

# ***Aplicación de técnicas de Análisis Inteligente de Datos para el catastro de sistemas de abastecimiento de agua mediante georadar (GPR)***

*David Ayala-Cabrera; Rafael Pérez-García; Joaquín Izquierdo; Manuel Herrera*

*FluIng-IMM, Universitat Politècnica de València (UPV)*

*E-mail: {daaycab; rperez; jizquier; mahefe}@upv.es*

## **1 Resumen**

El desconocimiento de trazados y características de los componentes de sistemas de abastecimiento de agua (SAA), es un factor relevante que desfavorece o encarece la gestión de los mismos. Este trabajo está enfocado al uso de procedimientos de fácil aplicación que permiten visualizar, de manera rápida y por parte de personal no altamente cualificado, las características de los componentes en SAA. Se han empleado metodologías que permiten alcanzar dicho conocimiento sin alterar las condiciones y características del entorno (métodos no destructivos), a través del estudio de bases de datos del subsuelo obtenido con Georadar (ground penetrating radar - GPR). Con el empleo del GPR, se busca minimizar los costes en el mapeo del subsuelo y obtener la mayor cantidad de información del mismo.

En este estudio se realiza la captura de imágenes de GPR de tuberías de cuatro diferentes materiales (PVC, polietileno, fibrocemento y fundición) enterrados en suelo seco y se realiza el análisis a través de la interpretación de la amplitud de onda. Los datos obtenidos de la prospección son pre-procesados con una metodología multi-agente que denominamos *carrera de agentes*. Posteriormente se aplica un proceso de agrupación no supervisado denominado “hierarchical agglomerative clustering”. Los resultados obtenidos de este sistema son trasladados al espacio original y visualizados como imágenes. Tras la aplicación de la metodología, se demuestra la viabilidad de agrupar en un clúster los datos obtenidos en la prospección, disminuyendo la cantidad de información y conservando la fiabilidad de la misma. Los resultados obtenidos son prometedores como metodología no subjetiva y repetible que permite eficientemente la visualización de tuberías enterradas con el objetivo de generar bases y parámetros para el entrenamiento de sistemas inteligentes para la caracterización de componentes de SAA.

## **2 Introducción**

El Georadar (ground penetrating radar - GPR) se ha empleado como metodología no-destructiva para analizar componentes y anomalías de Sistemas de Abastecimiento de Agua (SAA). Los componentes más analizados son las tuberías, siendo aquellas de material metálicos las más frecuentemente evaluadas. En el caso de las fugas de agua, la identificación ha sido un poco más incipiente. La información sobre los componentes, cambios realizados en la red y anomalías son necesarios para el control y gestión de forma productiva en los SAA [1]. La gestión técnica puede enfocarse a metas tales como: catastro de redes, detección de conexiones ilegales, planificación en sistemas de abastecimiento, simulación y operación de redes, instalaciones interiores en redes de abastecimiento, mantenimiento, rehabilitación y renovación de redes de abastecimiento, detección y control de fugas, aplicaciones de Sistemas de Información Geográfica (SIG) para SAA y evolución de los contaminantes en la red. Estudios recientes como el de la Agencia de Protección Ambiental de Estados Unidos (U.S. Environmental Protection Agency - EPA), muestran el uso de herramientas no-destructivas como metodologías que favorecen la gestión técnica en SAA en comparación con otro tipo de metodologías (métodos destructivos) [2]. Sin embargo, aunque la captura de información por métodos no-destructivos es exitosa, el gran volumen de información, el uso y la interpretación requieren de alto nivel de habilidad y experiencia por parte del personal.

En este aspecto, múltiples trabajos han sido desarrollados con vistas a la localización y detección de componentes y anomalías para SAA con GPR. Encontramos trabajos en los cuales se aplican metodologías adaptadas de otros métodos no-destructivos tal como remoción de fondo (background removal) y migración [3], en el cual se buscaba limpiar las imágenes de tuberías metálicas en los datos de GPR obtenidos; trabajos en fugas

en los cuales se emplean las Transformadas de Hilbert y Fourier [4]. También se encuentran trabajos que emplean la Transformada de Hough, en el reconocimiento de patrones de las hipérbolas que se forman en la imagen del radar [5] y en la segmentación y limpieza de las imágenes de tuberías enterradas [6] o trabajos en la mejora de la visualización de tuberías [7]. De igual manera, se encuentran trabajos basados en sistemas inteligentes, para la detección automática de tuberías, en imágenes de GPR. Destaca el uso de redes neuronales [8], que ha dado paso a estudios basados en las máquinas de soporte vectorial [9] y a la aplicación de lógica difusa en el reconocimiento de patrones para el procesamiento de imágenes GPR [10,11] o localización de tuberías plásticas, mediante la aplicación de metodologías multi-agente [12], entre otros. El éxito de la aplicación de estas metodologías depende en gran medida de la limpieza de las imágenes que se obtengan con procesos previos a la clasificación. En la mayoría de ellos, se tiene como meta la detección de hipérbolas o de objetos concretos en la imagen obtenida.

