

Ciencia de datos en ciencia e ingeniería del agua

María Castrillo Melguizo

Grupo de Computación Avanzada y e-Ciencia

castrillo@ifca.unican.es

9 de octubre de 2024

Contenido

1. Grupo de Computación Avanzada y e-Ciencia (IFCA (CSIC-UC))
2. Situación actual de los datos del sector del agua
3. Aplicaciones de la ciencia de la ciencia de datos en la gestión y la ingeniería del agua
 - Soft-sensors
 - Estratificación térmica en embalses

Take-home message

Promoción de la innovación

- Transformación digital en el sector del agua

Metodología

- Abordaje de un problema de ciencia de datos

Ciencia abierta

- Datos en abierto

Grupo de Computación Avanzada y e-Ciencia

Instituto de Física de Cantabria

Centro de investigación mixto del Consejo Superior de Investigaciones Científicas (CSIC) y la Universidad de Cantabria (UC).

Fundado en 1995, el objetivo fundamental del IFCA era la investigación de alta calidad en los ámbitos de la Astrofísica y la Estructura de la Materia.

Actualmente cuenta con una plantilla de unas 100 personas.



Grupo de Computación Avanzada y e-Ciencia

Dos líneas de investigación complementarias:

- Servicios y herramientas de computación avanzada
- Ciencia de datos (imagen médica, medio ambiente, etc.)

Infraestructura:

- Computación en la nube (renovado 2023)

- > 15000 cores
 - 100 GPUs (T4 V100)
 - 500TB almacenamiento
- HPC (en renovación)
 - Parte de la RES
 - 2.5PB almacenamiento

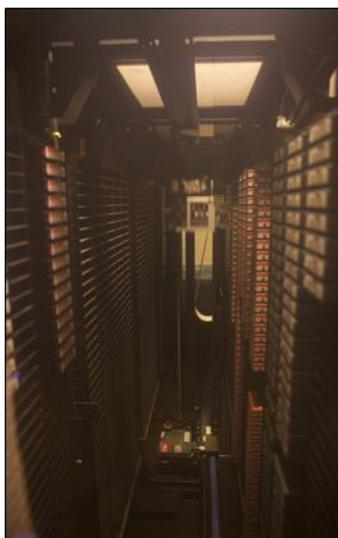
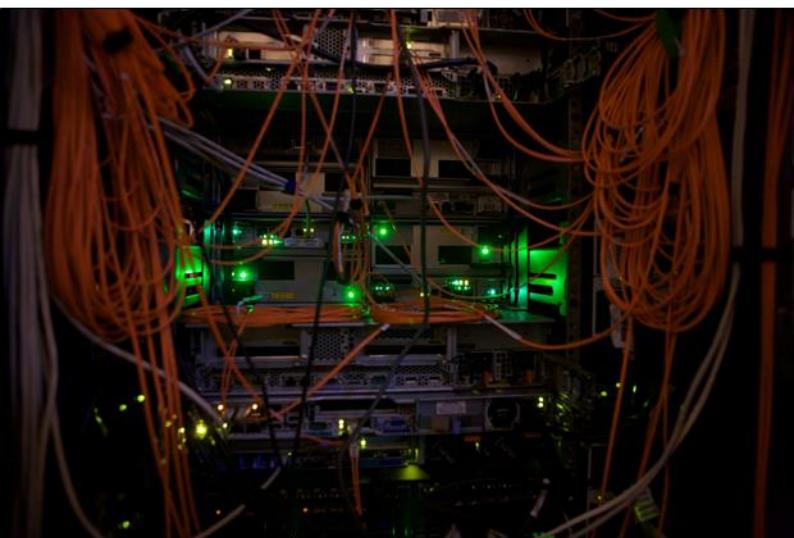


<https://computing.ifca.es/>

Miembros:

- **Permanentes:** Fernando Aguilar, María Castrillo, Lara Lloret, Álvaro López, Jesús Marco, David Rodriguez
- **Contratados:** Ibán Cabrillo, Jaime Céspedes (f. doctores), Miriam Cobo (FPU), Pablo Izquierdo (PCD2022), Andrés Heredia (FACE), Ignacio Heredia (AI4EOSC), Jaime Iglesias (iMagine), José González (f. doctores), Esmeralda Madrazo (JAE Intro), Miguel Ángel Nuñez Vega, Aida Palacio (AI4EOSC), María Peña (JAE Intro), Judith Sainz-Pardo (FACE), Susana Rebolledo (AI4EOSC), Carmen Marcos (iMagine)

Grupo de Computación Avanzada y e-Ciencia

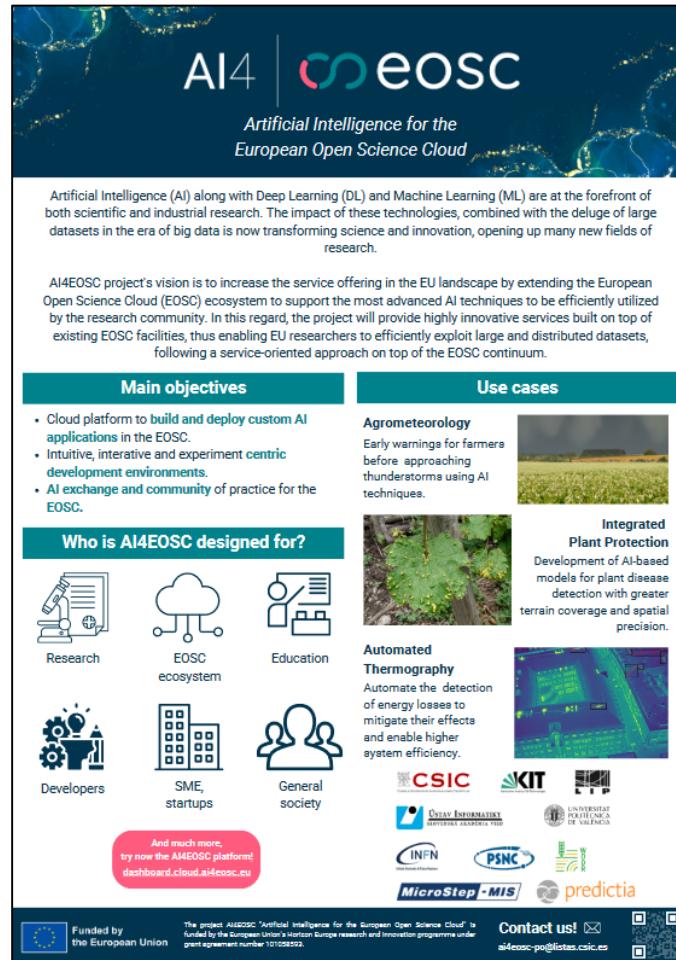


Grupo de Computación Avanzada y e-Ciencia

Proyectos relevantes:



Desarrollo y validación de un sistema de información digitalizado para la evaluación integral de hábitos alimentarios en grupos poblacionales (PDC2022-133861-C21 and PDC2022-133861-C22)



The image shows the official website for the AI4EOSC project. At the top, it features the project logo 'AI4 eosc' with the subtitle 'Artificial Intelligence for the European Open Science Cloud'. Below this is a brief description of the project's vision and objectives. The main content area is organized into three columns: 'Main objectives', 'Use cases', and 'Who is AI4EOSC designed for?'. The 'Main objectives' section lists the development of a cloud platform for AI applications, an intuitive development environment, and a community of practice. The 'Use cases' section highlights four areas: Agrometeorology (warning farmers before thunderstorms), Integrated Plant Protection (disease detection), Automated Thermography (energy loss automation), and Research, EOSC ecosystem, Education, Developers, SME startups, and General society. A red button at the bottom encourages users to try the platform. The footer includes logos for partners like CSIC, KIT, INFN, PSNC, and predictia, along with funding information from the European Union.

Situación actual de los datos del sector del agua

¿Puede la ciencia de datos ayudar a la gestión sostenible del agua?

- First, it creates the demand for more good information.
- Second, good data changes behavior.
- Finally, data ignites innovation.

The New York Times

Opinion

OP-ED CONTRIBUTOR

Water Is Broken. Data Can Fix It.

By Charles Fishman
March 17, 2016

Give this article Share Bookmark 42

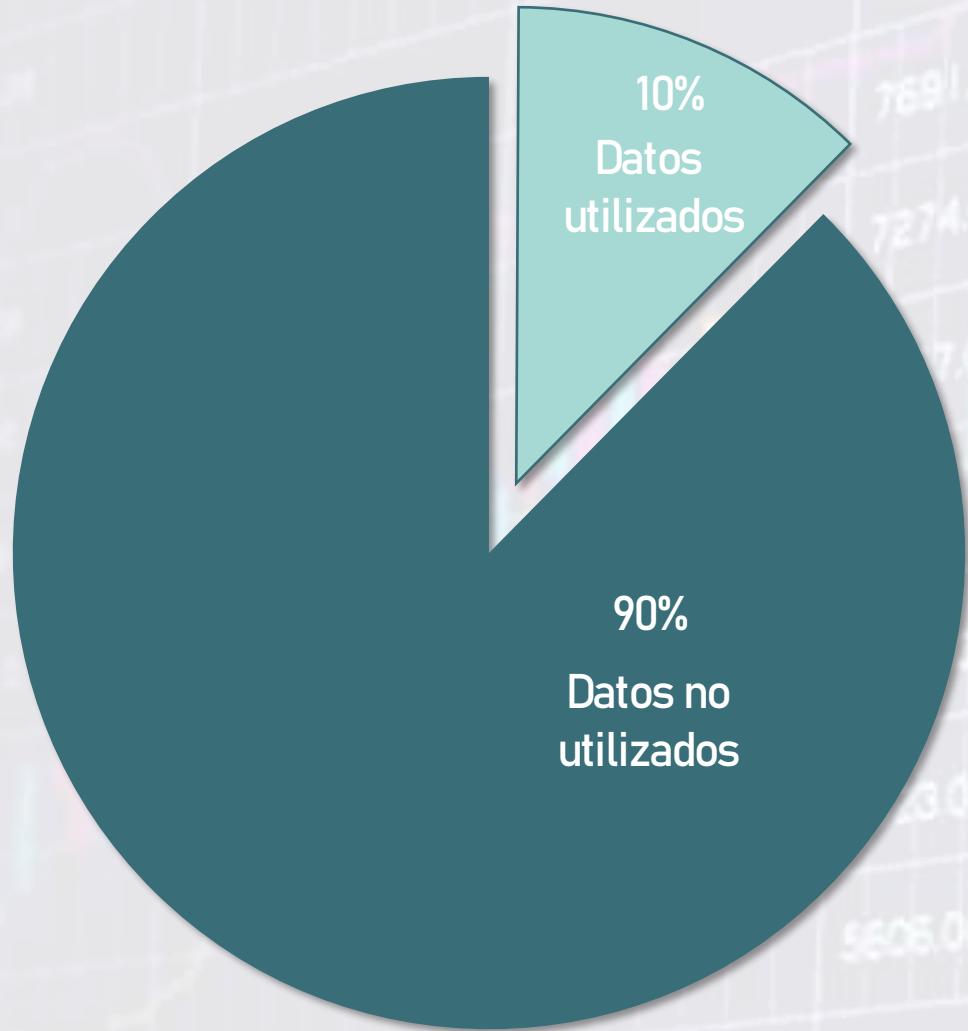


Sprinklers in an agricultural field in California. Max Whittaker for The New York Times



“Utilities are becoming overwhelmed with data.”

Leveraging Other Industries – Big Data Management
The Water Research Foundation. 2018.



- ✓ Falta de herramientas de analítica de Big Data
- ✓ Falta de personal especializado
- ✓ Baja calidad de los datos
- ✓ Seguridad de los datos

Encuesta realizada a 50 mayores compañías de agua en EEUU.
Leveraging Other Industries – Big Data Management
The Water Research Foundation. 2018.

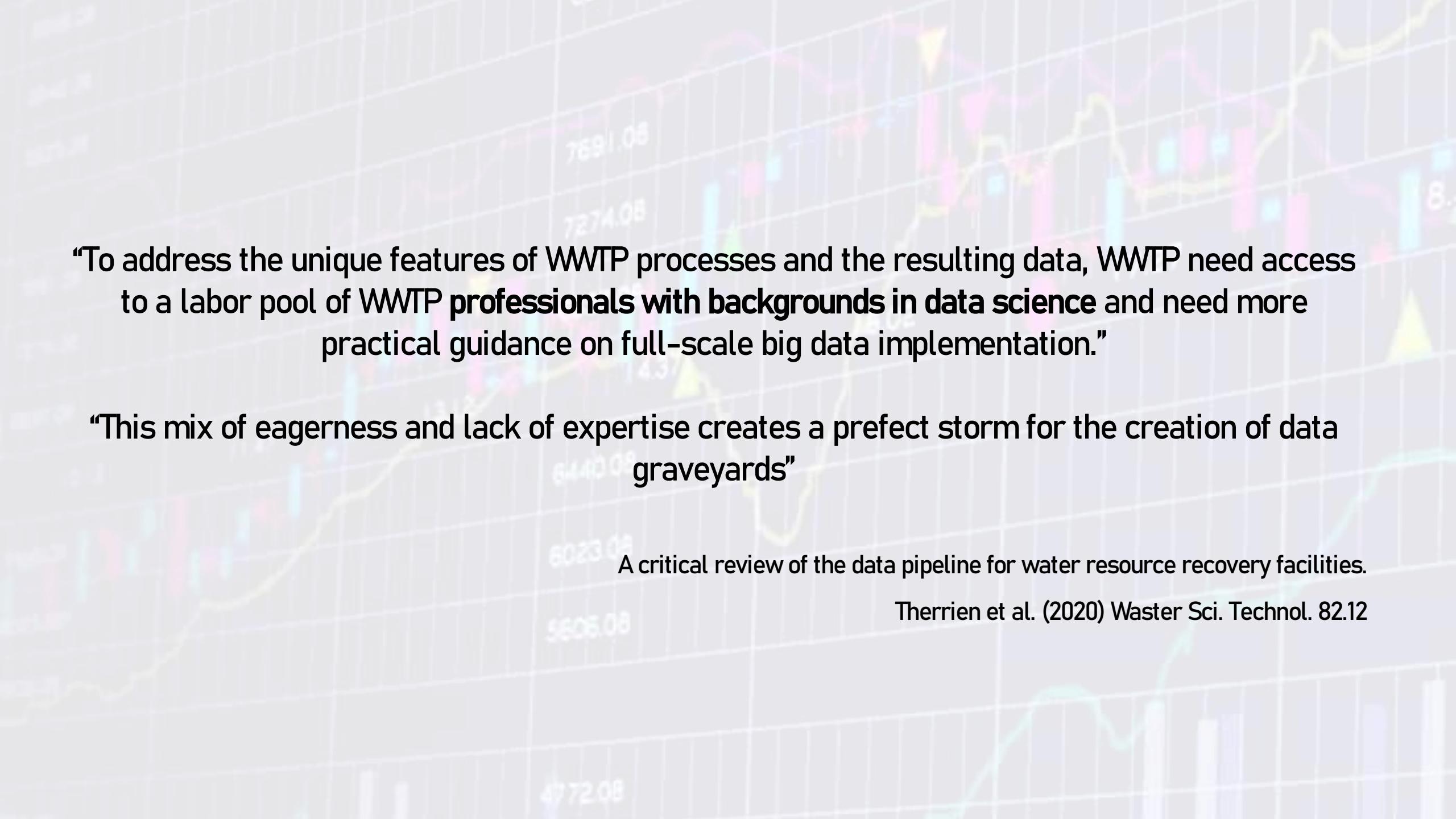
“-Water professionals tend to lack the expertise in computer engineering required to develop adequate ML-derived models;

-Computer engineers do not possess the field-specific expertise needed to develop adequate ML-derived models for water systems on their own either;

-The tools and interfaces delivered to water professionals to explore their data are too difficult to use and lack adequate long-term support for them to be used in practice.”

A critical review of the data pipeline for water resource recovery facilities.

Therrien et al. (2020) Waster Sci. Technol. 82.12



"To address the unique features of WWTP processes and the resulting data, WWTP need access to a labor pool of WWTP professionals with backgrounds in data science and need more practical guidance on full-scale big data implementation."

"This mix of eagerness and lack of expertise creates a perfect storm for the creation of data graveyards"

A critical review of the data pipeline for water resource recovery facilities.

Therrien et al. (2020) Water Sci. Technol. 82.12

Nuevos roles en el sector del agua

Añadir e integrar nuevos dispositivos en los sistemas de control.

Gestionar y optimizar la recogida de datos en función de los distintos objetivos.

Gestionar la calidad de los datos obtenidos a partir de sensores, haciendo uso de herramientas de analítica de datos.

Aumentar los flujos de datos con metadatos, de una manera automatizada.

Homogeneizar los metadatos relativos a la operación de la planta, los algoritmos usados y el mantenimiento de sensores.

Nuevos roles en el sector del agua

Añadir e integrar nuevos dispositivos en los sistemas de control.

Ingenieros de datos

Gestionar y optimizar la recogida de datos en función de los distintos objetivos.

Gestionar la calidad de los datos obtenidos a partir de sensores, haciendo uso de la analítica de datos.

Gestores de datos

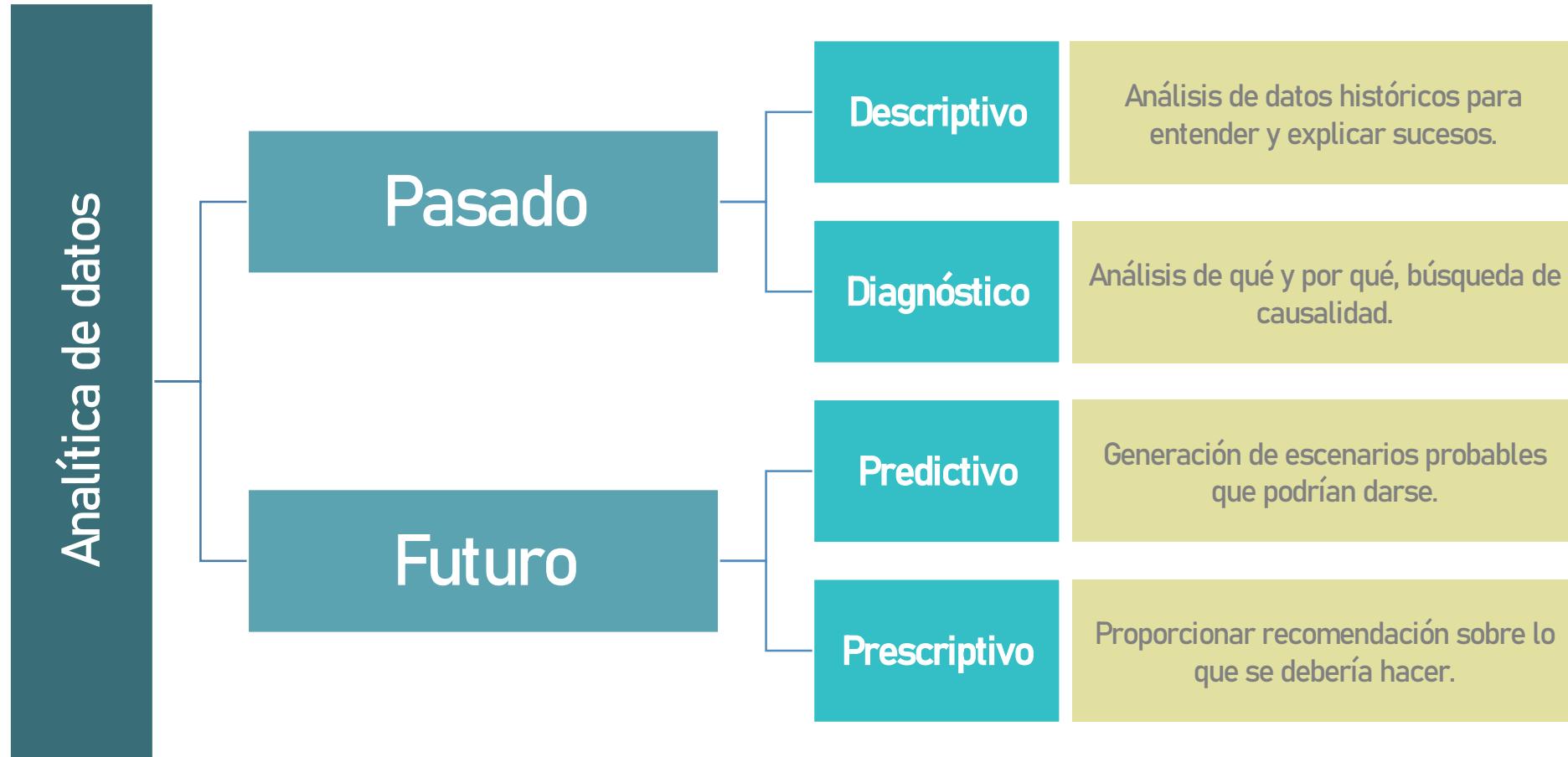
Aumentar los flujos de datos con metadatos, de una manera automatizada.

