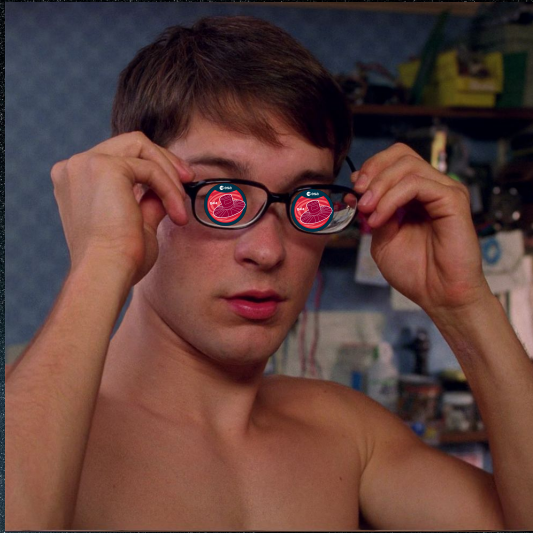








The Magellanic Clouds



# The Magellanic Clouds

# Neural Network classifier for the generation of clean Magellanic Cloud samples

Óscar Jiménez-Arranz

[ojimenez@fqa.ub.edu](mailto:ojimenez@fqa.ub.edu)

Institut de Ciències del Cosmos (ICCUB)

# Neural Network classifier for the generation of clean Magellanic Cloud samples

PhD Supervisors:

M. Romero-Gómez

X. Luri

Óscar Jiménez-Arranz

[ojimenez@fqa.ub.edu](mailto:ojimenez@fqa.ub.edu)

Institut de Ciències del Cosmos (ICCUB)

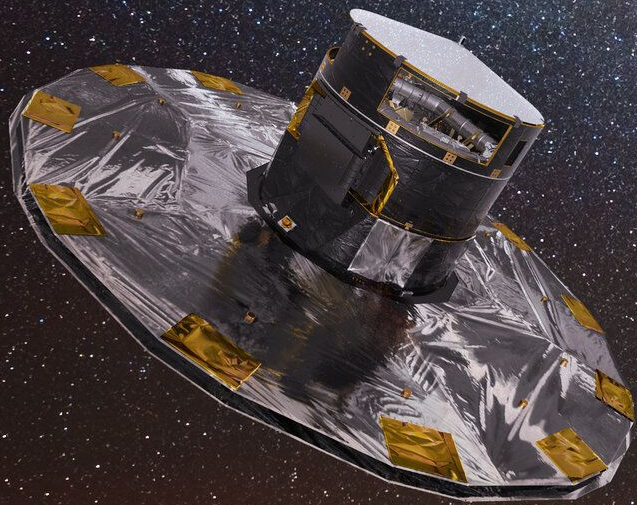
In collaboration with:  
P.J. McMillan, T. Antoja, L. Chemin,  
S. Roca-Fàbrega, E. Masana, A. Muros

# Neural Network classifier for the generation of clean Magellanic Cloud samples

PhD Supervisors:  
M. Romero-Gómez  
X. Luri

Óscar Jiménez-Arranz  
ojimenez@fqa.ub.edu  
Institut de Ciències del Cosmos (ICCUB)

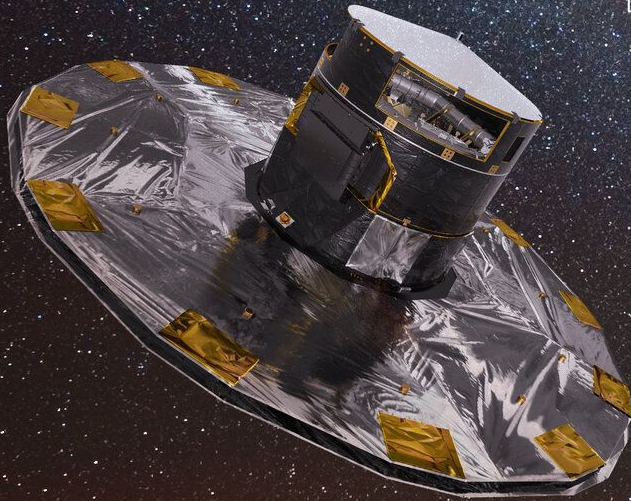
# Gaia (ESA)





# Gaia (ESA)

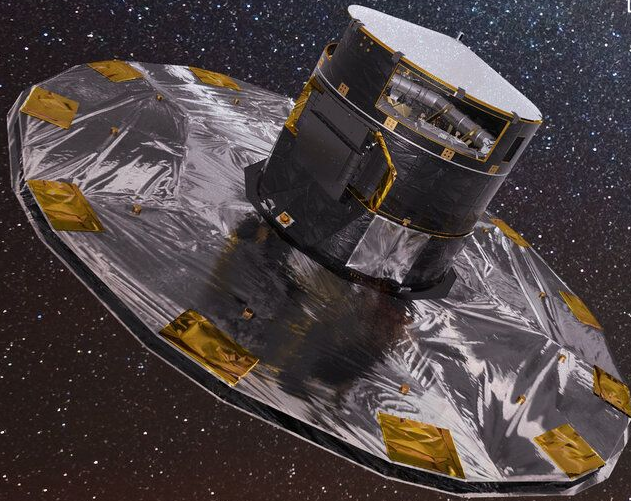
**Gaia** (2013 - present) has **astrometric**, **photometric** and **spectroscopic** data for almost **2.000M** stars.



# Gaia (ESA)

Position and  
velocities

Gaia (2013 - present) has astrometric, photometric and spectroscopic data for almost 2.000M stars.

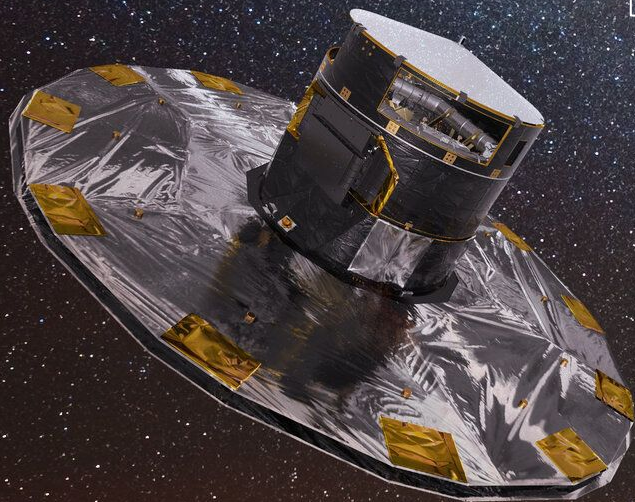


# Gaia (ESA)

Colors of the  
stars

Position and  
velocities

**Gaia** (2013 - present) has **astrometric, photometric and spectroscopic data** for almost **2.000M stars**.



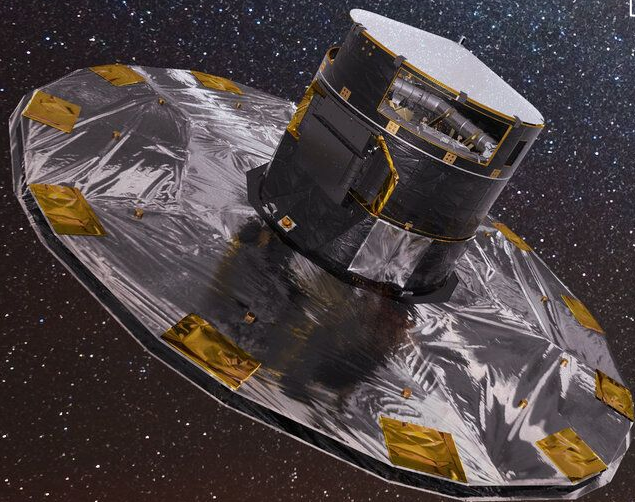
# Gaia (ESA)

Colors of the stars

Position and velocities

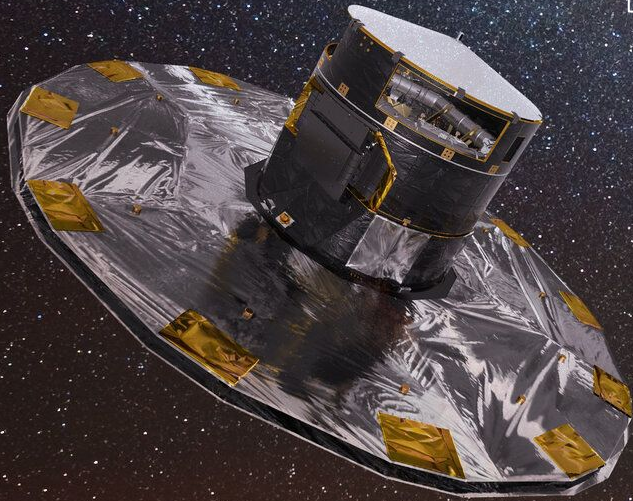
**Gaia** (2013 - present) has **astrometric, photometric and spectroscopic** data for almost **2.000M** stars.

**Spectra**

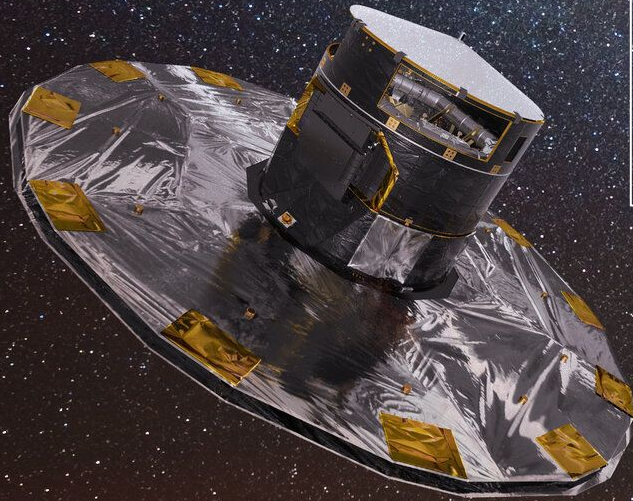


# Gaia (ESA)

**Gaia** (2013 - present) has **astrometric**, **photometric** and **spectroscopic** data for almost **2.000M stars** (**1%** of the **MW stars**).



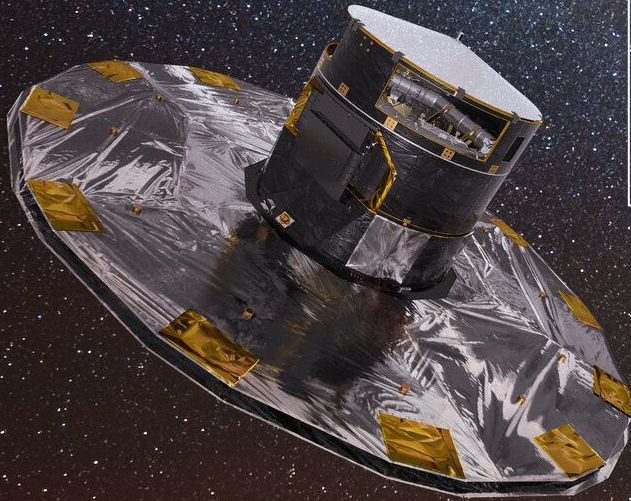
# Gaia (ESA)



**Gaia** (2013 - present) has **astrometric**, **photometric** and **spectroscopic** data for almost **2.000M stars** (**1%** of the **MW stars**).

Predecessor: **Hipparcos** (1989) with **100k stars**.

# Gaia (ESA)



**Gaia** (2013 - present) has **astrometric**, **photometric** and **spectroscopic** data for almost **2.000M stars** (**1%** of the **MW stars**).

Predecessor: **Hipparcos** (1989) with **100k stars**. Ratio of 1 : 20.000 (!!!)

**Hipparcos**





# GAIA'S REACH

The Gaia spacecraft will use parallax and ultra-precise position measurements to obtain the distances and 'proper' (sideways) motions of stars throughout much of the Milky Way, seen here edge-on. Data from Gaia will shed light on the Galaxy's history, structure and dynamics.

Gaia

Hipparcos



Sun

Galactic Centre

Previous missions could measure stellar distances with an accuracy of 10% only up to 100 parsecs\*

# GAIA'S REACH

The Gaia spacecraft will use parallax and ultra-precise position measurements to obtain the distances and 'proper' (sideways) motions of stars throughout much of the Milky Way, seen here edge-on. Data from Gaia will shed light on the Galaxy's history, structure and dynamics.

## Hipparcos

Previous missions could measure stellar distances with an accuracy of 10% only up to 100 parsecs\*

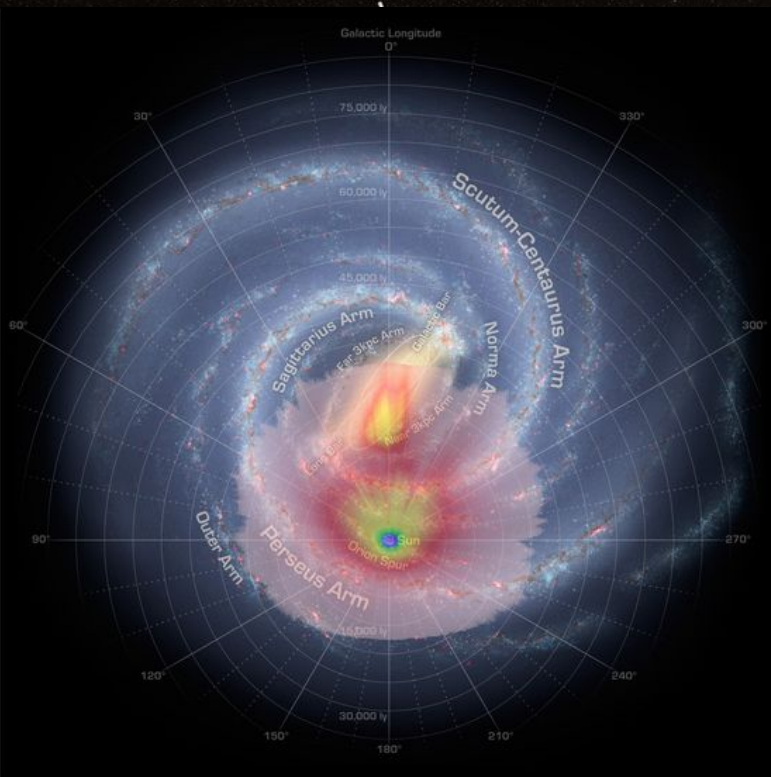


Sun

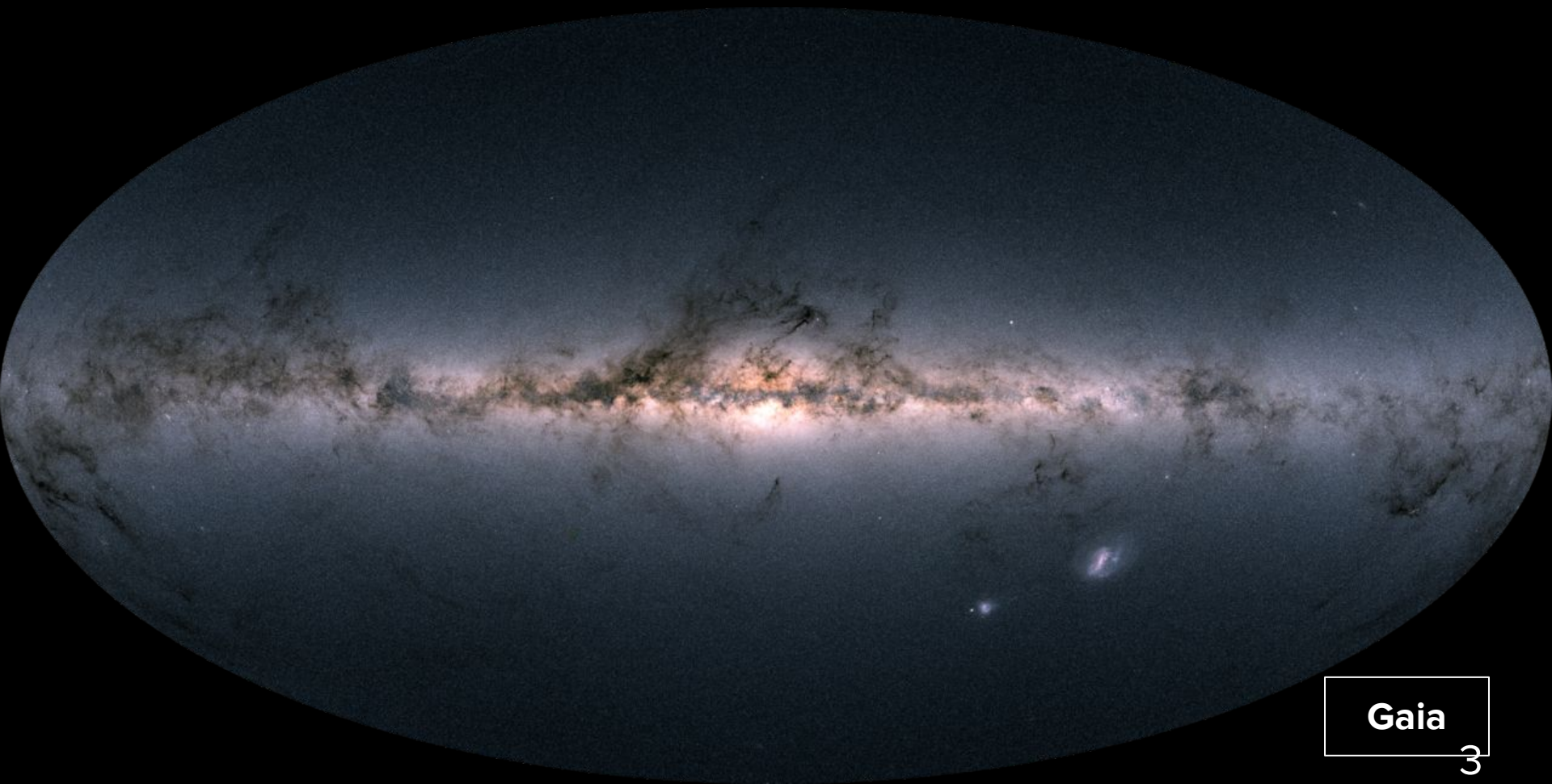
Galactic Centre

## Gaia

Gaia's distance 10% w

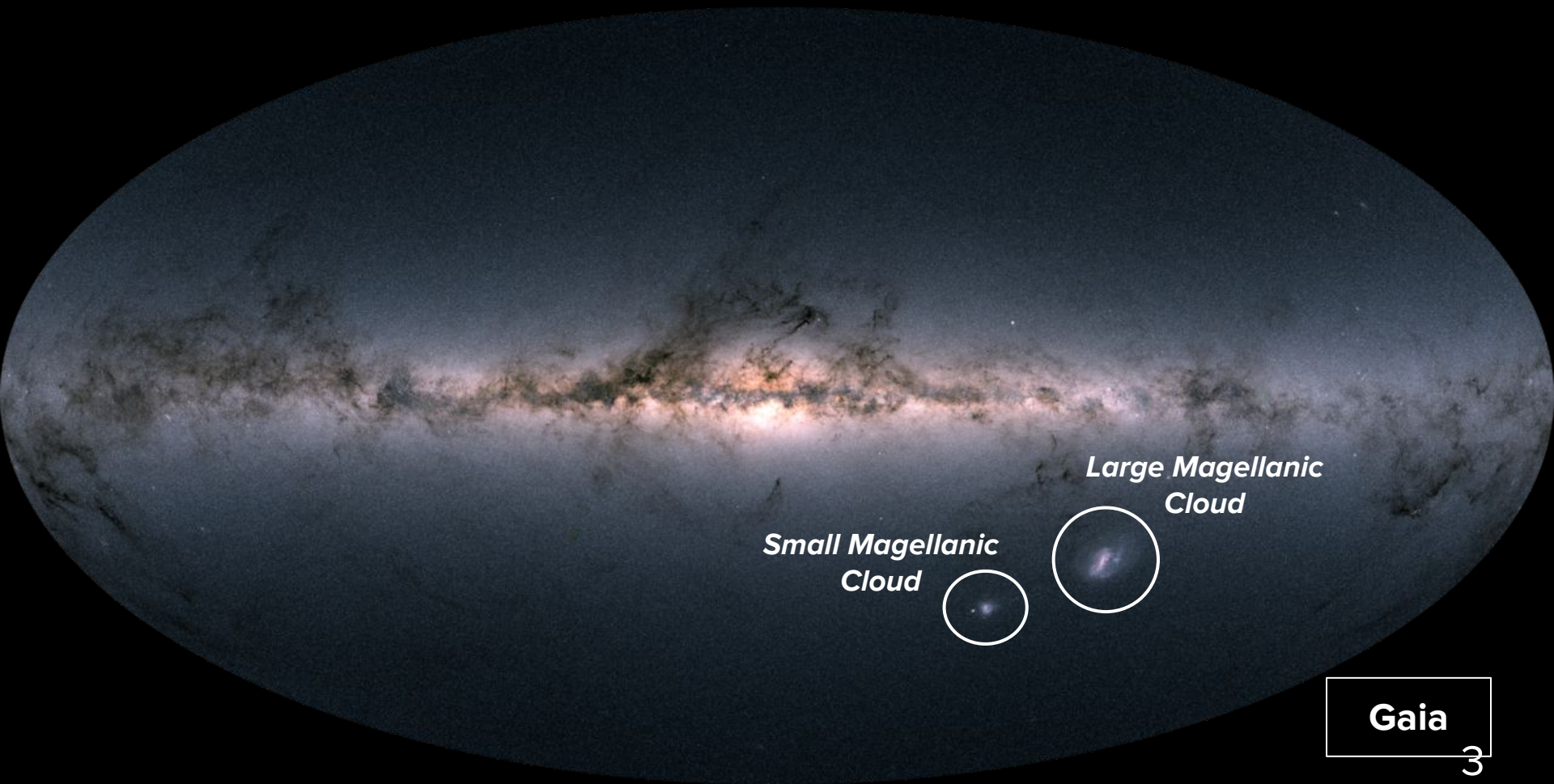


\*1 parsec = 3.26 light years



Gaia

3



*Small Magellanic  
Cloud*

*Large Magellanic  
Cloud*

*Large Magellanic Cloud*



*Small Magellanic Cloud*



Gaia

*Large Magellanic Cloud*



*Small Magellanic Cloud*



*What makes the MCs interesting?*

Gaia

## *Large Magellanic Cloud*



## *Small Magellanic Cloud*



*What makes the MCs interesting?*

- LMC is the **closest spiral galaxy** to the MW

## Large Magellanic Cloud

## Small Magellanic Cloud



*What makes the MCs interesting?*

- LMC is the **closest spiral galaxy** to the MW  
(astrometric information for million stars)

Gaia



## *Large Magellanic Cloud*



## *Small Magellanic Cloud*



*What makes the MCs interesting?*

- LMC is the **closest spiral galaxy** to the MW (astrometric information for million stars)
- In **strong interaction**

Gaia

## Large Magellanic Cloud

## Small Magellanic Cloud

The **MCs** are the **perfect laboratory** for **testing methodologies** and **models** designed for the study of **external** and **interacting galaxies**

*What makes the MCs interesting?*

- LMC is the **closest spiral galaxy** to the MW (**astrometric information for million stars**)
- In **strong interaction**

# My PhD Journey

- 1) **Kinematic analysis of the Large Magellanic Cloud using Gaia DR3**  
(Ó. Jiménez-Arranz et al. 2023a)

# My PhD Journey

- 1) **Kinematic analysis of the Large Magellanic Cloud using Gaia DR3**  
(Ó. Jiménez-Arranz et al. 2023a)
  - **Neural network classifier** for the selection of **clean LMC samples**

# My PhD Journey



- 1) **Kinematic analysis of the Large Magellanic Cloud using Gaia DR3**  
(Ó. Jiménez-Arranz et al. 2023a)
  - **Neural network** classifier for the selection of **clean LMC samples**
  - Kinematic analysis of the **in-plane velocities** for the LMC

# My PhD Journey

## 1) Kinematic analysis of the Large Magellanic Cloud using Gaia DR3

(Ó. Jiménez-Arranz et al. 2023a)

- Neural network classifier for the selection of clean LMC samples
- ~~Kinematic analysis of the in-plane velocities for the LMC~~

Irrelevant today

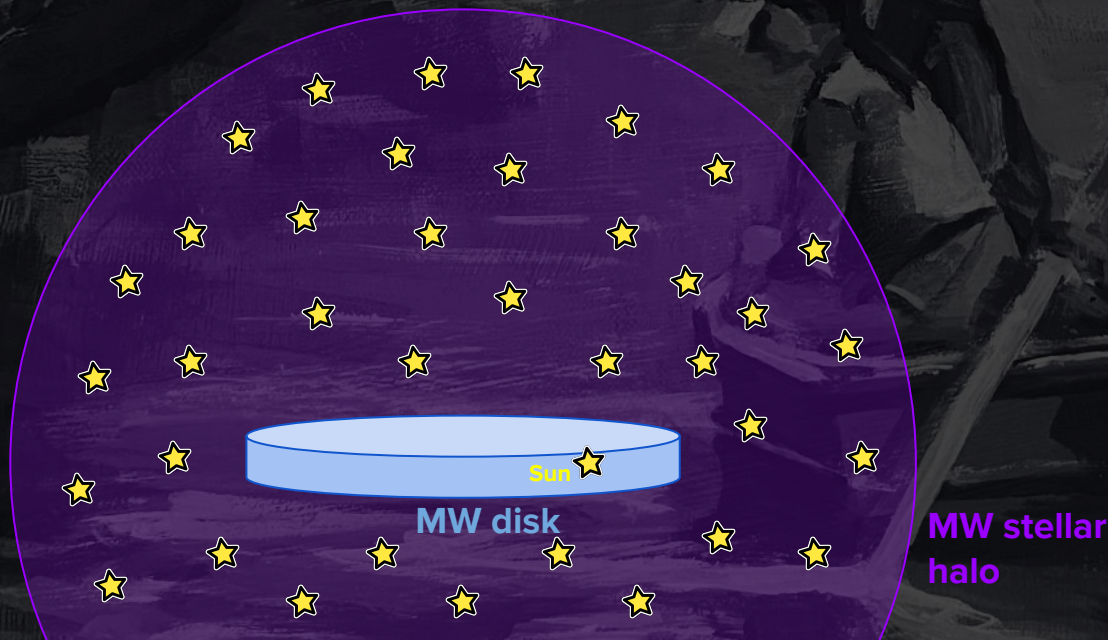


# My PhD Journey

## 1) Kinematic analysis of the Large Magellanic Cloud using Gaia DR3

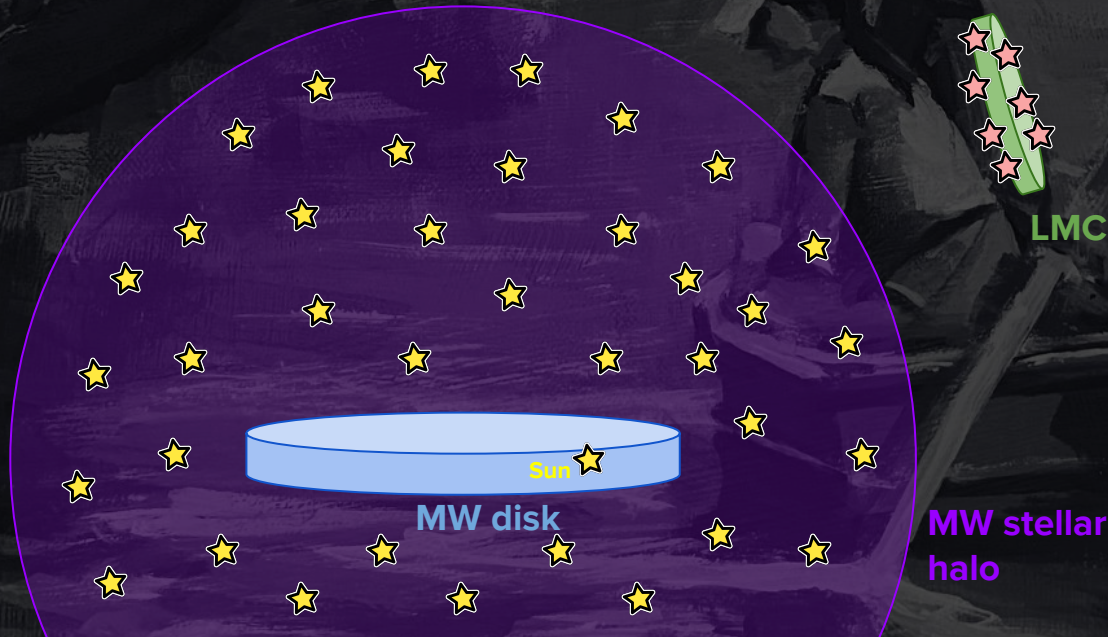
(Ó. Jiménez-Arranz et al. 2023a)

