

Impact of water-quality modelling simplifications on the optimal positioning of sensors in looped water distribution networks

Stefania Piazza (✉ stefania.piazza@unikore.it)

University of Enna Kore: Università degli Studi di Enna 'Kore' <https://orcid.org/0000-0002-4528-4727>

Mariacrocetta Sambito

University of Enna Kore: Università degli Studi di Enna 'Kore'

Gabriele Freni

University of Enna Kore: Università degli Studi di Enna 'Kore'

Research Article

Keywords: Advection, Dispersion, Random walk model, Water quality, Water distribution network

Posted Date: April 6th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1439022/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

To develop an efficient monitoring system for looped piped water systems, an adequate model for evaluating water quality is necessary. Several literature studies have used advective-reactive models to analyse water quality, neglecting diffusive transport, which is claimed to be irrelevant in turbulent flows. Although this may be true in simple systems such as linear transport pipes, the presence of laminar flows in looped systems may be significant, especially at night and in the peripheral parts of the network. The present paper compares the current EPANET advective model, an updated version of EPANET in which diffusion-dispersion equations were implemented and a newly developed diffusive-dispersive model under dynamic conditions that uses the random walk method. The results demonstrated the inadequacy of the advective model in reproducing data from the experimental campaign. Additionally, a significant difference in sensor positioning was found based on the numerical model used.

Highlights

- Distribution network water quality models may be unreliable at low flow.
- Backward dispersion processes are significant in laminar and transitional flow.
- The use of advection-dispersion equations better explains network contamination.
- Contamination detection efficiency is improved by the use of complete equations.

Introduction

In recent years, the problem of monitoring water quality in urban water systems has been the subject of numerous scientific studies ((Aisopou, et al., 2012) - (Li, et al., 2017)). Such studies typically integrate monitoring and modelling to obtain a reliable picture of the systems, reducing the costs of real-time monitoring (Sambito and Freni, 2021).

In a water distribution network (WDN), water quality may be compromised due to the accidental introduction or deliberate release of contaminants, resulting in a public health issue ((Ostfeld, et al., 2008) – (Preis and Ostfeld, 2008)). To better identify contamination events, it is essential to determine the optimal position of the probes to minimize equipment costs (Villegz, et al., 2016) and maximize detection efficiency (Weickgenannt, et al., 2008) while at the same time reducing contamination risks (Murray, et al., 2009).

The choice of a fixed or mobile monitoring system affects measurements differently, as it has been reported that the use of sensors installed within the water supply helps to monitor water more continuously than sampling modes ((Perelman and Ostfeld, 2013) – (Oliker and Ostfeld, 2016) – (Sankary and Ostfeld, 2017)). It is therefore critical to use real-time monitoring systems, although the installation and operating costs are greater than those of a system that uses static control (Creaco, et al., 2019).

Current state-of-the-art water distribution system analysis usually adopts a simplified approach to water quality modelling, neglecting dispersion and diffusion and focusing on either simplified or detailed reaction kinetics. In fact, Ohar et al. (2015) solved the sensor optimization problem by using the EPANET-MSX (Multi-Species eXtension) model to simulate the contamination of three literature networks (Net3, BWSN 1 and Dover) with three organophosphates that were simulated using detailed contamination kinetics. Furthermore, the authors

considered quantitative measures of the population affected by these contaminants as a result of such intrusions. Yang and Boccelli (2016) used the same modelling approach, but incorporated dynamic water quality models to more realistically simulate the response of water quality parameters to a network-scale contamination event. The authors also evaluated current risk assessment assumptions for sensor placement and the performance of event detection algorithms (EDAs). The results demonstrated that current EDA assessment approaches, as well as contamination warning system (CWS) design assumptions, may not adequately represent the real evolution of events in a distribution system under common low flow conditions.

Abokifa et al. (2020) used the EPANET model and the WUDESIM model (advective-dispersive model for a single species) of their own development to model the behaviour of chlorine within water networks. The authors found a better result of modelling water quality using the full model, which takes into account the dispersion process, as not considering such processes can not only overestimate the residual chlorine concentrations in dead-end branches, but it can lead to partially masking the deterioration in water quality due to reduced demand.

Numerous studies have highlighted the importance of considering diffusive-dispersive processes in water distribution systems. First, in 1953–1954, Taylor numerically analysed dispersive phenomenon in laminar and turbulent flow regimes to determine the value of the dispersion coefficient and show how it is related to flow regime quantities ((Taylor G., 1953) - (Taylor G. I., 1954)).

Axworthy and Karney (1996) found that dispersive-diffusive transport processes became important when the flow velocity was low and the Reynolds number was less than 50,000, which is common in urban water distribution networks at night. This was demonstrated by combining the limiting velocity and the dispersion of the coefficient to adjust the sensitivity of the dispersion to the flow velocity at a particular position. After analytical applications, they discovered that as speed increased, dispersive effects tended to disappear and became negligible; thus, an advective transport model should be suitable for any analytical need.

Romero-Gomez and Choi (2011) realized that the presence of a solute persists long after a tracer pulse has passed a fixed downstream position and revealed that the dispersion velocity near the end of the pulse is greater than the velocity near the front of the pulse. This result occurs because low-speed regions close to the wall strongly impede solute transport due to the non-slip boundary condition, and such conditions differ for dispersion upstream and downstream of the contaminant injection. For this reason, they specified the dispersion coefficient while considering the effect of flow direction on dispersion. This approach was used in their study because it highlights the difference between mass flows backwards and forwards from a specific position, which result in different dispersion velocities that lead to solute transport in both directions.

Furthermore, the importance of these transport processes has been studied and experimentally validated by Piazza et al. (2020), who observed that their importance decreased in the presence of turbulence in the pipes, which produces a very different result in terms of sensor positioning.

For this reason, the goal of this work is to demonstrate the limitations of the advective model in reproducing experimental results of contamination events when laminar and transitional regimes are significantly present in the system, as well as how the positioning of quality sensors changes depending on the numerical model used. Three different numerical models were used to reproduce the experimental data maintaining constant the hydraulic features: an EPANET advective model (Rossman, et al., 1994), in which the solute transport mechanisms are relatively simple, of the plug-flow type; an updated version of the EPANET model in which the

diffusion-dispersion equations proposed by Romero-Gomez and Choi (2011) are inserted; and a new diffusive-dispersive model that solves the Advection-Diffusion Equation (ADE) in the two-dimensional case under dynamic conditions. The latter model is discussed in the following section.

Experimental Setup And Conditions

The experimental network of Enna University is a closed water-distribution network composed of high-density polyethylene pipes (HDPE 100), with a nominal pressure of 16 bar (PN16), a DN (nominal diameter) of 63 mm, and a thickness of 5.8 mm. The net is divided into three meshes, each of which contains pipe windings with a radius of 2.0 m and a length of approximately 45 m. Eight “users” are connected to the main network through eight internal nodes consisting of multilayer polyethylene pipes with diameters of 20 mm (internal diameter of 16 mm), which simulate the domestic users of the real network (Fig. 1, Fig. 2).

