

Capítulo 4

Identificación de las enanas blancas del halo

Tal y como hemos visto en capítulos precedentes, la función de luminosidad de las enanas blancas ha resultado ser una herramienta idónea para el estudio de muchas de las propiedades del disco galáctico: la edad del disco, la historia pasada del ritmo de formación estelar local, etc. Análogamente, la existencia de una función de luminosidad observacional del halo nos proporcionaría información muy valiosa sobre la edad del halo, la duración del ritmo de formación estelar, o incluso información que nos restringiría la forma posible de la función inicial de masa (IMF). Lamentablemente no se dispone aún de una determinación fiable de dicha función de luminosidad.

La población de enanas blancas del halo ha sido, desde hace prácticamente una década, objeto de un interés continuo, tanto desde el punto de vista teórico (Mochkovitch et al. 1990; Tamanaha et al. 1990) como desde el punto de vista observacional. En este sentido, cabe mencionar que Liebert, Dahn y Monet (1989) estudiaron una muestra de enanas blancas con movimientos propios altos, de la cual obtuvieron la primera — y hasta la fecha la única — función de luminosidad de las enanas blancas del halo. Posteriormente, Flynn, Gould, y Bahcall (1996) y Méndez et al. (1996) estudiaron el contenido de enanas blancas del *Hubble Deep Field* (HDF). Por otra parte, la colaboración *MACHO* propuso que la mayor parte de la materia oscura del halo galáctico se podría hallar en forma de enanas blancas (Alcock et al. 1997). No obstante, los resultados obtenidos han sido negativos. Flynn, Gould, y Bahcall (1996) encontraron que el número de objetos estelares en el HDF con $V - I > 1,8$ es inferior a tres, mientras que Méndez et al. (1996) identificaron seis objetos con $0 < V - I < 1,2$ que podrían tratarse de enanas blancas, pese a que los autores recomiendan rechazar estos objetos como candidatos a enanas blancas debido a sus colores. En cualquier caso ambos valores pueden ser considerados como límites superiores realistas al número de enanas blancas presentes en el HDF.

Centrándonos de nuevo en la función de luminosidad de las enanas blancas del

halo, cabe decir que ésta fue obtenida a partir de una muestra de enanas blancas con paralajes conocidas, movimientos propios elevados ($2,5'' \text{ yr}^{-1} \geq \mu \geq 0,8'' \text{ yr}^{-1}$), utilizando una magnitud visual de corte de $m_V = 19^{\text{mag}}$, y suponiendo que tan sólo aquellas enanas blancas con velocidades tangenciales superiores a 250 km s^{-1} eran miembros del halo (Liebert et al. 1989). Tan sólo cinco enanas blancas conforman la función de luminosidad, por lo que consecuentemente, resulta una estadística muy pobre. Además, en la muestra de Liebert et al. (1989), no existe ninguna enana blanca brillante del halo. La ausencia, o no, de tales enanas blancas brillantes tiene importantes consecuencias con respecto a la IMF de las estrellas del halo tal y como muestran las últimas IMFs modificadas presentadas por Adams y Laughlin (1996) y por Chabrier, Ségretain, y Mera (1996), que predicen un número muy escaso de tales enanas blancas brillantes si la edad del halo es suficientemente grande (Isern et al. 1998). Si la ausencia de enanas blancas brillantes del halo en la muestra de Liebert et al. (1989) es debida a un sesgo observacional o no, es algo que debe ser analizado con mayor detenimiento.

En este capítulo mostraremos como un algoritmo de inteligencia artificial puede ser utilizado con éxito en la identificación de las enanas blancas del halo presentes en los catálogos existentes. Así mismo, presentaremos, a partir del proceso de identificación anterior, una función de luminosidad preliminar de las enanas blancas del halo, siendo finalmente comparada ésta con las predicciones teóricas.

4.1. Metodología

4.1.1. Algoritmos de redes neuronales

Con el advenimiento de grandes bases de datos, la necesidad de técnicas eficientes que mejoren las estrategias de clasificación automática ha conducido a un número considerable de nuevos descubrimientos en este campo. Entre estas técnicas, quizás las más prometedoras son las basadas en algoritmos de inteligencia artificial. Estos algoritmos, introducidos originariamente como modelos descriptivos de los sistemas de redes neuronales biológicos, se han aplicado con éxito en campos tan diversos como el de reconocimiento de dialectos, análisis financiero, análisis clínico, biología (ver Kohonen 1990 para un estudio excelente) o astronomía. Como ejemplos en este campo cabe mencionar que Bazell y Peng (1998) utilizaron estas técnicas para discernir estrellas de galaxias de manera automática, Naim et al. (1995) las utilizaron para clasificar galaxias de acuerdo con su morfología, Serra-Ricart et al. (1996) encontraron la fracción de binarias en cúmulos de estrellas, y Hernández-Pajares & Floris (1994) recurrieron a estas técnicas para clasificar poblaciones a partir del catálogo de datos de *Hipparcos*.

El término algoritmos de redes neuronales hace referencia, de manera bastante vaga, a modelos matemáticos en los que existe una cierta estructura jerárquica, consistente en una serie de nodos (análogos a las neuronas) con múltiples conexiones

(análogas a las dendritas y axones de las células neuronales). Estas conexiones tienen por lo general parámetros adaptables de manera que se pueda modificar la señal que pasa a través de ellos. La gran flexibilidad de los esquemas de estos algoritmos nos permite una amplia generalidad en cuanto al tipo de base de datos a tratar. Por otra parte, una característica común a todas las técnicas de clasificación basadas en algoritmos de redes neuronales consiste en la existencia de un proceso de aprendizaje bastante similar a la forma en que un experto humano realizaría la clasificación de manera manual. Estas técnicas de clasificación se suelen dividir en dos grandes grupos: los procesos de aprendizaje *supervisados* y los *no supervisados*. La diferencia entre ambos reside en el conocimiento *a priori* o no, respectivamente, de los grupos o clases que puedan existir en nuestro espacio de datos. Por lo tanto, en los métodos supervisados, dado un nuevo dato, se determina cual de los grupos ya preexistentes es el que optimiza su clasificación, mientras que en los no supervisados se trata de hallar los grupos “naturales” existentes en los datos sin utilizar ninguna restricción externa adicional. La principal ventaja que poseen los métodos de clasificación no supervisados es que requieren una manipulación mínima de los datos de entrada y, en consecuencia, es de suponer que los resultados que se obtengan sean más realistas.

