GENERATION OF BINARY PERMEABILITY FIELDS FOR THE SIMULATION OF ESCAPE PATHS IN GEOLOGICAL CO_2 SEQUESTRATION PROJECTS

ANDREA COMERLATI[1], ALBERTO BELLIN[2], AND MARIO PUTTI[1]

[1]DMMMSA - Department of Mathematical Methods and Models for Scientific Applications, University of Padova, via Belzoni, 7, 35131 Padova, Italy

[2]DICA - Dipartimento di Ingegneria Civile ed Ambientale, University of Trento, via Mesiano, 77, 38050 Trento, Italy

Abstract

The emission of greenhouse gases, and particularly of carbon dioxide (CO_2), is steadily increasing since the beginning of the industrial revolution. Many studies evidenced that the earth's climate dynamics may be seriously impacted if the actual increasing trend of CO_2 emission in atmosphere is not reversed very soon. However, reducing the emissions had been proven very difficult for most of the occidental countries, and an elusive goal for countries with a rapid economical growth. Given this problematic scenario, sequestration in deep geologic formations may reduce CO_2 emission in atmosphere with little impact on the industrial activities. Deep saline aquifers are potentially well suited for storing large volumes of CO_2, though the general lack of specific studies assessing the feasibility of such an operation. In particular, more studies are needed to identify and properly characterize aquifers sealed by continuous clay geological units at the regional scale. An important aspect that should be addressed prior any application in this field is the role of hydraulic property variations of the confining layer e.g., by estimating the probability of having small fractures or zones of high permeability interrupting the continuity of the sealing cap.

To this aim, we present an algorithm for the generation of stochastic binary permeability fields composed of a low permeability matrix with high permeability inclusions simulating local interruption in the continuity of the sealing cap. We investigate the performance of the Indicator Kriging technique as a generation tool and study the possibility of conditioning the resulting realization to available data. The features and limitations of the algorithm are assessed numerically.

1. INTRODUCTION

There is a growing concern for the influence that a higher than actual concentration of greenhouse gases in the atmosphere may have on the earth's climate in the next century. During the past 60 years the atmospheric carbon dioxide concentration has risen from the preindustrial levels of 280 ppm to the present level of 375 ppm [Keeling and Whorf, 2004], primarily because of the massive use of fossil fuels as primary source of energy. In just one

year from 1999 to 2000 the global CO_2 emission from fossil fuel increased of 1.8% reaching 6.6 GtC [1] [Marland *et al.*, 2003]. Solid and liquid fuels account for almost 76.8% of the total fossil fuel consumption, while natural gases account for 19.3% of the total emission, reflecting a gradually increasing utilization of natural gas. In addition, emissions from cement production, which is a good proxy of the civil construction market, have doubled since the mid 70s and now represent 3.4% of the global CO_2 release from fossil-fuel burning. On the other hand, there are evidences from climate models that the burst of atmospheric concentrations of CO_2 may be the most important cause of the global warming, currently estimated at 0.3 to 0.6 °C during the last 150 years [Ledley *et al.*, 1999].

The 1997 Kyoto protocol [UNFCCC, 1997], ratified by most of the industrialized countries, includes a program for reducing the CO_2 emissions. Governments are then facing with the problem of reducing the emissions by minimizing the negative impact on the economy. In addition to actions promoting energy saving and production from renewable sources, CO_2 sequestration may significantly contribute to achieving the challenging Kyoto's goals. Oil industry have developed the technology for pumping CO_2 in deep reservoir for enhancing oil recovery. The same technology can be used to pump CO_2 in depleted hydrocarbon fields or saline aquifers, such that only a major concern remains: the long term performance of confining fields.

