Gaussian processes on galaxy

Finding Galaxies in the Shadows of Quasars with Gaussian Processes

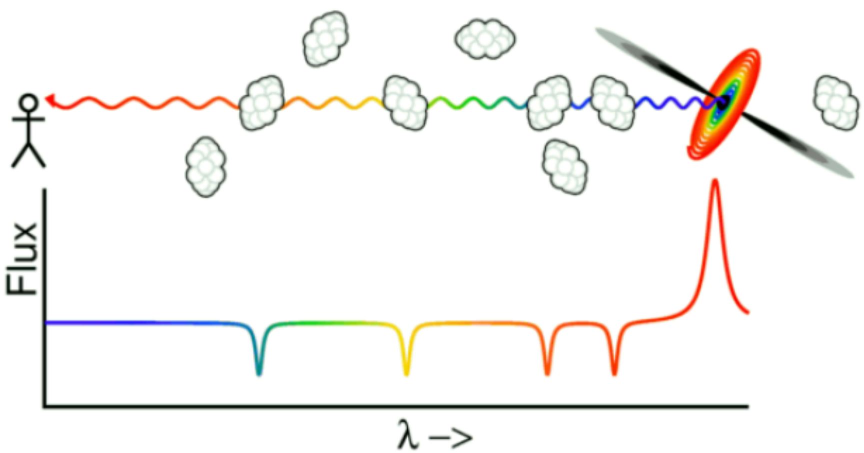
Roman Garnett GARNETT@WUSTL.EDU
Washington University in St. Louis, St. Louis, MO 63130, United States

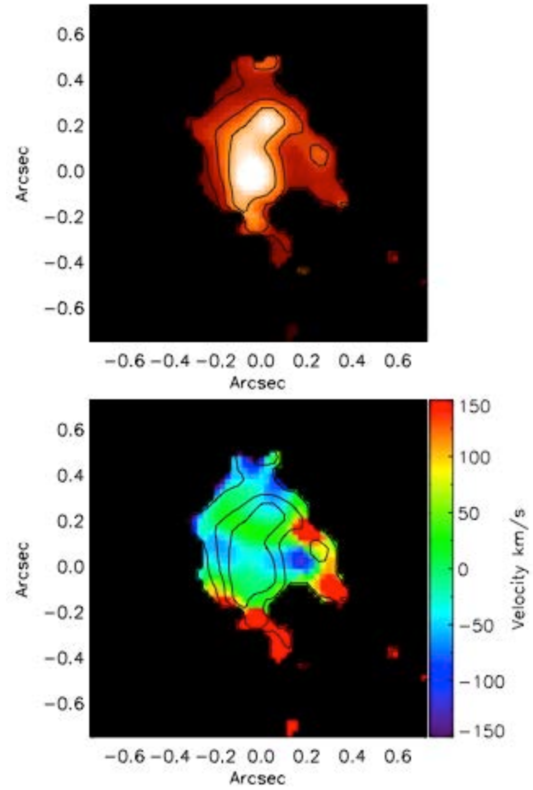
Shirley Ho SHIRLEYH@ANDREW.CMU.EDU
Jeff Schneider JEFF.SCHNEIDER@CS.CMU.EDU
Carnegie Mellon University, Pittsburgh, PA 15213, United States

(ICML 2015)

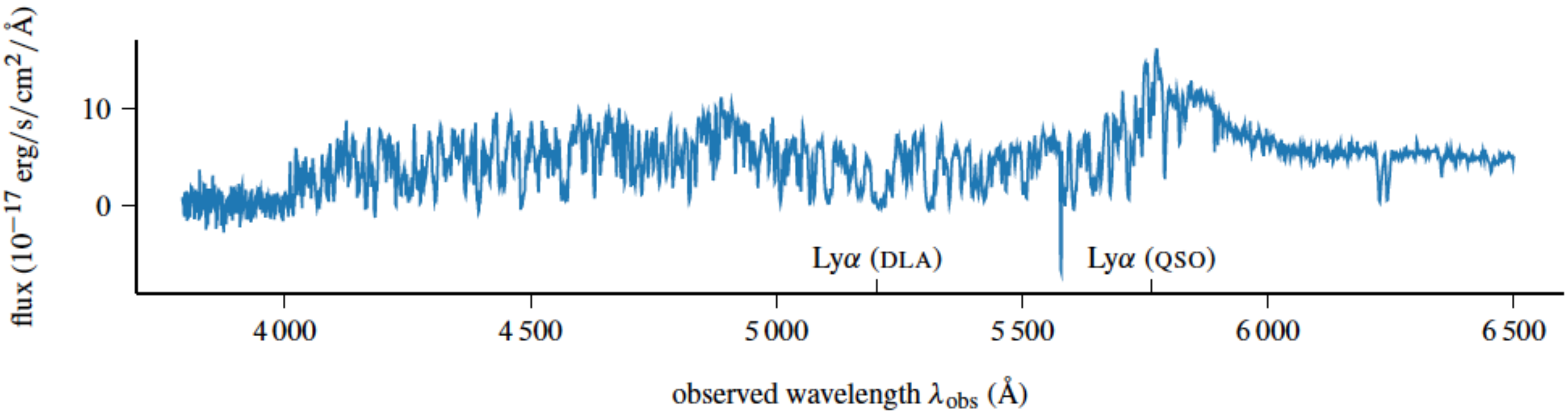
Gaussian processes on galaxy (2)

