Water in Celtic Countries: Quantity, Quality and Climate Variability (Proceedings of the Fourth InterCeltic Colloquium on Hydrology and Management of Water Resources, Guimarães, Portugal, July 2005).
IAHS Publ. 310, 2007.

A new method for groundwater plume detection under uncertainty

LUÍS MIGUEL NUNES[1], MARIA DA CONCEIÇÃO CUNHA[2], LUÍS RIBEIRO[3] & JOÃO AZEVEDO[3]

1 *Faculty of Marine and Environmental Sciences, University of Algarve, Campus de Gambelas, 8000 Faro, Portugal*
lnunes@ualg.pt

2 *Civil Engineering Department, University of Coimbra, Pinhal de Marrocos, 3030 Coimbra, Portugal*

3 *Instituto Superior Técnico, Lisbon Technical University, Av. Rovisco Pais, 1096 Lisboa Codex, Portugal*

Abstract Groundwater contamination plume characterization is a very hard task to perform and usually requires a large number of sampling sites. In this article a method for optimizing a monitoring network for plume detection and delimitation is proposed. It is assumed that an extensive sampling campaign has already been conducted and that only a few sampling sites should be included in the optimal monitoring network. The objective function incorporates the prior knowledge about concentration variability in the form of its density function, and also a measure of spatial coverage (space-filling method) in order to best distribute the stations over the field. The method was applied to a synthetic case study with 160 sampling locations, and a final optimal monitoring network with 40 stations was obtained. A simulated annealing optimization algorithm was used to solve this very difficult combinatorial problem, which has more than 8.6×10^{37} possible solutions.

Key words groundwater; monitoring; optimization; plume

INTRODUCTION

Groundwater monitoring network design and operation has become a subject of major concern in the last two decades. In fact, groundwater resources are very important natural resources that have not only been overexploited in some cases, but also submitted to increasing contamination of different origins (agriculture, landfill, effluent disposal, hydrocarbon leaks, etc.). Groundwater assessment, protection, remediation and containment of contamination are the keys to preserving these important resources. It has been recognized that monitoring networks play a significant role in the successful development of such tasks. The importance attributed to this subject has given rise to abundant contributions to the literature in the field. In general terms, designing and operating groundwater monitoring networks involve siting monitoring points and devising sampling strategies. An analysis of the literature shows that four types of approach have been used to deal with such problems. The first attempts were centred on the application of geostatistical methods (Rouhani, 1985; Rouhani & Hall, 1988). That period also saw the extensive use of simulation approaches (Massmann & Freeze, 1987; Meyer & Brill, 1988). Later, transfer function methods start to emerge (Andricevic & Foufoula-Georgiou, 1991). Optimization

methods are among the most popular approaches applied nowadays, even if the earliest attempts at using such techniques date back to the 1980s (Olea, 1984; Hudak & Loaiciga, 1992; McKinney & Loucks, 1992; Meyer *et al*., 1994; Lee & Ellis, 1996; Reed *et al*., 2000; Shlomi *et al*., 2004). Different objective functions and constraints have been incorporated into the optimization models. Various examples of objectives in this field are: minimization of costs (installation, operation and maintenance, environmental costs, etc.; Reed *et al*., 2001); minimization of the number of wells (Meyer *et al*., 1994); minimization of the volume of contaminated groundwater (Mahar & Datta, 1997); maximization of the number of detections (Meyer *et al.*, 1994); and minimization of pollutant concentration at given points (Shlomi *et al*., 2004). In several studies, the objective of cost minimization is replaced, for the sake of simplicity, by surrogate goals such as the minimization of the error of kriged values of concentration (Lee & Ellis, 1996). In such circumstances, constraints representing the maximum budget allocated to the monitoring network are introduced either directly or in the form of a maximum number of wells. The optimization models in this field are usually mixed integer nonlinear models, given the nonlinear form of the objective function and of some constraints, and they contain discrete variables, given the combinatorial nature of the decision variables. This means that classic optimization methods, such as linear, nonlinear or integer linear programming, are not appropriate to solve those models. Modern heuristics like simulated annealing, genetic algorithms and tabu search, have proved successful in solving problems with the mathematical characteristics mentioned. The work by Lee & Ellis (1996) compares eight heuristic algorithms used to define the optimal design of a groundwater monitoring network, and the conclusions pointed to the superior performance of the simulated annealing and the tabu search heuristics.

