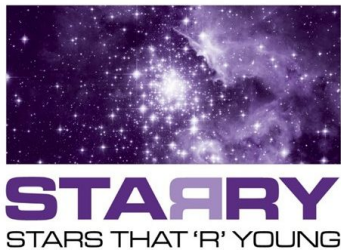


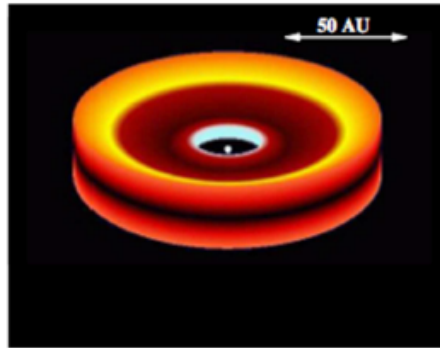
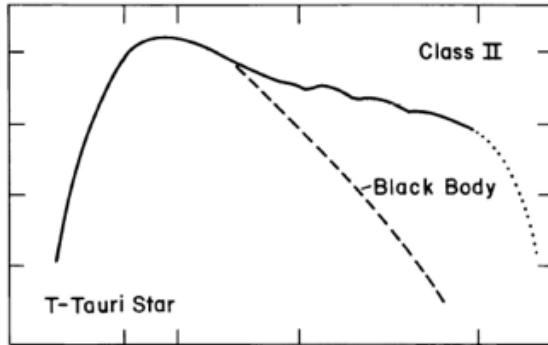
New catalogue of Pre-Main Sequence objects using AI

Miguel Vioque
University of Leeds

R. D. Oudmaijer (University of Leeds, UK), M. Schreiner (Desupervised, Denmark), D. Baines (ESAC, Spain), and R. Pérez-Martínez (Isdefe, Spain)



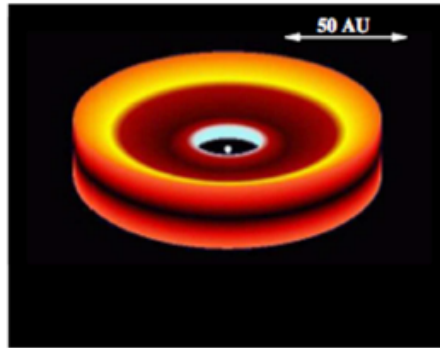
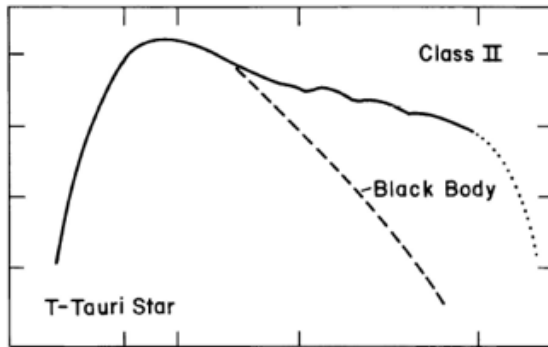
Looking for new High-Mass Pre-Main Sequence (PMS) objects



Main characteristics of PMS objects:

- Infrared excesses
- $H\alpha$ emission
- Photometric variability

Looking for new High-Mass Pre-Main Sequence (PMS) objects



Main characteristics of PMS objects:

- Infrared excesses
- $H\alpha$ emission
- Photometric variability

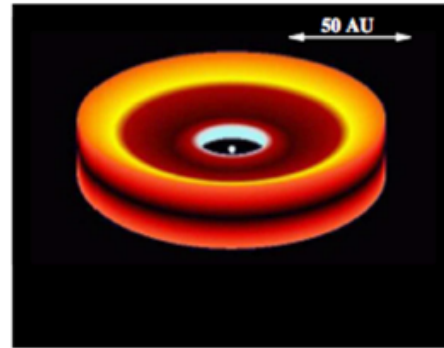
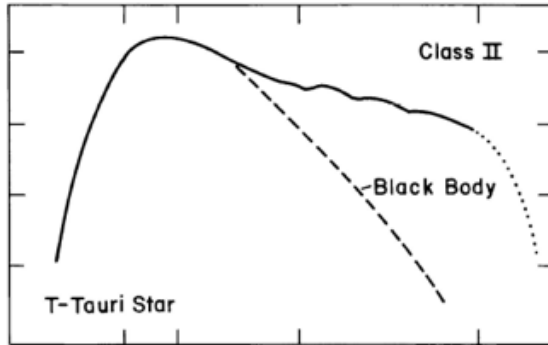
~250 high-mass PMS
known at the moment

Looking for new High-Mass Pre-Main Sequence (PMS) objects



**Perform an homogeneous
selection, distance and
position independent!**

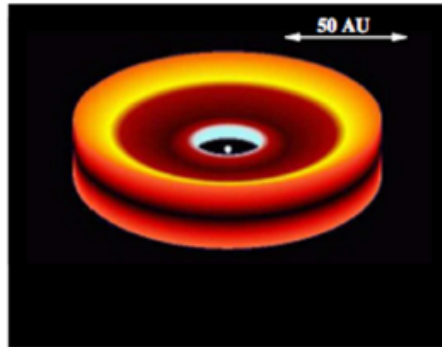
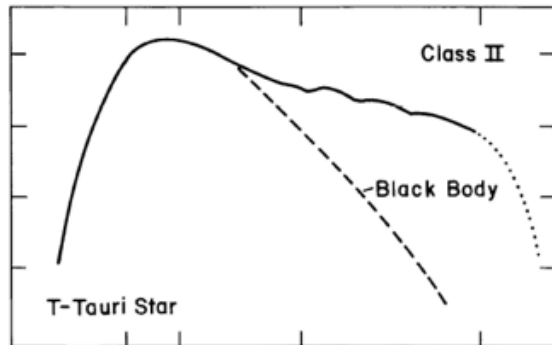
Looking for new High-Mass Pre-Main Sequence (PMS) objects



Main characteristics of PMS objects:

- Infrared excesses
- $H\alpha$ emission
- Photometric variability

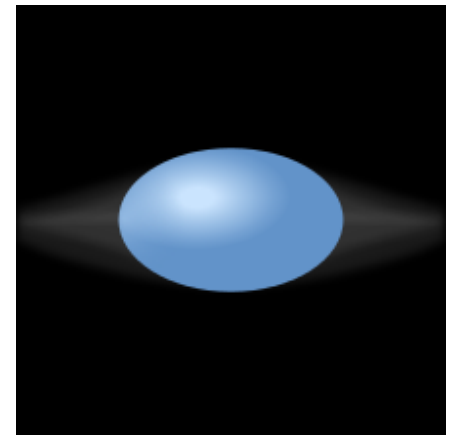
Looking for new High-Mass Pre-Main Sequence (PMS) objects



Main characteristics of PMS objects:

- Infrared excesses
- $H\alpha$ emission
- Photometric variability

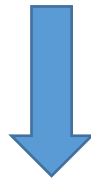
High mass PMS objects (Herbig Be stars) are very similar to **Classical Be stars**



We used an Artificial Neural Network

Selection of the **characteristics**:

- From Gaia: B_p , G , R_p and **2 variability indicators**
- From AllWISE: J , H , K_s , $W1$, $W2$, $W3$, $W4$
- From IPHAS & VPHAS+: $r - H_\alpha$



Create **all possible colours**



Remove all linear dependency
(PCA)

**Distance and
position
independent!**

Cross-match Gaia DR2 x AllWISE x IPHAS and VPHAS+

Master Sample = 4,151,538 sources

Construction of the **Training Set** (3 classes):

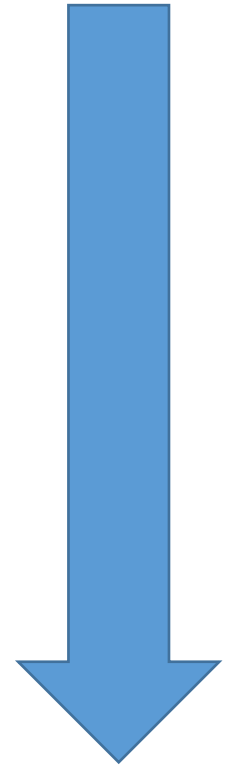
- **848** Pre-Main Sequence objects
 - **163** are Herbig Ae/Be stars
(high mass end, all available)
- **775** Classical Be stars (all available)
- **471,111** random sources with all the characteristics

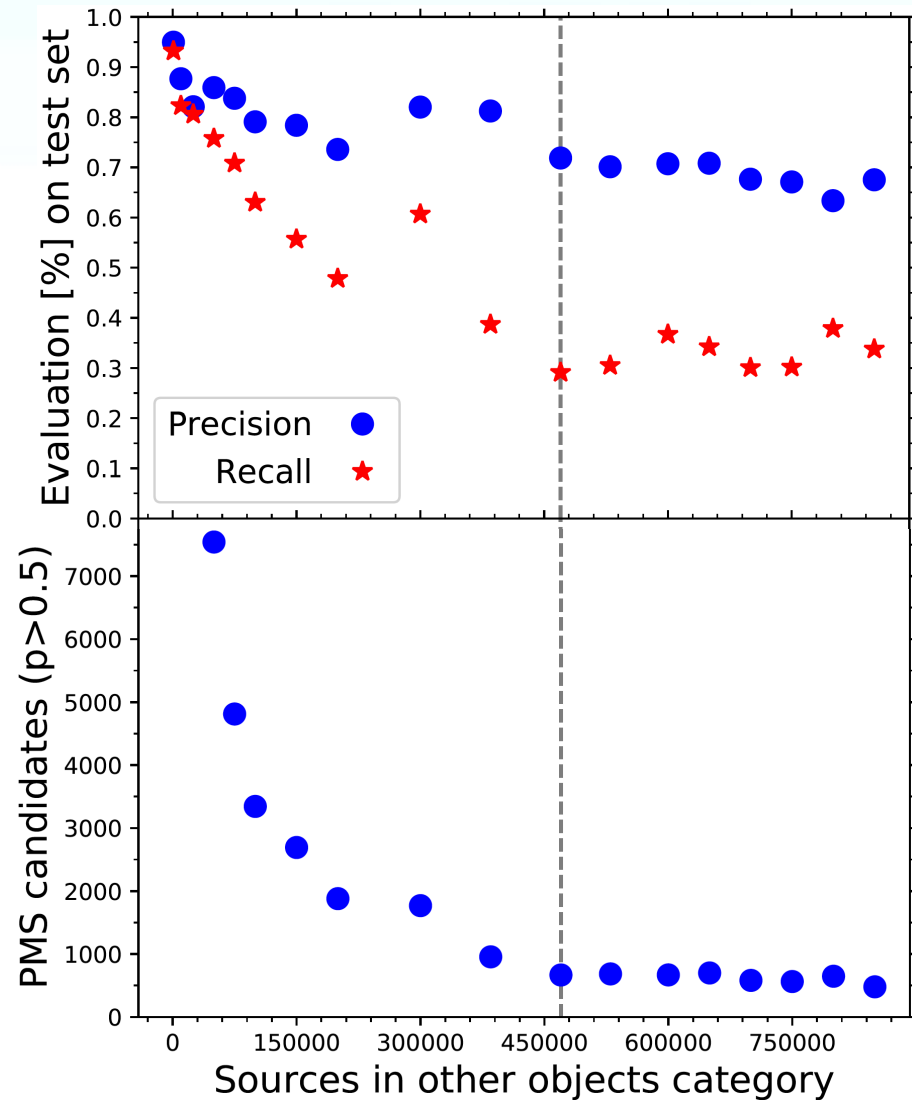
Cross-match Gaia DR2 x AllWISE x IPHAS and VPHAS+

Master Sample = 4,151,538 sources

Construction of the **Training Set** (3 classes):

- **848** Pre-Main Sequence objects
 - **163** are Herbig Ae/Be stars
(high mass end, all available)
- **775** Classical Be stars (all available)
- **471,111** random sources with all the characteristics





Size of other objects category

There is a stabilization point and after this the algorithm generalizes properly.

