doi: https://doi.org/10.4995/WDSA-CCWI2022.2022.14833

WATER DISTRIBUTION NETWORK DISRUPTIVE EVENTS. GENERATION AND EXPLOITATION OF AN INCIDENT HUB TO INCREASE THE NETWORK PREPAREDNESS

David Ayala-Cabrera[1], Jorge Francés-Chust[2], Sotudeh Hoseini-Ghafari[3], Gemma Stanton[4] and Joaquín Izquierdo[5]

[1,3,4]CWRR-School of Civil Engineering, University College Dublin, Dublin (Ireland)
[2]Business Development, Aguas de Bixquert, S.L., Xàtiva, Valencia (Spain)
[5]FluIng-IMM, Universitat Politècnica de València (Spain)

[1] *david.ayala-cabrera@ucd.ie*, [2] *jorge@abxat.com*, [3] *sotudeh.hoseinighafari@ucdconnect.ie*, [4] *gemma.stanton@ucdconnect.ie*, [5] *jizquier@upv.es.*

Abstract

This paper seeks to develop increased knowledge about disruptive events in a water distribution network (WDN) through the experience acquired by previous anomalous events in the system. This work explores the various relationships between several parameters in an incident hub (specifically water loss events) in a Spanish real, small WDN. The incident hub consists of basic elements recorded during an incident (e.g. breakdown, maintenance activity, among others) and the corresponding causes that generated the incident (e.g. breakage due to excess pressure, breakage due to tree roots, etc.), as well as the management times of the incident (e.g. awareness time, isolation, and repair time). The utility collected and stored these data, which were completed with direct interviews with the system operator. Measurements were performed at pressure and flow sensors, which allowed evaluating the effects of both the incident itself and the actions taken to solve it. The records of the incidents are categorised depending on the nature of the data they contain to facilitate mapping their causes and effects. To characterise the disruptive events, a feature extraction process has been proposed using a temporal-spatial approach combined with a migration proposed that describes parameters' behaviour in the spatial dimension for a certain period of time. The characteristics obtained in the previous lessons of the incidents contained in the incident hub are compared with potential causes obtained with different control parameters. The objective is to determine the potential causal relationship of the incident that allows its characterisation. The results of this characterisation are presented and analysed in this contribution. The outcomes are promising in the sense of a clear ability to provide WDNs with key parameters that foster prediction and classification processes.

Keywords
Hub of incidents, Water distribution networks, Spatial-temporal analysis, Network preparedness, Resilience, Protection of critical infrastructures, Intelligent data analysis.

1 INTRODUCTION

Water distribution networks (WDNs) are critical infrastructures which are exposed to multiple challenges that can stop guaranteeing their correct operation concerning the satisfaction of their basic objectives, that is, ensuring the provision of an adequate quantity of water demanded, the continuity of the service and its water quality. These systems need to be prepared both to face these different challenges and to avoid disruptions or reduce their potential impacts. Utilities are currently focusing on increasing the systems' preparedness as this is seen as a powerful tool that can help them deal with these challenges. The preparedness of the system has normally been used in the context of emergencies. However, events not considered emergencies (or rarely considered as such) can cause situations that can cause or increase possible negative effects on the network.

The resilience of infrastructures normally refers to their capabilities to face certain threats. In this sense, these capabilities can be referred to as absorptive, adaptive and restorative. The absorptive capacity refers to the ability of the system to deal with abnormal events without any intervention other than its capabilities. The adaptive and restorative capabilities are more related to the interventions deployed to mitigate the effects of the anomalous event (adaptive capacity) and to fix it from a long-term perspective (restorative capacity). However, there is an additional ability of the system that many authors refer to (*e.g*. see [7]), which is the ability to learn from the events that have occurred. The events that have occurred can provide the system with information to not only face events with similar characteristics but also to face new potential challenges.

