A Cellular Automata Model of Soil Bioremediation

S. Di Gregorio*
Università della Calabria,
Dipartimento di Matematica,
Cosenza

R. Serra†
M. Villani
Centro Ricerche Ambientali Montecatini,
via C. Menotti 48,
I-48023 Marina di Ravenna

The remediation of contaminated soils is one of the major environmental problems in industrial countries today. Among the different techniques that can be applied, *in situ* bioremediation, which relies upon the use of indigeneous microorganisms to degrade the contaminant, is one of the most attractive, both from an environmental and an economic viewpoint. A full-scale bioremediation process requires a number of laboratory and pilot-scale tests in order to assess the feasibility of the remediation, to define potential health threats, and to find optimal operating conditions. Scaling up from the laboratory to the field can greatly benefit from the development of reliable mathematical models, which need to deal with several interacting physical, chemical, and biological phenomena.

A macroscopic cellular automata (CA) model is presented here, which describes the major phenomena that take place in bioremediation. The reasons for using macroscopic CA are discussed. The model is composed of the following three layers, each layer depending on the others.

1. A *fluid dynamical* layer, which describes multiphase flow through the soil.
2. A *solute description* layer, which deals with solute transport, adsorption/desorption, and chemical reactions.
3. A *biological* layer, which describes biomass growth and its interaction with the different chemicals.

The model has been tested in a pilot plant in the case of contamination by phenol. The values of the phenomenological parameters have been determined by the use of genetic algorithms (GAs). The model has proven capable of carefully describing experimental results for a wide range of experimental conditions. It is therefore an application of CA models to a real-world problem of high social and economic relevance.

*Electronic mail address: dig@unical.it.
†Electronic mail address: rserra.cramont@arcobaleno.com.

1. Introduction

Although less debated than air and water pollution, soil contamination is one of the major environmental problems in industrial countries. Different technologies are available, including physical and chemical ones such as soil washing, use of vapor, heating, and vitrification [10, 25]. The traditional approach, based upon landfills, is being discouraged in industrial countries: actually, their use is not a way to really solve the pollution problem, but rather a way to displace it in space and time. The use of biotechnological methods is particularily attractive due both to environmental and economical reasons. Different biological methods include use of bioreactors, landfarming, composting, biopiles, and *in situ* bioremediation, which is the subject of this paper.

In situ bioremediation [2, 24, 26] makes use of the capability of many microorganisms such as bacteria, fungi, and actynomycetes to degrade organic contaminants directly in the polluted soil. In most cases, the degrading activity of indigeneous bacteria is stimulated to accelerate natural phenomena by providing appropriate nutrients (e.g., oxygen, phosphorus, and nitrogen) or creating appropriate physico-chemical conditions (e.g., pH, moisture content), or both. A typical way of providing the desired nutrients uses aqueous solutions that are circulated through the contaminated soils by artificial rain, or using injection and drainage wells. The model described here is tailored to this bioremediation technique and focuses upon the biostimulation of indigeneous bacteria by percolating an aqueous nutrient solution through the soil.

These models can be very useful for estimating the results of field-scale operations starting from laboratory or pilot plant data, while the prevailing approach today is still largely empirical. A few models exist which aim at describing bioremediation, making use of partial differential equations (PDEs). A different approach is taken here, based upon the use of macroscopic CA. The reasons for this choice are discussed in section 2 and further commented on in section 9.

The good agreement between model prediction and experimental results which has been achieved shows that CA can be successfully applied to the bioremediation problem. This is therefore an example of the application of cellular methods to a real-world problem having both a deep scientific appeal and a high practical importance from both environmental and economic viewpoints.

The model has a layered structure, which has a sound physical basis. The fluid dynamical layer describes multiphase flow in a porous medium, the solute description layer describes the fate of the solute, and the third layer describes the biomass dynamics of transport, growth and decay, and metabolic activities. The overall structure of the model is described in section 2, while sections 3 through 5 are dedicated to the description of the three layers of the model.

Dealing with such a complex and only partly known set of phenomena, it is necessary to introduce some phenomenological parameters, which cannot be directly measured, but need to be estimated by comparing the model predictions with experimental results, and using an optimization method to minimize the difference between simulations and experiments. The optimization method chosen in this case is that of genetic algorithms (GAs) [19, 20]. These do not require that the function to be optimized has any special property and are not bound to get stuck in local optima; such as, for example, gradient descent techniques. The GA used is described in section 6.

Models are useful if they compare well with experimental results which can be obtained from a pilot plant or in the field. The model testing that has been performed so far, albeit limited, provides meaningful indications. The major limitations are due to the size of the equipment (a pilot plant of about 1 m^3) and the choice of a single contaminant, namely phenol, being one of the most frequently encountered in practice. On the other hand, an unusually high number of data has been recorded, which allow a thorough testing of the model; moreover, different experimental conditions have been examined, thus allowing testing of the model in cases that had not been used for parameter optimization.

The comparison of the model behavior with experimental results is presented in sections 7 and 8. The former describes cases used for parameter adaptation and the latter shows how the model generalizes to slightly different experimental conditions, without further parameter tuning.

An important issue is that of the model generalization properties, that is, its capability to describe cases that were not used for parameter adaptation without further adjustments. Some results concerning this aspect are shown and discussed in section 8.

Conclusions and suggestions for further improvements are summarized in section 9.

2. The model

Although many models describing multiphase flow in porous media and contaminant transport have been published (for a review see [21]), still few models exist describing the whole bioremediation process, including its biological aspects [6, 9, 11, 12, 21, 23].

All these models make use of an approach based upon PDEs, which involve a limiting operation where the volume of the elementary cell ΔV and the time increment Δt, shrink to zero. However, the elementary cell must still comprise several pores, so this limiting approach is legitimate when a representative volume element can be found (see [5] for a comprehensive discussion). The models are also fairly complicated, so that analytical solutions do not exist for the most interesting cases,

and it is necessary to resort to numerical simulations. So the overall process involves first a limiting operation $\Delta V \to 0, \Delta t \to 0$, and then a discretization scheme where finite approximations are introduced.

We explore an alternative approach based on a straightforward discrete scheme, namely that of CA which, as it is well known, rely upon discrete space cells, discrete time steps, and discrete state space. CAs, which were introduced in the 1950s by von Neumann to study self-reproducing systems [8], have been extensively used since the 1980s for modeling complex dynamics and physical systems [30–32].

While most attempts to apply CA to modeling macroscopic physical phenomena have dealt with a "microscopic" approach, where the state variables can take only a limited number of different discrete values [18, 31], we take here a rather "macroscopic" attitude, considering the cells as portions of space which include several pores. It is therefore meaningful to assign to each cell a more complex state space, which is the cartesian product of different subspaces, which may refer, for example, to the water content, the concentration of a given chemical, the density of a given class of bacteria, and so on. Some of these variables are usually considered as continuous, and we will admit their use.

