Impact of an incomplete solute mixing model on sensor network design
Pedro Romero-Gomez, Kevin E. Lansey and Christopher Y. Choi

ABSTRACT
Incomplete mixing models have recently been shown to better represent solute transport at junctions of pressurized water systems, compared to a complete mixing assumption. The present work incorporated an incomplete solute mixing model into a methodology for sensor network design. Water quality simulations conducted using both mixing models were carried out to generate pollution matrices that provided the input data for the set covering optimization formulation. Multiple contamination and detection scenarios were simulated by considering both a minimum hazard level of the contaminant and a maximum volume of contaminated water consumed. Examination and comparison of outcomes demonstrated that the water quality solver used may impact sensor network designs in three ways by altering: (i) the minimum number of monitoring stations required for full detection coverage, (ii) the optimal layout of stations over the water network and (iii) the detection domain of some stations.

Key words | contamination, optimization, sensor, solute mixing, water system

INTRODUCTION
Early warning monitoring systems deployed in municipal water networks seek to protect the public health and to reduce the effects of contamination incidents. Because each water system is unique, a persistent problem is locating water quality monitoring stations in such a way that they will provide large spatial detection coverage, detect contaminant intrusions promptly and minimize the system’s vulnerability. Potential health effects have spurred research on the technical guidelines for designing sensor networks, i.e. finding a set of locations that can provide, with the highest possible certainty, information on the water quality status over the entire piping system.

The starting point of most Sensor Network Design (SND) methodologies is the definition of contamination scenarios characterized by timing and location of pollutant intrusions. Their resulting impacts are evaluated by hydraulic and water quality (WQ) simulations. Optimal sensor locations can then be determined to detect these intrusions based on one or more objective functions. Core assumptions in these methodologies are that the scenarios are representative of possible events and that the simulation models are representative of the system. Here, we examine the impact of an incomplete mixing model on SNDs.

Approaches to SND have evolved along with the capabilities of the computational solvers used to predict hydraulic variables (e.g. water demand, pressure, tank level, etc.) and water quality variables (e.g. water age, tracer, chlorine concentration, etc.). In an early SND study, Lee & Deininger (1992) analyzed the flow pathways using hydraulic simulations and maximized the percentage of the total water demand that was actually covered or monitored by a set of stations using integer programming. In a subsequent study, Kumar et al. (1997) eliminated the need for integer programming by performing hydraulic simulations that produced a means of ranking nodes according to the demand covered and selected the locations with highest rank as monitoring stations. They also demonstrated that the optimal sensor location lay on the node with the lowest water quality.
(or with the lowest chlorine concentration as predicted by simulations), thus recognizing the role of water quality behavior in SND.

Kessler et al. (1998) fully employed the results of their water quality simulations as the primary input data for optimal placement of monitoring stations capable of detecting accidental contaminations in water networks. Their optimal locations were a function of the volume of contaminated water consumed prior to detection, defined as the Level of Service (LOS). A small LOS prevented the pollutant from further propagating to other zones of the network at the expense of more monitoring stations. A binary pollution matrix summarized the contamination status for each contemplated scenario and was the basis for formulating a set-covering optimization problem. In a similar fashion, Ostfeld et al. (2004) implemented a Minimum Hazard Level (MHL) within the optimization procedure in order to examine the impact of dilution in the water network. Additionally, the sensing capacity of monitoring stations (Minimum Detection Level, MDL) was taken into account as SND parameters.

Research efforts subsequent to these have dealt with solving multi-objective formulations, reducing the computational time required when the method was applied to real-world water networks, and more realistically considering water utility needs. Ostfeld et al. (2008) summarized the state-of-the-art methodologies used in solving the problem presented by various researchers in the Battle of Water Sensor Networks (BWSN). Despite the latest advances in SND methodologies, a research gap remains with respect to the integration and evaluation of more accurate water quality solvers. Using a WQ model that assumed incomplete mixing at cross junctions, Romero-Gomez et al. (2008a) produced steady-state-based designs. They found that making such a mixing assumption had significant effects on the layout of sensors over a small network as well as over a highly-interconnected piping system with multiple cross junctions.

The present study thoroughly examines the impact of water quality solvers on sensor network designs intended for networks with multiple types of four-way junctions subject to transient hydraulics. In the present study, in an effort to evaluate alternative contamination events under the incomplete mixing assumption, we utilized the formulation developed by Ostfeld et al. (2004) in conjunction with the improved WQ model, AZRED. These outcomes are then compared against designs based on the conventional complete mixing assumption.

