

# Detección de anomalías para el control de la calidad de aguas termales

Abel Pampín Rodríguez

Elena Hernández Pereira

Óscar Fontenla Romero

## Resumen

El proyecto AQUAPRED aplica técnicas avanzadas al sector termal, para mejorar la calidad de sus tratamientos terapéuticos, a través de sistemas de control y predicción de la calidad de sus aguas y del modelado de su hidrobioma.

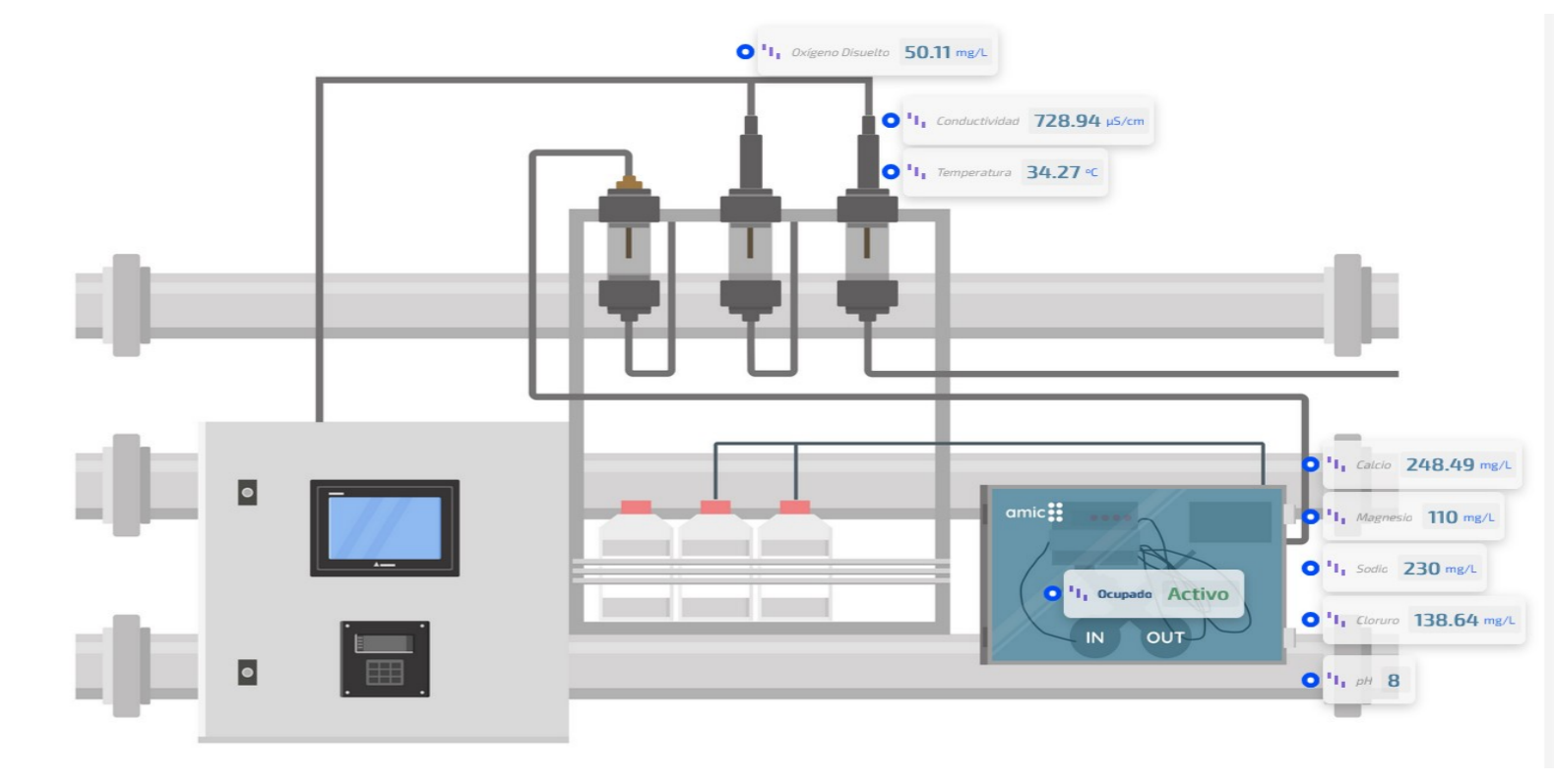
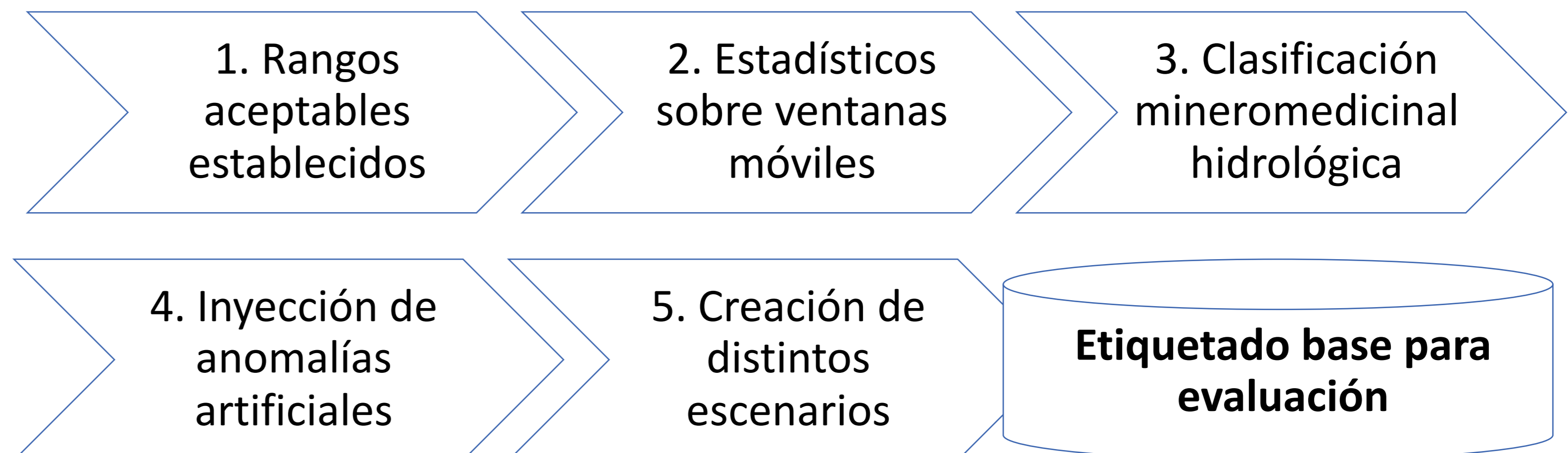
Uno de estos sistemas es un módulo de detección de anomalías para monitorizar la calidad de las aguas termales. Este módulo aplica un modelo de ML de detección de anomalías (AD) basado en distancias (LOF) entrenado con observaciones normales del conjunto de datos. Mediante extracción de características y diferentes estrategias de etiquetado e inyección de anomalías artificiales obtenemos un modelo robusto y eficiente sobre múltiples escenarios de evaluación.

