



TESIS DOCTORAL

CARACTERIZACIÓN DE SISTEMAS
ESTELARES EN ESPACIOS DE
N-DIMENSIONES: SIMULACIONES Y
APLICACIÓN AL CATÁLOGO
ASTROMÉTRICO UCAC4.

LAURA MARÍA SAMPEDRO HERNÁNDEZ
Instituto de Astrofísica de Andalucía - CSIC

Tesis por memoria
presentada en la Universidad de Granada
para optar al grado de DOCTOR EN FÍSICA Y MATEMÁTICAS

Director de tesis:
EMILIO JAVIER ALFARO NAVARRO

Junio, 2016

A Alberto, a mi familia y a mis amigos.

No hay nada más grande en este Universo....

DERECHOS DE AUTOR

El doctorando Laura María Sampedro Hernández y el director de la tesis el Emilio Javier Alfaro Navarro, garantizamos al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de los autores al ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, a 25 de Abril de 2016

Director de la tesis

Doctorando

Fdo.: Emilio Javier Alfaro Navarro

Fdo.: Laura María Sampedro Hernández

D. Emilio Javier Alfaro Navarro, Científico Investigador del Instituto de Astrofísica de Andalucía, del Consejo Superior de Investigaciones Científicas (IAA-CSIC),

CERTIFICA:

que la presente tesis: ***“Caracterización de Cúmulos Estelares en Espacios de N-Dimensiones: Simulaciones y Aplicación al Catálogo Astrométrico UCAC4 ”*** ha sido realizada en el Instituto de Astrofísica de Andalucía, del Consejo Superior de Investigaciones Científicas (IAA-CSIC), por Doña Laura María Sampedro Hernández bajo su dirección y que constituye su Tesis Doctoral para optar al grado de Doctor en Física y Matemáticas.

Granada, 25 de Abril de 2016

D. Emilio Javier Alfaro Navarro

Fdo.:

AGRADECIMIENTOS

Quiero empezar agradeciendo a mi supervisor, Emilio J. Alfaro Navarro. Primero por haber confiado en mí y haberme dado la oportunidad de iniciarme en este gran trabajo y, desde luego, por todos estos años de paciencia, en los que con su gran didáctica me ha enseñado tanto. Por esas mañanas y tardes de debate de estadística, cúmulos y reflexiones que tanto me han aportado.

A mis padres Eloy y Mayte, por vuestro apoyo incondicional durante toda mi vida en los buenos y en los malos momentos. Por creer en mí y en los ánimos que me disteis en aquellos momentos en los que yo dudaba de mí misma. Por entenderme y acompañarme en mis viajes en busca de mi sueño durante todos estos años y por los esfuerzos que habéis hecho para que yo lo consiguiese.

A Alberto... Por tu apoyo incondicional en todos los momentos, por tu paciencia conmigo, por quererme cuando menos lo merecía pero cuando más lo necesitaba, por tu comprensión, por tu visión de la vida tan optimista y llena de aventuras, por todas las horas y horas de debate conmigo hablando sobre mis cumulitos y, desde luego, por ser quien eres. Has hecho de esta tesis uno más de nuestros muchos paseos agarrados de la mano.

A mis hermanos Alejandro y Hugo. A pesar de la distancia siempre estamos ahí para apoyarnos en los momentos en los que más lo necesitamos o para echarnos unas risas jugando al juego del laberinto y de los pingüinos. A mi Lali por siempre estar ahí y escucharme e intentar ayudarme en todos mis problemas con todo lo que estaba a su alcance. Por nuestros Lali, hija, Lali... que no falten nunca. Al resto de mi familia, gracias por estar a mi lado cuando os he necesitado. Por esas tardes de veranos interminables en la finca todos reunidos en los que respiraba vida y felicidad, todos juntos. A mis abuelos. Sé que siempre estáis ahí cuando os necesito.

A los que son y, considero familia, de Alberto. A todos! Por cómo me habéis apoyado estos años y el cariño que me habéis demostrado siempre. Por esas sobremesas que se juntan casi con la cena de conversaciones y risas y comida y más risa y más comida.... Por hacerme sentir como una más desde el principio y por querer estar a mi lado en los momentos más complicados.

A Rebeca y a Lifong. Siempre habéis confiado en mi. Sois una parte muy importante de mi vida y vuestro apoyo ha sido mi flotador en muchos momentos. Los difíciles momentos de la vida, han demostrado que independientemente de lo que pase o lo lejos que estemos, de alguna manera estaremos juntas apoyándonos en todo.

A Mariluz, Angel, Javi, Luisfe, Laura, Alvaro, Patricia, Chuchi, José, Gelu y Vanesa. Sois mis amigos y compañeros de fatigas desde que iniciamos la carrera. Durante todos estos años habéis demostrado estar ahí para cualquier cosa que necesite, desde el principio. Por todas las risas tan apreciadas por mi que nos echamos cada vez que estamos juntos. Por hacer lo imposible por verme cada vez que subo a Salamanca y esperarme con los brazos abiertos y con una sonrisa.

A Mirian, Pedro, Adrián :) , Pablo, Alba y Dario, mis amigos y compañeros de juegos en todos los fines de semana que hemos podido, principalmente los primeros años en los que nuestra casa era

literalmente la casa del juego. Por apoyarme en los momentos más complicados (Adrián también... si sólo con que sonría es suficiente) durante estos años de tesis y escucharme cuando lo he necesitado.

A Ari, Vasi y Fer aunque en la distancia siempre habéis estado ahí con nuestros Skypes, compartiendo todos los momentos importantes de nuestras vidas: viajes, nacimiento de Leo, experiencias, problemas...

A Rafa mi compi de despacho desde el principio. Muchas gracias por todas las conversaciones y apoyos de estos últimos cuatro años y medio. A Zaira, Rubén, Sara, Carolina, Sol, Lorena, Fran, Guillermo... y toda la demás gente del IAA. A Mayte por el apoyo que me ha dado en estos últimos meses de tesis, cuando más lo necesitaba. La verdad es que da gusto ir a trabajar a un sitio en el que te sientes como en casa. Y más te vale.... :)

A Wilton Dias, Hekto Monteiro y Claudia Mendes. Por hacerme tan sencilla mi estancia en Brasil. Por reírnos con mi portañolo!. Por esperarme siempre con una sonrisa y facilitarme la vida incluso con un océano de por medio.

Summary

This PhD-thesis has been devoted to the problem of determining (potential) members in open stellar clusters. As emphasized by Gilmore et al. 2012, these systems are ideal laboratories for many astrophysical researches since they represent perfect targets for the understanding of the formation of stars as much as the formation and destruction of the clusters themselves along with the understanding of the evolution of the stellar content and structure of the Milky Way.

One of the fundamental steps that must be faced in any study regarding open clusters is, precisely, the identification of its members. Unfortunately, it is most cases unfeasible to guarantee if a star truly belongs to a given cluster or not. Actually, it only turns out possible to estimate a probability of being so based on the available information for that star in the field. The definition itself of a cluster member has an inherent probabilistic nature. Although a number of efforts have been made up-to-date to nail down this issue, the problem still remains unsolved attracting the interest of many studies related with these stellar systems.

Taking into account the amount of incoming high-quality data from the many new astronomical surveys, it seems natural to devote time to the development of new tools capable to handle all these data in a self-consistent manner. In particular, for the open clusters and the membership estimations, it would be desirable that these tools would either combine as many variables as possible and weight the different bits of information based on the specific data quality.

With this idea in mind, in this thesis has been developed a new methodology capable to overcome the analysis of the membership probabilities of open clusters using an unlimited number of N-variables. By construction, the method does not take into account the nature of these variables. It only assumes the variables to be more densely distributed for the cluster members than for the field stars. Physical variables in the phase space (such as proper motions, radial velocities or positions) and/or the metallicity of the stars are supposed to meet this condition for an ideal cluster. If this idea holds, it turns out possible to define a metric, in an N-dimensional space (N-D), where the distribution of distances to the cluster centroid of the members will be more densely distributed than for the field stars.

Therefore, this methodology is able to estimate the distance of every star to the central over-density in an N-D space. Assuming the distribution of distances to be model as a combination of two 1-dimensional (1-D) Gaussian functions (one for the members and one for the field stars), it becomes possible to express a N-D problem as a 1-D. After an iterative process to determine the parameters describing the distributions (Wolfe, 1970), the new N-D centroid and the distribution of distances, the membership probabilities are calculated for the stars in the sample. During the final determination of members, it is applied the Bayes minimum error rate decision rule which minimizes the error in the classification (Kulkarni & Harman, 2011). This theorem establishes a threshold in the probability of 0.5 from which the sample is easily classified as members or field stars. Based on the resulting sample,

it becomes feasible to describe the properties of the clusters in the N-D space.

In this thesis, we designed a set of simulations to characterize the robustness and applicability of this new methodology. The simulations recreated a region of the sky with two coexisting stellar populations (a Galactic field and a stellar cluster), each described by a different distribution function in the phase space. In particular, the simulations were recreated in both the position and the proper motion space using an ample range of parameters directly derived from real catalogues as Dias et al. (2002) (DAML02) and Dias et al. (2014) (D14). Likewise, the simulated uncertainties (associated to the variables) were empirically drawn from the aforementioned catalogues. The reason of relying on real data was to assure the simulations to be as realistic as possible. It is worth mentioning at this point that two sets of uncertainties for the proper motions were implemented in these simulations: a) uncertainties simply described by normal (Gaussian) distributions and b) magnitude-dependent uncertainties. Finally, in order to test the impact of incomplete data in the membership analysis (due to selection effects), different levels of subsampling were also used artificially diluting the cluster signal into the field.

Taking advantage of the versatility of the method, it was performed a triple membership analysis including $N = 1, 2$ and 4 variables on these simulations. In the meantime, to be able to compare the goodness of these results in a more generic fashion, we decided to include in the same analysis two already existing well-known methodologies (a parametric approach presented in Cabrera-Caño & Alfaro (1985) and a non-parametric approach presented in Cabrera-Caño & Alfaro (1990)). Finally, it was necessary to carry out an estimation of any potential outliers in the proper-motion space (previous to any analysis) applying OUTKER technic (Cabrera-Caño & Alfaro, 1985).

The results obtained in this thesis prove the potential and versatility of this new methodology identifying members in (simulated) open clusters, sometimes even surpassing the other methodologies using a fixed number of variables. A discussion of these results and an in-depth description of the methodology has been published in the paper “Stellar Open Cluster’s Membership Probabilities: an N-Dimensional Geometrical Approach (Sampedro & Alfaro, 2016)”.

Apart from the design itself and testing of this new analytical tool, this PhD-thesis also aimed at deriving a new membership catalogue of open clusters. The catalogue (Sampedro et al., in prep.) contains the results of 4 different methodologies (the one developed in this thesis using $N=2$ and $N=4$ variables along with the parametric and non-parametric ones introduced above) for the sample of clusters listed in DAML02. The cluster radii were systematically calculated based on the radial density profiles. These radii were used to set the searching radius when downloading the data from the UCAC4 dataset (Zacharias et al., 2013). A total of 1876 clusters were analyzed and compared with those from DAML02 and D14, showing a remarkable high agreement.

These results will be published in two catalogues. On the one hand, a catalogue solely including the parameters describing the distributions (as much for the clusters as for the fields) either in the position as in the proper motion space. Basically, the number of members, the means, the dispersions and the correlation coefficients. In addition, the catalogue will include the number of stars classified as outliers, the new values for the radii (as defined in this thesis), the distances, the ages and the color

excess extracted from the DAML02. On the other hand, another catalogue including the astronomical coordinates, the proper motions, the uncertainties, a flag for the identification of outliers, the membership probabilities for each of the four methodologies and a cluster/field classification of the stars. Non detected clusters (for any methodology) are set to -1.

The results of this thesis may surely serve for incoming researches needing a determination of the stellar members belonged to the open clusters of our Galaxy.

Resumen

Esta tesis versa sobre la determinación de los miembros potenciales de los cúmulos abiertos estelares. Estos sistemas son considerados como uno de los laboratorios más importantes para la investigación astrofísica (Gilmore et al. 2012). Constituyen los laboratorios ideales para los estudios de formación estelar, de formación y destrucción de cúmulos, evolución estelar, y estructura y evolución de nuestra Galaxia, entre otros.

Uno de los pasos iniciales fundamentales a cualquier estudio que requiera de los cúmulos abiertos es, precisamente, la determinación de los miembros que los constituyen. La definición de miembro del cúmulo tiene una inherente naturaleza probabilística. No se puede asegurar que una estrella es miembro o no del cúmulo, pero podemos indicar que una estrella tiene una determinada probabilidad de serlo, en función de los datos disponibles sobre las estrellas en la vecindad del cúmulo. Numerosos esfuerzos se han hecho para intentar resolver este problema. Sin embargo, hoy en día sigue siendo una cuestión abierta y de mayor interés en todos los estudios relacionados con estos sistemas estelares.

Dada la gran cantidad de nuevos cartografiados que están llevándose a cabo o en preparación y, del gran número y precisión de las variables observadas que esperamos estos datos nos aporten, parece natural el intentar diseñar nuevas herramientas que permitan realizar un análisis de pertenencia que tenga en cuenta el mayor número de variables disponibles, aunque sean de distinta naturaleza y, a la vez sean, lo suficientemente flexibles como para rechazar o incluir variables dependiendo de su calidad y de la cantidad de información que contengan. Esta característica permite adecuar los estudios de membresía a la disponibilidad de las variables y adaptarlos a la base de datos que estemos utilizando. En este sentido, hemos querido dar un paso más allá, desarrollando una nueva técnica de análisis de pertenencia la cual utiliza un número no fijo de variables para realizar los análisis de pertenencia.

Esta nueva metodología determina los miembros potenciales de los cúmulos abiertos de estrellas a partir de un conjunto de N -variables físicas, medidas sobre las estrellas en la vecindad del cúmulo. El método no tiene en cuenta la naturaleza de estas variables siempre y cuándo se cumpla la hipótesis fundamental de que las mismas estén más densamente concentradas para los miembros del cúmulo que para las estrellas de campo. Así, las variables físicas que conforman el espacio de fase, y/o, por ejemplo, la metalicidad, cumplen este requisito para un cúmulo ideal. De aquí que consideremos que, definiendo una métrica en ese espacio de N -dimensiones (N -D), la distribución de distancias al centroide del cúmulo de las estrellas miembros también estará más densamente concentrada que la de las estrellas de campo.

Por tanto, la nueva metodología determina las distancias entre cada estrella y la sobre-densidad central en un espacio de N -D, en función de las N -variables que estemos usando. Asumimos que la distribución de distancias puede modelarse como la combinación de dos Gaussianas mono-dimensionales (1-D), una para los miembros del cúmulo y otra para las estrellas de campo, convirtiendo así un

problema N-D en uno 1-D. A partir de ahí, y mediante un proceso iterativo de determinación de los parámetros de las distribuciones (Wolfe, 1970), estimación del nuevo centroide N-D y de la distribución de distancias, se calculan las probabilidades de pertenencia al cúmulo de las estrellas de la muestra. Para la determinación de los miembros de los cúmulos se aplicó la regla de decisión de Bayes que minimiza el error en la clasificación (Kulkarni & Harman, 2011). Esta regla propone un valor umbral en la probabilidad de pertenencia de 0.5 a partir de la cual se clasifica la muestra en miembros del cúmulo y estrellas de campo. A partir de la determinación de los miembros realizada se pueden describir las propiedades del cúmulo en el espacio de N-variables.

Diseñamos un conjunto de simulaciones para evaluar las capacidades de la nueva metodología desarrollada, así como su rango de aplicación. Estas simulaciones recrean una región de cielo que incluye dos poblaciones estelares, con diferentes funciones de distribución en el espacio de fase: un campo Galáctico y un cúmulo estelar. Las simulaciones se llevaron a cabo en los espacios de posiciones y de movimientos propios (un sub-espacio del espacio de fase) usando diferentes conjuntos de parámetros que abarcan la mayoría de la casuística observada. Los valores de los parámetros fueron tomados de los principales y más completos catálogos de cúmulos abiertos estelares. En particular, se hizo uso de los trabajos Dias et al. (2002) (DAML02) y Dias et al. (2014) (D14). A partir de ellos hemos analizado las distribuciones estadísticas de los diferentes parámetros, considerando que las simulaciones de entrada serán más representativas del mundo real, si se ajustan al cúmulo 'típico' descrito por los valores de las distribuciones de los parámetros reales. De igual forma los errores simulados también deberían ser similares a aquellos que observamos en los catálogos disponibles actualmente. Aquí haremos dos distinciones: a) consideraremos que los errores de los movimientos propios de la muestra siguen una distribución Gaussiana, con independencia de otras variables estelares, y b) que la distribución de errores en los movimientos propios varía con la magnitud de la estrella. Por último, se diseñaron simulaciones afectadas de diferentes grados de sub-muestreos observacionales. La intensidad y la naturaleza del sub-muestreo es variable, por lo que si queremos simular un cúmulo realista tendremos que tener en cuenta también estos posibles sesgos y trabajar con datos sin un contraste de densidad del cúmulo bien definido y con diferentes porcentajes de muestreo.

Estas simulaciones y la gran flexibilidad que presenta la nueva metodología en el uso de diferentes conjuntos de variables, nos permitieron realizar tres análisis de pertenencia diferentes, en los que $N = 1, 2$ y 4 variables, fueron utilizadas. La aplicación de otras metodologías, una paramétrica (Cabrera-Caño & Alfaro, 1985) y otra no-paramétrica (Cabrera-Caño & Alfaro, 1990), ampliamente utilizadas durante décadas, nos permite la comparación de sus resultados con los obtenidos por la nueva metodología cuando son aplicadas al mismo conjunto de variables. Cabe mencionar que antes de la realización de estos análisis de pertenencia, es necesaria la determinación de posibles *outliers* en el espacio de los movimientos propios. Para su determinación hemos aplicado la técnica OUTKER desarrollada por Cabrera-Caño & Alfaro (1985).

Los resultados obtenidos muestran el potencial de la nueva metodología en cuanto a la determinación de los miembros de los cúmulos abiertos, siendo capaz de obtener similares o incluso mejores resultados que aquellas metodologías que tienen un número fijo de variables para realizar los análisis de pertenencia. Estos resultados, junto con la descripción completa de la nueva técnica desarrol-

lada, fueron publicados en un artículo titulado: *Stellar Open Clusters' Membership Probabilities: an N-Dimensional Geometrical Approach* (Sampedro & Alfaro, 2016).

Uno de los principales objetivos de esta tesis consiste en la creación de un nuevo catálogo de análisis de pertenencia de cúmulos abiertos (Sampedro et al. en preparación). Para ello, se aplicaron las metodologías anteriormente mencionadas así como la nueva metodología para 2 y 4 variables, a los cúmulos listados en el catálogo DAML02. Para ello, se analizaron los perfiles de densidad radial de todos estos cúmulos con el objetivo de conocer el radio de los mismos, el cual fue usado como radio de extracción de los datos del cartografiado UCAC4 (Zacharias et al., 2013). Un total de 1876 cúmulos fueron analizados con cuatro análisis de pertenencia diferentes. Sus resultados fueron comparados con los de los trabajos DAML02 y D14 mostrando un elevado grado de acuerdo, fundamentalmente con D14.

Los resultados obtenidos serán publicados en forma de dos catálogos. En uno de ellos se listarán los parámetros que describen las distribuciones de las poblaciones de cúmulo y campo, en los espacios de posiciones y de movimientos propios (número de miembros, medias, dispersiones y coeficientes de correlación). Además, se listarán el número de *outliers* determinados, el radio de los cúmulos determinados en esta tesis y, la distancia, edad y exceso de color extraídos del catálogo DAML02. Individualmente para cada cúmulo, se aportará otro catálogo con sus coordenadas, movimientos propios, errores, un indicador que tomará el valor de 1 si la estrella ha sido clasificada como un *outlier* ó 0 en caso contrario, las probabilidades de pertenencia obtenidas por cada metodología y, unos indicadores de ceros y unos los cuales indican si las metodologías han determinado que la estrellas es de campo o de cúmulo, respectivamente. Si alguna metodología no ha determinado cúmulo este resultado vendrá dado por el valor -1.

Con el trabajo realizado en esta tesis, esperamos poder ayudar en las investigaciones astrofísicas futuras que requieran la determinación de las estrellas pertenecientes a los cúmulos abiertos de nuestra Galaxia.

Index

1	Introducción	1
1.1	Motivación de esta tesis.	1
1.2	Cúmulos abiertos.	2
1.3	Principales líneas de investigación con datos astrométricos y espectroscópicos.	3
1.4	Principales catálogos astrométricos de cúmulos estelares.	4
1.5	Metodologías de análisis de pertenencia en cúmulos abiertos.	7
2	Metodologías	11
2.1	Nueva metodología geométrica N-Dimensional.	11
2.1.1	Introducción.	11
2.1.2	Métrica en un espacio de N-Dimensiones	11
2.1.3	Probabilidades de pertenencia Bayesianas.	13
2.2	Método No-Paramétrico	16
2.3	Método Paramétrico	21
2.3.1	Metodología Paramétrica 2-D	21
2.3.2	Metodología Paramétrica 1-D	23
2.4	Otras técnicas aplicadas.	23
2.4.1	Determinación de los <i>outliers</i> . Procedimiento OUTKER.	23
3	Simulaciones	25
3.1	Introducción	25
3.2	Parámetros del sistema de cúmulos estelares en los catálogos de Dias y colaboradores.	26
3.3	Simulaciones con error Gaussiano de σ_μ constante.	29
3.4	Simulaciones con error Gaussiano con σ_μ dependiente de la magnitud.	32
3.5	Ausencia de sobre-densidad en el espacio de posiciones.	36
3.5.1	Simulaciones con distribuciones aleatorias en el espacio de posiciones para la población de cúmulo.	36
3.5.2	Efectos de sub-muestreo en la muestra.	38
4	Resultados obtenidos en las simulaciones	39
4.1	Introducción	39

4.2	Resultados obtenidos en las simulaciones.	40
4.2.1	Error Gaussiano con σ_μ constante.	40
4.2.2	Error Gaussiano con σ_μ dependiente de la magnitud.	47
4.2.3	Efecto de la simulación de distribuciones aleatorias para la población de cúmulo.	55
5	Catálogo de cúmulos reales.	59
5.1	Catálogo DAML02.	59
5.2	Determinación de los radios de los cúmulos abiertos listados en DAML02.	60
5.2.1	Introducción	60
5.2.2	Problemas encontrados durante la determinación de los radios de los cúmulos.	63
5.2.3	Métodos en la determinación de los radios.	68
5.2.4	Procedimiento para la obtención del radio del cúmulo a partir de un ajuste polinómico al PDR.	71
5.3	Resultados obtenidos por las metodologías descritas en esta tesis.	73
5.4	Comparación con otras bases de datos	91
5.4.1	Comparación con el trabajo D14.	91
5.4.2	Comparación con el catálogo DAML02.	97
6	Conclusions	105
6.1	Main conclusions of this PhD thesis.	105
6.2	Personal request.	107
6.3	Personal conclusion.	107
7	Conclusiones	109
7.1	Principales conclusiones de esta tesis.	109
7.2	Requerimiento personal.	111
7.3	Conclusión personal.	111
8	Trabajo Futuro	113
A	Apéndice 1	115
A.1	Comportamiento de C y de M en función de la DC obtenidos para las metodologías aplicadas.	115
B	Apéndice 2	123
B.1	Efectos del sub-muestreo.	123
C	Apéndice 3	125
C.1	Descripción de los parámetros de las distribuciones de ambas poblaciones obtenidos por los diferentes análisis de pertenencia..	125
D	Publicaciones	127

Lista de Tablas

2.1	Ejemplo de archivo de configuración utilizado por la metodología geométrica.	17
3.1	Valores de los parámetros fijos en todas las simulaciones, de las distribuciones de las poblaciones de cúmulo y campo, en los espacios de posiciones y de movimientos propios.	30
3.2	Valores de los parámetros variables en las simulaciones, de las distribuciones de las poblaciones de cúmulo y campo, en los espacios de posiciones y de movimientos propios.	30
3.3	Esquema de los valores que toman los parámetros variables de las configuraciones simuladas para un porcentaje de estrellas de cúmulo.	30
4.1	Diferencias y dispersiones de los valores medios obtenidos respecto de los simulados, de los parámetros asociados a las distribuciones espaciales para los cuatro casos de sub-muestreo simulados.	49
4.2	Diferencias y dispersiones de los valores medios obtenidos respecto de los simulados, de los parámetros asociados a las distribuciones de movimientos propios para los cuatro casos de sub-muestreo simulados.	49
5.1	Diez nuevos centroides en el espacio de posiciones	68
5.2	Características de la distribución de las diferencias en la determinación de los centroides en el espacio de movimientos propios entre las metodologías aplicadas en esta tesis.	86
5.3	Características de la distribución de las diferencias en la determinación de los centroides en el espacio de movimientos propios entre las metodologías aplicadas en esta tesis y el trabajo de D14.	93
5.4	Características de la distribución de las diferencias en la determinación de los centroides en el espacio de movimientos propios entre las metodologías aplicadas en esta tesis y el trabajo de DAML02.	98
5.5	Número de objetos analizados y, centroides y dispersiones en movimientos propios en mas/año de los 19 cúmulos inéditos obtenidos.	101
C.1	Parámetros referentes a la población de cúmulo, determinados por todas las metodologías aplicadas, para todos los cúmulos analizados. Los mismos parámetros se publicarán para la población de campo.	126

Lista de Figuras

2.1	Diagrama de flujo esquemático del análisis de pertenencia realizado por la metodología geométrica.	12
2.2	Efecto de la determinación del centroide de las variables en la distribución de distancias obtenida.	15
2.3	Ejemplo de ajuste a una distribución de distancias.	16
2.4	Funciones de densidad de probabilidad de la metodología no-paramétrica en los espacios de posiciones y de movimientos propios.	21
2.5	Ejemplo de distribución de probabilidad en el espacio de los movimientos propios obtenida por la metodología paramétrica.	23
3.1	Distribuciones del número de objetos, distancias y radios de una muestra de 1646 cúmulos reales a partir de los cuales se obtuvieron los valores utilizados en las simulaciones.	27
3.2	Distribución de las dispersiones de las estrellas de campo en el espacio de movimientos propios a partir de las cuales se realizarán las simulaciones.	28
3.3	Distribución de las dispersiones en movimientos propios de los cúmulos listados en el trabajo de D14.	29
3.4	Ejemplo de una simulación en los espacios de posiciones y de movimientos propios. . .	31
3.5	Errores en movimientos propios en función de la magnitud para una muestra de 21 cúmulos reales.	33
3.6	Distribuciones de los errores en movimientos propios y de las magnitudes de una muestra formada por 21 cúmulos reales	33
3.7	Ejemplos de distribuciones de errores en movimientos propios y de magnitudes simuladas. 34	
3.8	Ejemplos de simulaciones en el espacio de movimientos propios y comparación con el cúmulo NGC 2360.	35
3.9	Perfil de densidad radial y distribución en el espacio de posiciones de la región del cúmulo ASCC 102.	37
3.10	Distribuciones en el espacio de posiciones del cúmulo NGC 2506.	37
3.11	Ejemplo de simulación de una distribución aleatoria en el espacio de posiciones y comparación con el cúmulo NGC 5138.	38

4.1	Comportamiento de las figuras de mérito en función de la DC obtenido en las <i>Sim1</i> . . .	42
4.2	Dependencia con el porcentaje de estrellas de cúmulo simuladas de los valores de ambas figuras de mérito respecto de la DC para MTG2.	43
4.3	Distribución de probabilidades de pertenencia para un 20% y un 80% de estrellas de cúmulo simuladas, obtenidas por MTG1, MTG2 y MTG4.	45
4.4	Errores obtenidos de ambas figuras de mérito en función de la DC	46
4.5	Influencia de la presencia de sub-muestreo en la muestra en el comportamiento de las figuras de mérito en las <i>Sim1</i>	48
4.6	Dependencia con el porcentaje de las estrellas de cúmulo simuladas de los valores de C y M respecto de la DC en las <i>Sim2</i>	51
4.7	Comparación del comportamiento de las dos figuras de mérito en función de la DC , obtenido por las metodologías MT1 y MTG1, en las <i>Sim1</i> y <i>Sim2</i>	52
4.8	Comparación del comportamiento de las dos figuras de mérito en función de la DC , obtenido por las metodologías MT2 y MTG2, en las <i>Sim1</i> y <i>Sim2</i>	53
4.9	Comparación del comportamiento de las dos figuras de mérito en función de la DC , obtenido por las metodologías MT4 y MTG4, en las <i>Sim1</i> y <i>Sim2</i>	54
4.10	Influencia de la presencia de sub-muestreo en la muestra en el comportamiento de las figuras de mérito en las <i>Sim2</i>	56
4.11	Comportamiento de las figuras de mérito en función de la DC obtenido en las simulaciones con distribuciones de cúmulo aleatorias en el espacio de posiciones.	57
5.1	Ejemplos de perfiles de densidad radial de los cúmulos Berkeley 90 y NGC 2451B . . .	61
5.2	Ejemplo de sub-estimación de radio publicado en DAML02 para el cúmulo NGC 2099. . .	63
5.3	Ejemplo de sobre-estimación de radio publicado en DAML02 para el cúmulo IC 2944. . .	64
5.4	Ejemplo de ausencia de sobre-densidad para el cúmulo NGC 2319.	65
5.5	Ejemplo de depresión en la densidad en la región central del cúmulo IC 5146.	66
5.6	Ejemplo de centroide en posiciones no correspondiente con la posición central del cúmulo IC 361.	67
5.7	Ejemplo de contaminación por otro cúmulo en la región del cúmulo Trumpler 16. . . .	69
5.8	Ejemplos de ajustes por polinomio al PDR de los cúmulos Trumpler 14 y FSR 0941. . .	72
5.9	Radios obtenidos por ajuste a un polinomio y por el ajuste de King de 3 parámetros frente al radio de los cúmulos listados en DAML02.	73
5.10	Distribución de los radios muestrales.	74
5.11	Fracción de estrellas analizadas respecto al número de estrellas analizadas para 2137 cúmulos listados en DAML02	75
5.12	Comparación de los centroides y dispersiones en el espacio de movimientos propios de la muestra de 982 cúmulos, obtenidos por MT2 y MT4 respecto a MTG2.	77
5.13	Dispersiones de la muestra de 982 cúmulos obtenidas por las metodologías aplicadas en el espacio de movimientos propios en función del número de estrellas analizadas. . . .	78
5.14	Distribuciones en los espacios de posiciones y de movimientos propios del cúmulo Ruprecht 3.	78

5.15	Explicación de la diferencia encontrada en el número de cúmulos obtenidos entre MTG2 y MT2.	80
5.16	Ajustes realizados a una simulación por las metodologías MTG2 y MT2 a las distribuciones de distancias y de movimientos propios.	81
5.17	Porcentaje de miembros de cúmulo clasificados por MT4 y MTG4 en <i>Sim3</i> en función del porcentaje simulado.	82
5.18	Número de miembros obtenidos por cada metodología para la muestra de 1876 cúmulos reales, en función del número de estrellas analizadas.	83
5.19	Número de estrellas analizadas en función del diámetro angular de los cúmulos analizados.	84
5.20	Distribución de los movimientos propios en las regiones de los cúmulos Alessi 19 y Melotte 22.	84
5.21	Diferencias entre los centroides obtenidos por MTG2 y MT2 en el espacio de movimientos propios.	85
5.22	Implicación en los resultados obtenidos por la existencia de absorciones no homogéneas a lo largo de la región del cúmulo Czernik 38.	87
5.23	Análisis de pertenencia del cúmulo NGC 2632.	89
5.24	Análisis de pertenencia del cúmulo Blanco 1.	90
5.25	Ejemplo de catálogo final obtenido para el cúmulo NGC 7226.	92
5.26	Diferencias en los centroides en movimientos propios entre las metodologías aplicadas en esta tesis y los publicados en D14.	94
5.27	Centroides obtenidos por MTG2 frente a los publicados en D14 en función del número de miembros de cúmulo obtenidos por MTG2.	95
5.28	Comparación del número de miembros obtenidos por las metodologías aplicadas en esta tesis en función del número de miembros obtenidos por D14.	96
5.29	Dispersiones en el espacio de movimientos propios obtenidas por MTG2 y MT4 respecto a las obtenidas por D14.	97
5.30	Dispersiones en el espacio de movimientos propios obtenidas por MTG2 y MT4 respecto a las obtenidas por D14 una vez sustraídos los errores de UCAC4.	98
5.31	Diferencias en los centroides en movimientos propios respecto al valor publicado en DAML02.	99
A.1	Comportamiento de las figuras de mérito en función de la <i>DC</i> obtenidos por MT1. . .	116
A.2	Comportamiento de las figuras de mérito en función de la <i>DC</i> obtenidos por MTG1. .	117
A.3	Comportamiento de las figuras de mérito en función de la <i>DC</i> obtenidos por MT2. . .	118
A.4	Comportamiento de las figuras de mérito en función de la <i>DC</i> obtenidos por MTG2. .	119
A.5	Comportamiento de las figuras de mérito en función de la <i>DC</i> obtenidos por MT4. . .	120
A.6	Comportamiento de las figuras de mérito en función de la <i>DC</i> obtenidos por MTG4. .	121
B.1	Efectos del sub-muestreo en las figuras de mérito en función de la <i>DC</i> para las <i>Sim2</i> . .	124

1

Introducción a los Cúmulos Abiertos

1.1 Motivación de esta tesis.

Entender nuestra Galaxia ha sido uno de los retos intelectuales más importantes del ser humano. Respuestas a las preguntas de dónde venimos o a dónde vamos, pasan por entender el origen, la evolución tanto química como dinámica, la distribución espacial, así como entender la evolución de las estrellas que conforman la Vía Láctea. Numerosos estudios han arrojado luz sobre muchas de estas, en principio, incógnitas. En particular, estudios astrofísicos realizados a partir del estudio de unos sistemas físicos específicos, han contribuido a incrementar notablemente nuestro conocimiento y entendimiento de nuestra Galaxia. Estos sistemas físicos son los *Cúmulos Abiertos*, sobre los cuales versa esta tesis, siendo considerados como uno de los laboratorios más importantes para la investigación astrofísica de nuestra Galaxia (Gilmore et al. 2012).

Uno de los pasos iniciales fundamentales a cualquier estudio que requiera de los cúmulos abiertos es, precisamente, la determinación de los miembros que los constituyen. A lo largo del siglo XX se han diseñado diferentes métodos de estimación de la probabilidad de pertenencia al cúmulo, cada uno de los cuáles define de alguna manera lo que es un cúmulo estelar. A pesar de los numerosos esfuerzos que se han hecho para intentar resolver este problema, sigue siendo hoy en día una cuestión abierta y de gran interés en todos los estudios relacionados con estos sistemas estelares.

Los análisis de pertenencia son muy dependientes del tipo y calidad de los datos disponibles. La promesa de Gaia nos obliga a pensar en herramientas que aprovechen al máximo la diversidad de variables físicas, la precisión de las mismas y la ingente cantidad de estrellas que se van a observar. Éste es el objetivo de mi trabajo de tesis, desarrollar un nuevo método de determinación de probabilidades de pertenencia a un cúmulo, especialmente diseñado para los datos de Gaia.

Su aplicación a simulaciones me ha permitido caracterizar esta nueva metodología usando diferentes conjuntos de variables y describir los cúmulos obtenidos a partir de los análisis de pertenencia realizados. Además, la necesidad de aplicar diferentes técnicas a datos reales, nos ha conducido a una mejor caracterización de estas simulaciones, trabajo que me ha enseñado la importancia de realizar unas simulaciones lo más similares posible a los datos que vayamos a utilizar para poder entender los

resultados obtenidos.

La aplicación de diferentes técnicas de determinación de miembros de los cúmulos abiertos a grandes bases de datos, me ha aportado una visión más amplia de lo que es un cúmulo estelar y de qué variables físicas características presentan una mayor incertidumbre.

Esta tesis versa sobre la determinación de los miembros físicos de los cúmulos abiertos de estrellas. Empezamos describiendo estos sistemas estelares.

1.2 Cúmulos abiertos.

Desde un punto de vista físico, un cúmulo abierto es una agrupación de decenas a millares de estrellas que nacen dentro de nubes moleculares gigantes, nubes de gas y polvo distribuidas en el disco de la Galaxia, a partir de procesos de colapso y fragmentación de las mismas (Elmegreen et al. 2000; Lada 2010), que tienen lugar en un corto periodo de tiempo de unos varios millones de años (Leisawitz, Bash, & Thaddeus, 1989). Las estrellas que pertenecen a los cúmulos están dinámica y gravitatoriamente ligadas, presentando una dispersión de velocidades en torno a un movimiento común. Por tanto, comparten una misma región del espacio y, al nacer de la misma nube molecular, podemos asumir que además comparten una misma composición química, edad y distancia (Bruzual, 2010).

Los cúmulos abiertos también se caracterizan por mostrar una morfología muy variada. Así, pueden presentar desde distribuciones de tipo radial, donde en el centro del cúmulo existe una mayor densidad de estrellas que decae conforme aumenta la distancia a su centro, hasta distribuciones caracterizadas por la presencia de sub-estructuras a lo largo de todo el cúmulo. Estudios más recientes (Smith et al., 2011) parecen indicar que la distribución estelar inicial, la dispersión de velocidades, así como la evolución de las condiciones iniciales que presentan los cúmulos en el momento de la expulsión del gas, son factores claves para la supervivencia del mismo.

Desde el punto de vista de esta tesis, un cúmulo abierto es un conjunto de N -cuerpos en los que se han medido N -variables que están más densamente distribuidas para las estrellas que pertenecen al mismo que para las estrellas que no pertenecen a él, comúnmente conocidas como estrellas de campo. Precisamente, las características comunes a las estrellas que forman parte de los cúmulos nos permitirán diferenciarlas de las de campo. Por tanto, la definición de miembro del cúmulo, tiene una clara naturaleza probabilística. No se puede asegurar que una estrella es miembro o no del cúmulo, pero podemos indicar que una estrella tiene una determinada probabilidad de serlo, en función de los datos que estemos usando en el análisis de pertenencia. Cabe indicar que, las variables utilizadas en esta tesis son aquellas que cumplen esta condición como posiciones, movimientos propios y velocidades radiales, aunque la metodología es ampliable a otras variables que serán medidas por Gaia, como paralaje, metalicidad, etc.

En este punto es interesante indicar que la clasificación de una muestra en estrellas de campo o miembros de un cúmulo abierto está supeditada a las variables que se estén utilizando en el análisis de pertenencia. Cabe indicar que no siempre la utilización de unas variables puede llevar a la correcta

clasificación de la muestra. Dependiendo del tipo y calidad de las variables disponibles puede ser interesante ver previamente que tipo de metodología es la más adecuada.

Precisamente éste es uno de los problemas que tienen la mayoría de las técnicas existentes actualmente: su análisis de pertenencia está supeditado a la disponibilidad de los datos y en muchos casos al tipo de distribución de los mismos en el espacio de variables. Como ya se ha indicado, los cúmulos abiertos están caracterizados por presentar una gran variabilidad en sus distribuciones espaciales, desde distribuciones con simetría radial a la presencia de diferentes sub-estructuras, que hace difícil hablar de un centro del cúmulo, a lo largo de toda su extensión. En función de la metodología y de la definición de cúmulo que ésta asuma, puede llegarse o no a una solución realista.

1.3 Principales líneas de investigación con datos astrométricos y espectroscópicos.

Se considera que un alto porcentaje de entre un 70-90% (Lada & Lada, 2003) de las estrellas que pueblan el disco de la Vía Láctea iniciaron su vida en cúmulos estelares (Lada & Lada 1991; Lada & Lada 2003; Proszkow & Adams 2009). Por tanto, la gran mayoría de las estrellas que pueblan nuestra Galaxia nacieron previamente en estas guarderías estelares (Portegies Zwart, McMillan, & Gieles, 2010), emancipándose posteriormente dependiendo de la masa inicial del cúmulo y de las condiciones de su entorno.

Los cúmulos abiertos son usados en multitud de investigaciones astrofísicas al presentar una gran variedad de edades y distancias. Además, estos parámetros junto con la metalicidad pueden ser mejor determinados para estos objetos que para estrellas individuales (Paunzen & Netopil 2006; Monteiro, Dias, & Caetano 2010), lo que convierte a estos sistemas en catas singulares para el estudio de la evolución estelar y de la estructura, evolución y formación de las galaxias que los albergan (Phelps & Janes 1994; Frinchaboy & Majewski 2008; Gilmore et al. 2012; Oliveira et al. 2013).

Los cúmulos abiertos han sido utilizados para estudiar y determinar la posición de los brazos espirales así como su cinemática (van den Bergh & McClure 1980; Janes, Tilley, & Lynga 1988; Dias & Lépine 2005; Camargo, Bica, & Bonatto 2013; Junqueira et al. 2015) y para establecer la existencia o no de una estructura de la metalicidad en el disco Galáctico (Lépine et al. 2011; Heiter et al. 2013).

En nuestra Galaxia los cúmulos se utilizan para establecer restricciones a los modelos de evolución estelar (Schaller et al. 1992; Girardi et al. 1996; Landin et al. 2006), para analizar las propiedades de la formación estelar tales como la universalidad de la función inicial de masa, la presencia o no de segregación de masa primordial (Bonnell & Davies 1998; Allison et al. 2009) o la ruptura del equilibrio dinámico (Smith et al., 2011). Dado que el fin último de los cúmulos estelares es diluirse en el disco Galáctico, el número de estos sistemas por intervalo de edad puede proporcionarnos una idea de las escalas temporales de destrucción de un cúmulo y de la relevancia de los diferentes procesos que intervienen en su dilución (Krumholz et al., 2014).

Poco es conocido también sobre la evolución de la estructura espacial de los cúmulos abiertos.

Simulaciones numéricas (Bonnell, Bate, & Vine, 2003) y observaciones (Larson, 1995) de cúmulos abiertos en las primeras etapas de su formación, demuestran que la formación estelar ocurre a lo largo de patrones definidos por las regiones más densas de las nubes moleculares a partir de las cuales se originan. Durante las primeras etapas de la evolución de un cúmulo, Sánchez & Alfaro (2009) y Alfaro & Sánchez (2011), encuentran una tendencia con la edad en la evolución espacial de los cúmulos de distribuciones con presencia de sub-estructuras a distribuciones radiales. Diferentes estudios numéricos también apoyan esta hipótesis, tales como Allison et al. (2010), Cartwright & Whitworth (2004) o Parker et al. (2014). Diferentes técnicas han sido desarrolladas para determinar la estructura espacial de los cúmulos abiertos. Entre ellas resaltamos el trabajo de Cartwright & Whitworth (2004), el cual se basa en la estimación del valor de un parámetro denominado Q . Para la determinación de este parámetro utilizan la técnica del MST (Minimum Spanning Tree, Gower & Ross 1969) y a partir de su valor, son capaces de clasificar los cúmulos en las dos clases de distribuciones espaciales.

Una tarea previa a estos estudios es la determinación de los miembros físicos del cúmulo. Esta parte es crucial en la determinación de los principales parámetros astrométricos asociados a los cúmulos (Kharchenko, Pakulyak, & Piskunov 2003; Frinchaboy & Majewski 2008; Dias et al. 2012), principalmente para aquellos con pocos miembros o con un gran diámetro angular, en los que la contaminación de estrellas de campo, puede sesgar los resultados finales obtenidos para estos parámetros (Baumgardt, Dettbarn, & Wielen, 2000). La idea inicial no puede ser más sencilla: hay que separar los miembros del cúmulo de las estrellas de campo. Sin embargo, diferentes problemas pueden emerger haciendo que esta clasificación en dos poblaciones sea complicada y, en muchos casos, sesgada, incompleta y/o errónea.

Así un aspecto fundamental a la hora de plantearse la separación de una muestra de estrellas, en la vecindad de un cúmulo estelar, entre miembros del cúmulo y estrellas de campo es el de conocer qué variables físicas medidas sobre las estrellas tenemos disponibles y que herramientas matemáticas son las más adecuadas a este tipo de variables. Veamos primero cuales son las características fundamentales de los principales cartografiados ampliamente utilizados en los estudios de membresía de los cúmulos abiertos de la Vía Láctea.

1.4 Principales catálogos astrométricos de cúmulos estelares.

A partir de finales de los años 90, fundamentalmente con el lanzamiento del satélite Hipparcos (Kovalevsky, 1998), se puede considerar que comenzó la era de las grandes bases de datos. Durante casi cuatro años el satélite Hipparcos tomó datos astrométricos de paralajes, posiciones y movimientos propios de casi 1.2×10^5 objetos predeterminados, con una precisión del orden de los miliares de segundos (mas) para $H_p < 9$ mag, con unos errores sistemáticos inferiores a 1 mas. Los datos publicados son referidos al ICRS (International Celestial Reference System, Ma 1997). Por otro lado, el experimento Tycho observó más de un millón de objetos con magnitudes hasta $V \sim 11.5$ mag, teniendo una completitud hasta $V \sim 10.5$ mag, con una precisión astrométrica de 7 mas. La unión de estos dos catálogos dió lugar al conocido como Catálogo de Hipparcos.

En el año 2000, con la publicación del catálogo Tycho-2 (Høg et al., 2000), se incrementó el número de objetos con datos de posiciones y movimientos propios hasta los 2.5×10^6 , de las estrellas más brillantes del cielo ($V = 11.5$) con unos errores sistemáticos inferiores a 1 mas y a 0.5 mas/año, respectivamente. Con un número similar de objetos observados, Ducourant et al. (2006) publicó un catálogo en la región del cielo $+11^\circ \leq \delta \leq +18^\circ$, con datos de posiciones y movimientos propios en el sistema ICRS. La precisión en posiciones de este catálogo es de 50-70 mas y en movimientos propios de 1.5-6 mas/año.

Diez años después de la publicación del catálogo Tycho-2, la publicación del catálogo PPMXL (Roeser, Demleitner, & Schilbach, 2010) incrementó el número de objetos observados hasta los 9×10^8 . Este catálogo procede de la combinación de los catálogos USNOB-1 (United States Naval Observatory, Monet et al. 2003) y 2MASS (Two Micron All Sky Survey, Skrutskie et al. 2006) donde, tanto las posiciones como los movimientos propios publicados, a diferencia del catálogo USNOB-1, el cual contiene datos para más de mil millones de objetos, están en el sistema de referencia ICRS. La completitud alcanzada por PPMXL llega a $V \approx 20$ mag, con fotometría 2MASS para unos 410 millones de objetos. Los errores medios de los movimientos propios van desde los 4 mas/año hasta más de 10 mas/año con unos errores sistemáticos de entre 1-2 mas/año, siendo superiores para agrupaciones de estrellas, mientras que los errores en posiciones tienen un rango de valores desde los 100 mas hasta los 300 mas. El número de observaciones utilizadas para determinar los movimientos propios de las estrellas varía entre 2 hasta 6, dependiendo si la estrella no tenía o sí, datos en USNOB-1.

En 2011 se publicó el catálogo SPM4 (The Southern Proper Motion Catalog, Girard et al. 2011) con datos para aquellos objetos situados en declinaciones inferiores a -20° . Este catálogo contiene coordenadas y movimientos propios absolutos para 103 millones de objetos, siendo completo hasta una magnitud $V = 17.5$. Los errores en movimientos propios son del orden de los 2-3 mas/año, con unas incertidumbres sistemáticas del orden de 1 mas/año. Posteriormente, a principios del 2013 se publicó el catálogo UCAC4 (United States Naval Observatory CCD Astrograph Catalog, Zacharias et al. 2013), con datos para 113 millones de objetos, de los cuales 105 millones tienen datos de movimientos propios absolutos, siendo completo hasta magnitud $R = 16$. La precisión en las posiciones obtenidas es del orden de 15-100 mas, dependiendo de la magnitud del objeto. En cuanto a los movimientos propios, los errores publicados son del orden de entre 1-10 mas/año dependiendo de la magnitud y del número de observaciones realizadas, con unos errores sistemáticos entre 1-4 mas/año.

Está claro que la aplicación de la tecnología CCD en las observaciones astrométricas ha producido un gran salto en la precisión obtenida, pero sobre todo en la obtención de un mayor número de objetos observados. Sin embargo, hay que tener presente que los datos aportados por estas grandes bases de datos, proceden de un proceso automático de reducción, no pudiendo ignorar que podemos encontrar movimientos propios de baja calidad o incluso erróneos, fundamentalmente para estrellas débiles que son las que menos épocas de observación tienen. En el trabajo de Teixeira et al. (2014) realizan una comparación entre sus datos, obtenidos con el telescopio Valinhos Observatory (Brasil), con los de las bases de datos PPMXL, UCAC4 y SPM4. En este trabajo resaltan la problemática que tienen estas grandes bases de datos (fundamentalmente PPMXL), en cuanto a la determinación de los movimientos propios se refiere, sobretodo para estrellas débiles. Además, Vityazev & Tsvetkov

(2014) encuentran una diferencia sistemática de 4 mas/año entre los movimientos propios publicados en UCAC4 y PPMXL. De hecho, Vickers, Roeser, & Grebel (2016) encuentran para una muestra de 10^3 fuentes extragalácticas, movimientos propios distintos de cero y en torno a 0.26 mas/año en α y de -2.9 mas/año en δ , proponiendo un método de corrección para los datos presentes en PPMXL.

La era Hipparcos se acaba veintitrés años después del lanzamiento de este satélite para dar lugar a la era Gaia (Perryman et al. 2001; Lindegren et al. 2008). El satélite Gaia lanzado a finales del 2013, es la misión fundamental de la Agencia Espacial Europea (ESA). Sus objetivos científicos fundamentales consisten en estudiar la estructura y evolución cinemática, dinámica y química de nuestra Galaxia. Se espera que los datos aportados por Gaia mejoren los obtenidos por el satélite Hipparcos en un factor 50-100 en la precisión de las posiciones y de un factor 10^4 en cuanto al número de estrellas se refiere. De hecho, Gaia observará en torno a 10^9 de los objetos más brillantes de nuestra Galaxia, en un rango de magnitudes de $V \sim 6$ hasta 20 mag, con una precisión sin precedentes del orden de unos pocos a decenas de micro-arcosegundos (μas) y de $\mu\text{as/año}$ para posiciones y movimientos propios respectivamente, para magnitudes $V \leq 15$ mag y, precisiones de sub-mas para 20 mag (de Bruijne, 2012). Gaia aportará paralajes para unas 2×10^7 estrellas con errores inferiores al 1% y para más de 10^8 estrellas con errores inferiores al 5% (Perryman, 2005). En cuanto a los datos espectroscópicos, este satélite determinará velocidades radiales para unas 10^8 estrellas, con errores de 1 km/s para las estrellas más brillantes y de 15 km/s a 17 mag (Wilkinson et al. 2005; Perryman 2005).

La primera publicación de datos de Gaia se espera que tenga lugar en otoño del 2016 y estará formada por más de 10^5 estrellas comunes con la misión Hipparcos. Este proyecto conocido como *HTPM* por sus siglas en inglés (Hundred Thousand Proper Motions project), y propuesto por Mignard (2009), espera alcanzar precisiones del orden de los 14-100 $\mu\text{as/año}$ (Michalik et al., 2014). Durante los cinco años (extensible a uno más) que durará la misión, Gaia observará de media cada objeto entre 80 y 200 veces (Wyrzykowski & Hodgkin, 2012), obteniendo una cantidad de datos equivalente a grabar 1.5 millones de CD-ROMs.

Aunque las variables principales de medida de Hipparcos y Gaia son las mismas, es decir, posiciones, movimientos propios y paralajes, los casos científicos son diferentes (Eyer et al., 2012). Mientras Hipparcos estaba orientado fundamentalmente hacia el estudio estelar, Gaia tiene como objetivo el estudio de la estructura y evolución Galáctica. En ambos casos, los cúmulos abiertos juegan un papel fundamental.

El espacio de fase puede ser descrito completamente por seis parámetros: tres de posición y tres de movimiento. Estas variables son tradicionalmente la ascensión recta α , la declinación δ , el paralaje π , el movimiento propio en ascensión recta μ_α , el movimiento propio en declinación μ_δ y la velocidad radial v_r . Sólo los cinco primeros parámetros son considerados clásicamente como parámetros astrométricos. La velocidad radial es calculada preferentemente por técnicas espectroscópicas debido a la, en general, falta de precisión cuando es calculada a partir de observaciones llevadas a cabo en pocos años (Dravins, Lindegren, & Madsen, 1999).

Aunque Gaia aportará velocidades radiales, como ya se ha comentado, numerosos proyectos espectroscópicos han sido realizados desde tierra para la estimación de las velocidades radiales, con el

objetivo de complementar los datos aportados por el satélite Gaia. Entre estos proyectos destacamos RAVE (the Radial Velocity Experiment, Steinmetz et al. 2006) y GES (The Gaia-ESO survey, Gilmore et al. 2012; Randich, Gilmore, & Gaia-ESO Consortium 2013).

RAVE es un cartografiado espectroscópico cuyo objetivo consiste en medir velocidades radiales y parámetros atmosféricos estelares de 10^6 estrellas situadas en el hemisferio sur. Utiliza un espectrógrafo multi-objeto de media resolución ($R \sim 7500$) en la región del ultravioleta cercano (8410 - 8795 Å) con un rango de magnitudes $9 < I < 12$, obteniendo unos errores en las medidas de velocidades radiales cuya moda es de unos 2km/s, teniendo el 80% de las estrellas observadas errores inferiores a los 3.4 km/s. A su vez, GES es un cartografiado que obtiene espectros de alta y muy alta resolución ($R \sim 20000$ y $R \sim 47000$) para unas 10^5 estrellas, hasta magnitud $V = 19$. La precisión en velocidades radiales esperada para este cartografiado es de 0.26 km/s (Jackson et al., 2015).

Otros cartografiados como APOGEE (The Apache Pont Observatory Galactic Evolution Experiment, Allende Prieto et al. 2008) u OCCASO (The Open Cluster Chemical Abundances from Spanish Observatories, Casamiquela et al. 2016) aportan valores de velocidades radiales para estrellas situadas en el 'red clump' o estrellas gigantes.

Por tanto, dada la cantidad de información existente actualmente y, la mayor y con mejor precisión que se espera obtener en la era Gaia, parece una evolución natural el desarrollar nuevas herramientas que nos permitan utilizar toda esta información de una manera fácil y flexible, para en nuestro caso, determinar de una forma rigurosa los miembros de los cúmulos abiertos de estrellas. Antes de detallar la metodología que ha sido desarrollada en esta tesis, precisamente con este objetivo, queremos indicar las técnicas que durante las últimas décadas han sido y son utilizadas para la determinación de los miembros de los cúmulos abiertos.

1.5 Metodologías de análisis de pertenencia en cúmulos abiertos.

La determinación de las estrellas que pertenecen a los cúmulos abiertos es un paso inicial fundamental para los subsecuentes análisis que se quieran realizar con ellos. La separación entre los miembros de los cúmulos de estrellas y las comúnmente conocidas como estrellas de campo, tiene una naturaleza estadística basada en las funciones de distribución de las variables a utilizar. Obviamente, las funciones de distribución de estas variables deben contener la información necesaria para que, a partir de un análisis estadístico, podamos diferenciar entre ambas poblaciones de estrellas.

Esto se consigue haciendo uso de características comunes a una determinada población, que en este caso sería la de cúmulo. Como las estrellas de cúmulo nacen de un mismo núcleo denso de una nube molecular (Lada & Lada, 2003), podemos aproximar que todas las estrellas tienen la misma edad, que comparten una pequeña región del espacio, que se encuentran a una misma distancia y que tienen aproximadamente la misma composición química inicial. Desde un punto de vista cinemático, los miembros de un cúmulo presentan una pequeña dispersión en torno a un movimiento común.