Este trabajo se presenta con el objetivo de generar herramientas para el análisis y simplificación de las bases de datos obtenidas con GPR, buscando facilitar la toma de decisiones en la gestión técnica de SAA, iniciando con el aspecto catastro de redes. El objetivo específico es obtener un reducido número de clústers que capturen las características más relevantes del suelo testeado. Se pretende que los objetos o anomalías presentes en el suelo testeado, sean agrupados de manera natural en los clúster y no que estos objetos u anomalías presentes sean el objetivo de nuestra búsqueda, intentando con ello ampliar el espectro de posibilidades de búsqueda y la rápida interpretación sin el requerimiento de gran habilidad y experiencia por parte del personal para la interpretación.

Este artículo ha sido organizado de la siguiente manera. En esta sección se ha introducido brevemente el trabajo desarrollado y se introduce a las diferentes metodologías de transformación de las señales, aplicadas a localización de tuberías en imágenes de GPR. La siguiente sección presenta el sistema propuesto. Posteriormente, son presentados, los resultados más representativos del planteamiento realizado. Finalmente, una sección de conclusiones cierra el documento.

### 3 Sistema propuesto

La arquitectura del sistema propuesto (ver Figura 1) puede ser dividida en tres interrelacionados procesos: a) pre-proceso, b) hierarchical agglomerative clustering, y c) reconstrucción y visualización. El primer proceso emplea la metodología que el autor ha denominado *carrera de agentes* [13], que usa metodología multi-agente para el análisis de las señales contenidas en los archivos de GPR obtenidos. En esta metodología son generados dos espacios, enmarcados por el comportamiento de los agentes, denominados calentamiento y carrera. Emplearemos este último espacio como input en el proceso siguiente. En el segundo proceso evaluaremos la viabilidad de implementación de un proceso no supervisado que agrupe los datos obtenidos de la prospección de una manera natural y permita una rápida y fiable interpretación de los mismos. El método que probaremos en este documento será *Hierarchical Agglomerative Clustering* (HAC). En el HAC, evaluaremos las más comunes subrutinas y emplearemos la que presente mejor representación en grupos de la variación del suelo. Finalmente, el último proceso consistirá en el traslado de los datos que conforman cada clúster obtenido al espacio original, con el objetivo de visualizar los resultados como imágenes.

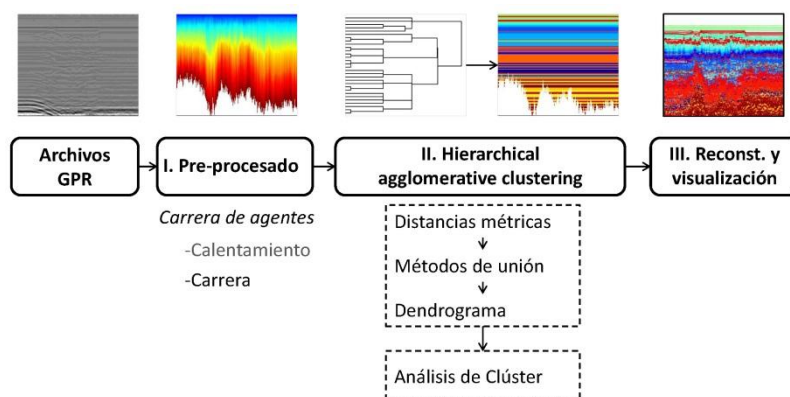


Figura 1 Sistema propuesto

El sistema propuesto será determinado de manera iterativa, variando las distancias métricas y los métodos de unión del proceso II. El proceso III será aplicado para cada iteración. Las imágenes obtenidas tras la aplicación de estos dos procesos será el punto de referencia para determinar las subrutinas a proponer para el proceso II. Los procesos que componen este sistema y la selección de los mismos serán presentados más detalladamente a continuación.

### 3.1 Pre-procesado

El algoritmo *carrera de agentes* está basado en la teoría de juegos y en él se usa el paradigma multi-agente [14]. El input será el radargrama resultante de la prospección GPR, que consiste en una matriz de tamaño  $m \times n$ . Las  $n$ -trazas generadas en la prospección con GPR son empleadas en este trabajo como pistas paralelas que recorrerán los  $n$ -agentes. La carrera será una prueba de resistencia entre los agentes comprendida por dos fases: a). *calentamiento* y b). *competencia*. El proceso estará determinado por un tiempo total  $t = t_w + t_r = m$ , siendo  $t_w$  el tiempo de calentamiento y  $t_r$  el tiempo de competición. Los desplazamientos de los agentes en  $t_r$  estarán condicionados por el cambio de tendencia de la traza que recorren. La carrera terminará cuando se cumpla el tiempo  $t$  y el agente ganador será el que tenga mayor desplazamiento en este tiempo.

El output de este proceso será una matriz de tamaño  $ml \times n$ , donde  $ml$  = número máximo de desplazamientos. Las columnas de esta matriz, describen el movimiento de los agentes en relación a la competición. En este documento se empleará el espacio de competición, como input para el proceso de Hierarchical agglomerative clustering (ver, Figura 1 - proceso I).