Cross-match Gaia DR2 x AllWISE x IPHAS and VPHAS+

Master Sample = 4,151,538 sources

Construction of the **Training Set** (3 classes):

- **848** Pre-Main Sequence objects
 - **163** are Herbig Ae/Be stars
(high mass end, all available)
- **775** Classical Be stars (all available)
- **471,111** random sources with all the characteristics

Cross-match Gaia DR2 x AllWISE x IPHAS and VPHAS+

Master Sample = 4,151,538 sources

Construction of the **Training Set** (3 classes):

- **848** Pre-Main Sequence objects
 - **163** are Herbig Ae/Be stars (high mass end, all available)
- **775** Classical Be stars (all available)
- **471,111** random sources with all the characteristics

There is a large contamination between categories!



This algorithm cannot assess itself

Construction of the **Training Set** (3 classes):

- **848** Pre-Main Sequence objects
 - **163** are Herbig Ae/Be stars (high mass end, all available)
- **775** Classical Be stars (all available)
- **471,111** random sources with all the characteristics

There is a large contamination between categories!



Architecture & Methodology

In order to deal with the **small Training Set**
and **the large contamination:**

Bootstrap (x30)

Balanced class weights

Architecture & Methodology

In order to deal with the **small Training Set**
and **the large contamination**:

Bootstrap (x30)

Balanced class weights

Chosen architecture:

- 2 hidden layers of **580 neurons** each
- L2 **regularization** (0.01) and 50% **dropout**
- Early-stopping when **precision** gets to a maximum
(10% Cross-Validation)
- Test Set size is kept to 10%



