Contaminant source localization via batch-sequential Bayesian global optimization

G. Pirot, University of Lausanne, Switzerland
T. Krityakierne, Mahidol University, Bangkok, Thailand
D. Azzimonti, Istituto Dalle Molle di studi sull’Intelligenza Artificiale, Switzerland
D. Ginsbourger, Idiap Research Institute and University of Bern, Switzerland
P. Renard, University of Neuchâtel, Switzerland

Key words: Surrogated-based optimization, Gaussian Processes, parallelization, aquifer, inverse problem.

Introduction

Driven by the concept of polluter pays, the topic of contaminant localization is not recent. One possibility to localize the source of a contaminant is to consider an inverse problem formulation where the source location and release history are inferred from concentration samples. In that case, reasonable information about the contaminant plume is required and the method can be adapted to uncertain geology. Here we build upon [1] and appeal to Bayesian optimization, a powerful approach that smartly explores the parameter space relying on figures of merit such as the Expected Improvement (EI) criterion [2], that trade off exploitation of available results and space exploration. We will present the results of [1], where the efficiency of the method was assessed on synthetic cases presenting realistic hydrogeological contrasts and connected structures, and complement them with novel investigations on the “batch-sequential” case where several simulations are run at each iteration.

Efficient global optimizer based on expected improvement criteria

The employed expected improvement algorithms (Figure 1) are initialized by performing $n_0$ evaluations of the objective function $f$ (possibly in parallel) at locations defined by a space filling design (e.g., a Latin Hypercube Design). Then, at each iteration, based on the $n$ previous evaluations of the objective function, the objective function $f$, approximated by a Gaussian Process (GP), is predicted at unexplored locations of the decision space by ordinary kriging with a Matérn covariance function (here $\nu = 3/2$, in the notation of [3]).

![Figure 1: Illustration of the first two iterations of a sequential EI algorithm for one scenario; the sub-figures in the left column illustrate the prediction mean of $f$ over the two-dimensional decision space at each iteration; the blue dots indicate the decision space locations where $f$ was previously evaluated; the sub-figures in the center column illustrate the prediction variance of $f$ over the two-dimensional decision space at each iteration; the sub-figures in the right column illustrate the expected improvement map over the two-dimensional decision space at each iteration; the red dot denotes the decision space location with the maximum EI value. Adapted from [1].](image-url)
the 2D decision space represent the coordinates of the potential contaminant source. The mean prediction and the associated standard deviation are used jointly to compute the EI criterion. The maximum value on the EI map indicates the next point of the decision space to explore, i.e. where the objective function \( f \) will be evaluated at the next iteration. The number of explored locations \( n \) in the decision space is incremented. Here, a budget depletion stopping criterion is used (on this example the algorithm stops after 100 evaluation of \( f \) have been performed). We will discuss and illustrate how such algorithm can be parallelized, and investigate efficiency gains depending on controllable algorithmic settings.

**Results and conclusions**

So far the purely sequential method was tested on synthetic cases generated for two different geologies and two sources of contaminant [1]. Flow and transport simulations were performed for each case. Observations of concentrations as function of time were recorded at 25 downstream monitoring wells. The objective function \( f \) was defined as the \( \ell^2 \) norm of the misfit between observations and simulations. To assess the efficiency of the method, the exploration was replicated for 100 different initial designs. Up to a prescribed tolerance the minimum of the objective function was generally reached in less than 50 evaluations (Figure 2). Novel investigations will illustrate how parallelization may increase efficiency of the overall approach.

![Figure 2: Solution exploration results for 2 out of 4 scenarios over the cost functions; A for geology 1 and contaminant initial location at (100, 10); B for geology 2 and initial contaminant location at (89, -36). Adapted from [1].](image)

**References**

