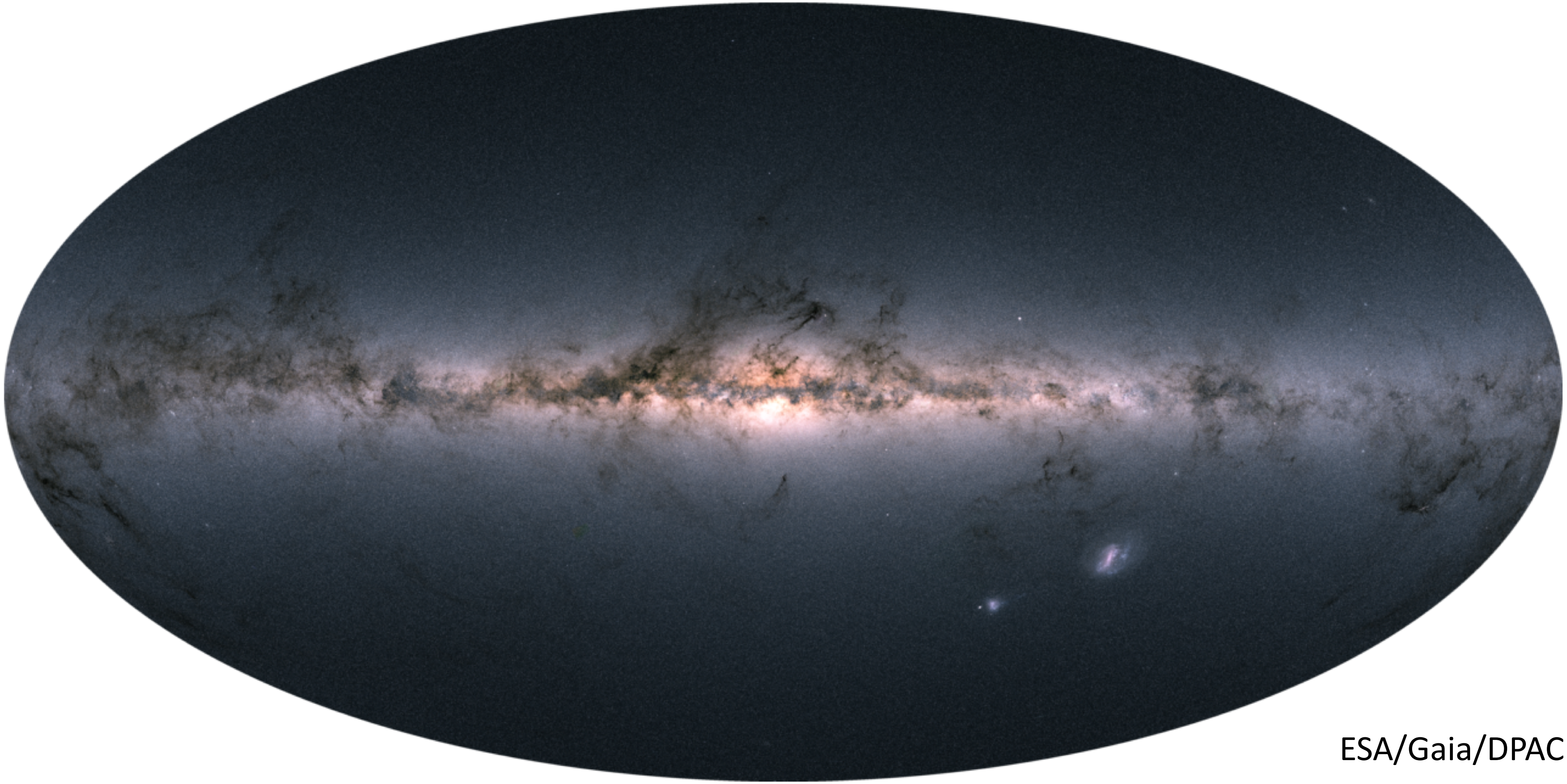
Modeling the 3D Dust Map of the Milky Way Galaxy as a Gaussian Process

