Bayesian Calibration of Simulators with Structured Discretization Uncertainty

Oksana A. Chkrebtii

Department of Statistics, The Ohio State University

Joint work with Matthew T. Pratola (Statistics, The Ohio State University)

Probabilistic Scientific Computing: Statistical inference approaches to numerical analysis and algorithm design, ICERM
June 7, 2017

THE OHIO STATE UNIVERSITY

Probabilistic Numerics

This is an active new field that challenges historical perspectives on numerical analysis.

It is important for this community to develop new methods with an eye to overcoming challenges that lay ahead.
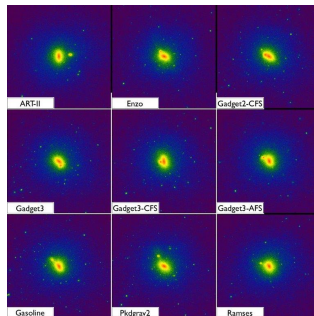
This talk focuses on calibration for forward problems defined by the solution of ordinary and partial differential equations. If you're not already convinced that probabilistic numerics is useful in this setting ...

THE OHIO STATE UNIVERSITY

Example - galaxy simulation (Kim *et al.*, 2016)
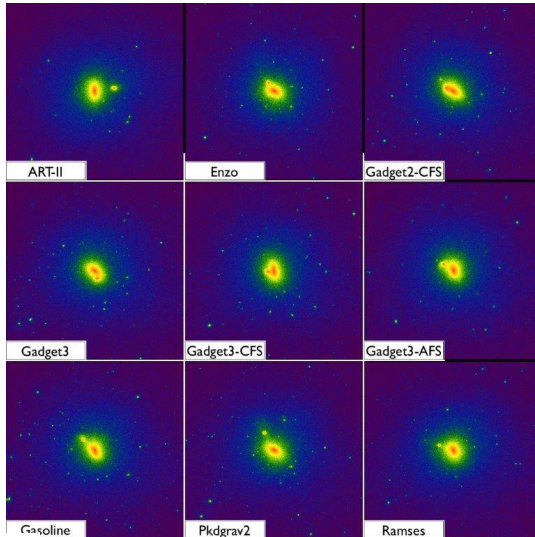
Example - galaxy simulation (Kim *et al.*, 2016)

- These are *not* realizations of a field (the model is deterministic)
- The initial conditions and inputs are held fixed
- How do we evaluate these numerical solver outputs?

Perspectives on Probabilistic Numerics

Probability measures on numerical solutions via randomization

- Conrad *et al* (2015/16), Lie *et al* (2017)
- defined outside of the Bayesian framework, but resulting algorithms overlap

Bayesian uncertainty quantification for differential equations

- Skilling (1991), Chkrebtii *et al* (2013/16), Arnold *et al* (2013)
- defined outside of the numerical analysis framework, but resulting methods can be analogous in some sense

Bayesian numerical methods

- Hennig & Hauberg (2013/14), Schober *et al* (2014).
- computationally efficient probabilistic GP based methods; can recover numerical solvers in the mean

THE OHIO STATE UNIVERSITY

Calibrating stochastic computer models

Regardless of the perspective, the deterministic but unknown forward model is replaced by a stochastic process: for fixed inputs, the output is a random variable with (often) unknown distribution.
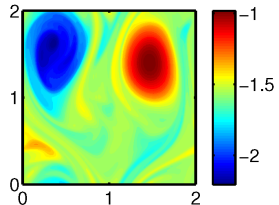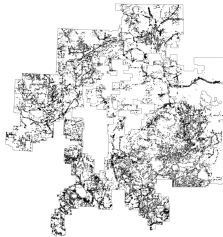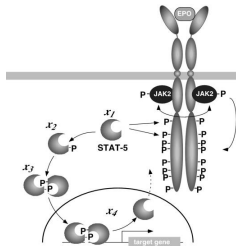
Pratola & Chkrebtii (2017+) describe a hierarchical framework to calibrate stochastic simulators with highly structured output uncertainty/variability.

