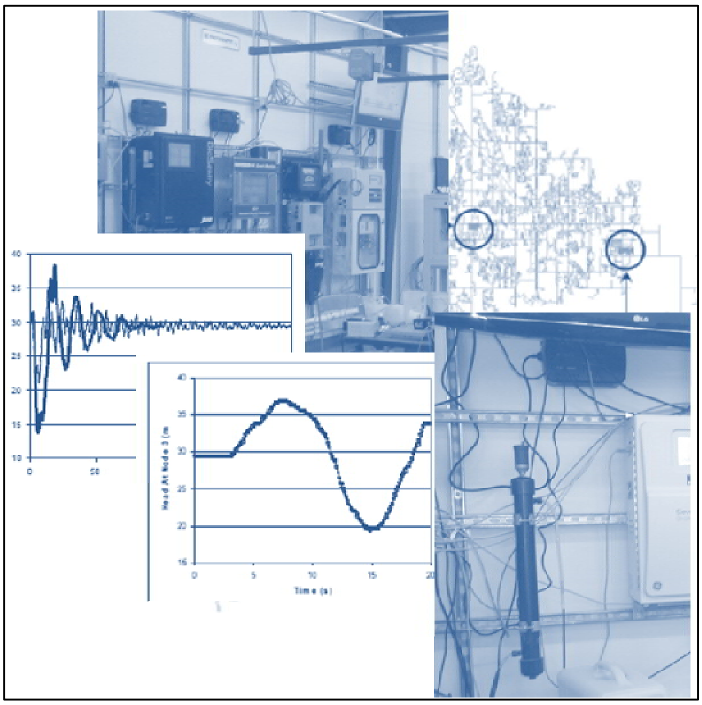
Prepared
enabling change

Optimal Macro-Location methods for sensor placement in Urban Water Systems

Literature Review



Centre for Water Systems, University of Exeter
August 2011

SEVENTH FRAMEWORK

Prepared
enabling change

SEVENTH FRAMEWORK PROGRAMME

COLOPHON

Title
Optimal Macro-Location methods for sensor placement in Urban Water Systems

Report number
PREPARED 2011.004

Deliverable number
3.5.1

Author(s)
Kyle E. Thompson, Lydia Vamvakeridou-Lyroudia, Zoran Kapelan and Dragan Savic (UNEXE)

Quality Assurance
Jean-Luc Bertrand-Krajewski (INSA)

Document History

Version	Team member	Status	Date update	Comments
1.0	Thompson, K.E.	Draft	14/03/2011	
1.1	Vamvakaridou-Lyroudia, L.S. Kapelan, Z.	Draft		Review/Additions
1.2	Kapelan, Z.	Draft	25/03/2011	Review/Additions/changes in the structure
1.3	Thompson, K.E.	Draft		Additions/changes in the structure
1.3	Vamvakeridou-Lyroudia L.S. Kapelan, Z.	Draft	12/04/2011	Review/Additions
1.4	Thompson, K.E.	Draft	May2011	Additions/Corrections
1.5	Vamvakaridou-Lyroudia, L.S. Kapelan, Z.	Draft	June 2011	Review/Corrections
1.6	Vamvakaridou-Lyroudia, L.S.	Final	August 2011	Review

This report is: PU = Public

Summary

This report provides a critical state-of-the-art literature review on the subject of optimal sensor placement in Urban Water Systems (UWS) satisfying the requirements of Deliverable D3.5.1 within Work Package (WP) 3.5 of the FP7-EC Large Scale Integrating Project PREPARED. It presents a summary of existing sensor macro-location design methodologies intended to facilitate the collection of relevant and efficient measurements in UWS.

The optimal placement of a limited number of sensors within an urban water network ('ideal' locations at which measurements of selected quantities should be taken) is a necesssary step in the application of intelligent and cost-effective monitoring for current and prospective water supply and sanitation systems. The definition of an 'optimised' sensor network is dependent on the intended purpose of the sampling scheme and the resulting sensor data. Design methodologies in literature are typically catered toward one of a number of distinct agendas and the field is consequently segmented into a range of subsidiary groups, i.e., methods to determine optimal placement schemes for effective contaminant detection, methods to determine optimal schemes for model calibration, and methods for leak and burst detection or source tracing; each formulation of which may be largely irrelevant outside of its own context. It should be noted that *model calibration* is the focus of this review and directly addresses the mandate of the Deliverable D3.5.1; however, in accordance with the description of work (DoW) and the targets of PREPARED, and in aid of an inclusive and cohesive review of sensor macro-location literature, optimal macro-location is broached here from a more generalised perspective.

Numerical models are instrumental to the future of cost-effective monitoring and management for urban water systems. Unfortunately, numerous sources of uncertainty in these models remain a key concern for their practical viability, and indeed the need for calibration measures in both hydraulic and water quality simulations is commonly highlighted by researchers and practitioners in the field. Model calibration requires data collected on a physical reference system and hence the efficacy of a calibration procedure is directly dependent on the sampling scheme. The prevalent methodologies underpinning modern sensor placement

algorithms for optimal calibrations are based largely on Uncertainty Analysis via the First-Order Second-Moment method with ultimate design suitability being assessed through subjective measures of parametric and prediction variance tensors. The performance of a sampling design depends on both the spatial distribution and the *quantity* of measurements/devices to be optimally located. Importantly, sampling design literature reveals a tendency for diminishing returns accompanying iterative increases in the number of sensing devices/locations on a network and it is therefore important that practitioners are able to discern an appropriate price point for the number of optimised measurement locations.

This report begins with a brief introduction to the history of hydraulic and water quality modelling for both water distribution systems (WDS) and urban waste water systems (UWWS) and further details the key challenges relating to their practical application. Section 3 discusses model calibration and highlights its connection to sampling design (i.e. macro-location of sensors). Sampling design methods and algorithms are introduced in sections 4 and 5 for both model calibration and 'other' sampling objectives respectively. Within section 5, detailed accounts of existing macro-location approaches and algorithms that are largely independent of their objective functions are provided on the grounds of their applicability to general macro-location problems. Finally, section 6 concludes the report and identifies future actions with regards to PREPARED objectives in linked work areas.

The report is a useful resource for researchers involved in sensor network design, including those involved in the development of relevant tools in subsequent PREPARED work packages

Contents

1 Introduction

This report consists Deliverable 3.5.1 of the FP7 EC integrated project PREPARED and addresses optimal macro-location methods for sensor placement in Urban Water Systems (UWS). In accordance with the D3.5.1 description of work, it will review existing methods in sensor placement ('sampling design') literature with a focus on model calibration for both water distribution systems and urban waste water (sanitation) systems. Suitable methodologies are to be critically evaluated with respect to the requirements of Work Area 3 (WA3) and linked work packages (WP4.2 and WP4.4) (PREPARED, 2010).

1.1 PREPARED: Enabling Change

There can be little doubt that climate change is one of the most crucial and challenging issues confronting the global community in the 21st Century. Amongst the potential impacts of progressive climate change are impending disturbances to the availability of water resources on a regional basis, including unprecedented flooding patterns, droughts, water scarcity and sea level variation (Bates et al., 2008). Significant changes to the temperature and capacity of natural water resources present important challenges to water utilities and regulatory bodies as they endeavour to maintain adequate clean water quality and quantity to meet consumer demands. Similar challenges apply to appropriate operating conditions and sanitation measures within current and future urban waste water systems; hydrological deviation may be a causal factor in the spread of new diseases and losses in local biodiversity (Parmesan et al., 2003). The threat of climate change is further compounded by continuously growing demand on correspondingly scarce potable water sources and there is a mounting ambition for governments and water utilities to adapt, as intelligently as possible, to the charges of future water management (Vorosmarty et al., 2000).

The PREPARED project aims to confirm and demonstrate that the water supply and sanitation systems of cities and their catchments can adapt and be resilient to the challenges of climate change, while simultaneously being cost effective and carbon efficient (PREPARED, 2010). PREPARED is principally driven by the requirements of industry and seeks first to augment the management of current technological infrastructures to cope with prospective supply and demand patterns, and second to assist in the planning and assessment of further developments. This will be achieved through the application of pragmatic, robust methods and technologies intended to better understand and administer clean water distribution and urban wastewater systems; enhanced information and control will guide decreased energy consumption, reduced burst leakage, improved clean water supply quality, improved waste water treatment works (WWTW) efficiency, reduced combined sewer overflow (CSO) event frequency/severity, decreased maintenance and expansion costs, and all manner of accompanying augmentations for both WDS and UWWS management procedures (Schutze et al., 2003). A continuous and well-informed appraisal of network capacities and limitations also reduces the potential for over- or under-investment in response to evolving estimates of climate change impact and consumer-driven demand growth.

PREPARED aims to provide meaningful tools and knowledge to assist policy-makers and utilities in their responsibility to recognise and respond to the biophysical impacts of their management decisions. The PREPARED project will deal with early warning systems, as well as short- and long-term response strategies for urban areas. More specifically, it will (PREPARED 2010):

- Address issues related to the management of water, waste water and storm water potentially impacted by climate change.
- Optimise, examine and implement adaptive solutions that will contribute towards international integration and coordination.
- Reduce the greenhouse gas emissions of the water and waste water sector (minimise the carbon and water footprint).
- Improve the resilience of water utilities in the face of climate and consumer demand prospects.
- Contribute to the development of the European Union (EU) knowledge base concerning the water supply and sanitation sector.
- Contribute to the integration of adaptation strategies into EU policies. Priority will be given to adaptation measures that will generate net social and/or economic benefits irrespective of uncertainty in future IPCC forecasts.

As part of the effort to meet these ends, an essential component of the PREPARED project and the overall ambition of this work area (Work Area 3) is the development of a functional 'toolbox' for the real-time monitoring and modelling of water distribution and urban wastewater systems. The real-time control strategy ultimately implemented to provide decision support for water utilities will necessitate the regular gathering of data regarding relevant quantities on the physical system (i.e. pressure heads at junction nodes), ideally from well-chosen spatial and temporal locations; the assessment of which forms the subject of this literature review (D.3.5.1). The optimal macro-location of sensors (or 'sampling design`) is a developing field with several tested and trusted methodologies that could be effectively implemented on a large and practical scale for a pemonstration city network in accordance with the objectives of PREPARED (Savic et al., 2009).

1.2 Overview of Optimal Macro-Location of sensors

Optimal 'macro-location' in the context of water sensor networks refers to the determination of an optimal spatial distribution at which measurements of some quantity/quantities would best be taken (i.e. an ideal sampling design incorporating a limited number of sensor devices within a UWS). Optimality is defined according to particular operational requirements reliant on the proposed objective(s) of a sampling project; naturally, the 'best' data points on a network are axiomatically influenced by the intended usage of the data posterior to any given sampling scheme i.e. an optimal distribution for effective contaminant detection could be dissimilar to an optimal distribution with regards to efficient model calibration. The optimisation of data sampling locations in UWS has been previously addressed by researchers with respect to several distinct agendas broadly classified as follows (Bush and Uber, 1998):

- ***Ambient monitoring*** for baseline system characteristics
- ***Detection*** of problem sources on the network such as contamination events, leaks, etc.
- ***Compliance*** for the maintenance of system performance
- ***Research (calibration)*** of Network Models

Deliverable D3.5.1 is concerned primarily with sensor placement schemes that facilitate the optimal calibration and monitoring of clean water distribution and urban waste water system models. Sensor placement naturally impacts the efficiency and effectiveness of calibration procedures for both hydraulic (quantitative) systems models and dependent models for water quality indicators (Lee and Deininger, 1992; Freni et al.,, 2010). An optimal sampling design provides the 'best', most informative data for the purposes of model calibration in each case (Savic et al.,, 2009). The report also includes optimal sensor placement methodologies for early warning (contamination problem), as this aspect is also part of the aims of PREPARED.

The optimisation of sensing locations on an urban water network should naturally be considered in conjunction with the *quantity* of sensing devices/measurement locations, as both have been shown to affect the resulting model calibration accuracy (Walski, 1983). The quantity (number) of measurements is closely related to the cost of sampling and hence is a significant factor in the uptake and viability of model calibration procedures in industry. Generally, sampling design literature reveals a tendency for diminishing returns accompanying iterative increases in the number of sensing devices/locations on a network (Kapelan et al.,, 2003, Krause et al.,, 2006; Krause et al.,, 2009); an excessive amount of data may provide little improvement over a reduced (and appreciably cheaper) sampling procedure. It is therefore paramount that practitioners are able to discern an appropriate 'price point' for the number of optimised measurements in urban water networks. The inclusion of a cost objective in macro-location optimisation is addressed explicitly within this report.

D3.5.1 is closely linked to WP3.1 (Sensors: testing, calibration and evaluation of uncertainties) and WP3.3 (Data validation). It is also closely linked to WP3.6 (PREPARED, 2010) which directly addresses uncertainty issues in urban water systems. Naturally, the scopes of the two are different; this report targets the methods and algorithms previously developed for the optimal macro-location of sensors, while WP3.6 tackles uncertainty from a generalised perspective. The toolbox for optimal macro-location of sensors will be implemented in the real-time modelling and control platform (WP4.2), while methodologies for water quality monitoring (contamination warning) will be included in the early warning and control systems toolbox for clean water urban water systems (WP4.4).

1.3 Report Structure

Section 2 includes an introduction to hydraulic and water quality modelling for urban water systems and discusses the challenges associated with the development of effective and valuable schematic (mathematical) models. Sources of uncertainty resulting in a variety of modelling errors are identified (links to Deliverable 3.6.1) (PREPARED,

2011) and the need for parameter calibration procedures in both hydraulic models and dependent water quality models for WDS and UWWS is discussed in the context of relevant literature.

In section 3, the calibration procedures for steady-state and transient models are examined with the intent of identifying and categorising distinct calibration approaches. The calibration of water system models naturally requires data collected on a reference system and hence is directly dependent on sampling design.

Section 4 formally introduces sampling design and the sensor macro-location problem with specific regard to optimal placement schemes for *model calibration*. It introduces the groundwork and established theory underpinning key design criteria in sampling design literature in accordance with the calibration objective (A, D, V optimality); sources of error regarding comparison data arising from the sensors themselves and other stochastic elements of experimental processes are discussed and formalised. Optimisation procedures subsequently applied to these design criteria in sampling design (SD) calibration literature are referenced, along with recent work on stochastic SD procedures.

Section 5 addresses macro-location literature and methodologies developed for purposes other than model calibration; mainly referencing prevalent macro-location literature surrounding the recent Battle of the Water Sensor Networks (BWSN) aimed at optimal contaminant detection. Detailed accounts of existing macro-location approaches and algorithms that are largely independent of their objective functions are provided on the grounds of their applicability to general macro-location problems.

Section 6 is the conclusion of this document and summarises the current state of research with respect to macro-location of sensors. Some limitations of current sensor placement methodologies that have been identified by researches are highlighted in this section. The future path with regards to PREPARED objectives in linked work areas is briefly discussed.

Finally, Appendix A contains a summary of core optimisation procedures and algorithms developed thus far by researchers for a range of sampling design goals. For generic optimisation procedures listed in this table, objective functions for model calibration would likely be based on the A, D and V optimality criteria.

2 Challenges in Water Systems Modelling

Effective modelling of clean Water Distribution Systems (WDS) and Urban Wastewater Systems (UWWS) is of key interest to water utilities in the face of global climate change and growing demand on limited water resources (Jamieson et al 2007, Savic et al 2009). Numerical models of urban water systems can provide designers and operators with valuable data and predictions regarding pertinent aspects of the system state under past, current, future, or hypothetical operating conditions. A model may be used to discern desired quantities in the real system in absence of direct measurement and thus facilitate cost-efficient online monitoring and control (a major objective within PREPARED). Additionally, numerical models of urban water systems may be implemented as an accessory in design measures to improve the quality of clean water or wastewater services with significant financial implications (AWWA 1999).