In sedimentary basins the long term assessment of CO_2 sequestration requires a careful characterization and modeling of hydraulic property variations of the clayey confining units overlying brine aquifers acting as storage units. In the Northern Adriatic sedimentary basin discontinuities of the sealing properties are typically due to small fractures and sandy inclusions. These features creates preferential pathways for CO_2 escaping. Unfortunately, a very limited amount of information is typically available for characterizing the earth's subsurface, mostly consisting in well-logs and seismic data, such that the deterministic mapping of hydraulic property variations is an elusive goal which is abandoned in favor of a geostatistical model of spatial variability. In this paper we develop a systematic approach to stochastically generate realistic representations of the confining units by embedding high permeability inclusions in a otherwise homogeneous low permeability matrix. The realizations are conditioned to available data, such that uncertainty is reduced in relation to the amount information provided by the measurements. The development of a geostatistical approach based on the Indicator Kriging is presented and a few sample tests are used to show the applicability of the proposed approach to realistic problems.

2. GEOSTATISTICAL PROCEDURE FOR THE BINARY FIELD GENERATION

2.1. **Unconditional binary field generation.** Let us consider a two-dimensional bimodal formation composed of sandy (highly conductive) inclusions embedded in an otherwise homogeneous low conductivity matrix. Since the thickness of the confining units is small compared to their areal extension we assume that the permeability k is isotropic and constant in the vertical direction. The spatial variability of k in the horizontal direction is modeled as a binary field [Goovaerts, 1997; Deutsch and Journel, 1998; Deutsch, 2002]:

$$k(x) = k_1\, I(x) + k_2\, [1 - I(x)] \tag{1}$$

[1] GtC: Giga tons of carbon. 1 GtC = 3.67 $GtCO_2$

where k_1 and k_2 are the permeability of the matrix and the inclusions, respectively, and $x \in \mathbb{R}^2$. The function $I(x) : \mathbb{R}^2 \to \{0, 1\}$ is a binary indicator which is set to 0 in the inclusions (lithotype or facies 2) and to 1 in the matrix (lithotype or facies 1). The indicator function is statistically defined by its mean μ and covariance function $C(h)$, (or its semivariogram $\gamma(h) = C(0) - C(h)$), where h is the two-point lag. It can be shown that the variance σ^2 of the indicator is given by [Papoulis and Pillay, 2002]:

$$\sigma^2 = \mu\,(1 - \mu) \tag{2}$$

Note that the value of μ represents the probability of occurrence of the facies 1. In other words, at a given position x within the computational domain we have the probability μ of observing the matrix and $1 - \mu$ of observing the inclusion.

2.1.1. *Numerical generation procedure.* The binary permeability field (1) is generated by using HYDRO_GEN, a correlated random field generator developed by [Bellin and Rubin, 1996], suitably adapted to the generation of binary random functions. We consider a rectangular domain covered by a structured grid formed by N nodes and we denote by x_1 the node located at the bottom left corner. A random number $0 \leq \varepsilon_1 \leq 1$ is generated from a uniform distribution and the corresponding indicator is built as:

$$I(x_1) = \begin{cases} 1 & \text{if } \varepsilon_1 \leq \mu \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The probability of observing the facies 1 at point x_2 must be conditioned to $I(x_1)$ as follows:

$$P^c(x_2) = \mu + \lambda_1(x_2)\ [I(x_1) - \mu] \tag{4}$$

where the weight $\lambda_1(x_2)$ is given by $\lambda_1(x_2) = C(x_1, x_2)/C(x_1, x_1)$. The indicator $I(x_2)$ is calculated by replacing μ with $P^c(x_2)$ in equation (3). This procedure is repeated visiting row by row and from left to right the remaining grid points as suggested in the work by Bellin and Rubin [1996]. At the generic point x_n the conditional probability $P^c(x_n)$ is given by:

$$P^c(x_n) = \mu + \sum_{j=1}^{n-1} \lambda_j(x_n)\ [I(x_j) - \mu]\,; \qquad n = 1, \ldots, N \tag{5}$$

where the coefficients λ_j $(j = 1, 2, \ldots, n-1)$ are the solution of the following simple Kriging problem represented by the following linear system:

$$\sum_{j=1}^{n-1} \lambda_j\, C(x_j, x_l) = C(x_l, x_n); \qquad l = 1, \ldots, n-1 \tag{6}$$

This procedure has been implemented within the code HYDRO_GEN [Bellin and Rubin, 1996].