A data Hub is seen in this paper as a structured data set/digital repository (collections of data) that allows quick, digital access to information among users that may be relevant for a certain use, e.g. [1]. A data Hub with an architecture designed to centralise incident data (incident Hub) can provide the system with information on how to avoid/prepare for; how to respond to; and define a series of actions, all of these to better deal with incidents of similar characteristics and also avoid/minimise the impacts of other events with greater relevance. For example, in the field of cybersecurity, there are incident Hubs (public or private) that cover aspects such as the incident response process to improve it in the cloud service [2], cyber-attack reports on critical infrastructure [3], or offering a coordinated response to a variety of cyber security issues [4]. To increase the capacity of the system to learn about previous events, this paper presents a data Hub created from incidents in a real WDN in Spain. In addition, this paper proposes a method to extract the acquired knowledge contained in the Hub. This information can be used to establish maintenance models [5] that can minimise the impact or prevent the incident from occurring. The relationships and dependencies existing in the spatial dimension of the control parameters are contrasted with the patterns obtained from the Incident Hub to characterise the system. This ultimately increases the system resilience derived from the obtained increase in the network preparedness to cope with such potential events [6]. This characterisation also favours the elaboration of good practices in WDNs that can help in two crucial aspects: the sustainability of the system through the preservation of resources, and the mitigation of impacts due to extreme events.

2 PROPOSED FRAMEWORK

This section presents the proposed framework to increase network preparedness in WDNs through the knowledge gained from previous incidents that occurred in the system (Figure 1). This is through the generation and potential exploitation of an Incident Hub. In the proposed framework in this paper, for the incident Hub architecture, two types of cores are distinguished: i) causes/times core, and ii) effects on the system core. These cores are, initially, associated with three types of incidents: 1) water leaks, 2) maintenance or operational activity, and 3) additional demand of water. Details about the creation and categorisation of the components of the incident Hub are described in Section 3.

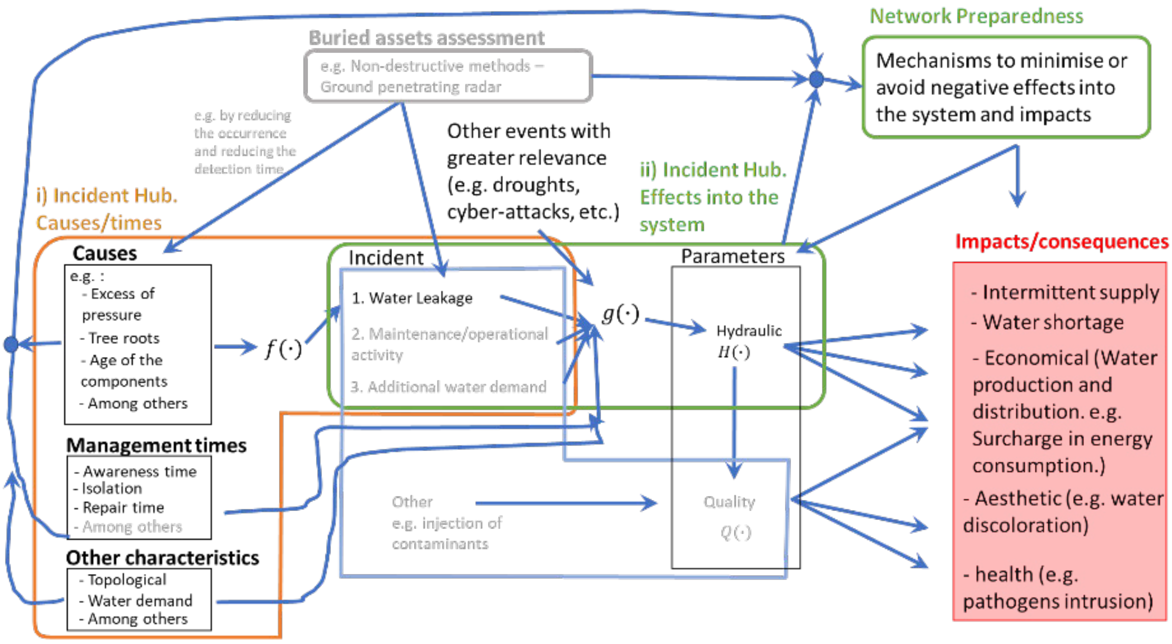


Figure 1. Incident Hub. Proposed framework to increase system preparedness through knowledge gained from previous incidents

3 INCIDENT HUB. GENERATION

The network under study corresponds to a real small WDN located in Spain. Let us consider a WDN as a graph $\boldsymbol{G} = \{\boldsymbol{V}, \boldsymbol{E}\}$, where $\boldsymbol{V}$ is the set of vertices/nodes (*i.e.* tanks, reservoirs, junction nodes) with a total of $\boldsymbol{n_n}$ nodes, and $\boldsymbol{E}$ is the set of edges/links that join these nodes (*i.e.* pipes, valves, pumps) with a total of $\boldsymbol{n_l}$ links. In this case, the network under study is represented by $\boldsymbol{n_n} = \mathbf{188}$ nodes, *i.e.* 4 tanks, 2 reservoirs, and 182 junction nodes; and $\boldsymbol{n_l} = \mathbf{230}$ links, *i.e.* 212 pipes, 16 valves (it should be mentioned that only a few isolation valves whose states are closed were considered in this initial representation of the network), and 2 pumps. It is worth mentioning that the network corresponds to a schematic configuration and that its spatial coordinates do not correspond precisely to the real georeferencing.