Training the Neural Network

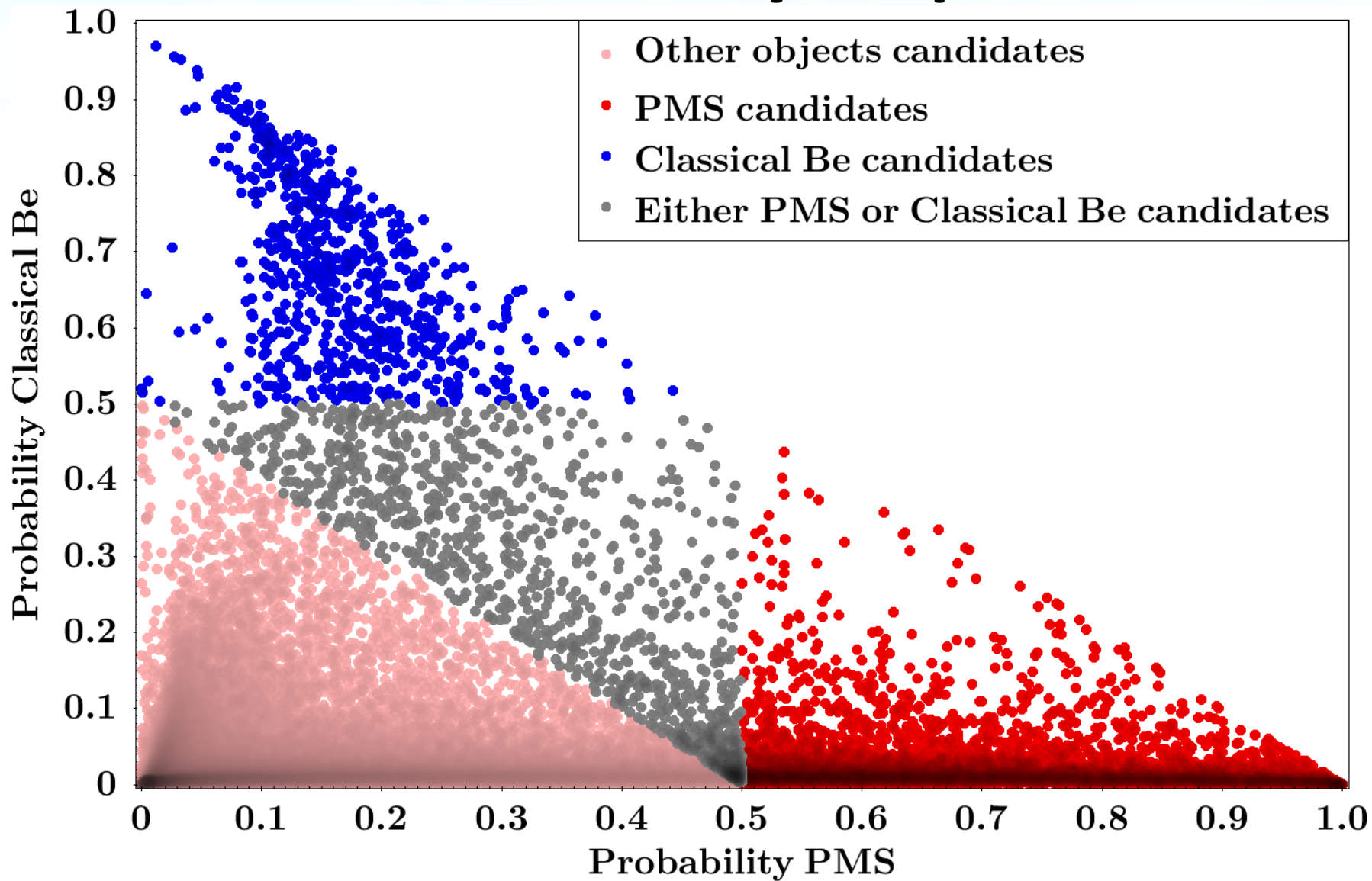


Trained Neural Network

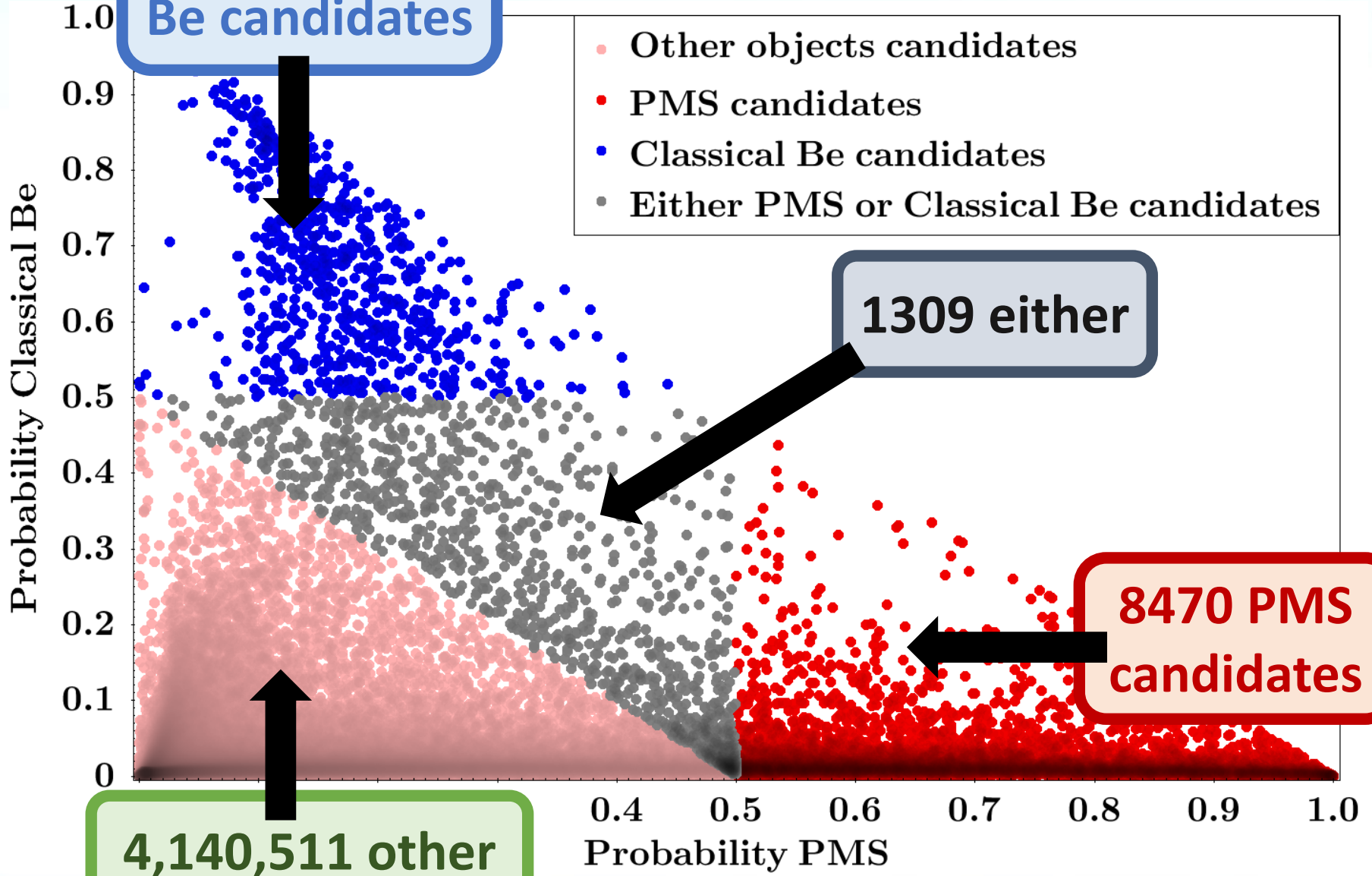


Master Sample = 4,151,538 sources

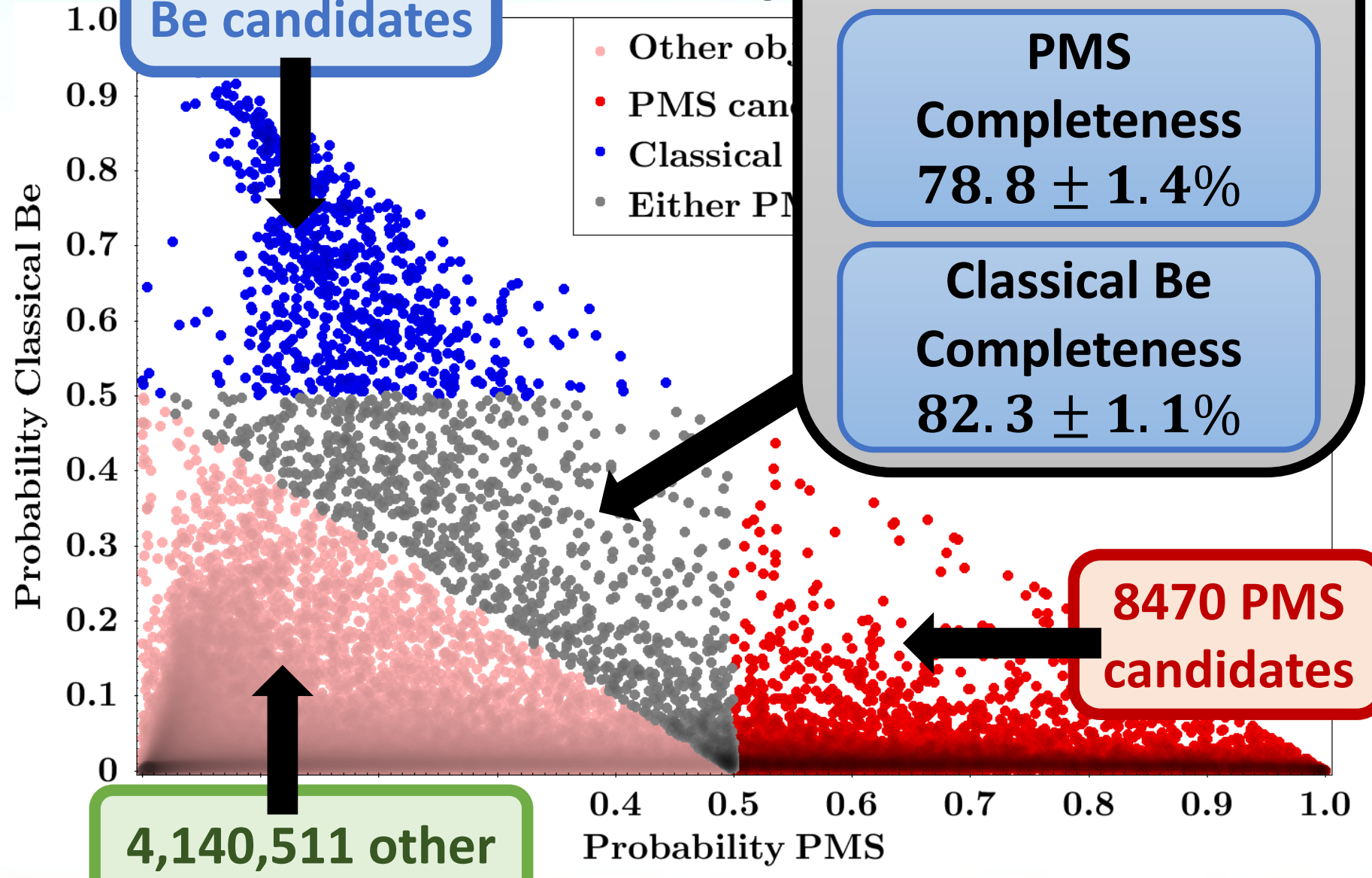
Probability Map



Probability Map



Probability



693 Classical Be candidates

Evaluation on Test Set

