

Abstract

Geophysics and hydrogeology are scientific disciplines that play important roles in understanding the subsurface environment. By proposing and implementing new methods to explore and characterize the subsurface, it is possible to make predictions, manage resources and mitigate hazards in better ways. The data acquired in geophysical and hydrogeological surveys are used to infer relevant properties of the subsurface by solving inverse problems. A probabilistic framework to solve these problems that accounts for different uncertainty sources is provided by Bayesian inversion. Unfortunately, the estimation of the posterior probability density function (PDF) provided by widely used standard Markov chain Monte Carlo (MCMC) methods can be limited when inverse problems are non-linear and high-dimensional. In practice, they may fail to properly sample the posterior PDF in allowed computational times by getting trapped in local minimas. Alternative methods relying on tempering overcome these difficulties by enhancing the freedom of exploration using a temperature variable reducing the influence of the likelihood function in Bayes' theorem. Sequential Monte Carlo (SMC) is a group of methods that build a sequence of importance sampling steps between distributions with gradually decreasing temperatures, going from the prior to the posterior PDF. These particle approaches use a set of evolving particles and their associated weights to approximate the posterior PDF. An appealing property of SMC methods is that they also offer an estimation of the evidence (marginal likelihood), a key quantity for comparing and ranking competitive conceptual models. The challenge of SMC and other tempering methods of pre-defining a suitable sequence of temperatures is overcome by adaptive SMC (ASMC) that adaptively finds the most suitable temperature decrement. In this thesis, we implement the ASMC method in challenging synthetic geophysical and hydrogeological test cases where the prior information is encoded either by deep generative neural networks or training images explored by multiple-point statistics methods to handle geologically-realistic prior PDFs. We find that ASMC is considerably more efficient to locate and sample the posterior PDF than a state-of-the-art adaptive MCMC method and parallel tempering, a popular MCMC method relying on tempering. We demonstrate that the ASMC evidence estimations are robust regarding proposal schemes and inversion parameters. Still, ASMC is computationally costly when considering expensive forward solvers in geophysics and hydrology. To seek important computational gains while targeting the original posterior PDF and evidence, we propose a multifidelity ASMC approach relying on surrogate

solvers that are updated as the inversion progresses. For a multivariate Gaussian test example, we show that this method accelerates the inversion while reaching similar accuracy as methods relying on high-fidelity solvers only. We find that ASMC is very powerful to solve complex geophysical and hydrogeological inverse problems, especially when non-linearity and multi-modality challenge standard MCMC approaches.

Key words : Bayesian inversion, Bayesian model selection, sequential Monte Carlo, geostatistics, geophysics, hydrogeology, multifidelity methods.

Résumé

La géophysique et l'hydrogéologie sont des disciplines scientifiques qui jouent un rôle important dans la compréhension de l'environnement souterrain. En proposant et en mettant en œuvre de nouvelles méthodes pour explorer et caractériser le sous-sol, il est possible de faire des prévisions, de gérer les ressources et d'atténuer les risques de manière plus efficace. Les données acquises lors des études géophysiques et hydrogéologiques sont utilisées pour déduire les propriétés pertinentes du sous-sol en résolvant des problèmes inverses. L'inversion bayésienne offre un cadre probabiliste pour résoudre ces problèmes en tenant compte des différentes sources d'incertitude.

Malheureusement, l'estimation de la fonction de densité de probabilité a posteriori fournie par les méthodes standard de Markov chain Monte Carlo (MCMC) largement utilisées peut-être restreinte lorsque les problèmes inverses sont non linéaires et de haute dimension. Dans la pratique, elles risquent de ne pas parvenir à échantillonner correctement la fonction de densité de probabilité a posteriori dans les temps de calcul autorisés en se retrouvant piégées dans des minima locaux. D'autres méthodes reposant sur le tempering permettent de surmonter ces difficultés en augmentant la liberté d'exploration à l'aide d'une variable de température réduisant l'influence de la fonction de vraisemblance dans le théorème de Bayes. La méthode sequential Monte Carlo (SMC) est un groupe de méthodes qui construisent une séquence d'étapes d'échantillonnage préférentiel entre des distributions dont les températures diminuent progressivement, en passant de la fonction de densité de probabilité a priori à la posteriori. Ces approches particulières utilisent un ensemble de particules évolutives et leurs poids associés pour approximer la fonction de densité de probabilité a posteriori. Une propriété intéressante des méthodes SMC est qu'elles offrent également une estimation de l'évidence (vraisemblance marginale), une quantité clé pour comparer et classer des modèles conceptuels compétitifs. Le défi du SMC et d'autres méthodes de tempering consistant à prédéfinir une séquence appropriée de températures est surmonté par le SMC adaptatif (ASMC) qui trouve de manière adaptative la décroissance de température la plus

appropriée. Dans cette thèse, nous mettons en œuvre la méthode ASMC dans des cas d'essai géophysiques et hydrogéologiques synthétiques difficiles où les informations préalables sont encodées soit par des réseaux génératifs de neurones profonds, soit par des images d'entraînement explorées par des méthodes statistiques à points multiples pour traiter des priors géologiquement réalistes. Nous constatons que la méthode ASMC est considérablement plus efficace pour localiser et échantillonner la fonction de densité de probabilité a posteriori qu'une méthode MCMC adaptative et que le *parallel tempering*, une méthode MCMC populaire reposant sur le *tempering*. Nous démontrons que les estimations de l'évidence par ASMC sont robustes en ce qui concerne les schémas de proposition et les paramètres d'inversion. Cependant, la méthode ASMC est coûteuse en termes de calcul si l'on considère les solveurs utilisés en géophysique et en hydrologie. Pour obtenir des gains de calcul importants tout en ciblant la fonction de densité de probabilité a posteriori et l'évidence originale, nous proposons une approche ASMC multifidélité reposant sur des solveurs substitués qui sont mis à jour au fur et à mesure que l'inversion progresse. Pour un exemple de test gaussien multivarié, nous montrons que cette méthode accélère l'inversion tout en atteignant une exactitude similaire à celle des méthodes reposant uniquement sur des solveurs de haute fidélité. Nous constatons que la méthode ASMC est très puissante pour résoudre des problèmes inverses géophysiques et hydrogéologiques complexes, en particulier lorsque la non-linéarité et la multimodalité défient les approches MCMC standard.

Mots clefs : Inversion bayésienne, sélection de modèles bayésiens, Monte Carlo séquentiel, géostatistique, géophysique, hydrogéologie, méthodes multifidélité.