Bayesian experimental design of tracer studies to monitor wastewater leakage from sewer networks

J. Rieckermann,1 M. E. Borsuk,2 D. Sydler,1 W. Gujer,3 and P. Reichert1

Received 14 September 2009; revised 11 February 2010; accepted 10 March 2010; published 6 August 2010.

Despite more than a decade of research, the magnitude of wastewater leakage from defective sewer systems into groundwater supplies is still largely unknown, partly because reliable measurement methods are lacking. Although recently suggested in-sewer tracer studies present a promising solution, it is unclear how to optimally design such studies in network settings. In this study we present a formal experimental design procedure, which uses Bayesian data analysis to improve the diagnosis of sewer leakage by combining tracer test data with prior knowledge on network topology and condition. From a simulation study, we show that (1) if a single sewer section is expected to have high leakage, that section should be distinguished in measurement layouts through isolated tests or appropriate overlapping of multiple tests; (2) if multiple sections are expected to have high leakage, layouts with tests that cover more than one high-leakage section should be avoided; and (3) if a robust experimental design is desired, a balanced layout of tests that overlap multiple sections of high leakage, yet minimizes stretch length, should be chosen. This design will have the additional benefit of inducing covariance in the posterior distribution of exfiltration estimates, which can be used to advantage in subsequent studies. We apply these guidelines to a case study of a catchment in Zurich, Switzerland, and find that optimal layout design can improve the anticipated gain of information substantially relative to designs based on practical considerations alone. Remaining concerns regarding the procedure include (1) the generally poor understanding of the mechanisms governing sewer leakage, which can hamper reliable prior information on exfiltration; (2) the currently low measurement precision of sewer tracer studies, which might only allow for the detection of large leaks; and (3) the need for numerical implementation of the Bayesian inference procedure, which requires careful tuning and long computation times. In general, we were able to demonstrate that the incorporation of prior information through a Bayesian procedure adds significant value to experimental design, especially in situations with few “hard” data but good site-specific knowledge, which is common in water resources research.


1. Introduction

Anthropogenic pollution continues to contaminate surface water courses and groundwater supplies. One source of particular concern in urban areas is the leakage of wastewater from sewers into aquifers used for potable water supply. Because wastewater contains pathogens, pharmaceuticals, and personal care products, this situation can present serious public health risks [Böhlke et al., 2006; Kent et al., 2007; Masciopinto et al., 2008; Nakada et al., 2008].

Hence, the assessment of wastewater leakage from deteriorated sewers has received increased attention during the last decade [Bertrand-Krajewska et al., 2006; Bishop et al., 1998; Vollersen and Hvitved-Jacobsen, 2003; Wolf et al., 2004]. Unfortunately, knowledge about sewer leakage and its environmental impact in particular settings is still limited. As an example, Wolf and Hötzl [2006] estimate sewer leakage from groundwater quality measurements in the catchment of Rastatt, Germany, to be in the very wide range of 0.2 to 1001 mm/a. Interestingly, a comprehensive monitoring program in the same system originally suggested rather small leakage rates of 2 mm/a [Klinger et al., 2005]. In a recent review, Rutsch et al. [2008] conclude that exfiltration is a heterogeneous and dynamic process that is difficult to measure (e.g., from groundwater chemical analysis or pressure testing). To overcome this undesirable situation, Rieckermann et al. [2005] suggest quantifying leakage directly from in situ sewer tracer tests. Such experiments (1) are relatively inexpensive, (2) have measurement errors of only a few percent (of the dosed
tracer), (3) can monitor leakage under realistic flow conditions, and (4) do not disturb existing leaks during the measurement process (as do pressure tests). Unfortunately, the tracer method shares two disadvantages of groundwater quality measurements. First, the measured tracer loss is an integrated measure, which provides no information about the exact location of the leaks in the investigation reach. Second, the expected measurement accuracy, which is estimated to be a few percent of the labeled flow, might not allow for the detection of leaks with such small loss rates as reported by Klinger et al. [2005].

[5] To avoid use of limited resources on inconclusive or poorly performed tests, Rieckermann et al. [2005] investigated the specific design of one single tracer experiment (e.g., regarding the amount of tracer mass dosed and the specific timing of the individual injections) in a previously selected sewer reach. In this study we are interested in how to optimally design tracer test layouts to efficiently screen a network for losses, i.e., how to efficiently choose appropriate reaches and combine multiple experiments. The answer is generally not straightforward because the best layout will depend on the location and magnitude of leaks relative to the inflows, which can differ considerably from system to system. While a few tracer tests might therefore be enough if the network is deteriorated and has a simple structure, a larger number and different layouts might be necessary if the sewer system is well maintained and/or has a complex topology.

[4] We propose a Bayesian experimental design procedure of tracer tests to optimally monitor leakage from specific sewer networks. To this aim, we formally combine the results of tracer tests with a priori information on the location of damages and the magnitude of exfiltration in the studied reaches. Specifically, we present the following analysis procedure:


[6] 2. A Bayesian data analysis procedure to infer exfiltration rates from the combination of prior knowledge and candidate tracer test layouts in a network.

[7] 3. An optimal experimental design procedure to select the best layout from among the set of candidate layouts.

[8] This paper is structured as follows. In section 2, we develop the methodology. We briefly describe (1) the tracer test; (2) a mass balance model; (3) the assessment of prior knowledge; (4) the combination of experimental results, prior knowledge, and model by Bayesian data analysis; (5) the numerical technique used for implementing the Bayesian method; and (6) the optimal experimental design criterion for layout selection. In section 3, we illustrate our approach with a simple example. Section 4 describes the application of the technique to a case study in Zurich, Switzerland. Finally, we discuss the results and draw conclusions in sections 5 and 6.

