Identification of the unknown pollution source in the Alsatian aquifer (France) through groundwater modelling and Artificial Neural Networks applications

Philippe ACKERER**, Maria Laura FODDIS*, Augusto MONTISCI*** and Gabriele URAS *.

* Department of Land Engineering (DIT)– Section of applied Geology and applied geophysics, University of Cagliari, via Marengo 3, 09123, Cagliari, ITALY (E-mail: laura.foddis@tiscali.it, urasg@unica.it).
* Laboratory of Hydrology and Geochemistry of Strasbourg (LHyGeS), University of Strasbourg, 1, rue Blessig, 67084, Strasbourg Cedex, FRANCE (E-mail: ackerer@unistra.fr).
** Department of Electrical and Electronic Engineering (DIEE), University of Cagliari, via Marengo 3, 09123, Cagliari, ITALY (E-mail: amontisci@diee.unica.it).

Abstract

Groundwater is the main source of drinking water and it has a vital importance in developed and developing countries. Computational models of groundwater are required to ensure its suitable management. Such models are useful to design remediation strategies in polluted aquifers.

This work aims at studying the spreading of a dangerous chemical - carbon tetrachloride (CCl4) - that contaminated a part of the Alsatian aquifer (France) because of a tanker accident in 1970. The exact amount of the chemical infiltrated is unknown and this constitutes the main issue for its individuation and remediation.

The purpose of this study is to define the behaviour of the unknown pollution source in terms of temporal variations, injection rates and duration of the activity.

An Artificial Neural Network (ANN) has been developed to identify the characteristics of the source and to solve the groundwater inverse problem: on the basis of known contamination concentrations data in pumping wells, the pollution source temporal evolution is reconstructed. The ANN learns to solve a problem by developing a memory, associating a large number of input patterns examples, with resulting set of outputs or effects. ANN are characterized by a flexible structure capable of approximating almost al input-output relationships.

To develop an ANN, it is necessary to generate a data set of patterns for training, validation and test procedure. In the case studied, the simulation models of solute transport in saturated groundwater flow are generated considering different scenarios of the source behaviour. These models examples are created with the software TRACES (Transport or RadioActiver Elements in the Subsurface) developed at the IMFS (Fluid and Solid Mechanics Institute) of Strasbourg "HOTEIT and ACKERER (2003)". Conceptual model and model design of carbon tetrachloride pollution in the Alsatian aquifer has been the subject of various studies developed by the Institute de Mécaniques des Fluides et des Solides of Strasbourg. The numerical simulation model used to generate the necessary patterns for the ANN is based on data measured between 1970 and 2004.

Keywords

Groundwater pollution, pollution source identification; groundwater modelling, Artificial Neural Networks.

INTRODUCTION

Surface water and groundwater pollution is a major issue in the global context that involves both developed and developing countries. Groundwater is an important resource for the production of drinking water. Its control is of extreme importance to protect global health. This implies the necessary development of an effective protection and the monitoring of key zones, especially in those areas where the geological characteristics of the soils strata allow relatively easy penetration of anthropogenic pollution into the groundwater. A sensible management aimed at protecting the groundwater quality and at safeguarding the groundwater resources has consequently a vital importance for life support systems.

This work focuses on groundwater resources contaminations. In this field, it should be underlined that in some cases, pollution may results from contamination whose origins are generated at different times and places where these contaminations have been actually found. To tackle such situations of pollution, it is necessary to develop techniques that allow identifying unknown contaminant sources behaviour in time and space. In general, the identification and delineation of the source of a contaminant plume is important for improving subsurface remediation and site management decisions in many contaminated groundwater sites.

The problem of identifying an unknown pollution source, based on known contaminant concentrations measurement in the studied areas, is part of the broader group of issues, called inverse problems. Various studies dealing with the resolution of inverse problems in groundwater contaminations have been recently carried out by several authors using ANN:, "TAPESH K AJMERA and A.K.RASTOGI (2008)", "SINGH R.M. and DATTA B. (2007)" "SISH et al. (2004)", "IQBAL and GUANGBAI (2003)", "A.FANNI et al (2002)", "MAHAR P.S. and DATTA B. (2000)", "ZIO (1997)".

The purpose of this study is to define the behaviour of an unknown pollution source that, because an accident in 1970, has polluted with carbon tetrachloride (CCl4), one of the largest aquifer in Western Europe and main sources of drinking water in the Alsace Region (France): the Alsatian aquifer "STENGERA A. and WILLINGERB M. (1998)".

