SMaRT-OnlineWDN D3.1: Detailed specification of the performance criteria and application of sensor placement

Olivier Piller, Jochen Deuerlein, Thomas Bernard, Fereshte Sedehizade

SPONSORED BY THE

Federal Ministry of Education and Research

Deliverable 3.1

Detailed specification of the performance criteria and application of sensor placement

Dissemination level: Public

WP3
OPTIMAL LOCATION OF SENSORS
15 November 2013

SMaRT-Online WDN

Online Security Management and Reliability Toolkit for Water Distribution Networks

ANR reference project: ANR-11-SECU-006
BMBF reference project: 13N12180

Contact persons

Fereshte SEDEHIZADE	Fereshte.Sedehizade@bwb.de
Olivier PILLER	Olivier.Piller@irstea.fr

WP 3 – OPTIMAL LOCATION OF SENSORS

D3.1 Detailed specification of the performance criteria and application of sensor placement

List of Deliverable 3.1 contributors:

From Irstea

Olivier Piller (olivier.piller@irstea.fr) [WP3 leader]

From 3S Consult

Jochen Deuerlein (deuerlein@3sconsult.de)

From IOSB

Thomas Bernard (thomas.bernard@iosb.fraunhofer.de)

From BWB

Fereshte Sedehizade (Fereshte.Sedehizade@bwb.de)

Work package number	3.1	Start date:		01/04/2012
Contributors	Irstea	3S Consult	IOSB	BWB
Person-months per partner	3	1	1	1
Keywords Sensor placement – Computer-aided designs – Models – Early warning detection system – Calibration – Multi-objective optimisation				
Objectives To specify suitable formulations for optimal sensor placement for both early warning detection system design and online transport equation calibration				

Summary

Drinking water distribution networks are exposed to malicious or accidental contamination. Several levels of responses are conceivable. One of them consists of installing a sensor network to monitor the system in real time. Once a contamination has been detected, it is also important to take appropriate counter-measures. The *SMaRT-Online^{WDN}* project relies on modelling to predict both hydraulics and water quality. An online model makes it possible to identify the contaminant source and perform a simulation of the contaminated area. The sensor system is intended for detection by an early warning system and for online calibration of the transport model.

The main objective of this deliverable report is to specify which performance criteria should be considered to place water quality and water quantity sensors for both early detection and model calibration.

Firstly, a review of previously published research on water security and model calibration is presented. Then, the experiences of partners in the two previous German National projects, *STATuS* and *IWaNet*, and in the FP7 European project, SecurEau, are reported. Next, formulations and objectives for early warning detection are proposed. Following, problem formulations that aim to minimize the estimator variance for calibration are specified. Finally, a summary of the conclusions is given.

List of Figures:

Contents:

1 Background

Drinking water distribution networks are exposed to malicious or accidental contamination. As a consequence, they should be protected. One useful means of protection is to resort to an early warning detection system by placing water quality sensors at strategic and important sites. Computer-aided designs by modelling and analysing the fate of contaminations are powerful aids for decision makers.

There are two main types of online water quality sensors: a biosensor system (very accurate but expensive), or a multi-probe sensor with standard outputs (less sensitive but cheaper). The first ones are actually only employed for the monitoring of the water quality at “critical facilities”. The second ones may be installed at many points in the system. The early warning system should comprise both kinds of sensors. Obviously, only placement for sensor of the second kind (cheap with standard output) is sought for the detection system design.

Once a water contamination incident has occurred, it is important to take appropriate counter-measures. Also, waiting for confirmation might not be desirable, as it would be more prudent to firstly take appropriate responses to the contamination threats, and start to study the problem sources and extent. In the *SMaRT-Online*^{WDN} project, an important tool is a reliable model for water quantity and for water quality. At regular short time steps (*e.g.*, 5 min), the hydraulic model should be calibrated for the demand, and the transport model for the velocity. For the calibration of this online model, there is a need for measurement data. Where to place sensors and meters to improve the online demand calibration is an important question.

The problem of water quality and quantity sensors placement is schematically represented in Figure 1. Online simulation and alarm generation are important tasks for the *SMaRT-Online*^{WDN} project. The initial sensor network consists of multi-probe sensors (yellow circles) and biosensors (pink crosses). The two brown circles represent additional sensor placements to supplement an existing sensor network. In this project, the sensor system is considered for improving both the early warning detection and the online calibration of the transport model.

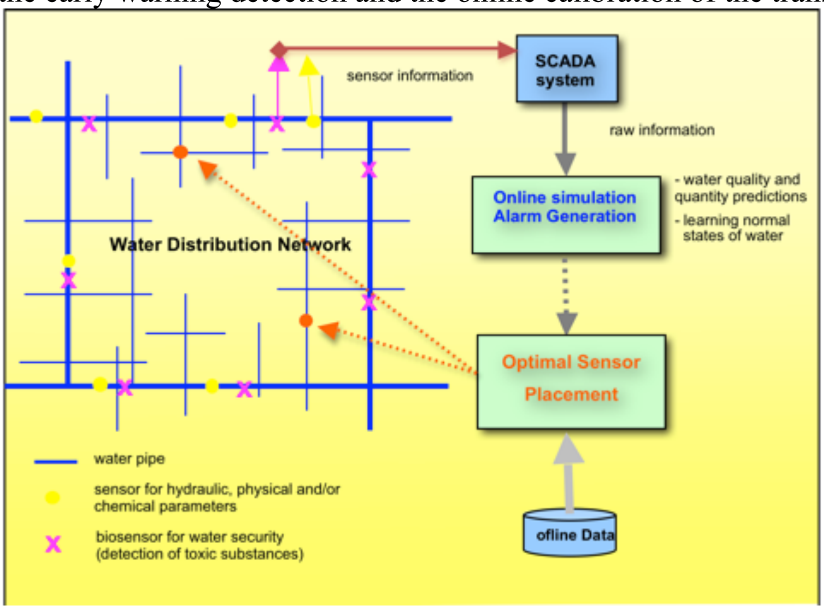

Figure 1: Double goals of early warning detection and online calibration for optimal sensor placement in *SMaRT-Online*^{WDN}.

In the literature, there are many papers that were published on both subjects in an offline context. Recently, a few were proposed for the pseudo-real-time context but never with such a short time step for calibration update. It is also relevant to report experience by *SMaRT-Online^{WDN}* partners in previous National and European projects. We will present first a short review of the literature then experience by partners.

