Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

SEQUENTIAL IMPORTANCE SAMPLING FOR HYBRID MODEL BAYESIAN INFERENCE TO SUPPORT BIOPROCESS MECHANISM LEARNING AND ROBUST CONTROL

Wei Xie
Keqi Wang
Hua Zheng

Mechanical and Industrial Engineering
Northeastern University
360 Huntington Ave
Boston, MA 02115, USA

Ben Feng

Statistics and Actuarial Science
University of Waterloo
200 University Ave W
Waterloo, ON N2L 3G1, CANADA

ABSTRACT

Driven by the critical needs of biomanufacturing 4.0, we introduce a probabilistic knowledge graph hybrid model characterizing the risk- and science-based understanding of bioprocess mechanisms. It can faithfully capture the important properties, including nonlinear reactions, partially observed state, and nonstationary dynamics. Given very limited real process observations, we derive a posterior distribution quantifying model estimation uncertainty. To avoid the evaluation of intractable likelihoods, Approximate Bayesian Computation sampling with Sequential Monte Carlo (ABC-SMC) is utilized to approximate the posterior distribution. Under high stochastic and model uncertainties, it is computationally expensive to match output trajectories. Therefore, we create a linear Gaussian dynamic Bayesian network (LG-DBN) auxiliary likelihood-based ABC-SMC approach. Through matching the summary statistics driven through LG-DBN likelihood that can capture critical interactions and variations, the proposed algorithm can accelerate hybrid model inference, support latent state monitoring, and facilitate mechanism learning and robust control.

1 INTRODUCTION

The biopharmaceutical manufacturing industry is growing rapidly and it plays a critical role to ensure public health and support economy. *However, biomanufacturing often faces critical challenges, including high complexity, high variability, and very limited process observations.* As new biotherapeutics (e.g., cell and gene therapies) become more and more personalized, it requires more advanced manufacturing protocols. For example, the seed cells, extracted from individual patients or donors, can have different optimal culture policies. Therefore, the production process involves a complex stochastic decision process (SDP) with output trajectory dynamics and variations influenced by biological/physical/chemical (a.k.a. *biophysicochemical*) reactions occurring at molecular, cellular, and system levels.

In general, there are two main categories of biomanufacturing process modeling methodologies in the existing literature: mechanistic and data-driven approaches. The ordinary/partial differential equations (ODE/PDE) mechanistic models are developed based on biophysicochemical mechanisms. They have good interpretability and show generally higher extrapolation power than data-driven models. However, existing mechanistic models often fail to rigorously account for *uncertainties*, i.e., inherent stochasticity and model estimation uncertainty. For example, batch-to-batch variation, known as a major source of bioprocess uncertainty (Mockus et al. 2015), is ignored in deterministic mechanistic models. Therefore, mechanistic models may not fit well to the observations collected from real systems in many situations, which also limits their power in terms of mechanism learning, process monitoring, and robust control to support flexible on-demand manufacturing. On the other hand, data-driven approaches often use general

978-1-6654-7661-4/22/$31.00 ©2022 IEEE