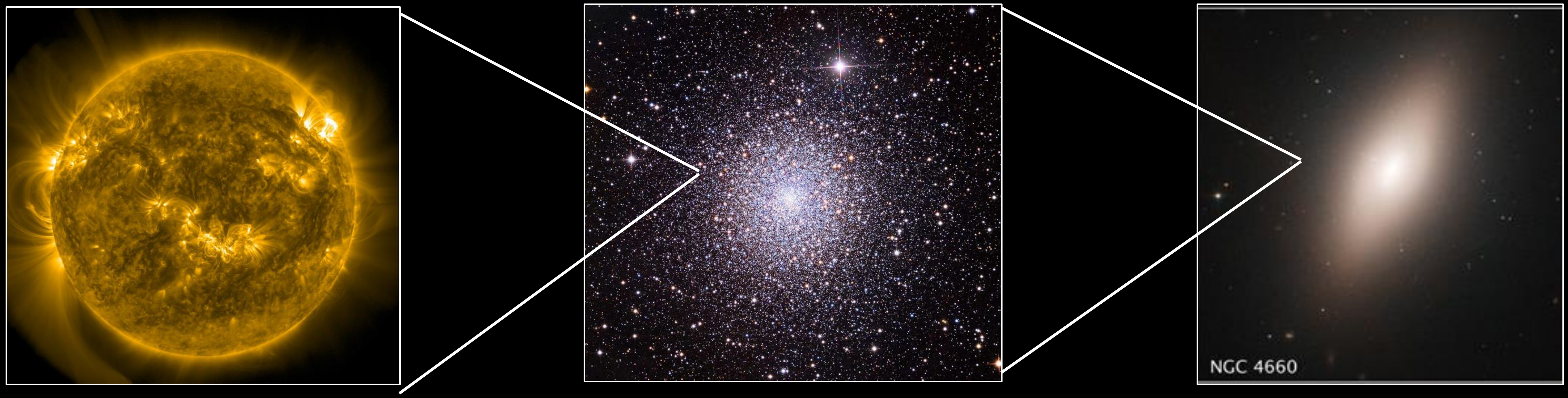


STUDYING THE UNIVERSE WITH ASTROSTATISTICS

STAMPS Colloquium , Carnegie Mellon University, October 11, 2024

UNIVERSITY OF TORONTO

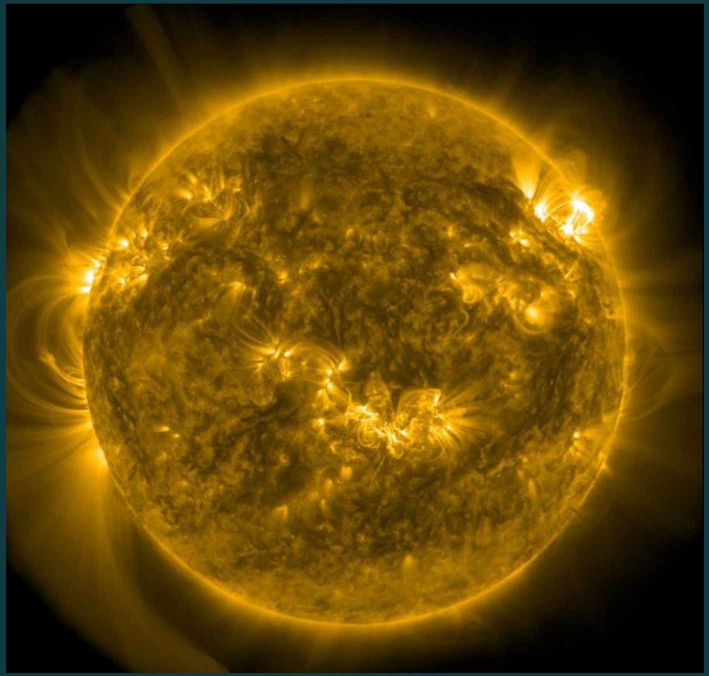
Gwendolyn Eadie, Assistant Professor of Astrostatistics
David A. Dunlap Department of Astronomy & Astrophysics
Department of Statistical Sciences

DUNLAP INSTITUTE *for* ASTRONOMY & ASTROPHYSICS
DUNLAP INSTITUTE *for* ASTRONOMY & ASTROPHYSICS

Statistical Sciences
UNIVERSITY OF TORONTO

DATA SCIENCES INSTITUTE

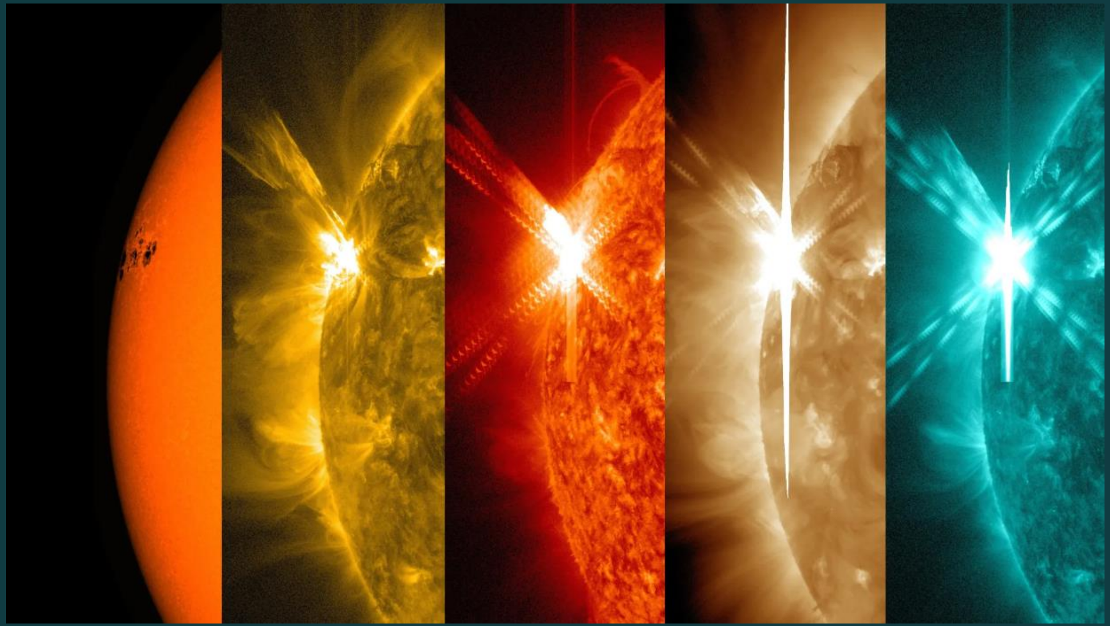
ART



STELLAR FLARES IN HIDING

using hidden Markov models to find stellar flares in time series data from TESS

SOLAR FLARES



Solar flare on May 5. 2015, in 5 different wavelengths (visible to UV)
Credit: NASA/Solar Dynamics Observatory/Wiessinger

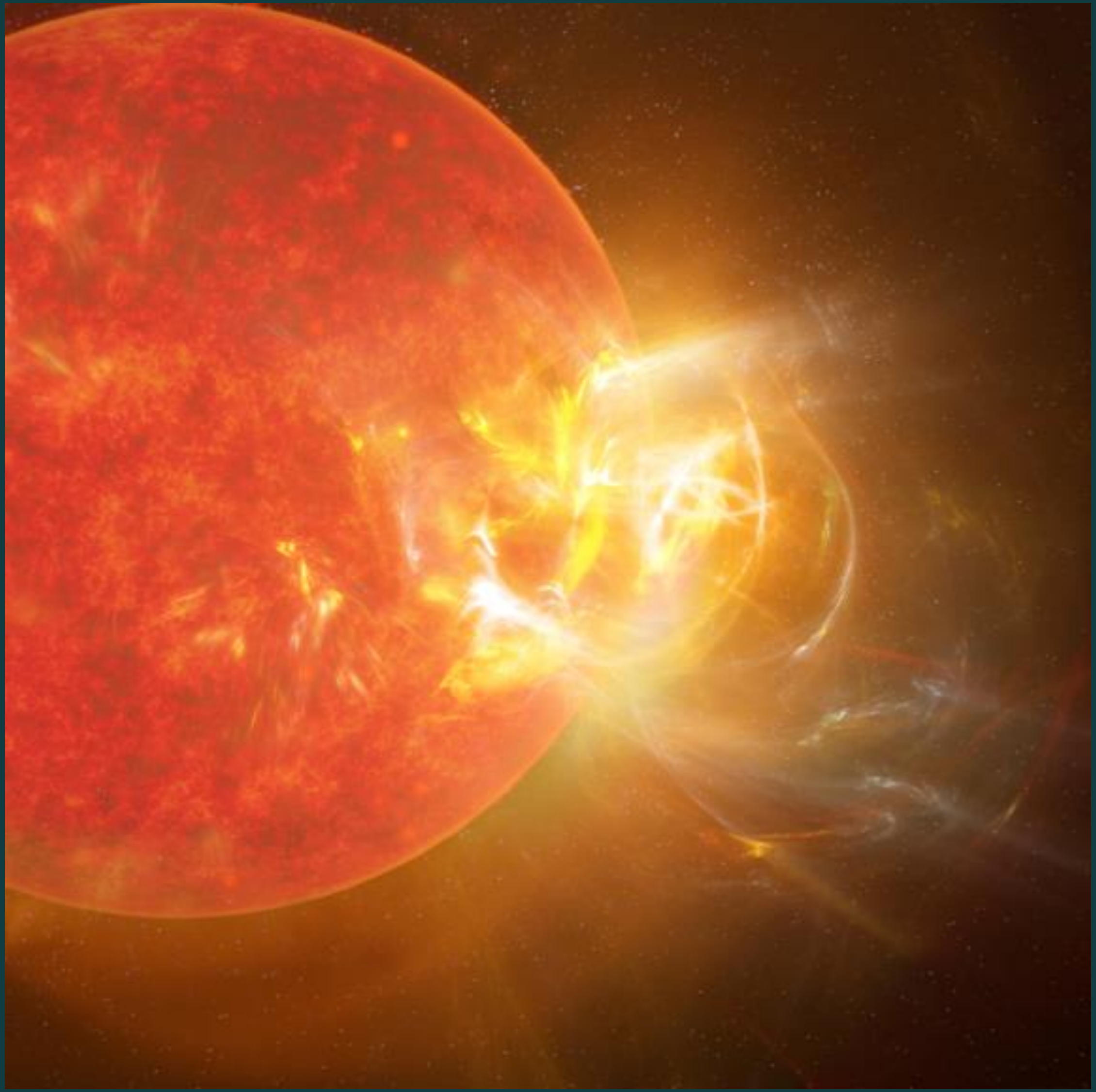
- Bursts of radiation and material from the surface of the Sun
- Related to magnetic activity
- Frequency of flaring related to solar cycle

STELLAR FLARES

- Like solar flares, but on other stars
- Related to:
  - magnetic activity of star
  - Stellar age
  - Stellar rotation

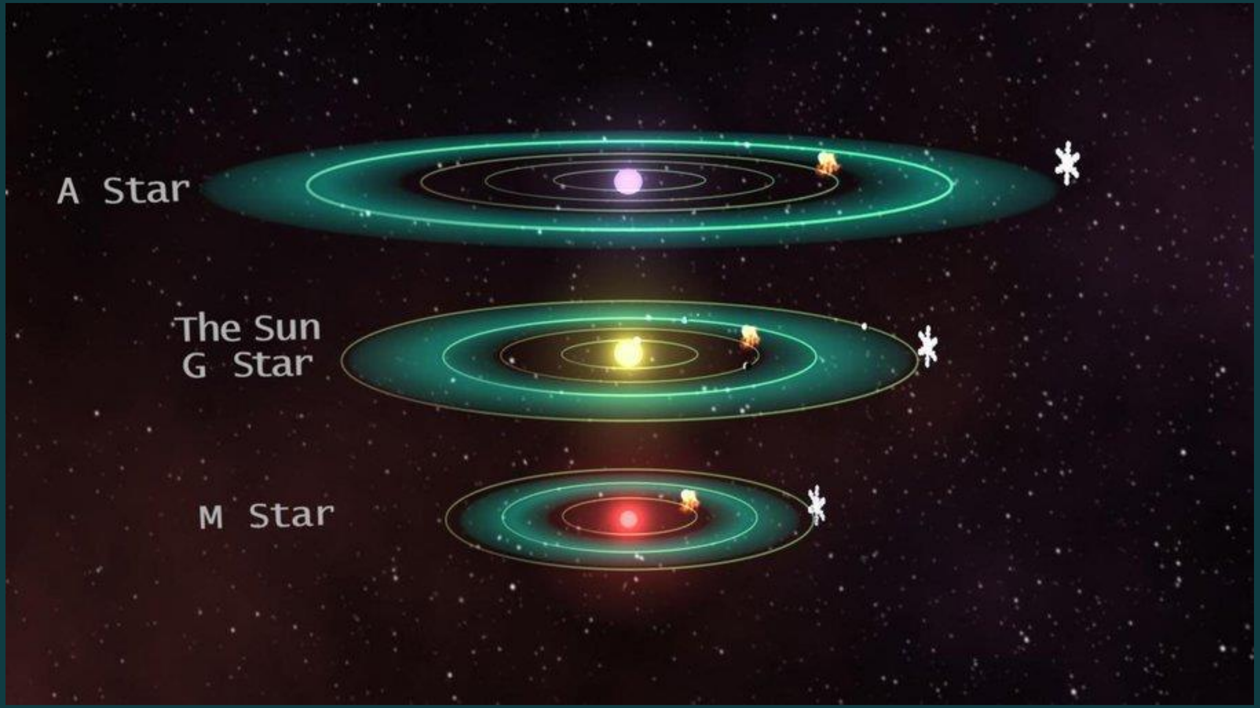
STELLAR FLARES

→ M-dwarf stars



M-DWARF STARS

- 70% of all stars in our Galaxy
- Low-mass stars ($0.08M_{\odot} < M < 0.6M_{\odot}$)
- Burning hydrogen into helium
- When $\mathrm{M} < 0.35M_{\odot}$, fully convective cores
- Long lifetimes (up to trillions of years)
- Open questions:
  - Size of habitable zone
  - Level of activity vs. age
  - Frequency of flaring events

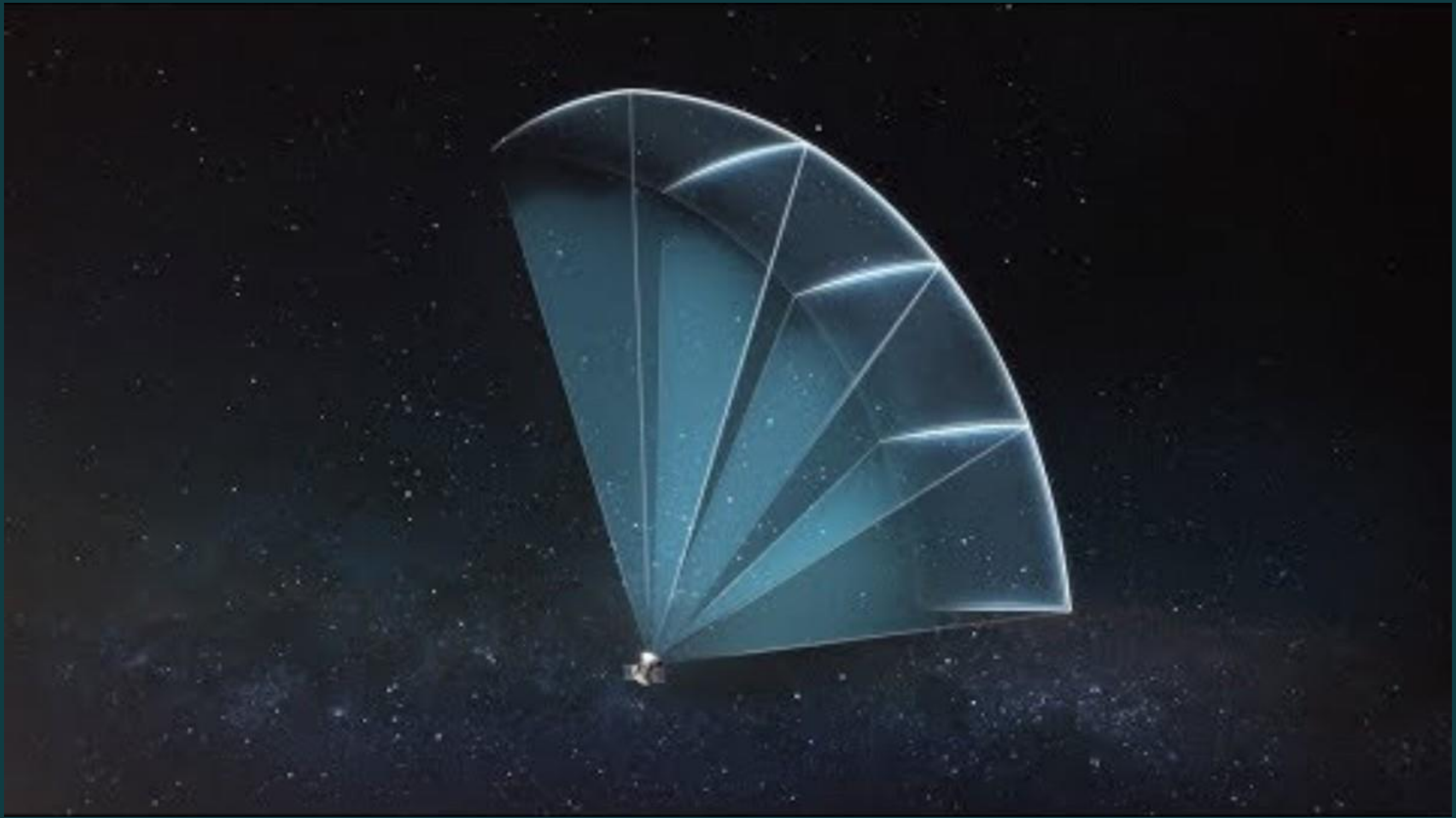


https://astrobiology.nasa.gov/news/where-is-the-habitable-zone-for-m-dwarf-stars/

IMAGE CREDIT: NASA/JPL-CALTECH/MSSS.

The estimated habitable zones of A stars, G stars and M stars are compared in this diagram.

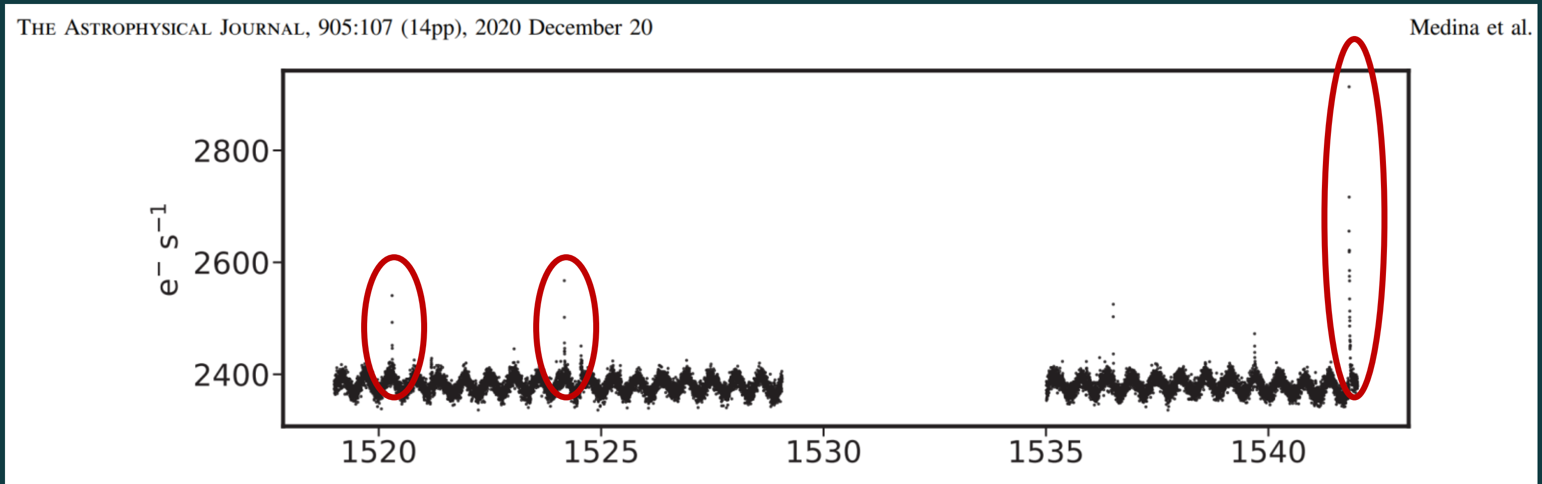
TRANSITING EXOPLANET SURVEY SATELLITE



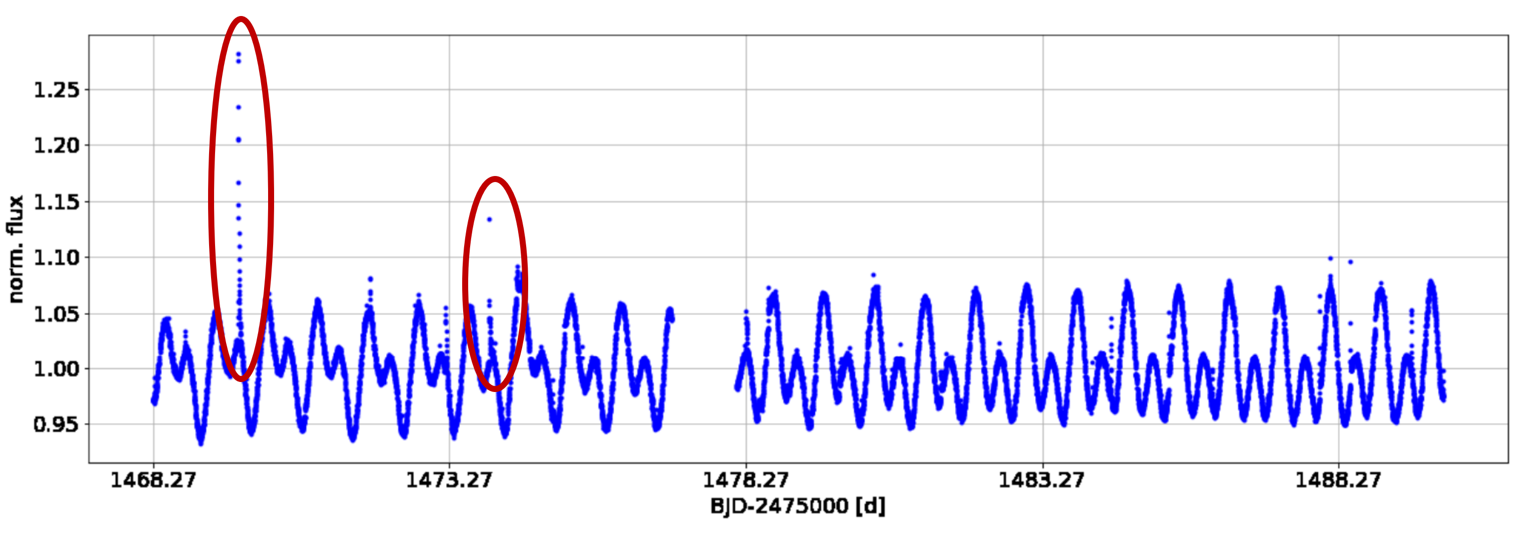
https://www.youtube.com/watch?mute=1;v=evHF_mnIdj4