1.1 Main previous published research in water security

1.1.1 Demand coverage method.

One of the first formulations is the demand coverage method (DCM) by Lee and Deininger (1992) for placing monitoring station (MS). They propose an integer linear programming problem with node covering constraints Eq. (1) for the maximisation of the nodal demand covered by the monitoring stations. A MS covers a node if a sufficient fraction of water flows from this node to the MS for a demand scenario. The design relies on the principle that the water quality (trihalomethanes, bacteria, and chlorine residual) decreases with time and distance from the source. That is, good quality at a water sample entails a good quality at immediate upstream nodes. They extend the formulation to a multi-flow maximum coverage problem to handle multiple scenarios of demand. Assembling coverage constraints requires for each scenario: (1) to be run at steady state; (2) to make a flow path analysis that results in the water fraction matrix; then, (3) to indicate if a node is covered, given a minimum water-fraction criterion (*e.g.*, 50%) in the K knowledge-carrying matrix (Eq. (1)).

$$\max f(x,y) = \sum_{j=1}^{n} d_j y_j$$

(DCM)
subject to:

$$\sum_{i=1}^{n} x_i \le N_s \text{ (cardinality constraint)}$$

$$\mathbf{K}^T\mathbf{x} - \mathbf{y} \ge \mathbf{0}_n \tag{1}$$

$$x_i \in \{0,1\} \text{ and } y_j \in \{0,1\}$$

Where x_i is one if there a sensor at node i, else 0; y_j is an auxiliary variable which indicates if node j is covered by the MS or not; and d_j is the demand at node j. The maximum coverage problem is NP-hard (non-deterministic polynomial-time hard). The greedy algorithm for DCM at each stage chooses a set, which contains the maximum weight of uncovered elements. This algorithm achieves an approximation ratio of 1 – 1/e. The greedy algorithm has performance guarantee when the maximum coverage problem has the same unit cost as the sensor (only cardinality constraint), but this is not generally the case when costs are different (when the cardinality constraint is replaced by a budget constraint). An improved greedy algorithm may be applied as in Khuller *et al.* (1999). Lee and Deininger (1992) solve the demand coverage problem with an ILP solver, while Kumar *et al.* (1997) use a greedy algorithm and Al-Zahrani and Moied (2001) a genetic algorithm (GA).

One drawback of the method for its application to water security is that it is only based on water quantity under steady state. Even if multi-flow scenarios are proposed, transient transport equations are needed to fully comprehend the complexity of water quality indicators and contaminant propagation in the water distribution network. Woo *et al.* (2001) run a water quality model, then modify the objective costs of (DCM) to give a higher weight to nodes with low disinfectant concentration.

Several authors propose to go further and account for the travel time between nodes to design an early warning detection system (EWDS). Regarding the contaminant plume, monitoring stations should be designed to raise an alarm within an acceptable time to limit the pollution domain. Suggested solutions are: (1) solving a set covering problem for detection of a pollution domain within a maximum consumed polluted water volume criterion or (2) solving a MILP problem related to the p-median problem that is devoted to the optimisation of impact factors such as the average time to detection. The EWDS design requires generating contaminant events with use of an offline transport model for the contaminant plume dynamics and impact assessment.

1.1.2 Coverage of pollution events.

Kessler *et al.* (1998) formulate a set-covering problem (SCP) to find the optimal layout for detecting a random pollution event. The optimal design satisfies a given level of service to the consumers that is defined by the maximum volume of consumed polluted water prior to detection. The method involves the construction of an auxiliary graph with travel time as arc weights; the Floyd-Warshall shortest path finding algorithm is used to demarcate the domain of pollution arising out of the Level of Service. Finally, a set cover problem is solved with the aim of minimising the design cost. They assign lower costs at nodes with higher detection redundancy. The methodology is demonstrated on a small illustrative case and on a midsize water network. The 0-1-pollution matrix is built on the auxiliary graph. In the paper this matrix has as many rows as non-resource nodes in the graph. Ostfeld and Salomons (2004) observe that it is preferable to calculate the domain of pollution with the complete transport model instead of the auxiliary graph in order to better take into account the water dilution and the water quality changes. Their notion of randomised pollution matrix (P Eq. (2)) is a generalisation of the Kessler *et al.* (1998) ones. A contamination event may comprise several intrusions and start at any time. Quality of the design solution is studied with regards to its likelihood of detection and the detection redundancy.

$$\min_{x_i \in \{0,1\}} c(x) = \sum_{i=1}^{n} c_i x_i$$

(SCP)
subject to:

$$\mathbf{P}^T \mathbf{x} \geq \mathbf{1}_m \tag{2}$$

The set-covering problem is closely related to the maximum coverage problem and they are both NP-hard. The SCP formulation searches for a safe cover for every pollution scenario. Because of the budget constraint, realistic sensor networks are unlikely to satisfy such a severe requirement; some (location, scenario) pairs may be left undetected if they impact fewer nodes or population. Uber *et al.* (2004) introduce an auxiliary binary variable $y_{ik} = 1$ if location i is protected for scenario k by at least one sensor, and zero otherwise. They deduce the following maximum coverage problem (MCP) with weaker assumption for the cover:

$$\max f(x,y) = \sum_{j=1}^{m} w_j y_j$$

(MCP)
subject to:

$$\sum_{i=1}^{n} x_i \leq N_s \text{ (cardinality constraint)}$$

$$\mathbf{P}^T \mathbf{x} - \mathbf{y} \geq \mathbf{0}_m \tag{3}$$

$$x_i \in \{0,1\} \text{ and } \mathrm{y}_j \in \{0,1\}$$

Solution of the set-covering problem is achieved by Kesser *et al.* (1998) using a graph heuristic algorithm suggested in the Christofides's book (1975), by Ostfeld and Salomons (2004) using GAs and by Uber *et al.* (2004) with a greedy algorithm.

1.1.3 P-median problem.

Propato *et al.* (2005) propose a MILP (Mixed Integer Linear Programming) formulation. Their generic objective function consists of costs that are impact factors for an EWDS such as averaged time to detection, likelihood of detection, etc. The formulation of this problem is the following:

$$\min f(z,\delta) = \sum_{j=1}^{m} w_j z_j$$

(MILP)

subject to:

$$\sum_{j=1}^{n} \delta_j \geq n - N_s \quad \text{(cardinality constraint)}$$

$$z \geq \mathbf{P}^T(\delta - \mathbf{1}_n) + \mathbf{1}_m \tag{4}$$

$$z_j \geq 0 \text{ and } \delta_i \in \{0,1\}$$

By changing the variables ($\delta_i = 1 - x_i$ and $z_j = 1 - y_j$) it is possible to show that this MILP formulation is close to the MCP problem of Eq. (3). Nevertheless, there are two small differences: the binary constraint on the auxiliary variable y_j is relaxed to be $y_j \leq 1$ that is more tractable; and conceptually the P matrix is built differently from Eqs. (4) and (3). No simplification upon a given level of service, *i.e.*, no maximum volume of consumed polluted water prior to detection is defined for the P assembling. The factor of adjustment is rather the minimum objective function value given the number of sensors. Berry *et al.* (2006) propose another variant with a formulation mathematically equivalent to the p-median facility location problem. They report scalability challenges due to 1) the need to use a large number of attack scenarios to be representative spatially and temporally and 2) the use of a small water quality-reporting step. They solve with a Greedy Randomised Adaptive Search Procedure (GRASP) and they quantify how close to optimality the solution is with the MIP Cplex solver or with LP bounds. Propato and Piller (2006) also solved with the Cplex MIP solver and observe near optimality for a greedy algorithm.

To discuss the convenience and the potential of each approach regarding designing an EWSN, the Battle of the Water Sensor Networks (BWSN) was held as part of the Eighth annual WDSA symposium in Cincinnati in 2006. Among diverse conclusions of the common paper by all the BWSN participants (Ostfeld *et al.*, 2008), one may conclude that there is not a single formulation/solving method solution that was superior to the others; better solutions were ones combining strength of the algorithm with engineering judgement and intuition. Interestingly, they suggest several future research directions such as: definition of the pollution matrix and better contaminant event generation to better represent the network complexity; graph simplification or water quality model simplification without reducing the model prediction power; dual use of sensors (not only for security goal but also for model calibration, *etc.*); inclusion of risks; sensor reliability and alarm generation with false positive and false negative classification; and finally incorporation of operational conditions. To a greater or lesser extent, all these research directions are explored in the *SMaRT-Online*^{WDN} project.

