Ultra-Short Time Batching and Unscented Kalman Inversion for Calibration of Expensive Chaotic Models

Teo S. Price-Broncucia[a,*], Rebecca E. Morrison[a]

[a]*Computer Science, University of Colorado Boulder, 1111 Engineering Drive ECOT 717, 430 UCB Boulder, CO 80309, USA*

Abstract

Computer models traditionally used for weather and climate prediction have extremely high computational costs. While reduced models exist, their utility is limited in part because their calibration poses a host of difficulties, including chaotic dynamics that prevent the use of adjoint methods, computation costs that become unreasonable when sampling approaches require many forward runs with long ergodic trajectories, and large existing code bases that necessitate black box approaches. Recently, Unscented Kalman Inversion (UKI) methods have shown promise for such models by providing approximate derivatives of parameters in order to reach convergence using relatively few forward model runs. Previous UKI applications have required expensive ergodic trajectories. Inspired by recent work in consistency testing for climate models, we instead sample many ultra-short model trajectories to greatly reduce the calibration cost of a canonical chaotic test case.

Keywords: Climate Models, Parameter Estimation, Black-Box Optimization

1. Introduction

Modern Global Climate Models (GCMs) can require hundreds of thousands of core hours for a single simulation. The high costs prevent the use of such models in most educational and industrial applications. In the scientific sphere, these costs preclude activities requiring many model runs, such as uncertainty quantification and the exploration of rare events. Such tasks require computationally cheaper alternative models but development and adoption of less expensive, or reduced, models is contingent on effective and efficient parameter calibration.

The calibration of GCMs poses multiple problems. First, they often involve large existing code bases, which are developed by domain experts and may not support the machinery required for automatic calculation of gradients. The Community Earth System Model, a widely used GCM, contains over one million lines of Fortran code. Second, they exhibit strong chaotic dynamics. Combined with the size of the code-base, this precludes the use of adjoint methods for gradient calculation [1] and requires the use of long trajectories to ensure stable ergodic averaging [2]. Third, even reduced weather and climate models can still be computationally expensive, requiring on the order of an hour of wall-clock to compute a typical trajectory length, preventing the use of sampling-based calibration methods like Markov Chain Monte Carlo.

Recent work in the field of climate model consistency testing has shown that comparing ensembles of ultra-short simulations can achieve similar performance to the traditional use of multi-

*Student Contributor

Email addresses: teo.pricebroncucia@colorado.edu (Teo S. Price-Broncucia), rebeccam@colorado.edu (Rebecca E. Morrison)

hundred year simulations [3]. These short simulations each simulate just a few hours, resulting in a vast speedup from traditional approaches and enabling comparison before chaotic effects take over. While the goals of consistency testing—i.e., to ensure that changes to compilers, hardware, etc. do not not impact the scientific conclusions of the model—are narrower than those of calibration, these results give some confidence that information about large climate models can be obtained from ultra-short model runs.

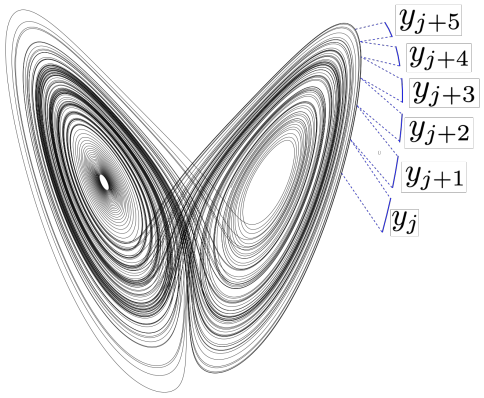

Figure 1: Demonstration of splitting a single long chaotic trajectory into many ultra-short samples for use in calibration.

Recent deep-learning models for weather prediction [4] have also used short time intervals of observation data. However, these approaches do not apply to black box models, and the goals are generally different from our work, with a focus on building deep-learning models directly from observations instead of the physics-based reduced models we consider here.

In this work, we aim to reduce the computational cost of parameter calibration when ground-truth data is generated by a high-fidelity computer model, and the model to be calibrated is too expensive for sampling approaches, exhibits chaotic dynamics, and precludes the use of adjoint methods. By considering a single long observation as a collection of many ultra-short simulations (fig. 1), our approach is able to reduce the computational costs of calibrating parameters of chaotic models by more than 90% when compared to the use of full ergodic trajectories.

2. Methodology

In this work we extend Unscented Kalman Inversion (UKI) to perform calibration, that is, to determine an unknown set of parameters θ of a known model $\mathcal{G}(\theta, \mathbf{u_0}, \Delta t)$ by comparing outputs (or statistics of those outputs like higher moments) of said model to known quantities $\mathbf{y}_{\text{true}}$ (e.g., high-fidelity model outputs). Beginning from initial state $\mathbf{u_0}$ and running the model forward for Δt time, we denote the D-dimensional full output of the model as $\mathbf{z} \in \mathbb{R}^D$. The transformation from full model output $\mathbf{z}$ to the C-dimensional space of known outputs $\mathbf{y} \in \mathbb{R}^C$ is described by an observation operator as $\mathbf{y} = \varphi(\mathbf{z})$.