Por el contrario, las estrellas de campo pueblan todo el disco galáctico, presentando una distribución más aleatoria en el espacio de posiciones. El origen de estas estrellas de campo es muy heterogéneo, procediendo la gran mayoría de ellas de cúmulos embebidos formados en diferentes momentos de la historia de la Vía Láctea, que al expulsar el gas no alcanzan un equilibrio energético, dejando de estar ligados gravitacionalmente y conllevando la consecuente evaporación del cúmulo. Al no estar ligadas gravitacionalmente, la distribución de velocidades de las estrellas de campo viene descrita por la distribución de velocidades del disco Galáctico. Esta distribución es aproximada por una función elipsoidal anisotrópica, caracterizada por una gaussiana en 3D, siendo comúnmente conocida como distribución de velocidades de Schwarzschild K. (1907). Las dispersiones que caracterizan a este elipsoide de velocidades son mayores que las dispersiones de velocidades de las estrellas que pertenecen a los cúmulos. Esto es debido a que las dispersiones de los movimientos propios de las estrellas de campo son el resultado de la rotación diferencial Galáctica, de movimientos peculiares y de errores observacionales entre otros (Dias, Lépine, & Alessi, 2001), mientras que la dispersión observada de las estrellas de cúmulo es prácticamente debida a errores observacionales al ser su dispersión interna, en la mayoría de los casos, despreciable.

Son precisamente estas diferencias en las funciones de distribución de las variables asociadas tanto al cúmulo como al campo en las que se basan las técnicas, que usan estas variables, para separar ambas poblaciones. El cálculo de las probabilidades de pertenencia al cúmulo radica en la estimación de las funciones de densidad de probabilidad de las variables que se estén utilizando. Existen dos formas diferentes de estimar estas funciones de densidad de probabilidad, lo cual nos permite dividir las metodologías existentes en dos grupos: paramétricas y no-paramétricas.

Las métodos paramétricos asumen un modelo para las funciones de distribución de las variables que se estén utilizando, a partir de las cuales se calculan las probabilidades de pertenencia. El objetivo fundamental de estas metodologías consiste en la determinación de los parámetros de los modelos, en general, a partir de la maximización de la función likelihood (Sanders, 1971). Uno de los modelos históricamente más utilizados es el propuesto por Vasilevskis, Klemola, & Preston (1958), el cual asume que los movimientos propios vienen descritos por funciones de distribución Gaussianas circulares para las estrellas de cúmulo y Gaussianas bivariantes elípticas para las estrellas de campo. Diferentes metodologías se han desarrollado tanto para mejorar los modelos así como las técnicas aplicadas para estimar los parámetros que los definen. Ejemplos son: Zhao & He (1990) que introducen los errores de los movimientos propios directamente en la función de densidad de probabilidad, Zhao & Shao (1994) que mejora la anterior metodología introduciendo un coeficiente de correlación en la distribución de las estrellas de campo, Cabrera-Cañó & Alfaro (1985), Uribe, Barrera, & Brieva (2006) y Monteiro, Dias, & Caetano (2010) que proponen un proceso iterativo Wolfe (Wolfe, 1970), el algoritmo EM (Expectation Maximization algorithm, Dempster et al. 1977) y la técnica CE (Cross Entropy technique, Rubinstein, R. Y. 1997), respectivamente, para estimar los parámetros de los modelos o Su et al. (1995) que además de utilizar los movimientos propios, incluye las funciones de densidad superficial como pesos en las funciones de distribución de los movimientos propios. Otras metodologías incluyen, además de los movimientos propios, información adicional en la estimación de las probabilidades de pertenencia como fotometría, por ejemplo Sarro et al. (2014) o Monteiro, Dias,

& Caetano (2010) que hacen uso de las posiciones y de las variables fotométricas de la estrellas.

El problema subyacente de las metodologías paramétricas es que, precisamente, asumen un determinado comportamiento de las variables que se estén utilizando para calcular las probabilidades de pertenencia al cúmulo. Aunque en muchos casos, la asunción de un modelo da lugar a una buena separación entre las estrellas de campo y cúmulo, diversos efectos pueden producir que la distribución de las variables se aleje del modelo asumido. Tal y como apuntan Cabrera-Caño & Alfaro (1990), la asunción de un modelo para el comportamiento de las variables que estemos utilizando, puede no ser la correcta si tenemos datos con errores grandes o si la distribución cinemática de las estrellas de campo no está bien descrita por la distribución asumida.

Estos problemas han conducido al desarrollo de metodologías no-paramétricas que no asumen un modelo para las funciones de densidad de probabilidad de las variables que utilizan, sino que estas últimas son calculadas empíricamente a partir de funciones Kernel (KDE, Kernel Density Estimation, Silverman, B.W. 1986). Ejemplos de metodologías no-paramétricas que utilizan funciones KDE, en particular Kernels gaussianos, son los trabajos: Cabrera-Caño & Alfaro (1990), el cual además de los movimientos propios utiliza las posiciones de las estrellas para incrementar la distancia estadística entre las estrellas de cúmulo y de campo, Galadi-Enriquez, Jordi, & Trullols (1998) incluye además información fotométrica o Priyatikanto & Arifyanto (2015) quienes introducen los errores de los movimientos propios como pesos en las funciones Kernel. Otros estudios utilizan, además de las posiciones, las variables fotométricas como Dias et al. (2012), Perren, Vázquez, & Piatti (2015), Krone-Martins & Moitinho (2015) y Caetano et al. (2015).

A pesar de las diferencias entre ambos tipos de metodologías, resultados similares han sido obtenidos en diferentes trabajos. Javakhishvili et al. (2006) obtiene resultados similares en una comparativa llevada a cabo tras realizar un estudio de pertenencia basado en una metodología no-paramétrica de carácter acumulativo, con los obtenidos usando la metodología propuesta por Vasilevskis, Klemola, & Preston (1958) para un estudio de 240 cúmulos. Balaguer-Núñez et al. (2004) aplican tanto metodologías paramétricas como no-paramétricas al cúmulo NGC 1817 obteniendo resultados similares usando los dos tipos de técnicas.

En esta tesis hemos desarrollado una nueva metodología de naturaleza paramétrica que trata de resolver problemas asociados a ambos tipos de metodologías. Así, esta nueva metodología no asume un modelo para las variables físicas a utilizar, sino que lo asume para la distribución de distancias calculadas a partir de ellas. La existencia de errores en las variables que estemos utilizando o la existencia de una gran contaminación de estrellas de campo en la región del cúmulo a analizar, puede hacer que las soluciones de los análisis de pertenencia realizados no sean correctas. El poder incluir en el análisis otras variables de una forma fácil y flexible, nos permite incrementar la distancia estadística entre ambas poblaciones de estrellas, optimizando el análisis de pertenencia y pudiendo adaptarlo a las mejores variables del cartografiado que queramos usar.

En el siguiente capítulo comenzaremos describiendo en detalle la nueva metodología geométrica desarrollada en esta tesis. Además, se describirán otras dos metodologías de análisis de pertenencia, una no-paramétrica (Cabrera-Caño & Alfaro, 1990) y otra paramétrica (Cabrera-Caño & Alfaro, 1985)

que han sido ampliamente aplicadas a lo largo de este trabajo así como una técnica de determinación de *outliers* (OUTKER, Cabrera-Caño & Alfaro 1985), paso previo a la realización de los análisis de pertenencia.

2

Metodologías.

2.1 Nueva metodología geométrica N-Dimensional.

2.1.1 Introducción.

En esta tesis desarrollamos una nueva metodología de determinación de probables miembros de un cúmulo estelar a partir de un conjunto de N-variables físicas, medidas sobre las estrellas en la vecindad del cúmulo. El método no tiene en cuenta la naturaleza de estas variables siempre y cuándo se cumpla la hipótesis fundamental de que las mismas estén más densamente concentradas para los miembros del cúmulo que para las estrellas de campo. Así, las variables físicas que conforman el espacio de fase, y/o por ejemplo, la metalicidad, cumplen este requisito para un cúmulo ideal. De aquí que consideremos que, definiendo una métrica en ese espacio de N-dimensiones (N-D), la distribución de distancias al centroide del cúmulo de las estrellas miembros también estará más densamente concentrada que la de las estrellas de campo.

Asumimos que la distribución de distancias puede modelarse como la combinación de dos Gaussianas mono-dimensionales, convirtiendo así un problema N-D en uno 1-D. A partir de ahí, y mediante un proceso iterativo de determinación de los parámetros de las distribuciones, de cálculo de probabilidades de pertenencia, y de estimación del nuevo centroide N-D y de la distribución de distancias, se calculan las probabilidades de pertenencia finales al cúmulo de las estrellas de la muestra, a partir de las cuales, se caracterizan las propiedades del cúmulo en el espacio de N-variables.

2.1.2 Métrica en un espacio de N-Dimensiones

Dada la diferente naturaleza de las variables físicas que pueden ser utilizadas en este problema, y la consiguiente diversidad en el rango de valores que puedan presentar cada una, se hace necesaria la normalización de las mismas. Sin embargo, hemos de notar que la presencia de *outliers* pudiera alterar de forma considerable el proceso de normalización y el posterior análisis (Zhao et al., 1982), por lo que el paso inicial debe ser el de la detección y posterior eliminación de estos objetos.

Para la detección de los *outliers*, hemos utilizado el procedimiento OUTKER propuesto por

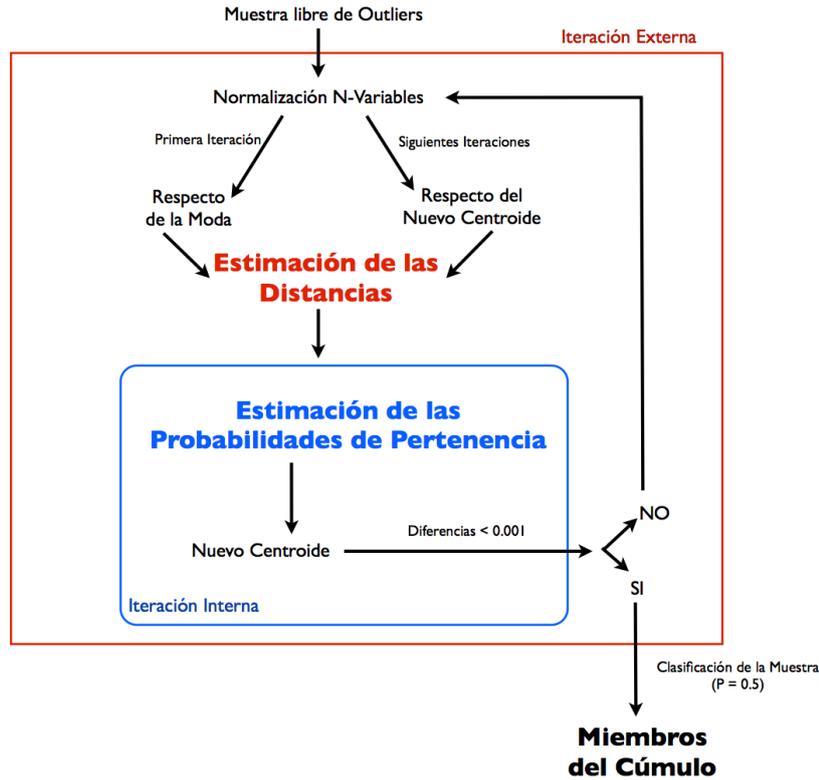


Figura 2.1: Diagrama de flujo esquemático de la nueva metodología propuesta para el análisis de pertenencia a un cúmulo estelar de una muestra de estrellas en su vecindad.

Cabrera-Caño & Alfaro (1985) para la distribución de los movimientos propios (2-D), el cual es fácilmente simplificable a 1-D. Una descripción más detallada de este procedimiento se presenta en el apartado 2.4.1.

Con una muestra libre de *outliers*, la estimación de las probabilidades de pertenencia implica dos procesos iterativos, imbricados uno dentro del otro. La figura 2.1 detalla el diagrama de flujo del método. En el bucle externo se calculan las distancias entre cada estrella y la sobre-densidad central en un espacio N-D. Una vez calculadas, el bucle interno hace uso de ellas para calcular las probabilidades de pertenencia al cúmulo de cada una de las estrellas de la muestra.

En el bucle externo y en la primera iteración del procedimiento, las variables utilizadas en el análisis de pertenencia son normalizadas a partir de su moda y su dispersión, de acuerdo con la expresión:

$$X_{i,j} = \frac{x_{i,j} - x_{j,0}}{\sigma_{x_j}} \quad (2.1)$$

donde la variable matemática $X_{i,j}$ se refiere a la variable física j de la estrella i ($x_{i,j}$), obtenida por la normalización expresada en la ecuación 2.1; $x_{j,0}$ es la moda de la distribución de la variable j y σ_{x_j} es su desviación estándar, calculadas una vez eliminados los *outliers*.

Una vez normalizadas las variables, las distancias de cada estrella (dis_i) al primer centro de refer-

encia se obtienen como el módulo del vector (X_{i_N}) dado por:

$$dis_i = \sqrt{(X_{i_N})^T(M)(X_{i_N})} \quad (2.2)$$

En este trabajo se ha utilizado la métrica euclídea M , la cual otorga el mismo peso a las N-variables utilizadas. Sin embargo, esto puede ser modificado permitiendo asignar diferentes pesos a las N-variables implicadas en el análisis de pertenencia, más acordes con las características del problema.

2.1.3 Probabilidades de pertenencia Bayesianas.

Una vez calculadas las distancias y considerando que éstas se distribuyen como dos Gaussianas, se estiman las soluciones del modelo y las probabilidades de pertenencia de cada estrella por un proceso iterativo (Wolfe, 1970), dentro de lo que hemos denominado bucle interno. Este proceso aborda la clasificación de una muestra inicial de objetos en diferentes clases, haciendo uso de la información disponible para estos objetos. En nuestro caso consideramos que sólo hay dos clases: cúmulo y campo. La variable que nos permite la determinación de las funciones de densidad de probabilidad de ambas poblaciones, la estimación de las probabilidades y la posterior clasificación, son las distancias dadas por la ecuación 2.2 y dependientes de las variables que conforman el vector X_{i_N} y de la métrica elegida. Suponiendo que estas distribuciones son Gaussianas mono-dimensionales y, considerando que los subíndices c y f hacen referencia a las clases cúmulo y campo, respectivamente, podemos describir el modelo de la función de densidad de probabilidad de la distribución de distancias como:

$$\phi_i(dis_i) = n_c \phi_{i,c}(dis_i) + n_f \phi_{i,f}(dis_i) \quad (2.3)$$

donde n_c y n_f son las probabilidades *a priori*, es decir, el número de estrellas de cada grupo respecto del total de la muestra, y $\phi_{i,c}(dis_i)$ y $\phi_{i,f}(dis_i)$, son las funciones de densidad de probabilidad condicionadas para la estrella i relativas al cúmulo y al campo, cuyas expresiones son:

$$\phi_{i,c}(dis_i) = \frac{1}{\sigma_c \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{dis_i - \mu_c}{\sigma_c}\right)^2\right) \quad (2.4)$$

$$\phi_{i,f}(dis_i) = \frac{1}{\sigma_f \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{dis_i - \mu_f}{\sigma_f}\right)^2\right) \quad (2.5)$$

donde dis_i es el valor de la distancia dada por la ecuación 2.2 y, μ_c , σ_c , μ_f y σ_f son los parámetros (medias y desviaciones estándar) de los modelos de ambas poblaciones.

Una vez estimados los parámetros del modelo, la probabilidad de pertenencia al cúmulo viene dada por la expresión:

$$P_i(c/dis_i) = \frac{n_c \phi_{i,c}(dis_i)}{\phi_i(dis_i)} \quad (2.6)$$

Con estas probabilidades se vuelven a calcular los parámetros del modelo de acuerdo a las ecuaciones:

$$n_c = \frac{\sum_{i=1}^{N_T} P_i}{N_T} \quad (2.7)$$

$$\mu_c = \frac{\sum_{i=1}^{N_T} P_i dis_i}{\sum_{i=1}^{N_T} P_i} \quad (2.8)$$

$$\sigma_c = \sqrt{\frac{\sum_{i=1}^{N_T} P_i dis_i^2}{\sum_{i=1}^{N_T} P_i} - \mu_c^2} \quad (2.9)$$

Este proceso se repite de forma iterativa hasta que se alcanza una convergencia en los parámetros del modelo definida por una diferencia (ϵ) entre una iteración y su anterior, inferior a una milésima cuando trabajamos en mas/año. Esto nos garantiza, para las simulaciones analizadas, que la clasificación de la muestra es la misma para las dos iteraciones y que, por tanto, el bucle interior puede darse por finalizado. Este método de resolver el problema de determinación de parámetros de una mezcla de Gaussianas y la posterior clasificación de la muestra en las diferentes clases fue propuesto por Wolfe en 1970 e introducido en la estadística astronómica por Cabrera-Caño y Alfaro en 1985.

Las probabilidades de pertenencia resultantes se utilizarán para calcular la media pesada de cada una de las N-variables usadas en la estimación de las distancias con el objetivo de re-determinar el centroide del cúmulo en el espacio N-D. El nuevo centroide se compara con el anterior (considerad que para la iteración cero el centroide estaba definido por la moda de las variables). Si la diferencia es mayor que un cierto valor límite ($\epsilon > 0.001$), se repite el proceso hasta que la convergencia es alcanzada. El nivel de tolerancia de 0.001, utilizados en ambos bucles, nos garantiza para las simulaciones analizadas, que la clasificación de la muestra es la misma (tanto en el bucle interior como en el exterior) y el proceso ha convergido realmente.

Cabe señalar que, aunque en la primera iteración las distancias son calculadas usando la moda, es decir, respecto de la sobre-densidad observada en el espacio N-D, en las siguientes iteraciones las distancias son calculadas respecto al centroide del cúmulo estimado. Esta forma de proceder es debida a que en la primera iteración no hay ninguna clasificación previa de las estrellas en las dos categorías, y por tanto, no conocemos el centroide del cúmulo. Como primera hipótesis consideramos que las distancias a cualquier punto origen deben tener el máximo del histograma en el centroide del cúmulo. Si por las características de la muestra este máximo corresponde a las estrellas de campo, la situación se corrige en las siguientes iteraciones donde se obliga (por hipótesis de partida), a que la dispersión de las estrellas de cúmulo sea inferior a la de campo. Una vez se ha obtenido la primera distribución de probabilidades obtenemos el centroide del cúmulo como el momento de orden cero de esta distribución, que se convierte así en el nuevo origen para la determinación de distancias de la siguiente iteración.

La utilización de la moda en el cómputo de las distancias en la primera iteración de la metodología, produce una distribución de distancias mejor comportada que la obtenida si usásemos la media o la mediana. Un ejemplo de las distribuciones de distancias obtenidas se puede observar en la figura

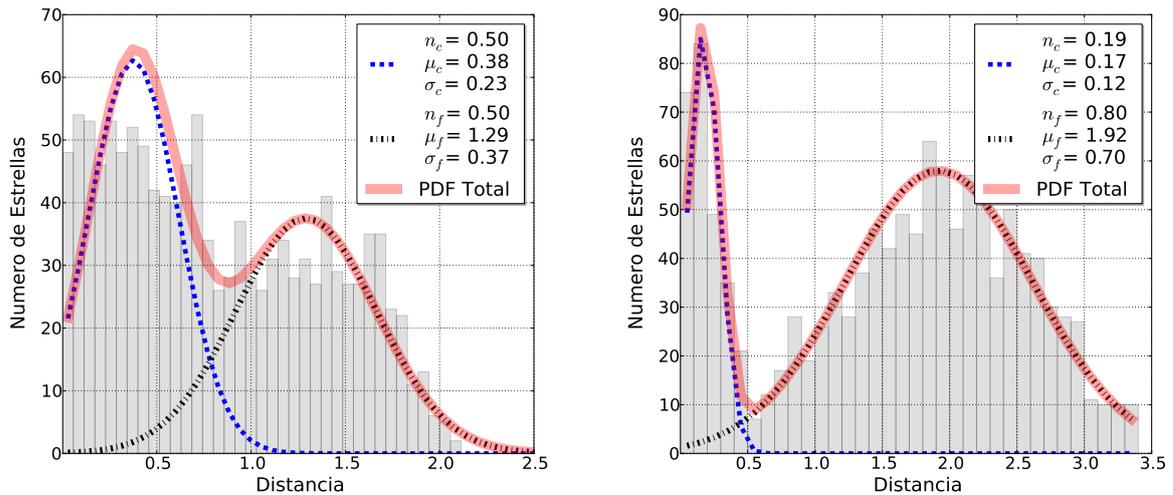


Figura 2.2: Efecto de la determinación del centroide de las variables en la distribución de distancias obtenida. En el gráfico izquierdo se muestra la distribución de distancias haciendo uso de la media de las variables en el cálculo del centroide, mientras que, en el gráfico derecho se ha realizado respecto de la moda de las variables. La utilización de la moda produce una distribución de distancias a partir de la cual la metodología puede realizar una mejor clasificación de la muestra en cúmulo y campo.

2.2, donde en el gráfico de la izquierda se muestra la distribución de distancias obtenida cuando se ha utilizado la media y en el de la derecha, la moda. Es evidente que el valor de las medias de las variables no es representativo de la población de cúmulo.

En la figura 2.3 se muestra un ejemplo de una distribución de distancias y el ajuste a dos Gaussianas 1-D, en el que las posiciones y los movimientos propios son las variables simuladas utilizadas para llevar a cabo el análisis de pertenencia (la descripción completa de las simulaciones se puede encontrar en el capítulo 3). La línea azul discontinua representa el ajuste Gaussiano para la distribución de distancias asociada a las estrellas de cúmulo, mientras que la línea negra punteada discontinua representa la distribución de distancias asociada a las estrellas de campo. La función de densidad de probabilidad del modelo de la muestra total se ha representado por una línea roja continua superpuesta. La leyenda relata los valores de los parámetros de las Gaussianas de ambas poblaciones, que mejor ajustan a la distribución de distancias y que fueron utilizados para el cálculo de las probabilidades de pertenencia. Como se puede observar en la figura, el ajuste Gaussiano de la distribución de distancias del cúmulo muestra una media más pequeña y una dispersión más estrecha que aquellas que caracterizan la distribución de distancias del campo. Éste es un resultado esperado dado que las distancias están calculadas con respecto al centroide del cúmulo y, por tanto, las estrellas asociadas al mismo tendrán distancias más pequeñas. Además, la Gaussiana asociada con la población de cúmulo presenta una menor dispersión debido a que las variables que estamos utilizando están más densamente distribuidas para el cúmulo que para el campo.

Por tanto, al final del proceso la metodología determina las probabilidades de pertenencia al cúmulo de cada una de las estrellas de la muestra. Una vez calculadas estas probabilidades de pertenencia,

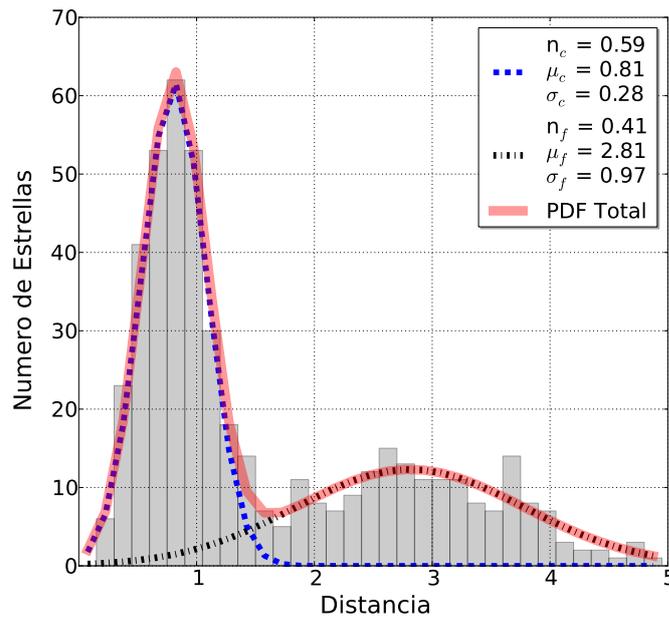


Figura 2.3: Ejemplo de ajuste a una distribución de distancias por dos Gaussianas mono-dimensionales. La leyenda muestra los mejores ajustes obtenidos por la metodología para ambas distribuciones. La línea azul discontinua hace referencia a la distribución del cúmulo y la línea negra punteada discontinua a la distribución de las estrellas de campo. La función de densidad de probabilidad de la muestra total está superpuesta en rojo.

es necesario tomar una decisión en el valor umbral de la probabilidad a partir del cual una estrella será clasificada como miembro del cúmulo. Intentando conseguir la menor subjetividad y la mayor homogeneización posible en la toma de decisiones, utilizamos la regla de decisión de Bayes que minimiza el error en la clasificación (Kulkarni & Harman, 2011), que propone un valor umbral en la probabilidad de pertenencia de 0.5 para separar la muestra en ambas poblaciones. Así, una estrella con un valor en la probabilidad de pertenencia superior a 0.5 será clasificada como miembro del cúmulo. Obviamente, la teoría de decisiones nos permite establecer otros criterios dependiendo de las necesidades del estudio.

El nuevo método de determinación de pertenencia a un cúmulo se ha programado en Python. La metodología requiere la introducción de unos valores iniciales razonables del modelo de mezcla de las distancias. Para facilitar su uso lo hemos configurado de tal forma que, todos los parámetros necesarios además de las posiciones en el catálogo de las variables a utilizar para la aplicación del método, se introducen en un fichero inicial. Un ejemplo del mismo se muestra en la tabla 2.1.

2.2 Método No-Paramétrico

Esta metodología, diseñada por Cabrera-Caño & Alfaro (1990), tiene el mismo objetivo, la clasificación de una muestra inicial de estrellas en dos grupos: miembros del cúmulo y estrellas de campo, a partir en este caso de la distribución de movimientos propios y coordenadas angulares. Determina las funciones

Tabla 2.1: Ejemplo de archivo de configuración utilizado por la metodología geométrica.

VARIABLES	POSICIÓN
Id	0
X	1
Y	2
μ_x	3
μ_y	4
VR	5
PARALAJE	6

VARIABLES PARA LA ESTIMACIÓN DE LOS <i>OUTLIERS</i>	POSICIÓN
μ_x	3
μ_y	4

VARIABLES PARA LA ESTIMACIÓN DE LAS DISTANCIAS	POSICIÓN
X	1
Y	2
μ_x	3
μ_y	4

PARÁMETROS INICIALES DEL AJUSTE DE LA DISTRIBUCIÓN DE DISTANCIAS	VALORES
LIM	5000
ϵ	0.001
μ_{x_c}	3
σ_{x_c}	0.5
n_c	0.4
μ_{x_f}	2.2
σ_{x_f}	1
n_f	0.6

UMBRAL DE LA PROBABILIDAD PARA CLASIFICAR LA MUESTRA	VALOR
P	0.5

de densidad de probabilidad de cada clase de una forma empírica, utilizando kernel Gaussianos, sin hacer ninguna hipótesis acerca de los modelos subyacentes a cada población. La única condición que se debe cumplir es que las variables utilizadas estén más densamente concentradas para las estrellas de cúmulo que para las de campo.

Para poder llevar a cabo este análisis se necesitan dos muestras de objetos: un conjunto pre-clasificado obtenido haciendo uso de información *a priori*, y otro que se quiere clasificar a partir de las características observadas de estos objetos. El análisis discriminante compara las características observadas (variables) de cada uno de los objetos no clasificados con aquellos que ya han sido previamente clasificados y, mediante una toma de decisiones definidas por unos umbrales de probabilidad en diferentes espacios (lo que podríamos llamar nuestra definición de cúmulo), se asignan a una clase w_i .

Para obtener el conjunto pre-clasificado inicial, y dado que los miembros potenciales de los cúmulos de estrellas están situados, con una mayor probabilidad, en regiones de alta densidad de probabilidad en el espacio de medida, podemos hacer una primera clasificación inicial de la muestra en las dos clases haciendo uso de esta información. Así, una vez calculada la función de densidad de probabilidad de la muestra, se puede tomar un valor en la misma a partir del cual realizar la clasificación. Como hipótesis de partida, el valor de la función de densidad de probabilidad escogido se corresponde con aquel que contiene el 40% de las estrellas de la muestra con una mayor densidad de probabilidad, siendo inicialmente clasificadas como estrellas de cúmulo. Como resultado se obtiene la pre-clasificación inicial de la muestra en las dos clases, requerida para realizar el análisis discriminante. La elección del 40% es un valor intermedio que no da un peso muy significativo a cualquiera de los dos grupos, y que introduce un menor coste computacional no teniendo, en la mayoría de los casos, un mayor efecto en el proceso de clasificación llevado a cabo por este método.

En esta metodología las funciones de densidad de probabilidad se estiman empíricamente a partir del uso de funciones kernel Gaussianas, donde se excluye del cálculo de la función de densidad de probabilidad aquel objeto sobre el que se esté calculando la misma. Consideramos aquí que las variables de posición (x e y) y las variables cinemáticas (los movimientos propios, μ_x y μ_y), son variables independientes, lo cual permite calcular la función de densidad de probabilidad conjunta de la siguiente manera:

$$f(\mu_x, \mu_y, x, y) = f_\mu(\mu_x, \mu_y) \cdot f_r(x, y) \quad (2.10)$$

donde:

$$f_\mu(\mu_{x_i}, \mu_{y_i}) = \frac{1}{(n-1)h_\mu^2} \sum_{j=1, i \neq j}^n \frac{1}{2\pi} \exp\left(-\frac{1}{2} \left[\left(\frac{\mu_{x_j} - \mu_{x_i}}{h_\mu} \right)^2 + \left(\frac{\mu_{y_j} - \mu_{y_i}}{h_\mu} \right)^2 \right] \right) \quad (2.11)$$

$$f_r(x_i, y_i) = \frac{1}{(n-1)h_r^2} \sum_{j=1, i \neq j}^n \frac{1}{2\pi} \exp\left(-\frac{1}{2} \left[\left(\frac{x_j - x_i}{h_r} \right)^2 + \left(\frac{y_j - y_i}{h_r} \right)^2 \right] \right) \quad (2.12)$$

representan las funciones de densidad de probabilidad calculadas a partir de los movimientos propios y de las posiciones, respectivamente. Estas dos ecuaciones son válidas tanto para las estrellas de cúmulo

como para las estrellas de campo, de forma que al final, este método calcula 4 funciones de densidad de probabilidad diferentes, cuyas ecuaciones son:

$$f_{\mu}(\mu_{x_i}, \mu_{y_i}) = \frac{1}{(N1)h_{\mu}^2} \sum_{j=1, j \neq i, j \in \text{Cúmulo}}^n \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left[\left(\frac{\mu_{x_j} - \mu_{x_i}}{h_{\mu}}\right)^2 + \left(\frac{\mu_{y_j} - \mu_{y_i}}{h_{\mu}}\right)^2\right]\right) \quad (2.13)$$

$$f_{\mu}(\mu_{x_i}, \mu_{y_i}) = \frac{1}{(N2)h_{\mu}^2} \sum_{j=1, j \neq i, j \in \text{Campo}}^n \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left[\left(\frac{\mu_{x_j} - \mu_{x_i}}{h_{\mu}}\right)^2 + \left(\frac{\mu_{y_j} - \mu_{y_i}}{h_{\mu}}\right)^2\right]\right) \quad (2.14)$$

$$f_r(x_i, y_i) = \frac{1}{(N1)h_r^2} \sum_{j=1, j \neq i, j \in \text{Cúmulo}}^n \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left[\left(\frac{x_j - x_i}{h_r}\right)^2 + \left(\frac{y_j - y_i}{h_r}\right)^2\right]\right) \quad (2.15)$$

$$f_r(x_i, y_i) = \frac{1}{(N2)h_r^2} \sum_{j=1, j \neq i, j \in \text{Campo}}^n \frac{1}{2\pi} \exp\left(-\frac{1}{2}\left[\left(\frac{x_j - x_i}{h_r}\right)^2 + \left(\frac{y_j - y_i}{h_r}\right)^2\right]\right) \quad (2.16)$$

donde n es el número de estrellas de la muestra y $N1$ y $N2$ son factores de normalización cuyo valor dependerá de si la estrella i ha sido clasificada previamente como miembro del cúmulo o de campo. Sea n_c el número de estrellas de cúmulo, entonces:

- si la estrella i ha sido clasificada como estrella de cúmulo:

$$N1 = n_c - 1 \quad (2.17)$$

$$N2 = n - n_c \quad (2.18)$$

-si la estrella i fue clasificada como estrella de campo:

$$N1 = n_c \quad (2.19)$$

$$N2 = n - n_c - 1. \quad (2.20)$$

Si nos fijamos en las ecuaciones 2.13 - 2.16, la contribución de una estrella j a la función de densidad de probabilidad de otra estrella i depende de si en la muestra pre-clasificada, la estrella j fue clasificada como estrella de campo o estrella de cúmulo. De esta forma podemos conocer como la estrella i está distribuida respecto de las estrellas que describen las clases de cúmulo y campo y, a partir de aquí, poder ser clasificada en una de las dos clases. Como se puede observar en estas mismas ecuaciones aparecen los parámetros de suavizado h_{μ} y h_r . La elección de estos parámetros juega un papel fundamental para establecer un balance entre la precisión y la incertidumbre de las funciones de densidad de probabilidad obtenidas. En el trabajo inicial se optó por obtener los valores de h_{μ} y h_r que maximizaban la verosimilitud para cada par de variables (Cabrera-Caño & Alfaro, 1990). Para el caso particular de los movimientos propios, la ecuación a maximizar es la siguiente:

$$\log(L(h_\mu)) = \sum_{i=1}^n \log(f_\mu(\mu_{x_i}, \mu_{y_i})) \quad (2.21)$$

Un ejemplo de las funciones de densidad de probabilidad calculadas por esta metodología, se representan en la figura 2.4. A partir de ellas, se calcula la probabilidad de pertenencia al cúmulo de cada objeto, considerando que la muestra total se puede definir a partir de su función de densidad de probabilidad y que cada objeto sólo puede pertenecer a una y sólo a una clase, de acuerdo con la expresión:

$$f(X_j) = \sum_{i=1}^2 p(w_i) f(X_j/w_i) \quad (2.22)$$

donde $f(X_j)$ es la función de densidad de probabilidad de la muestra total, $p(w_i)$ es la probabilidad *a priori*, es decir, la probabilidad de que un objeto pertenezca a la clase w_i sin conocer sus propiedades, y $f(X_j/w_i)$ es la probabilidad condicionada, es decir, la probabilidad de que dado un vector de medida X_j , ese objeto pertenezca a la clase w_i . Las probabilidades de pertenencia son calculadas a partir de la aplicación del teorema de Bayes:

$$p(w_i/X_j) = \frac{f(X_j/w_i)p(w_i)}{f(X_j)} \quad (2.23)$$

donde $p(w_i/X_j)$ es la probabilidad de que un objeto pertenezca a la clase w_i cuando su vector de medidas es X_j , $p(w_i)$ es la probabilidad *a priori* y $f(X_j/w_i)$ es la probabilidad condicionada.

La hipótesis de considerar las posiciones y los movimientos propios como variables independientes, nos permite calcular tres probabilidades de pertenencia diferentes, en función de las variables utilizadas en el cálculo de las funciones de densidad de probabilidad: $P(x, y)$ calculada haciendo uso únicamente de las posiciones de las estrellas; $P(\mu_x, \mu_y)$ denominada *probabilidad cinemática*, calculada usando únicamente los movimientos propios, y $P(x, y, \mu_x, \mu_y)$ calculada utilizando tanto las posiciones como los movimientos propios, la cual se denomina *probabilidad conjunta*. Estas dos últimas probabilidades son las utilizadas para realizar una nueva clasificación de las estrellas. Para hacer esta clasificación es necesario elegir un umbral en el valor de la probabilidad a partir del cual los objetos serán clasificados en uno de los dos grupos. Al igual que en la metodología geométrica el criterio utilizado es la regla de decisión de Bayes que minimiza el error en la clasificación. Consideramos que una estrella pertenece al cúmulo, si su probabilidad cinemática y su probabilidad conjunta son mayores que 0.5, dando, de esta forma, un mayor peso a las variables cinemáticas que a las posicionales. Al final del proceso, se obtiene una nueva clasificación de las estrellas, la cual es comparada con la clasificación anterior. Si ambas clasificaciones no coinciden, se hace una nueva iteración del proceso donde ahora la pre-clasificación a utilizar es la clasificación final obtenida. En caso contrario, la convergencia es alcanzada y tendremos la muestra final clasificada en las dos poblaciones de estrellas.

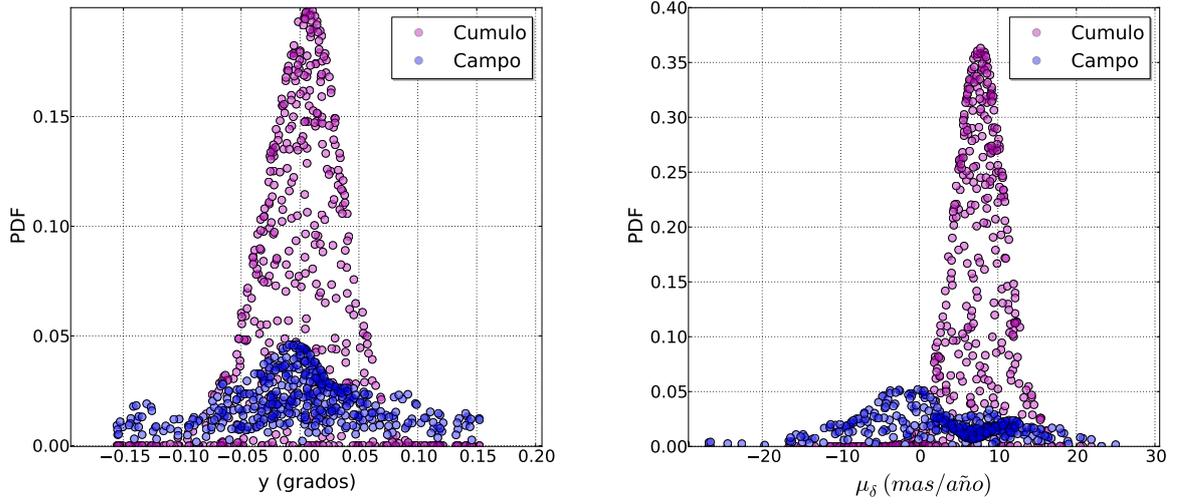


Figura 2.4: Funciones de densidad de probabilidad estimadas por la metodología no-paramétrica en el espacio de posiciones (gráfico izquierdo) y en el de movimientos propios (gráfico derecho).

2.3 Método Paramétrico

2.3.1 Metodología Paramétrica 2-D

Este método, cuya versión utilizada en esta tesis es la desarrollada por Cabrera-Caño & Alfaro (1985), calcula las probabilidades de pertenencia al cúmulo a partir de un modelo de la muestra formado por dos Gaussianas bivariantes, que describen las distribuciones de los movimientos propios (μ_x , μ_y) de las estrellas de campo y cúmulo, respectivamente:

$$\phi(\mu_x, \mu_y) = n_f \phi_f(\mu_x, \mu_y) + n_c \phi_c(\mu_x, \mu_y) \quad (2.24)$$

donde n_f y n_c son las probabilidades *a priori* de pertenecer al campo o al cúmulo, respectivamente, y $\phi_f(\mu_x, \mu_y)$ y $\phi_c(\mu_x, \mu_y)$ son las funciones de densidad de probabilidad descritas por las ecuaciones:

$$\phi_c(\mu_x, \mu_y) = \frac{1}{2\pi\sigma_{cx}\sigma_{cy}\sqrt{1-\rho_c^2}} \exp\left(\frac{-1}{2\sqrt{1-\rho_c^2}} \left[\left(\frac{\mu_x - \mu_{cx}}{\sigma_{cx}}\right)^2 + \left(\frac{\mu_y - \mu_{cy}}{\sigma_{cy}}\right)^2 - 2\rho_c \left(\frac{\mu_x - \mu_{cx}}{\sigma_{cx}}\right) \left(\frac{\mu_y - \mu_{cy}}{\sigma_{cy}}\right) \right]\right) \quad (2.25)$$

$$\phi_f(\mu_x, \mu_y) = \frac{1}{2\pi\sigma_{fx}\sigma_{fy}\sqrt{1-\rho_f^2}} \exp\left(\frac{-1}{2\sqrt{1-\rho_f^2}} \left[\left(\frac{\mu_x - \mu_{fx}}{\sigma_{fx}}\right)^2 + \left(\frac{\mu_y - \mu_{fy}}{\sigma_{fy}}\right)^2 - 2\rho_f \left(\frac{\mu_x - \mu_{fx}}{\sigma_{fx}}\right) \left(\frac{\mu_y - \mu_{fy}}{\sigma_{fy}}\right) \right]\right) \quad (2.26)$$

donde σ_{cx} , σ_{cy} , σ_{fx} y σ_{fy} son las dispersiones de las estrellas de cúmulo y de campo, μ_{cx} , μ_{cy} , μ_{fx} y μ_{fy} son los centroides de los movimientos propios de las estrellas de cúmulo y campo, y por último, ρ_c y ρ_f son los coeficientes de correlación de los dos tipos de estrellas.

Esta metodología, al igual que la nueva metodología geométrica propuesta en esta tesis, utiliza el proceso de estimación iterativa Wolfe, ya descrito previamente. Brevemente recordar que este proceso se basa en el cálculo de los parámetros que definen las funciones de densidad de probabilidad de ambas poblaciones haciendo uso de las probabilidades de pertenencia. Así, partiendo de unos valores iniciales razonables de los parámetros de los dos modelos y, aplicando el teorema de Bayes, las probabilidades de pertenencia de las estrellas son determinadas. Estas probabilidades de pertenencia son utilizadas para re-calcular los parámetros de los modelos, haciendo uso de las ecuaciones siguientes:

$$n_c = \frac{\sum_{i=1}^{N_T} P_i}{N_T} \quad (2.27)$$

$$\mu_{cx} = \frac{\sum_{i=1}^{N_T} P_i \mu_{x_i}}{\sum_{i=1}^{N_T} P_i} \quad (2.28)$$

$$\mu_{cy} = \frac{\sum_{i=1}^{N_T} P_i \mu_{y_i}}{\sum_{i=1}^{N_T} P_i} \quad (2.29)$$

$$\sigma_{cx} = \sqrt{\frac{\sum_{i=1}^{N_T} P_i \mu_{x_i}^2}{\sum_{i=1}^{N_T} P_i} - \mu_{cx}^2} \quad (2.30)$$

$$\sigma_{cy} = \sqrt{\frac{\sum_{i=1}^{N_T} P_i \mu_{y_i}^2}{\sum_{i=1}^{N_T} P_i} - \mu_{cy}^2} \quad (2.31)$$

$$\rho_c = \left(\frac{\sum_{i=1}^{N_T} P_i \mu_{x_i} \mu_{y_i}}{\sum_{i=1}^{N_T} P_i} - \mu_{cx} \mu_{cy} \right) / \sigma_{cx} \sigma_{cy} \quad (2.32)$$

donde P_i es la probabilidad de pertenencia de la estrella i y N_T , el número total de estrellas. Para la obtención de los parámetros de las distribuciones de campo, las probabilidades de pertenencia utilizadas, son las respectivas a la pertenencia al campo.

El criterio de convergencia utilizado es de nuevo de 0.001, de forma que, si la diferencia entre los valores de cada uno de los parámetros entre dos iteraciones consecutivas es superior a este valor, otra iteración es realizada pero ahora haciendo uso de los nuevos parámetros estimados. Si, por el contrario, esta diferencia es inferior a una milésima, el método converge y obtenemos las probabilidades de pertenencia de cada una de las estrellas de la muestra, junto con los parámetros de los modelos que definen las dos poblaciones. En la figura 2.5 se muestra la distribución de las probabilidades de pertenencia al cúmulo, obtenidas en una simulación, en el espacio de los movimientos propios.

De nuevo, al igual que en las anteriores metodologías, se ha utilizado la regla de decisión de Bayes para minimizar el error de clasificación con el objetivo de clasificar la muestra inicial en miembros de cúmulo y estrellas de campo.

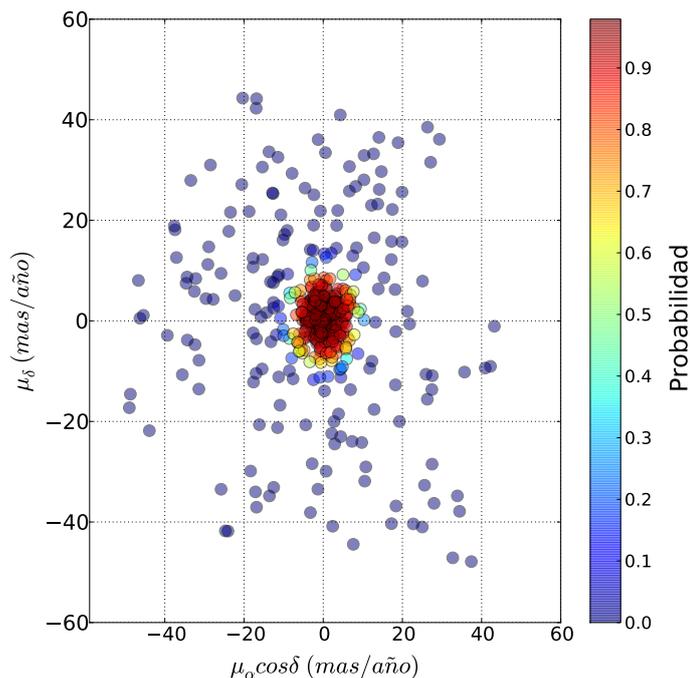


Figura 2.5: Ejemplo de distribución de probabilidad en el espacio de los movimientos propios, obtenida en una de las simulaciones realizadas, por la metodología paramétrica.

2.3.2 Metodología Paramétrica 1-D

Este método es una adaptación del anterior cuando se utiliza una única variable. Así, aproxima la distribución de la muestra total como una mezcla de dos funciones Gaussianas mono-variantes pesadas por la probabilidad de pertenencia *a priori* de ambas clases.

En este caso, al igual que en la metodología geométrica, el número de parámetros a determinar es de 5: los centroides y dispersiones para ambos grupos, y el porcentaje de estrellas perteneciente al cúmulo. El algoritmo de resolución del problema y las tomas de decisiones se llevan a cabo utilizando los mismos criterios y valores que en los casos anteriores.

2.4 Otras técnicas aplicadas.

2.4.1 Determinación de los *outliers*. Procedimiento OUTKER.

La determinación de los posibles *outliers* y su posterior eliminación de la muestra, es un paso inicial fundamental para un correcto análisis de pertenencia. Su grado de importancia varía de unas metodologías a otras. Así, la metodología no-paramétrica no vería muy afectados sus resultados ya que estos objetos contribuirían poco a la determinación de las funciones de densidad de probabilidad de ambas poblaciones. En el caso de las metodologías paramétricas, estos objetos pueden sesgar el

análisis de pertenencia al ensanchar la distribución de las estrellas de campo y, consecuentemente, clasificar un mayor número de estrellas como miembros del cúmulo (Zhao et al. 1982; Cabrera-Caño & Alfaro 1985). En el caso de la metodología geométrica, la no eliminación de los *outliers* influye desde el principio en el proceso de normalización de las variables.

Por tanto, previamente a cualquier análisis de pertenencia realizado en esta tesis, se determinan y eliminan aquellas estrellas con una alta probabilidad de ser *outliers*. Sobre la muestra resultante se aplicarán las diferentes metodologías anteriormente descritas.

La técnica aplicada para su determinación es conocida como procedimiento OUTKER, propuesto por Cabrera-Caño & Alfaro (1985) para la distribución de los movimientos propios (2-D), el cual fue simplificado a 1-D para poder determinar los *outliers* cuando el análisis de pertenencia se realiza utilizando una única variable. Es importante hacer notar que un *outlier* es uno de los pocos conceptos matemáticos que no están rigurosamente definidos. De hecho, no se puede decir de una forma determinista que un objeto es un *outlier*, sobre todo, si la única información que tenemos es la propia muestra de objetos. Sin embargo, podemos determinar la probabilidad de que un objeto de la muestra sea un *outlier* de la función de densidad de probabilidad definida por la muestra en sí misma. La mejor aproximación a la definición de *outlier* es la de un objeto localizado en una región de baja densidad de probabilidad del espacio. Para su detección, el procedimiento OUTKER compara la densidad de probabilidad observada para cada objeto con la función de densidad de probabilidad esperada para la muestra total. Así, obtenemos la probabilidad de ser un *outlier* para cada objeto de la muestra. Dadas estas probabilidades podemos decidir que objetos eliminamos o no de la muestra inicial. De nuevo, el valor de la probabilidad elegido viene dado por la regla de decisión de Bayes para minimizar el error de la clasificación. Aquellos objetos con una probabilidad de ser un *outlier* superior a 0.5 serán eliminados del análisis. Dado que las muestras utilizadas están seleccionadas espacialmente aplicamos el purgado de datos solamente a las variables cinemáticas.

3

Simulaciones.

3.1 Introducción

En este capítulo describimos un conjunto de simulaciones diseñadas para evaluar las capacidades de la nueva metodología desarrollada, así como su rango de aplicación. Estas simulaciones recrean una región de cielo que incluye dos poblaciones estelares, con diferentes funciones de distribución en el espacio de fase: un campo galáctico y un cúmulo estelar. Las simulaciones se llevaron a cabo en los espacios de posiciones y de movimientos propios (un sub-espacio del espacio de fase) usando diferentes conjuntos de parámetros que abarcan la mayoría de la casuística observada. ¿Cómo elegir este rango de parámetros? Para ello acudimos a las grandes bases de datos de cúmulos estelares y catálogos de movimientos propios actualmente disponibles. En particular, hacemos uso de los resultados del trabajo de Dias et al. (2014) (D14) donde, utilizando datos del catálogo astrométrico UCAC4 (Zacharias et al., 2013), Dias y colaboradores analizaron una muestra de 1805 cúmulos abiertos. En D14 se enumeran los parámetros estimados asociados a las distribuciones de las variables astrométricas para las dos poblaciones: es decir, centroide, matriz de covarianza y fracción de estrellas para cada población, así como el radio estimado del cúmulo. Estos datos junto con los publicados en Dias et al. (2002) (a partir de ahora DAML02), conforman el espacio de parámetros del sistema de cúmulos abiertos galácticos. Sin pérdida de representatividad hemos eliminado unos pocos cúmulos donde el número inicial de objetos en el campo es superior a 5000 y la muestra queda entonces reducida a 1646 cúmulos.

Con estos datos hemos analizado las distribuciones estadísticas de los diferentes parámetros, considerando que las simulaciones de entrada serán más representativas del mundo real, si se ajustan al cúmulo 'típico' descrito por el valor central de las distribuciones de los parámetros reales. De igual forma, los errores simulados también deberían ser similares a aquellos que observamos en los catálogos disponibles actualmente. Aquí haremos dos distinciones: a) consideraremos que los errores de los movimientos propios de la muestra siguen una distribución Gaussiana, con independencia de otras variables estelares, y b) que la distribución de errores en los movimientos propios varía con la magnitud de la estrella.

Pero quizá la principal fuente de error en el análisis de pertenencia a cúmulos estelares a partir

de datos de los catálogos astrométricos disponibles, viene más de la distribución de las estrellas en el cielo que de la distribución en movimientos propios. Como veremos más adelante, la mayoría de los cúmulos están sub-muestreados y el sub-muestreo no es homogéneo. Esto implica que en algunas ocasiones no se llegue a observar un claro contraste en densidad superficial entre el cúmulo y el campo. La intensidad y la naturaleza del sub-muestreo es variable, por lo que si queremos simular un cúmulo realista tendremos que tener en cuenta también estos posibles sesgos y trabajar con datos sin un contraste de densidad del cúmulo bien definido y con diferentes porcentajes de muestreo.

3.2 Parámetros del sistema de cúmulos estelares en los catálogos de Dias y colaboradores.

En este apartado analizamos la distribución de los parámetros de las poblaciones estelares en el campo del cúmulo derivados del análisis de Dias y colaboradores. En los gráficos de la figura 3.1 mostramos la distribución del número de estrellas en la muestra, de su distancia al Sol y de su radio en parsecs (pc). La mayoría de estas distribuciones muestran una gran asimetría por lo que el valor central definido por la media, la mediana o la moda no son coincidentes y debemos tomar decisiones acerca de cuál sería el más representativo para cada parámetro en nuestras simulaciones. Para el número de objetos en el catálogo de entrada para cada cúmulo, hemos tomado un valor representativo de 500 objetos. La distribución de este parámetro para 1646 cúmulos del trabajo D14 se muestran en el gráfico superior izquierdo. De igual forma, la moda en la distribución de distancias de los objetos catalogados se encuentra alrededor de los 1250 pc, mientras que su radio típico es de 2 pc (ver gráficos superior derecho e inferior de la figura 3.1). Con estos valores obtenemos un radio angular típico de 0.093° o equivalentemente 5.5 minutos de arco (ma).

El espacio de movimientos propios caracterizado por los 1646 cúmulos de la muestra con análisis de pertenencia previo, se describe por dos Gaussianas bivariantes, una circular para el cúmulo y otra elíptica para el campo. La distribución de la dispersión de movimientos propios para la población de campo en nuestra muestra patrón se representa en la figura 3.2. El gráfico de la izquierda presenta el cociente entre ambas componentes diagonales de la matriz de covarianza de los movimientos propios del campo y el de la izquierda muestra como se distribuye el elemento (1,1) de esta matriz.

La dispersión de velocidades de las estrellas del cúmulo vendrá caracterizada primeramente por la esperada para un equilibrio virial:

$$\sigma_{vel}^2 \simeq \frac{GM}{R} \quad (3.1)$$

donde G es la constante de gravitación universal, M es la masa total del cúmulo y R es su radio. Para los valores representativos de nuestro cúmulo típico (Masa = $250 M_\odot$, $R = 2$ pc) se obtiene un valor de $\sigma_{vel,mean} = 0.42$ km/s por grado de libertad, que será considerado como el valor representativo a partir del cual se simulará la dispersión interna de velocidades del cúmulo. Considerando que nuestro cúmulo típico se encuentra a una distancia de 1250 pc, obtenemos un valor para la dispersión interna

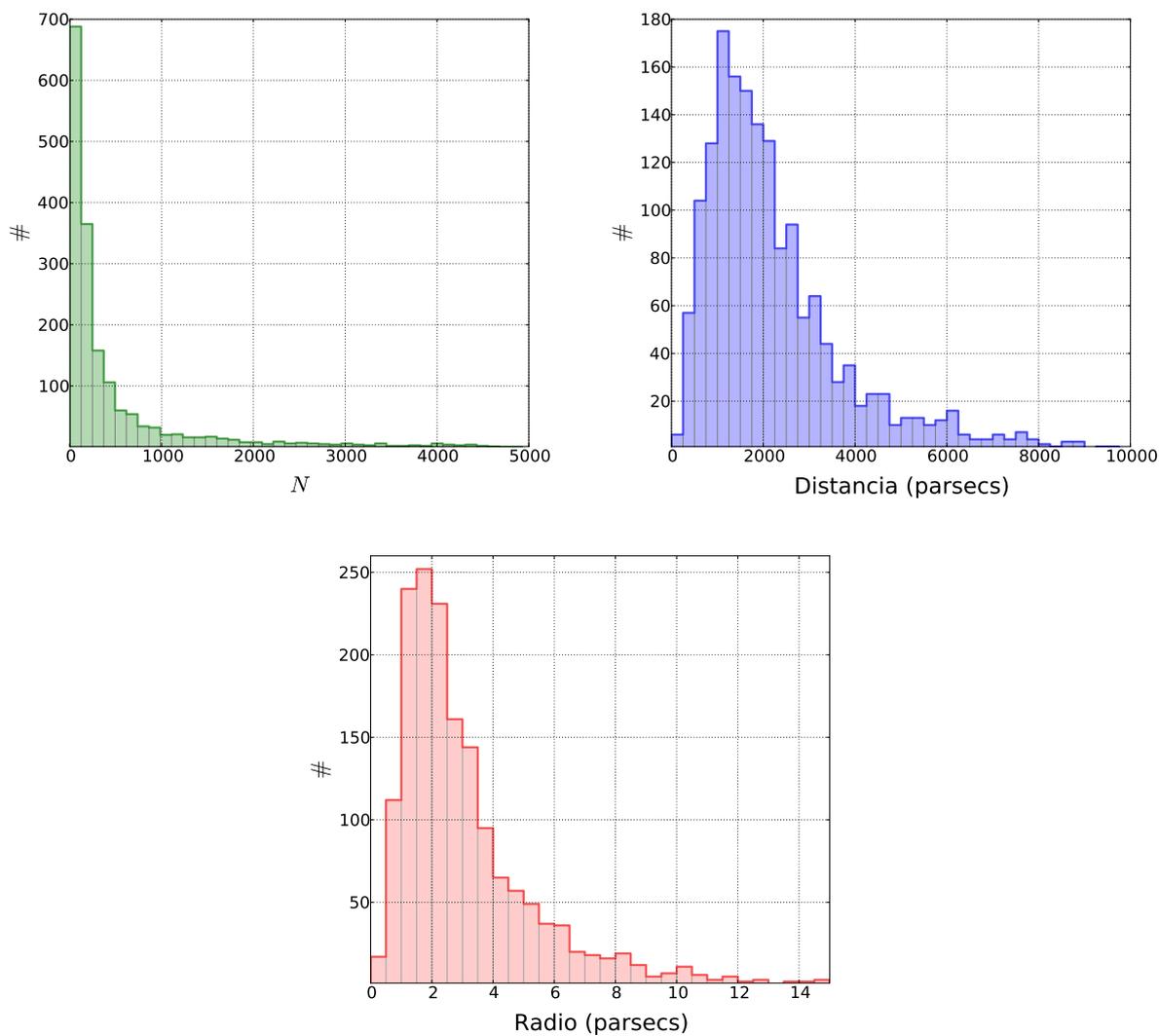


Figura 3.1: En los gráficos superiores se muestran las distribuciones del número de objetos y de distancias (en pc) de la muestra de 1646 cúmulos obtenida a partir de los trabajos DAML02 y D14. En el gráfico inferior se representa la distribución de los radios en pc. A partir de estos valores determinamos que un cúmulo típico se encuentra a 1250 pc y que tiene un radio de unos 5.5 minutos de arco.