1.2 Main previously published research in model calibration for WDNs

1.2.1 Parameter calibration problem.

Network parameters that are used in the hydraulic and transport models are often rough estimates. This is mainly because the distribution network is buried underground since a long time and the consumption at any time and location is random. The measurement values (pressure, flow, tank levels, concentration, conductivity, *etc.*) may be used to calibrate the parameters of both the hydraulic and the transport models, and to estimate the unknown state of the network. Given a set of network measurements, is it possible to derive all water quality and quantity unknowns of the network? Carpentier and Cohen (1991) defined the observability problem as the one of determining whether the available set of measurements provides sufficient information for the state estimation. They define two levels of observability (topological, which may be assessed with graph theory, and algebraic, by means of analysing the sensitivity matrix). The quality of the estimation of these parameters, which drives the quality of the predictions, depends on the position, number and nature of the measurements. This choice must ensure the observability of the network but also prevent small errors in measurement resulting in incorrect estimations of the parameters (Piller *et al.*, 1999).

Several fitness functions may be selected for the parameter calibration. For a review of them, application of hydraulic modelling and suggestion of additional fitness functions and entropy-based criteria one may consult de Schaetzen (2000) or still Savic *et al*. (2009). A least-squares formulation that minimises the deviations between some predicted values and corresponding observations is a standard approach for overdetermined systems. Its weighted form has the advantage of dealing with errors in observations. When the latter ones come from exponential family distribution (*e.g.*, the Gaussian) the weighted least-squares (WLS) minimisation problem corresponds to a maximisation of a likelihood function. The WLS problem reads:

$$\text{(WLS)} \qquad \min_{\mathbf{x}\in\mathbb{R}^n} \ \tfrac{1}{2}\int_{t_0}^{t_f}\left(\mathbf{S}\mathbf{y}(\mathbf{x},t)-\mathbf{y}^{mes}(t)\right)^T \mathbf{W}\left(\mathbf{S}\mathbf{y}(\mathbf{x},t)-\mathbf{y}^{mes}(t)\right)dt$$

subject to:

$$\underline{\mathbf{x}} \le \mathbf{x} \le \overline{\mathbf{x}} \tag{5}$$

Where **x** is the vector of parameters to determine; **y**(**x**,t) is the hydraulic and transport state that is implicitly defined by the hydraulic and transport equations; **S** is the selection matrix to select the state vector components that corresponds to the measurements; $\mathbf{y}^{\mathrm{mes}}$ is the vector of measurements; and **W** is a diagonal weight matrix. Most authors have considered a simpler form of (WLS) without the time dependency (Kapelan *et al.*, 2002) or with a quadrature formula for the integral (Piller, 1995).

Preis *et al.* (2011) use the Huber function to modify the (WLS) least-squares criterion for large residuals to be least absolute deviations (L^1 norm). This way the parameter estimation is more robust against outliers. More efficient solving methods are gradient type methods that use derivatives of the criterion; for example, Piller (1995) applies the Levenberg-Marquardt (LM) method for solving WLS calibration problems for water distribution networks. However, as the criterion may exhibit several local minima (and maxima) a genetic algorithm (GA) was preferred for the first iteration steps in a hybrid GA/LM approach (Kapelan, 2002). In that situation, Piller *et al.* (2010) propose to convexify the LS criterion with addition of a Tikhonov regularisation term that penalises departure from a prior solution.

For estimating the Jacobian of the residuals or calculating the LS gradient several methods may be explored.

1.2.2 Sensitivity estimations.

Four main approaches may be used for sensitivity estimations: finite differences, automatic differentiation, sensitivity equations and adjoint method. The finite difference techniques can be used with a large number of hydraulic software to approximate the sensitivity; they are easy to implement but suffer from a lack of accuracy. Automatic differentiation is a family of techniques for computing the derivatives of a function defined by a computer program. Even though this method is accurate and fast, it produces lengthy and complex computer codes. Interestingly, for the steady state problem, the sensitivity equation method leads to explicit formulations for the derivatives of hydraulic state with respect to the demand and roughness hydraulic parameters (Piller, 1995). For calculating a row of the Jacobian, a linear system is solved that possesses the same structure (same Schur matrix) as the original problem. Fabrie *et al.* (2010) derive ad hoc sensitivity equations from the direct transport-reaction problem. For each pipe, the formulation is a 1D hyperbolic partial differential equation with a similar advection and reaction structure to the original problem. Solving economies are therefore possible. Gancel *et al.* (2006) solve, in coupled manner, the slow transient hydraulic equations and its sensitivity equations. Finally, the adjoint method, largely used in ground water hydrology, consists of solving the adjoint equations for determining Lagrangian multipliers, which help to calculate the least-squares criterion gradient. Kapelan *et al.* (2003) report this method yields accurate results and is also time-efficient. In the case of steady state both methods are equivalent.

Sensitivity analysis allows the determination of how "sensitive" our model is to change in the values of these parameters. They have been successfully applied to hydraulic sensitivity (Bargiela and Hainsworth, 1989), hydraulic calibration (Piller, 1995; de Schaetzen, 2000; Kapelan, 2002; Gancel, 2006) and hydraulic and water quality sampling design (Bush and Uber, 1998; Piller *et al.,* 1999). For the latter, it gives the most sensitive nodes where it would be most profitable to perform the necessary measures for calibration.

1.2.3 Minimising the variance of estimators.

From optimal design in Statistics (*e.g.*, Optimal_Design, 2013), there are various forms of optimality criteria that are used for sampling design. If $\hat{x}$ is a least-squares estimate of x (a solution of problem WLS), a deviation from y^{mes}, or equivalently another measurement error realisation, will lead to a different estimate. Under the assumption of a mean-zero independent and identically distributed (iid) error, a first-order estimate of the Covariance matrix of the $\hat{x}$ parameters is given by:

$$\mathbf{V}_{\hat{x}}(\delta) = s^2 \left(\mathbf{J}^T \mathbf{S}_\delta{}^T \mathbf{W}_\delta \mathbf{S}_\delta \mathbf{J} \right)^{-1} \quad (6)$$

Where the design δ is the set of all the observations; $\mathbf{S}_\delta$ has the same definition as in Eq. (5); $\mathbf{W}_\delta$ is composed of suitable positive weights for ensuring the weighted residuals to be iid; **J** is the Jacobian of **y** with respect to **x**; and s^2 is an estimate of the variance of the residual error.

An attempt to minimise the confidence volume for the parameters is then to seek the design δ that minimises the determinant of the Covariance matrix Eq. (6) or equivalently that maximises the determinant of the information matrix. This leads to the popular D-optimality design that consists of seeking:

$$\max_{\delta} = \det\left(\mathbf{J}_\delta{}^T \mathbf{W}_\delta \mathbf{J}_\delta \right) \quad (7)$$

Where δ is the decision variable to determine (where to locate measurements and their nature); $\mathbf{J}_\delta$ is the Jacobian matrix of the model residuals for choice δ; and $\mathbf{W}_\delta$ is the inverse of the covariance matrix for the corresponding observations. In practice the $\mathbf{J}_\delta$ matrix is extracted from the full Jacobian matrix at parameter estimates x_0:

$$\mathbf{J}_\delta = \mathbf{S}_\delta \mathbf{J}\left(x_0\right)$$

With $\mathbf{J}(x_0)$ is the Jacobian matrix of y in x_0. Similarly, the observation covariance matrix is extracted from the diagonal covariance matrix of potential measurements:

$$\mathbf{W}_\delta = \mathbf{C}_\delta^{-1} = \mathbf{S}_\delta \mathbf{C}^{-1} \mathbf{S}_\delta^T$$

The D-optimality criterion results in minimising the generalised variance of parameters.