Homogeneizar los metadatos relativos a la operación de la planta, los algoritmos y el mantenimiento de sensores.

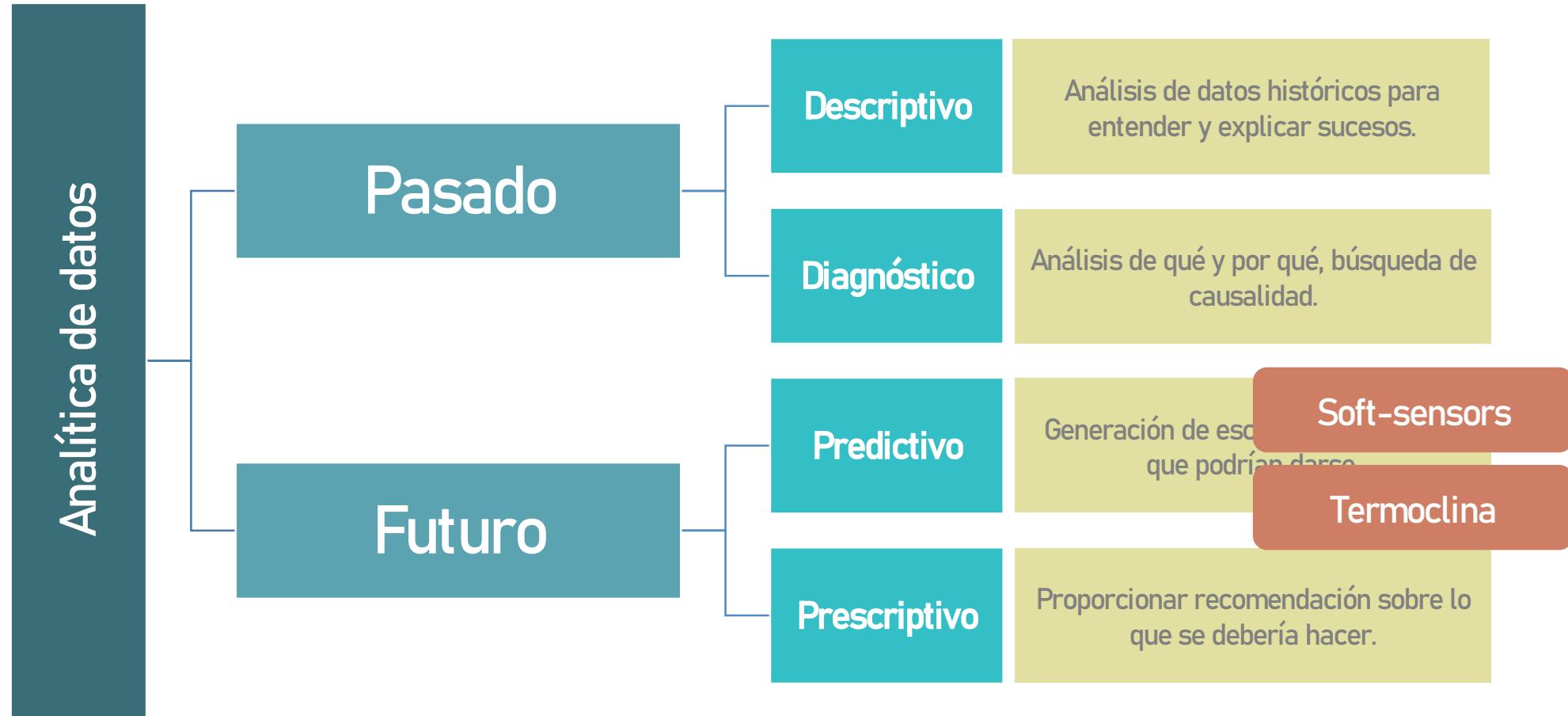
Analistas de datos

Aplicaciones de la ciencia de datos en el sector del agua

Aplicaciones de la ciencia de datos en el sector del agua

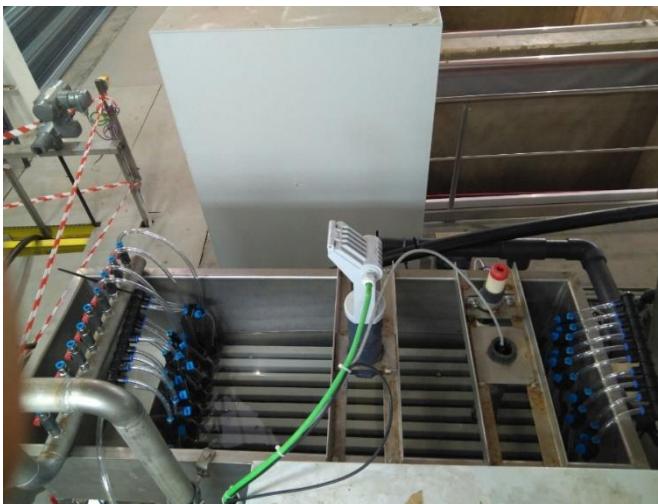


Aplicaciones de la ciencia de datos en el sector del agua



Soft-sensors

Soft-sensors



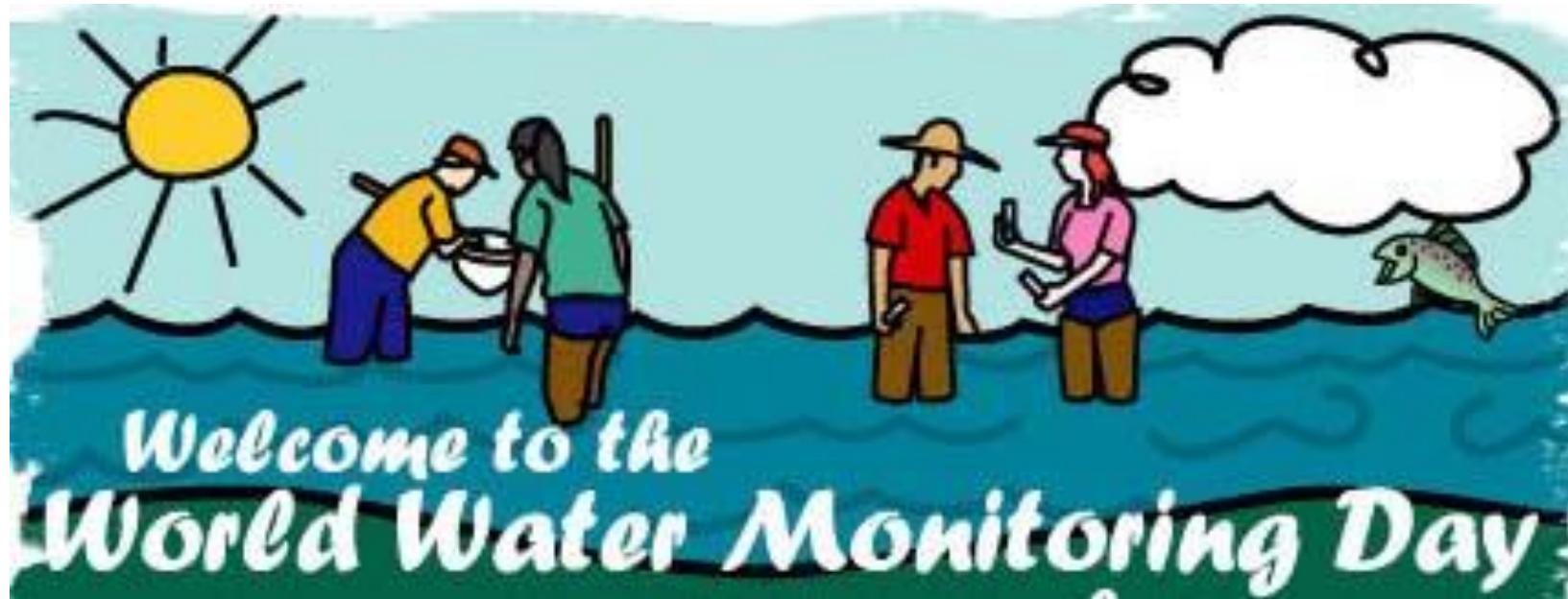
Soft-sensors



MEDIR

- Cuanto más mejor
- Cuanto antes mejor
- Cuanto más preciso mejor

Soft-sensors



Cómo medimos en agua



Quimiometría

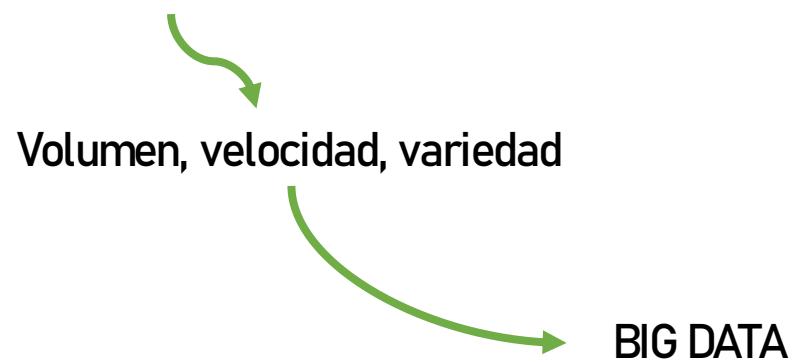
Disciplina que aplica las técnicas matemáticas de la estadística a los problemas analíticos de la identificación y cuantificación de las sustancias químicas.

En la actualidad, esta disciplina ha ganado importancia debido sobre todo por cantidad de información que obtenemos a través de los equipos instrumentales (los cuales generan una **gran cantidad de datos** numéricos) y por el incremento en la **capacidad de cálculo** de los ordenadores actuales.

Quimiometría

Disciplina que aplica las técnicas matemáticas de la estadística a los problemas analíticos de la identificación y cuantificación de las sustancias químicas.

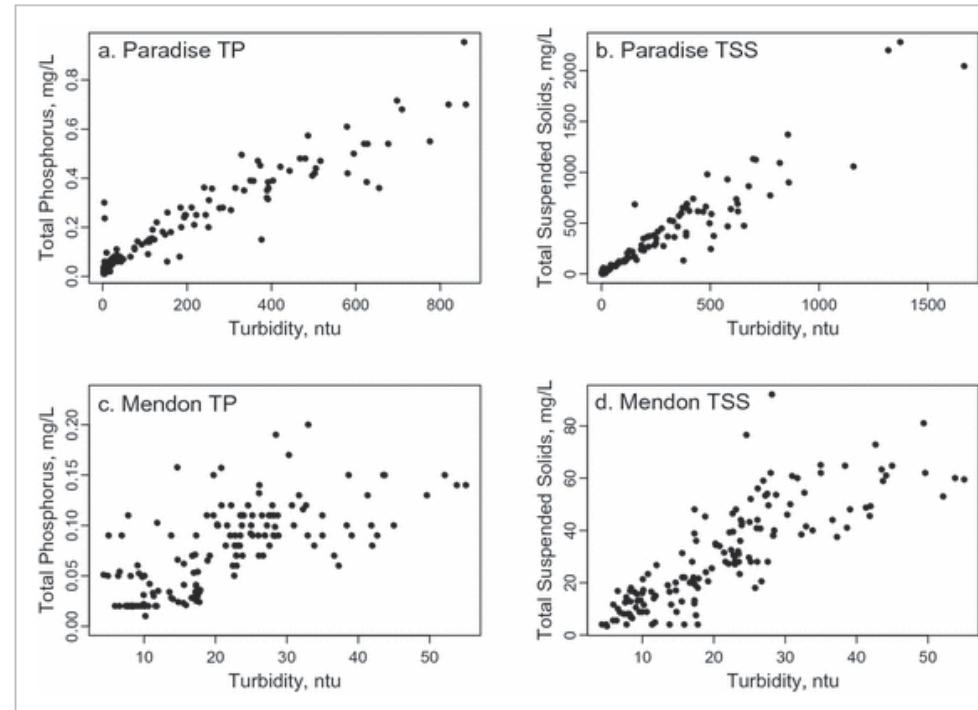
En la actualidad, esta disciplina ha ganado importancia debido sobre todo por cantidad de información que obtenemos a través de los equipos instrumentales (los cuales generan una **gran cantidad de datos** numéricos) y por el incremento en la **capacidad de cálculo** de los ordenadores actuales.



Germán Tortosa Muñoz (2011) Manual práctico de Quimiometría. Digital CSIC.

Quimiometría

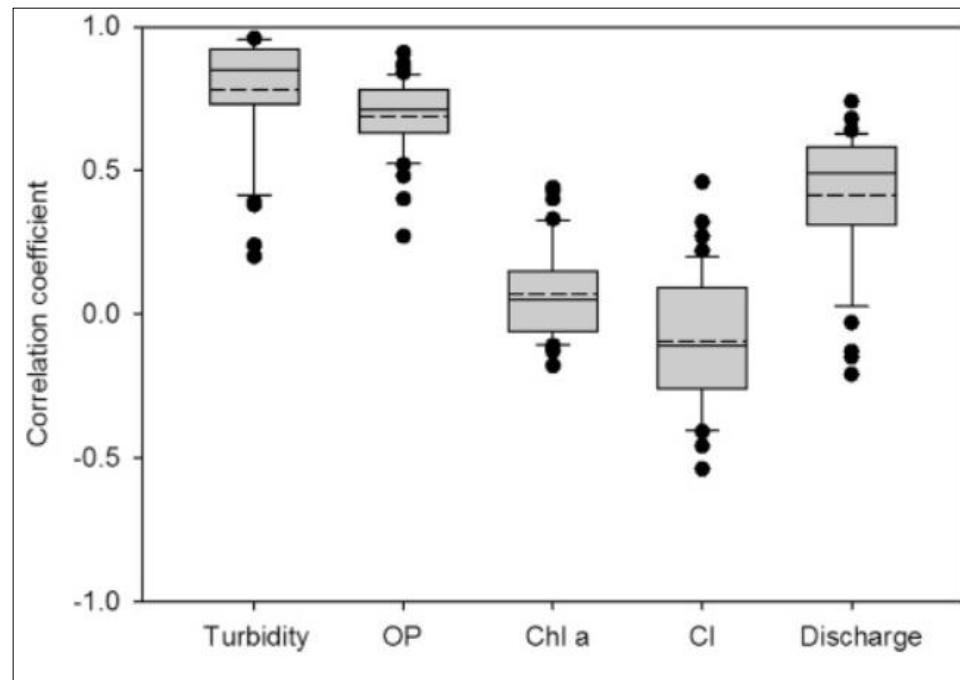
Regresión lineal simple: típicamente relación lineal entre el fósforo total y la turbidez.



Jones, A.S., Stevens, D.K., Horsburgh, J.S., Mesner, N.O., 2011. Surrogate measures for providing high frequency estimates of total suspended solids and total phosphorus Concentrations¹. JAWRA J. Am. Water Res. Assoc. 47 (2), 239e253.
<https://doi.org/10.1111/j.1752-1688.2010.00505.x>.

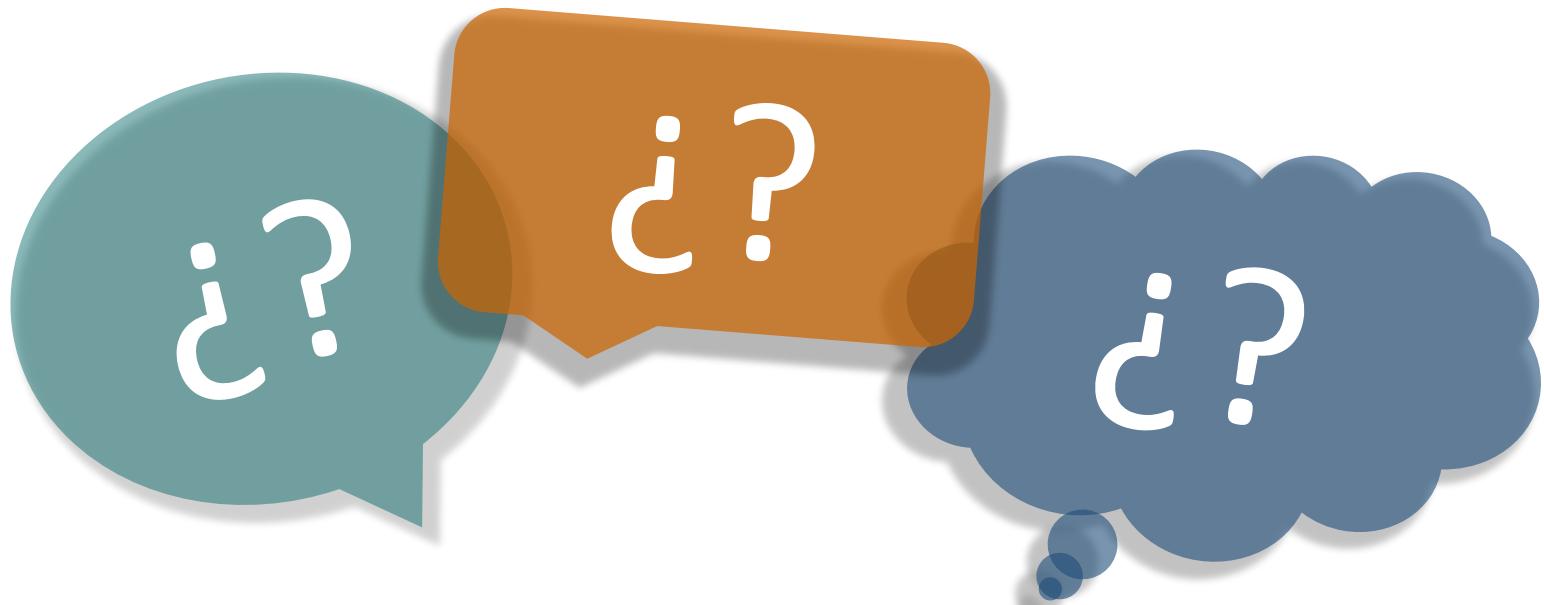
Quimiometría

Regresión lineal múltiple: ejemplo de relación lineal entre el fósforo total y 5 variables subrogadas.