THE OHIO STATE UNIVERSITY

Calibration problem

We wish to estimate the unknowns, $\theta \in \Theta$, given observations,

$$y(x_t) = A\{u(x_t, \theta)\} + \varepsilon(x_t), \quad x_t \in \mathcal{X}, \ \ t = 1, \ldots, T,$$

of the deterministic state $u_t = u(x_t, \theta)$ transformed via an observation process A, and contaminated with stochastic noise ε.

The likelihood defines a discrepancy between the model and the data:

$$f(y_{1:T} \mid \theta) \propto \rho\{y_{1:T} - A(u_{1:T})\}$$

THE OHIO STATE UNIVERSITY

The Bayesian paradigm

Bayesian inference is concerned with modeling degree of belief about an unknown quantity via probability models.

For example, we may not know $\theta \in \Theta$ but we may have some prior belief about, e.g., its range, most probable values,

$$\theta \sim \pi(\theta).$$

We seek to update our prior belief by conditioning on new information, $y_{1:T} \in \mathcal{Y}$, e.g., data, model evaluations, via Bayes' Rule:

$$p(\theta \mid y_{1:T}) = \frac{p(y_{1:T} \mid \theta)\,\pi(\theta)}{\int p(y_{1:T} \mid \theta)\,\pi(\theta)\,d\theta} \propto p(y_{1:T} \mid \theta)\,\pi(\theta).$$

THE OHIO STATE UNIVERSITY

A Hierarchical model representation

Hierarchical modelling enables inference over the parameters of a stochastic state,

$$
\begin{aligned}
[y_{1:T} \mid u_{1:T}, \theta] &\propto \rho \{y_{1:T} - A(u_{1:T})\} \\
[u_{1:T} \mid \theta] &\sim p(u_{1:T} \mid \theta) \\
[\theta] &\sim \pi(\theta).
\end{aligned}
$$

When $p(u_{1:T} \mid \theta)$ is not known in closed form, exact inference may still be possible via Monte Carlo, using forward-simulation from the model. However, this is often computationally prohibitive.

THE OHIO STATE UNIVERSITY

For probabilistic numerics

If the state is deterministic but defined implicitly by a system of differential equations, our uncertainty about the solution can be modelled probabilistically,

$$
\begin{aligned}
[y_{1:T} \mid u_{1:T}, \theta] &\propto \rho\left[y_{1:T} - A\left(u_{1:T}\right)\right] \\
[u_{1:T} \mid \theta] &\sim \text{a probability measure representing uncertainty} \\
&\quad\ \ \text{in the solution given discretization of size } N \\
[\theta] &\sim \pi(\theta).
\end{aligned}
$$

We use the Bayesian uncertainty quantification approach to model this middle layer.

Bayesian UQ for differential equations

Given θ and for linear operator D consider the initial value problem,

$$\begin{cases} Du & = f(x, u), & x \in \mathcal{X}, \\ u & = u_0 & x \in \partial\mathcal{X}. \end{cases}$$

We may have some prior knowledge about smoothness, boundary conditions, etc., described by a prior measure,

$$u \sim \pi, \quad x \in \mathcal{X}$$

We seek to update our prior knowledge by conditioning on model interrogations, $f_{1:N}$ via Bayes' Rule,

$$p(u(x) \mid f_{1:N}) = \frac{p(f_{1:N} \mid u(x))\, \pi(u(x))}{\int p(f_{1:N} \mid u(x))\, \pi(u(x))\, du(x)} \propto p(f_{1:N} \mid u(x))\, \pi(u(x))$$

THE OHIO STATE UNIVERSITY

Prior uncertainty in the unknown solution

The exact solution function u is deterministic, but unknown. We may describe our prior uncertainty via a probability model defined on the space of suitably smooth derivatives, e.g.,

$$u \sim \mathcal{GP}(m^0, C^0), \qquad m^0 : \mathcal{X} \to \mathbb{R}, \ C^0 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$$

with the constraint $m^0 = u^0, x \in \partial X$.

