



UNIVERSIDAD NACIONAL DE EDUCACION A DISTANCIA

ESCUELA TECNICA SUPERIOR DE INGENIERIA INFORMATICA

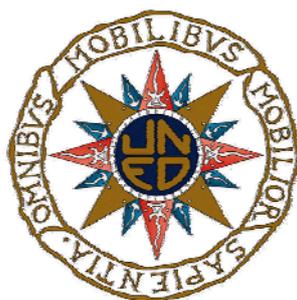
Trabajo de Fin de Máster en Inteligencia Artificial Avanzada: Fundamentos, Métodos y Aplicaciones. Especialidad Sistemas Inteligentes de Diagnóstico, Planificación y Control.

**DISEÑO DE UN PROTOTIPO DE SISTEMA EXPERTO PARA LA
GENERACIÓN DE LOS INFORMES DE CALIDAD RELATIVOS A LOS
OBJETOS VARIABLES DETECTADOS POR LA MISIÓN ESPACIAL GAIA**

FERNANDO GARCIA SEDANO

Dirigido por: LUIS MANUEL SARRO BARO

Curso: 2.011 – 2.012



Diseño de un Prototipo de Sistema Experto para la generación de los Informes de Calidad
Relativos a los Objetos Variables detectados por la Misión Espacial GAIA

Trabajo Fin de Máster

Fernando García Sedano

Dirigido por: LUIS MANUEL SARRO BARO

Tribunal Calificador:

Presidente: D./D^a

Secretario: D./D^a

Vocal: D./D^a

Fecha de Lectura y Defensa:

Calificación:

Resumen

El objetivo de este Trabajo Fin de Máster es el diseño de un SE diseñado para la tarea de evaluar la calidad de los muestreos o “surveys” de variabilidad obtenidos por la misión espacial GAIA. GAIA es una misión de la AGENCIA espacial Europea (ESA) con fecha prevista de lanzamiento en el año 2.013, con el objetivo de recoger datos para entender el origen y evolución de nuestra Galaxia. Esto implica la observación de unos 10^9 objetos celestes (el mayor y más preciso mapa galáctico realizado hasta la fecha) con posiciones, distancias y características espectrales básicas, convirtiéndose de este modo en la sucesora de la misión Hipparcos, aunque a una escala mucho mayor.

En el capítulo 1 se describe brevemente la misión espacial GAIA, así como su estructura organizativa, y se describen los procesos de observación de objetos en astrofísica como paso previo al diseño del SE. Se presenta un encuadre de este trabajo en la estructura de GAIA, así como los objetivos del mismo y el efecto de los sesgos.

En el capítulo 2 se hace un análisis bibliográfico de las aplicaciones de los SE en los distintos campos de actividad de nuestra sociedad del siglo XXI, y posteriormente se profundiza en las aplicaciones realizadas hasta la fecha de SE en astrofísica, así como la utilidad de los mismos.

El capítulo 3 detalla el diseño del SE, incidiendo en el Modelo de Conocimiento y mediante la metodología **CommonKADS**. Se describen las estrellas variables objetivo, se presenta una descripción detallada del dominio, y se presenta la descomposición inferencial de la diferentes tareas en las que se descompone la tarea objetivo del SE diseñado.

En el capítulo 4 se detalla el diseño de una de las tareas en las que se ha descompuesto la tarea objetivo del SE, detallando la estructura del código computable obtenido finalmente, definiendo la implementación de los diferentes parámetros astrométricos, extinciones, magnitudes absolutas, etc. Así mismo se generan los “surveys” sintéticos de las estrellas variables objetivo, y se efectúa la validación de los mismos.

Finalmente, el capítulo 5 contiene un breve resumen de los resultados obtenidos, de las dificultades encontradas, de las posibles mejoras a aplicar y un breve resumen final.

Palabras Clave

Estrellas variables, “surveys” de variabilidad, GAIA, LMC, SMC, Galaxia, Vía Láctea, sesgos, funciones de densidad de probabilidad (pdf), Sistemas Expertos, Modelo de Conocimiento, Conocimiento Inferencial.

Abstract

The aim of this Master is to design an Expert System tailored to the task of assessing the Quality of the GAIA mission variability “survey”. GAIA is a European Space Agency (ESA) scientific mission with a planned launch date in 2.013 that aims the collecting key data for the deep understanding of the origin and past evolution of our Galaxy. This implies the observation and processing of 10^9 celestial objects (the largest and most accurate Galactic map up to now) with positions, distances, and basic spectral characteristics measured, thus becoming the successor of the Hipparcos mission, but on a much larger scale.

In chapter 1, is described briefly the framework of GAIA mission, as well as its organizative structure, and also the observational principles used in Astrophysics as first step to the SE design. It is also showed the GAIA context in which this work is carried out, and also the main targets of this work and bias effects.

Chapter 2 has in charge to carry out an analysis of the SE applications in all those general areas of our society in XXI century, as well as an analysis of any application in the astrophysics area of the Expert Systems and their utilities.

In chapter 3 is detailed de design of the Expert System as main target of this work, focusing in the knowledge model and using **CommomKADS** methodology. Target variable stars used in this work are also described, and a detailed description of the domain in which lay the SE to be designed is carried out, as well as task decomposition and inferencial knowledge of all those task in which main task of the SE designed is decomposed.

Chapter 4 is intended to detail the implementation of the first task in which main task of the expert system is decomposed, using ‘R’ language, and with a deep description of the code structure obtained, defining the implementation procedures of all those astrometric parameters involved in variable stars, absolute magnitudes, and so on. Final target of this chapter is to obtain and validate three different synthetic surveys of the target variable stars.

Finally, chapter 5 contains a short summary of the obtained results, improvements to apply in potential next versions and difficulties found in this work.

Key words

Variable stars, variability surveys, GAIA, LMC, SMC, MW, bias, probability density functions, Expert Systems, Knowledge Model, and Inferencial knowledge.

INDICE

Capitulo 1: <i>INTRODUCCION</i>	1
1.1. La Misión Espacial GAIA	1
1.2. Fundamentos Observacionales en Astrofísica	3
1.3. Encuadre del Trabajo Fin de Máster en la estructura de GAIA.....	6
1.4. Objetivos del Trabajo Fin de Máster	7
Capitulo 2: <i>REVISION BIBLIOGRAFICA DEL AREA</i>	11
2.1 Introducción	11
2.2 Aplicaciones Generales de la IA y de los SE.....	12
2.3 Aplicaciones de la IA y de los SE en Astrofísica	15
Capitulo 3: <i>DISEÑO DEL SISTEMA EXPERTO</i>	19
3.1 Introducción	19
3.2 Conocimiento de la Tarea	19
3.3 Estrellas variables objetivo	24
3.4 Conocimiento del Dominio.....	26
3.5 Conocimiento Inferencial.....	29
Capitulo 4: <i>IMPLEMENTACIÓN PARCIAL DEL SISTEMA EXPERTO</i>	37
4.1. Introducción	37
4.2. Estructura del código ejecutable	37
4.3. Parámetros Astrométricos (t, alpha y delta)	39
4.3.1. Gran Nube de Magallanes (LMC).....	39
4.3.2. Pequeña Nube de Magallanes (SMC).....	41
4.3.3. La Vía Láctea	42
4.4. Cálculo de Extinciones (Av) y Magnitudes Absolutas Extinguidas (mv).....	43
4.5. Cálculo de otros parámetros de GAIA.....	44
4.6. “Survey” de Cefeidas.....	46
4.7. Validación del “Survey” de Cefeidas	49
4.7.1. Histograma de periodos.....	50
4.7.2. Representación de las coordenadas (Coordenadas Ecuatoriales)	52
4.7.3. Representación de las coordenadas (Coordenadas Galácticas).....	53
4.7.4. Visualización de objetos en Banda G	54
4.7.5. Visualización de los valores de Extinción (Av).....	55

4.7.6. Diagrama HR	56
4.7.7. Diagrama logP-MV.....	57
4.7.8. Diagrama logP-VI.....	58
4.7.9. Diagramas PLC.....	59
4.7.10. Diagrama de Amplitudes.....	60
4.7.11. Diagrama distancia-error de paralaje	61
4.7.12. Diagramas de Lutz-Kelker.....	62
4.7.13. Histograma de distancias	63
4.8. “Survey” de RR Lyrae	64
4.9. Validación del “survey” de RR Lyrae.....	66
4.9.1. Histograma de periodos.....	67
4.9.2. Representación de las coordenadas (Coordenadas Ecuatoriales)	70
4.9.3. Representación de las coordenadas (Coordenadas Galácticas).....	70
4.9.4. Visualización de objetos en Banda G	70
4.9.5. Visualización de los valores de Extinción (A_v).....	71
4.9.6. Diagrama HR	71
4.9.7. Diagrama logP-MV.....	72
4.9.8. Diagrama logP-VI.....	73
4.9.9. Diagramas PLC.....	73
4.9.10. Diagrama de Amplitudes.....	74
4.9.11. Diagrama distancia-error de paralaje	76
4.9.12. Diagramas de Lutz-Kelker.....	76
4.9.13. Histograma de distancias	76
4.10. “Survey” de LPV (“Long Period Variable”).....	77
4.11. Validación del “survey” de LPV.....	79
4.11.1. Histograma de periodos.....	79
4.11.2. Histograma de color (K_s) y series	80
4.11.3. Representación de las coordenadas (Coordenadas Ecuatoriales)	82
4.11.4. Representación de las coordenadas (Coordenadas Galácticas).....	82
4.11.5. Visualización de objetos en Banda G	83
4.11.6. Visualización de los valores de Extinción (A_v).....	83
4.11.7. Diagrama HR	84
4.11.8. Diagrama logP-MV.....	84
4.11.9. Diagrama logP-VI.....	85

4.11.10.	Diagramas PLC	85
4.11.11.	Diagrama de Amplitudes	86
4.11.12.	Diagrama distancia-error de paralaje	87
4.11.13.	Diagramas de Lutz-Kelker	88
4.11.14.	Histograma de distancias	88
Capítulo 5:	<i>CONSIDERACIONES FINALES</i>	89
5.1	Introducción	89
5.2	Resultados Obtenidos	89
5.3	Trabajos Pendientes	90
5.4	Resumen final	92
APENDICE A	95
a.	Antecedentes de este trabajo.....	95
b.	Estructura del DPAC.....	96
APENDICE B	99
a.	Sistemas de Coordenadas Astronómicas	99
b.	Coordenadas horizontales	99
c.	Coordenadas ecuatoriales absolutas	101
d.	Coordenadas Galácticas	102
e.	Transformación de ecuatoriales a galácticas (y viceversa)	103
APENDICE C	107
a.	Determinación de valores atípicos	107
APENDICE D	109
a.	Teoría estadística de Cópulas	109
ANEXO E	115
a.	Código R	115
Abreviaturas y Acrónimos	117
Bibliografía	119

Lista de tablas

Tabla 1	15
Tabla 2	39
Tabla 3	41
Tabla 4	43
Tabla 5	43
Tabla 6	46
Tabla 7	98
Tabla 8	98

Lista de Figuras

Figura 1: Vista Artística de GAIA	1
Figura 2a: Diagrama de Hertsprung-Rusell	5
Figura 2b: Estrellas variables en el diagrama H-R	6
Figura 3: Estructura de Subtareas	21
Figura 4: Estructura de la Vía Láctea	22
Figura 5: Tipos espectrales y períodos de variables	25
Figura 6: Descripción del Dominio: Objeto ‘Estrella’	27
Figura 7: Descripción del Dominio: Objeto ‘survey’	27
Figura 8: Descripción del Dominio: Objeto ‘pdf’	27
Figura 9: Descripción del Dominio: Relación ‘pertenece_a’	28
Figura 10: Descripción del dominio: Relación ‘genera’	28
Figura 11: Subtarea ‘generación de Muestras Controladas’	29
Figura 12: Subtarea ‘Aplicación de Algoritmos de QA a surveys’	32
Figura 13: Subtarea ‘Evaluación de H_0 ’	33
Figura 14: Subtarea ‘Detección de Observaciones Atípicas’	36
Figura 15: Histograma de periodos (cefeidas sintéticas)	50
Figura 15a: Histograma de períodos LMC (cefeidas sintéticas)	51
Figura 15b: Histograma de períodos SMC (cefeidas sintéticas)	51
Figura 15c: Histograma de períodos MW (cefeidas sintéticas)	52
Figura 15d: Histograma de períodos de cefeidas	52
Figura 16: Coordenadas Ecuatoriales (cefeidas sintéticas)	53
Figura 17: Coordenadas Galácticas (cefeidas sintéticas)	54
Figura 18: Diagrama visible (cefeidas sintéticas)	54
Figura 19: Diagrama de extinciones (cefeidas sintéticas)	55
Figura 20a: Diagrama HR (cefeidas sintéticas)	56

