

Borse di studio GARR
Orio Carlini

Machine learning applicato all'astrofisica stellare

Francesco Turini

Tutor: Scilla Degl'Innocenti

Università di Pisa, dipartimento di Fisica

- **Astrofisica stellare**

- La sua importanza, confronto teoria osservazione, le survey.

- **Il problema da risolvere**

- I dati, l'obiettivo, lo stato dell'arte, l'approccio di questo progetto.

- Identificazione delle popolazioni stellari: il clustering

- L'importanza degli errori, **Potential-based** clustering.

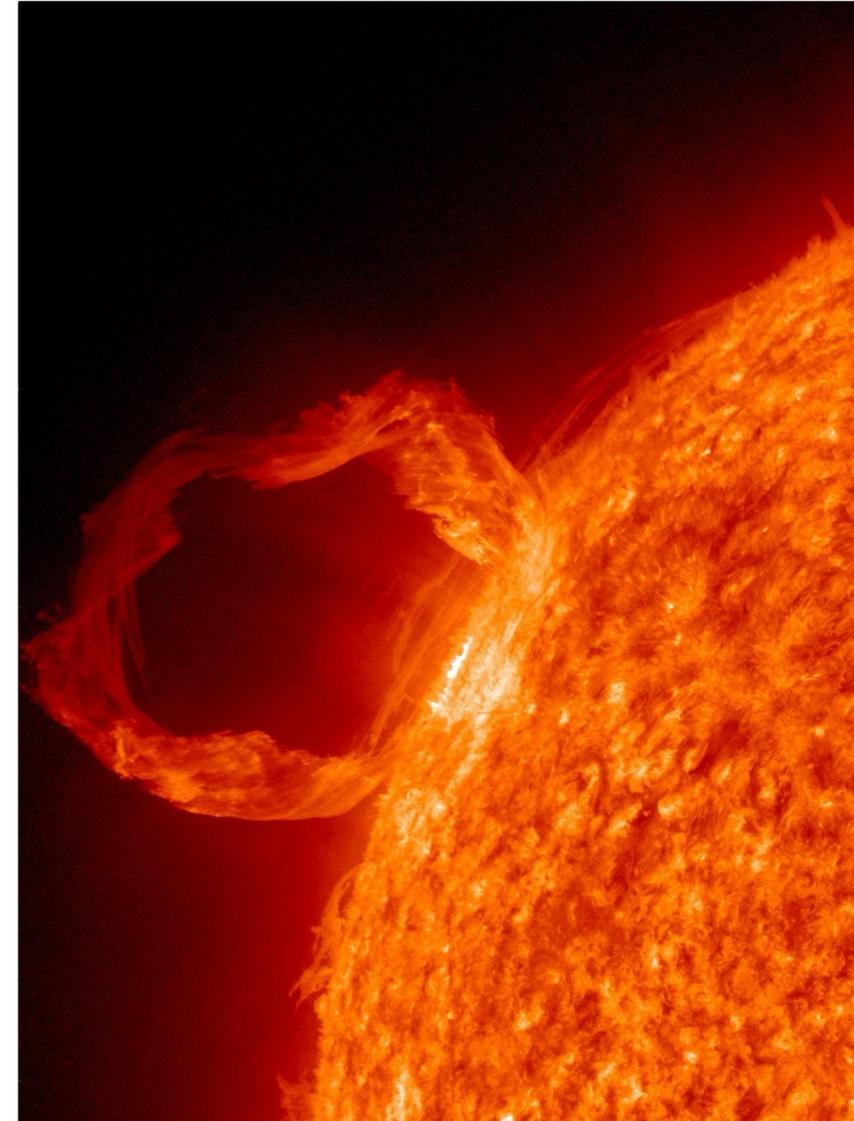
- **Prossimi step.**

Astrofisica stellare

Spazio, ultima frontiera!

Cosa è?

È quella branca della fisica che si occupa dello studio della **formazione** e dell'**evoluzione** delle stelle.



[1i]

Perché è importante?



[2i]

Lo studio delle stelle ci permette di:

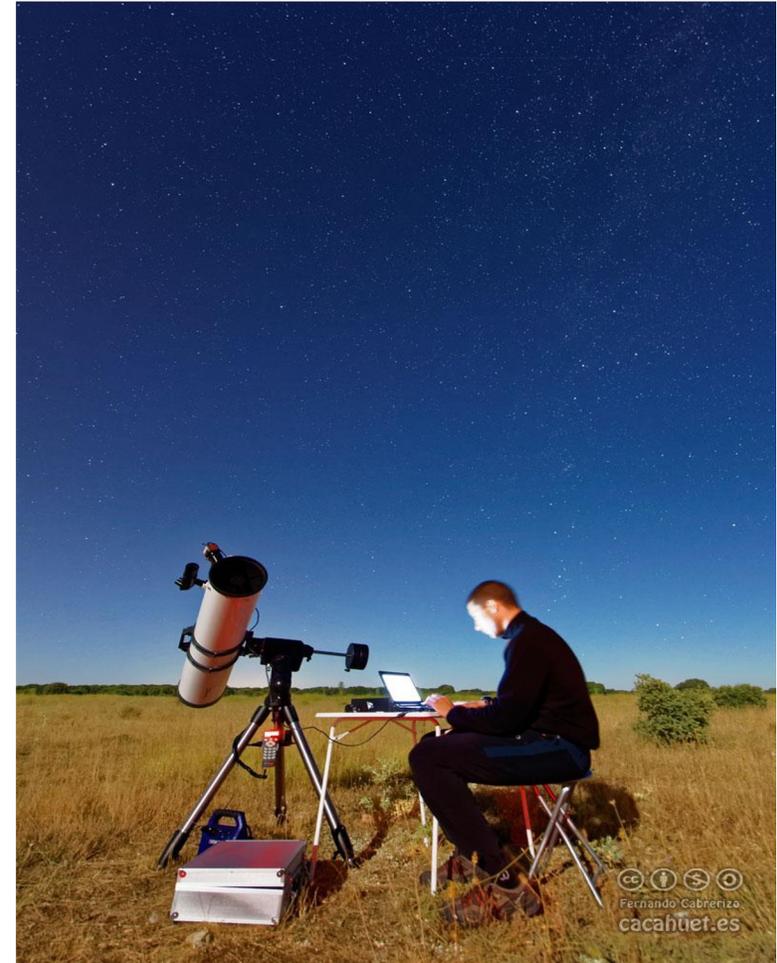
- Indagare la **struttura** e la storia della nostra **galassia** (Archeologia Galattica)
- Sviluppare **modelli cosmologici**
- Determinare l'età dell'universo (Universo locale)