Lauren Anderson

Flatiron Institute NYC

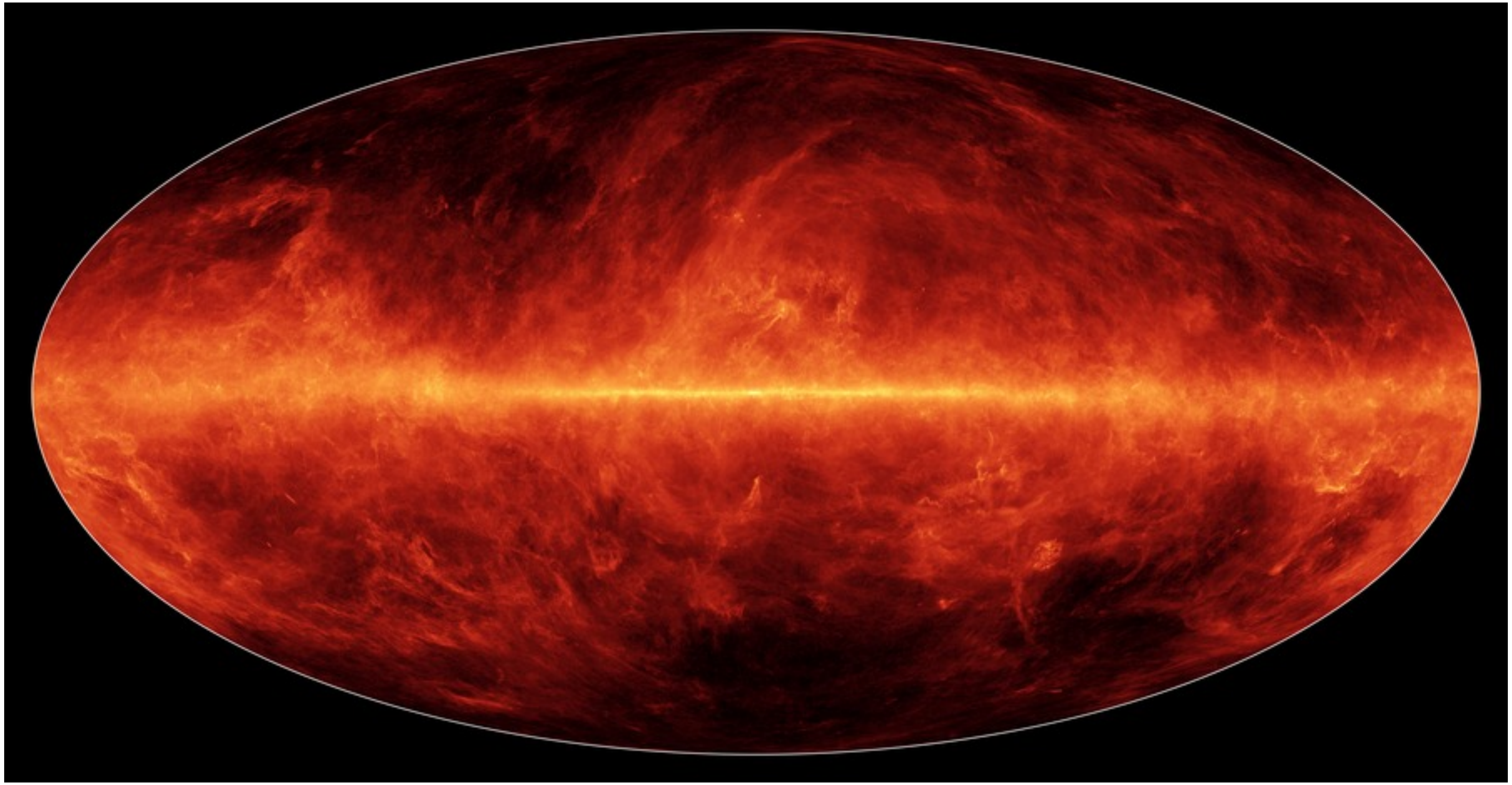
Andrew Miller (Columbia), Boris Leistedt (NYU), David Blei (Columbia), David Hogg (NYU)

