

Herramientas para la implementación de mantenimiento proactivo en alcantarillados urbanos utilizando confiabilidad de inundación y conceptos de entropía de información

Tools for the implementation of proactive maintenance in urban sewer systems by the use of flooding reliability and entropy of information concepts

*Santiago Sandoval*¹, Andrés Torres¹, Nelson Obregón¹*

¹Facultad de Ingeniería. Pontificia Universidad Javeriana. Carrera 7 N.º. 40 - 62, Bogotá. Colombia

(Recibido el 22 de marzo de 2012. Aceptado el 6 de noviembre de 2012)

Resumen

Las bases de datos de alcantarillados urbanos existentes en ciudades de países en vía de desarrollo presentan deficiencias en los registros de las características físicas de las tuberías, lo cual limita su utilización como soporte para la implementación de herramientas de gestión de dicha infraestructura. El objetivo de este trabajo consistió en proponer y emplear en 13 subcuencas pertenecientes a la cuenca El Salitre de Bogotá, comprendiendo un área de 4515 Ha, con un total de 12842 tuberías pluviales y combinadas, árboles de decisión y entropía de información para identificar tuberías con demanda de actividades de mantenimiento en la red como limpieza, rehabilitación, reemplazo, *etc.*, a partir de confiabilidades de inundación estimadas (para 2337 tuberías con cuyas características físicas se encontraban registradas de manera precisa). El modelo clasificatorio presentó una capacidad predictiva promedio de 62 %. Las tuberías fueron clasificadas mediante 28 reglas, según los valores de confiabilidad, identificando 3383 tuberías (26 % sobre el total) con una priorización alta de mantenimiento proactivo.

----- *Palabras clave:* Gestión de alcantarillados urbanos, árboles de decisión, confiabilidad de inundación, mantenimiento proactivo, entropía de información.

* Autor de correspondencia: telefax: 011 + 57 1 320 83 20, ext. 5259, correo electrónico: sandoval.s@javeriana.edu.co. (S. Sandoval)

Abstract

Existent databases of urban sewerage networks in developing countries have shown imprecisions in data related to physical characteristics of sewer pipes. This fact limits the appropriateness of databases for being used as a support for asset management tools for sewerage. This paper is aimed at proposing the use of decision trees and information entropy for prioritizing sewer pipes by their cleaning, rehabilitation, replacement activities demand, based on 13 sub basins belonging to El Salitre basin of Bogotá, with a total area of 4515 Ha and 12842 pluvial and combined sewer pipes. The starting point was the flooding reliability estimate of 2337 sewer pipes with well specified physical characteristics. The classificatory model showed a capacity for prediction of roughly 62 %. Sewer pipes were classified by using 28 rules defined as a function of the flooding reliability. As a result, 3383 sewer pipes (26 % of the total) were found to have a high-prioritization proactive maintenance demand.

----- *Keywords:* Urban sewerage management, decision trees, flooding reliability, preventive maintenance, information entropy

Introducción

[1] reportan herramientas de soporte para la toma de decisiones con diferentes niveles de alcance en la gestión: (i) Herramientas de análisis de desempeño, las cuales se basan en pronósticos de desempeño de la red, utilizando métodos de estadística como Cadenas de Markov [5, 16]; (ii) Herramientas de análisis de decisión [2], las cuales integran las fases de recopilación de información [26], modelación de desempeño [6, 11] y estrategias de gestión [22]; (iii) Herramientas para la gestión integral [23], las cuales involucran todas las fases de gestión, teniendo en cuenta tecnologías apropiadas, su implementación y monitoreo. Aunque el primer tipo de herramientas, en efecto, puede representar una base para la toma de decisiones objetiva e informada en gestión patrimonial, al brindar pronósticos del estado de un sistema de alcantarillado con base en estimaciones de estados anteriores (lo cual constituye una ventaja en términos de la antigüedad de las inspecciones requeridas para su implementación), éste se basa en información inexistente o escasamente reportada en las bases de datos de alcantarillados en nuestro medio (e.g. inspecciones directas, tipo de suelo, exposición a daños). En cuanto a los

modelos mencionados propuestos para niveles más avanzados de gestión (herramientas (ii) y (iii)), los cuales parecen ser más robustos, al tener en cuenta variables operativas e incluso de costos, éstos presentan las mismas desventajas que las indicadas anteriormente en cuanto a la información disponible, siendo aún más crítica la integración de bases de datos, ya que no es necesario únicamente información relacionada con la tubería (incluyendo historial de fallas) sino también datos del entorno (e.g. tipo de suelo y tipo de vía o tráfico) y datos socio-económicos (e.g. presupuesto y costos de rehabilitación).

Por otra parte, se han identificado como escasas las experiencias en cuanto a la formulación de estrategias proactivas de gestión de alcantarillados mediante indicadores de operación en la red que puedan reflejar fenómenos locales de inundación en las ciudades por un funcionamiento inadecuado de las tuberías, específicamente desde el concepto de confiabilidad de inundación. Este tipo de indicadores puede permitir la planeación de actividades proactivas de limpieza en la red, que a su vez, representan soluciones de bajo costo a los problemas de inundación en las ciudades [8].

En el ámbito internacional, diferentes autores han puesto en evidencia la relación entre el

deterioro del nivel de servicio del alcantarillado y la probabilidad de inundación de un casco urbano [3, 8, 30] desarrollaron una metodología para medir, con base en registros históricos de inundaciones, el nivel de servicio brindado por la empresa gestora del alcantarillado en una cuenca urbana en términos de protección contra el riesgo de inundación pluvial, proponiendo estrategias de gestión. Este enfoque puede en cierto sentido omitir las causas físicas particulares de cada inundación, al no incorporar en el análisis variables hidrológicas detalladas. Por otro lado, metodologías basadas en registros históricos de reportes de fallas necesitan de información generalmente inexistente en cuencas urbanas en países en vía de desarrollo [28]. Adicionalmente, la implementación de modelos matemáticos que permitan representar comportamientos altamente no-lineales (*e.g.* árboles de decisión, redes neuronales artificiales, métodos *Kernel*, *SVM*, *Machine Learning*) se ha detectado como escasa en investigaciones relacionadas con la gestión patrimonial de alcantarillados [4, 18, 31] y más aún en el contexto de la utilización de indicadores de desempeño como herramientas para la implementación de mantenimiento proactivo (*e.g.* confiabilidad de inundación).

