

# Mejora del reconocimiento de usos finales del agua mediante la simplificación de la traza de caudal: un caso de estudio

## *Simplifying water consumption flow traces for improving end use recognition: a case study*

Pastor-Jabaloyes, L.<sup>a1</sup>, Arregui, F.J.<sup>a2</sup>, Cobacho, R.<sup>a3</sup>

<sup>a1</sup>ITA-Grupo de Ingeniería y Tecnología del Agua, Dpto. de Ingeniería del Agua y Medio Ambiente, Universitat Politècnica de València, Camino de Vera s/n, València 46022. e-mail: <sup>a1</sup>laupasja@ita.upv.es, <sup>a2</sup>farregui@ita.upv.es, <sup>a3</sup>rcobacho@ita.upv.es

Recibido: 05/02/2018

Aceptado: 30/05/2018

Publicado: 31/10/2018

**Citar como:** Pastor-Jabaloyes, L., Arregui, F.J., Cobacho, R. 2018. Simplifying water consumption flow traces for improving end use recognition: a case study. *Ingeniería del agua*, 22(4), 195-208. <https://doi.org/10.4995/la.2018.9476>

### RESUMEN

El éxito de estrategias para la desagregación y clasificación automática de los consumos de agua en usos finales depende de un adecuado filtrado previo de las trazas de caudal registradas. Se propone un nuevo algoritmo de filtrado, cuyos parámetros de entrada se ajustan mediante un proceso de calibración automático por evento de consumo, asegurando la adaptabilidad y simplificación de la traza filtrada a la original. Esta herramienta se aplica a un caso de estudio mediante el análisis de 5210 eventos de consumo, procedentes de una campaña de monitorización en una ciudad española. Los resultados muestran que el filtro es capaz de simplificar sustancialmente las trazas de caudal manteniendo la información esencial. En media, las trazas de caudal de eventos más complejos pueden definirse con menos del 10% de los puntos de las trazas originales. Además, el análisis realizado permite identificar diversas estrategias para mejorar y optimizar el proceso de filtrado.

**Palabras clave** | filtrado de la señal de caudal; demanda de agua residencial; microcomponentes del agua; contadores inteligentes; desagregación en usos finales del agua.

### ABSTRACT

*The success of automatic water end use disaggregation and classification strategies greatly depends on the filtering and signal conditioning of the flow traces recorded. The work presented proposes a new filtering algorithm of water consumption flow traces. To improve the performance of the filter, the parameters driving the process are found per event by an automatically calibration procedure. These parameters are selected to ensure the maximum adaptability and simplification of the filtered flow traces. The methodology has been tested with 5210 consumption events obtained from a measurement campaign conducted in a Spanish city. The results obtained show that the filtering algorithm is capable of significantly simplifying the original flow traces while maintaining their main characteristics. On average, it has been found that the most complex events can be described using only 10% of the input data. This analysis can be used to make more efficient the filtering procedure proposed*

**Key words** | filtering of the flow trace; residential water demand; water microcomponents; smart metering; water end use disaggregation.

## INTRODUCCIÓN

La mejora del uso eficiente del agua en áreas urbanas es uno de los objetivos señalados a nivel internacional (UNEP, 2011) para afrontar los retos futuros que plantean la creciente demanda de la población, las limitaciones impuestas por el recurso hídrico disponible y el impacto del cambio climático. La gestión de la demanda es uno de los instrumentos que permiten la consecución de este objetivo, cuyo principal dato de entrada radica en el conocimiento de cuándo, cómo y quién hace uso del agua.

El análisis de usos finales del agua es una potente herramienta que brinda soporte efectivo a la gestión y modelación de la demanda de agua urbana. El desarrollo de los contadores inteligentes de alta resolución ha hecho viable la caracterización precisa de los perfiles de consumo de usos finales (Cominola *et al.*, 2015), los cuales podrían ofrecer retroalimentación *ad-hoc* para cada usuario que promueva cambios en su comportamiento hacia la eficiencia (Fielding *et al.*, 2013). No obstante, este tipo de estudios requiere análisis de *trazas de caudal* –series temporales en las que los datos de caudal se registran a intervalos cuya duración es del orden de segundos– que conllevan una gran inversión en tiempo y capital humano, no viable desde el punto de vista económico (Nguyen *et al.*, 2013a). Por esta razón, se hace necesario el desarrollo de herramientas automáticas de desagregación y clasificación en usos finales.

El filtro de las trazas de caudal es el paso previo a las operaciones que integran las herramientas automáticas de análisis de usos finales del agua. El principal propósito que persigue es suavizar la señal, eliminando ruido de la misma, y depurar los inicios y finales de evento, así como los cambios de rasante. En consecuencia, el filtrado condiciona la desagregación, a través de la cual es posible determinar las características físicas de los eventos individuales (volumen, caudal medio y duración, entre otros). A su vez, las características físicas de los eventos de consumo constituyen el dato de entrada de la operación de clasificación. Por este motivo, su desempeño depende de que éstas sean representativas del uso final que debe asociarse a cada evento. En definitiva, el filtro de la traza de caudal afecta al resto de procesos y de ahí su importancia.

Las herramientas comerciales disponibles para la desagregación de usos finales del agua, Trace Wizard® (DeOreo *et al.*, 1996), Identiflow® (Kowalski y Marshallsay, 2003) y BuntBrainForEndUses® (Arregui, 2015), están descritas someramente en la literatura científica y no se ha encontrado evidencia sobre si emplean algún tipo de filtro. Por otro lado, Larson *et al.* (2012) proponen para la identificación de usos finales un enfoque alternativo basado en datos recogidos a través de sensores de presión. En este caso, las estrategias de filtrado utilizadas se describen con detalle, si bien el tipo de señal que procesan es distinto al obtenido a través de los contadores de agua, por lo que no son aptos para la depuración de la traza de caudal.

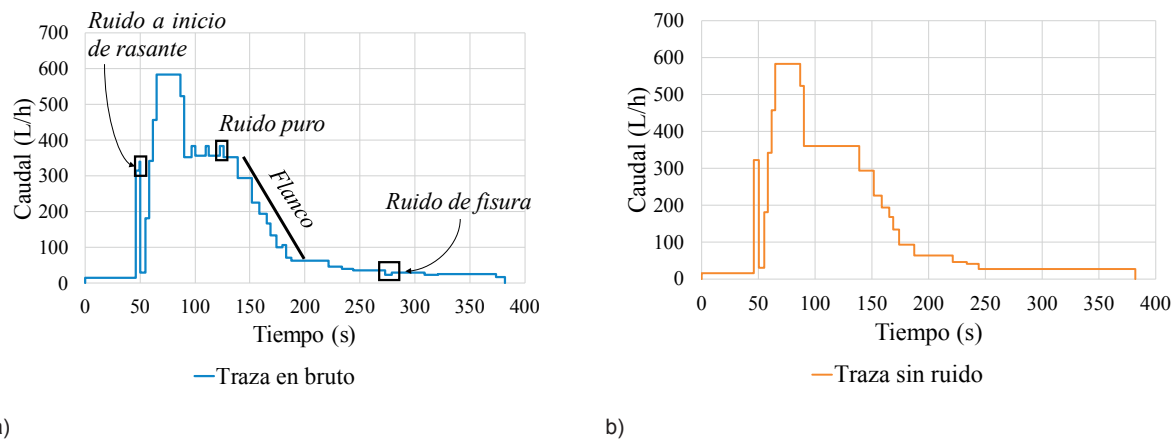
El nuevo filtro para la traza de caudal, objeto del presente artículo, toma como punto de partida el trabajo llevado a cabo por Nguyen *et al.* (2013b). El filtro empleado en este trabajo, *gradient vector filtering*, se basa en examinar el cambio de gradiente a lo largo de la serie de caudal. Las fluctuaciones que no son debidas al inicio o fin de un evento son identificadas mientras el gradiente asociado quede por debajo de un cierto umbral. Al aplicar este filtro a la colección de datos disponible para el presente estudio, se observaron ciertas imprecisiones que dificultaban la posterior desagregación de un evento solapado. Es por ello que se hizo necesario el desarrollo de un filtro que superara dichas imprecisiones y mejorar la calidad de la señal de caudal.

