

The Limitations of AGN Reverberation Mapping:

Bayesian Analysis Shows as Few
as 10% of MgII Sources Clearly Reverberate

Illuminating the Active Universe, ESO Garching Sep 15 2026



About Me (Hugh McDougall)

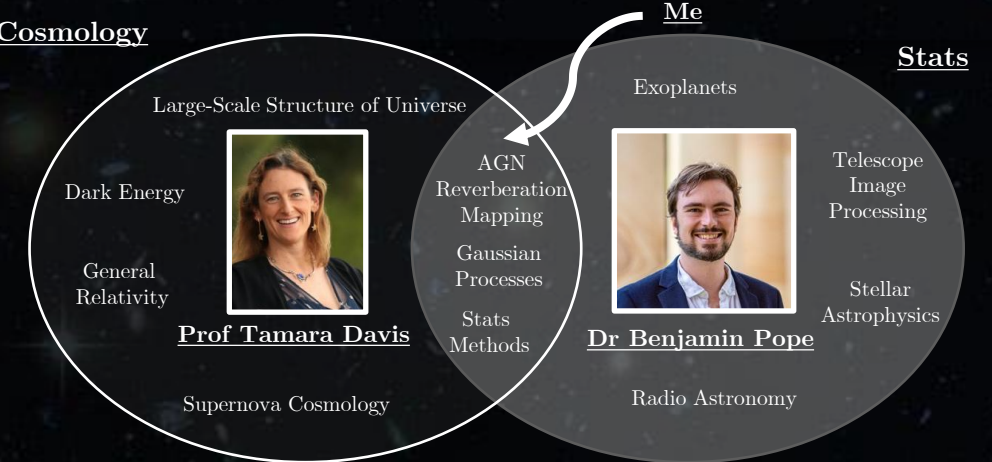


Final Year PhD Student (71 Days left)

- **Astro-Statistics**
- **Bayesian Stats Methods:**
 - Gaussian Processes
 - **JAX + NumPyro**
- **Cosmology**
 - Active Galactic Nuclei
 - Gravitational Waves (Sometimes)

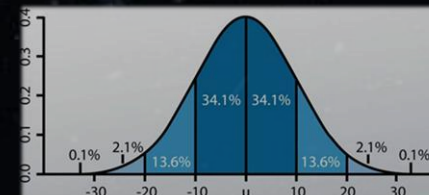


Cosmology



My Science

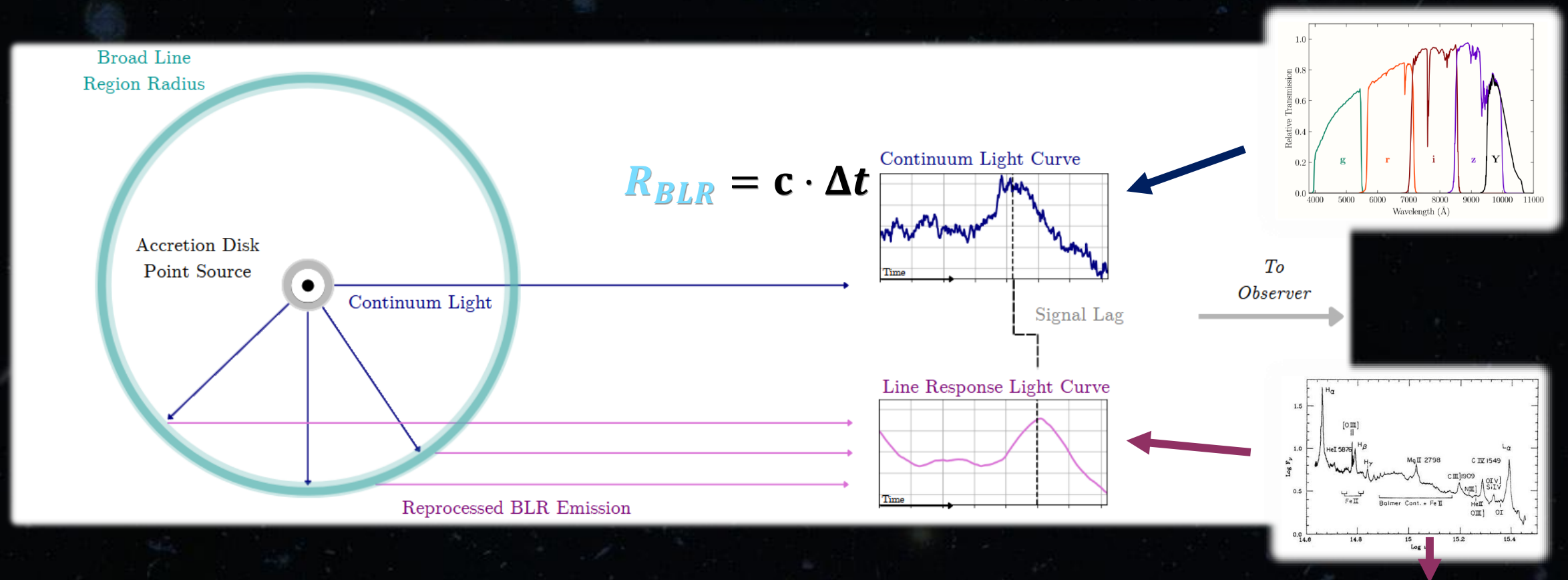
Reverberation mapping of supermassive black holes (more to come)



My Methods

Finding ways to make difficult Bayesian models tractable

Reverberation Mapping



Accretion disk
flickers

Flickers “echo” in
Broad Line Region

Light-Travel lag is
ruler for system size

Radius +
Doppler Broadening
gives us mass

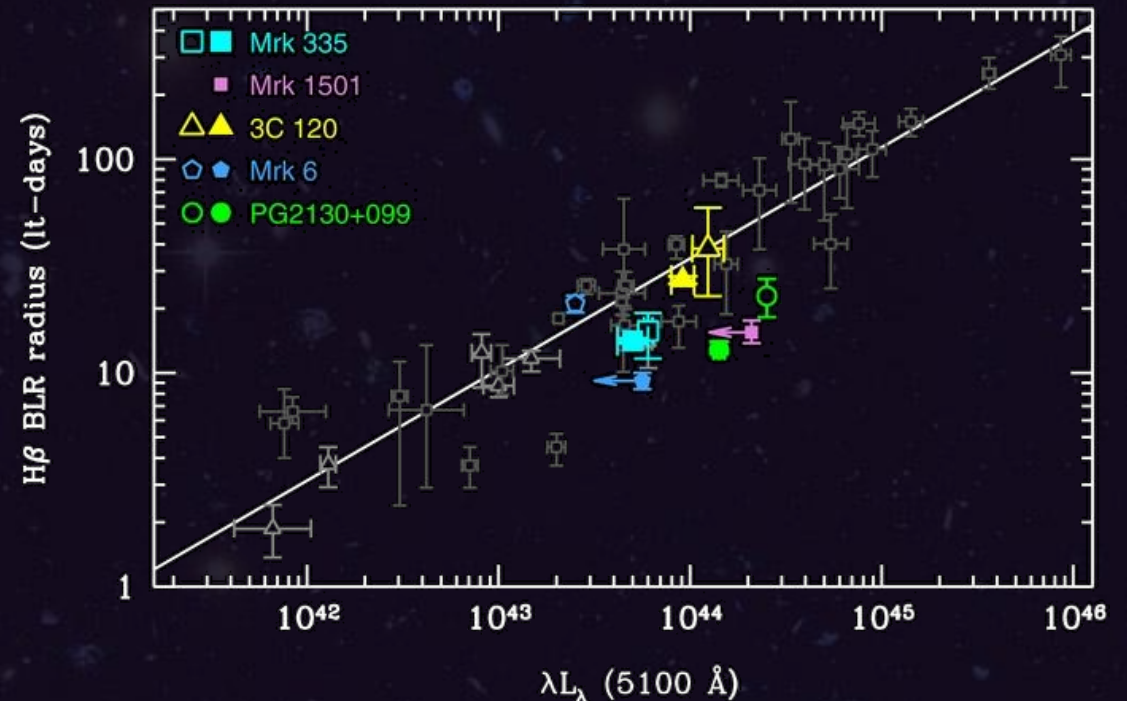
Reverberation Mapping – Old Methods

For short well-sampled light curves we can interpolate & Cross-Correlate Light Curves



Works for “traditional” RM surveys with good SNR, short lags, good coverage.