We will therefore use the term "cellular automata" in a wider sense than that of lattice gases (although formal compliance with the CA definition could be assured by resorting to a "discretized" set of values). Moreover, in order to describe phenomena which happen in a vessel of finite dimensions with physical boundaries, our CAs will not be spatially homogeneous. A rigorous formalization of the generalized CA concept is given in [17].

This kind of model is similar, in some respects, to lattice Boltzmann models [29]. However, our model, which aims at describing large-scale phenomena, makes direct use of macroscopic variables such as water saturation and concentration of a given chemical, while the applications of lattice Boltzmann models to flow in pororus media which have been developed so far describe phenomena taking place at the pore level [1]. Macroscopic CA models are instead related to the so-called compartment models [7], where the space is divided into cells of finite size, which are assumed to be in local equilibrium.

These macroscopic CA models should be effective in simulating large systems because they allow one to choose a cell size which is appropriate for the scale of the simulation. Moreover, as has already been pointed out, they avoid the need for a limiting operation, followed by the introduction of a discrete approximation. On the other hand, the use of a CA approach introduces a dependence upon the discrete space topology which is chosen. Examples from the simulation of landslides and lava flows [4, 13] provide encouraging results on the capability of the model to describe complex physical phenomena. Of course, the validity of this approach should be judged by agreement with the experimental data.

We now describe the bioremediation CA model. As was mentioned before, the model has a layered structure.

- The lower, or *fluid dynamical* layer describes the flow of different fluid phases in a porous medium.
- The second, or *solute description* layer, describes the physical and chemical processes affecting the solutes, including hydrodynamic dispersion, chemical reactions with other solutes or with chemicals which may be present in the soil, and adsorption and desorption to and from the pore walls.
- The third, or *biological* layer, describes the growth of microorganisms and their interactions with their environment.

This paper is concerned with the description of a specific case, namely that of a phenol-contaminated soil. The major simplifications of the general model structure, which have been applied in this case, are the following.

- *Static soil*. The soil is considered to be an unchanging matrix, not affected by the flow; moreover the growth of microorganisms is supposed not to modify the void space in an appreciable way (i.e., no pore clogging).
- *Two phase flow*. Only the flow of a gaseous phase (air) and a liquid phase (aqueous solution) have been considered; due to the high solubility of phenol in water, no separate phase is present.
- *No chemical reactions*. Phenol is assumed to undergo only biologically induced transformations; the introducton of chemical reactions would be straightforward, if it were needed.
- *Simplified population structure*. The bacterial population has been divided into three classes, according to their behavior with respect to the chosen contaminant, and no attempt has been made to provide a more detailed description of the different species involved and of their mutual interactions.

In spite of these limitations, the model is very ambitious, as it aims at describing a wealth of different interacting phenomena.

The specification of the CA involves its topology, its state space, and its transition function (for a more detailed and rigorous formal specification, see [17]).

- *Topology*. The model is three-dimensional, and the space is divided into cubic cells, so that each cell has six neighbors (up, down, north, east, south, and west), one for each face.
- *State space*. The cartesian product of different subspaces is used

$$Q = Q_w \times Q_{\mathrm{cap}} \times Q_{\mathrm{gra}} \times Q_{\mathrm{mph}} \times Q_{\mathrm{cph}} \times Q_{\mathrm{nrb}} \times Q_{\mathrm{rndb}} \times Q_{\mathrm{rdb}}$$

where Q_w is the water content in the cell, Q_{cap} is the "capillary water" substate, that is, the water which is held by capillary forces (see section 3), Q_{gra} is the "gravitational water" (see section 3), Q_{mph} accounts for the phenol mass adsorbed or precipitated in the cell per unit mass of dry solid, Q_{cph} accounts for the phenol concentration in the cell water, Q_{nrb} accounts for the concentration of bacteria which are not resistant to the phenol, Q_{rndb} accounts for the concentration of those bacteria which are resistant to the phenol but are not able to degrade it, and Q_{rdb} refers to those bacteria which are resistant to and can degrade phenol.

- *Transition function.* The main features of the deterministic transition function, that is, the law which determines the state of a cell at time $t + 1$ from the state of its neighboring cells at time t, are described in sections 3 through 5. The precise specification of the transition functions must be given in terms of algorithms, which play in CAs the same role of equations in continuous models; as they are fairly long and cumbersome to read, a complete description of the algorithms is not given here, but can be found in [17].

In order to describe the boundaries of the experimental vessel, some cells are given a particular transition function; for example, the upper layer can take an inflow of water from the outside, while the lower layer describes the collecting apparatus at the bottom of the pilot plant [17].

3. The fluid dynamical layer

We now consider what happens in a porous soil when water saturation S (which is defined as the ratio of volume of water to the volume of voids) increases. When water saturation is low, a minimum quantity of water surrounds the grains of soil, it is immobile and can be removed only by strong heating of the soil. At a slightly greater saturation level water forms rings (pendular rings) around the grain contact points, influenced essentially by intermolecular forces. As water saturation increases, the pendular rings expand until a water continuum is reached (water funicular saturation), so that water can flow. If water saturation increases further, all the pores will be filled, except air, if any, entrapped in the largest pores. The above qualitative description [5] refers to the case where water is the wetting phase.

In order to capture the phenomenology described, we introduced the following three kinds of water in the model [15].

- *Immobile* (or irreducile) water, which cannot flow.
- *Capillary* (or diffusion water), which can flow according to a diffusion-like dynamics, tending to equalize its content with neighboring cells.
- *Gravitational* (or transport) water, which can flow more rapidly under the action of gravity.

All the water is regarded as immobile until a certain threshold is met (there are no connected paths below this threshold). Above this level, water is capillary (a thin layer close to the pore walls) until a second threshold is found. Above this latter threshold, water is considered gravitational (if appropriate, part of the water can be regarded as fissural).

Moreover, as in some cases macroscopic fissures can be found, these could also be described in a way similar to gravitational water, but with a faster flow rate, provided that "fissural water" is introduced. In the following, fissures are not taken into account.

According to the model, the motion of capillary water tends to equalize its value in neighboring cells. The mechanism of equalization distributes the excess water of a given cell among those of its neighbors with a water content that is smaller than the local average water content, thus giving rise to dynamics similar to those described by diffusion equations. The maximum flow of capillary water is computed as follows.

Let V_i denote the neighborhood of cell i, which is composed of N_i cells, and let q_i denote the capillary water in cell i. An average $\langle q \rangle_i$ between all the cells of the neighborhood is first computed:

$$\langle q \rangle_i = \frac{1}{N} \sum_{j \in V_i} q_j.$$

The neighboring cells having a water content that is higher than this average are identified and discarded. Let $V_i^{(1)}$ be the subset of V_i composed of the $N_i^{(1)}$ cells which have not been eliminated. Then the average between the remaining cells is computed:

$$\langle q \rangle_i^{(1)} = \frac{1}{N_i^{(1)}} \sum_{j \in V_i^{(1)}} q_j.$$

The process is iterated until no more cells are found with a water content higher than the average. The value of this latter average is attributed to the central cell and to all the cells which have "survived" (see Figure 1 for a two-dimensional example).

