Multi-Time Unscented Kalman Inversion for Calibration of Expensive Chaotic Models

Teo Price-Broncucia
Graduate Student, Dept. of Computer Science, Univ. of Colorado, Boulder, USA

Rebecca Morrison
Asst. Professor, Dept. of Computer Science, Univ. of Colorado, Boulder, USA

ABSTRACT: The computer models traditionally used for weather and climate prediction have extremely high computational costs, which impede their use in infrastructure design decisions, scientific exploration, and uncertainty quantification. While reduced models exist, they have limited utility because they are unable to reliably emulate the behavior of the relevant detailed models. In part, this is because their calibration poses a host of difficulties, including chaotic dynamics that prevent the use of adjoint methods, computational costs that become unreasonable when sampling approaches require many forward runs, and large existing code bases that require black box approaches. Recently, methods based on Kalman ensembles have shown promise by providing approximate derivatives of parameters in black box models in order to reach convergence using relatively few forward model runs. Another set of strategies for reducing computational costs are multi-fidelity methods, which combine various resolution model runs. Building on these approaches, and inspired by work in consistency testing for climate models, we propose an approach that utilizes ultra-short model runs to help improve calibration performance for reduced models, while still using full length runs to ensure convergence for the relevant quantities of interest. In contrast to other coarsening approaches, we are able to use the same model resolution throughout calibration, key when model parameters are resolution dependent. Most importantly, we show a strong reduction in computational cost on both classical chaotic test cases and reduced sub-grid models as part of a full global circulation model.

1. INTRODUCTION

The history of computer models used for weather and climate prediction is deeply intertwined with the history of computers themselves. From test problems run on ENIAC (Easterbrook et al. (2011)) to modern versions run on the largest supercomputers in the world, these models have consistently pushed the limits of available processing power. A single modern long-term Global Climate Model (GCM) simulation can require hundreds of thousands of core hours, equivalent to weeks on a large supercomputer. The extremely high computational requirements prevent the use of such models in most educational and industrial applications. In the scientific sphere, these costs preclude activities re-quiring many model runs, such as uncertainty quantification and exploration of rare events.

These difficulties motivate a long-term goal: to develop computationally cheaper alternative models for climate and weather prediction. The development and adoption of these less expensive, or reduced, models can only occur if their parameters can be effectively calibrated in a reasonable amount of time. In this work, we aim to reduce the computational cost of calibration for a specific subset of reduced models—those that are too expensive for sampling approaches, exhibit chaotic dynamics, and that preclude the use of adjoint methods.

Our proposed method incorporates information from ultra-short simulations, and demonstrates a

roughly 50% reduction in computation time while still utilizing long runs to ensure reliable long-term behavior. Our approach takes advantage of the access to highly granular ground-truth data provided by the expensive detailed model. We integrate ideas from consistency testing in climate modeling while building on recent work in Kalman calibration methods that can handle expensive black-box chaotic models.
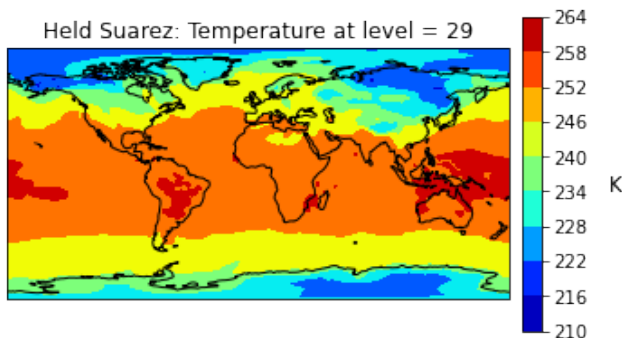

Figure 1: The uppermost temperature field used as initial conditions for the Held-Suarez GCM. Due to grid coarseness, many GCM phenomena are captured using sub-grid models.