Brighter AGN are Bigger:
A Radius-Luminosity Relationship



**Provides a single-epoch
Radius + Mass Estimator**

Modern RM Surveys – Industrial Scale

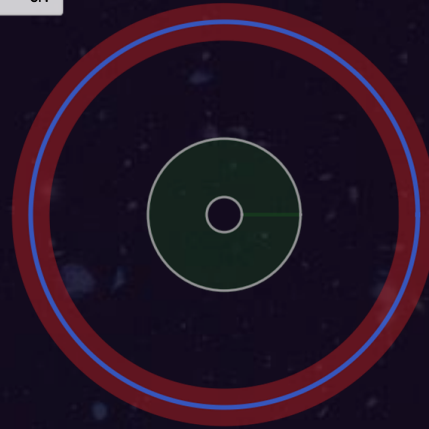
Established / Past



Upcoming

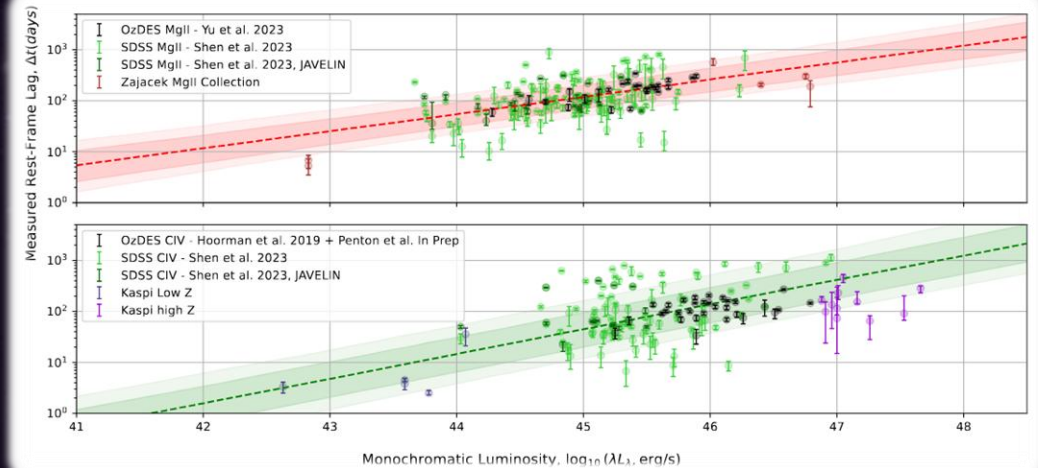


— $H\beta$
— MgII
— CIV



*Different R-L Relations
for Each Line*

*Different Emission Lines May
Emit From Different Radii*



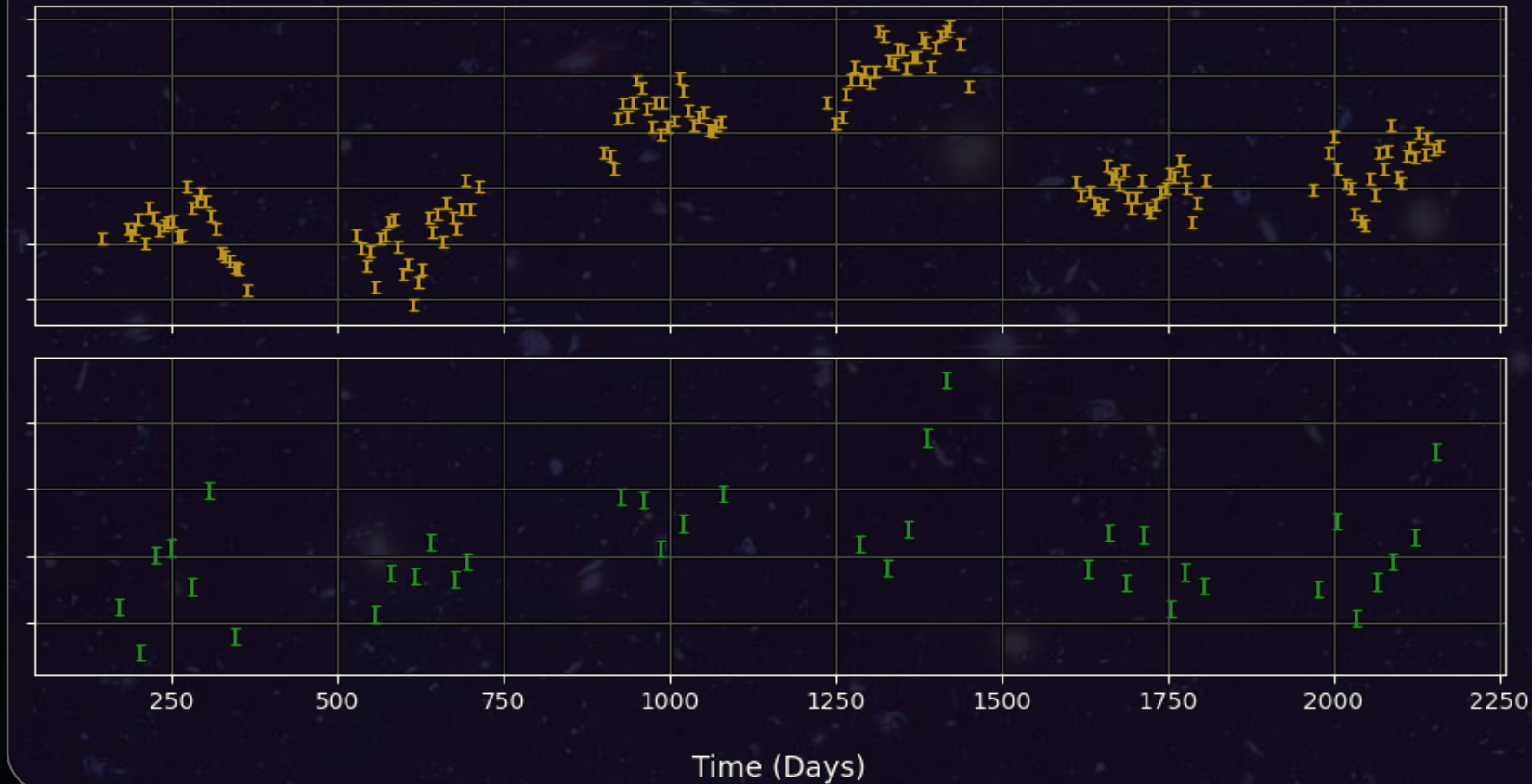
Examine AGN
populations and
changes over time

Measure Masses At
High-z

See if R-L relation
looks different at
High-z

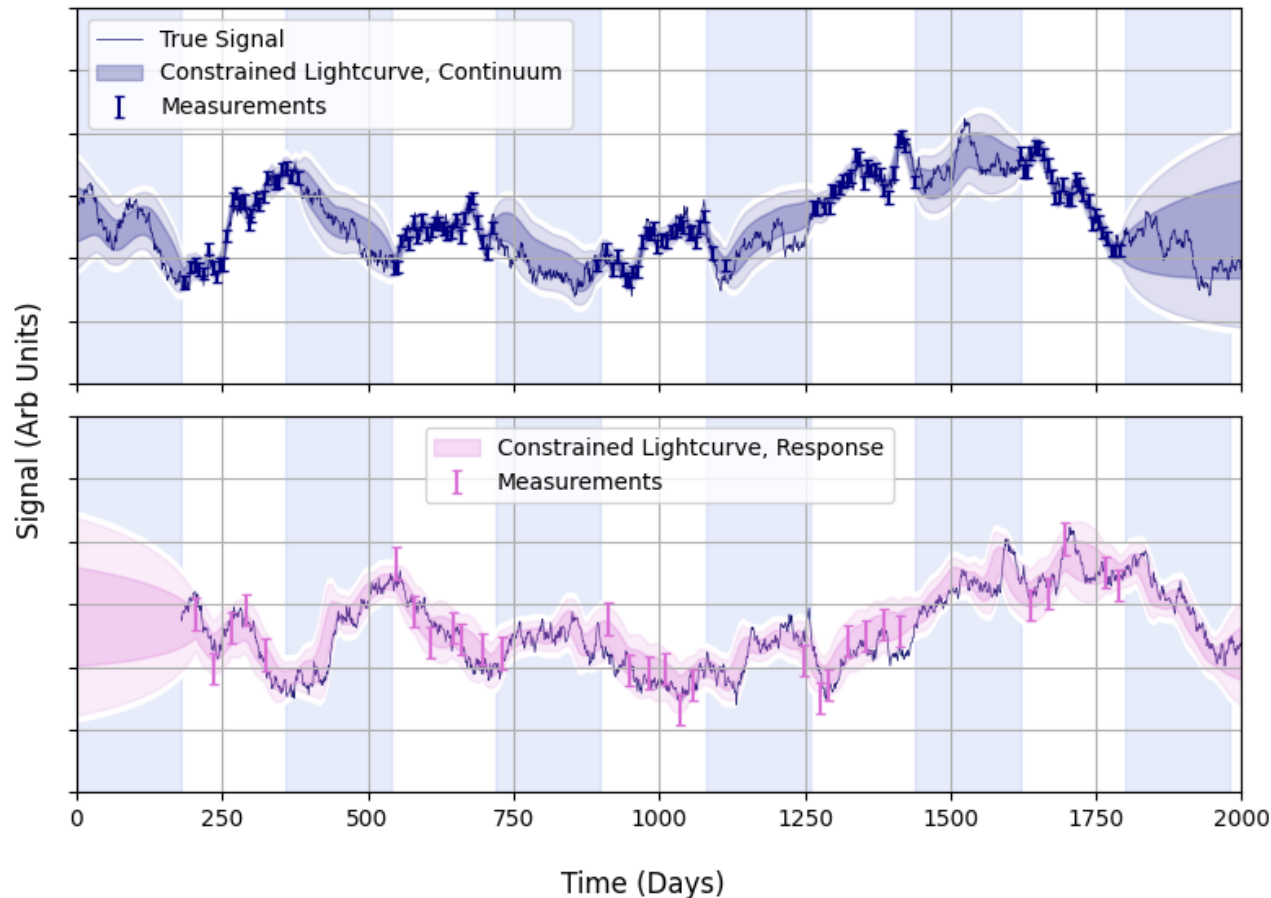
“Industrial Scale” RM, ≈ 1000 sources over wide redshift range.
“Statistically significant” sample, Multi-year surveys.