The vast size and complexity of most urban water networks requires that the construction of adequate models be approached with care by researchers and practitioners, although, the challenges associated with model building are somewhat alleviated by a raft of modern commercial and open-source utilities developed specifically to aid WDS and UWWS model building practices (‘Epanet’: Rossman, 2000; ‘SWMM’: Rossman, 2004). In any case the reliability of model output predictions remains a principal concern afflicting effective analysis of ‘real world’ water distribution and waste water systems (Kapelan et al.,, 2007). Modelling discrepancies can arise from structural (epistemic) uncertainty in the models themselves i.e. uncertain mathematical relationships between physical quantities stemming from incomplete or deliberately approximated knowledge of the system, and from uncertainty in values for model parameters e.g. Hazen-Williams roughness coefficients (Walski, 2003).

Significant differences exist between WDS and UWWS modelling approaches and the specifics of uncertainty are naturally dependent on the operating conditions of individual networks. A brief discussion of both WDS and UWWS modelling is continued in the next sections.

2.1 Water Distribution Systems (WDS) Modelling

A Water Distribution System is essentially a network of pipes, pumps, tanks, valves and other appurtenances intended to transport (and ensure the quality of) drinking water at sufficient pressure and volume for end consumers, be they domestic, commercial or industrial users with differing service requirements. The scale and complexity of such systems, combined with the importance of a consistent and high-quality service, have all contributed to the significant challenges hindering effective and efficient management processes. The cost of obtaining useful data from a WDS is often prohibitive, and comprehensive studies on (typically) vast water networks are highly impractical, particularly when data is desired under extreme or disruptive operating conditions (Walski et al., 2003). In lieu of empirically derived data and expert 'guesswork', accurate forecasting and virtual sensor systems based on numerical modelling have become the foundation of vital tools for the future of well-informed and cost-effective decision making within the water distribution industry (Savic et al., 2009). In addition, the recent drive towards 'real-time' data collection and analysis has opened the door to significantly more accurate and continuously evolving models, which facilitate effective and reactive control processes (Hall et al., 2007). WDS modelling may be divided into two components: quantitative modelling for the pressures and flows in the bulk fluid, and water quality modelling to track contaminant concentrations and processes.

2.1.1 WDS Quantitative-Hydraulic Modelling

Quantitative modelling of water distribution systems originated in the 1930's on the back of a mathematical technique developed for moment distribution analyses -- which was adapted to solve for pressures or flows in closed-loop water distribution systems (Cross, 1936). Later, and subsequent to the dawn of the computing age in the 1950s and '60s, the so-called 'Hardy Cross' method was readily applied to larger water distribution networks and began to reveal a number of prohibitive limitations; including lengthy convergence times and an occasional propensity to fail given particular network conditions (Ormsbee, 2006). In response, numerous alternatives were investigated including a noteworthy improvement to the original approach by Cross via application of the Newton-Raphson algorithm to simultaneously solve for flow adjustment factors (Flower, 1969). Other approaches include the simultaneous node method (Martin and Peters, 1963; Shamir and Howard, 1968), the linear method (Wood and Charles, 1972) and the more recent 'gradient' method first suggested in 1987 (Todini and Pilati, 1987) and implemented in the now prevalent 'Epanet' modelling software (Rossman, 2000).

Mass and energy equations, balanced to estimate velocities or pressure heads are naturally dependent on defining characteristics and quantitative relationships within the 'real' system, and require parameter values which in practice are difficult or impossible to determine with any accuracy. Poor estimates of these values (and of the structure of the model itself) can lead to serious error and resulting loss of modelling usefulness and efficacy (Kapelan, 2002), further elaborated in section 2.2.3. Hydraulic flow solvers implemented in WDS networks form the basis for dependent interactions regarding pollutants and chemical substances within the fluid body Falkoner, 1992; Rossman, 2004) i.e. water quality modelling.

2.1.2 WDS Water Quality Modelling

Water quality models first appeared in the early 1980s (Wood, 1980; Chun and Selznick, 1985; Grayman et al, 1988). They have been implemented as a tool for analysing flow paths, for contaminant source identification, for general tracer modelling (Shang et al., 2002; Ostfeld, 2005), and more recently for the assessment of sampling designs with regards to tracer calibration (Preis and Ostfeld, 2006, Ostfeld et al, 2008). Since water quality models rely on hydraulic flows, the usefulness and reliability of water quality modelling predictions is dependent on the accuracy of underlying hydraulic models and solvers. Accordingly, tracer studies can provide detailed information on water travel times and flow paths throughout a water distribution network which may in turn be used to verify the accuracy of hydraulic model flows (Kennedy et al., 1991; Clark et al., 1993).

Water quality parameters are affected by diverse mechanisms governing the behaviour of numerous pollutants (both conservative and non-conservative substances) contained within the bulk fluid. Models for conservative contaminants may be directly dependent on the flows determined by a quantitative hydraulic model in addition to some water quality modelling assumptions e.g. complete and instantaneous mixing at junctions (Austin et al., 2009); however, the accurate modelling of a non-conservative substance (such as free chlorine) requires an appropriate reaction model and associated additional model parameters (Nagatani et al., 2008). The most commonly modelled non-conservative contaminants are disinfection by-products, such as total trihalomethane and halogenated acetic acids (Savic et al., 2009).

2.1.3 Sources of Uncertainty in WDS Modelling

Water distribution systems contain a large number of components leading to models that are inherently complex and reliant on a substantial number of model parameters (including one roughness coefficient, and perhaps one wall-reaction coefficient, per pipe segment) which must be estimated or determined via direct measurement of a physical system. Empirical data in this context is fundamentally susceptible to numerous sources of error, including (Fillion and Karney, 2003; AWWA 1999):

- *Inaccurate data regarding pipe roughness coefficients*; aging effects, erroneous design specifications and the highly stochastic changes to the interior surface of pipes during extended use ensure that roughness properties are difficult to predict accurately without experimentation (Sharp et al., 1988). It is practically impossible and prohibitively expensive to test each pipe in a water network to obtain data on physical characteristics, and hence a significant amount of uncertainty remains in the specification of related modelling parameters. Providing sufficiently accurate roughness coefficients for even moderately-sized networks can be a considerable undertaking.

- *Inaccurate data regarding water quality reaction coefficient; t*he potential for error within water quality parameters is equally acute as many water quality models require separate coefficients describing reactions that take place within the bulk fluid and reactions that take place between the fluid and the pipe wall i.e. two additional parameters for each pipe segment in the network (AWWA 1999). Obtaining bulk reaction coefficients is usually not difficult and there are trusted procedures available to acquire the data. Unfortunately, wall reaction coefficients can be a great deal more challenging, likely being unique for each pipe segment (Osman et al., 2002), leading to uncertainty issues in WDS modelling.

- *Outdated or estimated pump characteristics*; pump response curves are usually supplied by the manufacturer, although in practice they do not typically operate at this efficiency (Walski et al., 2003). The problem is compounded by the deterioration of performance over time due to a host of stochastic and highly complex interactions i.e. cavitation damage and various forms of wear associated with the everyday operation of the device(s). Re-testing of pumping units would ideally be completed at regularly intervals, but in practice this is often avoided due to the cost and inevitable service disruptions (AWWA, 1999).

- *Errors in measurement of component dimensions and elevations*; the dimensions and locations of physical structures on the network are not generally known with precision and manufacturer specifications may be inaccurate or erroneous (i.e. old pipes often vary in diameter across their length; AWWA, 1999). Elevation data for pipes is often obtained via a contour map of the ground (AWWA, 1999), under the 'fairly good' assumption that pipes lie a fixed distance below the surface -- a method plainly susceptible to error and a significant source of inaccuracy affecting head and pressure estimations for water system models.

- *Numerous other sources*; these include '*skeletonisation*' of the Water Distribution Network, a common modelling procedure whereby particular non-critical pipes are modified or removed from a given model i.e. pipes in series and in parallel may be merged into a single representative pipe via the application of appropriate hydraulic theory (AWWA, 1999). Pipes with small diameters and pipes leading to dead-ends are often removed entirely, with the demand being reapportioned to nearby nodes (AWWA, 1999). Most significantly of all, uncertain and simplified *nodal demand* patterns and their *diurnal variation* may significantly contribute to inaccuracies (Kapelan 2010, Kapelan et al, 2003), while *unknown valve and tank settings* are also significant sources of uncertainty in hydraulic modelling (Laucelli et al, 2010). These and other sources of uncertainty are addressed rigorously by the interrelated Work Package (WP3.6) and Deliverable (D3.6.1) (PREPARED, 2011).

It is therefore unlikely that numerical models will be sufficiently reliable in their first instance. Rather, it is widely accepted that water models must in general be 'calibrated' before becoming viable for practical application

(Walski, 1983). Model calibration is discussed in detail in section 3. It is based on the collection of real time performance data from the network (with the help of sensors), and hence is related directly to optimal macro-location (to be addressed in Chapters 4 and 5).

2.2 Urban Waste Water Systems (UWWS) Modelling

The urban waste water system (UWWS) is charged with the removal of domestic, commercial and industrial wastewater ('dry weather flow') as well as surface rainwater influent, with the goal of maintaining suitable sanitation and reducing the risk of flooding in urban environments. Urban waste water systems (UWWS) typically consist of three distinct components; the sewer system, the treatment plant(s) and the receiving water body (Vanrolleghem et al., 2005). The diversity of UWWS components and processes has encouraged segregation of modelling concepts between each of the three major subsystems i.e. they have most often been analysed individually, as opposed to being considered as a complete system (Muschalla et al., 2008). That being said, some research in recent years has investigated a range of integrated modelling approaches (Vanrolleghem et al., 2005) intended to capture the UWWS sub-system interactions in a more discerning manner. Modelling concepts for waste water treatment processes and receiving water quality are not directly discussed in this report, as sensor macro-location is most relevant to sewer systems.

Sewer networks around the world exist in one of two forms; they are either a separate system or a combined system. Separate systems have two distinct pipe/channel networks, one intended for the transportation of relatively clean urban rainwater runoff, and the other intended for domestic, commercial and industrial wastewater product en route to a treatment facility (and ultimately the receiving water body). A combined system has a single pipe/channel network designed to cope with both the waste water product and the urban surface run-off simultaneously. Despite the disadvantages of a combined system (Beychok, 1971), they are notably cheaper to construct by virtue of fewer pipes and are the most common form of sewer system found in major cities, particularly in Europe and the United States (EPA, 1999).

Environmental regulations, such as the EC Water Framework Directive, have created a legal requirement for sewer operators in Europe to predict the flow quality as well as quantity of sewer system releases. In the United Kingdom, sewer systems are managed and maintained by private water companies subject to assessment by the Office for Water Regulation (OFWAT) and the Environment Agency (EA). Sewer flow quantity and quality modelling procedures may provide suitable evidence that the sewer system is meeting accepted environmental standards to these governing bodies, and it is therefore of principal importance that they be sufficiently reliable (Schellart et al., 2008). As a result, active, online monitoring and control are ongoing developments in UWWS research and the use of models for planning and analysis is now standard practice for many waste water utilities (Schutze et al., 2005).

2.2.1 Quantitative Modelling in Sewer Systems

Quantitative modelling of sewer flows is typically based on the Saint-Venant 'shallow water' equations, derived directly via vertical integration (applicable because the horizontal velocity field is nearly constant throughout the depth of the fluid) of the Navier-Stokes equations describing the conservation of mass and energy (momentum) for Newtonian fluids (Randall, 2006). They are a system of partial differential equations requiring potentially time-consuming numerical integration. The Saint-Venant equations assume that unsteady flow in open channels is one-dimensional, that the fluid is homogenous and incompressible, and that the pressure distribution is hydrostatic, along with some other geometric limitations (Strelkoff, 1970). Since the solutions of these equations (or more accurately, their approximations) can be extremely computationally demanding, several further simplifications are commonly applied to reduce the equations in situations where particular conditions are present or may be reasonably assumed (Havlik, 1996). Often in cases where explicit Saint-Venant derived simulations are operationally prohibitive or notably inefficient, less complex surrogate models may be employed to reduce computation time i.e. The KOSIM model (Meirlaen et al., 2002).

In addition to the modelling of shallow channel flows, the sewer system is closely related and dependent on the urban rainwater run-off from linked geographic catchments areas. The amount of rainfall reaching the sewer in urban areas is affected by numerous factors including the geography of permeable and impervious areas (and dependent filtration processes), depression storage behaviour and evaporation losses (Schutze et al., 2002). Surface run-off models consider these processes various degrees of detail, within which some of all of the above processes may or may not be taken into account (Fuchs, 1996). Subsequent to approximations of losses, the net rainfall is transported on the surface until it reaches entry points to the main sewer network. This process is often modelled by one or a cascade of several linked linear or non-linear reservoirs (Schutze et al., 2002).

2.2.2 Water Quality Modelling in Sewer Systems

Water quality modelling within urban sewer systems is diverse field involving complex biochemical processes and pollutant/sediment transport models. In addition to classical water quality concerns, sedimentation behaviour and resuspension of pollutants is a significant factor in the determination of overall sewerage quality (Jack et al., 1996). Sediment transport is influenced by geometry, spatial and temporal variations in hydraulic conditions (significant in sewers) and all-manner of other interactions (Ashley and Verbanck, 1996). Sedimentation mechanisms are complex and contain poorly understood transitions between 'active' sediment layers and consolidated/storages layers, along with inevitable difficulties regarding unknown inputs of 'gross solids' and their resulting interactions with sediment layers (jack et al., 1996; Schutze et al., 2002, Bouteligier et al., 2004).

Sources of uncertainty in UWWS models are numerous and diverse, suffering from increased complexity in both quantitative and water quality interactions. Some principal sources of error for both quantitative and qualitative modelling are discussed in the following section.

2.2.3 Uncertainty in UWWS Models

Urban Waste Water Systems (UWWS) are typically large, convoluted systems incorporating multiphase open channel flows and highly complex suspended solid transport and water quality related decay mechanisms, which naturally endure numerous potential sources of uncertainty accompanying numerical modelling schemes. Major difficulties associated with UWWS modelling include:

- *Inaccurate rainfall predictions and data*; a precise determination of the magnitude and frequency of rainfall events which unmistakably provide a key input into the physical waste water system is a significant challenge and source of potential discrepancy in numerical modelling. *Rain gauge measurements* required to gather the necessary empirical data are subject to systematic errors due to wind effects (Willems 2001), inaccurate gauge calibration curves (Stransky et al. 2007) and random errors due to a range of other processes i.e. mechanical problems such as clogging. Additionally, the *spatial and temporal resolution of rain gauge devices* can have a prominent impact on the quality of the data collected and the subsequent performance of the urban drainage and sewer flow models. Most gauges collect rainfall data on a small area ('point') and therefore the total rainfall profile must be derived via interpolation of the data collected at these selected points (Zhu and Shaofeng 2004). Similarly, temporal resolution may be an issue whereby rain gauges are unable to record sufficient data points in time with suitable accuracy and is typically dependent on the intensity of rainfall, particularly with regards to the commonly used bucket gauges. The efficacy of the interpolation procedure in time and space logically depends on the quality and quantity of point gauge data and upon the assumptions made during the interpolation process, but invariably introduces error into relevant model input data.

