An expert system for real-time well field management


ABSTRACT

Well field management in urban areas faces challenges such as pollution from old waste deposits and former industrial sites, pollution from chemical accidents along transport lines or in industry, or diffuse pollution from leaking sewers. One possibility to protect the drinking water of a well field is the maintenance of a hydraulic barrier between the potentially polluted and the clean water. An example is the Hardhof well field in Zurich, Switzerland. This paper presents the methodology for a simple and fast expert system (ES), applies it to the Hardhof well field, and compares its performance to the historical management method of the Hardhof well field. Although the ES is quite simplistic it considerably improves the water quality in the drinking water wells. The ES knowledge base is crucial for successful management application. Therefore, a periodic update of the knowledge base is suggested for the real-time application of the ES.

Key words | decision support, expert system, groundwater, urban well field, well field management

INTRODUCTION

Groundwater is the most important drinking water resource in Switzerland. The natural reservoirs typically lie in valleys and are mainly fed by infiltrating water from rivers. However, the majority of the Swiss population live in valleys close to rivers. Thus, pressure on the quality of the drinking water resource is high. Industrial sites, transportation lines for hazardous materials, sewage lines and old waste disposal sites result in a high potential for groundwater contamination and constitute a considerable challenge for well field operation. In this paper we present the methodology for the design of an expert system (ES) for real-time well field management and apply it to a case study: Hardhof well field in Zurich, Switzerland.

Well field management became popular in the 1980s with the optimization of pump and treat schemes in groundwater remediation (e.g., Gorelick 1983; Gorelick & Voss 1984). The solutions to the optimization problems were obtained analytically for simple groundwater flow configurations. However, these analytical solutions are not applicable to general groundwater flow problems with complex geometry. Alternative management strategies for reservoirs have been proposed for synthetic problems (e.g., Gordon et al. 2000). The first successful applications of real-time management of underground reservoirs were implemented in the oil industry (e.g., Saputelli et al. 2006). With the increasing availability of online measurements of hydraulic head and water quality indicators, real-time management of groundwater becomes relevant (e.g., Bauser et al. 2010; Cheng et al. 2011).

Site description

The Hardhof well field is fed directly by the river Limmat in the north and west. In the south, it is artificially
recharged with river bank filtrate in three infiltration basins and 12 infiltration wells (Figure 1). Through this recharge, the capacity of the well field is increased and a hydraulic barrier against potentially contaminated water from the former waste disposal site, Herdern, is formed. The dimension of the solute plume spreading from the Herdern site westwards was determined by Jäckli (1992). In 2001, Kaiser analyzed the origin of the water drawn in the four horizontal wells (Kaiser 2001). Based on these studies the water in the four drinking water wells can be assigned either to river bank filtrate (low electrical conductivity (EC) close to that of the river Limmat) or to city water (groundwater coming from below the city area and possibly passing through the Herdern site, characterized by elevated EC). A detailed description of the Hardhof well field including technical data can be found in Bauser et al. (2010).

Central to this well field is the fact that the drinking water quality is influenced not only by pumping rates in the production wells but also to a considerable part by the artificial recharge. Although ES may be applied for the management of any well field this work focuses on well fields heavily influenced by artificial recharge.

For this study, daily measurements of hydraulic head in 85 locations were available for the years 1992 to 2011. The locations of the piezometers are given in Huber et al. (2011). Furthermore, the daily pumping rates of all wells and basins as well as daily discharge data of the rivers Limmat and Sihl, precipitation and climate data were available.

METHODOLOGY

An ES is a collection of logic rules mimicking expert knowledge. It is a widely applied decision support system (e.g., Shu-Hsien 2005). The principal layout of an ES follows a tree of if-then relationships, leading from a given situation to a decision. The algorithm follows human reasoning and therefore it is intuitive and easy to understand. In order to establish the if-then relationships, a knowledge base has to be built. In our case, this was done by analyzing the daily time series of model input and output data.

For the design of the ES we proceed in four steps:

- Phase 1: Modeling the well field.
- Phase 2: Building the knowledge base; expert knowledge is gathered, and the control and actuating variables that influence the control variable are determined.
- Phase 3: Setting up the rule base of the ES; the knowledge is organized in a set of if-then relationships leading from a given situation to a decision.
Phase 4: Identifying parameters and adjusting the rule base; the ES is iteratively calibrated on the design period and validated on a new set of data.

**Phase 1: Modeling of the well field**

The design of a new management strategy often cannot be undertaken in the real well field because flawless operation of the well field cannot be guaranteed during the design phase. Therefore, a model of the well field is needed which reproduces the relevant processes of the real well field with an appropriate accuracy over the management horizon.