Modern RM Surveys – Seasonal Gaps



Seasonal gaps: Can only see AGN for $\frac{1}{2}$ of each year

Recovering Lags with Gaussian Process Models

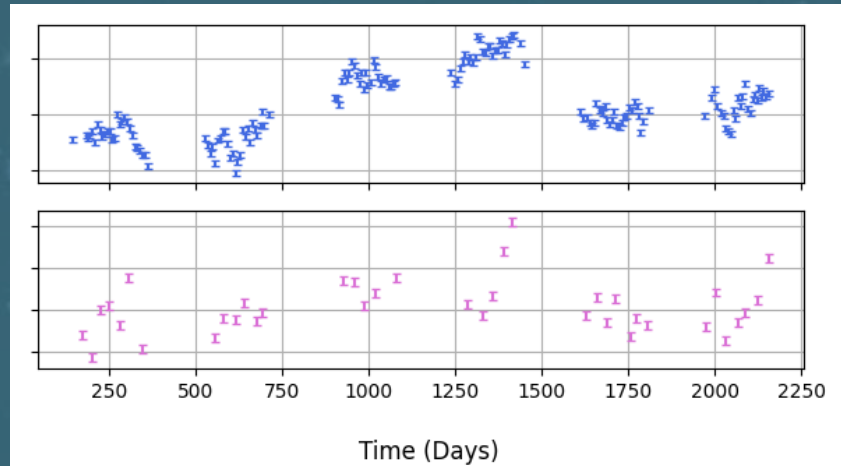


Stochastic Structure of AGN allows us to forward model them:

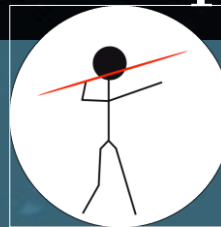
Light curves can be described as a
Damped Random Walk
(Or Other Gaussian Process)

Basis for **Bayesian Fitters**
(e.g. JAVELIN)

Modern RM Surveys: The Aliasing Problem

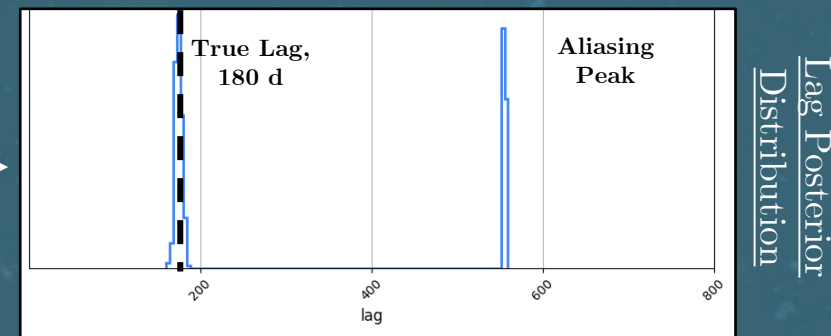


Multi-Year Surveys $\Rightarrow$
Seasonal Gaps in Data



JAVELIN

Fits with MCMC



Multiple Peaks in Lag Posterior
Distribution, “Aliasing”

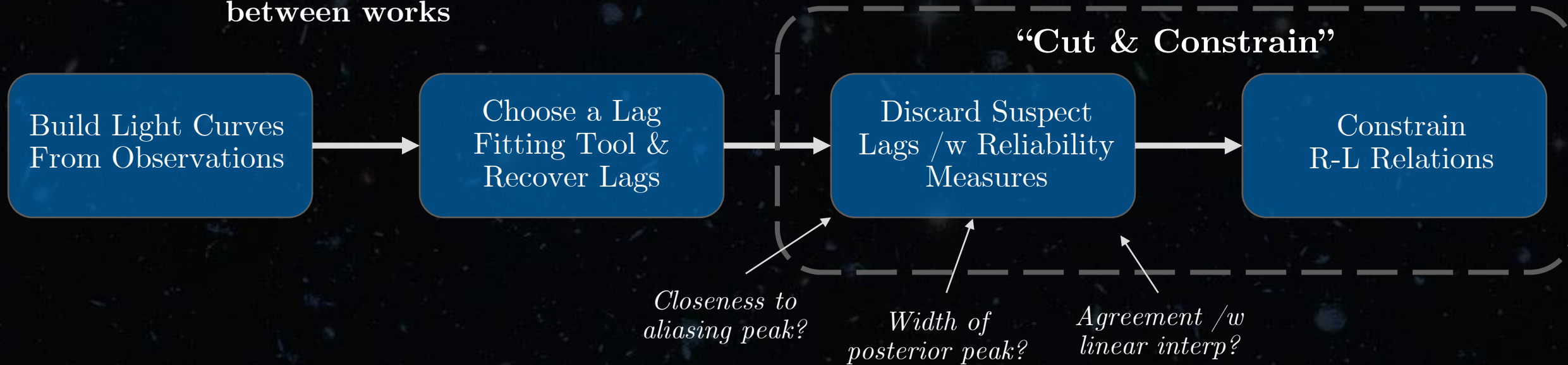
Multi-Year Surveys $\Rightarrow$ False Positive Lags

Existing Method: “Cut and Constrain”



Vastly different approaches
between works

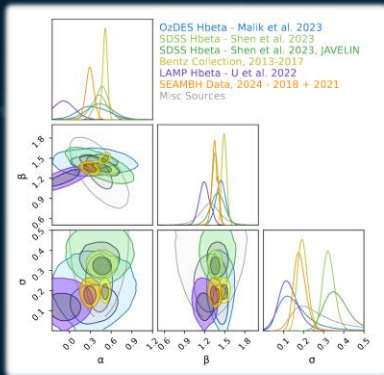
Patchwork mix of tools and significance
measures from disparate tools and statistical
frameworks leads to **different results**



Different RM Surveys are in Tension: Why?

