A PARALLEL MARKOV CHAIN MONTE CARLO METHOD FOR PROBABILISTIC CHARACTERIZATION OF POLLUTANT SOURCES IN GROUNDWATER SYSTEMS

Kumar Mahinthakumar, North Carolina State University, 919-303-1403, gmkumar@ncsu.edu

1. G. (Kumar) Mahinthakumar, North Carolina State University
2. Sarat Sreepathi, North Carolina State University
3. Shruthi, Sarode

Groundwater pollutant source characterization is an important problem for effective groundwater management and remediation. Inverse modeling is frequently used to solve groundwater source identification problems from downstream contaminant concentration measurements. Due to the inherent uncertainty in hydraulic conductivity distribution it is difficult to characterize a s ource deterministically from given flow conditions and concentration measurements. However, it is possible to obtain a probabilistic distribution of parameters characterizing the pollutant source if the uncertainty in hydraulic conductivity can be quantified. One can formulate this problem using Bayes' theorem by expressing the posterior probability distribution of the parameters as a product of prior distribution with a likelihood function. A popular method to sample the posterior distribution is the Markov chain Monte Carlo (MCMC) method. There has been very little work published in the use of MCMC methods for groundwater source identification problems (the only known work to our knowledge is by Prof. Kitanidis group at Stanford) due to the computationally prohibitive nature of MCMC for this application; many groundwater flow and transport simulations are required for the computation of each point in the chain and many points may be required for sampling a large parameter space. While parallel computation could be used to alleviate this problem, parallelism is typically limited to the independent groundwater flow and transport simulations involved in computing the covariance matrix of the likelihood function or fine grained parallelism within a single groundwater simulation as the MCMC method requires that each point in the chain be computed sequentially.

We propose a parallel MCMC method where multiple parallel chains are created and evolved using a niched co-evolutionary strategy that minimizes redundancy among chains. Parallelism is also exploited in the covariance matrix computation within each chain. This multilevel parallelism strategy will facilitate scaling computations to a large number of processors as would be required on leadership class facilities such as the Jaguar PF supercomputer at the Oak Ridge National Laboratory. Preliminary results for a simple two-dimensional test problem on a small number of processors show promise of the algorithm. In this case, a single source identification problem involving 5 parameters characterizing the source (two coordinates of the source location, release concentration, release time, and release duration) is solved for a random Gaussian hydraulic conductivity field using concentration measurements at 6 monitoring wells. The problem was solved using 64 processor cores of a departmental cluster at North Carolina State University in a MATLAB parallel computing environment. Work is currently underway to implement this algorithm for three-dimensional problems using Fortran90 and MPI for leadership class systems.