2. Methods

2.1. Tracer tests

[5] The basic idea of the QUEST method (Quantification of Exfiltration from Sewers with artificial Tracers) [Rieckermann, 2005] is to add a known amount of tracer $m_{in}$ at the upstream end of a sewer reach and measure the mass downstream $m_{out}$. Tracer loss is then quantified by the recovery ratio

$$ r = \frac{m_{out}}{m_{in}} $$

This dimensionless ratio can be measured with an accuracy of about 1%-3%, as demonstrated in laboratory [Bolognesi et al., 2007] and field experiments [Rieckermann et al., 2007; Rieckermann et al., 2005]. To relate recovery ratios in a network to inflows and leakage rates, we employ a simple mass balance model.

2.2. Mass Balance Model

[10] We model tracer recovery ratios in a network of generic sewer units that possess inflows and losses (Figure 1). Sewerage flows are routed according to the network structure. Given the typical treelike structure of sewer networks, we assume exactly one downstream connection for each unit, whereas multiple upstream connections are possible. We further assume that the discharge in the $i$th unit $Q_i$ is constant, all inflows $Q_{in,i}$ are added at the beginning of the unit and all exfiltration $Q_{ex,i}$ is subtracted at its end. This ensures that each unit is represented by a single flow $Q_i$. Then, the tracer recovery ratio $r_i$ is calculated as

$$ r_i = \frac{Q_i - Q_{ex,i}}{Q_i}. $$

The recovered mass of tracer $m_{tracer}$ can be calculated sequentially based on the dosed mass $m_{tracer}$ as

$$ m_{tracer} = m_{tracer,i}r_i \text{ for } i \geq 1 \text{ with } m_{tracer} = m_{tracer,0}. $$

Figure 1. Generic mass balance model to predict tracer recovery ratios in a sewer network with a series of inflows and losses.
The overall recovery ratio for a tracer test over several combined sections is thus given by

$$r_m = \frac{m_{\text{tracer}}^{\text{out}}}{m_{\text{tracer}}^{\text{in}}} = \prod_{i \in \mathcal{K}} r_i,$$  \hspace{1cm} (4)

where \( k \) is the number of sewer units that the tracer pulse passes. With this model, recovery ratios for various layouts of multiple tracer tests can be properly calculated from the network topography, discharge, and exfiltration values. When we assume that the measurement error of a single tracer experiment is normally distributed [Rieckermann et al., 2005], the likelihood function of the model is

$$f_{\text{model}}(r|\theta) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left( -\frac{(r - r_m(\theta))^2}{2\sigma^2} \right),$$  \hspace{1cm} (5)

where \( r \) is the observed tracer recovery, \( \theta = \{Q_{\text{in}}, Q_{\text{ex}}\} \) represents the inflows and leakage rates, \( \sigma \) is the standard deviation of the measurement error, and \( r_m(\theta) \) is the modeled recovery ratio (equations 2–4). In the case that a layout consists of several independent tracer tests (see Figure 5, in section 3), the resulting likelihood function is a product of such expressions. In addition to the likelihood function, we must specify prior knowledge on inflows and exfiltration rates for Bayesian statistical inference [Gelman et al., 1995].

### 2.3. Assessment of Prior Knowledge

Prior knowledge can be specified as a specific value when a quantity is well known, but should be represented by a probability distribution when a quantity is subject to uncertainty. While information on inflows can be readily estimated from the number of connected households or from discharge measurements in the sewer, exfiltration is generally more difficult to assess [Blackwood et al., 2001; Davies et al., 2001; Decker, 1997; Ellis et al., 2003; Fenz et al., 2005; Lerner, 2003; Rauch and Stegner, 1994; Ulmann, 1994; Vollertsen and Hvitved-Jacobsen, 2003; Wakida and Lerner, 2005]. [Rutsch et al., 2008] provide a comprehensive review. The problem is that exfiltration magnitude depends on many processes and factors that are generally not well known [Rutsch et al., 2008; Wolf and Hötzl, 2006]. These include the hydraulic head and flow dynamics, the wastewater composition, the sealing of leaks caused by particles and/or bacterial growth, the hydraulic conductivity of the soil and backfill material (e.g., preferential flow paths), as well as the dynamics of groundwater level at the leak. To obtain consistent results, it can be helpful to employ a general model for sewer leakage that can be parameterized using site-specific information [Christakos, 2003]. We will use a physically based model concept of leakage to infer prior knowledge on leakage rates from more easily accessible background information and qualitative expert opinion.

General model for sewer leakage: Seepage from rivers and sewers can be described by the Darcy equation [Ellis et al., 2003; Rauch and Stegner, 1994; Rutsch et al., 2008; Vollertsen and Hvitved-Jacobsen, 2003]. Assuming the groundwater level is below the sewer pipe bottom, the main relationships are

$$Q = k_{\text{clog}} A_{\text{eff}} I, \text{ with } k_{\text{clog}} < k_{\text{soil}}$$  \hspace{1cm} (6)

$$I = \frac{h_{\text{fill}} + h_{\text{wall}}}{h_{\text{wall}}},$$  \hspace{1cm} (7)

where \( Q \) is the leakage rate, \( I \) is the hydraulic gradient, \( h_{\text{fill}} \) is the water level, \( h_{\text{wall}} \) is the thickness of the sewer pipe or the layer of soil clogged with fine particles or bacteria, \( A_{\text{eff}} \) is the effective area of the sewer leak, and \( k_{\text{clog}} \) and \( k_{\text{soil}} \) are the hydraulic conductivity of the clogged layer and the soil, respectively (Figure 2). Site-specific knowledge can then be used to estimate the values or distributions of these individual parameters.

