis primeros años como investigadora, enfocada en el estudio del comportamiento maternal y éxito de crianza de lobas marinas, identificamos que la duración de los viajes de alimentación en el mar tenía un gran impacto en la mortalidad de las crías. Aunque reconocíamos la importancia de lo que sucedía en el mar con estos organismos, estos resultados nos motivaron a sumergirnos en el mar para comprender lo que ocurría cuando las lobas y leonas marinas no estaban a nuestra vista. Inicialmente, investigamos la composición de la dieta de ambas especies mediante una combinación de métodos. Utilizamos la identificación de restos duros en fecas y analizamos las firmas isotópicas de  $d_{13}C$  y  $d_{15}N$  tanto de las hembras como de sus presas, incorporando esta información en modelos de mezcla bayesianos. Además, evaluamos la segregación entre especies y sexos a mediano plazo (2-3 años) y la especialización trófica a nivel individual mediante el análisis isotópico de las vibrisas (tejido inerte de crecimiento continuo) de hembras y machos. Para conocer las áreas de alimentación y el comportamiento de buceo de las hembras de ambas especies, utilizamos telemetría satelital y sensores de buceo. En el caso de las leonas marinas, que se alimentaban en áreas costeras de la plataforma continental, evaluamos el solapamiento con pesquerías artesanales e industriales de fondo costero, encontrando un alto grado de superposición en el uso de recursos y diferentes niveles de solapamiento espacio-temporal. Además, aplicamos la técnica de isótopos estables en cráneos de lobos y leones marinos del Holoceno y del período 1941-2012 para evaluar si la partición de recursos entre especies ha sido constante a lo largo del tiempo, considerando los cambios en las abundancias poblacionales y en la estructura de tamaño de las presas debido a la explotación pesquera. Encontramos segregación de los nichos isotópicos entre especies durante el Holoceno medio, que persistieron a principios de la segunda mitad del siglo XX, diferenciación que se redujo con el tiempo. Esta convergencia gradual de los nichos isotópicos se debe probablemente a la disminución del tamaño promedio de los peces demersales debido a la pesca intensiva de las clases de mayor tamaño. Esto ha aumentado la accesibilidad de estos peces a los organismos con una boca más pequeña, como los lobos marinos de ambos sexos y las hembras de leones marinos. Por otro lado, evaluamos la captura incidental de lobos y leones marinos en la pesquería industrial de arrastre de fondo costero durante un período de tres años, estimando una captura anual de aproximadamente 80 leones marinos y 25 lobos finos. También observamos que la interacción con las pesquerías resultaba en el enredo de artes de pesca en los animales. En el caso de los leones marinos, cerca del 50% de las lesiones fueron graves, y más del 60% de los objetos provenían de la pesca artesanal y recreativa costera. En el caso de los lobos finos, cerca del 40% de los objetos procedían de la pesca industrial pelágica, con la cual esta especie comparte sus áreas de alimentación. Estos incidentes de enredo afectan el éxito de estos organismos e incluso pueden llevar a su muerte. Para abordar la problemática de la interacción entre los leones marinos y las pesquerías artesanales, llevamos a cabo una investigación participativa con pescadores y otros actores locales. Juntos desarrollamos un nuevo arte de pesca selectivo con el objetivo de evitar la interacción entre los leones marinos y la pesquería con palangre.

## **Medidas de calidad para fusión de imágenes. Ana Georgina Flesia**

Profesora Titular Diploma en Estadística, Universidad Nacional de Córdoba Investigadora Independiente-Conyctet

La fusión de imágenes se considera una metodología de mejora eficaz ampliamente incluida en los sistemas de imágenes de alta calidad. No obstante, al igual que otras técnicas de mejora, la evaluación de la calidad de los resultados se realiza en estudios de evaluación subjetiva de muestras pequeñas que son muy limitadas a la hora de predecir la calidad percibida por el ser humano de los resultados generales de la fusión de imágenes. Los métodos simples, ciegos, universales y perceptivos para evaluar la calidad de las imágenes fusionadas siguen siendo un reto, parcialmente resuelto sólo en aplicaciones particulares. En esta charla discutiremos medidas de fidelidad para fusión de imágenes basadas en descomposiciones multiescala, en especial la medida MS-QW de Martínez et al (2019) y la medida perceptual de Tang et al (2020).

## **Deep learning for semantic segmentation; some medical applications. Badih Ghattas**

Aix Marseille School of Economics (AMSE), Université d'aix-Marseille.

# PONENCIAS

## Regresión en Variedades. Alejandro Cholaquidis<sup>1\*</sup>

<sup>1</sup> Centro de Matemática, Facultad de Ciencias, Universidad de la República. \*acholaquidis@cmat.edu.uy

Los datos en la esfera o en el toro son ejemplos clásicos de datos en variedades. Los primeros se usan, por ejemplo, para representar fenómenos a nivel terrestre, y los segundos para representar fenómenos que tengan un comportamiento cíclico o periódico (por ejemplo, ciclos circadianos). Otro ejemplo clásico son los datos de viento (dirección e intensidad) que por su naturaleza se pueden representar en un cilindro. Existen una enorme cantidad de fenómenos que por sus características es necesario representarlos en espacios más generales que  $\mathbb{R}^d$  para no perder información de los mismos. En estos casos una alternativa a  $\mathbb{R}^d$  es pensar que los datos toman valores en una variedad, es decir, una estructura que si bien localmente, en un entorno pequeño de un punto, es como  $\mathbb{R}^d$ , a escala mas grande aparece la curvatura. En el caso de la esfera esto hace que el camino más corto que une dos puntos no sea la recta que los une (que no pertenece a la esfera) sino un arco de circunferencia en la misma. Lo mismo pasa en el cilindro o en el toro. Este salto en generalidad supone grandes desafíos, es claro que no se pueden promediar datos ya que en general no vamos a obtener un punto en la variedad, con lo cual las herramientas clásicas de la probabilidad como el Teorema Central del Límite, requieren de adaptaciones. Incluso la definición misma de esperanza no es trivial, no siempre está garantizada su existencia o unicidad. En la charla veremos, a modo de motivación inicial, algunos ejemplos de datos en variedades, y, más en general, datos en espacios métricos (como lo son, por ejemplo, los árboles filogenéticos que se usan para modelar la evolución de un virus, un lenguaje, etc.) Los problemas de regresión y clasificación se han llevado al marco de estadística en variedades, pero típicamente se asume que las variables explicativas toman valores en una variedad, y el output es una variable real. Esto se debe a que los métodos de regresión clásicos como vecinos mas cercanos, kernel, etc., se pueden extrapolar de manera “sencilla” a este escenario. Por ejemplo, en el clasificador de vecinos más cercanos dado un imput  $X_i$  en una variedad, para obtener el output  $Y_i$  hay que considerar distancias geodésicas a  $X_i$ , y el valor estimado para  $Y_i$  es el promedio de los output correspondientes a los k-vecinos más próximos de  $X_i$  en la muestra. Dicho promedio es un promedio de números reales y por lo tanto se puede realizar sin inconvenientes. El problema se vuelve mucho más complicado cuando el output está en una variedad, ya que allí no se puede promediar. En su lugar hay que tomar medias de *Frechet*, cuya existencia y unicidad está garantizada para variedades de curvatura negativa, o en casos muy particulares. Recordemos que la media de Frechet de una variable aleatoria  $X$  a valores en un espacio métrico  $(M, d)$  se define como el conjunto de los  $x$  que minimizan  $f(x) = E(d^2(X, x))$ . Veremos que bajo ciertas condiciones esta media de Frechet existe, y se puede estimar, a partir de una generalización del estimador clásico de vecinos más cercanos.

# ¿Qué es la inferencia causal? ¿Cómo podemos incorporarla al desarrollo de modelos de aprendizaje supervisado? Gabriel Illanes<sup>1</sup>

<sup>1</sup>\* IDATHA, Pyxis Ecosystem

\*gabriel.illanes@idatha.com

La propuesta de esta charla es explorar los conceptos básicos de inferencia causal, mostrar cómo nos puede ayudar a generalizar cualquier modelo de aprendizaje supervisado, y compartir experiencias de IDATHA en proyectos que incorporaron estas ideas. Esta charla se puede enmarcar en el eje temático 1 del evento. Las básicas del tema las pueden encontrar en (Pearl, Glymour, and Jewell 2016).

A modo de resumen: decimos que las variables  $X_1, \dots, X_n$  son una *causa directa* de la variable  $X_{n+1}$  si existe una función  $f_{X_{n+1}}$  tal que

$$X_{n+1} = f_{X_{n+1}}(X_1, \dots, X_n).$$

A modo de visualización, podemos pensar que cada variable  $X_1, \dots, X_n, X_{n+1}$  es un nodo de un grafo dirigido, en el cual el nodo  $X_i$  apunta a el nodo  $X_j$  si  $X_i$  es causa directa de  $X_j$ . Cuando el grafo resultante es acíclico, lo denominamos *grafo causal* o *red bayesiana*.

Podemos decir que el principal objetivo de la inferencia causal es encontrar la red bayesiana asociada a un conjunto de variables, encontrando todas las causas directas existentes entre ellas. Vale destacar que, en principio, no necesitamos conocer o estimar las funciones  $f_{X_i}$ , solo saber si existen o no. Para atacar este problema, resaltamos dos tipos de estrategias. En la primera, detectamos si distintos pares de variables son condicionalmente independientes a no con respecto a subconjuntos del resto de las variables, y proponer redes bayesianas coherentes a nuestros hallazgos (Tsamardinos, Brown, and Aliferis 2006). En la segunda, imponemos hipótesis sobre el tipo de relaciones existente entre nuestras variables, y estimamos simultáneamente la existencia y la forma de las funciones  $f_{X_i}$  Zheng et al. (2018).

La inferencia causal puede usarse para complementar cualquier trabajo de clasificación supervisada. Usualmente, en un problema de aprendizaje supervisado, nosotros planteamos hipótesis de la forma

$$Y = f(X_1, \dots, X_n),$$

ya que queremos predecir el valor de la variable  $Y$ , sabiendo los valores de las variables  $X_1, \dots, X_n$ . El problema es que, al hacer esto, estamos asumiendo la red bayesiana del modelo: todas (o varias) las variables  $X_i$  apuntarán a  $Y$ , pero negamos la posibilidad de que las variables  $X_i$  sean causadas directamente por algunas de las demás, o incluso que  $Y$  sea causa directa de alguna de las variables  $X_i$ . Por lo tanto, al incorporar inferencia causal a nuestros modelos de aprendizaje supervisado, mejoramos su interpretabilidad y obtenemos una comprensión global de nuestras variables y sus interacciones; como mínimo, es fácil ver que construir la red bayesiana es la mejor manera de realizar selección de variables a la hora de predecir  $Y$ . Además, cuando disponemos de la red bayesiana y estimaciones de todas las funciones  $f_{X_i}$ , tenemos acceso a herramientas de estudio de escenarios sintéticos: *intervenciones* (útiles para la creación de métricas de impacto libres de sesgos) y *contrafácticos* (útiles para la evaluación *a posteriori* de toma de decisiones).