This yields a joint prior on the fixed but unknown state and its derivative(s)

$$\begin{bmatrix} u \\ Du \end{bmatrix} \sim \mathcal{GP} \left(\begin{bmatrix} m^0 \\ Dm^0 \end{bmatrix}, \begin{bmatrix} C^0 & C^0 D^* \\ DC^0 & DC^0 D^* \end{bmatrix} \right)$$

THE OHIO STATE UNIVERSITY

Interrogating the model recursively

❶ Draw a sample from the marginal predictive distribution on the state at the next discretization grid point $s_{n+1} \in \mathcal{X}$, $1 \leq n < N$

$$u(s_{n+1}) \sim p\left(u(s_{n+1}) \mid \mathsf{f}_{1:n}\right)$$

❷ Evaluate the RHS at $u(s_{n+1})$ to obtain a model interrogation,

$$\mathsf{f}_{n+1} = f\left(s_{n+1}, u(s_{n+1})\right)$$

❸ Model interrogations as "noisy" measurements of Du:

$$\mathsf{f}_{n+1} \mid Du, f_{1:n} \sim \mathcal{N}\left(Du(s_{n+1}), \Lambda(s_n)\right)$$

Sequential Bayesian updating

Updating our knowledge about the true but unknown solution given the new interrogation trajectory f_{n+1}

$$\begin{bmatrix} u \\ Du \end{bmatrix} \mid f_{n+1} \sim \mathcal{GP} \left(\begin{bmatrix} m^{n+1} \\ Dm^{n+1} \end{bmatrix}, \begin{bmatrix} C^{n+1} & C^{n+1}D^* \\ DC^{n+1} & DC^{n+1}D^* \end{bmatrix} \right)$$

where,

$$m^{n+1} = m^n + K^n \left(f_{n+1} - m^n(s_{n+1}) \right)$$
$$C^{n+1} = C^n - K^n DC^{n*}$$
$$K^n = C^n D^* \left(DC^n + \Lambda(s_n) \right)^{-1}$$

This becomes the prior for the next update.

THE OHIO STATE UNIVERSITY

Bayesian UQ for differential equations

Due to the Markov property, we cannot condition the solution on multiple trajectories $f_{1:N}^j, j = 1, \dots, J$ simultaneously. In fact, the posterior over the unknown solution turns out to be a continuous mixture of Gaussian processes,
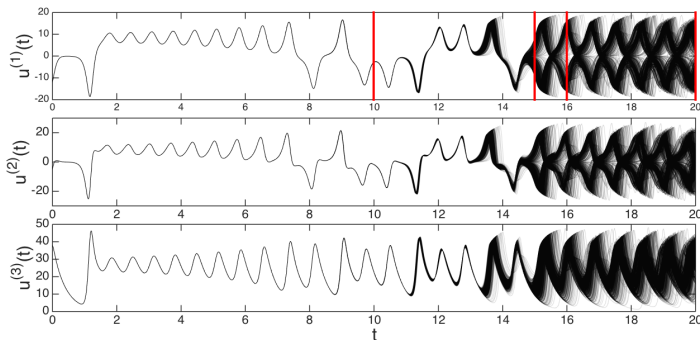
$$[u \mid \theta, N] = \int \int [u, Du \mid f_{1:N}, \theta, N] \, d(Du) \, df_{1:N}.$$

Samples from this posterior can be obtained via Monte Carlo.

Example - Lorenz63 forward model

A probability statement over probable trajectories given fixed model parameters and initial conditions for the Lorenz63 model:

1000 draws for the probabilistic forward model for the Lorenz63 system given fixed initial states and model parameters in the chaotic regime.