### Site-specific knowledge

The various sources of information to which a sewer operator has access can significantly improve estimates of leakage rates in a specific catchment. These include

1. measurements or maps of soil type, permeability, and groundwater table
2. measurements of groundwater quality parameters
3. maps of sewer damages and defects
4. information on sewer type, construction material, and age
5. information from detailed sewer surveys (e.g., visual inspections, CCTV monitoring, and pressure testing)
6. The goal here is for this prior knowledge to be formulated as a probability distribution for each of the parameters of the Darcy equation so that it can be propagated

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Unit</th>
<th>Parameterization</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_{\text{wall}} )</td>
<td>Pipe thickness/Thickness of the clogging layer</td>
<td>[m]</td>
<td>Lognormal distribution ( \mu = 0.02 ), ( \sigma_{\text{wall}} = 0.1 )</td>
</tr>
<tr>
<td>( h_{\text{soil}} )</td>
<td>Water level in the individual sewer section</td>
<td>[m]</td>
<td>Lognormal distribution ( \mu ) and ( \sigma_{\text{soil}} ) from database</td>
</tr>
<tr>
<td>( A_{\text{eff}} )</td>
<td>Effective area of a sewer leak</td>
<td>[m²]</td>
<td>Lognormal distribution ( \mu ) and ( \sigma_{\text{wall}} ) from database</td>
</tr>
<tr>
<td>( k_{\text{clog}} )</td>
<td>Hydraulic conductivity of the clogging layer during dry weather flow.</td>
<td>[m s⁻¹]</td>
<td>Lognormal distribution ( \mu = 2 \times 10^6 ), ( \sigma_{\text{clog}} = 0.4 )</td>
</tr>
</tbody>
</table>
uncertain quantity where \( x \) is the one that maximizes this anticipated gain of information (AGI) for a particular layout.

The anticipated gain of information (AGI) is then calculated as the median of \( G \) over the full prior on \( \Theta \)

\[
AGI = \text{median}_{\Theta \in \Theta} \{ G \},
\]

where \( \Theta \) is the full parameter space.

While the expected gain of information (calculated as the mathematical expectation, or probability weighted average, of the gain of information) is traditionally used [Lindley, 1956], we chose the median as a more robust measure that is less strongly influenced by extreme values. As this choice of a design criterion probably influences the selection of the best candidate layout, a brief description of other possibilities is provided in section 5.

2.6. Approximating Posteriors and AGI With Markov Chain Monte Carlo

Of the various possible algorithms for implementing Bayesian inference (e.g., importance sampling), Markov Chain Monte Carlo (MCMC) algorithms have the advantage that they approximate the posterior distribution without any assumptions on its shape [Gelman et al., 1995]. We used a Metropolis-Hastings MCMC routine [Hastings, 1970] in the following iterative procedure (Figure 4). First, a single vector of parameters, \( \theta \), is drawn from the prior distribution, for which a distribution of hypothetically measured recovery ratios, \( f_{\text{model}} (r|\theta) \), is calculated. Then, the MCMC is used to estimate \( f_{\text{post}} (r|\theta) \) from the prior knowledge and hypothetical recovery ratios. Next, \( G \) is computed for each point of the sample and recorded. After many iterative samplings over \( \theta \), the anticipated gain of information AGI is calculated as the median of the stored \( G \) values.

2.7. Computational Implementation

The sewer mass balance model was implemented in C++ in the UNCSIM simulation environment [Reichert, 2005], which implements algorithms for systems analysis, random sampling, and Bayesian inference using MCMC. The Darcy equation for estimating exfiltration rates from sewer damages was implemented in the simulation environment R [R Development Core Team, 2007], which was also used for statistical analysis, for database access, and for interfacing the different models and UNCSIM routines.

As the gain of information for a tracer layout is computed from many repetitions of MCMC with very different correlation structures, an automated adjustment of the MCMC jump distribution was required. Therefore, we implemented a preloop (3000 points) in which the normal jump distribution was adjusted iteratively before starting the computations.

Figure 3. Hypothetical prior (dotted gray line) and posterior (solid black line) distributions on exfiltration rates. We use the ratio between the [0.025, 0.975] interquantile ranges \( \xi_{\text{post}} \) and \( \xi_{\text{pri}} \) as a measure for the gain of information (G).

Through to a prior distribution of exfiltration values. Extensive literature on methods for expert elicitation of Bayesian priors exists [e.g., O’Hagan et al., 2006], and examples for this particular application are provided in the case study below.

2.4. Bayesian Data Analysis

According to Bayes’ theorem, prior knowledge on an uncertain quantity \( \theta \), \( f_{\text{pri}} (\theta) \), is combined with measured data \( r \) and a model-based likelihood function \( f_{\text{model}} (r|\theta) \) to yield a representation of posterior knowledge according to the equation

\[
f_{\text{post}} (\theta|r) \propto f_{\text{pri}} (\theta)f_{\text{model}} (r|\theta).
\]

Gelman et al. [1995] provide details.