The network is powered by four polyethylene tanks, three of which are hydraulically connected upstream of the pumping system and one which is positioned centrally to the network, where the collected flows are conveyed to the user nodes, which feed the other three via a recirculation pump. Overall, the tanks can store up to 8 m³ of water (Fig. 2). The pumping system, which consists of four vertical multistage centrifugal electric pumps and an air vessel that allows pressure stabilization, behaves like a constant load tank, keeping the pressure constant and equal to a preset value between 1 and 6 bar with a tolerance of 0.05 bar; the pumping system varies the speed of the pumps.

The system flows in the pipes are monitored by 5 electromagnetic flow meters installed in various sections of the network (trunks 4–5, 6–7, 7–8, 9–10 and 11–12). Pressure cells and multi-jet water meters are present in each node. Furthermore, Wi-Fi real-time remote-controlled conductivity probes were positioned at each node and connected with the monitoring appliances via a central computer, which was also capable of regulating flows supplied to users via remote-controlled valves. The conductivity probes have an accuracy ranging from 1,300 µs to 40,000 µs and outputted the conductivity values in µs, the total dissolved solids (TDS), and the salinity. The salinity values were derived from the practical salinity scale (PSS- 78). The real-time Wi-Fi remote control system consisted of a Wi-Fi router that was connected to 8 Arduino cards that were outfitted with conductivity probes. The Arduino cards were programmed with a sample code to detect the desired sizes. All data were collected with the Wi-Fi data acquisition cards and sent to the server. Experiments were carried out by varying the outflows in nodes 6, 7, 8 and 10 (kept constant in each experiment) with flows ranging from 80 L/h to 400 L/h to obtain different flow regimes in the pipes. No leakages were applied in the present study.

To carry out contamination experiments, the network was equipped with a 100-litre polyethylene tank and an injection pump connected to one of the network nodes chosen at random via a flexible PVC pipe approximately 50 m in length and 40 mm in diameter (Fig. 3). Sodium chloride was chosen as the contaminant because it is easy to find, inexpensive, nontoxic and easy for the probes to detect. The experiments were carried out considering all branches of the network open.

Methods

2.1 Method

The modelling analysis was carried out using the state-of-the-art advective EPANET model, an upgraded version of the EPANET model (advective-diffusive-dispersive model), which includes the diffusion and dispersion equations proposed by Romero-Gomez and Choi (2011) and a new diffusive-dispersive model developed by the authors called EPANET-DD (dynamic-dispersion).

The classic advective EPANET model solves the advective transport equation by solving a mass balance of the fundamental plug-flow substance that accounts for the advective transport and might include kinetic reaction processes (Eq. 1). Using this approach, the substance mass is assigned to discrete volume elements once all the connections in the network have been partitioned. Thereafter, the concentration within each volume segment is eventually subjected to reactions and then transferred to the adjacent downstream segment. If the latter is a junction node, the incoming mass and flow volumes combine with those already present at the network nodes. Once these processes have been exhausted for all network elements, the concentration is calculated and released in the first pipe segments, with flow leaving the node. In this case, the effect of longitudinal dispersion is totally neglected, as it is not considered important in most operating conditions (Rossman, et al., 1993).

$$\frac{\partial C_i(x, t)}{\partial t} = -u_m \frac{\partial C_i(x, t)}{\partial x} - KC_i(x, t) \quad (1)$$

In the present application, in which a conservative tracer is used, the reaction term is neglected.

In contrast, the advective-diffusive-dispersive model, which includes the diffusion and dispersion equations proposed by Romero-Gomez and Choi (2011), solves the transport equation by highlighting the differences between mass flows backwards and forwards from a specific position, which result in different dispersion speeds that lead to solute transport in both directions (Eq. 2).

$$\frac{\partial C}{\partial t} = \frac{1}{\Delta x} (\varphi_b - \varphi_f) - u_m \frac{\partial C}{\partial x} \quad (2)$$

in which

$$\varphi_b = -E_b \frac{\partial C}{\partial x} \Big|_b \text{ and } \varphi_f = -E_f \frac{\partial C}{\partial x} \Big|_f \quad (3)$$

$$E_b = E_b(0) \exp(-xpT) + \beta_b(T) E^* \quad (4)$$

$$E_f = E_f(0) \exp(-xpT) + \beta_f(T) E^* \quad (5)$$

where E_b and E_f (Eq. 4 and Eq. 5) are the dispersion parameters backwards and forwards with respect to the flow direction, u_m is the flow average velocity and $\beta_b(T) = \beta_f(T) = 3.705 \cdot T$.

The dimensionless travel time (T) was calculated as in Eq. (6). This parameter indicates how far the dispersion coefficient has progressed towards achieving stable conditions.

$$T = \frac{4D_{AB} \bar{t}}{d^2} = 4 \frac{x^*}{S_C \cdot R} \quad (6)$$

in which

$x^* = \frac{L}{d}$ is a dimensionless pipe length that defines the location of solute migration, L , with respect to the pipe diameter, d ;

$R = \frac{u_m \cdot d}{\nu}$ is the Reynolds number, which accounts for the mean flow velocity's (u_m) geometric dimensions (d) and fluid conveying properties (kinematic viscosity, ν);

$S_C = \frac{\nu}{D_{AB}}$ is the Schmidt number, which accounts for the solute properties (solute diffusion coefficient, D_{AB});

and $\bar{t} = \frac{L}{u_m}$ is the time, which is defined as the ratio between the location of solute migration, L , and the flow velocity (u_m).

The previously reported expression of the coefficient $\beta_b(T) = \beta_f(T)$ was determined, as the authors found that for short travel times ($T < 0.01$), the dispersion rates were amplified by 25% more than the numerical results using the formulation proposed by Lee (2004), where $\beta(T) = 1 - \exp(-16T)$.

The EPANET-DD model solves the equations under quasi-steady flow conditions, solving the hydraulic problem under steady flow conditions with the EPANET-MATLAB-Toolkit (Eliades, et al., 2016) and the advection-diffusion-dispersion equation under dynamic flow conditions in the two-dimensional case with the classical random walk method (Delay, et al., 2005), implementing the diffusion and dispersion equations proposed by Romero-Gomez and Choi (2011).

As demonstrated in the literature ((Kinzelbach and Uffink, 1991) - (Delay, et al., 2005)), the use of this combined method is possible due to the similarities between the Fokker-Planck-Kolmogorov equation and the advection-dispersion equation. In fact, the two equations are essentially identical unless there is a conceptual difference between the parameters of the two equations, as the parameters present in the Fokker-Planck-Kolmogorov equation are independent of time, resulting from the stationary hypothesis. To overcome this problem and address the issues related to discontinuities that could cause local mass conservation errors (LaBolle, et al., 1996), Delay et al. (2005) provided a new equivalence, making this analogy valid again. This methodology can be easily applied to any flow model because the mass of the solute is discretized and transported by the particles in the random walk. Consequently, the mass conservation principle is automatically satisfied because the particles cannot suddenly disappear.