The pollution source behaviour at the accident location is unknown. The objective of this work is to identify this unknown pollution source in terms of temporal variations, injection rates and duration of activity. To solve the inverse problem related to the case of the Alsatian aquifer contamination, we will try to describe the unknown pollution source behaviour using the ANN approach.

General description of the Upper Rhine Graben and the Alsatian aquifer

The Upper Rhine Graben is a segment of the European Cenozoic rift system that developed in the north-western forelands of the Alps. It is extended over 300 km, from Basel (Switzerland) in the south to Frankfurt (Germany) in the north, with an average width of approximately 40 km. It is flanked, in the south, by the Vosges and Black Forest (Schwarzwald) mountains, to the west and the east, respectively" BERTRAND G. et al (2006)", Fig. 1.

Alsatian aquifer is located in the southern part of the Upper Rhine valley. The aquifer surface is over 3000 km^2 and contains a volume of alluvial about 250 billion m^3. It represents one of the largest fresh water reserves in Europe. The groundwater reservoir contains about 50 billion m^3 of water, with an annual renewal of 1.3 billion m^3. This large aquifer has a vital importance since it supplies to 75% of the drinking water requirements, 50% of the industrial water needs and 90% of the irrigation water needs in Alsace.

The Alsatian part of the Rhine aquifer has a surface length of 160 km and a maximum width of 20 km.

The groundwater reservoir is part of a complex hydrosystem, which includes frequent exchanges between the rivers and the aquifer which vary with the seasons, and are caused by the proximity between the surface and the groundwater. This aquifer is highly exposed to contamination from rivers "STENGER A. and WILLINGER. M. (1998)".

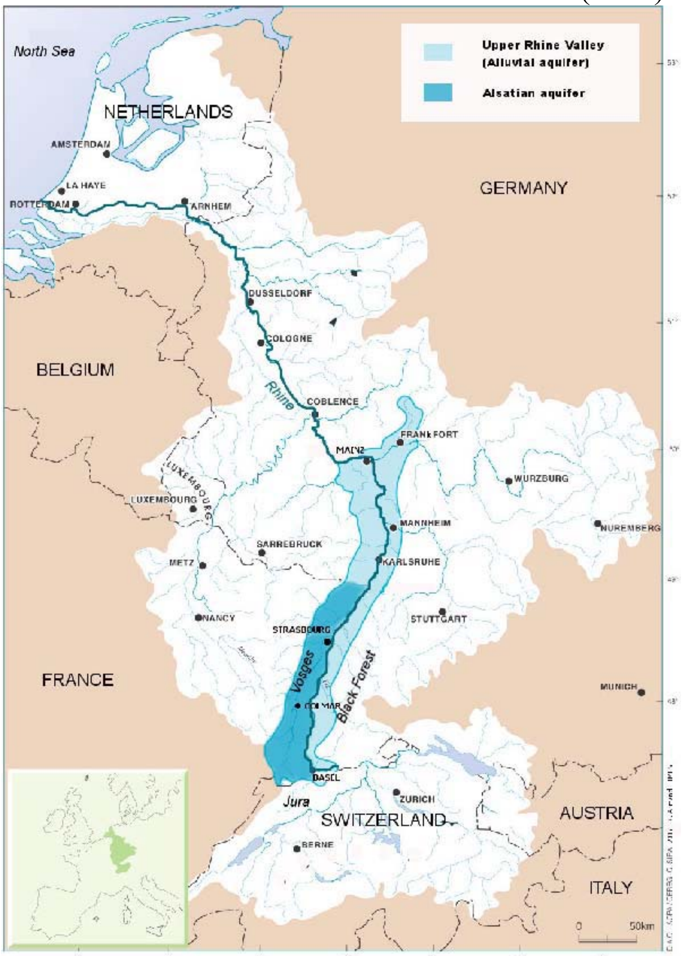

Figure 1: Geographic situation of the Alsatian aquifer.

Alsatian aquifer is an extensive alluvial aquifer with a layered structure composed by a random superposition of different alluviums (clay, sand fine to rough, gravels, coarse…). This permeable alluvial has a thickness of a few meters at the Vosgean edge, and 150 m to 200 m in the centre of the Rhine plain.