2. BACKGROUND

Methods of model reduction vary widely; they may simplify the modeled systems based on some domain knowledge, increase the coarseness of spatial or temporal resolution, or replace the model entirely using emulators, like Gaussian Processes, or even deep-learning methods. Inferring the best parameters of these models from ground-truth output is referred to as calibration. If a reduced model cannot be calibrated effectively it will be of little use. Instead of using real-world observations as ground-truth, reduced models may rely on outputs from the detailed (high-fidelity) model for calibration, which tends to be much less expensive than direct observations (Sacks et al. (1989)). Further, detailed model output can provide a level of numerical, spatial, and temporal granularity impossible with physical observations.

Different calibration techniques address unique goals and constraints. Bayesian sampling methods like Markov Chain Monte Carlo (MCMC) (Hastings (1970)) require many forward runs of the model but provide a posterior probability distribution for parameters. Modifications to MCMC, like Hamiltonian Monte Carlo (HMC) (Duane et al. (1987)), have been developed to reduce the number of samples, but instead require gradient information from the model. Many stochastic gradient descent methods exist (Goodfellow et al. (2016)) which rely on the adjoint method to efficiently estimate the gradient. Recently, a group of Kalman inversion methods, outlined in the next section, have demonstrated promise for the types of models considered in this work.

Weather and climate models pose a number of challenges for calibration. First, they often involve large existing code bases, which are developed by domain experts and do not support the machinery required for automatic calculation of gradients. The Community Earth System Model, the GCM codebase used in this work, contains over one million lines of Fortran code. Second, they exhibit strong chaotic dynamics. This precludes the use of adjoint methods for gradient calculation (Lea et al. (2000)) and requires the use of long trajectories to ensure stable ergodic averaging (Strogatz (2018)). Third, even reduced weather and climate models can still be computationally expensive, easily requiring 10 minutes to compute a typical trajectory length.

An important inspiration for this work comes from the field of consistency testing in climate modeling, which seeks to ensure that changes occurring during the development cycle of a large climate model do not adversely affect the simulation. When changes occur in the scientific model, compilers, or even hardware, models cannot be expected to be bit-for-bit equivalent even if they are scientifically equivalent. Traditionally, consistency testing has been an onerous process involving several multi-hundred year simulations. However, recent work has shown that comparing ensembles of ultra-short simulations can achieve similar performance (Milroy et al. (2018)). These short simulations each simulate just a few hours, resulting in a vast speedup from traditional approaches.

Many important climate phenomena are unresolved at the tractable mesh sizes used in simulations (see Fig. 1). For this reason GCM's con-

tain a number of sub-grid models, or physics parameterizations, for mechanisms like cloud formation and precipitation. Calibration is particularly important for these embedded models because of their complexity and semi-empirical nature (Stensrud (2007)). As model resolution changes, the degree to which a particular phenomena is captured by the GCM's core dynamics varies, and so sub-grid model parameters will also vary. In contrast to many existing multi-fidelity methods (Peherstorfer et al. (2018)), our approach uses the same model configuration and resolution throughout calibration, circumventing this issue.

3. METHODOLOGY
3.1. *Unscented Kalman Inversion*

While a multi-time approach could theoretically be used in a variety of calibration methods, in this work it was implemented in conjunction with Unscented Kalman Inversion (UKI) as described by (Huang et al. (2021)). This choice was made considering the unique constraints of weather and climate models outlined above. UKI is able to treat models as black boxes, requires sufficiently few iterations to enable calibration of expensive models, and seems to function well for chaotic models, possibly due to the the averaging of the unscented transform. To make the nature of our modifications clear we first outline the key steps of UKI here.

UKI adopts the Unscented Kalman Filter to the purpose of inversion, that is, to determine an unknown set of parameters θ of a known model $\mathscr{G}(\theta)$ 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 or observations). We denote the full output of the model as $\mathbf{z} \in \mathbb{R}^{D \times T}$ where D is the dimension of the full model output at each time-step and T is the number of time-steps. Due to chaotic effects, in ergodic contexts the time-averaged observations are used:

$$\bar{\mathbf{y}} = \overline{\varphi(\mathbf{z})} = \overline{\varphi(\mathscr{G}(\theta))}, \qquad (1)$$

where $\varphi(\mathbf{z})$ is known as the observation operator. For brevity φ and the overbar are generally omitted, yielding $\mathbf{y} = \mathscr{G}(\theta)$.