PMS
Completeness
 $78.8 \pm 1.4\%$

Classical Be
Completeness
 $82.3 \pm 1.1\%$

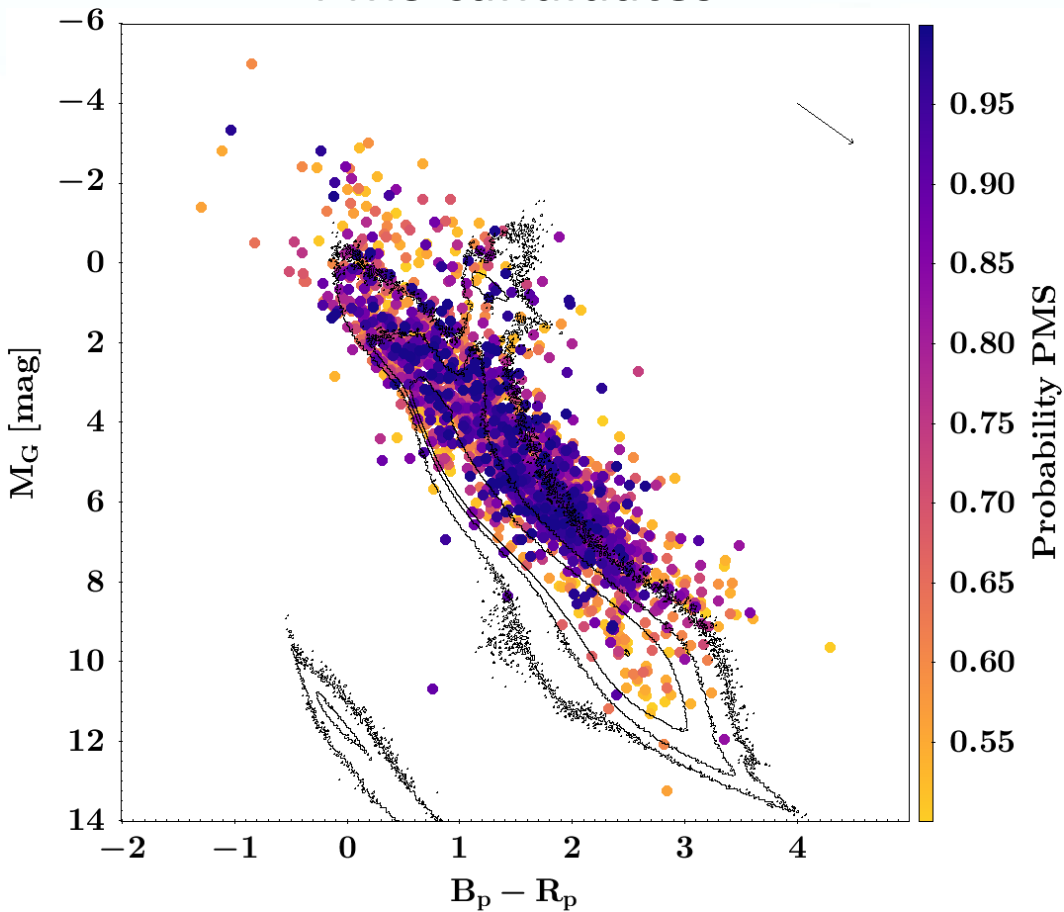
8470 PMS candidates

4,140,511 other

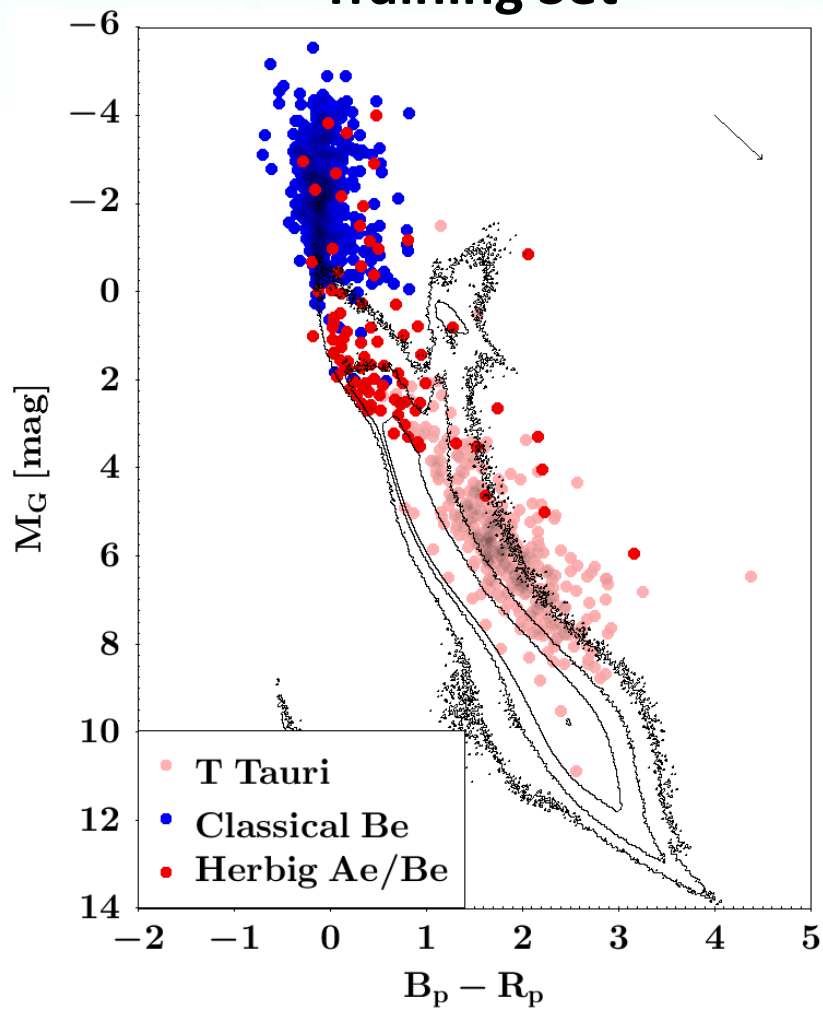
Probability PMS

HR diagram

PMS candidates

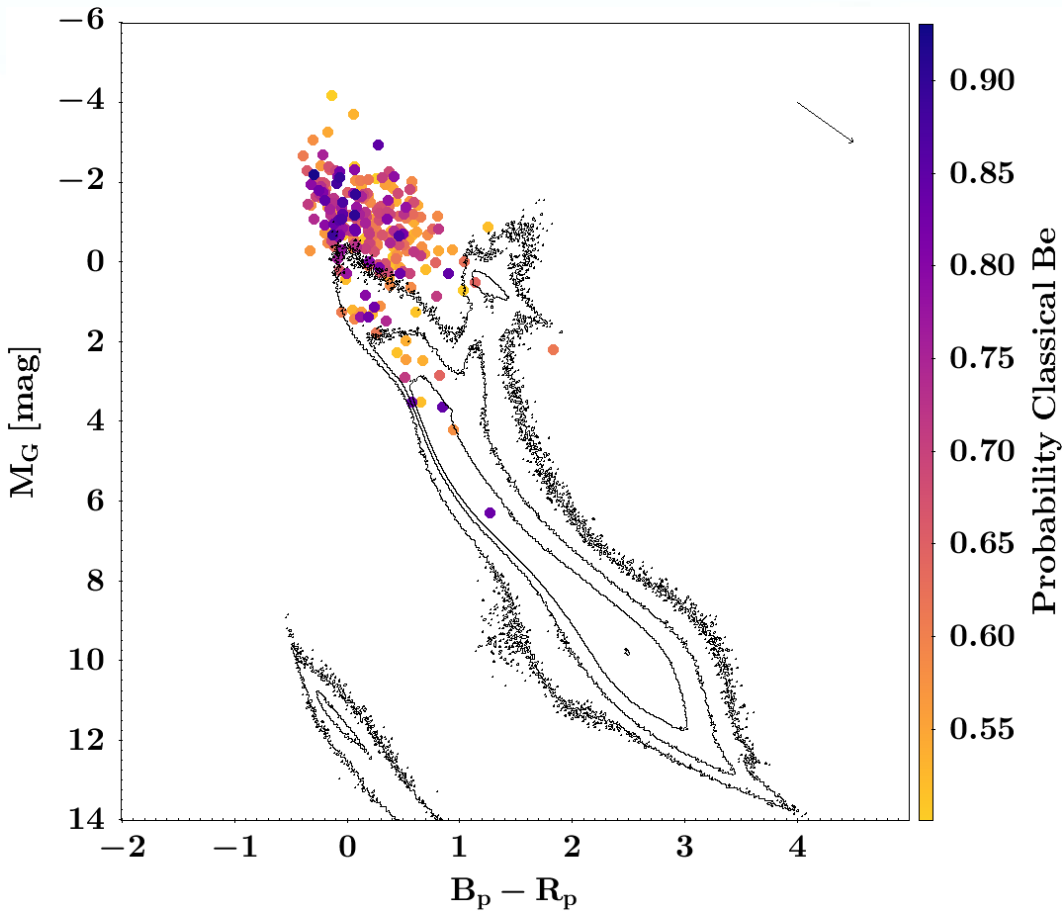


Training Set

