Thermodynamic integration via Replica Exchange Hamiltonian Monte Carlo for faster sampling and model comparison.

Damian M. Ndiwago[1], Remko Nijzink[3], Christophe Ley[2], Stanislaus J. Schymanski[3], Jack S. Hale[1].

[1] Institute of Computational Engineering, Faculty of Science, Technology and Medicine, University of Luxembourg, Esch–sur–Alzette, Luxembourg.
[2] Department of Mathematics, Faculty of Science, Technology and Medicine, University of Luxembourg, Esch–sur–Alzette, Luxembourg.
[3] Catchment and Ecohydrology Group (CAT), Environmental Research and Innovation (ERIN), Luxembourg Institute of Science and Technology (LIST), Belvaux, Luxembourg.

24/04/2023, The European Geosciences Union (EGU) General Assembly 2023
Session HS3.1
Hydroinformatics: data analytics, machine learning, hybrid modelling, optimisation

Abstract

Hydrologists often need to choose between competing hypotheses or weight the predictions of different models when averaging models. Several criteria for choosing and weighting models have been developed, which balance model complexity and goodness of fit by penalising the number of model parameters. The penalty is explicit for information theory approaches or implicit for Bayesian model selection based on marginal likelihood and, by extension, the Bayes factor. The Bayes factor is the ratio of the marginal likelihoods of two competing models. In addition, the Bayes factor is consistent in contrast to information-theoretic-based approaches. However, marginal likelihood estimation is computationally intensive and slow for dynamic models with multiple modes. This study uses Replica Exchange Hamiltonian Monte Carlo and thermodynamic integration for fast, simultaneous calculation of marginal likelihood and identification parameters of dynamic rainfall-runoff models. Using synthetic data, the method selected the true model in our numerical experiments. The technique was also applied to real data from Magela Creek in Australia. The selected model was not the model with the highest or lowest number of parameters for real data. The method is implemented using the differentiable programming software "TensorFlow Probability". This implementation can be applied to other types of models for fast simultaneous parameter estimation and model comparison.

keywords: marginal likelihood, Bayesian, thermodynamic integration, rainfall-runoff.