Bayesian Learning of Stochastic Dynamical Models

Peter Lu and Pierre F. J. Lermusiaux

Massachusetts Institute of Technology, Department of Mechanical Engineering, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Abstract

A new methodology for rigorous Bayesian learning of high-dimensional stochastic dynamical models is developed. The methodology performs parallelized computation of marginal likelihoods for multiple candidate models, integrating over all state variable and parameter values, and enabling a principled Bayesian update of model distributions. This is accomplished by leveraging the dynamically orthogonal (DO) evolution equations for uncertainty prediction in a dynamic stochastic subspace and the Gaussian Mixture Model-DO filter for inference of nonlinear state variables and parameters, using reduced-dimension state augmentation to accommodate models featuring uncertain parameters. Overall, the joint Bayesian inference of the state, model equations, geometry, boundary conditions, and initial conditions is performed. Results are exemplified using two high-dimensional, nonlinear simulated fluid and ocean systems. For the first, limited measurements of fluid flow downstream of an obstacle are used to perform joint inference of the obstacle's shape, the Reynolds number, and the $O(10^5)$ fluid velocity state variables. For the second, limited measurements of the concentration of a microorganism advected by an uncertain flow are used to perform joint inference of the microorganism's reaction equation and the $O(10^5)$ microorganism concentration and ocean velocity state variables. When the observations are sufficiently informative about the learning objectives, we find that our posterior model probabilities correctly identify either the true model or the most plausible models, even in cases where a human would be challenged to do the same.

Keywords: Bayesian data assimilation, learning, GMM-DO, dynamical system, stochastic PDEs, ocean and weather prediction

1. Introduction

Stochastic dynamical systems [1] are everywhere, from oceanic and ecological systems, power grids and communications networks, to financial markets and social networks. The mathematical tools that have been developed for investigating stochastic dynamical systems are thus highly versatile and have been applied in a wide range of fields [2 3 4 5 6 7 8 9 10 11 12 13].

Quantitative investigations of a stochastic dynamical system typically assume that the mathematical model formulated for the system is an accurate description of its governing processes. Uncertainty in the system's state variables are often assumed to originate solely from uncertainty in the system's initial and boundary conditions, and stochastic forcings with known statistical properties. This assumption of absolute model formulation validity however is not always defensible. For example, when dealing with complex systems for which governing equations have not yet been derived from known first principles, the assumption is surely inappropriate. In general, uncertainty in model formulation can originate from the choice of state variables themselves, from the functional forms of the model equations, boundary conditions or initial conditions, and from the definition of the (spatial) domain of integration. Both the deterministic and stochastic components of the model formulation can be uncertain. In what follows, when possible, we will refer to model formulation uncertainty simply as model uncertainty.

Model uncertainty can be difficult to quantify and is thus often ignored. This is not damaging when model uncertainty is insignificant. For example, when we model ballistic dynamics on Earth, one can have confidence in Newton's laws of motion. In other cases however, it can lead to significant underestimation of uncertainty. [13], [14], and [15] review poignant examples from statistics in which ignorance of model uncertainty resulted in overconfidence in state estimates, which subsequently led to tragically flawed conclusions.