El algoritmo de filtrado propuesto complementa el análisis del gradiente con comprobaciones de volumen y saltos de caudal. R (R Core Team, 2013) es el lenguaje de programación seleccionado para escribir el código del filtro, el cual no necesita para ser ejecutado ningún paquete adicional de R. Consta de 10 parámetros de entrada, los cuales hacen del mismo una herramienta flexible, capaz de enfrentar una amplia casuística de eventos. En tanto que trabajar con los valores por defecto establecidos por la experiencia de los autores no asegura en cualquier caso la adaptación adecuada de la traza de caudal filtrada a la traza original, se ha desarrollado una metodología de calibración automática, que está descrita con detalle en Pastor-Jabaloyes *et al.* (2018). En el presente artículo se testeará el desempeño del filtrado de la señal, ajustados los parámetros de entrada mediante un proceso de calibración automático, en una muestra de 5210 eventos registrados en una campaña de monitorización realizada en una ciudad española. Además, se llevará a cabo un análisis descriptivo de los valores que adoptan los parámetros de entrada del filtro en un intento de dilucidar cuáles son aquellos que conducen a soluciones de filtrado óptimas, así como encontrar estrategias que permitan aumentar la eficiencia del proceso de calibración.

## MATERIAL Y MÉTODOS

### Arquitectura del filtro

Esencialmente, el algoritmo de filtrado tiene que emular las simplificaciones mentales que un ser humano lleva a cabo cuando procesa un evento de consumo. Por ejemplo, cuando un evento solapado es recortado manualmente, el analista no confunde los saltos de caudal debidos al ruido de la señal con aquellos que corresponden al inicio o fin de un evento individual. Ignorar los primeros implica una simplificación. En línea con esta idea, la subrutina propuesta calcula en primer lugar las características físicas de la traza de caudal (ej. cambios de gradiente o magnitud de un salto de caudal). A continuación, estas características son comparadas con una serie de umbrales. Finalmente, una simplificación de la traza se aplica siempre y cuando ciertas condiciones basadas en dichos umbrales se cumplan. En total, ha sido necesario definir 10 umbrales, los cuales constituyen los parámetros de entrada del algoritmo de filtrado.



**Figura 1** a) traza de caudal en bruto de un evento solapado con los distintos tipos de ruido señalados sobre ella; b) traza de caudal obtenida después de la depuración del ruido.

El proceso general descrito se detalla seguidamente. El filtro se ha dividido en cuatro bloques. El primer bloque de código tiene como objetivo suprimir los saltos de caudal que corresponden al ruido de la señal. Se han considerado tres tipos principales de ruido (Figura 1-a): i) *Ruido puro*, que se define como aquel volumen de agua por debajo de un umbral que está contenido entre dos saltos de caudal consecutivos de igual magnitud y distinto signo; ii) *Ruido de fisura*, que conceptualmente es similar al anterior, se diferencia en el hecho de que el volumen contenido corresponde a un hueco en vez de a un saliente; iii) Ruido en el inicio o fin de un flanco, entendido este último como una sección con pendiente completa, formada por varios saltos de caudal consecutivos, que se sitúa entre dos secciones horizontales. Por tanto, para caracterizar estos tipos de ruido se hace necesario determinar el volumen y el gradiente que corresponden a un salto de caudal. Dado un vector  $\mathbf{tc}=(tc_1, tc_2, \dots, tc_i, \dots, tc_m)$ , que contiene la traza de caudal expresada en litros por hora (L/h) registrada para cada instante  $t_i$  en milisegundos (ms), el vector de saltos de caudal  $\mathbf{s}$  y el vector de ventanas temporales  $\mathbf{vt}$  se definen como:

$$s_j = tc_{j+1} - tc_j, \quad 1 \leq j < n \quad (1)$$

$$vt_j = t_{j+1} - t_j, \quad 1 \leq j < n \quad (2)$$

Donde el índice  $j$  se refiere a aquellos puntos de  $\mathbf{tc}$  que satisfacen la siguiente condición:

$$tc_{i+1} - tc_i \neq 0, \quad 1 \leq i < n \quad (3)$$

A partir de los vectores  $s$  y  $vt$ , el vector volumen  $v$  se define como:

$$\begin{cases} v_j = |s_j| \cdot vt_j, & \text{si } s_j > 0 \\ v_j = |s_j| \cdot vt_{j-1}, & \text{si } s_j < 0 \end{cases} \quad 1 \leq j < n \quad (4)$$

Por otro lado, para calcular el vector gradiente  $g$  es necesario crear un parámetro de escala ( $pl$ , valores entre 130 (h·ms)/L y 170 (h·ms)/L; valor por defecto de 150 (h·ms)/L) para asegurar que el gradiente aumenta lentamente en función de la magnitud del salto de caudal ( $vt_i > 6000$  ms), con lo que:

$$g_j = pl \cdot s_j / vt_j, \quad 1 \leq j < n \quad (5)$$

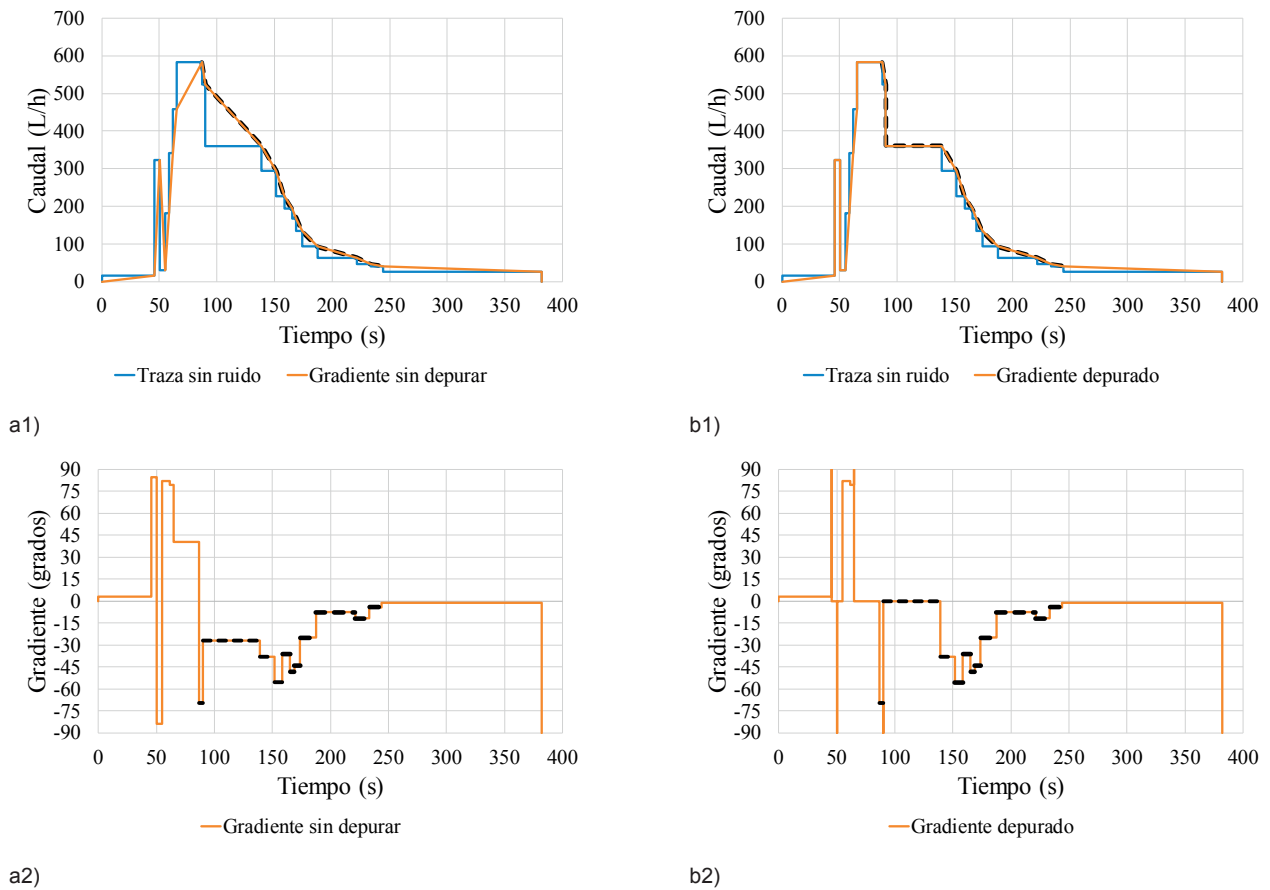
Además de  $pl$ , en este bloque se han establecido cuatro parámetros más: a) *volumen máximo* ( $p2$ ) contenido por el ruido puro o de fisura (rango de valores permitidos entre 0.1 L y 0.18 L; valor por defecto de 0.16 L); b) *salto de caudal máximo* ( $p3$ ) para tener en cuenta la posibilidad de existencia de ruido de fisura (puede tomar valores entre 40 L/hora y 120 L/hora; valor por defecto de 80 L/h); c) *diferencia máxima de gradiente* ( $p4$ ) expresada en grados, a partir del cual se considera que se produce realmente un cambio de gradiente y comienza o finaliza un flanco (el límite superior es 75 grados y el inferior 30 grados; valor por defecto de 40 grados); d) *porcentaje de un flanco completo* ( $p5$ ) que supone el primer o último salto de caudal que pertenece al mismo, de forma que si simultáneamente es superado el umbral de porcentaje y se cumplen las condiciones de ruido de fisura, el salto no sea suprimido, ya que en caso contrario el flanco podría quedar distorsionado (valores entre 0.001% y 10%; valor por defecto del 5%). El resultado final obtenido se muestra en la Figura 1-b (*Traza sin ruido*).

