

DETECCIÓN DE FALLOS EN REDES DE PRESIÓN PERTENECIENTES A REDES DE SUMINISTRO DE AGUA MEDIANTE FILTROS ESTÁTICOS LINEALES BASADOS EN PCA

M. Pereira, J. Blesa, V. Puig y D. Muñoz de la Peña
 DI, Universidad Loyola Andalucía, mpereira@uloyola.es
 IRI, Universitat Politècnica de Catalunya (UPC), {joaquim.bleesa, vicenc.puig} @upc.edu
 DISA, Universidad de Sevilla, dmunoz@us.es

Resumen

Este artículo se centra en la elaboración de una serie de modelos estáticos lineales que relacionan los datos de un conjunto de sensores de presión con un subconjunto de ellos que denominaremos base. Gracias a estos modelos se pueden detectar comportamientos anómalos en el funcionamiento de dichos sensores mediante la comparación de los datos reales con los estimados. Para la detección de anomalías en la base se empleará otro subconjunto secundario para su chequeo que denominaremos base secundaria. Como caso de estudio se han utilizado datos almacenados de los sensores de presión de una sección de la red de suministro de aguas de la ciudad de Madrid.

Palabras clave: Detección de fallos, modelado y simulación de redes de suministro de agua, algoritmos basados en componentes principales

1. Introducción

Los sistemas de tuberías subterráneos son infraestructuras de gran importancia debido a la fuerte dependencia que la sociedad tiene de ellas a pesar de que actualmente están relativamente anticuadas. Este hecho, además de su instalación bajo tierra, hacen que su monitorización sea muy compleja. Por si esto no fuera ya un reto, la variedad de materiales utilizados en estas tuberías provocan la aparición de diferentes clases de mecanismos que las deterioran complicando el problema aún más (ver [14]).

Las pérdidas debidas a fallos en tuberías, tales como las clásicas fugas, no se limitan únicamente a pérdidas directas asociadas con la desaparición del líquido sino que incluyen otros aspectos nocivos, indirectos y sociales, tales como la interrupción del suministro o deterioro de imagen pública y credibilidad (ver [1, 11]). Estos efectos tienen un gran impacto en los propietarios y en la sociedad. Como añadido, el coste total del fallo es dependiente del tiempo que tarde la reparación, de manera que la detección rápida de estas averías es una parte importante de la gestión de los sistemas de suministro de aguas.

Hay una amplia variedad de técnicas que pueden ser

usadas a la hora de detectar fallos en tuberías. Estas son llamadas comúnmente "non-destructive testing/evaluation" (NDT/NDE), es decir, técnicas de chequeo y evaluación no destructiva [1]. Su objetivo principal es el de cuantificar el deterioro de las redes de tuberías y encontrar fallos potenciales tempranos. Dichas estrategias son fuertemente dependientes de los sistemas de monitoreo y medida [7, 8, 16]. Los métodos más comunes para estos propósitos son los acústicos, es decir, aquellos basados en detectar el sonido emitido por la fuga en la tubería usando hidrófonos que necesitan tener acceso al interior de ésta. Por tanto estaríamos ante un conjunto de métodos invasivos. Además de esto, no están preparados para un despliegue espacial amplio y para el monitoreo continuo debido a su alto coste energético, alto muestreo y alto coste computacional.

Una amplia gama de las técnicas NDE se basan en encuestas geofísicas de la superficie terrestre y no son válidas para despliegues a largo plazo en redes de tuberías. Ejemplos de estas técnicas son los sistemas basados en visión y radares de penetración terrestre (GPR). Además, estas técnicas tienen un alto consumo energético y también son invasivas. Todas estas metodologías tienen sus ventajas e inconvenientes dependiendo de la tecnología utilizada [3, 7, 11, 12].

