



UNIVERSIDAD DE LA REPUBLICA
FACULTAD DE CIENCIAS
Instituto de Física
Departamento de Astronomía

Inferencia de la función de masa inicial de grupos estelares co-móviles jóvenes del entorno solar

Tesis de Maestría en Física, Opción Astronomía

Por: **Lic. Rafael Bertolotto**

Instituto de Física, Departamento de Astronomía, Udelar

Tutor: Dr. Juan José Downes (Instituto de Física, Departamento de Astronomía, Udelar ;

Programa de Desarrollo de las Ciencias Básicas)

Co-tutor: Dr. Genaro Suárez (American Museum of Natural History, NY, EEUU)

Montevideo, julio de 2024

Resumen

El entorno solar está poblado por grupos de estrellas jóvenes y co-móviles. Su juventud sugiere que se trata de remanentes de grupos estelares que se están dispersando en el disco galáctico lo cual combinado con su cercanía los vuelve candidatos ideales para el estudio de la Función de Masa Inicial (IMF).

En este trabajo, desarrollamos e implementamos un método para detectar miembros de estos grupos a partir de su distribución en un espacio cinemático de cinco dimensiones construido con datos del sondeo Gaia DR3 e inferir sus IMF en el rango de masas $0,01 < m/M_{\odot} < 5$. Aplicamos el método a diez grupos conocidos, detectando un total de 1229 candidatos cinemáticos y fotométricos a miembros, recuperando 66 % de los miembros conocidos y detectando 604 nuevos candidatos de los cuales estimamos que entre 105 y 287 sean contaminantes. Finalmente, obtuvimos las diez IMF, nueve de las cuales son inéditas, y calculamos la IMF normalizada promedio de los diez grupos. Hallamos que las parametrizaciones de la IMF como funciones log-normales para masas $m/M_{\odot} < 1$ son consistentes con las estimadas por [Chabrier \(2003\)](#) y [Suárez et al. \(2019\)](#), tanto a partir de los miembros de la literatura como a partir de nuestras detecciones. Además, nuestros resultados presentan un mejor acuerdo con la IMF de [Salpeter \(1955\)](#) en el rango de masas $m/M_{\odot} > 1$ que las funciones de masa obtenidas de los miembros de la literatura. Finalmente, hallamos indicios de contaminaciones en promedio en torno a $1 M_{\odot}$ y de incompletitud para objetos con masas $m/M_{\odot} < 0,02$ que serán estudiados en futuros trabajos. Las incompletitudes observadas pudieran ser consecuencia de un proceso de dispersión cinemático que amerita futuros estudios.

Este trabajo representa un avance importante en el estudio de los grupos jóvenes del entorno solar tanto por ser el primero en reportar las funciones de masa de nueve de estos grupos como por proporcionar información sobre su posible origen.

Abstract

The solar neighborhood is populated by groups of young, co-moving stars. Their youth suggests that they are remnants of stellar groups dispersing in the galactic disk, which combined with their proximity makes them ideal candidates for the study of the Initial Mass Function (IMF).

In this work, we develop and implement a method to detect members of these groups based on their distribution in a five-dimensional kinematic space constructed from Gaia DR3 survey data, and to infer their IMF in the mass range $0,01 < m/M_{\odot} < 5$. We apply the method to ten known groups, detecting a total of 1229 kinematic and photometric candidate members, recovering 66% of the known members and detecting 604 new candidates from which we estimate that between 105 and 287 are contaminants. Finally, we obtained the IMF for the ten groups, nine of which are new, and calculated the average normalized IMF of the ten groups. We found that IMF parameterizations as log-normal functions for masses $m/M_{\odot} < 1$ are consistent with those estimated by [Chabrier \(2003\)](#) and [Suárez et al. \(2019\)](#) from both literature members and our detections. Additionally, our results show better agreement with the IMF of [Salpeter \(1955\)](#) in the mass range $m/M_{\odot} > 1$ than the mass functions obtained from literature members. Finally, we found evidence of contamination on average around $1 M_{\odot}$ and incompleteness for objects with masses $m/M_{\odot} < 0,02$ that will be studied in future work. The observed incompleteness could be the result of a kinematic dispersion process that needs further studies.

This work represents a significant advance in the study of young groups in the solar neighborhood, both as the first to report the mass functions of nine of these groups and by providing information on their possible origin.

Agradecimientos

A Juan José y Genaro a quienes agradezco no solamente por todo el esfuerzo, la paciencia y el trabajo que invirtieron y compartieron conmigo en este proyecto, ni solamente por haber sido actores fundamentales de mi formación académica como personal, sino también por las largas conversaciones, risas y cervezas. Saber que este trabajo no es solamente mío sino también suyos es para mi una fuente de felicidad y honor.

A Miguel Campiglia, Carlos Román Zúñiga, Sebastián Bruzzone y Pau Ramos. Por haber dedicado su tiempo y esfuerzo a leer y evaluar esta tesis así como por sus comentarios y preguntas pertinentes que ayudaron a enriquecer este trabajo.

A las instituciones de la UdelaR que me permitieron realizar este trabajo: el Instituto de Física, Pedeciba, la Comisión Académica de Posgrado y ClusterUy.

A Cecilia, Luis, Mauro, Bruno y Rodrigo. Por todas las conversaciones compartidas, sus ideas, preguntas y sabiduría, que contribuyeron a enriquecer y desarrollar este trabajo de manera directa como indirecta.

A mis demás amigos Román, Juan, Guillermo, Jorge, Diego, Lucas, Gaspar, Rodrigo, Juliana, Marcio, Romina, Mateo, Natalia, Iaron, Iván y Leandro. Por todos los momentos compartidos en estos años.

A mis padres Fernando y Lucia y a mis hermanos Nicolas y Ricardo, por nunca haber dejado de quererme y apoyarme en los tiempos más felices como en sus momentos más difíciles.

Contenido

Resumen	I
Abstract	II
Agradecimientos	III
1. Los Grupos Jóvenes Cercanos y el estudio de la IMF	1
1.1. Introducción	1
1.1.1. La Función de Masa Inicial	2
1.1.2. Los grupos co-móviles jóvenes del entorno solar	4
1.1.3. Ventajas y desafíos del estudio de la IMF de los Grupos Jóvenes	5
1.2. Motivaciones y objetivos	6
2. La muestra de estudio	8
2.1. Hacia un muestreo completo del Entorno Solar con Gaia DR3	8
2.1.1. Selección del Entorno Solar en Gaia DR3	8

2.1.2.	Método para estimar la completitud mediante la Función de Selección	10
2.1.3.	Completando la muestra de estudio	14
2.1.4.	Sobre la calidad de los datos	15
2.2.	Los NYMGs conocidos del entorno solar	18
2.2.1.	La base de datos MOCA y el catálogo maestro	18
2.2.2.	La cinemática de los NYMGs	20
3.	Detección de Grupos	24
3.1.	El método de detección	24
3.1.1.	El algoritmo DBSCAN	24
3.1.2.	Esquema general del algoritmo de detección	25
3.2.	Detección de fluctuaciones espurias del campo	28
3.2.1.	Construcción de campos sintéticos	28
3.2.2.	Frecuencia de detección de sobre-densidades espurias	29
3.3.	Candidatos a grupos en las detecciones significativas	30
4.	Inferencia de las IMFs de los NYMGs estudiados	35
4.1.	Inferencia Bayesiana de masas a partir de la fotometría	35
4.1.1.	Las IMFs inferidas	37
4.1.2.	La Función de Masa normalizada media	42
5.	Análisis, discusión y conclusiones	45

5.1. Métodos y herramientas desarrollados en este trabajo	45
5.2. Los NYMGs detectados y su comparación con la literatura	46
5.3. Las IMFs de los NYMGs	47
5.4. Conclusiones científicas generales	49
5.5. El futuro del trabajo	50
A. Apéndices	51
A.1. Estimación del parámetro ϵ	51
A.2. Los NYMGs detectados	54
A.3. Las IMFs de los NYMGs detectados	57
A.4. Estimación cinemática de la pureza de las detecciones	59
A.5. Desarrollo de un método para evaluar la pureza y completitud de la IMF	61
Abbreviations	63

Capítulo 1

Los Grupos Jóvenes Cercanos y el estudio de la Función de Masa Inicial

1.1. Introducción

La conversión de gas a estrellas en el medio interestelar resulta en una diversidad de objetos de tipo estelar con masas que van desde centésimas de la masa del Sol hasta decenas de su masa. Estudiar la proporción de objetos con distintas masas es crucial tanto desde un punto de vista teórico, ya que la forma de la distribución de masas es el consecuencia de los procesos físicos que participan de la formación estelar ([Hennebelle & Chabrier, 2008](#)), como desde un punto de vista práctico, ya que es un parámetro fundamental para, por ejemplo, inferir la masa de poblaciones no resueltas como galaxias y cúmulos lejanos (e.g. [Hopkins, 2018](#); [Offner et al., 2014](#)).

Aunque la distribución de masas estelares es una propiedad esencial para la caracterización de toda población estelar, aún persisten dudas sobre su origen físico y continúa siendo necesaria su determinación observacional en poblaciones estelares con distintas propiedades.

1.1.1. La Función de Masa Inicial

La Función de Masa Inicial (del inglés *Initial Mass Function* (IMF)), se define como la distribución logarítmica de las masas estelares que resulta de un único evento de formación estelar (e.g., [Hopkins, 2018](#)). La Figura 1.1 muestra las cuatro parametrizaciones usuales de la IMF cuyos correspondientes expresiones funcionales y rango de masa son:

- Ley de potencia de Salpeter ([Salpeter, 1955](#)):

$$\Psi(\log m) \propto m^{-x} \text{ o en la escala lineal } \Psi(m) \propto m^{-\alpha} \text{ con } x = 1,35 \text{ y } x = \alpha - 1 \text{ para } m > M_{\odot}$$

- Log-Normal ([Miller & Scalo, 1979](#); [Scalo, 1986](#); [Chabrier, 2003](#)):

$$\Psi(\log m) \propto \begin{cases} e^{-\left(\frac{\log m - \log m_c}{2\sigma^2}\right)^2} & \text{si } m < 1M_{\odot} \\ m^{-x} & \text{si } m > 1M_{\odot} \end{cases} \quad (1.1.1)$$

- Leyes de potencia a trazos de Kroupa ([Kroupa, 2001](#)):

$$\Psi(m) \propto m^{-\alpha_i} \begin{cases} \alpha_1 = 0,3 \pm 0,7 & \text{si } 0,01 < m/M_{\odot} < 0,08 \\ \alpha_2 = 1,3 \pm 0,5 & \text{si } 0,08 < m/M_{\odot} < 0,50 \\ \alpha_3 = 2,3 \pm 0,3 & \text{si } 0,50 < m/M_{\odot} < 1,00 \\ \alpha_4 = 2,3 \pm 0,7 & \text{si } 1,00 < m/M_{\odot} \end{cases} \quad (1.1.2)$$

- Trapped power law ([De Marchi et al., 2010](#)):

$$\Psi(m) \propto m^{-x} [1 - e^{(-m/m_c)^{\gamma+x}}] \quad (1.1.3)$$

donde $m_c = 0,15$, $x = 0,97 \pm 0,17$ y $\gamma = 1,54 \pm 0,64$ en el intervalo de masas $0,01 < M/M_{\odot} < 10$.

A pesar de la extensa investigación realizada desde la primera determinación observacional de la IMF por [Salpeter \(1955\)](#), la pregunta central sobre su universalidad, es decir, si la IMF de todas las poblaciones estelares son realizaciones de una misma distribución, aún es objeto de intensa investigación. Las revisiones de [Bastian et al. \(2010\)](#), [Offner et al. \(2014\)](#) y [Hopkins \(2018\)](#) muestran que la IMF en el rango de masas $0,1 < M/M_{\odot} < 10$ es aproximadamente consistente para diferentes poblaciones cercanas en la Vía Láctea, aunque se destacan algunas excepciones como la región Upper Sco ([Slesnick et al., 2008](#)).

En el dominio de masas sub-estelares ($0,01 < M/M_{\odot} < 0,072$) la situación continúa siendo muy incierta pues son muy pocas las IMF que en ese rango de masas muestran bajas incertidumbres (e.g. [Suárez et al., 2019](#); [Bayo et al., 2011](#); [Peña Ramírez et al., 2012](#)) y las limitaciones observacionales no han permitido todavía determinar la IMF para un número significativo de poblaciones.

En el extremo de altas masas, [Dib \(2023\)](#) encontraron una correlación entre la pendiente α de la ley de potencia de Salpeter en el extremo de alta masa de la IMF y la densidad superficial σ_{\star} de los cúmulos jóvenes descrita por la ley de potencias $\alpha = 1,31\sigma_{\star}^{-0,095}$ lo que sugiere una dependencia entre el proceso de formación de estrellas masivas y las propiedades del entorno en el que estas se forman.

La determinación observacional de la IMF es un proceso complejo que, esencialmente, puede ser dividido en dos grandes partes: la búsqueda de todas las estrellas miembro de una dada población estelar y la determinación de sus masas individuales. Ambas determinaciones son problemas en astrofísicos complejos y de gran relevancia. En resumen, determinar la membresía es un proceso complejo porque una estrella miembro y una estrella que no lo es pueden mostrar algunos observables en común y porque la probabilidad de detección es diferente para estrellas miembro de distintas masas debido a limitaciones observacionales. Ambas dificultades conducen a incompletitudes y contaminación que deben ser caracterizadas. Por otro lado, la determinación directa de las masas estelares es posible únicamente para estrellas en sistemas binarios, mientras que para estrellas aisladas depende de modelos de estructura, evolución y atmósferas estelares que no son igualmente precisos para estrellas de distintas masas.

Así, hasta qué punto los sesgos observacionales y metodológicos han influido en nuestra comprensión de la IMF sigue siendo incierto ([Hopkins, 2018](#)) y se necesitan estudios observacionales más precisos, estadísticamente robustos y en una variedad de poblaciones estelares para explorar la universalidad de la IMF. Para reducir el efecto de las incertidumbres observacionales en el estudio de la IMF, es preferible inferir esta distribución en poblaciones cercanas.

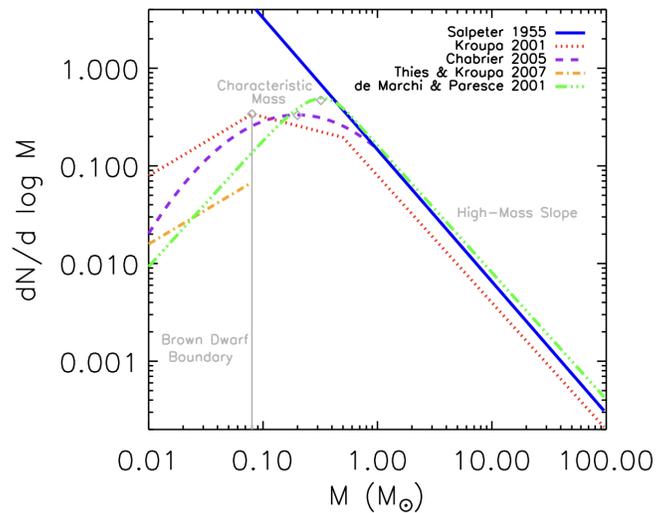


Figura 1.1: Parametrizaciones de la IMF. Tomado del artículo de revisión de [Offner et al. \(2014\)](#). La línea vertical señala el límite sub-estelar de masas.

1.1.2. Los grupos co-móviles jóvenes del entorno solar

Desde el primer descubrimiento de estrellas T-Tauri dispersas, co-móviles y cercanas mediante el sondeo Pico Dos Días ([Torres et al., 1999](#)), numerosos grupos de estrellas co-móviles y no ligadas gravitacionalmente han sido encontrados en la vecindad solar (e.g., [Torres et al., 2008](#); [Gagné & Faherty, 2018](#); [Kerr et al., 2021](#)), cuyas edades varían entre unos pocos Maños¹ y un Gaño², y con un número de miembros de entre unas pocas decenas hasta unos pocos miles. Entre estos grupos se encuentran los Grupos Jóvenes y co-Móviles Cercanos, (del inglés *Nearby Young Moving Groups* (NYMG)). En este estudio los definimos como aquellos grupos más jóvenes que 100 Maños y a distancias menores de 200 pc. Actualmente, alrededor de 20 grupos cumplen estas características, con un total de unos pocos cientos de miembros estelares conocidos.

Dado que las edades de estos grupos son un orden de magnitud menor que sus tiempos dinámicos entorno al centro galáctico, es poco probable que estos grupos se hayan formado mediante procesos resonantes como algunos otros grupos más viejos y más masivos. Al descartar este primer potencial origen, la hipótesis más aceptada respecto al origen de estos grupos sugiere que son remanentes de asociaciones

¹ 1 Maño = 10^6 años

² 1 Gaño = 10^9 años

estelares que están siendo dispersadas a través del disco galáctico (Gagné et al., 2021) como consecuencia de que sus estrellas se formaron sin estar ligadas gravitacionalmente dentro de su nube molecular parental, como se espera de la mayoría de las poblaciones estelares (Lada & Lada, 2003).

A pesar del progreso significativo en la identificación de nuevos NYMG y sus miembros, existe un único estudio de la IMF en el contexto de los NYMG y corresponde al trabajo de Gagné et al. (2017), quienes infirieron la IMF del grupo TW-Hya a partir de una muestra de 55 miembros con masas $5 M_{\text{Jup}} < m < 2 M_{\odot}$. Encontraron una IMF bien descrita por una distribución log-normal con una masa característica $m_c = 0,21^{+0,11}_{-0,06} M_{\odot}$, consistente con resultados previos en el campo ($m_c = 0,20 \pm 0,05 M_{\odot}$; Chabrier et al., 2005) pero con un ancho característico $\sigma = 0,8^{+0,2}_{-0,1}$ dex que es significativamente mayor que los $0,3 - 0,55$ dex usualmente encontrados en asociaciones jóvenes (Bastian et al., 2010). Los autores señalan esta discrepancia y enfatizan que sus resultados podrían estar sesgados por un efecto observacional o físico desconocido, especialmente en el régimen de baja masa, lo que podría explicar el alto valor de σ . Como veremos la determinación de las IMFs en los NYMGs es una tarea demandante que constituye un aporte importante a la descripción del entorno solar.

1.1.3. Ventajas y desafíos del estudio de la IMF de los Grupos Jóvenes

Las edades jóvenes y cortas distancias hacen que los NYMG sean candidatos ideales para la determinación de la IMF. Respecto de su edad, porque sólo deberían haber perdido sus estrellas más masivas como consecuencia de la evolución estelar. En cuanto a la distancia, su cercanía permite observaciones con mayor SNR de sus miembros menos masivos comparado con poblaciones estelares más lejanas. Por lo tanto, la IMF observada debería ser producto de la IMF de la población original y de la posible dispersión cinemática de las estrellas

Existen algunas limitaciones observacionales y características intrínsecas de los NYMGs que desafían el estudio de su IMF pues afectan principalmente nuestra capacidad para identificar a los miembros de cada grupo: (i) sus cortas distancias al Sol hacen que las proyecciones de los NYMG cubran regiones extensas del cielo (ver Figura 2.6) y las diferencias porcentuales entre las distancias a estrellas de un mismo grupo pueden ser significativas; (ii) debido a que la mayoría de estos grupos son asociaciones dispersas con un pequeño número de miembros (ver Tabla 2.2), su distribución en el espacio de fases no siempre se diferencia de la distribución de las demás estrellas del entorno solar, lo cual significa que algunos NYMG son muy

difíciles de detectar cinemáticamente y *(iii)* si estos grupos realmente se están dispersando en el disco galáctico, algunos de sus miembros pueden haber perdido completamente su memoria cinemática, lo que hace imposible determinar su membresía. Este último efecto debería traducirse en una incompletitud de masa en nuestra medida de la IMF, lo que significa que si las incompletitudes observacionales están bien estimadas, la existencia de una incompletitud en la IMF medida sería indicativa de la dispersión.

1.2. Motivaciones y objetivos

Es sabido que la mayoría de las poblaciones estelares se dispersan rápidamente durante las primeras decenas o cientos de Maños de su evolución y que grupos densos como, por ejemplo, las Pleyades son excepcionales (e.g. [Lada & Lada, 2003](#)). Entonces, es razonable suponer que los NYMG representan remanentes de poblaciones estelares jóvenes que todavía se están dispersando en el disco galáctico y surgen naturalmente preguntas como las siguientes, que motivan el presente trabajo: ¿Son los NYMGs representantes del proceso por el cual las estrellas jóvenes nacidas en cúmulos se dispersan desde las regiones de formación estelar para poblar el disco galáctico? ¿Es la IMF de los NYMGs similar a la distribución de masas de las estrellas de campo del entorno solar y de grupos en regiones de formación estelar? ¿Hasta que punto la distribución de masas de un NYMG es representativa de la IMF de su población original ante la posible pérdida de memoria cinemática de algunos de sus miembros originales? En una era con sondeos astrométricos y fotométricos de todo el cielo, estas preguntas pueden investigarse con más detalle que antes.

Por otro lado, la revisión de la literatura muestra indicios claros pero no cuantificados de incompletitudes en los catálogos disponibles de miembros de los NYMGs ([Rafael Bertolotto, 2021](#), trabajo de licenciatura). Todo lo anterior justifica firmemente la necesidad de emprender un estudio estadístico robusto de la IMF en los NYMGs además de un estudio de su evolución dinámica y cinemática. Un estudio robusto de la IMF que de cuenta explícita de la completitud y pureza de las muestras en un conjunto de NYMGs es el objetivo general del presente trabajo.

Dicho estudio es crucial para entender la evolución de la función de masa medida en cúmulos estelares jóvenes debida a los procesos dinámicos y cinemáticos de dispersión. La importancia de tal estudio es aún mayor considerando el hecho de que hasta la fecha solamente se ha inferido la IMF de un único NYMG ([Gagné et al., 2017](#)). En este trabajo, desarrollamos y aplicamos los primeros pasos de dicho estudio, con

los siguientes objetivos específicos:

1. Desarrollar e implementar un método de detección de NYMG basado en la detección de sobredensidades en espacios cinemáticos y fotométricos para mejorar los catálogos disponibles.
2. Cuantificar la completitud y pureza de las muestras resultantes.
3. Calcular las IMF de un conjunto representativo de los NYMGs detectados.
4. Comparar las IMF obtenidas entre sí y con la IMF representativa de la población de campo del entorno solar y grupos estelares jóvenes en regiones de formación estelar.
5. Explorar si las posibles incompletitudes de las IMFs respecto de un modelo universal de IMF podrían ser explicadas por un proceso de dispersión.

Capítulo 2

Construcción de la muestra de estudio

2.1. Hacia un muestreo completo del Entorno Solar con Gaia DR3

Realizamos nuestra búsqueda de nuevos miembros de los NYMG utilizando el catálogo *Gaia Data Release 3*; (Gaia Collaboration et al., 2018) (GDR3), que ofrece la mejor combinación hasta la fecha de completitud espacial y fotométrica para el vecindario solar y además proporciona datos fotométricos y astrométricos muy precisos. Una caracterización completa del sondeo se puede encontrar en Gaia Collaboration et al. (2022). Como ejemplo, objetos de brillo $G \sim 18 \text{ mag}$ poseen incertidumbres promedio del $\sim 10\%$ en paralaje, $\sim 0,1\%$ en brillo y $\sim 15\%$ en movimientos propios. Finalmente, como muestra de validación, compilamos una lista de miembros espectroscópicos de los NYMG y candidatos cinemáticos con membresía altamente probable, principalmente del catálogo *Montreal Open Clusters and Associations* (MOCA) compilado por Gagné & Faherty (2018), al que complementamos con resultados adicionales de la literatura, como explicamos en las siguientes secciones.

2.1.1. Selección del Entorno Solar en Gaia DR3

Nuestra muestra inicial incluye a las 5.244.458 fuentes de GDR3 con paralajes $\varpi \geq 5 \text{ mas}$, lo que corresponde a distancias $d \leq 200 \text{ pc}$ si asumimos $d = 1/\varpi$. Luego, aplicamos la corrección de punto cero para

ϖ proporcionada por [Lindegren et al. \(2021\)](#) e implementada en el paquete de Python *gaiadr3_zeropoint*.