- Neural network classifier for the selection of clean LMC samples



# My PhD Journey

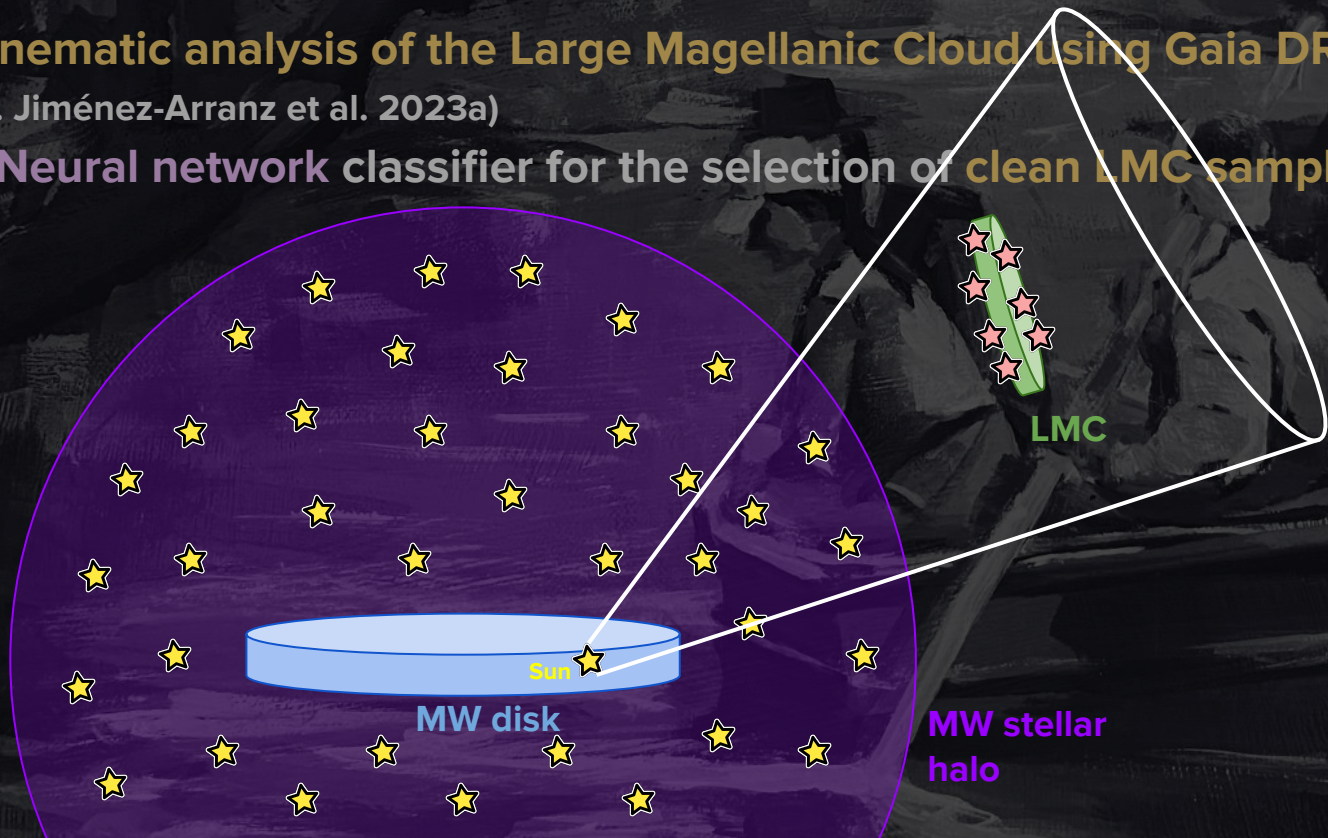
- 1) **Kinematic analysis of the Large Magellanic Cloud using Gaia DR3**  
(Ó. Jiménez-Arranz et al. 2023a)
  - **Neural network classifier for the selection of clean LMC samples**





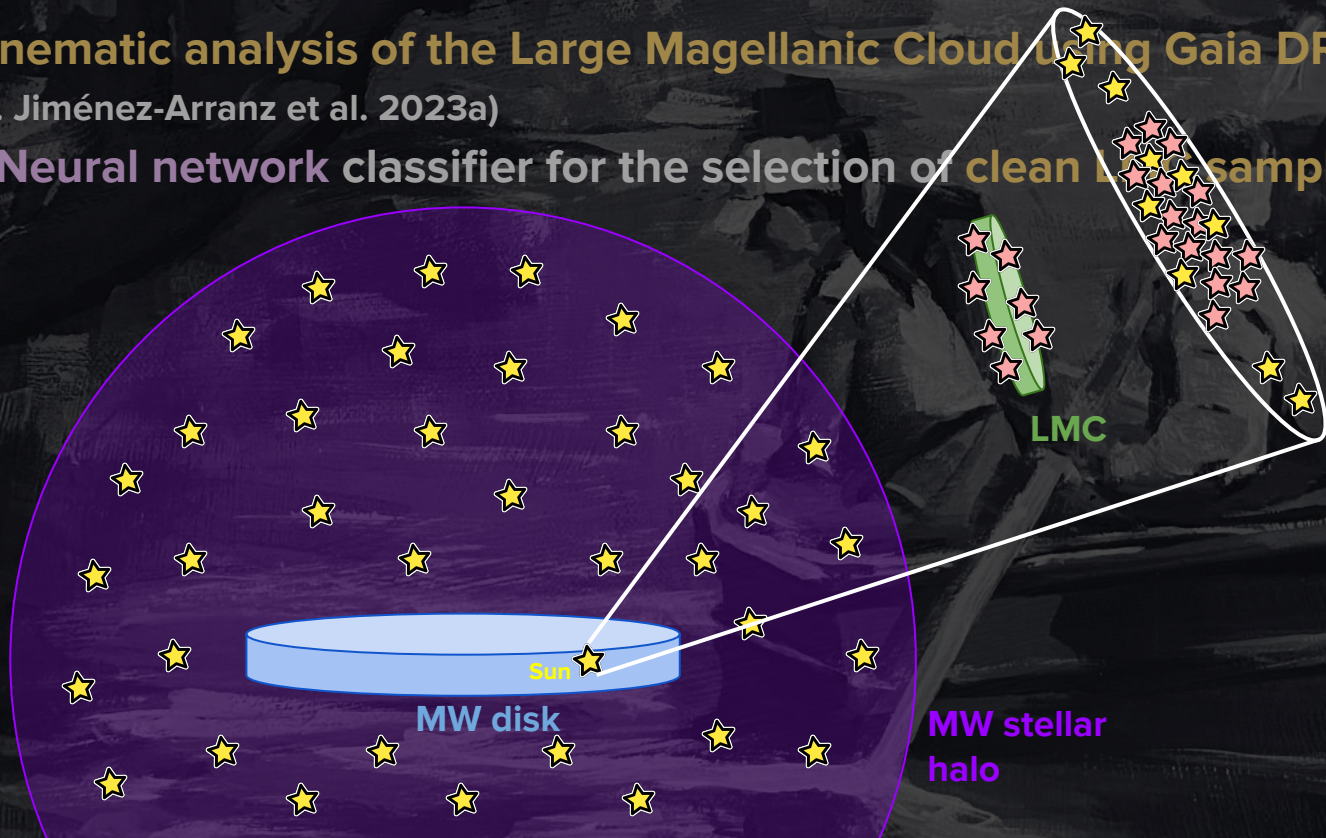
# My PhD Journey

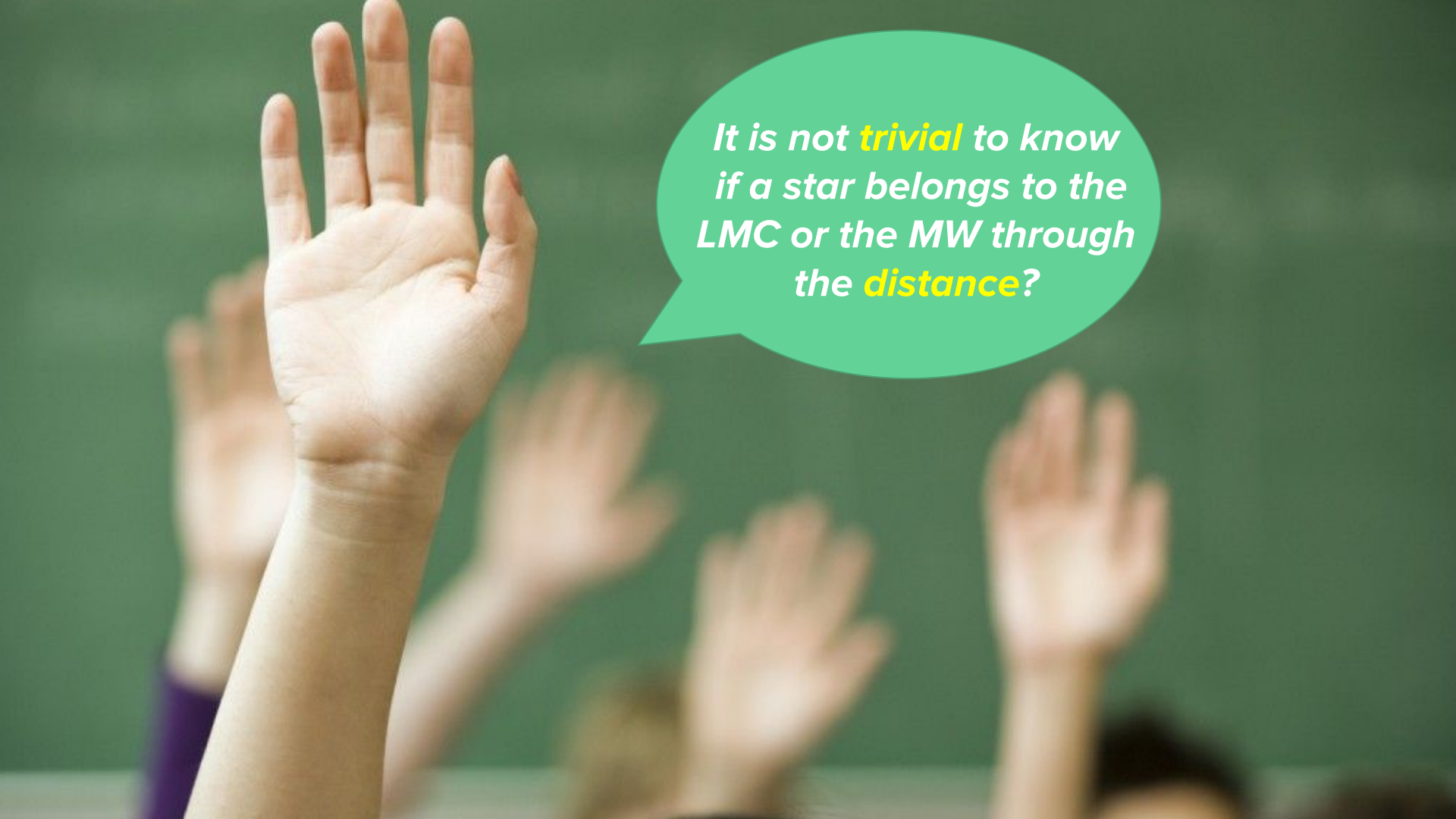
- 1) **Kinematic analysis of the Large Magellanic Cloud using Gaia DR3**  
(Ó. Jiménez-Arranz et al. 2023a)
  - **Neural network classifier for the selection of clean LMC samples**



# My PhD Journey

- 1) Kinematic analysis of the Large Magellanic Cloud using Gaia DR3 (Ó. Jiménez-Arranz et al. 2023a)
  - Neural network classifier for the selection of clean LMC samples




A hand is raised in the foreground, palm facing forward. In the background, several other hands are raised, though they are out of focus. A green speech bubble is positioned in the upper right quadrant, containing text. The background is a solid green color.

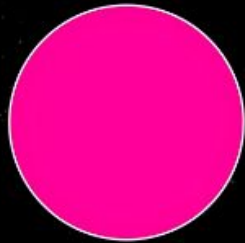
*It is not **trivial** to know  
if a star belongs to the  
LMC or the MW through  
the **distance**?*

A hand is visible at the top left, holding a white rectangular sign. On the sign is a large, hand-drawn speech bubble with a thick black outline. Inside the bubble, the text "YES, BUT..." is written in a bold, sans-serif font. "YES," is in black, and "BUT..." is in a dark red color. The background of the entire image is a blurred green wall.

**YES, BUT...**



It is not **trivial** to know  
if a star belongs to the  
LMC or the MW through  
the ~~distance?~~  
**parallax**

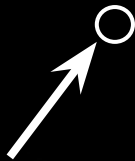








**Which is the  
distance of this  
star?**



LMC/MW classifier



# LMC/MW classifier

- Classification problem

# LMC/MW classifier

- **Classification problem** → **Supervised learning**

# LMC/MW classifier

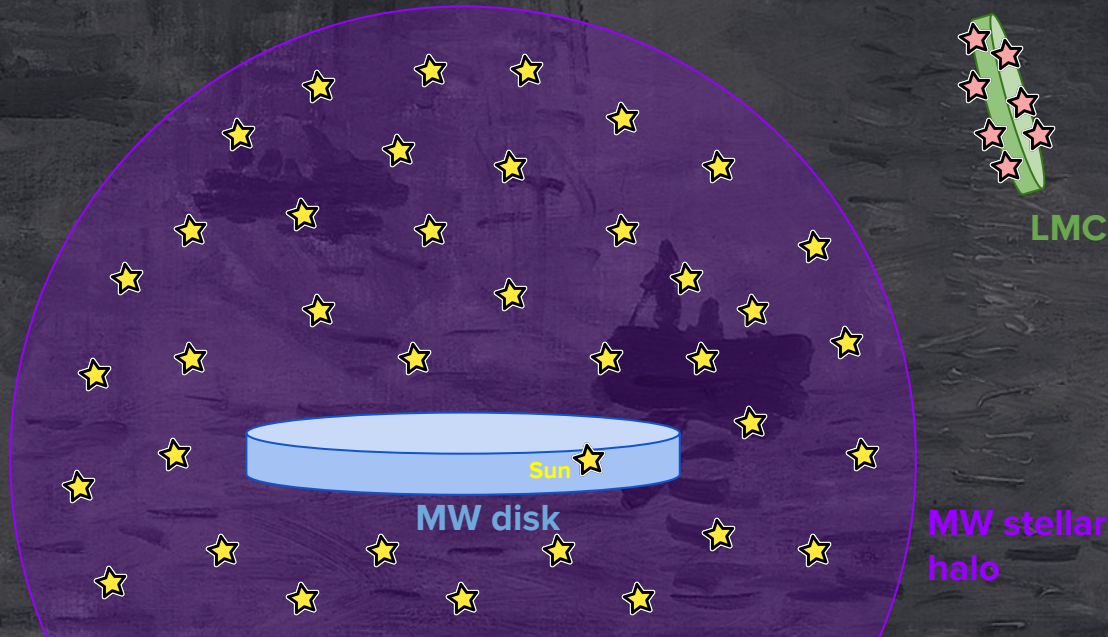
- **Classification problem** → **Supervised learning** → **Training sample**

## LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**

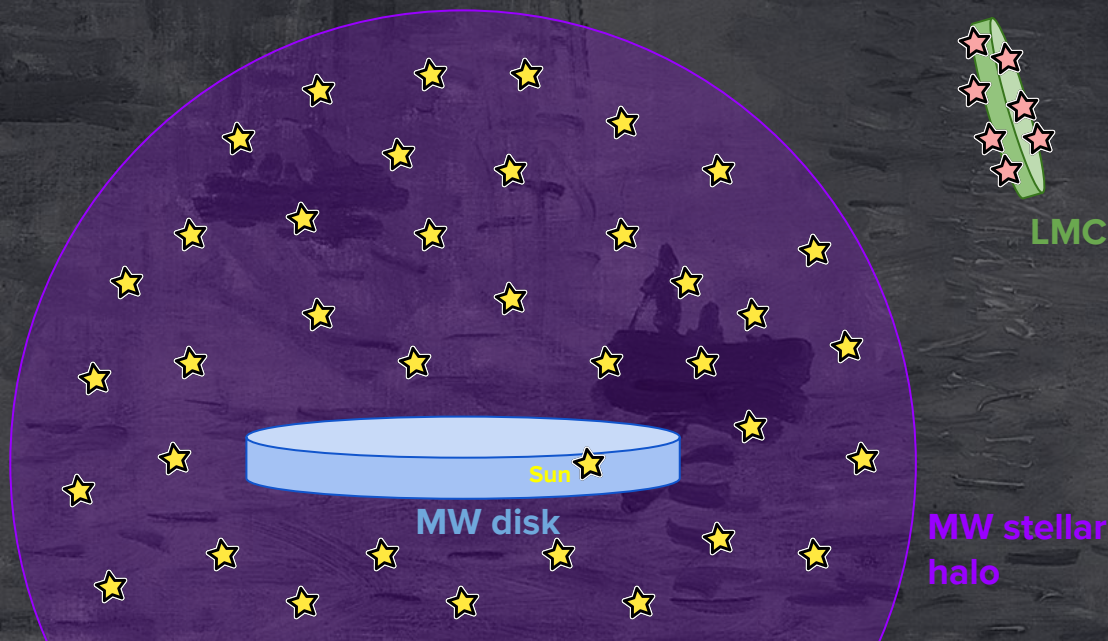
# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**



# LMC/MW classifier

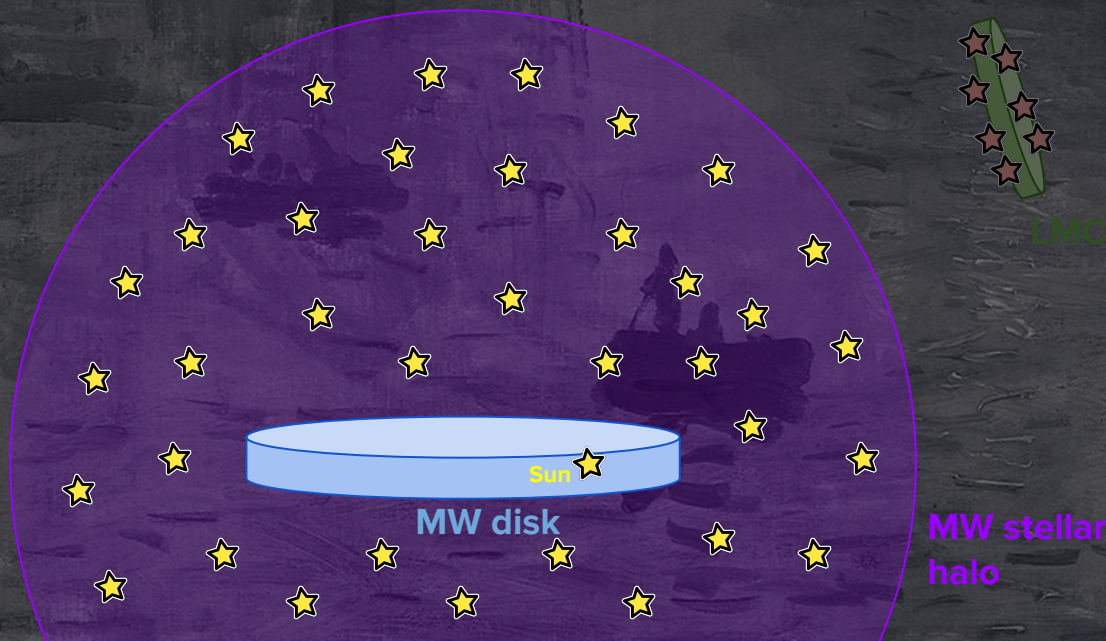
- **Classification problem** → **Supervised learning** → **Training sample**





# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**



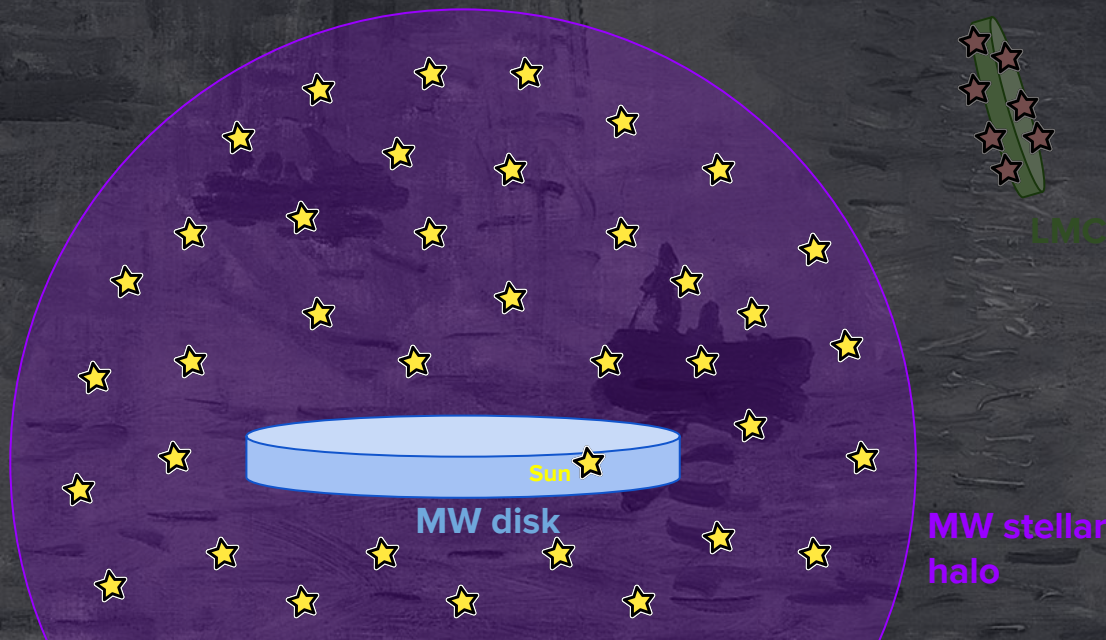
MW training sample



MW stellar halo

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**

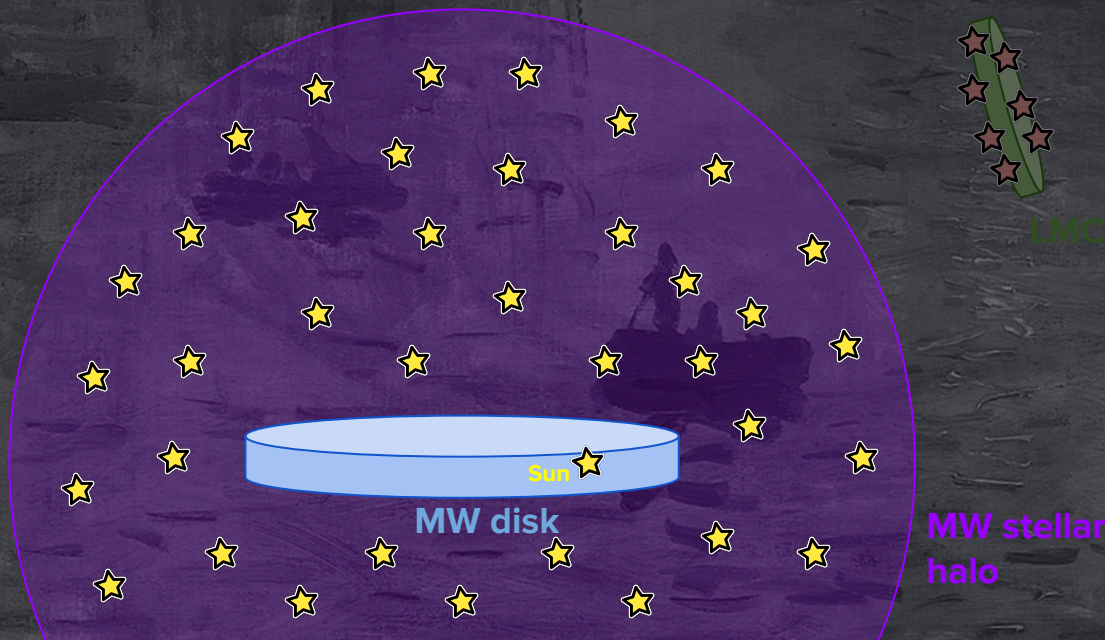


MW training sample:

- Realistic **galactic model**

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**

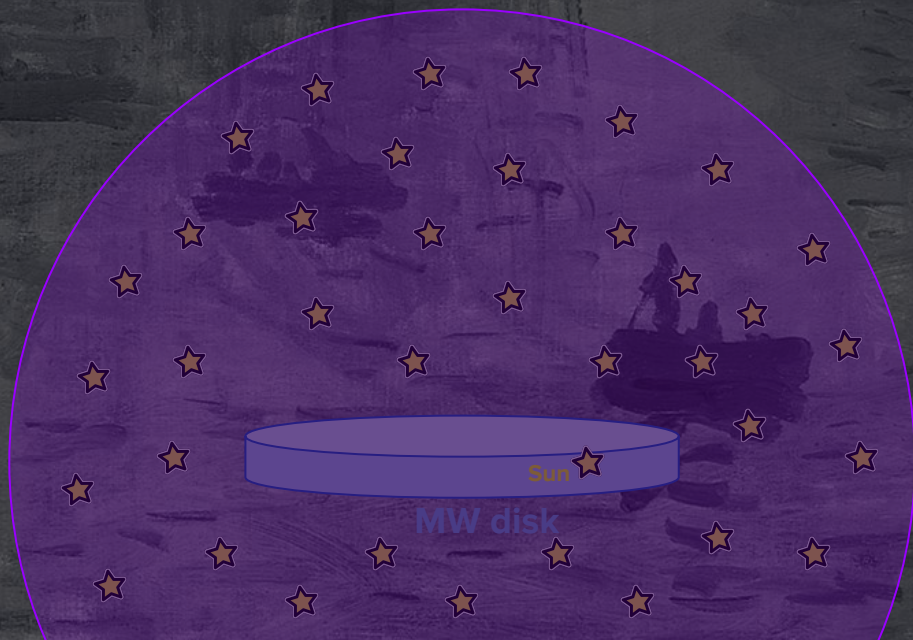


## MW training sample:

- Realistic **galactic model**
- Realistic **number of stars**

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**



**LMC training sample:**

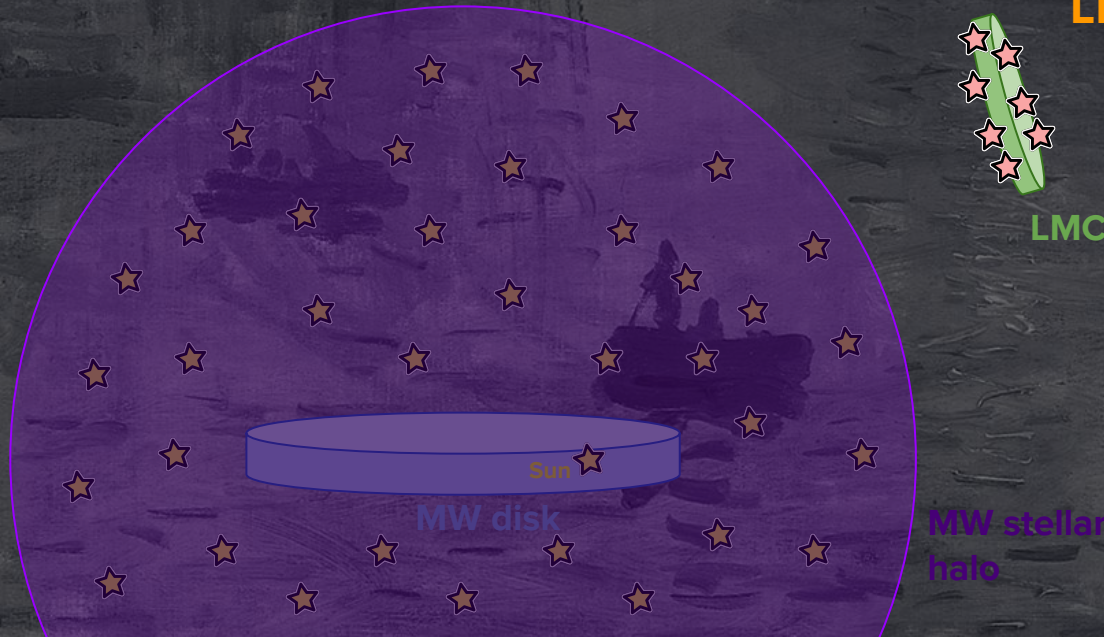


LMC

MW stellar  
halo

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**

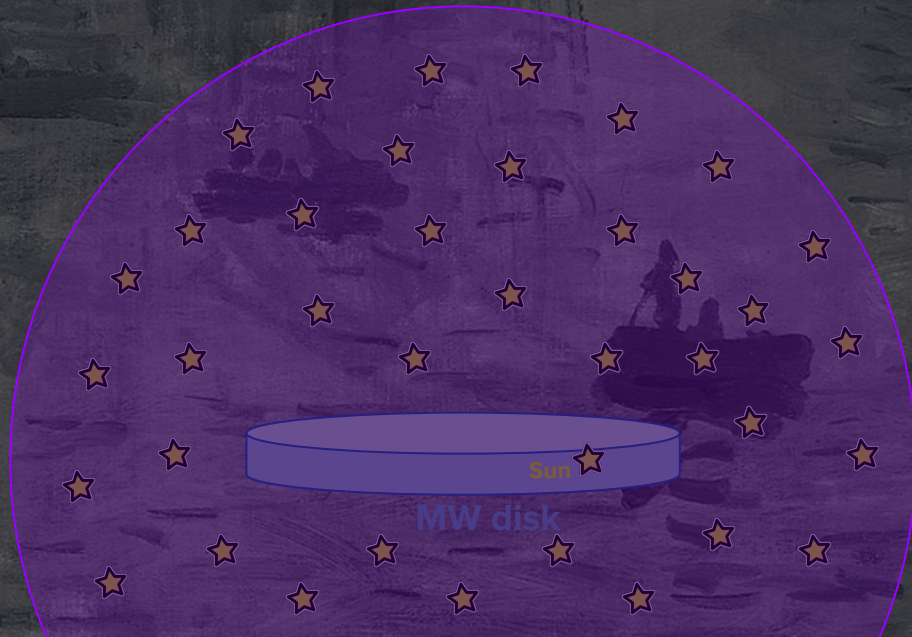


## LMC training sample:

- **Catalogue-based**  
(previous observations)

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**



## LMC training sample:

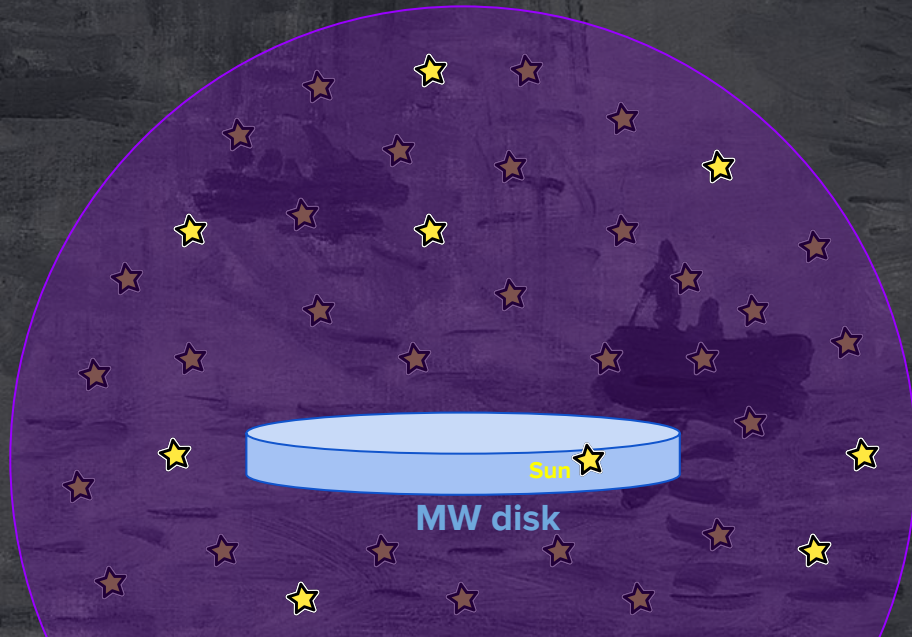
- **Catalogue-based**  
(previous observations)
- **Very few stars**

LMC

MW stellar  
halo

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**



LMC

## LMC training sample:

- **Catalogue-based**  
(previous observations)
- **Very few stars**

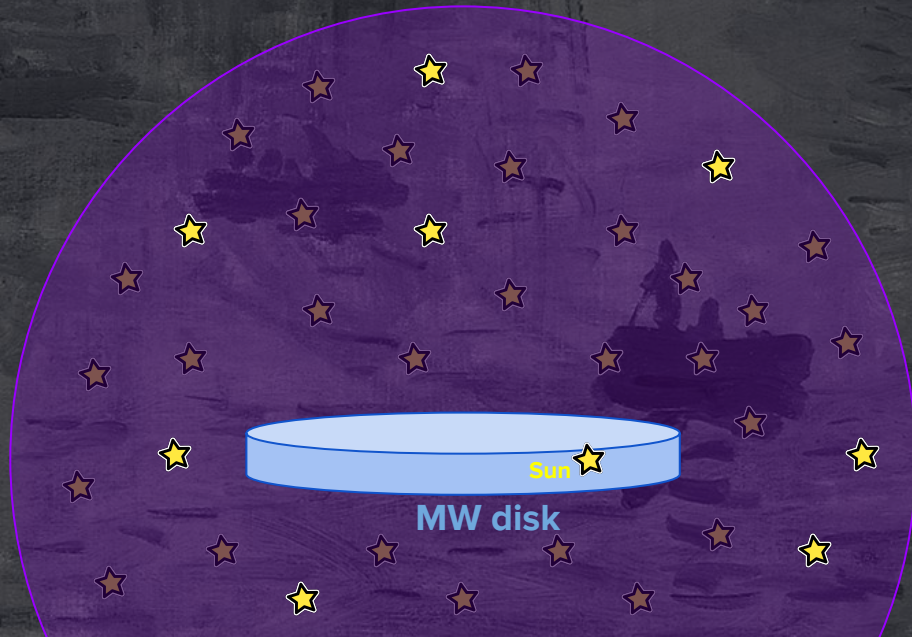
## MW training sample:

- We selected a **20% fraction** of the sample to balance

MW stellar halo

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**



## LMC training sample:

- **Catalogue-based**  
(previous observations)
- **Very few stars**

**280k stars**

## MW training sample:

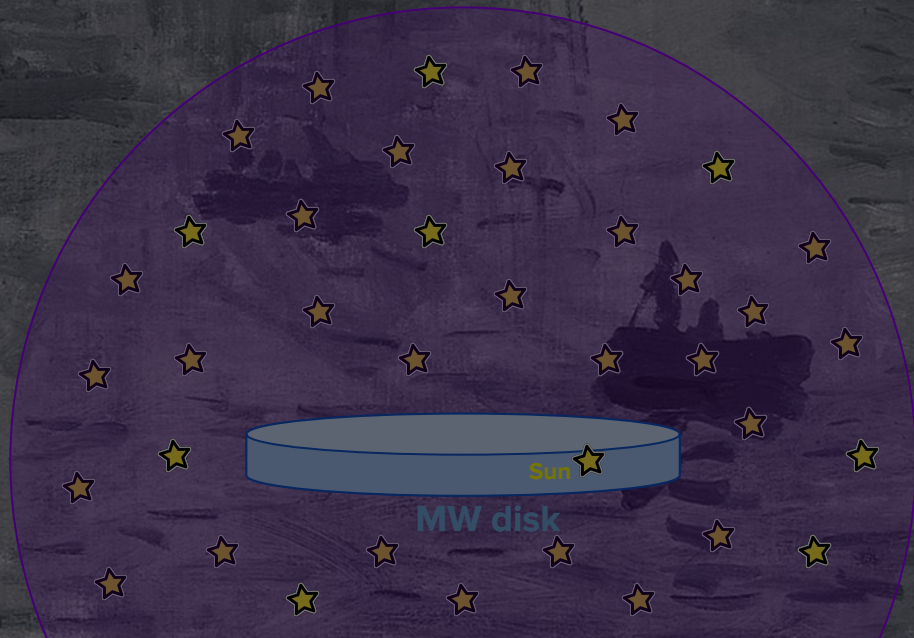
- We selected a **20% fraction** of the sample to balance

**1.3M stars**



# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**



## LMC training sample:

- Catalogue-based (previous observations)
- Very few stars

**280k stars**

## MW training sample:

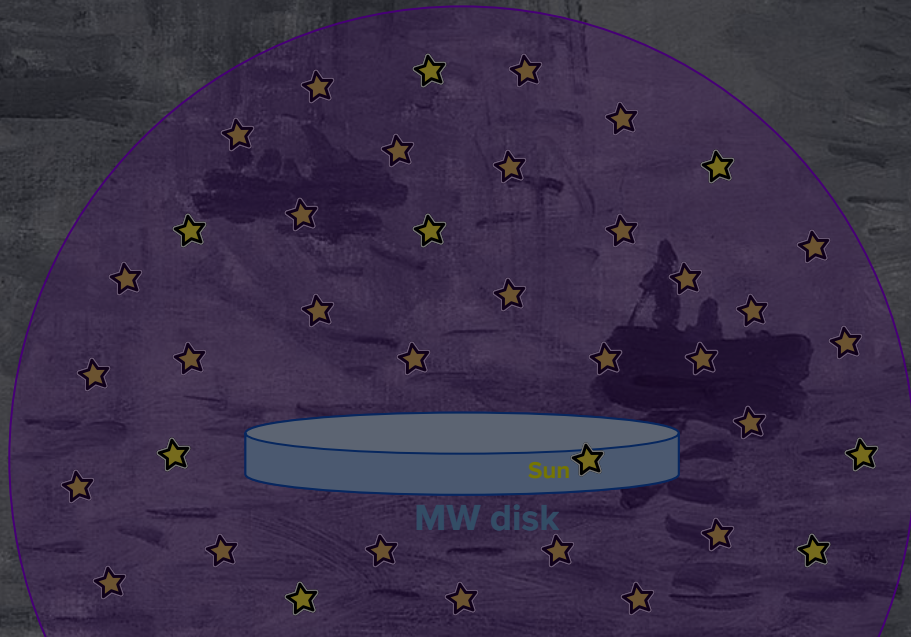
- We selected a 20% fraction of the sample to balance

**1.3M stars**

# LMC/MW classifier

- **Classification problem** → Supervised learning → Training sample

**60% training** | **40% test**



## LMC training sample:

- Catalogue-based (previous observations)
- Very few stars

**280k stars**

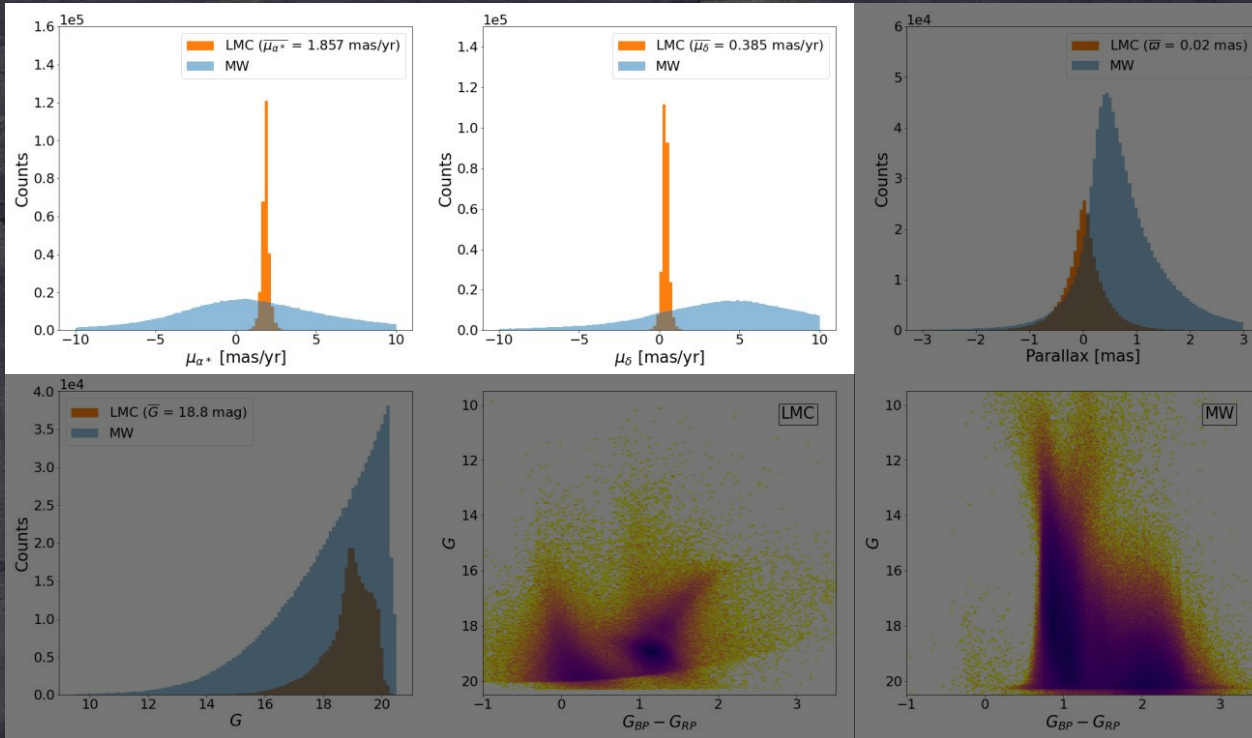
## MW training sample:

- We selected a 20% fraction of the sample to balance

**1.3M stars**

## LMC/MW classifier

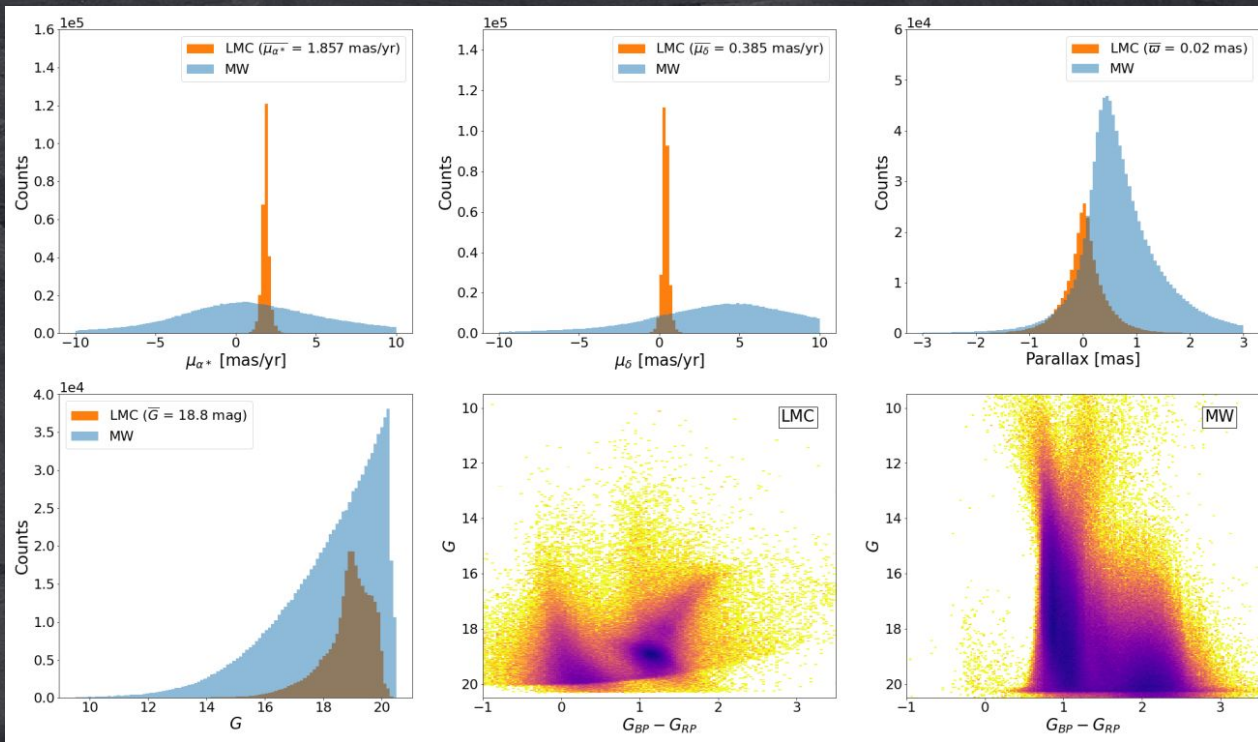
- Classification problem → Supervised learning → Training sample



LMC training sample  
MW training sample

## LMC/MW classifier

- Classification problem  $\rightarrow$  Supervised learning  $\rightarrow$  Training sample

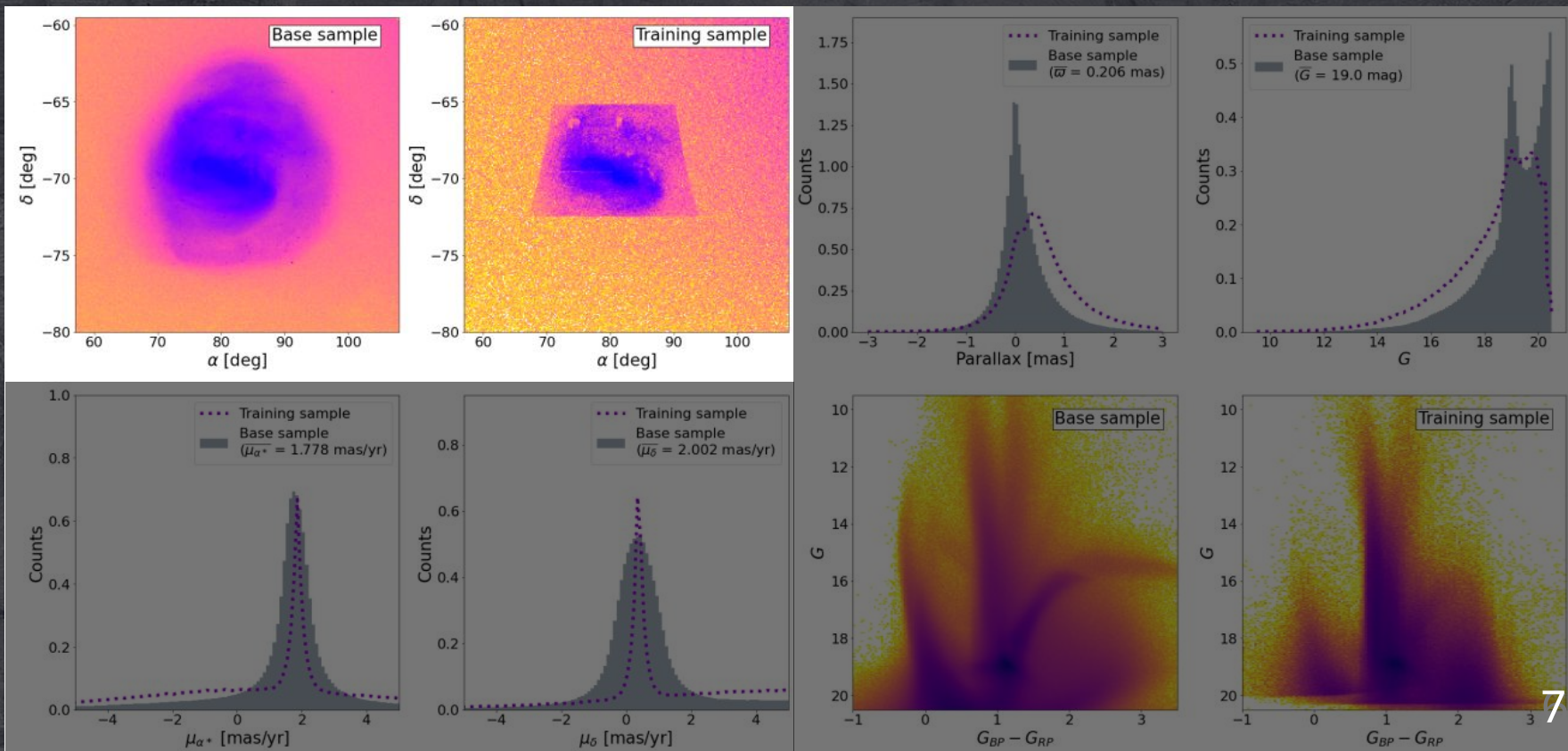


LMC training sample  
MW training sample

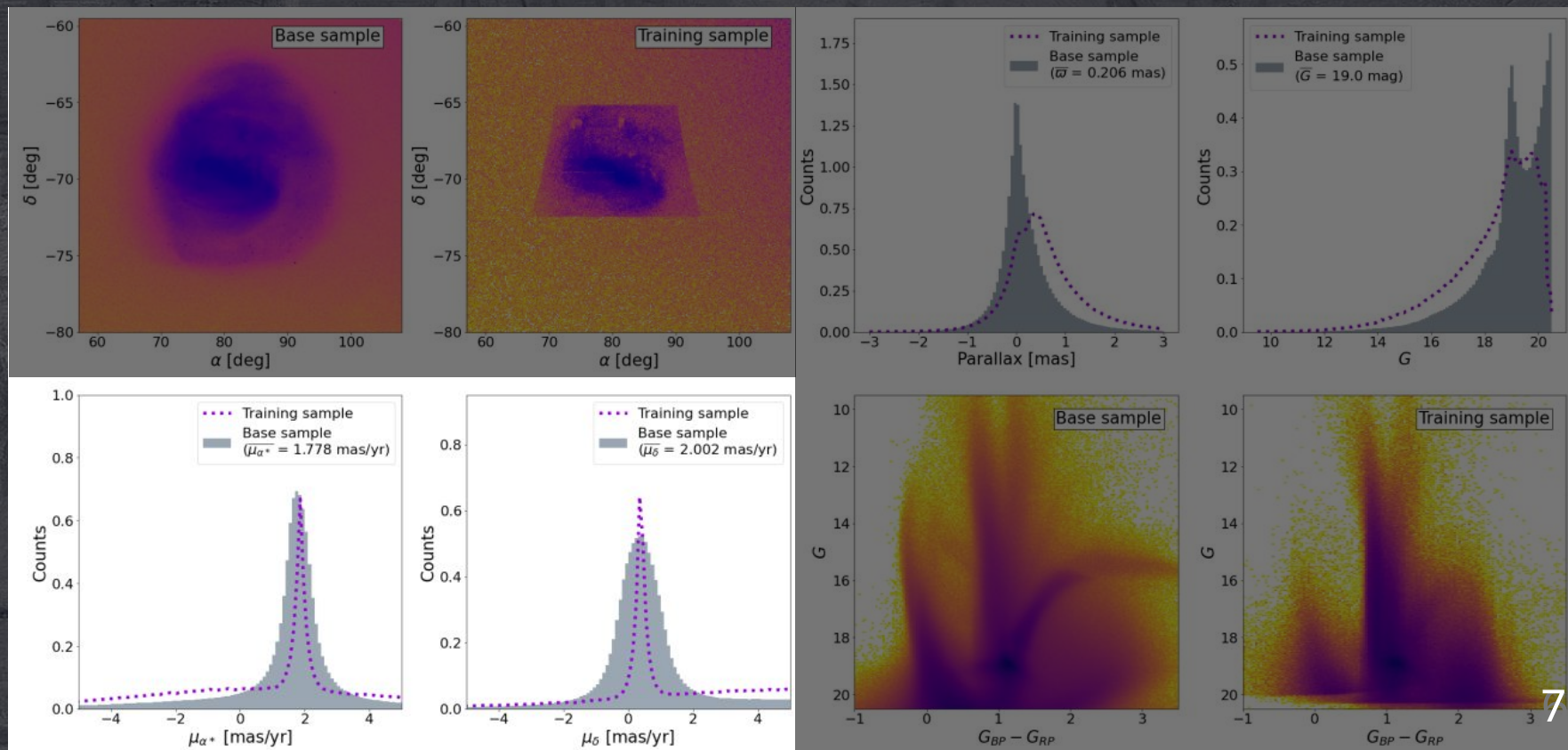
A hand is raised in the foreground, palm facing forward. In the background, several other hands are raised, though they are out of focus. A green speech bubble is positioned to the right of the hand, containing text. The background is a dark green wall, likely a chalkboard.

