Inverse modelling for flow and transport in porous media

Mauro Giudici Università degli Studi di Milano Dipartimento di Scienze della Terra Sezione di Geofisica Via Cicognara 7, I-20129 Milano, Italy Mauro.Giudici@unimi.it phone +39 02 503 18478 fax +39 02 503 18489

Introduction

The problem of parameter identification for flow and transport model in porous media is discussed in this communication. First, a general framework for the development and application of environmental models is discussed, following the review by Giudici (2001). Then the forward and inverse problems for discrete models are described in detail, introducing fundamental concepts (uniqueness, identifiability, stability, conditioning). The importance of model scales is reviewed and is shown its link with the stability and conditioning issues. Finally some remarks are given to the use of several independent sets of data in inverse modelling.

The general framework for the development and application of environmental models

The process of the development and application of an environmental model is sketched in figure 1. The complete description of this scheme can be found in Giudici (2001). Here I recall some issues.

Data on the physical system are fundamental at every step of model development and application. In fact they guide the definition of the goals of the model and of several other steps.

They are obviously very important for model calibration and validation, as well. Model calibration is the process of fixing the values of the model parameters, e.g., hydraulic conductivity or diffusivity, so that model predictions honor data (measurements and information). This is usually done with inverse modelling, i.e. solving an inverse problem.



Figure 1 – Scheme of the process of development and application of an environmental model.

In several applications, e.g. for geophysical exploration or medical imaging, the goal of the inverse problem is to obtain a map which shows the spatial and, in some cases, temporal distribution of some physical parameters. Among the parameters of interest, we recall seismic velocity for the seismic tomography (Nolet, 1987) and electrical resistivity for the electrical tomography (Griffiths and Barker, 1993; Loke and Barker, 1996).

Instead in environmental applications, this is only one part of the problem, because the results of the inverse problem must be used to solve forward problems that answer the specific questions of the final users. This implies that the whole process of model development and application must take into account the goals of the model and this is a must also for the model calibration.

We mentioned the importance of data. In environmental modelling we have to take care of some problems:

• Data include not only measurements of physical quantities (hard and soft data, see below), but also qualitative or descriptive information. We differentiate between hard and soft data: if we consider water contamination, the measurements of the concentration of chemical species in water samples are "hard data", since they are measurements of physical quantities strictly related with the process under study. The measurements of water conductivity are related to the presence of dissolved salts and are "soft data", since they are related to the process under study, but the relationship with pollutant concentration is the result of an interpretation. Furthermore we could have qualitative information on the presence of industrial or agricultural activities which produce or handle toxic species: this is a qualitative information, which is nevertheless very important for the development and application of a transport model, mainly for risk analysis. Other data and information come from the geological studies and are sometimes available as quantitative "soft" data, e.g. from the results of filed and laboratory tests, or qualitative information on the characteristics of geological formations.

• The environmental data are often affected by strong errors. In fact, it is very difficult to perform experiments under controlled conditions in the field and the acquisition of specific data for model development could be very expensive. As a consequence one is often forced to work only with already available data, which are often sparse or irregularly spaced, are lacking in a check of the measurement procedures and in the accuracy of measurement devices. Therefore the model calibration in environmental sciences has often to do with large uncertainties on the data.

A general abstract description of the forward and inverse problem for discrete models

This section is devoted to the introduction of a formalism that allows us to describe a general discrete model and is based on Giudici (2003).

We consider the 1-D stationary balance equation for single-phase flow in porous media as an example:

(1) $K_{i-1/2}(H_{i-1}-H_i) + K_{i+1/2}(H_{i+1}-H_i) = F_i$, i=1,...,N-1In particular we have assumed that the 1-D domain has been discretised with N+1 nodes, where the values of piezometric head $(H_i, i=0,...,N)$ are computed or assigned as Dirichlet boundary conditions. The physical parameters appearing in (1) are the so called internode conductivities $(K_{i-1/2}, i=0,...,N-1)$ and the source terms $(F_i, i=1,...,N-1)$ include all the sources and sinks in a cell.

We introduce the following notation:

- *u* is a *N*-dimensional array of parameters that describe the state of the system; it consists, e.g., of the values of pressure head or solute concentration in pore water at the nodes of a finite difference grid.
- *m* is a *M*-dimensional array of model parameters; it consists, e.g., of the values of internode conductivity or diffusion coefficient; among these parameters we include also those used to fix the boundary and initial conditions and the forcing terms, e.g. sources and sinks.

f(m,u)=0 are L state equations, e.g., the continuity equation (mass balance equation) for each cell or element of the discrete grid, like (1).

Definition (Forward problem) Given *m*, find *u* such that

(2) f(m,u)=0In explicit form u=g(m)

Some further notation.

d is a *D*-dimensional array of measurements, e.g., pressure head or solute concentration in pore water at some points of the domain

r(m,u,d) is a residual function, which links the model and the state parameters to measured data. Simple examples of residual functions are $r(m,u,d) = \sum_j c_j^{\ u} u_j - \sum_i c_i^{\ d} d_i$, which describes the interpolation process or r(m,u,d) = f(m,d), which is called equation error.