This model allows us to determine the position of the solute particles that move inside the network in the x and y directions as a function of the different flow regimes that occur inside the network, as shown in equations (7) and (8):

$$x = x + \frac{3}{2} u_x \left(1 - \left(\frac{y}{\frac{d}{2}} \right)^2 \right) dt + \sqrt{2 \cdot E_{forb}} \cdot dt \quad (7)$$

$$y = y + u_y dt + \sqrt{(E_f + E_b) \cdot dt} \quad (8)$$

where u_x corresponds to the component along the x axis of the flow velocity, u_y corresponds to the component along the y axis of the flow velocity, dt is the duration of the contamination event, d is the pipe diameter, and E_f and E_b are the forwards and backwards diffusion coefficients, respectively, as defined by Romero-Gomez and Choi (2011). The diffusion coefficient used in Eq. (7) assumes the forwards or backwards values depending on whether the flow direction is positive or negative. The above equation was developed considering laminar flow conditions, in which the velocities in the network are relatively low. This allows the particles to move freely along

the y axis. This characteristic is also highlighted by the presence of the term in round brackets, $\left(1 - \left(\frac{y}{\frac{d}{2}}\right)^2\right)$

, which multiplies the x component of the velocity u_x . In fact, as the velocity along the x direction increases and the flow rate changes, the particles tend to move along the preferred flow direction, and the term in brackets disappears from the equation.

To confine the particles inside the pipe section, the previous equations are solved considering the following boundary conditions (Eq. 9 and Eq. 10).

$$y = -2 \cdot y_{max} - y \text{ for } y < -y_{max} \quad (9)$$

$$y = 2 \cdot y_{max} - y \text{ for } y > y_{max} \quad (10)$$

where the particle position along y is limited above and below by the physical presence of the pipe wall. The parameters $-y_{max}$ and y_{max} coincide with the value of the pipe radius and take on a positive and negative value since the x axis has been placed at the centre of gravity with respect to the cross-section of the pipe. By means of these two boundary conditions, the particles are not only prevented from escaping from the pipe but are also reflected, which prevents the particles from settling along the wall. These conditions are called the boundary reflection condition.

At this point, the contaminant concentration has been determined through Eq. (11), in which the concentration value at the previous time has been increased by an amount that corresponds to the concentration per unit of

particles $(C \cdot n)$ passing through the control volume $\left(\frac{L}{\Delta x} \cdot \pi \frac{d^2}{4}\right)$, where L is the pipe length, Δx is the section

number of the pipe, and $\pi \frac{d^2}{4}$ is the cross-sectional area of the pipe.

$$C = C + \frac{C \cdot n}{\frac{L}{\Delta x} \cdot \pi \frac{d^2}{4}} \quad (11)$$

The models were applied to the experimental network of Enna University – UKE (see (Piazza, et al., 2020) for details) suitably calibrated previously from the hydraulic point of view.

The roughness coefficient (reported in the Table 2) was calibrated according to the flow rate measured upstream of the network ($1.44 \text{ m}^3 / \text{h}$) and the diameter of each pipeline, calculating and iterating the uniform flow rate in

order to coincide with the measured flow rate upstream of the network. Numerous experimental tests were conducted on the network, varying the pressure set at the pumping system (3.5–4.5 bar) and the flow rates drawn from the network nodes (between 5 and 15 L/min for nodes 5, 8 and 11).

The Table 1, Table 2 and Table 3 show the calibrated roughness values of the pipes and the standard deviation[1] (σ) values determined for the pressures at the nodes 6, 7, 9, 10, the flow rates flowing into the network and the flow rates tapped at the nodes 5, 8, 11.

Table 1
Standard deviation between the pressures measured in the network and simulated numerically.

	Node 6	Node 7	Node 9	Node 10
σ [mH2O]	0.01	0.15	0.05	0.09

Table 2
Pipes roughness and standard deviation between the flow rates measured in the network and simulated numerically.

	Link 5	Link 6	Link 7	Link 9	Link 10	Link 11	Link 13
Roughness [mm]	1	1	1	1	1	1	1
σ [m ³ /h]	0.12	0.12	0.08	0.11	0.11	0.11	0.15

Table 3
Standard deviation between the measured tapped flow rates and the numerically simulated flow rates.

	Node 5	Node 8	Node 11
σ [L/min]	0.45	0.07	0.07

To calibrate the relative dispersion coefficients for the two models used in the present study, three statistical parameters were determined: the Nash-Sutcliffe efficiency (NSE), Kling-Gupta efficiency (KGE) and coefficient of determination (R^2). The Nash-Sutcliffe efficiency (NSE) coefficient (Nash and Sutcliffe, 1970) is a hydrology metric that measures how well a model simulation predicts an outcome variable. It is defined as one minus the ratio of the error variance of the modelled time series divided by the variance of the observed time series, as shown in Eq. (12):

$$NSE = 1 - \frac{\sum (y_i - y_{i, sim})^2}{\sum \left(y_i - \bar{y} \right)^2} \quad (12)$$

where y_i and $y_{i, sim}$ correspond to the measured and simulated values of the variable, respectively, and \bar{y} is the average of the measured values of y . If $NSE = 1$, there is a perfect correspondence between the model and the

observed data; if $NSE = 0$, the model has the same predictive capacity as the average of the time series in terms of the sum of the square errors. If $NSE < 0$, the observed mean is a better predictor of the model.

The Kling-Gupta efficiency (KGE) coefficient (Gupta, et al., 2009) is a metric that measures the goodness of fit (Eq. (13)). It consists of three main components: the correlation coefficient between the observations and simulations r ; the ratio between the standard deviation of the simulated values and the standard deviation of the observed values, and the ratio between the average of the simulated values and the average of the observed values.

$$KGE = 1 - \sqrt{(r - 1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2} \quad (13)$$

Similar to the NSE coefficient, $KGE = 1$ indicates perfect agreement between the simulations and observations. For KGE values ≤ 0 , analogous to what the authors observed for NSE values, all negative values below the threshold $KGE = 0$ indicate results with poor model performance.

The coefficient of determination (R^2) is a measure of the goodness of fit of a statistical model. It is defined as the squared value of the linear correlation coefficient. The R^2 value ranges between 0 and 1. A value of zero indicates that there is no correlation between two data series. On the other hand, higher values of the coefficient indicate a better fit of the model. However, it is not always true that large R^2 values result in a good model fit, as the linear correlation coefficient could produce a perfectly positive or negative relationship (Moriasi, et al., 2007).

2.2 Optimization problem

The optimization problem was solved by using the Monte Carlo method, which is based on probabilistic procedures and can solve problems that present analytical difficulties that would otherwise be difficult to overcome (Tarantola, 2004).