La charla durará 20 minutos, y comenzará con una breve introducción del presentador, de IDATHA, y por qué comenzamos a estudiar inferencia causal. Luego, pasaremos a las definiciones básicas de la inferencia causal y conversaremos sobre la paradoja de Simpson. La charla finalizará con ideas sobre como puede ayudar la inferencia causal a cualquier problema de clasificación supervisada. Nos gustaría que esta charla sirva como incentivo a fomentar el estudio de inferencia causal en la comunidad, y como motivación para el comienzo de colaboraciones de trabajos entre UdelaR e IDATHA.

# Tres resultados para implementar la teoría asintótica extremal en el caso de muestras grandes de datos fuertemente dependientes. Gonzalo Perera<sup>1</sup>, Carolina Crisci<sup>1</sup>, Lía Sampognaro<sup>1</sup>

<sup>1</sup>Departamento MEDIA, CURE, Universidad de la República. gperera@cure.edu.uy

La clásica teoría de Fisher-Tippet-Gnedenko (FTG, en lo que sigue) establece que el máximo de una muestra de  $n$  datos *iid*, converge en distribución, cuando  $n$  tiende a infinito, a una distribución extremal. Eso se traduce en la práctica en que, si se dispone de una muestra en la que cada dato resulta de maximizar una cantidad grande de registros, suele intentarse ajustar dicha muestra a una distribución extremal. Los trabajos de, entre otros, Leadbetter, Lindgren y Rootzen permiten extender FTG cuando los  $n$  datos, en lugar de iid, son estacionarios y débilmente dependientes. Sin embargo, en aplicaciones Ambientales estos resultados pueden no ser aplicables, ya que hay situaciones donde claramente los datos exhiben dependencia fuerte. Para adaptarse a esa realidad, un resultado de Crisci-Perera (2022) muestra que, bajo dependencia fuerte la distribución asintótica del máximo es una mezcla de distribuciones extremales, lo cual aporta resultados fuera del marco FTG. Esto propone, en la práctica, intentar ajustar mezclas de distribuciones extremales a muestras que consisten en máximos de gran cantidad de registros que presenten dependencia fuerte. Para que ello sea posible, un segundo resultado de Crisci-Perera-Sampognaro (2023b) permite, estimar las componentes (pesos y parámetros distribucionales) de una muestra fuertemente dependiente que sigue una mezcla de distribuciones extremales. Para poder cerciorarse de la validez del ajuste, un tercer resultado de Crisci-Perera-Sampognaro (2023a) desarrolla un test del tipo Kolmogorov-Smirnov para muestras fuertemente dependientes. La conjunción de estos tres resultados permite por lo tanto ajustar distribuciones de muestras de datos extremales en contextos hasta ahora no cubiertos por FTG y extensiones, y que puede ser de utilidad en el análisis estadístico de datos ambientales.

Crisci, C. and Perera, G. (2022) Asymptotic Extremal Distribution for Non-Stationary, Strongly-Dependent Data. *Advances in Pure Mathematics*, 12, 479-489. doi:10.4236/apm.2022.128036. Crisci, C., Perera, G. and Sampognaro, L. (2023a). Goodness-of-Fit Test for Non-Stationary and Strongly Dependent Samples. *Advances in Pure Mathematics*,13, 226-236. doi:10.4236/apm.2023.135016 Crisci, C., Perera, G. and Sampognaro, L. (2023b). Estimating the Components of a Mixture of Extremal Distributions under Strong Dependence. *Advances in Pure Mathematics*, 13, 425-441. <https://doi.org/10.4236/apm.2023.137027>

## Estimación univariada de la densidad adaptando el método root-unroot con una función núcleo. Mathias Bourel<sup>1</sup> y Jairo Cugliari<sup>2</sup>

1 Universidad de la República, Uruguay, mbourel@fing.edu.uy 2 Université Lyon 2, France, jairo.cugliari@univ-lyon2.fr

La estimación de una densidad puede plantearse como un problema de regresión gracias a una transformación adecuada. A menudo conocida como transformación root-unroot (Brown et al 2010), esta estrategia permite utilizar cualquier método de regresión. La construcción utiliza una partición del soporte de la densidad para crear clases, como el histograma, y estimar las frecuencias absolutas de cada clase. Luego, se ajusta una curva para estimar la forma de la densidad a partir de la raíz de las frecuencias. La curva se vuelve a escalar a la escala original y luego se transforma para obtener una densidad. En este trabajo, utilizamos el procedimiento root-unroot para la estimación de la densidad, pero ayudándonos de una función núcleo para mejorar la estimación. Realizamos varias simulaciones utilizando densidades univariantes para comparar su error cuadrático medio con el de otros métodos (Bourel y Cugliari 2019) y proponemos finalmente una banda de confianza basada en este estimador.

L Brown, T Cai, R Zhang, L Zhao, H Zhou, 2010. The root–unroot algorithm for density estimation as implemented via wavelet block thresholding. *Probab. theory related fields* 146, 401–433.

M Bourel, J Cugliari, 2019. Bagging of density estimators. *Comput. Stat.* 34, 1849–1869

# Métodos algebraicos para el estudio de tiempos y localizaciones de salida en difusiones estocásticas. Didier Henrion<sup>1</sup>, Mauricio Junca<sup>2</sup>, Mauricio Velasco<sup>3\*</sup>

<sup>1</sup> LAAS-CNRS, Toulouse, France

<sup>2</sup> Universidad de los Andes, Bogotá, Colombia

<sup>3\*</sup> Universidad Católica del Uruguay (UCU), Montevideo, Uruguay

\*mauricio.velasco@ucu.edu.uy

Suponga que el vector  $X(t)$  es la solución de una ecuación diferencial estocástica iniciando en un punto  $x_0$  de un dominio  $D$  y que  $t$  es el primer instante (stopping time) en el que la solución toca la frontera de  $D$ . En esta charla describiremos nuevas herramientas algorítmicas para estimar los momentos polinomiales  $E[g(X_t)]$  y la densidad de la localización del punto de salida en  $\partial D$  cuando el conjunto  $D$  es semialgebraico. Estas herramientas provienen de la geometría algebraica real (el llamado método de momentos o de sumas-de-cuadrados) y de la teoría de kernels reproductores (el kernel de Christoffel Darboux) y tienen el potencial de aplicarse a difusiones analíticamente intratables y de extenderse a muchos otros contextos de interés en probabilidad. La charla será una introducción autocontenida a este círculo de ideas.

**Estimación de las direcciones de simetría de una distribución y aplicaciones.**  
**Alejandro Cholaquidis<sup>1</sup>, Juan Cuesta-Albertos<sup>2</sup>, Ricardo Fraiman<sup>1</sup>, Manuel Hernandez-Banadik<sup>3\*</sup>**

<sup>1</sup> Centro de Matemática, Facultad de Ciencias, Universidad de la República, Uruguay <sup>2</sup> Departamento de Matemáticas, Estadística y Computación. Facultad de Ciencias. Universidad de Cantabria. España <sup>3\*</sup> Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay manuel.hernandez@fcea.edu.uy

La simetría, en general, es una propiedad importante en la estadística y ha sido ampliamente estudiada. Recientemente, en [Hudecová y Siman 2021, 2022] se han propuestos distintos test de hipótesis para testear si la muestra de datos proviene de una distribución que es simétrica respecto de una dirección dada, pero, hasta donde sabemos, no existe un estimador de la dirección. En este trabajo, proponemos un estimador y probamos su consistencia. Además presentamos dos aplicaciones metodológicas. Por un lado, mostramos que se puede mejorar significativamente la estimación de una densidad utilizando las direcciones de simetría estimadas; y por otro lado, presentamos este estimador como un insumo para el trabajo en análisis de imágenes médicas.

S. Hudecova, M. Siman. Testing axial symmetry by means of directional regression quantiles. *Electronic Journal of Statistics* 15 (2021) 2690-2715. S. Hudecova, M. Siman. Testing axial symmetry by means of integrated rank scores. *Journal of Nonparametric Statistics* (2022) 1-17.

# Modelos de estimación de variables de calidad de agua en el Río Negro a partir de sensores remotos. Juan Manuel Barreneche<sup>1\*</sup>, Bruno Guigou<sup>1</sup>, Federico Gallego<sup>1,2</sup>, Andrea Barbieri<sup>1,2</sup>, Virginia Fernández<sup>1,2</sup>

<sup>1</sup> División Información Ambiental. Ministerio de Ambiente <sup>2</sup> Facultad de Ciencias. Universidad de la República \*jumanbar@gmail.com