Groundwater contamination plumes characterization is a very difficult task to perform, usually requiring a large number of sampling sites. In this article a method for optimizing a monitoring network for plume detection and delimitation is proposed. It is assumed that an extensive sampling campaign was already conducted, and that only a few sampling sites must be included in the optimal monitoring network. This extensive monitoring campaign was generated by geostatistical conditional simulations of the permeability field and by running a flow and transport model on the simulated fields. The objective function incorporates the prior knowledge about concentration variability in the form of the relative number of detections at each location (number of detections divided by the number of simulations), and also a measure of spatial coverage (space-filling method), in order to best distribute the stations over the field. No statistical assumption is made on the data, so the number of relative detections is considered a variable and not a probability, hence permitting several mathematical operations, not possible with probabilities.

The method was applied to a synthetic case-study with 160 sampling locations, and a final optimal monitoring network with 40 stations was obtained. These results are compared with those obtained by accounting only for the relative number of detections at each location (this empirical approach in a statistical framework would correspond to the probability of detection). Simulated annealing optimization algorithm was used to solve this very difficult combinatorial problem, which has more than 8.6×10^{37} possible solutions).

METHOD

Stochastic simulations

A groundwater flow scenario was built in ASMWin (Chiang *et al.*, 2003), simulating a steady-state condition. A rectangular non-confined aquifer was discretized in 40 × 15 cells of 10 m size. The following homogeneous hydraulic parameters were used: hydraulic conductivity, K = 864 m day^{-1}, and porosity equal to 10%. The base of the aquifer is at 20 m. No pumping and no recharge situations were considered. Flow is from left to right. Boundary conditions of the first kind (Dirichlet type) prescribe the head value at the left side as 27 m and at the right side, 24 m. Figure 1 shows the aquifer grid model. An advective–dispersive solute transport situation was simulated using the following parameters: molecular diffusion coefficient = 5×10^{-10} m^2 s^{-1}; longitudinal dispersivity coefficient = 4.5 m, transverse dispersivity coefficient = 1.125 m. An instantaneous injection of 40 kg of pollutant was used. The total simulation time was set at 32 days. No retardation effects were considered. A total of 100 stochastic simulations of K were performed using the following parameters: mean value in $\log_{10}$ (in m s^{-1}) = –2; standard deviation = 0.5; correlation length/field width = 0.1 (in both X and Y directions). The generation of stochastic conductivity distributions is based on a lognormal distribution. Each matrix thus generated is one realization of an aquifer with the given statistical properties. Figure 1 shows pollutant particles locations, at the end of one of the simulations. Particles were then transformed into concentrations and allocated to each grid centre.

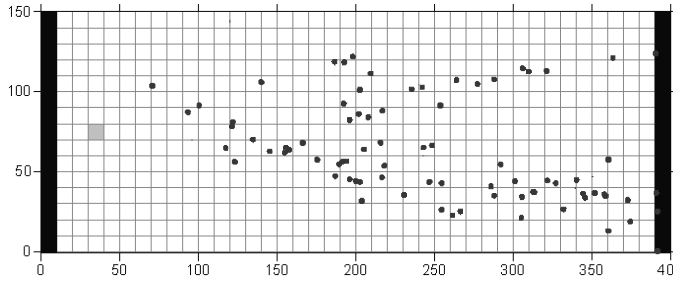

Fig. 1 Result from one of the transport simulations.