Conceptually, the method is based on the possibility of sampling an assigned probability distribution, $F_{(X)}$, using numbers drawn at random (random numbers); that is, the possibility of generating a sequence of events $X_1, X_2, \dots, X_n, \dots$, distributed according to $F_{(X)}$. Instead of using numbers drawn at random, the authors use a sequence of numbers obtained through a well-defined iterative process; these numbers are referred to as pseudorandom because, although they are not random, they have statistical properties similar to those of true random numbers.

The Monte Carlo method yields reliable answers in the study of complex real systems, although the solution obtained is never exact in a statistical sense, as it is subject to uncertainty, which decreases as the number of statistical samples increases.

In the present case, the Monte Carlo method was used to solve the optimization problem for the positioning of three water quality sensors within the laboratory network of the University of Enna "KORE", which is fully described in the following section. Given the uncertainty in the position, magnitude and duration of the contamination, 1000 simulations were performed, with the contamination parameters (mass of contaminant, duration of contamination and contamination node) set at random. User demands in all nodes were fixed at 2.5 l/min. The inlet head was fixed to 1.5 bar. At the same time, an experiment was carried out in node 5, with a duration of 3 minutes and a mass of 460 grams (leading to a constant concentration of 4600 mg/l).

Three objective functions were used:

- F_1: Detection likelihood, i.e., the probability that a sensor configuration will detect the contamination;
- F_2: Detection time, i.e., the average time between contamination and detection in 200 simulations;
- F_3: Detection redundancy, i.e., the probability that the contamination is detected by two sensors within 20 minutes.

The objective functions were slightly modified from those presented in (Preis and Ostfeld, 2008) to comply with the smaller dimensions of the analysed network, and they were equally weighted in the selection of the optimal sensor location.

[1] Standard deviation of zero means that there is no variability between the data.

Results And Discussion

The three numerical approaches discussed above (advective, advective-dispersive based on the Romero-Gomez and Choi formulation, and the EPANET-DD) were applied to the University of Enna “KORE” laboratory network and compared to experimental data. For the advective model, calibration is limited to the hydraulic aspects so no further calibration is needed for conservative solutes; for both advective-diffusive-dispersive models, the forwards and backwards dispersion coefficients, E_f and E_b , were calibrated. The calibration process was carried out through a simple trial, and an error procedure was used to maximize the Kling-Gupta efficiency (KGE), the Nash-Sutcliffe efficiency (NSE) and the determination coefficient (R^2) convergence criterion over the measured and simulated concentrations (Table 5).

Figure 4 compares the experimental results (dotted line), obtained by considering all the branches of the network open, and the three different modelling approaches (advective model: grey line; Romero-Gomez and Choi model: dashed line; and random walk model: continuous line), obtained by contaminating the UKE network at node 5 with a sodium chloride concentration of 4600 mg/l for 3 minutes, monitoring nodes 6 (a), 7 (b), 8 (c), 9 (d), 10 (e) and 11 (f) of the UKE network and evaluating the effect generated by the different Reynolds values present in the network. The test was 3 hours, and the flow rates tapped at the nodes were set to achieve different flow regimes in the network, as shown in Table 4. It was observed that the three flow regimes (laminar, transition and turbulent) occurred simultaneously within the network as a function of the tapped flow rates. In particular, at node 6, which is immediately downstream of the contaminant inlet node, there is a turbulent flow regime. At greater distances from the contaminant entry node, the authors observe a variation in the flow regime, which oscillates between transition and laminar flow. Furthermore, some nodes of the network are affected by both flow regimes (Fig. 4e, Fig. 4f), as they converge to two branches of the network (Fig. 1).

The numerical analysis was carried out considering the following calibration coefficient values: backwards and forwards coefficients equal to 0.17 and 0.51, respectively. These coefficients provide a better fit between the simulated and measured data.

It was observed that the advective model was unable to represent all the experimentally monitored nodes. This demonstrates the inadequacy of the advective model in reproducing experimental data. Although some nodes have high values for the parameters KGE, NSE and R^2 , graphically, there is no correspondence between the

simulated and measured data to justify these values. In fact, as shown in Fig. 4, this model works well only for one node of the network (Fig. 4a), which is supported by a high value of the KGE, NSE and R^2 coefficients, which are reported in Table 5. This node, which was directly downstream of the node where sodium chloride was injected, was characterized by a turbulent flow regime with a Reynolds number of 4112. Under these conditions, the advective model was able to centre the peak of the contamination and the time interval in which it occurs, despite having a higher peak concentration. Analysing the remaining nodes shown in Fig. 4b, Fig. 4c, Fig. 4d, Fig. 4e, Fig. 4f, the advective model reproduces anticipated events with respect to the experimental data, which in some cases corresponded to a few minutes, but in the case of Fig. 4c, the event was anticipated by approximately an hour. Furthermore, as shown in Fig. 4c and Fig. 4d, the model underestimates the persistence of the contamination, as the event is quickly exhausted.

The results of the advective-dispersive model, which is based on the formulations of Romero-Gomez and Choi, obtained using the previously calibrated backwards and forwards coefficients differs more from other models and from the trend of the experimental data. In fact, as shown in Fig. 4, although the model can reproduce the dispersive effect of the contaminant in the network, it cannot reproduce the behaviour of the experimental data. This is easily seen for the nodes in Fig. 4c, Fig. 4d, and Fig. 4e, where the model overestimated the contamination peaks and the persistence of the contaminant within the distribution network and reproduced a significantly anticipated event in Fig. 4c but delayed events for all other nodes. The model performed slightly better for the node in Fig. 4f, which has a higher coefficient of determination R^2 than that calculated for the other nodes. In this node, the coexistence of the transition flow regime and the laminar flow regime, which have Reynolds values of 3598 and 200, respectively, was observed. The transition flow regime most likely dominates because the dispersive effects are not overpowering. Furthermore, for the aforementioned node, the model was able to centre the contamination peak, despite having overestimated its mass.

Using the same backwards and forwards coefficients used in the Romero-Gomez and Choi model, it was observed that the results obtained by applying the new EPANET-DD model outperform the previous two models used in this study. This is supported by the high values of the KGE, NSE and R^2 coefficient, as shown in Table 5. In fact, for all the monitored nodes in Fig. 4, the model can centre the contamination peak. Furthermore, considering Fig. 4b and Fig. 4f, which have transition flow regimes with Reynolds numbers of 3598, the model perfectly fits the experimental data departing from the classic bell-shaped trend typical of a Gaussian distribution. It is worth noting that at node 9 in Fig. 4d, which has a laminar flow regime with a Reynolds number of 514, the model perfectly fit the ascending branch of the experimental data but failed to reproduce the descending branch of the curve. This is due to a separation between the contamination behaviour at the edge and in the centre of the pipeline caused by the transition from turbulent to laminar flow. For the previous node, in Fig. 4e, the model has a gap with the experimental data in the descending part of the curve but perfectly reproduced the terminal part of the pollutograph.

Table 4
Reynolds Number and Flow Regime.