Incident Hub. Causes/times core. To obtain relevant data about the incidents that affected in some manner the operation of the WDN (e.g. breakdown, operational or maintenance activities, water demands by users external to the network under study), a series of interviews with the operator was conducted. In the first interviews, a template was generated where information on basic aspects of the incidents was collected. The template was completed and is currently populated in an online manner by the system operator as incidents occur. Afterwards, the collected data is completed through regular interviews. The evaluation period in this paper corresponds from July 2021 to May 2022.

The aspects collected in this core of the Hub can be categorised as 1) the causes that led to the anomaly in the system, and 2) the management times of the incident. In this paper, the causes can be seen as a function $\boldsymbol{f}(\cdot)$ that characterises those factors, and the relationships among them, that can lead to a water leak incident. The main interest of this paper is to advance in the identification of these potential relationships. The effects of anomalous events can eventually be reflected, in the first instance, in hydraulic and water quality parameters, such as $\boldsymbol{H}(\cdot)$ and $\boldsymbol{Q}(\cdot)$. However, the effects of these parameters are the result of the contribution of multiple factors $\boldsymbol{g}(\cdot)$ (occurring simultaneously or sequentially). This paper considers those management elements as factors whose contribution can avoid, increase or minimise the negative effects that can be reflected in the hydraulic and quality parameters of the system (see Figure 1).

Other aspects in this core, such as topological characteristics of the system, and demand (from the own system or due to sales of water to other neighbour utilities), were located and were also collected in this Hub. They were considered as a contributor to $\boldsymbol{g}(\cdot)$. Topological aspects of the network were compiled into the hydraulic model of the network. Considering that the hydraulic model is a simplification of the real model, the incidents were allocated in the hydraulic model at the point that best fit their real position (according to the operator criteria).

Functions $\boldsymbol{f}(\cdot)$ and $\boldsymbol{g}(\cdot)$ jointly characterise the system in terms of a specific anomalous event and can be fed from previous experiences. This will increase the capacity of the WDN to learn and consequently avoid/minimise impacts on the system itself or other related ones.

Incident Hub. Effects on the system. Testing resilience is an effective manner to prepare the system to deal with events that may affect its proper operation. This preparation can be in terms of avoiding or minimising the negative effects that these anomalies can potentially cause. Testing resilience through simulated events, although an effective tool, has considerable limitations in terms of the selection of the event to be simulated, the number of parameters to be assessed, and in terms of the difficulty to incorporate the coexistence of more than one event simultaneously. This adds to the difficulty of obtaining a hydraulic model that can adequately capture the behaviour of the system.

The effects of an anomalous event on the system can be estimated through relationships established with its basic parameters (*i.e.* hydraulic and quality) and system characteristics (*e.g.* topological configuration). The valuable information provided by these stress tests on the system can, however, be complemented by extracting the characteristics of the effects from the basic parameters (*e.g.* measured through sensors) in events that have already occurred. In this manner, it is possible to consider a variety of other factors that are not easily incorporated into the simulation or that can normally be ignored. In this line, this work proposes the incorporation of a core for the Incident Hub that contains the effects of the anomalous event in terms of pressure and flow measured through the sensors available in the network during the period of study.

4 CATEGORISATION OF THE INCIDENTS BASED ON EXPERT KNOWLEDGE

The total number of incidents recorded during the evaluation period was 70 cases distributed in 56 water leak cases, 11 reported cases of maintenance or operational activities (it should be mentioned that not all maintenance operations were recorded), and three cases of additional water consumption events (water sold to another WDN, this last case still in progress).