Astrostatistics in the Era of Large Surveys



ESA/Gaia/DPAC

Dust lives mostly in the plane of the galaxy



ESA/NASA/JPL-Caltech

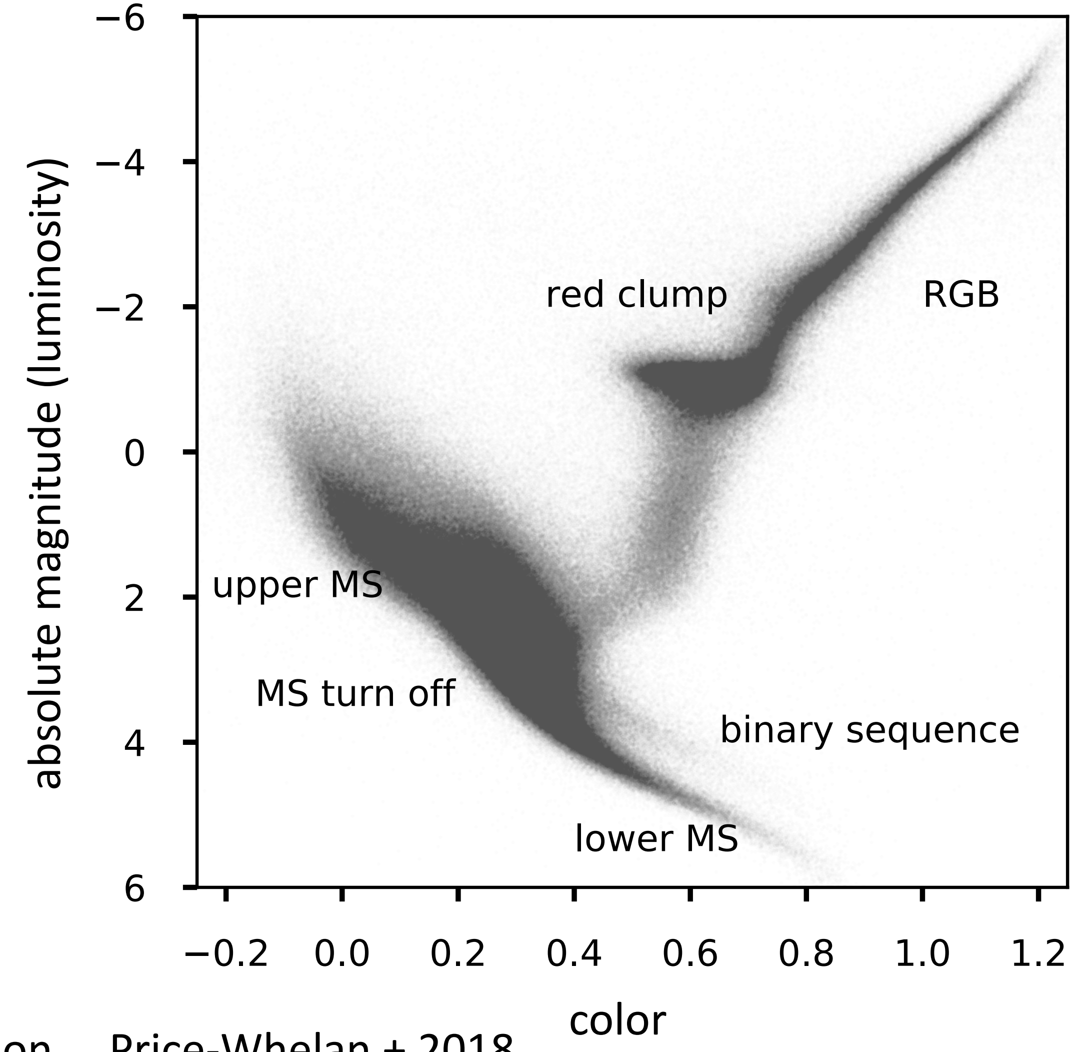
We can get a better view of dust distribution in other galaxies

NASA, ESA, S. Beckwith (STScI), and
The Hubble Heritage Team (STScI/AURA)

We infer the integrated density of dust along a line of sight by how much star light it absorbs

Loke Kun Tan
(StarryScapes)