Confronto teoria osservazione

Lo studio delle stelle si basa sul confronto

teoria (**modelli simulati**)

osservazioni (**survey**).



[3i]

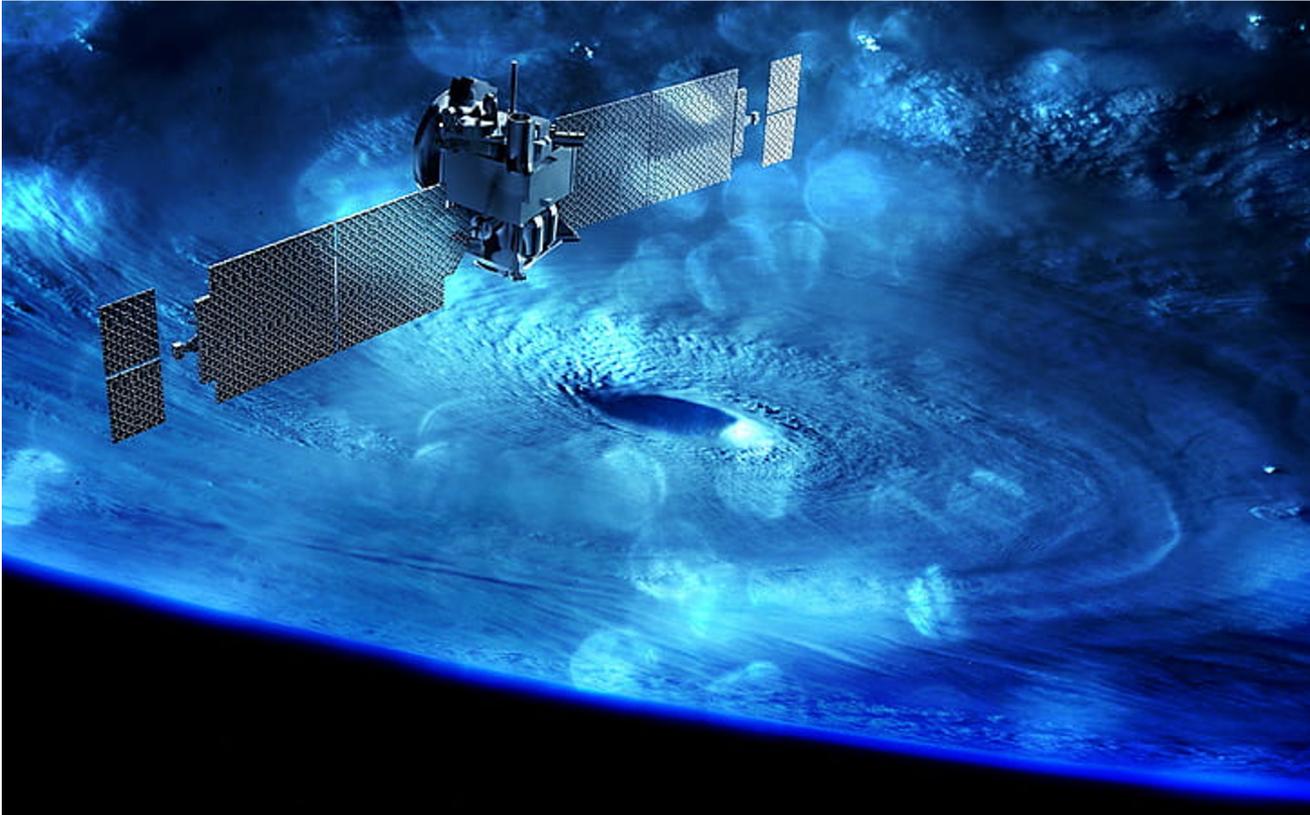
I modelli

Modelli: esistono vari programmi (FRANEC, MESA) che date:

- le equazioni di struttura stellare
- i vari rate delle reazioni nucleari
- delle **condizioni chimico-fisiche iniziali simulano** i vari processi all'interno della stella.