2.2. **Conditional binary field generation.** Given M points with coordinates x_l, $l = 1, \ldots, M$, where the lithotype has been observed, we generate a random indicator field conditional to the observations by following the procedure described in the book of Kitanidis [1997]. According to this procedure an unconditional field $I(x)$ is first generated over a regular grid including the observation nodes; then the following residuals are computed as $\widetilde{I}(x_l) = \overline{I}_l - I(x_l)$, where $\overline{I}_l$ is the known value of the indicator at the observation

TABLE 1. Parameters of the computational grid.

Parameter	Symbol	Value
Integral scale	λ_I	1.0
Grid spacing	Δx, Δy	$0.25\,\lambda_I$
Field dimension	l_x, l_y	$40\,\lambda_I$
Mean	μ	0.95
Variance	σ^2	0.0475
Covariance	$C(h)$	$\sigma^2\, e^{-h}$

node x_l. The residuals, which are assumed to possess the same spatial correlation of the unconditioned indicator $I(x)$, are then distributed over the computational grid and added to the unconditional field:

$$\widetilde{P}^c(x_n) = P^c(x_n) + \sum_{l=1}^{M} \lambda'_l(x_n)\,[\overline{I}_l - P^c(x_l)]; \qquad n = 1, \ldots, N \tag{7}$$

where, λ'_l are the weights obtained by solving a Kriging system similar to (6), but using the M observation nodes instead of the n grid nodes belonging to the search neighborhood of the generation procedure, such that n should be replaced with M, and N is the total number of grid nodes. Note that in equation (6) $P^c(x_n)$ is the probability of observing the facies 2 at the grid node x_n, conditional to the observations at the locations x_l. Finally, the value of the indicator at the location x_n is assigned according to the equation (3), by replacing x_1 with x_n, ε_1 with ε_n, and μ with $\widetilde{P}^c(x_n)$, where ε_n is still drawn from a uniform distribution.

It can be shown that for $x_n \to x_l$ the conditional probability (7) converges to $\widetilde{P}^c(x_n) \to \overline{I}_l$, while for $|x_n - x_l| >> \lambda_I$, where λ_I is the integral scale of the indicator field, we obtain that $\sum_{l=1}^{M} \lambda'_l(x_n)\,[\overline{I}_l - P^c(x_l)] \approx 0$, showing that the conditional probability of the indicator function (7) at point x_n converges to $P^c(x_n)$, i.e. the unconditional value. As a consequence, the conditioned field built as described above does not change with respect to the unconditional field.

3. NUMERICAL EXAMPLES

The geostatistical approach described above is tested by comparing the a posteriori statistics of the binary random field with the corresponding imposed statistics summarized in Table 1. The posterior statistics are computed by averaging over N_r independent Monte Carlo realizations of the random binary field and the comparison is limited to the mean:

$$\overline{\mu} = \frac{1}{N_r} \sum_{i=1}^{N_r} \widehat{\mu}_i \tag{8}$$

where $\widehat{\mu}_i$ is the mean value of the $i-$th Monte Carlo realization, and the covariance function (or the semivariogram):

$$\overline{C}(h) = \frac{1}{N_r} \sum_{i=1}^{N_r} \widehat{C}_i(h); \qquad \text{or} \qquad \overline{\gamma}(h) = \frac{1}{N_r} \sum_{i=1}^{N_r} \widehat{\gamma}_i(h) \tag{9}$$