OBSERVING STELLAR FLARES

Brightness



Medina et al (2020), ApJ

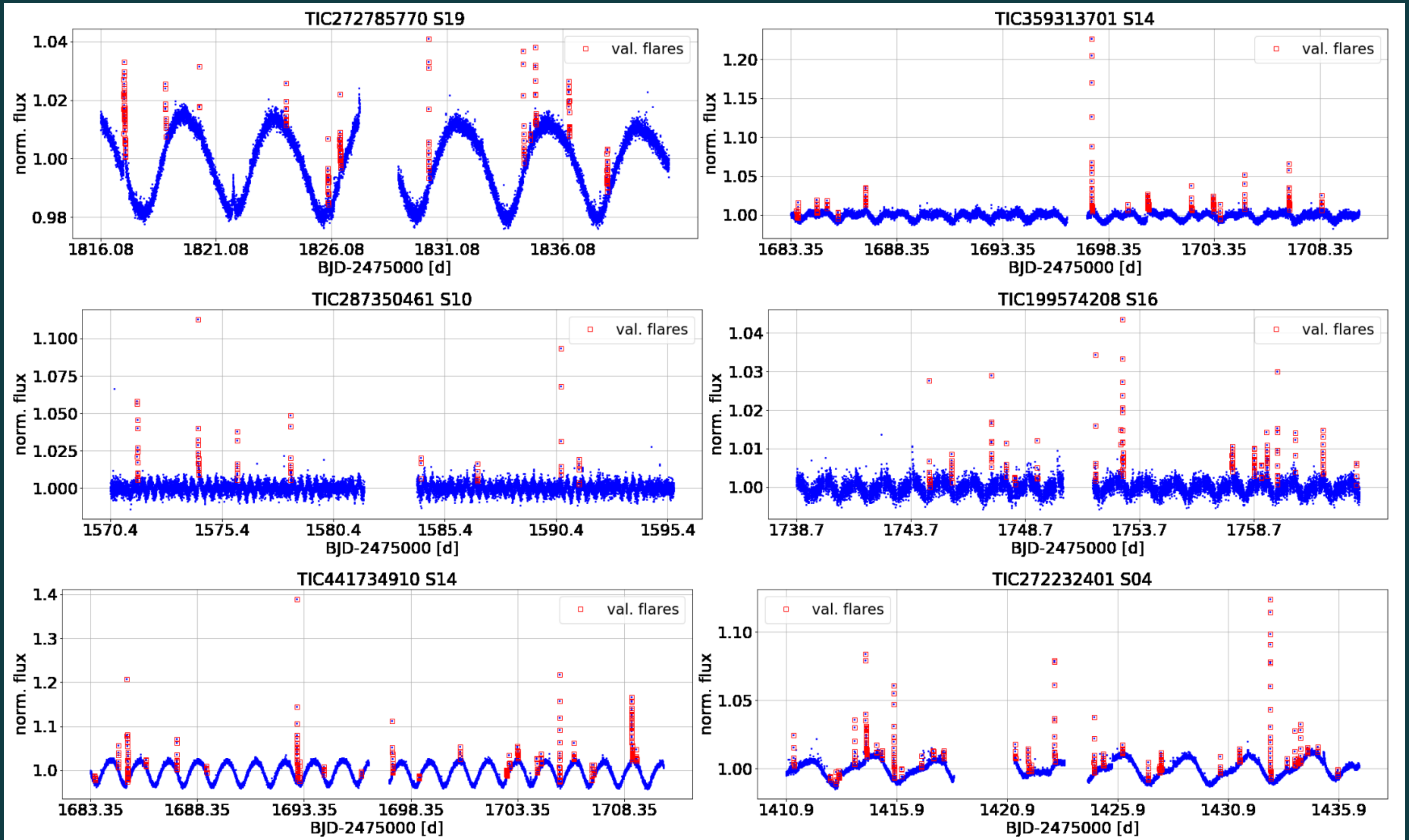


Stelzer et al (2022), A&A

Time (days)

OBSERVING STELLAR FLARES WITH TESS DATA

Brightness



Time (days)

Stelzer et al (2022), A&A

IDEALIZED FLARE TEMPLATE

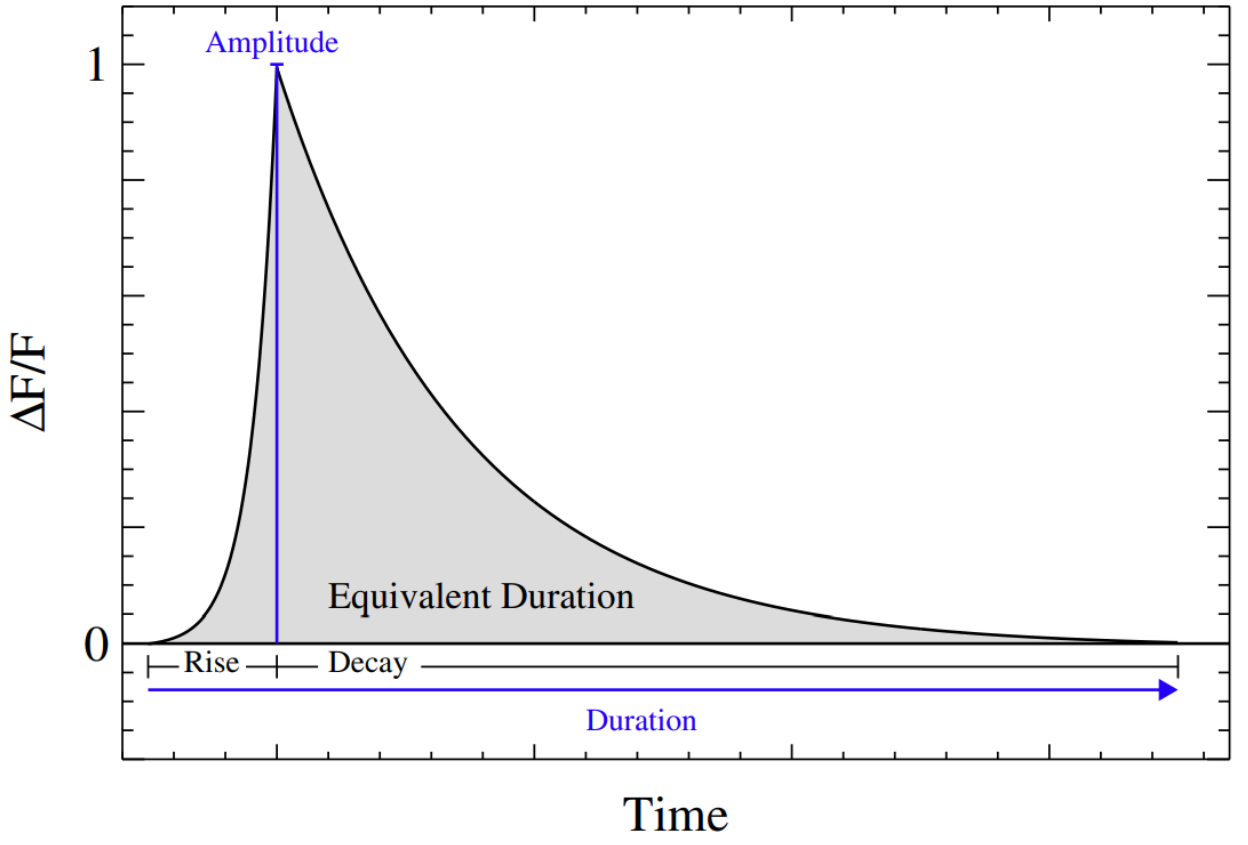


Figure 6. Idealized classical flare light curve showing measured quantities: amplitude, rise and decay times, duration, and equivalent duration.

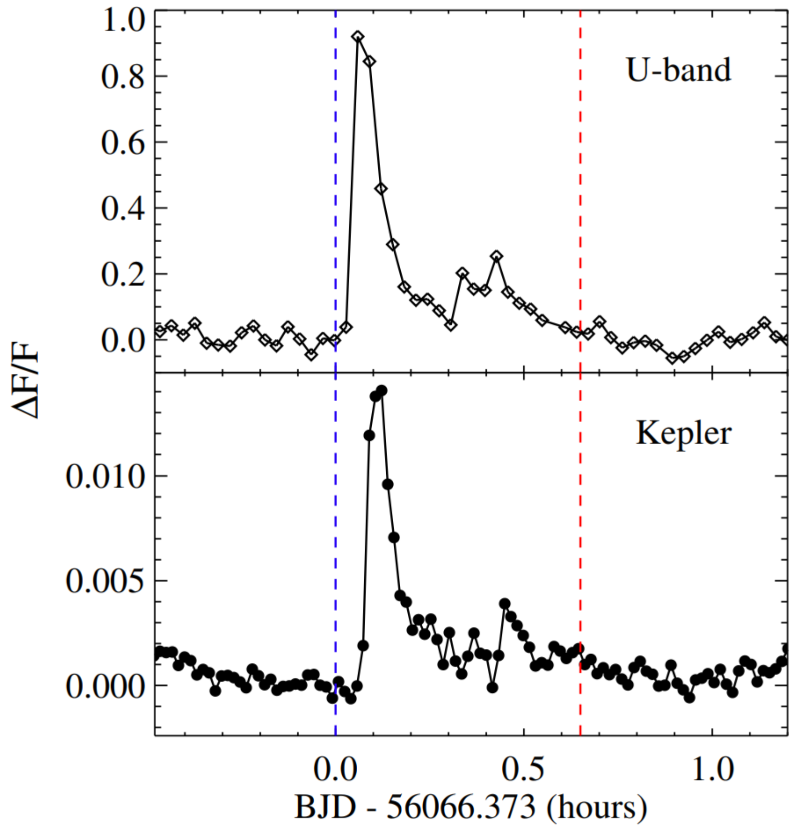


Figure 7. Simultaneous *Kepler* and *U*-band (from the NMSU 1 m telescope at APO) observations of a flare on GJ 1243. The start and stop times used to compute the equivalent durations are indicated by vertical dashed lines.

Hawley et al (2014), 797:121 ApJ

FLARE DISTRIBUTIONS

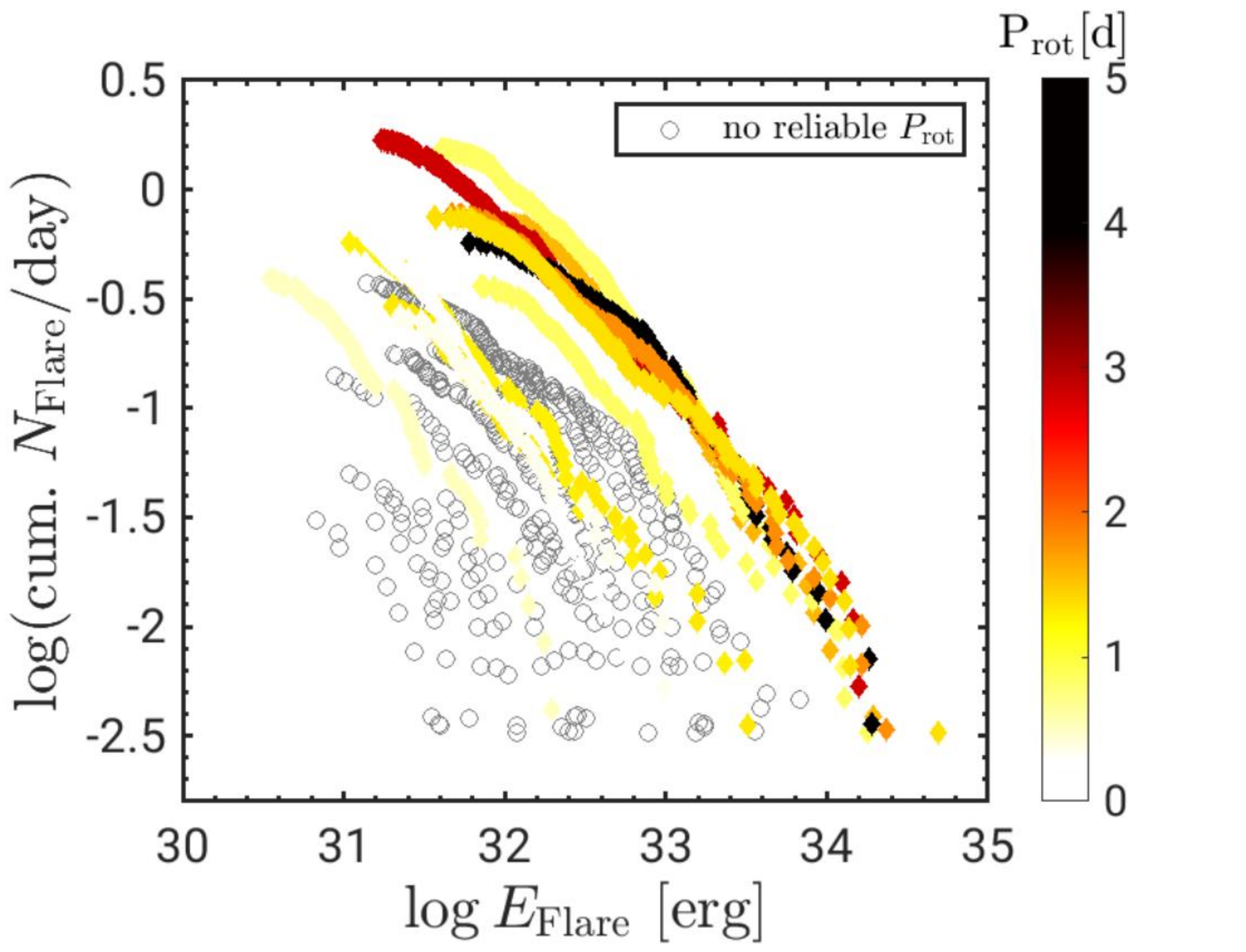
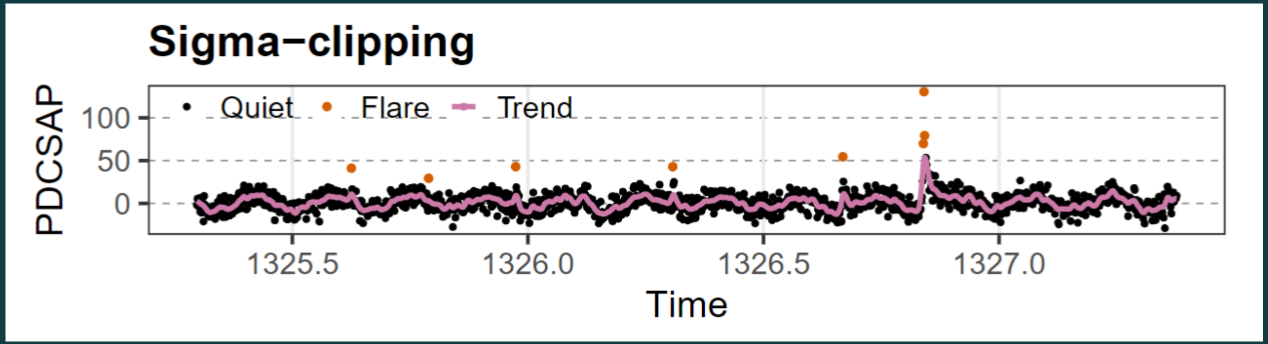


Fig. 12. Cumulative distribution of flare energies for all 35 flaring stars; the 12 stars with reliable P_{rot} are highlighted in color.

Stelzer et al (2022), A&A 665, A30

APPROACHES IN THE LITERATURE

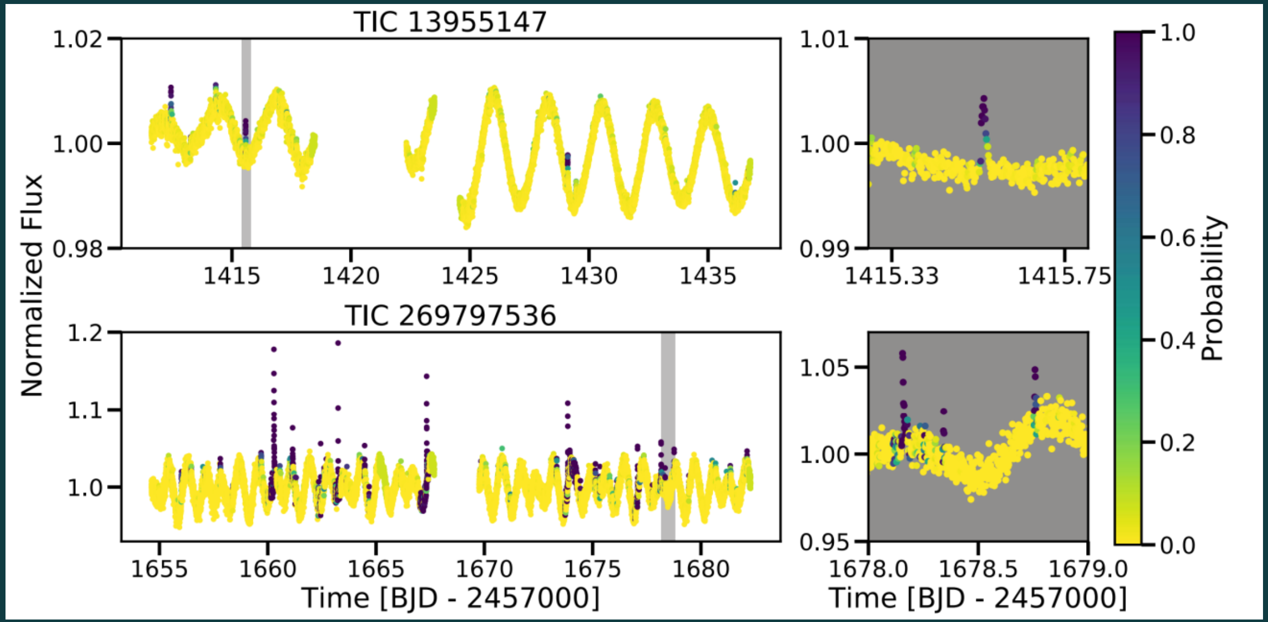
- **"Sigma-clipping"**
  (e.g. Walkowicz et al. 2011; Osten et al. 2012; Hawley et al. 2014; Davenport 2016; Yang et al. 2017; Günther et al. 2020)



- **Convolutional neural networks** applied to young stars
  (e.g. *stella,* Feinstein et al 2020, AJ)

- **Change-point detection**
  (e.g., Chang et al 2015 ApJ)

- Once flares identified → template fitting



Feinstein et al 2020, AJ

MEASURING STELLAR FLARES

The Astrophysical Journal, 905:107 (14pp), 2020 December 20 Medina et al.

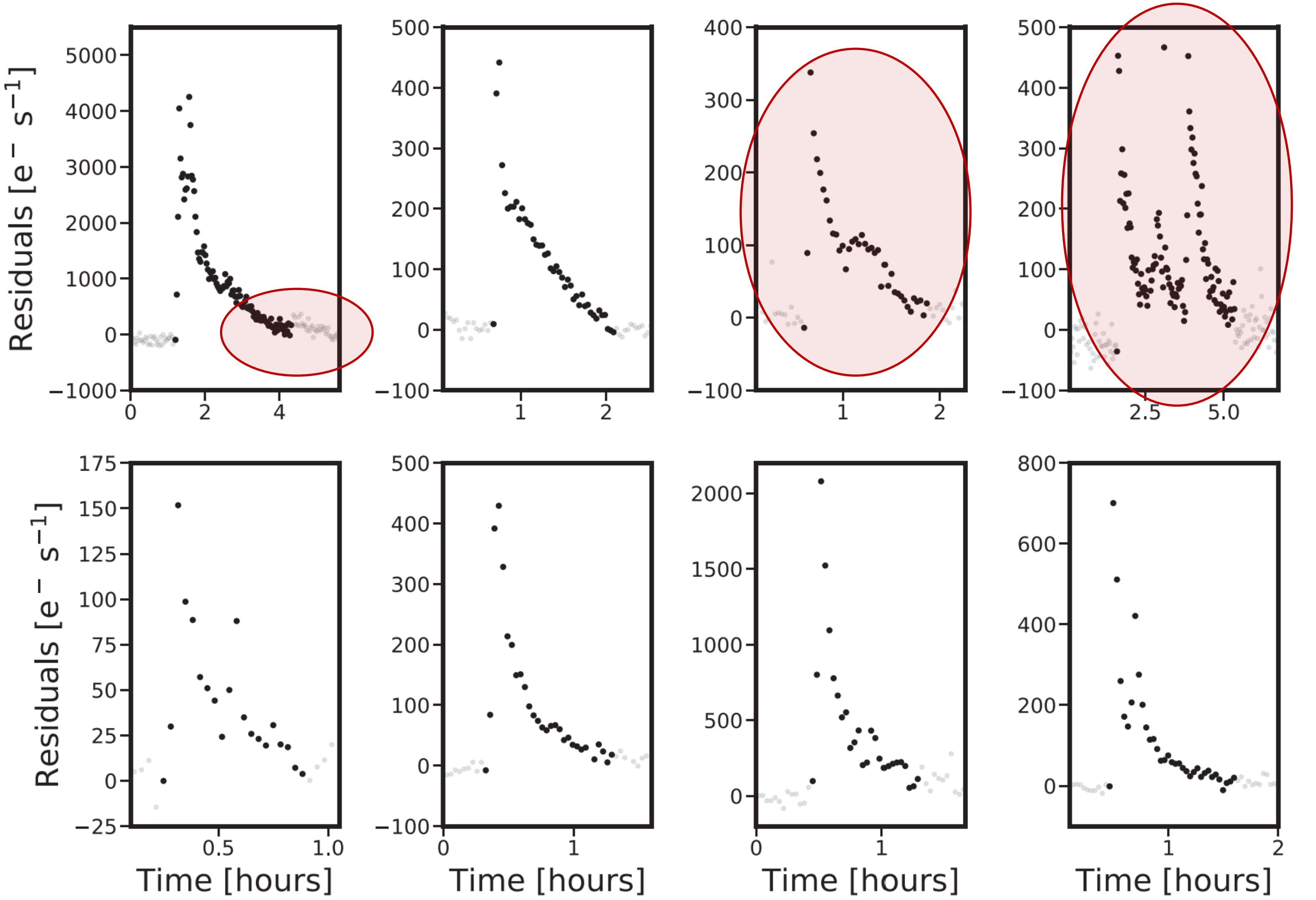


Figure 4. Examples of flares detected with our flare-finding algorithm for six different stars in our sample. Flares come from Proxima Centauri, SCR J0754-3809, UPM J0409-4435, APMPM J0237-5928, L 87-10, and Gl 54.1. The black points denote points used to calculate the flare energy.

