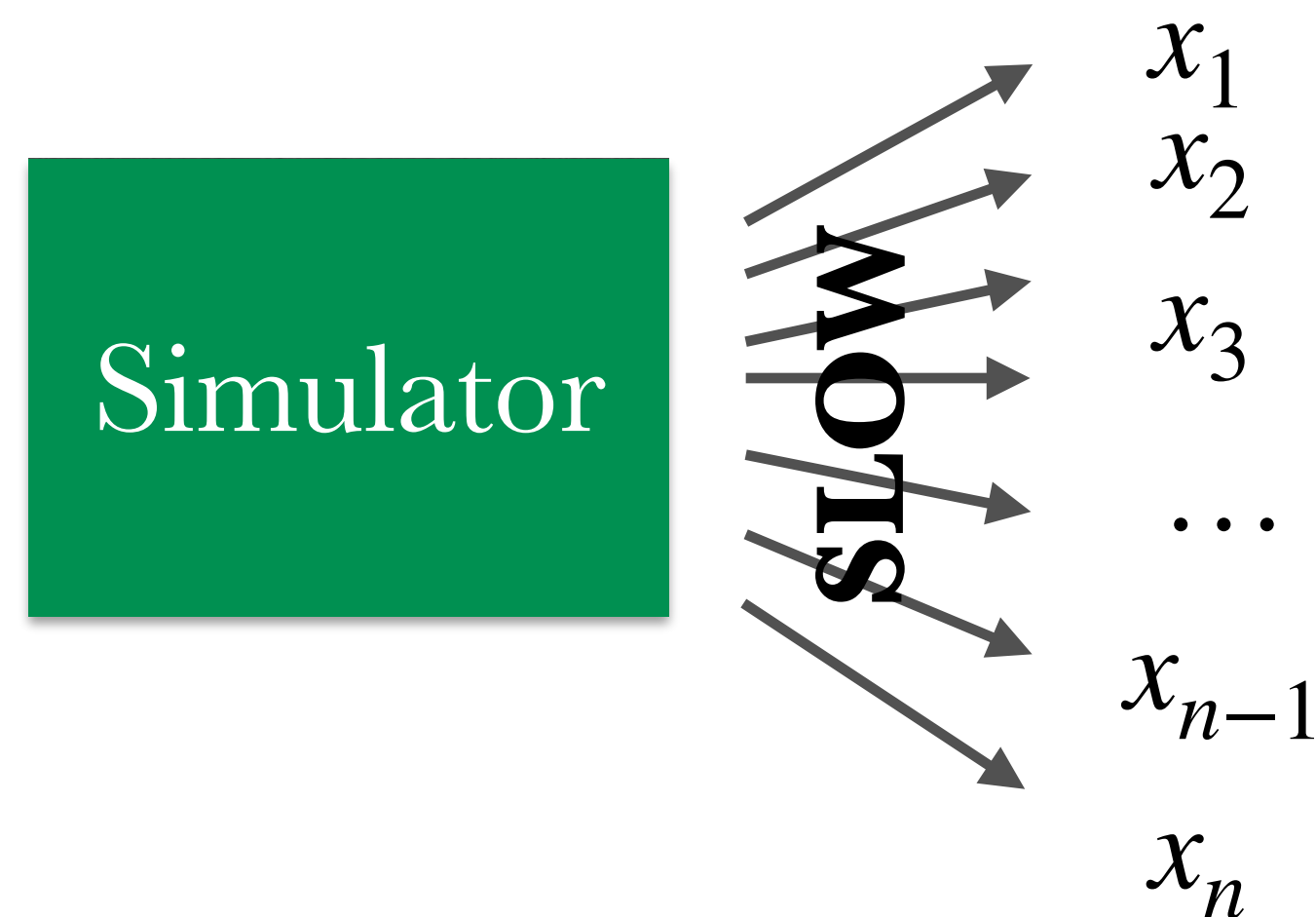


1. Surrogate Modeling for Fast Simulation

(Generation)

Surrogate Models with Generative AI

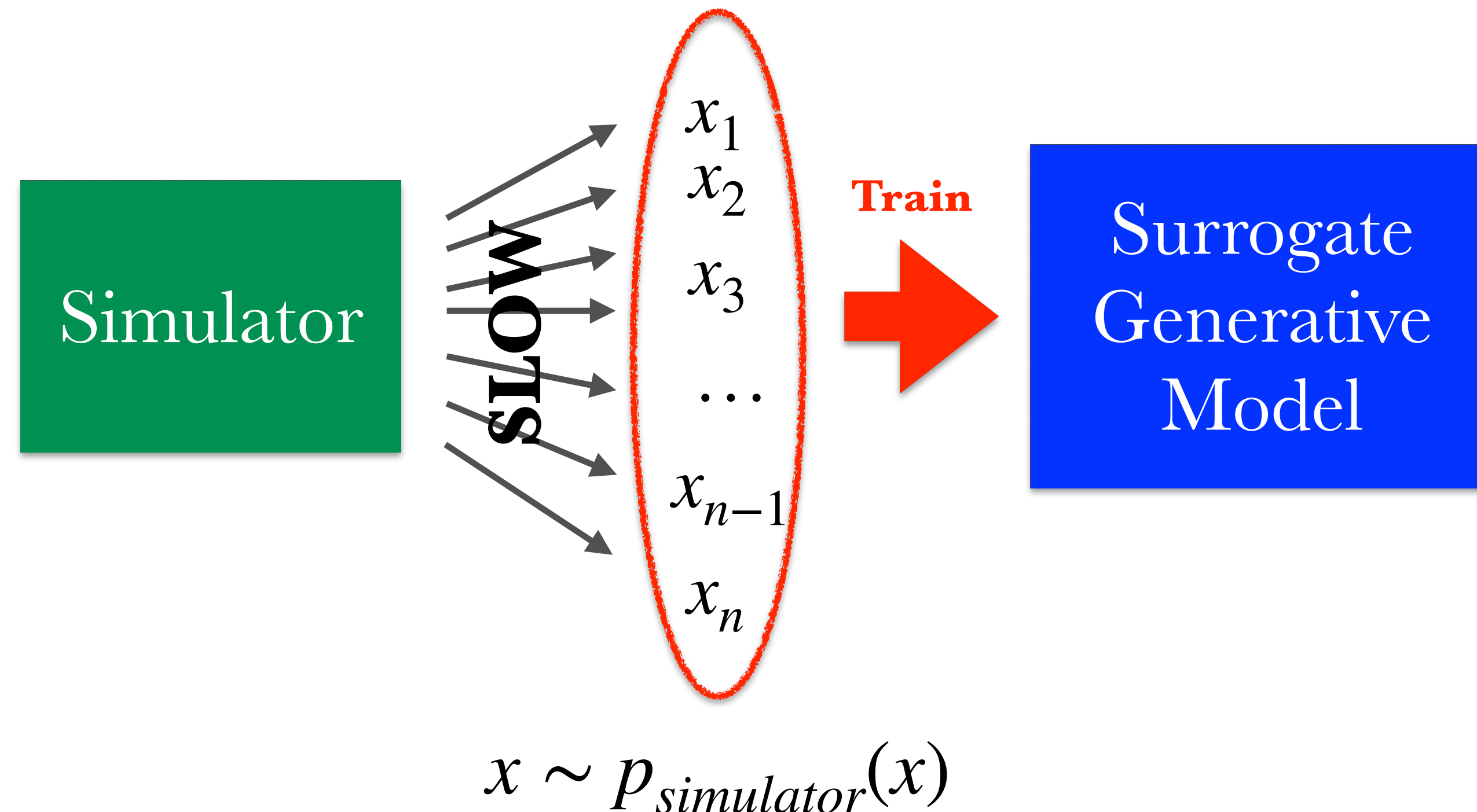
- Simulations are key to science! They are where theory meets experiment, often essential to measurements of all kinds
- But they are often very slow — enter generative AI!



$$x \sim P_{\text{simulator}}(x)$$

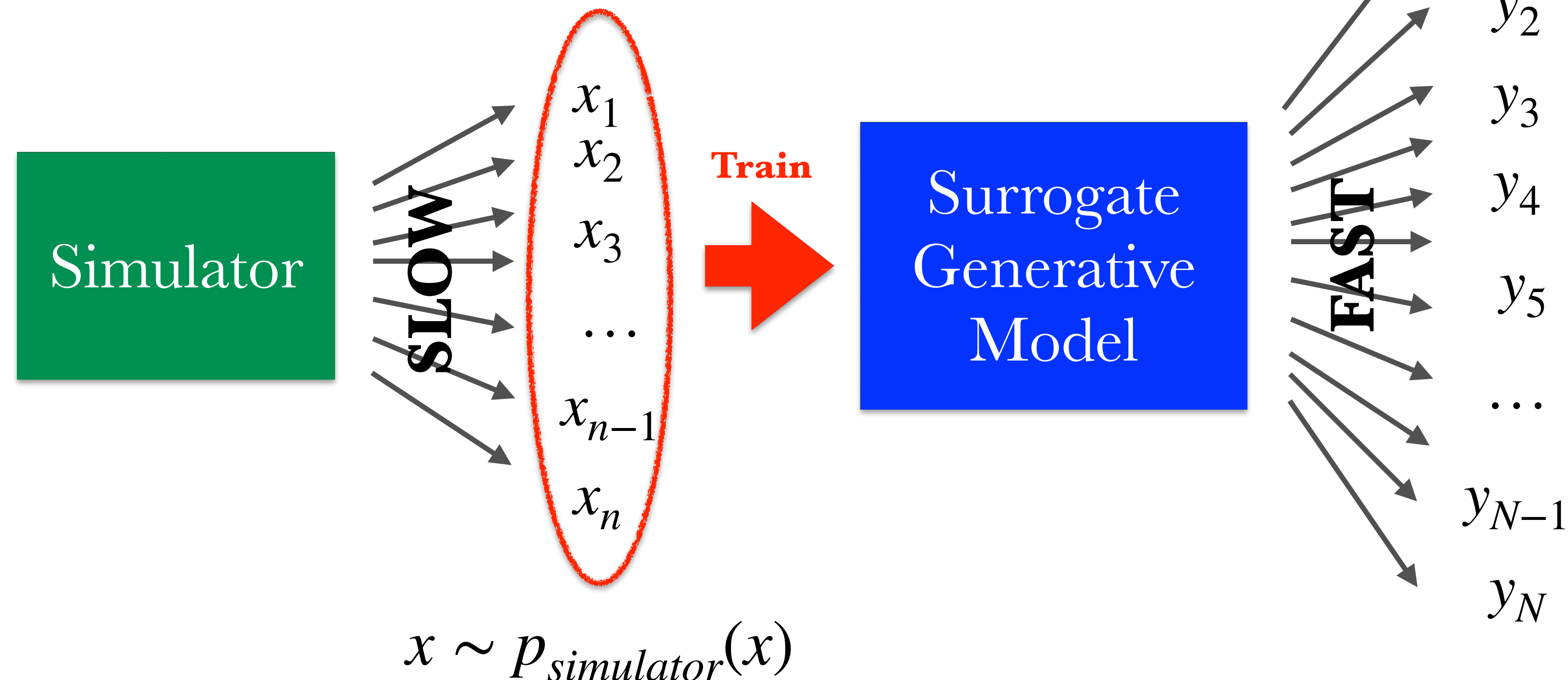
Surrogate Models with Generative AI

- Simulations are key to science! They are where theory meets experiment, often essential to measurements of all kinds
- But they are often very slow — enter generative AI!



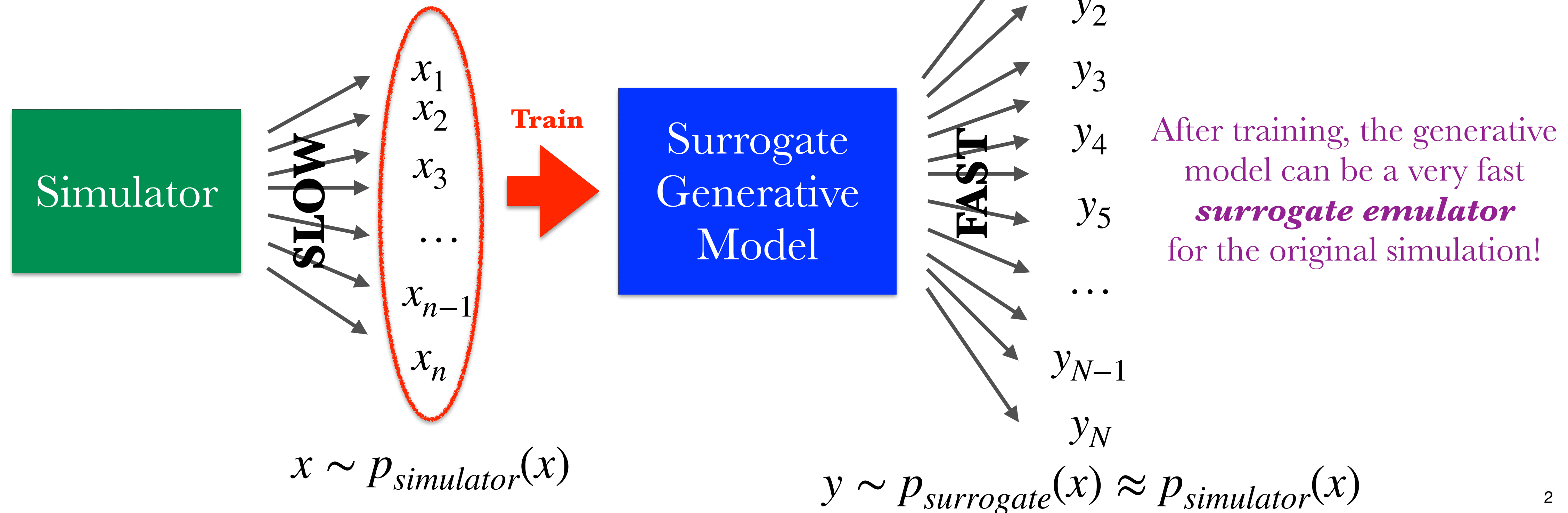
Surrogate Models with Generative AI

- Simulations are key to science! They are where theory meets experiment, often essential to measurements of all kinds
- But they are often very slow — enter generative AI!



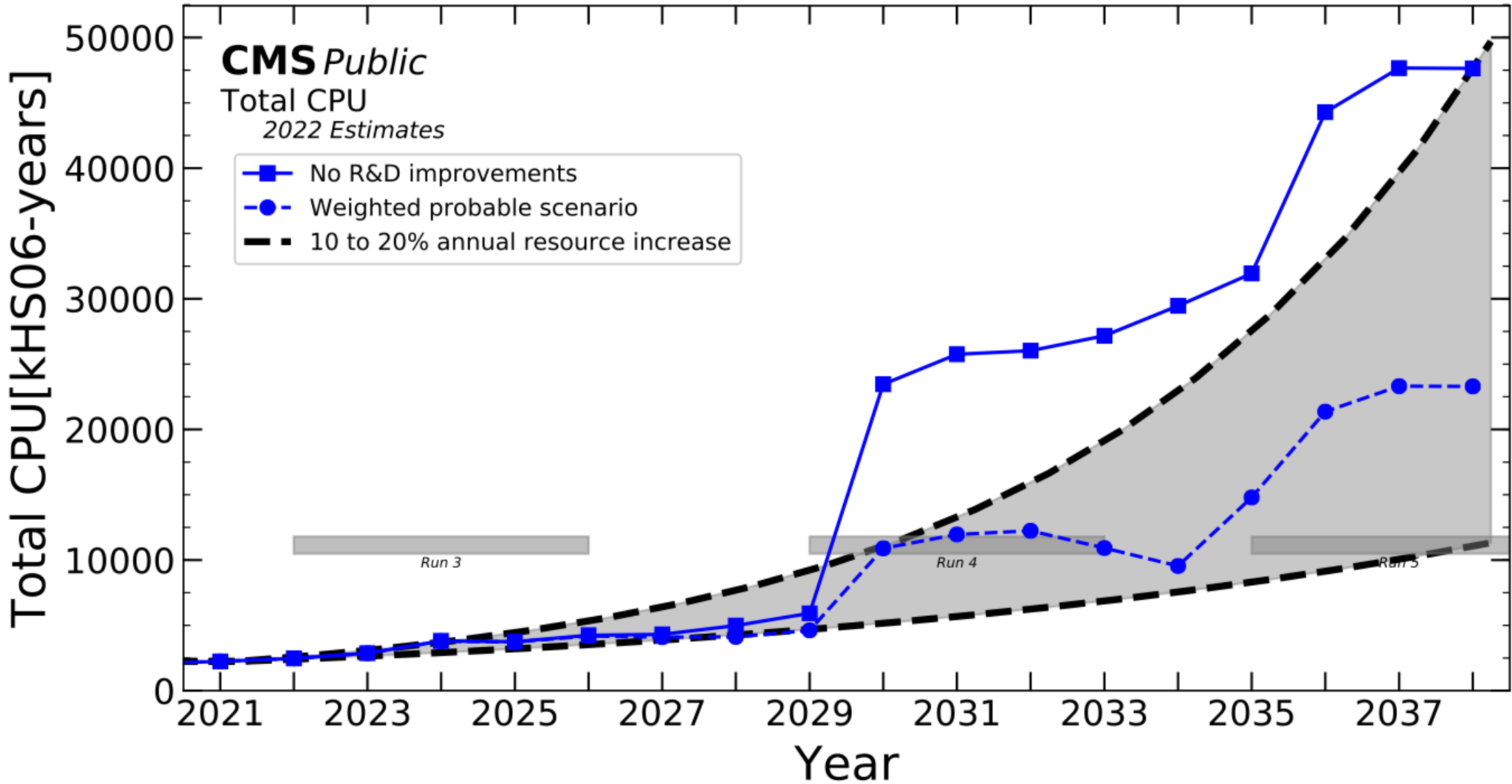
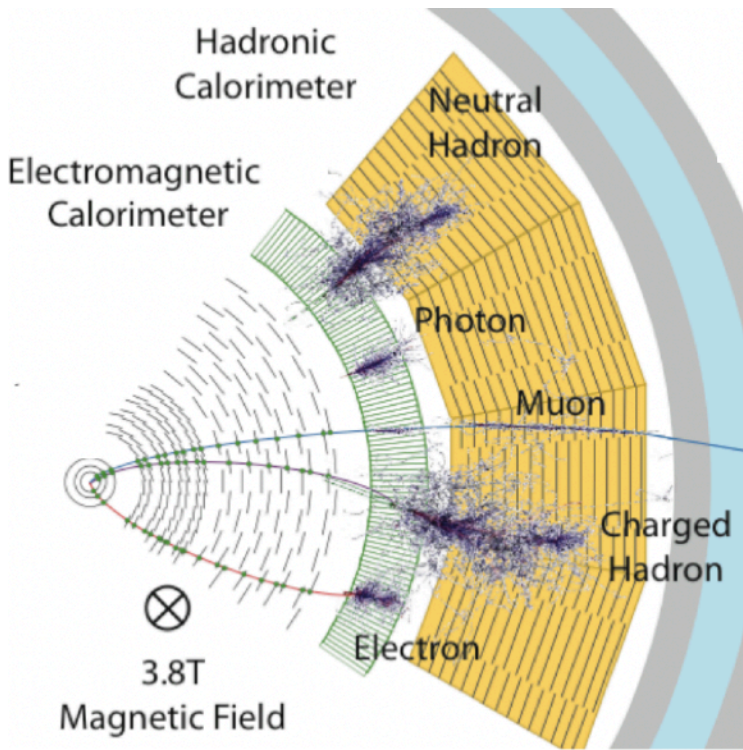
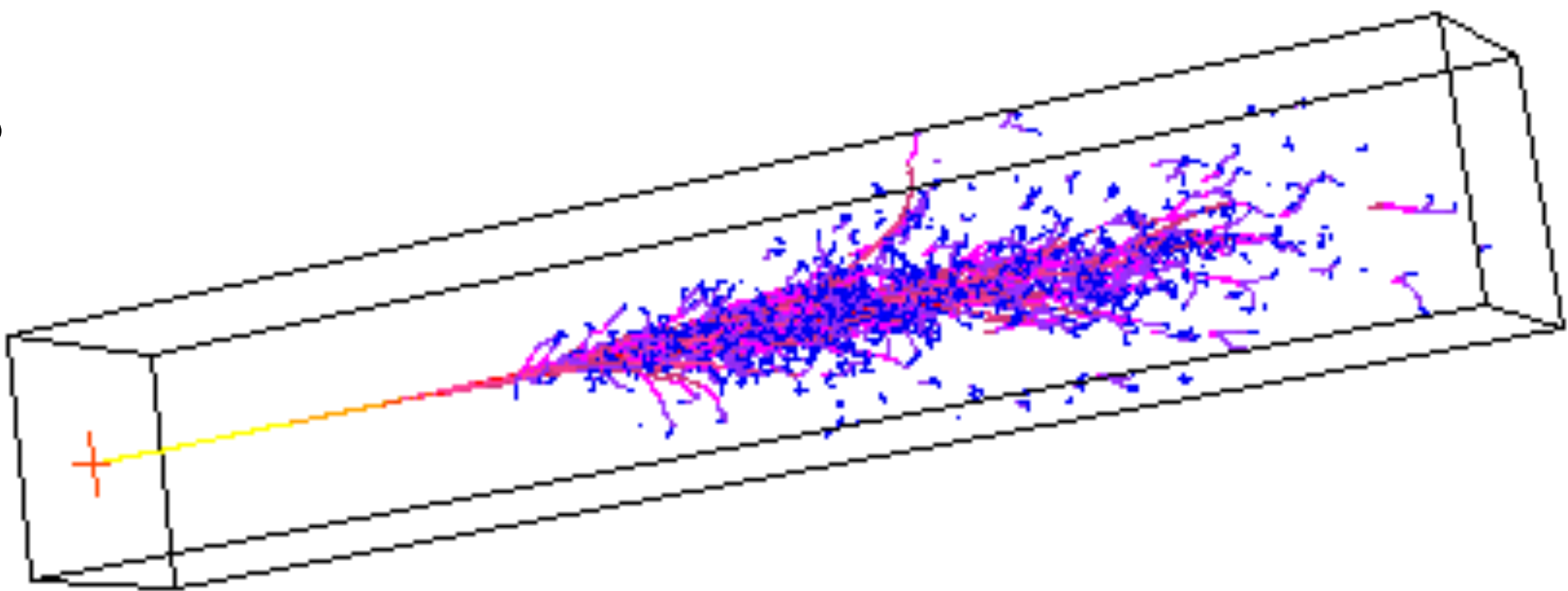
Surrogate Models with Generative AI

- Simulations are key to science! They are where theory meets experiment, often essential to measurements of all kinds
- But they are often very slow — enter generative AI!

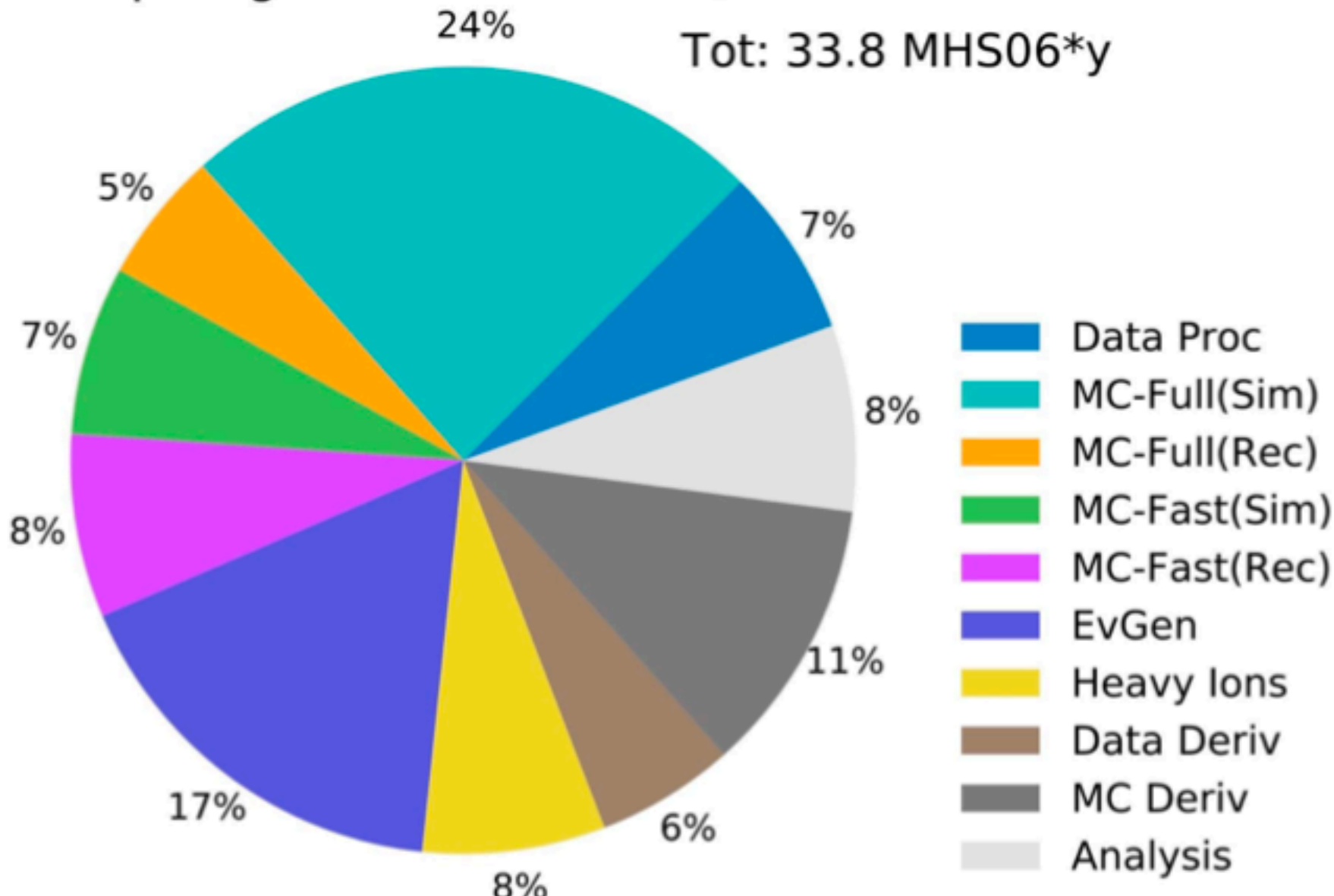


Surrogate Models for Calorimeter Simulation

Simulating calorimeter showers with GEANT4 is the most expensive part of the LHC compute budget



ATLAS Preliminary
2022 Computing Model - CPU: 2031, Conservative R&D
Tot: 33.8 MHS06*y



CaloGAN and CaloFlow

Early work showed that generative models such as **GANs** and **normalizing flows** could be *fast* and *high fidelity* surrogate models for GEANT4 showers

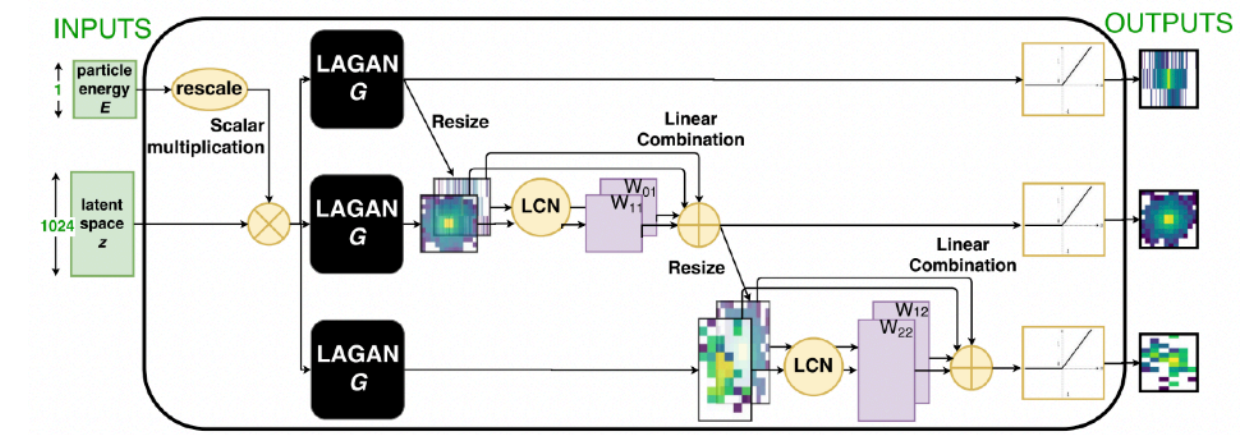
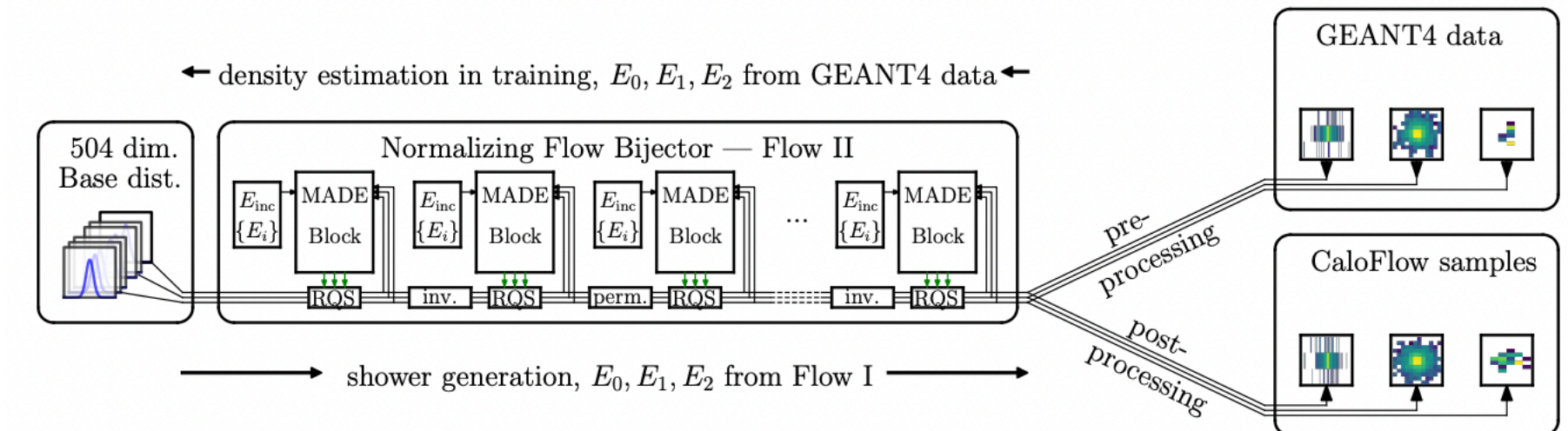
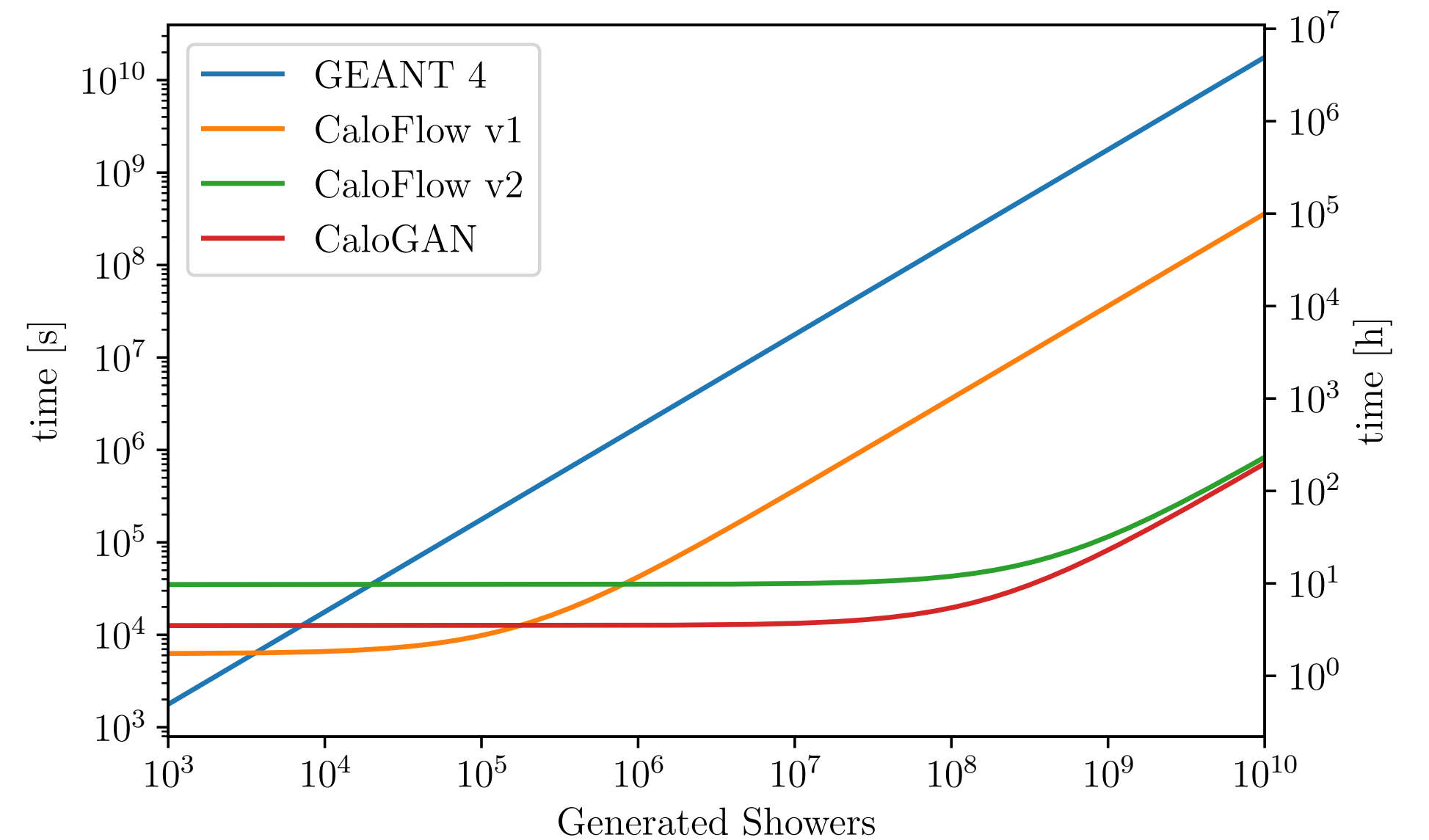


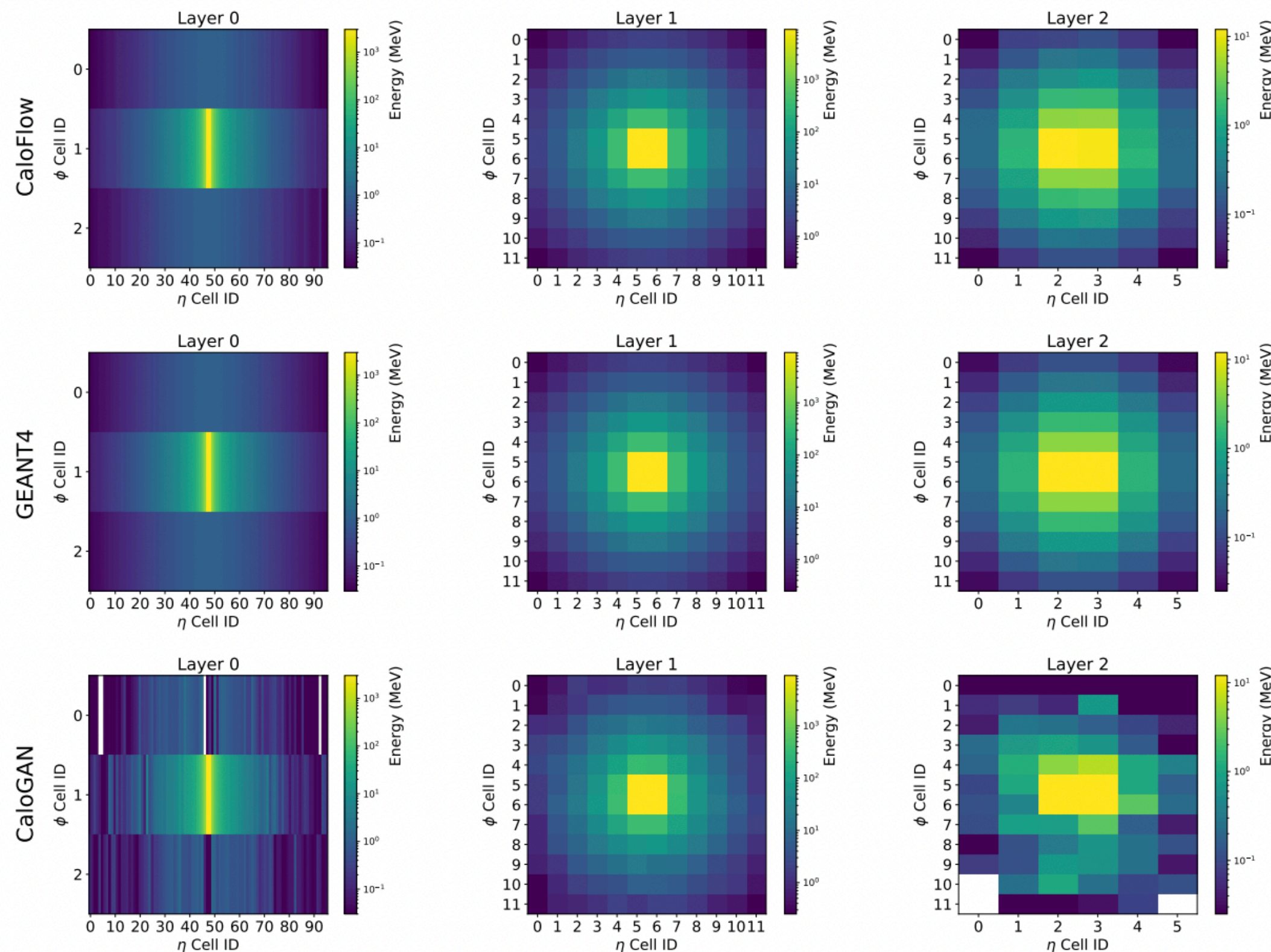
FIG. 4: Composite Generator, illustrating three stream with attentional layer-to-layer dependence.

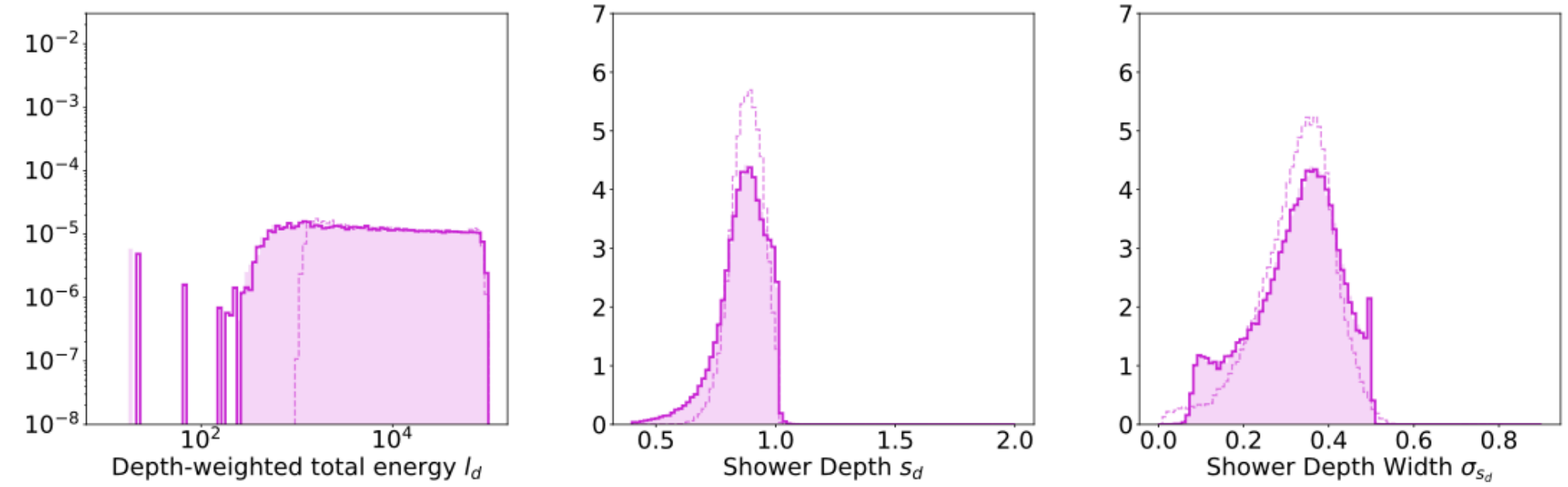
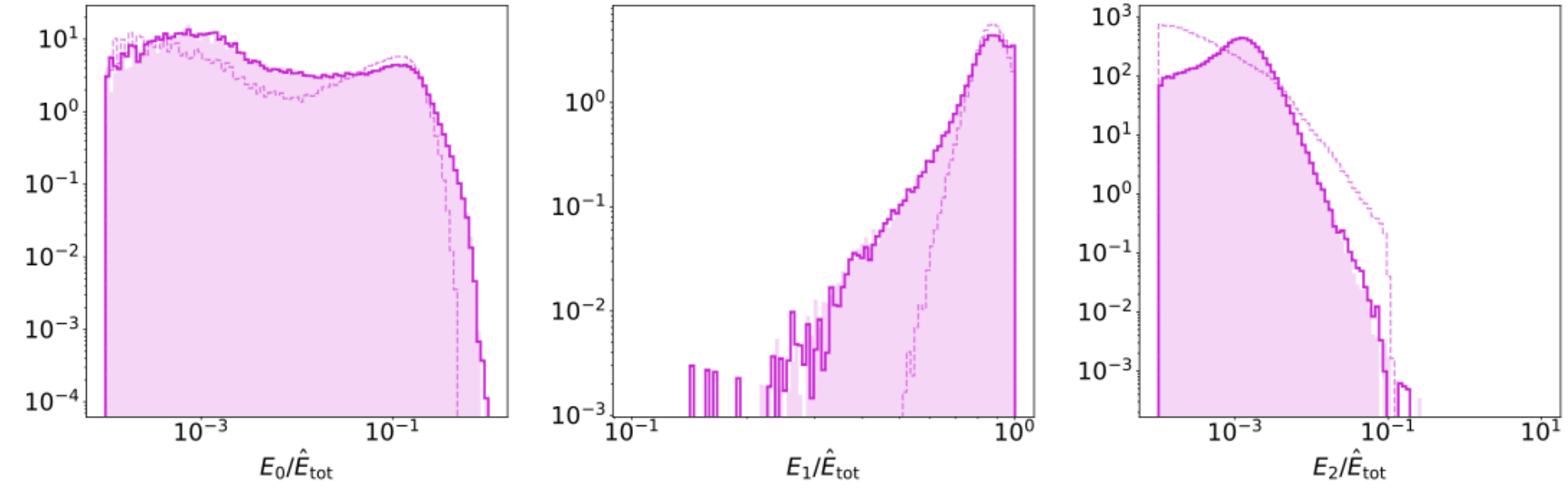
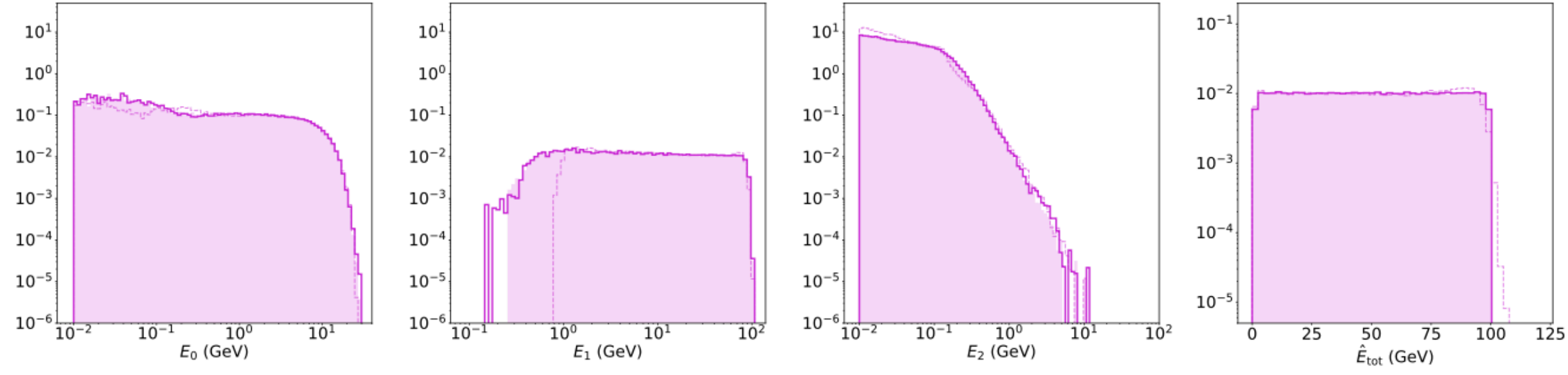


Krause & DS, [2106.05285](#), [2110.11377](#)

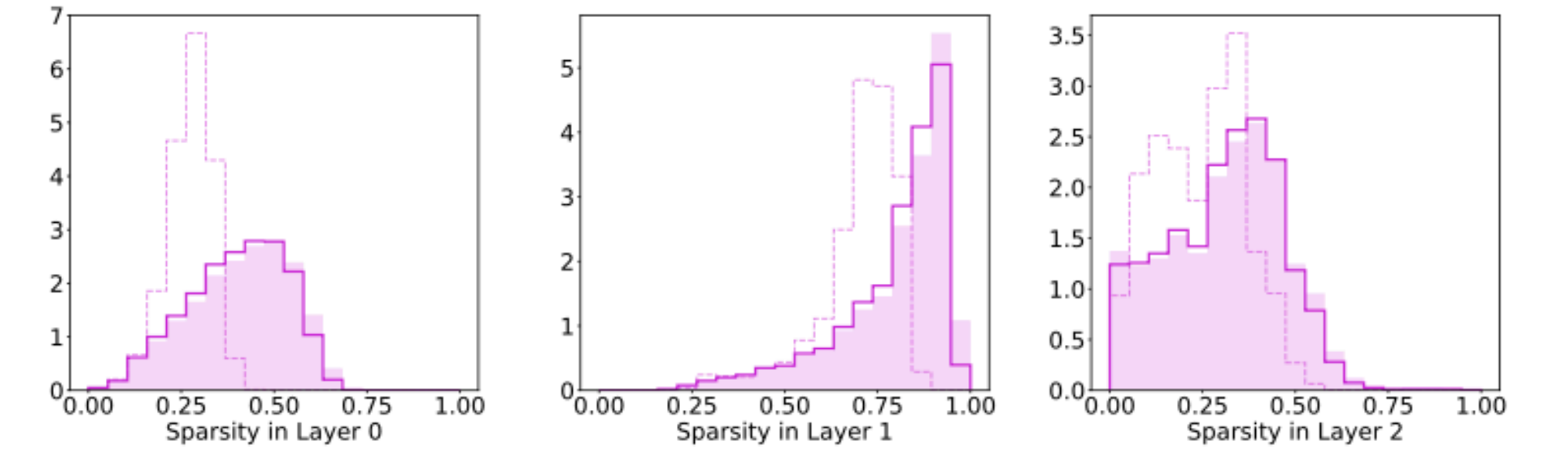
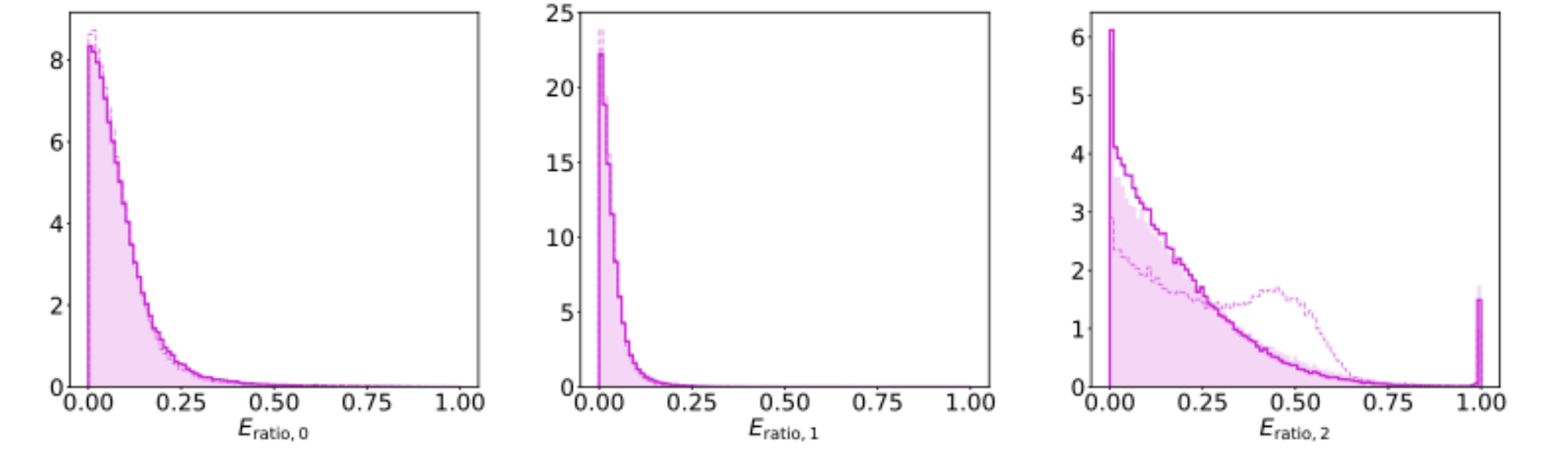
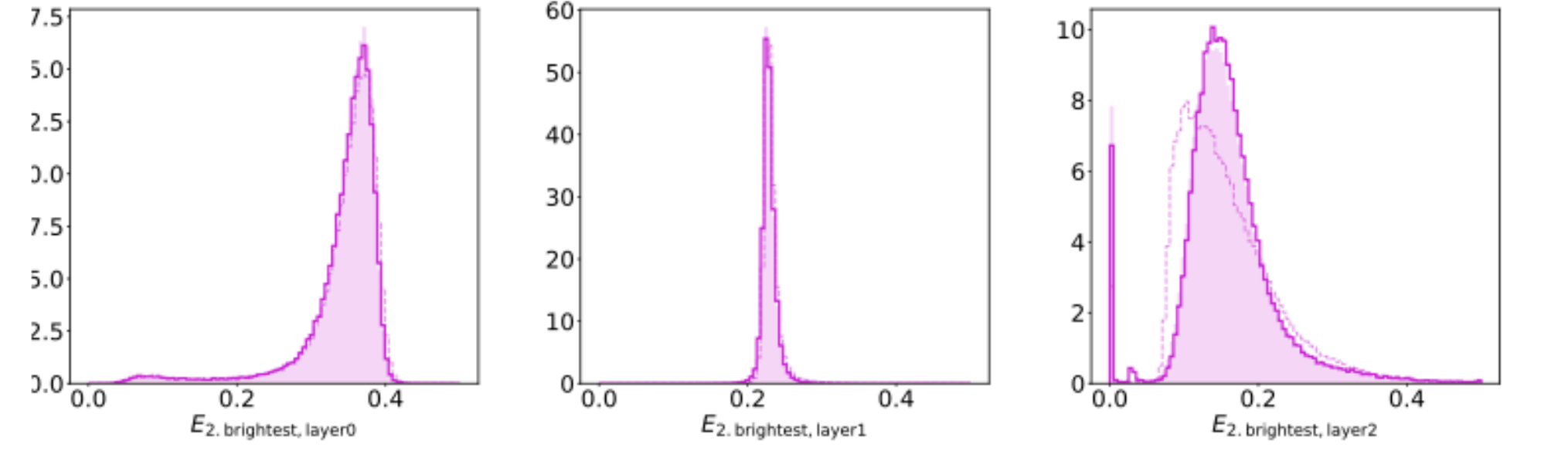
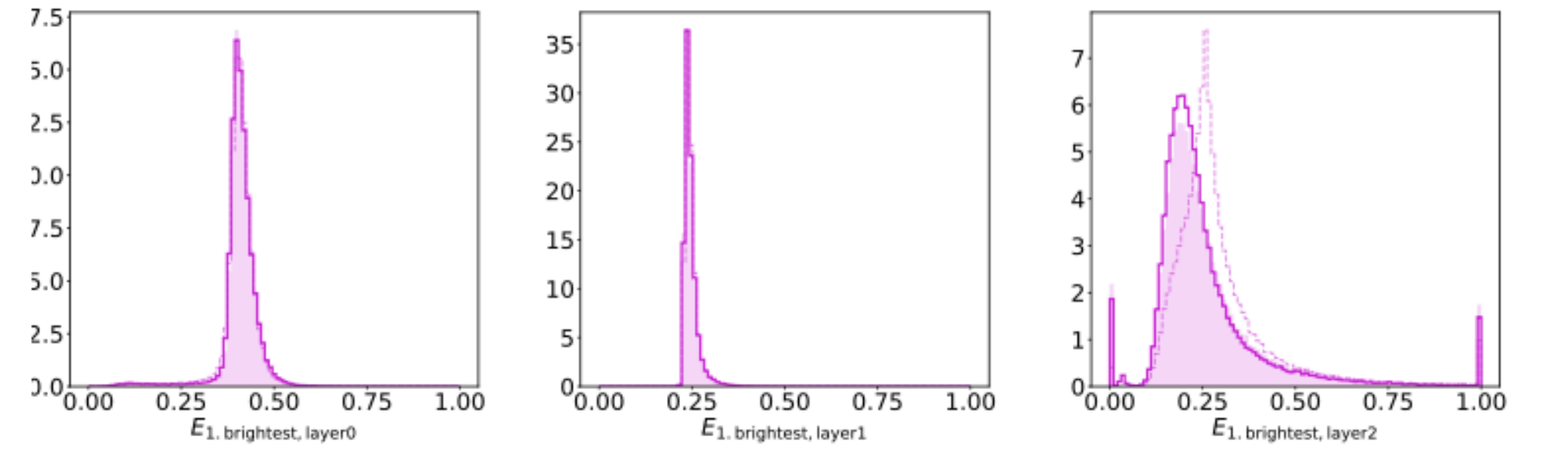


$\sim 10^4 \times$ faster than GEANT4!





γ GEANT
 γ CaloGAN
 γ CaloFlow



γ GEANT
 γ CaloGAN
 γ CaloFlow

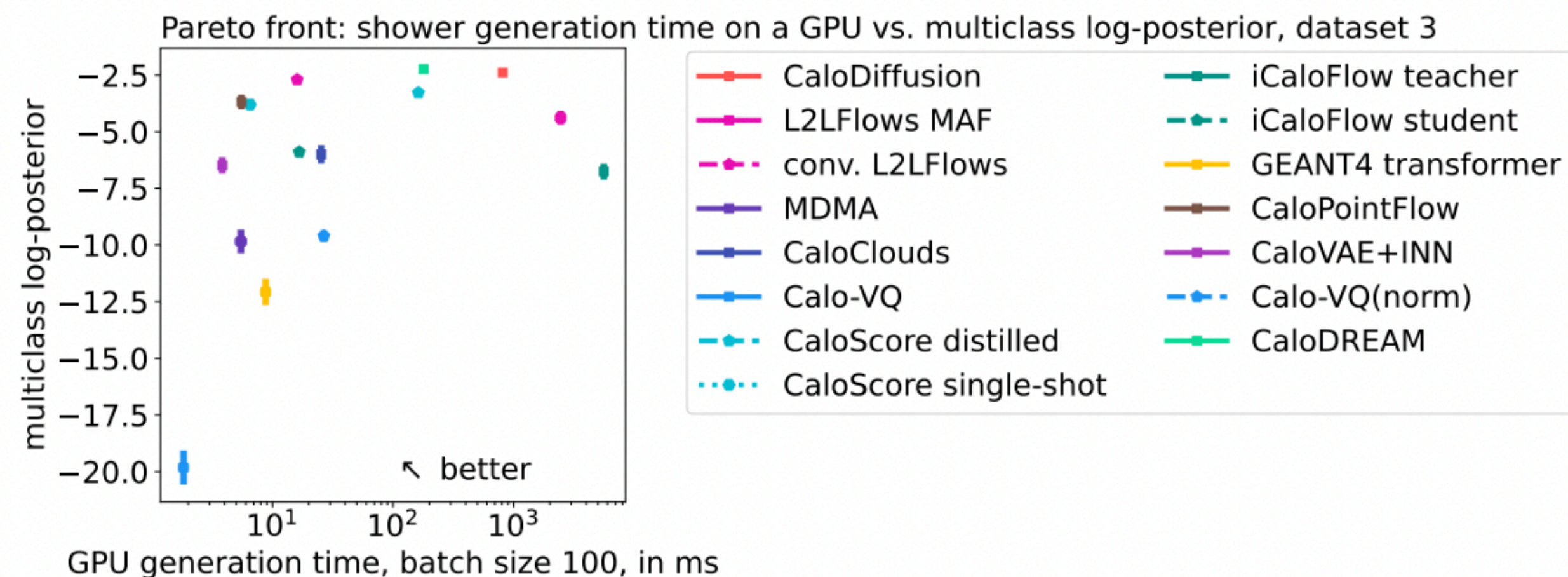
CaloChallenge

- Public data challenge 2022-2024
(organizers: Krause, Faucci Giannelli, Kasieczka, Nachman, Salamani, **DS**, Zaborowska)
- 3 GEANT4 training datasets of increasing dimensionality
- 70+ participants
- 31 submissions
(GANs, VAEs, flows, diffusion, ...)

CaloChallenge 2022: A Community Challenge for Fast Calorimeter Simulation



“Calorimeter Simulation”, generated via midjourney, 2022



Claudius Krause^{1,2} (main editor) ,
 Michele Faucci Giannelli^{3,4} (editor) ,
 Gregor Kasieczka⁵ (editor) , Benjamin Nachman⁶ (editor) ,
 Dalila Salamani⁷ (editor) , David Shih⁸ (editor) ,
 Anna Zaborowska⁷ (editor) ,

Oz Amram⁹ , Kerstin Borrás^{10,11} , Matthew R. Buckley⁸ ,
 Erik Buhmann⁵ , Thorsten Buss^{5,10} ,
 Renato Paulo Da Costa Cardoso⁷, Anthony L. Caterini¹² ,
 Nadezda Chernyavskaya⁷ , Federico A.G. Corchia^{13,14} ,
 Jesse C. Cresswell¹² , Sascha Diefenbacher⁶ ,
 Etienne Dreyer¹⁵ , Vijay Ekambaram¹⁶, Engin Eren¹⁰ ,
 Florian Ernst^{2,7} , Luigi Favaro² , Matteo Franchini^{13,14}

[<https://arxiv.org/abs/2410.21611>]

2. Measuring the Phase Space Density of Stars

(Density estimation)

In collaboration with:



Matt Buckley

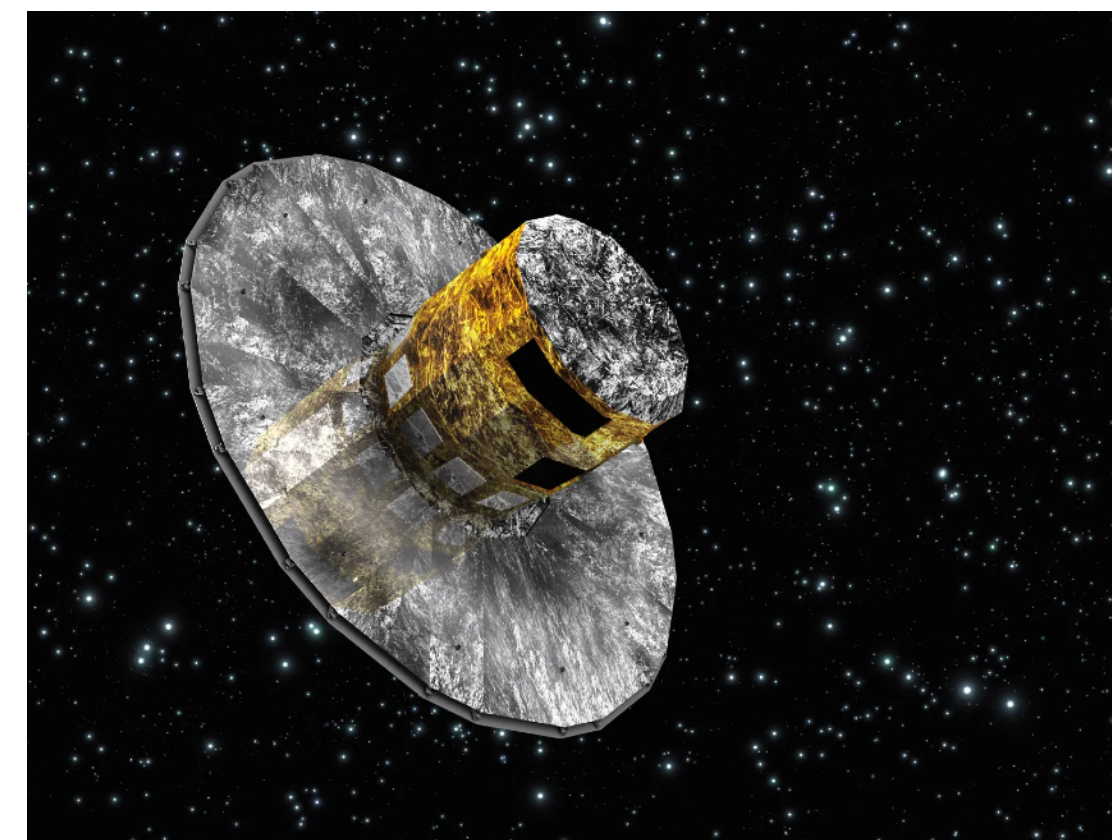
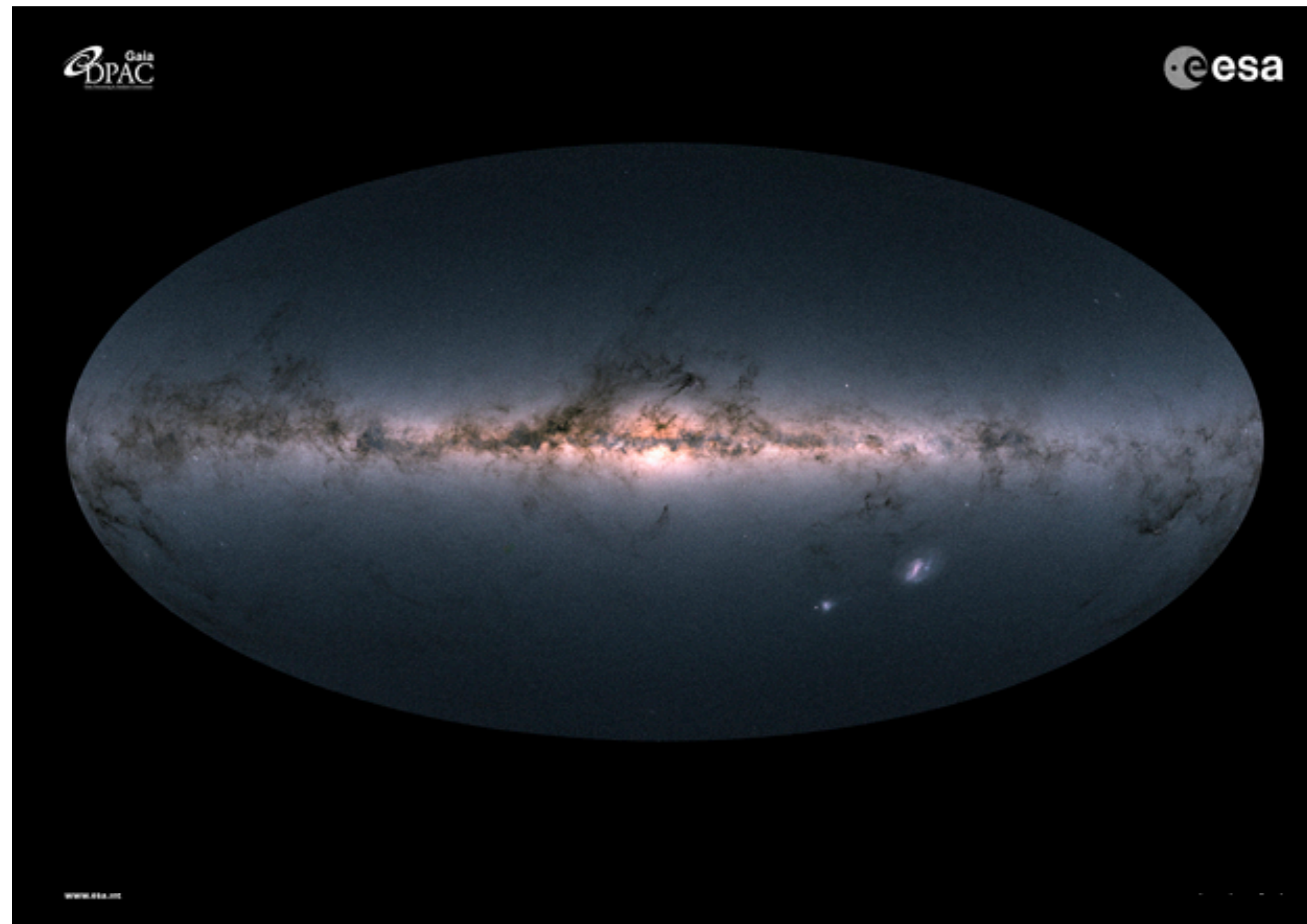


Sung Hak Lim



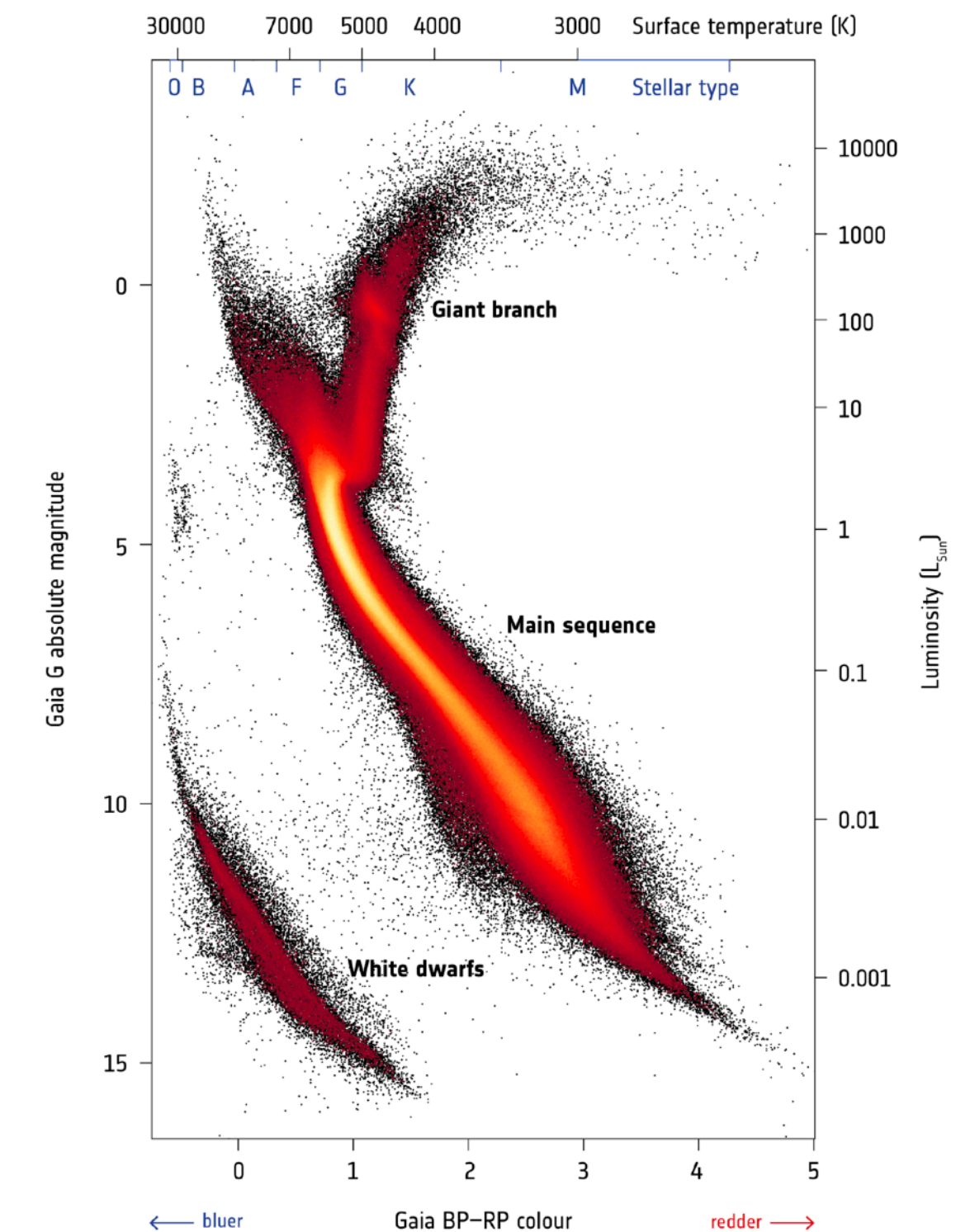
Eric Putney

Gaia space telescope



- Lifetime: 2013-2025
- Currently on 3rd Data Release (DR3)

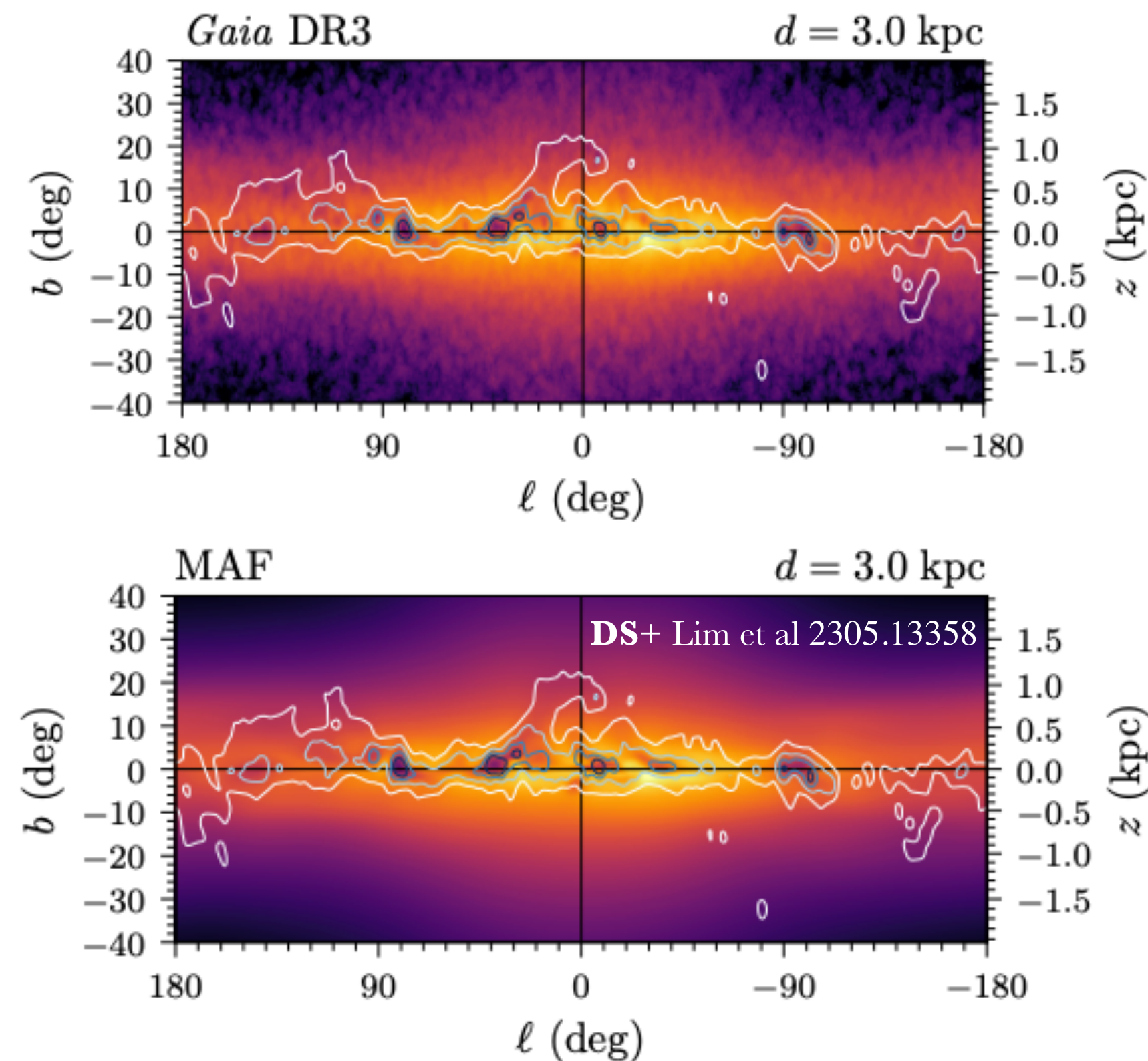
→ GAIA'S HERTZSPRUNG-RUSSELL DIAGRAM



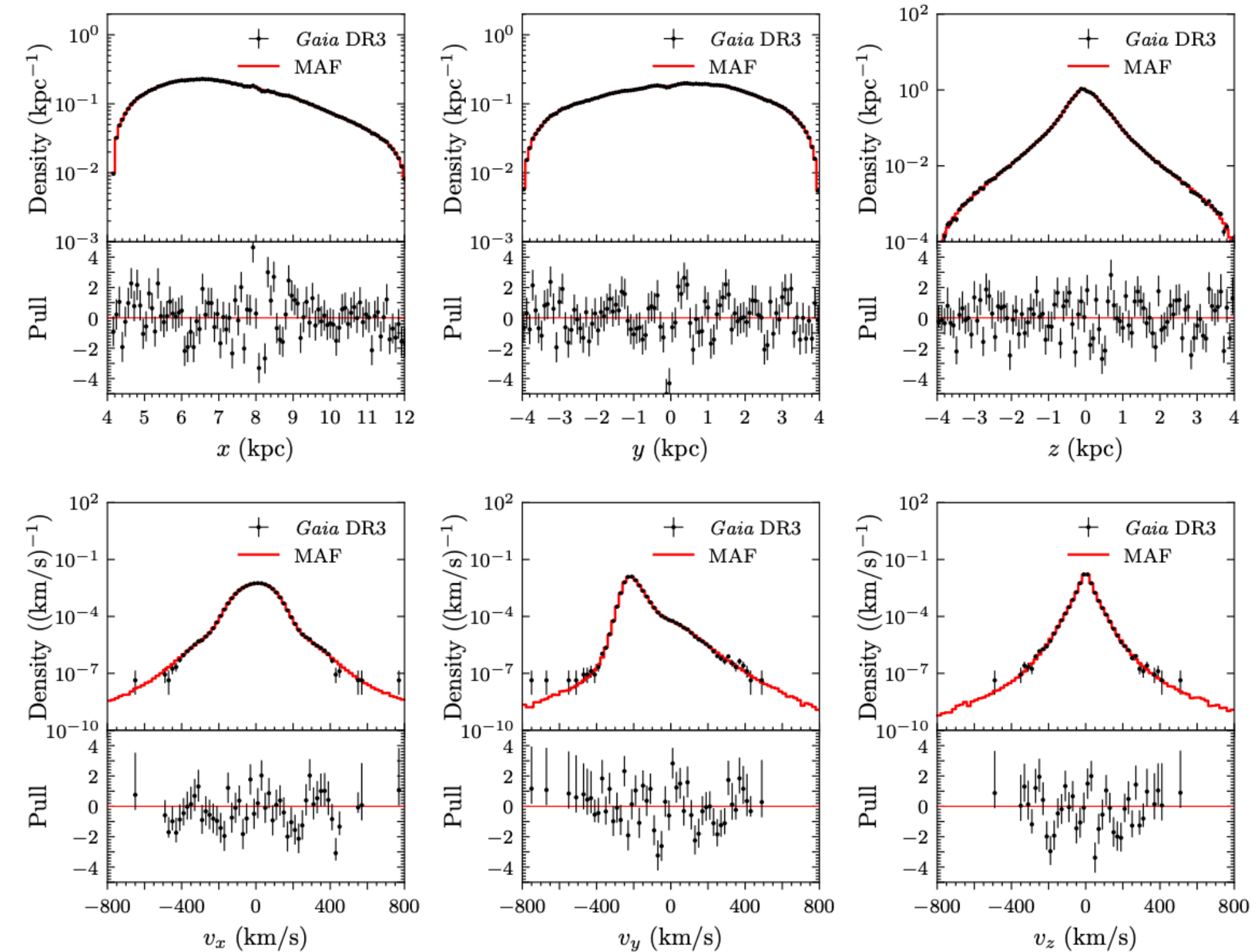
- **Angular positions**, **proper motions**, **color & magnitude** of over **1 billion stars** in our Galaxy
- **Full 6d kinematics** (distances and radial velocities) for a smaller subset of nearby stars (**~30M** within **~4 kpc**)

Stellar Phase Space Density from Gaia DR3

- The phase space density of stars is a key observable for Galactic dynamics
- Can be learned well with a normalizing flow!



DS+ Lim et al 2305.13358



A PINN for Dark Matter

Green et al 2011.04673, 2205.02244; Naik et al 2112.07657;
An et al 2106.05981; Kalda et al 2310.00040 2507.03742
Buckley, Lim, Putney & DS 2205.01129, 2305.13358, 2412.14236, 2512.09989

A PINN for Dark Matter

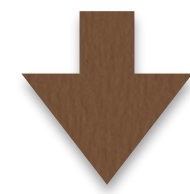
Green et al 2011.04673, 2205.02244; Naik et al 2112.07657;
An et al 2106.05981; Kalda et al 2310.00040 2507.03742
Buckley, Lim, Putney & DS 2205.01129, 2305.13358, 2412.14236, 2512.09989

Fit normalizing flow to Gaia stars
with measured 6d kinematics

A PINN for Dark Matter

Green et al 2011.04673, 2205.02244; Naik et al 2112.07657;
An et al 2106.05981; Kalda et al 2310.00040 2507.03742
Buckley, Lim, Putney & DS 2205.01129, 2305.13358, 2412.14236, 2512.09989

Fit normalizing flow to Gaia stars
with measured 6d kinematics

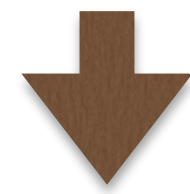


Learn phase space density $f(\vec{x}, \vec{v})$

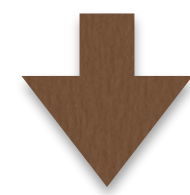
A PINN for Dark Matter

Green et al 2011.04673, 2205.02244; Naik et al 2112.07657;
An et al 2106.05981; Kalda et al 2310.00040 2507.03742
Buckley, Lim, Putney & DS 2205.01129, 2305.13358, 2412.14236, 2512.09989

Fit normalizing flow to Gaia stars
with measured 6d kinematics



Learn phase space density $f(\vec{x}, \vec{v})$



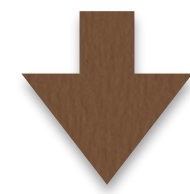
Solve eq. Collisionless Boltzmann Equation
for gravitational potential parametrized with NN

$$\vec{v} \cdot \frac{\partial f}{\partial \vec{x}} - \nabla \Phi(x) \cdot \frac{\partial f}{\partial \vec{v}} = 0$$

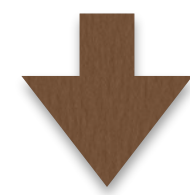
A PINN for Dark Matter

Green et al 2011.04673, 2205.02244; Naik et al 2112.07657;
An et al 2106.05981; Kalda et al 2310.00040 2507.03742
Buckley, Lim, Putney & DS 2205.01129, 2305.13358, 2412.14236, 2512.09989

Fit normalizing flow to Gaia stars
with measured 6d kinematics



Learn phase space density $f(\vec{x}, \vec{v})$



Solve eq. Collisionless Boltzmann Equation
for gravitational potential parametrized with NN

$$\vec{v} \cdot \frac{\partial f}{\partial \vec{x}} - \nabla \Phi(x) \cdot \frac{\partial f}{\partial \vec{v}} = 0$$

Derivatives of potential give
acceleration field and mass density

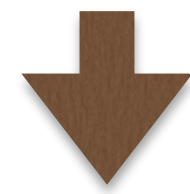
$$\vec{a}(x) = - \nabla \Phi(x)$$

$$4\pi G \rho(x) = \nabla^2 \Phi(x)$$

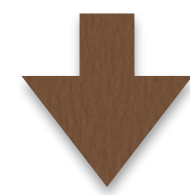
A PINN for Dark Matter

Green et al 2011.04673, 2205.02244; Naik et al 2112.07657;
An et al 2106.05981; Kalda et al 2310.00040 2507.03742
Buckley, Lim, Putney & DS 2205.01129, 2305.13358, 2412.14236, 2512.09989

Fit normalizing flow to Gaia stars
with measured 6d kinematics



Learn phase space density $f(\vec{x}, \vec{v})$



Solve eq. Collisionless Boltzmann Equation
for gravitational potential parametrized with NN

$$\vec{v} \cdot \frac{\partial f}{\partial \vec{x}} - \nabla \Phi(x) \cdot \frac{\partial f}{\partial \vec{v}} = 0$$

Derivatives of potential give
acceleration field and mass density

$$\vec{a}(x) = - \nabla \Phi(x)$$

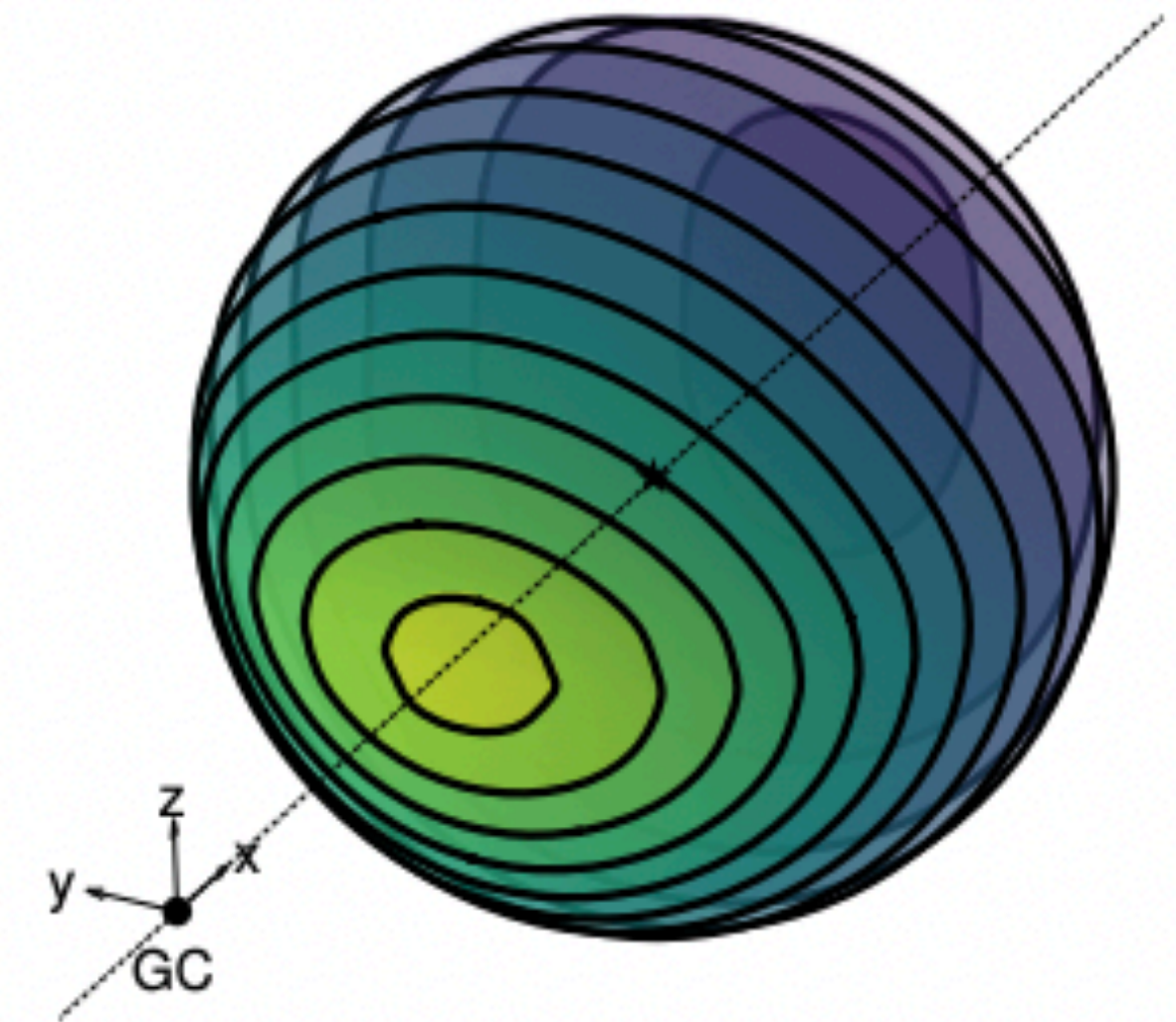
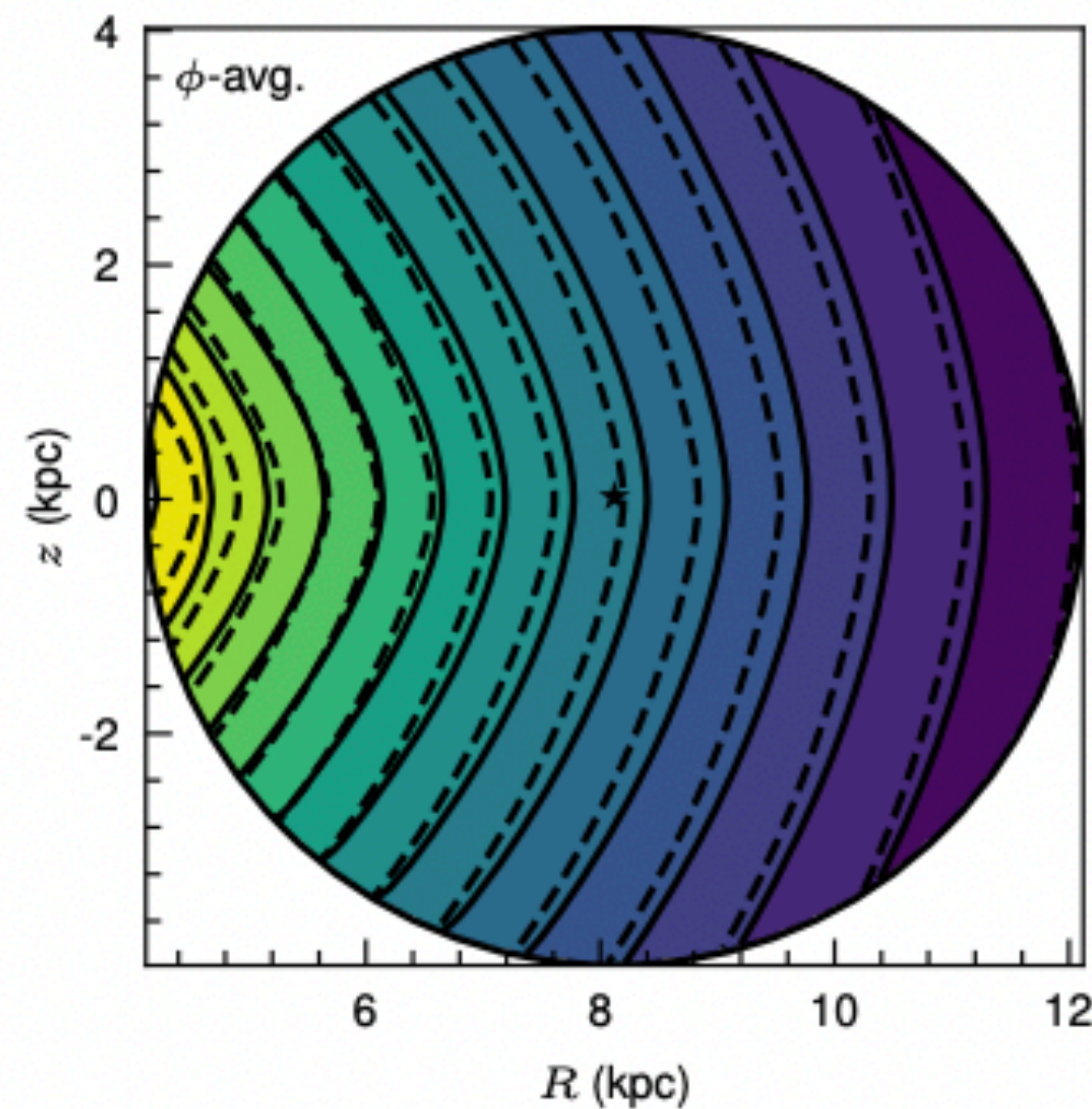
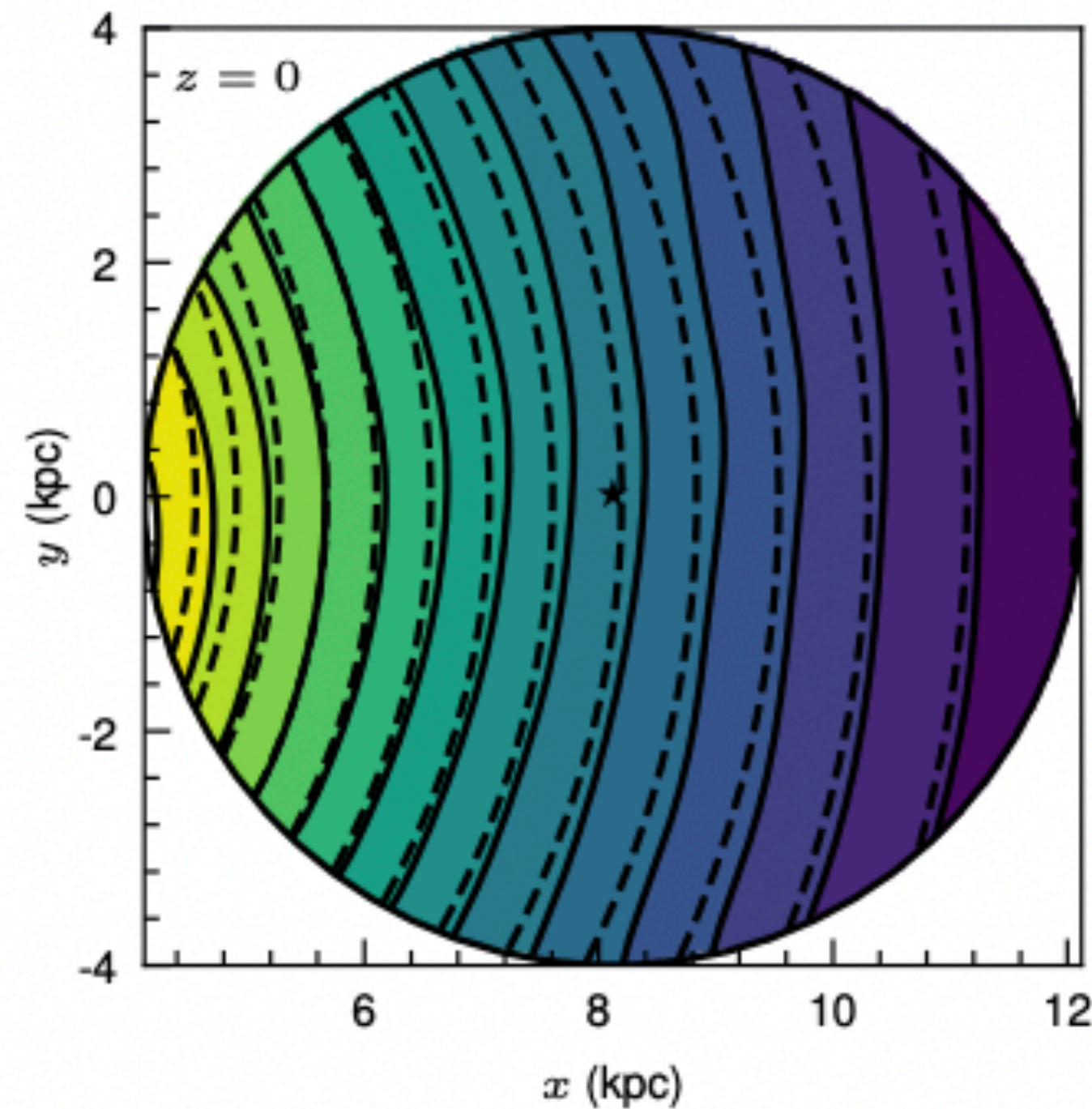
$$4\pi G\rho(x) = \nabla^2 \Phi(x)$$

**Unbinned, symmetry-free, data-
driven, fully 3d measurement
with ~ 4 kpc of Solar location!**

Results: Potential

DS+ Putney et al 2512.09989

Solid: our $\Phi(x)$
Dashed: MWPotential2014 $\Phi(x)$



- Generally good qualitative agreement with gala's MWPotential2014!
- Expected symmetries (azimuthal, north/south, ...) recovered

Results: accelerations

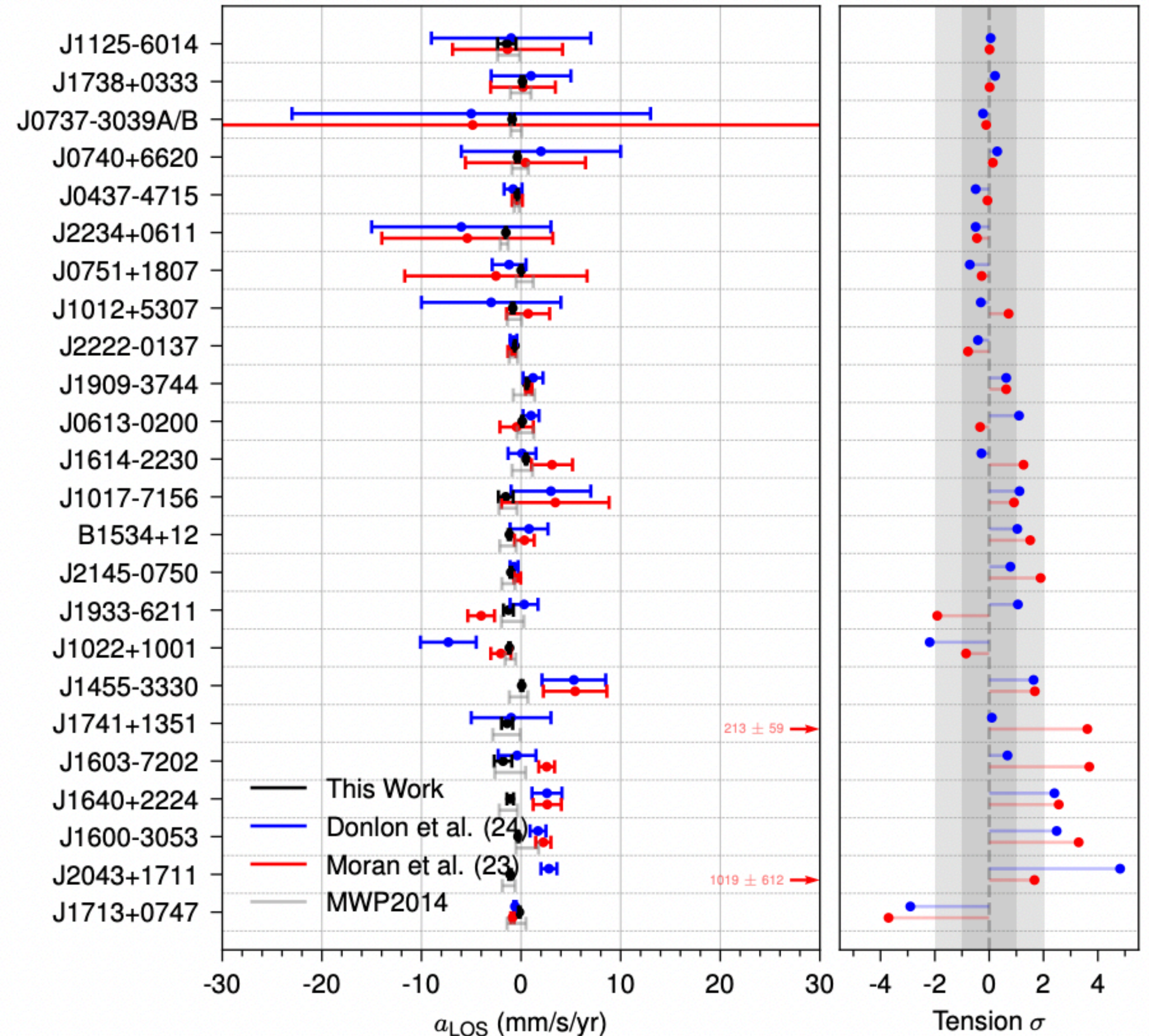
DS+ Putney et al 2512.09989

Also excellent overall agreement with recent measurements of LOS Galactic accelerations using binary pulsars.