Motivation for Project: a data-driven prior of distances to stars

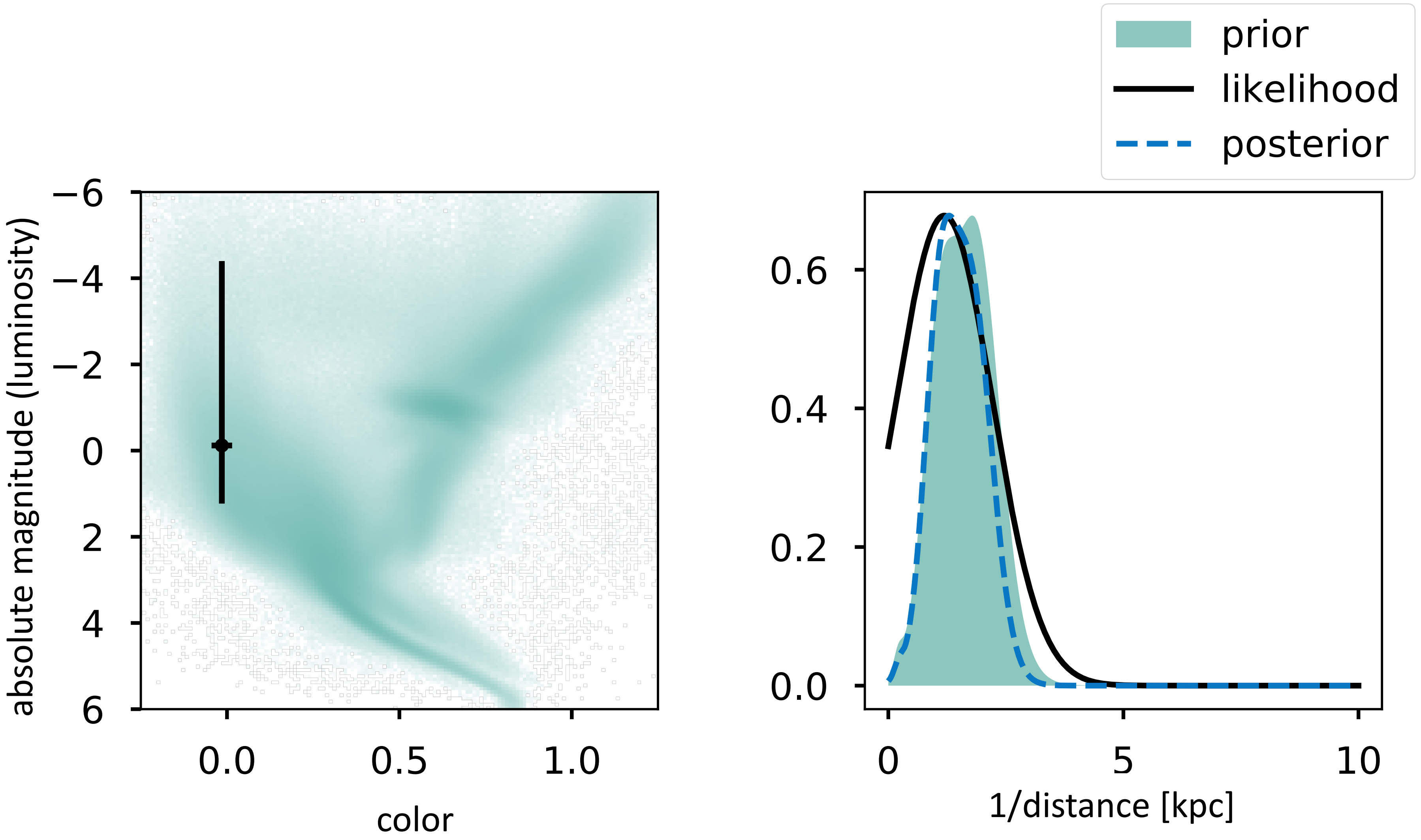


Anderson,… Price-Whelan + 2018

The color-magnitude diagram (CMD)

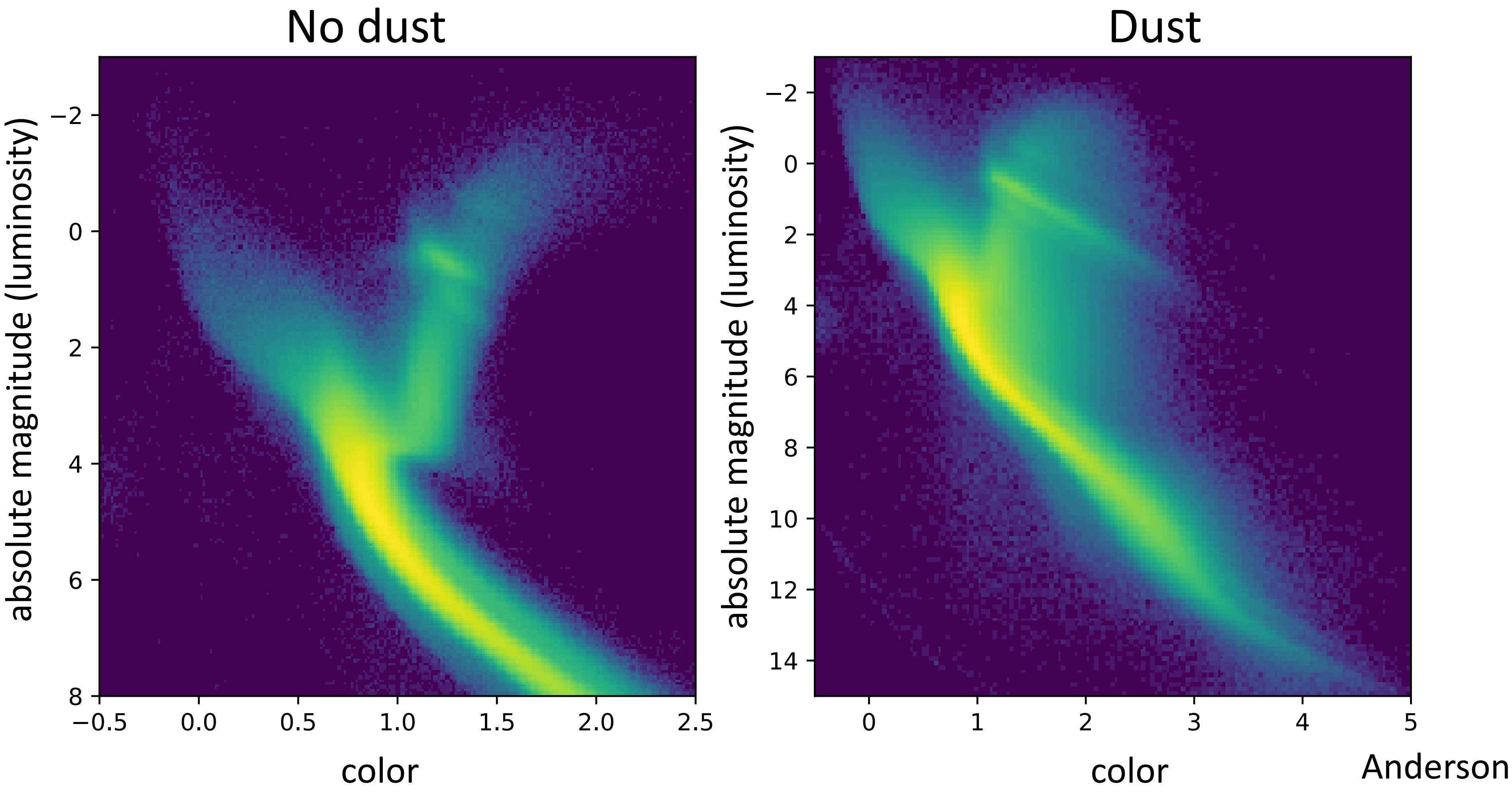
A space where star properties are highly correlated including information about distance