Las redes de sensores subterráneas inalámbricas (WSDN) ofrecen una plataforma muy adaptable para la monitorización de amplias redes de tuberías gracias a su escalabilidad y capacidad de monitorear de manera continuada. Existe una amplia variedad de este tipo de redes [2, 19, 17, 13, 5, 21, 15]. Sin embargo, dichas redes dependen usualmente de sensores de presión que introducen puntos de fuga similares a los de las líneas de servicio [18] o contaminación en los tubos de cemento de asbesto [20]. Además, el alto coste de estos sistemas comparado con las técnicas de topografía tradicionales hacen que su amplio despliegue no sea económicamente factible. Esto hace necesario un sistema de monitoreo de bajo coste, sencillo de instalar y no invasivo.

Este artículo presenta un sistema de monitorización no invasivo para redes de sensores de presión en redes de suministro de agua. Dicho sistema se basa en el estudio de los datos proporcionados por una red de sensores de presión y aprovechándose de la repetitividad de

la información en ciertos sensores puede detectar anomalías de manera sencilla y con un bajo coste computacional.

2. Formulación del problema

Las medidas de presión de las redes de suministro de agua en tuberías están fuertemente correladas, es decir, se necesitan muy pocos sensores de presión para estimar el comportamiento de la totalidad de la red. Dicho comportamiento es bastante estable en el tiempo debido al carácter periódico del consumo y a la geometría fija de la red. Ejemplos que justifiquen este hecho pueden encontrarse en [4].

Suposición 1 *La red de distribución de agua tendrá un número de sensores de presión superior a r , siendo r el número mínimo de sensores necesarios para reconstruir el comportamiento de la red con un grado de repetitividad nulo en la información.*

Dado un conjunto de n medidas discretas

$$\mathbf{v}_j = [v_j(0) \cdots v_j(n-1)]^T \quad j = 1, \dots, m$$

asociadas a m sensores de presión de una red de distribución de agua en tuberías, y suponiendo que dichos datos no han sido previamente validados y/o reconstruidos, se presenta una metodología para la detección de errores en los sensores asociados a cada serie temporal \mathbf{v}_j .

Cada elemento del vector \mathbf{v}_j se corresponde con una medida del instante de tiempo k sensible a posibles errores de medición. Por lo tanto, se tendrá una matriz $M \in \mathcal{R}^{n \times m}$ cuyas columnas serán los vectores \mathbf{v}_j .

Lo comentado al inicio de esta sección y la suposición 1 serán la base principal para el cálculo de un conjunto de estimadores o filtros estáticos lineales que permitirán vincular el comportamiento de los $m - r$ nodos de presión a r nodos que conformarán lo que se denominará *base*. Por tanto, se podrán detectar evoluciones temporales anómalas en estos $m - r$ sensores que podrán estar asociadas a fallos de los sensores, fugas de agua, etc. En cualquier caso, serán situaciones críticas que deberán ser resueltas por los correspondientes operarios y que podrán acotarse geográficamente con facilidad.

3. Determinación de la base y filtros estáticos lineales

El cálculo de los estimadores o filtros estáticos se puede descomponer en la siguientes etapas:

1. Selección de los sensores más representativos a nivel de información para la reconstrucción de los $m - r$ sensores. A estos sensores se los denominará *base principal*.

2. Cálculo de los modelos estáticos lineales que relacionan los $m - r$ sensores con la base principal.
3. Selección de un conjunto de sensores que permitan la reconstrucción de la base principal. A estos c sensores se los denominará *base secundaria*.
4. Cálculo de los modelos estáticos lineales que relacionan la base principal con la secundaria.

Previo a las etapas anteriores, es conveniente la realización de una normalización de los datos para eliminar posibles problemas de escalado en las medidas, por ejemplo una normalización *zscore*.