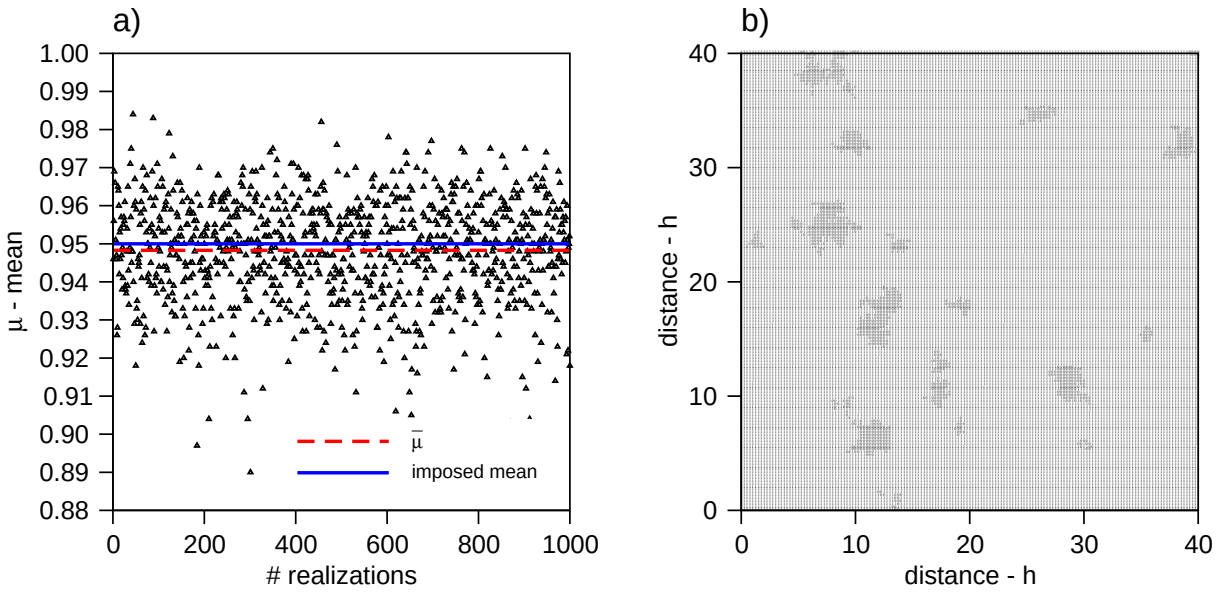

FIGURE 1. Comparison between computed mean over all the 1000 realizations and the imposed mean (a). Example of a unconditioned binary field realization (b).