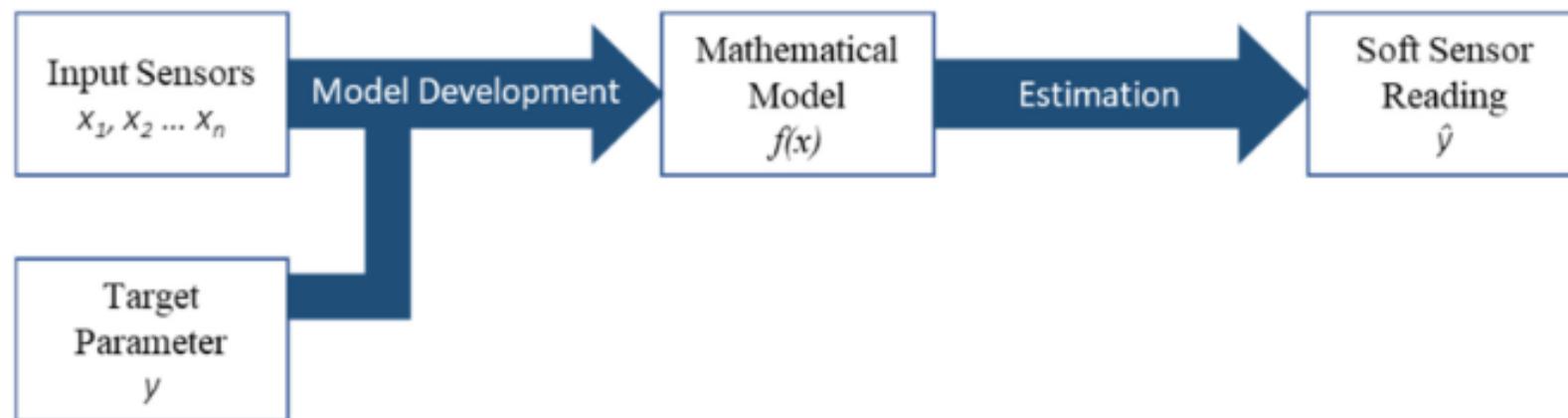
Quimiometría

Regresión no-lineal múltiple: relación no-lineal entre la variable objetivo y varias variables subrogadas.



Soft-sensors

- Sensores basados en software
- Proporcionan una estimación de una variable difícil de cuantificar, utilizando como input variables fáciles de medir, como datos procedentes de sensores robustos, de bajo coste y ampliamente disponibles.



Soft-sensors



Recolección de datos
(proceso y medio ambiente)

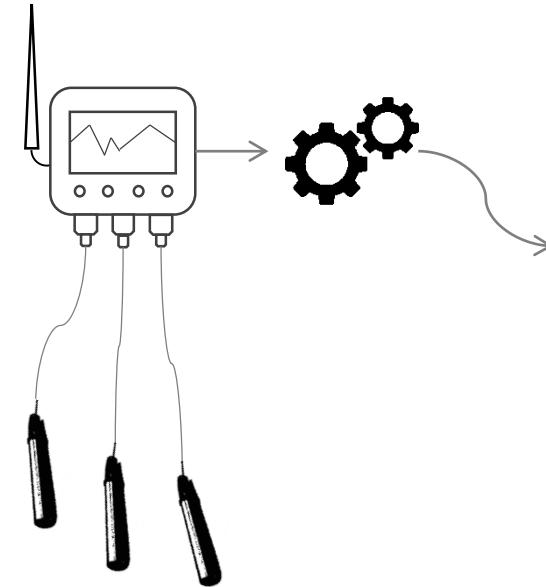
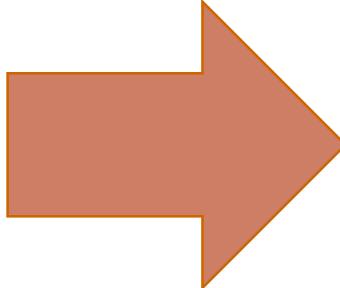
Entrenamiento y
calibrado de modelos
de machine learning

Despliegue de soft-sensors

Determinación y
visualización en
continuo de
concentraciones de
contaminantes

Notificaciones de
alerta y activación de
respuesta

Soft-sensors



In-situ and high frequency
surrogate measures

- Nitrogen
- Phosphorous
- Chlorophyl
- Etc.

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



<https://www.miteco.gob.es/es/agua/temas/concesiones-y-autorizaciones/vertidos-de-aguas-residuales/que-es-vertido/>



<https://www.laverdad.es/murcia/regantes-pagaran-agua-20191009084113-nt.html>

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

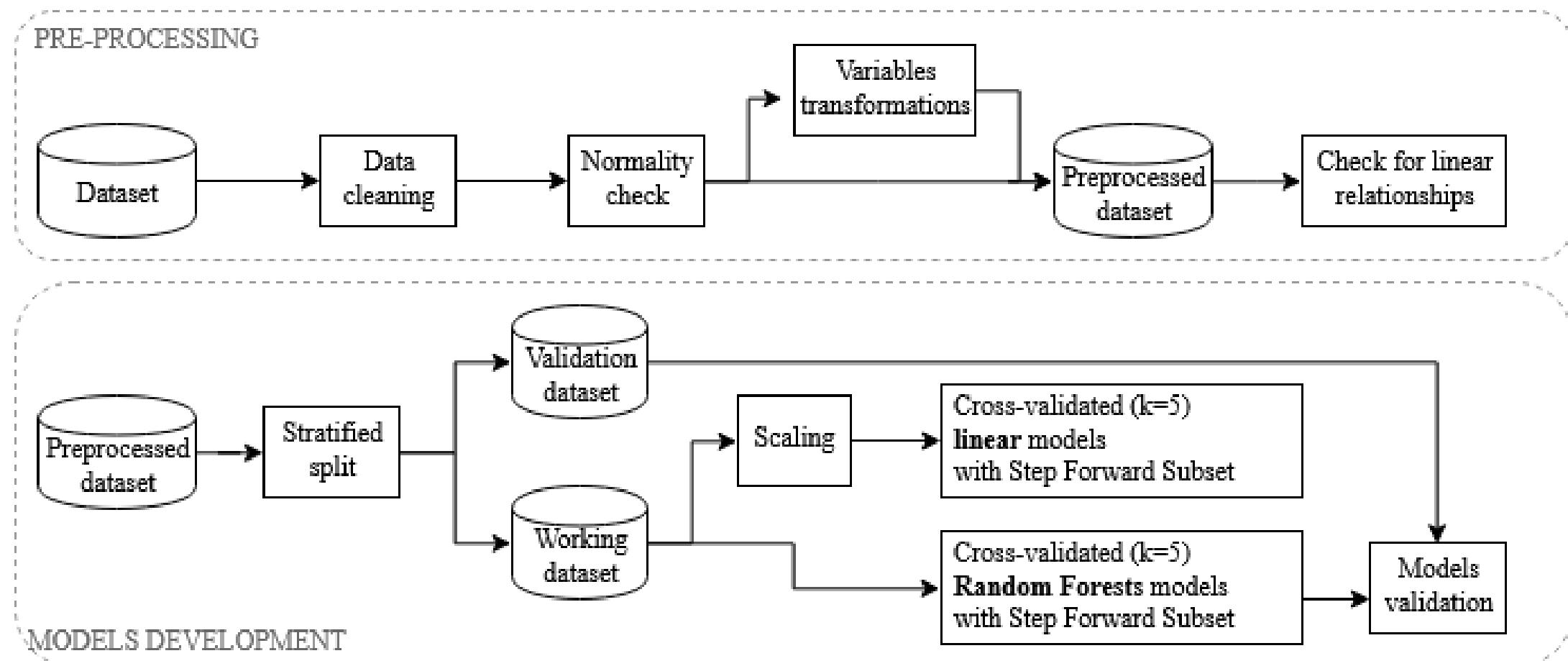


<https://www.ospar.org/work-areas/hasec/eutrophication>

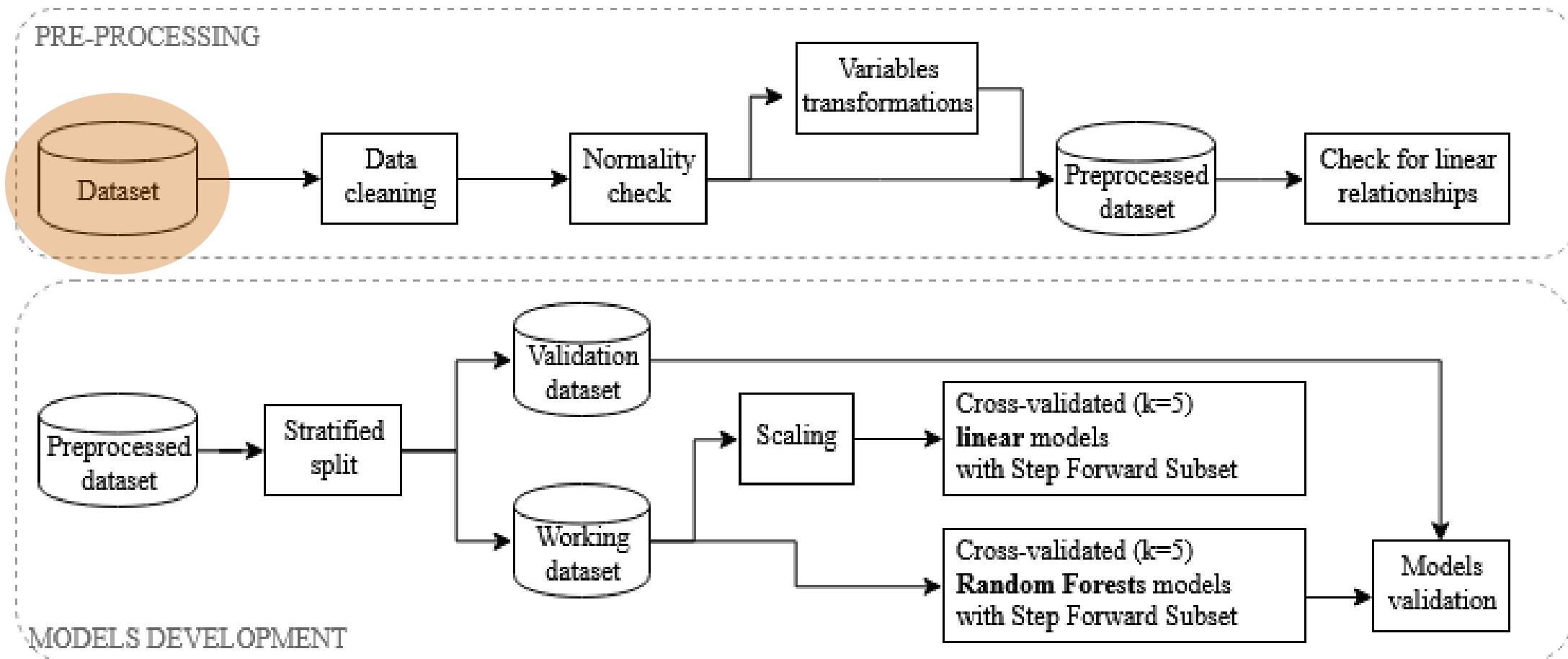


<https://ec.europa.eu/jrc/en/science-update/fighting-marine-eutrophication-requires-nutrient-input-reductions>

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



The Environmental Information Data Centre (EIDC) is a NERC Data Centre hosted by the [UK Centre for Ecology & Hydrology](#) (UKCEH). We manage nationally-important datasets concerned with the terrestrial and freshwater sciences.



Latest updates

EIDC @EIDC_data Just published: Topsoil physico-chemical properties from the UKCEH Countryside Survey, Great Britain, 2019 doi.org/10.5285/aaada1...



36m

EIDC @EIDC_data Now available from the EIDC: United Kingdom Butterfly Monitoring Scheme: phenology 2019. doi.org/10.5285/4bc7bb...

Oct 2, 2020

EIDC @EIDC_data

Find & use data

Find and access the data & services we provide in our data catalogue

Deposit data

We'll guide you through the deposit process with a dedicated team of Data Centre Operatives

Support

We provide support to NERC-funded researchers in data management planning and deposit to an appropriate data centre

Citing Data

Gain credit for your work

About

Who we are and what we do

Contact us

Get in touch for help and advice

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Bowes, M. J. et al/

Hourly physical and nutrient monitoring data for the River Enborne, Berkshire (2009-2012)

<https://doi.org/10.5285/11d712e0-7456-4ea9-8af8-fe81a666e91b>

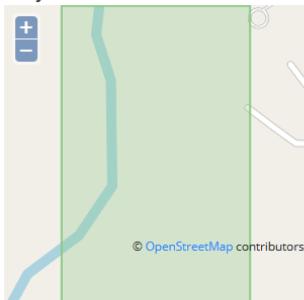
Cite this dataset

This data set comprises of hourly physical and nutrient monitoring data of the River Enborne near Brimpton (National grid reference SU568648), from November 2009 to February 2012. Parameters measured are total reactive phosphorus, nitrate, conductivity, temperature, dissolved oxygen, pH and total chlorophyll. The accompanying hourly averaged flow data (from the EA flow gauging station at the same site) are also supplied. The monitoring programme was funded by the EPSRC, through the LIMPIDS project.

Publication date: 2015-02-17

Where/When

Study area



Provenance & quality

The hourly nutrient data was produced by the Centre for Ecology & Hydrology (CEH) fieldwork staff at Wallingford. Phosphorus data was produced in the field using a Systea Micromac C auto-analyser. Nitrate was monitored by UV absorption using a Hach Lange Nitratex Plus probe. The other physical parameters and total chlorophyll were monitored using a YSI 6600 sonde. The monitoring site installation, telemetry / data collection and maintenance / calibration of the YSI probe was conducted by the Environment Agency's National Water Quality Instrumentation Service. Data processing was carried out by CEH and the University of Reading. Full details of the field procedures

1 citation



Get the data

- Download the data
- Supporting documentation

Licence terms and conditions apply

Format of the data: Comma-separated values (CSV)

You must cite: Bowes, M. J.; Gozzard, E.; Newman, J.; Loewenthal, M.; Halliday, S.; Skeffington, R. A.; Jarvie, H. P.; Wade, A.; Palmer-Felgate, E. (2015). Hourly physical and nutrient monitoring data for the River Enborne, Berkshire (2009-2012). NERC Environmental Information Data Centre.
<https://doi.org/10.5285/11d712e0-7456-4ea9-8af8-fe81a666e91b>

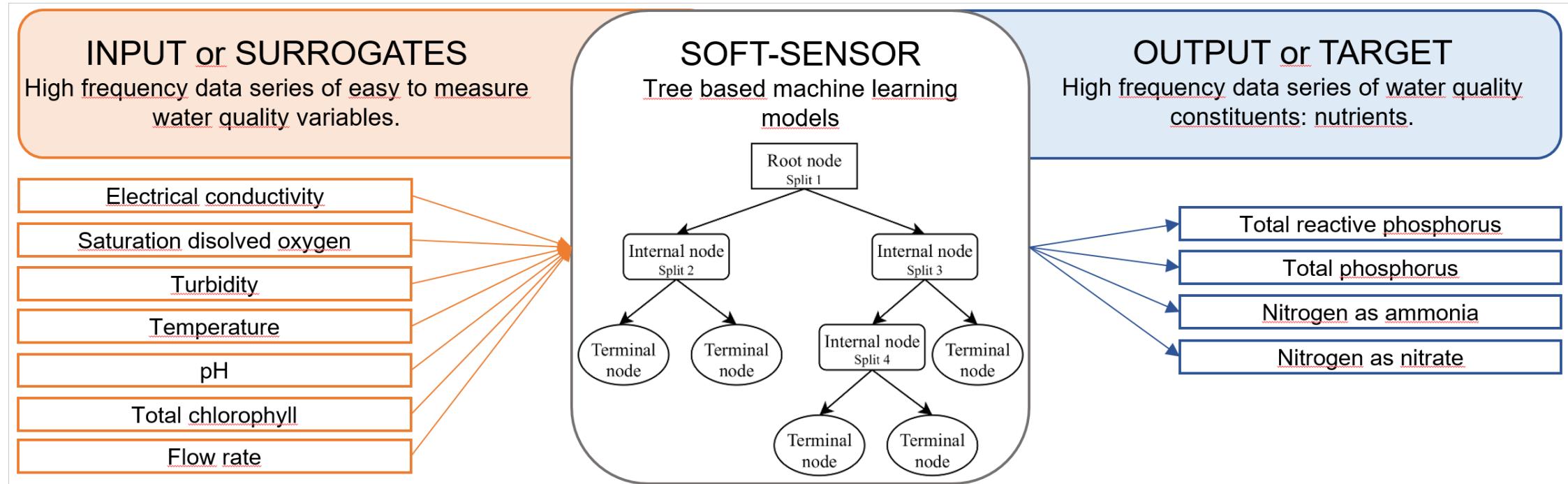
[BibTeX](#) [RIS](#)

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

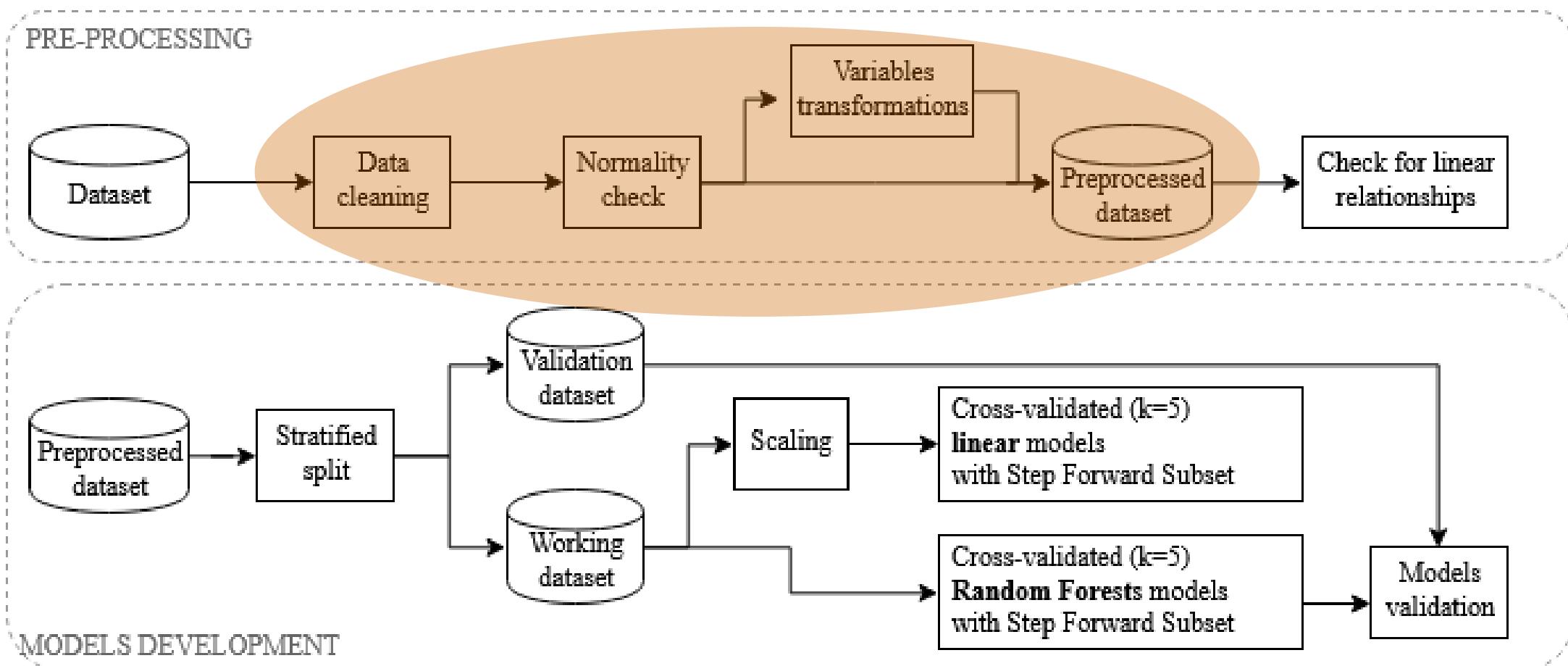
Date / time,Flow (hourly average)(m³/s),Total Reactive Phosphorus (mg P/l),Conductivity (uS/cm),Nitrate (mg NO₃ /l),Turbidity (NTU),Total chlorophyll (ug/l),Dissolved Oxygen (% saturation),pH,Temperature (C)