Figura 20b: Diagrama HR (cefeidas)	57
Figura 21a: Diagrama logP-MV (cefeidas sintéticas)	57
Figura 21b: Diagrama logP-MV cefeidas	58
Figura 22: Diagrama logP-VI (cefeidas sintéticas)	59
Figura 23a: Diagrama PLC Galaxia (cefeidas sintéticas)	59
Figura 23b: Diagrama PLC LMC (cefeidas sintéticas)	59
Figura 23c: Diagrama PLC SMC (cefeidas sintéticas)	60
Figura 24a: Diagrama de Amplitudes (cefeidas sintéticas)	60
Figura 24b: Diagrama de Amplitudes de cefeidas	61
Figura 25: Diagrama de errores de paralaje	61
Figura 26: Diagrama de Lutz-Kelker (I)	62
Figura 27: Diagrama de Lutz- Kelker (II)	63
Figura 28: Histograma de distancias	63
Figura 29: Histograma de periodos (RR Lyrae sintéticas)	67
Figura 29a: Histograma de períodos LMC (RR Lyrae sintéticas)	68
Figura 29b: Histograma de períodos SMC (RR Lyrae sintéticas)	69
Figura 29c: Histograma de períodos Galaxia (RR Lyrae sintéticas)	69
Figura 29d: Histograma de períodos LMC/SMC RR Lyrae	69
Figura 29e: Histograma de períodos Galaxia RR Lyrae	70
Figura 30: Diagrama HR (RR Lyrae sintéticas)	72
Figura 31: Diagrama logP-MV (RR Lyrae sintéticas)	72
Figura 31a: Distribución MV RR Lyrae	72
Figura 32: Diagrama logP-VI (RR Lyrae sintéticas)	73
Figura 32a: Diagrama logP-VI Galaxia RR Lyrae	73
Figura 33a: Diagrama PLC Galaxia (RR Lyrae sintéticas)	74
Figura 33b: Diagrama PLC LMC (RR Lyrae sintéticas)	74
Figura 33c: Diagrama PLC SMC (RR Lyrae sintéticas)	74

Figura 34: Diagrama de Amplitudes (RR Lyrae sintéticas)	75
Figura 34a: Diagrama de Amplitudes LMC RR Lyrae	75
Figura 34b: Diagrama logP-A _I y A _I -A _V SMC RR Lyrae	75
Figura 34c: Diagrama de Amplitudes Galaxia RR Lyrae	76
Figura 35: Histograma de períodos (LPV sintéticas)	80
Figura 36: Diagrama K _s -logP (LPV sintéticas)	80
Figura 37: Histograma K _s (LPV sintéticas)	81
Figura 37a: Distribución K _s LMC LPV	81
Figura 37b: Distribución K _s SMC LPV	82
Figura 38: Diagrama HR (LPV sintéticas)	84
Figura 39: Diagrama logP-MV (LPV sintéticas)	84
Figura 40: Diagrama logP-VI (LPV sintéticas)	85
Figura 41a: Diagrama PLC Galaxia (LPV sintéticas)	85
Figura 41b: Diagrama PLC LMC (LPV sintéticas)	86
Figura 41c: Diagrama PLC SMC (LPV sintéticas)	86
Figura 42: Diagrama de Amplitudes (LPV sintéticas)	87
Figura 42a: Diagrama de Amplitudes LPV	87

Lista de Figuras de los Anexos

Figura 1: estructura Organizativa del DPAC	97
Figura 2: Coordenadas Horizontales	100
Figura 3: Coordenadas Ecuatoriales	101
Figura 4: Hora Siderea	102
Figura 5: Coordenadas Galácticas	103
Figura 6: Transformación de coordenadas Galácticas a Ecuatoriales	104

Capítulo 1: *INTRODUCCION*

1.1. La Misión Espacial GAIA

GAIA es el nombre de una de las misiones de la ESA incluida dentro del programa científico a largo plazo **ESA Horizon 2000**, cuyo lanzamiento está previsto a comienzos del año 2013, situándose en una órbita de Lissajous alrededor del sistema Sol-Tierra, en el punto L2 de LaGrange. Se espera que tenga una vida operativa de 5 años, con una posible ampliación de 2 años más.

GAIA, como su antecesora Hipparcos, es esencialmente una misión de astronomía cuyo objetivo es obtener un catálogo final de unas 10^9 estrellas de hasta magnitud 20, es decir: las obtenidas por la misión Hipparcos incrementadas en un factor de 10.000.

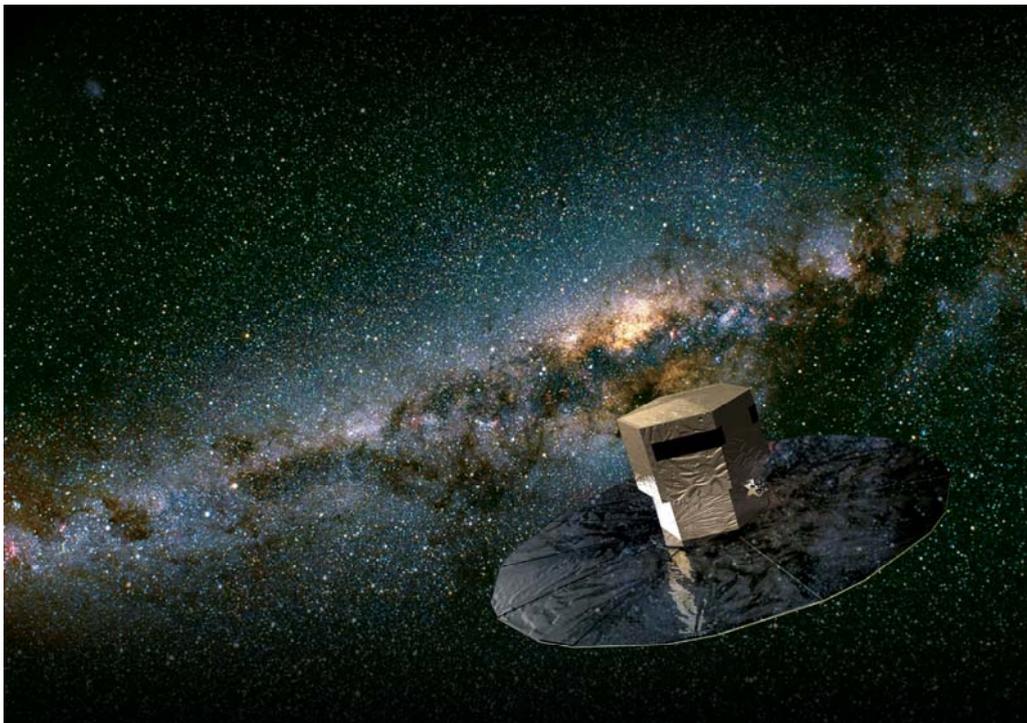


Figura 1: *Vista Artística del Satélite GAIA con el escudo solar desplegado y en vuelo.*

Las características más sobresalientes de GAIA son las siguientes:

- Obtención de medidas astrométricas de las estrellas (distancia, posición, movimientos anuales) con una precisión de 20μ as (microsegundos de arco) de hasta magnitud 15.

- Obtención de medidas astrométricas de las estrellas (distancia, posición, movimientos anuales) con una precisión de 200 μ as (microsegundos de arco) desde magnitud 15 hasta magnitud 20.
- Obtención de medidas fotométricas (observación multicolor y multiépoca) de cada objeto medido.
- Obtención de medidas de velocidad radial de las estrellas observadas (hasta magnitud 17 como mínimo).
- Cada objeto observado se monitorizará en promedio unas 70 veces a lo largo de un periodo de 5 años, midiendo también las variaciones fotométricas.
- Detección de nuevos planetas extrasolares, nuevas enanas blancas y marrones, nuevas estrellas variables y sistemas múltiples, nuevos asteroides, estos últimos dentro de nuestro sistema solar.
- Aportar nuevas pruebas de la Ley de la Relatividad General enunciada por Albert Einstein.

Con estos datos, el objetivo final de la misión GAIA es trazar un mapa tridimensional de nuestra galaxia (y de algún otro objeto mas allá de ella) extremadamente preciso. También hará un mapa de sus movimientos, que dará pistas sobre el origen y evolución de nuestra galaxia. Las medidas fotométricas servirán para conocer las propiedades físicas detalladas de cada estrella observada, caracterizando su luminosidad, temperatura, gravedad, así como la composición de las mismas. Este exhaustivo censo estelar proporcionará los datos astrofísicos básicos para abordar un amplio rango de problemas importantes relacionados con el origen, estructura, y evolución e historia de nuestra galaxia. Se podrán medir simultáneamente un gran número de cúasares y galaxias, así como de planetas extrasolares, y otros cuerpos del sistema solar.

Particularmente importante es la medición de las distancias de los cuerpos astrofísicos que llevará a cabo GAIA: esta determinación de la distancia con la mayor precisión posible, permite hacer física sobre los cuerpos celestes: sin conocer la distancia, dos cuerpos celestes pueden ser fotométricamente iguales; una vez conocida la distancia, es posible caracterizarlos individualmente.

Las distancias pueden medirse desde tierra, pero a causa de la atmósfera principalmente, dichas medidas son poco precisas: es necesario realizar estas medidas desde el espacio para que su precisión sea adecuada. GAIA medirá con mayor precisión que Hipparcos.

La sonda GAIA generará cerca de un millón de Gigabytes de información (1 Petabyte), que deberá ser analizada posteriormente por el Consorcio de Análisis y Procesamiento de Datos (DPAC). Además, cuando la misión haya finalizado, seguirán trabajando durante tres años para confeccionar los datos astrofísicos a partir de dicha información, y crear los catálogos que se publicarán alrededor de 2.020 y servirán como punto de referencia para los datos astrométricos de la Astrofísica en el presente siglo XXI, generando datos para un largo tiempo de estudio.

1.2. Fundamentos Observacionales en Astrofísica

En astrofísica es importante realizar observaciones en diferentes longitudes de onda: la longitud de onda está relacionada con la energía. Una menor longitud de onda implica una mayor energía, y la energía está a su vez asociada con la masa. La ley de Wien, deducida experimentalmente por Wilhelm Wien en 1.893, relaciona de forma inversamente proporcional la emisión de energía de un cuerpo con el concepto de temperatura, que se mide en grados Kelvin (°K).

Todos los procesos que ocurren en el Universo están asociados a la temperatura, a la energía y a la masa. Los procesos más energéticos, que involucran a masas mayores, generarán ondas de longitud de onda muy pequeña. Otros procesos menos energéticos lo harán con ondas de longitud de onda mayor. Se deben registrar todos los procesos para comprender el universo, y por lo tanto es necesario observar el universo en todas las longitudes de onda. Los objetos cósmicos, como el sol emiten en varias longitudes de onda (si bien su pico se encuentra en la zona visible del espectro electromagnético).

Obviamente, el espectro electromagnético es un continuo mucho más amplio que la radiación visible. Más allá de la luz visible, concretamente del color rojo, a medida que incrementamos la longitud de onda, se pasa a la radiación infrarroja y, posteriormente a las ondas de radio. Si se disminuye la longitud de onda, a partir de luz visible violeta, aparece la radiación ultravioleta, a continuación los rayos X y en el último lugar, los rayos gamma.

El ojo humano es sensible, pues, a una limitada fracción del espectro electromagnético, con la mayor sensibilidad centrada en el color verde. Una estrella variable suele cambiar de luminosidad de forma distinta según la longitud de onda en que se estudia. Por ello suelen estudiarse en infrarrojo (utilizando el filtro I), en color rojo (utilizando el filtro R), en luz visual (efectuando estimaciones de magnitud visuales o bien usando el filtro V), en luz azul (filtro B) o en ultravioleta (filtro U). GAIA utiliza su propia banda (banda G) para realizar las observaciones; dicha banda está ya perfectamente caracterizada (G, G_{BP}, G_{RP}, G_{RVS}) por CU5, dependiendo de los instrumentos fotométricos instalados en el satélite: fotómetros BP (“Blue Photometer”), RP (“Red Photometer”) y RVS (“Radial Velocity Spectrometer”) [Jordi-06] [Jordi-10].

También es muy importante en astrofísica el papel que juegan las estrellas variables (particularmente las Cefeidas): se utilizan como indicadores de distancia. Antes de avanzar en este punto, debemos distinguir la magnitud aparente (sea visual, infrarroja o en la longitud de onda en la que se trabaje) de la magnitud absoluta. La magnitud absoluta aísla el factor distancia con el fin de que nos pueda indicar la luminosidad real del objeto; entendemos por magnitud absoluta, la magnitud aparente que tendría un objeto situado a una distancia de 10 parsecs (1 parsec equivale a 3,2616 años luz, o 3×10^{14} km). El Sol tiene una magnitud aparente de $-26,7$ pero su magnitud absoluta es de $4,72$.

