Detección temprana de anomalías en redes de distribución de agua mediante Matrix Profile: comparación entre enfoques multivariantes y estrategias bivariantes basadas en órdenes admisibles

Raluca Iulia Cozma Mircescu[1,2][0009−0008−9886−8516], Gabriel Viscarret Atienza[1], Humberto Bustince[2][0000−0002−1279−6195], Francisco Javier Fernández[2][0000−0003−4427−3935], and Cédric Marco-Detchart[2,3][0000−0002−4310−9060]

[1] dateando S.L., Pol. Ind. Berroa 19, 31192 Tajonar, Navarra, España
rcozma@dateando.tech, gviscarret@dateando.tech
[2] Dpto. de Estadística, Informática y Matemáticas, Universidad Pública de Navarra, Campus Arrosadia, 31006 Pamplona, España
cozma.179091@e.unavarra.es, bustince@unavarra.es, fcojavier.fernandez@unavarra.es, cedric.marco@unavarra.es
[3] IdiSNA, Instituto de Investigación Sanitaria de Navarra, Hospital Universitario de Navarra, 31008 Pamplona, España

Resumen El Agua No Registrada (ANR) supone pérdidas millonarias a nivel global, impulsando la búsqueda de métodos de detección temprana de anomalías en redes de distribución. Este trabajo presenta la aplicación y comparación de dos estrategias bivariantes (caudal Q, presión P) sobre datos reales del SCADA de la empresa dateando (Navarra), en entornos dinámicos no etiquetados con elevada incertidumbre epistémica. En este trabajo se compara la aproximación *mstump* estándar para búsqueda de *top-k discords* con una estrategia basada en órdenes admisibles. Frente a la naturaleza compensatoria del *mstump*, que enmascara precursores críticos cuando una variable normaliza la desviación de la otra, el uso de órdenes admisibles jerarquiza las variables según cinco escenarios operativos físicamente interpretables (fuga, anomalía de control, dominancia de Q, dominancia de P y referencia), segregando eventos incipientes antes de su escalada a fallos graves en la regulación de presión (PRV). Los resultados demuestran que el método propuesto iguala la eficacia diagnóstica del estándar, incorpora un escenario de anomalía de control inaccesible para éste, e identifica criterios objetivos de selección de ventana temporal sin requerir etiquetas ni validación externa.

Keywords: Matrix Profile · Top-k discords · Órdenes admisibles · Incertidumbre · Hidráulica urbana.

1. Introducción

El Agua No Registrada (ANR), desperdiciada principalmente por fugas y roturas de tuberías, genera pérdidas globales de aproximadamente 39.000 millones de USD al año [5]. En Navarra, el ANR alcanza el 22 % [4], lo que pone de manifiesto la necesidad de desarrollar sistemas capaces de anticiparse a los fallos.

En este contexto, esta investigación se desarrolla dentro de *Aquando*, la plataforma de gestión inteligente del agua de la empresa *dateando*. El objetivo es mejorar las capacidades de detección de anomalías de la plataforma, actualmente basadas en umbrales fijos, mediante enfoques más flexibles y adaptativos capaces de responder a la variabilidad real de las redes hidráulicas. Para ello, se propone un enfoque que busca un equilibrio entre precisión, interpretabilidad y eficiencia operativa, tres requisitos que los modelos de *Deep Learning* raramente cumplen de forma simultánea en entornos industriales [3,7]. El enfoque propuesto combina la información de caudal y presión mediante *Matrix Profile*[4] y órdenes admisibles, considerando las variables caudal y presión y priorizando sobre otra según el contexto operativo. El resultado es un método capaz de detectar señales tempranas de fallo de forma robusta, sin necesidad de datos etiquetados y adaptado a las condiciones reales de operación de la red [6].

2. Preliminares

Esta investigación analiza series temporales de caudal (Q) y presión (P) recogidas por el sistema SCADA de *dateando* en depósitos y distintos sectores de la red de distribución de Navarra, con mediciones cada 5 minutos. Se trata de datos procedentes de un entorno dinámico y no etiquetado, caracterizado por estacionalidad cíclica, una tendencia global estable y ruido operativo, lo que dificulta distinguir el comportamiento normal de las anomalías. Para abordar este problema, se realiza una limpieza de datos seguida de una descomposición STL (*Seasonal-Trend decomposition using Loess*) [1], que permite separar tendencia, estacionalidad y ruido. De este modo, se establece una línea base del comportamiento normal de la red. A partir de esta referencia, el objetivo no es únicamente detectar fallos evidentes, sino identificar precursores, es decir, pequeñas desviaciones que aparecen antes de una avería y que actúan como señales tempranas de degradación del sistema.

Para ello, se emplea *Matrix Profile* (MP), una estructura de datos que, dado un fragmento de señal de longitud L, mide su grado de similitud respecto al fragmento más parecido dentro de toda la serie. Los fragmentos más disimilares, denominados *top-k discords*, se consideran candidatos a anomalía. Se comparan dos formas de combinar las señales Q y P:

Agregación estándar (*mstump*[5]): Promedia las distancias de Q y P antes de identificar el *discord*. La limitación de este enfoque es que, si P se mantiene

[4] https://github.com/matrix-profile-foundation/matrixprofile
[5] https://github.com/stumpy-dev/stumpy

Cuadro 1. Escenarios operativos y criterio prioritario $V_1 = \mathbf{A}_{[1,\cdot]} \cdot [d_Q, d_P]^\top$.

