Bayesian approach to environmental problem based on PFLOTRAN package

Orest Dorosh[1], Henryk Wojciechowicz[1], Piotr Kopka[1]

Bayesian Young Statisticians Meeting (BAYSM), Milan June, 5-6, 2013
Paper no. 11

[1] National Centre for Nuclear Research, Nuclear Energy Department (EJ1), Swierk-Otwock, Poland
orest.dorosh@ncbj.gov.pl
henryk.wojciechowicz@ncbj.gov.pl
piotr.kopka@ncbj.gov.pl

Abstract

Presented result of applying Bayes interface for searching an isolated source of contamination in soil using program PFLOTRAN.

Keywords: CFD, PFLOTRAN, Bayesian approach, Environmental hazards.

1 Introduction

Due to the progress in the development of computing clusters and computational algorithms quite widely is used dedicated software for modelling problems related to the transport and dispersion of contamination in the environment [1], allowing for simulation physical processes in big domains and high resolution grids.

In particular, for the dispersion of contamination in the porous media we have used program PFLOTRAN a massively parallel 3-D reservoir simulator [2, 3], developed at LANL/ORNL, USA. PFLOTRAN can model multiphase reactive flows in geologic formations based on continuum scale mass and energy conservation equations. It employs the PETSc (Portable, Extensible Toolkit for Scientific Computation) a numerical modular package and efficient Newton-Krylov solver framework. We present here a computational problem related to the identification of source of contamination in porous media, by solving inverse problem, basing on the Bayesian approach.

2 Problem

Our goal is to localize source of contamination in soil with underground water flow using information from data of set of sensors. For this inverse problem we use statistic method [4] coupled with software package PFLOTRAN.
The area of calculation is a rectangular parallelepiped of size $5000 \times 2500 \times 100$ meters. For our invers problem we generated syntetic data by PFLOTRAN. Then statistic methods are applied for localization of source comparing dato in sencor position.

3 Bayesian approach

As "experimental data" we used the results of calculations with position of the well (the contamination source) at place $(x, y) = (1050\,\text{m}, 1350\,\text{m})$. The map of contamination concentration is presented on figure 1,a. The contamination is spread from lower part of well in depth from 20 to 65 meters below level of ground. The values of contaminate concentration taken in eight different points of calculation domain are considered as "experimental data" (on the figure 1, a sencors are marked as green crosses).
The algorithm of source searching consists of the following steps:

1. For prior distribution choose set of possible well positions. The probability for choosing the possitionn of well is proportional to prior probability distribution for the position.

2. Perform simulation with PFLOTRAN for each of well positions.

3. Compare the data from sensors set with the "experimental data" in "sensors" location, and take some subset (ten in our case) locations of sources that have lowest value of fitting function.

4. Update and normalize posterior distribution and use it as a prior distribution for the next stage.

For every stage we performed 50 simulations with different locations of contamination source. With fitting function $f = \sum_{i=1}^{8} (\eta^i_{obs.} - \eta^i_{calc.})^2$ (where η is a concentration of contaminats, and summation is by sensors) we choose ten sources locations to calculate posterior distribution, according to the formula:

$$\rho_{posteriors}(\vec{r}) = \sum_{i=1}^{10} \rho_{prior}(\vec{r}) \cdot \exp\left(-\frac{(x - x_i)^2}{2\sigma_x^2} - \frac{(y - y_i)^2}{2\sigma_y^2} - \frac{(z - z_i)^2}{2\sigma_z^2} \right) \quad (1)$$

where (x_i, y_i, z_i) is the location of source from ten best. The variances are: $\sigma_x = 500\,\text{m}$, $\sigma_y = 500\,\text{m}$ and $\sigma_x = 10\,\text{m}$. For the first stage we used uniform prior distribution.