01/11/2009 00:00,0.381,NaN,NaN,NaN,5.9,3.6,62,7.68,12.4
01/11/2009 01:00,0.376,NaN,533,NaN,6.1,3.5,64,7.71,12.4
01/11/2009 02:00,0.366,NaN,532,NaN,6.3,7,64,7.7,12.4
01/11/2009 03:00,0.357,NaN,531,NaN,5.9,3.7,65,7.73,12.4
01/11/2009 04:00,0.349,NaN,534,NaN,5.9,3.4,65,7.71,12.4
01/11/2009 05:00,0.342,NaN,538,NaN,5.8,3.6,64,7.71,12.4
01/11/2009 06:00,0.335,NaN,540,NaN,5.8,3,64,7.73,12.4
01/11/2009 07:00,0.311,NaN,538,NaN,5.8,3.5,63,7.7,12.3
01/11/2009 08:00,0.313,NaN,532,NaN,5.2,3.3,64,7.7,12.4
01/11/2009 09:00,0.32,NaN,523,NaN,5.3,3.8,64,7.73,12.5
01/11/2009 10:00,0.325,NaN,514,NaN,5.3,3.9,65,7.71,12.6
01/11/2009 11:00,0.321,NaN,506,NaN,5.2,4.3,65,7.7,12.9
01/11/2009 12:00,0.33,NaN,498,NaN,5.2,4.1,65,7.69,12.9
01/11/2009 13:00,0.338,NaN,490,NaN,5.1,4.1,64,7.67,12.9
01/11/2009 14:00,0.358,NaN,480,NaN,5.2,4.9,64,7.68,13
01/11/2009 15:00,0.423,NaN,471,NaN,5.8,4.9,64,7.65,12.9
01/11/2009 16:00,0.598,NaN,470,NaN,6.1,4.2,63,7.64,12.8
01/11/2009 17:00,0.768,NaN,475,NaN,7.9,4.8,61,7.62,12.6
01/11/2009 18:00,0.93,NaN,476,NaN,9.7,4.5,60,7.6,12.5
01/11/2009 19:00,1.061,NaN,471,NaN,12.5,61,7.58,12.4
01/11/2009 20:00,1.168,NaN,455,NaN,15.9,4.6,63,7.63,12.4
01/11/2009 21:00,1.266,NaN,457,NaN,19.5,1,64,7.62,12.3
01/11/2009 22:00,1.341,NaN,460,NaN,22.3,5.3,64,7.63,12.1
01/11/2009 23:00,1.356,NaN,450,NaN,23.3,5.7,65,7.65,12
02/11/2009 00:00,1.317,NaN,439,NaN,22.8,5.4,67,7.64,11.9
02/11/2009 01:00,1.254,NaN,413,NaN,22.3,5.3,67,7.62,11.8
02/11/2009 02:00,1.171,NaN,404,NaN,21.3,5.3,67,7.59,11.7

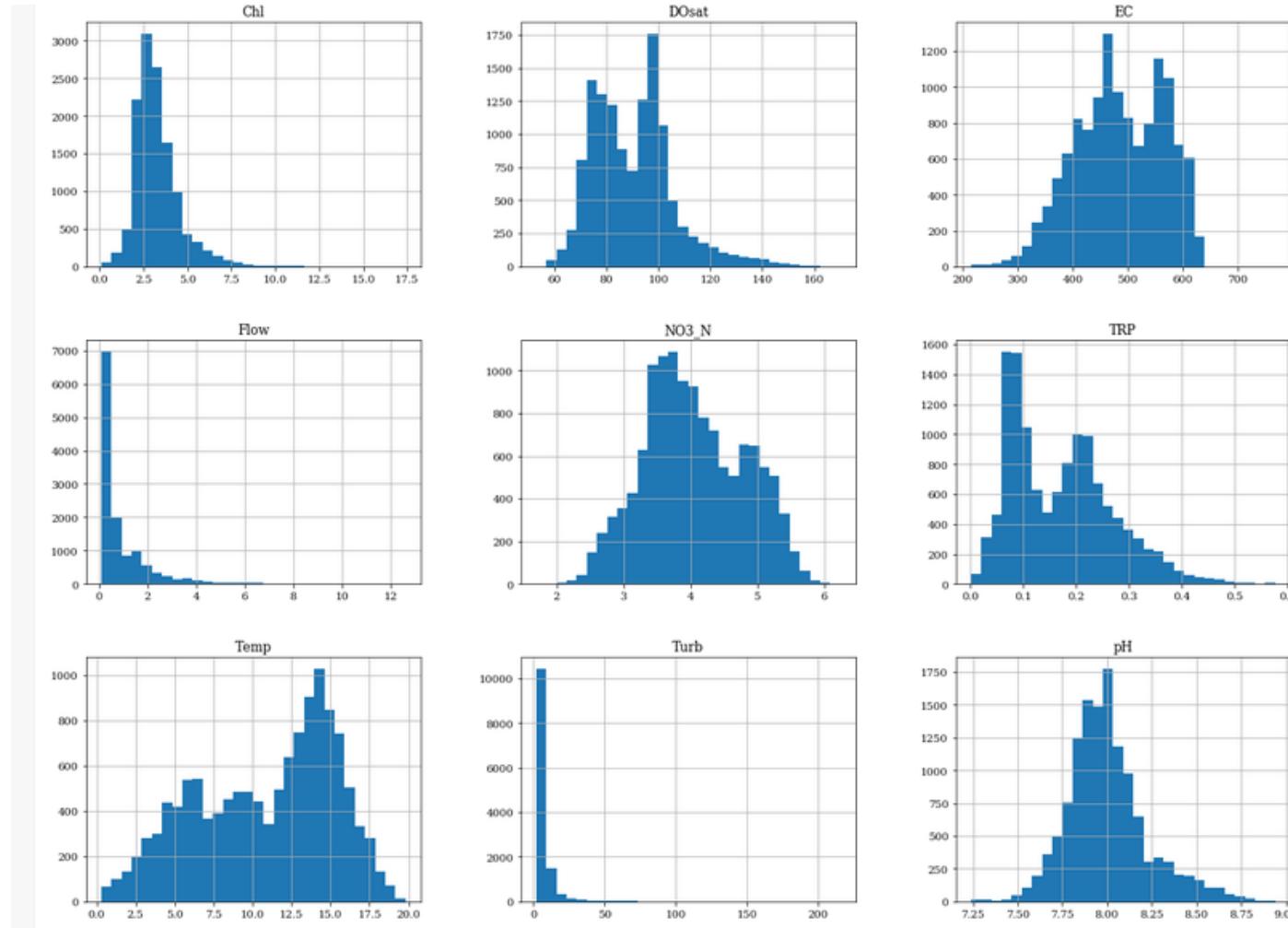
Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

- Datasets curados por la Universidad de Reading:
 - Revisión de puntuación, ortografía, unidades, etc.
 - Sustitución o eliminación de *outliers*.
- Limpieza NaN
- Comprobación de secuencia temporal
 - ❖ Enborne dataset
 - 20.412 observaciones
 - 12.723 tras eliminar NaN
 - ❖ The Cut dataset
 - 15.636 observaciones
 - 8934 tras eliminar NaN



Fig. 3. Photographs of the river monitoring stations. River systems (top row – taken in March 2010): (a) River Enborne at Brimpton (Great Britain National Grid Reference, SU56776477), (b) River Kennet at Clatford (SU16016885) and (c) The Cut at Bracknell (SU85837050). Monitoring stations (bottom row – taken in April 2010): (a) River Enborne at Brimpton (SU56776477), (b) River Kennet at Clatford (SU16016885) and (c) The Cut at Bray (SU91397866).

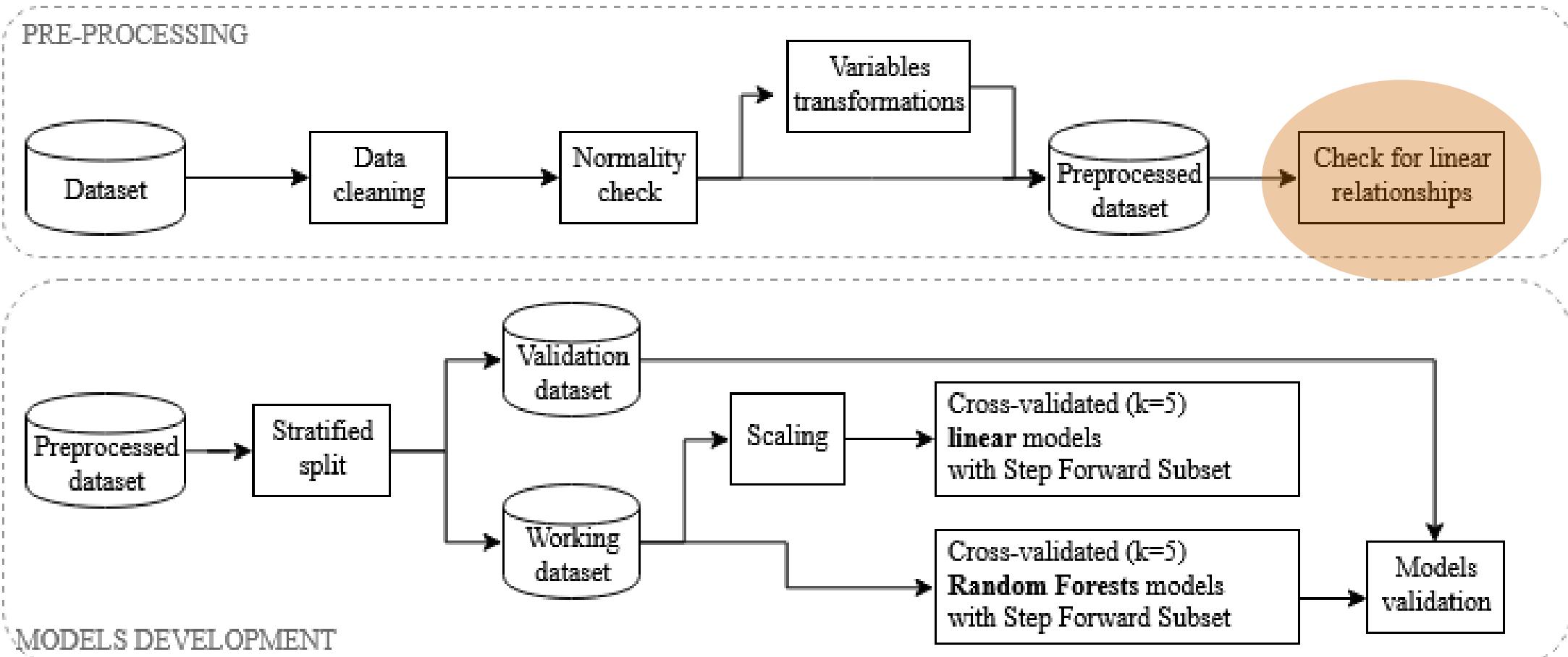
Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

Table 1

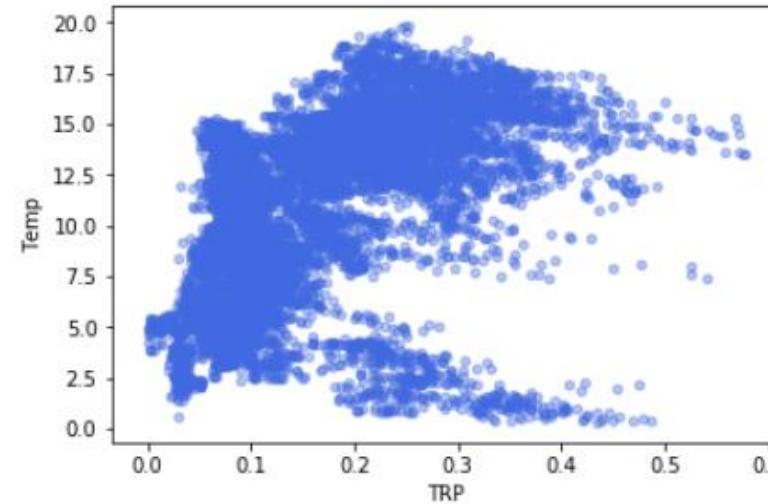
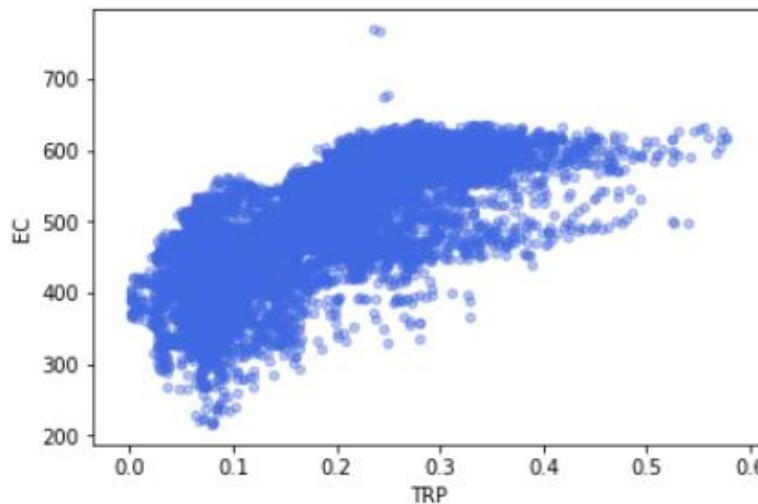
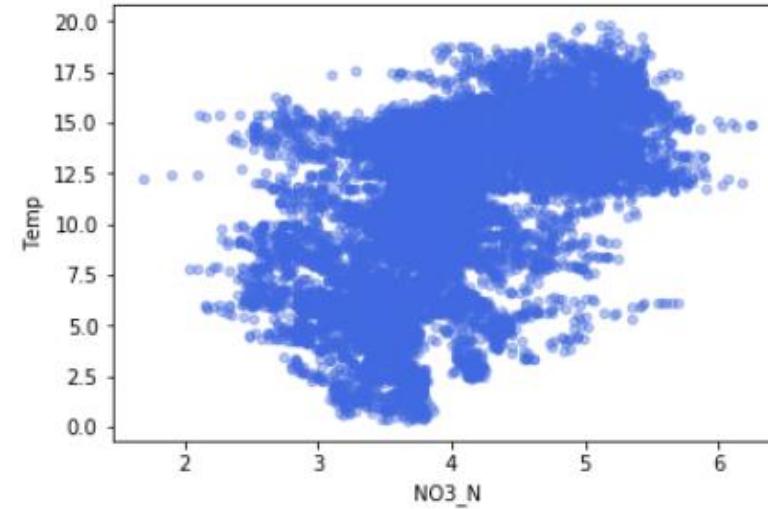
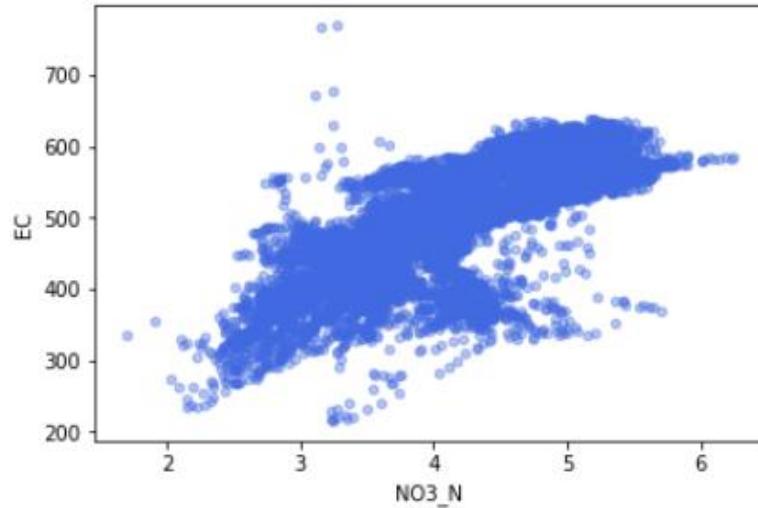
Mean, standard deviation (sd), minimum value (Min) and maximum value (Max) of the variables contained in each dataset.

Parameter	Unit	Enborne (n = 12,723)			The Cut (n = 8934)		
		Mean ± sd	Min	Max	Mean ± sd	Min	Max
Flow	$\text{m}^3 \text{s}^{-1}$	0.97 ± 1.32	0.11	12.60	0.24 ± 0.30	0.06	4.64
Temp	°C	10.85 ± 4.41	0.30	19.80	13.48 ± 4.00	2.20	22.9
pH	—	8.00 ± 0.22	7.24	8.94	7.72 ± 0.13	7.23	8.39
DOSat	%	90.28 ± 15.58	53.00	170	76.04 ± 20.20	27.00	176.00
Turb	NTU	8.05 ± 10.28	2.00	215.8	15.47 ± 74.09	0.70	1255.40
EC	$\mu\text{S cm}^{-1}$	486.53 ± 79.64	215	769	958.88 ± 97.73	367.00	1115.00
Chl	$\mu\text{g L}^{-1}$	3.30 ± 1.33	0.10	17.4	4.51 ± 4.76	0.80	92.9
TRP	mg P L^{-1}	0.17 ± 0.10	0.00	0.58	0.59 ± 0.22	0.15	1.45
TP	mg P L^{-1}	—	—	—	0.64 ± 0.21	0.22	1.49
NO_3-N	mg N L^{-1}	4.06 ± 0.75	1.70	6.25	—	—	—
NH_4-N	mg N L^{-1}	—	—	—	0.60 ± 0.24	0.03	1.43

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

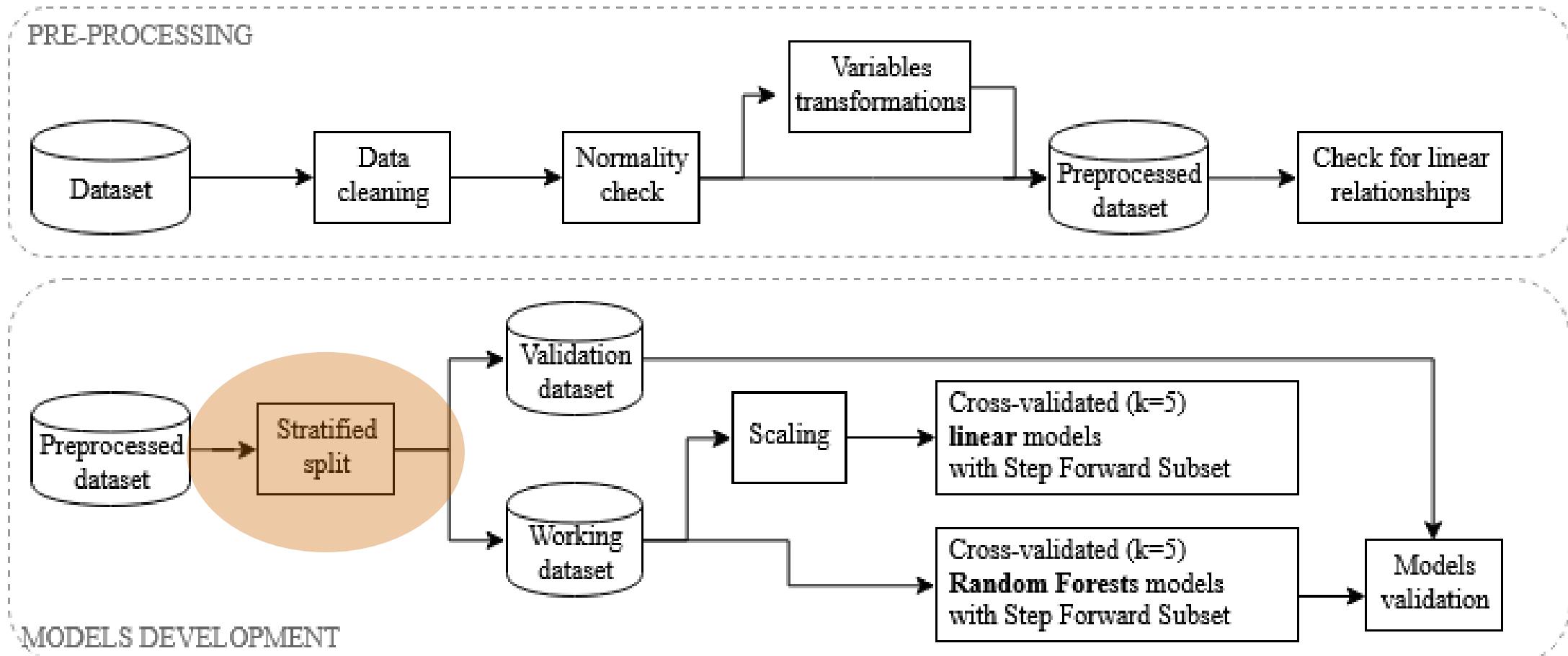


Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

Pearson's r for each of the target variables with respect to the seven predictors, ordered by the absolute value of r .