H β

Differences in Targets



*Different types of AGN in
different surveys
(Old News)*

MgII

Differences in Methods

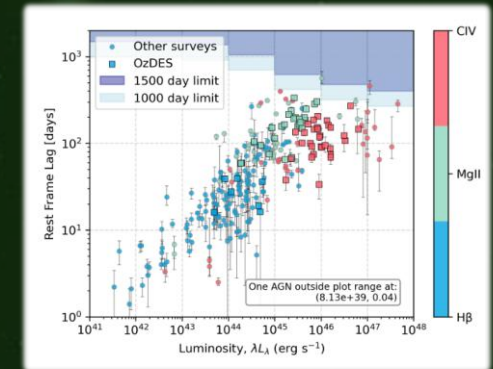


*Changing Cut and Constrain
choices gives different results*

Methods Dominate
Results

CIV

Differences in Survey Length

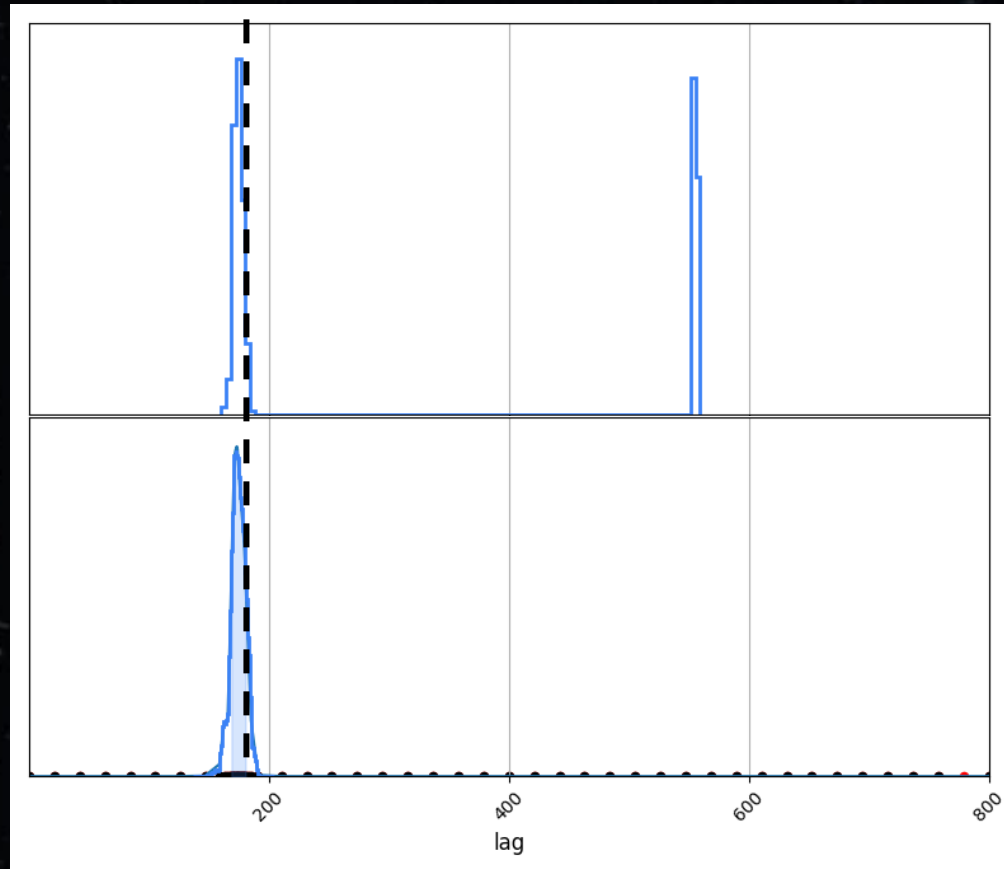


*Can only see lags $\ll$ Survey
Length ("Lag Ceiling")*

Step 1: Trying to Recreate JAVELIN

Lag Posterior
Distribution from
JAVELIN

True Lag Posterior
Distribution

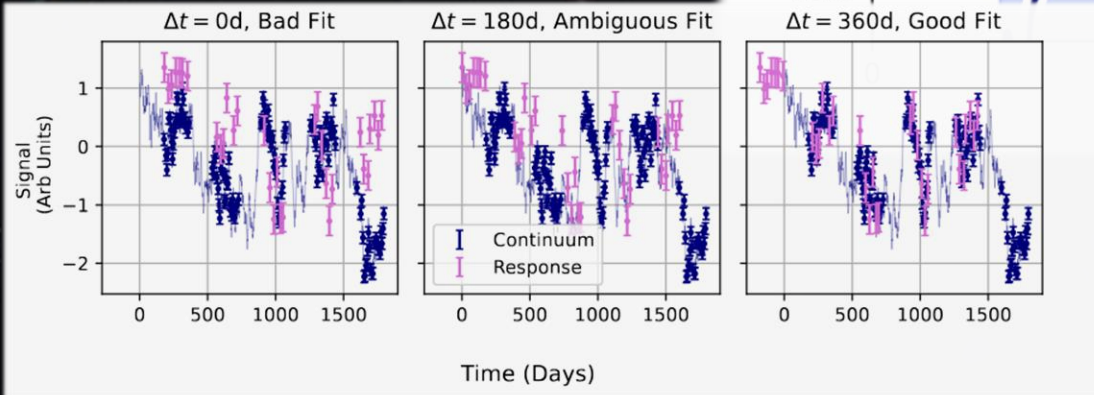
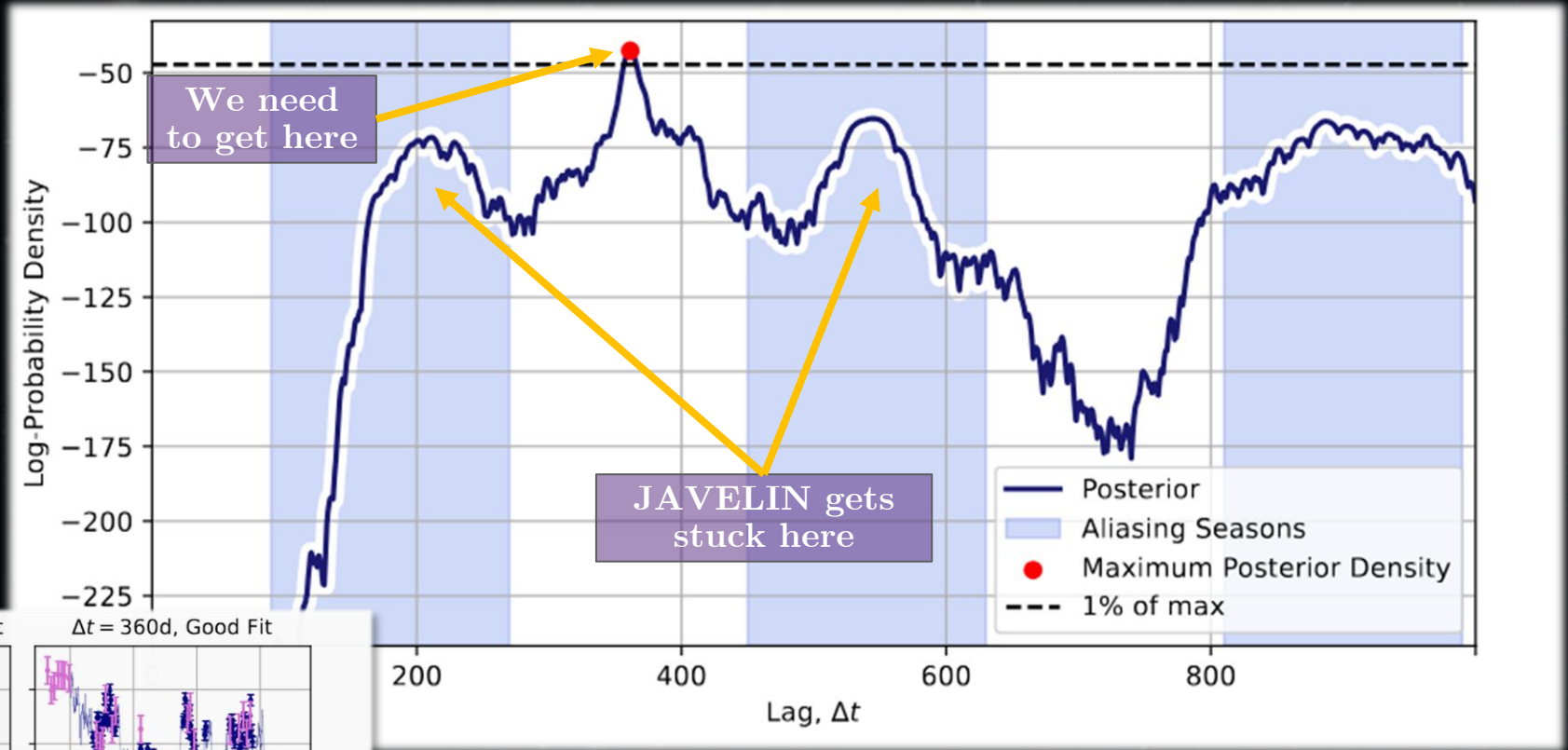


Same data, same stats model,
different aliasing peaks.
Why?

Reason: Inconvenient Likelihood

Lag-Axis Log Posterior distribution is extremely noisy and multimodal

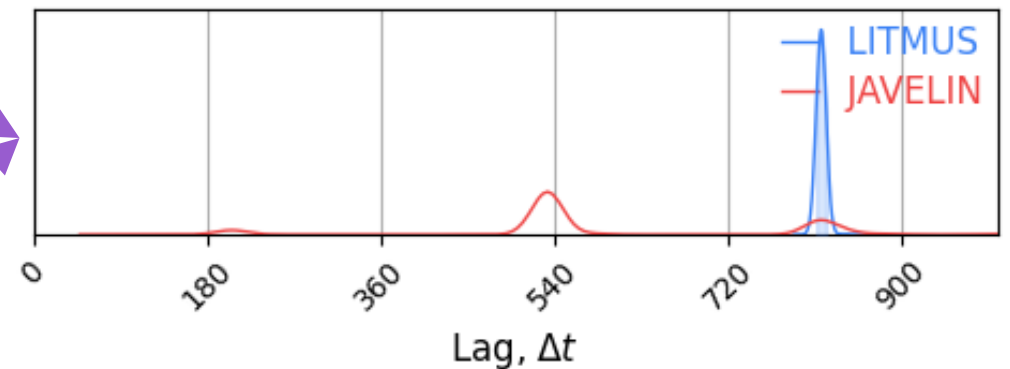
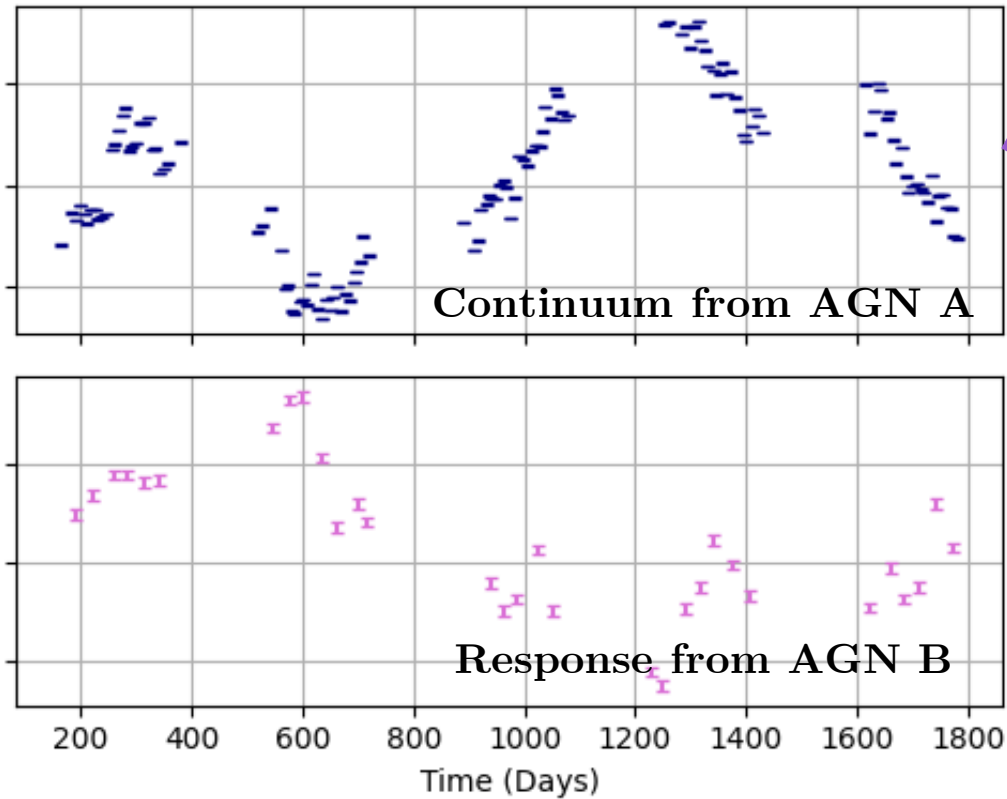
JAVELIN's MCMC fitter (emcee) can't handle multimodal distributions



Off-season Lags with no Overlap are Ambiguous:
Locally Optimal fits

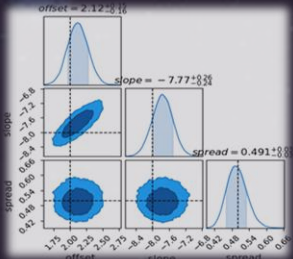
False Positives

We can get a well constrained lag by random chance even if we combine light curves from different AGN!