3.1. Determinación de la base principal

Dada la matriz M de datos históricos de medida, se realizará una descomposición en componentes principales (PCA). Esta metodología realiza una proyección ortonormal del espacio de datos a un espacio de dimensión inferior

$$M \approx W \cdot X$$

donde $W \in \mathcal{R}^{n \times r}$ es la matriz de pesos de las componentes principales y $X \in \mathcal{R}^{r \times m}$ es la matriz de los factores principales y r es el número de componentes principales. Para obtener la matriz W se realiza la descomposición en autovalores de la matriz de covarianza

$$C = \frac{1}{m} M \cdot M^t = Z_C T_C Z_C^t$$

donde $T_C \in \mathcal{R}^{m \times m}$ es una matriz diagonal que contiene los autovalores de C y $Z_C \in \mathcal{R}^{m \times m}$ contiene los autovectores de C . La matriz W corresponde a las r columnas asociadas a los autovalores de mayor módulo indicando cuantas columnas de la matriz M son independientes y conforman la base principal. Existen diferentes métodos para definir el valor de r . En este caso escogemos los r como el número de autovalores C con un módulo superior a la media.

Una vez determinada la dimensión de la base, es necesario escoger los sensores que la conforman. De nuevo, existen diferentes formas de resolver este problema de diseño. A continuación describimos una solución heurística basada en el análisis de los valores singulares de la matriz de históricos M :

- i Descomposición en valores singulares de la matriz $M = U^t D V$.
- ii Para cada sensor j se realizan los pasos del iii al vi.
- iii Se obtiene la matriz de datos M_j eliminando la columna j .

$$M_j^t = [v_1, \dots, v_k] \quad \forall k \neq j$$

- iv Descomposición en valores singulares de la matriz reducida M_j^r

$$M_j^r = U^t D_j^r V$$

- v Ampliamos el vector D_j^r con valores nulos hasta igualar la dimension de D .

$$D^{ra} = [D_j^r; 0]$$

- vi Calculamos el error entre D y D^{ra} .

$$E(j) = \|D^{ra} - D\|_2$$

- vii Se seleccionan como base los r nodos con mayor error.

Además de la base principal, es necesario calcular una base secundaria para detectar comportamientos erráticos en los nodos de la base. Para ello se realizará un algoritmo heurístico similar al anterior. En este caso, la matriz M inicial se reducirá eliminándose dos de las columnas asociadas a la base principal. Se obtendrá una nueva base de r elementos asociada a la columna no suprimida. Este proceso se realiza para cada nodo de la base principal. Se tendrán por tanto r bases secundarias.

3.2. Cálculo de los modelos estáticos

Una vez definidas de manera adecuada las correspondientes bases, se calculan los modelos estáticos lineales de los $m - r$ nodos de la red además de los correspondientes a los r elementos de la base. Debido a su sencillez y a la falta de restricciones, se utilizan mínimos cuadráticos para el calculo de los coeficientes de los filtros. El modelo se define de la manera siguiente:

$$\hat{v}_j(k) = \sum_{i=0}^r c_{ji} b_i(k) \quad \forall j$$

donde \hat{v}_j es la estimación de la medida del sensor j en el instante k , c_{ji} es el coeficiente que relaciona el sensor j con el componente i -ésimo de la base y $b_i(k)$ es su correspondiente medida. Este método da buenos resultados y permite obtener el modelo estático con un coste computacional mínimo ideal para su uso en tiempo real y con grandes bases de datos.

Para estimar los parámetros c_{ji} se utilizan mínimos cuadrados minimizándose el ajuste al histórico de datos disponible. Se minimiza por lo tanto la suma de los cuadrados de los residuos definido como

$$\sum_{j=0}^{m-r} (v_j - \sum_{i=0}^r c_{ji} b_i(k))$$

Para más detalles sobre identificación por mínimos cuadrados ver [10, 6, 9].