Nuestro procedimiento de detección requiere dos muestras: la primera, que llamamos Subconjunto de los NYMG (S_{NYMG}), es la muestra sobre la que realizaremos la detección de NYMGs en el Capítulo 3 y la segunda, que llamaremos Subconjunto Auxiliar (S_{aux}), se utilizará como muestra auxiliar para determinar el nivel de completitud de S_{NYMG} según el procedimiento que explicaremos en la Sección 2.1.3. A continuación describimos como se realizó la selección de ambas muestras y la Tabla 2.1. indica los números de fuentes en cada una.

La muestra S_{aux}

La muestra S_{aux} contiene un total de 594.297 fuentes obtenidas aplicando las siguientes máscaras a la muestra inicial:

1. Descartamos fuentes con errores fraccionales de paralaje $\sigma_{\varpi}/\varpi > 0,1$. Esta máscara permite conservar únicamente objetos con paralajes suficientemente buenos para las distancias estimadas como $1/\varpi$ sean consistentes con las distancias foto-geométricas de [Bailer-Jones et al. \(2021\)](#) como se muestra en la sección 2.1.4.
2. Descartamos todas las fuentes cuyas posiciones en el Diagrama de Color-Magnitud Absoluta (DCMA) M_G vs. $BP - RP$ no son consistentes con el locus de la Secuencia Principal (SP) y Pre-Secuencia Principal (PSP) y que por tanto no pueden ser parte de los NYMG. Por ejemplo, dentro de los objetos descartados se encuentran enanas blancas. Las estrellas en las ramas de las sub-gigantes o gigantes que pudieran ocupar posiciones similares a las de estrellas de la PSP forman parte de la contaminación que trataremos posteriormente.
3. Descartamos todas las fuentes cuya Relación Señal a Ruido (SNR) cumple $\log_{10}(F_G/\sigma_{F_G}) < 2,2$ donde F_G es el flujo en la banda G y σ_{F_G} es el error asociado. Como se muestra en la Figura 2.1, la mayoría de estas fuentes poseen valores de $M_G > 12$ lo cual corresponde principalmente a fuentes que se distribuyen en un locus no estelar del DCMA como se muestra en la Figura 2.2. Además, la mayoría de estas fuentes presentan valores de movimientos propios muy cercanos a cero, lo que sugiere problemas en las medidas astrométricas y/o que se ubican a distancias mucho mayores a las del entorno solar.

4. Descartamos fuentes con velocidades tangenciales en el *Local Standard of Rest* (LSR) con norma $[(v_{\alpha}^{LSR})^2 + (v_{\delta}^{LSR})^2]^{1/2} > 20 \text{ km.s}^{-1}$ para mantener solamente fuentes cuya cinemática es consistente con la de los NYMG conocidos (ver Sección 2.2.2). Esta máscara recorta sensiblemente los tiempos de cómputo sin descartar objetos de interés.

La muestra S_{NYMG}

La muestra S_{NYMG} contiene un total de 495.079 fuentes y resulta de aplicar las siguientes máscaras a S_{aux} :

1. $\log_{10}(\text{phot_bp_rp_excess_factor}) \leq 0,001 + 0,039(BP - RP)$ o $0,012 + 0,039(BP - RP) \leq \log_{10}(\text{phot_bp_rp_excess_factor})$ sugerido por [Gaia Collaboration et al. \(2021\)](#) para descartar fuentes con fotometría inconsistente debido a la contaminación de fuentes cercanas y otros efectos.
2. Utilizando el *Re-normalized Unit Weighted Error* (RUWE), aplicamos la máscara $\text{RUWE} \geq 1,4$ como recomienda [Fabricius et al. \(2021\)](#). Esta máscara ayuda a descartar fuentes con mala astrometría, a menudo causada por multiplicidad, lo cual afecta sistemas estelares dobles o múltiples tanto de la SP como PSP.

La Figura 2.2 muestra la distribución en el DCMA de las fuentes de la muestra de partida, la muestra S_{aux} y la muestra S_{NYMG} . En la siguiente sección estimamos la fracción de miembros eliminados por estas máscaras.

2.1.2. Método para estimar la completitud mediante la Función de Selección

Las máscaras empleadas son necesarias para detectar a los NYMG porque eliminan un gran número de contaminantes y detecciones espurias. Sin embargo, es posible que las máscaras empleadas para construir S_{NYMG} excluyan algunas estrellas de la SP y la PSP del entorno solar. Si esto ocurriera se podrían estar perdiendo miembros de las asociaciones lo cual produciría incompletitudes en la IMF de los NYMG. Por esta razón, proponemos inferir el nivel de completitud de S_{NYMG} para luego completar la muestra con objetos sintéticos.

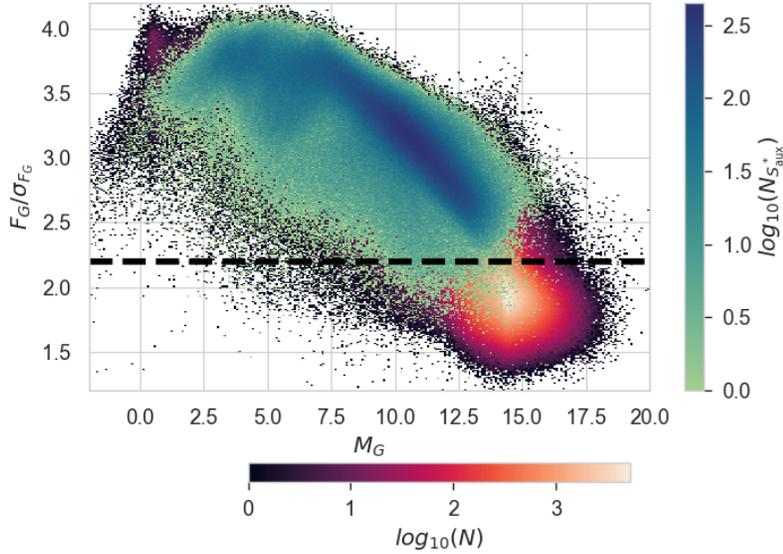


Figura 2.1: SNR del flujo en G vs. M_G de las estrellas de la muestra de partida y de S_{aux} sin la máscara (S_{aux}^*) en la SNR. La línea a trazos indica el valor de $\log_{10}(F_G/\sigma_{F_G}) = 2,2$, el cual ayuda a filtrar la muestra inicial.

A continuación desarrollamos un método para inferir la completitud de una muestra arbitraria de GDR3 haciendo uso de la *Gaiainlimited Selection Functions* (GSF). La GSF nos proporciona las herramientas para calcular la función de selección general de GDR3 (Cantat-Gaudin et al., 2023), que corresponde a un mapa continuo $\mathcal{F}(q)$ en un espacio de cuatro dimensiones $q = (\alpha, \delta, G - RP, G)$ con (α, δ) las coordenadas ecuatoriales y G y RP magnitudes en el sistema de Gaia DR3, que devuelve para un q dado la probabilidad $\mathcal{F}(q) = P_q(\text{GDR3})$ de que una fuente con $q \in \Delta q = (q - dq, q + dq)$ haya sido incluida en GDR3. Como explican Rix et al. (2021), para cada q , $\mathcal{F}(q)$ es una probabilidad adimensional cuyo valor está entre 0 y 1, por lo que $\mathcal{F}(q)$ no es una distribución de probabilidad sobre todo q . La GSF también nos proporciona las herramientas para estimar la función de selección $\mathcal{F}_C(q) = P_q(C|\text{GDR3})$ de una submuestra C definida por alguna condición arbitraria dentro del catálogo GDR3, es decir, la probabilidad de que una fuente con $q \in (q - dq, q + dq)$ pertenezca a C sabiendo que está en GDR3.

Supongamos que estamos interesados en estudiar un conjunto específico de estrellas que llamaremos la Muestra de Interés Verdadera (MIV). En nuestro caso particular, la MIV correspondería a las estrellas reales de la SP y PSP del vecindario solar con cinemática similar a la de los NYMG conocidos. De manera general, para identificar la MIV dentro de algún catálogo, se emplean máscaras como lo hemos hecho

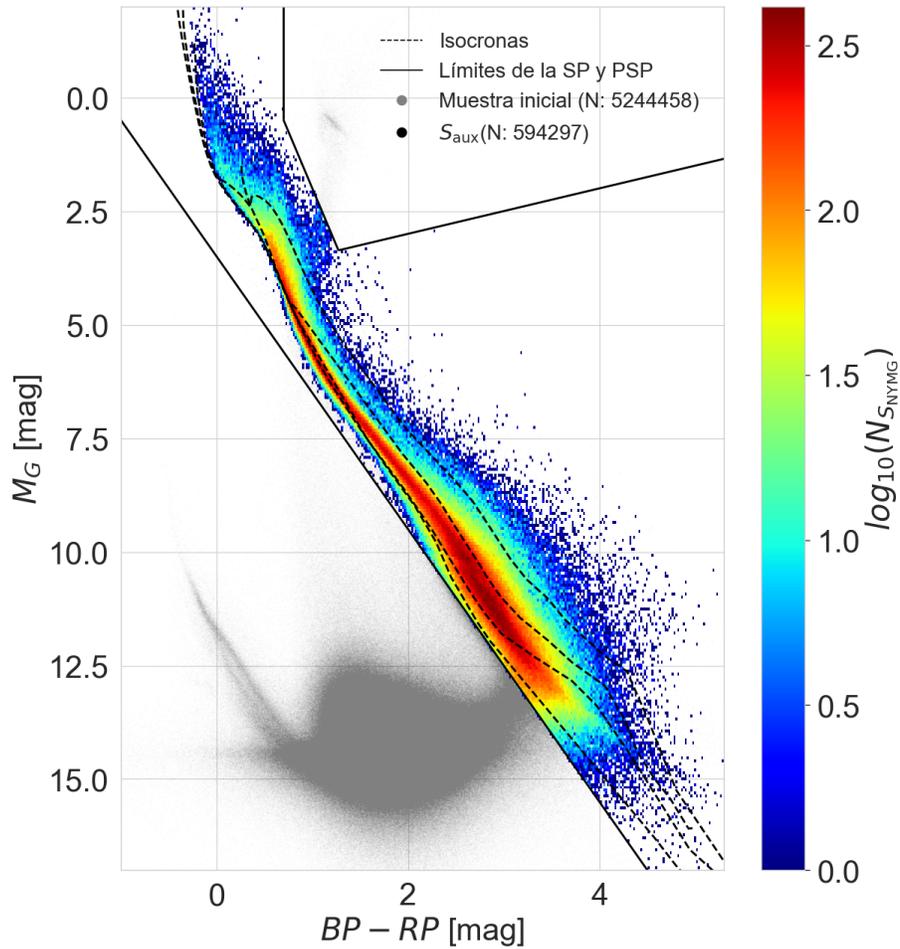


Figura 2.2: DCMA de la muestra inicial (puntos grises) y la muestra S_{NYMG} (escala de colores que indica la densidad de puntos). Las líneas sólidas indican los límites del locus de la MS y PMS y las líneas discontinuas las isocronas de los modelos estelares de Phillips et al. (2020), Baraffe et al. (2015) y Marigo et al. (2017) para edades de 10 Maños, 30 Maños, 100 Maños y 1 Gaño.

para identificar la muestra S_{NYMG} de GDR3. Estas máscaras se pueden clasificar en tres categorías: (i) las *máscaras de selección* que definen la MIV dentro del catálogo de partida (en nuestro caso GDR3), (ii) las *máscaras de calidad* que no definen la muestra de interés, sino que aseguran un cierto grado de calidad en los datos, y (iii) las *máscaras mixtas* que aseguran un cierto nivel de calidad y a la vez participan en la selección de la muestra. En nuestro caso específico, las máscaras $\varpi \geq 5$ y $\text{RUWE} < 1,4$ son ejemplos de máscaras de selección y máscaras de calidad, respectivamente. Una máscara mixta es, por ejemplo, la máscara $\sigma_{\varpi}/\varpi < 0,1$ ya que esta máscara, que se usa con el fin de mejorar la calidad de la muestra del entorno solar, también pudiera descartar fuentes fuera del vecindario solar cuyos paralajes no fueron bien medidos.

Asumiendo que no se aplican ninguna máscaras mixtas, definimos la Muestra de Interés Observada (MIO) como la muestra resultante de aplicar únicamente las máscaras de selección al catálogo de partida y la Muestra de Interés Estudiada (MIE) la muestra resultante de aplicar todas las máscaras. Con estas definiciones, el nivel de completitud en un bin Δq corresponde a la fracción $f_{\Delta q} = n_{\Delta q}^{\text{MIE}}/n_{\Delta q}^{\text{MIV}}$ donde $n_{\Delta q}^{\text{MIE}}$ y $n_{\Delta q}^{\text{MIV}}$ corresponden al número de objetos del MIE y del MIV en el bin Δq respectivamente. La completitud $f_{\Delta q}$ se puede estimar en términos de probabilidades como $f_{\Delta q} \sim P_q(\text{MIE}|\text{MIV})$, que es la probabilidad de que una fuente dentro de Δq pertenezca a MIE sabiendo que pertenece a MIV. Usando la regla de Bayes obtenemos:

$$f_{\Delta q} \sim P_q(\text{MIE}|\text{MIV}) = \frac{P_q(\text{MIV}|\text{MIE})P_q(\text{MIE})}{P_q(\text{MIV})} \quad (2.1.1)$$

Si la cantidad de contaminantes en MIE es insignificante, entonces casi todas las fuentes en $\text{MIV} \cap \text{MIE}$ deberían pertenecer a la MIV, lo que significa que la probabilidad $P_q(\text{MIV}|\text{MIE}) \sim 1$. En cuanto al prior de la Ecuación 2.1.1, es claro que todas las fuentes del MIE pertenecen a la muestra de GDR3, por lo tanto $P_q(\text{MIE}) = P_q(\text{MIE} \cap \text{GDR3}) = P_q(\text{MIE}|\text{GDR3})P_q(\text{GDR3})$. Finalmente, podemos estimar la probabilidad de normalización $P_q(\text{MIV})$ como $n_{\Delta q}^{\text{MIV}}/n_{\Delta q}$ donde $n_{\Delta q}$ es el número real de fuentes del Universo dentro de Δq . Luego, notamos que $n_{\Delta q}^{\text{MIV}}/n_{\Delta q} \sim n_{\Delta q}^{\text{MIO}}/n_{\Delta q}^{\text{GDR3}}$ donde $n_{\Delta q}^{\text{MIO}}$ y $n_{\Delta q}^{\text{GDR3}}$ son el número de fuentes en Δq que pertenecen a las muestras MIO y GDR3 respectivamente. Esta última relación se puede estimar como la probabilidad $P_q(\text{MIO}|\text{GDR3})$. Así, podemos re-escribir la Ecuación 2.1.1 como:

$$f_{\Delta q} \sim P_q(\text{MIE}|\text{MIV}) \approx \frac{P_q(\text{MIE}|\text{GDR3})}{P_q(\text{MIO}|\text{GDR3})} P_q(\text{GDR3}) \quad (2.1.2)$$

De esta manera, todos los términos en el lado derecho de la Ecuación 2.1.2 pueden ser calculados usando el GSF. En caso de que algunas de las máscaras usadas para seleccionar el MIE sean Máscaras Mixtas (MM), no podemos repetir el mismo razonamiento ya que dejaría de cumplirse la condición $n_{\Delta q}^{\text{MIV}}/n_{\Delta q} \sim n_{\Delta q}^{\text{MIO}}/n_{\Delta q}^{\text{GDR3}}$. Esto se debe a que no podemos saber qué fracción de las fuentes descartadas por una MM del MIO fueron incluidas erróneamente en dicho conjunto. Sin embargo, si podemos usar todas las MM junto con las máscaras de selección para seleccionar el MIO y luego usar la Ecuación 2.1.2. Con este cambio, estamos limitados a estimar la completitud de la muestra resultante de las máscaras de calidad. Este razonamiento puede ser generalizado para cualquier muestra de cualquier catálogo si conocemos las funciones de selección de dicho catálogo.

2.1.3. Completando la muestra de estudio

Haciendo uso del método desarrollado en la Sección 2.1.2, podemos estimar el nivel de completitud $f_{\Delta q}$ mediante la Ecuación 2.1.2. Para esto, notamos que en nuestro caso particular los conjuntos MIE y MIO corresponden a las muestra S_{NYMG} y S_{aux} . Las MM de S_{aux} corresponden a las máscaras $\log_{10}(F_G/\sigma_{F_G}) \geq 2,2$ y $\sigma_{\varpi}/\varpi \leq 0,1$. Luego, definimos el conjunto *real* de estrellas de la SP y PSP del Vecindario Solar como el conjunto S_{NYMG} Real (S_{real}), equivalente al conjunto MIV. Así, la Ecuación 2.1.2 se transforma en:

$$f_{\Delta q} \sim P_q(S_{\text{NYMG}}|S_{\text{real}}) \approx \frac{P_q(S_{\text{NYMG}}|\text{GDR3})}{P_q(S_{\text{aux}}|\text{GDR3})} P_q(\text{GDR3}) \quad (2.1.3)$$

Para completar la muestra S_{NYMG} , primero la dividimos en bins Δq utilizando un bin de 1 mag y 0.1 mag respectivamente para G y $G - RP$ y un nivel de healpixel¹ de 4 en (α, δ) que equivale a la división de la esfera en 192 healpixels. Luego, para cada Δq , calculamos $f_{\Delta q}$ a través de la Ecuación 2.1.3

¹HEALPix es el acrónimo en inglés de Hierarchical Equal Area isoLatitude Pixelization (pixelización jerárquica de igual área a latitud constante) de una esfera. Es la pixelización que subdivide la superficie de una esfera de manera que cada píxel cubre la misma superficie que cualquier otro píxel.

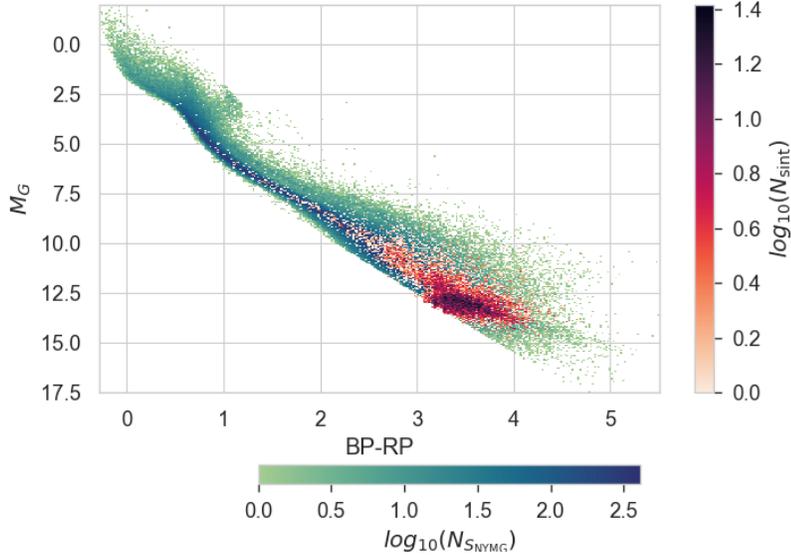


Figura 2.3: DCMA de las fuentes de S_{NYMG} y de las estrellas sintéticas (colores calientes) de uno de los diez S_{comp} (colores fríos).

y muestreamos $N_{\Delta q}/f_{\Delta q} - N_{\Delta q}$ estrellas de la distribución de estrellas de la muestra S_{NYMG} con $q \in \Delta q$ donde $N_{\Delta q}$ corresponde a la cantidad de estrellas de S_{NYMG} en Δq . El catálogo resultante corresponde a la muestra S_{NYMG} más 14329 fuentes sintéticas y lo llamamos muestra S_{NYMG} Completado (S_{comp}). Para lograr resultados más robustos estadísticamente, construimos diez S_{comp} . La Figura 2.3 muestra la distribución en el DCMA de fuentes sintéticas y reales de uno de los diez S_{comp} . Notamos que la mayoría de las estrellas sintéticas tienen valores altos de M_G , lo que muestra que el nivel de completitud disminuye a medida que G aumenta, como era de esperar.

2.1.4. Sobre la calidad de los datos

Estimación de distancias mediante $d = 1/\varpi$

Estimar distancias usando $d = 1/\varpi$ es el método más simple pero menos preciso (Bailer-Jones, 2015). Por otro lado, las distancias foto-geométricas d_{pg} estimadas por Bailer-Jones et al. (2021) proporcionan una determinación más robusta para el catálogo completo de GDR3. Ambas estimaciones de distancia muestran un buen acuerdo cuando $\sigma_\varpi/\varpi < 0,1$, donde σ_ϖ es la incertidumbre en ϖ . Este resultado surge de la

Muestra	$N_{d_{\varpi}}$	$N_{d_{pg}}$	N_{match}
Muestra inicial	5.244.458	2.328.436	2.320.981
S_{aux}	594.297	577.003	568.703
S_{NYMG}	495.079	482.417	475.854

Tabla 2.1: Número de estrellas en cada muestra usando las distancias d_{ϖ} y d_{pg} , y número de estrellas con ambas distancias.

comparación de dos muestras: la muestra inicial que definimos en la Sección 2.1.1 a partir de $d_{\varpi} = 1/\varpi < 200$ pc) y una muestra similar que llamaremos $GDR3_{d_{pg}}$ definida a partir de $d_{pg} < 200$ pc.

Aplicamos los mismos procedimientos utilizados para seleccionar las muestras S_{aux} y S_{NYMG} a partir de la *muestra inicial* (Sección 2.1.1) a la muestra $GDR3_{d_{pg}}$, y generamos las muestras $S_{aux_{pg}}$ y $S_{NYMG_{pg}}$. El número de fuentes para cada muestra se indica en la Tabla 2.1. La diferencia en las muestras con diferentes estimaciones de distancia es menor al 5% y ocurre para estrellas con $G > 17$. La fracción de diferencias aumenta con G a medida que la SNR disminuye, como se muestra en la Figura 2.4. Concluimos que, en nuestro caso, es equivalente estimar distancias con cualquiera de los dos métodos. La estimación de la distancia como el inverso del paralaje será una ventaja práctica en el capítulo 3.

Pérdida despreciable de miembros debido a las máscaras en distancia y SNR

Debemos estimar cuántas estrellas que pertenecen al vecindario solar, pertenecen a la SP o PSP y poseen cinemática consistente con los NYMG, están siendo descartadas por las máscaras $\log_{10}(\text{SNR}) > 2,2$ y $\sigma_{\varpi}/\varpi < 0,1$. Para ello consideramos una nueva muestra, que corresponde a S_{aux} y que contiene todos los objetos que poseen distancias $1/\varpi$ y d_{pg} pero a los que no se les aplicó las máscaras en SNR y σ_{ϖ}/ϖ . Así, si $N_{match,aux}^*$ corresponde al número de fuentes de esa muestra y $N_{match,aux}$ es el número de fuentes de esa muestra pero luego de aplicar las máscaras de SNR y σ_{ϖ}/ϖ , entonces podemos estimar la completitud de la muestra después de aplicar ambas máscaras como $N_{match,aux}/N_{match,aux}^*$. Encontramos que esta razón es igual a 0,987, lo que significa que estamos descartando solamente el 1,3% de las fuentes de interés. Si bien esta fracción es pequeña, como la mayoría de las fuentes que se pierden de esta máscara corresponden a las estrellas menos brillantes, esto significa que pudiéramos estar descartando algunos objetos miembros de NYMGs de baja masa.