The preceding computation refers to the maximum (capillary) water that can flow from the central cell. At each CA time step, only a fraction k_{cap} of the excess water so computed actually flows, in order to take into account the system kinetics. As it is well known [5] that the flow rate of the water phase depends upon its saturation level S, it has been assumed that the dependence of the flow rate coefficient k_{cap} upon saturation is S-shaped, as suggested by experimental data concerning the so-called relative permeability curves.

The gravitational water can flow under the action of gravity from a cell to the cell below it. A transport coefficient k_{grav} for each cell is calculated; the flow of water to the inferior cell is proportional to

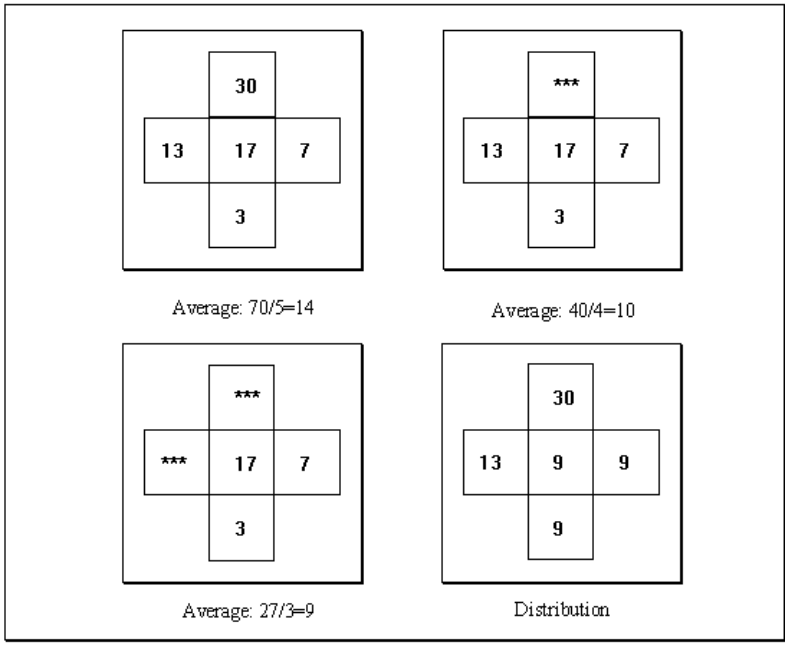

Figure 1. An example of the algorithm for capillary water.

the smallest of the two k_{grav} and in each case cannot be higher than the receptivity (i.e., the quantity of water that a cell can accept) of the inferior cell. For the same reasons just discussed for capillary water, the gravitational flow rate coefficient also depends upon saturation, although its kinetic parameters allow a faster flow.

It is worth mentioning here that this model is capable of describing both the fast initial outflow and the long slow percolation tail (see section 7), thus encompassing phenomena with widely different time constants.

4. The dynamics of chemicals

This model layer describes the behavior of the different chemicals transported by the aqueous phase (i.e., nutrients, contaminants, others). We will usually refer to the specific case of contaminants, but the same approach is adopted to describe the transport of nutrients; such as, for example, dissolved oxygen, nitrogen, and methanol.

Solutes are transported by the water flowing in the system, they interact with the pore walls and diffuse in the water which was already present in the cell. The adsorption/desorption process has been modeled by first-order kinetic equations, where the adsorption rate to the pore

walls is proportional to the contaminant bulk concentration, and the desorption rate is proportional to the contaminant adsorbed concentration.

Immobile water is also present, and it must be taken into account when modeling contaminant and nutrient transport. For the sake of definiteness, we now describe the contamination event. Actually, in the test apparatus, the amount of water used for contamination was smaller than the overall immobile water of the soil, so that the phenomenon of contamination of existing water played a crucial role. It has been assumed that water can be either contaminated or not contaminated, and that only contaminated water flows from one cell to another (of course, if it were necessary to describe different phenomena, like the flow of pure water through a contaminated soil, this simplification should be avoided). It has also been assumed that, as time passes, the contamination front advances within the cell, so that a larger portion of water becomes contaminated (of course, the computation of the new contaminant concentration is done in such a way that assures mass conservation).

As has already been pointed out, in the version of the model used for comparison with experimental results, it was not necessary to introduce chemical reactions. Their introduction would be straightforward, as they take place within each cell and do not involve transport among neighbors.

5. Biological phenomena

We identified three classes of microorganisms that were measured during the experimental phase. Using bacterial counts on selected colture media it was possible to estimate the density of different kinds of bacteria [3]. Some counts (T-bacteria) were performed on a rich colture medium, others (R-bacteria) on the same colture medium to which a given amount of phenol was added, and the third kind of count (D-bacteria) on a minimum salt medium where the only carbon source was phenol. The idea is that the first kind of count provides an estimate of the "Total" number of bacteria, the second kind of count includes a selective pressure which favors those bacteria which are able to "Resist" a given phenol concentration, while in the third colture medium only those bacteria able to "Degrade" phenol can grow. From these counts it is possible to estimate the densities of NR-bacteria ("nonresistant," NR = T – R) and of RND-bacteria ("resistant but not degrading," RND = R – D).

It is well known that plate count techniques do not provide precise measures of the microorganisms present in the soil, which may sometimes be underestimated even by two orders of magnitude [25], and that the bacterial species which prevail on rich colture media are likely to be different from those which prevail in an oligotrophic ecosystem

such as a contaminated soil [28]. Moreover, terms such as "resistant" or "degrading" have no absolute value, but refer to the concentration of phenol used in the experiments. In spite of these limitations, plate count techniques on selective colture media are useful for estimating the densities of the different bacterial species and to monitor their change in time.

The model of bacterial growth takes into account three classes of bacteria, and for all three classes standard Volterra growth equations have been used as a starting point. As phenol is toxic for bacteria, a death term has been added to the biomass growth equation, proportional to the phenol concentration times the bacterial density. The coefficient of this death term is high for NR-bacteria and low for R- and D-bacteria. The equation for this last kind of bacteria also includes a growth term, which is proportional to the bacterial density times a Monod function of the phenol (if bacteria are exposed for the first time to phenol, this term becomes active only after a suitable lag time, required to develop the necessary metabolic pathways). As transition functions, which are expressed by procedures, may appear somewhat cumbersome, we prefer to cast the major assumptions used to model the dynamics of D-bacteria in a familiar form, that is, using difference equations, in the following way:

$$B(t+1) = B(t)(1 + K_N) - \beta B(t)^2 + \alpha B(t)\frac{F(t)}{A + F(t)} - K_M B(t)F(t)$$

$$F(t+1) = F(t) - \alpha' B(t)\frac{F(t)}{A + F(t)}$$

where B is the concentration of degrading bacteria, F is the phenol concentration in the soil, α and A are Monod coefficients, α' is the phenol consumption rate, β is the coefficient of the generic death term for bacteria, K_N is the coefficient of the birth term, and K_M is the coefficient of the death term caused by phenol. The equations for ND- and RND-bacteria have a similar form, without the Monod term.