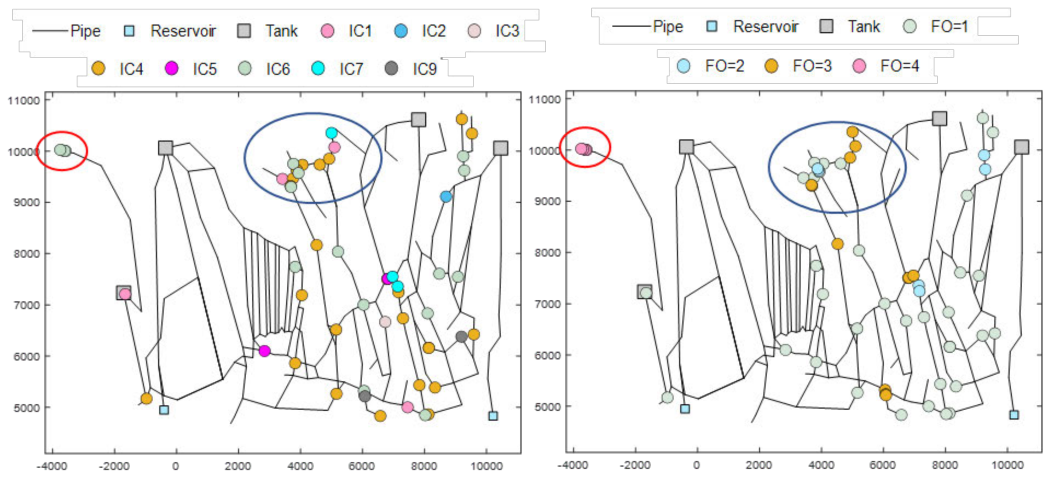
Based on the causes of each event provided by the utility's operator experience, the incidents observed during the evaluation period were classified into eleven potential incident categories (IC). The eleven selected categories were:

- **IC1. Settlement.** (4 cases) This category corresponds, among others, to the differential settlement of walls or the ground due to heavy vehicle traffic.
- **IC2. Asset fissure/crack**. (1 case) This category was assigned when no other breakage category was attributable. According to the experience of the operator, this type of case has been presented before.
- **IC3. Tree roots.** (2 cases).
- **IC4. Pressure inference.** (26 cases) This category includes all those labels provided by the operator that included pressure exclusively (high pressure, 14 cases; and pressure oscillation, 1 case), and also those that included one or more other causes besides pressure. Other causes besides pressure include 1) 2 cases - settlement (land settlement, heavy vehicles traffic); 2) 2 cases - Age of the component; 3) 5 cases - defective component;

4) 1 case - tree roots and material wear; and 5) 1 case - inappropriate pipe (low-density material).

- **IC5. Age/corrosion.** (2 cases).
- **IC6. Pipe defects**. (17 cases). This category includes inappropriate pipe wall thickness, contraction and expansion of the pipe, and poor quality of the component, among others.
- **IC7. Exposed components.** (2 cases). This category refers to incidents occurring in components of the network that are exposed to the surface.
- **IC8. Maintenance activity.** (10 cases).
- **IC9. Weather conditions/component frozen.** (2 cases).
- **IC10. Additional consumption.** (3 cases).
- **IC11. Others.** (1 case).

The spatial location of the categories that correspond to incidents that resulted in water leaks (IC1-IC7, and IC9) is presented in Figure 2a. In addition, the quantification of the frequency of occurrence (FO) of water leaks in a particular pipe is presented in Figure 2b.



a) b)

Figure 2. Incident Hub. a) Incident categories (IC), and b) frequencies of occurrence (FO) of a leak in the same pipe.

Figure 2 shows in its two insets a marked tendency of a particular zone to present water leaks (blue circle). This particular area presents diversity in incident categories (see Figure 2a) and in particular shows a high concentration of the category associated with pressure (IC4). Similarly, this zone also shows a tendency to leak water in at least two of its pipes (Figure 2b). An area that does not present diversity in terms of categories of incidents and has the highest frequency of occurrence is found at the end of the network (red circle). Other zones that present an FO of three and high diversity can be observed in Figure 2b. In general, the network shows a tendency to leak due to pressure and has its occurrences concentrated in specific pipes.

5 EXTRACTION OF KNOWLEDGE FROM PREVIOUS LESSONS - CAUSES

This section proposes a method for the extraction of knowledge from previous leakage events in WDNs and in particular the cause element of the incident Hub, *i.e.*, to increase the knowledge in

$\boldsymbol{f}(\cdot)$. To this end, this section proposes the spatial analysis of the data collected through the use of a data migration process [7]. Two types of data are distinguished at this point, 1) data referring to the knowledge generated by the event already occurred (previous lessons), and 2) parameter data (potential cause - not exclusively hydraulic), whose potential contribution to the event under study is to be identified. On the part of the information provided by the previous events, in this document, the results obtained with the frequency of occurrence of the leak event during the sampling period are used in the first instance to conduct the migration process. Regarding the parameters to be analysed, in particular pressure, this paper proposes a clustering process based on pressure [8] as a process prior to the migration process.