Definidos los componentes del modelo estático, la propuesta para la detección de comportamientos anómalos en los distintos sensores es la siguiente:

- i Se calcula el error absoluto máximo entre los datos reales v_j y los estimados por los filtros \hat{v}_j para un conjunto de datos de validación no incluidos en los datos de ajuste por mínimos cuadrados. Se determinan los umbrales como un 25 % superior a dicho error. Estos umbrales son configurables y se emplean para la fijar la precisión de la detección. Se calcularán tanto para la base como para el resto de sensores y siempre con valores no anómalos. Para la estimación de los sensores base se emplea la base secundaria.

$$Umbral_j = 1,25 \text{ máx } |v_j(t) - \hat{v}_j(t)| \quad t = 0, \dots, k$$

- ii Se calcula el error instantáneo $|v_j(k) - \hat{v}_j(k)|$ para el instante de tiempo k y para los r sensores o nodos base. Si alguno de ellos sobrepasa el umbral, se asumirá que dicho el sensor base en cuestión esta en situación anómala y no se chequean los $m - r$ sensores restantes.
- iii En caso de no haber error en los r sensores base, se calcula el error instantáneo $|v_j(k) - \hat{v}_j(k)|$ para el instante de tiempo k y para los $m - r$ sensores. Si alguno de ellos sobrepasa el umbral, se asumirá que dicho sensor esta en situación anómala.

4. Caso de estudio

Como caso de estudio de este procedimiento se van a utilizar datos reales de un sector de la red de distribución de aguas de Madrid. Este sector tiene alrededor de 1000 nodos pero dispone únicamente de 15 sensores de presión espacialmente distribuidos por todo el sector con distancias que pueden llegar hasta los dos kilómetros entre sensores.

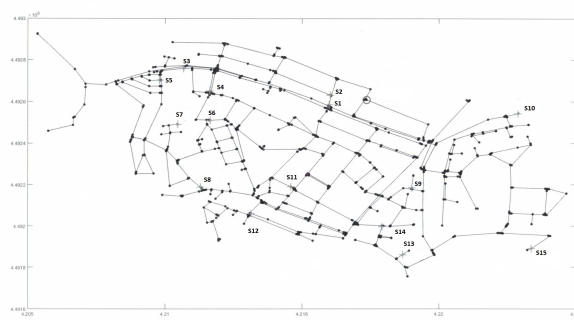


Figura 1: Mapa de la red de suministro de aguas considerada.

| Nodo | Error |
|-----------|---------------|
| 1 | 226,88 |
| 2 | 211,25 |
| 3 | 176,94 |
| 4 | 208,38 |
| 5 | 169,82 |
| 6 | 271,82 |
| 7 | 278,30 |
| 8 | 315.70 |
| 9 | 282,26 |
| 10 | 235,72 |
| 11 | 312,51 |
| 12 | 342.08 |
| 13 | 284,31 |
| 14 | 295,04 |
| 15 | 530 |

Tabla 1: Errores en el algoritmo de cálculo de la base principal. En negrita se marcan los 3 sensores de mayor error que conforman la base.

| Nodo | E. medio | E. máximo | Desv.std |
|------|----------|-----------|----------|
| 1 | -0,0032 | 0,6592 | 0,2307 |
| 2 | -0,0037 | 0,7053 | 0,2539 |
| 3 | -0,0041 | 0,7668 | 0,2838 |
| 4 | -0,0030 | 0,6469 | 0,2212 |
| 5 | -0,0036 | 0,6488 | 0,2620 |
| 6 | -0,0017 | 0,3826 | 0,1338 |
| 7 | -0,0002 | 0,2490 | 0,0907 |
| 9 | -0,0024 | 0,4570 | 0,1796 |
| 10 | -0,0040 | 0,8391 | 0,2811 |
| 11 | -0,0013 | 0,5579 | 0,1901 |
| 13 | -0,0027 | 1,3321 | 0,3383 |
| 14 | -0,0022 | 1,5367 | 0,4540 |

Tabla 2: Errores medio, máximo y desviación estándar del modelo estimador.