Optimization function and algorithm

The space-filling method was applied in monitoring network design by Morris & Mitchell (1995) and Royle & Nychka (1998), but has had very little application since then. In space-filling designs some criterion based on a metric is used to evaluate the goodness of a space-covering design. The most common criteria are those that use the following criterion:

$$C_{p,q}(D) = \left(\sum_{u \in C} d_p(x,D)^q \right)^{1/q} \quad (1)$$

which is the average of distances between candidate stations and the other stations

already included in the subset. The exponent q is usually > 0. One possible metric is:

$$d_p(x,D)=\left(\sum_{u\in D}\|x-u\|\right)^{1/p} \tag{2}$$

where when $p < 0$, $d_p(x, D)\rightarrow 0$ as x converges to a member of D. The coverage design is the subset of ω elements in D from the Ω elements in C, $D \subset C$, that minimize the criterion $C_{p,q}(D)$. It is clear here that in this criterion the only information needed is the spatial coordinates of the stations. This criterion forces the spatial distribution of stations to the most regular design—the network must cover the domain such that the function of the metric is minimized. It is clear that if the values of p and q are set to 1, the criterion reduces to the minimization of the Euclidean distance connecting all stations. However the aforementioned authors conclude that such setting does not lead to good spatial coverage. A second function is still needed if one wants also to include information from previous measurements. Consider a measured value of some variable under study, $g(x_{i,j})$, at a location $x_{i,j}$, a given reference value (e.g. legal limit, or analytical detection limit), g_{ref}, the relative number of detections is determined by:

$$\begin{aligned} a_s(x_{i,j}) &= \begin{cases} 1 & if \quad g_s(x_{i,j}) \geq g_{ref} \\ 0 & \text{otherwise} \end{cases} \quad s=1,\ldots,m \\ r(x_{i,j}) &= \frac{1}{m}\sum_{s=1}^{m} a_s(x_{i,j}) \end{aligned} \tag{3}$$

with m the total number of conditional simulations. The value of $r(x_{i,j})$ is considered a variable, not a probability, therefore allowing mathematical operations, like additions, that would make no sense in a statistical framework. The statistical perspective would require the calculation of the conditional probabilities of finding a concentration value at a location given the concentrations observed in neighbouring locations. This problem is well solved in the context of the Bayesian maximum entropy methods (Christakos, 1990; Christakos & Li, 1998; D'Or *et al.*, 2001). The use of $r(x_{i,j})$ reflects the empirical need to include in the monitoring network those stations that detected contamination more often. However, in most practical cases it is also necessary to consider stations in areas where contamination events are less frequent (in the borders of the contamination plume). To do so it is proposed that the optimized network should have a maximum spatial coverage, by minimizing the coverage criterion, and also that the prior information be considered. Hence, the unconstrained problem may be stated as:

$$\min \quad \frac{C_{p,q}(D)}{\sum_i \sum_j r(x_{i,j})} \tag{4}$$

The method proceeds by: (i) selecting a subset of stations, ω, from the large set, Ω; (ii) calculate the value of the objective function; (iii) select or discard the design according to the simulated annealing algorithm; (iv) swap one station between the large set and the design set; (v) repeat steps (i)–(iv); (vi) stop the search of new designs according to the simulated annealing algorithm and present the final solution. A simulated annealing (SA) heuristic optimization algorithm was implemented in Fortran 90 and thoroughly tested. The implementation followed the description presented in

Nunes *et al.* (2004a). Simulated annealing has had many applications in water management studies and in related fields, and proved to be a good optimization algorithm for difficult combinatorial problems. More information on the basics of simulated annealing may be found in Kirkpatrick *et al.* (1983) and Aarts & Korst (1990).