2.5. Bayesian Experimental Design

To select the optimal layout, we define a measure for the gain of information resulting from the tracer test data. We then calculate the median value of this measure over the prior distribution on leakage (and inflows) to obtain a posterior “anticipated gain of information” for each layout (i.e., before collecting any actual data). The optimal layout is the one that maximizes this anticipated gain of information. Our measure of the gain of information for a particular sewer section \( i \) is based on a comparison of the spread of the posterior relative to the prior marginal distributions on sewer leakage

\[
g_i = 1 - \frac{\xi(Q_{\text{ex},i}^{\text{post}})}{\xi(Q_{\text{ex},i}^{\text{pri}})}
\]

where \( \xi(Q_{\text{ex},i}^{\text{pri}}) \) and \( \xi(Q_{\text{ex},i}^{\text{post}}) \) are the [0.025, 0.975] interquantile ranges of the prior and posterior distributions of \( Q_{\text{ex},i} \), respectively (see Figure 3). The interquartile range was chosen because it is a relatively robust measure of uncertainty. As we are primarily interested in reducing the uncertainty in sewer sections with high losses, we propose a weighted statistic combining measurements over \( n \) sections to express the overall gain of information (\( G \)) for a particular layout

\[
G = \frac{\sum E[Q_{\text{ex},i}^{\text{post}}] / \xi_{\text{post}}}{\sum E[Q_{\text{ex},i}^{\text{pri}}] / \xi_{\text{pri}}}.
\]

The anticipated gain of information (AGI) is then calculated as the median of \( G \) over the full prior on \( \Theta \).
actual Markov Chain (11,000 points). In this procedure, we approximated the correlation matrix with calculated correlations of the MCMC sample of the preloop and scaled the standard deviations of the jump distribution until an adequate acceptance rate was obtained (0.23–0.44, suggested by Gelman et al. [1995], p. 334f).

For the Zurich case, 150 repetitions were performed for each of the five layouts. The computations were performed on four to six 3 GHz CPUs and around 5–80 GB of data were stored per analysis. Although the code was run in parallel for the individual layouts, an optimal experimental design analysis took several days of computation time.

3. Numerical Test Case on a Simple Network

3.1. Scenario Analysis

We investigated the performance of 14 possible layouts (Figure 5) on a model system of three consecutive sewer units. The 14 different layouts were evaluated for (1) inflows, $Q_{in,i}$, of 0, 20, and 50 L s$^{-1}$, (2) leaks, $Q_{ex,i}$, with most probable values of 0.4 and 2 L s$^{-1}$, and (3) varying knowledge on the magnitude of the leaks (maximal possible values of 5 and 10 L s$^{-1}$) (see Table 1). For each of the 196 combinations, the AGI was calculated from a sample of 200 values of $G(r)$. Each of the $G(r)$ was based on a MCMC with 25,000 points (i.e., model runs). This scenario was used to explore fundamental questions such as: How does the position and magnitude of inflows and leaks influence the optimal experimental design? Is it recommended to overlap tracer tests? How do multiple in-sewer dilutions influence the gain of information? To what extent might correlations between sewer sections be induced by the analysis?

3.2. Results

In Figure 6, we present the results of a single MCMC for 6 of the 196 scenarios. In the graphs, we plot the marginal prior and posterior distributions from which $g_i$ is computed. Also, bivariate plots of leakage rates are presented to illustrate how the correlation structure of the joint posterior distribution changes with different layouts and priors. Figure 7 shows the AGI for the 98 scenarios with constant discharge and selected hypothetical exfiltration.

From our set of results, we can deduce the following general guidelines for the design of tracer leakage experiments in a network setting (Figure 7).
1. If possible, the greatest gains in information can be achieved by measuring each sewer section individually. This is a rather intuitive result illustrated by layout 13, which achieves gains of approximately 67%–77%. In most cases, such comprehensive studies will not be possible; however, these values provide a baseline against which other studies can be gauged.

2. When a single sewer section is expected to have high leakage, it is advisable to distinguish that section in measurement layouts. This is illustrated for column 5 by layouts 2–4 and 11–12; for column 6 by layouts 8, 10, and 12; and for column 7 by layouts 5–7. Distinguishing a section can occur either by testing it directly (e.g., layouts 2, 3, 5, 6, 8, and 10–12) or by not including it in one of multiple nested layouts (e.g., 4 and 7). In all such layouts, uncertainty can be reduced by 50% or more in the posterior distribution relative to the prior. How the two lower leakage sections are covered does not seem to make much difference (e.g., compare layouts 2, 3, 11, and 12 for column 5).

3. When multiple sections are expected to have high leakage, it is best to avoid layouts that do not distinguish the high-leakage sections. Such layouts, illustrated for column 2 by layouts 5–7, for column 3 by layout 8, and for column 4 by layouts 2–4, lead to the lowest gains of information (less than 40%). It is better in such cases to focus measurements on the two high-leakage sections (layouts 10–12 for columns 3–5, respectively) or to combine a high leakage with a low leakage section (layouts 2–4 for columns 2 and 3, respectively, and layouts 5–7 for columns 3 and 4, respectively). If many sections are expected to have equally high leakage, then some sections should be studied in isolation, even if others are ignored (layouts 10–12 for column 1). This is because combining leaky sections does not give any information about whether one, the other, or both are actually leaky.

4. When a robust experimental design is desired, a balanced layout that overlaps sections yet minimizes stretch length should be selected. This type of layout is exemplified by layout 9, which achieves reasonable, though not optimal, gains for all leakage configurations. All other two-test experimental layouts lead to poor performance in at least one possible configuration. This layout is robust because each of the sections can be distinguished from the others through inductive reasoning (i.e., Bayesian inference).