Motivation for Project: a data-driven prior of distances to stars



Anderson,… Price-Whelan + 2018

Motivation for Project: a data-driven prior of distances to stars



Anderson+ 2019 in prep

Astrophysical Motivation

Learn about dynamical structures like spiral arms and star forming regions

Learn how much attenuation lies along each line of site

Modeling Challenges of using a Gaussian Process

Measurements are integrals of the dust density field

We have 100s of millions of measurements

GPs scale as N^3 in computation and N^2 in memory

The distance estimates are also noisy

Gaussian Process toy example

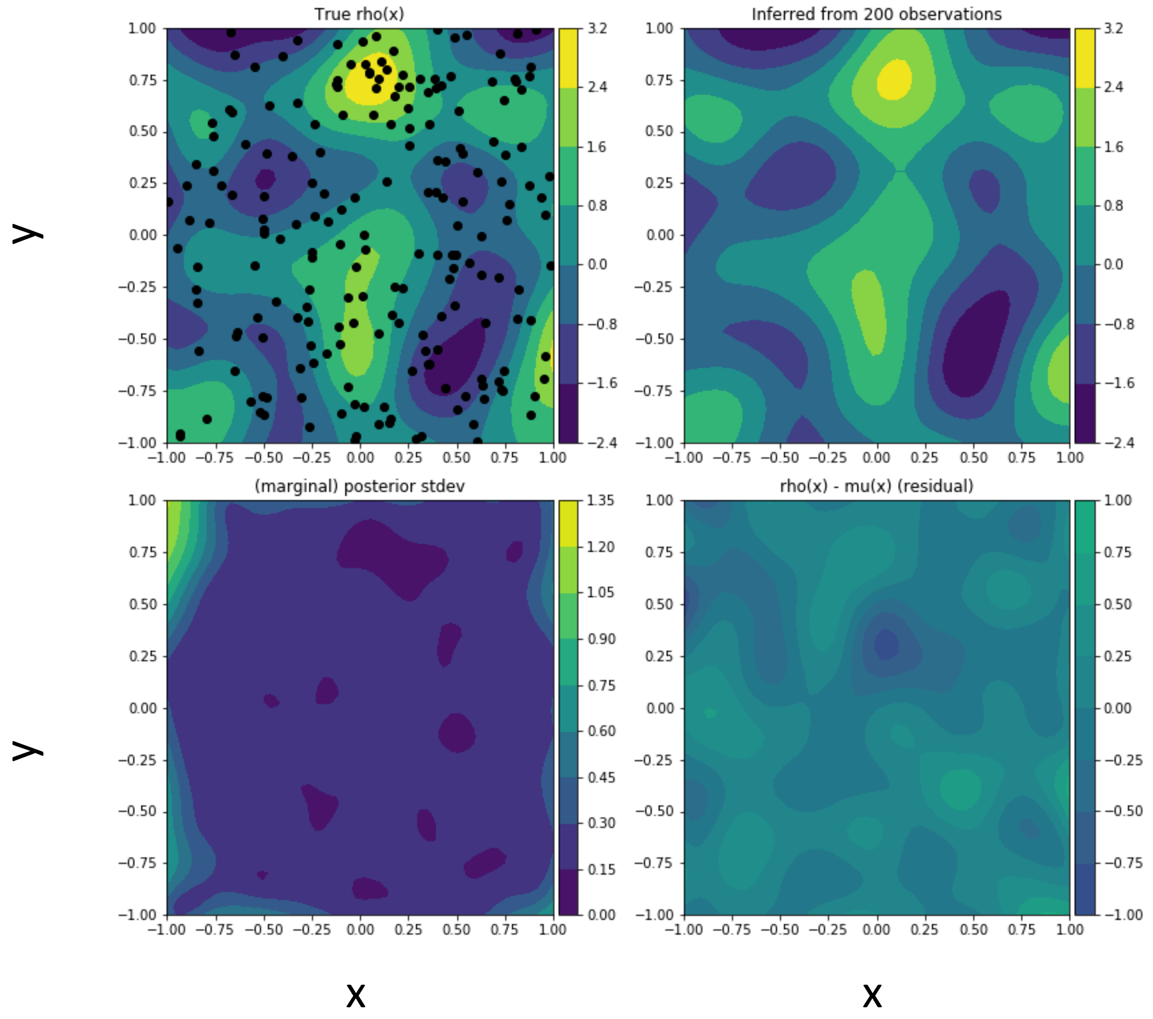
$$\boldsymbol{\rho} \sim \mathcal{N}(0, \mathbf{K_x})$$