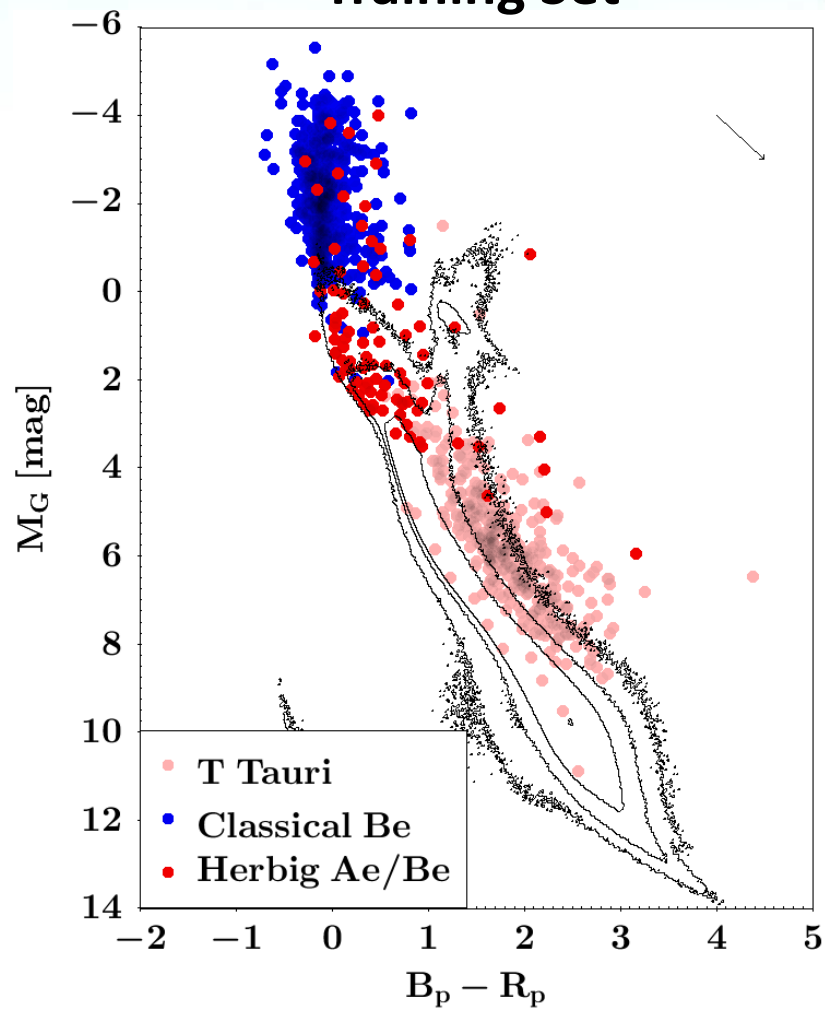


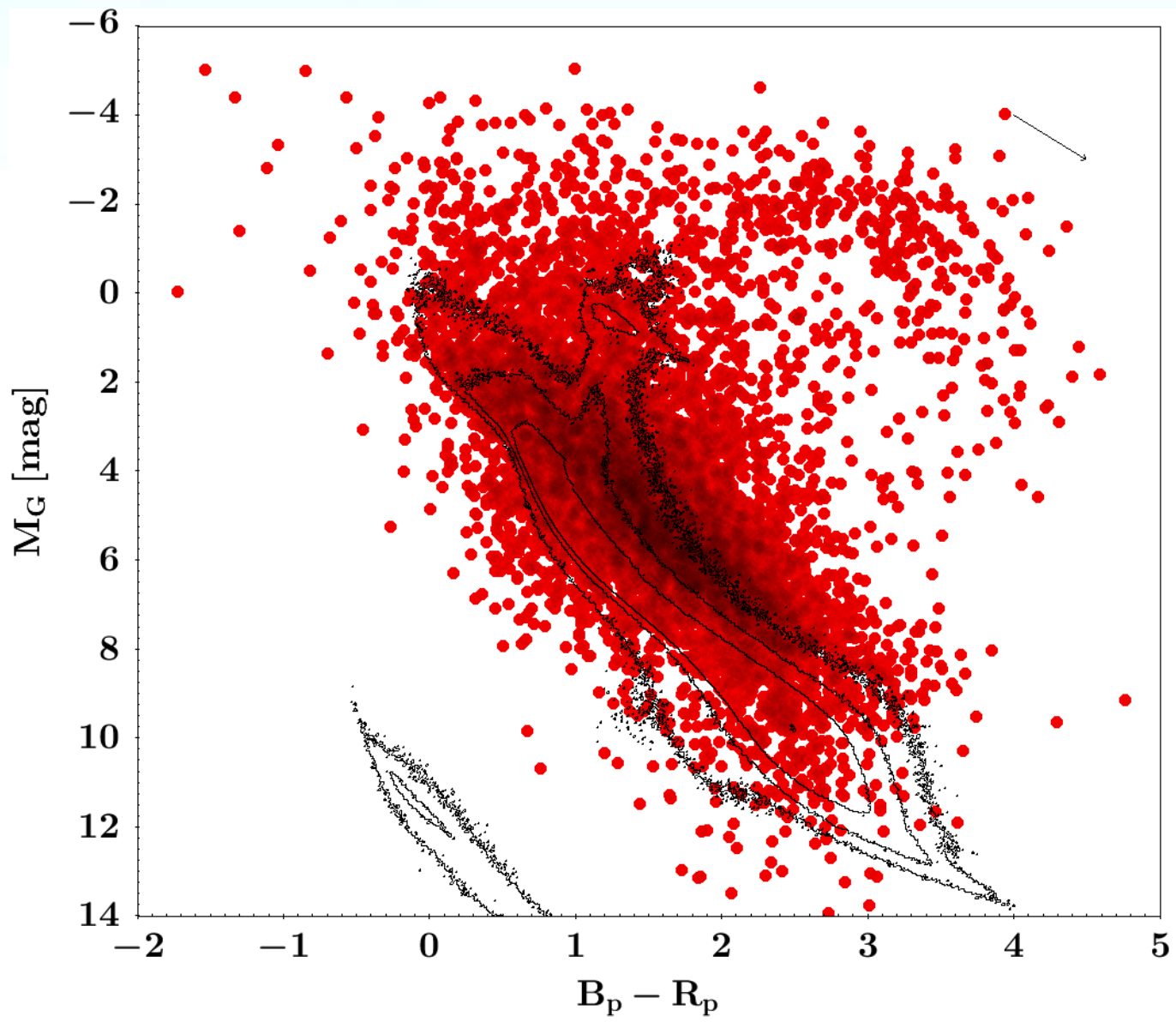
HR diagram

Classical Be candidates

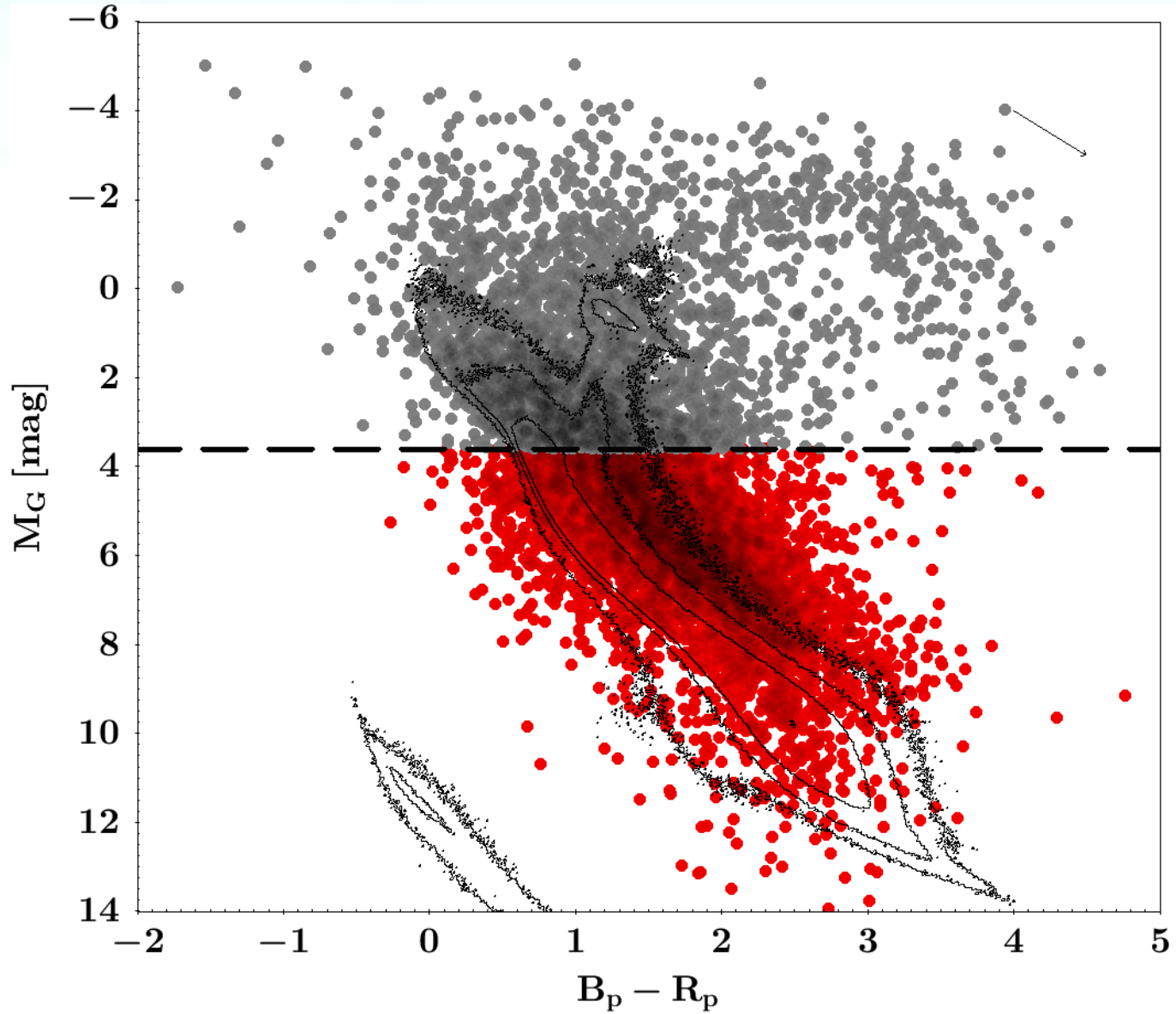


Training Set

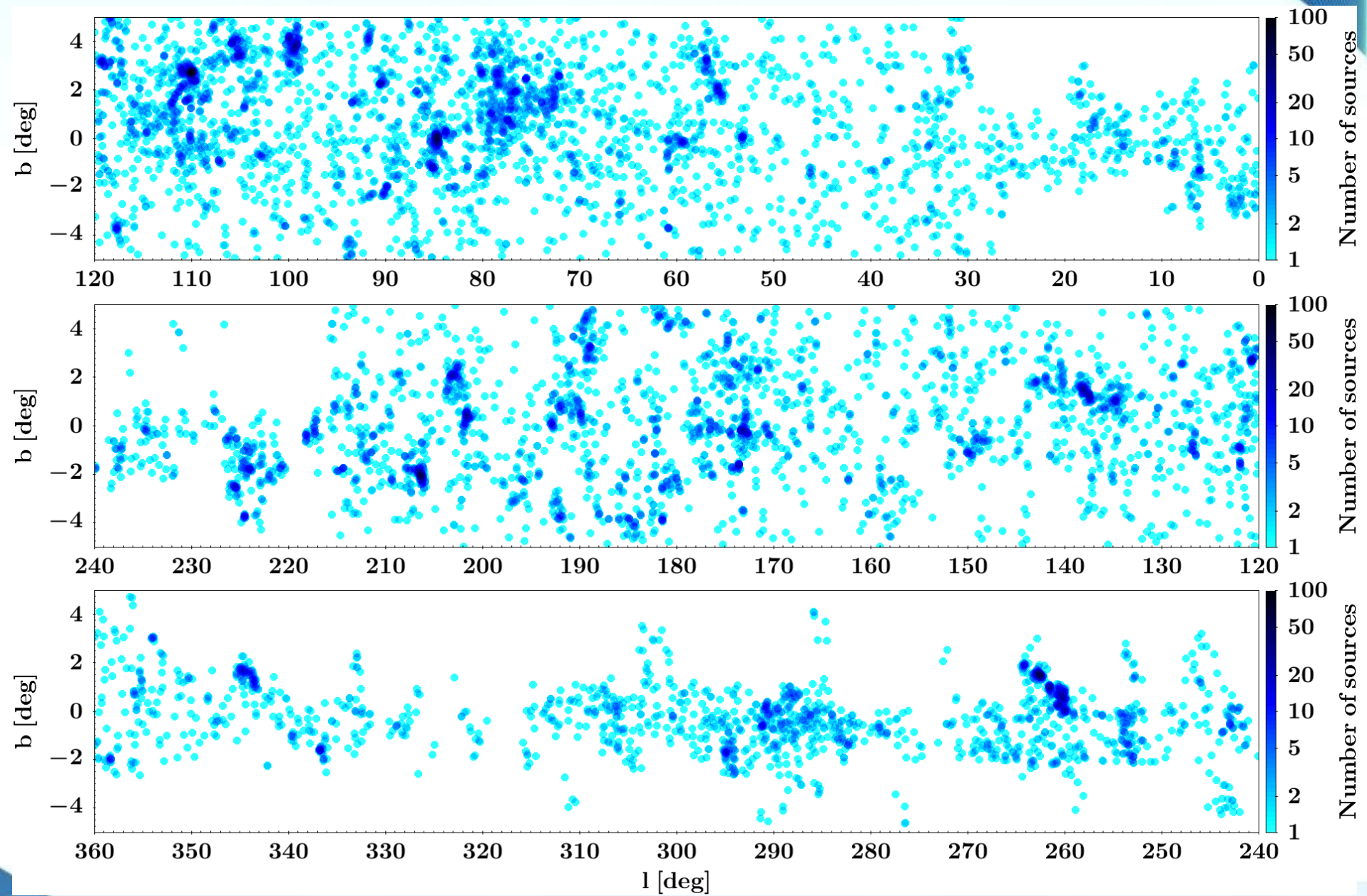




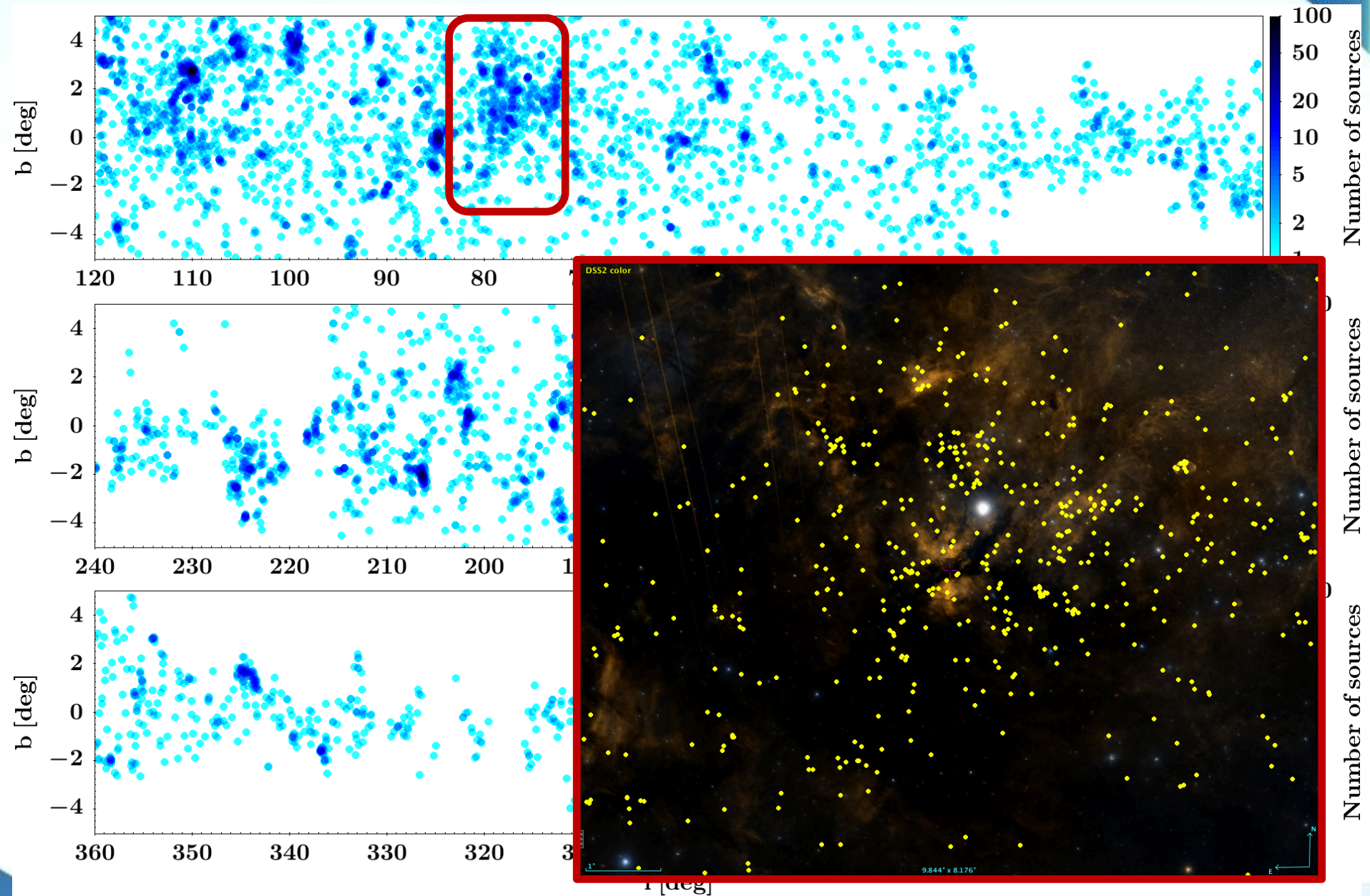
3131 potential high mass (682 with good Gaia solution)