Este papel de indicadores de distancia se basa en la denominada ley periodo-luminosidad que vincula directamente la magnitud absoluta de una estrella variable con el valor de su periodo de pulsación (descubrimiento realizado por Henretta Swan en 1.912); un aumento de luminosidad de una estrella variable en función de su periodo es compatible con la teoría de la pulsación estelar, según la cual la luminosidad depende del radio, y este último del periodo; la consecuencia más importante de la relación periodo-luminosidad, es que proporciona una herramienta para calcular la magnitud absoluta de una estrella variable, y una vez conocida esta, calcular la distancia a la que está calculando su diferencia respecto a su magnitud aparente a través de la expresión:

$$M_x = m_x + 5 - 5 \times \log d$$

Donde:

M_x : Magnitud absoluta en la banda x

m_x : Magnitud aparente en la banda x

d : Distancia (en parsecs)

x : Banda (K, K_u, K_a, V, etc.)

Por último, conviene mencionar el hecho de que Las estrellas nacen, evolucionan y mueren. Se puede adelantar que las estrellas variables se encuentran en determinadas fases de inestabilidad en su evolución.

Para ayudar en la comprensión de la evolución estelar, se utiliza el denominado diagrama de Hertzsprung-Rusell (o H-R). Se trata de un gráfico en el que en el eje de ordenadas se sitúa una medida de luminosidad absoluta (luminosidad real de la estrella en unidades, donde 1 corresponde a la luminosidad del Sol, o bien en Magnitud absoluta). En el eje de abscisas se sitúa la temperatura superficial de la estrella, el espectro de la misma, o color ya que los tres indicadores están directamente relacionados; aquí trabajaremos con el espectro o tipo espectral.

Las estrellas de tipo espectral O son azules, con una temperatura superficial de más de 30.000 °K. Las de tipo B son también azuladas (o mejor, blanco-azuladas) con temperaturas entre 10.000 °K y 30.000 °K. Las de tipo espectral A son blancas y de temperatura superficial entre 8.000 °K y 10.000 °K. Las de tipo F son blanco-amarillentas, con temperatura alrededor de los 7.000 °K. Nuestro Sol es de tipo G, color amarillo y temperatura entre 5.000 °K y 6.000 °K. Las estrellas de tipo espectral K son anaranjadas, con temperatura entre 4.000 °K y 5.000 °K. Finalmente, las estrellas de tipo M son rojas con temperaturas inferiores a los 4.000 °K. Cada tipo espectral se divide en 10 subtipos indicados por un número, así una estrella puede ser de tipo A0 o A9, siendo esta última muy similar a una estrella de tipo F0.

Lo primero que puede observarse en este diagrama H-R es que las estrellas se agrupan en determinadas posiciones. La más importante por el número de estrellas es la “secuencia

principal”, también hay acumulación en la zona superior derecha (gigantes y súper gigantes rojas) y en la parte inferior central-izquierda (enanas blancas).

La evolución estelar típica se inicia con una acumulación de gas, denominada “protoestrella” que inicia la combustión nuclear de fusión del hidrógeno que contiene. Se trata de una etapa de inestabilidad hasta que la fuerza de gravedad queda estabilizada con la emisión de energía de la estrella.

Cuando se estabiliza la estrella entra en la “secuencia principal” donde permanece la mayor parte de su “vida”. La cantidad de hidrógeno de la “protoestrella” determina la situación en la secuencia principal y es un factor determinante sobre la evolución posterior de la estrella. Cuanto más masiva es una estrella, menos dura su vida. Se entiende que el Sol tiene una vida de 10.000 millones de años y una estrella de masa equivalente a 30 soles en la secuencia principal, vivirá sólo unos 10 millones de años.

Las estrellas de masas similares a la solar permanecen un 80% aproximadamente de su vida en la secuencia principal. Posteriormente el hidrógeno del núcleo se agota y la estrella inicia una fase de inestabilidad en la que finalmente inicia la combustión de helio en el centro de la estrella a la vez que consume hidrógeno en las capas más cercanas a la superficie. La estrella se dilata y se convierte en una gigante roja. Cada vez que se agota el combustible en el núcleo, la estrella experimenta una convulsión, con contracciones y expansiones, pérdida de masa superficial e inicio de la combustión de elementos más complejos en el núcleo. La pérdida de masa puede provocar la formación de una nebulosa planetaria y, una vez perdida gran parte de la masa, la estrella se consume en una contracción imparable que la lleva a la fase de “enana blanca” y de “estrella de neutrones”.

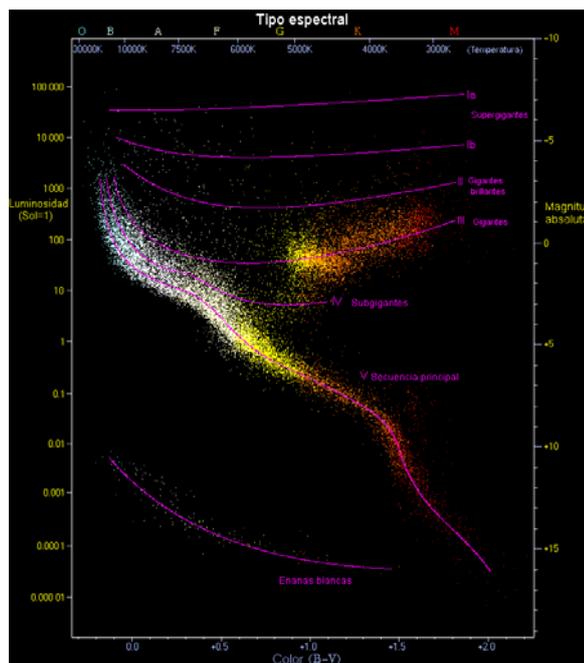


Figura 2a: Diagrama de Hertzsprung-Rusell. El color de cada estrella se corresponde con su tipo espectral.

Esta muerte más o menos controlada de la estrella, contrasta con la evolución de las estrellas masivas (con masa superior a unas 8 veces el Sol en la secuencia principal): en este caso evolucionan de forma mucho más rápida a la fase de gigantes y súper gigantes rojas o azules. De ahí la combustión en el núcleo pasa de estar basada en el hidrógeno a basarse en el helio, y posteriormente el carbono, neón, oxígeno, silicio y finalmente el hierro. El hierro no genera energía en su descomposición nuclear, por lo que la estrella se descompensa totalmente y se colapsa en una explosión de supernova gravitacional (de tipo II o similares). El resultado de esta colosal explosión es una estrella de neutrones o, en las estrellas tremendamente masivas, un agujero negro. La figura 2a muestra el diagrama H-R indicado en este párrafo, y la figura 2b, la situación de las estrellas variables tratadas en este trabajo dentro de dicho diagrama.

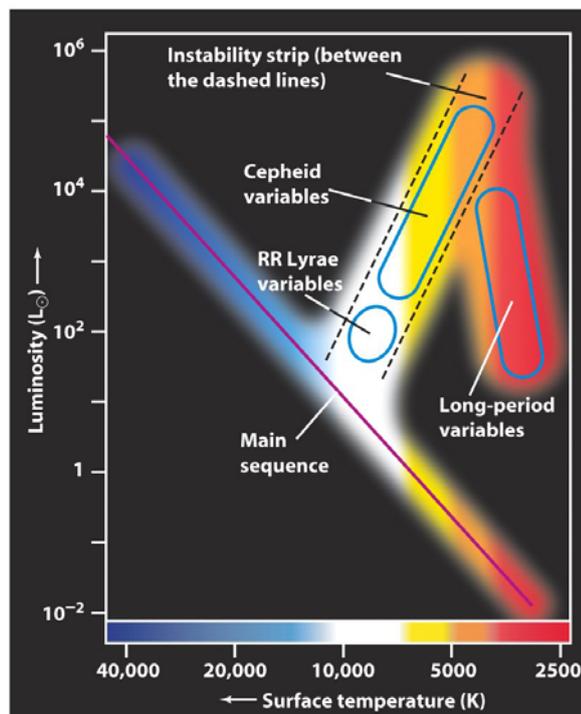


Figura 2b: Posicionamiento de las *Estrellas Variables* (Cefeidas, RR Lyrae y Long Period en el Diagrama de Hertzsprung-Rusell. El color es proporcional a la luminosidad de la estrella representada.

En este trabajo es esencial el papel de las estrellas variables; en el punto 3.3 se detallan las estrellas variables objetivo que se manejan.

1.3. Encuadre del Trabajo Fin de Máster en la estructura de GAIA.

Como en el caso de los PFC mencionados en el apéndice A, este Trabajo Fin de Máster se inscribe dentro del CU7 en un paquete de trabajo denominado “*Quality Assessment*”, cuya misión es analizar estadísticamente los resultados de otras misiones espaciales o proyectos dedicados a la exploración sistemática del cielo y que hayan

producido bases de datos de objetos variables (este tipo de datos se denominan “surveys” en inglés), y compararlos con las bases de datos que vaya produciendo GAIA a lo largo de la misión en las diferentes frecuencias de observación utilizadas; en el punto anterior, ya se han descrito brevemente los fundamentos en los que se basa la observación de este tipo de objetos variables.

El objetivo principal de CU7 es detectar y explicar diferencias sistemáticas en las propiedades de las diferentes clases de objetos variables (Cefeidas, RR-Lyrae, LPV...) detectados por GAIA que pueblan nuestra galaxia.

También en el apéndice A se describen con más detalle los trabajos realizados en el Departamento de Inteligencia Artificial de la U.N.E.D relacionados con la misión espacial GAIA, así como la estructura del DPAC.

1.4. Objetivos del Trabajo Fin de Máster

El objetivo de este Trabajo Fin de Máster es *diseñar un Prototipo de Sistema Experto (SE en adelante) capaz de Generar los Informes Relativos a los Objetos Variables detectados por la Misión Espacial GAIA, entendiendo por informes los parámetros obtenidos por GAIA que puedan constituir indicios de la calidad del sistema.*

A esta tarea, por simplicidad, la denominaremos abreviadamente: “*Análisis de Calidad de GAIA*”. Para el diseño de dicho SE, se usará la metodología **CommonKADS**, descrita en detalle en [Schreiber-99]. No es objeto de este TFM la implementación completa del mismo, aunque si se implementa una de las subtarefas necesarias para alcanzar el objetivo antes indicado; sin embargo, para realizar el diseño de dicho SE prototipo, es esencial conocer el proceso de catalogación de los objetos, el funcionamiento de GAIA en este aspecto, y los *sesgos* y errores que GAIA puede cometer en sus observaciones. Esto exige la realización de numerosos experimentos, tanto con datos sintéticos como con datos reales.

Para alcanzar el objetivo de este trabajo, es imprescindible conseguir una familiarización adecuada con el dominio en el que se va a trabajar, familiarización que se consigue en base a entrevistas realizadas a un experto, Luis Manuel Sarro Baro, astrofísico director de este trabajo. De esta manera se pueden identificar los conceptos claves, que tareas hay que realizar en ese dominio de trabajo, y como pueden descomponerse esas tareas; con todos estos conocimientos, es con los que se podrá luego diseñar el Sistema Experto objetivo: de las entrevistas realizadas al experto, se extraen los datos en los que se basa este trabajo.

En el caso particular de GAIA, CU7 como ya se ha indicado, es responsable de los estudios de variabilidad global; dos de los paquetes más importantes de CU7 son el “*Quality Assessment*” o QA ya mencionado y el “*Bias Estimation*” o BE; este último, se

desgajo del primero hace ya algún tiempo. El objetivo de CU7 es comparar los resultados de otros estudios o Bases de Datos anteriores (OGLE, Hipparcos) con los datos obtenidos por GAIA.

Inicialmente, esta comparación se hacía únicamente sobre números: se comparaba por ejemplo el número de variables de un tipo que detectaba OGLE o Hipparcos, y el número de variables de ese tipo que detectaba GAIA, pero esto carecía de sentido por múltiples motivos; quizás, el más evidente era el número tan diferente de estrellas estudiados en estas Bases de Datos y las estudiadas por GAIA. Para resolver esto, el experto responsable del WP denominado GWP-M-721 propuso que sería más útil comparar las funciones de densidad de probabilidad (pdf de sus siglas en inglés: “*probability density function*”) de cada tipo de estrellas, suponiendo que si todo era correcto, las citadas pdf de cada tipo de estrellas serían las mismas en las diferentes Bases de Datos disponibles.

Siendo la idea útil, pronto se encontró que era necesario refinarla, ya que para los distintos tipos de estrellas, las pdf calculadas en cada Base de Datos se encuentran afectadas por varios *sesgos* diferentes:

- *Sesgo observacional*, que viene dado por los diferentes puntos de observación a partir de la que se han realizado las Bases de Datos: un objeto de determinada magnitud puede detectarse con un brillo débil bien porque esté cerca y su brillo sea intrínsecamente débil, o bien por ser un objeto muy brillante pero muy lejano. Este es uno de los principales sesgos que sufre la pdf.
- *Sesgo de muestreo temporal*, con el que se han obtenido: por ejemplo, OGLE observaba las estrellas una vez al día (durante la noche), y GAIA realizará una media de 80 observaciones en 5 años, lo que implica que en el caso de GAIA se perderán muchas frecuencias, que también dependerán de la posición.