4 out of 24 pulsars show $>2\sigma$ discrepancies with both measurements — could be evidence for localized disequilibrium in the Milky Way.

Our error bars include:

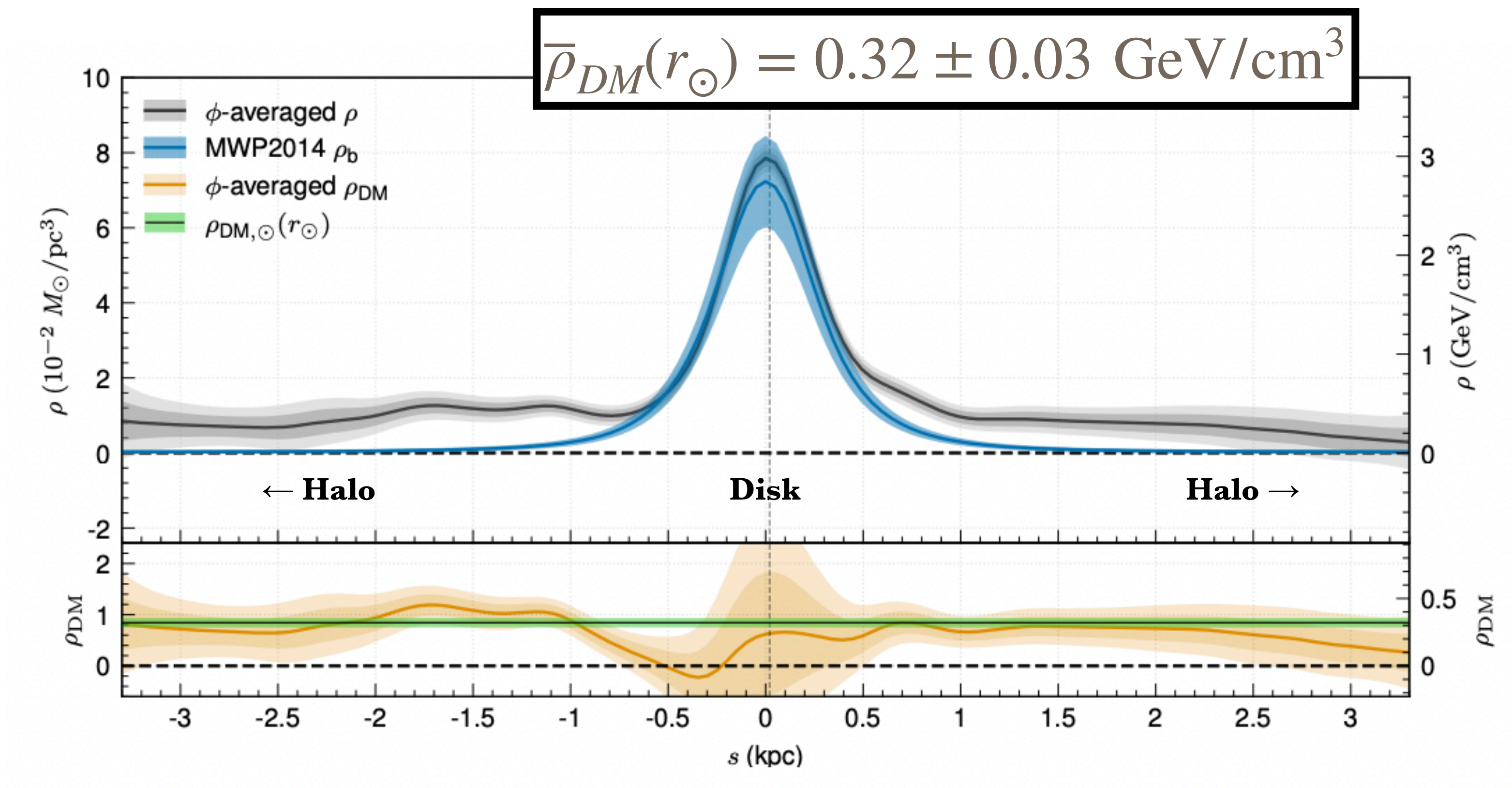
- finite training statistics
- training variance
- Gaia measurement errors



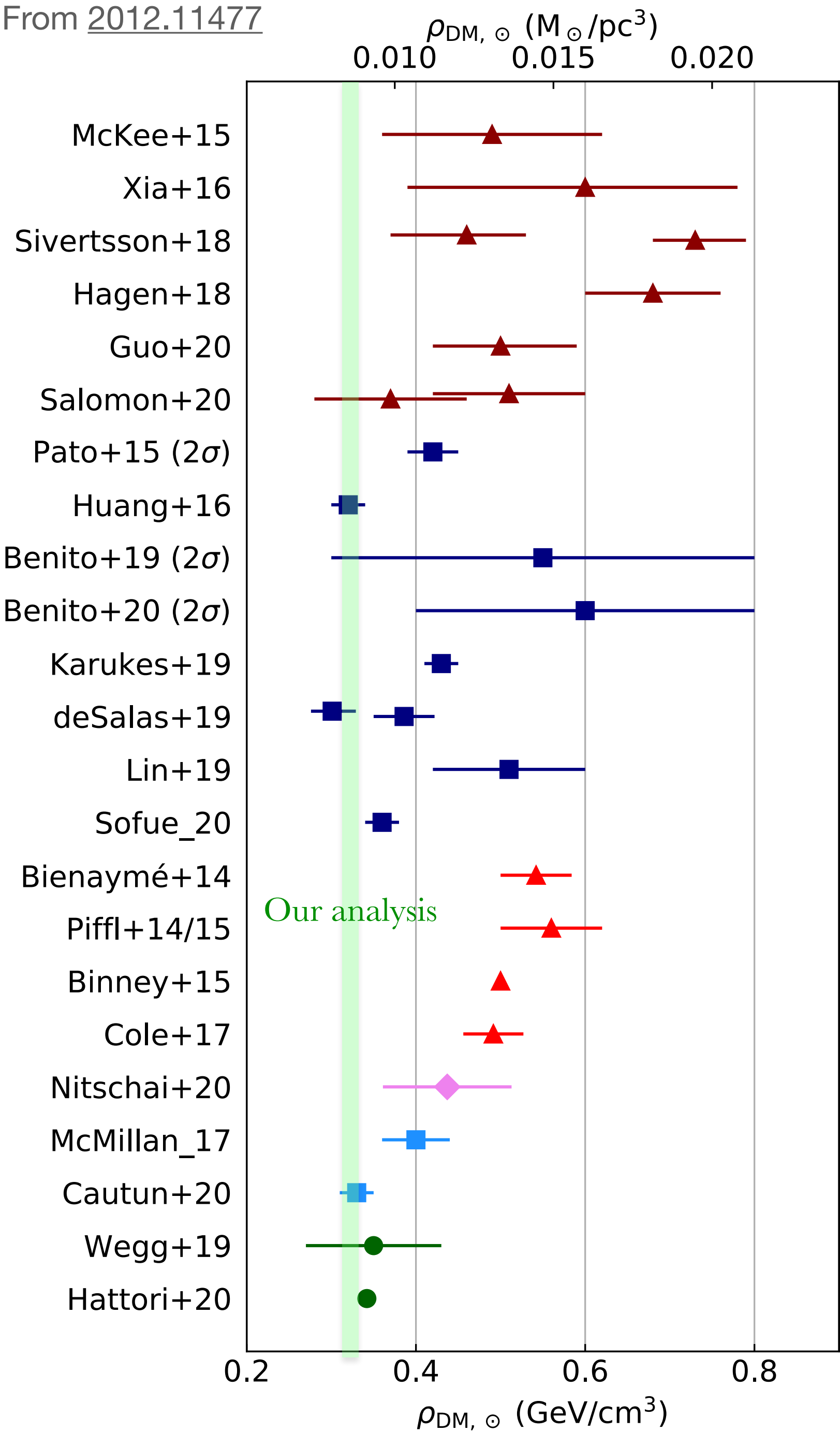
Results: mass density

DS+ Putney et al 2512.09989

From 2012.11477

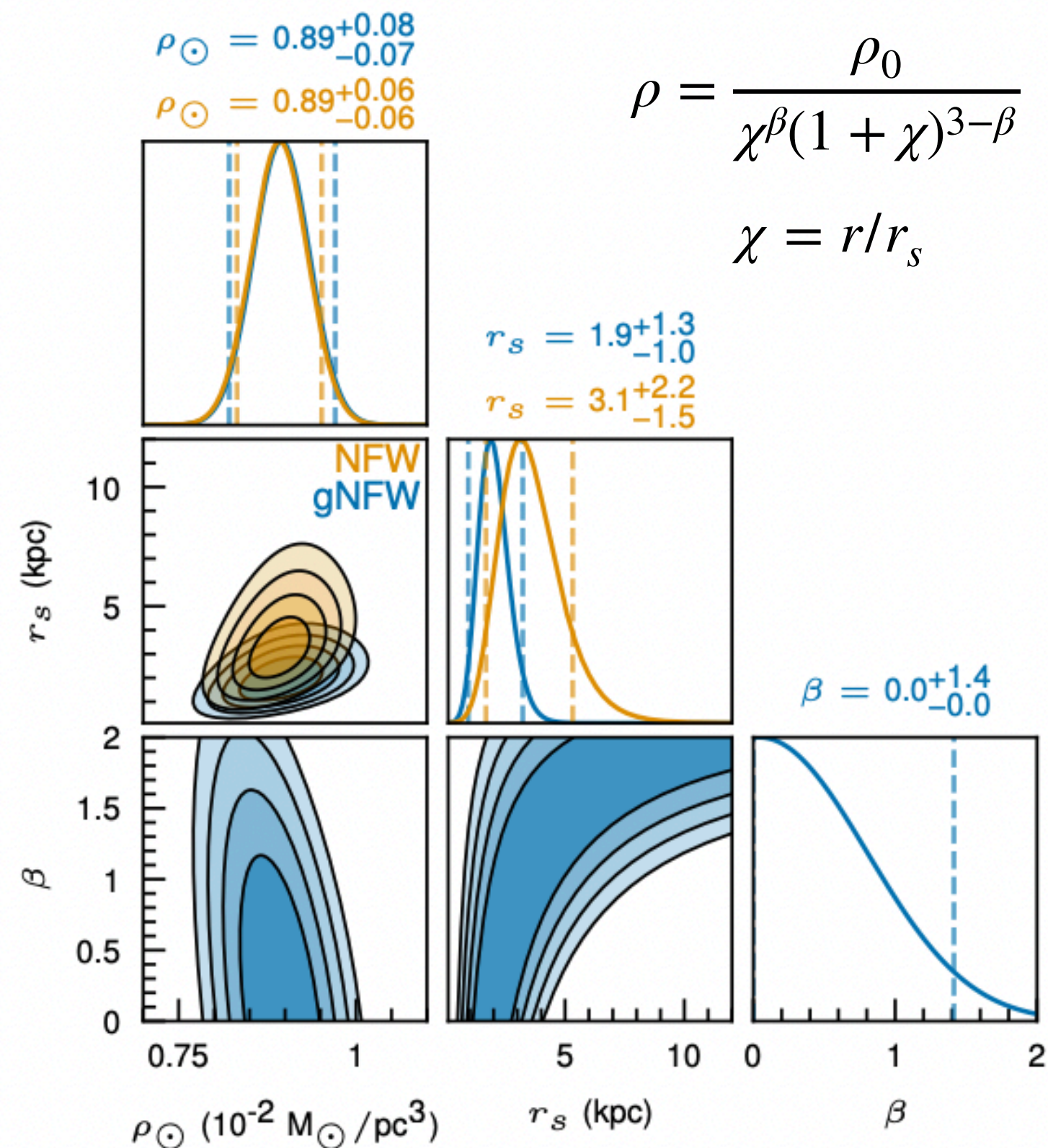


mass density vs arc height off disk, azimuthally averaged

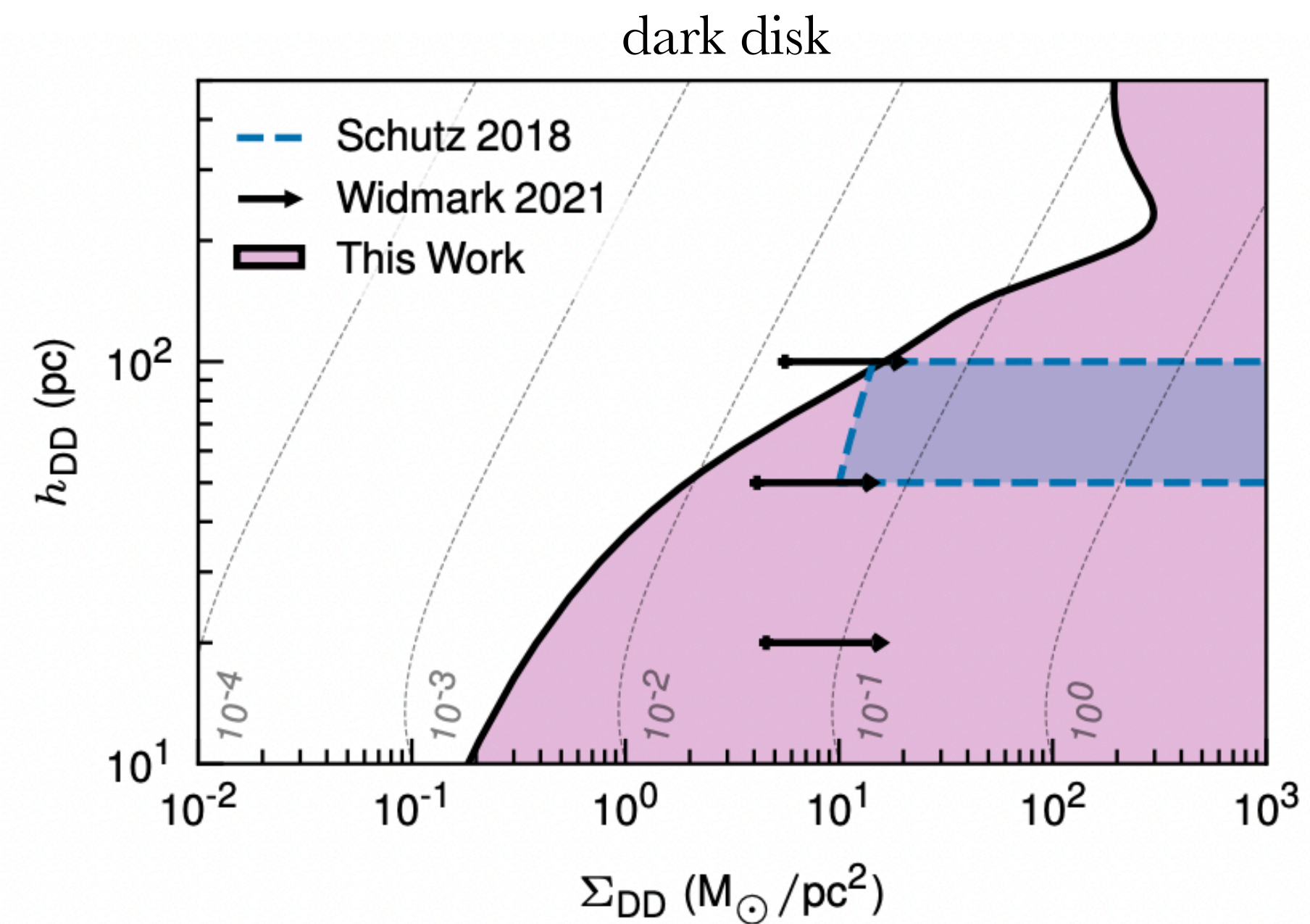


Results: constraints on models

DS+ Putney et al 2512.09989



prefer small scale radius
and highly cored profile

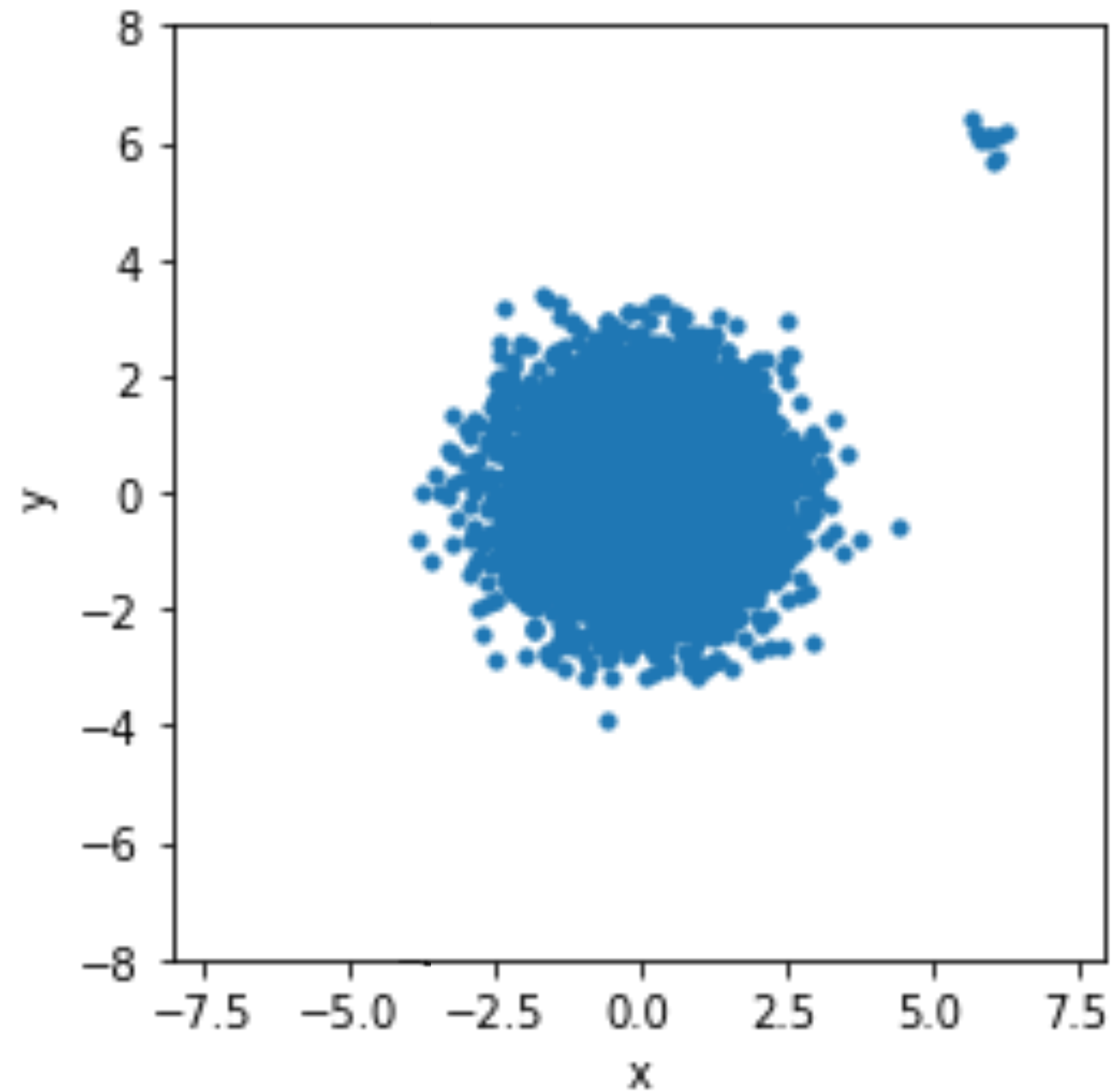


strongest constraints to
date on thin dark disk

3. Anomaly Detection

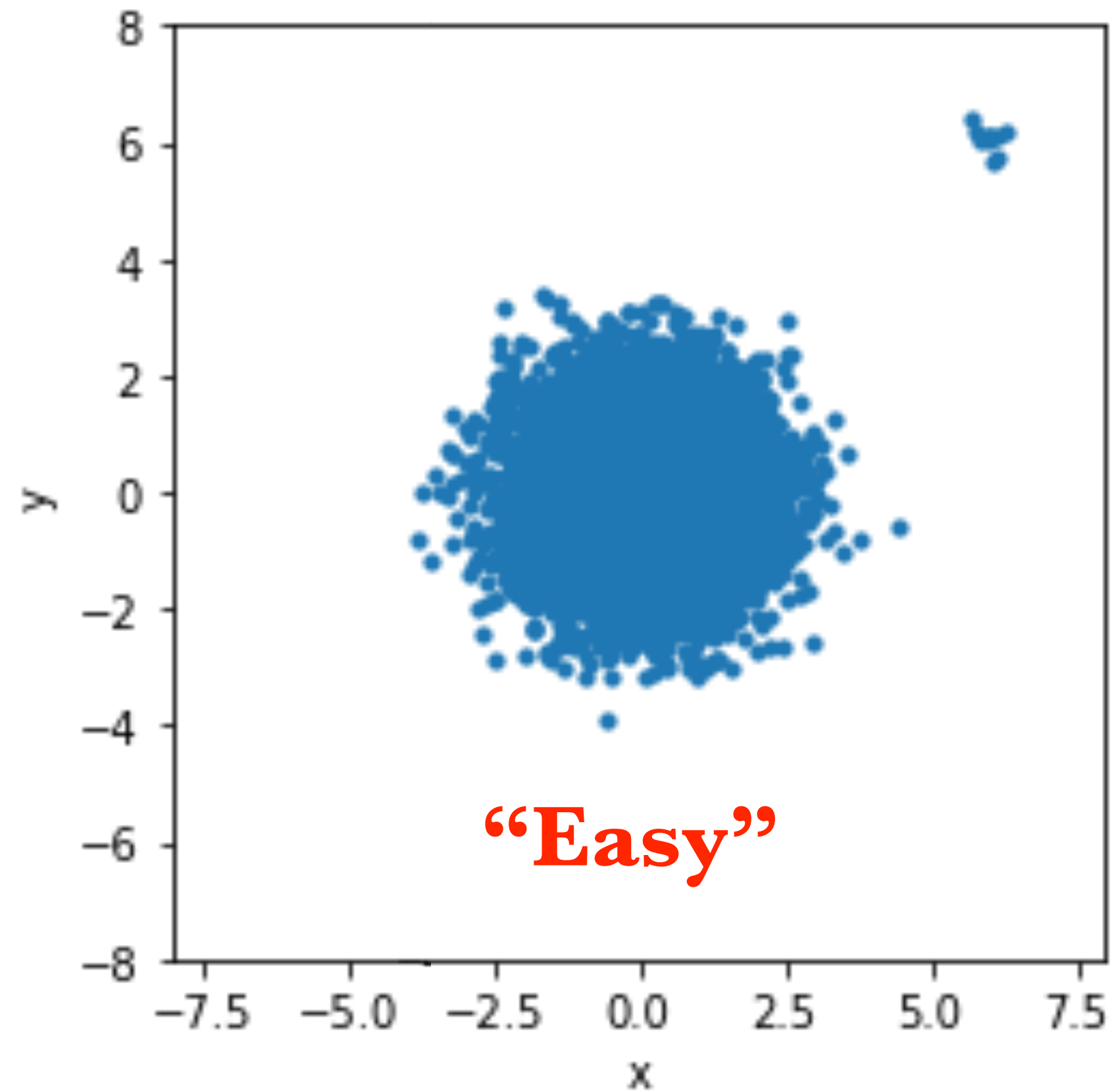
(Generation and density estimation)

Types of Anomalies: Outliers vs Overdensities



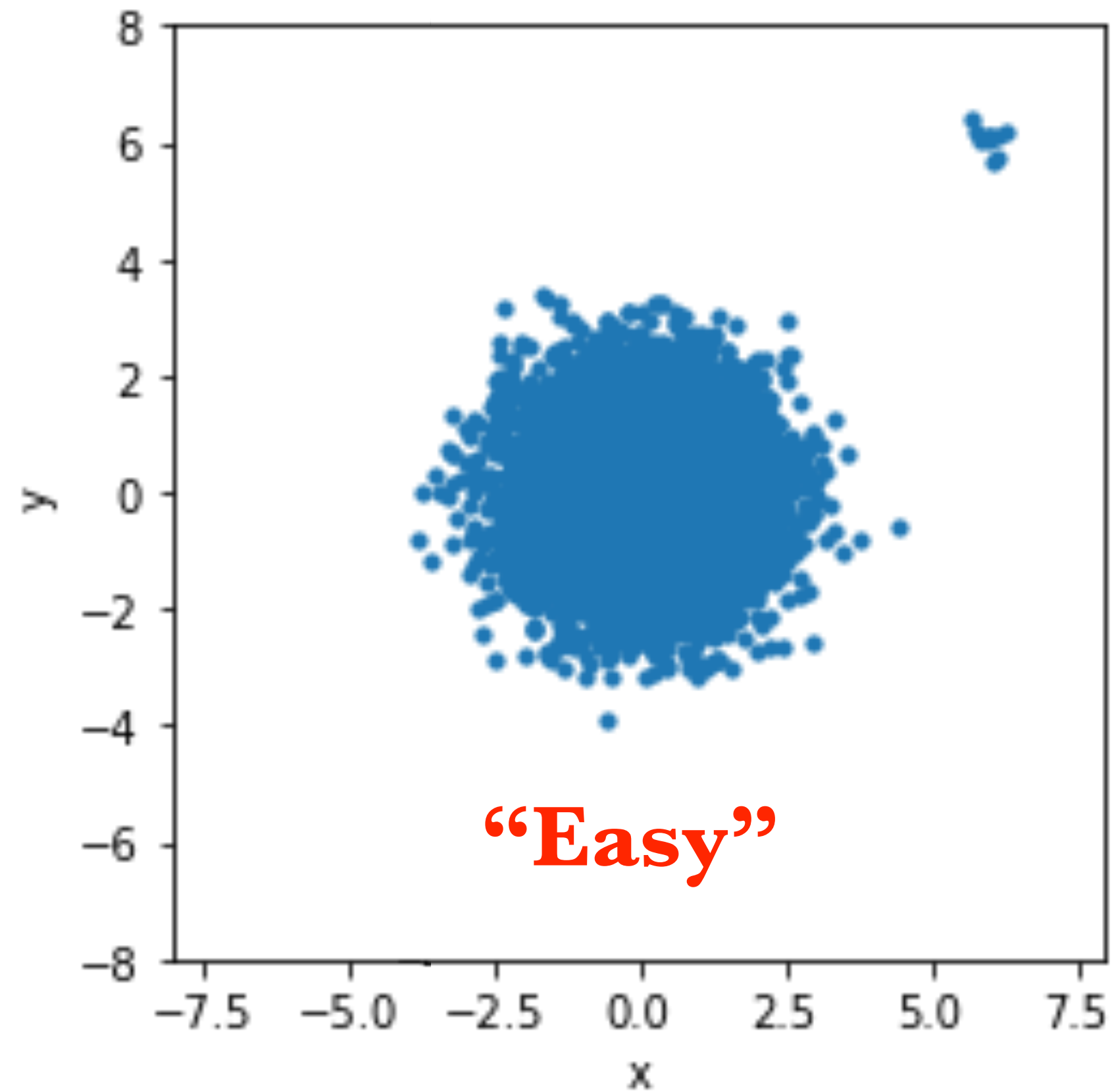
Outliers

Types of Anomalies: Outliers vs Overdensities

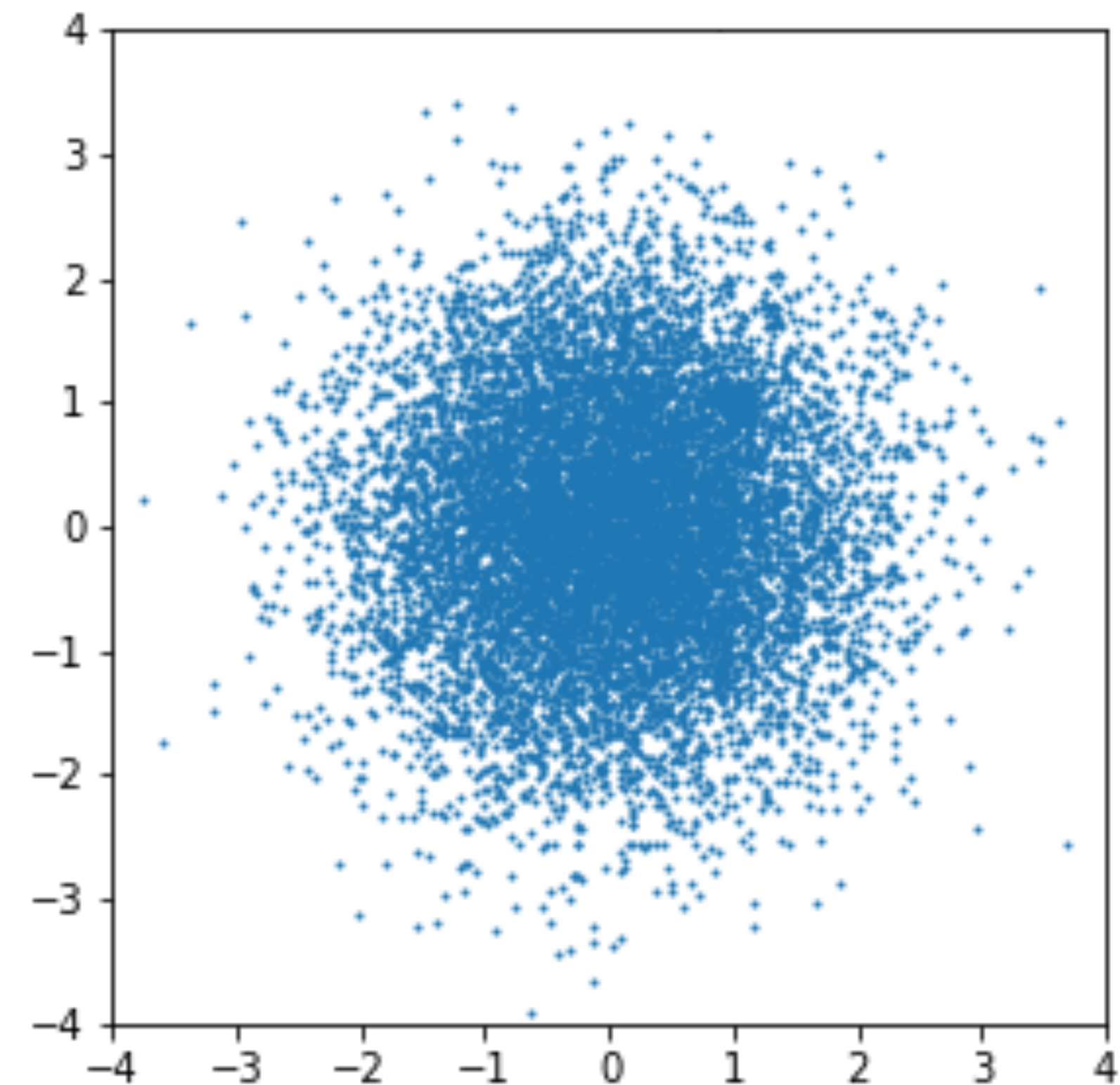


Outliers

Types of Anomalies: Outliers vs Overdensities

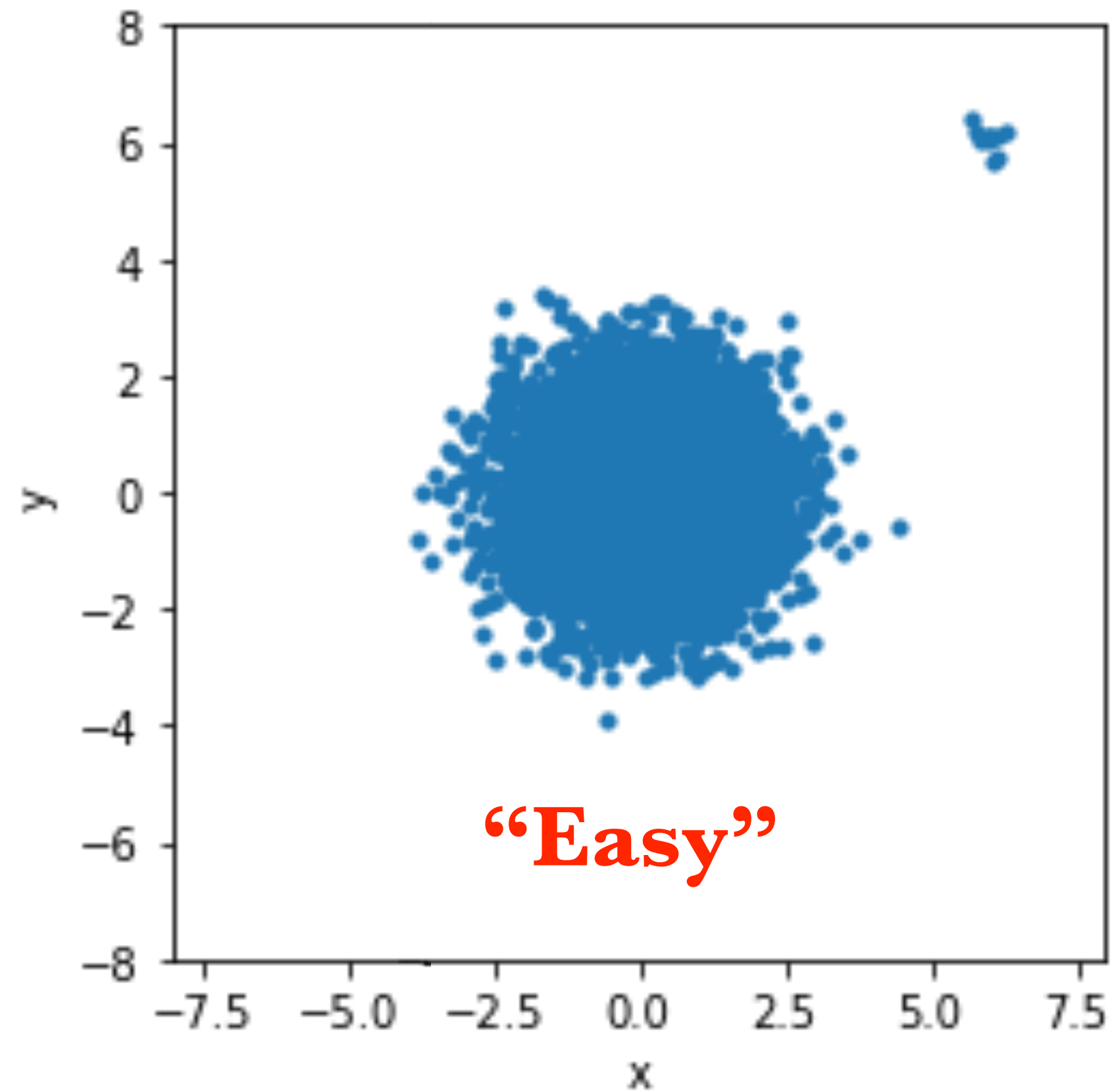


Outliers

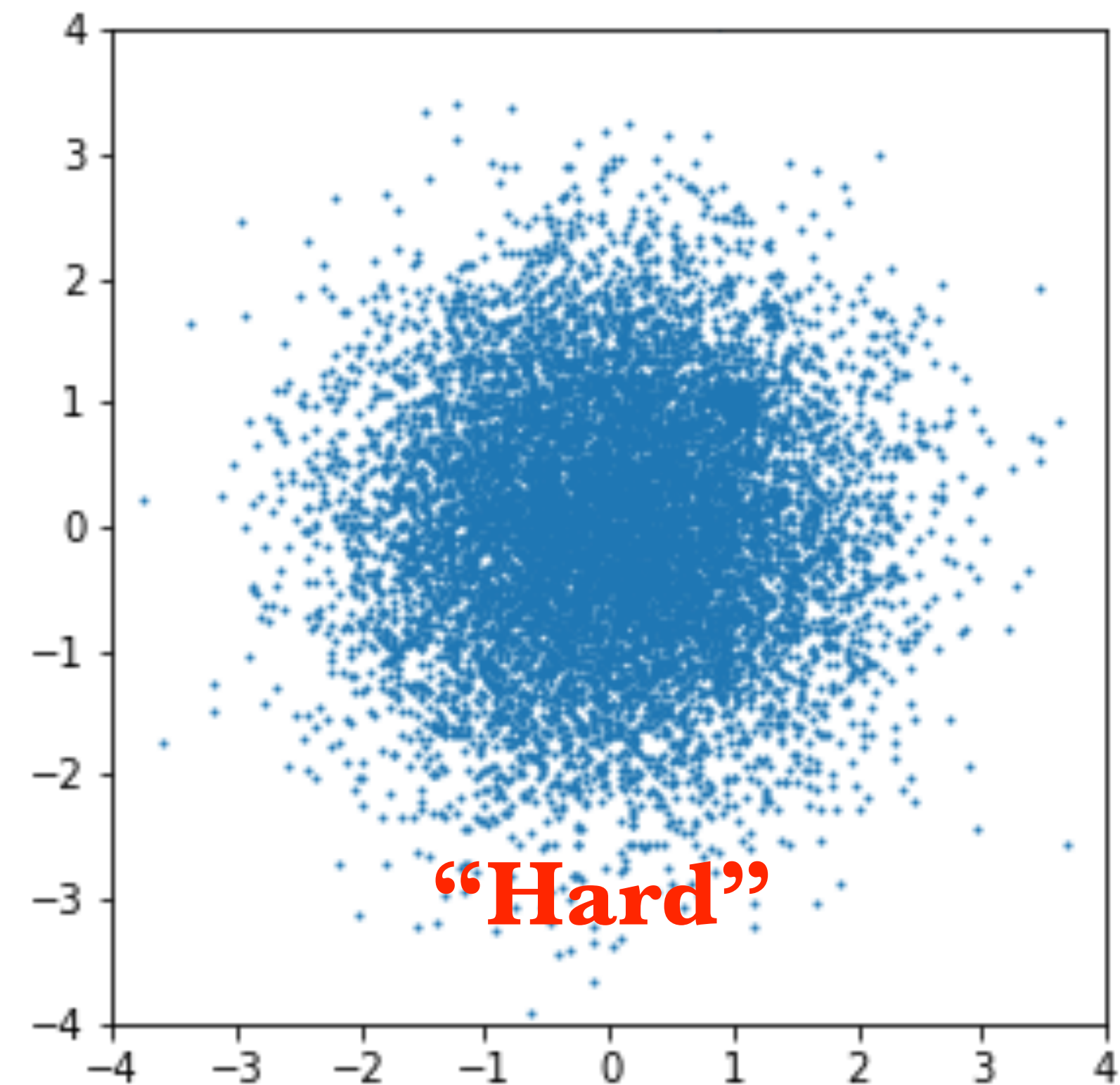


Overdensities

Types of Anomalies: Outliers vs Overdensities

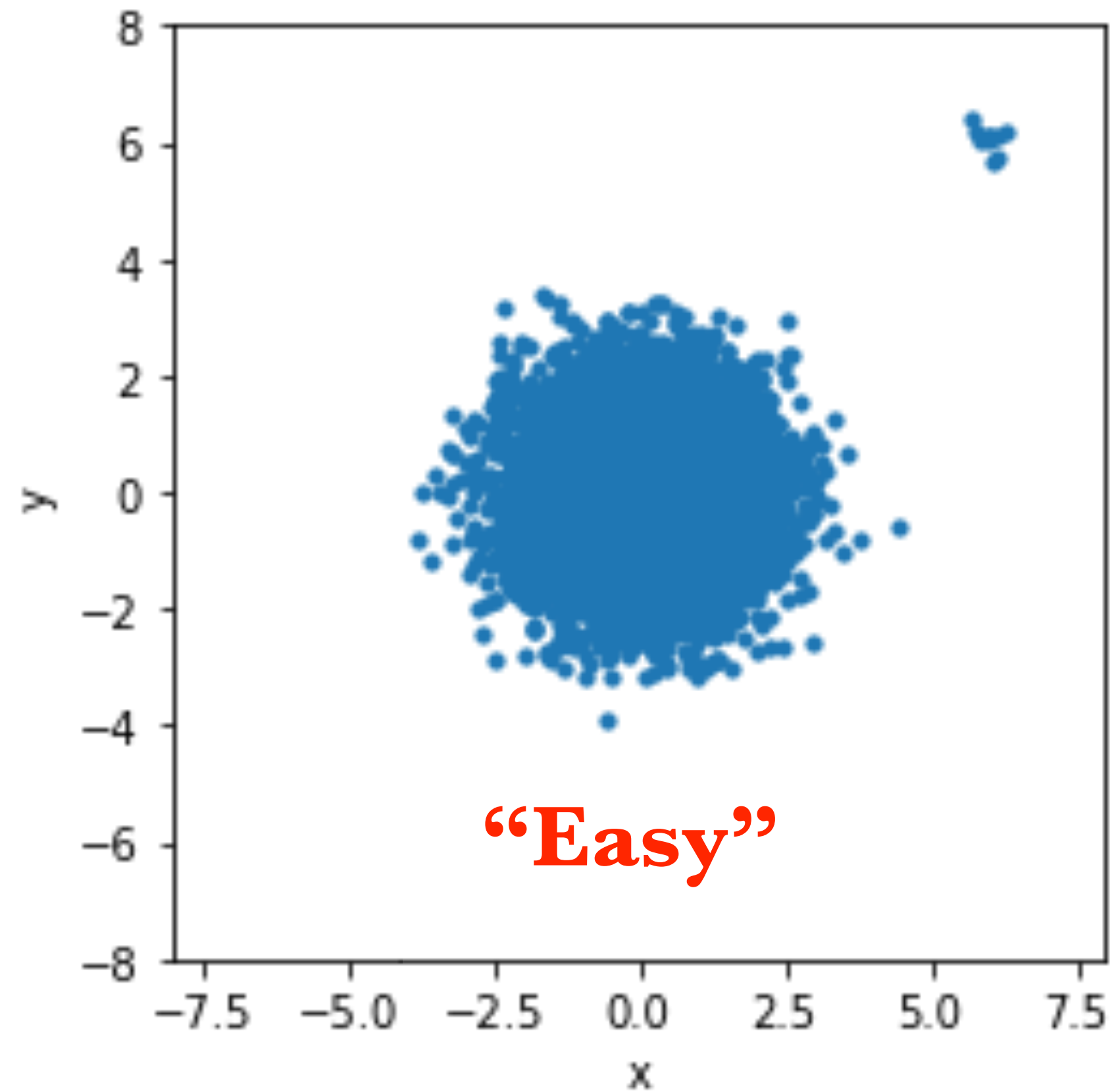


Outliers

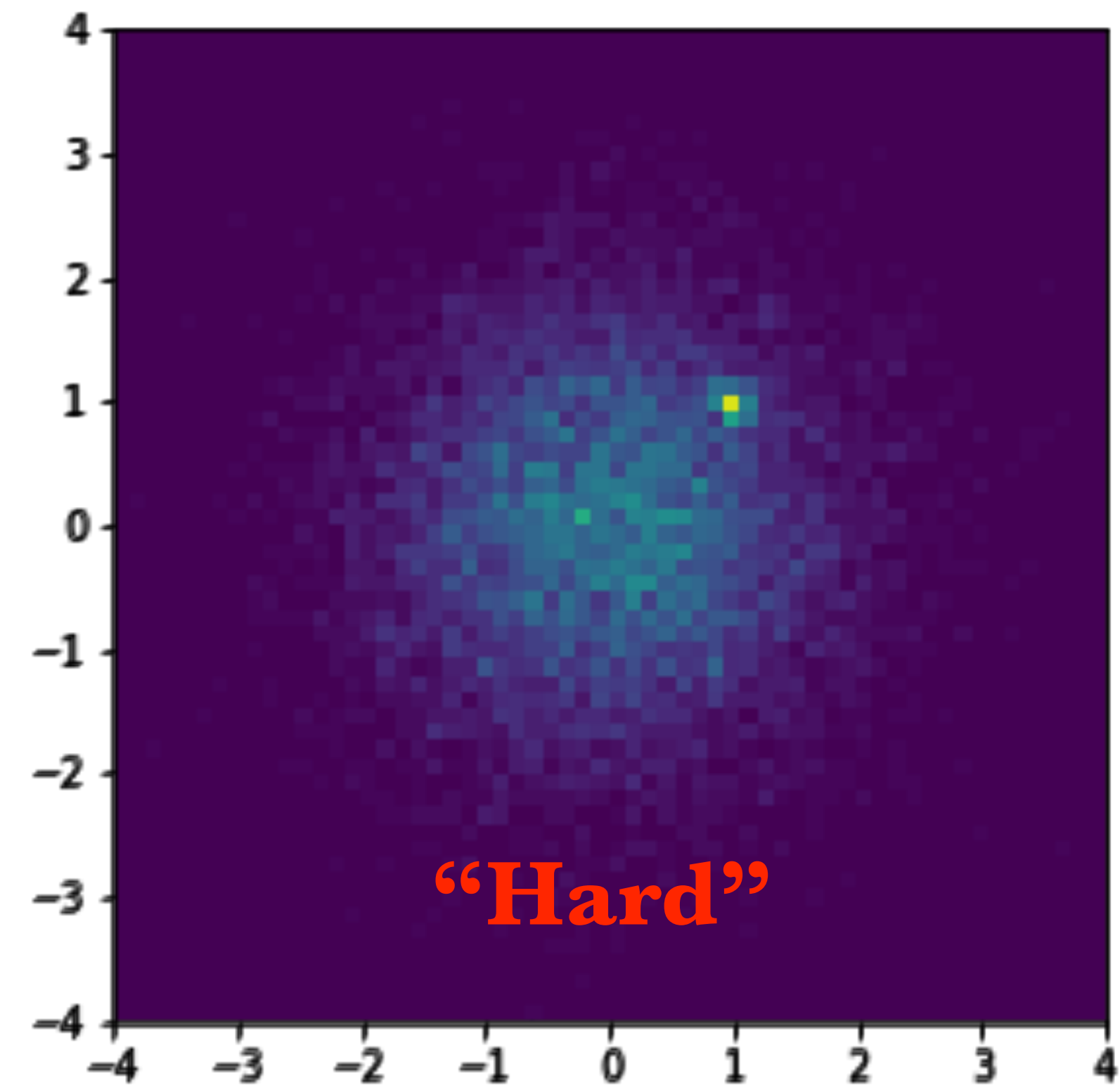


Overdensities

Types of Anomalies: Outliers vs Overdensities

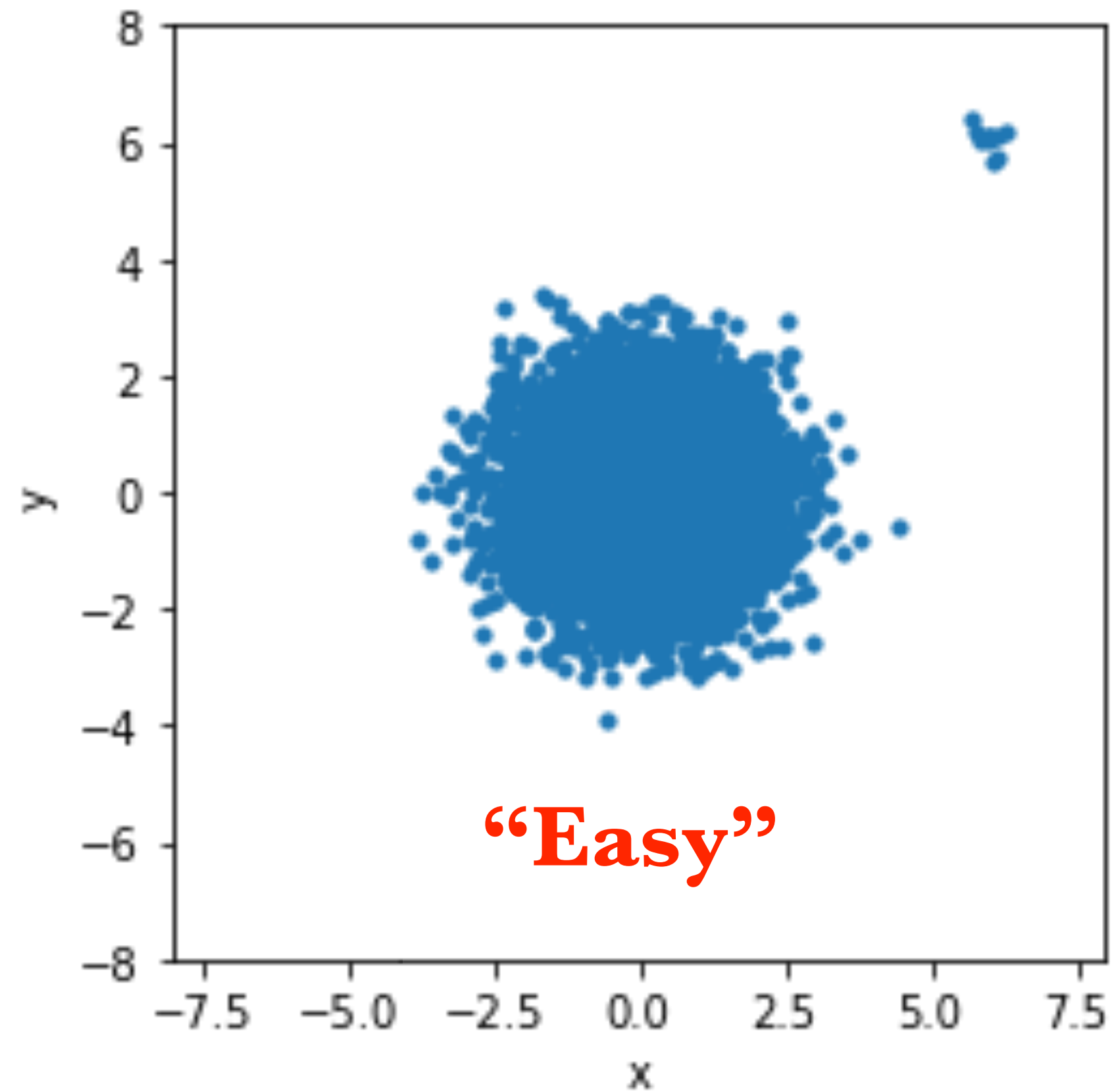


Outliers

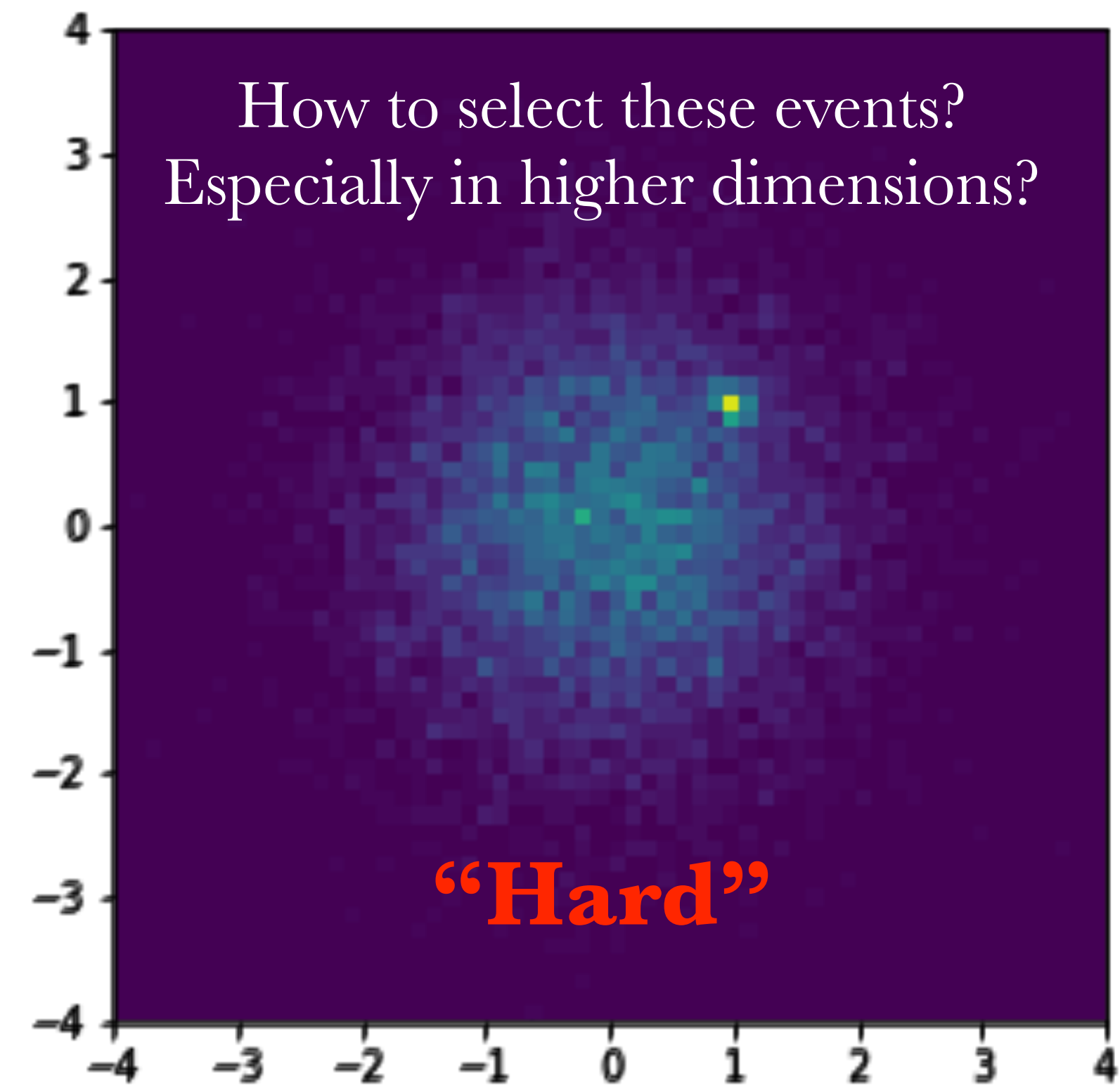


Overdensities

Types of Anomalies: Outliers vs Overdensities



Outliers

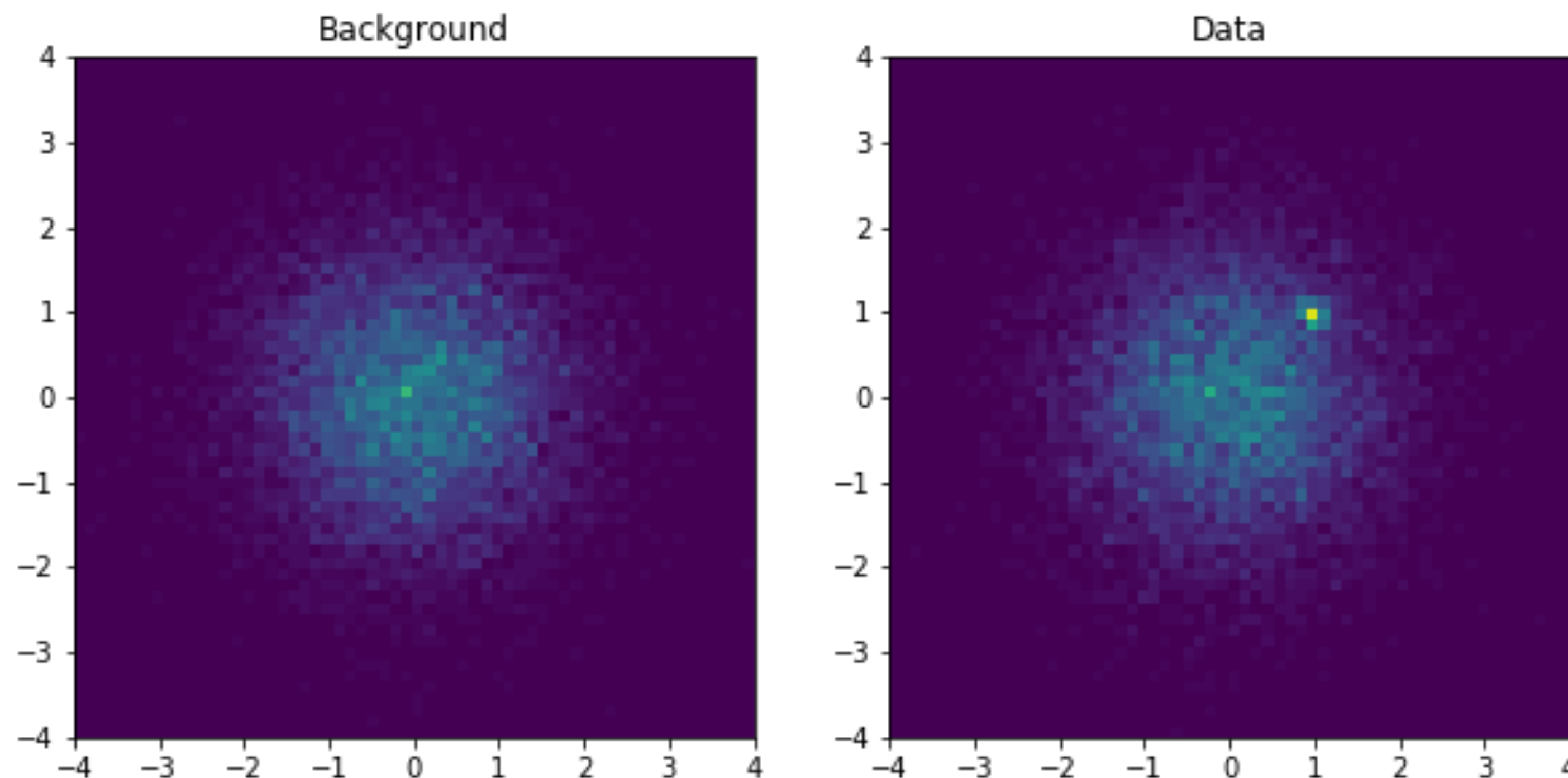


Overdensities

Overdensity anomaly score

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}$$

One strategy: learn the ***overdensity anomaly score*** from data.

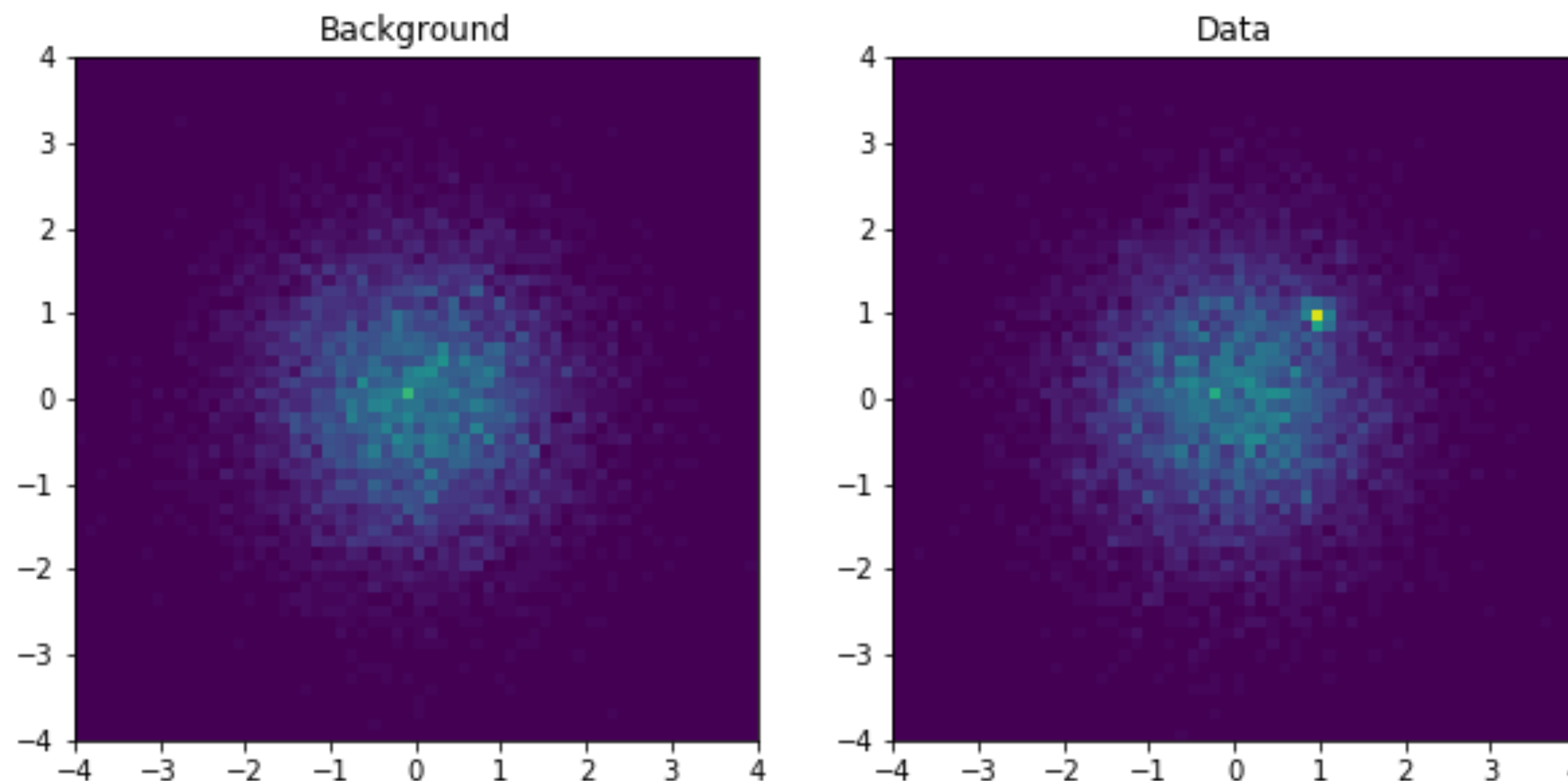


Overdensity anomaly score

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}$$

One strategy: learn the ***overdensity anomaly score*** from data.

Can use
density estimation
or
classifiers

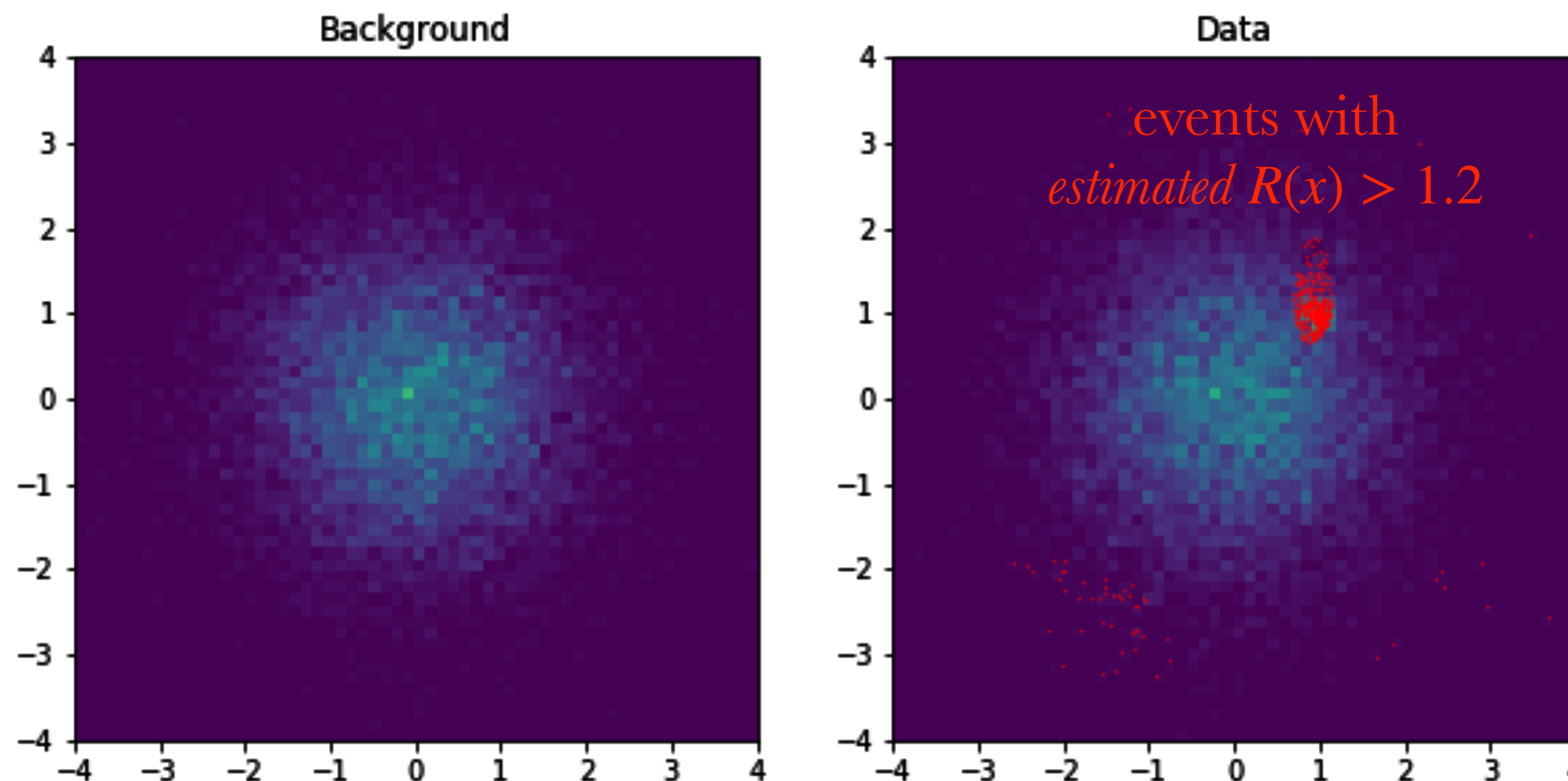


Overdensity anomaly score

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}$$

One strategy: learn the ***overdensity anomaly score*** from data.

Can use
density estimation
or
classifiers

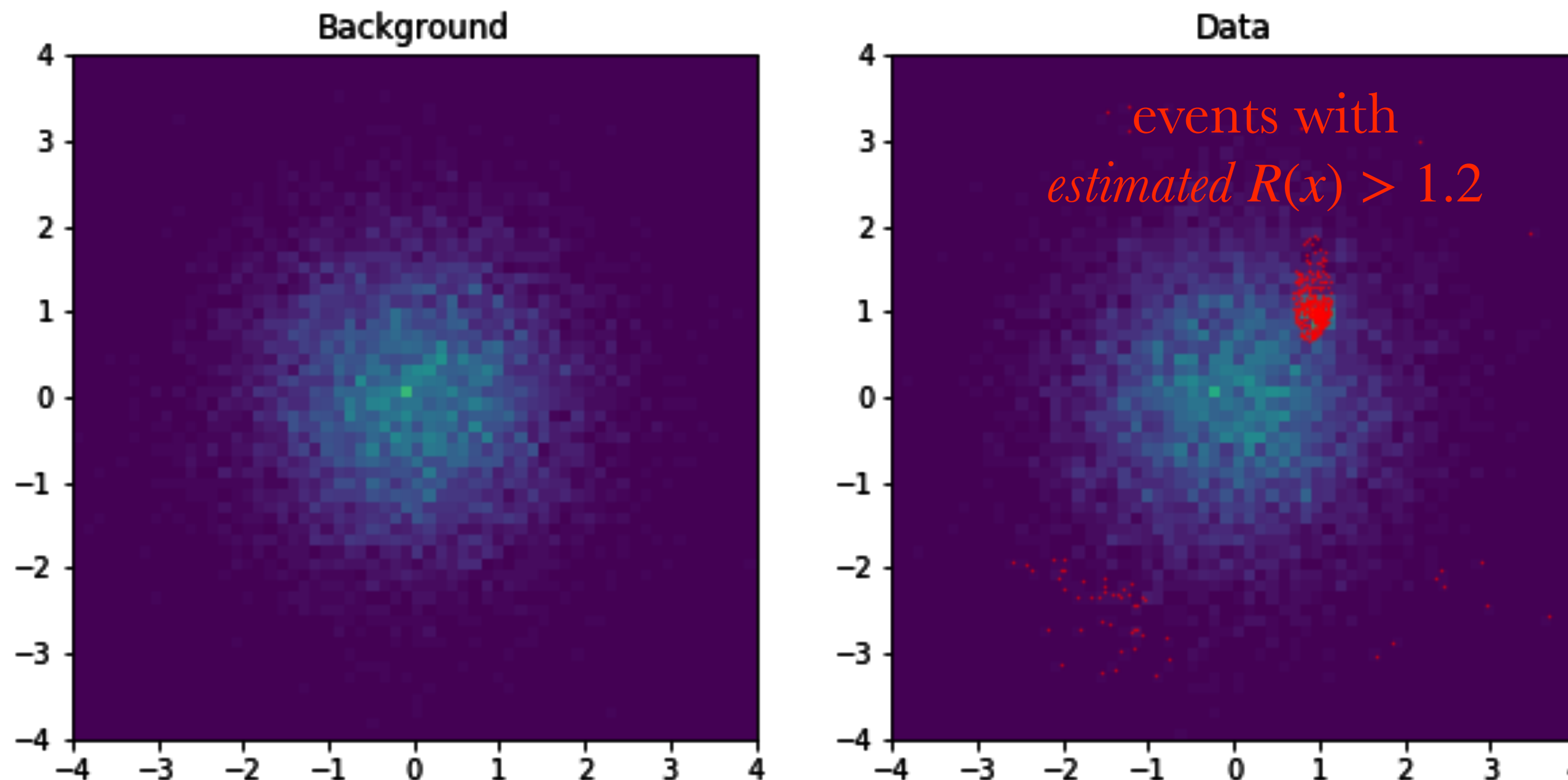


Overdensity anomaly score

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}$$

One strategy: learn the ***overdensity anomaly score*** from data.

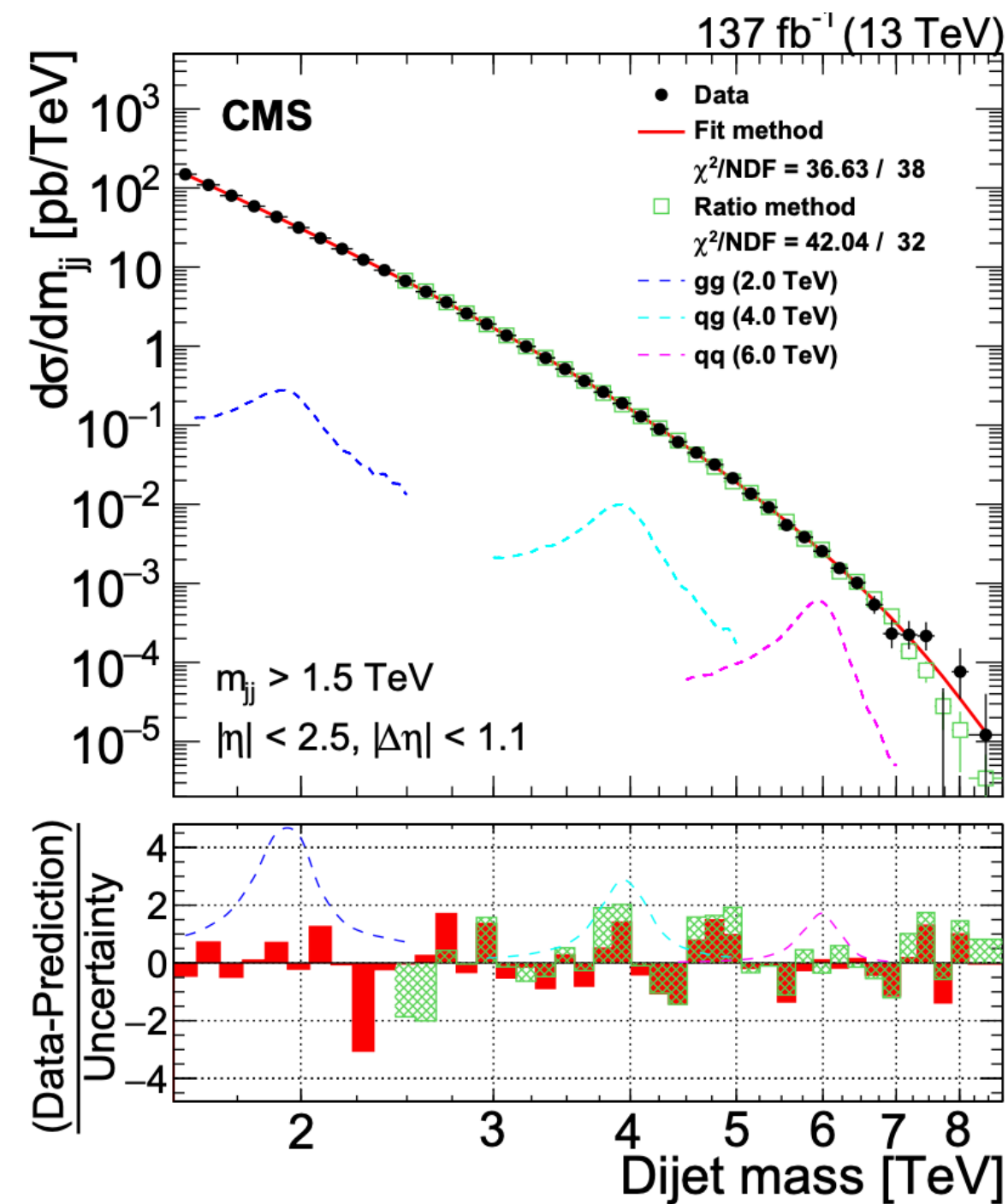
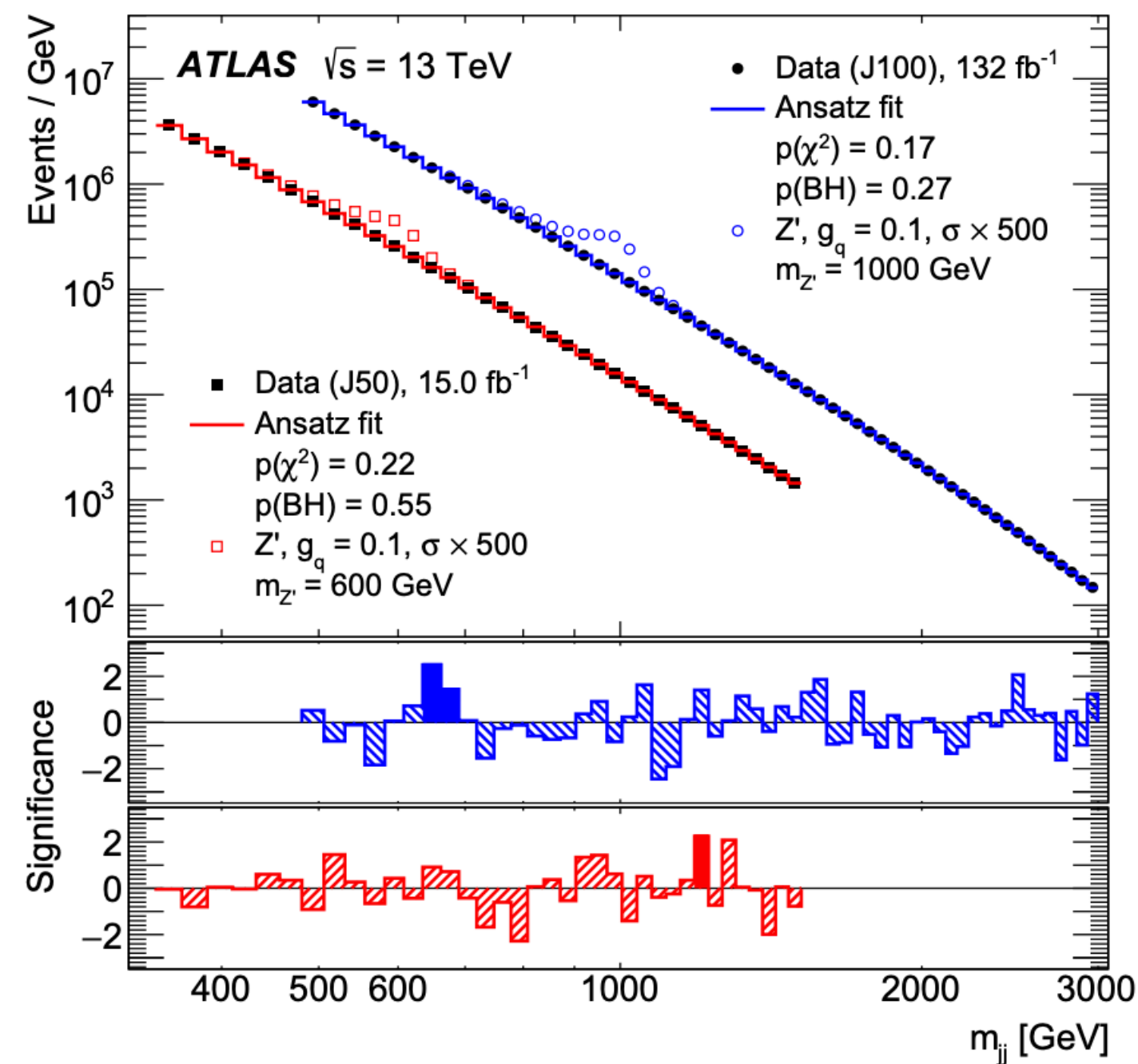
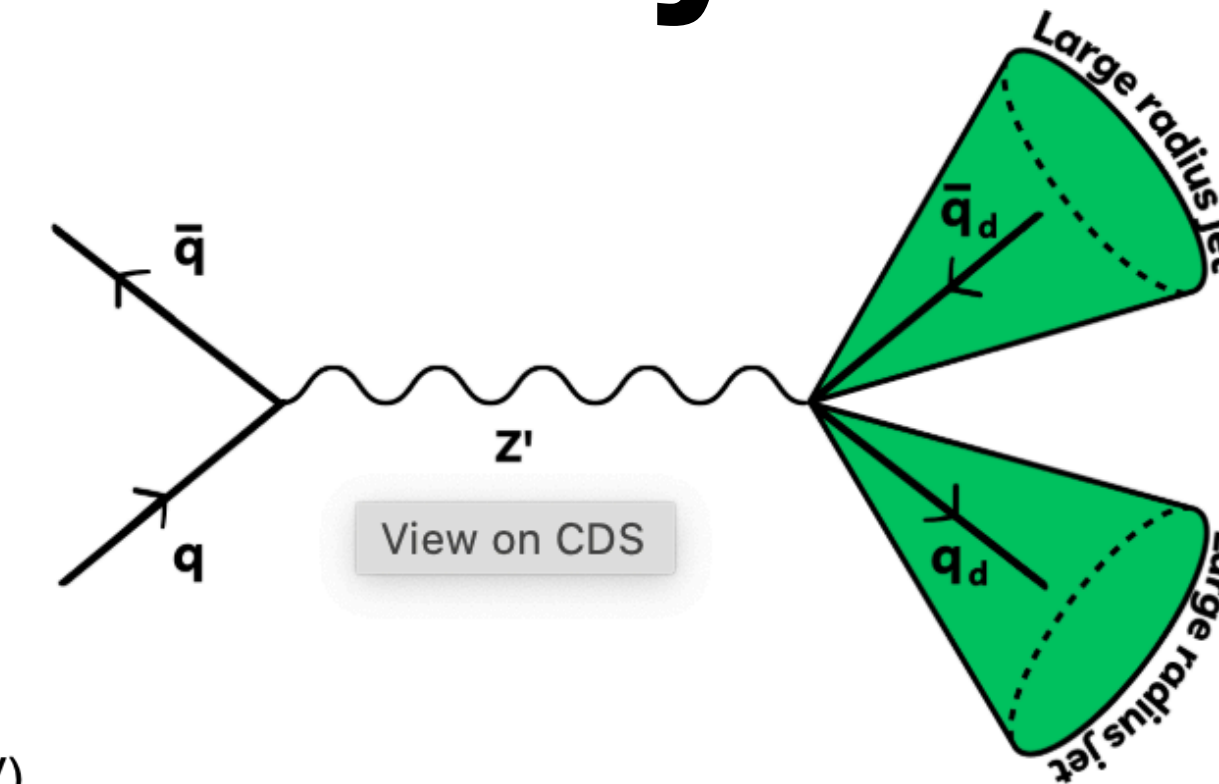
Can use
density estimation
or
classifiers



But where does the background come from?

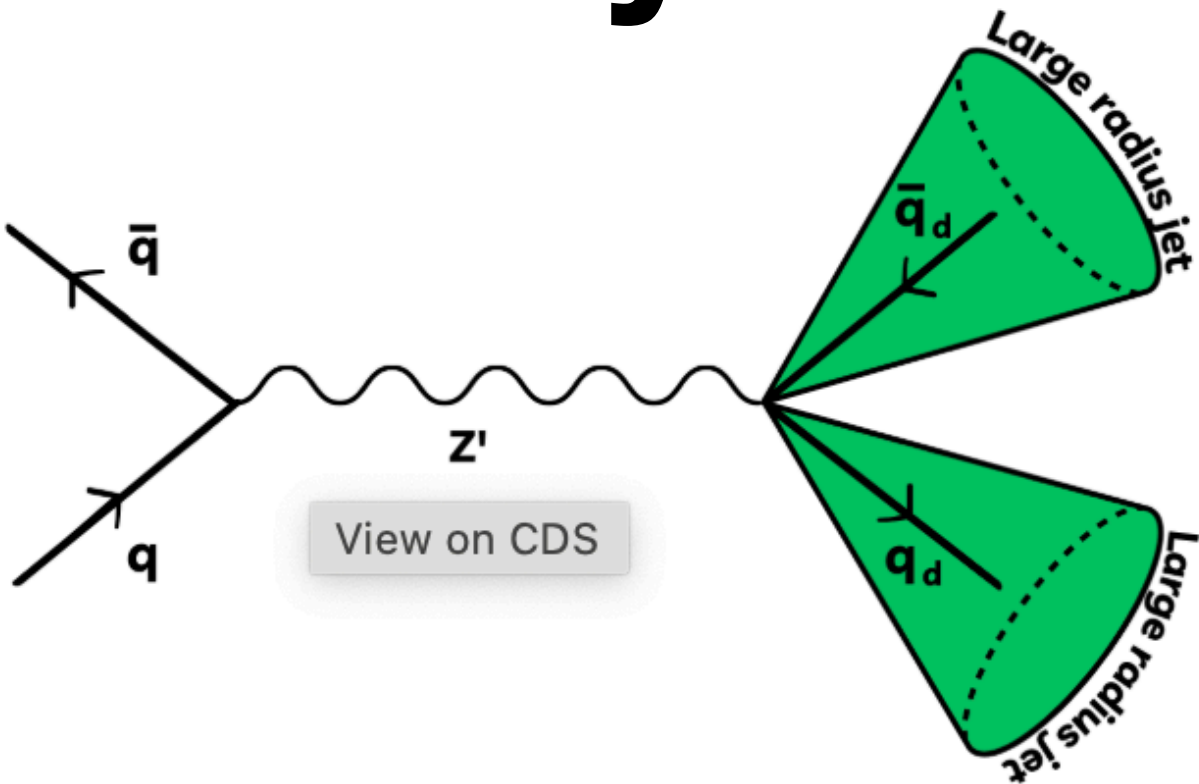
Resonant AD: A Sandbox for Overdensity Anomalies

Consider the humble dijet resonance search

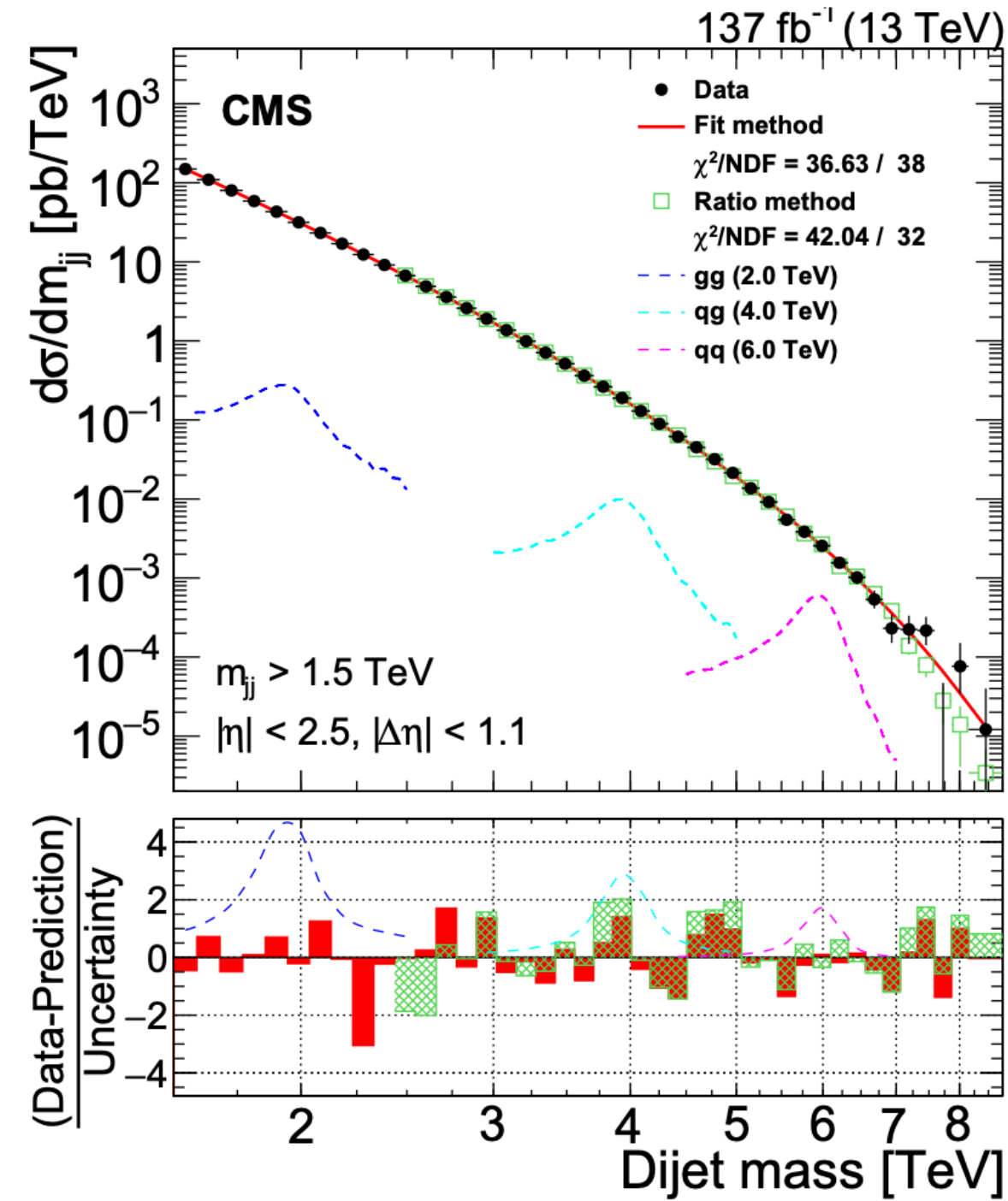
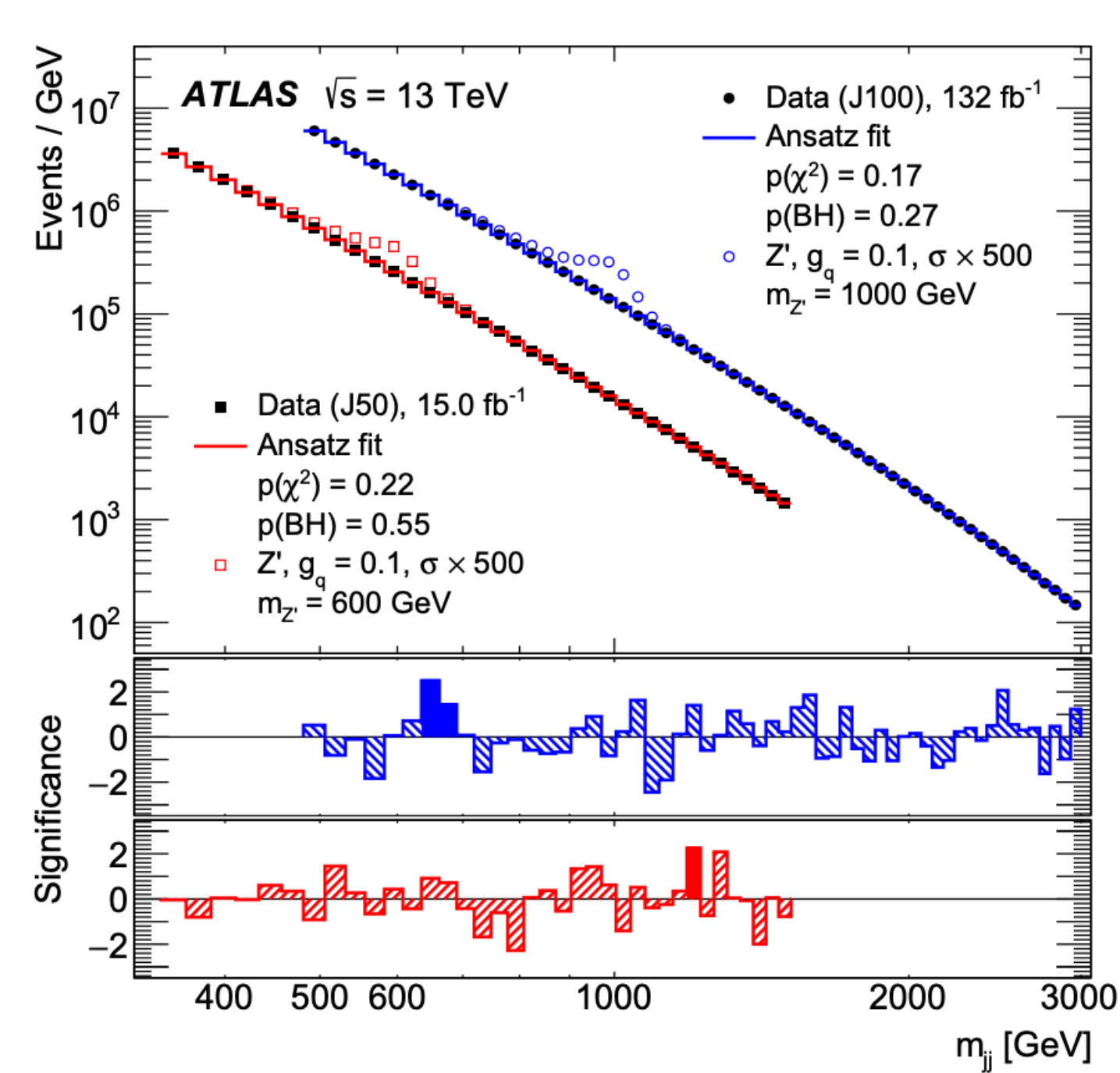


