



Predicción a tiempo real de mezclas de aguas de diferentes orígenes en redes de distribución de agua de consumo

Las mezclas de agua de diferentes orígenes en las redes de distribución de agua potable son una realidad en muchas áreas urbanas, especialmente aquellas que son densamente pobladas. Los cambios en el origen del agua pueden provocar alteraciones en la percepción organoléptica de los consumidores, así como afectaciones en la estructura del sistema de distribución. El presente estudio se centra en la zona del área metropolitana de Barcelona, que tiene tres orígenes principales de agua potable, el río Ter, el río Llobregat y agua desalada del mar Mediterráneo. Mediante el análisis de parámetros fisicoquímicos de estas aguas se desarrollaron y validaron modelos de predicción de origen del agua, tales como modelado con redes neuronales, clasificación y regresión. Los modelos mostraron una buena capacidad de predicción del origen del agua y del porcentaje de cada origen en caso de mezclas binarias.

Palabras clave

Agua de consumo, mezclas, modelización, redes neuronales, red de distribución.

REAL-TIME PREDICTION OF MIXED WATER FROM DIFFERENT SOURCES IN DRINKING WATER NETWORKS

Water mixtures of different origins in drinking water distribution networks are a reality in many urban areas, especially those that are very populated. Changes in water origin can cause alterations in the organoleptic perception of the consumers, as well as changes in the structure of the distribution system. The present study focuses on the Barcelona metropolitan area, which has three main drinking water sources, the Ter river, the Llobregat river and desalinated water from the Mediterranean Sea. By analyzing physicochemical parameters of these waters, water origin prediction models, such as neural network modeling, classification and regression, were developed and validated. The models showed a good prediction capacity of the water origin and the percentage of each origin in case of binary mixtures.

Keywords

Drinking water, mixtures, modeling, neural networks, water supply network.

Albert Serra Compte

project manager de Cetaqua, Centro Tecnológico del Agua

Bernat Joseph Duran

project manager de Cetaqua, Centro Tecnológico del Agua

Victor Saenger

project manager de Cetaqua, Centro Tecnológico del Agua

Susana González Blanco

responsable técnica de Cetaqua, Centro Tecnológico del Agua

Fernando Valero

responsable de I+D+i y Control de Procesos del Ens d'Abastament d'Aigües Ter Llobregat

Pere Emiliano

técnico de I+D+i Control de Procesos del Ens d'Abastament d'Aigües Ter Llobregat

Verónica García

responsable de laboratorio ETAP SJD de Aigües de Barcelona, Empresa Metropolitana de Gestió del Cicle Integral de l'Aigua

Isabel Pérez

responsable de laboratorio de Química Inorgánica de Aigües de Barcelona, Empresa Metropolitana de Gestió del Cicle Integral de l'Aigua

Miquel Paraira Faus

director de Calidad del Agua de Aigües de Barcelona, Empresa Metropolitana de Gestió del Cicle Integral de l'Aigua



1. INTRODUCCIÓN

En las últimas décadas, la disponibilidad de agua superficial como fuente de agua potable se ha visto reducida significativamente en un gran número de áreas urbanas, principalmente debido a impactos antropogénicos y presiones ambientales. Una de las medidas implementadas para garantizar el suministro de agua potable en aquellas áreas donde hay diferentes fuentes de agua, es aumentar las interconexiones en la red de distribución para que cada una de las fuentes tenga el potencial de suministrar agua en la mayor parte de la población servida. Este es el caso del área metropolitana de Barcelona, cuyas fuentes principales de agua son el río Llobregat y el río Ter. Además, en los últimos años, ha aumentado la proporción de agua desalada proveniente del mar Mediterráneo. Estas tres fuentes que proporcionan agua de características muy dispares entre ellas y se mezclan en la red de distribución de Barcelona y su área metropolitana, pueden provocar cambios en la percepción organoléptica de los consumidores.

