FY24 R&TD Innovative Spontaneous Concepts (ISC)

Accelerated MCMC for CARDAMOM

Principal Investigator: Otto Lamminpää (398); **Co-Investigators:** Anthony Bloom (329)

Strategic Focus Area: Innovative Spontaneous Concepts

Objectives: Our objective was to replace the CARDAMOM MCMC algorithm with the dimension-reduced sampler demonstrated in OCO-2 retrieval algorithm by Lamminpaeae et al (2019). The main work was to interface the MCMC sampler with the CARDAMOM framework. The CARDAMOM land model infrastructure had been refactored, which was leveraged to allow interfacing with the proposed approach, with likelihood function metrics stored in memory for use by the novel MCMC technique.

Background: The JPL-led CARDAMOM Bayesian land model-data integration framework is instrumental for synthesizing synergistic land surface observations of the terrestrial carbon and water cycles, including observations of MODIS vegetation and fire indices, multi-platform fluorescence products, terrestrial water storage anomalies from GRACE, and land-atmosphere CO2 fluxes inferred from OCO platforms. The CARDAMOM framework critically relies on Markov Chain Monte Carlo (MCMC) samplers to retrieve (i) land model parametrization which minimizes model-data inconsistencies, and (ii) full uncertainty metrics critical for testing and resolving science hypotheses and for supporting mission formulation analyses. Throughout the wide use of CARDAMOM across multiple NASA-funded and/or academic efforts, a critical system bottleneck is the current Bayesian MCMC samplers used in CARDAMOM: while these provide stable output, the operational use of CARDAMOM is compromised by the relatively slow computational speed (~48 hours per pixel), which makes global-scale analyses at low or moderate resolutions expensive or outright prohibitive.

Significance/Benefits to JPL and NASA: The innovation in the proposed effort was to leverage OCO-2 retrievals MCMC algorithms and know-how to accelerate CARDAMOM use and applications across the lab and the broader community. When successful, this effort will enable new JPL-led science and mission formulation capabilities and will enable currently unattainable CARDAMOM applications. The improved capability will allow for CARDAMOM to be used for formulating competitive JPL-led or JPL-participating ROSES proposals in terrestrial carbon and water sciences. Furthermore, NASA's AIST thrust on Earth System Digital Twins will directly benefit from a faster CARDAMOM model.

Approach: As MCMC suffers from the curse of dimensionality, we use Likelihood Informed methods to find a low dimensional subspace (LIS) in the model input space, in which the sampling leads to significantly faster convergence. The forward model induces this structure, as the corresponding Jacobian matrix has only a few large singular values in its SVD. As Jacobians are not available, we have introduced a new gradient-free method for computing the LIS:

$$H = \frac{1}{N(N-1)} \sum_{j<i} \left(\mathcal{L}(x_i) - \mathcal{L}(x_j) \right)^2 \frac{(x_i - x_j)(x_i - x_j)^T}{||x_i - x_j||^2}$$

Our MCMC sampling proceeds as follows:

1. Sample the input space with a Sobol sequence $\{x_i\}_1^N$ to evenly cover it.
2. Obtain a prior $\mathcal{N}(x_a, L_a L_a^T)$ from mean and covariance of $\{x_i\}_0^N$
3. Evaluate CARDAMOM likelihood $\mathcal{L}(x_i)$ over $\{x_i\}_1^N$
4. Compute $L_a^T H L_a$ over $\{x_i\}_1^N$ and use SVD to obtain LIS and CS bases P_r and $P_\perp$
5. Adaptive Metropolis on low dimensional LIS, for t = 1, … ,T:
    1. Propose $u_r \sim \mathcal{N}(0, C_t)$ and project to $x_r^{t+1} = x_r^t + P_r u_r$
    2. Marginalize over CS: $\tilde{\mathcal{L}}(x_r^{t+1}) = \frac{1}{N} \sum_{i \in CS} \mathcal{L}(x_r^{t+1} + x_\perp^{(i)})$
    3. Accept with probability $\alpha = \min(1, \frac{\tilde{\mathcal{L}}(x_r^{t+1})}{\tilde{\mathcal{L}}(x_r^t)})$
    4. Sample CS from earlier points $x_\perp^{(i)}$, set $x^{t+1} = x_r^{t+1} + x_\perp^{t+1}$
    5. Every k iterations, update C_t to match accepted points so far.

Results: Our proposed MCMC method (**blue**) compared with points taken from a converged CARDAMOM run (**red**). Top panel shows the chain and lower panel the histograms from 3 different state parameters. We get good agreement (Fig 1), more information (bi-modal, Fig 2), and insufficient agreement (Fig 3) with operational MCMC all at the same time, which shows that further research, comparison, and tuning is still needed.