El segundo bloque de código lleva a cabo el análisis del gradiente. Tal y como el cálculo del gradiente ha sido definido, puede conducir a error con respecto a la detección del inicio o el fin de una sección con pendiente. Esto se observa en el ejemplo de la Figura 2-a1, donde una bajante partida en dos tramos se interpretaría como un único flanco (línea discontinua), dado que el gradiente se mantiene en valores relativamente altos a lo largo de la misma (Figura 2-a2). La solución adoptada para corregir posibles errores es forzar a que el algoritmo tenga en cuenta en el cálculo del gradiente un punto auxiliar cuando se cumple que el salto de caudal y la ventana temporal están por encima de cierto umbral. También se considera un punto adicional en aquellos tramos donde se produce un cambio de signo en el gradiente. En consecuencia, ha sido necesario definir 4 parámetros: a) *Salto de caudal mínimo* ( $p6$ , valores entre 40 L/h y 120 L/h; valor por defecto de 100 L/h); b) *ventana temporal mínima* ( $p7$ , valores entre 5000 ms y 14000 ms; valor por defecto de 10000 ms); c) Otros dos parámetros referentes a la ventana temporal creados *ad-hoc* para el tratamiento de pulsos de duración corta (menor a 5 min): *ventana temporal mínima* ( $p8$ , rango de variación entre 3000 ms y 11000 ms; valor por defecto de 6000 ms) y *porcentaje mínimo* ( $p9$ ) que dicha ventana temporal supone respecto a la duración total del pulso (puede oscilar entre el 0.001% y 10%; valor por defecto del 5%). El resultado obtenido al término de esta operación se muestra en la Figura 2-b1 (*Gradiente depurado*).

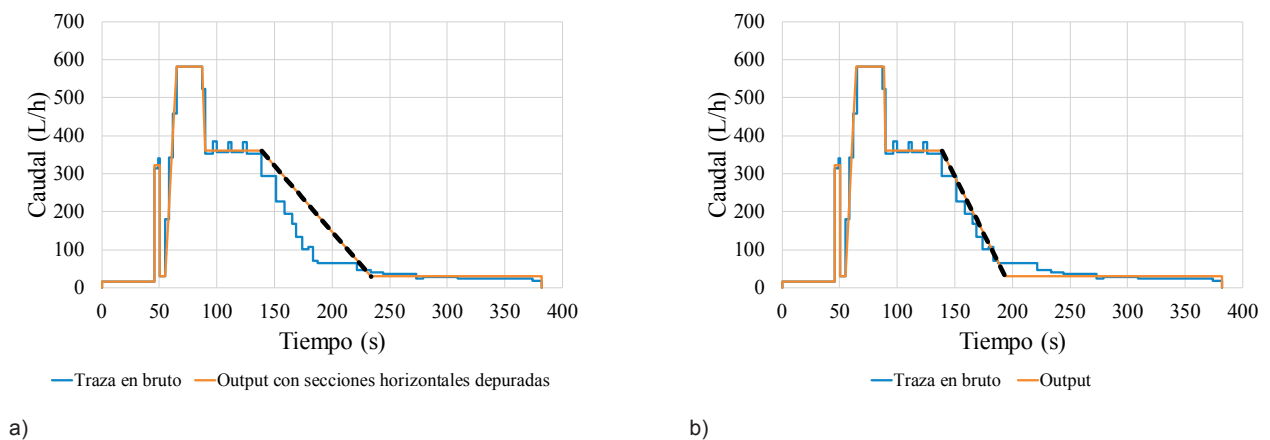
Una vez corregido el gradiente, el segundo bloque finaliza al clasificar cada tramo como *horizontal* o *con pendiente*. Para ello es necesario establecer el último parámetro de entrada al filtro, el *gradiente máximo* ( $p10$ ) por debajo del cual se considera que un tramo es horizontal (puede tomar valores entre 3 y 30 grados; valor por defecto de 10 grados).

El tercer bloque de código detecta las secuencias de secciones clasificadas en la misma categoría (*horizontal* o *con pendiente*). Una vez delimitadas, si el salto de caudal asociado es menor que  $p3$ , esa secuencia se suprime por no ser su magnitud lo suficientemente importante como para considerarla un flanco de inicio o fin de un evento de consumo individual.

Por último, el cuarto bloque de código rectifica la traza de caudal de salida (*Output*) para aproximar su volumen al de la traza en bruto. En primer lugar, dada la misma ventana temporal, el algoritmo iguala el caudal de una sección horizontal con el caudal medio alcanzado en la traza original. Para las secciones restantes, su pendiente se ve incrementada para compensar un exceso o defecto de volumen contenido. La Figura 3 ilustra un ejemplo de esto, en la que se corrige un exceso de volumen (Figura 3-a, línea discontinua). El resultado obtenido al final del proceso completo se muestra en la Figura 3-b (*Output*).



**Figura 2** | a1) y a2) muestran, respectivamente, el gradiente sobre la traza de caudal y su magnitud expresada en grados, antes de realizar la corrección en el cálculo del gradiente (parámetro de escala adoptado de 150 L/(h·ms)). b1) y b2) son análogas a las dos anteriores, pero parten del gradiente depurado.



**Figura 3** | a) traza de caudal con los tramos horizontales corregidos; b) resultado final del proceso completo de filtrado.