Para la realización y ajuste de los modelos se han utilizado 24 horas de evolución temporal (2 min por muestra). Adicionalmente se han utilizado 12 horas para validar la bondad de los modelos y otras 12 horas para realizar una prueba de la metodología propuesta de detección de comportamientos anómalos en los sensores. En total se han necesitado 48 horas de datos.

Tras aplicar el procedimiento definido en la Sección 3.1, se obtiene que la base la compondrán 3 sensores de los 15 totales, es decir, $r = 3$. Aplicando el algoritmo de dicha sección, se obtienen los errores expuestos en la Tabla 1.

Los errores más grandes los proporcionan los nodos [15, 12, 8] (ver Tabla 1) y por tanto serán los que conformen la base principal. Mediante el algoritmo expuesto en la Sección 3.1 se calcula también la base secundaria. En este caso estaría formada por los nodos [11, 13, 14] para cualquiera de los tres sensores base.

Definidas la bases (principal y secundarias) se calculan los filtros estáticos para los $m - r$ sensores restantes y para la base principal (Sección 3.2).

El error medio, máximo y varianza de las estimaciones con los filtros están expuestos en la Tabla 2 para los $m - r$ sensores y en la Tabla 3 para los r sensores que conforman la base. Puede observarse que el error de estimación máximo lo presentan los sensores 14 y 13 con aproximadamente un metro y medio de error. Como ejemplo gráfico de la estimación conseguida, podemos ver la Figura 2 que muestra el ajuste conseguido en el caso del sensor 14. En rojo se representan los datos reales y en azul intermitente los estimados.

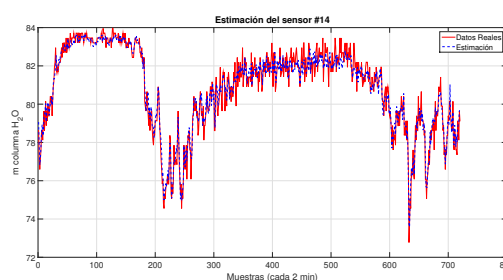


Figura 2: Estimación del sensor 14 respecto a los datos reales.

Para comprobar la detección de errores en los comportamientos de los sensores se han estudiado tres escenarios:

- i Escenario 1: Sensor que a partir de un determinado instante k pasa a dar un valor constante de presión coincidente con el último valor correcto.
- ii Escenario 2: Sensor que a partir del instante k cambia su comportamiento. En este caso, dicho sensor pasa a devolver los datos de otro sensor en vez de los suyos.
- iii Escenario 3: Sensor de la base principal que a partir del instante k cambia su comportamiento. Escenario similar al segundo pero aplicado a los sensores base.

En el primer escenario, el sensor 5 deja de funcionar correctamente a partir del instante $k = 460$. El resultado de esta simulación puede verse en la Figura 3. En color azul intermitente se representa la estimación obtenida por los filtros, en amarillo continuo se representan los datos reales y en rojo grueso los puntos o muestras catalogados como erróneos. Puede observarse que en este escenario su margen de acierto es de 54,78 % puesto que el sistema no detecta el error mientras la presión se mantiene en torno al valor que devuelve el sensor defectuoso.

En el segundo escenario los resultados obtenidos pueden observarse en la Figura 4. Los colores de la curva

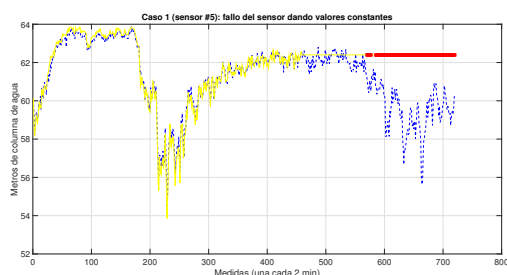


Figura 3: Detección de anomalía en el sensor 5 (escenario 1): estimación (azul), dato real(amarillo), puntos anómalos detectados(rojo)

| Nodo | E. medio | E. máximo | Desv.std |
|------|----------|-----------|----------|
| 15 | 0,225 | 2,228 | 0,624 |
| 12 | 0,045 | 0,644 | 0,197 |
| 8 | 0,025 | 1,506 | 0,404 |

Tabla 3: Errores medio, máximo y desviación estándar del filtro estático estimador de las base.

se mantienen igual que en gráfica anterior. En este caso el porcentaje de acierto es del 100 % ya que el comportamiento del sensor cambia de manera drástica.