## Conjunto de datos y etiquetado

- Análisis de laboratorio realizados por el balneario Hervideros de Cofrentes.
- Muestras diarias de sus piscinas durante gran parte del año 2024. (275 muestras)
- Parámetros físico-químicos relevantes (variables):

Temperatura	Cloro libre	Turbidez	pH	CO <sub>2</sub>
Tª ambiente	Cloro combinado	Humedad relativa	CO <sub>2</sub> <sup>in</sup>	CO <sub>2</sub> <sup>ex</sup>

### ETIQUETADO:



## Escenarios

Mediante estrategias de etiquetado con:

- Diferentes rangos y umbrales.
- Distintos tamaños de ventana y técnicas estadísticas.
- Combinando distintos tipos de anomalías artificiales (*puntuales vs. contextuales*).
- Afectando a una o más variables (*univariado vs. multivariado*).
- Mayor o menor grado de contaminación.

... creamos los siguientes escenarios:

E1: Puntual multivariado

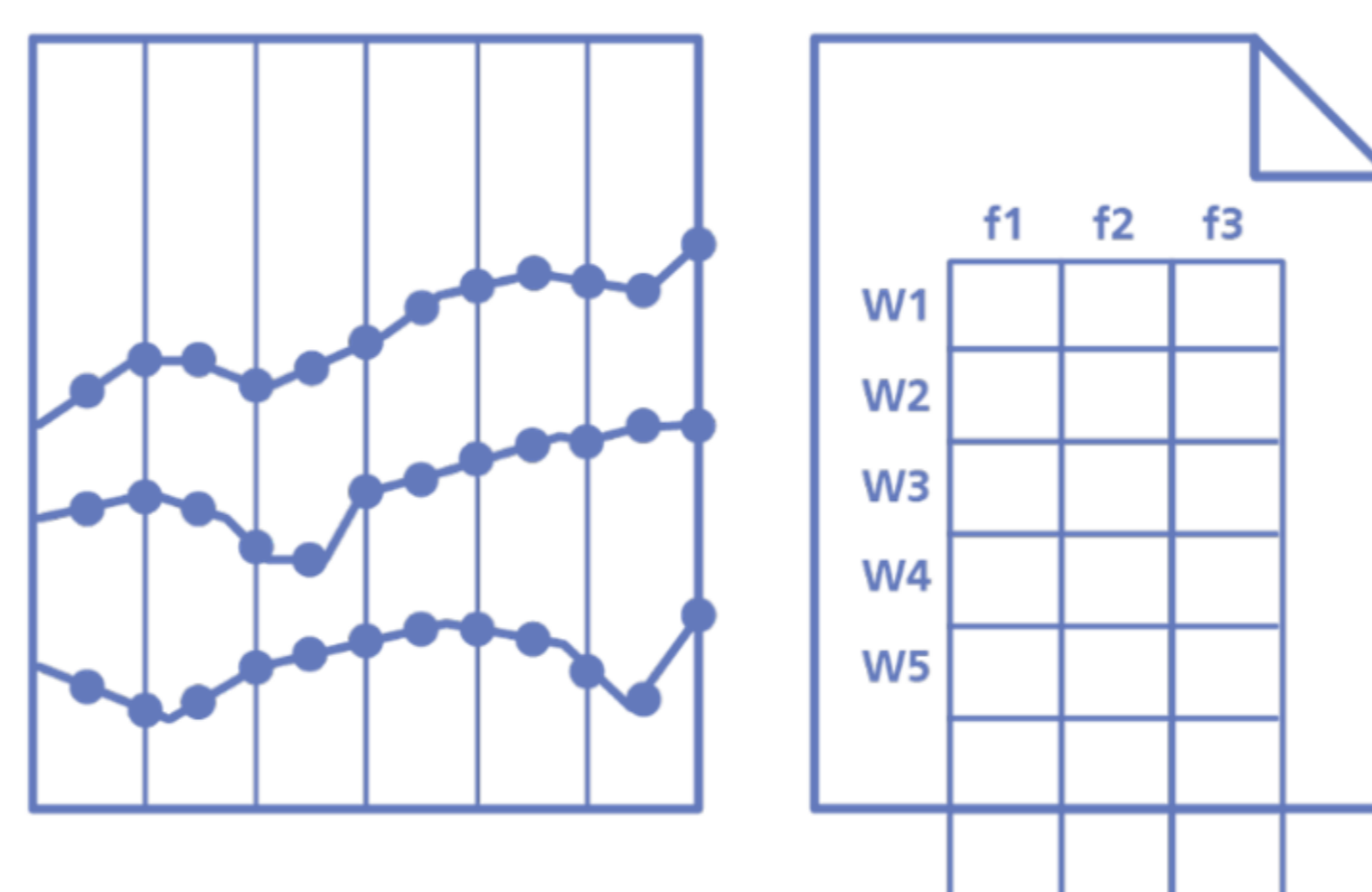
E2: Contextual multivariado

E3: Mixto multivariado

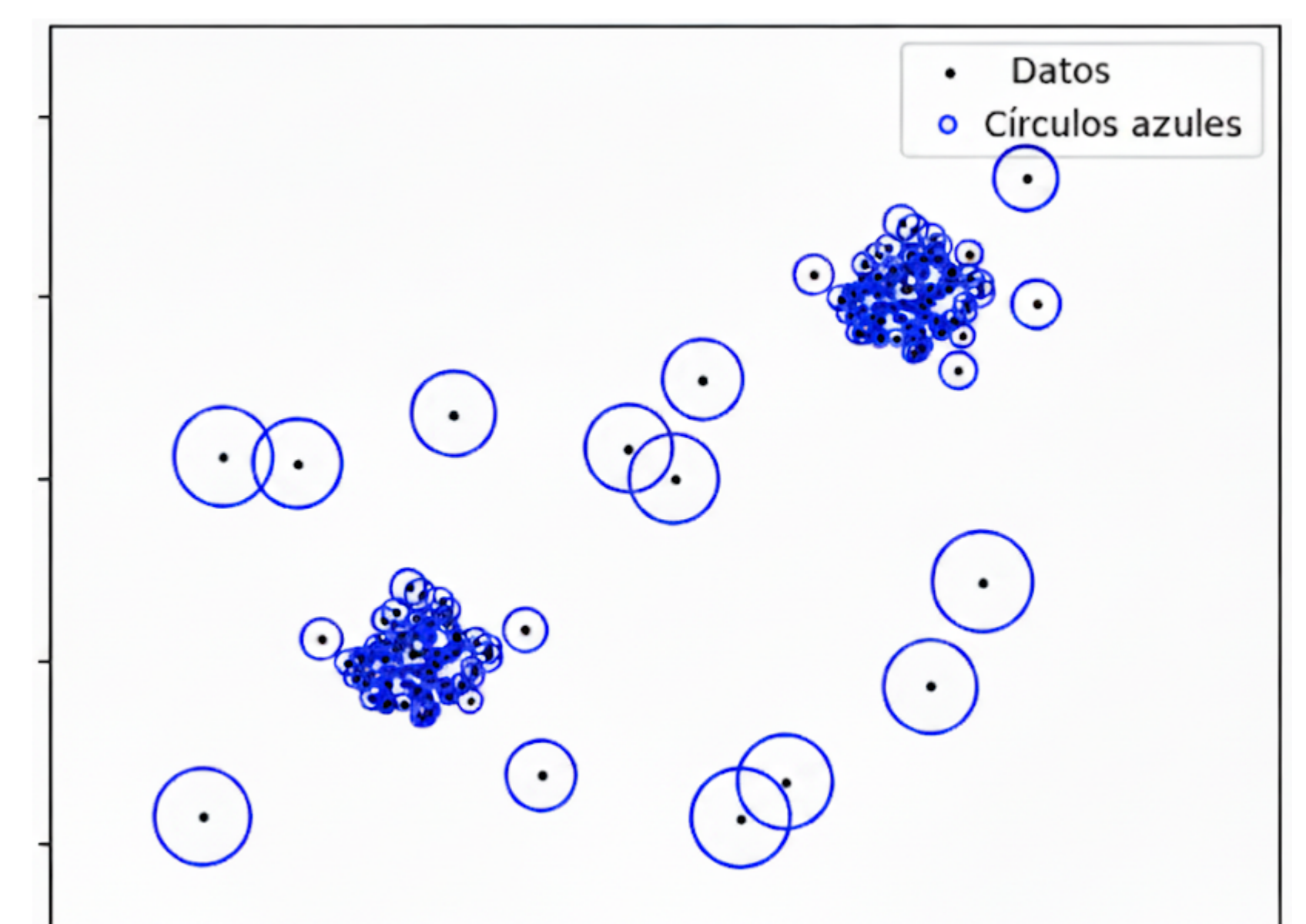
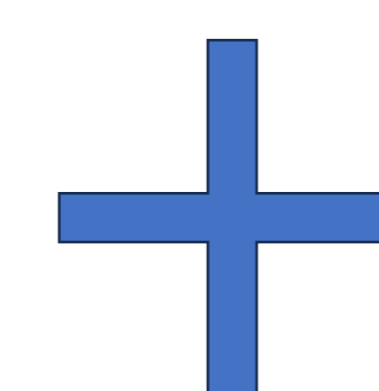
E4: Mixto univariado

## Modelo y sistema propuesto

W1 W2 W3 W4 W5 (W6) → (W6)



Extracción de características



Local Outlier Factor (LOF)

## Resultados y conclusiones

Escenario	Recall	MCC
	Media (DE)	Media (DE)
Puntual	0.70 (0.04)	<b>0.61 (0.03)</b>
Contextual	0.69 (0.05)	<b>0.64 (0.03)</b>
MixtoMulti	0.65 (0.06)	<b>0.57 (0.02)</b>
MixtoUni	0.56 (0.06)	<b>0.55 (0.04)</b>

- Resultados estables y similares en todos los escenarios.
- Mayor sensibilidad en escenarios puntuales → apropiado para un modelo estático como LOF.
- Aplicación en un caso → conjunto de datos sin etiquetar.
- Falta de dinámicas marcadas.