Description of doctoral project and research results achieved to date

Groundwater is a resource for drinking water and hence needs to be protected from contaminations. However, many well catchments include an inventory of known and unknown risk sources which cannot be eliminated, especially in urban regions. As a matter of risk control, all these risk sources should be monitored. Many national regulations and UN guidelines already suggest monitoring as a measure of risk control, but make no statements on how to assess or design monitoring under the fact of uncertainty. Finding optimal positions of monitoring wells for such purposes is challenging because there are various parameters (and their uncertainties) that influence the reliability and optimality of a suggested monitoring location and of overall monitoring networks. Knowledge about the catchment and the risk sources can reduce this uncertainty, but it still exists. There could be uncertainty in ambient flow directions due to alternating hydraulic situations, e.g., high or low discharge of an infiltrating river inside the catchment, or there could be uncertainty in hydrogeology, e.g., in the conductivity field, which cannot be known at each point of the catchment.

In many well catchments monitoring networks already exist, but often they have grown historically. Each monitoring well could have a different task, e.g., to measure the hydraulic head of an aquifer, or to find the shape of an already existing contaminant plume. Mostly, the existing monitoring networks are not designed to protect the production wells from potential risk sources. In case of failure of a risk source, i.e., when contamination enters the aquifer unnoticed, the contamination could pass the whole monitoring network without being detected before affecting the quality of the water produced at the pumping wells. In the best case, this would just provoke a shut down of the production wells; in the worst case, it would lead to adverse health effects of the consumers. In order to prevent these negative effects, an optimal monitoring network should detect all potential contaminations that could affect the production wells, and should do so as early as possible. The remaining time between the detection of the contaminant by the monitoring network and its arrival at the production well is the reaction time available for the well operator to install counter measures. Blindly increasing the reaction time at a constant high detection probability would lead to a cost explosion for the monitoring network. Thus, the challenge is to find best compromises (so-called Pareto-optimal solutions) between detection probability, early-warning time, and installation or operation costs. Regarding the high number of risk sources and the multi-objective character of the problem, it is hardly possible to find optimal monitoring networks without computational help.

The goal of this project is to contribute to the reliability of drinking water resource protection through optimal early-warning monitoring networks. Technically, we want to develop a tool that can optimize early-warning monitoring networks for real catchments. The three overall objectives are (1) to detect all contaminations emanating from all risk sources inside the catchment, (2) to detect them as early as possible, and (3) to find networks with low installation and operation costs. Since this tool should be used by water suppliers, it needs to be as simple as possible, but as complex as required to minimize computing time with sufficient accuracy.

The tool consists of two main parts: (1) transport simulation for all possible plumes that could emanate from risk sources, and (2) the optimization of possible monitoring networks. To also include uncertainty in the transport simulation, different hydraulic scenarios and aquifer realizations are taken into consideration. The transport simulation is done by a particle racking random walk
(PTRW) algorithm (Lagrangian approach) and the optimization is done by a genetic algorithm.

The transport simulation can be split into two parts. The first part is to identify the well catchment in order to know the area in which risk sources could affect the production well. The second part is to simulate the possible contaminant plumes from the risk sources which are still considered. Both parts are done for all hydraulic scenarios and aquifer realizations.

The optimization problem is multi-objective and the objectives are competitive. That means, one can not find an overall optimal solution because such a solution does not exist: a monitoring network cannot monitor each risk source in a catchment, providing a good early warning time, but be cheap at the same time. Instead, one can find many so-called Pareto-optimal solutions [1]. A Pareto-optimal solution is a trade-off between the individual competing objective functions. It is defined as a solution that cannot be improved in any objective without degrading at least one other objective. All Pareto-optimal solutions together form the Pareto front. The challenge is to find a good approximation of the Pareto front in little time, covering the complete Pareto front. An approximation of a Pareto front that is partly very good but does not cover the edges of the front, fails in diversity of the solutions. Often, it is a competing problem itself to get a good and diverse approximation [1].

The schedule of this work was clearly defined due to the problem sketch and through a cooperation with water suppliers (deadlines, which need to be held).

At first, we had to know which optimization algorithm should be used. As a result, we created a mixture of the genetic algorithm NSGA-II [2] and useful ideas taken from the multi-objective optimization framework BORG [3].

The second task was to get a fundamental understanding about the factors that coordinate the optimal position for a single monitoring well that is used to monitor a single risk source. Therefore, we used the well-known 2D advection-dispersion equation and developed analytical solutions for detection probability as a function of monitor locations. The probabilities result from considering different uncertainties in the system (e.g., uncertainty in ambient flow direction, uncertainty in the exact risk source position, uncertainty in relevant transport parameters like transverse dispersivities).