Los sistemas acuáticos continentales son parte activa del mantenimiento de la biodiversidad y la regulación biogeoquímica. Representan un aspecto fundamental para la calidad de vida de la sociedad garantizando el acceso a un componente fundamental como lo es el agua. No obstante, las diferentes intervenciones antrópicas determinan un escenario de constante vulnerabilidad que afecta la calidad del agua en los sistemas acuáticos. Por esta razón el monitoreo de los parámetros biofísicos relacionados con el agua se constituye en uno de los principales medios para comprender la dinámica de estos sistemas y contribuir así en la mejora de su gestión. En general, el uso más extendido para el monitoreo de la calidad del agua se basa en muestreos a campo. Esta aproximación provee de mediciones de buena precisión en sitios puntuales y espaciados en el tiempo, pero no ofrece coberturas a nivel de grandes áreas o de alta frecuencia. En contraposición, el sensoramiento remoto permite registrar variaciones espacio temporales en sistemas acuáticos completos, a la vez que reduce los esfuerzos en términos de costo y tiempo, proporcionando una fuente de datos complementaria a los enfoques tradicionales de monitoreo. Esto es posible a partir de la información espectral contenida en las imágenes satelitales y el desarrollo de modelos bio-ópticos que registran los efectos de la interacción de la radiación solar con los componentes ópticamente activos del agua. La estimación satelital de variables de calidad de agua es un aporte novedoso para el análisis y la gestión ambiental, no solo a nivel internacional, sino que también en Uruguay. Si bien existen modelos globales y regionales para la estimación de parámetros de la calidad del agua, sus ajustes son variables y dependen de las características de los cuerpos de agua. Es por esta razón que, para acceder a este tipo de datos, se hace necesario el desarrollo de modelos locales para cuerpos de agua interiores de Uruguay. El objetivo de este trabajo fue generar modelos locales para el Río Negro y contrastarlos con modelos globales y regionales de estimación de tres parámetros de calidad de agua (clorofila-a, turbidez y materia orgánica disuelta coloreada, CDOM). Para ello, entre junio de 2021 y febrero de 2022 se realizaron 32 jornadas de monitoreo radiométrico y de medición de parámetros de calidad del agua en el Río Negro. Durante las mismas, se recolectaron pares de datos radiométricos (firmas espectrales) y de las variables de interés, resultando en 184 pares de datos de clorofila-a, 97 pares de datos de turbidez y 34 pares de datos de CDOM. Con los datos radiométricos, se simuló las reflectancias para el satélite Sentinel 2 y se calibraron algoritmos locales de estimación satelital para los tres parámetros de calidad de agua. Los modelos ajustados usaron como variable predictora ( $x$ ) distintas combinaciones de bandas de Sentinel 2. Para la clorofila-a se realizó un ajuste polinomial basado en  $x = \frac{(B5+B3-B4)}{(B5+B4+B3)}$ . En el caso de turbidez se realizó una regresión lineal con  $x = (B5 + B6)/2$ . Finalmente, se ajustó para CDOM un modelo exponencial ( $y = a * e^{xb}$ ) con  $x = (B2 + B4)/2$ . Para evaluar el desempeño de los modelos desarrollados se utilizaron datos independientes para las tres variables de calidad de agua estudiadas y las siguientes métricas: Precisión Simétrica Media (“Mean Symmetric Accuracy,” MSA); Precisión Simétrica Mediana ( $\epsilon$ ); Porcentaje Simétrico Signado ( $\beta$ ), Pendiente (“Slope,”  $S$ ) e Intercepto ( $I$ ). Estas mismas métricas se utilizaron para evaluar el desempeño de otros algoritmos internacionales o regionales referenciados por la literatura científica. Para la clorofila-a, se evaluaron 3 algoritmos propuestos por: Mishra y Mishra (2012), Pahlevan et.al. (2020) y Drodz et.al. (2019). Para turbidez, se evaluaron 3 algoritmos propuestos por: Zhan et.al. (2022), Delegido et.al. (2019) y Maciel (2022). Por último, para el CDOM, se evaluaron 2 algoritmos desarrollados por: Chen et.al. (2017) y Ficek et.al. (2011). Los resultados mostraron diferencias importantes entre los modelos de estimación de las variables de calidad de agua. Como era de esperar, en términos generales, los modelos desarrollados localmente para el Río Negro fueron más precisos en la estimación de los tres parámetros de calidad de agua analizados, en las 5 métricas consideradas. Para la clorofila, el modelo desarrollado localmente mostró un  $MSA : 0,47$ ;  $\epsilon : 0,31$ ;  $\beta : -0,05$ ;  $Pendiente : 0,9$ ;  $Intercepto : 0,06$ . Para la turbidez, las métricas fueron:  $MSA : 0,18$ ;  $\epsilon : 0,13$ ;  $\beta : -0,04$ ;  $Pendiente : 1,06$ ;  $Intercepto : -0,09$ . Finalmente, para el CDOM, las métricas fueron:  $MSA : 0,78$ ;  $\epsilon : 0,78$ ;  $\beta : 0,78$ ;  $Pendiente : 0,62$ ;  $Intercepto : 0,33$ . Este trabajo es uno de los primeros desarrollados a nivel nacional donde se comparan diferentes modelos basados en datos satelitales para la estimación de variables de calidad de agua. Se visualiza que los modelos locales capturan condiciones particulares de los sitios analizados que permiten mejorar las estimaciones y tener una confianza mayor sobre

las variables analizadas satelitalmente.

# Prospección del uso de imágenes de drones para el monitoreo detallado del pastizal nativo. Amparo Quiñones <sup>1\*</sup>, Andrés Hirigoyen <sup>1</sup>, Jean Savian <sup>1</sup>

<sup>1</sup> Instituto Nacional de Investigación Agropecuaria \*aquinones@inia.org.uy

La vegetación puede ser estudiada de forma precisa, frecuente y en grandes superficies gracias a la amplia disponibilidad de sensores y a los importantes avances en el procesamiento de datos mediante la Inteligencia Artificial. Es así como las imágenes satelitales se han transformado en un insumo esencial para la gestión de los Pastizales del Río de la Plata (p.ej. Paruelo et al. 2016, Baeza et al. 2022). El uso de drones para el monitoreo detallado del pastizal (o campo natural) permanece menos explorado. Los drones poseen una serie de ventajas que justifican su creciente uso a nivel mundial, como contar con cámaras de alta resolución (RGB, multi o hiperespectrales), permiten generar modelos digitales de terreno y vegetación, son flexibles en su operación, autónomos y soportan distintos sensores (de Castro et al. 2021, Lyu et al. 2022). En este marco, se busca incorporar las imágenes aéreas de alta resolución al monitoreo de la vegetación del experimento de largo plazo (ELP) “Manejo del pastoreo y su efecto en los servicios ecosistémicos del campo natural,” que integra la plataforma agroambiental de INIA. Nuestro propósito es complementar o sustituir los métodos tradicionales de relevamiento de la vegetación mediante el análisis de imágenes aéreas con técnicas de machine learning. Se espera reducir el tiempo y mano de obra destinados a la observación directa y aumentar la superficie de colecta de datos. El ELP comenzó en 2021 y está ubicado en la Unidad Experimental Palo a Pique de INIA Treinta y Tres. Los tratamientos reflejan un gradiente de intensidad de pastoreo, con alturas objetivo de 4, 8, 12 y 16 cm además de un tratamiento sin pastoreo (Savian et al. 2021). En 2021, la riqueza media fue de 42 especies/m<sup>2</sup> siendo las monocotiledóneas más abundantes y frecuentes *Paspalum notatum*, *Carex* sp., *Andropogon ternatus* y *Axonopus fissifolius* y las dicotiledóneas *Eryngium horridum* (cardilla) y *Baccharis trimera* (carqueja) (datos sin publicar). Tanto cardilla como carqueja son especies poco consumidas por el ganado y se excluyen de las mediciones de altura. Por tanto, estimar su área es prioritario para caracterizar los tratamientos. Consecuentemente, nuestro primer objetivo es estimar de forma precisa la cobertura de cardilla. La cardilla integra la familia Apiaceae, es una roseta perenne, de mediano-alto porte, sus hojas son lineales, largas y espinosas y crece formando manchones. Es una especie relevante en términos ecológicos y agronómicos y avanzar en su sensado remoto contribuirá al desarrollo de estrategias para su manejo sustentable (Quiñones et al. 2020). Esta planta tiene una fisionomía característica y se distingue fácilmente en imágenes de dron, lo que permite adoptar como marco de trabajo el Análisis de Imágenes Basado en Objetos Geográficos (GEOBIA por sus siglas en inglés) (Castilla y Hay 2008). El GEOBIA es empleado en el análisis de imágenes, especialmente en el mapeo de clases de uso en ambientes urbanos, forestales, agrícolas y zonas naturales. La unidad básica de clasificación son los objetos, definidos como grupos de píxeles espectralmente homogéneos y espacialmente continuos. Los principales pasos del GEOBIA son: i) segmentación, para lo que se usan mayormente algoritmos de crecimiento en regiones o de detección de bordes ii) extracción de atributos/características de los segmentos, que una vez dotados de información pasan a denominarse objetos iii) generación de un modelo de clasificación, usualmente random forest o support vector machines. Para la clasificación se utiliza información de la geometría, textura y contexto de los objetos (Kucharczyk et al. 2020). En un primer sondeo, en agosto del 2022 se realizó un vuelo con un dron comercial (Mavic Air 2, DJI) en una de las parcelas pastoreadas a 4 cm. El ortomosaico resultante (RGB, 1 ha de extensión), tuvo un tamaño de píxel de 0.7 cm. Para evaluar diferentes opciones de segmentación se usó el software QGIS y su extensión de OTB, particularmente el algoritmo mean-shift. Este algoritmo segmenta la imagen en regiones de alta densidad. Para ello, la imagen se representa como un espacio de características cuyas dimensiones se corresponden con la intensidad de los canales (p.ej. RGB) y donde cada píxel es un punto en dicho espacio. La segmentación comienza con un número inicial de puntos, a los que aplica una ventana de búsqueda y calcula su centro de masa. Cada punto se desplaza iterativamente hacia el centro hasta no poder moverse dentro de la ventana. Por último, los puntos que convergen al mismo centro se agrupan y forman un segmento. En una superficie reducida del ortomosaico probamos una serie de combinaciones de los 3 parámetros del algoritmo, logrando una segmentación que respeta la forma de la planta. Posteriormente, en los más de 20000 segmentos, se extrajo el valor de 12 atributos espectrales (media, máximo, mínimo, desvío estándar por banda) y un atributo de geometría (tamaño). Luego, se seleccionaron aleatoriamente 100 objetos que fueron clasificados como cardilla / no cardilla. Con esta información, y mediante el paquete caret de R, se entrenó un modelo random forest, que luego fue usado para establecer la clase de los restantes objetos. De la implementación de esta simple rutina se destacan 3 aspectos positivos. Primero, con pocos intentos de

parametrización se logró que los segmentos delimitaran la cardilla. Este paso puede optimizarse al ensayar nuevas combinaciones de valores en los parámetros del algoritmo. Segundo, el modelo de clasificación tuvo un desempeño aceptable, siendo la exactitud de 0,67 (proporción de aciertos en la predicción de los valores observados) y el coeficiente de Kappa de 0,20 (concordancia entre las clasificaciones del modelo y las etiquetas reales más allá de lo esperado por azar). Estas métricas pueden mejorar al incluir más datos de entrenamiento y nuevos atributos. Tercero, la inspección visual realizada por un experto de campo de las predicciones sobre el conjunto total de datos es satisfactoria. A futuro se planifica realizar nuevos vuelos en el sitio experimental y profundizar el desarrollo de las etapas de segmentación y clasificación. Resulta interesante evaluar como inciden en la clasificación de cardilla la calidad de la imagen (RGB vs multiespectral), la altura del pastizal, la estación del año y el tamaño de píxel.