The availability of nutrients can affect the kinetic parameters of the growth equation. Based on bibliography and on experimental data, it has been possible to limit consideration to the effect of oxygen only; as the phenol degradation pathway is aerobic, it has been assumed that the coefficient of the Monod growth term is a growing function of the available oxygen concentration.

In some experiments [3] it was shown that bacteria can survive for a long time even in very harsh conditions; for example, very dry soil, giving rise to survival forms with a very limited metabolic activity. When water is introduced into the system, these bacteria undergo a gradual reorganization, coming back to "full life" after a certain lag phase. In a similar way, when a new carbon source is provided as food for the microorganism, it may take some time to activate a set of enzymes

capable of degrading it. In order to describe these phenomena it has been convenient to introduce in the model a kind of "clock" which measures, for example, the time elapsed since the last time the water saturation exceeded a minimal threshold in a given cell. The bacterial activity then depends upon these clocks, as the kinetic coefficients are turned on or off in an appropriate way. For example, the Monod parameters α and α' are 0 until a certain time has elapsed since the introduction of phenol in the container in order to model the lag phase necessary to develop the corresponding metabolic pathway.

As far as bacterial motion is concerned, we took an "Occam razor" approach and kept the model complexity at a minimum level, compatible with experimental data. Therefore in this version of the model the bacteria are assumed to be immobile, bound to the pore walls, and bacterial transport in the aqueous phase is supposed to be negligible (although it is known that other different behaviors are possible [21, 23]).

6. Search techniques

As has been observed, there are some parameters in the model that cannot be directly measured or derived from first principles (see Table 1). It is therefore necessary to estimate these parameters on a pilot plant, and to use the estimated values to simulate field-scale operations. Therefore a suitable optimization technique is required, which adapts the model parameters by comparing the simulation results with the experimental data.

Parameter	Estimation Method
immobile water threshold	measurement
threshold between capillary and gravitational water	fitting (*genetic*)
shape factors for dependence of k_{cap} upon saturation	fitting (*genetic*)
shape factors for dependence of k_{grav} upon saturation	fitting (*genetic*)
initial water saturation	measurement
porosity	measurement
adsorption/desorption kinetic coefficients	fitting (*genetic*)
rate of diffusion in immobile water	fitting (*genetic*)
Monod coefficents	fitting a Monod equation to the observed population growth curve

Table 1. The major model parameters and their estimation methods.

GAs [19, 20], which are not bound to get stuck in local minima (the way gradient descent techniques can) and which do not require that the cost function to be minimized possess any peculiar property (i.e., continuity, differentiability, etc.), have been used. In this work we used a binary coding of the parameter values, and the classic genetic operators of single point crossover and point mutation. Each set of parameter values is coded as an individual, and a random initial population is generated. The GA then tries to produce new generations of fitter individuals, that is, those able to give a closer agreement between simulations and experiments. The fitness function of an individual (i.e., a set of parameter values) is defined as the (normalized) square of the euclidean distance between two M-dimensional vectors: $f = \|\mathbf{O} - \mathbf{S}\|^2/\|\mathbf{O}\|^2$. Here the components of $\mathbf{S}$ are the values, at M time steps, obtained in the simulation and the components of $\mathbf{O}$ are the values, at the same time steps, obtained in the pilot plant experiments.

As genetic search may sometimes be affected by premature convergence to suboptimal solutions, it is convenient to limit the competition to subsebts of the whole population. The "individuals" are therefore regarded as if they were distributed in space, interacting only with their neighbors. More precisely, the individuals are arranged on a two-dimensional grid with wrap around, thus giving rise to a toroidal topology. The five-membered neighborhood of each individual is composed of itself and by its north, east, south, and west neighbors. At each step an individual is chosen and selection, crossover, and mutation only act within its neighborhood. The children strings that have been generated from the genetic operators, compete only with the individuals belonging to the chosen neighborhood and have a chance of replacing some of them. At the following step of the GA, a new individual will be chosen, and the same procedure will be applied. Further details can be found in [16].

The layered structure of the model is convenient as it allows one to optimize subsets of parameters in different phases, for example, the parameters of the fluid dynamical layer can be adapted using the results of the contamination experiments only. In this way a large search space can be broken into more manageable portions.

7. Comparison with experimental results

The model described in the previous sections has been tested versus a set of careful experiments on pilot-scale constructed soils, suitably equipped. Most of them have been done with two identical vessels, called MR1 and MR2, while some auxiliary tests have been performed on two smaller twin vessels called MR3 and MR4. The soil came from an industrial site that had been contaminated by phenol for several years, and which had already undergone a successful field-scale bioremedia-

tion intervention. Two kinds of experiments have been performed. In the first experiment (contamination), an aqueous phenol solution was percolated through the soil, while in the following bioremediation phase an aqueous nutrient solution was percolated in order to stimulate the activity of indigeneous bacteria. Further details can be found in [3].

In the contamination experiments, the following variables were measured in time.

- The amount of drainage water which percolated through the soil to the bottom of the container.
- The concentration of phenol in the drainage water.
- The concentration of phenol in the soil at different depths.
- The soil temperature.

The two twin pilot plants could be contaminated in two different ways, that is, either with an instantaneous flooding from the upper side of the container, or with a slower flow generated by a spraying system.

Bioremediation made use of a nutrient aerated solution that was sprayed on the soil. Although most field operations are performed in a closed loop, where percolated water is collected by wells, mixed with oxygen and nutrients, and reinjected in the field, the tests were performed in an open loop fashion to ease model validation. In the bioremediation experiments the same parameters described for the contamination experiment were measured, plus microbiological counts.

The model predictions compare well with experimental results, in a set of conditions having widely separated time scales and a range of different operating conditions. In this section we compare the results of some experiments with the model predictions. The model parameters have been adaptad to achieve the best possible agreement with experimental data.

Figures 2 and 3 compare model and experimental results concerning the total percolated water on a long and short time scale for the contamination experiment. It can be seen in both cases that the two differ at most by a few percent. These data refer to the case where the vessel was flooded with the phenol solution. Both one- and three-dimensional models perform in a similar way.

The phenol concentration in drainage water was also carefully described by the model for the flooding case as shown in Figure 4.