Le survey



Survey: sono campagne di prese dati, fatte tramite vari **telescopi** con **sensibilità diverse** alle varie lunghezze d'onda della luce

[5i]

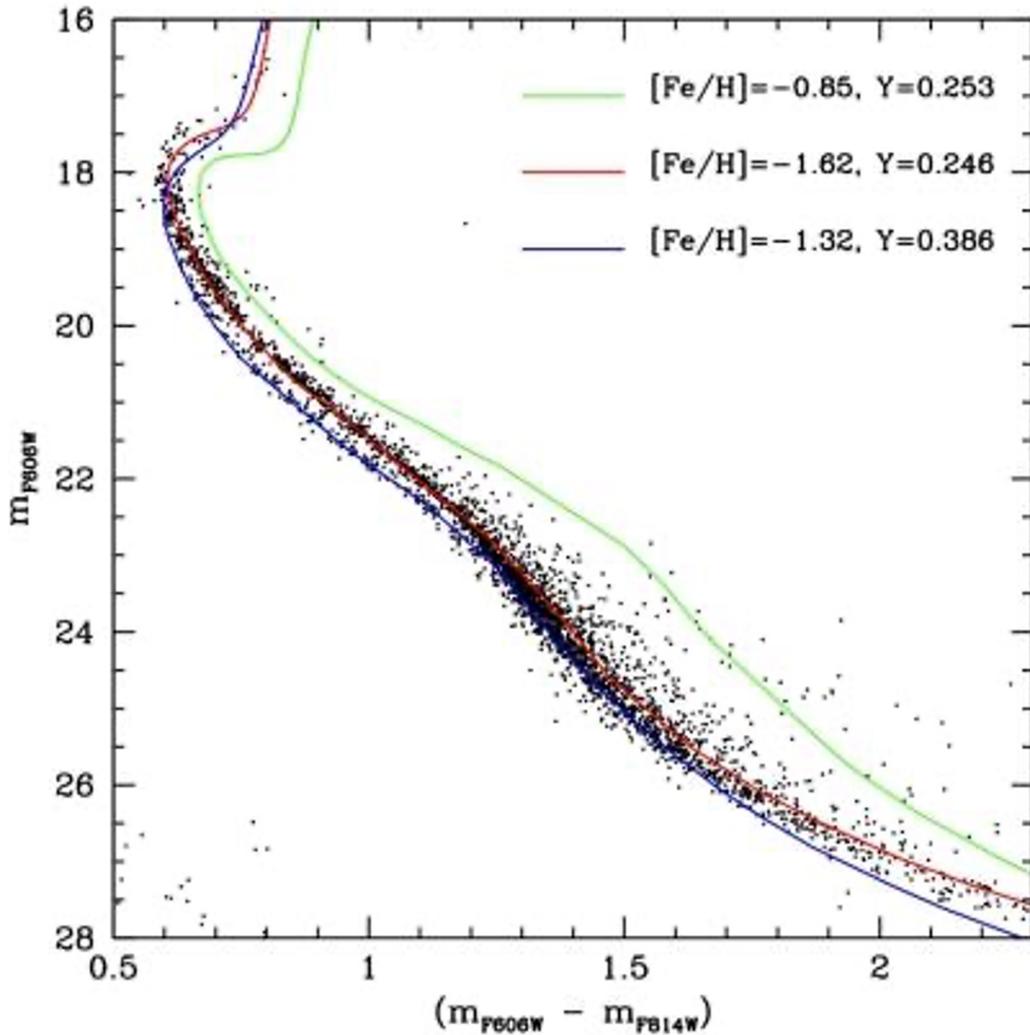
Le survey: GAIA, APOGEE, HST

GAIA: una survey moderna (dr3=2022), con un **ampio dataset**, precisa nella misura della posizione, della distanza e dei moti propri.

APOGEE: una survey che fa misure spettroscopiche (**chimica**), dataset molto meno ampio rispetto a GAIA.

Hubble Space Telescope: filtri sensibili alle differenze di abbondanze di alcuni elementi per stelle di RGB (**chromosome map**).

Il confronto



Dati i dati delle survey si va a cercare la **simulazione** che **meglio rappresenta i dati**.

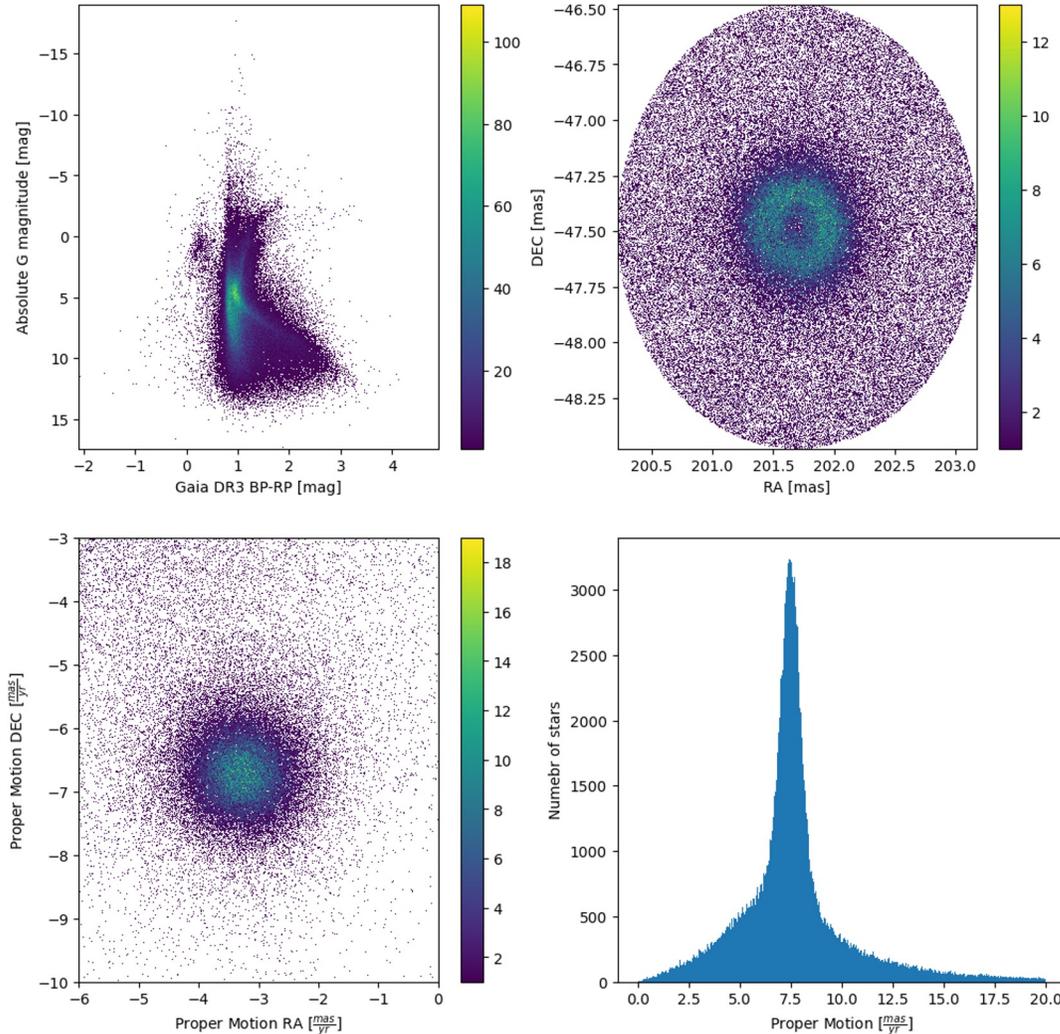
Quindi bisogna fare più simulazioni con **vari set di parametri** per poi andarle a confrontare

