Bayesian Calibration of Geothermal Reservoir Models via Markov Chain Monte Carlo

Tiangang Cui

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Engineering Science, The University of Auckland, 2010.

Abstract

The aim of the research described in this thesis is the development of methods for solving computationally intensive computer model calibration problems by sample based inference. Although our primary focus is calibrating computer models of geothermal reservoirs, the methodology we have developed can be applied to a wide range of computer model calibration problems.

In this study, the Bayesian framework is employed to construct the posterior distribution over all model parameters consistent with the measured data, accounting for various uncertainties in the calibration process. To construct the posterior distribution for computer model calibration problems, several methods such as the additive bias framework of Kennedy and O'Hagan (2001) and the enhanced error model (Kaipio and Somersalo, 2007) are investigated.

Then, the solutions of computer model calibration problems are given by estimating the expected value of statistics of interest over the posterior distribution. Markov chain Monte Carlo (MCMC) sampling, Metropolis-Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) in particular, is empoyed to explore the posterior distribution, and Monte Carlo integration is used to calculating the expected values.

To be able to automatically adjust the proposal densities used in the MH algorithm, the state-of-the-art adaptive MCMC algorithms (Haario et al., 2001; Atchade and Rosenthal, 2005; Andrieu and Moulines, 2006; Roberts and Rosenthal, 2007) are investigated in this thesis. A group components adaptive Metropolis (GCAM) algorithm has been designed, which combines the features of the adaptive Metropolis algorithm (Haario et al., 2001) and the adaptive Metropolis-within-Gibbs algorithm

(Roberts and Rosenthal, 2009) to improve both statistical and computational efficiencies of sampling involving computationally demanding target densities.

In sampling the posterior distribution of the computer model calibration problem, the main computational cost of the MH algorithm is that the posterior density has to be evaluated at each iteration. The delayed acceptance MH (DAMH) algorithm (Christen and Fox, 2005) can be used to speed up the sampling. An approximate posterior distribution is necessary to run the DAMH algorithm, and the enhanced error model (Kaipio and Somersalo, 2007) provides a potential candidate.

However, constructing the enhanced error model requires evaluating the coarse model and the fine model for a large number of input points *a priori*. Also, the enhanced error model built over the prior distribution is not accurate enough, and does not show desirable statistical efficiency in the DAMH algorithm .

To overcome this difficulty, we combine the DAMH algorithm, the enhanced error model and techniques in the adaptive MCMC together to give an adaptive DAMH (ADAMH) algorithm, which allows construction of the enhanced error model from the posterior distribution adaptively. Sufficient conditions for the algorithm to converge to the target distribution are provided, and several adaptive approximations have also been designed under these conditions.

For test cases based on a well discharge test model with a synthetic data set and a measured data set, the ADAMH algorithm and the adaptive approximations show significantly better computational and statistical efficiencies than the non-adaptive DAMH algorithm. For the synthetic data set, the approximations are compared under various schemes, and best statistical efficiency is achieved by using the full covariance matrix in the approximate posterior distribution. For both the synthetic data set and the measured data set, the model predictions follow the data reasonably well.

We are able to run the ADAMH algorithm for 9,000 iterations in about 30 days time for a 3D natural state geothermal reservoir model. This originally would cost the standard MH algorithm about eight months of computing time. The calibration result shows good agreement between the estimated temperature profiles and the measured data.

Acknowledgements

During my PhD, I have had cause to be grateful for the advice, support and understanding of many people. In particular, my supervisors, Colin Fox and Michael O'Sullivan, have been constant sources of good ideas and sound advice. Without their guidance, encouragement, and support I could never have accomplished what I did. I am truly grateful for them taking me as their student and giving me the opportunity to work with them during these years.

I have benefited from the numerous useful discussions and other productive conversations with other past and present members of the Department of Engineering Science staff and the Department of Mathematics staff, including Adrian Croucher, Jari Kaipio, Geoff Nicholls, Al Parker, Angus Yeh and Sadiq Zarrouk, to all of whom I am very grateful. Especially, I wish to express my gratitude to Geoff Nicholls, for introducing me into the field of Bayesian inference. I would also like to thank Susie Bayarri for her advice and many useful discussions during the period I visited the SAMSI.

Furthermore, it is a pleasure to thank Hyuck Chung, Emily Clearwater, Wen Duan, Ziming Guan, Imran Ishrat, Yilin Jia, Eylem Kaya, Charles Moliere, Juliet Newson, John O'Sullivan, Al Parker, Andrea Raith, Sepideh Rastin, Christian Schwarzl, Seda Tolun, Yu Wang, Angus Yeh, Lei Zhang, and my other fellow students and post-docs who have made this thesis a much more rewarding experience than it would be without them, and who have also taken part in my personal and professional growth over these years.

I would also like to thank our administrative staff and IT support team, Shobha Herle, Nicola Kovacevich, Annette Warbrooke, Kim Williams, Sajy Augusty, Percy

Barboza, and Rao Cherukuri, for their patience and support over these years. Special thanks go in particular to Sandy Wilson, who helped me to settle down in the department during my multiple visits to the University of Otago.

I am extremely thankful to have received the generous support from the New Zealand Institute of Mathematics and its Application (NZIMA) during my study, without this financial support, this work would not have been possible. I give special thanks to Margaret Woolgrove for organizing all the paperwork for my scholarship every year.

Finally, special mention is deserved by my father and mother, Jianzhi Cui and Huijun Ge, who have encouraged and supported me in many different ways during this project, and of course, long before that.

Contents

List of Acronyms

ADAMH	adaptive delayed acceptance Metropolis-Hastings
AM	adaptive Metropolis
AMWG	adaptive Metropolis-within-Gibbs
CLT	Central Limit Theorem
DAMH	delayed acceptance Metropolis-Hastings
GCAM	grouped components adaptive Metropolis
GMRF	Gaussian Markov Random Field
GP	Gaussian process
GPS	Gaussian process surrogate, see Section 4.4
IACT	integrated autocorrelation time
MAP	maximum *a posteriori*
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
ML	maximum likelihood
MT	magnetotellurics
RBF	Radial basis function
SCAM	Single components adaptive Metropolis
SVD	Singular value decomposition
i.i.d.	Independent and identically-distributed
masl	Meter above the sea level