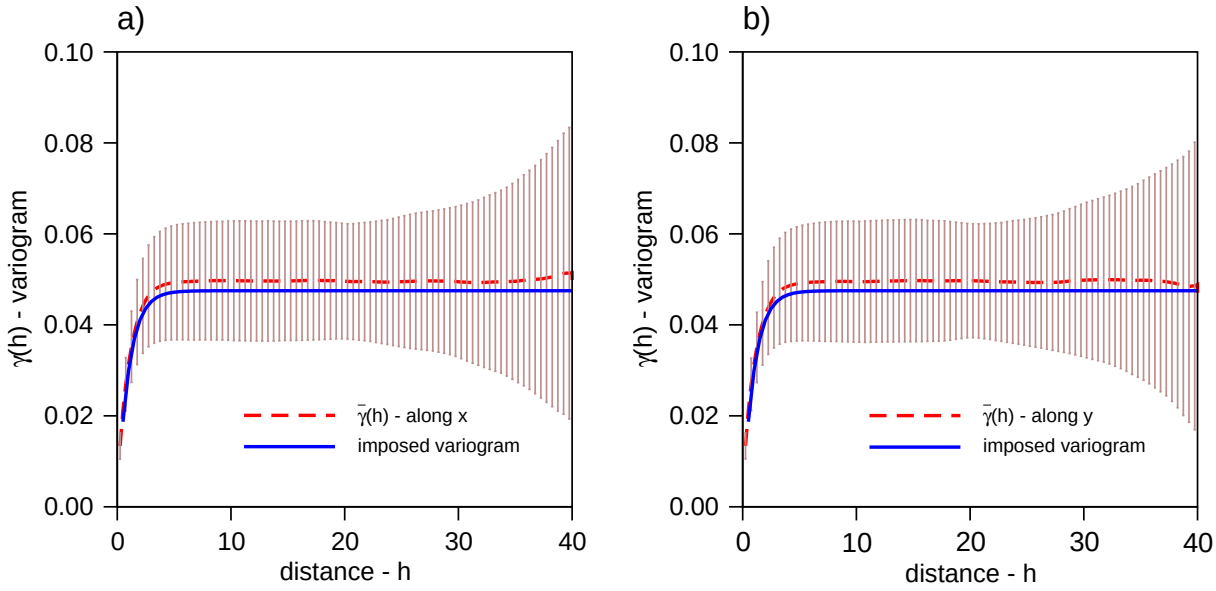

FIGURE 2. Comparison between the average variogram of the 1000 unconditioned generations and the imposed variogram along the x-direction (a) and the y-direction (b), respectively. Vertical bars denote the parameter Γ as defined above.

TABLE 2. An example of lithotype observations coded into the indicator variable $\overline{I}$.

x	1.0	1.0	5.0	10.0	15.0	20.0	25.0	30.0	35.0	38.0
y	1.0	3.0	30.0	37.0	15.0	20.0	20.0	25.0	10.0	5.0
$\overline{I}$	1.0	0.0	1.0	1.0	0.0	1.0	0.0	1.0	1.0	0.0

where $\widehat{C}_i(h)$ and $\widehat{\gamma}_i(h)$ are respectively the spatial covariance function and the semivariogram of the $i-$th Monte Carlo realization, both computed assuming statistical stationarity. Furthermore, h indicates the two-point separation distance, and N_r is fixed to a value that ensures convergence of both $\overline{C}(h)$ and $\overline{\gamma}(h)$, such that no differences are appreciated in both $\overline{C}(h)$ and $\overline{\gamma}(h)$ by a further increase of N_r. Computations are performed with N_r =1000. Once $\overline{C}(h)$ (or $\overline{\gamma}(h)$) is known along a direction, for example along the coordinate directions, the deviation between the entire set of back-calculated variograms and average variogram can be defined as:

$$\sigma_\gamma(h) = \sqrt{\frac{1}{N_r}\sum_{i=1}^{N_r}\left(\overline{\gamma}(h) - \widehat{\gamma}_i(h)\right)^2} \tag{10}$$

3.1. **Unconditional generation.** In this example we utilized N_r=1000 Monte Carlo realizations of the binary random field with the statistical properties shown in Table 1. Figure 1 (b) shows an example of the binary random field. The mean $\overline{\mu}$ as defined through by the equation (8) is plotted in Figure 1 (a) together with the expected mean (continuous line) and the spatial means in each one of the 1000 independent Monte Carlo realizations. The mean is in a good match with the expected value and no correlation is observed between the spatial means in different realizations, showing the absence of bias in the proposed methodology. A good agreement is also observed between $\overline{\gamma}(h)$ and the theoretical semivariogram as shown in Figures 2 (a) and (b) along the longitudinal and the transverse directions, respectively. Furthermore, the vertical bars indicate the interval of confidence of $2\,\sigma_\gamma(h)$, which increases in amplitude with the two-point separation distance since the number of pairs utilized for computing the spatial semivariogram decreases as h increases.

3.2. **Conditional generation.** An example of random binary field conditioned to the observations listed in the Table 2 is shown in Figure 3 (b). The overall behavior is in line with the observations, although the inclusions are not well delimited as in the unconditional field of Figure 1 (b). The origin of the scattering observed at the fringes of the inclusions needs to be studied very carefully because it may introduce spurious effect into the flow simulations. Figures 3 (a) 4 (a) and (b) show that, because of the observations, the theoretical statistics are not reproduced by the ensemble of conditional fields. This is because conditioning on the observations is equivalent to excluding from the plausible representations of the reality the realizations not compatible with the observations, and this is reflected through the conditional moments (estimated by averaging over the N_r independent Monte Carlo realizations), which differ from the theoretical values.