UKI computes a deterministic ensemble of model runs at each calibration iteration to generate an approximate gradient of parameters to outputs. Whereas previous works relied on long ergodic trajectories and their associated time-averaged outputs, we split $\mathbf{y}_{\text{true}}$ and $\mathbf{z}$ into "ultra-short" sub-trajectories of length δt, yielding $\frac{\Delta t}{\delta t} = N$ samples. We select a value of δt that is long enough to enable measurable spread in ensemble members, but too short for the exponential effects of chaotic dynamics to dominate. At each UKI iteration, parameter updates are averaged for B random samples from the N total samples. (For a full description of UKI, see [5]). We term this approach *ultra-short time batching*, outlined as:

1. Split true full output $\mathbf{z}_{\text{true}}$ and transformed output $\mathbf{y}_{\text{true}}$ into N samples of length δt.
2. Estimate noise in observations to set UKI parameter Σ_ν, analogous to setting step size.
3. Repeat until convergence:
   (a) Run UKI iteration to compute B separate parameter updates using B random samples.
   (b) Average the parameter updates to get new estimate of parameters θ.

We find batch sampling across the entire trajectory advantageous for two reasons. First, in order to simulate structural differences between a high-fidelity model and a reduced one, Gaussian noise is added to $\mathbf{y}_{\text{true}}$. Batching provides a better estimate of the true gradient under noise, enabling faster convergence to good parameters. Second, depending on the point in the trajectory, a model may be more or less sensitive to certain parameters. For instance, a climate model parameter that controls land-ice reflectivity may have a bigger impact during the southern hemisphere summer due to the asymmetrical distribution of land-ice on earth. Batching averages these impacts.

3. Test Problem: Lorenz '96

As a test problem, we consider the Lorenz '96 model [6]. The model emulates the fast and slow evolving variables present in GCMs and, at the parameter values used here, exhibits strong chaotic behavior. The impact of chaotic dynamics on the loss landscape is seen in fig. 2; this chaotic dependence on parameters can make calibration difficult [1]. However, the loss landscape for a single ultra-short trajectory appears relatively smooth.

The model consists of K slow variables X_k, each coupled to J fast variables $Y_{j,k}$. Both types are periodic, such that $X_{k+K} = X_k$ and $Y_{j+J,k} = Y_{j,k+1}$. The differential equations, defined with notation from [7], are:

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - Xk + 1) - X_k + F - hc\bar{Y}_k \quad (1)$$

$$\frac{dY_{j,k}}{dt} = c\left(-bY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - Y_{j,k} + \frac{h}{J}X_k\right) \quad (2)$$

$$\bar{Y}_k = \frac{1}{J}\sum_{j=1}^{J} Y_{j,k}. \quad (3)$$

In this work we set $K = 36$, $J = 10$ for a total of 396 variables. The ground truth is generated with $[h, F, c, b] = [1, 10, \log(10), 10]$. We begin from a prior estimate with mean $[0.1, 5, 2, 7]$ and independent variances $[1, 10, 0.1, 1]$. The model is run until $t = 1\text{e}4$.

For the ultra-short batching we set $\delta t = 0.01$, resulting in $N = 1\text{e}6$ total samples. Ground-truth samples are perturbed using additive, i.i.d. Gaussian noise with diagonal variance $\sigma^2 = 0.0001 \cdot \overline{\mathbf{z}_{\text{true}}}$, i.e., noise is applied to both the initial state $\mathbf{u_0}$ of each sample, and the transformed output $\mathbf{y} = \varphi(\mathbf{z})$, where

$$\varphi(\mathbf{z}) = \frac{1}{K}\sum_k \left(X_k, \bar{Y}_k, X_k^2, X_k\bar{Y}_k, \overline{Y^2}_k\right). \quad (4)$$

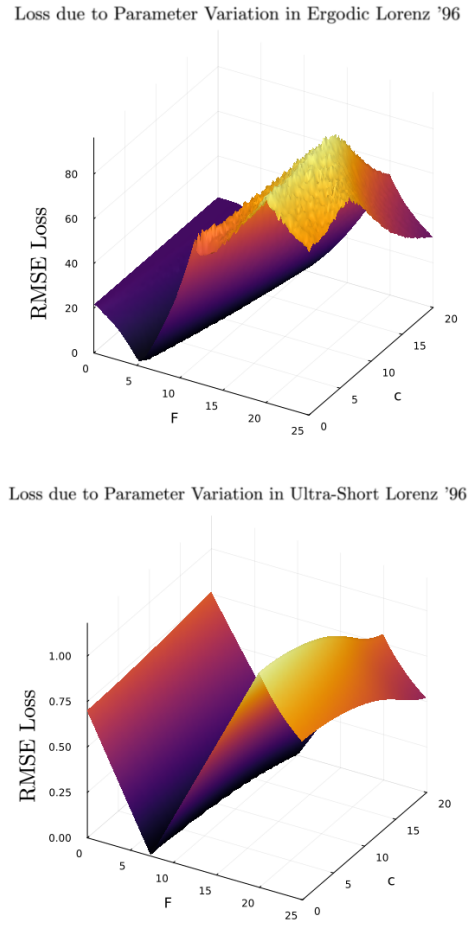

Figure 2: The RMSE loss due to perturbed parameters (F and c) for ergodic and ultra-short Lorenz '96 trajectories. Chaotic dynamics result in a very rough loss-landscape, inhibiting calibration.