THE OHIO STATE UNIVERSITY

Example - Lorenz63 forward model
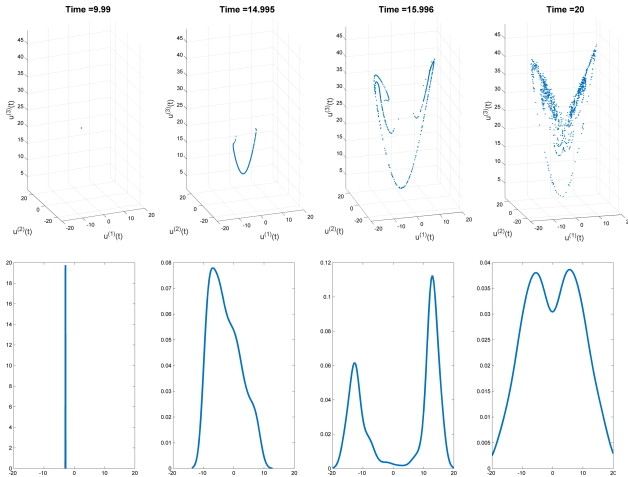
1000 draws from forward model for Lorenz63 system at four fixed time points.

For probabilistic numerics

If the state is deterministic but defined implicitly by a system of differential equations, our uncertainty about the solution can be modelled probabilistically,

$$
\begin{aligned}
[y_{1:T} \mid u_{1:T}, \theta] &\propto \rho\left[y_{1:T} - A\left(u_{1:T}\right)\right] \\
[u_{1:T} \mid \theta] &\sim \text{a probability measure representing uncertainty} \\
&\quad\text{in the solution given discretization of size } N \\
[\theta] &\sim \pi(\theta).
\end{aligned}
$$

We use the Bayesian uncertainty quantification approach to model this middle layer.

Computer model emulation

We interrogate the model at M regimes, $\theta_{1:M} = (\theta_1, \ldots, \theta_M)^\top$. Each interrogation is comprised of an ensemble of K output samples.

Let $\tilde{u}_{1:T}^{1:K}(\theta_{1:M})$ denote the ensemble of K output simulations for each of the regimes $\theta_{1:M}$.

$$
\begin{aligned}
\left[y_{1:T} \mid \tilde{u}_{1:T}^{1:K}(\theta_{1:M}), u_{1:T}, \delta_{1:T}, \theta \right] &\propto \rho \left[y_{1:T} - A\left(u_{1:T}(\theta) \right) - \delta_{1:T} \right] \\
\left[\tilde{u}_{1:T}^{k}(\theta_m) \mid u_{1:T}^{k}, \theta \right] &\sim \mathcal{N}\left(u_{1:T}^{k}(\theta_m), \Lambda \right) \quad k = 1, \ldots, K \\
\left[u_{1:T}^{k} \mid \theta \right] &\sim \text{generative stochastic model} \\
\left[\theta, \delta \right] &\sim \pi(\theta, \delta).
\end{aligned}
$$

THE OHIO STATE UNIVERSITY

A Hierarchical model representation

Challenges of MCMC sampling from the posterior include:

- Emulation based on *MK* model evaluations is computationally expensive
- Models are evaluated at multiple spatio-temporal locations and over multiple states

Our approach:

- Dimension reduction over the second (output) layer of the hierarchical model
- We include the dimension reduction specifications within the hierarchical model, resulting in a fully probabilistic approach

THE OHIO STATE UNIVERSITY

Example: Exact vs Emulation Based Inference
in a Model of Protein Dynamics

Inference for a model of protein dynamics

JAK-STAT chemical signaling pathway model describes concentration of 4 STAT factors by a delay differential equation system on $t \in [0, 60]$,

Illustration of the JAK-STAT mechanism
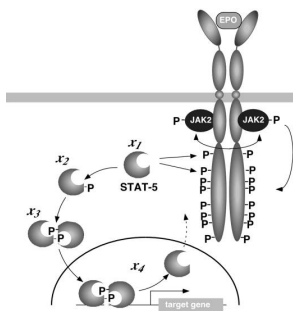
$$\frac{d}{dt} u^{(1)}(t, \theta) = -k_1 \, u^{(1)}(t, \theta) \, EpoR_A(t) + 2 \, k_4 \, u^{(4)}(t - \tau, \theta)$$