A traditional Kalman Filter attempts to determine some underlying true state of a system $\mathbf{x} \in \mathbb{R}^L$

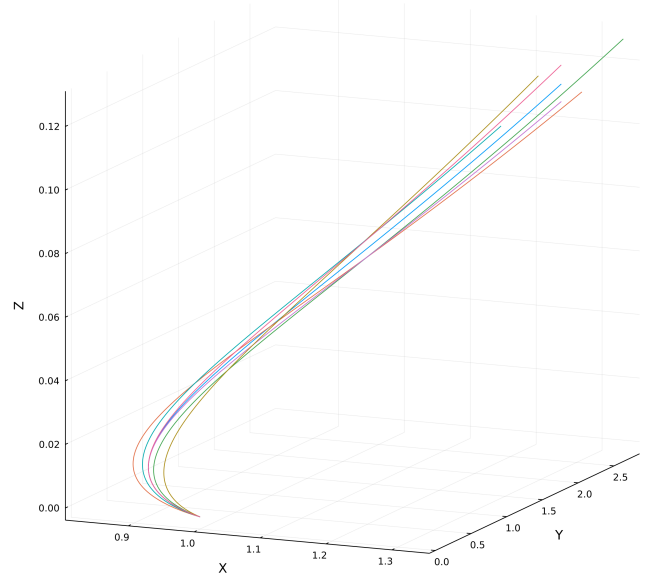Lorenz '63 Trajectories, Timespan = 0.1, Θ ± .7

Figure 2: A selection of Lorenz '63 ultra-short trajectories resulting from small perturbations to the model parameters. At this time-scale the signature exponential sensitivity of chaotic trajectories has yet to appear.

from observations $\mathbf{y} \in \mathbb{R}^N$. Then the system at time $k+1$ is described as:

$$\mathbf{x}_{k+1} = \mathscr{M}(\mathbf{x}_k) + w_k, \quad \mathbf{y}_{k+1} = \mathscr{H}(\mathbf{x}_{k+1}) + v_k, \ (2)$$

where $\mathscr{M}$ and $\mathscr{H}$ are known as the process and observation functions, respectively and w_k and v_k are process noise and observation noise, respectively. In UKI the unknown state $\mathbf{x}$ is actually the parameters $\theta \in \mathbb{R}^{N_\theta}$, now being *observed* through the model in question. Instead of time k referring to a time-step of the model it now refers to an iteration of the UKI algorithm. Thus:

$$\begin{aligned} x = \theta, \ \mathscr{M}(\theta) &= \theta + w, \ \mathscr{H}(\mathbf{x}) = \mathscr{G}(\theta) + v, \\ w &\sim \mathscr{N}(0, \Sigma_w), \quad v \sim \mathscr{N}(0, \Sigma_v). \end{aligned} \quad (3)$$

Prior knowledge about the parameters is expressed as $\theta \sim \mathscr{N}(\mathbf{m}_0, C_0)$. In our implementation the artificial evolution error covariance is set as $\Sigma_w = C_0$, a common choice in UKI literature. The artificial observation error covariance, Σ_v is set to reflect variance in model outputs due to chaotic behavior and thus reflects the scale of simulation outputs. Based on our use of the hyperparameters

given in Huang et al. (2021), constant weights are defined as:

$$a = \min\left(\frac{4}{N_\theta}, 1\right), \quad c_0 = c_j = \sqrt{a^2 N_\theta},$$

$$W_0 = 4 - a^2 - \frac{1}{a^2}, \quad W_j = \frac{1}{2a^2 N_\theta}. \quad (4)$$

A UKI iteration updates the estimated mean ($\mathbf{m}_n$) and covariance (C_n) of θ. It should be noted that C_n is not a true covariance, but instead functions to estimate the sensitivity of model outputs to θ. To avoid confusion we shall refer to it as the approximate covariance. With that clarification, and starting values and constants established above, a UKI iteration consists of the following:

1. Predict updated mean and approximate covariance:

$$\hat{\mathbf{m}}_n = \mathbf{m}_{n-1}, \quad \hat{C}_n = C_{n-1} + \Sigma_w \quad (5)$$

2. Generate deterministic ensemble of θ values, also known as sigma points:

$$\theta_n^0 = \hat{\mathbf{m}}_n,$$

$$\theta_n^j = \hat{\mathbf{m}}_n + c_j[\sqrt{\hat{C}_n}]_j,$$

$$\theta_n^{j+N_\theta} = \hat{\mathbf{m}}_n - c_j[\sqrt{\hat{C}_n}]_j, \quad (6)$$

$$\text{for } (1 \le j \le N_\theta),$$

where $[\sqrt{C}]_j$ is the j'th column of the Cholesky factor of C.

3. Analyze:

$$\hat{\mathbf{y}}_n^j = \mathscr{G}(\theta_n^j),$$

$$\hat{C}_n^{\theta p} = \sum_{j=0}^{2N_\theta} W_j(\theta_n^j - \hat{\mathbf{m}}_n)(\hat{\mathbf{y}}_n^j - \hat{\mathbf{y}}_n^0)^T,$$

$$\hat{C}_n^{pp} = \sum_{j=0}^{2N_\theta} W_j(\hat{\mathbf{y}}_n^j - \hat{\mathbf{y}}_n^0)(\hat{\mathbf{y}}_n^j - \hat{\mathbf{y}}_n^0)^T + \Sigma_v, \quad (7)$$

$$\mathbf{m}_n = \hat{\mathbf{m}}_n + \hat{C}_n^{\theta p}(\hat{C}_n^{pp})^{-1}(\mathbf{y}_{\text{true}} - \hat{\mathbf{y}}_n^0),$$

$$C_n = \hat{C}_n + \hat{C}_n^{\theta p}(\hat{C}_n^{pp})^{-1}(\hat{C}_n^{\theta p})^T.$$

Let us summarize the purpose of these steps. First, the artificial process noise is added to the approximate parameter covariance. This sets a lower floor for the spread of the ensemble and helps ensure the approximate covariance matrix be positive definite. Second, using the current approximate covariance and mean estimate, an ensemble of parameter values is created. Third, the forward model is run with those parameters. With the model outputs, the local gradient of parameters to outputs is estimated. Finally, the parameter mean and approximate covariance estimates are updated.

3.2. Multi-Time Modifications

In order to reduce the computational expense of the UKI algorithm, we create a second ground-truth output corresponding to an ultra-short model run. This can be selected directly from the original ground-truth trajectory. For clarity, we now distinguish between ergodic and ultra-short model runs as follows:

$$\mathbf{y}_{\text{true}} = \mathscr{G}_{\text{long}}(\theta) = \mathscr{G}(\theta, t_{\text{long}}),$$

$$\mathbf{y}_{\text{true}}^* = \mathscr{G}_{\text{short}}(\theta) = \mathscr{G}(\theta, t_{\text{short}}), \quad (8)$$

$$t_{\text{short}} \ll t_{\text{long}}.$$

In the short-time domain we compare trajectories before chaotic sensitivity has taken over (see Fig. 2). Because this negates the need for time-averaging, and because the impact of parameter changes on trajectories may be relatively small, $\varphi(\mathbf{z})$ does not apply time-averaging when used for ultra-short trajectories. Based on this new $\mathbf{y}_{\text{true}*}$ we also create a new Σ_v^* to account for scale differences between ultra-short and ergodic timescale outputs.