*How **representative**  
the **training sample** is in  
comparison to the **real**  
**data**?*

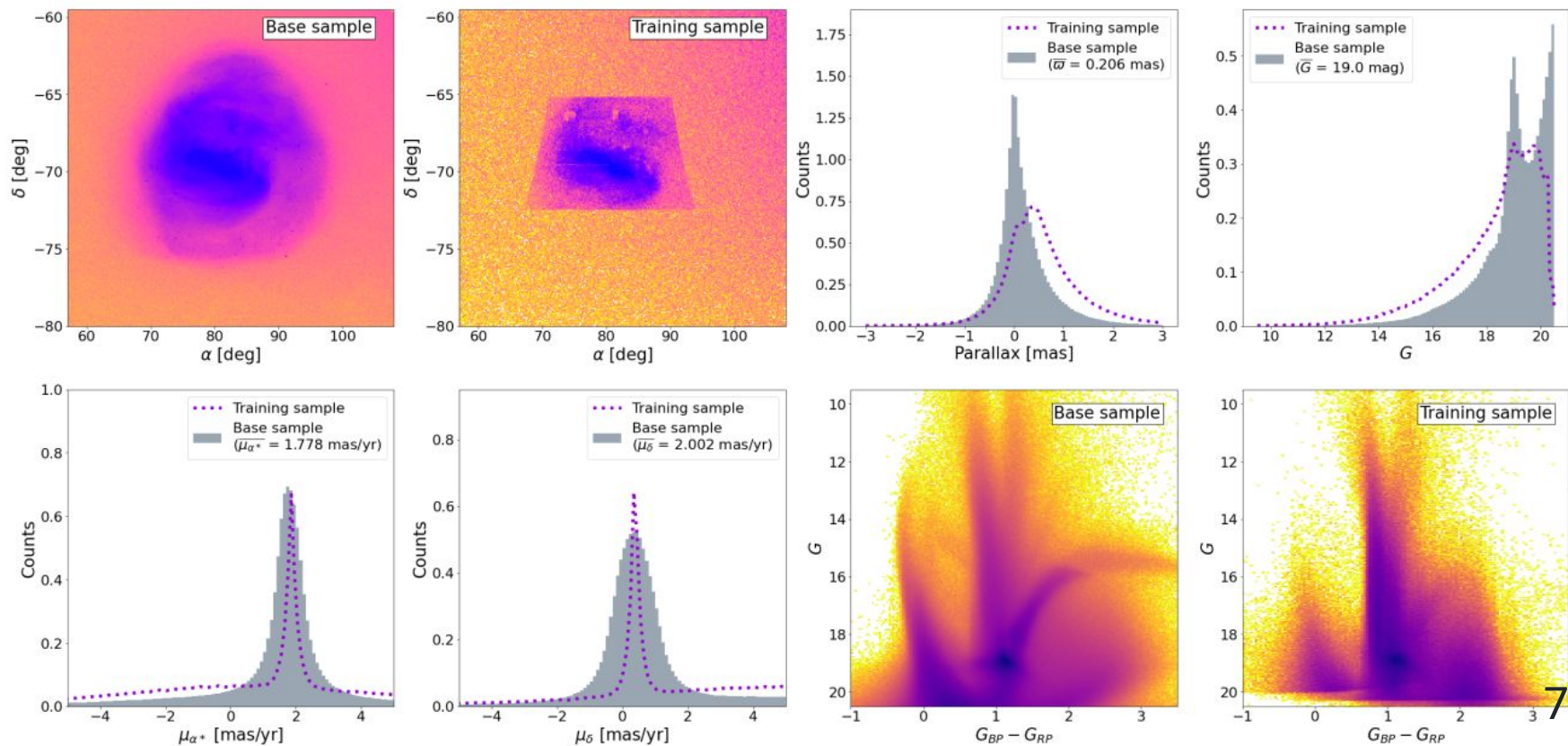
# LMC/MW classifier



# LMC/MW classifier



# LMC/MW classifier





## LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**

## LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**
- **Neural Network**

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**
- **Neural Network**
  - **Input: Gaia astrometry and photometry (11 variables)**

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**
- **Neural Network**
  - **Input: Gaia astrometry and photometry (11 variables)**

Position and  
velocities

Colors of the  
stars

# LMC/MW classifier

- **Classification problem** → **Supervised learning** → **Training sample**
- **Neural Network**
  - **Input: Gaia astrometry and photometry (11 variables)**
  - **Output: Probability P of being a LMC star**

# LMC/MW classifier

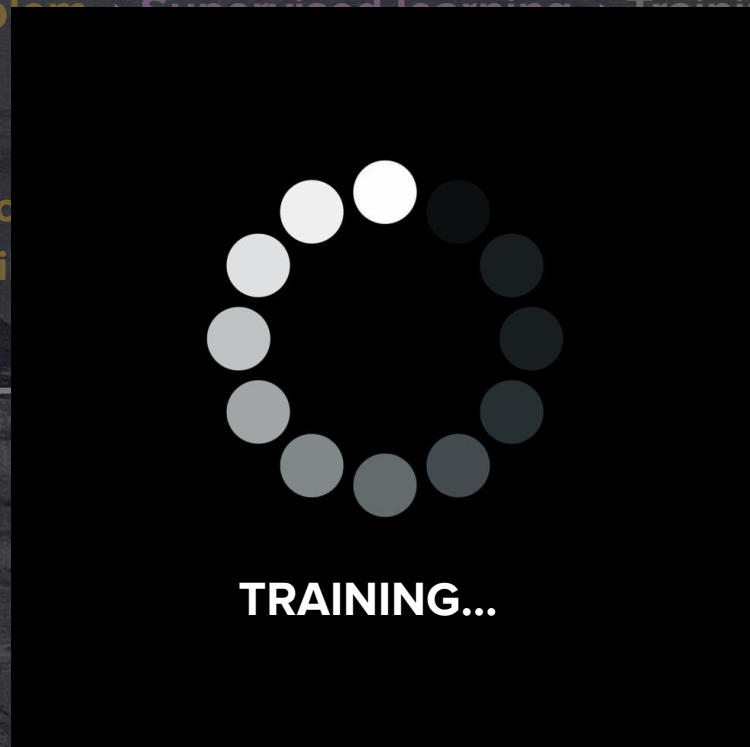
- **Classification problem** → **Supervised learning** → **Training sample**
- **Neural Network**
  - **Input:** Gaia **astrometry** and **photometry** (11 variables)
  - **Output:** **Probability P** of being a **LMC star**



# LMC/MW classifier

- Classification problem: Supervised learning: Training sample
- Neural Network
  - Input: Gaia astrometry
  - Output: Probability

MW 0



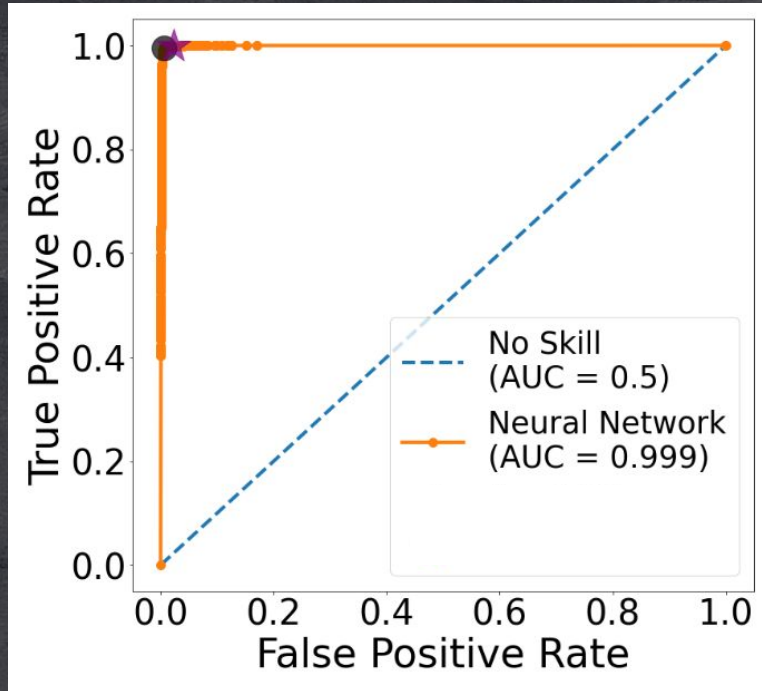
# LMC/MW classifier

Applying it to the **test sample**



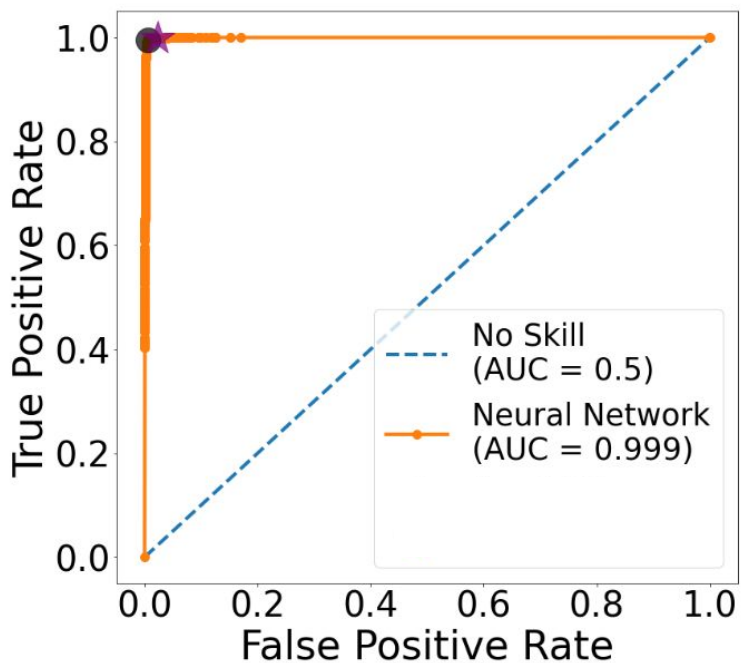
# LMC/MW classifier

ROC curve

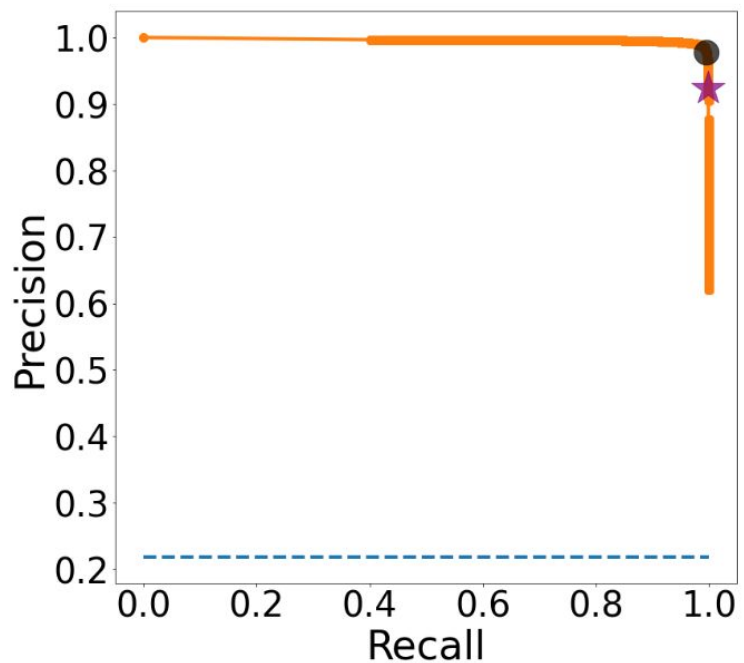


# LMC/MW classifier

ROC curve



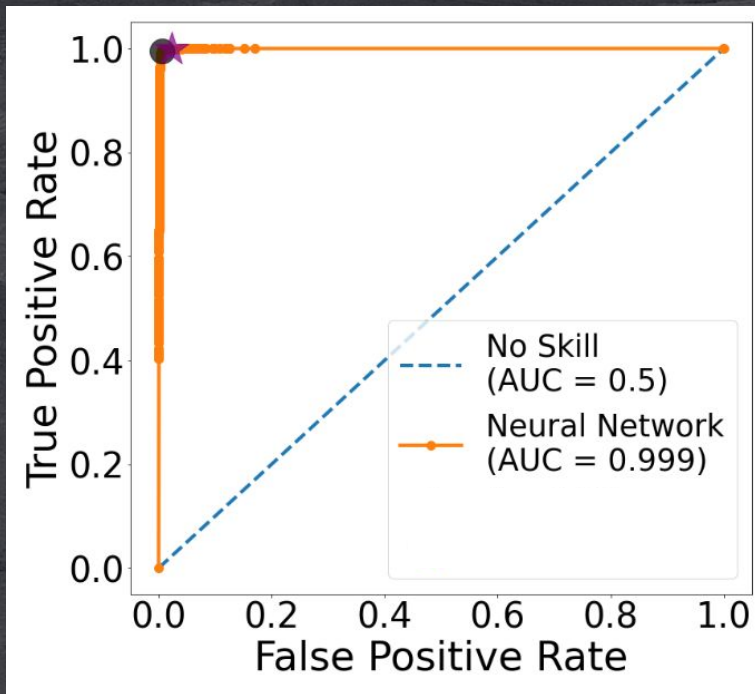
Precision-Recall curve



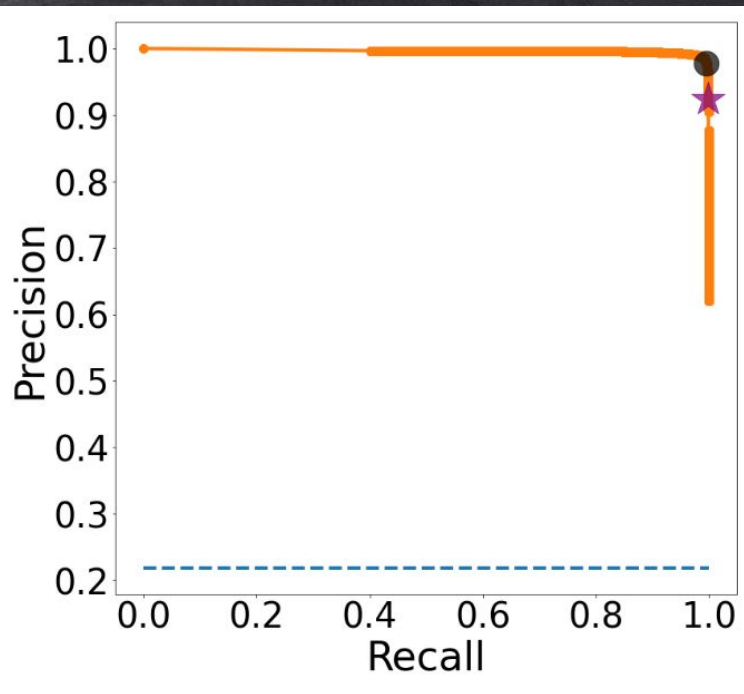
# LMC/MW classifier



ROC curve



Precision-Recall curve

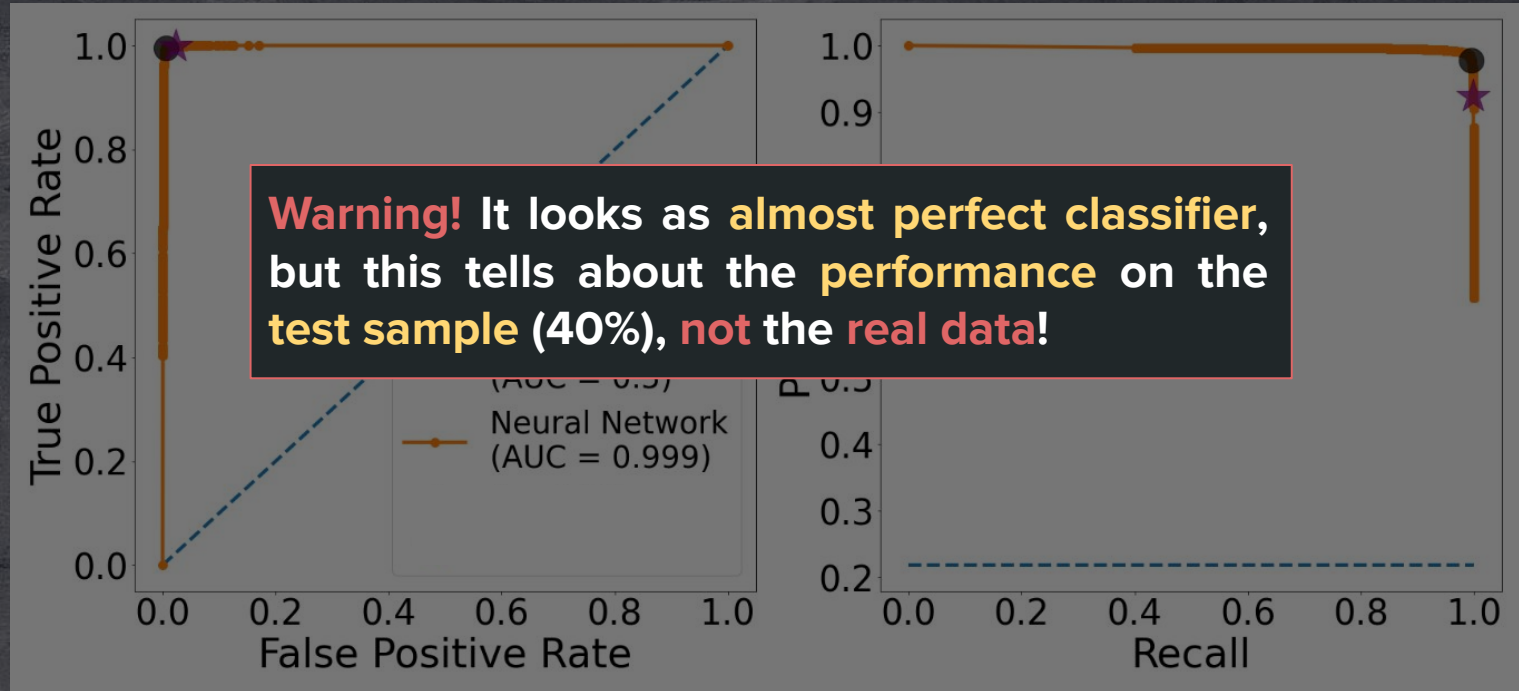


# LMC/MW classifier



ROC curve

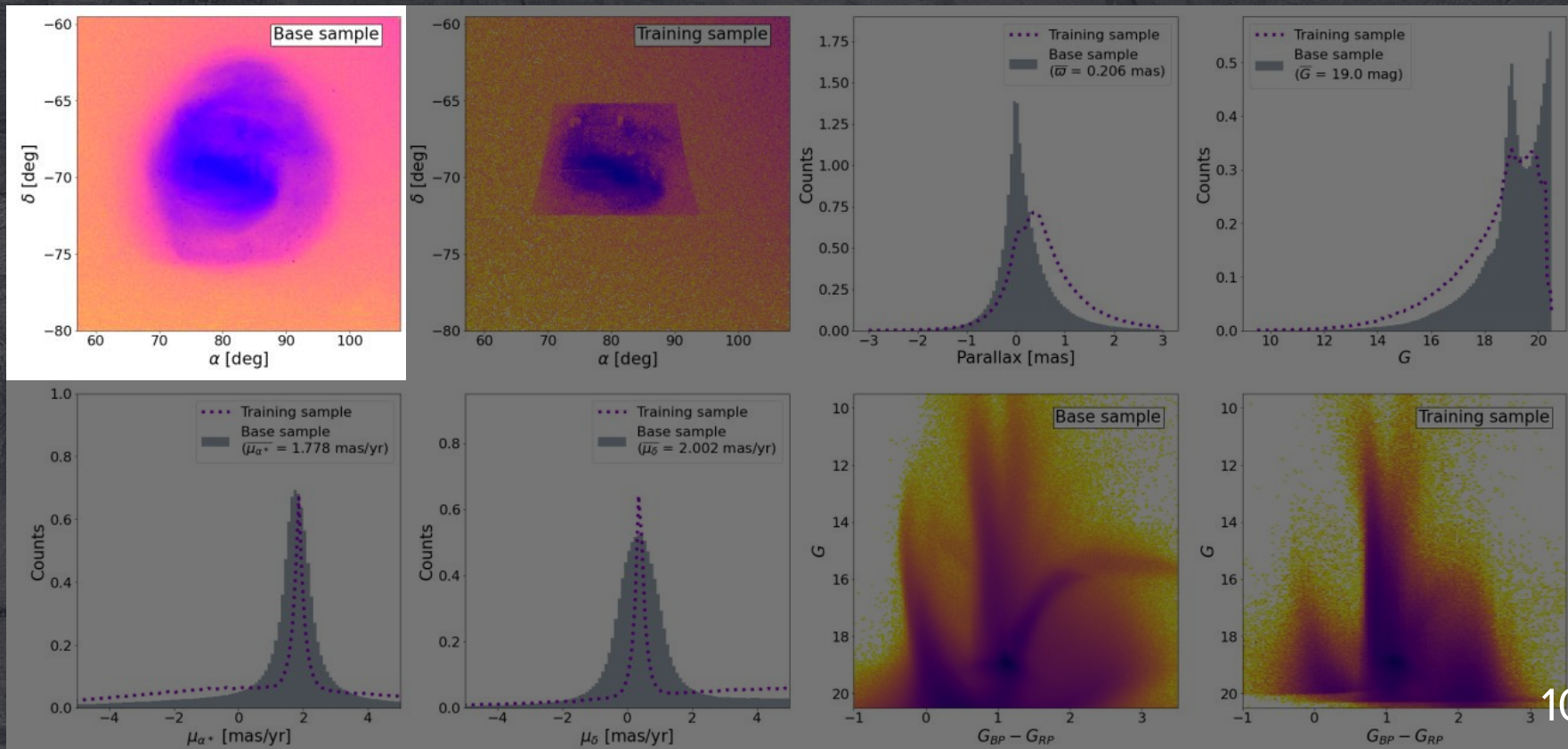
Precision-Recall curve



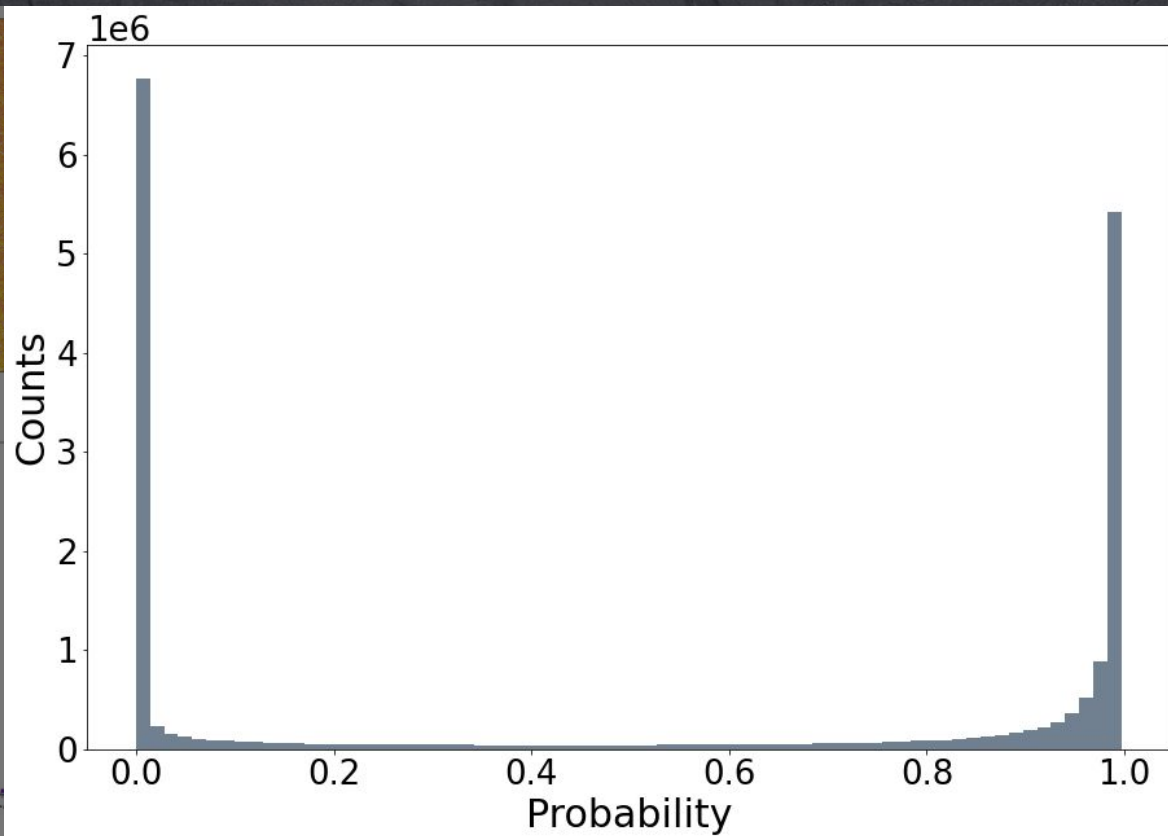
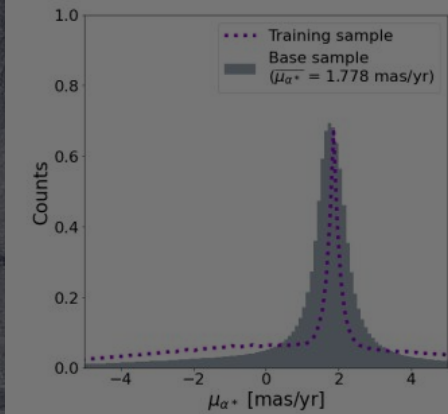
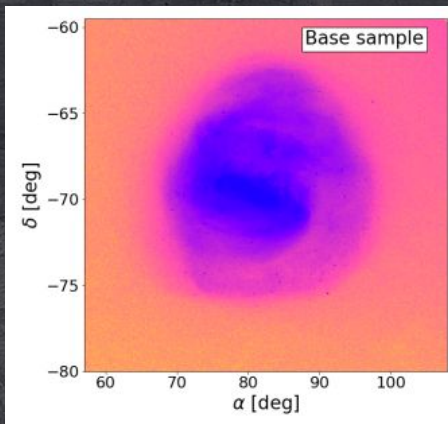
# LMC/MW classifier

Applying it to the **data**

# LMC/MW classifier



# LMC/MW classifier

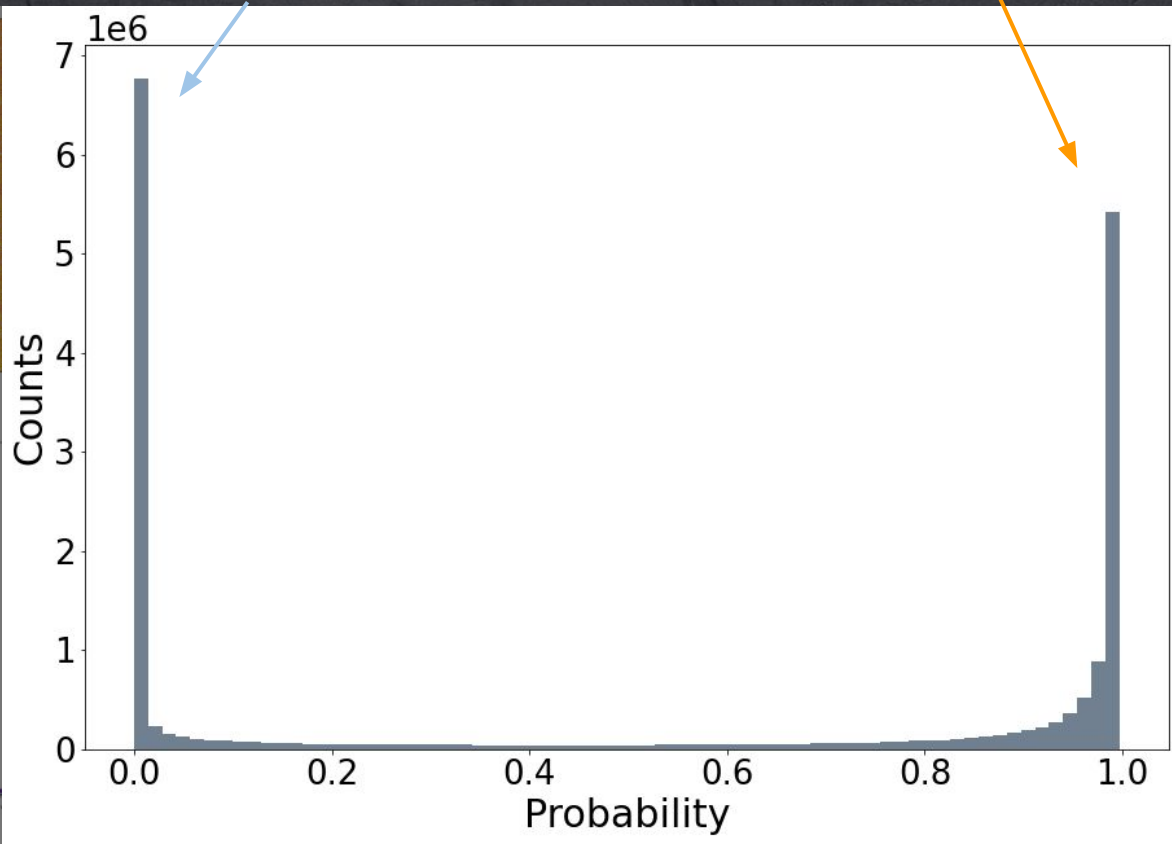
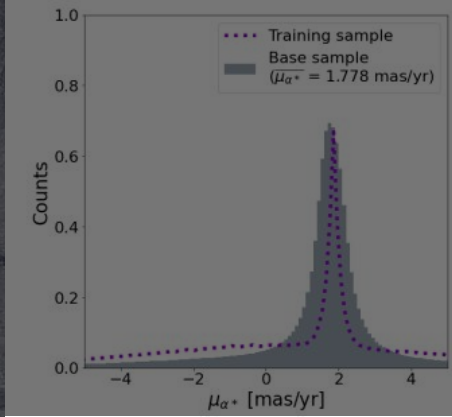
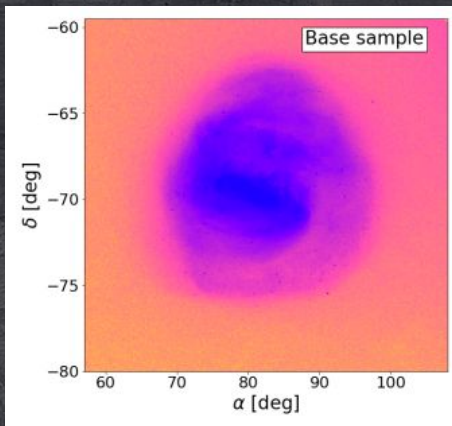