	<i>Link 4</i>	<i>Link 6</i>	<i>Link 7</i>	<i>Link 9</i>	<i>Link 10</i>	<i>Link 11</i>	<i>Link 12</i>	<i>Link 13</i>
Reynolds (Re)	4112	200	3598	1542	514	2056	1542	3598
Flow Regime	Turbulent	Laminar	Transition	Laminar	Laminar	Transition	Laminar	Transition

Table 5
Comparison of statistical parameters (Kling-Gupta efficiency, Nash-Sutcliffe efficiency, R^2) for the advective, Romero-Gomez and Choi (2011) and random walk models.

<i>Node</i>	<i>Advective model</i>			<i>Romero-Gomez and Choi (2011) model</i>			<i>Random Walk model</i>		
	KGE	NSE	R^2	KGE	NSE	R^2	KGE	NSE	R^2
6	0.44	0.52	0.29	-0.60	-0.72	0.21	0.63	0.69	0.49
7	0.25	0.59	0.68	-0.08	-0.15	0.12	0.81	0.84	0.76
8	-0.55	-1.50	0.08	0.01	0.35	0.04	0.45	0.43	0.92
9	0.22	0.18	0.43	-1.58	-5.57	0.13	0.29	0.35	0.17
10	0.34	-0.01	0.19	-4.35	-14.81	0.09	-0.15	-0.54	0.55
11	-0.30	-0.62	0.05	-0.94	-1.18	0.79	0.42	0.76	0.90

When the optimization problem is solved using the three models previously analysed, very different optimal configurations were obtained. In fact, using the advective model, two possible configurations were obtained for the positioning of the sensors, as presented in Fig. 6. Table 6. shows the characteristics and performance of the optimal configurations for the three models. By solving the optimization problem using the advective model, it was possible to optimize two objective functions (F_1 and F_2) simultaneously using a single sensor configuration (5–6–10). To maximize the objective function F_3 , node 11 must also be considered. With three sensors in optimal positions, 66% of the contamination episodes were detected on average (maximum value of function F_1), and 52% were detected by at least 2 sensors within 20 minutes (maximum value of function F_3). The optimal average detection time was approximately 8 minutes. Figure 5a shows the Pareto fronts obtained for the three objective functions.

Table 6
Numerical analysis: Results of the Optimization Problem.

<i>Advective model</i>			<i>Romero-Gomez and Choi (2011) model</i>			<i>Random Walk model</i>		
Objective Functions	Optim. Values	Sensor node index	Optim. Values	Sensor node index	Optim. Values	Sensor node index	Optim. Values	Sensor node index
Detection likelihood (F_1)	0.66	5 6 10	0.65	6 8 9	0.95	6 7 10		
Detection time [s] (F_2)	517.13	5 6 10	858.54	6 8 9	538.96	6 7 10		
Redundancy (F_3)	0.52	5 6 11	0.54	5 6 7	0.70	6 7 10		

The optimal solutions considering dispersion, using the Romero-Gomez and Choi (2011) formulations, differ from the advective case by considering both the optimal sensor location and the objective function values (Table 6.). As in the previous case, the objective functions F_1 and F_2 were both maximized by the same configuration of sensors (6–8–9), with values that are slightly lower than the previous values. The contamination event had a detection likelihood of 65%, and the optimal value of the detection time was slightly more than 14 minutes. For the objective function F_3, the optimal redundancy value of 54% was obtained from a clearly different configuration than the first method and included nodes 5–6–7. This is related to the importance of backwards diffusive propagation, which is most evident in laminar velocity profiles. Using the aforementioned dispersive model, in which the dispersion was turbulent or transitional rather than laminar, the backwards diffusive propagation was less evident, as the dispersive behaviour was homogeneous across all network pipelines. Therefore, in this case, the discriminant that determines the phenomenon of redundancy is a function of higher circulating water volumes (Fig. 6). The Pareto fronts are presented in Fig. 5b.

By solving the optimization problem using the dynamic dispersive model (EPANET-DD), the three objective functions were optimized using a single configuration of three sensors, positioned at nodes 6–7–10, which yielded better values than the other models (Table 6). In fact, with this configuration, 95% of the contamination episodes were detected on average, and 70% were detected by at least 2 sensors within 20 minutes. The optimal average detection time was approximately 9 minutes, which was slightly lower than what was obtained using the advective model. The Pareto fronts are presented in Fig. 5c.

Conclusions

In this study, the inadequacy of the advective model in reproducing the results of an experimental contamination campaign as a function of different flow regimes was demonstrated, as in 5 out of 6 nodes, the trend of the simulated data differs from that of the experimental data. The model performance was evaluated using three distinct coefficients (KGE, NSE, R^2). Two other models (an advective-dispersive model based on the formulations of Romero-Gomez and Choi and a new model called EPANET-DD) were evaluated with respect to experimental data, and the performance of the latter model was compared with the results obtained from the advective model. Furthermore, the performance of the models was evaluated for solving the problem of the optimal positioning of water quality sensors in terms of the detection likelihood, detection time and redundancy. In summary, the main conclusions of this study are as follows:

- The advective model works well only in locations close to the contamination node, where it can intercept the peak of the contamination even for with lower values.
- In all other cases, the contamination event was anticipated and had a shorter duration than that detected by the experimental campaign.
- The Romero-Gomez and Choi model can represent the dispersive behaviour of the contaminant, but it poorly represents the experimental data in terms of delay or anticipation of the contamination peak and overestimation of the contaminant mass.
- The new EPANET-DD model produces the best results in terms of adaptability with the experimental data. It simultaneously represents the peak time and provides the same precision as the Romero-Gomez and Choi model.
- When the advective approach was used to solve the optimization problem, the sensors were positioned in areas with high Reynolds numbers, where the flow regimes are predominantly turbulent and transition.
- Using the dispersive approach of Romero-Gomez and Choi, the sensors were positioned in a linear pattern and covered most of the network.
- The EPANET-DD model provided the best performance, with a contamination event detection likelihood of 95%, a redundancy of 70% and a detection time of approximately 9 minutes.
- The importance of backwards diffusive propagation is most evident in laminar velocity profiles. This is because the fluid threads of the velocity profile closest to the wall have low velocities, which determines that the nodes that are in a laminar area of the network, which has a greater backwards diffusivity; thus, in this profile, the fluid threads are more sensitive to contaminant detection.
- Depending on the model used to solve the optimization problem, different configurations for sensor positions are obtained, as are different detection efficiencies for the objective functions. In fact, the parameters determined by the advective model are much lower than those determined by the dynamic dispersive model (EPANET-DD).

Declarations

Ethical Approval: Not applicable.

Consent to Participate: Not applicable.

Consent to Publish: Not applicable.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

Author contributions: G.F. designed the research; G.F. wrote the code; S.P., M.S. and G.F. performed the research; S.P. analysed the data; and S.P., M.S. and G.F. wrote the paper.

Availability of data and materials: https://github.com/gabrielefreni/epanet_dd.git.