### INCOMPLETE SOLUTE MIXING AT JUNCTIONS

Water quality solvers are mathematical models that describe the underlying transport mechanisms of constituents in piping networks. These solvers have been developed on the basis of the convenient but potentially erroneous assumption that solute mixing is complete and instantaneous at the pipe junctions. Figure 1(a) depicts a typical cross junction in order to explain solute transport at a cross junction involving two adjacent inflows and two outflows. The tracer concentrations in both outlet pipes would be equal under the complete mixing assumption. As Fowler & Jones (1991) first pointed out, completely mixed conditions can differ greatly from measurements obtained directly during laboratory and field observations. By making preliminary computational and experimental comparisons of impinging flows with different qualities, van Bloemen Waanders et al. (2005) found that the assumption of complete mixing failed to represent the mass transport at cross junctions. Romero-Gomez et al. (2006, 2008b), Ho et al. (2006) and Austin et al. (2008) examined tracer mixing at cross and double-T junctions involving two adjacent inflows and two outflows; all came to the general conclusion that the expected complete mixing does not take place under most flow configurations. Instead, the limited spatio-temporal interaction that occurs between the two inflows actually produces a rather incomplete tracer mixing, with most of the incoming tracer being deflected to the adjacent outlet and only a small amount crossing the junction to the opposite outlet.

In an effort to improve on solvers that rely on the perfect mixing assumption, researchers at the University of Arizona recently incorporated incomplete mixing at junctions into the WQ model. Named AZRED (“AZ” refers to the University of Arizona and “RED” means “network” in Spanish), this solver integrated experimental datasets of tracer mixing at single cross, double-T and Y junctions (as shown in Figure 1) under multiple flow configurations.

The experimental databases summarized the level of mixing as a function of (i) the ratio of incoming flow rates, (ii) the
ratio of outgoing flow rates and (iii) the type of junction. These databases, along with functions designed to retrieve and to process the data, were incorporated into the EPANET C/C++ code. EPANET performs extended period simulations of the hydraulics and water quality of pressurized water systems (Rossman 2000). Thus, AZRED takes advantage of most of EPANET’s built-in functions in order to implement incomplete mixing at the junction types depicted in Figure 1. The performance of AZRED on a laboratory-scale water network has been experimentally examined and validated (Song et al. 2009) and applied to a large-scale network (Choi et al. 2008). The present study uses AZRED as the WQ modeling tool for providing the input data needed to carry out sensor network designs over two exemplary water networks.

**PROCEDURE**

**Pollution matrix**

As the first step of the methodology adopted from Ostfeld et al. (2004), a pollution matrix is constructed. This pollution matrix has binary elements in which the ith row lists all nodes contaminated when an intrusion at node i occurs and, consequently, its jth column lists all pollution nodes that contaminate the jth node. Two parameters are defined to establish the criteria for determining whether a node becomes contaminated or not. First, a node is considered contaminated when its concentration becomes greater than a Minimum Hazard Level (MHL). Second, it is assumed that a contaminant continues to propagate throughout the water network until a volume of contaminated water (the Level of Exposure, LOE) is consumed.

Once these two parameters have been established, then the pollution matrix can be constructed in the following manner. Here, we assume that each network node will be impacted individually at the simulation’s starting time. This procedure is general and allows for any contamination location and timing.

1. Set all the elements \( l_{ij} \) of the pollution matrix equal to zero.
2. Introduce a contaminant intrusion at the ith node of the water network at time zero.
3. Carry out an Extended-Period WQ Simulation (EPS) using EPANET or AZRED. At each quality time step, calculate the cumulative volume of contaminated water consumed (withdrawn from the network) at nodes in which the concentration is greater than the MHL. Enter \( l_{ij} = 1 \) if the concentration at the jth node is greater than the MHL.
4. If the total contaminated water volume consumed is equal to or greater than LOE, stop the simulation; otherwise, continue until the end of the simulation.
5. Reinitialize the network and return to Step 2 to set the intrusion at the \((i + 1)\)-th node.
6. Repeat Steps 2–5 for the \( N \) nodes considered to be potential intrusion locations.