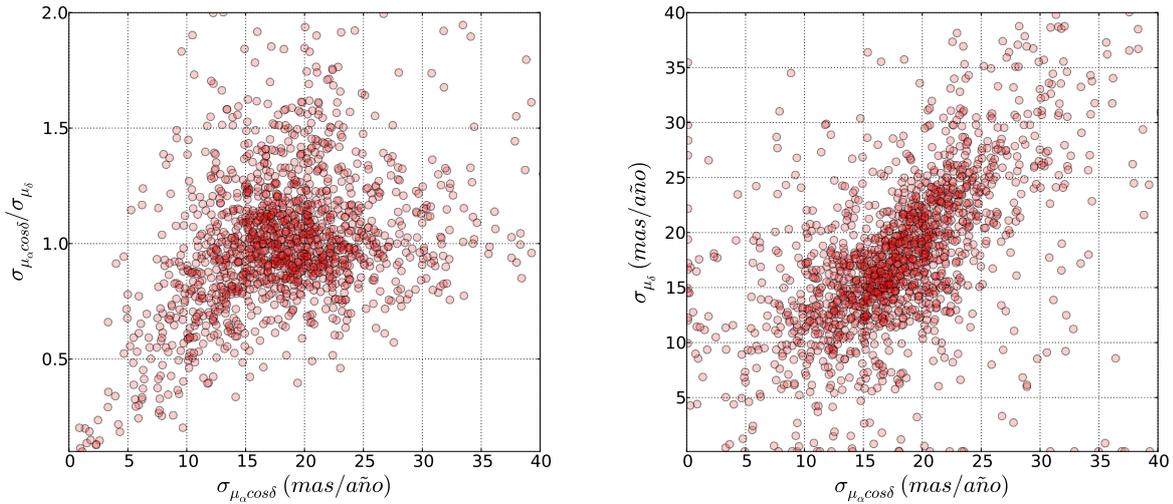


Figura 3.2: Distribución de las dispersiones de las estrellas de campo de los 1646 cúmulos, donde se justifica el rango de valores utilizados para estos parámetros en nuestras simulaciones.

del cúmulo en movimientos propios de $\sigma_{\mu c} = 0.07$ mas/año, valor calculado a partir de la ecuación:

$$\sigma_{\mu} = \frac{\sigma_{vel,mean}}{kd} \quad (3.2)$$

donde d es la distancia al cúmulo y k es una constante que toma el valor 4.74 km/s cuando la distancia se mide en parsecs, la velocidad radial en km/s y la velocidad tangencial en arco-segundos/año (as/año).

Sin embargo, la precisión de los movimientos propios catalogados (ej. UCAC4) está muy por encima del valor intrínseco de la dispersión de velocidades de origen dinámico por lo que la distribución observada de los movimientos propios de la población de cúmulo, está principalmente dominada por errores observacionales. Una vez más, utilizando los valores catalogados en el trabajo D14, y considerando que la dispersión de los movimientos propios de la población de cúmulo es representativa de la incertidumbre real de los movimientos propios catalogados, observamos que la desviación estándar de los movimientos propios de los cúmulos listados en nuestro catálogo patrón, presenta un máximo en torno a 3 mas/año (figura 3.3), valor que será tomado como valor representativo de la incertidumbre de los movimientos propios como base para nuestras simulaciones.

En lo que sigue describiremos las simulaciones realizadas, donde detallaremos primero los parámetros utilizados en cada aproximación. Con estas simulaciones pretendemos describir de una forma realista, cómo es un cúmulo estelar, en este sub-espacio del espacio de fase, en la era pre-Gaia.

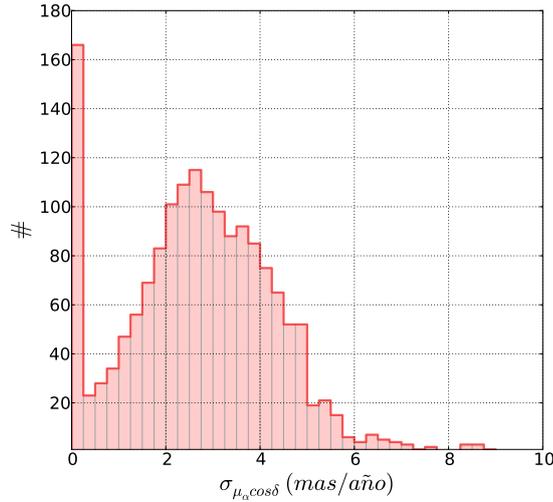


Figura 3.3: Distribución de la coordenada x de la dispersión de velocidades de los cúmulos listadas en D14. Como puede observarse, la distribución presenta un máximo en torno a los 3 mas/año.

3.3 Simulaciones con error Gaussiano de σ_μ constante.

La distribución de las posiciones de las estrellas del cúmulo viene modelada por una distribución Gaussiana circular con $\sigma_{pos} = 0.031^\circ$, es decir, un tercio del radio angular del cúmulo típico. El cúmulo siempre estará centrado en la posición (0, 0), que es también la posición central de la muestra. La distribución espacial de las estrellas de campo viene descrita por una distribución aleatoria dentro de un campo cuadrado de lado 0.3° .

Los movimientos propios de ambas poblaciones de estrellas son descritos por funciones de distribución Gaussianas bivariantes: circular para el cúmulo y elíptica para el campo. La distribución de los movimientos propios de las estrellas de campo está siempre centrada en el (0,0) mas/año y su matriz de covarianza se toma diagonal con valores que varían entre 10, 15, 20 y 25 mas/año para la componente $\sigma_{\mu_{f,x}}$, y 1 ó 1.2 veces $\sigma_{\mu_{f,x}}$ para la dispersión en el eje y ($\sigma_{\mu_{f,y}}$). Estos valores son representativos de la distribución de las estrellas de campo en la muestra de 1646 cúmulos (ver figura 3.2).

La frecuencia relativa de estrellas de cúmulo la hacemos variar entre 20%, 40%, 60% y 80% del número total de la muestra. La distribución de las estrellas de cúmulo en el plano de los movimientos propios la obtenemos a partir de una función Gaussiana de desviación estándar, la correspondiente a su equilibrio virial. La posición del centroide del cúmulo en el VPD (Vector Point Diagram) mantiene fija su componente $\mu_{x,c} = 0$ mas/año, mientras que la componente $\mu_{y,c}$ la hacemos variar entre los valores (1., 5. y 7.5) mas/año.

Para introducir los errores en la muestra, procedemos de la siguiente manera: el movimiento propio inicial de cada estrella (obtenido usando su dispersión interna de velocidades), ha sido reemplazado por un número aleatorio tomado de una distribución Gaussiana de media igual al valor del movimiento

Tabla 3.1: Valores de los parámetros fijos en todas las simulaciones, de las distribuciones de las poblaciones de cúmulo y campo, en los espacios de posiciones y de movimientos propios.

Variables	Values
Número total de estrellas	$N_{Total} = 500$
Centroide del Cúmulo y del Campo en Posiciones	$(x_c, y_c) = (x_f, y_f) = (0,0)$ deg
Centroide del Cúmulo en Movimientos Propios (Coord - X)	$\mu_{c,x} = 0$ mas/año
Centroide del Campo en Movimientos Propios	$(\mu_{f,x}, \mu_{f,y}) = (0,0)$ mas/año
Dispersión Interna del Cúmulo en Movimientos Propios	$\sigma_{\mu c,x} = \sigma_{\mu c,y} = \sigma_{\mu c} = 0.07$ mas/año
Errores en Movimientos Propios	$\sigma_{\mu} = 3$ mas/año
Radio Projectado del Cúmulo	$R_{Cúmulo} = 0.093$ deg

Tabla 3.2: Valores de los parámetros variables, en las simulaciones de las distribuciones de las poblaciones de cúmulo y campo, en los espacios de posiciones y de movimientos propios.

Variables	Values
Centroide del Cúmulo en Movimientos Propios (Coord - Y)	$\mu_{c,y} = (1, 5, 7.5)$ mas/año
Dispersión del Campo en Movimientos Propios (Coord-Y)	$\sigma_{\mu f,x} = (10, 15, 20, 25)$ mas/año
Dispersión del Campo en Movimientos Propios (Coord-X)	$\sigma_{\mu f,y} = (1, 1.2) \cdot \sigma_{\mu f,x}$
Porcentaje de Miembros de Cúmulo	20%, 40%, 60%, 80%

propio inicial y con una desviación estándar igual al error (Priyatikanto & Arifyanto, 2015), que en este caso es de 3 mas/año. Al final, un total de 96 simulaciones construidas de esta forma, representan un banco de prueba para evaluar la bondad de la nueva metodología.

En las tablas 3.1 y 3.2 se muestra un resumen del conjunto de parámetros que caracterizan nuestras simulaciones para las cuatro variables del sub-espacio de fase. La tabla 3.1 resume los parámetros que son fijos en las simulaciones y la tabla 3.2 los variables. La tabla 3.3 muestra un esquema de los parámetros variables simulados para un porcentaje dado de estrellas de cúmulo. Un ejemplo de simulación se muestra en la figura 3.4, donde los gráficos superior e inferior presentan las distribuciones de campo y cúmulo en los espacios de posiciones y de movimientos propios, respectivamente.

Dada que la nueva metodología propuesta está diseñada para trabajar con un diferente número de variables, estas simulaciones han permitido la realización de distintos análisis utilizando sub-espacios del espacio de fase con $N = 1, 2$ y 4 dimensiones. En el caso de 1 y 2 dimensiones nos hemos restringido a los datos cinemáticos.

Tabla 3.3: Esquema de los valores que toman los parámetros variables de las configuraciones simuladas para un porcentaje de estrellas de cúmulo.

Simul.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
$\mu_{c,y}$	1.	5.	7.5	1.	5.	7.5	1.	5.	7.5	1.	5.	7.5	1.	5.	7.5	1.	5.	7.5	1.	5.	7.5	1.	5.	7.5
$\sigma_{\mu f,x}$	10	10	10	10	10	10	15	15	15	15	15	15	20	20	20	20	20	20	25	25	25	25	25	25
$\sigma_{\mu f,y}$	10	10	10	12	12	12	15	15	15	18	18	18	20	20	20	24	24	24	25	25	25	30	30	30

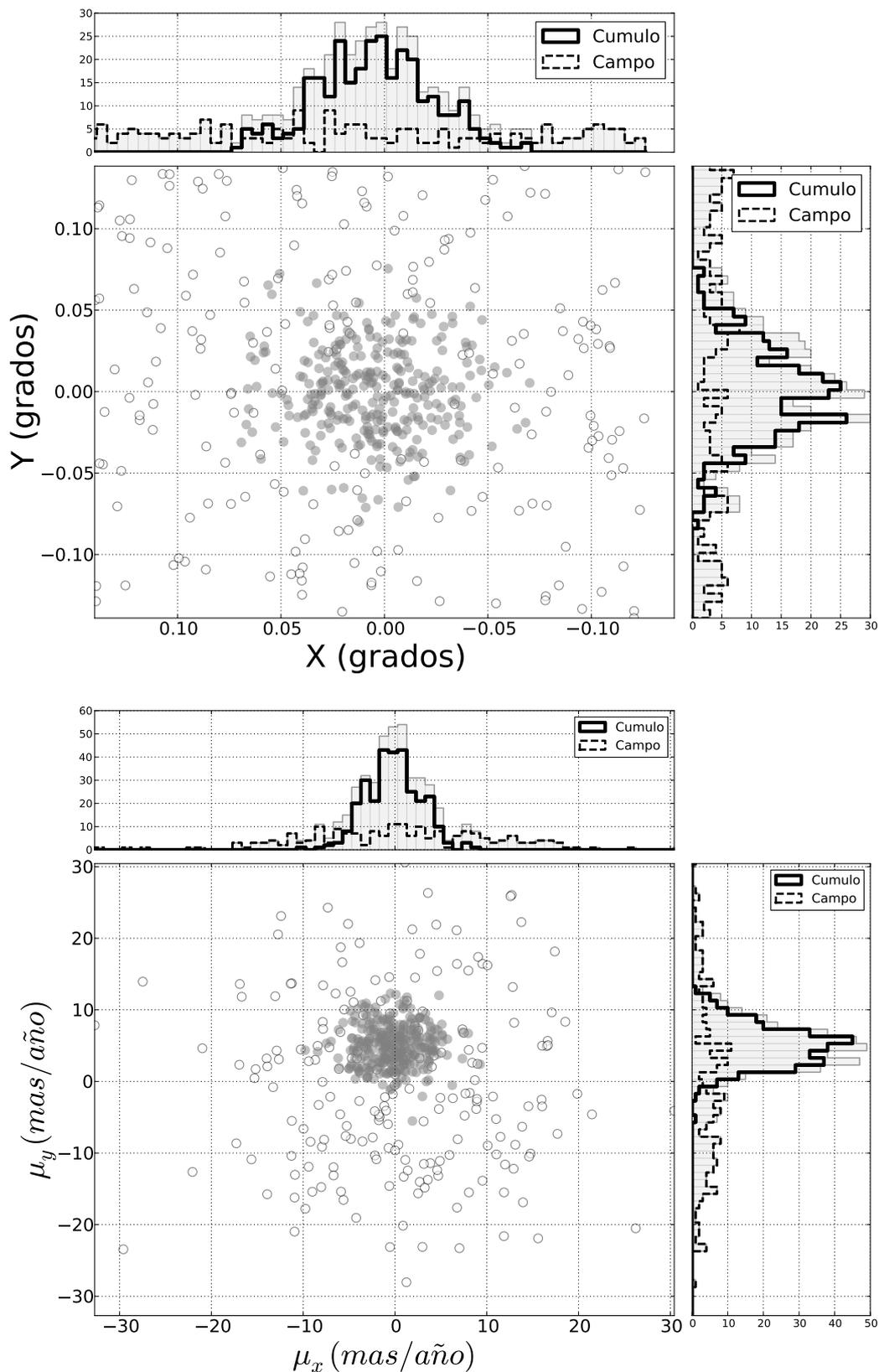


Figura 3.4: Ejemplo de una simulación en los espacios de posiciones y movimientos propios, gráficos superior e inferior, respectivamente. Los miembros del cúmulo siguen distribuciones circulares bivariantes en ambos sub-espacios. Las estrellas de campo tienen una distribución aleatoria en posiciones y distribuciones Gaussianas bivariantes en movimientos propios. En esta simulación, se está representando el caso cuando se simula un porcentaje de un 60% de estrellas de cúmulo, respecto de la muestra total, con $\mu_{c,y} = 5.0$, $\sigma_{\mu f,x} = 10$ mas/año y $\sigma_{\mu f,y} = 10$ mas/año.

3.4 Simulaciones con error Gaussiano con σ_μ dependiente de la magnitud.

Uno de los objetivos fundamentales de esta tesis consiste en la generación de un nuevo catálogo de miembros de los cúmulos abiertos listados en DAML02, haciendo uso de los datos del cartografiado UCAC4. La ausencia de cúmulos control que nos permitan determinar cómo las diferentes metodologías son capaces de separar una muestra en estrellas de cúmulo y campo, hace necesario el aproximar las simulaciones realizadas lo máximo posible a los datos que vamos a utilizar. Así, estas simulaciones nos permiten cuantificar en que medida las diferentes metodologías, para la determinación de probabilidades de pertenencia, son capaces de seleccionar adecuadamente los miembros potenciales de la población de cúmulo, en función de las características de la distribución de la muestra completa.

Con este objetivo se diseñaron unas nuevas simulaciones basándonos en las anteriormente descritas, pero en las que hacemos un tratamiento diferente de los errores en movimientos propios. Dado que ya nos basamos en los resultados publicados en los principales y más completos catálogos de cúmulos reales, que a su vez usan datos de este mismo cartografiado, en este apartado nos centraremos en describir las diferencias respecto a las simulaciones anteriores, siendo la diferencia fundamental, la introducción de los errores en función de la magnitud de las estrellas.

Se tomó una muestra aleatoria de 21 regiones de cúmulos usando datos del cartografiado UCAC4. Si representamos los errores en movimientos propios frente a la magnitud, obtenemos el gráfico 3.5. La figura muestra que, a pesar de la gran dispersión de los datos, se observa una tendencia exponencial que incrementa el error de los movimientos propios cuanto más débiles son las estrellas. Los histogramas de errores en movimientos propios y magnitud aparente de las estrellas de estos 21 cúmulos se muestran en la figura 3.6. El gráfico izquierdo muestra la distribución de los errores en movimientos propios, observándose un máximo en torno a los 4 mas/año. En cuanto a la distribución de magnitudes (gráfico derecho), se observa que el catálogo UCAC4 parece completo hasta la magnitud $R = 16$. Este último gráfico también nos muestra que la gran mayoría de las estrellas se encuentran distribuidas en el intervalo de magnitudes entre $R = 9$ y 16 mag.

Estas distribuciones fueron tenidas en cuenta en la realización de las simulaciones con el objetivo fundamental de introducir los errores de los movimientos propios de las estrellas de una manera más realista. Para ello partimos de una distribución de tipo Salpeter (Salpeter, 1955), usando un intervalo de masas entre 1-10 M_\odot . Para no hacer ninguna hipótesis en cuanto a la edad y a la extinción del cúmulo, y dado que sólo estamos interesados en su distribución, se decidió mover simplemente la muestra obtenida al intervalo de magnitudes deseado. Una vez conocida la frecuencia de estrellas por intervalo de magnitud, a partir de ellas se asignaron los errores en movimientos propios. Cabe señalar en este punto que la simulación de la distribución de las magnitudes que hemos realizado tiene el único objetivo de introducir los errores en los movimientos propios, no queriendo inferir ningún estudio a posteriori a partir de la distribución de las estrellas en función de su magnitud, tales como segregación de masas o cualquier otro.

Una vez determinada la distribución de objetos por intervalo de magnitud, se hizo uso de los datos reales procedentes de las 21 regiones de cúmulos, para asignar los errores en movimientos propios a

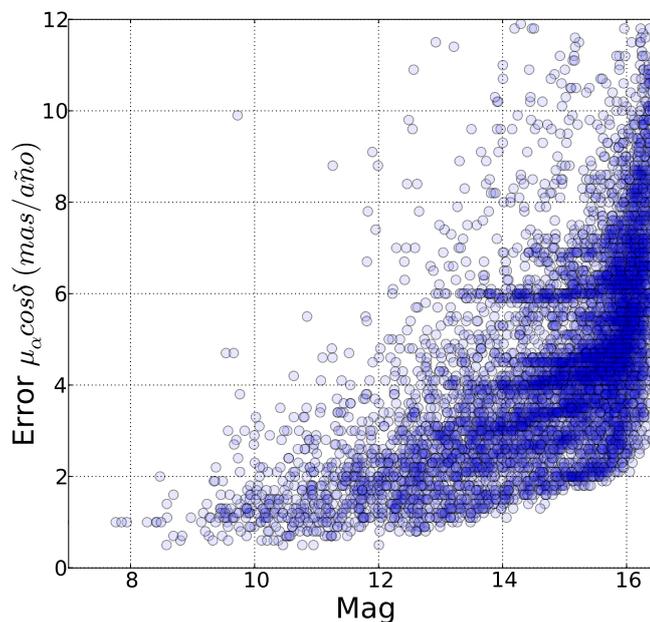


Figura 3.5: Errores en movimientos propios en función de la magnitud para una muestra de 21 cúmulos reales, utilizando datos del cartografiado UCAC4.

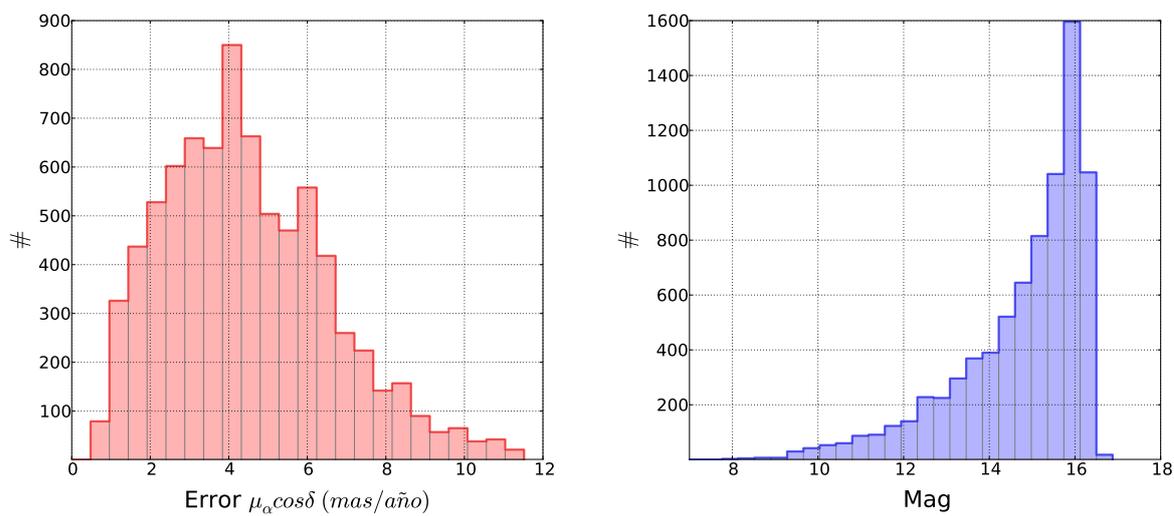


Figura 3.6: Distribuciones de los errores y de las magnitudes de una muestra formada por 21 cúmulos usando datos de UCAC4.

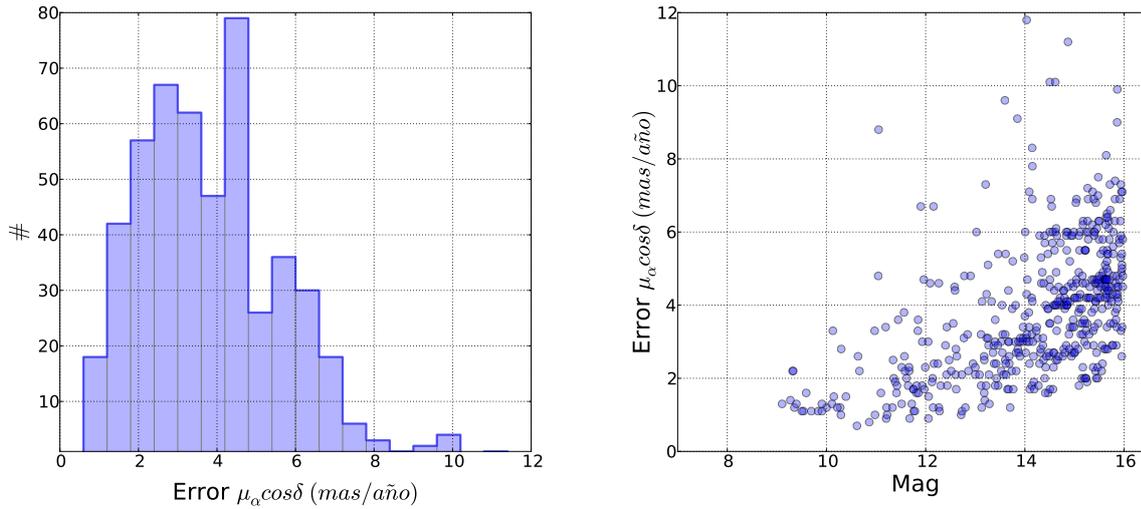


Figura 3.7: Ejemplos de distribuciones de errores en movimientos propios y de magnitudes simuladas.

cada una de las estrellas. Para el rango de magnitudes entre 9 y 16 mag, se tomaron intervalos de 0.5 magnitudes y, para cada uno de estos intervalos, se cogieron aleatoriamente errores reales de los movimientos propios de la muestra de datos reales. Las distribuciones resultantes de la aplicación de este proceso se muestran en la figura 3.7.

Una vez que tenemos los errores en movimientos propios para cada estrella, se incluyeron en los datos de la misma manera que en las simulaciones anteriores. El movimiento propio inicial de cada estrella fue reemplazado por un número aleatorio tomado de una distribución Gaussiana centrada en el valor propio inicial y con una dispersión igual al error del movimiento propio asignado a cada estrella, que a su vez dependerá de su magnitud. En la figura 3.8 se muestra un ejemplo de las diferentes simulaciones en el plano de los movimientos propios.

En este segundo grupo de simulaciones hemos decidido añadir unas nuevas configuraciones, en las que ampliamos los parámetros simulados relacionados con las dispersiones intrínsecas de las estrellas de campo en movimientos propios. Así, el rango de valores simulados se amplía siendo la matriz de covarianza diagonal con valores que varían entre (4, 6, 8, 10, 15, 20 y 25) mas/año para la componente $\sigma_{\mu_{f,x}}$. La componente $\sigma_{\mu_{f,y}}$, es de nuevo 1 ó 1.2 veces ($\sigma_{\mu_{f,x}}$). Las razones fundamentales que nos han llevado a esta ampliación son dos. Por un lado, el objetivo de este segundo grupo de simulaciones consiste en tratar de caracterizar cómo las metodologías son capaces de clasificar las estrellas de cúmulo y campo, en función de las características de las distribuciones de ambas poblaciones, al ser aplicadas a la muestra de cúmulos listadas en el trabajo DAML02, usando para ello datos de UCAC4. Aunque en las primeras simulaciones se utilizaron los intervalos de valores más frecuentes, se decidió explorar aquellas regiones en las que las distribuciones de cúmulo y campo son más similares y que, por tanto, pueden producir una mayor dificultad en la clasificación de las estrellas en las dos poblaciones. El segundo motivo procede de los resultados obtenidos tras aplicar las metodologías a las primeras simulaciones, y que se detallarán en el capítulo 4. En las simulaciones se observa que existe una

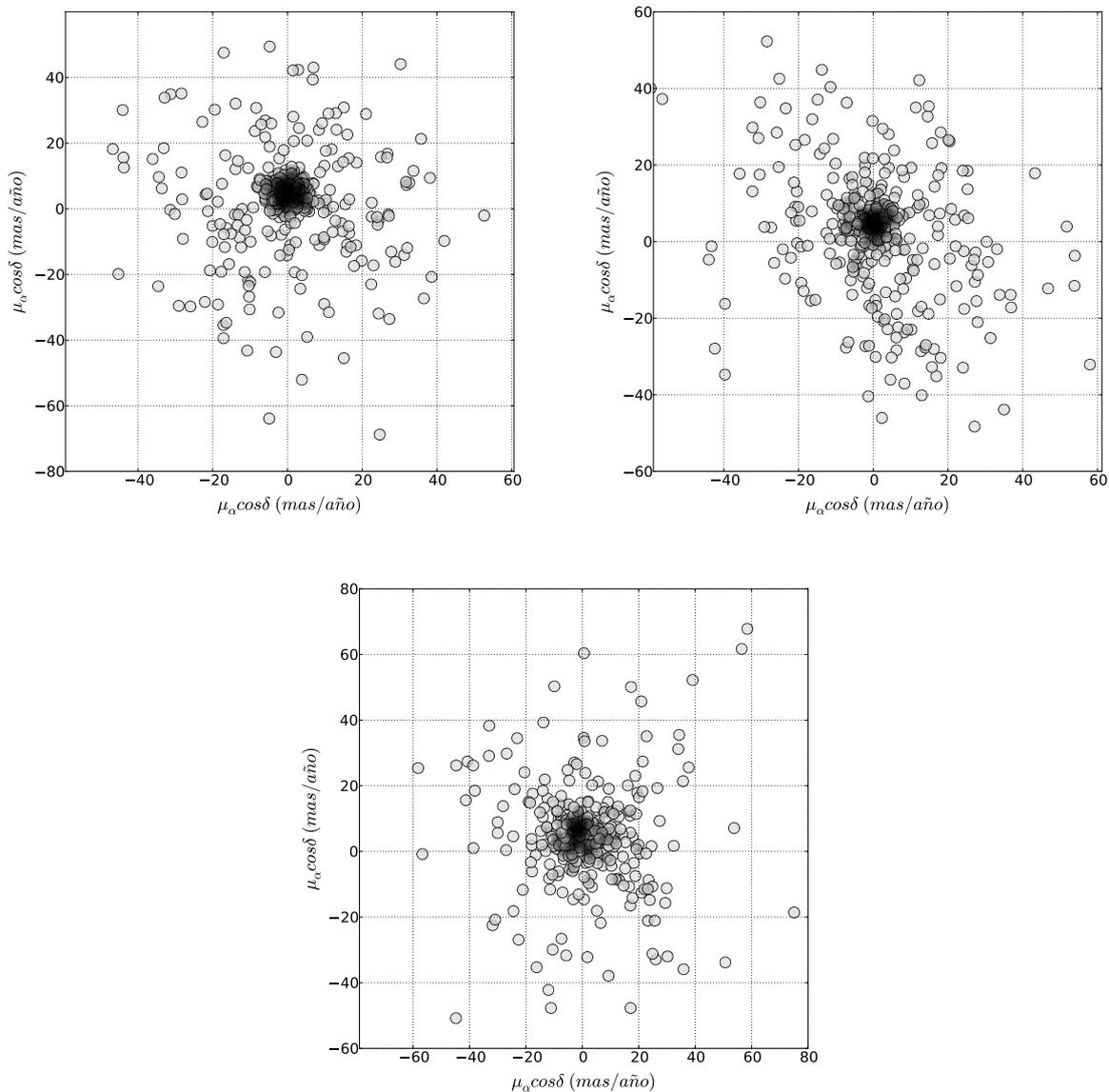


Figura 3.8: Ejemplos de ambos tipos de simulaciones en el espacio de movimientos propios y comparación con el cúmulo NGC 2360. En los gráficos superiores se muestra la misma configuración simulada en la que los errores tienen un comportamiento Gaussiano independiente de otras variables estelares (izquierda) y dependiente de la magnitud de las estrellas (derecha), respectivamente. En el gráfico inferior se muestra la distribución de los movimientos propios del cúmulo NGC 2360.

tendencia, a obtener mejores resultados cuanto mayor es la dispersión de los movimientos propios de las estrellas de campo, es decir, se recupera un mayor número de estrellas de cúmulo y disminuye la mala clasificación realizada por las metodologías. Suponiendo que esta tendencia se mantiene, se prefirió ampliar el estudio a aquellas regiones del cielo donde la dispersión de las estrellas de campo es más parecida a la de cúmulo.

El resto de los parámetros que describen las distribuciones de ambas poblaciones y que no han sido descritos en esta sección, toman los mismos valores que los detallados en la sección anterior. Al final, un total de 168 simulaciones se han realizado para describir el comportamiento de las diferentes metodologías en la clasificación de la muestra en las dos poblaciones.

3.5 Ausencia de sobre-densidad en el espacio de posiciones.

3.5.1 Simulaciones con distribuciones aleatorias en el espacio de posiciones para la población de cúmulo.

Una de las características críticas a la hora de buscar los miembros probables de un cúmulo es, precisamente, la existencia de una sobre-densidad en el espacio de posiciones frente a una distribución aparentemente uniforme de las estrellas de campo, en dicho espacio. Sobre la base de esta premisa se diseñaron las simulaciones descritas previamente, donde la distribución espacial del cúmulo se describía a partir de una función Gaussiana bivalente, responsable de la sobre-densidad observada en el espacio de posiciones.

Sin embargo, cuando se trabaja con datos reales esta sobre-densidad puede tener un bajo o nulo contraste. En la figura 3.9, se muestra el perfil de densidad radial y la distribución en posiciones del cúmulo ASCC 102, donde se observa una densidad superficial casi constante. Esta ausencia de sobre-densidad puede producirse por varios efectos: el límite de magnitud del telescopio y la incompletitud en los datos en las variables a utilizar es un hecho en catálogos reales como UCAC4. No todas las estrellas listadas para cada muestra a analizar tienen datos de movimientos propios (ver figura 3.10), las cuales son, por tanto, eliminadas de los análisis de pertenencia a realizar. Se observa una mayor ausencia de datos en movimientos propios en la región del cúmulo, posiblemente debidos a efectos de *blending* del telescopio o a presencia de absorciones. Por otro lado, la distribución en posiciones de los cúmulos abiertos es altamente variable, siendo también características distribuciones aleatorias que se muestran como una meseta en los perfiles de densidad radial.

Debido a estas distribuciones observadas en los cúmulos reales se decidió explorar el comportamiento bajo estas condiciones, de las metodologías que hacen uso de las posiciones para realizar sus análisis de pertenencia, que son la metodología no-paramétrica y la metodología geométrica para $N = 4$ variables (posiciones y movimientos propios). Así, los parámetros asociados a los movimientos propios de las estrellas tanto de cúmulo como de campo, se mantienen idénticos a los descritos en la sección anterior. En cuanto a las posiciones, la distribución de las estrellas de campo permanece siendo una distribución aleatoria, mientras que la distribución de las estrellas de cúmulo sigue el mismo

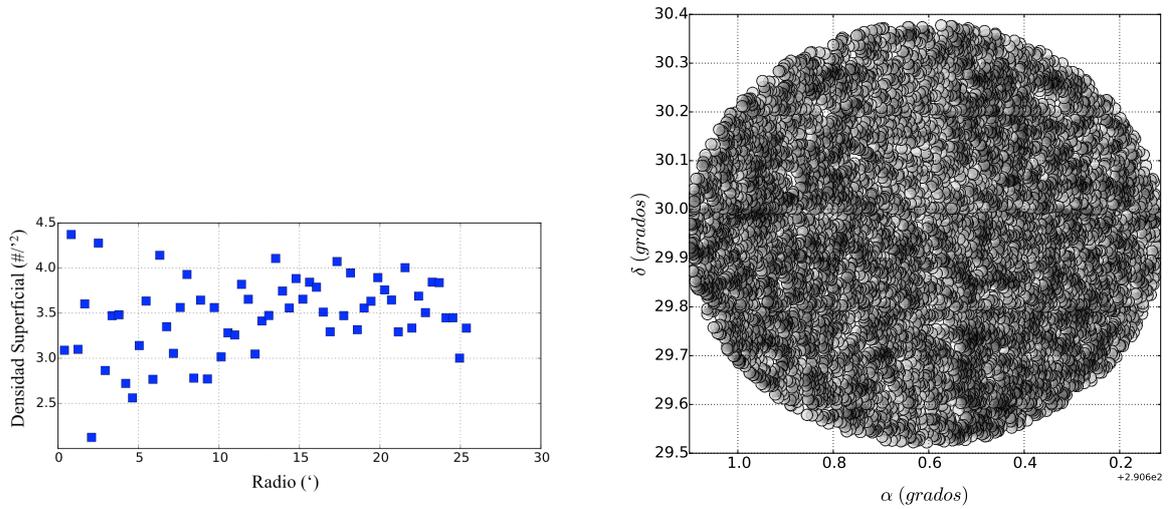


Figura 3.9: Perfil de densidad radial y distribución en el espacio de posiciones de la región del cúmulo ASCC 102 usando la base de datos de UCAC4. Se puede observar la ausencia de sobre-densidad existente para este cúmulo.

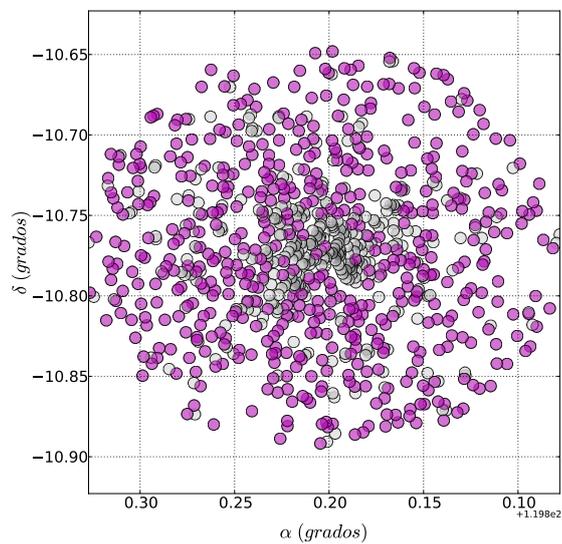


Figura 3.10: Distribución de las posiciones del cúmulo NGC 2506 usando datos de UCAC4. La muestra total se representa en negro y, superpuesta en magenta, aquellas estrellas con datos de movimientos propios. Como se puede observar, en la región del cúmulo hay un mayor número de estrellas sin datos en movimientos propios, estrellas que son eliminadas de los análisis de pertenencia.

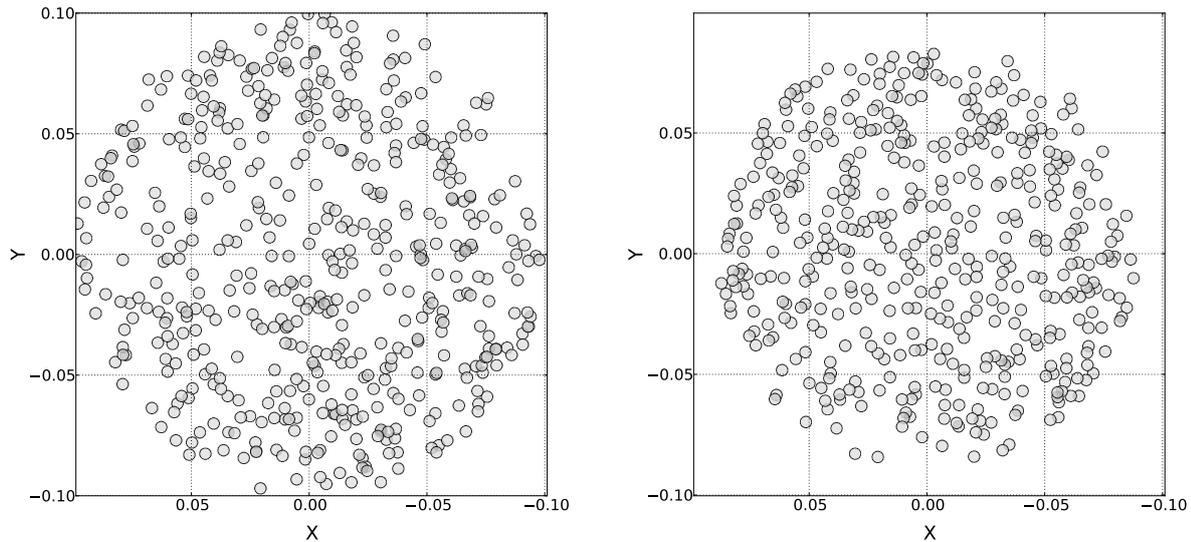


Figura 3.11: Comparación entre una simulación donde la distribución de las posiciones es aleatoria (gráfico izquierdo) y el cúmulo NGC 5138 (gráfico derecho), el cual se representa en distancias angulares.

tipo de distribución pero superpuesta a la de campo, manteniéndose el mismo radio del cúmulo. Un ejemplo de las simulaciones realizadas y la comparativa con el cúmulo real NGC 5138 se muestra en la figura 3.11.

3.5.2 Efectos de sub-muestreo en la muestra.

En las simulaciones anteriores nos hemos centrado en estudiar la ausencia de sobre-densidad en el espacio de posiciones. Sin embargo, también la ausencia de datos de movimientos propios sesga el espacio cinemático. Con el objetivo de analizar este posible sesgo introducido por un sub-muestreo observacional de la población, hemos realizado un sub-muestreo de un 50%, un 20% y un 10% con respecto a cada una de las muestras iniciales simuladas de 500 estrellas, siendo estos porcentajes las proporciones del número de estrellas finales que son recuperadas de la muestra inicial. Así, para el sub-muestreo más pequeño, el del 50%, la muestra a analizar estará formada por 250 estrellas, para el del 20% por 100 estrellas y, finalmente por 50 estrellas para el sub-muestreo del 10%, manteniendo la proporción inicial de estrellas de campo y cúmulo. Este estudio se ha realizado tanto en las simulaciones en las que los errores en los movimientos propios tenían una dispersión constante como en las que este error era dependiente de la magnitud de las estrellas.

Todos estos estudios nos permiten explorar a partir de qué características de las distribuciones de campo con referencia a la de cúmulo, los efectos de sub-muestreo pueden llevar a que las metodologías realicen una mala clasificación de las estrellas en las dos poblaciones.

Los resultados obtenidos, al aplicar las metodologías a las simulaciones descritas en este capítulo, se detallan a continuación.

4

Resultados obtenidos en las simulaciones.

4.1 Introducción

En este capítulo se detallan los resultados obtenidos tras la aplicación de las diferentes metodologías descritas en el capítulo 2, a los cúmulos simulados. El análisis de las simulaciones nos permitirá conocer como se comporta cada uno de los métodos en función de los parámetros que definen las simulaciones.

Para poder cuantificar y caracterizar los resultados obtenidos en las simulaciones hemos introducido dos parámetros de control, que nos muestran la bondad de las diferentes metodologías en separar ambas poblaciones a partir de la muestra inicial. Estas dos figuras de mérito son la *Complejitud* (C) y el *Coefficiente de mal clasificados* (M), viniendo definidas por las siguientes expresiones:

$$C = \frac{N_{c,met}}{N_{c,real}} \quad (4.1)$$

$$M = \frac{N_{c \rightarrow f,met} + N_{f \rightarrow c,met}}{N_{Total}} \quad (4.2)$$

donde $N_{c,met}$ es el número de estrellas de cúmulo recuperado por una metodología, $N_{c,real}$ es el número total de estrellas de cúmulo simuladas, $N_{c \rightarrow f,met}$ es el número de estrellas que han sido simuladas como miembros pero que han sido clasificadas como estrellas de campo, $N_{f \rightarrow c,met}$ es el número de estrellas que han sido simuladas como estrellas de campo, pero que la metodología las ha clasificado como estrellas de cúmulo y, finalmente, N_{Total} es el número total de estrellas simuladas.

De acuerdo con estas ecuaciones, C es el porcentaje de miembros de cúmulo simulados que son recuperados por una metodología, mientras que M es el porcentaje de estrellas que han sido mal clasificadas, bien porque se han simulado como estrellas de campo y han sido clasificadas como miembros del cúmulo, o viceversa.

Es evidente que la bondad en la clasificación de las estrellas en cúmulo y campo depende de las características de las funciones de distribución que definen a las dos poblaciones en el sub-espacio de fase. Una medida del grado de similitud entre las funciones de distribución que describen ambas poblaciones

de estrellas, es precisamente la distancia de Chernoff (Chernoff, 1952). Esta distancia estadística, que a partir de ahora se denotará por DC , ha sido calculada haciendo uso de los parámetros de las distribuciones tanto de posiciones como de movimientos propios simuladas. Su expresión analítica general viene dada por la ecuación:

$$DC = \frac{1}{2} \alpha_c \alpha_f (\mu_c - \mu_f)^T [\alpha_c \Sigma_c + \alpha_f \Sigma_f]^{-1} (\mu_c - \mu_f) + \frac{1}{2} \log \frac{|\alpha_c \Sigma_c + \alpha_f \Sigma_f|}{|\Sigma_c|^{\alpha_c} |\Sigma_f|^{\alpha_f}} \quad (4.3)$$

donde α , μ y Σ son los porcentajes, medias y covarianzas de ambos grupos de estrellas, y los super-índices T y -1 se refieren al vector transpuesto y a la matriz inversa, respectivamente.

La combinación de los valores de ambas figuras de mérito (C , M) en función de la DC , obtenidos para cada simulación, nos permitirá describir el comportamiento de las diferentes metodologías en la determinación de los miembros potenciales de los cúmulos abiertos, en función de las características de las funciones de distribución de las variables utilizadas. Cabe recordar que las metodologías son aplicadas a muestras libres de *outliers*. Comenzamos con los resultados obtenidos en las simulaciones en las que las desviaciones típicas en el espacio de movimientos propios son constantes.

4.2 Resultados obtenidos en las simulaciones.

4.2.1 Error Gaussiano con σ_μ constante.

Nos referiremos a estas simulaciones como *Sim1*. Las simulaciones realizadas y el potencial de la nueva metodología en la utilización de diferentes conjuntos de variables, nos permiten llevar a cabo tres análisis de pertenencia diferentes en los que una, dos y cuatro variables fueron utilizadas ($N = 1, 2$ y 4 variables). La aplicación de las metodologías descritas en la sección 2, nos conducirá a la comparación de los resultados obtenidos por la nueva metodología con aquellos obtenidos por otra metodología cuando son aplicadas al mismo conjunto de variables. Los métodos y número de variables utilizadas se combinan de la siguiente forma: a) metodología geométrica usando una única variable (MTG1); b) metodología paramétrica con una variable (MT1); c) método geométrico con dos variables (MTG2); d) metodología paramétrica con dos variables (MT2); e) metodología geométrica usando cuatro variables (MTG4) y; f) metodología no-paramétrica en 4 dimensiones (MT4). Los conjuntos de variables utilizados son: en el caso $N = 1$ variable la componente (μ_y) de los movimientos propios, en el caso de $N = 2$, los movimientos propios (μ_x , μ_y) y en el caso de $N = 4$ variables, tanto las posiciones de las estrellas como sus movimientos propios (x , y , μ_x , μ_y).

La figura 4.1 muestra los resultados obtenidos para las seis combinaciones de métodos y número de variables en función de la DC . Esta figura está dividida en seis gráficos: los dos superiores detallan los resultados obtenidos para 1 variable, los dos intermedios para el caso de 2 variables, y los dos inferiores

para el caso de 4 variables. Cada gráfico tiene información sobre las dos figuras de mérito C (en el eje vertical izquierdo) y M (en el derecho) en función de DC . Los valores obtenidos por el método geométrico se representan por una línea continua magenta mientras que los derivados del modelo con el que se esté realizando la comparación, se representan por una línea discontinua negra. Las áreas coloreadas alrededor del valor central representan la dispersión (1σ) de los resultados obtenidos dentro de intervalos de 0.15 en unidades de DC .

La figura 4.1 muestra que si DC aumenta los resultados mejoran en el sentido de que C aumenta y M disminuye y, además, esto ocurre para todas las combinaciones de método y número de variables. Se observa también una mayor convergencia de los resultados al aumentar DC y una menor dispersión de los mismos. El coeficiente de mal clasificados, M , decrece al aumentar el número de variables utilizadas en los análisis de pertenencia. Debemos hacer notar que la M obtenida por el método geométrico decrece desde un 25% para una variable hasta un 5% para cuatro variables, para los valores más bajos de DC . Igualmente, la completitud, C aumenta con el número de variables. Sin embargo, la C obtenida por la metodología geométrica muestra un comportamiento casi constante ($C > 90\%$) para cualquier número de variables utilizadas y dentro del rango de DC analizadas. Es de notar que la metodología geométrica obtiene resultados similares o incluso mejores que aquellos obtenidos por el resto de las metodologías usando el mismo conjunto de datos para estas simulaciones.

Para $N = 4$ variables, tanto la metodología no-paramétrica como la geométrica, son capaces de determinar los miembros de los cúmulos de una forma muy eficaz. Los valores alcanzados tanto por C como por M indican que, la mayoría de las estrellas que han sido clasificadas como miembros, realmente lo son y que, además, un porcentaje muy pequeño de estrellas de campo han sido clasificadas como estrellas de cúmulo, inferior al 5% (unas 25 estrellas) para $DC > 1$. Para prácticamente todo el rango de valores de DC , la M obtenida por la metodología geométrica no supera el 5%.

La columna derecha de la figura 4.1 muestra un cambio abrupto en el comportamiento de M para unos valores de la DC entre 0.9 y 1. Este comportamiento, tal y como se observa en la figura 4.2, es debido al cambio en la proporción de estrellas de cúmulo simuladas, al pasar de un 20% a un 40%. En esta figura se muestran los valores de C y de M obtenidos por MTG2, para los cuatro porcentajes de estrellas de cúmulo simulados: un 20% en magenta, un 40% en azul, un 60% en verde y un 80% en negro. Se observa que los valores más pequeños de la DC ($DC < 0.9$), se corresponden sólo con el menor de los porcentajes de estrellas de cúmulo simulado, el del 20%. Además, mientras que C tienen un comportamiento prácticamente constante, se puede ver que cuanto mayor es el porcentaje de estrellas de cúmulo en la muestra, menor es el valor de M obtenido. Estos resultados muestran que la proporción del número de estrellas de cúmulo en la muestra, tiene una gran influencia en las figuras de mérito. En particular, la C obtenida por la metodología MT1 (figura 4.1 gráfico superior izquierdo), está especialmente afectada por este porcentaje, siendo muy baja cuando el número de estrellas de cúmulo es del orden del 20%.

Los resultados mostrados en las figuras anteriores se corresponden con aquellas simulaciones en las que los diferentes métodos aplicados han determinado la existencia de cúmulo. Sin embargo, se debe señalar que en algunos casos los métodos no convergen y no se llega a una solución para todas las

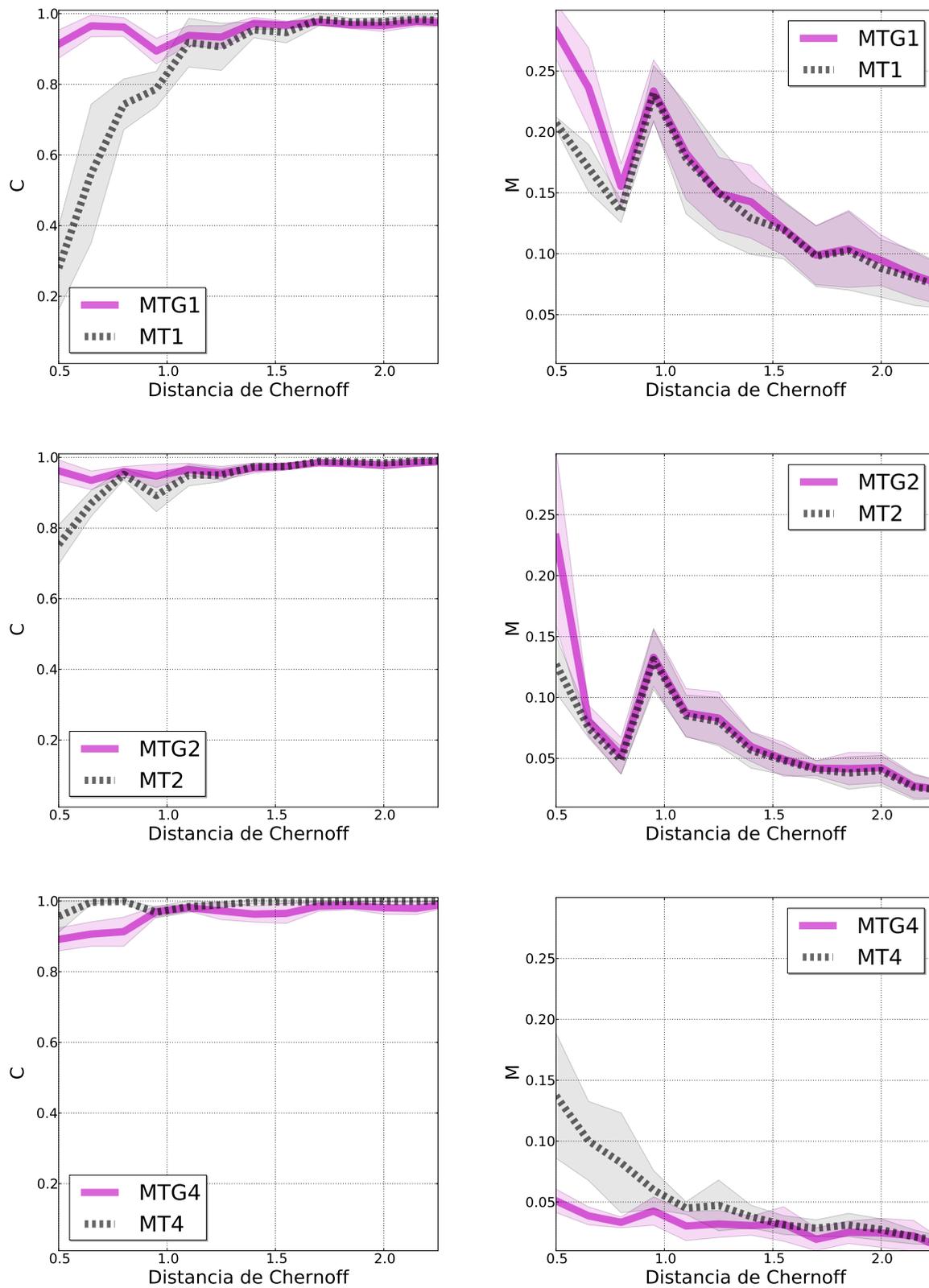


Figura 4.1: Comportamiento de ambas figuras de mérito con el número de variables utilizadas en el análisis de pertenencia, respecto a la DC . Las líneas continuas magenta refieren a los resultados obtenidos por la nueva metodología, mientras que las líneas discontinuas negras muestran los resultados obtenidos por los otros modelos. Las áreas coloreadas representan la dispersión (de 1σ) de los resultados en intervalos de 0.15 unidades de DC . Se observa una mejora en los resultados con el incremento tanto con el número de variables utilizadas como con la heterocedasticidad de las funciones de distribución medidas por la DC entre las dos poblaciones.

