



O papel de R en Investigación Mariña: da xenónmica a estudo global dos océanos

Isabel Fuentes-Santos



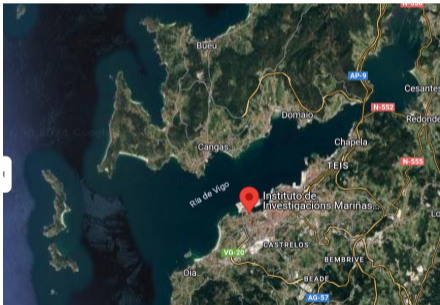
Santiago de Compostela, 24 outubro 2024

- 1 Introducción
- 2 Software estadístico no IIM
- 3 Algunas aplicación de R
- 4 Que más podemos aportar?
- 5 Conclusión



- 1 Introducción
- 2 Software estadístico no IIM
- 3 Algunas aplicación de R
- 4 Que más podemos aportar?
- 5 Conclusión

- O **IIM-CSIC** é un dos principais centros de investigación mariña da Península Ibérica.





Cambio global e outros impactos antropoxénicos sobre o océano



Observación e modelaxe de sistemas oceánicos e costeiros



Comprender as interaccións entre o océano e a vida



Biodiversidade e conservación para uns ecosistemas resilientes



Enfoque ecosistémico das pesquerías e da acuicultura



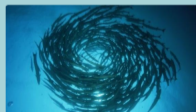
Organización da vida: dos xenes aos biomas



Alimentos do océano: trazabilidade, saúde e seguridade



Estudo e modelaxe de biosistemas e bioprocesos



Economía circular: uso integral de recursos mariños

- **RETO:** avanzar no coñecemento do océano:
 - características físicas,
 - alimento que tomamos del
 - e procesos que conectan todo o sistema
- Colaboración coas AAPP, o sector produtivo e a sociedade.
- Alineado coas misións da Axenda 2030.

Bioteconoloxía e Acuicultura

- Bioteconoloxía.
- Acuicultura.
- Inmunoloxía e xenómica.
- Patoloxía.

Oceanografía

- Procesos oceánicos e cambio global.
- Bioxeoquímica mariña.
- Xeoquímica orgánica.
- Fotobioloxía e toxicoloxía do fitoplancto.

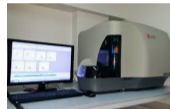
Ecoloxía e Recursos Mariños

- Pesquerías.
- Ecoloxía mariña.

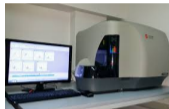
Tecnoloxía de alimentos

- Enxeñería de bioprocesos.
- Química de produtos mariños.
- Bioxeoquímica de alimentos.
- Microbioloxía e tecnoloxía de produtos mariños.
- Reciclado e valorización de residuos.

R Que papel ten R no IIM?



- O carácter multidisciplinar do IIM implica traballar cunha ampla variedade de datos.
- Distintas ferramentas tanto para recoller como para **tratar e analizar os datos**.



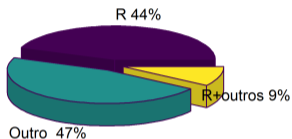
- O carácter multidisciplinar do IIM implica traballar cunha ampla variedade de datos.
- Distintas ferramentas tanto para recoller como para **tratar e analizar os datos**.

● Revisión do papel de R no IIM

- Revisión bibliográfica 2021-2024 (160 traballos).
- Software e técnicas estatísticas utilizadas.
- Paquetes de R específicos.



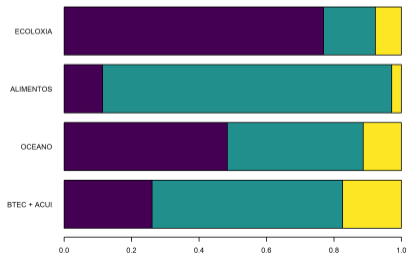
- 1 Introducción
- 2 Software estadístico no IIM**
- 3 Algunas aplicación de R
- 4 Que más podemos aportar?
- 5 Conclusión



- Úsase R en máis do 50% dos traballos.
- Os traballos do departamento de **Tecnoloxía de Alimentos** teñen menos carga estatística.
- En **Biotecnoloxía** combínase o uso de software estatístico e xenómico.

• Paquetes de R máis usados

- **Visualización:** `ggplot2`.
- **Modelos e inferencia:** `mgcv`, `nlme`, `INLA`.
- **Acceso a datos:** `rearddap`, `rnaturalearth`, `worms`.
- **Xenómica:** `Bioconductor`, `qiime2R`.





- Tests de comparación clásicos (ANOVA, PERMANOVA, T test, Wilcoxon...).
- Regresión lineal simple e múltiple.
- Regresión non lineal (MATLAB).
- Modelos harmónicos para series de tempo (MATLAB).
- Análise de datos composicionais (StatGraphics).