MatLab's `griddata` function was used in this document to represent sparse data in a spatially homogeneous manner by constructing it in a grid. The resulting matrix has dimension $\boldsymbol{m} \times \boldsymbol{n}$. This matrix covers the maximum and minimum spatial coordinates of all vertices, $\boldsymbol{n_n}$, (network nodes). Let us denote the resulting spatial matrices as $\boldsymbol{D}_{\mathrm{pl}}$ and $\boldsymbol{D}_{\mathrm{pc}}$, for the previous lessons and from the potential cause; respectively.

Migration process. To identify the spatial relationships among the elements, this section proposes the condensation of information through its migration. This is an iterative process (based on [7]) and essentially consists of the transmission of information to a central cell, from its neighbours for a given matrix $\boldsymbol{D_p}$ ($\boldsymbol{D}_{\mathrm{pl}}$or $\boldsymbol{D}_{\mathrm{pc}}$) as $\boldsymbol{M}_{\mathrm{p},\boldsymbol{i},\boldsymbol{j}} = \left(\left(\sum_{\boldsymbol{k}=\boldsymbol{1}}^{\boldsymbol{k}=\boldsymbol{4}} \boldsymbol{D}_{\mathrm{p},\boldsymbol{cell_k}}\right) + \boldsymbol{D}_{\mathrm{p},\boldsymbol{i},\boldsymbol{j}}\right)$. A diamond-shaped configuration was proposed to transmit information from neighbours, *i.e.* $\boldsymbol{cell} = \{(\boldsymbol{i}+\boldsymbol{1},\boldsymbol{j}), (\boldsymbol{i},\boldsymbol{j}-\boldsymbol{1}), (\boldsymbol{i},\boldsymbol{j}+\boldsymbol{1}), (\boldsymbol{i}-\boldsymbol{1},\boldsymbol{j})\}$ with $\boldsymbol{k} = \{\boldsymbol{1}, \dots, \boldsymbol{4}\}$ in this paper. $\boldsymbol{D}_{\mathrm{p}}$ is updated with the information from the $\boldsymbol{M}_{\mathrm{p}}$ of the previous iteration (after $\boldsymbol{M}_{\mathbf{p}}$ has been normalised) and thus the coverage of the transferred information is increased. In this paper, three numbers of iterations were explored: 1) number of iterations when the correlation coefficient between $\boldsymbol{D}_{\mathrm{p}}$ and $\boldsymbol{M}_{\mathrm{p}}$ starts to be constant, 2) number of iterations equal to $\boldsymbol{n}$, and 3) number of iterations equal to $\boldsymbol{m}$.

This iterative process transmits information from cells with lower values to cells with higher values. To evaluate which space best characterises the event, three variants were evaluated, 1) "to-max"; it corresponds to the matrix $\boldsymbol{D}_{\mathrm{p}}$ preserved in its initial condition, 2) "to-min"; it consists of the variation of $\boldsymbol{D}_{\mathrm{p}}$ towards minima (*i.e.* $(-(\boldsymbol{D}_{\mathrm{p}} - \max(\boldsymbol{D}_{\mathrm{p}}))) + \min(\boldsymbol{D}_{\mathrm{p}})$), and 3) "to-both", which consists of the difference for $\boldsymbol{M}_{\mathrm{p}}$ resulting from variants 1 and 2.

Previous lessons and potential causes, migration.

Figure 3 presents the migration process for the matrix of previous lessons corresponding to spatially located water leaks and their corresponding frequencies of occurrence in a particular pipeline. Variants of the migration space and also the number of iterations (50, 142 and 308) were selected, looking if any of them could represent the past event from which information is to be extracted. Considering that the WDN model is both a skeletonized and schematic configuration of the system, it can be assumed that in the analysis it is necessary to incorporate some uncertainty regarding the spatial location of the event. For this reason, in this work, it is considered that the "to-max" configuration can better represent the evaluated period when compared with "to-min" and "to-both" for the matrix of previous lessons.