### 3.2 Hierarchical Agglomerative Clustering

Hierarchical agglomerative clustering (HAC) es uno de los procesos más comúnmente empleados para resumir estructuras de datos. Un árbol jerárquico es un conjunto jerarquizado de particiones representado por un diagrama de árbol o dendograma. El seccionado de un árbol en un nivel particular produce una partición en grupos disjuntos. Si son creados dos grupos en las diferentes particiones (los resultados de la división en diferentes niveles), entonces los grupos son disjuntos o un grupo contiene al otro totalmente [15]. Este proceso de agrupamiento jerárquico puede ser visualizado en forma de dendograma, donde se ilustra cada paso en el proceso de agrupamiento con una combinación en el árbol (ver figura 1 - Proceso II.). El proceso se divide en dos fases: 1) construcción del dendograma, y 2) análisis de clúster. En la primera fase, se construye el dendograma. En esta fase se tienen dos aspectos a evaluar que son: a) distancias métricas, y b) método de unión. Estos aspectos pueden ser relacionados mediante la correlación *cofenética*, con la cual se obtiene que tan cercanas son las distancias de entrada con las de salida del sistema planteado. La segunda fase corta el dendograma obtenido de acuerdo al criterio planteado. El criterio empleado en este documento será un corte buscando una división natural de los datos. Para esto, se empleará el coeficiente de *inconsistencia*. Las componentes de las fases serán explicadas más detalladamente a continuación.

#### 3.2.1 Distancias métricas

En el sistema propuesto empleamos tres tipos de distancias métricas, que son comúnmente usadas en HAC. Tomando la matriz de datos  $X$  de tamaño  $ml \times n$ , resultante del pre-procesado y tratándola como  $ml(1 \times n)$ , siendo el vector columna determinado como  $x_1, x_2, \dots, x_{ml}$ , las diferentes distancias entre los vectores  $x_s$  y  $x_t$ , donde estos últimos son los vectores a comparar, se definen las distancias métricas como:

1). Distancia Euclídea:

$$D_{Eu}^2(x_s, x_t) = (x_s - x_t)(x_s - x_t)'$$

Ecuación (1)

2). Distancia Seudoeuclídea:

$$D_{Seu}^2(x_s, x_t) = (x_s - x_t)V^{-1}(x_s - x_t)',$$

Ecuación (2)

donde  $V$  es una matriz diagonal  $n \times n$  cuyo  $j$ -ésimo elemento diagonal es  $S(j)^2$ , siendo  $S$  el vector de desviación estándar.

3). Distancia Coseno:

$$D_{Co}^2(x_s, x_t) = 1 - \frac{x_s x_t'}{\sqrt{(x_s x_s')(x_t x_t')}} .$$

Ecuación (3)

### 3.2.2 Métodos de unión

Una vez que la proximidad entre objetos del conjunto ha sido calculada, se puede determinar cómo los objetos en el conjunto de datos se deben agrupar en clúster, utilizando los métodos de unión. Los métodos de unión toman la información de la distancia generada en el método de distancia métrica aplicado y enlaza pares de objetos que están cercanos entre sí en clúster binarios (el clúster se compone de dos objetos). Los métodos de unión continúan su proceso agrupando los clúster obtenidos con nuevos clúster cercanos hasta crear un gran clúster en el que todos los objetos en el conjunto de datos original son unidos en un árbol jerárquico. Existen muchas posibilidades para seleccionar los valores de similitud. Entre los métodos de unión más comunes se tienen: 1) método del mínimo, 2) método de la media, y 3) método del máximo [16]. La siguiente notación se utiliza para describir los vínculos utilizados por los métodos:

1). Método del mínimo (single linkage):

$$Z(r, s) = \min \left( \text{dist}(x_{ri}, x_{sj}) \right), i \in (1, \dots, n_r), j \in (1, \dots, n_s) .$$

Ecuación (4)

2). Método de la media (average linkage):

$$Z(r, s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} \text{dist}(x_{ri}, x_{sj}) .$$

Ecuación (5)

3). Método del máximo (complete linkage):

$$Z(r, s) = \max \left( \text{dist}(x_{ri}, x_{sj}) \right), i \in (1, \dots, n_r), j \in (1, \dots, n_s) .$$

Ecuación (6)

Para las ecuaciones 4, 5 y 6 se tiene que:

$r$  y  $s$  son dos clúster;

$n_r$  y  $n_s$  es el número de objetos en el clúster  $r$  y  $s$ , respectivamente;

$x_{ri}$  es el  $i$ -ésimo objeto en el clúster  $r$ ;

$x_{sj}$  es el  $j$ -ésimo objeto en el clúster  $s$ .

### 3.2.3 Coeficiente de correlación cofenética

Después de vincular los objetos de un conjunto de datos en un árbol de agrupamiento jerárquico, pueden ser comprobadas las distancias (es decir, las alturas) en el árbol de reflejar la distancia original con exactitud. La correlación cofenética de un árbol de clúster se define como el coeficiente de correlación lineal entre la distancia cofenética obtenida del árbol y las distancias originales (o diferencias) que se utilizan para construir el árbol. Por lo tanto, es una medida de la fidelidad del árbol que representa las diferencias entre las observaciones [17]. El valor de salida,  $c$ , es el coeficiente de correlación cofenética. La magnitud de este valor debe ser muy cercana a 1 para que la solución sea de alta calidad. Esta medida puede ser usada para comparar alternativas de solución de clúster obtenidos mediante diferentes algoritmos. La correlación cofenética se define por:

$$c = \frac{\sum_{i < j} (Y_{ij} - y)(Z_{ij} - z)}{\sqrt{\sum_{i < j} (Y_{ij} - y)^2 \sum_{i < j} (Z_{ij} - z)^2}},$$

Ecuación (7)

donde

$Y_{ij}$  es la distancia entre los objetos  $i$  y  $j$  en  $Y$ , la cual es la matriz de distancia métricas;

$Z_{ij}$  es la distancia cofenética entre los objetos  $i$  y  $j$  para  $Z$ , siendo  $Z$  el output del método de unión empleado;

$y$  y  $z$  son la media de  $Y$  y  $Z$ , respectivamente.