In the present case, drinking water quality has to be maintained at an appropriate level. The relevant information for management is the distribution of hydraulic head in the aquifer and the quality of the water in the four drinking water wells. The management horizon in this case study is 1 day. The well field was modeled with a transient three dimensional finite element groundwater flow model (SPRING® 3.4, Delta 2006) and calibrated with a modified pilot point method similar to Alcolea et al. (2008). Because of computational constraints (large number of nodes) and the large uncertainty of the input data for solute transport (i.e., the initial and boundary conditions), fully coupled flow and solute transport has not been feasible for the Limmat valley aquifer up to now. Bauser et al. (2010) therefore implemented a particle tracking tool which they coupled to the groundwater flow model to verify the origin of the water in the drinking water wells. Particles are tracked back from the four horizontal wells along a quasi-stationary flow field until they either pass the virtual boundary between Hardhof and city area (along Highway A1 in Figure 1) or it is clear that the particle path lines do not lead to the city area, hereafter called fraction of city water. They further showed that the fraction of city water can be used as an indicator for water quality in the drinking water wells. A detailed description of the particle tracking tool is given in Bauser et al. (2010). More details on the model can be found in Hendricks Fransen et al. (2011). We have used their deterministic central model in the present study.

**Phase 2: Building of the knowledge base**

A throughout process oriented understanding of the system was necessary for the modeling of the system. In this phase, knowledge about the system is organized in a control oriented manner: the control variable as well as the relevant disturbances are determined. In the case of well field management, the control variable can be the water quality produced or the pumping rates in the wells. In this work we focus on the drinking water quality. Relevant disturbances with regard to drinking water quality in a well field are processes which dominate the flow field (i.e., recharge and abstraction of groundwater) or sources of pollution (e.g., periodicity or threshold behavior). In order to be controllable, the disturbances have to be known. Unknown or minor disturbances are treated as uncertainties of the model.

Figure 2 shows the concept of a control oriented system description where \( \Sigma \) denotes the model of the system to be managed. The simulated water quality (the output of the model \( \Sigma \)) is fed back to the input of the ES. We will refer to it as the feedback variable.

Already, the major sources and sinks of the system are known from phase 1, the modeling of the system. Delay times between the disturbances and the control variable may be found through an analysis of Pearson’s cross-correlation between each major disturbance and the control variable. The delay times of the system should be verified for example with an estimation of travel times in the aquifer and taken into account in the input of the ES.

**Phase 3: Setting up of the rule base**

In order to cope with multiple inputs to the ES we propose combining the impacts of the disturbances and the feedback variable on the simulated water quality in one artificial variable. This variable contains the
information about how probable a reduced water quality in the current management horizon is. First, each of the
$i = 1, \ldots, N$ disturbances and the feedback variable are attributed a weight $w_i$. As a first guess this weight corre-
sponds to the maximum of Pearson’s cross-correlation. The weight $w_i$ accounts for the relative impact of the
input variable $i$ to the ES on the output of the model. The following steps are accomplished for each manage-
ment horizon:

- For a first guess, the input variables are classified approximately according to their 30th, 60th and 90th
  percentile and attributed a value $v_i$. For example invalid measurements are classified as ‘no measurement’ and
  attributed the value 0. Values below the 30th percentile are classified as low and attributed the value 1. See
  Table 1 for an overview of the classes and the corre-
  sponding values.
- The values of the input variables are then multiplied with the weights of the variables and summed up over all input
  variables $\sum(w_iv_i)$. The resulting value is a measure for the probability of an impairment of the drinking water quality
  in the given time step.
- The resulting value is again classified into low, medium, high and extreme, and an infiltration scheme
  is attributed. The infiltration scheme determines the amount for artificial recharge and (if necessary) its distri-
  bution to the artificial recharge infrastructure(s). For the determination of the infiltration scheme, expert
  knowledge again plays a central role. The artificial recharge modifies the flow field of the groundwater and the expert has to understand in which way it has
to be changed in order to maintain an appropriate water quality.

Phase 4: Identification of parameters

The well field management is now simulated with the ES management and the parameters of the ES (weights,
values of the classes) are adapted manually where needed. If the calibration of the ES is satisfactory, a validation
period is simulated.

RESULTS AND DISCUSSION

An ES was designed according to the methodology described above to manage the water quality in the Hardhof
well field. The simulation of the well field management is based on historical data. The following paragraphs describe
the design phases 2 to 4 (phase 1, the modeling of the well field, is described in detail in Bauser et al. (2010), and
Hendricks Franssen et al. (2011)).