4.1.2. Mapa auto-organizativo de Kohonen

Uno de los métodos de clasificación no supervisada más destacados es el denominado mapa auto-organizativo (*Self-Organizing Map, SOM*) de Kohonen. Una amplia descripción de este método y de sus numerosas aplicaciones se puede encontrar en Kohonen (1997). Para nuestro propósito, resumiremos aquí los aspectos principales de esta técnica. La idea básica del mapa de Kohonen consiste en proyectar un espacio multidimensional (S) de datos de entrada en un espacio bidimensional (Λ) de salida (Figura 4.1). El espacio bidimensional Λ es un espacio discretizado en un número finito de nodos. Cada nodo del espacio Λ tiene asociado un vector (neurona), $\vec{m}_i \in \mathbb{R}^n$, en el espacio de datos S . Una vez inicializados de manera aleatoria los c nodos o neuronas, el esquema del algoritmo es el siguiente:

1. Para cada uno de los vectores, \vec{x}_j , de la base de datos se determina cual es el nodo k cuya neurona, \vec{m}_k , mejor aproxima al vector de entrada. Esta neurona se denomina “neurona ganadora” y la condición de aproximación viene determinada por la distancia mínima: $d(\vec{m}_k, \vec{x}_j) \leq d(\vec{m}_i, \vec{x}_j)$, $\forall i \in \{1, \dots, c\}$. Donde d representa una distancia (habitualmente la distancia euclídea).
2. Una vez determinada qué neurona es la ganadora se procede a actualizar la neurona ganadora y también a aquellas neuronas vecinas que estén dentro de un determinado radio, N_k . Denominando t al índice indicador de cada paso del algoritmo, el proceso de actualización queda de la siguiente forma:

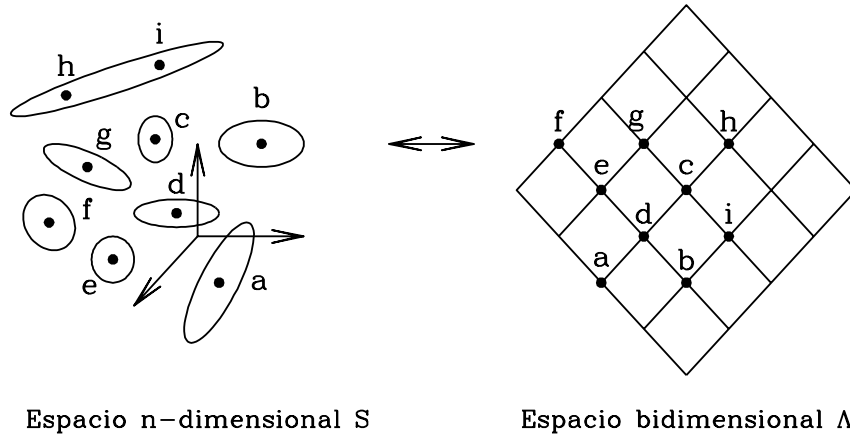


Figura 4.1: Mapa auto-organizativo de Kohonen: el espacio de entrada multidimensional S queda proyectado en un espacio bidimensional Λ de manera que grupos con propiedades similares en el espacio S quedan representados en nodos vecinos del espacio Λ .

$$\vec{m}_i(t+1) = \begin{cases} \vec{m}_i(t) + h_{ki}(t) [\vec{x}_j(t) - \vec{m}_i(t)] & i \in N_k(t) \\ \vec{m}_i(t) & i \notin N_k(t) \end{cases} \quad \forall i \in \{1, \dots, c\} \quad (4.1)$$

La función $h_{ki}(t)$ se suele denominar función vecinal, siendo un tipo de función *kernel* dependiente del tiempo y que decrece con la distancia entre los nodos k e i . Habitualmente se suele expresar en términos de la función gaussiana,

$$h_{ki}(t) = \alpha(t) \cdot \exp\left(-\frac{\|\vec{m}_k - \vec{m}_i\|^2}{2\sigma^2(t)}\right), \quad (4.2)$$

donde $\alpha(t)$ es una función monótona decreciente, $0 < \alpha(t) < 1$, denominada “ritmo de aprendizaje”. La forma exacta de esta función no es relevante: $\alpha(t)$ puede ser tanto lineal, como exponencial, o inversamente proporcional a t . Una opción razonable resulta: $\alpha(t) = A(1 - t/B)$, donde A y B son constantes. Mientras que el término $\sigma(t)$ denota la anchura del *kernel*, es decir, el radio de nodos vecinos, N_k . Del hecho de actualizar no sólo la neurona ganadora, sino también un conjunto de neuronas vecinas, resulta una reordenación de los nodos de manera tal que los centroides que sean cercanos en el espacio bidimensional Λ representan propiedades similares en el espacio de datos S .

3. Los pasos anteriores se van repitiendo hasta que la configuración de nodos sea estable.

A diferencia de otras técnicas de clasificación — clásicas, como la agrupación jerárquica (*hierarchical clustering*); basadas en algoritmos de redes neuronales, como el perceptrón multicapa; o los algoritmos genéticos —, el mapa auto-organizativo de Kohonen destaca tanto por su sencilla implementación como por la facilidad posterior con la que pueden ser visualizados e interpretados los diferentes grupos obtenidos.

En definitiva, el mapa auto-organizativo de Kohonen nos permite un proceso de reducción de la dimensionalidad del espacio de datos de entrada, al mismo tiempo que una identificación de los posibles grupos del espacio de datos conjuntamente con una clasificación automática de los objetos individuales.

4.1.3. El catálogo observacional: un análisis en componentes principales

El catálogo de McCook y Sion (1987) — del cual ya hemos hablado en 2.2 — es una recopilación de datos observacionales referentes a 1279 enanas blancas identificadas espectroscópicamente. Con objeto de clasificar las poblaciones estelares que presumiblemente se hallan en este catálogo, hemos de adoptar un conjunto de variables representativas de sus propiedades. Ahora bien, cuanto mayor sea el número de variables que adoptemos tanto menor será el número de objetos que tengan determinadas todas estas variables, mientras que por el contrario, con un número excesivamente reducido de variables corremos el riesgo de despreciar información valiosa del espacio de datos. Cuántas y cuáles son las variables que necesitamos utilizar es la doble pregunta que se nos plantea. Como primer criterio hemos optado por tomar un número mínimo de variables, de esta manera nos aseguramos que el número de objetos a analizar del catálogo sea máximo. Las variables que hemos adoptado son las siguientes: magnitud visual absoluta, M_V , movimiento propio, μ , coordenadas galácticas, (l, b) , paralaje, π , y un índice de color, $B - V$. El número de objetos que presentan determinaciones observacionales de todas y cada una de las variables del conjunto anterior es de 232, número sensiblemente inferior al número de objetos del catálogo completo. No obstante, se trata de un número suficiente como para permitirnos una clasificación segura, siendo además, el conjunto de variables que hemos adoptado, representativo de las propiedades generales — cinemáticas y fotométricas — de una población estelar.