- *Poorly represented urban drainage processes*; within urban drainage modelling (especially for water quality modelling), the state-of-the-art may be considered somewhat less advanced in comparison to similar fields (Deletic et al., 2009). Complex surface run-off processes, particularly with regards to water quality modelling, are a major cause of difficulty; urban sewer systems are typically encompassed within larger rainwater catchments and hence accumulate rainfall and polluted surface run-off from wider hydrological areas. The rainwater/surface sludge enters the sewer system by infiltration through permeable and semi-permeable surfaces, surface flows and a variety of other means affecting input into the physical UWWS. The inclusion of a detailed urban run-off model may help to ease the errors introduced by ignoring or naively estimating this form of wet weather input, however such surface run-off models naturally suffer from their own sources of error (Kunstmann et al., 2002). Further compounding the problem, the state of knowledge regarding uncertainties in urban drainage models appears to be quite poor (Deletic et al., 2009), perhaps due to a lack of clarity and consensus on the way in which the results of model uncertainty analyses are obtained, presented and used. The significant computational effort associated with rigorous uncertainty quantification in urban drainage models (Muschalla et al., 2009) is cited as one potential impediment to the field.

- *Inaccurate urban waste predictions and data*; in addition to rainwater flow, effluent from domestic, commercial and industrial users represents a major input (and the primary source of pollution) into the urban waste water system. Similarly to demand uncertainties for water distribution networks, client effluent patterns in UWWS are highly stochastic in nature, consisting of varying effluent mixtures from numerous industrial processes along with domestic installations, and are extremely difficult to quantify precisely in terms of their volumetric and temporal distributions (Butler 1993). On a large scale however, moderately predictable diurnal usage patterns may be determined and provide reasonable estimates of required input data, nevertheless some degree of uncertainty is inherently introduced into model the inputs.

- *The determination of physical characteristics;* similarly to those problems encountered in WDS modelling, UWWS models contain a range of physical parameters (i.e. roughness coefficients) along with dimensional data regarding the widths, heights and vertical placement of the structures comprising the UWWS. Many of these parameters are difficult and/or costly to measure accurately and typically generate some form of uncertainty in model inputs as discussed in the previous section on sources of uncertainty in WDS. The problem is perhaps more acute in UWWS application since the parameters such as channel roughness coefficients tend to exhibit higher degrees of spatial and temporal variability according to a number of affecters including, in particular, complex and stochastic sediment behaviours (Crabtree 1989).

- *Numerous other sources;* both water distribution and urban waste water systems are hugely complex hydrodynamic and inherently stochastic systems with a significant potential for uncertainty from systemic, experimental and data transmission sources. A complete assessment of urban water system uncertainty is outside the scope of this report, and is to has been formally addressed by deliverable D3.6.1 (PREPARED 2011).

Sources of uncertainty within UWWS models are typically more severe and diverse juxtaposed against those associated with modern WDS modelling. Indeed the increased system complexity of both the quantitative and water quality processes occurring within sewer systems further highlight the need for effective and efficient data gathering and model calibration measures in advance of feasible implementation at the industrial level.

3 Model Calibration

3.1 Introduction

Models of Urban Water Systems suffer from numerous sources of uncertainty and invariably rely on physical parameters which must be sufficiently determined prior to any successful application of the model in a realistic setting (Section 2). Furthermore, the necessity and importance of effective calibration procedures in practical modelling scenarios has been clearly identified by numerous researchers and practitioners in the field (Walski 1983, Ormsbee 1989).

A calibration process modifies chosen model parameters (e.g. pipe roughness) until predictions presented by a model match closely with measured values (or values calculable from measured values) on the real network, under a range of operating conditions (scenarios) within acceptable error bounds. If parameter calibration is unable to obtain sufficiently small error bounds (as determined by a post calibration uncertainty analysis or validation procedure) then the model itself may need to be restructured to remove epistemic uncertainties inherent in the mathematical description (Walski, 2003).

Parameter values for an uncalibrated hydraulic model are regressed according to data recorded on the physical system. Intuitively, the set of experimental data used for model calibration is one of many **possible** sets that **could have** been collected; measured responses will vary according to uncontrolled operating conditions, sensor inaccuracies and other stochastic elements comprising the physical system. Formally, a model generates y output quantities from vectors of input variables, x, and parameters, θ.

$$y = f(x,\theta) \tag{3.1.1}$$

In a calibration procedure the vector of parameters are to be regressed from known measured values, y* (Datta et al., 2004).

$$y^* = f(x,\theta) + \varepsilon \tag{3.1.2}$$

The error term, ε, captures the discrepancy between experimental values y* and the model predicted values y. For a particular vector of measured quantities, θ is selected so as to minimise the error term in equation 3.1.2.

3.2 Steady-state and Extended Period Simulation Model Calibration

Numerous calibration procedures for urban water networks have been developed since the 1970s. Generally, calibration models may be grouped into three categories:

Iterative calibration methods are based on custom trial-and-error procedures and typically rely on the skill and experience of the operator for effective calibration runs (Walski, 1983). In these procedures, unknown parameters are updated using output variable quantities (e.g. pressure heads) obtained by solving the set of steady-state mass balance and energy equations for the hydraulic network. Simplification (e.g. skeletonisation) is usually required as only small calibration problems (with a small number of calibration parameters) are workable in practice (Savic et al., 2009). The convergence rate of the iterative models is generally very slow (Bhave, 1988) and these types of procedure are all but obsolete from a research standpoint (Savic et al., 2009).

Explicit calibration models are based on solving an extended set of steady-state mass-balance and energy equations (Ormsbee and Wood, 1986; Ferreri et al., 1994). This extended set consists of initial equations describing a steady-state network model and a number of additional equations derived from available head and flow measurements. The extended set of equations is to be solved explicitly and hence the number of unknown calibration parameters is limited by the number of available measurements. In the event that the number of unknown calibration parameters is larger than the number of available measurements, the calibration parameters must be grouped. The key disadvantages and limitations of explicit calibration methods are as follows (Savic et al., 2009):

1. The number of calibration parameters to be determined must be equal to the number of measurements.
2. Measurement errors are not taken into account.
3. There is no way to quantify uncertainty in the estimated calibration parameters.

4. These methods require considerable expertise, experience and sophisticated solution tools.

Implicit calibration processes employ an objective function to minimise the differences between measured and model predicted variables. This is done via some optimisation technique(s) coupled with a hydraulic solver. Ormsbee (1989) used an implicit calibration method for WDS multiple steady-state and extended period simulation network models. Further improvements and applications of implicit calibration were reported by Lansey and Basnet (1991), Pudar and Liggett (1992), Savic and Walters (1995), Reddy et al. (1996), Lingireddy and Ormsbee (1999), Greco and Del Giudice (1999), Todini (1999), de Schaetzen (2000), Lansey et al. (2001), Kapelan et al. (2004), Boccelli and Uber (2004, 2005), van Bloemen Waanders (2004), Wu and Sage (2006) and Kapelan et al. (2007).

More recently, the Battle of the Water Calibration Networks (Ormsbee et al., 2010) lead multiple authors to propose calibration algorithms for water distribution network models; tending to build on existing ideas with a focus on greater computational efficiency. These proceedings include papers by Bros and Kalungi (2010), Burd et al., (2010), Kang and Lansey (2010), Shen and McBean (2010), Wu and Walski (2010), Alvisi and Franchini (2010), Johnson et al., (2010), Koppel and Vassiljev (2010), Kim el al. (2010), Diao et al., (2010), Chang et al., (2010), Prasad (2010), Laucelli et al (2010) and Asadzadeh et al., (2010).

3.3 Calibration of Transient Hydraulic Models

So far, sampling design for the calibration of transient models has been given limited attention by the researchers. Pudar and Liggett (1992) and Liggett and Chen (1995) suggested measuring pressures at the most sensitive locations in a WDS. Vitkovsky and Simpson (1997) analysed the effects of the number of pressure measurement locations and length of pressure measurement records on the values of estimated calibration parameters. Analysis was performed on the small network synthetic case study, with the following conclusions:

1. The average difference between estimated friction factors and their 'true' values is related to the number of measurement sites and increases with reduction in the number of measurement sites.

2. When using perfect transient pressure data, no gains were experienced from longer pressure records. However, it is expected that when imperfect data is used, the longer the records the greater the chance that correct calibration parameter values will be found.

More recently, Vitkovsky et al. (2003) successfully employed the Shuffled Evolution Complex (SCE-UA) algorithm for the deterministic calibration of transient hydraulic models. The same problem was solved by Kapelan et al. (2003) who developed a new hybrid search algorithms based on Genetic Algorithms and Levenberg-Marquardt methods.

3.4 Calibration of Drinking Water Quality Models

The calibration of a water quality model requires an extended period simulation to predict hydraulic behaviour (Savic et al., 2009). Most of the water Quality Model calibration studies assume a priori that the underlying hydraulic model flows and demands are correct, which is not always the case (Jonkergouw et al., 2008). It has been shown that the reliability of results from water quality model parameter calibration is not only largely dependant on the quality of the calibration data (Maier et al., 1999, Rhode et al 2007), but also on the accuracy of the underlying hydraulic model. It is important to note that a hydraulic model may produce accurate pressure values, and be 'well calibrated' in some sense without any guarantee that the hydraulic flows relevant to water quality processes are equally acceptable.

Implicit calibration models use an optimisation algorithm to minimise a given error statistic between simulated and observed data until the algorithm convergences or a predefined termination criterion is reached. Although the implicit calibration of hydraulic models dates back at least as far as 1974 (Shamir, 1974), implicit models for the calibration of water distribution system chlorine decay models have only appeared more recently (Zierolf et al., 1998; Shang et al., 2002; Munavalli and Mohan Kumar, 2003b; Shang, 2005; Munavalli and Mohan Kumar, 2005; Yang et al., 2006; and Huang and McBean, 2006). Both stochastic and deterministic (gradient-based) optimisation techniques have been used to calibrate water quality model parameters, although the latter is far more common (Savic et al., 2009).

3.5 Calibration of Urban Wastewater System Models

The calibration of urban waste water models has been confronted in much the same way as that for clean water distribution systems and seeks plainly to minimise discrepancies with respect to a physical reference system, most commonly via implicit methods (di Piero et al., 2005). The collection of adequate data for the purposes of UWWS model calibration remains problematic, however the calibration processes themselves are mature and well defined in literature for all three sanitation sub-systems (as well as for integrated models), (Belia et al., 2009); the sewer system, treatment plant and receiving water processes have been extensively modelled and a range of techniques have been employed with the aim of calibrating model parameters for relevant biological and chemical processes within these systems (Petersen and Gernaey 2002, De Clercq et al., 1999, Gronewold et al., 2009).

The determination of model parameters (and their sensitivity to output variations) acutely depends on the operating state / boundary conditions of the system during the data collection process. Considering that it is generally only practical to effectively calibrate parameters that are somewhat sensitive to model output, it is entirely possible that particular insensitive model parameters may be vastly incorrect or unphysical subsequent to a given calibration procedure which, in conjunction with remaining uncertainty in the well-calibrated parameters, leads to poor predictive capabilities of a calibrated model given inputs and boundary conditions far removed from those utilised in the calibration process. This can become a serious concern in the face of effective UWWS model calibration due to the vastly different influent conditions observed for wet weather and dry weather flow. UWWS models should be calibrated appropriately for their intended application, and some researchers have even approached the optimal parameterisation as a multi-objective problem (di Piero et al., 2005).

UWWS models (and particularly integrated models) may suffer difficulties stemming from the shear complexity and prohibitive calculation times of existing UWWS hydraulic and water quality models (Meirlaen et al., 2002). In cases where ‘complete’ hydrodynamic models based on the de Saint Venant equations are indeed prohibitively demanding, they may be effectively and efficiently substituted by faster surrogate models. These surrogate models contain less information in their physical concept, which must be remunerated with more and better data in the calibration process (Meirlaen et al., 2002). However, since data collection is an expensive and time consuming task (Vanrolleghem et al., 1999) some researchers propose that instead of collecting all of the data needed to calibrate a surrogate model, empirical data are used merely to calibrate the ‘complete’ schematic model; and once this model is validated it may be subsequently used to generate new virtual data for the calibration of the surrogate model(s). In this way, some of the information contained with the complex model is imparted to the surrogates.

3.6 Relevance to Sampling Design

Within PREPARED, measurements taken from urban water networks will be implemented for the calibration of steady-state hydraulic models and water quality models intended for the real-time monitoring and operation of those networks. In this respect, D3.5.1 is closely linked with WP3.1 (Sensors: testing, calibration and evaluation of uncertainties) and WP3.3 (Data validation).

This review is the foremost task in the development of a prototype optimal sensor placement tool (Deliverable 3.5.2), to be demonstrated at a later stage in the project. However the final demonstration of applications and models largely depends on the PREPARED cities, which will be involved in WP3.5 as demonstration cities, and have not yet been defined. The specific models, data and problems these cities may have, will ultimately determine the content of the modelling and calibration approach.

4 Sampling Design for Calibration

4.1 Introduction

The calibration of water distribution system (WDS) and urban waste water system (UWWS) models necessitates the collection of data from a physical water network. Due to the scale of most urban water systems, it is practically impossible to install sensors or to take measurements at all of the candidate locations. In order to obtain information for an estimate of the model parameters, data must be collected from a subset of carefully chosen locations designed to maximise the performance with respect to specific design criteria. Prior to 1998, few authors had investigated optimal sampling design for urban water systems (Bush and Uber, 1998). However, Walski (1983), Lee and Deininger (1992), Stukel et al. (1987), Liggett and Chen (1994) amongst others had demonstrated the significant impact of sensor location on inferences regarding the system state. In general, the aim of the sampling procedure for water distribution model calibration is to determine (Kapelan et al., 2003):

1. The model prediction variables (pressures, flows, water quality parameters, etc) to be observed.
2. Where in the system to observe them (location).
3. When to observe them (duration and frequency).
4. What conditions should exist during the observation (e.g., demand condition(s), etc).

Naturally, problem (2) is the focus of this review; however, all of the other factors play an important role in the definition of an optimal sampling design. The sampled data used for the calibration of water system models should yield an optimal agreement between model outputs/predictions and the physical system.

At the outset of a calibration procedure the vector of parameters are assumed uncertain and must be regressed from measured values, y* (Section 3.3.1).

$$y^* = f(x,\theta) + \varepsilon \tag{4.1}$$

where $f(x, \theta)$ are values generated by a given model dependent on some input quantities x, and some model parameters, θ. Logically, the optimal value(s) of the parameter vector, θ, depend(s) on the measurements y* which are subject to random and other errors, i.e. uncertain. During the calibration process, the uncertainty in y* is propagated to the calibration parameter θ uncertainties. The magnitude of variation in recorded data, *and the extent of the impact of those uncertainties on model parameters*, and ultimately model predictions for a particular operating state are potentially reliant on the location of the sensor devices (Lansey et al, 2001; Kapelan et al, 2005).

4.2 Optimal Macro-location of Sensors

'Optimal macro-location' has been thus far been defined as the 'ideal placement' of a limited number of sensors within an urban water network (i.e. a number of optimal locations at which measurements of some quantity/quantities should be taken). It is important to note that most commonly, within the realms of literature for sampling design, the uncertain spatial parameter is not optimised on a continuum of potential locations, but rather selected from a finite domain of predefined 'suitable' placement options. Precise definitions of the sensor placement (SP) problem, as accepted near-unanimously by relevant sampling design literature, are discussed with respect to clean water distribution systems and urban waste water systems in the sections that follow.