Coincidence can give a “Best Fit” Lags that isn’t necessarily a *good* fit.

Consistent Bayesian Model Comparison



$$\rho(\theta) = \mathcal{L}(D|\theta) \times \pi(\theta)$$

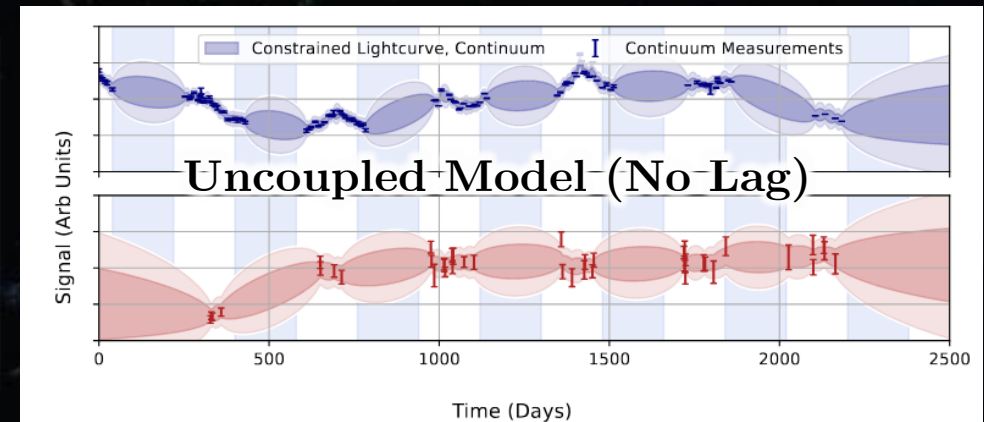
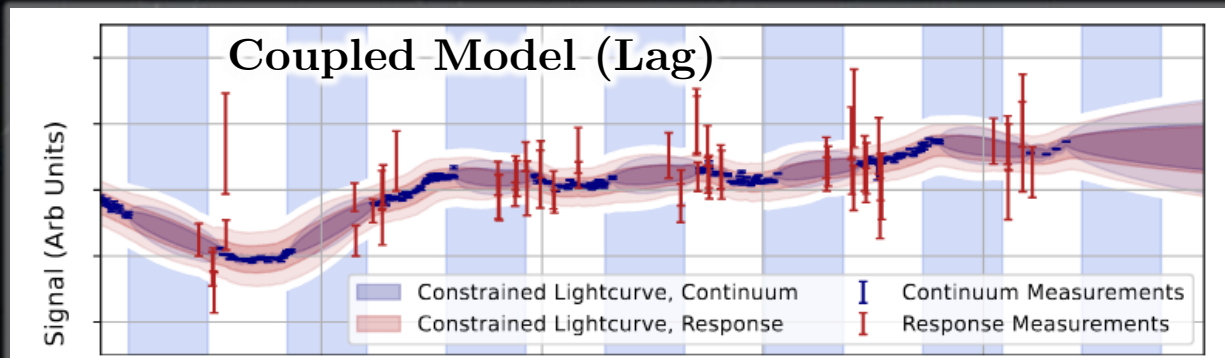
Integrates To

Probability Mass (Evidence)

$$\mathcal{Z} = \int \rho(\theta) d\theta$$

Compare Models

$$\text{BF}_{\text{Lag}} = \mathcal{Z}_{\text{Lag}} / \mathcal{Z}_{\text{No Lag}}$$



LITMUS: A New RM Tool

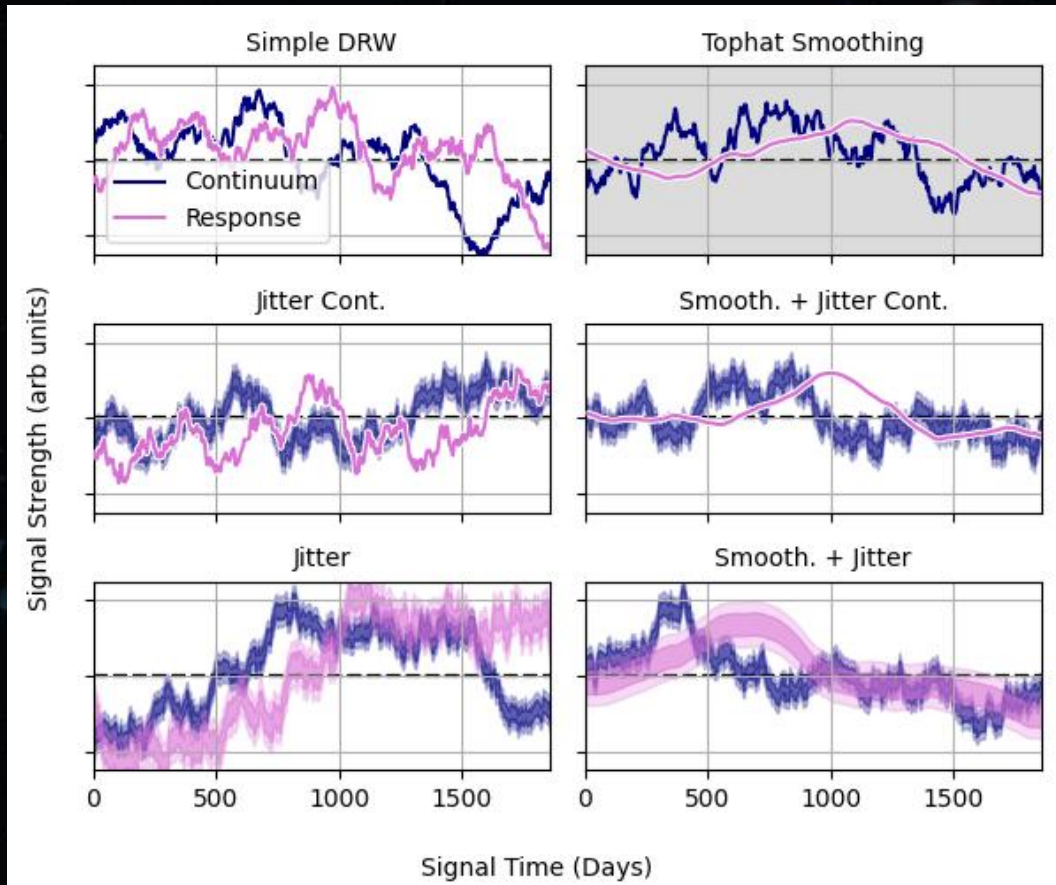
Mission Statement:

1. Measure Lags Properly
 - Bayesian GP modelling
 - Fix the multimodal issue
2. Proper Significance Measures
 - Bayesian model comparison:
Coupled vs uncoupled models
 - We need to be able to **integrate**
as well as explore the posterior

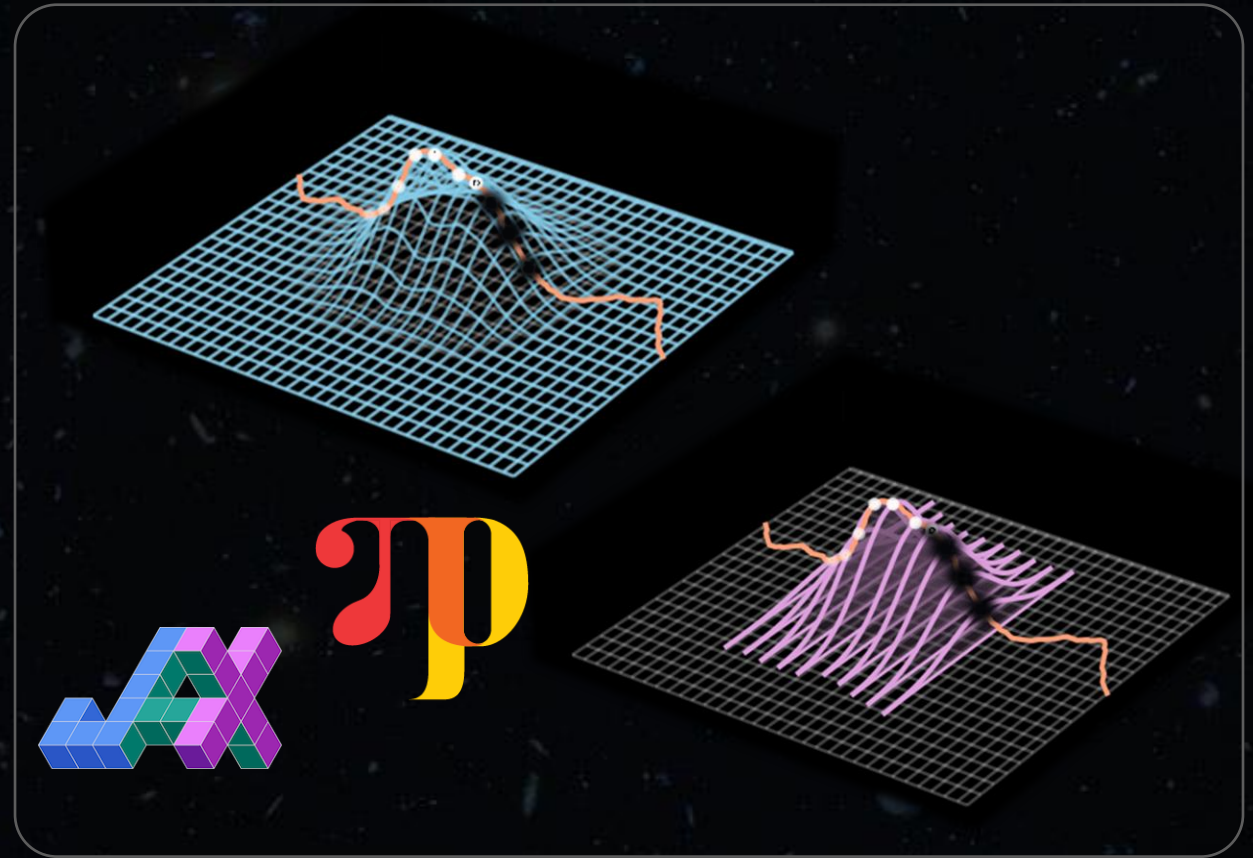


LITMUS: Other Features

Modular design lets us test different light curve models



Super-Fast Custom-Made Fitters with **JAX** and **NumPyro** (Laplace Approx Quadrature)



Applying to Real Data



Even after doing the stats rigorously, only a tiny fraction of OzDES targets have clear lags. Why?