Phase 2: Building of the knowledge base

The goal of the ES is to maintain a high water quality in the Hardhof well field by adapting the amount and distri-
bution of water infiltrated in basins I to III and in the injection wells S1 to S12. Due to its location close to the
southern boundary of the well field, well C is the most vulnerable one from which to withdraw city water.
This is confirmed by the measurement of EC in the

<table>
<thead>
<tr>
<th>Class</th>
<th>Value</th>
<th>Initial guess</th>
<th>Adjusted values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$Q_c(t)$</td>
<td>$f_r(t-1)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[m$^3$/d]</td>
<td>[-]</td>
</tr>
<tr>
<td>No value</td>
<td>0</td>
<td>$\leq 0$</td>
<td>$\leq 0$</td>
</tr>
<tr>
<td>Low</td>
<td>1</td>
<td>$\leq 2'500$</td>
<td>$\leq 0.05$</td>
</tr>
<tr>
<td>Medium</td>
<td>10</td>
<td>$\leq 6'500$</td>
<td>$\leq 0.10$</td>
</tr>
<tr>
<td>High</td>
<td>100</td>
<td>$\leq 12'000$</td>
<td>$\leq 0.20$</td>
</tr>
<tr>
<td>Extreme</td>
<td>1,000</td>
<td>$&gt; 12'000$</td>
<td>$&gt; 0.20$</td>
</tr>
<tr>
<td>Weight $w$</td>
<td>–</td>
<td>0.5</td>
<td>0.8</td>
</tr>
</tbody>
</table>
drinking water wells, with highest values in well C. For the ES, we thus only use the fraction of city water in well C $f_{r_C}$ from the model output. For the decision about the infiltration scheme of the next time step the fraction of city water of time step $t - 1$ is fed back to the input of the ES. The desired water quality is zero percent of city water in well C.

In this study we have assumed that the Herdern deposit constitutes a constant source of pollution. Therefore, we can concentrate on the flow field as the major influence on the quality of the drinking water. The factors influencing the flow field are recharge and abstraction from the aquifer. The major abstractions in the Hardhof area are effected by the Zurich water works. The rates are determined by the demand of the city of Zurich and regarded as known disturbances in this work. So is the distribution of the abstraction in the four drinking water wells. Recharge has two important components in the Hardhof area: infiltration from the river Limmat and artificial recharge. The rate of infiltration from the river depends highly on the water level in the river ($h_L$). The water level can be determined on a daily basis but it cannot be influenced by the Zurich water works, it is therefore a known disturbance. The artificial recharge in the Hardhof area is determined by the Zurich water works. It can be used by management to act on the system in order to maintain an appropriate drinking water quality. The degrees of freedom for management can be reduced from $12 + 3$ (12 injection wells + 3 recharge basins) to 6 because the 12 injection wells cannot be operated individually but in the following groups: group 1 contains S1 to S6, group 2 corresponds to well S7, group 3 contains S8, S9, and S10, and group 4 contains S11 and S12. Group 1 lies downstream of the well field and is never operated.

The cross-correlation analysis between EC measurements in the river Limmat, the horizontal wells, the infiltration infrastructures, and a representative piezometer in the city revealed the typical lag times (in days) between the measurement locations.

The following disturbance variables were found to have the highest correlation with EC in well C on a given day $t$: Abstraction in well C on day $t$, $Q_c(t)$, fraction of city water in well C on day $t - 1$, $f_{r_C}(t-1)$, and water level of the river Limmat on day $t-14$, $h_L(t-14)$. The corresponding correlations between the input variables and the water quality measure are given in Table 1 in the form of the weights $w$ of the variables (initial guess). The lag times correspond to the average travel times of the water found with tracer tests (e.g., Kaiser 2001).

**Phase 3: Setting up of the rule base**

In phase 2, two known disturbances and one feedback variable were identified and their influence on water quality was estimated (initial guess of the weights $w_i$ in Table 1). The initial estimates of the boundaries between the classes for each ES input variable are determined based on the cumulative frequency distribution of daily values from 2004 for each variable (Table 1). The initial estimate for the infiltration scheme is given in Table 2. The infiltration focuses on injection well groups 2, 3 and 4 and basins I, II and III.

An exemplary path through the ES algorithm for the computation of the infiltration scheme of day $t$ could look like this (the numbers refer to the initial guess of the parameters of the ES): the river water level on day $t-14$ was...
below 397 m a.s.l. It is therefore classified as low and attributed the value 1. This value is multiplied with the weight 0.2 yielding 0.2. Then, the abstraction rate of day \( t \) is classified as high, attributed the value 100, multiplied with the corresponding weight of 0.8 (yielding 80), and summed up with the result of the previous decision (yielding \( 80 + 0.2 = 80.2 \)). The same procedure is repeated for the fraction of city water in well \( C \) on day \( t-1 \). The resulting value is again classified. The class is usually higher or the same as the highest classified input variable. A classification of low means that the danger of attracting city water into the well is low and less infiltration is needed in order to maintain the hydraulic barrier.