Definition (Inverse problem – direct formulation) Given d, find m^{dir} such that $r(m^{dir}, g(m^{dir}), d) = 0$ This definition implicitly assumes that we neglect:

- the modelling errors, including discretization and parameterization errors, are negligible, i.e., we assume that (1) describes perfectly the physical phenomenon under study;
- 2. the forward problem is exactly solved, e.g. no numerical approximation is introduced;
- 3. the measurement errors are negligible.

When some of these three conditions fail, we cannot guarantee that a solution to the inverse problem exist, in particular this is the case for overdetermined problems. Therefore we are led to the following new definition:

Definition (Inverse problem – indirect formulation) Given *d* and a norm ||, find *m*^{ind} such that $|r(m,g(m),d)| \ge |r(m^{ind},g(m^{ind}),d)| \forall m$.

Typical examples of norms used in practice are the l2 norm, which leads to the least-squares problem, and the l1 norm, which correspond to finding the minimum average absolute error. See, e.g., Menke (1989) for a description of the maximum-likelihood method; in the framework of that method, it is proven that the l2 norm corresponds to the assumption that errors (measurement and modeling errors)follow a gaussian distribution and that the l1 norm is more robust and can account for some outliers in the data set, i.e. few data with a great error.

With the formalism introduced above we can give some important definitions.

Definition (Uniqueness)

Let m and m' be two solutions of the inverse problem corresponding to the same data d; we say that the solution to the inverse problem is unique if m=m'.

Definition (Identifiability)

Let *m* and *m*' be two arrays of model parameters; the problem is identifiable if g(m)=g(m')

Definition (Stability)

Let *m* and *m*' be two solutions of the inverse problem corresponding to the data *d* and *d*', respectively; the solution to the inverse problem is stable if $|d-d'| \rightarrow 0$ implies $|m-m'| \rightarrow 0$.

Definition (Conditioning)

Let *m* and *m*' be two solutions of the inverse problem corresponding to data *d* and *d*', respectively, such that |d-d'| < C/m-m'| (Lipschitz condition). We say that the inverse problem is well conditioned if *C* is small

In the literature one can find several results for identifiability and uniqueness (see Giudici, 2001, for a review).

Here I spend few words for stability and conditioning.

Most of the methods applied for computing solutions to the inverse problems consists of applications of continuous operations, so that the stability of the inverse technique is guaranteed. Unfortunately many of the practical inverse problems are often ill-conditioned, so that small errors in the data prevent the effective computation of the inverse solution: this effect is often misinterpreted as an instability of the discrete inverse problem.

Space and time scales and their relation with the stability problem

The space and time scales at which a given process is modelled control the ability of the model to reproduce real features and the discretization of the domain for the numerical model. A necessary condition to be confident on model forecast is that the model scales are consistent with the space and time distribution of the field measurements.

Decreasing the grid spacing can improve the ability of the model to reproduce fine-scale features. However modelling physical processes with a fine grid requires a detailed knowledge of the physical system, which is a difficult task if the natural system is heterogeneous or anisotropic, as is often the case for geological formations or turbulent flows.

Therefore the modeller of the physical system must work with numerical techniques that provide "good solutions" even for grids with "large spacing".

"Large spacing" means that (1) the grid elements or cells could be larger than the optimal value obtained from the purely mathematical theory, but nevertheless consistent with the distribution of the available measurements and (2) the grid elements could be larger than the volume over which phenomenological laws are validated with laboratory or field experiments. For instance we validate Darcy's law with samples whose size is of the order of 1 m, whereas typical grid elements for the study of regional ground water flow could have side length of hundreds of meters.

"Good solutions" are those solutions which satisfy the physical principles that are at the base of the model, i.e. the conservation principles.

These remarks have important consequences for model calibration and therefore for the solution to the inverse problem, in particular for the inverse problem in the discrete case, which is the most important for applications. From the above remarks, the following questions arise.

- 1. Can we use field tests to infer the values of model parameters, even if the relevant scales are often different?
- 2. Stability is usually an overwhelming problem for continuous domains. Is it a problem also for discrete models? Or does numerical instability arise from ill conditioning? How does ill conditioning depend upon the grid size?

The first question is related to scaling problems, which are discussed in a comprehensive review by Renard and de Marsily (1997) for the study of dynamics of fluids in porous media. Scaling of physical parameters are discussed in relation with measurement scales by Cushman (1986) and Beckie (1996). Some discussions on this topic, connected with the inverse problem, can be found also in Giudici et al. (1995) and Ginn and Cushman (1992), whereas a real case study is presented by Bersezio et al. (1999).

A simple example shows why space and time scales are important for the stability and illconditioning problems. The example refers to modelling ground water flow, when typical values of hydraulic gradient are of the order of 10^{-3} .

The accuracy of measuring devices is of the order of 10^{-2} m if measurements are taken manually, whereas it is smaller by one order of magnitude (10^{-3} m) if automatic devices are used (e.g. pressure transducers). However measurement errors are due also to other sources, for instance the presence of pumping wells nearby the piezometers: these errors might be rather great, of the order of 1 m.