It is also interesting to note that the experimental data have shown a continuous decrease in phenol concentration in percolating water, with the exception of day 8, where an increase was observed. By simply looking at the experimental data set, doubts were raised about the quality of the data showing a concentration increase. However, the model also reproduces this behavior, which seems therefore related to a real

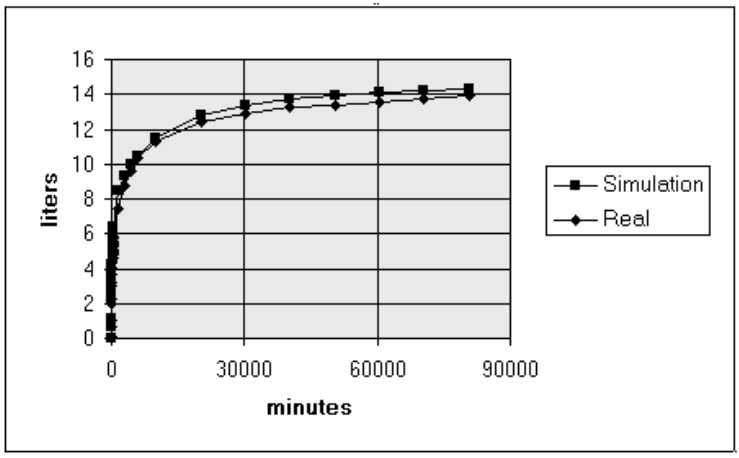

Figure 2. Total percolated water (MR1, flooding).

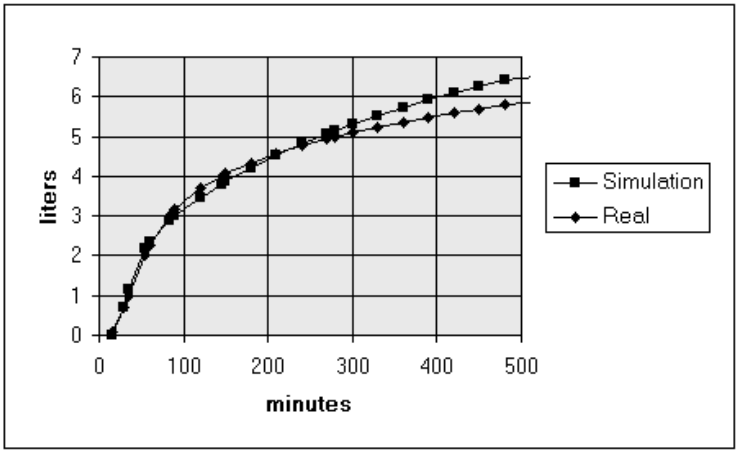

Figure 3. Total percolated water (MR1, flooding).

physical phenomenon. If bacterial activity is turned off, one observes a slight increase in phenol concentration in drainage water, which lasts for some time. The decrease which is experimentally observed is due to the fact that bacteria spontaneously start to degrade phenol in an appreciable amount about one week after contamination. This is a case where modeling allowed us to understand the meaning of experimental data which would otherwise have been overlooked.

The data for phenol concentration in soil are less reliable, due to their high spatial heterogeneity (as sampling perturbs the system, the limited

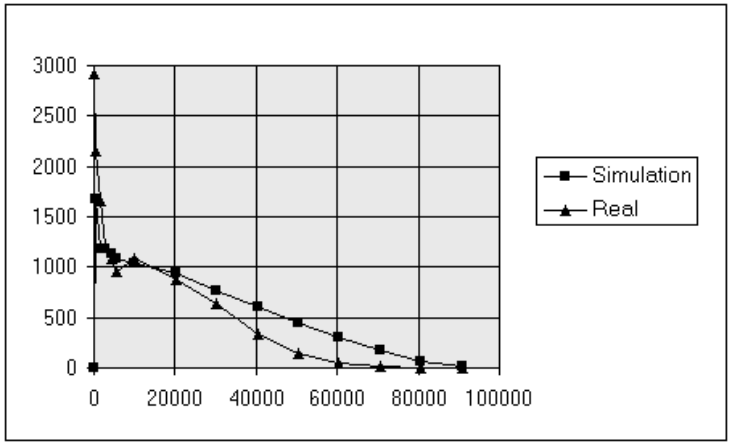

Figure 4. Phenol concentration in the drained water (MR2, flooding).

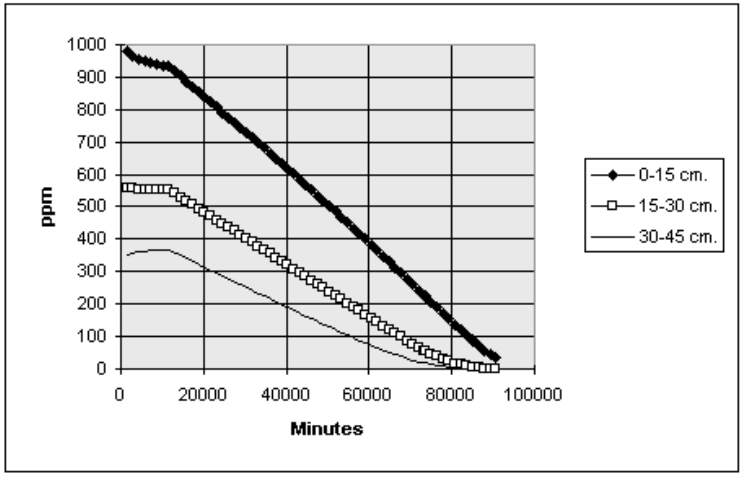

Figure 5. Simulation of the phenol concentration in three soil layers of MR2.

dimensions of the pilot plant did not allow us to take a large number of samples). The most important experimental fact is that, although phenol concentration decreases with increasing depth, the time when phenol disappears (that is, becomes smaller than the detection limit) is about the same at every depth. The model reproduces the same behavior, as can be seen in Figure 5. This is due to the fact that in the upper layers there are both higher phenol concentrations and more oxygen available to the microrganisms, which in turn leads to greater biomass growth and to faster phenol degradation.

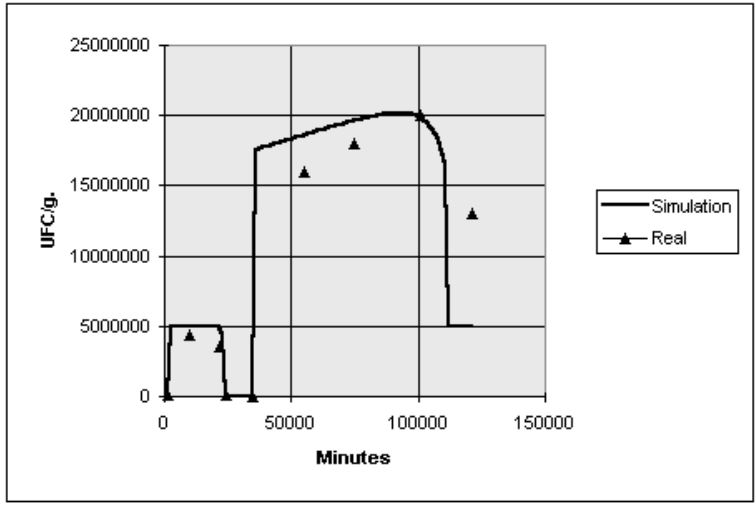

Figure 6. Concentration of D-bacteria in MR4 (upper layer, simulated and experimental data).