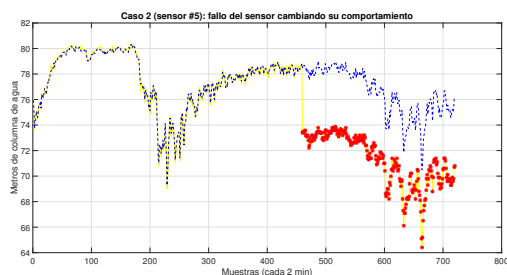


Figura 4: Detección de anomalía en el sensor 5 (escenario 2): estimación (azul), dato real(amarillo), puntos anómalos detectados(rojo)

En el escenario 3 se presenta un posible fallo en una de las bases (sensor 12) por lo que se utiliza la base secundaria. El rendimiento en este caso también sería del 100 %.

5. Conclusiones

Este artículo demuestra como a través de la elaboración de una serie de modelos estáticos lineales que relacionan los datos de un conjunto de sensores con un subconjunto de ellos puede conseguirse localizar de manera bastante eficiente y sencilla comportamientos anómalos en redes de sensores de presión pertenecientes a una red de distribución de agua. Este tipo de metodologías resultan de gran interés en redes con grandes cantidades de datos donde realizar estimaciones puede llegar a ser bastante complejo.

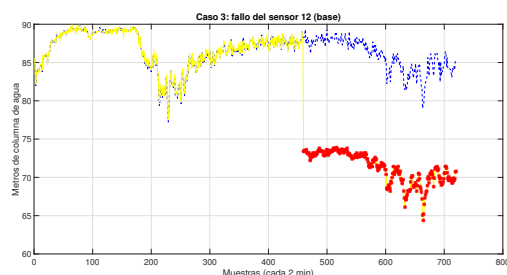


Figura 5: Detección de anomalía en el sensor 12 (escenario 3): estimación (azul), dato real(amarillo), puntos anómalos detectados(rojo)

Agradecimientos

Este trabajo ha sido financiado por los proyectos DPI2016-76493-C3-2-R y DPI2016-76493-C3-3-R del MINECO (España).

English summary

FAILURE DETECTION ON WATER DRINKING NETWORKS BASED ON LINEAR STATIC FILTERS AND PCA

Abstract

This paper is focused on the design of a set of static and linear models which relate the data of a subset of pressure sensors in a drinking water network with the rest of them. This subset of sensors is called principal base. Thanks to these models wrong behaviours in these devices can be detected comparing real data from sensors with estimated data obtained from these models. In case of fault in a sensor belonging to the principal base, we propose the use of another base, different from the previous one, called secondary base. This second set is needed to detect anomalies in the first one. The proposed methodology was applied to a section of the water drinking network of the city of Madrid.

Keywords: Failure detection, water drinking networks modelling and simulation, principal component based algorithm.

Referencias

[1] H. Al-Barqawi and T. Zayed. Condition rating model for underground infrastructure sustainable