A pesar de que, como se mostró anteriormente, en el ámbito internacional existe una diversidad de modelos como soporte para realizar mantenimiento proactivo de las redes de alcantarillado, y que a su vez, algunos de éstos utilizan específicamente indicadores de desempeño hidrológico e hidráulico [7, 27], la mayoría se basan en información completa y oportuna para ciudades (*e.g.* por medio de diferentes métodos de inspección: [15]). Debido a lo anterior, la utilización de dichos modelos en nuestro contexto resulta limitada en términos de: (i) la escasez y baja calidad de información, baja cobertura de inspecciones en la red, *etc.* y (ii) la capacidad de los modelos descritos para representar comportamientos altamente no-lineales, que permitan incorporar la variabilidad y complejidad en sistemas de drenaje urbano de ciudades en crecimiento constante con una

influencia directa, en muchos casos desapercibida, sobre el sistema de alcantarillado [20]. Por lo tanto es necesario continuar investigando en este tema y generar herramientas propias, adaptadas a las necesidades, calidad y cantidad de información disponible.

Los primeros acercamientos en nuestro país hacia la búsqueda de este tipo de herramientas han sido abordados acudiendo a modelos matemáticos y probabilísticos que pretenden representar qué tan confiable es el funcionamiento del sistema bajo varios enfoques [19, 29]. [28], siguiendo el lineamiento de [29], realizaron un estimativo de la confiabilidad de inundación aplicado a una zona específica de Bogotá para más de 12000 tuberías de alcantarillado pluvial y combinado, apoyándose en herramientas computacionales (*R Development Core Team*, 2011) y en métodos de generación aleatoria (simulación de Montecarlo). Este trabajo tuvo como limitante el no poder reportar resultados de confiabilidad de inundación para la mayor parte de las tuberías estudiadas. Lo anterior, debido a que algunos valores asociados a las características físicas de un conjunto de tuberías (diámetros, materiales, pendientes), necesarios para estimar los valores de confiabilidad de inundación (concepto definido en la siguiente sección de este documento), no se reportaban en las bases de datos. Esta particularidad de las bases de datos existentes es un factor limitante a la hora de utilizar la confiabilidad de inundación como indicador de gestión de la red de alcantarillado. Teniendo en cuenta lo anterior, surge como iniciativa para la realización de este trabajo, el proponer y emplear en una zona de estudio, una herramienta que permita inferir estimativos de confiabilidad de inundación en tuberías donde no es posible realizar su cálculo directo, debido a potenciales deficiencias en la información disponible sobre las características de la red de alcantarillado en un sistema de drenaje urbano (13 subcuencas pertenecientes a la cuenca El Salitre de Bogotá, comprendiendo un área de 4515 Ha, con un total de 12842 tuberías pluviales y combinadas).

Materiales y métodos

La zona de estudio seleccionada para este trabajo (también seleccionada en los trabajos [28] comprende 13 subcuencas pertenecientes a la cuenca El Salitre en la ciudad de Bogotá: Bonanza, Chicó Norte, Colsubsidio, Juan Amarillo Izquierda, Las Delicias, La Vieja, Pinos, Río Negro, Río Nuevo, Salitre, San Marcos, Subcuenca 59 y Tabora. Esta zona abarca una superficie de 4515 Ha, y trabaja con 12842 tuberías pluviales y combinadas. La información utilizada para el cálculo de la confiabilidad de inundación fue proporcionada por la Empresa de Acueducto y Alcantarillado de Bogotá (EAAB),

incluyendo la ubicación georreferenciada (utilizando el Software *ARCGIS*) de cada tubería, con sus características como diámetros, materiales, longitud, pendientes, cotas claves iniciales y finales, cotas rasantes iniciales y finales entre otras, además de las curvas IDF para cada sector de Bogotá. Cabe recalcar que dicha información en muchos casos (cerca del 80 % del total de la zona de estudio) se reportaba como inexistente “NA” (dato no reportado) para varias características de las tuberías, principal motivación del presente trabajo. A continuación se muestra un mapa de la zona de estudio con la descripción de cada Subcuenca (pluvial o combinada):

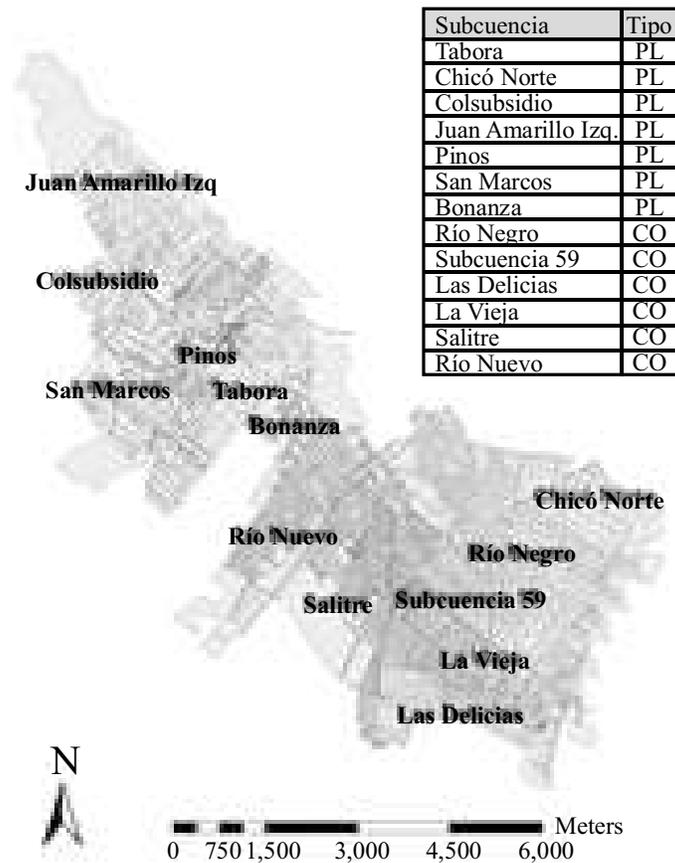


Figura 1 Zona de estudio con la descripción de cada subcuenca (tipo de alcantarillado pluvial (PL) o combinado (CO)). Fuente: [28]

La confiabilidad de inundación (*CI*) puede ser definida como la probabilidad de excedencia de la capacidad hidráulica de cada tubería sometida

a una carga hidrológica de diseño. Al comparar los dos caudales en cada uno de los tramos, se puede obtener el margen de seguridad (*MS*) de

una tubería como la resta entre la capacidad hidráulica y la carga hidrológica. En ese sentido puede plantearse *CI* como la probabilidad de que *MS* sea superior a cero [25].

Para el presente estudio se contó con los resultados de confiabilidad de inundación obtenidos por [28], para periodos de retorno de 3, 5 y 10 años, resultados que reportaban un estimativo numérico de confiabilidad para 2337 tuberías pertenecientes a la zona de estudio; para las 10505 tuberías restantes (valor de confiabilidad etiquetado como NA) no fue posible calcular de manera directa la confiabilidad de inundación [28] (figura 2 a).