4. Results and Discussion

We find that calibration of the Lorenz '96 model requires substantially less computation with ultra-short time batching than full ergodic runs. As seen in fig. 3, the ultra-short time batching

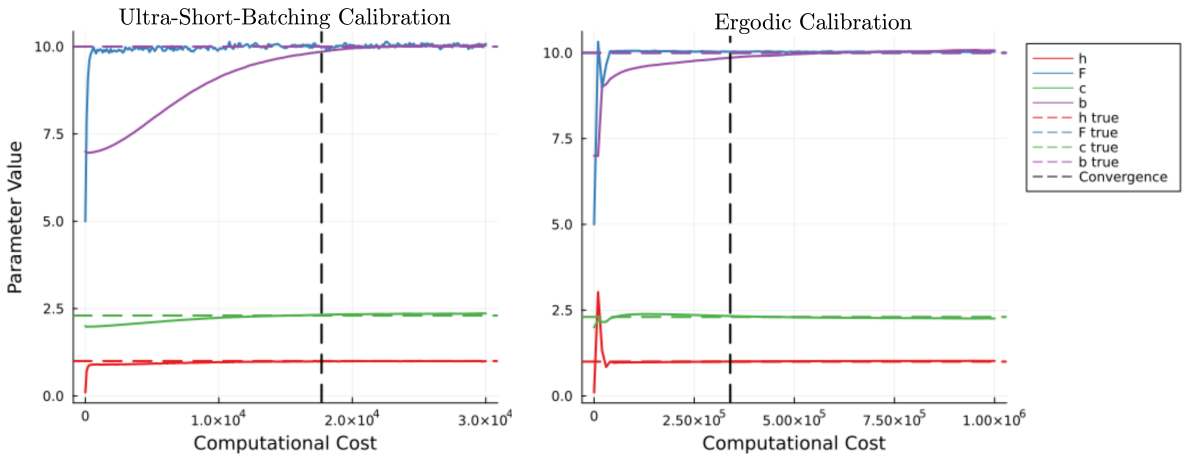

Figure 3: Lorenz '96 calibration results and relative computational costs. Using UKI, the ultra-short approach converges (within 1.5% of θ_{true}) using just 5.2% of the computational cost of the ergodic approach. One unit of computational cost is equivalent to an ensemble of forward model runs with trajectory lengths of $t = 1$.

method achieves parameter convergence (all parameters within 1.5% of true values) using just 5.2% of the computational cost required when using ergodic trajectories. Alternative hyper-parameters of UKI may result in more efficient ergodic performance, but are unlikely to erase the demonstrated improvement entirely.

This substantial cost improvement makes ultra-short time batching a compelling addition to a calibration toolkit for the expensive chaotic models found in climate and weather prediction. If the values of interest for calibrated models are time-averaged quantities, e.g., average yearly temperature, it is simple to test the parameters found using ultra-short time batching with additional ergodic model runs. If needed, further ergodic calibration can be done with less wasted computation. Future work includes the development of adaptive methods for noise estimation (which in turn controls the UKI step size), analogous to deep-learning methods that adjust the learning rate, and tests of more realistic climate models with substantial model inadequacy, a feature of nearly all reduced models employed in practice.

References

[1] D. J. Lea, M. R. Allen, T. W. N. Haine, Sensitivity analysis of the climate of a chaotic system, Tellus A: Dynamic Meteorology and Oceanography 52 (5) (Jan. 2000).

[2] S. H. Strogatz, Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering, CRC Press, 2018.

[3] D. J. Milroy, A. H. Baker, D. M. Hammerling, E. R. Jessup, Nine time steps: ultra-fast statistical consistency testing of the Community Earth System Model (pyCECT v3.0), Geoscientific Model Development 11 (2) (2018) 697–711.

[4] FourCastNet: A Global Data-driven High-resolution Weather Model using Adaptive Fourier Neural Operator-sArXiv:2202.11214 [physics] (Feb. 2022).

[5] D. Z. Huang, T. Schneider, A. M. Stuart, Unscented Kalman Inversion, arXiv:2102.01580 [cs, math] (Apr. 2021).

[6] E. N. Lorenz, Predictability: A problem partly solved, Vol. 1, Reading, 1996, issue: 1.

[7] T. Schneider, S. Lan, A. Stuart, J. Teixeira, Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations, Geophysical Research Letters 44 (24) (2017) 12,396–12,417.