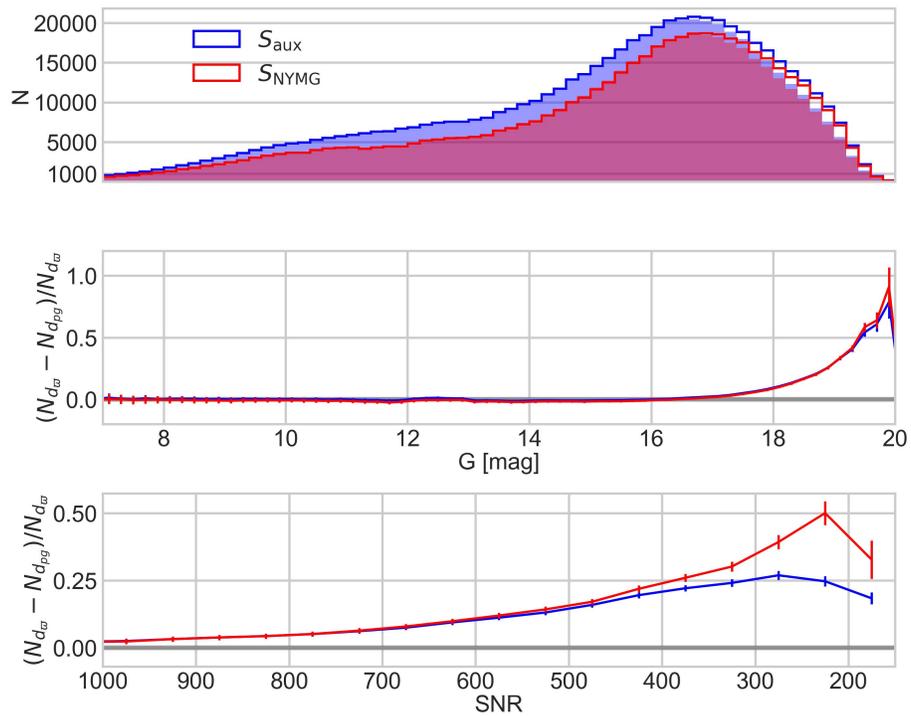


Figura 2.4: **Panel superior:** Distribución de magnitudes G para la muestra de estrellas con distancias $d = 1/\varpi < 200$ pc (histogramas de línea sólida) y distancias $d_{pg} < 200$ pc de [Bailer-Jones et al. \(2021\)](#) (histogramas sombreados). Los histogramas azules representan las muestras S_{aux} y $S_{aux_{pg}}$ y los histogramas rojos las muestras S_{NYMG} y $S_{NYMG_{pg}}$ descritas en la Sección 2.1.1. **Panel intermedio:** Residuos entre el número de fuentes en S_{NYMG} y $S_{NYMG_{pg}}$ (rojo) y entre S_{aux} y $S_{aux_{pg}}$ (azul) como función de G . **Panel inferior:** Mismos residuos como función de SNR. Las barras de error corresponden al error de Poisson propagado.

Efecto de los errores

La Figura 2.5 muestra los errores cinemáticos en $(X, Y, Z, v_{\alpha, \text{LSR}}, v_{\delta, \text{LSR}})$ propagados a partir de $(\alpha, \delta, \mu_{\alpha}, \mu_{\delta}, \varpi)$. Para las velocidades cartesianas, se consideraron los errores en la velocidad del Sol en el LSR reportadas por Schönrich et al. (2010) $(\Delta U_{\odot}, \Delta V_{\odot}, \Delta W_{\odot}) = (0,72, 0,47, 0,37) \text{ km s}^{-1}$. Se utilizó el paquete Astropy (Astropy Collaboration et al., 2022) para calcular la cinemática en el LSR el cual se basa en los resultados de Schönrich et al. (2010).

Observamos que los errores en posición son menores a 1 pc para objetos con $6 < G/mag < 16$, aumentan hasta 4 pc para objetos brillantes ($G < 6 \text{ mag}$) y aumentan hasta 14 pc para objetos débiles ($G > 16 \text{ mag}$). Un comportamiento análogo se observa para los errores en velocidad. Este comportamiento de los errores debe ser consecuencia de una baja SNR para objetos poco brillantes y de *crowding* o *blending* para estrellas muy brillantes. Los errores propagados de Schönrich et al. (2010) se traducen en errores sistemáticos en las velocidades tangenciales de $\Delta v_{\alpha, \text{LSR}} = 0,7 \text{ km s}^{-1}$ y $\Delta v_{\delta, \text{LSR}} = 0,4 \text{ km s}^{-1}$. Los resultados se muestran en la Figura 2.5 de donde concluimos que los errores cinemáticos son despreciables para la mayoría de las fuentes de S_{NYMG} pero son significativos para los objetos más tenues ($G > 18 \text{ mag}$) que corresponden a un $\sim 15 \%$ de la muestra total. Este comportamiento de los errores será relevante en el análisis de la componente menos masiva de los NYMGs estudiados.

2.2. Los NYMGs conocidos del entorno solar

2.2.1. La base de datos MOCA y el catálogo maestro

Diremos que una fuente puntual es candidato a miembro de un NYMG si sus propiedades fotométricas y cinemáticas son consistentes con la de los otros miembros del NYMG.

Como muestra de control empleamos el catálogo maestro que incluye al catálogo MOCA que compila la información disponible sobre los grupos (NYMGs y otros tipos de poblaciones) detectados y caracterizados por Gagné & Faherty (2018) e información de numerosos trabajos previos. Además, hicimos la recopilación de candidatos y miembros confirmados de NYMGs de los siguientes trabajos: Makarov & Urban (2000); Zuckerman & Song (2004); Torres et al. (2006); Viana Almeida et al. (2009); Cruz et al.

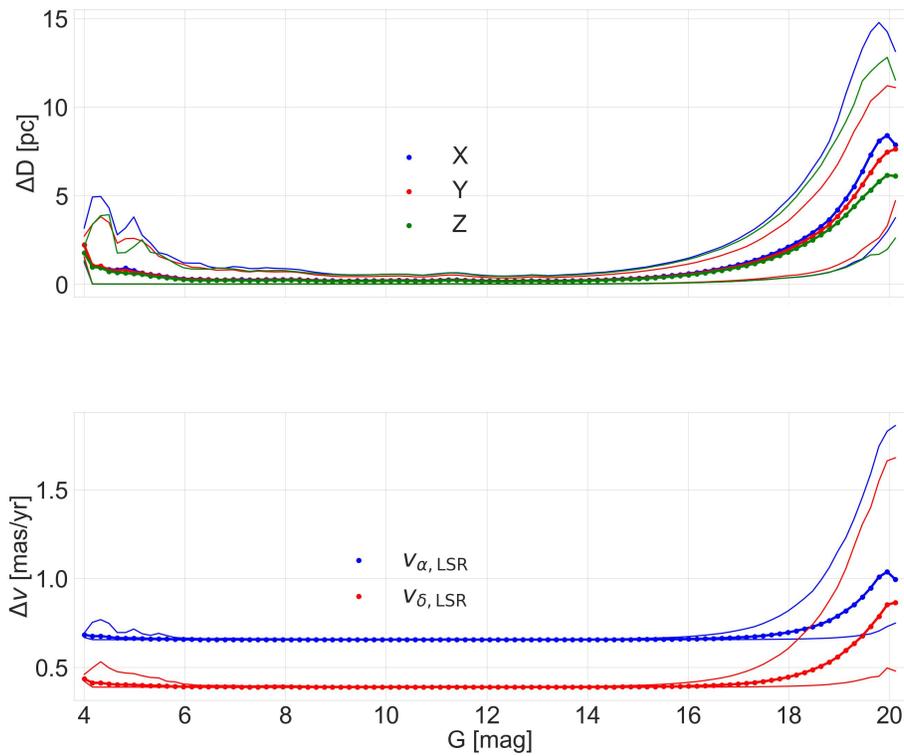


Figura 2.5: Errores cinemáticos propagados para las posiciones galácticas (panel superior) y las velocidades tangenciales del LSR (panel inferior) como función de G para la muestra S_{NYMG} . Los puntos conectados muestran los valores medios para cada bin de G de tamaño 0,25 mag y las líneas delgadas muestran los bordes del percentil 90 % para cada bin de G .

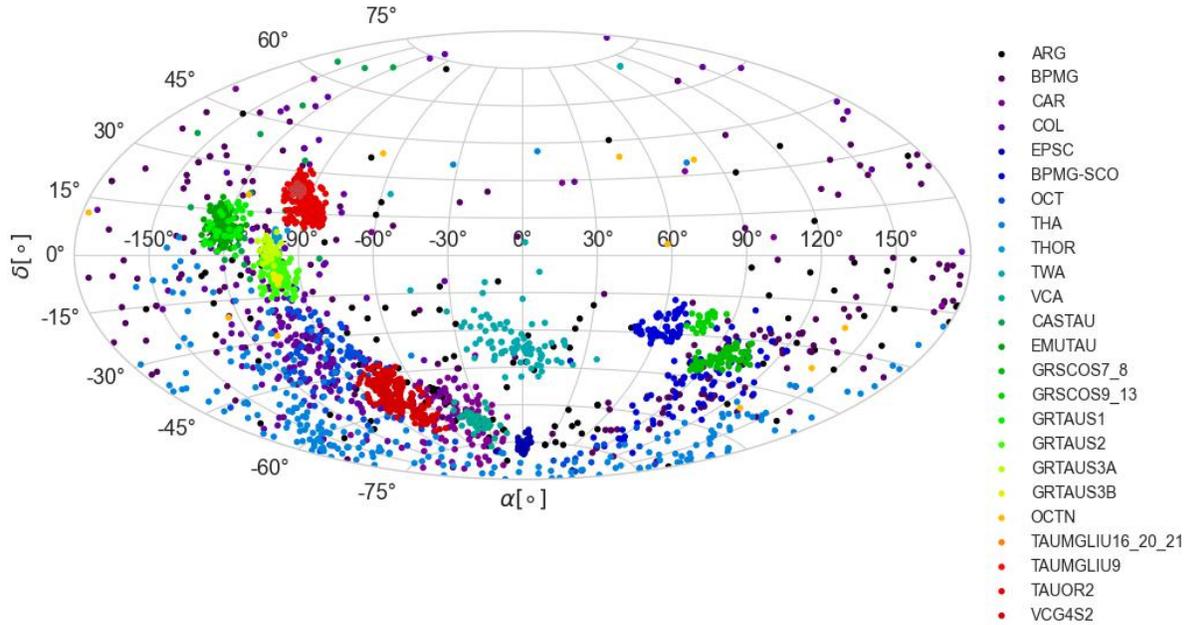


Figura 2.6: Proyección de Aitoff de la distribución espacial de los NYMG de S_{lit} en coordenadas ecuatoriales.

(2009); da Silva et al. (2009); Galvez-Ortiz et al. (2010); Shkolnik et al. (2012); De Silva et al. (2013); Dawson et al. (2013); Moor et al. (2013); Malo et al. (2013, 2014); Malo et al. (2014); Riedel et al. (2014); Elliott et al. (2014); Kraus et al. (2014); Ducourant et al. (2014); Galvez-Ortiz et al. (2014); Murphy & Lawson (2015); Gagné et al. (2015). Nos referiremos a la muestra resultante de miembros reportados en la literatura como Muestra de Candidatos de la Literatura (S_{lit}), cuyas estadísticas generales se resumen en la Tabla 2.2. La distribución en el cielo de los NYMGs conocidos se muestra en la Figura 2.6. En cuanto a las fuentes provenientes de MOCA, solamente consideramos aquellas con probabilidades de membresías mayores a 0,9. Muchas de las estrellas de los NYMG en MOCA provienen del proyecto BANYAN (i.e. Gagné et al., 2021) que realiza el cálculo de probabilidades en base a la distribución en el espacio de fase (X, Y, Z, U, V, W) empleando una velocidad radial que no es medida sino inferida estadísticamente.

2.2.2. La cinemática de los NYMGs

Dado que las dimensiones espaciales de muchos NYMG son del mismo orden de magnitud e incluso a veces más grandes que sus distancias al Sol, se espera que sus distribuciones de movimientos propios

NYMG	Edad [Maños]	N	N_{VR}	N_{lit}	N_{MOCA}
ARG	45	146	109	69	38
BPMG	26	221	151	29	55
CAR	45	132	95	51	20
CASTAU	50	30	10	13	30
COL	42	191	112	54	46
EPSC	4	37	21	0	12
GRTAUS1	52	36	16	9	36
GRTAUS2	31	116	52	84	116
GRTAUS3A	19	37	15	24	37
GRTAUS3B	14	12	6	10	12
GRTAUS8	5	10	4	0	10
OCT	35	111	96	1	53
OCTN	55	12	11	6	12
TAUMGLIU9	8	25	10	0	25
TAUOR2	42	234	81	220	234
THA	51	236	126	31	62
THOR	25	13	1	0	3
TW-Hya	10	89	42	11	30
VCA	89	54	28	0	1
VCG4S2	38	191	63	3	191
118TAU	10	20	10	0	20
BPMG-SCO	25	141	92	36	123
GRSCOS7_8	14	72	31	25	72
GRSCOS9_13	21	30	16	10	30
TAUMGLIU16_20_21	38	60	23	53	60
EMUTAU	56	145	39	1	145

Tabla 2.2: Información general de los NYMGs conocidos de la muestra S_{lit} . Edad: edades reportadas por MOCA; N : número total de candidatos conocidos; N_{VR} : número de candidatos con velocidad radial de GDR3, N_{lit} : número de candidatos de la literatura que no se encontraron en MOCA, N_{MOCA} : número de candidatos que provienen únicamente de MOCA.

sufren de dos fenómenos: (i) los NYMG pueden cubrir regiones muy grandes del cielo, lo que significa que incluso si todos los miembros de un grupo comparten el mismo vector de velocidad, sus movimientos propios pueden ser muy diferentes, y (ii) dos miembros de una misma asociación pueden estar a distancias muy diferentes del Sol, lo que significa que incluso si tienen la misma posición en el cielo y el mismo vector de velocidad cartesiana, sus vectores de movimiento propio pueden ser diferentes. Aunque sin la Velocidad Radial (VR) no es posible corregir el sesgo introducido por el primer efecto, el sesgo producido por el segundo sí se puede corregir calculando la velocidad tangencial v_T de cada fuente a partir del paralaje ϖ y el movimiento propio μ mediante:

$$v_T = \frac{1}{\varpi} \mu \quad (2.2.1)$$

Además de los efectos anteriores, la distribución de movimientos propios de los NYMG también puede ser transformada por el movimiento del Sol tanto en el espacio de movimientos propios como en el espacio de velocidades tangenciales. Por esta razón, trabajaremos en el LSR.

La Figura 2.7 muestra la distribución de los NYMGs del S_{lit} tanto en el espacio de movimientos propios como en el de velocidades tangenciales y tanto en el marco heliocéntrico como en el LSR. Observamos que la cinemática en el marco heliocéntrico está mucho más dispersa que en el LSR y que la distribución de los NYMG cambia radicalmente al pasar de movimientos propios a velocidades tangenciales. Sin embargo, al usar la estimación de la velocidad del Sol de Schönrich et al. (2010) para calcular la cinemática en el LSR, los errores son sistemáticamente mayores en el LSR que en el marco heliocéntrico. Esto se debe a que el error en la estimación de la velocidad cartesiana del Sol (Schönrich et al., 2010) se propaga a las velocidades tangenciales calculadas en el LSR como se mostró en la Figura 2.5.

Así, justificamos la búsqueda de NYMGs en el espacio 5-dimensional de posiciones cartesianas (X, Y, Z) y velocidades tangenciales $(v_{\alpha, \text{LSR}}, v_{\delta, \text{LSR}})$ en el LSR de la muestra S_{comp} donde la calidad de la fotometría y astrometría es suficientemente buena. Adicionalmente, hemos presentado el catálogo S_{lit} que nos permite comparar nuestros resultados con la literatura.

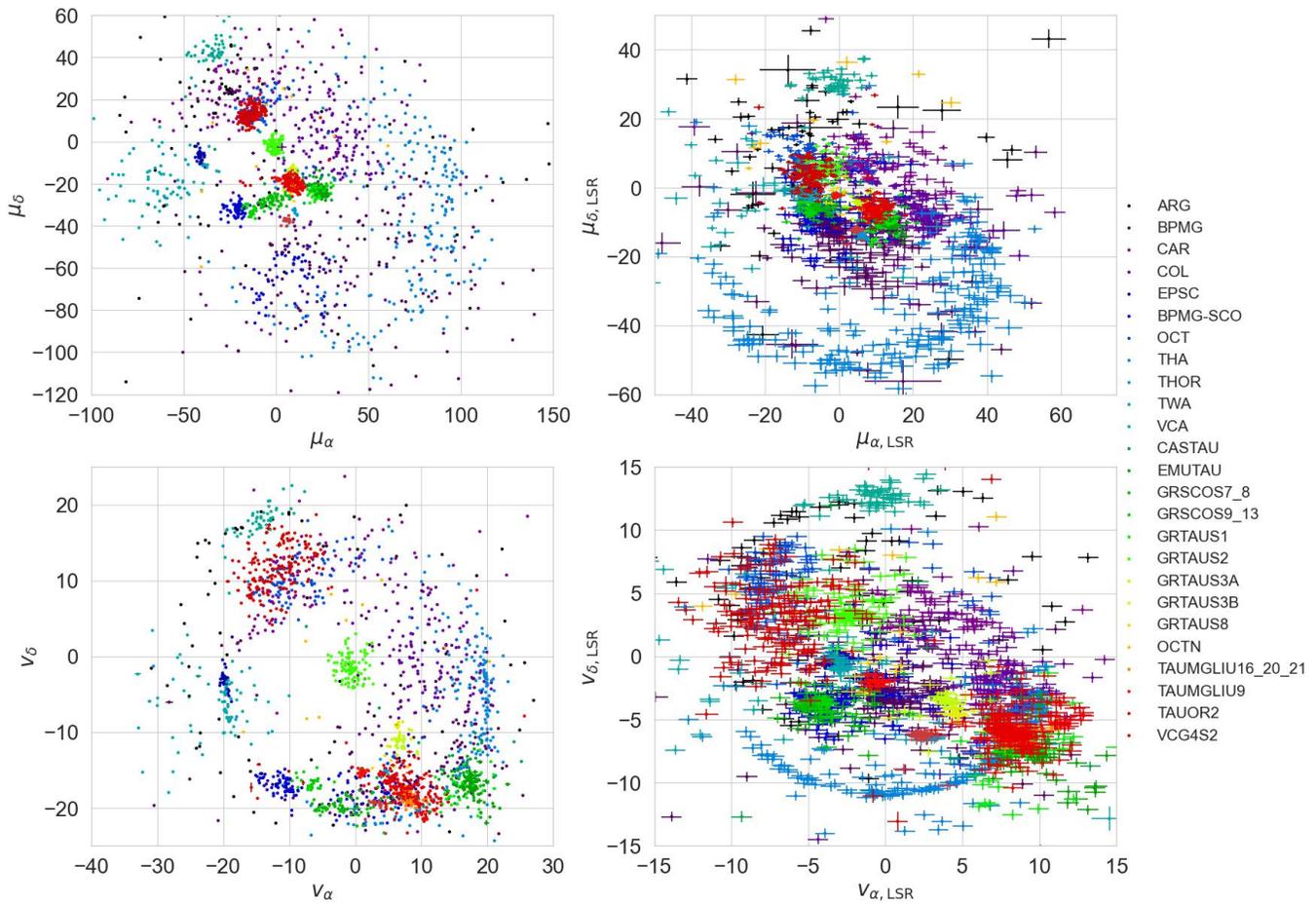


Figura 2.7: Distribuciones del movimiento propio (panel superior) y las velocidades tangenciales (panel inferior) de los NYMGs incluidos en la muestra S_{lit} . Panel de la izquierda: coordenadas galácticas heliocéntricas. Panel de la derecha: sistema ecuatorial LSR.

Capítulo 3

Detección de Grupos

En este capítulo describimos el método que desarrollamos e implementamos para la detección de NYMG y los resultados de los grupos detectados a partir de este método y estudiados en este trabajo. Los métodos y códigos fueron desarrollados exprofeso para esta tesis salvo que se indique lo contrario. Los códigos se encuentran disponibles en <https://github.com/rafaelbertolotto/searchForNYMGs>. Parte de su ejecución es computacionalmente demandante, supera las capacidades de los computadores personales actuales y se realizó empleando el servicio de cómputo científico de *ClusterUy* (Nesmachnow & Iturriaga, 2019).

3.1. El método de detección

3.1.1. El algoritmo DBSCAN

El método y códigos correspondientes que empleamos en este trabajo para la detección de NYMGs se basan en el algoritmo conocido como *Density-Based Spatial Clustering of Applications with Noise* de Ester et al. (1996) (DBSCAN), que es un algoritmo de clustering diseñado para detectar sobre-densidades de puntos en un espacio N-dimensional. El DBSCAN opera con dos hiperparámetros: la distancia máxima entre dos puntos para considerarlos vecinos (ϵ) y el número mínimo de vecinos necesarios para que un punto

sea etiquetado como punto núcleo (N_{\min}). Luego, el DBSCAN identifica como grupos a los conjuntos de puntos núcleos que son vecinos y a sus vecinos asociados. Una vez elegidos ε y N_{\min} , los grupos detectados por DBSCAN serán más densos que un cierto umbral $\rho_{\varepsilon, N_{\min}} = N_{\min}/V_{\varepsilon}$, donde V_{ε} es el volumen de una hipersfera de dimensión igual al número de dimensiones del espacio en las que se está usando DBSCAN y depende del parámetro ε . Esto convierte a DBSCAN en un algoritmo de clustering basado en estimación de densidad (Malzer & Baum, 2020; Ratzenböck et al., 2022).

Como veremos en la Sección 3.1.2, el espacio que emplearemos es el espacio de fase que incluye posiciones cartesianas y velocidades tangenciales. La detección de NYMGs como sobredensidades presenta tres inconvenientes principales: (i) la densidad de los NYMGs en el espacio de fase varía mucho de grupo a grupo, (ii) algunas estrellas de campo podrían generar sobredensidades espurias y (iii) algunas estrellas de campo podrían coincidir con las sobredensidades generadas por los NYMGs. Por estas razones es necesario explorar el espacio de hiperparámetros (ε, N_{\min}) para asegurarnos de que todos los grupos conocidos sean detectados y establecer la validez de las detecciones. Una forma de hacerlo es empleando el *Hierarchical DBSCAN* (HDBSCAN; Campello et al., 2013), que construye un dendrograma de grupos detectados con DBSCAN a diferentes niveles de densidad y usa diferentes métodos estadísticos para seleccionar los grupos finales. Aunque el HDBSCAN se ha utilizado para buscar NYMG (Kerr et al., 2021), podría ser menos eficiente que DBSCAN en la detección de estructuras con números muy diferentes de miembros (Malzer & Baum, 2020). Adicionalmente, en un estudio de subestructuras en la asociación Sco-Cen, Ratzenböck et al. (2022) encuentran mayores tasas de recuperación con DBSCAN que con HDBSCAN. Además, dado que los hiperparámetros de HDBSCAN no correlacionan directamente con la densidad local mínima de los grupos detectados, como sí lo hace DBSCAN, los resultados de éste último son más fáciles de interpretar. Por estos motivos preferimos usar DBSCAN en lugar de HDBSCAN.

3.1.2. Esquema general del algoritmo de detección

Es sabido que un NYMG forma una sobre-densidad en el espacio de fase de 6 dimensiones que forman las posiciones (X, Y, Z) y velocidades (U, V, W) en el sistema cartesiano de coordenadas galácticas. Esta sobredensidad se superpone a la distribución característica de las estrellas de campo en el espacio de fase. Además los miembros de un NYMG son coetáneos por lo que también se distribuyen a lo largo de su correspondiente isocrona en un DCMA. Lo anterior sugiere el uso del espacio de fase que forman

(X, Y, Z, U, V, W) pero, dado que solo el 44% de las fuentes en la muestra S_{NYMG} poseen VR de Gaia DR3, no podemos buscar NYMG en ese espacio sin reducir significativamente la completitud de los grupos resultantes. Por lo tanto, siguiendo a [Ratzenböck et al. \(2022\)](#), restringimos nuestra búsqueda al espacio de 5 dimensiones que llamamos Espacio Fase Restringido (EFR), que incluye las posiciones (X, Y, Z) y las velocidades tangenciales en el LSR $(v_{\alpha}^{\text{LSR}}, v_{\delta}^{\text{LSR}})$. Como vimos en la Sección 2.2.2, las velocidades tangenciales reducen la dispersión de los NYMGs en el espacio de movimientos propios y al trabajar en el LSR impedimos que el movimiento del Sol distorsione su distribución.

Para ejecutar el DBSCAN en un espacio de coordenadas con diferentes unidades físicas, es necesario que todas las coordenadas sean re-escaladas y adimensionadas. Para ello definimos el Espacio Fase Restringido Escalado (EFRE) como:

$$EFRE = (X/\varepsilon_r, Y/\varepsilon_r, Z/\varepsilon_r, v_{\alpha}^{\text{LSR}}/\varepsilon_v, v_{\delta}^{\text{LSR}}/\varepsilon_v) \quad (3.1.1)$$

donde hemos introducido el factor de escala para los espacios de posiciones cartesianas (ε_r) y el factor de escala para el espacio de velocidades tangenciales en el LSR (ε_v) . En el Apéndice A.1, mostramos cómo la búsqueda de sobredensidades en el EFRE usando $\varepsilon = 0,847\sqrt{2}$ es una buena aproximación para utilizar el DBSCAN de manera independiente en los espacios de posición y velocidad. Con este resultado concluimos que unas buenas estimaciones de las densidades mínimas para que se detecte una sobre-densidad son $\rho_r \sim N_{\text{min}}/(\frac{4}{3}\pi\varepsilon_r^3)$ y $\rho_v \sim N_{\text{min}}/(\frac{4}{3}\pi\varepsilon_v^2)$ en los espacios de posición y velocidad, respectivamente.