Challenges:

- End of flare ill-defined
- Compound flares
- Energy of flare difficult to measure
- Smaller/fainter energy flares difficult to detect

Medina et al (2020), 905:107 ApJ

Arturo Esquivel
PhD Candidate, statistics, UofT

OUR APPROACH: HIDDEN MARKOV MODEL

$$Y_t | S_t \sim f_t + Z_t | S_t$$

Gaussian Process to model quasi-periodic trend (*celerite*, Foreman-Mackey et al 2017)

$$f | \mu, K \sim Celerite(\mu, K)$$

Flaring Channel:

→ Quiet

→ Firing

→ Decay

Probability of transiting between states

Esquivel et al (2024), submitted to ApJ

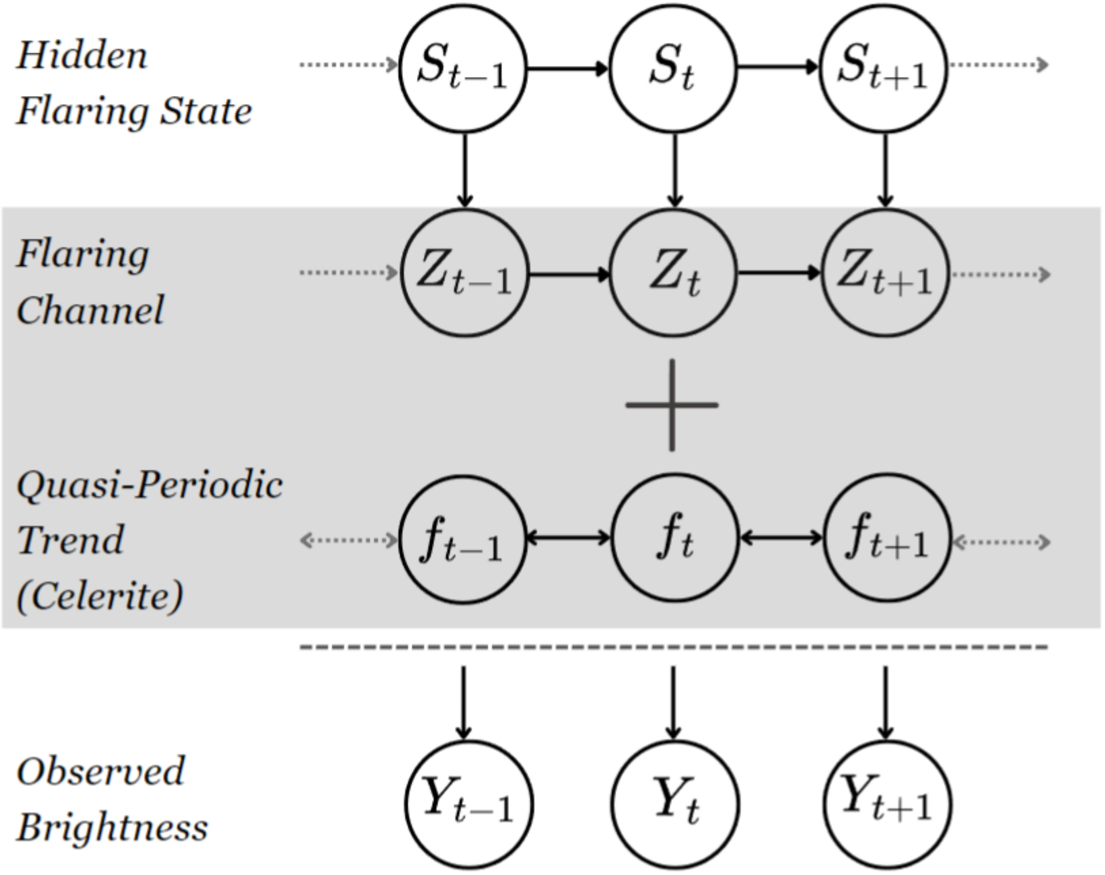


Figure 1. Graphical representation of our complete model for observed brightness decomposed into its various parts. The Quasi-Periodic Trend reflects the average brightness of the star when it is not flaring. The Flaring Channel represents the extra brightness due to the state of the star. The Hidden Flaring State represents the (unobserved) state of the star (Q,F, or D).

OUR APPROACH: HIDDEN MARKOV MODEL

		to		
		Q	F	D
	Q	remain quiet	start firing	*forbidden*
from	F	*forbidden*	increased firing	start decaying
	D	return to quiet	start compound flare	decaying continues

$$Y_t | S_t \sim f_t + Z_t | S_t$$

Quiet state $Z_t | Q, Z_{t-1}, \sigma^2 \sim N(0, \sigma^2)$

Firing $Z_t | F, Z_{t-1}, \lambda\sigma^2 \sim N(Z_{t-1}, \sigma^2) + Exp(\lambda)$

Decay $Z_t | D, Z_{t-1}, r, \sigma^2 \sim N(rZ_{t-1}, \sigma^2)$

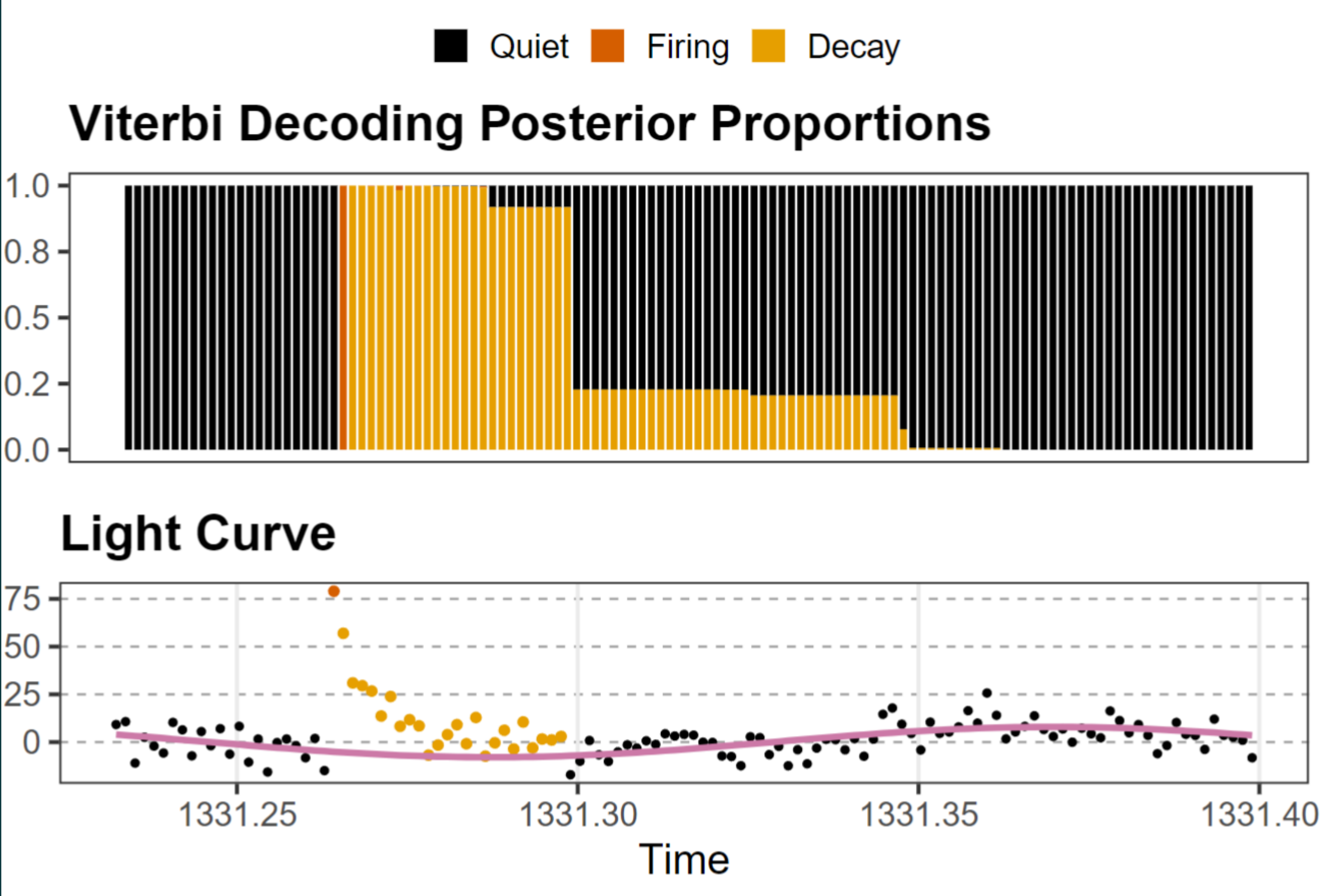
Bayesian approach

Esquivel et al (2024), submitted to ApJ

COMPUTATION

- Bayesian approach using *Stan* (Carpenter et al 2017)
- Split time series data into 2000-step chunks
- Gaussian Process and HMM parameters sampled simultaneously
- For each sample of parameters, state decode (Viterbi algorithm)
- distribution of states for each time step

POSTERIOR DISTRIBUTION OF STATES



Esquivel et al 2024, submitted to ApJ

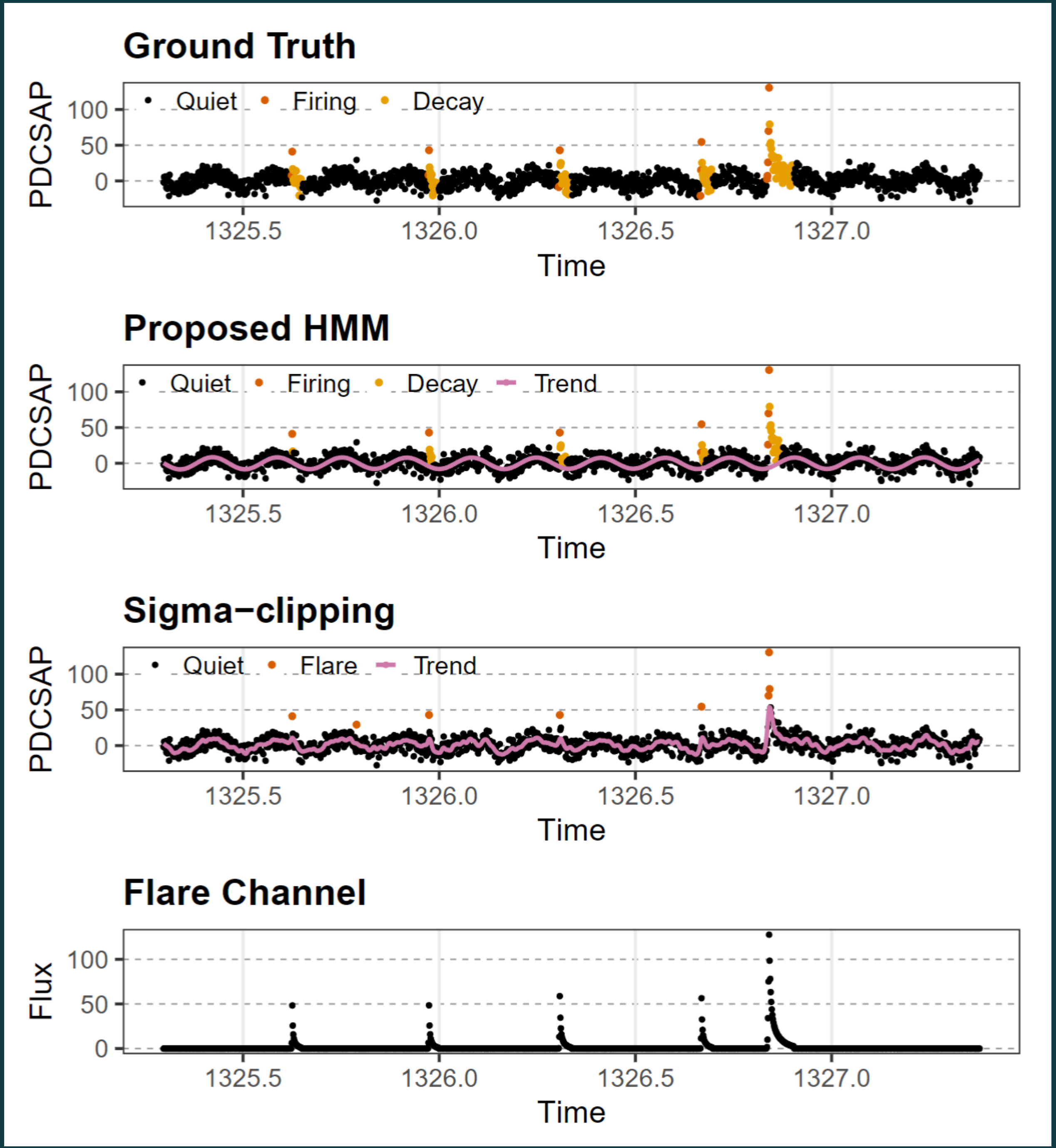
Potential advantages:

- Uncertainty estimate of flare duration
- Uncertainty estimate of flare energy
- Pick up lower energy flares

INJECTION TESTS

- M-dwarf star TIC 031381302 (TESS)
  → Sections that are very quiet
  → Section with known flares
- Injected small and large simulated flares
- Sensitivity $= \dfrac{true\ flares\ detected}{number\ of\ true\ flares}$
- Positive Predictive Value

$$= \frac{true\ flares\ detected}{number\ of\ flares\ detected}$$



INJECTION TESTS

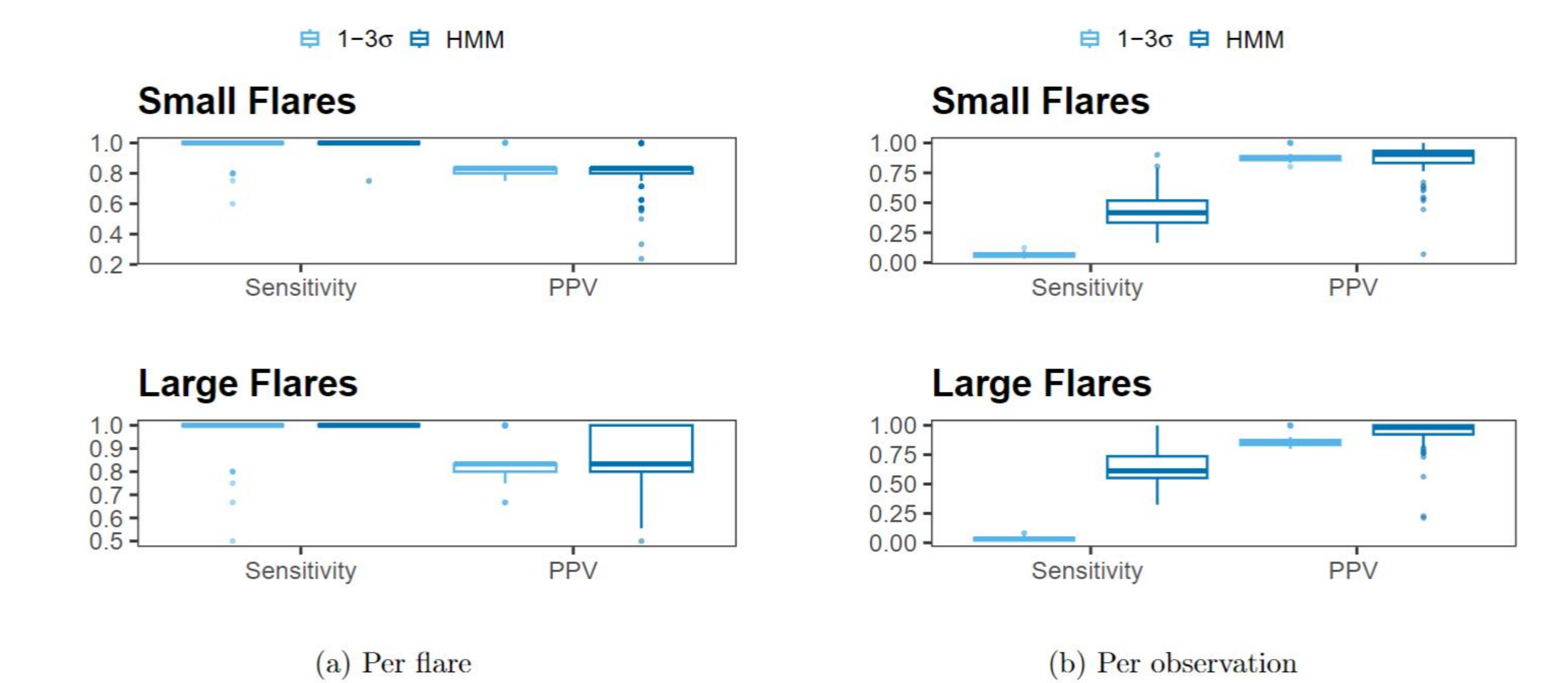
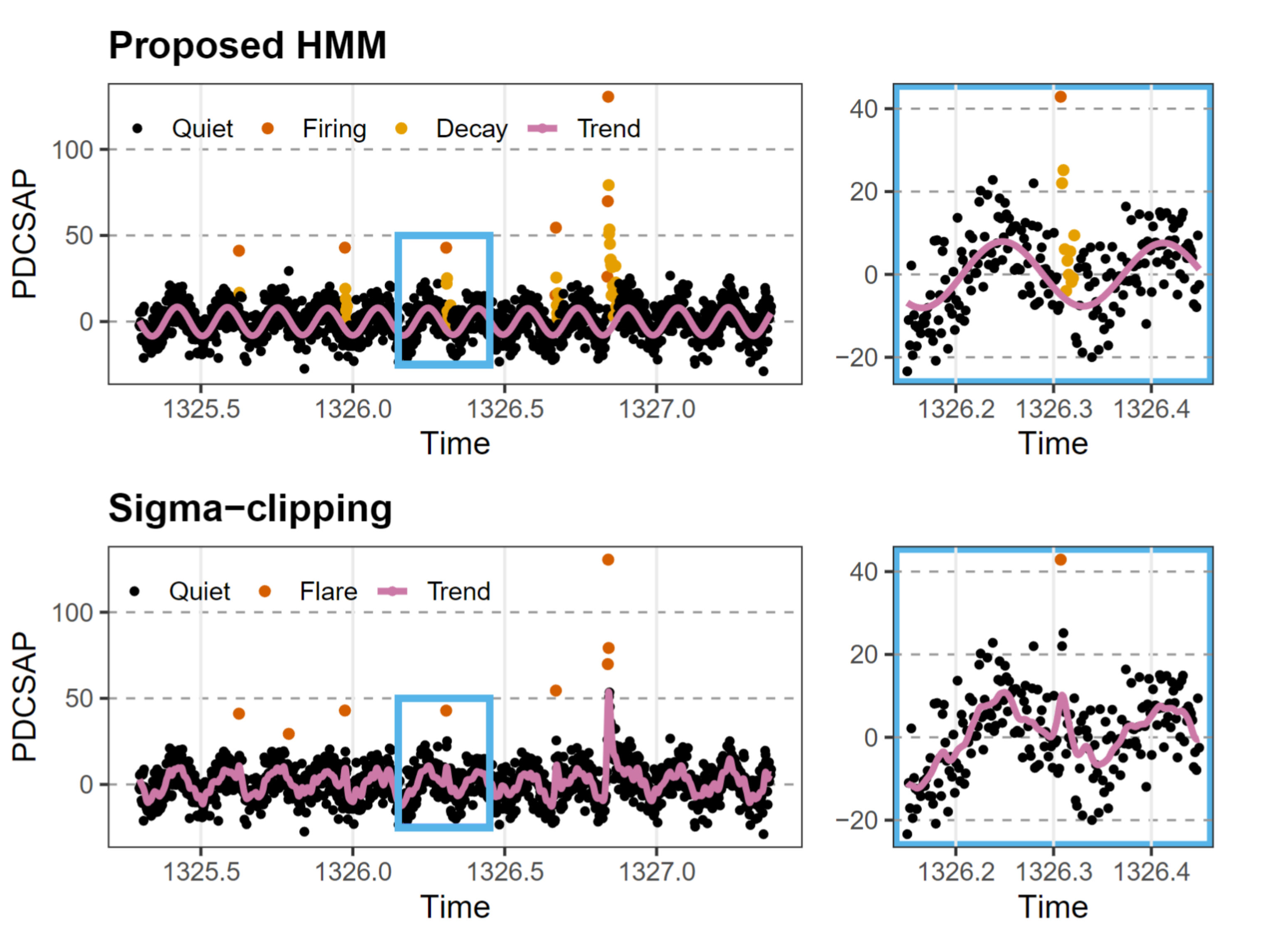


Figure 5. Flare recovery sensitivity and positive predictive value (PPV) distributions across 100 small and large flares injections, using the 1-3σ rule and *celeriteQFD* (HMM). *(a)* shows results from detecting flare occurrences (equations (6) and (7) are used). *(b)* shows results on a per-observation basis (equations (8) and (9) are used).