Resonant AD: A Sandbox for Overdensity Anomalies

Consider the humble dijet resonance search

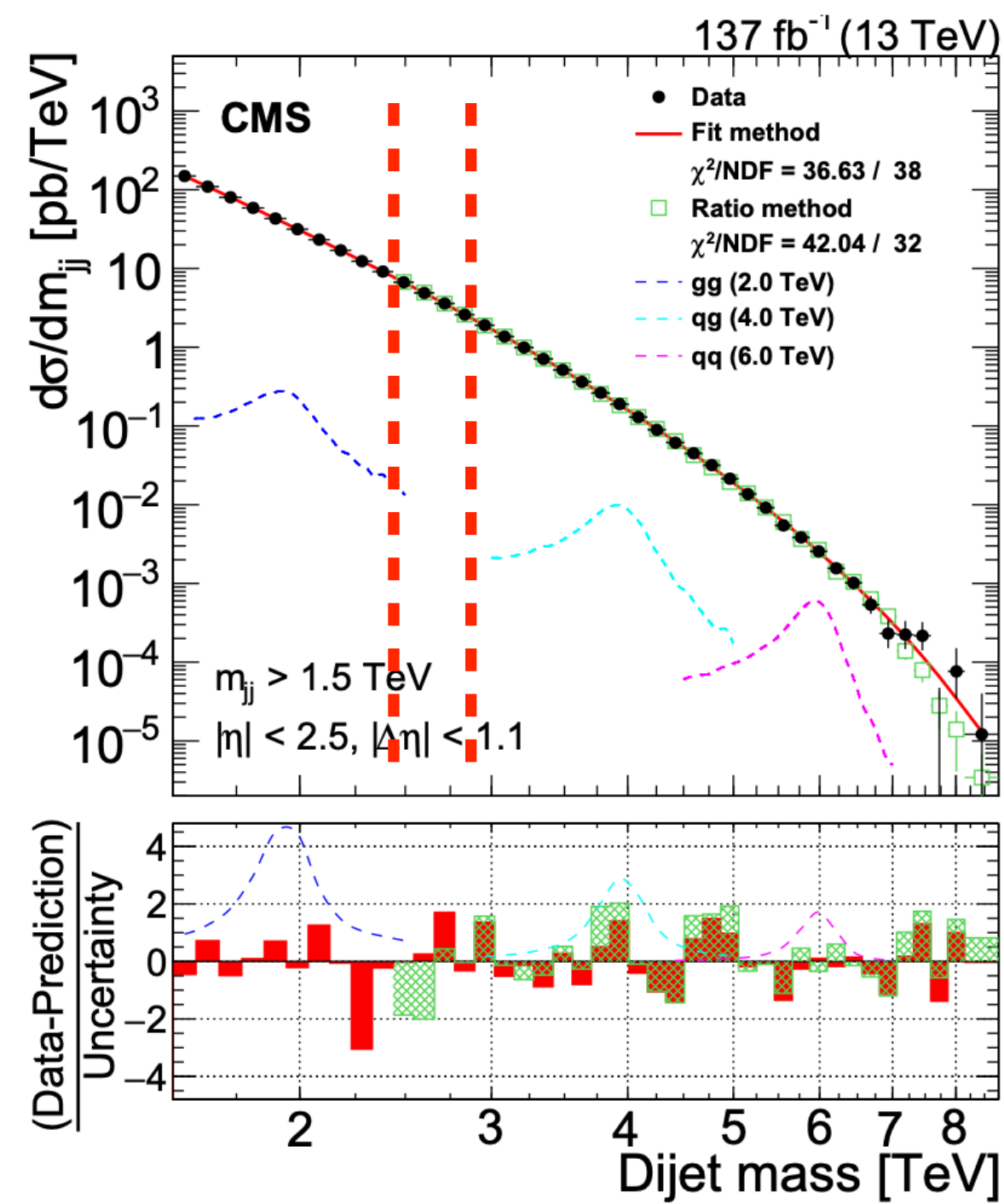
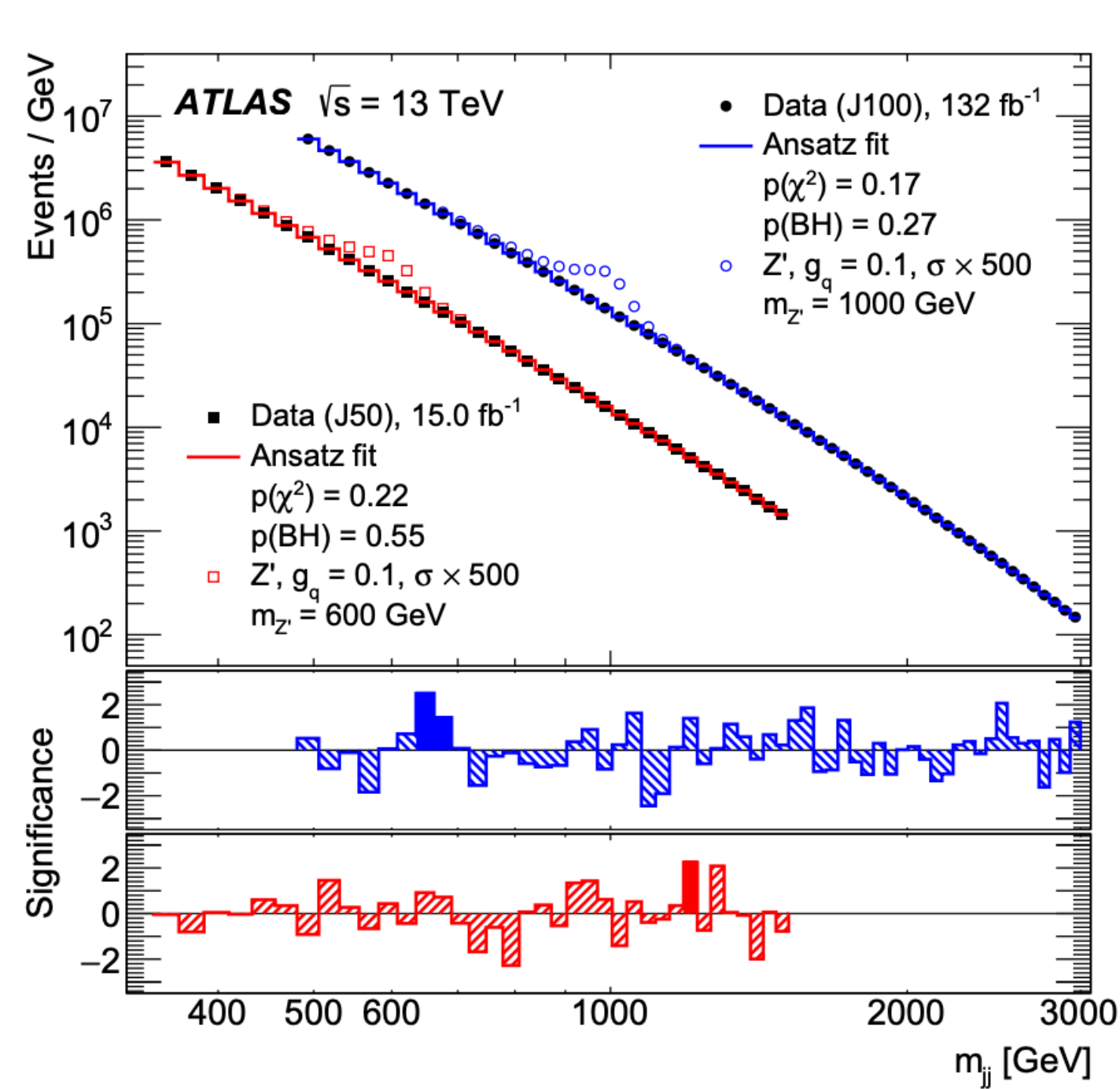
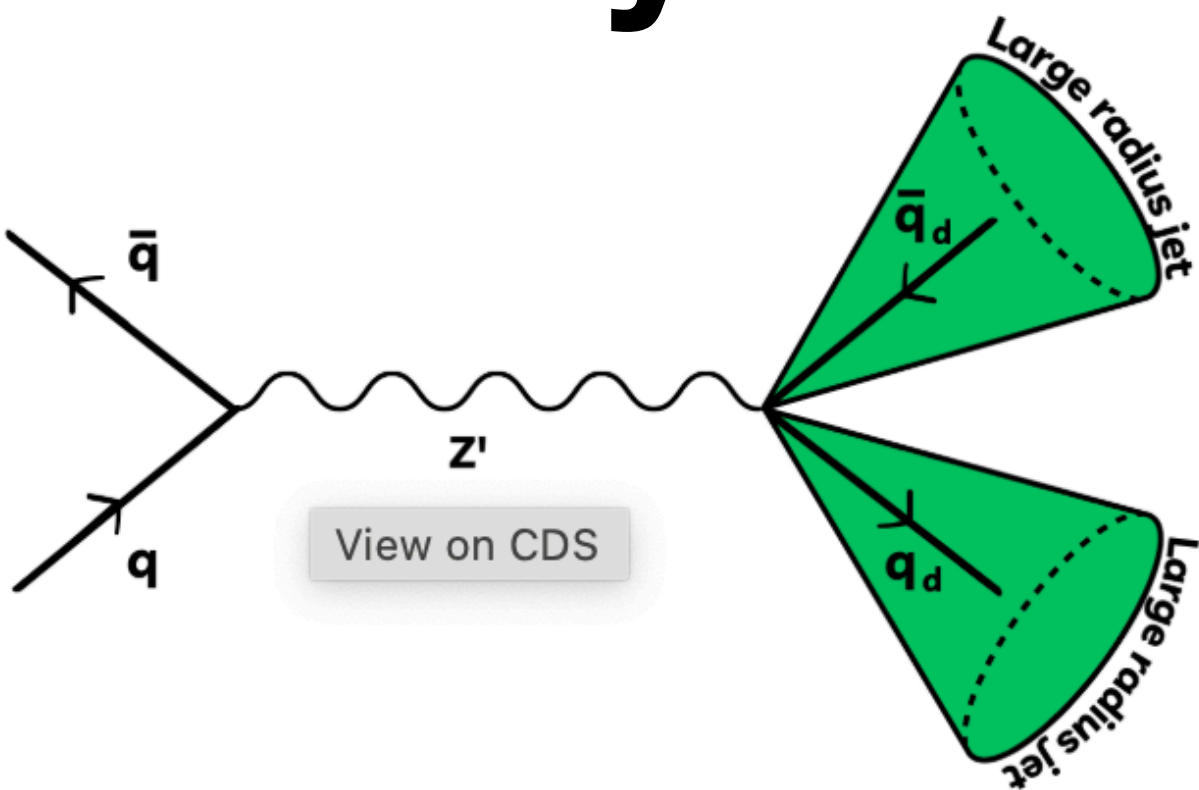


- The “bump hunt”: look for a bump over **smooth background**



Resonant AD: A Sandbox for Overdensity Anomalies

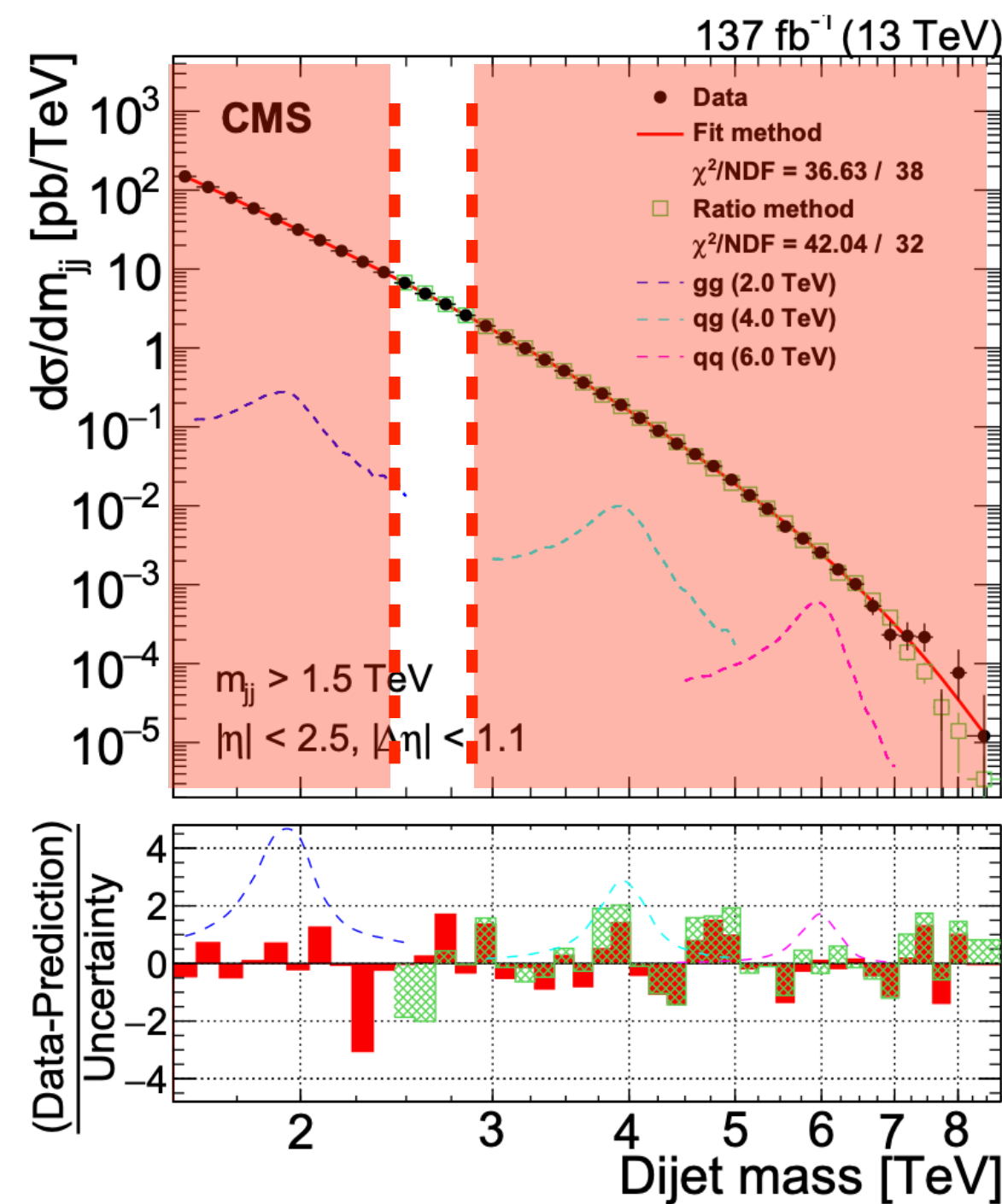
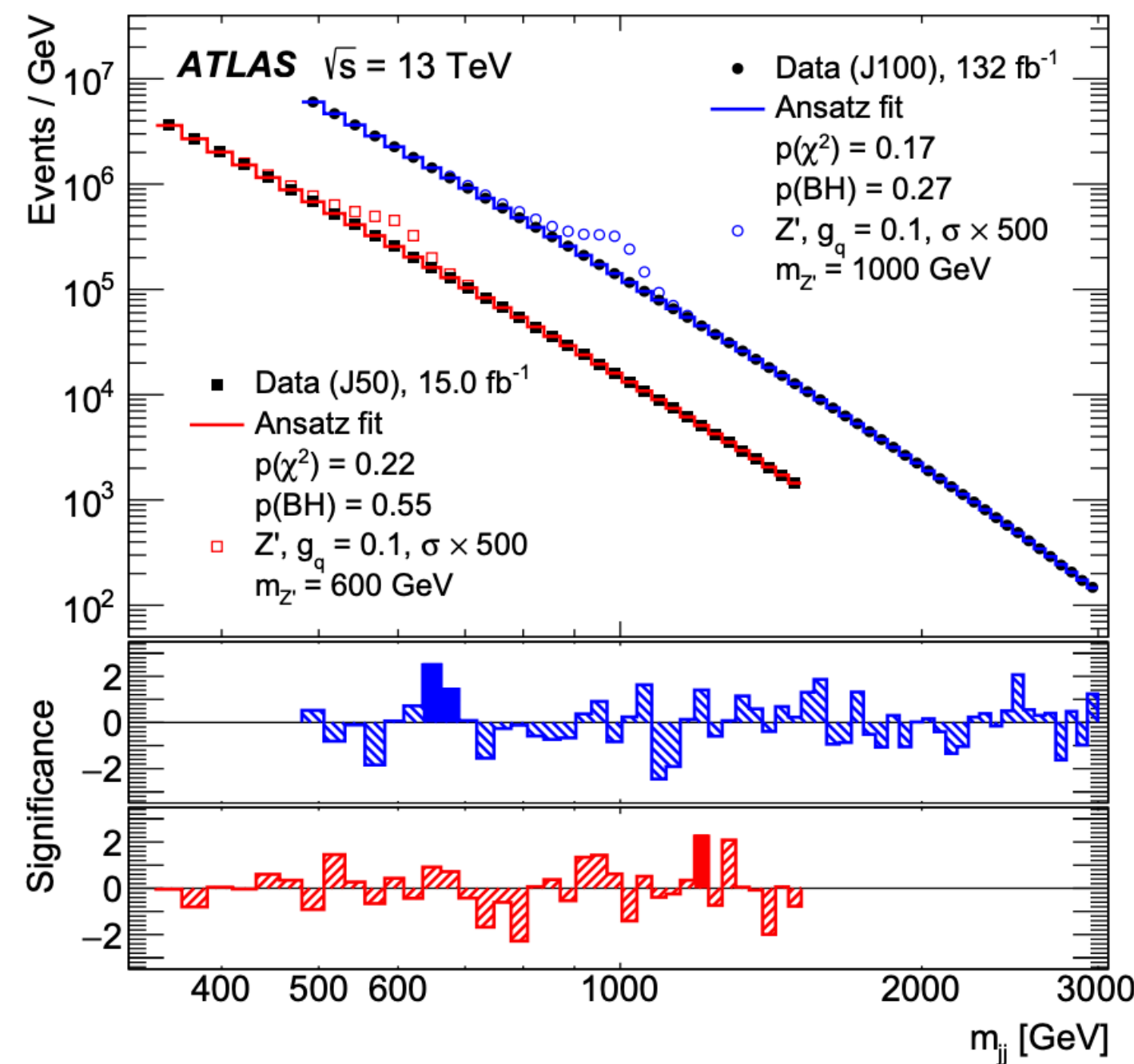
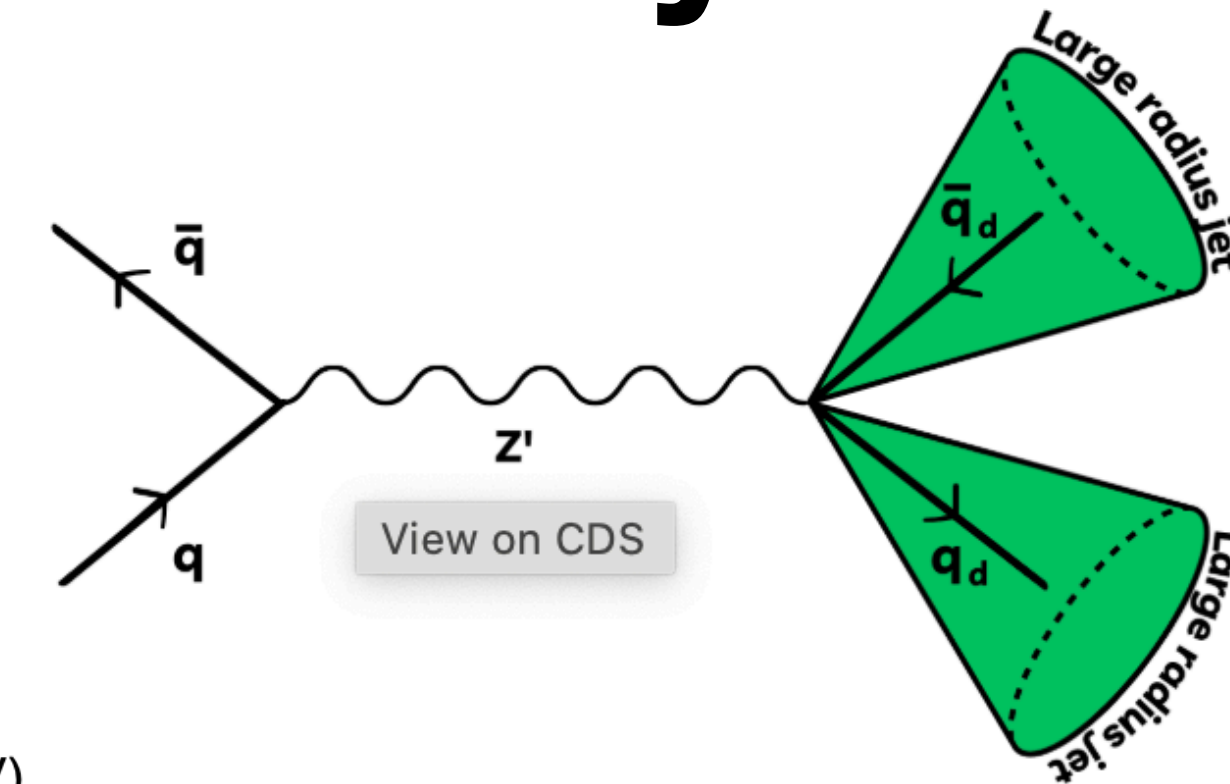
Consider the humble dijet resonance search



- The “bump hunt”: look for a bump over **smooth background**
- **Sliding window** search

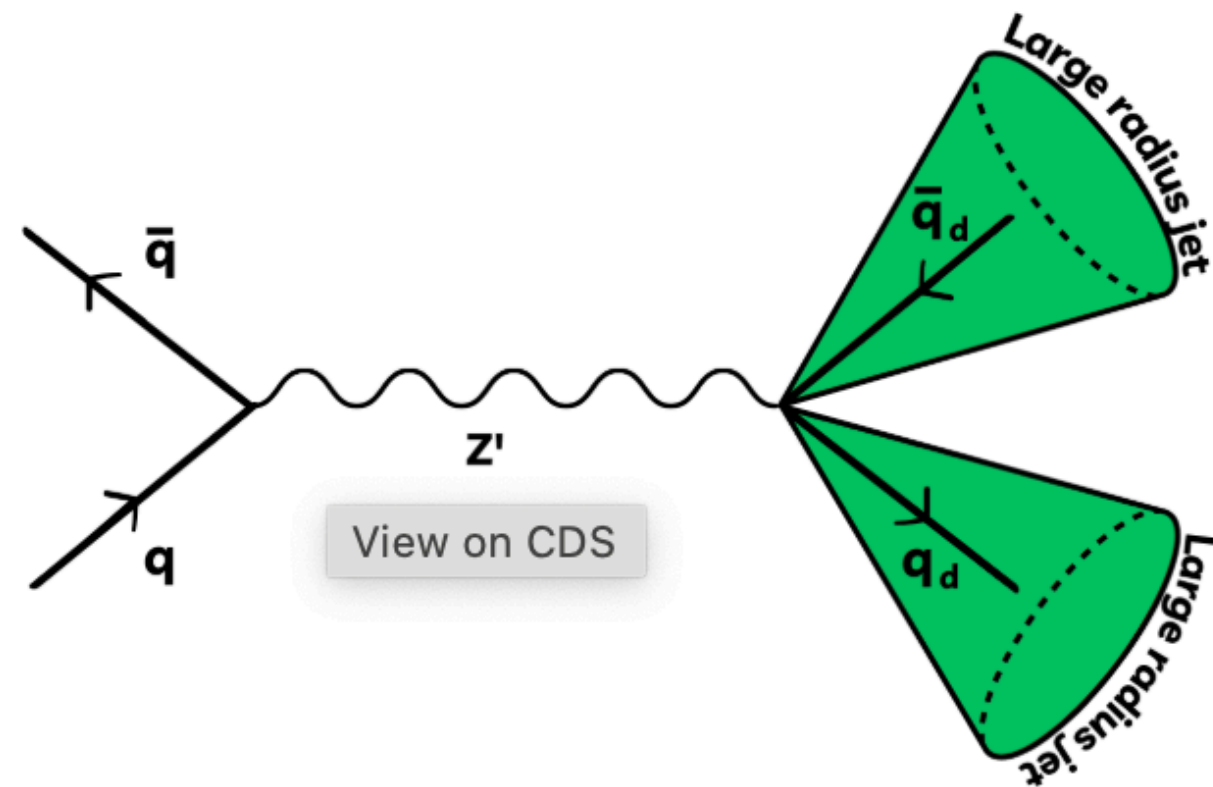
Resonant AD: A Sandbox for Overdensity Anomalies

Consider the humble dijet resonance search



- The “bump hunt”: look for a bump over **smooth background**
- **Sliding window** search
- Data-driven **background** estimation from **sidebands**

Resonant AD: A Sandbox for Overdensity Anomalies

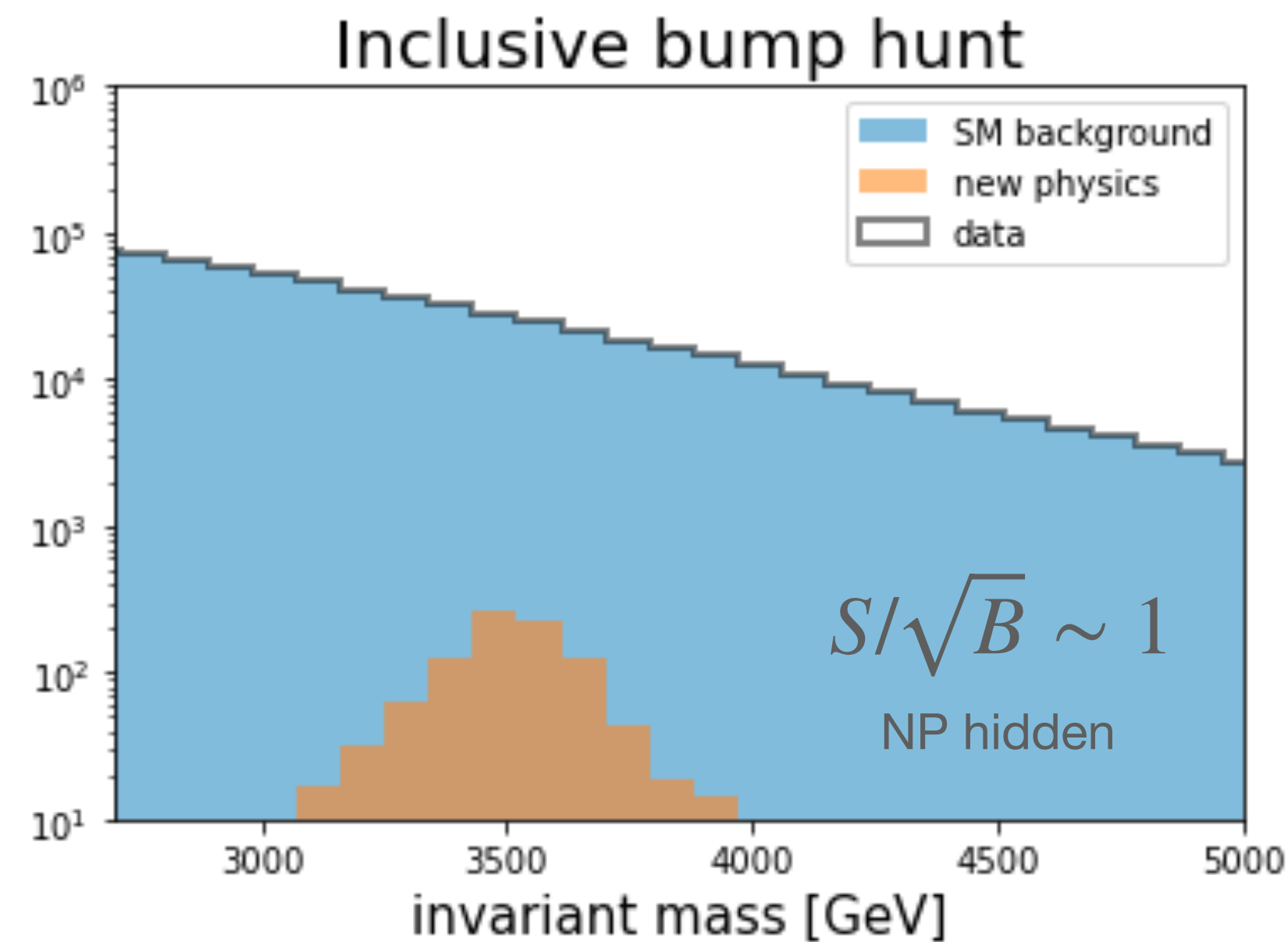


There could be new physics
hiding in **dijet resonances**

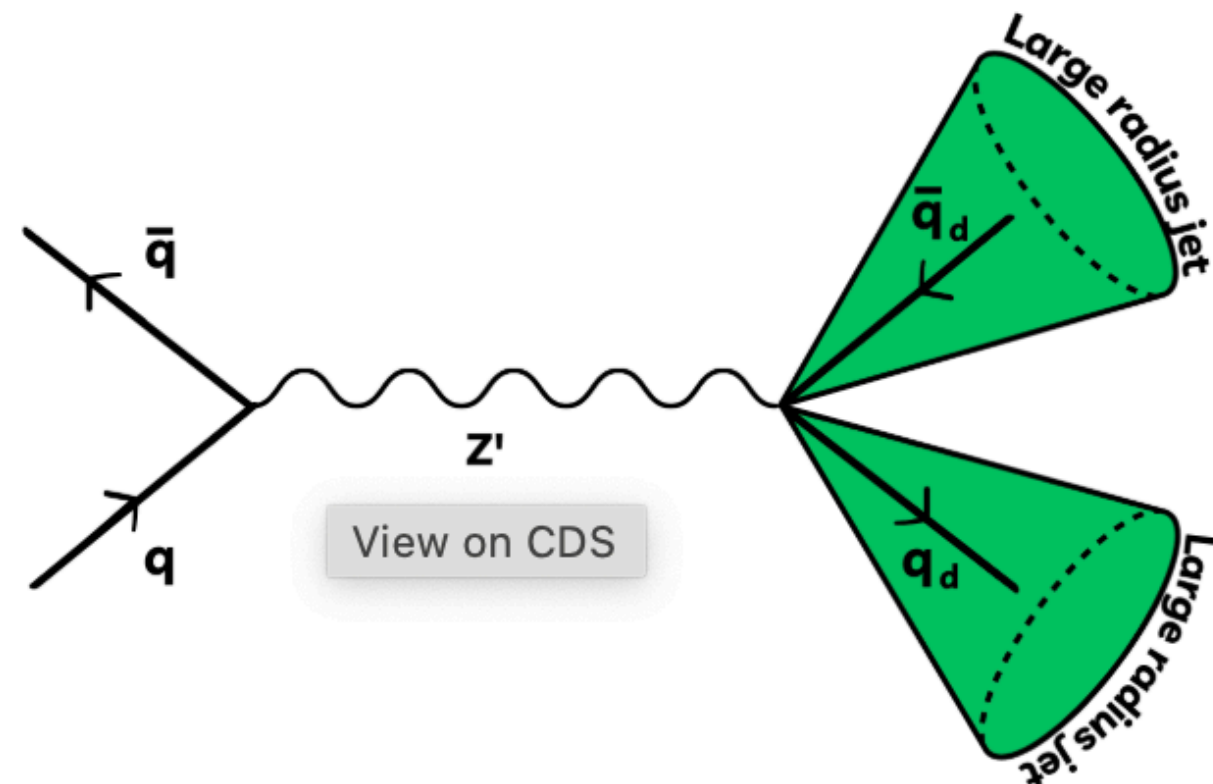
$$Z' \rightarrow XY$$

$$X \rightarrow \text{jets}, Y \rightarrow \text{jets}$$

- Boosted hadronically decaying particles
- Extra dimensions
- non-minimal RPV SUSY
- Dark showers
- ...



Resonant AD: A Sandbox for Overdensity Anomalies



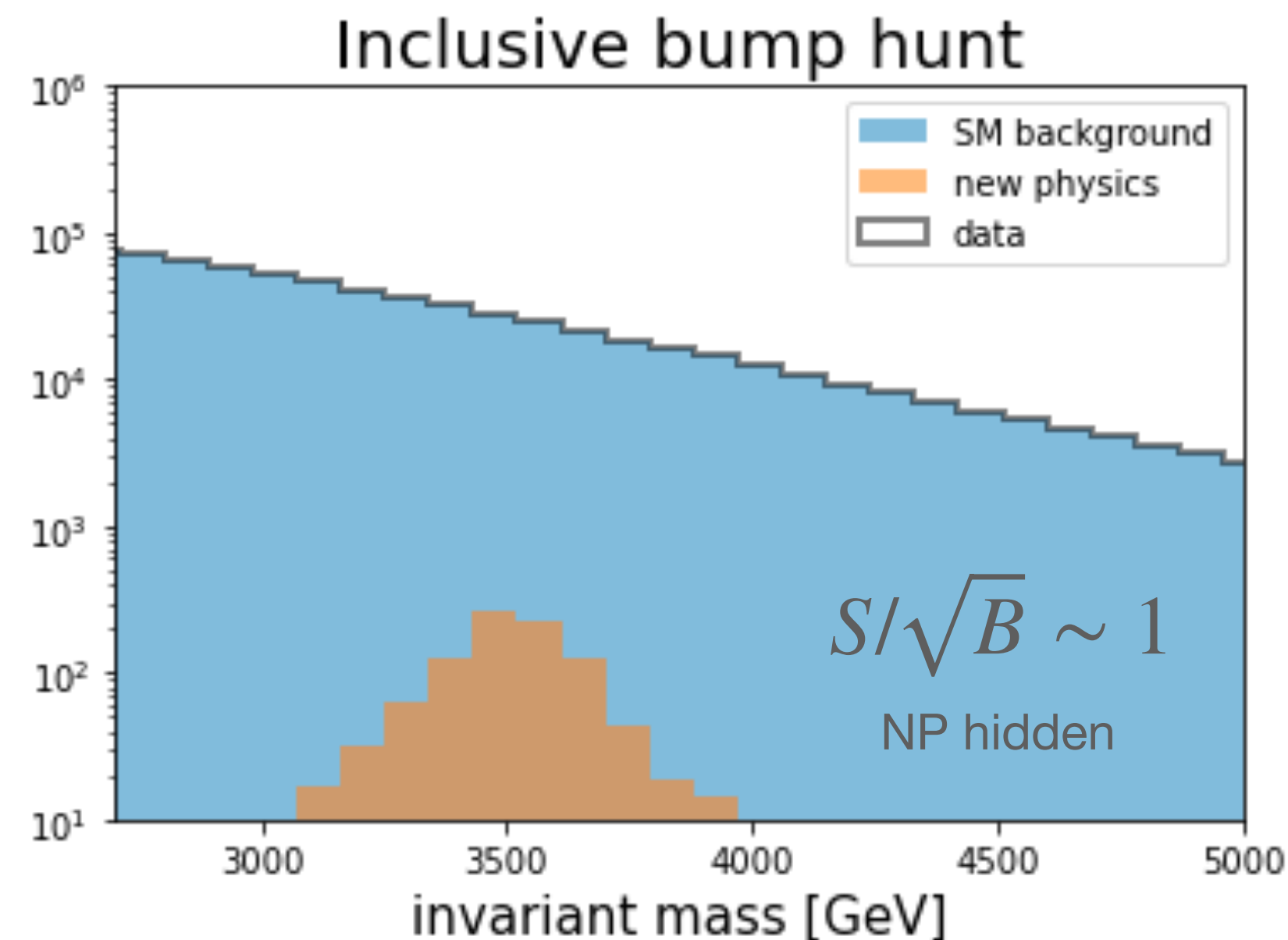
There could be new physics
hiding in **dijet resonances**

$$Z' \rightarrow XY$$

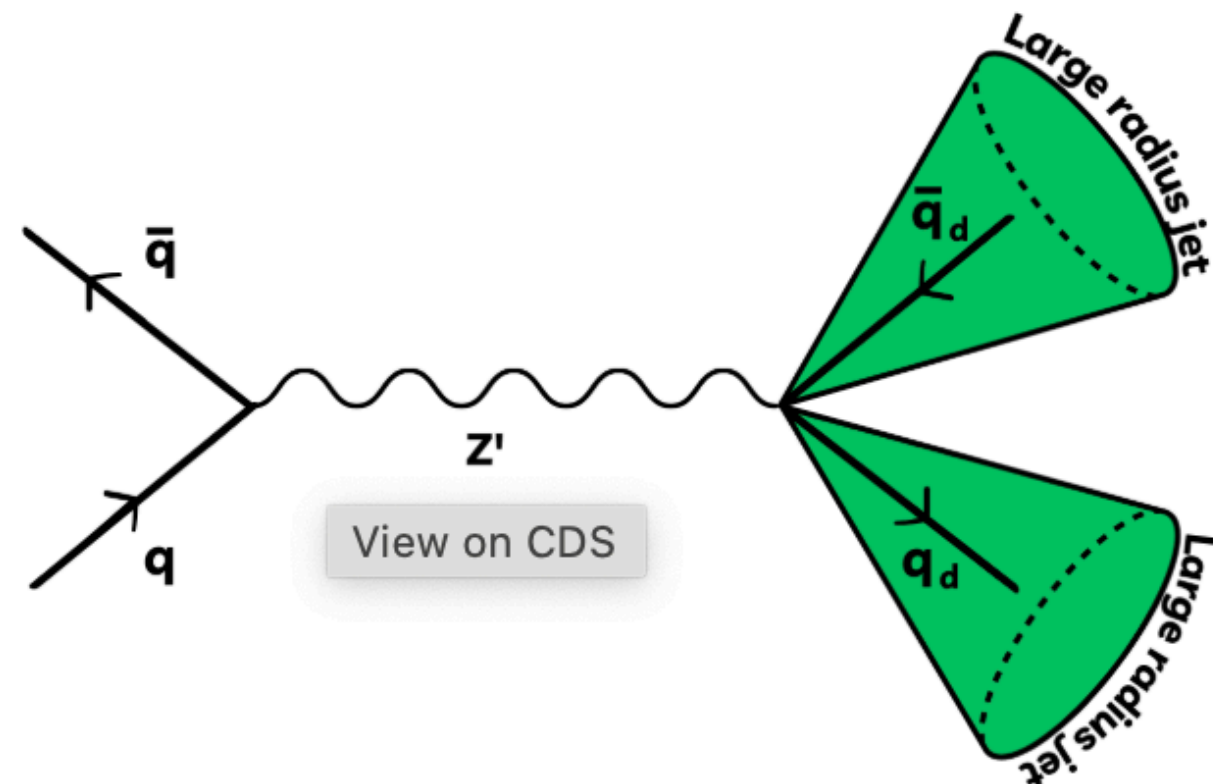
$$X \rightarrow \text{jets}, Y \rightarrow \text{jets}$$

- Boosted hadronically decaying particles
- Extra dimensions
- non-minimal RPV SUSY
- Dark showers
- ...

Additional features can distinguish
new physics from QCD background
 $x = \text{masses, substructure}$



Resonant AD: A Sandbox for Overdensity Anomalies



There could be new physics *hiding* in **dijet resonances**

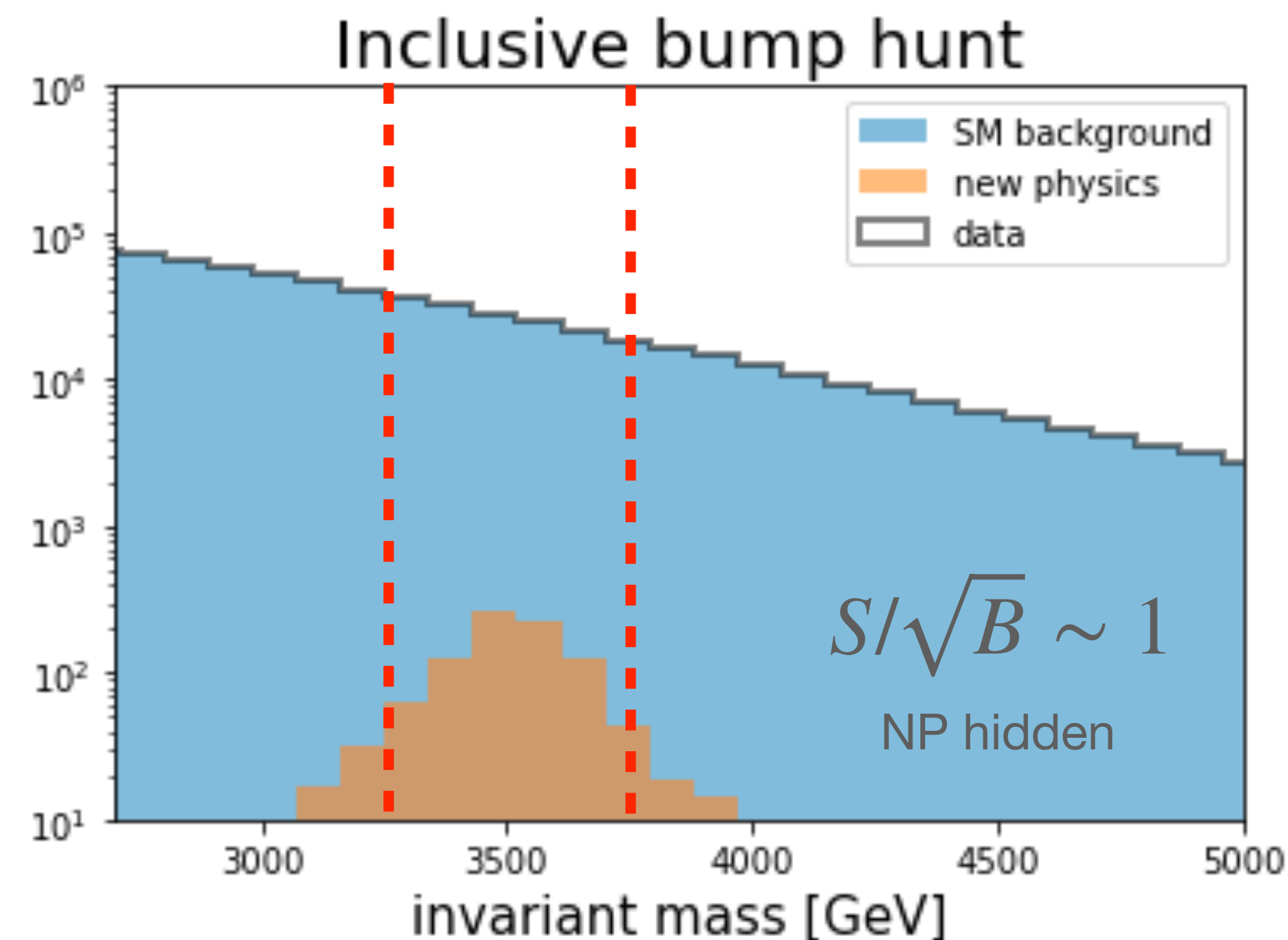
$$Z' \rightarrow XY$$

$$X \rightarrow \text{jets}, Y \rightarrow \text{jets}$$

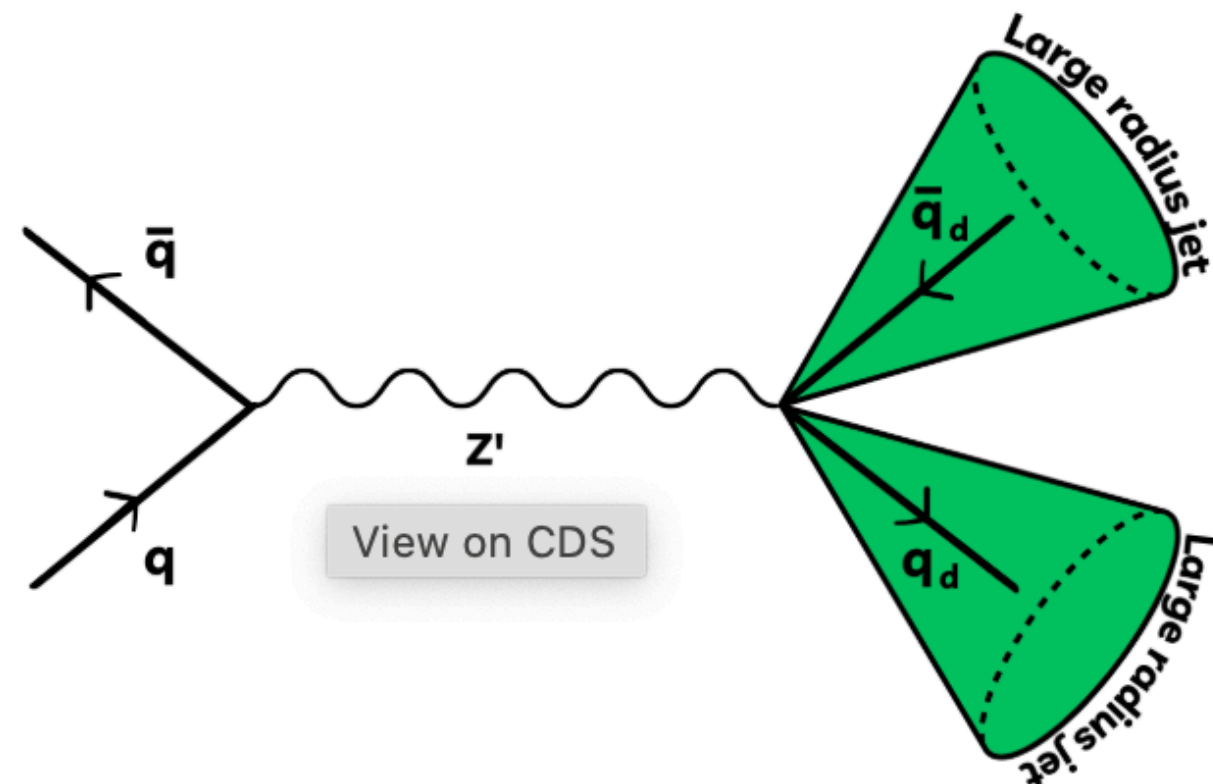
- Boosted hadronically decaying particles
- Extra dimensions
- non-minimal RPV SUSY
- Dark showers
- ...

Additional features can distinguish new physics from QCD background
x=masses, substructure

Use data in **sidebands**
to access **background**



Resonant AD: A Sandbox for Overdensity Anomalies



There could be new physics *hiding* in **dijet resonances**

$$Z' \rightarrow XY$$

$$X \rightarrow \text{jets}, Y \rightarrow \text{jets}$$

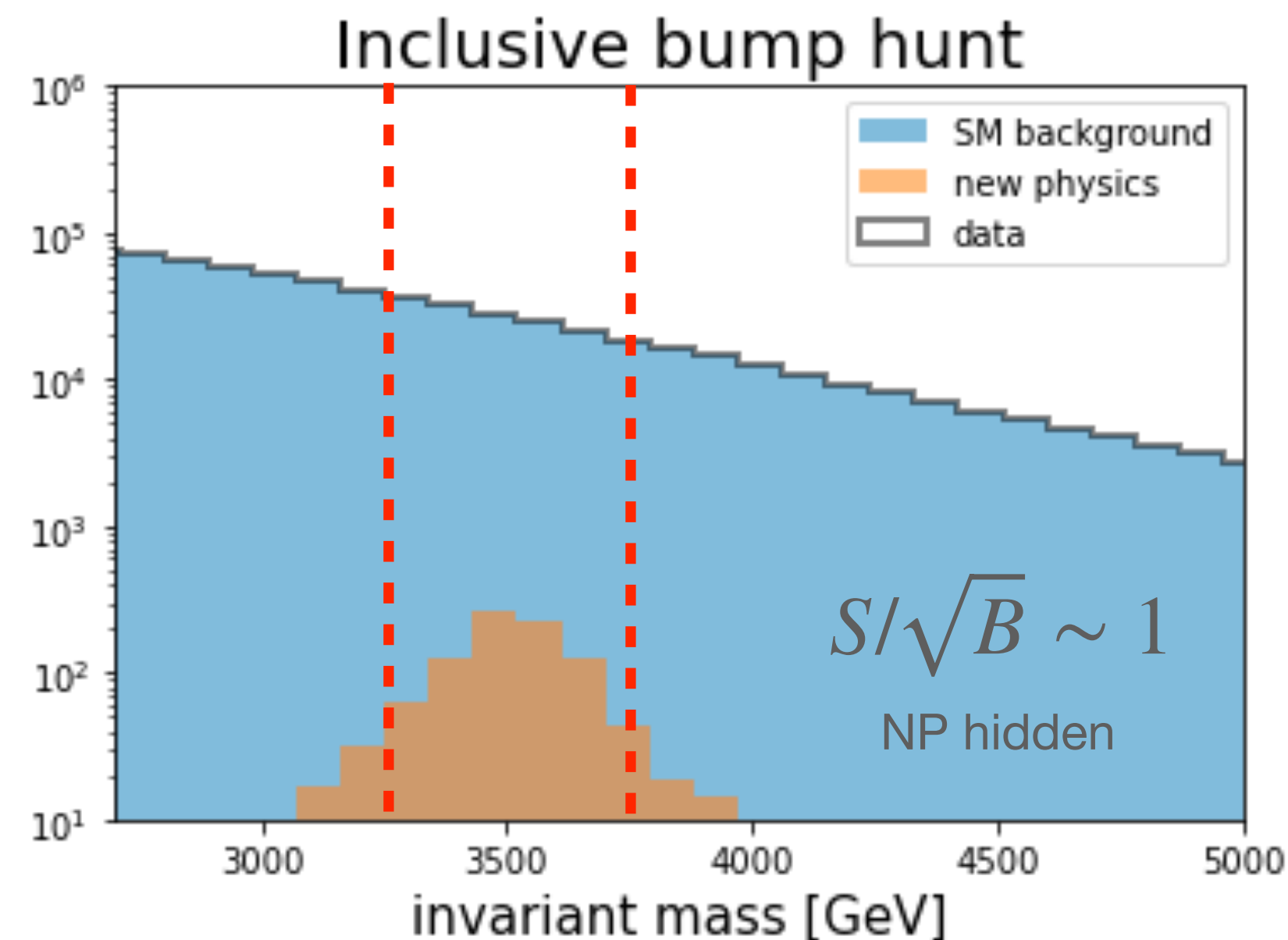
- Boosted hadronically decaying particles
- Extra dimensions
- non-minimal RPV SUSY
- Dark showers
- ...

Additional features can distinguish new physics from QCD background
x=masses, substructure

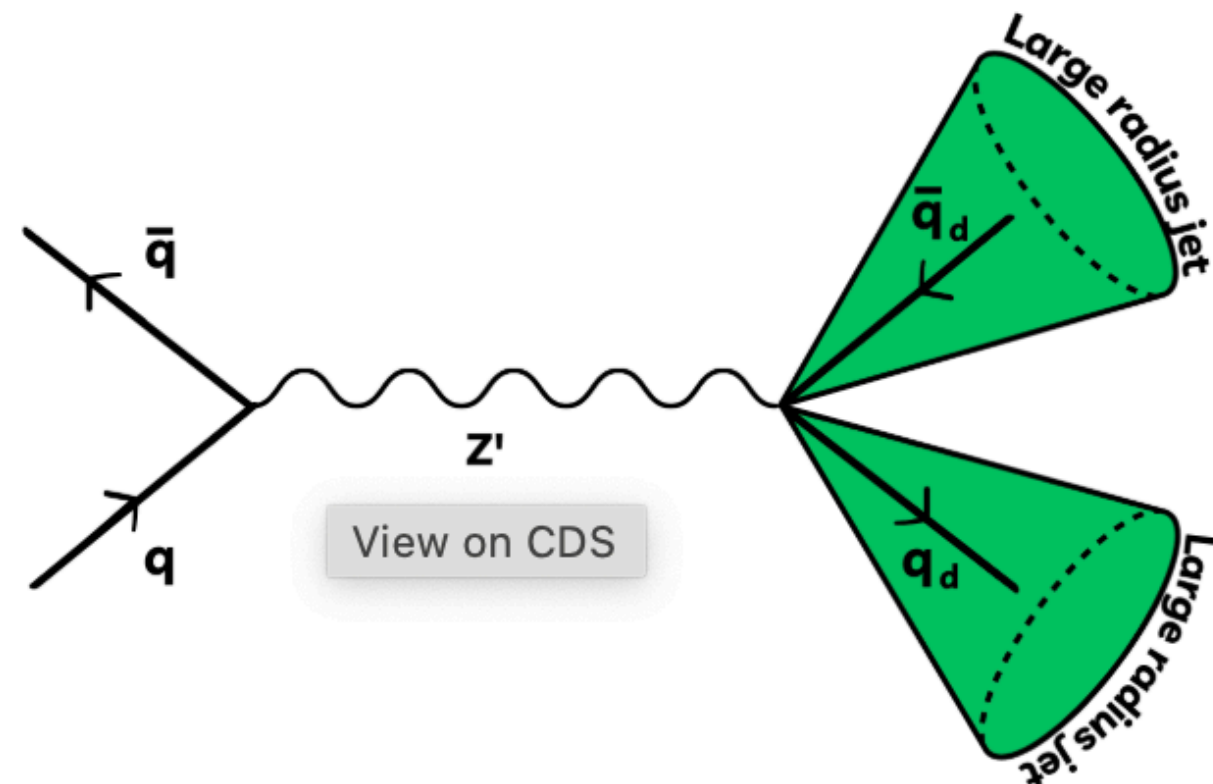
Use data in **sidebands** to access **background**

Learn anomaly score

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}$$



Resonant AD: A Sandbox for Overdensity Anomalies



There could be new physics *hiding* in **dijet resonances**

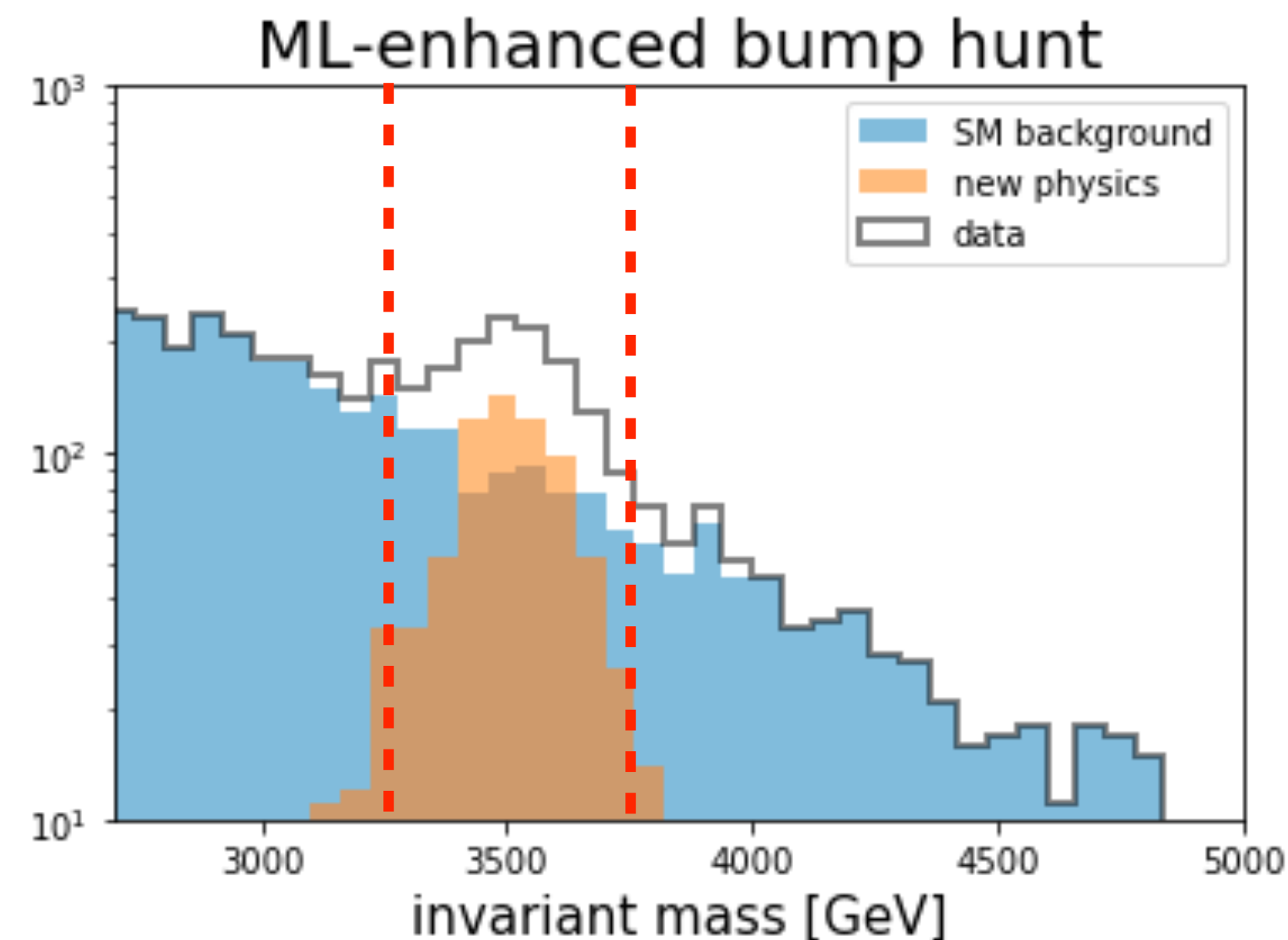
$$Z' \rightarrow XY$$

$$X \rightarrow \text{jets}, Y \rightarrow \text{jets}$$

- Boosted hadronically decaying particles
- Extra dimensions
- non-minimal RPV SUSY
- Dark showers
- ...

Additional features can distinguish new physics from QCD background
 $x = \text{masses, substructure}$

$$\text{Cut } R(x) > R_c$$

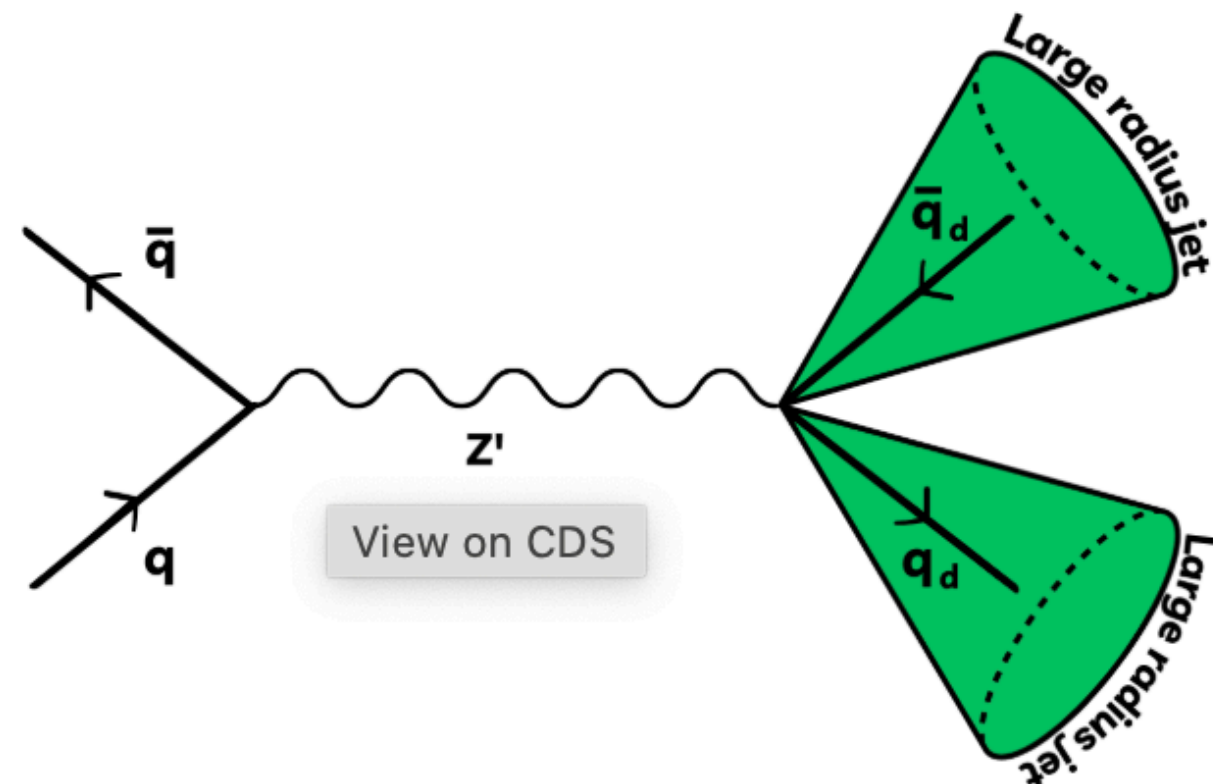


Use data in **sidebands**
to access **background**

Learn anomaly score

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}$$

Resonant AD: A Sandbox for Overdensity Anomalies



There could be new physics *hiding* in **dijet resonances**

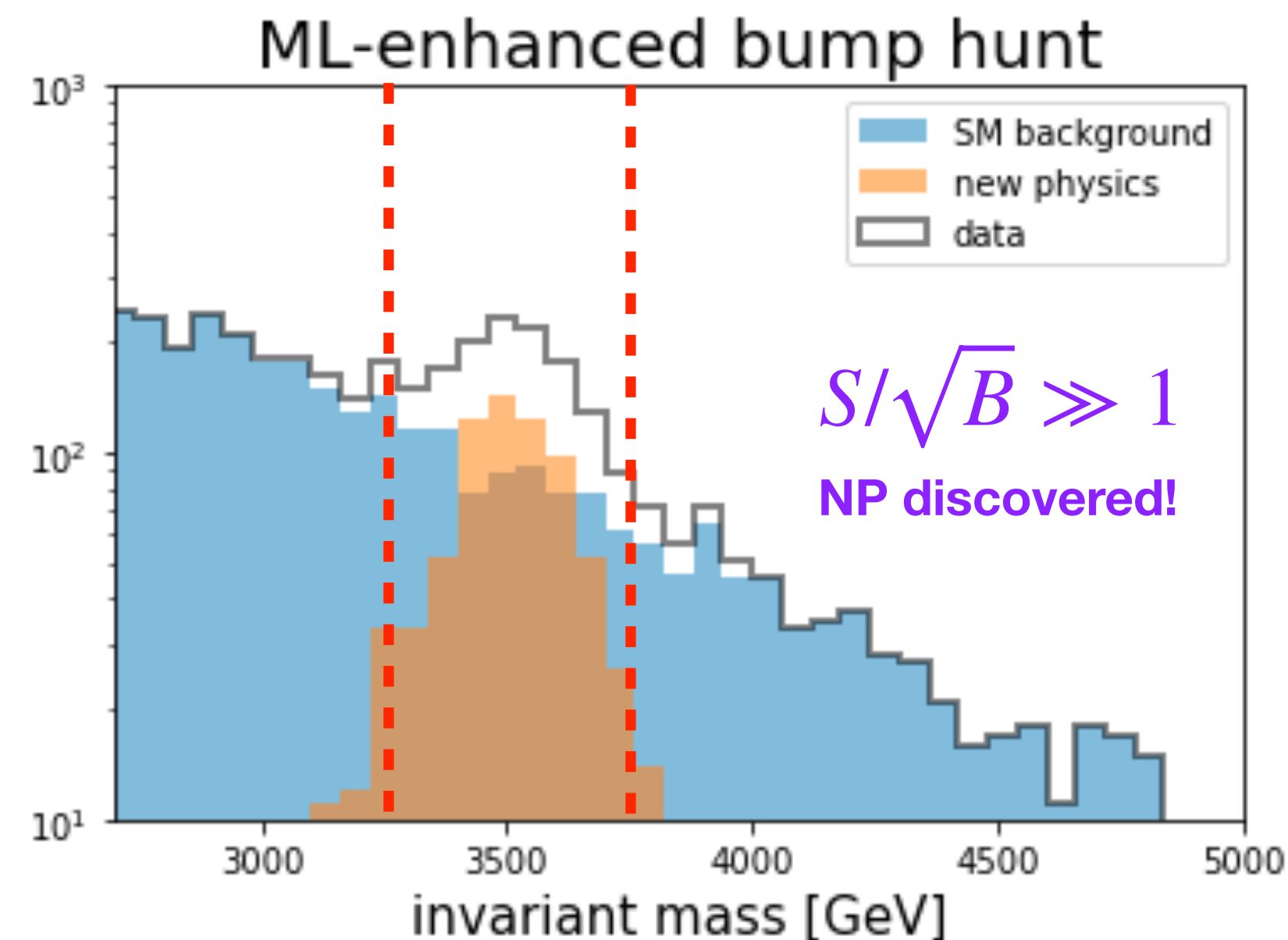
$$Z' \rightarrow XY$$

$$X \rightarrow \text{jets}, Y \rightarrow \text{jets}$$

- Boosted hadronically decaying particles
- Extra dimensions
- non-minimal RPV SUSY
- Dark showers
- ...

Additional features can distinguish new physics from QCD background
 $x = \text{masses, substructure}$

$$\text{Cut } R(x) > R_c$$

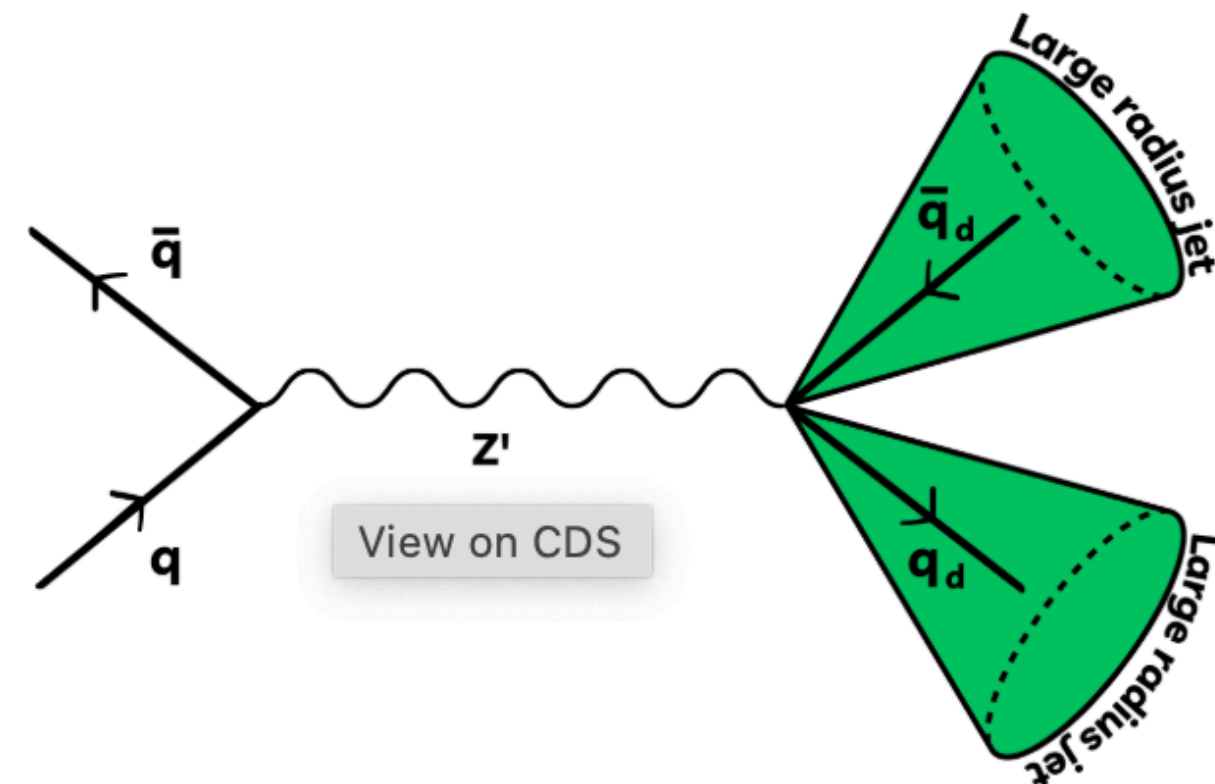


Use data in **sidebands** to access **background**

Learn anomaly score

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}$$

Resonant AD: A Sandbox for Overdensity Anomalies



There could be new physics *hiding* in **dijet resonances**

$$Z' \rightarrow XY$$

$$X \rightarrow \text{jets}, Y \rightarrow \text{jets}$$

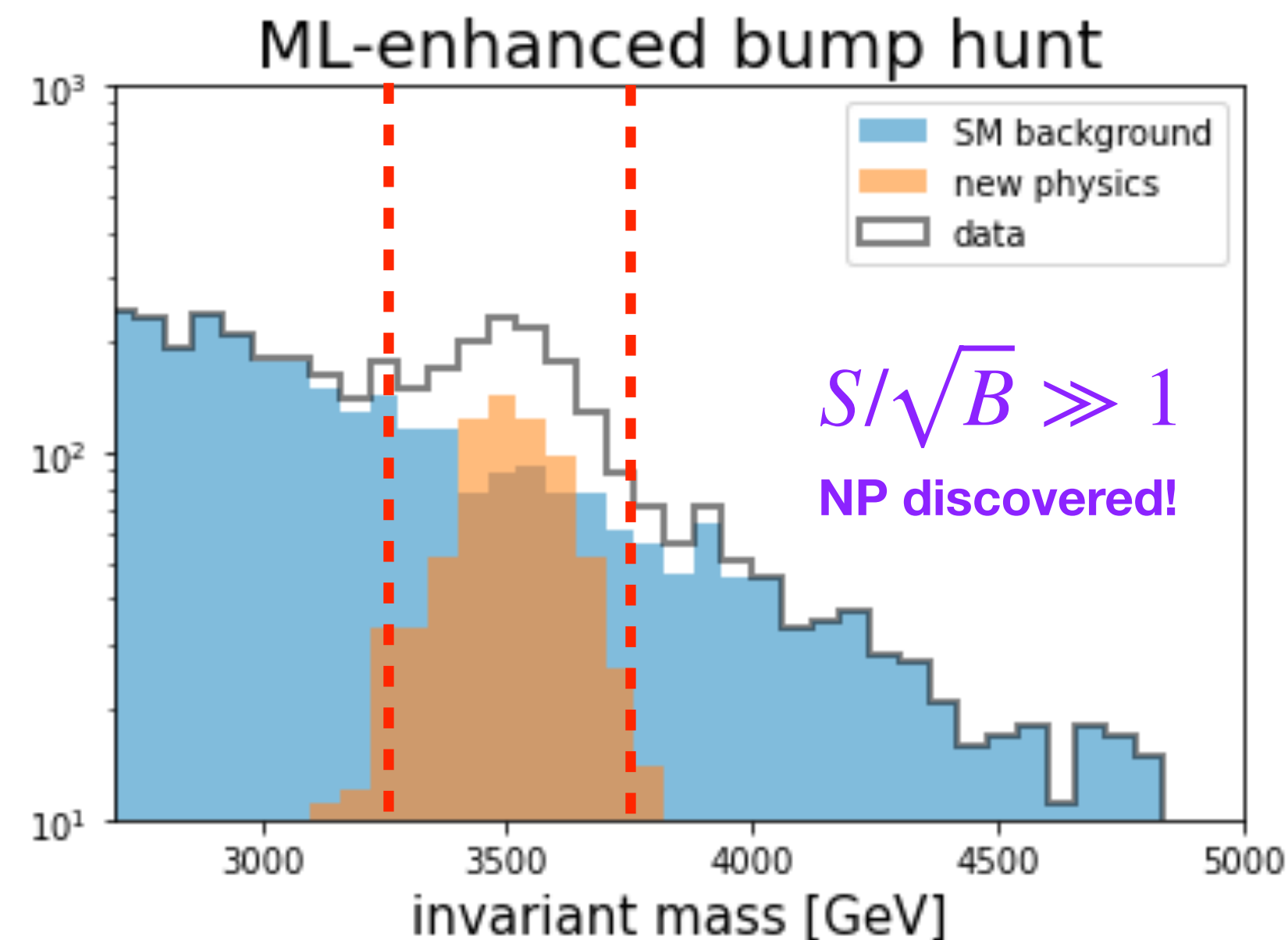
- Boosted hadronically decaying particles
- Extra dimensions
- non-minimal RPV SUSY
- Dark showers
- ...

Additional features can distinguish new physics from QCD background
x=masses, substructure

$$\text{Cut } R(x) > R_c$$

$$\frac{S}{\sqrt{B}} \rightarrow \frac{\epsilon_S}{\sqrt{\epsilon_B}} \frac{S}{\sqrt{B}}$$

“Significance improvement factor”



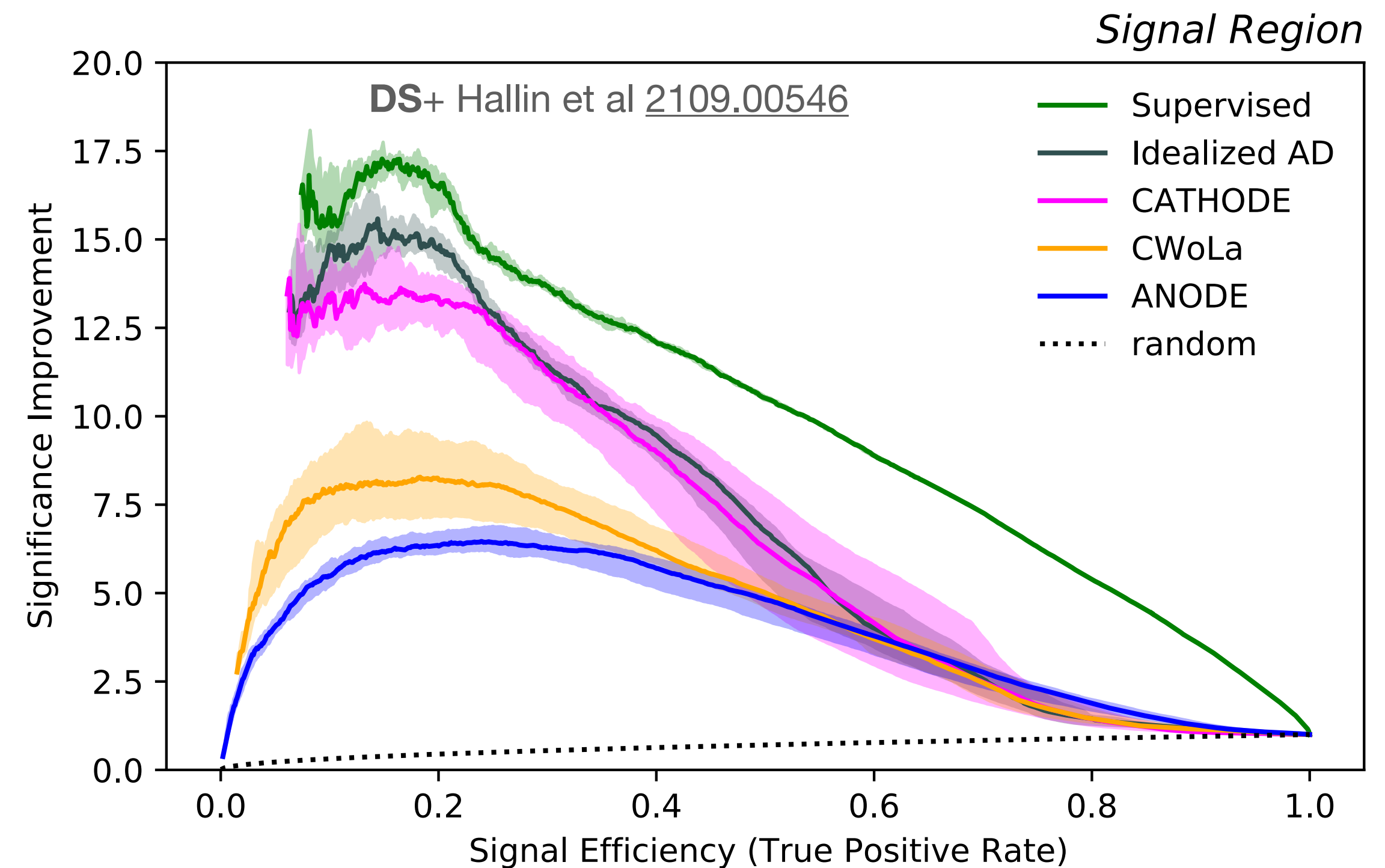
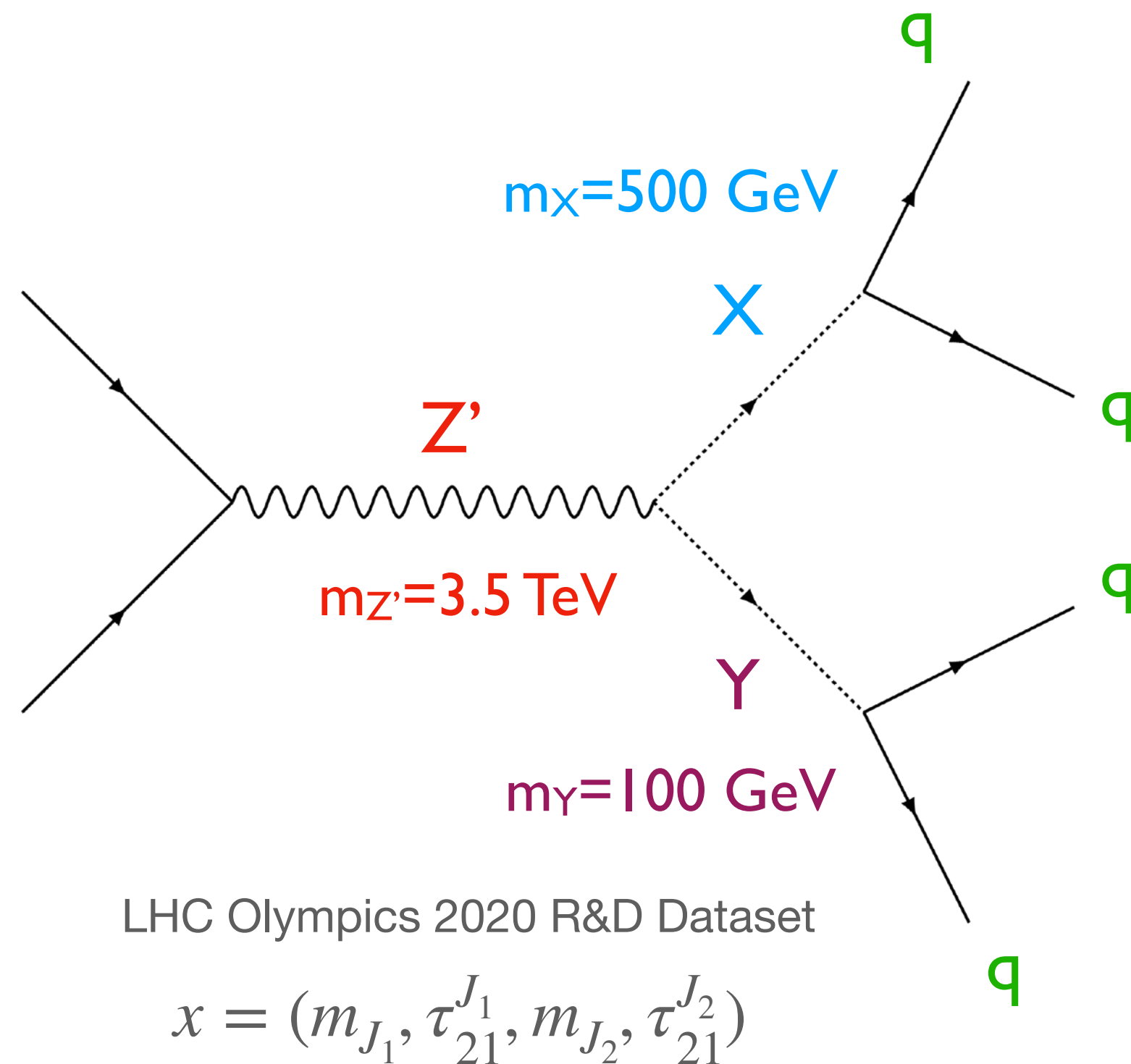
Use data in **sidebands** to access **background**

Learn anomaly score

$$R(x) = \frac{p_{data}(x)}{p_{bg}(x)}$$

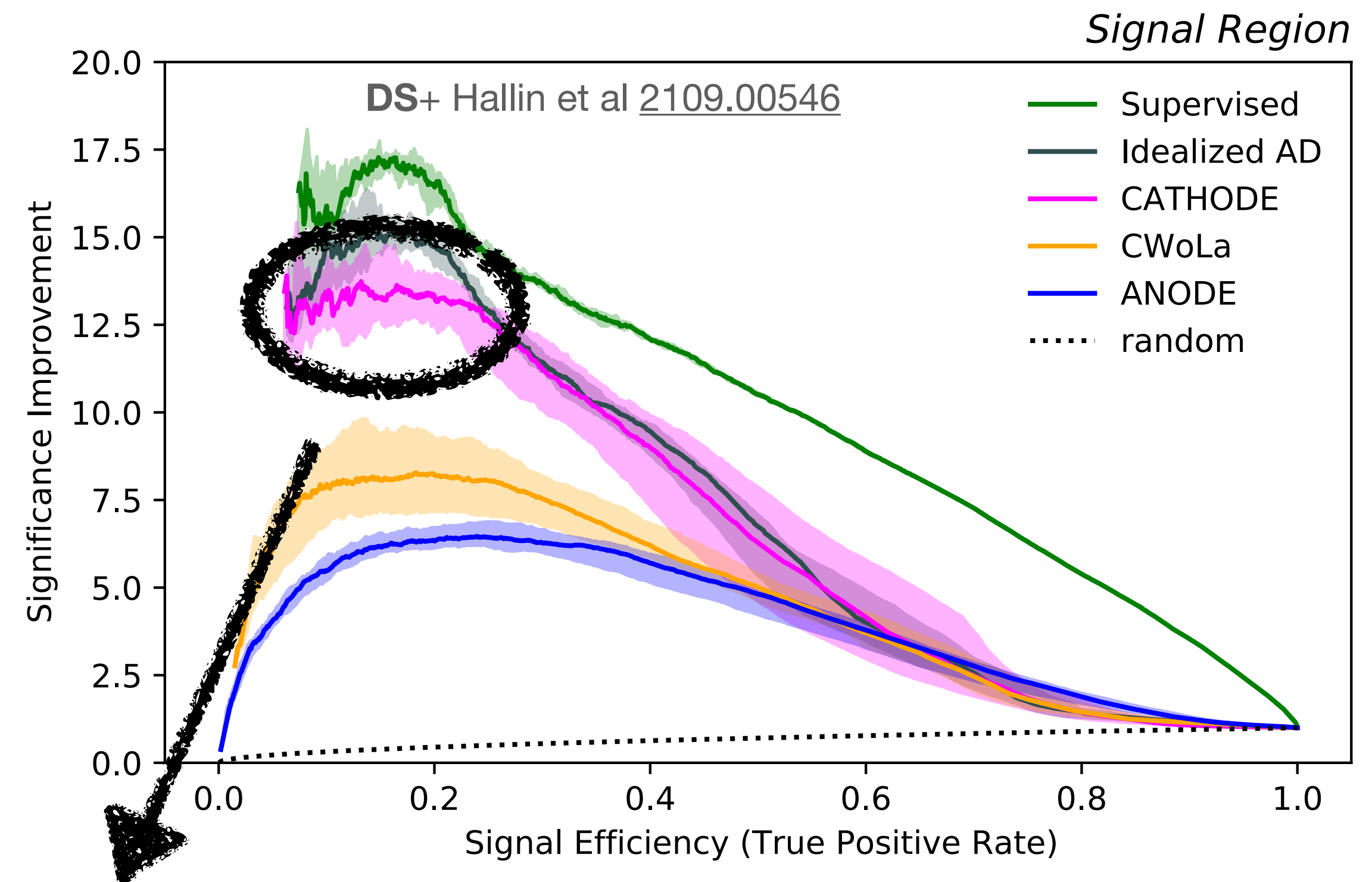
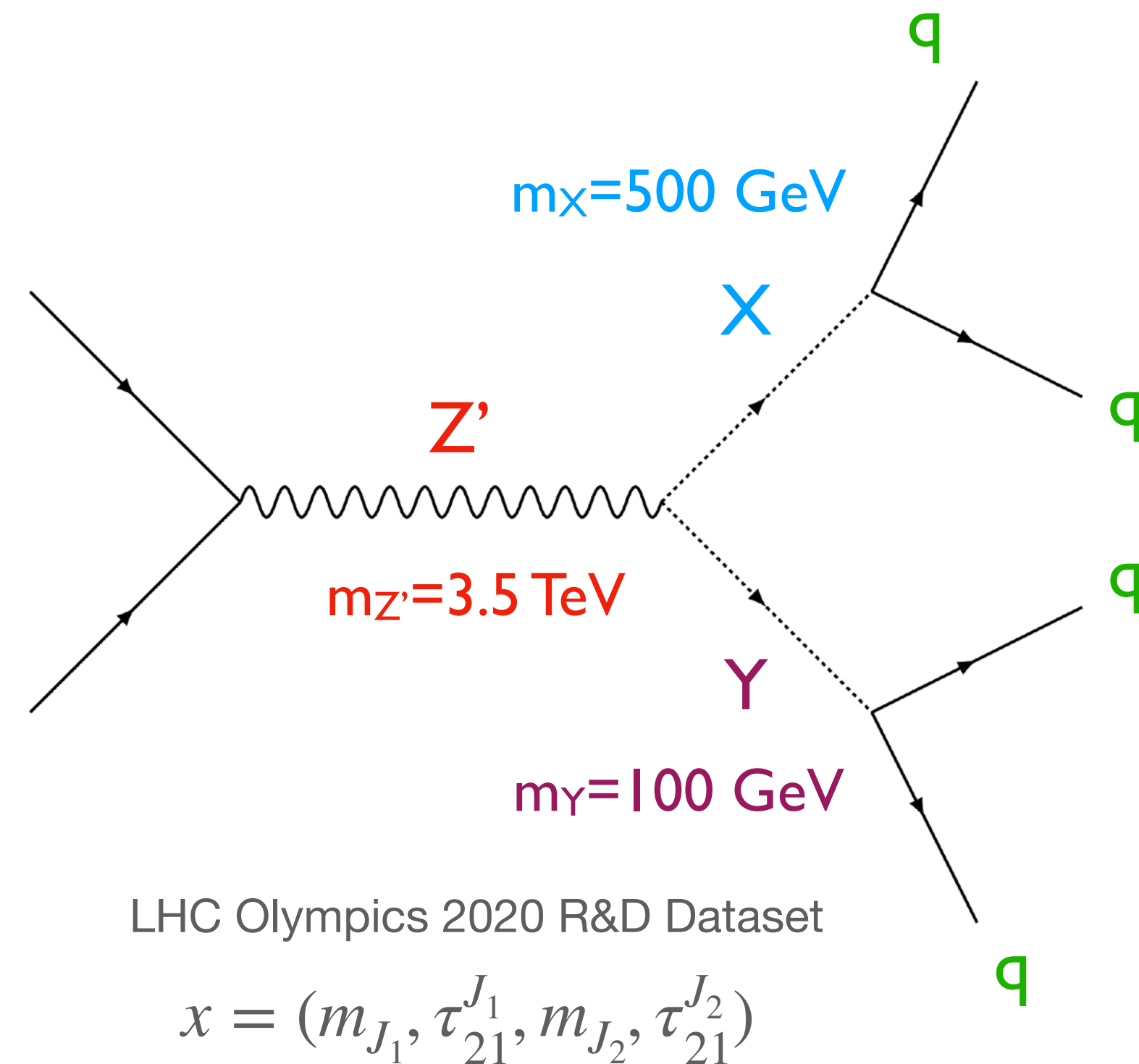
Overdensity score: performance

New methods can achieve impressive performance gains over the inclusive bump hunt.



Overdensity score: performance

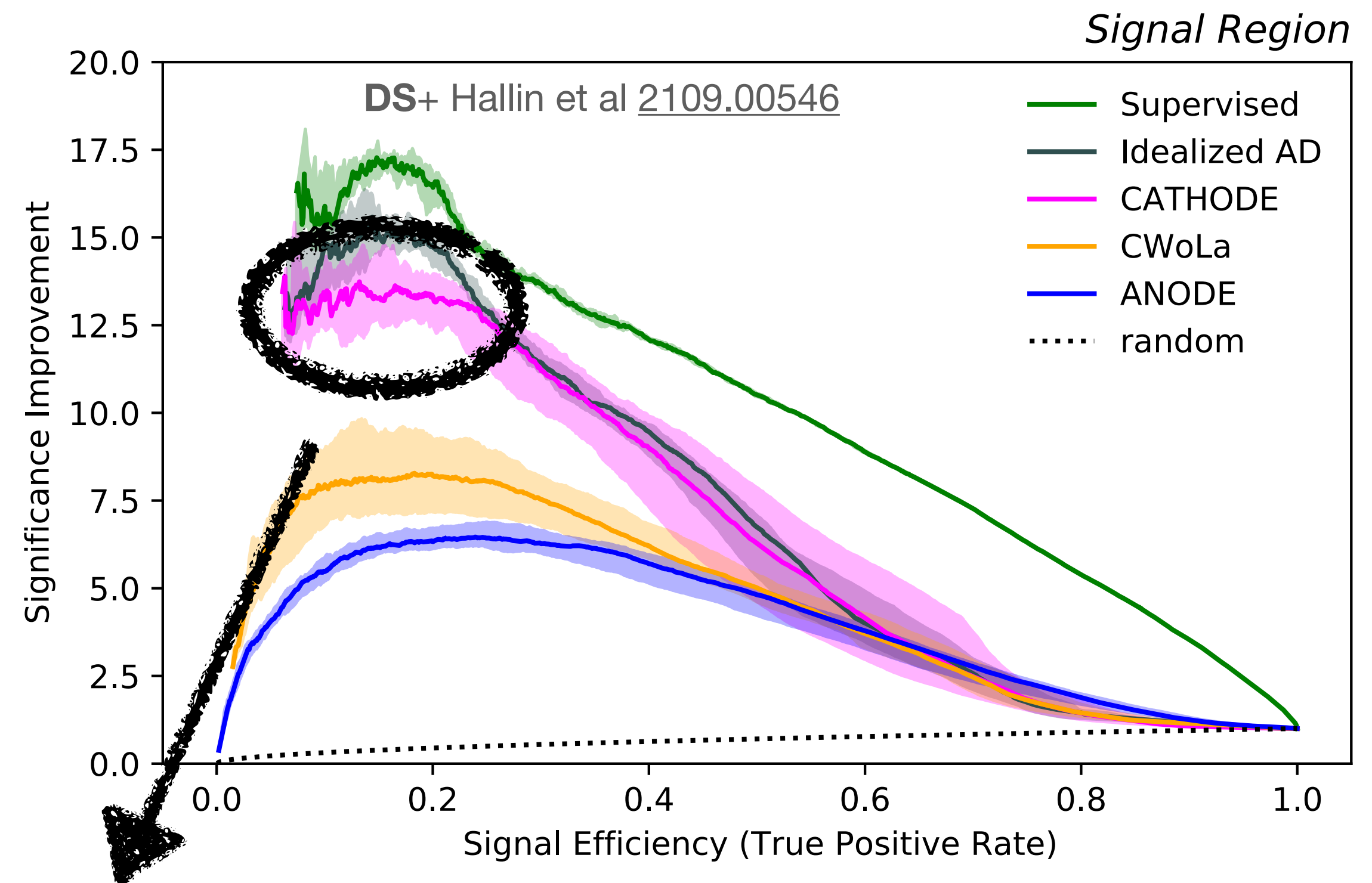
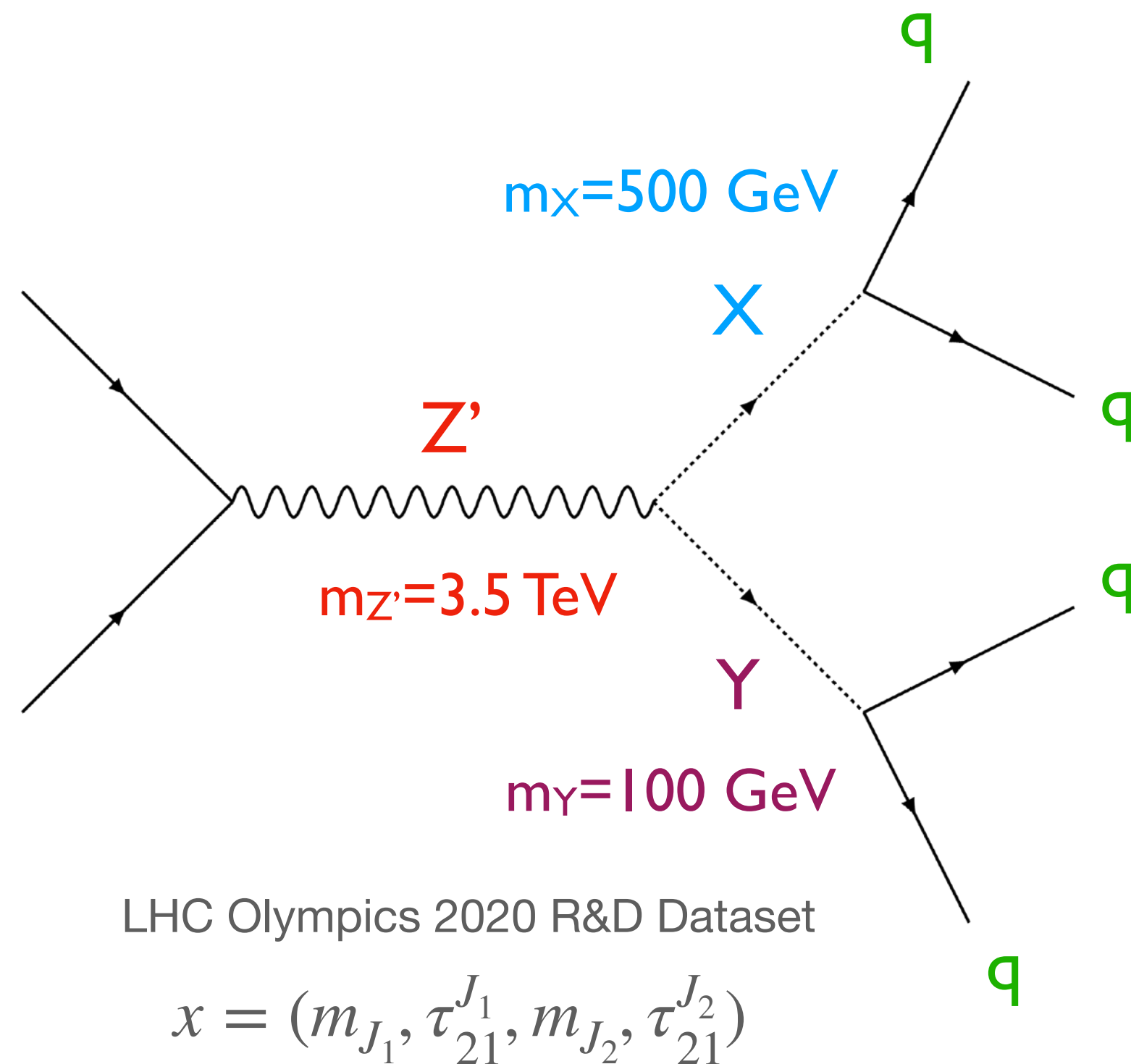
New methods can achieve impressive performance gains over the inclusive bump hunt.



On this signal, $\sim 2\sigma$ inclusive dijet $\Rightarrow$ up to $\sim 30\sigma$ with CATHODE method [DS+ Hallin et al 2109.00546]

Overdensity score: performance

New methods can achieve impressive performance gains over the inclusive bump hunt.

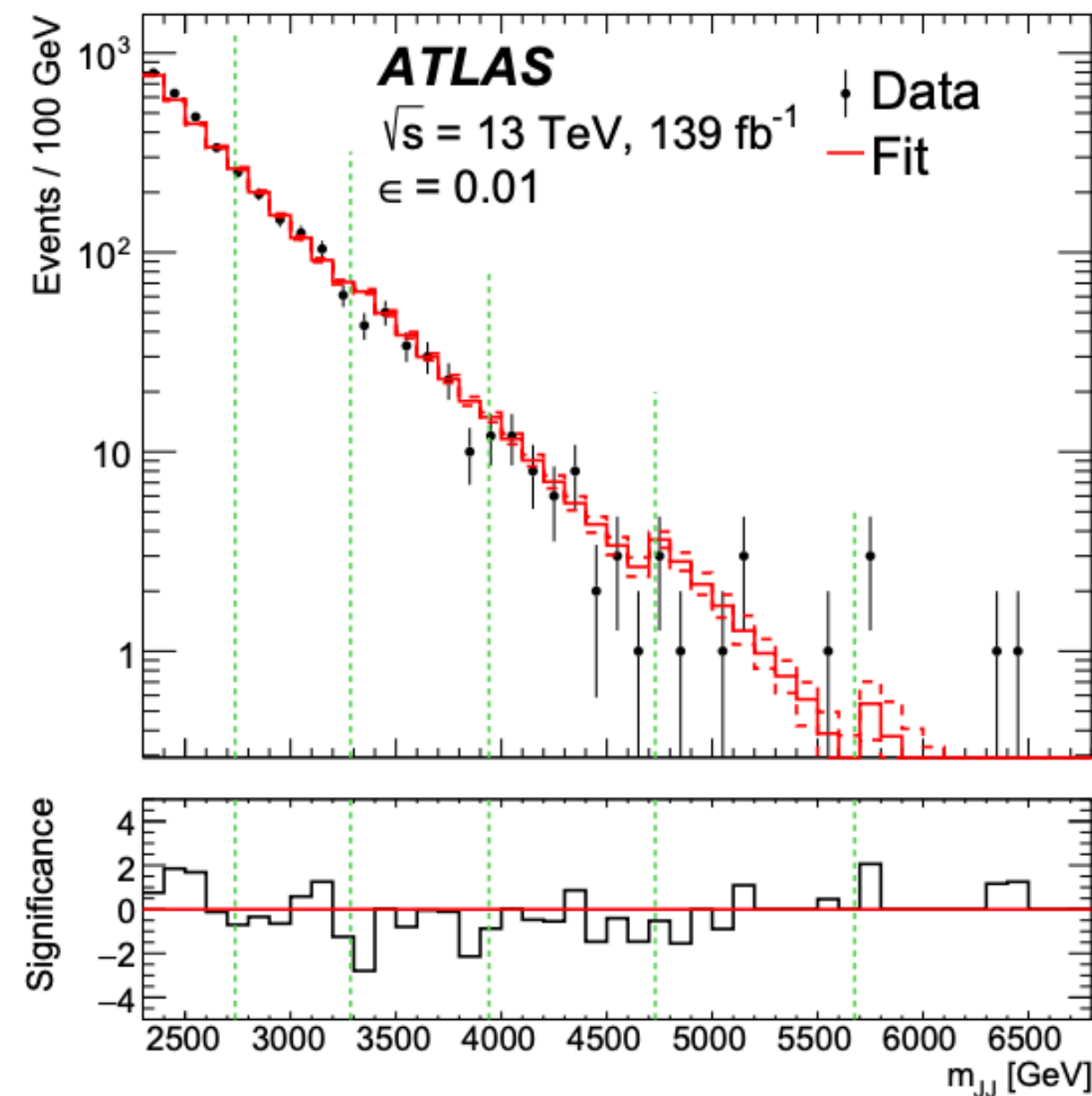


On this signal, $\sim 2\sigma$ inclusive dijet $\Rightarrow$ up to $\sim 30\sigma$ with CATHODE method [DS+ Hallin et al 2109.00546]

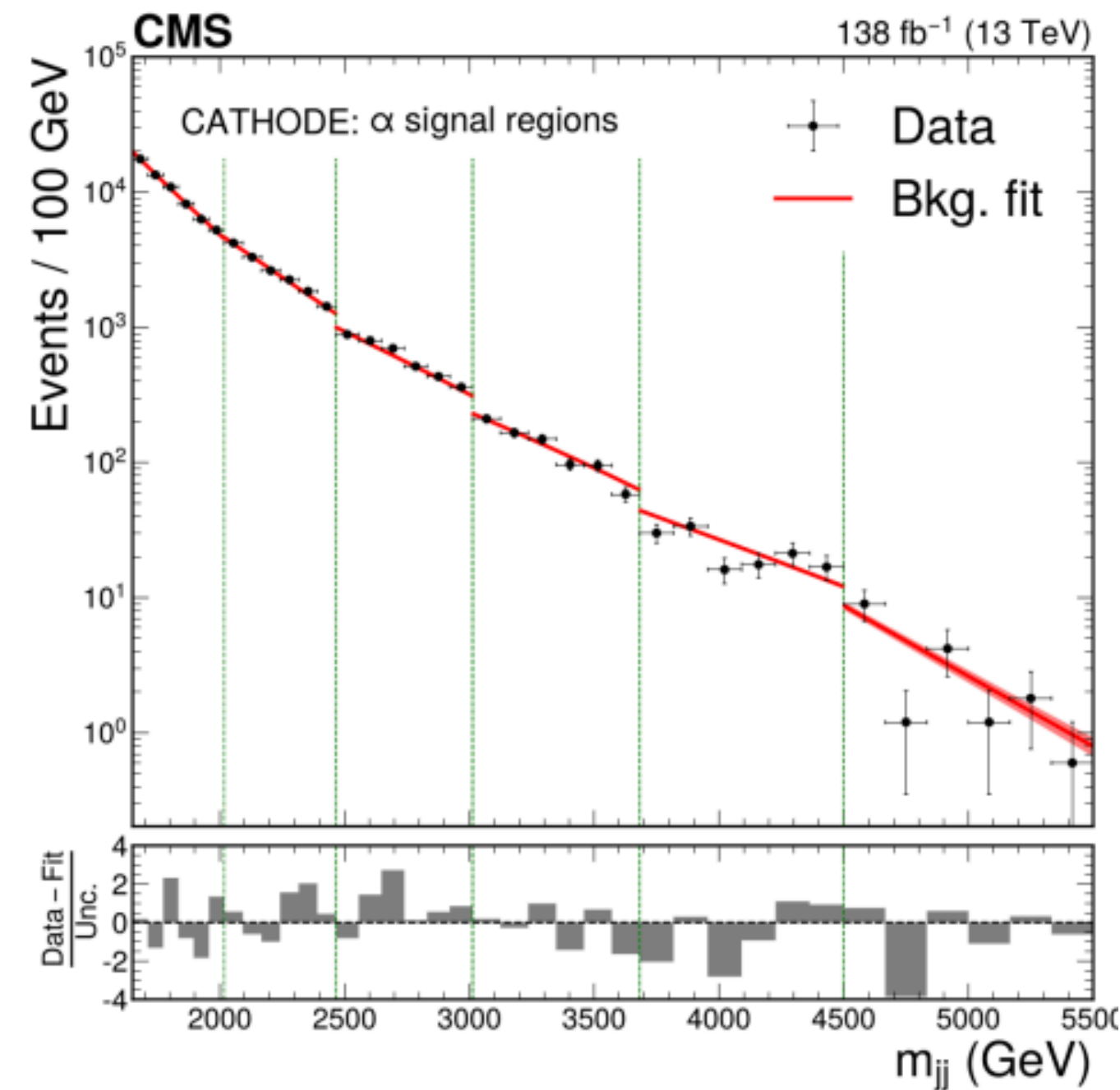
New physics could be hiding in the data right now!