**Mathematical formulation**

After the pollution matrix is constructed, the next step is to find the minimum number of sensors that detect a prescribed proportion of all of the intrusions within the water network.
(100% for the examples in this study, full coverage). This problem type, known as a Set Covering Problem (SCP), has wide applications and has been the subject of numerous studies (e.g. Beasley 1987; Beasley & Jörnsten 1992). The SND formulation seeks to find the minimal set of columns (sensor locations) that include at least one value of 1 in all the rows of the pollution matrix (i.e. nodes impacted by each intrusion). The decision variable takes the form of a binary vector \( X_j \) that indicates all sensor locations. If a sensor should be installed at node \( j \), the \( j \)th element of vector \( X \) is equal to 1; otherwise \( X_j \) is equal to 0. Mathematically, the problem can be defined as follows:

\[
\text{Minimize } \sum_{j=1}^{N} C_j X_j
\]

subject to \[ \sum_{j=1}^{N} \lambda_{ij} X_j \geq 1 \quad i = 1 \ldots N \]

\[ X_j (0, 1), \quad j = 1 \ldots N \]

where \( \lambda_{ij} \) is the binary matrix \((i,j)\) term. According to Ostfeld et al. (2004), in order to maximize the overlapping of sensor coverage (so that a contaminant intrusion may be detected by more than one sensor, if possible), a cost \( C_j \) can be assigned to each column by using the following formula:

\[ C_j = G - \sum_{i=1}^{N} \lambda_{ij} \text{ where } G > N. \]

This minimization problem is solved by using the software Premium Solver for Excel (Frontline Systems, Inc., Incline Village, NV). The solution procedure is carried out on a user-friendly spreadsheet interface on which solution settings are selected (constraints, decision variables, variable type, etc.). Given that the SCP problem is non-convex and NP-hard, a global minimum is not guaranteed.

**Modeling modifications to account for incomplete mixing**

Conventionally, junctions in hydraulic models are represented as single nodes at which links of different diameter may connect. However, to account for incomplete mixing, an AZRED user should modify the configuration of four-way junctions. First, the base demand at the junction node is set to zero. Next, four nodes, all at the same elevation and also having no water demand, are added around the junction. These dummy nodes connect to the junction node through short links (relative to the overall network size) with diameters that fall between the smallest and largest diameters of the links that are physically connected to the junction. These changes are consistent with the physical conditions under which the incomplete mixing experiments were conducted.

In AZRED’s Graphical User Interface (or GUI), the added pipes are drawn to resemble the junction type that they represent, i.e. an N-type arrangement (as shown in Figure 1(b)) should follow a straight line from which two nearly perpendicular links branch out in opposite directions. The combination of angles formed by the links directs AZRED to search for the level of mixing in the correct database according to the junction type.

**ILLUSTRATIVE EXAMPLES**

**A simple network configuration**

Net1 of the EPANET examples consists of 12 pipes, 9 nodes (only 8 with base demands), a water source, a pump and an elevated tank. The base demands are subject to a transient pattern throughout a simulation period of 24 h. The remaining characteristics are readily accessible for downloading from the EPANET software package. A few of the original network’s features were modified. The initial quality at all nodes was set equal to zero so that only the injected pollutant had an effect on nodal concentrations. Node 22 was modified to conform to the AZRED requirements for four-way junctions. Since the remaining junctions connected two or three pipes, they were not changed. Eight nodes are defined as possible contaminant sources. The resulting network configuration is shown in Figure 2.

Pollution matrices were generated for MHL values of (in \( \text{mg L}^{-1} \)): 0.5, 1, 1.5, 2, 3 and 4 and LOE values (in \( \text{m}^3 \)) equal to 100, 200, 500, 1000 and 2000. Thus, pollution matrices were prepared and corresponding optimal SNDs were determined for 30 MHL–LOE combinations (six MHL values by five LOE values). Each intrusion was represented by a constant injection mass rate of 10 000 \( \text{mg min}^{-1} \) over the entire
simulation period, starting from time zero for all intrusion scenarios.

Comparing EPANET and AZRED results showed that the designs differed following one of three cases: (i) the minimum number of sensors is different, (ii) the minimum number of sensors is the same but some locations are different and (iii) the minimum number of sensors and the optimal locations are the same but some sensors have different detection domains.

Of the 30 designs that were obtained for Net1, five (indicated as MHL–LOE combinations A, B, C, D and F in Figure 3) required a different number of sensors owing to the WQ solver used, whereas one design (combination E) required the same number of sensors but they were placed at different locations. In the former five results, EPANET-based designs required more sensors than AZRED-based designs, with the greatest difference (two stations) resulting from combination A (MHL = 1.5 mg L⁻¹, LOE = 1000 m³). Such a difference was deemed significantly large with respect to the maximum number of sensors possible for this network (eight possible stations). Under incomplete mixing, the presence of the cross junction at node 22 produced a tendency for the contaminant to be concentrated at one of the two downstream nodes (nodes 23 and 32). Greater levels of mixing (as produced by EPANET) will cause the contaminant to spread more widely at lower concentrations, making detection more difficult and requiring more monitoring stations.