With these new values, and the ability to run the model for an arbitrary trajectory length, we can now run an iteration of the UKI algorithm at either the ergodic or ultra-short timescale. We briefly describe alternative methods of combining iterations:

1. **Simple Alternating:** A simple version consists of running a UKI iteration using the ultra-short ground-truth, model, and Σ_v^*, then swapping to the ergodic versions for the subsequent iteration and back again. Because the accuracy of ergodic outputs are considered at every second iteration, this approach helps provide confidence in the eventual calibrated parameter values. This approach seems to

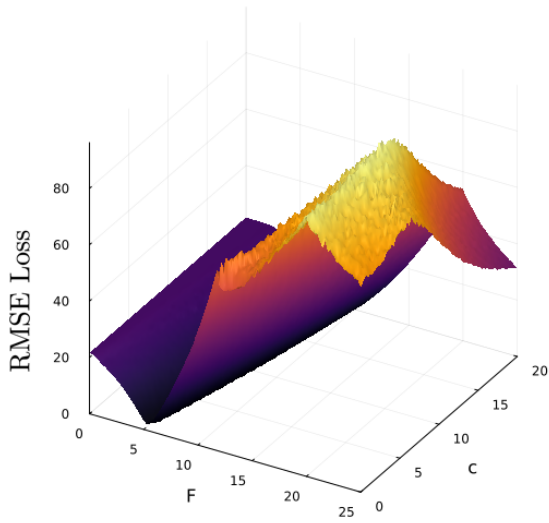Loss due to Parameter Variation in Ergodic Lorenz '96

Figure 3: The RMSE loss due to perturbed parameters (F and c) for ergodic Lorenz '96 trajectories. Chaotic dynamics result in noise in the loss-landscape, inhibiting calibration.

2. **Block Alternating:** Alternatively, one can take multiple iterations at each timescale before swapping. In our experience, this can help if the loss-landscape is less well behaved and the algorithm takes longer to converge to correct parameter values.

3. **Pretuning:** Distinctly, one may first calibrate in the short-time setting until parameter estimates stabilize, and then use those estimates as the prior mean for ergodic calibration. One should use the original covariance estimate to ensure the ergodic calibration can still explore the parameter space effectively. In contrast to the simple alternating case, pretuning can still work when sensitivity of certain parameters is substantially different at the two timescales.

Loss due to Parameter Variation in Ultra-Short Lorenz '96

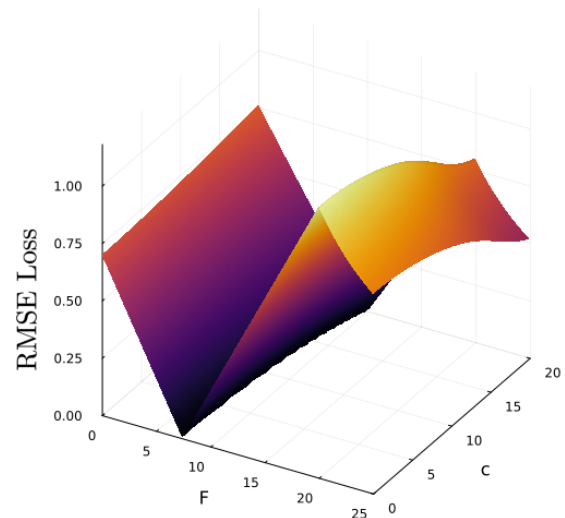*Figure 4: The RMSE loss due to perturbed parameters (F and c) for ultra-short Lorenz '96 trajectories. Because chaotic effects have yet to dominate, the landscape is smooth.*

3.3. Multi-Time UKI Challenges and Disadvantages

Many of the issues associated with Multi-Time UKI concern the introduction of additional hyperparameters, and are similar in nature to those found in the UKI methods itself. Main difficulties include:

1. **Estimating Internal Variability of Ultra-Short Trajectories:** Characterizing the internal variability of short runs of the model, in the form of Σ_ν^*, is important for performance of the multi-time approach. Authors of the original UKI paper (Huang et al. (2021)) first suggest splitting an extra-long model run and estimating the covariance from those sub-trajectories; if this approach is ineffective, they then suggest a heuristic of 1-3% of the output magnitude. The former approach cannot be applied for ultra-short trajectories since they do not undergo time-averaging. The latter is applicable but seems to over-estimate variability for ultra-short trajectories. We have found better success with 0.01% of the ground-truth magnitude.