History of the pollution by CCl_4 in the aquifer

Alsace is a region where groundwater have a very important role in water supply. On the other side, Alsace is a heavily industrialised area, so the quality of groundwater is jeopardized by industrial contamination. This sort of pollution is harmful for drinking water.
In 1970 a tanker truck containing carbon tetrachloride (CCl_4) property of a Dutch company, capsized in the north of Benfeld (a small town located about 35 km south of Strasbourg, eastern France). In spite of the efforts of the firemen to control the spilling of the chemical, an important quantity of it could not be recovered. According to a note of SGAL (Service Géologique d'Alsace-Lorraine) of December 21th 1971 about 4000 litres of CCl_4 were spread in the area of the accident, infiltrating into the ground or disappearing by evaporation (Hamond, 1995). In 1991, the analyses carried out by BRGM (Bureau de Recherche Géologique et Minière) showed abnormal quantities of CCl_4 in the supplies of drinking water located downstream of Erstein (about 60 µg/l). These quantities exceeded the safe limits recommended by the World Health Organization (2 µg/l). This high level of CCl_4 concentration has caused a serious problem in the

region by contaminating the most important drinking water source in the area “ASWED and ACKERER (2008)”.

MATERIAL & METHODS

In the framework of this study, the inverse problem is solved using Artificial Neural Network (ANN) technologies (CARCANGIU et al, 2007). The ANN learns to solve a problem by developing a memory capable of associating a large number of example input patterns, with a resulting set of outputs or effects. In order to solve the inverse problem and to identify the unknown pollution source behaviour, an ANN is developed.

In order to train the ANN, it is necessary to generate a consistent data set of patterns (flux and transport model in porous media) for training, validation and test. To generate the patterns, we have used the numerical simulation software TRACES (Transport or RadioActiver Elements in the Subsurface) developed by “HOTEIT and ACKERER (2003)”, that combines the mixed-hybrid finite element and discontinuous finite element to solve the hydrodynamic state and mass transfer in the porous media.

Various scenarios of the contamination source behaviour have been constructed using an Alsatian aquifer 3D model, developed by the Fluid and Solid Mechanics Institute (IMFS) of Strasbourg. The model created by IMFS was calibrated using measured data of carbon tetrachloride concentration, that were collected during 12 years (from 1992 to 2004) and simulations were performed from 1970 to 2024.

This model has been the base to generate the ANN patterns. Using the TRACES software, various scenarios have been constructed, modifying the source behaviour characteristics in terms of temporal variation, injection rates and duration of activity.

Carbon tetrachloride modelling pollution in the Alsatian aquifer

Conceptual model and model design of carbon tetrachloride pollution in the Alsatian aquifer has been the subject of various studies developed by the Fluid and Solid Mechanics Institute of Strasbourg. The contaminated zone is enclosed within a 3D domain of 6 km width, 20 km length, and about 110 m depth. This aquifer domain is discretized using a 3D triangular prismatic grid with 25388 nodes and 45460 elements.

The domain is divided into 11 layers and each layer has a depth of about 10m. The numerical model TRACES is based on steady-state groundwater flow and transient solute transport in 3D heterogeneous media. The source is discretized into four layers in the vertical direction. The depth of the contaminated zone is about 35 m. The concentration of CCl_4 is imposed in the first eight mesh elements (each layer has two mesh elements) situated vertically beneath the accident location. The thicknesses of the layers are 16, 4, 5, and 5m from top layer at the groundwater surface to the lowest layer, respectively. The horizontal discretization is the same for all layers “ASWED and ACKERER (2008)”.

Artificial Neural Networks technologies

Over the past decades, Artificial Neural Networks (ANN) have become increasingly popular as a problem solving tool and have been extensively used as a predicting and forecasting tool in many disciplines. ANN has the ability to solve extremely complex problems with highly non-linear relationships. ANN’s flexible structure is capable of approximating almost any input-output relationships.

The ANN learns to solve a problem by developing a memory, associating a large number of example input patterns, with resulting set of outputs or effects. These are characterized by a flexible structure capable of approximating almost input-output relationships.

An artificial neural network consists of a number of interconnected processing element called neurons, which are logically arranged in two or more layers and interact with each other through weighted connections. The number of neurons in each layer and the number of layers in the network depend on the nature of the problem. There is no unique guiding theory for the suitable selection of the number of neurons and the number of layers. The scalar weights determine the nature and strength of the influence between the interconnected neurons. Each neuron is connected to all the neurons in the next layer.
The ANN consists of an input layer (where input data are presented to the network) and an output layer that holds the response of the network to the input. Additional intermediate layers, also termed as hidden layers, allow these networks to represent and compute complicated associations between inputs and outputs. The structure of a typical neural network is shown in Fig.1.