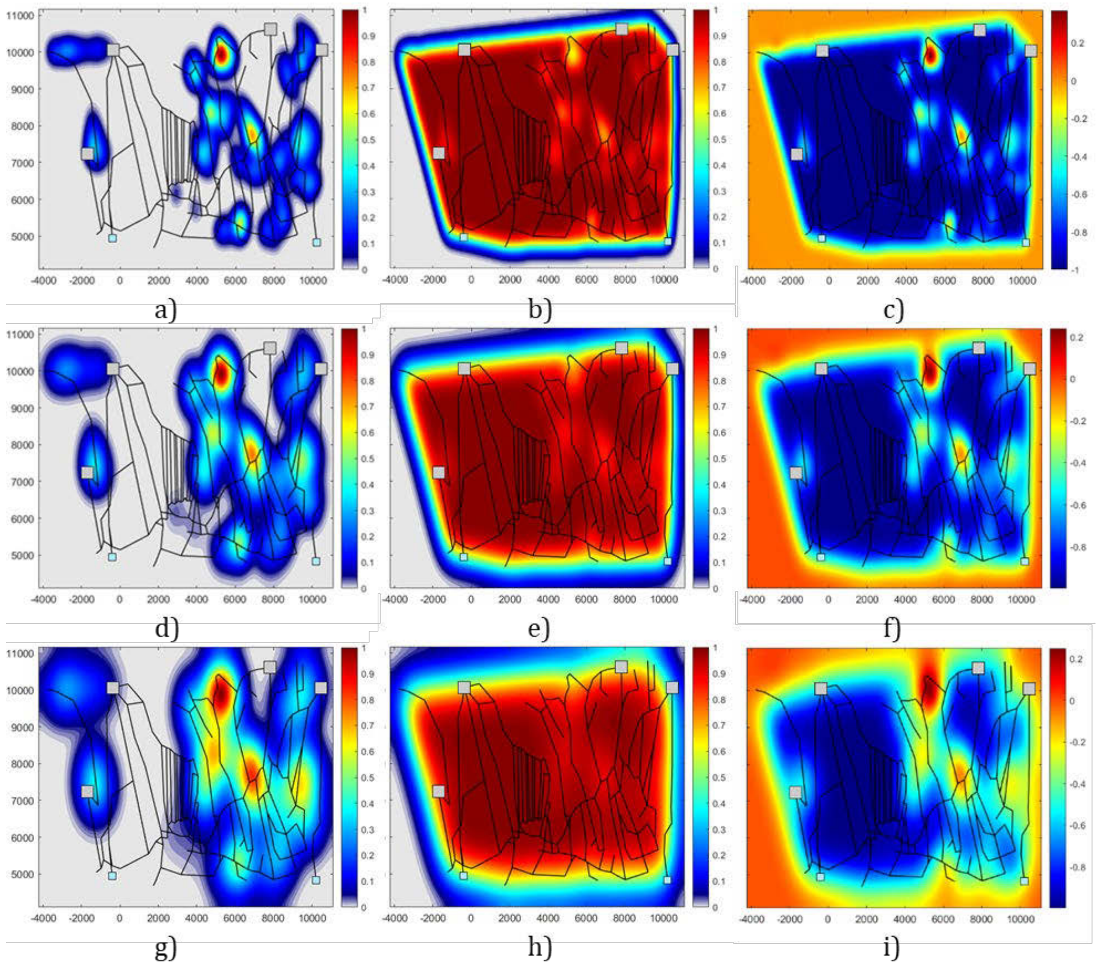


Figure 3. Previous lessons. Incidents of water leaks and frequency of their occurrences - migration. Variant: (a, d, g) to-max; (b, e, h) to-min; (c, f, i) to-both; (a-c) 50 iterations; (d-f) 142 iterations; (g-i) 308 iterations

For the analysis of the matrix of potential causes, a simulation of the hydraulic model of an arbitrary day was conducted. Each of the pressures obtained at each node was classified into clusters according to the percentage of time that each node spent in a particular range of pressures (as in [8]). The results of the conglomerate as a function of the pressure and the percentage of time that each node spent in this pressure cluster obtained are presented in Figure ***4***.