El primer paso a realizar, previo a cualquier aplicación de un algoritmo de clasificación, es determinar si existen relaciones lineales entre el conjunto de variables utilizado para describir nuestros datos. En este sentido, una de las técnicas más generalizadas es el denominado análisis en componentes principales — una descripción detallada de esta técnica, conjuntamente con los códigos de los algoritmos, se puede encontrar en Murtagh y Heck (1987) —. La idea básica del análisis en componentes

Cuadro 4.1: Análisis en componentes principales.

MATRIZ DE CORRELACIÓN

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
λ_1	1.0000					
λ_2	0.7048	1.0000				
λ_3	0.7560	0.7117	1.0000			
λ_4	0.4168	0.1158	0.2342	1.0000		
λ_5	0.0568	0.0832	0.0688	-0,0312	1.0000	
λ_6	-0,0220	0.0430	0.0323	0.0614	0.1159	1.0000

VALORES PROPIOS

	Val. propios	Porcentaje	Por. acumulado
λ_1	2.5844	43.0730	43.0730
λ_2	1.1136	18.5599	61.6329
λ_3	0.9855	16.4249	78.0578
λ_4	0.8351	13.9175	91.9753
λ_5	0.2780	4.6330	96.6084
λ_6	0.2035	3.3916	100.0000

VECTORES PROPIOS

Variable	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
M_V	0.5746	-0,0892	-0,0235	-0,0643	-0,0455	-0,8094
H	0.5298	0,0374	0,2609	0,2407	0,7061	0,3056
$B - V$	0.5561	-0,0085	0,1325	0,1310	-0,6924	0,4203
π	0.2720	-0,1329	-0,7508	-0,5065	0,1396	0,2620
b	0.0697	0,7114	0,3217	-0,6208	-0,0019	0,0111
l	0.0334	0,6832	-0,4967	0,5280	-0,0240	-0,0778

principales consiste en determinar, mediante rotaciones en \mathbb{R}^n , los ejes más representativos, es decir, aquellos ejes en los que la proyección de los datos tengan varianza máxima. En definitiva, la técnica de cálculo se reduce a un problema de valores y vectores propios. Los k valores propios mayores nos indicaran cuáles son los k ejes más representativos, mientras que los vectores propios nos indican qué combinación de las variables nos dan estos ejes. Si existe alguna variable que es un combinación lineal de las restantes variables del conjunto el análisis en componentes principales podrá detectarlo al obtenerse un valor propio nulo.

En la Tabla 4.1 mostramos los resultados obtenidos mediante el análisis en componentes principales aplicado al subconjunto de 232 enanas blancas del catálogo de McCook y Sion (1987) que poseen determinaciones en todas las variables utilizadas. No obstante, hemos creído más conveniente adoptar el movimiento propio reducido, definido como $H = M_V - 5 \log \pi + 5 \log \mu$, en vez del movimiento propio μ , dado que los grupos resultantes son más fáciles de visualizar. En la subtabla superior de la Tabla 4.1 mostramos la matriz de correlación de las variables utilizadas. La distancia entre variables es directamente proporcional a la correlación entre ellas: $d^2(j, k) = 2(1 - \rho_{jk})$, donde $d^2(j, k)$ es la distancia entre las variables j y k , y ρ_{jk} el correspondiente coeficiente de la matriz de correlación. Del análisis de la matriz de correlación no se observa ninguna similitud destacable entre variables. En la subtabla central presentamos los valores propios obtenidos, conjuntamente con el porcentaje que representan y el porcentaje acumulado. En primer lugar observamos que no existe ningún valor propio nulo, por lo que descartamos la existencia de combinaciones lineales entre las variables. En segundo lugar podemos señalar que, si bien el porcentaje acumulado de los cuatro primeros valores propios ya representa un valor superior al 90 %, los dos últimos también aportan información — aunque en menor grado: un 4,6 % y 3,4 %, respectivamente —, por lo que hemos creído oportuno no descartar ninguna variable. Finalmente, en la subtabla inferior mostramos las proyecciones en los ejes principales de cada una de las variables utilizadas. Los resultados que obtenemos aquí son los mismos que apuntábamos en la discusión previa, siendo resumidamente los siguientes: el análisis en componentes principales de la submuestra observacional nos muestra que no existe ninguna combinación lineal clara entre las variables utilizadas, al mismo tiempo que restringir el número de variables implicaría una pérdida de información.

Una desventaja del análisis en componentes principales se encuentra en el hecho de que únicamente es válido para variedades lineales. No obstante, podemos averiguar dependencias cuadráticas generando nuevas variables como el producto de éstas. Para no alargar aquí la discusión mencionaremos que los resultados obtenidos con diversas combinaciones cuadráticas y logarítmicas, al igual que incrementando o disminuyendo el número de variables utilizadas, no aportan ninguna diferencia significativa con respecto a la conclusión antes mencionada, por lo que concluimos que el conjunto de variables $(M_V, H, B - V, \pi, l, b)$ no presenta combinaciones lineales ni cuadráticas entre sus variables y describe de manera completa las propiedades del