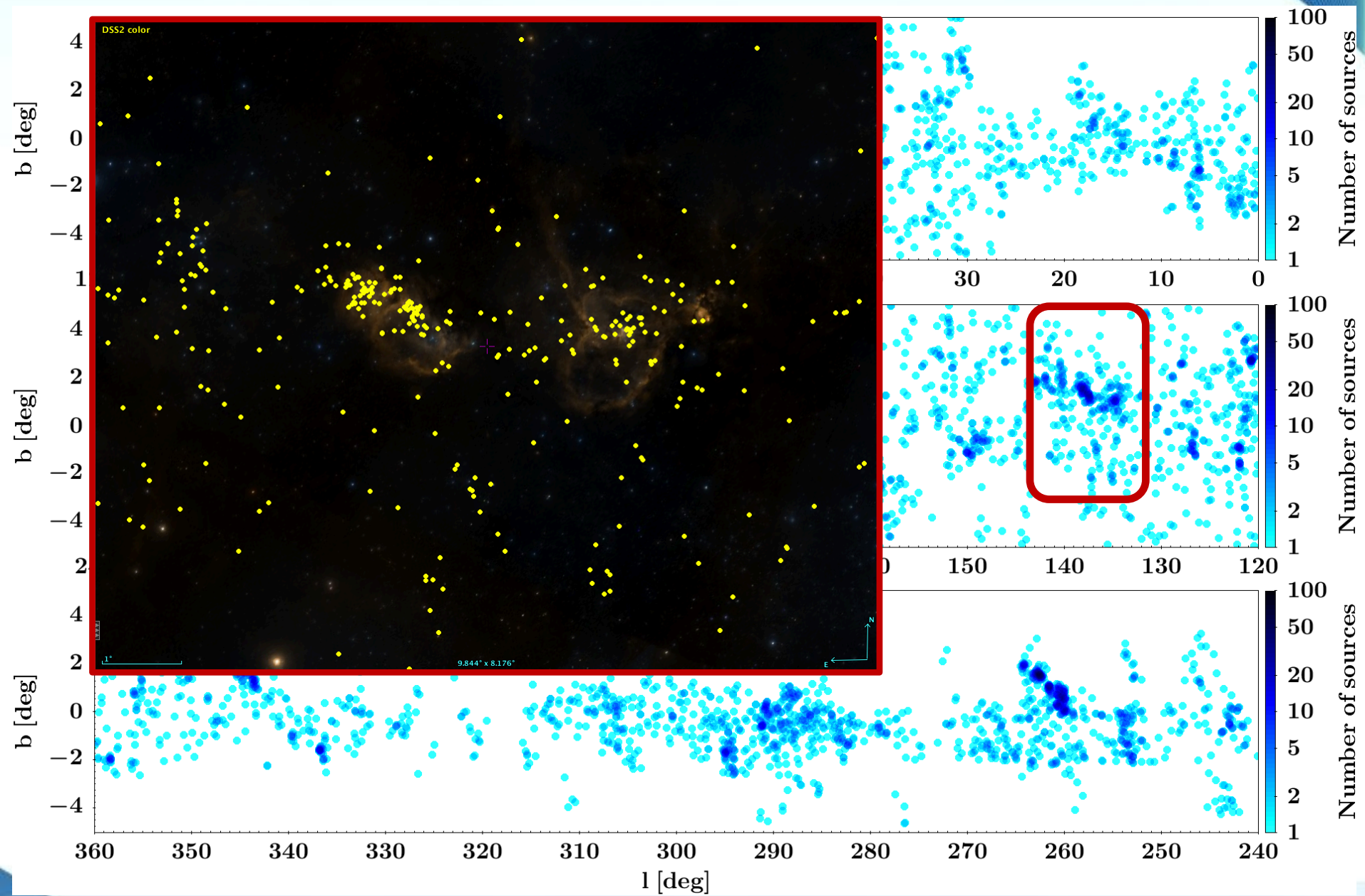
Coordinates



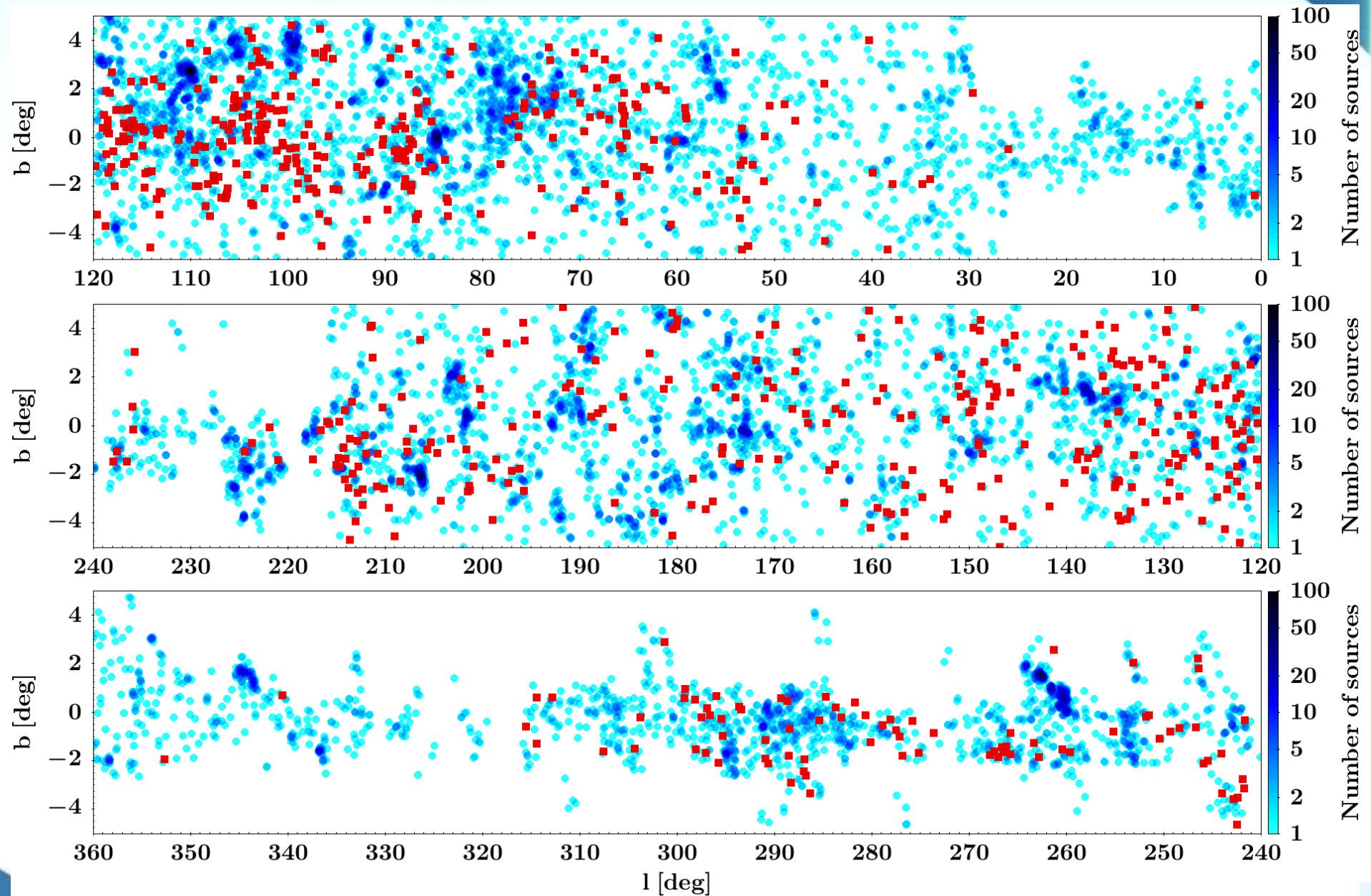
Coordinates



Coordinates

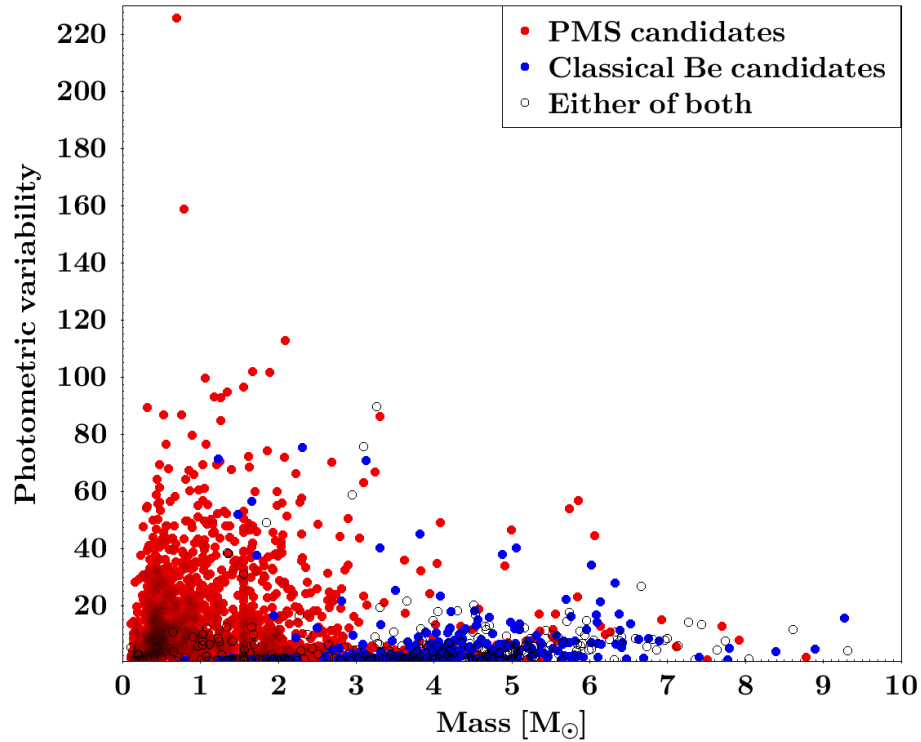


Coordinates

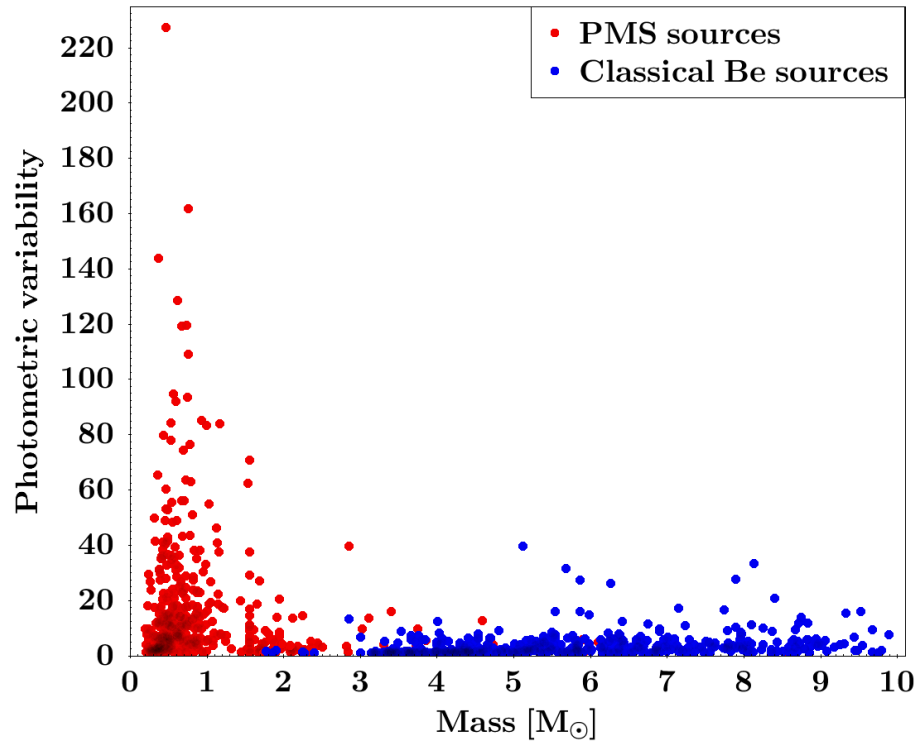


Variability vs. Masses (lower limits)

Candidates

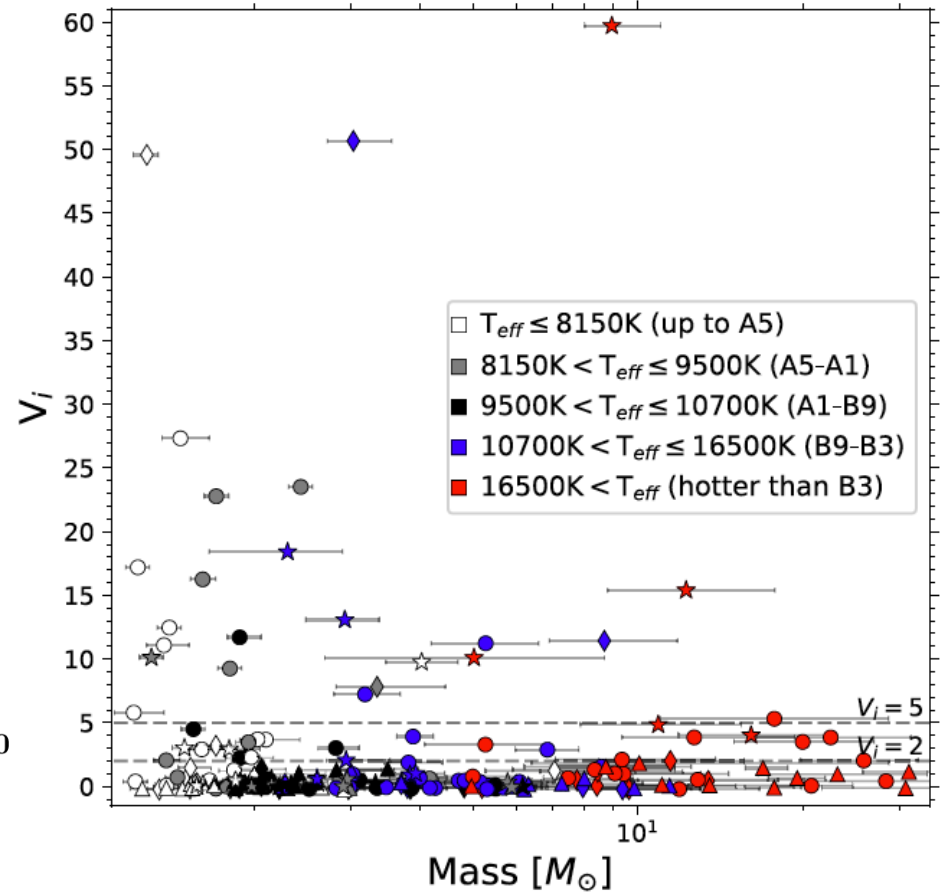
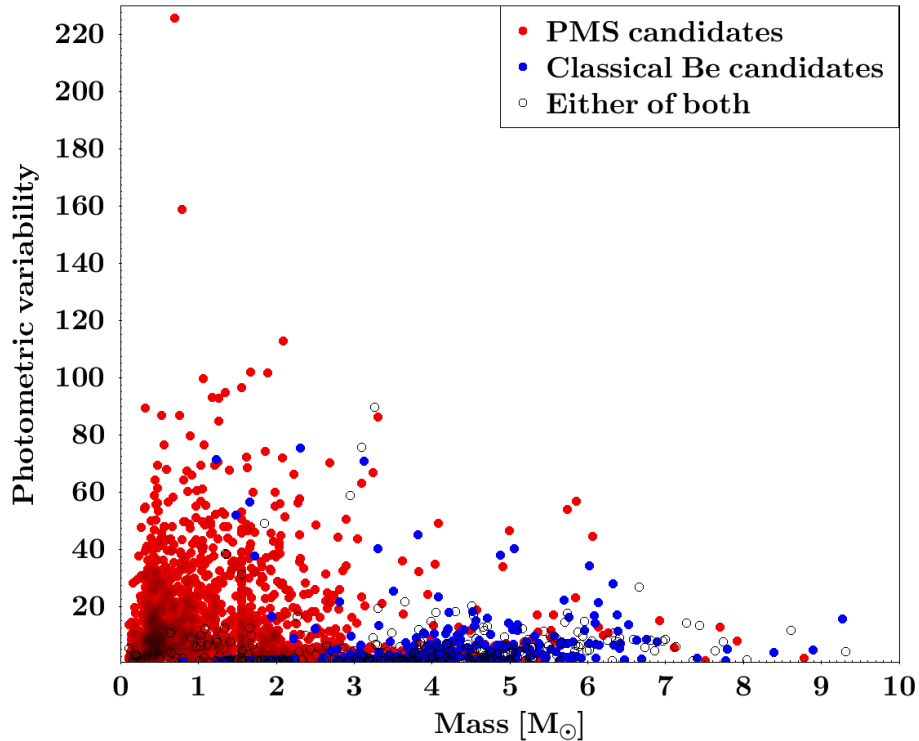


Training Set



Variability vs. Masses (lower limits)

Candidates



Vioque et al. 2018

Results:

- We retrieve **8470** new PMS candidates. **3131 (682)** potential high-mass ones (~ 250 known at the moment).
- We retrieve **693** new Classical Be stars candidates.
- We retrieve **1309** candidates of belonging to either one of the two categories.

Completeness
 $78.8 \pm 1.4\%$

Completeness
 $82.3 \pm 1.1\%$

Near- and mid- infrared excesses are the most important characteristics followed by $H\alpha$ and variability **which are equally important**

Architecture & Methodology

In order to deal with the **small Training Set**
and **the large contamination**:

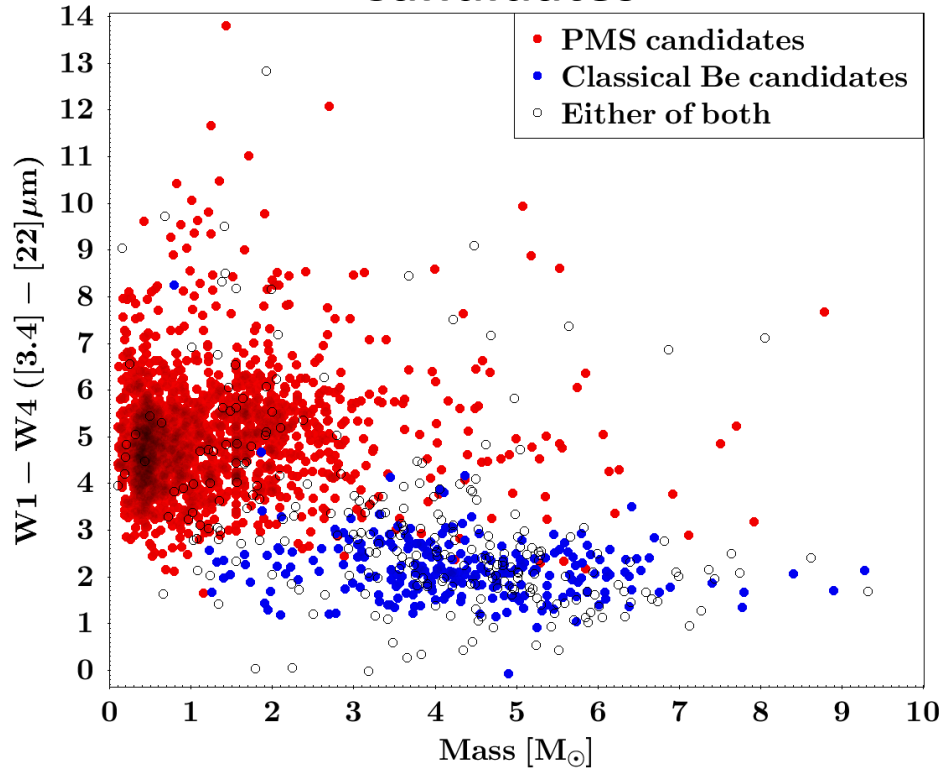
Bootstrap (x30)