Resonant AD with Overdensity Score @ LHC

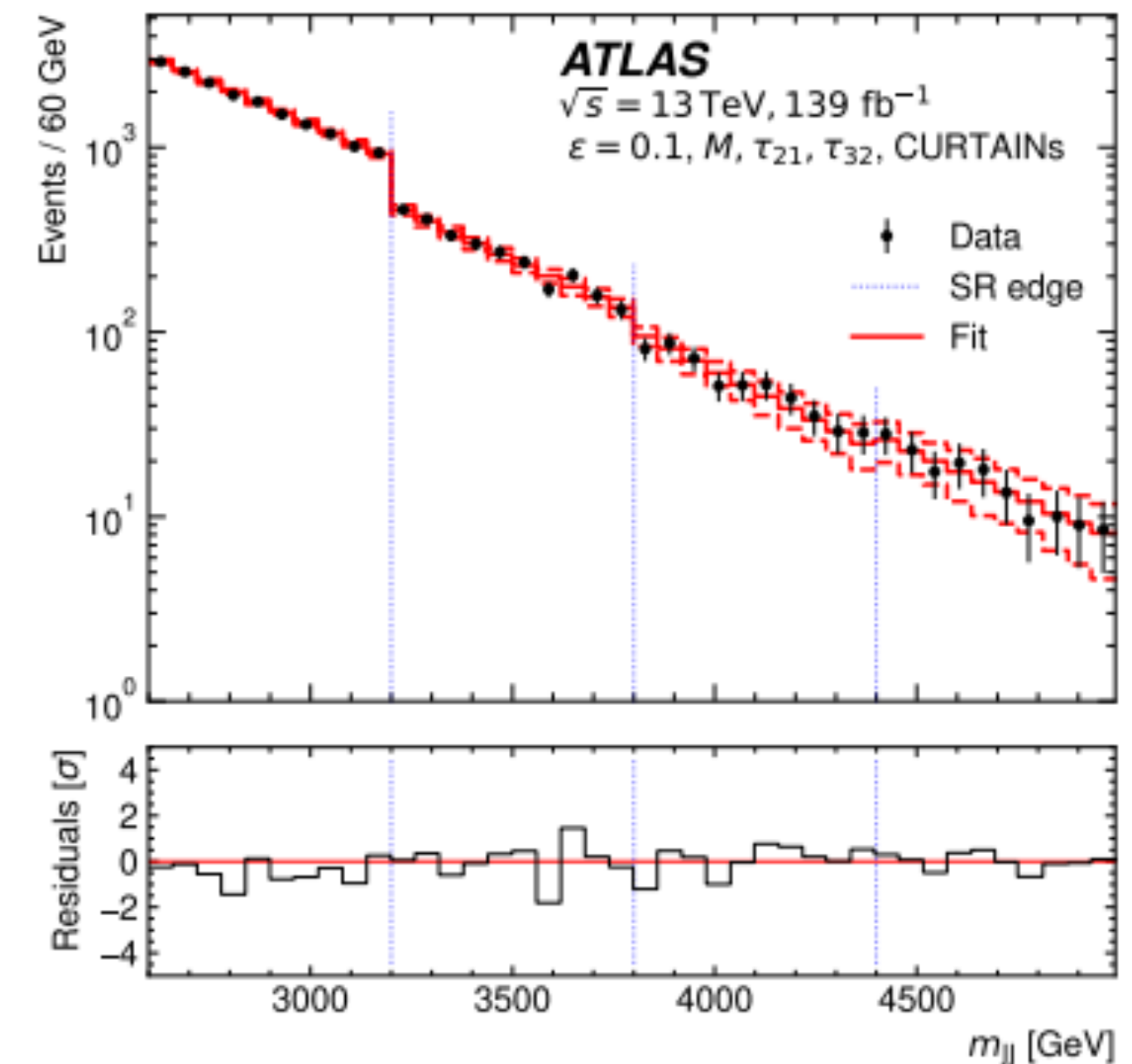
Overdensity-based methods are now becoming analyses at the LHC!



CWoLa Hunting
ATLAS dijet anomaly search
2005.02983

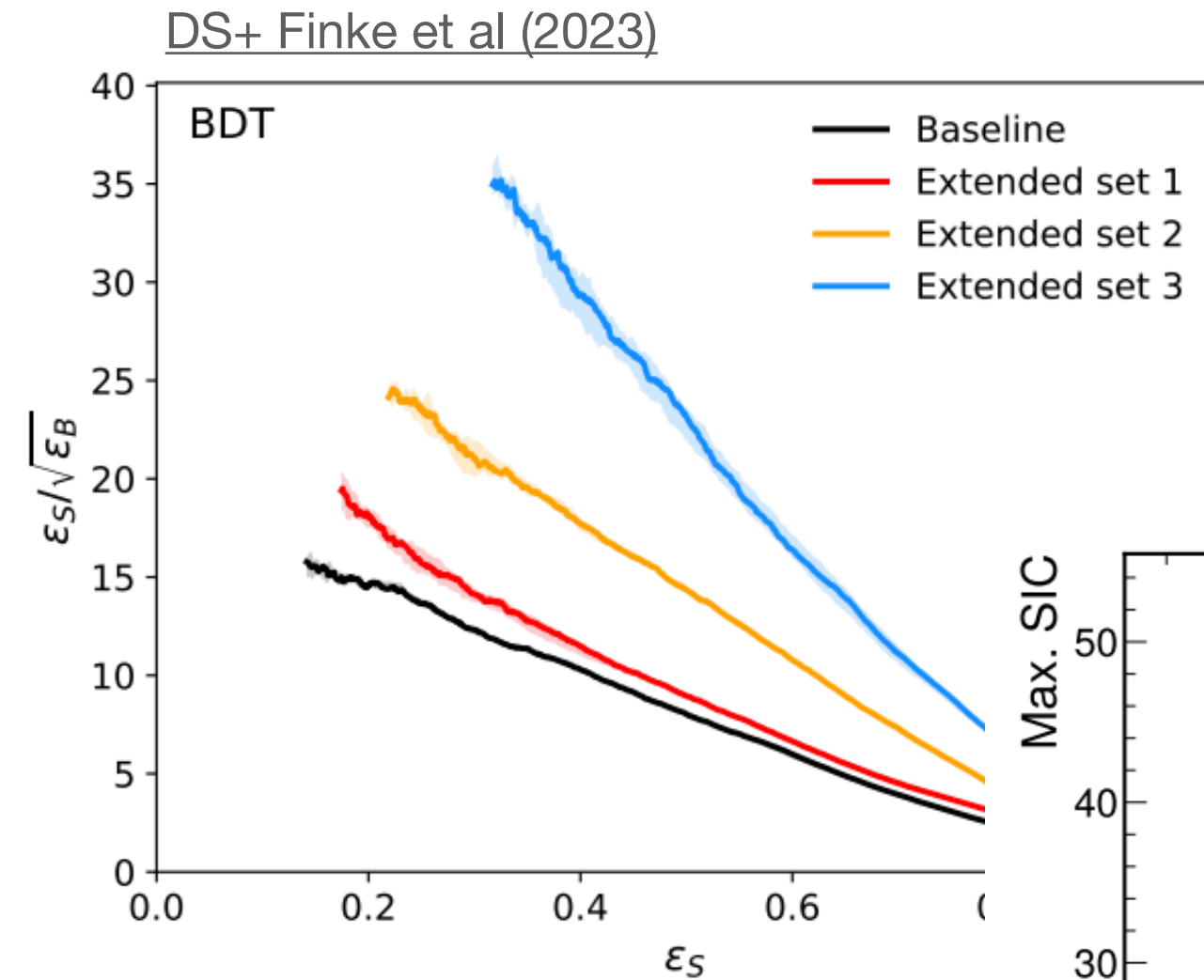


CATHODE
CMS dijet anomaly search
2412.03747

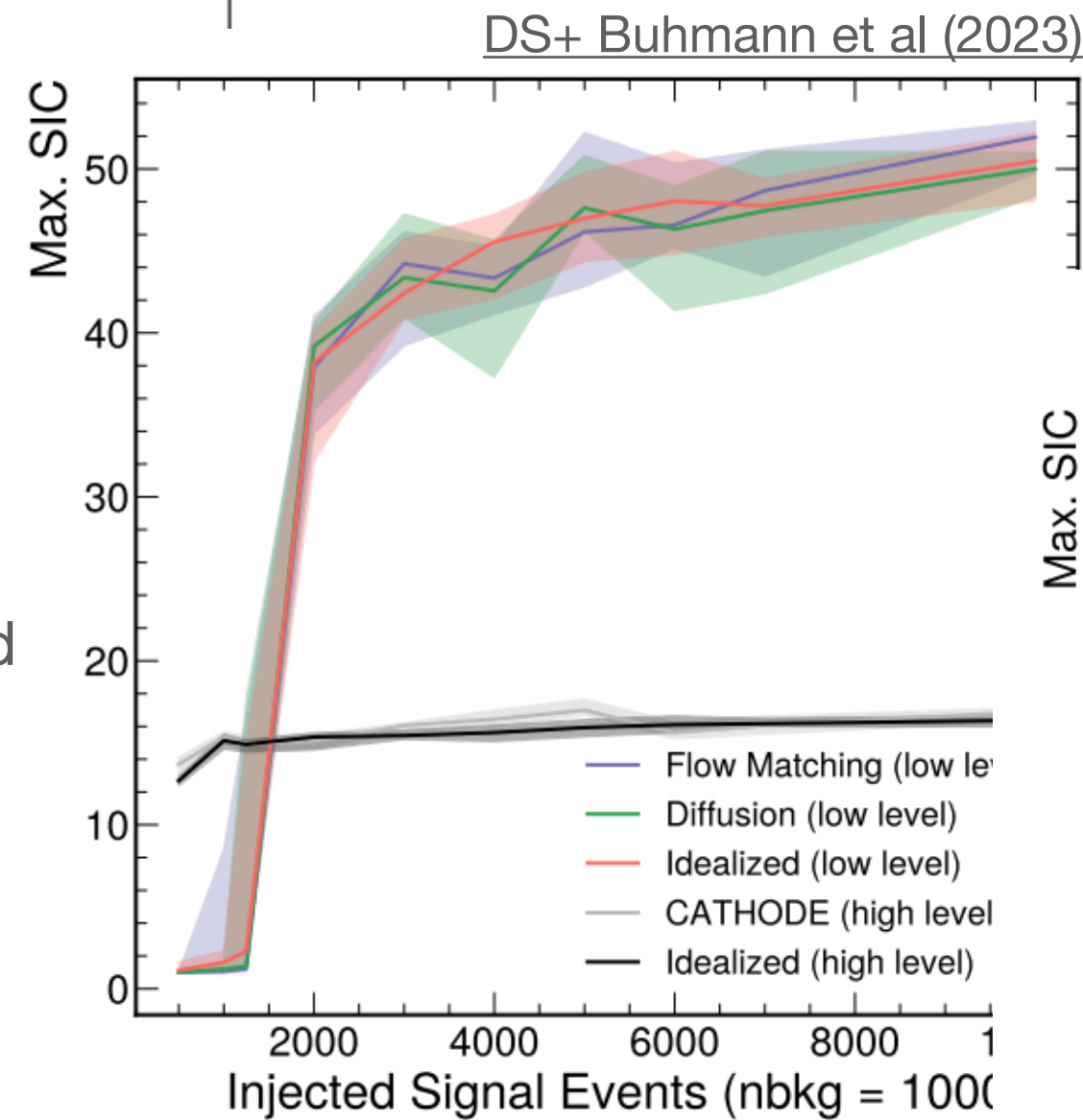


CURTAINS
ATLAS dijet anomaly search
2502.09770

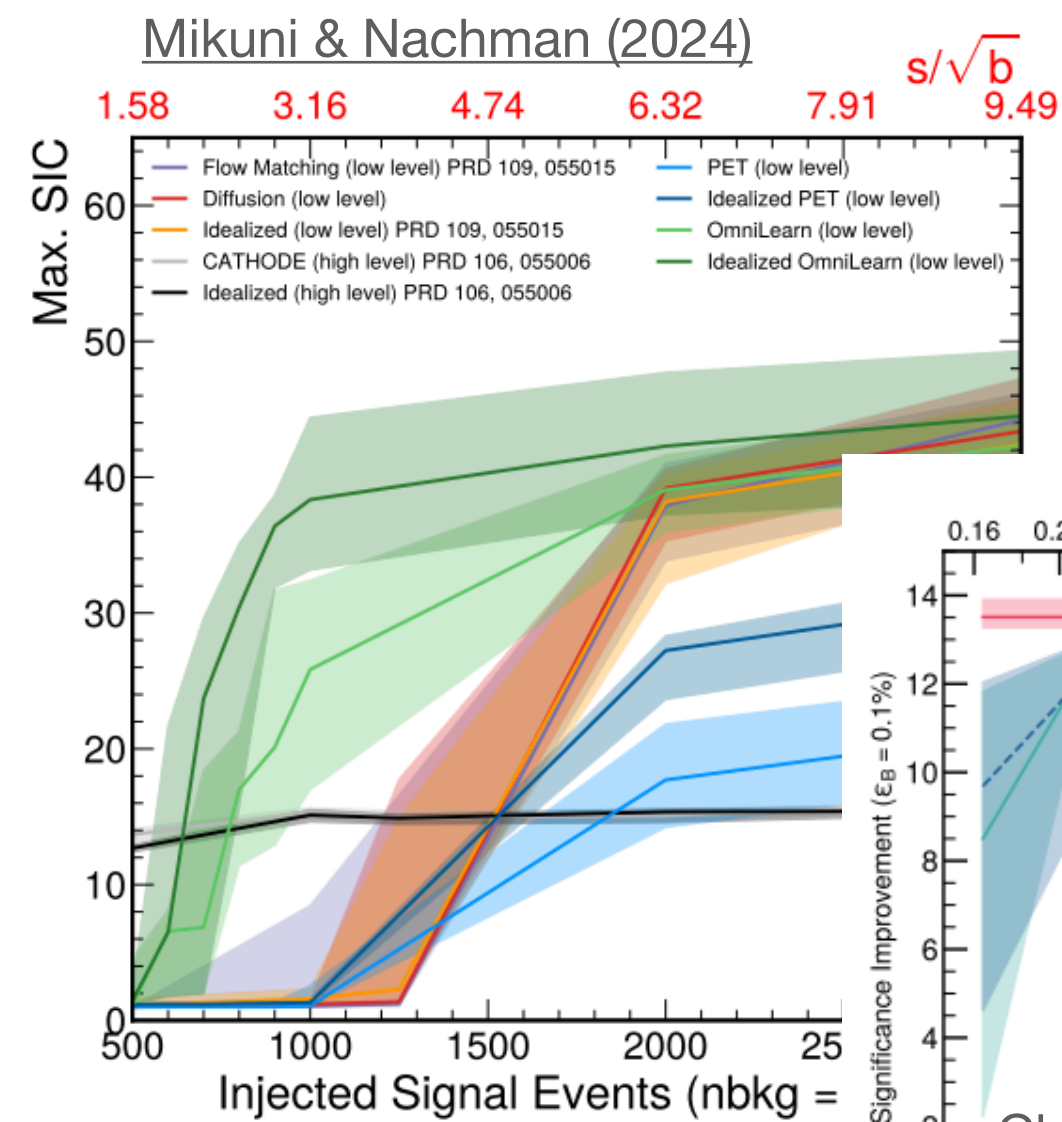
Many more improvements on the way



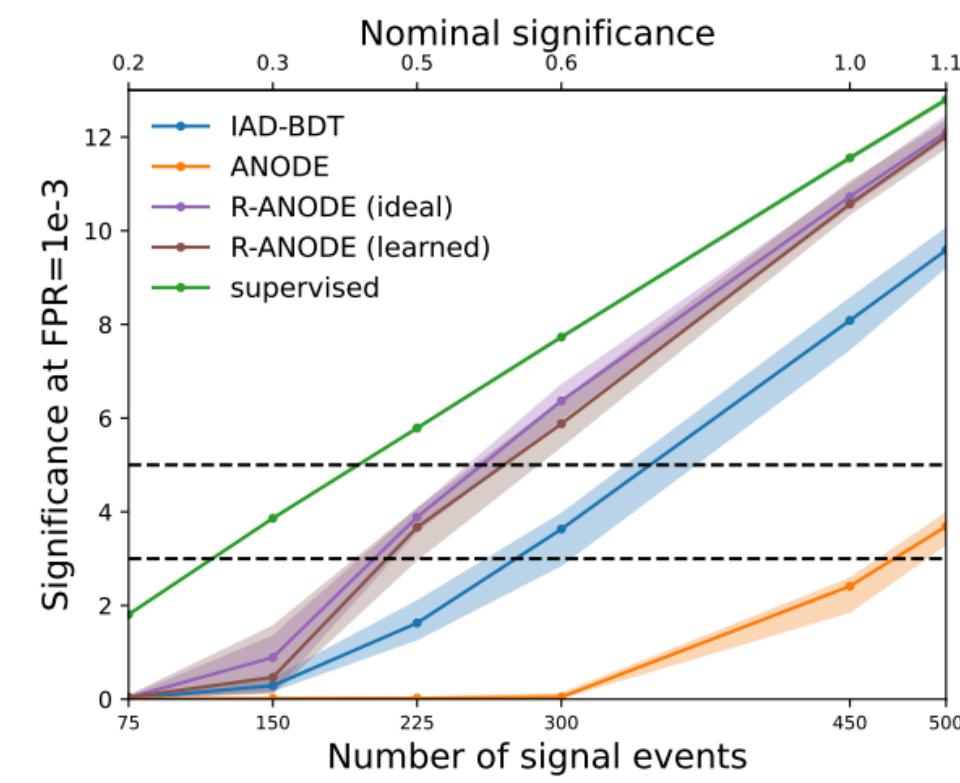
Improved and more robust performance with BDTs and more features



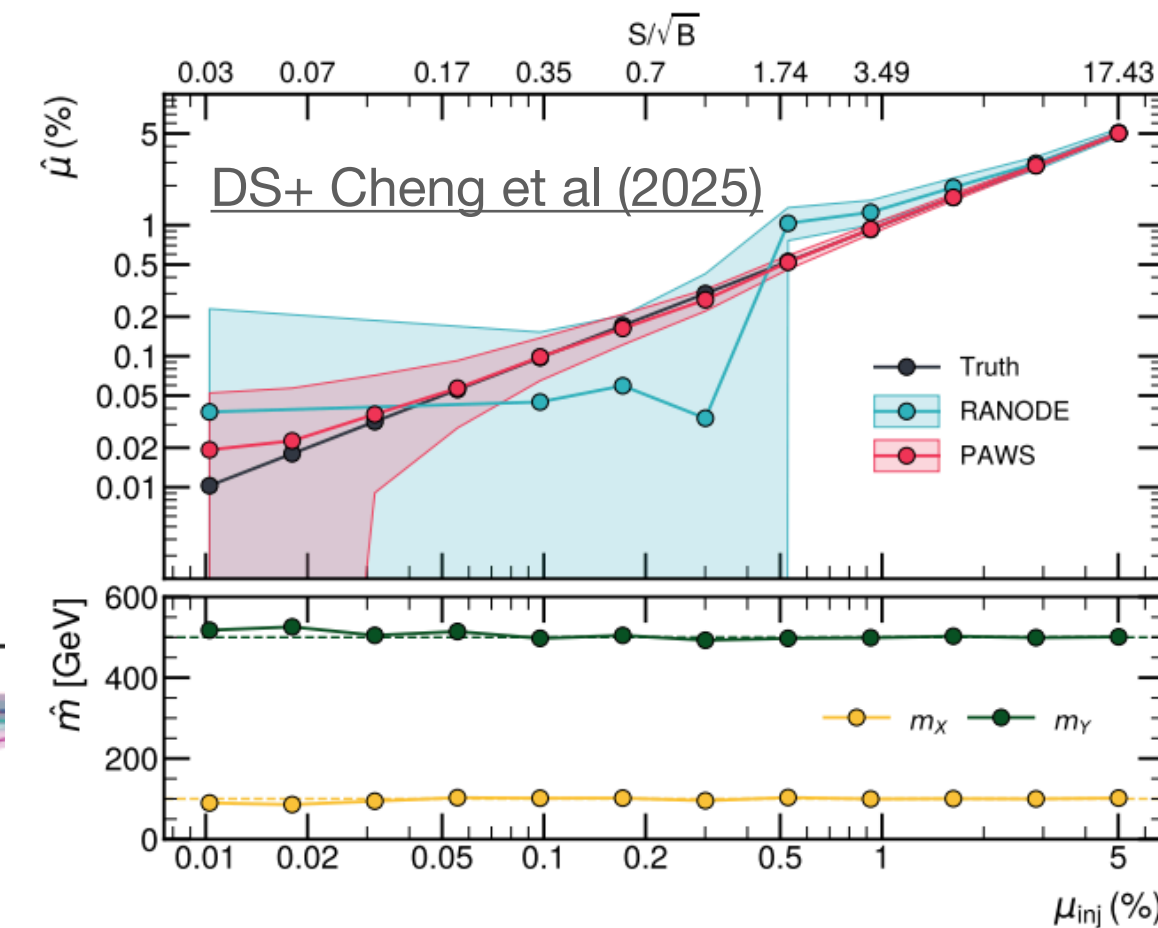
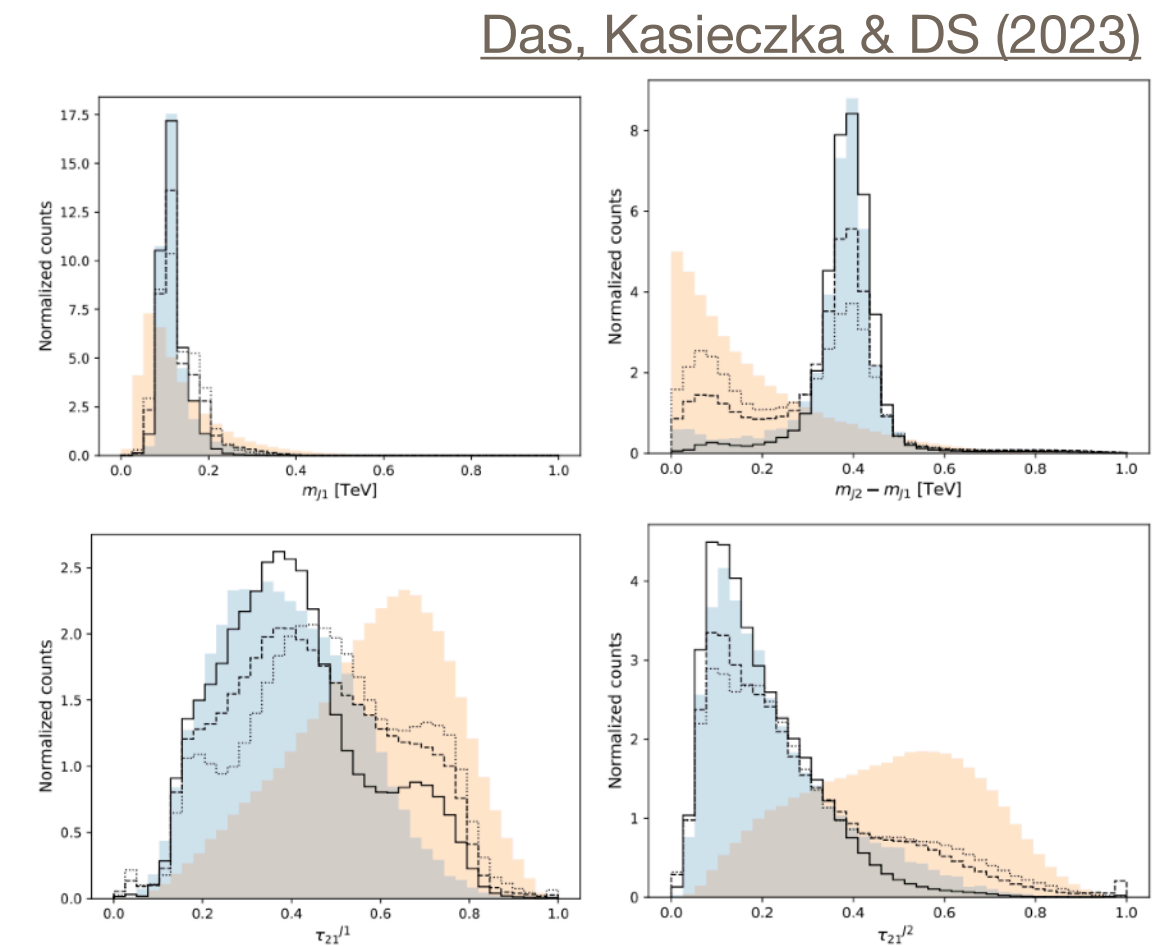
Full phase space



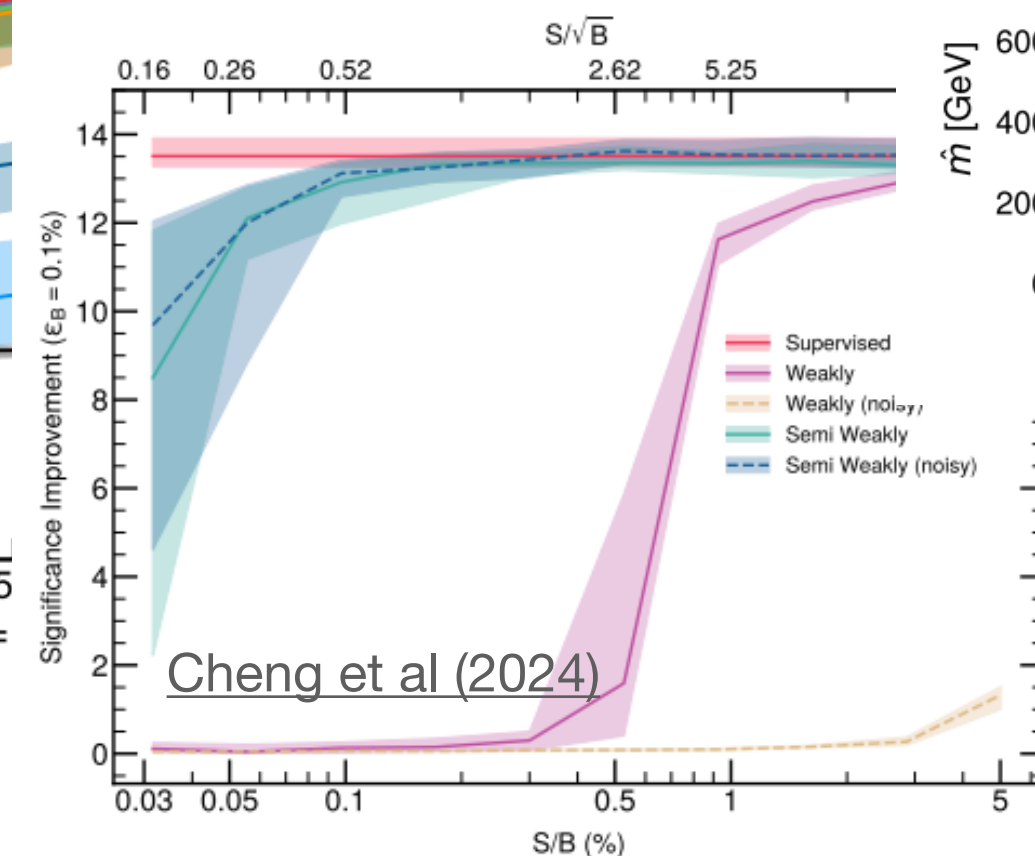
Pretraining with foundation models



Residual ANODE: directly fitting for unknown signal component



“Generator based inference”



Incorporating signal priors

ML for New Physics Searches

The LHC Olympics 2020

A Community Challenge for Anomaly
Detection in High Energy Physics



Gregor Kasieczka (ed),¹ Benjamin Nachman (ed),^{2,3} David Shih (ed),⁴ Oz Amram,⁵ Anders Andreassen,⁶ Kees Benkendorfer,^{2,7} Blaz Bortolato,⁸ Gustaaf Brooijmans,⁹ Florencia Canelli,¹⁰ Jack H. Collins,¹¹ Biwei Dai,¹² Felipe F. De Freitas,¹³ Barry M. Dillon,^{8,14} Ioan-Mihail Dinu,⁵ Zhongtian Dong,¹⁵ Julien Donini,¹⁶ Javier Duarte,¹⁷ D. A. Faroughy,¹⁰ Julia Gonski,⁹ Philip Harris,¹⁸ Alan Kahn,⁹ Jernej F. Kamenik,^{8,19} Charanjit K. Khosa,^{20,30} Patrick Komiske,²¹ Luc Le Pottier,^{2,22} Pablo Martín-Ramiro,^{2,23} Andrej Matevc,^{8,19} Eric Metodiev,²¹ Vinicius Mikuni,¹⁰ Inês Ochoa,²⁴ Sang Eon Park,¹⁸ Maurizio Pierini,²⁵ Dylan Rankin,¹⁸ Veronica Sanz,^{20,26} Nilai Sarda,²⁷ Uroš Seljak,^{2,3,12} Aleks Smolkovic,⁸ George Stein,^{2,12} Cristina Mantilla Suarez,⁵ Manuel Szwec,²⁸ Jesse Thaler,²¹ Steven Tsan,¹⁷ Silviu-Marian Udrescu,¹⁸ Louis Vaslin,¹⁶ Jean-Roch Vlimant,²⁹ Daniel Williams,⁹ Mikael Yunus¹⁸

<https://arxiv.org/abs/2101.08320>

<https://arxiv.org/abs/2105.14027>

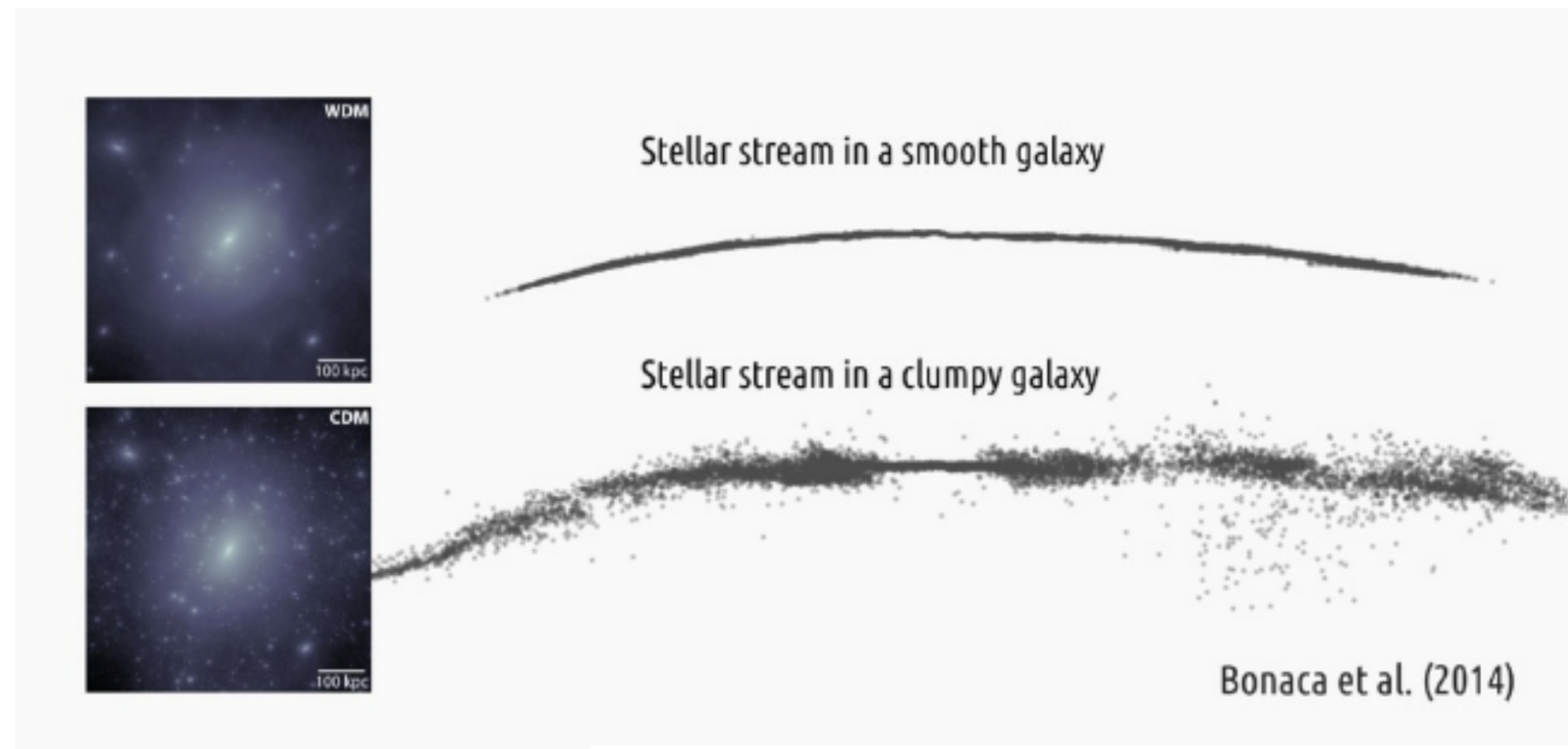
The Dark Machines Anomaly Score Challenge: Benchmark Data and Model Independent Event Classification for the Large Hadron Collider

T. Aarrestad^a M. van Beekveld^b M. Bona^c A. Boveia^e S. Caron^d J. Davies^c
A. De Simone^{f,g} C. Doglioni^h J. M. Duarteⁱ A. Farbin^j H. Gupta^k L. Hendriks^d
L. Heinrich^a J. Howarth^l P. Jawahar^{m,a} A. Jueidⁿ J. Lastow^h A. Leinweber^o
J. Mamuzic^p E. Merényi^q A. Morandini^r P. Moskvitina^d C. Nellist^d J. Ngadiuba^{s,t}
B. Ostdiek^{u,v} M. Pierini^a B. Ravina^l R. Ruiz de Austri^p S. Sekmen^w
M. Touranakou^{x,a} M. Vaškevičiūtė^l R. Vilalta^y J.-R. Vlimant^t R. Verheyen^z
M. White^o E. Wulff^h E. Wallin^h K.A. Wozniak^{α,a} Z. Zhang^d

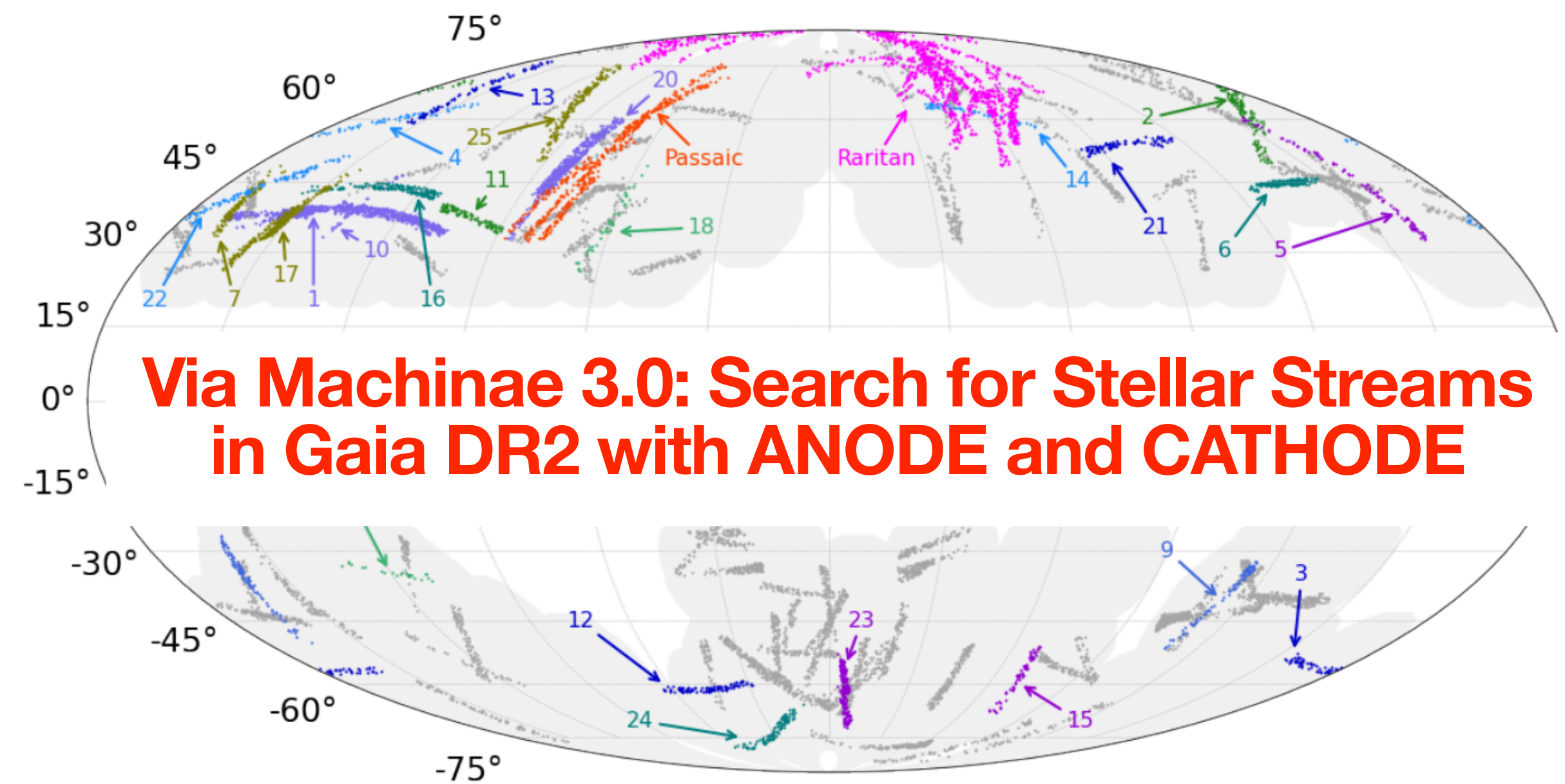
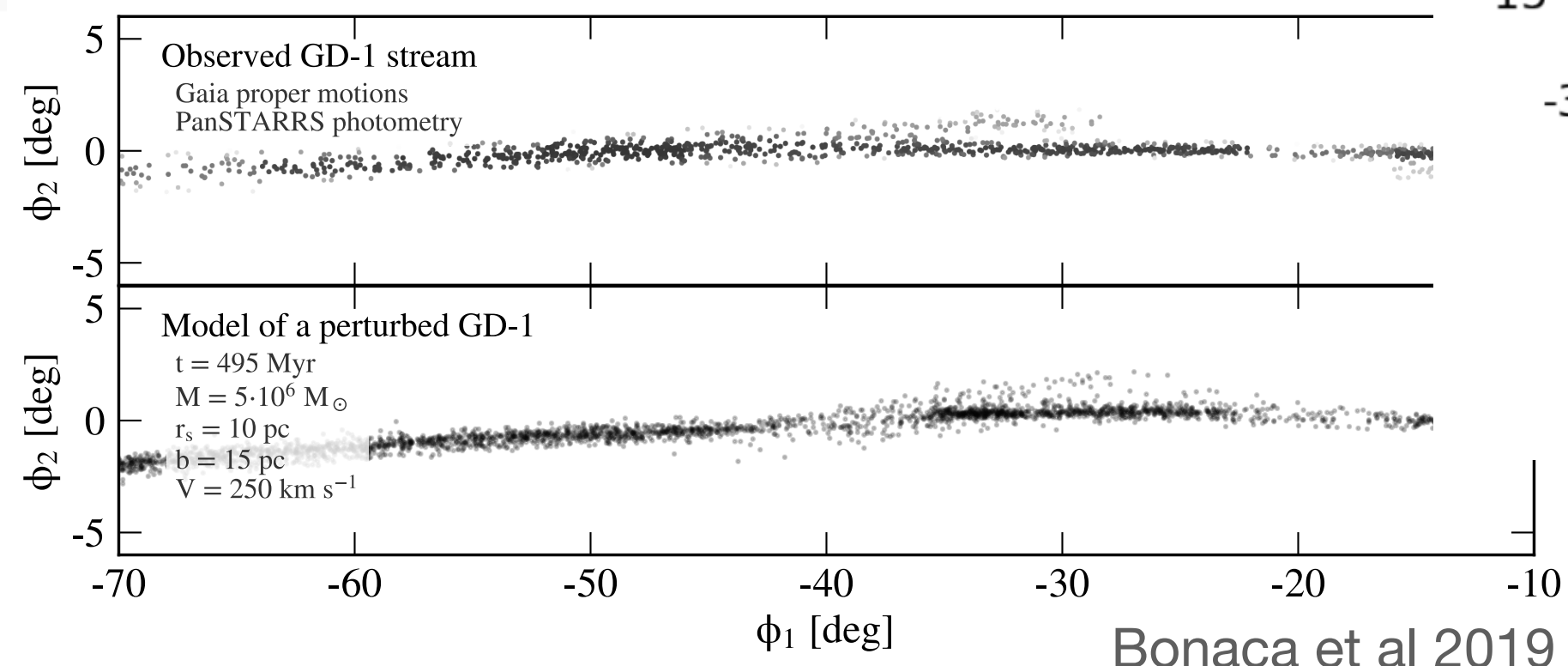
A lot of new ideas for model-agnostic searches!

From LHC to Astro: Search for Stellar Streams

The same resonant AD techniques can be used to search for **stellar streams** in **astro data**



Streams are potentially unique astrophysical probes of dark matter

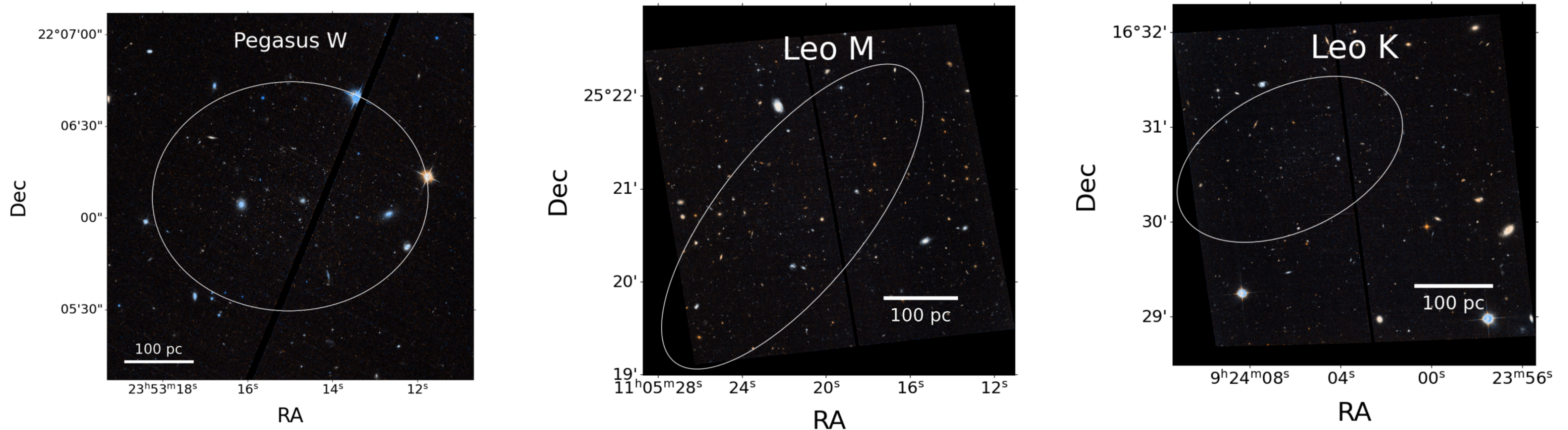


Via Machinae 3.0: Search for Stellar Streams in Gaia DR2 with ANODE and CATHODE

[Hallin, DS, Krause, Buckley 2509.08064]

Spinoff: discovering new ultra-faint dwarf galaxies

[DS+ McQuinn et al (2023a), (2023b)]



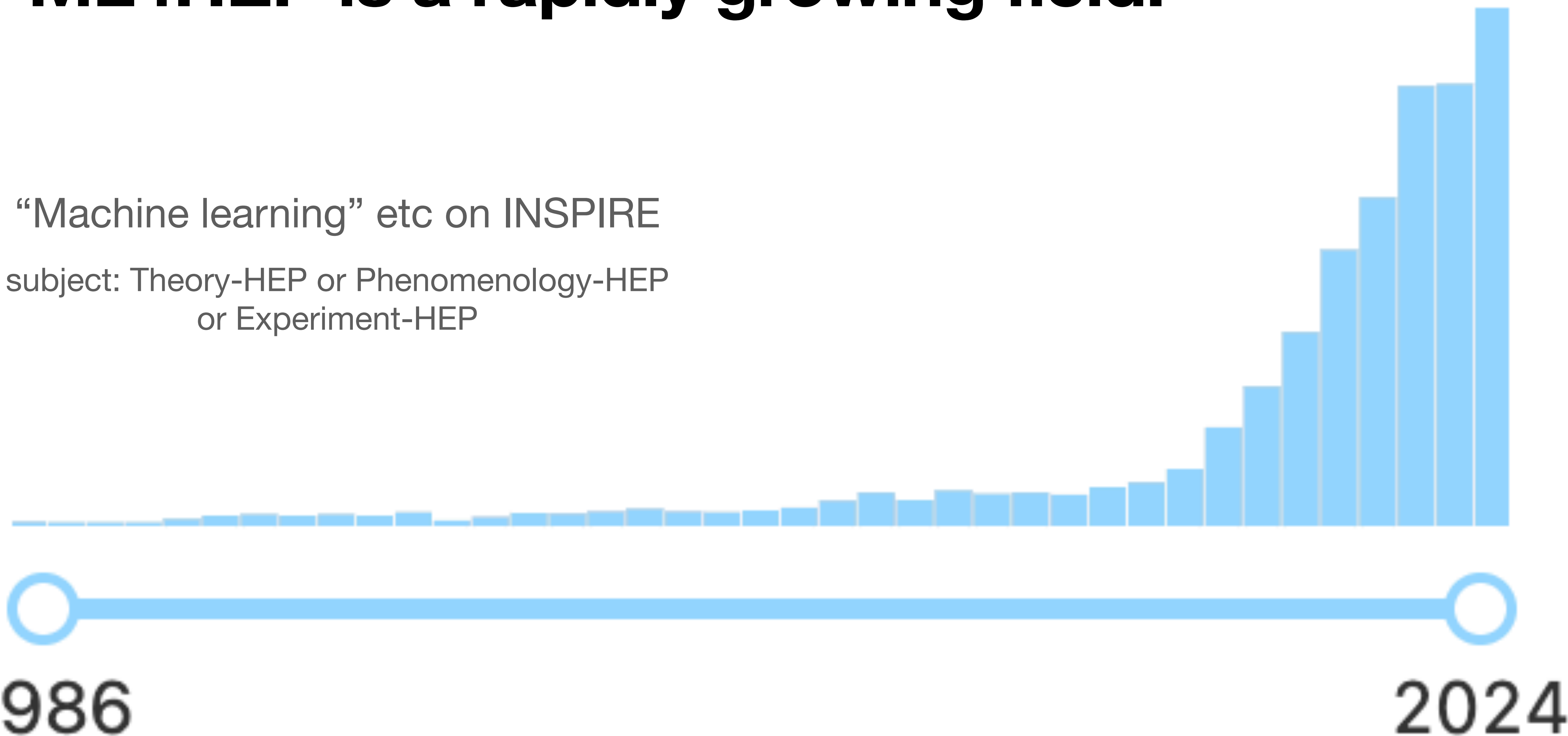
- Helped astronomers to discover 3 new ultra-faint “field” dwarf galaxies
 - First as overdensities of star counts in DESI Legacy Survey data
 - Then confirmed with dedicated Hubble time

Backup

ML4HEP is a rapidly growing field!

“Machine learning” etc on INSPIRE

subject: Theory-HEP or Phenomenology-HEP
or Experiment-HEP

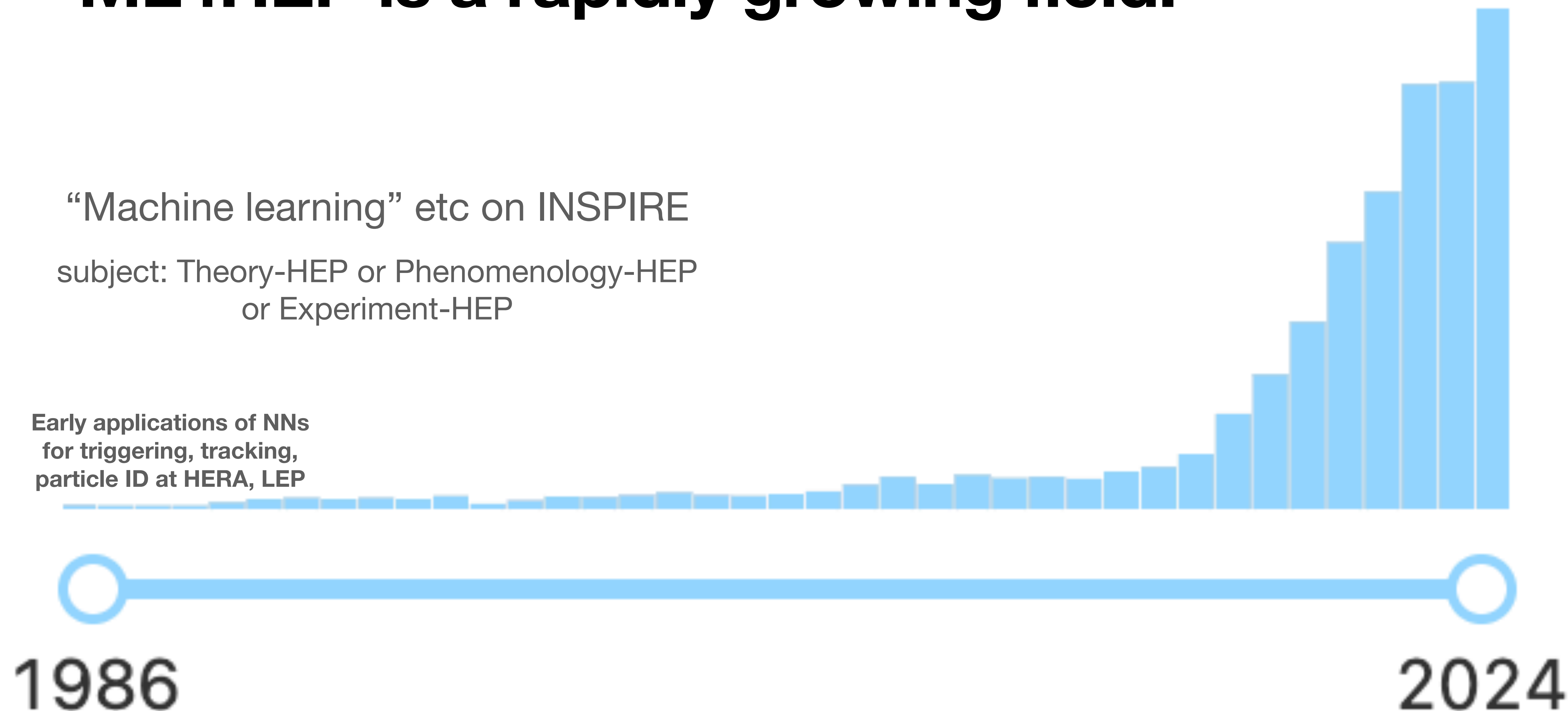


ML4HEP is a rapidly growing field!

“Machine learning” etc on INSPIRE

subject: Theory-HEP or Phenomenology-HEP
or Experiment-HEP

Early applications of NNs
for triggering, tracking,
particle ID at HERA, LEP



ML4HEP is a rapidly growing field!

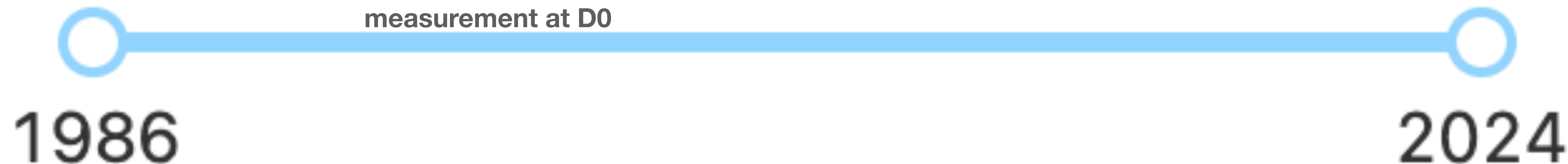
“Machine learning” etc on INSPIRE

subject: Theory-HEP or Phenomenology-HEP
or Experiment-HEP

Early applications of NNs
for triggering, tracking,
particle ID at HERA, LEP

1997

NN for top quark mass
measurement at D0



ML4HEP is a rapidly growing field!

“Machine learning” etc on INSPIRE

subject: Theory-HEP or Phenomenology-HEP
or Experiment-HEP

Early applications of NNs
for triggering, tracking,
particle ID at HERA, LEP

1997

NN for top quark mass
measurement at D0

2012

Start of the “Deep Learning Revolution”

2024

ML4HEP is a rapidly growing field!

“Machine learning” etc on INSPIRE

subject: Theory-HEP or Phenomenology-HEP
or Experiment-HEP

4000+ papers and counting!

Early applications of NNs
for triggering, tracking,
particle ID at HERA, LEP

1997

NN for top quark mass
measurement at D0

2012

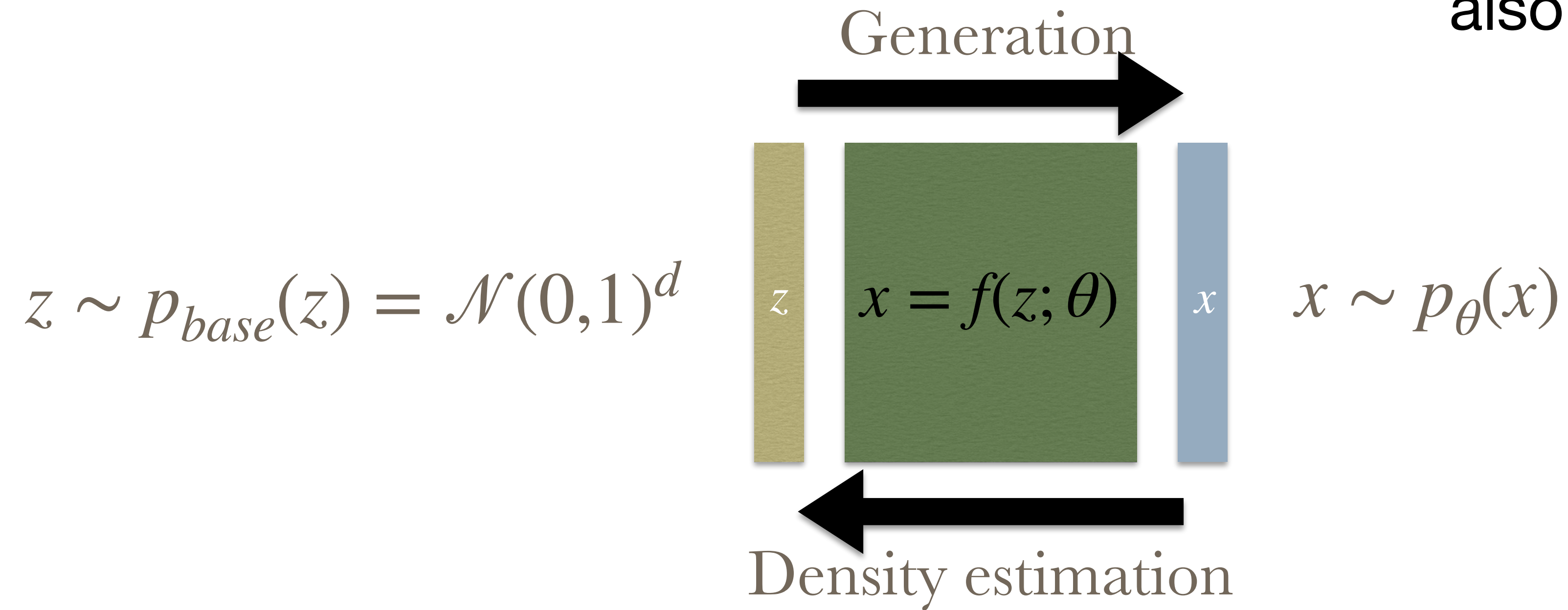
Start of the “Deep Learning Revolution”

1986

2024

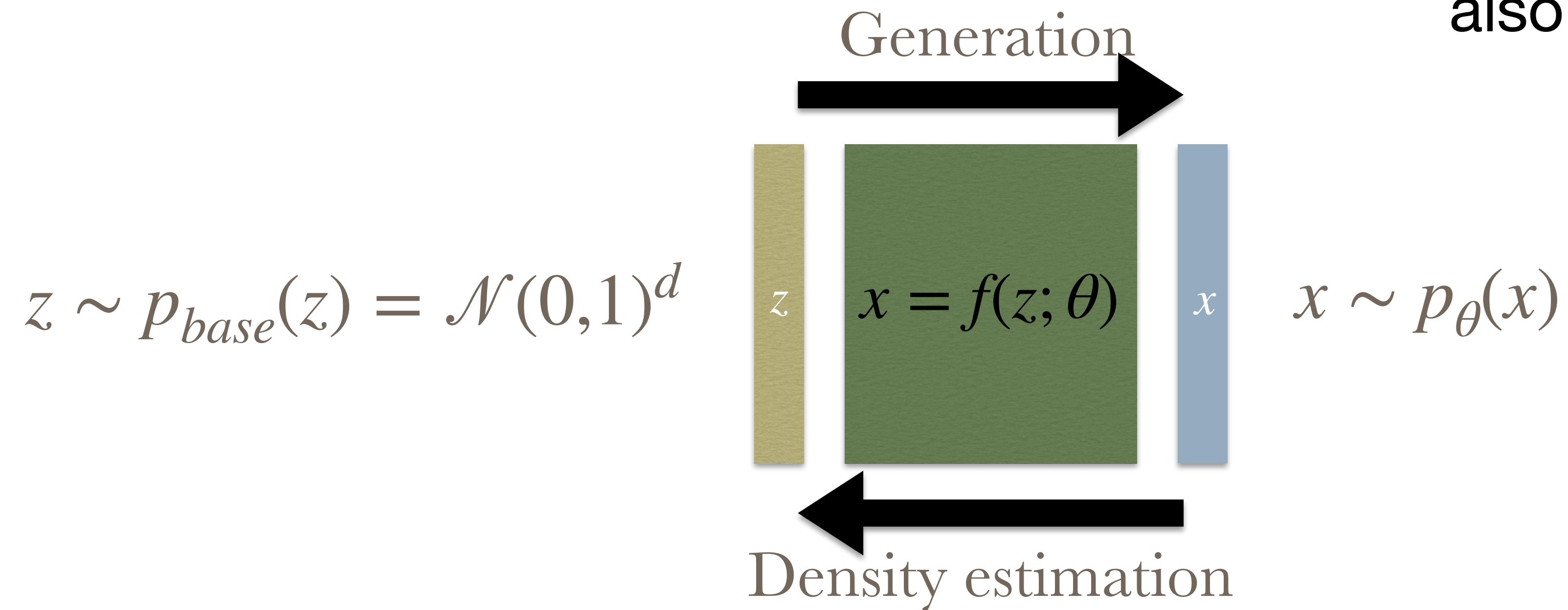
Example: Normalizing Flows

Powerful class of generative models that are also density estimators



Example: Normalizing Flows

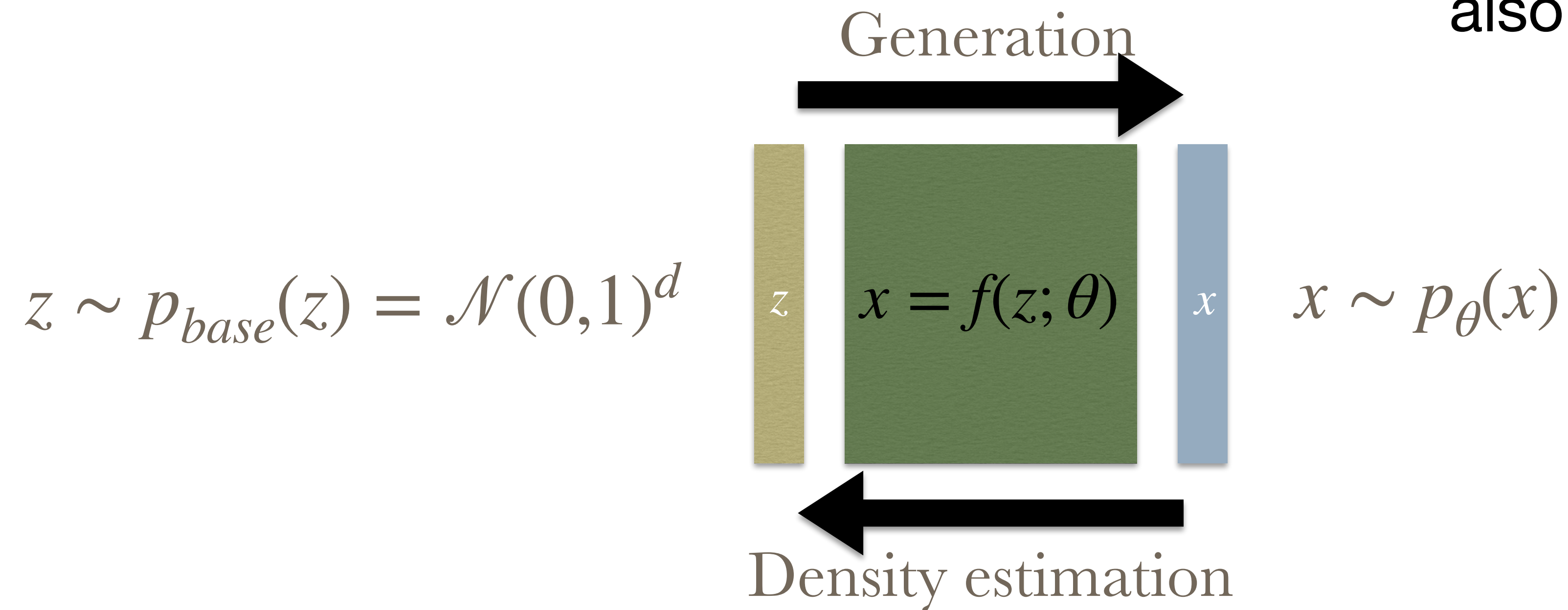
Powerful class of generative models that are also density estimators



- Family of invertible maps $f(z; \theta)$ parametrized by neural networks

Example: Normalizing Flows

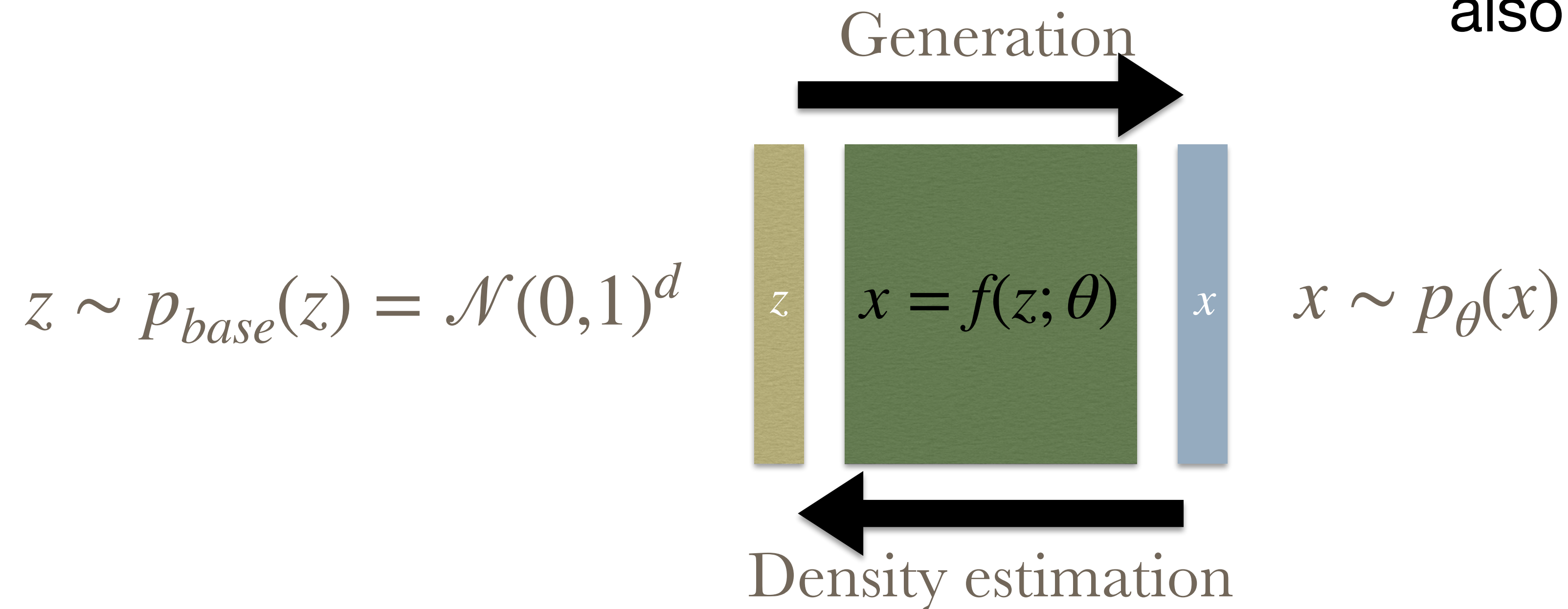
Powerful class of generative models that are also density estimators



- Family of invertible maps $f(z; \theta)$ parametrized by neural networks
- Base density and data density related by Jacobian $p_{\theta}(x) = p_{base}(z = f_{\theta}(x)) \left| \frac{\partial z}{\partial x} \right|$

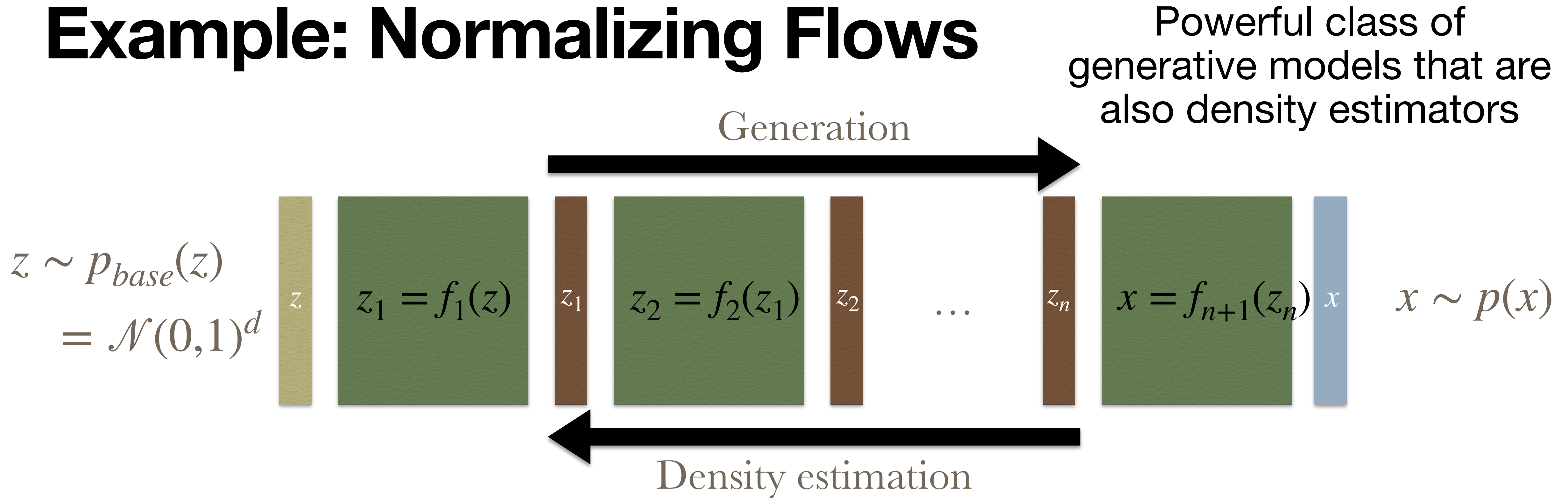
Example: Normalizing Flows

Powerful class of generative models that are also density estimators



- Family of invertible maps $f(z; \theta)$ parametrized by neural networks
- Base density and data density related by Jacobian $p_{\theta}(x) = p_{base}(z = f_{\theta}(x)) \left| \frac{\partial z}{\partial x} \right|$
- Train with maximum likelihood objective $L = - \sum_{x_i \in data} \log p_{\theta}(x_i)$

Example: Normalizing Flows



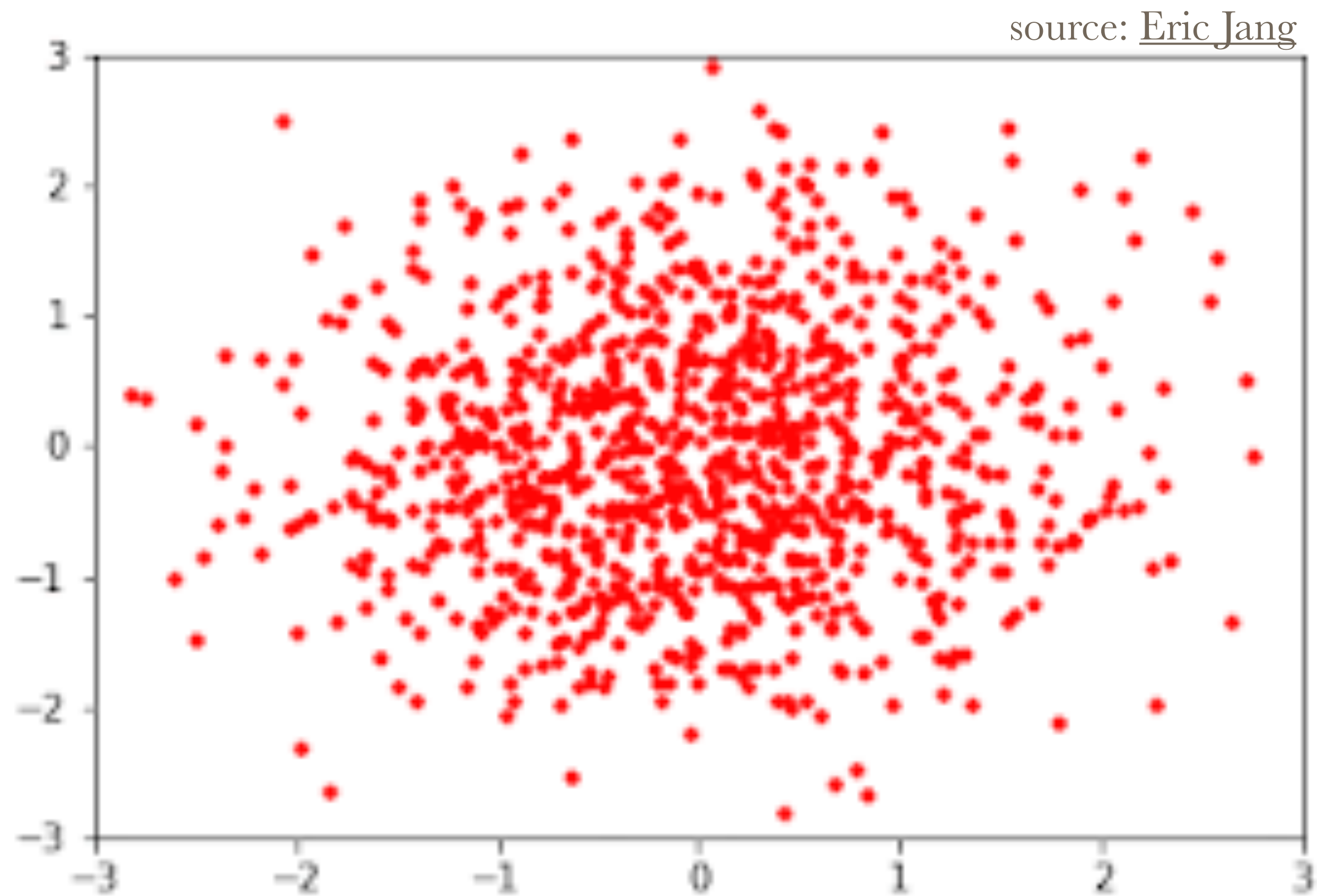
- Family of invertible maps $f(z; \theta)$ parametrized by neural networks

- Base density and data density related by Jacobian $p_{\theta}(x) = p_{base}(z = f_{\theta}(x)) \left| \frac{\partial z}{\partial x} \right|$

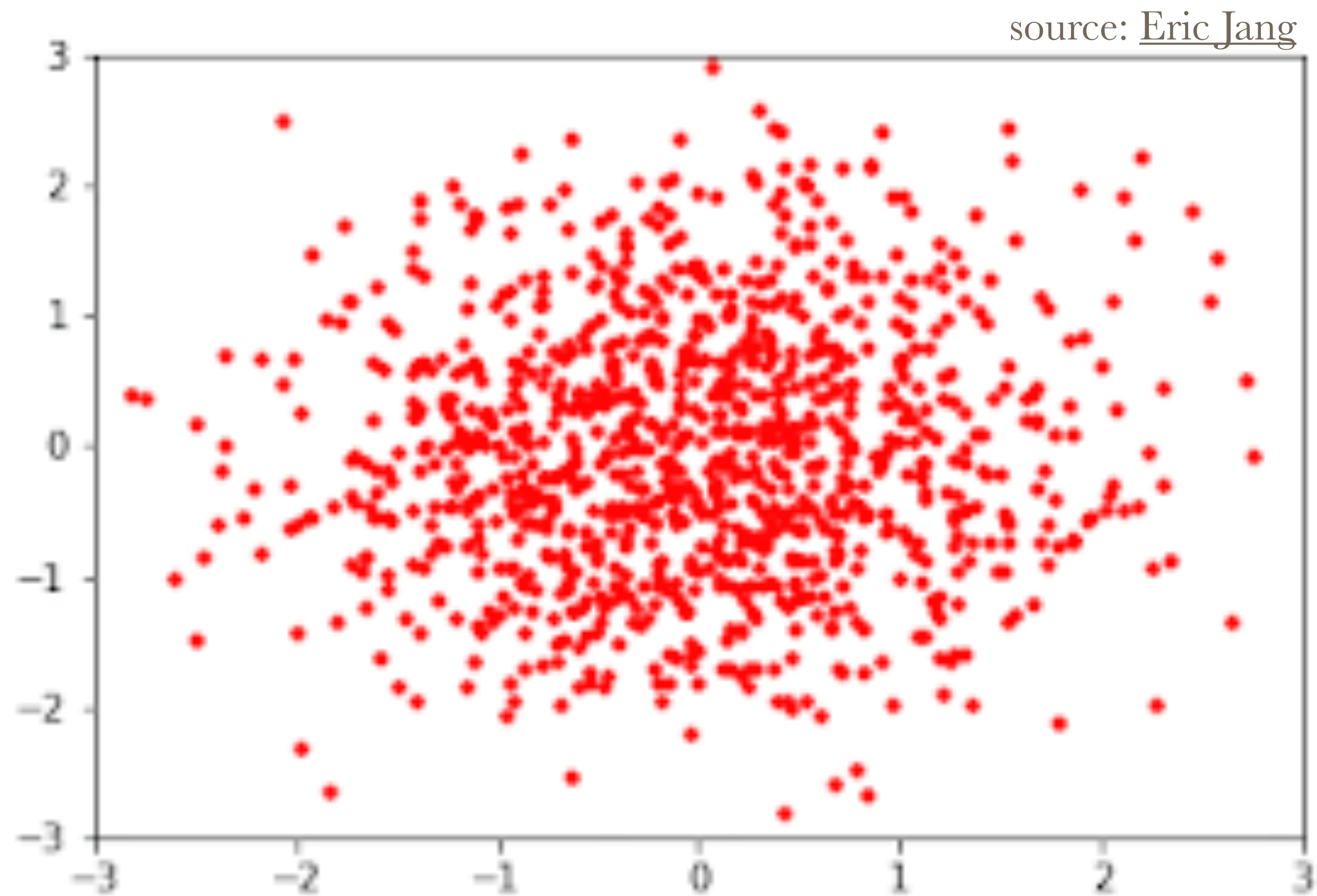
- Train with maximum likelihood objective $L = - \sum_{x_i \in data} \log p_{\theta}(x_i)$

- **Compose multiple maps for greater expressivity**

Example: Normalizing Flows

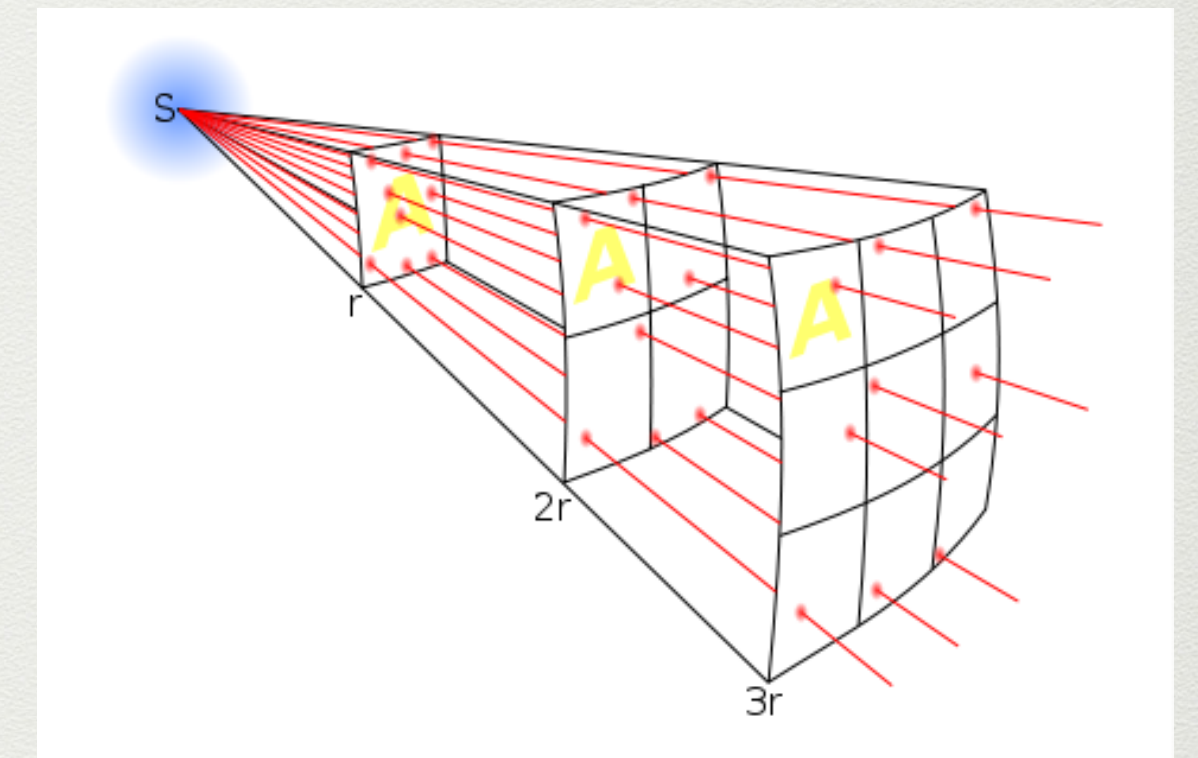


Example: Normalizing Flows

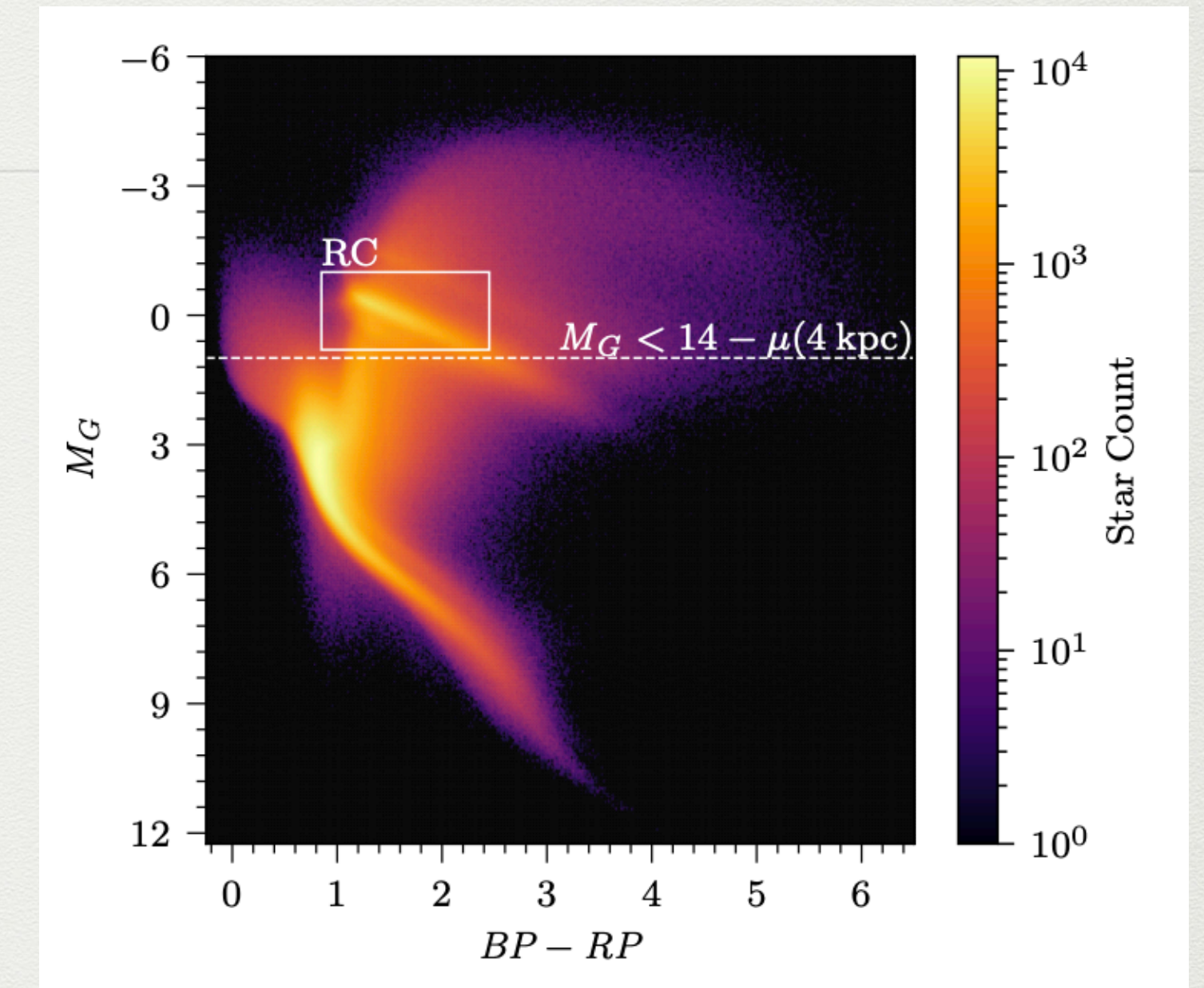
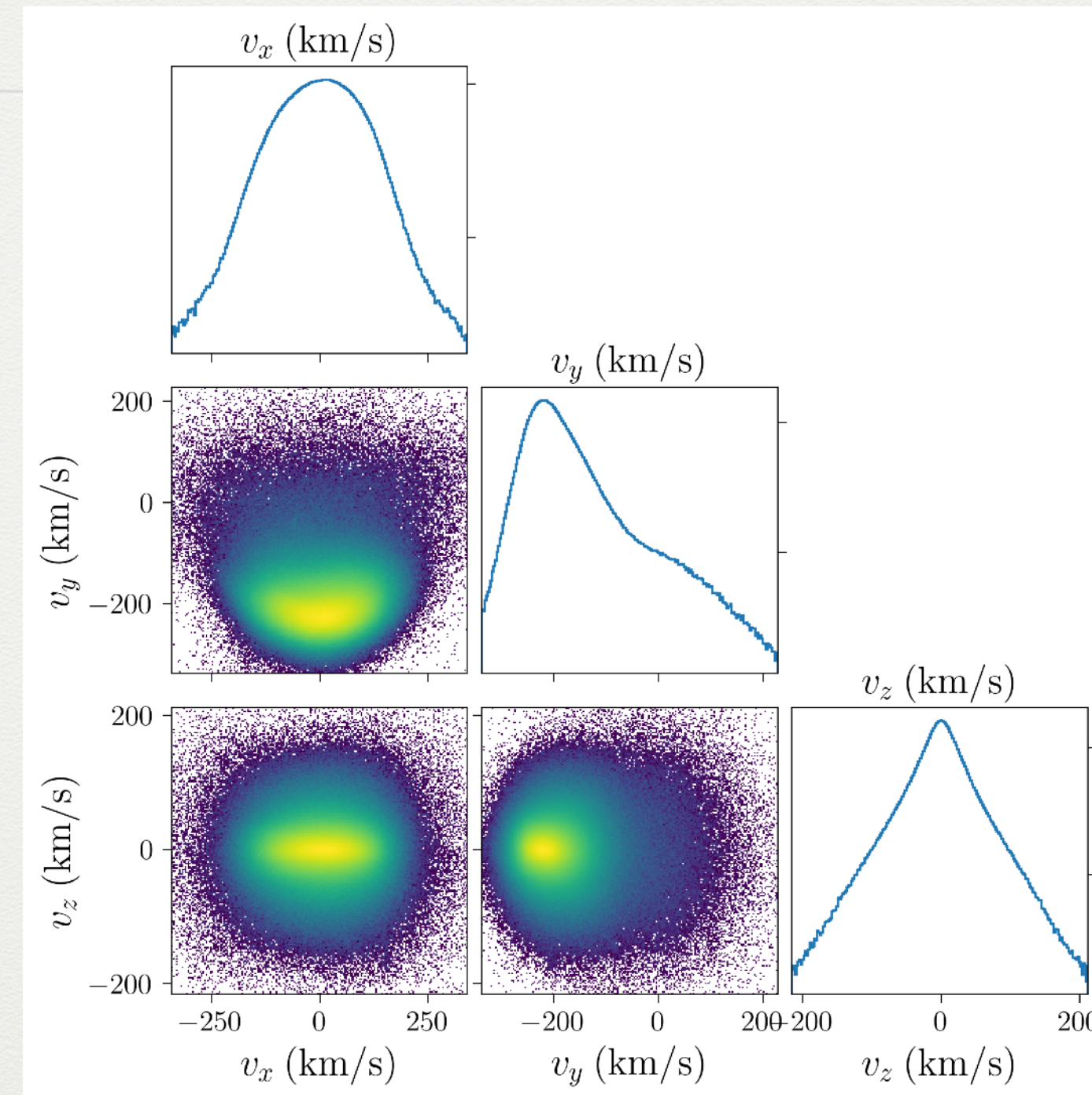
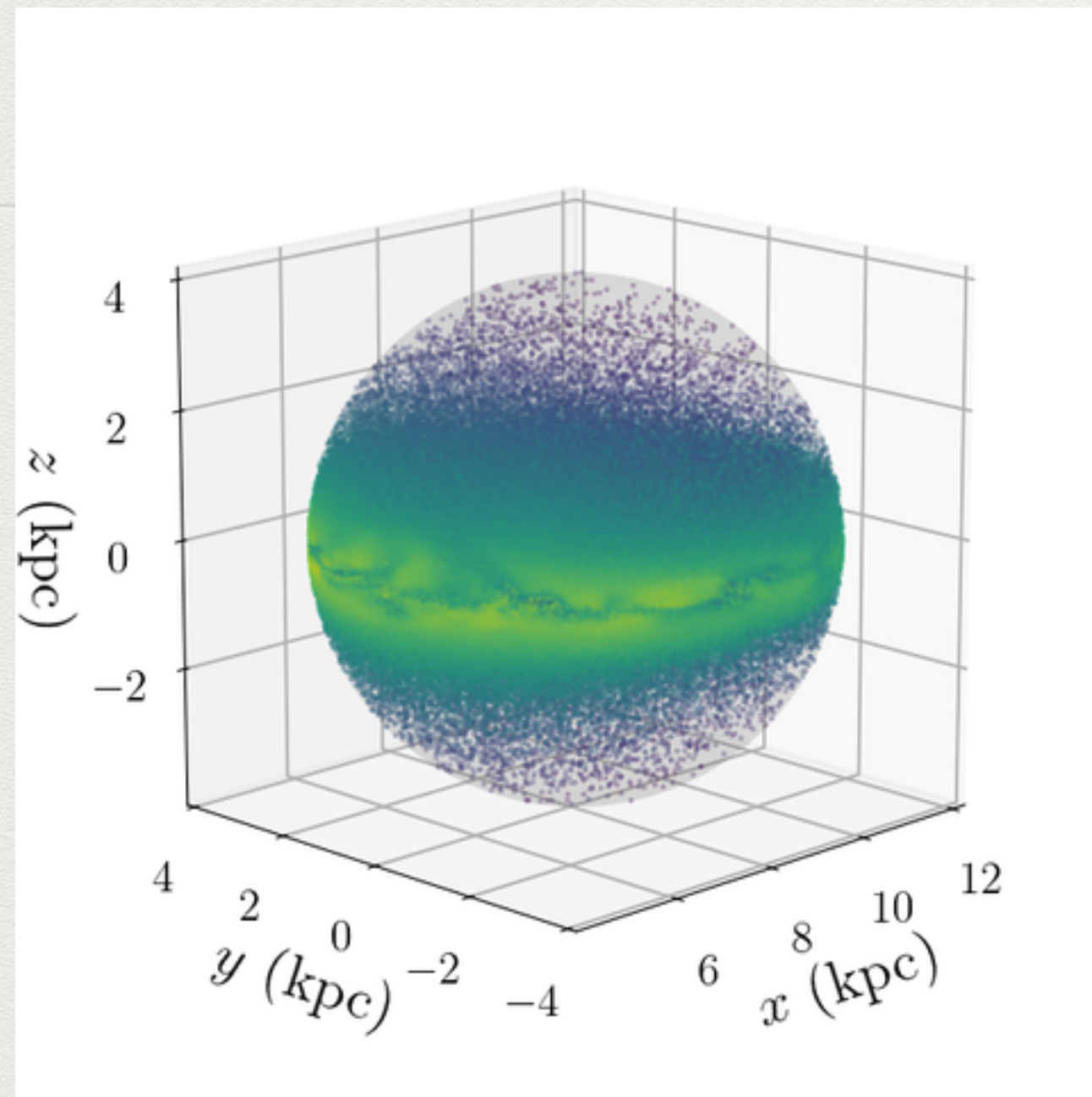


Gaia data

- $\sim 35\text{M}$ stars in DR3 with 6d kinematics, located within ~ 4 kpc
- Want a **complete** and **unbiased** subset
- Gaia is complete up to observed magnitude $G_{obs} < 14$
- However, more distant stars are dimmer
- **For an unbiased subset, require all stars to be bright enough to have been observable anywhere within 4 kpc volume**



Gaia data



- **This results in 6M stars**
- Includes Red Clump stars (65%) — old, bright red giants believed to be in equilibrium, widely used already in Jeans analysis

Back to the CBE

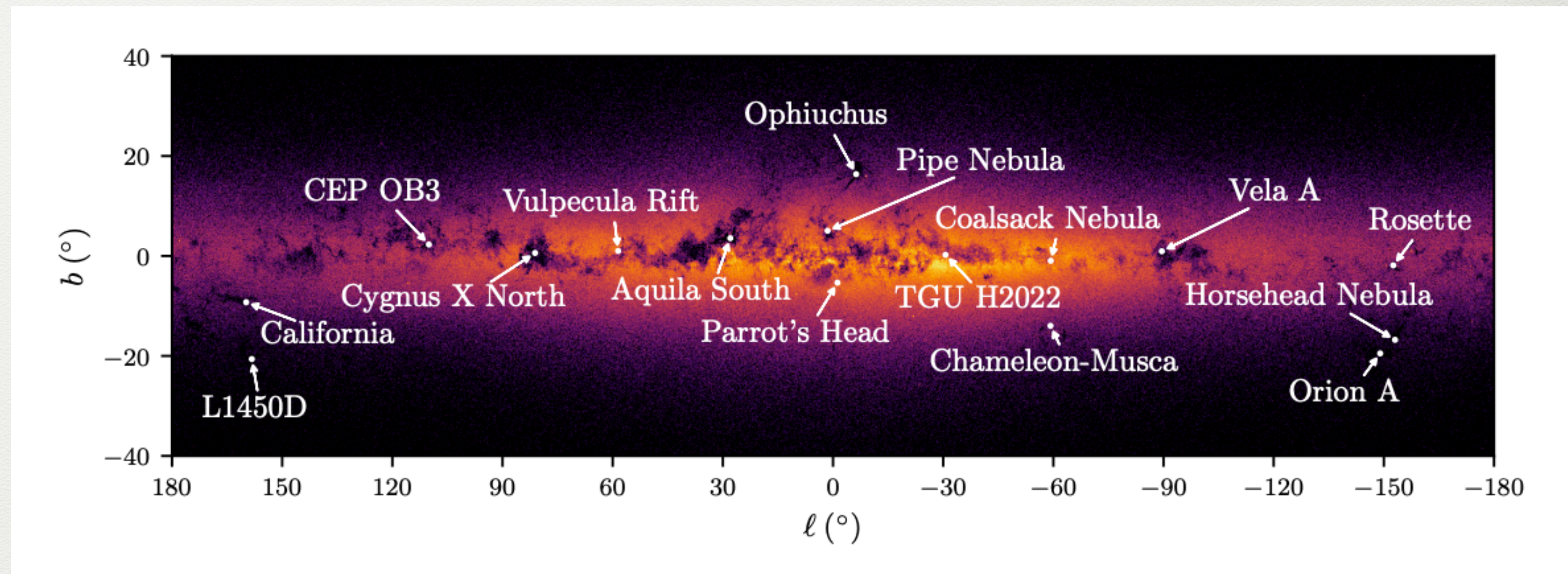
$$\vec{v} \cdot \frac{\partial f}{\partial \vec{x}} + \vec{a}(x) \cdot \frac{\partial f}{\partial \vec{v}} = 0$$