Escenario	$\det(\mathbf{A})$	V_1 (criterio prioritario)
Fuga/Rotura	$-2{,}00$	$d_Q + d_P$ (magnitud total)
Anomalía de Control	$+2{,}00$	$d_Q - d_P$ (incoherencia Q-P)
Dominancia Q	$+0{,}68$	$0{,}8\, d_Q + 0{,}2\, d_P$
Dominancia P	$-0{,}68$	$0{,}2\, d_Q + 0{,}8\, d_P$
Referencia	$+0{,}98$	$0{,}7\,(d_Q + d_P)$

estable, su comportamiento puede compensar una desviación relevante en Q, ocultando posibles precursores.

Agregación con órdenes admisibles: En lugar de promediar, se emplean órdenes admisibles [2], definidos como relaciones de orden total que refinan la estructura de un orden parcial dado, para asignar una prioridad explícita a cada variable a través de la transformación $\mathbf{V} = \mathbf{A} \cdot [d_Q,\, d_P]^\top$. La primera componente V_1 determina el ranking de anomalías, mientras que V_2 actúa únicamente como criterio de desempate. La matriz $\mathbf{A}$ se selecciona según el escenario operativo que se desee analizar (Tabla 1), de forma que una anomalía en la variable prioritaria no quede enmascarada por la estabilidad de la otra.

3. Resultados y Discusión

La validación se llevó a cabo con datos reales de un sector de la red de distribución de agua de Navarra, recogidos entre agosto y septiembre de 2023, con un total de 17 568 muestras. Durante este periodo se produjo un colapso crítico entre los días 11 y 14 de septiembre, precedido por señales sutiles de inestabilidad a finales de agosto.

El análisis en distintas escalas temporales (L entre 1 y 5 días) muestra que la detección de anomalías depende fuertemente de la ventana elegida. Mientras que el fallo masivo de septiembre domina la mayoría de escalas, la configuración de $L = 3$ días destaca por identificar claramente el precursor de finales de agosto como el evento más singular. Esto sugiere que las 72 horas constituyen el intervalo clave para capturar las primeras señales de inestabilidad antes del fallo.

El método propuesto va un paso más allá del *mstump*, detectando la anomalía y clasificándola al mismo tiempo. Los escenarios de Fuga/Rotura y Referencia reproducen los resultados del estándar, validando la coherencia interna del método. Sin embargo, el escenario de Anomalía de Control aporta un valor diferencial, identificando como evento principal el precursor de finales de agosto, asociado a un comportamiento anómalo en los sistemas de regulación (válvulas PRV). Esto indica que el método es capaz de detectar pérdidas de coherencia entre variables antes de que se produzca un fallo evidente. Al evitar la compensación entre señales, se hacen visibles problemas que los métodos basados en promedios tienden a ocultar. Finalmente, la ventana de $L = 4$ días se identifica como el mejor compromiso global, ya que todos los escenarios coinciden en señalar el mismo

evento crítico. Ventanas más cortas fragmentan la anomalía, mientras que ventanas más largas la diluyen al mezclarla con el comportamiento habitual de la red. En conjunto, el enfoque permite avanzar hacia un mantenimiento proactivo, donde cada alerta se asocia a una causa física clara sin necesidad de etiquetado previo.

4. Conclusiones y Trabajo Futuro

La incorporación de órdenes admisibles al *Matrix Profile* mejora el diagnóstico respecto a enfoques basados únicamente en magnitud estadística. En particular, el intervalo de 3 a 4 días se identifica como la escala más relevante del sistema. La ventana de 3 días permite capturar precursores sutiles que aparecen varios días antes del fallo, mientras que la de 4 días confirma la degradación estructural al obtener consenso entre las señales de caudal y presión. Esta combinación permite distinguir de forma más clara entre variaciones operativas habituales y anomalías reales, proporcionando una detección con margen suficiente para intervenir. Como trabajo futuro, se plantea automatizar la selección de la ventana óptima y explorar técnicas de análisis espectral que permitan caracterizar mejor los fallos de control, así como reducir la sensibilidad del método al ruido en los datos.

Acknowledgments. Este trabajo ha sido financiado por la ayuda de «Doctorados Industriales 2025» del Gobierno de Navarra (referencia 0011-1408-2025-000005).

Disclosure of Interests. Los autores declaran que no tienen intereses en competencia relevantes no declarados para el contenido de este trabajo.

Referencias

1. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I.: STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics **6**(1), 3–73 (1990)
2. De Miguel, L., Bustince, H., Fernandez, J., Induráin, E., Kolesárová, A., Mesiar, R.: Construction of admissible linear orders for interval-valued Atanassov intuitionistic fuzzy sets with an application to decision making. Inf. Fusion **27**, 189–197 (2016)
3. Fu, G., et al.: The role of deep learning in urban water management. Water Res. **223**, 118983 (2022)
4. INE: Estadística sobre el Suministro y Saneamiento del Agua. Instituto Nacional de Estadística, Madrid (2024)
5. Liemberger, R., Wyatt, A.: Quantifying the global non-revenue water problem. J. Water Supply: Res. Technol.-Aqua **68**(1), 1–14 (2019)
6. Najafi, H., Huang, J.J.: Reliable anomaly detection in water systems. J. Water Process Eng. **76**, 108207 (2025)
7. Syed, T.A., et al.: Smart water management with digital twins. Water **16**(23), 3451 (2024)