[1]

Il problema da risolvere

Houston, we have a problem

I dati dalle survey: Omega Centauri



Omega Centauri è plausibilmente un nucleo di una galassia nana catturata dalla Via Lattea.

Al suo interno vi è stata osservata una presenza di **multi popolazioni stellari**.

Come si analizzano?

Lo scopo del progetto:

1) Identificare le stelle che fanno parte di un **ammasso**

2) Identificare le varie **popolazioni stellari** all'interno dell'ammasso

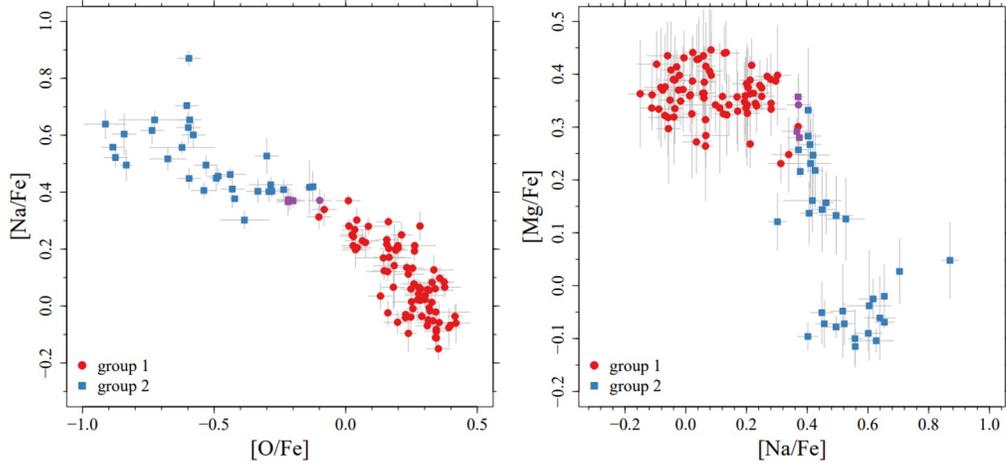


[6i]

3) Classificare la singola popolazione (età, metallicità ...)

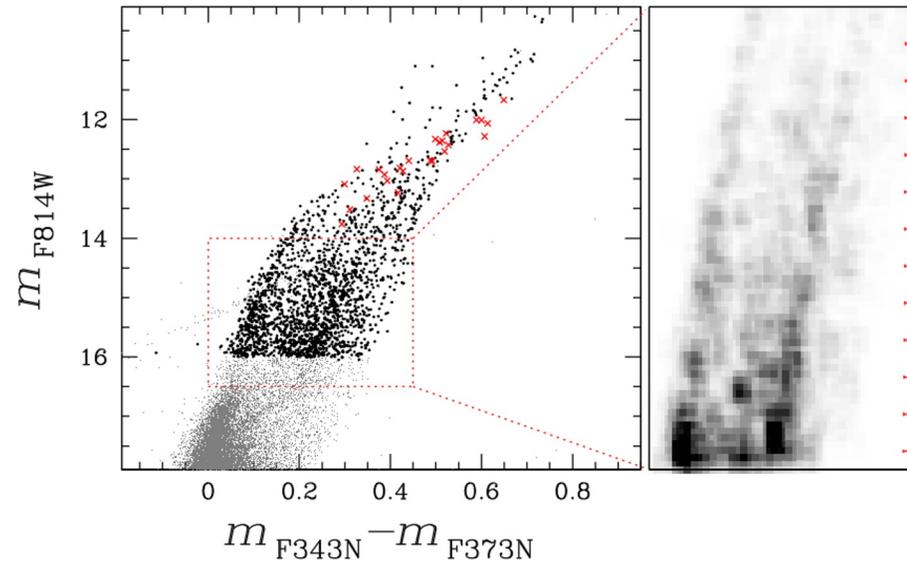
4) Determinare le caratteristiche delle **singole stelle** (massa)

Lo stato dell'arte?