Baeza S., Vélez-Martin E., De Abelleira D., Bancharo S., Gallego F. et al. (2022). Two decades of land cover mapping in the Río de la Plata grassland region: The MapBiomias Pampa initiative. *Remote Sensing Applications: Society and Environment*, 28, 100834.

Castilla G., Hay G.J. (2008). Image objects and geographic objects. In: Blaschke T., Lang S., Hay G.J. (Eds.) *Object-Based Image Analysis: Spatial Concepts for Knowledge-Driven Remote Sensing Applications*. Springer: Berlin. 91-110.

de Castro A. I., Shi Y., Maja J. M., Peña J. M. (2021). UAVs for vegetation monitoring: overview and recent scientific contributions. *Remote Sensing*, 13(11), 2139.

Kucharczyk M., Hay G. J., Ghaffarian S., Hugenholtz C. H. (2020). Geographic object-based image analysis: a primer and future directions. *Remote Sensing*, 12(12), 2012.

Lyu X., Li X., Dang D., Dou H., Wang K., Lou, A. (2022). Unmanned Aerial Vehicle (UAV) remote sensing in grassland ecosystem monitoring: a systematic review. *Remote Sensing*, 14(5), 1096.

Paruelo J.M., Texeira M., Staiano L., Mastrángelo M., Amdan L., Gallego F. (2016). An integrative index of ecosystem services provision based on remotely sensed data. *Ecological Indicators*, 71, 145-154.

Quiñones A., Lattanzi F., Saldain N., Lezama F. (2020). Weed control in natural grasslands: a case study using a perennial native forb from the South American Campos. *Austral Ecology*, 45(7), 909-920.

Savian J., Quiñones A., Pravia M., Cardozo G., Guido A. et al. (2021). Un nuevo experimento de largo plazo en INIA Treinta y Tres con foco en la sostenibilidad del campo natural. *Revista INIA*, 67. INIA: Montevideo. 29-31.

# Análisis preliminares de datos obtenidos con escáner láser terrestre para caracterización de la estructura de un bosque nativo ribereño en gradientes ambientales. Andrea Cardoso<sup>1\*</sup>, Carolina Crisci<sup>1</sup>, Nestor Mazzeo<sup>2</sup>, Gonzalo Perera<sup>1</sup>, Santiago Verón<sup>3</sup>

<sup>1</sup> Departamento de Modelización estadística de datos e inteligencia artificial, Centro Universitario Regional este, Universidad de la República. Rocha, Uruguay.

<sup>2</sup> Departamento de Ecología y rehabilitación de sistemas acuáticos, Centro Universitario Regional Este, Universidad de la República. Maldonado, Uruguay.

<sup>3</sup> Instituto Clima y Agua, Instituto Nacional de Tecnología Agropecuaria. Buenos Aires, Argentina

\* andrea10del85@gmail.com

Resumen El escáner láser terrestre (TLS por sus siglas en inglés) es una herramienta de teledetección terrestre que utiliza tecnología LIDAR (Ligth detecting and ranging) y a través de la emisión de un haz láser mide las distancias y almacena información de todos los puntos que toca. La aplicación del TLS en los muestreos de bosque abre la posibilidad estudiar la estructura de los árboles que lo componen con un detalle y una precisión que mediante los métodos tradicionales sería imposible medir (Disney 2019). Algunas aplicaciones interesantes del muestreo con TLS en bosque nativo se relacionan con el poder de extracción de información morfológica de los árboles de forma no destructiva, la posibilidad de modelar la arquitectura de los árboles (a nivel de troncos y ramas, dimensiones de las copas e interacciones entre copas vecinas), relevar la ubicación relativa de cada árbol en el bosque, distinguir entre la biomasa leñosa de la foliar y cuantificar de forma precisa el volumen de la biomasa aérea de los árboles (Dassot et al. 2011, Disney et al. 2019, Mummbe et al. 2021). El análisis de las características morfológicas o rasgos de los árboles en los diferentes gradientes ambientales es relevante para el conocimiento de las funciones ecosistémicas del bosque (Malhi et al. 2018, O’Sullivan et al. 2021, Maynard et al. 2022). Además de grandes ventajas, la aplicación del TLS sobre bosques nativos como los nuestros con alta densidad de plantas, múltiples especies y formas de crecimiento, presenta desafíos. Estos se centran principalmente en la etapa de procesamiento y análisis, por el gran volumen de datos que se generan (por ejemplo manejo de nubes de puntos de 3-5 gb de tamaño) y los tiempos de procesamiento (Liang et al. 2016). En este sentido mi proyecto de doctorado tiene como objetivo principal la caracterización de la estructura del bosque ribereño de la cuenca alta del río Negro en diferentes gradientes ambientales y su relación con una función ecosistémica fundamental como es la reserva de carbono en la biomasa aérea leñosa. Se aplica para su abordaje la teoría ecológica basada en rasgos (TDT) que provee herramientas para analizar como los cambios en el ambiente van a causar cambios en la distribución de los rasgos y como esas variaciones en los rasgos afectan las funciones ecosistémicas (Enquist et al. 2015, de Bello et al. 2021). Particularmente en las Jornadas presentaré los resultados preliminares del procesamiento y análisis de los rasgos de árboles en el marco de la TDT del bosque ribereño de la cuenca alta del río Negro obtenidos con TLS en diferentes gradientes ambientales (objetivo 3 de la tesis). El bosque ribereño de la cuenca alta del río Negro se encuentra ubicado al noreste del Uruguay, cubriendo una superficie total de bosque de 43954 has. Para su muestreo se utilizó como base un diseño sistemático que definió 30 sitios distribuidos a lo largo del bosque (MGAP, 2022). Sobre estos se identificaron 10 sitios que presentaban buenas condiciones para el acceso con el equipo, abarcaban la diversidad de especies arbóreas existentes a lo largo del bosque y presentaban características contrastantes respecto a los gradientes ambientales. Se analizaron los gradientes ambientales según la distancia al cuerpo de agua principal, el tipo de manejo del bosque y los usos del suelo aledaño al bosque. Se desarrollaron un total de 7 salidas de campo, destinando un día para cada sitio de muestreo en el que se realizaban tres parcelas (de 6 m de radio cada una) distribuidas en una perpendicular al cuerpo de agua principal.

La captura de los datos en campo se realizó con un TLS del tipo FARO focus S150, de retorno simple y alcance de hasta 150 m. Se realizaron múltiples escaneos por parcela y se usaron esferas de referencia para luego unir los escaneos relevando completamente los árboles de la parcela. El procesamiento de los datos capturados con TLS es semiautomático para cada una de las parcelas de estudio. Los pasos generales son 1) pre procesamiento y co-registración los escaneos, 2) clasificación en nube de puntos del suelo y de vegetación, 3) segmentación de nubes de árboles completos, 4) clasificación de nube de puntos de madera y de hojas para cada árbol, 5) obtención de un modelo cuantitativo estructural (QSM) por árbol a partir del que se calculan los rasgos. Algunos ejemplos de rasgos son diámetro a la altura del pecho (DBH), altura, largo de tronco

principal, diámetro y altura de copa, radios y largos de ramas de n orden. A pesar que existe evidencia sobre la mayor precisión de medición de arboles con TLS respecto de los muestreos tradicionales en diferentes bosques a nivel mundial (Disney 2019), el uso de esta herramienta en nuestros bosques nativos es reciente y no cuenta con resultados anteriores que permitan su aplicación directa. En este sentido, se realizó la evaluación de los QSM obtenidos para los árboles a través del DBH. Para un conjunto de árboles medidos a campo y de nubes de árboles que pudieron ser homologadas se cuantificó un error cuadrático medio entre el DBH obtenido de ambas formas. Cabe mencionar que este tipo de procesamiento de datos para obtener resultados a nivel de árbol (“grano de estudio”) se requiere de aproximadamente 3 días de trabajo con computador potente por parcela. Finalmente, se presentarán resultados preliminares de distribuciones de algunos rasgos de arboles en los diferentes gradientes ambientales, esperando identificar diferencias en función de los mismos.

Dassot, Mathieu, Thiéry Constant, y Meriem Fournier. «The Use of Terrestrial LiDAR Technology in Forest Science: Application Fields, Benefits and Challenges». *Annals of Forest Science* 68, n.o 5 (agosto de 2011): 959-74. <https://doi.org/10.1007/s13595-011-0102-2>.

De Bello, F. Carmona C. Dias A. Götzenberger L. Moretti M. Berg M. 2021. *Handbook of Trait-Based Ecology. From Theory to R tools*. Cambridge University Press. p. 293. DOI: 10.1017/9781108628426.  
Disney M. 2019. Terrestrial LiDAR: a three-dimensional revolution in how we look at trees. *New Phytologist* 222: 1736–1741

Disney, M., A. Burt, K. Calders, C. Schaaf, y A. Stovall. 2019. Innovations in Ground and Airborne Technologies as Reference and for Training and Validation: Terrestrial Laser Scanning (TLS). *Surveys in Geophysics* 40, n.o 4: 937-58. <https://doi.org/10.1007/s10712-019-09527-x>.

Enquist, Brian J., Jon Norberg, Stephen P. Bonser, Cyrille Violle, Colleen T. Webb, Amanda Henderson, Lindsey L. Sloat, y Van M. Savage. 2015. Scaling from Traits to Ecosystems. En *Advances in Ecological Research*, 52:249-318. <https://doi.org/10.1016/bs.aecr.2015.02.001>.

Malhi, Yadvinder, Tobias Jackson, Lisa Patrick Bentley, Alvaro Lau, Alexander Shenkin, Martin Herold, Kim Calders, Harm Bartholomeus, y Mathias I. Disney. 2018. New perspectives on the ecology of tree structure and tree communities through terrestrial laser scanning. *Interface Focus* 8, n.o 2: 20170052.

Maynard, Daniel S., Lalasia Bialic-Murphy, Constantin M. Zohner, Colin Averill, Johan Van Den Hoogen, Haozhi Ma, Lidong Mo, et al. 2022. Global Relationships in Tree Functional Traits. *Nature Communications* 13, n.o 1: 3185. <https://doi.org/10.1038/s41467-022-30888-2>.

MGAP. 2022. Iniciativa para el Rio Negro. Fortalecimiento de las capacidades de gestión para la protección del bosque nativo de la cuenca. Informe de Dirección General Forestal. p. 119.

