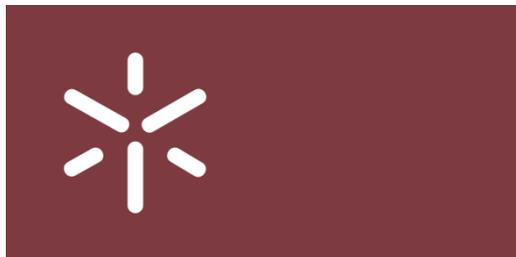




Relación entre presiones antrópicas y ecosistemas en los ríos (MARS – Project)

Tano Gutiérrez-Cánovas



Universidade do Minho



Directiva Marco del Agua: Estrategia europea de evaluación y mejora ambiental de las masas de agua

Año 2000

- Evaluar **estado ecológico** y recuperar condiciones de “buen estado ecológico” (2027)
- Foco en detectar y gestionar el **impacto** dominante

2015-2020:

- Más del **50%** de las aguas europeas están **degradadas**
- No hay **mejoras** importantes entre 2009-2015

Más del 39% de las aguas europeas tienen 2 o más tipos de impacto

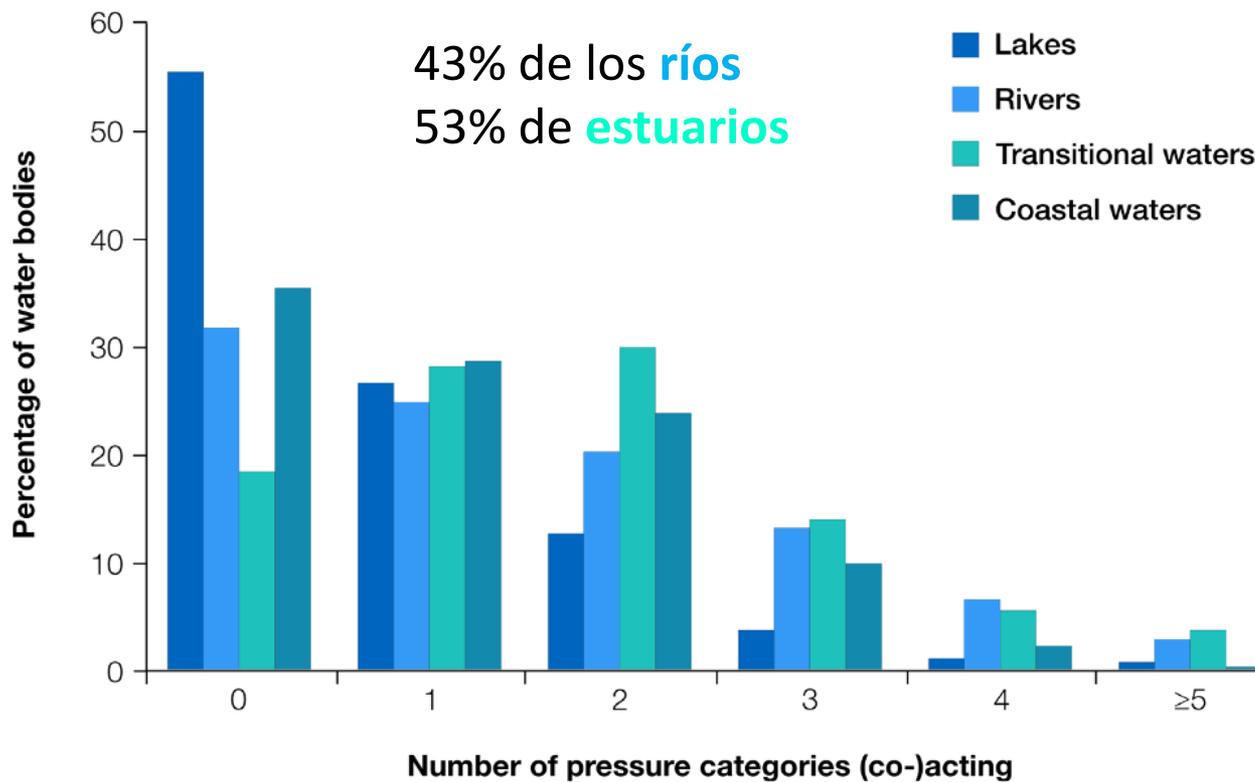


Figure 1: Percentage of water bodies affected by no, one or multiple pressure categories. Pressure categories cover point source pollution, diffuse pollution, water abstraction, physical alteration, hydrological alteration, continuity disruption and other pressures.