Identificazione di popolazione tramite **chromosome map** [3]

Identificazione di popolazioni tramite anti-correlazioni nelle **abbondanze chimiche** di alcuni elementi [2]



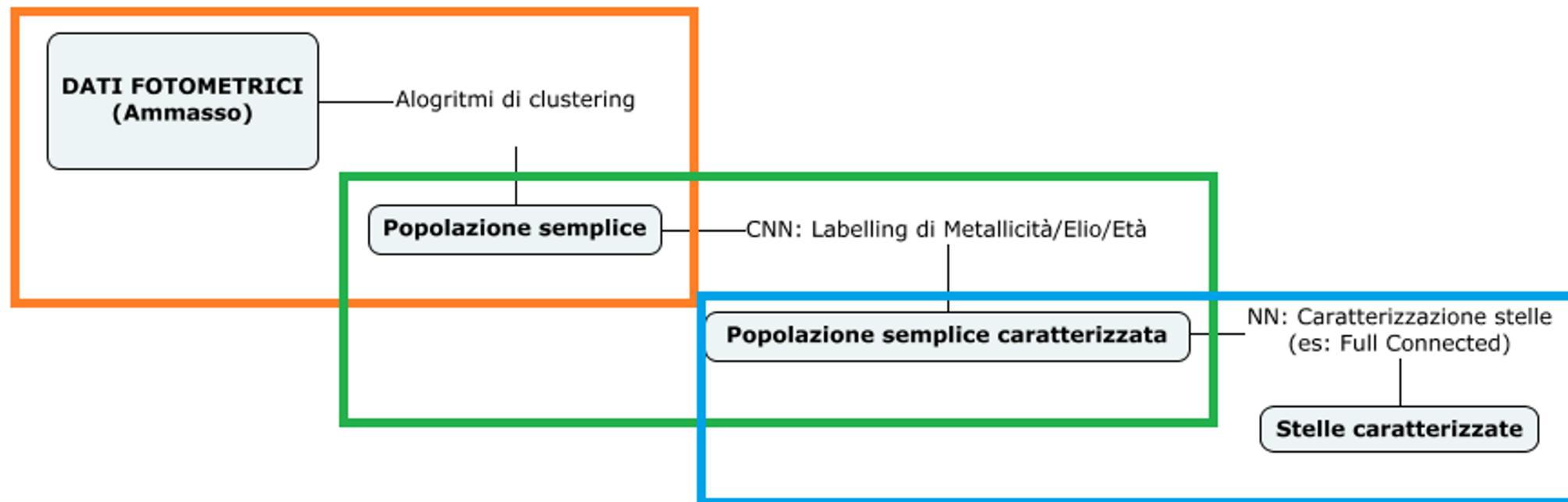
Pochi dati

Sia l'anti-correlazione negli elementi chimici sia la chromosome map **separano "bene" i dati**. Solo che i dati relativi ad esse sono a disposizione per un **numero limitato** di stelle.



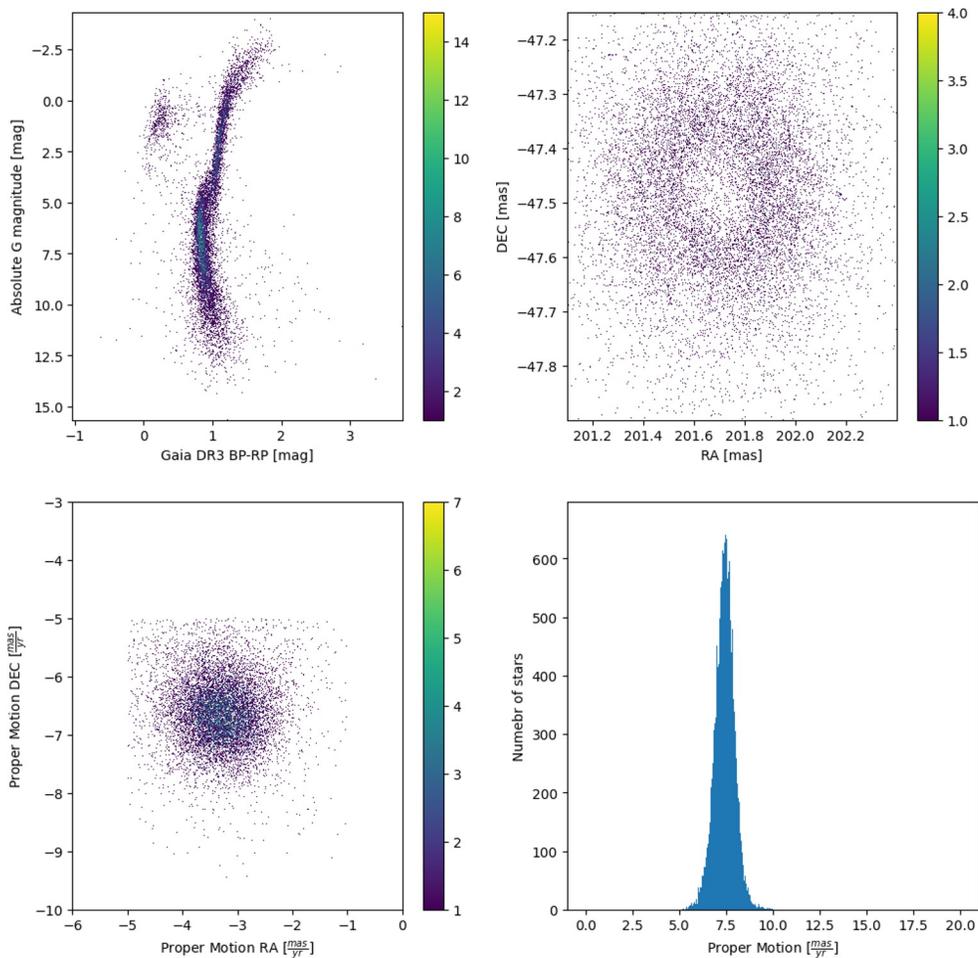
Es: per Omega Centauri vi è un **fattore di circa 150** tra dati **GAIA** ed **APOGEE**.