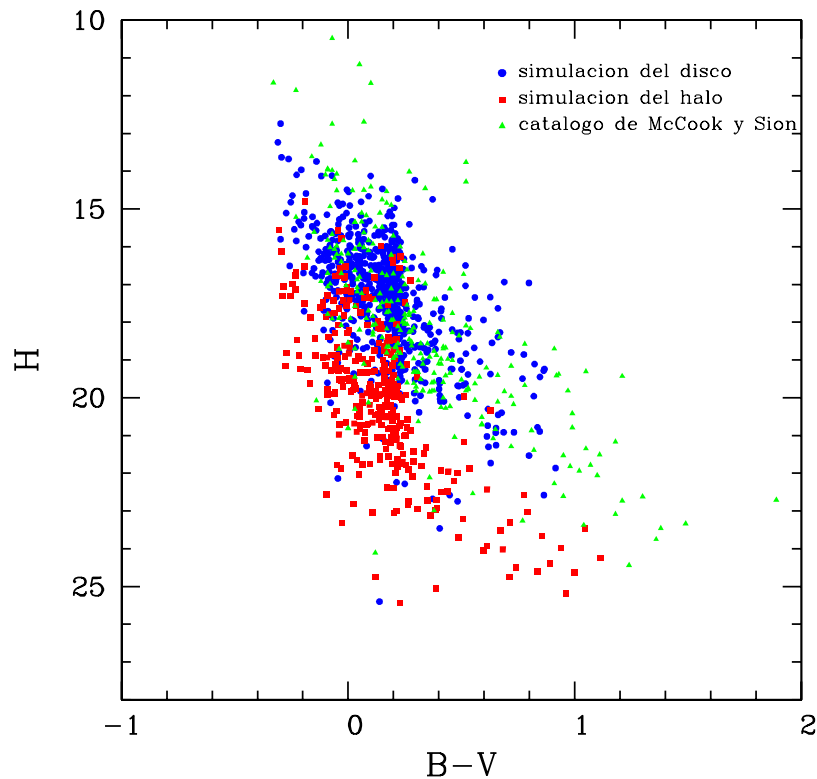


Figura 4.2: Diagrama movimiento propio reducido y color para las simulaciones MC del disco (*círculos azules*), del halo (*cuadrados rojos*) y la submuestra observacional (*triángulos verdes*).

conjunto del catálogo observacional.

4.1.4. Las poblaciones sintéticas de enanas blancas

Las técnicas de clasificación estadística de una base de datos observacional finalizan normalmente con la detección de grupos en el espacio de entrada y su posterior análisis. Dado que nuestro interés consiste en detectar diferentes poblaciones estelares simultáneamente con el proceso de agrupación, utilizaremos una población de estrellas sintéticas, las cuales mezclaremos con los datos de entrada. Estas estrellas servirán como trazadoras y permitirán etiquetar los grupos detectados por el algoritmo como halo, disco, o población intermedia (para más detalles de esta estrategia véase Hernández-Pajares et al. 1994).

En cuanto al conjunto de estrellas sintéticas que utilizaremos como trazadoras, hemos empleado nuestro simulador Monte Carlo para su construcción. En el capítulo

lo 2 hicimos una descripción detallada de la simulación de la población del disco, mientras que en el capítulo 5 se halla una descripción detallada de la población del halo. No obstante, para mayor claridad, resumiremos aquí los valores más importantes utilizados en este capítulo. En la simulación del disco hemos adoptado un IMF estándar (Scalo 1998), una edad del disco de 13 Gyr, y hemos distribuido las estrellas dentro de una esfera de 200 pc de radio centrada en el Sol, de manera uniforme con respecto al plano galáctico y exponencialmente decreciente con respecto al eje z según la ley de altura patrón descrita en 2.1. En cuanto al halo, hemos supuesto que su formación ocurrió hace 14 Gyr a partir de una intensa etapa de formación estelar de 1 Gyr de duración. Las estrellas del halo son así mismo distribuidas dentro de una esfera de 200 pc de radio centrada en el Sol de acuerdo con una distribución radial de densidades dada por la expresión $\rho(r) \propto (a^2 + R_\odot^2)/(a^2 + r^2)$, donde r es la coordenada radial galactocéntrica, $a \approx 5$ kpc y $R_\odot = 8,5$ kpc, de acuerdo con los modelos típicos de un halo isoterma con simetría esférica. Las velocidades de las estrellas trazadoras del halo se han generado aleatoriamente siguiendo distribuciones normales tanto para la componente radial como para la componente tangencial de la velocidad, con las dispersiones de velocidades dadas por Marković y Sommer–Larsen (1997); la velocidad de rotación V_c adoptada es de 220 km s^{-1} . Finalmente, para reproducir de manera adecuada las propiedades del catálogo observacional, hemos restringido nuestras simulaciones, tanto del disco como del halo, al mismo espacio de valores determinado por el subconjunto de 232 enanas blancas que presentaban todas las determinaciones en sus variables. Estas restricciones adicionales son las siguientes: $8,5 \leq M_V \leq 16,5^{\text{mag}}$, $\mu \leq 4,1'' \text{ yr}^{-1}$ y $0,006'' \leq \pi \leq 0,376''$.

En la Figura 4.2, puede comprobarse cómo los resultados de nuestro simulador Monte Carlo, para un diagrama de movimiento propio reducido y color, imitan muy bien los valores de la muestra observacional. Como puede observarse en este diagrama, las dos poblaciones simuladas son, además, fácilmente discernibles. De manera similar podemos producir diagramas para cada par de variables utilizadas, tal y como mostramos en los paneles de la Figura 4.3, verificándose en todos ellos de manera clara como las poblaciones sintéticas creadas con nuestro simulador Monte Carlo quedan en perfecto acuerdo con los datos reales.

Un valor añadido al procedimiento de utilizar estrellas trazadoras descrito anteriormente consiste en que, de esta manera, podemos comprobar no sólo la precisión de nuestro algoritmo de clasificación, sino que además podemos poner a prueba la calidad de nuestras simulaciones Monte Carlo.