Hay por lo tanto que determinar que frecuencias se van a poder recuperar con GAIA, dependiendo del tipo de estrella y de su posición en el cielo: este trabajo es responsabilidad del WP BE, que realiza mediante simulaciones de distintos periodos de muestreo de periodos simulados, mediante técnicas de análisis temporal. Las BB.DD que proporcionan “surveys” reales son OGLE, Hipparcos, MASS y MACHO, que son las BB.DD de surveys de mayor calidad: de estas BB.DD se obtendrán sus correspondientes pdf para compararlas con las de GAIA.

En definitiva, el SE objeto de este trabajo es, partiendo de unos datos de entrada (básicamente observaciones de periodos de los diferentes objetos astrofísicos), obtener como resultado una predicción de las correlaciones de los objetos variables detectados por la misión espacial GAIA. En este trabajo veremos cómo sistematizar esto para que un Sistema Experto lo realice para todo tipo de estrellas variables con análisis adaptados a cada una de sus clases, centrandó este trabajo en el Modelo de Conocimiento según la metodología **CommonKADS** tal y como esta descrita en [Schreiber-99].

Por otro lado, en el estado actual de la misión, los informes que entregará QA para llevar a cabo la evaluación de calidad de GAIA se resumirán en decidir si los “surveys” de GAIA y los de otras BB.DD provienen de la misma población de estrellas, así como en la determinación de observaciones atípicas.

Capítulo 2: *REVISION BIBLIOGRAFICA DEL AREA*

2.1 Introducción

La experiencia demuestra que una de las mayores dificultades para la resolución de problemas técnicos o científicos tiene su origen en la incorrecta comprensión del mismo y de su entorno para poder plantear la mejor solución. El presente trabajo se ha realizado proponiendo como método de optimización, que la resolución del problema descrito en el punto 1.4 de Generar los Informes Relativos a los Objetos Variables detectados por la Misión Espacial GAIA se considere como un proceso de elaboración de un modelo de acuerdo con la Ingeniería del Conocimiento que facilite la comprensión del dominio del problema y la comunicación con los expertos, antes de llegar a un modelo de la solución expresado en un lenguaje matemático algebraico.

Existe un paralelismo claro entre Ingeniería del Conocimiento (IC en adelante) e ingeniería del Software (IS en adelante); según [Palma-00], se puede definir un Sistema Basado en el Conocimiento (SBC en adelante) como:

“Un sistema software capaz de soportar la representación explícita del conocimiento de un dominio específico y de explotarlo a través de los mecanismos apropiados de razonamiento para proporcionar un comportamiento de alto nivel en la resolución de problemas”

En definitiva, la IC se caracterizan por:

- Manejar conocimiento humano incompleto, inconsistente, impreciso e incierto.
- Razonamiento y respuestas del sistema inseguros.
- Conocimiento evolutivo, posibilidad de aprendizaje.
- A diferencia de la IS, no es aplicable el ciclo de vida en cascada.
- La figura del “experto” es fundamental.

A su vez, es necesario diferenciar entre SBC y Sistemas Expertos (SE en adelante):

- Los SBC están orientados estructuralmente: se centra en el conocimiento sobre un dominio y un problema, habiendo una separación entre el conocimiento codificado en la máquina y los mecanismos deductivos (motor de inferencias).
- Los SE están orientados funcionalmente: se analiza el comportamiento de un experto cuando resuelve problemas, y se intenta emularle; además puede incrementar su experiencia y tiene capacidad para intercambiar información con el usuario del SE.

Históricamente se ha utilizado también el nombre de “Sistemas Expertos”, con una distinción sutil: al hablar de “Sistemas Basados en Conocimiento” en general, simplemente se está haciendo referencia a que lo más importante del sistema es el conocimiento que almacena y gestiona, es decir, una visión estructural. Sin embargo, el término “Sistema Experto” se refiere a un sistema que imita la actividad de un experto humano para resolver una determinada tarea en un dominio específico, es decir: una visión funcional del sistema.

2.2 Aplicaciones Generales de la IA y de los SE

La tecnología representada por los SE actuales, surge de las técnicas de IA que han sido objeto de amplias e intensivas investigaciones desde finales de los años 50. Las investigaciones referidas tuvieron su punto de arranque en las matemáticas para apoyar el razonamiento simbólico, desarrollándose IPL, el primer lenguaje simbólico orientado al procesamiento de listas. Más tarde surgió LISP, reconocido como lenguaje artificial. La investigación en el campo los SE comenzó a mediados de los años sesenta con un alcance limitado y se orientaron hacia juegos o temas altamente académicos e idealizados.

Posteriormente se iniciaron desarrollos en los campos de la medicina, química, industria y la administración. Los SE se emplean para ejecutar una variedad de tareas que en el pasado solamente podían llevarse a cabo por un número limitado de personas expertas. A través de la aplicación de las técnicas de Inteligencia Artificial (IA), los SE captan el conocimiento básico que permite a una persona desempeñar una tarea como un experto en la misma frente a problemas complicados. En [Palma-00] puede encontrarse una amplia descripción de la evolución historia de la IA y de los SBC y SE.

Entre las características más relevantes de los SE, que los distingue de la mayoría de las aplicaciones tradicionales de la computación, es su capacidad para enfrentar problemas que constituyen un reto del mundo real, por medio de la aplicación de procesos que reflejan el discernimiento y la intuición humanos. En el futuro, a medida que se produzcan nuevas arquitecturas de equipos que soporten de una manera directa la ejecución de SE y se perfeccione la tecnología de IA, es razonable esperar un desarrollo de sistemas que se aproximen asintóticamente al comportamiento humano en muchas áreas. Como consecuencia de las investigaciones anteriores, se han ido implementando SE específicos orientados a la resolución de tareas en varios campos. Entre los campos donde se aplican los SE destacan:

- Medicina: Gran parte de los SE que se han desarrollado se aplican en el área de la medicina, donde su función es realizar diagnósticos de enfermedades basados en síntomas y/o en el cálculo de probabilidades.
- Finanzas y seguros: Es un campo importante debido a las grandes inversiones realizadas por entidades financieras, bancarias y aseguradoras, y los riesgos de pérdidas elevadas que pueden producirse. Es un campo en alza en estos momentos.

- **Industria:** El reto de los SE industriales se centra en la necesidad de que se comuniquen con dispositivos sensores, bases de datos, dispositivos de mando y accionamiento en tiempo real. También existen aplicaciones orientadas al mantenimiento industrial y de maquinaria.
- **Electrónica:** Se orientan al diseño, diagnóstico y reparación. El uso de SE se debe a la creciente complejidad de los circuitos y al gran número de parámetros a considerar en los mismos.
- **Defensa:** En actividades de monitorización, seguimiento, diseño, planificación, y control, tanto de armamento como de estrategias de defensa. Especialmente importante en esta área es la incorporación de la Teoría de Juegos, como puede comprobarse en [Brams-89].
- **Aeronáutica y espacio:** Suelen ser aplicaciones orientadas al control de la posición de los satélites y la interpretación de sus imágenes.
- **Agricultura:** Diagnóstico y tratamiento de tierras, control de plagas y creación de nuevos herbicidas.
- **Arqueología:** Para determinar la antigüedad de los restos arqueológicos, composición de las rocas.
- **Geología:** Se emplea en la exploración de suelos, localización de recursos naturales, etc.
- **Meteorología:** Se aplica en la previsión meteorológica, y en el control y administración de recursos hidrográficos.
- **Química:** Estudio de las propiedades de compuestos, interpretación de los resultados obtenidos en los análisis, planificación de los procesos de síntesis.
- **Investigación:** aplicación de SE para diversos objetivos científicos en matemáticas, simulación, etc.
- **Cultura y Entretenimiento:** son muy conocidas las aplicaciones en las que un SE actúa como jugador de un determinado juego, típicamente el ajedrez.

A continuación se muestra en la tabla 1 algunas aplicaciones reales concretas que constituyen la materialización de los puntos anteriores:

Nombre del SE	Descripción	Campo de Aplicación
<i>Airplan</i>	Lanzamientos y despegues de aviones en portaviones	Defensa
<i>Arches</i>	Antigüedad de restos	Arqueología
<i>Argos-2</i>	Simulación de tomas de decisión de un robot	Investigación
<i>Cards</i>	Jugador de bridge	Cultura y Entretenimiento
<i>Casnet</i>	Diagnóstico y tratamiento del glaucoma.	Medicina
<i>Cat</i>	Evaluación de amenazas en aeronáutica	Aeronáutica y espacio

Nombre del SE	Descripción	Campo de Aplicación
<i>Csa</i>	Gestión de Centrales Nucleares	Industria
<i>DAA</i>	Diseño de circuitos integrados VHDL	Electrónica
<i>Delta</i>	Reparación de material móvil ferroviario	Industria
<i>Director Ideal (Banco de Santander)</i>	Valoración de la situación financiera de una empresa o cliente	Finanzas y seguros
<i>AL1, AL5</i>	Ajedrez	Cultura y entretenimiento
<i>Drendal</i>	Establecimiento de la estructura molecular de materia	Química
<i>Emycin</i>	Versión independiente del dominio de <i>Mycin</i> , que contiene todo lo de este último, excepto su conocimiento sobre las enfermedades infecciosas de la sangre, facilitando el desarrollo de aplicaciones de diagnósticos de otro tipo, como es el caso de Puff para enfermedades pulmonares.	Medicina
<i>Expert</i>	Evolución de <i>Casnet</i> . Diagnóstico para oftalmología, endocrinología y también en reumatología	Medicina
<i>Expert Edge</i>	Concesión de créditos y préstamos	Finanzas y seguros
<i>Hearsay II</i>	Interpretación del lenguaje hablado	Investigación
<i>IGC</i>	Control de Calidad	Industria
<i>Macsyma</i>	Cálculo Matemático	Investigación
<i>MetaDrendal</i>	Evolución de Dendral para enfermedades infecciosas	Medicina
<i>Mycin</i>	Diagnostico y terapia de enfermedades bacterianas	Medicina
<i>Planpower</i>	Optimización de Planes de Inversión	Finanzas y seguros
<i>Politics</i>	Estrategia a utilizar en las conversaciones de desarme	Defensa
<i>Priling Advisor</i>	Interpretación de resultados de Sondeos geológicos	Geología
<i>Prospector</i>	Gestión de yacimientos minerales	Geología
<i>RI</i>	Configuración Automática de equipos VAX	Electrónica
<i>Reactor</i>	Gestión de estados de alarma	Defensa
<i>Tax Genus</i>	Optimización de pago de impuestos	Finanzas y seguros

Nombre del SE	Descripción	Campo de Aplicación
<i>Teireisias</i>	Construcción de Bases de Conocimiento	Investigación
<i>Underwriting Advisor</i>	Análisis de riesgos	Finanzas y seguros

Tabla1: Implementación de SE en diversos campos; se muestra el nombre del SE, su utilidad y campo genérico de aplicación.

2.3 Aplicaciones de la IA y de los SE en Astrofísica

Este trabajo pretende ser una aplicación práctica de la IA en general, y de los SE en general a la Astrofísica, un campo en el que las aplicaciones de IA están todavía en un estado muy incipiente. Hablando en términos generales, pueden establecerse dos líneas principales en las cuales se puede avanzar de forma más inmediata en la aplicación de la Inteligencia Artificial a la astrofísica:

- La tarea genérica de Clasificación (supervisada o no supervisada)
 - Juega un papel central en la Astrofísica.
 - Su razón de ser se basa en la existe una cantidad de datos almacenados ingentes en archivos digitales que hay que clasificar, y que superan la capacidad de análisis de los astrofísicos, lo que implica la necesidad de automatizar la tarea como una tarea genérica de clasificación.

- SE (Ingeniería del Conocimiento)
 - Como primer paso para su aplicación, se necesita crear una ontología (o vocabulario común) del conocimiento astrofísico que permita unificar el conocimiento necesario para realizar entre otras la tarea genérica de clasificación y la determinación de esa tarea de manera automática.
 - Posteriormente, se requiere el modelado del conocimiento implícito del astrofísico a la hora de crear los archivos y de procesar la información que hay en ellos.

Un ejemplo claro y a la vez esencial, es el estudio de espectros en Astrofísica: los astrofísicos fundamentan gran parte de su trabajo en la observación de objetos cósmicos (astronómicos) en diferentes longitudes de onda y a partir de los datos que se obtienen de esas observaciones se elaboran mapas de espectro de dichos objetos, que son como una firma propia y a la vez única de cada uno de ellos, y sirven tanto para determinar su composición como sus movimientos (traslación, giro), existencia de campos magnéticos, etc.