Another criterion is the A-optimality that tries to minimise the trace of the determinant of parameter covariance matrix:

$$\min_\delta = tr\left(\mathbf{J}_\delta^T \mathbf{W}_\delta \mathbf{J}_\delta\right)^{-1} \tag{8}$$

This criterion results in minimising the average variance of the parameter estimates.

Bush and Uber (1998) suggest three simple sensitivity-based methods with simpler but practical measures (no determinant, no inverse matrix). Their findings were that both tracer and pressure measurements might improve parameter confidence by a factor of 2. Cheung *et al.* (2005) developed a multi-objective sampling method framework. Their method aims simultaneously to maximise the model precision (D-optimality), to minimise the number of sensors and to maximise the sensor spreading. Kang and Lansey (2010) analyse the trade-off between the model accuracy and the model precision for nodal demand uncertainty and pressure prediction uncertainty.

1.3 Experience of partners

1.3.1 The *STATuS* Project

Within the collaborative research project *STATuS* that was funded by the German Federal Ministry of Education and Research (13N10623, 2009 – 2013), a risk-based approach to water network security was elaborated by the partners. 3S worked on the development of a risk map of the entire water supply system. The risk map is created by the use of hydraulic simulation model and graph decomposition (Deuerlein, 2006; Deuerlein, 2008). Eventually, for each consumer (demand node) of the network, the relative risk of being affected by a contamination event was calculated as well as for each node and hydrant the risk of being the possible intrusion point of the contamination. For the development of the model, a number of contamination scenarios were simulated. As parameters, the input flow (contaminant and carrier medium), the place and time of intrusion and the duration were varied.

It is well known that the risk of an event is defined as the product of its probability of occurrence and the impact. In the context of *STATuS* a simple approach was taken where the impact is estimated by the total volume of water (flow) that leaves the intrusion node. In a steady-state version only the flows have to be considered whereas in the extended period version the integral over time must be calculated. One of the most important outcomes was that the risk is not a continuous function of input mass. There is a jump in the risk when the contaminant intrusion by pumping and flow reversal in the connected pipes reaches a superposed distribution main.

Other work that is related to *SMaRT-Online*^{WDN} concerns field tests that were carried out in order to get a better understanding of the transport and mixing mechanisms in a real network as well as the capabilities of common multi-parameter sensors for detection of deteriorating water quality. As a surrogate for the contaminant, the conductivity of water from different sources was studied.

Although development of a new sensor placement algorithm was not part of *STATuS*, some recommendations for the placement of sensors were derived from the results and graph theoretical properties of the network (Deuerlein *et al.* 2010):

- Sensors should be placed at so called path nodes (nodes with degree > 2) only
- Full observability of pipes and nodes of the graph theoretical forest is impossible except if sensors were to be installed at every leaf node.

- Event scenarios for sensor placement could focus on intrusion at path nodes. Other events at inner path nodes are covered as part of the latter. Two main cases can be distinguished: If the path includes a sink node the contamination is local only and transported to the sink node. In the other case the contaminant is transported to one of the end nodes (path node). From there the global contamination event is identical with the contamination of the path node directly.
- Graph theoretical bridge elements are well suited for sensor placement since a sensor on a bridge pipe separates the network into two parts without any ambiguity.

The risk-based approach could be explored in the *SMaRT-Online*^{WDN} project in order to better sample the contaminant events. The *STATuS* recommendation for placing sensors will help to reduce the number of candidate locations. This will be examined in the deliverable 3.2.

1.3.2 The *IWaNet* project

Within the collaborative research project *IWaNet* (funded by the German Ministry for Education and Research, 01ISO9014B, 2009 – 2011) a hybrid system consisting of a deterministic hydraulic simulation model, artificial neural networks (ANNs) and multi-parameter sensors with GPRS data transfer was developed. The intention was to use ANNs that were trained by simulations for network monitoring and control optimisation (ANN in conjunction with Genetic Algorithms). One of the tasks of 3S was to develop a mathematical model for finding the optimal sensor locations (Pinzinger *et al.*, 2011). A hybrid method was developed that takes into account quality measurements (conductivity, temperature, pH) as well as hydraulic parameters (pressure, flow). For detection of contaminants a mono objective integer linear programming (ILP) algorithm was implemented. As objective function the maximum coverage of Pollution Events (MCP § 2.1) was used. By defining a maximum travel time, the competing objective of minimising time to detection was also considered in a simplified way as a constraint.
For solution of the same problem also a Greedy algorithm was implemented. The results were almost as good as those of the ILP but the running time could be reduced by a magnitude. The optimal locations for hydraulic measurements were calculated by a second Greedy-Algorithm that is based on sensitivities. Using the inverse matrix of the Jacobian of the hydraulic network equations delivers, for example, the sensitivity of the pressure at a node i with respect to a change of outflow at node k. The node with biggest sensitivity to all other nodes is selected as the sensor location. Then, the system is modified (the sensor node is fixed) and the second most sensitive node is selected. This procedure is repeated until the maximum number of available sensors is reached. In conclusion, the *IWaNet* project has tested sampling methods for designing both an EWDS and an optimal sensor set for calibration.

1.3.3 The SecurEau project.

The FP7 *SecurEau* project (EC n° 217976) was aimed at the security and decontamination of drinking water distribution systems following a deliberate contamination. This 4-year collaborative project ended in February 2013 and gathered 14 partners. More specifically, this project focused on appropriate responses for rapidly restoring the use of the drinking water network after a deliberate contamination. To attain this goal, the *SecurEau* responses (*SecurEau*, 2013) were on 4 different levels:

1. ***Setup of an early warning system.*** Selection of unspecific (multi-probe) sensors able to detect unexpected brutal changes to the water quality. Determination of optimal sensor locations;
2. ***Grab/automatic sampling for identification or detection of the contaminant in a rapid and accurate manner.*** Has the early warning system raised an alarm? If there is high probability of a real contamination with potential risks it is necessary to identify it. It could be chemical, biological, radiological or nuclear;

3. ***Contamination source identification and definition of the areas contaminated***. Sorption/Reaction phenomena have been studied for a large array of contaminants to give an indication of the contaminated media (water and bulk flow, pipe wall and biofilm, sediments...) the level of importance and the transport. Additionally, contaminant source identification problem is solved to pre-localise the sources of the problem;
4. ***Decontamination procedures and control of the decontamination efficiency.*** New methods to decontaminate polluted installations including an integrated approach to neutralise water. The control of the efficacy of decontamination, by using sensors (deposits measurements) and coupons installed previously in representative areas of the network.

For ***the setup of an early-warning system***, a multi-objective problem was formulated. Several objectives were defined. Some of them are early warning specific; others were introduced to mitigate the decontamination procedure; while the last ones decrease the population vulnerability and the financial cost. Two groups of constraints were considered in order to select sensor designs ready for use by water utilities. The first group is for the operational and capital costs. The second group encompasses all the location restrictions and limitations. A novel formulation is derived that reduces the problem size in term of unknowns and constraints that leads to a Nonlinear Integer Programming problem formulation. A Monte Carlo simulation, that generates contaminant scenarios, achieves the cost evaluation. For each scenario, an extended period simulation is used to simulate the fate and transport of the contaminant, from its entry point to the taps of consumers. A conservative contaminant was selected for its stronger impact with respect to the concentration. Uniform probability distribution and equiprobability were applied to the contamination scenarios. Finally, it was proposed to simplify the network graph for overcoming large execution time but to get a better predictive hydraulic model. Engineering expertise was used and it is based on node and link aggregations.