This network is simple enough to allow for a detailed observation of the contaminant pathways. For example, when contamination is introduced at node 21 (Figure 2), the complete mixing outcome was that clear (or non-contaminated) water from node 12 reduced the contaminant concentration in the pipes leaving node 22, dispersing the contaminant but at lower concentrations. In contrast, in the incomplete mixing analysis in AZRED most of the contaminant entering node 22 was deflected towards node 32 with little mixing. The transient concentration patterns at node 32, as obtained using both WQ solvers, are shown in Figure 4. Node 32’s AZRED concentration exceeds the MHL value of 2 mg L⁻¹ established for combination C while the complete mixing model does not. Consequently, this difference was reflected in the pollution matrices generated by both WQ solvers.

---

**Figure 2** | Example network 1 (modified Net1.net from EPANET examples).

**Figure 3** | Number of stations based on EPANET and AZRED at different MHL–LOE combinations. (A) MHL = 1.5 mg L⁻¹, LOE = 1000 m³, (B) MHL = 1.5 mg L⁻¹, LOE = 2000 m³, (C) MHL = 2 mg L⁻¹, LOE = 500 m³, (D) MHL = 2 mg L⁻¹, LOE = 1000 m³, (E) MHL = 2 mg L⁻¹, LOE = 2000 m³, (F) MHL = 3 mg L⁻¹, LOE = 100 m³.

**Figure 4** | EPS contaminant concentration at node 32, based on EPANET and AZRED solvers. Shaded area indicates concentrations higher than 2 mg L⁻¹ (dotted horizontal line, MHL setting for combination C).
solvation of parameters (combination C, Figure 3) requires only three monitoring stations in order to achieve full detection coverage (Figure 5(b)). In a broader context, such a difference directly increases the cost of the network. Detection domains in Figure 5 show which nodes are covered by each station; for instance, injections at nodes 11 and 12 will be detected by the station placed at node 11 for both EPANET and AZRED-based designs.

Figure 5(c, d) depict conditions (combination E: MHL = 2 mg L\(^{-1}\), LOE = 2000 m\(^3\)) under which the number of required stations remained the same but their locations were different due to the mixing assumption. To provide full coverage under this combination, the monitoring station that was placed at node 11 using the EPANET results was shifted to node 21 based on AZRED outcomes. In addition to providing full coverage, a monitoring station at node 21 increases sensor overlap as compared to the optimal design that included a sensor at node 11. A monitoring system’s ability to detect contaminant intrusions at more than one station is particularly desirable when the probability of sensor failure is taken into consideration as part of the SND methodology.

Figure 5(e, f) present the EPANET- and AZRED-based designs, respectively, for MHL = 1 mg L\(^{-1}\) and LOE = 500 m\(^3\), in which the number and location of stations are the same but the detection domains are different. Although the EPANET-based design maximized sensor overlap as compared to the optimal design that included a sensor at node 11. A monitoring system’s ability to detect contaminant intrusions at more than one station is particularly desirable when the probability of sensor failure is taken into consideration as part of the SND methodology.

**Net3 from EPANET examples**

Net3 is available with the EPANET download and consists of 117 links, 92 nodes, 3 elevated tanks and 2 pumps that connect to 2 reservoirs (a lake and a river). Eight nodes from the network are four-way junctions (Table 2) and were modified to each of the four alternative AZRED junction types (Figure 1). The location and characteristics of the modified junctions are listed in Table 2. Contamination scenarios were developed with each of the 84 nodes being equally likely as the contamination source. The quality time step and parameter were set at 2 min and “Chemical”,
respectively. The injection mass rate was set equal to 20000 mg min\(^{-1}\) for all the scenarios. The settings for the MHL value (in mg L\(^{-1}\)/C\(_0\)) were 1, 2, 5, 10, 20, 50 and 100, and the LOE settings (in m\(^3\)) were 50, 100, 200, 500, 1000, 2000 and 5000. Pollution matrices were thus developed for all 49 MHL–LOE combinations.

The tradeoff between the level of exposure and the number of monitoring stations for MHL = 1 mg L\(^{-1}\) is presented in Figure 6(a). The LOE level indirectly implies the time till a contamination event is detected. Higher LOE values allow for more consumption (or usage) of contaminated water and, consequently, longer times to detection and further pollutant propagation. Thus, fewer stations are needed and, as seen in Figure 6(a), the number of stations decreases monotonically with increasing LOE. This decreasing trend is consistent with Ostfeld et al. (2004) and occurred for all sensor arrangements.
regardless of the mixing model applied. However, AZRED simulations generally required one more station for most LOE values. Therefore, if complete mixing at junctions is applied, some nodes will not be covered by any of the sensors. This monitoring deficiency could become critical if the potential entry points were assigned a likelihood of contamination.