Este estudio se sistematizó en el siglo XIX en el que dichos espectros estaban en placas fotográficas. A la hora de analizar los datos que revelaban esas placas se utilizaba una enorme cantidad de recursos humanos (de personas). Se cuenta la anécdota en la que

en la Universidad de Harvard, se contrataron mujeres que en principio no tenía conocimientos en la materia, y que eran mucho más metódicas y cuidadosas en la clasificación que los hombres, para clasificar las fotos de objetos, de donde surgió la primera clasificación de objetos cósmicos: la “Clasificación de Harvard”.

A lo largo del siglo XX surgieron los grandes telescopios, los satélites de observación y sobre todo la digitalización de las observaciones junto con avances en los sistemas de almacenamiento y proceso que permitieron almacenar todo ese conocimiento en archivos digitales electrónicos. El ritmo de crecimiento de los datos ha sido imparable y también la necesidad de automatizar su tratamiento masivo. El primer paso para materializar esta automatización serán los observatorios virtuales, en los que se almacenarán en un repositorio común todas las observaciones realizadas por los diferentes instrumentos de observación repartidos por el planeta y el espacio, que luego los investigadores podrán analizar con el apoyo de herramientas inteligentes de recuperación de la información y clasificación que hagan uso de técnicas de inteligencia artificial.

En relación al asunto de la Ingeniería del Conocimiento aplicada a la astrofísica, podemos preguntarnos si es posible extraer de una manera formal y sistemática ese conocimiento, por ejemplo en forma de ontología (un vocabulario consensuado expresado en un lenguaje formal con capacidades de expresar conocimiento en el campo y con estructuras que facilitan la inferencia por parte de agentes inteligentes que realizan tareas en el campo), para luego utilizarlo para procesar la ingente información obtenida de las observaciones del universo: la falta de la mencionada ontología obstaculiza la utilidad de los observatorios virtuales, ya que dependiendo de los catálogos de observaciones un mismo objeto astronómico puede denominarse de diferentes maneras: tengamos en cuenta que los datos astronómicos son recopilados por diferentes científicos, con diferentes perspectivas, en diferentes países y a veces con diferentes metodologías. Bien pudiera ocurrir (y de hecho ocurre) que en muchas ocasiones se realizan observaciones redundantes, o bien datos que potencialmente pudieran ser muy útiles, finalmente se pierden entre la inmensidad de las BB.DD que los almacenan.

Un proceso de clasificación debidamente automatizado, que si puede ser realizado mediante técnicas de IA, puede liberar al investigador en tiempo y pensamiento para realizar otras actividades más productivas y que requieren una mayor creatividad, aspecto este último en el que las técnicas de Inteligencia Artificial todavía tienen capacidades limitadas: la aplicación de SE en astrofísica es por lo tanto un nicho a explotar para la Inteligencia Artificial en general y los SE en particular.

Para esto es esencial obtener el conocimiento del experto (en este caso del astrofísico), y posteriormente poner ese conocimiento en una máquina mediante técnicas de IA (a través de la ingeniería del conocimiento): buscar métodos, tareas, conocimiento declarativo y procedimental, etc.

Existen ya algunos trabajos en astrofísica donde se ha empezado a utilizar la IA:

- Cálculo de periodos de estrellas variables RR Lyrae por computador [Lafler-64], una de las primeras aplicaciones en Astrofísica, más cercana a la IS que a la IC, pero que contiene elementos interesantes.
- Pre-reducción (o calibración) de elementos CCD en receptores de imágenes de objetos celestes [Johnston-97].
- Clasificación automatizada de estrellas variables por determinación de periodos [Sanchez-05].
- Clasificación de estrellas variables CSE731: “StarClass Enterprise Project”
<https://sites.google.com/site/cse731expertsystemproject/project>

Este trabajo es un paso más en la dirección de la aplicación de SE en astrofísica, si bien como puede apreciarse de los antecedentes indicados más arriba, la aplicación de estas técnicas de forma generalizada está todavía en un estado muy inicial, y en algunos casos más cercana a la algorítmica que a la Ingeniería del Conocimiento.

Capítulo 3: *DISEÑO DEL SISTEMA EXPERTO*

3.1 Introducción

En base a entrevistas realizadas al experto, se define la tarea que debe de realizar el SE y se va desgranando el procedimiento y las técnicas que han de ser aplicados para llevarla a cabo, centrándose en lo que constituye su núcleo: el Modelo de Conocimiento. Este consta de tres partes esenciales:

- Conocimiento de la Tarea: se describen los objetivos que se persiguen, y como pueden alcanzarse dichos objetivos mediante subtareas.
- Conocimiento de inferencias: describe los pasos básicos para realizar una tarea.
- Conocimiento del dominio: describe los conceptos (u objetos) con los que se debe de trabajar, así como las relaciones existentes entre los mismos

Para nuestro trabajo dentro del WP QA, es fundamental entender los *sesgos* descritos en el punto 1.4, y comprender en detalle porque la pdf de OGLE y GAIA para el mismo tipo de objetos variables es diferente (si es que efectivamente lo es). Naturalmente, otra posible causa para que existan diferencias entre la pdf observada por GAIA y la de otra BB.DD, puede ser que GAIA cometa errores: es teóricamente posible que si OGLE es capaz de detectar una determinada frecuencia de una familia de variables (por ejemplo de las Cefeidas) en una región dada del cielo en la que GAIA no lo hace, sea posible por varias causas:

- GAIA no las detecta (por múltiples motivos: porque está a una distancia que excede el campo de detección de GAIA, o bien porque su brillo sea muy débil).
- GAIA las detecta, pero no las cataloga como variables.
- GAIA las detecta y las cataloga como variables, pero dentro de otro grupo al asignarlas un periodo erróneo (por ejemplo, catalogando Cefeidas como RR-Lyrae), por ejemplo por que las observaciones sufran una relación señal ruido muy pobre.

Una vez identificada la tarea de alto nivel a realizar, se describe el método para realizar la misma, y se plantea una descomposición en subtareas más simples, antes de llegar al último paso: la descomposición inferencial de dichas subtareas.

3.2 Conocimiento de la Tarea

La tarea principal a realizar por parte de nuestro SE (Analizar Calidad de GAIA) constará de cuatro subtareas, que se muestran en la figura 3. A continuación se describe en lenguaje formal la tarea principal y el método para llevar a cabo la misma, siempre según la metodología CommonKADS:

Task Analizar Calidad de GAIA;

Goal:

Determinación de la corrección de las observaciones realizadas por la misión espacial GAIA;

Roles:

Input:

BB.DD reales (OGLE, Hipparcos...);
Restricciones;
Especificaciones;

Output:

Decisión relativa a igualdad de pdf de GAIA y otras BB.DD;
Observaciones atípicas;

Spec:

Generar los Informes Relativos a los Objetos Variables detectados por la Misión Espacial GAIA, entendiendo por informes los parámetros obtenidos por GAIA que puedan constituir indicios de la calidad del sistema;

End Task Análisis de Calidad de GAIA;

Tan importante como la definición de la tarea, es la definición del método mediante el que se realiza la tarea indicada; es decir: la estrategia de razonamiento. El método también nos indica cual es la descomposición elegida en subtareas de la tarea de alto nivel, así como el flujo de control en el que intervienen las subtareas identificadas. En nuestro caso:

1. Se generan los “surveys” sintéticos y los reales, estos últimos partiendo de una BB.DD; normalmente los “surveys” reales se obtendrán por filtrado (de acuerdo a unos requerimientos de dicha BB.DD).
2. Aplicación de Algoritmos de QA a los “surveys” para obtención de pdf
3. Evaluación de H_0 partiendo de las pdf de los “surveys” comparados.
4. Detección de las observaciones atípicas en los “surveys” bajo análisis.

Task-Method Comparación-pdf-detección-observaciones-atípicas;

Realizes: Análisis de Calidad de GAIA;

Decomposition:

Sub-Tasks:

Generar muestras controladas;
Aplicar algoritmos de QA a “surveys”;
Evaluar hipótesis nula H_0 ;
Detectar observaciones Atípicas;

Roles:

Intermediate:

Estrellas variables: “Observaciones bajo análisis”;
Surveys: “Conjunto de estrellas variables etiquetadas”;
pdf: “Funciones de densidad de probabilidad”;

Control-Structure:

Generar muestras controladas (requerimientos \rightarrow survey_real);
Generar muestras controladas (requerimientos \rightarrow survey_sintético);
Aplicar algoritmos de QA a surveys (survey_real + survey_sintético \rightarrow pdf1 + pdf2);
Evaluar hipótesis nula H_0 (pdf1 + pdf2 \rightarrow TRUE / FALSE);
Detectar observaciones atípicas (survey_real / survey_sintético \rightarrow survey_hallazgos + survey_sesgos);

END Task-method;

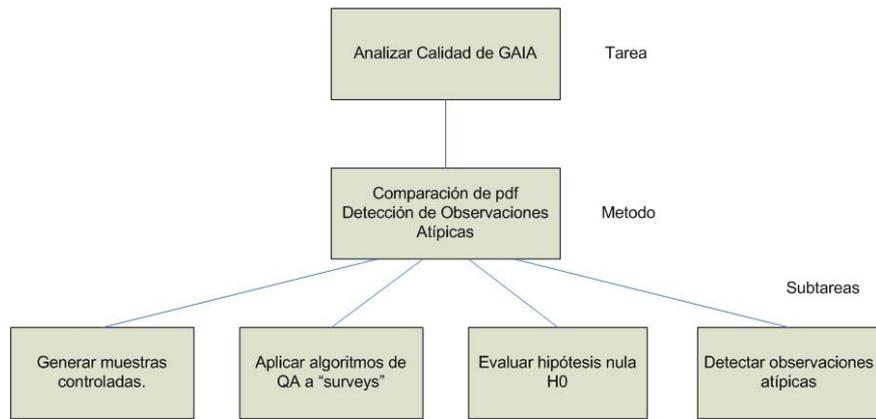


Figura 3: Estructura de subtareas en las que se descompone este trabajo. La tarea principal es el análisis de Calidad de GAIA, que a su vez se descompone en 4 subtareas: Generar muestras controladas, Aplicar algoritmos de QA (cálculo de la pdf), Evaluar hipótesis nula H_0 , y Detectar observaciones atípicas.

1. Generación de Muestras Controladas

Como ya se ha indicado, esta subtarea permite generar bien un conjunto de datos sintéticos para la Gran Nube de Magallanes (LMC), la Pequeña Nube de Magallanes (SMC), el disco fino galáctico, el disco grueso galáctico, y el bulbo galáctico (o abultamiento central), o bien partiendo de una BB.DD real, se filtra esta para obtener un “survey” real con la variabilidad deseada, que posteriormente puedan utilizarse para evaluar los sesgos que GAIA introduce en su análisis:

- Por tratarse de muestras limitadas en tamaño
- Por efecto de la extinción (atenuación del brillo estelar a causa del polvo cósmico que se encuentra durante su trayecto).
- Por introducción de ruido a causa de los equipos o procedimientos de medida.
- Por errores de paralaje.
- Por la frecuencia de medición específica de la misión GAIA (una media de 80 en 5 años).

Los objetos celestes sobre los que se generaran los “surveys” son: LMC (“Large Magellanic Cloud”), SMC (“Small Magellanic Cloud”), y el disco fino, disco grueso y bulbo de nuestra Galaxia, la Vía Láctea (en la figura 4 pueden verse los diferentes elementos modelados de esta última). Puesto que a causa de alguno de los motivos antes indicados, en el proceso de recuperación de frecuencias se perderán estrellas poco brillantes, con elevado ruido, etc., mediante la utilización de estas muestras controladas, se pretende saber donde están esas observaciones en el espacio de parámetros sobre el que se va a trabajar. Estas muestras controladas se generan con código escrito en lenguaje de programación R. Es responsabilidad de BE el manejo de estas muestras

controladas. En este trabajo, y en el caso de los “surveys” sintéticos para cada estrella se define el siguiente espacio de parámetros:

- **type**: indicativo del lugar de origen de la estrella (LMC, SMC, disco fino, disco grueso, y bulbo galáctico).
- **t**: distancia (coordenadas ecuatoriales).
- **alpha**: ascensión recta (coordenadas ecuatoriales).
- **delta**: declinación (coordenadas ecuatoriales).
- **MV**: Magnitud absoluta de la estrella (10 pc).
- **VI**: Color en banda V-I.
- **logP**: logaritmo en base 10 del periodo de la estrella.
- **P**: periodo (en días) de la estrella variable.
- **Amp**: amplitud del brillo de la estrella.
- **l**: longitud (coordenadas galácticas).
- **b**: latitud (coordenadas galácticas).
- **Av**: extinción aplicable a la estrella en la banda V.
- **mv**: magnitud absoluta una vez aplicada la extinción en la banda V. esto es:

$$MV + Av = mv$$

- **v**: magnitud aparente en la banda V (banda visible).
- **g**: magnitud en la banda G (banda GAIA).
- **SNR**: Relación señal ruido. Calculado como:

$$SNR = \frac{\text{Amplitud de la curva de luz}}{\text{Error fotométrico } (v_{ph})}$$

- **sigma**: error de paralaje.
- **rnd.phase**: diferencia de fase aleatoria.
- **K_s**: Índice de color (solo para variables del tipo LPV).
- **Serie**: serie a la que pertenece la estrella sintetizada (solo para variables del tipo LPV).