MARS
PROJECT

Managing Aquatic
ecosystems and
water Resources
under multiple Stress

Contaminación difusa
Alteraciones hidromorfológicas

Impactos múltiples: implicaciones para gestores

Imaginemos dos tipos de impacto comunes:

1. Contaminación difusa: **nitratos**
2. Alteraciones hidromorfológicas: **regulación de caudal**

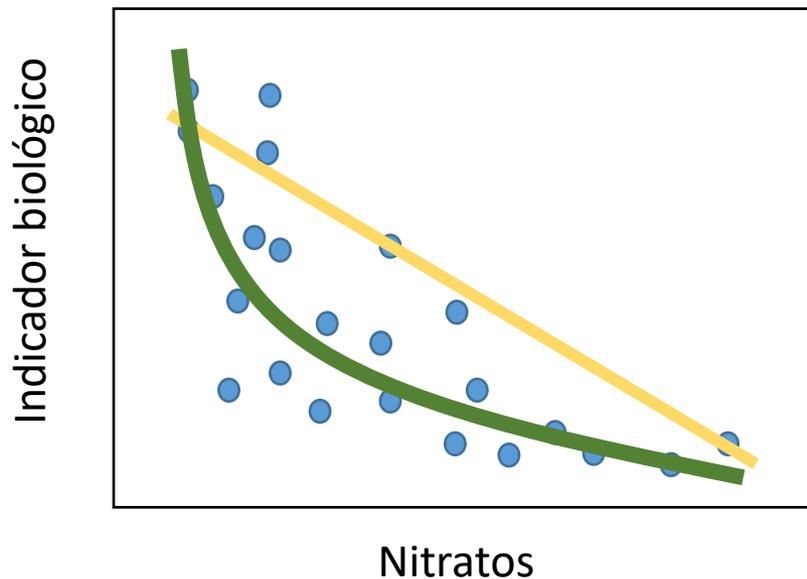
Tipos de efecto conjunto:

Estrategia de gestión:

- **dominante: $1 + 0 = 1$** Mitigar **dominante**
- **aditivo: $1 + 1 = 2$** Mitigar según coste-beneficio
- **sinérgico: $1 + 1 = 3$** Mitigar **ambos**
- **antagonista: $1 + 1: 1,5$** Estrategia adaptativa

¿Cómo identificamos los impactos más importantes y su efecto conjunto?

- **GLM/GLMM**: modelos lineales, regresión lineal, modelos mixtos
- **Machine learning**: CART, Random Forest (RF), Boosted Regression Trees (BRT)