- Just one scalar equation for three acceleration components.
Underdetermined system?
- But acceleration only depends on position! So actually many equations (one for each velocity) at the same position.
Highly overdetermined system!
- Many ways to solve. Our approach: minimize MSE weighted by phase space density

$$\sum_{\vec{v} \sim f(\vec{v}|\vec{x})} \left| \vec{v} \cdot \frac{\partial f}{\partial \vec{x}} + \vec{a}(x) \cdot \frac{\partial f}{\partial \vec{v}} \right|^2$$

Problem: interstellar dust

- But wait! What are all those blotches in the number density?
- Interstellar dust!
- Blocks many lines of sight
- Stars are reddened and dimmed — can fall out of the Gaia dataset
- Affects the completeness of the dataset, biases the phase space density



New version:



Mapping Dark Matter Through the Dust of the Milky Way Part I:
Dust Correction and Phase Space Density

[Eric Putney](#), [David Shih](#), [Sung Hak Lim](#), [Matthew R. Buckley](#) (Dec 18, 2024)

e-Print: [2412.14236](#) [astro-ph.GA]

- New idea: can simultaneously determine Galactic potential and “dust efficiency” function

$$f_{obs}(x, v) = \epsilon(x) f_{corr}(x, v)$$

- **Key point: dust efficiency doesn't depend on star's velocity, just its position!**
- Dust-corrected PSD should satisfy equilibrium CBE

$$\vec{v} \cdot \frac{\partial f_{corr}}{\partial \vec{x}} + \vec{a}(x) \cdot \frac{\partial f_{corr}}{\partial \vec{v}} = 0$$

New version:



Mapping Dark Matter Through the Dust of the Milky Way Part I:
Dust Correction and Phase Space Density

[Eric Putney](#), [David Shih](#), [Sung Hak Lim](#), [Matthew R. Buckley](#) (Dec 18, 2024)

e-Print: [2412.14236](#) [astro-ph.GA]

- New idea: can simultaneously determine Galactic potential and “dust efficiency” function

$$f_{obs}(x, v) = \epsilon(x) f_{corr}(x, v)$$

- **Key point: dust efficiency doesn't depend on star's velocity, just its position!**
- Dust-corrected PSD should satisfy equilibrium CBE

$$\vec{v} \cdot \frac{\partial f_{corr}}{\partial \vec{x}} + \vec{a}(x) \cdot \frac{\partial f_{corr}}{\partial \vec{v}} = 0 \quad \Rightarrow$$

New version:



Mapping Dark Matter Through the Dust of the Milky Way Part I:
Dust Correction and Phase Space Density

Eric Putney, David Shih, Sung Hak Lim, Matthew R. Buckley (Dec 18, 2024)

e-Print: [2412.14236](#) [astro-ph.GA]

- New idea: can simultaneously determine Galactic potential and “dust efficiency” function

$$f_{obs}(x, v) = \epsilon(x) f_{corr}(x, v)$$

- **Key point: dust efficiency doesn't depend on star's velocity, just its position!**
- Dust-corrected PSD should satisfy equilibrium CBE

$$\vec{v} \cdot \frac{\partial f_{corr}}{\partial \vec{x}} + \vec{a}(x) \cdot \frac{\partial f_{corr}}{\partial \vec{v}} = 0 \quad \Rightarrow \quad \vec{v} \cdot \frac{\partial \log f_{obs}}{\partial \vec{x}} - \vec{v} \cdot \nabla \log \epsilon - \nabla \Phi \cdot \frac{\partial \log f_{obs}}{\partial \vec{v}} = 0$$

New version:



Mapping Dark Matter Through the Dust of the Milky Way Part I: Dust Correction and Phase Space Density

Eric Putney, David Shih, Sung Hak Lim, Matthew R. Buckley (Dec 18, 2024)

e-Print: [2412.14236](#) [astro-ph.GA]

$$\vec{v} \cdot \frac{\partial \log f_{obs}}{\partial \vec{x}} - \vec{v} \cdot \nabla \log \epsilon - \nabla \Phi \cdot \frac{\partial \log f_{obs}}{\partial \vec{v}} = 0$$

- Different velocity dependence — still overdetermined system!
- Can simultaneously determine $\Phi(x)$ and $\epsilon(x)$

New version:



Mapping Dark Matter Through the Dust of the Milky Way Part I: Dust Correction and Phase Space Density

Eric Putney, David Shih, Sung Hak Lim, Matthew R. Buckley (Dec 18, 2024)

e-Print: [2412.14236](#) [astro-ph.GA]

$$\vec{v} \cdot \frac{\partial \log f_{obs}}{\partial \vec{x}} - \vec{v} \cdot \nabla \log \epsilon - \nabla \Phi \cdot \frac{\partial \log f_{obs}}{\partial \vec{v}} = 0$$

- Different velocity dependence — still overdetermined system!
- Can simultaneously determine $\Phi(x)$ and $\epsilon(x)$

$$L(\theta, \varphi) = \sum_{(x,v) \sim f(x,v)} \left| \vec{v} \cdot \frac{\partial \log f_{obs}}{\partial \vec{x}} - \vec{v} \cdot \nabla \log \epsilon_{\theta}(x) - \nabla \Phi_{\varphi}(x) \cdot \frac{\partial \log f_{obs}}{\partial \vec{v}} \right|^2$$

New version:



Mapping Dark Matter Through the Dust of the Milky Way Part I:
Dust Correction and Phase Space Density

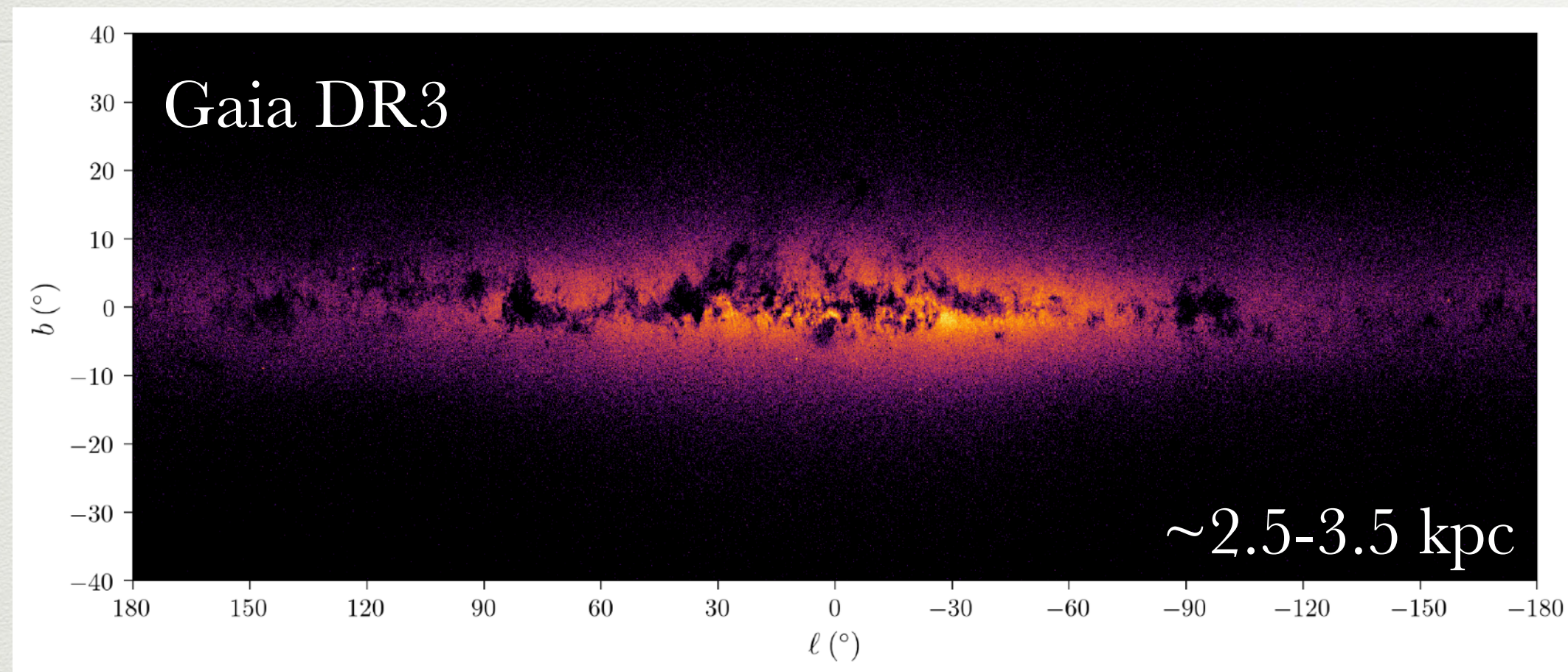
[Eric Putney](#), [David Shih](#), [Sung Hak Lim](#), [Matthew R. Buckley](#) (Dec 18, 2024)

e-Print: [2412.14236](#) [astro-ph.GA]

$$L(\theta, \varphi) = \sum_{(x, v) \sim f(x, v)} \left| \vec{v} \cdot \frac{\partial \log f_{obs}}{\partial \vec{x}} - \vec{v} \cdot \nabla \log \epsilon_{\theta}(x) - \nabla \Phi_{\varphi}(x) \cdot \frac{\partial \log f_{obs}}{\partial \vec{v}} \right|^2$$

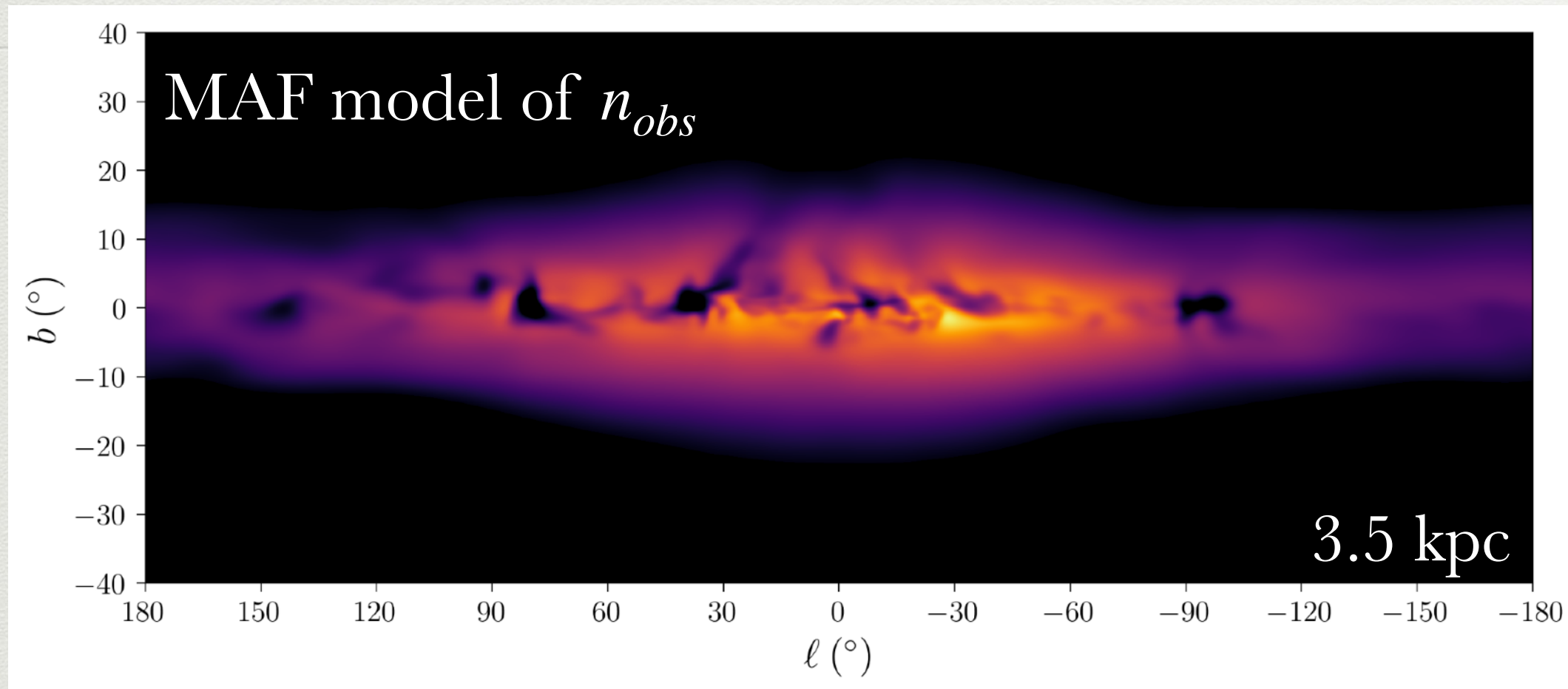
- Parametrize both $\epsilon(x)$ and $\Phi(x)$ with NNs (simple MLPs)
- Advantages
 - Much more computationally efficient compared to previous pointwise approach
 - More physical — guarantees curl-free accelerations

Results: Dust-Corrected PSD



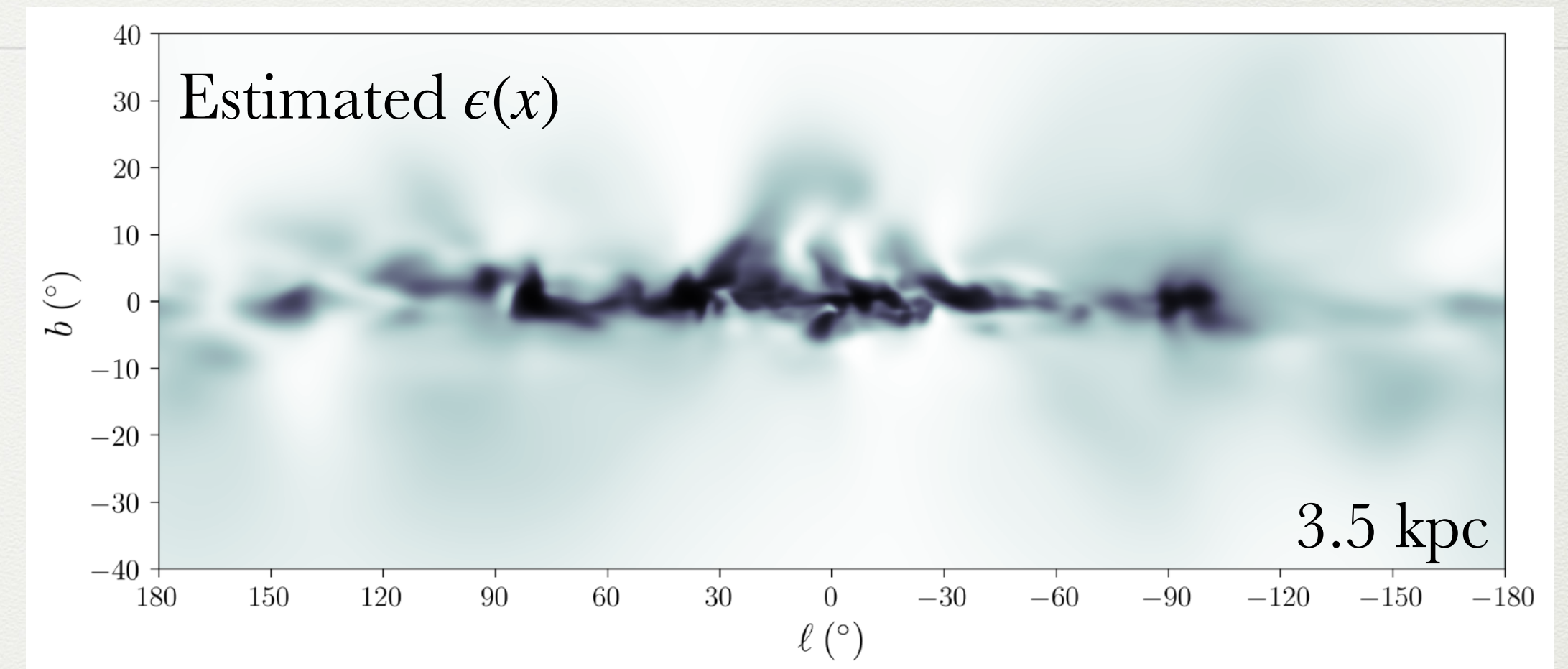
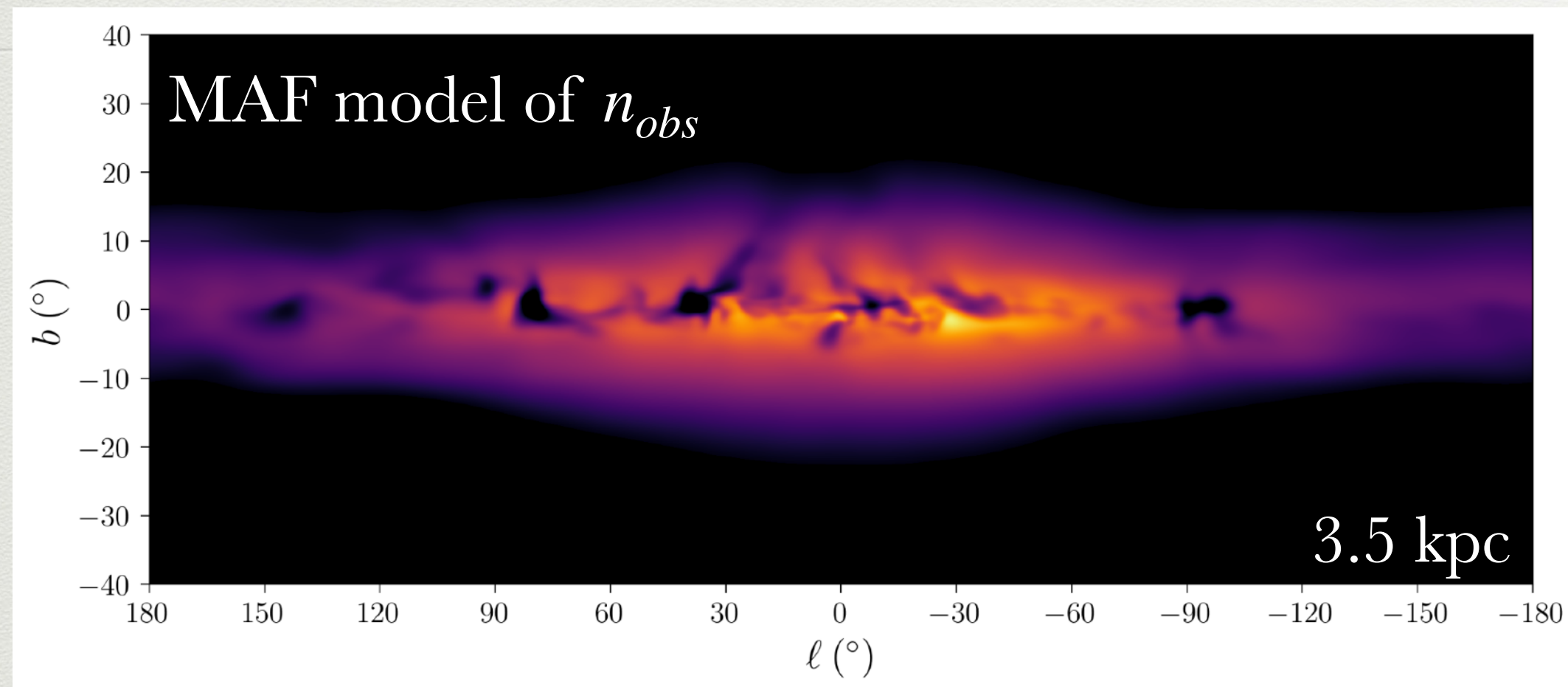
Observed data is biased by dust

Results: Dust-Corrected PSD



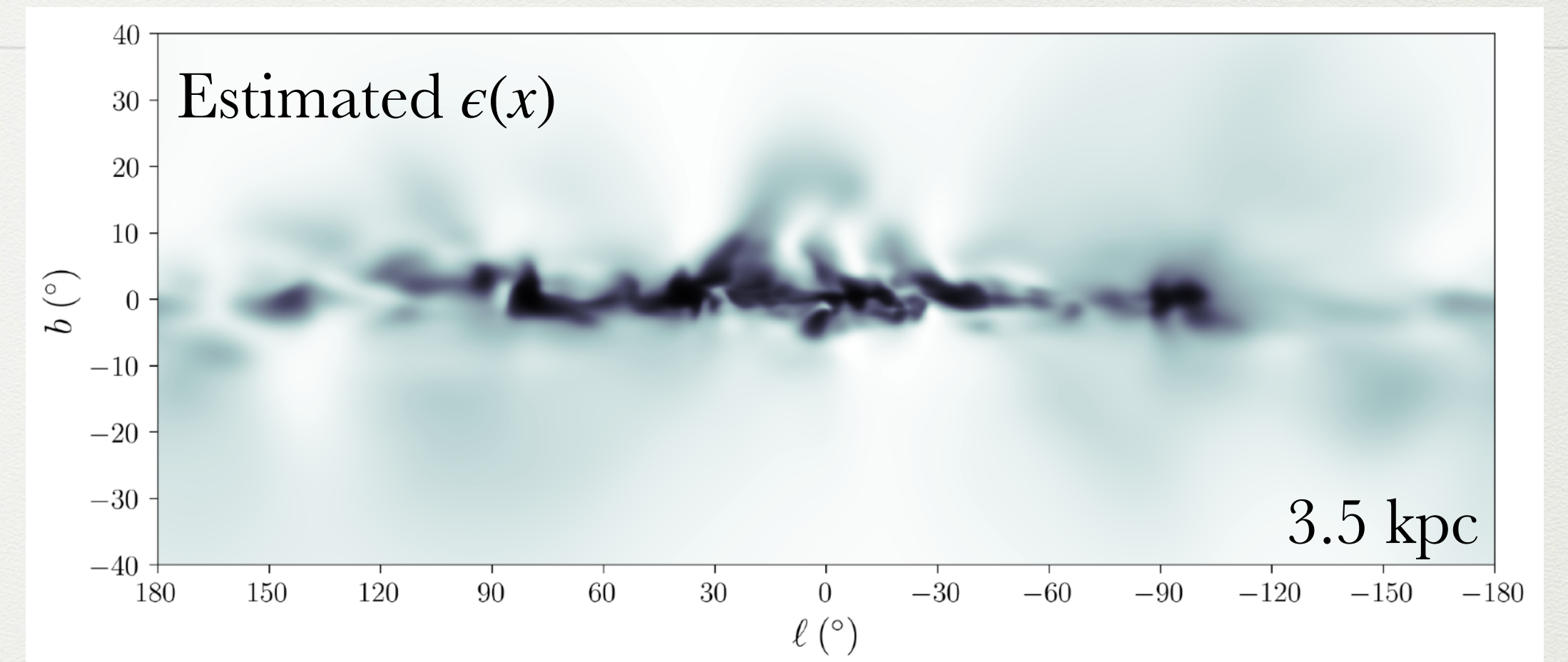
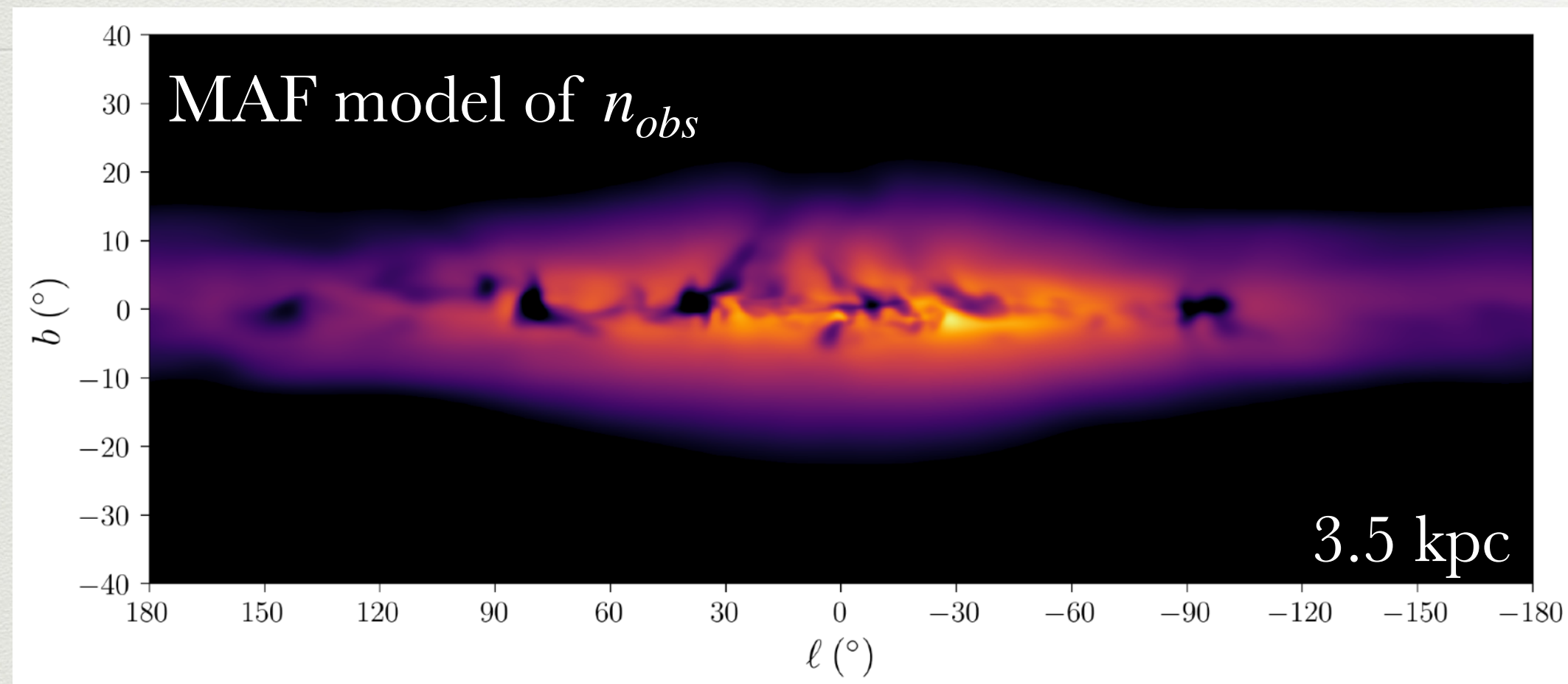
Learned PSD is biased by dust

Results: Dust-Corrected PSD

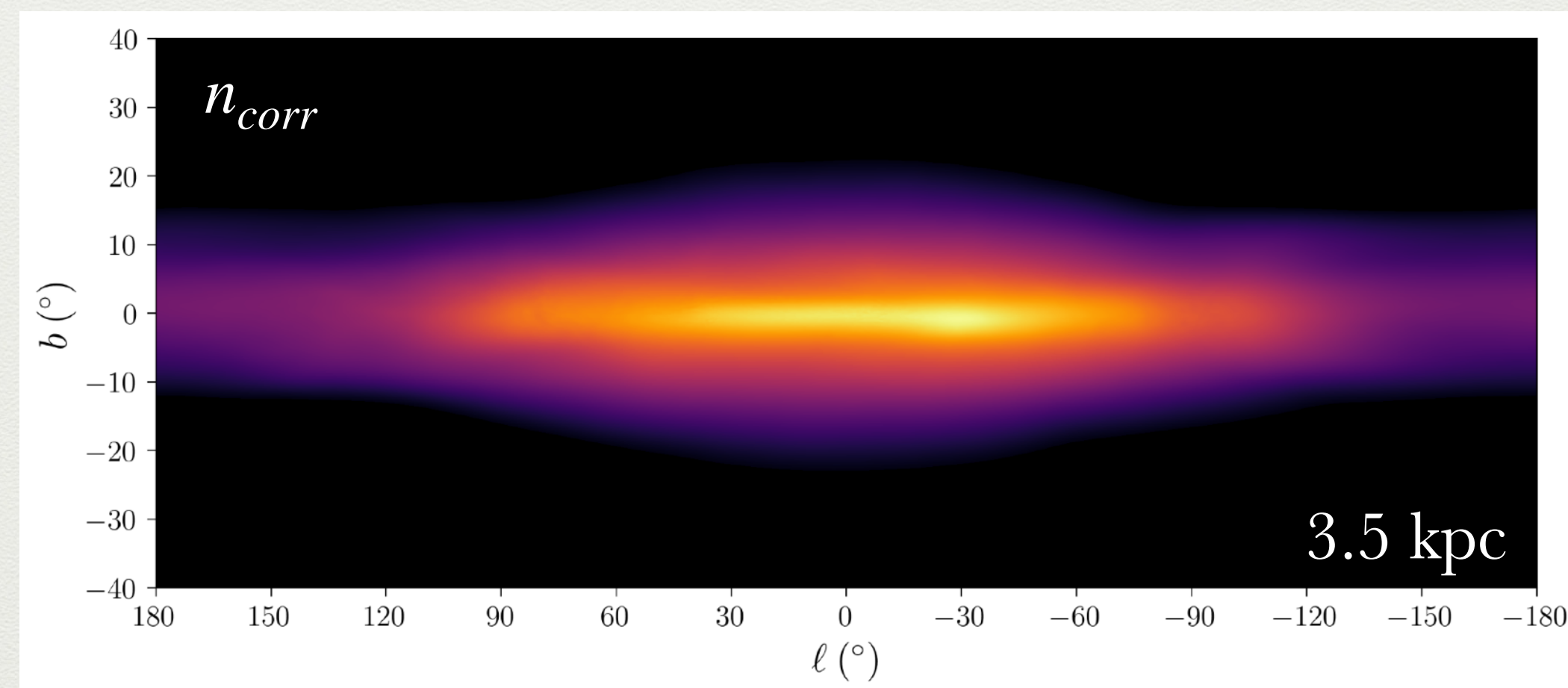


Estimate $\epsilon(x)$ from CBE

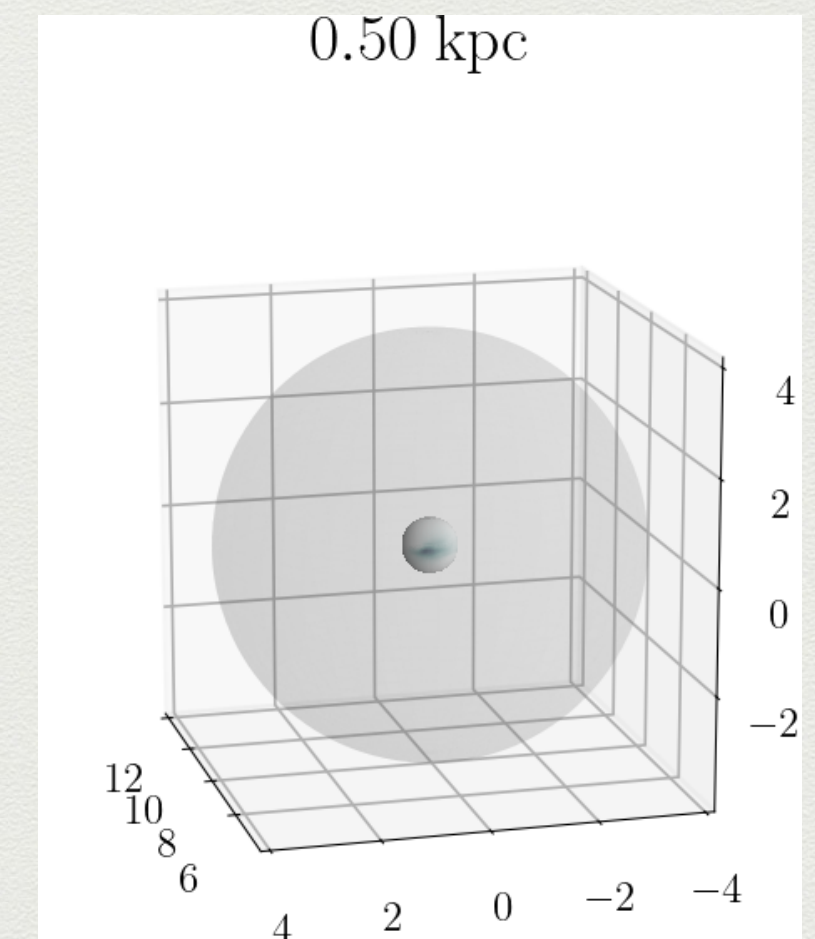
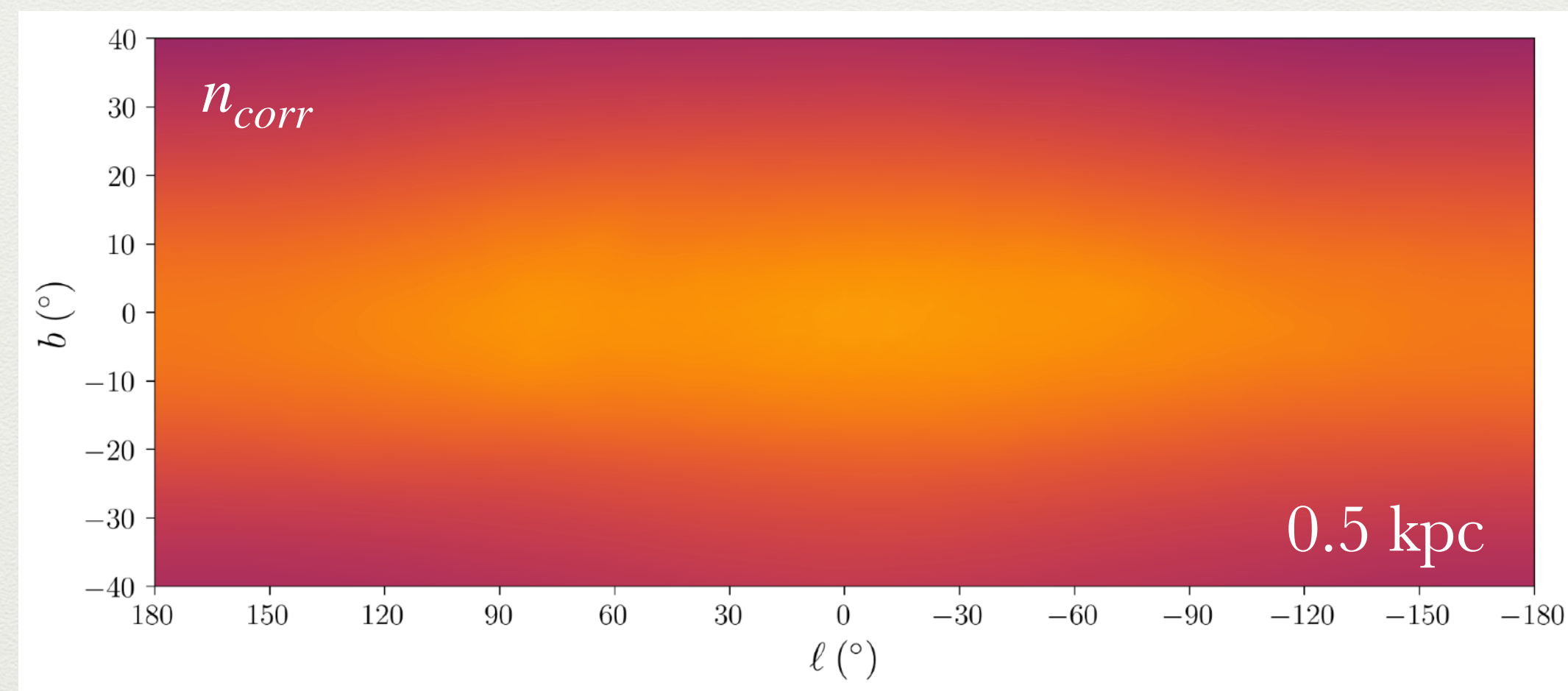
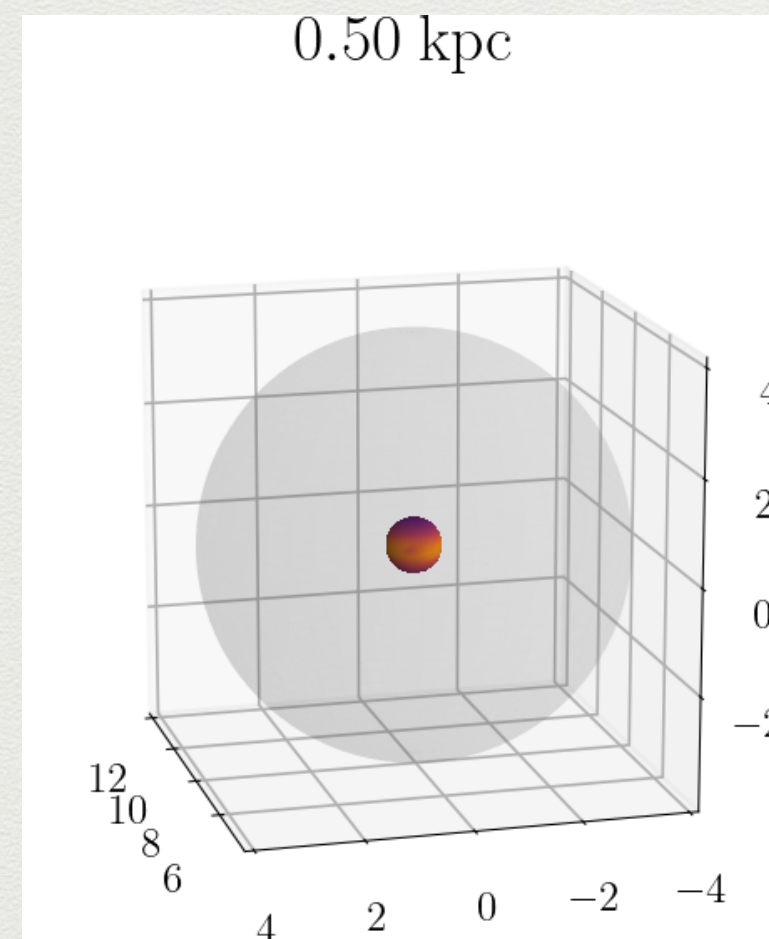
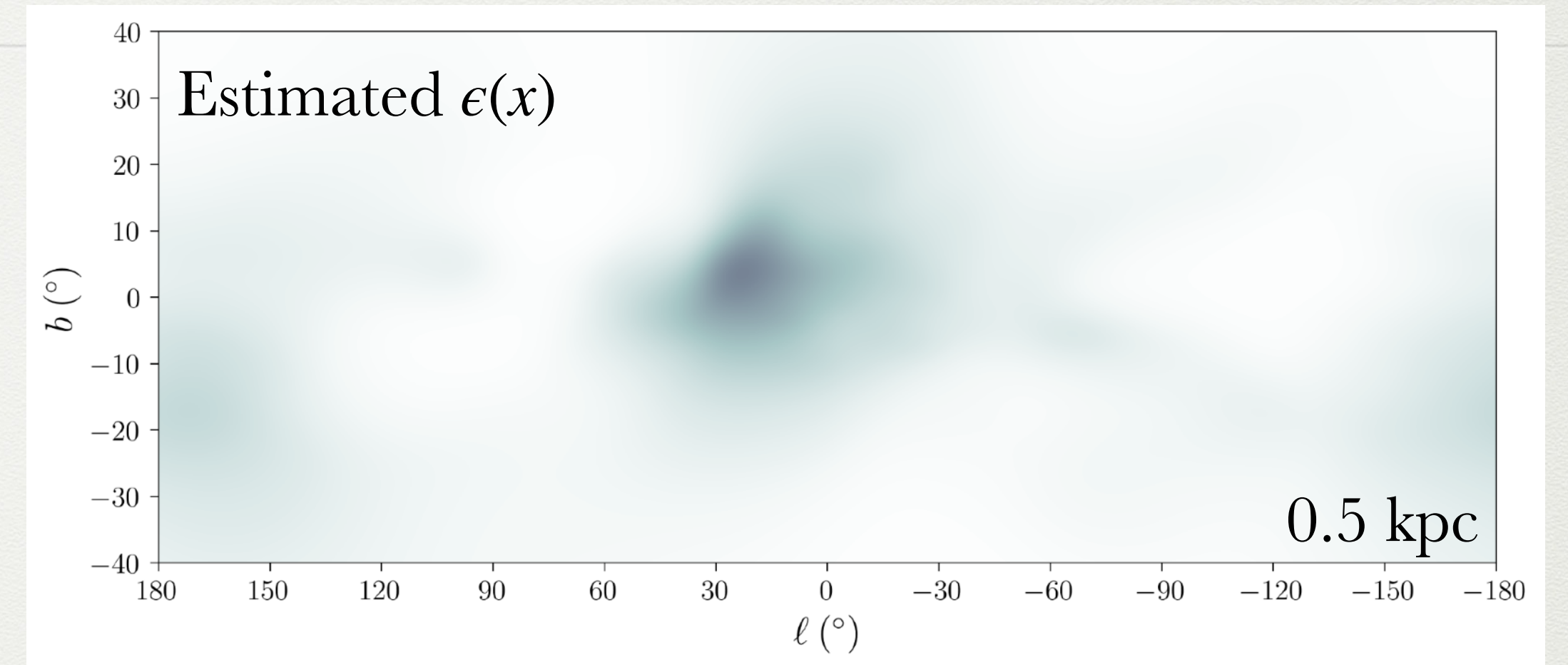
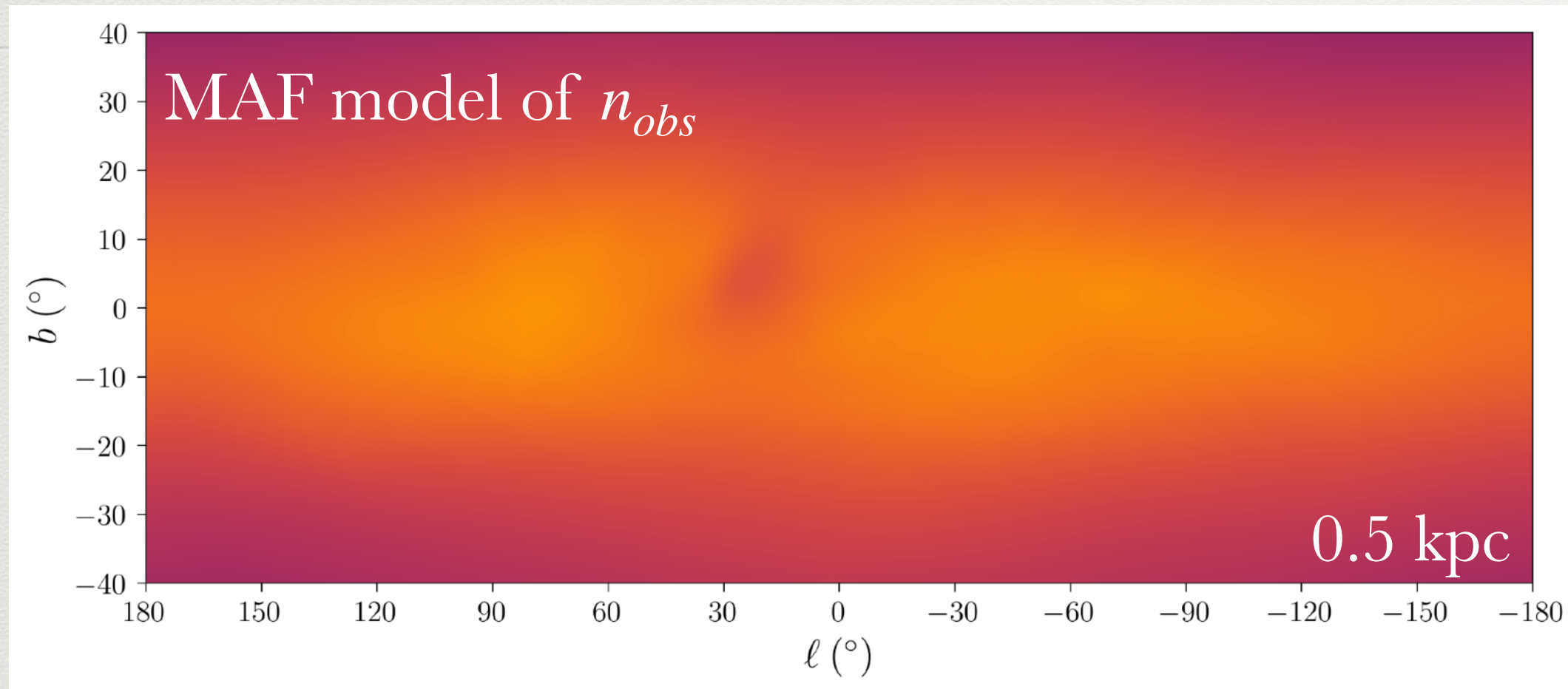
Results: Dust-Corrected PSD



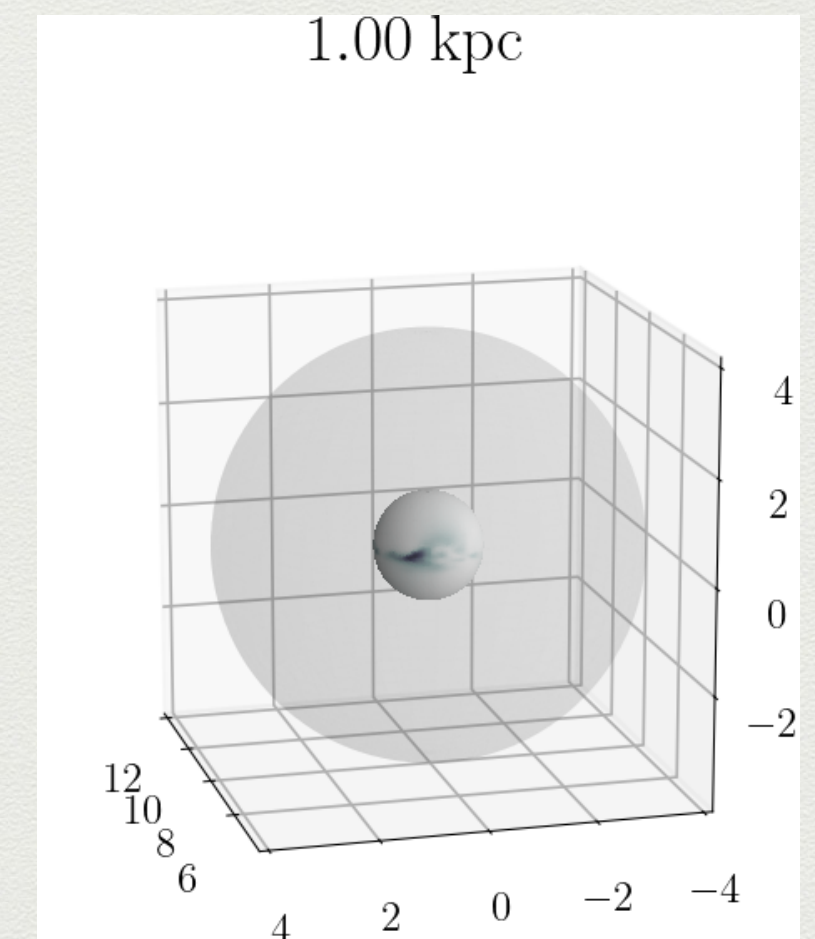
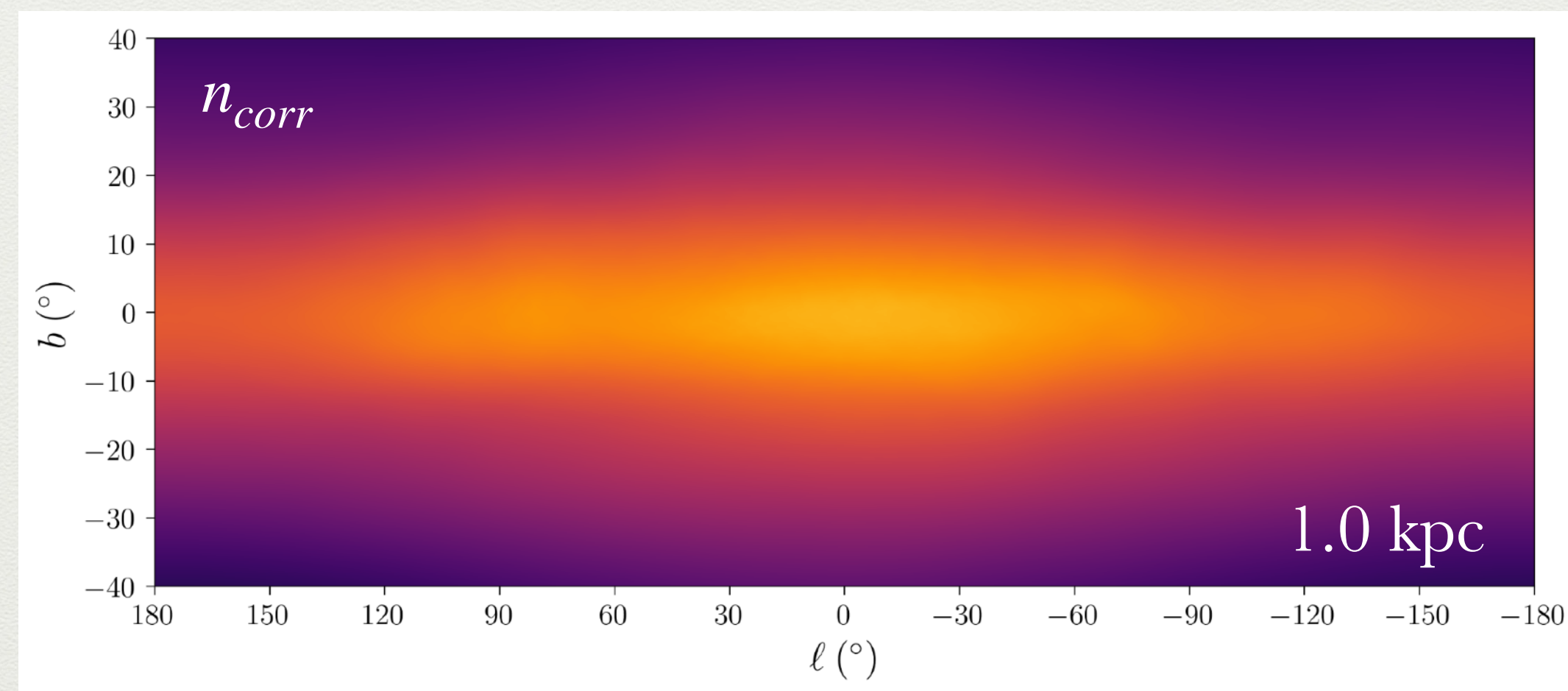
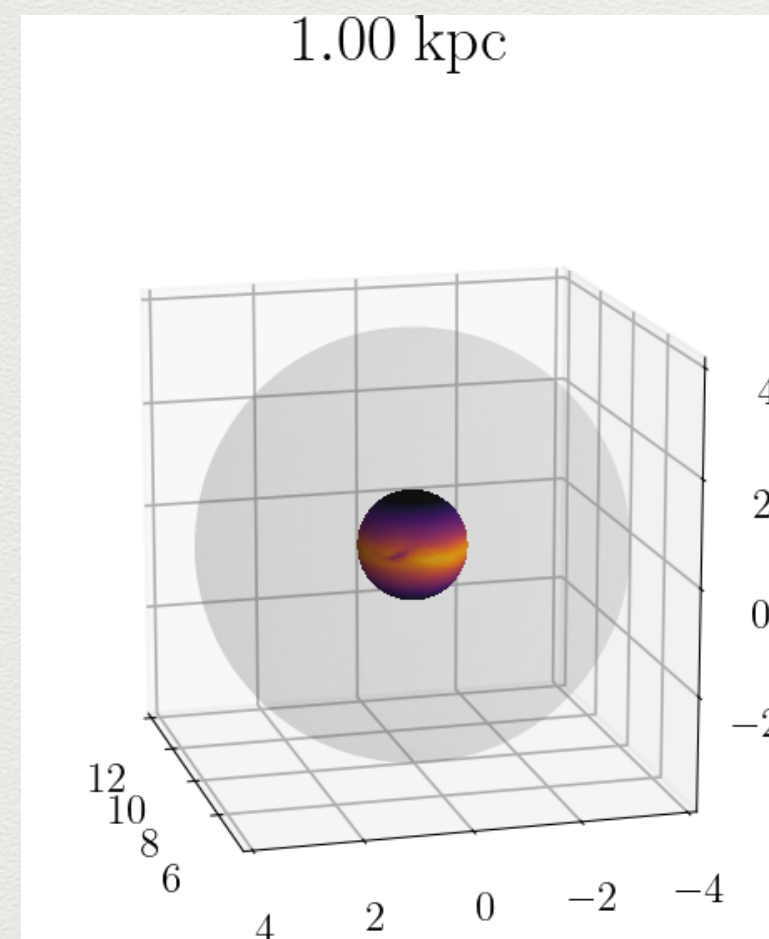
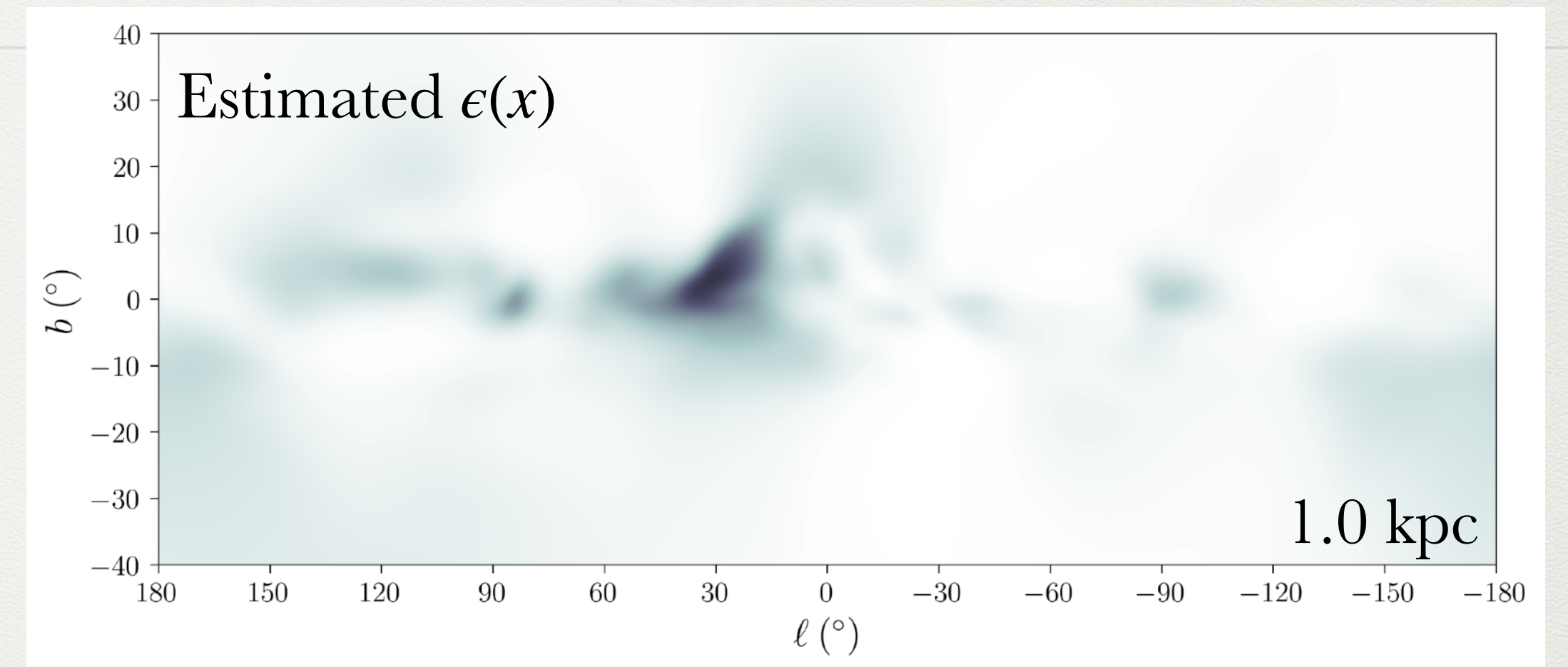
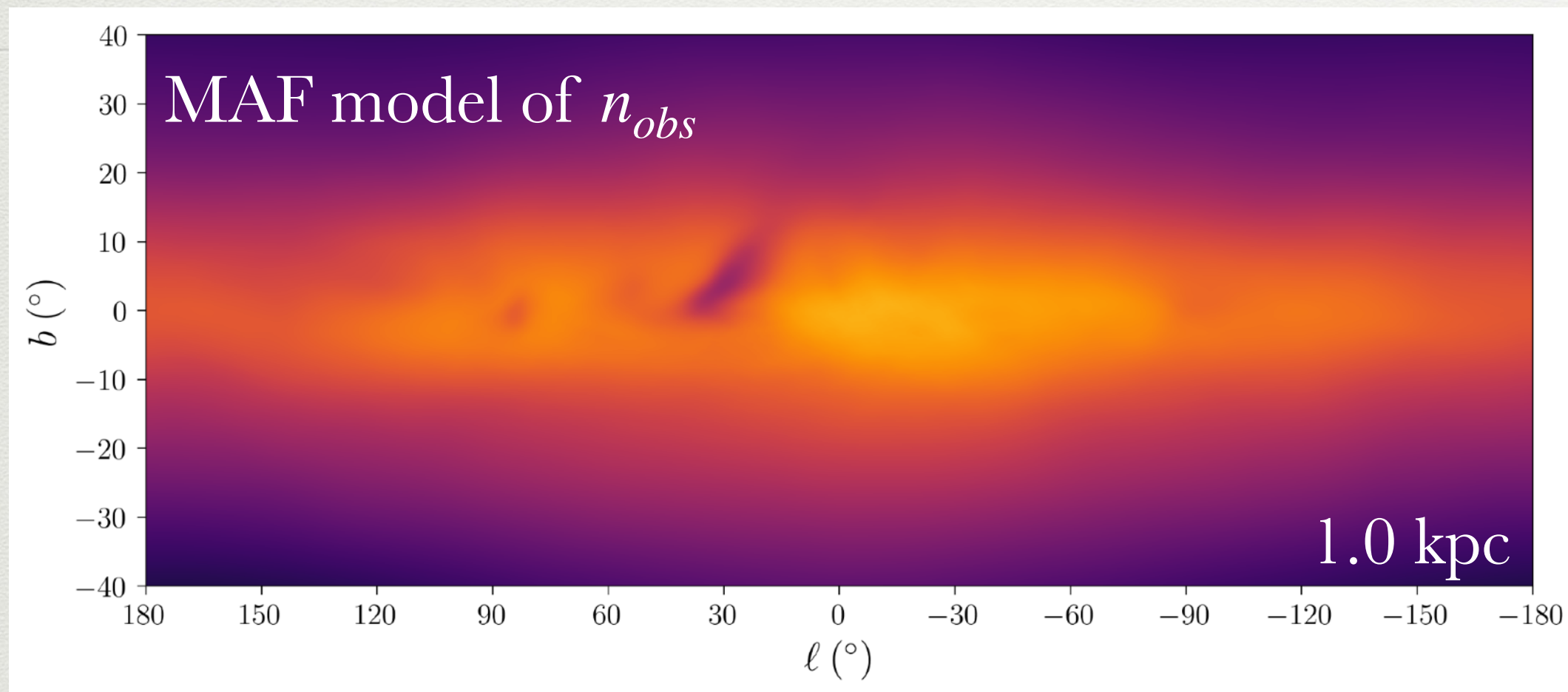
Corrected density
$$n_{corr}(x) = n_{obs}(x)/\epsilon(x)$$



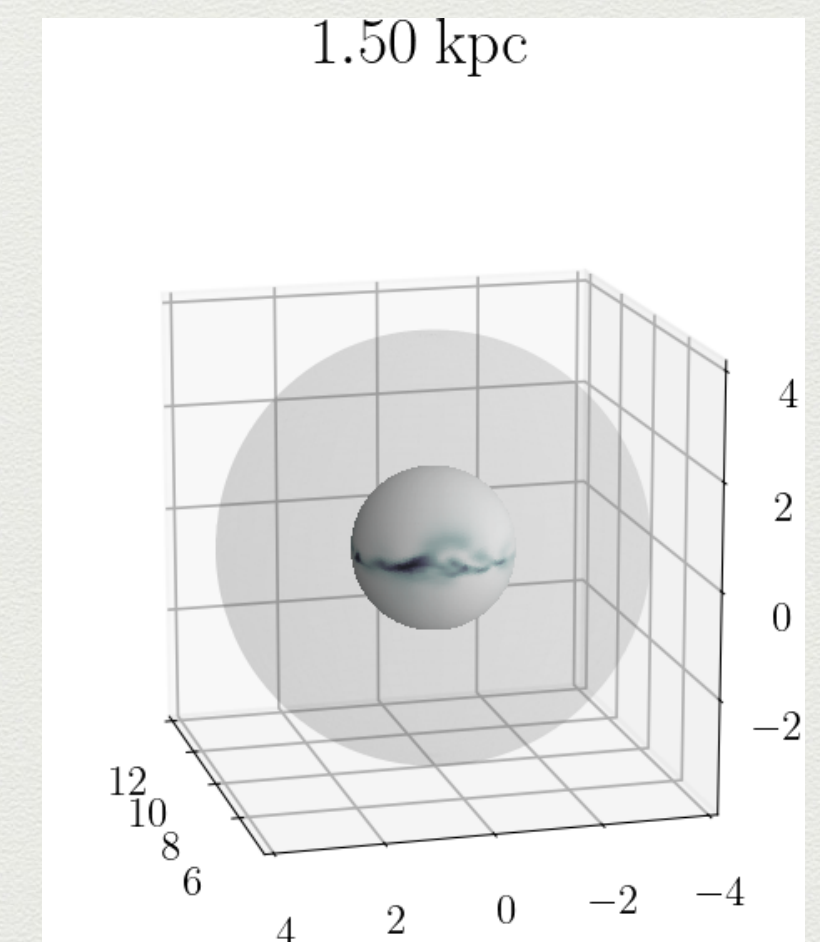
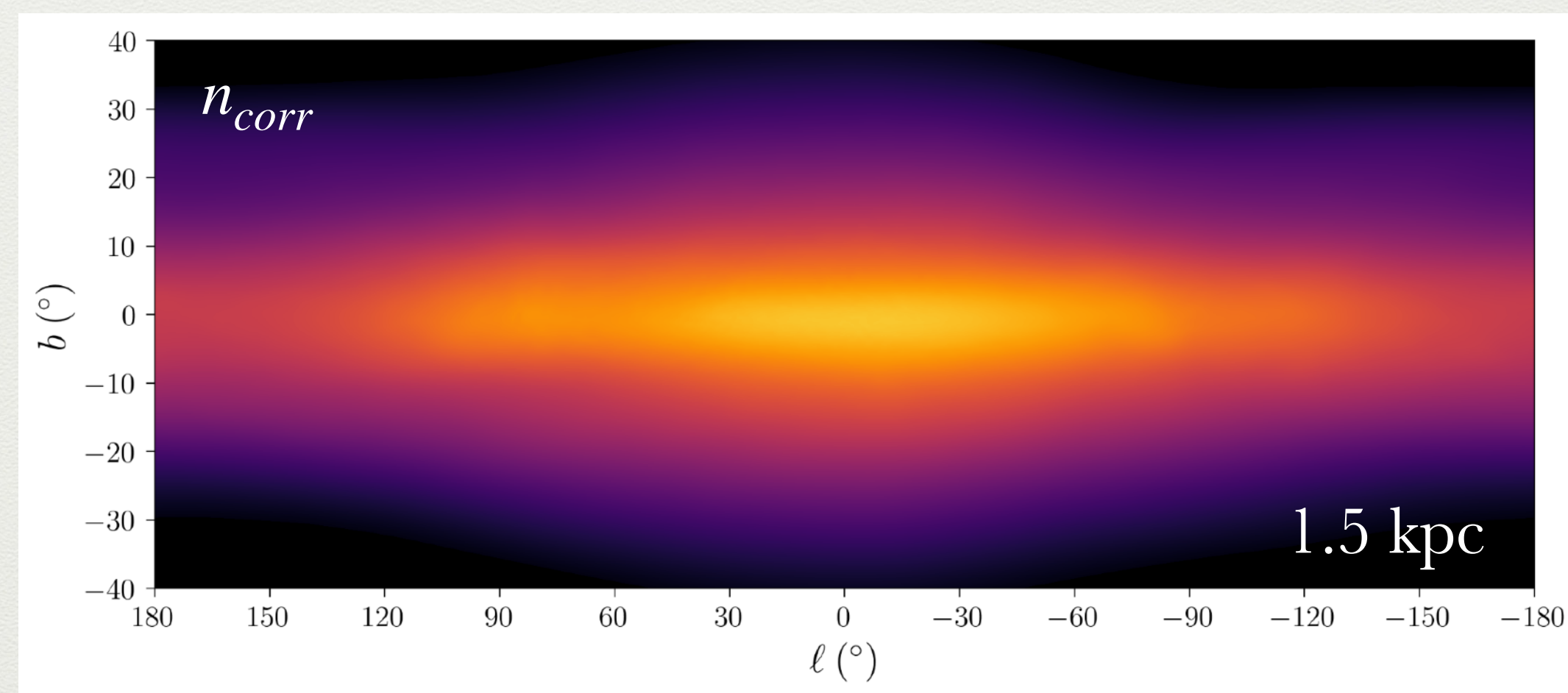
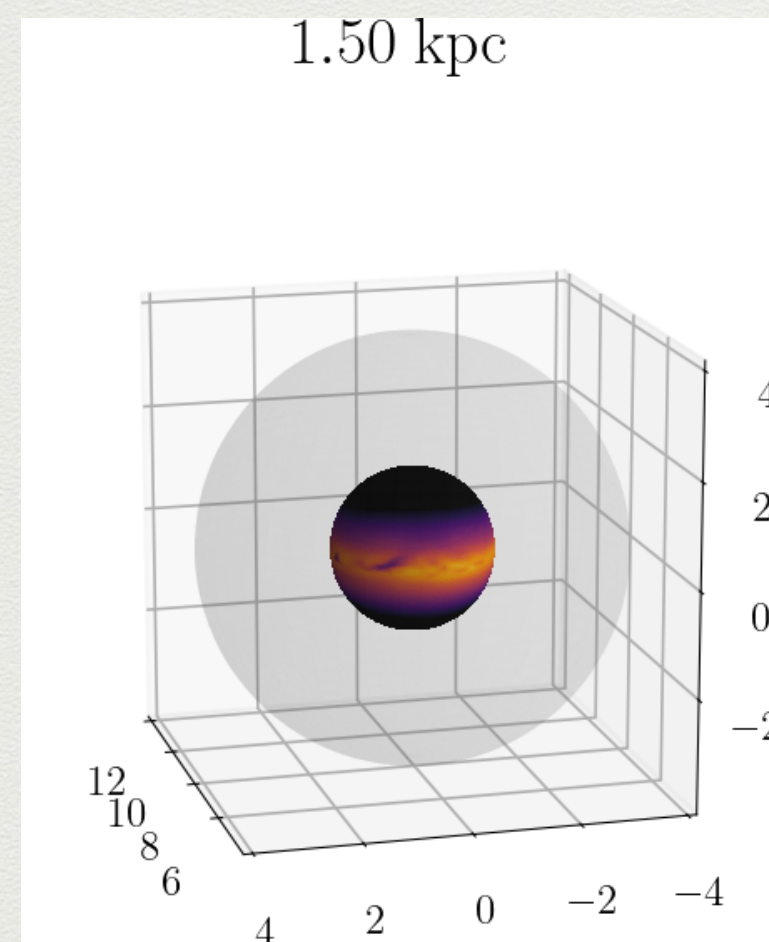
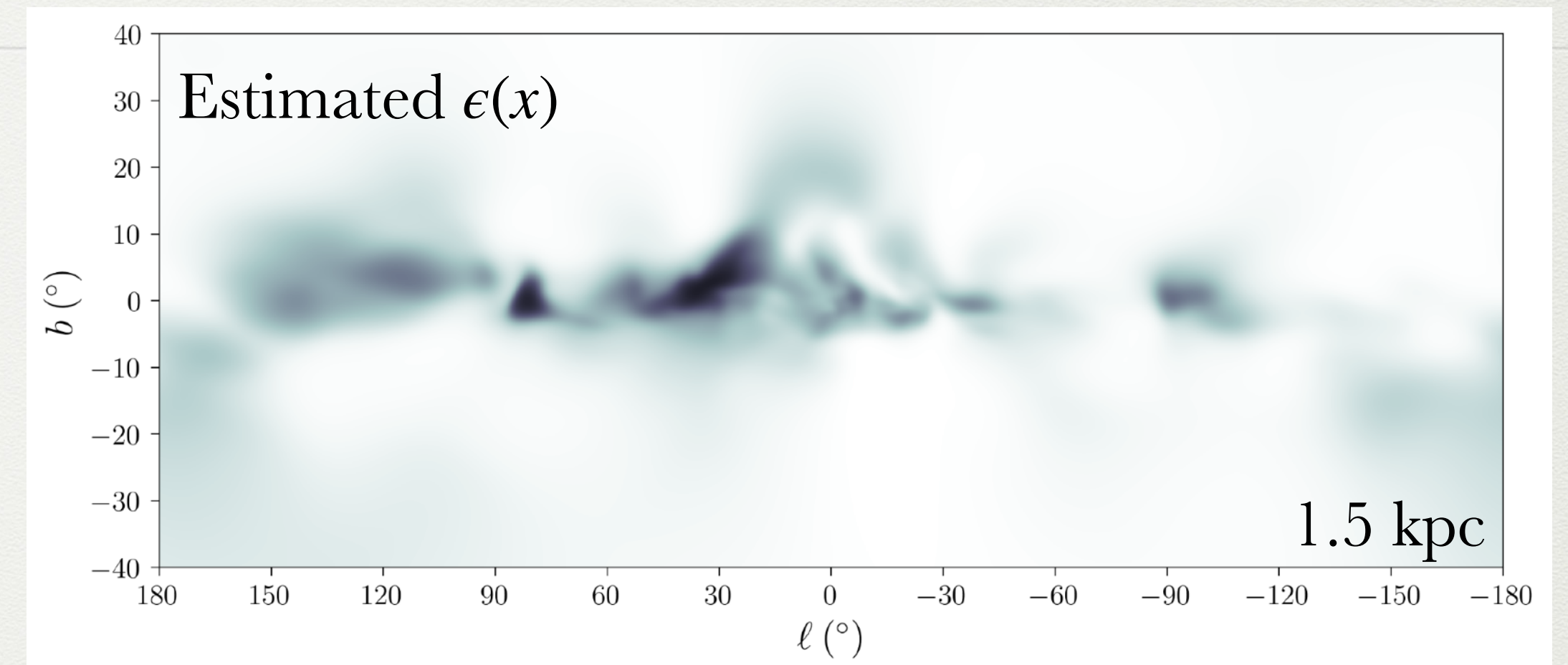
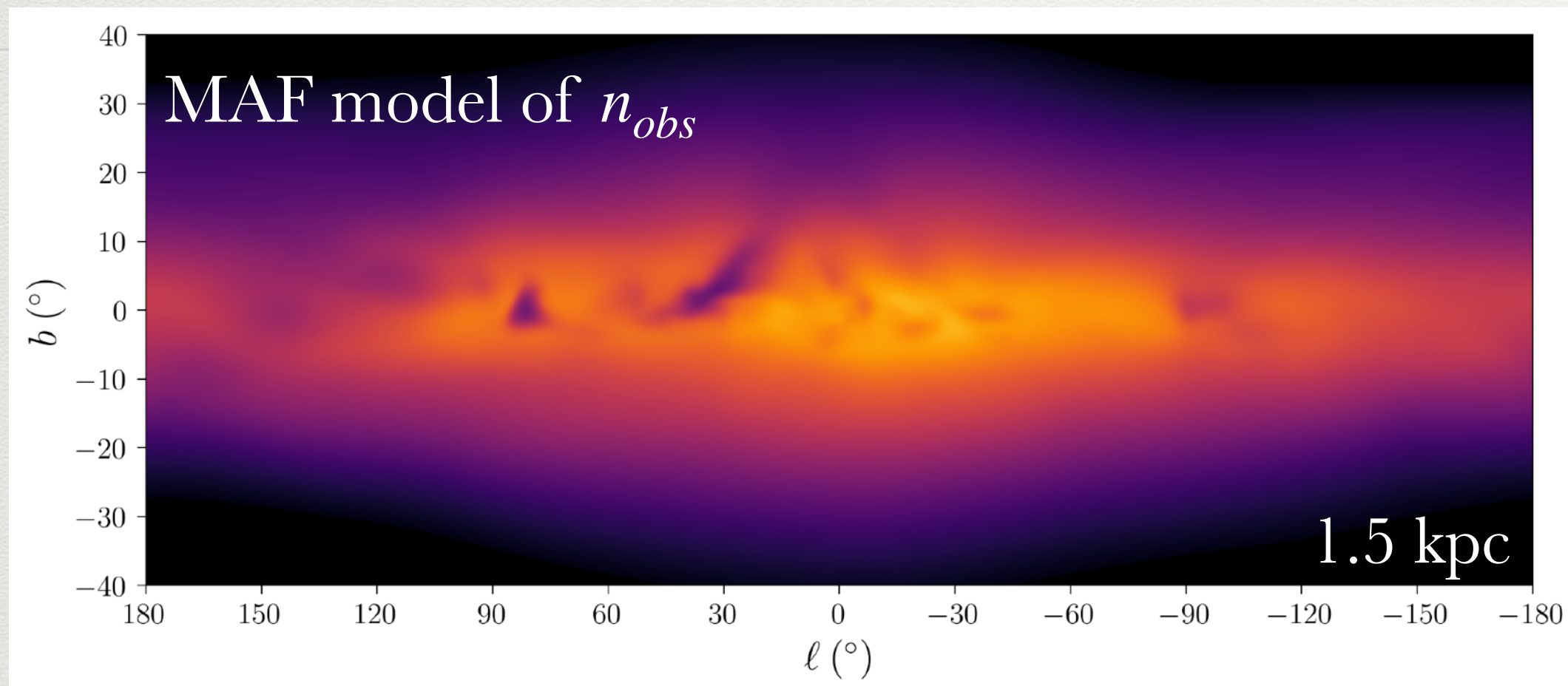
Results: Dust-Corrected PSD



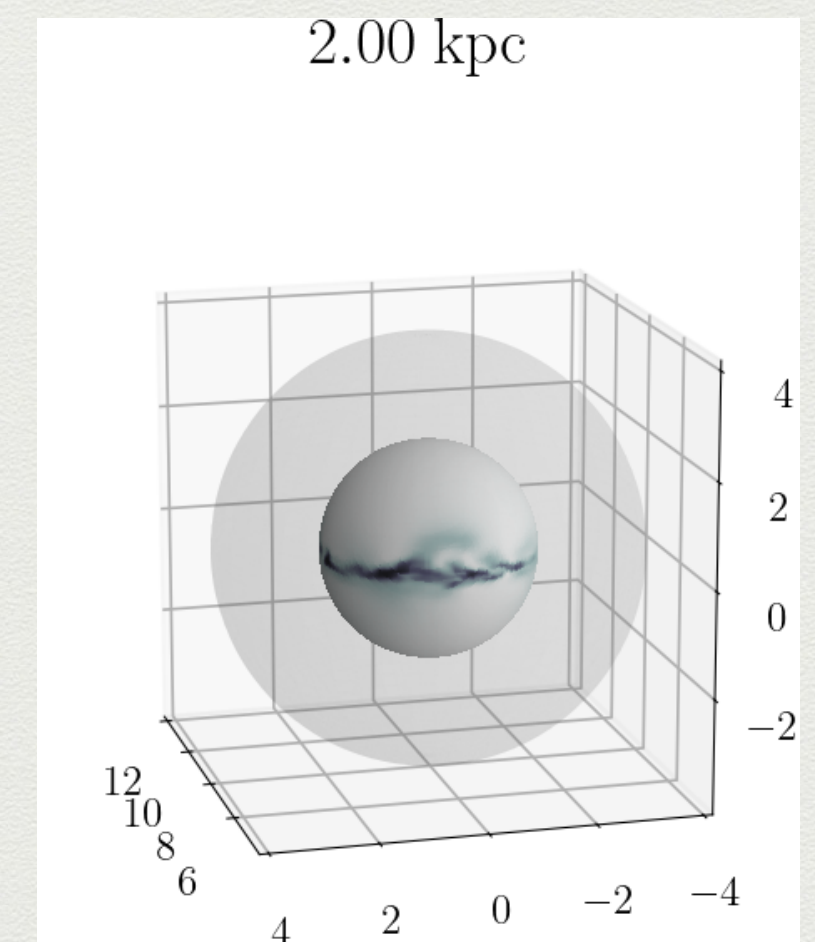
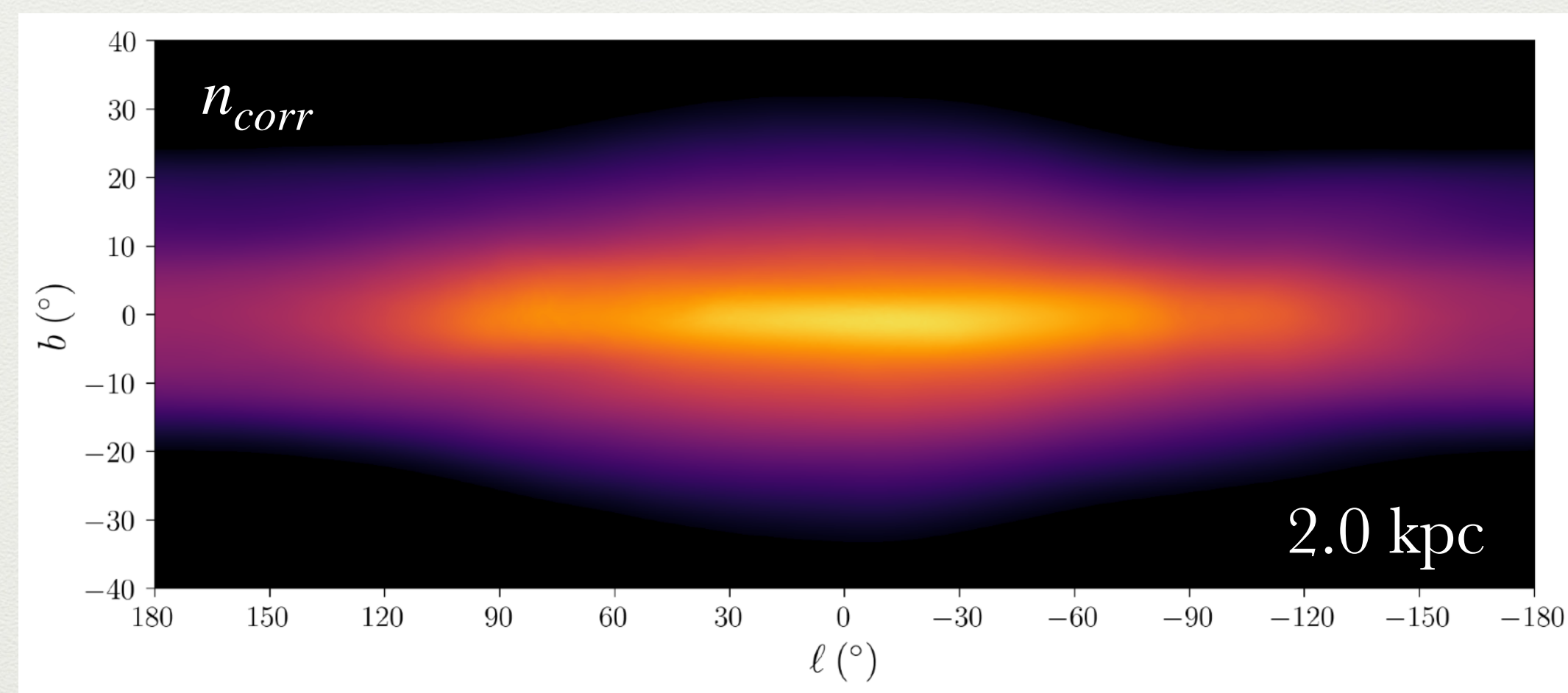
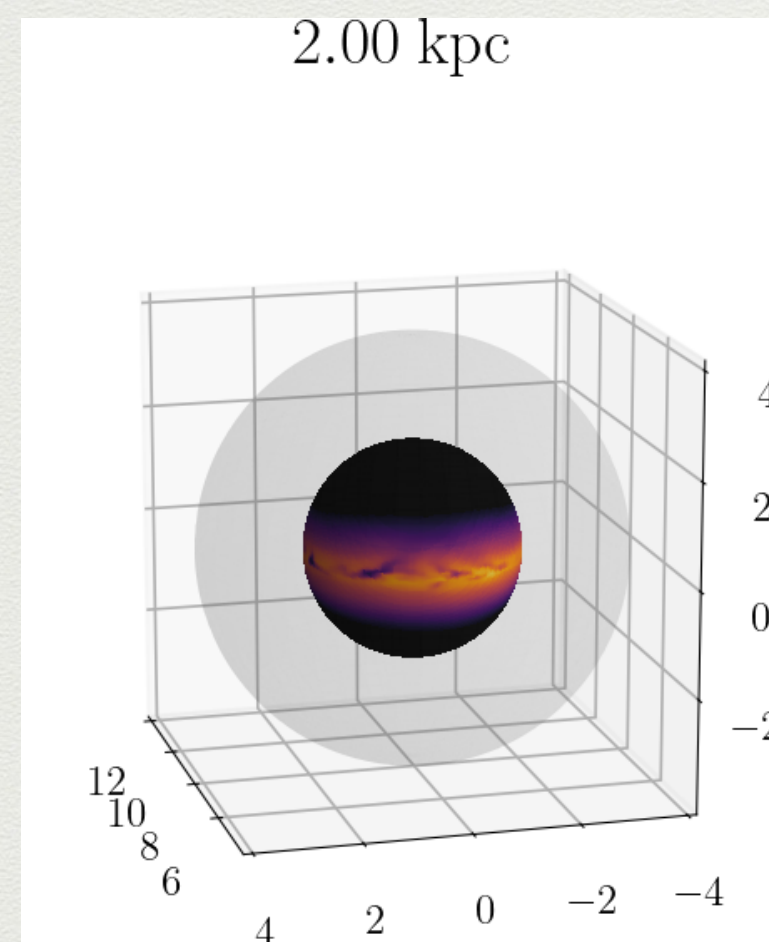
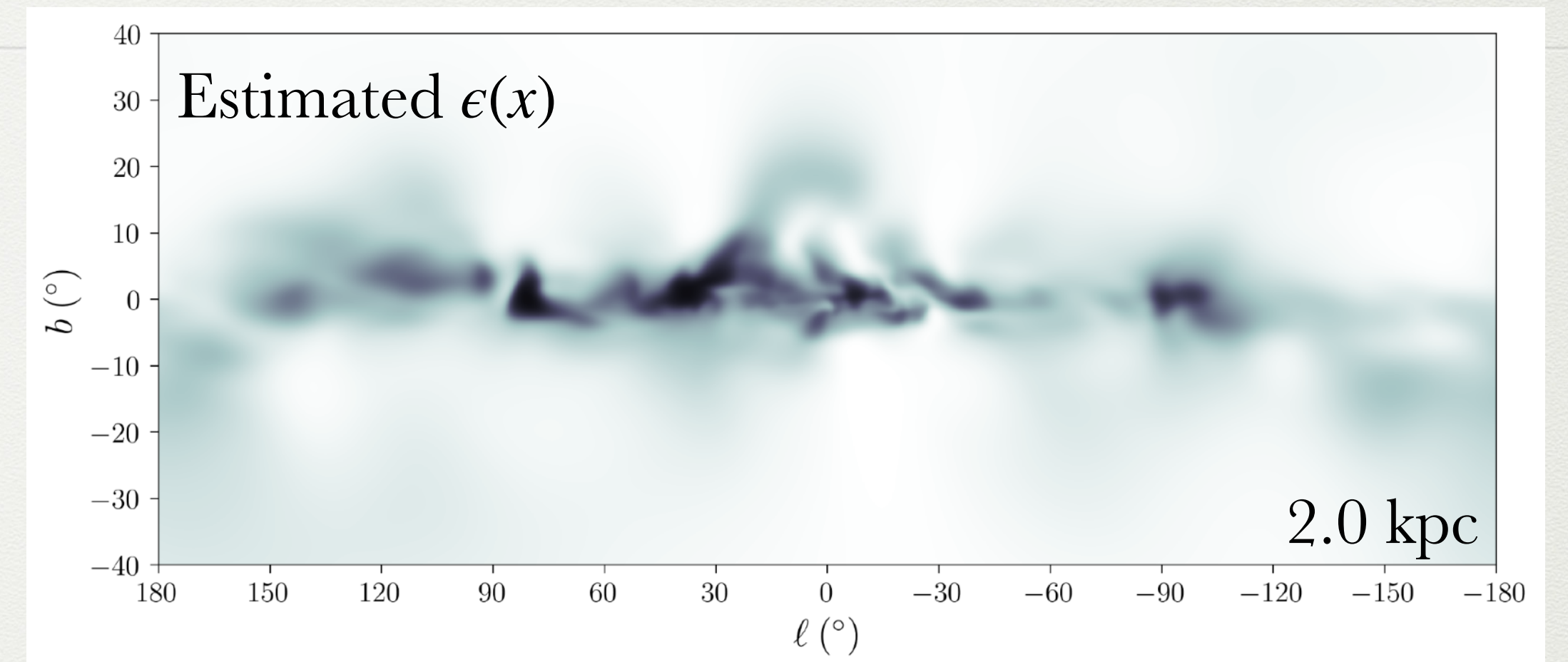
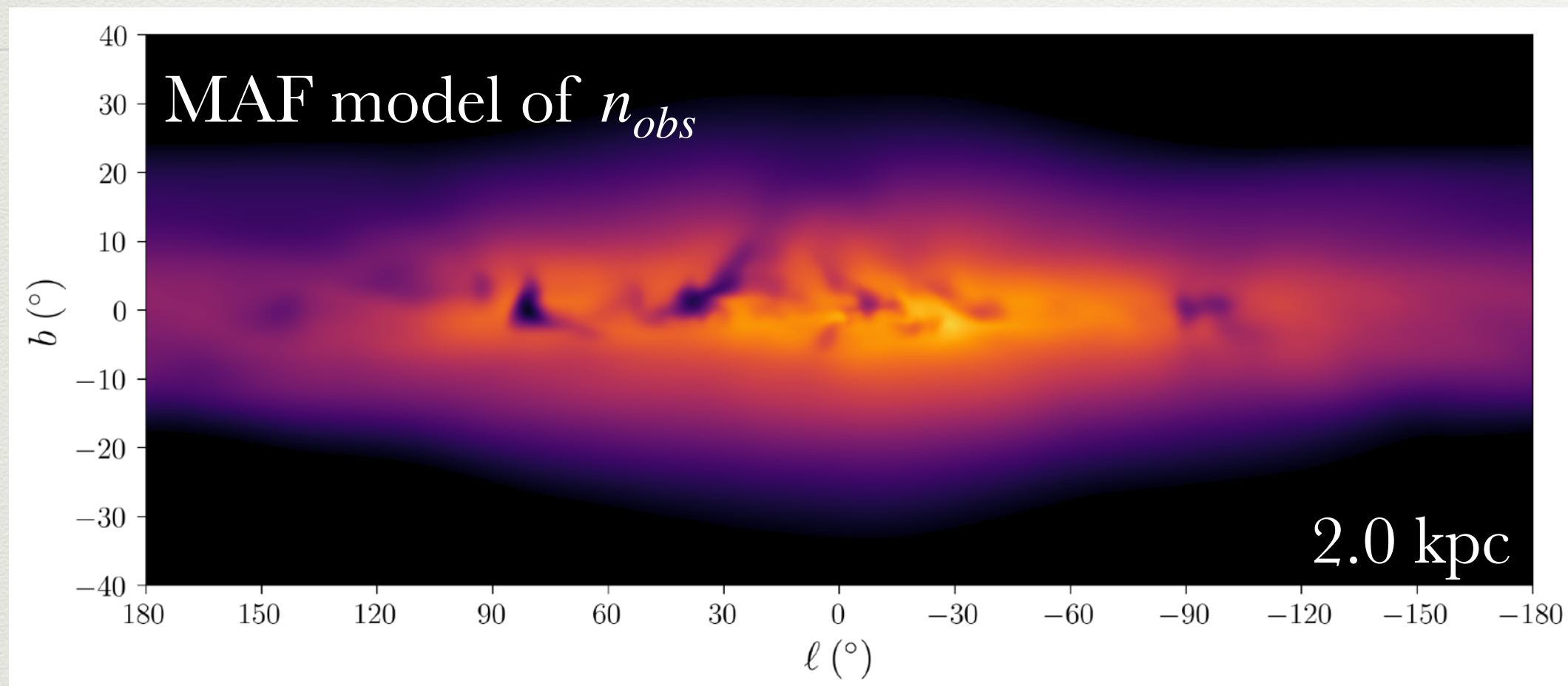
Results: Dust-Corrected PSD



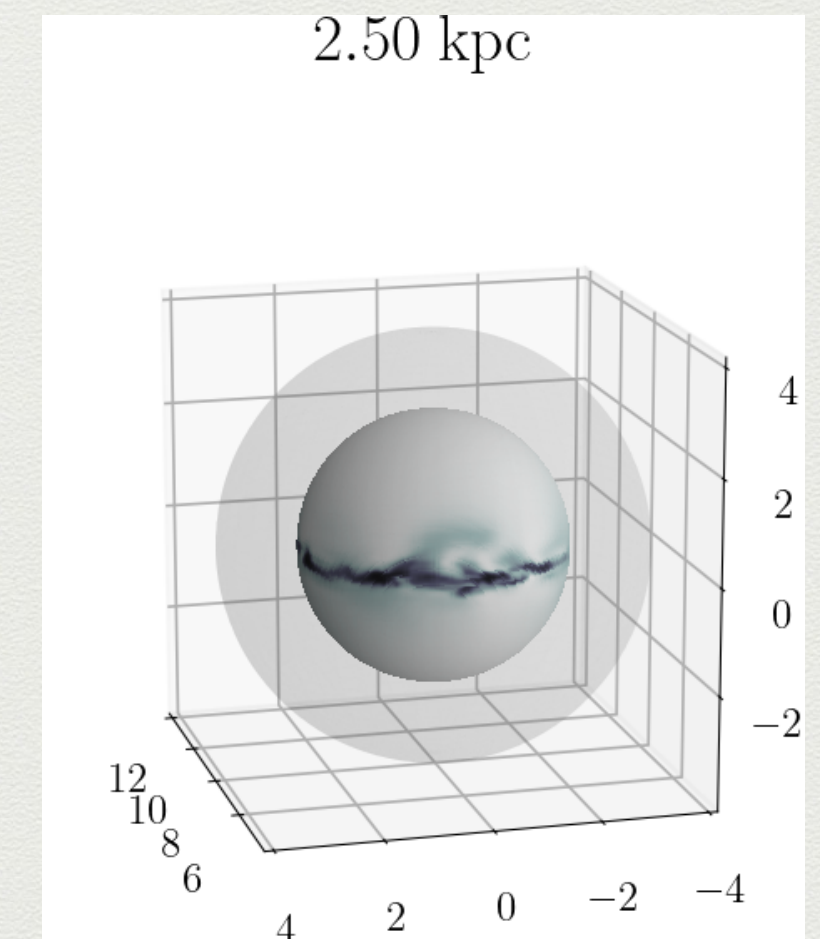
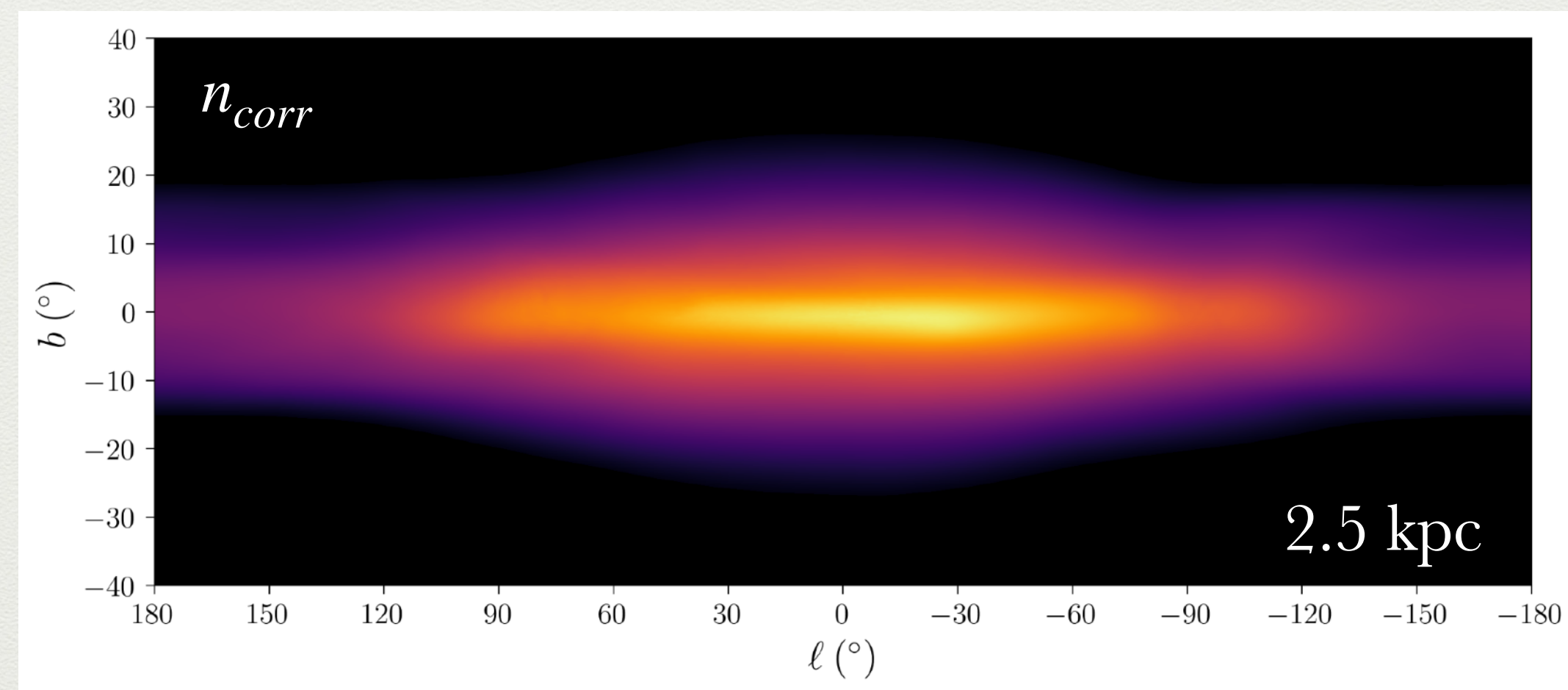
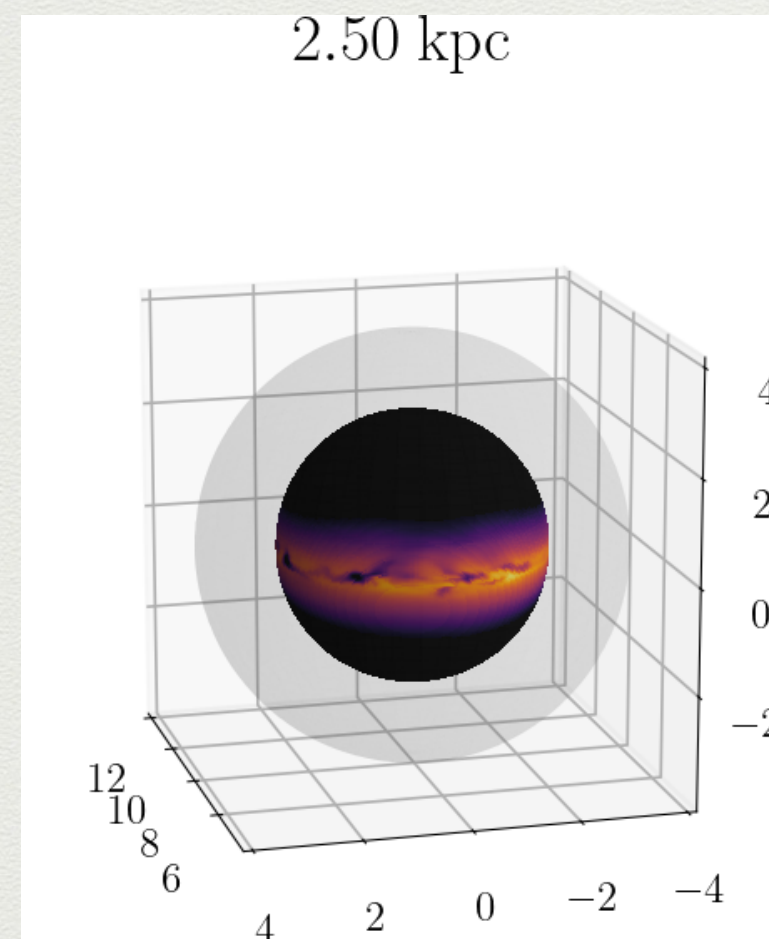
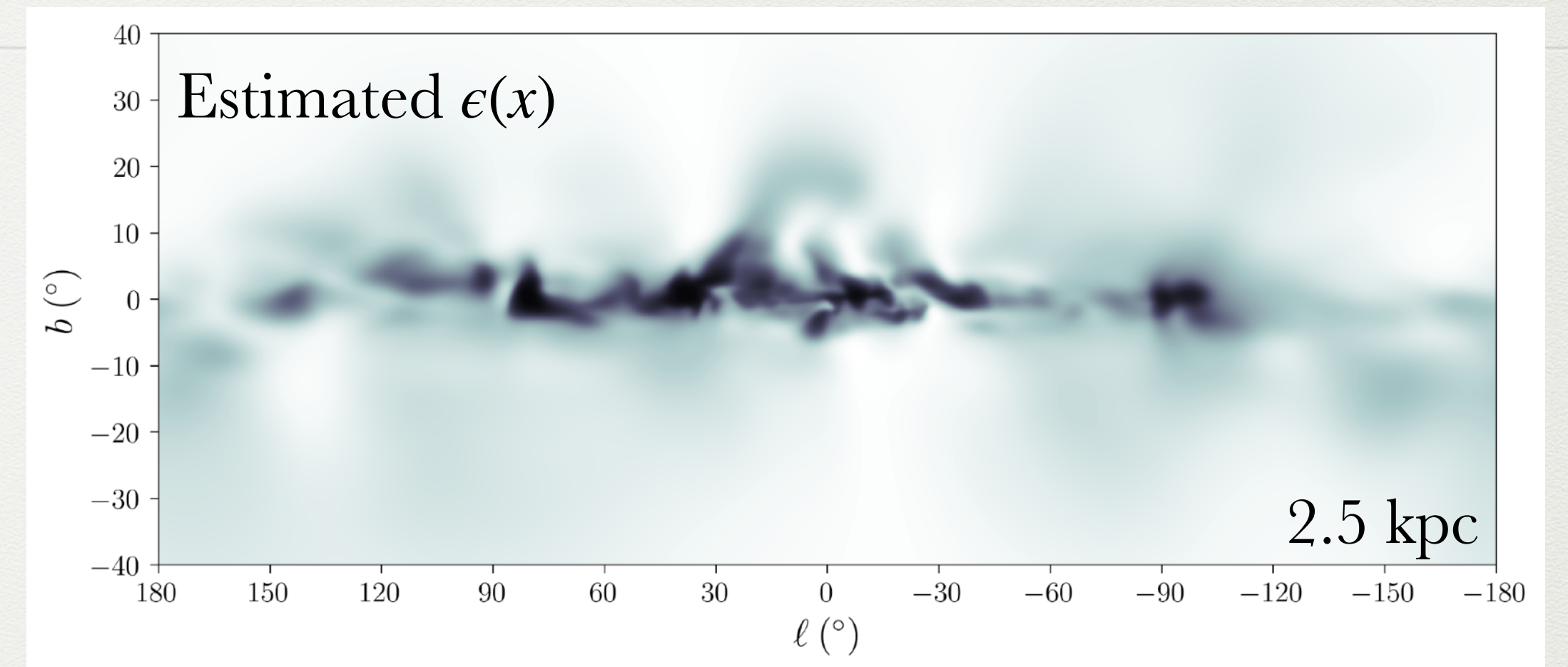
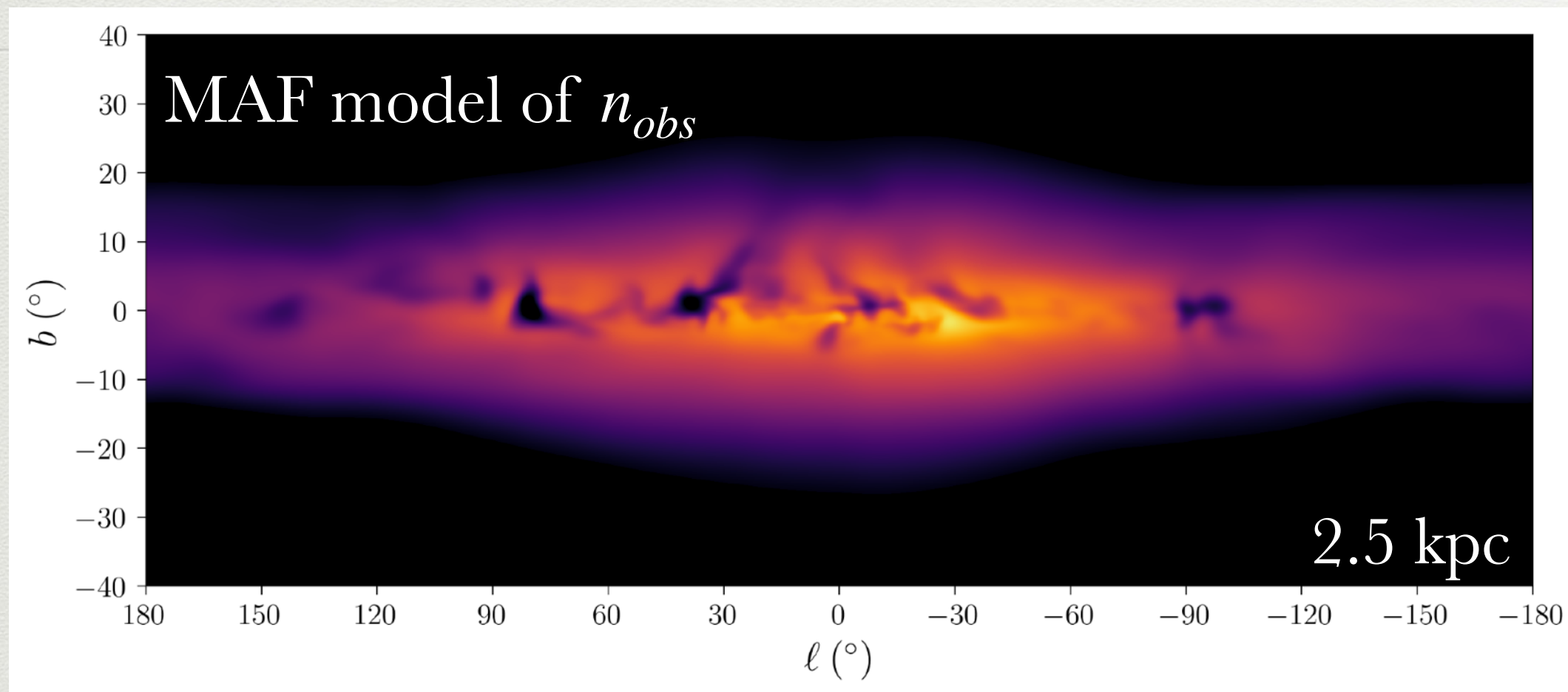
Results: Dust-Corrected PSD