Figure 6(b) illustrates the tradeoff between the Minimum Hazard Level (MHL) and the number of monitoring stations for LOE = 100 m$^3$. Lower MHLs will result in more sensitive sensors that can detect at lower concentrations. Thus, more dilute contaminations could be identified and fewer stations are likely to be necessary for the same protection. At the other end of the spectrum (MHL > 20 mg L$^{-1}$ in this exemplary case), few nodes become contaminated and monitoring stations are only needed at those locations.

Comparing SNDs for the alternative WQ modeling assumptions for the 49 MHL–LOE combinations, differences in designs occurred primarily for conservative parameter sets, i.e. when detection capacity was enhanced (MHL ≤ 10 mg L$^{-1}$) or when the propagation of polluted water was restricted (LOE ≤ 200 m$^3$) (Table 3). No specific trend was identified in relation to the minimum number of sensors needed for either WQ solver.

To understand the spatial differences in SND, the frequencies that a node was selected by both water-quality-model-based SNDs were computed and compared for the 49 MHL–LOE combinations. It was found that nodes with the highest frequency corresponded to the locations that were more sensitive to the solute mixing model used to generate the pollution matrices. Figure 7 shows that sensitive nodes generally lie near four-way junctions, a finding that suggests the effect will be stronger at locations in close proximity to a junction and weaker further downstream from four-way junctions. Circle sizes in Figure 7 represent the magnitude of the difference in frequencies; the small circles indicate the node was selected one more time by one of the models and the larger circles are differences greater than one. This analysis also reveals that certain patterns will occur according to the water quality model used. For instance, nodes 211 and 213 were consistently interchangeable (as were nodes 247 and 251) due to the influence of a YU-type junction (node 255) and the elevated tank (tank 2) that produced highly dynamic hydraulics in this region.

### Table 2

<table>
<thead>
<tr>
<th>Node ID</th>
<th>Type</th>
<th>Junction diameter (D)</th>
<th>Dimensionless distance (L/D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>121</td>
<td>YN</td>
<td>24</td>
<td>2.0</td>
</tr>
<tr>
<td>120</td>
<td>N</td>
<td>10</td>
<td>1.0</td>
</tr>
<tr>
<td>119</td>
<td>X</td>
<td>20</td>
<td>N/A</td>
</tr>
<tr>
<td>115</td>
<td>YN</td>
<td>10</td>
<td>3.0</td>
</tr>
<tr>
<td>193</td>
<td>N</td>
<td>12</td>
<td>2.5</td>
</tr>
<tr>
<td>189</td>
<td>X</td>
<td>12</td>
<td>N/A</td>
</tr>
<tr>
<td>255</td>
<td>YU</td>
<td>10</td>
<td>1.0</td>
</tr>
<tr>
<td>111</td>
<td>YN</td>
<td>12</td>
<td>2.0</td>
</tr>
</tbody>
</table>
CONCLUSIONS

This study incorporated a newly developed incomplete solute mixing model in an optimal sensor network design. We analyzed various contamination and detection scenarios over two exemplary networks. The first example system allowed for detailed descriptions of the contaminant pathway and its ultimate impact on each design while the second exemplary network provided the global trends of the impact on the sensor configurations due to the WQ solvers. Application results demonstrated that, while the quantitative impact of incomplete mixing varies, EPANET-based designs did not provide the desired level of protection under incomplete mixing conditions. For example, several AZRED-based designs allocated more monitoring stations than EPANET-based design; this strongly implies that an EPANET-based sensor arrangement will fail to detect some contaminant intrusions.

Overall, these results demonstrate that water quality modeling tools and their accuracy play a central role in SND. Developing and implementing SND approaches should necessarily entail a continuous re-examination in view of the new advances in water quality modeling in pressurized piping systems. More accurate solute transport predictions will make SNDs account for a better representation of the water quality behavior of a network. This in turn should increase the effectiveness and reliability of early warning monitoring systems as well as the likelihood that intentional or accidental contamination events can be detected early enough to initiate countermeasures that can protect water customers.

ACKNOWLEDGEMENTS

This work is supported by the Environmental Protection Agency/Department of Homeland Security (under Grant No. 615383D).

REFERENCES