### 3.2.4 Análisis de clúster

El árbol de conglomerados jerárquicos, naturalmente, puede dividir los datos en clústers distintos y bien separados. Esto puede ser particularmente evidente en un diagrama dendrograma creado a partir de los datos en grupos de objetos que están densamente empaquetados en ciertas áreas y no en otras. El coeficiente de inconsistencia de los enlaces en el árbol de grupo puede identificar estas divisiones en las similitudes entre los objetos que cambian abruptamente. Podemos utilizar este valor para determinar que la función del clúster que crea sus límites [17]. De esta manera se define una etiqueta para cada vínculo del dendrograma. Esta etiqueta muestra la cantidad de pares de clúster similares. Con esta medida, que puede unirse a grupos, si el valor de la inconsistencia es menor que cierto umbral determinado. El coeficiente de inconsistencia que caracteriza cada vínculo un árbol de clúster mediante la comparación de su longitud con la longitud media de otros enlaces en el mismo nivel del dendrograma. Cuanto mayor sea el valor de este coeficiente, menor similitud entre los clústeres conectados por el enlace. Para cada enlace,  $k$ , el coeficiente de inconsistencia se calcula como:

$$IC(k) = (L(k) - W1(k))/W2(k)$$

Ecuación (8)

donde

$IC$  y  $L$  son vectores  $(m - 1) \times 1$ ;

$IC$ : coeficiente de inconsistencia;

$W1$ : Distancia de todos los vínculos incluidos en el cálculo;

$W2$ : Desviación estándar de todos los vinculo incluidos en el cálculo;

$L$ : Longitudes de los vínculos.

Para los nodos hoja, los nodos que no tienen más nodos debajo de ellos, el coeficiente de inconsistencia es 0. El umbral utilizado en este trabajo fue 1.

## 4 Estudios experimentales

Esta sección se presenta la aplicación del sistema propuesto para la generación natural de clúster, descrito en la sección anterior. El caso de estudio corresponde a la captura en imágenes de GPR de cinco tuberías de materiales diferentes usualmente empleadas en SAA. Los materiales de tubería empleados en los ensayos fueron PVC, polietileno (PE), fibrocemento (Fib) y hierro fundido (Fund). Las características de las tuberías empleadas se presentan en la Tabla 1. Las imágenes de GPR se obtuvieron enterrando las tuberías en suelo seco en un tanque de laboratorio (ver, figura 2). La prospección fue realizada enterrando cada tubería por separado y realizando una prueba adicional de referencia (sin tubería).

Tabla 1 Características de las tuberías enterradas

Material	Abreviatura	Diámetro Interno [mm]	Diámetro Externo [mm]
PVC	PVC	100	110
Polietileno	PE1	35	50
Polietileno	PE2	76	90
Fibroceso	Fib	80	96
Fundición	Fund	86	98

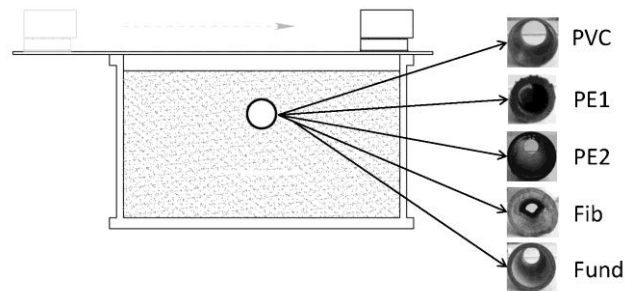


Figura 2 Configuración de captura de imágenes con GPR

El equipo GPR empleado en las prospecciones corresponde a una antena comercial monoestática, con una unidad de control GSSI SIR 3000 y una frecuencia central de la antena de 1,5 GHz. La captura de las imágenes de GPR, consistió en el deslizamiento de la antena sobre una superficie plana, bajo las configuraciones establecidas y una adicional sin tubería (Blanco). Las mediciones se realizaron de manera transversal a la tubería. Los parámetros del equipo fueron constantes en la totalidad de los muestreos, con el fin de observar las respuestas del medio a las variaciones de configuración planteadas al mantener constantes los parámetros de captura. A manera de imágenes en escala de gris, en la Figura 3, se presentan los resultados obtenidos de las pruebas realizadas sobre las configuraciones planteadas.