En el caso concreto de Barcelona, la percepción organoléptica ha sido históricamente diferente para el agua proveniente del río Llobregat que para el agua proveniente del río Ter, siendo determinante en ello el impacto de las explotaciones mineras de potasa, ubicadas en la cuenca del río Llobregat. Las diferentes características del agua pueden afectar de forma diferente a la estructura de la red de distribución, así como provocar precipitaciones en la cara interna de las tuberías cuando se producen mezclas, hecho que puede dañar la estructura de la red. Además, estos cambios en el origen del agua pueden alterar el funcionamiento de los equipos de

» Las mezclas de agua en la red de distribución pueden provocar cambios en la percepción organoléptica de los consumidores. Por eso, este estudio desarrolla un modelo de predicción del origen del agua para optimizar su gestión

medición en continuo de la calidad del agua, así como incentivar reclamaciones por parte de los consumidores, por los cambios en la percepción organoléptica anteriormente mencionados.

Por estos motivos, conocer el origen del agua en los diferentes puntos de la red de distribución se convierte en un objetivo cada vez más relevante. Determinar el origen del agua y el porcentaje de mezclas en cada zona de la red permitiría optimizar la gestión en la misma y dar respuesta a cambios en las características del agua servida, debidas a cambios en el origen del agua.

Así, el objetivo principal de este estudio fue el de desarrollar un modelo de predicción del origen del agua y el porcentaje de mezclas utilizando parámetros medibles a tiempo real y, que, por lo tanto, se pudiera implementar en puntos de interés de la red de distribución para conocer el origen y el porcentaje de mezclas en cada momento.

2. MATERIAL Y MÉTODOS

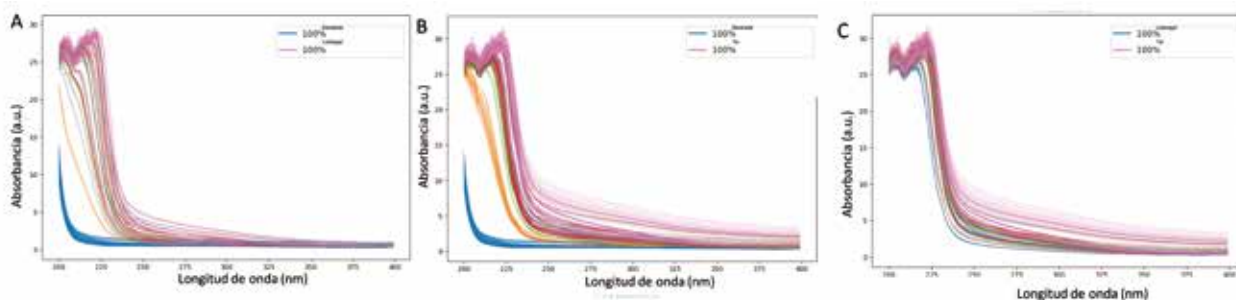
2.1. RECOGIDA Y ANÁLISIS DE MUESTRAS

Las muestras se recogieron a la salida de tres plantas potabilizadoras, cuyos puntos de captación corresponden con las tres principales fuentes de agua potable del área de Barcelona. Estas fueron la estación de tratamiento de agua potable (ETAP) de Cardedeu, que recoge agua del río Ter; la ETAP de Abrera,

que recoge agua del río Llobregat; y la instalación desaladora de agua marina (IDAM) del Prat, que recoge agua del mar Mediterráneo. El muestreo se realizó durante el año 2020, desde enero a marzo y desde julio a noviembre, cubriendo tres estacionalidades diferentes. Por motivos de la COVID-19, no se pudieron recoger muestras entre los meses de marzo a julio, por lo que no se obtuvieron muestras representativas de la primavera. En total, se realizaron 24 campañas de muestreo. En cada una de ellas se recogieron 5 L de agua de la salida (agua tratada) de cada una de las plantas. Las muestras se trasladaron al laboratorio para su análisis el mismo día de muestreo.

Una vez en el laboratorio, a partir de las muestras recogidas en los tres puntos, se prepararon mezclas binarias de agua en diferentes proporciones. Tanto las muestras de agua puras (sin mezclar) como las mezclas binarias se caracterizaron con diferentes parámetros fisicoquímicos. Según experiencias previas para diferenciar el origen del agua, los parámetros más relevantes que se utilizaron para caracterizar el agua fueron el espectro UV de 200 a 400 nm y la conductividad. Además, de forma complementaria, se caracterizó pH, potencial redox y UV 254. Todos estos parámetros se midieron con sondas con capacidad de reportar datos a tiempo real. En cada uno de los análisis se realizaron un mínimo de 3 réplicas analíticas.