In order not to allow a depletion of the aquifer, the minimum total infiltration rate is set equal to \( 1-1.5 \) times the total abstraction rate in the Hardhof well field on a given day, depending on the classification of the infiltration scheme \((Q_{\text{infiltration}} = f \cdot Q_{\text{abstraction}}, \text{with } f \in [1,1.5], \text{Table 2}).\)

**Phase 4: Identification of parameters**

The ES management is now simulated with historical data from 2004. Iteratively, the parameters of the ES are tuned in order to reduce the fraction of city water in the drinking water wells. Thereby, the visualizations of particle path lines for selected days are consulted. A significant reduction of the fraction of city water is achieved after three iterations. A first validation period in January 2005, however, exhibits a very low groundwater table which never occurred during 2004 and where the ES management does not perform satisfactorily.

Phases 2 to 4 are completed again with an extended knowledge base from January 1992 to January 2005. Although the statistics of the prolonged time series change, the delays between the input variables and the water quality remain the same. After three iterations of manual parameter tuning, the water quality computed with the ES management between January 2004 and February 2005 was again reduced considerably compared to the historical management scheme (Figure 3). The final adjusted parameters of the ES are given in Tables 1 and 2. The weights of the adjusted parameter set do not refer to the statistics of the knowledge base any more but have a very similar weight. As an alternative to the correlations, the input variables of the ES could be given the same weight as an initial guess.

**Figure 4** shows the particle path lines before (on the left) and after (on the right) tuning of the parameters. On the right side of Figure 4 no particle path lines computed with the ES management cross the boundary line between Hardhof area and city area, whereas this is not the case in the left hand figure. Accordingly, the fraction of city water is also smaller after the tuning of the parameters than before tuning. The figure further shows the particle path lines of the historical management where the fraction of city water was elevated (also reflected in the elevated EC measurement). Compared to the historical management, the
infiltration rate in basin III was increased by a factor of 2 using the adjusted ES. While with the historical management scheme, less water was infiltrated than abstracted for several days before 4 May 2004 (path lines for this day are depicted in Figure 4), the ES management infiltrates 50% more water than is abstracted during the same period and thus succeeds in maintaining the fraction of city water at a low level (see Figure 3).

The validation period was prolonged to 8 months from February to August 2005. The ES with an enlarged knowledge base and tuned parameters performed well during the validation phase (Figure 3). Except for two peaks of 3% of city water in well D in June 2004 and June 2005, all peaks of city water in the drinking water wells are reduced significantly with the ES management scheme. The two peaks are caused by particle path lines crossing the boundary line twice (see Figure 5). Even though the path lines pass through the potentially contaminated city area, the water quality in the well is not endangered since the water originates from one of the injection wells of the water works. The fraction of city water could be reduced by increasing the infiltration rates. This is not recommendable, however, because excessive infiltration could lead to water logging in the Hardhof area.

The ES management holds a comparison with the optimal control presented in Bauser et al. (2012). Even though the infiltration scheme found with the ES is not optimal, it nevertheless reduces the amount of city water in the drinking water wells to an acceptable level with a reasonable simulation time (because no iterative model runs are needed).

Applicability to other settings

The design procedure of the ES presented here may be applied to an arbitrary well field management problem, given that the problem can be solved. The essential limitation of the ES is the knowledge base. The management results can only get as good as the knowledge base is wide. If essential processes influencing the well field are unknown the application of the presented method yields unreasonable results.

A drawback of the ES is the need for manual adjustment of the parameters. The procedure is tedious and prone to conceptual errors (i.e., if the understanding of the system is not profound enough). A periodic updating of the knowledge base and the parameters of the ES with the newest data is recommendable for management application in a real well field. The implementation of the ES in a real well field is assumed to be straight forward because the rule base is intuitively understandable without background knowledge in control engineering, which enhances its acceptance among professionals operating the well fields.
Model uncertainty and unknown disturbances

In the presented methodology we do not consider model uncertainties and unknown disturbances. However, the model is assumed to yield a conservative estimate of the water quality (Bauser et al. 2010). So we base management decisions on a cautious estimate of the fraction of city water in the wells.

CONCLUSIONS

In this paper, the methodology for the design of an ES for well field management is presented. The applicability of the method was demonstrated in a study of the Hardhof well field.

It has been shown that, given that the main processes influencing the well field are known, the ES is an efficient alternative for well field management. We propose periodic updating of the knowledge base and the parameters of the ES to newly available data in order to maintain good performance.

Although the ES management is not optimal, it is close to it, as shown by a comparison with Bauser et al. (2010, 2012). Furthermore it features a small computation time and a simple, intuitive structure which are key advantages for implementation in a well field.

REFERENCES


Delta, H. 2006 Software SPRING® 3.4. Ingenieurgesellschaft GmbH, Parkweg 67, 58453 Witten, Germany.


First received 3 January 2012; accepted in revised form 12 March 2012