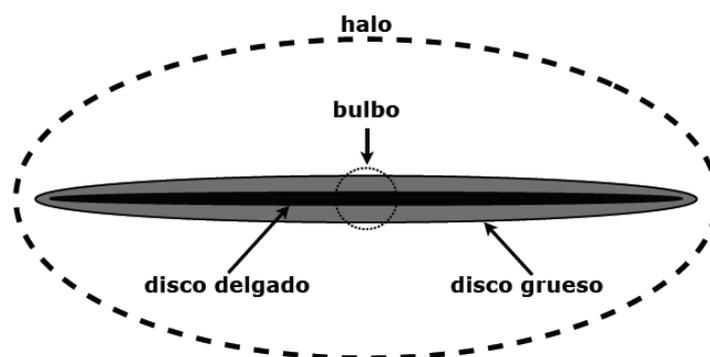


Figura 4: Estructura de la Vía Láctea: bulbo galáctico (línea discontinua fina) en el núcleo, disco fino (en negro), disco grueso (en gris) y halo (línea discontinua gruesa). La corona galáctica no está representada en esta figura

No obstante, según avance el proyecto GAIA, se irán añadiendo nuevos parámetros al espacio de parámetros, como la descomposición de Fourier de las frecuencias de las estrellas variables observadas (este muy importante), así como Temperatura, gravedad y quizás otras.

BE ejecutará los algoritmos que han desarrollado para la detección de frecuencias (proceso que se denomina caracterización), y a partir de ahí se aplican los algoritmos desarrollados por QA para obtener las gráficas (pdf) correspondientes.

2. Aplicar algoritmos de QA a “surveys”

Una vez que se disponga de los resultados de BE sobre los datos sintéticos antes indicados, se utilizarán los algoritmos de QA para procesar dichas salidas, y obtener las pdf correspondientes. Esta fase se realizará mediante R, en lugar del código de QA. En definitiva: se generan muestras controladas en la subtarea 1 de las que por una parte se conocen los valores reales, y se asignan artificialmente valores orientados para pasarlos por el software de GAIA, y se analiza como difiere el valor obtenido del original. Esto naturalmente puede hacerse para cualquier “survey”, no solo el de GAIA, pues en todo caso se conocen las características observacionales de los instrumentos de a bordo (por ejemplo, de OGLE), pero siempre teniendo en cuenta que en el caso de los “surveys” sintéticos los valores de los parámetros deben extenderse por los límites del espacio de parámetros para analizar la sensibilidad en dichos límites.

En esta fase se realiza el cálculo de las pdf sobre los grupos de variables (reales o sintéticas) analizados para obtener las correspondientes funciones de distribución de probabilidad de las diferentes poblaciones estelares observadas o sintetizadas.

Finalmente, se procede al análisis de los datos obtenidos, que consta de dos subtareas bien diferenciadas:

3. Evaluar hipótesis nula H_0

Con las pdf obtenidas en la subtarea anterior, tanto para los “surveys” reales como para los sintéticos (recordemos aquí que dichas pdf son multivariantes) se pueden comparar ambas pdf para posteriormente calibrar los sesgos introducidos por GAIA para los diferentes datos generados.

En definitiva: se verificara si es estadísticamente posible afirmar que la hipótesis nula H_0 : *“Las pdf obtenidas de BB.DD reales y del procesamiento de*

los “surveys” sintéticos por parte de BE provienen de la misma población estelar para las mismas familias de variabilidad”.

4. Detectar observaciones atípicas.

Tan importante como verificar si estadísticamente hablando H_0 es cierta, es detectar las observaciones atípicas obtenidas de los surveys bien sintéticos o bien que provengan de BB.DD reales; la localización de atípicos se realiza mediante métodos estadísticos; dichas observaciones atípicas pueden ser debidas a dos razones bien distintas:

- Las observaciones atípicas se deben a *sesgos* o errores sistemáticos ya explicados en apartados anteriores y que son específicos de GAIA.
- Puede deberse a observaciones reales, denominadas *hallazgos*, que ponen de manifiesto nuevos objetos, o la existencia de objetos de corta vida que forman parte de la evolución específica de las estrellas variables analizadas. No es en absoluto descartable que GAIA detecte *hallazgos*, uno de los objetivos de la misión.

Llegados a este punto, surge una pregunta interesante: ¿Cómo podemos distinguir una observación atípica afectada de *sesgo* de un *hallazgo*? El criterio utilizado para ello es la relación señal ruido de la observación en cuestión: una observación atípica será en general un *hallazgo* cuando su relación señal ruido (SNR en nuestro espacio de parámetros) sea elevada. Dicho umbral será determinado por equipos de trabajo específicos de GAIA.

3.3 Estrellas variables objetivo

Como ya se ha mencionado en el capítulo anterior, una de las subtareas a implementar, es la generación de “surveys” sintéticos de determinadas familias de estrellas variables, que posteriormente pueda ser usado para evaluar la calidad de los procesos de astronomía de la Misión Espacial GAIA en base a los procedimientos indicados por el experto.

La obtención de dichos “surveys” se basará también en la información contenida en artículos científicos astronómicos de investigadores e instituciones de reconocido prestigio; no obstante, como ya se ha indicado, los parámetros sintéticos obtenidos no siempre se pretende que sean reales, en el sentido de que siempre respondan a las observaciones astronómicas disponibles, sino que en ocasiones pueden forzarse a valores no reales para con ello evaluar mejor la capacidad astrométrica de GAIA. En particular, se generarán “surveys” para las siguientes familias de estrellas variables: Cefeidas, RR Lyrae (únicamente R Rab, modo fundamental) y LPV (“Long Period Variable”).

Las **Cefeidas** son estrellas variables tipo F-K (ver figura 2a) muy masivas con un valor de luminosidad extremadamente variable, con periodos que oscilan entre 1 y 70 días, cuyo

mecanismo de variabilidad viene dado por la contracción y expansión radial de la propia estrella en torno a un valor de radio medio dependiendo de su masa; según este principio, la contracción de la estrella implicaría una menor radiación de energía hacia el exterior, y por lo tanto un aumento de la temperatura interna, y como consecuencia última un aumento de las reacciones nucleares de fusión, lo que provoca un aumento de energía que primero compensa la contracción que la ha producido, y luego obliga a una expansión de la estrella haciéndola más transparente, circunstancia que aumenta la radiación enviada al exterior, lo que en consecuencia enfría la estrella, repitiéndose así el ciclo descrito; de este modo, la radiación de energía (y por tanto su luminosidad) es máxima cuando su radio es mínimo y viceversa. Se denominan cefeidas por el nombre de su estrella prototipo: δ -Cephei, de la constelación Cepheus (o Cefeo).

Las **RR-Lyrae** son estrellas variables de tipo espectral A generalmente, menos luminosas y masivas que las cefeidas y de más corto periodo (0.05 a 12 días) que las cefeidas de tipo B-F (ver figura 2a), en las que comienza la fusión del Helio en su núcleo, y aunque su mecanismo de variabilidad es muy similar (contracciones y expansiones de su radio alrededor de un valor de radio medio), en este caso aquel es debido a la presencia en la estrella de elementos pesados (Oxígeno, Carbono, Hierro...); este atributo se denomina *metalicidad* de la estrella (representando habitualmente con la letra *Z*), y está directamente relacionado con la edad de la misma: una metalicidad alta, es característica de estrellas jóvenes. Su estrella prototipo es la RR de la constelación Lyrae (o Lira). A veces también se denominan “Cefeidas de Cúmulo”, como en la figura 5.

Las **LPV** son estrellas gigantes o súper gigantes rojas tipo M, R, C ó N (ver figura 2a), con periodos de variación en su luminosidad muy amplia y largos en general, que pueden variar entre unos pocos días y algunos miles de días. Una estrella prototipo de esta clase, es Mira (ω -Ceti) de la constelación Cetus (o La Ballena).

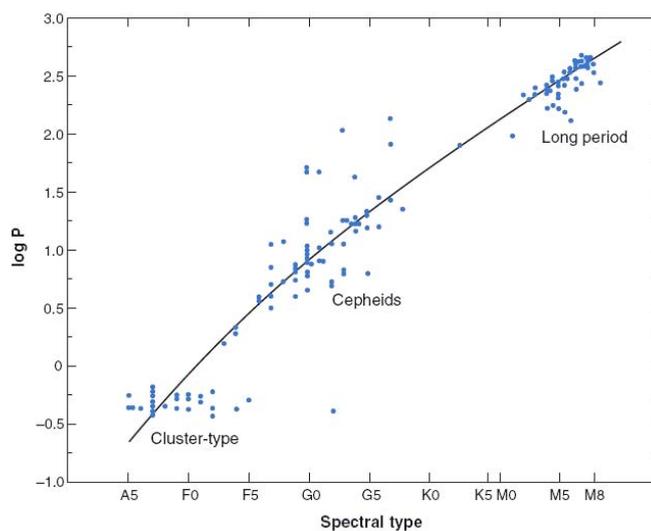


Figura 5: Tipos espectrales y Periodos de las estrellas variables objeto de este trabajo: Cefeidas, RR Lyrae (también llamadas “Cluster-type”) y Long Period Variable. El periodo aparece en escala logarítmica.

La figura 5 muestra de manera gráfica los tipos espectrales y periodos de las familias de estrellas variables antes indicadas.

No obstante, podemos adelantar que la clasificación de la figura 5 no es absoluta: las últimas investigaciones en el campo de objetos variables muestran que los diferentes tipos de estrellas variables presentan unos periodos que si bien mantienen sus valores medios en las zonas indicadas en la citada figura, también poseen valores que se solapan entre sí, por ejemplo en las LPV.

Con las subtarefas definidas en 3.2 y con la descripción cualitativa del punto 3.3 extraída del experto, ya estamos en condiciones tanto de modelar el dominio como de diseñar en detalle las subtarefas en las que hemos descompuesto la tarea de alto nivel, trabajo que se aborda en los apartados siguientes.

3.4 Conocimiento del Dominio

En base a las entrevistas realizadas al experto, podemos especificar el dominio en el que hemos de trabajar, comenzando por los conceptos que forman parte del mismo, y representándolos mediante diagramas de clase en lenguaje UML. Nuestro dominio consta de los siguientes conceptos (u objetos):

- Estrella
- Estrella variable (subclase de estrella)
- Cefeida, RR Lyrae y LPV (subclases de estrella variable).
- “Survey”.
- “Survey_sintético”, “survey_real” y “survey_atípicos” (subclase de “survey”).
- “survey_hallazgos” y “survey_sesgos” (subclase de “survey_atípicos”).
- pdf (“probability density function”), o function de densidad de probabilidad de un determinado “survey” para un conjunto de parámetros del espacio de parámetros definidos en el punto 3.2.

El diagrama UML de dichos objetos se muestra en las figuras 6 (objeto estrella con sus respectivas subclases) 7 (objeto “survey” con sus respectivas subclases) y 8 (objeto pdf), donde también se muestran los atributos de dichos objetos. Puede observarse en las figuras antes indicadas como la representación de conceptos en IC es muy similar a la de la metodología OO en IS, si bien los conceptos (u objetos) de IC no disponen de métodos, sino únicamente de atributos, por lo que parece más correcto usar en IC la denominación ‘concepto’ a la de ‘objeto’, precisamente por ausencia de métodos en los primeros; no obstante por simplicidad, y ya que para representar los conceptos en este trabajo se ha utilizado una herramienta de IS para OO, nos referiremos en este trabajo indistintamente a los conceptos como ‘conceptos’ u ‘objetos’

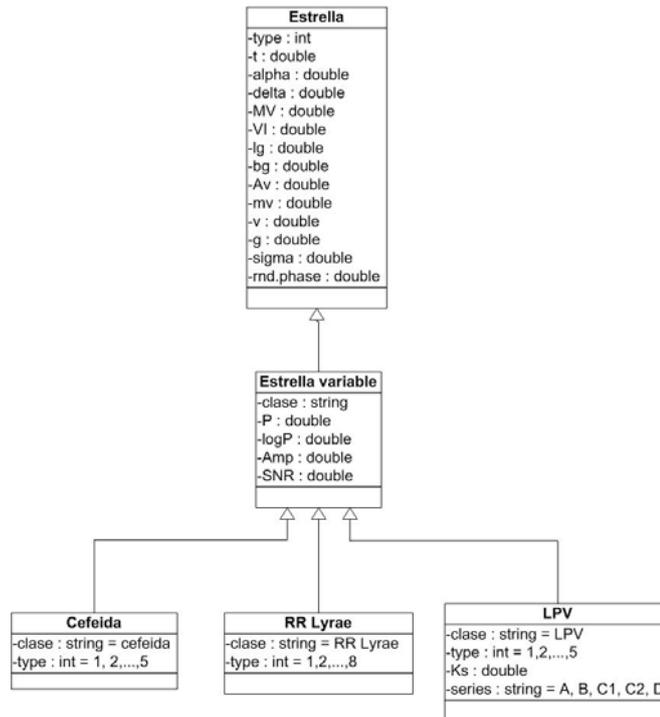


Figura 6: Objetos del dominio; objeto estrella, en el que pueden verse también las subclases del mismo estrella variable, y Cefeida, RR Lyrae y LPV.