2. **Poorly Aligned Covariance Between Er-**

work well when the loss-landscapes are reasonably aligned between the short and long time-domains. It can be less successful if the sensitivity to different parameters is substantially different at the two timescales.

godic and Ultra-Short Trajectories: Ultra-short trajectories, depending on the part of the domain, can be much more sensitive to certain parameters. This leads to covariance estimates that don't function well when used for ergodic iterations. Pretuning can be a good alternative as it does not re-use the covariance matrix.

3. **Difficulty Enforcing Strong Prior Distributions:** Similar to traditional UKI, priors over the parameter space are difficult to enforce. This is especially true if they involve absolute bounds such [0, 10]. Previous UKI work involves the use of transforms to ensure only parameter values within the prior bounds are used, but this approach can modify the relative density of the prior distribution substantially.

For the first two categories, there is potential to develop better adaptive schemes, similar to those that have been implemented in Bayesian calibration (Haario et al. (2006)) and gradient based optimization (Kingma and Ba (2017)).

3.4. *Multi-Time UKI Advantages*

The ability of Multi-Time UKI to reduce computational cost is demonstrated through upcoming examples; we highlight more conceptual benefits of the approach here.

Figs. 3 and 4 are created using a multi-scale (2 level) Lorenz '96 model (Lorenz (1996)) with a setup and observation operator as in Cleary et al. (2021). The loss is the Root Mean Square Error (RMSE) between model output using true parameters ($\mathbf{y}_{\text{true}} \in \mathbb{R}^5$) and perturbed parameters ($\mathbf{y}'$), where

$$\text{RMSE}(\mathbf{y}_{\text{true}}, \mathbf{y}') = \sqrt{\frac{1}{N} \sum_{i=0}^{i=N} (y_{\text{true},i} - y_i')^2} \quad (9)$$

The relative smoothness of a loss-landscape for ultra-short trajectories becomes muddied with the presence of chaotic noise in the ergodic loss-landscapes. This noise impacts the ability of UKI to effectively estimate local gradients.

Other key benefits include the ability to utilize the same resolution throughout calibration, and the ability to vary the emphasis on ergodic model outputs using different approaches highlighted in section 3.2.

4. NUMERICAL RESULTS
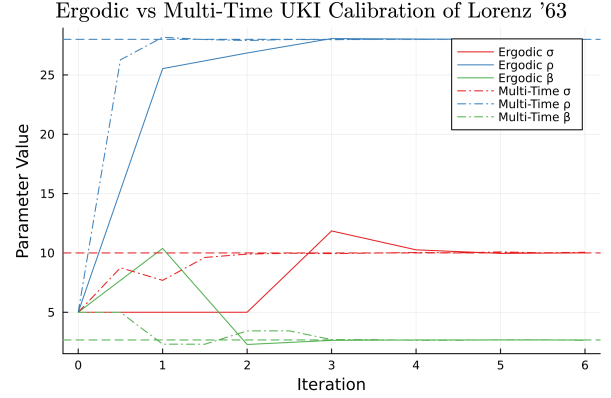
4.1. *Lorenz '63*

Figure 5: Calibration performance for the Lorenz '63 model using Ergodic UKI (solid lines) and Multi-Time UKI (dot-dash lines). True parameter values are indicated by horizontal dashed lines.

We first consider the 3-dimensional Lorenz '63 system (Lorenz (1963)), a simplified description of a fluid warmed from above and cooled from below, that is one of the most studied chaotic systems. The Lorenz '63 differential equations are,

$$\frac{dx}{dt} = \sigma(y - x),$$
$$\frac{dy}{dt} = x(\rho - z) - y, \quad (10)$$
$$\frac{dz}{dt} = xy - \beta z,$$

with parameters $\theta = (\sigma, \rho, \beta) \in \mathbb{R}^+$. We seek to recover $\theta_{\text{true}} = (10, 28, 8/3)$ starting from $\theta_0 = (5, 5, 5)$ and $C_0 = 2\mathbb{I}$. We define $\varphi(\mathbf{z})$ as

