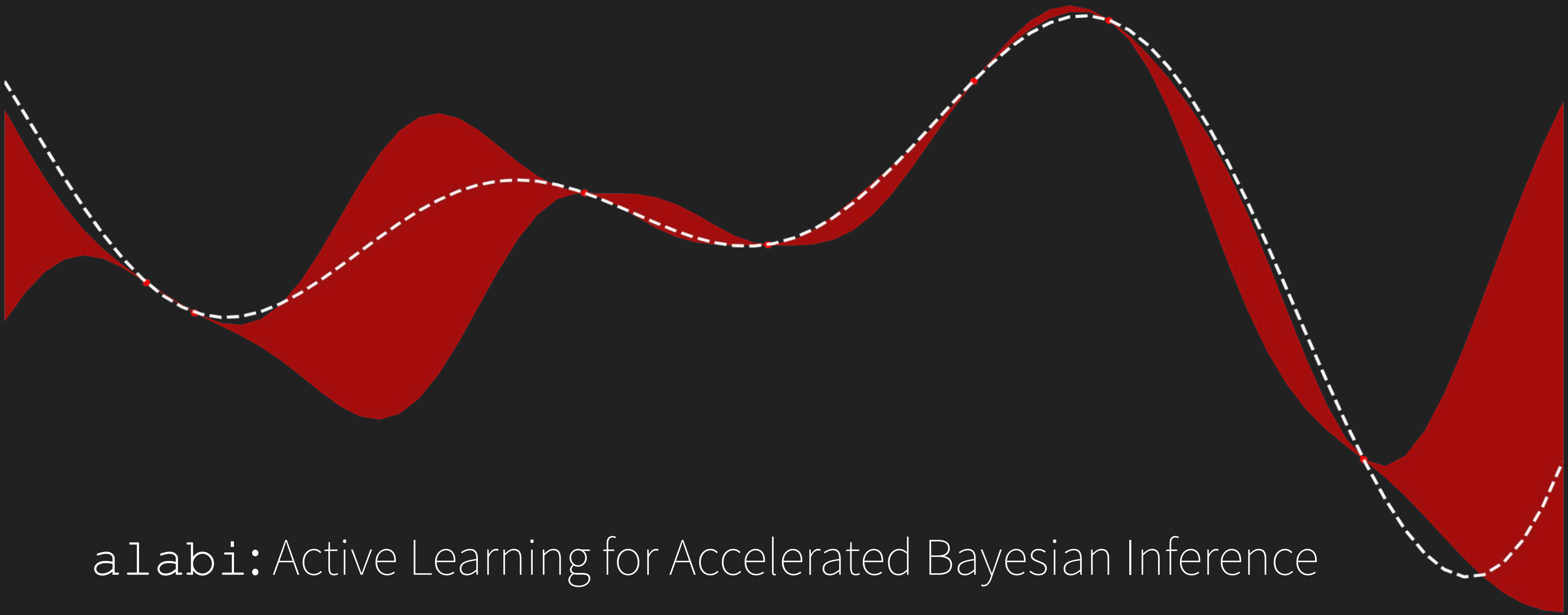
Jessica Birky

University of Washington

IAU Symposium 362



`alabi`: Active Learning for Accelerated Bayesian Inference

Model Inference: a Universal Problem

Astronomers want to make ***predictions*** about the Universe using ***models***

Theorists: Derive forward model f

$$Y = f(\theta)$$

Observers: Measure observations Y, infer parameters θ

$$\theta = f^{-1}(Y_{obs})$$

Model Inference: a Universal Problem

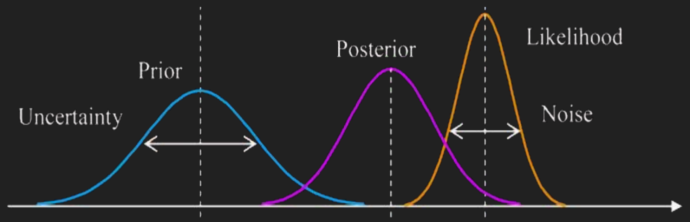
$$\theta = f^{-1}(Y_{obs})$$

Generally we'd like to be able to ***interpret θ***, and assess the confidence in our predictions from ***uncertainties of θ***

Bayesian inference provides a framework for this

Bayesian Inference: Overview

BAYES THEOREM



$$\ln P(\theta) \propto \ln Like(\theta) + \ln Prior(\theta)$$

POSTERIOR *LIKELIHOOD* *PRIOR*

Can use Markov Chain Monte Carlo (MCMC) to evaluate posterior samples

Bayesian Inference: Challenges

But performing MCMC requires evaluating your likelihood function *many times*

Drawing sufficient nwalkers per dimension & drawing independent samples means computing many samples, and throwing many away

If your likelihood function is expensive, this inefficiency adds up!