River Enborne		The Cut							
TRP	NO ₃ -N	TRP	TP	NH ₄ -N	Temp	Chl	Turb	Flow	
EC	0.72	EC	0.76	EC	0.44	EC	0.44	Temp	0.29
Flow	-0.67	Temp	0.55	Flow	-0.30	Flow	-0.29	Chl	-0.17
DOsat	-0.60	Flow	-0.54	DOsat	-0.22	DOsat	-0.22	Turb	0.16
Temp	0.52	Chl	-0.33	Turb	-0.15	Turb	-0.12	Flow	-0.16
Chl	-0.37	DOsat	-0.31	Chl	-0.12	Temp	0.10	EC	0.12
Turb	0.11	pH	0.16	Temp	0.08	Chl	-0.09	DOsat	0.02
pH	-0.11	Turb	-0.06	pH	0.02	pH	0.01	pH	0.00

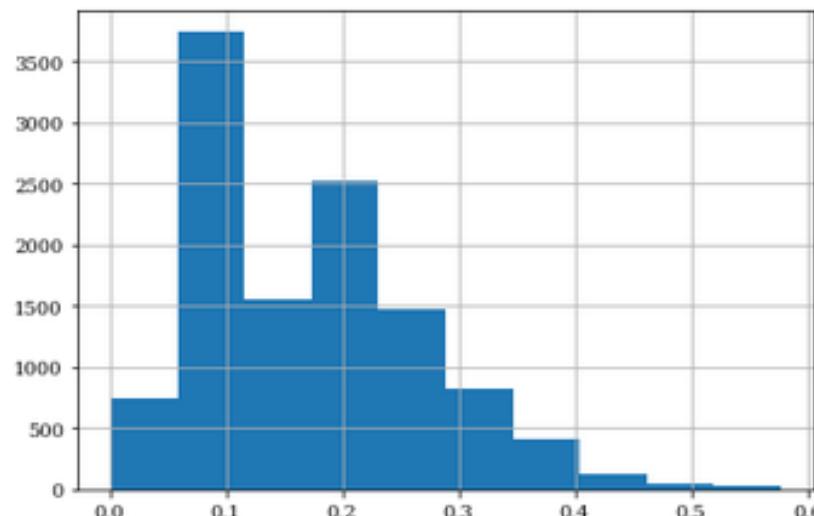
Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



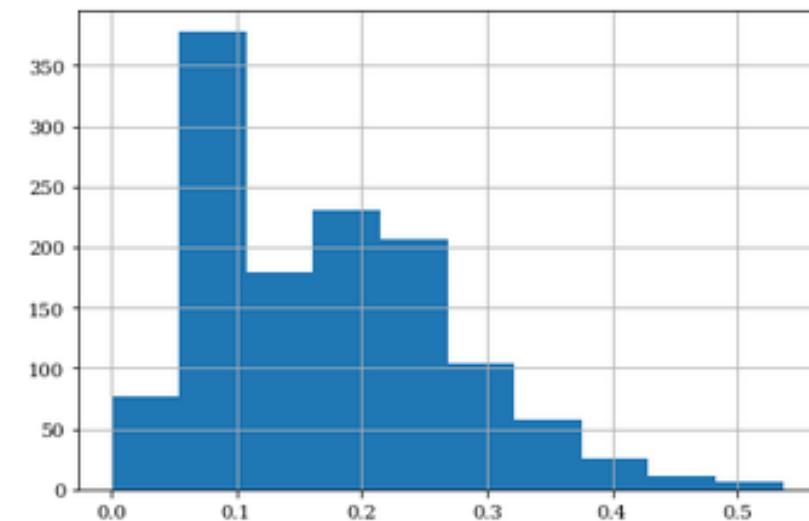
Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Partición estratificada de la fracción de test:

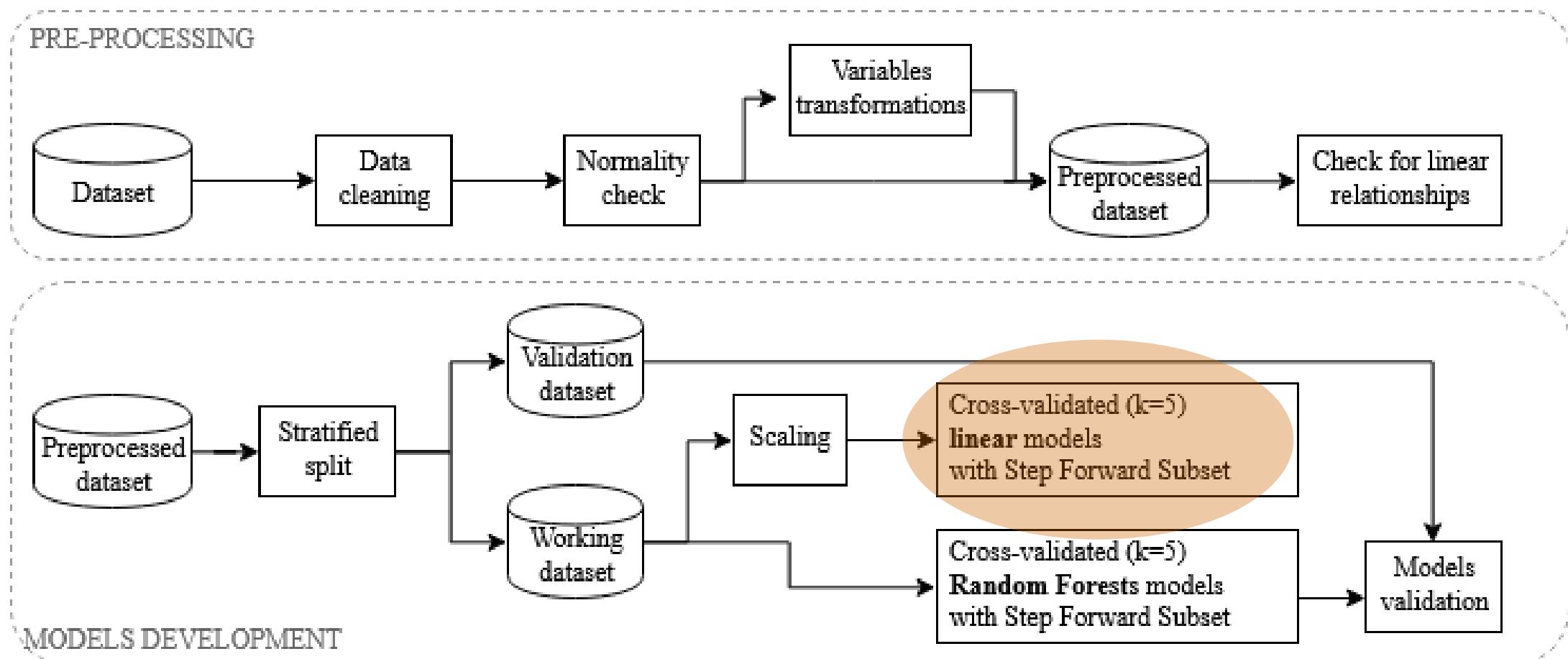


Histograma de TRP en fracción de trabajo (train+val) (90%).



Histograma de TRP en fracción de test (10%).

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

¿Por qué es interesante seleccionar las variables?

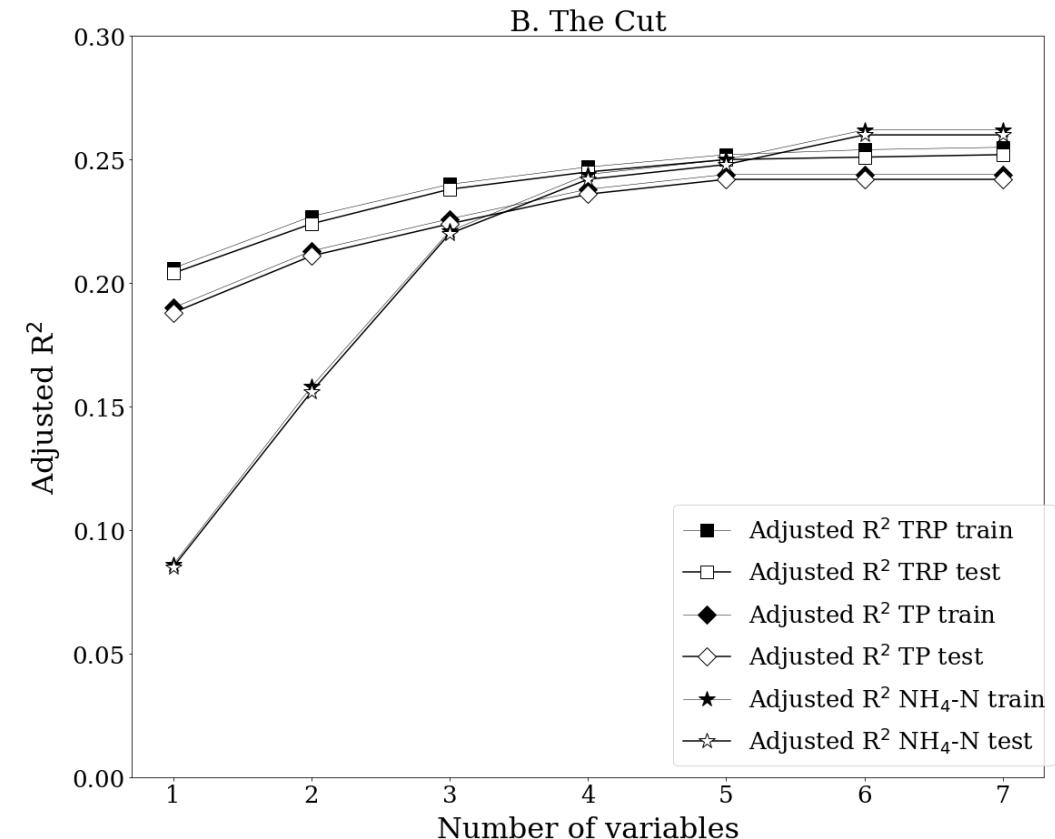
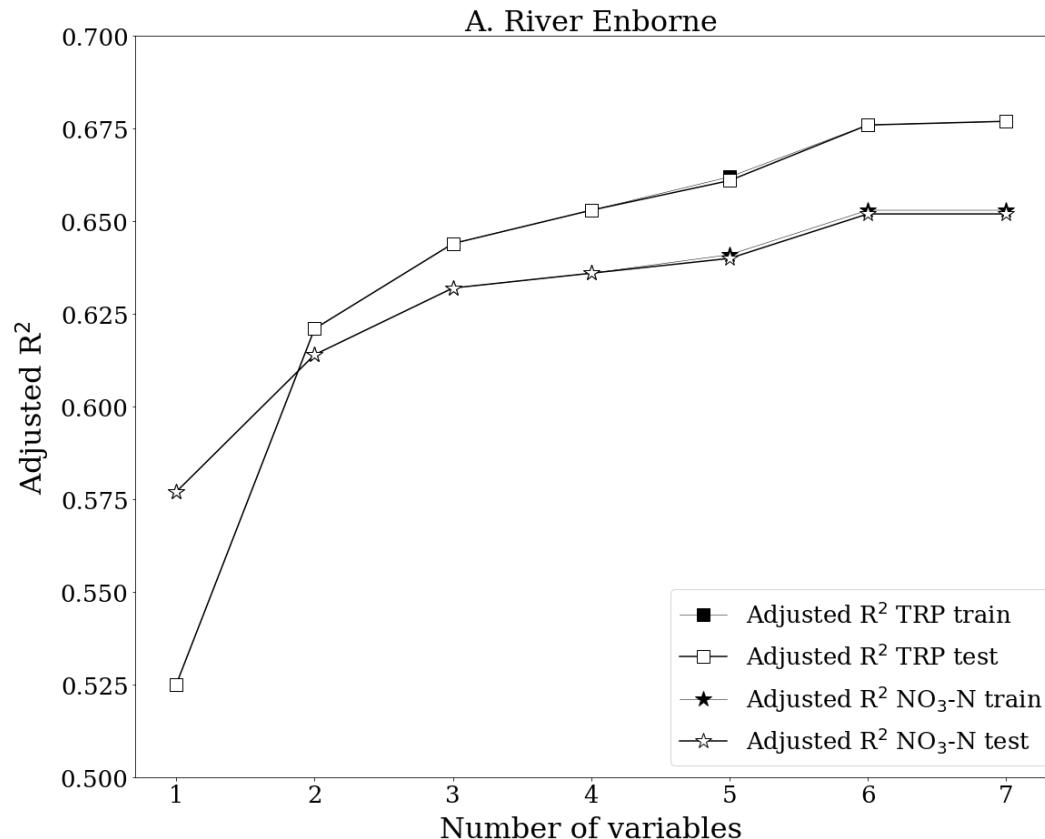
- Modelos más sencillos
- Con menos tendencia al sobreajuste
- Más generalizables
- Más interpretables
- Menos demandantes de recursos computacionales

Además, en nuestro caso particular:

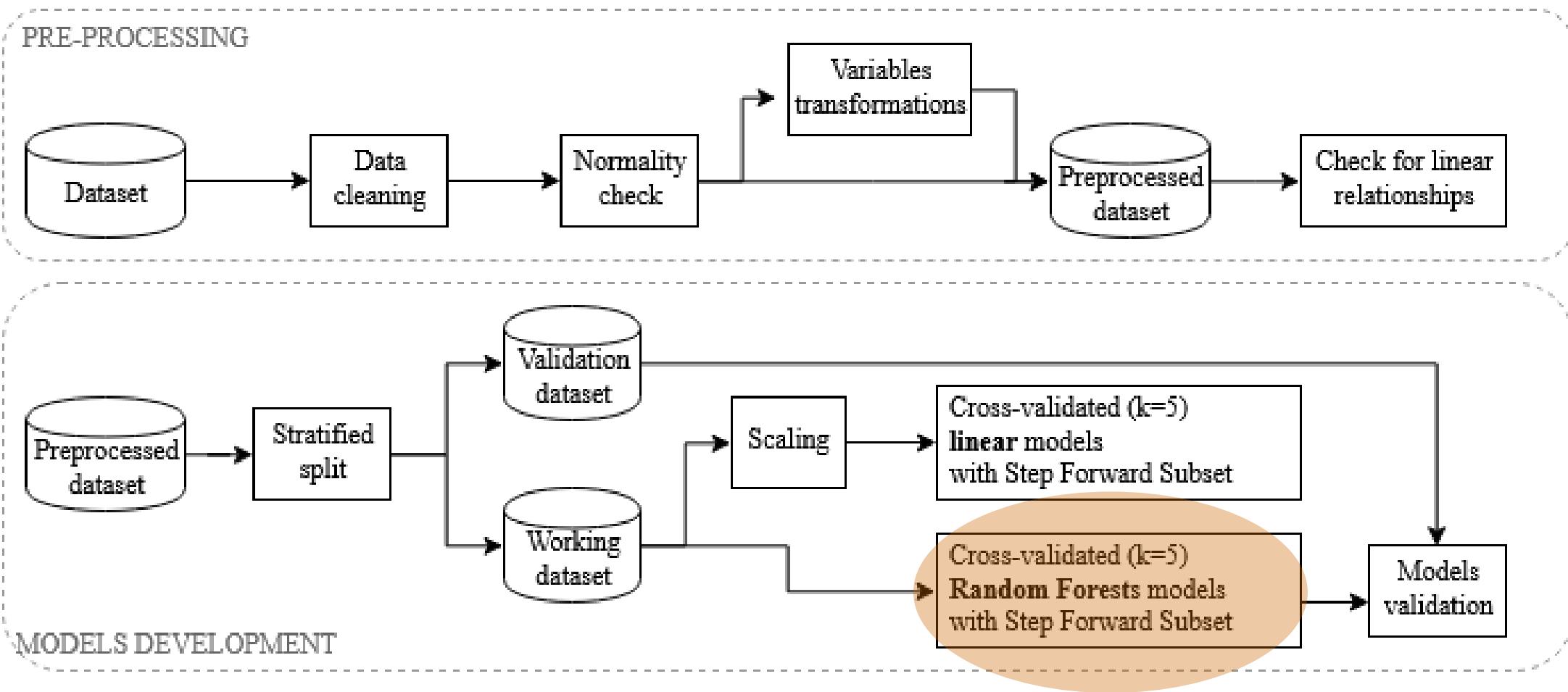
- Soft-sensor más sencillo y económico
 - en inversión inicial
 - en mantenimiento

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

$$y_i(w, x_i) = w_0 + w_1 x_{i,1} + \dots + w_n x_{i,n}$$



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

Elección del modelo a utilizar

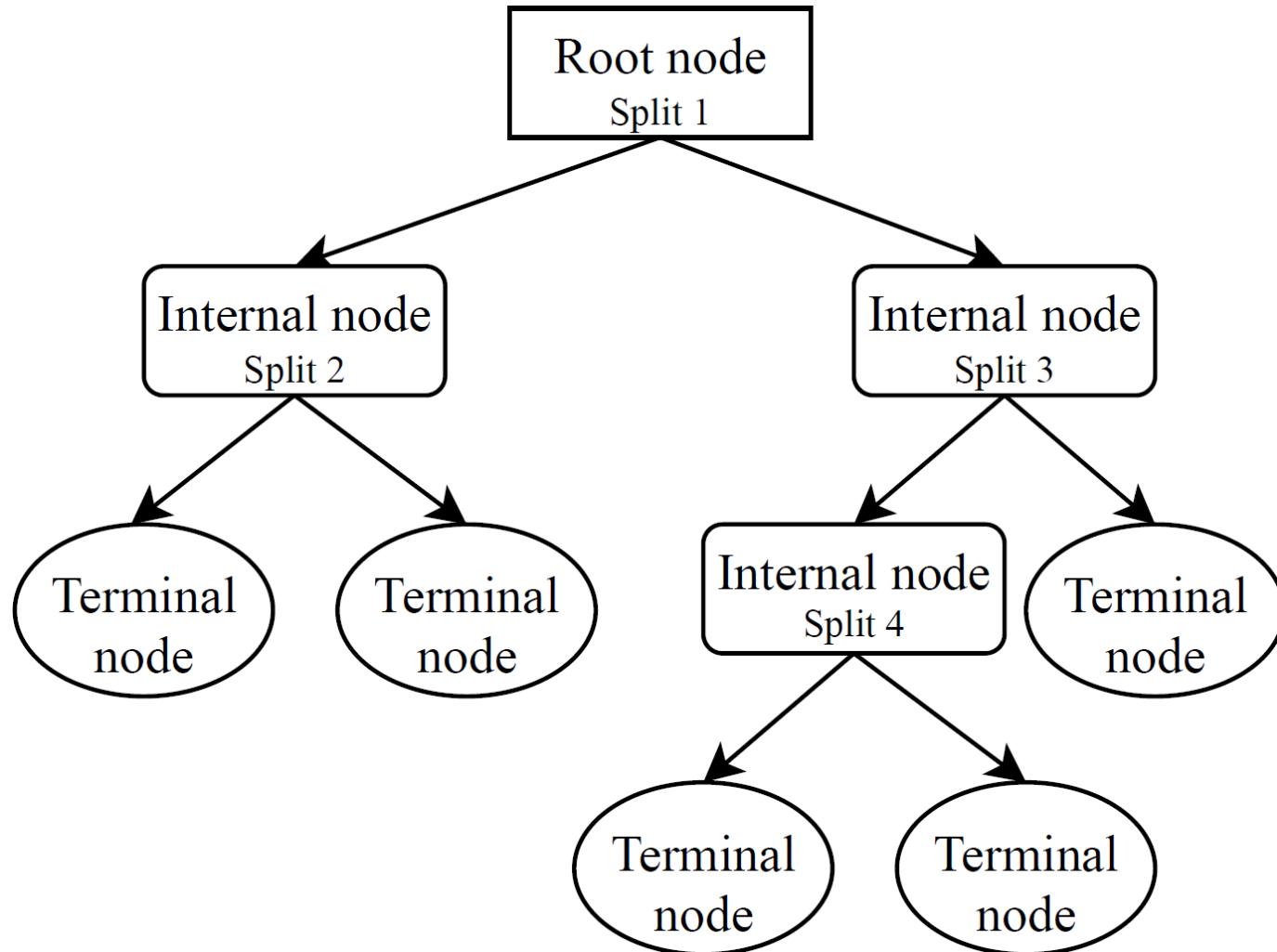
Adecuación del modelo a los datos disponibles.