$$(\mathbf{K_x})_{ij} = k_{\boldsymbol{\theta}}(x_i, x_j)$$

$$k_{sqe}(x_i, x_j) = \sigma^2 \exp\left(-\frac{1}{2}(x_i - x_j)^{\intercal} L^{-1}(x_i - x_j)\right)$$

$$y_i \,|\, x_i \sim \rho(x_i) + \sigma_i \cdot \epsilon$$

$$\begin{pmatrix} \mathbf{y} \\ \rho_* \end{pmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \mathbf{K_x} + \Sigma_\mathbf{y} & \mathbf{K_{x,*}} \\ \mathbf{K_{*,a}} & \mathbf{K_*} \end{pmatrix}\right)$$



Modeling Challenges of using a Gaussian Process

Measurements are integrals of the dust density field

We have 100s of millions of measurements

GPs scale as N^3 in computation and N^2 in memory

The distance estimates are also noisy

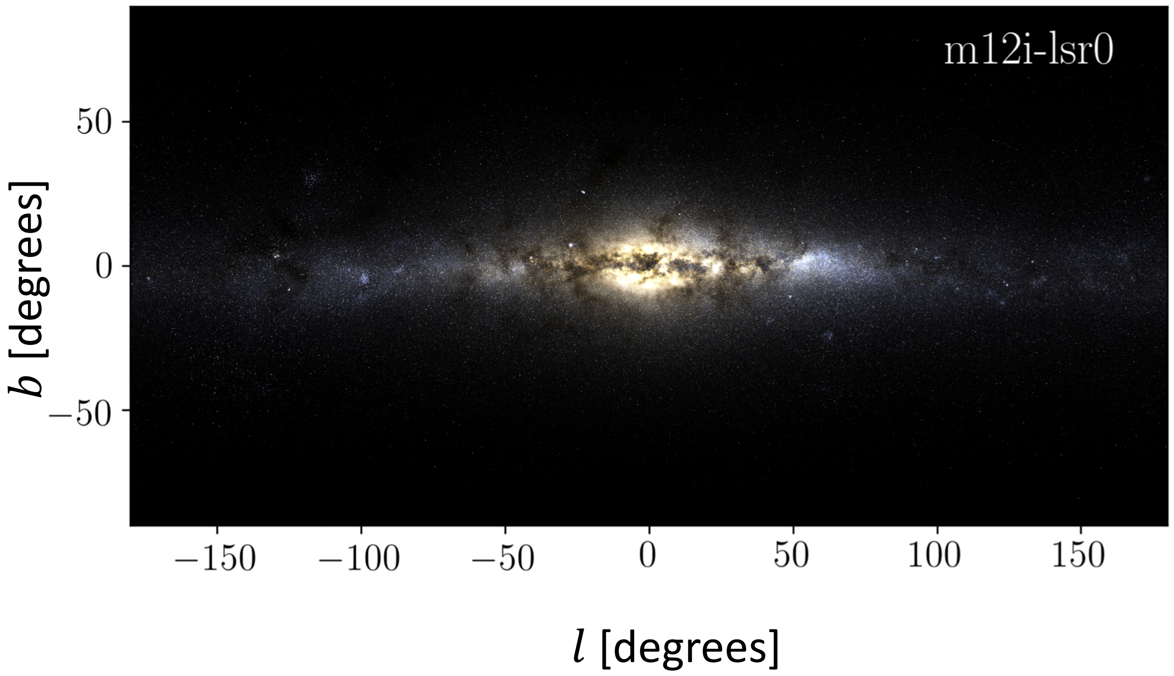
Big Data: 100s of Millions of Measurements