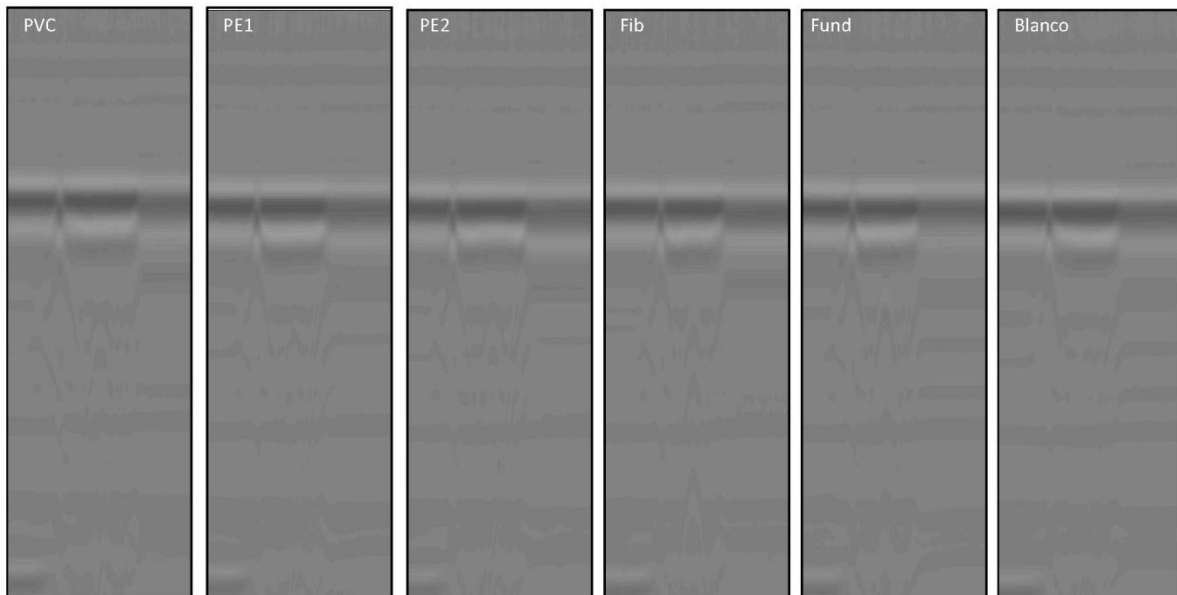


Figura 3 Imágenes en bruto obtenidas con GPR

En las imágenes de los datos en bruto mostrados en la Figura 3, se puede notar en las tuberías de PVC y PE que el contraste de las fronteras no es lo suficientemente fuerte para ser observado de manera inmediata, lo que es el

resultado de la baja permitividad, característica de estos materiales. Esto causa una baja intensidad de color que no permite una demarcación fuerte en las tuberías de PVC y PE (casi invisibles). Por otra parte, las tuberías Fib y Fund, aunque su contraste es mayor, presentan dificultad de visualización. En todas las imágenes presentadas, aunque no es imposible encontrar la anomalía, la dificultad de interpretación de los datos es alta y requiere de una gran pericia y experiencia por parte del personal, siendo muy difícil diferenciar en si hay tubería o no. Esta dificultad de interpretación ratifica la necesidad de establecer metodologías que faciliten la misma. Los datos de estas imágenes serán empleados en el sistema propuesto en este documento, para ello emplearemos las 3 distancias métricas y los 3 métodos de unión planteados anteriormente. Los seis ensayos realizados en el laboratorio, serán procesados y evaluados mediante la correlación cofenética en las nueve combinaciones de subrutinas del sistema propuesto (ver, Tabla 2).

Tabla 2 Combinaciones de distancias métricas y métodos de unión propuestos en el sistema

	Mínimo	Media	Máximo
Euclídea	SEu	AEu	CEu
Seudoeuclídea	SSeu	ASeu	CSeu
Coseno	SCo	ACo	CCo

Al evaluar las subrutinas planteadas en el sistema propuesto para cada una de las configuraciones de los ensayos propuestos, se obtuvo el coeficiente de correlación cofenética para cada configuración. Las correlaciones cofenéticas obtenidas, se presenta en la Figura 4.