Why so few recoveries?

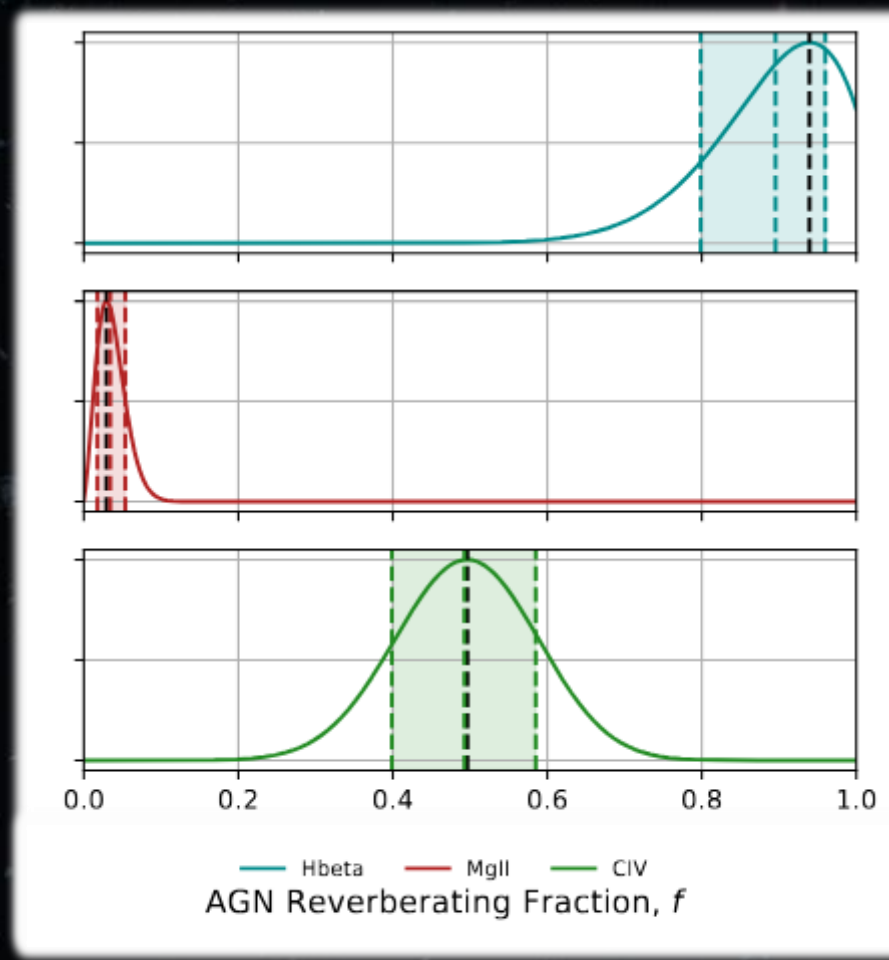
Do all AGN actually have lags? We can estimate $f = \frac{N_{\text{has a lag}}}{N_{\text{Total}}}$ using Bayes Factors:

$$\mathcal{L}(D|f) \propto \prod_i f \cdot \frac{\text{BF}_i}{1 + \text{BF}_i} + (1 - f) \frac{1}{1 + \text{BF}_i}$$

$\text{BF}_i = \text{Bayes Factor for Source } i$

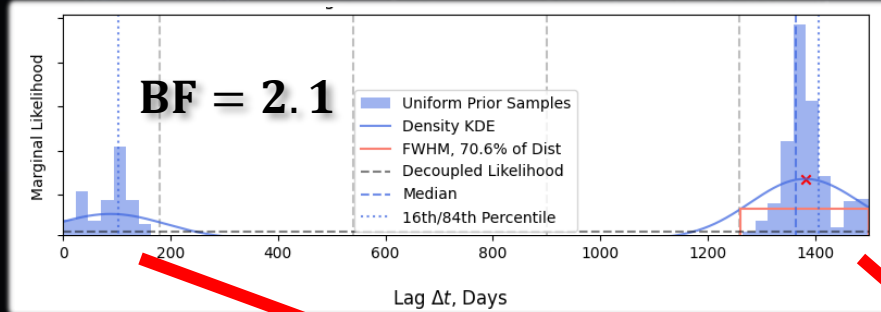
Two Conclusions:

1. Many AGN don't have lags
2. SNR too low to see lags for most that do

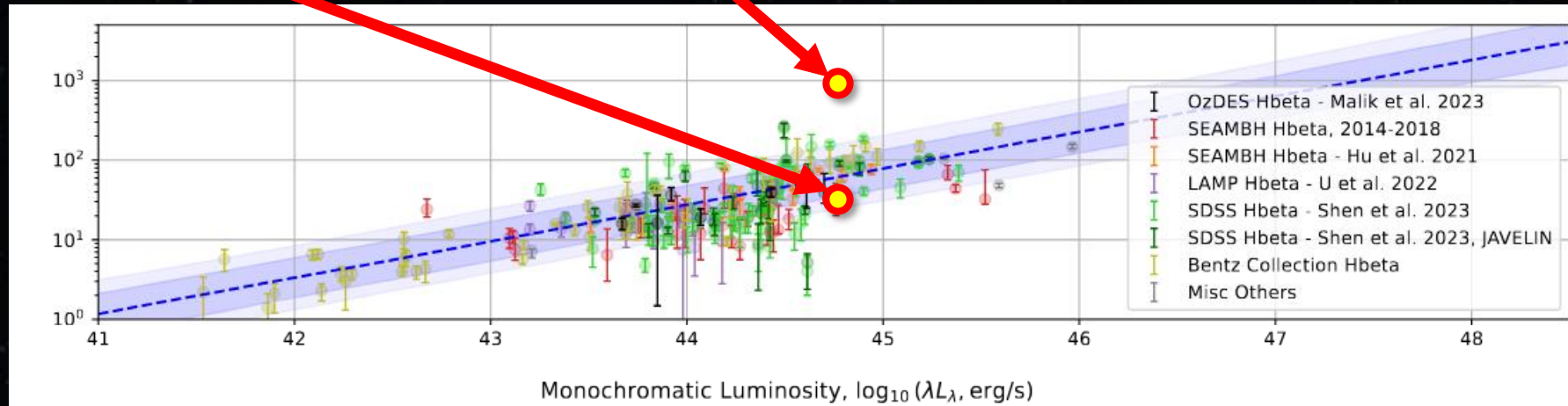
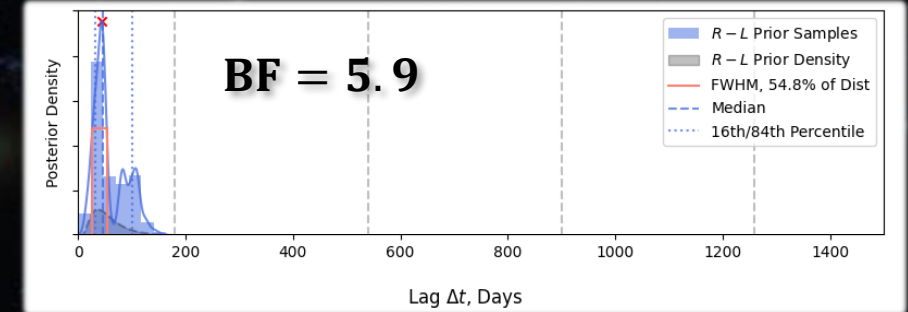


R-L Relation As a Prior Fixes SNR + Aliasing Issues

Fit with Uniform Lag Prior



Fit with R-L Informed Prior

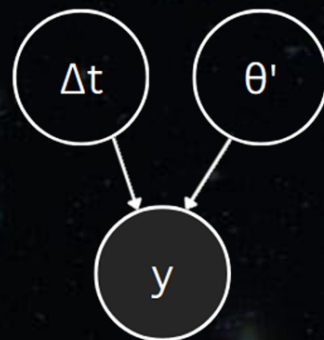


How do we do this without confirmation bias?