Ventajas e inconvenientes de los modelos lineales

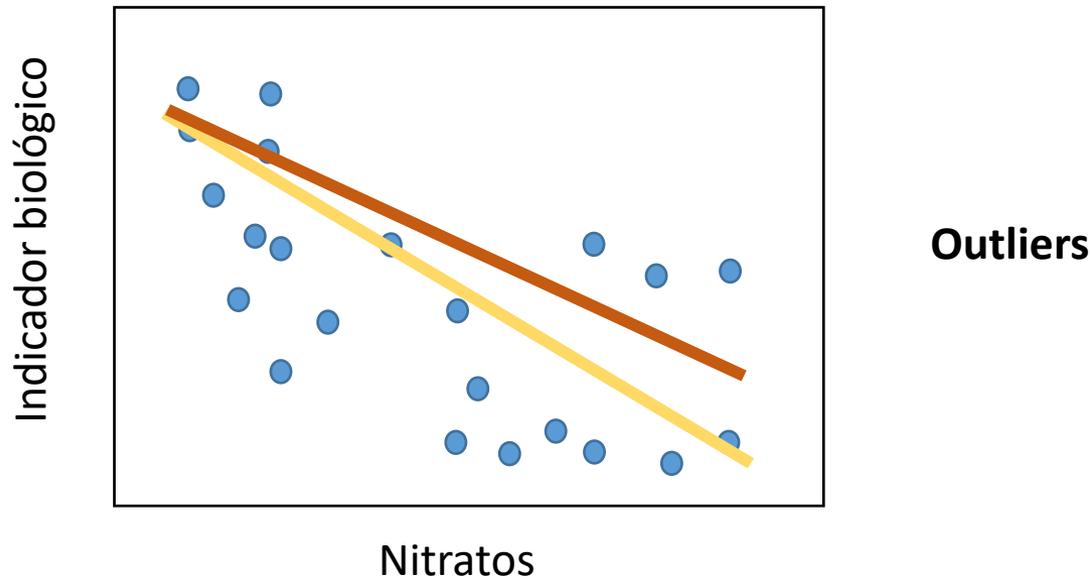
GLM/GLMM: Generalized Linear Mixed Models



- Modelos **transparentes** y conocidos
- Fácilmente **interpretables**
- Relativamente fiables con “bajo” número de **observaciones** (≤ 100)
- Buenos para testar **hipótesis**
- Medidas **repetidas** en espacio o tiempo



- **Poco flexibles**
- Sensibles a **datos incompletos** (NAs)
- Problemas con **valores extremos**
- Limitación respecto al número de **predictores**
- **Asunciones** estadísticas



Ventajas e inconvenientes de técnicas Machine Learning

Random Forest, Boosted Regression Trees

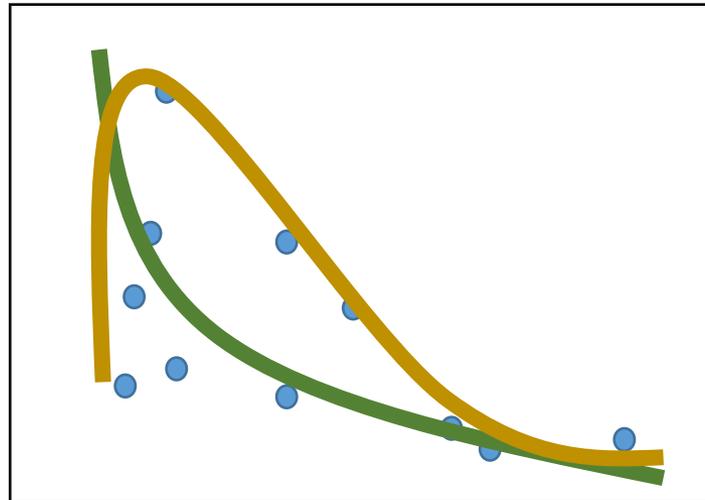


- Modelos muy **flexibles**
- Relaciones **no-lineales**
- **Jerarquización** de los impactos con muchos predictores
- Puede gestionar **datos incompletos** (NAs) - autocompletado
- Respuestas y predictores de distinta **naturaleza** (categóricos, binarios)



- Modelos **poco transparentes** y **menos conocidos**
- Tendencia a exagerar su **capacidad predictiva** (Overfitting)
- Necesitan un alto número de **observaciones**
- No son buenos para testar **hipótesis**
- Problemas con **medidas repetidas**