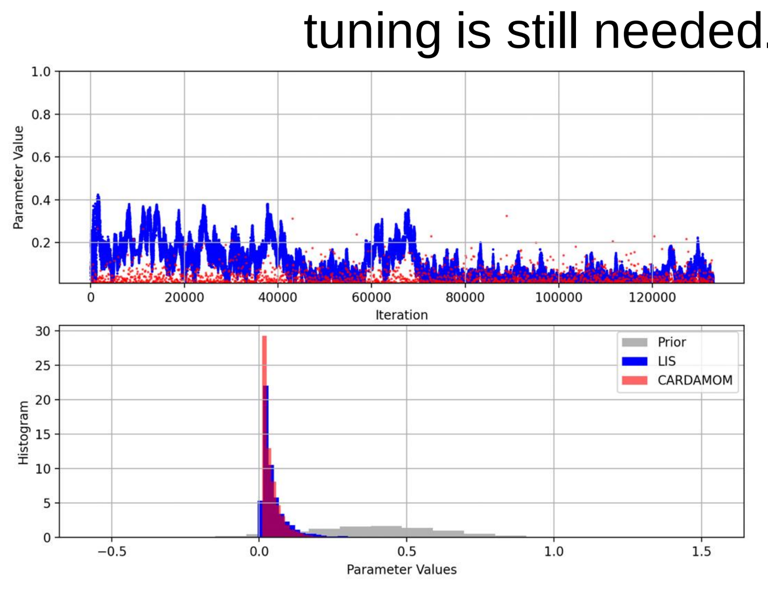

Figure 1

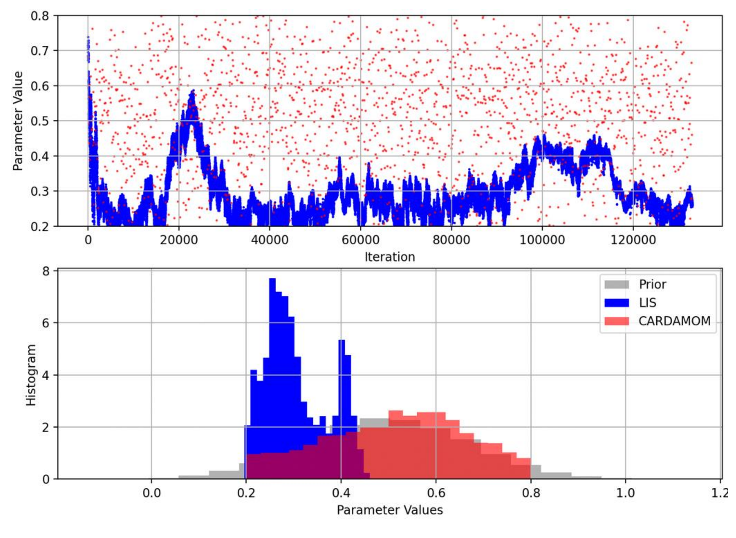

Figure 2

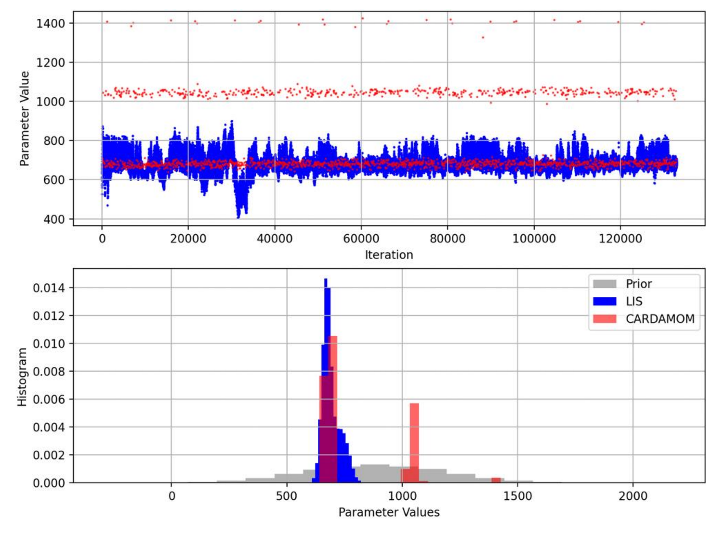

Figure 3

References:

Will be updated before printing.

PI/Task Mgr. Contact Information:
Otto.m.lamminpaa@jpl.nasa.gov

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

RPD-000 Clearance Number: CL#00-0000