FIGURA 1. Espectro UV de las muestras analizadas: A, desaladora El Prat con Llobregat Abrera; B, desaladora El Prat con Ter-Cardedeu; y C, Llobregat-Abrera con Ter-Cardedeu. Las líneas azules y lilas corresponden a agua pura de cada origen; mientras que las líneas naranjas, rojas y verdes corresponden a las mezclas binarias en cada uno de los casos correspondientes.



2.2. TRATAMIENTO DE DATOS Y MODELIZACIÓN

Para la predicción del origen del agua y el porcentaje de mezcla en caso de mezclas binarias, se probaron diferentes estrategias de modelado, incluyendo, redes neuronales y clasificación más regresión.

Para la modelización basada en redes neuronales, se utilizó un modelo de red neuronal simple *feed-forward* con función de activación ReLU, donde hiperparámetros tales como el número de capas en la red, el número de neuronas por capa, así como el *learning rate* inicial, entre otros, fueron seleccionados utilizando un algoritmo GridSearch.

La segunda estrategia de modelado combinó clasificación más regresión. El modelo utiliza en primer lugar un análisis clasificatorio para discernir entre muestras puras y mezclas. Si se trata de muestras puras, un segundo clasificador permite determinar el origen de la muestra de estudio. En caso de tratarse de una mezcla, se utiliza un análisis de regresión basado en *random forest* para determinar el porcentaje de mezcla.

Los dos modelos se desarrollaron y entrenaron con los datos obtenidos en 19 de los 24 muestreos realizados. Se utilizaron los datos de todos los parámetros fisicoquímicos medidos (conductividad, pH, UV254, potencial redox y espectro UV). Ambos

modelos se realizaron mediante el programa Python.

Una vez los modelos estuvieron desarrollados, estos se validaron para determinar su desempeño a la hora de predecir el origen del agua y el porcentaje en caso de mezclas. La validación de los modelos se realizó con datos de muestras analizadas que no se habían utilizado para el desarrollo de los modelos. Se pasaron las mediciones de los parámetros de las muestras 'a ciegas' y se comparó el valor predicho por los modelos con el valor real (de origen de agua o porcentaje de mezcla). Según la comparación del valor predicho con el valor real, se calcularon diferentes errores para determinar de forma cuantitativa el rendimiento de los modelos, específicamente se determinó el error relativo, el error máximo y el error medio absoluto.

3. RESULTADOS

3.1. CARACTERIZACIÓN DE LOS DIFERENTES ORÍGENES DEL AGUA

Los resultados del análisis del espectro UV para cada una de las fuentes de agua y las mezclas en proporciones conocidas se representan en la **Figura 1**.

El agua de origen de desaladora presentó un espectro claramente diferenciado de los otros dos orígenes (Ter y Llobregat) al no tener materia

orgánica, corresponde a las líneas azules de las **Figuras 1A** y **1B**. En ambos casos, mezcla desaladora-Llobregat y mezcla desaladora-Ter, se observó una clara evolución en el espectro desde la muestra de desaladora pura, a medida que se fue incrementando el porcentaje de los otros orígenes (Ter y Llobregat).

La comparación entre los espectros provenientes del agua del río Ter y el río Llobregat se puede observar en la **Figura 1C**. Los espectros de ambos orígenes mostraron una forma parecida, pero no idéntica. El espectro de Llobregat (líneas azules) presentó valores de absorbancia más bajos en la mayoría de las longitudes de onda comparadas con las muestras puras de Ter (líneas lilas), **Figura 1C**. Las mezclas de ambos orígenes se situaron en medio de los valores de las muestras puras.

Otro de los parámetros más relevantes para determinar el origen del agua fue la conductividad. Como ejemplo, se muestran los valores de conductividad de un día específico de análisis (**Figura 2**). Las muestras del Ter tuvieron una conductividad alrededor de 520 $\mu\text{S}/\text{cm}$ (valores del Ter oscilaron entre 380 y 520 $\mu\text{S}/\text{cm}$ en la mayoría de los días de análisis), mientras que los valores del agua proveniente del Llobregat estuvieron alrededor de 630 $\mu\text{S}/\text{cm}$ (las muestras de Llobregat oscilaron entre



FIGURA 2. Evolución de la conductividad durante un día de análisis de las diferentes muestras de agua, siguiendo la secuencia: Ter (puntos rojos), Llobregat (puntos verdes), desaladora (puntos azules). Las mezclas binarias corresponden a los puntos de colores intermedios (proporciones especificadas en la leyenda).