2 Formulations and objectives for early warning detection

2.1 Sensor design objective definitions

Half of the four conflicting objectives by Ostfeld *et al.* (2008) that were part of The Battle of the Water Sensor Networks (BWSN) are retained.

The *average time to detection* criterion is the simplest. It is defined as:

$$Z_1(\delta) = \frac{1}{Nsimu} \sum_{j=1}^{Nsimu} t_j(\delta) \tag{9}$$

Where, δ is a feasible sensor design (number and location); Nsimu is the number of contamination events to consider; 1/Nsimu is the probability of a contamination event; and t_j is the minimal detection time of the j^{th} contamination for the given sensor location δ. It is worth noting that Z_1 corresponds to the expected time to detection for equiprobability of contamination events. Consequently, Z_1 is the mathematical expectation of the minimal detection time for this assumption. More general minimal detection time involving non-perfect sensors will be considered in the *SMaRT-Online*^{WDN} project (in the D3.2).

The exposed population affected (i.e., exposed population that becomes infected or symptomatic) as defined in the BWSN paper is not proposed here. This is because the information of the population supplied at network nodes is not always available from the water utility and especially because the calculation requires contaminant specific parameters such as median lethal dose, which is only possible for some targeted scenarios (impossible for general contamination purpose). The BWSN volume of contaminated water consumed prior to detection was not selected either. It depends on the strong assumption about the demand time pattern that should be the same for the same category of demand at every location in the system. A more robust objective design (as proposed in the *SecurEau* project, 2013) is to consider the *average fraction of population exposed prior to detection*.

The latter is estimated as the ratio of the connections at exposed nodes over the total number of connections in the WDN:

$$Z_2(\delta) = \frac{1}{Nsimu} \sum_{j=1}^{Nsimu} f_j(\delta) \tag{10}$$

Where f_j is the fraction population exposed to the j^{th} contamination for the given sensor design δ.

The *likelihood of detection* is the average number of detections for a given sensor design. Its complement of one is the average number of failed detection that is:

$$L_D(\delta) = \frac{1}{Nsimu} \sum_{j=1}^{Nsimu} d_j(\delta) = 1 - Z_3(\delta) = 1 - \frac{1}{Nsimu} \sum_{j=1}^{Nsimu} (1 - d_j)(\delta) \tag{11}$$

Where d_j is one if the j^{th} contaminant event is detected by δ else zero. Contrary to the two first designs (Z_1 and Z_2) that should be minimised, the likelihood of detection L_D should be maximised. Alternatively, the *average number of failed detections* Z_3 is to be minimised.

Two additional specific design objectives were defined in the *SecurEau* project (2013) to facilitate the decontamination procedure. The *average contaminated network water volume* is the average volume of contaminated water that is inside the system at the time of warning. It is estimated by:

$$Z_4(\delta) = \frac{1}{Nsimu} \sum_{j=1}^{Nsimu} v_j(\delta) \tag{12}$$

Where v_j is the volume of contaminated water in the system for the j^{th} event before detection by δ. The second additional objective to minimise is the *average contaminated network pipe surface*:

$$Z_5(\delta) = \frac{1}{Nsimu} \sum_{j=1}^{Nsimu} s_j(\delta) \tag{13}$$

Where s_j is the contaminated surface for the j^{th} event before detection by δ.

In order to protect a population that is at risk (*e.g.*, a hospital; a school; a vulnerable customer - registered in the Safeguard scheme; and to a certain extent a normal consumer) the *average fraction of population exposed at risk* criterion is used. It is defined as:

$$Z_6(\delta) = \frac{1}{Nsimu} \sum_{j=1}^{Nsimu} r_j(\delta) \tag{14}$$

Where r_j is the fraction of population at risk that is exposed to the j^{th} contamination before detection by the given sensor design δ. This criterion differs from the Z_2 *fraction of population exposed* (Eq. (10)) as the definition risk may differ from the connection number. An example of risk definition is 5, for presence of a hospital or a school; 3 for a safeguard consumer; and 1 for a normal consumer.

For operational reasons, the traffic information is proposed in order to account for the installation cost that is less (about 10 times) for a location with no traffic than for a location with traffic. The installation cost is defined as:

$$Z_7(\delta) = IC(\delta) \tag{15}$$

Where *IC* is the installation cost with respect to the design δ.

Finally, the *weighted linear sum criterion* is proposed in an attempt to make a trade-off between the different objectives:

$$Z_8(\delta) = \sum_{\alpha=1}^{6} w_\alpha \frac{Z_\alpha(\delta)}{Z_\alpha(\varnothing)} + w_7 \frac{IC(\delta)}{|\delta| UIC_{\text{traffic}}} \tag{16}$$

Where the w_α, $\alpha = 1,\ldots,7$ are positive weights that relates the importance of an objective in the design; their sum is one; $|\delta|$ is the sensor cardinality (*i.e.*, the number of sensors for the design set δ); $\varnothing$ is the empty sensor design set; and UIC is the traffic unitary installation cost.
In summary, the list of the objectives retained for early-warning detection is:

- the average time to detection;
- the fraction population exposed;
- the likelihood of detection;
- the average contaminated network water volume;
- the average contaminated network pipe surface;
- the average fraction of population exposed at risk;
- the installation cost;
- the weighted linear sum of the previous objectives.

2.2 Constraints

Two groups of constraints have to be considered in order to select sensor designs ready for use by water utilities.

The first group is for the operational and capital costs. Two French *SMaRT-Online*^{WDN} end-users planned and started to install and operate water quantity and quality sensors. With full sectorisation of 14 sectors, the CUS water utility will install 48 new remote monitoring points for a cost of 40 k€/station in 2014/2015. The eight stations funded by the ANR in the *SMaRT-Online*^{WDN} project were already set up in 2012/2013. The VEDIF water utility has instrumented a pilot site on the Villejuif subnet with 20 Kapta sensors and 9 flow meters at the beginning of the project and globally, the installation of 70 sensors started on the SEDIF network in 2013 and will stop in March 2014. 110 new quality sensors will be installed on the VEDIF network in 2014. The global capital cost for VEDIF is about 420 k€. To simplify, the operational cost is considered to be a linear function of the number of sensors. The installation cost for a monitoring station will greatly vary from one location to another. Nevertheless, an average installation cost may be given and used because of the amount of sensors to install. The capital cost may also be considered as a linear function of the sensor number. In this study, operational and capital costs are represented and valuated by the number of sensors.

The second group encompasses all the location restrictions and limitations. Some locations are selected by the water utility. This leads to defining a preselected sensor set P. Other locations should be avoided because they lead to technical and financial limitations: installation costs are too expensive and/or technical requirements such as minimum velocity (for optimisation of chlorine sensor working) are not met. This defines the feasible sensor set F that is a superset of the preselected sensor set P.

2.3 Formulation

Here, starting from the (MILP) formulation by Propato *et al.* (2005) a novel formulation is derived that reduces the problem size.