5. The power of inductive reasoning exploited by layout 9 actually has greater potential than implied by our measure of gain of information. When two or more sections are included in a single tracer experiment, the measurements induce a negative covariance in the posterior distributions of exfiltration in those sections (e.g., see Figures 6c, 6d, and 6e). As easily shown by linear error propagation, covariance in estimates of individual sections will reduce uncertainty about total leakage in the network. Covariance also means that obtaining further data about the exfiltration rate in one of the sections will yield information (through induction) about the other section. For example, in Figure 6c, if exfiltration in section 2 is later found out to be on the low end of the distribution, then exfiltration in section 3 will very likely be on the high end (because layout 2 revealed information on the sum of the two sections).

Our gain of information measure (equation 9) does not account for covariance in the posterior distributions and thus underestimates the true reduction in uncertainty result-

---

**Figure 5.** Layouts of different tracer experiments tested in the preliminary investigations. The direction of flow in the three consecutive units is from left to right. Each arc represents a single tracer experiment. Each block of layouts represents a set of variations on a general strategy.
Table 1. Overview of the Prior Distributions and the Tested Scenarios in the Simple Network Investigations

<table>
<thead>
<tr>
<th>Description</th>
<th>Parameterization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exfiltration</td>
<td></td>
</tr>
<tr>
<td>(Q_{\text{Ex,1}}): large leakage (L s(^{-1}))</td>
<td>Triangular distribution: (z_{\text{min}} = 0, z_{\text{med}} = 2, z_{\text{max}} = 10)</td>
</tr>
<tr>
<td>(Q_{\text{Ex,2}}): small leakage (L s(^{-1}))</td>
<td>Triangular distribution: (z_{\text{min}} = 0, z_{\text{med}} = 0.4, z_{\text{max}} = 5)</td>
</tr>
<tr>
<td>Infloows</td>
<td></td>
</tr>
<tr>
<td>(Q_{\text{In,1}}): large inflows (L s(^{-1}))</td>
<td>50</td>
</tr>
<tr>
<td>(Q_{\text{In,2}}): medium inflows (L s(^{-1}))</td>
<td>20</td>
</tr>
<tr>
<td>(Q_{\text{In,3}}): no inflows (L s(^{-1}))</td>
<td>0</td>
</tr>
<tr>
<td>Scenarios for inflows</td>
<td></td>
</tr>
<tr>
<td>constant discharge</td>
<td></td>
</tr>
<tr>
<td>increasing discharge</td>
<td></td>
</tr>
<tr>
<td>Scenarios for exfiltration</td>
<td></td>
</tr>
<tr>
<td>large exfiltration</td>
<td></td>
</tr>
<tr>
<td>large exfiltration but small leakage in 1</td>
<td></td>
</tr>
<tr>
<td>large exfiltration but small leakage in 2</td>
<td></td>
</tr>
<tr>
<td>large exfiltration but small leakage in 3</td>
<td></td>
</tr>
<tr>
<td>small exfiltration but large leakage in 1</td>
<td></td>
</tr>
<tr>
<td>small exfiltration but large leakage in 2</td>
<td></td>
</tr>
<tr>
<td>small exfiltration but large leakage in 3</td>
<td></td>
</tr>
<tr>
<td>Tracer test layouts</td>
<td></td>
</tr>
<tr>
<td>see Figure 5</td>
<td></td>
</tr>
</tbody>
</table>

We examined all combinations of scenarios of inflows, scenarios for exfiltration and tracer test layouts.

4. Case Study

[37] Zurich, with its approximately 400,000 inhabitants, has 925 km of public sewers that are divided into approximately 25,000 sewer reaches. To handle information on this vast network, the Zurich sewer operator (Entsorgung und Recycling Zürich (ERZ)) maintains a comprehensive sewer information system, in which data are stored and managed. This not only includes the topology of the network but also the information on the structural state of the sewer (e.g., location, magnitude, and type of known leaks), sewer maintenance, as well as detailed discharge information. ERZ is interested in innovative sewer rehabilitation strategies, and the performance of tracer tests for leakage detection is an appealing option for them. For us, the excellent database on the sewer system made it an ideal test case to investigate what can be learned by combining the results from tracer tests with prior information.

4.1. Zurich Subcatchment

[38] ERZ was interested to test a subcatchment with a presumably high risk of sewer leakage because of an elevated number of known damages. As a compromise between practical considerations and computational efforts, we concentrated on the area shown in Figure 8 as a test catchment. This consists of 21 sewer sections, which in our analysis were merged to 6 "street" units (Figure 8) because these are more relevant lengths for sewer rehabilitation and tracer measurements.

[39] We deliberately chose an area with a wide range of different flow rates: in street 6, the discharge is about 800 L s\(^{-1}\), whereas in streets 1 to 5, the discharge is only 0.5 to 7 L s\(^{-1}\). Because dilution reduces the sensitivity of a tracer experiment [Rieckermann, 2005], we scaled the variance of the tracer measurement error (\(\sigma_{\text{abs}}^2\)), based on the ratio of downstream to upstream flow:

\[
\sigma_{\text{abs}}^2 = \sigma_{\text{rel}}^2 + \left(\frac{Q_{\text{down}}}{Q_{\text{up}}}\right)^2 \sigma_{\text{rel}}^2,
\]

where \(\sigma_{\text{abs}}\) and \(\sigma_{\text{rel}}\) are the absolute and relative standard deviation, respectively. We used values of \(\sigma_{\text{abs}} = 0.01\) and \(\sigma_{\text{rel}} = 2 \times 10^{-2}\) based on our field tests showing errors in the 1%-3% range with the higher values occurring under conditions of increasing flow [Bolognesi et al., 2007; Rieckermann et al., 2007; Rieckermann et al., 2005; Rutsch, 2006].