$$\mathbf{y} = \varphi(\mathbf{z}) = (x, y, z, x^2, y^2, z^2). \quad (11)$$

We set $t_{\text{long}} = 500$ for ergodic iterations, discarding the transient up to $t = 20$. For ultra-short trajectories we use $t_{\text{short}} = 0.01$ without any discard. For the ergodic case, Σ_ν is estimated by splitting the ground-truth, $\mathbf{y}_{\text{long}}$, trajectory into 10 sub-trajectories and calculating covariance between the

sub-trajectories as in Huang et al. (2021). Process noise in the ultra-short case (Σ_v^*) is estimated as

$$\Sigma_v^* = (0.0001)\mathbb{I}y_{\text{short}}. \qquad (12)$$

Iterations are handled as in the "Simple Alternating" case described above, beginning with the ultra-short iteration. With a computation cost 50,000 times less than an ergodic iteration, for the purposes of plotting and discussion an ultra-short iteration is considered zero iterations long. Fig. 5 shows that Multi-Time UKI requires just 3 iterations to match the true parameters vs 5 iterations for Ergodic UKI (using a tolerance of 5%.) This is a substantial 40% speedup.
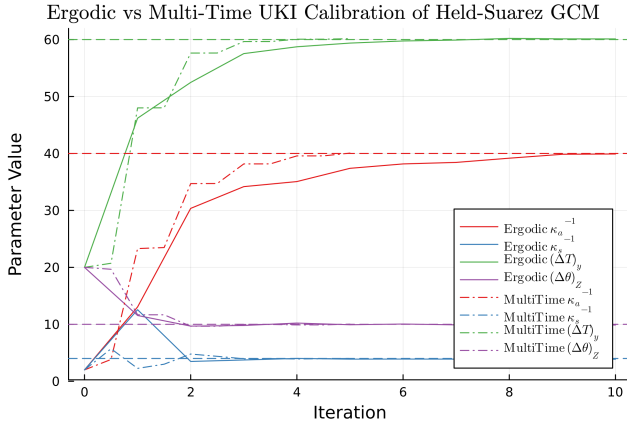
4.2. *Held-Suarez GCM Test Case*

Figure 6: Calibration performance for the Held-Suarez model using Ergodic UKI (solid lines) and Multi-Time UKI (dot-dash lines). True parameter values are indicated by horizontal dashed lines.

We now test our method using the Held-Suarez test case (Held and Suarez (1994)), an idealized sub-grid parameterization scheme for a GCM. We adopted the implementation developed for the Community Atmospheric Model (CAM) in CESM. The Held-Suarez model provides temperature and wind-speed derivatives for each grid cell at each time-step via a Newtonian relaxation to an equilibrium field. These derivatives are then used by the atmospheric solver in a Eulerian spectral-transform dynamical core.

The Held-Suarez model requires substantial computational resources to run. A 400-day simulation takes roughly 6 minutes when running on

96 cores. While substantially less than the multiple hours required for the full-physics version of CAM with similar resources, this is still too long to feasibly implement MCMC-type methods.

We calibrate four parameters that control the equilibrium field and the speed of relaxation, $(\kappa_a^{-1}, \kappa_s^{-1}, \Delta T_y, \Delta\theta_Z)$. Analogous to the Lorenz '63 example, we aim to recover true parameter values $\theta_{\text{true}} = (40, 4, 60, 10)$ with prior values $\theta_0 = (2, 2, 20, 20)$ and $C_0 = \mathbb{I}$. We use $\mathbf{y} = \varphi(\mathbf{z}) \in \mathbb{R}^{64}$ corresponding to temperature values averaged in the zonal and vertical dimensions. We set trajectories of $t_{\text{long}} = 400$ days for ergodic iterations, discarding a transient of $t = 200$ days. We set $t_{\text{short}} = 3$ hours without transient for ultra-short trajectories.