In order to develop a robust optimal sampling design it is necessary to quantify and define measures of effectiveness suitable to assess any given sensor layout on a water distribution or waste water network. The purpose of a particular sampling design within the context of this report is to provide the 'best' empirical data from a physical system to be used for the calibration of uncertain parameters within a numerical model of that system. Logically, the efficacy of a calibration may be determined by the amount of uncertainty remaining in model parameters subsequent to an application of the procedure. The average reduction in parameter uncertainty, subsequently related to prediction uncertainty, is generally dependent on the location of sensors.

4.2.1 Problem formulation for Water Distribution Systems

The set of potential locations for sensor placement in Water Distribution Systems is most commonly defined as the finite set of all nodes (junctions) on the given network (Bush and Uber, 1998, Ostfeld et al. ,2008):

Box 1: Definition of the general Sensor Placement Problem

Network junctions (nodes) are usually considered to be the complete set **N** of viable locations for sensors. A limited number of sensors, **S**, must then be located on a subset of network nodes $\mathbf{N_S}$ ($|\mathbf{N}_S| = \mathbf{S}$), where $|\mathbf{N}_S|$ represents the cardinality of the subset $\mathbf{N}_S$.

Optimal design, in general, aims to find a “best” solution or “set of best solutions” amongst all other potential (feasible) solutions to a given problem. The search space (the set of all feasible solutions) may be vast and largely unknown at the outset. The general sensor placement problem is combinatorial with a potentially enormous corresponding search space, given by equation 4.2.1 below.

$$C_s = \binom{|Y|}{N} = \frac{|Y|!}{N!(|Y|-N)!} \tag{4.2.1}$$

where |Y| is the cardinality of the set of possible sensor locations and N is the number of sensor stations to be placed. Optimal designs should avoid situations where ‘useless’ data are collected, for example, when the head loss in the system is low causing head loss and velocity to be of a similar order of magnitude as the errors in measurement (Walski, 2000). Similarly, ‘bad’ data, i.e. data containing gross errors, should be discarded in the model assembling process before it can be used unintentionally in model calibration (Savic et al. 2009). As has been stated, optimal designs should also avoid unnecessary expense in collecting redundant data, i.e., data whose information contribution is already contained within another measurement.

4.2.2 Problem Formulation for Urban Waste Water Systems

The vast majority of existing literature on the subject of optimal macro-location for sensor placement addresses clean water distribution systems.

There are very few references in sampling design literature for optimal sensor placement in UWWS. The reasons for this appear to be manifold (Dempsy et al. 1997, Borgeois et al. 2001, Korving and Clemens 2005):

1) Sewer systems are largely gravity driven and often lack effective means of control in response to model feedback (curbing industrial drive for online monitoring).
2) Sewerage utilities are not required to provide the same level of service as clean water distribution networks i.e. they do not need to ensure particular minimum pressures or water quality indicators across network nodes, again resulting in slower take-up of online modelling and monitoring technologies.
3) Sensor networks in wastewater systems experience a multitude of complex practical concerns which have tended to perturb the advancement of the field e.g. clogging or impact damage by debris, complex sediment behaviour (Crabtree 1989), potentially explosive atmospheres and greatly varying water levels (perhaps necessitating a floating sensor platform) and host of other concerns (Winkler et al 2005).
4) In recent years (where much of the interest in sensor placement is found), research agendas have moved toward sensor placement for the detection of deliberate or accidental contamination in potable water networks; this is less of an issue for waste water systems due to the lack of direct human consumption.

The lack of availability of sufficient and reliable empirical data (as a result of (3)), has previously resulted in sewer models often being left uncalibrated. Sensor devices for real-time control in current SCADA systems may typically placed at a small number of strategic locations i.e. upstream of a channel / weir for a combined sewer outlet (Korving and Clemens 2005). In spite of the open-channel flows and the complicated mix of fluids, suspended solids and sediment in the network, the sensor placement problem for sewers may be naively formulated in much the same way as that for the Water Distribution System described in section 4.2.1. However, as indicated, sensor placement in real-world sewer systems are driven by a host of significant practical concerns and placement schemes should take into account factors such as variable clogging risk or sediment behaviour across the network.

4.3 Early Sensor Placement Methods

Researchers and practitioners concerned with the optimal sensor placement problem initially attempted to develop guidelines or quantifiable relationships for macro-location and the resulting calibration uncertainty. Walski (1983) was amongst the first to directly address the issues of the sampling design in the context of WDS model calibration, and he suggested to:

- Locate pressure measuring devices near points of high demand.
- Perform pressure tests on the perimeter of the skeletonised network, away from known heads.
- Perform multiple fire-flow tests with as large as possible fire flows at test hydrants.
- Collect both head and flow measurements.

Subsequently, the majority of the relevant study has been based on numerous forms of sensitivity analysis; a process designed to apportion the uncertainty in the output of a given mathematical model to distinct sources of variation in the input of that model, either quantitatively or qualitatively. Perhaps the simplest sensitivity criteria are defined solely by the Jacobian Matrix of model outputs with respect to calibration parameters (Bush and Uber, 1998), but may also include inferences from parameter or prediction covariance matrices. These criteria have been used either to rank potential locations in a heuristic sensitivity-based method (Ferreri et al. 1994; Bush and Uber, 1998; Piller et al., 1999), or to define a general optimisation problem (Lee and Deininger, 1992; Meier and Barkdoll, 2000; de Schaetzen et al. 2000).

4.4 Quantifying Calibration Uncertainty (Variance Methods)

Subsequent to the work of Bush and Uber (1998) and Lansey (2001), the variance (and corresponding confidence regions) surrounding uncertain model parameters and predictions has become the defining criterion with regards to optimal sampling design. Covariance matrices for model parameters and predictions in sampling design literature are generally approximated via the First-Order Second-Moment (FOSM) method (e.g. Piller et al 1999, Ahmed et al 1999, Nagar and Powell 2000, Meier and Barkdoll 2000, Lansey et al 2001, Lansey 2003, Kapelan et al 2003, Kapelan et al 2005), which is detailed in section 4.4.1 below. Sampling designs for calibration may be specifically assessed on the A, D and V optimality criteria which depend on the covariance matrices for parameter and prediction values.

4.4.1 The First-Order Second-Moment (FOSM)

Experimental data collected by a particular sampling design will contain some degree of error affecting the accurate calibration of model parameters. Assuming that a model itself is error-free, the parameter variance results from the propagation of these errors in measured quantities. The covariance matrix for the parameter vector may be estimated via a first order approximation as follows (Bard 1974, Wong 1984, Lansey et al. 2001, Kapelan et al 2003):

$$Cov_\theta = \sigma^2 \cdot Cur^{-1} \tag{4.4.1}$$

where σ^2 is the error variance in measured quantities and Cur^{-1} is the curvature matrix defined by:

$$Cur = J^T WJ \tag{4.4.2}$$

Hence:

$$Cov_\theta = \sigma^2 \cdot \left(J^T WJ\right)^{-1} \tag{4.4.3}$$

Where J is the Jacobian Matrix of Derivatives $dy_i / d\theta_j$ indicating the sensitivity of model predictions y_i (which directly correspond to measurements y_i *) to changes in parameters θj. W accounts for the relative difference in measurement variance and/or error covariance i.e. the data quality. If all of the measurements are independent and have a comparable error variance (not unreasonable if the sensing devices are similar), then the weights are unity (identity) and the covariance matrix can be further approximated by (Lansey 2001):

$$Cov_\theta = \sigma^2 \cdot \left(J^T J\right)^{-1} \tag{4.4.4}$$

Intuitively, the degree of uncertainty in the model parameters is further propagated onto predictions later generated from the calibrated model. The covariance matrix for the vector of model-predicted values Z_i is given by a first-order approximation. It should be pointed out that vector Z_i is ***not*** *necessarily the same as the set of measurements taken by the sampling design. For instance,* the model may predict the piezometric head at a node where no measurements were taken. (Bard 1974, Kapelan et al 2005):

$$Cov_z = Cov_\theta \cdot J^T J \tag{4.4.5}$$

4.4.2 Alphabetic Optimality Criteria

A particular sampling design naturally aims to minimise the variance/uncertainty in predicted quantities generated by a model. For a statistical model with more than one parameter (a vector of parameters), the variance of the parameter vector is a matrix (see section 4.4.1 above) and hence the problem of "minimising" the variance (and compressing the result into a single number) is a subjective one. Alphabetic Optimality criteria are single-number metrics which each attempt to capture a different aspect of the suitability of a particular experimental design (Kiefer and Wolfowitz 1959). It should be noted that a 'best design' is typically more complex than can be summarised by a single number and the analyst must take care to assess the results in context.

4.4.3 D-Optimality

The D-optimality criterion discussed by Kiefer (Kiefer, 1958) was the first of a list of criteria that later became known as "alphabetic optimality". The confidence region for a parameter dependent on **N** predictors is a measure of the accuracy of the parameter estimate obtained from a particular experimental design. The volume of a confidence region around a parameter estimator generated from a particular set of predictors is determined by the confidence level. The true optimum value of a parameter would be found inside of the confidence region in parameter space 100*(1 – A)% of the time. It follows that an experimental design which minimises the N-dimensional volume of the confidence region is optimal in some sense.

Box 2: The D-Optimality Criterion

The volume of this confidence region is proportional to the square root of the determinant of the covariance matrix for parameters; minimising the determinant $\|Cov_\theta\|$ results in a confidence region of minimum N-dimensional hypervolume in parameter space. A design meeting this criterion is said to be D-optimal.

$$\text{Confidence Volume} \propto \left(|Cov_\theta|\right)^{\frac{1}{2}}$$

=>

$$\text{Min}\,\{|Cov_\theta|)\} \qquad (4.4.6)$$

where Cov_θ is the covariance matrix for the model parameter vector θ. It is explicitly defined in section 4.4.1 (equation 4.4.1).

4.4.4 A-Optimality

The A-Optimality criterion considers only the variances of the parameter estimates (ignoring co-variances between estimates) and as such is dependent on the diagonal elements (the “trace”) of the covariance matrix.

Box 3: The A-Optimality Criterion

The A-optimality Criterion is satisfied by minimising the Trace of the covariance matrix of parameters.

Thus, the A-optimality Criterion can be written as follows:

$$\text{Min}\,\{Tr(|Cov_\theta|)\} \qquad (4.4.7)$$

where $Tr(Cov_\theta)$ is the trace of the parameter covariance matrix defined by equation 4.4.1; minimising this value equates to minimising the average of variance of the parameters (and their confidence *intervals*).

4.4.5 V-Optimality

In addition to quantifying the model parametric uncertainty resulting from a particular sampling design, it is often more useful to realise the effect of that parameter uncertainty on the confidence of any predictions ultimately generated by a model (Savic et al. 2009) e.g. how much of the parameter/input uncertainty makes its way into the model-predicted values.

A first order Taylor expansion about the mean value of the parameters provides a good estimate of the prediction covariance matrix inferred from the covariance matrix of the estimators, given by equation 4.4.4.

Box 4: The V-Optimality Criterion

The V-optimality Criterion is satisfied by minimising the Trace of the covariance matrix of *model predictions*.

Consequently, the A-optimality Criterion can be written as follows:

$$\text{Min}\,\{\text{Tr}(|Cov_z|)\} \qquad (4.4.8)$$

where Cov_z is the covariance matrix of the prediction variables and is defined explicitly in section 4.4 (equation 4.4.5). The optimal macro-location criteria presented in this section have been implemented as objective functions for many of the implicit sensor placement methodologies to be detailed in this report.

4.5 Sequential Heuristic Algorithms

As indicated, quantitative approaches in sampling design most often rely on measures of sensitivity derived from Jacobian matrices (Ferreri et al., 1994; Bush and Uber, 1998; Piller et al., 1999), and many of the resulting sensor optimisation algorithms have employed sequential ranking-type procedures. Within these algorithms, a new measurement location leading to the current greatest improvement in sampling design (SD) accuracy is added to the previously selected set of sensor locations, until a preset number of devices have been placed. To obtain optimal solutions in general, these 'greedy' sequential procedures would rely on the optimal set for N measurement locations always being a superset of the optimal set for N-1 locations; unfortunately, this has been shown to be an erroneous assumption in general (Kapelan et al. 2003). Despite the apparent limitations of this type of sequential approach, the relative computational ease of simple sequential algorithms lends credence to their viability for optimal sampling design on large water networks.

Three sequential algorithms for the placement of measurements in a network (Bush and Uber, 1998) are detailed in this and the following two sections. Each of these methods aims to generate a set $\mathbf{R_k}$ of ranked nodes/measurements, where rank specifies the suitability of a particular location and measurement type for the estimation of model parameters.

All methods analysed in this ection begin with a set of unranked network locations $\mathbf{U_k}$ and proceed to select locations from this set according to desirability, placing them in a ranked set $\mathbf{R_k}$. Nodes selected at the iteration step **k** are assigned the rank **k**, hence nodes selected first are those found most desirable against the selection criteria. The algorithms are motivated by D-optimality (as described in 4.4) but do not attempt to solve the D-optimal problem directly. Minimising parameter uncertainty by most accepted measures in sampling design literature corresponds to minimising the variance of those parameters; or equivalently maximising the determinant of the inverse of the variance matrix (proportional to the fisher information for the design). The inverse variance matrix is approximated via a first order expansion as detailed in section 4.4 (FOSM method) as follows (Bard 1974):

$$|V_\theta^{-1}| = \det\left[\left[\frac{\partial f}{\partial \theta}\right]_{\theta*}^T V_e^{-1}\left[\frac{\partial f}{f\theta}\right]_{\theta*}\right] \qquad (4.5.1)$$

where f is the vector of model predicted values and θ is the vector of model parameters. V_e^{-1} depends on the covariance matrix of the residual errors. Equation 4.5.1 loosely suggests that the best measurements to include in a sampling design are those for which normalised model predictions are sensitive to changes in estimated parameters, and this is the basis for the method max-sum, max-min, and weighted-sum algorithms.

4.5.1 The Max-Sum Method

The max-sum design algorithm ranks measurements relative to the sum of the normalised magnitude of their sensitivity coefficients ($df/d\theta|_{\theta*}$).

Intuitively, taking absolute values allows both positive and negative coefficients to contribute equally and is consistent with design theory.

$$MS_i = \sum_{j=1}^{n} N_{ij} \tag{4.5.2}$$

where MS_i determines the rank of each node (higher values are preferred). $N_{i,j}$ is the sensivity co-efficient $df/d\theta|_{\theta^*}$ for measurement I and parameter j. To its benefit, the max-sum design is the simplest of the methods and the easiest to implement.

One disadvantage of the approach lies in its sole consideration of the aggregate sensitivity across all variables, for instance by considering a situation, where several measurement predictions (x_1, x_2, x_3) are sensitive to a single parameter b_1, while only one (x_4) is sensitive to another parameter b_2; in the event that measurements $x_{1\text{-}3}$ are ranked higher than x_4, the resulting design would estimate b_1 accurately and b_2 relatively poorly.

4.5.2 The Max-Min Method

The max-min design algorithm tackles the problems introduced by the max-sum method by attempting to estimate all parameters accurately. The selection process considers the effects of the measurement locations already ranked at each stage by determining how well each parameter is estimated relative to the other parameters by the predictors in the already ranked set $\mathbf{R_k}$. The location to be ranked at each iteration $\mathbf{k}$ is selected based on maximum sensitivity to this least-well estimated parameter.

$$S_j^k = \sum_{iER_k} N_{ij} \qquad j = 1,\ldots,n \tag{4.5.3}$$

where $\mathbf{S^k_j}$ is the sum of sensitivities indicating how well parameter j is estimated relative to other parameters by the already selected sensor locations. Perhaps one disadvantage of this approach is the complete focus on improving the confidence in a single poorly estimated parameter at each stage, essentially ignoring any opportunity for potentially large confidence gains on other parameters that are already relatively well estimated by $\mathbf{R_k}$.