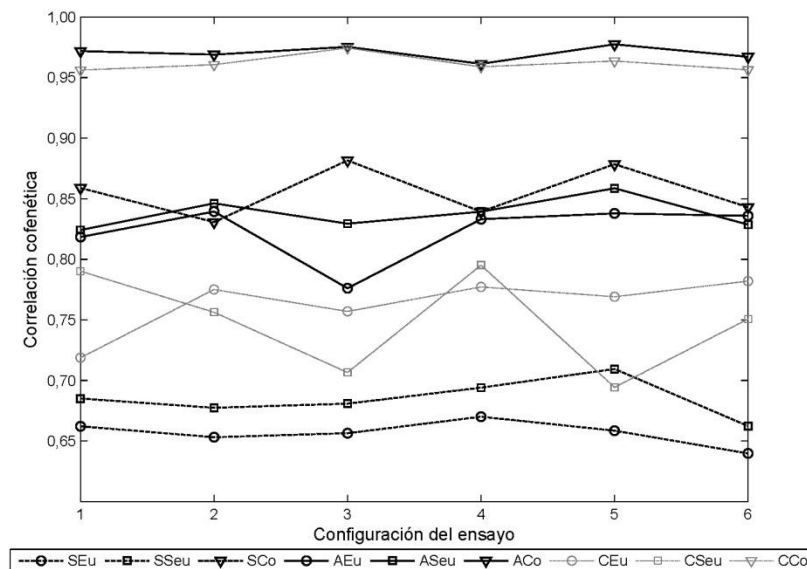


Figura 4 Coeficientes de correlación cofenética

En la Figura 4, se puede observar que el uso de la distancia métrica coseno incrementa el coeficiente de correlación cofenética en las diferentes configuraciones de ensayos planteados. Podemos observar de igual manera que las combinaciones ACo y CCo presentan los valores más altos. Siendo ACo la que presenta en las condiciones del suelo planteadas los coeficientes más altos y constantes de las subrutinas planteadas. De lo cual se podía inferir que la combinación de coseno en el cálculo de las distancias y Average en el método de unión, es adecuada para imágenes de GPR, en la obtención de clúster, mediante HAC. En la Figura 5 se presentan las imágenes de los clúster obtenidos (en su espacio original) de la combinación ACo para los datos de los ensayos obtenidos.

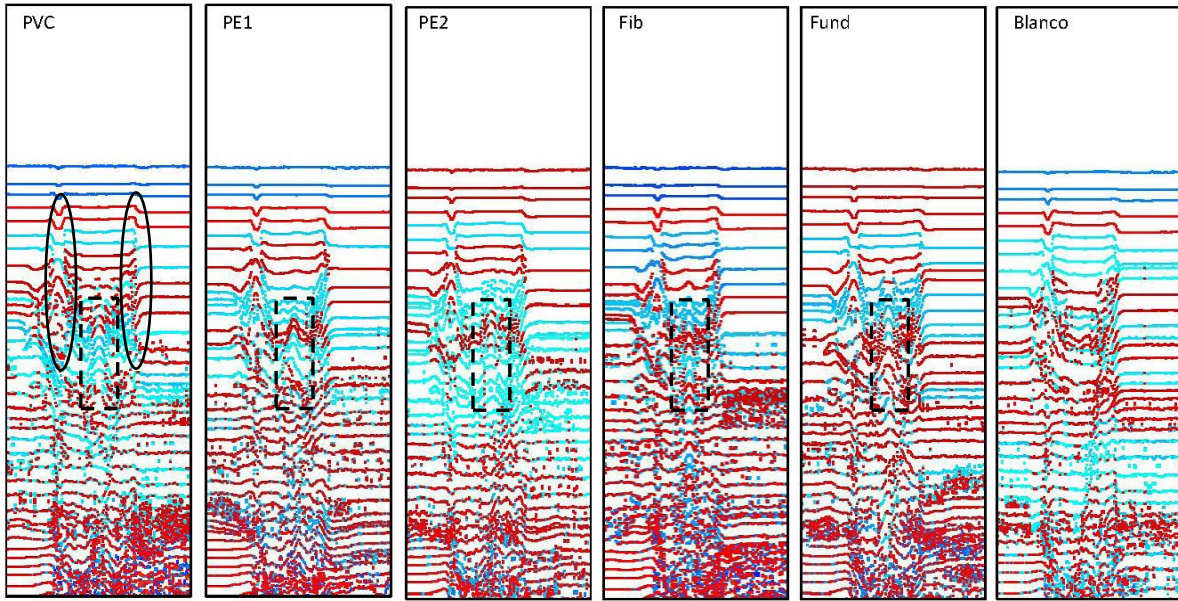


Figura 5 Imágenes resultantes del procesamiento

En las imágenes procesadas, con el sistema propuesto (Figura 5), se observa una considerable mejora en las formas y características de las imágenes, al compararlas con sus respectivas imágenes en bruto (Figura 3). En ambas imágenes las hipérbolas que determinan las diferentes tuberías adicionales son demarcadas. Por otra parte, en las imágenes de cada ensayo se observa la demarcación de las paredes del tanque que contienen el suelo seco y como las imágenes de las tuberías están demarcadas en la parte central, acorde la configuración planteada en la Figura 2. Al tener en cuenta que los clúster representados en la figura anterior, son como capas que pueden removerse, en la Figura 6, son presentadas las capas que aparecen demarcadas con la adición de las tuberías en los ensayos.