To constrain the model parameters as

$$\kappa_a^{-1} > 1, \quad \kappa_a^{-1} > \kappa_s^{-1} > 1$$
$$0 < \Delta T_y, \quad 0 < \Delta\theta_Z, \qquad (13)$$

we define the following transforms (where θ are parameters used in the UKI algorithm),

$$\kappa_a^{-1} = 1 + |\theta_1|, \quad \kappa_a^{-1} = \frac{1 + |\theta_2| + |\theta_1\theta_2|}{1 + |\theta_1|}$$
$$\Delta T_y = |\theta_3|, \quad \Delta\theta_Z = |\theta_4|. \qquad (14)$$

Once again, iterations are handled with the "Simple Alternating" case, beginning with the ultra-short iteration. Using a 5% tolerance, Multi-Time UKI requires just 3 iterations to match the true parameters vs 9 iterations for Ergodic UKI (Fig. 6). This is a substantial 66% speedup. Using a stricter 1% tolerance, Multi-Time UKI requires 5 iterations while Ergodic UKI remains the same, still resulting in a 45% speedup.

5. CONCLUSION

Multi-Time UKI demonstrates tangible opportunities for speedups, even when used with already performant algorithms. The underlying ideas of Multi-Time UKI are largely agnostic to the calibration algorithm used, enabling their adaptation to new developments in the field. While our implementation of UKI is limited to approximating sensitivities, recent work has opened the door to estimating approximate posteriors using Kalman inversion (Huang et al. (2022)). A Multi-time approach

could function as a type of preconditioner for such methods, preventing costs from growing too large.

Adaptive methods to estimate the artificial evolution and observation error covariance, control the swapping between trajectory lengths, and even manage instability could all improve performance while lowering the calibration expertise required. The adoption of batching methods used in deep-learning may enable a lower proportion of ergodic length iterations, in turn allowing for the calibration of far more complex physics parameterizations.

6. ACKNOWLEDGEMENTS

7. REFERENCES

Cleary, E., Garbuno-Inigo, A., Lan, S., Schneider, T., and Stuart, A. M. (2021). "Calibrate, emulate, sample." *Journal of Computational Physics*, 424, 109716.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). "Hybrid Monte Carlo." *Physics Letters B*, 195(2), 216–222.

Easterbrook, S., Edwards, P., Balaji, V., and Budich, R. (2011). "Guest Editors' Introduction: Climate Change - Science and Software.." *IEEE Software*, 28, 32–35.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning* (November) Google-Books-ID: omivDQAAQBAJ.

Haario, H., Laine, M., Mira, A., and Saksman, E. (2006). "DRAM: Efficient adaptive MCMC." *Statistics and Computing*, 16(4), 339–354.

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57(1), 13.

Held, I. M. and Suarez, M. J. (1994). "A Proposal for the Intercomparison of the Dynamical Cores of Atmospheric General Circulation Models." *Bulletin of the American Meteorological Society*, 75(10), 1825–1830.

Huang, D. Z., Huang, J., Reich, S., and Stuart, A. M. (2022). "Efficient Derivative-free Bayesian Inference for Large-Scale Inverse Problems." *arXiv:2204.04386 [cs, math]* arXiv: 2204.04386.

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

Kingma, D. P. and Ba, J. (2017). "Adam: A Method for Stochastic Optimization, <http://arxiv.org/abs/1412.6980> (January). arXiv:1412.6980 [cs].

Lea, D. J., Allen, M. R., and Haine, T. W. N. (2000). "Sensitivity analysis of the climate of a chaotic system." *Tellus A: Dynamic Meteorology and Oceanography*, 52(5), 523–532.

Lorenz, E. N. (1963). "Deterministic Nonperiodic Flow." *Journal of the Atmospheric Sciences*, 20(2), 130–141.

Lorenz, E. N. (1996). "Predictability: A problem partly solved." Vol. 1. Issue: 1.

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

Peherstorfer, B., Willcox, K., and Gunzburger, M. (2018). "Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization." *SIAM Review*, 60(3), 550–591.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). "Design and Analysis of Computer Experiments." *Statistical Science*, 4(4), 409–423.

Stensrud, D. J. (2007). *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models* (May) Google-Books-ID: lMXSpRwKNO8C.

Strogatz, S. H. (2018). *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (May) Google-Books-ID: 1kpnDwAAQBAJ.