# Applying it to the data

## LMC/MW classifier

MW

LMC



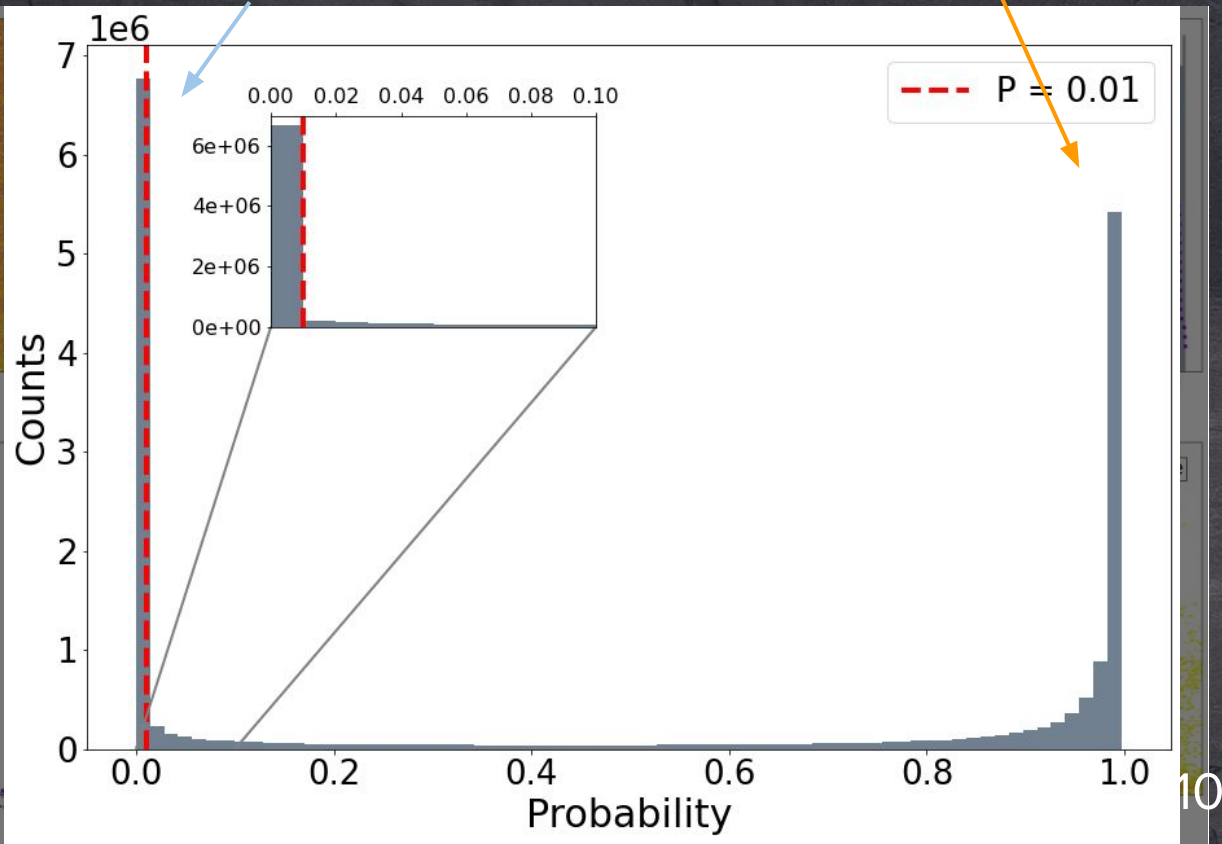
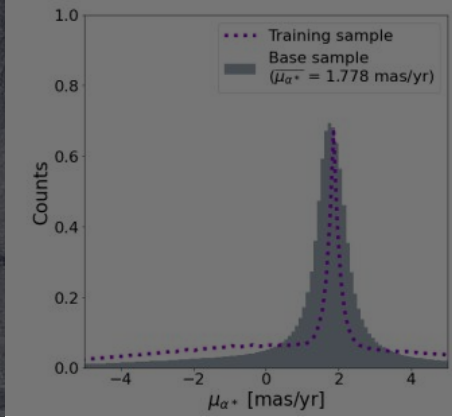
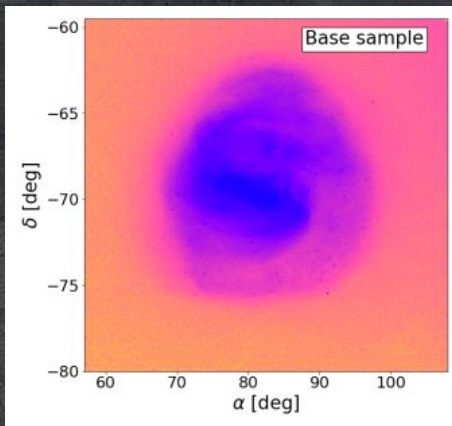


Applying it to the data

# LMC/MW classifier

MW

LMC

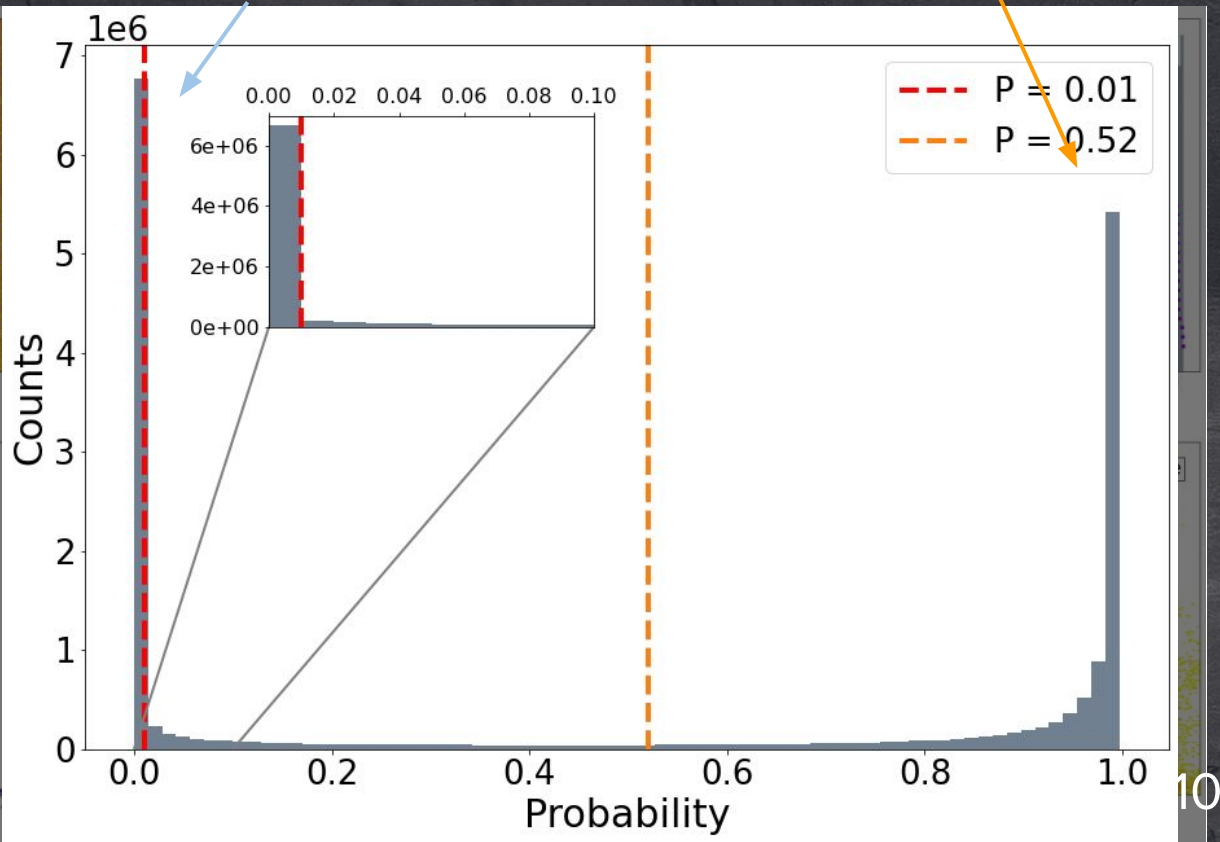
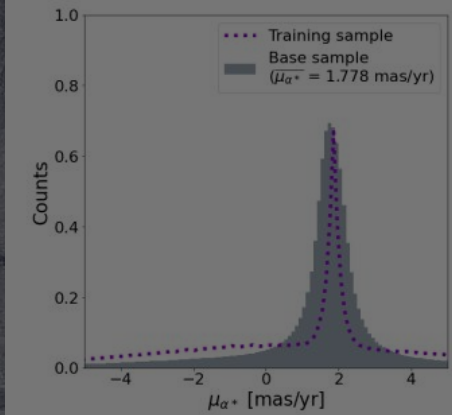
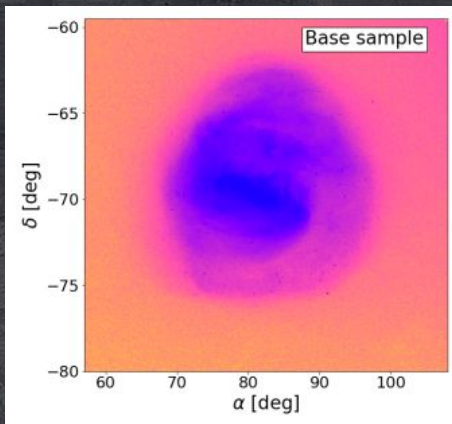


Applying it to the data

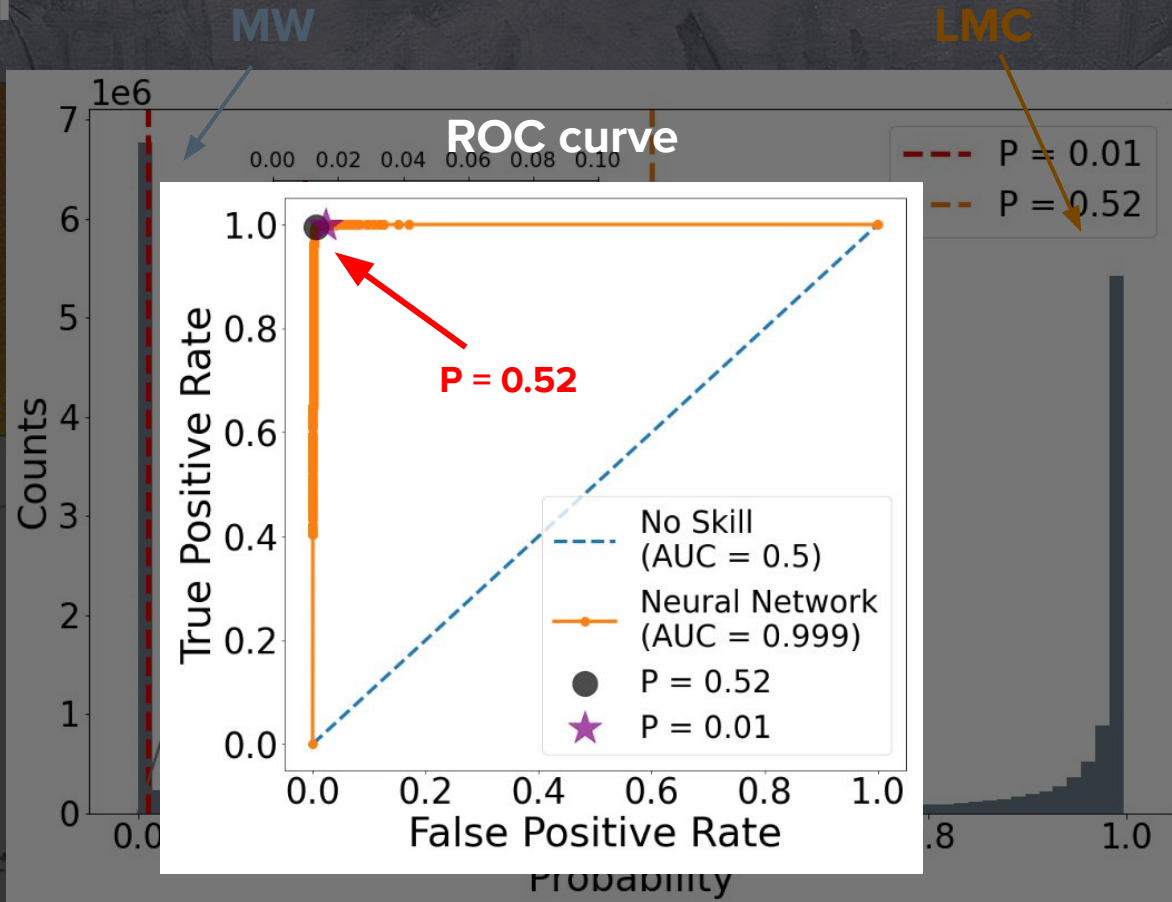
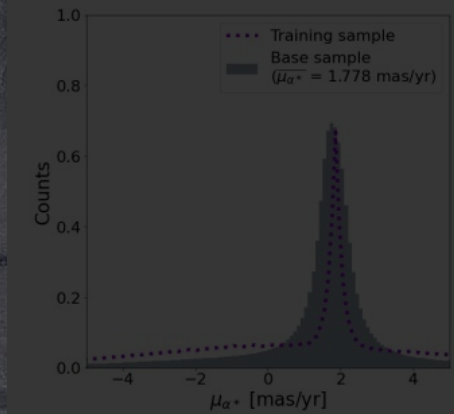
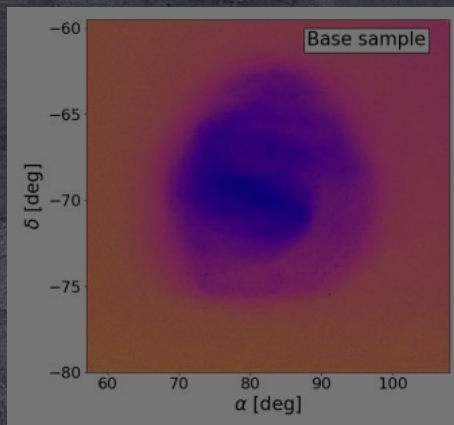
# LMC/MW classifier

MW

LMC



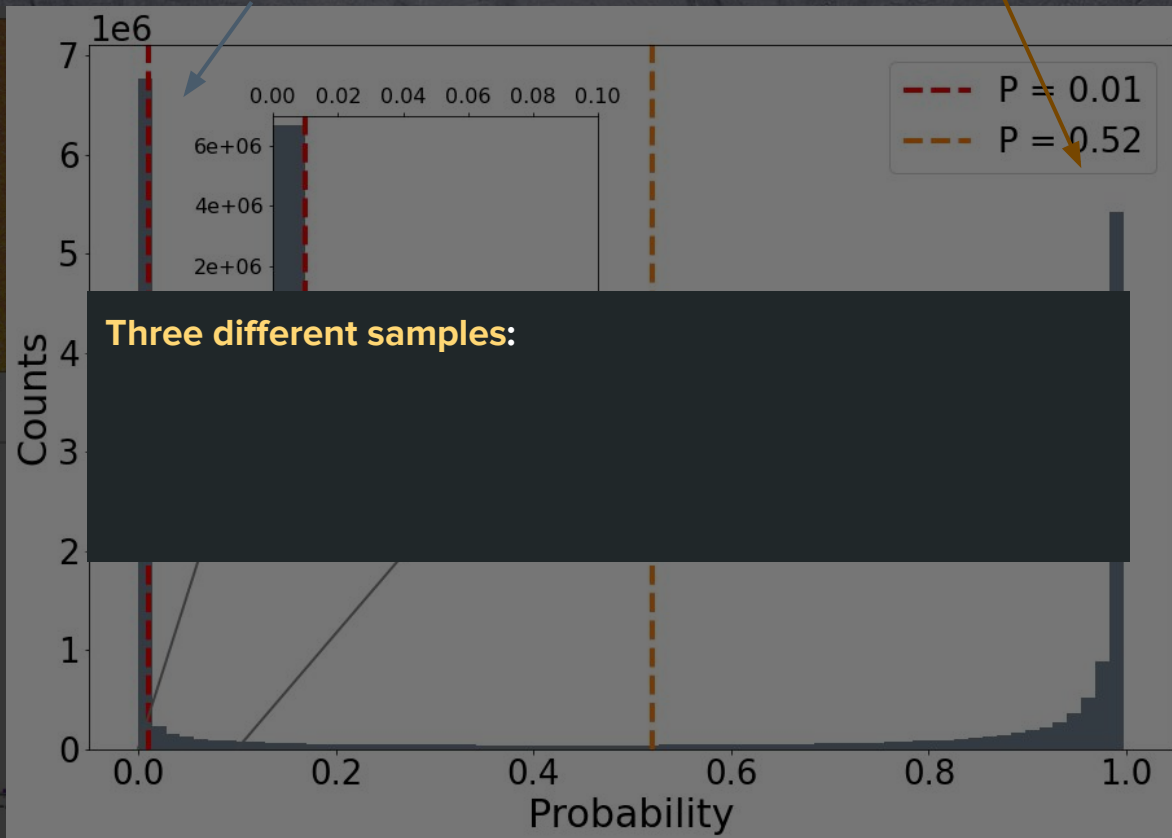
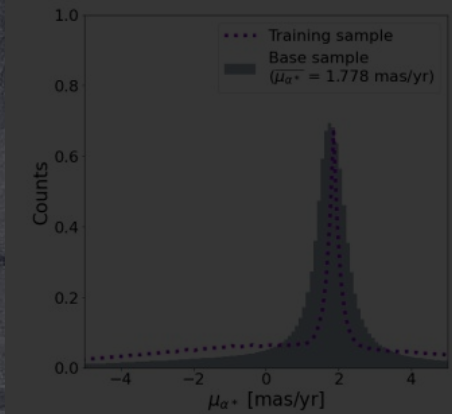
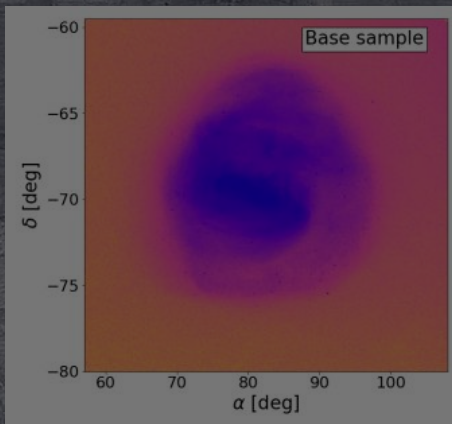
# LMC/MW classifier



# LMC/MW classifier

MW

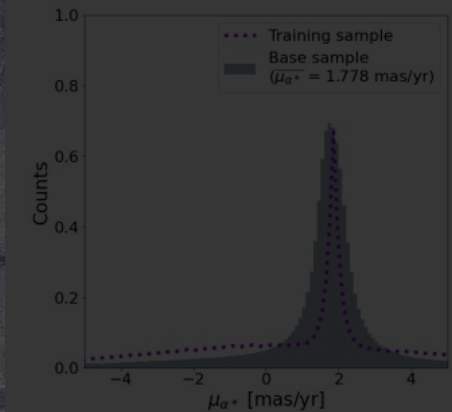
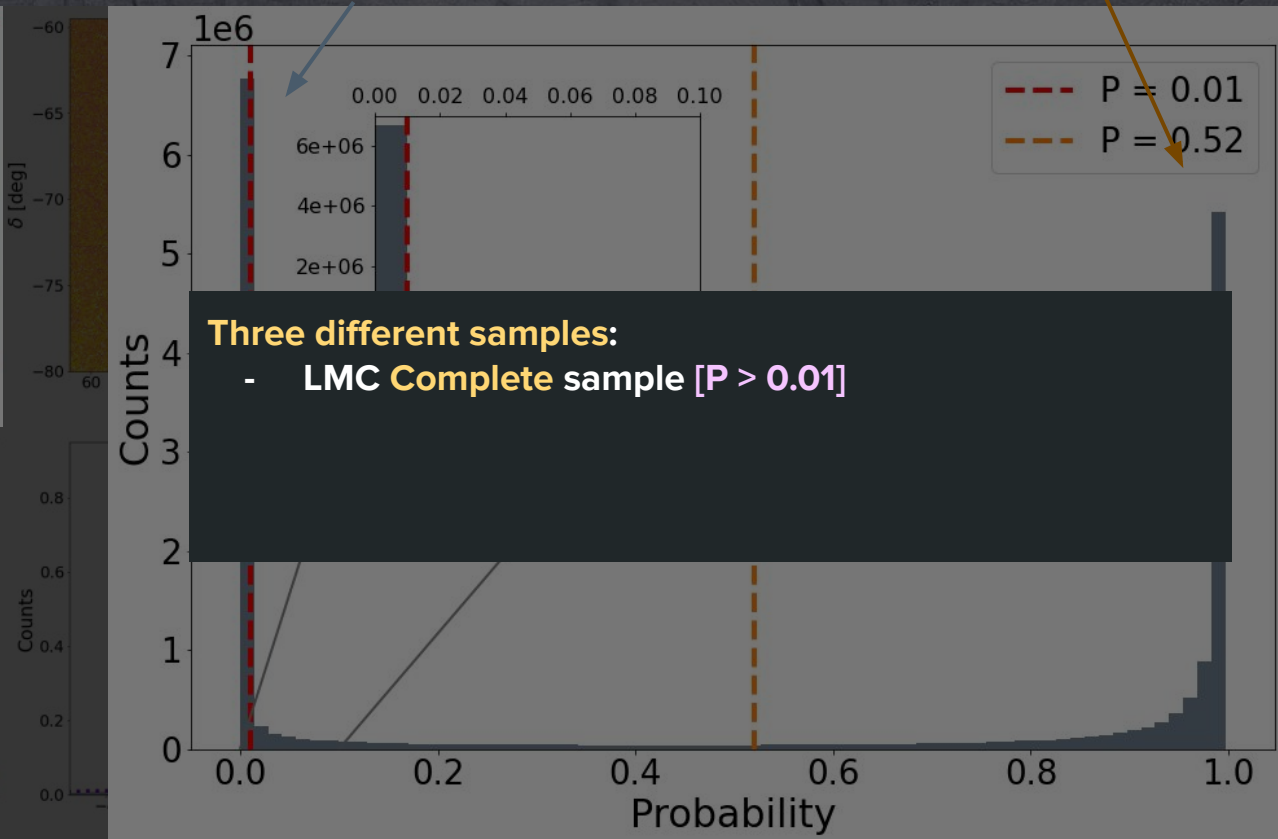
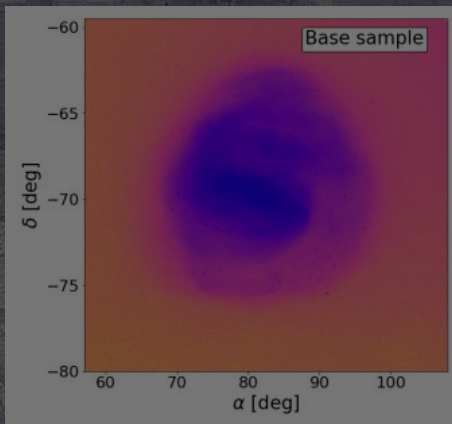
LMC



# LMC/MW classifier

MW

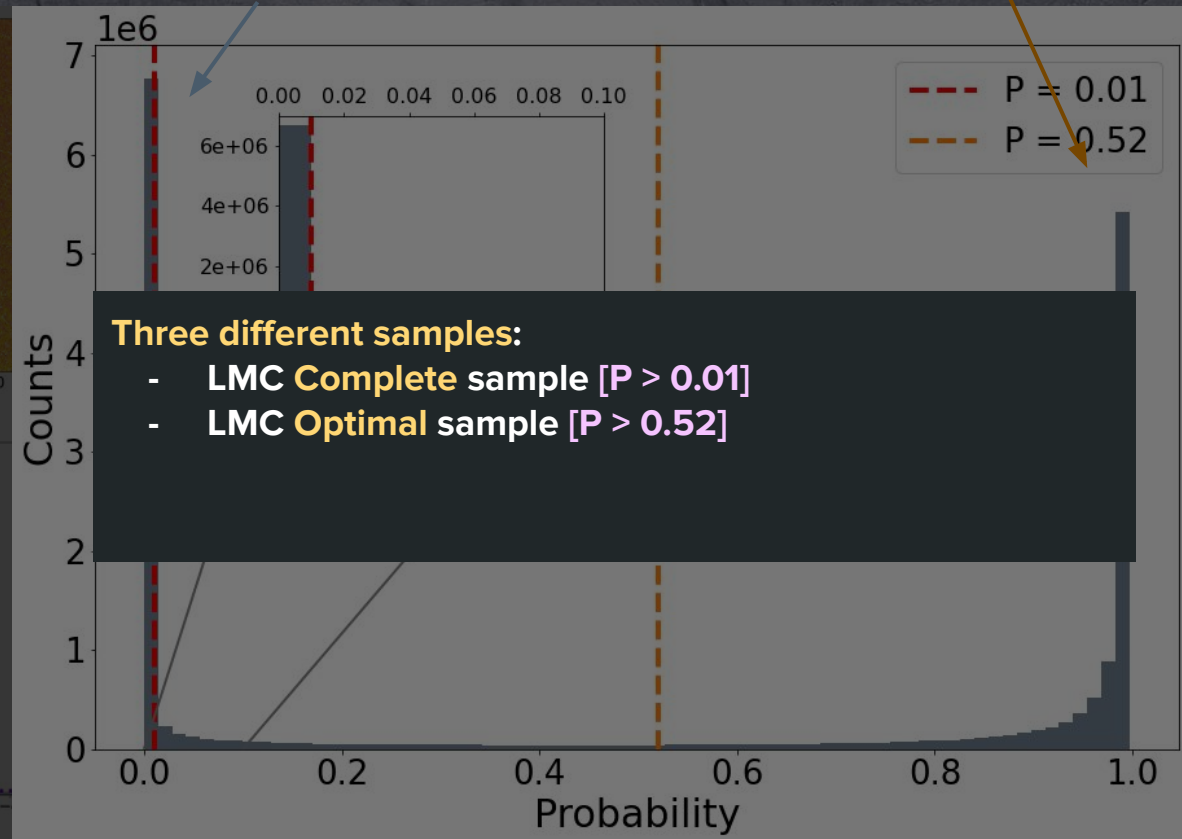
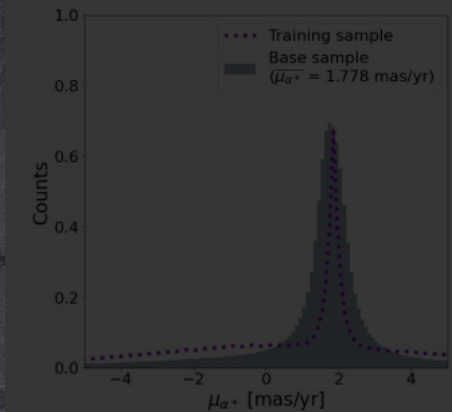
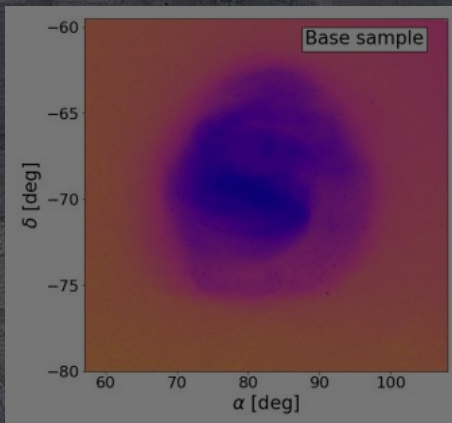
LMC