Esquivel et al 2024, submitted to ApJ

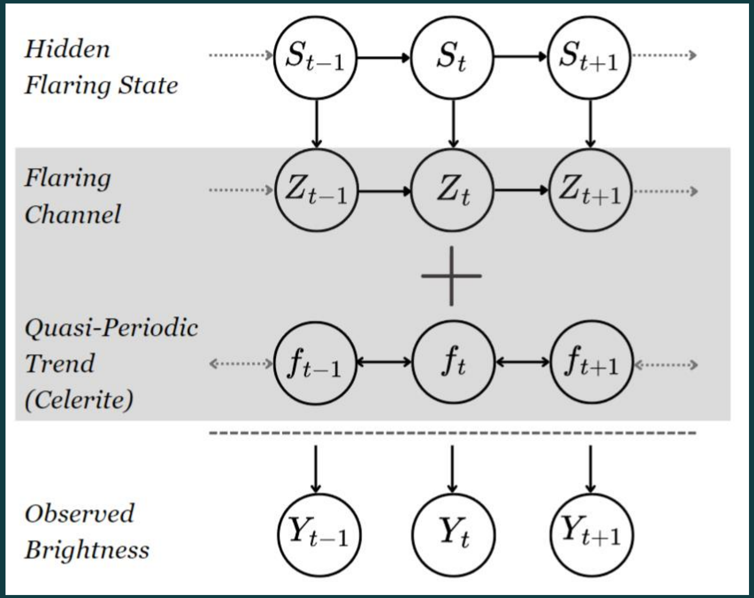
FINDING REAL FLARES

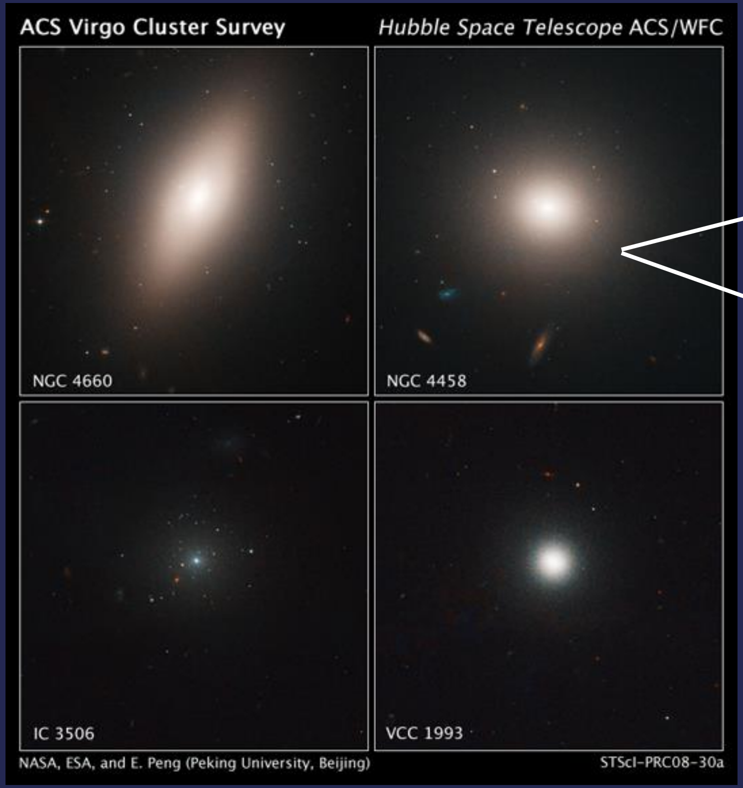


Esquivel et al 2024, submitted to ApJ

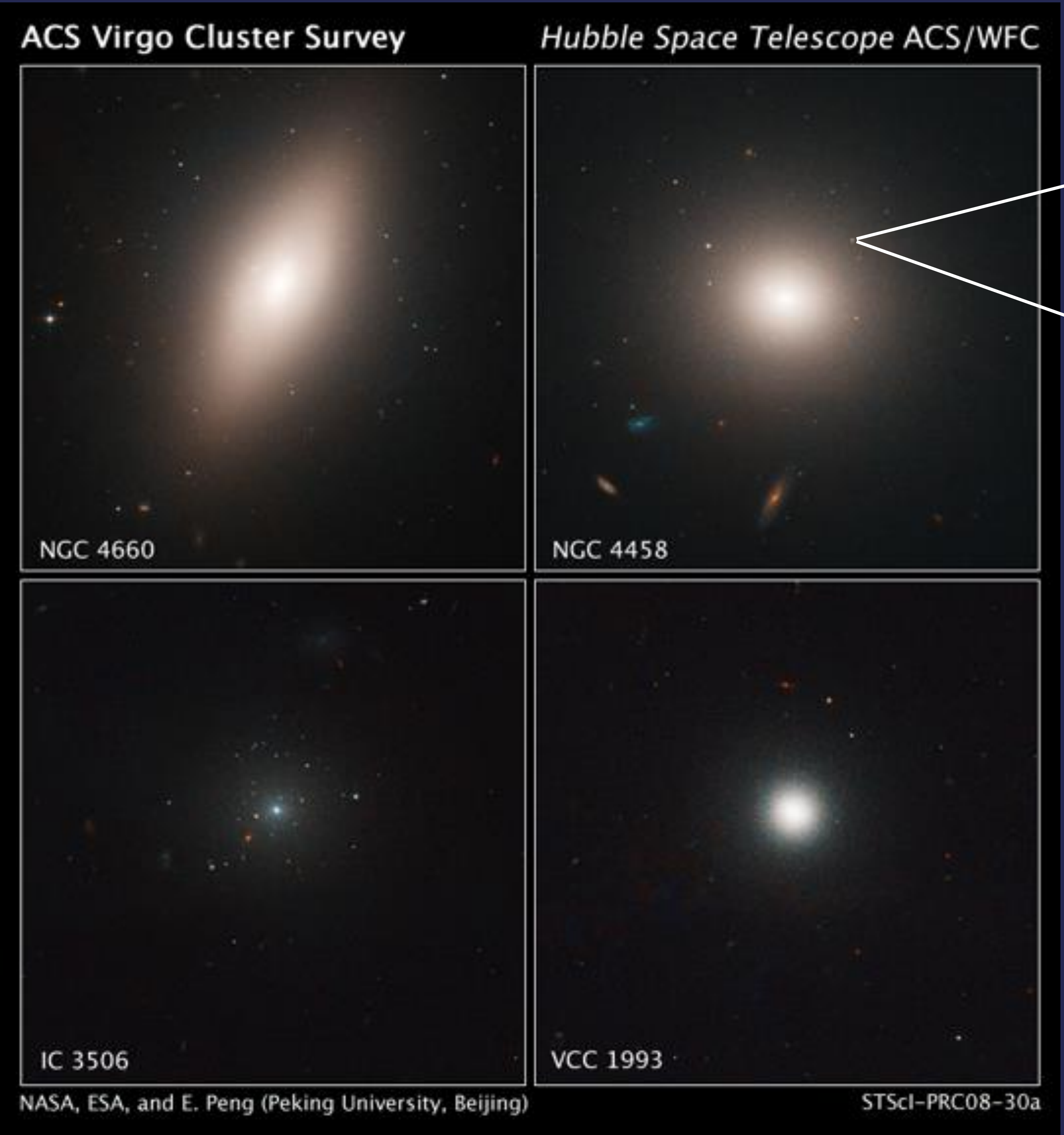
SUMMARY AND FUTURE WORK

- 3-state Hidden Markov Model to detect stellar flares
- At least as good as sigma-clipping
- Advantages:
  → Duration of flare events
  → Energy recovery
  → no need for iterative procedures, manual identification
  → improve population studies
- Disadvantages:
  → computationally expensive
- Speeding up GP portion of model (Rodrigo Barradas Herrera, stats PhD student)
- Application to population of M-dwarfs (Phil vanLane, astro PhD candidate)
- Gaia data release of time series





Studying the relationship between globular cluster populations and their host galaxies



Almost every galaxy in the universe has a population of *Globular Clusters* (GCs, or old star clusters)

https://www.nasa.gov/mission_pages/hubble/science/virgo_cluster.html

What we observe in the local universe:

Large galaxies have GCs, but some smaller (dwarf) galaxies do not.

In between these extremes, there is a transition.

Big Questions:

Why do some galaxies lack GCs?

What is a good empirical model for the distribution of galaxies with and without GCs?

What can this tell us about GC evolution in the context of galaxy formation and evolution?

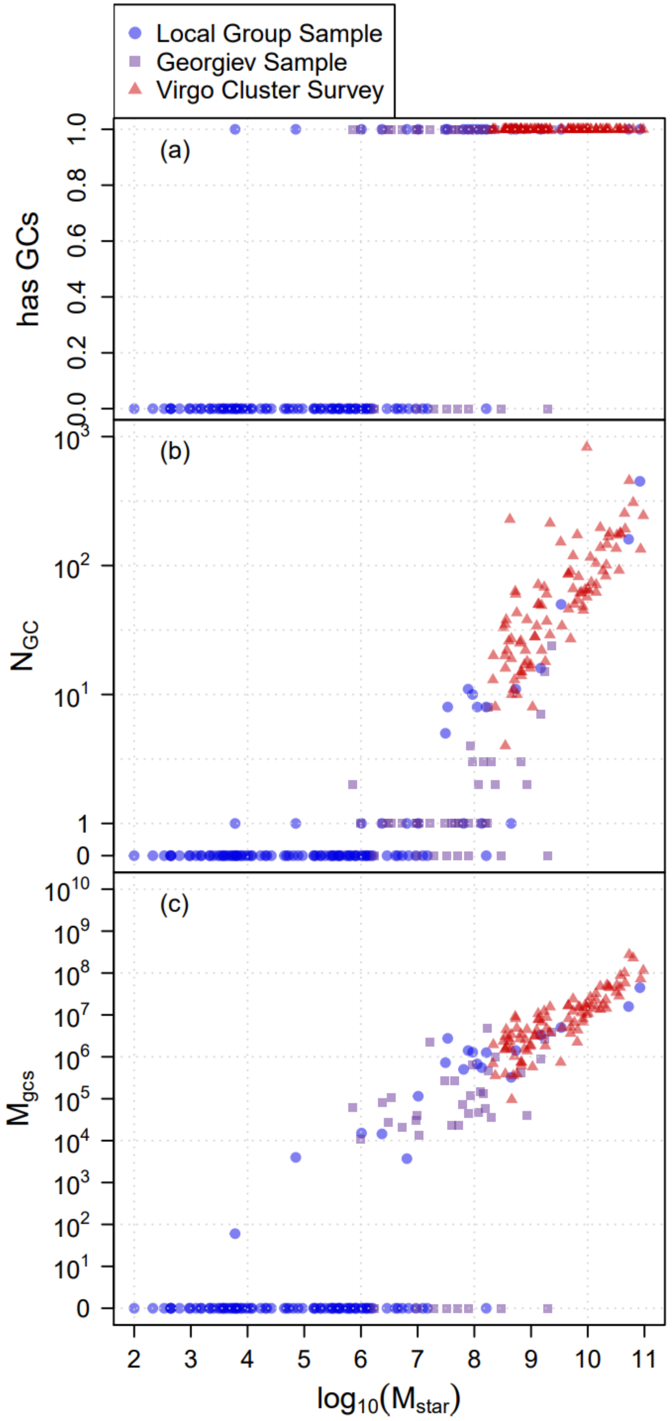
Data: GCs and host galaxy stellar mass

- Has GCs? = 1
- No GCs? = 0

Number of GCs

Mass in GCs

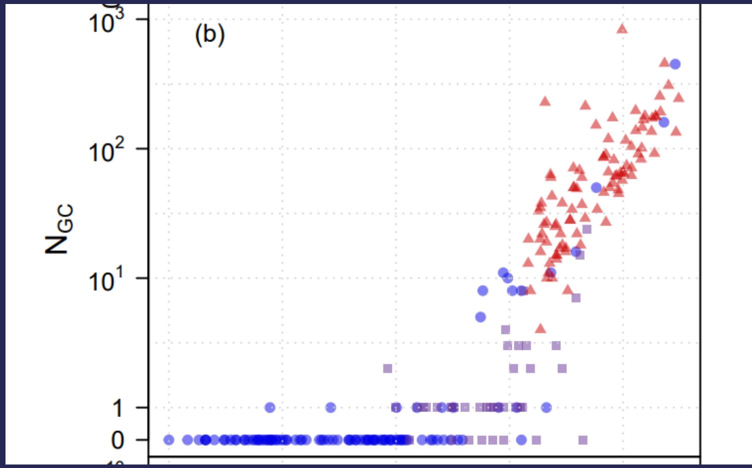
Eadie, Harris, and Springford (2022), ApJ 926 162



Empirical statistical models

What’s often done: Poisson Regression

Number of GCs



THE ASTROPHYSICAL JOURNAL, 972:104 (10pp), 2024 September 1 https://doi.org/10.3847/1538-4357/ad6147

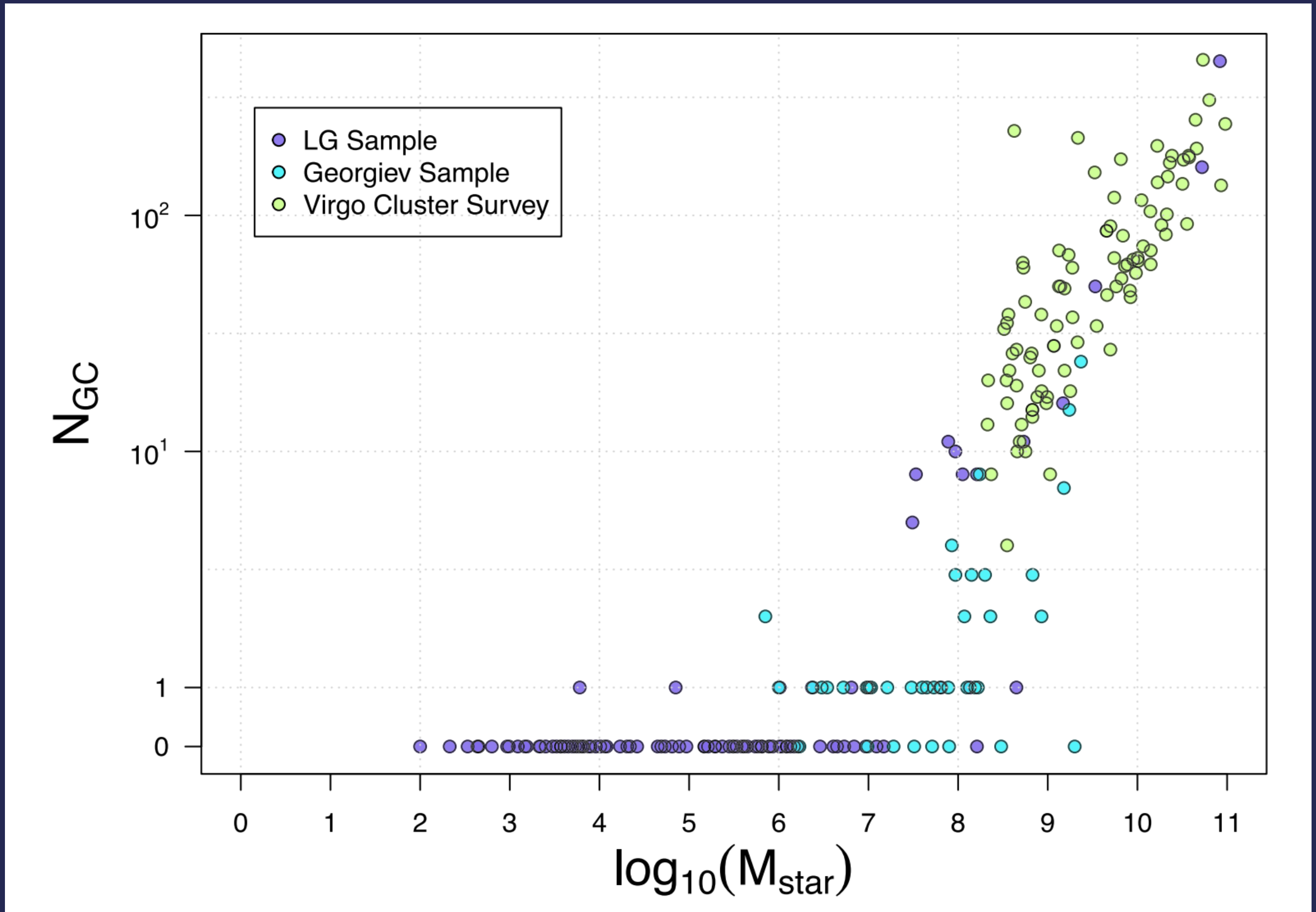
Should Zeros Count? Modeling the Galaxy–Globular Cluster Scaling Relation with(out) Zero-inflated Count Models

Samantha C. Berek[1,2,3,5], Gwendolyn M. Eadie[1,3,4], Joshua S. Speagle (沈佳士)[1,2,3,4], and Shu Yan Wang[4]
[1] David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada; sam.berek@mail.utoronto.ca
[2] Dunlap Institute for Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada
[3] Data Sciences Institute, University of Toronto, 17th Floor, Ontario Power Building, 700 University Avenue, Toronto, ON M5G 1Z5, Canada
[4] Department of Statistical Sciences, University of Toronto, 9th Floor, Ontario Power Building, 700 University Avenue, Toronto, ON M5G 1Z5, Canada
Received 2024 March 1; revised 2024 June 20; accepted 2024 June 28; published 2024 August 28

Investigation of count models for regression:

- Poisson regression
- Negative binomial
- Negative binomial with shape parameter
- *And their zero-inflated counterparts*

Sam Berek
PhD Candidate



Zero-inflated versions of the models, too:

$$f_{\mathrm{ZIP}}(y) = \begin{cases} \pi_0 + (1-\pi_0) f_{\mathrm{P}}(y=0|\lambda) & \text{if } y = 0, \\ (1-\pi_0) f_{\mathrm{P}}(y|\lambda) & \text{if } y > 0, \end{cases}$$

$$f_{\mathrm{ZINB}}(y) = \begin{cases} \pi_0 + (1-\pi_0) f_{\mathrm{NB}}(y=0|\lambda, \phi) & \text{if } y = 0, \\ (1-\pi_0) f_{\mathrm{NB}}(y|\lambda, \phi) & \text{if } y > 0, \end{cases}$$

Poisson model

$$E[Y] = \lambda,$$
$$\mathrm{Var}[Y] = \lambda$$

Negative binomial (NB) model

$$E[Y] = \lambda,$$
$$\mathrm{Var}[Y] = \lambda + \frac{\lambda^2}{\phi}.$$

NB model with shape parameter

$$E[Y] = \lambda,$$
$$\mathrm{Var}[Y] = \lambda + \frac{\lambda^2}{\phi}.$$

$$\ln(\phi_i) = \gamma_0 + \gamma_1 X_i$$

Berek, **Eadie,** Speagle, & Wang (2024), ApJ 972 104

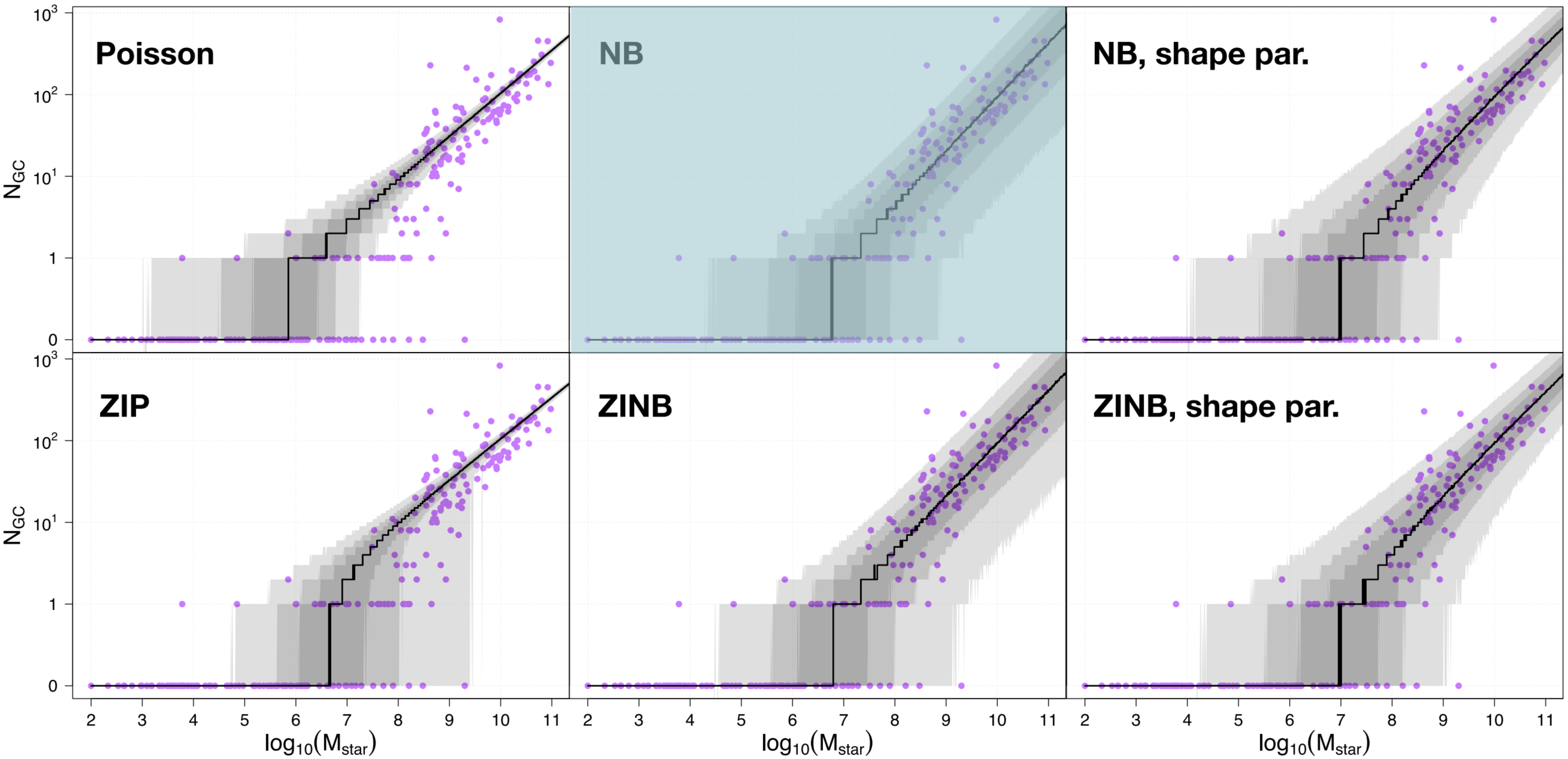


Figure 3. Predictive intervals for the six models. Shaded regions, from outer to inner, show the ranges in which 95%, 75%, and 50% of new data would be expected to lie. These regions should also encompass 95%, 75%, and 50% of the plotted galaxy data, if the given model is a good fit to the data.