A Bayesian Hierarchical Model

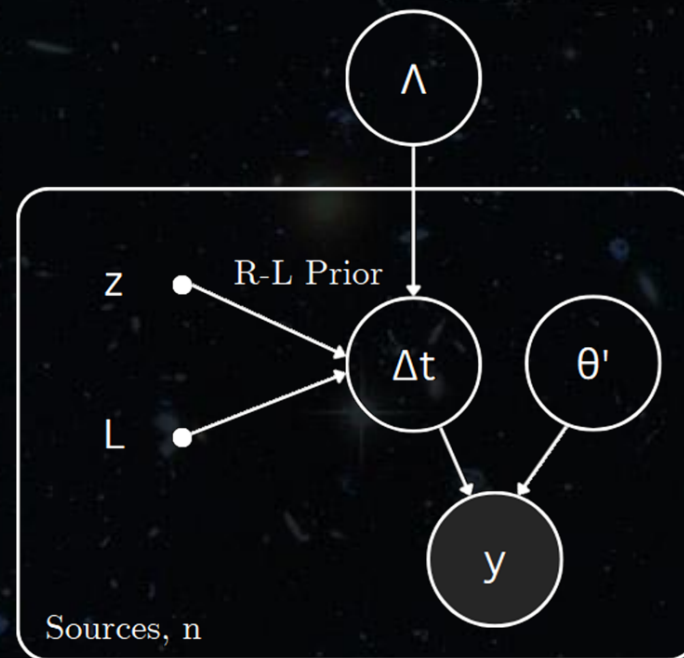
Old Way: Fit Lags one at a time, discard all but the most convincing ones

Uniform Prior



Single-Source Model

$$P(\Delta t^n, \phi'^n | d^n, \text{Lag}) \propto \pi(\phi'_n) \pi(\Delta t_n) \mathcal{L}_{\text{Lag}}(d^n | \Delta t_n, \phi'_n)$$



Hierarchical Model

New Way: Lags & Population Hyper-parameters fit simultaneously in one giant Bayesian model

Highly Multimodal *and* high dimensional: Can't be fit with normal methods

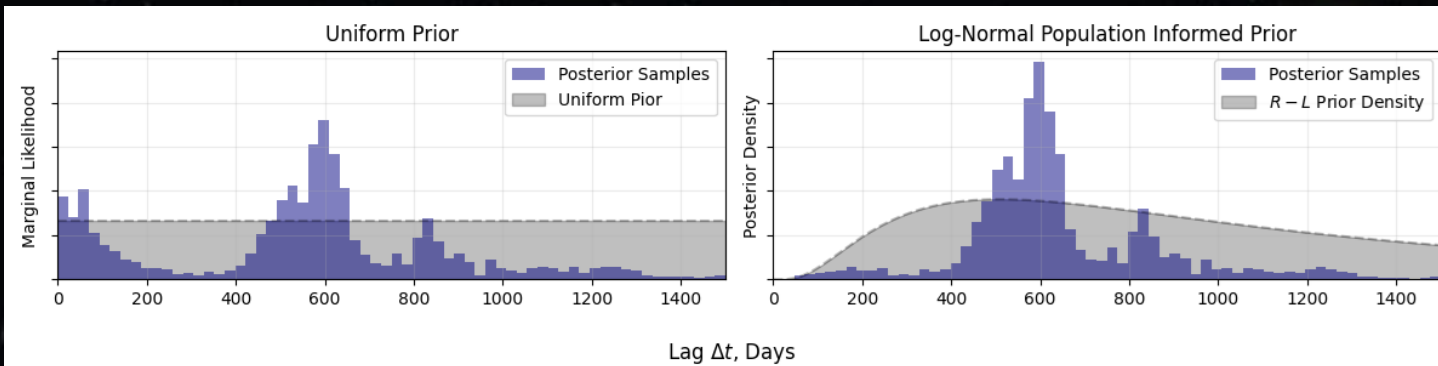
$$P(\Lambda, f, \{\Delta t^n, \phi'^n\} | \{d^n\}) \propto \pi(\Lambda, f) \prod_n \pi(\phi'_n) \left(\overset{\text{RL Informed Prior}}{f \cdot \pi(\Delta t_n | \Lambda)} \mathcal{L}_{\text{Lag}}(d^n | \Delta t_n, \phi'_n) + (1 - f) \cdot \mathcal{L}_{\text{No Lag}}(d^n | \phi'_n) \right)$$

Reverb / non reverb frac Lag & Lag-Free Likelihoods

The Trick To Making This Tractable: Importance Sampling

We already have information about the posterior shape from our vague prior fits.

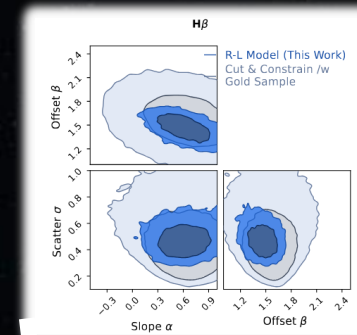
Importance Sampling lets us re-weight source-by-source fits to new priors



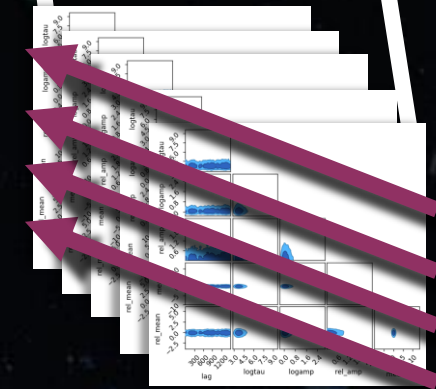
It also lets us marginalize over source parameters with Monte Carlo Integration:

$$\mathcal{Z}_{\text{Lag}}^n(d^n|\Lambda) = E \left[\frac{\pi(\Lambda)}{\pi_\phi} \right]_{\text{Post.}(\Lambda)} \approx \mathcal{Z}_{\text{Lag}}^\phi(d^n) \sum_{\text{Fid.Post Samps}} \frac{\pi(\Delta t_n|\Lambda)}{\pi^\phi(\Delta t)}$$

Population Params (≈ 4)

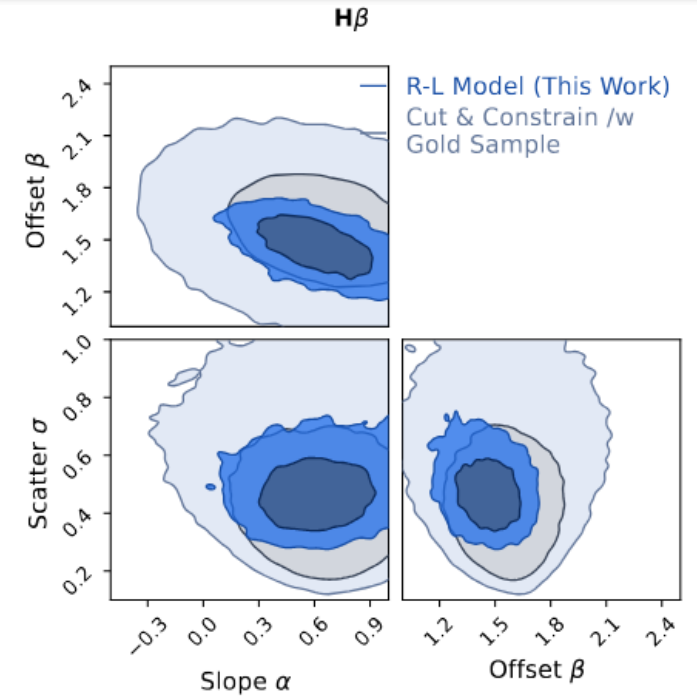
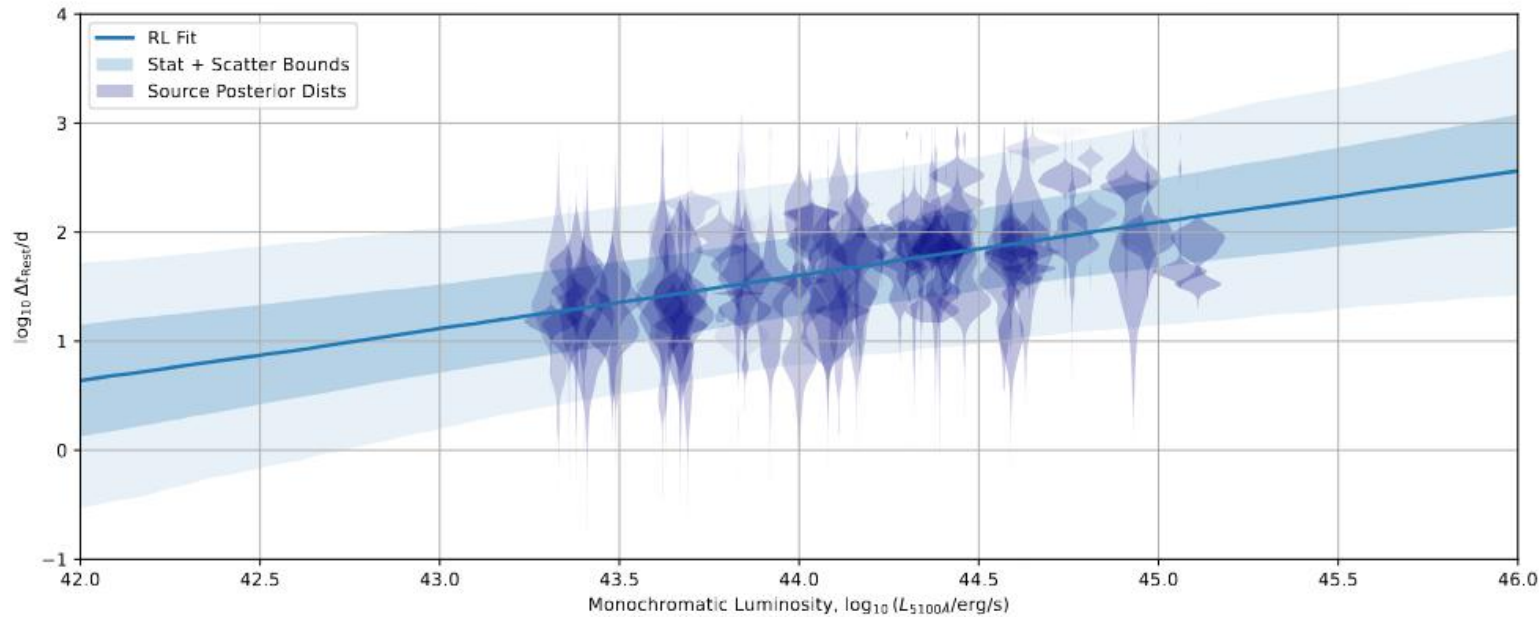