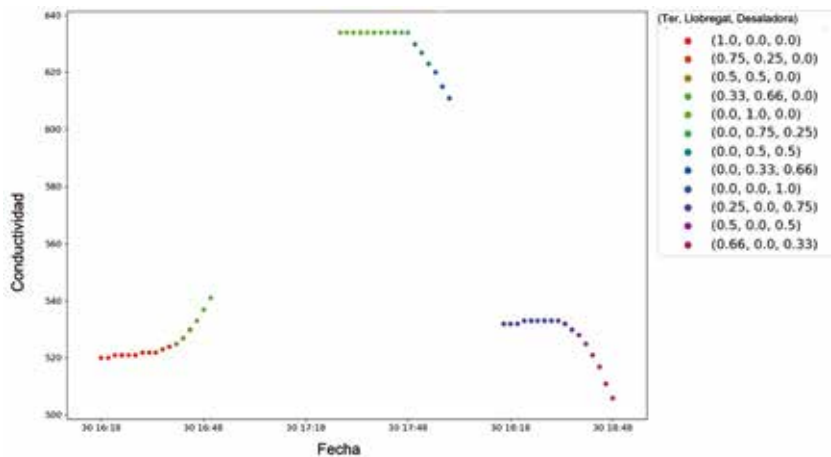


FIGURA 3. Comparativa predicción (líneas naranjas) con proporción real (líneas azules) en 12 muestras diferentes (cada cuadrado) utilizadas para la validación. Modelo utilizado basado en redes neuronales.

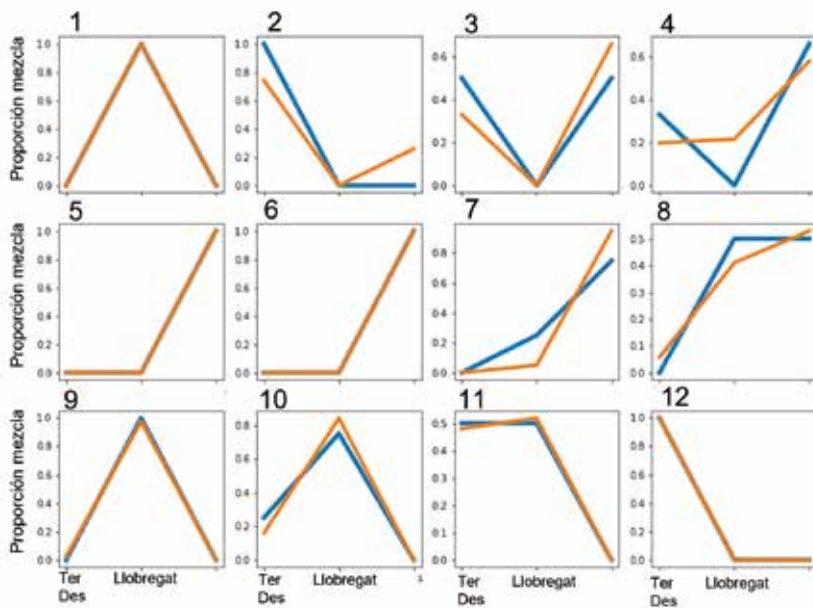


TABLA 1

ERROR ABSOLUTO, ERROR MÁXIMO MEDIO Y ERROR RELATIVO MEDIO DE CADA UNO DE LOS MODELOS DESARROLLADOS. MODELO 1: REDES NEURONALES. MODELO 2: CLASIFICACIÓN MÁS REGRESIÓN. COLOR VERDE OSCURO INDICA UN VALOR DE ERROR INFERIOR AL 10%; COLOR VERDE CLARO, VALOR DE ERROR POR DEBAJO DEL 20%; Y COLOR NARANJA, VALOR DE ERROR SUPERIOR AL 20%.