La eficiencia de DBSCAN aumenta si el cociente entre la densidad de estrellas de los NYMGs y la densidad de estrellas del campo aumenta ([Ester et al., 1996](#)). Por esta razón, las muestras S_{comp} excluyen todas las estrellas fuera de los loci SP y PSP, lo que reduce la densidad del campo sin reducir las densidades de los grupos.

Para disminuir aún más la densidad de estrellas del campo, dividimos cada una de las muestras S_{comp} en 10 sub-muestras según su posición respecto de isocronas en el DCMA M_G vs. $BP - RP$. La primera sub-muestra, llamada S_{comp}^{10} , incluye todas las fuentes que caen por encima de la isócrona de 10 Maños, considerando una tolerancia de 0,5 magnitudes hacia brillos más débiles.

De esta manera, S_{comp}^{10} incluye todas las fuentes más jóvenes que 10 Maños y considera las incertidum-

bres fotométricas. La selección no incluye un límite más brillante porque podría sesgar la muestra al excluir fuentes afectadas por extinción. Repetimos el mismo procedimiento para nueve isocronas adicionales con edades entre 20 Maños y 100 Maños a intervalos de 10 Maños y generamos las sub-muestras correspondientes S_{comp}^A , donde A es la edad de la isocrona. En todos los casos, usamos las isocronas de los modelos ATMO 2020 de Phillips et al. (2020) para masas $0,01 \leq M/M_{\odot} \leq 0,015$ y edades menores a 20 Maños y para masas $0,01 \leq M/M_{\odot} \leq 0,03$ y edades mayores a 20 Maños, siguiendo las conclusiones de Phillips et al. (2020). Luego, para edades y masas mayores a los límites anteriores, utilizamos los modelos de Baraffe et al. (2015) hasta $0,75 M_{\odot}$ y los modelos de Marigo et al. (2017) para masas $0,75 < M/M_{\odot} < 10$. Notamos que la sub-muestra correspondiente a una dada edad incluye a todas las sub-muestras de edades más jóvenes. Debido a que el algoritmo de detección se ejecuta independientemente sobre cada sub-muestra S_{comp}^A , si un grupo es detectado en una dada muestra, también debería ser detectado en todas las muestras de mayor edad. La ventaja de este procedimiento es doble: confirma la detección de un grupo dado en diferentes sub-muestras y disminuye la contaminación cuando el grupo es detectado en la sub-muestra más joven.

En resumen, el conjunto completo de hiperparámetros del algoritmo de detección está formado por: la edad A de la isocrona utilizada para la selección fotométrica, los factores de escala ε_r y ε_v , y el número mínimo de vecinos N_{min} . Para explorar la detección de NYMG con diferentes edades y densidades en el EFRE ejecutamos el algoritmo sobre cada sub-muestra S_{comp}^A para un conjunto de combinaciones de los hiperparámetros. Específicamente, usamos ε_r desde $5 pc$ hasta $30 pc$ en pasos de $2,5 pc$; ε_v desde $0,25 km s^{-1}$ hasta $5 km s^{-1}$ en pasos de $0,25 km s^{-1}$; y N_{min} desde 2 hasta 30 en pasos de 2.

Finalmente, para cada sobre-densidad detectada, se realizó un test de hipótesis con un nivel de significancia del 99 %, donde nuestra hipótesis nula \mathcal{H}_0 es: *Este grupo no es un NYMG, sino una fluctuación estocástica de la distribución del campo de las estrellas viejas de la Vía Láctea* donde dicho campo no contiene cúmulos más masivos y viejos que los NYMG. Luego descartamos todos los grupos que verifican \mathcal{H}_0 . La implementación y resultados de la prueba de hipótesis se presenta en la Sección 3.2.1.

3.2. Detección de fluctuaciones espurias del campo

3.2.1. Construcción de campos sintéticos

Para evaluar la significancia estadística de los grupos detectados por DBSCAN, estimamos la probabilidad de identificar sobre-densidades resultantes de fluctuaciones estocásticas de densidad en el campo. Para lograr esto, ejecutamos el algoritmo en una muestra sintética generada a partir de un modelo de campo que imita la distribución tanto en el EFR como en el DCMA M_G vs. $BP - RP$.

Como modelo del EFR, utilizamos el catálogo sintético *Gaia Object Generator* (Antiche et al., 2014) (GOG), que se genera aplicando los modelos de error de GDR3 a las poblaciones estelares galácticas predichas por los modelos de Besançon (Robin et al., 2012). Como GOG no incorpora cúmulos jóvenes ni NYMG, sirve como un catálogo ideal para construir un modelo del campo viejo de la Vía Láctea. Después de aplicar las máscaras descritas en la Sección 2.1.3 a GOG y completar la muestra utilizando el método descrito en la Sección 2.1.3, empleamos un *Kernel Density Estimator* (KDE) con un kernel Gaussiano¹ para modelar la distribución en el EFR de la muestra resultante. Finalmente, generamos a partir de este modelo un número de fuentes igual al número de fuentes en S_{comp} .

Para modelar la distribución de fuentes de campo en el DCMA, elegimos muestrear la distribución de S_{comp} en ese diagrama para evitar la discretización que el catálogo GOG muestra en el color fotométrico y que no reproduce adecuadamente los diagramas observados. Hicimos esto creando distribuciones acumuladas de M_G dentro de bins de $BP - RP$ de 0,1 magnitudes a lo largo de todo el rango de $BP - RP$ y generando el mismo número promedio de fuentes encontradas en las muestras S_{comp} dentro de cada bin. Este procedimiento asume que las distribuciones EFR y DCMA son independientes y que la gran mayoría de las fuentes en las muestras S_{comp} son estrellas de campo cuya distribución en el DCMA no cambia significativamente debido a la presencia de NYMG o cúmulos jóvenes.

¹La implementación del kernel Gaussiano no se puede hacer directamente en el espacio de 5 dimensiones que hemos definido pues las posiciones cambian entre $-200 pc$ y $200 pc$ y las velocidades entre $-20 km/s$ y $20 km/s$. La diferencia de un orden de magnitud entre ambos rangos numéricos genera problemas numéricos en el cálculo de la KDE. Por esta razón, primero multiplicamos las velocidades por un factor de 10 y luego aplicamos un kernel de ancho 5 para las 5 coordenadas (las tres posiciones y las dos velocidades modificadas). Luego de calculada la KDE las velocidades se regresaron a sus cantidades originales dividiendo entre un factor de 10.

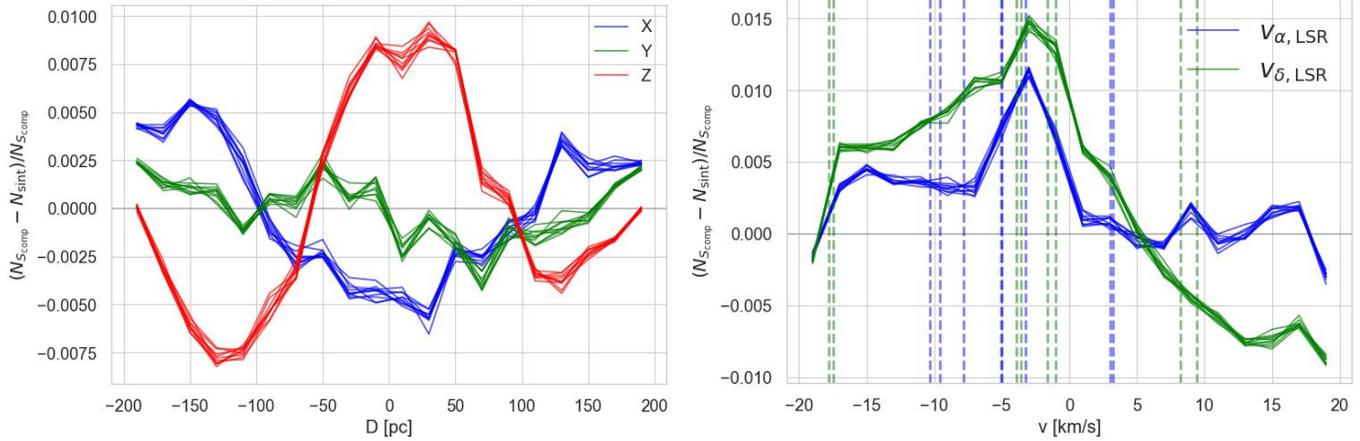


Figura 3.1: Distribución de residuos entre diez campos sintéticos y una de las muestras S_{comp} en el espacio de posiciones (panel izquierdo) y en el de velocidades tangenciales en el LSR (panel derecho). Las líneas discontinuas indican las velocidades medias de los cúmulos abiertos conocidos según la base de datos MOCA.

La Figura 3.1 muestra los residuos entre diez campos sintéticos y una de las diez muestras S_{comp} . Encontramos que las diferencias no superan el 1% en el espacio de posición y el 1,5% en el espacio de velocidad tangencial. Notablemente, encontramos que las regiones del espacio de velocidades en las que ocurren las mayores diferencias coinciden con las velocidades medias de cúmulos abiertos conocidos presentes en la muestra S_{comp} y compilados en la base de datos MOCA. Los residuos notablemente pequeños subrayan la eficiencia de GOG reproduciendo el campo viejo de la Galaxia observado en GDR3 incluso dentro del volumen limitado de nuestro estudio.

3.2.2. Frecuencia de detección de sobre-densidades espurias

Generamos diez campos sintéticos y aplicamos nuestro algoritmo de detección a cada uno de ellos para cada combinación de los hiperparámetros $(A, N_{\text{min}}, \varepsilon_r, \varepsilon_v)$. Para cada campo sintético y cada combinación de hiperparámetros, obtuvimos el número K de miembros detectados en cada sobredensidad. Construimos la distribución de K para cada sobre-densidad e identificamos $K_{\text{min},\alpha}$, como el cuantil α de la distribución.

Finalmente, calculamos el valor medio de $K_{\text{min},\alpha}$ en los diez campos sintéticos para cada combinación de hiperparámetros. Este valor medio representa el número mínimo promedio de miembros requerido para que una sobredensidad sea clasificada como un grupo con un nivel de significancia α . En este trabajo,

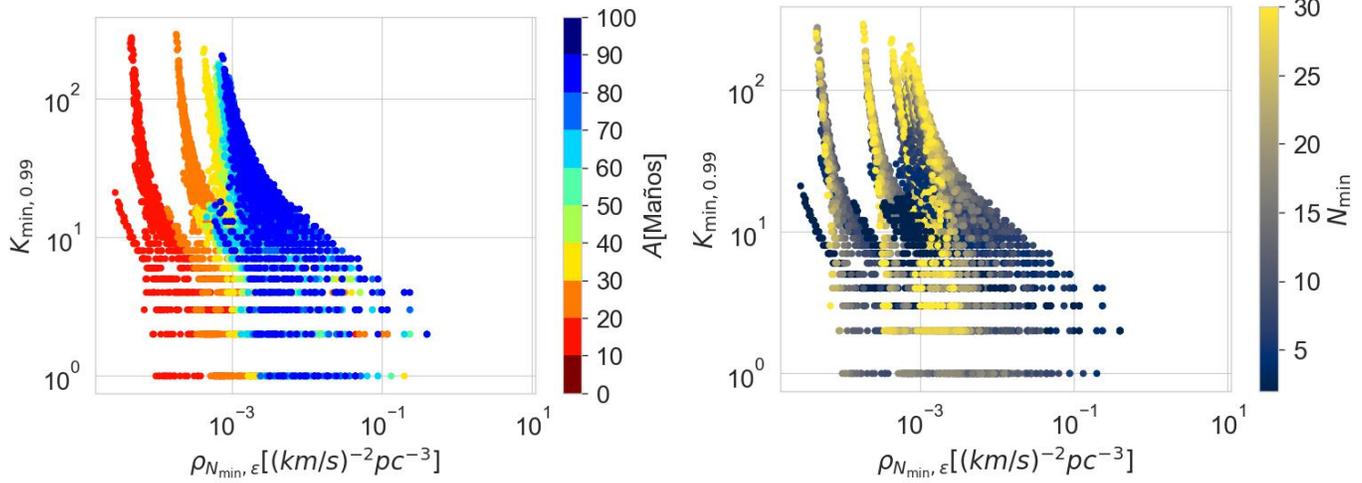


Figura 3.2: Valores de $K_{\min,0,99}$ como función de $\rho_{N_{\min},\epsilon} = N_{\min}/(\epsilon_r^3 \epsilon_v^2)$ para todas las combinaciones de los hiperparámetros. El mapa de colores indica la edad de la isocrona para la selección fotométrica (gráfico izquierdo) y los valores de N_{\min} (gráfico derecho).

definimos $\alpha = 0,99$. Como no esperamos que un NYMG tenga más de mil objetos, se descartan todas las combinaciones de hiperparámetros que produjeron $K_{\min,0,99} \geq 10^3$.

La Figura 3.2 muestra los valores finales de $K_{\min,0,99}$ como función de $\rho_{N_{\min},\epsilon} = N_{\min}/(\epsilon_r^3 \epsilon_v^2)$ (que es proporcional al umbral de densidad del DBSCAN) para todas las edades A consideradas. Como era de esperar, $K_{\min,0,99}$ disminuye a medida que $\rho_{N_{\min},\epsilon}$ aumenta independientemente de la edad de la isocrona. Otro fenómeno independiente de la edad es que al disminuir $\rho_{N_{\min},\epsilon}$, debe aumentar N_{\min} para que $K_{\min,0,99} < 10^3$. Esto significa que para que una detección sea significativa con un umbral de densidad pequeño, la sobredensidad detectada debe tener por lo menos un punto núcleo con un alto número de vecinos. Finalmente, notamos que a mayores edades, mayores deben ser los umbrales de densidad del DBSCAN, lo cual es consistente con que al aumentar la edad de la isocrona, aumenta el número de fuentes de la SP lo cual se traduce en un incremento de la densidad del campo.

3.3. Candidatos a grupos en las detecciones significativas

Aplicamos el algoritmo de detección a las diez muestras S_{comp} , considerando sobredensidades con $\alpha \geq 0,99$ como detecciones significativas, tal como se explica en la Sección 3.2.2. Los resultados incluyen

NYMG conocidos y nuevas sobredensidades. En este trabajo, nos centramos en inferir la IMF de los siguientes NYMGs conocidos: BPMG, BPMG-Sco, THA, EMUTAU, GRТАUS3A, GRТАUS2, VCA, GRS-COS7-8, COL, GRТАUS1, CAR y TW-Hya indicados en la Tabla 2.2. Estos grupos fueron seleccionados porque son representativas de los distintos NYMGs, son un número considerable de NYMGs, sólo uno de ellos posee una estimación previa de la IMF ((TW-Hya) [Gagné et al., 2017](#)) y tienen un número significativo de miembros conocidos en S_{lit} (≥ 36 miembros), lo que permite estimar la completitud y contaminación de las correspondientes sobredensidades detectadas. Esta estimación es crucial para validar nuestro método y calcular con precisión la IMF. El estudio de las restantes sobredensidades detectadas es parte de un trabajo en desarrollo derivado del presente estudio.

Para este estudio, consideramos para cada uno de los diez grupos y para la selección fotométrica la isocrona que mejor se ajusta a la distribución en el DCMA de los miembros del grupo según S_{lit} y descartamos los miembros que quedaran fuera del locus de selección. Este método nos permitió descartar contaminación evidente de la SP en algunos casos como BPMG donde existía una doble secuencia de objetos de la PSP y de la SP. De aquí en adelante, consideraremos a estos grupos de S_{lit} luego de haber aplicado esta selección fotométrica.

Todas las sobredensidades detectadas surgen de combinaciones específicas de hiperparámetros y algunas pueden incluso ser detectadas utilizando diferentes combinaciones de hiperparámetros. Los criterios para seleccionar la combinación de hiperparámetros adecuada para un dado NYMG se basan en maximizar la pureza y recuperación de las fuentes presentes en S_{lit} y en minimizar el nivel de fragmentación (NF). Estas tres últimas cantidades se definen de la siguiente manera:

1. Definimos la pureza como $f_p = N_{TD}/N_D$ donde N_D es el número de objetos de S_{NYMG} que forman la sobredensidad detectada y N_{TD} es el número de esos objetos que fueron clasificados como miembros del NYMG que además pertenecen al grupo según S_{lit} .
2. Definimos la recuperación como $f_r = N_{TD}/N_T$, donde N_T corresponde al número total de miembros del NYMG de acuerdo a S_{lit} .
3. Definimos el NF como el número de sobredensidades detectadas que contienen miembros de un mismo NYMG en S_{lit} . Esta situación surge cuando el umbral de densidad $\rho_{N_{min},\varepsilon}$ es demasiado alto respecto a la densidad del grupo o debido a la falta de miembros en los datos.

Elegimos las detecciones que cumplen con un NF menor a 3 y que minimizan la función de costo:

$$f_{rp} = \sqrt{w_r(1 - f_r)^2 + w_p(1 - f_p)^2} \quad (3.3.1)$$

donde $w_r = 1$ y $w_p = 0,5$ son pesos elegidos arbitrariamente. Los valores elegidos priorizan la recuperación ante la pureza ya que la pureza se basa en la información de S_{lit} que pudiera estar contaminada o sesgada. La Tabla 3.1 muestra los resultados generales de las detecciones y que maximizan f_{rp} .

Detectamos un total de 1229 candidatos cinemáticos, de los cuales 625 son candidatos recuperados de la literatura, lo cual representa un 66 % de todos los 940 candidatos de los grupos estudiados de literatura. Esto significa que hemos identificado 604 nuevos candidatos cinemáticos, lo cual representa un 49 % de todos los candidatos detectados. Sin embargo, según el análisis del Apéndice A.4, estimamos que entre 105 y 287 de estos nuevos candidatos (entre 17 % y 48 %) son estrellas contaminantes con velocidades tangenciales similares a la de los grupos detectados pero que probablemente tengan VR diferentes.

Logramos niveles altos de recuperación y pureza respecto a la literatura ($f_r, f_p > 0,6$) para los grupos GRTAUS3A, GRTAUS2, THA, EMUTAU y VCA. Además estos valores se alcanzaron con un NF de 1 con la excepción de EMUTAU (NF= 2). Luego, solamente logramos alcanzar valores de recuperación en el intervalo $0,32 < f_r < 0,4$ para las poblaciones TW-Hya, BPMG y CAR pero sí un nivel alto de pureza respecto a la literatura ($0,6 < f_p < 0,88$). En el caso de COL, obtuvimos una recuperación alta de 0.64 pero una baja pureza de $f_p = 0,36$. Finalmente, en cuanto a BPMG-SCO, solamente logramos alcanzar una recuperación de $f_r = 0,44$ y una pureza respecto de la literatura de $f_p = 0,21$.

La Figura 3.3 muestra el DCMA y diagramas cinemáticos del grupo BPMG que es uno de los que muestra menor tasa recuperación. Figuras análogas para los restantes grupos se muestran en el Apéndice A.2. Notamos que los casos de menores recuperaciones (BPMG, CAR, TW-Hya y BPMG-SCO) corresponden sistemáticamente a las poblaciones con mayor dispersión en el espacio de velocidades tangenciales. Esto pudiera explicar porque es tan difícil re-detectar todos los miembros de estos grupos, particularmente aquellos que además muestran una mayor dispersión espacial.

Ahora, en la Figura 3.4 se muestran los valores de f_r como función de la dispersión de velocidades tangenciales $\sigma_{v_{LSR}}$ y la distancia media de cada grupo, notamos que a mayores distancias, los valores de

NYMG	ε_r [pc]	ε_v [km/s]	N_{\min}	f_r	f_p	f_{rp}	NF	m_c [M_{\odot}]	σ	α	A_{IMF} [Maños]	A_{lit} [Maños]	N_{nuevos}
TW-Hya	10	3	6	0.40	0.88	0.60	1	$0,32 \pm 0,17$	$0,80 \pm 0,28$	$-0,41 \pm 0,88$	10	10	5
GRTAUS3A	10	1.4	4	0.75	0.86	0.27	1	$0,20 \pm 0,06$	$0,40 \pm 0,35$	$-1,26 \pm 1,97$	10	19	4
BPMG-SCO	5	2.2	6	0.44	0.21	0.79	2	$0,16 \pm 0,06$	$0,30 \pm 0,30$	$-2,68 \pm 0,48$	10	25	130
BPMG	10	2	4	0.39	0.85	0.62	2	$0,18 \pm 0,07$	$0,45 \pm 0,33$	$-3,08 \pm 0,70$	10	26	57
GRTAUS2	12.5	4	30	0.73	0.81	0.31	1	$0,22 \pm 0,08$	$0,55 \pm 0,30$	$-2,26 \pm 0,13$	20	31	15
COL	17.5	0.8	6	0.64	0.36	0.58	2	$0,16 \pm 0,06$	$0,30 \pm 0,32$	$-2,63 \pm 0,44$	20	42	150
CAR	12.5	0.8	2	0.32	0.60	0.73	2	$0,16 \pm 0,05$	$0,35 \pm 0,33$	$-0,65 \pm 0,23$	20	45	24
THA	15	1	2	0.66	0.65	0.42	1	$0,14 \pm 0,05$	$0,35 \pm 0,33$	$-2,54 \pm 0,43$	20	51	141
EMUTAU	12.5	2	30	0.90	0.73	0.22	2	$0,14 \pm 0,05$	$0,20 \pm 0,33$	$-1,63 \pm 0,08$	20	56	62
VCA	10	2.8	20	0.89	0.81	0.17	1	$0,18 \pm 0,05$	$0,25 \pm 0,35$	$-2,29 \pm 1,03$	50	89	16

Tabla 3.1: Resultados generales de las detecciones de los NYMGs estudiados que minimizan la función de costo f_{rp} . Las columnas ε_r , ε_v y N_{\min} indican los hiperparámetros correspondientes a la detección de estos grupos y la columna frag. indica el NF. Luego, las columnas m_c , σ y α indican respectivamente la masa característica, el ancho característico y la pendiente de los ajustes de las IMF inferidas. Las columnas A_{IMF} y A_{lit} indican las edades de la isocrona con la que se infirieron las masas y reportada en la literatura respectivamente. La columna N_{nuevos} indica el número de nuevos candidatos fotométricos y cinemáticos a miembros detectados.

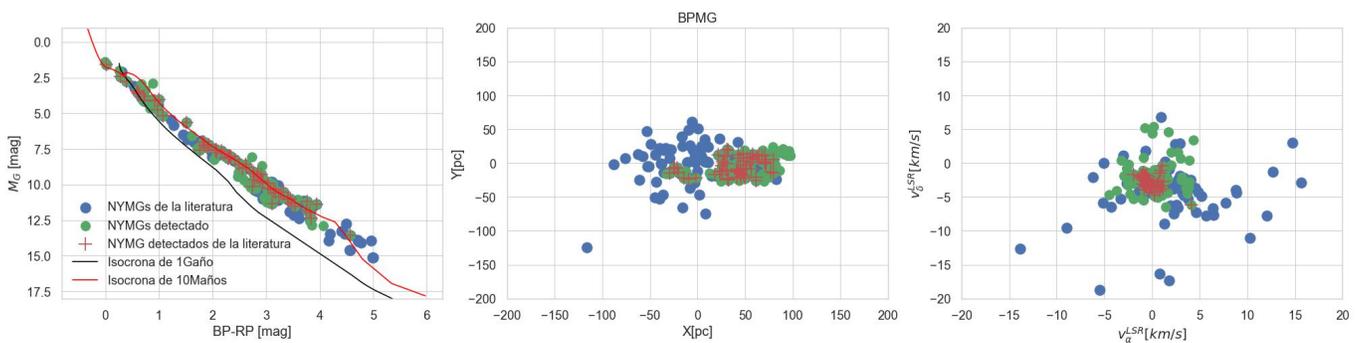


Figura 3.3: Distribuciones en el DCMA (izquierda), en el plano Y vs. X (centro) y en el espacio de velocidades tangenciales (derecha) del grupo BPMG. Los puntos azules indican los miembros del grupo según la literatura (muestra S_{lit}), los puntos verdes indican las detecciones de este trabajo y las cruces rojas los miembros de la literatura recuperados con nuestras detecciones. En el DCMA, la curva negra indica la isocrona de 1Gaño y la curva roja de la isocrona utilizada para la inferencia de las masas.