# LMC/MW classifier

MW

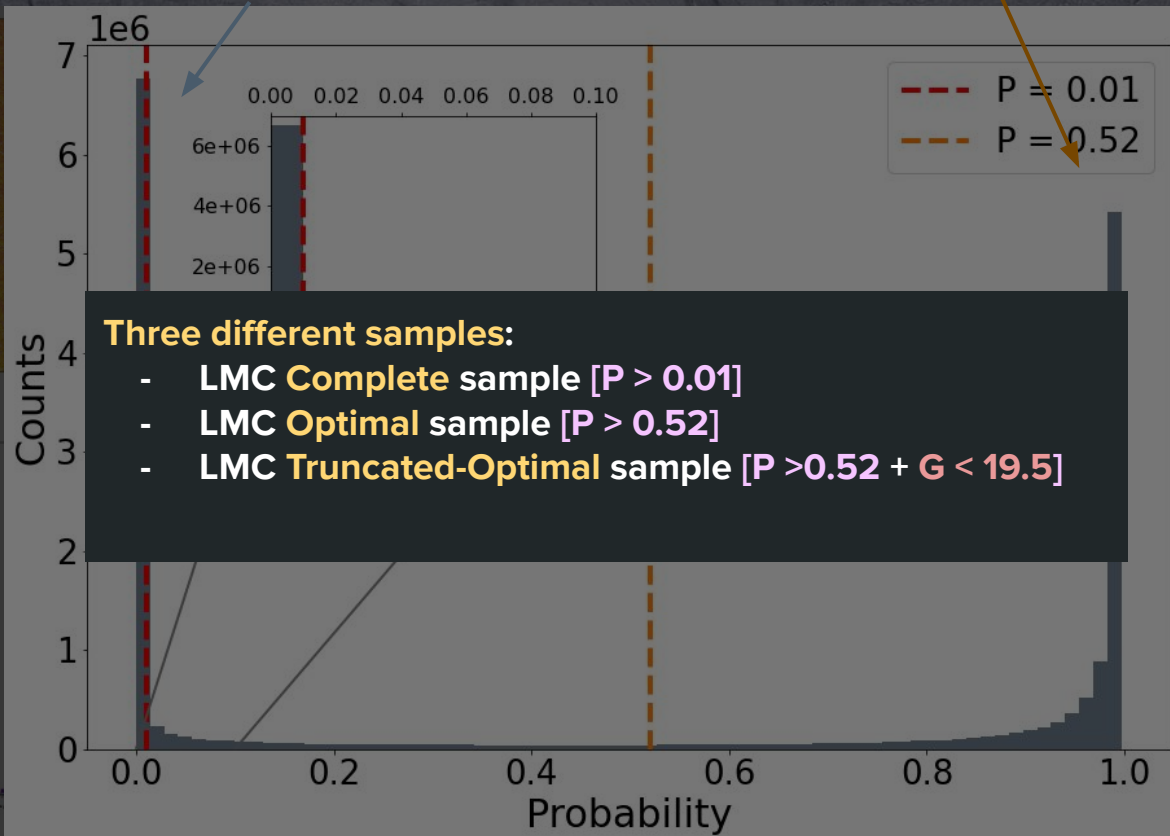
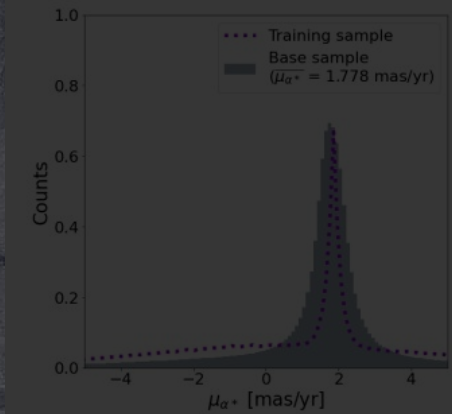
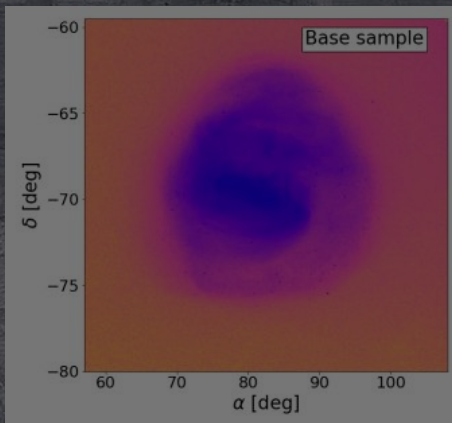
LMC



# LMC/MW classifier

MW

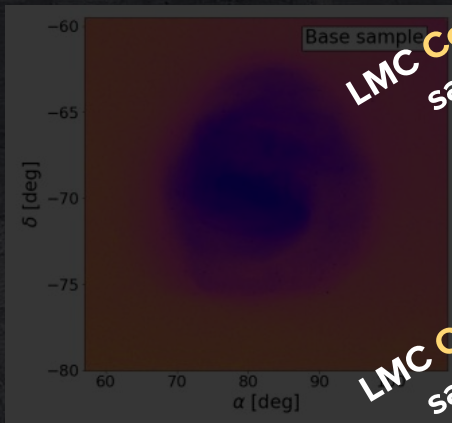
LMC



## Three different samples:

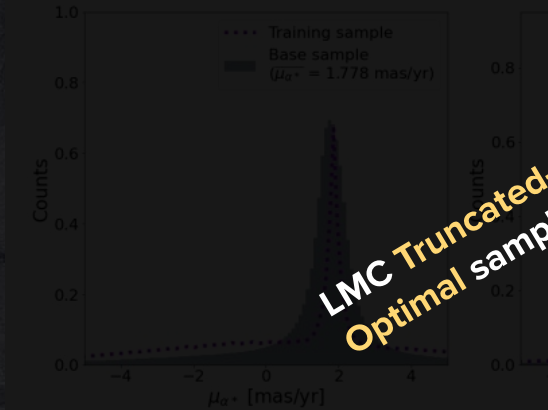
- LMC Complete sample [ $P > 0.01$ ]
- LMC Optimal sample [ $P > 0.52$ ]
- LMC Truncated-Optimal sample [ $P > 0.52 + G < 19.5$ ]

# LMC/MW classifier

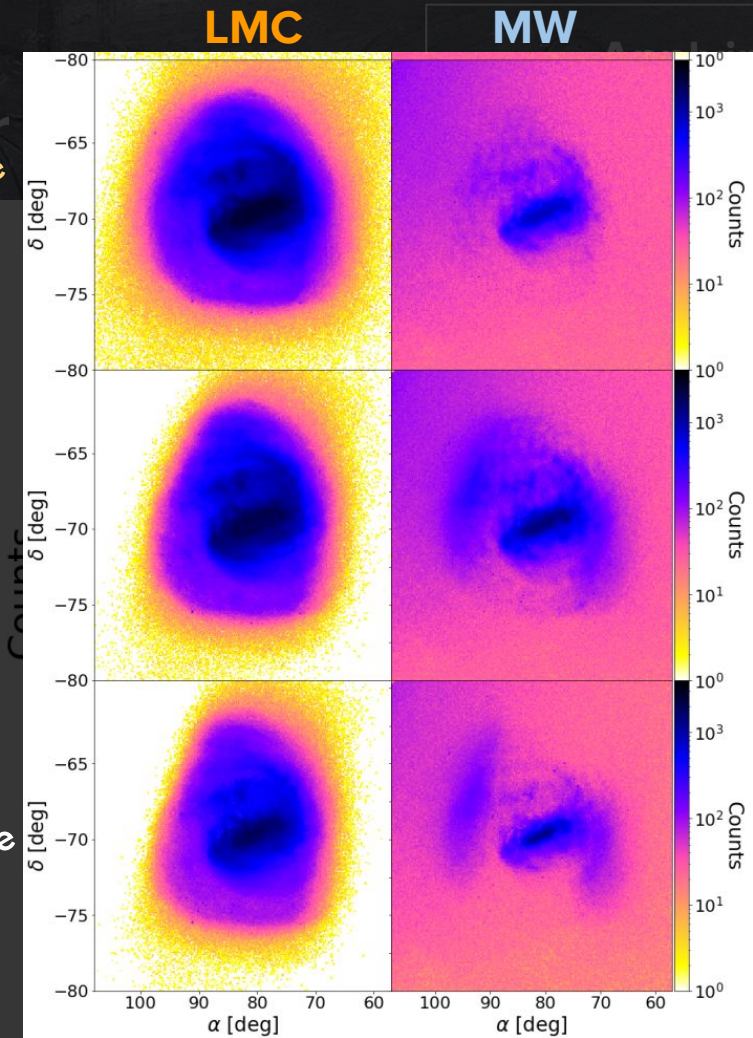


LMC Complete sample

LMC Optimal sample



LMC Truncated-Optimal sample

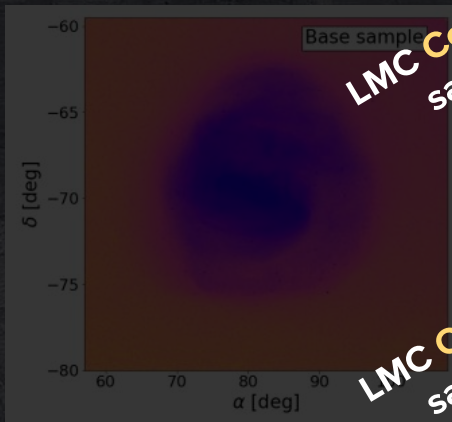


Adding it to the data



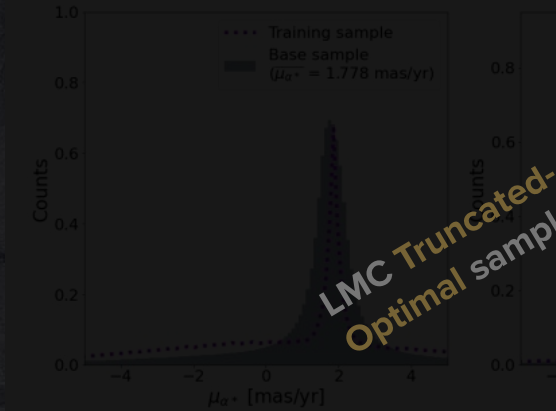


# LMC/MW classifier

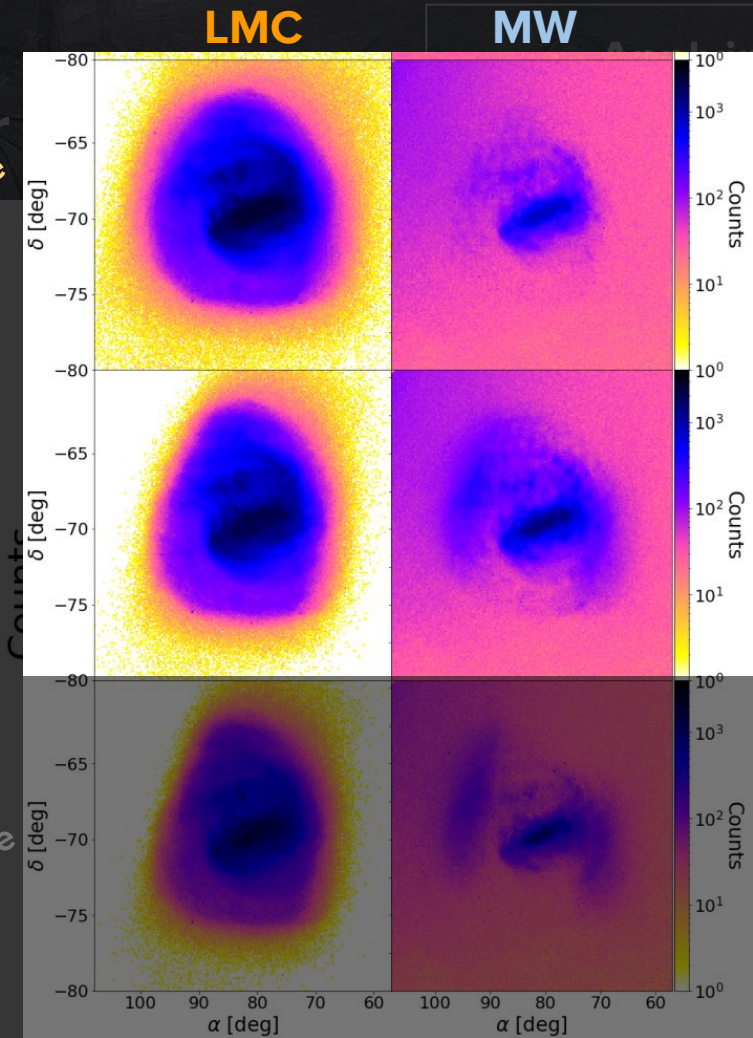


LMC Complete sample

LMC Optimal sample



LMC Truncated-Optimal sample



fitting it to the data







Base sample



Complete sample



Optimal sample



Truncated-optimal sample

**18M stars  
(LMC+MW)**



Base sample

**18M stars**  
**(LMC+MW)**



Complete sample

**12M stars**



Optimal sample

**10M stars**



Truncated-optimal sample

**6M stars**



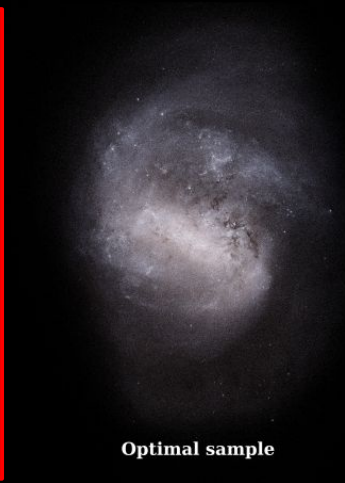
Base sample

**18M stars  
(LMC+MW)**



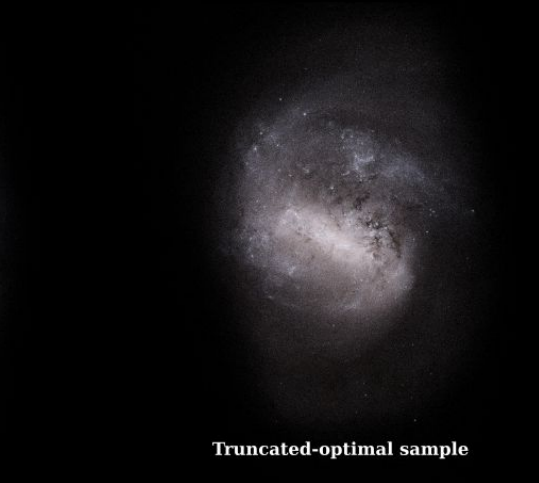
Complete sample

**12M stars**



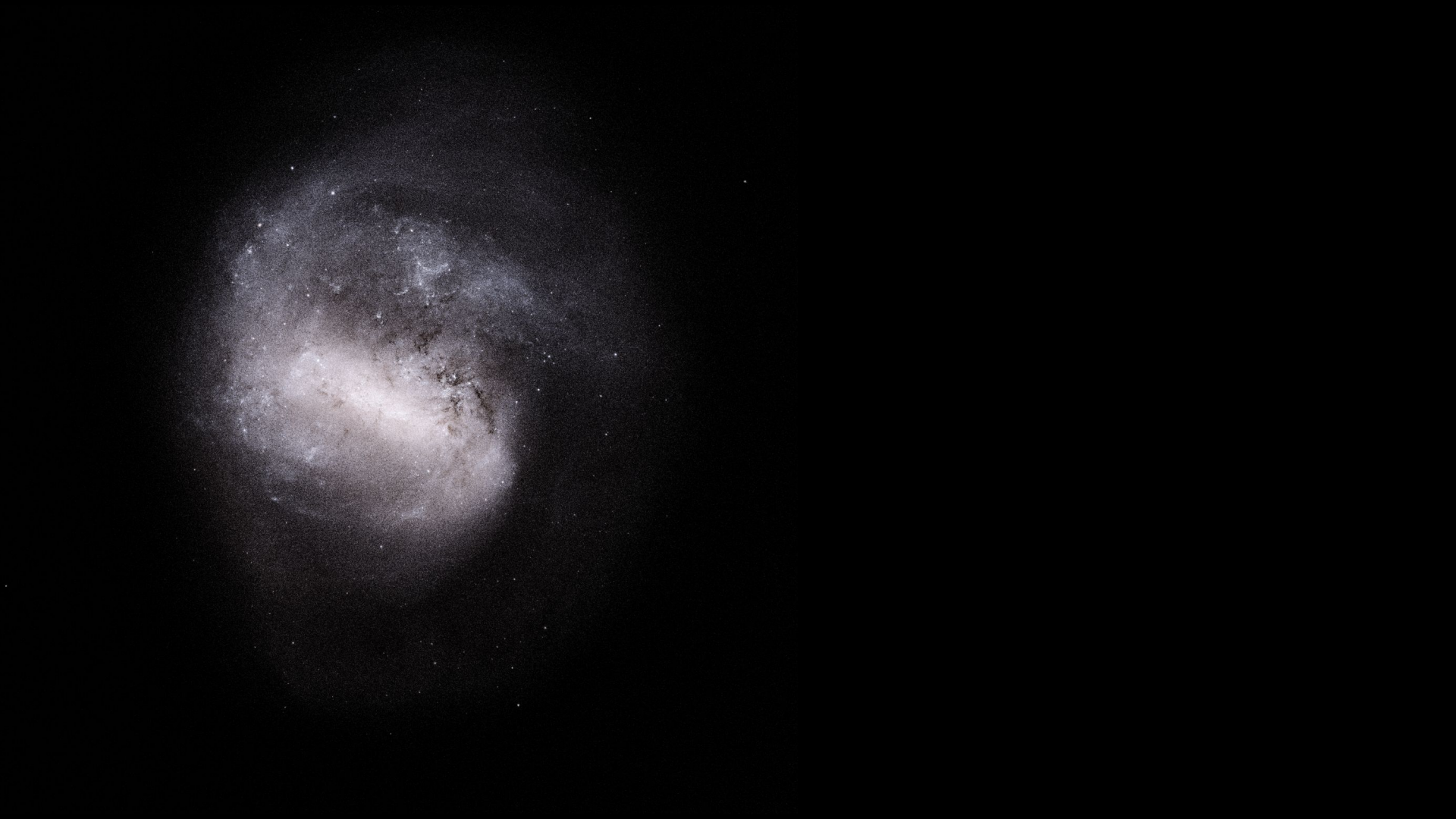
Optimal sample

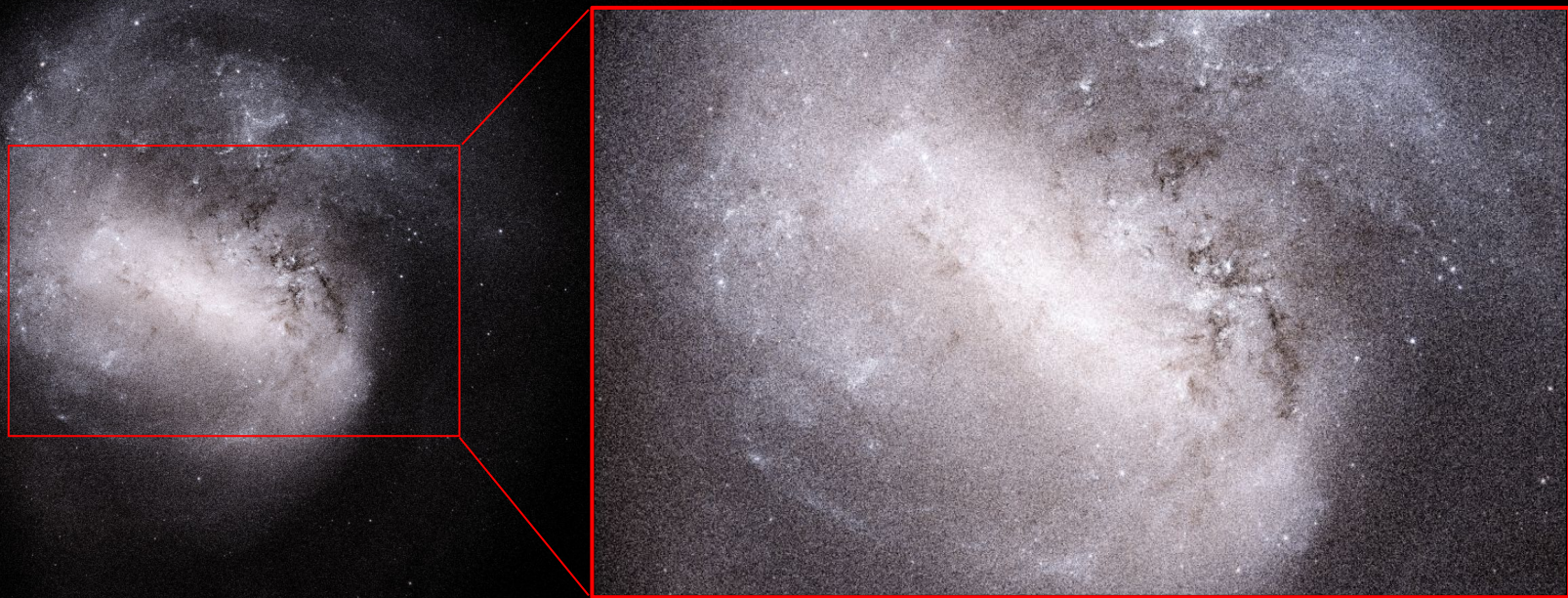
**10M stars**




Truncated-optimal sample

**6M stars**





A hand is raised in the foreground, with several other hands raised in the background, all against a green chalkboard. A green speech bubble is positioned in the upper right area of the image.

*How can we be sure that the classifier **works well** for the **real data**?*



# LMC/MW classifier

External validation

# LMC/MW classifier

- Three **external independent** classifications

# LMC/MW classifier

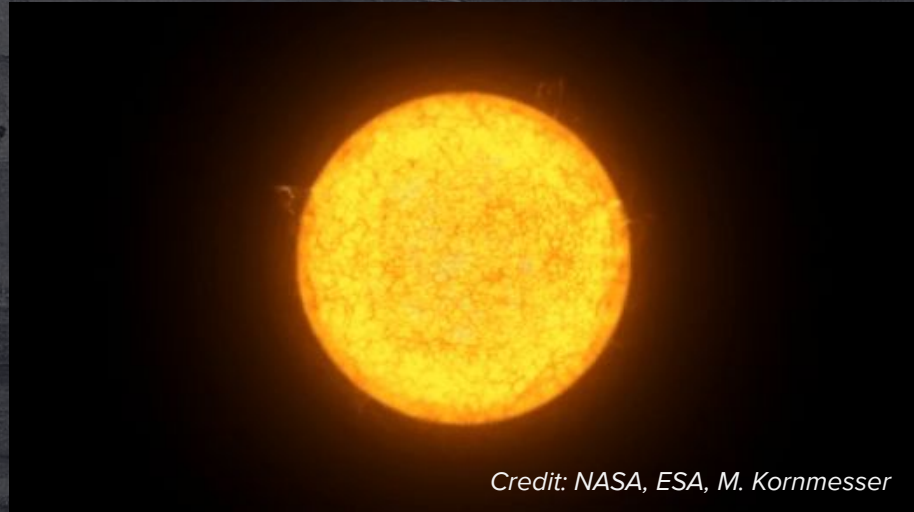
- Three **external independent** classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)

# LMC/MW classifier

- Three **external independent** classifications
    - LMC Cepheids (4.5k stars)
    - LMC RR Lyrae (21k stars)
- ← Variable stars

# LMC/MW classifier

- Three **external independent** classifications
    - LMC Cepheids (4.5k stars)
    - LMC RR Lyrae (21k stars)
- ← Variable stars



*Credit: NASA, ESA, M. Kornmesser*

# LMC/MW classifier

- Three **external independent** classifications
  - LMC Cepheids (4.5k stars) ← Variable stars
  - LMC RR Lyrae (21k stars) ← Variable stars
  - LMC/MW StarHorse (1M / 3M stars)

# LMC/MW classifier

- Three external independent classifications

- LMC Cepheids (4.5k stars)
- LMC RR Lyrae (21k stars)
- LMC/MW StarHorse (1M / 3M stars)

Variable stars

Bayesian inference  
(distance)

## LMC/MW classifier

- Three external independent classifications
  - LMC Cepheids (4.5k stars) ← Variable stars

Stars classified as LMC	LMC Cepheids (4 467)	LMC RR-Lyrae (21 271)	LMC StarHorse (985 173)	MW StarHorse (2 940 282)
NN complete	4 407 (98.7%)	20 223 (95.1%)	970 719 (98.5%)	722 750 (24.6%)
NN optimal	4 160 (93.1%)	17 860 (84.0%)	832 733 (84.5%)	627 619 (21.3%)
NN truncated-optimal	4 160 (93.1%)	14 750 (69.3%)	832 733 (84.5%)	627 619 (21.3%)



## LMC/MW classifier

- Three external independent classifications
  - LMC Cepheids (4.5k stars) ← Variable stars

Stars classified as LMC	LMC Cepheids (4 467)	LMC RR-Lyrae (21 271)	LMC StarHorse (985 173)	MW StarHorse (2 940 282)
NN complete	4 407 (98.7%)	20 223 (95.1%)	970 719 (98.5%)	722 750 (24.6%)
NN optimal	4 160 (93.1%)	17 860 (84.0%)	832 733 (84.5%)	627 619 (21.3%)
NN truncated-optimal	4 160 (93.1%)	14 750 (69.3%)	832 733 (84.5%)	627 619 (21.3%)

## LMC/MW classifier

- Three external independent classifications
  - LMC Cepheids (4.5k stars) ← Variable stars

Stars classified as LMC	LMC Cepheids (4 467)	LMC RR-Lyrae (21 271)	LMC StarHorse (985 173)	MW StarHorse (2 940 282)
NN complete	4 407 (98.7%)	20 223 (95.1%)	970 719 (98.5%)	722 750 (24.6%)
NN optimal	4 160 (93.1%)	17 860 (84.0%)	832 733 (84.5%)	627 619 (21.3%)
NN truncated-optimal	4 160 (93.1%)	14 750 (69.3%)	832 733 (84.5%)	627 619 (21.3%)

## LMC/MW classifier

- Three external independent classifications
  - LMC Cepheids (4.5k stars)

Variable stars

Stars classified as LMC	LMC Cepheids (4 467)	LMC RR-Lyrae (21 271)	LMC StarHorse (985 173)	MW StarHorse (2 940 282)
NN complete	4 407 (98.7%)	20 223 (95.1%)	970 719 (98.5%)	722 750 (24.6%)
NN optimal	4 160 (93.1%)	17 860 (84.0%)	832 733 (84.5%)	627 619 (21.3%)
NN truncated-optimal	4 160 (93.1%)	14 750 (69.3%)	832 733 (84.5%)	627 619 (21.3%)



# LMC/MW classifier

- Three external independent classifications
  - LMC Cepheids (4.5k stars) ← Variable stars

Stars classified as LMC	LMC Cepheids (4 467)	LMC RR-Lyrae (21 271)	LMC StarHorse (985 173)	MW StarHorse (2 940 282)
NN complete	4 407 (98.7%)	20 223 (95.1%)	970 719 (98.5%)	722 750 (24.6%)
NN optimal	4 160 (93.1%)	17 860 (84.0%)	832 733 (84.5%)	627 619 (21.3%)
NN truncated-optimal	4 160 (93.1%)	14 750 (69.3%)	832 733 (84.5%)	627 619 (21.3%)

# LMC/MW classifier

- Three **external independent** classifications
  - LMC Cepheids (4.5k stars) ← Variable stars
  - LMC RR Lyrae (21k stars) ← Variable stars
  - LMC/MW StarHorse (1M / 3M stars) ← Bayesian inference (distance)
- Gaia's **line-of-sight velocity** (20-30k)

# LMC/MW classifier

- Three **external independent** classifications

- LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW StarHorse (1M / 3M stars)
- ← Variable stars

← Bayesian inference  
(distance)

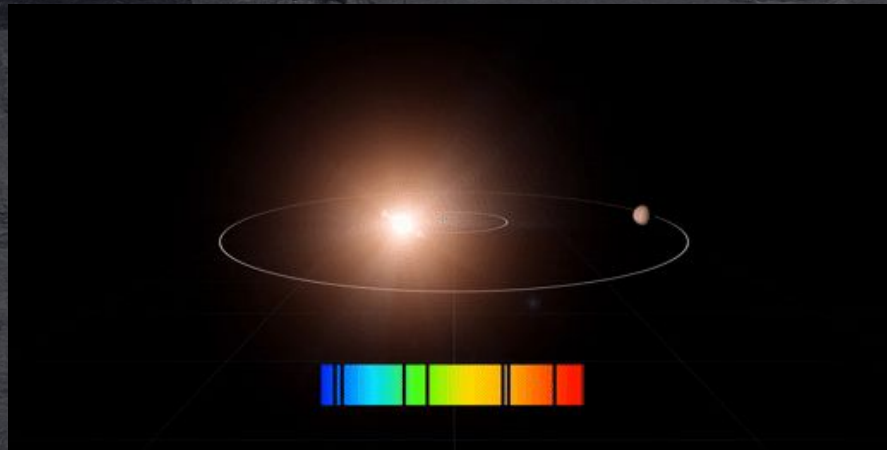
- Gaia's **line-of-sight velocity** (20-30k)

↗  
Only for the  
brightest stars!