- Start with section of data very close to our sun and use inducing point methods (Snelson et al 2007)
  - choose $m < n$ positions to "summarize your data"
  - Computation scales as $O(nm^2)$ in computation and as $O(nm)$ in memory
  - For $n \sim 10^8$ this is still prohibitive

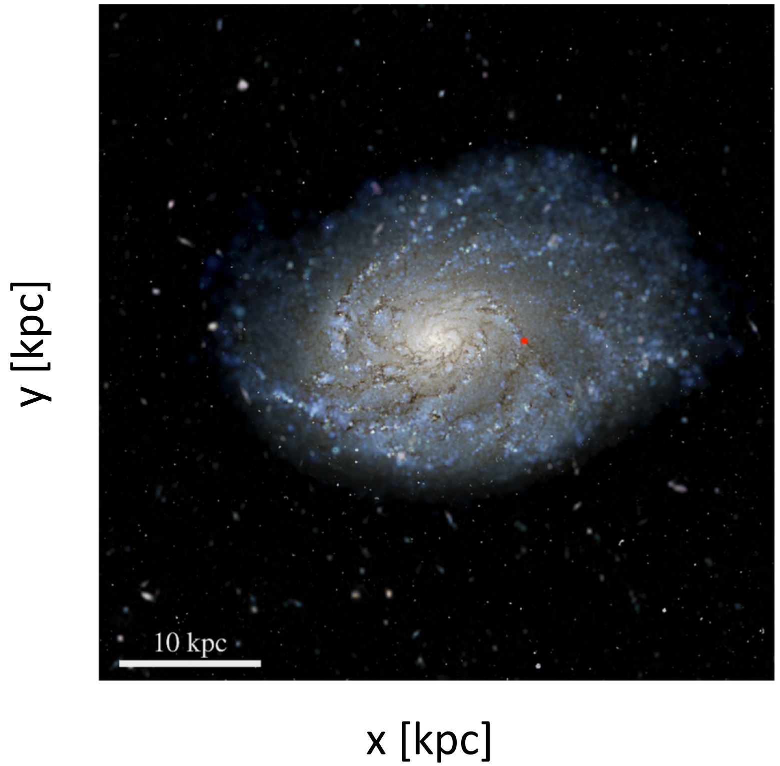
- Break up the sky, correlation length is very small relative to the galaxy size
- Measurements are highly correlated due to passing through local neighborhood of dust, try to model very precisely then remove from data

Domain Simulation: synthetically observed galaxy simulation

$\{d_i, a_i, \theta_i, \sigma_{d,i}, \sigma_{a,i}\}$



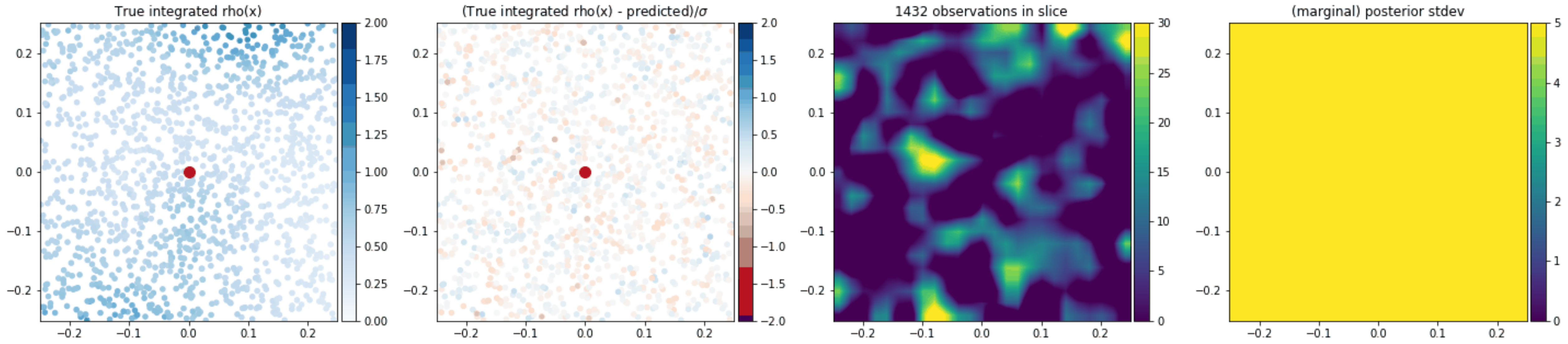
$\rho(\boldsymbol{X})$



Note: not active learning, in fact we are hoping to use the data to inform the simulations/models in a very apples to apples comparison.