[40] As the optimal tracer layout depends on the location and magnitude of inflows as well as expected leaks, we will present the formulation of prior knowledge for this case study in detail, emphasizing how prior exfiltration rates were assessed using expert knowledge and CCTV inspections.

4.2. Formulation of Prior Knowledge

4.2.1. Prior Knowledge on Exfiltration

[41] General model: Although there is some evidence that the Darcy equation yields low estimates for sewer losses under typical operating conditions [Rutsch, 2006; Wolf et al., 2006], this bias should not significantly alter our experimental design procedure, which focuses on relative gains in information across layouts rather than absolute exfiltration estimates. Additionally, we found the use of the
Figure 6. Prior (dotted gray lines) and posterior (histograms and solid lines) distributions of exfiltration (all values in liters per second) of selected scenarios (see Table 1). Contour lines in the bivariate plots indicate highest probability density regions (5%, 50%, and 95%). Figures 6a–6e represent scenarios with constant discharge (50 L s$^{-1}$) and different tracer measurement layouts. The scenario in Figure 6f has increasing discharge (20 L s$^{-1}$ per sewer unit) but the same tracer measurement layout as Figure 6b.
Darcy equation to significantly aid in the application of the ERZ database to exfiltration estimation.

[42] For reasons of safety and reliability, tracer experiments should not be performed under storm conditions. Therefore, it was reasonable to assume the hydraulic conductivity to be constant, with other parameter values representing dry weather conditions [Wolf and Hötzl, 2006] (see Figure 2). The effective area of the sewer leaks, \( A_{\text{eff}} \), was determined based on knowledge of the structural state of the sewers and expert interviews (see below).

[43] Site-specific knowledge base: The sewer information system of ERZ contains a comprehensive archive of damage records, which are coded according to the German ATV M143 standard [ATV-DVWK, 1999]. In addition to these individual defects, the operators document the structural state of the whole sewer section as a further attribute. The classification of the individual damages is performed manually and based on standardized optical inspection by CCTV. Each sewer section is inspected at least once every 10 years, whereas critical sections (e.g., in groundwater protection areas) are investigated more often.

[44] Following the procedure of Stiber et al. [1999], we conducted several expert interviews to elicit the relation between the operator’s description of each defect (i.e., ATV code and damage grade) and the effective area of the damage. This resulted in lognormal distributions for each of the most common pairs of ATV damage codes and damage grades, which best represented the experts’ belief on the size of the different leaks (not shown). Based on this assessment, exfiltration rates were computed as follows. First, the size of the total effective leak area \( A_{\text{eff},i} \) for the entire street \( i \) was calculated by summing up the lognormal random variables characterizing each of its leaks. Then, the exfiltration rate for each individual street was calculated with equation 6.

The uncertainties in the input variable \( A_{\text{eff},i} \) and the other model parameters were propagated by means of Monte Carlo simulation to estimate the distribution of \( Q_{\text{Ex},i} \). For use in the MCMC algorithm, we used least squares to fit triangular distributions to the samples of \( Q_{\text{Ex},i} \) (Figure 8). Lognormal distributions might be a conceptually superior choice but caused difficulties in the MCMC algorithm because the long tails were inadequately sampled, causing artifacts in the gain of information. We feel that triangular distributions are an adequate practical choice for the present purpose. For details on the elicitation and utilization of prior knowledge in this study, see Sydler [2005].

Figure 7. The anticipated gain of information (AGI) for tested layouts applied to our simple network of three consecutive sewer units (see Table 1). Dark shading indicates the most informative layouts. The rows distinguish the 14 tracer layouts and the columns show different exfiltration scenarios. Large cracks in the sewer pictograms (top row) represent large leakage (see Table 1 for details).
Because the actual losses in a sewer might be much higher than predicted by the Darcy equation, we repeated our procedure after multiplying the prior exfiltration rates reported in Figure 8 by a factor of 5. This was done to investigate the robustness of the result with respect to prior estimates of the magnitude of exfiltration.

### 4.3. Preselection of Tracer Measurement Layouts

Figure 9 shows the tested tracer measurement layouts, which were selected based on the guidelines developed from the simple network and to demonstrate the benefits of our approach relative to naively chosen layouts. Figure 9a contains six experiments to cover each section individually and should provide the reference for optimal results; Figure 9b uses three experiments that cover all the sections but were selected without the benefit of prior information or our experimental design guidelines. Intuitively, this seems to be a good layout because only one measurement station is needed at the end of the catchment, and because section 6, which has the highest flow and pressure depth, should probably be covered thoroughly. Figure 9c uses three experiments that are arranged according to guideline 3 (focus on the highest flow and pressure depth).