Propato *et al.* (2005) have formulated a Mixed Integer Linear Programming (MILP) problem to solve the sensor design problem. This is equivalent to a maximum coverage problem (MCP) with the contribution of the jth contaminant event to objectives Eqs (9-15) given by the following expression:

$$q_j^{\alpha}(x,y) = \sum_{i=1}^{N_j-1} c^{\alpha}{}_{ij} y_{ij} \tag{17}$$

Where q_j is the contribution of interest for the jth event; N_j is the number of nodes belonging to the jth contaminant plume (if N_j is 1, q_j is zero); c_{ij} is the partial impact cost for not monitoring the ith node (in the temporal sequence) for the jth contaminant pollution; x is the decision variable that depends on the sensor design δ, is binary and is involved only in the constraints Eq. (3) in combination with y; and y_{ij} is a positive continuous auxiliary variable that is introduced for making the objective function linear. The binary y_{ij} variable was relaxed to lie between 0 and 1. The advantage of such a formulation is that standard solving methods may be built. A disadvantage is the number of variables and constraints which makes this problem difficult to solve when numerous contaminant simulations and all the potential node locations are to be explored. Actually, the number of auxiliary variables and constraints is of the order of N^2, where N is the number of nodes. That is for *SMaRT-Online*^{WDN} water utility end-users several millions of variables and constraints.

A novel and derived formulation is proposed here to overcome the size problem. The auxiliary variable y_{ij} is removed from the formulation using Boolean functions instead:

$$q_j^{\alpha}(\delta) = \sum_{i=1}^{N_j-1} c^{\alpha}{}_{ij} y_{ij}(\delta), \text{ with } \mathrm{y}_{ij}(\delta) = \prod_{k=1}^{i} \left[1 - x_{kj}(\delta)\right] \in \{0,1\} \tag{18}$$

Where δ is the sensor design variable with sensor node identifiers as components; x_{kj} is a Boolean function that is 1 if there is a sensor at the kth node in the temporal sequence for the jth simulation, else zero; and y_{ij} is also a Boolean function of δ. Assuming that junction nodes are numbered from 1 to N, the mth-member or element of δ verifies:

$$\delta_m \in \{1,2,\cdots,N\},\ \mathrm{m}=1,\cdots,N_s \tag{19}$$

Where $N_s = |\delta|$ is the number of sensors to place.

The sensor design multi-objective problem may be formulated as:

$$\min_{\delta}\left[Z_\alpha(\delta) = \frac{1}{\mathrm{Nsimu}} \sum_{j=1}^{Nsimu} q_j^{\ \alpha}(\delta) \right]^T_{\alpha=1,\cdots,7} \tag{20}$$
$$\text{subject to: } \mathrm{P} \subset \delta \subset \mathrm{F},\ |\delta| = N_s$$

Where P is the pre-selection set; F is the feasible set; Z_α is one of the seven objective functions defined in Eqs. (9-15); q_{ij} is defined by Eq. (18); and δ is the sensor design (decision) variable that is a subset of F and a superset of P. This problem consists of a Nonlinear Integer Programming problem that is multi-objective. The solution to this problem is a set of Pareto points.

By combining all the objectives in the weighted linear sum of the objectives Z_8, a mono-objective version problem can be written as:

$$\min_{\delta} Z_8(\delta),\ \text{subject to: } \mathrm{P} \subset \delta \subset \mathrm{F},\ |\delta| = N_s \tag{21}$$

Where arg min stands for argument of the minimum. The solution to this problem is proposed as an alternative way to find a solution to the multi-objective problem. The weights have to be calibrated by the decision maker (main rules look difficult to establish). Clearly, when solving a mono-objective programming problem, global information is lost but the decision is made easier.

3 Optimal designs for minimising the variance of estimators

3.1 Parameter classes and measurements

The nature of measurements and their locations determine if a network is observable and influence strongly on the numerical behaviour for the calibration problem. The dispersion of the parameter estimation error is also dependant on the sensor design to a greater or lesser degree.

Any water quality or water quantity output may be a candidate to be observed. But the measurements may be preselected with cheaper location-specific costs and with the easiness of installation.

As discussed in the background chapter, the nodal demands for short periods of time are rough estimates. Accordingly, they will constitute the unknowns that we seek to identify. Based on the nature of premise occupation and water use metering analysis, consumers may be grouped in few classes with the same demand multiplier time pattern. For example, one will distinguish domestic, residential and industrial consumer classes. Then, the consumers are aggregated at nodes. Few consumers of different classes can be aggregated at the same node. This reads:

$$\mathbf{d}(t) = \mathbf{G}_d \mathbf{x}(t) \tag{22}$$

Where **d** is the nodal demand; $\mathbf{G}_d$ is the nj by nd class matrix of nodal demand allocation; and **x**(t) is the demand class of size nd. This way parameter observability may be obtained as in Cohen and Carpentier (1991). The definition of such classes is a difficult task. A trade-off should be made between the model error resulting from simplification and the parameter uncertainty. Indeed, few parameters are easy to calibrate but errors in the model can be significant. Several authors have examined this question as well as the rational use of probability theory and automatic clustering (Mallic *et al.*, 2002; Moughton *et al.*, 2006).

The parameter calibration problem may be extended to the offline identification of relative pipe roughnesses, diameters, valve states, pipe kinetics and dispersion coefficients. In the *SMaRT-Online*^{WDN} project we will focus mainly on online demand calibration.

The vector $\mathbf{y}^{all}$ of all potential measurements may be retrieved using the following nonlinear regression model assuming **x** is known and there is no error in the model:

$$\mathbf{y}^{all}(t) = \mathbf{y}(\mathbf{x},t) + \varepsilon^{all}(t) \tag{23}$$

Where **y**(**x**,t) is the corresponding vector-values calculated from the water quantity and the water quality models; and ε^{all} is the error or noise vector that we will assume distributed with mean zero and diagonal covariance matrix **C**. In Eq. (23) $\mathbf{y}^{all}$ and ε^{all} are random variables and **y**(**x**,t) is the expectation of $\mathbf{y}^{all}(t)$.

Some measurements are of a different nature; the accuracy may also depend of the level of the measurement values. Also, it is worth introducing the confidence we have in the measurements:

$$\left| y_i^{all} - y_i \right| \le \Delta y_i \tag{24}$$

With Δy_i the confidence we have for the i^{th} measurement at a given level. We assume these coefficients are proportional to the model variance of the model error and let **W** be the weight diagonal matrix with i^{th} term the inverse of the i^{th} measurement confidence squared. More specifically, we assume **C** of this form:

$$\mathbf{C} = \sigma^2 \operatorname{diag}\left(\Delta y_i^{\,2}\right) = \sigma^2 \mathbf{W}^{-1}$$

The latter can be used to homogenise Eq. (23) for the error model to be iid. Thus by reducing by the confidence we get:

$$\mathbf{Y}^{all}(t) = \mathbf{Y}(\mathbf{x},t) + \mathbf{E}(t) \quad \text{with } \mathbf{E}(t) = \mathbf{W}^{\frac{1}{2}} \varepsilon(t) \tag{25}$$

The model error or Eq. (25) has a mean-zero and there is homoscedasticity, *i.e.*: the variance error matrix is:

$$\operatorname{var}\left(\mathbf{E}(t)\right) = \sigma^2 \mathbf{I}_m$$

For a sensor design δ, we observe with one realisation at time t some components of $\mathbf{y}^{all}$. Using Eq. (23) and the definition of measurements in δ, this could be written as the reduced nonlinear regression equation:

$$\mathbf{y}_\delta^{mes} = \mathbf{S}_\delta \mathbf{y}(\mathbf{x},t) + \varepsilon_\delta(t) \tag{26}$$