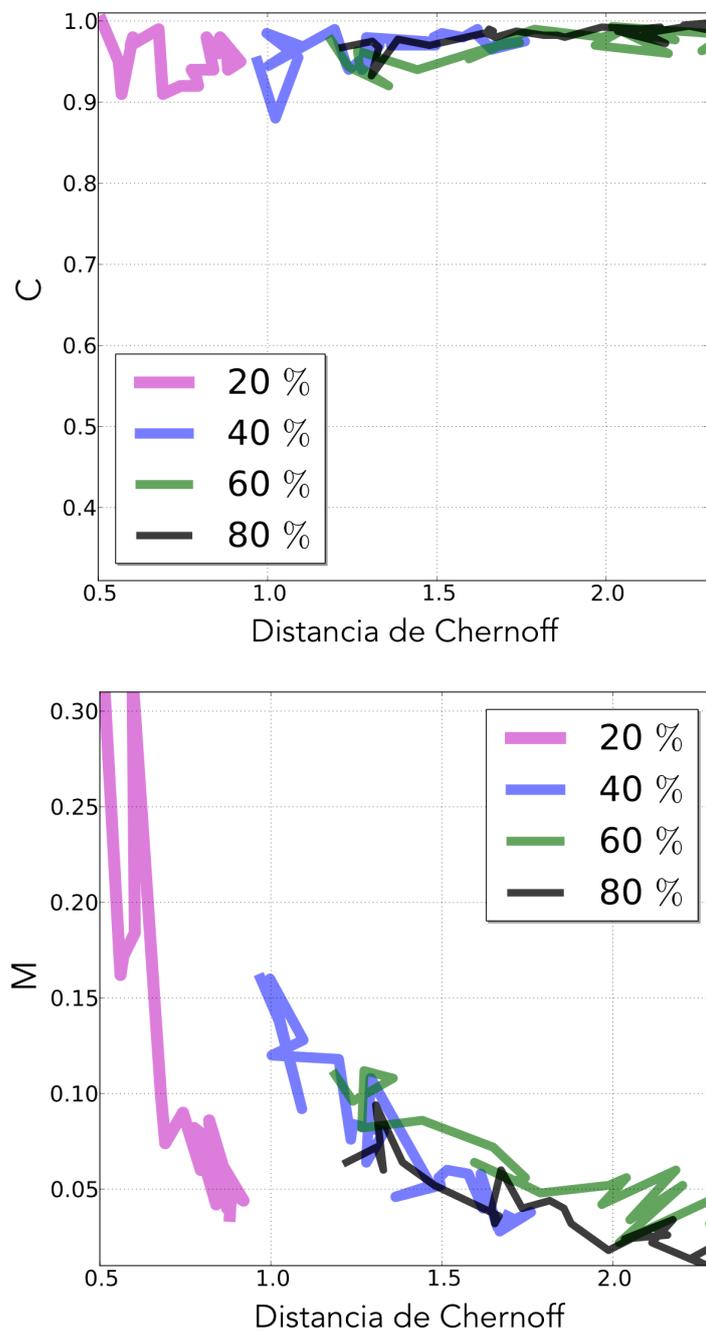


Figura 4.2: Dependencia con el porcentaje de estrellas de cúmulo simuladas de los valores de C y de M de acuerdo a la DC obtenidos por MTG2. Los resultados obtenidos para un número de estrellas de cúmulo igual al 20% de la muestra total se representan en magenta, el 40% en azul, el 60% en verde y el 80% en negro. Los valores más bajos de DC ($DC < 0.9$) se corresponden con el porcentaje más bajo de estrellas de cúmulo simulado. Para los porcentajes más altos, DC aumenta y se observa una disminución en los valores de M obtenidos.

simulaciones generadas con un mismo modelo (un mismo conjunto de parámetros). En estos casos, la falta de solución se ha producido únicamente cuando el porcentaje de estrellas de cúmulo simulado es del 20%.

Las distribuciones de probabilidad de pertenencia obtenidas por la metodología geométrica están influenciadas tanto por el número de variables utilizadas en el análisis como por el porcentaje de estrellas de cúmulo simuladas. En la figura 4.3 se muestran dichas distribuciones de probabilidad, obtenidas por este método con $N = 1, 2$ y 4 variables. Estos resultados se corresponden con un número de estrellas de cúmulo igual al 20% de la muestra total en el gráfico superior y, del 80% de la muestra en el gráfico inferior, manteniendo los mismos valores de los parámetros que definen las distribuciones de cúmulo y campo. Se observa que para la proporción de estrellas de cúmulo del 20%, la probabilidad más alta obtenida es del orden del 90%, mientras que para el 80% llega al 100%. Para ambos casos, los valores más altos de la probabilidad alcanzados se obtienen en los estudios en los que el número de variables utilizadas es $N = 4$. Además, la contaminación en la asignación de probabilidades de pertenencia intermedia es de nuevo inferior para 4 variables.

Es evidente que los indicadores C y M están afectados de una incertidumbre en su determinación. Para estimarla se aplicó la técnica del Bootstrap (Efron, 1979). Con este fin, se realizaron 50 re-muestreos de cada uno de los modelos simulados, donde se considera que una medida de la incertidumbre de estos parámetros viene dada por la dispersión de los valores obtenidos para los 50 muestreos. Los errores obtenidos son inferiores al 10% para la gran mayoría de las simulaciones, tomando valores inferiores al 5% para $DC > 1$. Su distribución respecto de DC se presenta en la figura 4.4, donde en el gráfico de la izquierda se muestran las incertidumbres obtenidas para MTG1, en el gráfico intermedio las obtenidas para MTG2, y en el gráfico de la derecha para MTG4. Los círculos magenta y los hexágonos azules representan los errores de C y M , respectivamente. Se puede observar una menor incertidumbre en la estimación de las dos figuras de mérito conforme aumenta la DC .

Como se comentó en el capítulo 3, con el objetivo de estudiar un posible sesgo introducido por sub-muestreos observacionales de la población, hemos tomado sub-muestras de un 50%, un 20% y un 10% respecto de la muestra total, de cada una de las simulaciones realizadas (250, 100 y 50 objetos respectivamente), y para los cuatro porcentajes de estrellas de cúmulo. Con estas simulaciones pretendemos visualizar el efecto de la falta de completitud de las muestras en la obtención de los resultados.

La figura 4.5 presenta las figuras de mérito C y M (gráficos superior e inferior respectivamente), para un sub-muestreo del 50% (línea azul punteada), del 20% (línea magenta discontinua), y del 10% (línea negra continua), en función de DC , cuando utilizamos el método geométrico. Los resultados se muestran de una forma promedio para los tres análisis de pertenencia realizados usando $N = 1, 2$ y 4 variables. Como norma general se observa que los valores promedios de ambas figuras de mérito para el sub-muestreo correspondiente al 50%, son similares a los obtenidos con el sub-muestreo del 20% y a los resultados obtenidos en la muestra completa. Aunque los valores promedios obtenidos en el

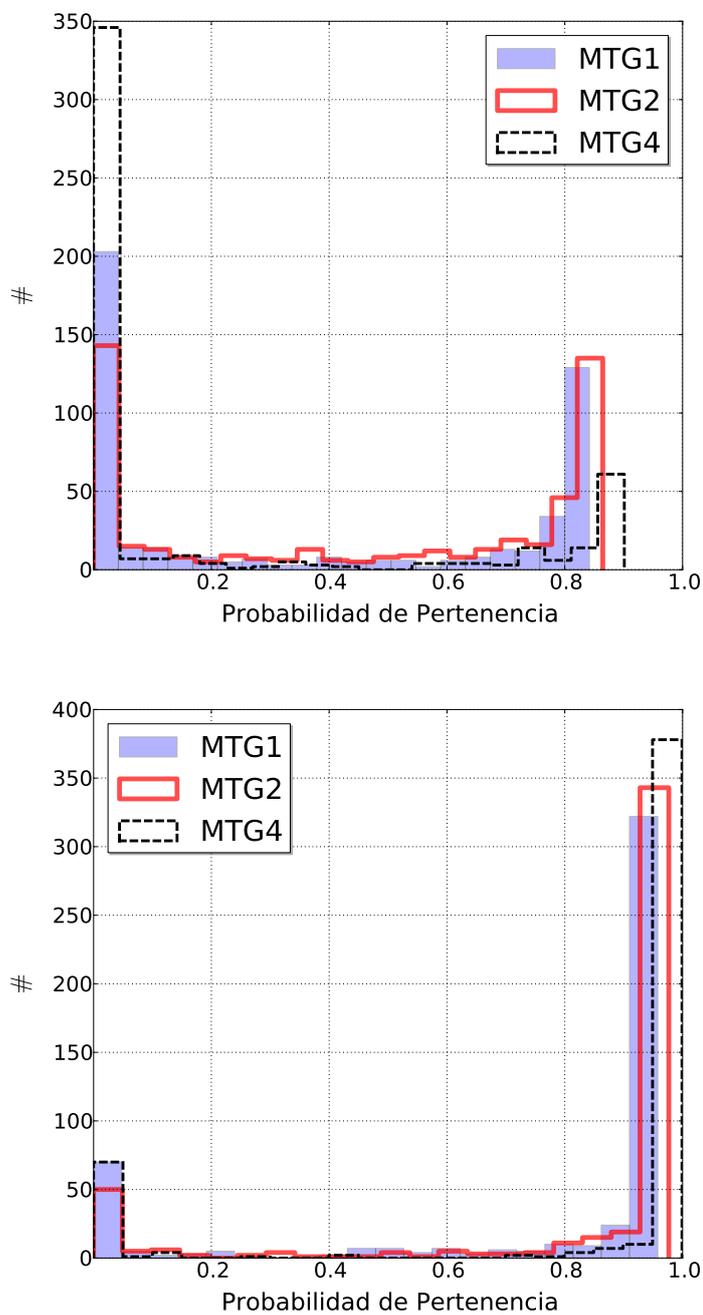


Figura 4.3: Distribución de las probabilidades de pertenencia al cúmulo obtenidas para MTG1, MTG2 y MTG4. Estos resultados se corresponden con el mismo modelo simulado en el que el único parámetro variable es la fracción de estrellas de cúmulo de un 20% arriba, y de un 80% abajo.

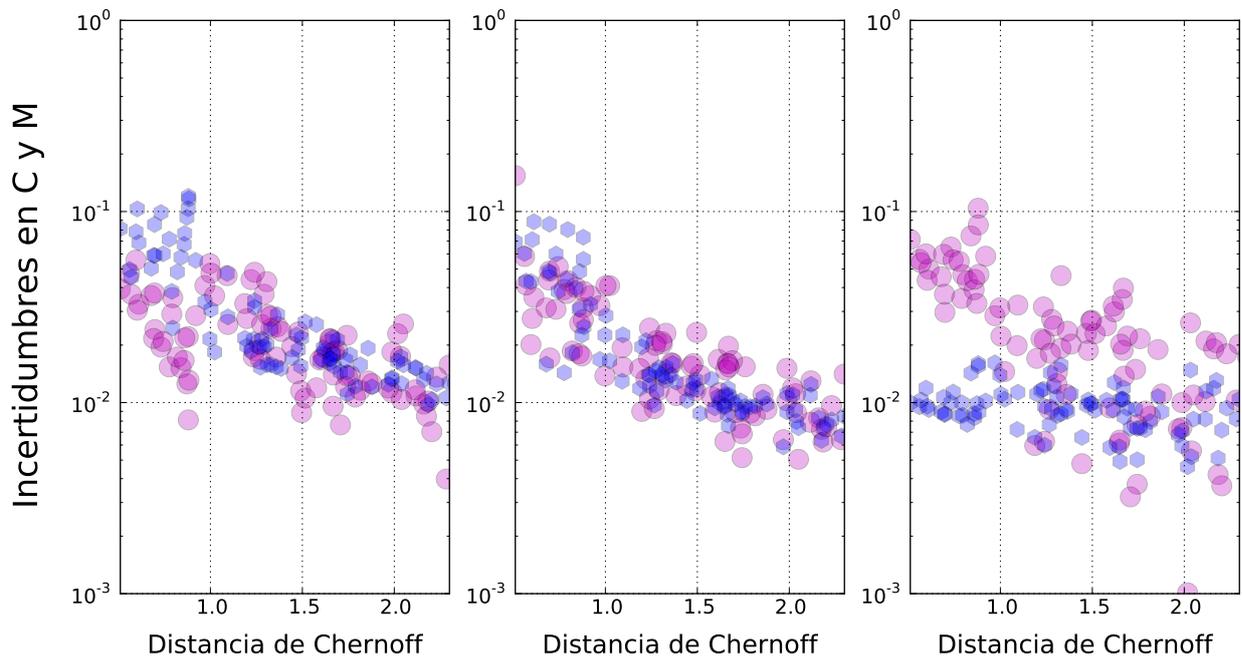


Figura 4.4: Distribución de los errores de C y M obtenidos para MTG1 (gráfico de la izquierda), MTG2 (gráfico del medio) y MTG4 (gráfico de la derecha), en función de la DC , a partir de la aplicación de la técnica del Bootstrap. Los círculos magenta y los hexágonos azules representan las incertidumbres en los errores de C y M , respectivamente.

sub-muestreo del 10% no son muy diferentes a los obtenidos para los otros dos casos, la dispersión de estos valores (región sombreada correspondiente a 1σ) sí es mayor, indicando una mayor inestabilidad en la bondad de la clasificación.

Las tablas 4.1 y 4.2 muestran el promedio (sobre las diferentes proporciones de estrellas de cúmulo) de las diferencias entre los parámetros del cúmulo simulado y del recuperado ($\Delta\bar{x}_c, \Delta\bar{\sigma}_{x_c}, \Delta\bar{y}_c, \Delta\bar{\sigma}_{y_c}, \Delta\bar{\mu}_{c,x}, \Delta\bar{\sigma}_{\mu_{c,x}}, \Delta\bar{\mu}_{c,y}, \Delta\bar{\sigma}_{\mu_{c,y}}$), junto con sus dispersiones ($\sigma_{\Delta\bar{x}_c}, \sigma_{\Delta\bar{\sigma}_{x_c}}, \sigma_{\Delta\bar{y}_c}, \sigma_{\Delta\bar{\sigma}_{y_c}}, \sigma_{\Delta\bar{\mu}_{c,x}}, \sigma_{\Delta\bar{\sigma}_{\mu_{c,x}}}, \sigma_{\Delta\bar{\mu}_{c,y}}, \sigma_{\Delta\bar{\sigma}_{\mu_{c,y}}}$), obtenidos por MTG1, MTG2 y MTG4 y para los tres casos de sub-muestreo, incluyendo los test iniciales con 500 objetos, en los espacios de posiciones y de movimientos propios respectivamente. Para estimar dichos parámetros, en ambos espacios, se hizo uso de las siguientes expresiones:

$$Z = \frac{\sum_i z_i P_i}{\sum_i P_i} \quad (4.4)$$

$$\sigma_Z = \frac{\sum_i P_i (z_i - z)^2}{\sum_i P_i} \quad (4.5)$$

donde Z y σ_Z refieren a los parámetros de cúmulo obtenidos por una metodología, pesados por las probabilidades de pertenencia (P_i) asignadas a las estrellas de la muestra.

Otra vez se manifiesta que la inclusión de un mayor número de variables y un muestreo más completo mejoran los resultados y recuperan los parámetros que definen la distribución del cúmulo en el sub-espacio de fase.

4.2.2 Error Gaussiano con σ_μ dependiente de la magnitud.

En lo que sigue nos referiremos a estas simulaciones como *Sim2*. El objetivo último de esta tesis consiste en la obtención de un catálogo de cúmulos abiertos con asignación de probabilidades de pertenencia a cada una de las estrellas de la muestra a analizar, obtenidas tras la aplicación de las metodologías detalladas en el capítulo 2. Para conocer y poder describir cómo las diferentes metodologías llevan a cabo la clasificación en cúmulo y campo, su aplicación a simulaciones lo más parecidas a la muestra que vamos a utilizar, es fundamental.

Debido a que las metodologías geométrica y paramétrica requieren de unos parámetros iniciales para iniciar el proceso iterativo, el valor de estos parámetros se ha calculado de la siguiente manera. Para la metodología paramétrica se han tomado los valores medios y las dispersiones de las muestras a analizar (una vez eliminados los *outliers*), como valores representativos de la población de campo. Los valores de entrada para las dispersiones de la población de cúmulo son 5 veces inferiores a las de campo. En cuanto al valor de los centroides de los cúmulos se ha tomado, como hipótesis de partida para ambas metodologías, las modas de las distribuciones.

En la figura 4.6 se pueden ver los valores de C y M (gráficos superior e inferior respectivamente), obtenidos por MTG2, para los cuatro porcentajes de estrellas de cúmulo simulados (20% en magenta, 40% en azul, 60% en verde y 80% en negro), en función de DC . Es de notar que el intervalo de DC

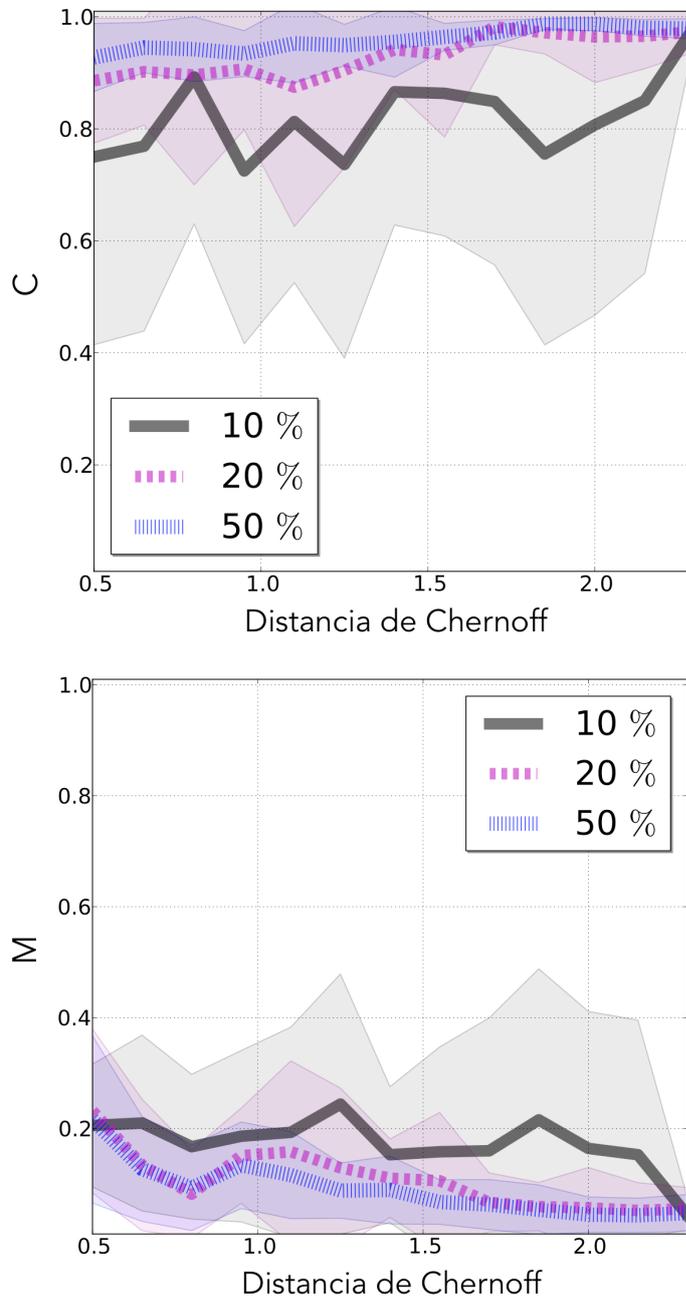


Figura 4.5: Influencia en la determinación de C y de M de los diferentes sub-muestras observacionales simulados, respecto de la DC . Los resultados referidos a un 50%, a un 20% y a un 10% de la muestra se representan por una línea punteada azul, por una línea magenta discontinua y por una línea continua negra, respectivamente. Las zonas coloreadas hacen referencia a una dispersión de 1σ .

Tabla 4.1: Diferencias y dispersiones de los valores medios obtenidos respecto de los simulados, para $N = 1, 2$ y 4 variables, de los parámetros asociados a las distribuciones de posiciones para los cuatro casos de sub-muestreo simulados.

Muestreo	Nº Var	Coordenadas Espaciales del Cúmulo (deg)			
		$\Delta\bar{x}_c, \sigma_{\Delta\bar{x}_c}$	$\Delta\bar{\sigma}_{x_c}, \sigma_{\Delta\bar{\sigma}_{x_c}}$	$\Delta\bar{y}_c, \sigma_{\Delta\bar{y}_c}$	$\Delta\bar{\sigma}_{y_c}, \sigma_{\Delta\bar{\sigma}_{y_c}}$
100%					
	1 Var	0.000, 0.003	-0.019, 0.011	0.000, 0.003	-0.019, 0.011
	2 Var	0.000, 0.002	-0.011, 0.009	0.000, 0.002	-0.011, 0.009
	4 Var	0.000, 0.001	0.001, 0.001	0.000, 0.001	0.001, 0.001
50%					
	1 Var	0.000, 0.004	-0.019, 0.012	0.000, 0.004	-0.019, 0.012
	2 Var	0.001, 0.004	-0.011, 0.009	0.000, 0.003	-0.011, 0.009
	4 Var	0.001, 0.003	0.001, 0.002	0.001, 0.002	0.001, 0.002
20%					
	1 Var	0.001, 0.011	-0.019, 0.013	0.000, 0.009	-0.019, 0.013
	2 Var	0.001, 0.006	-0.010, 0.011	0.000, 0.005	-0.012, 0.012
	4 Var	0.000, 0.006	0.002, 0.004	0.000, 0.005	0.001, 0.004
10%					
	1 Var	0.005, 0.028	-0.018, 0.019	0.000, 0.022	-0.021, 0.017
	2 Var	-0.001, 0.013	-0.014, 0.016	0.003, 0.015	-0.015, 0.017
	4 Var	0.002, 0.014	0.003, 0.007	0.000, 0.010	0.002, 0.008

Tabla 4.2: Diferencias y dispersiones de los valores medios obtenidos respecto de los simulados, para $N = 1, 2$ y 4 variables, de los parámetros asociados a las distribuciones de movimientos propios para los cuatro casos de sub-muestreo simulados.

Muestreo	Nº Var	Coordenadas Cinemáticas del Cúmulo (mas/yr)			
		$\Delta\bar{\mu}_{c,x}, \sigma_{\Delta\bar{\mu}_{c,x}}$	$\Delta\bar{\sigma}_{\mu_{c,x}}, \sigma_{\Delta\bar{\sigma}_{\mu_{c,x}}}$	$\Delta\bar{\mu}_{c,y}, \sigma_{\Delta\bar{\mu}_{c,y}}$	$\Delta\bar{\sigma}_{\mu_{c,y}}, \sigma_{\Delta\bar{\sigma}_{\mu_{c,y}}}$
100%					
	1 Var	-0.013, 0.695	-5.693, 3.442	0.142, 0.316	0.031, 0.379
	2 Var	0.001, 0.147	0.033, 0.272	0.093, 0.373	-0.114, 0.400
	4 Var	-0.025, 0.109	-0.147, 0.279	0.027, 0.151	-0.231, 0.317
50%					
	1 Var	-0.143, 0.846	-5.531, 3.634	0.244, 0.561	-0.039, 0.712
	2 Var	-0.020, 0.385	0.006, 0.420	0.142, 0.433	-0.154, 0.534
	4 Var	-0.026, 0.229	-0.137, 0.382	0.049, 0.289	-0.232, 0.391
20%					
	1 Var	-0.023, 1.541	-5.771, 4.226	0.782, 3.086	-0.170, 3.454
	2 Var	0.012, 0.491	0.003, 0.651	0.275, 0.759	-0.083, 0.928
	4 Var	-0.062, 0.517	-0.104, 0.496	0.137, 0.436	-0.119, 0.534
10%					
	1 Var	0.763, 5.752	-6.262, 6.056	2.491, 9.565	-0.696, 5.546
	2 Var	0.722, 3.964	-0.651, 2.400	1.595, 3.641	-0.851, 2.663
	4 Var	0.017, 1.839	-0.135, 1.457	0.198, 1.079	-0.147, 1.317

es diferente al obtenido en las simulaciones previas (ver por ejemplo la figura 4.2). Este cambio en el intervalo observado de DC se debe principalmente a la inclusión de errores más realistas que aumentan la dispersión de ambas poblaciones en el espacio de movimientos propios y, por lo tanto, disminuyen su heterocedasticidad.

Al contrario que en los resultados obtenidos con *Sim1*, la nueva distribución de C y M con DC (ver 4.6), presenta un mayor solapamiento entre los porcentajes simulados DC (ver figura 4.2), aunque al menor de los porcentajes le sigue correspondiendo los valores más bajos de DC . En estas simulaciones los gráficos son mucho más ruidosos que en las simulaciones anteriores indicando una menor estabilidad de las metodologías, fundamentalmente para aquellas simulaciones con menor heterocedasticidad y para los menores porcentajes de estrellas de cúmulo. Para DC alrededor de 0.6, el parámetro M varía entre un 10% y un 40%, y de igual forma la completitud se encuentra en el rango entre el 55% y el 90%. De nuevo se observa que a medida que DC aumenta todos los métodos analizados en esta tesis recuperan mejor al cúmulo aumentando la completitud (C) y disminuyendo el coeficiente de mal clasificados (M). Los resultados obtenidos con el resto de métodos y números de variables, se pueden ver en el apéndice A, donde se muestran de forma comparativa los resultados obtenidos para *Sim1* y *Sim2*.

Los gráficos 4.7, 4.8 y 4.9 muestran el comportamiento de las dos figuras de mérito en función de la DC , para $N = 1, 2$ y 4 variables, respectivamente. En cada una de las figuras se comparan los resultados obtenidos por cada metodología respecto a los obtenidos por las mismas en las *Sim1*. Tal y como era de esperar, la introducción de errores peor comportados produce un empeoramiento en los resultados obtenidos de una manera generalizada. Particularmente diferentes son los obtenidos por los métodos MT2 y MTG2. En la figura 4.8 se puede ver una diferencia en los valores de ambos parámetros en torno al 20% en C y al 10-15% en M para valores de $DC < 1$, manteniéndose en torno al 10% para valores más altos ($DC > 1$). La inclusión de las posiciones en los análisis de pertenencia revierte en una mejor clasificación de la muestra en cúmulo y campo (ver figura 4.9). Todas las metodologías presentan una mejora en sus resultados conforme la DC aumenta.

Si comparamos los resultados obtenidos por el método geométrico respecto a los obtenidos por las otras metodologías, centrándonos únicamente en estas segundas simulaciones, se puede observar que la C obtenida por el método geométrico es mejor para $N = 1$ y 2 variables, especialmente para $DC < 1$, proporcionando unos valores muy similares a los de MT4. La M obtenida por MTG1 y MTG2 es similar a la obtenida por MT1 y MT2, siendo mejor la obtenida por MTG4, fundamentalmente para $DC < 1$.

Los resultados mostrados en las figuras anteriores se corresponden con aquellas simulaciones en las que los diferentes métodos aplicados han determinado la existencia de cúmulo, pero hay realizaciones de los modelos donde no se ha producido esta determinación. En general podemos decir que esto ocurre cuando la proporción de estrellas en la muestra es del 20%. Sin embargo, hay que hacer notar que MT1 es la que muestra un mayor número de fallos, con un total de 17 simulaciones donde no ha encontrado cúmulo, 10 de ellas correspondientes al porcentaje del 20%.

Utilizando las simulaciones *Sim2* se ha estudiado también el efecto del posible sub-muestreo pre-

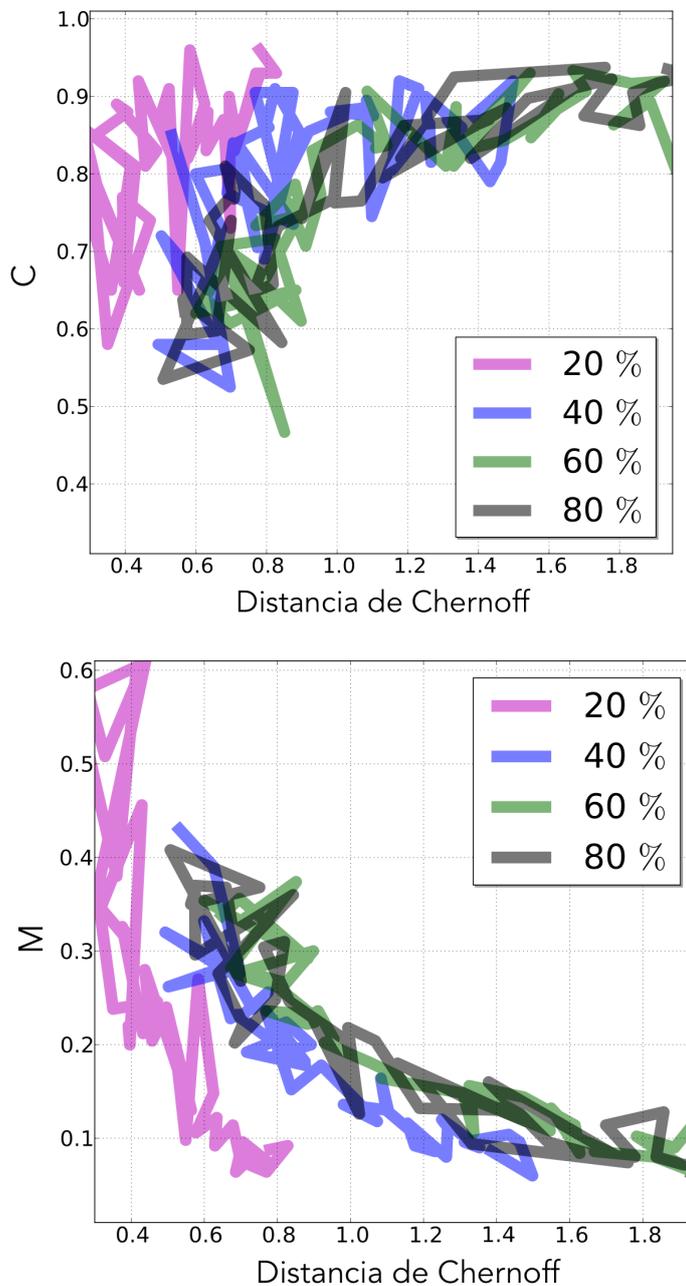


Figura 4.6: Dependencia con el porcentaje de estrellas de cúmulo simuladas de los valores de C y de M respecto a la DC . Los resultados obtenidos para un número de estrellas de cúmulo igual al 20% de la muestra total se representan en magenta, el 40% en azul, el 60% en verde y el 80% en negro. Estos resultados son los obtenidos por la nueva metodología para $N = 2$ variables. Los valores más bajos de DC ($DC < 0.5$) se corresponden con el porcentaje más bajo de estrellas de cúmulo simulado. Para los porcentajes más altos, la DC aumenta y se observa una disminución en los valores de M y un aumento en los valores de C obtenidos.

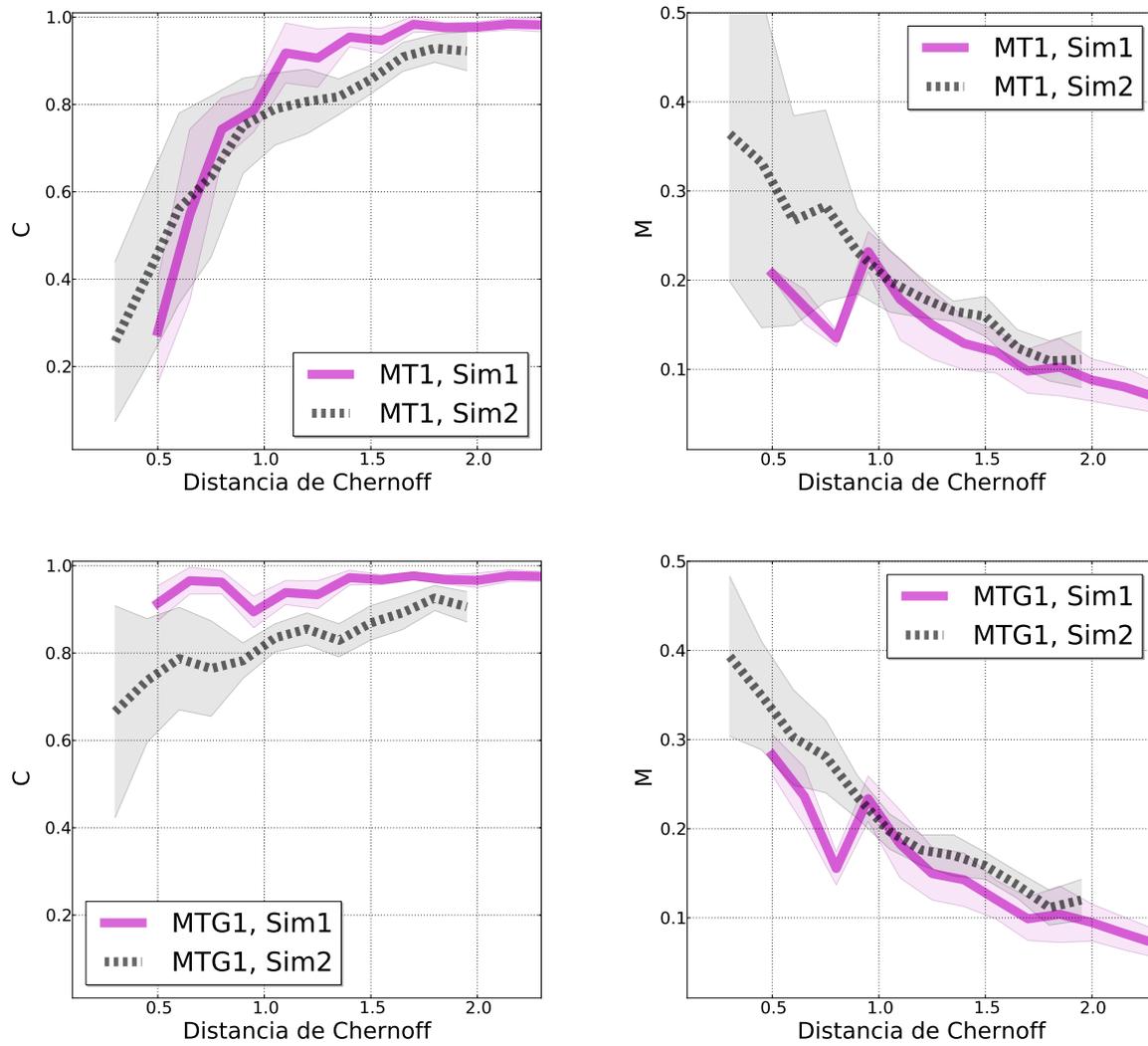


Figura 4.7: Comparación de los valores de ambas figuras de mérito obtenidos por las metodologías MT1 y MTG1, respecto a la DC en las $Sim1$ y $Sim2$. En este caso, las líneas continuas magenta refieren a los resultados obtenidos por las metodologías en $Sim1$, mientras que las líneas discontinuas negras muestran los resultados obtenidos en las $Sim2$. Las áreas coloreadas representan de nuevo la dispersión (de 1σ) de los resultados en intervalos de 0.15 unidades de distancia.

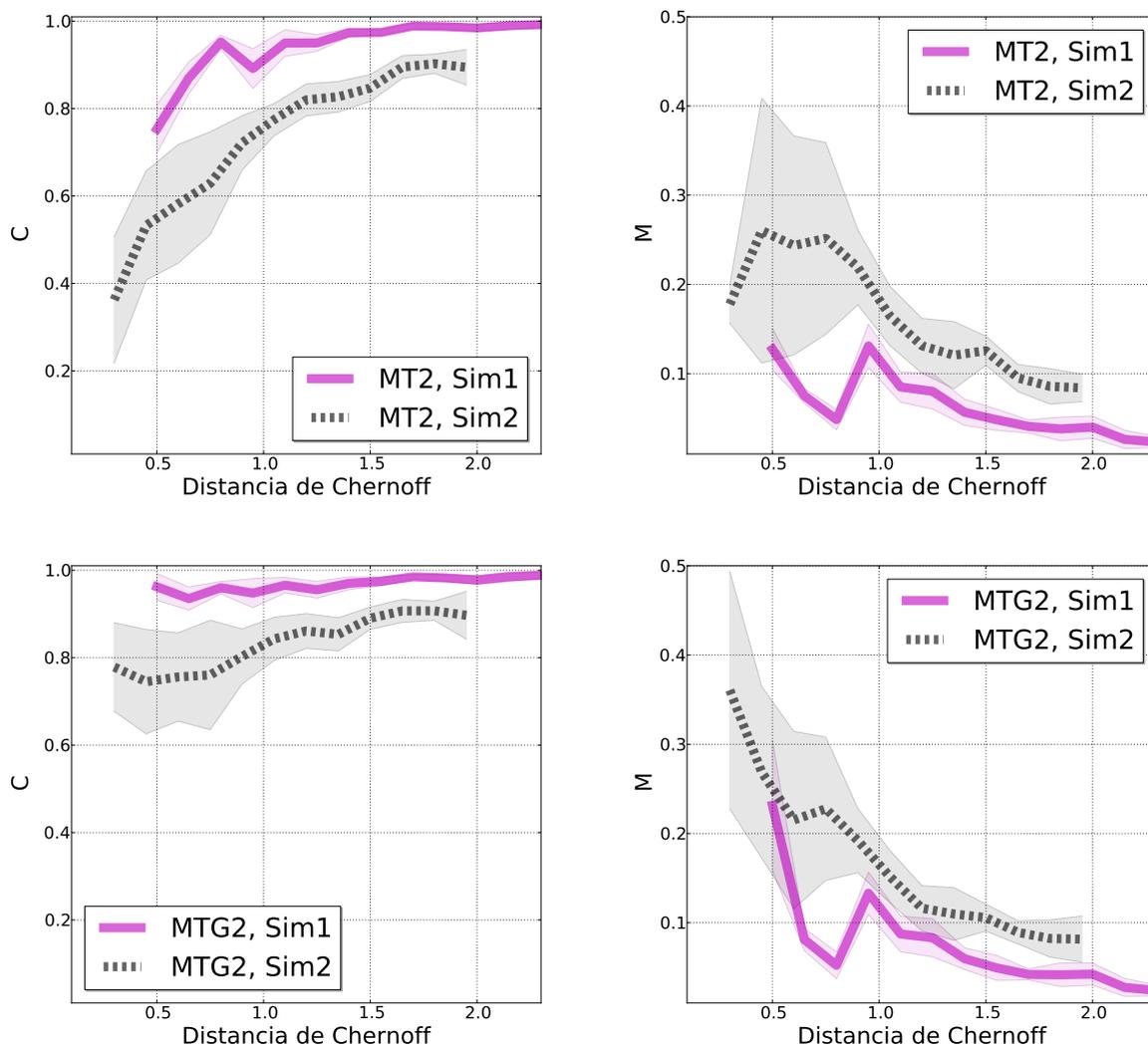


Figura 4.8: Comparación de los valores de ambas figuras de mérito obtenidos por las metodologías MT2 y MTG2, respecto a la DC en las $Sim1$ y $Sim2$. En este caso, las líneas continuas magenta refieren a los resultados obtenidos por las metodologías en $Sim1$, mientras que las líneas discontinuas negras muestran los resultados obtenidos en las $Sim2$. Las áreas coloreadas representan de nuevo la dispersión (de 1σ) de los resultados en intervalos de 0.15 unidades de distancia.

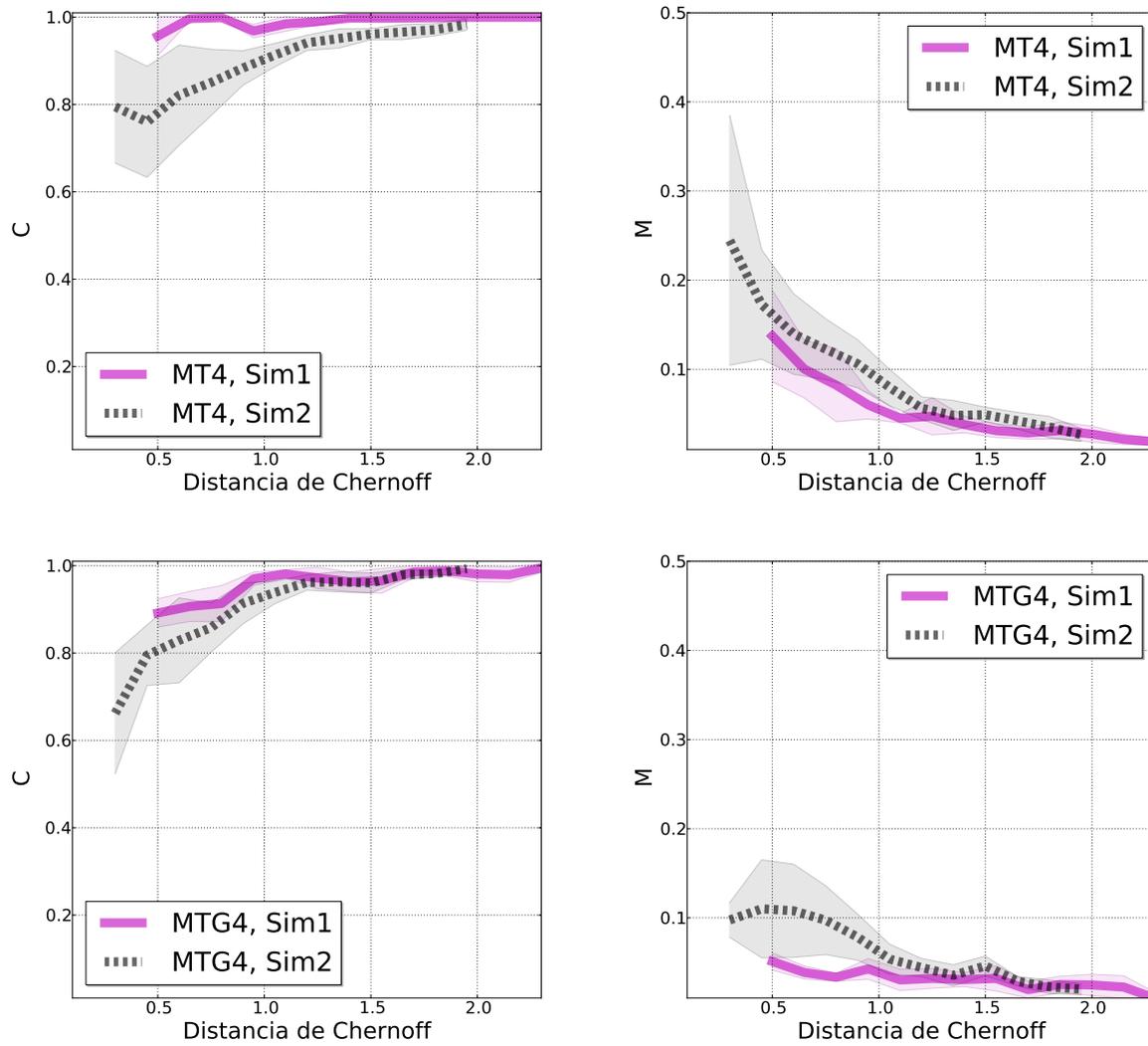


Figura 4.9: Comparación de los valores de ambas figuras de mérito obtenidos por las metodologías MT4 y MTG4, respecto a la DC en las $Sim1$ y $Sim2$. En este caso, las líneas continuas magenta refieren a los resultados obtenidos por las metodologías en $Sim1$, mientras que las líneas discontinuas negras muestran los resultados obtenidos en las $Sim2$. Las áreas coloreadas representan de nuevo la dispersión (de 1σ) de los resultados en intervalos de 0.15 unidades de distancia.

sente en datos reales para todas las metodologías descritas en el capítulo 2. Los resultados presentan un empeoramiento gradual, para todos los métodos utilizados, conforme el muestreo es más incompleto. En la figura 4.10 se muestran, como ejemplo, los resultados obtenidos por MTG2, para los cuatro porcentajes de estrellas de cúmulo. En el apéndice B se presentan los resultados obtenidos por las diferentes aproximaciones en función del número de variables utilizadas para realizar el análisis de pertenencia. Estos gráficos muestran una mejora en los resultados obtenidos tanto cuando la DC es mayor entre las poblaciones de cúmulo y campo como cuando el sub-muestreo presente en la muestra es menor (sub-muestreo del 50%). Se observa, en general, que los resultados obtenidos por la metodología geométrica son mejores que los obtenidos por el resto de metodologías al usar el mismo número de variables. La inclusión de un mayor número de variables para la realización de los análisis de pertenencia revierte en la obtención de unos mejores resultados tanto en los valores de C como de M , así como en una menor dispersión de sus valores obtenidos (zonas sombreadas).

4.2.3 Efecto de la simulación de distribuciones aleatorias para la población de cúmulo.

En todos los casos previos hemos siempre considerado que el cúmulo era reconocible por una sobre-densidad espacial respecto a las estrellas de campo y que esta densidad podría ser modelada por una distribución Gaussiana bivalente circular. Sin embargo, los cúmulos reales no necesariamente se comportan de esta forma, y lo que es peor, nuestra representación del cúmulo, los catálogos astrométricos que lo definen, adolecen de sesgos muestrales que conducen a representaciones muy alejadas de la realidad. Por ello se ha decidido realizar un tercer tipo de simulaciones (nos referiremos a estas simulaciones como *Sim3*) donde la distribución espacial de las estrellas de cúmulo viene caracterizada por una función aleatoria. De esta forma hemos comparado los métodos MTG4 y MT4, ya que son los únicos que incorporan las variables espaciales en sus análisis.

Los resultados obtenidos, que se muestran en la figura 4.11, claramente manifiestan que cuando no se cumplen las hipótesis fundamentales del modelo, éste no puede llevar a soluciones realistas. El método geométrico se basa en que en un espacio de N -D, la distancia de los miembros del cúmulo al centroide del mismo será inferior a la de las estrellas de campo, lo cual implícitamente conlleva la existencia de una sobre-densidad no uniforme, sino con cierta dependencia radial. De otra forma, la inclusión del espacio de coordenadas sólo introduce ruido y no permite una clasificación creíble. Esto se manifiesta en la alta dispersión de los valores medios obtenidos por MTG4. Por el contrario, MT4 mantiene unos valores en concordancia con los de las simulaciones *Sim2*, sólo que ahora DC ha disminuido debido a la similitud (baja heterocedasticidad) de las distribuciones espaciales de las poblaciones de cúmulo y campo. Esto es debido a que la definición de cúmulo implícita en la selección de miembros por MT4, permite una redistribución del peso de los subconjuntos de coordenadas y movimientos propios en función de su grado de concentración.

Estas simulaciones muestran la importancia de elegir las variables adecuadas para cada problema en cuestión, o por lo menos si trabajamos en modo industrial, conocer la calidad de la clasificación que estamos realizando. Estas ideas nos ayudarán a describir los resultados obtenidos por las metodologías cuando son aplicadas a catálogos de cúmulos reales.

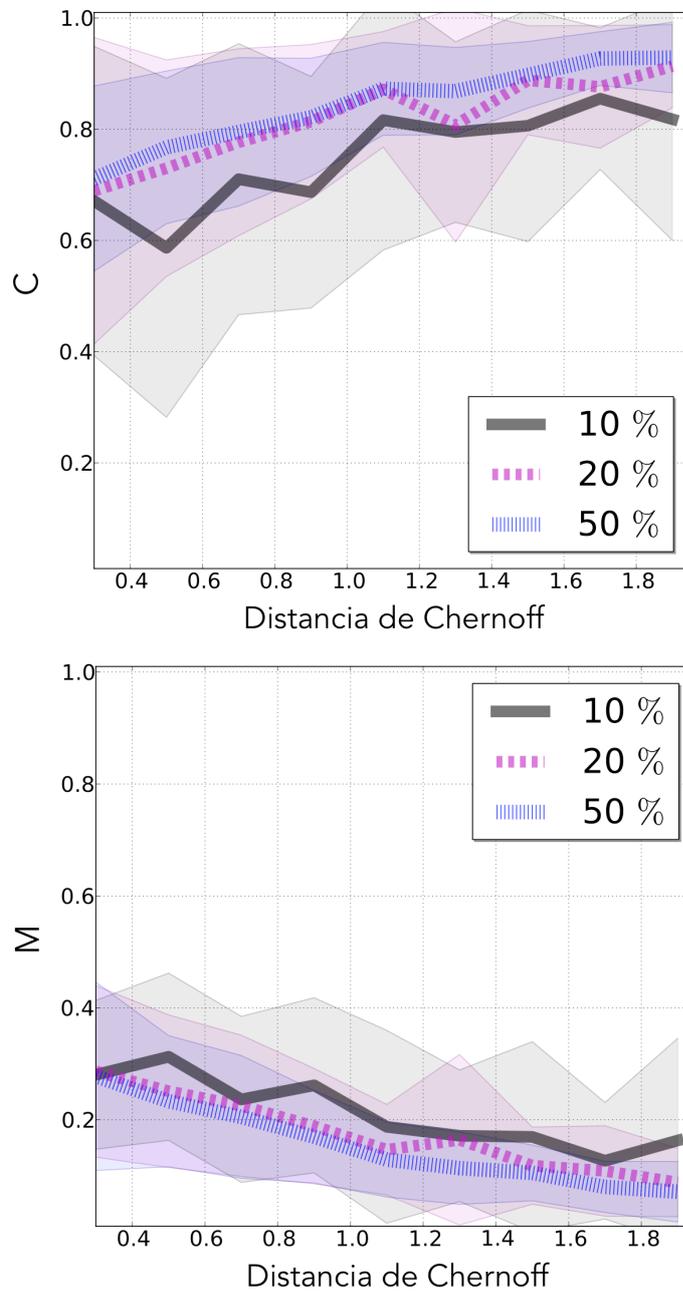


Figura 4.10: Influencia en la determinación de C y de M de los diferentes sub-muestras observacionales simulados, respecto de la DC . Los resultados promedios para $N = 1, 2$ y 4 variables referidos a un 50%, a un 20% y a un 10% de la muestra se muestran por una línea punteada azul, por una línea magenta discontinua y por una línea continua negra, respectivamente. Las zonas coloreadas hacen referencia a una dispersión de 1σ .

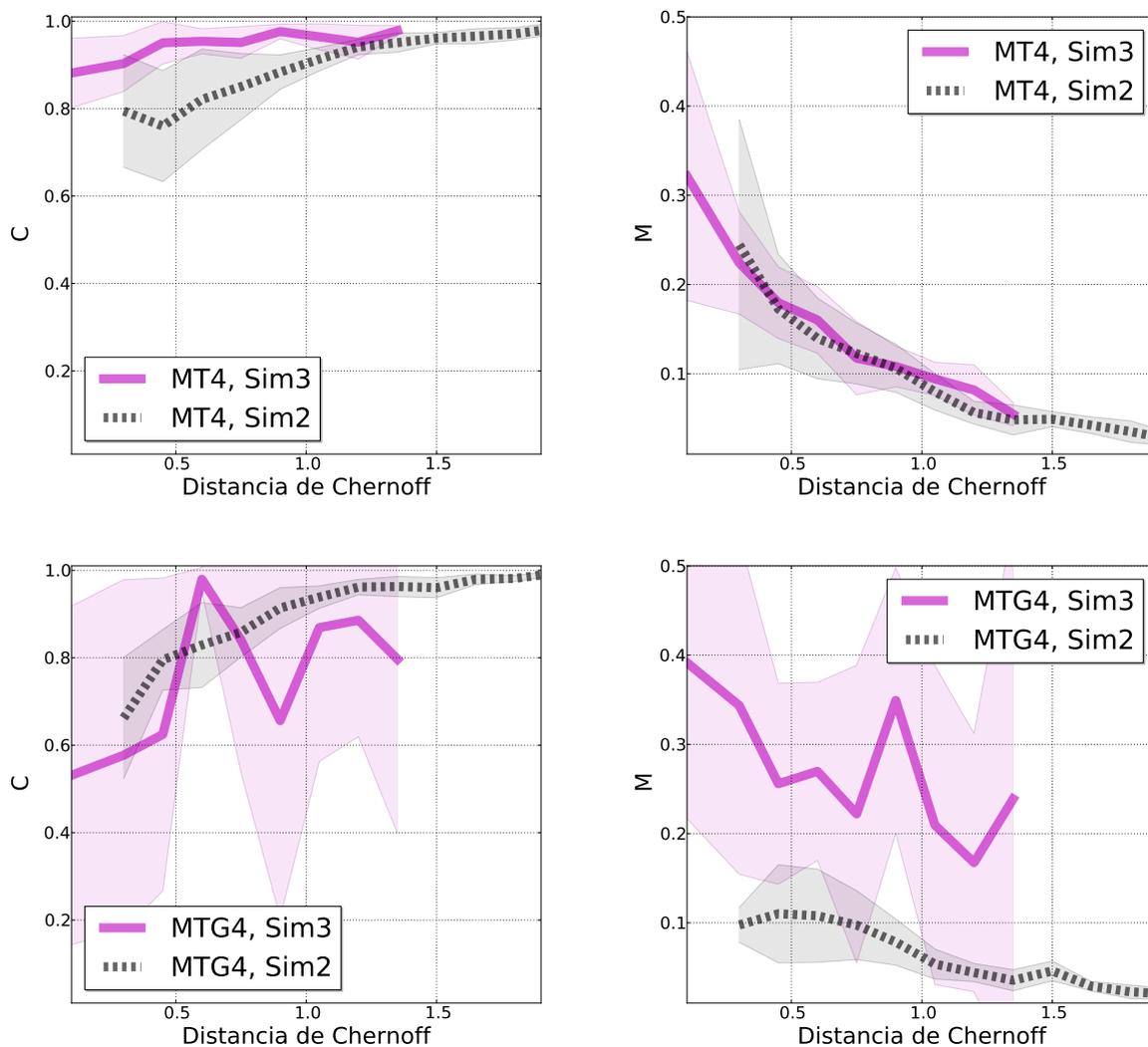


Figura 4.11: Comportamiento de ambas figuras de mérito respecto a la DC , cuando la distribución de las posiciones de las estrellas de cúmulo es aleatoria. En los gráficos superiores se muestran los resultados obtenidos por MT4, mientras que en los inferiores por MTG4. Las líneas continuas magenta refieren a los resultados obtenidos en las *Sim3*, mientras que las líneas discontinuas negras muestran los resultados obtenidos en las *Sim2*. Las áreas coloreadas representan la dispersión (de 1σ) de los resultados en intervalos de 0.15 unidades de distancia. Se observa un empeoramiento en los resultados siendo principalmente acusado para MTG4.

5

Catálogo de cúmulos reales.

En este capítulo se presentan los resultados obtenidos tras la aplicación de las diferentes metodologías a los 2167 cúmulos listados en el catálogo DAML02, usando los datos procedentes del cartografiado UCAC4. Al estudiar los perfiles de densidad radial (PDR) de los cúmulos a analizar, usando estos datos, se observa que los radios publicados en DAML02 no se ajustan a los PDR observados en todos los casos. Por tanto, una tarea previa a la realización del análisis de pertenencia, es la estimación de los radios de los cúmulos que a su vez serán los radios muestrales de los datos del cartografiado UCAC4.

Comenzaremos detallando las características del catálogo DAML02 así como el procedimiento a partir del cual se determinaron los radios de los cúmulos. Los resultados obtenidos por los diferentes análisis de pertenencia, así como su comparación con los publicados en el trabajo de D14 y en DAML02, serán analizados en la última parte de este capítulo.

5.1 Catálogo DAML02.

El catálogo DAML02 está constituido por 2167 cúmulos abiertos de nuestra Galaxia, siendo una actualización de los catálogos previos de Lynga & Palous (1987) y Mermilliod (1995) (incluido en la base de datos WEBDA¹), con la adición de resultados obtenidos por otros autores.

Para cada cúmulo el catálogo incluye: el nombre, coordenadas angulares, un indicador con la clasificación del cúmulo, el diámetro angular (para el 99.7% de los cúmulos), el centroide de movimientos propios junto con su error (para el 97.2%), el número de estrellas catalogadas como miembros, distancia (para el 94%), exceso de color (para el 93.3%), edad (para el 92.8%) y para muchos de ellos datos de sus velocidades radiales y metalicidades. Los cúmulos listados en este catálogo tienen un rango de diámetros proyectados de entre 0.2 y 1400 minutos de arco (ma), en un intervalo de distancias de entre 25 y 14871 pc y con edades comprendidas entre $\log(t) = 6$ y 10.1.

¹<http://obswww.unige.ch/webda/>

El radio estimado de los cúmulos en este catálogo tiene un origen heterogéneo. Así, algunos de estos radios provienen de la inspección visual de imágenes de DSS (*Digitized Sky Survey*) por Dias W. S. y colaboradores, y otros de trabajos como Bonatto & Bica (2010), Bonatto & Bica (2010b), Camargo, Bonatto, & Bica (2010), Glushkova et al. (2010), Tadross (2012) y Carraro, Beletsky, & Marconi (2013).