(A likelihood that takes ~10s of seconds per model to run can end up taking ~weeks-months to run an MCMC even on a cluster node)

ALABI: Active Learning for Accelerated Bayesian Inference

$$\ln P(\theta) \propto \ln Like(\theta) + \ln Prior(\theta) \approx g(\theta)$$

POSTERIOR *LIKELIHOOD* *PRIOR* *GAUSSIAN PROCESS SURROGATE MODEL*

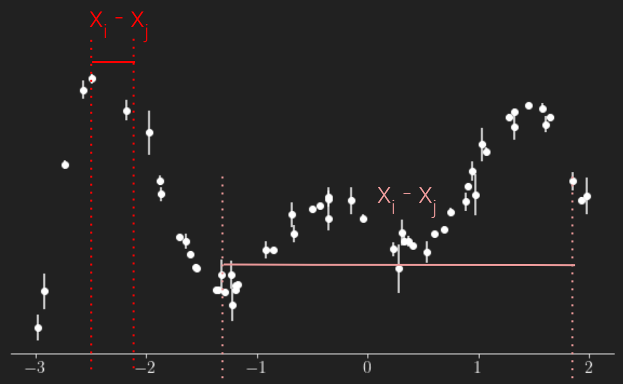
For models which are ***computationally expensive***, we can approximate the posterior using a ***surrogate model***

This can be done with two additional steps: ***Gaussian Process + Active Learning***

Methods: Gaussian Process Surrogate Model

GAUSSIAN PROCESS (GP)

Assumes that ***neighboring points are correlated*** according to a ***kernel function***

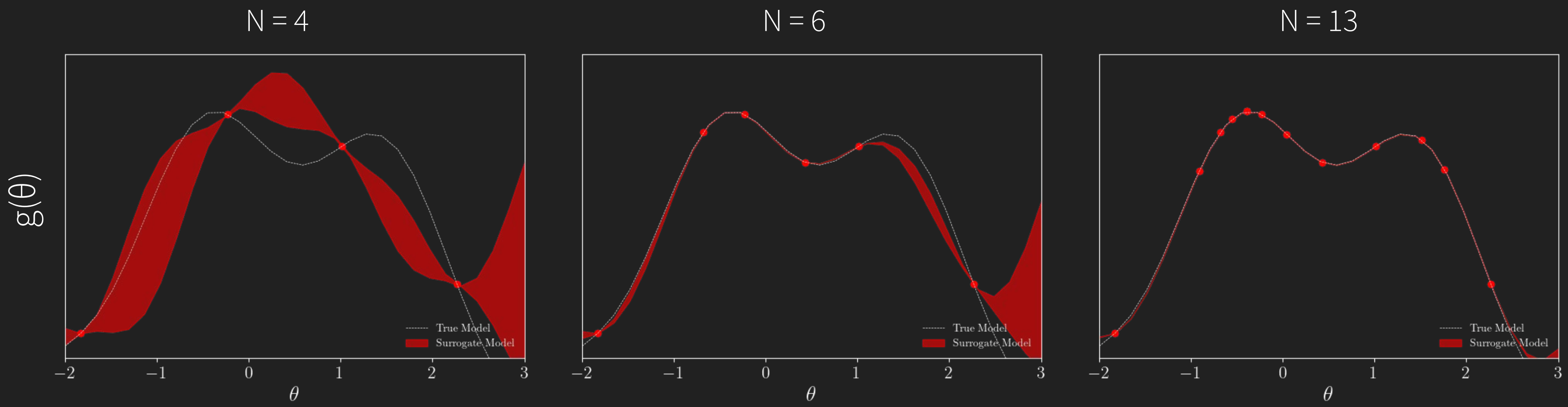


Highly correlated

Weakly correlated

This kernel would then allow us to define our regression model as a probability distribution over functions, which has a Gaussian ***mean and covariance***