References

1. Abokifa A, Xing L, Sela L (2020) Investigating the Impacts of Water Conservation on Water Quality in Distribution Networks Using an Advection-Dispersion Transport Model. *Water* 12(4):1–18. doi:10.3390/w12041033
2. Aisopou A, Stoianov I, Graham N (2012) In-pipe water quality monitoring in water supply systems under steady and unsteady state flow conditions: A quantitative assessment. *Water Res* 46(1):235–246. doi:10.1016/j.watres.2011.10.058
3. Axworthy DH, Karney B (1996) Modelling Low Velocity/High Dispersion Flow in Water Distribution Systems. *J Water Resour Plan Manag* 122:218–221. doi: 10.1061/(ASCE)0733-9496(1996)122:3(218)
4. Creaco E, Campisano A, Fontana N, Marini G, Page P, Walski T (2019) Real time control of water distribution networks: A state-of-the-art review. *Water Res* 161:517–530. doi:10.1016/j.watres.2019.06.025
5. Delay F, Ackerer P, Danquigny C (2005) Simulating Solute Transport in Porous or Fractured Formations Using Random Walk Particle Tracking: A Review. *Vadose Zone J* 4(2):360–379. doi:10.2136/vzj2004.0125
6. Eliades D, Kyriakou M, Vrachimis S, Polycarpou M (2016) EPANET-MATLAB Toolkit: An Open-Source Software for Interfacing EPANET with MATLAB. *14th International Conference on Computing and Control for the Water Industry (CCWI)*. doi:10.5281/zenodo.831493
7. Gupta H, Kling H, Yilmaz K, Martinez G (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J Hydrol* 377:80–91. doi:10.1016/j.jhydrol.2009.08.003
8. Kinzelbach W, Uffink G (1991) The random walk method and extensions in groundwater modelling. *Transp Processes Porous Media* 761–787. doi:10.1007/978-94-011-3628-0_17
9. Knoben W, Freer J, Woods R (2019) Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrol Earth Syst Sci* 1–7. doi:10.5194/hess-2019-327
10. LaBolle E, Fogg G, Tompson A (1996) Random-walk simulation of transport in heterogeneous porous media: Local mass-conservation problem and implementation methods. *Water Resour Res* 32(3):583–593. doi:10.1029/95WR03528
11. Lee Y (2004) *Mass dispersion in intermittent laminar flow*. Ph.D. thesis, University of Cincinnati, Cincinnati
12. Li T, Winnel M, Lin H, Panther J, Liu C, O'Halloran R, Zhao H (2017) A reliable sewage quality abnormal event monitoring system. *Water Res* 121:248–257. doi:10.1016/j.watres.2017.05.040
13. Moriasi D, Arnold J, Van Liew M, Bingner R, Harmel R, Veith T (2007) Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Trans ASABE* 50(3):885–900. doi:10.13031/2013.2
14. Murray R, Haxton T, Janke R, Hart W, Berry J, Phillips C (2009) Sensor network design for drinking water contamination warning systems: A compendium of research results and case studies using TEVA-SPOT. *EPA/600/R-09/141, Office of Research and Development*. National Homeland Security Research Center
15. Nash J, Sutcliffe J (1970) River flow forecasting through conceptual models, Part I - A discussion of principles. *J Hydrol* 10(3):282–290. doi:10.1016/0022-1694(70)90255-6

16. Ohar Z, Lahav O, Ostfeld A (2015) Optimal sensor placement for detecting organophosphate intrusions into water distribution systems. *Water Res* 73:193–203. doi:10.1016/j.watres.2015.01.024
17. Olikar N, Ostfeld A (2016) Inclusion of Mobile Sensors in Water Distribution System Monitoring Operations. *J Water Resour Plan Manag* 142(1). doi:10.1061/(ASCE)WR.1943-5452.0000569
18. Ostfeld A, Uber JG, Salomons E, Berry JW, Hart WE, Phillips CA, Walski T (2008) The Battle of the Water Sensor Networks (BWSN): A Design Challenge for Engineers and Algorithms. *J Water Resour Plan Manag* 134(6):556–568. doi:10.1061/(ASCE)0733-9496(2008)134:6(556)
19. Perelman L, Ostfeld A (2013) Operation of remote mobile sensors for security of drinking water distribution systems. *Water Res* 47(13):4217–4226. doi:10.1016/j.watres.2013.04.048
20. Piazza S, Blokker E, Freni G, Puleo V, Sambito M (2020) Impact of diffusion and dispersion of contaminants in water distribution networks modelling and monitoring. *Water Supply* 20(1):46–58. doi:10.2166/ws.2019.131
21. Preis A, Ostfeld A (2008) Multiobjective Contaminant Sensor Network Design for Water Distribution Systems. *J Water Resour Plan Manag* 134(4):366–377. doi:10.1061/(ASCE)0733-9496(2008)134:4(366)
22. Romeo-Gomez P, Choi CY (2011) Axial Dispersion Coefficients in Laminar Flows of Water-Distribution Systems. *J Hydraul Eng* 137(11):1500–1508. doi:10.1061/(ASCE)HY.1943-7900.0000432
23. Rossman L, Boulos P, Altman T (1993) Discrete volume-element method for network water quality models. *Journal of Water Resources Planning and Management*, 119(5), 505–517. doi:/10.1061/(ASCE)0733-9496(1993)119:5(505)
24. Rossman L, Clark R, Grayman W (1994) Modeling Chlorine Residuals in Drinking-Water Distribution Systems. *J Environ Eng* 120(4):803–820. doi:10.1061/(ASCE)0733-9372(1994)120:4(803)
25. Sambito M, Freni G (2021) Strategies for Improving Optimal Positioning of Quality Sensors in Urban Drainage Systems for Non-Conservative Contaminants. *Water* 13(7):934. doi:10.3390/w13070934
26. Sankary N, Ostfeld A (2017) Inline Mobile Water Quality Sensors Deployed for Contamination Intrusion Localization. *Computing and Control for the Water Industry*
27. Tarantola A (2004) Monte Carlo Methods. In A. Tarantola, & S. f. Mathematics (Ed.), *Inverse Problem Theory and Methods for Model Parameter Estimation* (pp. 41–55). doi:10.1137/1.9780898717921.ch2
28. Taylor G (1953) Dispersion of soluble matter in solvent flowing slowly through a tube. *Proceedings of the Royal Society of London, Series A, Mathematical, Physical & Engineering Sciences*, 219(1137), 186–203. doi:10.1098/rspa.1953.0139
29. Taylor GI (1954) The dispersion of matter in turbulent flow through a pipe. *Proc. Roy. Soc. London Ser. A*, 223(1155), 446–468. doi:10.1098/rspa.1954.0130
30. Villegz K, Vanrolleghem P, Corominas L (2016) Optimal flow sensor placement on wastewater treatment plants. *Water Res* 101:75–83. doi:10.1016/j.watres.2016.05.068
31. Weickgenannt M, Kapelan Z, Blokker M, Savic DA (2008) Optimal Sensor Placement for the Efficient Contaminant Detection in Water Distribution Systems. *Water Distribution Systems Analysis*. doi:10.1061/41024(340)99