Balanced class weights

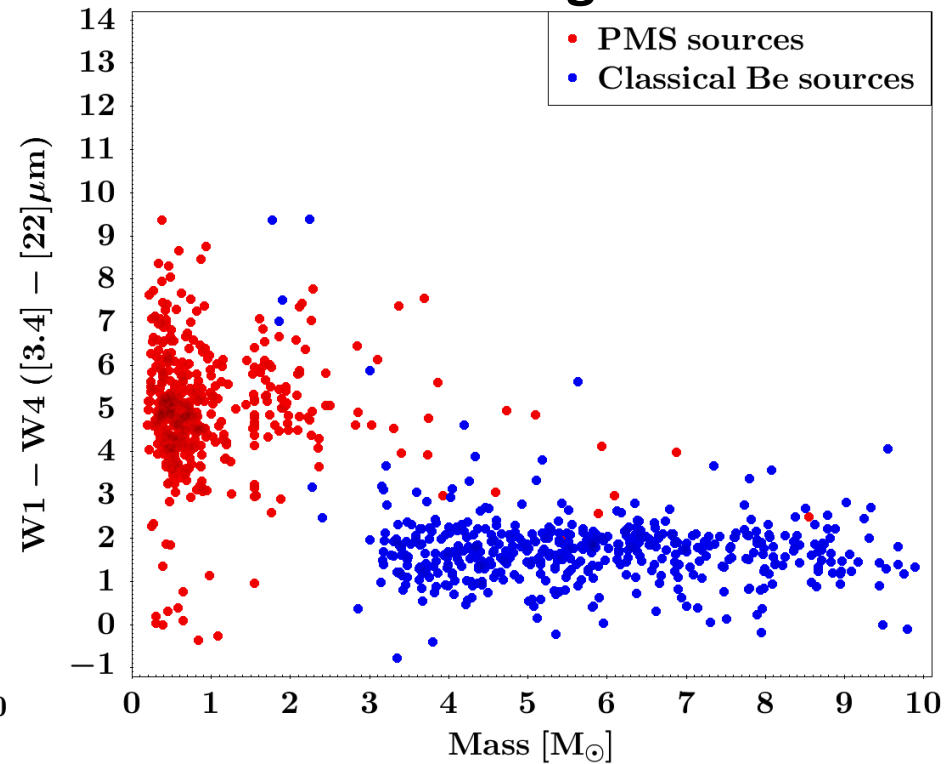
Original Dataset	1	2	3	4	5	6	7	8	9	10
Bootstrap 1	8	6	2	9	5	8	1	4	8	2
Bootstrap 2	10	1	3	5	1	7	4	2	1	8
Bootstrap 3	6	5	4	1	2	4	2	6	9	2

Mid-IR excess vs. Masses (lower limits)

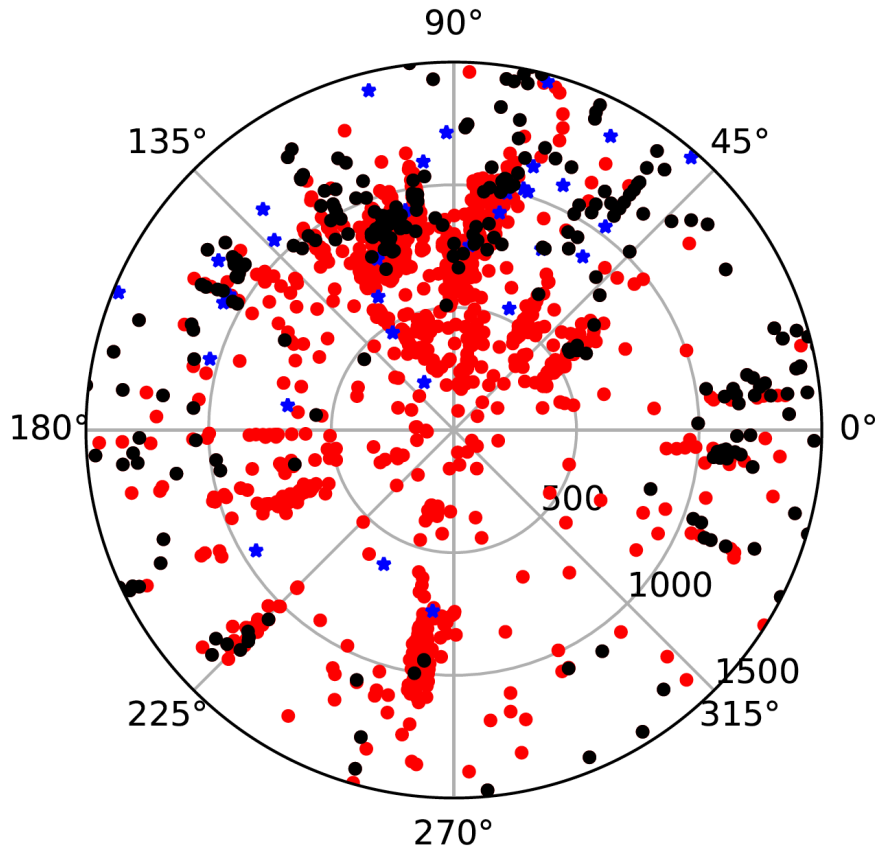
Candidates



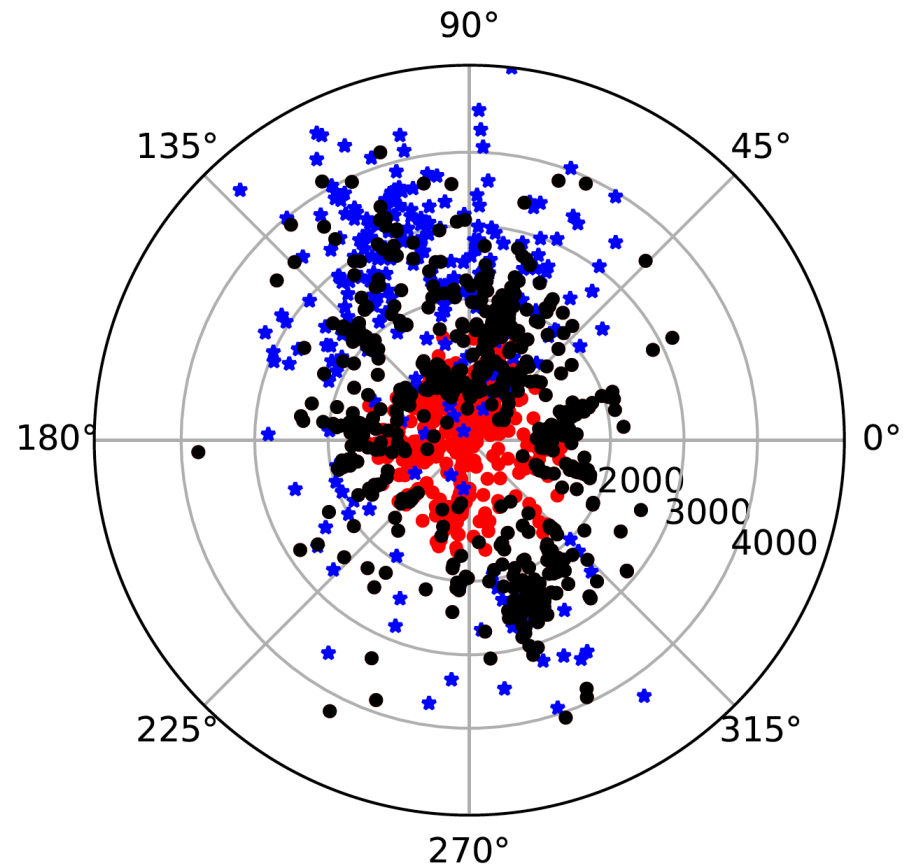
Training Set



Spatial distribution

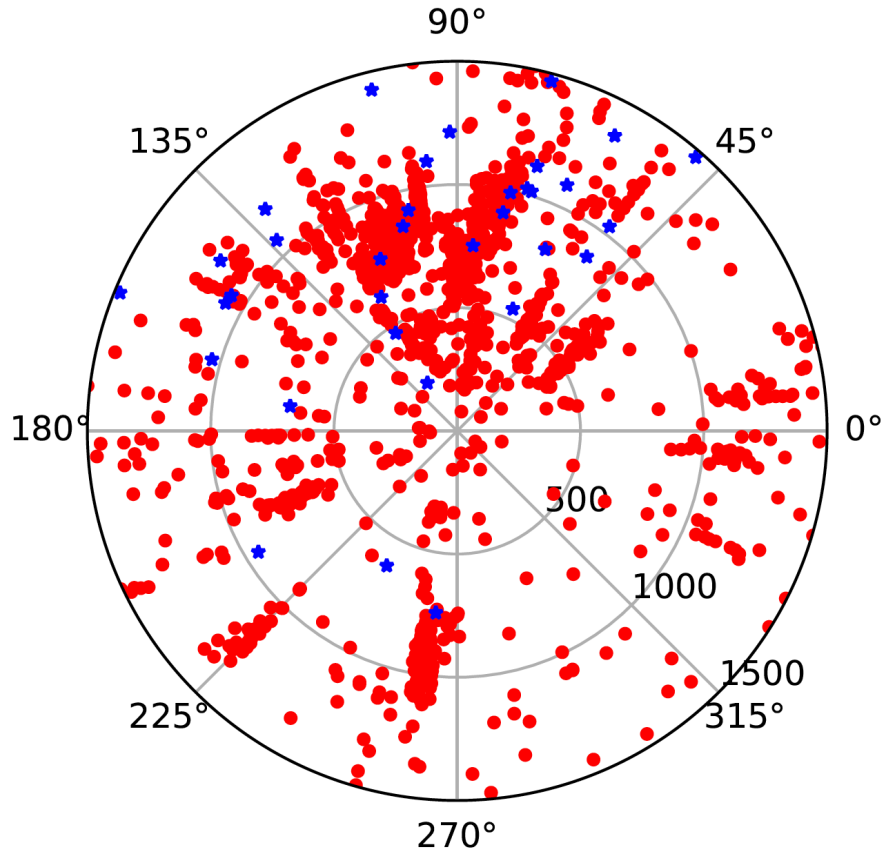


Within 1.5 Kpc

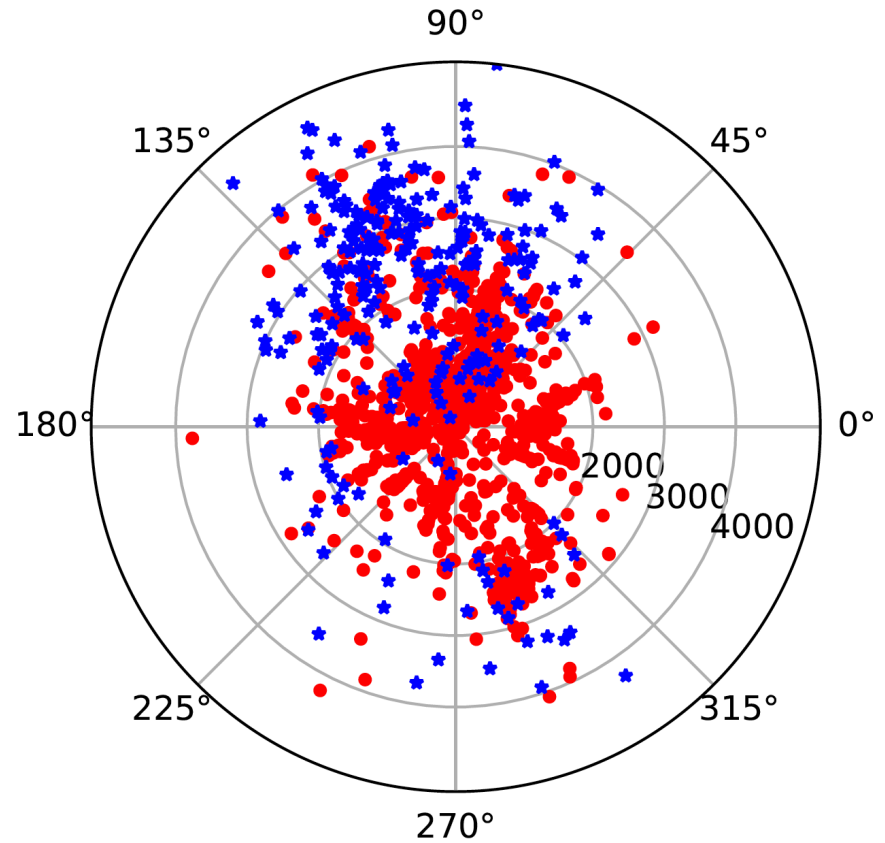


All (up to 5 Kpc)