Los datos de movimientos propios en DAML02 proceden de diversas fuentes y, la pertenencia al cúmulo se determinó por diferentes técnicas y autores. En particular, 1594 cúmulos proceden del trabajo de Dias et al. (2014), 179 cúmulos de Kharchenko et al. (2013), 91 de Kharchenko et al. (2012), 78 de Baumgardt, Dettbarn, & Wielen (2000), 54 de Kharchenko et al. (2005), 37 de Dias, Lépine & Alessi (2001, 2002), 14 de Kharchenko, Pakulyak, & Piskunov (2003), 10 de Platais, Kozhurina-Platais, & van Leeuwen (1998), 8 de Alessi, Moitinho, & Dias (2003) y los demás de pequeñas compilaciones como por ejemplo Dias et al. (2006) y de Magrini et al. (2010).

5.2 Determinación de los radios de los cúmulos abiertos listados en DAML02.

5.2.1 Introducción

Un cúmulo abierto está formado por dos regiones fundamentales: una región central más densa denominada núcleo y un halo (o corona) correspondiente con la zona más externa y extensa con una menor densidad de estrellas, la cual puede contener un número sustancial de miembros del cúmulo (hasta el 75% según Nilakshi et al. (2002)). La definición funcional de estas dos regiones es la siguiente. El radio del núcleo r_c se define como aquella distancia a la cual la densidad de estrellas decae a la mitad respecto de la sobre-densidad central. De la misma manera, el radio de marea r_t se define como aquella distancia a la cual la densidad superficial del cúmulo decae a cero, o se iguala al nivel de la densidad superficial de las estrellas de campo. Dado que el radio de marea engloba también a la región de la corona, éste es considerado el radio del cúmulo.

La existencia de un cúmulo en el espacio de posiciones se presenta como una sobre-densidad respecto a una distribución constante de estrellas de campo. A partir de un perfil de densidad radial y, presuponiendo una distribución de estrellas de campo homogénea en posiciones, se podría definir el radio de un cúmulo como aquella distancia donde la densidad de estrellas se iguala a la densidad de estrellas de campo.

Esta definición es válida para cualquier tipo de cúmulo, bien sean cúmulos con PDRs donde se observa una sobre-densidad central que decrece con la distancia al centro, como para cúmulos caracterizados por la presencia de una 'meseta' con una densidad aproximadamente constante y mayor que la observada para el campo. Un ejemplo de ambos tipos de distribuciones se representa en la figura 5.1, donde se muestran los PDRs de los cúmulos Berkeley 90 y NGC 2451B, en los gráficos superior e inferior, respectivamente.

Por otra parte, conviene indicar que para evitar distorsiones espaciales en la representación de cúmulos con gran radio angular, las coordenadas ecuatoriales (α, δ) han sido transformadas a coor-

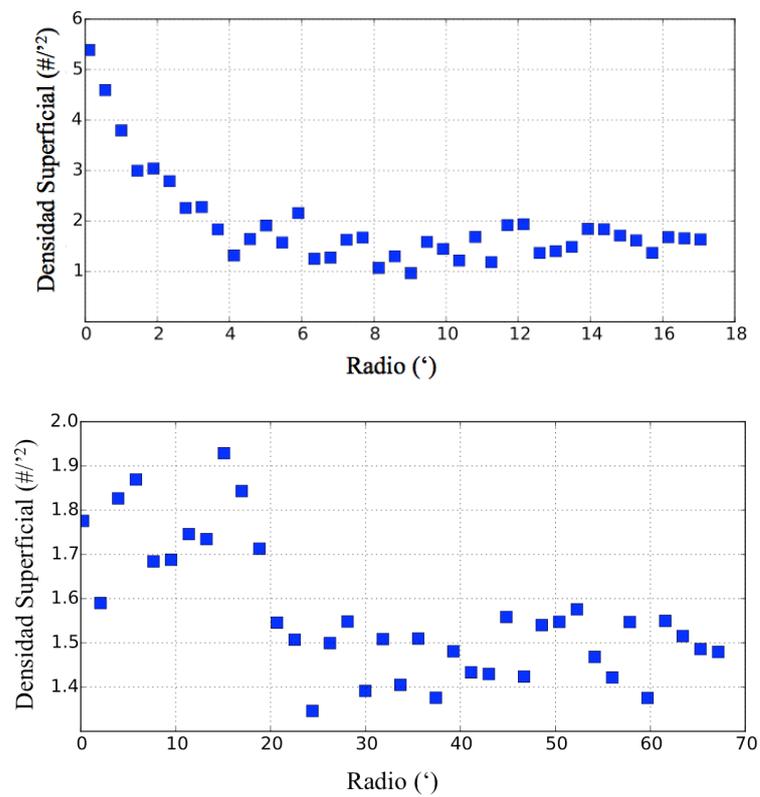


Figura 5.1: Ejemplos de PDRs de los cúmulos Berkeley 90 (panel superior) e NGC 2451B (panel inferior). Mientras Berkeley 90 es un cúmulo con estructura radial, NGC 2451B es un cúmulo caracterizado por la presencia de una meseta en el PDR observado.

denadas tangenciales (ξ, η) de acuerdo con las expresiones dadas por Green (1985):

$$\xi = \frac{\cos\delta \sin(\alpha - \alpha_0)}{\sin\delta_0 \sin\delta + \cos\delta_0 \cos\delta \cos(\alpha - \alpha_0)} \quad (5.1)$$

$$\eta = \frac{\cos\delta_0 \sin\delta - \sin\delta_0 \cos\delta \cos(\alpha - \alpha_0)}{\sin\delta_0 \sin\delta + \cos\delta_0 \cos\delta \cos(\alpha - \alpha_0)} \quad (5.2)$$

donde α y δ son las coordenadas ecuatoriales de la estrella y, α_0 y δ_0 son las coordenadas del centroide del cúmulo. En este trabajo el radio de los cúmulos ha sido calculado respecto a estas últimas coordenadas.

A la hora de elegir la muestra de estrellas de un catálogo para realizar su análisis de pertenencia al cúmulo es importante determinar el radio del mismo. Si elegimos la muestra con un radio muestral inferior al real, estaremos dejando fuera a estrellas del cúmulo y la selección ya estaría sesgada. Sin embargo, si elegimos un radio muestral muy superior al radio real, estamos incluyendo muchas más estrellas de campo, y sesgando las probabilidades a priori del campo hacia valores más altos, lo que podría redundar en una selección de estrellas de cúmulo inferior a la real. Tal y como se señala en el trabajo de Sánchez, Vicente, & Alfaro (2010), al menos para el método paramétrico utilizando movimientos propios, la elección ideal es tomar el radio muestral como el radio del cúmulo. Esto puede parecer un círculo vicioso, donde los miembros definen el radio del cúmulo y la selección de objetos dentro de ese radio definen la mejor muestra para el análisis de pertenencia. Como ya hemos comentado, la elección del mejor radio muestral fue analizada para el método paramétrico pero no para las otras metodologías utilizadas en este trabajo de tesis. No obstante aplicamos esta receta como válida para la elección de nuestra muestra.

Aunque la definición de radio parezca obvia, su determinación mediante datos reales no está carente de ciertas dificultades, haciendo que su determinación tenga un carácter subjetivo difícil de introducir en un procedimiento automático. La gran variedad en la forma y en el número de estrellas son características típicas de los cúmulos abiertos. En muchos casos, la edad de los cúmulos es inferior al tiempo de relajación necesario para alcanzar el equilibrio dinámico, dando lugar a distribuciones no homogéneas características de la estructura de la nube molecular de la cual proceden. Por otro lado, las fuerzas de marea de la Galaxia (Fujii & Baba 2012; Fukushige & Heggie 2000), las interacciones tanto con otros grupos de estrellas como con por brazos espirales o las nubes de gas y polvo (Gieles et al., 2006), pueden producir la evaporación del cúmulo al perturbar sus distribuciones cinemáticas (Lynga & Palous 1987; Carlberg & Sellwood 1985) y, consecuentemente, sus distribuciones espaciales.

Las propias características de los datos que se estén utilizando pueden suponer otra limitación a la correcta determinación del radio. El estudio de los cúmulos es muy dependiente de la calidad de los datos usados. Si consideramos los grandes catálogos astrométricos actuales, vemos que cada uno de ellos presenta características peculiares que ya pueden sesgar el análisis de pertenencia posterior. Algunos efectos de selección pueden venir, por ejemplo, del rango espectral de las observaciones o de su magnitud límite. Sharma et al. (2006), Bukowiecki et al. (2011) y Tadross (2012) encuentran que los radios de los cúmulos utilizando datos de 2MASS (Jarrett et al., 2000) son normalmente mayores

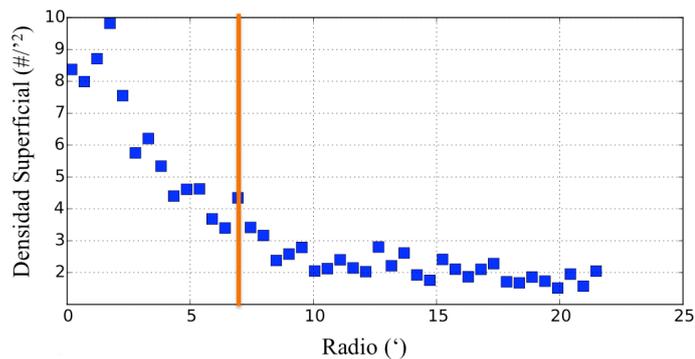


Figura 5.2: Perfil de densidad radial del cúmulo NGC 2099 usando datos de UCAC4. La línea anaranjada muestra el radio publicado en el catálogo DAML02.

que los encontrados a partir de catálogos estelares en el visible (UCAC4).

De la misma manera se pueden encontrar variaciones en los datos de un mismo cúmulo provenientes de estudios realizados por diferentes autores. Un ejemplo de este problema se discute precisamente en Sánchez, Vicente, & Alfaro (2010), donde se exponen siete valores diferentes para el radio del cúmulo NGC 2323, en función de la bibliografía utilizada. Como consecuencia, aunque el cúmulo tenga un radio físico determinado, el valor del radio empírico puede ser diferente en función de los datos que se estén usando para su estudio. Por tanto, su determinación es específica y debe ser optimizada para el conjunto de datos que se quiera utilizar. Su extrapolación a otros trabajos puede dar lugar a resultados no realistas del problema a abordar, además de tener una clara influencia en los análisis de pertenencia, problema directamente relacionado con esta tesis.

5.2.2 Problemas encontrados durante la determinación de los radios de los cúmulos.

En esta tesis hemos realizado una revisión del radio de los cúmulos del catálogo DAML02, tomando como base los radios publicados en dicho trabajo. Esta revisión es necesaria dado que los valores listados del radio para cada cúmulo proceden de un compendio de diferentes autores. Es importante enfatizar que su determinación no es homogénea y es dependiente del estudio realizado en la determinación de los mismos. Como consecuencia, al calcular los PDR de estos cúmulos, se observa que no todos los radios acotan la región en la que se encuentran las sobre-densidades relativas a cada uno de ellos. En la figura 5.2 se muestra a modo de ejemplo el PDR para el cúmulo NGC 2099 y, superpuesta, una línea vertical anaranjada indicando el radio listado en DAML02. Como se puede observar, éste sub-estima el radio del cúmulo para los datos del cartografiado UCAC4. Otro ejemplo, pero en este caso de sobre-estimación, se muestra en la figura 5.3 para el cúmulo IC 2944.

Por otro lado, la falta de contraste en los PDR estimados para cúmulos reales, dificulta la precisa determinación del radio del mismo. Con esta afirmación no se pretende insinuar la no existencia de cúmulo en la región del espacio analizada, sino que con los datos utilizados a veces éste no se observa.

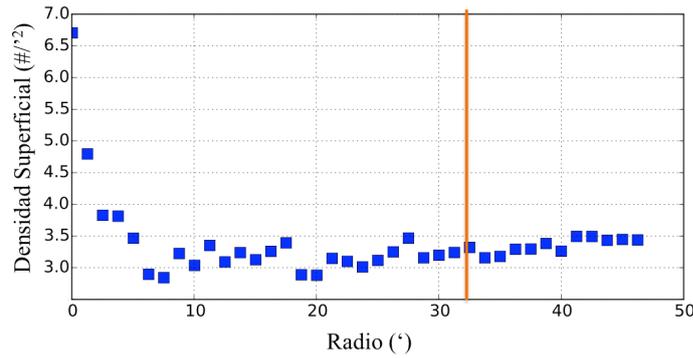


Figura 5.3: Perfil de densidad radial del cúmulo IC 2944 usando datos de UCAC4. La línea anaranjada muestra el radio publicado en el catálogo DAML02.

Un ejemplo de este caso sería el cúmulo NGC 2319 para el que, en la figura 5.4, se muestran su PDR y su distribución en el espacio de posiciones. Esta ausencia de sobre-densidad imposibilita la correcta determinación del radio físico asociado a este cúmulo. De forma análoga, se encuentra un número no despreciable de cúmulos donde, en vez de observar una sobre-densidad en la región del cúmulo, se observa un 'pozo' de densidad. Un ejemplo de este caso se ilustra en la figura 5.5 para el cúmulo IC 5146, donde claramente es visible una depresión de la densidad en la región central del mismo.

Otro factor importante en la correcta estimación del radio es la precisión en la posición del centroide del cúmulo. Su determinación puede ser complicada para aquellos cúmulos con pocas estrellas o con sub-estructuras. Sin embargo, el utilizar un centroide no correspondiente con su centro produce una estimación mayor del radio, introduciendo un mayor número de estrellas de campo en el análisis. En la figura 5.6 se muestra la importancia de este factor para la correcta estimación del PDR, en el cúmulo IC 361. Como se puede observar en la figura, la imprecisa determinación del centroide produce un ensanchamiento del PDR y como consecuencia, el radio estimado es mayor que el obtenido (gráfico superior) si el centroide estuviera bien determinado (gráfico intermedio). Para aquellos cúmulos en los que el máximo de la sobre-densidad no se correspondía con el centro del cúmulo catalogado, se realizó una inspección visual usando tanto los datos de UCAC4, como la herramienta SKY VIEW² (banda g de Sloan). De entre un total de 30 cúmulos con PDRs que podrían indicar un centroide mal estimado, 10 nuevos centroides han sido determinados, los cuales se encuentran detallados en la tabla 5.1. Para los 20 cúmulos restantes se han mantenido los radios publicados en DAML02.

La existencia de otros cúmulos cercanos en proyección ha sido otro problema encontrado. No se han observado cúmulos lo suficientemente cercanos para que se solapen sus PDRs. Sin embargo, la presencia de otros cúmulos cercanos en proyección, sí puede afectar a la estimación de la densidad del campo a partir de la cual se estima el radio del cúmulo (subsección 5.2.3). En el gráfico superior de la figura 5.7 se muestra el PDR del cúmulo Trumpler 16 en la que se puede observar un cúmulo 'contaminante' que se corresponde con el cúmulo Trumpler 14 (su posición se indica con una flecha roja).

²<http://skyview.gsfc.nasa.gov/current/cgi/titlepage.pl>

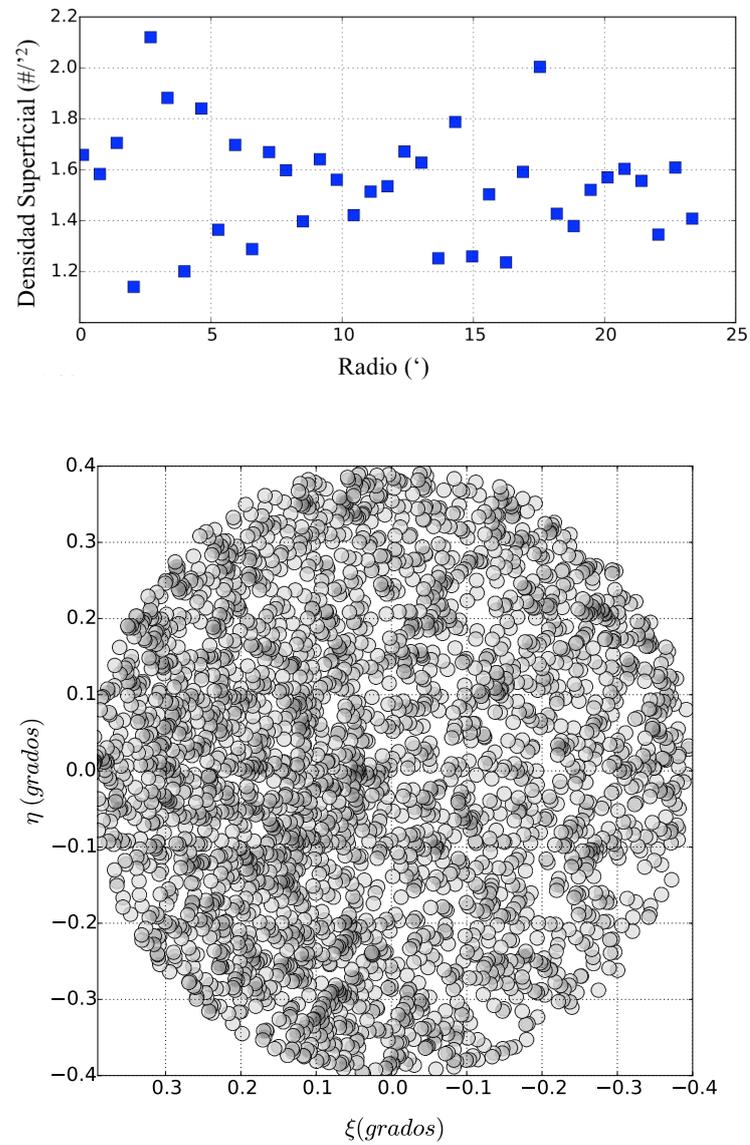


Figura 5.4: Perfil de densidad radial del cúmulo NGC 2319 y su distribución en el espacio de posiciones. Ambos reflejan la no existencia de sobre-densidad asociada a los datos de este cúmulo.

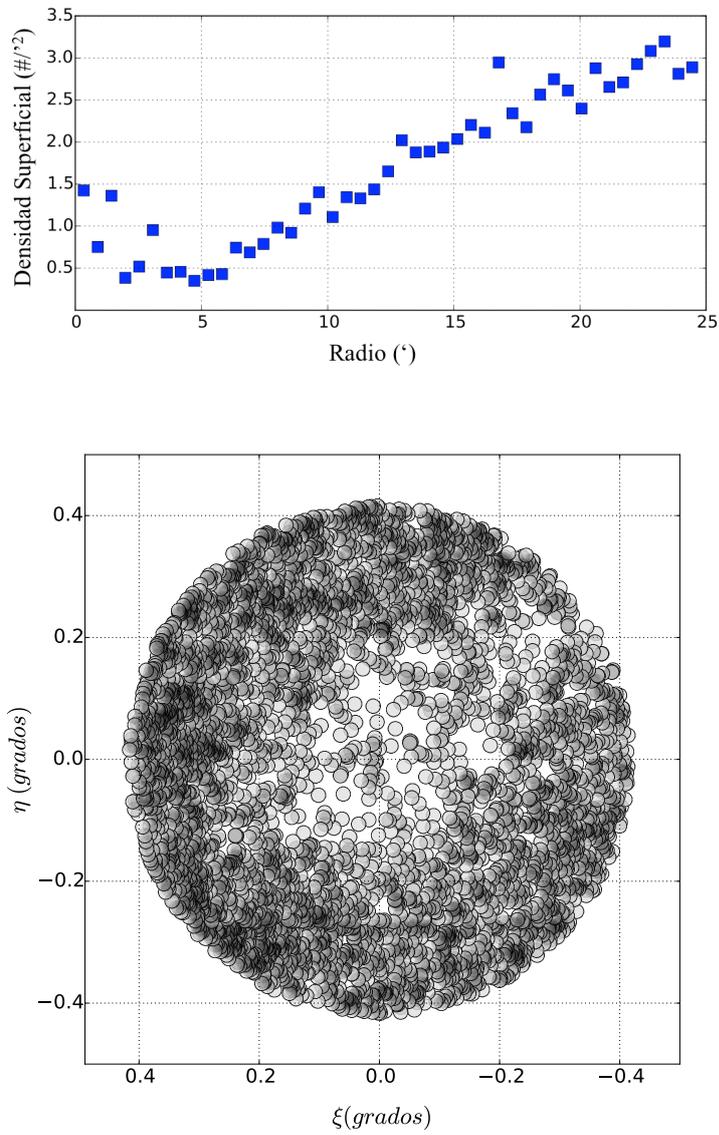


Figura 5.5: Ejemplo de depresión en la densidad en la región central del cúmulo IC 5146 usando datos de UCAC4.

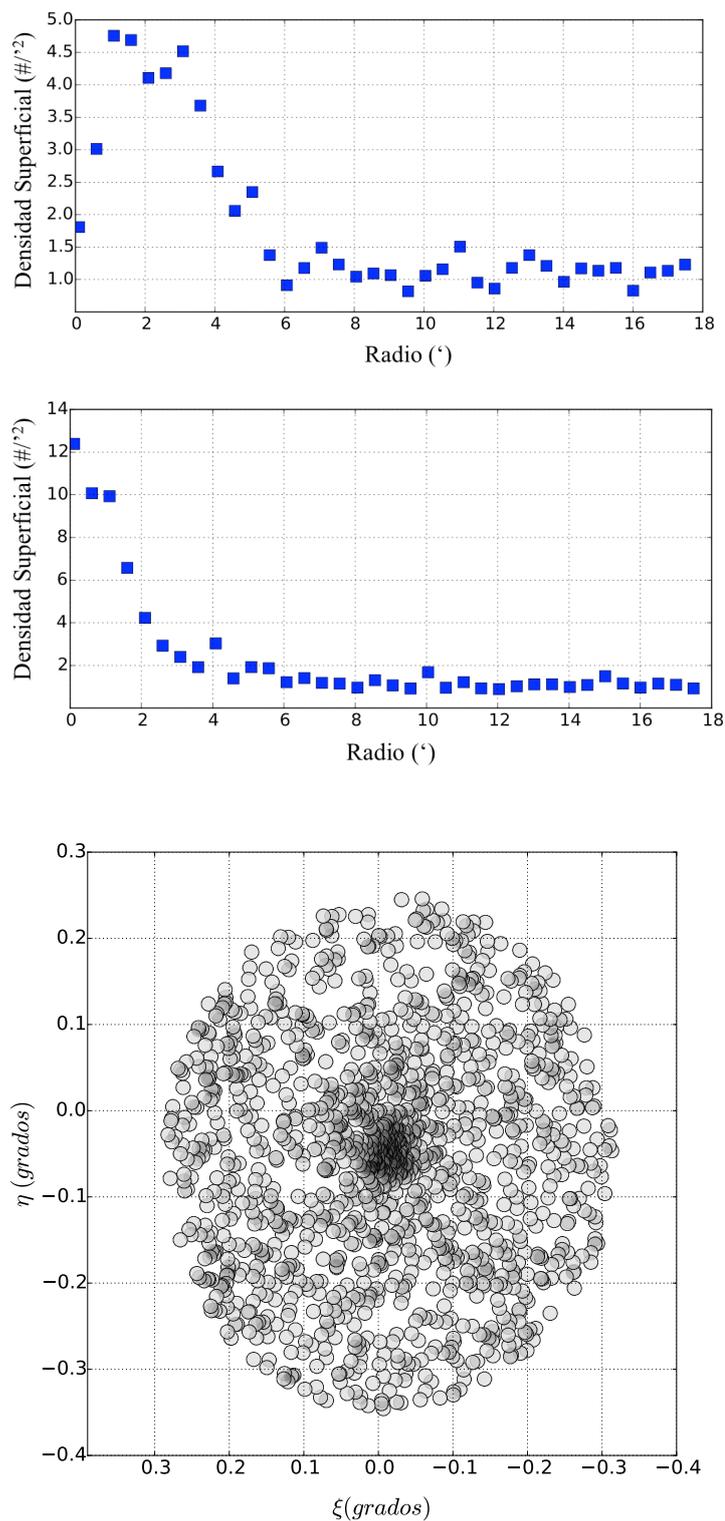


Figura 5.6: Perfiles de densidad radial y distribución en el espacio de posiciones del cúmulo IC 361. Se observa que el centroide dado en DAML02 no se corresponde con la posición central del cúmulo. Como consecuencia, el PDR asociado (gráfico superior) es más ancho que el estimado a partir de la sobre-densidad central del cúmulo (gráfico intermedio), dando lugar a un radio de mayor valor.

Tabla 5.1: Ascensiones rectas y declinaciones de los nuevos (subíndices 1) y antiguos (subíndices 0) centroides determinados para 10 cúmulos.

Nombre	α_1 (h:m:s)	δ_1 (g:m:s)	α_0 (h:m:s)	δ_0 (g:m:s)
Berkeley 28	06 52 07	02 54 47	06 52 12	02 56 00
Berkeley 39	07 46 42	-04 40 06	07 42 46	-04 36 00
Berkeley 43	19 15 32	11 16 20	19 15 36	11 13 00
Berkeley 45	19 19 05	15 42 47	19 19 12	15 43 00
Berkeley 50	20 10 01	34 57 58	20 10 24	34 58 00
BH 208	16 59 32	-37 07 20	16 59 36	-37 05 00
IC 1311	20 10 46	41 10 27	20 10 18	41 13 00
IC 1369	21 12 06	47 46 04	21 12 06	47 44 00
IC 361	04 18 54	58 15 00	04 19 00	58 18 00
Ruprecht 164	11 30 25	-60 45 10	11 30 51	-60 44 00

La ausencia de sobre-densidades, la absorción variable, las desviaciones del centroide y la contaminación por cúmulos cercanos imposibilitan la automatización de la estimación del radio de los cúmulos sin supervisión visual previa. En esta tesis hemos aplicado diferentes técnicas intentando eliminar, en la medida de lo posible, la toma de decisiones personalizadas, pero que en muchos casos ha sido imposible evitar. Tanto las técnicas que comúnmente han sido más utilizadas como el procedimiento en la determinación de los radios realizado en esta tesis, se describen a continuación.

5.2.3 Métodos en la determinación de los radios.

La razón de la ausencia de una herramienta global para la determinación de los radios estriba en la variedad, tanto de las diferentes geometrías de los cúmulos abiertos como en las características de los datos que se estén usando. Sin embargo, su correcta determinación es esencial para todos aquellos estudios que necesiten de los miembros de estos sistemas estelares. Por tanto, la pregunta sería cómo calcular el radio de los cúmulos de la manera más automática y objetiva posible.

Dos técnicas han sido y son actualmente utilizadas para calcular el radio de los cúmulos. Son los conocidos como los modelos de King de 2 y 3 parámetros (King, 1962). Por una parte, en 1962 King publicó un trabajo en el que proponía una ley empírica para la determinación de los radios de los cúmulos globulares (cúmulos con un gran número de estrellas y simetría esférica bien definida), con la que pretendía ajustar tanto las partes externas como las partes internas de los mismos. Esta ley empírica dependía de tres parámetros: la densidad superficial central del cúmulo, el radio del núcleo y el radio de marea (el considerado radio del cúmulo). Su expresión matemática es:

$$f_{3p}(r) = k \left(\frac{1}{\left(1 + \left(\frac{r}{r_c}\right)^2\right)^{\frac{1}{2}}} - \frac{1}{\left(1 + \left(\frac{r_t}{r_c}\right)^2\right)^{\frac{1}{2}}} \right)^2 \quad (5.3)$$

donde r es la distancia al centro del cúmulo, k es una constante, r_c es el radio del núcleo y r_t es el radio de marea.

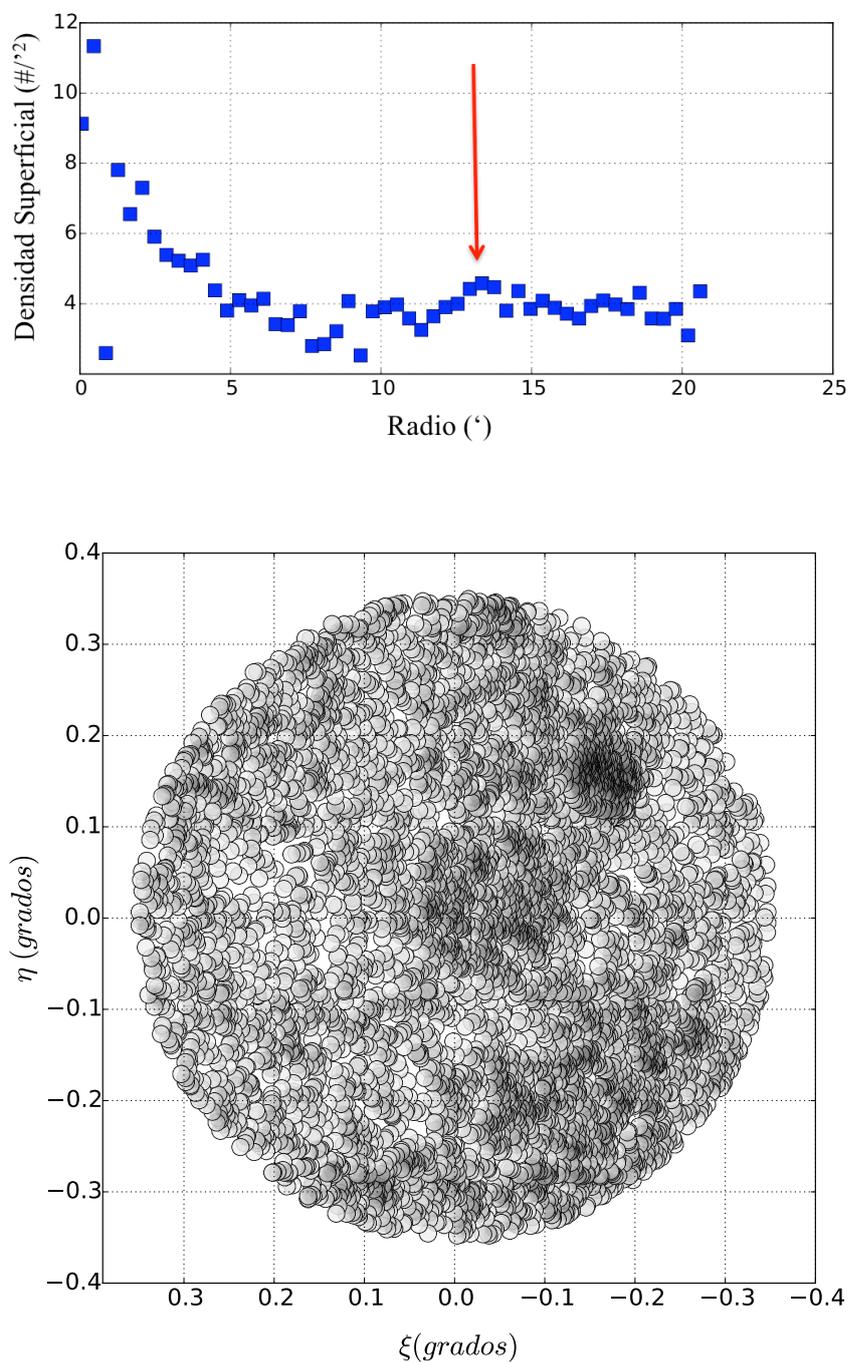


Figura 5.7: Ejemplo de cómo la existencia de cúmulos cercanos puede afectar en el cálculo del radio. El cúmulo central se corresponde con Trumpler 16, mientras que el cúmulo 'contaminante' es Trumpler 14.

Según otro trabajo de King (1966), esta ley empírica puede no ajustar a las partes externas de los cúmulos (como por ejemplo para el cúmulo NGC 1758 estudiado en el trabajo de Galadi-Enriquez, Jordi, & Trullols (1998)), que son las que pueden verse fuertemente afectadas por fuerzas de marea, interacciones gravitatorias con otros sistemas estelares o el paso de los cúmulos por nubes moleculares o brazos espirales. Sin embargo, las partes internas de un cúmulo podrían permanecer ajenas a estos fenómenos pudiendo realizar un ajuste de acuerdo con la ecuación (modelo de King de 2 parámetros):

$$f_{2p}(r) = f_{bkg} + \frac{f_0}{1 + \left(\frac{r}{r_c}\right)^2} \quad (5.4)$$

donde r es la distancia al centro del cúmulo, f_{bkg} es la densidad de campo y r_c es el radio del núcleo del cúmulo.

En ambos trabajos, pero explícitamente en King (1962), se propone su aplicación a cúmulos abiertos aplicando su ley empírica a tres cúmulos abiertos ricos en estrellas y con un alto grado de simetría esférica: M 67, NGC 188 y NGC 7789. Dado que la gran mayoría de los cúmulos no verifican estas características, diferentes autores han aplicado variaciones de estos dos modelos. Así, Bukowiecki et al. (2011) utiliza el modelo de King de 2 parámetros para determinar el radio del núcleo y, posteriormente, introduce la siguiente ecuación para determinar el radio del cúmulo:

$$r_t = r_c \sqrt{\left(\frac{f_0}{3\sigma_{bkg}} - 1\right)} \quad (5.5)$$

donde r_c es el radio del núcleo determinado por el ajuste de King de dos parámetros, f_0 es la densidad central y σ_{bkg} es la dispersión del campo.

Otra aproximación al mismo problema proviene de Tadross (2012) quien de nuevo utiliza el modelo de King de 2 parámetros para determinar el radio del núcleo, pero si la masa del cúmulo total es conocida (M_c), el radio de marea se podría estimar haciendo uso de la ecuación (Jeffries, Thurston, & Hambly, 2001):

$$r_t = 1.46(M_c)^{1/3} \quad (5.6)$$

Trabajos como Maia, Corradi, & Santos (2010) y Tadross (2005), aplican el modelo de King de 2 parámetros para determinar el radio del núcleo del cúmulo, mientras que el radio del cúmulo es calculado por inspección visual del PDR del cúmulo en cuestión. Otros por el contrario (Subramaniam, Carraro, & Janes, 2010), ajustan los PDR mediante inspección visual, estimando a partir de ellos el radio del núcleo y calculando a partir de la ecuación 5.4 el radio del cúmulo.

Ambas metodologías fueron aplicadas a los cúmulos presentes en DAML02. Para el modelo de King de 2 parámetros, el radio del cúmulo fue calculado utilizando la ecuación (5.5) propuesta en el trabajo de Bukowiecki et al. (2011). Sin embargo, en muchos casos los radios obtenidos no se ajustaban bien a los valores esperados. Hay que tener presente que el ajuste del PDR usando un modelo de King se realiza mediante ajustes de mínimos cuadrados o χ^2 . Este ajuste puede dar lugar a valores de r_t muy

grandes que, aunque sean una buena solución matemática, no son representativos del radio real del cúmulo. Tal y como señala Kaluzny & Udalski (1992), la función observada $f_{3p}(r)$ puede dar lugar a una variedad de soluciones.

Por estos motivos y con el objetivo de minimizar en lo posible la toma de decisiones, hemos diseñado un procedimiento en el que los PDRs son ajustados por polinomios de bajo grado. A continuación detallamos el proceso que nos lleva a la obtención de los radios finales que, posteriormente, serán utilizados como radios muestrales para la descarga de los datos de UCAC4.

5.2.4 Procedimiento para la obtención del radio del cúmulo a partir de un ajuste polinómico al PDR.

Teniendo en cuenta los problemas asociados en la estimación de los radios por ajustes de funciones de King de 2 y 3 parámetros, hemos diseñado un procedimiento basado en ajustes por mínimos cuadrados de los PDRs por polinomios de bajo grado. Su objetivo fundamental consiste en la determinación del radio de los cúmulos minimizando, en lo posible, la toma de decisiones en cuanto al valor de dicho radio. Para su aplicación, definimos el radio del cúmulo como aquella distancia a la que su densidad radial se iguala a la densidad de campo (ρ_{bkg}) más 2 veces su dispersión (σ_{bkg}):

$$\rho_{cluster} = \rho_{bkg} + 2\sigma_{bkg} \quad (5.7)$$

Con esta definición estamos teniendo en cuenta variaciones estadísticas en la densidad de las estrellas de campo que pueden afectar a la determinación del radio del cúmulo. En el caso específico en el que la distribución de las estrellas de campo sea homogénea, la diferencia entre nuestro radio y aquel que hubiera sido obtenido considerando igualdad en las densidades de cúmulo y campo, será mínima.

Como base de nuestro procedimiento se tomó el radio publicado en el catálogo DAML02, ampliándolo 15 ma para asegurarnos de estar cogiendo toda la región del cúmulo y, además, poder realizar el ajuste a la densidad del campo. Del radio total, los 6 ma más alejados del centroide del cúmulo fueron utilizados para calcular la media y la dispersión del campo circundante. Una vez calculados estos parámetros y, eliminando estos 6 ma, hicimos el ajuste de los PDRs por polinomios de bajo grado (hasta un máximo de tres) de la muestra resultante. El grado del polinomio utilizado se obtuvo haciendo uso del *Criterio de Información Akaike* (Akaike, 1974), que compensa la bondad del ajuste utilizado con la complejidad del modelo, con el objetivo de minimizar fluctuaciones en el ajuste realizado. Un ejemplo de ajuste por este procedimiento se puede observar en el gráfico izquierdo de la figura 5.8.

Esta técnica no siempre obtiene buenos resultados, lo cual no nos permite eliminar toda la subjetividad relativa a la estimación de los radios. La elección de un polinomio de bajo grado en los ajustes a los PDRs, tiene como objetivo eliminar posibles oscilaciones en el ajuste que dan lugar a valores mayores de los radios, aunque no siempre fue posible (ver gráfico derecho de la figura 5.8). Por tanto, este procedimiento requiere de una inspección visual y de la toma de decisiones en cuanto al valor

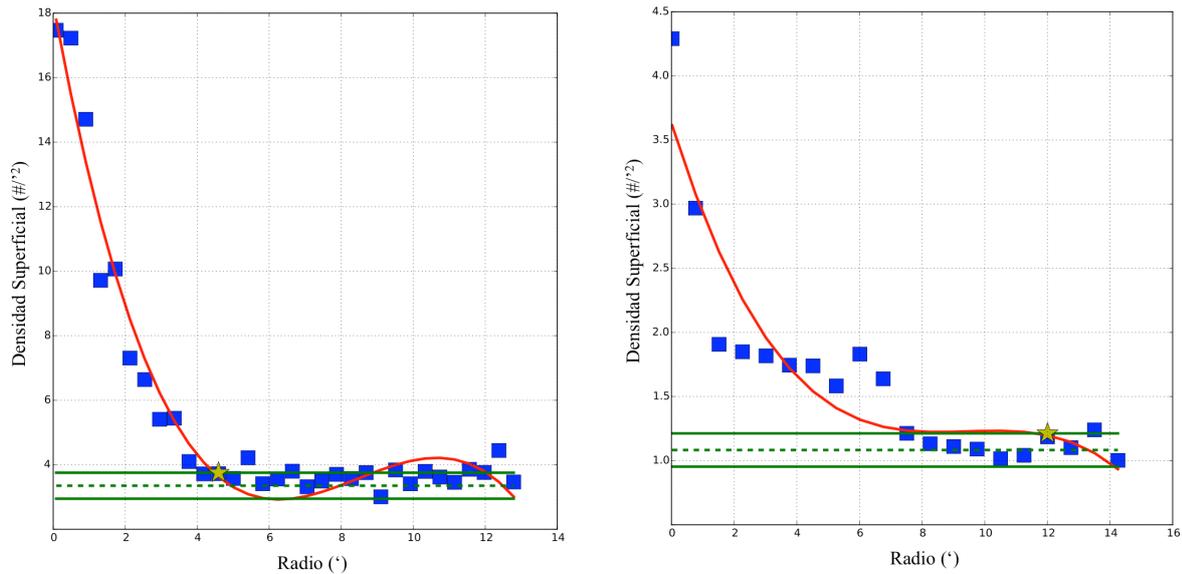


Figura 5.8: Ejemplos de buen y mal ajuste (en rojo) por polinomios a los PDRs. En el gráfico izquierdo se representa el ajuste realizado para el cúmulo Trumpler 14 y en el derecho para el cúmulo FSR 0941. La estrella amarilla indica el radio del cúmulo estimado por el ajuste, la línea verde punteada horizontal el valor medio de la densidad de campo y las líneas verdes continuas horizontales el valor de la densidad a 2σ 's.

final del radio del cúmulo. Para todos los cúmulos, pero fundamentalmente dirigidos a aquellos donde no se obtuvo un buen ajuste, se hicieron dos PDRs diferentes al ya realizado utilizando dos muestreos diferentes: uno eligiendo el número de intervalos como la raíz cuadrada del número de objetos dividido por 2 y el otro dividido por 4.

Para aquellos cúmulos en los que tanto el ajuste al PDR como la visualización del mismo, utilizando diferentes muestreos no proporcionaban una solución clara, se decidió mantener el radio catalogado en DAML02 pero extendiéndolo 0.5' como medida profiláctica.

En los gráficos de la figura 5.9 se muestra en azul el radio final que se usará para realizar los análisis de pertenencia frente al radio listado en DAML02. En magenta se representan los radios obtenidos por el ajuste a los PDR por un polinomio y haciendo uso del modelo de King de 3 parámetros (gráficos izquierdo y derecho, respectivamente). Aunque ninguna de las dos técnicas presenta un elevado grado de acuerdo con el radio final elegido, sí se encuentra una mayor concordancia con los resultados obtenidos a partir del ajuste por polinomios de los PDRs de los cúmulos. Se observa que el modelo de King tiende a determinar radios mucho mayores que los listados en el catálogo DAML02 o los encontrados al realizar el ajuste del PDR por polinomios y, por tanto, mucho mayores a los radios finales utilizados.

En resumen, los radios muestrales proceden, en primer lugar, del ajuste por polinomios al PDR de los cúmulos (para un 34% de los cúmulos). Si el ajuste realizado no es bueno, pero existe una sobre-densidad clara con respecto al campo (y no coincide con la catalogada en DAML02), se adoptó un valor del radio del cúmulo, haciendo uso de los otros dos PDRs auxiliares construidos para tal fin (para un 26% de los cúmulos). En el resto de los cúmulos (un 40%) se mantuvo el radio catalogado en

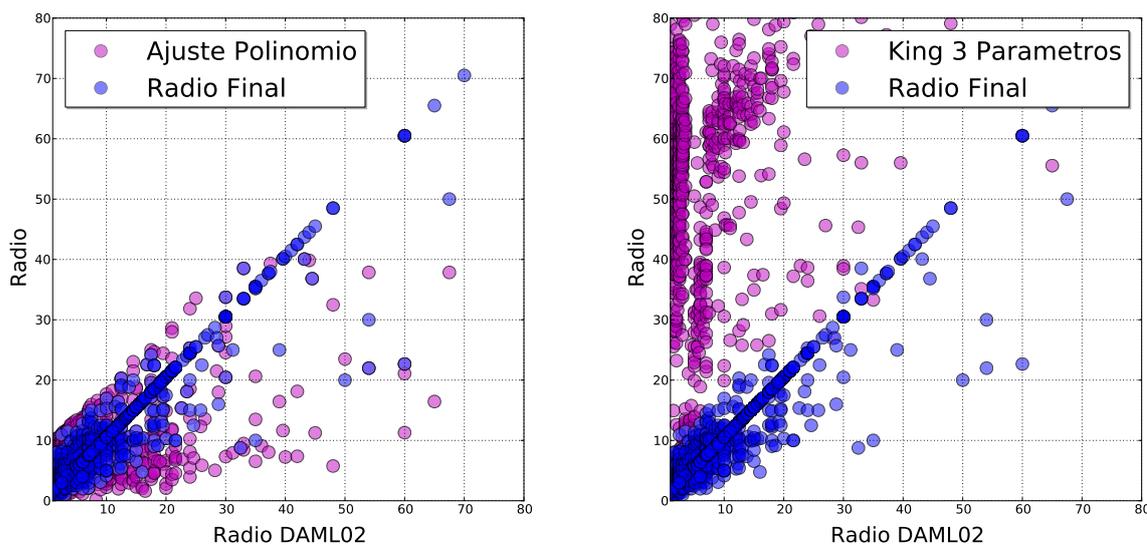


Figura 5.9: En la figura se muestran en azul los radios finales utilizados en función de los radios listados en DAML02. En magenta, en el gráfico izquierdo se representa los radios obtenidos por el ajuste polinómico y, en el gráfico de la derecha, el obtenido tras aplicar un modelo de King de 3 parámetros.

DAML02 extendiéndolo 0.5' por no poder hacer una determinación precisa del mismo. En la figura 5.10 se muestra la distribución de los radios finales en ma, teniendo un valor de la mediana en torno a 2 ma y con un 85% de los cúmulos con un valor inferior a los 10 ma. Los radios finales elegidos serán los radios utilizados en la selección de la muestra a partir de la base de datos UCAC4 y a la que le serán aplicadas las diferentes metodologías descritas en esta tesis, para realizar el análisis de pertenencia. Los resultados obtenidos se detallan a continuación.

5.3 Resultados obtenidos por las metodologías descritas en esta tesis.

Una vez determinados los radios de los cúmulos listados en el catálogo DAML02, se aplicaron las metodologías descritas en el capítulo 2. Para cada uno de estos cúmulos y, a partir de los radios ya calculados, se procedió a la descarga automática de los datos del cartografiado UCAC4 haciendo uso de la herramienta Vizier (Ochsenbein, Bauer, & Marcout, 2000). Se utilizaron los centroides en posiciones listados en DAML02 salvo aquellos cúmulos a los que hubiese sido necesario calcular de nuevo el centroide (ver tabla 5.1 en la subsección 5.2.2), siendo este último el que se utilizó.

Posteriormente, y antes de proceder a los análisis de pertenencia, se eliminaron aquellas estrellas que cumplieran algunas de estas condiciones:

- Estrellas sin datos de movimientos propios.
- Estrellas con errores superiores a 12 mas/año. Según Zacharias et al. (2013) aquellos objetos en

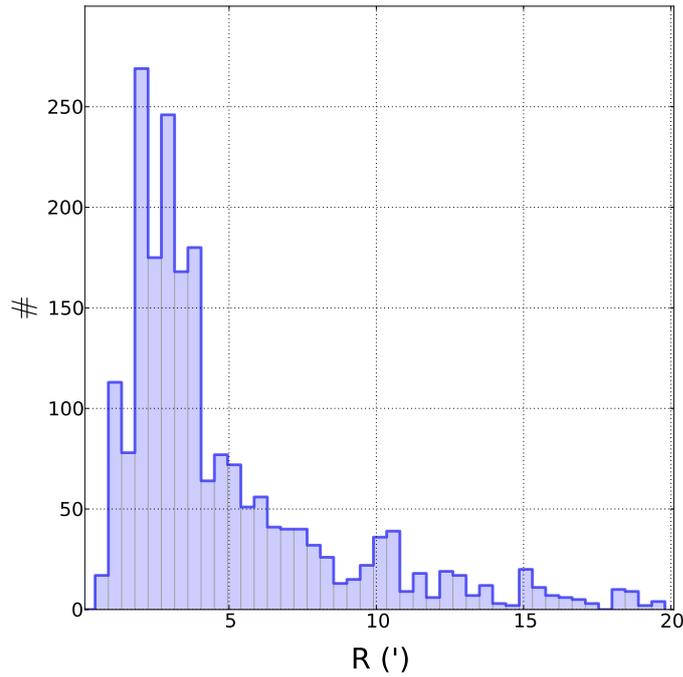


Figura 5.10: Distribución de los radios muestrales a partir de los cuales se seleccionará la muestra a analizar de la base de datos UCAC4.

UCAC4 con errores superiores a este valor pueden no corresponder a estrellas reales.

- Estrellas con valores de movimientos propios en el rango $|\text{pm}| > 100 \text{ mas/año}$. Estas estrellas son posiblemente objetos del halo con baja probabilidad de ser miembros del cúmulo (Tadross, 2005). Las estrellas de alta velocidad afectan a la correcta determinación de los *outliers* presentes en la muestra a analizar.

En la figura 5.11 se muestra el porcentaje entre el número de estrellas que se utilizaron para realizar los análisis de pertenencia (N_{inif}), (tras aplicar los criterios anteriores), respecto al número total de estrellas dentro de la región del cúmulo (N_{ini}), para un total de 2137 cúmulos. Aproximadamente el 93% de las muestras tienen más de un 70% de objetos con datos de movimientos propios, con una centena de cúmulos con la información completa.

Por otra parte, cúmulos con un N_{inif} inferior a 20 (261 cúmulos) y superior a 10^4 (30 cúmulos) fueron descartados, dando lugar a una muestra final formada por 1876 cúmulos. Para estos cúmulos, se realizaron cuatro análisis de pertenencia diferentes haciendo uso de los datos en posiciones y en movimientos propios. Las metodologías aplicadas fueron MT2, MT4 y la nueva metodología geométrica para $N = 2$ y 4 variables (MTG2 y MTG4, respectivamente). En este estudio, y al igual que en las simulaciones, los *outliers* se determinaron en el espacio cinemático haciendo uso de la herramienta OUTKER. Para evitar efectos geométricos, MTG4 y MT4 utilizaron las distancias angulares (ξ, η) definidas en la subsección 5.2.2.

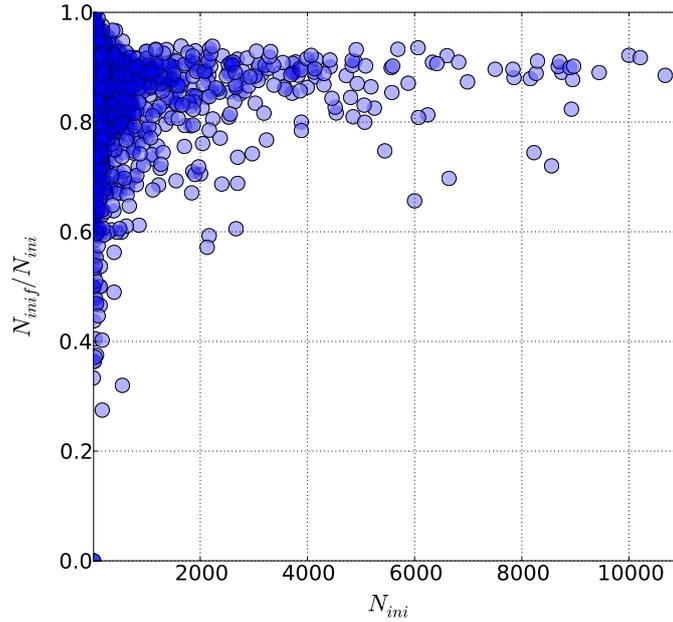


Figura 5.11: Fracción de estrellas finales respecto al número inicial de estrellas (N_{inif}/N_{ini}) frente al número inicial de estrellas (N_{ini}) en las regiones de los 2137 cúmulos catalogados en DAML02.

Dado el elevado tiempo que llevaría el analizar los resultados obtenidos para cada cúmulo de manera individual, impusimos unos criterios iniciales de selección que debían cumplir los resultados obtenidos por cada una de las metodologías aplicadas. De esta forma pretendemos realizar una primera purga de resultados espurios que puedan emerger de su aplicación. Así, aquellos análisis de pertenencia que determinen un número de miembros de cúmulo inferior al 20%, no serán considerados en lo sucesivo. Esta decisión se basa en dos hipótesis. Cabe suponer que al estar utilizando el radio de los cúmulos para la descarga de los datos, un porcentaje superior del 20% de la muestra debería ser clasificado como miembro potencial. Como se comentará más adelante en esta sección, esta hipótesis podría no cumplirse para aquellos cúmulos cercanos con elevados valores de radio proyectado, donde la contaminación de campo podría ser tan elevada que el porcentaje correspondiente al cúmulo sea inferior al indicado. La segunda hipótesis hace referencia a nuestras simulaciones. Se observa que las muestras que tienen menos de un 20% de estrellas de cúmulo afectadas por sub-muestreos, no producen buenos resultados en los análisis de pertenencia realizados por las diferentes metodologías. De la misma manera, dado que las estrellas de cúmulo están más densamente distribuidas que las de campo en el espacio de movimientos propios, aquellos análisis de pertenencia cuyos resultados determinen lo contrario, serán considerados soluciones erróneas y eliminados del catálogo final.

De la muestra final de 1876 cúmulos se encontraron 982 (un 52%) con resultados válidos obtenidos por todas las metodologías. En la figura 5.12 se muestran los centroides y las dispersiones de estos cúmulos (gráficos superior e inferior, respectivamente) en el espacio de movimientos propios. En estos gráficos se representan los centroides obtenidos por las metodologías MT2 y MT4 en función de los centroides obtenidos respecto de la metodología MTG2. Como se puede observar, la determinación

de los centroides estimados por estos métodos presenta un elevado grado de acuerdo, especialmente aquellos obtenidos por MTG2 y MT2 (gráfico superior izquierdo). Se aprecia en el gráfico superior derecho la existencia de un cúmulo con una diferencia en la determinación del centroide realizada por la metodología MT4 muy diferente a la obtenida por MTG2. Este cúmulo es Blanco 1 y más adelante se explicarán con más detalles sus resultados y el por qué de esta diferencia.

Las dispersiones obtenidas en el plano de los movimientos propios por MT2 y MTG2 (gráfico inferior izquierdo de la figura 5.12) son de nuevo bastante similares, siendo superiores las obtenidas cuando las posiciones son incluidas en los análisis de pertenencia (gráfico inferior derecho). Este efecto se ilustra en la figura 5.13 donde se representan, en el lado izquierdo, las dispersiones obtenidas en movimientos propios por MTG4 (negro) y MT4 (rojo) y, en lado derecho, las obtenidas por MTG2 (azul) y MT2 (magenta) respecto a N_{inif} . Se puede observar un gradiente en los valores obtenidos. Las metodologías que usan únicamente movimientos propios (gráfico derecho) determinan cúmulos con menores dispersiones en este espacio. La inclusión de las posiciones revierte en la determinación de cúmulos con una mayor dispersión, fundamentalmente MTG4 (círculos negros). Cabe añadir que, para aquellos cúmulos con ausencia de sobre-densidad, las posiciones actuarían como variables no informativas, sesgando las probabilidades obtenidas, tal y como se observó con los datos simulados.

Se observa que a partir de un $N_{inif} > 1000$ estrellas (figura 5.13) y hasta aproximadamente $N_{inif} < 3000$ estrellas, los valores de las dispersiones en movimientos propios obtenidos por las diferentes metodologías (exceptuando MTG4) tienden a converger en torno a los 4-6 mas/año, precisamente el valor medio de los errores presentes en UCAC4. Para $N_{inif} > 3000$ estrellas la dispersión en movimientos propios obtenida por MTG2 y MT2 aumenta conforme crece N_{inif} hasta alcanzar un valor asintótico en torno a los 6 mas/año. Sin embargo, la dispersión obtenida por MT4 disminuye manteniéndose en torno a los 4 mas/año.

Para 12 de los 1876 cúmulos, todos los análisis determinan que la población de cúmulo es inferior al 20% del total de la muestra. Para todos ellos, N_{inif} toma valores muy bajos, no superando prácticamente las 50 estrellas a analizar. Estos cúmulos son: FSR 0334, Saurer 4, Czernik 11, Loden 27, Alessi 53, Ruprecht 3, Berkeley 76, Ruprecht 35, Majaess 79, FSR 1471, NGC 1663 y Pismis 5. Para 5 de estos 12 cúmulos ninguna de las cuatro metodologías aplicadas determinó cúmulo. Estos cúmulos son Czernik 11, Loden 27, Ruprecht 3, Berkeley 76 y Pismis 5. Un ejemplo de las distribuciones en los espacios de posiciones y de movimientos propios, en particular para el cúmulo abierto Ruprecht 3, se muestra en la figura 5.14. Como se puede observar, es imposible determinar los miembros del cúmulo con esta muestra de estrellas con la que, hasta la misma visualización del cúmulo se puede poner en duda.