Liang, Xinlian, Ville Kankare, Juha Hyyppä, Yunsheng Wang, Antero Kukko, Henrik Haggrén, Xiaowei Yu, et al. 2016. Terrestrial Laser Scanning in Forest Inventories. *ISPRS Journal of Photogrammetry and Remote Sensing* 115:63-77. <https://doi.org/10.1016/j.isprsjprs.2016.01.006>.

O’Sullivan, Hannah, Pasi Raunonen, Pekka Kaitaniemi, Jari Perttunen, y Risto Sievänen. 2021. Integrating terrestrial laser scanning with functional-structural plant models to investigate ecological and evolutionary processes of forest communities. *Annals of Botany* 128, n.o 6: 663-84.

Muumbe, Tasiyiwa Priscilla, Jussi Baade, Jenia Singh, Christiane Schullius, y Christian Thau. 2021. Terrestrial Laser Scanning for Vegetation Analyses with a Special Focus on Savannas. *Remote Sensing* 13, n.o 3: 507. <https://doi.org/10.3390/rs13030507>.

# Una propuesta de modelización para predecir la distribución del Caracol Negro (*Pachycymbiola brasiliana*) como insumo para una pesquería artesanal sustentable. Ignacio Santillán<sup>1</sup>, Ximena Lagos<sup>2</sup>, Fabrizio Scarabino<sup>3</sup>, Angel M. Segura<sup>1</sup>

<sup>1</sup> Modelización Estadística de Datos e Inteligencia Artificial (MEDIA) CURE, UdelaR.

<sup>2</sup> Centro Interdisciplinario para el Manejo Costero Integrado del Cono Sur. Centro Universitario de la Región Este (CURE-UdelaR)

<sup>3</sup> Centro Universitario de la Región Este (CURE-UdelaR), Sede Rocha, Uruguay.

\* rignaciosantillan@gmail.com

A nivel mundial, la gestión pesquera basada únicamente en información científica ha sido insuficiente para evitar el colapso de numerosos stocks pesqueros. El colapso de las pesquerías es aún más relevante en zonas donde la información científica es escasa y fragmentada. Esto último aplica a la mayoría de las especies bentónicas de Uruguay, donde la actividad pesquera se desarrolló históricamente con un limitado conocimiento de las características intrínsecas de dichos recursos. Entre los recursos no tradicionales explotados en Uruguay, se destacó en las costas de La Paloma (Rocha), entre 1991 y 1998, la pesca de caracol negro (*Pachycymbiola brasiliana*), un gasterópodo de gran tamaño. La pesquería del caracol negro siguió una dinámica de “boom and bust”: su extracción comenzó de forma abrupta, sin estudios previos y colapsó a los pocos años mostrando signos de sobreexplotación. En la actualidad, se observan indicios de recuperación en las agregaciones de caracol negro en las costas de La Paloma. Además, debido al interés de comerciantes y pescadores existe la potencialidad de retomar la pesquería. Es por esto que para garantizar una extracción sostenible es necesario generar información precisa sobre la distribución de la especie y establecer recomendaciones para una potencial pesquería artesanal sustentable del caracol negro. Ante la falta de datos empíricos sobre la distribución del caracol negro, intentar predecir su presencia utilizando algoritmos automatizados puede generar información objetiva y replicable. En este sentido, la modelización de la distribución de especies (MDE) comprende un conjunto de herramientas y técnicas estadísticas que permiten predecir la distribución geográfica y ambiental de una especie en función de variables explicativas. La MDE implica la construcción de algoritmos que relacionan diferentes variables con registros de presencia/ausencia (P/A). En este contexto, el presente estudio se enfoca en la construcción de modelos predictivos para la presencia del caracol negro (*P. brasiliana*) en las costas de La Paloma. La base de datos utilizada abarca desde 1972 hasta 2016 ( $n = 408$ ) y fue compilada a partir de información de campañas de investigación recopiladas en publicaciones previas. Debido al desbalance en los datos,  $P = 81(19, 85)$  y  $A = 327(80.15)$ , se aplicó la técnica Synthetic Minority Over-sampling Technique (SMOTE) para crear una base de datos balanceada y se comparó con resultados de aplicar los algoritmos sin el balanceo de la muestra de entrenamiento. Además se generaron modelos con todas las variables y un conjunto solo con variables de fácil obtención (remotas, ej. sedimentos, SST satelital). Las variables explicativas incluyeron la ubicación geográfica, condiciones ambientales (temperatura superficial y profundidad) y esfuerzo de captura. Las variables con datos faltantes fueron interpoladas mediante el algoritmo rImpute del paquete Randomforest de R. La profundidad fue extraída de las cartas náuticas digitales ROU3 (SOHMA, 1997) y ROU19 (SOHMA, 2012); los datos de sedimentos marinos fueron extraídos a partir de datos espaciales de sedimentos de fondo (DINACEA, 2017) y además se interpolaron datos de sedimentos a partir de la carta náutica ROU19 mediante rImpute. Por último, se extrajeron datos diarios y mensuales de temperatura superficial del mar satelital (MUR-GHRSST) con resolución espacial de  $0,01^\circ$  a partir de 2002. Para los datos previos y faltantes se realizó una interpolación mediante rImpute. Se entrenaron con un subconjunto de datos de aprendizaje (75) tomados de forma aleatoria modelos lineales Generalizados (GLM) y de Machine Learning: Árboles de Clasificación y Regresión (CART), AdaBoost (AB), Supported Vector Machine (SVM) y RF. Sobre este último se aplicaron 3 variantes: RF con pesos diferenciales entre clases (RFwc), con punto de corte (RFcf) y con remuestreo estratificado (RFst) para lidiar con el desbalance. Se evaluó la capacidad predictiva con datos de prueba (25%). La capacidad de predicción de cada modelo se evaluó utilizando: Precisión (ACC), Tasa verdaderos positivos (TPR), Tasa de Falsos positivos (FPR) y Área bajo la curva (AUC). El modelo RF alcanzó la mayor precisión de predicción tanto en los modelos completos, así como en los modelos remotos, con un promedio del 82.62% y 82.04% respectivamente. Mientras que el modelo RF entrenado con todas las variables presentó el  $FPR = 6.27\%$  y RFCF entrenado en base a datos balanceada presentó un  $FPR = 17.38\%$  para los modelos con variables

remotas. Mientras que SVM entrenado a partir de la base de datos balanceada presentó el mejor TPR tanto para los modelos completos como para los modelos remotos, 87.46% y 90.06% respectivamente. Por otra parte, el modelo RFst entrenado a partir de la base de datos balanceada presentó el mayor valor de AUC tanto para los modelos completos, así como para los modelos remotos con un promedio de 0.72. Por último, los modelos con menor precisión son los modelos GLM y CART para ambas bases de datos y para los modelos completos y remotos. En conjunto, este estudio presenta modelos predictivos valiosos que pueden apoyar la gestión pesquera y la toma de decisiones. En síntesis, desarrollar una modelización de la distribución de caracol negro puede ser de utilidad para la gestión de este recurso y constituye una metodología objetiva que puede ser replicada en el estudio de otros recursos pesqueros de interés para el país. En instancias posteriores se integrará el conocimiento experto de investigadores y gestores utilizando técnicas de métodos de meta-learning que permiten integrar información de diferentes fuentes. Estos algoritmos se utilizarán para generar una aproximación a la distribución del caracol negro en las costas de La Paloma (Rocha) y en las costas de Uruguay. El enfoque interdisciplinario y transdisciplinario que integra el machine learning y el conocimiento ecológico local es una estrategia innovadora y participativa que permite avanzar hacia el desarrollo de pesquerías sostenibles.

## Aplicando Machine Learning para predecir la condición reproductiva de peces. Andrés Flores<sup>1</sup> , Rodrigo Wiff<sup>2,3</sup>, Carl R. Donovan<sup>4,5</sup> , Patricio Gálvez<sup>6</sup>

<sup>1\*</sup> Investigador independiente, Viña del Mar, Chile

<sup>2</sup> Center of Applied Ecology and Sustainability (CAPES-UC), Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>3</sup> Instituto Milenio en Socio-Ecología Costera (SECOS), Chile

<sup>4</sup> DMP Statistical Solutions UK Ltd, St. Andrews, Scotland

<sup>5</sup> School of Mathematics and Statistics, University of St Andrews, Scotland

<sup>6</sup> División de Investigación Pesquera, Instituto de Fomento Pesquero, Valparaíso, Chile \* andflores378@gmail.com

El conocimiento de los rasgos reproductivos de las poblaciones marinas explotadas es crucial para su gestión y conservación. El estado de madurez de los peces suele determinarse mediante métodos tradicionales como la macroscopía y la histología. El análisis macroscópico consiste en evaluar, a ojo desnudo, las características externas de las gónadas para asignar un estado de madurez, pero suele introducir bastantes errores por ser subjetiva, aún con observadores entrenados. Por el contrario, la histología es el método más preciso dado que analiza la condición de madurez a nivel celular, pero toma tiempo, costoso y necesita laboratorios especializados, la cual generalmente no está disponible para muchos stocks de peces alrededor del mundo. En este trabajo utilizamos el método de aprendizaje automático desde Random Forest (RF) para clasificar el estado reproductivo de los peces, utilizando como ejemplo los amplios datos de la merluza común (*Merluccius gayi gayi*) capturada en la zona centro-sur de Chile. Las gónadas colectadas aleatoriamente desde operaciones de pesca comercial y cruceros de investigación fueron clasificadas en tres clases de madurez: inmadura, madura-activa y madura-inactiva, basándose en la macroscopía y la histología. Se ajustó un clasificador para estas tres clases de madurez mediante RF, con las covariables continuas longitud total, índice gonadosomático, factor de condición, latitud, longitud y profundidad, junto con el mes como variable categórica. Finalmente, se comparó la condición reproductiva y la estimación de la proporción madura a la talla (*i.e.*, ojiva de madurez) entre los tres métodos. El modelo RF mostró una alta precisión (> 82%), con un nivel de concordancia mayor que 71% y un error inferior que 15% en comparación con la histología. Asimismo, la condición reproductiva a la longitud y las variaciones mensuales fueron similares entre RF y la histología. La ojiva de madurez no varió significativamente entre histología y RF ( $p < 0.05$ ). Además, los estadios de madurez de las hembras fueron predichos con precisión por RF en comparación a la macroscopía. La aplicación del RF proporciona una herramienta muy promisoría para clasificar los estadios de madurez cuando se dispone de covariables del proceso reproductivo, permitiendo mejorar retrospectivamente las predicciones cuando sólo se dispone de la clasificación macroscópica. Además, esta herramienta también puede ser usada como un complemento a la histología en peces cuando el tamaño de la muestra es limitada.