- Random forests resulta idóneo en datos con diferentes escalas y distribuciones.

Usabilidad por el usuario final

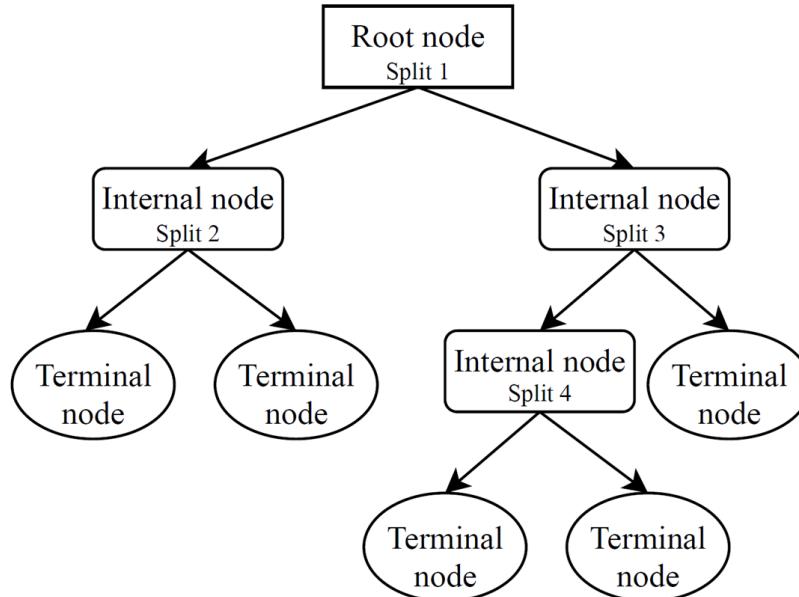
- Relativamente fácil de interpretar.

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

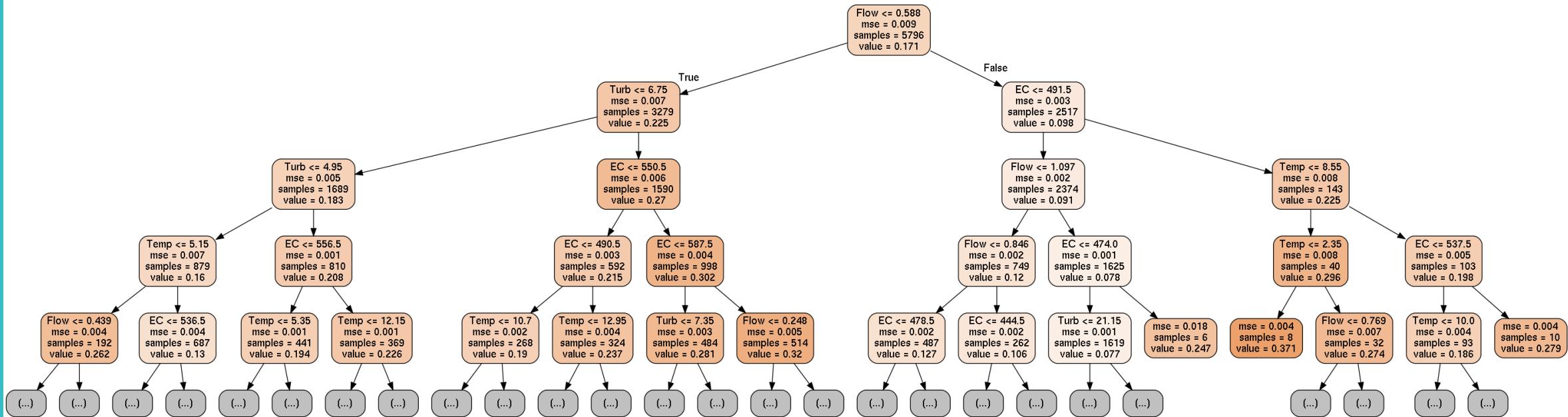


Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

Hyperparameter	Values
Bootstrap	True, False
Size of the random subsets	auto ⁽¹⁾ , sqrt ⁽²⁾ , 2
Depth of the trees	10, 20, 30
Minimum number of samples to split a node	6, 12, 20
Minimum number of samples to be at a leaf node	6, 12, 20



Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

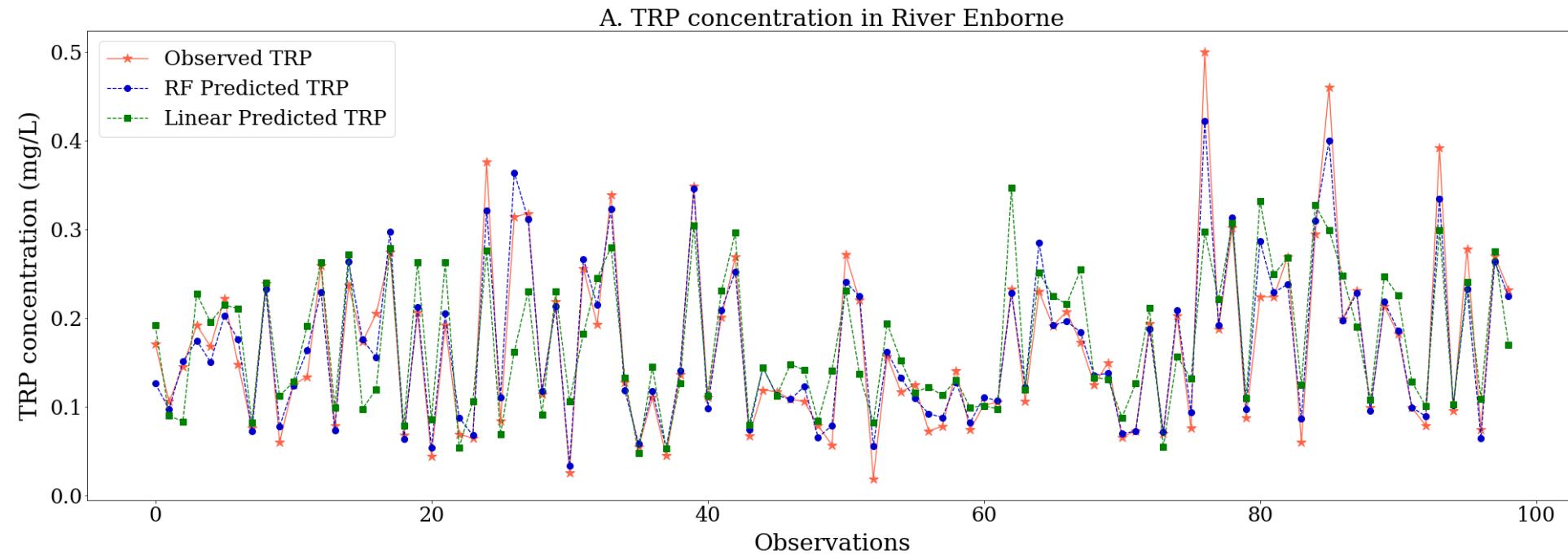


Soft-sensors: prueba de concepto para estimación de nutrientes en ríos

Test RMSE for each target variable and each subset of variables of size one to seven.

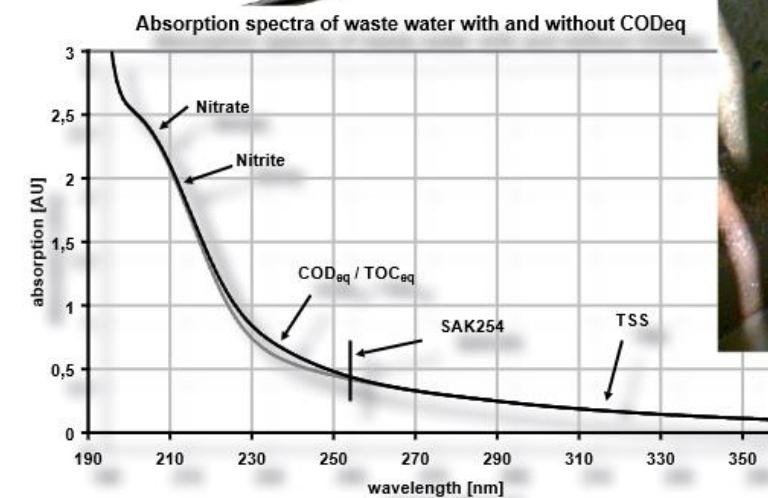
Order of variables	1st	2nd	3rd	4th	5th	6th	7th
TRP in River Enborne							
Linear model	EC	+ pH	+ Turb	+ Chl	+ Flow	+ DOsat	+ Temp
Test RMSE	0.067	0.059	0.058	0.057	0.056	0.056	0.055
RF model	EC	+ Flow	+ Temp	+ Turb	+ pH	+ DOsat	+ Chl
Test RMSE	0.063	0.045	0.035	0.027	0.024	0.023	0.025
NO₃-N in River Enborne							
Linear model	EC	+ Flow	+ Chl	+ Temp	+ pH	+ DOsat	+ Turb
Test RMSE	0.487	0.464	0.454	0.451	0.448	0.441	0.441
RF model	EC	+ pH	+ Flow	+ Temp	+ DOsat	+ Turb	+ Chl
Test RMSE	0.459	0.348	0.253	0.174	0.165	0.161	0.168
TRP in The Cut							
Linear model	EC	+ DOsat	+ pH	+ Turb	+ Flow	+ Chl	+ Temp
Test RMSE	0.194	0.191	0.189	0.188	0.188	0.188	0.188
RF model	EC	+ Turb	+ Temp	+ pH	+ Flow	+ DOsat	+ Chl
Test RMSE	0.191	0.183	0.141	0.122	0.113	0.112	0.106
TP in The Cut							
Linear model	EC	+ DOsat	+ pH	+ Turb	+ Flow	+ Chl	+ Temp
Test RMSE	0.190	0.187	0.186	0.184	0.184	0.184	0.184
RF model	EC	+ Turb	+ Temp	+ pH	+ Flow	+ DOsat	+ Chl
Test RMSE	0.194	0.169	0.143	0.118	0.106	0.105	0.104
NH₄-N in The Cut							
Linear model	Temp	+ Turb	+ Chl	+ EC	+ pH	+ DOsat	+ Flow
Test RMSE	0.228	0.219	0.211	0.208	0.207	0.205	0.205
RF model	Chl	+ Temp	+ Turb	+ pH	+ EC	+ Flow	+ DOsat
Test RMSE	0.214	0.189	0.149	0.121	0.102	0.116	0.098

Soft-sensors: prueba de concepto para estimación de nutrientes en ríos



Soft-sensors: datos espectrométricos

- Uso de espectros de absorción UV.
- Explotar relación entre la absorción a determinadas longitudes de onda y la concentración de la especie de interés.
- Posibilidad de procesar los datos: derivada, segunda derivada, etc.

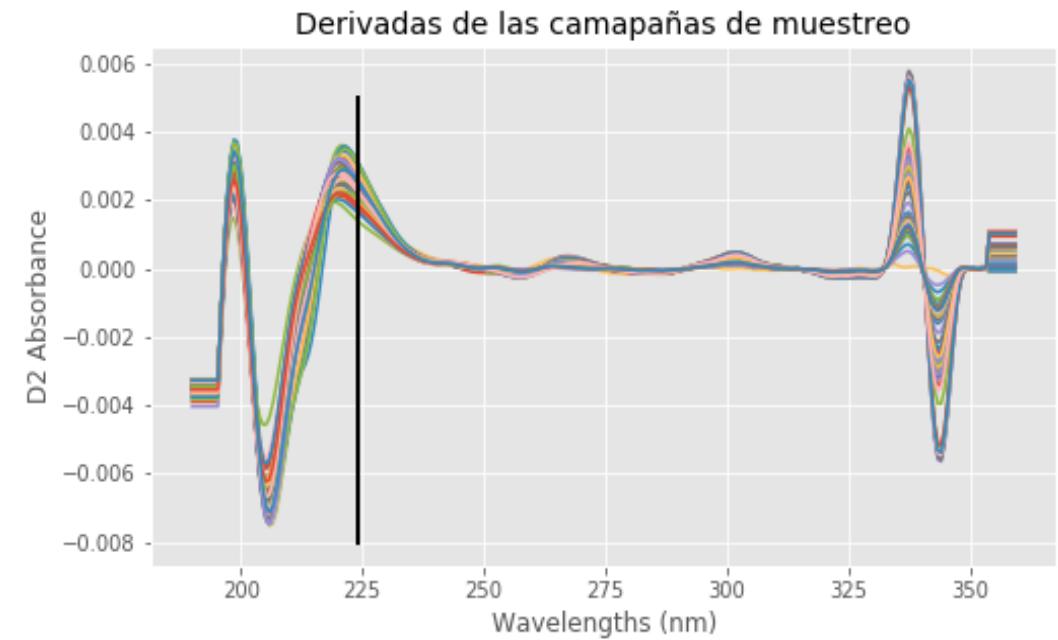
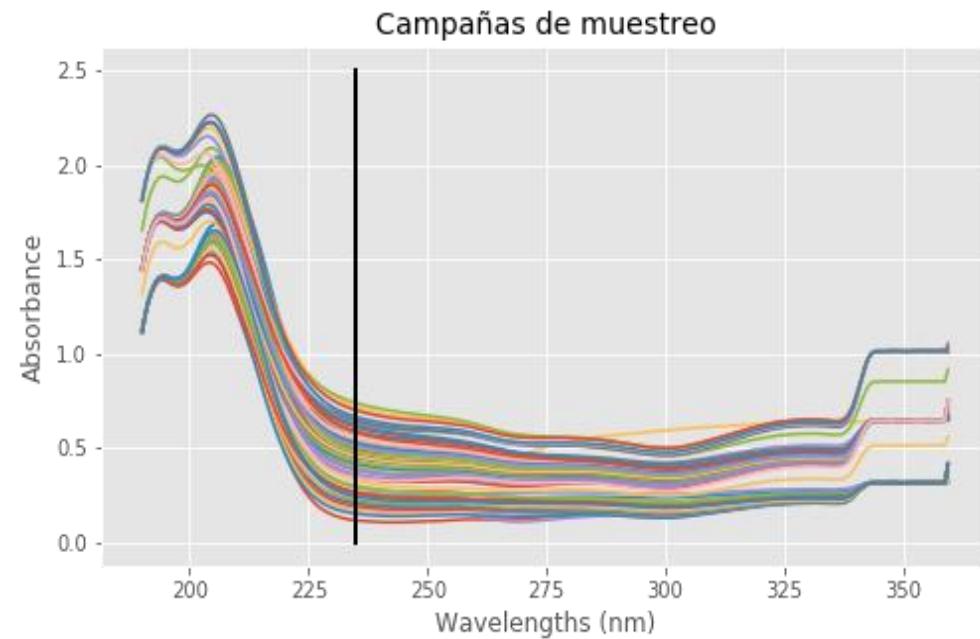


Zhan, Q., Xiang-Zhen, K., Rinke, K. (2019). High-frequency DOC measurements by SAC monitoring and quality controlling in Königshütte Reservoir in Harz Mountains, Germany. 10.13140/RG.2.2.30363.52004.

Soft-sensors: datos espectrométricos

```
[Spectrum]
Version=1
IDData=VAL1_V1_1_2020-06-22_21-37-47_020
IDDevice=D16D
IDDataType=SPECTRUM
IDDataTypeSub1=RAW
DateTime=2020-06-22 21:37:47
[Attributes]
IntegrationTime=256
Unit1=$05 $00 Pixel
Unit2=$03 $05 Intensity counts
Unit3=$f0 $05 Error counts
Unit4=$f1 $00 Status
[END] of [Attributes]
[DATA]
0 1069 0 0
1 1070 0 0
2 1084 0 0
3 1123 0 0
4 1187 0 0
5 1288 0 0
6 1451 0 0
7 1673 0 0
8 1951 0 0
9 2267 0 0
10 2627 0 0
11 3032 0 0
12 3457 0 0
13 3968 0 0
14 4549 0 0
15 5204 0 0
16 5970 0 0
```

Soft-sensors: datos espectrométricos



Soft-sensors: datos espectrométricos

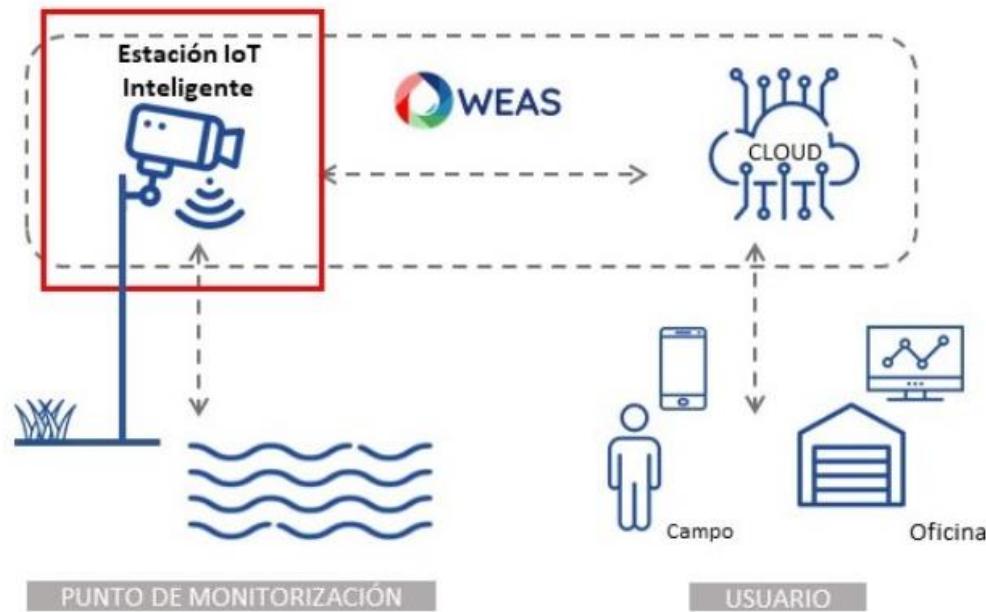
Métodos más utilizados:

- Métodos lineales:
 - Regresión lineal simple: longitud de onda con mayor coeficiente de correlación con la variable de interés.
 - Regresión lineal múltiple: conjunto de longitudes de onda que mayor información aporte al modelo.
 - PLS (Partial Least Squares): método de reducción de la dimensionalidad empleando combinaciones de las longitudes de onda
- Métodos no lineales:
 - SVM, NN, etc.