Figures

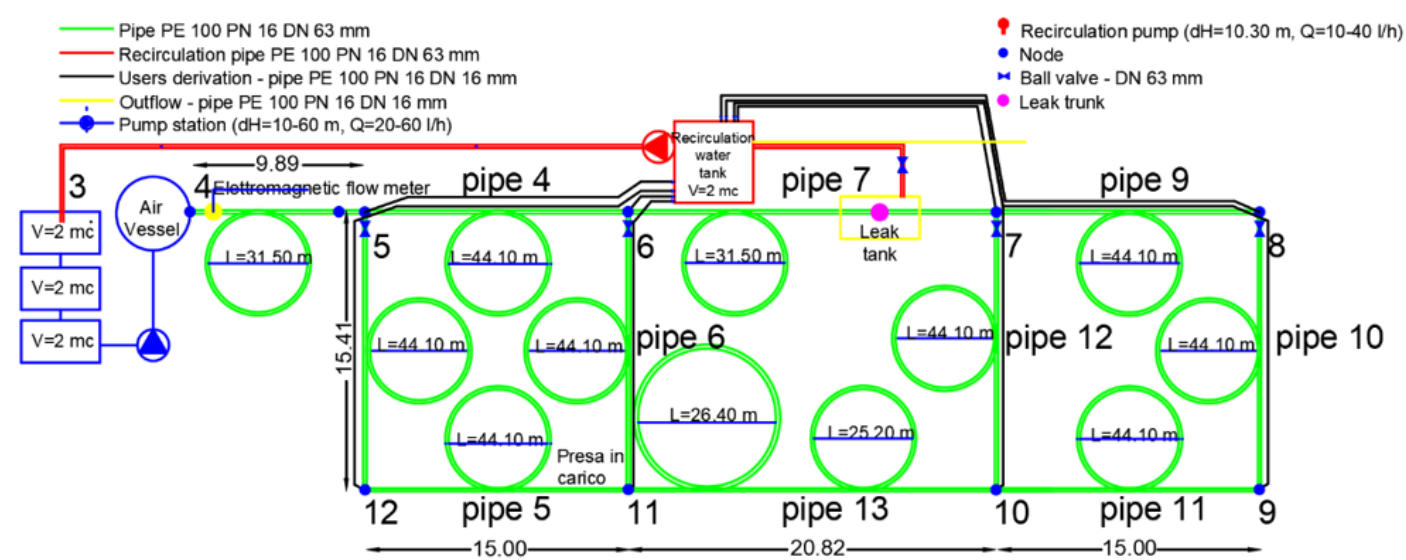


Figure 1

Layout of the water distribution network.



a)



b)



c)

Figure 2

Overview of the water distribution network (a), pumping system (b) and recirculation system (c).



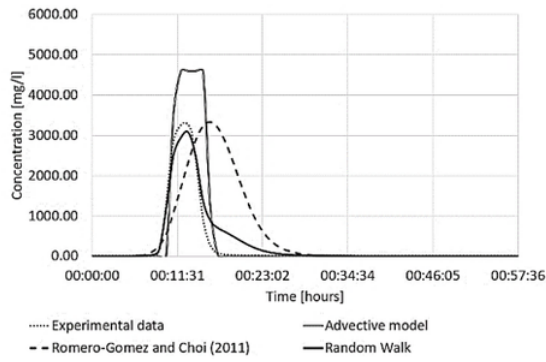
a)



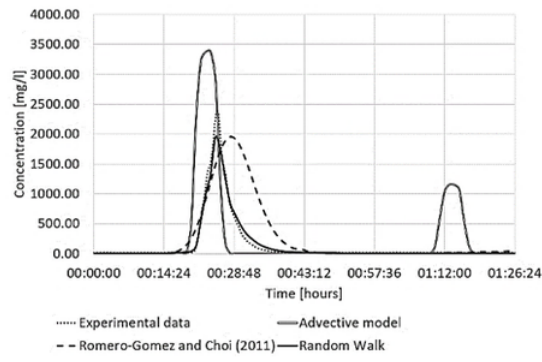
b)

Figure 3

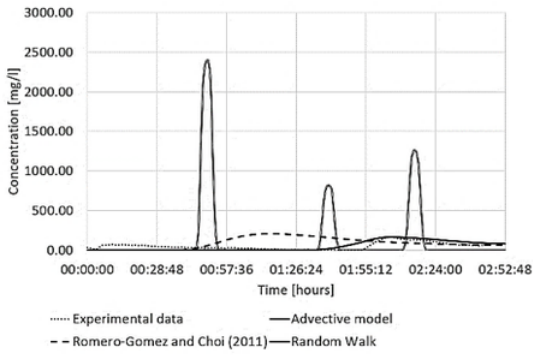
Installation of the pump-reservoir system (a) and connection to the node (b).



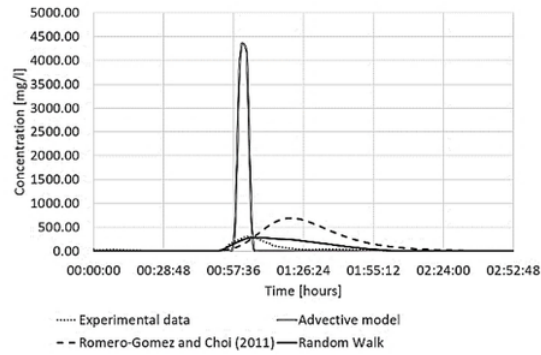
(a)



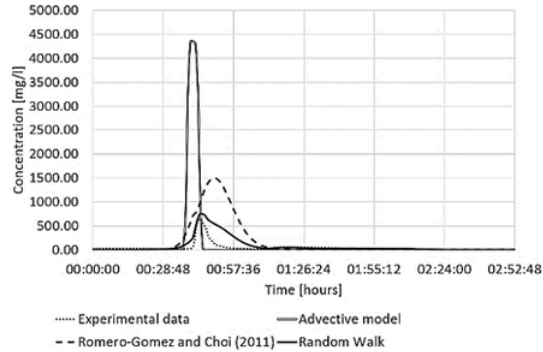
(b)



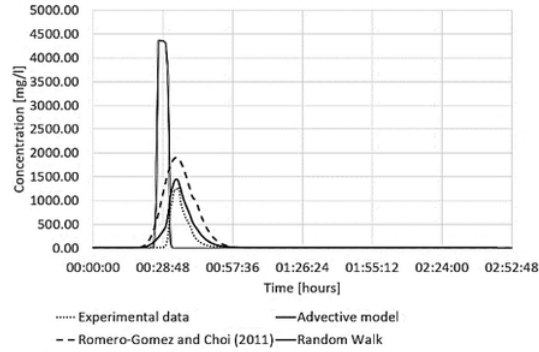
(c)



(d)