El algoritmo de filtrado propuesto puede ejecutarse automáticamente una vez establecidos los valores de los 10 parámetros de entrada. Estos hacen del filtro una herramienta flexible, capaz de enfrentar una amplia casuística de eventos, cuyas características físicas varían considerablemente dado que dependen del tipo de dispositivo y el estado de la instalación interna de la vivienda,

así como de los hábitos de consumo y la cultura de sus ocupantes. En tanto que trabajar con los valores por defecto no asegura en cualquier caso la adaptación adecuada de la traza de caudal filtrada a la traza original, surge la necesidad de ajustar los parámetros de entrada para cada evento, con el objetivo de que la calidad del resultado obtenido garantice la correcta consecución de la siguiente fase en el proceso de desagregación: el recorte.

Así pues, se ha desarrollado una metodología de calibración automática del algoritmo de filtrado, que está descrita con detalle en Pastor-Jabaloyes *et al.* (2018). A grandes rasgos, dicha metodología plantea el proceso de calibración como un problema de optimización, en el que se buscan aquellas combinaciones de parámetros de entrada que *minimizan el número de vértices* que describen la traza de caudal filtrada –la solución óptima sería la más *simple* posible, un rectángulo de altura el caudal medio de la traza de caudal original– y, a su vez, *maximizan la adaptación* a la traza original –la solución óptima en este caso sería la propia traza de caudal original. Dado el carácter no-lineal y multi-objetivo del problema planteado, se seleccionó un método heurístico de resolución: el algoritmo genético de clasificación elitista no dominada NSGA-II (en inglés, *the Elitist Non-Dominated Sorting Genetic Algorithm NSGA-II*). El resultado arrojado por el NSGA-II conforma un Frente de Pareto, del cual se escogen las soluciones –ya que, por las características del problema, varias combinaciones de parámetros pueden dar la misma solución como resultado– que cumplen unos determinados criterios de adaptación y simplificación con respecto a la traza de caudal original.

En este artículo se analizan aquellas combinaciones de parámetros que satisfacen los criterios del *Filtro-A* definido en Pastor-Jabaloyes *et al.* (2018) para cada evento incluido en la muestra de validación, en el que el peso dado a la simplificación es mayor que el concedido a la adaptación. Dicho análisis se lleva a cabo en un intento de dilucidar cuáles son los valores habituales de los parámetros de entrada al filtro, así como encontrar estrategias que permitan aumentar la eficiencia del filtrado, ya que actualmente, dentro del proceso completo de desagregación descrito en Pastor-Jabaloyes *et al.* (2018), esta fase es la que lleva asociado mayor coste computacional.

## Descripción de la muestra para la validación

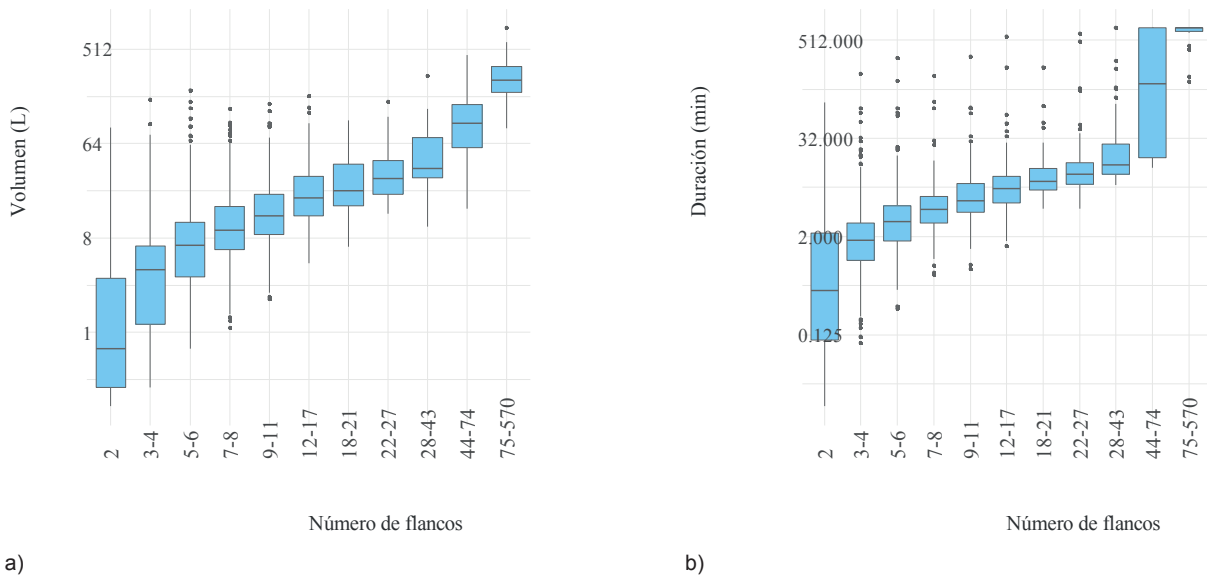
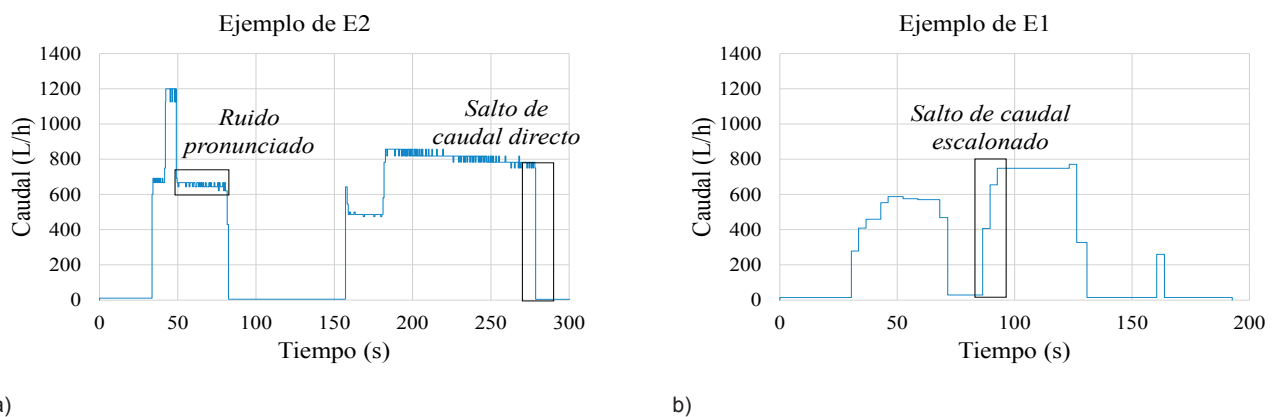
Los datos empleados para testear la adaptación del filtro a la traza de caudal original provienen de un estudio sobre demanda de agua residencial, cuya localización no puede ser desvelada por motivos de confidencialidad. Para dicho estudio se emplearon contadores Aquadis+ (ITRON) volumétricos de pistón rotativo (caudales nominales de 2.5 m<sup>3</sup>/h), con una resolución de 0.1 L por pulso, registrando los *loggers* el tiempo de ocurrencia de cada pulso con una resolución de 0.02 s. Los datos de consumo recabados en el estudio proceden de 74 viviendas residenciales sin usos exteriores dispuestas en cuatro edificios. Para el presente artículo se han seleccionado 10 casos que están distribuidos uniformemente con respecto al consumo medio diario de la muestra original. Del periodo de monitorización completo, que se extendió desde octubre de 2016 a enero de 2017, se han escogido 14 días consecutivos por vivienda. Ello corresponde a un total de 5210 eventos de consumo registrados. En la Tabla 1 se resumen los estadísticos básicos que describen tanto las características de las viviendas seleccionadas como de los eventos asociados. Hay que puntualizar asimismo que la existencia de eventos de larga duración (del orden de horas) y considerable volumen se debe a que la muestra incluye 3 viviendas con fuga continua.