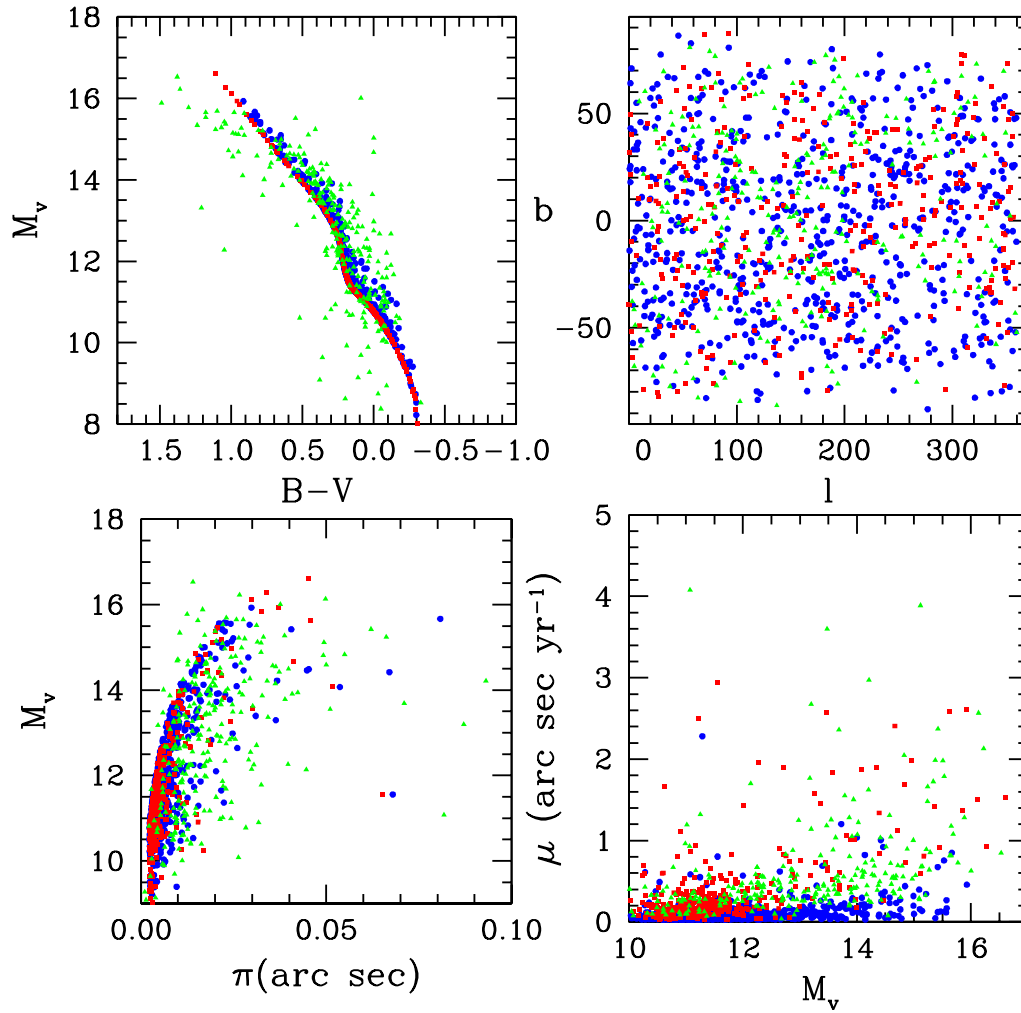


Figura 4.3: Diagramas para diferentes pares de variables de las muestras simuladas del disco (*círculos azules*), del halo (*cuadrados rojos*) y la submuestra observacional (*triángulos verdes*).

4.2. Resultados

4.2.1. Aplicación del algoritmo

El algoritmo de redes neuronales que hemos utilizado es el paquete de programas de dominio público SOM_PAK¹. El procedimiento que sigue el algoritmo consta, tal y como hemos visto, de varias fases. En una fase inicial los vectores representativos de cada nodo son inicializados de forma aleatoria. En una segunda fase, que consta del orden de unas 10^3 iteraciones, el ritmo de aprendizaje es alto — $\alpha(t) = 0,5(1-t/1000)$ —, al igual que el radio de actualización de nodos vecinos — $N_k = 3$ —, el cual se extiende más allá de la mitad del mapa. De esta manera se consigue una ordenación somera aunque global de los nodos. En la última fase, que consta del orden de unas 10^6 iteraciones, el ritmo de aprendizaje es menor — $\alpha(t) = 0,02(1 - t/10^6)$ — al igual que el radio de actualización que únicamente se extiende a la neurona vecina. El algoritmo afina la ordenación de los datos hasta alcanzar una configuración estable. Los diversos parámetros que entran en juego en el algoritmo — tal y como son el número de nodos, la función ritmo de aprendizaje y el radio de actualización de los nodos vecinos — se han de analizar de manera que la clasificación obtenida sea óptima y el error de cuantización mínimo. Por ejemplo, si el número de nodos es excesivo existirán prácticamente tantos grupos como datos, mientras que si el número de nodos es muy reducido todos los datos convergerán a un mismo grupo. En ambos casos la clasificación no tendría sentido.

En nuestro caso, hemos obtenido una clasificación óptima utilizando una red de 5×5 nodos con geometría hexagonal y una función de convergencia con un *kernel* gaussiano. En la Tabla 4.1 presentamos la distribución de objetos en cada nodo y el porcentaje de estrellas trazadoras que hay en cada uno de ellos. Las dos primeras columnas hacen referencia a las coordenadas bidimensionales del nodo, las columnas tercera, cuarta y quinta muestran el número de enanas blancas sintéticas del disco, sintéticas del halo, y observacionales, respectivamente, mientras que las dos últimas columnas hacen referencia al porcentaje de estrellas trazadoras. En primer lugar podemos observar que el conjunto de estrellas sintéticas simula de manera adecuada la población real, puesto que no existen grupos formados únicamente por estrellas sintéticas o sólo por objetos observacionales. Por tanto, al no ser el algoritmo capaz de discernir entre población sintética u observacional podemos concluir que nuestras simulaciones son completamente realistas. En segundo lugar, observamos como existen claramente grupos en los que las estrellas trazadoras de una u otra población son muy altos. Este hecho nos ha permitido etiquetar cada grupo como población del disco (D), o bien del halo (H), si el porcentaje de estrellas trazadoras resulta superior al 70 %. Los grupos que no superan este porcentaje de reconocimiento

¹El paquete de programas SOM se puede obtener en http://www.cis.hut.fi/nnrc/som_pak/ y ha sido preparado por el *SOM Programming Team of the Laboratory of Computer and Information Sciences, Helsinki University of Technology, Finland*.

Cuadro 4.2: Distribución de objetos y porcentajes del mapa auto-organizativo resultante.

MAPA AUTO-ORGANIZATIVO

n_x	n_y	n_{disco}	n_{halo}	$n_{\text{obs.}}$	% disco	% halo
0	0	0	19	7	0.0	100.0
0	1	3	17	5	15.0	85.0
0	2	8	5	24	61.5	38.5
0	3	12	0	8	100.0	0.0
0	4	23	1	16	95.8	4.2
1	0	1	24	2	4.0	96.0
1	1	2	16	3	11.1	88.9
1	2	8	0	8	100.0	0.0
1	3	29	0	4	100.0	0.0
1	4	11	0	8	100.0	0.0
2	0	0	31	6	0.0	100.0
2	1	6	3	8	66.7	33.3
2	2	16	0	15	100.0	0.0
2	3	27	0	2	100.0	0.0
2	4	32	0	5	100.0	0.0
3	0	1	29	7	33.3	66.7
3	1	8	0	11	100.0	0.0
3	2	22	0	7	100.0	0.0
3	3	14	0	2	100.0	0.0
3	4	30	1	8	96.8	3.2
4	0	2	9	12	18.2	81.8
4	1	26	0	24	100.0	0.0
4	2	19	0	7	100.0	0.0
4	3	22	0	18	100.0	0.0
4	4	23	0	15	100.0	0.0