| Nombre del modelo | Error absoluto medio (%) | Error máximo medio (%) | Error relativo medio (%) |
|-------------------|--------------------------|------------------------|--------------------------|
| Modelo 1 | 9 | 12 | 19 |
| Modelo 2 | 9 | 13 | 22 |

630 y 900 $\mu\text{S}/\text{cm}$ en la mayoría de los días de análisis). Finalmente, los valores de conductividad del agua de desaladora fueron de 530 $\mu\text{S}/\text{cm}$ (la conductividad del agua de desaladora fue muy constante durante los días de análisis). A diferencia de los valores del espectro UV, los valores más diferentes de conductividad fueron los vinculados a las muestras del Ter y Llobregat.

3.2. MODELOS DE PREDICCIÓN DEL ORIGEN DEL AGUA

Los resultados de validación (error absoluto, error máximo medio y error relativo medio) de los dos modelos se muestran en la **Tabla 1**.

Los modelos desarrollados reportaron valores de error muy parecidos entre ellos. En la mayoría de los casos, los valores estuvieron por debajo del 20% de error. Estos resultados nos indican que los modelos predicen de forma fiable el origen del agua y porcentaje de mezcla en mezclas binarias. Además, los dos modelos reportaron un comportamiento parecido, lo que indica que hay una clara relación en los datos que permite su modelado para determinar el origen del agua.

Como ambos modelos reportaron resultados parecidos, en las gráficas de la **Figura 3** se muestran solo los resultados de uno de ellos, el modelo basado en redes neuronales, ya que presentó unos valores de error ligeramente inferiores. Así, la **Figura 3** muestra la comparativa entre valores predichos por el modelo basado en redes neuronales con los valores reales, en 12 muestras utilizadas para la validación. Los resultados muestran que, para las muestras sin mezcla, el modelo se ajusta con exactitud a los valores reales (las líneas naranjas y azules se encuentran perfectamente solapadas, muestras 1, 5, 6, 9 y 12).

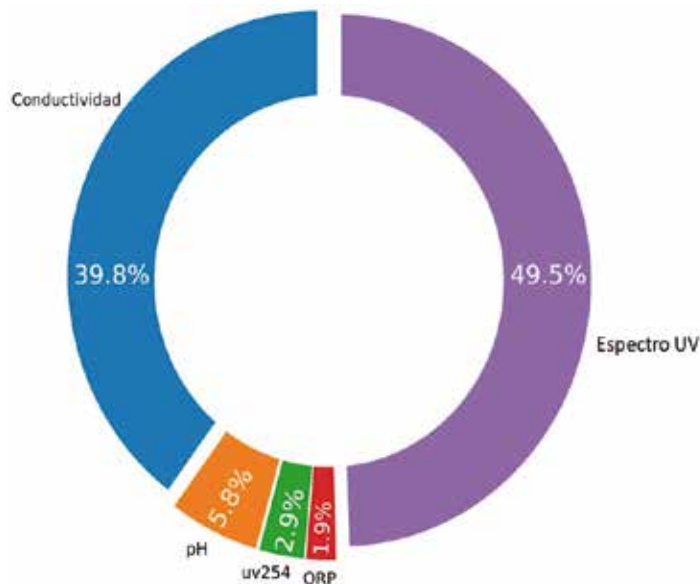
Solo la muestra 2, cuyo origen del agua corresponde al río Ter, presentó cierta variación entre los valores predichos y reales.

El modelo predijo con un error bajo las mezclas entre Ter y Llobregat (muestras 10 y 11), indicando que a pesar de que estas aguas mostraron unos espectros UV similares, el modelo fue capaz de diferenciar ambos orígenes y calcular el porcentaje de mezclas. El error en la predicción aumentó ligeramente en aquellas mezclas donde estaba presente agua de desaladora, tanto mezclada con agua del Ter (muestras 3 y 4), como mezclada con agua del Llobregat (muestras 7 y 8). Esto puede deberse a que, aunque el espectro de agua desalada pura era muy diferente a los otros dos orígenes, al no tener materia orgánica, cuando se produce mezcla, solo hay aporte de materia orgánica del otro origen (Ter y Llobregat). El efecto del agua desalada sobre el espectro UV de la mezcla con otro origen de agua sería de dilución y no de cambio de forma del espectro, lo que puede dificultar al modelo determinar el porcentaje de desaladora. Sin embargo, los errores de predicción fueron aceptables en todos los casos.