- Astronomers want to find DLA (Damped Lyman-α systems):
  - Large gaseous objects with neutral hydrogen gas
  - Emits little light and cannot be observed directly.



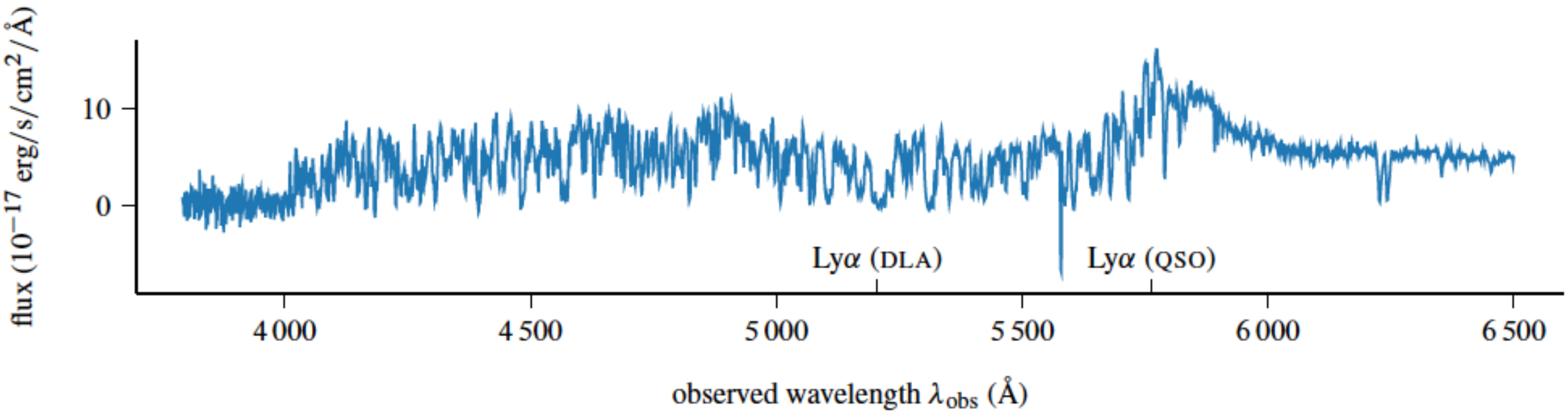


Finding DLAs



- DLAs can be found by quaser emission spectrum
- Usually: by astronomers looking at it

Finding DLAs



- Problem: Quaser spectrums are huge in number!
  - Sloan digital survey: 300 000
  - Millions of quasers observed
- How to automate discoveries?