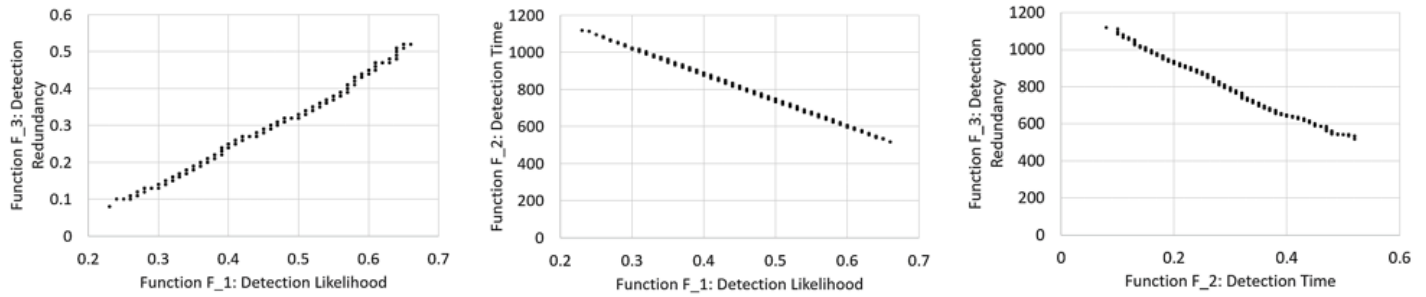
(e)



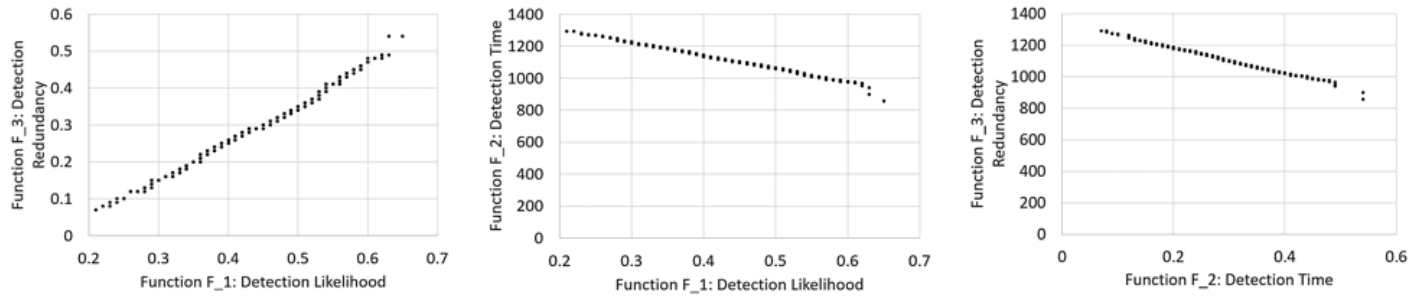
(f)

Figure 4

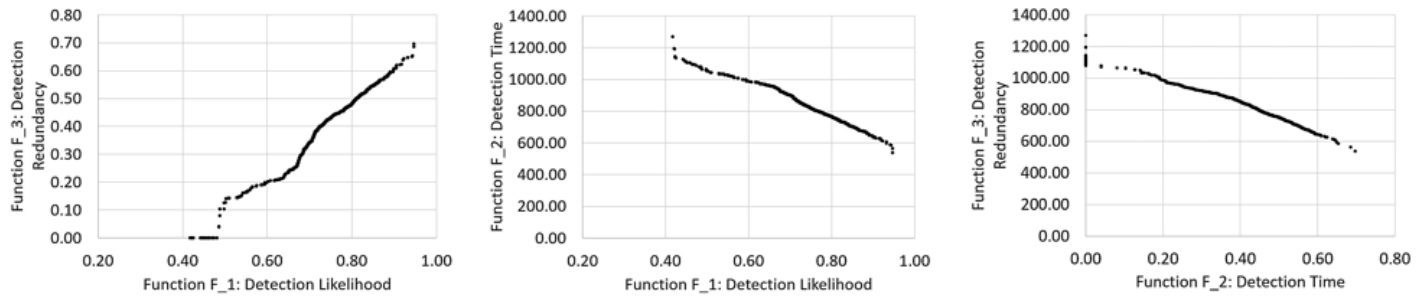
Comparison between the experimental data, the advective model, Romero-Gomez and Choi (2011) model (backwards coefficient = 0.17, forwards coefficient = 0.51), and random walk model for nodes 6 (a), 7 (b), 8 (c), 9 (d), 10 (e), 11 (f), obtained by contaminating node 5 at an operating pressure of 1.5 bar.



a)



b)



c)

Figure 5

The Pareto front including advection only (row a), dispersion (Romero-Gomez and Choi 2011) (row b), dispersion (random walk) (row c): F_1-F_3 (left), F_1-F_2 (centre), and F_2-F_3 (right).

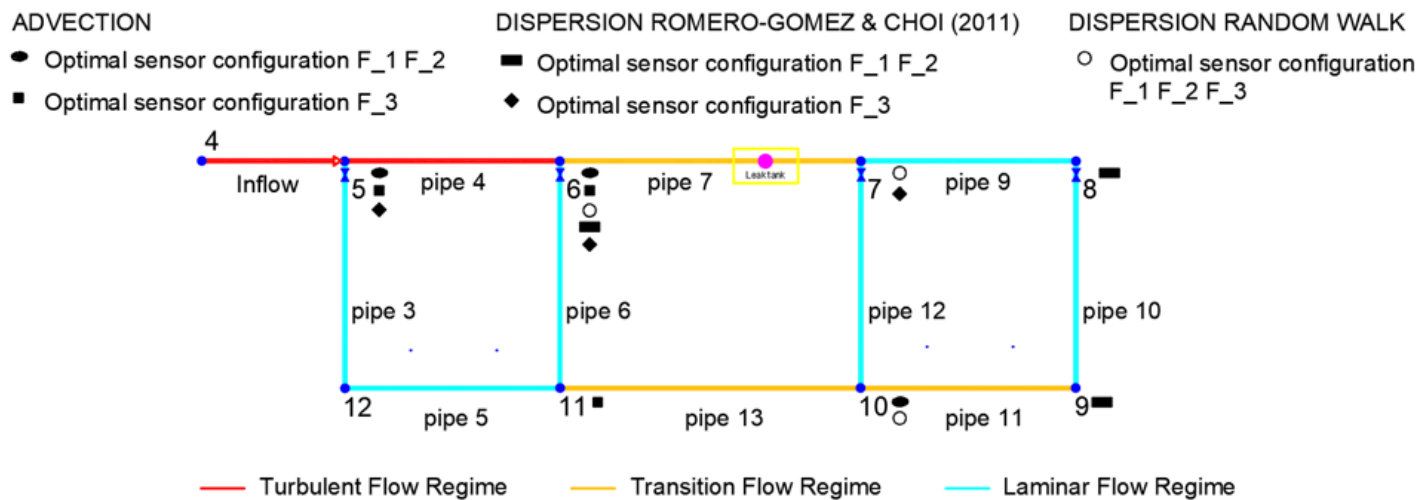


Figure 6

Optimal sensor positioning for the UNIKORE Laboratory Network.