# Revisión de métodos estadísticos y métricas para predecir contaminación fecal en playas recreativas. Victoria Vidal <sup>1\*</sup>, Lia Sampognaro<sup>1</sup>, Angel Segura<sup>1</sup>, Fernanda de León<sup>1</sup>, Carla Kruk<sup>1</sup>, Gonzalo Perera<sup>1</sup>, Carolina Crisci<sup>1</sup>

<sup>1</sup> Departamento Modelización Estadística de Datos e Inteligencia Artificial (MEDIA), CURE-Rocha, Universidad de la República, Ruta Nacional N°9 intersección Ruta N°15, Rocha, PC 27000, Uruguay  
\*victoria.vidal@cure.edu.uy

La contaminación por coliformes fecales (CCF) es uno de los principales y más frecuentes problemas ambientales de las zonas costeras (Mallin et al. 2000). La CCF varía ampliamente afectando la calidad de las playas y por lo tanto el uso recreativo de dichos espacios (He and He, 2008; Molina et al. 2014) lo que genera un riesgo para la salud humana, por el aumento de la probabilidad de contraer enfermedades (Hlavsa et al. 2011; Sabino et al. 2014). El desarrollo de modelos predictivos de CCF surge como un complemento de los sistemas de monitoreo tradicionales y su implementación es recomendada para la gestión de la calidad de playas recreativas (USEPA, 2012; OMS, 2018). En algunos casos, los sistemas de alerta temprana involucran programas de notificación al público que incluyen avisos emitidos en páginas web, aplicaciones para teléfonos móviles, y campañas de prevención y alerta (Searcy et al. 2018). Estos componentes de comunicación aportan a la buena gestión y se utilizan para proveer a los usuarios información accesible y fácilmente interpretable. El objetivo de este estudio fue 1) Realizar una revisión bibliográfica de los modelos estadísticos utilizados para la predicción de la calidad de agua en playas de uso recreativo y 2) Evaluar la implementación operativa de modelos estadísticos en sistemas de alerta temprana a nivel global. Se realizó una revisión de artículos académicos con una ventana temporal desde el 2000 a 2023. La búsqueda fue realizada en portales de acceso público (ej. Google Académico, Portal Timbó). Los dominios de este estudio se restringieron al agua superficial de playas recreativas, ya sean costeras, estuarinas o continentales. Se analizaron las métricas de capacidad predictiva con foco en aquellos artículos que presentaron las mismas calculadas sobre datos independientes de los utilizados para el entrenamiento de los modelos. Se recabó y sistematizó la información acerca del país de origen de las publicaciones, el tipo de ambiente estudiado (marino-costero, río, arroyo, estuarino), el año, estrategia de muestreo (escalas espacio-temporales analizadas) y de modelización (tipo de modelo, variables explicativas utilizadas, entre otras). Se realizó además una búsqueda de las distintas interfaces web utilizadas para notificar al público sobre las predicciones (utilizando modelos estadísticos) de CCF en playas recreativas. Se seleccionaron los programas de alerta que tuvieran modelos predictivos operativos. En este trabajo se registraron 65 artículos que incluyeron modelos desarrollados en 10 países de 4 continentes. La mayor parte de los artículos se concentraron en Norte América. Los ambientes de aguas continentales (lagos, lagunas, ríos y arroyos) fueron mayor representados (67%) que los sistemas marinos y estuarinos (33%). Se extrajeron métricas de 38 artículos de los cuales 14 fueron de regresión pero categorizando las predicciones (primero ajustan un modelo de regresión y luego se clasifica la salida del modelo en categorías (ej. excede o no excede el umbral de CCF permitido por la normativa)), 13 utilizaron sólo modelos de regresión (intenta predecir una variable continua,  $Y$ =concentración de coliformes), 4 fueron modelos de clasificación (intenta predecir una categoría de calidad de agua, ej. Playa abierta o Cerrada) y 7 utilizaron modelos de regresión y clasificación por separado. Los modelos de Regresión Lineal Múltiple (RLM) fueron los más utilizados durante todo el período considerado, seguido de las Redes Neuronales Artificiales (ANNs). Hacia el final del periodo analizado (2015-2023) comenzaron a registrarse otras estrategias de predicción basadas en algoritmos de Aprendizaje Automático (Machine Learning). Dentro de este grupo, los más utilizados en el periodo 2015 -2023 fueron los Árboles de Clasificación y Regresión (CART; Breiman et al. 1984) y los Bosques Aleatorios (Random Forest, RF, Breiman et al. 1984). La variable de respuesta  $Y$  (ya sea *Escherichia Coli*, *Enterococcus*, Coliformes fecales, totales o termotolerantes) en la mayoría de los casos fue transformada a escala logarítmica. Entre las métricas utilizadas para la evaluación del desempeño de modelos estadísticos de regresión estuvo el coeficiente de determinación ( $R^2$ ) para la RLM y la Raíz del Error Cuadrático Medio (RMSE) en general. En modelos de clasificación, se reportó la Exactitud (i.e. Accuracy), la Especificidad (Tasa de verdaderos negativos:  $TN/TN + FP$ ) y la Sensibilidad (Tasa de verdaderos positivos:  $TP/TP + FN$ ). Para los modelos de regresión las RLM presentaron un  $R^2$  ajustado promedio de 0,4 ( $min - max = 0,1 - 0,8$ ), una RMSE para el logaritmo de (CF) de 1,3(0,4 - 1,9). Para las ANNs, el  $R^2$  ajustado fue de promedio 0,7 ( $min - max = 0,04 - 0,9$ ), la RMSE 1,5 ( $min - max = 0,08 - 8,2$ ). Respecto a los modelos CART y RF presentaron una RMSE de 0,4 ( $min - max = 0,3 - 0,7$ ). Dentro de los modelos de clasificación

las ANNs presentaron una Exactitud de promedio 90,3 ( $min - max = 85,4 - 92,9$ ), Especificidad 95,2 ( $min - max = 90,0 - 99,5$ ), Sensibilidad 65,8 ( $min - max = 53,9 - 79,1$ ). Los modelos CART y RF de clasificación presentaron un Exactitud promedio de 86,0 ( $min - max = 73,0 - 93,0$ ), Especificidad 90,2 ( $min-max=71,0 - 98,0$ ) y Sensibilidad de 56,3 ( $min - max = 24,0 - 85,7$ ). Las variables explicativas más utilizadas para el desarrollo de los modelos fueron la temperatura del agua y del aire, la lluvia acumulada (24, 48, 72hs), tanto para ambientes costeros como continentales. La dirección e intensidad del viento, la turbidez del agua, la altura de ola y la salinidad también fueron importantes para ambientes costero-marinos. Se registraron un total de 10 interfaces web operativas en 6 países que actualmente utilizan modelos predictivos como herramientas para los sistemas de alerta temprana de calidad de playas. Cinco de ellos fueron llevados a cabo por instituciones gubernamentales a cargo de la gestión de la calidad ambiental, en coordinación con la academia. En otros casos, los modelos y las páginas web se crearon conjuntamente con organizaciones no gubernamentales o por iniciativas privadas. La inclusión de modelos estadísticos predictivos como parte de sistema de alerta temprana de calidad de playas fue menor en comparación con la variedad de modelos desarrollados en la literatura académica. La implementación de modelos predictivos, junto con interfaces web adecuadas y amigables, mejora el acceso y conocimiento de la calidad ambiental de las playas al público.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*/ Leo Breiman. . . [et al.]. Belmont, Calif: Wadsworth International Group, Belmont, Calif.

He, L. M. L., & He, Z. L. (2008). Water quality prediction of marine recreational beaches receiving watershed baseflow and stormwater runoff in southern California, USA. *Water research*, 42(10-11), 2563-2573.

Hlavsa, M. C., Roberts, V. A., &erson, A. R., Hill, V. R., Kahler, A. M., Orr, M., & Yoder, J. S. (2011). Surveillance for waterborne disease outbreaks & other health events associated with recreational water—United States, 2007–2008. *MMWR Surveill Summ*, 60(12), 1-32.

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(4), 1047-1056.

Molina, M., Hunter, S., Cyterski, M., Peed, L. A., Kelty, C. A., Sivaganesan, M., & 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. Organización Mundial de la Salud (OMS) (2018). WHO Recommendations on Scientific, Analytical & Epidemiological Developments Relevant to the Parameters for Bathing Water Quality in the Bathing Water Directive (2006/7/EC). 2018.

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 Br&ãõ, J., (2014). Routine screening of harmful microorganisms in beach s&s: implications to public health. En: *Science of The Total Environment*, 472, pp.1062–1069. doi:10.1016/j.scitotenv.2013.11.091

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.

USEPA, (2012). *Recreational Water Quality Criteria*. (No. OFFICE OF WATER 820-F-12- 058).