Finding DLAs

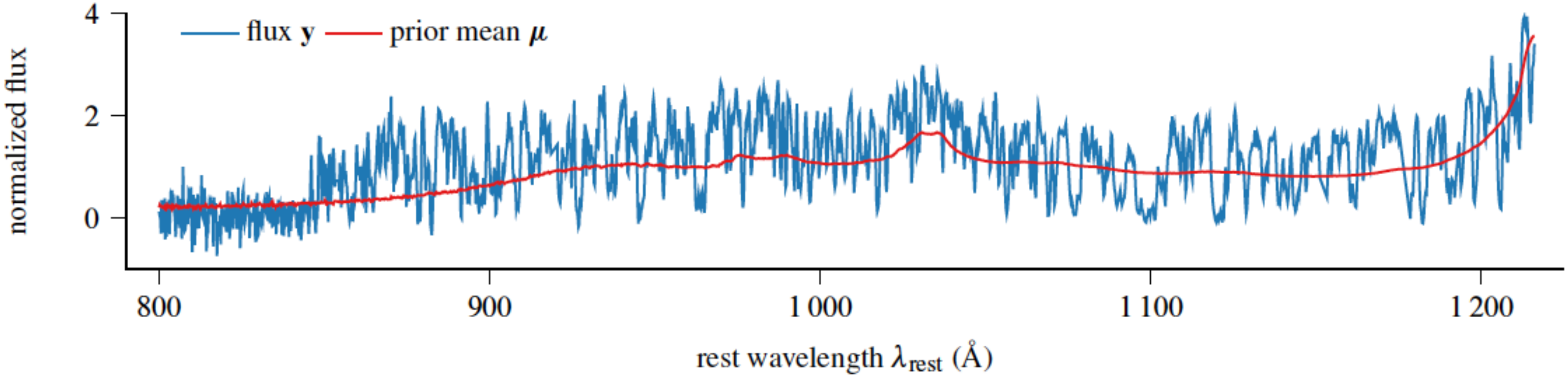
- Solution: compare probabilities of

$$p(\text{spectrum} \mid \text{DLA exists})$$

$$p(\text{spectrum} \mid \text{DLA does not exist})$$

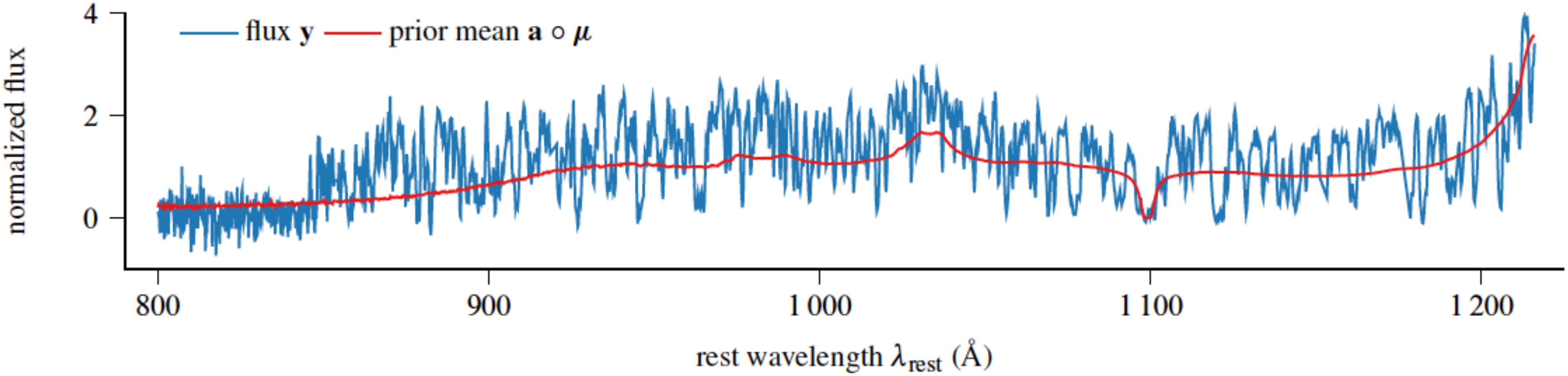
- How to define these non-trivial probabilities?

Case of no DLAs



- Gaussian process + noise
- log p(spectrum|¬DLA) = -2589.

Case of DLAs



- Gaussian process + unknown absorption
- log p(spectrum|DLA) = -2453 > -2589
  = p(spectrum|¬DLA)
- DLA exists!

Technically..

- If DLA does not exist

$$p(\mathbf{y}|\Theta, \neg\mathrm{DLA}) = \mathrm{N}(\mathbf{y}|\boldsymbol{\mu}, K+\Omega+N)$$

- If DLA exist

$$y(\lambda) = f(\lambda)e^{-\tau(z,N)} + \epsilon$$

Absorption frequency dependent!

$$\downarrow$$

$$p(\mathbf{y}|\Theta, \mathrm{DLA}, z, N) = \mathrm{N}(\mathbf{y}|\boldsymbol{a} \circ \boldsymbol{\mu}, A(K+\Omega)A+N)$$

Technically..

- If DLA does not exist

$$p(\mathbf{y}|\Theta, \neg\mathrm{DLA}) = \mathrm{N}(\mathbf{y}|\boldsymbol{\mu}, K+\Omega+N)$$

- If DLA exist

Unknown frequencies

$$\begin{aligned} p(\mathbf{y}|\Theta, \mathrm{DLA}) &= \int p(\mathbf{y}, z, N|\Theta, \mathrm{DLA}) dz dN \\ &= \int p(\mathbf{y}|\Theta, \mathrm{DLA}, z, N) p(z, N) dz dN \\ &= (\text{numerical integration}). \end{aligned}$$

– Requires understanding of Gaussian process machinery and statistics!