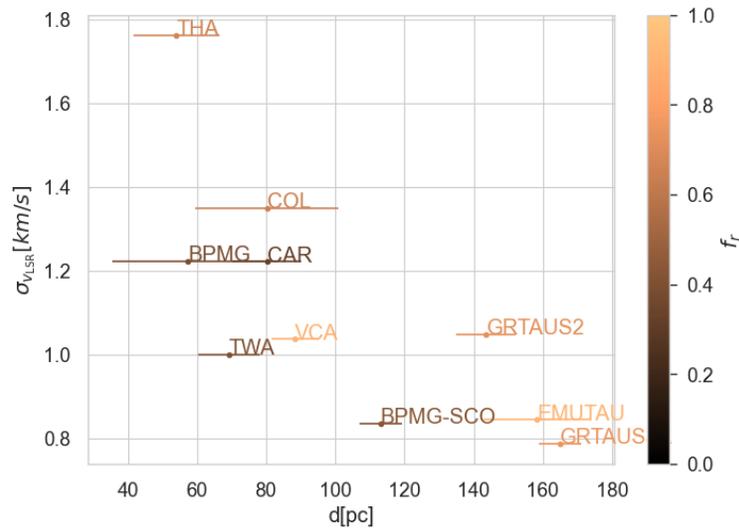


Figura 3.4: Desviación estándar de la velocidad tangencial $\sigma_{v_{LSR}}$ de cada NYMG detectado como función de la distancia promedio para los casos que minimizan f_{rp} . El mapa de colores indica los valores obtenidos de f_r y las barras de error la dispersión en distancias.

$\sigma_{v_{LSR}}$ disminuyen. Este fenómeno se debe a que a menores distancias, los miembros de un NYMG tenderán a ser proyectados en mayores regiones de la esfera celeste como se observa en la Figura 2.6, lo cual tiende a aumentar la dispersión de velocidades tangenciales del grupo. Esto significa que en los casos de BPMG, TW-Hya y CAR, la dificultad de detectarlos pudiera deberse a este sesgo puramente observacional. Pero ciertamente este no es el caso de BPMG-SCO que, de las figuras del Apéndice A.2 y a partir de los datos de la literatura es un grupo que parece estar dividido en dos sub-grupos en el espacio de posiciones.

Capítulo 4

Inferencia de las IMFs de los NYMGs estudiados

4.1. Inferencia Bayesiana de masas a partir de la fotometría

Estimamos la masa de cada estrella a partir de la fotometría y de los paralajes de GDR3, asumiendo su edad de S_{lit} y usando la regla de Bayes:

$$P(m|\overrightarrow{GBR}) = \frac{\mathcal{L}(\overrightarrow{GBR}|m)\mathcal{P}(m)}{\int_{m_{\min}}^{m_{\max}} \mathcal{L}(\overrightarrow{GBR}|m)\mathcal{P}(m)dm} \quad (4.1.1)$$

donde $\overrightarrow{GBR} = (BP - RP, M_G)$ representa la posición de una estrella en el DCMA. El prior $\mathcal{P}(m)$ es un modelo de la IMF, que asumimos como la IMF de [Chabrier \(2003\)](#) de estrellas individuales, descrita por una distribución log-normal con una masa característica $m_c = 0,2 M_{\odot}$ y $\sigma = 0,6$ para masas menores que $1 M_{\odot}$ y una ley de potencias de [Salpeter \(1955\)](#) para masas mayores. Con respecto al Likelihood $\mathcal{L}(\overrightarrow{GBR}|m)$, dado que BP y RP son medidos por un instrumento diferente al que mide G , podemos asumir que las incertidumbres en $BR = BP - RP$ y G son independientes. Además, asumiendo que G , BP y RP tienen errores gaussianos y que el error del paralaje es despreciable ya que todas las fuentes

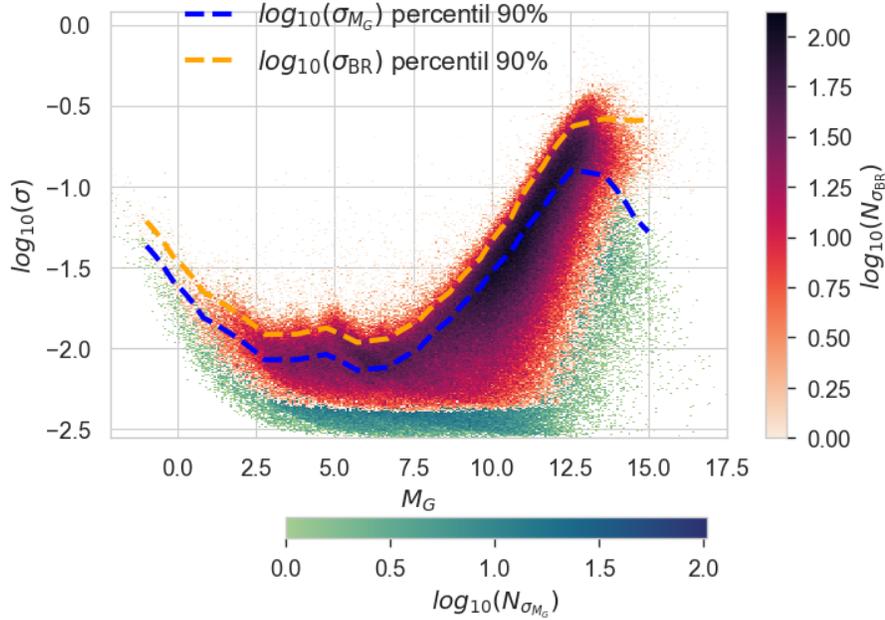


Figura 4.1: Errores σ_{M_G} y σ_{BR} como función de M_G de la muestra S_{NYMG} y la interpolación del percentil 90 % de estos.

cumplen con $\sigma_{\varpi}/\varpi < 0,1$, entonces $BP - RP$ y M_G también tendrán errores gaussianos, lo que implica que $\mathcal{L}(\overrightarrow{GBR}|m)$ se puede calcular como:

$$P(\overrightarrow{GBR}|m) = P(M_G|m)P(BR|m) = \frac{1}{2\pi(\sigma_G\sigma_{BR})^{1/2}} e^{-\frac{(M_G - M'_G)^2}{2\sigma_G^2}} e^{-\frac{(BR - BR')^2}{2\sigma_{BR}^2}} \quad (4.1.2)$$

donde M'_G y BR' son, respectivamente, la magnitud absoluta en la banda G y el color $BP' - RP'$ calculados a partir de la relación masa-luminosidad interpolada de los modelos de [Marigo et al. \(2017\)](#), [Baraffe et al. \(2015\)](#) y [Phillips et al. \(2020\)](#). La desviación estándar σ_G se estima como el error del percentil 90 % en M_G de las fuentes de S_{NYMG} dentro de un bin de M_G que contiene el valor de M_G medido. De manera similar, muestreamos los valores de los errores σ_{BP} y σ_{RP} en las magnitudes BP y RP , respectivamente, con los que calculamos $\sigma_{BR} = (\sigma_{BP}^2 + \sigma_{RP}^2)^{1/2}$. Los errores fotométricos y los interpolados se presentan en la Figura 4.1 donde notamos que estos errores aumentan para objetos muy brillantes o muy débiles y son consistentemente mayores para BR que para M_G . Finalmente encontramos la masa m que maximiza la posterior. Este enfoque nos permite estimar un error para cada masa estelar.

Para ilustrar el procedimiento, la Figura 4.2 muestra la recuperación de las masas individuales de las estrellas de una población sintética constituida por 1500 estrellas. Las masas iniciales se generaron a partir del modelo analítico de la IMF de Chabrier (2003) y se utilizaron los modelos evolutivos de Phillips et al. (2020), Baraffe et al. (2015) y Marigo et al. (2017) para una edad de 30 Maños para asignar las magnitudes sintéticas en las bandas de GDR3 a las que se les sumó los errores fotométricos característicos. A estos datos se aplicó el procedimiento de inferencia descripto con las ecuaciones 4.1.1 y 4.1.2 asumiendo la isocrona de 30 Maños. A partir de las masas inferidas se generó la IMF sintética. Encontramos que los errores en la inferencia de las masas crecen para objetos muy brillantes o muy ténues pero en todos los casos permanecen menores al 10 %. El acuerdo entre la IMF asumida para la construcción de la población sintética y la IMF inferida por nuestro método luego de asignar incertidumbres fotométricas realistas es excelente.

4.1.1. Las IMFs inferidas

Para cada uno de los candidatos a miembro de NYMG detectados y seleccionados en el Capítulo 3, ajustamos a la distribución de miembros en el DCMA la isocrona de los modelos de Phillips et al. (2020), Baraffe et al. (2015) y Marigo et al. (2017) e inferimos las masas de cada una de las estrellas del grupo a partir de esta isocrona según el método desarrollado en la Sección 4.1. Luego estimamos la IMF de cada NYMG ajustando un KDE a todas las masas inferidas del grupo. Siguiendo la misma metodología inferimos las masas y la IMF de cada grupo a partir de los miembros reportados en la literatura. Finalmente, para ambos casos, obtuvimos para cada grupo la log-normal que minimiza el *Root Mean Square Error* (RMSE) con el KDE de la IMF para masas $M < 1 M_{\odot}$ y ajustamos la ley de potencia que minimiza el RMSE con la KDE de la IMF para masas $M > 1 M_{\odot}$.

A continuación, estudiamos en detalle los casos de β -Pictoris y TW-Hya a modo de ejemplo y se presentan los demás casos en el Apéndice A.2. En la siguiente Sección 4.1.2 estudiamos la IMF promedio de todos los grupos y la correspondiente distribución de parámetros. Los resultados generales de los ajustes de todas las IMFs se presentan en la Tabla 3.1.

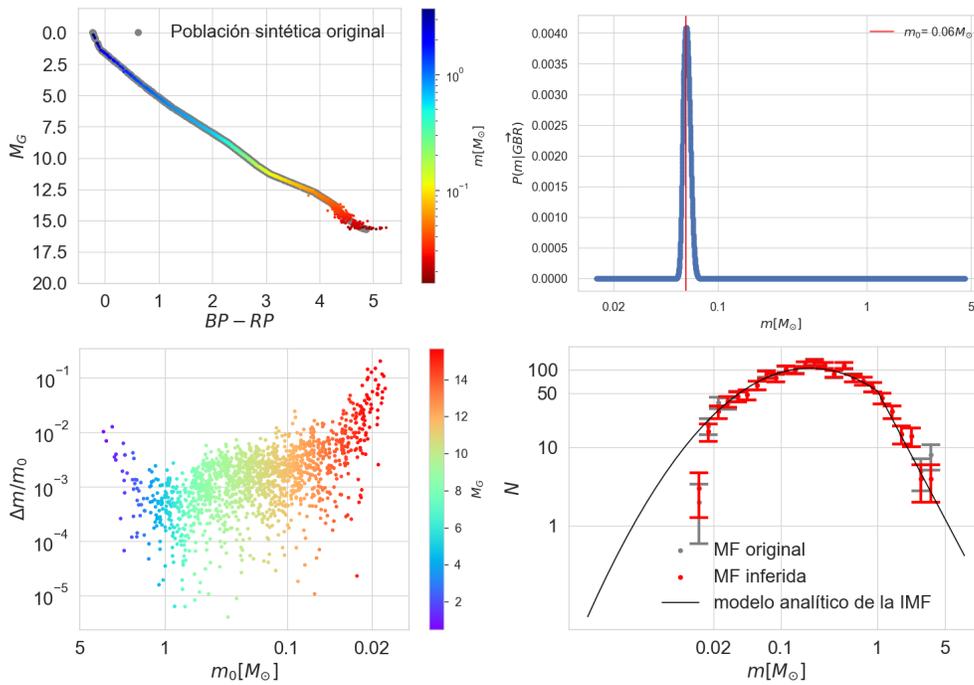


Figura 4.2: Ejemplo del procedimiento de inferencia de masas individuales sobre una población sintética de 1500 estrellas y calidad de las masas inferidas. *Izquierda superior*: DCMA de una población sintética antes (puntos grises) y después (puntos de colores) de aplicar errores fotométricos. Los puntos grises son más grandes que los de colores para efectos de visualización pero la secuencia de puntos grises posee menor dispersión. Nótese el aumento de la dispersión para objetos más débiles (menos masivos). *Derecha superior*: Posterior resultante de la inferencia de la masa de una de las estrellas sintéticas con masa original $m_0 = 0,06 M_{\odot}$. *Izquierda inferior*: Diferencias entre las masas inferidas y las respectivas masas originales (m_0). *Derecha inferior*: IMF original de la población sintética (puntos grises) e IMF inferida (puntos rojos). Los barras de errores corresponden a errores de conteo. La curva negra indica la IMF de [Chabrier \(2003\)](#) con la cual se construyó la población sintética.

El caso de β -Pictoris

En la base de datos MOCA se encuentra el grupo denominado BPMG, el cual corresponde al NYMG comúnmente conocido en la literatura como β -Pictoris. Por otro lado, también se puede encontrar el grupo GRSCOS1 con el siguiente comentario en MOCA: *This is a fragment of BPMG towards Upper Sco*. Basados en este comentario, decidimos renombrar a GRSCOS1 como BPMG-SCO e inferir la IMF considerando a BPMG y BPMG-SCO como un mismo grupo y las IMFs de ambos grupos por separado. La Figura 4.3 presenta los resultados de la estimación de las IMFs y los correspondientes ajustes para los grupos BPMG, BPMG-SCO y la unión de ambos grupos, tanto para las detecciones obtenidas en este trabajo como a partir de los miembros de la literatura.

Para BPMG, con nuestras detecciones obtuvimos resultados similares a los que se obtienen empleando los miembros de la literatura para masas $m > 0,1 M_{\odot}$. Encontramos una distribución plana en escala logarítmica en la IMF en el rango $0,1 \leq m/M_{\odot} \leq 1$ y un máximo local para $m \sim 1 M_{\odot}$ antes de una caída pronunciada en el rango $m > 1 M_{\odot}$. Por otra parte, en el rango de las estrellas de bajas masas y enanas marrones ($m < 0,1 M_{\odot}$), detectamos menos objetos que los reportados en la literatura. En cuanto a los parámetros del ajuste a una función log-normal, nuestras detecciones resultaron en una masa característica $m_c = 0,18 \pm 0,07 M_{\odot}$ y ancho característico $\sigma = 0,45 \pm 0,33$ ligeramente menores a los que se obtienen empleando los miembros de la literatura, aunque consistentes dentro de las incertidumbres: $m_c = 0,22 \pm 0,14 M_{\odot}$ y $\sigma = 0,95 \pm 0,23$. La diferencia entre ambos anchos característicos probablemente se deba a la falta de enanas marrones detectadas respecto a los miembros de la literatura.

Para el intervalo de altas masas ($m > 1 M_{\odot}$), hallamos que para ambas muestras el ajuste a leyes de potencia da pendientes más inclinadas que el famoso valor $\alpha = -2,35$ hallado por [Salpeter \(1955\)](#). Los valores encontrados fueron $\alpha = -2,92 \pm 0,23$ y $\alpha = -3,08 \pm 0,70$ para los miembros de la literatura y de nuestras detecciones, respectivamente.

Para el grupo BPMG-SCO encontramos resultados similares en el rango de altas masas pero para masas $m < 1 M_{\odot}$ hay más objetos en nuestras detecciones que en la literatura. No observamos el comportamiento plano de BPMG lo que produjo una log-normal más estrecha par ambas muestras. Encontramos que el exceso aparente que se observa alrededor de $m \sim 1 M_{\odot}$ en BPMG también se observa en BPMG-SCO y que la masa en la que se produce dicho exceso coincide con la posición aproximada del punto de entrada a la SP de la isocrona correspondiente. De aquí en adelante y en analogía con el *Turn-Off*, llamare-

mos a dicho punto el *Turn-On*, ya que corresponde a la etapa evolutiva de la correspondiente isocrona en la que se estabiliza la fusión de H . Es de esperar que se produzcan contaminaciones en las regiones del DCMA en las que la isocrona utilizada para seleccionar la población se confunde con la SP ya que permitiría que estrellas viejas del campo que casualmente comparten cinemática con el grupo sean identificadas erróneamente como miembros. Sin embargo, esta contaminación debería disminuir hacía el locus superior de la SP ya que al disminuir la magnitud de una estrella de la SP disminuye el límite superior de su edad y por lo tanto la probabilidad que hayan estrellas viejas del campo. Por este motivo es de esperar que la contaminación crezca desde bajas masas hacia el *Turn-On* antes de disminuir hacía masas mayores. Este fenómeno pudiera entonces explicar la presencia de estos excesos alrededor de la masa del *Turn-On*. A su vez, esta contaminación pudiera explicar porque obtuvimos una pendiente ligeramente más plana que la de [Salpeter \(1955\)](#) en el rango de altas masas.

Finalmente, cuando consideramos la unión de BPMG y BPMG-SCO, encontramos que la forma de las IMFs para los miembros de nuestra detección y para los miembros de la literatura resulta mucho más suave que para los grupos individuales. En el rango de altas masas seguimos obteniendo una pendiente más empinada que [Salpeter \(1955\)](#), lo cual probablemente sea consecuencia de la suma de los excesos de los dos grupos al rededor del *Turn-On*.

Destacamos que para las tres IMFs (BPMG, BPMG-SCO y la suma de ambas), los valores m_c y σ son consistentes dentro de las incertidumbres con los valores $m_c = 0,2 M_\odot$ y $\sigma = 0,6$ hallados por [Chabrier \(2003\)](#) para la población del disco galáctico tanto en el caso de las IMFs inferidas de nuestras detecciones como en el caso de las inferidas de S_{lit} .

El caso de TW-Hya

Los resultados para el grupo TW-Hya se presentan en la Figura 4.3. La IMF inferida a partir de los miembros de la literatura ($m_c = 0,14 \pm 0,06 M_\odot$ y $\sigma = 0,55 \pm 0,28$) resulta menos dispersa que la IMF reportada por [Gagné et al. \(2017\)](#) y tiene un mejor acuerdo con [Chabrier \(2003\)](#). En cuanto a la IMF inferida a partir de nuestras detecciones, notamos que, aunque detectamos menos miembros que la literatura, la IMF tiene una masa característica ($m_c = 0,32 \pm 0,17 M_\odot$) consistente con la hallada por [Suárez et al. \(2019\)](#) en el cúmulo joven 25-Ori ($m_c = 0,31 \pm 0,04 M_\odot$) pero hallamos una dispersión ($\sigma = 0,80 \pm 0,28$) mayor que el encontrado por estos autores ($\sigma = 0,46 \pm 0,05$). Destacamos que la población de 25-Ori es una población

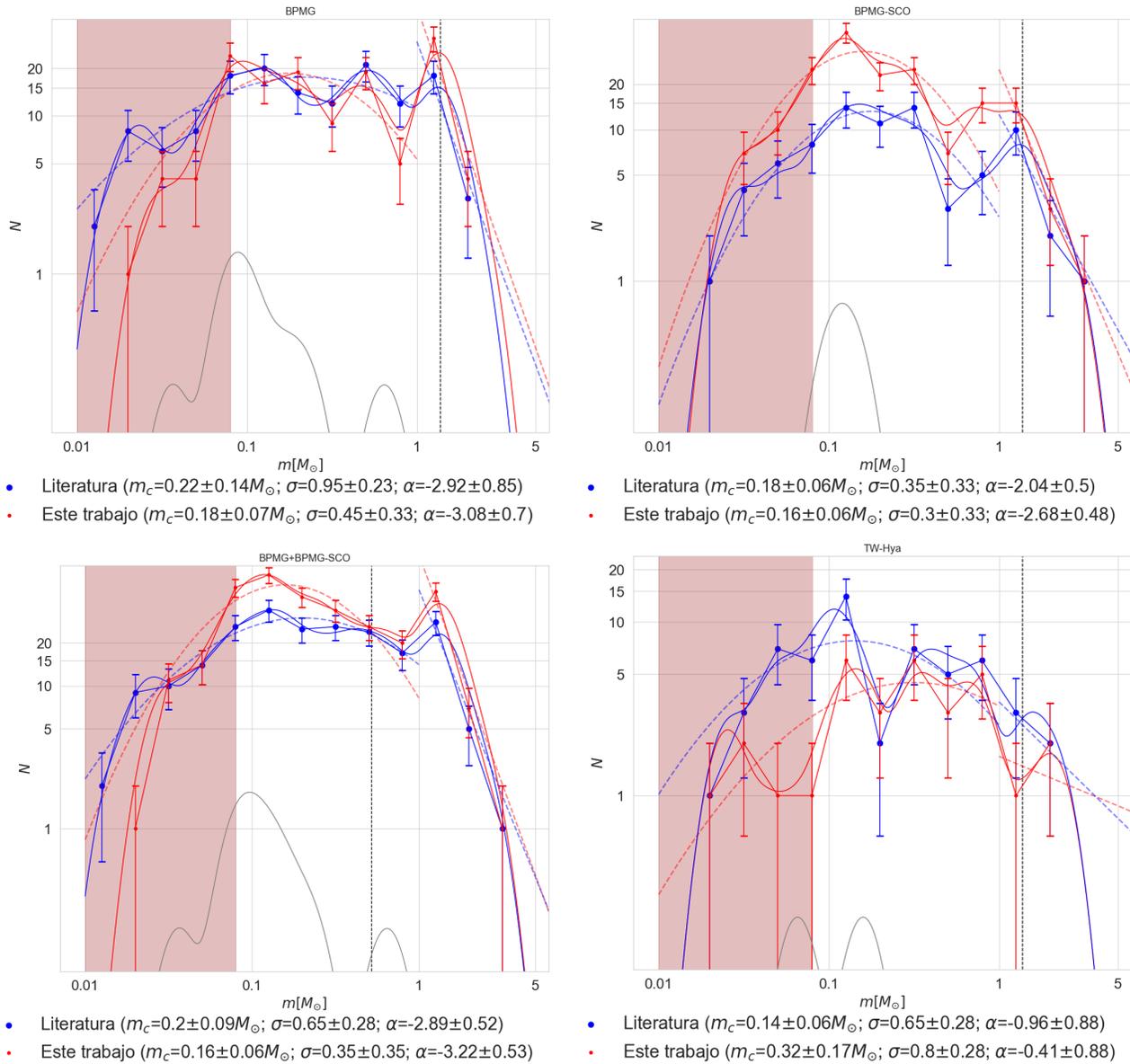


Figura 4.3: IMFs para los grupos BPMG (superior izquierda) y BPMG-SCO (superior derecha) y de la unión de ambos grupos (inferior izquierda) y TW-Hya (inferior derecha), inferidas a partir de las detecciones de este trabajo (rojo) y de las detecciones de la literatura (azul). Para cada uno de los dos casos, los puntos conectados corresponden al histograma de las masas con un bin de $\log_{10}(m/M_\odot) = 0,2$, las curvas sólidas al KDE de las masas, la curva a trazos corresponde a la log-normal que minimiza el RMSE con el KDE para masas $m < 1 M_\odot$ y la recta a trazos corresponde a la ley de potencias que minimiza el RMSE con el KDE para masas $m > 1 M_\odot$. La curva sólida gris indica el KDE promedio de las fuentes sintéticas detectadas como miembros del grupo. Finalmente, la región marrón indica el intervalo de masas de las enanas marrones y la recta vertical negra a trazos la masa aproximada del "Turn-On".

no ligada gravitacionalmente (Suárez et al., 2017, 2019), al igual que los NYMGs. En la región de altas masas ($m/M_{\odot} > 1$) hallamos pendientes significativamente más planas que la pendiente de Salpeter (1955) tanto usando los miembros de la literatura como a partir de nuestras detecciones. Los resultados de este grupo tienen un alto nivel de incertidumbre dada la escasa cantidad de objetos en la población en todo el rango de masas, lo cual se hace particularmente pronunciado en los extremos de la IMF.