Spatial distribution

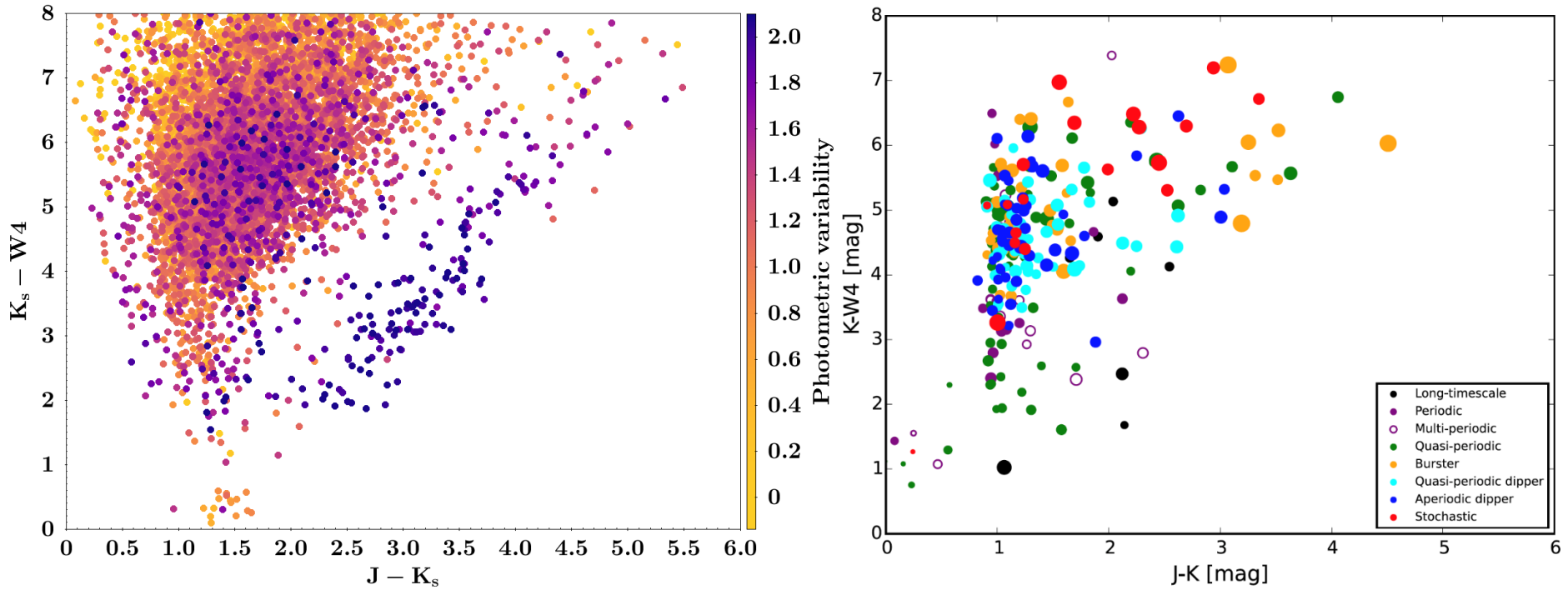


Within 1.5 Kpc



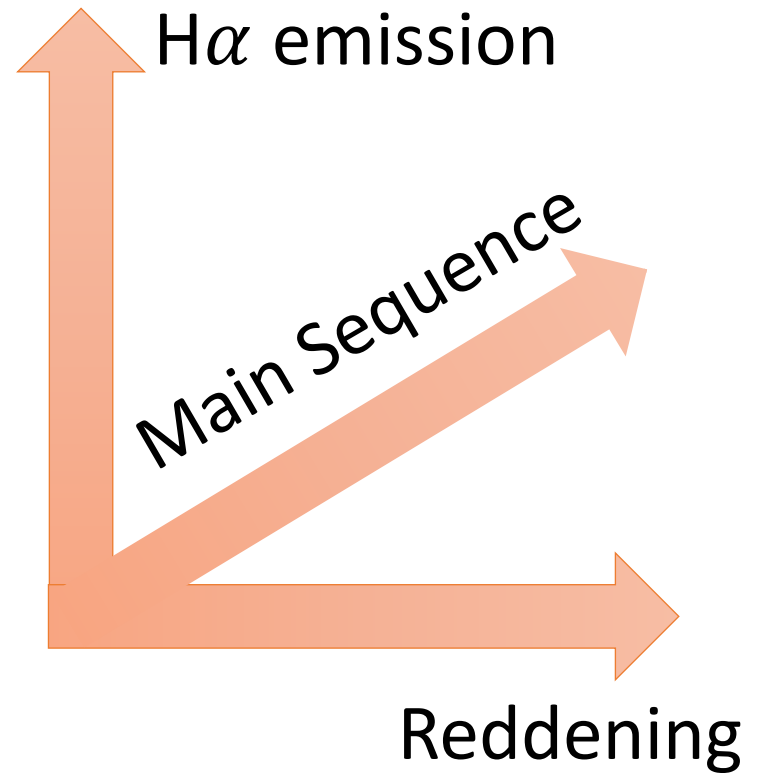
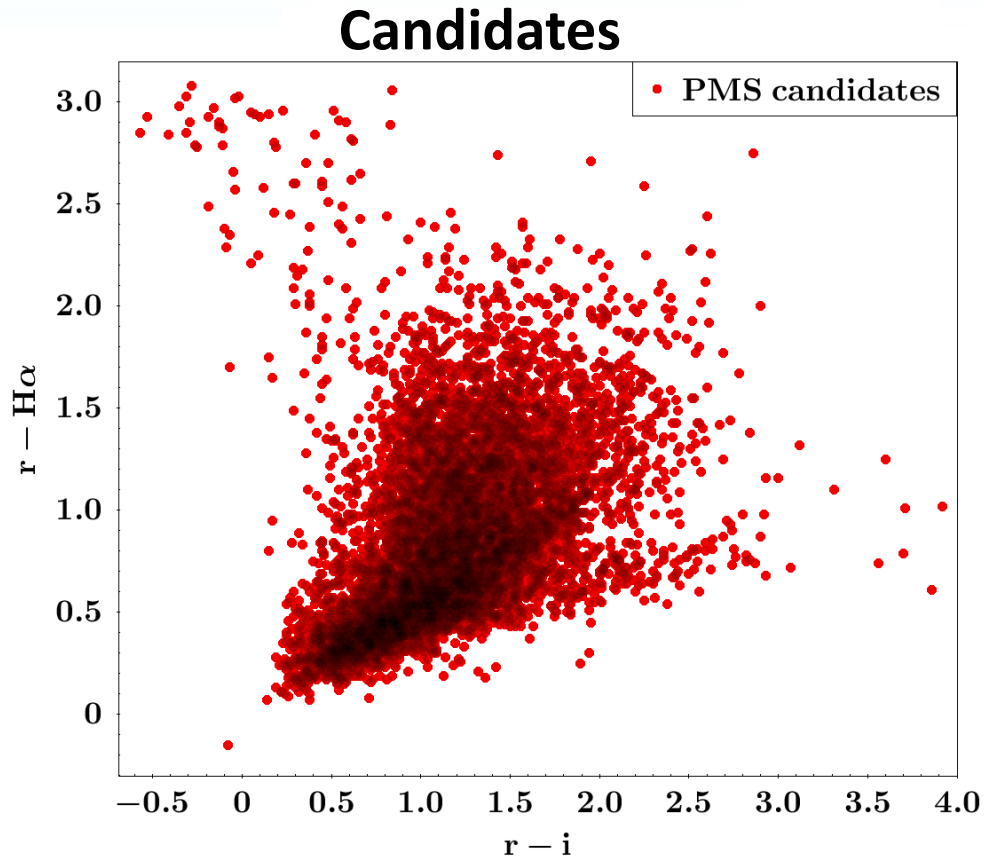
All (up to 5 Kpc)

Candidates

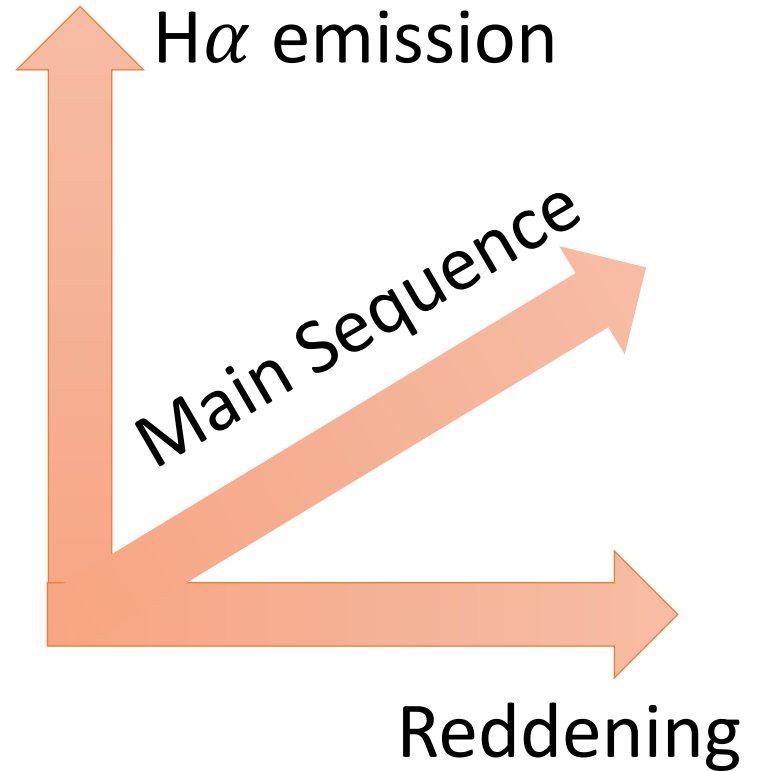
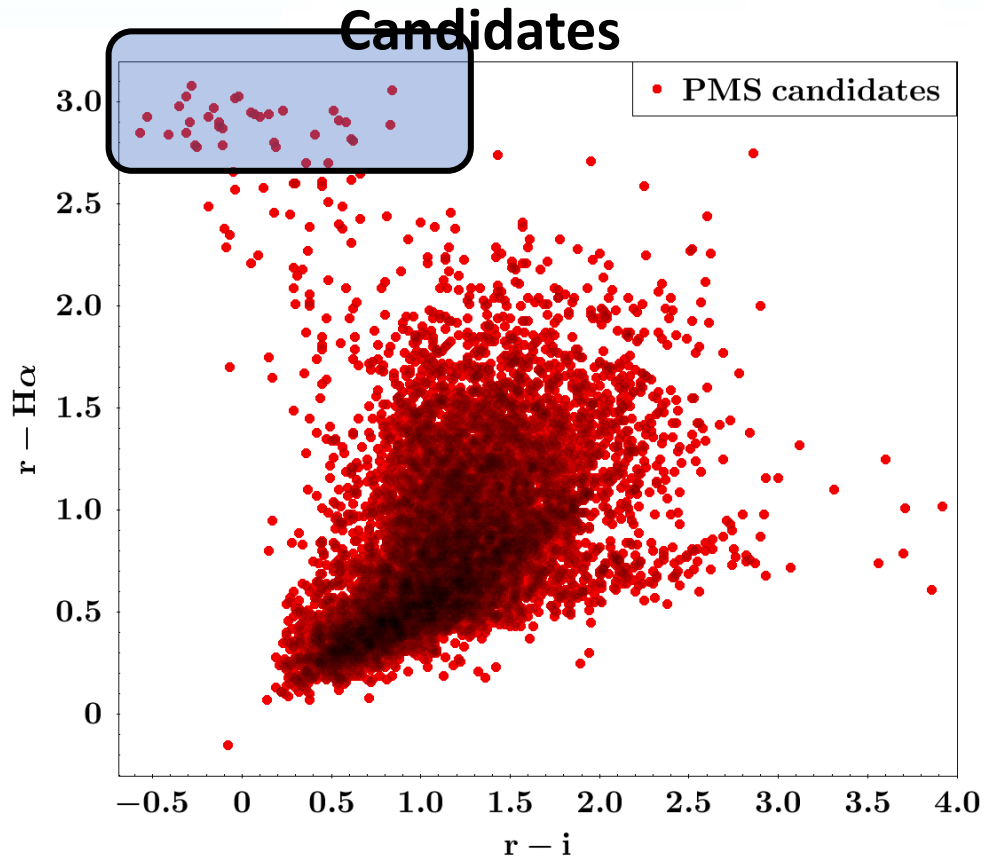


Cody & Hillenbrand, 2018

Caveats



Caveats



Caveats

Planetary Nebula!