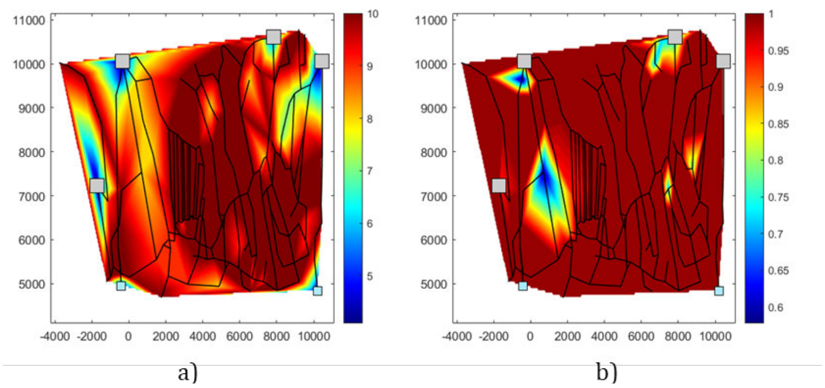


Figure 4. Potential Causes. Pressure-based clusters. a) Pressure-based cluster and b) percentage of time spent in a specific cluster

For the matrix of potential causes (pressure-based clusters), the results in the variants of the proposed migration spaces are presented in Figure 5.

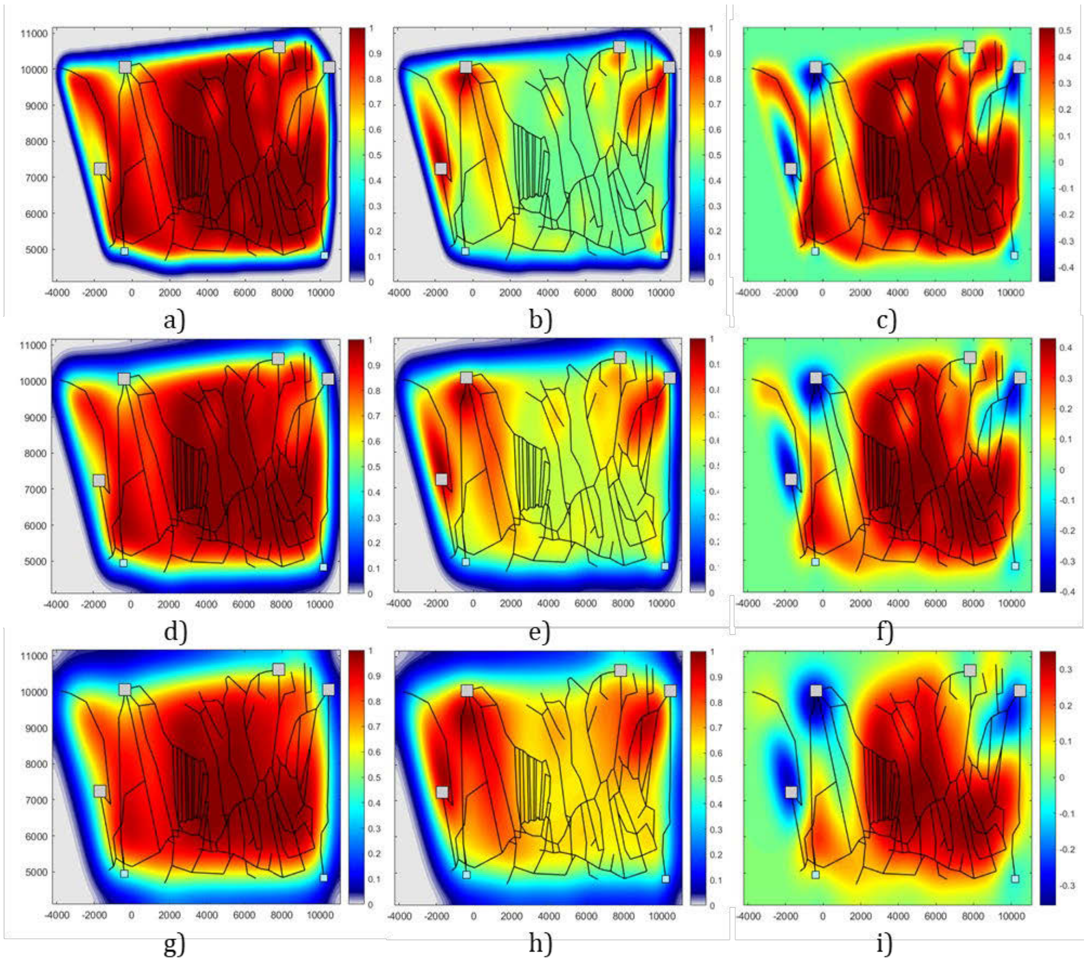


Figure 5. Potential causes. Pressure-based clusters - migration. Variant: (a, d, g) to-max; (b, e, h) to-min; (c, f, i) to-both; (a-c) 50 iterations; (d-f) 142 iterations; (g-i) 308 iterations

The results of going through the migration process for both the pressure-based cluster matrix and the percentage of time spent in a specific cluster were compared with the matrix results from previous lessons. These results are presented in Table 1 and Table 2 respectively.