4.1.2. La Función de Masa normalizada media

La Figura 4.4 muestra las distribuciones de los m_c , σ y α estimados de las poblaciones individuales. Notamos que tanto en el caso de nuestras detecciones como en el caso de los miembros de la literatura el valor promedio de m_c ($0,19 \pm 0,07 M_{\odot}$ y $0,17 \pm 0,07 M_{\odot}$, respectivamente) es consistente con el valor de Chabrier (2003) para la población del disco dentro de las incertidumbres, pero se aleja del valor obtenido por Suárez et al. (2019) para 25-Ori. Luego, observamos que tanto para la muestra de la literatura como para nuestras detecciones, el σ promedio ($0,42 \pm 0,31$ y $0,40 \pm 0,32$, respectivamente) es consistente con los valores de Chabrier (2003) y Suárez et al. (2019). En cuanto a los valores de α , notamos que la moda de la distribución se ubica en $\sim -2,5$ a partir de nuestras detecciones lo cual coincide con el valor hallado por Suárez et al. (2019) pero es ligeramente más empinada al hallado por Salpeter (1955).

Las incertidumbres de los parámetros indicados en la Figura 4.4 pudieran ser explicados por un lado por los bajos números de objetos en los rangos de baja y alta masa y, por otro, por potenciales contaminaciones e incompletitudes. Por ejemplo, como se vió en la Sección 4.1.1, algunos grupos como β -Pictoris presentan una aparente contaminación en el entorno de la masa del *Turn-On* cuando otras como TW-Hya poseen pocos miembros en los extremos de la IMF, lo cual participa particularmente en la incertidumbre de las estimaciones de α y σ . Destacamos sin embargo, que incluso con estas incertidumbres logremos obtener parametrizaciones consistentes con los resultados de Chabrier (2003), Suárez et al. (2019) y Salpeter (1955).

Luego, normalizamos las IMFs individuales a sus respectivas áreas y calculamos la IMF normalizada media resultante, la cual se presenta en la Figura 4.5. Esta distribución media nos permite, por un lado, reducir la incertidumbre debida a errores de conteo y por otro, reducir las incertidumbres producidas por incompletitudes y contaminaciones que ocurren en distintos rangos de masa para distintas poblaciones. Si por el contrario, existen incompletitudes o contaminaciones sistemáticas en determinados rangos de masa,

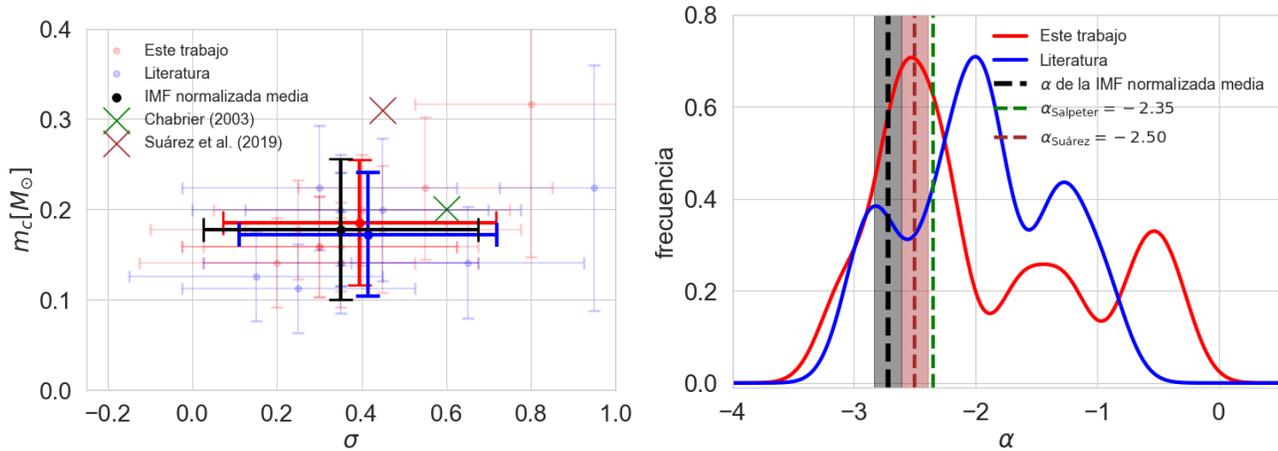


Figura 4.4: Distribución de los parámetros de ajuste de las IMF obtenidas. En los dos paneles el color rojo corresponde a los resultados obtenidos de nuestras detecciones, el azul a los obtenidos con datos de la literatura y el marrón a los resultados de la IMF de 25-Ori inferida por [Suárez et al. \(2019\)](#). *Izquierda*: parámetros m_c y σ de las funciones log-normales ajustadas para $m < 1 M_\odot$ de los grupos individuales (puntos más transparentes), de los valores medios de todos los grupos (puntos sólidos) y de la IMF normalizada media (punto negro). La cruz verde indica los valores de la IMF del disco de [Chabrier \(2003\)](#). *Derecha*: KDE de las pendientes del ajuste a leyes de potencia para $m > 1 M_\odot$. La recta a trazos verde indica el valor de α de la IMF de [Salpeter \(1955\)](#).

entonces estas deberían heredarse a la IMF normalizada media. Esto significa que si asumimos que los NYMGs se formaron de una misma IMF, entonces la IMF normalizada media nos permite estudiar las incompletitudes y contaminaciones provenientes de orígenes comunes a la mayoría de las poblaciones.

En la Figura 4.5, notamos que la IMF normalizada media es mucho más suave que la mayoría de las IMFs individuales, como era de esperar. Además, la IMF media heredó de las IMFs individuales la aparente escasez de objetos de baja masa, lo que se traduce en una log-normal más estrecha ($\sigma = 0,35 \pm 0,33$), aunque consistente dentro de las incertidumbres, con la obtenida por [Chabrier \(2003\)](#) y [Suárez et al. \(2019\)](#). Análogamente, esta IMF heredó de las IMFs individuales un aparente exceso de objetos en torno a la masa solar, lo que se traduce en una distribución más empinada en el rango de altas masas ($\alpha = -2,72 \pm 0,12$) respecto a [Salpeter \(1955\)](#) y [Suárez et al. \(2019\)](#). No obstante, la diferencia entre el α que obtuvimos y el obtenido por [Suárez et al. \(2019\)](#) no es muy grande y la masa característica obtenida ($m_c = 0,18 \pm 0,08$) es similar a la estimada por [Chabrier \(2003\)](#).

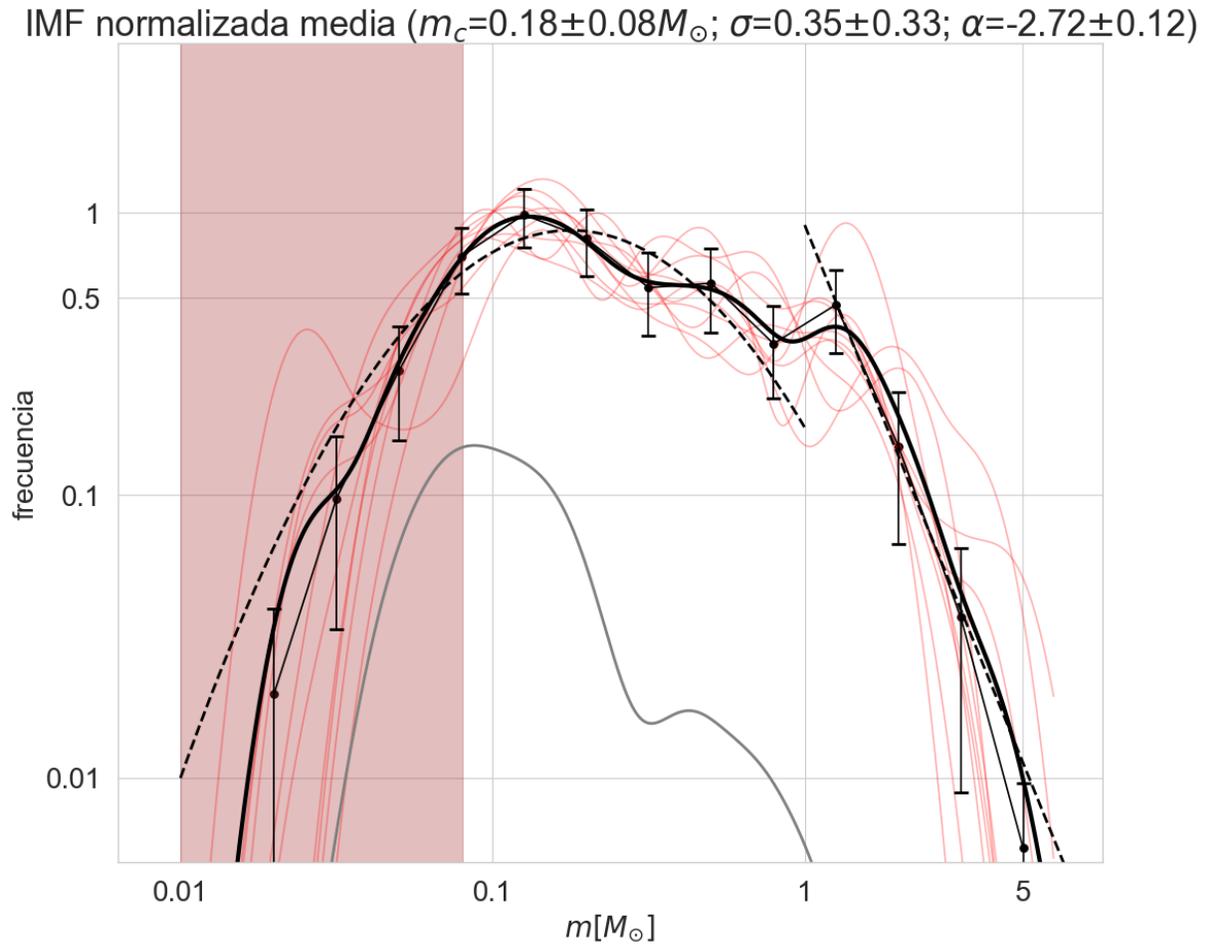


Figura 4.5: IMF normalizada promedio de todos los NYMGs estudiados en este trabajo. La curva negra sólida corresponde a la curva promedio de todos los KDE de las IMFs individuales normalizadas a sus respectivas áreas (curvas rojas) . La curva negra a trazos negra y la recta negra a trazos indican los ajustes de la IMF normalizada media a una función log-normal y a una ley de potencias en los rangos de baja y alta masa, respectivamente. La curva gris indica la IMF normalizada sintética media.

Capítulo 5

Análisis, discusión y conclusiones

En este capítulo presentamos el análisis, discusión y conclusiones de los resultados del trabajo. Dividimos el capítulo en una sección técnica referente a los métodos que desarrollamos e implementamos y una sección científica sobre el aporte de nuestro trabajo en la descripción de los NYMGs. Finalmente analizamos el futuro del proyecto tanto en el estudio de los NYMGs, como en la aplicación de los métodos desarrollados a otras poblaciones estelares jóvenes de interés ubicadas a mayores distancias.

5.1. Métodos y herramientas desarrollados en este trabajo

Para el estudio de la IMF de los NYMGs, hemos desarrollado en Python herramientas que permiten: (i) completar una submuestra de GDR3 (Secciones 2.1.2 y 2.1.3), (ii) generar campos sintéticos y obtener el número mínimo de miembros que debe tener una sobredensidad para ser estadísticamente significativa (Sección 3.2.2), (iii) aplicar el algoritmo de detección basado en el DBSCAN (Sección 3.1.2) e (iv) inferir masas estelares a partir de la fotometría de GDR3 (Sección 4.1). De los paquetes utilizados para construir estas herramientas, se destacan los paquetes Scikit-Learn (Pedregosa et al., 2011) que incluye la función para aplicar el DBSCAN, Gaiainlimited (Cantat-Gaudin et al., 2023; Castro-Ginard et al., 2023) para el manejo de las funciones de selección de GDR3 y Astropy (Astropy Collaboration et al., 2022) principalmente para la transformación de coordenadas y referenciales. Nuestros códigos también hacen uso de resultados

de los modelos de evolución estelar de [Phillips et al. \(2020\)](#), [Baraffe et al. \(2015\)](#) y [Marigo et al. \(2017\)](#) y del catálogo sintético GOG ([Antiche et al., 2014](#)).

5.2. Los NYMGs detectados y su comparación con la literatura

En este trabajo hemos estudiado los NYMGs: BPMG, BPMG-SCO, CAR, COL, EMUTAU, VCA, GRTAUSA, GRTAUS2 y TW-Hya. Con el método de detección desarrollado encontramos un total de 1229 candidatos cinemáticos y fotométricos a miembros de estas poblaciones. La muestra de candidatos reportados en la literatura (S_{lit}) contiene 940 objetos para estos NYMGs de los cuales 625 (66 %) fueron recuperados en nuestras detecciones. Así, nuestras detecciones incluyen un total de 604 *nuevos* candidatos cinemáticos y fotométricos a miembros, lo que equivale al 49 % de nuestras detecciones y a un aumento del 64 % de la muestra disponible en S_{lit} . Las detecciones que reportamos poseen tasas promedio de recuperación y pureza respecto de S_{lit} del 61 % y 67 % respectivamente.

Parte de los miembros de S_{lit} que no recuperamos y de los nuevos candidatos cinemáticos que aportamos, se explicarían por una combinación de la ausencia de velocidades radiales para la mayoría de las estrellas de GDR3, la cercanía de los NYMGs y la presencia en el entorno solar de grupos móviles más viejos y/o más masivos que los NYMGs. Según lo discutido en la Sección 2.2.2, la falta de velocidades radiales y la cercanía de los NYMGs nos obliga a emplear velocidades tangenciales en el LSR en vez de las velocidades cartesianas, lo cual es una mejora notoria respecto del uso de movimientos propios. Sin embargo, la cercanía de los NYMGs tiende a aumentar la regiones que ocupan los miembros de un grupo en la esfera celeste y esto aumenta la dispersión de las velocidades tangenciales. Esto puede producir que grupos coherentes en el espacio de fase tiendan a confundirse con la población de campo en el EFR cuando disminuye su distancia al Sol. Por la misma razón, es de esperar que algunos miembros cinemáticos detectados mediante velocidades cartesianas no sean detectados una vez que nos limitamos al EFR. Esto explicaría por qué alcanzamos menores niveles de recuperación para los grupos más cercanos ($\bar{f}_r = 0,55$ para $d < 100$ pc) que para los más alejados ($\bar{f}_r = 0,70$ para $d > 100$ pc).

Análogamente, la falta de la VR induce una degeneración en la información cinemática de todos los objetos de S_{NYMG} tal que algunos de los nuevos candidatos cinemáticos pudieran ser en realidad contaminantes del campo con velocidades tangenciales similares a la de los NYMGs pero con velocidades

radiales diferentes. Estos contaminantes pudieran tanto pertenecer al campo del disco galáctico como a grupos móviles o regiones de formación estelar que no son representados por nuestro modelo del campo, ya que GOG no los reproduce. Si bien esta contaminación puede estar presente en todo el DCMA, debería ser más importante en el locus de la SP.

Para estimar que tanto la falta de VR y la cercanía de los NYMGs pueden haber afectado la contaminación de los grupos detectados, realizamos el análisis de la pureza cinemática del apéndice A.4. De este análisis, estimamos que entre 105 y 287 de los 604 nuevos candidatos (entre 17 % y 48 %) son estrellas contaminantes con velocidades tangenciales similares a la de los grupos detectados pero que probablemente tengan VR diferentes.

Podemos entonces concluir que nuestro método ha probado ser eficiente para la detección de NYMGs pero que su eficiencia es proporcional a la distancia al grupo en el espacio de posiciones.

Ahora bien, es importante subrayar que todo el análisis sobre recuperación y pureza que hemos realizado se basa en la muestra S_{lit} . Si bien varios miembros de S_{lit} no recuperados en nuestras detecciones fueron incluidos en S_{lit} a partir de identificaciones en el espacio de fase (X, Y, Z, U, V, W) donde no existen degeneraciones cinemáticas, la amplia mayoría en realidad no poseen medidas de velocidad radial pues sus velocidades radiales fueron inferidas por el algoritmo BANYAN (Gagné & Faherty, 2018) para luego calcular sus probabilidades de membresía. Adicionalmente, algunos de los miembros cinemáticos de los grupos de S_{lit} fueron identificados por Torres et al. (2001) y Kerr et al. (2021), entre varios otros autores, haciendo uso de movimientos propios en vez de las velocidades tangenciales en el LSR. Así, algunos de los miembros cinemáticos incluidos en S_{lit} pudieran ser contaminantes. Como nos enfocamos en maximizar la recuperación respecto de la literatura, los contaminantes pudieran estar sesgando nuestro análisis pero siempre desfavoreciéndolo.

5.3. Las IMFs de los NYMGs

Hemos inferido las IMFs de los grupos considerados en el rango de masas $0,01 \leq m/M_{\odot} \leq 5$ de manera independiente para los miembros reportados en S_{lit} y para las detecciones resultantes de nuestro trabajo. Estas son las *primeras estimaciones* de IMFs de los NYMGs realizadas a la fecha, con la única

excepción de la IMF de TW-Hya inferida por [Gagné et al. \(2017\)](#).

Encontramos que en el rango de bajas masas ($m < 1 M_{\odot}$) las IMFs obtenidas a partir de nuestras detecciones son en promedio morfológicamente consistentes con las obtenidas a partir de S_{lit} . Esta similitud morfológica combinada con las tasas de recuperación obtenidas apoyan la idea de que los miembros de S_{lit} no identificados son objetos dispersos en el espacio de velocidades tangenciales debido a la cercanía de los grupos. En este mismo rango de masas, los valores obtenidos para m_c y σ son en promedio consistentes con los de la población de campo del disco galáctico reportados por [Chabrier \(2003\)](#). Además, encontramos que el promedio de σ es consistente con el valor obtenido por [Suárez et al. \(2019\)](#) en la agrupación joven (7 Maños) 25-Ori, aunque el promedio de m_c resulta menor pero consistente dentro de las incertidumbres a 1σ .

Para masas mayores ($m > 1 M_{\odot}$), la mayoría de las IMFs inferidas a partir de nuestras detecciones muestran valores de α que son consistentes con los valores de [Suárez et al. \(2019\)](#) para 25 Ori y [Salpeter \(1955\)](#) para el disco galáctico. No obstante, la mayoría de las IMFs inferidas a partir de S_{lit} , presentan pendientes más planas que [Salpeter \(1955\)](#) y [Suárez et al. \(2019\)](#). Esta diferencia pudiera ser explicada por el aumento de candidatas en el bin de masa alrededor de $1 M_{\odot}$ debido a una potencial contaminación en la muestra de nuestras detecciones.

La incertidumbre en nuestras estimaciones de los parámetros de la IMF tanto en la región de baja masa como la de alta masa están dominadas por una combinación de errores de conteo (principalmente en los extremos de alta y baja masa), incompletitud y contaminación en diferentes rangos de masas. Como explicamos en la Sección 4.1.2, la IMF normalizada media de todas las poblaciones nos permite, por un lado, reducir el impacto de los errores de conteo y, por otro, resaltar las incompletitudes y contaminaciones más frecuentes si asumimos que todas provienen de una misma distribución.

Encontramos que la IMF normalizada promedio presenta un valor m_c consistente con [Chabrier \(2003\)](#), un valor σ menor al de [Chabrier \(2003\)](#) y una pendiente α más empinada que la estimada por [Salpeter \(1955\)](#). Ahora, si estos NYMGs son en realidad remanentes de poblaciones estelares que se formaron a partir de una IMF común y se encuentran actualmente en proceso de dispersión, entonces sus distribuciones de masa inferidas solo pueden diferir entre sí y respecto de sus IMFs originales por efecto de la pérdida cinemática de miembros (Sección 1.1.2) y de los sesgos observacionales y metodológicos presentes en la detección de los grupos y la inferencia de sus IMFs. Luego, la IMF normalizada media nos permite estudiar

las contaminaciones e incompletitudes generadas por la dispersión cinemática y los sesgos observacionales y metodológicos.

Concluimos que el valor de σ obtenido pudiera ser un indicador de una falta sistemática de objetos de baja masa en nuestras detecciones y que el valor de α puede ser explicado por una falta de objetos de alta masa o por un exceso de objetos entorno a la masa solar. Es probable que la falta de objeto de baja masa se deba a que la máscara $\sigma_{\varpi}/\varpi < 0,1$ descarta más objetos tenues que brillantes. En menor medida, la ausencia de algunos objetos de baja masa puede ser consecuencia de la extinción que en estas poblaciones es baja $A_V \leq 0,2$ mag (e.g [Gagné et al., 2021](#)). Otra posible explicación es que los objetos de menor masa presentan una mayor dispersión cinemática en los datos en comparación a los de masas mayores debido al aumento de los errores para objetos con $G > 16$ mag y/o una mayor dispersión cinemática intrínseca de los objetos menos masivos, la cual pudiera haber producido una segregación de masas. Los resultados de este trabajo nos permiten cuantificar el impacto relativo de estos fenómenos en las IMFs inferidas, y es un trabajo importante que se encuentra en desarrollo.

En cuanto al empinamiento de la IMF en la región de altas masas, la explicación más plausible es que se deba al exceso de objetos observados entorno a la masa solar. Este exceso podría ocurrir por la contaminación de estrellas viejas del campo que se hace máxima alrededor de lo que hemos llamado *Turn-On* (Sección 4.1.2).

5.4. Conclusiones científicas generales

De la discusión de las Secciones 5.2 y 5.3, podemos concluir que nuestro trabajo representa un avance importante en el entendimiento físico y descripción observacional de los NYMGs. Más específicamente, el análisis de las IMFs individuales y de la IMF normalizada media nos ha permitido sugerir cuáles son las posibles fuentes de contaminación e incompletitud y en que rangos de masas ocurren, tanto en nuestras detecciones como en S_{lit} . Adicionalmente, destacamos que considerando las potenciales incompletitudes y contaminaciones, nuestras estimaciones de los parámetros de la IMF presentan en promedio un muy buen acuerdo con resultados para otras poblaciones ([Chabrier, 2003](#); [Suárez et al., 2019](#); [Salpeter, 1955](#)), lo cual apoya la hipótesis de que los NYMGs son remanentes de poblaciones estelares en proceso de dispersión. Un análisis dinámico apoyado por simulaciones de N-cuerpos es imprescindible para concluir sobre este

proceso.

5.5. El futuro del trabajo

Las posibles fuentes de contaminación e incompletitud discutidas motivan la implementación de las siguientes estrategias en los siguientes trabajos en desarrollo y futuros:

- Para estudiar el impacto de los errores cinemáticos en la dispersión en los datos de los grupos, estamos desarrollando herramientas que siguen el método que presentamos en el Apéndice A.5 que también nos permitirán evaluar los niveles de contaminación debidos a la población más vieja del campo.
- A partir de las herramientas construidas en este trabajo para simular grupos de estrellas y de un modelo numérico del potencial galáctico, podremos simular la evolución dinámica y cinemática de varios grupos sintéticos generados de diversas condiciones iniciales para inferir a nivel estadístico hasta que punto la forma de las IMFs inferidas pudieran ser explicadas por procesos de dispersión.

Luego, las herramientas desarrolladas y en desarrollo podrán ser utilizadas para inferir la IMF de otras poblaciones más lejanas. Esperamos además que podremos realizar estudios cinemáticos de los NYMGs y otras poblaciones con una mejor astrometría una vez que se publique el catálogo de Gaia DR4 en 2026. Además, se deberá estudiar si la combinación de las futuras velocidades radiales de este último catálogo con las velocidades radiales del Milky Way Mapper ([Almeida et al., 2023](#)) permite la construcción de un entorno solar con velocidades radiales suficientemente completo para mejorar la búsqueda de miembros en el espacio de fase.

Los resultados presentados en esta tesis forman parte de una publicación en preparación ([Bertolotto-Stefanelli, Downes & Suárez, 2024](#)).