Con el objetivo de obtener la mayor cantidad de información posible, tanto a la hora de evaluar el desempeño del filtro como de analizar las combinaciones de parámetros escogidas a través del proceso de calibración, los 5210 eventos de consumo han sido clasificados en distintas categorías atendiendo a su complejidad morfológica. El indicador escogido para evaluar dicha complejidad es el número de flancos presente en la traza de caudal filtrada. Otros candidatos considerados fueron el volumen y la duración, si bien finalmente se descartaron principalmente por no representar correctamente el grado de complejidad, ya que es relativamente fácil encontrar contraejemplos: tal es el caso de las duchas y las fugas, que, respectivamente, pueden llegar a consumir grandes volúmenes de agua y suelen ser de larga duración, pero que, al mismo tiempo, presentan normalmente configuraciones simples de la traza de caudal (únicamente dos flancos, uno de subida y otro de bajada). Por otro lado, la correlación existente al enfrentar el volumen y la duración con el número de flancos es clara, tal y como se muestra en la Figura 4, por lo que considerar adicionalmente alguno de estos indicadores resultaría redundante. En un intento de crear la mayor cantidad de categorías a la vez que el tamaño por cada una de ellas se mantenía por encima de 25 eventos, finalmente se han definido 11 niveles de complejidad (Figura 4, eje horizontal).



**Tabla 1** | Estadísticos básicos de las características de las viviendas y los eventos utilizados para la validación del filtro.

		Media	Mediana	1 <sup>er</sup> cuartil	3 <sup>er</sup> cuartil	Máximo	Mínimo
Datos por vivienda	Consumo diario (L/viv/día)	359.9	352	257.4	469.9	529.7	191.5
	Número de eventos	521	555	397	703	834	89
	Duración (min/evento)	9.7	2	0.5	4.2	721.5	0.02
Datos por evento	Volumen (L/evento)	10.2	3.8	0.7	8.9	816.9	0.2
	Caudal (L/h/evento)	224.7	178	68.4	349.5	1025.5	2.8

**Figura 4** | Distribución del volumen, a), y la duración, b), de los eventos que incluye cada categoría definida según número de flancos.**Figura 5** | a) ejemplo del ruido pronunciado y los saltos de caudal directos que predominan en los eventos del estudio utilizado en este artículo; b) ejemplo de un salto de caudal escalonado característico de las trazas registradas por otros equipos de medida diferentes a los empleados en el caso objeto de estudio.

Por último, cabe señalar que las características de los equipos de medición empleados y su configuración afectan a las características de las trazas de caudal en bruto recabadas. Así, las trazas registradas en el estudio utilizado en este artículo presentan saltos de caudal directos (no escalonados) al tiempo que un ruido pronunciado (Figura 5-a). Sin embargo, es importante señalar que

los autores han llevado a cabo validaciones complementarias de este mismo filtro con trazas generadas por equipos de medición de características diferentes, con resultados igualmente satisfactorios. Esto constituye una prueba de la flexibilidad de este algoritmo y de ahí que se le califique de *universal*. Más información al respecto puede encontrarse en Pastor-Jabaloyes *et al.* (2018).

## RESULTADOS

### Calidad del filtrado

Se han tenido en consideración cuatro indicadores para evaluar la calidad del filtrado obtenido. El primero de ellos es el error asociado al volumen total del evento, expresado como un porcentaje del volumen contenido por la traza en bruto. El segundo indicador es el índice de King-Gupta Efficiency (Gupta *et al.*, 2009), conocido como KGE, que evalúa la habilidad del filtro para respetar las tendencias marcadas por la traza en bruto. Este índice adimensional puede variar entre  $-\text{Inf}$  y 1, siendo este último el mejor valor. Se ha empleado el paquete de R *HydroGOF* para calcular dicho índice. Los dos últimos indicadores tenidos en cuenta miden el ahorro de memoria de almacenamiento: porcentaje de puntos suprimidos de la traza en bruto y reducción de los requerimientos de memoria, expresada en porcentaje respecto a la situación original y en bytes.

**Tabla 2** | Valores medios de los indicadores de desempeño por categoría de complejidad según número de flancos.

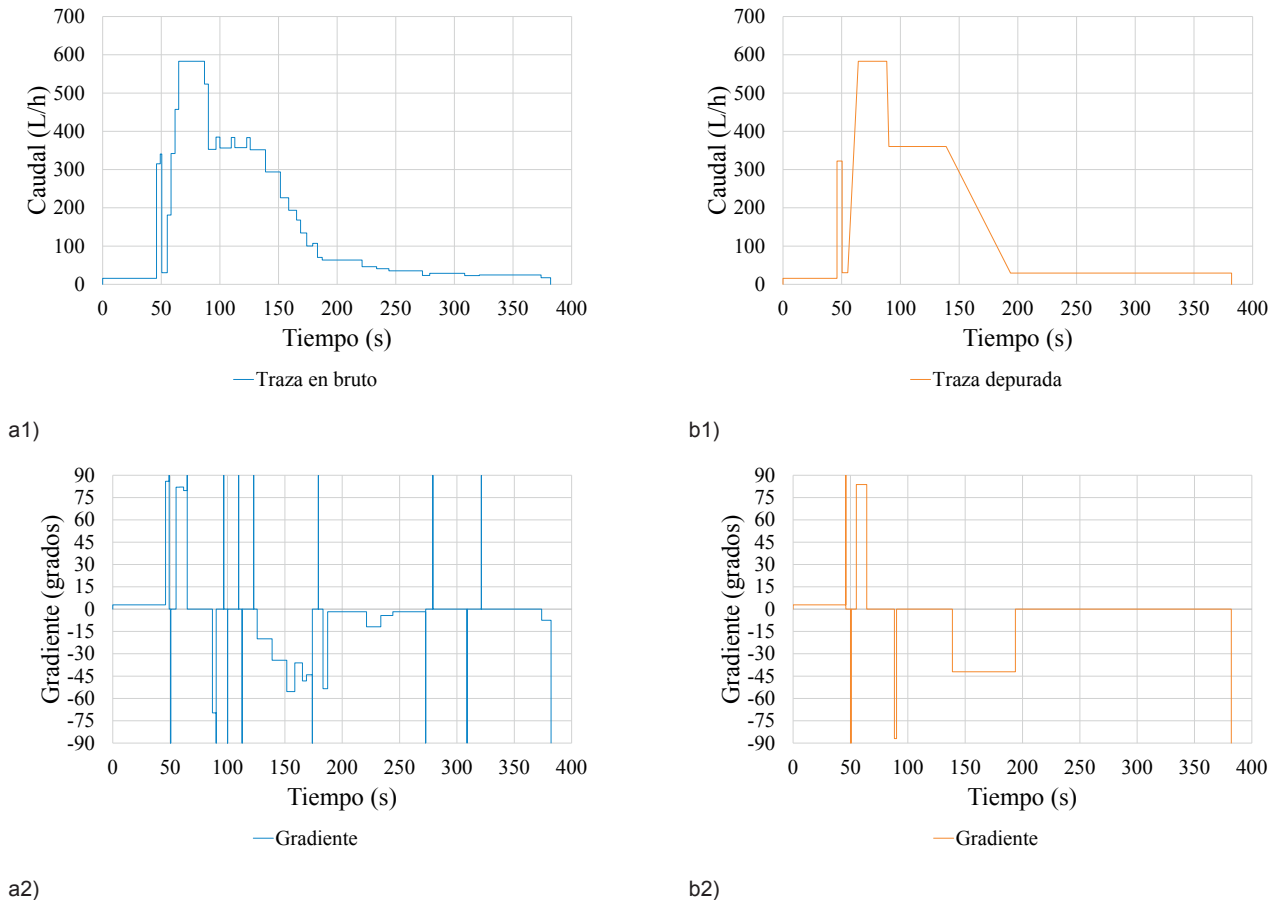
Número de flancos	2	3-4	5-6	7-8	9-11	12-17	18-21	22-27	28-43	44-74	75-570
<i>Número de eventos</i>	2293	1239	594	340	310	212	65	54	44	32	27
<i>Error del volumen total (%)</i>	0.07%	0.12%	0.10%	0.08%	0.06%	0.04%	0.04%	0.02%	0.03%	0.06%	0.04%
<i>KGE</i>	0.87	0.88	0.91	0.92	0.93	0.93	0.95	0.94	0.93	0.93	0.96
<i>Reducción del número de puntos (%)</i>	54.4%	73.7%	81.0%	84.0%	86.4%	87.1%	85.4%	86.1%	86.0%	88.9%	92.4%
<i>Reducción de los requerimientos de memoria (%)</i>	44.5%	66.9%	75.6%	79.8%	82.9%	84.3%	82.8%	84.0%	83.9%	87.8%	92.0%
<i>Reducción de los requerimientos de memoria (bytes)</i>	704	1564	2871	4189	5343	7113	8565	10289	15273	39970	93621