RESULTS AND DISCUSSION

The results shown here were obtained by considering that $p = -3.0$ and $q = 2.0$, and the initial number of monitoring stations, Ω, equals 160, while the final network should have only 40 stations ($\omega = 40$). Hence the combinatorial optimization problem is one of finding the "best" solution according to the objective function, by selecting iteratively networks with 40 stations out of the 160 available ones. If the problem had had to be solved exhaustively by testing all the possible combinations of stations it would require, with the 2GHz Pentium PC used in the calculus, more than 6.9×10^{29} years. The solution presented here, which is one of the best out of 10 runs, took almost seven hours to obtain. This shows the great advantage of using more efficient methods like SA. The disadvantage is that one will never know if the good quality solution obtained by the algorithm is the global optimum (the minimum of the minima), because optimality is only guaranteed in an almost infinite number of iterations (almost infinite time). Figure 2 shows the convergence curve for one of the optimization runs, where convergence to a minimum as temperature decreases is evident. A large jump from high objective function values at high temperatures, to very low values, is also evident. Simulated annealing is a method based in the physics of annealing processes when slowly cooling a material into its crystallized form; it is natural that it may show some characteristics also common in physics, like supersolid transition: a very fast drop in some characteristic of the material when cooled below some critical temperature (Andreev & Lifshitz, 1969; Chester, 1970; Kim & Chan, 2004). This behaviour is very uncommon in simulated annealing results, but is frequent in problems using space-filling functions. This may be related to the fractal characteristic of most space-filling natural networks, for which there seems to be a constant factor relating the dimension of the network (distance between stations or number of branches) to its coverage area (for more on this, see e.g. Rodríguez-Iturbe & Rinaldo (1997) and Sposito (1998) in hydrology, and West & Brown (2004) in life sciences). It may be here that the critical temperature is related in some way with specific spatial arrangements that best approximate the factorial nature governing natural networks. This may become an important indication as to which solutions may be the best candidates to constitute a smaller solution space, if one can devise a method to identify a set of solutions with the correct fractal nature (e.g. similar fractal dimension). No results on these lines are presented here, though.

Our results showed that if the empirical approach of including stations in the final network based on the highest relative number of detections is used, the resulting network will tend to be too concentrated in the centre of the plume. To obtain this network it is sufficient to order the stations by decreasing value of the relative number of detections and to choose the first 40 stations.

If the objective function (4) is used, the network is much more regularly distributed in space, covering not only the centre of the plume, but also areas where contamination

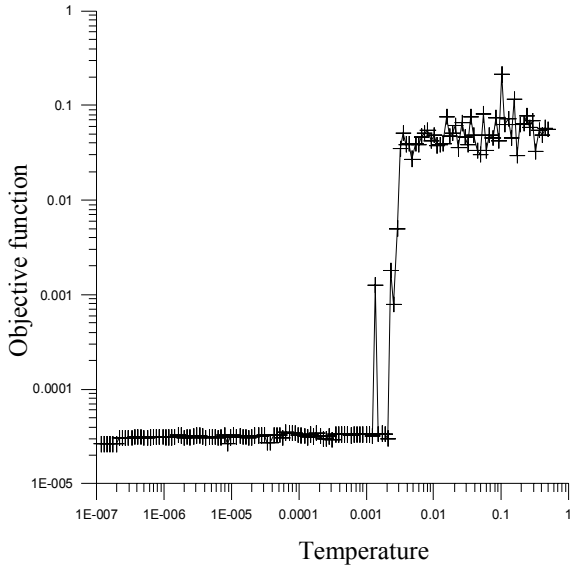

Fig. 2 Simulated annealing convergence curve.

levels are very low (see Figs 3 and 4). This is usually a much more realistic objective when dealing with the detection of contamination events from point or areal sources. Other methods have been proposed along this line using geostatistics and exploration costs (Nunes *et al.*, 2004a,b) with similar results. The advantage of the space-filling/relative number of detections method over the others relies in its speed, because no variogram fitting is needed, and because it is not constrained by ergodicity assumptions (which may not apply directly to contamination plumes due to evident three-dimensional concentration trends).