# LMC/MW classifier

- Three **external independent** classifications
  - LMC Cepheids (4.5k stars) ← Variable stars
  - LMC RR Lyrae (21k stars) ← Variable stars
  - LMC/MW StarHorse (1M / 3M stars) ← Bayesian inference (distance)
- Gaia's **line-of-sight velocity** (20-30k)

Only for the  
brightest stars!



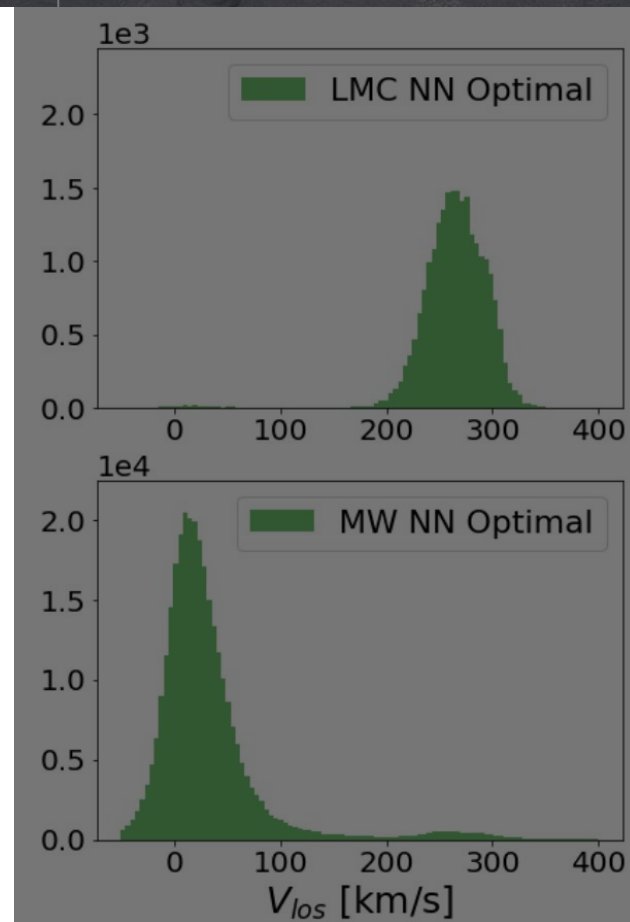
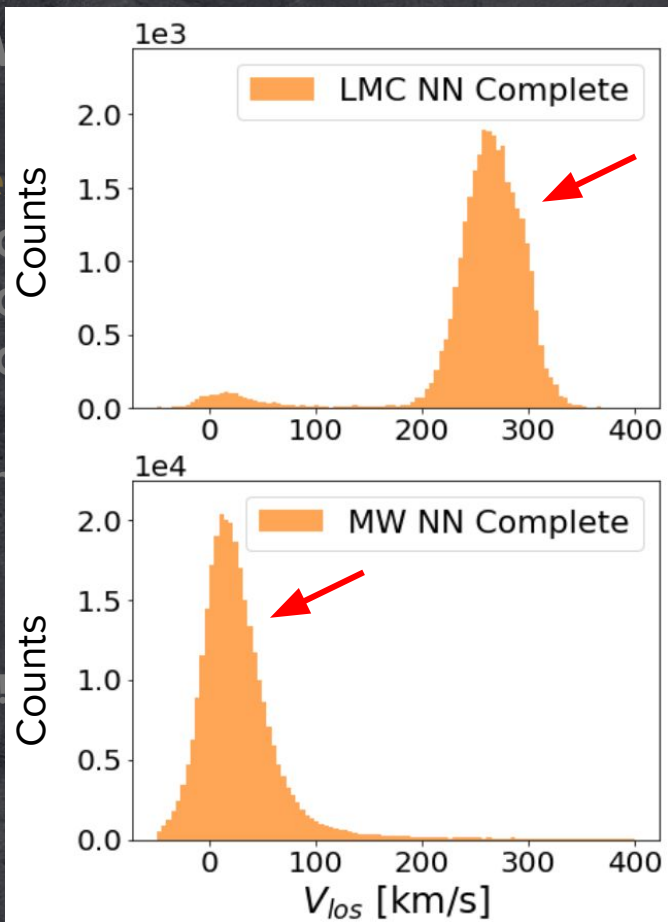
## External validation

### LMC/MW

- Three e
  - LMC
  - LMC
  - LMC

- Gaia's list

Only for the  
brightest stars!





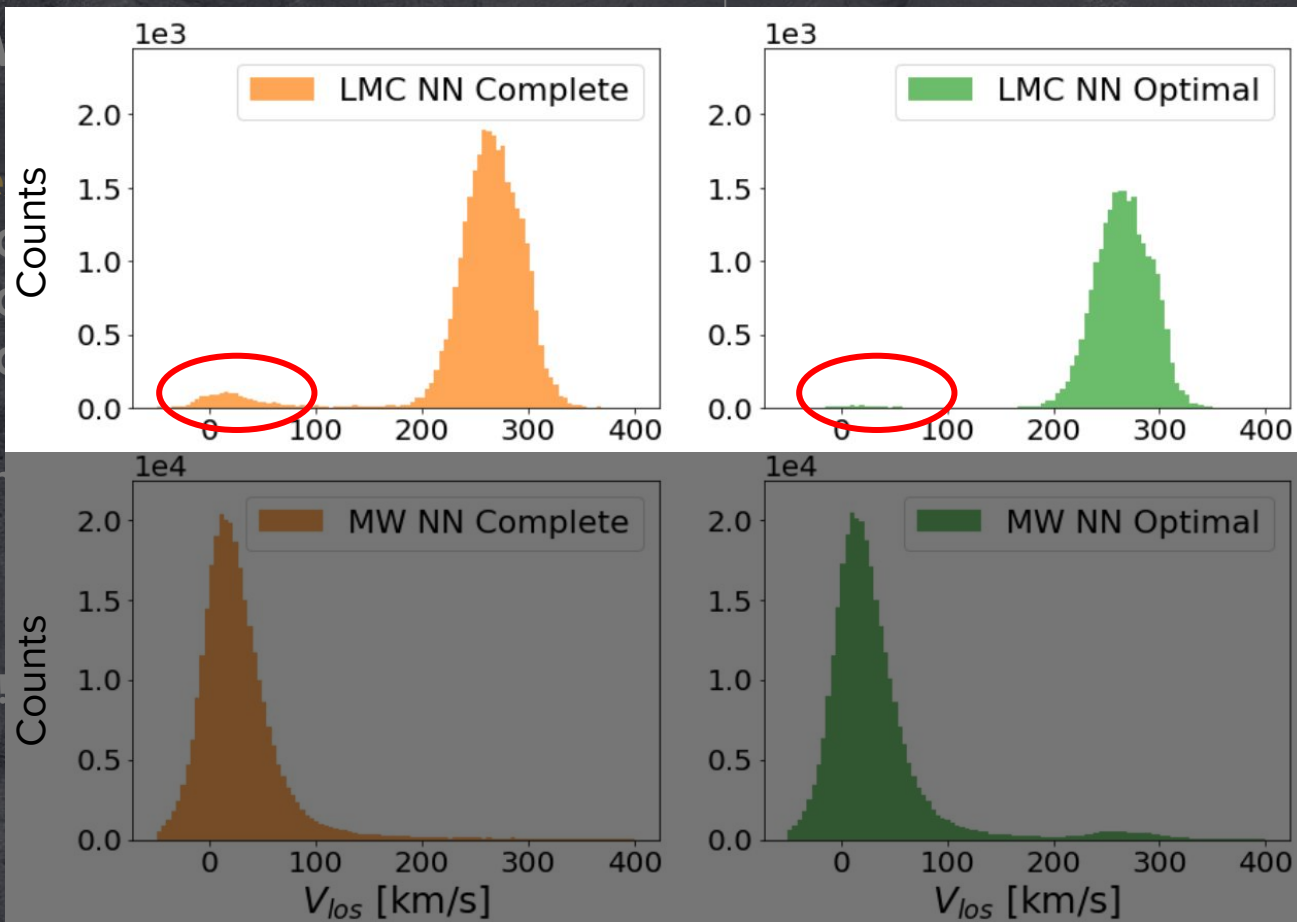
## External validation

### LMC/MW

- Three e
  - LMC
  - LMC
  - LMC

- Gaia's list

Only for the  
brightest stars!



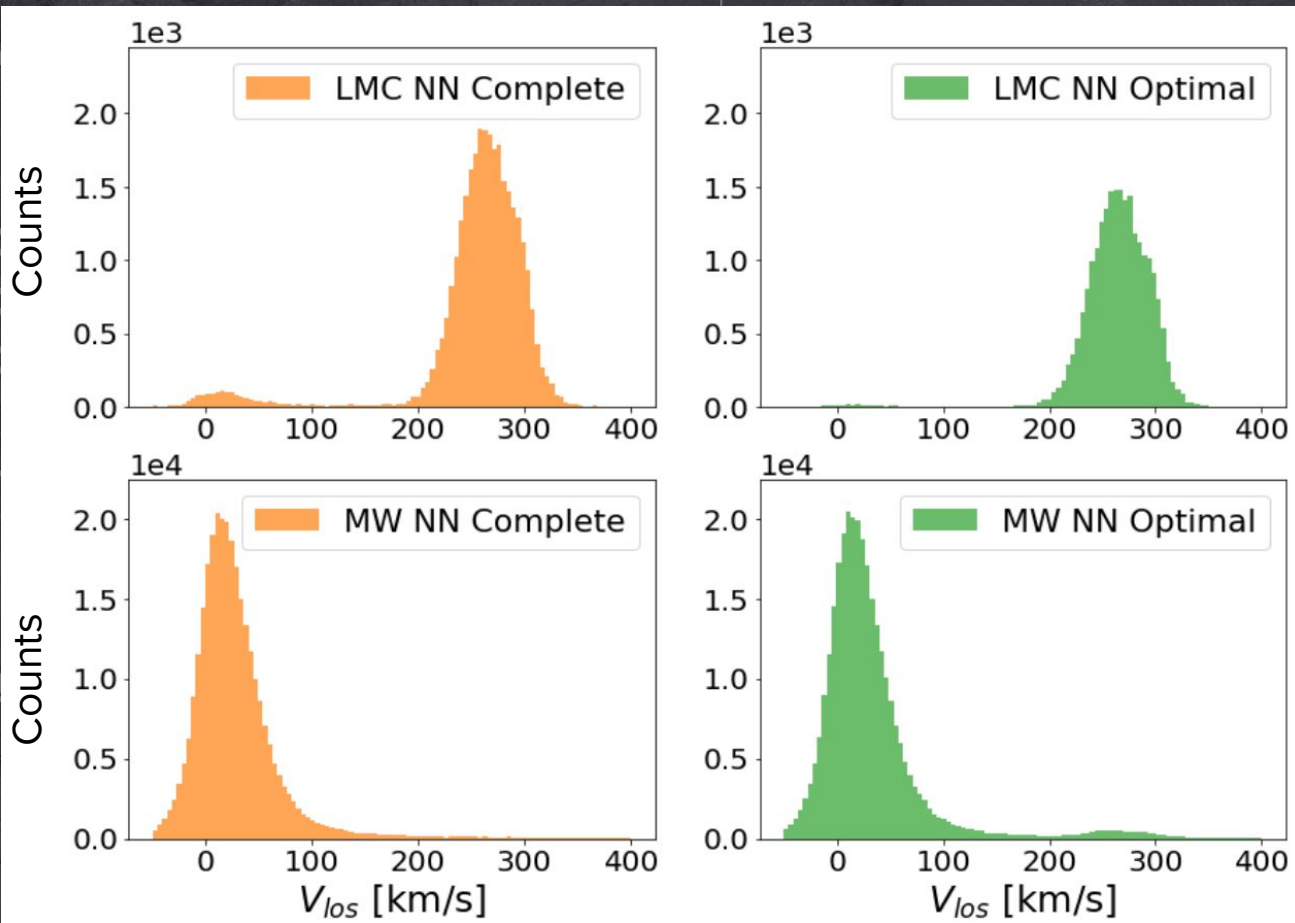


# LMC/MW

- Three e
  - LMC
  - LMC
  - LMC

- Gaia's list

Only for the brightest stars!

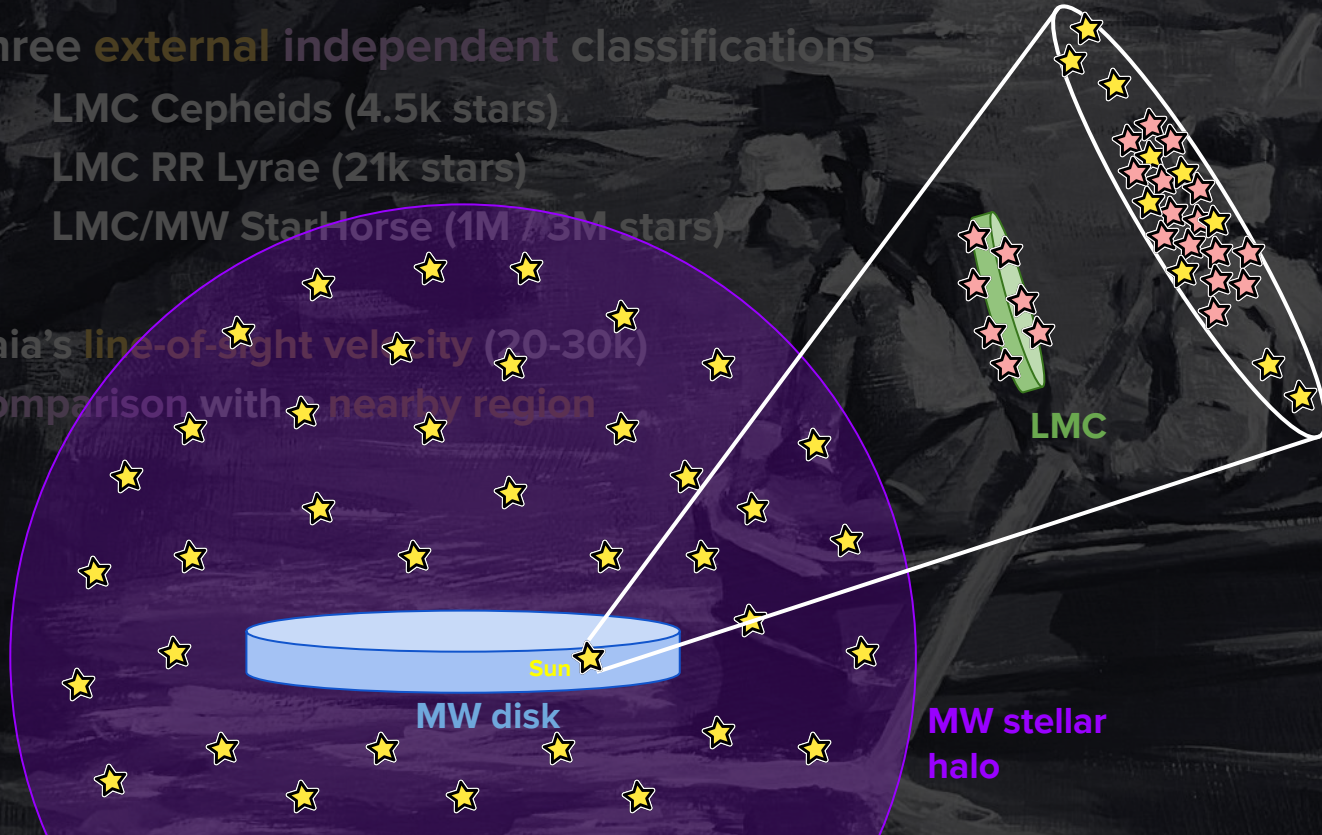


# LMC/MW classifier

- Three **external independent** classifications
  - LMC Cepheids (4.5k stars) ← Variable stars
  - LMC RR Lyrae (21k stars) ← Variable stars
  - LMC/MW StarHorse (1M / 3M stars) ← Bayesian inference (distance)
- Gaia's **line-of-sight velocity** (20-30k)
- **Comparison** with a **nearby region**

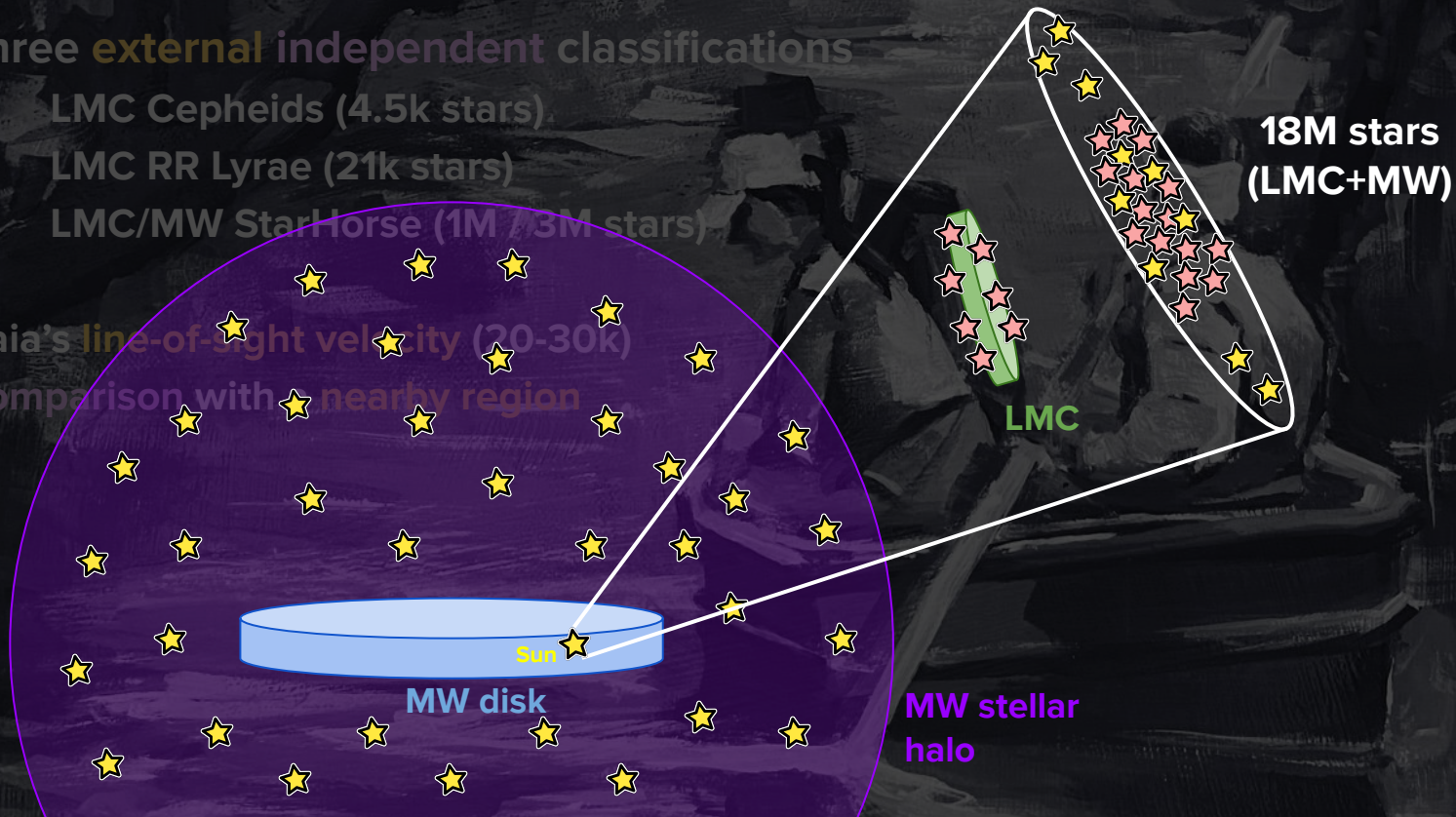
# LMC/MW classifier

- Three **external** independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's **line-of-sight velocity** (20-30k)
- Comparison with **nearly region**



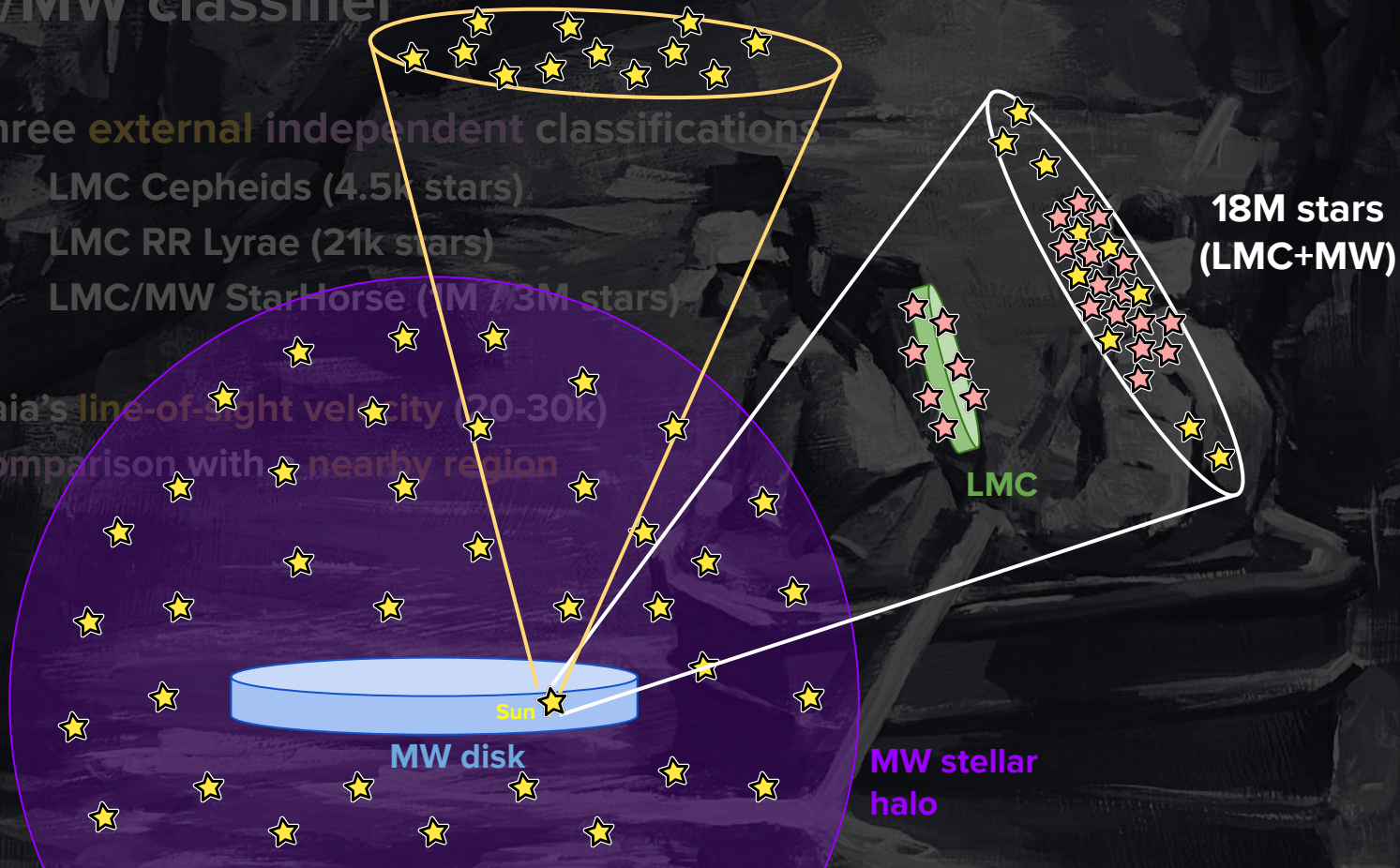
# LMC/MW classifier

- Three **external** independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's **line-of-sight velocity** (20-30k)
- Comparison with **nearby region**



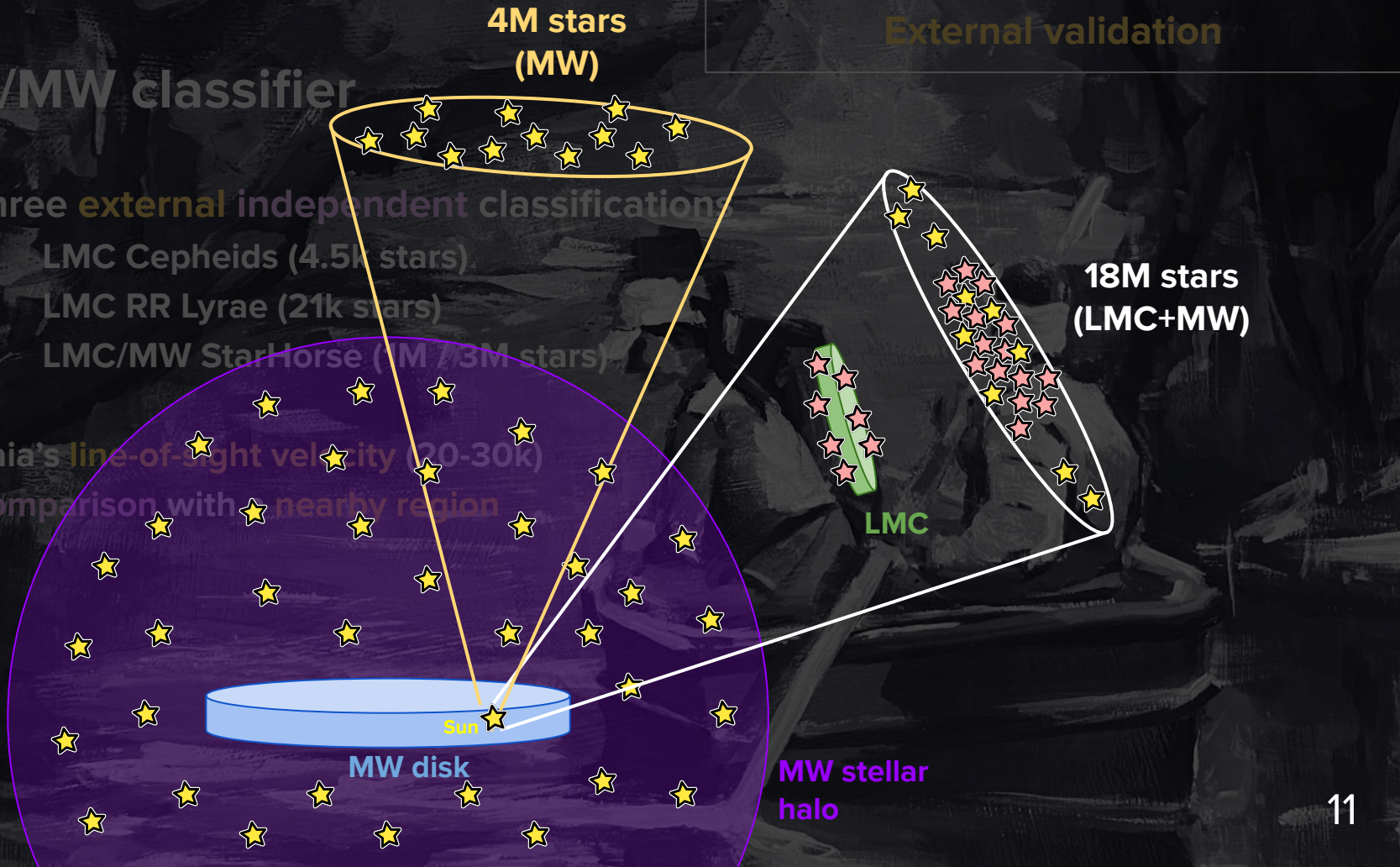
# LMC/MW classifier

- Three **external** independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's **line-of-sight velocity** (20-30k)
- Comparison with **nearly region**