La Tabla 2 resume los valores medios alcanzados para cada indicador de desempeño y categoría de complejidad. Se puede observar que el volumen contenido por el *output* es prácticamente el mismo que el contenido por la traza en bruto independientemente de la categoría, dado que el porcentaje de error está próximo a cero. Con respecto a la bondad del ajuste, el valor medio de KGE se sitúa en torno a 0.9. La elevada precisión alcanzada por el filtro indica que se consigue simplificar la traza en bruto respetando las tendencias generales que marca. No obstante, el objetivo no es reproducir con exactitud los datos de entrada, por tanto, es deseable que los valores de KGE estén por debajo de 0.99. En relación a los indicadores de ahorro de memoria, se observa un incremento con la complejidad del evento. Los resultados muestran que en media el 54.4% de los puntos de la traza en bruto son suprimidos para los eventos más simples (2 flancos), aumentando este porcentaje hasta el 92.4% para los eventos de mayor complejidad. Se observa un patrón similar en el indicador de reducción de requerimientos de memoria. En consecuencia, a través del uso de este filtro podría almacenarse de forma más eficiente las grandes cantidades de información que genera un estudio de usos finales del agua.

Con respecto a las mejoras que introduciría el resultado alcanzado en el proceso de desagregación, la eficacia a la hora de simplificar el gradiente puede observarse en el ejemplo de la Figura 6. Ésta muestra un evento que corresponde a una cisterna (mecanismo de válvula de llenado) solapada con un grifo. El gradiente calculado a partir de la traza en bruto resulta errático (Figura 6-a2): la magnitud de los cambios de gradiente debidos al ruido de la señal es similar a la de aquellos que corresponden al inicio o fin de un nuevo evento. Esto se debe a que los saltos de caudal que corresponden a ruido ocurren en periodos cortos de duración. Adicionalmente, la distorsión que tiene lugar en torno a los 175 segundos (Figura 6-a1) rompe en dos una única sección con pendiente. En cambio, cuando la traza depurada es el dato de entrada (Figura 6-b1) el inicio y el fin de un nuevo evento pueden



ser identificados con mayor facilidad, ya que las distorsiones se han suprimido, mientras que las tendencias generales se respetan. Además, características específicas de un determinado dispositivo pueden ser reconocidas. Por ejemplo, el mecanismo de válvula de llenado genera una cola en la traza de caudal de la cisterna, cuyo gradiente tiene menor magnitud y se sostiene durante más tiempo que el correspondiente a un grifo (Figura 6-b2).



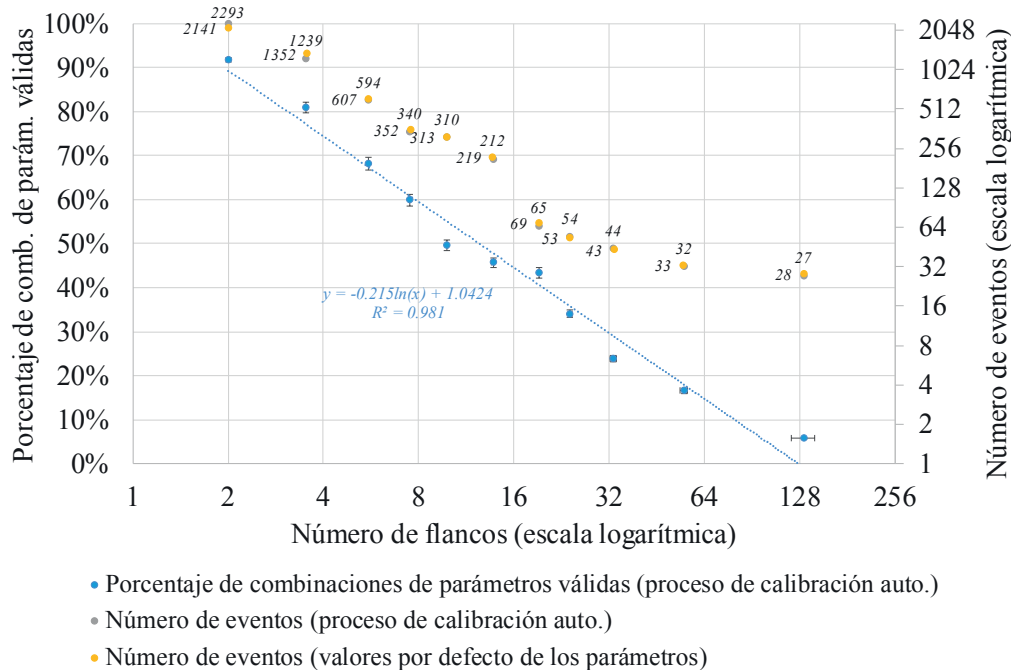
**Figura 6** | a1) y a2) muestran, respectivamente, la traza de caudal y la magnitud del gradiente expresado en grados, antes del filtrado. b1) y b2) son análogas a las dos anteriores, pero parten de la traza de caudal filtrada.

## Análisis de los valores de los parámetros de entrada seleccionados automáticamente

La selección de los valores de los parámetros del filtro se ha llevado a cabo por evento a través del proceso de calibración automático descrito anteriormente. El objetivo perseguido a través de este procedimiento es asegurar que las combinaciones de parámetros seleccionadas para cada evento proporcionen como solución una traza filtrada que cumpla unos requerimientos determinados de simplificación y adaptación con respecto a la traza original. A continuación, se analizan los resultados obtenidos con el fin de identificar posibles mejoras a implementar sobre el proceso completo de desagregación descrito en Pastor-Jabaloyes *et al.* (2018).

La primera parte de este análisis se ha realizado atendiendo a la cantidad de combinaciones válidas de parámetros obtenidas para cada evento. El propósito de este análisis es observar si existe alguna correlación entre: i) el porcentaje de soluciones escogidas –según los requerimientos establecidos de adaptación y simplificación– del total de soluciones que conforman el Frente de Pareto que proporciona el NSGA-II; y ii) el grado de complejidad del evento, que se evalúa a través del número de flancos. Los resultados obtenidos se muestran resumidos en la Figura 7, en el que cada punto de la primera serie –*Porcentaje de combinaciones de parámetros válidas (proceso de calibración auto)*– corresponde al promedio del número de flancos y porcentaje de combinaciones

válidas de parámetros por categoría de complejidad. Queda patente que existe una correlación logarítmica clara entre ambas variables, con un  $R^2$  de 0.98. Es decir, existe una gran cantidad de combinaciones de parámetros que dan como resultado una traza filtrada admisible para los eventos más simples (2 flancos), mientras que para los eventos más complejos (75 o más flancos) el número de combinaciones válidas obtenidas como resultado del proceso de calibración es en la mayor parte de los casos única.



