SPMC: Scalable Python Markov Chain Monte Carlo with application to Bayesian parameter inference in stochastic ecological models

Jonas Šukys[1] and Mira Kattwinkel[2]

[1] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Switzerland,
`jonas.sukys@eawag.ch`,
[2] University of Koblenz-Landau, Germany,
`kattwinkel-mira@uni-landau.de`

Keywords: Bayesian inference, Particle Markov Chain Monte Carlo, Individual Based Model, dynamic load balancing

1 Introduction

Individual or agent based models (IBMs/ABMs) have been successfully applied for modeling complex ecological dynamical systems [1]. However, model parameter estimation is often performed only ad-hoc by manual tuning or optimization, lacking a thorough assessment of the underlying uncertainties. Bayesian inference [2] can be used for both parameter estimation and uncertainty quantification, but its successful application to realistic scenarios has been hindered by the complex stochastic nature of IBMs leading to technical difficulties in likelihood estimation and entailing high computational costs. Likelihood estimation issues were recently tackled by applying Particle Markov Chain Monte Carlo (PMCMC) technique [3], where the hidden Markov structure of IMBs is used for efficient marginal likelihood approximations using time-series observations [4]. We present `SPMC`, a Python implementation of parallel PMCMC algorithm, which mitigates high computational costs by dynamically distributing particles (model evaluations) over multiple computational units. Due to minimal required changes in the original model code, `SPMC` is easily applicable to other IBMs.

2 Algorithms and implementation

PMCMC algorithm consists of recursive model parameter estimation steps using Markov Chain Monte Carlo (MCMC) posterior sampler, as described in [3, 4]. Each recursion step evaluates an unbiased statistical approximation of the proposal parameter likelihood using the actual observations provided by a time series. Likelihood approximation procedure involves multiple evaluations of the underlying model, so called particles. At each measurement time in the observations time series, model simulations are paused and all particles are re-sampled