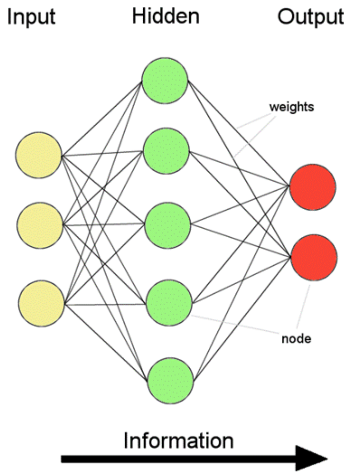

Figure 2: Structure of a typical neural network.

The ANN is a technology used for processing information whose characteristics are given the ability to: learning, generalization and approximation. An input is presented to the neural network and a corresponding desired or target response set at the output (supervised training). An error is composed from the difference between the desired response and the system output. This error information is fed back to the system and adjusts the system parameters in a systematic fashion (the learning rule). The process is repeated until the performance is acceptable. It is clear from this description that the performance hinges heavily on the data.

Pattern construction for training the ANN

A consistent data set of patterns, based on flux and transport model in porous media of carbon tetrachloride (CCl_4) contamination in the domain, is generated to train the ANN. To generate the patterns, we have used different scenarios of contamination source based on the 3D model of the Alsatian aquifer contamination. The estimation of the source term in the aquifer is influenced by the formation and fluid properties such as porosity, hydraulic conductivity and dispersivity coefficients.
Different source situations were generate by varying carbon tetrachloride concentration in the source area. 51 examples were constructed. For each scenario, to predict the temporal evolutions of the carbon tetrachloride concentration in the domain were calculating using the TRACES software: in total 45 pumping and observation wells were considered.
Firstly, taking into consideration the source activity, two hypothetical situations were designed: the activity of about 31 years and the activity of about 54 years. Finally, for each scenario, the source behaviour was changed in time and in every layer, being sometimes constant or variable. Each source scenario was calculated with the TRACES numerical simulator.

The examples obtained with TRACES from the study model consisted of matrix of size [m,n] of carbon tetrachloride concentration in time domain. Input matrices are composed by 4 columns: one column for each layer in the source. Output matrices are composed by 45 columns: one column for each well. In both matrices, rows represent time.
Matrices were too large to be processed through the ANN, requiring too many examples of inputs and thus a large network quite difficult to handle. For these reasons, it was necessary to elaborate a data pre-processing aimed at reducing the matrices size. Using Matlab, for each example, we have calculated a 2-D discrete Fourier transform (FFT). A preliminary feature extraction has been necessary such that both input and output layers of the neural network had a suitable dimension. To this end, among the frequency components of the FFT only the most significant have been considered, and the remaining ones have been set to zero. After that, only the not-null components have been kept to represent the time evolution of the contamination. The frequency components have been compared on the basis of their amplitude and that exceeding a prefixed threshold have been kept. The value of the threshold is determined by searching a crossover between the approximation of the acquired data and the dimension of the input and output matrices.
The approximation implies different issues in input and in output matrices. In fact, the aim of the proposed approach is to reconstruct the profiles of the source of pollutant, therefore a great precision is needed for the input, whereas the output has to be calculated on the basis of measurements, so that we only need outputs corresponding to different cases are distinguishable. On the other hand, while the input has only four time-varying concentrations, the output corresponds to 45 wells, therefore we can expect that a greater number of components are necessary to describe the output rather than the input. Several trials have been performed in order to select the suitable number of both input and output neurons, obtaining 9 components for the input and 102 components for the output.

Training and Inversion of the Neural Network model

The training of the ANN is a critical part of the proposed process. In fact a special care has to be spent to train the ANN in such a way that it is able to generalize the information contained in the training set. To this end, during the training phase the connection weights are modified in order to minimize the error on the training set, but at the same time the error is calculated also on a validation set, independent from the training set, and as the validation error begins to rise, the training process is interrupted.
This approach implies a validation set is available, but in this case the number of examples is too small and all of them are necessary to adequately describe the input-output relationship, therefore one cannot subtract a suitable subset (typically the 10-20% of the training set) to use it as validation set. To avoid this problem, the Leave-One-Out (LOO) technique has been adopted, which consists on performing as training as the examples of the training set, each time leaving one example out of the training set and using it as validation set. On the basis of the results obtained in all the trainings, the appropriate number of training iterations (epochs) is assumed. The second parameter that has to be defined for the training is the number of hidden neurons. Usually, this parameter is determined by means of a trial and error procedure, so that several trainings are performed assuming a growing number of neurons in the hidden layer. The better size of the hidden layer is that corresponding to the lower error in the validation set.
Once the training phase is completed, the inversion of the network can be performed. Knowing the output of the ANN, which derives from a set of measurements in the wells, the corresponding input can be calculated exploiting the method described in (CARCANGIU et al, 2007). By backward applying the pre-processing of the input data, the desired source profile is obtained.