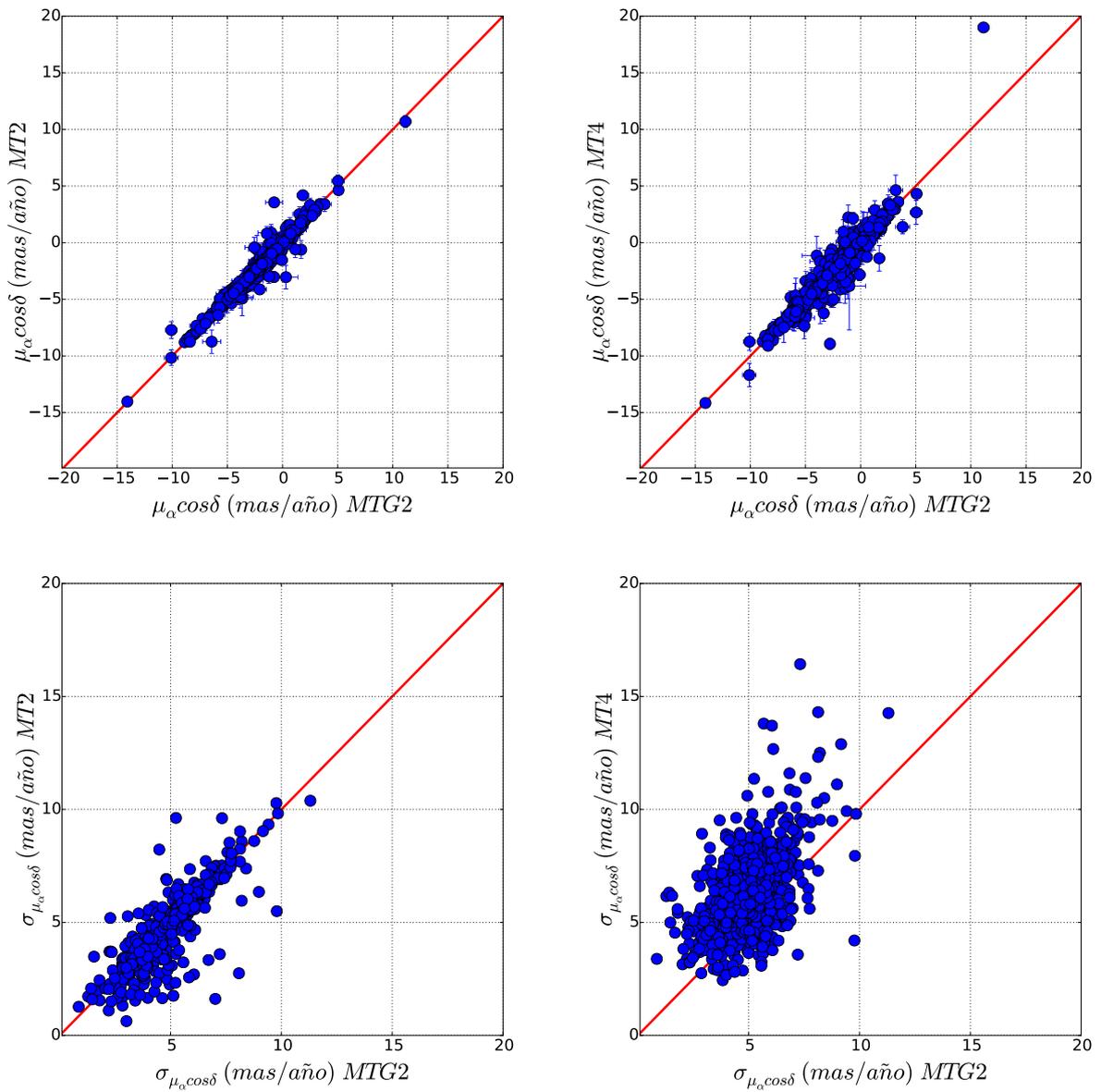


Figura 5.12: Comparación de los centroides y dispersiones, en el espacio de los movimientos propios, obtenidos por MT2 (izquierda) y MT4 (derecha) en función de los obtenidos por MTG2, para la muestra de 982 cúmulos con resultados válidos obtenidos por todos los métodos aplicados.

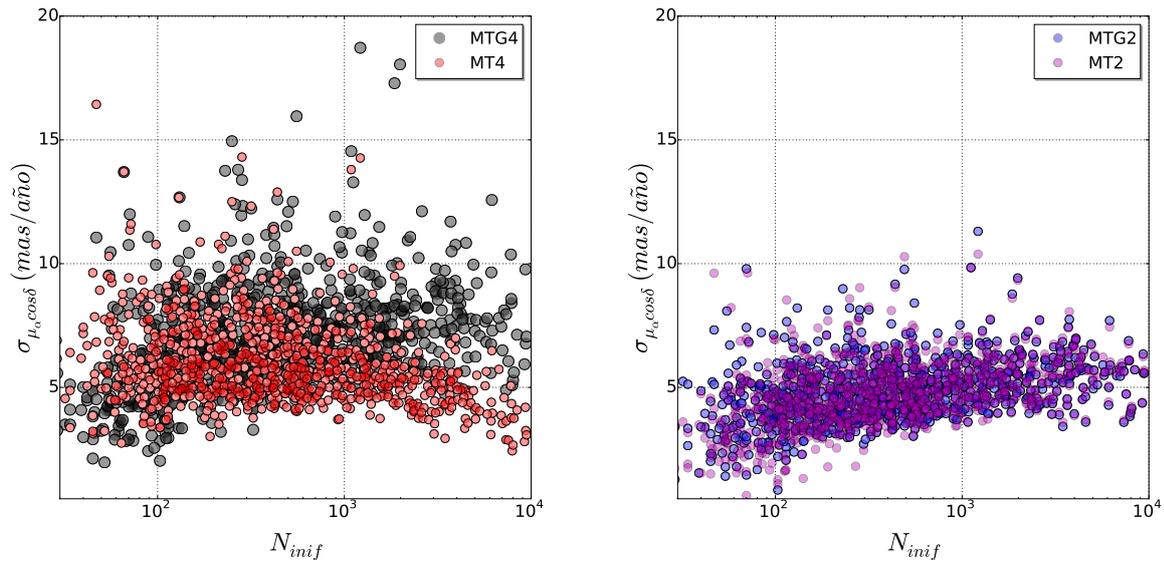


Figura 5.13: Dispersiones en el espacio de los movimientos propios de la muestra de 982 cúmulos en función del número de estrellas presentes en la muestra a analizar para cada cúmulo. En el gráfico izquierdo se muestran los resultados obtenidos por MTG4 y MT4 en negro y rojo, respectivamente. En el gráfico derecho se muestran los correspondientes a MTG2 y MT2 en azul y magenta. Se observa una tendencia a obtener mayores valores de las dispersiones conforme N_{inif} aumenta, exceptuando MT4 que para $N_{inif} > 3000$ estrellas, determina menores dispersiones que el resto de metodologías.

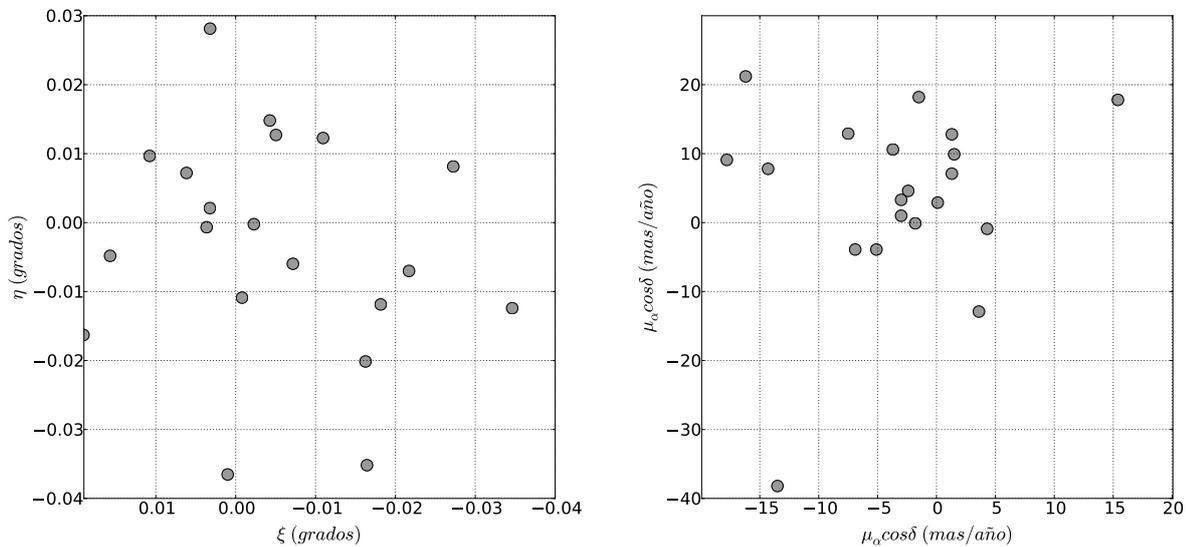


Figura 5.14: Distribuciones en los espacios de posiciones y de movimientos propios del cúmulo Ruprecht 3.

De manera individual, de los 1876 cúmulos analizados, MTG2 obtuvo resultados válidos para 1724 cúmulos, MT2 para 1585, MTG4 para 1058 y MT4 para 1692, después de eliminar aquellos con resultados considerados espurios. Por tanto, MTG2 no obtiene resultados en 152 casos, MTG4 en 818 y MT2 en 291. Para todos estos cúmulos no se cumple el criterio impuesto sobre el porcentaje del

20% de estrellas en la muestra clasificadas como estrellas de cúmulo. Esto es debido bien precisamente porque las metodologías convergen pero determinan un porcentaje inferior como miembros del cúmulo, bien porque esta convergencia no es alcanzada o bien porque toda la muestra ha sido clasificada en una única población. A efectos prácticos, para estas dos últimas causas, la metodología no es capaz de realizar una clasificación entre los dos grupos de estrellas y, por tanto, los resultados obtenidos para estos cúmulos no serán publicados. Por su parte, MT4 no obtiene resultados en 184 cúmulos. De estos, en 59 clasificó menos del 20% de la muestra como cúmulo y en 125 no cumplió el criterio de las dispersiones (las estrellas de cúmulo están más densamente distribuidas que las de campo en el espacio de los movimientos propios). La gran mayoría de las muestras, para las que las metodologías no han obtenido resultados válidos, se corresponden con regiones con $N_{inif} < 100$ estrellas.

La diferencia en el número de resultados válidos obtenidos entre MTG2 y MT2, ya se observó en las simulaciones (*Sim2*) en donde, de forma general pero particularmente para el porcentaje del 20% de estrellas de cúmulo simuladas, el número de miembros determinados por MTG2 era superior al obtenido por MT2 (gráfico izquierdo de la figura 5.15). Este comportamiento se ha encontrado en la aplicación a datos reales. Tal y como ilustra el gráfico derecho de la figura 5.15, para valores bajos de N_{inif} (inferiores a unas 250 estrellas), la metodología MTG2 tiende a clasificar un porcentaje mayor de estrellas como miembros potenciales del cúmulo que MT2, resultando en un mayor número de cúmulos que cumplen el criterio anteriormente indicado. Es precisamente para estos valores de N_{inif} donde las muestras están más afectadas por problemas de sub-muestreo (ver gráfica 5.11) causa que puede producir que la muestra a analizar tenga un porcentaje bajo de estrellas de cúmulo.

Consideramos que la explicación de estos resultados radica en la diferente naturaleza de los dos análisis de pertenencia. Aunque ambos utilicen el mismo conjunto de variables, los ajustes a partir de los cuales determinan las probabilidades de pertenencia, son distintos. Mientras que la metodología paramétrica hace un ajuste a la distribución de movimientos propios, la metodología geométrica lo hace a la distribución de distancias obtenidas a partir de estas variables (en este análisis de pertenencia). Desde un punto de vista general, el ajuste a la distribución de movimientos propios, referida a la población de campo, suele tener una mayor coincidencia a la referida al cúmulo, que el ajuste a la distribución de distancias (gráficos superiores de la figura 5.16). El resultado es una mayor determinación de miembros por MTG2 que por MT2 (gráficos inferiores donde los miembros determinados se representan en magenta) y, por tanto, MTG2 tenderá a cumplir en un mayor número de análisis la condición del 20% impuesta en la eliminación de resultados espurios.

De los 1876 cúmulos, MT4 obtiene resultados válidos para 1692. Si nos vamos a los resultados obtenidos en las simulaciones *Sim3* (simulaciones en las que la distribución de las posiciones de cúmulo son aleatorias), podemos dar una explicación a los resultados. En el gráfico izquierdo de la figura 5.17 se representa el porcentaje de estrellas clasificadas como miembros de cúmulo por MT4 en función del porcentaje simulado. Se observa que, para todos los porcentajes pero especialmente para los dos más bajos, MT4 determina un porcentaje mayor de miembros que el simulado. Al contrario que las otras metodologías, la no-paramétrica no impone la condición de que el valor de la dispersión de las estrellas de cúmulo tiene que ser menor a la de campo en el espacio de movimientos propios, sino que en ambos espacios tienen que estar más densamente distribuidas. Como consecuencia, 125 cúmulos no

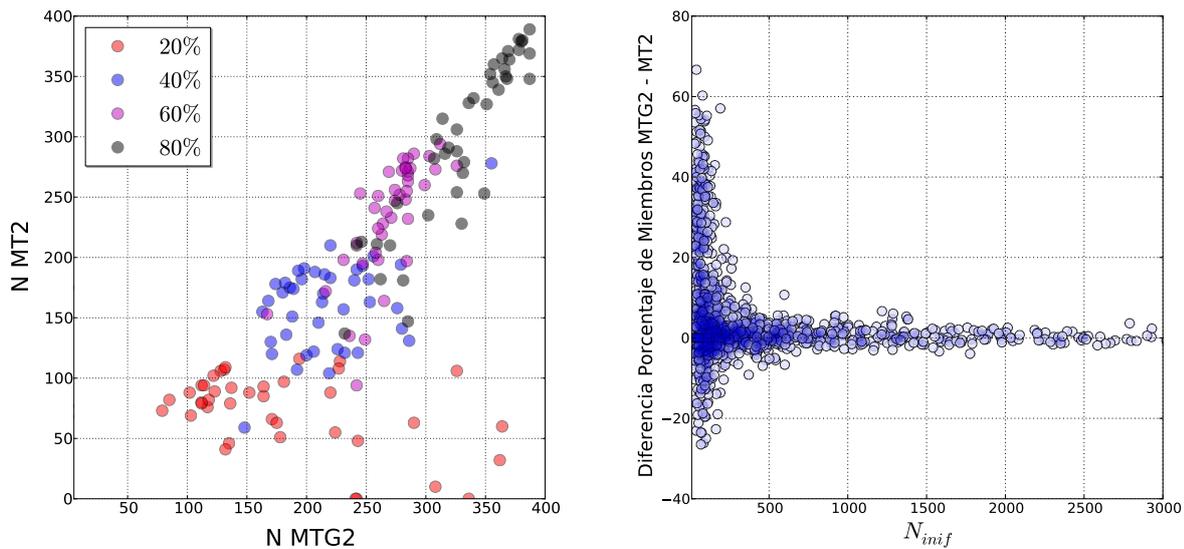


Figura 5.15: El gráfico derecho muestra que, en las simulaciones *Sim2*, el número de estrellas clasificadas como miembros de cúmulo por MT2 suele ser inferior al que clasifica MTG2. El gráfico derecho muestra la diferencia de los porcentajes de miembros obtenidos por MTG2 y MT2, en función del número de estrellas presentes en las muestras analizadas. Para valores bajos de N_{inif} , MTG2 determina en más casos un mayor porcentaje de miembros de cúmulo que MT2. Se observa una gran dispersión para muestras con menos de 250 objetos, tendiendo a converger conforme el número de estrellas aumenta.

cumplen el criterio en la relación de las dispersiones entre ambas poblaciones impuesto en la obtención de resultados válidos.

En las simulaciones *Sim3* ya observamos que el comportamiento de MTG4 era poco estable en la clasificación de la muestra (ver figura 4.11 del capítulo 4), cuando la distribución de las posiciones está carente de sobre-densidad. El gráfico derecho de la figura 5.17 muestra que, en general, o bien determina altos porcentajes de estrellas de cúmulo o bien inferiores a los simulados. De hecho, de las 168 configuraciones simuladas en las *Sim3*, en 67 no clasificó más del 20% de las estrellas de la muestra como miembros del cúmulo (un 40%). Este comportamiento también lo hemos observado en su aplicación a datos reales, en donde de los 1876 cúmulos analizados, obtiene resultados válidos para 1058, por clasificar menos de un 20% de la muestra como miembros del cúmulo.

En la figura 5.18 se compara el número de miembros determinados por las diferentes metodologías en función de N_{inif} . Se observa una mayor clasificación de estrellas de cúmulo a medida que aumenta N_{inif} . Particularmente, la metodología MTG4 es la que más estrellas clasifica como miembros del cúmulo de manera sistemática, resultado acorde con los obtenidos en las *Sim3*. Se observa un elevado grado de acuerdo en el número de miembros determinados por las metodologías MTG2 y MT2. Sin embargo, MT4 aunque determina un mayor número que estas dos últimas metodologías, a partir de $N_{inif} > 3000$ objetos, su comportamiento se invierte determinando un número menor de miembros en un mayor número de casos. Este comportamiento refleja lo observado en la distribución de las dispersiones obtenidas por MT4 para este rango de N_{inif} , en el espacio de movimientos propios (figura

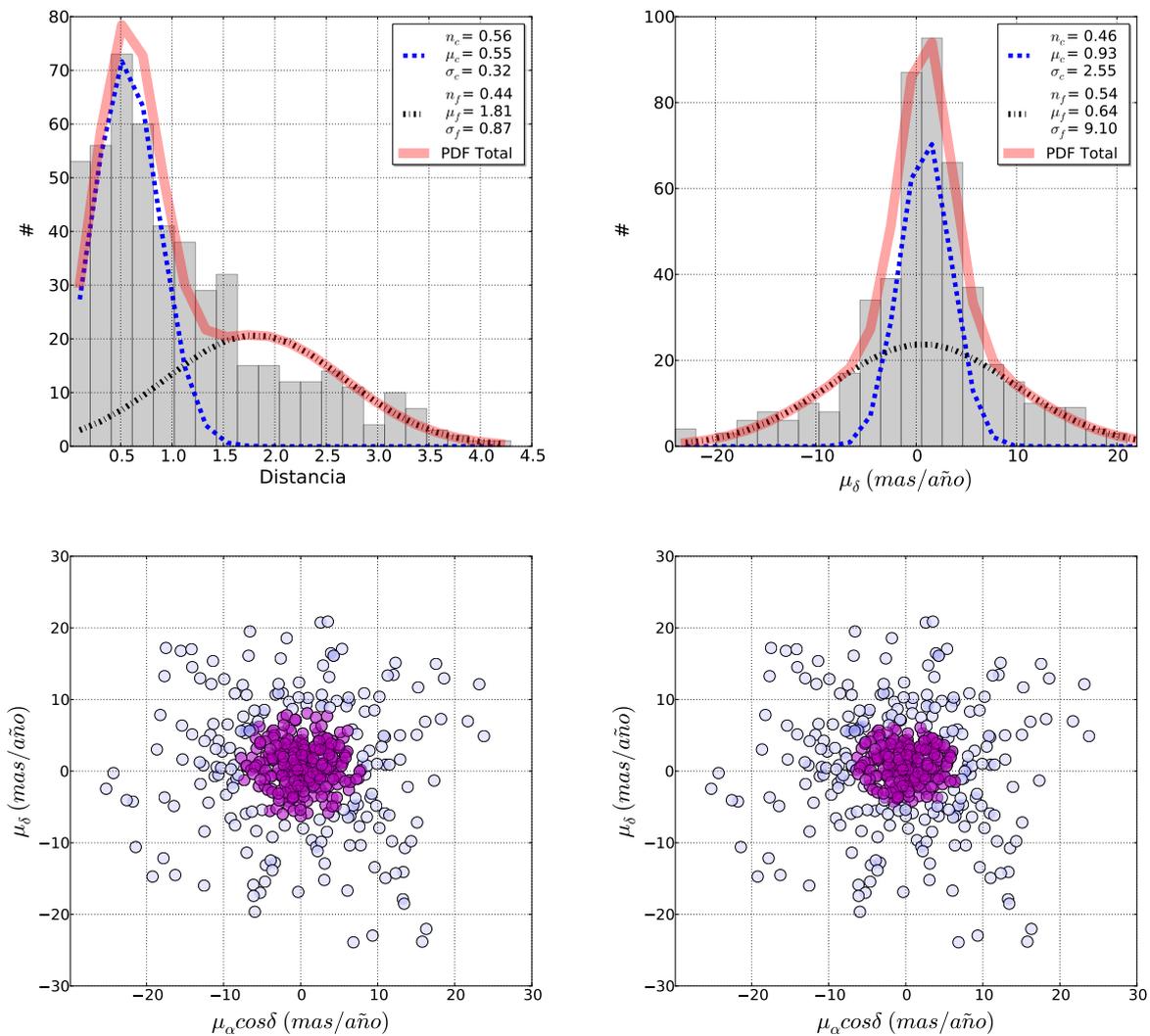


Figura 5.16: Los gráficos superiores muestran los ajustes realizados, a modo de ejemplo en una simulación, por las metodologías MTG2 y MT2 a las distribuciones de distancias y de movimientos propios, respectivamente. La línea discontinua azul hace referencia a la distribución de las estrellas de cúmulo mientras que la línea negra punteada a la distribución de las estrellas de campo. La línea roja continua refiere a la distribución de la densidad de probabilidad de la muestra total. Los gráficos inferiores muestran en magenta, los miembros determinados por las respectivas metodologías, en el espacio de movimientos propios.

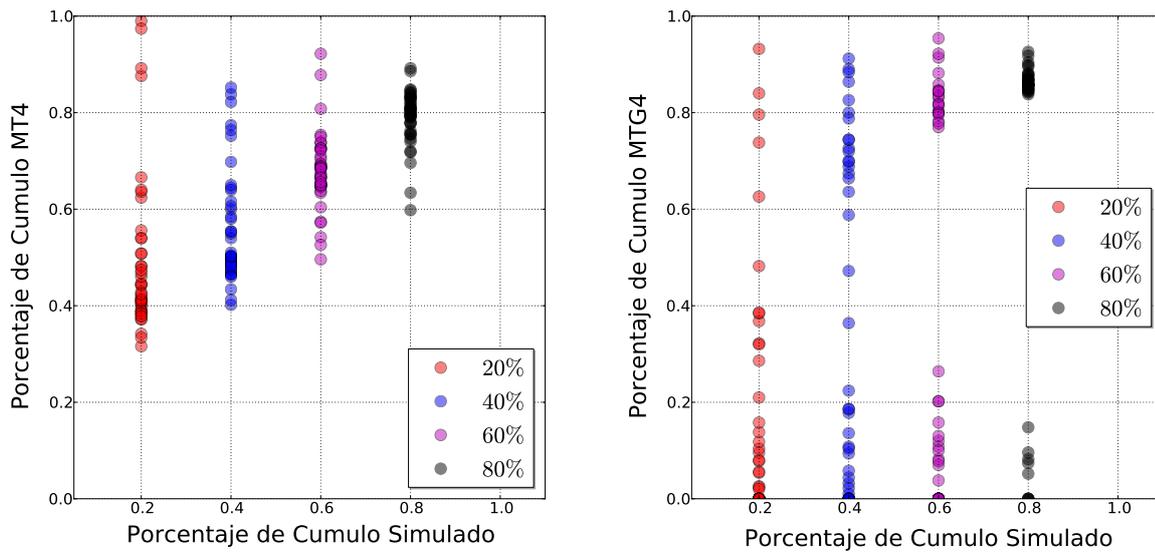


Figura 5.17: Porcentaje de miembros de cúmulo clasificados por MT4 y MTG4 en *Sim3* en función del porcentaje simulado. MT4 (gráfico izquierdo) determina un mayor porcentaje de miembros que el simulado, fundamentalmente para los 2 porcentajes de estrellas de cúmulo más bajos simulados (colores rojo y azul). MTG4 tiene un comportamiento inestable determinando menos del 20% del cúmulo (respecto de lo simulado) en 67 de las 168 configuraciones simuladas (un 40%).

5.13).

La gráfica 5.19 presenta aquellos cúmulos para los que se encuentra una mayor diferencia en el número de miembros determinados entre MT4 y MTG2 (similar para MT2). En ella se representa $N_{ini f}$ en función del diámetro angular de los cúmulos, donde los colores hacen referencia a la diferencia entre el número de miembros determinados en estos dos análisis. Se observa que las mayores diferencias se obtienen para aquellos cúmulos con mayores valores de estos dos parámetros. Cúmulos con elevados valores del diámetro angular pueden tener presencia de sub-estructuras en el espacio de posiciones, haciendo que aquellas regiones con mayor densidad sean clasificadas como cúmulo por la metodología MT4. Al mirar los PDR de estos cúmulos, observamos que la propia determinación del radio es imprecisa al no observarse una sobre-densidad clara. Hay que tener presente que estas sub-estructuras pueden no coincidir necesariamente con aquellas relacionadas con el cúmulo, sino que, pueden ser producidas por otros factores tales como extinción variable, etc.

Por otro lado, el número de estrellas analizado es elevado (superior a las 3000 estrellas), factor muy importante si tenemos en cuenta las incertidumbres en los movimientos propios con las que estamos trabajando. En general, para estos cúmulos se obtiene una distribución en movimientos propios como la que se muestra en la figura 5.20 para el cúmulo Alessi 19 (gráfico izquierdo). En estos casos, las metodologías que usan únicamente movimientos propios, clasificarán la región central de la distribución como cúmulo, pudiendo no corresponderse con la posición del cúmulo en este espacio. Solamente en aquellos casos en los que las poblaciones estén bien diferenciadas (ej. Melotte 22 gráfico derecho de la figura 5.20) o exista una sobre-densidad clara asociada al cúmulo, podremos considerar si los análisis

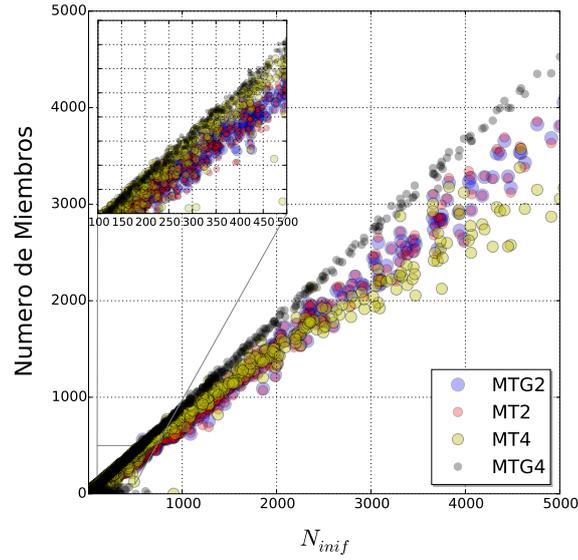


Figura 5.18: Número de miembros obtenidos por cada metodología para la muestra de 1876 cúmulos reales, en función del número de estrellas analizadas. Se observa que MTG4 (círculos negros), determina sistemáticamente más miembros que el resto de las metodologías. Se observa un elevado grado de acuerdo entre el número de miembros determinados por MTG2 (círculos azules) y por MT2 (círculos rojos). En amarillo se muestra el número de miembros determinado por MT4.

han realizado una buena clasificación de la muestra. Otro factor importante a tener en cuenta es que la dispersión de velocidades en cúmulos cercanos puede no ser despreciable frente a los errores de UCAC4 pudiendo incurrir en la no observación de las dos poblaciones en el espacio de movimientos propios.

Se observa de nuevo un elevado grado de acuerdo en la determinación de los centroides en el espacio de los movimientos propios. En la figura 5.21 se muestra a modo de ejemplo la diferencia en los centroides obtenidas entre los cúmulos comunes a MTG2 y MT2. Para hacer el ajuste a las elipses se utilizó el centroide, dispersión y coeficiente de correlación de la distribución resultante tras eliminar aquellos objetos con distancias superiores a 5σ 's. Los colores de las elipses hacen referencia a los valores de las dispersiones de 1, 2 y 3 σ 's, estando indicados en la tabla 5.2 los valores de las dispersiones correspondientes a 1σ . Los cúmulos con diferencias superiores a 3 σ 's y con $N_{inif} < 100$ han sido representados en color azul mientras que en rojo se han representado los que tienen $N_{inif} > 100$.

En la tabla 5.2, también se muestra el valor medio de estas diferencias ($\Delta\mu_\alpha \cos\delta$ y $\Delta\mu_\delta$), el número de cúmulos comunes entre cada metodología (denotado por N_0), y el porcentaje de muestras (respecto del número de cúmulos comunes), con $N_{inif} < 100$ (*Porcentaje 1*) y con $N_{inif} > 100$ (*Porcentaje 2*) que se encuentran a más de 3 σ 's de la distribución. Los valores muestran que, las dispersiones correspondientes a 1σ obtenidas entre metodologías, son mayores entre las metodologías MT4 y MT2, aunque siendo estas inferiores a 1 mas/año. Los valores de las dispersiones junto con los valores medios de las diferencias obtenidas, en torno a 0 mas/año, muestran un elevado grado de acuerdo en

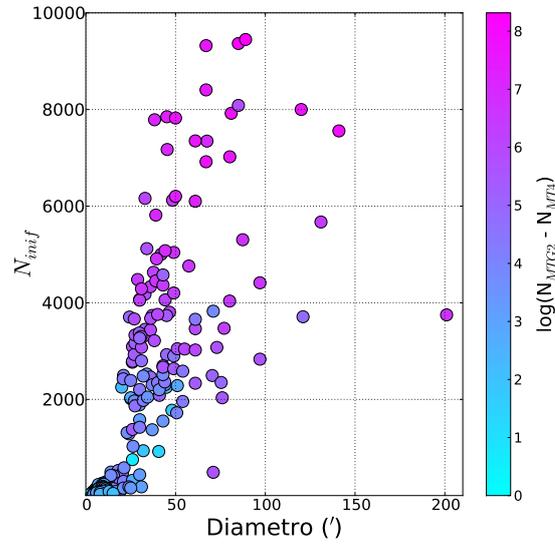


Figura 5.19: Número de estrellas analizadas en función del diámetro angular de los cúmulos analizados. Se observa que los cúmulos con mayores valores del radio proyectado, y por tanto, un mayor número de estrellas a analizar (N_{inif}), son en los que se encuentra una mayor diferencia en el número de miembros obtenidos por MTG2 y MT4.

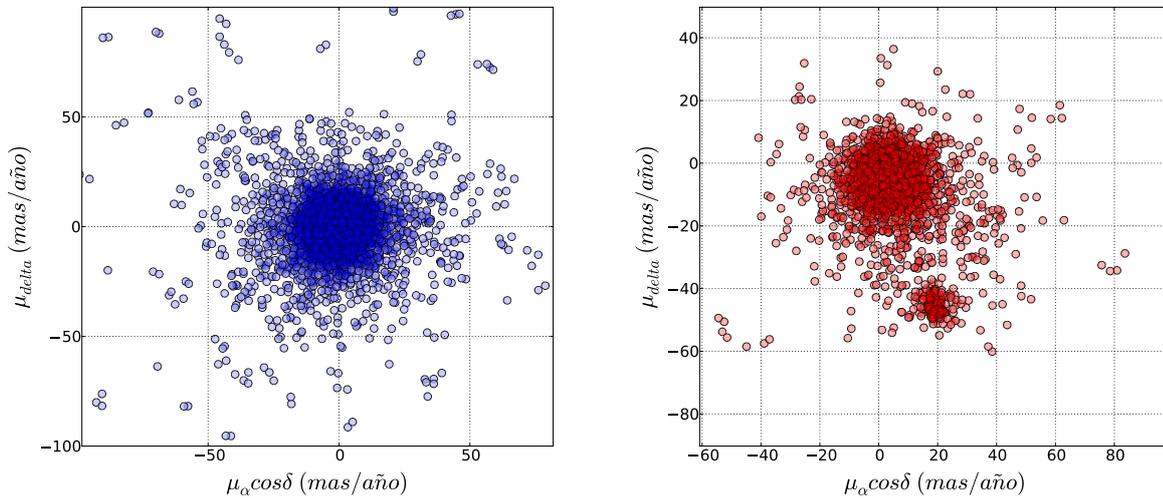


Figura 5.20: Distribución de los movimientos propios en las regiones de los cúmulos Alessi 19 y Melotte 22, gráficos izquierdo y derecho, respectivamente. En el caso de la distribución observada en la región del cúmulo Alessi 19 no se observa una separación entre ambas poblaciones de estrellas. Consecuentemente, MTG2 y MT2 clasificarán la región central de la distribución como cúmulo, pudiendo no corresponderse con la posición del cúmulo en este espacio.

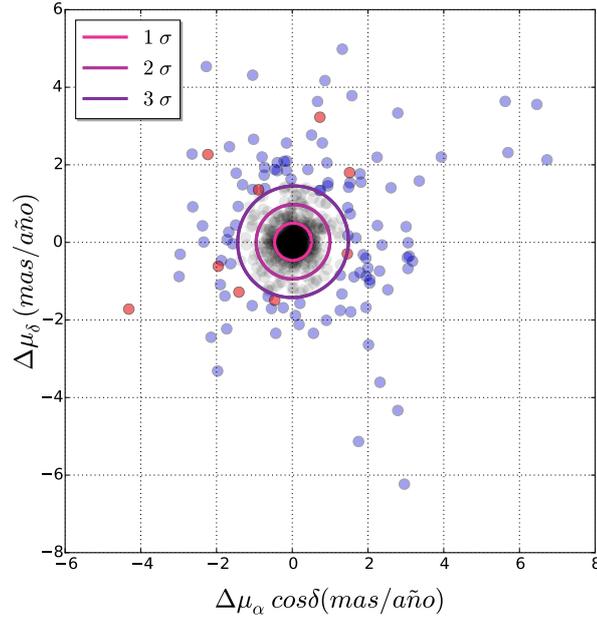


Figura 5.21: Diferencias entre los centroides obtenidos por MTG2 y MT2 en el espacio de movimientos propios. Cada cruz gris refiere a un cúmulo. Los círculos superpuestos refieren a aquellos cúmulos situados a más de 3σ 's cuya muestra está formada por menos o más de 100 objetos en azul o rojo respectivamente.

los resultados obtenidos por los diferentes análisis de pertenencia.

El número de cúmulos comunes es elevado, exceptuando aquellas comparaciones realizadas respecto de MTG4, al ser ésta la que menos cúmulos con resultados válidos obtiene. MTG2 y MT4 son las que muestran el mayor número de cúmulos comunes con 1589. La comparación entre MTG2 y MT2 ha obtenido el *Porcentaje 1* más elevado, de un 7.1 % de cúmulos. Hemos encontrado un *Porcentaje 2* de cúmulos del 2.6 % en las comparaciones entre MTG4 y MTG4, llegando a alcanzarse el 1% entre MTG2 y MT2 y, MT2 y MT4. Exceptuando la comparativa entre MTG4 y MT4, se observa que para el resto de metodologías la mayoría de los cúmulos que presentan unas diferencias en la determinación del centroide superiores a 3σ 's se corresponden a muestras con $N_{inif} < 100$ estrellas.

Los cúmulos correspondientes al *Porcentaje 1* han sido inspeccionados visualmente y las variaciones encontradas en la determinación de los centroides se deben exclusivamente al bajo número de estrellas presentes en las muestras analizadas.

Cúmulos más interesantes son aquellos en los que, aún teniendo un número mayor de estrellas en la muestra a analizar ($N_{inif} > 100$), los resultados de los análisis de pertenencia llevan a determinaciones de cúmulo en regiones diferentes en el espacio de movimientos propios (círculos rojos en la figura 5.21). Si nos fijamos en la tabla 5.2, estos cúmulos (*Porcentaje 2*), son una minoría respecto del total de cúmulos comunes, siendo el porcentaje máximo alcanzado del 2.6 %. Para estos cúmulos, la gran mayoría de las diferencias encontradas en la determinación del centroide, se deben exclusivamente a la naturaleza de los análisis de pertenencia en sí mismos. En el caso específico de MTG4, aquellos

Tabla 5.2: La tabla muestra los valores de los centroides y de las dispersiones, en mas/año, de las diferencias en la determinación de los centroides en el espacio de movimientos propios a 1σ ($\sigma_{\Delta\mu_{\alpha}\cos\delta}$, $\sigma_{\Delta\mu_{\delta}}$), obtenidas entre las metodologías, el número de miembros comunes (N_0) y los porcentajes de cúmulos (respecto del número de cúmulos comunes) que, situados a más de 3σ 's, tienen menos (*Porcentaje 1*) y más (*Porcentaje 2*) de 100 objetos en la muestra a analizar.

Metodologías	MTG2-MTG4	MTG2-MT4	MTG2-MT2	MT4-MT2	MTG4-MT4	MTG4-MT2
$\Delta\mu_{\alpha}\cos\delta$	0.030	0.012	0.021	0.023	0.016	0.016
$\Delta\mu_{\delta}$	0.087	0.032	0.016	-0.016	-0.047	-0.091
$\sigma_{\Delta\mu_{\alpha}\cos\delta}$	0.499	0.756	0.489	0.867	0.582	0.600
$\sigma_{\Delta\mu_{\delta}}$	0.496	0.736	0.478	0.821	0.564	0.600
N_0	1035	1589	1531	1489	1018	997
Porcentaje 1	4.4%	4.2%	7.1%	4.8%	3.3%	4.2%
Porcentaje 2	1.5%	1.1%	1.0%	1.0%	2.6%	1.4%

cúmulos que tengan absorciones no homogéneas revierten en una mala determinación de los centroides en posiciones, a partir de los cuales se calcula la distribución de distancias y por ende, en una mala determinación de los miembros del cúmulo. En general, la utilización únicamente de los movimientos propios o la introducción de las posiciones en los análisis de pertenencia, unido al bajo número de estrellas en las muestras analizadas, o la distribución de estas variables, producen por sí mismas diferencias en la determinación de los centroides. Algunos ejemplos de estos cúmulos son Czernick 38, Blanco 1 o NGC 2632, los cuales nos ayudarán a ilustrar estas diferencias encontradas y representadas por el *Porcentaje 2*.

Como ya sabemos, la distribución espacial de los cúmulos abiertos es muy variable, presentando desde distribuciones radiales con sobre-densidad central a distribuciones con presencia de subestructuras a lo largo de toda la región del cúmulo. Pero independientemente de la existencia o ausencia de sobre-densidad, la distribución de posiciones puede verse afectada por problemas de absorción interestelar no homogénea. Por la propia definición de cúmulo que asume MT4, esta metodología determinará la región más densa analizada como cúmulo. Así, la mayoría de los cúmulos en los que existe una discrepancia entre MT4 y MT2 o MTG2, se debe precisamente a este motivo. Un ejemplo se muestra en la figura 5.22 (gráfico superior) para el cúmulo Czernik 38, donde se puede observar una mayor existencia de estrellas en la región superior izquierda respecto al resto de la muestra. Como consecuencia, MT4 determinará que las estrellas pertenecientes a esta región, al ser la más densa, pertenecerán al cúmulo (círculos magenta en los gráficos inferior izquierdo). MTG2 al utilizar únicamente los movimientos propios determinará como cúmulo la región más densamente distribuida de este espacio (círculos azules gráfico inferior derecho). Las diferencias en las definiciones de cúmulo realizadas por las dos metodologías pueden llevar, como es el caso de este cúmulo, a su determinación en diferentes regiones en los espacios de posiciones y de movimientos propios.

Por otro lado, la existencia de una clara separación entre cúmulo y campo puede llevar no sólo a la determinación de cúmulo en diferentes regiones de la muestra, sino a la mala determinación del mismo. Ejemplos de estos cúmulos son NGC 2632 (figura 5.23) y Blanco 1 (figura 5.24). En el gráfico superior izquierdo y derecho de estas figuras, se muestran los espacios de movimientos propios de la región de cada uno de estos dos cúmulos. En los gráficos superiores derechos se muestran las determinaciones de los miembros realizada por MT4 (en rojo) y a modo de ejemplo la realizada por MT2 (en

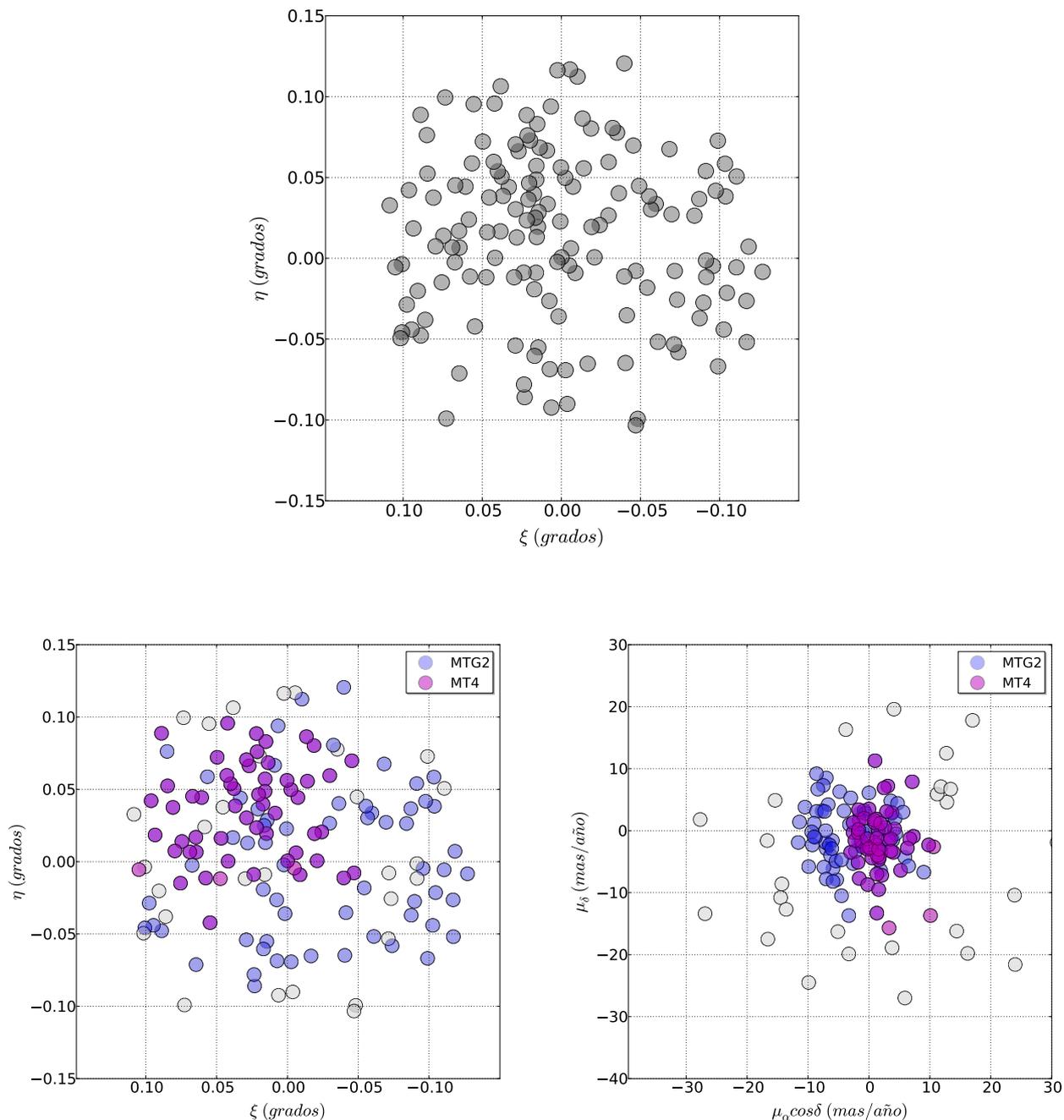


Figura 5.22: Implicación en los resultados obtenidos de la existencia de absorciones no homogéneas a lo largo de la región del cúmulo Czernik 38. En el gráfico superior se muestra la distribución en posiciones de este cúmulo, en la que se pueden observar regiones más densas que otras. Como consecuencia, MT4 a partir de su definición de miembro de cúmulo, determinará que las regiones más densas en ambos espacios, se corresponderán con el cúmulo (círculos rojos de los gráficos inferiores). MTG2 (círculos azules) al utilizar únicamente los movimientos propios para realizar los análisis de pertenencia determinará que el cúmulo se encuentra en la región más densa de este espacio. Las diferentes definiciones de cúmulo así como en este caso la presencia de absorciones, produce que la determinación de cúmulo sea diferente entre metodologías implicando la determinación del mismo en diferentes regiones de los espacios de posiciones y de movimientos propios.

azul), respectivamente (resultados similares a esta última se obtuvieron con MTG2 y MTG4). En los gráficos intermedios se muestran las distribuciones de probabilidad, en este espacio, obtenidas por las metodologías MT2 (izquierda) y MT4 (derecha), a partir de las cuales las metodologías determinan los miembros.

Para el cúmulo NGC 2632 (figura 5.23), MT4 determina que el cúmulo está formado por dos agrupaciones de estrellas situadas en diferentes regiones del espacio de movimientos propios, siendo una de ellas el cúmulo real. La obtención de este resultado es debido a la gran concentración de estrellas presentes en dos regiones diferentes de la muestra del cúmulo NGC 2632 y a la separación existente entre ellas. A su vez, MT2 (extrapolable a MTG2) determina como cúmulo la región de mayor sobre-densidad de la muestra observada en el espacio de movimientos propios, correspondiente a la distribución de estrellas de campo. En el caso del cúmulo Blanco 1 (figura 5.24), MT4 es capaz de clasificar la muestra correctamente en cúmulo y campo. Sin embargo, MTG2, MTG4 y MT2 determinan una región de cúmulo mayor con una mayor contaminación de estrellas de campo.

La explicación a estos ajustes erróneos realizados por las metodologías geométrica y paramétrica recae en los valores de los parámetros de entrada de las distribuciones de ambas poblaciones, necesarios para iniciar el análisis de pertenencia. La necesidad de automatizar el proceso para el análisis de los cúmulos presentes en el catálogo DAML02, nos lleva a automatizar también los valores de entrada de estos parámetros. Como ya se ha comentado, para la metodología paramétrica se han tomado los valores medios y las dispersiones de las muestras a analizar (una vez eliminados los *outliers*), como valores representativos de la población de campo. Los valores de entrada para las dispersiones de la población de cúmulo son 5 veces inferiores a las de campo. En cuanto al valor de los centroides de los cúmulos se ha tomado, como hipótesis de partida, las modas de las distribuciones.

En un principio podemos asumir que estos valores pueden ser aproximadamente representativos de ambas poblaciones. A partir de ellos, la propia distribución de probabilidad determinará la clasificación final de las estrellas. Sin embargo, para algunos cúmulos como son los casos de NGC 2632 y Blanco 1, estos valores de entrada no son representativos de ambas poblaciones, de hecho pueden tomar valores muy alejados a estos últimos, lo cual nos lleva a una mala clasificación de la muestra. Para NGC 2632, al estar ambas poblaciones lo suficientemente alejadas una de la otra, impide a la propia distribución de probabilidad que nos conduzca a la región en la que se encuentra el cúmulo.

En el caso específico de la metodología MT2, simplemente introduciendo unos parámetros más aproximados a las distribuciones de cúmulo y campo, es capaz de realizar la correcta determinación de los cúmulos Blanco 1 y NGC 2632 (ver gráficos inferiores de las figuras 5.23 y 5.24). A su vez MTG2 no es capaz de determinar la región del cúmulo Blanco 1, aunque introduzcamos unos valores de los parámetros más aproximados a las distribuciones de ambas poblaciones. Sin embargo, MTG2 sí es capaz de realizar una buena clasificación para la región del cúmulo NGC 2632.

Tanto NGC 2632 como Blanco 1, se encuentran a distancias muy cercanas de 187 pc y 269 pc respectivamente, con un radio proyectado de 35.5 minutos de arco. Este elevado valor produce la inclusión de un gran número de estrellas de campo en la muestra a analizar, haciendo que menos del 20% de la muestra sea referente al cúmulo (en torno al 14.5 % para NGC 2632 y al 19% para Blanco

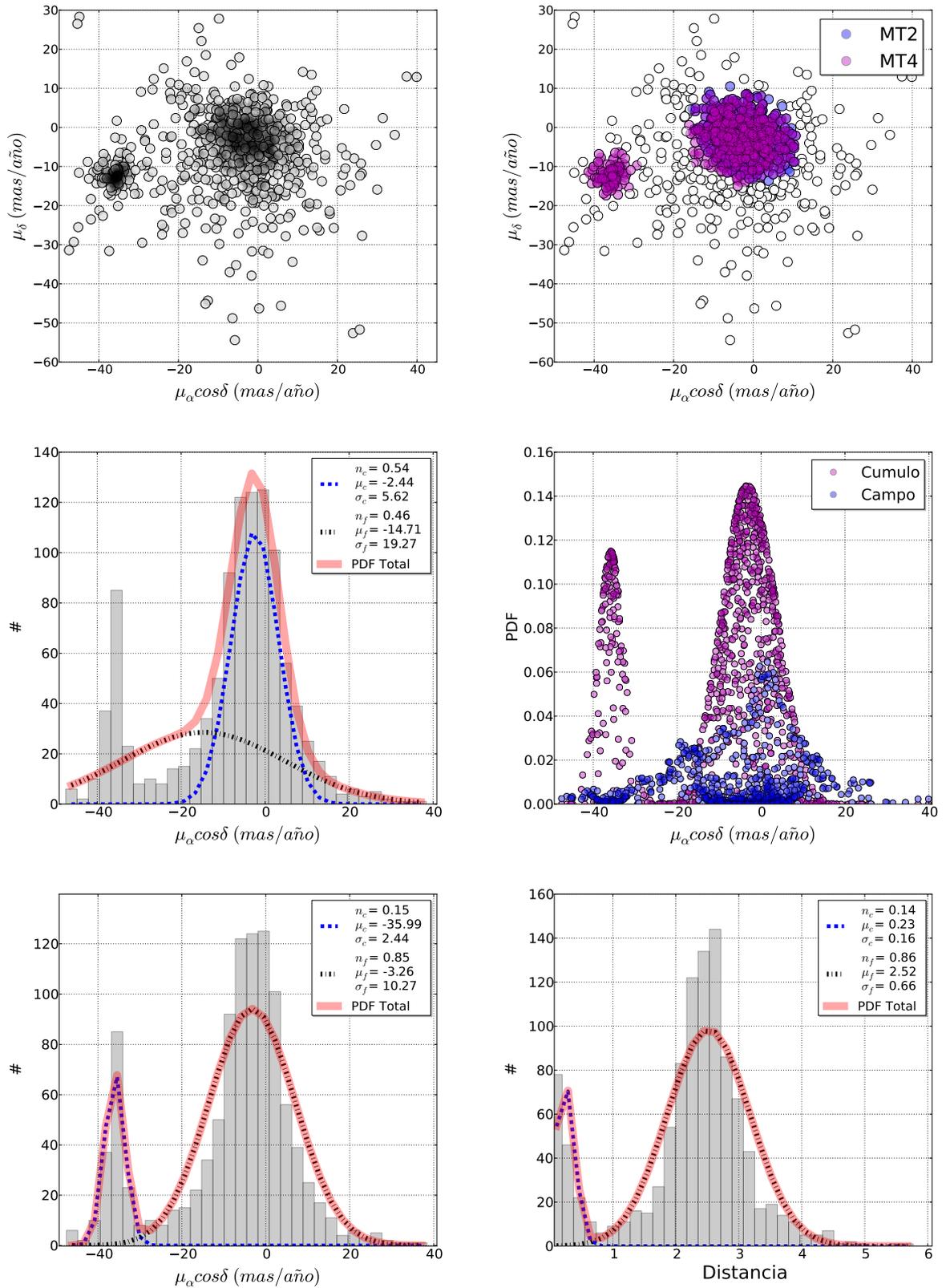


Figura 5.23: En los gráficos superiores se muestra la distribución de movimientos en la región del cúmulo NGC 2632 (izquierdo) y la determinación de miembros realizada por las metodologías MT2 en azul y MT4 en magenta (derecho). En los gráficos intermedios se muestran las distribuciones de probabilidad obtenidas por MT2 y MT4 (gráficos izquierdo y derecho, respectivamente), a partir de las cuales han clasificado la muestra. Los gráficos inferiores muestran el ajuste de MT2 y de MTG2 a la distribución de movimientos propios y distancias cuando los parámetros de entrada introducidos son más próximos a las distribuciones de ambas poblaciones de estrellas.

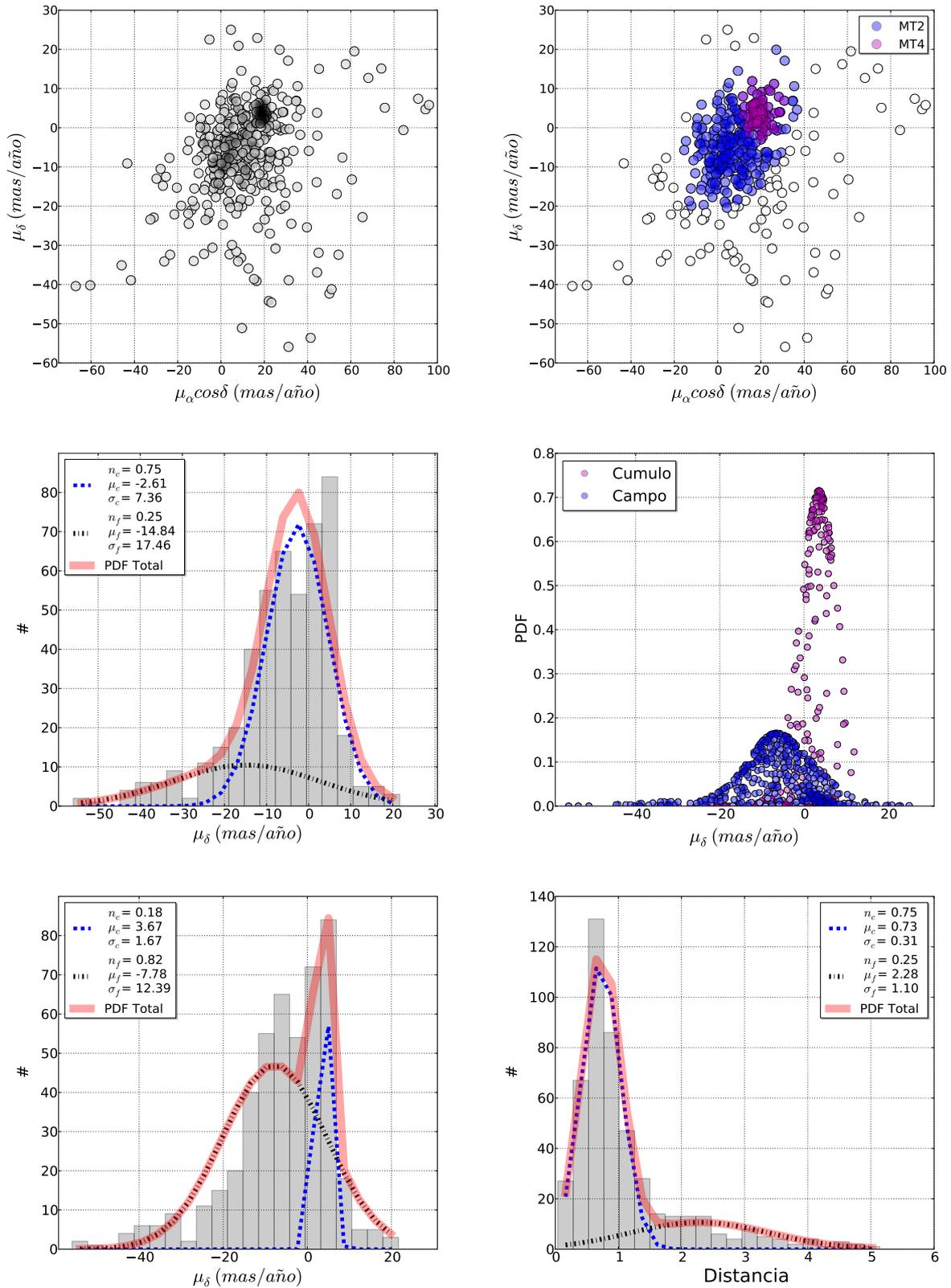


Figura 5.24: En los gráficos superiores se muestra la distribución de movimientos en la región del cúmulo Blanco 1 (izquierdo) y la determinación de miembros realizada por las metodologías MT2 en azul y MT4 en magenta (derecho). En los gráficos intermedios se muestran las distribuciones de probabilidad obtenidas por MT2 y MT4 (gráficos izquierdo y derecho, respectivamente), a partir de las cuales han clasificado la muestra. Los gráficos inferiores muestran el ajuste de MT2 y de MTG2 a la distribución de movimientos propios y de distancias cuando los parámetros de entrada introducidos son más próximos a las distribuciones de ambas poblaciones de estrellas.