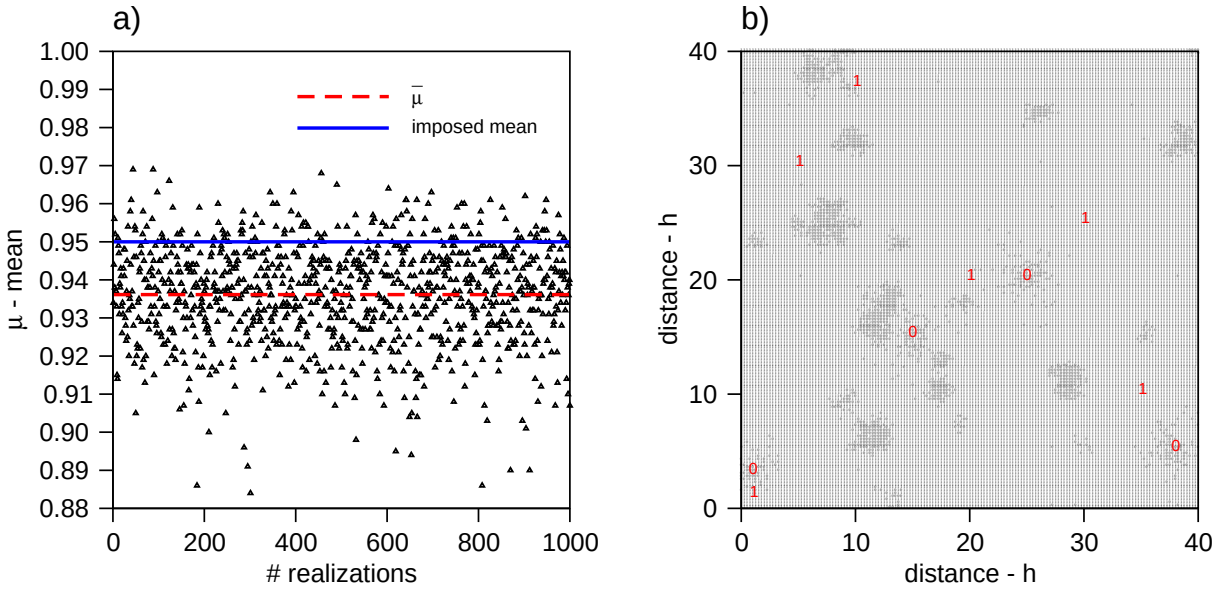

FIGURE 3. Comparison between computed mean over all the 1000 realizations and the imposed mean (a). Example of a conditioned binary field realization (b).