Soft-sensors: datos espectrométricos

AQUACORP

- Detección óptica (cámara)

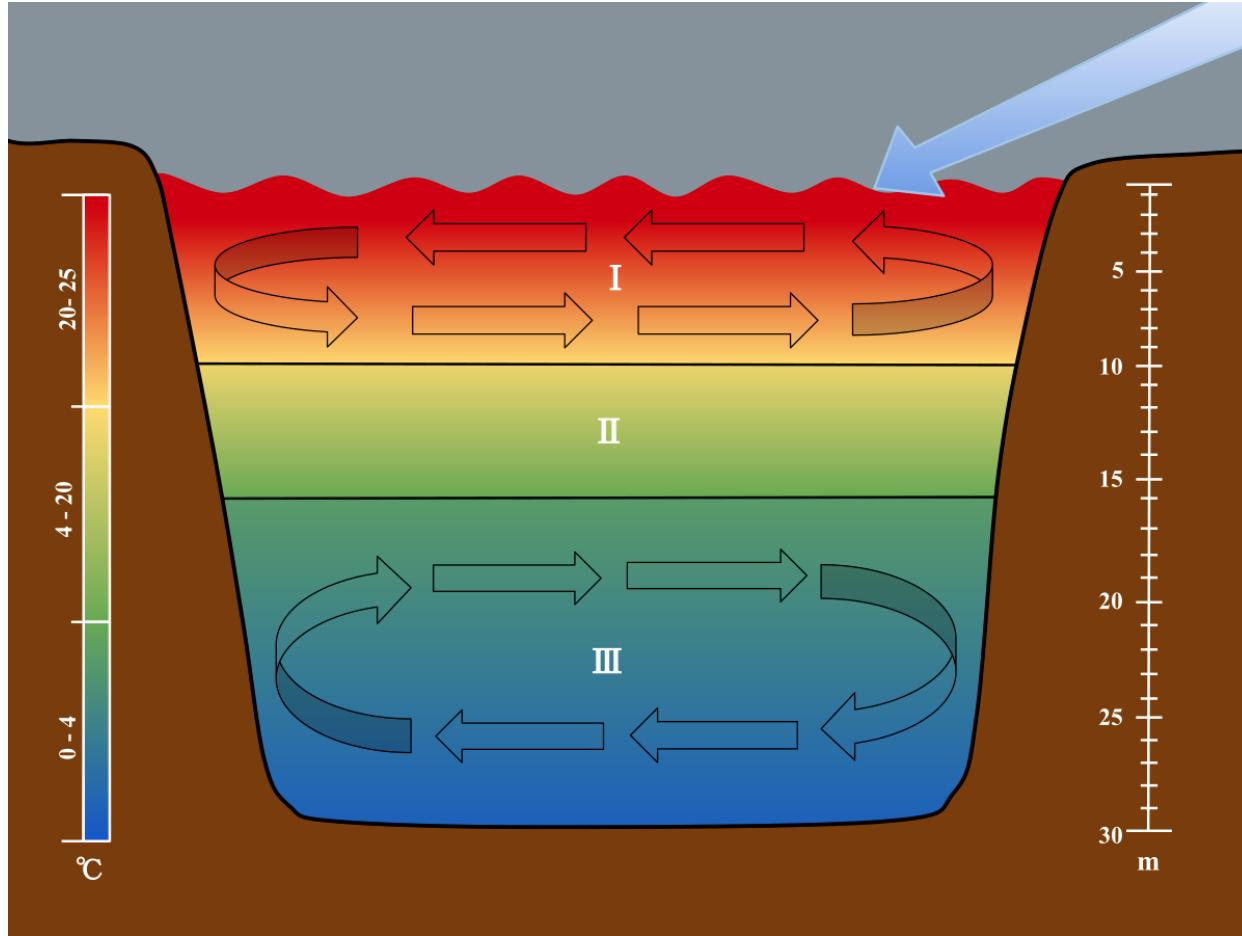


A color-coded legend is located on the right side of the table, showing a gradient from green to red. The text next to it reads: 'Los colores varían en función de la dificultad de medición, que depende de la variación para los valores de los parámetros y de la disponibilidad de las sondas.'

UPC	Color	[0 , 500]		
NTU	mS/cm	mgO/L	mgC/L	
Turbidez	Conductividad	DQO	Carbono Orgánico Total	
[1 , 300]	[0 , 20000]	[0 , 1500]	[0 , 500]	
mgSS/L	mV	mgO/L	mgNO ₃ /L	
Sólidos en Suspensión	REDOX	DBO	Nitratos	
[0 , 4000]	[0 , 1000]	[0 , 1000]	[0 , 100]	
Ud. pH	mgO/L	mgNH ₄ /L	mgNO ₂ /L	mgPO ₄ /L
pH	Oxígeno Disuelto	Amonio	Nitritos	Fosfatos
[0 , 14]	[0 , 20]	[0 , 100]	[0 , 100]	[0 , 10]

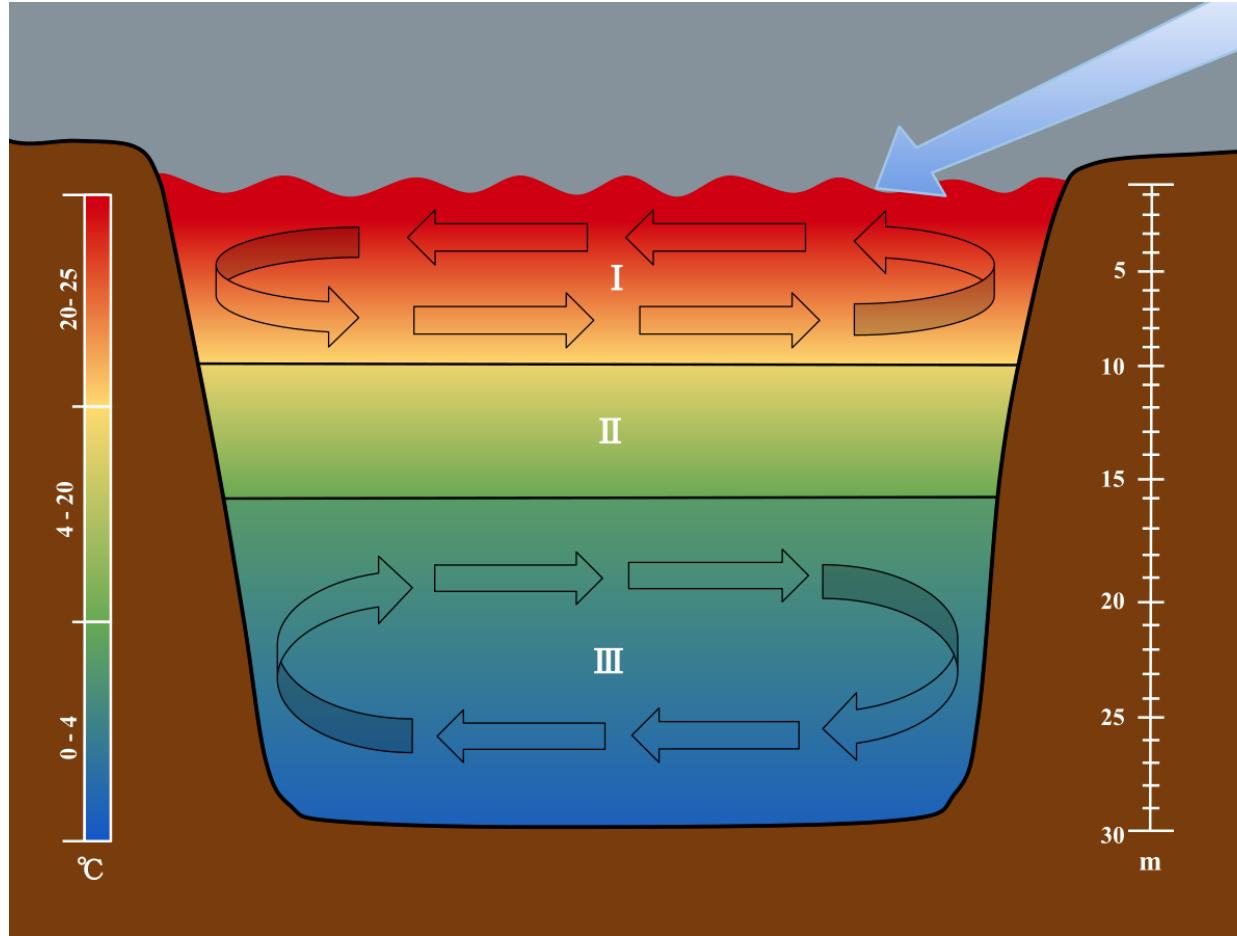
Estratificación térmica en embalses

Estratificación térmica en embalses



https://es.wikipedia.org/wiki/Estratificaci%C3%B3n_de_lagos

Estratificación térmica en embalses

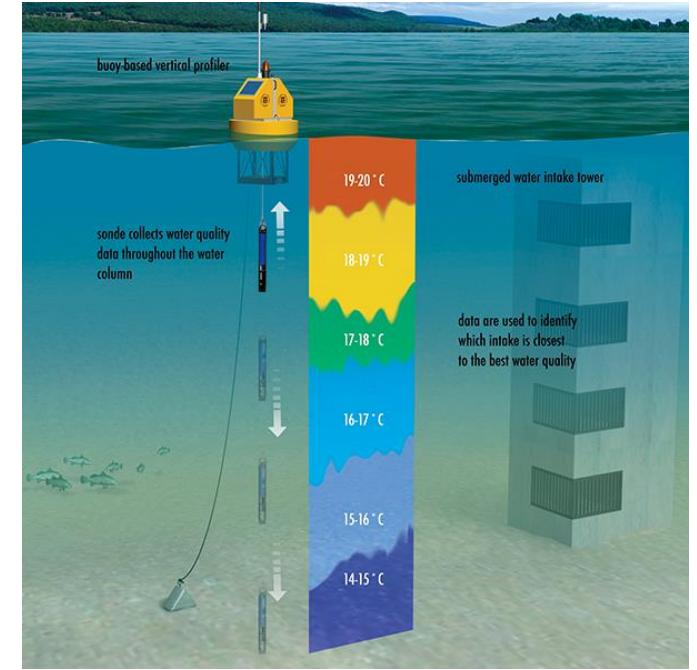
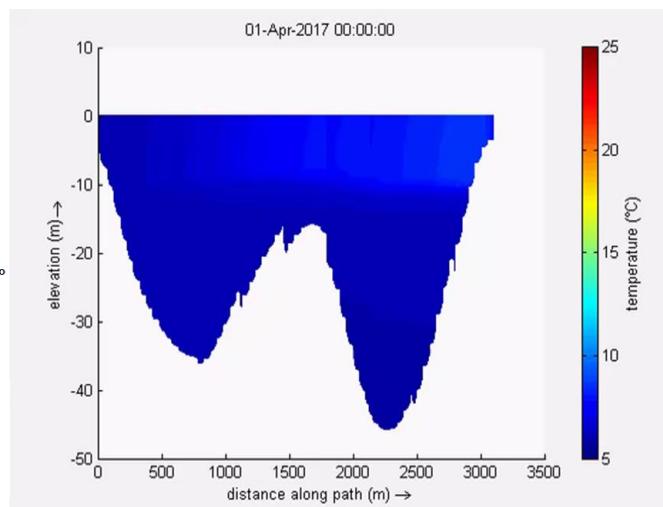
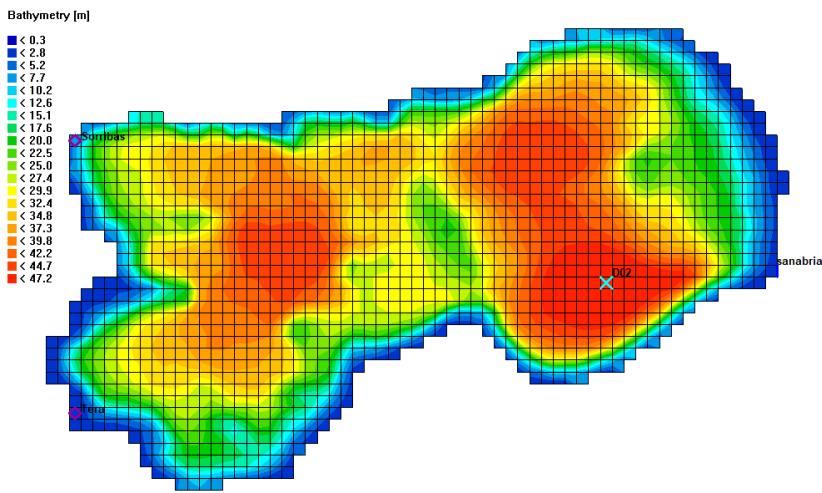


https://es.wikipedia.org/wiki/Estratificaci%C3%B3n_de_lagos

- Según bibliografía se considera que un embalse está estratificado cuando se dan gradientes de temperatura entre 0.2 to 2 °C m⁻¹.
- Cuándo empieza y cuándo acaba.
- A qué profundidad se encuentra la termoclina en cada momento.
- Qué temperatura tiene la termoclina.

Estratificación térmica en embalses

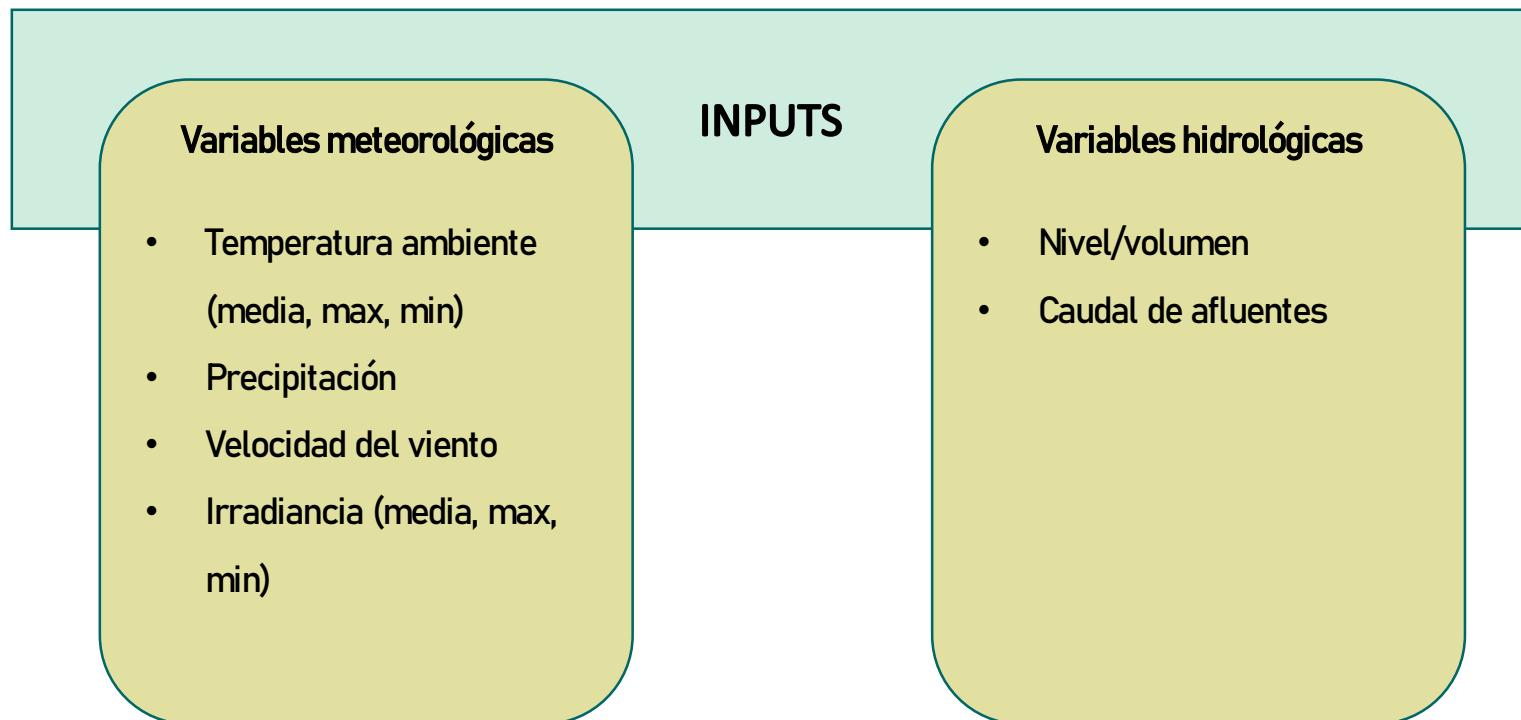
- **Monitorización**
Perfiles verticales diarios (o mayor frecuencia)
- **Modelos deterministas**
Modelos que, a través de una serie de ecuaciones matemáticas, describen la hidrodinámica de una masa de agua.



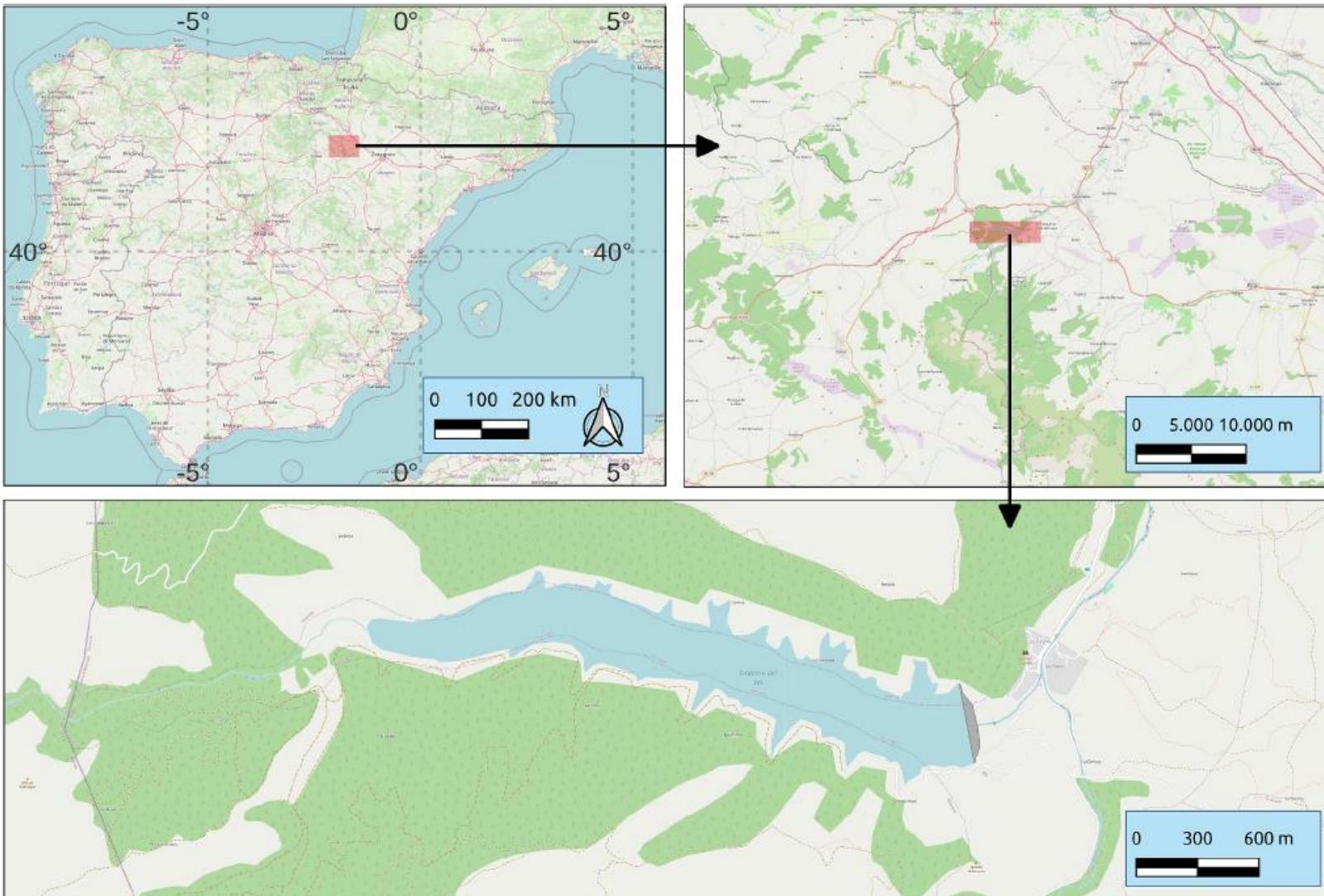
<https://www.ysi.com/6950fixedverticalprofiler>