As far as the bacterial density in soil is concerned, the experimental data refer to MR3 and MR4. This is due in part to the fact that bioremediation in MR1 and MR2 was faster than expected, so that the sampling frequency initially chosen was too low, and in part to the fact that improvements in the plate counting technique were introduced for MR3 and MR4. The experimental data concerning the time behavior of the bacterial densities in MR4 are shown in Figure 6.

8. Generalization properties

A very important concern is the way the model behaves in cases which were not used for parameter adaptation (the so-called generalization problem). Actually, in complex models like those considered here there is no theoretical guarantee that the parameter values which have been found represent global minima of the cost function. Therefore, the answer to the question concerning model generalization capabilities should be mainly empirical.

For this purpose, we show how the model behaves in cases which had not been used for parameter adaptation, and we show that it often achieves a satisfactory generalization. However, such empirical testing is of course limited by the nature of the experiments performed, and we cannot exclude that important disagreements exist in different kinds of experiments.

Figure 7 shows how the model describes the slow bioremediation phase, achieved by spraying 1 lt/day of an aqueous nutrient solution.

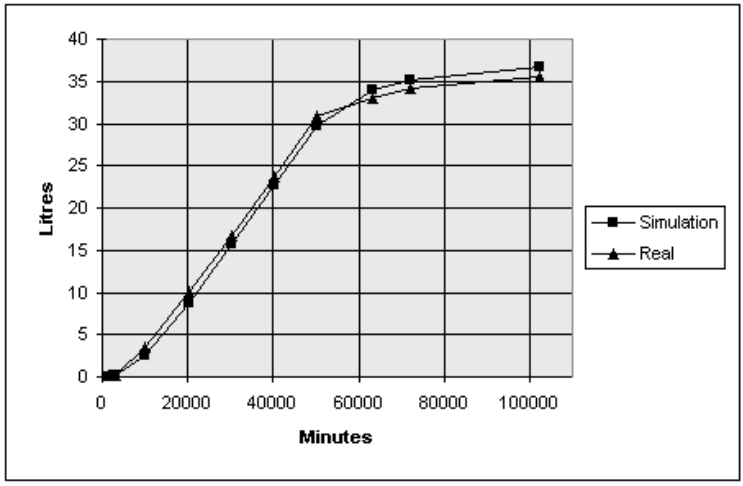

Figure 7. Water percolated during bioremediation.

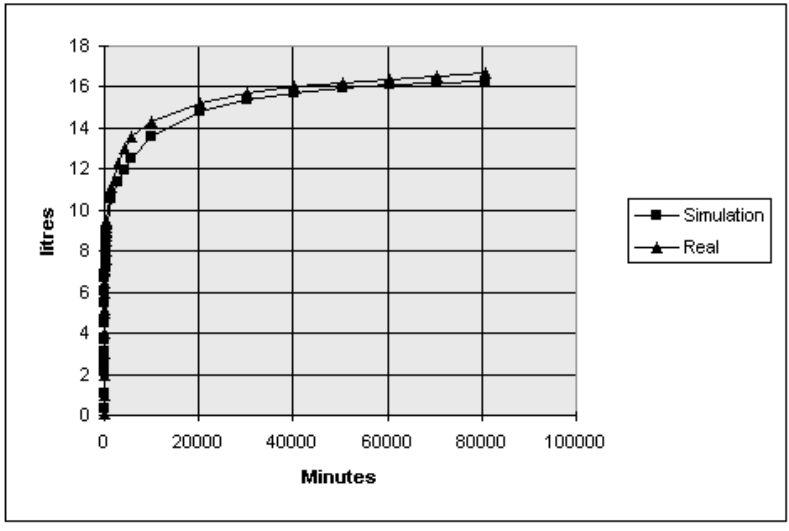

Figure 8. Water percolated after spraying 15 lt.

Recall that the model parameters were adapted for the case of sudden flooding (Figures 2 and 3)).

Figure 8 shows the model performance in predicting the long-term behavior for the contamination experiment, when contamination was performed with slow spraying (for historical reasons, the initial water saturation—which is a measurable parameter, see Table1—was higher in this case than in the flooding case).

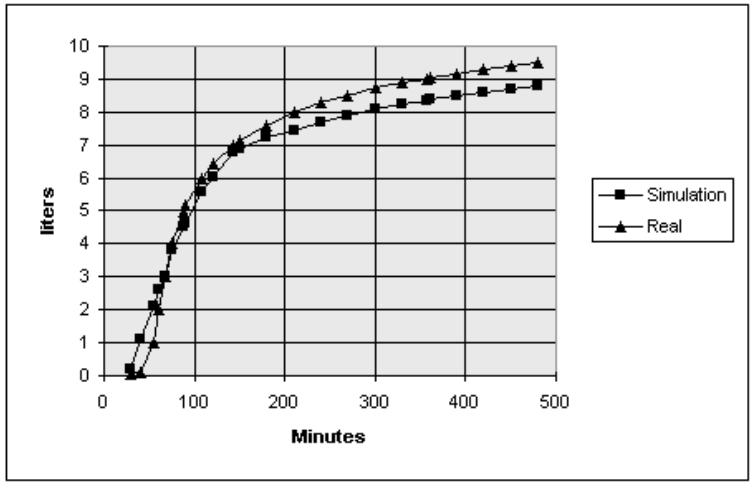

Figure 9. Total percolated water (MR2, rain). The simulation refers to the three-dimensional model with ponds.

It is interesting to note that for the case of contamination by slow rain the one-dimensional model was unable to achieve satisfactory performance on the short-time scale (even when we tried to independently adapt its parameters). The reason is physical, and is due to the fact that in the rain case ponds are formed, which gives rise to preferential flow paths. Due to the nonlinear relationship between water saturation and hydraulic conductivity these preferential paths provide a faster overall flow. This behavior could be successfully modeled only by a three-dimensional model (see Figure 9). In this case the parameters were the same as those used for the flooding case, but a spatial pond distribution was given which was based upon drawings that had been taken during the experiments.

It is also worth noting that it has been observed that the model provides a fairly accurate forecast of the percolated water even in the case when the slow rain lasts much longer, so that 60 lt. (i.e., four times the quantity of Figures 7 through 9) are supplied to the vessel.

As far as chemicals are concerned, the flow of phenol in the percolated water for the slow rain experiment (the same that Figures 8 and 9 refer to) is shown in Figure 10 (the parameters are of course the same as those for the flooding case of Figure 4). Note the accuracy of the model predictions, although in this case the total amount of percolated phenol is almost twice as much as that of the training case (Table 2). The difference is mostly related to the fact that a higher initial water saturation leads to different kinetics for phenol diffusion.