1).

Los resultados obtenidos de estos análisis de pertenencia se publicarán en forma de dos catálogos. Para cada uno de los cúmulos analizados, se listarán los parámetros que describen tanto a la población de cúmulo como de campo, en los espacios de posiciones y de movimientos propios. Además, se incluirán los radios estimados en esta tesis, junto con las distancias, edades y excesos de color listados en Dias et al. (2014). En el apéndice C se encuentra de forma detallada cada uno de estos parámetros únicamente para la población de cúmulo, habiendo sido calculados también para la población de campo.

Por otro lado, para cada uno de estos cúmulos se publicará un catálogo con los valores de posiciones y movimientos propios de cada estrella así como las probabilidades de pertenencia determinadas por cada metodología. En este último catálogo se incluirá un indicador que tomará el valor 1 para aquellas estrellas que hayan sido clasificadas como *outliers* en el espacio de movimientos propios. Además, también se incluirán unas columnas finales de ceros y unos indicando que estrellas han sido clasificadas por cada análisis como campo (0) y cúmulo (1), respectivamente. En el caso en el que alguna metodología no haya determinado cúmulo, se indicará con un -1. Un ejemplo de este último catálogo se pueden observar en la figura 5.25.

5.4 Comparación con otras bases de datos

Además de comparar y describir los resultados obtenidos entre las metodologías aplicadas a lo largo de esta tesis, hemos comparado estos resultados con aquellos publicados en otras grandes e importantes bases de datos, comúnmente utilizadas por la comunidad científica. Nos referimos concretamente a los trabajos Dias et al. (2014) (D14) y Dias et al. (2002) (DAML02).

5.4.1 Comparación con el trabajo D14.

En este trabajo se analizan 1805 cúmulos utilizando los datos del cartografiado UCAC4. Los radios muestrales usados para llevar a cabo el análisis de pertenencia proceden del catálogo DAML02. La metodología utilizada es de tipo paramétrico basada en el trabajo Uribe & Brieva (1994), la cual aproxima la distribución de los movimientos propios por una mezcla de dos Gaussianas bivariantes, las cuales incluyen los errores de las estrellas (Zhao & He, 1990) en la estimación de la probabilidad. Para la determinación y posterior eliminación de los *outliers*, ajustan la distribución de movimientos propios por un modelo Gaussiano, eliminando aquellas estrellas situadas a más de 3σ 's de la distribución.

En la figura 5.26 se muestran las diferencias entre los centroides obtenidos entre las metodologías aplicadas en esta tesis y los listados en el trabajo de D14. Se superponen unas elipses que se corresponden con las dispersiones de 1, 2 y 3σ 's estimadas a partir de la muestra resultante tras eliminar aquellos cúmulos que se encuentran a más de 5σ 's de la distribución (los valores de las dispersiones a 1σ están listados en la tabla 5.3). En esta figura los círculos negros representan las diferencias

#	UCAC4	Ra (deg)	erRa (mas)	Dec (deg)	erDec (mas)	PmRa (mas/yr)	ProbDuo (mas/yr)	erPmRa (mas/yr)	PmDec (mas/yr)	MemSa1	erPmDec (mas/yr)	MemDuo					
		ProbGeo_pm	ProbGeo_pospm	ProbSa1_pm	ProbSa1_pospm	ProbSa1_join	ProbDuo	MemGeo_pm	MemGeo_pospm	MemSa1	MemDuo						
727-080663	332.52901	115	55.38955	60	3.4	5.5	3.2	4.9	0	0.9793	-1.0	0.9999	1.0	0.8266	1	-1	1
727-080667	332.5341	42	55.38001	31	-6.1	4.2	-5.1	4.1	0	0.9872	-1.0	0.9945	1.0	0.8591	1	-1	1
727-080678	332.53793	124	55.38386	119	-12.6	5.6	0.2	5.4	0	0.9165	-1.0	0.9976	1.0	0.7529	1	-1	1
727-08068	332.53976	62	55.39255	59	-16.1	5.2	-7.7	5.0	0	0.5481	-1.0	0.9926	0.9989	0.4876	1	-1	1
727-080685	332.54257	121	55.39329	121	9.8	7.3	12.7	7.0	0	0.272	-1.0	0.9996	0.9999	0.0643	0	-1	1
727-08069	332.54686	30	55.37658	46	-13.2	5.1	4.4	5.8	0	0.8359	-1.0	0.9892	0.9999	0.5587	1	-1	1
727-080696	332.55	84	55.3761	83	77.9	5.2	-13.4	5.0	1	0.0	-1.0	0.0	0.0	0.0	0	-1	0
727-080697	332.55025	91	55.37168	159	-0.8	5.4	-8.1	6.2	0	0.9874	-1.0	0.9814	0.9996	0.7892	1	-1	1
727-080698	332.55075	156	55.3655	145	-4.0	6.2	-3.1	5.8	0	0.9901	-1.0	0.9973	0.9996	0.8908	1	-1	1
727-080707	332.55496	52	55.39923	58	-4.1	5.1	-2.0	4.8	0	0.9901	-1.0	0.9983	0.9984	0.8938	1	-1	1
727-080709	332.55682	23	55.38952	23	-8.1	5.2	4.2	4.9	0	0.9724	-1.0	0.999	0.9997	0.746	1	-1	1
727-080712	332.55982	40	55.3846	36	-3.9	5.2	-5.7	4.9	0	0.9894	-1.0	0.9924	0.9986	0.8609	1	-1	1
727-080715	332.56171	155	55.38792	149	9.1	6.2	-0.8	5.9	0	0.8999	-1.0	0.9995	0.9998	0.7845	1	-1	1
727-080721	332.56702	118	55.3827	119	-7.2	5.6	-2.1	5.4	0	0.9858	-1.0	0.9985	0.9995	0.8735	1	-1	1
727-080723	332.56812	23	55.37349	37	-8.2	3.6	-8.4	4.0	0	0.9733	-1.0	0.9839	0.9975	0.7305	1	-1	1
727-080726	332.56933	106	55.39882	95	4.0	5.8	-31.1	5.4	0	0.0016	-1.0	0.0198	0.0498	0.0	0	-1	0
727-080735	332.57525	34	55.35657	36	-7.3	5.1	20.3	4.9	0	0.1065	-1.0	0.9921	0.9722	3.0E-4	0	-1	1
727-080737	332.57658	79	55.38507	74	5.4	5.3	-2.4	5.0	0	0.9755	-1.0	0.9988	0.9983	0.8555	1	-1	1
727-080741	332.57915	46	55.35548	107	26.0	5.2	-19.8	5.4	0	0.0	-1.0	0.0077	0.0	0.0	0	-1	0
727-080745	332.58164	124	55.3777	120	-2.9	5.7	-1.3	5.4	0	0.9903	-1.0	0.9987	0.9985	0.897	1	-1	1
727-080746	332.58249	23	55.35832	31	6.3	5.2	9.3	4.9	0	0.8513	-1.0	1.0	1.0	0.3638	1	-1	1
727-080749	332.58474	170	55.39172	166	-6.9	6.5	-4.4	6.3	0	0.986	-1.0	0.9961	0.978	0.862	1	-1	1
727-080755	332.58936	15	55.37719	15	-5.4	5.3	-4.1	5.0	0	0.9888	-1.0	0.9962	0.9901	0.8767	1	-1	1
727-080759	332.59304	63	55.39504	74	-25.0	5.3	15.5	5.1	0	0.0	-1.0	0.3898	0.1327	1.0E-4	0	-1	0

Figura 5.25: Ejemplo de catálogo para un cúmulo individual. En este caso para el cúmulo NGC 7226, en el que aparecen para cada estrella de la muestra el identificador de UCAC4, las posiciones y los movimientos propios, junto con sus errores, un indicador que define si la estrella ha sido clasificada como *outlier*, las probabilidades de pertenencia obtenidas por cada metodología y cuatro columnas de 0's y 1's indicando si la metodología ha clasificado a la estrella como estrella de campo o miembro de cúmulo, respectivamente. En el caso en el que alguna metodología no haya determinado cúmulo, se representará con el valor de -1.

Tabla 5.3: La tabla muestra los valores de los centroides y de las dispersiones, en mas/año, de las diferencias en la determinación de los centroides en el espacio de movimientos propios a 1σ ($\sigma_{\Delta\mu_{\alpha}cos\delta}$, $\sigma_{\Delta\mu_{\delta}}$), obtenidas entre las metodologías y el trabajo D14, el número de miembros comunes (N_0) y porcentajes de cúmulos (respecto del número de cúmulos comunes) que, situados a más de 3σ 's, tienen menos (*Porcentaje 1*) o más (*Porcentaje 2*) de 100 objetos en la muestra a analizar.

Metodologías	D14 -MTG2	D14-MTG4	D14-MT4	D14-MT2
$\Delta\mu_{\alpha}cos\delta$	0.050	0.010	0.039	0.027
$\Delta\mu_{\delta}$	0.018	-0.048	0.002	0.006
$\sigma_{\Delta\mu_{\alpha}cos\delta}$	0.726	0.732	1.007	0.843
$\sigma_{\Delta\mu_{\delta}}$	0.723	0.772	0.916	0.862
N_0	1588	1014	1558	1481
<i>Porcentaje 1</i>	3.4%	2.8%	3.5%	4.8%
<i>Porcentaje 2</i>	1.4%	2.2%	0.9%	1.0%

en los centroides obtenidas para cada cúmulo y, los puntos rojos y azules superpuestos, a aquellos cúmulos que, estando a una distancia superior a 3σ 's, su N_{inif} contiene más o menos de 100 objetos, respectivamente.

En la tabla 5.3, se muestra el valor medio de estas diferencias ($\Delta\mu_{\alpha}cos\delta$ y $\Delta\mu_{\delta}$), los valores de las dispersiones a 1σ , el número de cúmulos comunes entre cada análisis y los cúmulos estudiados en D14 (denotado por N_0), y el porcentaje de muestras (respecto del número de cúmulos comunes), con $N_{inif} < 100$ (*Porcentaje 1*) y con $N_{inif} > 100$ (*Porcentaje 2*) que se encuentran a más de 3σ 's de la distribución. Los valores muestran que, las dispersiones correspondientes a 1σ obtenidas entre metodologías son mayores respecto a los obtenidos en la sección previa aunque siendo prácticamente inferiores a 1 mas/año. Los centroides de las diferencias están muy próximos a 0 mas/año. Este dato junto con los valores de las dispersiones indican que los resultados obtenidos por nuestros análisis de pertenencia producen resultados muy similares a aquellos listados en D14.

El análisis de pertenencia realizado en D14 por sí mismo puede llevar a diferencias en la determinación de los miembros de los cúmulos y, consecuentemente, a variaciones en la estimación de los centroides respecto a las nuestras. Hay que tener presente que el radio que se ha utilizado en este trabajo puede no coincidir con el utilizado en esta tesis. Este hecho podría ser otra causa de la obtención de una mayor dispersión en la determinación de los centroides. Por otro lado, de nuevo, los diferentes análisis de pertenencia realizados en los que diferentes conjuntos de variables son utilizados, la existencia de absorciones diferenciales y de un bajo número de estrellas presentes en la muestra, son las causas principales de las diferencias encontradas en la determinación de los centroides de los cúmulos en el espacio de movimientos propios. A modo de ejemplo, en la figura 5.27 se muestran los centroides determinados por MTG2 respecto a los obtenidos en D14. La muestra de los cúmulos se representa en azul, superponiendo en amarillo aquellos cúmulos en los que N_{inif} es superior a 1000 estrellas y en magenta aquellos en los que es inferior a 100. Se puede observar que la gran mayoría de los cúmulos con una mayor diferencia en la determinación del centroide se corresponden con cúmulos con N_{inif} bajo, exceptuando uno, Melotte 22, en el que nuestras metodologías no hacen una buena determinación de los miembros que pertenecen a este cúmulo.

Es la metodología MTG2 la que tiene un mayor número de cúmulos comunes con D14 (1588

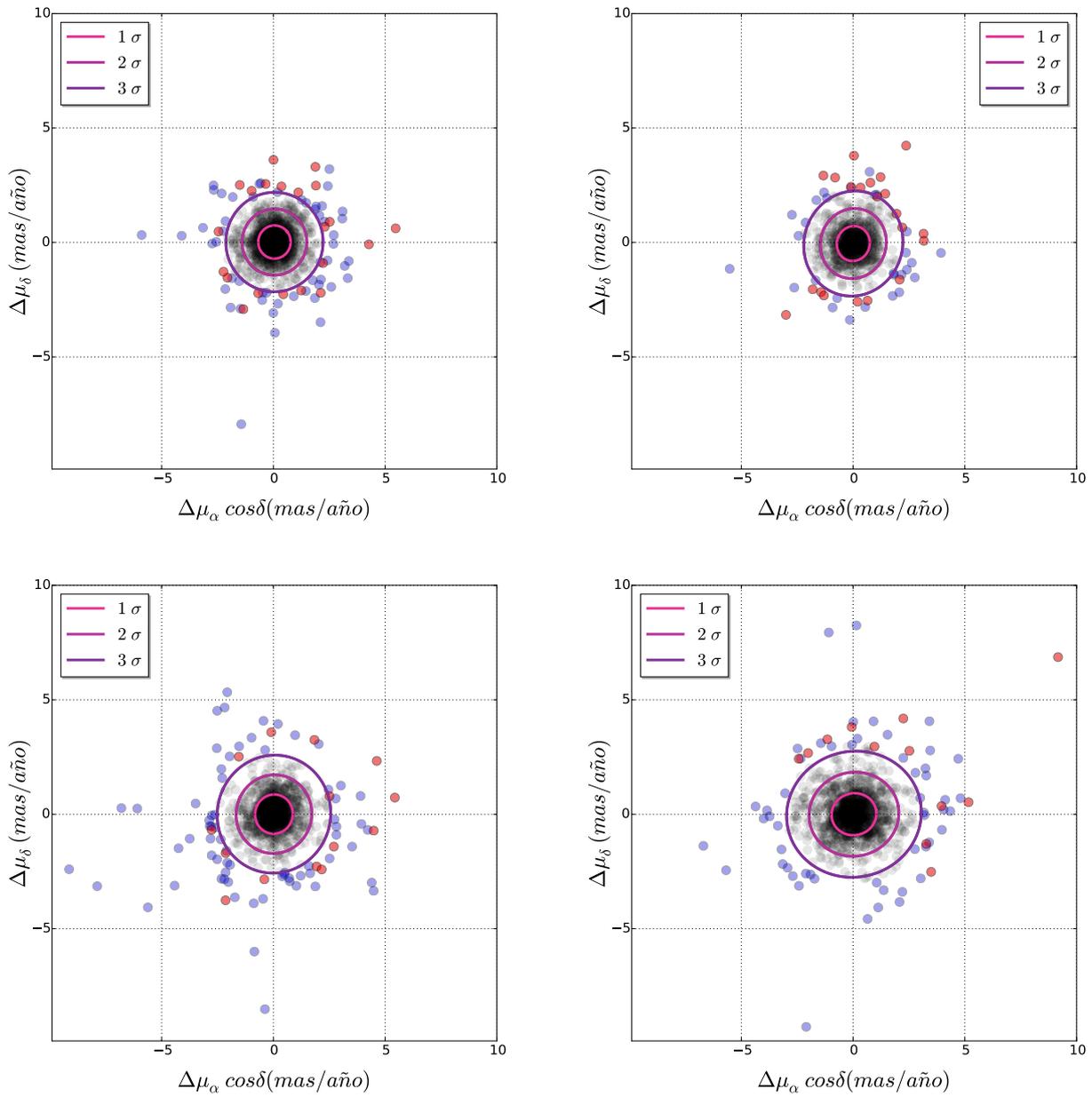


Figura 5.26: Diferencias en los centroides en el espacio de movimientos propios entre las metodologías MTG2 (gráfico superior izquierdo), MTG4 (gráfico superior derecho), MT2 (gráfico inferior izquierdo) y MT4 (gráfico inferior derecho) respecto a los valores de los centroides publicados en el trabajo D14.

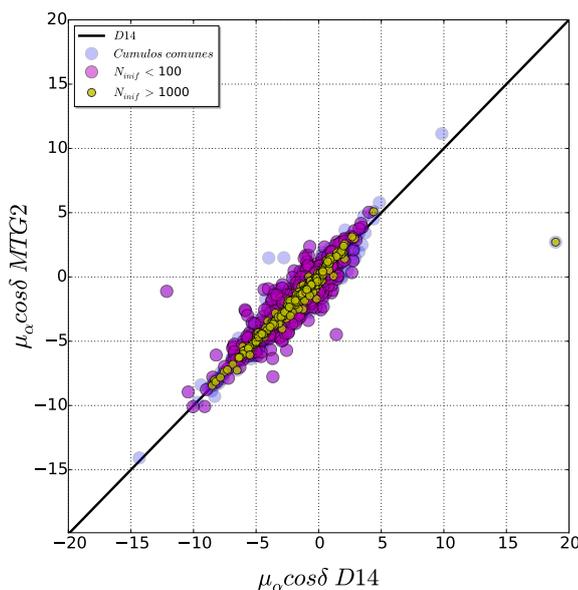


Figura 5.27: Centroides obtenidos por MTG2 frente a los publicados en D14 en función de N_{inif} . La muestra de los cúmulos se representa en azul, superponiendo en magenta aquellos cúmulos en los que N_{inif} es inferior a 100 y en amarillo en los que es superior a 1000. Se puede observar que la gran mayoría de los cúmulos con una mayor diferencia en la determinación del centroide se corresponden con cúmulos con un número bajo de N_{inif} , exceptuando uno, Melotte 22.

cúmulos), siendo MTG4 la que menos (1014) debido al bajo número de resultados válidos obtenidos por este análisis de pertenencia (ver tabla 5.3). Se ha obtenido el *Porcentaje 1* más elevado de un 4.8 % de cúmulos, encontrado un *Porcentaje2* del 2.2 % para MTG4, llegando a ser del 1% para MT4 y MT2. Por tanto, exceptuando el caso de la metodología MTG4, para el resto de metodologías la mayoría de los cúmulos que presentan unas diferencias en la determinación del centroide superiores a 3σ 's se corresponden a muestras con $N_{inif} < 100$ estrellas.

De los cúmulos referidos al *Porcentaje 2* (puntos rojos en la figura 5.26), cuatro son comunes a los cuatro análisis de referencia que son Melotte 22, FSR 0498, NGC 1664, Dutra Bica 12, y seis son comunes a tres de ellos, siendo estos cúmulos Markarian 38, Kronberger 4, FSR 0324, Berkeley 39, FSR 1153 y FSR 0869. Exceptuando Melotte 22, para el resto de los cúmulos, los radios utilizados por D14 y los empleados en nuestros análisis son diferentes. Una inspección visual de los perfiles de densidad radial para estos cúmulos muestran que nuestros radios ajustan mejor al PDR observado, exceptuando el cúmulo NGC 1664 donde consideramos que hemos sobre-estimado el valor del radio. Para este cúmulo se realizarán de nuevo los análisis de pertenencia determinando un radio que se ajuste mejor al perfil de densidad radial observado.

Curiosamente, cúmulos en los que sabemos que los análisis de pertenencia realizados no son los correctos por encontrarse el cúmulo y el campo en regiones bien diferenciadas del VPD, no han mostrado diferencias superiores a 3σ 's respecto a los resultados en D14. Estos cúmulos son Blanco 1 y Stock 2. El caso contrario es el cúmulo Melotte 22. Este cúmulo, con coordenadas aproximadas en

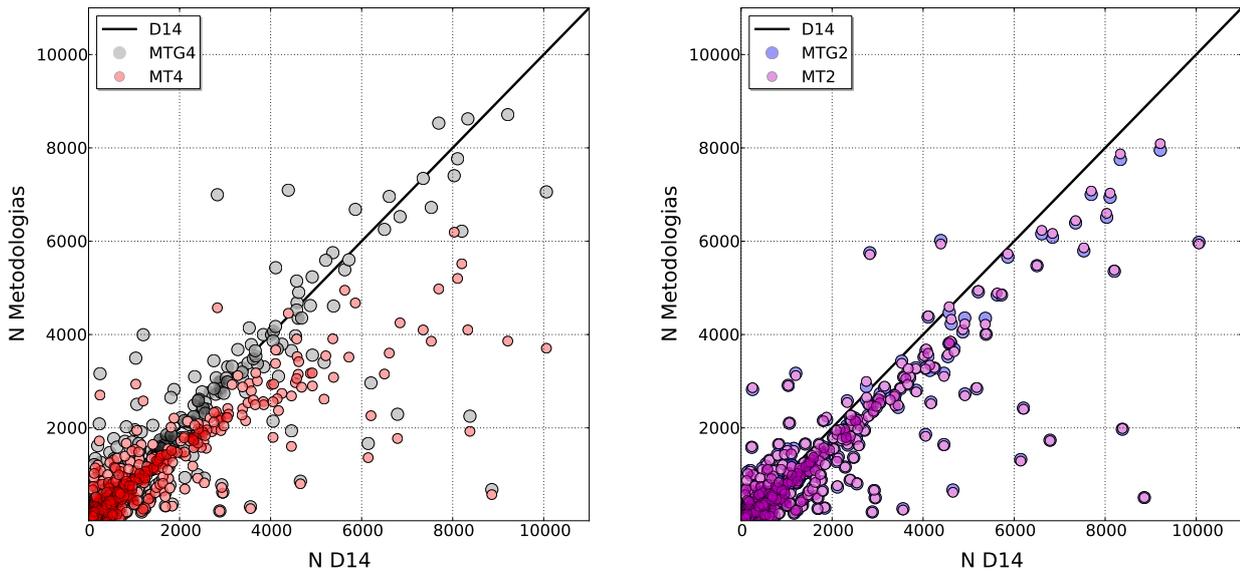


Figura 5.28: Comparación del número de miembros obtenidos por las metodologías aplicadas en esta tesis en función del número de miembros obtenidos por D14. En el gráfico izquierdo se representan las metodologías MTG4 en negro y MT4 en rojo. En el gráfico derecho se representan las metodogías MTG2 en azul y MT2 en magenta. Se observa que precisamente MTG4 es la que tiene una mayor coincidencia en el número de miembros determinados por D14.

los gráficos de la figura 5.26 de $(-16, 41)$ mas/año, no ha presentado diferencias superiores a 3σ 's en la determinación de su centroide por parte de las metodologías aplicadas en esta tesis, indicando que los análisis realizados no son los correctos, es decir, las metodologías han determinado una región del campo como cúmulo (mismo comportamiento que el ya descrito para el cúmulo NGC 2632).

En la figura 5.28 se compara el número de miembros obtenidos por las metodologías con respecto a los publicados por D14. En el gráfico de la izquierda se hace la comparativa respecto a las metodologías MTG4 y MT4 y, en el gráfico de la derecha respecto MTG2 y MT2. Curiosamente es MTG4 la que determina un número de miembros más similar al obtenido por D14. En general, las metodologías que usan movimientos propios en el análisis determinan un menor número de miembros conforme el número de los mismos aumenta, en comparación con los obtenidos por D14.

La comparación de las dispersiones en el espacio de movimientos propios se muestra en la figura 5.29. A modo de ejemplo se representan las dispersiones obtenidas por MTG2 y MT4 en función de las publicadas en D14. Como se puede observar, las dispersiones relativas a los análisis de pertenencia realizados en esta tesis son, para muchos cúmulos, superiores a las obtenidas por el trabajo D14, fundamentalmente cuando las posiciones de las estrellas se incluyen en el análisis. La explicación a este comportamiento recae en el análisis de pertenencia realizado en D14. En su análisis, ellos tuvieron en cuenta los errores en movimientos propios de las estrellas en el cálculo de las probabilidades de pertenencia. Con lo cual, las dispersiones publicadas en D14 se corresponderían con las dispersiones de los cúmulos una vez abstraídos los errores de UCAC4. Sin embargo, estas dispersiones no se corresponden con las dispersiones intrínsecas de los cúmulos ya que diferentes efectos, como la exis-

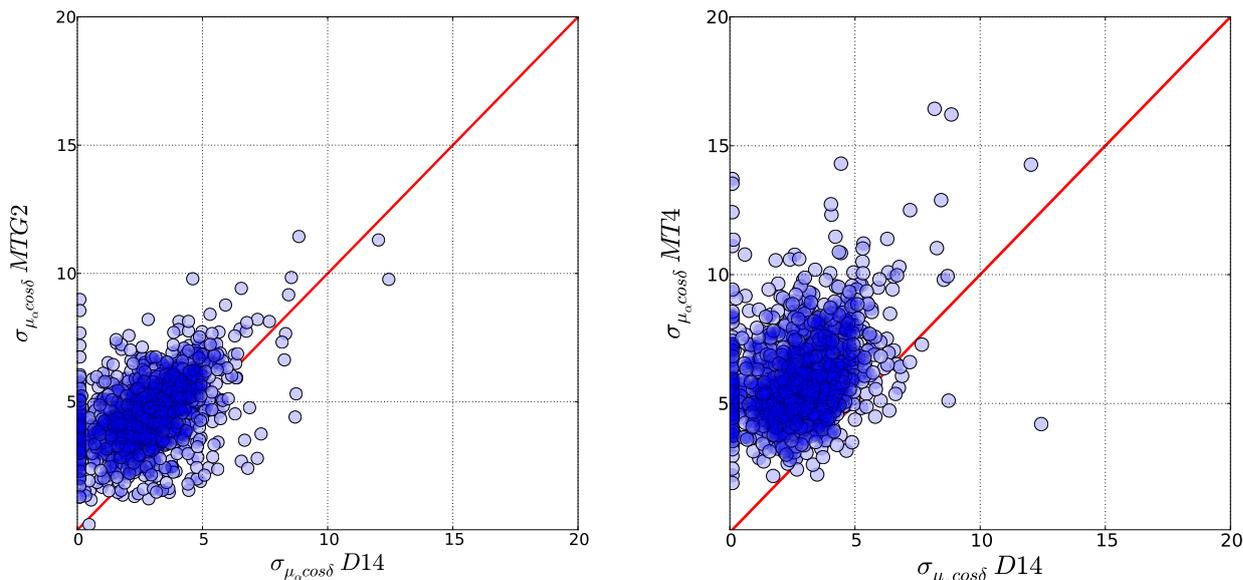


Figura 5.29: Dispersiones en el espacio de movimientos propios obtenidas por MTG2 y MT4 (gráficos izquierdo y derecho, respectivamente), respecto a las obtenidas por D14. La introducción de las posiciones revierte en un mayor valor de la dispersión en el espacio de movimientos propios.

tencia de sistemas binarios o múltiples, hacen que la dispersión del cúmulo sea del orden de los errores obtenidos. Si a las dispersiones obtenidas por nuestros análisis les eliminamos los errores presentes en UCAC4 (utilizando la moda de la distribución de los errores de unos 4 mas/año), se observa en la figura 5.30, una mayor correlación respecto a los resultados de D14 con las metodologías que usan únicamente movimientos propios (gráfico izquierdo). La introducción de las posiciones sigue dando lugar a un mayor valor de la dispersión obtenida (gráfico derecho).

5.4.2 Comparación con el catálogo DAML02.

La comparación con los resultados publicados en el catálogo DAML02 se muestran en la figura 5.31 y en la tabla 5.4. Se observa una mayor dispersión en la diferencia entre los centroides en movimientos propios con respecto a las obtenidas en comparaciones previas. Este resultado se debe a que DAML02 no es un catálogo homogéneo en el que se estén usando las mismas bases de datos o incluso las mismas variables y metodologías para realizar los análisis de pertenencia. A pesar de este hecho, el *Porcentaje* \geq sólo supera el 5% para MTG4.

La comparación con DAML02 reporta resultados interesantes. Aunque la comparación puede resultar complicada al estar utilizando diferentes bases de datos y diferentes análisis de pertenencia en el que variables como la fotometría pueden estar incluidas (Kharchenko et al., 2005), sí permite detectar para qué cúmulos los resultados obtenidos por las metodologías aplicadas en esta tesis presentan resultados muy diferentes a aquellos publicados por diferentes autores.

Dado el elevado número de cúmulos con diferencias en el centroide superiores a 3σ 's, se decidió

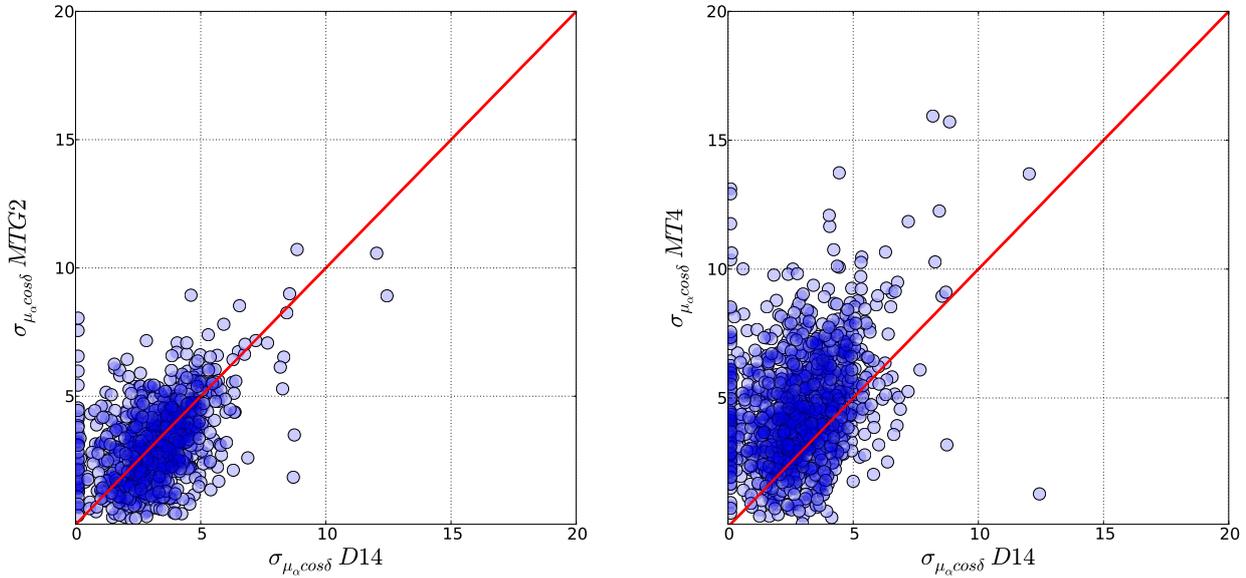


Figura 5.30: Dispersiones en el espacio de movimientos propios obtenidas por MTG2 y MT4 (gráficos izquierdo y derecho, respectivamente), respecto a las obtenidas por D14, una vez sustraídos los errores de UCAC4. Observamos que nuestras dispersiones tienen correlación con las obtenidas por este trabajo (gráfico izquierdo). La introducción de las posiciones revierte en un mayor valor de la dispersión en el espacio de movimientos propios (gráfico derecho)

Tabla 5.4: La tabla muestra los valores de los centroides y de las dispersiones, en mas/año, de las diferencias en la determinación de los centroides en el espacio de movimientos propios a 1σ ($\sigma_{\Delta\mu_\alpha \cos\delta}$, $\sigma_{\Delta\mu_\delta}$), obtenidas entre las metodologías y el trabajo DAML02, el número de miembros comunes (N_0) y porcentajes de cúmulos (respecto del número de cúmulos comunes) que, situados a más de 3σ 's, tienen menos (*Porcentaje 1*) o más (*Porcentaje 2*) de 100 objetos en la muestra a analizar.

Metodologías	DAML02 -MTG2	DAML02-MTG4	DAML02-MT4	DAML02-MT2
$\Delta\mu_\alpha \cos\delta$	0.202	0.238	0.212	0.152
$\Delta\mu_\delta$	-0.112	-0.135	-0.076	-0.132
$\sigma_{\Delta\mu_\alpha \cos\delta}$	1.873	1.966	2.030	1.946
$\sigma_{\Delta\mu_\delta}$	1.738	1.745	1.929	1.845
N_0	1709	1051	1674	1570
<i>Porcentaje 1</i>	3.1%	2.0%	3.0%	3.3%
<i>Porcentaje 2</i>	4.0%	5.5%	3.6%	4.3%

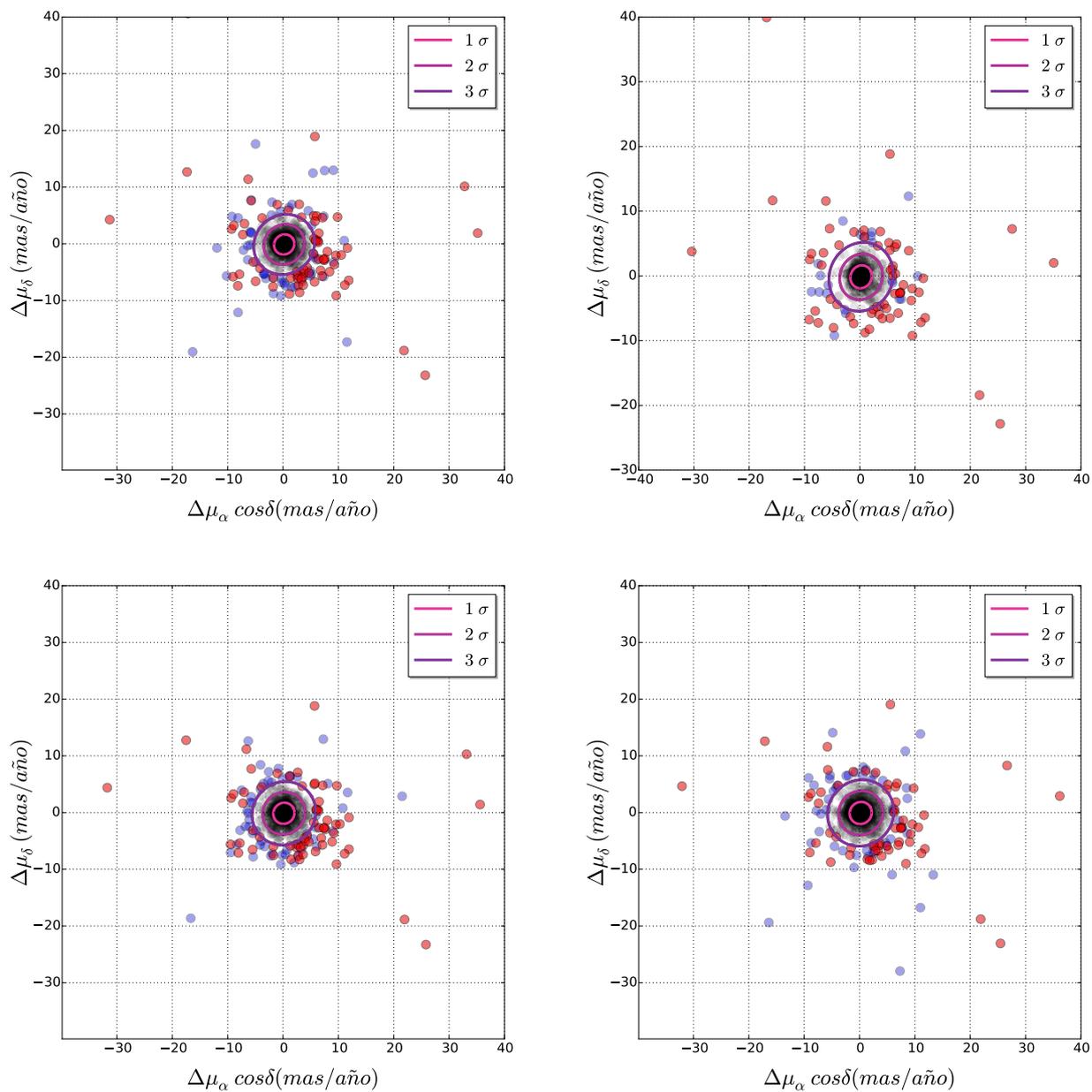


Figura 5.31: Diferencias en los centroides en el espacio de movimientos propios entre las metodologías MTG2 (gráfico izquierdo superior), MTG4 (gráfico derecho superior), MT2 (gráfico izquierdo inferior) y MT4 (gráfico derecho inferior) respecto a los valores de los centroides listados en el catálogo DAML02.

estudiar todos aquellos cúmulos comunes relacionados con el *Porcentaje 2*. Obtenemos un total de 53 cúmulos a analizar. Para más de la mitad de estos cúmulos no ha sido observada su existencia en el espacio de movimientos propios, siempre teniendo en cuenta que la base de datos que estamos utilizando es UCAC4. Ejemplos son ASCC 24, ASCC 74 o NGC 2489. Alrededor de un 30% presentan una elevada contaminación de estrellas de campo y es imposible determinar su existencia haciendo uso de sus posiciones y de sus movimientos propios. Ejemplos de estos cúmulos son: NGC 3766 o ASCC 66. Para unos pocos cúmulos la sobre-densidad asociada al cúmulo es observada en el espacio de movimientos propios, pero la elevada contaminación de estrellas de campo imposibilita su correcta determinación. Ejemplos de estos cúmulos son: Casado Alessi 1, Alessi 20, Collinder 135, NGC 2925, Stock1, Stock 12. Dos cúmulos, Trumpler 17 y Feigelson 1 presentan un bajo número de miembros listados en DAML02, 1 y 10, respectivamente. De los cúmulos con $N_{inif} < 100$ estrellas, no se ha observado su existencia en el espacio de movimientos propios para ninguno. Ejemplos de estos cúmulos son: Mamajek 2, Berkeley 45, Berkeley 43 ó Teutsch 50.

De los 2167 cúmulos listados en DAML02, 2102 tienen publicados centroides en el espacio de movimientos propios. De los 65 restantes, hemos obtenido 19 nuevos centroides. La metodología MTG2 ha obtenido 15 nuevos centroides, MTG4 7, MT4 18 y MT2 18. Estos cúmulos están listados en la tabla 5.5. Se observa que, de estos 19 cúmulos, aquellos con mayores valores de N_{inif} presentan una mayor concordancia en los valores de los centroides en movimientos propios obtenidos por los diferentes análisis.

La comparación realizada con los trabajos de D14 y DAML02 ha permitido detectar diez cúmulos Melotte 22, Blanco 1, NGC 2632, Mamajek 1, IC 2391, NGC 7092, Sock 2, Alessi 13, NGC 752 y Trumpler 10 en los que nuestros análisis de pertenencia no han hecho una correcta clasificación de las estrellas. Para todos estos cúmulos se observa una clara separación entre ambas poblaciones, siendo necesario realizar unos nuevos análisis de pertenencia antes de la publicación de los resultados.

Tabla 5.5: Número de objetos analizados y, centroides y dispersiones en movimientos propios en mas/año de los 19 cúmulos inéditos obtenidos.

Cúmulos	N_{inif}	MTG2			MTG4			MT4			MT2		
		$\mu_{\alpha} \cos \delta$, $\sigma_{\mu_{\alpha} \cos \delta}$	μ_{δ} , $\sigma_{\mu_{\delta}}$	μ_{δ} , $\sigma_{\mu_{\delta}}$	$\mu_{\alpha} \cos \delta$, $\sigma_{\mu_{\alpha} \cos \delta}$	μ_{δ} , $\sigma_{\mu_{\delta}}$	μ_{δ} , $\sigma_{\mu_{\delta}}$	$\mu_{\alpha} \cos \delta$, $\sigma_{\mu_{\alpha} \cos \delta}$	μ_{δ} , $\sigma_{\mu_{\delta}}$	μ_{δ} , $\sigma_{\mu_{\delta}}$	$\mu_{\alpha} \cos \delta$, $\sigma_{\mu_{\alpha} \cos \delta}$	μ_{δ} , $\sigma_{\mu_{\delta}}$	
AH03 J1725 34.4	23	-3.403, 1.264	-0.743, 1.051	-	-	-	-4.060, 5.401	-0.331, 4.533	-	-3.46, 1.240	-	0.791, 1.008	
Alessi 52	146	-5.783, 5.828	0.080, 4.541	-5.678, 6.261	-0.095, 5.290	-	-5.576, 5.07	0.063, 4.273	-	-5.861, 5.76	-	0.150, 4.430	
BH 208	48	-	-	-	-	-	-5.252, 8.108	-1.427, 7.604	-	-8.441, 1.848	-	-2.369, 6.590	
DC 1	80	0.764, 1.676	-0.339, 1.714	0.478, 4.997	1.026, 4.594	-	0.905, 4.542	-0.977, 4.290	-	0.360, 1.959	-	-0.539, 1.989	
Dutra Bica 12	306	1.270, 4.725	-3.82, 5.074	1.440, 8.076,	-3.408, 8.376	-	0.758, 7.782	-3.656, 8.462	-	1.228, 4.720	-	-3.814, 5.294	
FSR 0647	31	-2.638, 2.391	0.701, 3.951	-	-	-	-0.769, 13.527	-1.947, 7.700	-	-1.905, 6.609	-	0.362, 5.238	
FSR 0696	23	-	-	-	-	-	-7.274, 4.390	-3.138, 3.389	-	-	-	-	
FSR 0763	28	-	-	-9.000, 2.400	-5.300, 1.500	-	-12.677, 6.521	-4.592, 4.791	-	-9.326, 2.097	-	-2.667, 0.785	
FSR 0814	45	0.422, 2.256	-3.359, 2.045	-	-	-	0.683, 4.493	-3.399, 4.127	-	-0.899, 1.383	-	-5.389, 0.522	
FSR 0828	183	0.967, 3.533	-3.491, 3.292	1.565, 6.855	-3.925, 5.655	-	1.194, 4.831	-3.388, 5.07	-	0.959, 3.358	-	-3.498, 3.522	
FSR 1343	24	-	-	-7.339, 3.335	-3.479, 3.276	-	-3.299, 4.418	3.843, 3.961	-	-10.373, 0.452	-	-3.163, 2.083	
Juchert 10	79	-4.654, 4.591	-5.369, 4.183	-	-	-	-4.285, 5.001	-4.995, 5.130	-	-4.890, 2.048	-	-5.438, 3.509	
Kronberger 23	81	-0.923, 3.960	-0.445, 2.838	-	-	-	-0.785, 5.738	0.204, 5.327	-	-0.999, 4.112	-	-0.547, 2.794	
Majaess 30	51	0.311, 3.521	-3.274, 3.277	-	-	-	-1.378, 4.901	-2.601, 4.280	-	-0.710, 2.265	-	-3.722, 2.198	
Majaess 9	112	-2.321, 4.663	2.510, 3.527	-	-	-	-2.492, 3.728	1.632, 2.942	-	-2.324, 4.090	-	2.537, 3.549	
Mamajek 4	9366	-1.193, 5.575	-4.276, 6.176	-1.272, 6.770	-4.631, 7.595	-	-1.090, 3.255	-3.985, 3.660	-	-1.128, 5.619	-	-4.227, 6.417	
Ruprecht 147	648	-1.482, 4.592	-3.058, 5.115	-1.516, 6.177	-4.043, 7.899	-	-1.761, 5.456	-3.199, 5.803	-	-1.380, 4.825	-	-2.711, 4.963	
Teutsch 127	38	-1.437, 2.294	-2.863, 1.598	-	-	-	-0.257, 4.892	-3.704, 4.788	-	-3.389, 4.403	-	-1.988, 0.516	
Wit 3	40	0.176, 3.403	-3.887, 5.080	-0.875, 2.989	-7.952, 4.099	-	1.323, 7.175	-3.812, 5.413	-	0.159, 1.540	-	-2.608, 5.712	

6

Conclusions

6.1 Main conclusions of this PhD thesis.

- In this thesis, it has been presented a new methodology for the identification of (potential) members in stellar open clusters (Sampedro & Alfaro, 2016). This technique differs, from other already existing approaches, in its capability of combining, in a self-consistent manner, N-dimensional samples of physical variables. By construction, the method does not take into account the intrinsic nature of these N-variables. It only assumes the hypothesis that they will be more densely distributed for the cluster members than for the field stars.
- A number of realistic simulations have been designed and developed, spanning an ample set of physical parameters, to test and characterize the robustness and applicability of this new technique.
- In order to contextualize the quality of this new methodology, we have compared its results with those obtained from other approaches on the very same simulations. Several figures of merit have been defined to unbiasedly compare the goodness of the different methodologies.
- From these analysis, it has been demonstrated that this new N-Dimensional methodology successfully reaches similar (or even better) results than those obtained from other similar techniques, whenever the starting hypothesis of having different density distributions for cluster members and field stars is hold.
- We have analyzed the open clusters in the DAML02 catalogue (Dias et al., 2002) using the positions and the proper motions provided by the UCAC4 catalogue (Zacharias et al., 2013). A careful visual inspection of these catalogues led to the conclusion that the determination of the cluster radii is indeed a mandatory and complicated task that must be addressed previous to any membership analysis. In this thesis we propose a new functional method for the estimation of cluster radius.
- An observed inhomogeneous depth in the aforementioned catalogues (for the proper motions

and positions), made the application of the geometrical method unfeasible (when including all the four variables) since the starting hypothesis could not be met.

- The four methodologies discussed in this thesis, for the determination of membership probabilities, have been applied to a final sample of 1876 clusters from the DAML02 catalogue using the UCAC4 data. It was only possible to obtain simultaneous results for all the methodologies for a subsample of 982 clusters. Likewise, MTG2 yielded results for 1724 clusters, MT2 for 1585, MTG4 for 1058 and MT4 for 1692.
- A comparative analysis of these results, in terms of the first and the second moments (mean and dispersion) of the distributions, showed that the methodologies were highly coincident when the same number of variables was used. Including positional variables in the analysis led to an increment in the estimated dispersion of the proper motions in the clusters. This fact was even worst for the MTG4. In the meantime, we noticed that beyond a certain threshold in the number of initial objects, $N_{inif} > 3000$, MT4 was systematically underestimating both the number of cluster members and the dispersions in the proper motions when compared to MTG2 and MT2.
- The results obtained from these methodologies have been also compared to those from D14 and DAML02. On the one hand, we observed that the distributions of the differences for the cluster centroids between D14 and our methodologies were larger than those obtained between the four methodologies. The largest observed dispersion was about $\sigma \sim 1$ mas/year. The largest differences in the centroids were related with the initial values $N_{inif} < 100$. On the other hand, the comparison with DAML02 reported dispersions in the distributions of the differences for the cluster centroids of $\sigma \sim 2$ mas/year. The comparison with these works, specially with DAML02, has served to the identification of a list of clusters with unexpectedly large discrepancies between results (Melotte 22, Blanco 1, NGC 2632, Mamajek 1, IC 2391, NGC 7092, Stock 2, Alessi 13, NGC 752 y Trumpler 10). These disagreements mainly correspond to failed solutions due to incorrect guess values for the analysis of the clusters.
- These results will be published soon in the format of two catalogues (Sampedro et al. in preparation). A catalogue solely including the parameters describing the fitted distributions (as much for the clusters as for the fields) either in the position as in the proper motion space. Basically, the number of members, the means, the dispersions and the correlation coefficients along with the number of stars classified as outliers, the new values for the radii (as defined in this thesis), the distances, the ages and the color excess extracted from the DAML02.
- An another catalogue for each cluster including the astronomical coordinates, the proper motions, the uncertainties, a flag for the identification of outliers, the membership probabilities for each of the four methodologies and a cluster/field classification of the stars. Non detected clusters (for any methodology) are set to -1.

6.2 Personal request.

During this thesis, I have had to work with different datasets published either in Vizier ¹, in specific surveys as GES (Gaia ESO Survey, Gilmore et al. 2012; Randich, Gilmore, & Gaia-ESO Consortium (2013)) or even accessing hand-made catalogues from individual authors. I noticed that the data format could had not been more heterogeneous, being necessary to devote a fraction of time to the development of software just to be able to automatically download the data and read it. Understanding that different databases may have different philosophies in terms of the most convenient format of the data they want to provide, me, as a user, I would request that in the future these databases tend to homogenize formats so the usage of the data become straightforward. I would specially emphasize the usefulness of looking for standards in the definition of "non-data". Blank spaces or random symbols make the user to waste time and energy. In this sense, I would like to acknowledge the effort carried out by the Virtual Observatory (VO) in this direction and I hope in the near future all the astronomical databases would be easily accessible via VO-like tools like Topcat².

6.3 Personal conclusion.

This thesis has represented to me many years of hard work. During this time I got used to using several techniques for the membership analysis of stars in open clusters. I learnt all the associated problems to these analysis, the typical properties and biases of real data in large databases along with the inherent complexity in the definition and the study of stellar open clusters. This work leaves behind one of the few catalogues for open clusters available in the literature encompassing the results from several methodologies in a systematic and homogeneous fashion. The Gaia mission is not expected to deliver the final catalogues until 2021 and it will still take 2-3 years to reach the expected high precision of its data (Michalik et al., 2014). Therefore, while the Era of Gaia becomes true, I hope the ideas, the methodologies and the catalogues presented in this thesis may serve to contribute to the on-going understanding of the physical properties of open clusters in our Milky Way.

¹<http://vizier.u-strasbg.fr/viz-bin/VizieR>

²<http://www.star.bris.ac.uk/mbt/topcat/>

Conclusiones

En esta sección se detallan las principales conclusiones obtenidas a lo largo de esta tesis.

7.1 Principales conclusiones de esta tesis.

- En esta tesis se ha desarrollado una nueva metodología (Sampedro & Alfaro, 2016) para la determinación de los miembros potenciales de un cúmulo estelar a partir de un conjunto de N -variables físicas, medidas sobre las estrellas en la vecindad del cúmulo. El método es flexible en el número de variables a utilizar, no teniendo en cuenta la naturaleza de las mismas siempre y cuando se cumpla la hipótesis fundamental, de que estén más densamente concentradas para los miembros del cúmulo que para las estrellas de campo.
- Hemos diseñado, desarrollado y realizado diferentes tipos de simulaciones, para diferentes conjuntos de parámetros, que han permitido analizar la potencialidad y rango de aplicación de la nueva metodología propuesta.
- Hemos comparado esta nueva técnica con otras metodologías previas usando las mismas simulaciones, y definiendo figuras de mérito que permiten un análisis cuantitativo del comportamiento de los diferentes métodos.
- De este análisis podemos concluir que la nueva metodología se comporta igual o mejor que las previamente diseñadas, siempre que verifiquen las hipótesis de partida, es decir, que las estrellas de la población de cúmulo estén más densamente concentradas en ese espacio de variables que las estrellas de campo.
- Analizamos los cúmulos del catálogo Dias et al. (2002) (DAML02), a partir de los datos de posición sobre el cielo y movimientos propios que están en el catálogo UCAC4 (Zacharias et al., 2013). La inspección detallada de estos cartografiados nos indica que la determinación del radio del cúmulo es una tarea complicada, necesaria y previa a cualquier análisis de pertenencia que queramos hacer. Se propone un método funcional de búsqueda del radio del cúmulo.

- La distinta completitud alcanzada en la catalogación de las variables espaciales y de movimientos propios en este cartografiado hacen que el método geométrico, incluyendo cuatro variables, no pueda utilizarse en todos los casos debido a que los datos no verifican las hipótesis de trabajo.
- Se aplicaron los cuatro métodos de determinación de probabilidad de pertenencia a 1876 cúmulos listados en el catálogo DAML02 usando datos del cartografiado UCAC4. Para 982 cúmulos se obtuvieron resultados con los cuatro métodos. MTG2 proporcionó resultados válidos para 1724 cúmulos, MT2 para 1585, MTG4 para 1058 y MT4 para 1692.
- Un análisis comparativo muestra que los resultados de los primeros y segundos momentos de las distribuciones espaciales y en el espacio de movimientos propios, son bastante coincidentes para el mismo número de variables utilizadas. La inclusión de las variables posicionales en el análisis incrementa el valor de la solución para la dispersión de movimientos propios de los cúmulos. Este hecho, se acentúa para MTG4. A partir de un cierto valor del número de objetos iniciales, $N_{inif} > 3000$, el método MT4 tiende a proporcionar valores del número de miembros y de la dispersión de movimientos propios inferiores a los obtenidos por MTG2 y MT2.
- Los resultados obtenidos se compararon con los trabajos de D14 y DAML02. Respecto a D14, las distribuciones de las diferencias en los centroides obtenidas por este trabajo con respecto a nuestros análisis de pertenencia son mayores que las obtenidas entre nuestras metodologías. Aun así, la máxima dispersión encontrada es $\sigma \sim 1$ mas/año. Las mayores diferencias en los centroides se corresponden con valores iniciales, $N_{inif} < 100$. La comparación con DAML02 reportan unos valores en las dispersiones de las diferencias entre centroides mayores siendo $\sigma \sim 2$ mas/año. La comparación con estos trabajos, fundamentalmente con DAML02, nos ha permitido encontrar cúmulos (Melotte 22, Blanco 1, NGC 2632, Mamajek 1, IC 2391, NGC 7092, Stock 2, Alessi 13, NGC 752 y Trumpler 10) con una gran discrepancia en los resultados, que principalmente se deben a soluciones fallidas cuando los valores iniciales del proceso automatizado no son representativos de la realidad del cúmulo.
- Los resultados obtenidos serán publicados en forma de dos catálogos (Sampedro et al. en preparación). En uno de ellos se darán los parámetros que describen las distribuciones de las poblaciones de cúmulo y campo, en los espacios de posiciones y de movimientos propios (número de miembros, medias, dispersiones y coeficientes de correlación). Además, se listarán el número de *outliers* determinados, el radio de los cúmulos y, la distancia, edad y exceso de color extraídos del catálogo DAML02.
- Individualmente para cada cúmulo, se publicará otro catálogo con sus coordenadas, movimientos propios, errores, un indicador que tomará el valor de 1 si la estrella ha sido clasificada como un *outlier* o 0 en caso contrario, las probabilidades de pertenencia obtenidas por cada metodología, y unos indicadores de ceros y unos dependiendo de si los análisis de pertenencia han determinado que la estrella es de campo o de cúmulo, respectivamente. En el caso en el que alguna metodología no haya determinado cúmulo, se indicará con un -1.

7.2 Requerimiento personal.

A lo largo de mi tesis he tenido que trabajar con datos de diferentes fuentes publicados en Vizier¹ o provenientes de cartografiados como GES (Gaia ESO Survey, Gilmore et al. 2012; Randich, Gilmore, & Gaia-ESO Consortium (2013)) e incluso datos de cúmulos individuales publicados por diferentes autores. El formato de publicación de estos datos no puede ser más heterogéneo, teniendo que dedicar parte del tiempo en la realización de programas diferentes para la automatización en la descarga de los datos u obtener catálogos legibles por los programas a utilizar. Entiendo la diferente filosofía de las diferentes bases de datos, pero como usuaria de ellas, pediría que en lo sucesivo se intentara homogeneizar para que la descarga de los datos sea lo más sencilla posible. Indicar que propondría una forma común a todas las bases de datos para indicar la no existencia de datos. El uso de espacios en blanco o de símbolos aleatorios hacen al usuario perder tiempo y energía, que el formato en el que se encuentren sea el mismo o que al menos pueda serlo, o que tengas la seguridad de estar descargándote todos los datos, que se expongan o diseñen las herramientas necesarias para poder hacer una descarga recursiva de los mismos, que quede bien claro que variables se pueden o no utilizar y de que tabla, si hay que realizar cruces de tablas, que identificadores hay que usar.... En este sentido, querría agradecer el esfuerzo llevado a cabo por el Observatorio Virtual (VO) en esta dirección y espero que en un futuro cercano todas las bases de datos astronómicas sean fácilmente accesibles a partir de herramientas del VO como Topcat².