- water mains. *Journal of Performance of Constructed Facilities*, 20(2):126–135, 2006.
- [2] M. Cattani, C. A. Boano, D. Steffelbauer, S. Kaltenbacher, M. Günther, K. Römer, D. Fuchs-Hanusch, and M. Horn. Adige: an efficient smart water network based on long-range wireless technology. In *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*, pages 3–6. ACM, 2017.
- [3] S. Costello, D. Chapman, C. Rogers, and N. Metje. Underground asset location and condition assessment technologies. *Tunnelling and Underground Space Technology*, 22(5-6):524–542, 2007.
- [4] M. M. Gamboa-Medina and L. F. R. Reis. Sampling design for leak detection in water distribution networks. *Procedia Engineering*, 186:460–469, 2017. XVIII International Conference on Water Distribution Systems, WDSA2016.
- [5] O. Ghorbel, A. Ayadi, R. Ayadi, M. Aseeri, and M. Abid. Combined methods based outlier detection for water pipeline in wireless sensor networks. In *International Conference on Advanced Information Networking and Applications*, pages 214–224. Springer, 2019.
- [6] S. Kay. *Fundamentals of Statistical Signal Processing: Detection theory*. Fundamentals of Statistical Signal Processing. PTR Prentice-Hall, 1993.
- [7] Z. Liu and Y. Kleiner. State-of-the-art review of technologies for pipe structural health monitoring. *IEEE Sensors Journal*, 12(6):1987–1992, 2012.
- [8] Z. Liu and Y. Kleiner. State of the art review of inspection technologies for condition assessment of water pipes. *Measurement*, 46(1):1–15, 2013.
- [9] L. Ljung. *System Identification: Theory for the User*. Prentice Hall information and system sciences series. Prentice Hall PTR, 1999.
- [10] D. Luenberger. *Optimization by Vector Space Methods*. Professional Series. Wiley, 1997.
- [11] J. Makar and Y. Kleiner. Maintaining water pipeline integrity. In *AWWA Infrastructure Conference and Exhibition*, volume 12, pages 1–13, 2000.
- [12] D. Misiūnas. *Failure monitoring and asset condition assessment in water supply systems*. Vilniaus Gedimino technikos universitetas, 2008.
- [13] V. Puig, C. Ocampo-Martinez, R. Prez, G. Cembrano, J. Quevedo, and T. Escobet. *Real-time Monitoring and Operational Control of Drinking-Water Systems*. Springer Publishing Company, Incorporated, 1st edition, 2017.
- [14] B. Rajani and Y. Kleiner. Non-destructive inspection techniques to determine structural distress indicators in water mains. *Evaluation and Control of Water Loss in Urban Water Networks*, pages 21–25, 2004.
- [15] A. M. Sadeghioon, N. Metje, D. Chapman, and C. Anthony. Water pipeline failure detection using distributed relative pressure and temperature measurements and anomaly detection algorithms. *Urban Water Journal*, 15(4):287–295, 2018.
- [16] S. K. Sinha and M. A. Knight. Intelligent system for condition monitoring of underground pipelines. *Computer-Aided Civil and Infrastructure Engineering*, 19(1):42–53, 2004.
- [17] I. Stoianov, L. Nachman, S. Madden, and T. Tokmouline. Pipeneta wireless sensor network for pipeline monitoring. In *Proceedings of the 6th international conference on Information processing in sensor networks*, pages 264–273. ACM, 2007.
- [18] P. Tayefi, S. B. Beck, and R. A. Tomlinson. Fatigue failure of polyethylene electrofusion joints subject to contamination. In *Fracture, Fatigue, Failure, and Damage Evolution, Volume 5*, pages 197–202. Springer, 2015.
- [19] T. Tsung-Te Lai, W.-J. Chen, K.-H. Li, P. Huang, and H.-H. Chu. Triopusnet: automating wireless sensor network deployment and replacement in pipeline monitoring. In *2012 ACM/IEEE 11th International Conference on Information Processing in Sensor Networks (IPSN)*, pages 61–71. IEEE, 2012.
- [20] J. S. Webber, J. R. Covey, and M. V. King. Asbestos in drinking water supplied through grossly deteriorated a-c pipe. *Journal-American Water Works Association*, 81(2):80–85, 1989.
- [21] A. J. Whittle, M. Allen, A. Preis, and M. Iqbal. Sensor networks for monitoring and control of water distribution systems. 2013.



© 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution CC BY-NC-SA 4.0 license (<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>).