Third, we had to speed up both the optimization procedure and the transport simulation. To speed up the optimization procedure, we developed a method based on the analytical solutions mentioned above. Using the analytical solutions, we could restrict the search space of the optimization problem to only promising monitoring-well positions. Due to this restriction, the optimization needs less time for producing good results. Nevertheless, one could lose optimal positions by a restriction based on analytical estimations. Therefore, we also developed a method with which we can restrict the search space as much as possible while still guaranteeing that the restricted search space still contains all Pareto-optimal solutions. Compared to the initial search space, we can diminish the search space by around 90%. To improve the diversity of the Pareto front, we created specialized sub-spaces of the search space related to different properties of the possible monitoring-well positions (e.g., all positions that can provide a good early warning time form a sub-space). These sub-spaces are available for the search algorithm and the algorithm can adaptively explore a mixture of these spaces to speed up its convergence. Since it is relatively easy to create sub-spaces that satisfy one objective at a time, the edges of the Pareto front can be explored very well with this method.

To speed up the transport simulations and the post-processing that is necessary as interface to the optimization algorithm, we use a statistical approach to reconstruct the breakthrough curves. The classical approach using PTRW is to count the number of particles in the particular control volumes for given time intervals. For the classical approach, one needs at least around 30 particles in each control volume and time interval in order to make a valuable estimation of the breakthrough curve.
To guarantee this number, a high initial number of particles is necessary, which slow down the calculation. Instead, our statistical approach uses the inverse Gaussian distribution to approximate a break-through curve from the temporal moments of the arrival time distribution. These moments can be calculated by a much smaller number of particles, such that we can reduce the total amount of particles for the calculation. The inverse Gaussian distribution is known as an analytical solution to contaminant break-through curves under certain conditions.

The forth task was to investigate refined concepts for encoding the optimization problem, which enable water suppliers to optimize their monitoring networks with our method under realistic conditions. This includes technical problems due to realistic groundwater models, but also practical issues, e.g., how to prioritize risk sources within the catchment regarding their risk endangering the production wells, or how to handle risk sources that are unknown in location (hidden risk sources). Therefore, we developed two concepts, one for a qualitative risk-prioritization and one for handling unknown risk sources. For the qualitative risk-prioritization, all risk sources can be classified in four different risk classes, (1) almost tolerable risk, (2) medium risk, (3) severe risk, and (4) unknown risk. The forth risk class is a special class for the unknown risk sources. The objectives of detection probability and early-warning time are each split into objectives for each risk class. The costs are independent of the different risk classes and do not need to be split up. With that, the optimization algorithm is searching for monitoring networks that are optimal for all risk classes but also for networks, which are, for example, only optimal for the severe risk class.

Risk sources that are unknown in their location could occur everywhere in the catchment. It is impossible to optimize a monitoring network directly for a certain set of unknown risk sources. Instead, we introduced the concept of 'line of attack' and 'line of defense' based on terms from American football. We represent all unknown risk sources by artificial risk sources surrounding the production wells with a certain radius ('line of attack'). A monitoring network that can monitor these artificial risk sources is called 'line of defense' and is also able to monitor all risk sources (the known and the unknown) within the catchment. The drawback of this method is that the best achievable early-warning time is proportional to the radius of the circle used to represent the unknown risk sources. A large radius means a good achievable early-warning time, but unfortunately the area within the circle increases simultaneously and the inner area of the circle can not be controlled by the 'line of defense'. Thus, there is an implicit trade-off between early-warning time and the size of the uncontrolled area close to the production wells.

The fifth task was to figure out the influence of uncertainties in the model to the optimization results. Two different uncertainties were considered, parametric uncertainty (longitudinal and transverse dispersion, and effective porosity) and scenario uncertainty for different hydraulic conditions. The uncertainty analysis was done using a simple Monte-Carlo simulation. The key finding was that scenario uncertainty has a strong impact to the optimization results and the considered parametric uncertainty just a little. The main reason for the strong impact of scenario uncertainty is the dependency of the relevant risk-source inventory on the hydraulic scenario. Risk sources that can reach the production wells in one hydraulic scenario may be harmless in other hydraulic scenarios. The main conclusion is that at least scenario uncertainty has to be taken into account during the optimization procedure, because the optimization results are very sensitive to this kind of uncertainty. By doing so, one gets robust solutions that perform well in all different scenarios. When ignoring scenario uncertainty, one may get better results for any single scenario, but the solutions will likely fail for other scenarios when they occur.

References