L'approccio di questo progetto



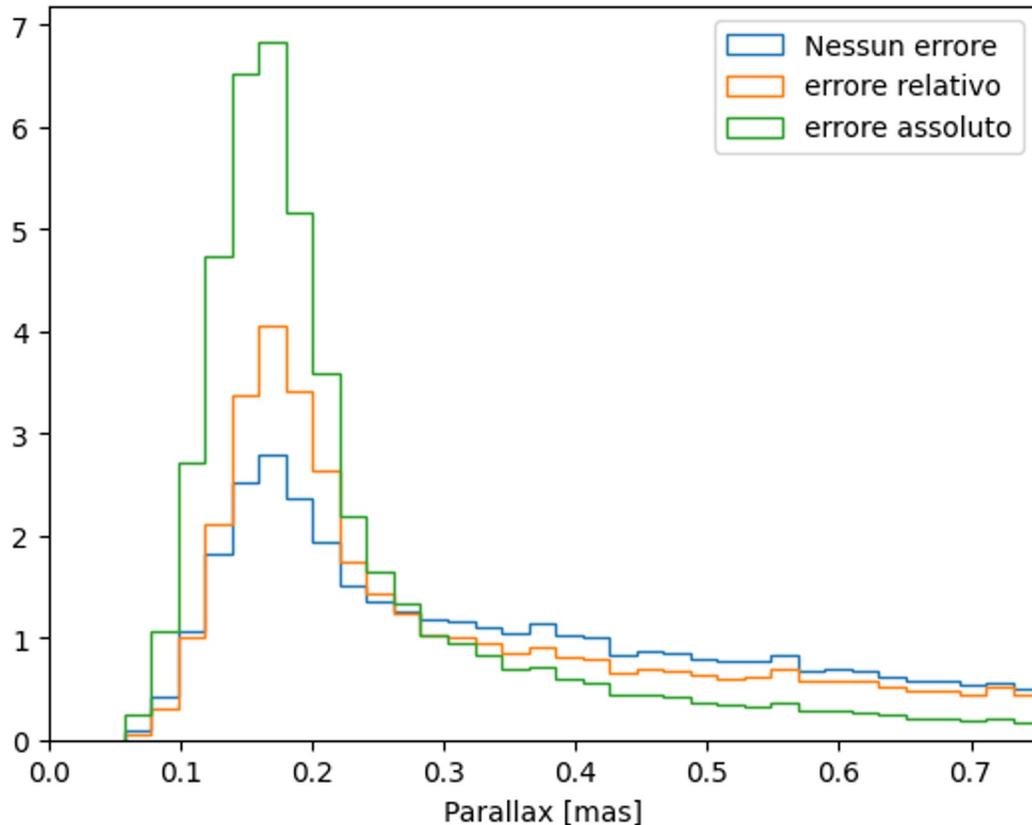
Identificazione delle popolazioni stellari

I primi passi: il clustering



- **Cut grezzi**
- Tentativo di identificare l'ammasso con vari **algoritmi standard** di clustering (k-means, fuzzy, DBSCAN, SVM)
- **Senza** però riuscire a **separare le stelle di campo.**

L'importanza degli errori



Varie distribuzioni della parallasse:

- Si popola con:

$$Parallax \xrightarrow{\left[\frac{1}{errParallax} \right]} Parallax$$

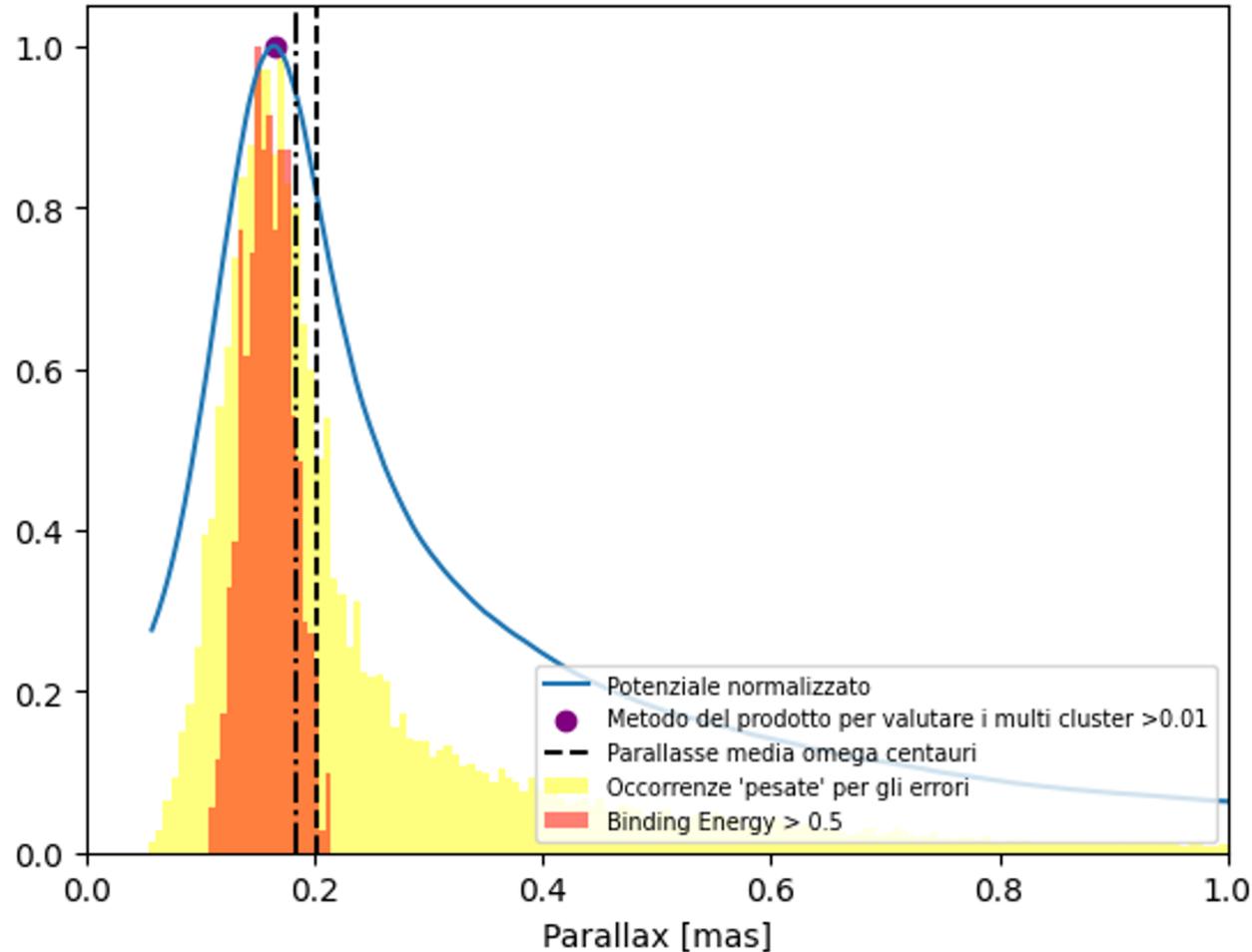
- Si nota che **l'errore evidenzia meglio l'ammasso**

Potential-Based Clustering

Si può sfruttare l'**errore di misura** anche in **algoritmi di clustering**?

- Si può modificare un algoritmo **Potential-Based** [4]
- Si definisce la massa di ogni singolo come: $mass = \frac{\min(error)}{error}$
- Ogni **feature** può definire una "**forza**" differente
- Questo algoritmo ottimizzato con algoritmi di hierarchical clustering permette di **separare meglio le stelle**.

Potential-Based Clustering sulla parallasse di Omega Centauri



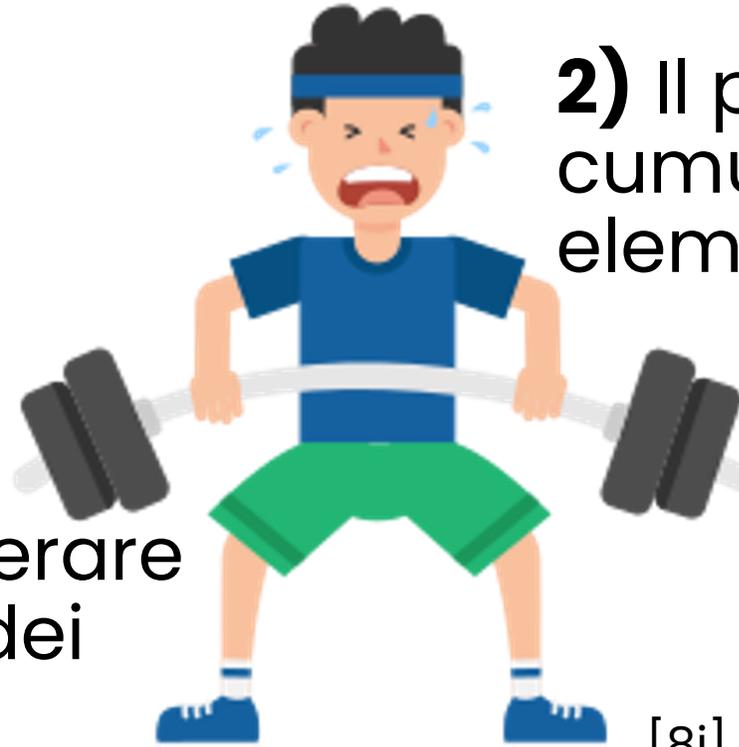
Inoltre si può definire:

- Metodo di **valutazione del numero dei cluster** (potenziale*parente)
- La **“Binding Energy”**

Potenzialità e punti di forza dell'algoritmo

1) Non è influenzato dalla "forma" dei dati

3) È possibile considerare le stelle per cui uno dei dati di interesse è mancante come non influenzato da una delle forze considerate