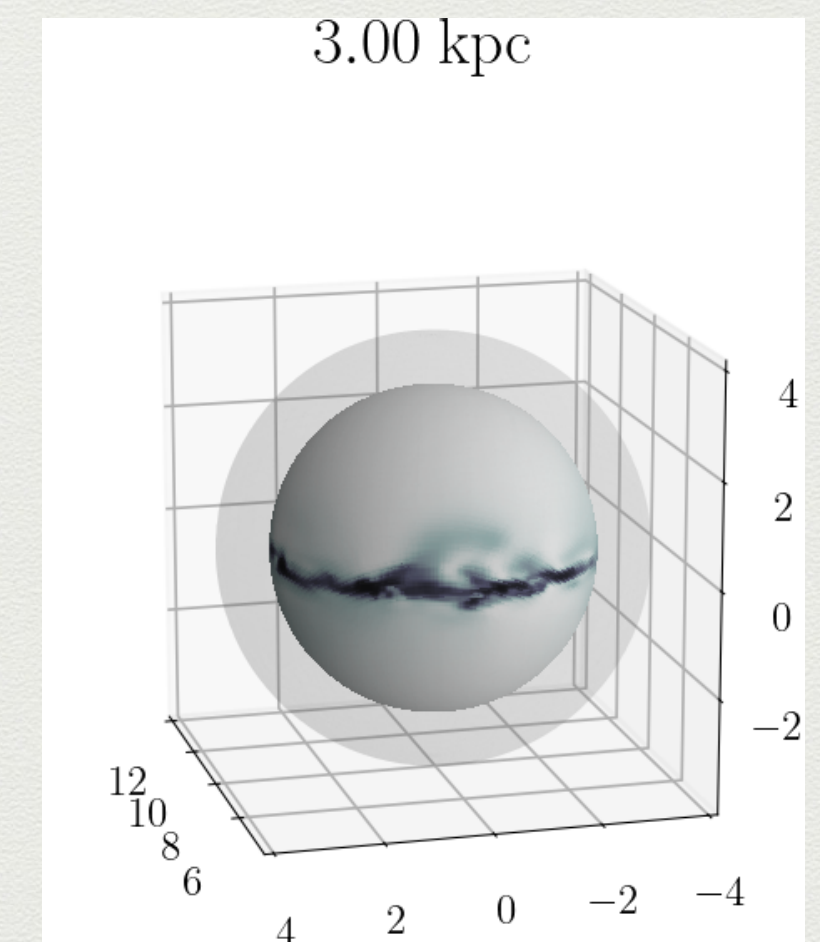
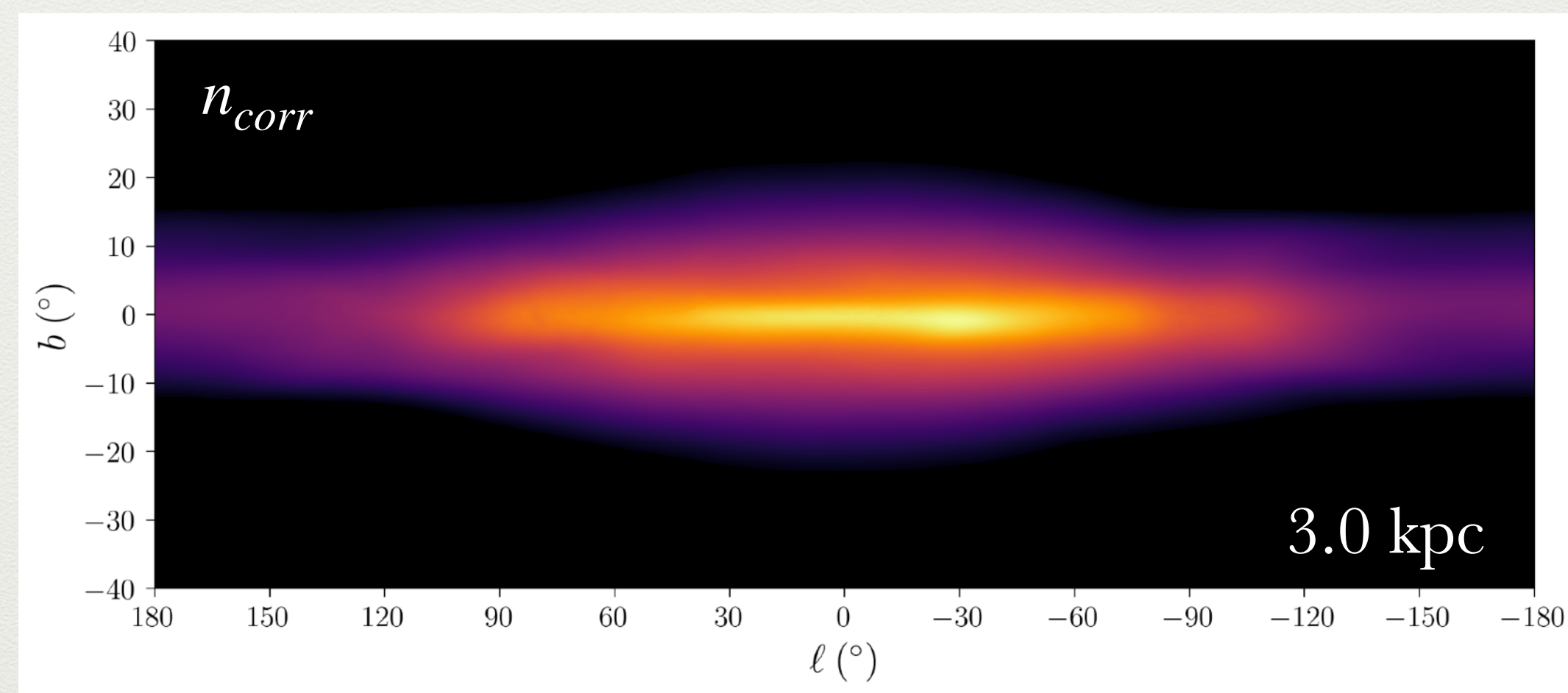
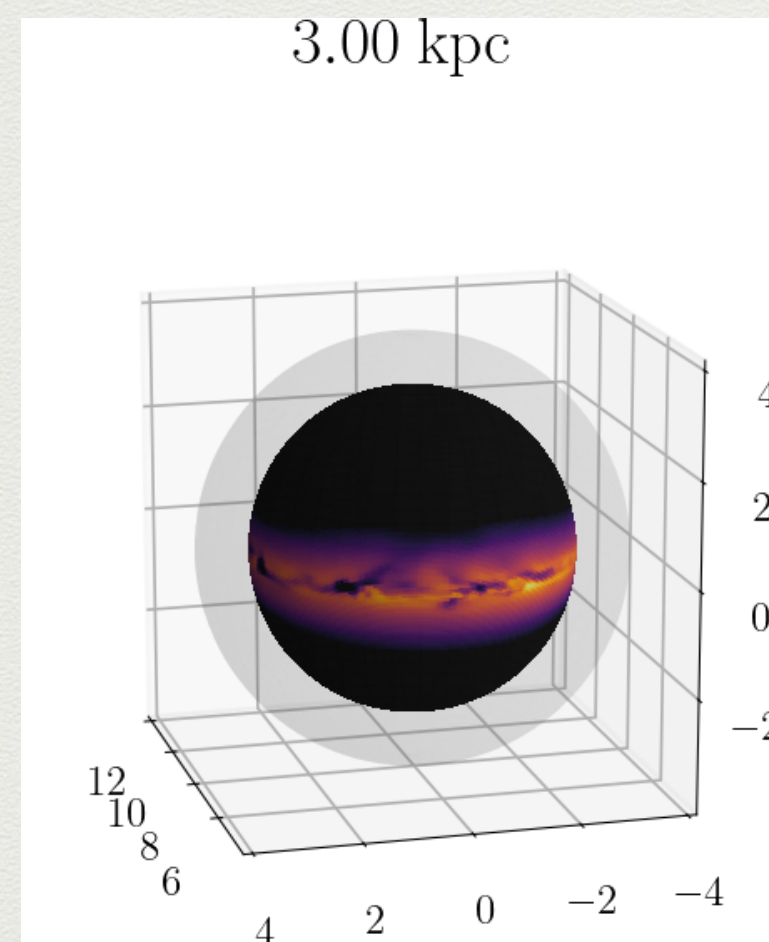
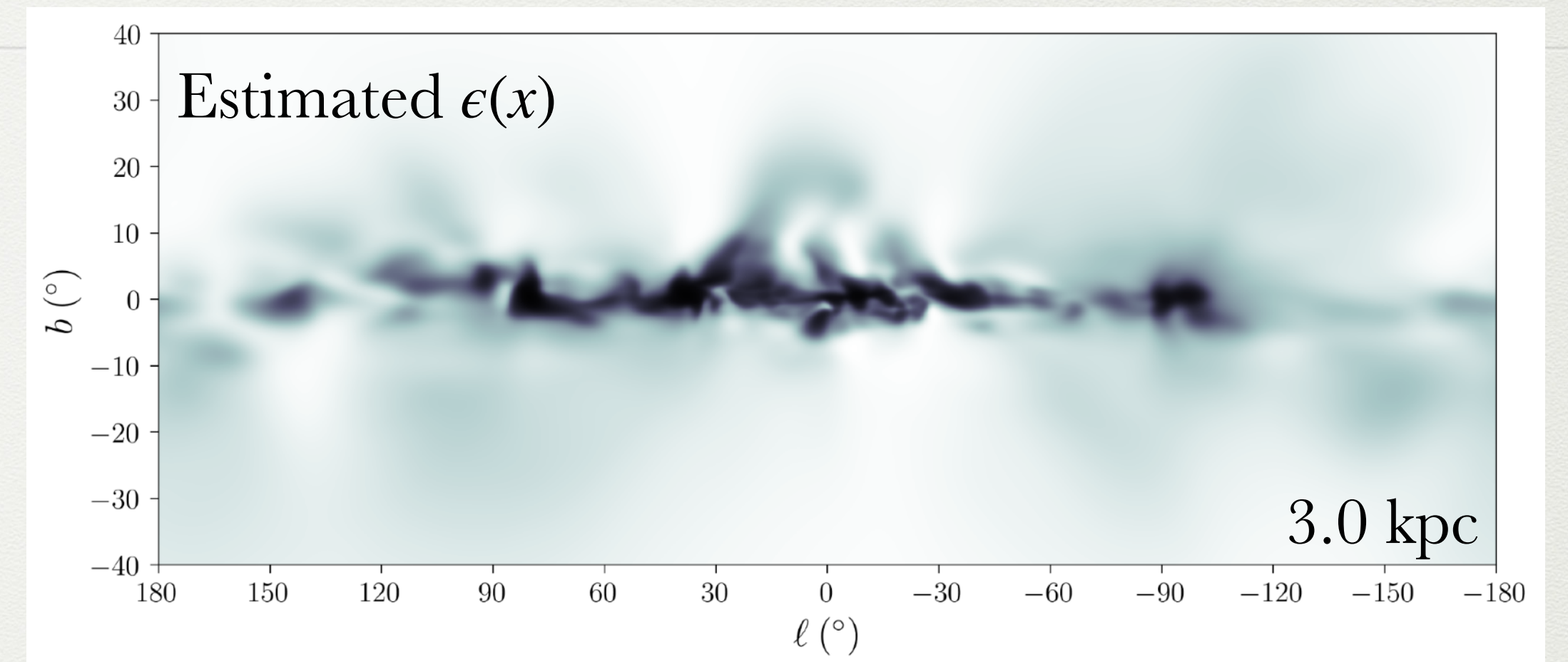
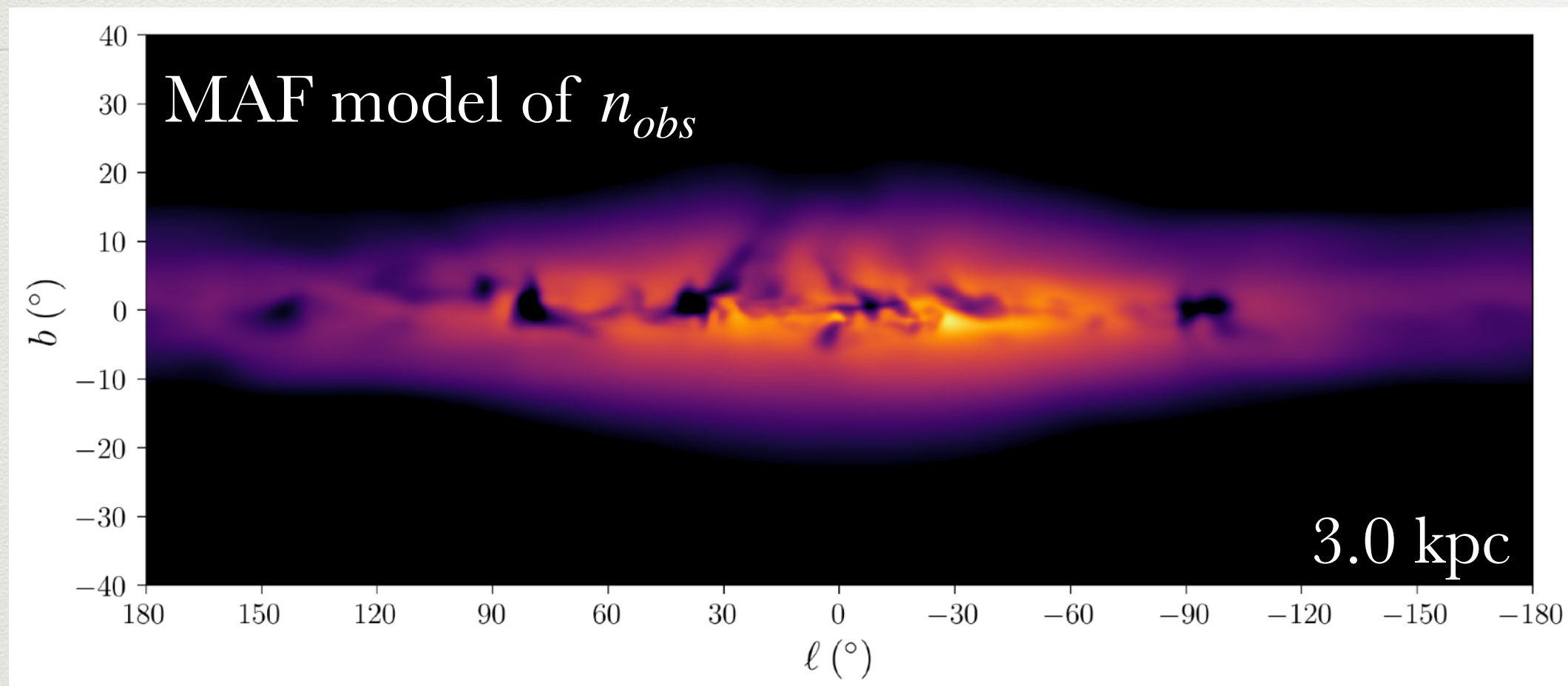
Results: Dust-Corrected PSD



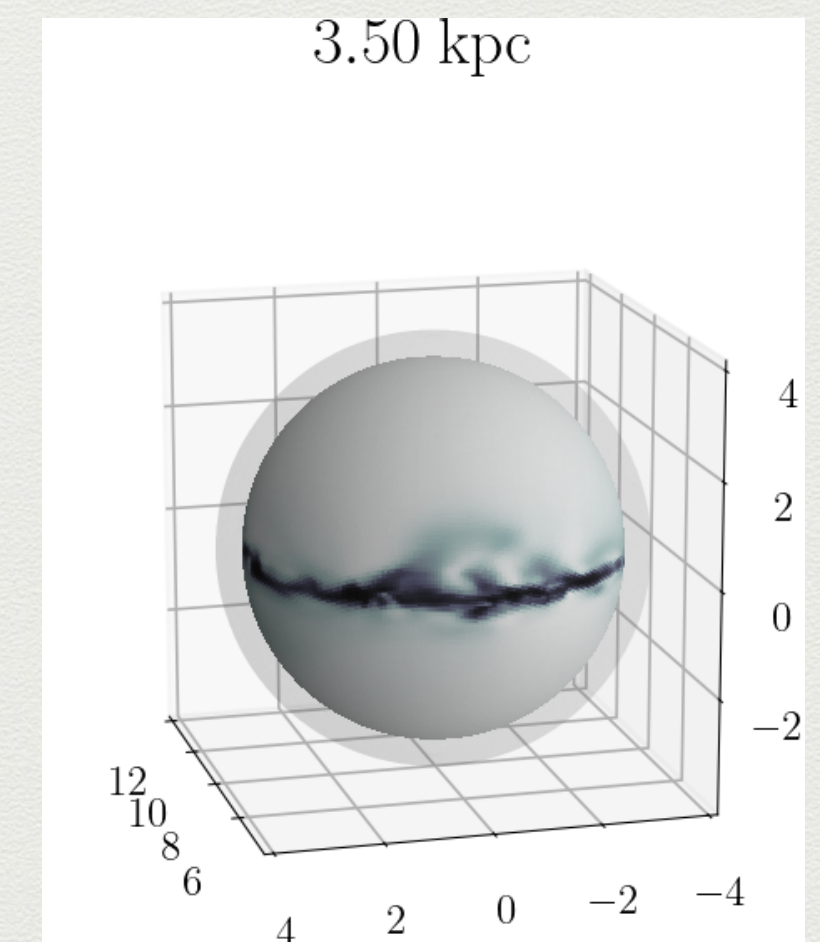
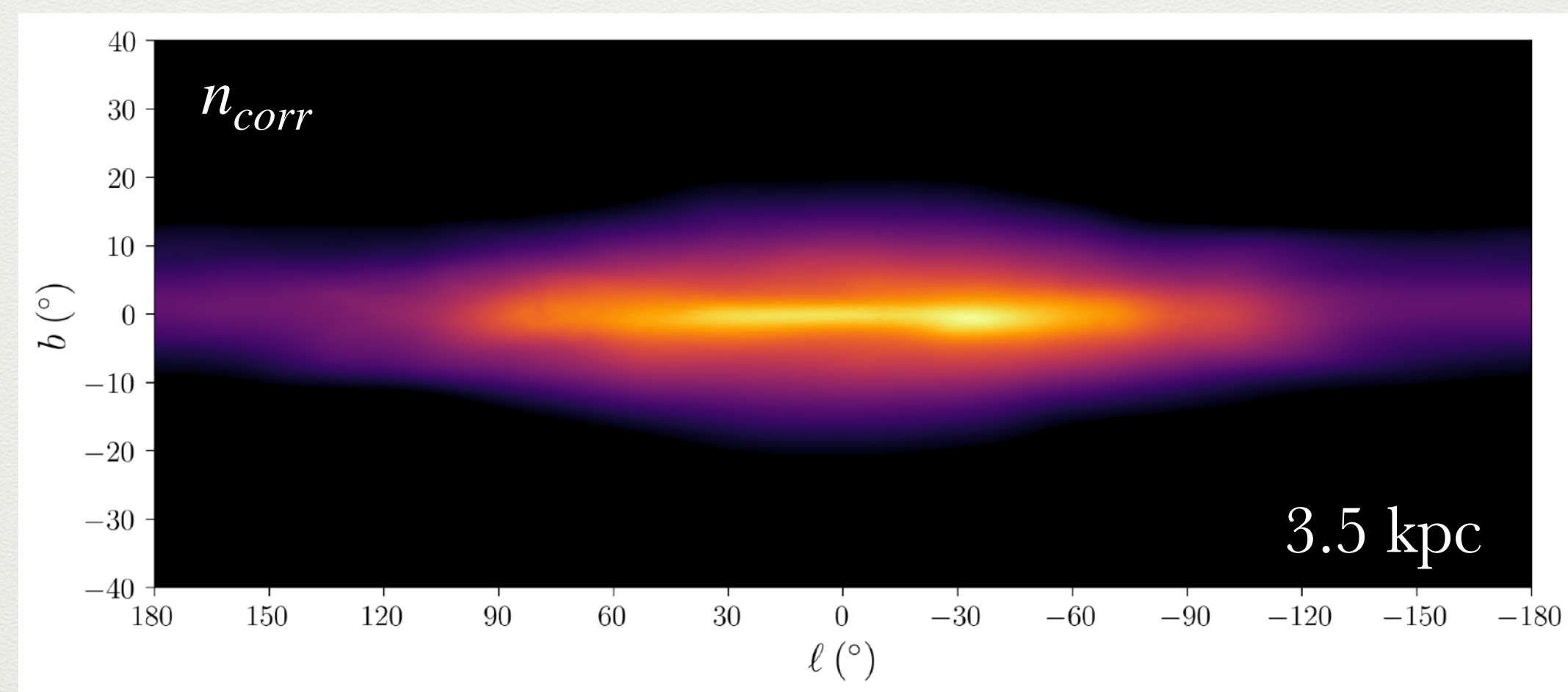
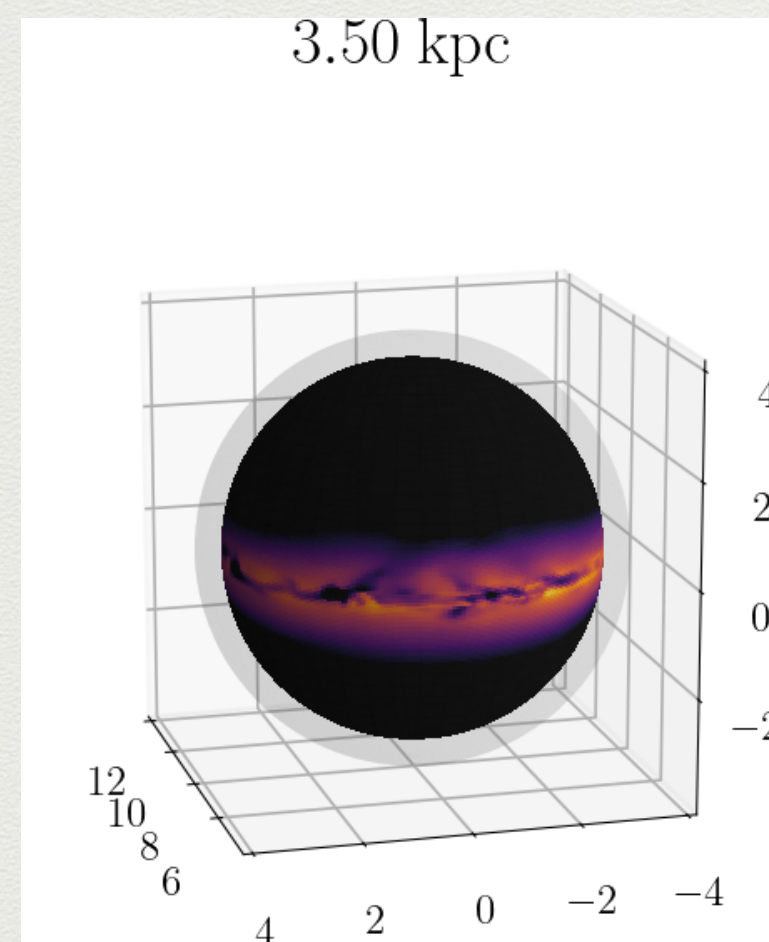
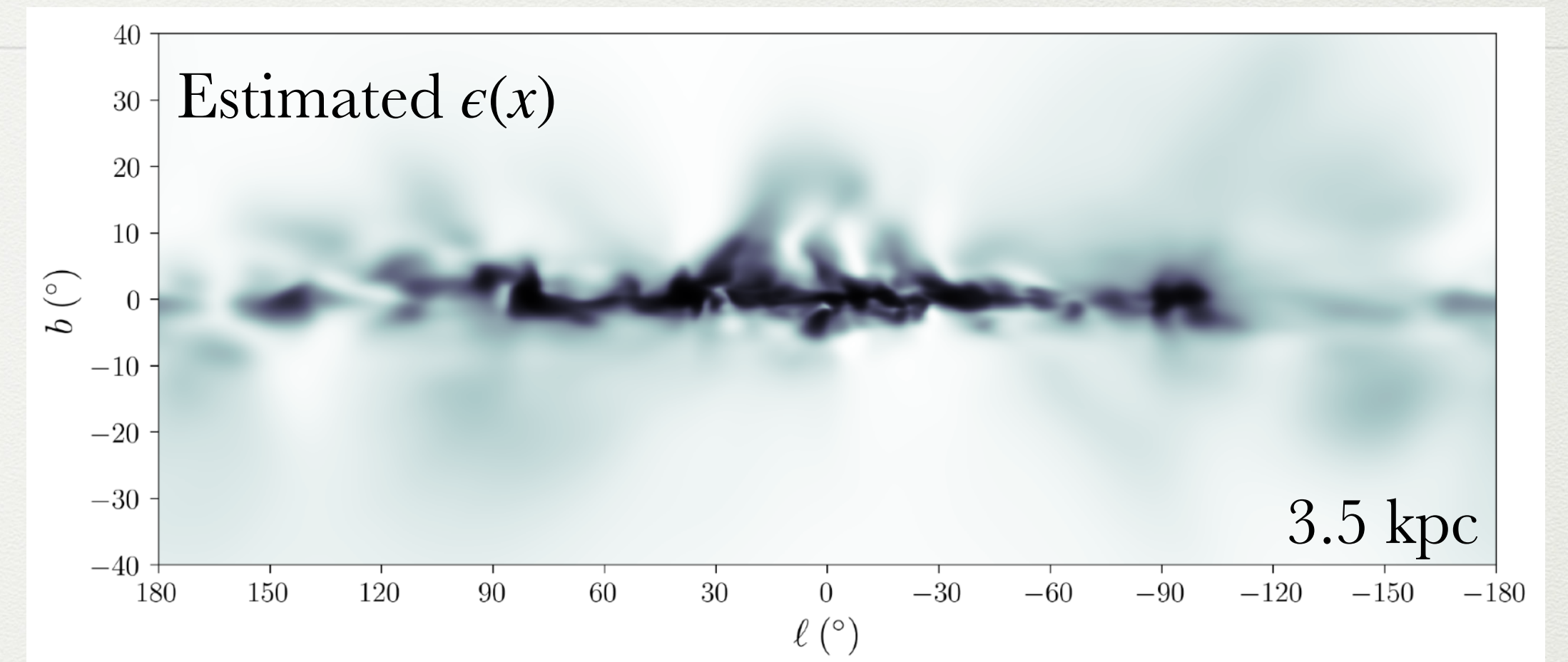
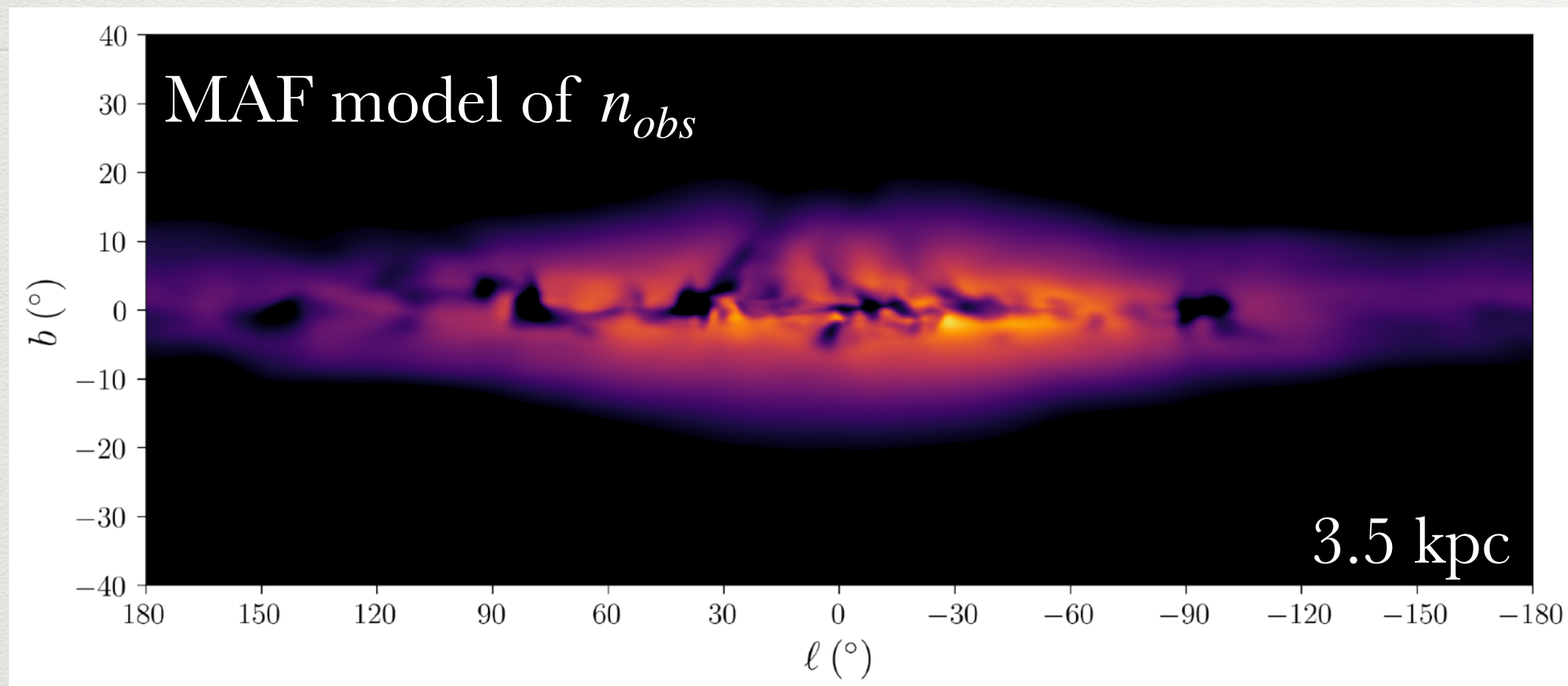
Results: Dust-Corrected PSD



Results: Dust-Corrected PSD

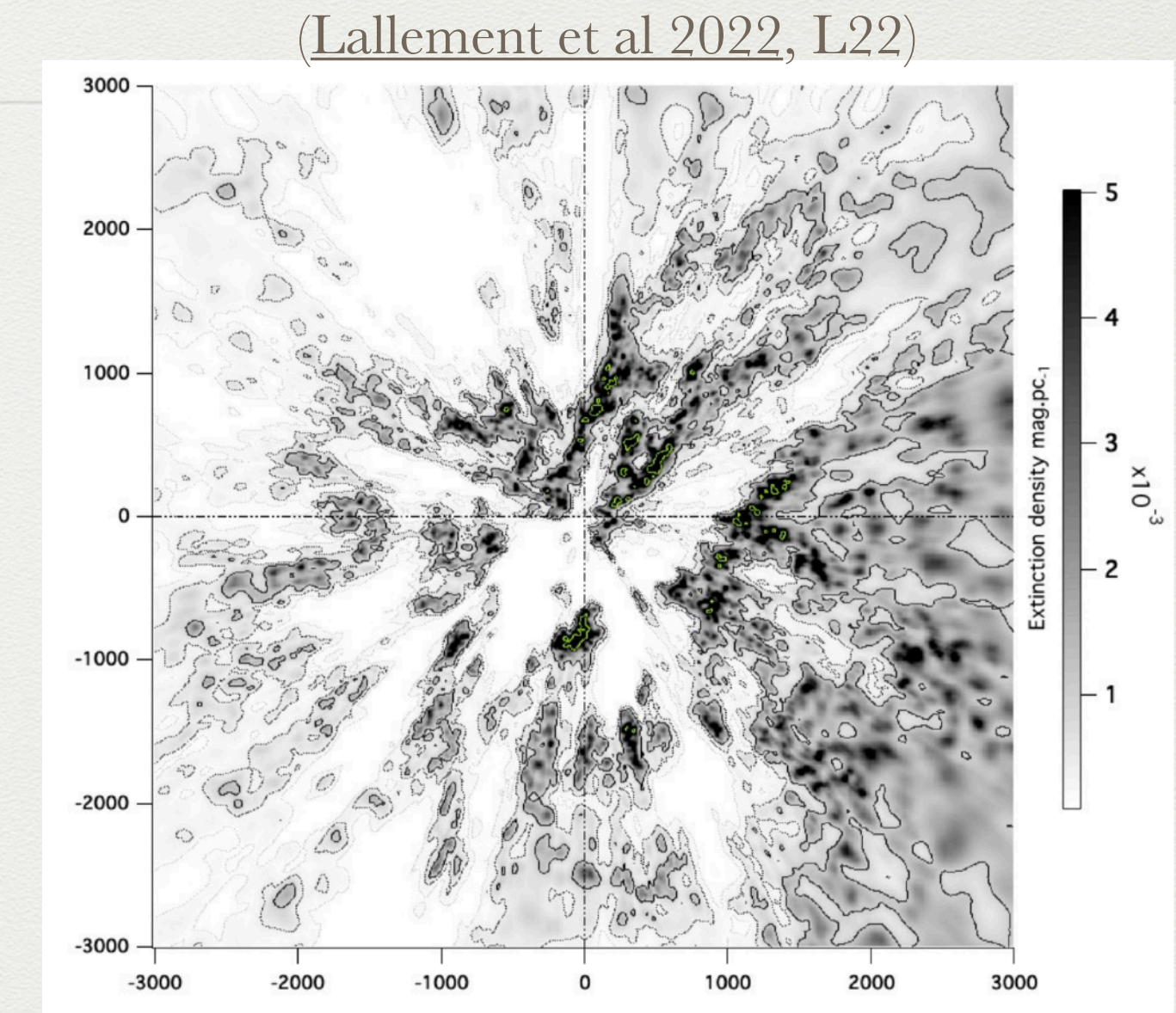


Results: Dust-Corrected PSD



Validating dust efficiency map

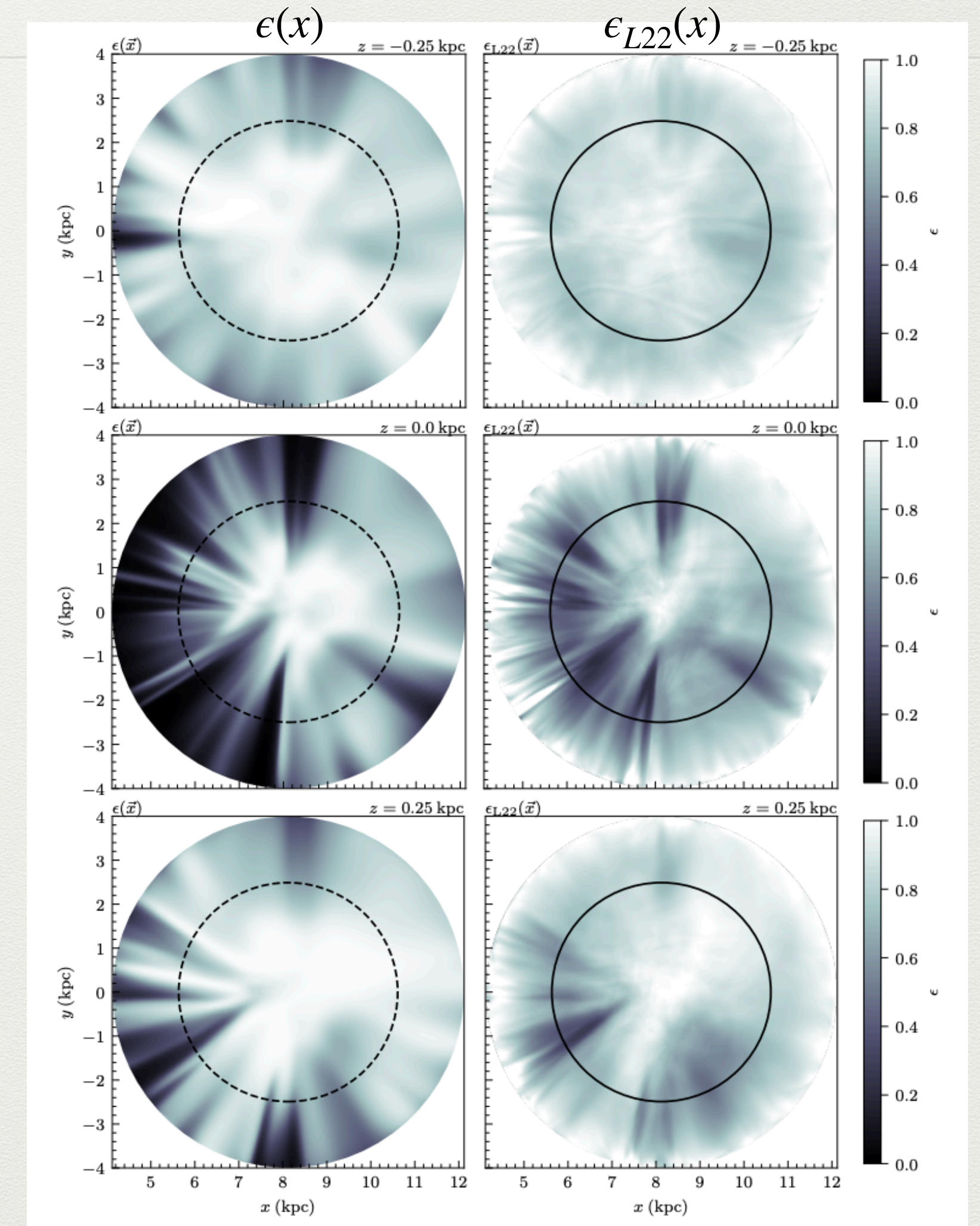
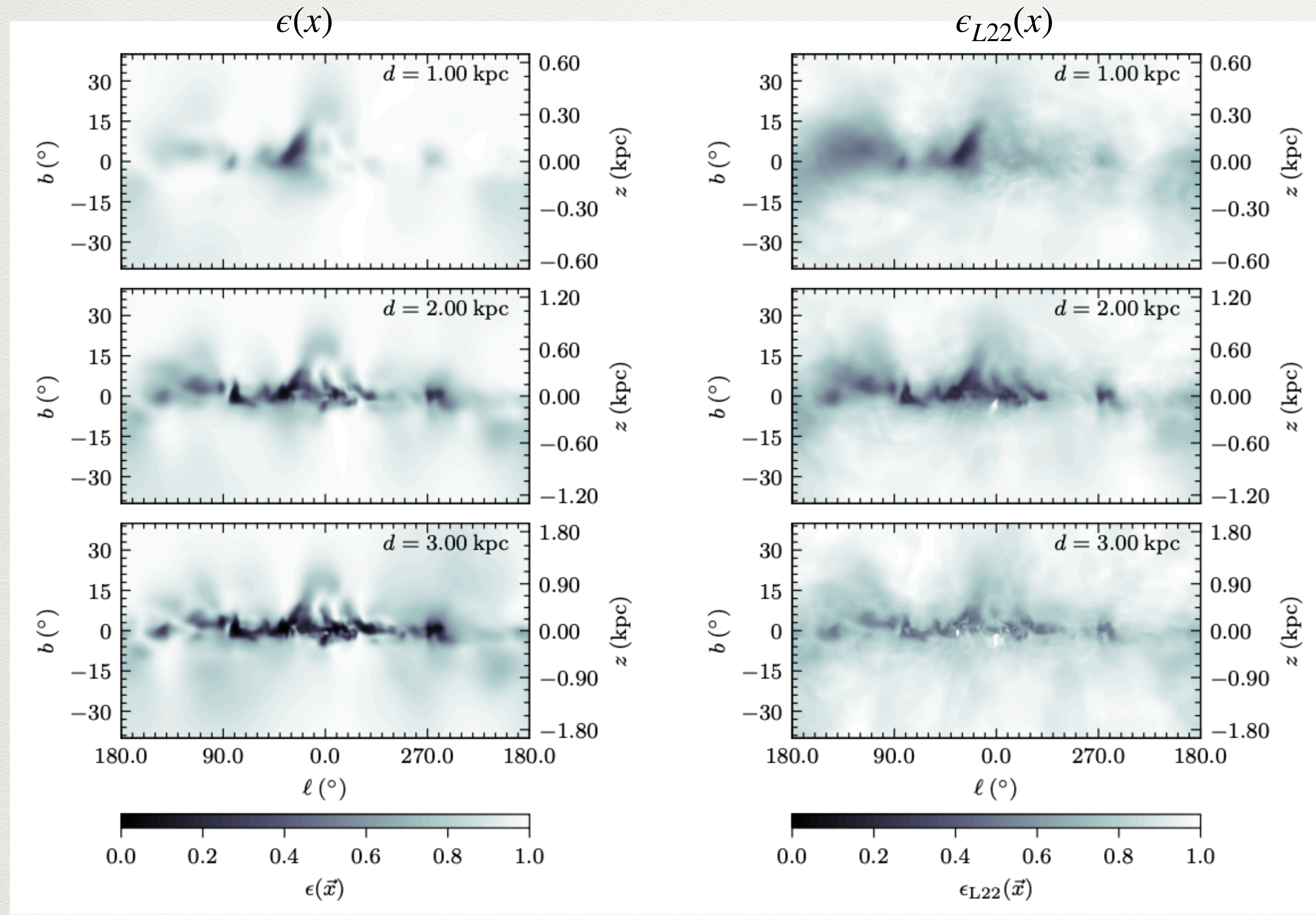
- How do we know that $\epsilon(x)$ is really correct?
- Astronomers have constructed extremely detailed and precise 3d dust maps using Gaia and other data
- However, these quantify how much an *observed* star is dimmed and reddened by dust.
- They cannot by themselves tell us about stars that are lost due to dust



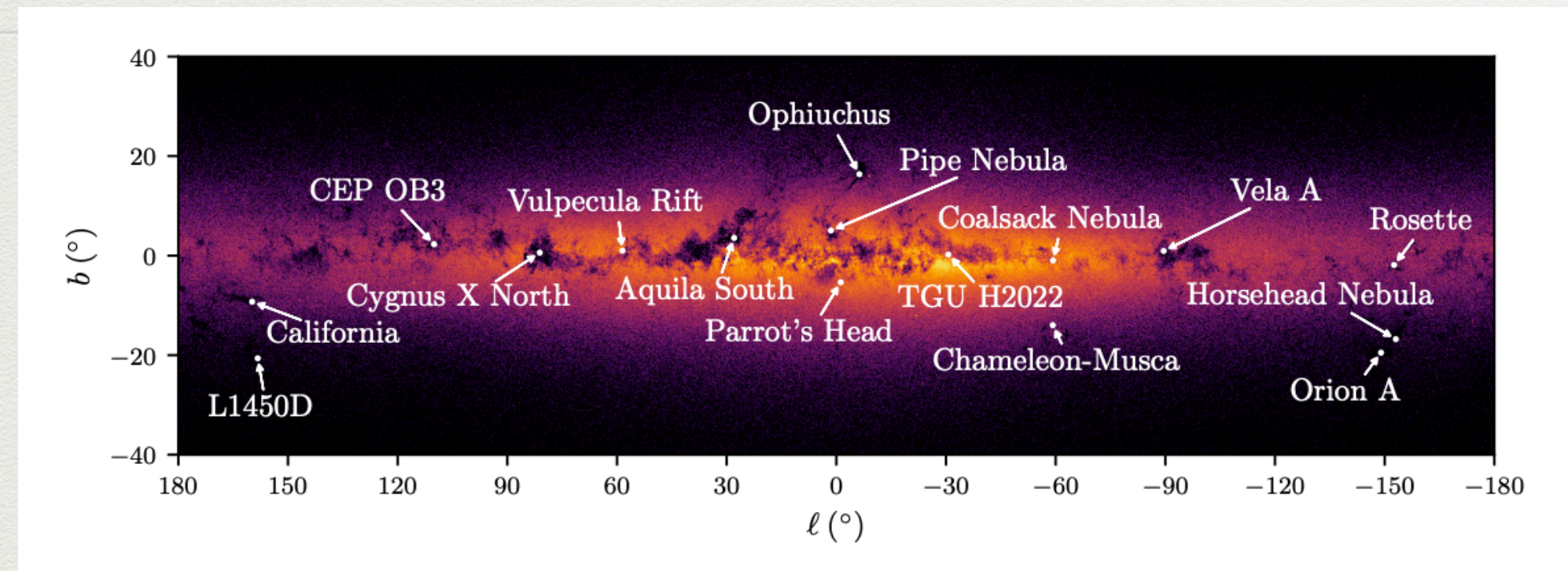
Validating dust efficiency map

- From L22 dust map we can construct an alternative efficiency map as follows:
 - Within ~ 2.5 kpc, Gaia is expected to be complete even accounting for dust extinction
 - Using L22 dust-corrected magnitudes, re-apply 4 kpc magnitude cut
 - Train flow on surviving stars to learn $n_{L22}(x)$
 - Compute $\epsilon_{L22}(x) = n_{L22}(x)/n_{obs}(x)$

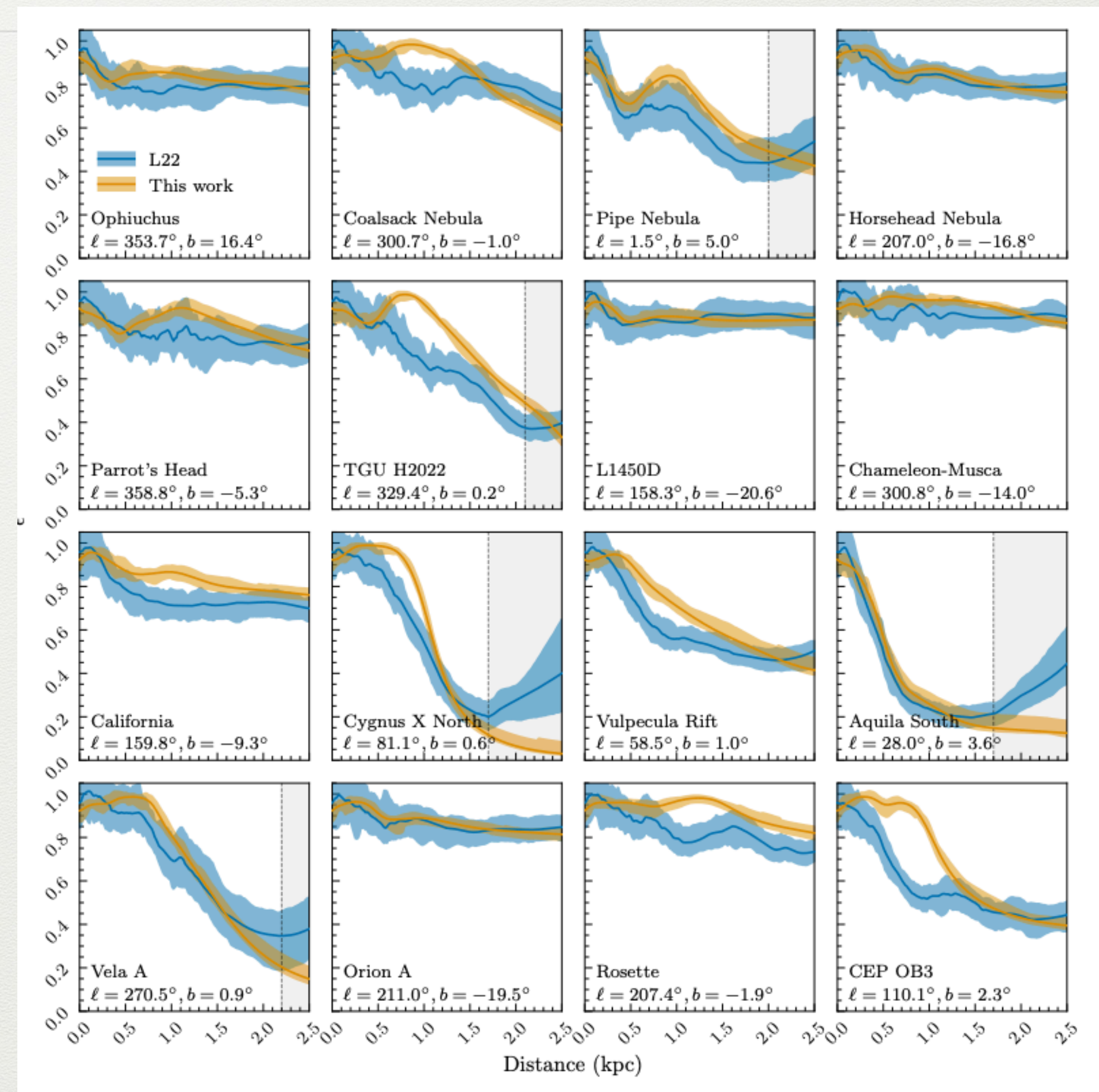
Validating dust efficiency map



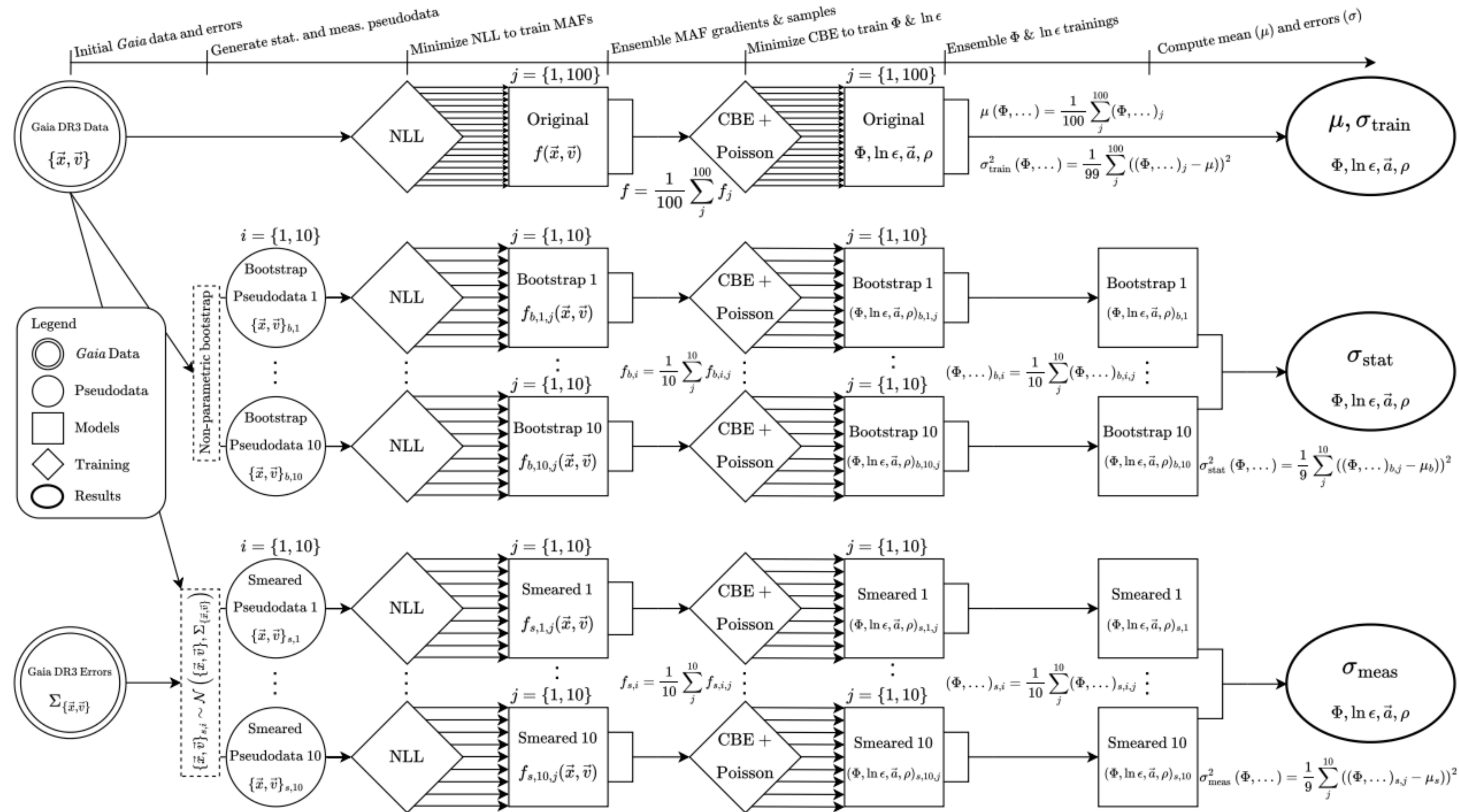
Validating dust efficiency map



(Mostly) reproduce known dust clouds!



Uncertainty quantification

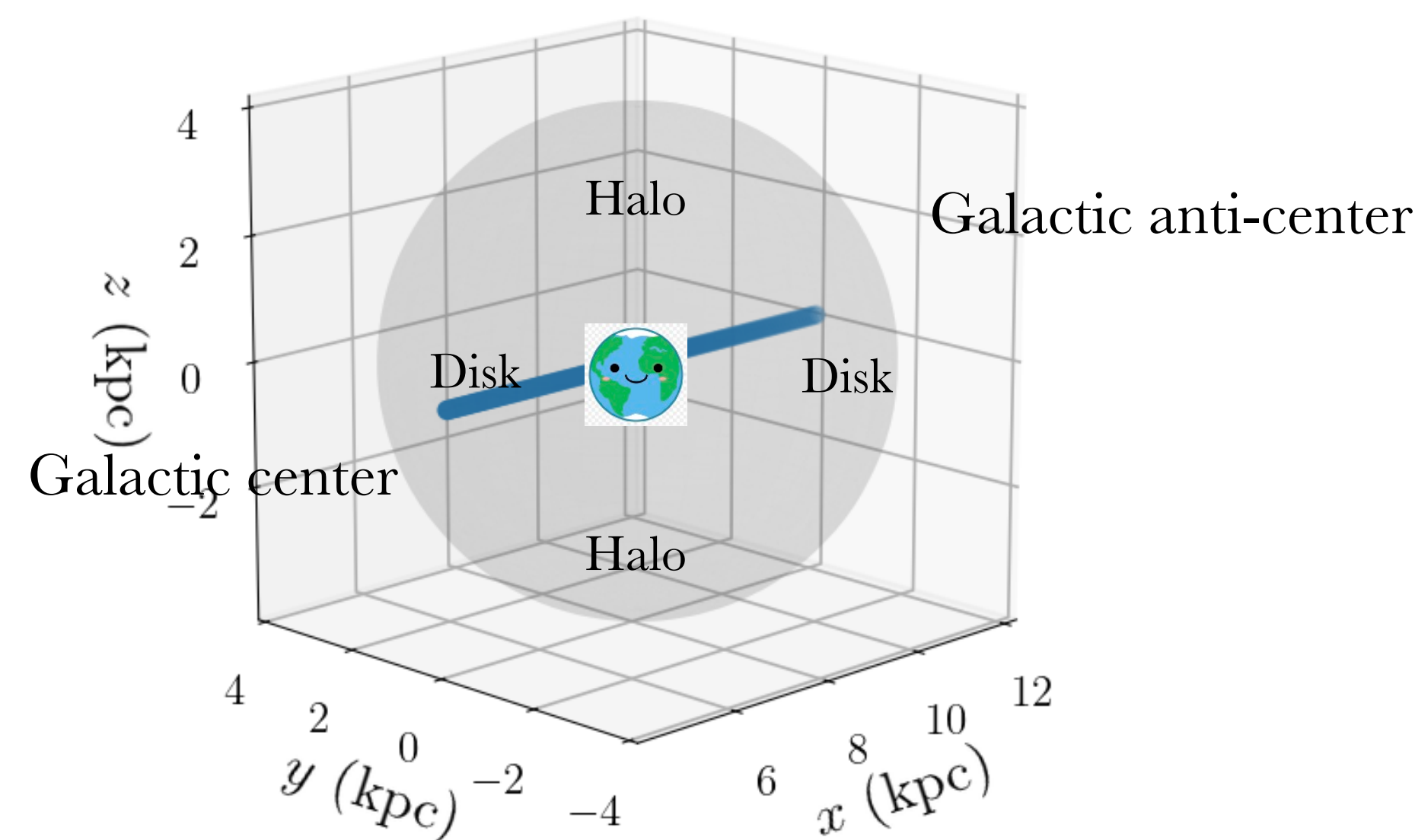


Quantified uncertainties:

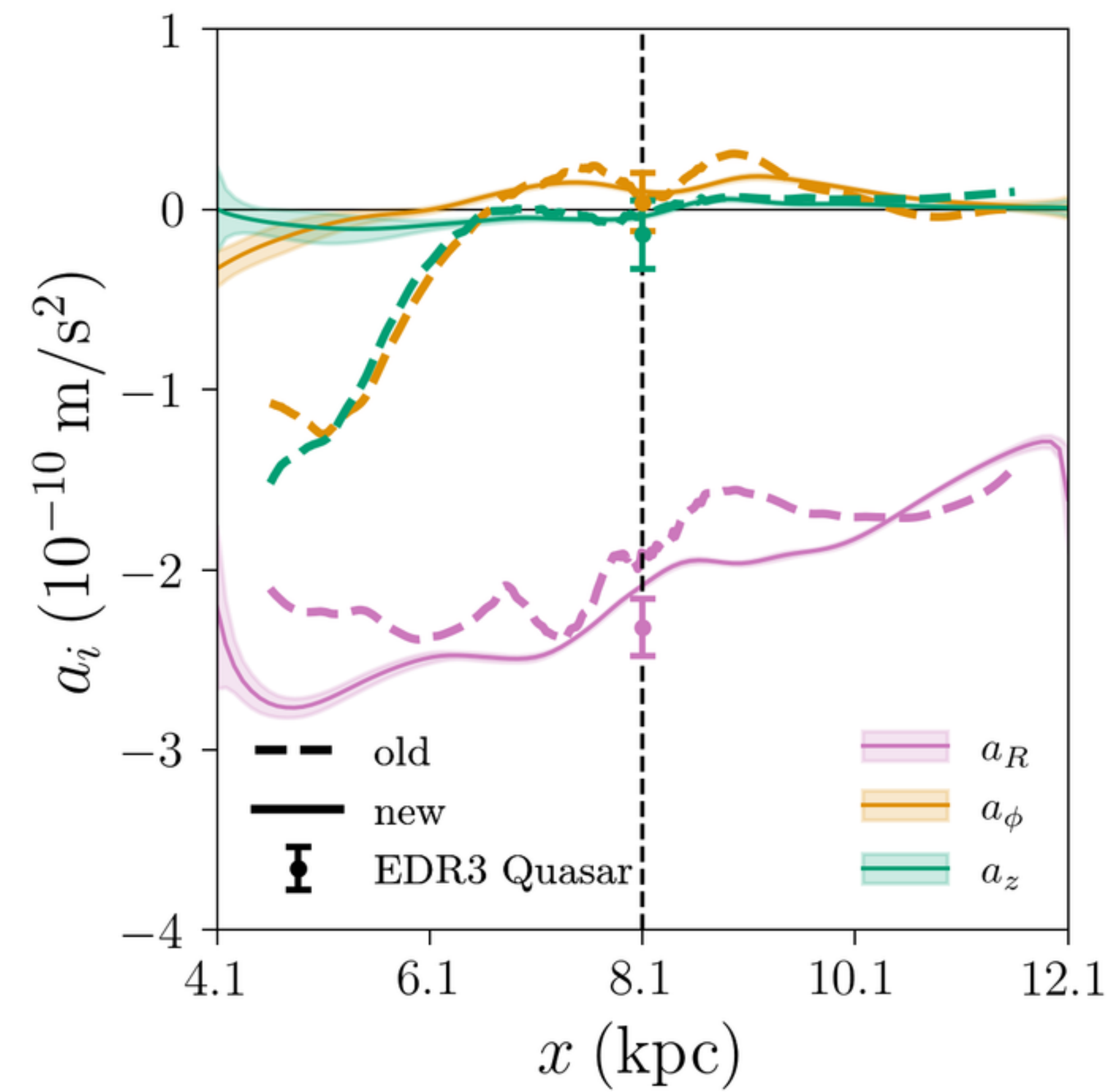
- finite training statistics
- training variance
- Gaia measurement errors

Results: accelerations

DS+ Putney et al 2511.xxxxx



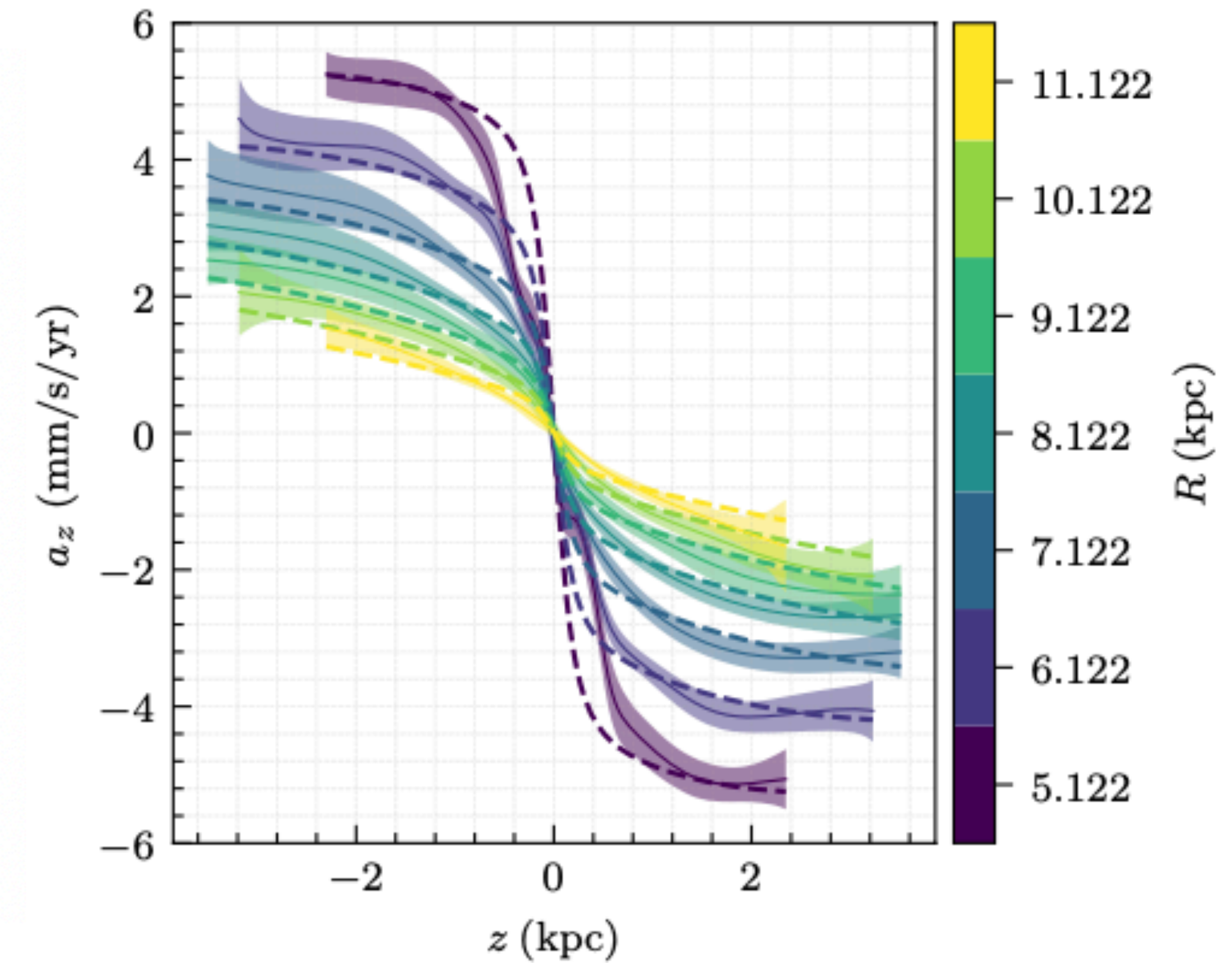
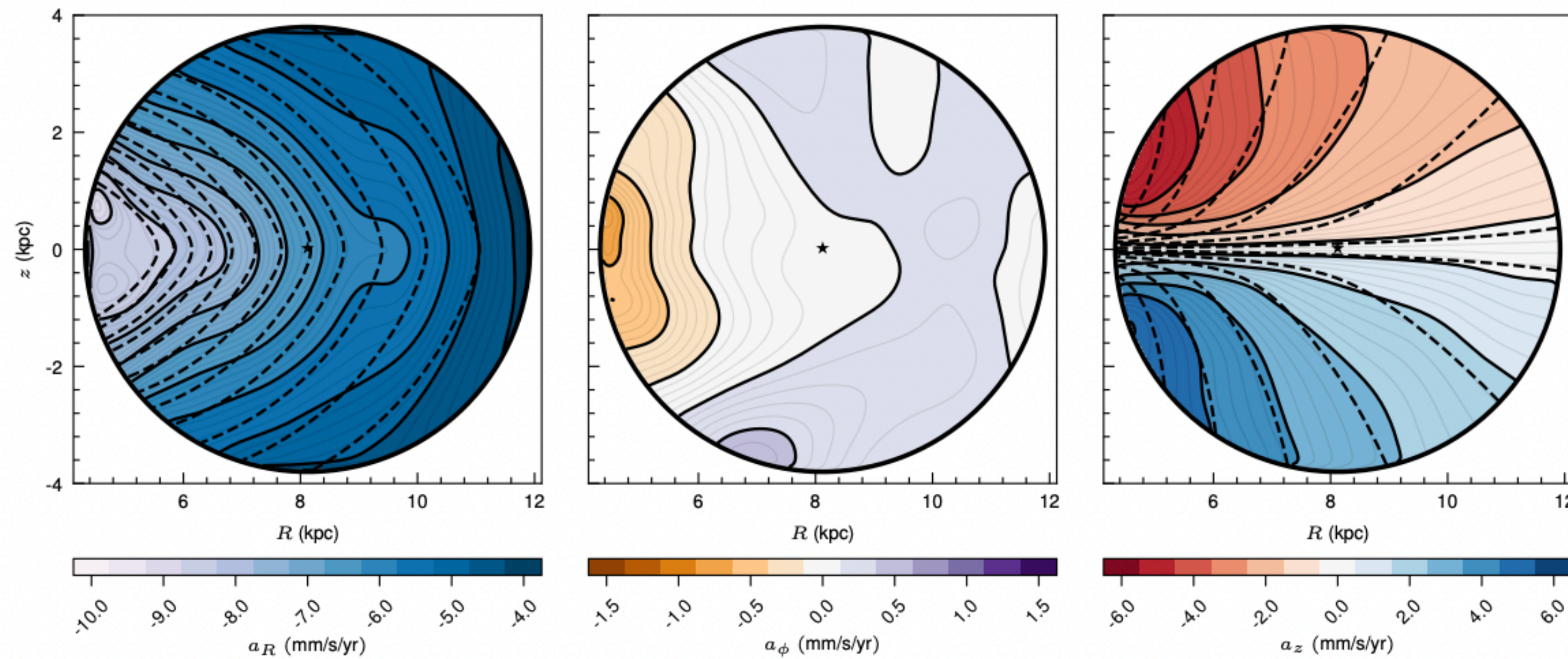
← Galactic center (dust clouds thicken)



Can now access accelerations in the disk!

Results: accelerations

DS+ Putney et al 2512.09989



- Generally good qualitative agreement with gala's MWPotential2014!

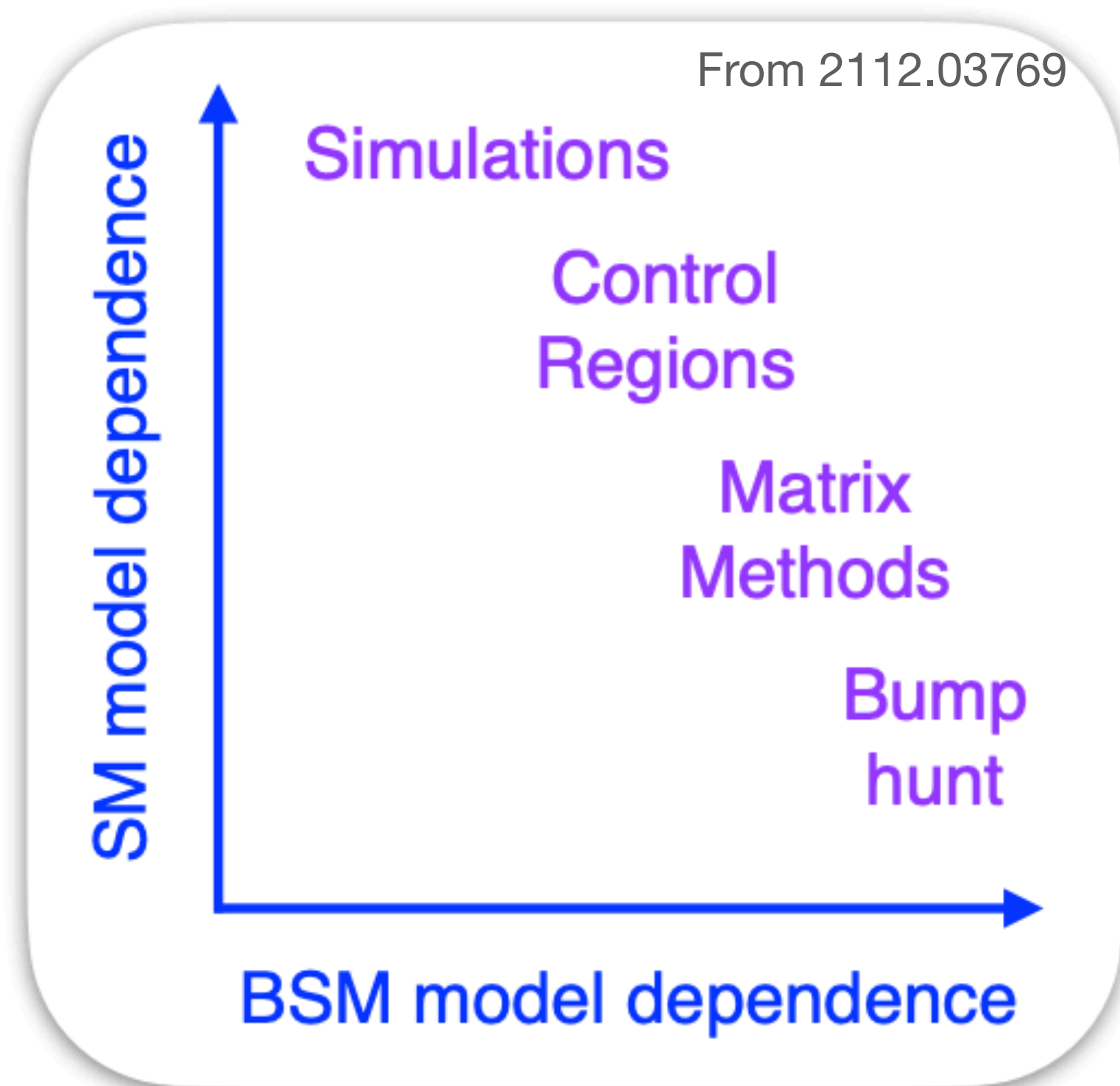
Two ingredients for a complete model-agnostic search

Two ingredients for a complete model-agnostic search

1. Anomaly score (signal sensitivity) — Outlier, Overdensity, ...

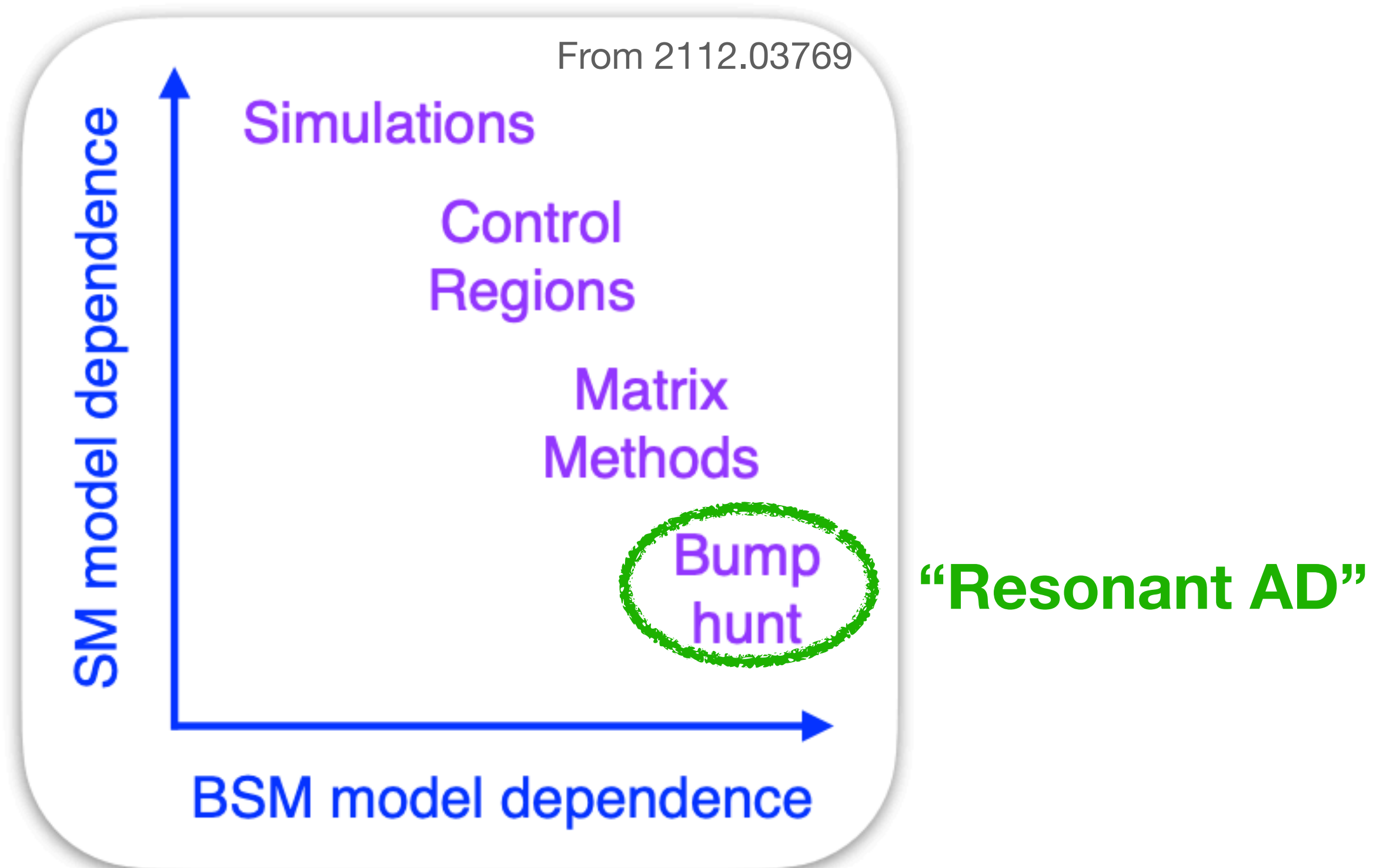
Two ingredients for a complete model-agnostic search

1. Anomaly score (signal sensitivity) — Outlier, Overdensity, ...
2. Background estimation



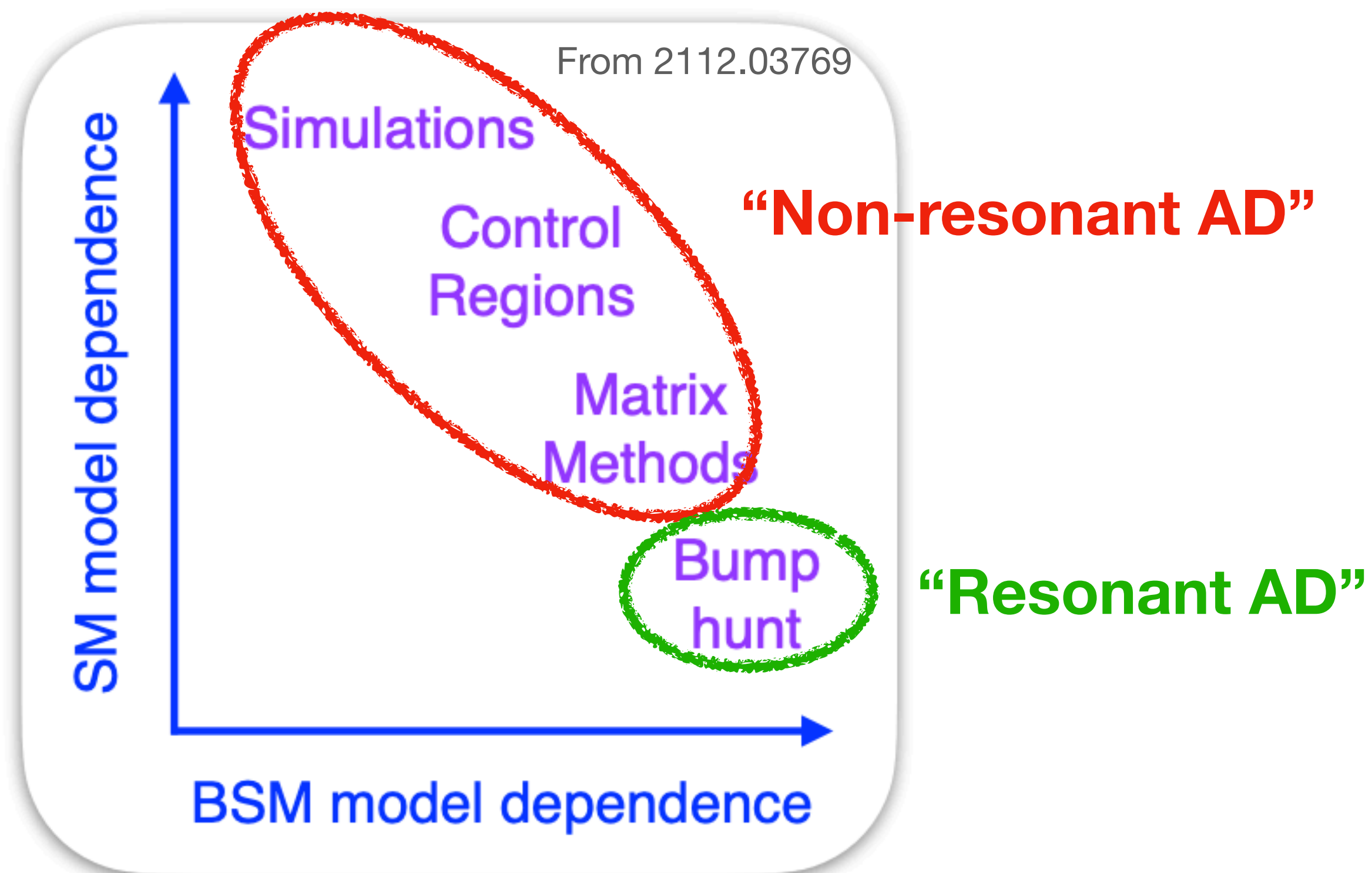
Two ingredients for a complete model-agnostic search

1. Anomaly score (signal sensitivity) — Outlier, Overdensity, ...
2. Background estimation



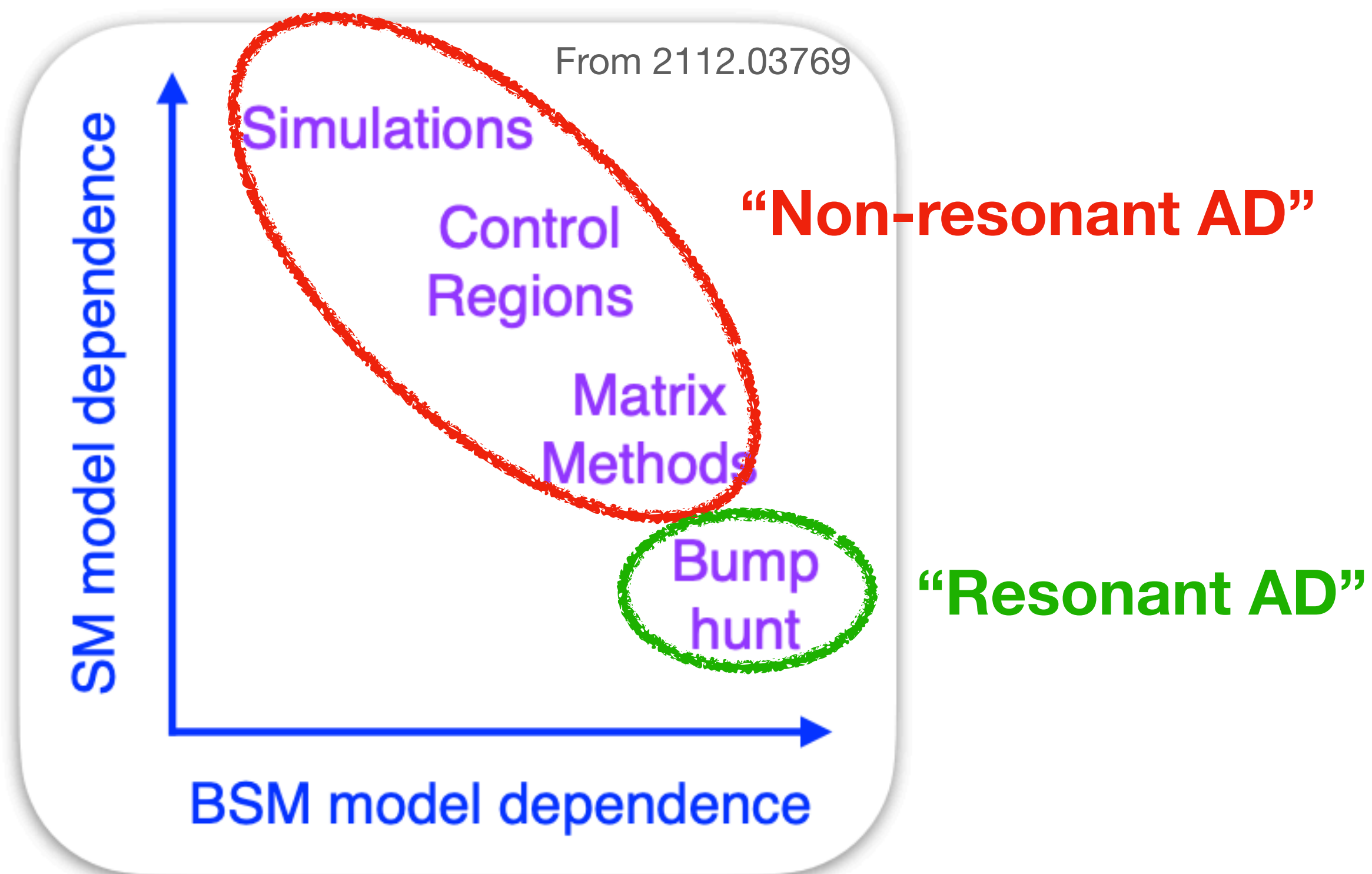
Two ingredients for a complete model-agnostic search

1. Anomaly score (signal sensitivity) — Outlier, Overdensity, ...
2. Background estimation



Two ingredients for a complete model-agnostic search

1. Anomaly score (signal sensitivity) — Outlier, Overdensity, ...
2. Background estimation

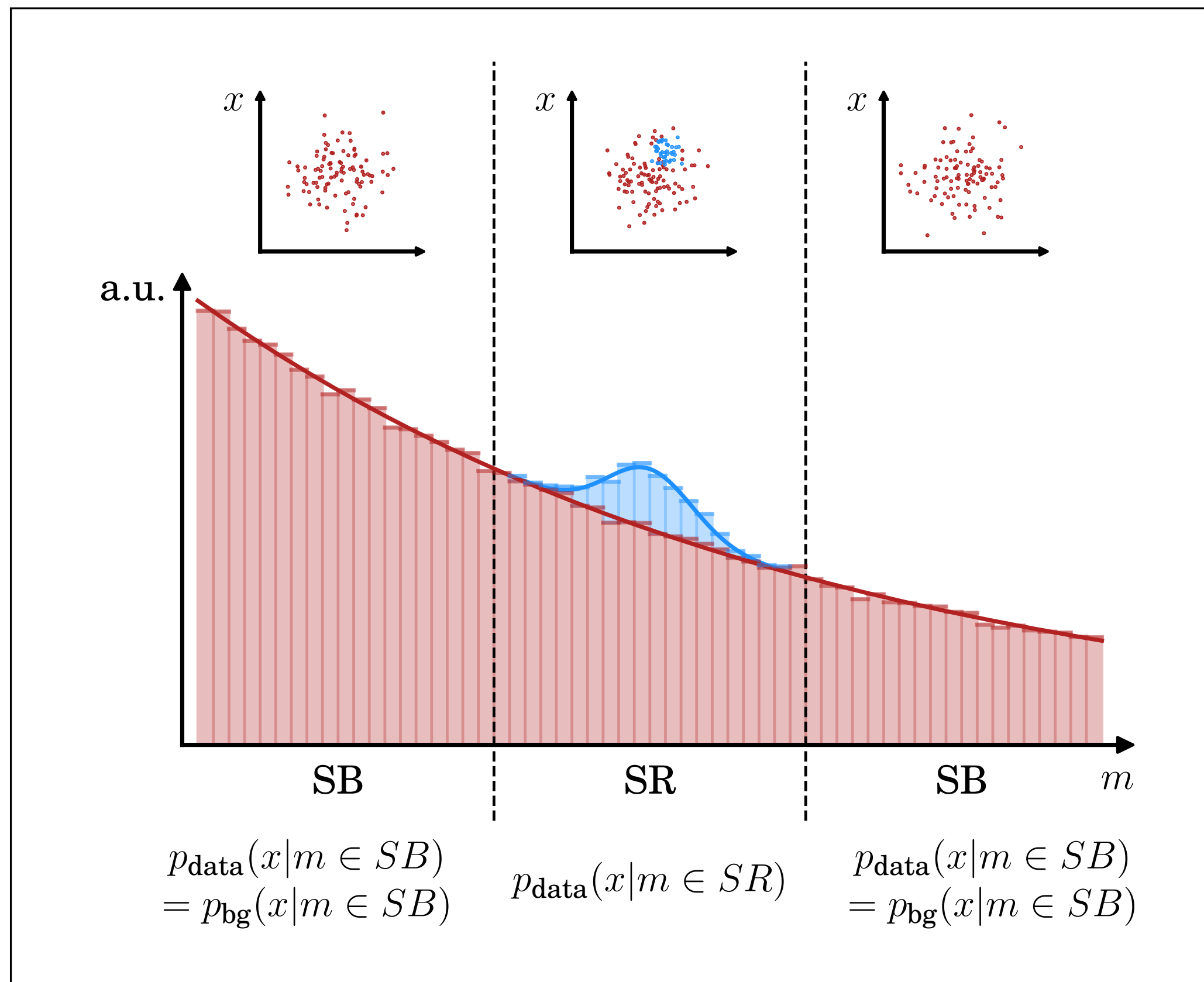


Many proofs-of-concept developed for both Resonant and Non-resonant AD.

So far, actual LHC analyses predominantly with Resonant AD.

Estimating the Overdensity Anomaly Score in Resonant AD

- **Many** proposals: CWoLa, ANODE, SALAD, CATHODE, CURTAINS, FETA, R-ANODE, ...
- General idea:
 - use ML method such as **normalizing flow** to learn multi-dimensional background in SB,
 - interpolate into SR,
 - compare with data in SR.