Takeaways:

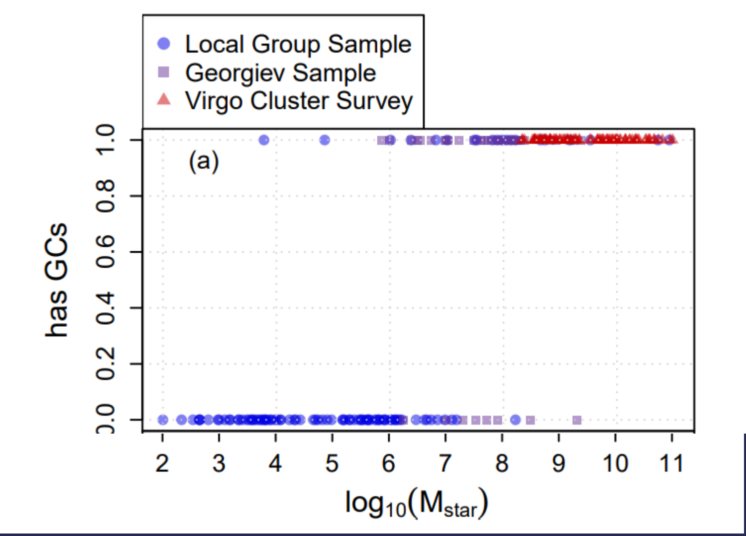
- Poisson regression is not a good model for GC counts as predicted by host galaxy stellar mass
- Negative binomial regression is much better
- Zero-inflation is not needed (at least in the parametric form we chose)
- Implications for simulations of galaxies with GC populations that inform other studies

Empirical statistical models

Logistic Regression

Has GCs? = 1

No GCs? = 0



Eadie, Harris, and Springford (2022), ApJ 926 162

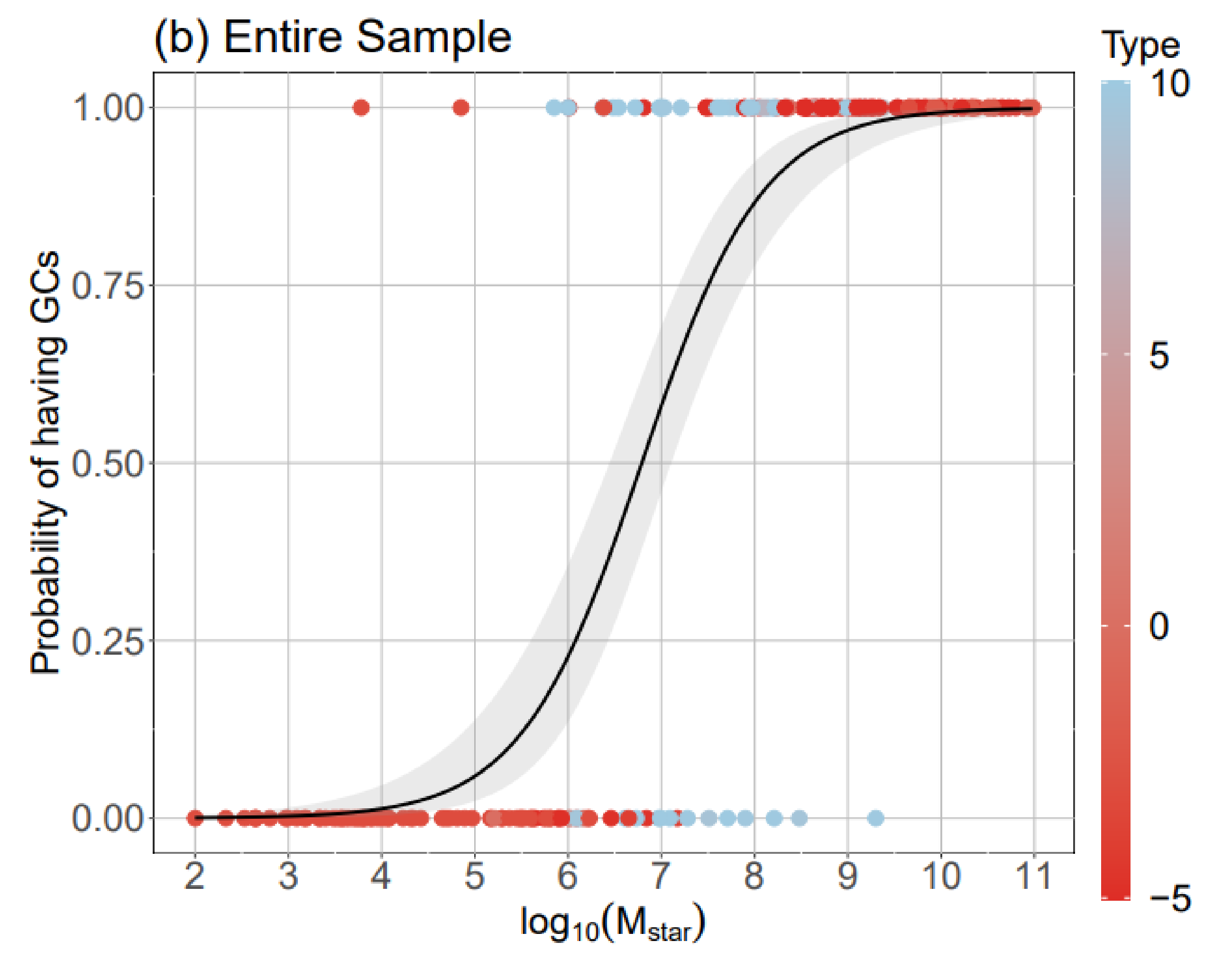
Logistic regression

$$\boldsymbol{p} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \log M_*)}}$$

$$\widehat{\beta_0} = -10.50\ (-13.33, -7.67)$$

$$\widehat{\beta_1} = 1.55\ (1.15, 1.94)$$

(brackets are 95% confidence intervals)

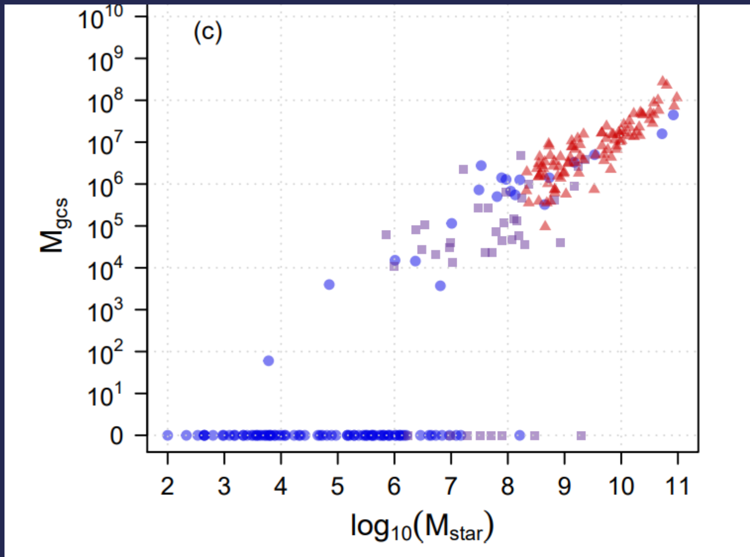


Eadie, Harris, and Springford (2022), ApJ 926 162

Empirical statistical models

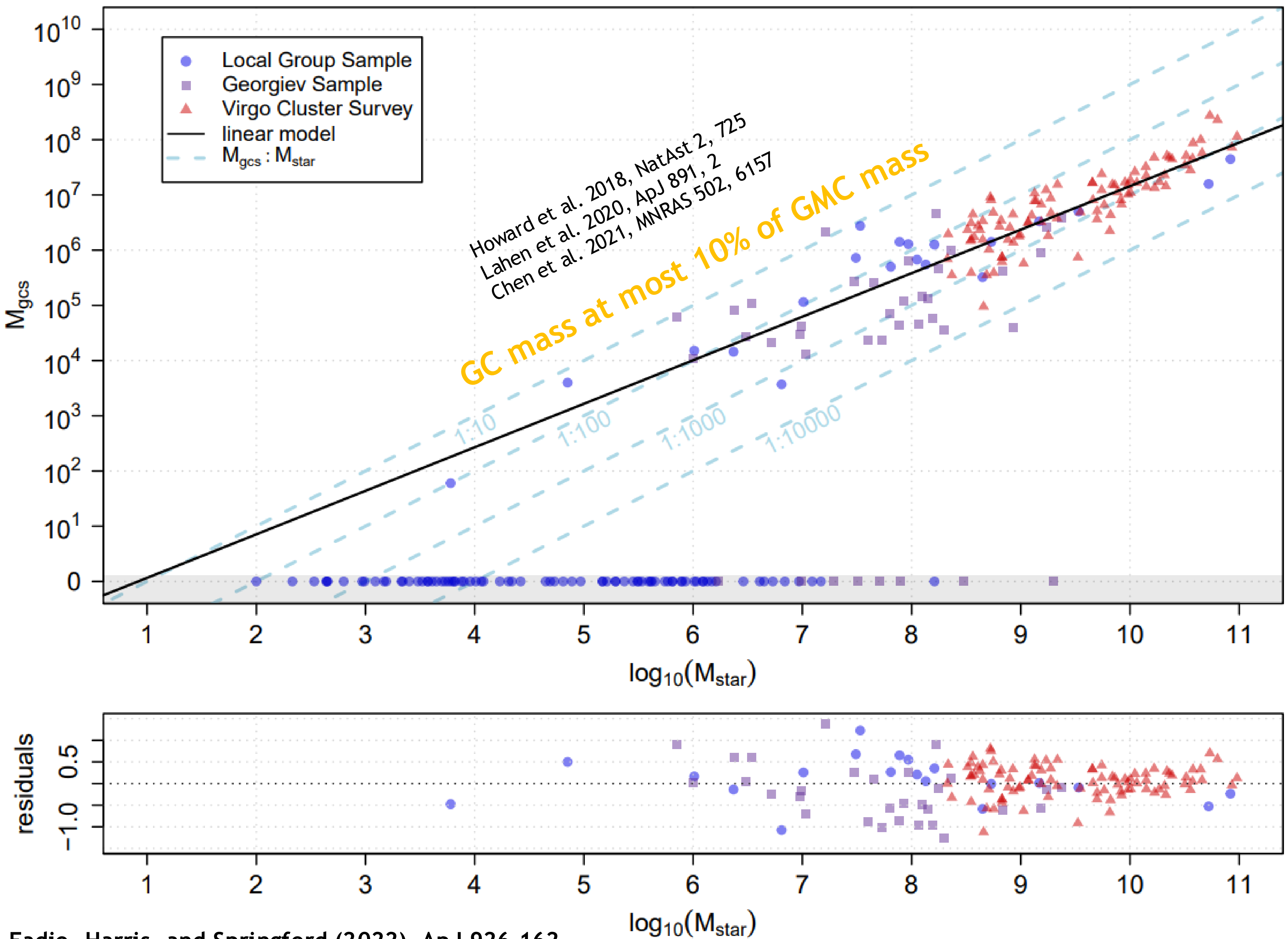
What's often done: Linear Regression

Mass in GCs



Eadie, Harris, and Springford (2022), ApJ 926 162

Linear Regression



Eadie, Harris, and Springford (2022), ApJ 926 162

Logistic Regression or Linear Regression?

- Covariates or independent variables are:
  - Galaxy stellar mass
- Response variable or dependent variable is
  - **Whether or not the galaxy has a GC population** → binary variable (0 or 1)
  - The **mass of the globular cluster population** → continuous variable

Which response variable should we choose?

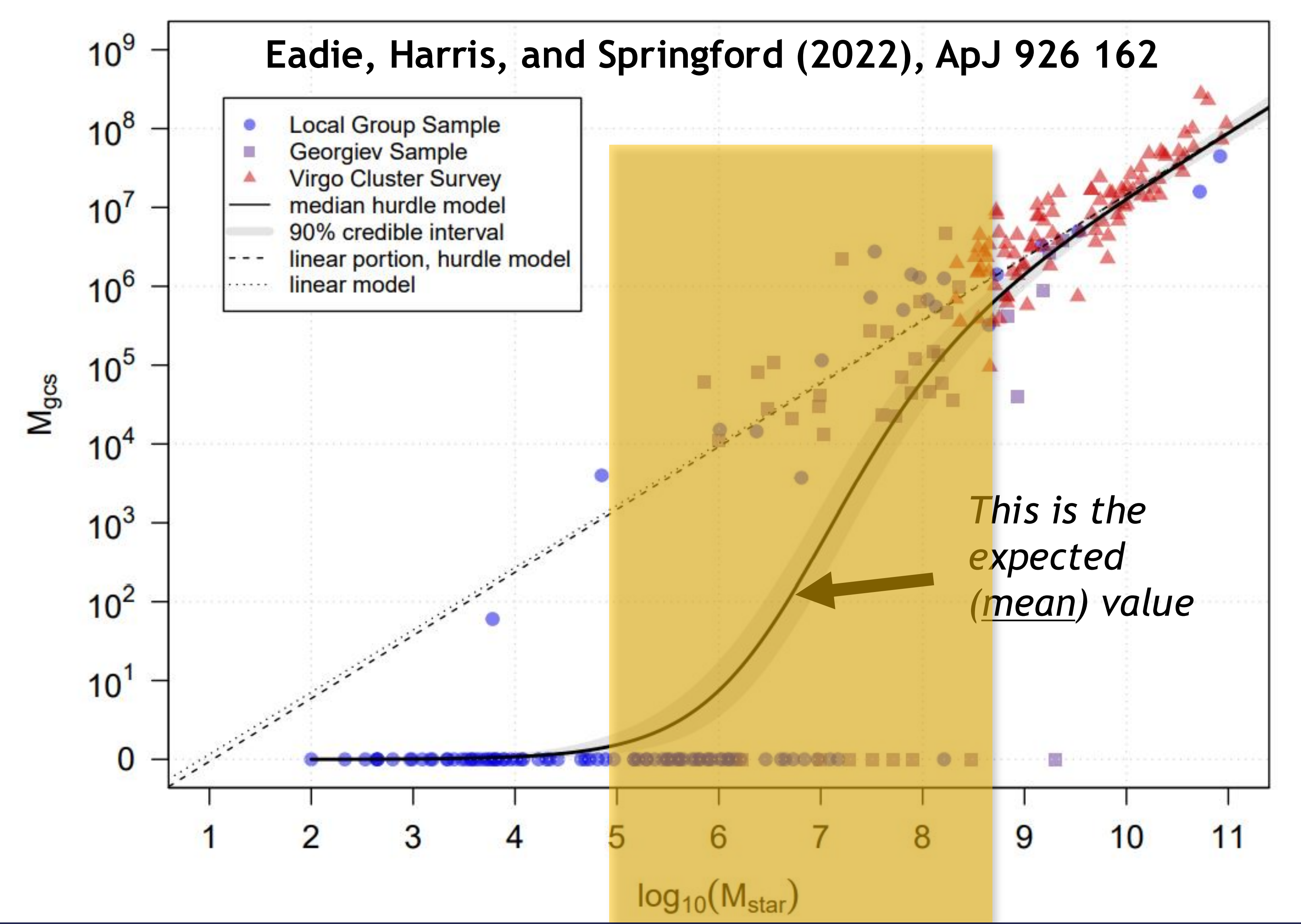
Hurdle Models

- another example of *generalized linear models* (GLMs)
- deal with **data containing significant numbers of zeros** (used routinely in e.g., applied statistics, biostatistics, ecology, environ. science)
- *Side note: GLMs are underused in astronomy but have shown promise (see De Souza et al 2015, Elliot et al 2015, Hattab et al 2019, Lenz et al 2016, see also textbook by Hilbe et al. 2017)*

Lognormal Hurdle Model

$$E\left[\log M_{gcs}\right] = \left(\frac{1}{1 + e^{-(\beta_0 + \beta_1 \log M_*)}}\right)(\gamma_0 + \gamma_1 \log M_*)$$

- use a Bayesian framework for parameter inference
- brms package *Bayesian Regression Models using Stan* (R Core Team 2020)
- Try both the default priors in brms (flat, improper priors) and diffuse priors → little difference



Transition region

If we observe a galaxy and see that is has a GC population, then we can predict its M_{gcs}!

If we observe a galaxy with some M_*, then we can get the probability of it having a GC population.

Transition region

- What we observe:
  - Galaxies with $M_* < 10^6 M_{solar}$ have no GCs today

- What we suspect:
  - maximum mass in GCs (from simulations) → 10% of galaxy M_*
  - If true, galaxies with $M_* < 10^6 M_{solar}$ would have GC population of $\sim 10^5 M_{solar}$
  - A GC with $\sim 20{,}000 M_{solar}$ has trouble surviving

- What is almost certainly true:
  - Including the zeros is important!
  - Initial GC mass, GCs initial location and orbit, structure of host galaxy and **dark matter halo**

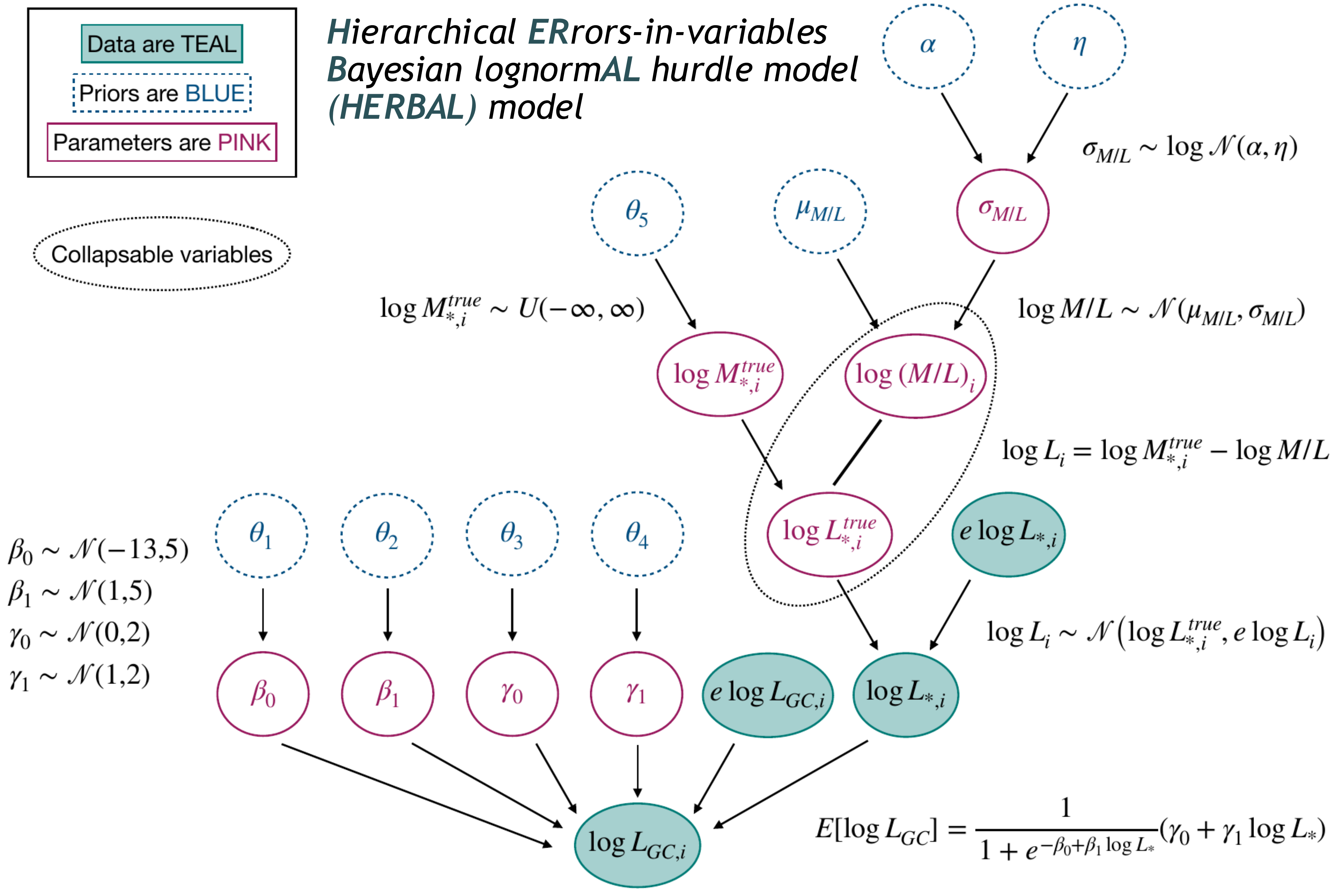
Hierarchical ERrors-in-variables BAyesian Lognormal hurdle model (HERBAL)

1. Include measurement uncertainties via an **errors-in-variables**
2. Incorporate a random effect term for unaccounted-for variance
3. Instead of stellar mass, use luminosity → estimate the mass-to-light ratio for each galaxy

Sam Berek
PhD Candidate

Berek, S. C., **Eadie, G. M.**, Speagle, J. S., & Harris, W. E. (2023), ApJ 955(1), 22.