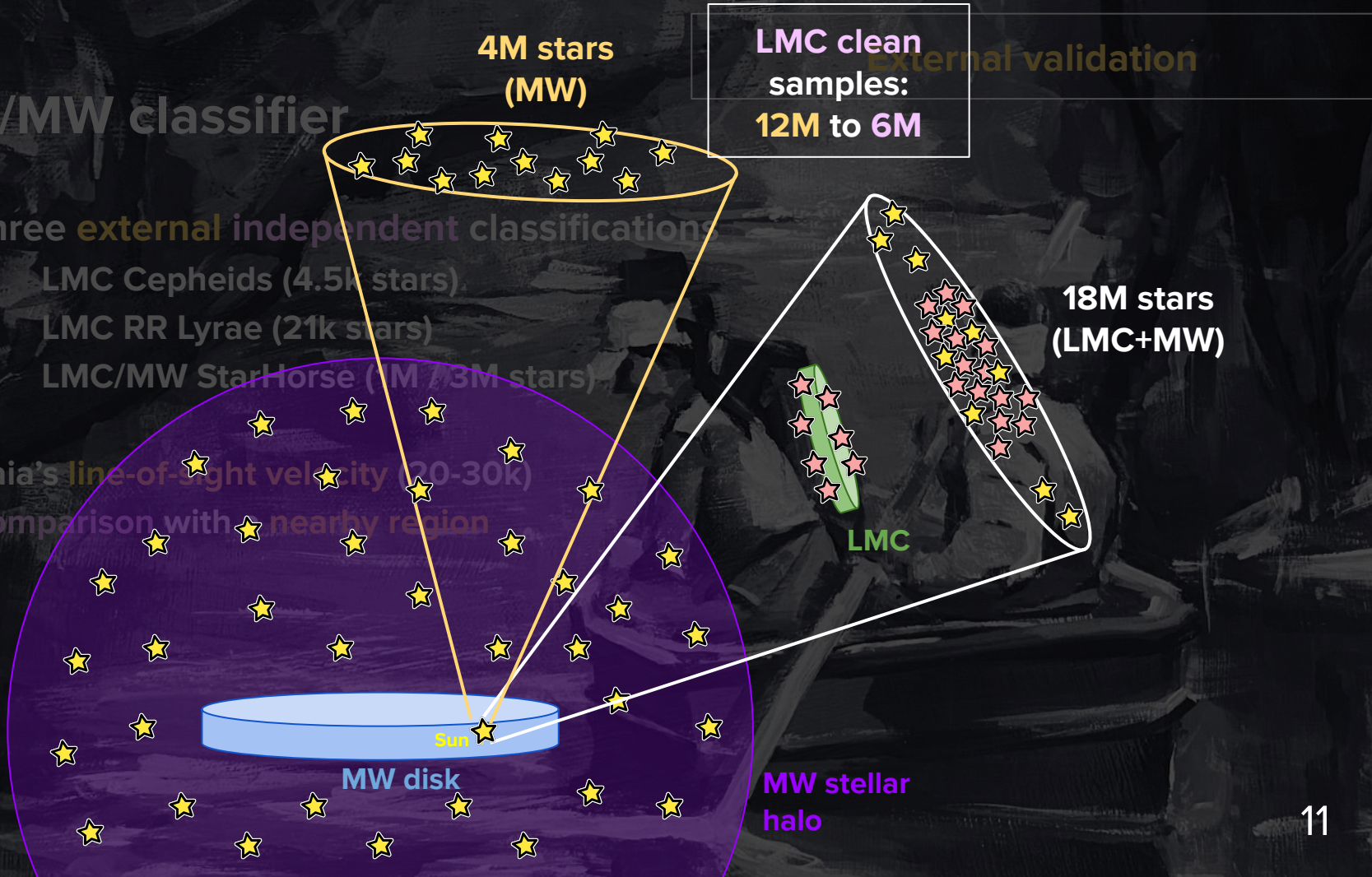
# LMC/MW classifier

- Three **external** independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's **line-of-sight velocity** (20-30k)
- Comparison with **nearly region**



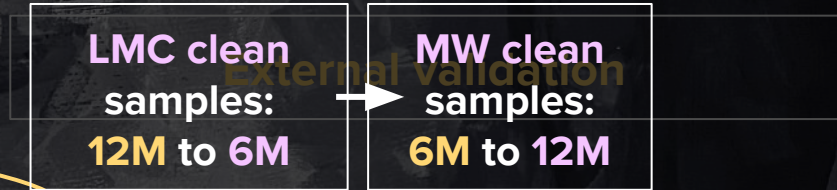
# LMC/MW classifier

- Three **external** independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's **line-of-sight velocity** (20-30k)
- Comparison with **nearly region**

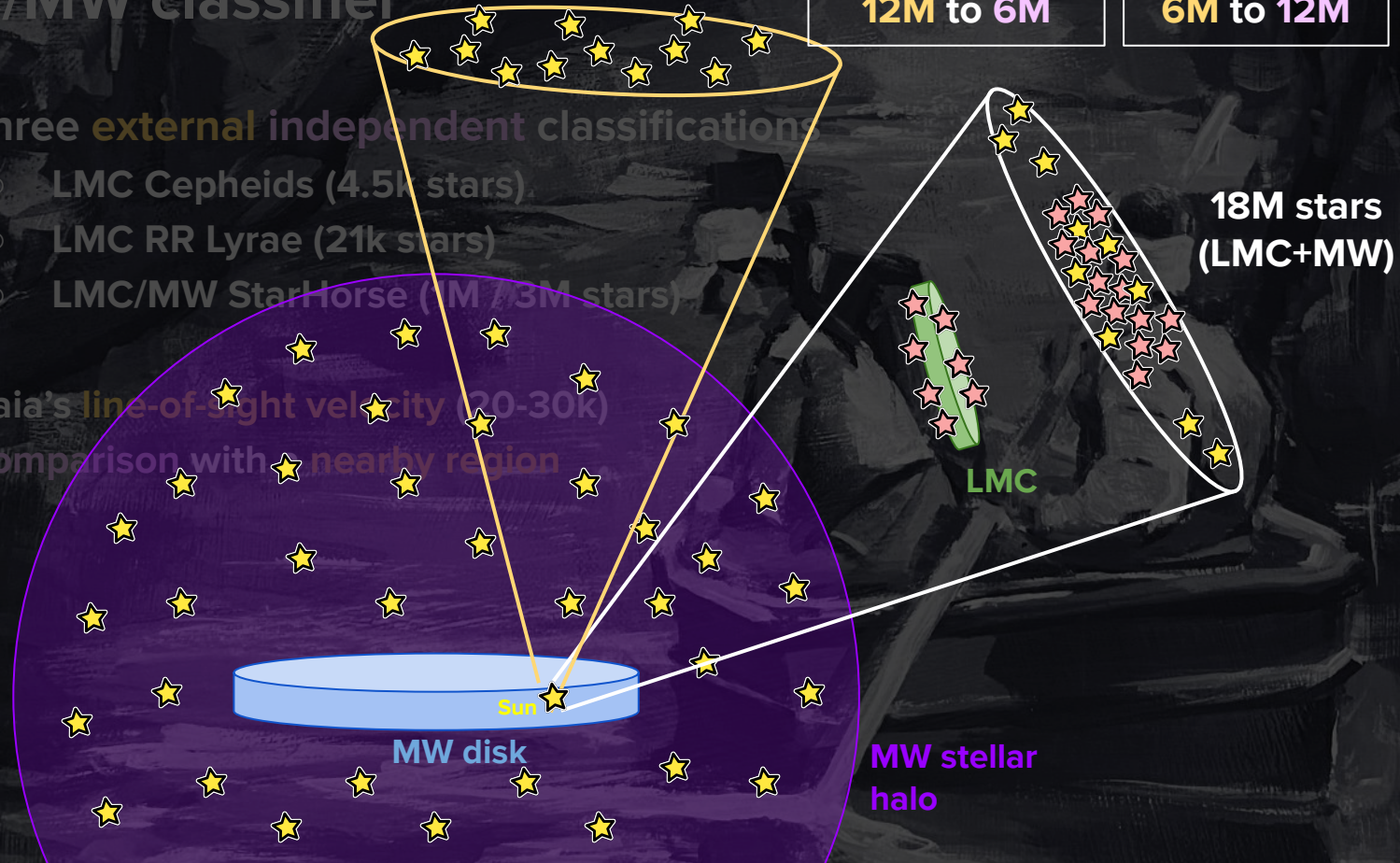




# LMC/MW classifier

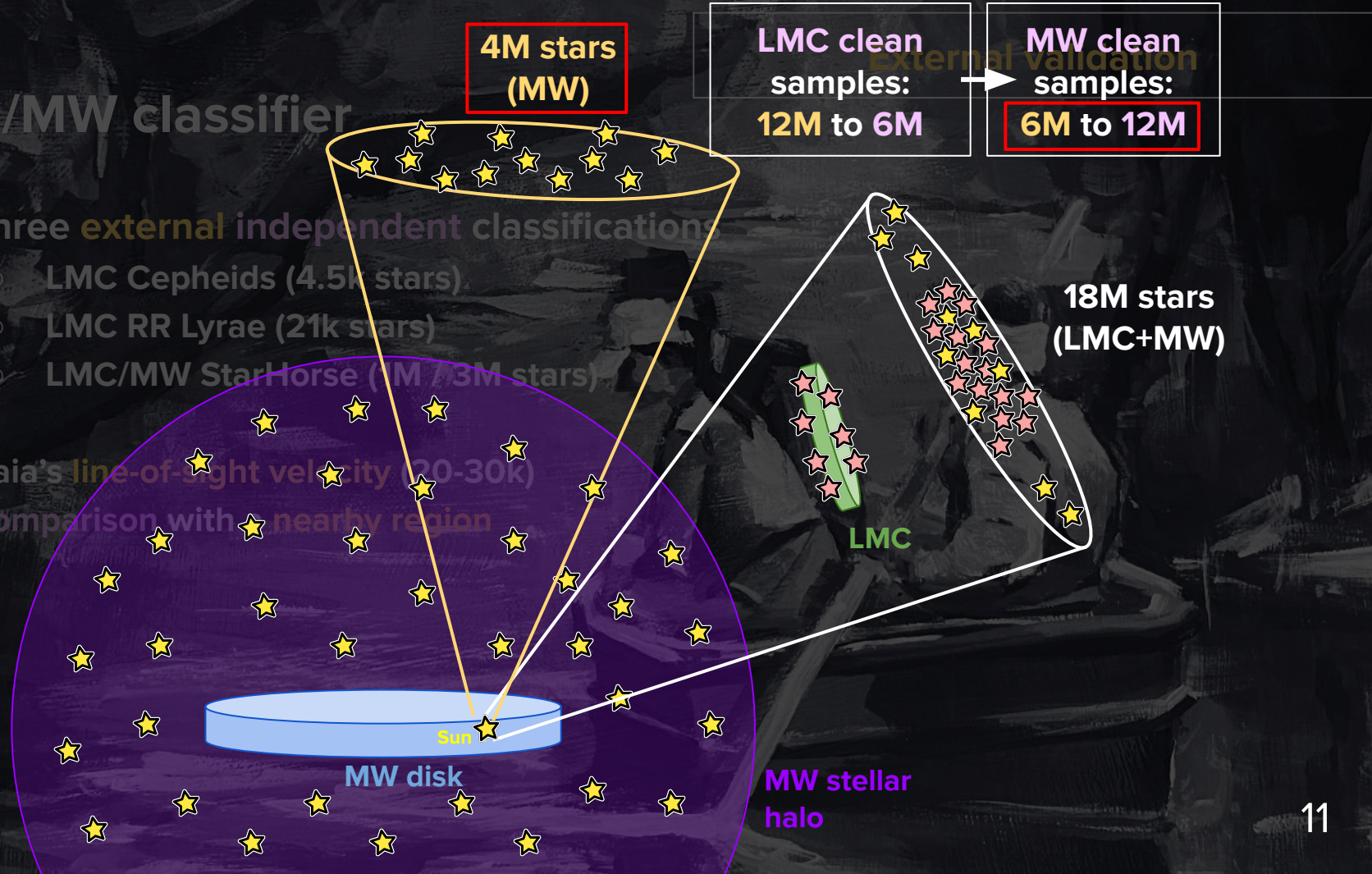


- Three **external** independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's **line-of-sight velocity** (20-30k)
- Comparison with **nearly region**



# LMC/MW classifier

- Three **external** independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's **line-of-sight velocity** (20-30k)
- Comparison with **nearly region**



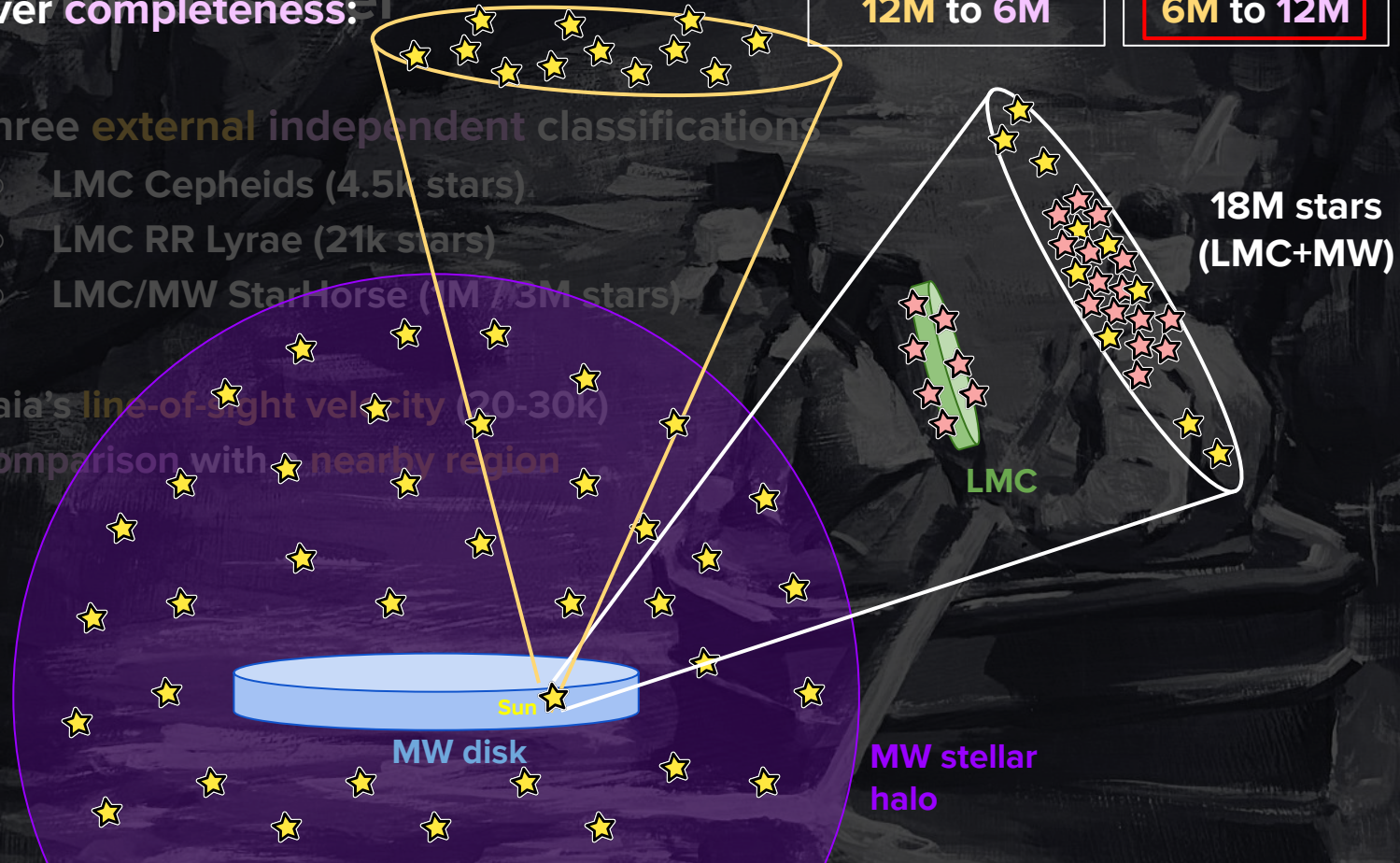
# LMC samples prioritise purity over completeness:

4M stars (MW)

LMC clean samples: 12M to 6M

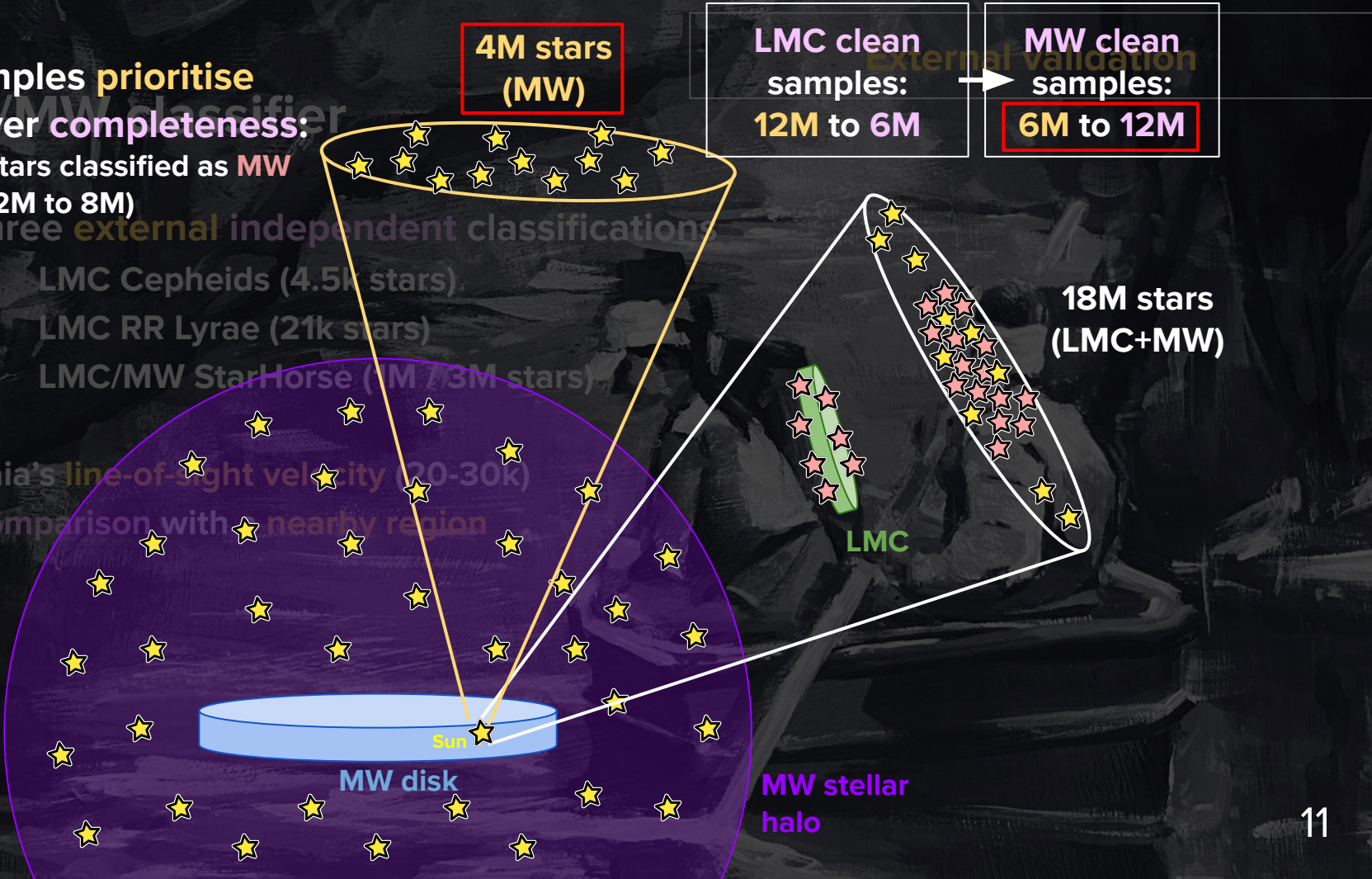
MW clean samples: 6M to 12M

- Three external independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's line-of-sight velocity (20-30k)
- Comparison with nearby region



LMC samples prioritise **purity** over **completeness**:  
too many stars classified as **MW**  
(excess of 2M to 8M)

- Three **external** independent classifications
  - LMC Cepheids (4.5k stars)
  - LMC RR Lyrae (21k stars)
  - LMC/MW Starhorse (1M / 3M stars)
- Gaia's **line-of-sight velocity** (20-30k)
- Comparison with **nearby region**

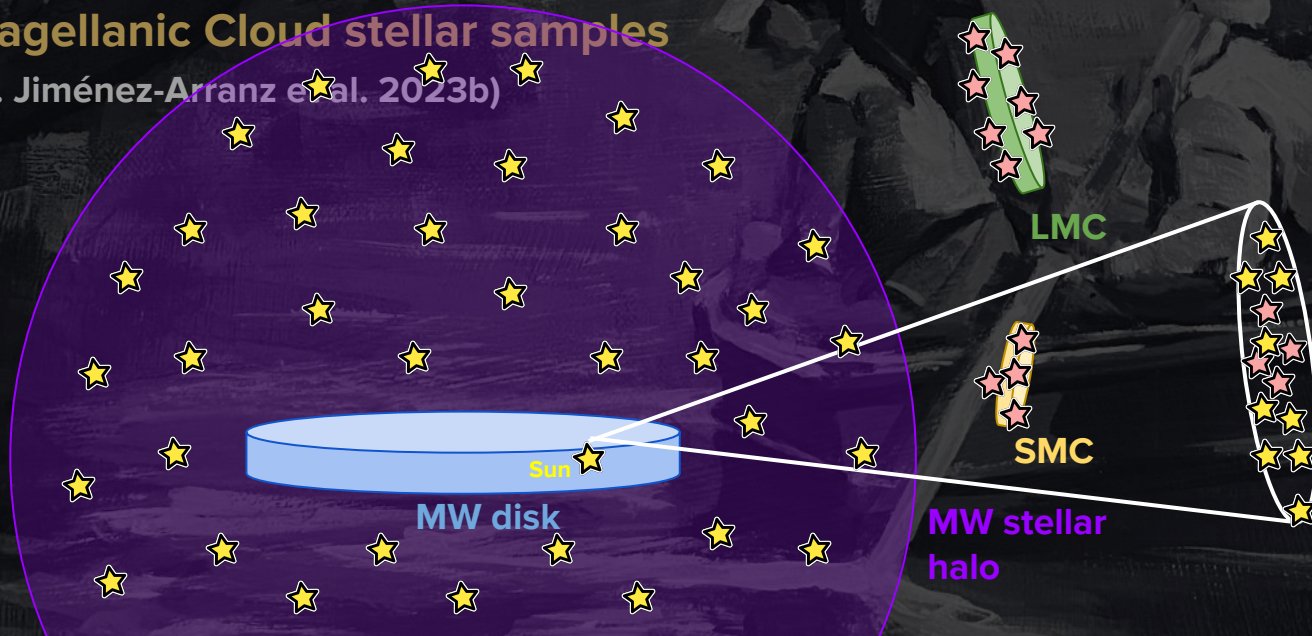


# My PhD Journey

- 1) Kinematic analysis of the Large Magellanic Cloud using Gaia DR3  
(Ó. Jiménez-Arranz et al. 2023a, w/ P. McMillan & S. Roca-Fàbrega)
  - Application of a neural network classifier for the generation of clean Small Magellanic Cloud stellar samples  
(Ó. Jiménez-Arranz et al. 2023b)

# My PhD Journey

- 1) Kinematic analysis of the Large Magellanic Cloud using Gaia DR3 (Ó. Jiménez-Arranz et al. 2023a, w/ P. McMillan & S. Roca-Fàbrega)
  - Application of a neural network classifier for the generation of clean Small Magellanic Cloud stellar samples (Ó. Jiménez-Arranz et al. 2023b)



# My PhD Journey

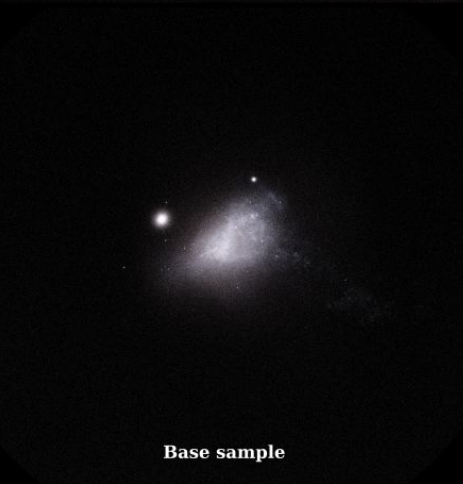
## 1) Kinematic analysis of the Large Magellanic Cloud using Gaia DR3

(Ó. Jiménez-Arranz et al. 2023a, w/ P. McMillan & S. Roca-Fàbrega)

Stars classified as SMC	SMC Cepheids (4 765)	SMC RR-Lyrae (2 922)	SMC StarHorse (193 402)	MW StarHorse (806 664)
NN complete	4 688 (98.4%)	2 814 (96.3%)	191 692 (99.1%)	125 200 (15.5%)
NN optimal	4 599 (96.5%)	2 694 (92.2%)	186 063 (96.2%)	110 704 (13.7%)
NN truncated-optimal	4 598 (96.5%)	821 (28.1%)	186 063 (96.2%)	110 704 (13.7%)



2 kpc



Base sample



Complete sample



Optimal sample



Truncated-optimal sample



2 kpc



Milky Way  
globular clusters



Base sample



Complete sample



Optimal sample



Truncated-optimal sample

# Conclusions



# Conclusions

The **Magellanic Clouds** are the **perfect laboratory** for testing **methodologies** and **models** designed for the study of **external** and **interacting galaxies**

# Conclusions

The **Magellanic Clouds** are the **perfect laboratory** for testing **methodologies** and **models** designed for the study of **external** and **interacting galaxies**

We need to **get rid of the Milky Way contaminants**: **LMC clean samples**

# Conclusions

The **Magellanic Clouds** are the **perfect laboratory** for testing methodologies and models designed for the study of **external** and **interacting galaxies**

We need to **get rid of the Milky Way contaminants**: **LMC clean samples**

**Artificial intelligence** (through **Neural Networks classifiers**) can help us using as much of the **Gaia data available**



**Gràcies!**

Óscar Jiménez-Arranz

[ojimenez@fqa.ub.edu](mailto:ojimenez@fqa.ub.edu)

Institut de Ciències del Cosmos (ICCUB)