Methods: Active Learning



GP is a regression model trained on N data points

More training data → less prediction uncertainty, better prediction accuracy

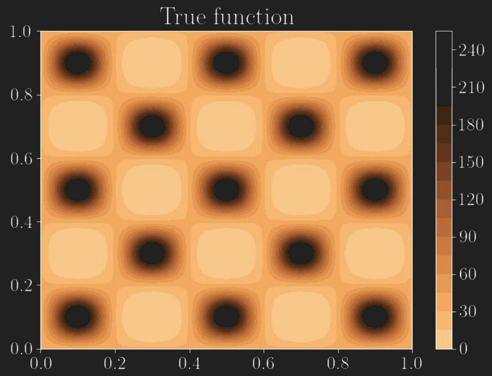
New points efficiently selected using active learning algorithm - iteratively optimize picking points with high probability and high uncertainty (Kandasamy et al. 2017)

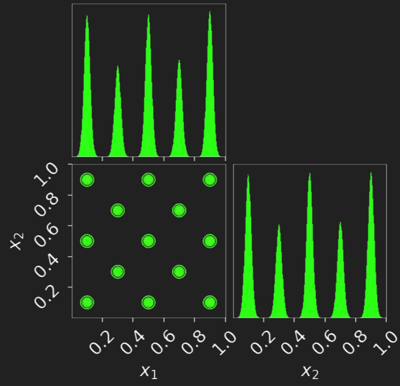
Methods: MCMC

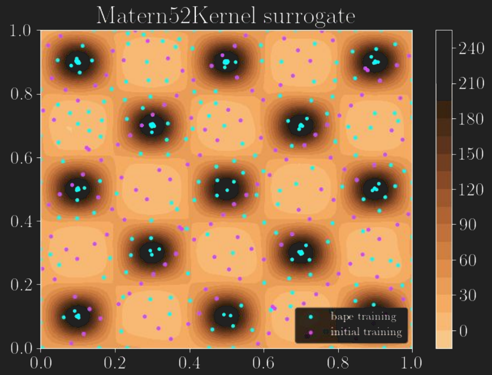
`alabi` easily interfaces with two different MCMC samplers:

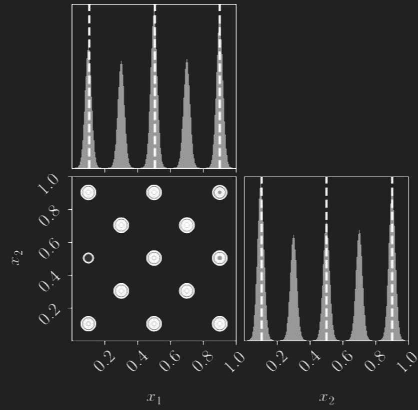
`emcee` (affine invariant sampler; Foreman-Mackey et al. 2013)

`dynesty` (nested sampler; Speagle 2020)