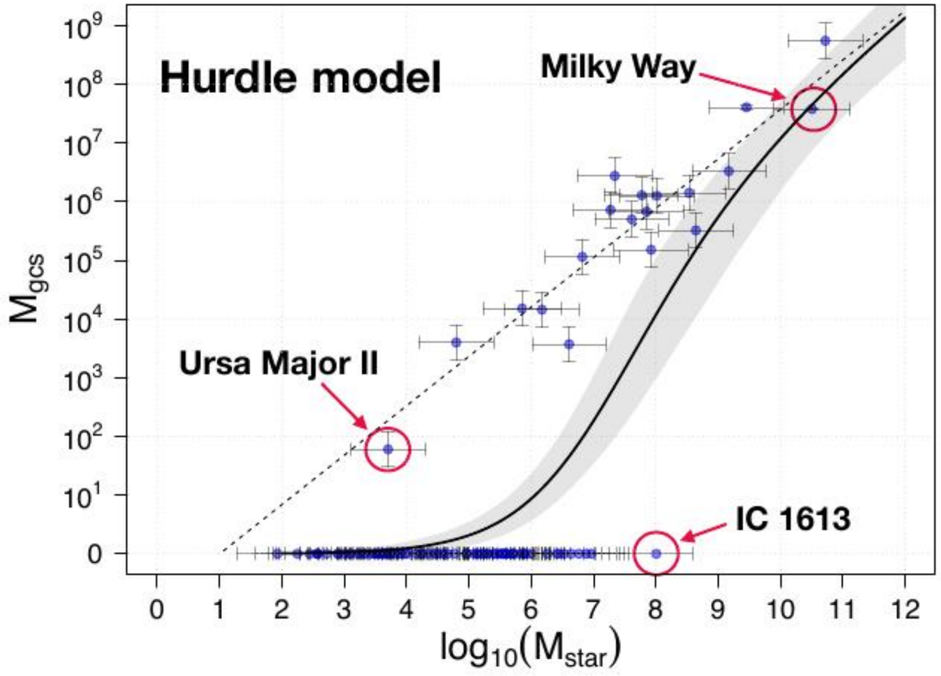
*Hierarchical **ER**rors-in-variables Bayesian lognorm**AL** hurdle model (**HERBAL**) model*

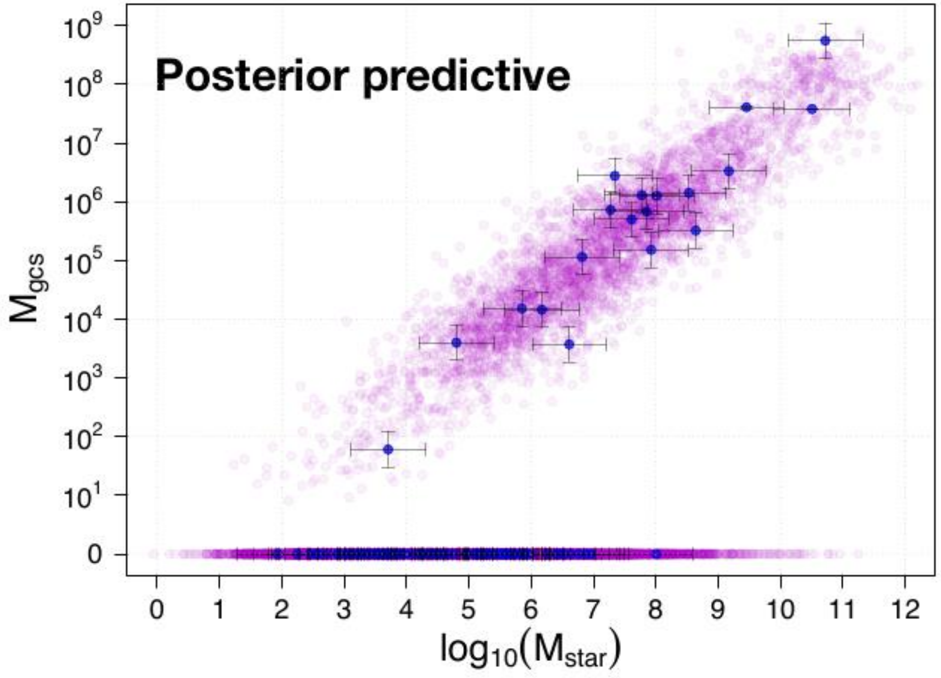


Sam Berek
PhD Candidate

Berek, **Eadie**, Speagle, & Harris (2023), ApJ 955(1), 22.

HERBAL model





Sam Berek
PhD Candidate

Good description of the data

Two points of interest: IC1613 & Ursa Major II

Even larger galaxies without GCs may exist

Berek, **Eadie,** Speagle, & Harris (2023), ApJ 955(1), 22.

GCs and their host galaxies

Summary:

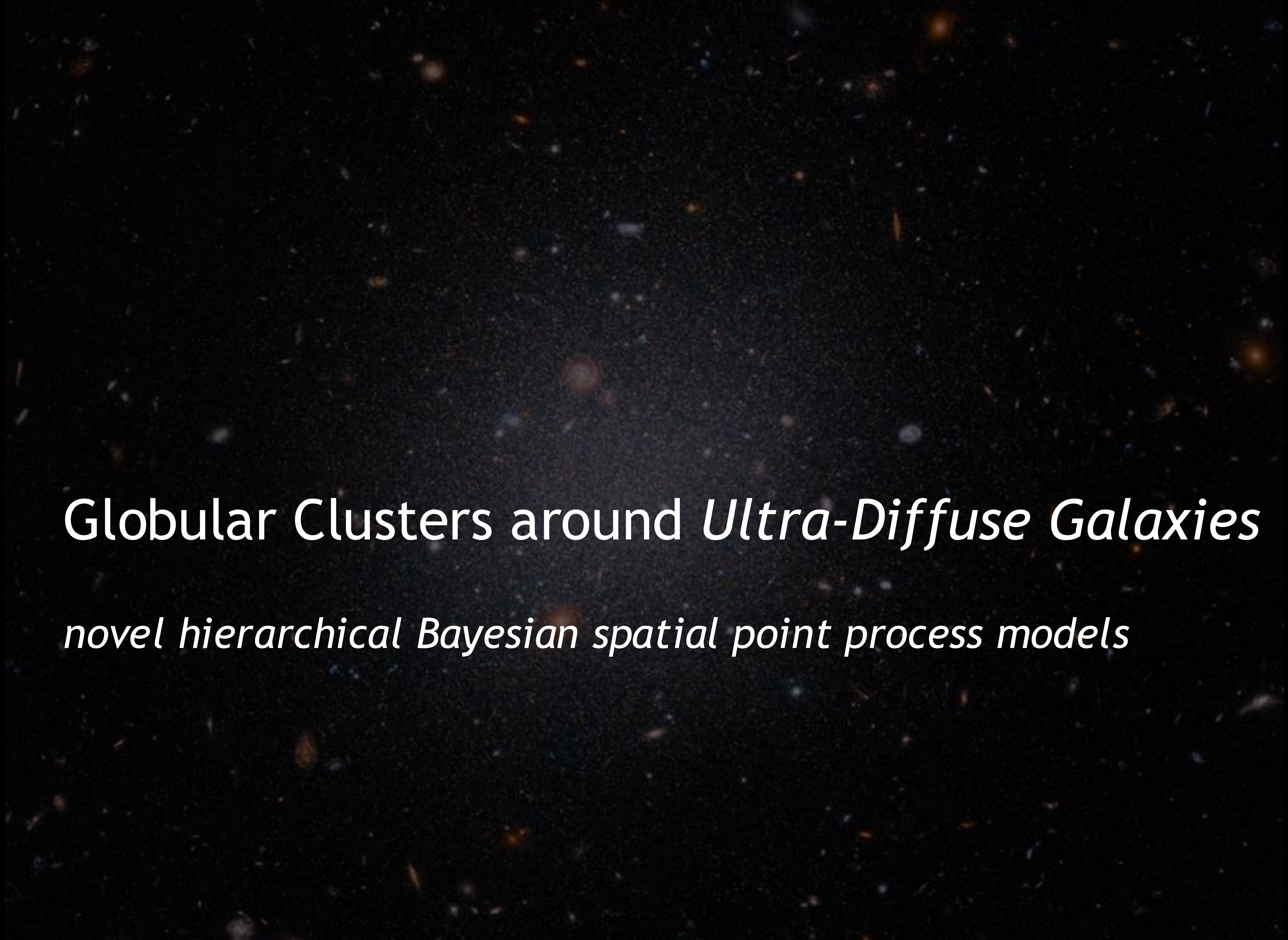
- Including zeros is important
- Potential information about formation and evolution history

Future work:

- Incorporating selection effects for galaxies beyond the Local Group

- Using both counts and GC population mass simultaneously

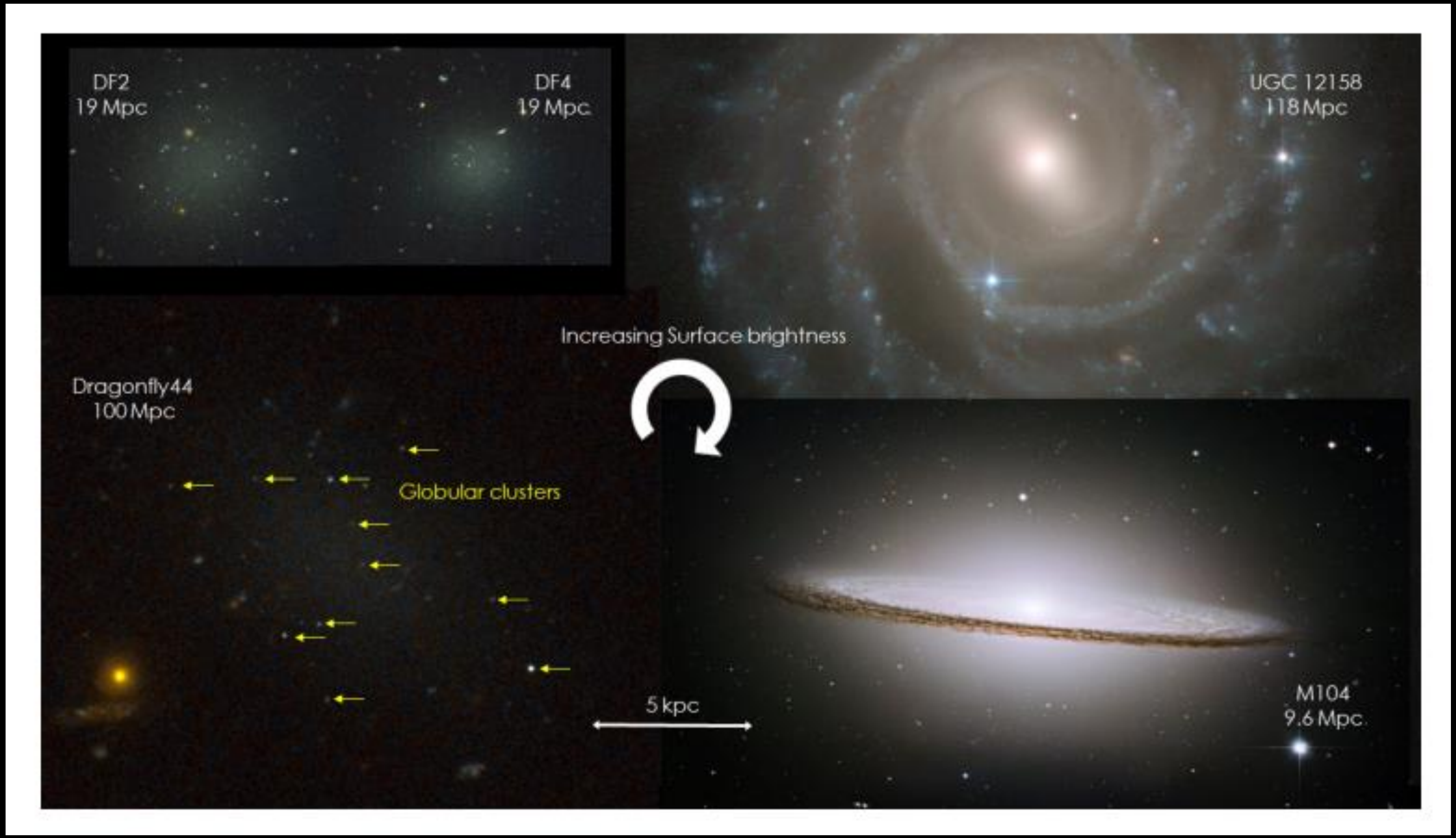
- Understanding GC populations in the early universe → JWST



Globular Clusters around *Ultra-Diffuse Galaxies*

novel hierarchical Bayesian spatial point process models

Ultra-Diffuse Galaxies (UDGs)



→ Can be as big as our galaxy, but way fewer stars

→ Very low surface brightness (brightness per area)

→ They have globular clusters!

Big Questions about UDGs

- How did these galaxies form?
- Why are they so big but have so few stars?
- How did their GCs form?
- Are their GC populations different from other galaxies?
- What can this tell us about galactic formation in the early universe?
- Why are some dark matter dominated, and others lacking dark matter?

Dragonfly 2 (DF2), image taken by Hubble Space Telescope. Credit: NASA, ESA, STScI, Zili Shen (Yale), Pieter van Dokkum (Yale), Shany Danieli (IAS) IMAGE PROCESSING: Alyssa Pagan (STScI) https://science.nasa.gov/missions/hubble/mystery-of-galaxys-missing-dark-matter-deepens/

Big *Challenges* with UDGs

- They are difficult to find (low surface brightness)
- There is significant measurement uncertainty in their GC populations
- They are sometimes found in clusters of galaxies → which GCs belong to UDGs and which belong to “regular” galaxies?

Discovering UDGs with Spatial Statistics

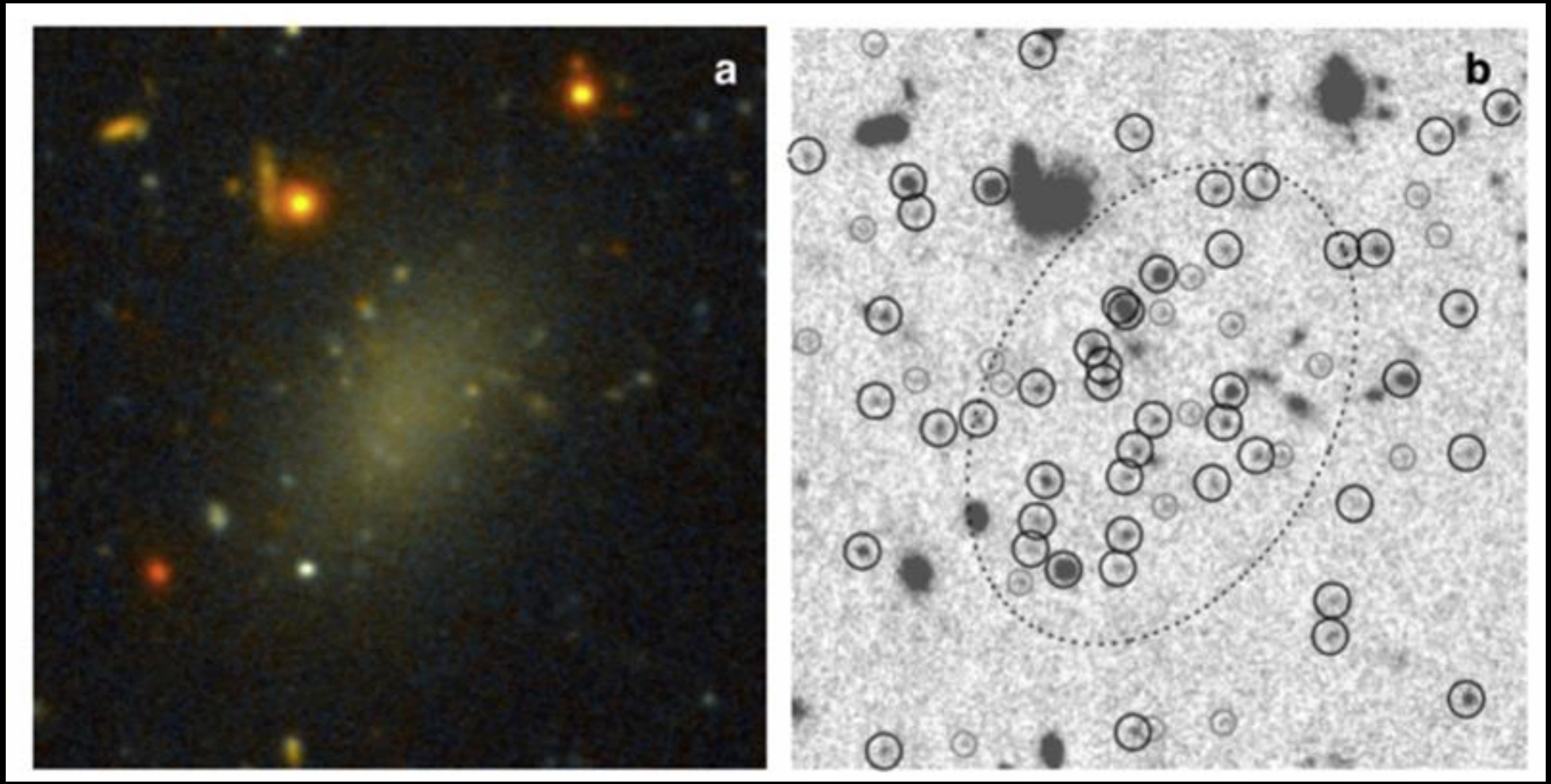


Figure from Van Dokkum et al (2016), ApJL, 828:L6

Dayi Li
PhD Candidate
Dept. of Statistical Sciences

The Astrophysical Journal, 935:3 (28pp), 2022 August 10 https://doi.org/10.3847/1538-4357/ac7b22

OPEN ACCESS

Light from the Darkness: Detecting Ultra-diffuse Galaxies in the Perseus Cluster through Over-densities of Globular Clusters with a Log-Gaussian Cox Process

Dayi David Li (李大一)[1], Gwendolyn M. Eadie[1,2], Roberto Abraham[2,3], Patrick E. Brown[1,4], William E. Harris[5], Steven R. Janssens[2,6], Aaron J. Romanowsky[7,8,9], Pieter van Dokkum[10], and Shany Danieli[11,12]

David Li
PhD Candidate
(statistics)

Log-Gaussian Cox Process (LGCP):

- (inhomogeneous) Poisson point process with some intensity $\lambda(s)$
- $\lambda(s)$ is generated by a Log-Gaussian random field → provides means for spatial correlation

Li, Eadie, et al (2022) ApJ 935:3

LGCP Model to discover new UDGs

$$
\begin{aligned}
X &\sim \mathrm{IPP}(\lambda(s)),\\
\log(\lambda(s)) &= X(s)'\beta + \mathcal{U}(s),\\
\mathcal{U}(s) &\sim \mathcal{GP}(\mathbf{0}, \sigma^2 C(\cdot)),\\
\mathrm{Cov}(\mathcal{U}(s), \mathcal{U}(t)) &= \sigma^2 C(|s - t|),\ s,\ t \in S.
\end{aligned}
$$

Inhomogeneous Poisson process

Log(intensity) = Fixed effect + Random effect

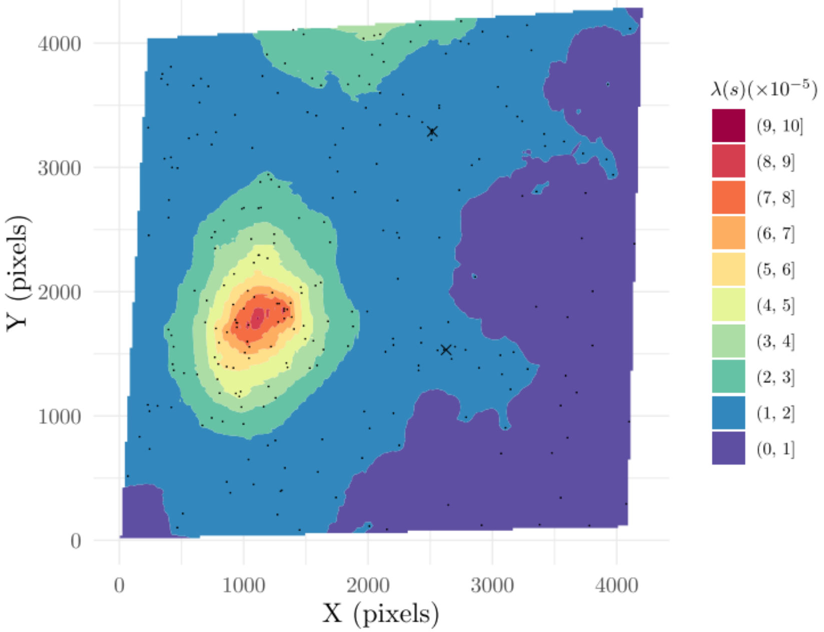
Randomness characterized by GP with some covariance structure

In a galaxy cluster, GCs can be part of
- A bright "normal" looking galaxy
- The intergalactic medium
- An ultra-diffuse galaxy