se han etiquetado como población intermedia (I). Por último, una comprobación adicional de la calidad global del esquema de clasificación utilizado se puede obtener a partir del análisis de las estrellas trazadoras que se hayan mal clasificadas. Estos resultados se resumen en la matriz de confusión

$$C = \begin{pmatrix} 0,98 & 0,03 \\ 0,02 & 0,97 \end{pmatrix},$$

donde el elemento de matriz C_{11} indica el porcentaje de estrellas trazadoras del disco clasificadas en grupos del disco, C_{21} es el porcentaje de estrellas trazadoras del disco clasificadas erróneamente en grupos del halo, y así sucesivamente. Como puede apreciarse, esta matriz es muy cercana a la matriz unidad, por lo que nuestro procedimiento de clasificación se muestra como eficaz y seguro.

En el panel superior de la Figura 4.4 mostramos una representación del mapa auto-organizativo de Kohonen donde cada grupo ha sido etiquetado siguiendo el criterio anterior. Así mismo, en el panel inferior se muestra una representación del denominado mapa de Sammon donde las distancias entre nodos son indicativas de la similitud entre grupos. En esta última representación podemos observar como los grupos pertenecientes al disco forman un conjunto homogéneo y claramente distanciado de los grupos identificados como halo, mediando entre ambos los grupos de población intermedia.

Una garantía adicional de que la clasificación es correcta consiste en comprobar, para aquellas enanas blancas observacionales para las que se ha determinado previamente la población a la que pertenecen, si se han identificado adecuadamente con el algoritmo neuronal. Este es el caso, por ejemplo, de las enanas blancas del disco viejo de la muestra de Liebert, Dahn, y Monet (1988), las cuales se hallan en su gran mayoría en los grupos (0, 2) y (2, 1), identificándolas, por tanto, como población intermedia. Puesto que las enanas blancas del disco viejo son objetos fríos que se han determinado en su mayoría a partir de catálogos observacionales basados en criterios de selección en movimientos propios, resulta un sesgo observacional importante de estos objetos hacia velocidades tangenciales altas, siendo, por lo tanto, lógico el resultado de clasificar dicha población como intermedia. Más interesante aún resulta cerciorarse de que la identificación de las enanas blancas utilizadas por Liebert et al. (1989) en su función de luminosidad del halo es la adecuada. En la Tabla 4.2 comprobamos como todas las enanas blancas se hallan en grupos del halo, excepto una que se ha clasificado como población intermedia. En resumen, podemos concluir que todas estas evidencias apuntan hacia una misma dirección: el procedimiento de clasificación utilizado es correcto.

Finalmente, en la Figura 4.5 hemos representado la distribución de diversas variables — velocidad tangencial, movimiento propio reducido, magnitud visual absoluta y paralaje — en función del mapa de Kohonen obtenido. Para ello hemos procedido de la siguiente manera. Para todos los objetos pertenecientes a cada nodo del mapa hemos determinado el valor medio de cada variable; de esta forma podemos

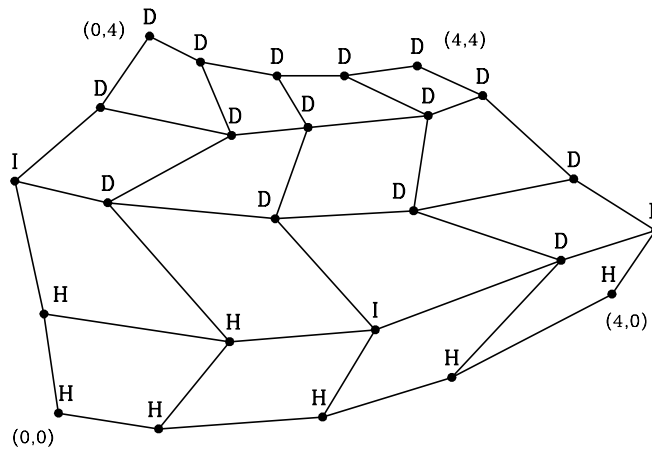
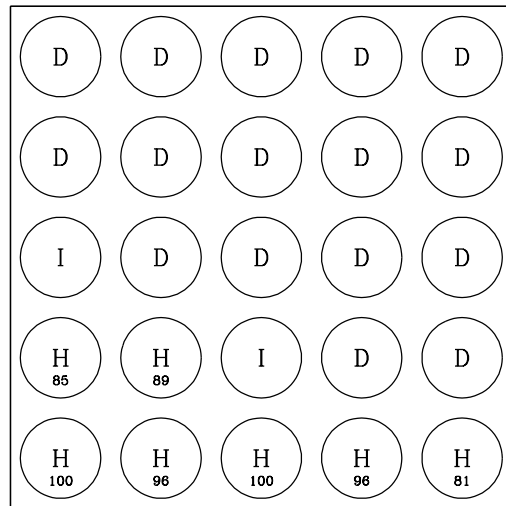


Figura 4.4: Mapa auto-organizativo de la muestra de enanas blancas (panel superior) y representación de Sammon del mismo (panel inferior).