4.5.3 The Weighted Sum Method

Much like the max-min design, this method ranks measurements by their contribution to individual parameters. However, the selection criterion is further extended to incorporate the effect of selections on all of the model parameters, weighted by their potential to provide information on parameters that are not well estimated by the current design in $\mathbf{R_k}$. This mitigates the solid focus on individual, poorly estimated parameters whilst still encouraging new information regarding their value.

$$W_i^k = \sum_{j=1}^{n} \frac{N_{ij}}{S_j^k} \quad (4.5.4)$$

Each node **I** in $\mathbf{U_k}$ is ranked at each iteration according to their weighting $\mathbf{W_i^k}$ defined by the sensitivity of the parameter **j** to that measurement, divided by the aggregate sensitivity defined in 4.5.3 above.

4.5.4 Shortest Path Algorithms

In addition to the iterative ranking procedures listed above, sampling design literature presents two design methodologies (de Schaetzen et al. 2000) intended to rank pressure monitoring points in water distribution networks based on shortest path algorithms (as nodal pressures further from sources are generally more sensitive to model parameters; Walski, 1983). Naturally, this method is conceptually limited to measurement types which demonstrate greater sensitivity as the distance to the source(s) increase. The final design is determined by an iterative ranking procedure whereby the most desirable nodes are selected sequentially.

At the outset of both procedures, the distance from the source to every other node in the network is calculated and the node furthest from the source is chosen and ranked according to the current iteration step. This, however, is where the two methods diverge; in the first approach, once a node has been selected, a dummy pipe of zero length is added to the network between the source and the node, effectively removing it from the viable choices in subsequent iterations. In the second approach however, all of the pipes from the source to the chosen node are set to zero length. The effect of the first algorithm is (in general) to emphasise the nodes located toward the interior of the network, whilst by contrast the second method should prefer nodes located further out. In the event that a network has multiple sources, a single “super source” is created and linked to each of the inflows via zero-length “dummy” pipes. This allows the algorithm to determine a “distance from source” with all sources included (but perhaps erroneously equally weighted).

4.6 Implicit Optimisation Procedures

In contrast to ‘greedy’, sequential ranking procedures, the alphabetic optimality criteria may be used to formulate objective functions for implicit optimisation procedures, perhaps in conjunction with other objectives, and indeed this has been readily pursued in modern sensor placement literature (Kapelan et al 2003, 2005; de Shaetzen et al. 2000). The vast majority of authors have implemented genetic or evolutionary algorithms (Michalewicz, 1999) to this end, tackling both single objective and multi-objective optimisations.

Within a genetic algorithm (GA), an initial population of ‘parent’ solutions is first generated. This is done either at random or, sometimes, preferably by incorporating previous knowledge on likely regions for optimal solutions in the search space e.g. information obtained via data mining techniques (Huang et al 2006). A more carefully considered initial population helps to prevent the algorithm from converging on sub-optimal local minima or maxima. In sampling design literature, solutions are vectors of node identities specifying the location of some number of sensing devices or measurements.

GAs subsequently apply a range of operators to the current (‘parent’) population to produce a set of offspring solutions that are likely to be superior to those in the previous generation. Typically this is done by assessing the suitability of solutions in the parent population and assigning them a fitness score; the “fitness” of a solution with respect to model calibration is commonly determined by the corresponding A, D and/or V optimality score, perhaps in conjunction with other objectives such as the sampling cost (Kapelan 2003). The best solutions are subsequently combined to produce new solutions for the next generation. Algorithms usually mimic nature by combining two parents into one or two offspring; however the combination of numerous parent solutions is entirely plausible.

The NSGA-II (Non-dominated Sorting Genetic Algorithm II), (Deb et al. 2001), is prevalent in sensor placement literature aimed at tackling the multiple objectives generally defined for practical optimisation scenarios (Kapelan et al., 2003; Weickgennant et al., 2009; Preis et al., 2009). The NSGA-II algorithm is similar in operation to the single-objective GA procedure, however it contains additional steps intended to progress solutions according to their worth in a multi-objective scenario (Deb et al., 2001).

For multiobjective algorithms in general, as also specifically for NSGA-II, a number of objectives are considered simultaneously via a modified fitness assessment based on a solution's location in n-dimensional Pareto space. Solutions are ordered according to the number of other solutions by which they are dominated. Subsequent sets of solutions move ("evolve") probabilistically towards the global Pareto front. Conceptually, the optimal evolution of a Pareto set involves in itself two potentially conflicting objectives; the distance to the optimal front is to be minimised whilst the diversity of the solutions in the population is to be maximised.

In addition to the NSGA-II, the Strength Pareto Evolutionary Algoritm (SPEA) algorithm has been implemented (Cheung et al 2005) to tackle objective functions based on the FOSM in sampling design literature with some success. The SPEA algorithm (Zitzler and Thiele, 1999) stores both a "population" and an "archive", with the latter consisting of the non-dominated front amongst all solutions considered so far. This procedure has performed favourably against major similar algorithms, such as the NSGA-II in a number of trials (Zitler et al., 2000).

A highly relevant multi-objective problem incorporating calibration accuracy and sampling cost was successfully solved via the application of a standard MOGA-type procedure intended to retrieve the desired Pareto front (Kapelan, 2005). A discussion of the inclusion of a cost objective alongside optimal model calibration is provided in section 4.8.

4.7 Stochastic Sampling Design

In the process of selecting measurements for estimating parameter values, the actual parameter values that are to be estimated must be known. The previously identified deterministic approaches to optimal sampling design require a fixed estimate for parameter values prior to any optimisation runs. Naturally, suitable initial estimates close to the optimal values are difficult to ascertain prior to the calibration process and indeed the problem may appear somewhat circular in nature. The most common solution is to apply an iterative design procedure whereby parameter estimates gleaned from a proceeding design are used as initial values to inform the subsequent design (Bush and Uber, 1998).

To overcome this problem, a stochastic sampling design problem approach is suggested (Behzadian et al., 2009), whereby calibration parameter values are assumed a priori uncertain and modelled using probability density functions. In this way, a more realistic representation of the (essentially unknown) calibration parameter values may be provided in aid of more robust sampling solutions. Understandably the stochastic sampling design problem is more difficult to solve than the deterministic one and places greater strain on the available computational capacity. The use of meta-models to reduce the computational times in optimisation processes may be implemented to improve the feasibility of stochastic studies (Blanning, 1975).

A meta-model is essentially a surrogate process model used in the calculation of design fitness values in place of more time-consuming (and 'realistic') simulations. Such a meta-model is gradually integrated into the optimisation process to substitute the increasing proportions of the simulation stage. One frequently used meta-model is the artificial neural network (ANN), favoured for its ability to approximate effectively a wide range of non-linear functions (Leshno et al., 1993). In a recent application of groundwater remediation design, an adaptive neural network combined with a single objective genetic algorithm negated the need for around ninety percent of the simulation procedures with no loss in accuracy in the optimal solutions (May et al., 2008).

4.8 Consideration of Costs

Macro-sensor Location problems reveal a tendency for diminishing returns accompanying iterative increases in the number of sensing devices/locations on a network (Kapelan et al., 2003, Krause et al., 2006, 2009). Consequently the collection of excessive amounts of data is not cost-effective, and it is paramount that practitioners are able to discern an appropriate 'price point' for the quantity of optimised measurements. Future Sensor placement tools to be developed in linked work packages within PREPARED aim to accurately assess cost and calibration trade-offs.

The calibration accuracy achieved for a given model and the cost of performing the required measurements at on a water network are both dependent on the quantity of measurements taken. Model calibration accuracy is a monotonically increasing function of the number of sensors/measurement locations and hence conflicts directly with an objective for minimum cost (Farmani et al., 2005). A number of authors have addressed cost with respect to optimal sampling design Meier and Barkdoll (2000), Vitkovsky et al. (2003), Kapelan et al. (2003). The quantity of sensor locations is generally considered to be an acceptable surrogate for sampling cost within the context of macro-location problems.

4.8.1 Multi-objective Optimisation

Design problems with a single objective criterion are analysed in order to identify the best solution, as the one which most closely meets the specified criterion. The process often entails minimising or maximising an objective function under a given set of constraints. More complex problems require a designer to work toward a variety of disparate objectives simultaneously -- many of which may be directly conflicting (Deb et al., 2005). In these cases it should be understood there is no single "best" solution. Further information regarding the relative importance of each objective must be specified –perhaps subjectively- by the designer before a final decision may be made (Kapelan et al., 2005). In the case of PREPARED objectives in this work area, calibration accuracy (resulting from optimal placement of N sensors) is to be assessed in conjunction with sampling cost.

Logically, quantitative optimisation for a given set of contradictory or independent objective functions may proceed by obtaining the set of solutions representing the best ("optimum") compromises involving the disparate design criteria.

Box 5: Solution assessment for multi-objective optimisation problems

A solution X_i to any multi-objective optimisation problem is said to "*dominate*" another solution X_j if it is superior with respect to any **one** of the optimisation objectives without being inferior with respect to **any** of the others. Put another way, solution X_i dominates solution X_j if it is at least as good in all objective cases and better in at least one case. A solution is "*non-dominated*" if it is not dominated by any other solution in the entire *search space*.

Globally non-dominated solutions are referred to henceforth as "Pareto Optimal" Solutions. Given the definition in Box 5, a solution is Pareto optimal if no other solution exists that is superior with respect to any one of the objectives without being inferior with respect to any of the others. Multi-objective optimisation algorithms in literature frequently rely on these criteria (Deb et al., 2002; Fieldsend et al., 2002; Kuntinee et al., 2004; Kapelan et al., 2003; Kapelan et al., 2005). Elements in the set of Pareto optimal solutions are said to be on the "Pareto Front" and represent the set of best trade-off designs from which a designer should make a final selection.

4.8.2 Assessing the Pareto Front

Multi-objective decision problems may be tackled directly via the analytical combination of the several objective functions into a single representative function, thereby reducing the problem to a single-objective optimisation (Deb et al., 2002). There are however several disadvantages to this method; most significantly that the relative importance or "weighting" of each objective must be specified prior to the optimisation procedure (Kapelan et al., 2005). Carefully constructed algorithms may still be able to recover the true Pareto front by varying objective weights however numerous optimisation runs would certainly be required (Cheung and Pillar, 2006).

Multi-objective optimisation methods intended to recover the Pareto Front directly have historically been based on the method of Lagrange multipliers (Carmichael et al., 1980). The majority of modern techniques centre on the concept of “Evolutionary” algorithms which work in parallel by sequential modification of an entire population of solutions tending towards the Pareto Front (provided that the solution fitness is correctly identified in a multi-objective setting). These methods have been popularised by the advent of extensive and available computing power (Kapelan et al., 2002; Gueli et al., 2006, Preis and Ostfeld, 2006; Wu and Walski, 2006; Austin et al., 2009; Preis et al., 2009; Weickgennant et al., 2010).

4.9 Chapter Summary

Optimal macro-location (and more generally ‘sampling design’) has been on the forefront of WDS research agendas since the 1980s with some initial focus on sensor placement in the face of model calibration concerns. Algorithms and procedures developed to solve these combinatorial optimisation problems may be broadly categorised into ‘greedy’ ranking procedures favoured by early research, and the more modern construction of implicit optimisation problems via the direct application of A, D and V optimality criteria. Once specified, the implicit optimisation problems derived from measures of calibration accuracy have been unanimously attacked with various single and multi- objective genetic algorithm procedures, such as the NSGA-II (Deb et al., 2001).

5 Sampling Design for Other Purposes

5.1 Introduction

Macro-location of sensors in the context of urban water systems is pursued in aid of numerous distinct agendas affecting the definition of an optimal sampling design. Optimal sampling design for *model calibration* remains the focus of interest within this review (PREPARED, Task 3.6.1), however, sampling design literature centring on dissimilar optimisation objectives (i.e. sensor placement for best contamination detection/burst detection and tracing, etc), remains a consideration for effective and sustainable management of urban water networks (Savic et al., 2009). Additionally, elements of research conducted for sampling design are conceptually applicable to processes aimed specifically at optimal model calibration. Most notably, the optimisation approaches and algorithms developed for (or merely implemented in) sampling design literature are distinct from the objectives to which they are applied.

A significant proportion of recent literature on the subject of sensor placement in urban waste water systems has been published in response to the Battle of the Water Sensor Networks (BWSN) challenge (Ostfeld et al. 2008, Ghimire and Barkdoll 2006, Guan et al. 2006, Gueli et al. 2006, Huang et al. 2006, Ostfeld and Salomons 2006, Ostfeld et al. 2006, Preis and Ostfeld 2006, Dorini et al. 2006, Eliades and Polycarpou 2006, Propato and Pillar 2006, Lansey et al. 2006, Walski et al. 2006, Watson et al. 2006, Wu and Walski 2006, Trachtman 2006, Krause et al. 2006). Accordingly, the following section details relevant algorithms for the optimal macro-location of sensors on urban water networks derived from this study.

5.1.1 Sampling Design for WDS contaminant detection

The BWSN exercise was a recent push in the field of sampling design, focusing on an effective response against intentional or accidental contamination of a Water Distribution Network. In the mock scenario, four pre-specified contamination events were assumed to occur randomly within the first 24 hours of a 96-hour simulation. The objective was to generate a sampling design that would best protect the downstream population. The majority of solutions were formulated as optimisation problems with four driving objectives:

1. *Maximise the detection likelihood*. This is simply as defined as the ratio of detected contamination events against the total number of events.

   Max {Detection Likelihood}

2. *Minimise the expected time to detection*. For each pollution event, the detection time is the time taken for any one of the sensors on the network to register a non-zero contaminant concentration. The expectation of this function across all possible contamination events provides a measure of its general performance.

   Min {Expected time of detection}

3. *Minimise the effect on the population*. For a specific contamination scenario, the population affected is a function of the ingested contaminant mass. The ingested contaminant mass, in turn, depends on the time of detection for the sensor network, as described above; two key assumptions are that no mass is ingested after detection and that all mass ingested during undetected events is not counted.

   Min {Expected Population effected prior to detection}

4. *Minimise the consumption/demand of total volumetric water demand* which exceeds a predefined concentration of contaminants.

   Min{Expected Demand of contaminated water prior to detection}

A possible weakness in the objective function formulation of the BWSN is identified by several authors (Eliades and Polycarpou, 2006); the undetected contaminations (specified probabilistically by 1) are not included in the calculation of the performance measures given by 2-4 hence the level of consumption at the undetected nodes is not properly taken into account. In order to combat this problem some authors introduced additional objective functions. Further information may be found in "The Battle of the Water Sensor Networks (BWSN): A Design Challenge for Engineers and Algorithms" (Ostfeld et al 2006).

5.2 Optimisations Approaches in Contaminant Sampling Design

The objective functions defined in the BWSN and related works are distinct from most measures of performance for model calibration, however, some of the computational optimisation procedures are conceptually relevant and potentially applicable to objective functions based on A, D and V optimality. Optimisation procedures applied in these studies have been adapted (or not) for sampling design or similar combinatorial design problems and are reviewed briefly in the follow sections.