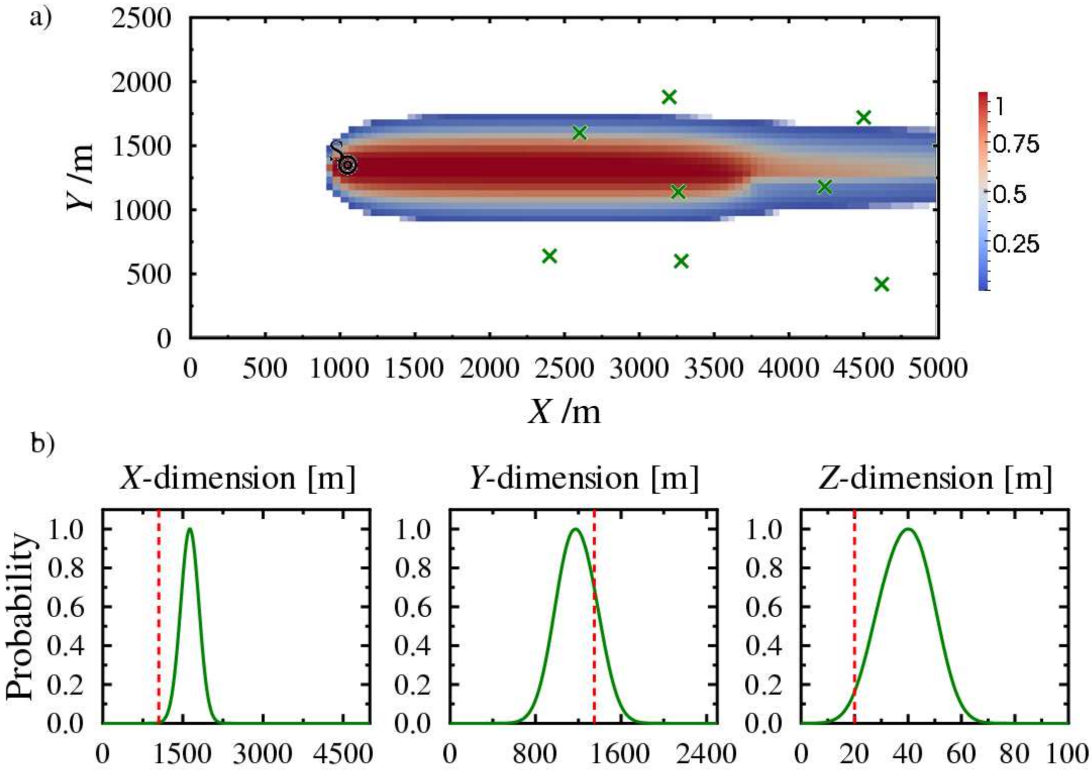

Figure 1: a) XY plane of calculation domain. **S** is a position of actual source of contamination. Green crosses are the position of the sensors. b) Three components of posterior distribution. Red dashed line — actual values of source position components.

4 Results and discussion

The resulting distribution (after 11 iteration steps) gives the most probable values for source position $(x, y, z) = (1640\,\text{m}, 1180\,\text{m}, 40\,\text{m})$ (figure 1, b). The difference with actual source location $\Delta\vec{r} = (590\,\text{m}, -170\,\text{m}, 20\,\text{m})$. The big difference for the X component can be explaned by the symmetry of the problem. Some sensors are out of the contamination flow (figure 1, a) and do not contribute in the process of the distribution recalculation. But information from simulation we performed can be helpful in putting sensors in new places for better collecting measurement data.

References

[1] P. Gousseaua, B. Blockena, T. Stathopoulosb, G.J.F. van Heijstc. **CFD simulation of near-field pollutant dispersion on a high-resolution**

grid: A case study by LES and RANS for a building group in downtown Montreal. *Atmospheric Environment*; 2011; 45(2); pp. 428-438.

[2] `http://ees.lanl.gov/pflotran/`

[3] **Quick Reference Guide: PFLOTRAN 2.0**. March 29, 2011.

[4] Laura C. Thomsona, Bill Hirst, Graha Gibsona, Steve Gillespie, Philip Jonathan, Kenneth D. Skeldona, Miles J. Padgett. **An improved algorithm for locating a gas source using inverse methods**. *Atmospheric Environment*; 2007; 41; pp. 11281134.