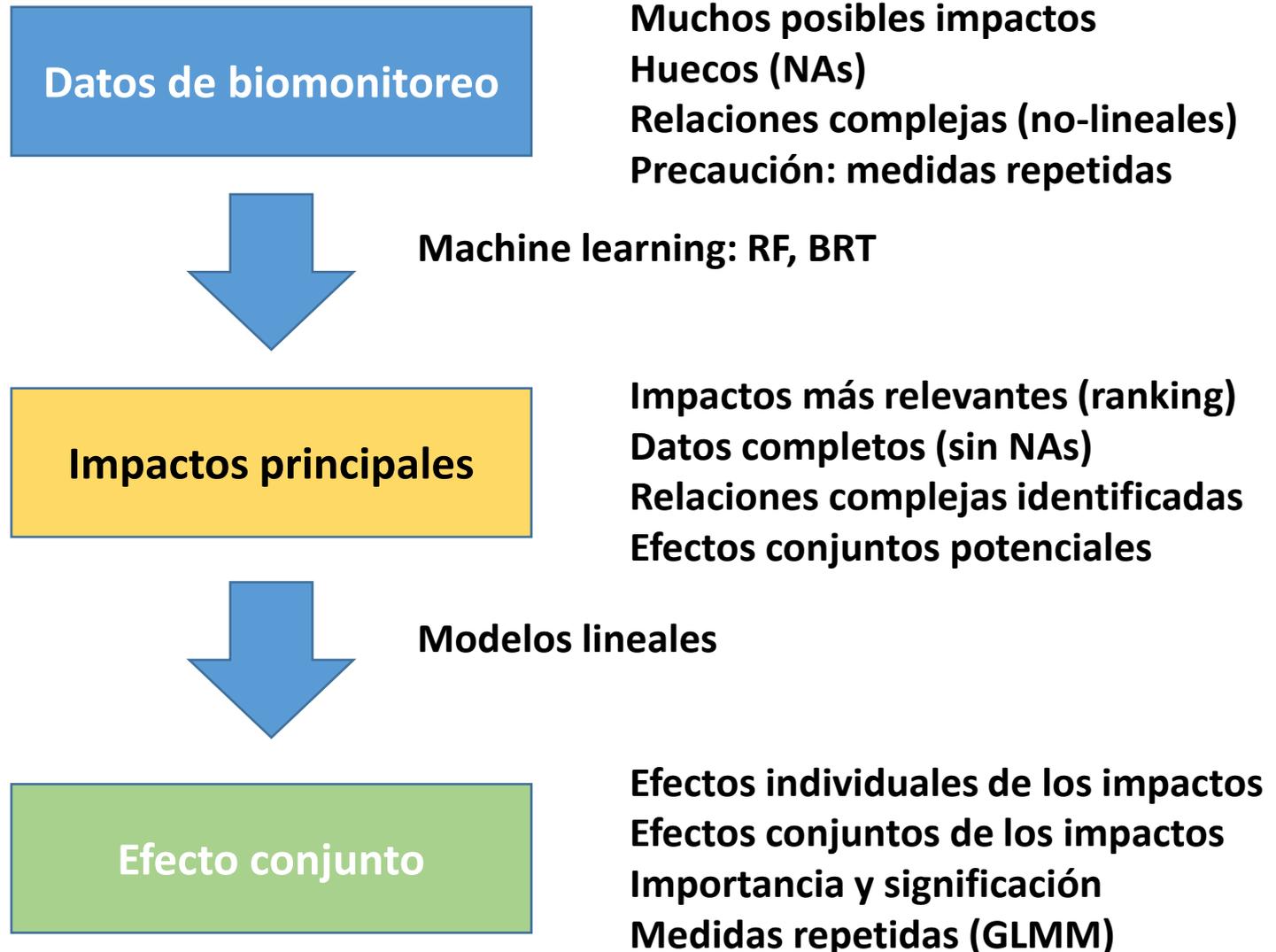
Indicador biológico



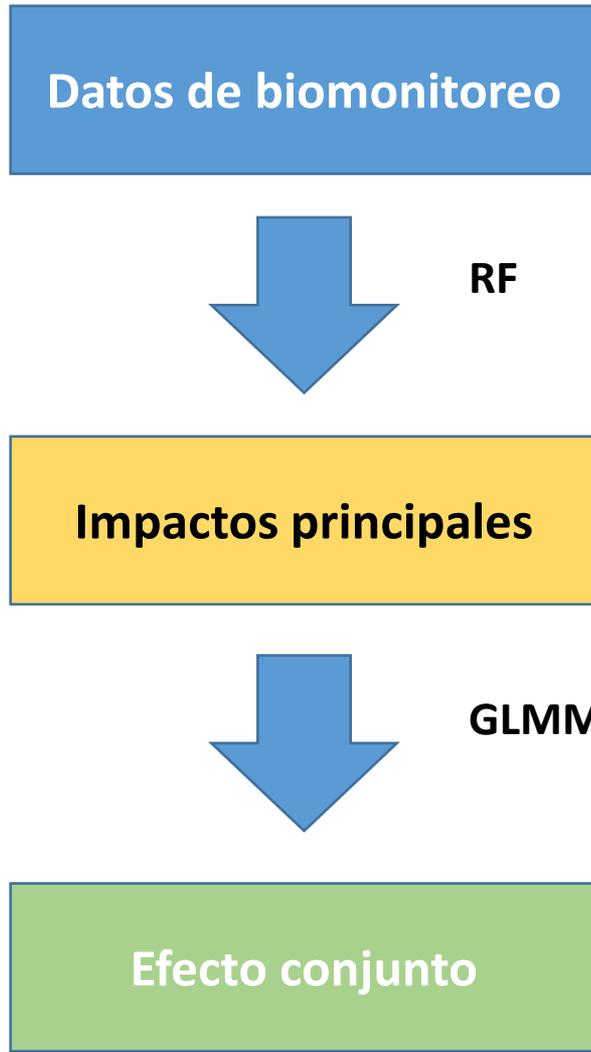
Overfitting

Nitratos

Uso complementario de modelos lineales y Machine Learning



Veamos un ejemplo en la cuenca del Ebro:

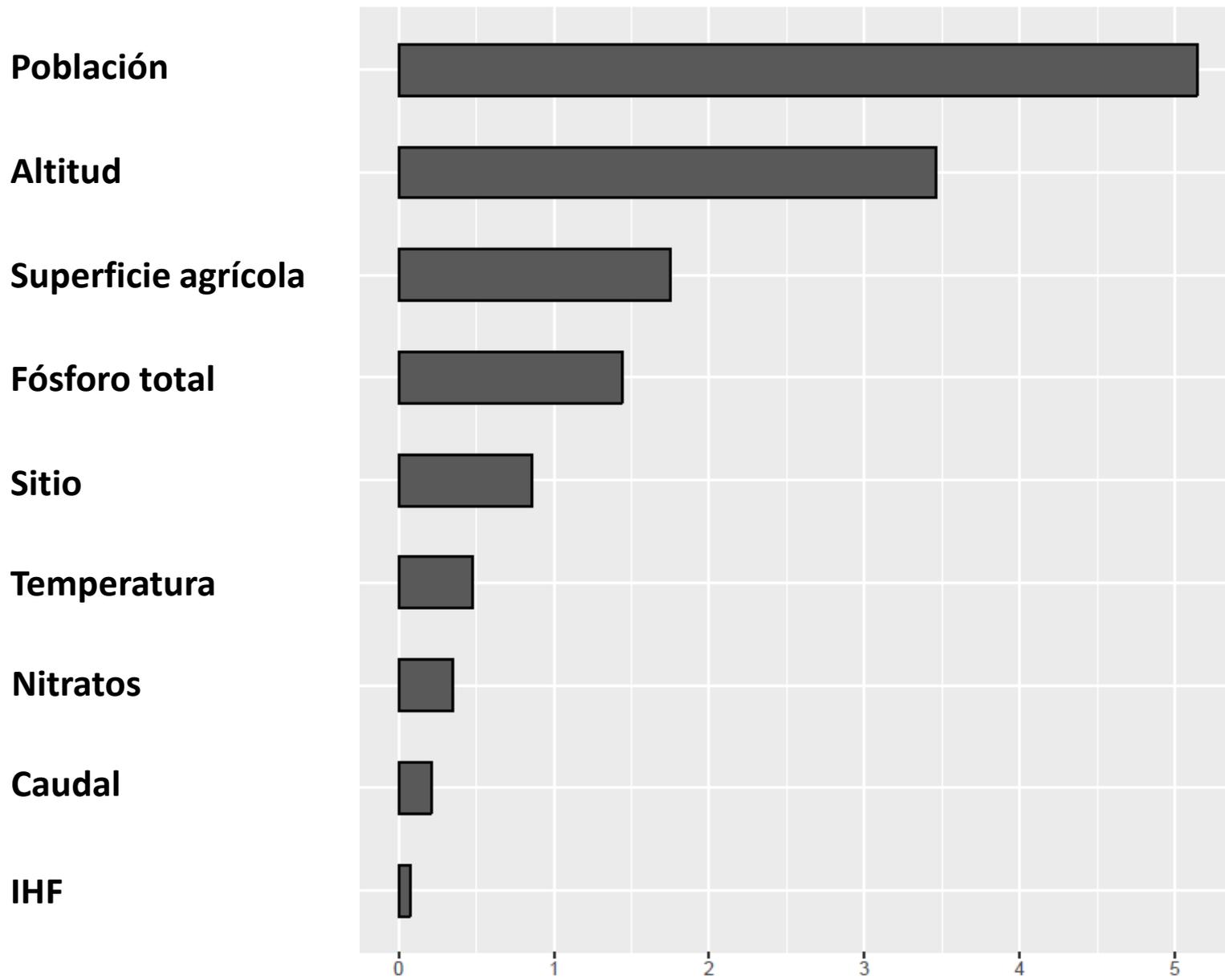


Indicador biológico: índice diatomeas (IPS)
8 candidatos: incluyendo impactos (agr, pob) y covariables naturales (altitud, caudal)
Huecos (NAs)
Relaciones complejas (no-lineales)
Medidas repetidas

Impactos: Población, superficie agricultura, fósforo
Covariable: altitud
Datos completos (sin NAs)
Relaciones complejas identificadas
Posibles interacciones: agr x alt, pob x alt

Efectos individuales de los impactos
Efectos conjuntos de los impactos
Importancia y significación
Medidas repetidas (GLMM)

Ranking de predictores con RF



Efectos conjuntos con modelo lineal

Table 1

Results of the models relating IPS, IVAM and IBMWP to single and combined abiotic variables. Standardized Effect Size (SES) and significance are shown. Significant variables are highlighted in bold. r^2_m : variance explained just by the fixed factors; r^2_c : variance accounted for by both fixed and random terms.

BQE	Abiotic variables	SES	P-value	r^2_m	r^2_c
IPS	Intercept	15.14	0.000	57.0	75.3