5.2.1 Iterative Deepening Optimisation

A novel incremental Macro-location procedure was developed in response to the BWSN exercise (Eliades and Polycarpou, 2006). The process begins by considering the simple single sensor problem with respect to some objectives, which may trivially solved with the corresponding Pareto front correctly and completely identified. Once the sensor is placed on one of the Pareto optimal locations, the placement of a second sensor is considered (whilst leaving the first in its chosen location). The resulting design, now containing **n**=2 sensors, is not necessarily Pareto optimal for the general two-sensor problem. This is because the suggested “local” Pareto frontier is somewhat defined by the placement of the first sensor. Put another way, the local Pareto frontier generated for the **n**=2 problem depends on the choice of location for the first sensor.

It is suggested that the basic approach can be improved by searching the Pareto frontier solutions in parallel as the algorithm expands from **n**=1 to **n**=**N** sensors. After obtaining the Pareto frontier for the **n**=1 sensor problem, all of the solutions on the frontier are stored in a list and individually combined with a second sensor to form a set of frontiers for the **n**=2 problem. Intuitively, all of the solutions on these frontiers are each combined with a third sensor and the process continues. It may be prudent to selectively sort or remove some of the solutions on the various frontiers according to a relevant heuristic at each stage. Naturally this reduces the computation time and memory whilst biasing the results toward some presumably favourable objective.

5.2.2 'SLOTS' Algorithm

'SLOTS' (Dorini et al., 2010) is a sensor placement routine developed in response to the BWSN and is intended to generate a good or optimal set of sensor placements on a given network. The algorithm begins with an arbitrary set of sensor locations N and continues as follows:

1. One sensor is moved from location to location while all of the other sensors are held fixed and performance changes are measured each time.
2. The sensor is placed in the location corresponding to the best improvement in performance.
3. Once this has been done for the first sensor, the algorithm moves onto the next sensor and the process is repeated.
4. After all of the sensors have been moved (or not), the algorithm starts again from the first sensor and follows the same sequence as the previous loop.
5. The algorithm terminates when there is no further improvement in performance from loop to loop, or a preset number of loops have been run.

SLOTS was tested on two case studies from the BWSN and performed favourably against several greedy algorithm solutions. The solutions to single-objective problems were often found to exceed 95% of the true optimum.

5.2.3 Genetic Algorithms

Evolutionary algorithms are at the forefront of research regarding optimisation procedures for all manner of sampling design problems and efficient contaminant detection is no exception. Well-known and generic optimisation procedures based on genetic algorithms are inherently applicable to combinatorial problems such as the macro-location of sensors and have not seen significant modifications to their operation. However, the BWSN inspired the introduction of at least one calibration-relevant modification intended to overcome 'computational inefficiencies' related to the study of large water distribution networks (Guan et al., 2006). The algorithm reduces computation time by working on a subset of the total available nodes. If **N** is a set of all the nodes in the system, the procedure begins by selecting a subset $\mathbf{N_{sub}}$ either at random or, preferably, by way of informed decision. At each subsequent generation a new sub-domain is created and will include the junctions found in the best solutions of the previous generation. The algorithm continues until all the potential sensor locations have been selected into the sub-domain at least once. This improves the chance of selecting a globally optimal design.

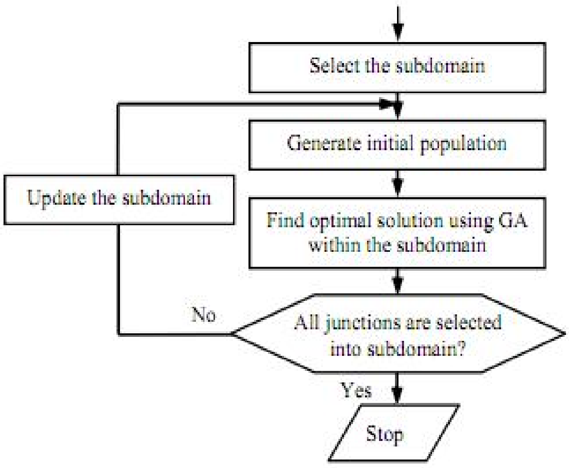


Figure 1. The ‘improved’ GA procedure. Modified extract from Guan et al. (2006)

The inclusion of an iterative sub-domain reduces the computational effort at each step but may be more prone to convergence on a local optimum than algorithms which operate on the entire population in parallel.

Research regarding genetic algorithms is an active and evolving field and in principle any one of number of generic optimisation procedures may be applicable to the sensor placement problem with objectives defined by A, D and V optimality or some other sensor placement criteria.

5.2.4 Integer Programming

Integer programming involves the solution of a set of equations and logical constraints whereby the domains of one or more variables (mixed integer programming) or all of the variables (pure integer programming) are limited to the set of integers only. If the domain is further restricted to binary digits, the equations are known as binary or (0, 1) integer programming problems and suit themselves well to a raft of optimisation tasks. Berry et al (2004), Berry et al (2005) formulate the BWSN optimisation as a (0, 1) integer programming problem. Each network node,

$\mathbf{N_i}$ is assigned a binary variable $\mathbf{X_i}$ which defines whether a sensor is placed on the node ($\mathbf{X_i}$=1) or not ($\mathbf{X_i}$=0). Whilst the objective functions considered for BWSN are generally unrelated to least-uncertainty calibration concerns, a similar (0, 1) integer programming approach may be applicable under a different set of constraints.

The solution of integer programming (IP) problems for systems with a large number of variables can be extremely computationally demanding (NP hard in general). Initial solution methodologies relied on the method of cutting planes (Ballas 1971), however, non-tailored (general) algorithms using this method can be extremely slow (Marchand et al., 2002). Another common approach aimed at reducing the necessary computation time is the "branch and bound" method. This is the approach employed in aid of a sampling design problem by Propato and Pillar (2006).

5.2.5 Cross Entropy Optimisation

The cross-entropy (CE) method is a generic Monte Carlo technique applicable to generic optimisation problems (Rubinstein and Kroese, 2004). Originally developed for rare event simulation and later adapted for optimisation (de Boer et al. 2005), the CE method is an iterative ('evolutionary') method essentially involving two steps:

1. A sample of random data is generated according to a specified random mechanism.
2. The random mechanism is augmented using the data, making it likely to produce a 'better' sample in the next iteration (a 'sample' being a parameter vector that maximises an arbitrary objective function)

The CE method has been applied with promising success to a number of troublesome combinatorial optimization problems, including the well-known maximal cut problem, the travelling salesman problem and the quadratic assignment problem (Rubinstein and Kroese, 2004). CE optimisation is based on importance sampling (Glynn et al., 1989) and approximates the optimal importance sampling Probability Density Function (PDF) by adaptively reducing the Kullback-Leibler divergence ('relative entropy') between the true optimum and current best guess (De Boer et al., 2005). Perhaps one disadvantage is the requirement to preselect a parametric family of PDFs within which the algorithm can search; a poor choice would logically limit the effectiveness of the procedural run.

A modified Cross Entropy-type procedure was developed for the BWSN (Dorini et al., 2006) for application in both single objective and multi-objective optimisation problems. Key departures from the standard CE procedure defined as follows:

- The updating mechanism is perturbed with an additive-asymptotically-zero noisy component.
- The process converges to non-degenerated probability matrices Solutions in each iteration are sorted with respect to the Pareto-Ranking procedure

5.3 Chapter Summary

Sampling design literature has moved decisively in recent years toward sensor placement with the goal of efficient and effective detection of contaminants in water distribution systems. The majority of research in this area has focused on the development of new and improved implicit objective functions (and combinations of functions) and typically relies on generic tried-and-tested optimisation procedures such as linear programming and genetic algorithms (source) with minor modification in some instances.

6 Conclusions

This report has reviewed relevant literature regarding optimal macro-location of sensors for Urban Water Systems (UWS) satisfying the requirements of Deliverable D3.5.1 within Work Package WP3.5 of the FP7-EC Large Scale Integrating Project PREPARED. It has focused on optimal sampling design for urban water *model calibration,* but additionally incorporates closely related optimisation procedures and processes found amongst macro-location literature for contaminant detection.

This report indicates that optimal sampling design for model calibration contains a mixture of tailored heuristic algorithms and general optimisation procedures (e.g. genetic algorithms) applied most often to objective functions derived from A, D and V (alphabetic) optimality criteria, or related sensitivity criteria. These criteria have regularly and successfully been combined with others in a multi-objective setting e.g. minimising calibration accuracy in addition to sampling cost (Kapelan et al., 2003) or (maximising) normalised sensitivity entropy (de Schaetzen et al., 2000). Interestingly, despite numerous methods available to discern parameter uncertainty in calibration procedures i.e. SCEM-UA (Kapelan et al., 2007, Cutore et al., 2008), SCE-UA (Duan et al., 1992, 1994, Vrugt et al., 2003) and other Monte-Carlo, Bayesian-type processes (covered explicitly in the Deliverable D3.6.1) (PREPARED, 2011), uncertainty assessment based on the FOSM method appears to be the only approach taken in sampling design literature thus far; perhaps due to demanding computation requirements of methods such as SCEM-UA, and the generally sufficient accuracy of FOSM for parameter estimation (Bard, 1974).

It is apparent that the vast majority of the literature on the subject of macro-location is proposed for sensor placement in clean Water Distribution Systems (WDS); generic calibration procedures are of course equally applicable to sewer models and hence some proportion of the sampling design algorithms would perform equivalently in Urban Waste Water System (UWWS) sensor placement application. Nonetheless designers must to be aware of practical concerns affecting sensor operation in a sewer environment (Winkler et al 2005).

It should also be noted that the sensor placement problem is approached in the majority of literature from a largely theoretical standpoint, applied to a somewhat homogeneous and idealised water network. Some authors, however (Berry et al., 2006, Murry et al., 2006), indicate potentially significant factors which may be of concern for the real-world application of online monitoring systems. For example, the cost of placement of sensor devices is highly dependent on the accessibility of a particular location on a network, where, a large digging operation may be required to access a particular junction for installation or maintenance of a sensor and hence should incur an appropriate penalty with respect to optimal sampling design (in contrast to past assumptions that cost is related only to measurement quantity and is not a function of location).

Additionally, in practice, the alphabetic optimality criteria may be sub-optimal in their own right since they treat all nodes equally – in general it may be beneficial to calibrate a model with respect to some nodes more than others e.g. it may be desirable to provide better accuracy in an industrial area requiring a high level of service and accurate pressures compared to other customers. Regardless, the methods presented in this paper provide a rigorous grounding framework for optimal design provided that theory is applied competently and in a suitable context.

6.1 Planned future work for PREPARED

Within WP3.5, some selected methodologies will be implemented for the development of a multi-objective optimisation tool for sensor placement intended to assess the Pareto front for trade-offs between sensor cost and the resulting model accuracy (Task 3.5.2). Appropriate methodologies will be demonstrated at one of the PREPARED cities (WP 4.4, Task 4.4.2). Outcomes from this WP (WP3.5) are closely linked with WA4 tasks: The optimal sensor placement prototype (Task 3.5.2) will be part of the OpenMI tool box (WP4.2) and will be used for the early warning in distribution systems (WP4.4 - Task 4.4.2).

The demonstration city/cities, within PREPARED, for macro-location sensor placement have not yet been defined. Thus, the development of the model will initially be based on generic principles, and tested/validated on benchmark networks from the related literature.

6.2 Appendices

Appendix A summarises optimal sensor placement algorithms covered in this review. Starting with the work by Bush and Uber (1998), these techniques are an assortment of computational optimisation procedures which intend to seek design solutions from a finite set of potential sensor locations. Generally they seek to minimise objective functions defined almost exclusively on A, D and V criteria (dependent on the covariance matrices for model parameters and predictions). Some of the procedures are multi-objective by design and hence may naturally incorporate a cost objective along with the appropriate calibration criteria. Other procedures would require a combined objective function, and potentially multiple algorithm runs to uncover portions of the Pareto front.

7 References

Ahmed KH, Hasan M.K., Burgess W.G., Dottridge J., Ravenscroft P., van Wonderen J.J. (1999), "The Dupi Tila aquifer of Dhaka, Bangladesh: hydraulic and hydrochemical response to intensive exploitation.". *Chilton PJ, Groundwater in the urban environment: selected city profiles.* ISBN: 9054108371, p19–30.

Ahyerrea, M., Chebboa, G., Tassina, B., Gaumea, E. (1998), "Storm water quality modelling, an ambitious objective?". *Water Science and Technology.* **37**(1), 205-213

Alvisi, S. and Franchini, M. (2010), "Calibration and Sensitivity Analysis Of The C-Town Pipe Network Model". *Water Distribution System Analysis 2010 (WDSA2010).* Tucson, AZ, USA, Sept. 12-15, 2010.

Asadzadeh, M., Tolson, B. McKillop, R. (2010), "A Two Stage Optimization Approach For Calibrating Water Distribution Systems." *Water Distribution System Analysis 2010 (WDSA2010).* Tucson, AZ, USA, Sept. 12-15, 2010.

Ashleya, R.M., Verbanckb, M.A. (1996), "Mechanics of sewer sediment erosion and transport". *Journal of Hydraulic Research.* **34**(6), 753 - 770

Austin, R.G, Choi, C.Y, Preis, A. Ostfeld, A. and Lansey, K. (2009), "Multi-objective Sensor Placements with Improved Water Quality Models in a Network with Multiple Junctions." *World Environmental and Water Resources Congress 2009,* USA.

AWWA, American Water Works Association (1999). *American Water Works Association,* Denver, USA. ANSI/AWWA Standard B408-03.

Ballas, E. (1971), "Intersection Cuts-A New Type of Cutting Planes for Integer Programming". *Operations Research.* **19**(1), 19-39.

Bard, Y. (1974), "Nonlinear parameter estimation". *Academic Press, New York.* ISBN: 978-01-2078-250-5

Bates, B., Kundzewicz, Z. W., Wu, S., Palutikof, J. (2008). "Climate change and water". *IPCC Technical Paper VI.* ISBN: 978-92-9169-123-4

Behzadian, K., Kapelan, Z., Savic, D.A. and Ardeshir A., (2009), "Stochastic Sampling Design using Multiobjective Genetic Algorithm and Adaptive Neural Networks", *Environmental Modelling and Software*, **24**(4), 530–541.

Berry, W. Jonathan, Carr, Robert D., Hart, William E., Leung, Vitus J., Phillips, Cindy A., Watson, Jean-Paul (2006) "On the placement of imperfect sensors in Municipal Water Networks". *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Berry, W. Jonathan, Fleischery, Lisa, Hart, William E., and Phillips, Cynthia (2004). "Sensor Placement in Municipal Water Networks*". J. Water Resour. Plng. and Mgmt.* **131**(1), 237-244.

Berry, W. Jonathan, Hart, William E., Phillips, Cynthia A., Watson, Jean-Paul (2005). "Scalability of Integer Programming Computations for Sensor Placement in Water Networks". *Proceedings of World Water and Environmental Resources Congress 2005,* USA.

Berry, Jonathan W., Hart, William E. , Phillips, Cynthia A., Uber , James G., and Watson, Jean-Paul (2005). "Validation and Assessment of Integer Programming Sensor Placement Models". *Proceedings of the 2005 World Water and Environmental Resources Congress,* USA.

Berry, J. W., Hart, W. E., Phillips, C. A., and Watson, J. P. (2006)."A facility location approach to sensor placement optimization." *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Beven, K., and Binley, A. (1992). "The Future of Distributed Models – Model Calibration and Uncertainty Prediction." *Hydrological Processes.* **6**(3), 279-298.

Beychok, M.R. (1971). "Performance of surface-aerated basins". *Chemical Engineering Progress Symposium.* **67**(107), 322–339.