Puede apreciarse que hacia la parte occidental son muy pocas las tuberías donde se logró estimar la confiabilidad de inundación, pero en las Subcuencas de Las Delicias, La Vieja, Subcuenca 59, Chicó Norte y Río Negro en su parte alta existe una cantidad de tuberías importante en donde sí fue posible ser definido este valor (aproximadamente 1800 tuberías).

Con el propósito de estimar la confiabilidad de inundación en elementos de la red con información topológica inexistente, se utilizaron las diferentes características de cada una de las tuberías tanto físicas como hidrológicas, así como su respectivo valor de confiabilidad como variable de salida del sistema (Tabla 1).

Tabla 1 Variables de estudio para la confiabilidad de inundación

Variables de entrada		Variable de Salida
Tipo: combinado (CO) o pluvial (PL)	Longitud (m)	
Pendiente tubería (%)	Cota rasante inicial (msnm)	
Profundidad a clave inicial (m)	Profundidad a clave final (m)	
Densidad de tuberías en la Subcuenca (m/m ²)	Fecha de Instalación (años)	
Área propia de drenaje (m ²)	Intensidad lluvia** (mm/h)	Confiabilidad de inundación
Cota clave Inicial (msnm)	Cota clave final (msnm)	
Cota rasante final (msnm)	Pendiente terreno (%)	
Diámetro (m)	Área Subcuenca (m ²)	
Coefficiente de Rugosidad	Número de daños (#)	

*La variable "Densidad de tuberías en la Subcuenca" se calculó como la longitud de tuberías de la red por m².

**La intensidad de la lluvia fue calculada tomando una duración típica de 15 minutos.

Con el fin de explotar los resultados de dichas tuberías donde sí pudo ser calculada la confiabilidad de inundación de manera directa, se implementó un modelo basado en árboles de decisión [21] utilizando la librería J48 [17] de *The WEKA Data Mining Software* [14]. La iniciativa de utilizar árboles de decisión en el presente estudio radica en la flexibilidad proporcionada por esta herramienta en términos de detectar reglas ocultas en bases de datos, aplicables para

características tanto numéricas como cualitativas de cada tubería de alcantarillado.

Otro concepto relevante en el presente estudio fue el estadístico *AMI* (*Average Mutual Information*), el cual corresponde a la información mutua promedio $I(X;Y)$ entre dos variables aleatorias X y Y , el cual se define como la cantidad de información promedio que se obtendría de X a partir de conocer el valor de Y . Matemáticamente, puede expresarse como [10]:

$$I(X; Y) \stackrel{\text{def}}{=} \sum (-\log P(X)) - (-\log P(X|Y)) = \sum \log \frac{P(X, Y)}{P(X)P(Y)} \quad (1)$$

Este estadístico fue utilizado para identificar cuales son las características físicas de una tubería de alcantarillado (*e.g.* diámetro, pendiente, cotas de instalación, longitud, área de drenaje, material) que presentan una mayor influencia sobre la confiabilidad de inundación. Específicamente, la prueba de correlación no lineal (AMI) se aplicó entre todas las variables de entrada (tipo, longitud, diámetro, coeficiente de rugosidad, pendientes de tubería y terreno, cotas claves y rasantes, profundidades a cotas claves, área de subcuenca y de drenaje, densidad de tuberías en la subcuenca, fecha de instalación, número de daños, intensidades de lluvia para periodos de retorno de 3, 5 y 10 años) y la confiabilidad de inundación para diferentes periodos de retorno de 3, 5 ó 10 años.

El siguiente paso abordado consistió en la implementación de árboles de decisión. El propósito de la utilización de esta herramienta es realizar estimativos de confiabilidad de inundación de aquellas tuberías de las que no se poseen suficientes valores característicos para calcular la confiabilidad de inundación, y por lo tanto etiquetados con “NA”. Se tuvieron presentes las mismas variables identificadas como más representativas según los resultados de la prueba AMI. Con el fin de implementar dentro del árbol variables no numéricas “NA”, ejercicio pertinente para lograr una representación de la problemática mediante árboles de decisión, se construyó una métrica de las variables asociadas a cada tubería. Se establecieron rangos (ALTO, MEDIO, BAJO) para todas las variables de estudio (entrada y salida), dividiendo el intervalo acotado por el valor máximo y mínimo de cada variable presente en

la zona de estudio en tres sub-intervalos de igual longitud. De esta manera, también se obtuvo un rango cualitativo de confiabilidad estimada (CI_ALTO, CI_MEDIO, CI_BAJO).

En el caso del material de la tubería, ya que los únicos materiales en la zona de estudio observados fueron concreto y PVC, y basándose en la rugosidad de dichos materiales, se adoptó la siguiente convención: (i) Concreto =ALTO y (ii) PVC=BAJO. Para aquellas tuberías donde se reportara un valor de “NA” en alguna de sus variables, se conservó esta asignación para su rango.

Habiendo definido el conjunto de variables mediante etiquetas, se utilizó la base de datos sin tuberías con “NA” como resultado de confiabilidad de inundación para construir un árbol de decisión que representara los datos de rangos de confiabilidad de una manera en que los estimativos a realizar fueran coincidentes con las confiabilidades de inundación observadas. El árbol fue construido con la herramienta *The WEKA Data Mining Software* [14], utilizando la librería J48 [17].

Por otro lado, en el árbol de decisión se reporta para cada posible salida el número de veces en que un patrón presenta una configuración específica y el número de veces en que la estimación es equivocada. Esto brinda la posibilidad de calcular para cada resultado un nivel de confianza en su estimación. Por lo tanto, en términos prácticos, esta herramienta brinda la posibilidad de presentar un resultado de confiabilidad de inundación para las tuberías restantes como “CI_ALTO”, “CI_MEDIO” o “CI_BAJO”, con un nivel de confianza asociado a cada resultado.

Para poder calcular la confianza asociada a cada rango de confiabilidad de inundación reportado en una tubería dada, puede definirse la siguiente ecuación:

$$NC = \frac{\text{(número de tuberías en las que fue correcta la predicción por la rama}_i)}{\text{número de tuberías clasificables mediante la rama}_i} \quad (2)$$

Donde NC es el nivel de confianza para la confiabilidad de inundación reportada.