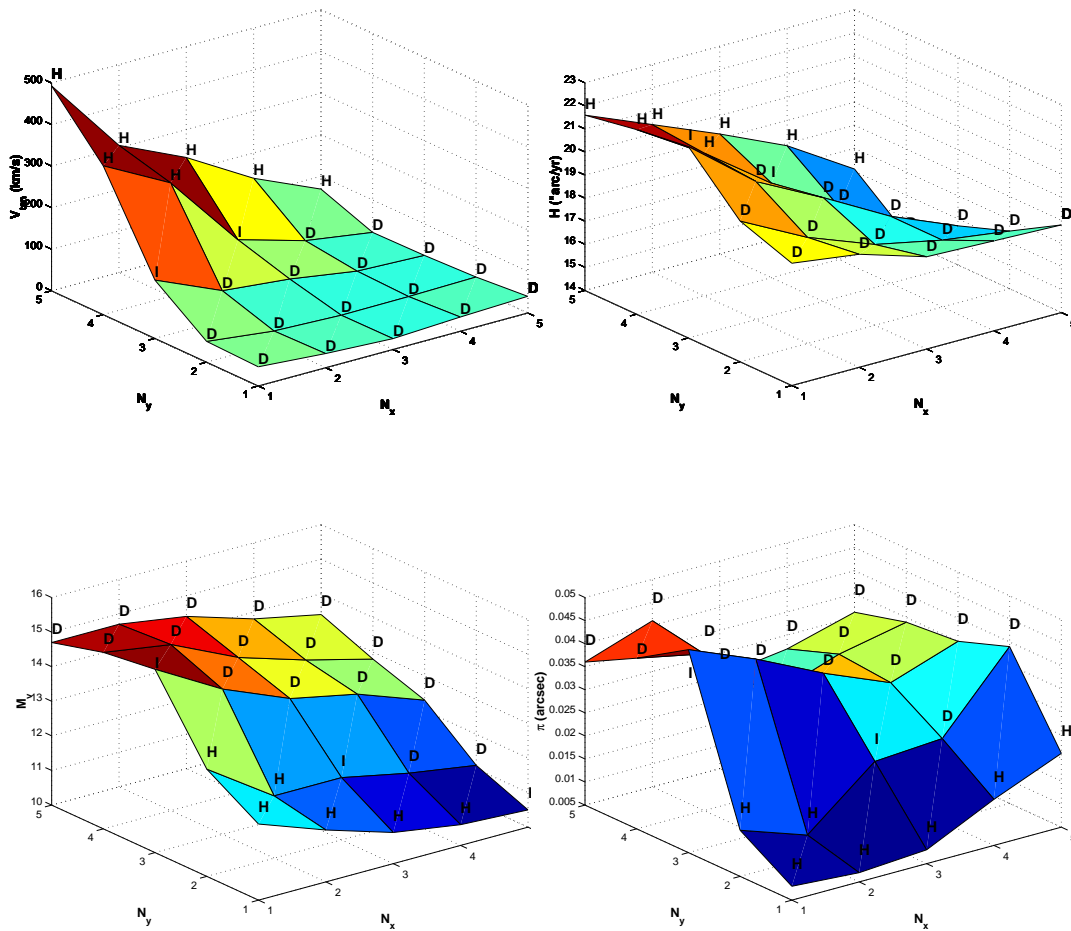


Figura 4.5: Distribución en el mapa de Kohonen de diversos valores medios: velocidad tangencial (panel superior izquierdo), movimiento propio reducido (panel superior derecho), magnitud visual absoluta (panel inferior izquierdo) y paralaje (panel inferior derecho).

Cuadro 4.3: Enanas blancas del halo identificadas por Liebert et al. (1989).

ENANAS BLANCAS DEL HALO
(LIEBERT ET AL. 1989)

Nombre	Grupo	Población
LHS 56	(1,0)	H
LHS 147	(1,0)	H
LHS 282	(0,2)	I
LHS 291	(1,0)	H
LHS 2984	(0,0)	H

apreciar el comportamiento en función de los grupos clasificados. Del análisis de cada panel observamos que los grupos del halo presentan velocidades tangenciales elevadas, movimientos propios altos y magnitudes visuales y paralajes pequeñas. Por tanto, la población identificada como del halo es un población que presenta elevadas velocidades tangenciales, es distante e intrínsecamente brillante.

4.2.2. Objetos candidatos a enanas blancas del halo

Con anterioridad habíamos identificado cada grupo si el porcentaje de estrellas trazadoras de una u otra población superaba el 70 %. No obstante, y con objeto de asegurar la clasificación, consideraremos candidatos a enanas blancas del halo a aquellos objetos pertenecientes a los grupos identificados como halo y cuyos vecinos no sean del disco. Es decir, consideraremos como población genuína del halo a los grupos (0, 0), (0, 1), (1, 0) y (2, 0). De este conjunto de cuatro grupos resultan 20 candidatos a enanas blancas del halo, cuyo nombre y propiedades más relevantes se especifican en la Tabla 4.3. Un rasgo interesante de estas enanas blancas es que todas ellas presentan $M_V \leq 14^{\text{mag}}$, mientras que solo cuatro de ellas presentan movimientos propios superiores a $1,0'' \text{ yr}^{-1}$, siendo su promedio $\langle \mu \rangle = 0,87'' \text{ yr}^{-1}$. Por otro lado, la mayoría de las enanas blancas identificadas como pertenecientes al halo presentan $\pi \leq 0,03''$, siendo el valor promedio $\pi \sim 0,01''$, por lo que resultan velocidades tangenciales superiores a 200 km s^{-1} para 11 de ellas, mientras que tan sólo un candidato presenta una velocidad tangencial inferior a 100 km s^{-1} . Concluimos, por tanto, subrayando la idea expresada anteriormente de que la población identificada del halo presenta velocidades tangenciales altas, es distante e intrínsecamente brillante.

Cuadro 4.4: Enanas blancas del halo identificadas mediante el algoritmo de red neuronal junto con su nombre, grupo y propiedades más relevantes. Las enanas blancas del halo, previamente identificadas por Liebert et al. (1989) se han señalado con un asterisco.

OBJETOS CANDIDATOS A ENANAS BLANCAS DEL HALO

Nombre	Grupo	M_V	μ ($'' \text{yr}^{-1}$)	π ($''$)	$B - V$	Tipo Esp.
LHS 2984*	(0,0)	11.62	0.930	0.015	0,03	DA
LHS 3007	(0,0)	13.06	0.636	0.028	0,29	DA
G 028-027	(0,0)	12.41	0.281	0.003	0,03	DQ
G 098-018	(0,0)	11.81	0.426	0.003	0,38	DA
G 138-056	(0,0)	13.34	0.692	0.006	0,37	DA
G 184-012	(0,0)	13.18	0.427	0.017	0,26	DC
LP 640-069	(0,0)	12.75	0.284	0.009	0,29	DA
LHS 56*	(0,1)	13.51	3.599	0.069	0,36	DA
LHS 147*	(0,1)	13.64	2.474	0.016	0,40	DC
LHS 151	(0,1)	13.46	1.142	0.053	0,33	DA
LHS 291*	(0,1)	13.39	1.765	0.012	0,11	DQ
LHS 529	(0,1)	13.94	1.281	0.046	0,64	DA
LHS 1927	(1,0)	11.41	0.661	0.009	0,11	DA
G 038-004	(1,0)	12.31	0.428	0.010	0,17	DA
LHS 3146	(2,0)	11.88	0.579	0.024	0,17	DA
G 021-015	(2,0)	11.54	0.390	0.015	0,05	DA
G 035-026	(2,0)	11.12	0.335	0.007	-0,14	DA
G 128-072	(2,0)	12.53	0.457	0.025	0,21	DA
G 271-106	(2,0)	11.77	0.396	0.014	0,18	DA
GR 363	(2,0)	11.39	0.133	0.003	-0,03	DA