$$\frac{d}{dt} u^{(2)}(t, \theta) = k_1 \, u^{(1)}(t, \theta) \, EpoR_A(t) - k_2 \left(u^{(2)}(t, \theta) \right)^2$$

$$\frac{d}{dt} u^{(3)}(t, \theta) = -k_3 \, u^{(3)}(t, \theta) + 0.5 \, k_2 \left(u^{(2)}(t, \theta) \right)^2$$

$$\frac{d}{dt} u^{(4)}(t, \theta) = k_3 \, u^{(3)}(t, \theta) - k_4 \, u^{(4)}(t - \tau, \theta)$$

$$u^{(i)}(t, \theta) = \phi^{(i)}(t), \qquad t \in [-\tau, 0], \; i = 1, \ldots, 4$$

Inference for a model of protein dynamics

States are observed indirectly through a nonlinear transformation:
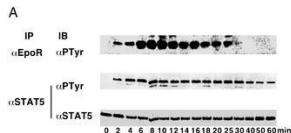
Experimental measurements

$$A^{(1)} = k_5\big(u^{(1)}(t;\theta) + 2u^{(3)}(t;\theta)\big)$$

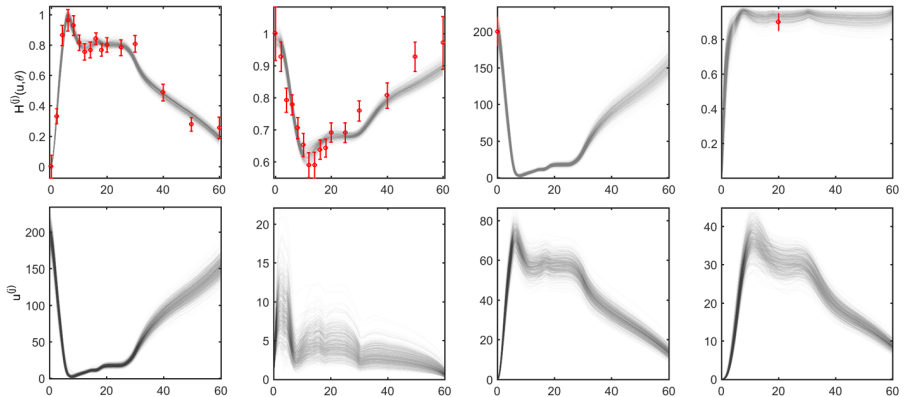$$A^{(2)} = k_6\big(u^{(1)}(t;\theta) + u^{(2)}(t;\theta) + 2u^{(3)}(t;\theta)\big)$$

$$A^{(3)} = u^{(1)}(t;\theta)$$

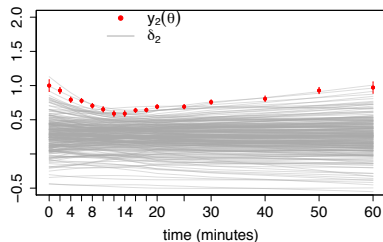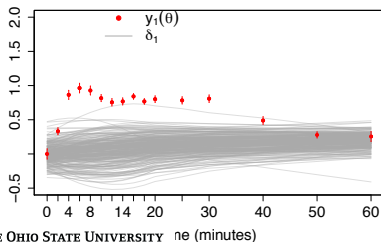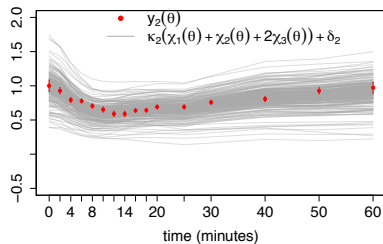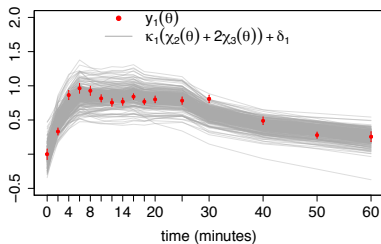$$A^{(4)} = \frac{u^{(3)}(t;\theta)}{u^{(2)}(t;\theta) + u^{(3)}(t;\theta)}$$