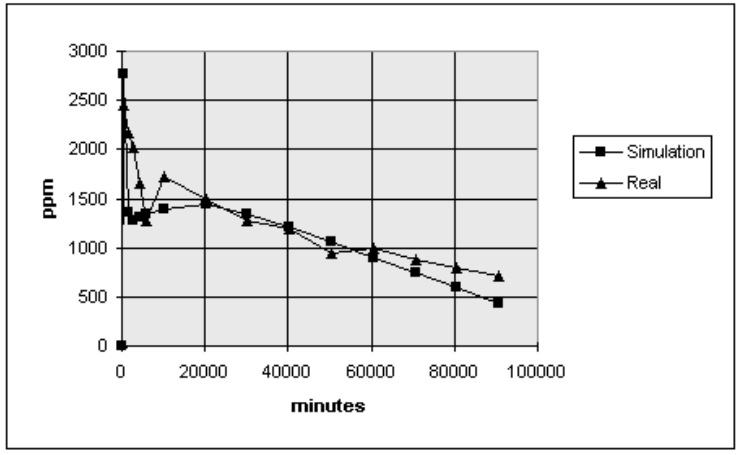

Figure 10. Phenol in the percolated water, slow rain.

Total phenol lost in drainage	Simulated	Real
MR1	8.58%	9.90%
MR2	15.77%	16.70%

Table 2. Total amount of phenol lost in drainage in two different experiments; MR1 (flooding) was used for training, MR2 (rain) for testing.

We now consider a different way of testing the model. It is well known that an empirical equation, called the Darcy–Buckingham law [22], which states that the water flow is proportional to the gradient of the hydraulic head, accurately describes the water flow through the soil under appropriate conditions. We reproduced in the model the same conditions as those described in the experiments which are known to follow this law. In particular, we simulated the downward flow in an unsaturated soil column, when the water level is kept constant above the upper surface. As in this case the so-called gravity flow approximation holds, the Darcy–Buckingham equation predicts that the flow rate q is equal to the Darcy coefficient k (for further details, see [27]). The comparison of these two quantities in the model simulations is shown in Figure 11.

From the data of Figure 11 a Darcy permeability for the saturated soil can also be computed which turns out to be equal to 2.2×10^{-3} cm/sec. This is in good agreement with the value of 2.6×10^{-3} cm/sec, which was measured for the soil and used in all the experiments of sections 7 and 8.

It is worth stressing that not all the generalization experiments achieved such good results. For example, when we flooded the ves-

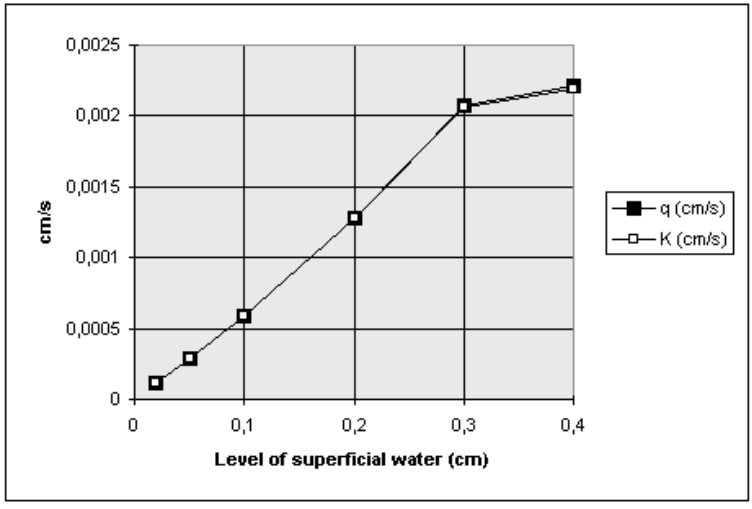

Figure 11. Relation between outcoming flow and Darcy coefficient, unsaturated region.

sel with 60 lt of water, instead of 15 lt, the kinetics of water percolation was rather different from that predicted by the model. The reason is probably that, in all the different experiments described in section 7, there was never such a high quantity of water within the cells, so the parameters have been tuned in regions of state space rather different from those which are involved in the 60 lt case. It should be recalled that this is a fairly general property of learning and adaptive systems: they tend to work much better in cases close to those used for training than in regions which are far away. The reason why the model performed so well in the cases shown in this section is probably that, although they involve fairly different experimental activities, the amount of water present in the cells at each time step remained bounded to values not very different from those used for training.

It should also be stressed that, due to the lack of reliable experimental data, it has not been possible so far to perform generalization experiments on the bacterial layer.

9. Conclusions

The agreement between experimental data and simulation results is remarkable, taking into account the complexity of the system under study and the fact that the phenomena are fully nonstationary.

This agreement provides an *a posteriori* test of the validity of the CA approach. These macroscopic CAs are in some respects similar to PDEs, and it can be guessed that similar results could be achieved

with this latter approach. However, developing a CA model has proven reasonably fast, and it is likely that there is no advantage in taking a PDE approach. Moreover, CAs allow a straightforward way to introduce phenomenological terms.

The models described here have been simulated on a MIMD parallel architecture, achieving an almost linear speed up (efficiency of the order of 80–90%) [14].

A major questionable aspect of this paper concerns the presence of adjustable parameters. This is certainly a limitation, which is however basically unavoidable in this kind of model. The testing was very demanding, involving different experimental settings and conditions, so we think that the good results achieved are not just a matter of fitting. In the model development phase, it was verified that satisfactory performance could not be achieved until the dominating physical phenomena are identified. For example, before understanding the crucial role of immobile water on the fate of the contaminant, it had not been possible to provide good results for the phenol concentration in drainage water, no matter which adsorption/desorption coefficients were chosen.

The generalization properties described in section 8 are rather interesting. It should be recalled that, as in most cases where adjustable parameters are used, good results can be achieved in regions of variable space close to those used for training while there is no guarantee of satisfactory performance when one tries to extrapolate to far regions.

In any case, a major improvement might involve reducing the number of parameters. Attempts are under way, for example, by using local equilibrium assumptions for the adsorption/desorption process and by eliminating the sharp threshold between capillary and gravitational water.

Note that the model is limited to a single contaminant, namely phenol. As there are several interacting phenomena to be taken into account, an Occam razor approach was chosen, thus avoiding the introduction of unnecessary terms. However, these terms might be necessary to deal with different cases, so further work is needed to confirm and complete the results described here. Anyway, all the model improvements which have been envisaged so far can be inserted in the framework presented. The major issue which is still open concerns the accuracy of the scaling of the model to the field, which of course requires expensive large-scale measurements. The tests which have been performed, albeit on a limited scale, are however very stringent, and have provided a thorough testing set.

We believe that these results, which show how even such a complex phenomenon can be precisely described by a simulation model, can facilitate the diffusion of model-based techniques in the sector of *in situ* bioremediation, which is still dominated by a highly empirical approach. We also believe that the given results show that CA can be

usefully applied to a real-world problem of very high environmental and economical importance.