Example: Classifying Anomalies Through Outer Density Estimation (CATHODE)

DS+ Hallin et al (2022)

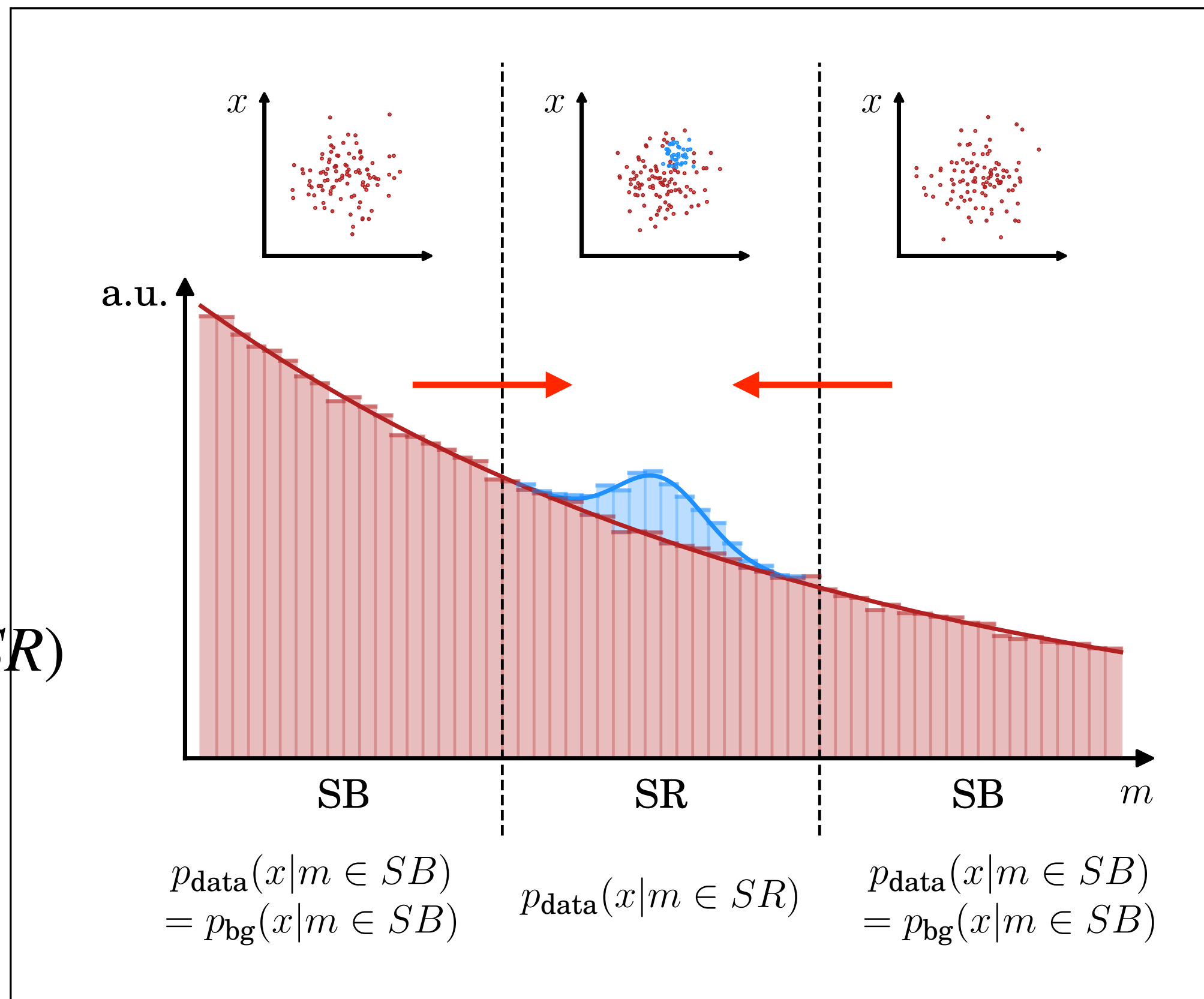
1. Fit normalizing flow to sidebands to learn $p_{bg}(x|m \in SB)$

2. Interpolate

$$p_{bg}(x|m \in SB) \rightarrow p_{bg}(x|m \in SR)$$

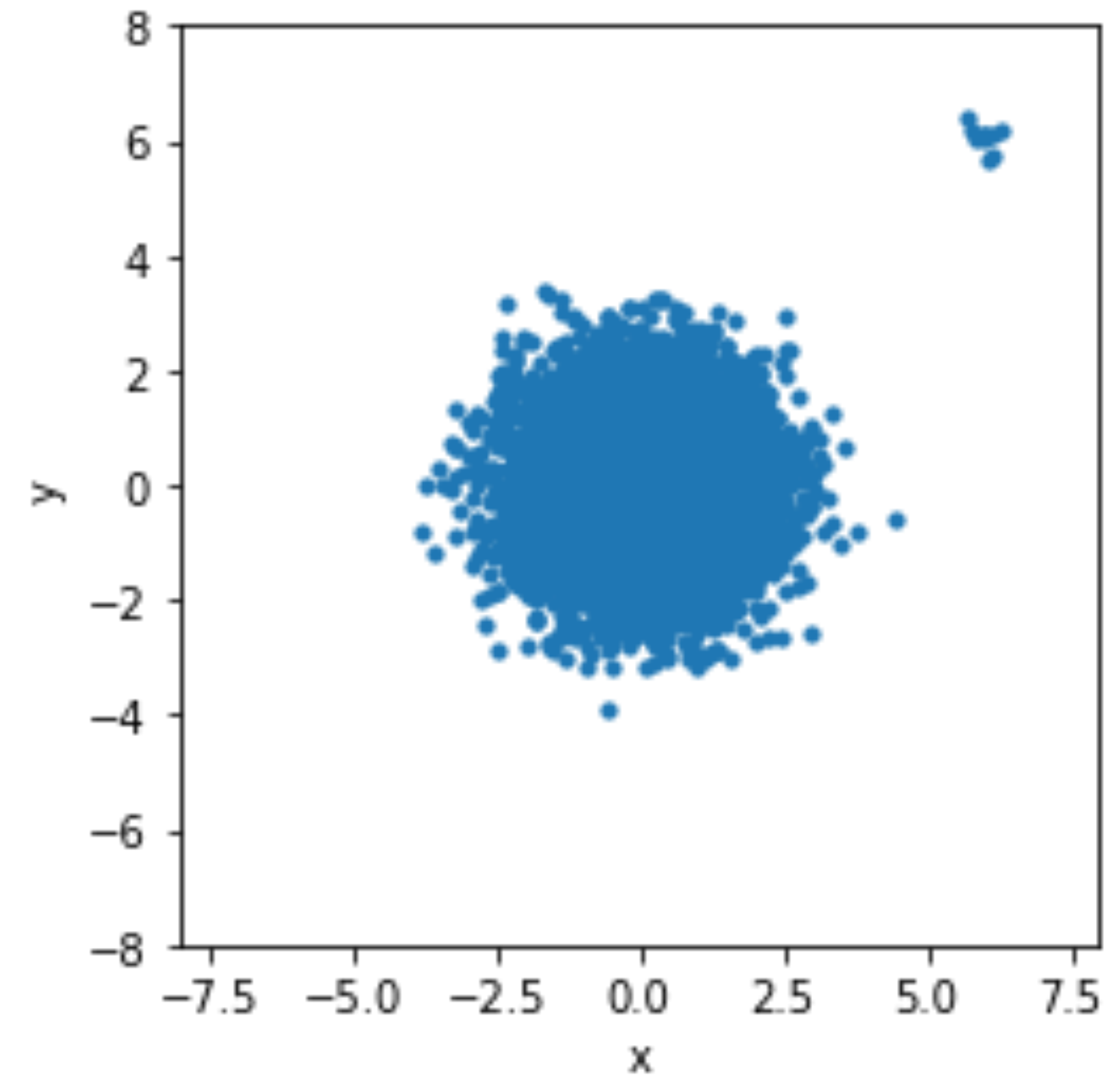
3. Sample background events in the SR from interpolated flow

4. Train classifier on data vs samples to learn $R(x)$



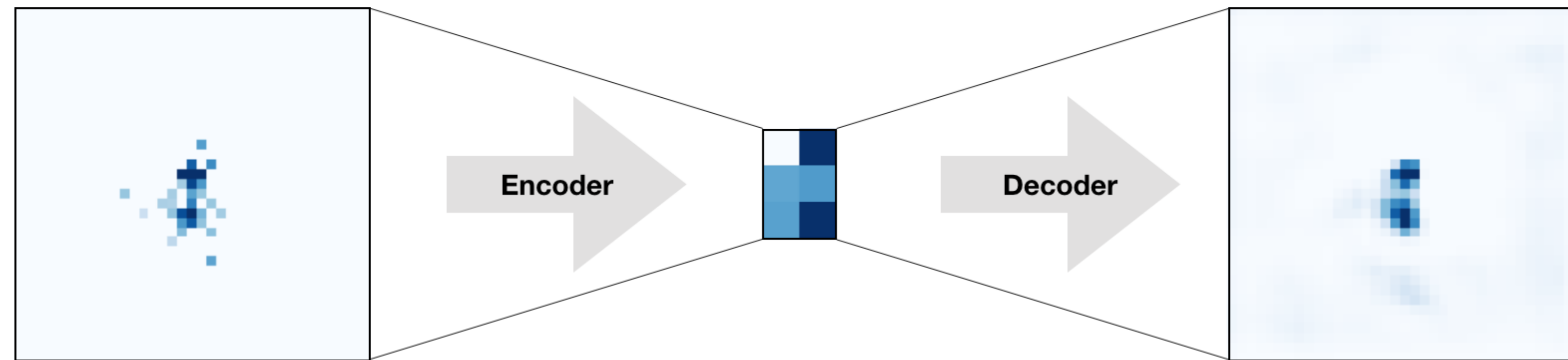
Outlier detection

- If our data looks like this, then outliers are easy to find
- Lots of conventional clustering techniques (k-means, kNN, DBSCAN, ...)
- But what if
 - data is higher dimensional?
 - ideal clustering features are not obvious?



Autoencoders

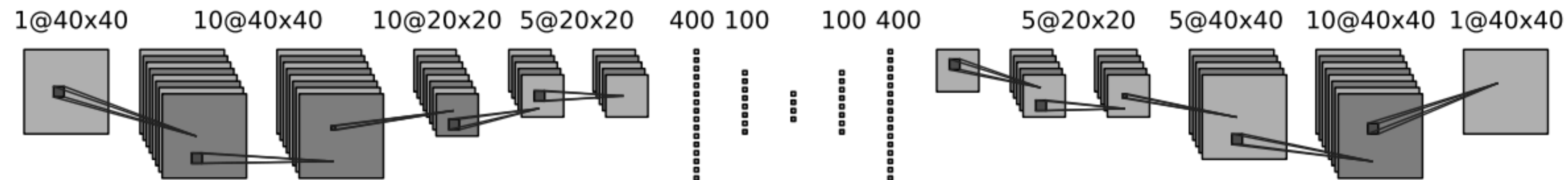
One deep-learning approach: train an **autoencoder** on the data



Farina, Nakai & **DS** 1808.08992

$$L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$$

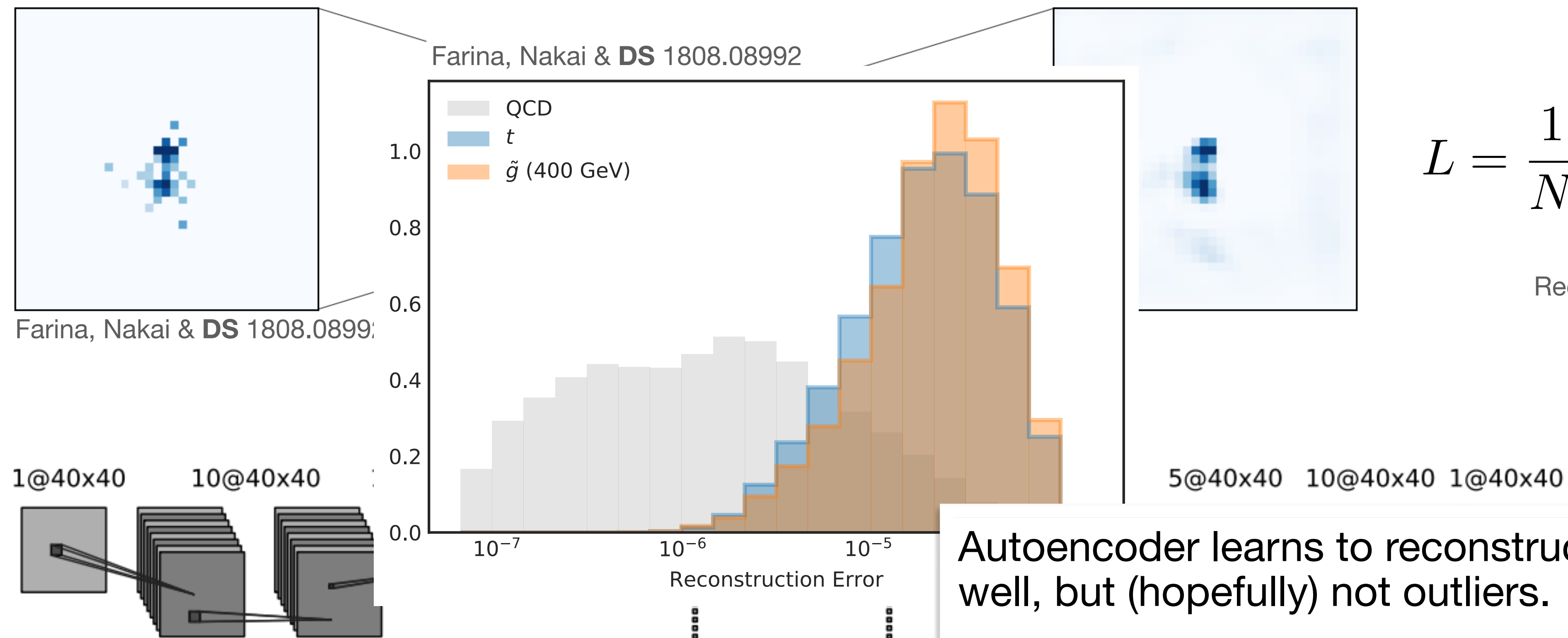
Reconstruction error loss



Heimel et al 1808.08979

Autoencoders

One deep-learning approach: train an **autoencoder** on the data



$$L = \frac{1}{N} \sum_{i=1}^N (x_i^{in} - x_i^{out})^2$$

Reconstruction error loss

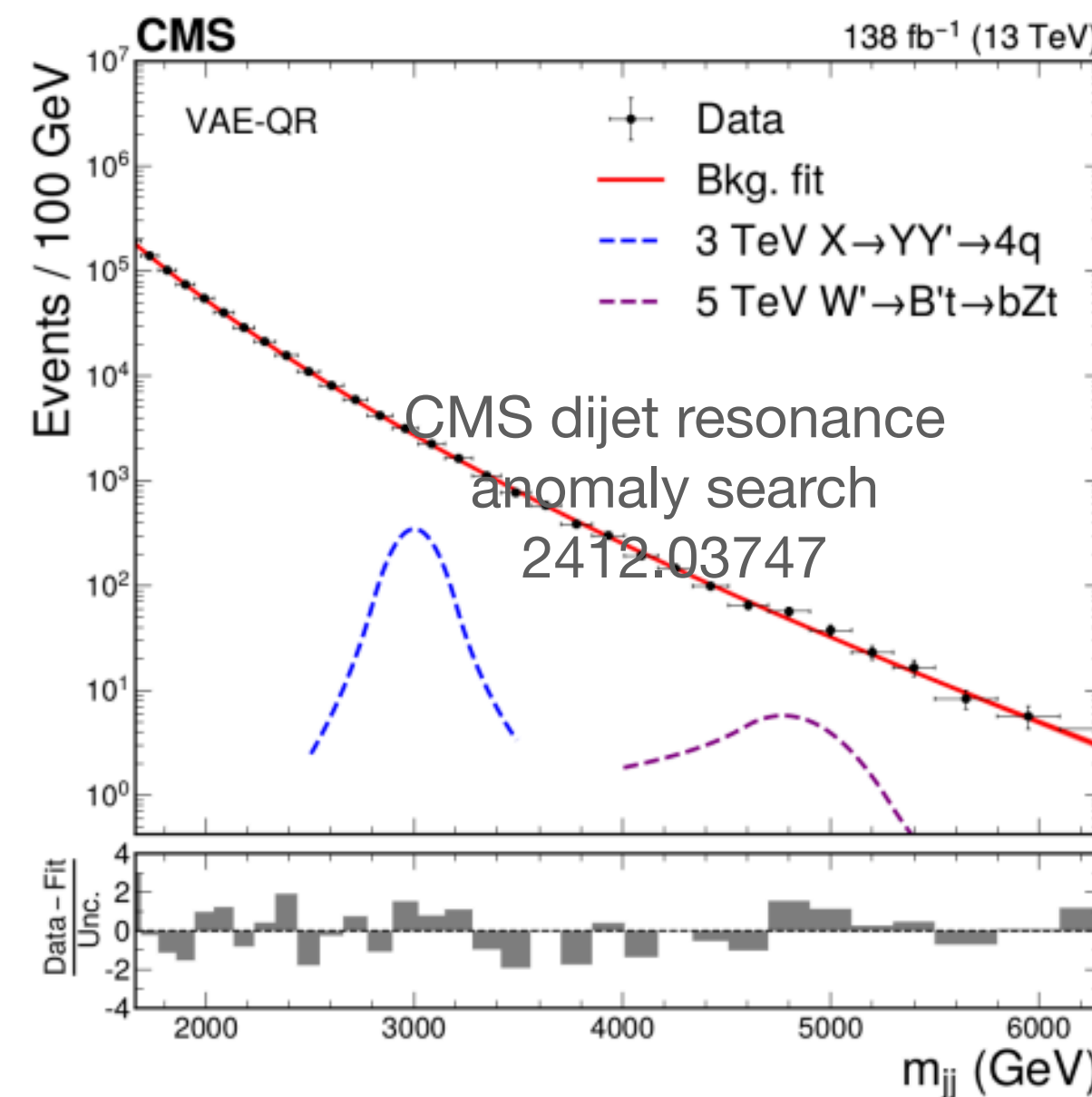
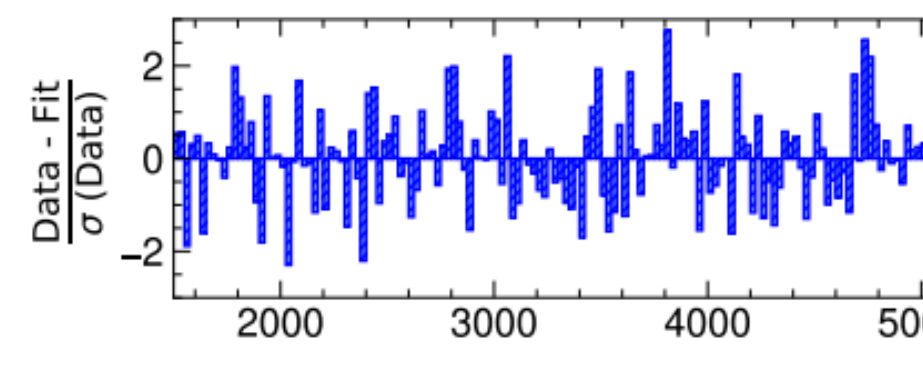
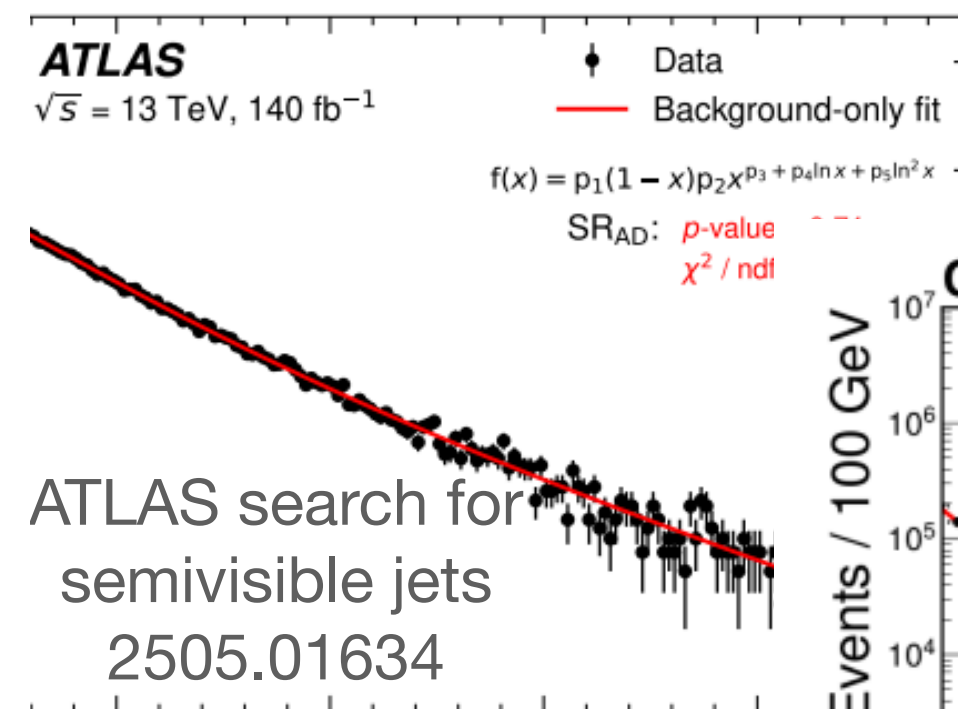
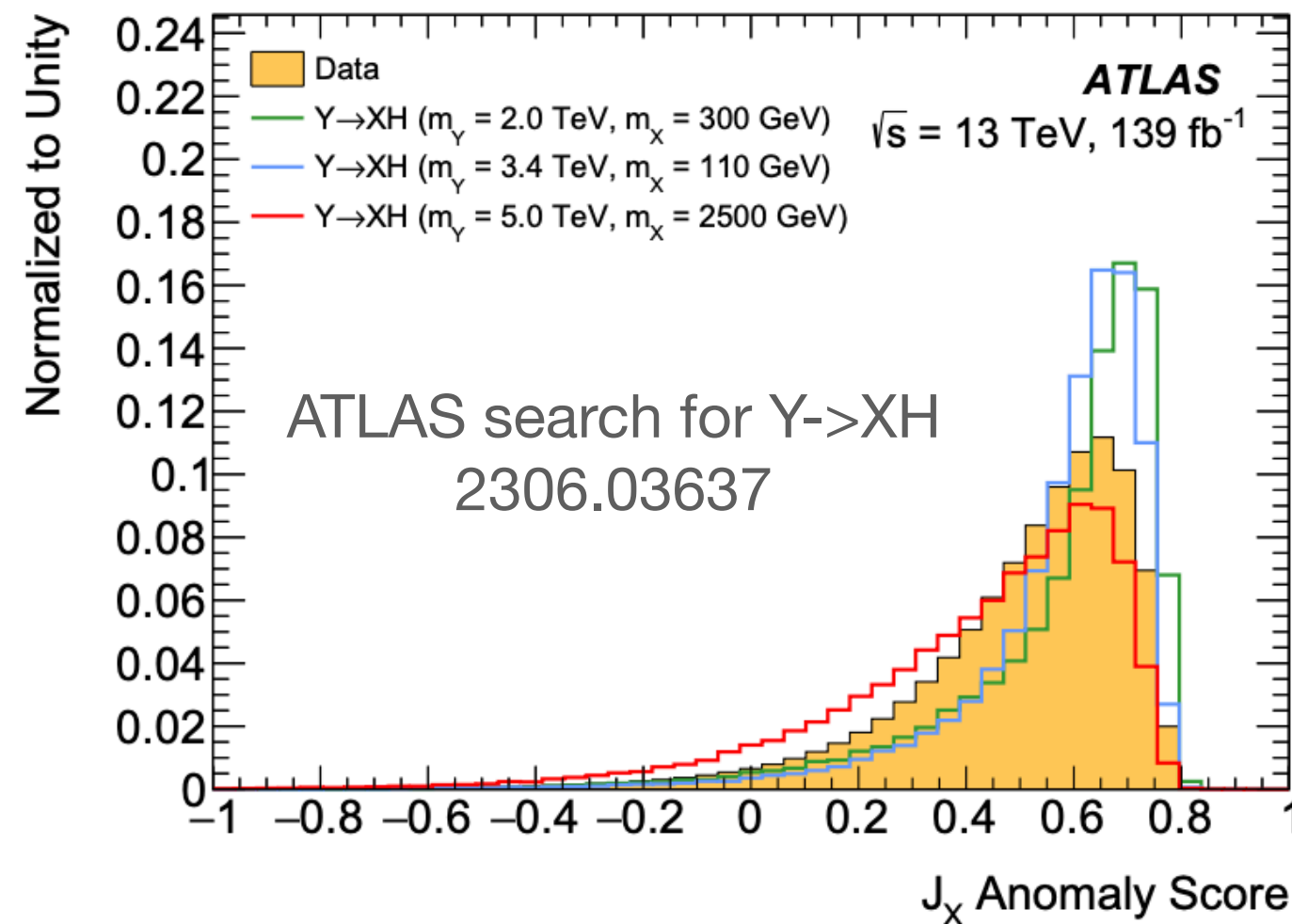
Autoencoder learns to reconstruct normal events well, but (hopefully) not outliers.

Can use reconstruction error as an ***anomaly score***

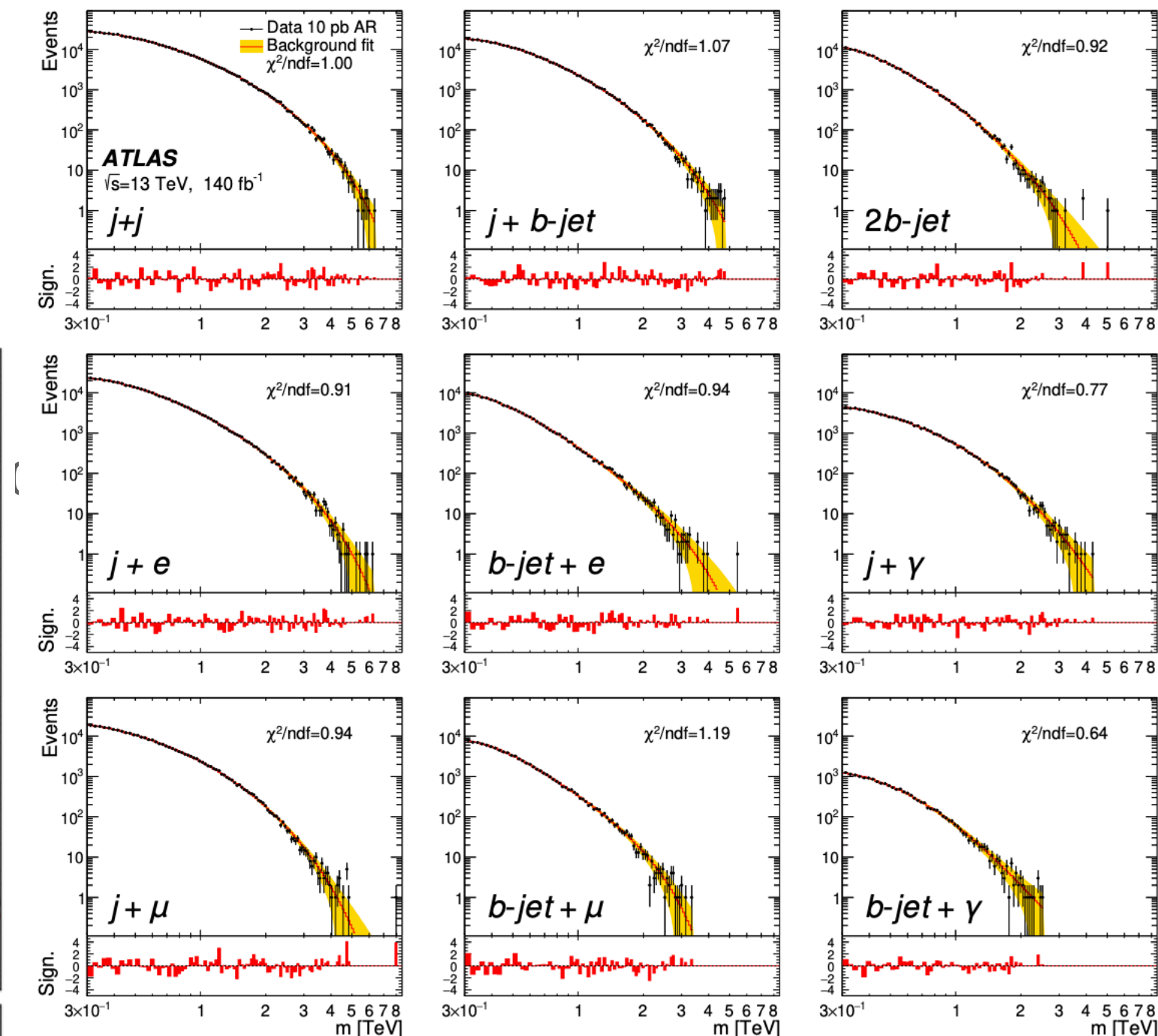
Heimel et al 1808.08979

AE-enhanced bump hunts @ LHC

Outlier detection + bump hunting with autoencoders are off to a good start at the LHC!



ATLAS search for anomalous jet+X
2307.01612

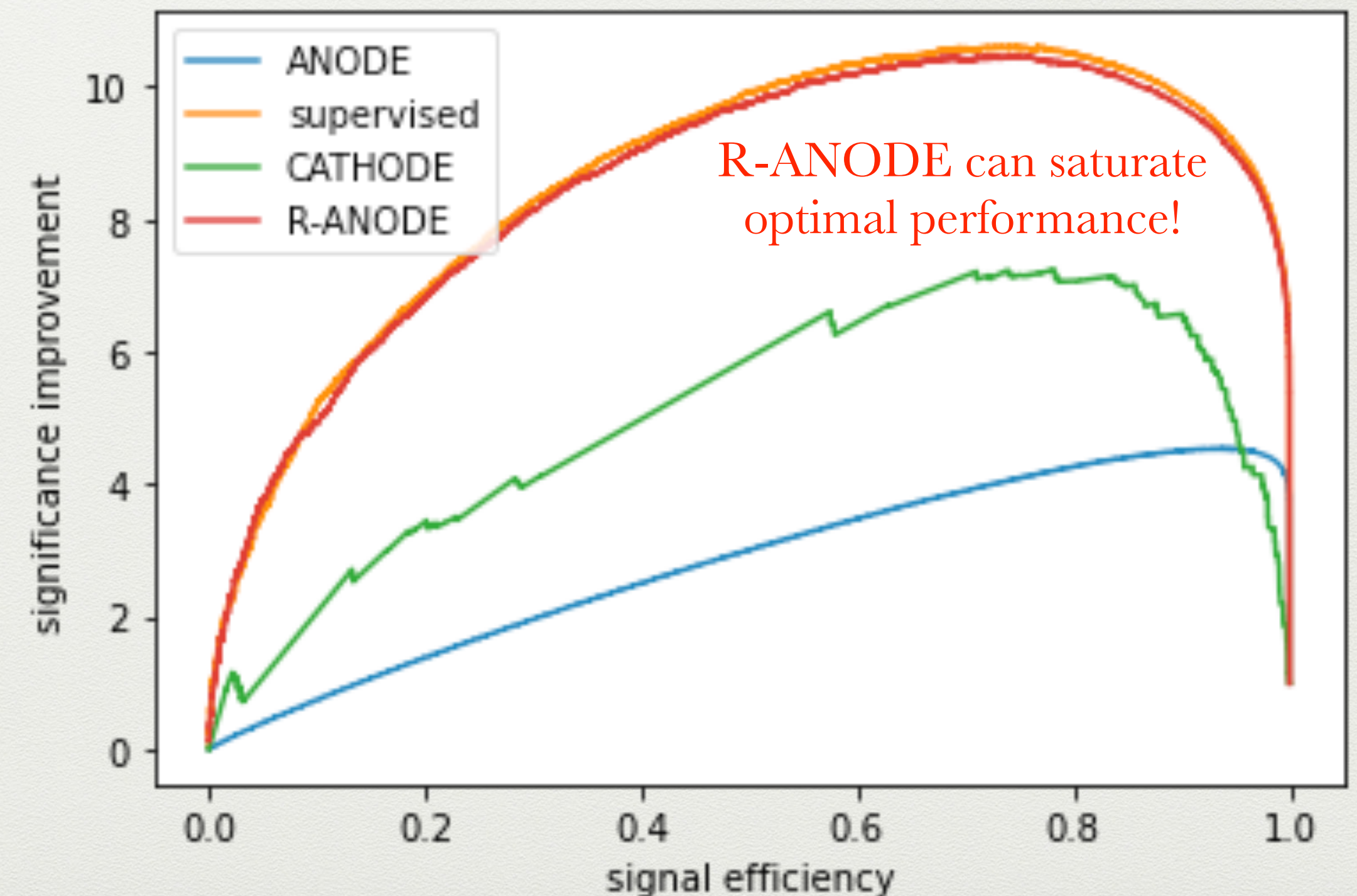


Residual ANODE

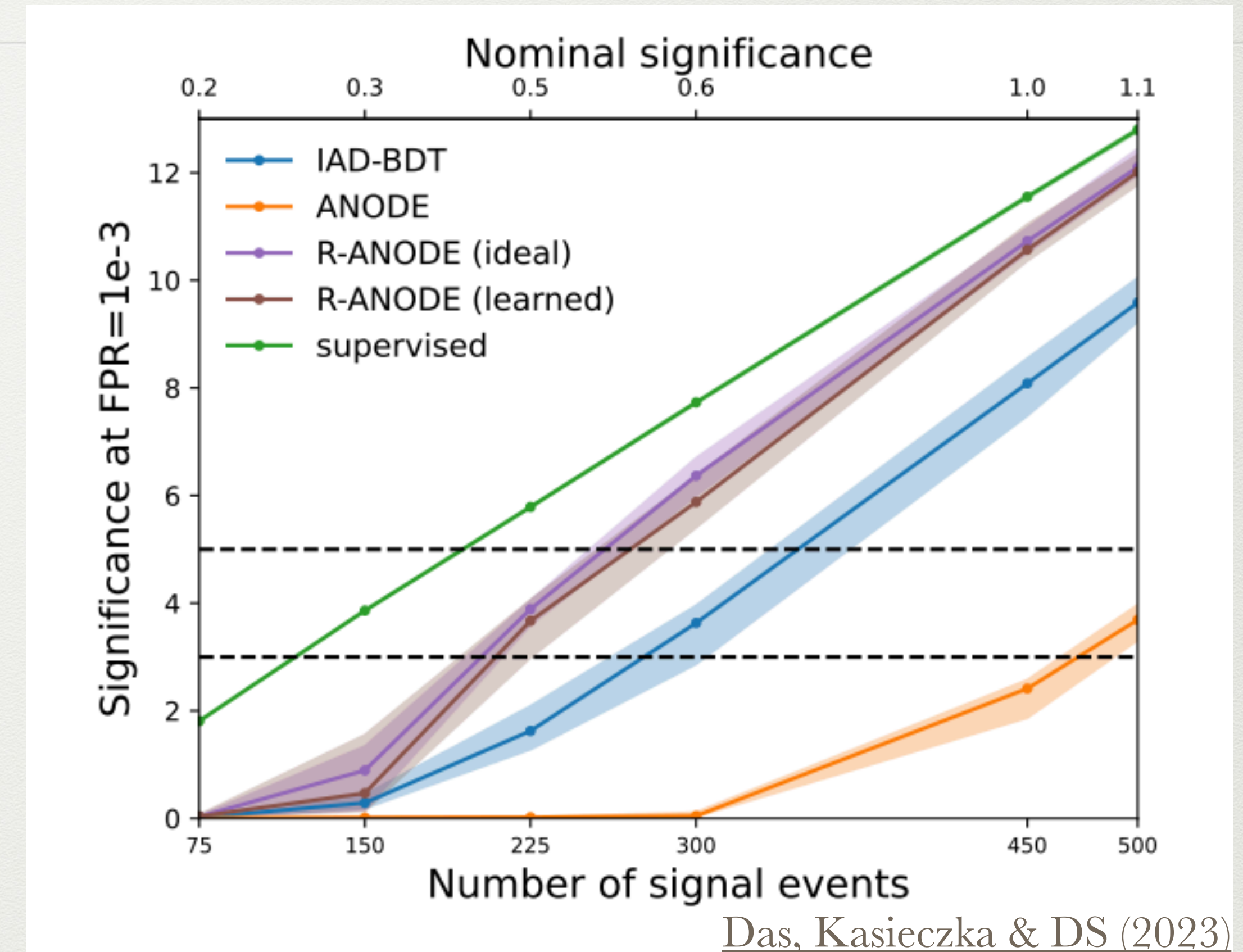
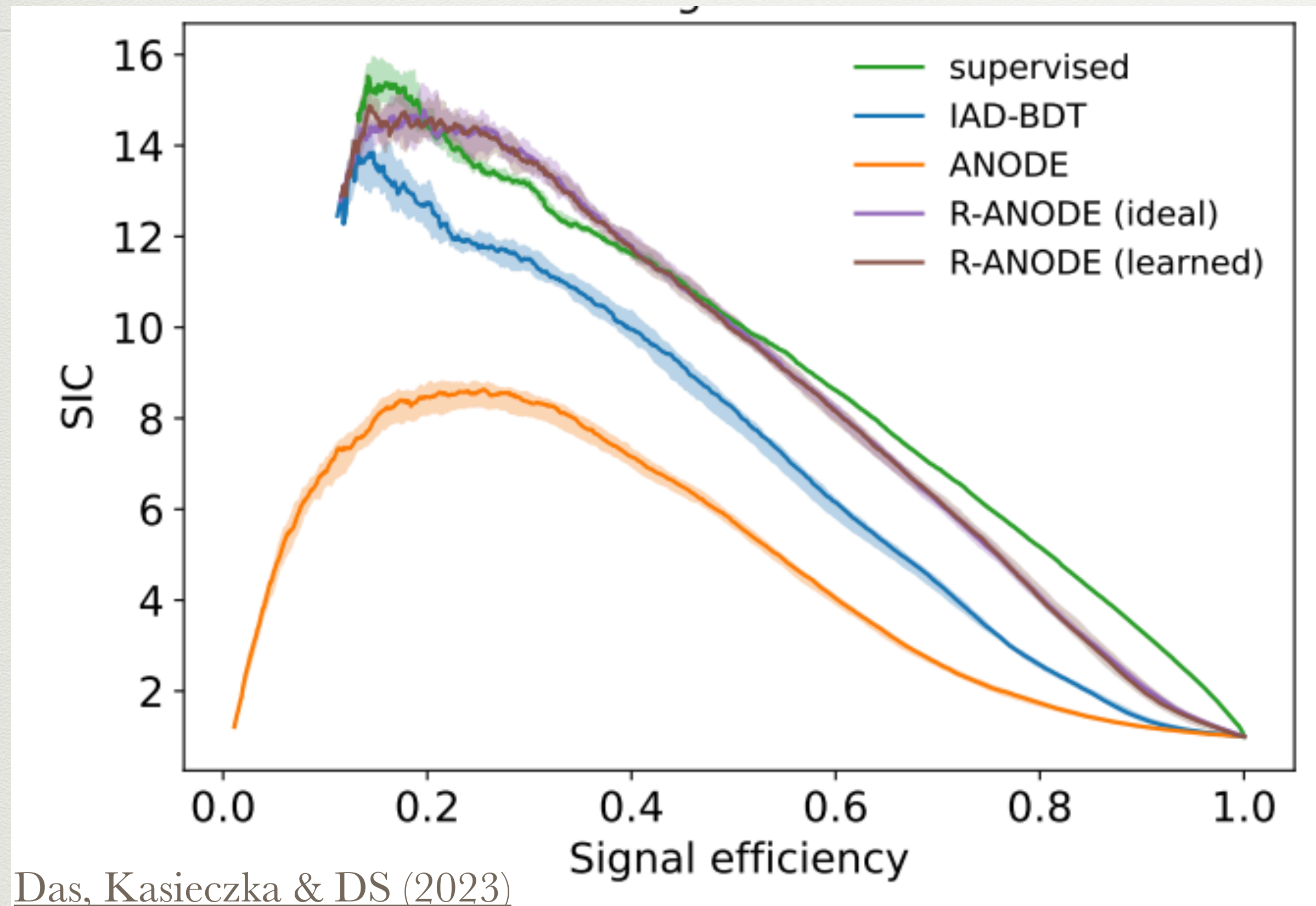
[Das, Kasieczka & DS (2023)]

Can do even better by freezing background density and only fitting a small “residual” density estimator to learn the signal in data

1. Fit $p_{bg}(x)$ to background
2. Freeze $p_{bg}(x)$ and fit $(1-w)p_{bg}(x) + wp_{sig}(x)$ to data

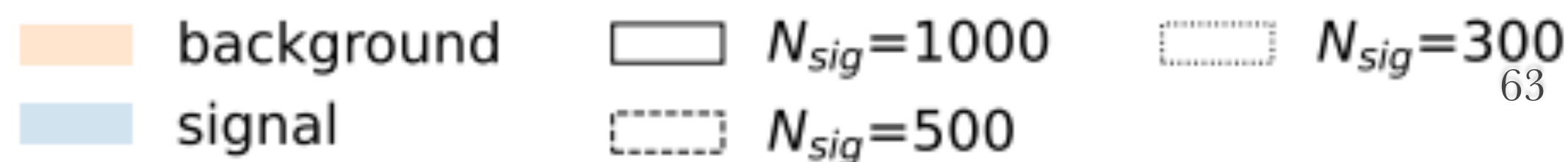
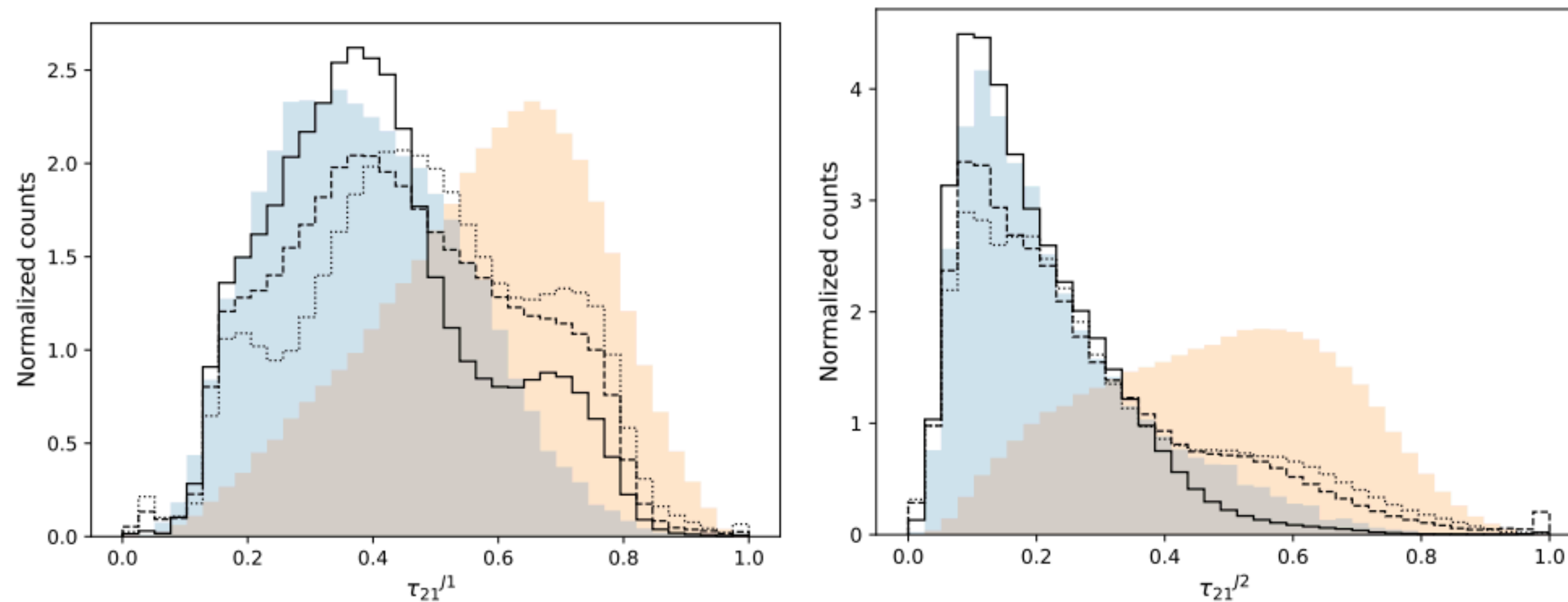
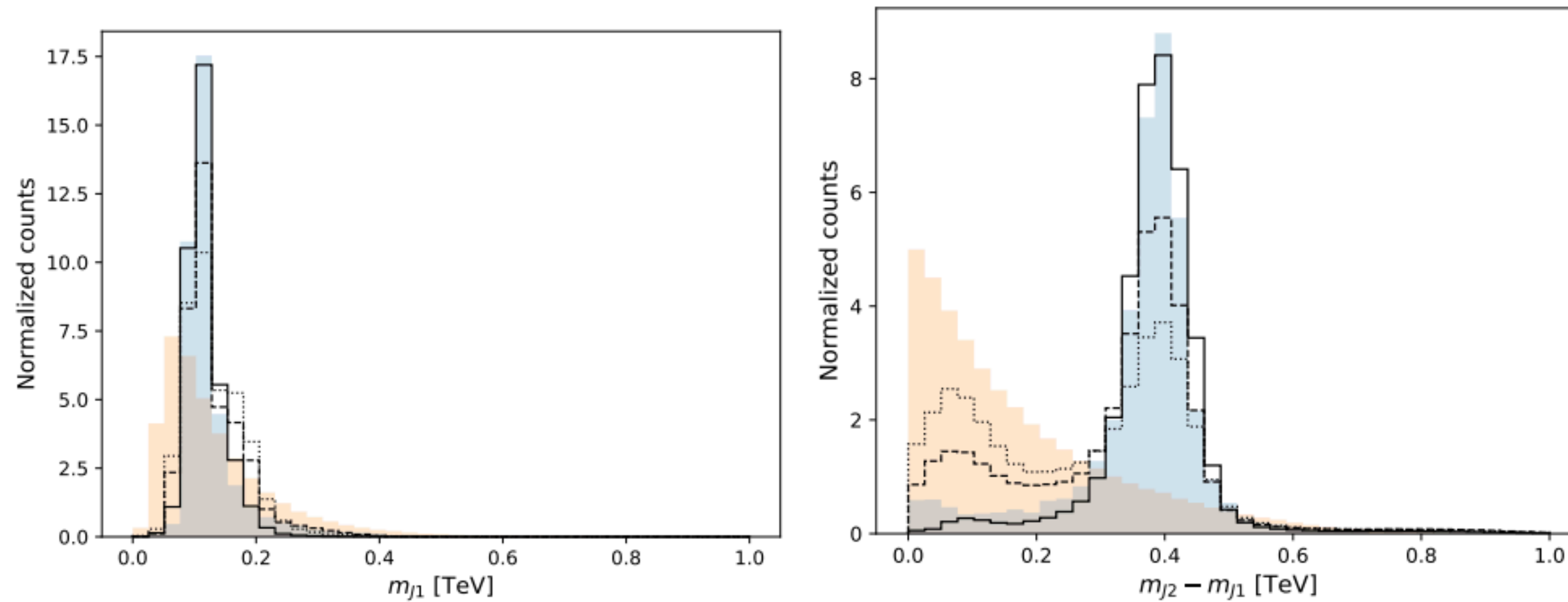


Performance on LHCO



R-ANODE exceeds classifier-based approaches
Closes/narrows the gap with fully supervised

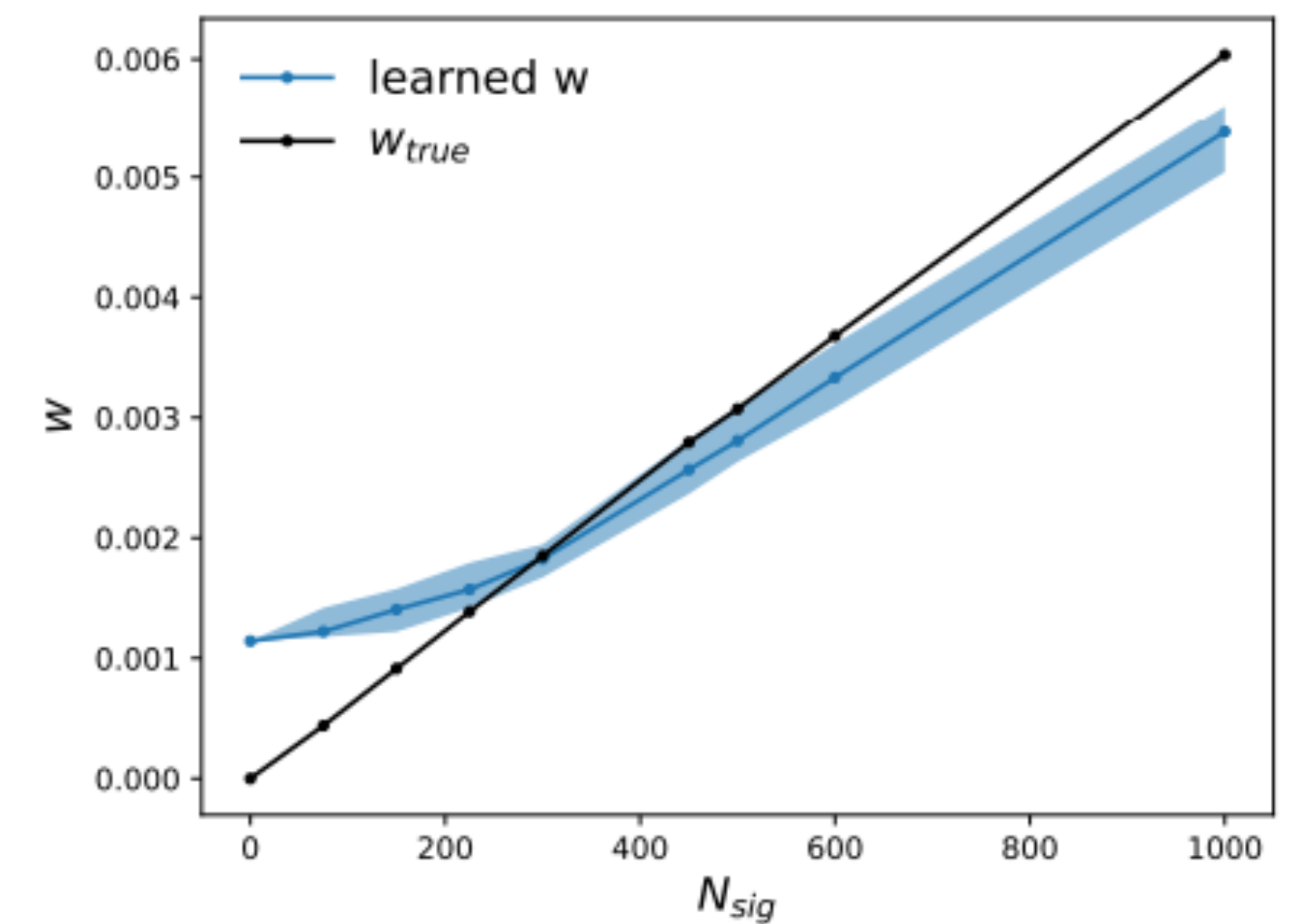
Interpretability with R-ANODE



$$p_{data}(x) = (1 - w)p_{bg}(x) + wp_{sig}(x)$$

In R-ANODE, fitting $p_{sig}(x)$ to the data in the signal region directly gives an empirical model for the signal. Can visualize this!

Can also directly infer the signal strength!

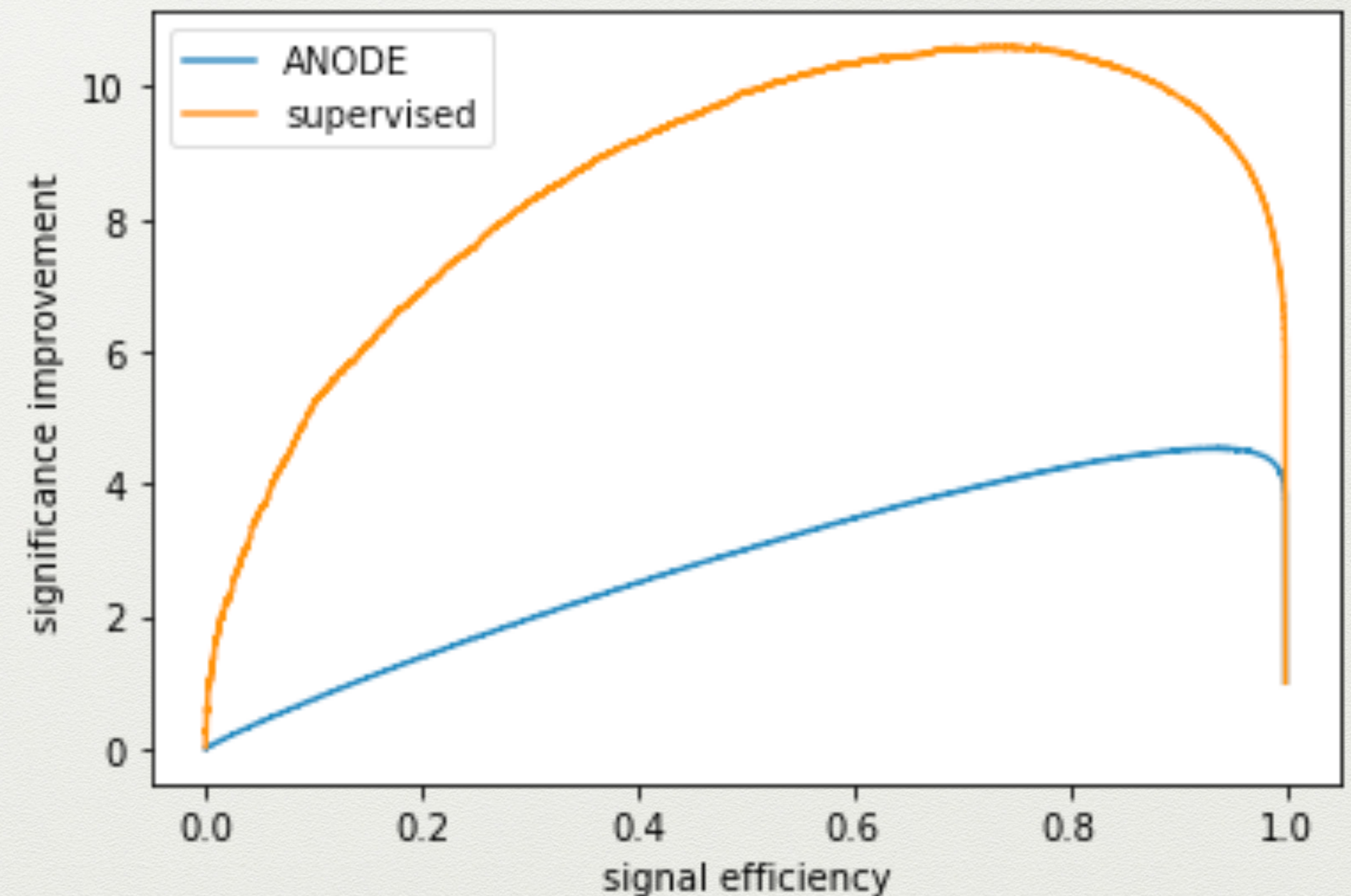


Is this the best one can do?

- For a given signal, the optimal classifier is:

the signal-to-background likelihood ratio (Neyman Pearson lemma):

$$R_{optimal}(x) = \frac{p_{sig}(x)}{p_{bg}(x)}$$

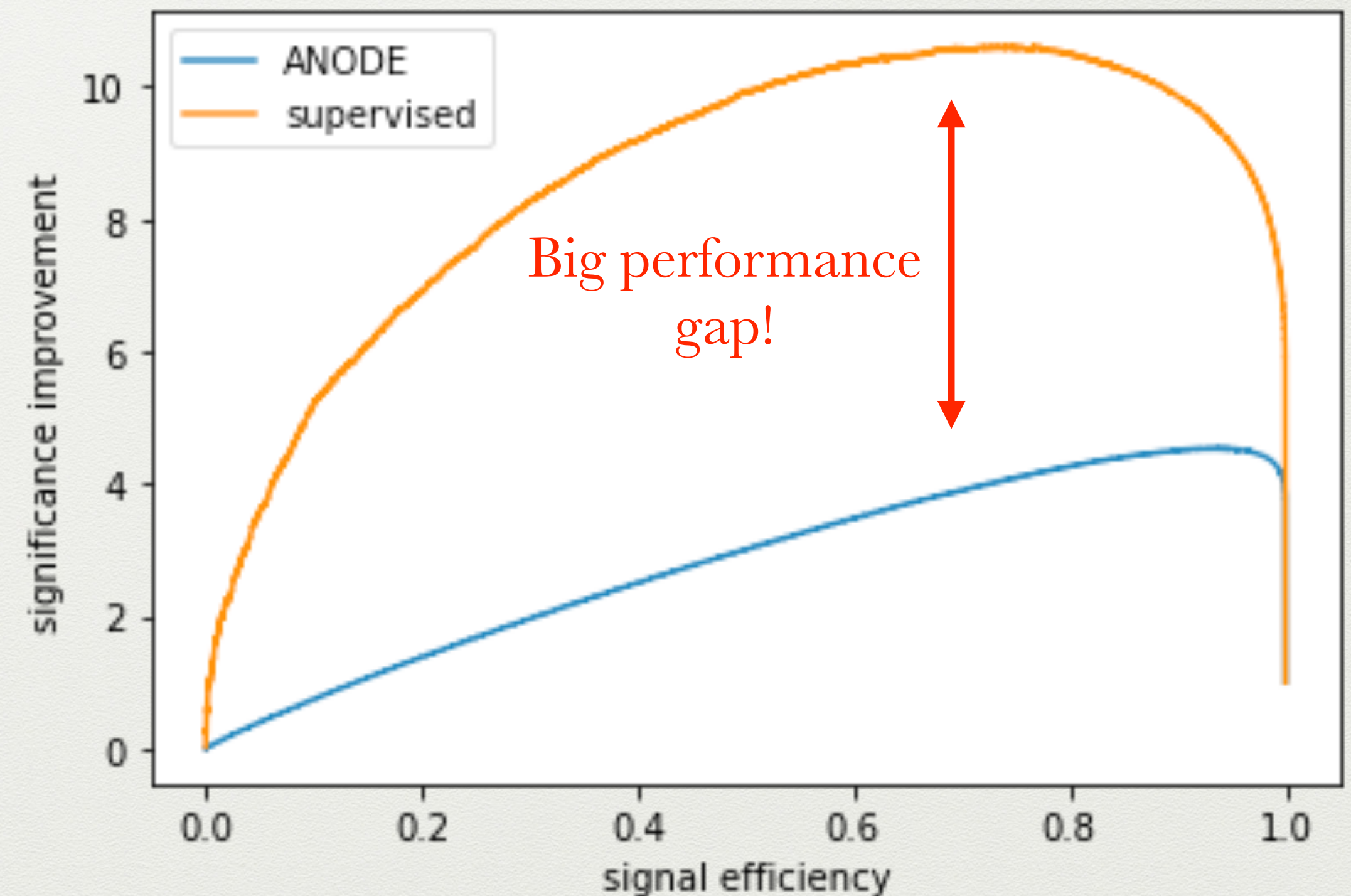


Is this the best one can do?

- For a given signal, the optimal classifier is:

the signal-to-background likelihood ratio (Neyman Pearson lemma):

$$R_{optimal}(x) = \frac{p_{sig}(x)}{p_{bg}(x)}$$



Is this the best one can do?

- Key realization: our data-to-background likelihood ratio should also be optimal!

$$p_{data}(x) = (1 - w) p_{bg}(x) + w p_{sig}(x)$$

$$\begin{aligned} R_{model-agnostic}(x) &= \frac{p_{data}(x)}{p_{bg}(x)} \\ &= (1 - w) + w \frac{p_{sig}(x)}{p_{bg}(x)} \\ &= const + w R_{optimal}(x) \end{aligned}$$

**optimal AND
model-agnostic
anomaly score
for any signal!**

Status of Nonresonant Anomaly Detection

- Nonresonant AD is highly motivated from many top-down scenarios such as SUSY
- But it is much harder — background estimation usually much more nontrivial than a bump hunt.
- So far only a handful of proofs-of-concept for complete analyses
- An interesting future direction!



Double decorrelated autoencoders

Mikuni, Nachman & **DS** 2111.06417

$$L[f_1, f_2, g_1, g_2] = \sum_i R_1(x_i)^2 + \sum_i R_2(x_i)^2 \\ + \lambda \text{DisCo}^2[R_1(X), R_2(X)]$$

“DisCo Decorrelation”

*Kasieczka & **DS** 2001.05310*

Szekely et al 0803.4101 et seq

Double decorrelated autoencoders

Mikuni, Nachman & **DS** 2111.06417

$$L[f_1, f_2, g_1, g_2] = \sum_i R_1(x_i)^2 + \sum_i R_2(x_i)^2 + \lambda \text{DisCo}^2[R_1(X), R_2(X)]$$

- Idea: Train *two* autoencoders and force them to be statistically independent of one another with a penalty term

“DisCo Decorrelation”
Kasieczka & **DS** 2001.05310

Szekely et al 0803.4101 et seq

Double decorrelated autoencoders

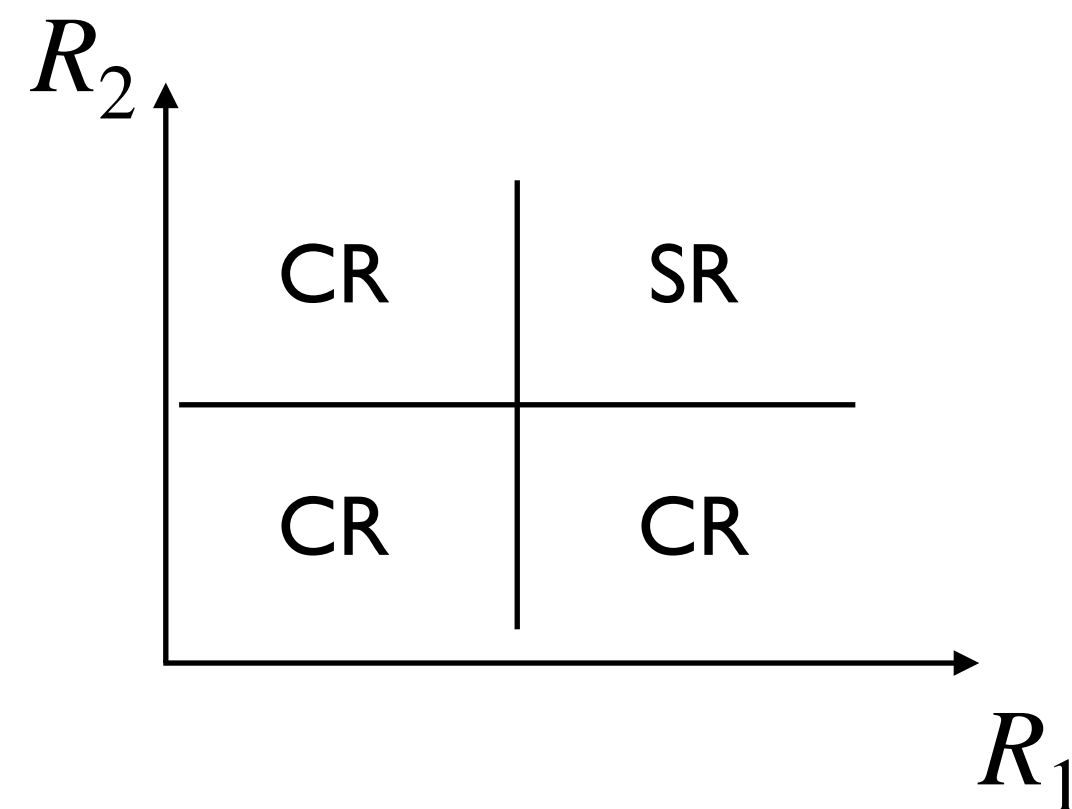
Mikuni, Nachman & **DS** 2111.06417

$$L[f_1, f_2, g_1, g_2] = \sum_i R_1(x_i)^2 + \sum_i R_2(x_i)^2 + \lambda \text{DisCo}^2[R_1(X), R_2(X)]$$

- Idea: Train *two* autoencoders and force them to be statistically independent of one another with a penalty term
- Use ABCD method for fully data-driven background estimation

“DisCo Decorrelation”
Kasieczka & **DS** 2001.05310

Szekely et al 0803.4101 et seq



$$N_{>, >}^{\text{predicted}}(\vec{c}) = \frac{N_{>, <}(\vec{c}) N_{<, >}(\vec{c})}{N_{<, <}(\vec{c})}$$

Double decorrelated autoencoders

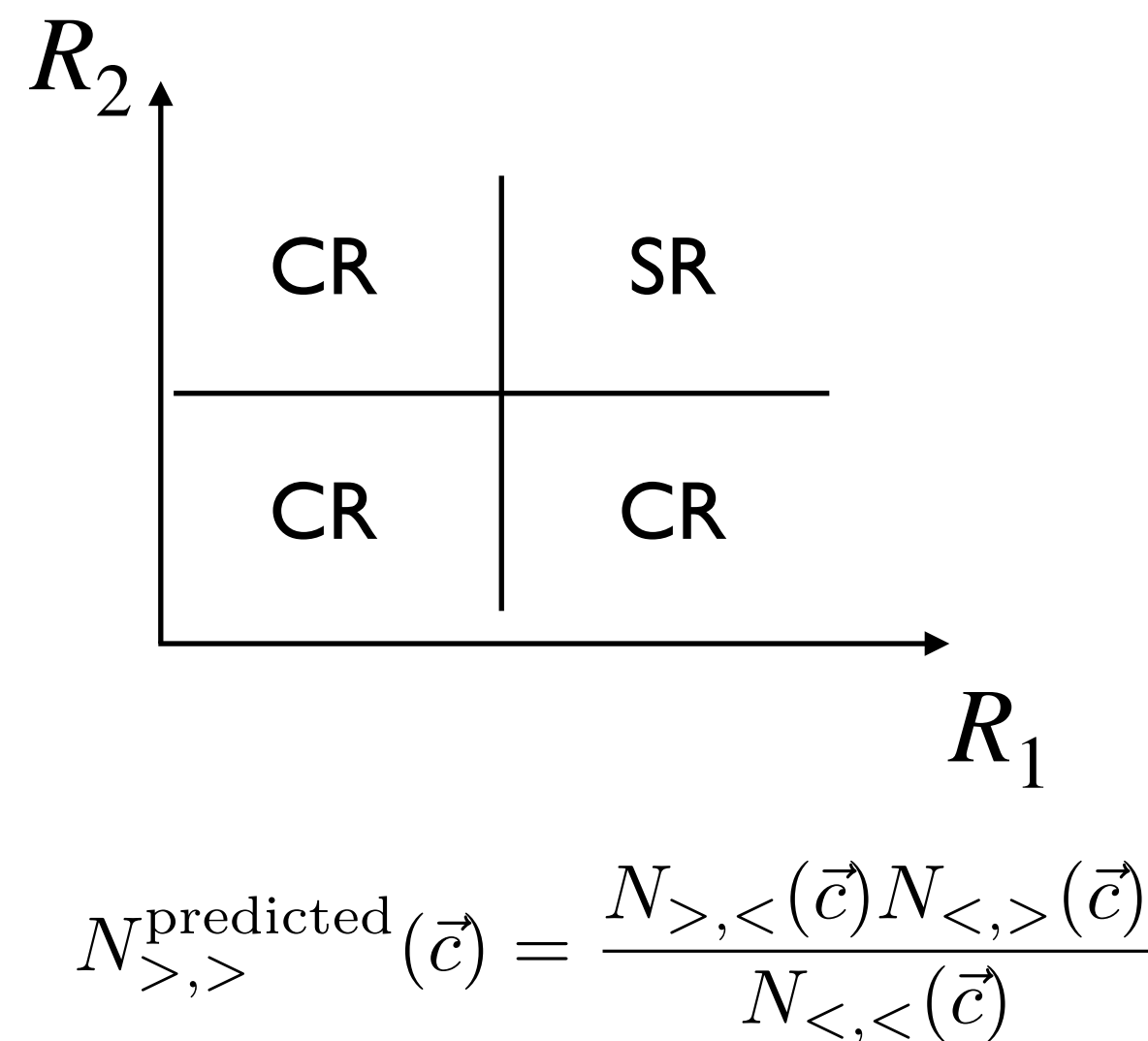
Mikuni, Nachman & **DS** 2111.06417

$$L[f_1, f_2, g_1, g_2] = \sum_i R_1(x_i)^2 + \sum_i R_2(x_i)^2 + \lambda \text{DisCo}^2[R_1(X), R_2(X)]$$

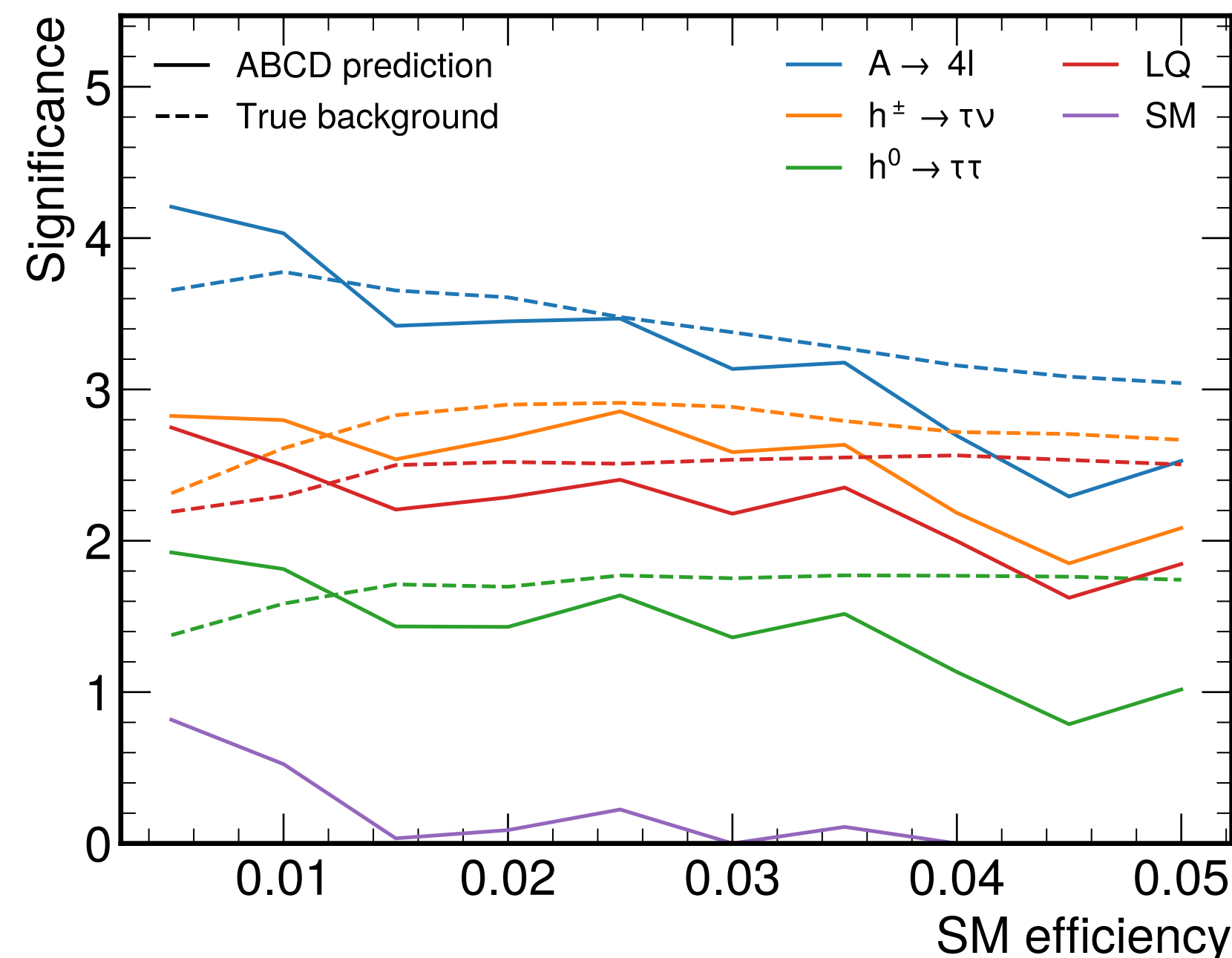
- Idea: Train *two* autoencoders and force them to be statistically independent of one another with a penalty term
- Use ABCD method for fully data-driven background estimation

“DisCo Decorrelation”
Kasieczka & **DS** 2001.05310

Szekely et al 0803.4101 et seq



The method works!



Double decorrelated autoencoders

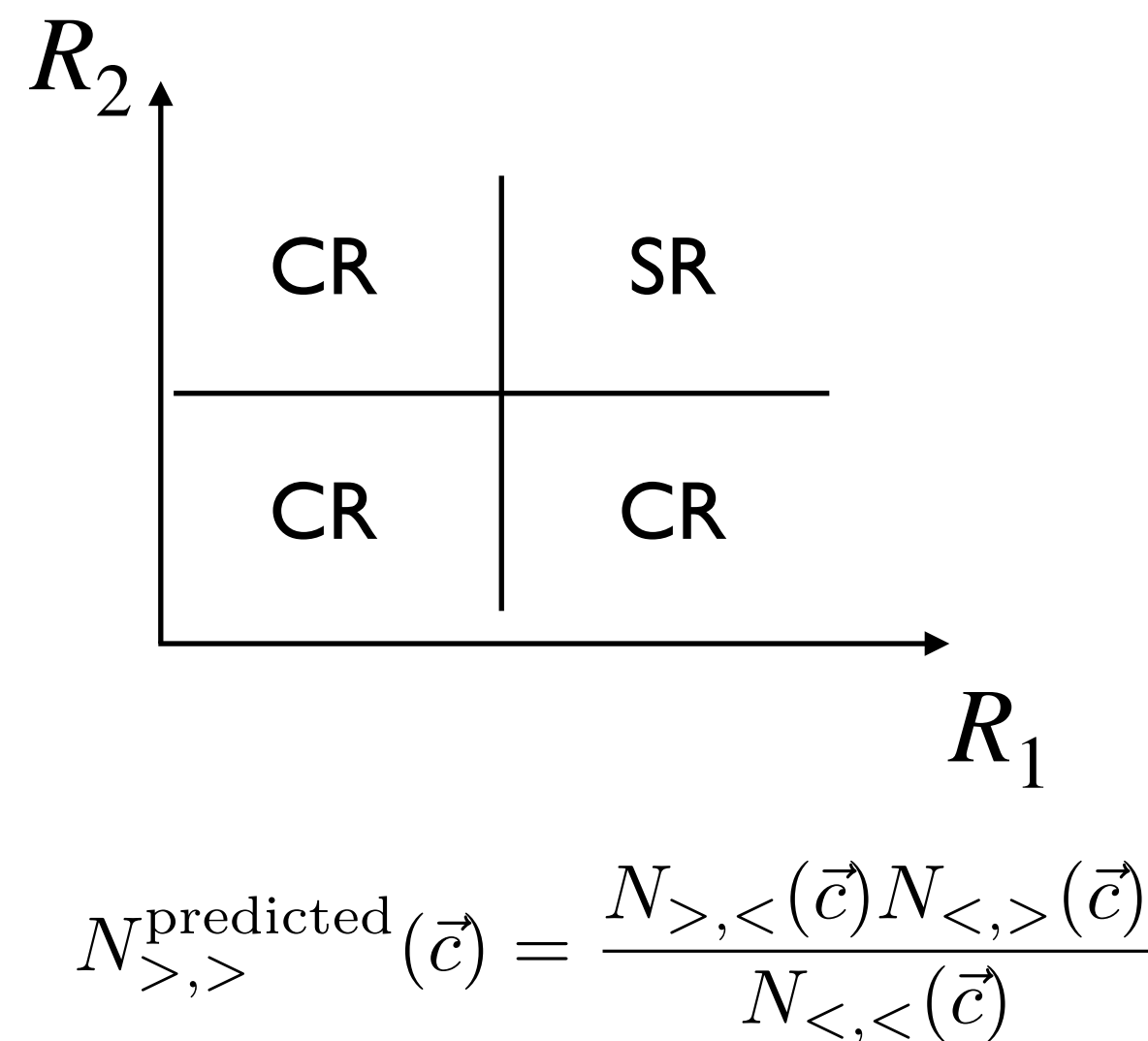
Mikuni, Nachman & **DS** 2111.06417

$$L[f_1, f_2, g_1, g_2] = \sum_i R_1(x_i)^2 + \sum_i R_2(x_i)^2 + \lambda \text{DisCo}^2[R_1(X), R_2(X)]$$

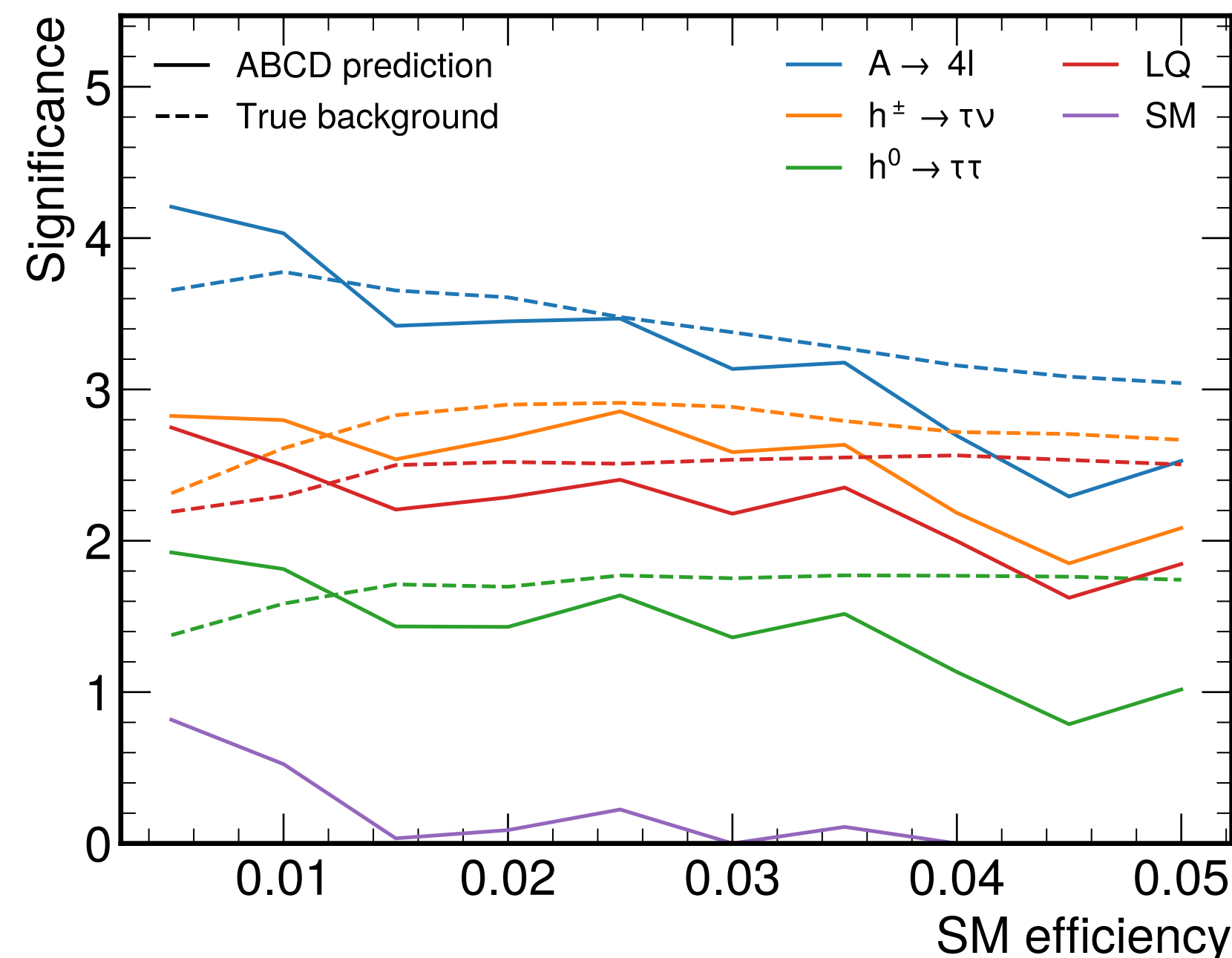
“DisCo Decorrelation”
Kasieczka & **DS** 2001.05310

Szekely et al 0803.4101 et seq

- Idea: Train *two* autoencoders and force them to be statistically independent of one another with a penalty term
- Use ABCD method for fully data-driven background estimation



The method works!

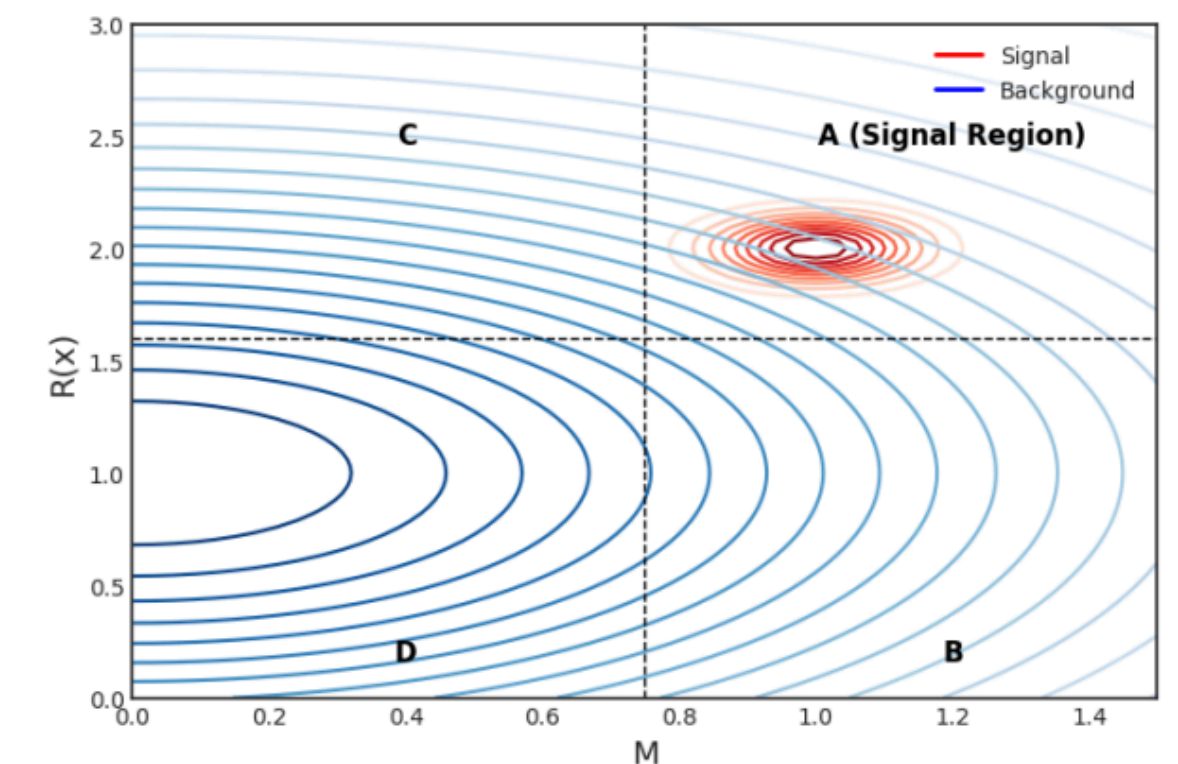


**Can also be used online
as an anomaly trigger**

Complete Optimal Non-Resonant AD (CONRAD)

Kasieczka, Raine, **DS** & Upadhyay 2404.07258

- Very preliminary proof-of-concept for non-resonant AD in jets + MET
 - Assume main background is $Z(\nu\nu)$ +MET for simplicity. γ +MET is a data-driven control sample.
 - Train normalizing flow on γ +MET to learn overdensity score $R(z)$ decorrelated with MET.
 - Estimate backgrounds with ABCD method between $R(z)$ and MET.

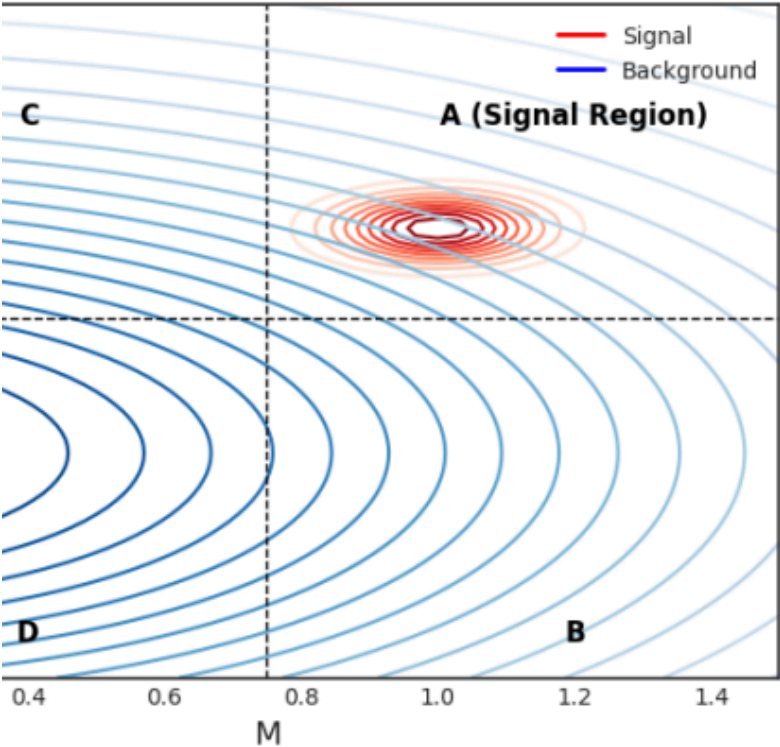
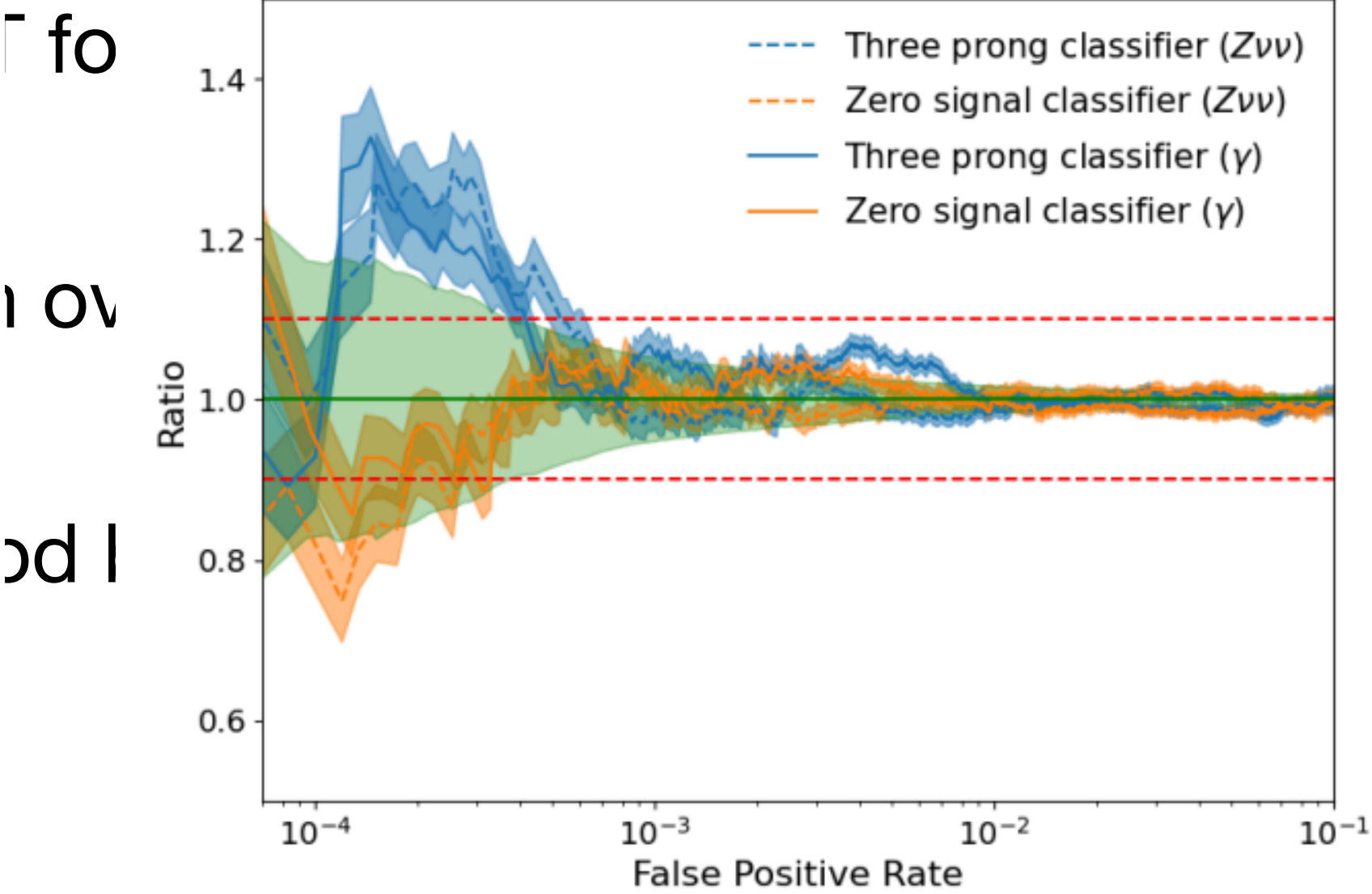
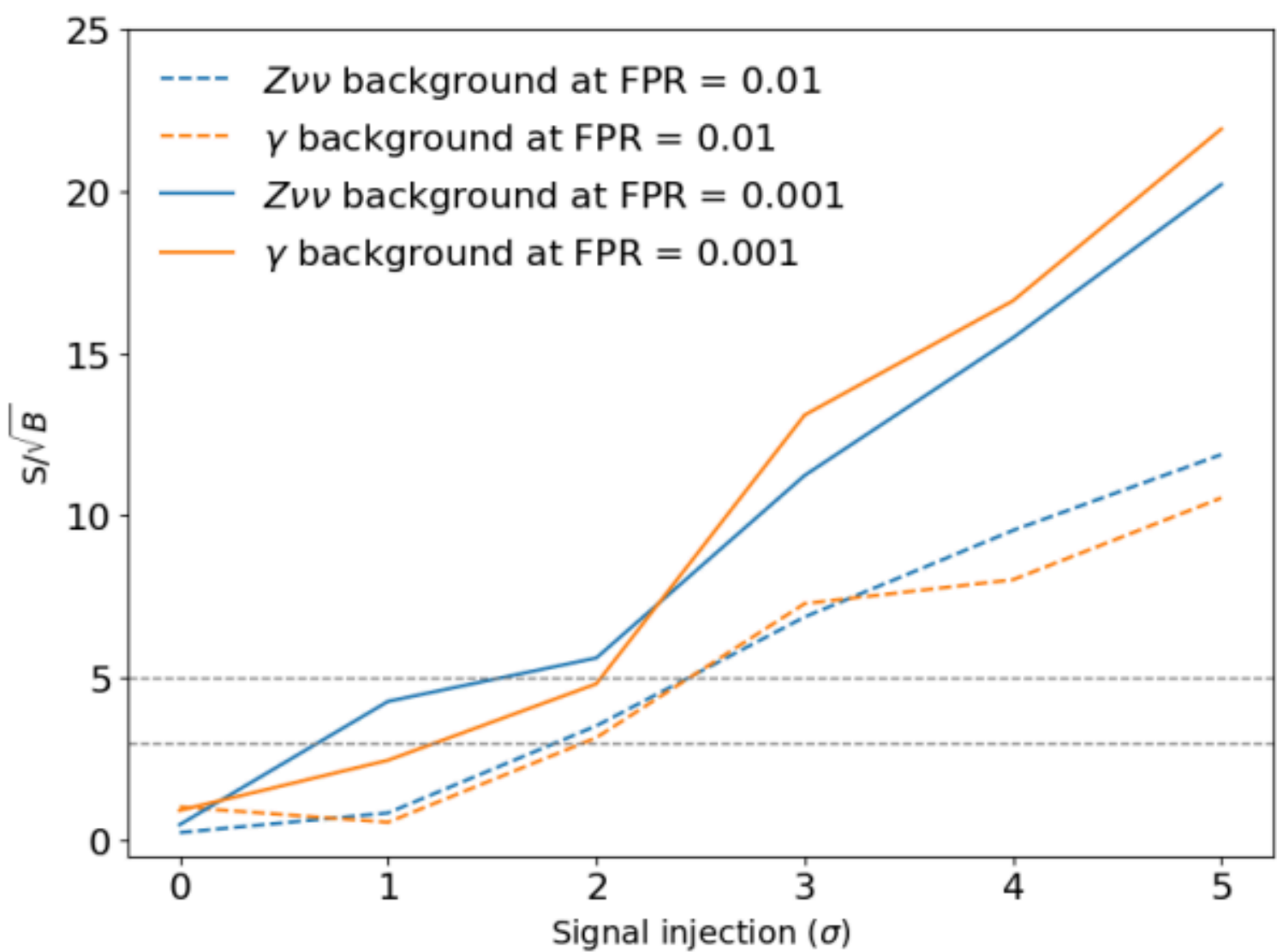


Complete Optimal Non-Resonant AD (CONRAD)

Kasieczka, Raine, **DS** & Upadhyay 2404.07258

- Very preliminary proof-of-concept for non-resonant AD in jets + MET

- Assumed data-c
- Train r decor
- Estima



$$Z' \rightarrow X(qqq)Y(inv)$$

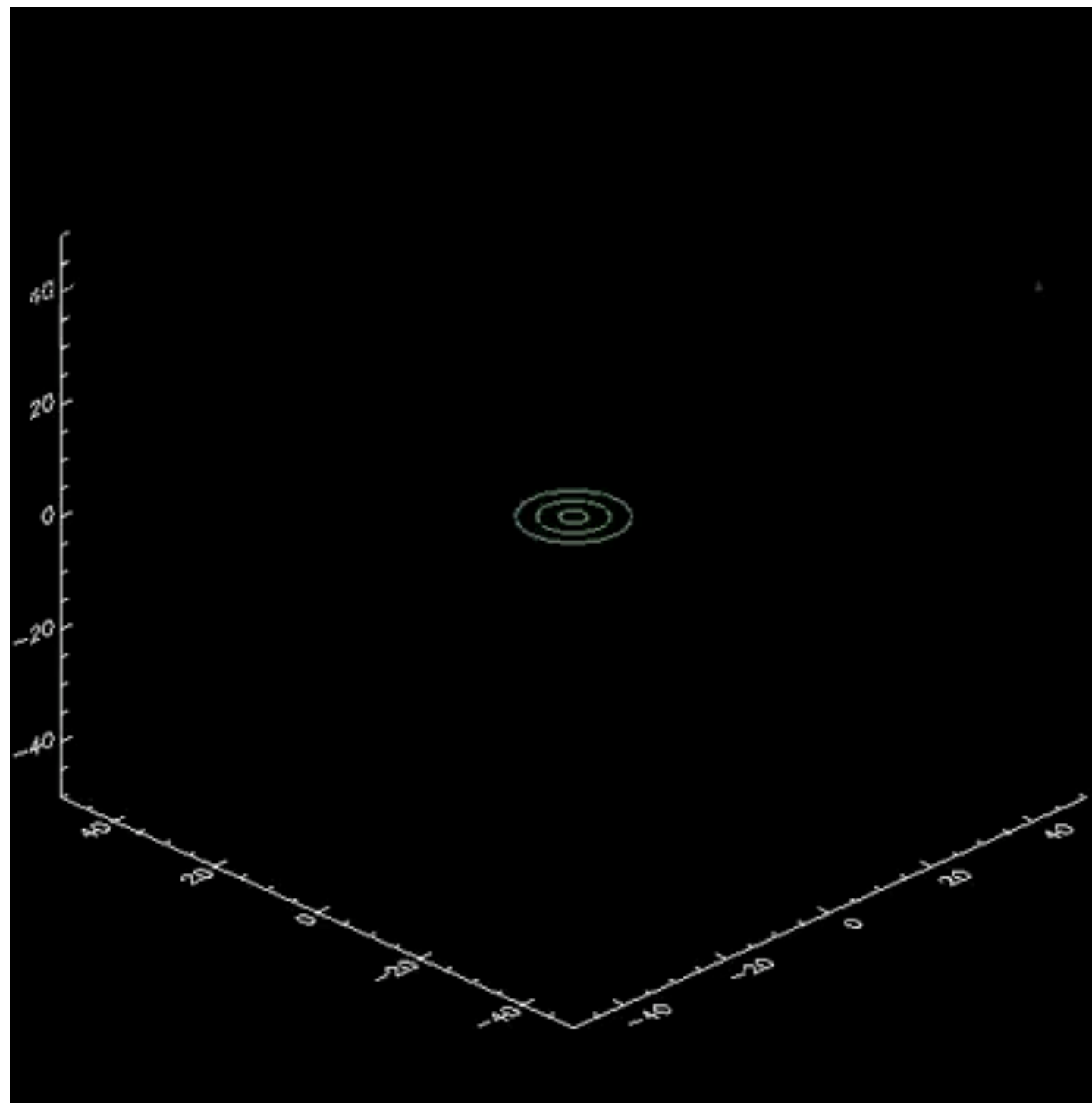
$$x = (m_J, \tau_{21}, \tau_{32})$$

Can achieve good signal sensitivity and ABCD closure for a benchmark signal!