7.3 Conclusión personal.

Este trabajo me ha supuesto años de duro trabajo en los que me he familiarizado con varias técnicas utilizadas en los análisis de pertenencia de las estrellas a los cúmulos abiertos. He conocido de primera mano las características intrínsecas de los análisis de pertenencia en sí mismos, las propiedades de los datos reales presentes en grandes cartografiados y, lo más importante, la propia problemática y definición de los cúmulos abiertos estelares. Con mi trabajo dejo uno de los pocos catálogos existentes de cúmulos abiertos, en el que diferentes análisis de pertenencia han sido realizados de una manera sistemática y homogénea. La publicación del catálogo final con los datos aportados por el satélite Gaia no será hasta el año 2021 y, hasta dentro de 2-3 años (Michalik et al., 2014) no se alcanzará la precisión esperada. Por tanto, mientras la era Gaia llega, espero que las ideas, metodologías y el catálogo presentado en esta tesis ayuden en los futuros estudios que utilicen de los cúmulos abiertos de nuestra Galaxia para su realización.

¹<http://vizier.u-strasbg.fr/viz-bin/VizieR>

²<http://www.star.bris.ac.uk/mbt/topcat/>

8

Trabajo Futuro

En esta sección se detallarán las principales líneas a seguir durante los próximos años:

- Uno de los principales objetivos a corto plazo es publicar el catálogo de los cúmulos analizados con las diferentes metodologías durante esta tesis. La idea fundamental es poner a disposición de la comunidad científica estos análisis con el objetivo de que puedan suponer un avance en las investigaciones que necesiten de los miembros de los cúmulos abiertos que hemos analizado de nuestra Galaxia. Previamente a esta publicación se realizarán unos nuevos análisis de pertenencia para los 10 cúmulos detectados en los que resultados no hacen una buena determinación de la población de cúmulo.
- Investigar la aplicación de la nueva metodología geométrica, desarrollada en esta tesis, a los datos que nos proporcionará la misión Gaia, objetivo para el cuál fue diseñada.
- Estudiar la ampliación de la aplicación de la metodología geométrica a otros sistemas estelares como los cúmulos globulares. En un principio se estudiarán las distribuciones de las variables que los describen para ver si es viable su aplicación.
- Actualmente, soy miembro del consorcio GES (Gaia ESO Survey). En este cartografiado, que proporcionará velocidades radiales de alta precisión, estoy involucrada en un proyecto de análisis de pertenencia en los cúmulos abiertos objetivos de este programa. Dada la gran flexibilidad de la metodología geométrica en el uso de diferentes conjuntos de variables, se analizará la membresía de los cúmulos en el espacio cinemático (velocidades radiales y/o movimientos propios), pudiendo introducirse además las posiciones de las estrellas en el análisis.
- Ampliar mis estudios a la utilización de la fotometría de las estrellas para la determinación de los miembros de los cúmulos. Estudiar como puedo introducir estas variables en los análisis y aprender a determinar, a partir de ellas, parámetros de los cúmulos como distancias, edades y metalicidades. Actualmente he pedido un proyecto con el objetivo de combinar la fotometría de los cartografiados S-PLUS+J-PLUS (PIs: Claudia M. Oliveira & Javier Cenarro). Si además

combinamos estos datos con los que nos proporcionará Gaia, se podría obtener uno de los catálogos con datos astrométricos y fotométricos más completos existentes de cúmulos abiertos.

- A partir de este último proyecto, estudios sobre la evolución química, segregación de masas y distribución espacial de los cúmulos, en función de la edad de los mismos, podrían ser objetivos a realizar en los próximos años.

A

Apéndice 1

A.1 Comportamiento de C y de M en función de la DC obtenidos para las metodologías aplicadas.

En las figuras A.1 - A.6 se muestran los modelos obtenidos para cada una de las metodologías: MT1, MTG1, MT2, MTG2, MT4 y MTG4, respectivamente. Las figuras de mérito C (gráficos superiores de cada una de las figuras) y M (gráficos inferiores), son representadas en función de la DC . En cada una de estas figuras, los gráficos de la columna de la izquierda hacen referencia a los resultados obtenidos por una metodología en las simulaciones referidas como *Sim1*, mientras que los gráficos de la columna de la derecha hacen referencia a los obtenidos en las *Sim2*.

Como se puede observar en las figuras, de una manera generalizada, la inclusión de los errores en función de la magnitud de las estrellas, produce peores resultados en la determinación de las dos figuras de mérito y, consecuentemente, en una peor diferenciación entre las poblaciones de cúmulo y campo simuladas. Ambos, el aumento de la DC como la introducción de un mayor número de variables, revierte en la obtención de unos mejores resultados. Se observa que los gráficos referentes a las simulaciones *Sim2*, presentan una menor estabilidad en la clasificación de la muestra en ambos grupos de estrellas, fundamentalmente para los valores más bajos de DC .

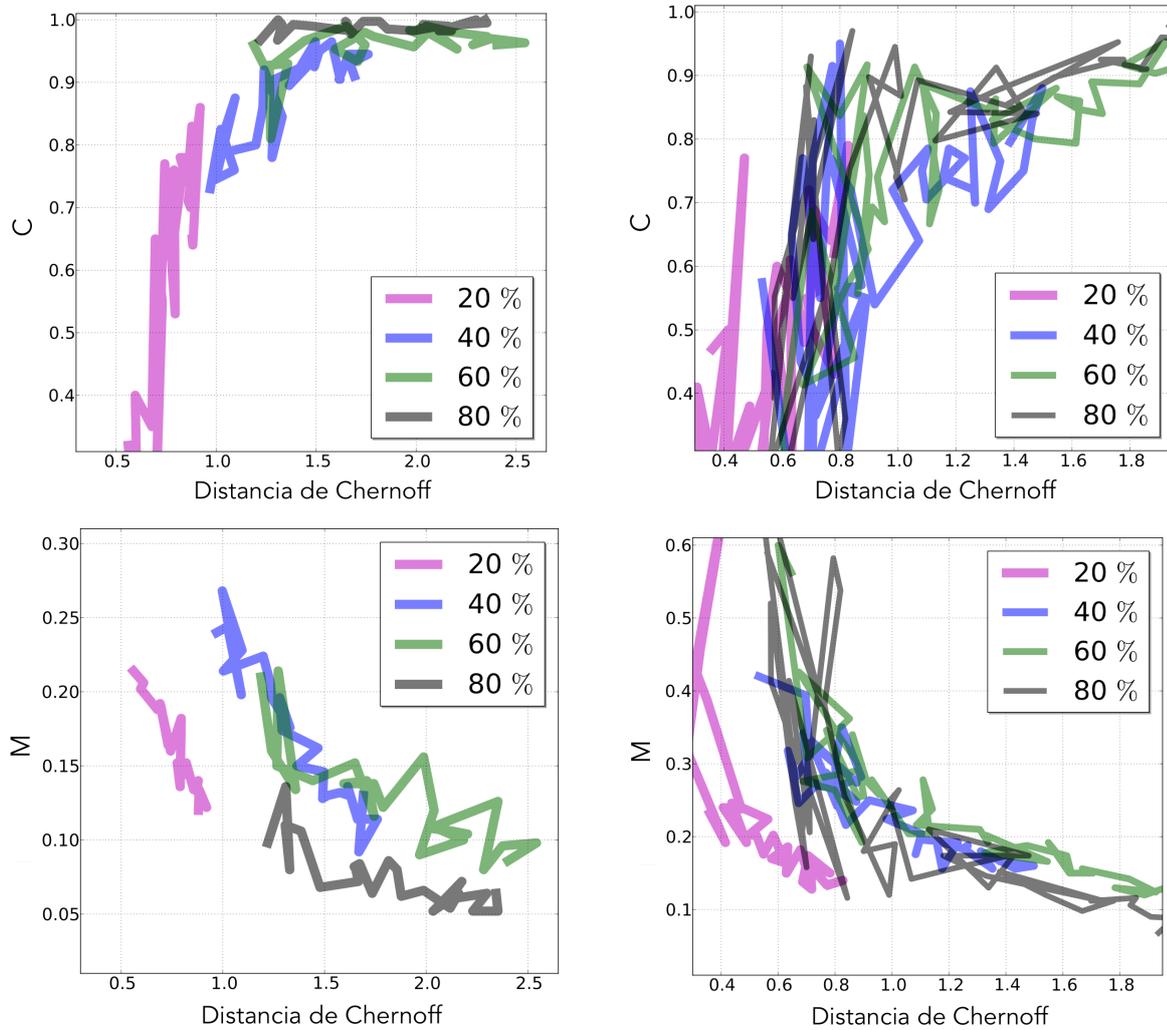


Figura A.1: Comportamiento de las figuras de mérito en función de la DC obtenidos por MT1. Los resultados obtenidos en las *Sim1* se muestran en la columna de la izquierda, mientras que los referidos a las *Sim2* se muestran en la columna derecha. Los colores hacen referencia al porcentaje de estrellas de cúmulo simuladas respecto del total de la muestra.

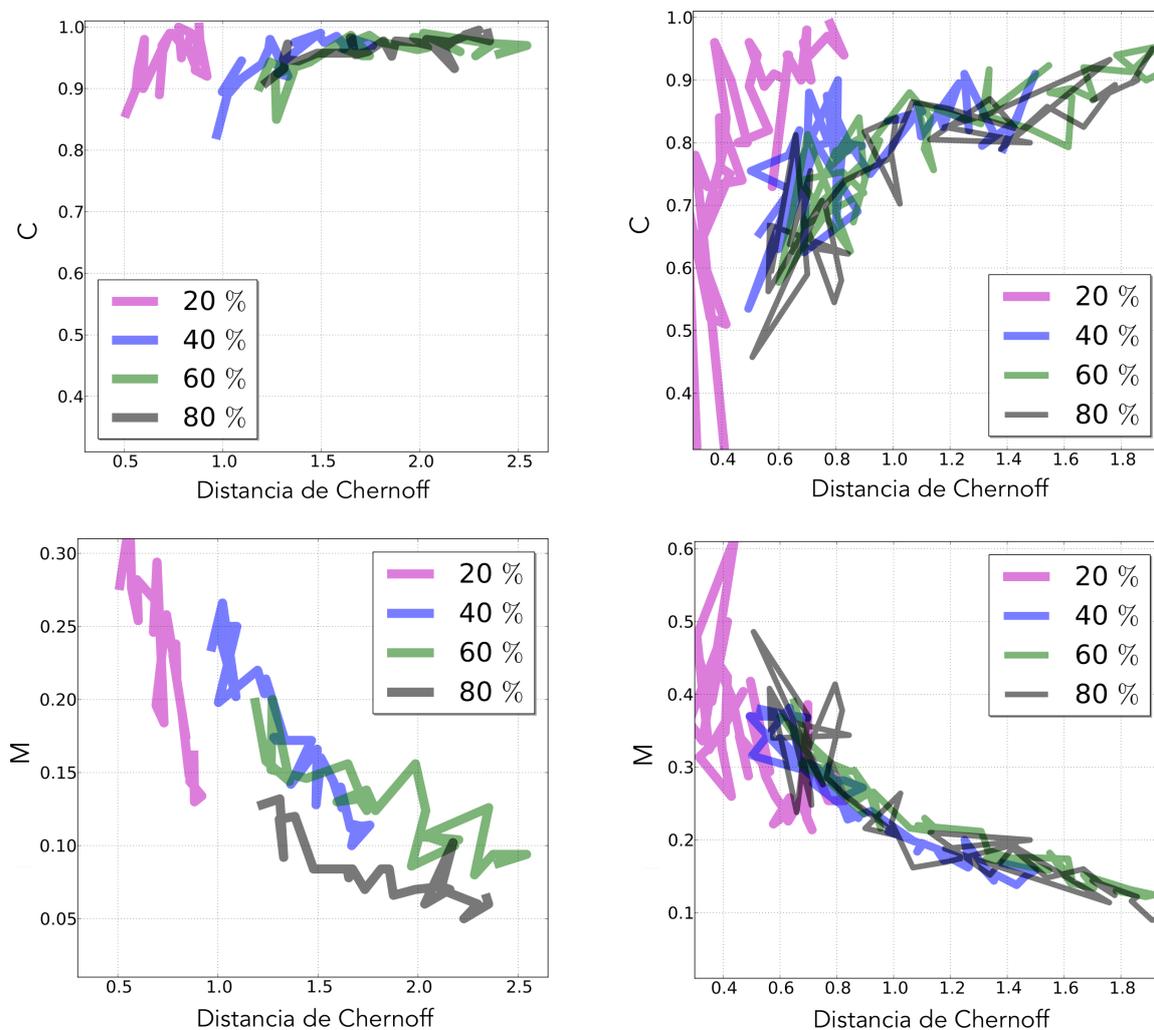


Figura A.2: Comportamiento de las figuras de mérito en función de la DC obtenidos por MTG1. Los resultados obtenidos en las $Sim1$ se muestran en la columna de la izquierda, mientras que los referidos a las $Sim2$ se muestran en la columna derecha. Los colores hacen referencia al porcentaje de estrellas de cúmulo simuladas respecto del total de la muestra.

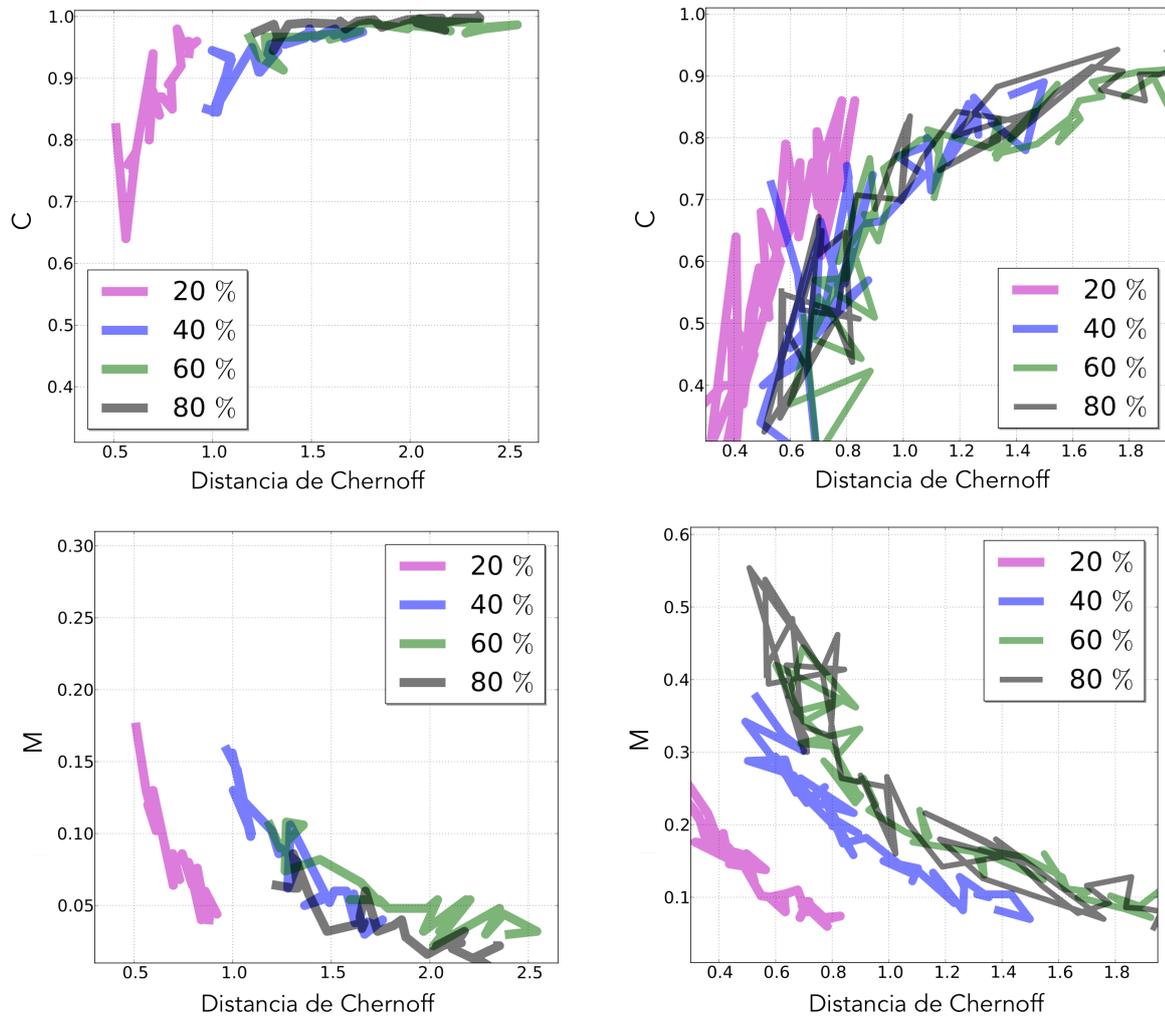


Figura A.3: Comportamiento de las figuras de mérito en función de la DC obtenidos por MT2. Los resultados obtenidos en las *Sim1* se muestran en la columna de la izquierda, mientras que los referidos a las *Sim2* se muestran en la columna derecha. Los colores hacen referencia al porcentaje de estrellas de cúmulo simuladas respecto del total de la muestra.

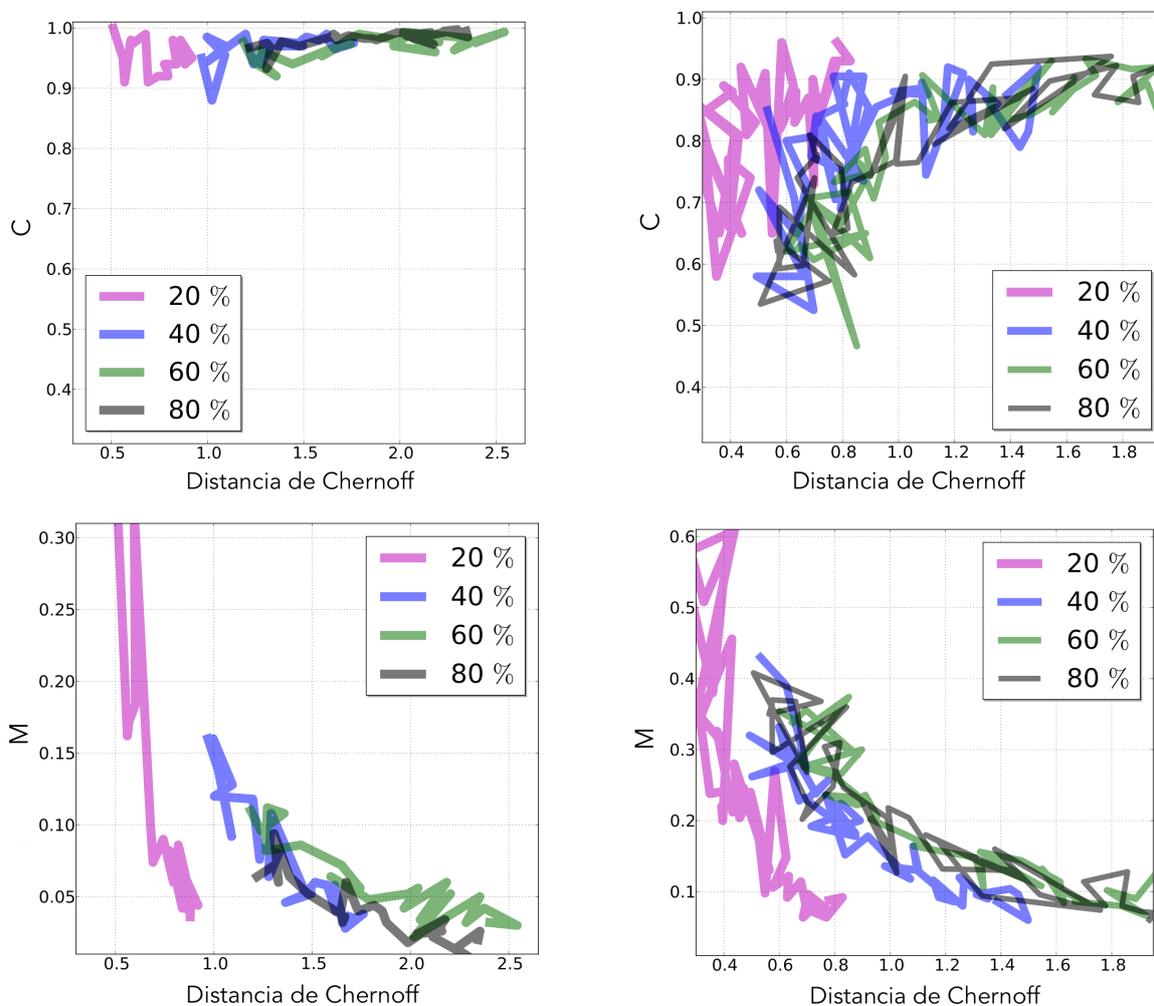


Figura A.4: Comportamiento de las figuras de mérito en función de la DC obtenidos por MTG2. Los resultados obtenidos en las $Sim1$ se muestran en la columna de la izquierda, mientras que los referidos a las $Sim2$ se muestran en la columna derecha. Los colores hacen referencia al porcentaje de estrellas de cúmulo simuladas respecto del total de la muestra.

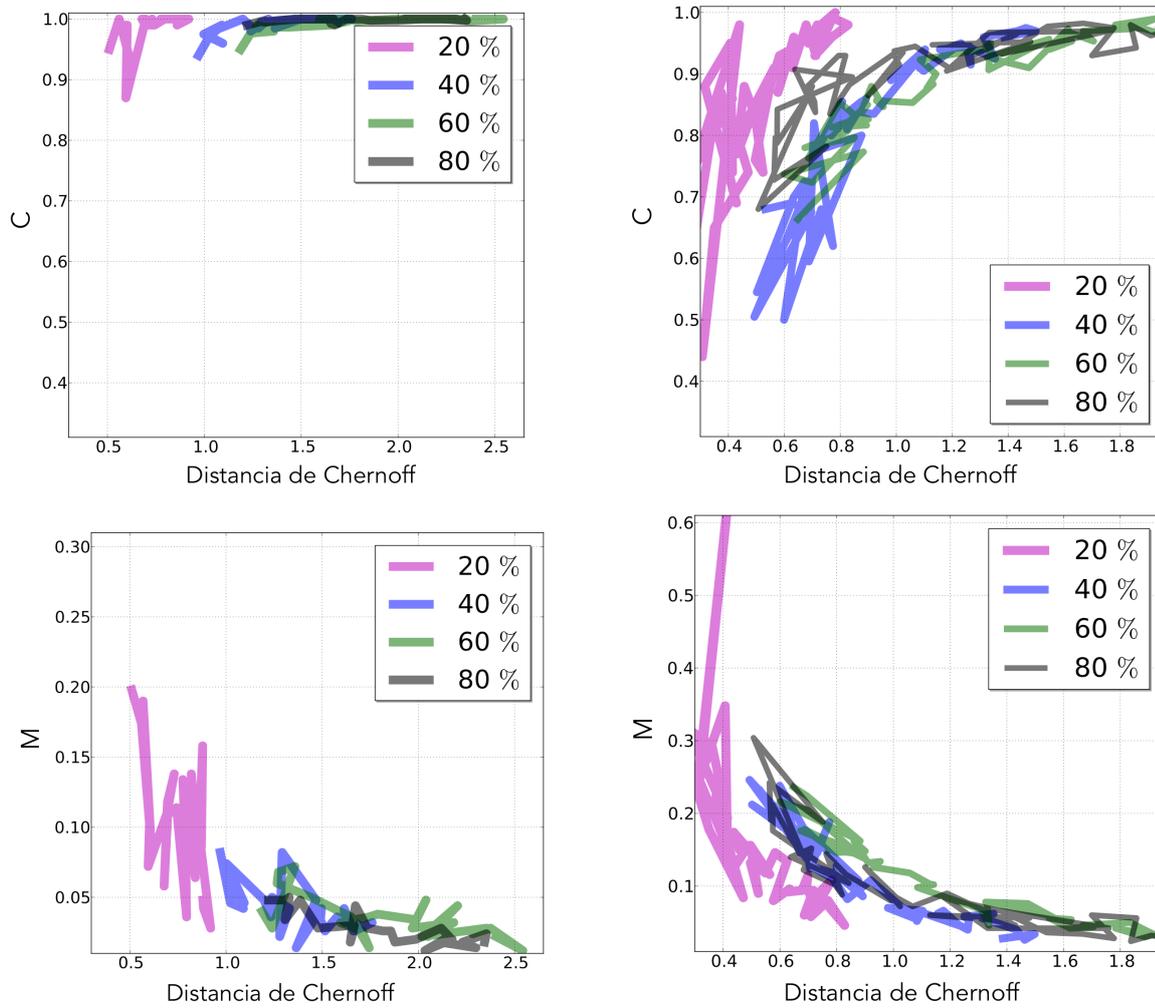


Figura A.5: Comportamiento de las figuras de mérito en función de la DC obtenidos por MT4. Los resultados obtenidos en las *Sim1* se muestran en la columna de la izquierda, mientras que los referidos a las *Sim2* se muestran en la columna derecha. Los colores hacen referencia al porcentaje de estrellas de cúmulo simuladas respecto del total de la muestra.

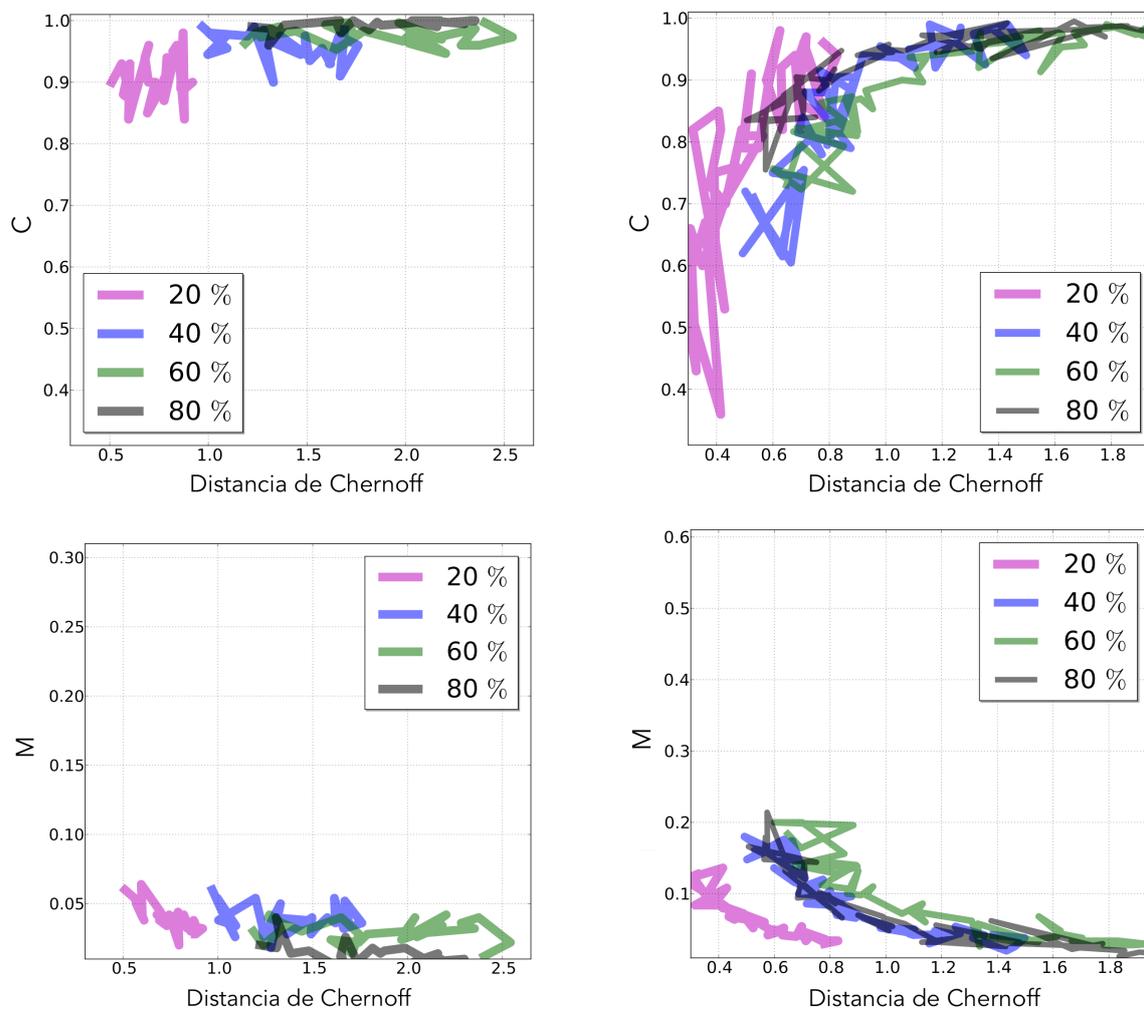


Figura A.6: Comportamiento de las figuras de mérito en función de la DC obtenidos por MTG4. Los resultados obtenidos en las $Sim1$ se muestran en la columna de la izquierda, mientras que los referidos a las $Sim2$ se muestran en la columna derecha. Los colores hacen referencia al porcentaje de estrellas de cúmulo simuladas respecto del total de la muestra.

B

Apéndice 2

B.1 Efectos del sub-muestreo.

La figura B.1 representa los efectos del sub-muestreo en la obtención de las figuras de mérito, C y M en función de DC , para las metodologías que usan $N = 1$ (gráficos superiores), 2 (gráficos intermedios) y 4 (gráficos inferiores) variables, en las simulaciones referidas como *Sim2*. Estos resultados se muestran de forma promedio para los cuatro porcentajes de estrellas de cúmulo simulados. Las líneas continuas hacen referencia a los resultados obtenidos por la nueva metodología geométrica (MTG1, MTG2 y MTG4), mientras que las líneas discontinuas hacen referencia a las diferentes metodologías aplicadas (MT1, MT2 y MT4), según el número de variables utilizado. Las zonas sombreadas hacen referencia a una dispersión de 1σ de los resultados obtenidos, utilizando un paso de 0.2 en unidades de distancia. Los colores hacen referencia a los diferentes sub-muestreos realizados respecto al total de la muestra.

Estos gráficos muestran una mejora en los resultados obtenidos tanto cuando la DC es mayor, como cuando el sub-muestreo presente en la muestra es menor (sub-muestreo del 50%). Se observa, en general, que los resultados obtenidos por la metodología geométrica son mejores que los obtenidos por el resto de metodologías siempre usando el mismo número de variables. De manera general, los resultados obtenidos por MTG1 y MTG2 mejoran los obtenidos por MT1 y MT2, respectivamente. Exceptuando para el sub-muestreo del 10%, en el que los resultados obtenidos por MT4 son mejores a los obtenidos por MTG4, el comportamiento de ambas metodologías es muy similar. La inclusión de un mayor número de variables para la realización de los análisis de pertenencia revierte en la obtención de unos mejores resultados tanto de C como de M , así como en la obtención de una menor dispersión (zonas sombreadas).

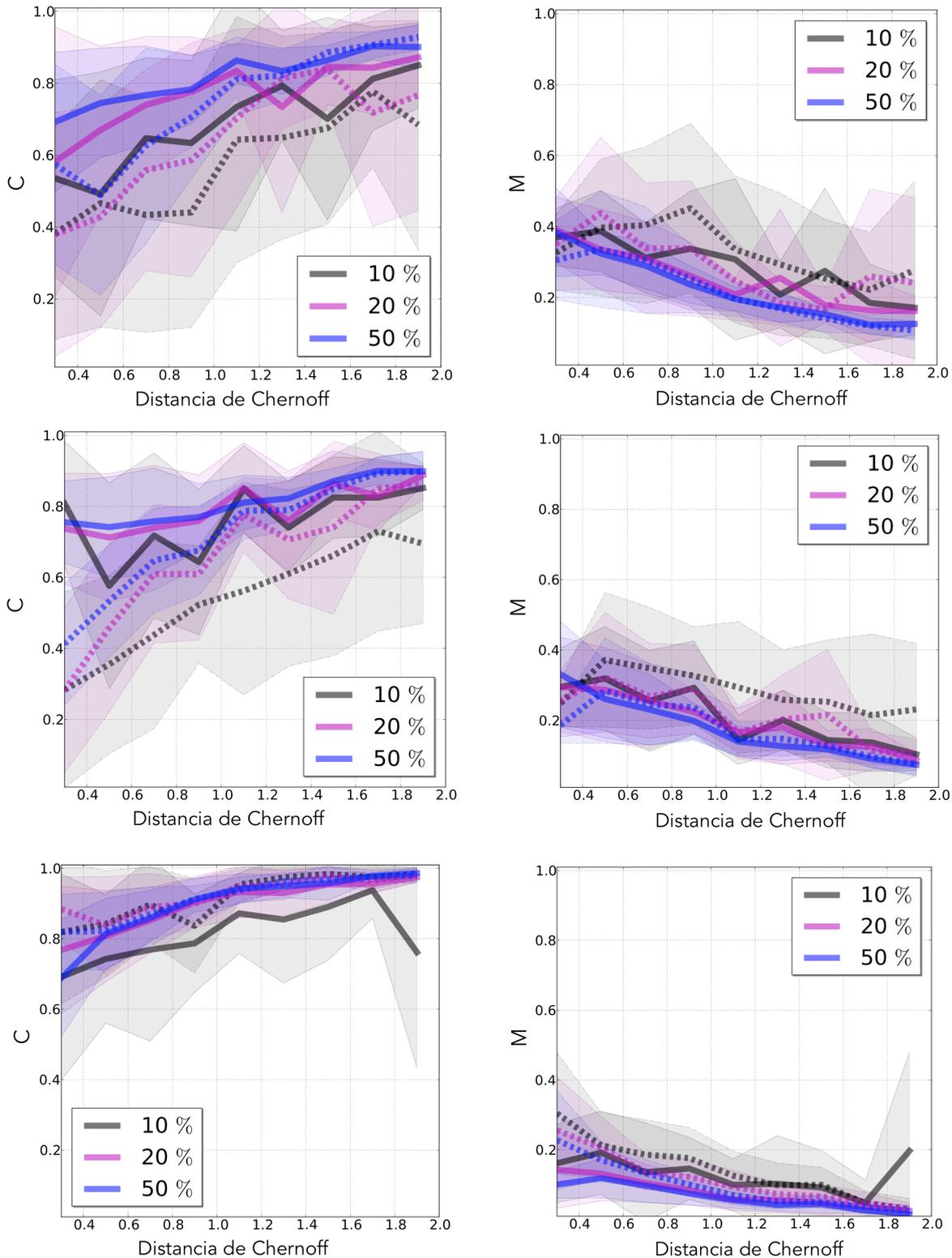


Figura B.1: Efectos del sub-muestreo de ambas figuras de mérito en función de la DC obtenidos en las simulaciones referidas como *Sim2*. En la columna de la izquierda y de la derecha se muestran los resultados obtenidos para C y para M , respectivamente. En este caso, los valores obtenidos son los promedios utilizando un paso de 0.2 en unidades de distancia. Los resultados muestran de nuevo una mejora conforme DC aumenta, así como cuando el número de variables utilizadas en el análisis es mayor. De manera general, los resultados obtenidos por MTG1 y MTG2 mejoran los obtenidos por MT1 y MT2, respectivamente. Exceptuando para el sub-muestreo del 10%, en el que los resultados obtenidos por MT4 son mejores a los obtenidos por MTG4, el comportamiento de ambas metodologías es muy similar.

C

Apéndice 3

C.1 Descripción de los parámetros de las distribuciones de ambas poblaciones obtenidos por los diferentes análisis de pertenencia..

En este apéndice se muestran la descripción de los parámetros calculados para cada cúmulo obtenidos por cada metodología. Estos parámetros son: el nombre del cúmulo, el número de estrellas iniciales, el número de estrellas finales que se analizarán, el número de outliers, el número y porcentaje, respecto del número de estrellas a analizar, obtenido por cada metodología, las coordenadas de las posiciones y de los movimientos propios del centroide de cada cúmulo obtenidas por cada metodología así como sus respectivas dispersiones, los coeficientes de correlación en movimientos propios, la distancia, el exceso de color y la edad (listadas en el catálogo DAML02), y por último los radios de los cúmulos en grados y en parsecs obtenidos por cada metodología. Los mismos parámetros, relativos a las posiciones y a los movimientos propios, han sido estimados para la distribución de campo.

Tabla C.1: Parámetros referentes a la población de cúmulo, determinados por todas las metodologías aplicadas, para todos los cúmulos analizados. Los mismos parámetros se publicarán para la población de campo.

Posición	Parámetro	Descripción	Tipo
1	Name	Nombre del cúmulo	Cadena de Caracteres
2	N_stars_ini	Número de estrellas Iniciales	Entero
3	N_stars_fin	Número de estrellas Finales	Entero
4	N_Outliers	Número de outliers en movimientos propios	Entero
5	N_MTG2	Número de miembros determinados por MTG2	Entero
6	N_MTG4	Número de miembros determinados por MTG4	Entero
7	N_MT4	Número de miembros determinados por MT4	Entero
8	N_MT2	Número de miembros determinados por MT2	Entero
9	perc.MTG2	Porcentaje de miembros determinados por MTG2	Decimal, 3
10	perc.MTG4	Porcentaje de miembros determinados por MTG4	Decimal, 3
11	perc.MT4	Porcentaje de miembros determinados por MT4	Decimal, 3
12	perc.MT2	Porcentaje de miembros determinados por MT2	Decimal, 3
13	x_c.MTG2	Centroide del cúmulo en posiciones (coord X) MTG2	Decimal, 3
14	sigx_c.MTG2	Dispersión del cúmulo en posiciones (coord X) MTG2	Decimal, 3
15	x_c.MTG4	Centroide del cúmulo en posiciones (coord X) MTG4	Decimal, 3
16	sigx_c.MTG4	Dispersión del cúmulo en posiciones (coord X) MTG4	Decimal, 3
17	x_c.MT4	Centroide del cúmulo en posiciones (coord X) MT4	Decimal, 3
18	sigx_c.MT4	Dispersión del cúmulo en posiciones (coord X) MT4	Decimal, 3
19	x_c.MT2	Centroide del cúmulo en posiciones (coord X) MT2	Decimal, 3
20	sigx_c.MT2	Dispersión del cúmulo en posiciones (coord X) MT2	Decimal, 3
21	y_c.MTG2	Centroide del cúmulo en posiciones (coord Y) MTG2	Decimal, 3
22	sigy_c.MTG2	Dispersión del cúmulo en posiciones (coord Y) MTG2	Decimal, 3
23	y_c.MTG4	Centroide del cúmulo en posiciones (coord Y) MTG4	Decimal, 3
24	sigy_c.MTG4	Dispersión del cúmulo en posiciones (coord Y) MTG4	Decimal, 3
25	y_c.MT4	Centroide del cúmulo en posiciones (coord Y) MT4	Decimal, 3
26	sigy_c.MT4	Dispersión del cúmulo en posiciones (coord Y) MT4	Decimal, 3
27	y_c.MT2	Centroide del cúmulo en posiciones (coord Y) MT2	Decimal, 3
28	sigy_c.MT2	Dispersión del cúmulo en posiciones (coord Y) MT2	Decimal, 3
29	px_c.MTG2	Centroide del cúmulo en movimientos propios (coord X) MTG2	Decimal, 3
30	sigpx_c.MTG2	Dispersión del cúmulo en movimientos propios (coord X) MTG2	Decimal, 3
31	px_c.MTG4	Centroide del cúmulo en movimientos propios (coord X) MTG4	Decimal, 3
32	sigpx_c.MTG4	Dispersión del cúmulo en movimientos propios (coord X) MTG4	Decimal, 3
33	px_c.MT4	Centroide del cúmulo en movimientos propios (coord X) MT4	Decimal, 3
34	sigpx_c.MT4	Dispersión del cúmulo en movimientos propios (coord X) MT4	Decimal, 3
35	px_c.MT2	Centroide del cúmulo en movimientos propios (coord X) MT2	Decimal, 3
36	sigpx_c.MT2	Dispersión del cúmulo en movimientos propios (coord X) MT2	Decimal, 3
37	py_c.MTG2	Centroide del cúmulo en movimientos propios (coord Y) MTG2	Decimal, 3
38	sigpy_c.MTG2	Dispersión del cúmulo en movimientos propios (coord Y) MTG2	Decimal, 3
39	py_c.MTG4	Centroide del cúmulo en movimientos propios (coord Y) MTG4	Decimal, 3
40	sigpy_c.MTG4	Dispersión del cúmulo en movimientos propios (coord Y) MTG4	Decimal, 3
41	py_c.MT4	Centroide del cúmulo en movimientos propios (coord Y) MT4	Decimal, 3
42	sigpy_c.MT4	Dispersión del cúmulo en movimientos propios (coord Y) MT4	Decimal, 3
43	py_c.MT2	Centroide del cúmulo en movimientos propios (coord Y) MT2	Decimal, 3
44	sigpy_c.MT2	Dispersión del cúmulo en movimientos propios (coord Y) MT2	Decimal, 3
45	ro.MTG2	Coefficiente de correlación del cúmulo determinado por MTG2	Decimal, 3
46	ro.MTG4	Coefficiente de correlación del cúmulo determinado por MTG4	Decimal, 3
47	ro.MT4	Coefficiente de correlación del cúmulo determinado por MT4	Decimal, 3
48	ro.MT2	Coefficiente de correlación del cúmulo determinado por MT2	Decimal, 3
49	Distance	Distancia listada en DAML02	Entero
50	ColorExcess	Exceso de Color listado en DAML02	Decimal, 2
51	Age	Edad listada en DAML02	Decimal, 2
52	Rad_deg.MTG2	Radio en grados obtenido por MTG2	Decimal, 3
53	Rad_pc.MTG2	Radio en parsecs obtenido por MTG2	Decimal, 3
54	Rad_deg.MTG4	Radio en grados obtenido por MTG4	Decimal, 3
55	Rad_pc.MTG4	Radio en parsecs obtenido por MTG4	Decimal, 3
56	Rad_deg.MT4	Radio en grados obtenido por MT4	Decimal, 3
57	Rad_pc.MT4	Radio en parsecs obtenido por MT4	Decimal, 3
58	Rad_deg.MT2	Radio en grados obtenido por MT2	Decimal, 3
59	Rad_pc.MT2	Radio en parsecs obtenido por MT2	Decimal, 3

D

Publicaciones

La siguiente lista de publicaciones representan las diferentes contribuciones realizadas durante el periodo de esta tesis.

1. **The Gaia-ESO Survey: Stellar content and elemental abundances in the massive cluster NGC 6705:** Cantat-Gaudin T., et al., 2014, A&A, 569, A17
2. **Stellar Open Clusters' Membership Probabilities: an N-Dimensional Geometrical Approach:** Sampedro L., Alfaro E. J., 2016, MNRAS
3. **Stellar Open Cluster Membership Catalog using UCAC4 Data:** Sampedro L., Alfaro E. J. Wilton D. & Moneteiro H., 2016 (in prep.)
4. **The Gaia-ESO Survey: Pre-Main Sequence Stars in the Young Open Cluster NGC3293:** Delgado, A., Sampedro L. et al., 2016, (submitted.)
5. **Preliminary Study of the Kinematic Structure in the Association Cygnus OB1:** Costado T., González M, Sampedro L. & Alfaro E., 2016, (in prep.)
6. **The Gaia-ESO Survey: Target Selection of Open Cluster Stars:** Braggaglia A. et al., 2016, (in prep.)

Bibliografía

Akaike H., 1974, ITAC, 19, 716

Alessi B. S., Moitinho A., Dias W. S., 2003, A&A, 410, 565

Alfaro E. J., Sánchez N., 2011, IAUS, 270, 81

Allende Prieto C., et al., 2008, AN, 329, 1018

Allison R. J., Goodwin S. P., Parker R. J., Portegies Zwart S. F., de Grijs R., Kouwenhoven M. B. N., 2009, MNRAS, 395, 1449

Allison R. J., Goodwin S. P., Parker R. J., Portegies Zwart S. F., de Grijs R., 2010, MNRAS, 407, 1098

Balaguer-Núñez L., Jordi C., Galadí-Enríquez D., Zhao J. L., 2004, A&A, 426, 819

Baumgardt H., Dettbarn C., Wielen R., 2000, A&AS, 146, 251

Bonatto C., Bica E., 2010, MNRAS, 407, 1728

Bonatto C., Bica E., 2010, A&A, 521, A74

Bonnell I. A., Davies M. B., 1998, MNRAS, 295, 691

Bonnell I. A., Bate M. R., Vine S. G., 2003, MNRAS, 343, 413

Bruzual A. G., 2010, RSPTA, 368, 783

Bukowiecki L., Maciejewski G., Konorski P., Strobel A., 2011, AcA, 61, 231

Cabrera-Cano J., Alfaro E. J., 1985, A&A, 150, 298

Cabrera-Cano, J., & Alfaro, E. J. 1990, A&A, 235, 94

Caetano T. C., Dias W. S., Lépine J. R. D., Monteiro H. S., Moitinho A., Hickel G. R., Oliveira A. F., 2015, NewA, 38, 31

Camargo D., Bonatto C., Bica E., 2010, A&A, 521, A42

Camargo D., Bica E., Bonatto C., 2013, MNRAS, 432, 3349

- Carlberg R. G., Sellwood J. A., 1985, *ApJ*, 292, 79
- Carraro G., Beletsky Y., Marconi G., 2013, *MNRAS*, 428, 502
- Cartwright A., Whitworth A. P., 2004, *MNRAS*, 348, 589
- Casamiquela L., et al., 2016, *MNRAS*,
- Chernoff, H. 1952, *Ann. Math. Stat.*, 23, 493
- de Bruijne J. H. J., 2012, *Ap&SS*, 341, 31
- Dempster, A., Laird, N. and Rubin, D.J.: 1977. *Journal of the Royal Statistical Society, Series B*, 39, 1.
- Dias W. S., Lépine J. R. D., Alessi B. S., 2001, *A&A*, 376, 441
- Dias W. S., Lépine J. R. D., Alessi B. S., 2002, *A&A*, 388, 168
- Dias W. S., Alessi B. S., Moitinho A., Lépine J. R. D., 2002, *A&A*, 389, 871
- Dias W. S., Lépine J. R. D., 2005, *ApJ*, 629, 825
- Dias W. S., Assafin M., Flório V., Alessi B. S., Líbero V., 2006, *A&A*, 446, 949
- Dias W. S., Monteiro H., Caetano T. C., Oliveira A. F., 2012, *A&A*, 539, A125
- Dias W. S., Monteiro H., Caetano T. C., Lépine J. R. D., Assafin M., Oliveira A. F., 2014, *A&A*, 564, A79
- Dravins D., Lindegren L., Madsen S., 1999, *A&A*, 348, 1040
- Ducourant C., et al., 2006, *A&A*, 448, 1235
- Efron B., 1979, *Ann. Statist.* 7 126
- Elmegreen B. G., Efremov Y., Pudritz R. E., Zinnecker H., 2000, *prpl.conf*, 179
- Eyer L., Dubath P., Saesen S., Evans D. W., Wyrzykowski L., Hodgkin S., Mowlavi N., 2012, *IAUS*, 285, 153
- Frinchaboy P. M., Majewski S. R., 2008, *AJ*, 136, 118
- Fujii M. S., Baba J., 2012, *MNRAS*, 427, L16
- Fukushige T., Heggie D. C., 2000, *MNRAS*, 318, 753
- Galadi-Enriquez D., Jordi C., Trullols E., 1998, *A&A*, 337, 125
- Gieles M., Portegies Zwart S. F., Baumgardt H., Athanassoula E., Lamers H. J. G. L. M., Sipiør M., Leenaarts J., 2006, *MNRAS*, 371, 793

- Gilmore G., et al., 2012, *Msngr*, 147, 25
- Girardi L., Bressan A., Chiosi C., Bertelli G., Nasi E., 1996, *A&AS*, 117, 113
- Girard T. M., et al., 2011, *AJ*, 142, 15
- Glushkova E. V., Koposov S. E., Zolotukhin I. Y., Beletsky Y. V., Vlasov A. D., Leonova S. I., 2010, *AstL*, 36, 75
- Gower J. C. & Ross G. J. S., 1969, *Minimum Spanning Tress and Single Linkage Cluster Analysis*, Royal Statistical Society.
- Green R. M., 1985, *Spherical Astronomy*, Cambridge University Press, Cambridge.
- Heiter U., Soubiran C., Netopil M., Paunzen E., 2013, *yCat*, 356,
- Høg E., et al., 2000, *A&A*, 355, L27
- Jackson R. J., et al., 2015, *A&A*, 580, A75
- Janes K. A., Tilley C., Lynga G., 1988, *AJ*, 95, 771
- Jarrett T. H., Chester T., Cutri R., Schneider S., Skrutskie M., Huchra J. P., 2000, *AJ*, 119, 2498
- Javakhishvili G., Kukhianidze V., Todua M., Inasaridze R., 2006, *A&A*, 447, 915
- Jeffries R. D., Thurston M. R., Hambly N. C., 2001, *A&A*, 375, 863
- Junqueira T. C., Chiappini C., Lépine J. R. D., Minchev I., Santiago B. X., 2015, *MNRAS*, 449, 2336
- Kaluzny J., Udalski A., 1992, *AcA*, 42, 29
- Kharchenko N. V., Pakulyak L. K., Piskunov A. E., 2003, *ARep*, 47, 263
- Kharchenko N. V., Piskunov A. E., Röser S., Schilbach E., Scholz R.-D., 2005, *A&A*, 438, 1163
- Kharchenko N. V., Piskunov A. E., Schilbach E., Röser S., Scholz R.-D., 2012, *A&A*, 543, A156
- Kharchenko N. V., Piskunov A. E., Schilbach E., Röser S., Scholz R.-D., 2013, *A&A*, 558, A53
- King I., 1962, *AJ*, 67, 471
- King I. R., 1966, *AJ*, 71, 64
- Krone-Martins A., Moitinho A., 2015, *ascl.soft*, 1504.001
- Krumholz M. R., et al., 2014, *prpl.conf*, 243
- Kovalevsky J., 1998, *ARA&A*, 36, 99
- Lada C. J., Lada E. A., 1991, *ASPC*, 13, 3
- Lada C. J., Lada E. A., 2003, *ARA&A*, 41, 57

- Lada C. J., 2010, *RSPTA*, 368, 713
- Landin N. R., Ventura P., D'Antona F., Mendes L. T. S., Vaz L. P. R., 2006, *A&A*, 456, 269
- Larson R. B., 1995, *MNRAS*, 272, 213
- Leisawitz D., Bash F. N., Thaddeus P., 1989, *ApJS*, 70, 731
- Lépine J. R. D., et al., 2011, *MNRAS*, 417, 698
- Lindgren L., et al., 2008, *IAUS*, 248, 217
- Lynga G., Palous J., 1987, *A&A*, 188, 35
- Ma C., 1997, *AAS*, 29, 16.13
- Maia F. F. S., Corradi W. J. B., Santos J. F. C., Jr., 2010, *MNRAS*, 407, 1875
- Magrini L., Randich S., Zoccali M., Jilkova L., Carraro G., Galli D., Maiorca E., Busso M., 2010, *A&A*, 523, A11
- Mermilliod J.-C., 1995, *ASSL*, 203, 127
- Michalik D., Lindgren L., Hobbs D., Lammers U., 2014, *A&A*, 571, A85
- Mignard F. The Hundred Thousand Proper Motion Project, Gaia Data Processing and Analysis Consortium (DPAC) technical note GAIA-C3-TN-OCA-FM-040
- Monet D. G., et al., 2003, *AJ*, 125, 984
- Monteiro H., Dias W. S., Caetano T.C., 2010, *A&A*, 516, A2
- Nilakshi, Sagar R., Pandey A. K., Mohan V., 2002, *A&A*, 383, 153
- Ochsenbein F., Bauer P., Marcout J., 2000, *A&AS*, 143, 23
- Oliveira A. F., Monteiro H., Dias W. S., Caetano T. C., 2013, *A&A*, 557, A14
- Parker R. J., Wright N. J., Goodwin S. P., Meyer M. R., 2014, *MNRAS*, 438, 620
- Paunzen E., Netopil M., 2006, *MNRAS*, 371, 1641
- Perren G.I., Vázquez R.A., Piatti A.E., 2015, *A&A*, 576, A6
- Perryman M. A. C., et al., 2001, *A&A*, 369, 339
- Perryman M. A. C., 2005, *ASPC*, 338, 3
- Phelps R. L., Janes K. A., 1994, *ApJS*, 90, 31
- Platais I., Kozhurina-Platais V., van Leeuwen F., 1998, *AJ*, 116, 2423
- Portegies Zwart S. F., McMillan S. L. W., Gieles M., 2010, *ARA&A*, 48, 431

- Priyatikanto R., Arifyanto M.I., 2015, *Ap&SS*, 355, 161
- Proszkow E.-M., Adams F. C., 2009, *ApJS*, 185, 486
- Randich S., Gilmore G., Gaia-ESO Consortium, 2013, *Msngr*, 154, 47
- Roeser S., Demleitner M., Schilbach E., 2010, *AJ*, 139, 2440
- Rubinstein, R. Y. 1997, *European Journal Operational Research*, 99, 89
- Salpeter E. E., 1955, *ApJ*, 121, 161
- Sampedro L., Alfaro E. J., 2016, *MNRAS*, 457, 3949
- Sánchez N., Alfaro E. J., 2009, *ApJ*, 696, 2086
- Sánchez N., Vicente B., Alfaro E. J., 2010, *A&A*, 510, A78
- Sanders W. L., 1971, *A&A*, 14, 226
- Sanjeev Kulkarni & Gilbert Harman, 2011, *Elementary Introduction to Statistical Learning Theory*.
John Wiley & Sons, Inc., Hoboken, New Jersey
- Sarro L. M., et al., 2014, *A&A*, 563, A45
- Schaller G., Schaerer D., Meynet G., Maeder A., 1992, *A&AS*, 96, 269
- Schwarzschild K., 1907, *Göttingen Nachr.*, 614
- Sharma S., Pandey A. K., Ogura K., Mito H., Tarusawa K., Sagar R., 2006, *AJ*, 132, 1669
- Smith R., Fellhauer M., Goodwin S., Assmann P., 2011, *MNRAS*, 414, 3036
- Skrutskie M. F., et al., 2006, *AJ*, 131, 1163
- Silverman, B.W.: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London (1986)
- Steinmetz M., et al., 2006, *AJ*, 132, 1645
- Su C.G., Fu C.Q., Zhao J.L., et al., 1995, *Acta Astrophys. Sin.* 15, 202
- Subramaniam A., Carraro G., Janes K. A., 2010, *MNRAS*, 404, 1385
- Tadross A. L., 2005, *AN*, 326, 19
- Tadross A. L., 2005, *BASI*, 33, 421
- Tadross A. L., 2012, *RAA*, 12, 158
- Tadross A. L., 2012, *NewA*, 17, 198

- Teixeira R., Galli P. A. B., Le Campion J. F., Fidencio M., Benevides-Soares P., Ferreira A. C. S., Ducourant C., Krone-Martins A. G. O., 2014, *RMxAC*, 43, 55
- Uribe A., Brieva E., 1994, *Ap&SS*, 214, 171
- Uribe A., Barrera R., Brieva E., 2006, *SerAJ*, 173, 57
- van den Bergh S., McClure R. D., 1980, *A&A*, 88, 360
- Vasilevskis S., Klemola A., Preston G., 1958, *AJ*, 63, 387
- Vickers J. J., Roeser S., Grebel E. K., 2016, *arXiv*, arXiv:1602.08868
- Vityazev V. V., Tsvetkov A. S., 2014, *MNRAS*, 442, 1249
- Wilkinson M. I., et al., 2005, *MNRAS*, 359, 1306
- Wolfe, J. H. *Multivariate Behavioral Research*, 5, 1970, 329, 350.
- Wyrzykowski L., Hodgkin S., 2012, *IAUS*, 285, 425
- Zacharias N., Finch C. T., Girard T. M., Henden A., Bartlett J. L., Monet D. G., Zacharias M. I., 2013, *AJ*, 145, 44
- Zhao J.-l., Tian K.-p., Xu Z.-h., Yin M.-g., 1982, *ChA&A*, 6, 293
- Zhao J. L., He Y. P., 1990, *A&A*, 237, 54
- Zhao J. L., Shao Z. Y., 1994, *A&A*, 288, 89