Acknowledgements

This work has been partially funded by the UE Esprit project (Capri initiative, subproject Caboto and project Colombo, #24907). We gratefully acknowledge the contribution of many friends and colleagues who contributed to the successful completion of the Caboto project and are presently involved in the Colombo project: Rocco Rongo and William Spataro performed the simulations on a parallel machine, using the Camel software environment developed by Giandomenico Spezzano and Domenico Talia; Alessandro Banti ran the pilot plant experiments; Federica Abbondanzi and Antonella Iacondini developed the microbiological counting techniques and performed the measurements, while Nello Lombardi set up the methods and performed the chemical analyses; Massimo Andretta, with his physical insight, provided useful criticism and suggestions for model development. We also wish to thank Adam Kleczkowski for carefully reading a first version of the manuscript and providing very useful suggestions for improvement.

References

[1] P. M. Adler, *Multiphase Flow in Porous Media* (Kluwer, Dordrecht, 1995).

[2] K. H. Baker and D. S. Herson, *Bioremediation* (McGraw-Hill, New York, 1994).

[3] A. Banti, R. Serra, F. Abbondanzi, M. Giorgini, A. Iacondini, and S. Mazzullo, "A Detailed Analysis of *in-situ* Bioremediation: Pilot Plant Experiments with a Phenol Contaminated Soil," Proceedings of First International Conference—The Impact of Industry on Groundwater Resources," (Associazione Georisorse e Ambiente, Torino).

[4] D. Barca, et al., "Cellular Automata for Simulating Lava Flows: A Method and Examples from the Etnean Eruptions," *Transport Theory and Statistical Physics*, **32** (1994) 195–232.

[5] J. Bear, *Hydraulics of Groundwater* (McGraw-Hill, New York, 1979).

[6] P. B. Bedient and H. S. Rifai, "Modeling *in situ* Bioremediation," in *In Situ Bioremediation*, edited by B. E. Rittman (National Research Council, Washington: National Academy Press, 1993).

[7] M. Bonazountas and D. Kallidromitou, "Mathematical Hydrocarbon Fate Modelling in Soil Systems," in *Principles and Practices for Petroleum Contaminated Soils* (Lewis Publishers, Boca Raton, 1993).

[8] A. W. Burks, *Essays on Cellular Automata* (University of Illinois Press, Urbana, 1970).

[9] C. Dacome, *Modelli di biorisanamento di suoli contaminati*, Ph.D. thesis, Università di Padova, 1997).

[10] E. De Fraja Frangipane, G. Andreottola, and F. Tatano, *Terreni Contaminati* (C.I.P.A., Milano).

[11] S. Dhawan, L. T. Fan, L. E. Erickson, and P. Tuitemwong, "Modeling, Analysis, and Simulation of Bioremediation of Soil Aggregates," *Environmental Progress*, **10** (1991) 251–260.

[12] S. Dhawan, L. E. Erickson, and L. T. Fan, "Model Development and Simulation of Bioremediation in Soil Beds with Aggregates," *Ground Water*, **31** (1993) 271–284.

[13] S. Di Gregorio, et al., "Un modello ad automi cellulari per la simulazione di frane: il caso di monte Ontake," in *Proceedings of ACRI 94*, edited by S. Di Gregorio and G. Spezzano (CRAI, Cosenza, 1994).

[14] S. Di Gregorio, R. Rongo, W. Spataro, and G. Spezzano, Talia, "High Performance Scientific Computing by a Parallel Cellular Environment," *Future Generation Computer Systems*, **12** (1996) 357–369.

[15] S. Di Gregorio, R. Rongo, R. Serra, W. Spataro, and M. Villani, "Simulation of Water Flow through a Porous Medium," in *Proceedings of ACRI 96*, edited by S. Bandini and G. Mauri (Springer-Verlag, Berlin, 1997).

[16] S. Di Gregorio, R. Serra, and M. Villani, "Environmental Application of Genetic Algorithms," in *Advances in Intelligent Systems* edited by F. C. Morabito (IOS Press, Amsterdam, 1997).

[17] S. Di Gregorio, R. Serra, and M. Villani, *Theoretical Computer Science*, in press.

[18] U. Frisch, B. Hasslacher, and Y. Pomeau, "Lattice Gas Automata for the Navier–Stokes Equation," *Physical Review Letters*, **56** (1998) 1505–1508.

[19] David E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison Welsey, 1989).

[20] J. H. Holland, *Adaptation in Natural and Artificial Systems* (University of Michigan Press, Ann Arbor, 1975).

[21] A. A. Jennings and A. Manocha, "Modeling Soil Bioremediation," in *Remediation of Hazardous Waste Contaminated Soils*, edited by D. L. Wise and D. J. Trantolo (Marcel Dekker, New York, 1994).

[22] W. A. Jury, W. R. Gardner, and W. H. Gardner, *Soil Physics* (John Wiley & Sons, 1991).

[23] F. J. Molz, M. A. Widdowson, and L. D. Benefield, "Simulation of Microbial Growth Dynamics Coupled to Nutrient and Oxygen Transport in Porous Media," *Water Resources Research*, **22** (1986) 1207–1216.

[24] R. D. Norris, et al., *Handbook of Bioremediation* (Lewis Publishers, Boca Raton, 1994).

[25] I. L. Pepper, C. P. Gerba, and M. L. Brusseau, *Pollution Science* (Academic Press, San Diego, 1996).

[26] B. E. Rittman, editor, *In situ Bioremediation* (National Research Council, National Academy Press, Washington, 1993).

[27] R. Serra, M. Villani, D. Oricchio, and S. Di Gregorio, "Recent Advances in Dynamical Models of Bioremediation," in *Proceedings of Acri 98*, edited by S. Bandini, R. Serra, and F. Suggi (Springer-Verlag, London, (in press)).

[28] J. H. Slater, and D. Lovatt, "Biodegradation and the Significance of Microbial Communities," in *Microbial Degradation of Organic Compounds*, edited by D. T. Gibson (Marcel Dekker, New York, 1984).

[29] S. Succi, R. Benzi, and F. Higuera, "The Lattice Boltzmann Equation: A New Tool for Computational Fluid Dynamics," *Physica D*, **47** (1991) 219–230.

[30] T. Toffoli, "Cellular Automata as an Alternative to (Rather than an Approximation of) Differential Equations in Modeling Physics," *Physica D*, **10** (1984) 117–127.

[31] T. Toffoli and N. Margolus, *Cellular Automata Machines* (MIT Press, Cambridge, 1987).

[32] S. Wolfram, "Universality and Complexity in Cellular Automata," *Physica D*, **10** (1984), 1–35.

[33] S. Wolfram, *Theory and Application of Cellular Automata* (World Scientific, Singapore, 1986).