**Figura 7** | Porcentajes de combinaciones válidas de parámetros para cada categoría de complejidad y número de eventos de cada una.

Por otro lado, el número de eventos por categoría de complejidad (eje secundario de la Figura 7) se ha calculado según los resultados del proceso de calibración lanzado para cada evento y los obtenidos al ejecutar el filtro con los valores por defecto definidos anteriormente. El hecho de que el número de eventos en cada categoría sea prácticamente el mismo indica que para esta mera clasificación de eventos (no así para el posterior filtrado previo a la fase de recorte) se hace innecesario llevar a cabo el proceso de calibración completo. Además, en el caso de los eventos más simples (2 flancos), para 2141 casos se ha obtenido el mismo resultado lanzando el filtro con los parámetros por defecto o ajustando estos a través del proceso de calibración. Este resultado es esperable, ya que se ha verificado que existe una gran cantidad de combinaciones válidas de parámetros para esta categoría de complejidad. Esto último tiene una gran importancia para la mejora futura del proceso de filtrado, en tanto que la primera categoría contiene siempre un gran número de eventos (casi la mitad del total para el caso objeto de estudio) y el evitar, con fiabilidad suficiente, la ejecución individual del NSGA-II en cada uno de ellos redundaría en una reducción muy significativa del tiempo que tarda en llevarse a cabo el proceso completo de desagregación.

La segunda parte de este análisis consiste en estudiar la distribución de los valores que escoge el proceso de calibración automático de cada uno de los diez parámetros del filtro para los eventos incluidos en cada categoría de complejidad. Las gráficas obtenidas se muestran en la Figura 8. Las conclusiones que se extraen de las mismas son:

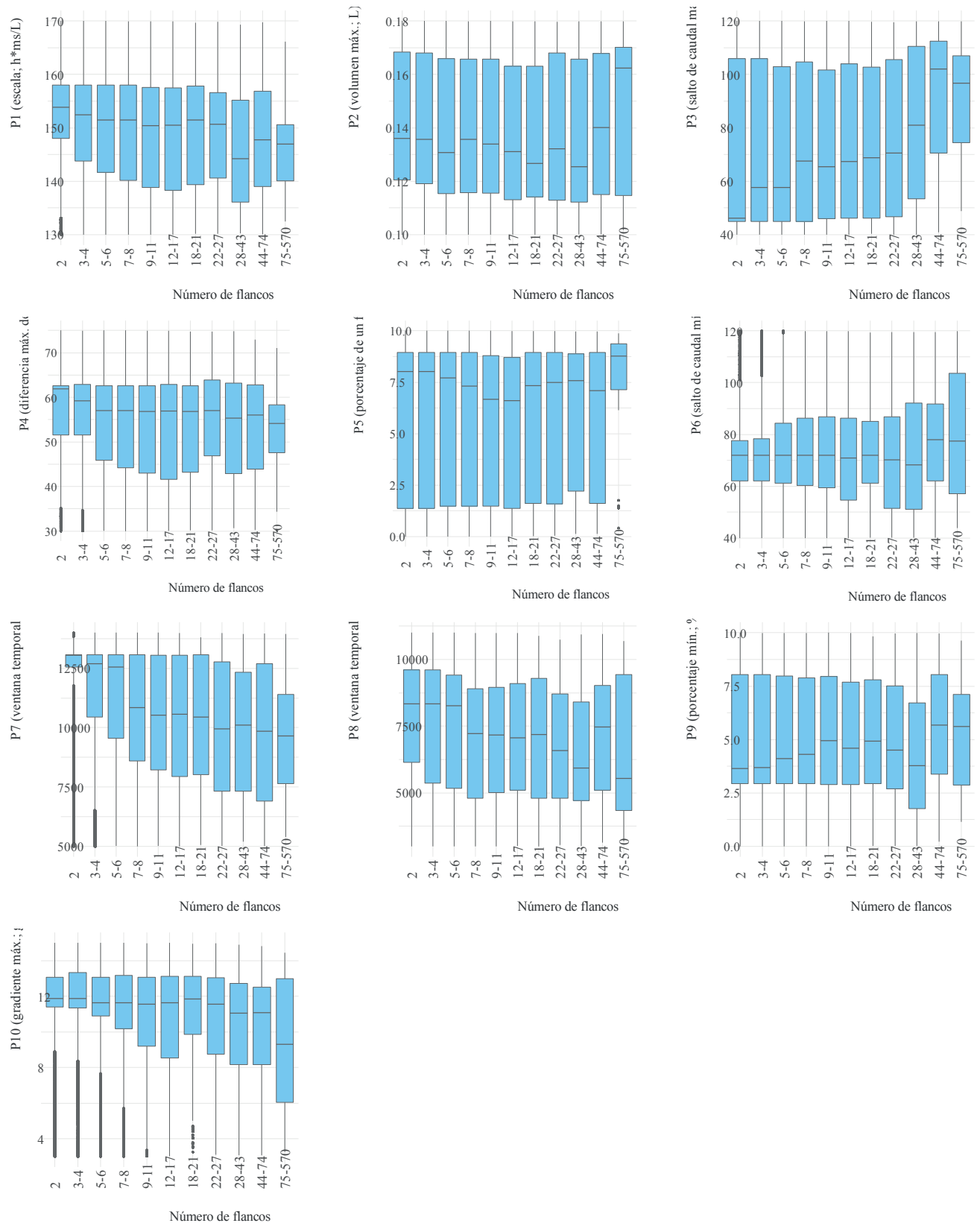
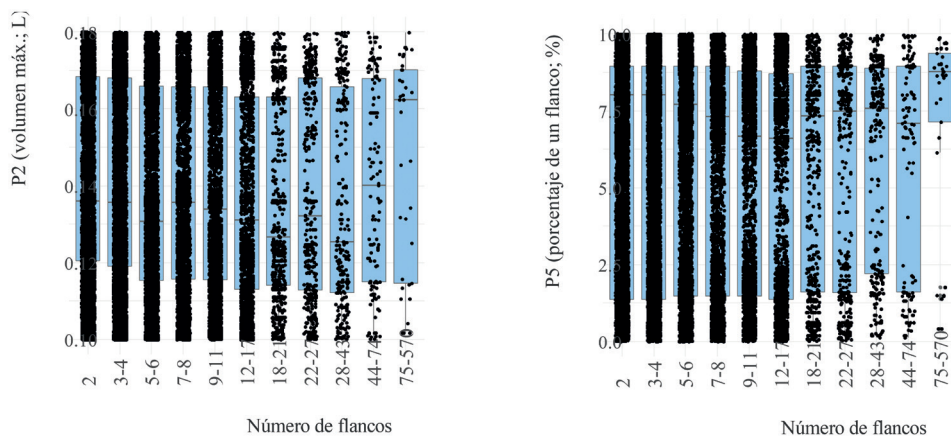


Figura 8 | Distribución de los valores escogidos a través del proceso de calibración para los diez parámetros del filtro por categoría de complejidad.