Apéndice A

Apéndices

A.1. Estimación del parámetro ϵ

Dado que estamos buscando NYMGs tanto en el espacio de posición como en el espacio de velocidad tangencial en LSR, consideramos dos parámetros ϵ diferentes: ϵ_r para el espacio de posición y ϵ_v para el espacio de velocidad. Entonces, las dos condiciones de que dos estrellas con vectores en el EFR $\vec{X}_0 = (\vec{r}_0, \vec{v}_0) = (X_0, Y_0, Z_0, v_{\alpha,0}^{LSR}, v_{\delta,0}^{LSR})$ y $\vec{X}_1 = (\vec{r}_1, \vec{v}_1) = (X_1, Y_1, Z_1, v_{\alpha,1}^{LSR}, v_{\delta,1}^{LSR})$ deben satisfacer para ser clasificadas como vecinas por el DBSCAN son:

$$\Delta v_{0,1}^2 = \|\vec{v}_1 - \vec{v}_0\|^2 \leq \epsilon_v^2 \quad (\text{A.1.1})$$

$$\Delta r_{0,1}^2 = \|\vec{r}_1 - \vec{r}_0\|^2 \leq \epsilon_r^2 \quad (\text{A.1.2})$$

Un posible enfoque es aplicar el DBSCAN por separado en el espacio de velocidad y en el espacio de posición. Este enfoque es útil para reducir la densidad del campo, aprovechando del hecho de que los NYMGs forman cúmulos más densos en el espacio de velocidad que en el espacio de posición. Además, los

resultados serían más representativos de las densidades de los grupos detectados en los dos espacios, ya que los cúmulos detectados deben tener al menos una estrella en una densidad local mayor que $N_{min}/(\frac{4}{3}\pi\varepsilon_r^3)$ y $N_{min}/(\pi\varepsilon_v^2)$ para cumplir con las Ecuaciones A.1.1 y A.1.2. Sin embargo, aplicar el DBSCAN por separado es computacionalmente costoso.

Afortunadamente, las condiciones dadas por las Ecuaciones A.1.1 y A.1.2 pueden aproximarse a una sola condición al observar que ambas ecuaciones implican que:

$$\sqrt{\left(\frac{\Delta r_{0,1}}{\varepsilon_r}\right)^2 + \left(\frac{\Delta v_{0,1}}{\varepsilon_v}\right)^2} \leq \sqrt{2} \tag{A.1.3}$$

Notemos sin embargo que la Ecuación A.1.3 no implica ni la Ecuación A.1.1 ni la Ecuación A.1.2. Para obtener una mejor aproximación de las condiciones dadas por las ecuaciones A.1.1 y A.1.2, podemos entonces realizar la búsqueda en todo el EFR considerando una nueva condición a partir de la Ecuación A.1.3 y definiendo los factores de escala $(\varepsilon_r, \varepsilon_v) = \alpha(\varepsilon_r, \varepsilon_v)$ donde α es un número real que maximiza la similitud de este enfoque con el primero. Entonces, si consideramos las posiciones de las dos estrellas \mathcal{X}_0 y \mathcal{X}_1 en el EFRE escaladas por ε_r y ε_v , podemos aproximar las Ecuaciones A.1.1 y A.1.2 a una sola condición:

$$d(\mathcal{X}_0, \mathcal{X}_1) \leq \alpha\sqrt{2} \tag{A.1.4}$$

Donde $d(\cdot, \cdot)$ es la función de distancia euclídea. Esto significa que una buena aproximación de aplicar por separado el DBSCAN al espacio de posición y velocidad consiste en escalar estos dos espacios y luego aplicar el DBSCAN al EFRE con $\varepsilon = \alpha\sqrt{2}$ utilizando la función de distancia euclídea.

Para encontrar el valor de α , consideramos dos estrellas con posiciones en el espacio de fases X_0 y X_1 . Según el primer enfoque, X_1 es un vecino de X_0 si $\Delta r_{0,1} \leq \varepsilon_r$ y $\Delta v_{0,1} \leq \varepsilon_v$. Geométricamente, esto significa que la velocidad de X_1 debe estar contenida dentro de una esfera \mathcal{S}_v de radio ε_v centrada en la velocidad de X_0 y, de manera equivalente, su posición dentro de una esfera \mathcal{S}_r . Por lo tanto, la región en el espacio de fases en la que X_1 debe estar contenida para ser un vecino de X_0 corresponde a la unión

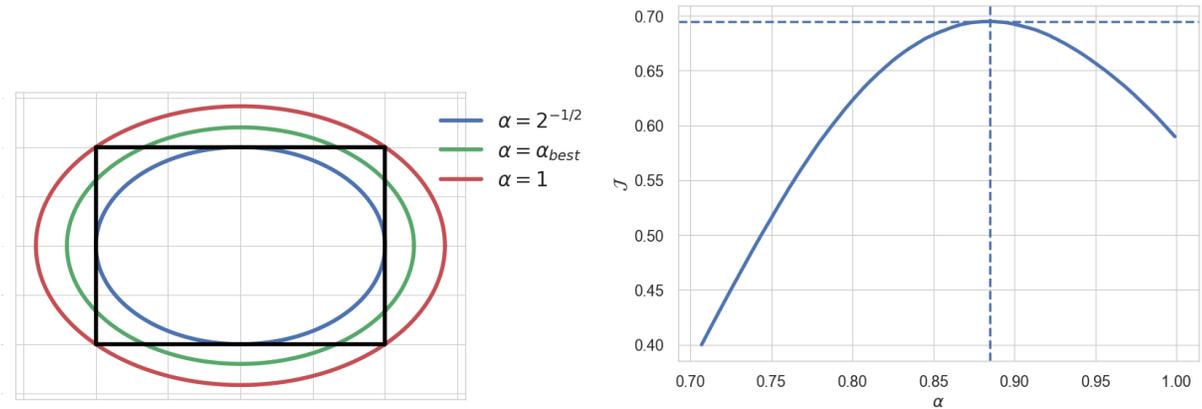


Figura A.1: *Izquierda:* Representación esquemática de U_{rv} (caja negra) y E_{rv} (elipses coloreadas) para diferentes valores de α en un caso bidimensional, donde el valor de α para la elipse verde corresponde al mejor valor de α en el caso bidimensional. *Derecha:* resultados numéricos de $J(\alpha)$ para la muestra S_{NYMG} (líneas sólidas). Las líneas punteadas indican tanto el valor máximo de J obtenido como el correspondiente valor de α .

$U_{rv} = \mathcal{S}_r \cup \mathcal{S}_v$. Como muestra la Figura A.1, es fácil ver que en el caso bidimensional, esta región es un rectángulo centrado en X_0 cuyos lados están alineados con los ejes cartesianos y con longitudes $2\varepsilon_r$ y $2\varepsilon_v$.

Ahora, según el segundo enfoque, X_1 es un vecino de X_0 si $d(\mathcal{X}_0, \mathcal{X}_1) \leq \alpha\sqrt{2}$, lo que implica que X_1 debe estar contenido en un elipsoide de 5 dimensiones E_{rv} cuyos ejes principales están alineados con los ejes cartesianos y están definidos por ε_r y ε_v .

El criterio que elegimos para maximizar la similitud entre ambos métodos es encontrar el valor de α que minimiza la diferencia relativa en tamaño entre $U_{rv} \cap E_{rv}$ y $U_{rv} \cup E_{rv}$, lo cual es equivalente a encontrar el valor de α que maximiza la similitud de Jaccard de las dos regiones. La similitud de Jaccard entre dos conjuntos, como se explica en [Ratzenböck et al. \(2022\)](#), se define como la razón entre el tamaño de su intersección sobre el tamaño de su unión. Si V_i es el volumen de $U_{rv} \cap E_{rv}$ y V_u el volumen de $U_{rv} \cup E_{rv}$, entonces buscamos el valor de α que maximice la similitud de Jaccard $J = V_i/V_u$.

A partir de la Figura A.1, es fácil notar que el valor que buscamos debe satisfacer $2^{-1/2} \leq \alpha \leq 1$. Utilizamos un método de Monte Carlo para calcular tanto V_i como V_u y por lo tanto J , para diferentes valores de α . La Figura A.1 muestra el gráfico de $J(\alpha)$ y el valor que buscamos es $\alpha = 0,847$ para $J = 0,654$.

A.2. Los NYMGs detectados

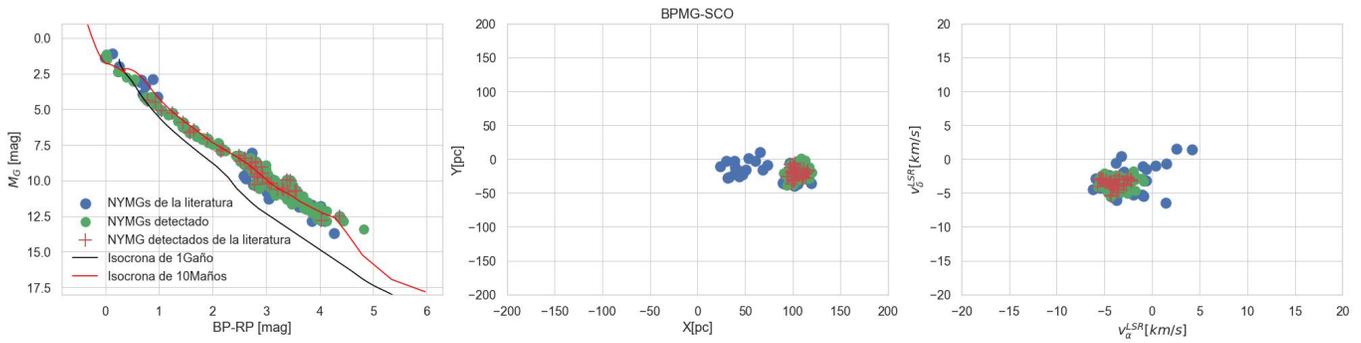


Figura A.2: Distribuciones análogas a las de la Figura 3.3 para la agrupación BPMG-SCO.

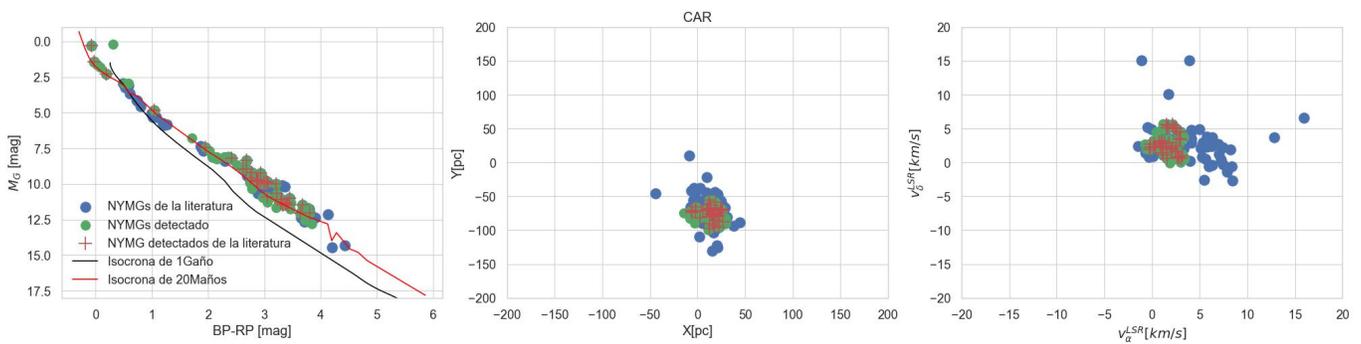


Figura A.3: Distribuciones análogas a las de la Figura 3.3 para la agrupación CAR.

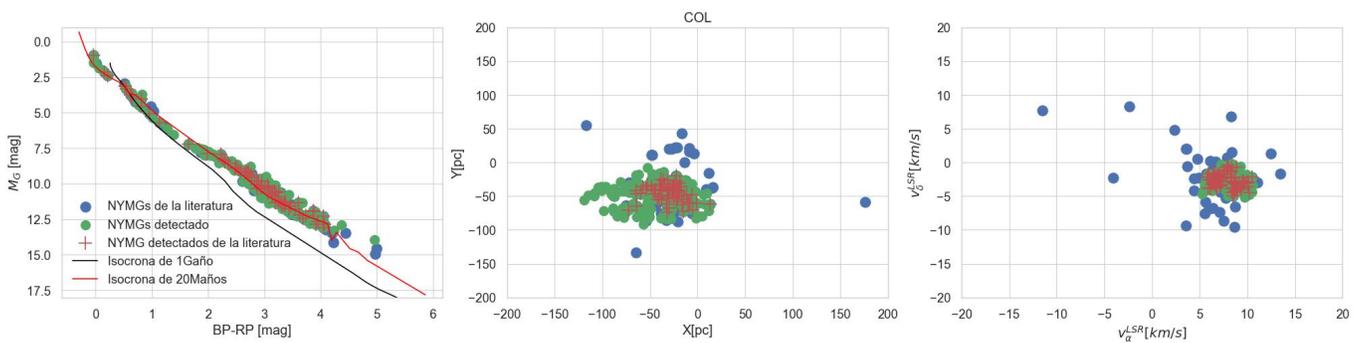


Figura A.4: Distribuciones análogas a las de la Figura 3.3 para la agrupación COL.

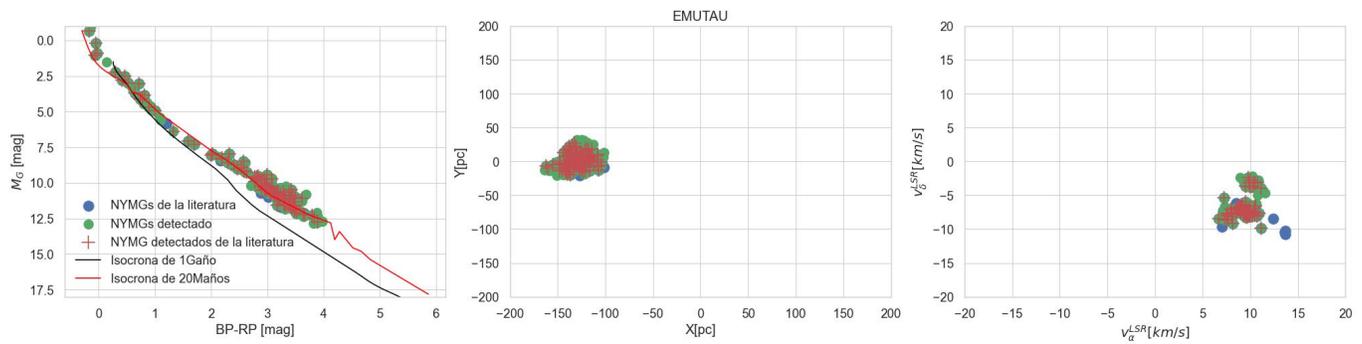


Figura A.5: Distribuciones análogas a las de la Figura 3.3 para la agrupación EMUTAU.

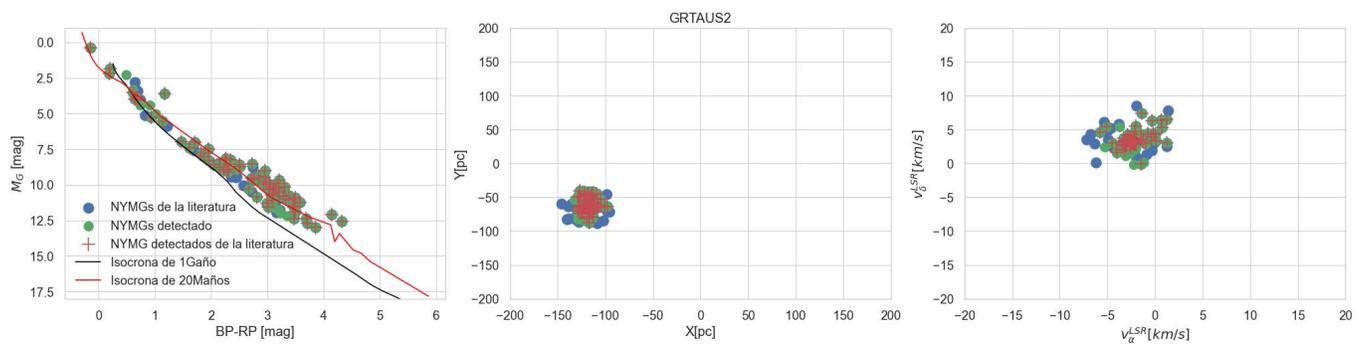


Figura A.6: Distribuciones análogas a las de la Figura 3.3 para la agrupación GRTAUS2.

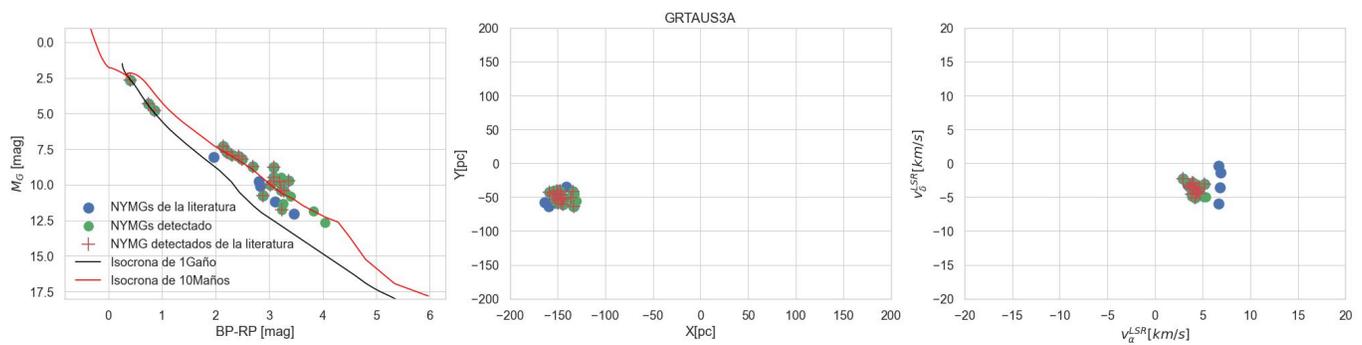


Figura A.7: Distribuciones análogas a las de la Figura 3.3 para la agrupación GRTAUS3A.

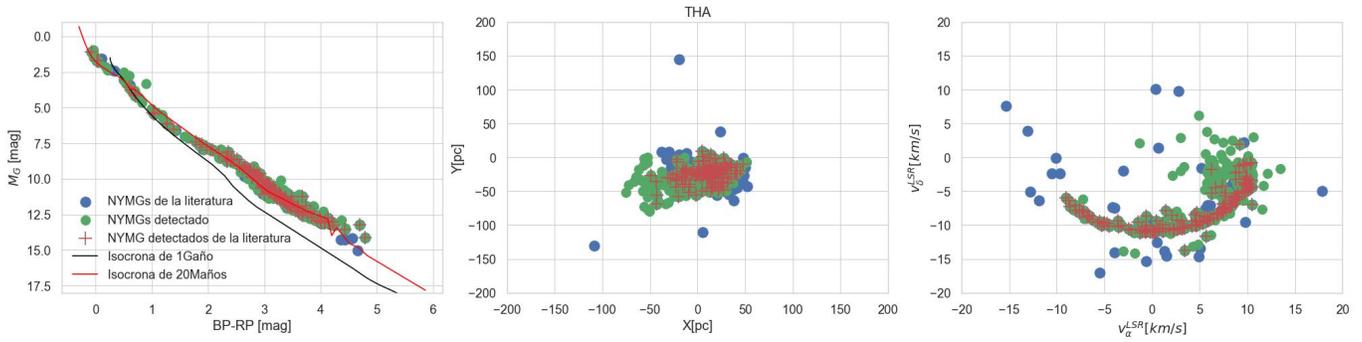


Figura A.8: Distribuciones análogas a las de la Figura 3.3 para la agrupación THA.

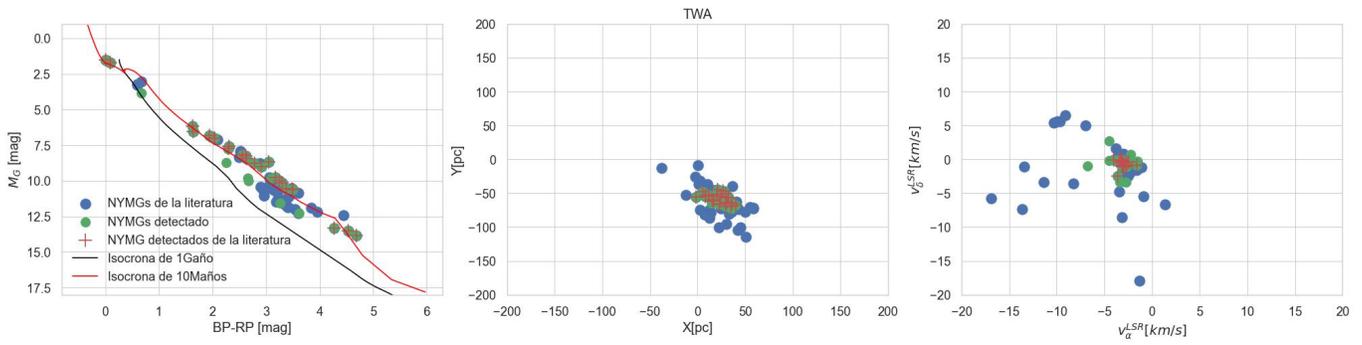


Figura A.9: Distribuciones análogas a las de la Figura 3.3 para la agrupación TW-Hya.

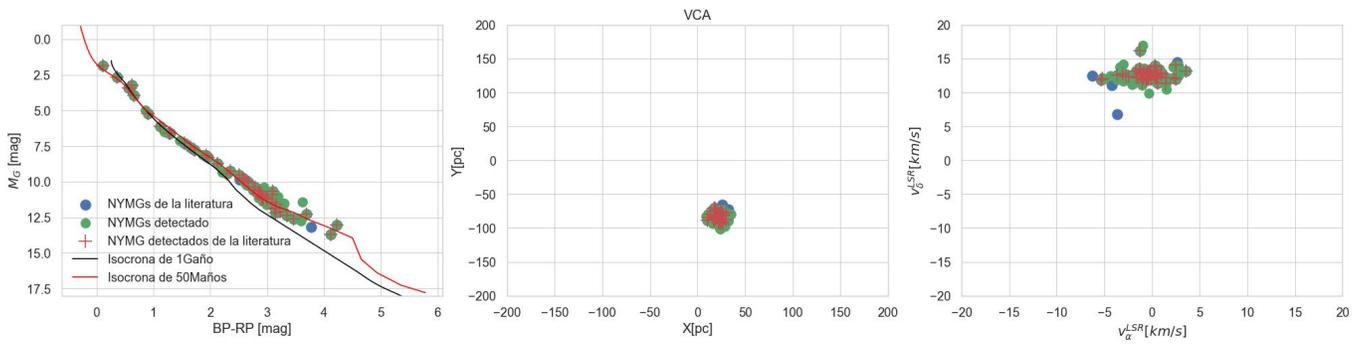


Figura A.10: Distribuciones análogas a las de la Figura 3.3 para la agrupación VCA.

A.3. Las IMFs de los NYMGs detectados

Las siguientes Figuras muestran las restantes IMFs obtenidas además de TW-Hya t β -Pictoris.

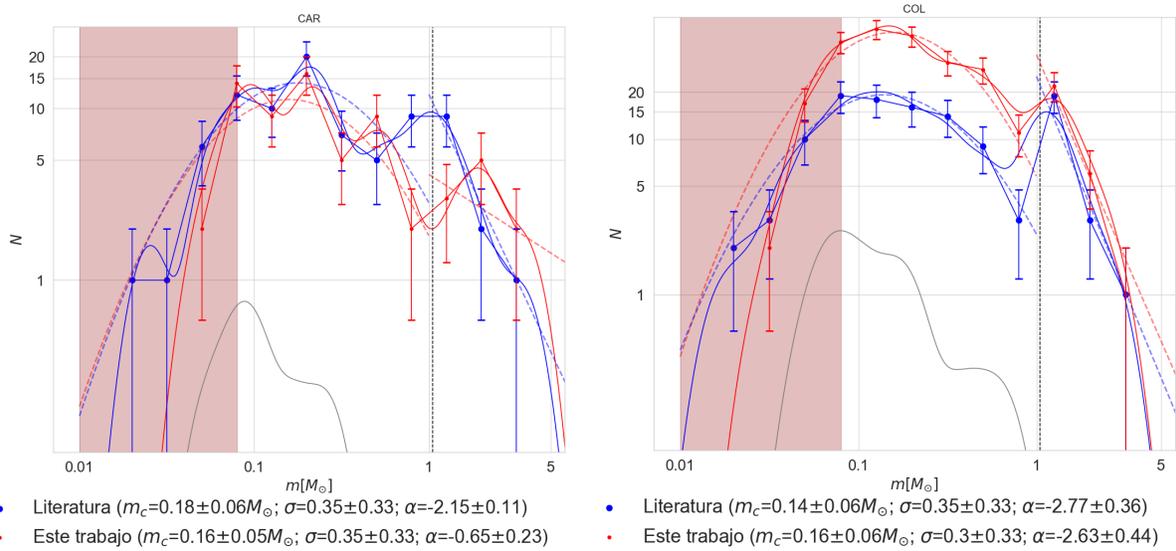


Figura A.11: IMFs inferidas análogas a las de la Figura 4.3 para la agrupación CAR (izquierda) y COL (derecha).