- 1 Introducción
- 2 Software estadístico no IIM
- 3 Algunas aplicación de R**
- 4 Que más podemos aportar?
- 5 Conclusión



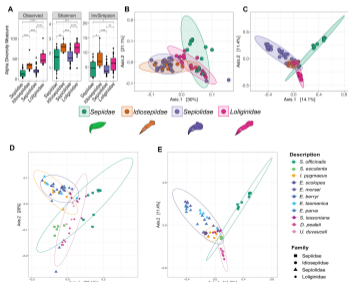
- Software específico para datos xenómicos.
- **Bioconductor**: software bioinformático compatible con R.
 - microbiome: secuenciación.
 - qvalue: control de falsos positivos en tests múltiples.
 - pcaMethods: análisis de componentes principales.
 - Mfuzz: clustering para series de tiempo de expresiones xénicas.
- qiime2R: integra **QIIME2** e R.

Host-Microbial Interactions | Research Article | 5 February 2024



Evolutionary history influences the microbiomes of a female symbiotic reproductive organ in cephalopods

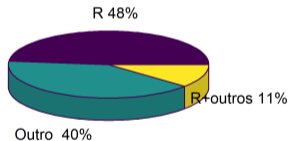
Authors: Nidhi Vijayan, Sarah J. McAnulty, Gustavo Sanchez, Jeffrey Jolly, Yuzuru Ikeda, Michele K. Nishiguchi, Elodie Réveillac, SHOW ALL (3 AUTHORS), Spencer V. Nyholm, AUTHORS INFO & AFFILIATIONS



- Caracterización do microbioma na glándula reproductiva das femias de calamar e sepia.
- **QIIME 2:** plataforma bioinformática para tratamento e análise de datos de microbioma.
- **Contribución de R**
 - qiime2R: artefactos QIIME2 a obxectos *phyloseq*.
 - phyloseq: alfa-diversidade.
 - vegan: beta-diversidade e Bray-Curtis PERMANOVA tests.
 - Ferramentas gráficas: ggplot2, RColorBrewer, dplyr, gpubr, reshape2.

Paquetes específicos

- `marmap`: acceso e análise de datos de batimetría e topografía.
- `rearddap`: acceso a datos de NOAA.
- `mixSIAR`: modelos *mixing* bayesianos.
- `heatwaveR`: definición, detección e análise de eventos extremos.
- `oce`: análise de datos oceanográficos (ADCP, CTD,...).
- `seacarb`: estimación de parámetros do sistema de carbono da auga de mar (acidificación oceánica).



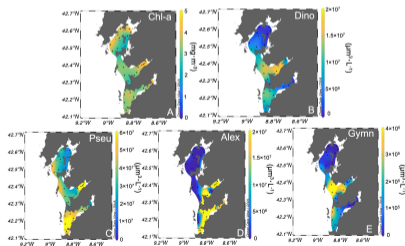
ORIGINAL RESEARCH article

Front. Mar. Sci., 16 February 2024

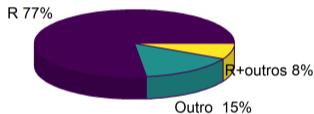
Sec. Marine Ecosystem Ecology

Volume 11 - 2024 | <https://doi.org/10.3389/fmars.2024.1330090>

Potentially toxicogenic phytoplankton patterns in the northwestern Iberian Peninsula



- Monitorización semanal de características hidrológicas e abundancia de fitoplancton tóxico na columna de auga (0 a 15 m) nas rías galegas (29 estacións, 2015-2022).
- **Obxectivo:** estimar o risco de mareas vermellas.
- **Contribución de R**
 - Análise exploratorio espacial e temporal.
 - Análise de correspondencias entre características hidrográficas e fitoplancton tóxico.
 - `mgcv`: GAM para estimar abundancia de fitoplancton tóxico
 - Transformación logarítmica da resposta.
 - Variables explicativas: temperatura da auga, Chl-a, UI, aporte de ríos, latitude e mes.
 - Selección de variables basada en AIC.



Pesquerías

- `VMStools`: acceso e análise de datos VMS.
- `CatDyn`: avaliación de stocks pesqueiros.

Ecoloxía

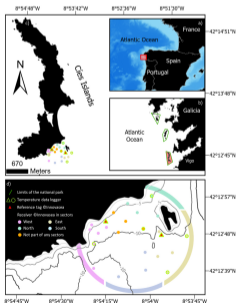
- `ade4`: análise de datos ecolóxicos.
- `adehabitatHR`: estimación do territorio ocupado por un animal.
- `BiodiversityR`: análise de ecoloxia comunitaria e sostibilidade.
- `cooccur`: análise de concorrencia de especies.
- `indicspecies`: relación entre especies e grupos de zonas.
- `tRophicPosition`: estimación bayesiana da posición na cadea trófica.
- `worrms`: rexistro mundial de especies mariñas.