	TP	-0.93	0.001	6.2	
	Nitrates	0.18	0.538		
	Agriculture	-1.09	0.004	11.0	
	Inhabitants	-1.34	0.001	26.1	
	IHF	-0.47	0.055		
	Temp	0.18	0.524		
	Qmean	0.33	0.340		
	Altitude	0.33	0.474	5.4	
	Agriculture x altitude	-0.44	0.211	0.0	
	Inhabitants x altitude	-0.24	0.437	0.0	

En resumidas cuentas:

1. El biomonitoreo y la gestión de las aguas deben de considerar **todos los impactos**
2. Identificar de forma correcta el **efecto conjunto** de varios impactos es fundamental para desarrollar la estrategia de gestión adecuada
3. Los modelos basados en “machine learning” **complementan** a los modelos lineales y ofrecen una gran potencia para explorar los datos
4. Su valor añadido es su capacidad para **identificar** los impactos dominantes en situaciones complejas



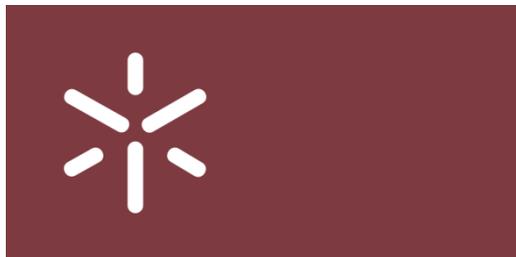
¡Muchas gracias!

Tano Gutiérrez-Cánovas
(cayeguti@um.es)

Tutorial con código en R disponible en GitHub
<https://github.com/tanogc/Multistressors-SIBECOL>



@tano_gc



Universidade do Minho