Estratificación térmica en embalses

- Estimación basada en datos



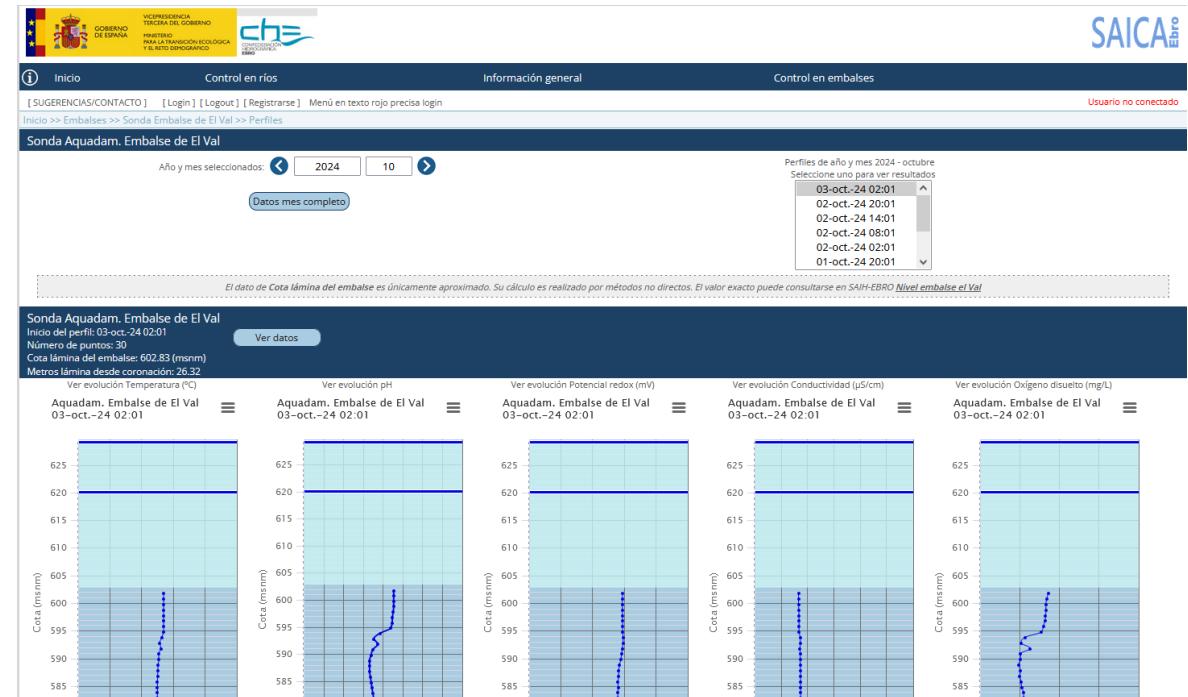
Estratificación térmica en embalses



Estratificación térmica en embalses

- Fuente de datos: SAIH Ebro y SAICA Ebro

The screenshot shows the homepage of the SAIH Ebro website. At the top, there is a navigation bar with links for 'SAIH Ebro', 'Confederación Hidrográfica del Ebro', 'El SAIH Ebro', 'Real time data', 'Forecasts', 'News', 'Datos a la carta', 'Contact', and 'Users'. Below the navigation bar, the main title 'Datos a la carta' is displayed in large blue letters. Underneath the title, there are sections for 'Datos' (with a note about being a sporadic user), 'Informes' (with a link to a form), and 'Contacto' (with a note about contacting them). A large graphic of a water drop containing binary code is centered on the page.



Estratificación térmica en embalses

- Aprendizaje no supervisado: clustering

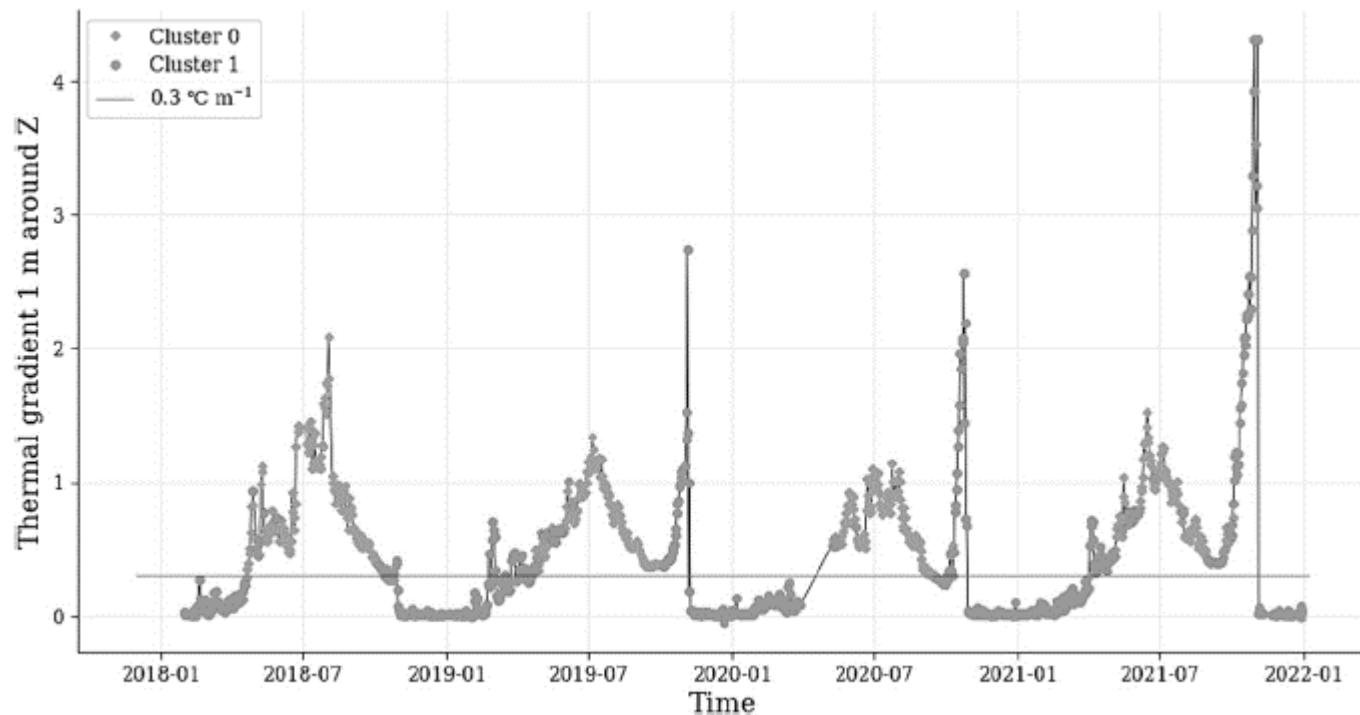


Fig. 4. Temperature gradient along 1 m with centre in Z between 2018 and 2021.

Estratificación térmica en embalses

- Aprendizaje no supervisado: clustering

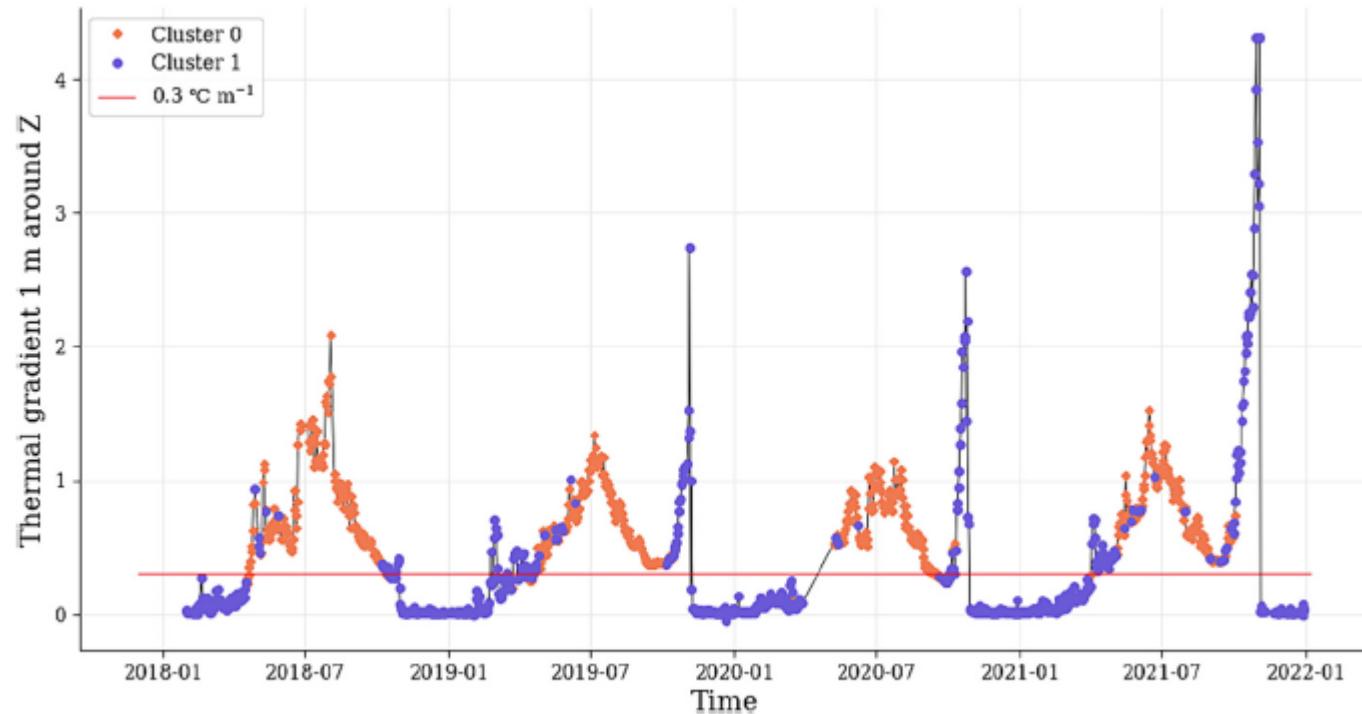
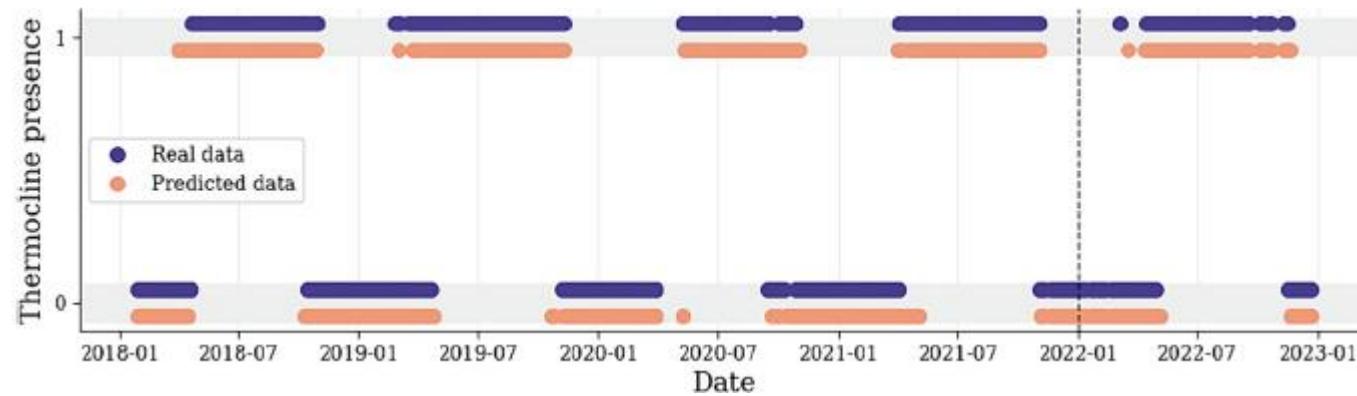


Fig. 4. Temperature gradient along 1 m with centre in Z between 2018 and 2021.

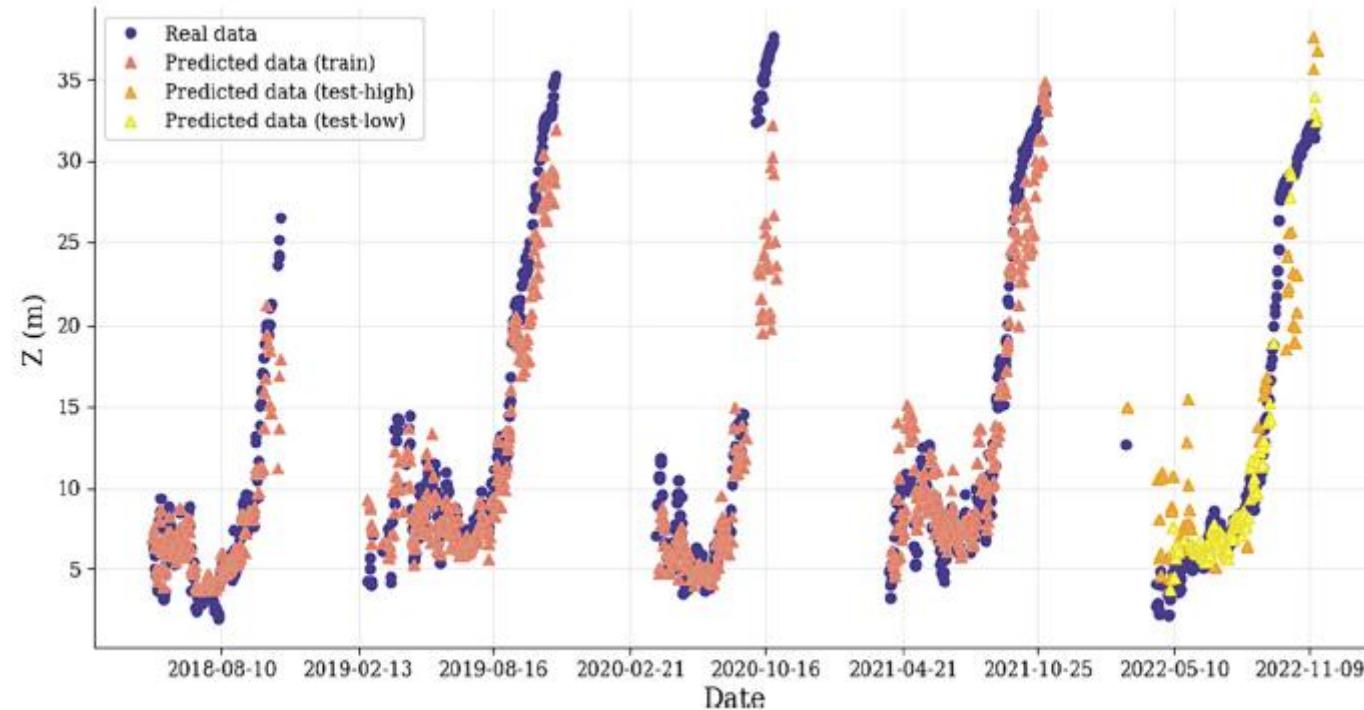
Estratificación térmica en embalses

- Detección de estratificación mediante modelo de clasificación



Estratificación térmica en embalses

- Estimación de la profundidad de la termoclina mediante modelo de regresión.



IA y agua desde otro prisma

Forbes TECNOLOGÍA

El hipercentro de datos de Meta en Talavera consumirá más de seiscientos millones de litros de agua berible en una zona en riesgo de sequía | Tecnología

por Adrián Cano
mayo 9, 2023



<https://forbes.com/es/tecnologia/el-hipercentro-de-datos-de-meta-en-talavera-consumira-mas-de-seiscientos-millones-de-litros-de-agua-berible-en-una-zona-en-riesgo-de-sequia-tecnologia/>

Tabla 11: Cálculo de demanda de agua y saneamiento

	Demanda de agua PSI		Aguas residuales PSI	
	Agua Industrial Data Center	Resto Urbanización	Agua Industrial Data Center	Resto Urbanización
Litros por Año	310.000.000	0	185.000.000	0
Caudal medio (litros/segundo)	10	0	6	0
Pico de demanda (litros/segundo)	190	0	115	0
	Agua Potable Data Center	Resto Urbanización	Agua Potable Data Center	Resto Urbanización
	17.000.000	440.000.000	10.000.000	396.000.000
Caudal medio (litros/segundo)	1	14	1	13
Pico de demanda (litros/segundo)	4	35	4	31
Estimación Total				
Litros por Año	327.000.000	440.000.00	195.000.000	396.000.000
Caudal medio (litros/segundo)	11	14	7	13
Pico de demanda (litros/segundo)	194	35	119	31

Global Villacreses S.L. Meta Data Center Campus Memoria Informativa Justificativa Marzo 2022

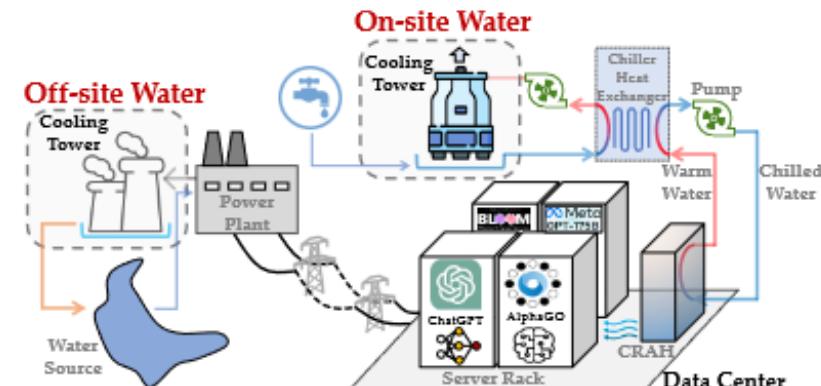
IA y agua desde otro prisma

- Los grandes Centros de Procesamiento de Datos o *Data Centers* consumen agua para su refrigeración.
- Water positive by 2030

AWS' four key strategies to become water positive are:

1. Improving water use efficiency across its operations
2. Using sustainable water sources (e.g., recycled water)
3. Returning water for community reuse
4. Supporting water replenishment projects for communities and the environment

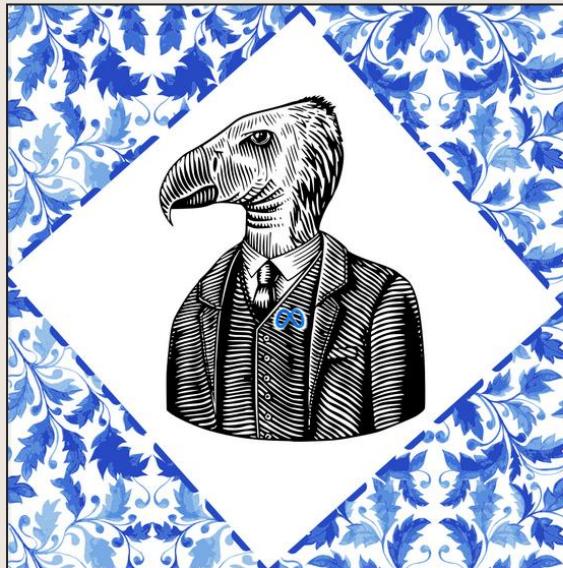
- *Inference: ChatGPT needs to “drink” a 500ml bottle of water for a simple conversation of roughly 20-50 questions and answers, depending on when and where ChatGPT is deployed.*



IA y agua desde otro prisma

TU NUBE SECA MI RÍO

INICIO SOMOS RECURSOS COLABORAR CONTACTO



25 de abril de 2023



19 de abril de 2023

El mega centro de datos de Meta en Talavera de la Reina afectará a aves