Puede apreciarse que NC será un resultado numérico ($0 \leq NC \leq 1,00$). Por lo tanto, con el fin de lograr consistencia entre las características de las tuberías, los estados de confiabilidad y el nivel de confianza asociable al estado de confiabilidad reportado para cada tubería, se hace necesario expresar dicho nivel de confianza en rangos, de manera análoga a las confiabilidades de inundación: (i) NC_BAJO ($0,00 \leq NC \leq 0,33$),

(ii) NC_MEDIO ($0,33 \leq NC \leq 0,66$), y (iii) NC_ALTO ($0,66 \leq NC \leq 1,00$).

La toma de decisión de qué tuberías deben tener prelación para mantenimiento proactivo debe presentarse en función de los rangos de confiabilidad de cada tubería y del nivel de confianza que se está dispuesto a aceptar en dicho estado de confiabilidad estimado. Con el fin de lograr exponer en conjunto estas dos etiquetas para cada una de las tuberías en una zona estudiada, puede acudir al tomador de decisiones para proponer un esquema como el que se muestra en la tabla 2.

Tabla 2 Ejemplo de esquema de priorización en función de CI y NC, proporcionado por el tomador de decisiones

		<i>Confiabilidad de inundación</i>			
		<i>CI_BAJO</i>	<i>CI_MEDIO</i>	<i>CI_ALTO</i>	
	NC_BAJO	<i>PR_MEDIA</i>	<i>PR_BAJA</i>	<i>PR_MEDIA</i>	
Nivel de confianza	NC_MEDIO	<i>PR_MEDIA</i>	<i>PR_MEDIA</i>	<i>PR_MEDIA</i>	Priorización
	NC_ALTO	<i>PR_ALTA</i>	<i>PR_ALTA</i>	<i>PR_BAJA</i>	

Donde, dependiendo del rango de confiabilidad de inundación de una tubería y del nivel de confianza para este resultado de confiabilidad, puede priorizarse una tubería de alcantarillado como “PR_ALTA”, “PR_MEDIA” o “PR_BAJA”, en términos del mantenimiento proactivo que requiere. Para este artículo se propone la priorización expuesta en la tabla 2, no sin antes aclarar que es finalmente el tomador de decisiones de la empresa encargada de la gestión del alcantarillado urbano quien debe establecer cuál es la priorización que desea asumir en función de múltiples variables como: (i) la disposición de presupuesto para esta actividad, (ii) el estado actual del sistema, (iii) la periodicidad de actividades de mantenimiento preventivas, entre otras. Inicialmente, se ejecutó el programa para las tuberías utilizadas en la construcción del árbol, con el fin de, representando gráficamente los resultados, comparar espacialmente los

rangos de confiabilidad observados (datos del trabajo de [28]).

Habiendo obtenido un árbol de decisión con una capacidad predictiva adecuada, el siguiente paso consistió en implementar el modelo para todas aquellas tuberías con confiabilidad “NA”, con el fin de inferir un estimador representativo del estado del sistema de alcantarillado en términos de gestión sobre toda la zona de estudio denominado “Priorización”. Para realizar la inferencia de este indicador, se cruzaron los estados de confiabilidad de inundación y los niveles de confianza para cada tubería, utilizando la información ejemplo consignada en la tabla 2.

Resultados y discusión

Los resultados al aplicar la prueba de correlación no lineal AMI se presentan en la tabla 3.

Tabla 3 Resultados prueba AMI

Características de cada tubería (X_i)	$I(X_i;ci3)$	$I(X_i;ci5)$	$I(X_i;ci10)$
tipo (CO o PL)	0,301	0,293	0,294
longitud (m)	0,137	0,137	0,137
Cota clave Inicial (msnm)	0,319	0,320	0,324
Cota clave final (msnm)	0,362	0,363	0,368
Pendiente tubería (%)	0,360	0,360	0,363
Cota rasante inicial (msnm)	0,361	0,358	0,364
Cota rasante final (msnm)	0,352	0,353	0,357
Pendiente terreno (%)	0,458	0,459	0,461
Profundidad a clave inicial (m)	0,056	0,056	0,056
Profundidad a clave final (m)	0,131	0,130	0,131
Diámetro (m)	0,046	0,048	0,054
Área Subcuenca (m ²)	0,193	0,193	0,194
Densidad de tuberías en la Subcuenca (m/m ²)	0,230	0,229	0,229
Fecha de Instalación (años)	0,192	0,189	0,190
Coefficiente de Rugosidad	0,652	0,652	0,653
Número de daños	0,027	0,025	0,025
Área propia de drenaje (m ²)	0,006	0,007	0,007
Intensidad lluvia TR3 (mm)	0,327	0,325	0,327
Intensidad lluvia TR5 (mm)	0,338	0,336	0,337
Intensidad lluvia TR10 (mm)	0,724	0,722	0,723

De acuerdo con la tabla 3, la relación entre cualquier variable de entrada y todos los valores de confiabilidad de inundación para los diferentes periodos de retorno estudiados (3, 5 ó 10 años) es del mismo orden, lo que permite afirmar que la influencia de las variables de entrada es la misma para una confiabilidad asociada a cualquier periodo de retorno estudiado. Además, se aprecia que la intensidad de la lluvia para un periodo de retorno de 10 años (TR10) es la variable de entrada más influyente sobre todas las posibles variables de salida. En ese sentido, resulta más conveniente definir el esquema conceptual del modelo utilizando la confiabilidad de inundación

para un periodo de retorno a 10 años (CI10) como variable de salida, y la intensidad de la lluvia asociada a un TR10 como única variable de entrada hidrológica.

Puede observarse que existe un conjunto de variables (cota clave final, cota rasante inicial y pendiente de la tubería) que presentan un resultado de la prueba *AMI* equivalente ($I(X_i;ci10)=0,36$) (ver tabla 3). Con el fin de seleccionar las variables de entrada a utilizar en la construcción del modelo, se establecieron dos criterios: (i) favorecer la parsimonia del modelo, al incluir más del 20 % de las características observables en una tubería

de alcantarillado, (ii) no excluir aquellas variables que reportaran el mismo *AMI* que alguna variable previamente seleccionada como representativa mediante el criterio (i). En ese sentido, se decidió incluir aquellas variables de entrada con valores de $I(X_i; c_i | 10)$ mayores a 0,36 (Cota rasante final, Tipo, Densidad de tuberías en la Subcuenca, Área Subcuenca, Fecha de instalación, Longitud, Profundidad a clave final, Profundidad a clave inicial, Diámetro, Número de daños y Área propia de drenaje). De lo anterior, se obtuvo que las variables más influyentes en *CI10* (variable de salida) de las tuberías estudiadas son: (i) Intensidad de la lluvia (ii) Coeficiente de rugosidad (iii) Pendiente del terreno (iv) Cota clave final (v) Cota rasante inicial (vi) Pendiente de la tubería, lo que permitió definir el esquema conceptual de modelación en función de éstas como variables de entrada en el modelo. Se observa que dentro de la lista anterior no se encuentran variables que han sido detectadas en la literatura como claves en gestión patrimonial de alcantarillados y que fueron evaluadas en el presente estudio como el diámetro, la longitud y la edad de la tubería, así como profundidades de instalación, tipo de alcantarillado y tipo de vía [12, 20, 20]. Lo anterior podría deberse a que en el presente estudio

se enfatizó en la confiabilidad de inundación como indicador único de demanda de mantenimiento de una tubería, sin contemplar variables relacionadas con el entorno como los tipos de alcantarillado y de vía, por ejemplo. Sin embargo, otras variables como el diámetro y la longitud de la tubería deberían tener una influencia sobre el indicador estudiado.