From LHC to Astro: Search for Stellar Streams

The same resonant AD techniques can be used to search for **stellar streams** in **astro data**

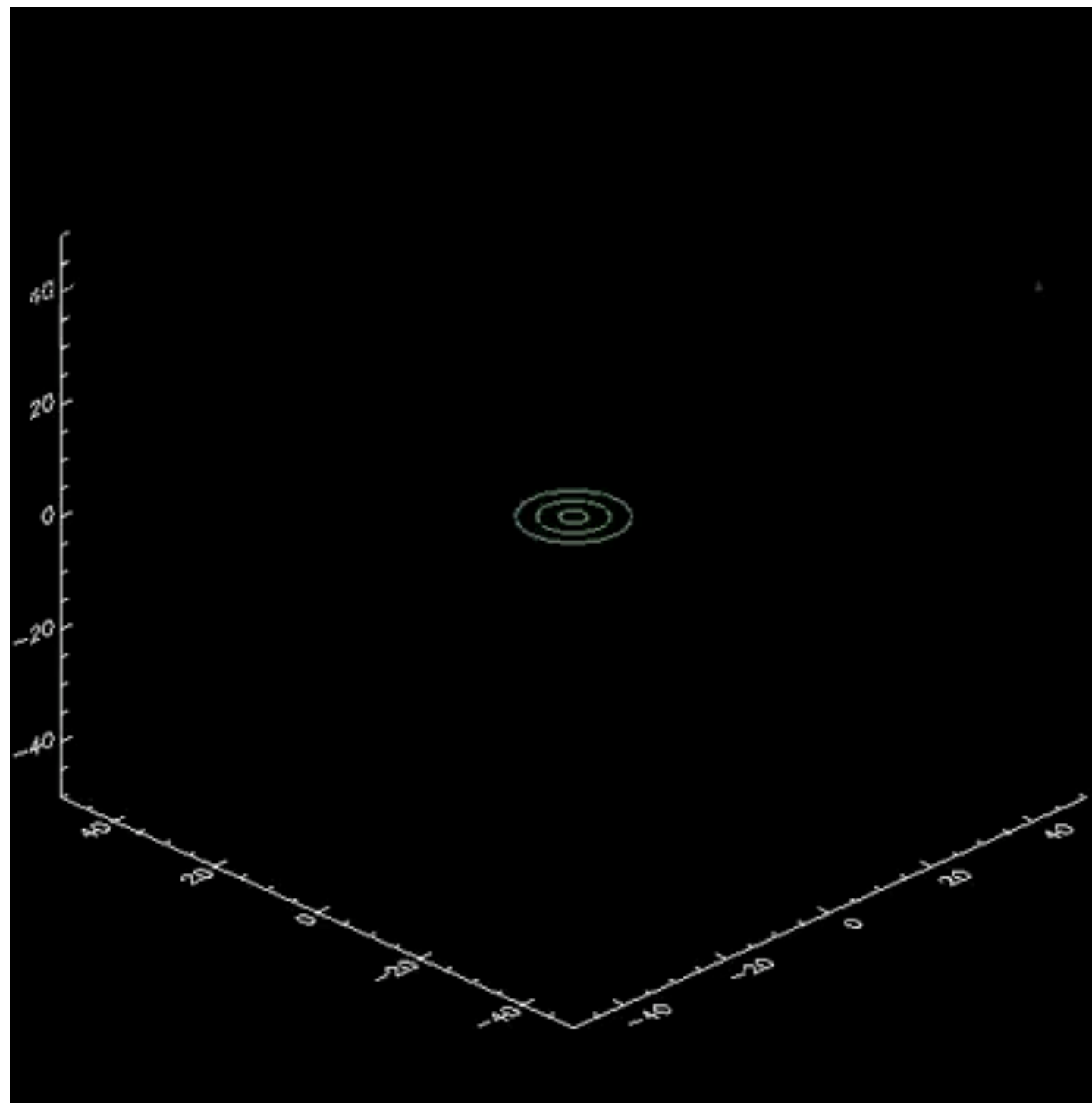
Stellar streams are tidally disrupted dwarf galaxies and globular clusters



From LHC to Astro: Search for Stellar Streams

The same resonant AD techniques can be used to search for **stellar streams** in **astro data**

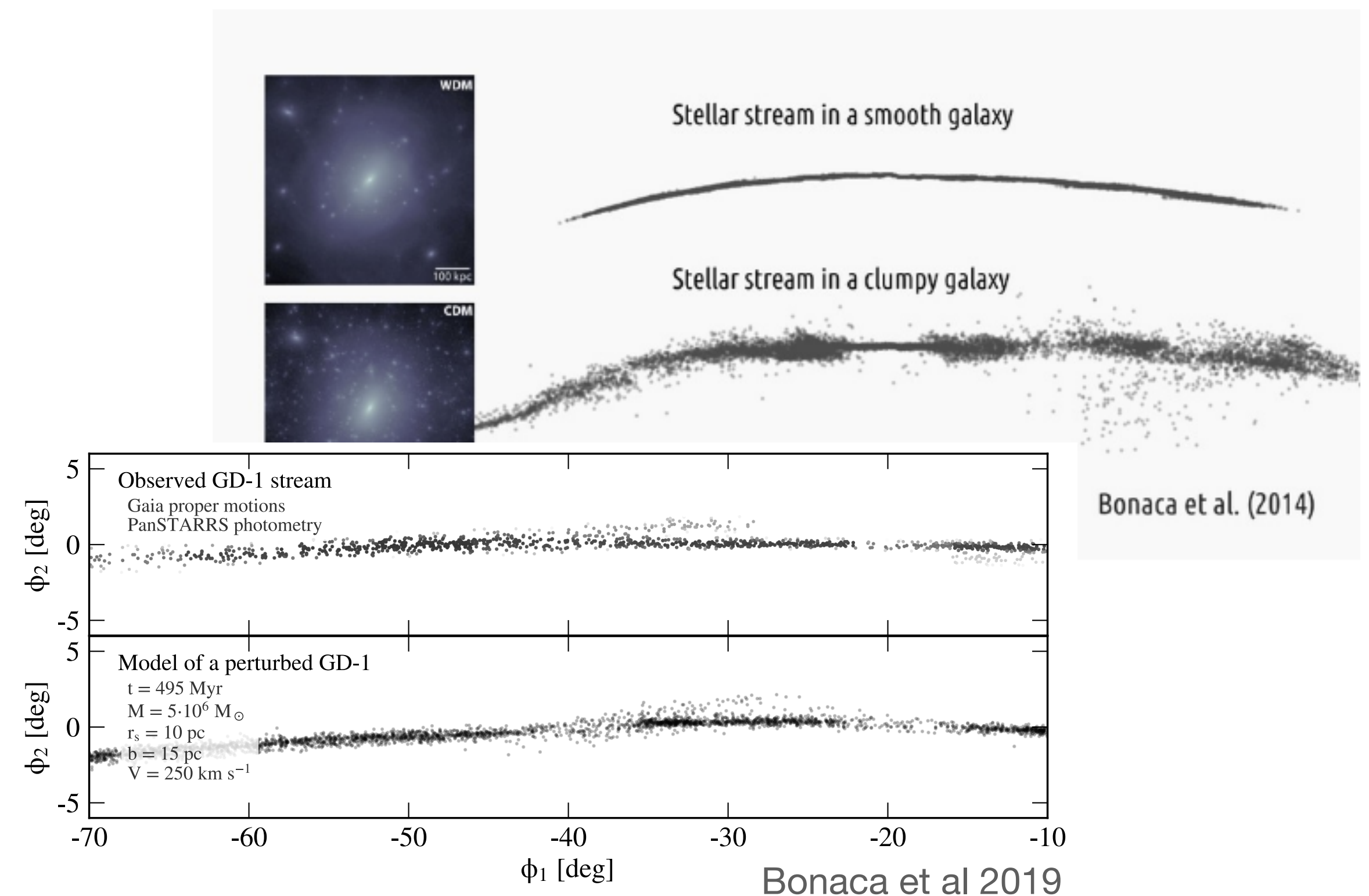
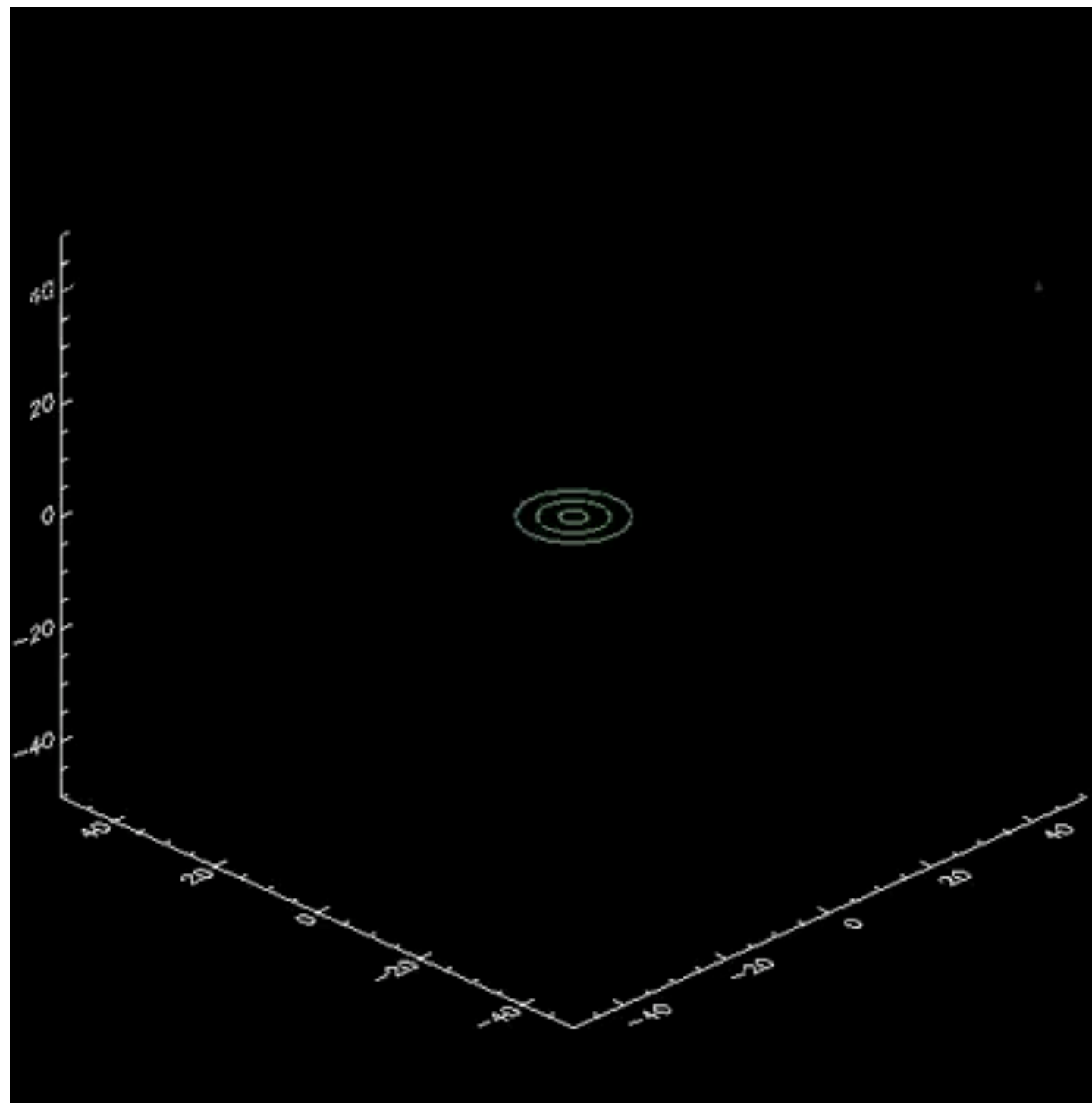
Stellar streams are tidally disrupted dwarf galaxies and globular clusters



From LHC to Astro: Search for Stellar Streams

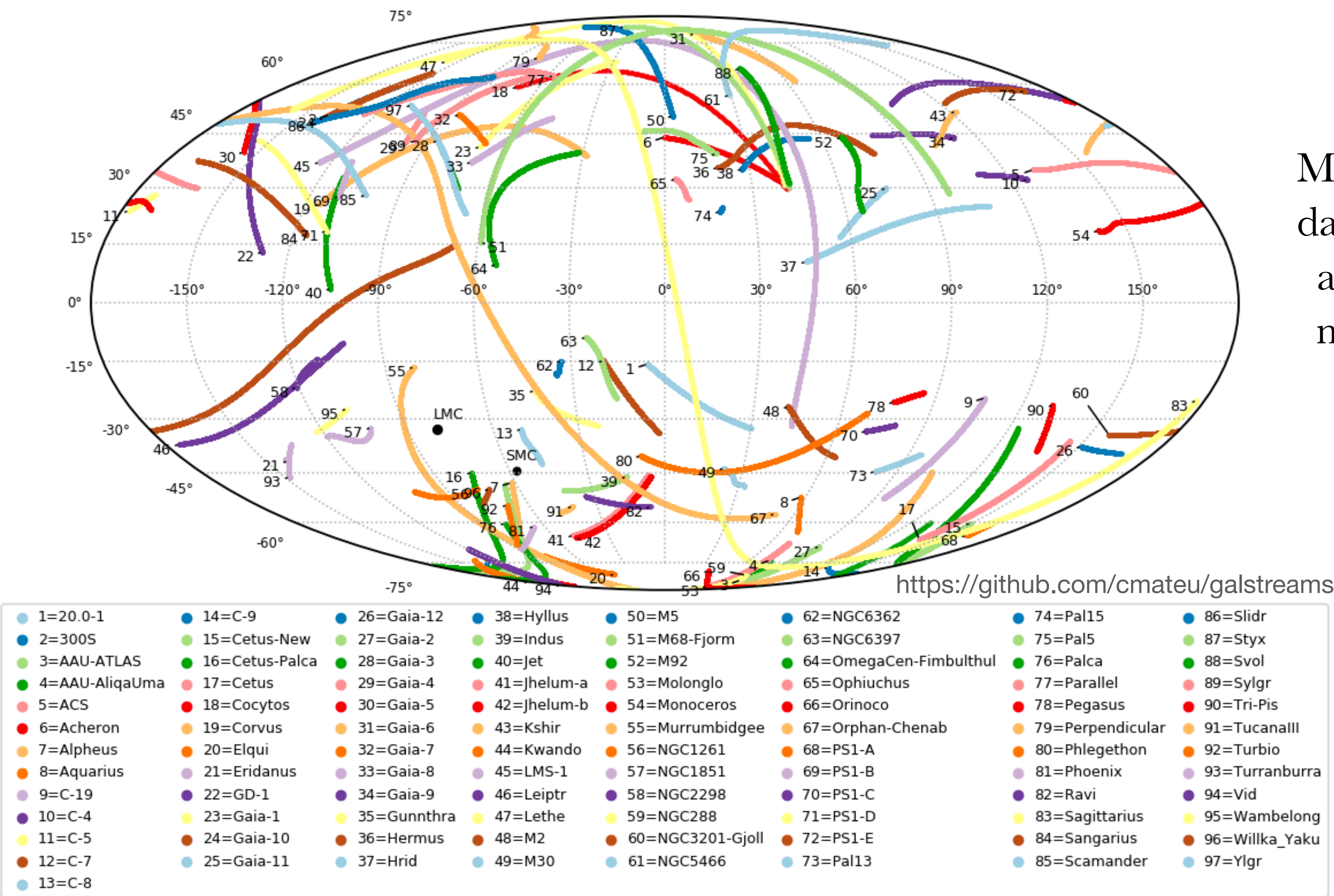
The same resonant AD techniques can be used to search for **stellar streams** in **astro data**

Stellar streams are tidally disrupted dwarf galaxies and globular clusters

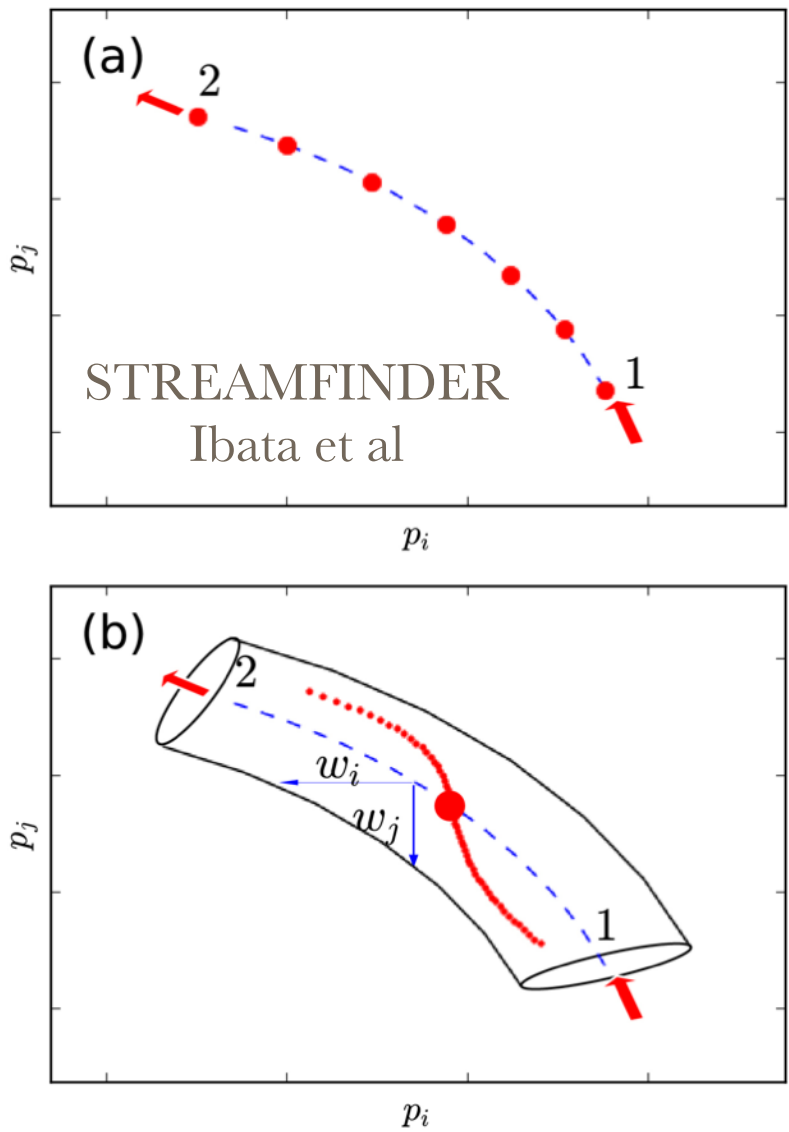


Streams are potentially unique astrophysical probes of dark matter

Known stellar streams of the Milky Way

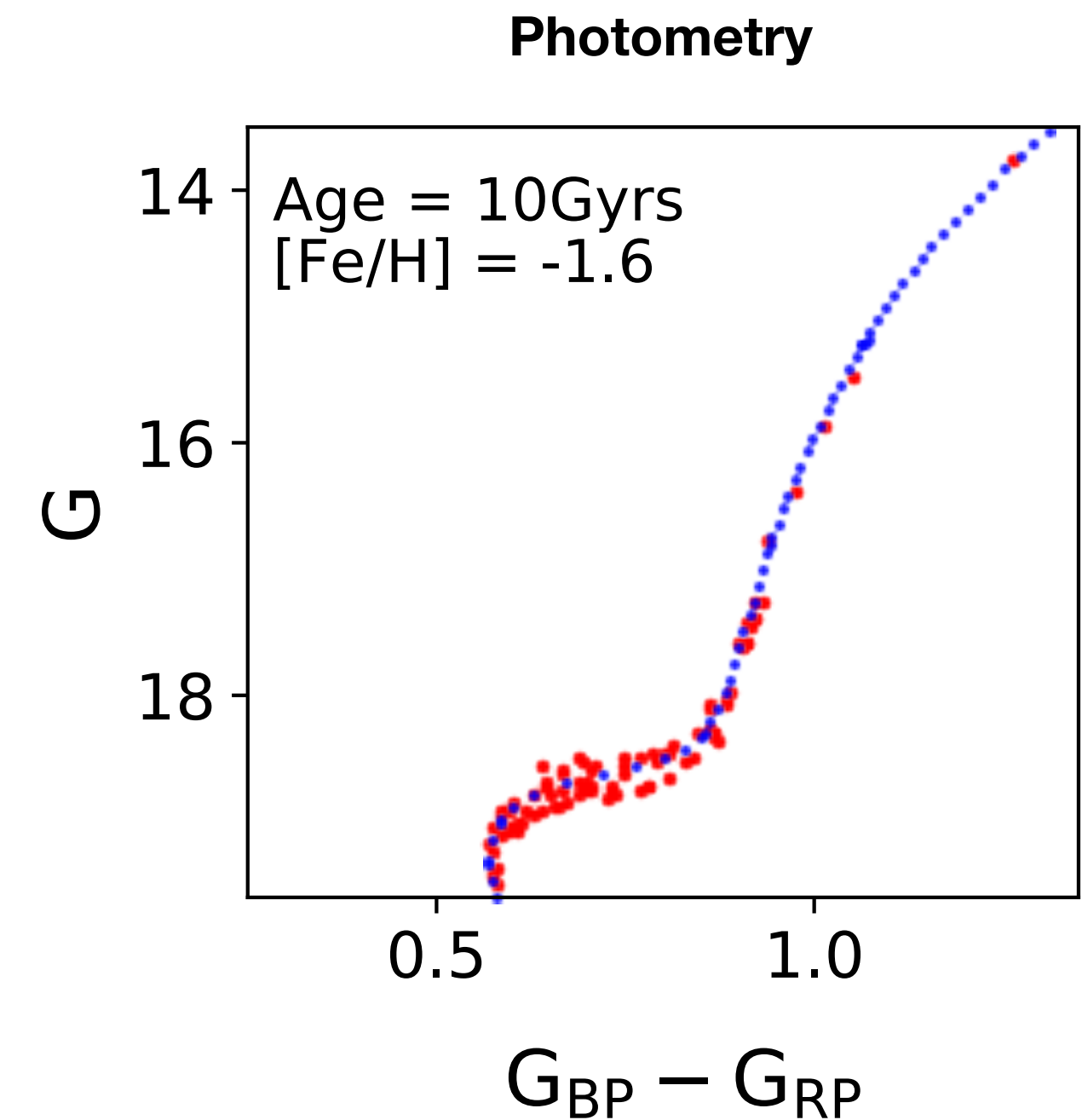
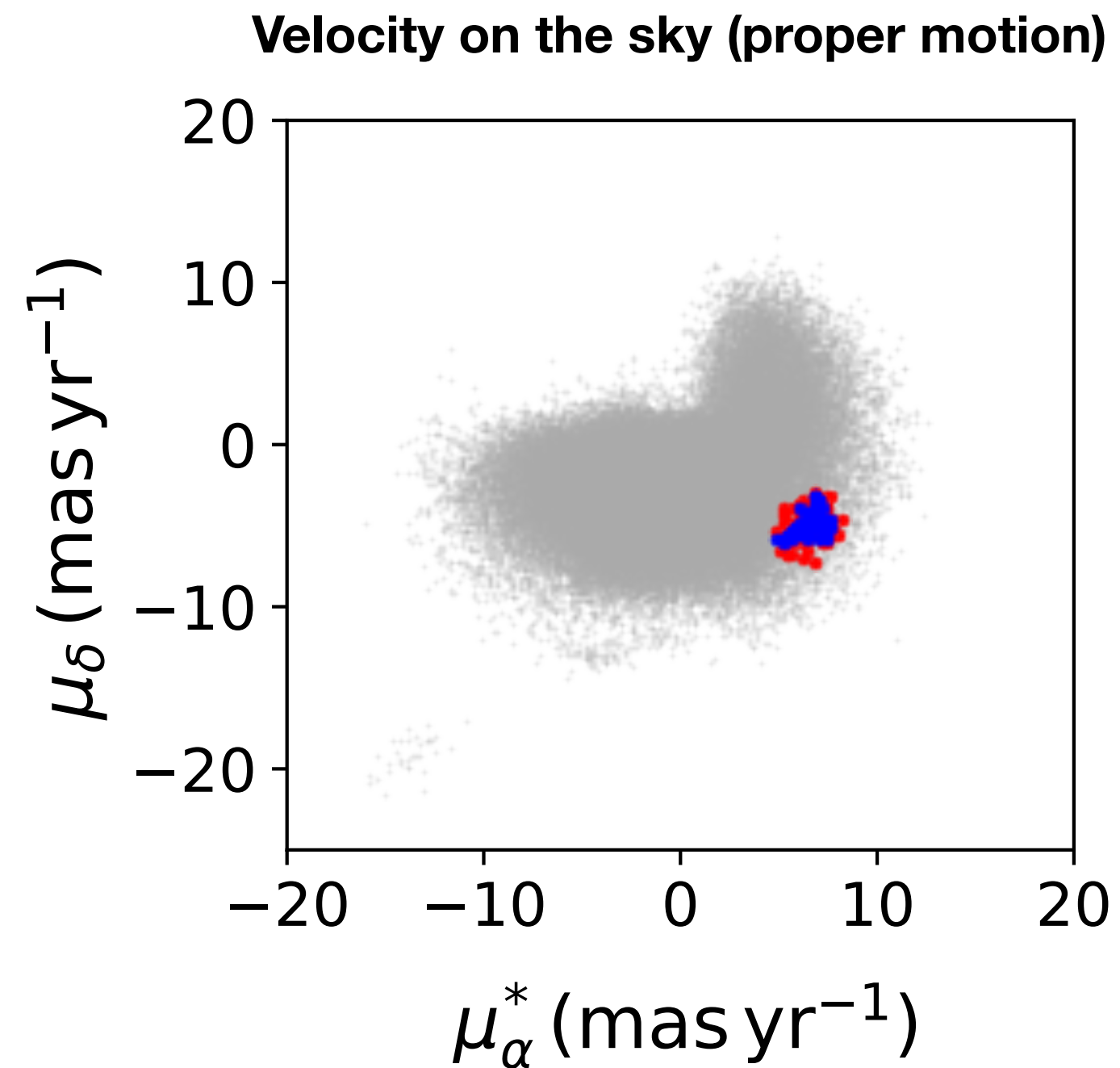
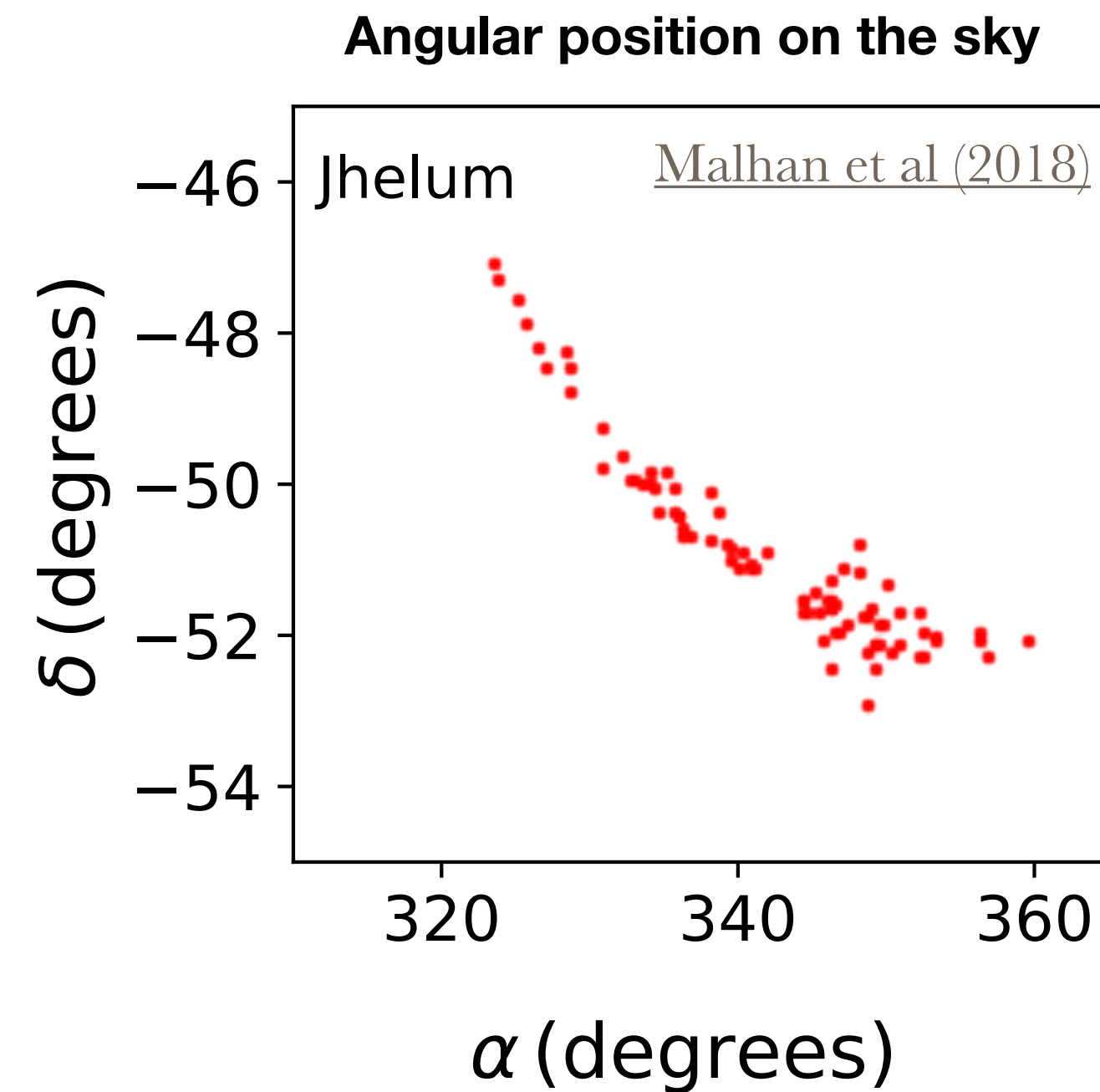


Many new streams found with data from **Gaia survey** using automated orbit-integration methods (model-dependent)



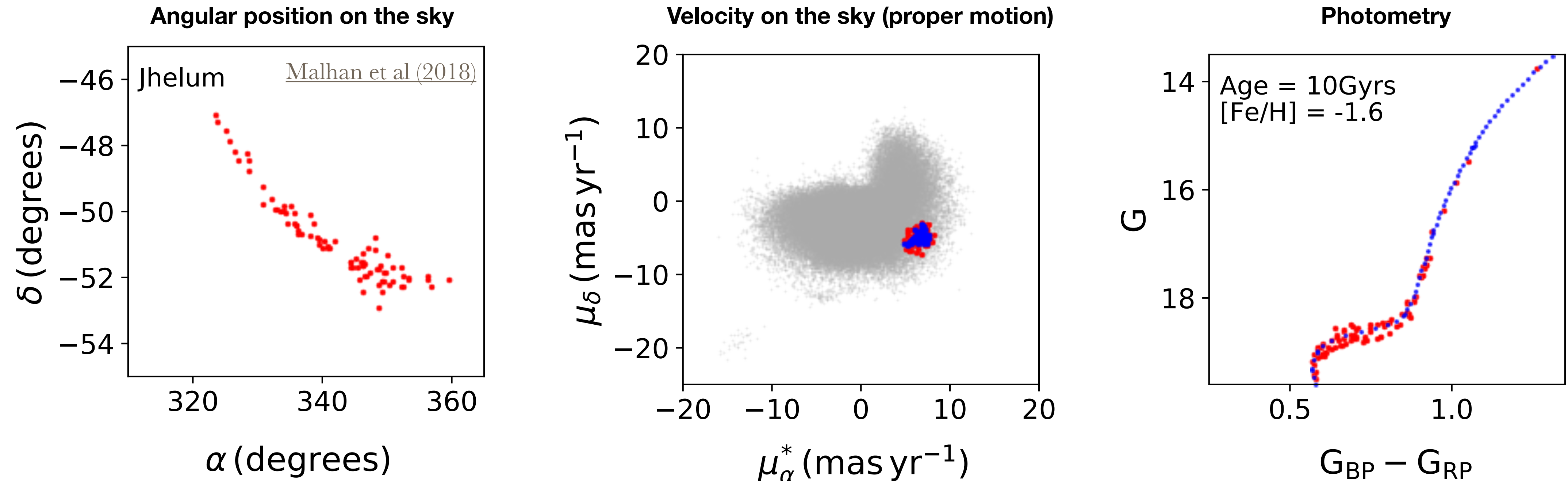
Via Machinae: Model-agnostic stream search

[DS, Buckley, Necib '23] [DS, Buckley, Necib, Tamasas '21]



Via Machinae: Model-agnostic stream search

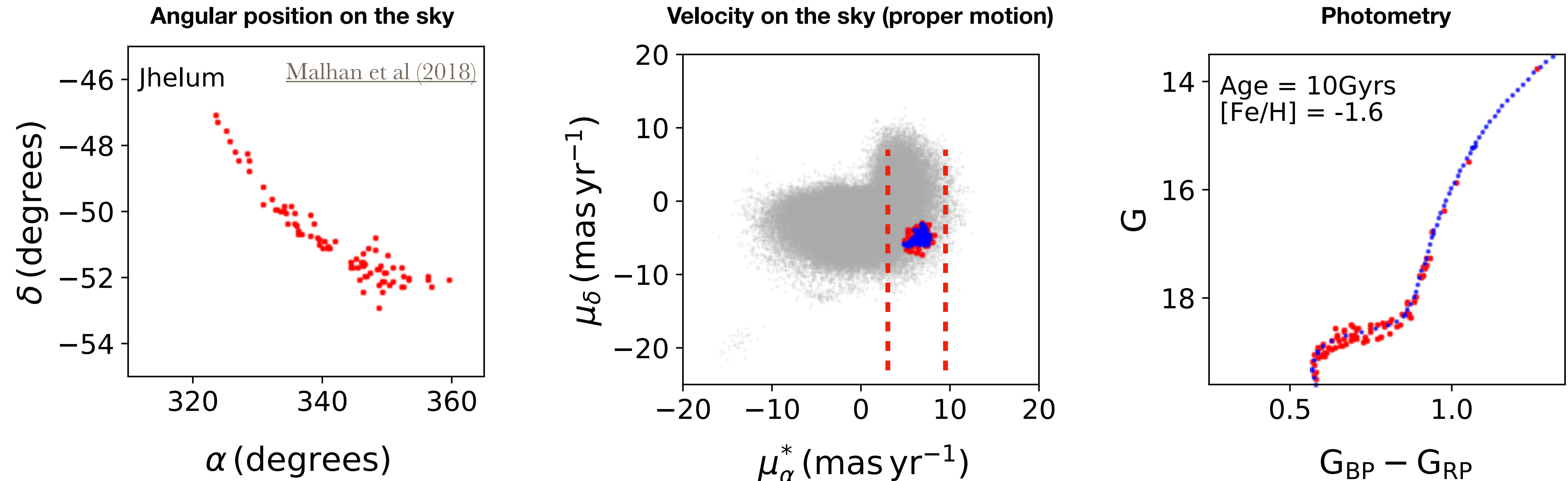
[DS, Buckley, Necib '23] [DS, Buckley, Necib, Tamasas '21]



- Streams are local overdensities in multiple features — ideal for **enhanced bump hunt methods!**

Via Machinae: Model-agnostic stream search

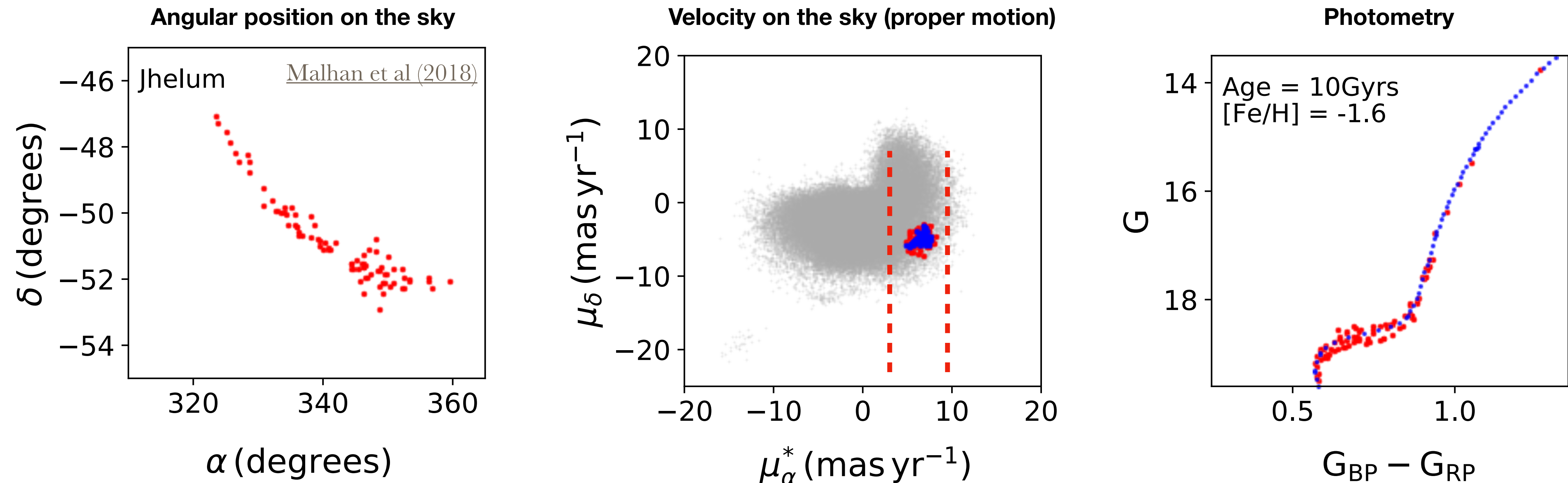
[DS, Buckley, Necib '23] [DS, Buckley, Necib, Tamasas '21]



- Streams are local overdensities in multiple features — ideal for **enhanced bump hunt methods!**
- Choose either proper motion coordinate as resonant feature

Via Machinae: Model-agnostic stream search

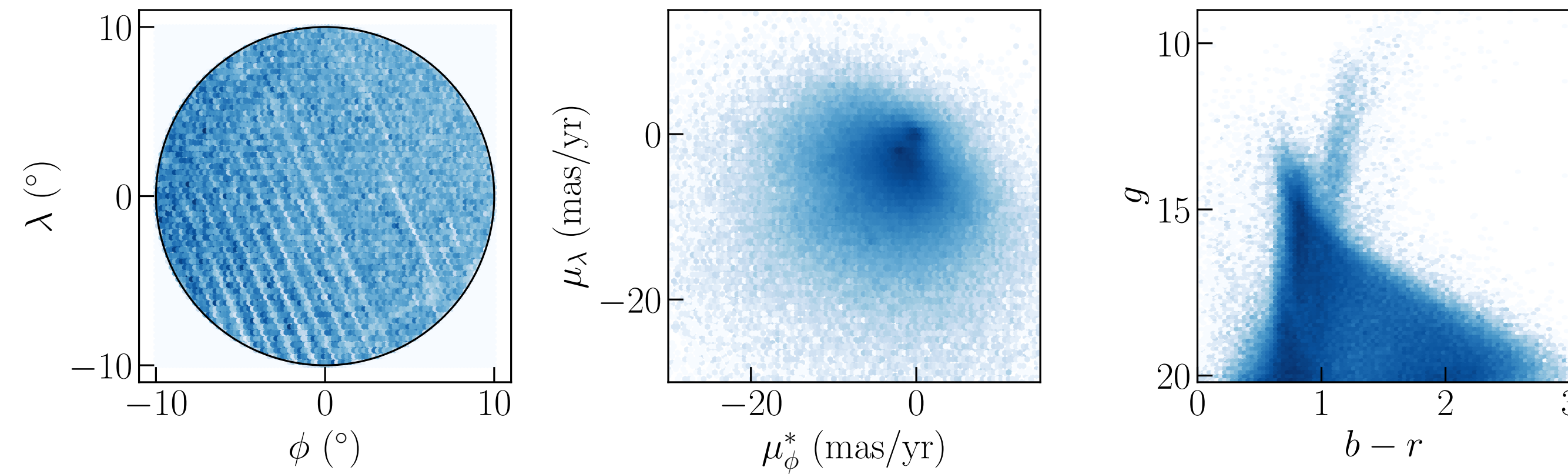
[DS, Buckley, Necib '23] [DS, Buckley, Necib, Tamasas '21]



- Streams are local overdensities in multiple features — ideal for **enhanced bump hunt methods!**
- Choose either proper motion coordinate as resonant feature
- Learn overdensity anomaly score (using normalizing flows) with remaining five features

Core method — illustrated with GD-1 stream

[DS, Buckley, Necib, Tamasas '21]

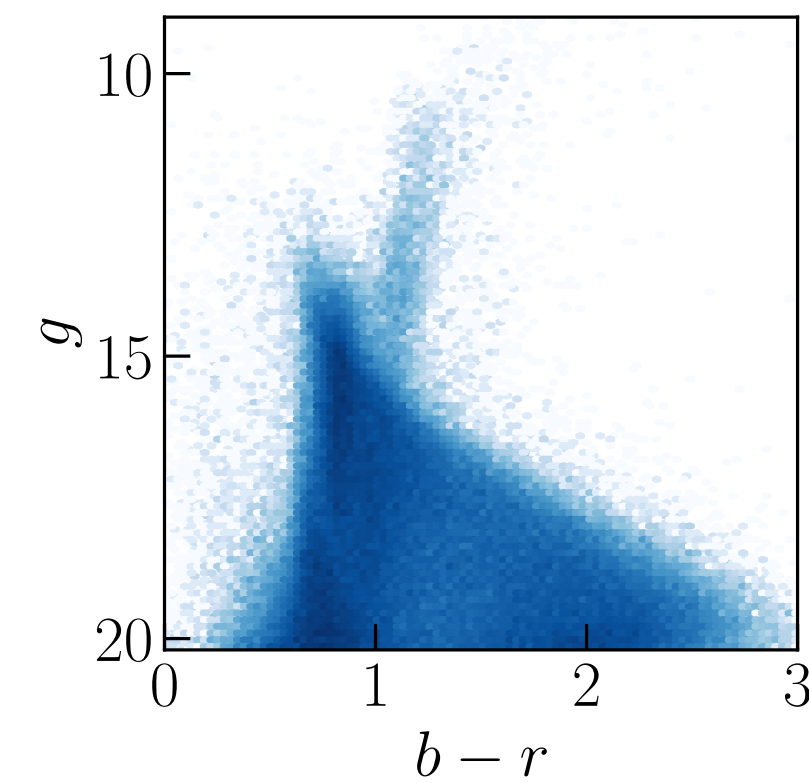
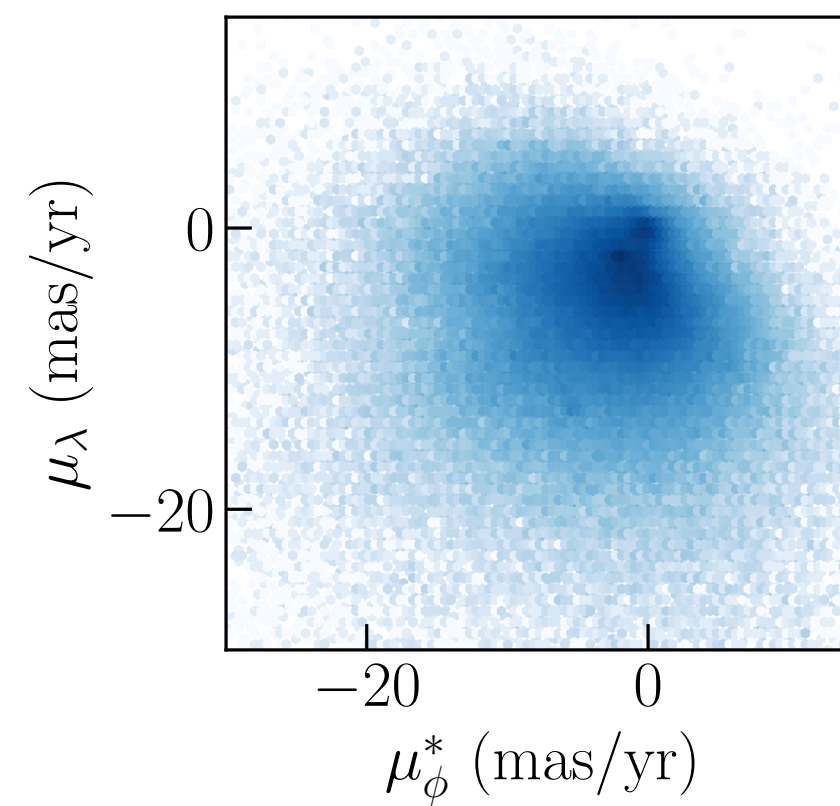
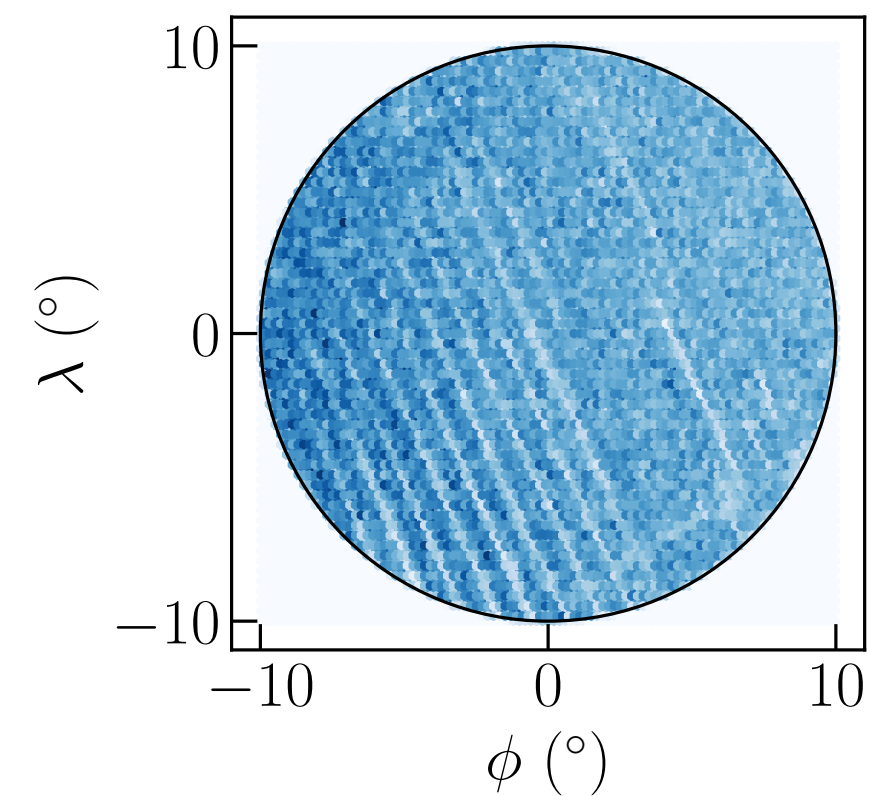


All stars in a patch of the sky
containing (part of) GD-1
(ra,dec)=(148.6,24.2)

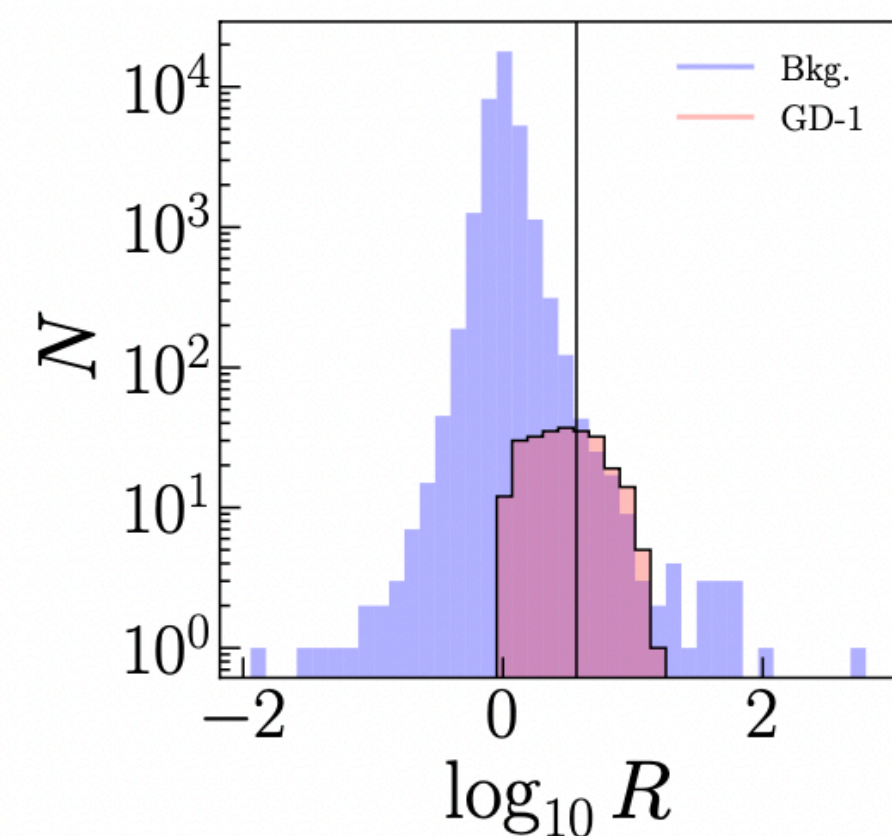
Core method — illustrated with GD-1 stream

[DS, Buckley, Necib, Tamasas '21]

Train ANODE on a
proper motion window
containing GD-1



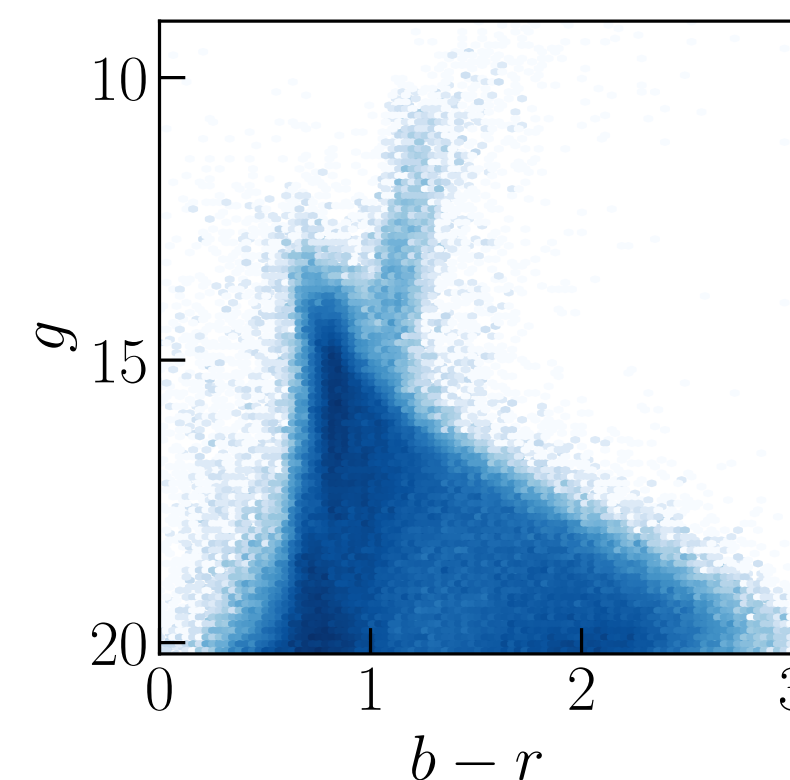
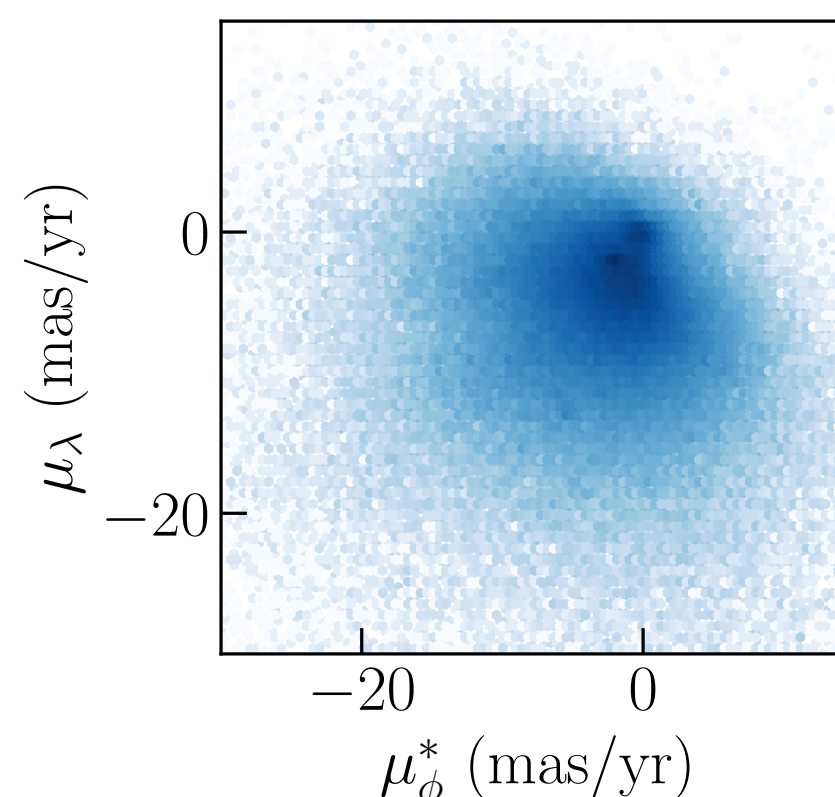
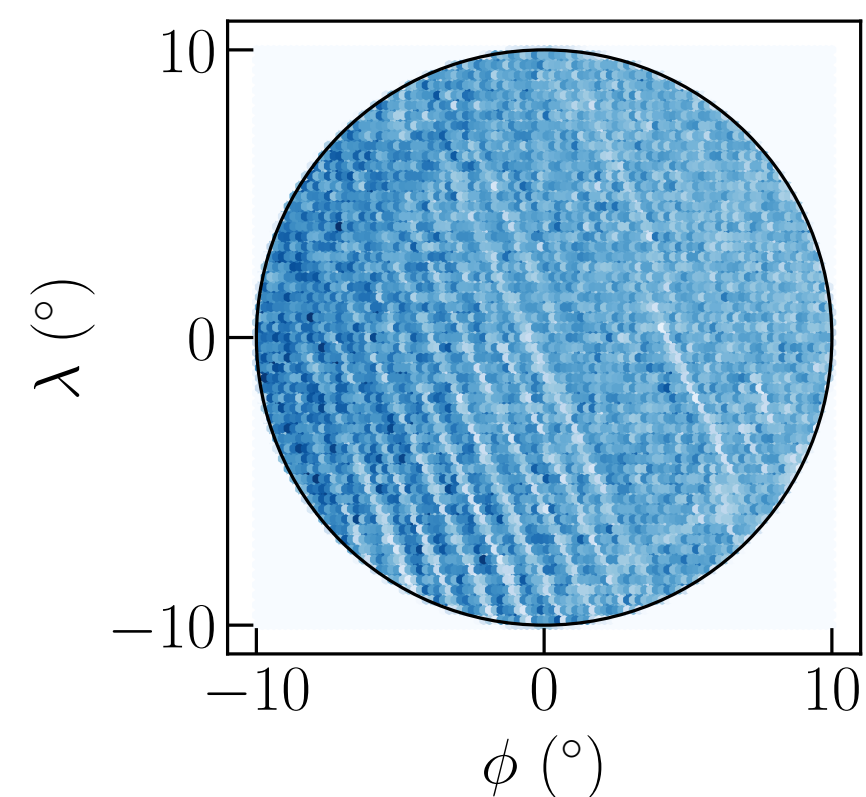
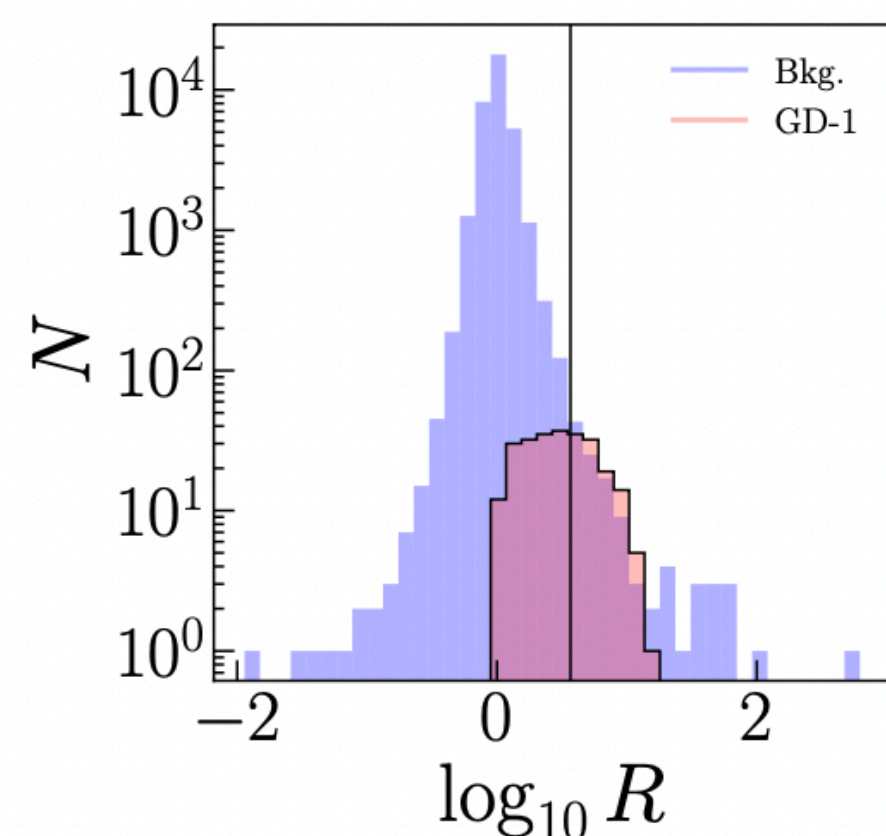
All stars in a patch of the sky
containing (part of) GD-1
(ra,dec)=(148.6,24.2)



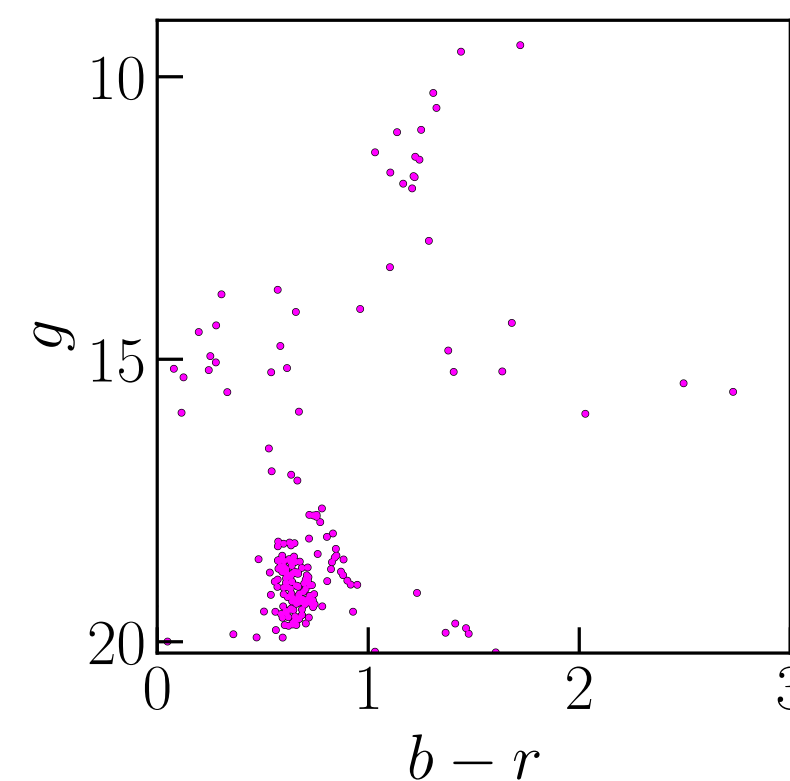
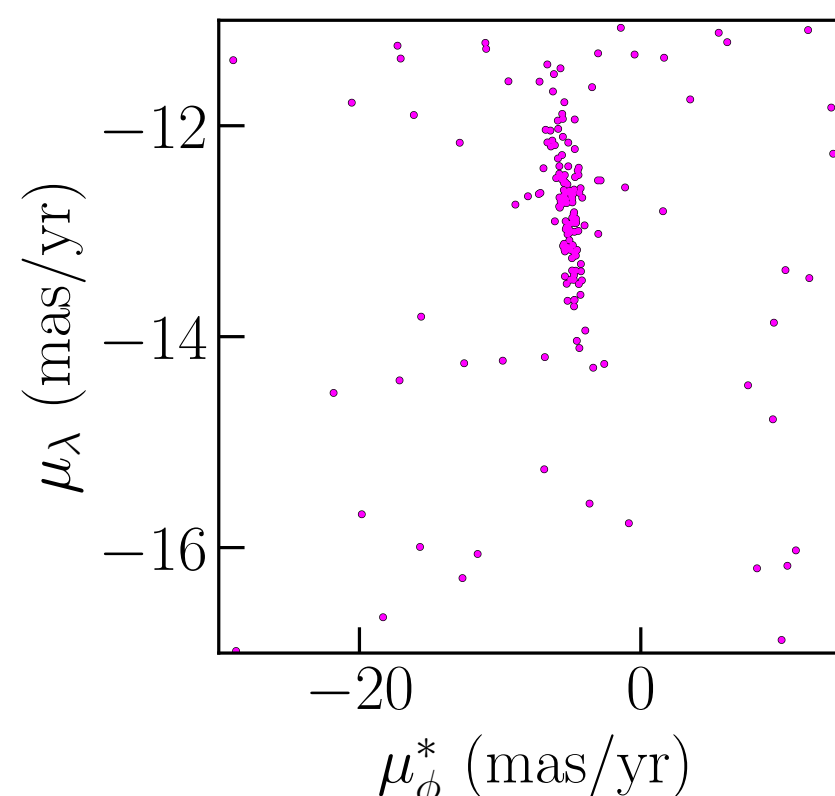
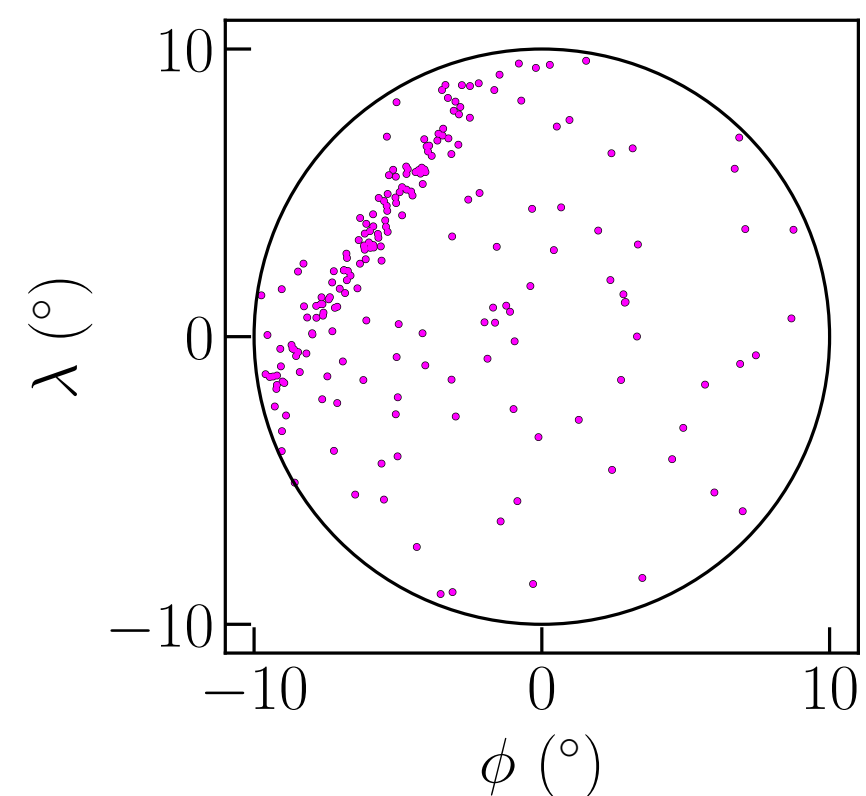
Core method — illustrated with GD-1 stream

[DS, Buckley, Necib, Tamasas '21]

Train ANODE on a proper motion window containing GD-1



All stars in a patch of the sky containing (part of) GD-1
(ra,dec)=(148.6,24.2)

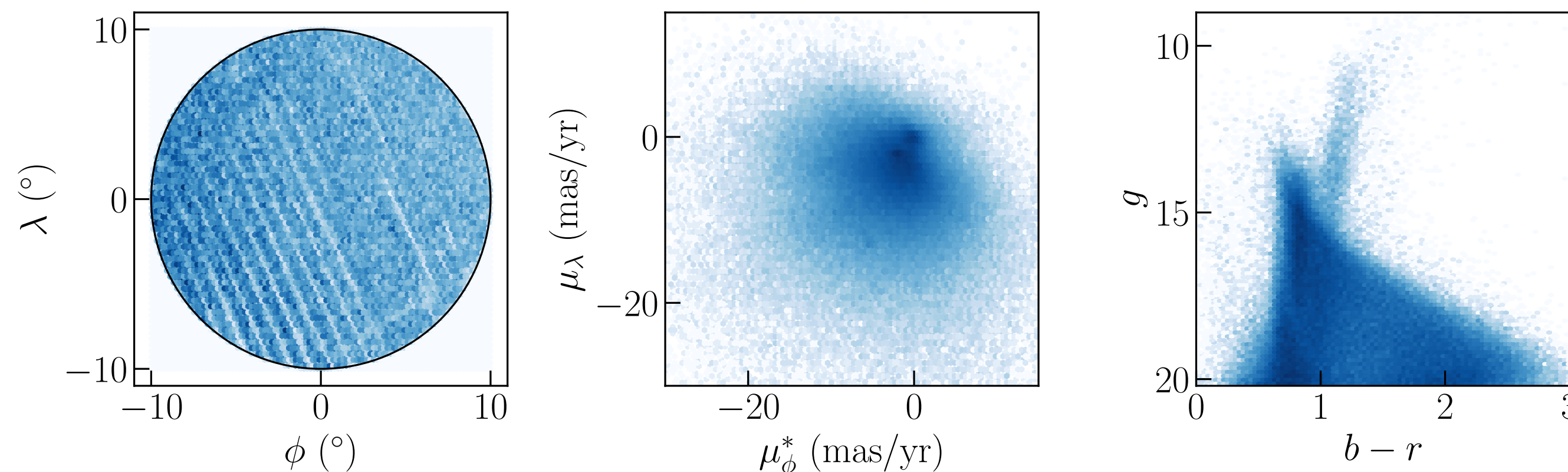


Stars in SR after cut on $R(x)$
obtained from ANODE

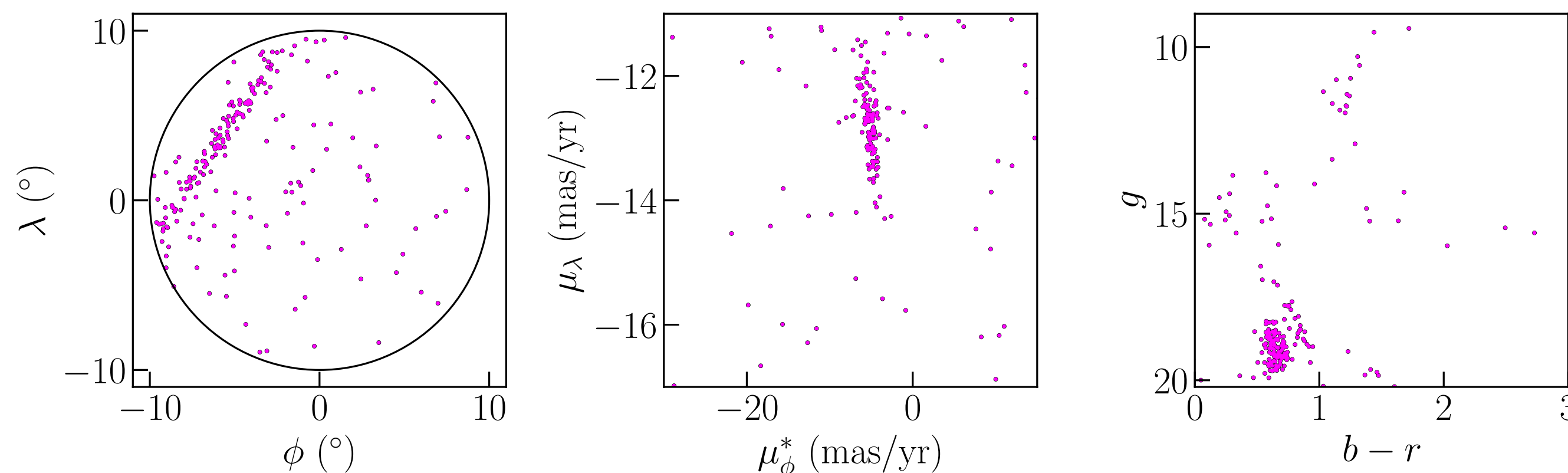
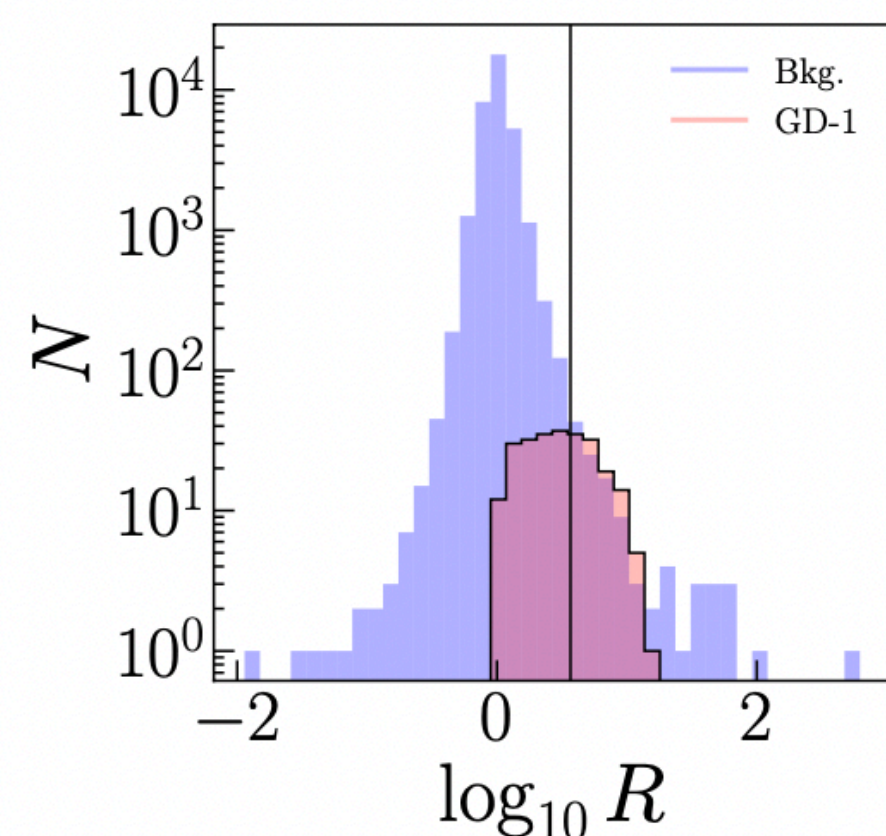
Core method — illustrated with GD-1 stream

[DS, Buckley, Necib, Tamasas '21]

Train ANODE on a proper motion window containing GD-1



All stars in a patch of the sky containing (part of) GD-1
(ra,dec)=(148.6,24.2)



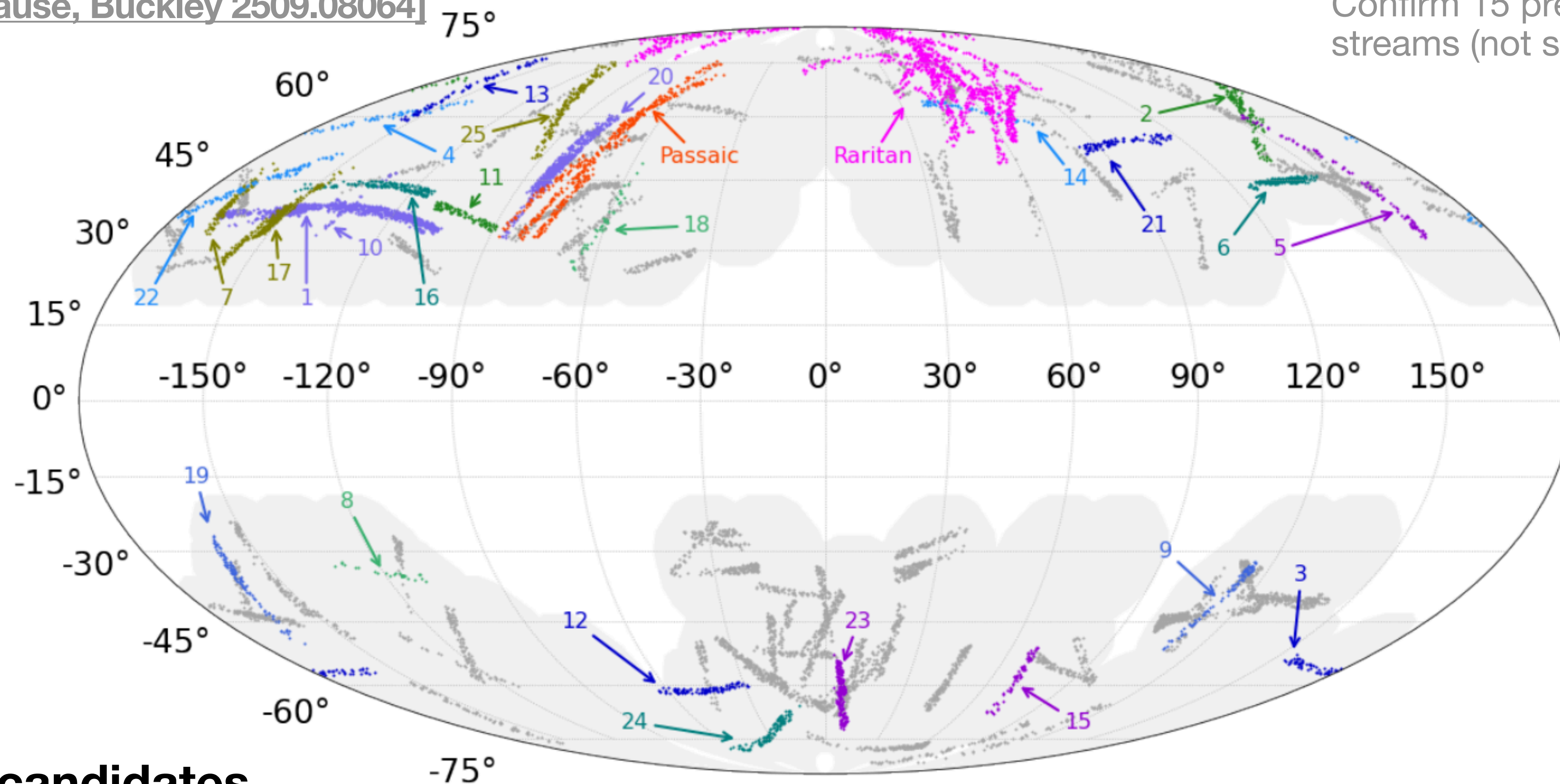
Stars in SR after cut on $R(x)$
obtained from ANODE

It works! Fully data-driven, simulation-independent

Via Machinae 3.0: Stream Atlas

[Hallin, DS, Krause, Buckley 2509.08064]

Confirm 15 previously discovered streams (not shown here)



New stream candidates



15 “gold plated”
— found by both
ANODE and
CATHODE methods



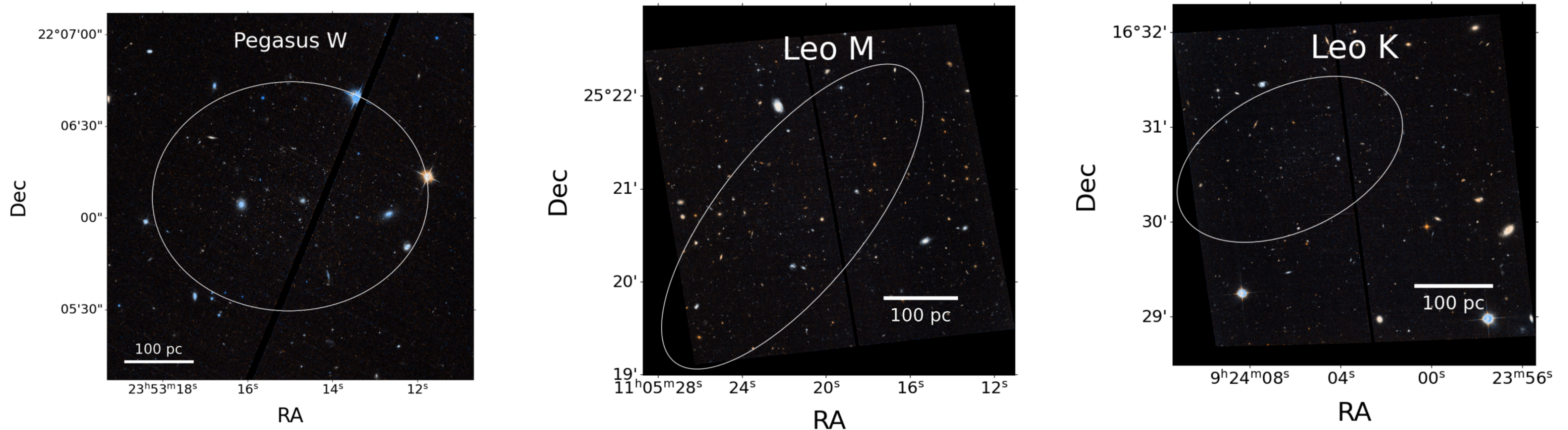
10 “silver plated”
— found by one method
but not the other with
high confidence



66 “bronze plated”
— found by one method
but not the other at
lower confidence

Spinoff: discovering new ultra-faint dwarf galaxies

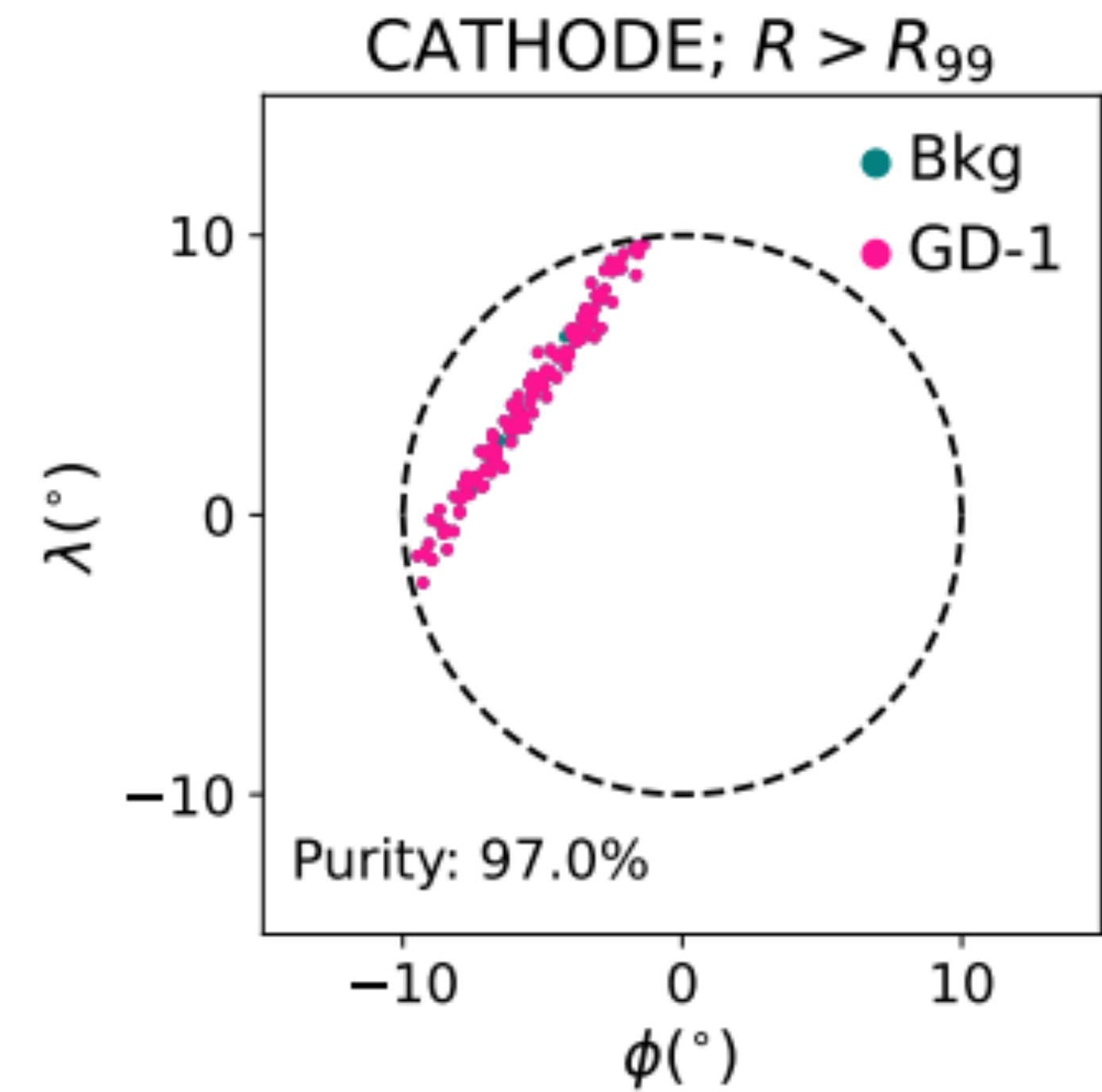
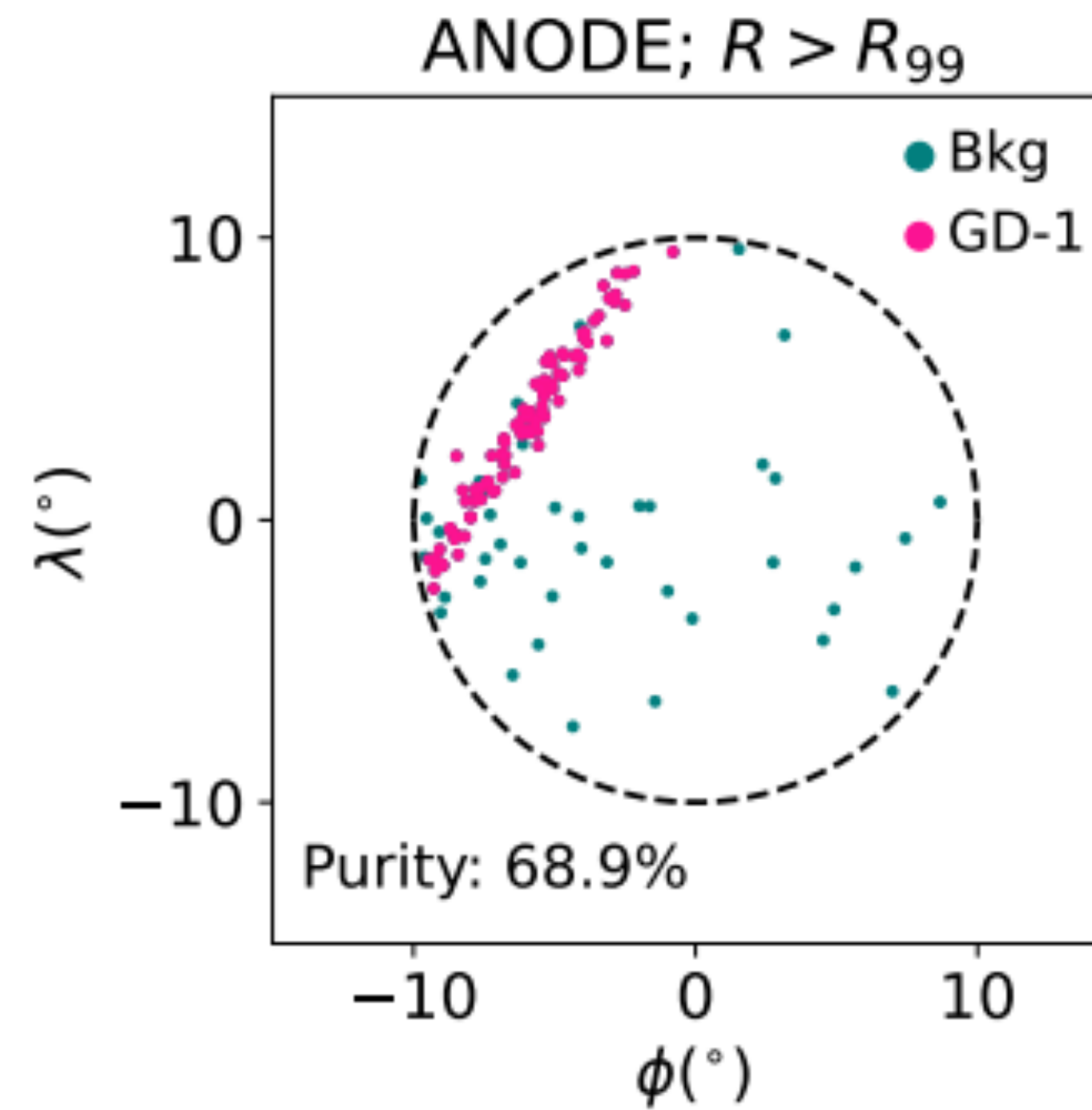
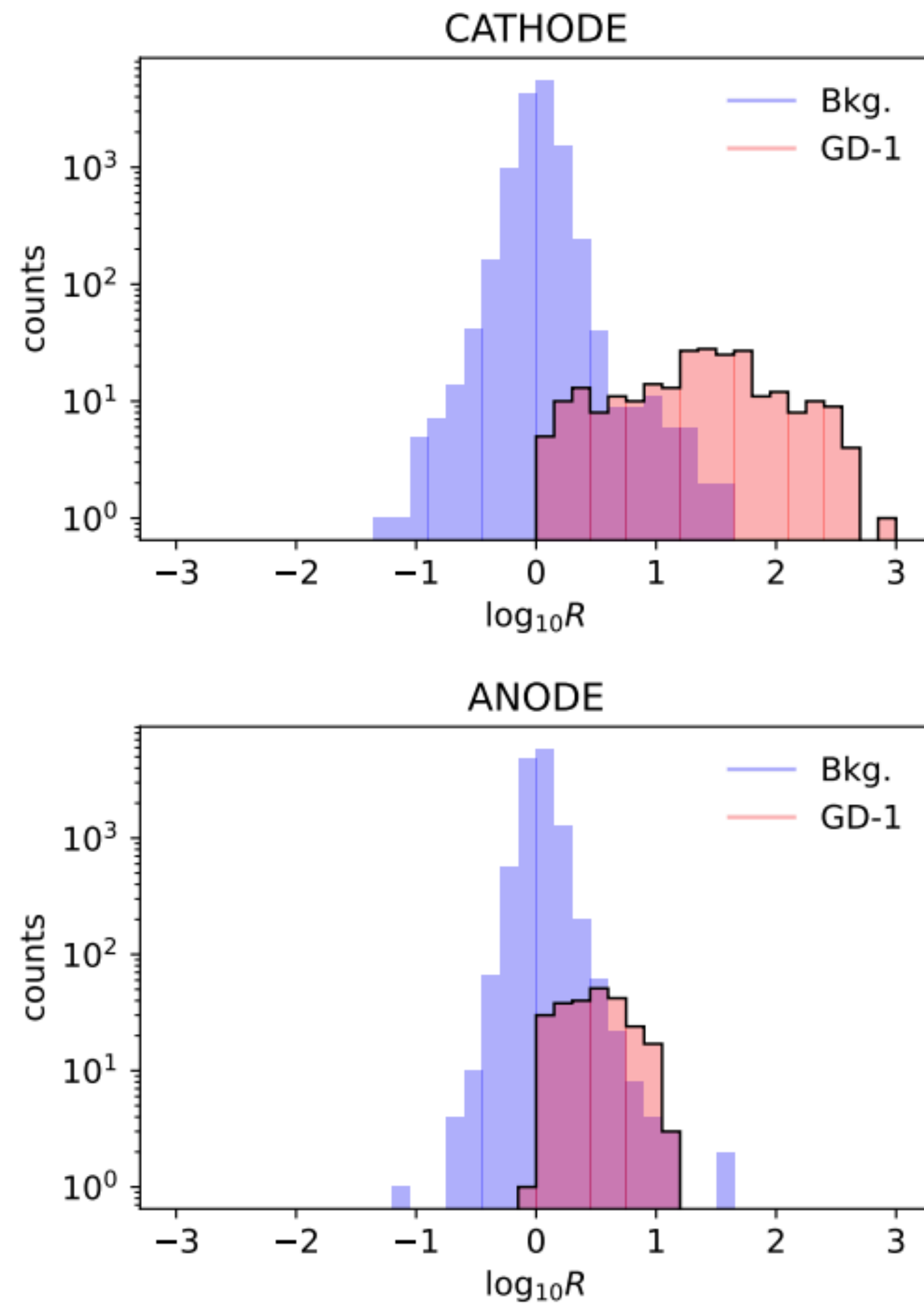
[DS+ McQuinn et al (2023a), (2023b)]



- Helped astronomers to discover 3 new ultra-faint “field” dwarf galaxies
 - First as overdensities of star counts in DESI Legacy Survey data
 - Then confirmed with dedicated Hubble time

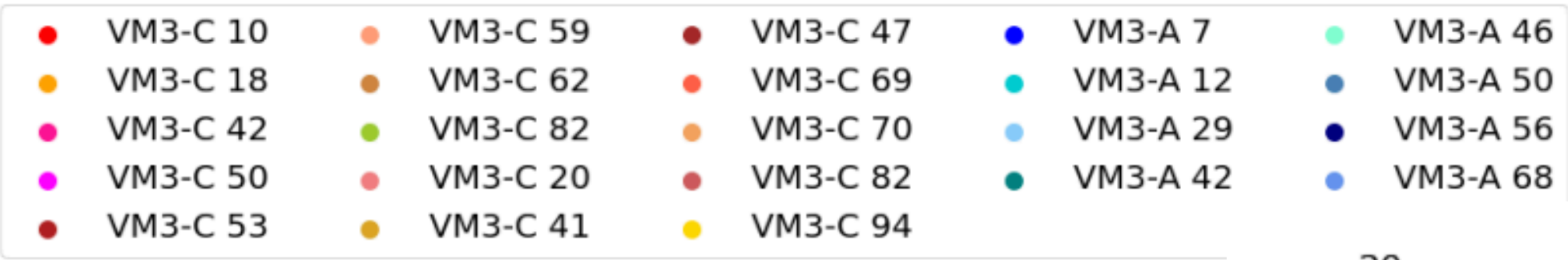
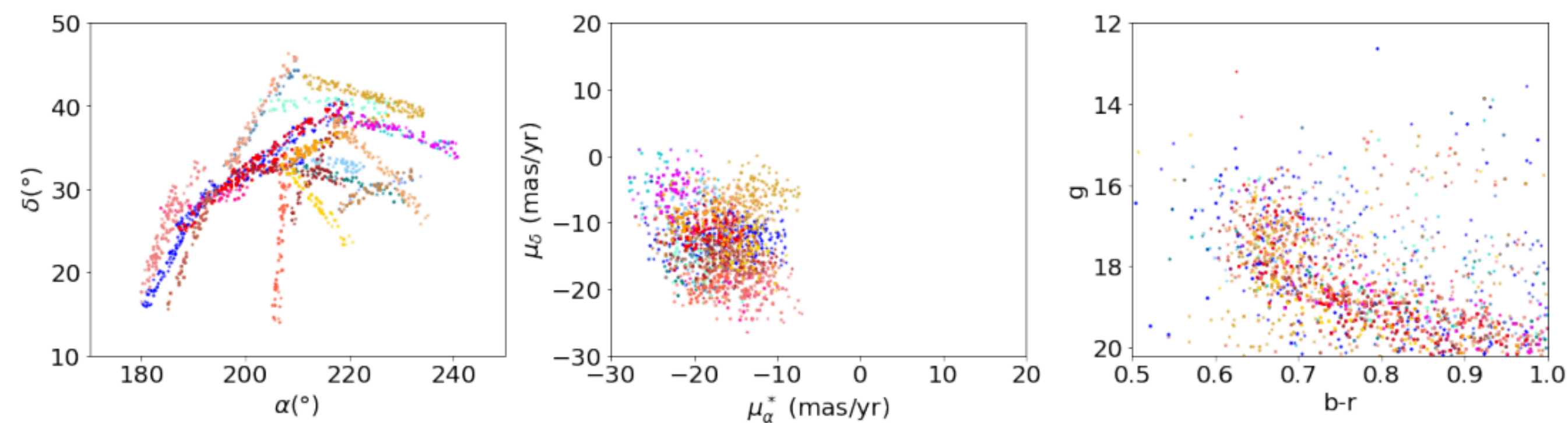
Improvements with CATHODE

[Hallin, DS, Krause, Buckley 2509.08064]



Passaic and Raritan Streams

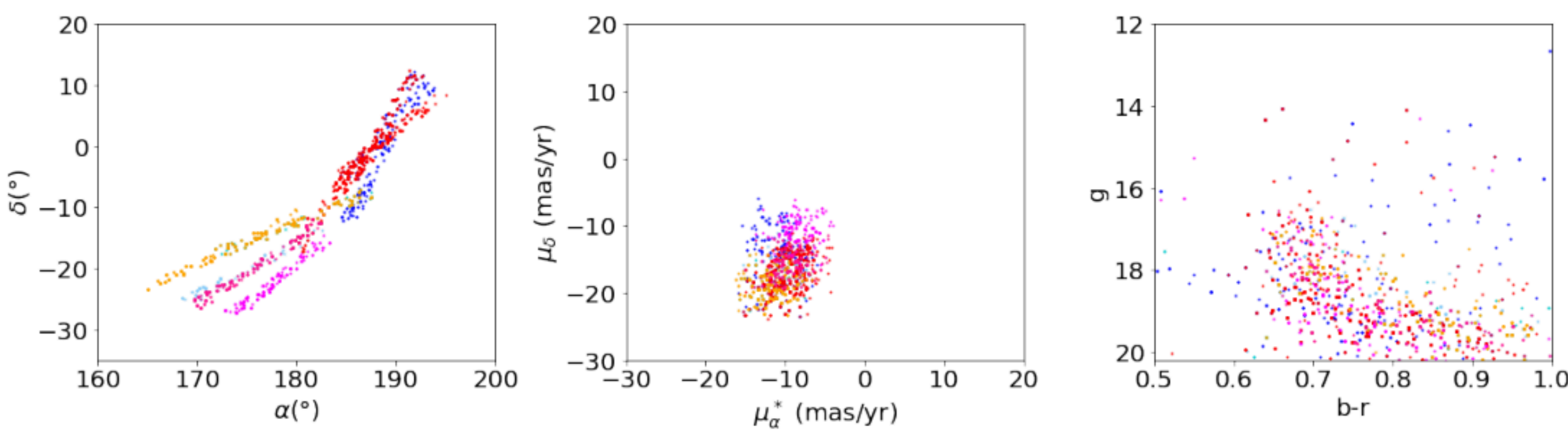
Raritan



Could be remnants of dwarf galaxies,
or result of exotic orbital dynamics
(chaos, bifurcation, fanning)

Find two large **clusters** of streams that don't look like anything that's been found before

Passaic



Collaborators around the world

Collaborators around the world

Map showing the locations of collaborators across various countries, including:

- United States
- Canada
- Greenland (Denmark)
- Iceland
- Norway
- Sweden
- Finland
- Estonia
- Latvia
- Lithuania
- Belarus
- Ukraine
- Georgia
- Azerbaijan
- Armenia
- Turkey
- Cyprus
- Lebanon
- Israel
- Jordan
- Saudi Arabia
- U.A.E.
- Yemen
- Oman
- Djibouti
- Iran
- Kazakhstan
- Mongolia
- China
- India
- Japan
- S. Korea
- Taiwan
- Philippines
- Malaysia
- Indonesia
- Timor-Leste
- Australia
- New Zealand
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanzania
- Kenya
- Somalia
- Ethiopia
- South Sudan
- Sudan
- Eritrea
- Chad
- Niger
- Mali
- Nigeria
- Ghana
- Benin
- Togo
- Equatorial Guinea
- Gabon
- Angola
- Namibia
- Bots.
- Zim.
- Zambia
- South Africa
- Lesotho
- eSwatini
- Mozambique
- Malawi
- Uganda
- Rwanda
- Burundi
- Tanz