Using a Bayesian approach and INLA, estimate posterior distribution

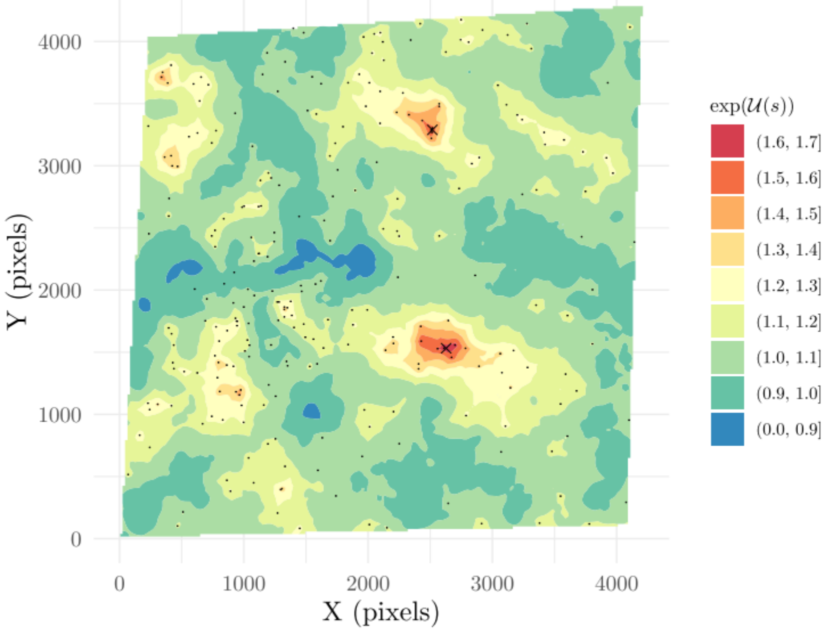
Finally, find regions in posterior of $U(s)$ that exceed some threshold

Intensity field $\lambda(s)$

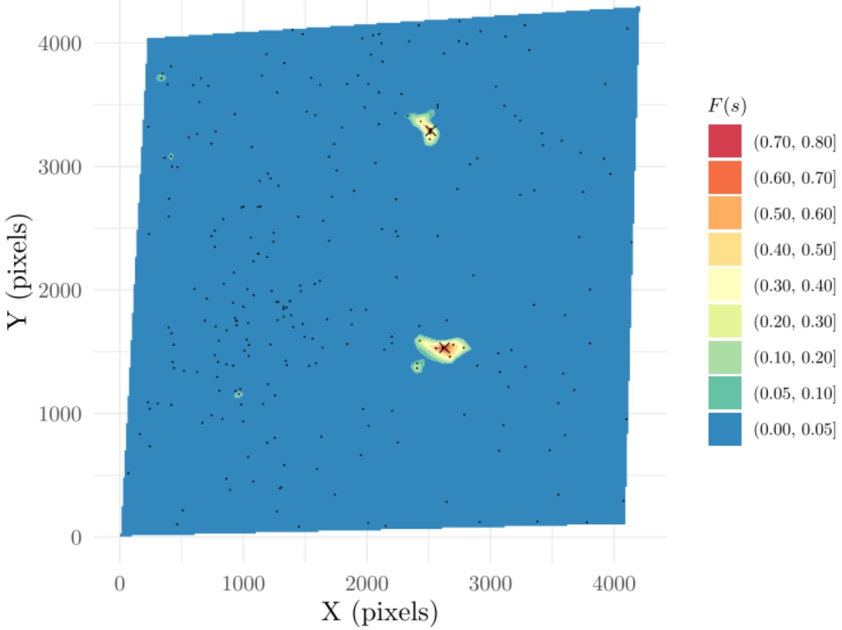


(a)

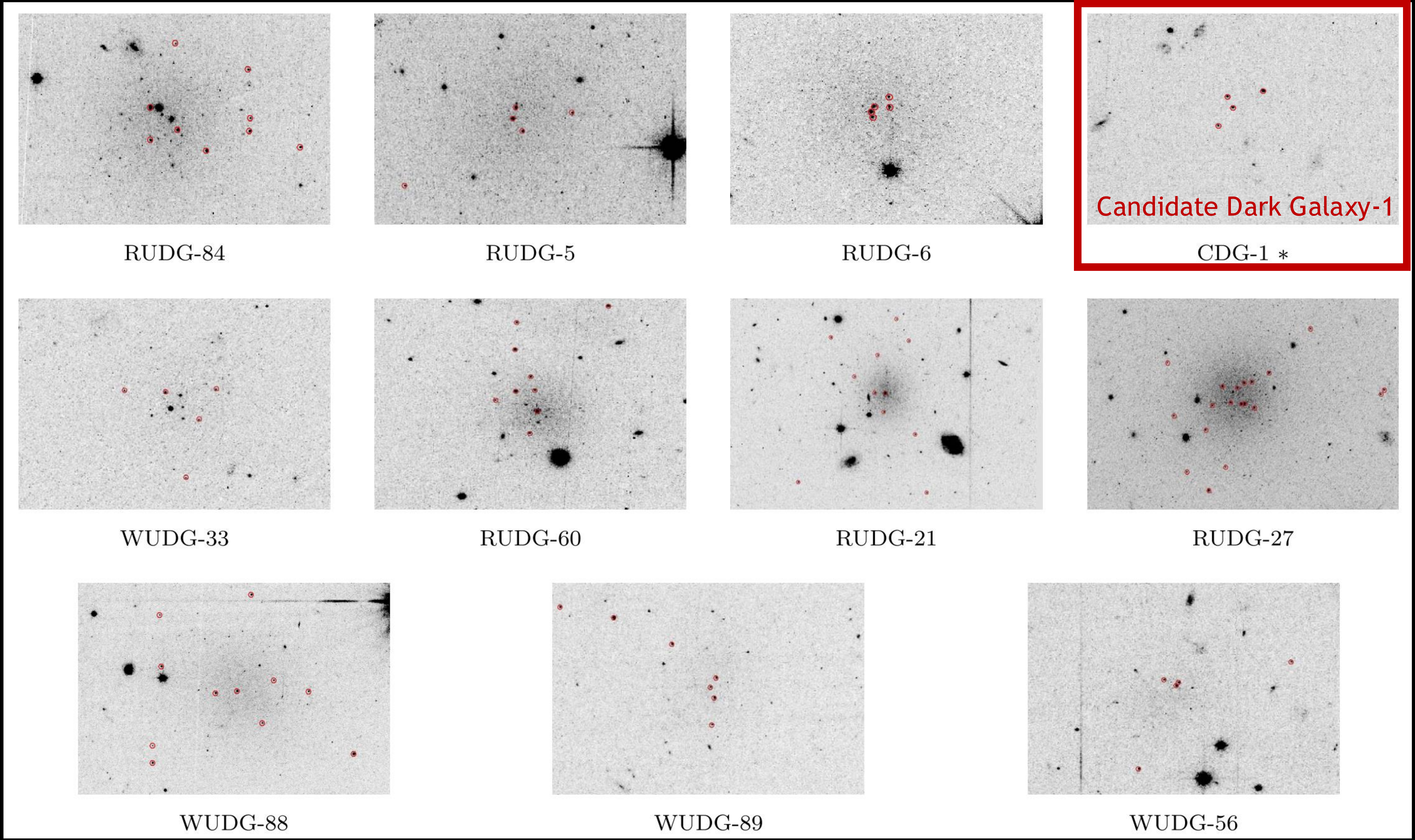
Random field $\exp(U(s))$



Detection of potential UDGs $F(s)$



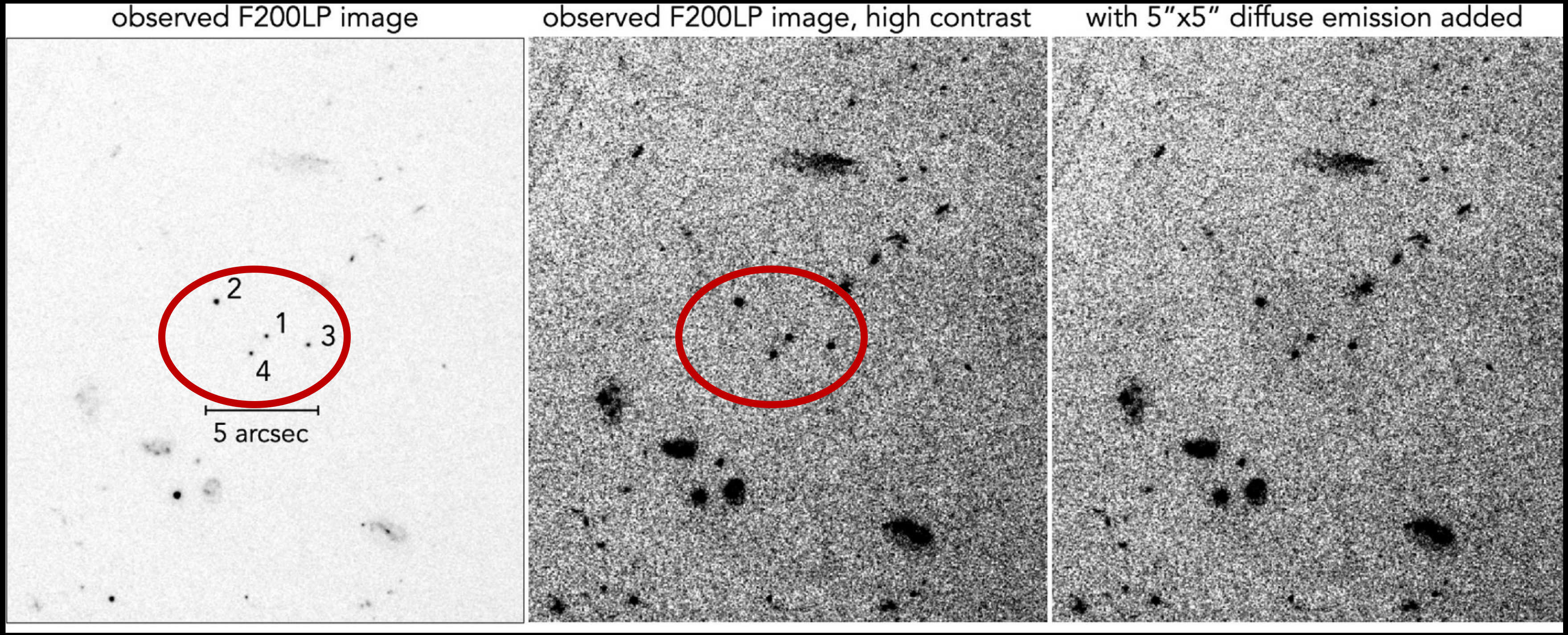
Li, Eadie, et al (2022) ApJ 935:3



David Li
PhD Candidate
(statistics)

Li, Eadie, et al (2022) ApJ 935:3

Observe CDG-1 with the Hubble Space Telescope!



Pieter van Dokkum et al (incl. **Eadie**) 2024
Res. Notes AAS 8:135

No diffuse emission is detected…

→ random alignment of GCs? (seems very improbable)

→ Not *actually* GCs? (also unlikely)

→ The "darkest" UDG detected to date?!

POISSON CLUSTER PROCESS MODELS FOR DETECTING ULTRA-DIFFUSE GALAXIES

accepted to AoAS!

BY DAYI LI[1,a], ALEX STRINGER[2,d], PATRICK E. BROWN[1,b], GWENDOLYN M. EADIE[1,3,c] AND ROBERTO G. ABRAHAM[3,e]

[1]Department of Statistical Sciences, University of Toronto , [a]dayi.li@mail.utoronto.ca , [b]patrick.brown@utoronto.ca , [c]gwen.eadie@utoronto.ca

[2]Department of Statistics and Actuarial Science, University of Waterloo , [d]alex.stringer@uwaterloo.ca

[3]David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto , [e]roberto.abraham@utoronto.ca

Include "marks" for the points (other covariates about GCS)

Include physical theories

David Li
PhD Candidate
(statistics)

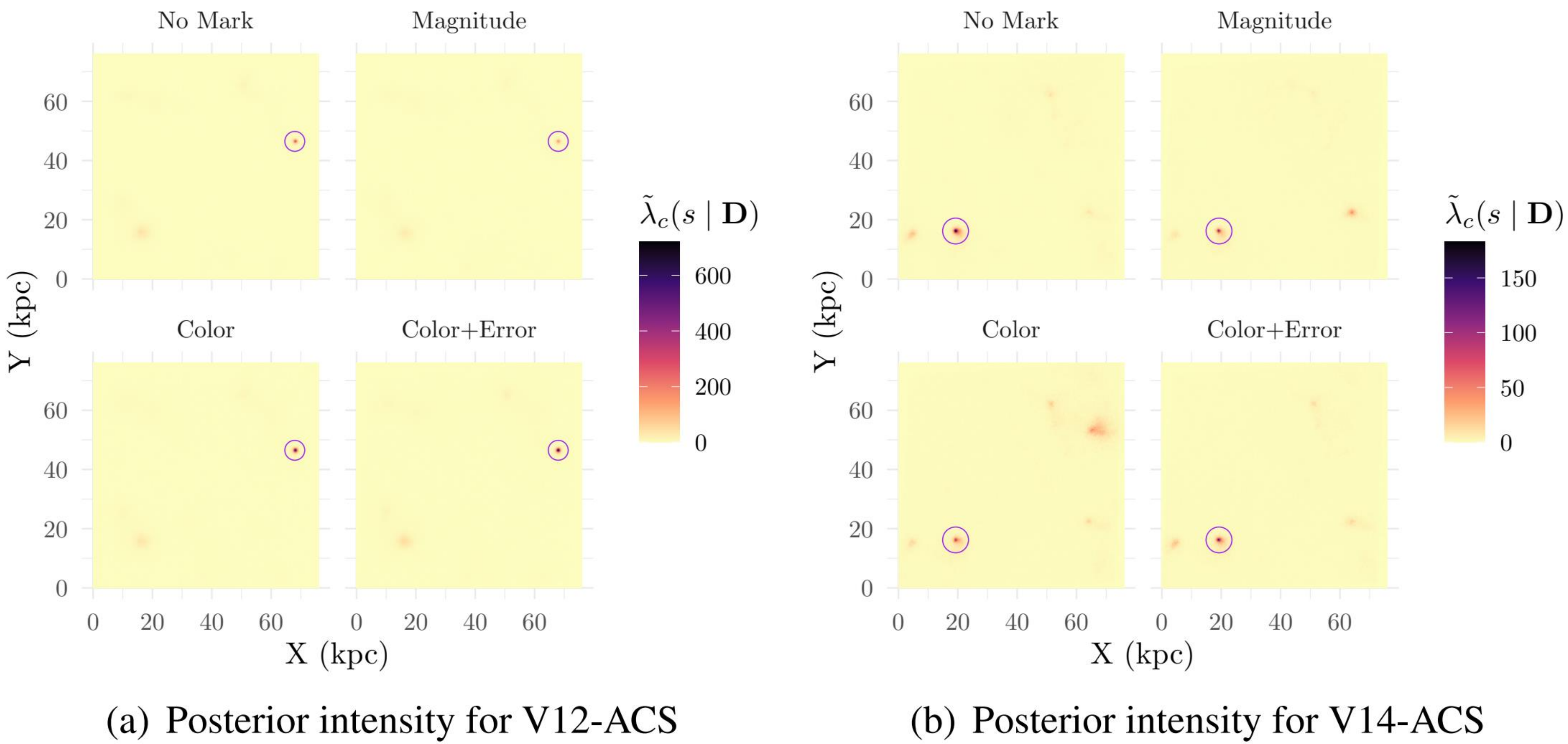
$$\mathbf{X}_c \sim \mathrm{HPP}(\lambda_c),$$

$$\Lambda(s) \mid \mathbf{X}_c, \boldsymbol{\Xi} = \beta_0 + \sum_{k=1}^{N_G} \text{Sérsic}(s; c_k^g, \boldsymbol{G}_k) + \sum_{j=1}^{N_U} \text{Sérsic}(s; \mathbf{c}_j^u, \boldsymbol{U}_j),$$

$$\mathbf{M}(\mathbf{x}_i) \mid \mathbf{x}_i, \mathbf{X}_c, \boldsymbol{\Xi}, \boldsymbol{\Sigma} \sim \sum_{m=0}^{N_U} p_{im} \mathcal{M}_{\Sigma_m},$$

$$p_{i0} = 1 - \Lambda_U(\mathbf{x}_i)/\Lambda(\mathbf{x}_i),\ p_{im} = \text{Sérsic}(\mathbf{x}_i; \mathbf{c}_m^u, \boldsymbol{U}_m)/\Lambda(\mathbf{x}_i),\ m \in [N_U]$$

$$\mathbf{X} \mid \Lambda \sim \mathrm{IPP}\{\Lambda(s)\}.$$



Candidate Dark Galaxy-2

Fig 3: Posterior results for our models in (a) V12-ACS and (b) V14-ACS. Circles indicate the CDG-2 location.

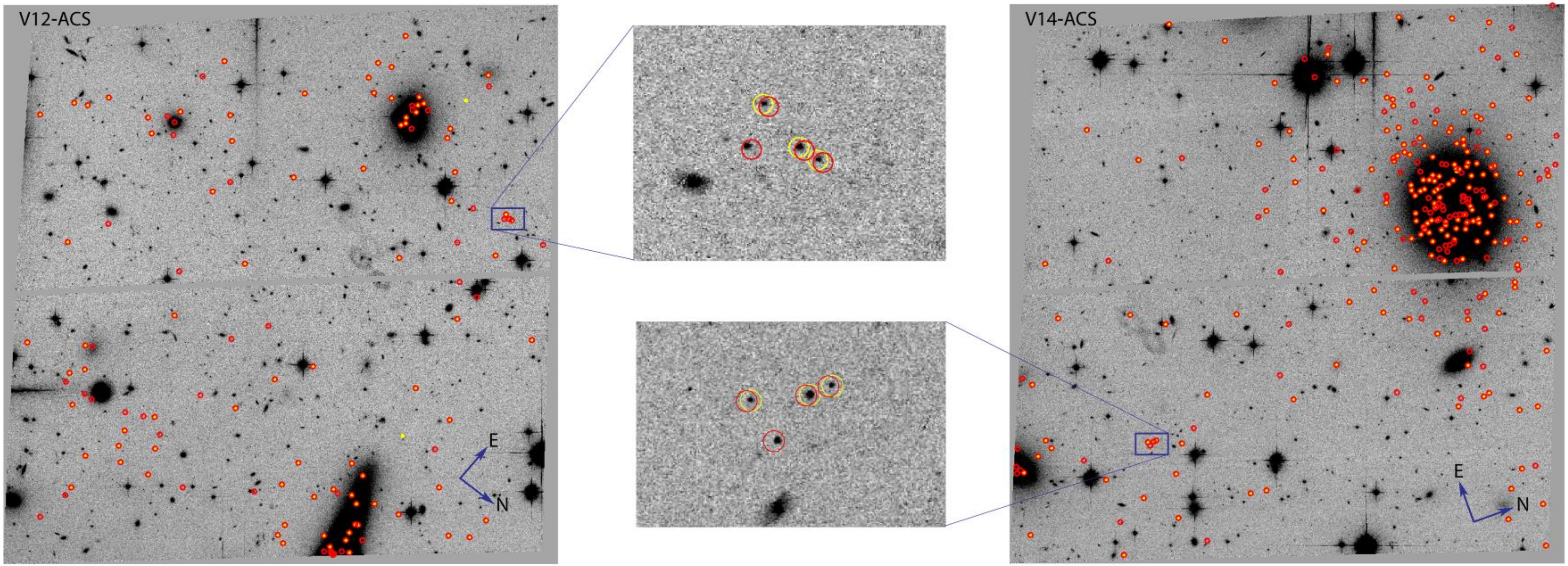


Figure 1. Spatial distributions of GC candidates in the F814W images V12-ACS (left) and V14-ACS (right). Red circles are GC candidates from DOLPHOT while yellow ones are from DAOPHOT. The GC candidates that constitute CDG-2 from the two images are enlarged in the middle.