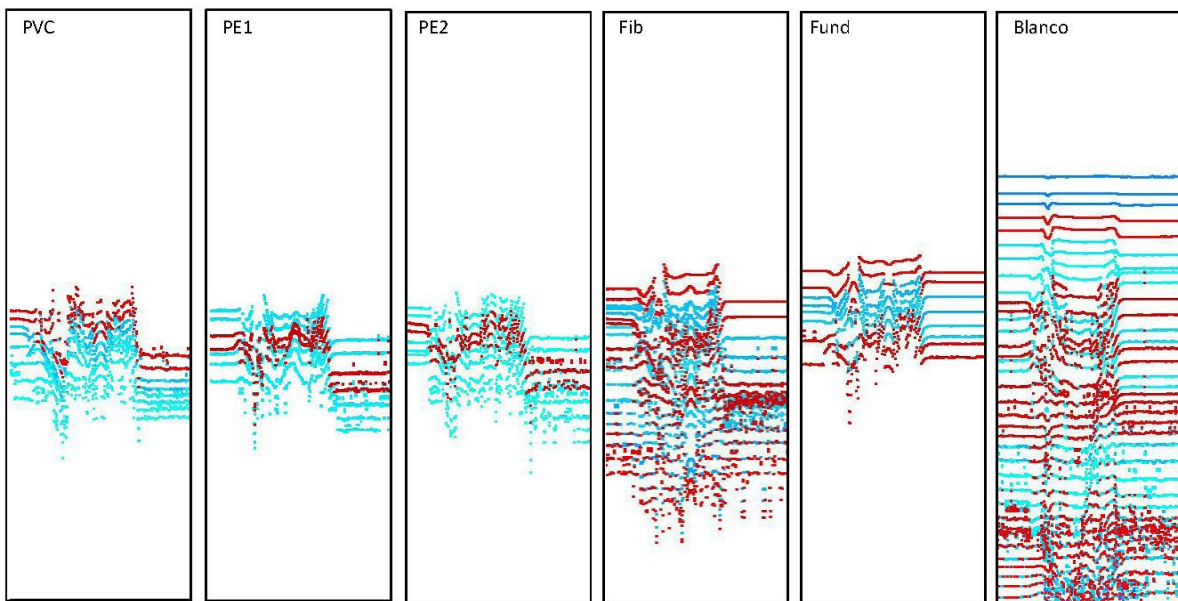


Figura 6 Imágenes resultante del procesamiento con la eliminación de capas sin tuberías

En Figura 6, se puede observar y contrastar con el blanco la marcación de las diferentes tuberías adicionales, tras el paso de los datos en bruto por el sistema propuesto. En estas imágenes, se visualiza con mayor claridad y menos ruido las tuberías de PVC y PE. En las imágenes posteriores al procesamiento de las tuberías, se puede observar formaciones de hipérbolas sólidas y definidas. El paso de las matrices en bruto por el



sistema propuesto, también aumenta la visualización de estas tuberías y disminuye el ruido de sus imágenes correspondientes. Prueba de ello es una mayor fuerza en la configuración de sus características, como por ejemplo las asíntotas de la hipérbola de la tubería en la imagen.

Tras el paso de los datos en bruto, resultado de las prospecciones por GPR realizadas en laboratorio, de diferentes materiales de tubería por separado y su posterior procesamiento por las metodologías planteada en este documento, se puede decir que:

**PVC.** Presenta mejoramiento en la visualización de la imagen y una mayor cantidad de las características propias del paso de la señal a través de la misma, al compararlo con los datos en bruto.

**PE.** Permite la visualización de este tipo de material, presentando una mejor demarcación de la misma al compararlos con los datos en bruto. Observamos también para este tipo de material que la mejora de la visualización se consigue aún en tuberías de diámetro tan pequeños como el PE1.

**Fib. y Fund.** Mejora la visualización de las características propias de la imagen, tal como las asíntotas y presenta mayor solidez en la imagen de la tubería. En la imagen de la tubería de Fibrocemento, se observan mayor cantidad de hipérbolas en comparación con las demás imágenes de tuberías adicionales. Esta cantidad de hipérbolas adicionales pueden ser el resultado de la incrustación que presenta la tubería al interior (Figura 2), lo cual es un resultado prometedor al plantear la posibilidad de inspeccionar al interior de las tuberías con GPR.

Si bien todo el procedimiento empleado se realiza de manera automática, la eliminación de capas se realizó de manera manual. Sin embargo, se debe tener en cuenta que este proceso de eliminación podrá realizarse de manera automática con la debida aplicación de procesos de reconocimiento de patrones.

El sistema propuesto, fue probado en 1140 imágenes, en total, de tuberías de los diferentes materiales (PVC, PE, Fib, Fund) y referencias prospectadas con GPR. Con la aplicación de este sistema se obtuvo en más del 95% de los casos, mejora considerable en la limpieza de las imágenes, conservándose en ellas sus propiedades y reduciéndose la cantidad de puntos que componen la imagen. Por otra parte, la disminución de puntos que componen la imagen es considerablemente superior a los puntos de la imagen en bruto (60% menos datos que la imagen en bruto). Lo cual representa un gran avance para posteriores clasificaciones por metodologías inteligentes, al manejarse menor cantidad de información, conservando las características principales de la imagen y obtenerse perfiles del terreno más limpios. Estos perfiles facilitarían la interpolación de imágenes en 3D y posteriores mapas de isolíneas.

## 5 Conclusiones

El sistema propuesto demuestra la viabilidad de agrupar en clúster los datos obtenidos en la prospección, disminuyendo la cantidad de información y conservando la fiabilidad de la misma, caracterizando el suelo de una manera natural y rápida acorde su composición, favoreciendo la interpretación de los archivos de GPR por parte de personal no altamente calificado y sin la necesidad de asumir parámetros del suelo. Los resultados de este sistema son prometedores, al agrupar y disminuir la cantidad de puntos que componen el radargrama inicial conservando sus propiedades principales y favoreciendo visualizaciones más claras de tuberías de sistemas de abastecimiento de agua con mejor identificación de los componentes en las mismas. La herramienta propuesta en este documento, sumada con el GPR como herramienta no-destruktiva, favorece la creación de perfiles del terreno que permitirán la creación de procedimientos automáticos que favorezcan la gestión técnica de los sistemas de abastecimiento.