With $\mathbf{S}_\delta$ is the selection matrix we introduced Eqs. (6) and (7). In the same manner, the following reduced form with iid errors may be obtained:

$$\mathbf{Y}_\delta^{mes} = \mathbf{S}_\delta \mathbf{Y}(\mathbf{x},t) + \mathbf{E}_\delta(t) \tag{27}$$

3.2 Influence of the measurement error on the Least-Squares estimation

Let $\hat{\mathbf{x}}$ be the solution of Eq. (26) at least-squares sense. $\hat{\mathbf{x}}$ should be a solution of the normal equation:

$$\mathbf{J}_\delta(\hat{\mathbf{x}})^T \mathbf{W}_\delta \left(\mathbf{S}_\delta \mathbf{y}(\hat{\mathbf{x}},t) - \mathbf{y}_\delta^{mes}(t)\right) = \mathbf{0}_p \tag{28}$$

Where $\mathbf{J}(\hat{\mathbf{x}})$ it the Jacobian of $\mathbf{y}$ evaluated at $\hat{\mathbf{x}}$. $\hat{\mathbf{x}}$ is also a random variable that depends on the measurement error ε_δ. With no measurement error, the least-square estimation is:

$$\hat{\mathbf{x}}_0 / \ \mathbf{J}_\delta(\hat{\mathbf{x}}_0)^T \mathbf{W}_\delta \left(\mathbf{S}_\delta \mathbf{y}(\hat{\mathbf{x}}_0,t) - \mathbf{S}_\delta \mathbf{y}(\mathbf{x},t)\right) = \mathbf{0}_p \tag{29}$$

In contrast, with a non-null measurement error the least-square estimation may be significantly different and it satisfies:

$$\hat{\mathbf{x}}_\varepsilon / \ \mathbf{J}_\delta(\hat{\mathbf{x}}_\varepsilon)^T \mathbf{W}_\delta \left(\mathbf{S}_\delta \mathbf{y}(\hat{\mathbf{x}}_\varepsilon,t) - \mathbf{S}_\delta \mathbf{y}(\mathbf{x},t) - \varepsilon_\delta\right) = \mathbf{0}_p \tag{30}$$

We will call the influence of the measurement error on the least-squares estimates, the deviation from the solution with no measurement error. At first-order estimates, the influence fulfils the linear equation:

$$\hat{\mathbf{x}}_\varepsilon - \hat{\mathbf{x}}_0 = \left(\mathbf{J}_\delta{}^T \mathbf{W}_\delta \mathbf{J}_\delta\right)^{-1} \mathbf{J}_\delta{}^T \mathbf{W}_\delta \varepsilon_\delta \tag{31}$$

With $\mathbf{J}_\delta$ is a Jacobian estimate that is assumed constant at the vicinity of the solution.

In equivalent manner, this could be rewritten as:

$$\hat{\mathbf{x}}_\varepsilon - \hat{\mathbf{x}}_0 = \left(\mathbf{T}_\delta{}^T \mathbf{T}_\delta\right)^{-1} \mathbf{T}_\delta{}^T \mathbf{E}_\delta = \mathbf{T}_\delta{}^+ \mathbf{E}_\delta \tag{32}$$

Where

$$\mathbf{T}_\delta = \mathbf{S}_\delta \mathbf{W}^{+\frac{1}{2}} \mathbf{J} \tag{33}$$

and $(\mathbf{W})^+$ is the pseudo-inverse of $\mathbf{W}$. $\mathbf{T}_\delta$ is the Jacobian of $\mathbf{S}_\delta\mathbf{Y}$ defined in Eq. (27). To calculate this matrix, we need a sensor design δ for the matrix $\mathbf{S}_\delta$, the measurement confidences for $\mathbf{W}$ and an estimate of **J** the Jacobian of the system at each potential measurement. The pseudo-inverse of $\mathbf{T}_d$ is the sensitivity matrix of the parameter estimate at least-squares sense. In Eq. (32), the influence of measurement errors, depends linearly on measurement errors.

3.3 Selected Criterion that minimises the variance of the least-squares estimator

The main idea (Piller, 1995) is to choose a sensor design δ that minimises the absolute value of the influence of measurement error Eq. (32) for measurement errors within the confidence limits defined in Eq. (24). Because of W and the appropriate change of variable the upper limit for E is:

$$|E_i(t)| \leq 1 \tag{34}$$

Which means that **E** belongs to the unit ball for the infinity (or maximum) norm.

For each design δ, we calculate:

$$\sup_{\mathbf{E}_\delta \in \overline{B}_\infty} (\mathbf{T}_\delta)^+ \mathbf{E}_\delta = \left\| (\mathbf{T}_\delta)^+ \right\|_\infty \tag{35}$$

Where $\overline{B}_\infty$ is the unit ball; and $\|\ \|_\infty$ is the infinity matrix norm which is simply the maximum absolute value row sum of the matrix.

The problem of optimal design for parameter calibration is formulated as:

$$\begin{aligned} &\min_\delta Z_\infty(\delta) = \left\| \mathbf{T}_\delta{}^+ \right\|_\infty \\ &\text{subject to: } \operatorname{rank}(\mathbf{T}_\delta) = p \end{aligned} \tag{36}$$

With $\mathbf{T}_\delta$ is defined in Eq. (33); and rank is the matrix rank operator. The full rank constraint the number of $\mathbf{T}_\delta$ columns is to ensure the algebraic observability.

In the *SMaRT-Online^{WDN}* project optimal solution for Eq. (36) will be compared with regards to their performance with the early-warning criteria specified in chapter 2.

4 Conclusions

This deliverable report specifies which performance criteria should be considered to place water quality and water quantity sensors for both early-warning detection systems and model calibration. The optimal designs that are proposed come from a thorough analysis of the literature and from the *SMaRT-Online^{WDN}* consortium experience.

For early-warning detection system, the eight following different objectives are defined to optimise: the average time to detection; the fraction population exposed; the likelihood of detection; the average contaminated network water volume; the average contaminated network pipe surface; the average fraction of population exposed at risk; the installation cost; and the weighted linear sum of the previous objectives. The solution is a two-step method. First, several pollution events are simulated and several impact costs if not measuring are worked out. Then, a multi-objective nonlinear integer-programming problem is to be solved to cover the pollution events under budget and limitation constraints. A trade-off is to be found, using Pareto-efficient fronts on conflicting objectives.

Placing sensors for model calibration relies on selecting designs that reduce the influence of measurement errors. The solution proposes to minimise the variance of the least-squares estimator. The objective function represents the sensitivity of the demand class parameter estimation to the measurement error. The full rank constraint restricts the design solution that leads to observability of parameters. The optimal sensor designs lead to calibration problem with the best-condition numbers. The confidence interval for parameters will be reduced compared to another sensor design with higher score.

5 Referred bibliography

Al-Zahrani, M. A., and Moied, K. (2001). "Locating Optimum Water Quality Monitoring Stations in Water Distribution Systems." World Water & Environmental Resources congress, EWRI01, ASCE, Orlando (Florida), USA.

Bargiela A., Hainsworth G. (1989). "Pressure and flow uncertainty in water systems.", Journal of Water Resources Planning and Management, Vol. 115, 2, March 1989, pp. 212-229.

Berry, J., Hart, W. E., Phillips, C. A., Uber, J. G., and Watson, J.-P. (2006). "Sensor Placement in Municipal Water Networks with Temporal Integer Programming Models." Journal of Water Resources Planning and Management, 132(4), 218-224.