Table 1. Correlation coefficients for previous lessons and pressure-based cluster

Previous lessons	Iter. 50			Iter. 142			Iter. 308		
Potential causes	Iter. 50	Iter. 142	Iter. 308	Iter. 50	Iter. 142	Iter. 308	Iter. 50	Iter. 142	Iter. 308
to-max	0.978	0.968	0.944	0.409	0.410	0.405	0.459	0.467	0.471
to-both	0.673	0.558	0.412	0.431	0.428	0.412	0.498	0.505	0.496
to-min	0.945	0.949	0.936	0.269	0.272	0.266	0.290	0.298	0.300

Table 2. Correlation coefficients for previous lessons and percentage of time spent in a specific cluster

Previous lessons	Iter. 50			Iter. 142			Iter. 308		
Potential causes	Iter. 50	Iter. 142	Iter. 308	Iter. 50	Iter. 142	Iter. 308	Iter. 50	Iter. 142	Iter. 308
to-max	0.991	0.984	0.963	0.389	0.392	0.388	0.433	0.443	0.449
to-both	0.841	0.803	0.727	0.425	0.444	0.467	0.480	0.509	0.546
to-min	0.988	0.980	0.956	0.336	0.336	0.328	0.371	0.377	0.377

Both parameters showed high correlation coefficients indicating that both the pressure and the frequency of time spent under a given pressure have a high contribution to the occurrence of the events evaluated. Likewise, correlations as high as those obtained for "to-max" with iter. 50 for the potential causes indicate that both high pressure and high times have a very big contribution in this particular network, even though the final cause labelled for the expert was a different one in some cases. Finally, both potential matrices can represent these events in a very approximate way for this system.

6 CONCLUSIONS

In this work, the generation of an Incident Hub and potential uses to increase the network preparedness and consequently its resilience has been presented. This paper presents the different causes collected in this Incident Hub. The results show that the network is prone to water leaks essentially due to high pressure in the system. In addition, this work proposes a method to extract knowledge from previous events. It is also shown how it is possible to transfer knowledge from previous events. The results are promising since the proposed methodology can be used for evaluating the contribution to the events of other parameters, as well as they can serve as bases for the training of intelligent systems.

7 REFERENCES

[1] (DRI), “Digital Repository Ireland,” https://www.dri.ie/, (accessed on: Apr. 2022).

[2] SITECORE “Functional documentation,” https://docs.stylelabs.com/, (accessed on: Apr. 2022).

[3] TISAFE “Incident Hub – Industrial Cybersecurity incidents Database,” https://hub.tisafe.com/, (accessed on: Apr. 2022).

[4] CIH “Cyber Incident Hub,” https://www.cyber-incident-hub.ch/, (accessed on: Apr. 2022).

[5] B. G. Mwanza and C. Mbohwa “The Impact of Maintenance Systems on Water Supply: A Case Study of a Utility Company,” in Proc. IEOM2016. Kuala Lumpur, Malaysia. (2016).

[6] D. Ayala-Cabrera, O. Piller, J. Deuerlein, M. Herrera, “Key Performance Indicators to Enhance Water Distribution Network Resilience in Three-Stages”. Water Utility Journal, vol.19, 2018, pp. 79–90.

[7] S. Hoseini-Ghafari, J. Francés-Chust, O. Piller, and D. Ayala-Cabrera, “Exploring a Spatial Dynamic Approach and Landmark Detection for Leakage/Burst event Characterisation in Water Distribution Networks,” in Proc. International Environmental Modelling and Software Society (iEMSs), Brussels, Belgium, 2022. Accepted.

[8] D. Ayala-Cabrera, M. Castro-Gama, C. Quintiliani, S. Hoseini-Ghafari, “Multi-step optimization approach to rehabilitate a deteriorated and intermittent water distribution network,” in Proc. 2nd International Joint Conference on Water Distribution System Analysis & Computing and Control in the Water Industry (WDSA/CCWI) - "The Battle of the Intermittent Water Supply", Valencia, Spain. Accepted.

[9] E. Lawson, S. Bunney, S. Cotterill, R. Farmani, P. Melville-Shreeve, D. Butler, "COVID-19 and the UK water sector: Exploring organizational responses through a resilience framework," Water and Environment Journal, vol. 36, no. 1, 2022, pp. 161-171.