4.2.3. Función de luminosidad preliminar de las enanas blancas del halo

Con los objetos identificados hasta el momento como enanas blancas del halo hemos construido su función de luminosidad. Para ello hemos utilizado el método habitual de $1/V_{\text{max}}$ adoptando los criterios de selección utilizados por Oswalt et al. (1996) — $m_V \leq 18,5^{\text{mag}}$ y $\mu \geq 0.''16 \text{ yr}^{-1}$ —. La función de luminosidad resultante se muestra en la Figura 4.6 (triángulos negros). Las barras de error se han determinado siguiendo el mismo procedimiento descrito en Liebert et al. (1989) y ya utilizado en 2.3.2. Es decir, la contribución al error de cada objeto es proporcional a su contribución a la densidad; siendo el error final en cada intervalo la suma en cuadratura de los errores individuales. El número de enanas blancas que finalmente han superado

los criterios de selección se muestra en la parte superior de cada barra de error. Por otro lado, la función de luminosidad resultante queda bastante lejos de representar una muestra completa, puesto que obtenemos un valor de $\langle V/V_{\max} \rangle = 0,115$. En consecuencia, la función de luminosidad obtenida debe ser considerada únicamente como preliminar aunque de todas formas representa una mejora considerable con respecto a la función de luminosidad de Liebert et al. (1989).

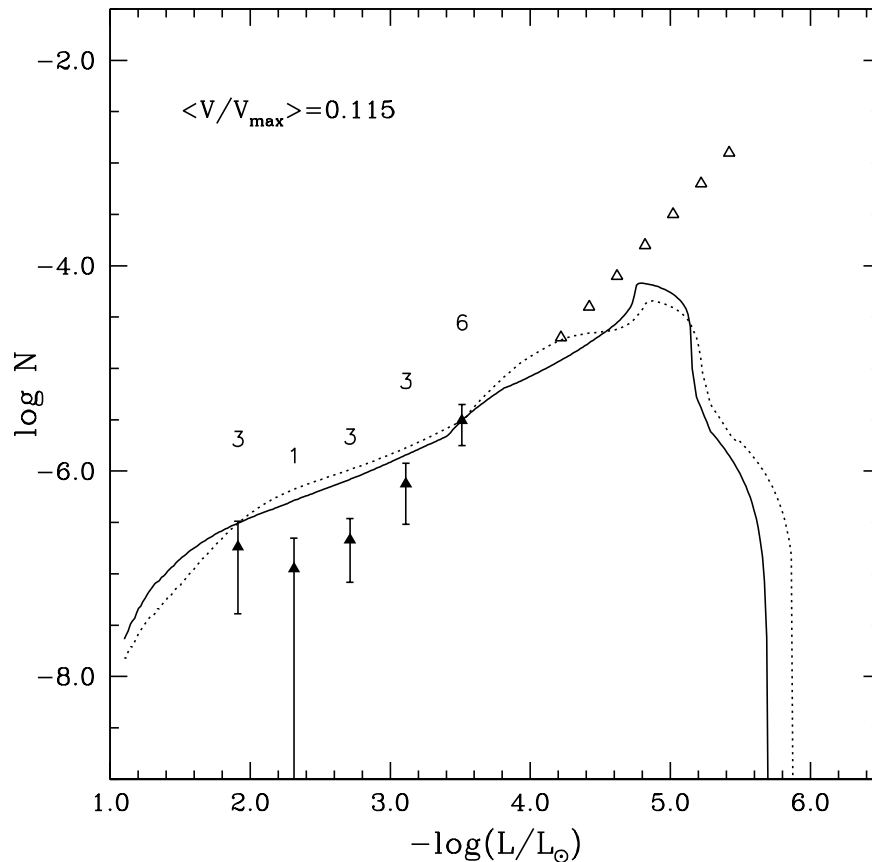


Figura 4.6: Función de luminosidad de las enanas blancas del halo (*triángulos negros*) determinada a partir de los objetos candidatos hallados en el presente estudio conjuntamente con el límite de detección de Liebert et al. (1988; *triángulos blancos*). Las líneas continúa y a puntos representan, respectivamente, funciones de luminosidad teóricas obtenidas a partir de una IMF estándar y una envoltura dominada por H y He.

De acuerdo con la discusión previa, uno de los rasgos más evidentes de esta función de luminosidad es la existencia de objetos candidatos del halo para los intervalos

más brillantes. En cuanto a la parte de la función de luminosidad más débil, el objeto identificado por Liebert et al. (1988), LHS 282, fue clasificado como perteneciente a la población intermedia, por lo que el correspondiente intervalo en su función de luminosidad se encuentra ausente en la nuestra. Otras diferencias de orden menor referentes al intervalo de luminosidades más débiles son debidos al procedimiento de distribución en intervalos y las magnitudes bolométricas asignadas a cada objeto individual. En este sentido, cabe mencionar que nosotros hemos utilizado las correcciones bolométricas de Bergeron, Wesemael y Beauchamp (1995), mientras que Liebert et al. (1989) utilizaron las correcciones bolométricas para un cuerpo negro.

En la Figura 4.6 también se muestra, con objeto de comparar la función de luminosidad obtenida, dos funciones de luminosidad teóricas calculadas a partir de una IMF estándar. La edad adoptada en ambos casos es de 14 Gyr, y la duración de la etapa de formación de 1 Gyr. La línea continua corresponde a una función de luminosidad en la que se ha utilizado para las enanas blancas un modelo de envoltura de H (Wood 1995), mientras que la línea punteada muestra una función de luminosidad calculada a partir de un modelo de envoltura dominado por He (Wood y Winget 1989). El método utilizado para determinar ambas funciones de luminosidad teóricas es el empleado por Isern et al. (1998), donde se ha utilizado el intervalo de menor error como criterio de normalización. En la Figura 4.6 también mostramos el límite de detección (triángulos blancos) de Liebert et al. (1988) para las enanas blancas muy débiles. Por último se ha determinado a partir de la función de luminosidad obtenida en el presente estudio, la densidad local numérica de enanas blancas del halo, resultando $9,0 \times 10^{-6}$ enanas blancas por pc^3 , resultado muy similar al obtenido por Liebert et al. (1989) de $1,3 \times 10^{-5}$ enanas blancas por pc^3 .