<table>
<thead>
<tr>
<th>street ID</th>
<th>Sewers</th>
<th>$Q_{in} \text{ [l s}^{-1}]$</th>
<th>$Q_{ex} \text{ [l s}^{-1}]$</th>
</tr>
</thead>
</table>
| 1         | circular  
d=0.3-0.45m   | 1.55                        | [0, 6e-04, 6e-02]           |
| 2         | circular  
d=0.45m      | 3.10                        | [0, 1e-04, 3e-02]           |
| 3         | egg-shape  
d1=0.6m  
d2=0.9m   | 0.55                        | [0, 2e-05, 4e-03]           |
| 4         | circular  
r=0.3-0.45m  | 1.70                        | [0, 4e-07, 2e-03]           |
| 5         | egg-shape  
d1=0.6m  
d2=0.9m   | 0.55                        | [0, 8e-05, 4e-03]           |
| 6         | rectangular  
r=0.3-0.45m  | 831.05                      | [0, 2e-05, 7e-05]           |

**Figure 8.** (a) Subcatchment in the city of Zurich for which the optimal tracer layout was identified. The locations of known damages are indicated with circles, the areas of which are proportional to the estimated mean size of the damage. (b) Simplified structure of the network used to visualize the tracer test layouts. The streets are numbered within the gray boxes.
measurements on the high-leakage sections 1 and 2 without too much concern for the other low leakage sections). Figure 9d uses two experiments that cover all sections and were selected by convenience and intuition (as with Figure 9b). Figure 9e uses two experiments that are intended to cover leaky sections 1 and 2, as well as be reasonably robust to the potential for leaks in other sections (guideline 4).

4.4. Results

[47] Optimized layouts (Figures 9c and 9e) achieve an anticipated gain of information that is much larger than naively chosen layouts (Figures 9b and 9d) (Table 2, top). The gains of information from both (about 13%) are similar to those resulting from the full measurement strategy (Figure 9a). The naively chosen layouts yield very little information (<1%) because the high-leakage sections 1 and 2 are not adequately distinguished.

[48] When the prior exfiltration estimates are increased by a factor of 5, it becomes clearer that Figure 9c, employing three tests, is slightly superior to Figure 9e which only employs two. However, the difference is not great, and the costs of performing the additional test should be considered in such a comparison. Additionally, because of its overlapping structure, Figure 9e induces a covariance in the posterior distributions for exfiltration from sections 1 and 2 that is not present in the results of Figure 9c. As mentioned above, this might add value if further measurement campaigns are anticipated.

5. Discussion

[49] Our experimental design procedure employing Bayesian methods can help avoid a waste of limited resources resulting from inconclusive tests. However, there are a number of issues that merit further discussion. These concern limitations imposed on the method by (1) the remaining imprecision of tracer tests, (2) the lack of sound prior knowledge on exfiltration, and (3) the computational resources required for Bayesian analysis. In addition, other criteria could have been used to select the optimal tracer test layout, including objective functions explicitly accounting

Table 2. Anticipated Gain of Information for Each Layout and Street in the Zurich Case Study

| Increase \( Q_{Ex} \) |  |  |  |  |  |  |
|-----------------|---|---|---|---|---|
| AGI street 1    | 0.21 | 0.00 | 0.19 | 0.00 | 0.22 |
| AGI street 2    | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 |
| AGI street 3    | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| AGI street 4    | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 |
| AGI street 5    | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| AGI street 6    | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| AGI total       | 0.13 | 0.00 | 0.12 | 0.00 | 0.13 |

| Increase \( Q_{Ex} \) |  |  |  |  |  |  |
|-----------------|---|---|---|---|---|
| AGI street 1    | 0.77 | 0.01 | 0.77 | 0.01 | 0.72 |
| AGI street 2    | 0.22 | 0.01 | 0.26 | 0.00 | 0.22 |
| AGI street 3    | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 |
| AGI street 4    | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| AGI street 5    | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 |
| AGI street 6    | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 |
| AGI total       | 0.52 | 0.01 | 0.53 | 0.01 | 0.50 |

*Each value represents the average of 150 MCMC-based estimates. The top table employs prior estimates for exfiltration directly from Darcy equation, while the bottom table multiplies such estimates by 5. Differences less than 0.01 cannot be distinguished from sampling uncertainty.
for environmental protection or monetary costs. We discuss each of these issues below.

[50] Our results indicate that for real-world settings the precision of tracer measurements might be a critical factor. If the estimates provided by the Darcy model are correct, current tracer substances and measurement equipment do not provide much information (uncertainty reductions of only 12%–13%) for sewers such as those investigated in Zurich. We estimate that the measurement error, which is approximately 1%–3% of the dosed tracer mass [Bolognesi et al., 2007; Rieckermann et al., 2005], must be lowered by a factor of 25 to 100 to reliably detect the “typical” (German) leakage rates reported by Klinger et al. [2005] or the leaks analyzed in our case study. However, as the sewers of Zurich specifically, and Switzerland in general, are in a very good condition by global standards, we expect that the method might be more informative in cities with possibly larger leaks. This is supported by our analysis for hypothetical exfiltration estimates that are a factor of 5 greater than those estimated by the Darcy equation; reductions in uncertainty were greater than 50%.

[51] Although it was useful for structuring the elicitation of a prior distribution on exfiltration, we recommend skepticism in using the Darcy equation to estimate the actual magnitude of leakage from sewers. While it performed satisfactorily in laboratory settings [Ellis et al., 2003], it has been found that for real-world systems, model predictions and measurements can differ by several orders of magnitude [Wolf et al., 2006]. The use of CCTV to assess the size of cracks is prone to error and the relationship between crack size and leakage is not clear. Yet CCTV inspections are the only routine method, so we do not see other options for informing expert priors in the near future. Fortunately, while estimates of actual exfiltration rates might be highly uncertain, our experimental design procedure is relatively insensitive to exfiltration magnitude. This is because the choice of an optimal layout is based on a comparison against alternative layouts rather than absolute performance.