En futuros trabajos se comparará el método de generación de clúster planteado en este documento con métodos adaptados de *K-means* o *support vector clustering*, evaluando la efectividad y fiabilidad de los mismos, en clasificaciones naturales del suelo. El algoritmo seleccionado será incorporado a sistemas más complejos de patrones de reconocimiento de datos con la finalidad de generar herramientas automáticas, para la gestión técnica de sistemas de abastecimiento mediante métodos no-destruktivos como el GPR.

## 6 Agradecimientos

Este trabajo ha sido desarrollado con el apoyo del proyecto IDAWAS, (Ref.DPI 2009-11591), financiado por el Ministerio de Ciencia e Innovación de España y proyecto complementario ACOMP/2011/188, financiado por la Consellería de Educación de la Generalitat Valenciana.

## 7 Referencias Bibliográficas

- [1]. Bimpas, M., Amditis, A., Uzunoglu, N.K. (2011). Design and implementation of an integrated high resolution imaging ground penetrating radar for water pipe rehabilitation. *Journal of Water Resources Management*, Springer, pp. Volume 25, No. 4, pp. 1239-1250.
- [2]. Thomson, J., Wang, L., Royer, M. (2009). State of Technology Review Report on Condition Assessment of Ferrous Water Transmission and Distribution Systems. Report EPA/600/R-09/055. EPA United States Environmental Protection Agency.
- [3]. Olhoeft, G.R. (1999). Maximizing the information return from Ground penetrating radar. *Journal of Applied Geophysics*, vol. 43, pp. 175-187.
- [4]. Tavera, M. (2008). Aplicación de georadar para la mejora del rendimiento de una red hídrica. Master Thesis. Polytechnic University of Valencia, DIHMA. Valencia (Spain).
- [5]. Simi, A., Bracciali, S. Manacorda, G. (2008). Hough Transform Based Automatic Pipe Detection for Array GPR: algorithm development and on-site test. In: Radar Conference, 2008. RADAR '08. IEEE.
- [6]. Ayala-Cabrera, D., Pérez-García, R., Herrera, M., Izquierdo, J. (2011). Segmentación y limpiado de imágenes de GPR de tuberías enterradas. In: IX SEREA, Seminario Iberoamericano de planificación, proyecto y operación de abastecimiento de agua. Morelia (México).
- [7]. Ayala-Cabrera, D., Herrera, M., Montalvo, I., Pérez-García, R. (2011). Towards the visualization of water supply system components with GPR images. *Mathematical and Computer Modelling*, In press. doi:10.1016/j.mcm.2010.12.045
- [8]. Al-Nuaimy, W., Huang, Y., Nakhkash, M., Fang, M.T.C., Nguyen, V.T., Eriksen, A. (2000). Automatic detection of buried utilities and solid objects with GPR using neural networks and pattern recognition. *Journal of Applied Geophysics*, vol. 43, pp. 157-165.
- [9]. Pasolli, E., Melgani, F., Donelli, M. (2008). Automatic detection and classification of buried objects in GPR images using genetic algorithms and support vector machines. In: International Geoscience and Remote Sensing Symposium (IGARSS), pp. 525-528.
- [10]. Ciu, Y., Wang, L., Xiao, J. (2010). Automatic feature recognition for GPR image processing. *Journal World Academy of Science, Engineering and Technology* 61, pp. 176-179.
- [11]. Delbo, S., Gamba, P., Roccató, D. (2002). A Fuzzy Shell Clustering Approach to Recognize Hyperbolic Signatures in Subsurface Radar Images. *Journal IEEE Transactions on Geoscience and Remote Sensing*, Vol. 38, No. 3, pp. 1447-1451.
- [12]. Ayala-Cabrera, D., Herrera, M., Izquierdo, J., Pérez-García, R. (2011). Location of Buried Plastic Pipes Using Multi-Agent Support Based on GPR Images. *Journal of Applied Geophysics*, under review.
- [13]. Ayala-Cabrera, D., Izquierdo, J., Montalvo, I., Pérez-García, R. (2011). Water Supply System Component Evaluation from GPR Radargrams using a Multi-agent Approach. In: International Conference on Mathematical Modelling in Engineering & Human Behaviour. Accepted, Valencia (Spain).
- [14]. Shoham, J., Leyton-Brown, K. (2009). Multiagent systems: algorithmic, game-theoretic and logical foundations. Cambridge University. In Press.
- [15]. Ghasemi, M., Sadoghi, H.S., Monsefi, R. (2010). A New Hierarchical Clustering Algorithm on Fuzzy Data (FHCA). *International Journal of Computer and Electrical Engineering*, Vol. 2, No.1, pp. 134-140.
- [16]. Lampropoulos, A.S., Tsihrintzis, G.A. (2004). Agglomerative Hierarchical Clustering for Musical Database Visualization and Browsing. In: 3rd Hellenic Conference on Artificial Intelligence, pp. 5-8, Samos Island.
- [17]. MathWorks, <http://www.mathworks.com/>