

# ¿Afirmaciones extraordinarias exigen evidencia extraordinaria? Reproducibilidad y datos abiertos en calidad de agua.

**Palabras clave:** Reproducibilidad, Ética de Investigación, Eutrofización, Ríos y Embalses

## Contexto

A finales del 2020 una revista internacional publicó un artículo de investigadores de Uruguay (BC2021<sup>1</sup>) donde se afirmaba que la eutrofización (proceso de enriquecimiento de los ambientes acuáticos con nutrientes) de origen agrícola no era una de las principales causas de la proliferación de cianobacterias en ríos y embalses artificiales. El artículo mostraba interpretaciones erróneas de la correlación estadística de causa y efecto, descuidando el conocimiento limnológico básico de la ecología de las floraciones de algas nocivas (FAN) y la investigación internacional sobre los efectos del uso de la tierra en la calidad del agua dulce hasta el momento. Inclusive al argumentar que el pH es una variable de predicción de las FAN ignoraban los paradigmas básicos de los flujos de carbono y el cambio en el pH debido a la productividad primaria. Como resultado, el artículo introducía la noción de que la formación de FAN no está relacionada con el uso de la tierra agrícola y el tiempo de residencia del agua lo cual implica un gran riesgo para la gestión de los cuerpos de agua superficiales. De la lectura del trabajo se podían identificar aspectos metodológicos débiles como usar ríos de características hidrológicas marcadamente diferentes. A raíz de esta conclusión y sumando a las discusiones teóricas de la disciplina sobre fallas en la interpretación del trabajo BC2021, a finales del 2021 un grupo de 53 investigadores nacionales representando a 18 instituciones diferentes publicamos una respuesta al artículo original sobre la que se basa el presente trabajo<sup>2</sup>.

## Datos Abiertos

El Observatorio Ambiental Nacional (OAN; <https://www.ambiente.gub.uy/oan/>) es un repositorio de datos ambientales del Ministerio de Ambiente de Uruguay (anteriormente Dirección Nacional del Medio Ambiente - DINAMA) que disponibiliza datos recopilados en programas de monitoreo nacionales de diversas instituciones públicas. En esta fuente de datos se basó la publicación de BC2021.

## Reproducibilidad

La reproducibilidad de un artículo científico se define como un trabajo que puede ser revisado y recreado de forma independiente, utilizando los mismos datos y análisis publicados por los autores originales<sup>3</sup>. Reproducible es diferente de replicable, robusto y generalizable<sup>4</sup>. La reproducibilidad se puede entender a nivel informático (códigos,

## ¿AFIRMACIONES EXTRAORDINARIAS EXIGEN EVIDENCIA EXTRAORDINARIA?

### REPRODUCIBILIDAD Y DATOS ABIERTOS EN CALIDAD DE AGUA

---

software, hardware, etc.), empírico (información sobre los experimentos y observaciones que permiten la adquisición y disponibilidad de datos) y estadístico (información sobre análisis estadísticos, valores umbral de detección, poder estadístico de las pruebas y parámetros de los modelos)<sup>5</sup>. En este trabajo, se analizaron los tres niveles de reproducibilidad sobre el artículo BC2021 mencionados anteriormente. El procedimiento de análisis, así como los códigos y paquetes utilizados en R se encuentran públicos en el siguiente repositorio: [https://github.com/Nalcan/Reply\\_BC2021](https://github.com/Nalcan/Reply_BC2021).

## Resultados y discusión

Identificamos fallas graves debido a la manipulación de datos errática e inexacta. El estudio no definió cómo se trataron los valores erróneos y faltantes del conjunto de datos original, cómo se introdujeron numéricamente las variables por debajo del límite de detección/cuantificación y detectamos la eliminación arbitraria de valores de pH > 7,5 (que no fueron valores atípicos) cuando se incluyeron valores extremos de otras variables ambientales.

Además, identificamos errores conceptuales y de procedimiento, como la construcción/evaluación sesgada de la capacidad de predicción del modelo. El estudio entrenó el modelo utilizando datos agrupados de una sección restringida del (gran) río Uruguay y del río Negro y sus embalses, pero luego probó la predictibilidad dentro del (pequeño) río Cuareim.

Este trabajo enfatiza la necesidad de buenas prácticas de gestión de datos abiertos, especialmente para bases de datos públicas en vista de la reproducibilidad externa. Destacamos la importancia de contar con bases de datos abiertas y la filosofía de reproducibilidad promovidas por las comunidades de usuarios de R como herramientas para hacer ciencia de calidad, democrática y reproducible.

## Bibliografía

1. Beretta-Blanco, A. & Carrasco-Letelier, L. Relevant factors in the eutrophication of the Uruguay River and the Río Negro. *Science of The Total Environment* **761**, 143299 (2021).
2. Alcántara, I. *et al.* A reply to “Relevant factors in the eutrophication of the Uruguay River and the Río Negro”. *Science of The Total Environment* 151854 (2021) doi:10.1016/j.scitotenv.2021.151854.
3. Bello, N. M. & Renter, D. G. Invited review: Reproducible research from noisy data: Revisiting key statistical principles for the animal sciences. *Journal of Dairy Science* **101**, 5679–5701 (2018).
4. Culina, A., Berg, I. van den, Evans, S. & Sánchez-Tójar, A. Low availability of code in ecology: A call for urgent action. *PLOS Biology* **18**, e3000763 (2020).
5. The Turing Way Community *et al.* The Turing Way: A Handbook for Reproducible Data Science. (2019) doi:10.5281/ZENODO.3233986.