Li, et al (incl. **Eadie**) 2024, *in prep*

- UDG detection is hampered by:
    - Photometric measurement uncertainties
    - GC membership uncertainty
    - Assumptions about underlying distribution of GCs
- *MATHPOP*
    - Novel statistical method to infer the globular cluster counts around UDGs and low surface brightness galaxies
    - Takes into account uncertainties and assumptions above

https://davidolohowski.github.io/MATHPOP/

DRAFT VERSION SEPTEMBER 16, 2024
Typeset using LATEX **twocolumn** style in AASTeX631

Submitted to ApJ

Discovery of Two Ultra-Diffuse Galaxies with Unusually Bright Globular Cluster Luminosity Functions via a Mark-Dependently Thinned Point Process (MATHPOP)

DAYI (DAVID) LI (李大一)[1,2,*] GWENDOLYN M. EADIE[1,2,3] PATRICK E. BROWN[1,4] WILLIAM E. HARRIS[5] ROBERTO G. ABRAHAM[3,6] PIETER VAN DOKKUM[7] STEVEN R. JANSSENS[8] SAMANTHA C. BEREK[2,3,6,†] SHANY DANIELI[9,‡] AARON J. ROMANOWSKY[10,11] AND JOSHUA S. SPEAGLE (沈佳士)[1,2,3,6]

https://davidolohowski.github.io/MATHPOP/

Li, **Eadie**, et al (2024, submitted to ApJ)

MATHPOP

Bayesian *MA*rk-dependently *TH*inned *PO*int *P*rocess

- Jointly model spatial distribution and brightness distribution
- Marks → brightness (magnitudes) of the GCs
- inhomogeneous Poisson Process
- Thinned process to account for observational and physical effects → GCs which are too faint to observe, or too far away to be part of galaxies
- Includes uncertainty in GC identification

https://davidolohowski.github.io/MATHPOP/

Li, **Eadie**, et al (2024, submitted to ApJ)

Bayesian *MA*rk-dependently *TH*inned *PO*int *P*rocess

Advantages of MATHPOP over traditional methods:

- Model difference sources of GCs (intergalactic medium, large galaxies, ultra diffuse galaxies)

- Includes the measurement uncertainty in brightness

- Includes the colour information and uncertainty in GC identification --> probabilistic catalogue of GCs

- No unphysical, negative counts of GCs

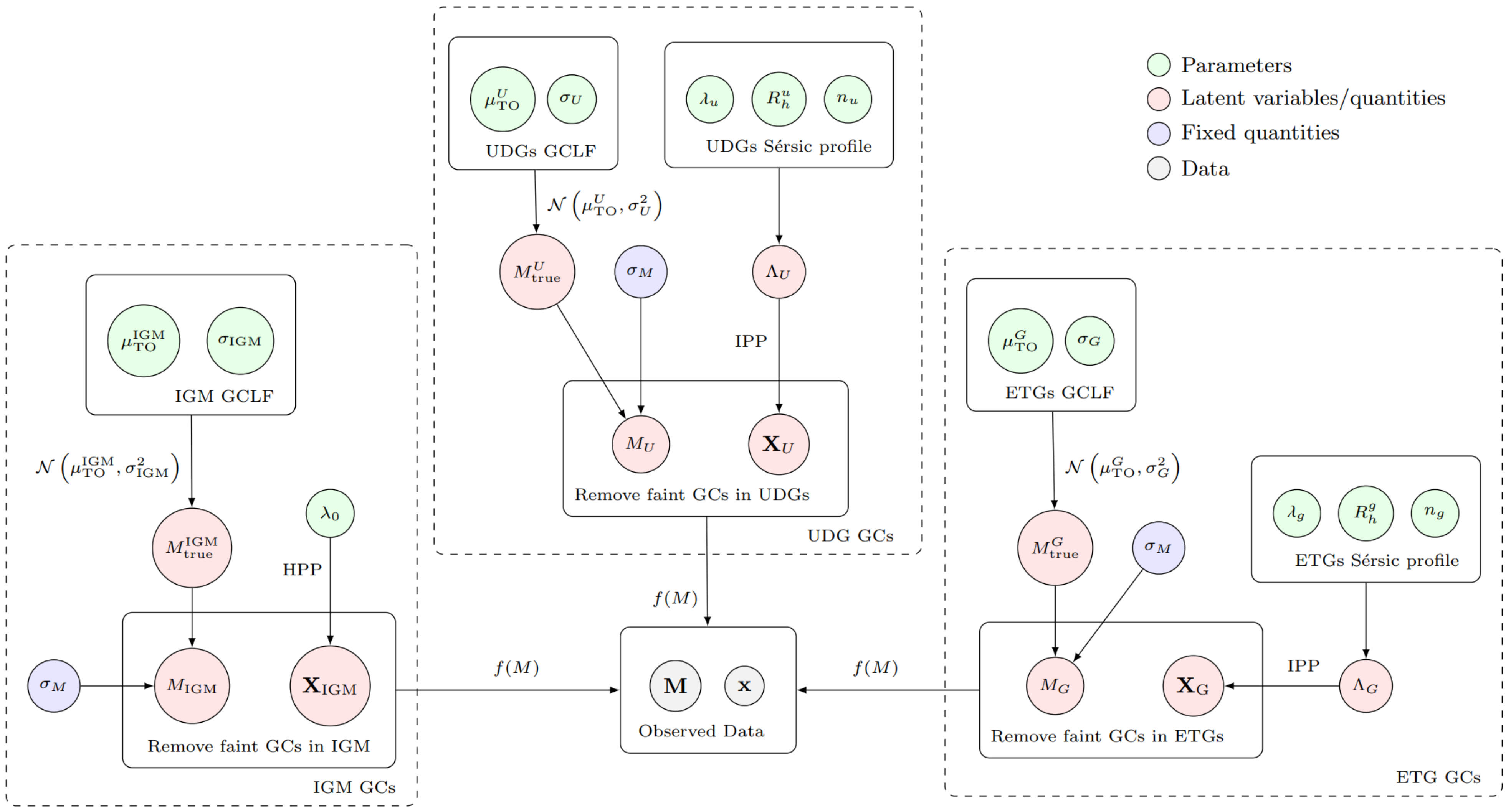
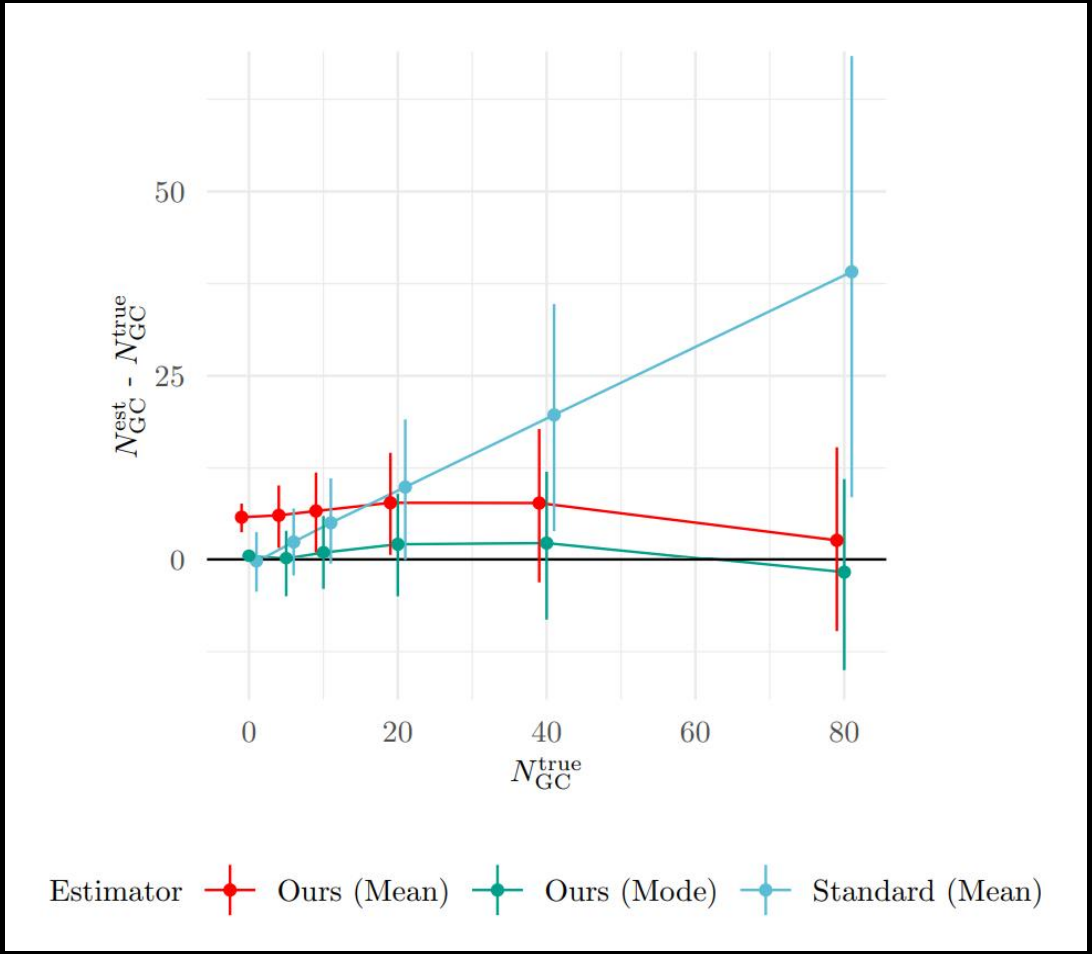


Figure 2. Graphical representation for the data generating process of the observed GC point pattern $\mathbf{x}$ and magnitudes $\mathbf{M}$ under MATHPOP. $\mathbf{x}$ and $\mathbf{M}$ come from independent unions of three different branches of data generating processes for GCs in IGM, UDGs, and luminous early type galaxies.

Li, **Eadie**, et al (2024, submitted to ApJ)

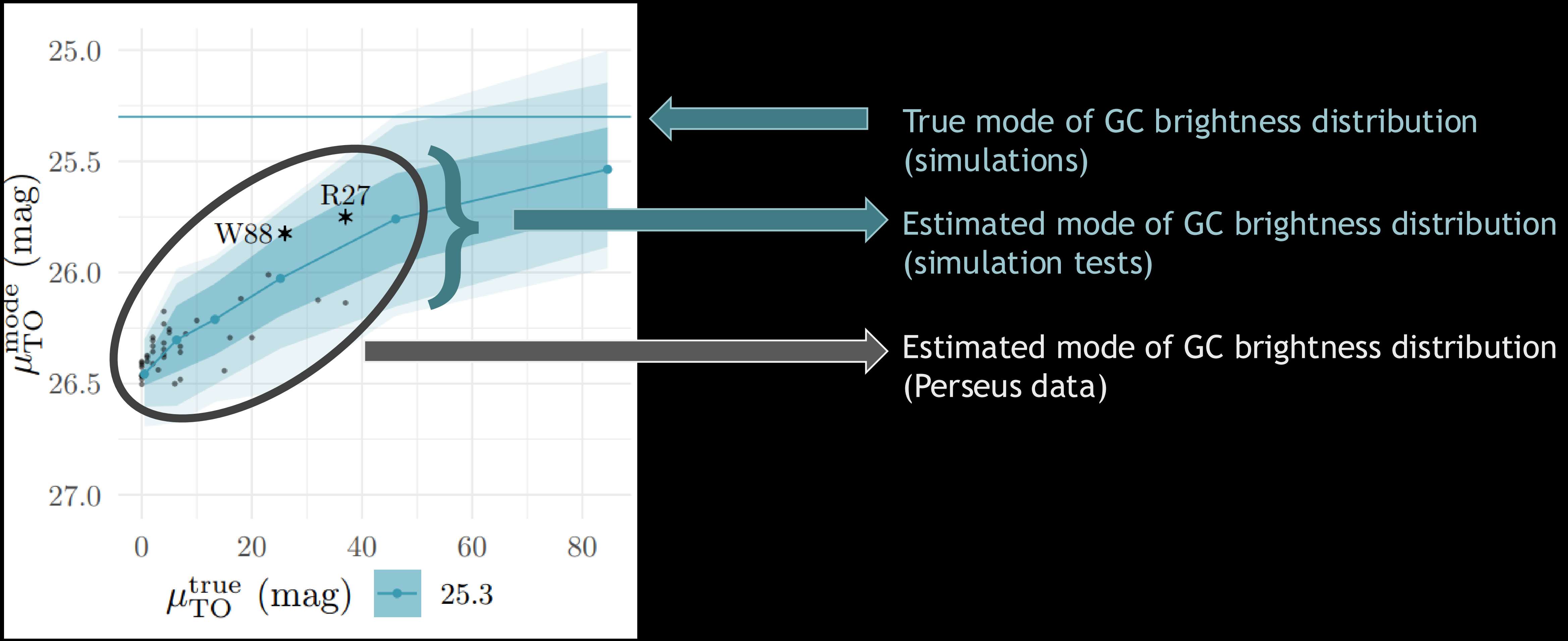
Simulation Tests

- Injected simulated GCs into real images
- Use MATHPOP to estimate number of GCs (N_{est})
- Use literature/standard method to estimate N_{est}
- Look at $N_{est} - N_{true}$ for each

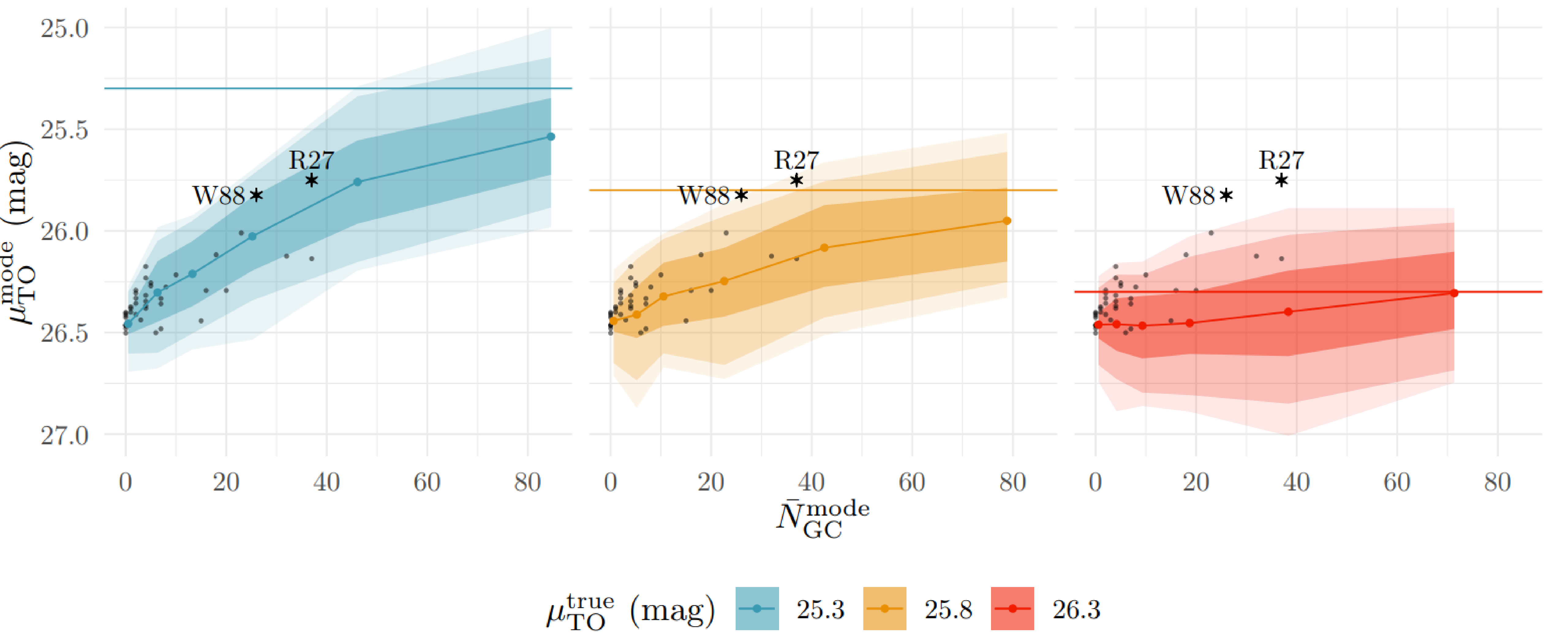


Li, **Eadie**, et al (2024, submitted to ApJ)

Simulation Tests and Application of MATHPOP to Perseus Cluster of Galaxies



Li, **Eadie**, et al (2024, submitted to ApJ)



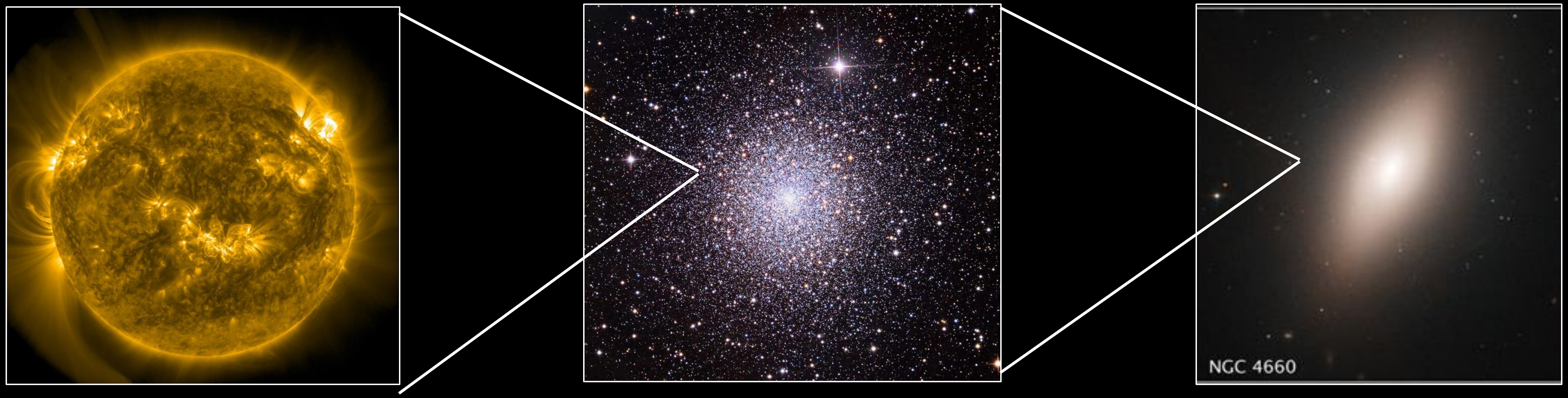
(c) Posterior mode of N_{GC} vs. Posterior mode of μ_{TO} from MathPop

Li, **Eadie**, et al (2024, submitted to ApJ)

Bayesian *MA*rk-dependently *TH*inned *PO*int *P*rocess

Takeaways

- MATHPOP → properly accounting for uncertainties, selection effects, GC identification
- Identified two UDGs with brighter than average GC populations
- Potential implications for distance estimates in astronomy
- We may be on the verge of discovering the "darkest" galaxy to-date!



STUDYING THE UNIVERSE WITH ASTROSTATISTICS

ART

Astrostatistics
Research
Team

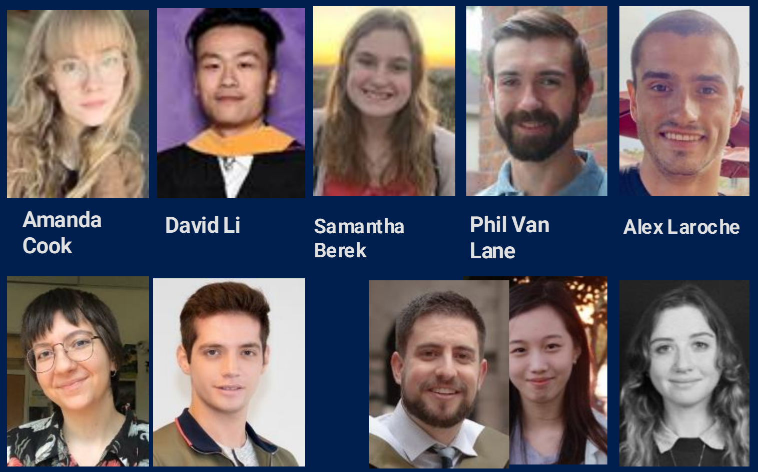
Faculty

Assistant Prof.
Gwen Eadie (me)

Assistant Prof.
Josh Speagle

Thank you!

PhD Students



Amanda Cook | David Li | Samantha Berek | Phil Van Lane | Alex Laroche

Anika Slizewski | Rodrigo Barradas Herrera | Arturo Esquivel | Jenny Su | Mairead Heiger

Postdocs

Biprateep Dey | Jacqueline Antwi-Danso | Mike Walmsley | Antonio Herrera Martin | Kevin McKinnen | Tanveer Karim

Statistical Sciences
UNIVERSITY OF TORONTO

DUNLAP INSTITUTE *for* ASTRONOMY & ASTROPHYSICS

NSERC CRSNG

CANSSI INCASS ONTARIO

DATA SCIENCES INSTITUTE

https://astrostatuoft.com/people.html

Gwendolyn Eadie, Assistant Professor
gwen.eadie@utoronto.ca

BACK-UP SLIDES

LOGISTIC REGRESSION

- Probability $\boldsymbol{p}$ of "success" (i.e., galaxy has a GC population) given some continuous covariate $\boldsymbol{X}$

$$\boldsymbol{p} = \frac{1}{1 + e^{-\boldsymbol{X\beta}}}$$

- Estimate the parameters $\boldsymbol{\beta}$ via maximum likelihood, where the likelihood is

$$L(\boldsymbol{\beta}; y) = \prod_{i}^{n} p_i^{y_i}(1 - p_i)^{1-y_i}$$

- Logistic regression is just one example of a broader class of models from statistics called *generalized linear models* (GLMs, McCullagh & Nelder 1989)

PRIORS

Table 3. Prior setting of the QFD part of the model

Parameter	Meaning	Prior distribution	Hyperparameter used
μ	Mean flux at quiet	$N(\mu_0, \sigma_0^2)$	$\mu_0 = 0, \sigma_0^2 = 100\sigma^2$
σ^2	variance of measurement noise	$\mathrm{invGamma}(\alpha, \beta)$	$\alpha = .01, \beta = .01$
$\log(\lambda)$	log average flux increasing during firing	$N(\mu_\lambda, \sigma_\lambda^2)$	$\mu_\lambda = 0, \sigma_\lambda^2 = 1e3$
$\mathrm{logit}(r)$	logit of decay rate	$N(\mu_r, \sigma_r^2)$	$\mu_r = 0, \sigma_r^2 = 1e3$
$p_{Q\|Q}, p_{Q\|F}$	transition probabilities from Quiet state	$\mathrm{Dir}(\alpha_Q)$	$\alpha_Q = (1, 0.1)$
$p_{F\|F}, p_{F\|D}$	transition probabilities from Firing state	$\mathrm{Dir}(\alpha_F)$	$\alpha_F = (1, 1)$
$p_{D\|Q}, p_{D\|F}, p_{D\|D}$	transition probabilities from Decay state	$\mathrm{Dir}(\alpha_D)$	$\alpha_D = (1, 0.1, 1)$

NOTATION

Table 1. Notation used in this paper.

	Meaning
Y_t	observed brightness of star at time t
$\boldsymbol{f}$	quasi-periodic trend of the time series (modeled with *celerite)*
f_t	*celerite*-modeled trend at time t
Z_t	flaring channel (time series without trend)
$\boldsymbol{\mu}$	mean flux in Quiet state
$\mathcal{K}$	kernel for *celerite*
σ^2	variance of measurement noise
S_t	state of time series at time t
Q	Quiet state
F	Firing state
D	Decay state
$\log(\lambda)$	log average flux increase during firing state
$\mathrm{logit}(r)$	logit of decay rate
$p_{Q\|Q}, p_{Q\|F}$	transition probabilities from Quiet state
$p_{F\|F}, p_{F\|D}$	transition probabilities from Firing state
$p_{D\|Q}, p_{D\|F}, p_{D\|D}$	transition probabilities from Decay state

Detecting stellar flares in photometric data using hidden Markov models

J. Arturo Esquivel,[1,2] Yunyi Shen,[3] Vianey Leos-Barajas,[1,2,4]
Gwendolyn Eadie,[1,2,5] Joshua S. Speagle,[1,2,5] Radu V Craiu,[1,2] Amber Medina,[6]
and James R. A. Davenport[7]

[1] Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada
[2] Data Sciences Institute, University of Toronto, Toronto, ON, Canada
[3] Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA
[4] School of the Environment, University of Toronto, Toronto, ON, Canada
[5] David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON, Canada
[6] Department of Astronomy, University of Texas-Austin, Austin, TX, USA
[7] Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA

Esquivel et al (2024), submitted to ApJ

https://doi.org/10.48550/arXiv.2404.13145

PhD Candidate, statistics, UofT