Bhave, P. R. (1988). "Calibrating Water Distribution Network Models." *Journal of Environmental Engineering.* ASCE, **114**(1), 120-136

Blanning, R.W., (1975). "The construction and implementation of metamodels". *Simulation*, **24** (6), 177–184.

Boccelli, D. L., Uber, J. G. (2004). "A markov chain monte carlo approach to network model calibration". *Water Quality Technology Conference and Exposition. American Water Works Association,* San Antonio, Texas, USA.

Boccelli, D. L., Uber, J. G., (2005). "Incorporating spatial correlation in a markov chain monte carlo approach for network model calibration". *Walton, R. (Ed.), World Water Congress.* **173**, 28–28.

Bourgeois, W., Burgess, J.E. and Stuetz, R.M. (2001). “On-line monitoring of wastewater quality: a review.” *Journal of Chemical Technology & Biotechnology.* **76**, 337-348.

Bouteligier, R., Vaes, G., Berlamont, J. (2002), “Global Solutions for Urban Drainage”. *Proceedings of 9th International Conference on Urban Drainage.*

Bros, Christopher M., and Kalungi, Paul (2010), “Battle of the Water Calibration Networks (BWCN)” *Water Distribution System Analysis 2010 (WDSA2010),* Tucson, AZ, USA, Sept. 12-15, 2010.

Burd, R, Zazula-Coetzee, B, Belrain, T. (2010), “Battle of the Water Calibration Networks (BWCN)” *Water Distribution System Analysis 2010 (WDSA2010),* Tucson, AZ, USA, Sept. 12-15, 2010.

Bush, C. A., and Uber, J. G. (1998). "Sampling design methods for water distribution model calibration." *Journal of Water Resources Planning and Management, ASCE.* **124**(6), 334-344.

Butler, D. (1993). "The influence of dwelling occupancy and day of the week on domestic appliance wastewater discharges". *Building and Environment*. **28**(1), 73-79.

Chang, K. Gao, J. Qu, S., Yuan, Y. (2010). “Research on Water Distribution Network Calibration Using Heuristics Algorithm ond Pipeline Resistance Coefficient Empirical Model”. *Water Distribution System Analysis 2010 (WDSA2010),* Tucson, AZ, USA, Sept. 12-15, 2010.

Cheung, P.B, Pillar, O. and Propato, M. (2009) “Optimal location of Water Quality Sensors in supply systems by multi-objective Genetic Algorithms.” *Hydraulics and Civil Engineering Research Unit*, Cemagref, Bordeaux, France.

Chun, D. G., Selznick, H. L. (1985). “Computer modeling of distribution system water quality”. *Torno, H. C., Computer Applications in Water Resources. ASCE.* **1**, 448–456.

Clark, R. L, D.E. Radcliffe, G. W. Langdale, R.R. Bruce. (1993). “Soil strength and water infiltration as affected by paratillage frequency”. *Transactions of the ASAE.* **36**(5), 1301-1305.

Corne, D.W., Knowles, J.D. (2000). “The Pareto-envelope based selection algorithm for multiobjective optimization”**.** *Proceedings of the Sixth*

International Conference on Parallel Problem Solving from Nature. ***1,*** 839-848.

Crabtree, R.W. (1989) “Sediment in sewers”. *Water and Environment Journal.* **3**(6), 569–578.

Cutore, P., Campisano, A., Kapelan, Z., Modica, C., and Savic, D. (2008). "Probabilistic prediction of urban water consumption using the SCEM-UA algorithm." *Urban Water Journal*. **5**(2), 125-132.

Datta, R. S. N., and Sridharan, K. (1994). "Parameter-Estimation in Water-Distribution Systems by Least-Squares." *Journal of Water Resources Planning and Management, ASCE.* **120**(4), 405-422.

De Boer, Pieter-Tjerk, Kroese, Dirk P., Mannor, Shie., Rubinstein, Reuven Y. (2006) “A Tutorial on the Cross-Entropy Method”. *Annals of Operations Research.* **134**, 19–67.

De Clercq, B., Coen, F., Vanderhaegen, B., and Vanrolleghem, P. A. (1999). "Calibrating simple models for mixing and flow propagation in waste water treatment plants." *Water Science and Technology.* **39**(4), 61-69.

De Schaetzen, W. B. F., Walters, G. A., and Savic, D. A. (2000). "Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms." *Urban Water*. **2**, 141-152.

Deb, K., S. Agrawal, A. Pratap, and T. Meyarivan (2000). “A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II”. *Parallel Problem Solving from Nature.* PPSN **VI**, 849–858.

Deb, K., Thiele, L., Laumanns, M., Zitler, E. (2005), “Scalable Test Problems for Evolutionary Multiobjective Optimization”. *Advanced Information and Knowledge Processing.* ***1****, 105-145.*

Dempsey, P., Eadon, A. and Morris, G., (1997). “SIMPOL: a simplified urban pollution modelling tool”. *Water Science Technology.* **36**(8-9), 83–88.

Diao, K., Zhou, Y., Li, J., Liu, Z. (2010), “Battle of the Water Calibration Networks (BWCN): A Component Status Changes Oriented Calibration Method For Zonal Management Water Distribution Networks”. *Water Distribution System Analysis 2010 – WDSA2010, Tucson, AZ, USA, Sept. 12-15, 2010*

Dorini, G., Jonkergouw, P., Kapelan, Z., di Pierro, F., Khu, S. T., and Savic, D. (2006). "An efficient algorithm for sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Dorini, Gianluca, Jonkergouw, Philip, Kapelan, Zoran and Savic, Dragan (2010) "SLOTS: Effective Algorithm for Sensor Placement in Water Distribution Systems". *J. Water Resour. Plng. and Mgmt.,* **136**, 620-629.

Di Pierro, F., Djordjevic, S., Kapelan, Z., Khu, S.-T., Savic, D., and Walters, G. A. (2005). "Automatic calibration of urban drainage model using a novel multi-objective genetic algorithm", *Water Science and Technology*, **52**(5), 43-52.

Duan, Q. Y., Sorooshian, S., and Gupta, V. (1992). "Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models." *Water Resources Research*, **28**(4), 1015-1031.

Duan, Sarooshian et al (1994) "Optimal use of the SCE-UA global optimisation method for calibrating watershed models". *Environmental Modelling & Software,* **12**(1), 27-34.

Eliades, D., and Polycarpou, M. (2006). "Iterative deepening of Pareto solutions in water sensor Networks." *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Falconer, I.R, Choice, A., Hosja, W. (1992). "Toxicity of edible mussels (Mytilus edulis) growing naturally in an estuary during a water bloom of the blue-green alga Nodularia spumigena." *Environmental Toxicology and Water Quality*, **7**(2), 119–123.

Farmani, R., Walters, G. A., and Savic, D. A. (2005). "Trade-off between total cost and reliability for Anytown water distribution network." *Journal of Water Resources Planning and Management*, *ASCE*, **131**(3), 161-171.

Felix S. Wong (1984) "First Order, Second Moment Methods". *Computers & Structures,* **20**(4), 779-791.

Ferreri, G.B., Napoli, E., Tumbiolo, A., (1994). "Calibration of roughness in water distribution networks". *Proceeding of second International Conference on Water Pipeline Systems, D. S. Miller*, Suffolk, U.K. **1**, 379-396.

Filion, Y. R. and Karney, B. W. (2003) “Sources of error in network modeling: question of perspective”. *Journal of American Water Works Association,* **95**(2), 119-130.

Ghimire, S. R., and Barkdoll, B. D. (2006a). “Heuristic method for the battle of the water network sensors: Demand-based approach.” *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Ghimire, S. R., and Barkdoll, B. D. (2006b). “A heuristic method for water quality sensor location in a municipal water distribution system: Mass related based approach.” *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Grayman, W. M., Clark, R. M., Males, R. M., (1988). “Modeling distribution-system water quality: Dynamic approach”. *Journal of Water Resources Planning and Management,* **114** (3), 295–312.

Greco, M., Del Giudice, G., (1999). “New approach to water distribution network calibration”. *Journal of Hydraulic Engineering,* **125** (8), 849–854.

Gronewold, A.D., Qian, S.S., Wolpert, R.L. and Reckhow, K.H., (2009). “Calibrating and validating bacterial water quality models: A Bayesian approach.” *Water Research,* **43**, 2688-2698.

Guan, J., Aral, M. M., Maslia, M. L., and Grayman, W. M. (2006). “Optimization model and algorithms for design of water sensor placement in water distribution systems.” *Proc. 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Gueli, R. (2006). “Predator–prey model for discrete sensor placement.” *Proc., 8th Annual Water Distribution System Analysis Symp.*, Cincinnati, USA, 2006.

Hallam, N. B., West, J. R., Forster, C. F., Powell, J. C., Spencer, I. (2002). “The decay of chlorine associated with the pipe wall in water distribution systems”. *Water Research,* **36**, 3479–3488.

Huang, J. J., McBean, E. A., and James, W. (2006). “Multiobjective optimization for monitoring sensor placement in water distribution systems.” *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Jack, A. G., Petrie, M. M., and Ashley, R. M. (1996). "The diversity of sewer sediments and the consequences for sewer flow quality modelling." *Water Science and Technology,* **33**(9), 207-214.

Jamieson, D. G., Shamir, U., Martinez, F., and Franchini, M. (2007). "Conceptual design of a generic, real-time, near-optimal control system for water-distribution networks." *Journal of Hydroinformatics*, **9**(1), 3-14.

Johnson, J.P., Ghimire, S.R., Barkdoll, B.D. (2010), "Flow-Sequential Sector-Specific Lumped Algorithm for Water Network Calibration". *Water Distribution System Analysis 2010 (WDSA2010).* Tucson, AZ, USA, Sept. 12-15, 2010.

Jonkergouw, P., Khu, S-T., Kapelan, Z. and Savic, D.A. (2008), "Water Quality Model Calibration under Unknown Demands", *ASCE Journal of Water Resources Planning and Management*, **134**(4), 326-336.

Kang, Doosan, and Lansey, Kevin (2010), "The Battle of The Water Calibration Networks (BWCN): Roughness And Demand Estimation Based On Weighted Least Squares (WLS) Method". *Water Distribution System Analysis 2010 (WDSA2010).* Tucson, AZ, USA, Sept. 12-15, 2010.

Kapelan, Z., D.A. Savic, and G.A. Walters (2003), "A Hybrid Inverse Transient Model for Leakage Detection and Roughness Calibration in Pipe Networks", *IAHR Journal of Hydraulic Research*, **41**(5), 481-492.

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, ASCE*. **129**(6), 466-479.

Kapelan, Z., Savic, D.A. and Walters, G.A. (2004). "Incorporation of Prior Information on Parameters in Inverse Transient Analysis for Leak Detection and Roughness Calibration" *Urban Water Journal.* **1**(2), 129-143.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2005). "Optimal sampling design methodologies for water distribution model calibration." *Journal of Hydraulic Engineering*, ASCE. **131**(3), 190-200.

Kapelan, Z. S., Savic, D. A., and Walters, G. A. (2007). "Calibration of water distribution hydraulic models using a Bayesian-Type procedure." *Journal of Hydraulic Engineering, ASCE*. **133**(8), 927-936.

Kapelan, Z. (2010), "Calibration of Water Distribution System Hydraulic Models", *LAP Lambert Academic Publishing*, ISBN 978-3-8383-5341-8, p. 284

Kennedy, M. S., Sarikelle, S., Suravallop, K., (1991). Travel time calibration in water distribution systems. *Water quality modeling in distribution systems*, AWWARF and US EPA, Cincinnati. **1**, 219–233.

Kiefer, J. (1958). "On the Nonrandomized Optimality and Randomized Nonoptimality of Symmetrical Designs". *Annals of Mathematical Statistics.* **29**, 675-699.

Kiefer, J. and Wolfowitz, J. (1959). "Optimum Designs in Regression Problems". *Annals of Mathematical Statistics.* **30**, 271-294.

Kim, J. H., Chung, G., and Yoo, D. G. (2010). "Calibration of C-Town Network using Harmony Search Algorithm." *Water Distribution System Analysis 2010 (WDSA2010).* Tucson, AZ, USA, Sept. 12-15, 2010

Koppel, T., and Vassiljev, A. (2010), "Calibration of Water Distribution Network for BWCN". *Water Distribution System Analysis 2010 (WDSA2010).* Tucson, AZ, USA, Sept. 12-15, 2010

Korving, H., van Noortwijk, J. M., van Gelder, P. H. A. J. M., and Parkhi, R. S. (2003). "Coping with uncertainty in sewer system rehabilitation." *Safety and Reliability,* Vol. **1** and **2**, 959-967.

Krause, A. and Guestrin, C. (2009), "Robust Sensor Placement for Detecting Adversarial Contaminations in Water Distribution Systems". *World Environmental and Water Resources Congress 2009: Great Rivers Proceedings of World Environmental and Water Resources Congress 2009*, USA.

Krause, A., Leskovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., and Fischbeck, P., (2006). "Optimizing sensor placements in water distribution systems using sub modular function maximization." *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA.

Kunstmann, H., Kinzelbach, W., Siegfried, T. (2002). "Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling". *Water Resources Research*, **38**(4), p1035

Lansey, K. E., El-Shorbagy, W., Ahmed, I., Araujo, J., Haan, C. T. (2001). “Calibration assessment and data collection for water distribution networks”. *Journal of Hydraulic Engineering.* **127** (4), 270–279.

Lansey, K. E., and Basnet, C. (1991). "Parameter-Estimation for Water Distribution Networks." *Journal of Water Resources Planning and Management, ASCE*, **117**(1), 126-144.

Laucelli, D., Vamakeridou-Lyroudia, L. S., Kapelan, Z., Savic, D., Berardi, L., Barbaro, G., and Giustolisi, O. (2010). "Calibration of Water Distribution System Using Topological Analysis." *Water Distribution System Analysis 2010 (WDSA2010),* Tucson, AZ, USA.

Lee, B. H., Deininger, R. A. (1992). “Optimal locations of monitoring stations in water distribution systems”. *Journal of environmental engineering,* **118** (1), 4–16.

Leshno, M., Lin, V.Y., Pinkus, A., Schocken, S., (1993). “Multilayer feed forward networks with a nonpolynomial activation function can approximate any function”. *Neural Networks,* **6**(1), 861–867.

Liggett, J. A., Chen, L. C. (1994). “Inverse transient analysis in pipe networks”. *Journal of Hydraulic Engineering,* **120** (8), 934–955.

Liggett, J. A., Chen, L. C. (1995). “Monitoring water distribution systems: the inverse method as a tool for calibration and leak detection”. *Cabrera, E., Vela, A. F. (Eds.), Improving Efficiency and Reliability in Water Distribution Systems,* **14**, 107–134.

Maier, H.R., Graeme, C.D. (1999). “Empirical comparison of various methods for training feed-Forward neural networks for salinity forecasting”. *Water Resources Research*, **35**(8), 2591-2596.

Marchanda, H., Martin, A., Weismantel, R., Wolsey, L. (2002). “Cutting planes in integer and mixed integer programming”. *Discrete Applied Mathematics,* **123** (1-3), 397-446.

May, R.J., Dandy, G.C., Maier, H.R., Nixon, J.B., (2008). “Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems”. *Environmental Modelling and Software*, **23** (10–11), 1289–1299.

Mays, L.W. (1989). “Reliability analysis of water distribution systems”. *ASCE, New York*. ISBN: 0-87262-712-8

Meier, R. W., Barkdoll, B. D. (2000). “Sampling design for network model calibration using genetic algorithms”. *Journal of Water Resources Planning and Management,* **126** (4), 245–250.