2) Il potenziale è cumulativo tra i vari elementi

4) **Sensibile all'errore** di misura

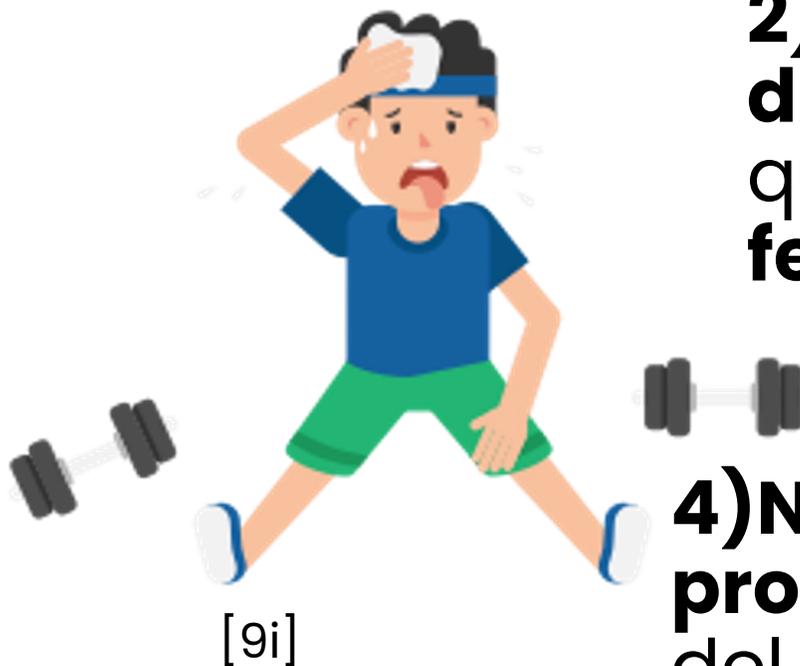
5) Può essere utilizzato anche per andare a **cercare correlazioni** stellari **non ancora identificate**

[8i]

Problemi dell'algoritmo

1) Relativamente lento, complessità $O(N^2)$

3) In caso di **manca**nza di una **feature** per un elemento **non** è ben definita la **distanza**



2) Difficile dare il **concetto di massa** di un elemento quando è definito da **più features**

4) **Non** è definita la **probabilità di membership** del singolo elemento (soluzioni: fuzzy, cut binding energy, cut potenziale)

Prossimi step

E ora?

Che cosa c'è da fare:

- Creare un **montecarlo** per testare meglio l'algoritmo



[10i]

- Applicarlo al **caso di Omega Centauri** e/o di NGC 2808

- Passare allo **sviluppo** e applicazione della **CNN** per la classificazione delle singole popolazioni stellari

Referenze

- [1] King, I. R., Bedin, L. R., Cassisi, S., Milone, A. P., Bellini, A., Piotto, G., Anderson, J., Pietrinferni, A., & Cordier, D. (2012). "Hubble Space Telescope Observations of an Outer Field in Omega Centauri: A Definitive Helium Abundance." *The Astronomical Journal*, 144(1), 5.
- [2] Valle, G., Dell'Omodarme, M., & Tognelli, E. (2022). "Robust statistical tools for identifying multiple stellar populations in globular clusters in the presence of measurement errors. A case study: NGC 2808." *Astronomy & Astrophysics*, 658(A141). doi:10.1051/0004-6361/202142454
- [3] A P Milone, A F Marino, A Renzini, C Li, S Jang, E P Lagioia, M Tailo, G Cordoni, M Carlos, E Dondoglio, A chromosome map to unveil stellar populations with different magnesium abundances. The case of ω Centauri, *Monthly Notices of the Royal Astronomical Society*, Volume 497, Issue 3, September 2020, Pages 3846–3859, <https://doi.org/10.1093/mnras/staa2119>
- [4] Lu, Y., & Wan, Y. (2013). "PHA: A fast potential-based hierarchical agglomerative clustering method." *Pattern Recognition*, 46(5), 1227–1239. doi:10.1016/j.patcog.2012.11.017

Borse di studio GARR
Orio Carlini

Crediti immagini

[1i] https://pxhere.com/it/photo/1360852?utm_content=shareClip&utm_medium=referral&utm_source=pxhere

[2i] https://pxhere.com/it/photo/862077?utm_content=shareClip&utm_medium=referral&utm_source=pxhere

[3i] <https://cacahuet.es/docs/2016/esos-locos-astronomos>

[4i] https://commons.wikimedia.org/wiki/File:CERN_Computer_Center_07.jpg

[5i] https://www.pickpik.com/mission-to-mars-mars-probe-space-travel-research-science-astronomy-59187#google_vignette

[6i] Man Working at his Desk Cartoon Vector.svg from Wikimedia Commons by Videoplasty.com, CC-BY-SA 4.0

[7i] https://pxhere.com/it/photo/888565?utm_content=shareClip&utm_medium=referral&utm_source=pxhere

[8i] Man Lifting Barbell Cartoon.svg from Wikimedia Commons by Videoplasty.com, CC-BY-SA 4.0

[9i] [Man Tired After Workout Cartoon.svg](#) from Wikimedia Commons by Videoplasty.com, [CC-BY-SA 4.0](#)

[10i] Amanda Wentworth: <https://www.flickr.com/photos/146278242@N03/31762655127>

Borse di studio GARR
Orio Carlini

Grazie per l'attenzione!

Backup Slides

L'algoritmo del potential-Based clustering

Ogni dato:

1. Genera un potenziale

$$\phi_{ij} = \frac{mass_j}{r_{ij}}$$

1. Vede un potenziale

$$\Phi_i = \sum_{j=1..N} \Phi_{ij}(r_{ij})$$

Questo algoritmo ha una complessità di ordine $O(N^2)$, ed è definito solo dalle distanze e dalle masse dei singoli dati. In questo caso è possibile definire anche una “**binding energy**”. Il suo senso è quello che un dato più preciso è più legato al cluster