Mediante la implementación del *software The WEKA Data Mining Software* [14], empleando la librería J48 [17], omitiendo aquellas tuberías con etiqueta de salida “NA” en el planteamiento del modelo, se construyó un árbol de decisión donde el número de ramas representa las posibles combinaciones de características presentadas por las tuberías. El árbol calculado mediante la herramienta computacional reportó el número de tuberías clasificadas correctamente y el número de tuberías clasificadas de manera incorrecta para cada una de sus ramas. Con estos valores se calculó numéricamente el nivel de confianza en la predicción realizada por cada una de las ramas del árbol, para luego asignar a cada valor de nivel de confianza una etiqueta “NC_ALTO”, “NC_MEDIO” o “NC_BAJO”, según correspondiese (figura 3).

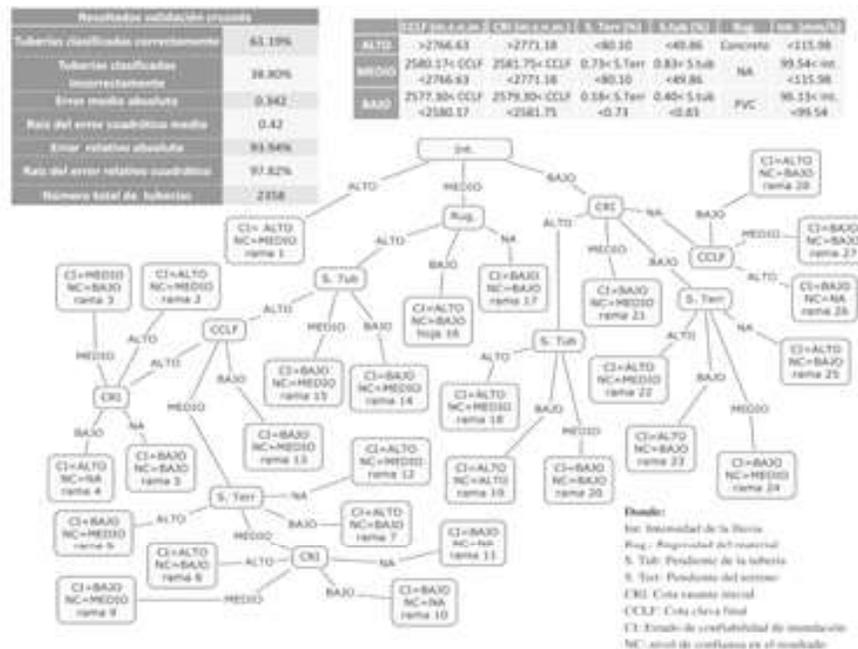


Figura 3 Árbol de decisión construido para inferir CI a partir de las características de una tubería de alcantarillado urbano

Puede apreciarse en la figura 3 que, según el modelo ejecutado, existen 28 reglas que clasifican la base de datos en 28 tipos de tuberías diferentes (árbol de 28 ramas). Además, se observa que la variable más influyente en cuanto a la confiabilidad de inundación es la intensidad de la lluvia, ya que es la primera instancia de clasificación en el árbol. Dicho resultado es consecuente con las pruebas *AMI* reportadas en la tabla 3. En cuanto a los niveles de confianza asociados a cada una de las ramas del árbol, se presentaron: “NC_MEDIO” (43 %), “NC_BAJO” (39 %), “NC_ALTO” (4 %) y 14 % no pudieron ser calculados (NC_NA). En general, puede afirmarse que los niveles de confianza asignados a la inferencia realizada acerca de la confiabilidad de inundación para cada tubería dependen de la rama con que se haya realizado dicha predicción. También es notable que alrededor del 38 % de las tuberías no fue clasificada de manera asertiva. Específicamente, dentro de las tuberías que no fueron clasificadas de manera asertiva, se observó que la configuración más frecuente presentada por éstas fue la siguiente (146 tuberías, 16 % del

total clasificado erróneamente. Ver convenciones adicionales figura 3): (i) CCLF: ALTA (ii) S. Tub: ALTA, (iii) CRI: ALTA (iv) S. Terr: ALTA, (v) Rug: ALTA, (vi) Int: ALTA. De acuerdo con lo anterior, es posible que exista una mayor variabilidad en el valor de CI para este tipo de tuberías, debido a que algunas de estas variables presentan una relación directa con CI (S. Tub) y otras una relación inversa (Rug., Int.) (Adaptado de [28]).

El árbol de decisión obtenido mediante *The WEKA Data Mining Software* [14], mostrado en la figura 3, fue reproducido utilizando un aplicativo computacional en el que al evaluar los datos de entrada para una tubería, se otorga la confiabilidad de inundación reportada como “CI_ALTO”, “CI_MEDIO” o “CI_BAJO”, además del nivel de confianza en el resultado obtenido, rotulado como “NC_ALTO”, “NC_MEDIO” o “NC_BAJO”. La figura 2 muestra, para efectos de comparación, los rangos de confiabilidad obtenidos por [28] y los resultados de confiabilidad simulados mediante el árbol de decisión.