- Los parámetros  $p3$  y  $p7$  están notablemente correlacionados con el grado de complejidad de los eventos. El parámetro  $p3$  interviene en la supresión del ruido. La magnitud de los saltos de caudal correspondientes al mismo es directamente proporcional al nivel de caudal alcanzado a lo largo de la traza de caudal original, debido a la resolución temporal con la que registran la ocurrencia de un pulso los *loggers* empleados en el caso objeto de estudio. Es decir, en tanto que es más probable que en los eventos complejos tenga lugar el solapamiento de usos, el nivel de caudal alcanzado en estos es mayor y, por consiguiente, también lo será la magnitud de los saltos de caudal que corresponden a ruido. Es por ello que la tendencia general, conforme se incrementa el grado de complejidad, es que el parámetro  $p3$  vaya en aumento. Por su lado,  $p7$  interviene en el cálculo del gradiente. Este parámetro toma para los eventos simples mayores valores que para los eventos complejos. Esto tiene relación con maximizar la simplificación de los eventos de menor grado de complejidad, ya que para bajos valores de  $p7$  se crean secciones horizontales auxiliares (ver Figura 2-b1) que podrían redundar en una simplificación insuficiente de los mismos.
- Se observa ausencia de tendencias en las gráficas que corresponden a los parámetros  $p2$ ,  $p5$ ,  $p8$  y  $p9$ . En el caso de  $p8$  y  $p9$  es esperable, ya que son parámetros creados para el correcto tratamiento de eventos de corta duración con solapamiento de usos (ej. una cisterna y un grifo). Al no ser este tipo de eventos tan frecuentes, estos parámetros tomarán cualquier valor para el resto de los casos. Con respecto a  $p2$  y  $p5$ , que intervienen en la supresión de ruido, se han añadido dos gráficas auxiliares en la Figura 9 donde se muestran la distribución de los casos sobre los gráficos de caja de bigotes. En estas dos gráficas observamos que, al incrementarse el grado de complejidad, los valores adoptados tienden hacia los extremos del rango de variación permitido. Al igual que en el caso de  $p3$ , esto guarda relación con la necesidad de aumentar los valores de los parámetros para eliminar ruido que se produce a un alto nivel de caudal cuando hay solapamientos de usos, pero que toman valores menores en caso contrario.



**Figura 9 |** Distribución de los valores escogidos a través del proceso de calibración para los parámetros  $p2$  y  $p5$  con los casos clasificados en cada categoría de complejidad dibujados sobre los gráficos de caja de bigotes.

- Finalmente, destacar que, para la mayor parte de los casos, los parámetros  $p1$ ,  $p4$ ,  $p6$  y  $p10$  se encuentran en rangos más pequeños que los establecidos. Por tanto, resulta conveniente explorar la posibilidad de reducir el espacio de búsqueda de soluciones en estos cuatro casos.

Las conclusiones extraídas permiten esbozar futuras mejoras en el proceso de filtrado con el objetivo de aumentar su eficiencia. Por ejemplo, podrían establecerse valores por defecto para algunos parámetros en función del grado de complejidad (ej.  $p2$  y  $p5$ ), así como reducir el rango en el que toman valores (ej.  $p3$  y  $p7$ ). Ello permitiría explorar otros métodos de resolución más eficientes del problema de optimización que se ha planteado en el proceso de calibración, como, por ejemplo, los métodos de búsqueda guiados, sin pérdida de la calidad del resultado obtenido.

## CONCLUSIONES

El análisis de los usos finales del agua es una poderosa herramienta que da soporte a la modelación y gestión de la demanda urbana de agua. No obstante, este tipo de estudios requiere una gran inversión en tiempo y capital humano. Esta es la razón por la cual se hace necesario el desarrollo de herramientas para la desagregación automática de las trazas de caudal. Para facilitar esta operación se ha creado un nuevo algoritmo de filtrado. Éste se ha testado a través del análisis de 5210 eventos de consumo, extraídos de unas mediciones realizadas durante octubre de 2016 en un abastecimiento de agua en España. Los resultados muestran que, en media, el algoritmo de filtrado propuesto es capaz de suprimir el 68.6% de los puntos de la traza en bruto, con el ahorro en memoria de almacenamiento que ello supone. Si bien, sus formas generales se respetan, tal y como queda demostrado al haberse alcanzado un valor medio del índice KGE en torno a 0.89. Además, el gradiente a lo largo de la traza de caudal se ve notablemente simplificado. En consecuencia, el uso del algoritmo permite identificar el inicio y final de los eventos individuales solapados con mayor facilidad, así como los cambios de gradiente. Finalmente, el análisis de las combinaciones de parámetros seleccionadas para cada evento por el proceso de calibración automático ha posibilitado la determinación de nuevas estrategias para su mejora futura en lo que se refiere a la reducción del tiempo de ejecución, tanto a través del uso de los valores por defecto para filtrar los eventos pertenecientes a las categorías de menor complejidad sin que el resultado pierda calidad, como de acotar de forma más precisa el espacio de búsqueda de soluciones.

El presente algoritmo de filtrado facilita en gran medida la desagregación automática de los eventos solapados. Por tanto, se potencia el desarrollo de nuevas estrategias basadas en la traza de caudal depurada, mientras que las ya existentes pueden ser utilizadas con mayor provecho. Por otro lado, el resultado obtenido después del filtrado permite reconocer claramente características específicas de algunos dispositivos de consumo de agua. Así pues, las trazas filtradas resultantes pueden ser tenidas en cuenta como dato de entrada en los procesos de categorización de eventos en usos finales del agua, reduciendo los errores de clasificación.

## AGRADECIMIENTOS

El trabajo presentado en este artículo ha sido posible gracias al Proyecto IMPADAPT/CGL2013-48424-C2-1-R del Ministerio de Economía y Competitividad de España con fondos FEDER y al VII Programa Marco de la Unión Europea, bajo el acuerdo de financiación no. 619172 (SmartH2O: an ICT Platform to leverage on Social Computing for the efficient management of Water Consumption).

## REFERENCIAS

- Arregui, F. (2015). New software tool for water End-Uses studies. *Presentation of 8th IWA International Conference on Water Efficiency and Performance Assessment of Water Services*, Cincinnati, USA.
- Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A.E. (2015). Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling & Software*, 72, 198-214, <https://doi.org/10.1016/j.envsoft.2015.07.012>
- DeOreo, W.B., Heaney, J.P., Mayer, P.W. (1996). Flow trace analysis to assess water use. *American Water Works Association*, 88, 79-90. <https://doi.org/10.1002/j.1551-8833.1996.tb06487.x>
- Fielding, K.S., Spinks, A., Russell, S., McCrea, R., Stewart, R.A., Gardner, J. (2013). An experimental test of voluntary strategies to promote urban water demand management. *Journal of Environmental Management*, 114, 343-351. <https://doi.org/10.1016/j.jenvman.2012.10.027>
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377, 80-91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>

- Kowalski, M., Marshallsay, D. (2003). A System for Improved Assessment of Domestic Water Use Components. *II International Conference Efficient Use and Management of Urban Water Supply*, International Water Association, Tenerife, Spain.
- Larson, E., Froehlich, J., Campbell, T., Haggerty, C., Atlas, L., Fogarty, J., Patel, S.N. (2012). Disaggregated water sensing from a single, pressure-based sensor: An extended analysis of HydroSense using staged experiments. *Pervasive and Mobile Computing*, 8, 82-102. <https://doi.org/10.1016/j.pmcj.2010.08.008>
- Nguyen, K.A., Zhang, H., Stewart, R.A. (2013a). Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-environment Research*, 7, 182-201. <https://doi.org/10.1016/j.jher.2013.02.004>
- Nguyen, K.A., Stewart, R.A., Zhang, H. (2013b). An intelligent pattern recognition model to automate the categorisation of residential water end-use events. *Environmental Modelling & Software*, 47, 108-127. <https://doi.org/10.1016/j.envsoft.2013.05.002>
- Pastor-Jabaloyes, L., Arregui, F.J., Cobacho, R. (2018). Water End Use Disaggregation Based on Soft Computing Techniques. *Water*, 10(1), 46. <https://doi.org/10.3390/w10010046>
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponible en: <http://www.R-project.org/>.
- UNEP (United Nations Environment Programme). (2011). Water: Investing in Natural Capital. UNEP, Towards a Green Economy: Pathways to Sustainable Development and Poverty Eradication, Nairobi.