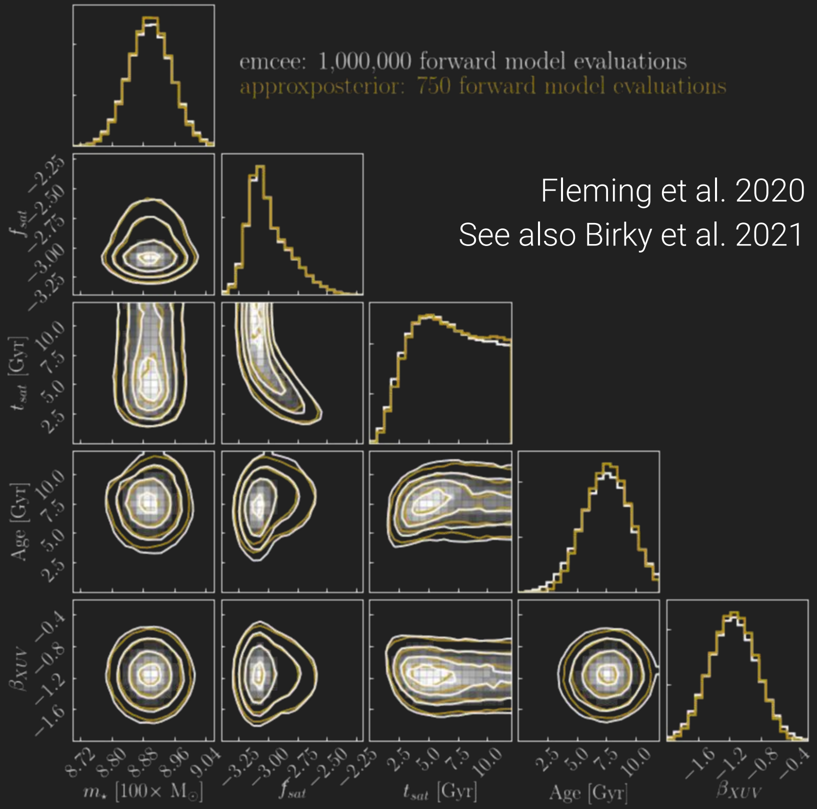
Applications

SCIENCE CASE:

TRAPPIST-1: 7 exoplanet hosting star

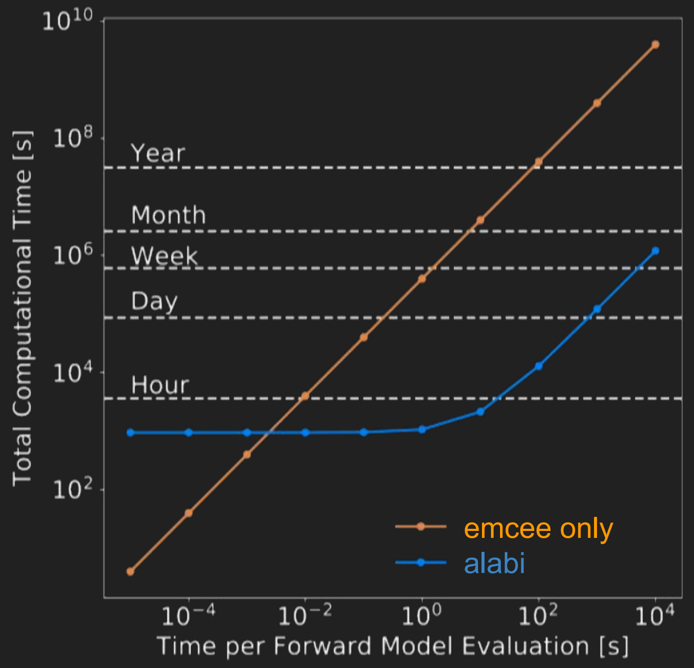
5 parameter model for modeling stellar evolution + XUV luminosity using `VPLanet` (Barnes et al. 2020)

~4000 core hours → ~4 core hours (weeks on cluster vs. hours on laptop)



Fleming et al. 2020
See also Birky et al. 2021

Performance



Fleming et al. 2018

GP surrogate model reduces MCMC computation time by orders of magnitude!

Conclusions

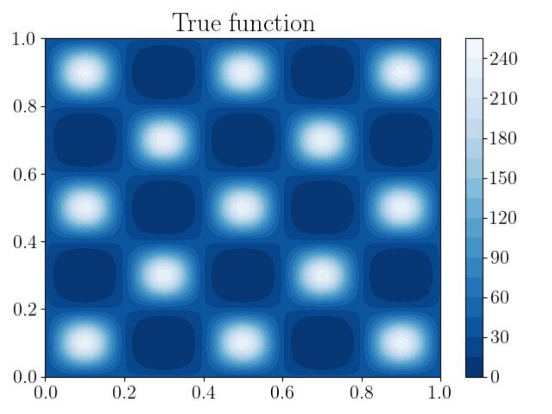
`alabi` presents a promising way forward for performing Bayesian Inference with computationally expensive models (demonstrated ~1000x faster)