Aquatic Conservation: Marine and Freshwater Ecosystems



RESEARCH ARTICLE | [Open Access](#) |

Drivers of behaviour and spatial ecology of the small spotted catshark (*Scyliorhinus canicula*)



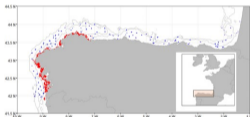
- **Telemetría acústica** para monitorizar 9 raias na contorna de Cíes (Set 2020 a Xuño 2022).
- **Obxectivo:** analizar o movemento das raias dentro da área de estudo e a **dirección** de entrada-saída da zona.
- **Contribución de R**
 - `adehabitatHR`: Kernel Utilization Distribution (KUDs) para estimar a zona de actividade de cada individuo (semanal).
 - `sunclac`: ciclos día - noite.
 - `mgcv`: GAMM, modelo AR1 e individuo como factor aleatorio.
 - Probabilidade de presenza: día do ano (cc), temperatura e sexo.
 - Zona de actividade: semana (cc).

JOURNAL ARTICLE

Combining fishery data through integrated species distribution models

Iosu Paradinas , Janine B Illian, Alexandre Alonso-Fernández, Maria Grazia Pennino, Sophie Smout

ICES Journal of Marine Science, Volume 80, Issue 10, December 2023, Pages



- INLA: para implementar un **Modelo Integrado de Distribución de Especies** aplicado a estimación de captura de linguado.

Journal of Animal Ecology



RESEARCH ARTICLE |  Open Access  

Atlantic cod individual spatial behaviour and stable isotope associations in a no-take marine reserve

Christopher T. Monk , Michael Power, Carla Freitas, Philip M. Harrison, Michelle Heupel, Anna Kuparinen, Even Moland, Colin Simpfendorfer, David Villegas-Rios, Esben M. Olsen

- Relación entre a migración e alimentación de atún.
- **Contribución de R**
 - `survival`: estimador Kaplan-meier da relación entre destino e nicho trófico.
 - `MCMCglmm`: axuste bayesiano de GLMM multivariantes.



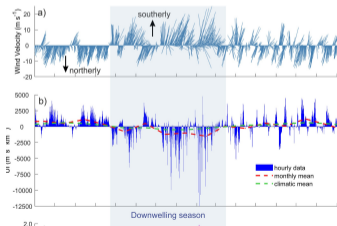
- 1 Introducción
- 2 Software estadístico no IIM
- 3 Algunas aplicación de R
- 4 Que más podemos aportar?**
- 5 Conclusión

JGR Oceans

Research Article  Open Access   

Wave Regime and Wave-Current Coupling in an Upwelling-Driven Bay: Seasonal and Inter-Annual Variability

Nicolás Villalacios-Robineau  Miguel Gilcoto, Paula C. Pardo, Eric Desmond Barton



● Metodoloxía actual

- Uso habitual do índice de afloramento para caracterizar o réxime de vento costeiro.
- Trátanse as direcciónn de vento, correntes e oleaxe como vectores.

● Proposta

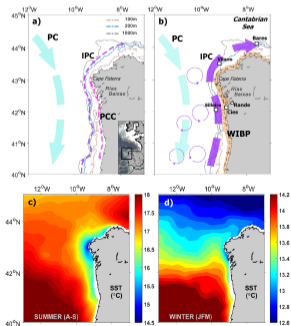
- Tratar vento, correntes e oleaxes como **datos direccionais**
- Usar as ferramentas implementadas no paquete **NPCirc**.

JGR Oceans

Research Article | [Open Access](#) |

Seasonal, Interannual and Long-Term Variability of Sea Surface Temperature in the NW Iberian Upwelling, 1982–2020

S. Piedracoba P. C. Pardo, X. A. Álvarez-Salgado, S. Torres



• Datos

- Temperatura media diaria, cunha resolución de $0.05^\circ \times 0.05^\circ$ entre 1982 e 2020.
- Temperatura medida en estacións de monitorización.

• Metodoloxía usada

- Regresión lineal para validar os datos de satélite.
- Modelos harmónicos de series de tempo.

• Propostas

- Usar técnicas de análise de datos espaciais e espaciotemporais (*spatial*, *npsp*).
- Tratar os ciclos anuais de temperatura como datos funcionais (*fda.usc*).

- 1 Introducción
- 2 Software estadístico no IIM
- 3 Algunas aplicación de R
- 4 Que más podemos aportar?
- 5 Conclusión**

- En Investigación Mariña atopamos gran diversidade de datos.
- Uso habitual de R combinado con outros programas.
- **Limitación**
 - A maioría dos investigadores non teñen acceso ós últimos avances en técnicas de tratamento e análise de datos.
 - Uso habitual de técnicas de estatística clásica e de ferramentas de representación gráfica.
 - Tendencia a transformar os datos para poder aplicar ferramentas coñecidas.
- Entretanto a **comunidade estatística** desenvolve ferramentas e software adaptado a distintos tipos de datos.
- **TAREFA PENDENTE:** fortalecer a comunicación entre a Investigadores Mariña e a Estatística.

GRACIAS!



PID2020-116587GB-I00

CODyNP

Complex Dynamics and Nonparametric Inference