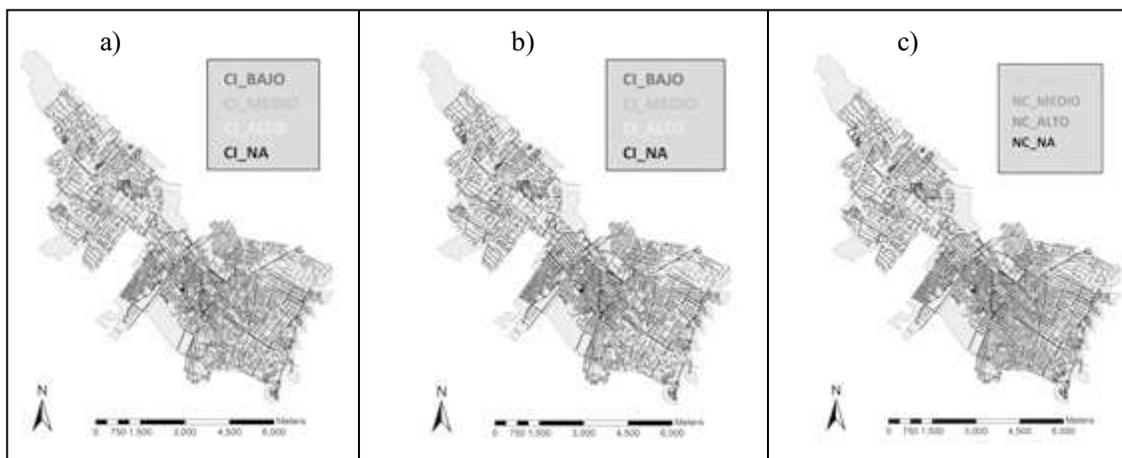


Figura 2 Mapas de estados de confiabilidad de inundación observados [28] (a), simulados (mediante el árbol de decisión) (b) y el nivel de confianza asociado a la inferencia de la confiabilidad de inundación mediante el árbol de decisión (c)

Hacia la parte sur de la zona de estudio se presentan algunas discrepancias entre lo observado y lo simulado (Subcuencas Salitre, La Vieja y Las Delicias; ver figura 2 a y b). Por otro

lado, hacia la parte noroccidental de la Subcuenca Río Negro, en donde [28] habían determinado un cúmulo de demanda de mantenimiento proactivo bastante importante, se mantuvo una similitud

notoria con lo modelado empleando el árbol de decisión (figuras 3 a y b). En cuanto a las tuberías con estado de confiabilidad “CI_MEDIO”, se obtuvo una asertividad de 0,83 %. Se reportó para los estados de confiabilidad de inundación “CI_BAJO” y “CI_ALTO”, 68 % y 60 % de resultados correctos, respectivamente. En general, el porcentaje de ocasiones en que el modelo realizó una predicción correcta para cualquier rango de CI fue cercano al 62 %. Dicho resultado, se hace plausible en el mapa de niveles de confianza que reporta en su mayoría tuberías con etiqueta “NC_MEDIO” (94 % de las tuberías) (figura 2 c), lo cual es del mismo orden de magnitud que resultados obtenidos en investigaciones similares para otras zonas de estudio [9, 13].

El mapa presentado en la figura 2 c proporciona un estimativo del comportamiento espacial de la capacidad predictiva del modelo, permitiendo ubicar para qué zonas existe una mayor capacidad predictiva, medida a través del nivel de confianza en el resultado reportado para los datos de partida en la construcción del árbol.

A continuación se muestran los resultados en forma de mapa, de los estados de confiabilidad de inundación, de los niveles de confianza y de la priorización del mantenimiento, obtenidos a partir de la implementación del modelo sobre toda la zona de estudio (figura 4).

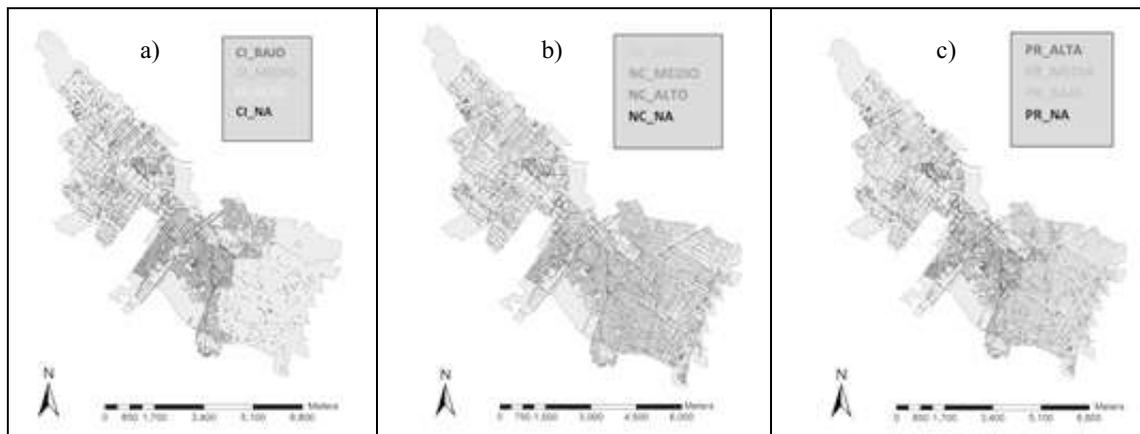


Figura 4 Estados de confiabilidad de inundación inferidos en toda la zona de estudio (a). Nivel de confianza en los resultados obtenidos (b) y priorización para el mantenimiento proactivo en la zona de estudio (c)

Puede notarse cómo la modelación permitió estimar la confiabilidad de inundación en prácticamente toda la zona de estudio (figura 4 a). Aún se siguen reportando 1455 valores de confiabilidad de inundación inexistentes (alrededor del 11 %) hacia la parte central de la zona estudiada, ya que existen tuberías con una combinación de variables de entrada que no pudieron ser clasificadas por el árbol. Existe un conjunto bastante importante de confiabilidades en estado “CI_BAJO” hacia la Subcuenca Salitre, diferente de las Subcuencas de La Vieja, Las Delicias, Río Negro y Subcuenca 59, que

presentan en general estados de confiabilidad “CI_ALTO”.

Es apreciable que en términos generales se obtuvo un nivel de confianza bastante alto “NC_ALTO” en las inferencias realizadas hacia la zona central de la zona de estudio (subcuencas Tabora, Pinos, Bonanza, Río Nuevo, Salitre). Hacia la parte oriental, los resultados de confiabilidad presentan un nivel de confianza clasificado en general como “NC_MEDIO”, lo que permite afirmar que las tuberías con las que fue construida el modelo tienden a reportar un nivel de confianza más bajo

en sus resultados que aquellas tuberías en donde no se conocía la confiabilidad de inundación. No se obtuvieron niveles de confianza “NC_BAJO” ($0,00 \leq NC \leq 0,33$) para la fase predictiva del modelo, al igual que para las tuberías utilizadas en la construcción del árbol. En cuanto a los resultados totales, se reportaron niveles de confianza “NC_MEDIO” (7260 tuberías), “NC_ALTO” (4078 tuberías) y “NC_NA” (1505 tuberías), para toda la zona de estudio (figura 4 b).