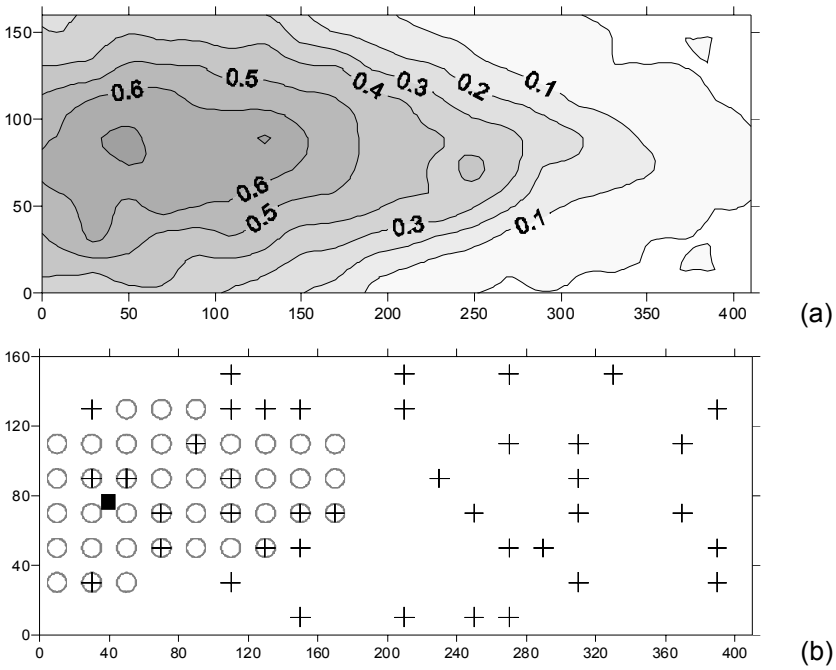

Fig. 3 (a) Relative number of detections. (b) Resulting monitoring networks: (+) – with the proposed objective function; (o) – with the empirical approach.

CONCLUSIONS

In this article a new objective function is proposed to design monitoring networks when the objective is to reduce the number of monitoring stations. Its advantage over other methods relies in its simplicity and speed, as well as in an apparently more reality-based approach, making it easier to explain to decision makers.

It was shown that the method generates monitoring networks that are more regularly distributed in space, allowing the monitoring of both high and low concentration areas. When compared to a more empirical approach, such as using the stations where contamination was detected more often, it is clearly much less clustered in the centre of the plume.

Some speculations were made regarding: (i) the relationship between the design of monitoring networks with space-filling objective functions by simulated annealing and supersolid characteristics; (ii) the fractal nature of natural networks and the possibility of using that information to reduce the solution space.

Acknowledgements This work was funded by Grant POCTI/ECM/36289/2000 from the Portuguese Fundação para a Ciência e Tecnologia.

REFERENCES

Aarts, E. & Korst, J. (1990) *Simulated Annealing and Boltzman Machines*. John Wiley & Sons, New York, USA.

Andreev, A. F. & Lifshitz, I. M. (1969) Quantum theory of defects in crystals. *Sov. Phys. JETP* **29**, 1107–1113.

Andricevic, R. & Foufoula-Georgiou, E. (1991) A transfer function approach to sampling network design for groundwater contamination. *Water Resour. Res.* **27**(10), 2759–2769.

Chester, G. V. (1970) Quantum theory of defects in crystals. *Physical Review A* **2**, 256–258.

Chiang, W.-H., Kinzelbach, W. & Rausch, R. (2003) *ASMWin – Aquifer Simulation Model for Windows* 6.0. Zurich: Institute of Hydromechanics and Water Resources Management (IHW) – Swiss Federal Institute of Technology: Zurich, Switzerland.

Christakos, G. & Li, X. (1998) Bayesian maximum entropy analysis and mapping: a farewell to kriging estimators? *Math. Geol.* **30**(4), 435–462.

Christakos, G. (1990) A Bayesian/maximum-entropy view to the spatial estimation problem. *Math. Geol.* **22**(7), 763–777.