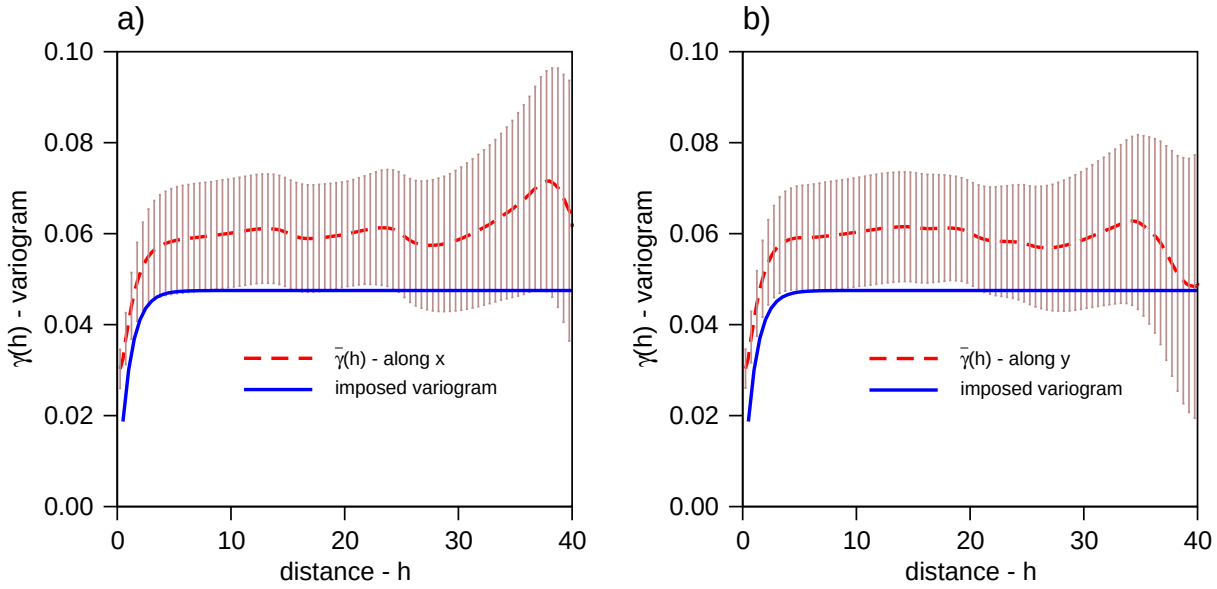

FIGURE 4. Comparison between calculated and imposed variograms of the 1000 conditioned generations along the x-direction (a) and the y-direction (b), respectively. Vertical bars denote the parameter Γ as defined above.

4. CONCLUSIONS

We presented a geostatistical approach for generating defects in the caprock of deep brine aquifers in the North Adriatic sea. The defects are in the form of localized reduction of the permeability (inclusions) which may compromise the possibility of using such

formations for CO_2 sequestration. The proposed methodology is easy to apply and is allows a straightforward inclusions of observations in borehole, from geophysical surveys and soft information from geological mapping. Using the generated caprock structures, conditional to the available information, it will be possible to obtain a reliable estimate of the carbon dioxide fluxes that may escape through these preferential flow paths. Using a Monte Carlo approach it would then be possible to assess the risk of CO_2 escape through layered sedimentary basins such as the Upper Adriatic Geological formations. Our methodology is shown to generate realistic representations of the permeability distribution in the caprock, which may be used to evaluate the amount of CO_2 that most probably may escape to the atmosphere and the associated interval of confidence. However, the observed "dispersion effect", by which a few scattered points appear around measurement locations when conditioning is performed, needs to be further studied.

Acknowledgments. This work has been partially supported by the Italian MIUR-PRIN projects "Numerical models for multiphase flow and deformation in porous media" (first and third authors) and "Nutrient and contaminant transport at the catchment scale: models for the management and the protection of water resources" (second author).

References

Bellin, A. and Rubin, Y. (1996). Hydro_gen: A new random number generator for correlated properties. *Stochastic Hydrology and Hydraulics*, **10**(4), 253–278.

Deutsch, C. V. (2002). *Geostatistical Reservoir Modeling*. Applied Geostatistics Series. Oxford University Press.

Deutsch, C. V. and Journel, A. G. (1998). *GSLIB: geostatistical software library and user's guide*. Applied Geostatistics Series. Oxford University Press.

Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Applied Geostatistics Series. Oxford University Press.

Keeling, C. D. and Whorf, T. P. (2004). Atmospheric CO_2 records from sites in the SIO air sampling network. In *Trends: A compendium of data on global change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Tenn., USA. http://cdiac.esd.ornl.gov/trends/trends.htm.

Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications in Hydrogeology*. Cambridge University Press.

Ledley, T. S., Sundquist, E. T., Schwartz, S. E., Hall, D. K., Fellows, J. D., and Killeen, T. L. (1999). Climate change and greenhouse gases. *EOS, Trans. AGU*, **80**(39), 453–458.

Marland, G., Boden, T. A., and Andres, R. J. (2003). Global, regional, and national fossil fuel CO_2 emissions. In *Trends: A compendium of data on global change*, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Tenn., USA. http://cdiac.esd.ornl.gov/trends/trends.htm.

Papoulis, A. and Pillay, S. U. (2002). *Probability, Random Variables and Stochastic Processes*. McGraw Hill, 4th edition edition.

UNFCCC (1997). Kyoto protocol to the United Nations framework convention on climate change. (FCCC/CP/1997/L.7/Add. 1).