Bush, C. A., and Uber, J. G. (1998). "Sampling Design Methods For Water Distribution Model Calibration." Journal of Water Resources Planning and Management, 124(6), 334-344.

Carpentier, P. and Cohen G. (1991). « State Estimation and Leak Detection in Water Distribution Networks ». Civil Engineering Systems, Vol. 8, pp. 247-257, 1991.

Cheung, P., Piller, O., and Propato, M. (2005). "Optimal Location of Water Quality Sensors in Supply Systems by Multiobjective Genetic Algorithms." Eight International Conference on Computing and Control in the Water Industry CCWI05 'Water Management for the 21st Century', University of Exeter, UK, printed by Centre for Water Systems, Volume 1/2, 203-208.

Christofides, N. (1975). Graph theory: An algorithmic approach. Academic Press, Inc., San Diego, Calif.

de Schaetzen, W. B. F. (2000). "Optimal Calibration and Sampling Design for Hydraulic Network Models," Engineering thesis from the University of Exeter, 277 pages, Exeter, UK.

Deuerlein, J. (2006). "Efficient supply network management based on linear graph theory", 8th Annual Water Distribution System Analysis Symposium; Cincinnati, Ohio, USA; August 27-30, 2006.

Deuerlein, J. W. (2008). "Decomposition model of a general water supply network graph", ASCE Journal of Hydraulic Engineering, 134, 6, 2008, 822 - 832.

Deuerlein, J., Wolters, A., Meyer-Harries, L., and Simpson, A. R. (2010). "Graph decomposition in risk analysis and sensor placement for water distribution network security", 12th Annual Symposium on Water Distribution Systems Analysis, WDSA 2010, American Society of Civil Engineers, Tucson, Arizona, 12-15 September, 2010.

Fabrie, P., Gancel, G., Mortazavi, I., and Piller, O. (2010). "Quality Modeling of Water Distribution Systems using Sensitivity Equations." Journal of Hydraulic Engineering, 136(1), 34-44.

Gancel, G. (2006). "Modélisation d'un problème inverse pour la qualité de l'eau dans les réseaux d'eau potable," Doctorat Mathématiques Appliquées, Université de Bordeaux 1 thesis from, defended on, 136 pages.

Gancel, G., Mortazavi, I., and Piller, O. (2006). "Coupled numerical simulation and sensitivity assessment for quality modelling in water distribution system." Applied Mathematics Letters, 19(12), 1313-1319.

Kang, D., and Lansey, K. (2010). "Optimal Meter Placement for Water Distribution System State Estimation." Journal of Water Resources Planning and Management, 136(3), 337-347.

Kapelan, Z. (2002). "Calibration of water distribution system hydraulic models," Engineering thesis from the Univ. of Exeter, 335 pages, Exeter, UK.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2003). "Multiobjective Sampling Design for Water Distribution Model Calibration." Journal of Water Resources Planning and Management, 129(6), 466-479.

Kessler, A., Ostfeld, A., and Sinai, G. (1998). "Detecting Accidental Contaminations In Municipal Water Networks." Journal of Water Resources Planning and Management, 124(4), 192-198.

Khuller S., Moss A., Naor J.S. (1999). “The budgeted maximum coverage problem.”, Information Processing Letters 70 (1999) 39–45.

Kumar, A., Kansal, M. L., and Arora, G. (1997). "Identification of Monitoring Stations in Water Distribution System." J. Envir. Engrg., 123(8), 746-752.

Lee, B. H., and Deininger, R. A. (1992). "Optimal Locations of Monitoring Stations in Water Distribution System." Journal of Environmental Engineering, 118(1), 4-16.

Mallick, K. N., Ahmed, I., Tickle, K. S., and Lansey, K. E. (2002). "Determining Pipe Groupings for Water Distribution Networks." Journal of Water Resources Planning and Management, 128(2), 130-139.

Moughton, L. J., Buchberger, S. G., Boccelli, D. L., Filion, Y. R., and Karney, B. W. (2006). "Effect of Time Step and Data Aggregation on Cross Correlation of Residential Demands." 8th Annual Symposium on Water Distribution Systems Analysis, Cincinnati, Ohio, USA, printed by ASCE, Volume 247/, 42-42.

Optimal_Design (2013), http://en.wikipedia.org/wiki/Optimal_design accessed on the 15th of November 2013.

Ostfeld, A., and Salomons, E. (2004). "Optimal Layout of Early Warning Detection Stations for Water Distribution Systems Security." Journal of Water Resources Planning and Management, 130(5), 377-385.

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, J.-P., Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S.-T., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S. R., Barkdoll, B. D., Gueli, R., Huang, J. J., McBean, E. A., James, W., Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G. B., Wu, Z. Y., and Walski, T. (2008). "The Battle of the Water Sensor Networks (BWSN): A Design Challenge for Engineers and Algorithms." Journal of Water Resources Planning and Management, 134(6), 556-568.

Piller, O. (1995). "Modeling the behavior of a network - Hydraulic analysis and sampling procedures for parameter estimation," Applied Mathematics thesis from the University of Bordeaux (PRES), 288 pages, Talence, France.

Piller, O., Brémond, B., and Morel, P. (1999). "A Spatial Sampling Procedure for Physical Diagnosis in a Drinking Water Supply Network." International Conference on Computing and Control for the Water Industry CCWI '99, Exeter, UK, printed by Research Studies Press ltd., Volume 1/2, 309-316.

Piller, O., Gilbert, D., and Van Zyl, J. "Dual Calibration for Coupled Flow and Transport Models of Water Distribution Systems." Water Distribution Systems Analysis 2010, 722-73.

Pinzinger, R., Deuerlein, J., Wolters, A., Simpson, A. R. (2011). "Alternative Approaches for Solving the Sensor Placement Problem in Large Networks" 13th Annual Symposium on Water Distribution Systems Analysis, WDSA 2011, ASCE, Palm Springs, USA, 22-26 May, 2011.

Preis, A., Whittle, A. J., Ostfeld, A., and Perelman, L. (2011). "Efficient Hydraulic State Estimation Technique Using Reduced Models of Urban Water Networks." Journal of Water Resources Planning and Management, 137(4), 343-351.

Propato, M., Piller, O., and Uber, J. G. (2005). "A sensor location model to detect contaminations in water distribution networks." World Water and Environmental Resources Congres (EWRI05), Anchorage, USA, May 15-19, 2005, 12.

Propato, M., and Piller, O. (2006). "Battle of the Water Sensor Networks." 8th annual Water Distribution System Analysis Symposium, University of Cincinnati, Cincinnati, Ohio USA, August 27-30 2006, printed by ASCE, 8.

Savic, D. A., Kapelan, Z. S., and Jonkergouw, P. M. R. (2009). "Quo vadis water distribution model calibration?" Urban Water Journal, 6(1), 3 - 22.

SecurEau project (2013). http://www.secureau.eu/, access on 03 November 2013.

Uber, J. G., Janke, R., Murray, R., and Meyer, p. (2004). "Greedy Heuristic Methods for Locating Water Quality Sensors in Distribution Systems." 2004 World Water and Environmental Resources Congress, Critical Transitions In Water And Environmental Resources Management, Salt Lake City, UT, printed by 2004, American Society of Civil Engineers, (CD-ROM).

Woo, H.-M., Yoon, J.-H., and Choi, D.-Y. (2001). "Optimal Monitoring Sites Based on Water Quality and Quantity in Water Distributions Systems." World Water & Environmental Resources congress, EWRI01, ASCE, Orlando (Florida), USA, 9.