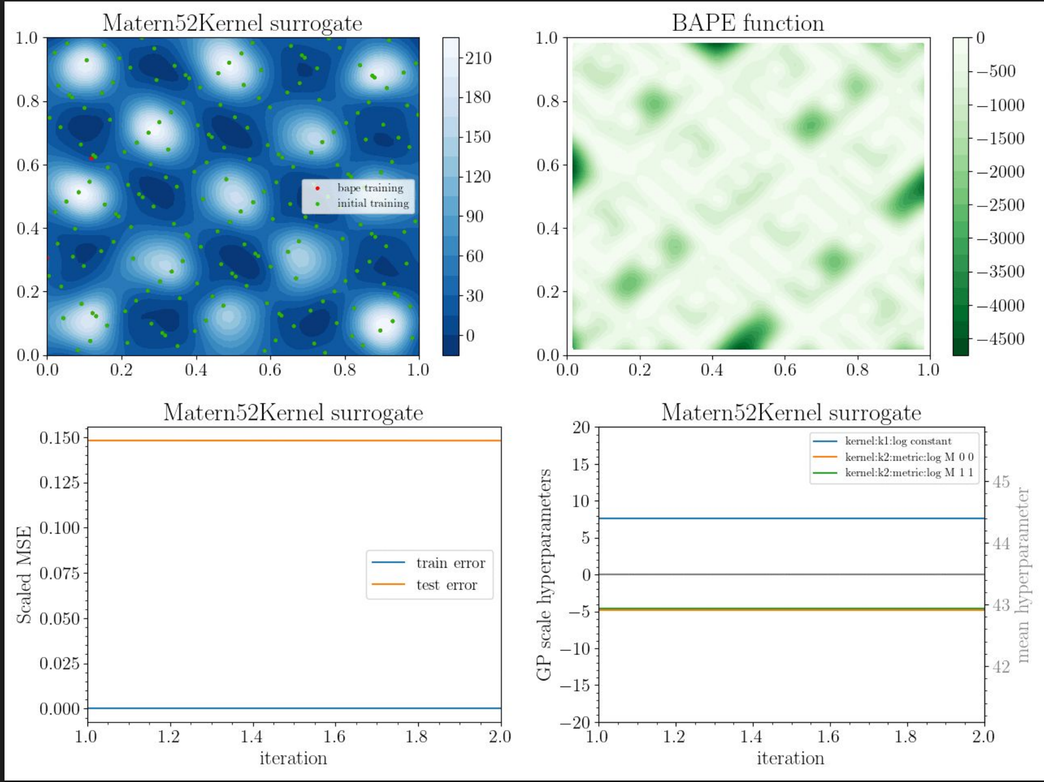
Future work: benchmark `alabi` on high dimensional problems, incorporate additional MCMC sampler compatibility (e.g. Hamiltonian MC)

`alabi` is an open source project and welcomes contributions from the community!

https://github.com/jbirky/alabi

Methods: Active Learning

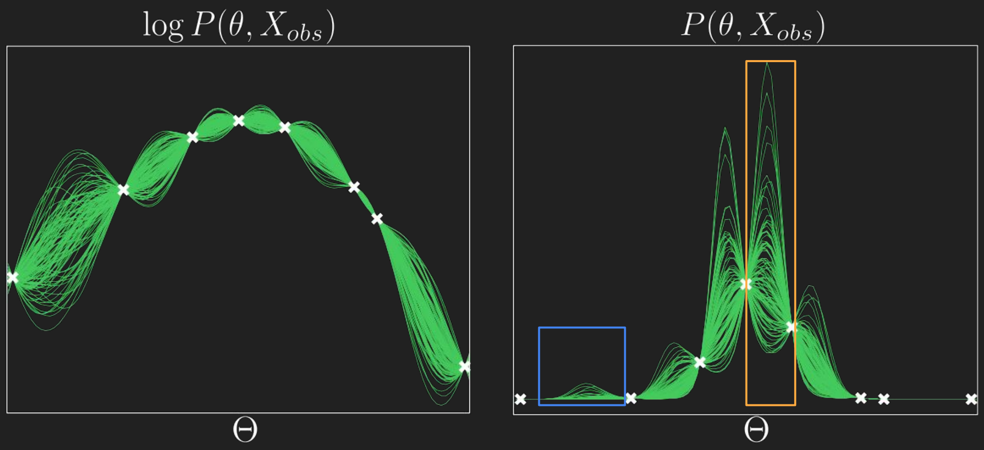




Methods: Active Learning

UTILITY FUNCTION

$$u_t^{\mathrm{EV}}(\theta_+) = \boxed{\exp(2\mu_t(\theta_+)} + \boxed{\sigma_t^2(\theta_+))(\exp(\sigma_t^2(\theta_+))} - 1)$$



Kandasamy et al. 2017

Favors selecting θ with ***high probability***

Favors selecting θ with ***high uncertainty***

Active learning optimizes $u_t(\theta)$ for niter