Observations are noisy measurements on the transformed states and forcing function at points $\mathbf{t} = \{t_{ij}\}_{i=1,\ldots,4;j=1,\ldots,n_i}$

$$y(\mathbf{t}) = A_{k_4,k_5}\,u(\mathbf{t}; k_1,\ldots,k_6,\tau,\phi,EpoR_A) \;+\; \varepsilon(\mathbf{t})$$
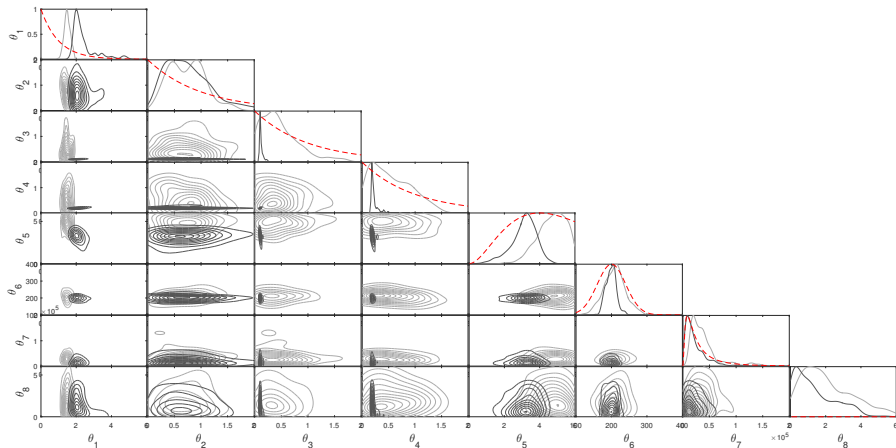
Results: exact inference

Results: emulation based inference

Results: emulation based inference

Kernel density estimates of the marginal stochastically calibrated posterior (gray) with
$M = 100$ model runs, and exact posterior (black) for the JAK-STAT system. Marginal prior
densities are shown as dotted lines.

THE OHIO STATE UNIVERSITY

Thank you!

References

1. Pratola, M.T., Chkrebtii, O.A. *Bayesian Calibration of Multistate Stochastic Simulators*. To appear in Statistica Sinica.

2. Chkrebtii, O.A., Campbell, D.A., Calderhead, B., Girolami, M. , *Bayesian Solution Uncertainty Quantification for Differential Equations* , Bayesian Analysis with discussion, 11(4), 1239-1267, 2016.

3. Arnold, A., Calvetti, D, Somersalo, E. *Linear multistep methods, particle filtering and sequential Monte Carlo*, Inverse Problems, 29, 2013.

4. Conrad, R., Girolami, M., Sarkka, S., Stuart, A., Zygalakis, K. *Statistical analysis of differential equations: introducing probability measures on numerical solutions*, Statistics and Computing, 2016.

5. Lie, H.C., Stuart, A. M., Sullivan, T. J. *Strong convergence rates of probabilisti integrators for ordinary differential equations*, arXiv preprint, 2017.

THE OHIO STATE UNIVERSITY

References

1. Hennig, P., Hauberg, S. *Probabilistic Solutions to Differential Equations and their Application to Riemannian Statistics*, In Proc. of the 17th int. Conf. on Artificial Intelligence and Statistics (AISTATS) (Vol. 33). JMLR, W&CP, 2014.

2. Schober, M., Duvenaud, D. K., Hennig, P. *Probabilistic ODE Solvers with Runge-Kutta Means*. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 27 (pp. 739747). Curran Associates, Inc, 2014.

3. Skilling, J. *Bayesian solution of ordinary differential equations* Maximum Entropy and Bayesian Methods, Seattle, 1991.

4. Kim *et al.*, *The AGORA High-resolution Galaxy Simulations Comparison Project*, The Astrophysical Journal Supplement Series, 2014.

5. Swameye *et al.*, *Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling*. PNAS, 2003.

THE OHIO STATE UNIVERSITY