D'Or, D., Bogaert, P. & Christakos, G. (2001) Application of the BME approach to soil texture mapping. *Stochastic Environ. Res. Risk Assess.* **15**, 87–100.

Hudak, P. F. & Loaiciga, H. A. (1992) A location modeling approach for groundwater monitoring network augmentation. *Water Resour. Res.* **28**(3), 643–649.

Kim, E. & Chan, M. W. H. (2004) Probable observation of a supersolid helium phase. *Nature* **427**(6971), 225–226.

Kirkpatrick, S., Gellat, C. D. Jr & Vecchi, M. P. (1983) Optimization by simulated annealing. *Science* **220** (4598), 671–680.

Lee, Y. M. & Ellis, J. H. (1996) Comparison of algorithms for nonlinear integer optimization: application to monitoring network design. *J. Environ. Engng* **122**(6), 524–531.

Mahar, P. S. & Datta, B. (1997) Optimal monitoring network and ground-water-pollution source identification. *J. Water Resour. Plan. Manage. ASCE* **123**(4), 199–207.

Massmann, J. & Freeze, R. A. (1987) Groundwater contamination form waste management sites: the interaction between risk-based engineering design and regulatory policy: 1. Methodology. *Water Resour. Res.* **23**(2), 351–367.

McKinney, M. A. & Loucks, D. P. (1992) Network design for predicting groundwater contamination. *Water Resour. Res.* **28**(1), 133–147.

Meyer, P. & Brill, E. D. Jr (1988) A method for locating wells in a groundwater monitoring network under conditions of uncertainty. *Water Resour. Res.* **24**(8), 1277–1282.

Meyer, P., Valocchi, A. J. & Eheart, J. W. (1994) Monitoring network design to provide initial detection of groundwater contamination. *Water Resour. Res.* **30**(9), 2647–2659.

Morris, M. D. & Mitchell, T. J. (1995) Exploratory designs for computational experiments. *J. Stat. Plan. Inference* **43**, 381–402.

Nunes, L. M., Cunha, M. C. & Ribeiro, L. (2004a) Groundwater monitoring networks optimization with redundancy reduction. *J. Water Resour. Plan. Manage. ASCE* **130** (1), 33–43.

Nunes, L. M., Cunha, M. C. C. & Ribeiro, L. (2004b) Optimal space-time coverage and exploration costs in groundwater monitoring networks. *J. Environ. Mon. Assess.* **93**(1–3), 103–124.

Olea, R. A. (1984) Sampling design optimization for spatial functions. *Math. Geol*. **16**(4), 365–391.

Reed, P., Minsker, B. & Goldberg, D. E. (2000) Designing a competent simple genetic algorithm for search and optimization. *Water Resour. Res*. **36**(12), 3757–3761.

Reed, P., Minsker, B. & Goldberg, D. E. (2001) A multi-objective approach to cost effective long-term groundwater monitoring using an elitist nondominated sorted genetic algorithm with historical data. *J. Hydroinfo.* **3**(2), 71–89.

Rodríguez-Iturbe, I. & Rinaldo, A. (1997) *Fractal River Basins, Chance and Self-Organization*. Cambridge University Press, USA.

Rouhani, S. & Hall, T. J. (1988) Geostatistical scheme for groundwater sampling. *J. Hydrol*. **103**, 85–102.

Rouhani, S. (1985) Variance-reduction analysis. *Water Resour. Res*. **21**, 837–846.

Royle, J. A. & Nychka, D. (1998) An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Comp. Geosc.* **24(**5), 479–488.

Shlomi, S., Ostfeld, A., Rubin, H. & Shoemaker, C. A. (2004) Optimal groundwater contamination monitoring using pumping wells. In: *World Water Congress 2004*, vol. 138, 104.

Sposito, G. (ed.) (1998) *Scale Dependence and Scale Invariance in Hydrology*. Cambridge Univ. Press, Cambridge, UK.

West, G. B. & Brown, J. H. (2004) Life's universal scaling laws. *Phys. Today* **57**(9), 36–42.