Finalmente, se investigó la contribución de cada uno de los parámetros fisicoquímicos utilizados en la predicción del origen del agua y porcentaje de mezclas, en el modelo de redes neuronales (**Figura 4**).

Los dos parámetros con mayor contribución al rendimiento del modelo fueron el espectro UV (49,5%) y la conductividad (39,8%). El espectro UV ya se destacó en iniciativas previas, tal y como se ha mencionado anteriormente, como el parámetro más útil para discernir entre diferentes orígenes de agua; junto con la conductividad, contri-

FIGURA 4. Porcentaje de importancia de cada parámetro en la predicción de origen del agua y porcentaje de mezcla. El porcentaje se calculó mediante un árbol de decisión del paquete *scikit-learn*, el cual cuenta con el parámetro *feature_importances* que permite obtener el porcentaje de varianza que cada variable explica al momento de construir al árbol.



buyeron en un 90% en la varianza explicada en el modelo. El resto de los parámetros, pH, UV254 y ORP, tuvieron contribuciones minoritarias, por debajo del 6% (**Figura 4**). Este análisis sirve para ajustar el número de sensores necesarios para la predicción del origen del agua, lo que permite la optimización del coste de la solución final.

4. CONCLUSIONES

Se concluye que:

- La caracterización de las muestras de agua mostró diferencias significativas entre el espectro UV del agua desalada con el espectro UV del agua del río Ter y el río Llobregat. La conductividad, sin embargo, se diferenció más en el caso del río Llobregat en comparación con agua de desaladora y agua del Ter.
- Se desarrollaron modelos de predicción de origen del agua y porcentaje de mezclas y se validaron con muestras ciegas. Los dos mo-

delos tuvieron errores bajos en la predicción (por debajo del 20%), lo que demuestra su capacidad para determinar el origen del agua.

- Los modelos tuvieron un mejor rendimiento determinando el origen del agua en muestras puras, en el caso de mezclas, especialmente cuando había agua de desaladora los errores de predicción aumentaron ligeramente.
- Los parámetros que más ayudaron a la predicción del origen del agua fueron el espectro UV y la conductividad.
- Se validaron modelos de predicción del origen del agua utilizando solo parámetros que pueden medirse a tiempo real y, por lo tanto, determinar en cada momento, en puntos específicos de la red, el origen del agua y porcentaje de mezclas.
- La aplicación de los modelos en la red de distribución se puede utilizar para mejorar la gestión de la red




» Se validaron modelos de predicción del origen del agua utilizando solo parámetros que pueden medirse a tiempo real y, por lo tanto, determinar en cada momento, en puntos específicos de la red, el origen del agua y porcentaje de mezclas

minimizando el impacto del cambio de agua en la percepción organoléptica de los consumidores.

Según los resultados obtenidos en el proyecto, los centros involucrados, Cetaqua, Aigües de Barcelona y Ens d'Abastament Aigües Ter Llobregat (ATL), acordaron la continuación del estudio con una nueva iniciativa que tiene tres objetivos principales. El primero, incrementar la capacidad

de los modelos, pasando de predicciones de origen del agua en mezclas binarias a mezclas ternarias, lo que se ajusta a la realidad cada vez más compleja de la red del área de Barcelona. El segundo objetivo se basa en profundizar en el efecto de las mezclas de agua y sus características fisicoquímicas en la percepción organoléptica del agua de los consumidores, a través de la realización de catas de agua con voluntarios. El

tercero busca automatizar los resultados obtenidos a través de modelos de predicción: reforzar el modelo de predicción del origen del agua en mezclas binarias, así como contemplar la incorporación de mezclas ternarias e implementar un modelo de predicción de la percepción organoléptica a través de parámetros fisicoquímicos del agua. Esta iniciativa se desarrollará durante el año 2023 y la primera mitad del 2024. 

MÁS ARTÍCULOS, REPORTAJES
Y NOTICIAS DEL SECTOR EN:

WWW.TECNOAQUA.ES