National Aeronautics and Space Administration

Accelerating MCMC to Operational Speeds

Principal Investigator: Amy Braverman (398); Co-Investigators: Jonathan Hobbs (398), Vijay Natraj (329), David Thompson (382), Youssef Marzouk (MIT)

Program: FY22 SURP
Strategic Focus Area: Uncertainty Quantification

Objectives

Our objective was to develop a Markov chain Monte Carlo (MCMC) retrieval algorithm for Earth remote sensing that is fast enough for operations. MCMC is computationally intensive, and generally too slow for operational use. In this project, we investigate a set of computational and theoretical modifications to standard MCMC algorithms to achieve the greater speed. We used the upcoming Surface Biology and Geology (SBG) mission as a motivating example.

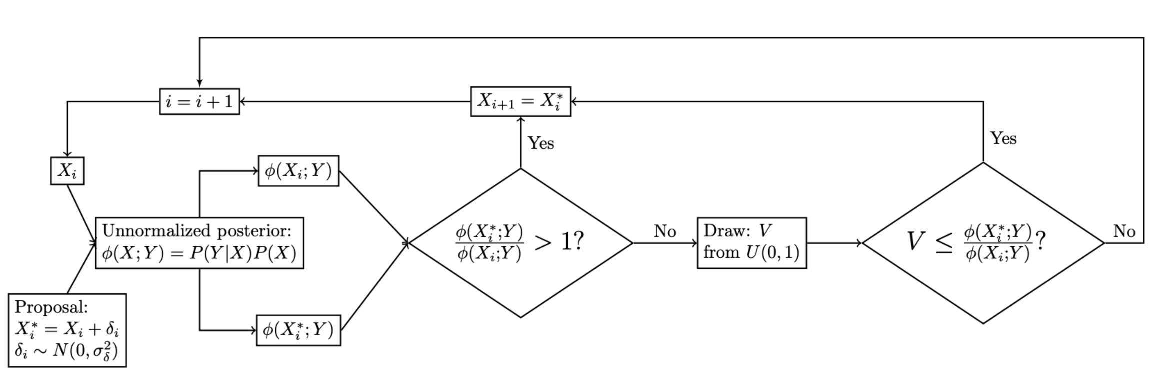
Background

A natural approach to the geophysical retrieval problem is to invoke Bayes' Rule: if X and Y are two jointly distributed random variables (or vectors) then the posterior distribution of X, $P(X|Y)$, is proportional to the product of the likelihood, $P(Y|X)$, and the prior distribution, $P(X)$:

$$P(X|Y) \propto P(Y|X)P(X).$$

$P(X|Y)$ reflects a new description of the behavior of X, based on what is learned from observing Y. By providing this posterior *distribution* of the geophysical state given the observed spectra, uncertainty information is incorporated directly into the retrieval output.

The traditional approach to computing $P(X|Y)$ for remote sensing problems, Optimal Estimation, assumes Gaussianity. This is a risky assumption because it's difficult to verify.



MCMC (see above) is a widely-used simulation-based method for sampling from posterior distributions. It does make require Gaussian assumptions, but it is slow. There are two reasons: high-dimensionality and the need to evaluate complex, computationally intensive forward models, F. Since $Y = F(X) + e$ this relationship defines $P(Y|X)$, and so must be evaluated on every iteration of MCMC.

Approach and Results:

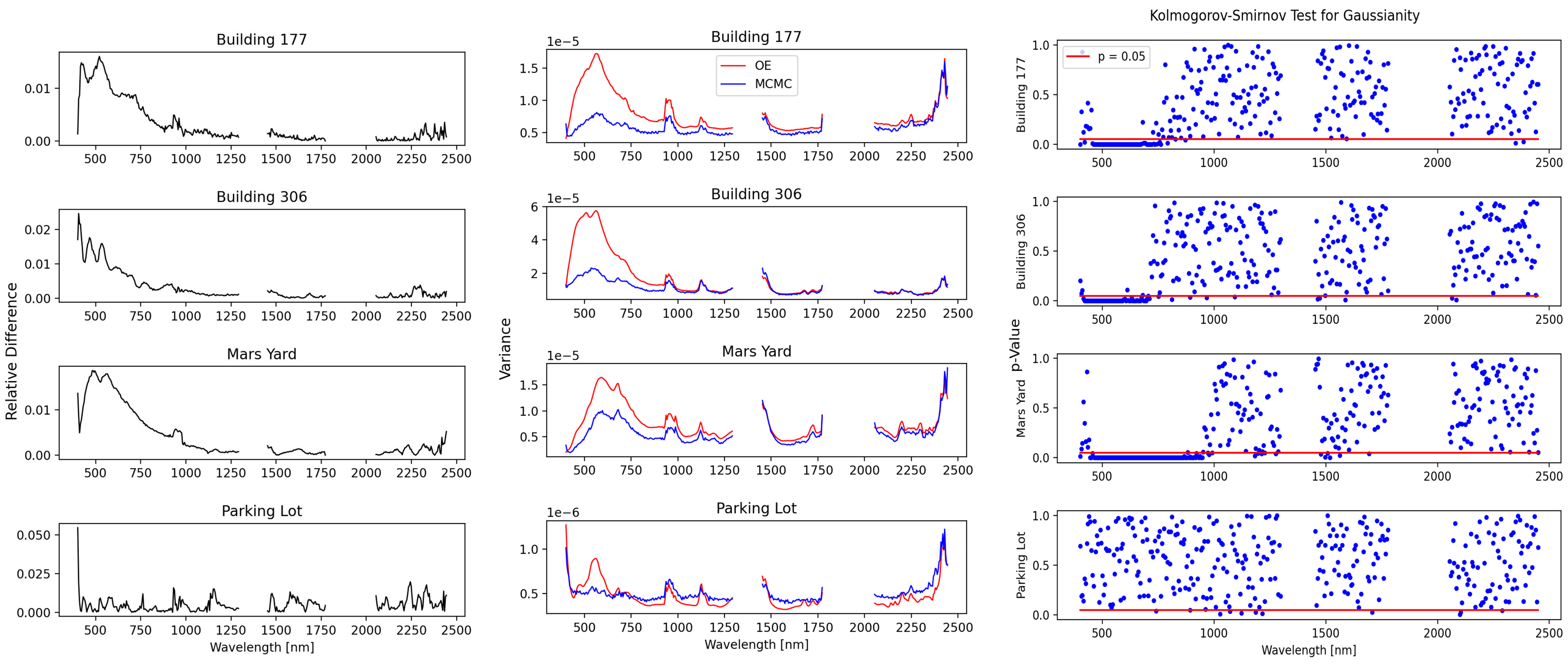
We developed a new sampling scheme for the "proposal" MCMC step, the **Block Metropolis Algorithm** (BMA), that exploits special structure in the SBG state vector $X = (X_{atm}, X_{surf})$. X_{atm} is a low-dimensional atmospheric state vector comprised of water vapor and aerosol optical depth. X_{surf} is a high-dimensional surface state vector (425-dimensional).

The key to BMA is that it samples separately from the atmospheric and surface components:

- ✓ X_{atm} is low-dimensional so sampling is fast.
- ✓ For a fixed value of X_{atm} the distribution $P(Y|X)$ approximately Gaussian because, for fixed X_{atm} $F(X)$ is a linear function.

Algorithm 1 Block Metropolis

0: Initialize $x^{(0)} = x_{\text{MAP}}$
for $i = 1 \ldots N$ **do**
Sample $x^{(i)}_{\text{atm}}$
Proposal $z_{\text{atm}} \sim \mathcal{N}(x^{(i-1)}_{\text{atm}},\ \Gamma^{(i)}_{\text{atm}})$
Metropolis accept/reject for $[x^{(i-1)}_{\text{refl}}, z_{\text{atm}}]$
Sample $x^{(i)}_{\text{refl}}$
Proposal $z_{\text{refl}} \sim \mathcal{N}(x^{(i-1)}_{\text{refl}},\ \epsilon_2\, \Gamma_{\text{Laplace}})$
Metropolis accept/reject for $[z_{\text{refl}}, x^{(i)}_{\text{atm}}]$
Compute $\Gamma^{(i+1)}_{\text{atm}}$
end for=0



Left: BMA posterior means for four case-study scenes acquired by AVIRIS-NG instrument. Center: BMA posterior variances compared to those obtained from Optimal Estimation (OE). Right: Results from tests of the null hypothesis that the BMA posterior distributions are Gaussian. The horizonal red line is at 0.05.

Conclusions: (1) BMA was fast enough to run in reasonable time for these four case study scenes. Full MCMC was not possible at all. (2) The posterior distribution of the state vector is non-Gaussian at most of the low wavelengths in three of the four scenes. (3) In that same wavelength region, Optimal Estimation overestimates the posterior variance.

Significance to JPL and NASA:

- ✓ BMA and similar approaches for other problems will increase the science return from analyses of past, present, and future JPL and NASA missions. By providing uncertainty information required to conduct formal statistical tests of scientific hypotheses, this work will raise the level of rigor that is possible in such studies.
- ✓ This will make JPL more competitive in winning new missions by providing retrieval algorithms that exploit observations more fully than do current methods.
- ✓ PL benefits from a continuing relationship with the UQ Group at MIT. The collaboration will stimulate basic research required to move operational retrieval algorithms for Earth remote sensing closer to state-of-the-art estimation and uncertainty quantification.

Publications:

Kelvin Leung, "Accelerating Bayesian Computation in Earth Remote Sensing Problems," Master of Science in Aeronautics and Astronautics Thesis, Massachusetts Institute of Technology, June 2021. Available at https://dspace.mit.edu/handle/1721.1/139062?show=full.

PI/Task Mgr. Contact Information:

Email: Amy.J.Braverman@jpl.nasa.gov

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

Clearance Number: CL#22-5127
Poster Number: RPC#SP20009