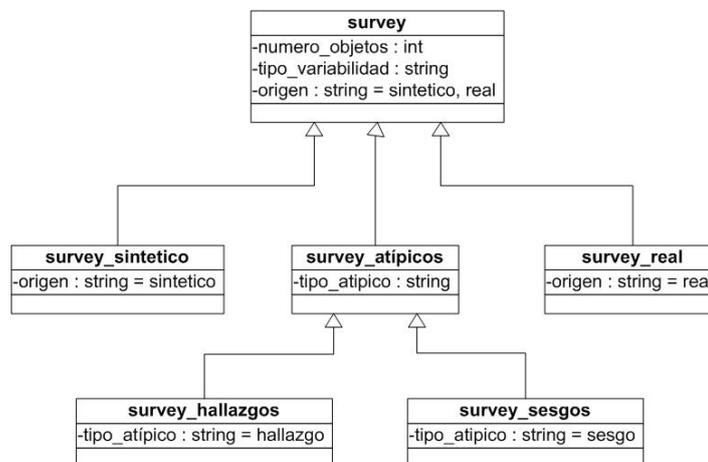


Figura 7: Objetos del dominio; objeto “survey”, en el que pueden verse también las subclases del mismo “survey_sintético”, “survey_real” y “survey_atípicos”, con las subclases de este último: “survey_sesgos” y “survey_hallazgos”.

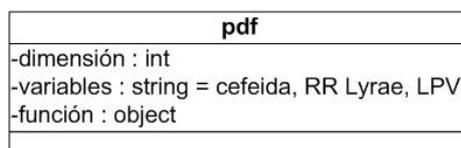


Figura 8: Objetos del dominio; objeto pdf.

Como atributos, aparece el espacio de parámetros de cada una de las familias de variables definidas en el punto 3.3, siendo la mayoría autoexplicativos, por lo que no necesitan descripción adicional. El objeto pdf se caracteriza mediante un objeto que es una función de densidad de probabilidad (atributo ‘función’).

Todos estos objetos presentan relaciones entre ellos, y son los que sirven de ‘entrada’ o ‘salida’ a las inferencias que se mostrarán posteriormente. Dichas relaciones se muestran en las figuras 9 y 10; las relaciones y sus atributos (cuando los poseen) son sencillos, por lo que no necesitan explicación adicional.

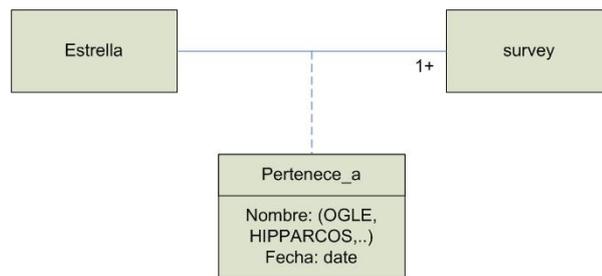


Figura 9: Relaciones del dominio; relación “pertenece_a”, en la que pueden verse también sus atributos, nombre (OGLE, Hipparcos, etc.) y fecha (del “survey”).

La definición mediante lenguaje formal de la relación ‘pertenece_a’ se muestra a continuación:

```

Binary-Relation pertenece_a;
  Inverse:
  Argument-1: estrella;
    Cardinality: 0-1;
  Argument-2: pdf;
    Cardinality: ANY;
  Attributes:
    Nombre: set of BB.DD;
    Fecha: date;
END Binary-Relation pertenece_a;
  
```

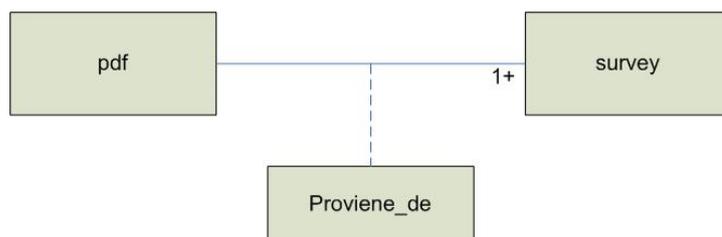


Figura 10: Relaciones del dominio; relación “genera”. Un mismo “survey” puede generar 1 o más pdf, dependiendo de los parámetros del espacio considerados.

Y la correspondiente a la relación ‘genera’ es la siguiente:

Binary-Relation proviene_de;
Inverse:
Argument-1: pdf;
Cardinality: 0-1;
Argument-2: survey;
Cardinality: 0-1;
Attributes:
END Binary-Relation proviene_de;

3.5 Conocimiento Inferencial

Finalmente, resta describir el nivel más bajo de descomposición funcional, en base a unidades básicas de procesamiento de información o inferencias. El modelado del conocimiento inferencial que se ha utilizado en este trabajo, está basado en [Schreiber-99] y [Palma-00]. La descomposición inferencial de la primera tarea “Generar muestras controladas” (o generación de “surveys”), que pueden ser sintéticos o no, se muestra en la figura 11.

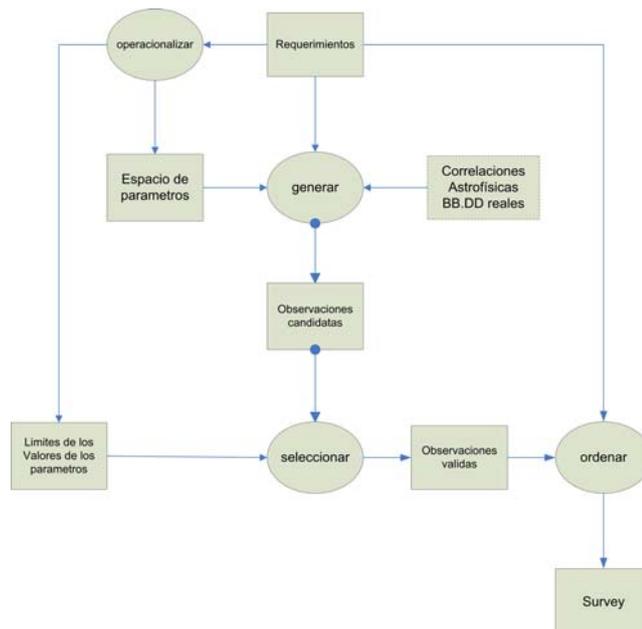


Figura 11: Conocimiento inferencial: subtarea “Generar muestras controladas”. Puede apreciarse que el esquema es muy similar al de la tarea genérica de Síntesis, puesto que realmente esta subtarea constituye una aplicación particular de la misma.

- **Requerimientos:** Constituyen los requisitos del “survey” a generar: si es sintético o real, tipo de estrellas que componen el “survey”, número de estrellas del survey, estrategia de ordenamiento, y límites deseados de cada uno de los componentes del espacio de parámetros, en función del uso que dicho “survey” tenga en las subtarefas posteriores. Estos requerimientos constituyen la entrada de la tarea.

- Operacionalizar: En este paso se define el espacio de parámetros del “survey” a sintetizar, así como los límites o restricciones de los mismos a observar en la selección de los valores finales.
- Generar posibles observaciones candidatas: Se generan / filtran todos los posibles valores de cada uno de los parámetros del espacio en el que se desea trabajar; dado que estos pueden ser infinitos, este paso debe de ser lo suficientemente específico para hacer que el número de valores generados sea manejable.
- Seleccionar observaciones válidas: se seleccionan las observaciones útiles para generar el “survey” final. Dicho “survey” es la salida de esta subtarea.
- Ordenar: clasifica las observaciones válidas generadas/filtradas según los criterios requeridos.
- Correlaciones Astrofísicas / BB.DD: Constituyen la base de conocimientos en función de la cual se generan los valores de los parámetros del “survey”, usándose las correlaciones para los “surveys” sintéticos, y las BB.DD para los surveys reales.

Task Generar muestras controladas;

Goal:

Genera “surveys” sintéticos / reales partiendo de requerimientos o BB.DD reales;

Roles:

Input:

BB.DD (OGLE, Hipparcos...);
Restricciones;
Especificaciones;

Output:

“survey_real” / “survey_sintético”

Spec:

Genera estrellas variables sintéticas del tipo especificado, u obtiene por filtrado un subconjunto de estrellas de una variabilidad especificada partiendo de una BB.DD real;

End Task Generar muestras controladas;

Task-Method proponer-revisar-filtrar;

Realizes: Generar muestras controladas;

Decomposition:

Inferences: Operacionalizar, generar, seleccionar, ordenar;

Roles:

Intermediate:

Límites valores de parámetros: “límites que deben de ser cumplidos”;
Espacio de parámetros: “definición de los parámetros a sintetizar o filtrar”;
Observaciones candidatas: “observaciones candidatas a incorporar al resultado”;
Observaciones válidas: “Observaciones que cumplen los requerimientos”;

Control-Structure:

operacionalizar (requerimientos → Espacio de parámetros + límites valores parámetros);
generar (requerimientos + espacio de parámetros → observaciones candidatas);
seleccionar (observaciones candidatas + límites valores parámetros → observaciones

válidas);

ordenar (observaciones válidas + requerimientos → survey);

END Task-method;

En el caso de la segunda subtarea “Aplicar Algoritmos de QA a surveys” o cálculo de la pdf, su descomposición inferencial es la mostrada en la figura 12. Sus elementos son los siguientes:

- Requerimientos: constituyen los requisitos de la pdf a generar, indicando el número de variables, las dimensiones, etc. Indica también el “survey” de partida. Estos requerimientos actúan como entrada del sistema.
- Especificar Métodos de cálculo de pdf: En este paso se analizan todos los métodos posibles de obtención de la pdf, en función de los requerimientos de partida.
- Seleccionar método de cálculo de la pdf: este paso efectúa la selección del método de cálculo de la pdf más adecuado en función de las restricciones impuestas.
- Función de transferencia “Obtener”: esta función de transferencia (representada en un recuadro con las esquinas redondeadas), se encarga de obtener el “survey” a partir del cual se obtendrá la pdf. Dicho “survey” puede obtenerse de una BB.DD real (OGLE, Hipparcos...), o puede proceder de la subtarea 1.
- Obtener pdf: en este paso, con el “survey” de observaciones a procesar disponible y el método de cálculo seleccionado, se obtiene la pdf objetivo.

Como puede observarse, en el desarrollo de esta tarea, aparece el concepto de función de transferencia, importante concepto, ya que dichas funciones de transferencia permiten el intercambio de información hacia y desde el exterior. Para llevar a cabo esta subtarea, es necesario disponer de conocimiento relativo a métodos estadísticos para seleccionar el método a utilizar en función de los requerimientos establecidos; la implementación de dichos métodos es objeto de un trabajo aparte. En lenguaje formal, definimos tanto la subtarea, como su método, de manera que quede completamente definida:

Task Aplicar algoritmos de QA a “surveys”;

Goal:

Dado un “survey”, aplica los algoritmos de QA al mismo de acuerdo a requerimientos establecidos;

Roles:

Input:

Survey_real; Survey_sintético;
Restricciones;
Especificaciones;

Output:

pdf

Spec:

Genera pdf multivariantes partiendo de los “surveys” de entrada con un subconjunto de parámetros definidos del espacio de parámetros;

End Task Aplicar algoritmos de QA a “surveys”;

Aparece también aquí la función de transferencia “obtener”, que se define formalmente a continuación:

Transfer-function obtener;
Type: Obtain;
Roles:
Input: categorías_disponible;
Output: categoría_confirmada;
END Transfer-function obtener;

Finalmente, el método para realizar la tarea se define formalmente a continuación:

Task-Method calculo-estadistico-pdf;
Realizes: Aplicar algoritmos de QA a “surveys”;
Decomposition:
Inferences: especificar, seleccionar, obtener;
Transfer-function: obtener;
Roles:
Intermediate:
Métodos de cálculo de pdf: “conjunto de método de cálculo de pdf”;
Método de cálculo pdf: “método confirmado de cálculo de la pdf”;
Control-Structure:
especificar (requerimientos → métodos de cálculo pdf);
seleccionar (métodos de cálculo pdf → método de cálculo pdf);
obtener (surveys-disponibles → survey_confirmado)
obtener (survey_confirmado + método de cálculo pdf → pdf);
END Task-method;

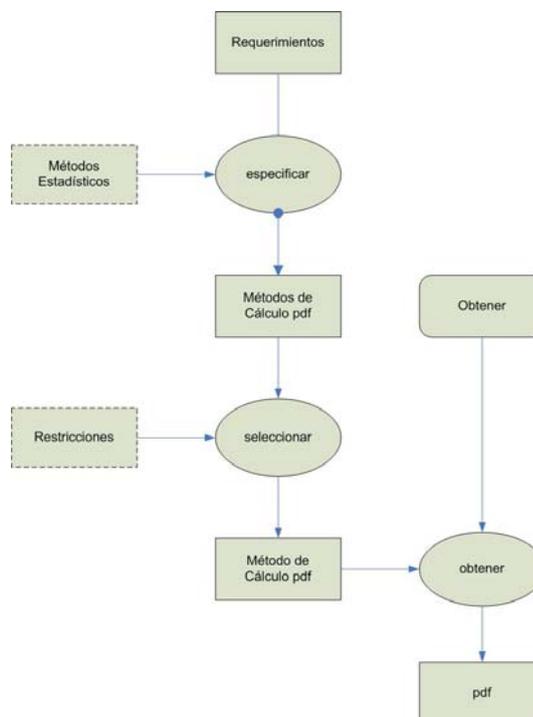


Figura 12: Conocimiento inferencial: subtarea “Aplicar algoritmos de QA a surveys”. Puede apreciarse que también en este caso el esquema utilizado es muy similar al de la tarea genérica de Síntesis, puesto que realmente esta subtarea constituye también una aplicación específica de la misma.