Sanderson et al 2018

500 pc cube around the synthetic sun



1e5 observations 25^3 inducing points   ℓ ~ 30 pc   τ ~ 100

Modeling Challenges of using a Gaussian Process

Measurements are integrals of the dust density field

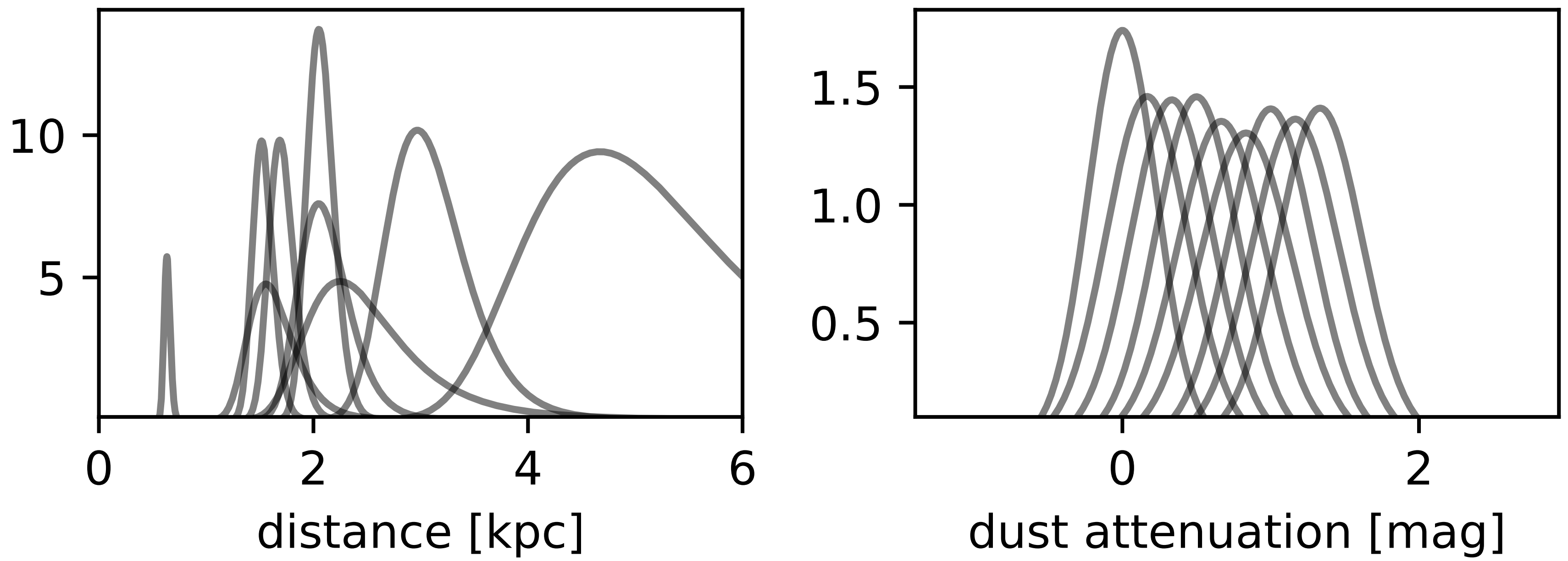
We have 100s of millions of measurements

GPs scale as N^3 in computation and N^2 in memory

The distance estimates are also noisy

Distance estimates are (very) noisy

Leistedt et al. in prep



Distances in astronomy are notoriously difficult to measure

Punting for now but would love to chat with people about approaches

Summary

3D dust map in the Milky Way using a Gaussian Process

Observations are integrals of the field

Big Data

Noisy observations of positions

Astro Hack Week 2019
part summer school, part unconference

http://astrohackweek.org

6th year, will be at Kavli Institute for Cosmology at Cambridge University, UK

Participants will learn the theoretical foundation of, and practical knowledge in, statistical and machine learning methods crucial to modern astronomical data analysis. We welcome participants to bring their own research projects with them; Astro Hack Week is a great place to apply new skills and methods, and work with others to move these projects forward. It is also an opportunity to work on something new, and we encourage participants to apply their technical knowledge to outstanding problems in the astronomical community