## Comparación de algoritmos de inteligencia artificial para la predicción de concentración de partículas finas en diferentes usos del suelo a partir de variables satelitales. Rodríguez Nuñez, Martín<sup>1,2,3\*</sup>, Mónica Balzarini<sup>2,3</sup>

<sup>1</sup> Facultad de Ciencias Agropecuarias, Departamento de estadística y biometría. Universidad Nacional de Córdoba (UNC), Argentina. <sup>2</sup> Facultad de Ciencias Exactas, Físicas y Naturales. Departamento de Química General. UNC, Argentina. <sup>3</sup> Consejo Nacional de Investigaciones Científicas y Técnicas. Córdoba, Argentina. \*martinrnu@gmail.com

La Organización Mundial de la Salud (OMS) estima que nueve de cada diez personas respiran aire con altos niveles de contaminante y como resultado siete millones mueren cada año en el mundo por causas relacionadas. El material particulado fino, partículas con un diámetro menor o igual a 2.5 micrómetros ( $PM_{2.5}$ ), es considerado uno de los contaminantes atmosféricos de mayor nocividad dada su capacidad de penetrar la barrera pulmonar y pasar al torrente sanguíneo, causando enfermedades cardiovasculares, respiratorias y cáncer. La presencia de estas partículas en la atmósfera tiene un impacto significativo en diversos aspectos que influyen en la vida humana, abarcando desde la salud hasta los ecosistemas y el clima. Este estudio explora las capacidades predictivas de diversos algoritmos de aprendizaje automático y profundo que utilizan combinaciones de variables satelitales para pronosticar la concentración de  $PM_{2.5}$  en áreas con distinto uso de suelo en la provincia de Córdoba, Argentina. Los datos de  $PM_{2.5}$  fueron recopilados mediante sensores de bajo costo ubicados en sitios con usos de suelo urbano y agrícola/área verde. Las concentraciones medidas de  $PM_{2.5}$  en la provincia de Córdoba durante el periodo de estudio oscilaron entre 0.53 (mínimo) y 95.28 (máximo), con una concentración promedio de 13.10 para el uso de suelo urbano y de 6.78 para el agrícola/área verde. Posteriormente, se descargaron múltiples variables satelitales con el fin de evaluar su efectividad al captar la variabilidad de la concentración de  $PM_{2.5}$ . Una vez identificadas las variables más relevantes para cada uno de los usos de suelo, se procedió a evaluar algoritmos predictivos pertenecientes al campo de la inteligencia artificial. En el contexto de un uso de suelo urbano, los resultados del estudio revelaron que la red neuronal recurrente apilada que incorpora unidades recurrentes con compuerta (GRU) fue el algoritmo de mayor capacidad, demostrando un rendimiento superior tanto en los conjuntos de datos de entrenamiento como en los de testeo. Los resultados durante el entrenamiento, obtenidos por medio de una validación cruzada de 10 veces, mostraron una raíz del error cuadrático medio (RMSE) de 1.63 y un error absoluto medio (MAE) de 1.24. Mientras que los resultados durante el testeo revelaron un RMSE de 3.23 y un MAE de 2.20. Por otra parte, para el uso de suelo agrícola/área verde, los resultados demostraron que el algoritmo de mayor capacidad predictiva fue el de bosques aleatorios. Los resultados obtenidos bajo la base de datos de entrenamiento fueron un RMSE de 3.23 y un MAE de 2.20. Mientras que bajo una base de datos de testeo fueron un RMSE de 2.54 y un MAE de 1.70. Los hallazgos de este estudio sugieren las características propias de la variabilidad de cada una de las series temporales en los distintos usos de suelo y de aprendizaje de cada uno de los algoritmos predictivos ajustados, son determinantes en el tipo de algoritmo que demuestra mayor exactitud para cada uno de los usos de suelo. Además, que la integración de datos satelitales y algoritmos de inteligencia artificial puede constituir una valiosa herramienta para la predicción de la concentración de  $PM_{2.5}$ .

## **Detección de cambios en mamografías y evolución de calcificaciones mamarias. Gabriel Wolfmann<sup>1</sup>, Ma. Lucrecia Ballarino<sup>2</sup>, Valeria S. Rulloni<sup>3\*</sup>**

<sup>1</sup> FCEFyN, Universidad Nacional de Córdoba, Argentina <sup>2</sup> Instituto Oulton, Córdoba, Argentina <sup>3</sup> FCEFyN, Universidad Nacional de Córdoba, Argentina \*vrulloni@unc.edu.ar

Las calcificaciones mamarias son un hallazgo frecuente y relevante en las mamografías. La mayoría de ellas tienen un origen benigno y en su mayoría no requieren ser controladas. Sin embargo, existen otras de origen maligno, como las microcalcificaciones, que pueden presentarse agrupadas, tener un aspecto sospechoso, un tamaño mas pequeño y que se deba a una lesión de alto riesgo, debiendo ser estrictamente controladas. El análisis temporal de las microcalcificaciones, independientemente de su origen, le da un indicio al personal de la salud sobre el avance de alguna lesión como puede ser un cáncer de mama. Para lograr una detección temprana en su desarrollo, es de suma importancia que las mujeres a partir de aproximadamente los 40 años, se realicen un estudio mamográfico anual. Debido a la falta de métodos objetivos y cuantitativos en la evaluación temporal de imágenes mamográficas, se verifica la importancia de crear una herramienta que permita la cuantificación de microcalcificaciones, que contribuyan al equipo médico a tomar una decisión. La realización del proyecto se resume en cinco etapas:

- i) Armado de base de mamografias de pacientes con microcalcificaciones. Por cada paciente se cuenta con imágenes en distintas etapas evolutivas de la manifestación de microcalcificaciones mamarias.
- ii) Construcción de un algoritmo de segmentación de las imágenes para obtener la región de interés (eliminación de zonas irrelevantes en el análisis: músculo pectoral y piel).
- iii) Detección de microcalcificaciones mediante técnicas de procesamiento de imágenes basadas en niveles de brillo.
- iv) Utilización de algoritmos no supervisados de agrupamiento.
- v) Análisis del cambio en el tiempo de los estudios mamográficos.

Los algoritmos con los que se han trabajado fueron desarrollado a partir del lenguaje de programación Python. Se alcanzaron muy buenos resultados en el algoritmo de segmentación semi automático de la mama, recorte de fondo, extracción de la piel y del músculo pectoral. Se logró obtener una buena segmentación de las microcalcificaciones, para su posterior evaluación. Esto se debe a la aplicación de distintos métodos utilizados como el filtro de área o excentricidad. Sin embargo, el gran avance estuvo cuando se aplicó el método de clustering para encontrar grupos de microcalcificaciones, eliminando cualquier ruido de falsos positivos, que se obtenía previo a dicha aplicación, como son las microcalcificaciones aisladas o las macrocalcificaciones. Se desarrolló un algoritmo de comparación entre imágenes sin la necesidad de usar métodos de corrección. Se logró resolver el problema mediante técnicas de agrupamiento. Se destaca por lo novedoso, ya que es una idea simple pero original que reemplaza el corrección de mamas. Finalmente, si bien es necesario un ajuste fino para las particularidades de cada paciente, se logró obtener una herramienta computacional para apoyo al personal de salud, en pos de la detección temprana de posibles síntomas cancerígenos. Esta herramienta devuelve la serie de imágenes con las microcalcificaciones agrupadas y dos archivos en formato CSV con los datos recabados de las agrupaciones, la imagen original, la imagen con las microcalcificaciones encontradas y el análisis en el tiempo. Mostraremos los resultados para algunos pacientes.

# Consistency of Nadaraya Watson estimator in non parametric regression model driven by fractional Brownian motion and sampled at random times. Tania Roa<sup>1\*</sup>, Meryem Slaoui<sup>1</sup>, Soledad Torres<sup>2</sup>

<sup>1\*</sup>Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez <sup>2</sup> CIMFAV - Facultad de Ingeniería. Universidad de Valparaíso \* tania.roa@uai.cl

In this paper, we study the consistency of the Nadaraya-Watson estimator, which, by choosing a kernel function, allows us to obtain an estimator that does not depend on the parameters to be estimated, but adapts to the shape of the function. For our work, we study a nonlinear regression model where observations are indexed in time according to a strictly increasing sequence of random variables defined by a renewal process. We prove consistency of the Nadaraya - Watson estimator. To illustrate the theoretical results obtained, a simulation study is presented to see how the proposed estimator behaves for different nonlinear functions and number of observations.

# Modelización de la madurez sexual de la raya *Sympterygia acuta* de un Área Protegida de la costa Atlántica de Uruguay: Cerro Verde e Islas de la Coronilla.

Ines Pereyra<sup>1\*</sup>, Cristina Oddone<sup>2</sup>, Angel Segura<sup>3</sup>

<sup>1</sup> Unidad de Gestión Pesquera Atlántica, Dirección Nacional de Recursos Acuáticos, La Paloma, Rocha, Uruguay. <sup>2</sup> Laboratório de Pesquisa em Chondrichthyes, Setor de Morfologia, Instituto de Ciências Biológicas, Universidade Federal do Rio Grande, Rio Grande-RS, Brasil. <sup>3</sup> Departamento de Modelización Estadística de Datos e Inteligencia Artificial, Centro Universitario Regional del Este (MEDIA-CURE), Rocha, Uruguay. \*ines.pereyra@gmail.com

*Sympterygia acuta* es una especie de raya endémica del Océano Atlántico Sudoccidental, que se distribuye desde Río de Janeiro, Brasil (22°S) hasta el Golfo de San Jorge, Argentina (45°43'S). En su rango de distribución es capturada incidentalmente por pesquerías industriales de arrastre multiespecíficas y dirigidas al langostino (*Pleoticus muelleri*). En Punta del Diablo (Rocha, Uruguay) es capturada incidentalmente por la pesquería artesanal de arrastre de fondo dirigida al langostino (setiembre-diciembre). Parte de la zona de operación de esta pesquería se encuentra dentro del Área Marina Protegida (AMP) de Cerro Verde e Islas de La Coronilla. A pesar de la relevancia de la zona como AMP y de que es una especie considerada En Peligro Crítico (RedList, UICN) y de prioridad alta para la conservación (Plan de Acción Nacional para la Conservación de Condriactios), no existen para esta área estudios específicos sobre aspectos reproductivos que permitan proveer de herramientas para su conservación y manejo. El presente trabajo tuvo como objetivo evaluar y comparar diferentes métodos estadísticos utilizados para estimar la madurez sexual de machos y hembras de *S. acuta*. Los individuos fueron obtenidos durante monitoreos de biodiversidad y evaluaciones del arte de pesca del langostino en los años 2011, 2012 y 2013. De cada ejemplar se registró: sexo, longitud total (LT, cm), ancho del disco (AD, cm) y peso total (PT, g). El grado de madurez sexual se evaluó mediante la observación macroscópica del tracto reproductivo y se determinó según Oddone et al. (2007) en tres estadios: inmaduros, en maduración y maduros. Conjuntamente, se registró peso de los ovarios (Pov, g), ancho de la glándula oviductal (AGO, mm) y ancho del útero (AU, mm) en hembras; y peso de los testículos (Ptest, g), ancho de la glándula del clasper (AGC, mm) y largo del clasper (LC, mm) en machos. El desarrollo sexual de las diferentes estructuras reproductivas (macho: LC, AGC y Ptest; hembra: AGO, AU y Pov) (variables explicativas) en función de la LT (variable de respuesta) se determinó mediante la utilización de modelos de regresión segmentada (MRS). La estimación de los parámetros se realizó siguiendo la metodología utilizada por Muggeo (2003), implementada en el paquete de R “Segmented” (Muggeo 2008). El tamaño al cual el 50% de los individuos de la población se encuentran sexualmente maduros (LTM50) se estimó para cada sexo a partir de un modelo de regresión logística (MRL), siendo la condición de madurez (0: inmaduros y 1: maduros) la variable de respuesta y la LT la variable explicativa. Asimismo, se utilizaron dos técnicas de aprendizaje automático para predecir la madurez de los individuos: árboles de clasificación y regresión (CART) y “random forest” (RF). Ambas se ajustaron para cada sexo por separado, siendo la condición de madurez (0: inmaduros y 1: maduros) la variable de respuesta y la LT, el AD y PT las variables explicativas. Los modelos se realizaron utilizando los paquetes estadísticos “rpart” (CART, Therneau et al. 2015) y “randomForest” (RF, Liaw & Wiener 2002) de R. La capacidad predictiva de CART, RF y MRL fue evaluada usando el error promedio de clasificación (EPC) mediante la estimación del error por el método de muestra de entrenamiento (2/3) y de prueba (1/3). Se realizaron 30 iteraciones para cada modelo, promediándose sus errores y calculándose su respectivo desvío estándar. Se muestrearon 149 machos (rango LT: 8,7-55,4 cm) y 103 hembras (rango LT: 8,5-54,1 cm), predominando los individuos inmaduros (machos: 84% y hembras: 94%). El LTM50 estimado para machos y hembras fue de 45,5 cm ( $IC_{95\%} = 44,5-46,4$  cm) y 52,0 cm ( $IC_{95\%} = 48,8-55,1$  cm) respectivamente. Los MRS ajustados para determinar el desarrollo sexual de las diferentes estructuras reproductivas de machos y hembras en función de la LT presentaron dos puntos de quiebre significativos en cuatro (LC, AGC, Ptest y AU) de las seis variables explicativas utilizadas. Para estos cuatro casos, la segunda pendiente siempre fue mayor que la primera y tercera ( $p - valor < 0,05$ ), indicando la existencia de dos fases de crecimiento lento (primera y tercera) y una de crecimiento rápido (segunda) de las distintas variables reproductivas en función de la LT. En cambio, los MRS ajustados a las relaciones entre el AGO/Pov vs. la LT solamente presentaron un punto de quiebre significativo, siendo la segunda pendiente mayor que la primera ( $p - valor < 0,05$ ), indicando la existencia de una fase inicial de crecimiento lenta, seguida de una rápida. Los modelos de CART (EPC: macho=0,2 y hembra=0,3) y RF

