Name: Myisha Hassan
Date: 11/5/25
Paper Title: Applying Gaussian Process Machine Learning and Modern Probabilistic Programming to Satellite Data to Infer CO2 Emissions
Author Names: Seongeun Jeong, Sofia D. Hamilton, Matthew S. Johnson, Dien Wu, Alexander J. Turner, and Marc L. Fischer
Year Published: 2025

Open questions for discussion in class:

- How scalable are fully Bayesian gaussian process models when applied to global-scale satellite data with millions of observations? Since the FBGP model is computationally expensive, so for massive global satellite datasets with millions of points, this seems kind of unfeasible. Are approximation methods the only practical way forward and will we lose some accuracy using them?
- Could similar GP based approaches be used to infer other greenhouse gases (like CH4 and N2O) with different measurement uncertainties?
- The paper contrasts the "Fully Bayesian" approach with the "Classical Bayesian" one, especially for estimating hyperparameters. Is the main takeaway that "classical" methods are just a poor approximation or are they still a reasonable choice if you don't have the computational budget for a full MCMC?

The topic areas covered by the paper are:

- Atmospheric inverse modeling for estimating CO2 emissions
- Application of Gaussian Process machine learning to environmental satellite data
- Using modern Probabilistic Programming Languages like PyMC and GPyTorch complex Bayesian modeling
- Modeling complex spatiotemporal covariance structures in observational data using GP kernels
- Bayesian hyperparameter estimation (noise, length scales) as an alternative to prescribing fixed values

The previous approaches to this problem were:

- Traditional atmospheric inverse models that fail to properly incorporate the spatial and temporal correlations present in satellite data
- Methods that lack ways to estimate key hyperparameters. Instead, parameters like model data mismatch (noise) and covariance length scales were prescribed using ad-hoc methods or values from other studies
- Classical Bayesian analytical solutions, which are less flexible and struggle to estimate all parameters (like the noise term) simultaneously with the state vector
- Standard linear inversion models which may not capture the complex nonlinear relationships in the system

Outline the basic new approach or approaches to this problem:

- It develops a Fully Bayesian Gaussian Process inversion system to infer emission scaling factors
- It treats the CO2 concentration as an unknown function drawn from a GP which is defined by a mean function (based on a chemical transport model) and a covariance function (kernel)
- It models the covariance structure using a spatiotemporal kernel (a product of a Matern 5/2 for space and a squared exponential for time)
- It uses modern PPLs (PyMC) with an MCMC sampler (NUTS) to simultaneously infer both the main parameters (the emission scaling factors) and all the model's hyperparameters (like noise level and kernel length scales)
- It validates the model using prior and posterior predictive checks (PPCs) to make sure the model's assumptions and results are consistent with the data

Critical assumptions made include:

- GEOS-Chem simulations (which informs the GP's mean function) approximate the true atmospheric CO2 field used for evaluation
- The selected kernels adequately capture both spatial and temporal dependencies
- Observation noise is Gaussian with stationary variance
- Scaling factors for emissions across sectors are constant over the modeled time
- Computational inference (via MCMC and NUTS) converges to the correct posterior

The performance of the techniques discussed in the paper was measured in what manner:

- The primary evaluation was a "supervised learning" (synthetic data) experiment
- The authors defined a set of true emission scaling factors and a true noise level which were hidden from the model
- Performance was measured by the model's ability to accurately get these known true parameters
- They compared the FBGP model's posterior distributions against the true values for both the scaling factors and the noise hyperparameter
- The FBGP results were benchmarked against two other methods: a simpler GP marginal log-likelihood optimization and a traditional Classical Bayesian method
- They used prior and posterior predictive checks to visually confirm that the model's simulated data distributions matched the observed (synthetic) data distribution
- Runtime and efficiency tradeoffs between FBGP, GP MLL, and CB methods

What background techniques are used in the paper that you are not familiar with:

- GEOS-Chem: This is a specific Chemical Transport Model (CTM). The details of how these atmospheric models work are new to me.
- NUTS (No-U-Turn Sampler): I'm familiar with MCMC and Hamiltonian Monte Carlo (HMC) in concept, but not the specific implementation and advantages of the NUTS algorithm
- Matern 5/2 kernel: I've mostly seen the standard RBF kernel using in class. The

specific properties of the Matern class of kernels are unfamiliar.
- GP MLL (marginal log-likelihood): I hadn't seen this specific optimization approach for GPs contrasted with a fully Bayesian MCMC method before.
- Fisher Information Matrix (FIM): The paper mentions using this to estimate the uncertainties for the GP MLL method, which is a new concept for me.

The following terms were defined:
- Gaussian Process (GP): nonparametric approach defining a prior probability distribution over functions, characterized by a mean and a kernel
- Kernel (Covariance Function): function that defines the covariance between any two points in the input space and is used to build the covariance matrix
- Probabilistic Programming Languages (PPLs): software (like PyMC, GPyTorch) used to build and infer Bayesian models
- FBGP (Fully Bayesian GP): paper's main model, which treats all hyperparameters (like noise and length scales) as random variables to be inferred
- GP MLL (GP using marginal log-likelihood): an alternative approach that finds a single point estimate for hyperparameters by maximizing the data likelihood, rather than sampling
- CB (Classical Bayesian): a reference method based on an analytical solution which struggles to estimate hyperparameters
- Prior/Posterior Predictive Check (PPC): methods used to validate a Bayesian model by sampling new data from the prior or posterior distributions and comparing it to the actual observed data
- NUTS (No-U-Turn Sampler): a specific & efficient MCMC algorithm (an extension of HMC) used to sample from the model's posterior distribution

I rate and justify the value of this paper as:

9/10. This paper is a clear example of why the fully Bayesian approach we're learning about in class is so important for real world problems. Instead of just picking a noise value or hyperparameters and hoping for the best, it shows how to infer them directly from data using Gaussian Processes and MCMC. I especially liked their use of a synthetic data experiment because it was a clever way to definitively prove that their model could find the 'truth' when it was known. I also liked how the authors compared multiple approaches (fully Bayesian GP, marginal log-likelihood GP, and classical Bayesian inversion) rather than just presenting one method. The computational cost seems like a massive barrier to scaling this up to global high-res satellite datasets which they acknowledge but don't fully solve. Some of the technical parts (like the part on spatiotemporal kernel construction and the NUTS sampling algorithm) a bit dense on first read… a bit of visuals or diagrams would've helped.