CONCLUSIONS

This work mainly focused on an accident which took place in the North-East of France in 1970. The objective of this work is to use the ANN methodology to solve the inverse problem for the case of carbon tetrachloride pollution in the Alsatian aquifer. Various numerical model scenarios of the CCl_4 contamination in Alsatian aquifer are generated to train the ANN.
Complex and heterogeneous hydrogeology systems are extremely difficult to model mathematically. However, it has been proved that ANN's flexible structure can provide simple and reasonable solutions to various problems in hydrogeology.
An ANN is developed for the definition of the groundwater pollution source that contaminated the Alsatian aquifer. The ANN generated is composed by 9 neurons in the input layer to represent the source concentration behaviour and by 102 neurons in the output layer to represent the pumping well concentrations behaviour. After the training procedure, the ANN has allowed to generalize the information contained in the training set. The mean squared error calculated in the validation set is 0,025, low enough to allow the inversion of the ANN to solve the inverse problem.
ANN can be inverted and by knowing the output of the ANN, which derives from a set of measurements in the pumping and monitoring wells, the corresponding input can be calculated. By backward applying the pre-processing of the input data, the desired source profile is obtained.

REFERENCES

- ASWED, (2008). Modélisation de la pollution de la nappe d'alsace par solvants chlores, thèse de doctorat. Université Louis Pasteur, Institut de Mécanique des Fluides et des Solides, UMR CNRS 7507, Strasbourg, pp 40-42.
- BERTRAND G., ELSASS P., WIRSING G., LUZC A. (2006). Quaternary faulting in the Upper Rhine Graben revealed by high-resolution multi-channel reflection seismic. C. R. Geoscience 338: pp. 574–580
- CARCANGIU S., DI BARBA P., FANNI A., MOGNASCHI M.E., MONTISCI A., (2007). Comparison of multi objective optimisation approaches for inverse magnetostatic problems. COMPEL: Int. J. for Computation and Maths in Electrical and Electronic Eng., Vol. 26, N. 2, pp. 293-305.
- FANNI A., URAS G., USAI M., ZEDDA M.K., (2002). Neural Network for monitoring. Groundwater. Fifth International Conference on Hidroinformatics, Cardiff, UK, 1-5 July 2002, pp. 687-692.
- HAMOND, M. C., (1995). Modélisation de l'extension de la pollution de la nappe phréatique d'Alsace par le tétrachlorure de carbone au droit et à l'aval de Benfeld, mémoire pour le DESS sciences de l'environnement. Institut de Mécanique des Fluides et des Solides, Strasbourg.
- HOTEIT H. and ACKERER P. (2003). TRACES user's guide V 1.00. Institut mécanique des fluides et des solides de Strasbourg.
- IQBAL Z. and GUANGBAI C. (2003). Inverse modelling to identify nonpoint source pollution using a neural network, Taihu lake watershed, China, environmental hydrology. The Electronic Journal of the International Association for Environmental Hydrology, http://www.hydroweb.com.
- MAHAR P.S. and DATTA B. (2000). Identification of pollution sources in transient groundwater systems. Water Resources Management vol 14, pp. 209–227.
- SINGH R. M., DATTA B., JAIN A. (2004). Identification of unknown groundwater pollution sources using artificial neural networks. Journal of Water Resources Planning and Management, n° 130, 6, pp. 506-514.
- SINGH R.M. and DATTA B. (2007). Artificial neural network modelling for identification of unknown pollution sources in groundwater with partially missing concentration observation data. Water Resource Management, Volume 21, Number 3 / March, 2007, pp. 557–572

- STENGERA A. and WILLINGERB M. (1998). Preservation value for groundwater quality in a large aquifer: a contingent-valuation study of the Alsatian aquifer. Journal of Environmental Management, Volume 53, Issue 2, June 1998, Pages 177-193.
- TAPESH K AJMERA and A. K. RASTOGI (2008). Artificial neural network application on estimation of aquifer transmissivity. Journal of Spatial Hydrology, Vol.8, No.2.
- ZIO E. (1997). Approaching the inverse problem of parameter estimation in groundwater models by means of artificial neural networks. Progress in Nuclear Energy, Elsevier Science Ltd, Vol 31, n° 3 pp. 303-315.

2010_04_07_07-28-03_paper_mlfoddis.doc