En términos generales, para el escenario de priorización propuesto, en el caso de estudio analizado, se encontró que alrededor del 5 % de las tuberías presentan una priorización “PR_BAJA”, ubicadas principalmente hacia las subcuencas de Juan Amarillo y Colsubsidio en la parte occidental y en Río Negro y Río Nuevo, ubicadas hacia el oriente de la figura 4 c. La priorización más alta, es decir las tuberías más críticas “PR_ALTA” (alrededor del 26 %, 3383 tuberías), se encontraron hacia la subcuenca Salitre, Río Nuevo en su parte sur, una zona específica de Río Negro y hacia el costado norte de Pinos y Tabora. Las tuberías en estado de priorización “PR_MEDIO” (7260 tuberías, alrededor del 57 % sobre el total) se identificaron en general hacia el costado oriental de la zona de estudio. En cuanto a las tuberías donde no pudo ser asignado un nivel de priorización debido a la imposibilidad de clasificarlas mediante el árbol de decisión (alrededor del 11 %), su priorización se reportó como “PR_NA”. Para este caso, se pudo apreciar una tendencia en su ubicación hacia la parte central de la zona de estudio al igual que con los rangos de confiabilidad de inundación y niveles de confianza no reportados (CI_NA y NC_NA).

Conclusiones

Para el caso estudiado, que consistía en inferir, con base en árboles de decisión, la confiabilidad de inundación en un sistema de alcantarillado urbano para un periodo de retorno de 10 años, basándose en estudios previos realizados por

[28] donde los resultados se vieron limitados debido a la falta de información detallada de la red de alcantarillados, se reportaron resultados medianamente satisfactorios en la medida en que el modelo presentó un ajuste de 62 % de las instancias clasificadas correctamente con respecto a los datos observados, permitiendo inferir la confiabilidad de inundación para cerca del 60 % de las tuberías de la red de alcantarillado estudiada, lo cual otorga mayor información a los gestores de dicha red comparado con el estudio inicial de [28] en el cual tan sólo se tenía un estimativo de confiabilidad de inundación para un 18 % de la red (2337 tuberías sobre un total de 12842).

Dentro de las variables detectadas mediante la prueba de información mutua promedio (AMI, por sus siglas en inglés) como influyentes sobre los valores de confiabilidad de inundación en las tuberías de alcantarillado no se encuentran variables que han sido detectadas en la literatura como claves en gestión patrimonial [12, 20, 24]. Dado que en el presente trabajo no se estudiaron las causas que explican el resultado obtenido, se recomienda profundizar sobre este aspecto en estudios posteriores, teniendo en cuenta elementos como la calidad y la cantidad de la información disponible en la zona de estudio (ver recomendaciones de [12]).

Los porcentajes de asertividad obtenidos para cada rango de confiabilidad de inundación (BAJO: 60 %; MEDIO: 0,83 %; ALTO: 68 %) sugieren que es necesario realizar una delimitación más fina de los niveles (más de tres niveles) o una representación continua de dicho indicador, tal como se realiza comúnmente en este tipo de estudios [32].

Los estimativos como los obtenidos en el presente trabajo, en conjunto con herramientas de apoyo como los sistemas de información georreferenciada [9], brindan la posibilidad de construir mapas que permitan a la empresa encargada de la gestión del sistema de alcantarillado en una cuenca urbana tomar decisiones más soportadas en cuanto a la priorización espacial

del mantenimiento proactivo de dicho sistema, con el fin de garantizar un manejo más eficiente de los recursos de la empresa destinados para esta finalidad.

Se recomienda a las empresas encargadas del desarrollo y gestión de la infraestructura de drenaje urbano consolidar bases de datos que contengan información topológica completa y de buena calidad (e.g. diámetros, pendientes, cotas). Especialmente, para el caso de estudio se recomienda enfatizar la revisión de la base de datos en los materiales de las tuberías, ya que es a partir de esta característica que puede definirse la confiabilidad de inundación directamente sin necesidad de utilizar metodologías adicionales para estimar este tipo de indicadores de gestión.

Mediante el uso de información asertiva acerca de la conectividad del sistema de alcantarillado en conjunto con información hidrológica, puede obtenerse un estimativo de la confiabilidad de inundación, el cual resulta funcional en términos de gestión del sistema de drenaje urbano. Por lo anteriormente expuesto, se recomienda a las empresas encargadas de dicha gestión realizar esfuerzos en la obtención y mejora de esta información, en especial lo que concierne el detalle de pozos y sumideros (e.g. tuberías de entrada y salida, diámetros, cotas).

El árbol de decisión propuesto se entrenó y validó para un caso específico de estudio (Cuenca el Salitre, Bogotá). Para futuras investigaciones en otros centros urbanos se recomienda que bajo enfoques similares al abordado en el presente trabajo, se plantee un árbol de decisión con base en información topológica e hidrológica local cuya variable de salida sea un indicador de gestión conocido (confiabilidad de inundación para este estudio).

Con el fin de verificar el desempeño de diferentes modelos e indicadores del estado del sistema de alcantarillado que se han venido reportando en la literatura, cuya implementación a su vez se identificó como inviable para el caso de estudio debido a las características de la información requerida como insumo para dichos

modelos, se recomienda incrementar el número de inspecciones sistemáticas, con un mayor cubrimiento y seguimiento en el tiempo de las tuberías del sistema de alcantarillado.

La confiabilidad de inundación puede ser un indicador funcional para la toma de decisiones en gestión patrimonial de alcantarillados urbanos. Con el fin de otorgar mejor representatividad del indicador mencionado para la zona de estudio, puede recomendarse ejecutar su cálculo a partir de registros de lluvias reales, teniendo en cuenta variabilidades espaciales (información de varias estaciones sobre la cuenca de estudio) y temporales en lugar de utilizar curvas IdF.

Teniendo en cuenta que todas aquellas tuberías en concreto con pendientes mayores a 49 % y que a su vez presentan una exposición hidrológica alta (intensidades de lluvia mayores a 116 mm/h en la zona de la tubería) presentan mayor incertidumbre en el cálculo de la confiabilidad de inundación, se recomienda prestar especial atención a este tipo de tuberías, toda vez que se requiera, por efectos de información incompleta o de baja calidad, estimar confiabilidad de inundación mediante el uso de árboles de decisión y conceptos de entropía de información.