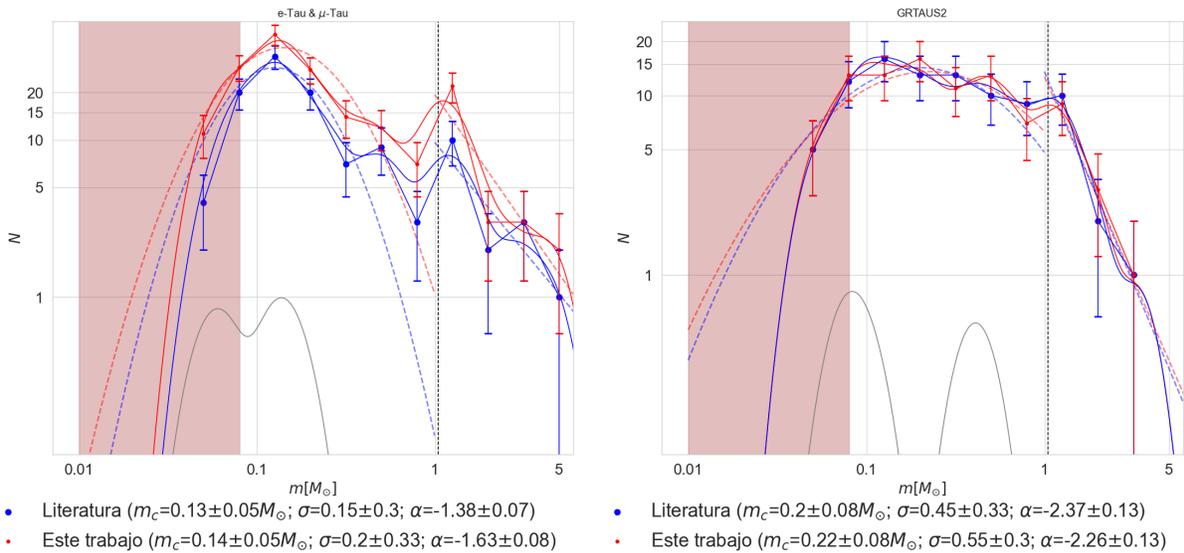


Figura A.12: IMFs inferidas análogas a las de la Figura 4.3 para la agrupación EMUTAU (izquierda) y GRTAUS2 (derecha).

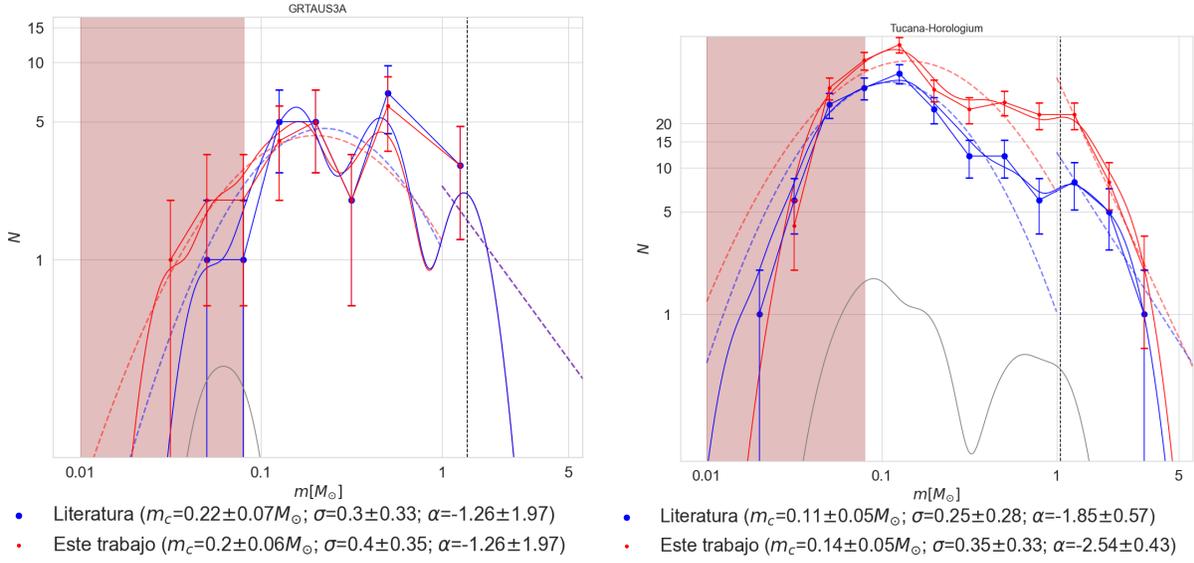


Figura A.13: IMFs inferidas análogas a las de la Figura 4.3 para la agrupación GRTAUS3A (izquierda) y THA (derecha).

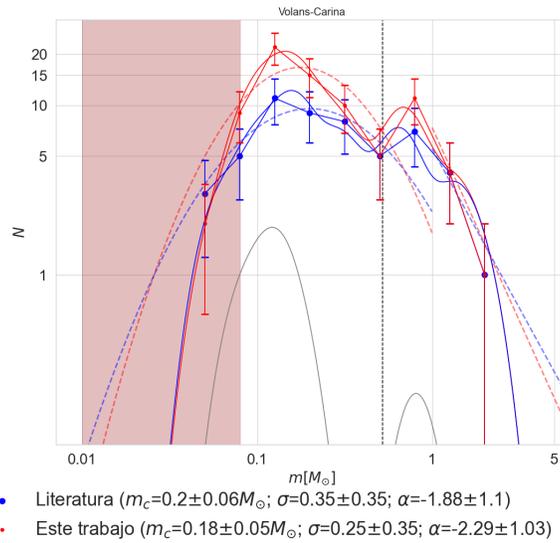


Figura A.14: IMFs inferidas análogas a las de la Figura 4.3 para la agrupación VCA.

A.4. Estimación cinemática de la pureza de las detecciones

Para estimar el nivel de contaminación de los candidatos a miembros no reportados en la literatura dado por la degeneración generada por la falta de VR, consideramos para cada grupo obtenido de nuestras detecciones el subconjunto de miembros que tienen velocidades radiales medidas en GDR3. De dicho subconjunto, calculamos la matriz de covarianza en velocidades cartesianas galácticas (U, V, W) así como la velocidad media. Luego, calculamos la distancia de Mahalanobis de los miembros del grupo de donde consideramos dos sub-muestras: (i) los candidatos con VR que se encuentran a una distancia de Mahalanobis de 1 (equivalente a un σ) y (ii) los candidatos con VR que se encuentra a una distancia de Mahalanobis de 4 (equivalente a 2σ). De estas dos muestras definimos las purezas cinemáticas a un σ y a 2σ como $f_{v,\sigma} = N_{\sigma,VR}/N_{VR}$ y $f_{v,2\sigma} = N_{2\sigma,VR}/N_{VR}$ respectivamente donde N_{VR} es el número de candidatos del grupo con VR y $N_{\sigma,VR}$ y $N_{2\sigma,VR}$ son los números de las submuestras definidas en (i) y (ii) respectivamente. A modo de ejemplo, la Figura A.4 muestra las distribuciones de velocidades cartesianas las diferentes submuestras para el caso particular de BPMG.

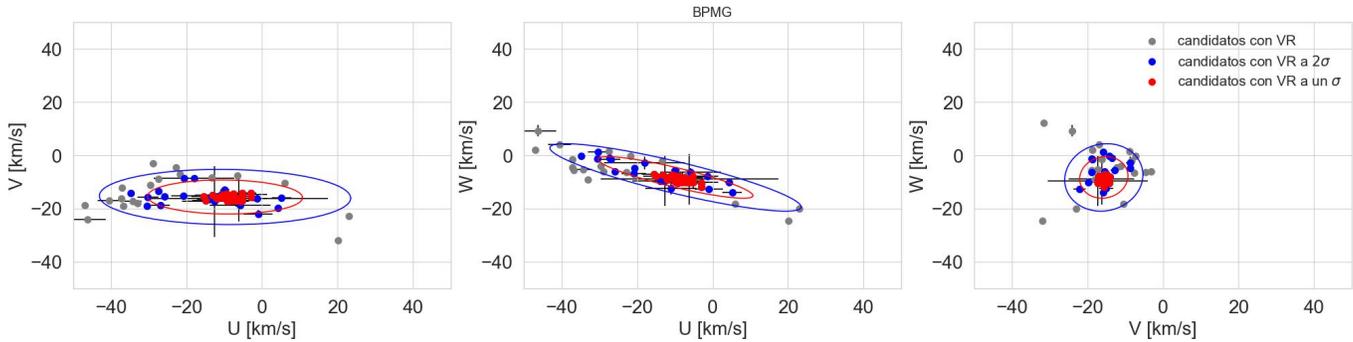


Figura A.15: Distribución de las velocidades cartesianas de los candidatos cinemáticos de nuestras detecciones del grupo BPMG con VR (puntos grises). Los puntos azules y rojos corresponden a los candidatos a menos de un σ y 2σ . Las elipses azules y rojas indican los límites a un σ y 2σ respectivamente. Las barras de errores negras corresponden a los errores propagados de los observables a las velocidades cartesianas.

Una vez calculadas las purezas cinemáticas para cada grupo y asumiendo que estas fracciones no cambiarían si se pudiera hacer este ejercicio con *todos* los candidatos, podemos estimar el número de miembros a un σ y 2σ como $N_{\sigma} = f_{v,\sigma}N$ y $N_{2\sigma} = f_{v,2\sigma}N$ respectivamente. La Tabla A.1 muestra los resultados generales de la estimación de la pureza cinemática de cada grupo. De este análisis, estimamos que entre 105 y 287 de los 604 nuevos candidatos (entre 17 % y 48 %) sean en realidad contaminantes del

NYMG	N	N_{VR}	N_r	N_{nuevos}	N_σ	$N_{2\sigma}$
TWA	29	20	24	5	1	3
GRTAUS3A	23	11	19	4	1	3
BPMG-SCO	171	88	41	130	57	112
BPMG	124	84	67	57	31	44
GRTAUS2	83	41	68	15	6	9
COL	226	133	76	150	92	131
CAR	60	38	36	24	12	18
THA	282	165	141	141	79	118
EMUTAU	167	49	105	62	29	50
VCA	64	35	48	16	9	11
Total	1229	664	625	604	317	499

Tabla A.1: Estadísticas generales de la estimación de la pureza cinemática. N: número de candidatos cinemáticos. N_{VR} : número de candidatos con VR. N_r : número de candidatos de S_{lit} recuperados. N_{nuevos} : número de candidatos no reportados en la literatura. N_σ : número de candidatos estimado a un σ . $N_{2\sigma}$: número de candidatos estimado a 2σ .

campo con velocidades tangenciales similares a la de los NYMGs detectados.

A.5. Desarrollo de un método para evaluar la pureza y completitud de la IMF

Para obtener una inferencia robusta de la IMF de una población estelar, es necesario estimar y corregir la incompletitud y contaminación presentes en la muestra de cada grupo. Al construir la muestra S_{comp} , se estimó y corrigió gran parte de las incompletitudes de origen observacional. Sin embargo, para lograr una estimación aún más completa de la contaminación e incompletitud generamos una muestra sintética constituida por NYMGs sintéticos y una población sintética de campo sobre la cual repetimos la detección de los NYMGs reales ya detectados. A partir de los resultados estimamos las fracciones esperadas de contaminación (fracción de objetos de campo sintéticos que fueron detectados como miembros de los NYMGs sintéticos) y la fracción de incompletitud (fracción de miembros que no fueron detectados como tales), ambos, como función de la masa estelar.

En cuanto a la completitud, vimos en la Figura 2.5 que los errores cinemáticos son mayores para fuentes menos brillantes que $G \sim 16$ mag, lo cual implica que los errores deberían tender a dispersar más los objetos de baja masa en el EFR que los de mayor masa. Esto significa que es menos probable que detectemos un objeto de baja masa que uno de alta masa, lo cual para la IMF se traduce en una mayor incompletitud en el rango de baja masa que en el de alta masa. El razonamiento anterior inspiró el siguiente procedimiento para estimar la incompletitud inducida por los errores cinemáticos: (i) modelamos la distribución en el EFR de las estrellas del grupo con $G < 16$ mag con un KDE, (ii) generamos una población sintética con posiciones en el EFR que siguen la distribución modelada por el KDE y con una distribución de masas uniforme en el logaritmo de la masa donde el número total de objetos se normaliza al número de fuentes del NYMG detectado con $G < 16$ mag, (iii) se asignan a las estrellas sintéticas errores cinemáticos generados estocásticamente de las distribuciones de errores de S_{NYMG} , (iv) se aplica el algoritmo de detección a la población sintética utilizando los hiperparámetros utilizados para detectar el NYMG estudiado, (v) se mide la fracción de miembros recuperados como función de la masa y (vi) repetimos los pasos anteriores diez veces de donde se calcula la recuperación promedia como función de la masa.

En cuanto a la contaminación, para que una estrella vieja del campo sea clasificada erróneamente como miembro por el algoritmo de detección, se debe cumplir que su posición en el DCMA coincida con la isocrona del grupo y que su cinemática se lo suficientemente parecida a la del grupo para que el DBSCAN la clasifique como un punto vecino de uno de los puntos núcleo del grupo. Basados en esta

observación, desarrollamos el siguiente procedimiento para estimar el nivel de contaminación del grupo: *(i)* generamos un campo sintético siguiendo el procedimiento descrito en la Sección 3.2.1, *(ii)* consideramos únicamente las fuentes sintéticas generadas que son vecinas de los puntos núcleos del grupo estudiado y las clasificamos como miembros, *(iii)* medimos la fracción de estrellas sintéticas clasificadas de esta manera como función de la masa y *(iv)* repetimos este procedimiento con diez campos sintéticos de donde calculamos la contaminación promedio como función de la masa.

Siglas

ε distancia máxima entre dos puntos para considerarlos vecinos

ε_r factor de escala para los espacios de posiciones cartesianas

ε_v factor de escala para el espacio de velocidades tangenciales en el LSR

N_{\min} número mínimo de vecinos necesarios para que un punto sea etiquetado como punto núcleo

S_{aux} Subconjunto Auxiliar

S_{comp} S_{NYMG} Completado

S_{lit} Muestra de Candidatos de la Literatura

S_{NYMG} Subconjunto de los NYMG

S_{real} S_{NYMG} Real

DBSCAN *Density-Based Spatial Clustering of Applications with Noise* de [Ester et al. \(1996\)](#)

DCMA Diagrama de Color-Magnitud Absoluta

EFR Espacio Fase Restringido

EFRE Espacio Fase Restringido Escalado

GDR3 *Gaia Data Release 3*; ([Gaia Collaboration et al., 2018](#))

GOG *Gaia Object Generator* ([Antiche et al., 2014](#))

GSF *Gaiainlimited Selection Functions*

IMF *Initial Mass Function*

KDE *Kernel Density Estimator*

LSR *Local Standard of Rest*

MIE Muestra de Interés Estudiada

MIO Muestra de Interés Observada

MIV Muestra de Interés Verdadera

MM Máscaras Mixtas

MOCA *Montreal Open Clusters and Associations*

NF nivel de fragmentación

NYMG *Nearby Young Moving Groups*

PSP Pre-Secuencia Principal

RMSE *Root Mean Square Error*

RUWE *Re-normalized Unit Weighted Error*

SNR Relación Señal a Ruido

SP Secuencia Principal

VR Velocidad Radial

Bibliografía

Almeida, A., et al. 2023, *ApJS*, 267, 44

Antiche, E., Masana, E., Julbe, F., & Borrachero, R. 2014, in *EAS Publications Series*, Vol. 67-68, *EAS Publications Series*, 355–355

Astropy Collaboration et al. 2022, *ApJ*, 935, 167

Bailer-Jones, C. A. L. 2015, *Publications of the Astronomical Society of the Pacific*, 127, 994

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, *The Astronomical Journal*, 161, 147

Baraffe, I., Homeier, D., Allard, F., & Chabrier, G. 2015, *Astronomy and Astrophysics*, 577, A42

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, *ARA&A*, 48, 339

Bayo, A., et al. 2011, *Astronomy and Astrophysics*, 536, A63

Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, in *Advances in Knowledge Discovery and Data Mining*, ed. J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Berlin, Heidelberg: Springer Berlin Heidelberg), 160–172

Cantat-Gaudin, T., et al. 2023, *A&A*, 669, A55

Castro-Ginard, A., et al. 2023, *A&A*, 677, A37

Chabrier, G. 2003, *PASP*, 115, 763

Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. H. 2005, *ArXiv Astrophysics e-prints*

- Cruz, K. L., Kirkpatrick, J. D., & Burgasser, A. J. 2009, *AJ*, 137, 3345
- da Silva, L., Torres, C. A. O., de La Reza, R., Quast, G. R., Melo, C. H. F., & Sterzik, M. F. 2009, *Astronomy and Astrophysics*, 508, 833
- Dawson, P., Scholz, A., Ray, T. P., Marsh, K. A., Wood, K., Natta, A., Padgett, D., & Ressler, M. E. 2013, *MNRAS*, 429, 903
- De Marchi, G., Paresce, F., & Portegies Zwart, S. 2010, *ApJ*, 718, 105
- De Silva, G. M., D’Orazi, V., Melo, C., Torres, C. A. O., Gieles, M., Quast, G. R., & Sterzik, M. 2013, *Monthly Notices of the Royal Astronomical Society*, 431, 1005
- Dib, S. 2023, *ApJ*, 959, 88
- Ducourant, C., Teixeira, R., Galli, P. A. B., Le Campion, J. F., Krone-Martins, A., Zuckerman, B., Chauvin, G., & Song, I. 2014, *Astronomy and Astrophysics*, 563, A121
- Elliott, P., Bayo, A., Melo, C. H. F., Torres, C. A. O., Sterzik, M., & Quast, G. R. 2014, *Astronomy and Astrophysics*, 568, A26
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD’96 (AAAI Press)*, 226–231
- Fabircius, C., et al. 2021, *A&A*, 649, A5
- Gagné, J., & Faherty, J. K. 2018, *ApJ*, 862, 138
- Gagné, J., Faherty, J. K., Moranta, L., & Popinchalk, M. 2021, *ApJ*, 915, L29
- Gagné, J., Lafrenière, D., Doyon, R., Malo, L., & Artigau, É. 2015, *ApJ*, 798, 73
- Gagné, J., et al. 2017, *The Astrophysical Journal Supplement Series*, 228, 18
- Gaia Collaboration et al. 2018, *A&A*, 616, A1
- . 2021, *A&A*, 649, A8
- . 2022, *Gaia Data Release 3: Summary of the content and survey properties*

- Galvez-Ortiz, M. C., et al. 2010, *Monthly Notices of the Royal Astronomical Society*, 409, 552
- . 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 3890
- Hennebelle, P., & Chabrier, G. 2008, *ApJ*, 684, 395
- Hopkins, A. M. 2018, *Publications of the Astronomical Society of Australia*, 35
- Kerr, R. M. P., Rizzuto, A. C., Kraus, A. L., & Offner, S. S. R. 2021, *ApJ*, 917, 23
- Kraus, A. L., Shkolnik, E. L., Allers, K. N., & Liu, M. C. 2014, *The Astronomical Journal*, 147, 146
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Lada, C. J., & Lada, E. A. 2003, *ARA&A*, 41, 57
- Lindgren, L., et al. 2021, *A&A*, 649, A2
- Makarov, V. V., & Urban, S. 2000, *Monthly Notices of the Royal Astronomical Society*, 317, 289
- Malo, L., Artigau, É., Doyon, R., Lafrenière, D., Albert, L., & Gagné, J. 2014, *ApJ*, 788, 81
- Malo, L., Doyon, R., Feiden, G. A., Albert, L., Lafrenière, D., Artigau, É., Gagné, J., & Riedel, A. 2014, *The Astrophysical Journal*, 792, 37
- Malo, L., Doyon, R., Lafrenière, D., Artigau, É., Gagné, J., Baron, F., & Riedel, A. 2013, *ApJ*, 762, 88
- Malzer, C., & Baum, M. 2020, in *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) (IEEE)*
- Marigo, P., et al. 2017, *ApJ*, 835, 77
- Miller, G. E., & Scalo, J. M. 1979, *ApJS*, 41, 513
- Moor, A., Szabó, G. M., Kiss, L. L., Kiss, C., Ábrahám, P., Szulágyi, J., Kóspál, Á., & Szalai, T. 2013, *Monthly Notices of the Royal Astronomical Society*, 1376
- Murphy, S. J., & Lawson, W. A. 2015, *MNRAS*, 447, 1267
- Nesmachnow, S., & Iturriaga, S. 2019, in *Supercomputing*, ed. M. Torres & J. Klapp (Cham: Springer International Publishing), 188–202

- Offner, S. S. R., Clark, P. C., Hennebelle, P., Bastian, N., Bate, M. R., Hopkins, P. F., Moraux, E., & Whitworth, A. P. 2014, *Protostars and Planets VI*, 53
- Pedregosa, F., et al. 2011, *Journal of Machine Learning Research*, 12, 2825
- Peña Ramírez, K., Béjar, V. J. S., Zapatero Osorio, M. R., Petr-Gotzens, M. G., & Martín, E. L. 2012, *The Astrophysical Journal*, 754, 30
- Phillips, M. W., et al. 2020, *Astronomy and Astrophysics*, 637, A38
- Ratzenböck, S., Großschedl, J. E., Möller, T., Alves, J., Bomze, I., & Meingast, S. 2022, *Significance Mode Analysis (SigMA) for hierarchical structures. An application to the Sco-Cen OB association*
- Riedel, A. R., et al. 2014, *The Astronomical Journal*, 147, 85
- Rix, H.-W., et al. 2021, *The Astronomical Journal*, 162, 142
- Robin, A. C., et al. 2012, *A&A*, 543, A100
- Salpeter, E. E. 1955, *ApJ*, 121, 161
- Scalo, J. M. 1986, *Fund. Cosmic Phys.*, 11, 1
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *Monthly Notices of the Royal Astronomical Society*, 403, 1829–1833
- Shkolnik, E. L., Anglada-Escudé, G., Liu, M. C., Bowler, B. P., Weinberger, A. J., Boss, A. P., Reid, I. N., & Tamura, M. 2012, *The Astrophysical Journal*, 758, 56
- Slesnick, C. L., Hillenbrand, L. A., & Carpenter, J. M. 2008, *ApJ*, 688, 377
- Suárez, G., Downes, J. J., Román-Zúñiga, C., Cerviño, M., Briceño, C., Petr-Gotzens, M. G., & Vivas, K. 2019, *MNRAS*, 486, 1718
- Suárez, G., et al. 2017, *AJ*, 154, 14
- Torres, C. A. O., Quast, G. R., da Silva, L., de La Reza, R., Melo, C. H. F., & Sterzik, M. 2006, *Astronomy and Astrophysics*, 460, 695

- Torres, C. A. O., Quast, G. R., de La Reza, R., da Silva, L., & Melo, C. H. F. 2001, in *Astronomical Society of the Pacific Conference Series*, Vol. 244, *Young Stars Near Earth: Progress and Prospects*, ed. R. Jayawardhana & T. Greene, 43
- Torres, C. A. O., Quast, G. R., Melo, C. H. F., & Sterzik, M. F. 2008, *Young Nearby Loose Associations* (Reipurth, B.), 757–+
- Torres, G., Neuhäuser, R., Latham, D. W., & Stefanik, R. P. 1999, in *American Astronomical Society Meeting Abstracts*, Vol. 195, *American Astronomical Society Meeting Abstracts*, 78.07
- Viana Almeida, P., Santos, N. C., Melo, C., Ammler-von Eiff, M., Torres, C. A. O., Quast, G. R., Gameiro, J. F., & Sterzik, M. 2009, *Astronomy and Astrophysics*, 501, 965
- Zuckerman, B., & Song, I. 2004, *Annual Review of Astronomy & Astrophysics*, 42, 685