Meirlaen, J., van Assel, J., Vanrolleghem, P.A. (2002). “Real Time Control Of The Integrated Urban Wastewater System Using Simultaneously Simulating Surrogate Models”. *Water Sci. Technol.,* **45**(3), 109-16.

Michalewicz, Z., Koziel, S., (1999). “Evolutionary Algorithms, Homomorphous Mappings, and Constrained Parameter Optimization”. *Massachusetts Institute of Technology, Spring (1999)*, **7**(1), 19-44.

Mulligan, A. E., and Brown, L. C. (1998), "Genetic algorithms for calibrating water quality models." *Journal of Environmental Engineering, ASCE,* **124**(3), 202-211.

Murry, R., Hart, W., and Berry, J., (2006), “Sensor Network Design for Contamination Warning Systems”. *Water Security Congress Conference,* September 10th - 12th 2006, Washington DC, USA (2006).

Muschalla, D., Schütze, M., Schroeder, K., Bach, M., Blumensaat, F., Klepiszewski, K., Pabst, M., Pressl, A., Schindler, N., Wiese, J., Gruber, G. (2008). “The HSG Guideline Document for Modelling Integrated Urban Wastewater Systems”. *11th International Conference on Urban Drainage (ICUD)*, Edinburgh, Scotland, UK, (2008).

Nagar, A. K., Powell, R. S. (2000). “Observability analysis of water distribution systems under parametric and measurement uncertainty.” *The proceedings of ASCE Joint Conference on Water Resources Engineering and Water Resources Planning and Management*, Minneapolis, U.S., (2000). ISBN 0784405174.

Ormsbee, L. E., Wood, D. J., (1986). “Explicit network calibration”. *Journal of Water Resources Planning and Management,* **112** (2), 166–182.

Ozdemir, O.N, and Ucak, A. (2002). “Simulation of Chlorine Decay in Drinking-Water Distribution Systems”. *Journal of Envir. Engrg,* **128**(1), 31-40.

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”. *ASCE Journal of Water Resources Planning and Management,* **134**(6), 556-568

Ostfeld, A., and Salomons, E. (2006). “Sensor network design proposal for the battle of the water sensor networks (BWSN).” *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA (2006).

Ostfeld, A., Uber, J., and Salomons, E. (2006), “Battle of the water sensor networks (BWSN): A design challenge for engineers and algorithms.” *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA (2006).

Ostfeld, A., and Salomons, E. (2004), “Stochastic Early Warning Detection System Model for Drinking Water Distribution Systems Security.” *Proc. World Water & Environmental Resources Congress*, ASCE, Reston, VA.

Parmesan, C., and Yohe, G. (2003). “A globally coherent fingerprint of climate change impacts across natural systems”. Nature **421**(2 January 2003), 37-42.

Petersen, B., Gernaey, K., Henze, M., and Vanrooleghem, P. A. (2002), "Evaluation of an ASM1 model calibration procedure on a municipalindustrial wastewater treatment plant." *Journal of Hydroinformatics*, **4**, 15-38.

Piller, O., Brémond, B., and Morel, P., (1999), “A spatial sampling procedure for physical diagnosis in a drinking water supply network”. *Savic, D. A., Walters, G. A. Water Industry Systems: Modelling and Optimization Applications.* University of Exeter, UK. **1**, 209–316.

Prasad, T.D. (2010), “A Clonal Selection Algorithm for the C-Town Network Calibration”. *Water Distribution System Analysis 2010 (WDSA2010),* Tucson, AZ, USA, Sept. 12-15, 2010

Preis, A., and Ostfeld, A. (2006), “Multiobjective sensor design for water distribution systems security.” *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Preis, A., Whittle, A.J., Ostfield, A. & Perelman, L. (2009), "On-line hydraulic state estimation in urban water networks using reduced models," *Proc. Computing & Control in the Water Industry (CCWI),* 2009

Preis, A., and Whittle (2009), "Multi-Objective Sensor Placement model for integrated monitoring of hydraulic and Water Quality Parameters". *Proc. World City Water Forum,* 2009

PREPARED (2010), at http://www.prepared-fp7.eu/prepared-home (last accessed 01-08-2011

PREPARED (2011). "Uncertainty Quantification and Reduction in Urban Water Systems Modelling: Evaluation report", Deliverable D3.6.1, accessible on: http://www.prepared-fp7.eu/prepared-publications (last accessed 18-3-2011).

Propato, M., and Piller, O. (2006), "Battle of the water sensor networks." *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA, 2006.

Randall, D.A., (2006). "The Shallow Water Equations". *Department of Atmospheric Science*, Colorado State University, Fort Collins, Colorado 80523.

Rode, M., Suhr, U., and Wriedt, G. (2007), "Multi-objective calibration of a river water quality model-information content of calibration data". *Ecological Modelling.* **204**, 129-142.

Rossman, L. A. (2000). "EPANET2 Users Manual". National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH 45268, USA.

Rossman, L. A. (2004). "Storm Water Management Model Users Manual-Version 5.0". National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH 45268, USA.

Pudar, R. S., Liggett, J. A., (1992), "Leaks in pipe networks". *Journal of Hydraulic Engineering.* **118**(7), 1031–1046.

Reddy, P. V. N., Sridharan, K., Rao, P. V., (1996), "WLS method for parameter estimation in water distribution networks". *Journal of Water Resources Planning and Management.* **122**(3), 157–164.

Rubinstein, R.Y., and D. P. Kroese. (2004), "The cross-entropy method: A unified approach to combinatorial optimization" *Monte-Carlo simulation and machine learning.* New York: Springer-Verlag.

Ruggaber, T.P, Talley, J.W., Montestruque, L.A. (2007). "Using Embedded Sensor Networks to Monitor, Control, and Reduce CSO Events: A Pilot Study". *Environmental Engineering Science*, **24**(2), 172-182.

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.*

Savic, D. A., Walters, G. A. (1995), "Genetic algorithm techniques for calibrating network models". *Technical Report 95/12,* Centre for Systems and Control Engineering, University of Exeter, UK.

Schellart, A.N.A, Tait, S.J., Ashley, R.M., Farrar, D., Hanson, D. (2008). "Uncertainty in deterministic predictions of flow quality modelling and receiving water impact". *11th International Conference on Urban Drainage*, Edinburgh, Scotland, UK.

Schutze, M.R., Butler, D., Beck, M.B. (2002), "Modelling, simulation and control of urban wastewater systems". *Springer; 1st Edition.* ISBN-13: 978-1852335533.

Schutze, M., Campisano, A., Colas, H., Vanrolleghem, P., Schilling, W. (2003). "Real-Time Control of Urban Water Systems". *International Conference on Pumps, Electromechanical Devices and Systems Applied to UrbanWater Management (PEDS 2003),* Valencia, Spain.

Shang, F., Uber, J. G., Polycarpou, M. M. (2002), "Particle backtracking algorithm for water distribution system analysis". *Journal of environmental engineering,* **128**(5), 441–450.

Sharp, W. W., Pfeffer, J., Morgan, M. (1991), "Insitu chlorine decay rate testing. Water quality modeling in distribution systems". *Water quality modeling in distribution systems*, AWWARF and US EPA, Cincinnati, 311–321.

Shen, H., and McBean, E. (2010). "Hydraulic Calibration for a small water distribution network." *Water Distribution System Analysis 2010 (WDSA2010),* Tucson, AZ, USA.

Smith, W.F. (2005). "Experimental Design for Formulation". *Chapter 5: "Designs for non-simplex shaped regions."* Philadelphia: ASA–SIAM, ISBN: 0898715806.

Strelkoff, T. (1970), "Numerical Solution of Saint-Venant Equations". *Journal of the Hydraulics Division*, **96**(1), 223-252.

Stukel, T. A., Reed, E C., Greenberg, E. R., and Iacobs, N. I. (1987), "Comparing three sampling designs for monitoring coliforms in small community water systems." *J. AWWA, **79**(11), 50-54.*

Todini, E. (1999). “Using a Kalman filter approach for looped water distribution network calibration”. *Savic, D. A., Walters, G. A. Eds, Proceedings Water Industry Systems: Modelling and Optimization Applications. **1**, 327–33.* University of Exeter, UK.

Trachtman, G. B. (2006). “A ‘strawman’ common sense approach for water quality sensor site selection.” *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, USA (2006).

Tzatchkov, V.G., Aldama, A.A., Arreguin, F.I. (2002), “Advection-Dispersion-Reaction Modeling in Water Distribution Networks”. *J. Water Resour. Plng. and Mgmt.* **128**(5), 334-343.

van Bloemen Waanders, B. G., (2004). “Application of optimization methods to calibration of water distribution systems”. *Sehlke, G., Hayes, D. F., Stevens, D. K., World Water Congress, ASCE Vol. **138**, 448–448*, Salt Lake City, Utah, USA.

Vanrolleghem, P. A., Benedetti, L., and Meirlaen, J. (2005), "Modelling and real-time control of the integrated urban wastewater system." *Environmental Modelling & Software*, **20**(4), 427-442.

Velez, C.A et al. (2007), “Optimisation of Urban Wastewater Systems Using Model-based Design and Control. Case study of Cali, Colombia”. *Conferencia latinoamericana de saneamient,* Latinosan (2007).

Vitkovsky, J. P., Simpson, A. R. (1997). “Calibration and leak detection in pipe networks using inverse transient analysis and genetic algorithms”. *Tech. Rep. R **157**,* Department of Civil and Environmental Engineering, University of Adelaide.

Vitkovsky, J. P., Liggett, J.A., Simpson, A. R., Lambert, M. F. (2003). “Optimal Measurement Site Locations for Inverse Transient Analysis in Pipe Networks”. *Journal of Water Resources Planning and Management,* **129** (6), 480–492.

Vorosmar, C.J., Green, P., Salisbury, J., Lammers, R.B. (). “Global Water Resources: Vulnerability from Climate Change and Population Growth”. *Science*, **289**(5477), 284-288

Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. (2003). "A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters". *Water Resources Research*, **39**(8), p. 1201.

Watson, J.P., Greenberg, H.J., and Hart, W.E. (2004). "A Multiple-Objective Analysis of Sensor Placement Optimization in Water Networks". *Critical Transitions in Water and Environmental Resources Management*, ISBN: 9780784407370

Weickgenannt M., Kapelan, Z., Blokker, M., and Savic, D.A. (2010). "Risk-based Sensor Placement for Contaminant Detection in Water Distribution Systems". *J. Water Resour. Plng. and Mgmt. vol.* ***136****, p629;*

Winkler, S., Gruber, G., and Pressl, A. (2005). "An integrated water quality monitoring network". *Assessing Infiltration and Exfiltration on the Performance of Urban Sewer Systems (2005),* ISBN: 9781843391494

Wong, F.S. (1985). "First-Order, Second-Moment Methods", *Computers & Structures,* (**20**)4, 779-791

Wood, D. J., (1980). "Slurry flow in pipe networks", *Journal of Hydraulics* **106** (1), 57–70.

Wu, Z.Y., and Sage, P. (2006). "Water loss detection via genetic algorithm optimization-based model calibration". *8th Annual Water Distribution System Symposium,* Cincinnati, Ohio, USA.

Wu, Z. Y., and Walski, T. (2006). "Multiobjective optimization of sensor placement in water distribution systems." *Proc., 8th Annual Water Distribution System Analysis Symp.,* Cincinnati, Ohio, USA 2006.

Wu and Walski (2010), "Progressive Optimization Approach for Calibrating Eps Hydraulic Models". *Water Distribution System Analysis 2010 – WDSA2010, Tucson, AZ, USA, Sept. 12-15, 2010*

Zitzler, E. and L. Thiele (1999). "Multiobjective evolutionary algorithms: A comparative case study and the strength pareto approach". *IEEE Transactions on Evolutionary Computation* **3**(4), 257–271

Zitzler, E., Deb, K., and Thiele. L. (2000). "Comparison of Multiobjective Evolutionary Algorithms: Empirical Results". *Evolutionary Computation*, **8**(2), 173-195.

Zhu, H., Shaofeng, J. (2004). "Uncertainty in the Spatial Interpolation of Rainfall Data" *Institute of Geographic Sciences and Natural Resources Research"*. CAS, Beijing 100101, China.

8. Appendix A: Taxonomy of Sensor Placement Algorithms

The following tables summarises the optimal macro-location procedures discussed in section 6.

KEY

- Column 4 indicates the intended area of application, they are: WDS (Water Distribution Systems), WWS (Waste Water Systems) or both.
- Column 5 indicates the objectives for which the algorithm has been applied in literature, they are: AM (Ambient Monitoring), DET (Contaminant Detection), COM (Compliance), CAL (Calibration).
- Column 6 indicates whether or not the procedure is able to recover a Pareto Front without a contrived objective function.

Table 1: Optimal Sensor Placement Heuristics

Method	Title of Paper	Author(s)/Reference (Year)	Original Application Area	Original Objective(s)	Multi-Objective	Comments
Max-Sum	"Sampling Design Methods for Water Distribution Model Calibration"	Bush and Uber (1998)	WDS	CAL	No	Difficult to combine with cost analysis.
Max-Min	"Sampling Design Methods for Water Distribution Model Calibration"	Bush and Uber (1998)	WDS	CAL	No	Difficult to combine with cost analysis.

Weighted-Sum	"Sampling Design Methods for Water Distribution Model Calibration"	Bush and Uber (1998)	WDS	CAL	No	Difficult to combine with cost analysis.
Iterative Deepening	"Iterative Deepening of Pareto Solutions in Water Sensor Networks"	Eliades and Polycarpou (2006)	WDS	DET	No	Performed well in the BWSN. May be prone to finding local optima.
Shortest Path Algorithms	"Optimal Sampling Design using Shortest Path Algorithms"	De Shaetzen et al (2000)	WDS	CAL	No	The benefit to calibration must be related to distance from some source. Works well for pressure devices on a WDS.
SLOTS Sensor Placement Algorithm	"SLOTS: an effective Algorithm for sensor placement in Water Distribution Systems"	Dorini et al (2010)	WDS	DET	No	Consistently outperforms greedy algorithms. Computationally Efficient.
Simple Genetic Algorithm		Kapelan et al (2005), Huang McBean et al (2006).	WDS	CAL-	No	Effectively searches vast search spaces in reasonable time.
Improved Genetic Algorithm	"Optimisation model and algorithms for Design of Water Sensor Placement in Water Distribution Systems"	Guan et al (2006)	WDS	DET	No	

Multi-Objective Genetic Algorithm (MOGA)	Kapelan et al (2003, 2005), de Schaetzen et al. (2000)	WDS	CAL / DET	Yes	Can directly incorporate a cost objective along with some appropriate calibration criteria.
Non-Dominated Sorting Genetic Algorithm-II (NSGA-II)	Kapelan et al. (2005), Preis and Ostfeld (2006), Preis and Whittle (2009), Ostfeld et al (2009), Weickgennant et al (2010)	WDS	CAL / DET	Yes	Can directly incorporate a cost objective along with some appropriate calibration criteria.
Strength Pareto Algorithm (SPEA)	Cheung and Pillar (2006)	WDS	DET	Yes	
Integer Programming	Berry et al. (2004), Berry et al (2006), Watson et al (2004)	WDS	DET	Yes	Success relies on a properly construction linear objective function not yet tackled in literature (has never been previously applied to SD for calibration). *Can* reliably find global optima.

Cross-Entropy Optimisation	Dorini et al (2006)	WDS	DET	Yes