Only 4 Params Remain



Marginalise out messy
source params

Results: Drastically Improved Performance



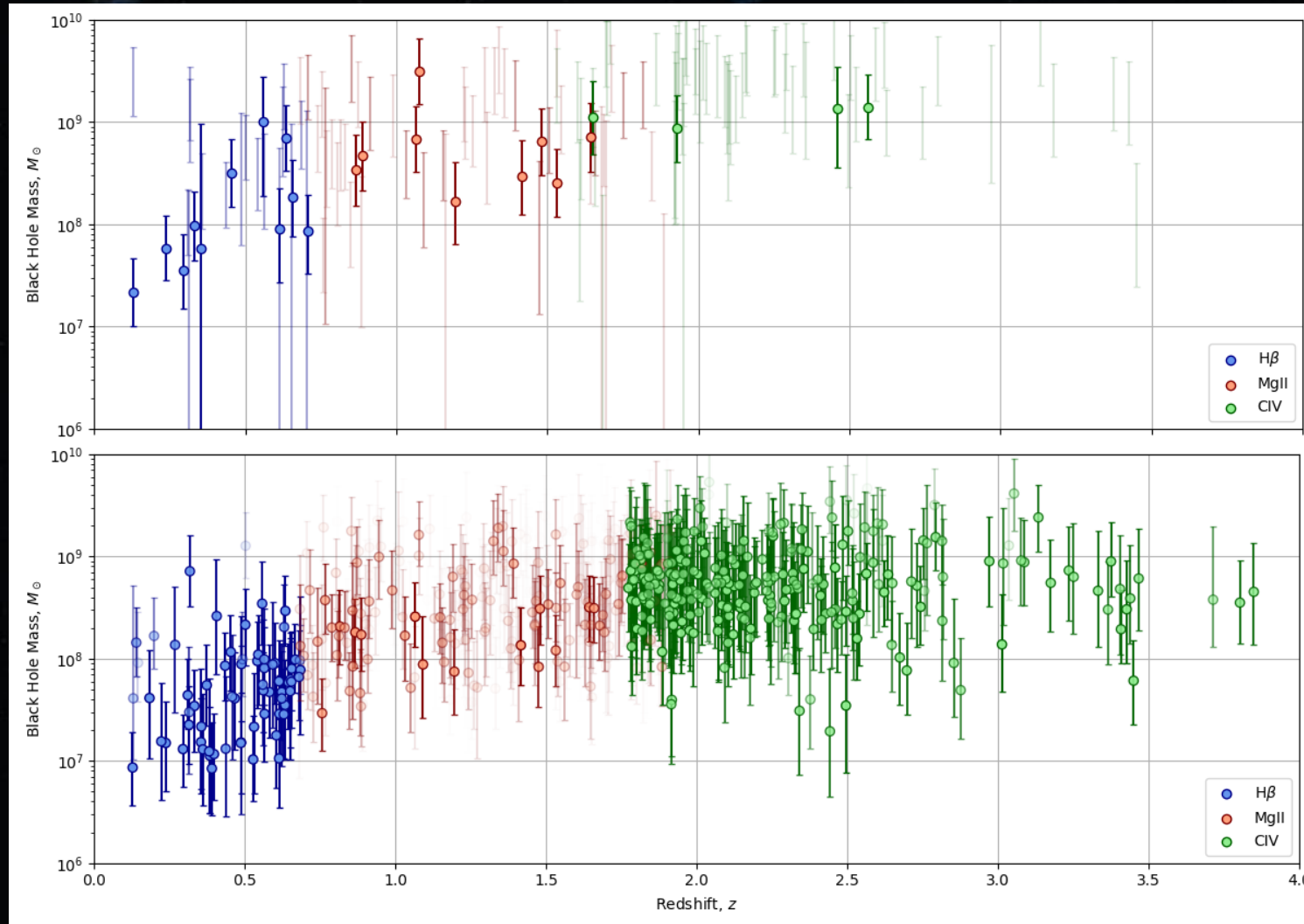
Without any new data:

- More than $2 \times$ increase in constraining power for R-L relation
- $\sim 5 - 10 \times$ increase in number of lags and masses

Results: Hugely Increased Sample Size

Old Method + **LITMUS**
(Cut and constrain):

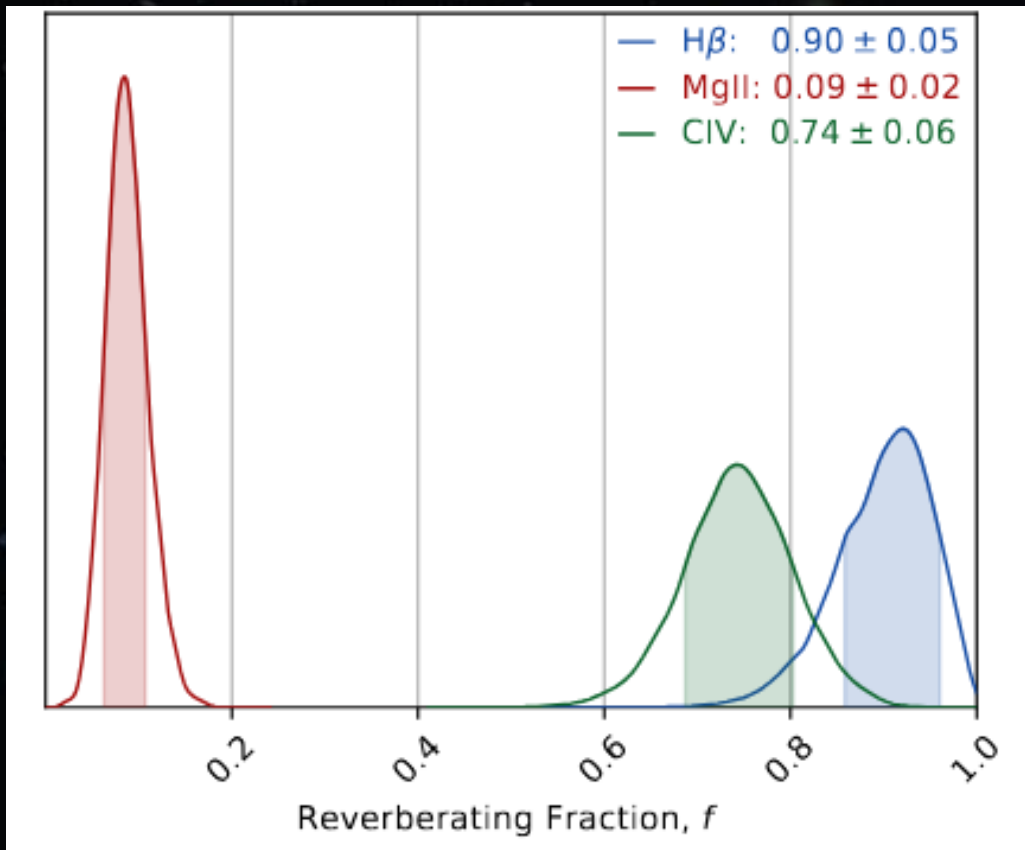
A few dozen SMBH Masses
with some false positive
contamination



New Method
(Hierarchical):

Hundreds of AGN masses

Results: Reverberating Fraction



$H\beta$

Nearly all sources have lags
(consistent with 100%)

$MgII$

$\sim \frac{2}{3}$ of sources are stationary
Of the $\frac{1}{3}$ that vary, only $\sim \frac{1}{3}$ have lags

CIV

$\approx 25\%$ are decoupled, even when
accounting for lags too long to see

The Future: More Data!



Similar AGN populations, different pipelines and systematics.

Will their results be the same?

Can be applied to many other RM types and Surveys!



(Acc. Disk RM)



Upcoming

Summary

Methods

1. New RM code (LITMUS) that fixes a major numerical issue and properly models false positives
2. A new AGN RM population fitting pipeline that models-in window effects. Drastically improved accuracy and precision over traditional methods
3. Both of these can be applied to non-BLR RM! Any time-lag estimation (disk RM, dusty torus RM)

Science

1. For OzDES MgII sources:
 1. $\sim 2/3$ of do not have any structured variability
 2. Only $\sim 9\%$ have RM lags
2. Possible explanations:
 1. Failure of our variability models, our spectral reduction or our reverberation model
 2. An actual physical feature of the MgII population

Thanks
Questions?