Referencias

1. E. Ana, W. Bauwens. "Sewer Network Asset Management Decision-Support Tools: A Review". *International Symposium on New Directions in Urban Water Management*. 2007. Paris: UNESCO. pp. 1-8.
2. S.T. Ariaratnam, C.W. MacLeod. "Financial outlay modeling for a local sewer rehabilitation strategy". *Journal of Construction Engineering and Management* 2002;128(6):486-495. *Journal of Construction Engineering and Management*. 2002. pp. 486-496.
3. S. Arthur, H. Crow, L. Pedezert, N. Karikas. "The holistic prioritisation of proactive sewer maintenance". *Water Science & Technology*. Vol. 59. 2009. pp. 1385 - 1396.
4. M. Aulinas, J. Nieves, U. Cortes, M. Poch "Supporting decision making in urban wastewater systems using a knowledge-based approach" 2011 *Environmental Modelling and Software*. Vol. 26. pp. 562-572.

5. H. Baik, H. Jeong, D. Abraham, "Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems". *Journal of Water Resources Planning and Management*. Vol. 132. 2006. pp. 15-24.
6. R. Baur, R. Herz. *Selective Inspection Planning with Aging Forecast for Sewer Types*. Efficient Water Management. 2001. pp. 389-396.
7. S. Bennis, J. Bengassem, P. Lamarre. "Hydraulic Performance Index of a Sewer Network". *Journal of Hydraulic Engineering*. Vol 129. 2003. pp. 504-510.
8. N. Caradot, D. Granger, J. Chaogier, F. Cherqui, B. Chocat. "Urban flood risk assessment using sewer flooding databases". *Water Science and Technology*. Vol. 64. 2011. pp. 832 – 840.
9. J. Chen, A. Hill, L. Urbano "A GIS-based model for urban flood inundation". *Journal of Hydrology*. Vol. 373. 2009. pp. 184-192.
10. T. Cover, A. Joy. *Entropy, relative entropy and mutual information*. Elements of information theory (2nd Ed). John Wiley & Sons, New Jersey. 2006. pp. 748.
11. P. Davis, S. Dhammika, D. Marlow, M. Mogolia, S. Gould, S. Burn. *Failure Prediction and Optimal Scheduling of Replacements in Asbestos Cement Water Pipes*. Australia: IWA Publishing. 2008. pp. 239-252.
12. R. Fenner. "Approaches to sewer maintenance: A review". *Urban Water*. Vol. 2. 2000. pp. 343-356.
13. D. Fernandez, M. Lutz. "Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis", *Engineering Geology*. Vol. 111. 2010. pp. 90-98.
14. M. Hall, E. Frank, G. Holmes, B. Fahringer, P. Reutemann, I. Witten. "The WEKA Data Mining Software: An Update". *SIGKDD Explorations*. Vol. 11. 2009. pp. 10-18.
15. Z. Liu, Y. Kleiner. "State of Art Review of Inspection Technologies for Condition Assessment of Water Pipes", *Measurement*. 2012. <http://dx.doi.org/10.1016/j.measurement.2012.05.032> (Revisado Julio 2012)
16. T. Micevski, G. Kuczera, P. Coombes. "Markov Model for Storm Water Pipe Deterioration". *Infrastructure Systems*. Vol. 8. 2002. pp. 49-56.
17. J. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco. 1993. pp. 302.
18. A. Quintero, D. Konare, S. Pierre. "Prototyping an Intelligent Decision Support System for improving urban infrastructures management". *European Journal of Operational Research*. Vol. 162. 2005. pp. 654-672.
19. J. Rodríguez, M. Díaz, J. Penagos, N. McIntyre, Č. Maksimović. "A serviceability model for supporting proactive maintenance in urban drainage systems". *Proceedings of the 12th International Conference on Urban Drainage*. Porto Alegre, Brazil. 11-16 September 2011.
20. J. Rodríguez, N. McIntyre, M. Díaz, Č. Maksimović. "A database and model to support proactive management of sediment related sewer blockages", *Water Research* (2012), <http://dx.doi.org/10.1016/j.watres.2012.06.037> (Revisado Julio 2012)
21. L. Rokach, O. Maimon. *Data mining with decision trees: theory and applications*. World Scientific Publishing, Singapore. 2008. pp. 244.
22. J. Ruwanpura, S. Ariaratnam, A. El-Assaly. "Prediction models for sewer infrastructure utilizing rule-based simulation". *Civ. Eng. Environ. Syst.* Vol. 21. 2004. pp. 169-185.
23. S. Saegrov. *CARE-S: Computer Aided Rehabilitation of Sewer and Storm Water Networks*. Londres: IWA Publishing. 2006. pp. 160.
24. B. Salman, O. Salem. "Modeling Failure of Wastewater Collection Lines Using Various Section-Level Regression Models." *J. Infrastruct. Syst.* Vol. 18. 2012. pp. 146-154.
25. S. Sandoval, A. Torres, K. Navarro. *Estimación de la confiabilidad de inundación en alcantarillados pluviales de Bogotá mediante métodos de generación aleatoria*. Evento: Agua 2011 ECOSISTEMAS Y SOCIEDAD, Seminario Internacional: "Inundaciones, Escasez de Agua y Ecosistemas, Acciones Frente al Cambio Climático". Cali. Colombia. 2011.
26. S. Stone, E. Dzuray, D. Meisegeier, A. Dahlgro. *Decision-Support Tools for Predicting the Performance of Water Distribution and Wastewater Collection Systems*. US EPA. 2002. pp. 97.
27. W. Tagherout, S. Bennis, J. Bengassem. "A Fuzzy Expert System for Prioritizing Rehabilitation of Sewer Networks". *Computer-Aided Civil and Infrastructure Engineering*, Vol. 26. 2011. pp. 146-152.
28. A. Torres, S. Sandoval, K. Navarro, L. Pulido. *Strategies of maintenance in storm sewers in Bogotá Colombia by using estimates of operational and flooding reliabilities*. Proceedings of the 12th International Conference on Urban Drainage. Porto Alegre. Brazil. 11-16 September. 2011. pp. 9.

29. A. Torres, Z. Méndez. *Estimación del riesgo en estructuras de drenaje pluvial mediante herramientas de inteligencia artificial*. Seminario Internacional: “La Hidroinformática en la Gestión Integrada de los Recursos Hídricos”. Ponencia: Libro: Agua 2003. Cartagena de Indias. Colombia. pp. 31-38.
30. T. Veldhuis, J. Clemens, F. Clemens. “The efficiency of asset management strategies to reduce urban flood risk”. *Water Science & Technology*. Vol. 64. 2011. pp. 1317-1324.
31. M. Yang, T. Su. “Automated diagnosis of sewer pipe defects based on machine learning approaches”. *Expert Systems with Applications* (2007). doi:10.1016/j.eswa.2007.08.013
32. Q. Zhou, P. Mikkelsen, K. Halsnæs, K. Arnbjerg. “Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits”. *Journal of Hydrology*, Vol. 414-415. 2012. pp. 539-549.