[52] We believe that the Bayesian approach to experimental design is an extremely useful concept, especially for problems with a large body of knowledge but only few “hard” data. However, numerical implementation of the MCMC procedure can be intensive. While, theoretically, a correctly constructed Markov chain will eventually converge to the posterior distribution, a poor choice of the jump distribution can make the algorithm inefficient. In experimental design, a large number of repetitions are required to compute the anticipated gain in information. Therefore, the jump distribution must be scaled automatically by the algorithm rather than actively by the analyst. This can be a challenge for certain parameter values and covariance structures (which are functions of the network topology and leakage characteristics). Additionally, computation time can limit the number of layouts tested, which places importance on preselection of layouts according to general guidelines. One method for overcoming computational limitations could be to chose a different algorithm for computing posterior distributions that does not rely on large numbers of repeated simulations. Alternative methods that perform nearly as well and require less CPU time include particle filters [Salamon and Feyen, 2009; Smith et al., 2008; Sun et al., 2009; van Delft et al., 2009] or, as our model is inherently linear, ensemble Kalman filters. Ensemble Kalman Filtering (EnKF) [Evensen, 2003; Nowak, 2009] has been used extensively in hydrological modeling, including surface water and complex groundwater models [Franssen and Kinzelbach, 2008; Nowak, 2009; Reichle et al., 2008]. Franssen and Kinzelbach [2008] provide an overview and Dowd [2007] directly compares MCMC and EnKF algorithms.

[53] With regard to the selection of optimal candidate layouts, we chose a relatively straightforward objective function based on the median gain of information between the prior and posterior distributions on sewer leakage. Other objective functions might be formulated that punish for undetected leakage, merit detected leakage, or account for experimental costs. Such functions lead to cost-optimal designs in O&M rules [e.g., Feyen and Gorelick, 2005]. In the study of Rieckermann et al. [2006], we argued that even more sophisticated decision analyses could be conducted that evaluate the relative merits of undertaking various field measurements, directly performing sewer repairs, or doing nothing at all. In addition to experimental costs, such analyses would need to account for economic information on groundwater contamination and infrastructure repair. The value of conducting tracer studies to reduce uncertainty could then be assessed in a truly holistic manner.

### 6. Summary and Conclusions

[54] Diffuse groundwater pollution from defective sewer systems is still largely unknown, partly because reliable measurement methods are lacking. Recently, in-sewer tracer experiments have been suggested as a promising solution. In the present study, we investigated how to optimally design such studies in a network setting and draw the following conclusions.

1. We developed and applied a Bayesian experimental design procedure to identify the optimal tracer measurement layout from among selected candidates for evaluating sewer leakage in a network. The procedure employs prior knowledge pertinent to leaks, such as information on inflows and structural damages.

2. Absolute estimates of the gain in information to be expected from tracer measurements should be viewed with skepticism because the underlying prior exfiltration estimates are typically uncertain. However, because the experimental design procedure is based on relative reductions in uncertainty across candidate layouts rather than absolute performance, the choice of an optimal layout is fairly insensitive to prior estimates.

3. On the basis of our analysis, we suggest the following general guidelines for experimental design of tracer studies: (i) when a single sewer section is expected to have high leakage, distinguish that section in measurement layouts through isolated study or appropriate overlapping; (ii) when multiple sections are expected to have high leakage, avoid layouts that cover more than one high-leakage section; and (iii) when a robust experimental design is desired, choose a balanced layout that overlaps sections of high leakage yet minimizes stretch length. This will have the additional benefit of inducing covariance in the posterior distribution of exfiltration, which can be used to advantage in subsequent studies.
4. Topics for further work include improvement of the precision of tracer experiments, further experience in calibrating estimations of leakage (e.g., from Darcy-type models) against measurements, consideration of algorithms for inferring posterior distributions that are more computationally efficient than MCMC, and incorporation of experimental design into a more holistic decision analysis of sewer monitoring and repair, while accounting for computational constraints.

5. We believe that Bayesian methods for experimental design can be equally useful in other areas of water resources research, where few “hard” data, but good site-specific knowledge, are available.

Acknowledgments. We thank the Swiss Federal Office for Education and Science (BBW) for financial support. The authors wish to acknowledge the cooperation of Entsorgung und Recycling Zürich (ERZ) in providing us with comprehensive expert information on the sewer network and access to their detailed databases. We are especially indebted to Markus Antener, Thomas Schneider, Walter Burgermeister, and Christoph Zemp for their full and efficient assistance before, during, and after the expert elicitation.

References

ATV-DVWK (1999), Registration, classification and evaluation of drainage systems outside buildings (Merkblatt M 149: Zustandsersassung, -klassifizierung und -bewertung von Entwässerungssystemen außerhalb von Gebäuden; Regelwerk Abwasser-Affabl), DWA, Hennef.


Kling, 1., et al. (2005), Leaky sewers: Measurements under operating conditions, paper presented at Fourth World Wide Workshop for Young Environmental Scientists (WWW-YES), Urban waters: Resource or risk?, Domaine de Chéroux, Vitry sur Seine, France, 10–13 May 2005.


M. E. Borsuk, Thayer School of Engineering, Dartmouth College, Hanover, NH 03755, USA.

W. Gujer, Eawag: Swiss Federal Institute of Aquatic Science and Technology and Swiss Federal Institute of Technology (ETH), Dübendorf, Switzerland.

P. Reichert, J. Rieckermann, and D. Sydler, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland. (joerg.rieckermann@eawag.ch)