The spacing of the discrete grid can vary from 10 m (for applications to small areas, e.g. for site remediation) to 10^2 or even 10^3 m (for regional studies, e.g. ground water management). If we take into account these values of spacing, the head difference between adjacent nodes varies between 10^{-2} m (for local applications) and 1 m (for regional studies).

A direct comparison shows that measurement errors could strongly affect the results of inversion for a local application, whereas they could be less severe for large scale models.

The role of multiple data sets

Several works, dealing with both statistical and deterministic inverse techniques, have shown the importance of using several independent data sets (for both stationary and transient conditions) to reduce uncertainty in the model calibration and validation (Sagar et al., 1975; Carrera and Neuman, 1986; Ginn et al., 1990; Snodgrass and Kitanidis, 1998; Parravicini et al., 1995; Giudici et al., 1995; Vàzquez Gonzàlez et al., 1997) for ground water hydrology. This is very important because, in principle, the parameters of the discrete models are non local quantities, in the sense that they depend upon the space distribution in a region wider than the area to which the parameters refers to and on the flow direction. The cartoon in figure 2 shows the distribution of hydraulic conductivity of a real porous medium, estimated by a detailed facies analysis and assigning a single value of hydraulic conductivity to each sedimentary facies.



Figure 2 – Hydraulic conductivity distribution of an aquifer analog (Bersezio et al., 1999); hydraulic conductivity is represented with the grey logarithmic scale ranging from 10^{-20} (black) to 10^{-2} (white) m/s.

When the domain is discretised with a grid of cell, several cells will be characterized by non-isotropic physical properties, in general with lower conductivity in the vertical direction. The complete tensor characteristics cannot be obtained from a single flow situation, which will not give information on the direction perpendicular to the actual flow direction. Therefore, parameters obtained from data sets corresponding to different flow situations are more confident and, as a consequence, the use of several independent sets of data improves the reliability of model forecasting.

References

- Beckie R, Measurement scale, network sampling scale, and groundwater model parameters, Water Resour Res., 32, 65-76, 1996.
- Bersezio R., Bini A. and Giudici M., Effects of sedimentary heterogeneity on groundwater flow in a quaternary pro-glacial delta environment: joining facies analysis and numerical modeling, Sedimentary Geology, 129, 327-344, 1999.
- Carrera J. and Neuman S. P., Estimation of aquifer parameters under transient and steady–state conditions: 3. application to synthetic and field data, Water Resour. Res., 22, 228-242, 1986.
- Cushman, J. H., On measurement, scale, and scaling, Water Resour Res., 22(2), 129-134, 1986.
- Ginn T. R. and Cushman J. H., A continuous-time inverse operator for groundwater and contaminant transport modeling: model identifiability, Water Resour. Res., 28, 539-549, 1992.
- Ginn T. R., Cushman J. H. and Houch M. H., A continuous–time inverse operator for groundwater and contaminant transport modeling: deterministic case, Water Resour. Res., 26, 241-252, 1990.
- Giudici M., Development, calibration and validation of physical models, in Geographic Information Systems and Environmental Modeling (K. C. Clarke, B. O. Parks and M. C. Krane, Eds.), 100-121, Prentice-Hall, Upper Saddle River (NJ), 2001.
- Giudici, M., Some problems for the application of inverse techniques to environmental modeling, *Contemporary Mathematics*, under revision, 2003.
- Giudici M., Morossi G., Parravicini G. and Ponzini G., A new method for the identification of distributed transmissivities, Water Resour. Res., 31, 1969-1988, 1995.
- Griffiths D.H. and Barker R.D., Two-dimensional resistivity imaging and modeling in areas of complex geology, *Journal of Applied Geophysics*, **29**, 211-226, 1993.
- Loke M.H. and Barker R.D., Practical techniques for 3D resistivity surveys and data inversion, *Geophysical Prospecting*, **44**, 499-524, 1996.
- Menke W., Geophysical data analysis: discrete inverse theory, Academic Press, San Diego, 1989.
- Nolet G. (ed.), Seismic tomography: with applications in global seismology and exploration geophysics, D. Reidel Pub Co, 1987.

- Parravicini G., Giudici M., Morossi G. and Ponzini G., Minimal a priori assignment in a direct method for determining phenomenological coefficients uniquely, Inverse Problems, 11, 611-629, 1995.
- Renard P. and Marsily G. de, Calculating equivalent permeability: A review, Adv. Water Res., 20, 253-278, 1997.
- Sagar B., Yakowitz S. and Duckstein L., A direct method for the identification of the parameters of dynamic non-homogeneous aquifers, Water Resour. Res., 11, 563-570, 1975.
- Snodgrass M. F. and Kitanidis P. K., Transmissivity identification through multi–directional aquifer stimulation, Stochastic Hydrology and Hydraulics, 12, 299-316, 1998.
- Vázquez González R., Giudici M., Parravicini G. and Ponzini G., The differential system method for the identification of transmissivity and storativity, Transport in Porous Media, 26, 339-371, 1997.