(EPC: macho=0,1 y hembra=0,2) fueron muy eficaces prediciendo la madurez de los individuos en machos y hembras. Los CART además proveen de una salida gráfica de fácil interpretación y de la estimación de un valor umbral ( $LT_{macho}$ : 45,5 cm; hembra: 48,8 cm) para clasificar a los individuos en maduros o inmaduros. En ambos sexos el orden de importancia de las variables explicativas fue la LT, el AD y en los machos el PT, no siendo esta última relevante para las hembras. En RF la variable explicativa más importante para explicar la madurez en ambos sexos fue la LT, seguida del PT y el AD, pero en este caso no se provee de una salida gráfica, ni de un valor umbral para clasificar a los individuos, por lo cual su aplicación a la gestión es más difícil. Los MRS y CART utilizados para estimar la madurez en machos y hembras de *S. acuta* presentaron resultados semejantes a los obtenidos con los métodos clásicos (MRL), por lo que se proponen como herramientas efectivas para la determinación de la talla de madurez sexual. Asimismo, las técnicas de aprendizaje automático (CART y RF) resultaron ser métodos muy eficaces en cuanto a su capacidad de predecir la madurez de los individuos, pudiendo ser una herramienta complementaria a los métodos clásicos y muy útil cuando se desean obtener predicciones cuantitativas eficientes. A pesar de ello su aplicación en estudios de ecología reproductiva (e.g. para determinar la madurez) no había sido evaluada hasta el momento. Este trabajo contribuye al conocimiento del desarrollo sexual de esta especie y particularmente se realizan aportes de métodos novedosos para la estimación de la madurez en condriictios.

# Una aplicación del Análisis de Sensibilidad al Aprendizaje Automático. Leonardo Moreno <sup>1\*</sup>

<sup>1</sup> IESTA, Departamento de Métodos Cuantitativos, FCEA, Udelar \*leonardo.moreno@fcea.edu.uy

Es conocido el impacto que están teniendo los métodos de aprendizaje automático para abordar diferentes tipologías de problemas en diversas áreas como por ejemplo en medicina, en el procesamiento de imágenes, en transporte, en seguridad o finanzas con amplias implicaciones sociales. Estas técnicas presentan en general una alta flexibilidad para afrontar problemas de complejidad con un elevado poder predictivo. Sin embargo, muchas veces son usadas para la toma de decisiones críticas sin dar ninguna explicación sobre la comprensión de cómo estas son obtenidas. En tal sentido investigaciones recientes apuntan a generar herramientas para brindar interpretabilidad a estos modelos, ver por ejemplo [8, 7]. En el trabajo describimos una aplicación de análisis de sensibilidad en dicha dirección. Dado un modelo de clasificación supervisada el objetivo es, dada una observación, identificar aquellas variables que más contribuyen en su clasificación. Para ello se generan perturbaciones aleatorias y mediante diferentes medidas de sensibilidad, como el índice de Sobol, se determina el impacto en la salida, ver [2]. Además, se utiliza un índice de sensibilidad basado en el estadístico de Cramer-Von Mises, ver [5, 3, 6], y su estimación, ver [4], mediante una nueva medida de asociación desarrollada por [1], que esta basada en estadísticos de rango. Se propone una generalización de dicho estadístico mediante proyecciones al azar al caso multivariado. Se realizan diferentes aplicaciones a datos reales, en particular a la clasificación de imágenes. Este trabajo es en colaboración con Agnes Lagnoux y Mehdi Zoutine (Toulouse, France).

[1] Sourav Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 116(536):2009–2022, 2021.

[2] Thomas Fel, Rémi Cadène, Mathieu Chalvidal, Matthieu Cord, David Vigouroux, and Thomas Serre. Look at the variance! efficient black-box explanations with sobol-based sensitivity analysis. *Advances in Neural Information Processing Systems*, 34:26005–26014, 2021.

[3] Ricardo Fraiman, Fabrice Gamboa, and Leonardo Moreno. Sensitivity indices for output on a riemannian manifold. *International Journal for Uncertainty Quantification*, 10(4), 2020.

[4] Fabrice Gamboa, Pierre Gremaud, Thierry Klein, and Agnès Lagnoux. Global sensitivity analysis: A novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374, 2022.

[5] Fabrice Gamboa, Thierry Klein, and Agnes Lagnoux. Sensitivity analysis based on cramer–von mises distance. *SIAM/ASA Journal on Uncertainty Quantification*, 6(2):522–548, 2018.

[6] Fabrice Gamboa, Thierry Klein, Agnes Lagnoux, and Leonardo Moreno. Sensitivity analysis in general metric spaces. *Reliability Engineering & System Safety*, 212:107–121, 2021.

[7] Alvarez-Castro I., Natalia Da Silva N., Moreno L., and Sosa A. Aprendizaje estadístico interpretable: una aplicación al precio de inmuebles en Montevideo, Uruguay (interpretable statistical learning: an application to property prices in Montevideo, Uruguay). In *Serie Documentos de Trabajo*, 22, 2022.

[8] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23(1):18, 2020.

## Predicción genómica mediante aprendizaje profundo. Maria Ines Faiello

La selección genómica es una metodología predictiva que busca seleccionar las mejores unidades de producción respecto de una característica medible de un organismo a partir de su información genómica. Para ello, se preentrena un modelo predictivo con una muestra de referencia la cual contiene datos genotípicos y fenotípicos. Posteriormente con este modelo entrenado se hacen predicciones para genotipos candidatos para los cuales únicamente se cuenta con información genotipo. Esta metodología ha revolucionando el mejoramiento genético ya que incrementa la ganancia genética por unidad de tiempo y ahorra recursos significativos en el fenotipado. Sin embargo, su implementación práctica es todavía compleja ya que requiere alta precisión en las predicciones para que su implementación sea exitosa. Los modelos utilizados tradicionalmente para predecir los fenotipos a partir de los genotipos son regresiones lineales con regularizaciones, tipo Ridge o Lasso, o utilizando modelos bayesianos. En este trabajo exploramos diferentes arquitecturas de redes neuronales, con el objetivo de mejorar las predicciones, ya sea en comparación a los modelos mas clásicos utilizados en el área o de aprendizaje automático.

En primer lugar evaluamos diferentes redes neuronales convolucionales (CNNs), donde mostramos que las conexiones residuales mejoran la performance de estos modelos, aunque en algunos casos las redes completamente conectadas (FCN) tienen mejor desempeño que las que las CNNs. Relacionamos este resultado con la relevancia biológica de las posiciones absolutas en el genoma, sugiriendo que la equivarianza traslacional de las CNNs puede no ser un sesgo inductivo adecuado para este problema. En segundo lugar evaluamos arquitecturas de redes neuronales soportadas en grafos (GNN), formulando la predicción de caracteres como regresión de nodos en un grafo poblacional. En dicho grafo cada nodo representa un individuo, y las aristas representan relaciones de parentesco, evaluadas genómicamente. Luego combinamos las arquitecturas de CNN y GNN. Con esta combinación de redes logramos un desempeño superior a todos los demás modelos en la predicción de producción de leche en ganado Holstein. Este resultado no fue obtenido para otros genotipos, confirmando que no existe un modelo que supere a todos los demás aún. Por último, realizamos de manera exploratoria predicciones basadas en transformers, donde aún nos queda mucho para mejorar y aprender.

Bello, Kevin, Bryon Aragam, and Pradeep Ravikumar. 2022. “DAGMA: Learning DAGs via M-matrices and a Log-Determinant Acyclicity Characterization.” In *Advances in Neural Information Processing Systems*.

Pearl, Judea, Madelyn Glymour, and Nicholas P Jewell. 2016. *Causal Inference in Statistics: A Primer*. John Wiley & Sons.

Tsamardinos, Ioannis, Laura Brown, and Constantin Aliferis. 2006. “The Max-Min Hill-Climbing Bayesian Network Structure Learning Algorithm.” *Machine Learning* 65 (October): 31–78. <https://doi.org/10.1007/s10994-006-6889-7>.

Zheng, Xun, Bryon Aragam, Pradeep Ravikumar, and Eric P. Xing. 2018. “DAGs with NO TEARS: Continuous Optimization for Structure Learning.” In *Advances in Neural Information Processing Systems*.