En cuanto a la tercera tarea: “Evaluación de la hipótesis nula H_0 ”, la descomposición inferencial es la mostrada en la figura 13. Esta es una tarea relativamente simple, tal y como se deduce de la descripción de la misma:

- Casos: representan la entrada a la tarea, en particular, las pdf que se comparan de acuerdo con la hipótesis nula H_0 , que en este caso es la indicada en el punto 3.1 apartado 3.
- Función de transferencia “obtener”: permite la obtención del método de evaluación multivariante para la veracidad de H_0 .
- Evaluar decisión: Una vez obtenido el método estadístico a utilizar, se aplica y se obtiene la decisión final, que en este caso será un “verdadero” o “falso” dependiendo de que H_0 sea verdadera o no lo sea, utilizando los métodos estadísticos de la Base de Conocimientos

Inicialmente, el método de evaluación de H_0 será en principio el estadístico T^2 de Hotelling, si bien esto no es relevante para la confección del modelo de conocimiento (aunque si lógicamente para la fase de implementación).

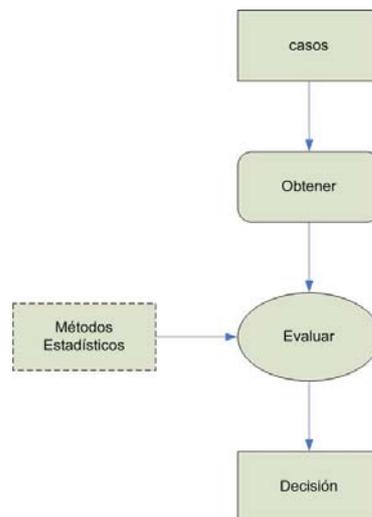


Figura 13: Conocimiento inferencial: sub tarea “Evaluar hipótesis nula H_0 ”. Puede apreciarse que en este caso el esquema base utilizada es el de la tarea genérica de Evaluación.

Task Evaluar hipótesis nula H_0 ;

Goal:

Dadas dos pdf, aplica métodos estadísticos para determinar si se cumple H_0 ;

Roles:

Input:

pdf1, pdf2;

Output:

decisión

Spec:

Dadas 2 pdf multivariantes determina si provienen de la misma población;

End Task Evaluar hipótesis nula H_0 ;

Task-Method prueba-de-hipótesis;

Realizes: Evaluar hipótesis nula H_0 ;

Decomposition:

Inferences: evaluar;

Transfer-function: obtener;

Roles:

Intermediate:

Control-Structure:

obtener (pdf-disponibles \rightarrow pdf_confirmada1 + pdf_confirmada2)

obtener (pdf_confirmada1 + pdf_confirmada2 \rightarrow decisión);

END Task-method;

Por último, resta la descomposición inferencial de la última subtarea: “Detectar Observaciones Atípicas”, que es la más compleja, según se describe a continuación:

- Caso: la subtarea recibe el “survey” sobre el que se pretenden detectar las observaciones atípicas; obsérvese que una observación atípica es un objeto de la clase “Estrella variable” en el que el valor de alguno de sus parámetros es anormalmente alto o bajo. El “survey” bajo análisis constituye la entrada de la tarea.
- Seleccionar parámetros: se seleccionan los parámetros de interés del “survey” de entrada para los que se desean detectar valores atípicos.
- Especificar límites atípicos: para los parámetros de interés seleccionados, se determinan, en base a los métodos estadísticos de la Base de Conocimiento los límites relevantes para calcular los atípicos. En el apéndice C se muestra un método muy sencillo de cálculo de valores atípicos.
- Especificar valor SNR umbral: se determina el valor SNR umbral para cada observación sobre la que se analizan los parámetros de interés, umbral que permite distinguir un *hallazgo* de un *sesgo* (ver punto 3.1 para más detalles).
- Comparar diferencias: Conociendo los límites de atípicos de cada parámetro de interés, se calculan las diferencias para cada observación en cuestión en el “survey” bajo análisis.
- Clasificar diferencias y obtener *hallazgos* y *sesgos*: Con el valor SNR umbral establecido, se clasifican las diferencias para determinar si estamos antes un *hallazgo* o un valor atípico debido a *sesgo*, añadiendo a la lista de resultados del análisis de las anteriores observaciones el resultado de la observación en curso.
- Función de transferencia Presentar: Se presentan las observaciones mediante dos conjuntos de observaciones, según sean son *hallazgos* o no. Es decir: se agrupan los resultados en dos “surveys”, uno de valores atípicos que se han clasificado como *sesgos*, y otro de valores atípicos clasificados como *hallazgos*.

La descomposición inferencial de esta subtarea se muestra en la figura 14, y la descripción formal de esta subtarea así como de su método, se indica a continuación:

Task Detectar observaciones atípicas;

Goal:

Dado un objeto “survey”, aplica métodos estadísticos para determinar si existen observaciones atípicas, clasificando estas en *hallazgos* o *sesgos*;

Roles:

Input:

survey;

Output:

Survey_hallazgos, survey_sesgos

Spec:

Dado un “survey”, determina si existen observaciones atípicas, y las clasifica y agrupa en dos nuevos objetos “survey”: “survey_hallazgos” y “survey_sesgos”;

End Task Detectar observaciones atípicas;

Task-Method clasificación-guiada-por-datos;

Realizes: Detectar observaciones atípicas;

Decomposition:

Inferences: seleccionar, comparar, especificar, clasificar;

Transfer-function: presentar;

Roles:

Intermediate:

Nueva observación: “análisis observación a observación del survey dado;

Parámetros: “definición de los parámetros a analizar”;

Límites atípicos: “límites de caracterización de observaciones atípicas”;

Diferencia: “diferencia entre el valor analizado y los límites de atípicos”;

Valor SRN umbral: “valor SNR que separa hallazgos de sesgos”;

Hallazgo: “survey de hallazgos survey_hallazgos”;

Sesgo: “survey de sesgos survey_sesgos”;

Control-Structure:

seleccionar (caso → parámetros);

especificar (parámetros → límites atípicos + valor SNR umbral);

Repeat

comparar (nueva observación + límites atípicos → diferencia);

clasificar (diferencia + valor SNR umbral → hallazgo + sesgo);

Until nueva observación = NULL;

presentar (hallazgo);

presentar (sesgo);

END Task-method;

La definición de las inferencias elementales, puede encontrarse en [Schreiber-99] .Con esto quedan ya especificadas las cuatro subtareas, y por tanto la tarea de alto nivel que debe de realizar el SE que se está diseñando.

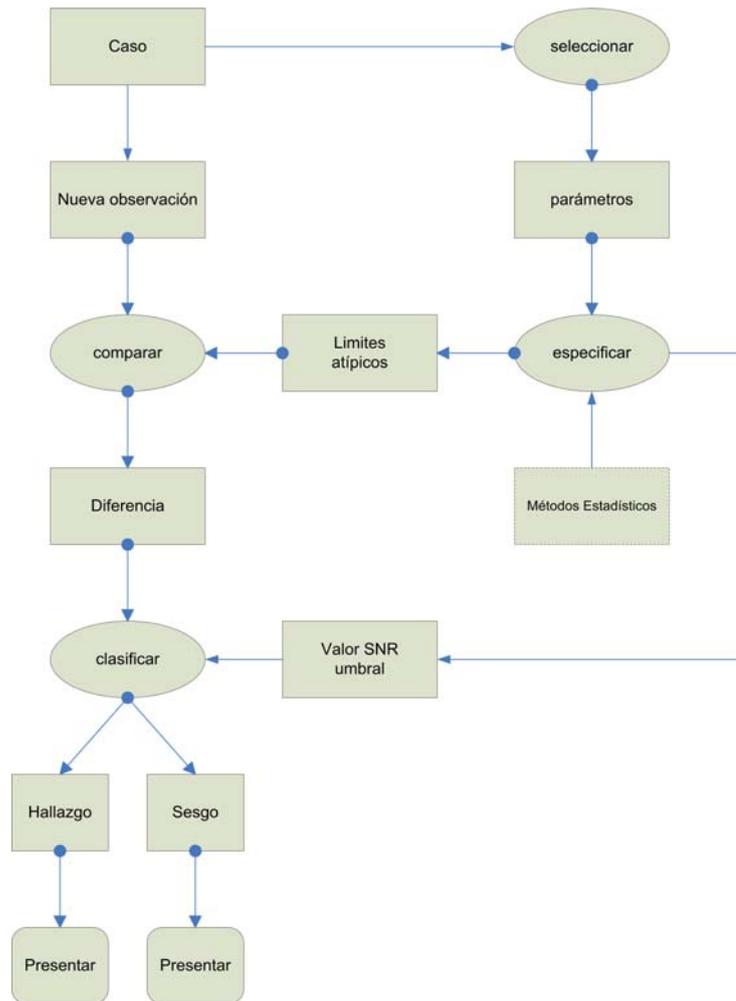


Figura 14: Conocimiento inferencial: subtarea “Detección de Observaciones Atípicas”. Puede apreciarse que en este caso el esquema en el que se ha basado es el de la tarea genérica de Monitorización, con algunas modificaciones relativas al cálculo de la SNR umbral.

Capítulo 4: *IMPLEMENTACIÓN PARCIAL DEL SISTEMA EXPERTO*

4.1. Introducción

Como ya se ha indicado en puntos anteriores, no es objeto de este trabajo la implementación completa del SE diseñado en el capítulo anterior, si bien si se incluye en el mismo la implementación de la primera subtarea: “Generación de muestras Controladas”, o generación de “surveys” sintéticos, de ahí el uso del adjetivo “parcial” que aparece en el encabezamiento de este capítulo.

La implementación de esta tarea, como ya se ha indicado, se ha apoyado en las entrevistas realizadas al experto, en el diseño conceptual descrito en el capítulo anterior, y en la aplicación de conocimientos estadísticos, que junto con los conocimientos astrofísicos específicos de los objetos variables y la programación (en este caso en ‘R’), conforman los tres pilares sobre los que se apoya la realización del diseño y realización completa del SE objeto de este trabajo.

4.2. Estructura del código ejecutable

El código de este trabajo está estructurado en varios ficheros escritos en ‘R’. Para cada uno de los “surveys” que se van a calcular para los tres tipos de estrellas variables modeladas, aparece un fichero denominado “*LMC-SMC-GLX-xxxx.R*” y otro denominado “*functions-xxxx.R*” donde según el tipo de estrellas sobre las que trabaja la cadena “xxxx” puede valer “*cepheids*”, “*RR_Lyrae*” o “*Long_Period*”. El primer fichero contiene el código que obtiene mediante simulación los parámetros de las estrellas sintéticas del “survey” sobre el que se está trabajando, mientras que el segundo contiene las funciones que son llamadas desde el primero.

Los datos obtenidos en cada una de las simulaciones ejecutadas, se almacenan en el fichero “*results-xxxx.dat*”, y en un fichero de datos específico de ‘R’ denominado “*xxxx.Rdata*”, donde la cadena “xxxx” puede tomar los valores antes indicados, y que sirve para almacenar los objetos ‘R’ de una forma más eficiente y sencilla de recuperar. El primer fichero contiene todo el espacio de parámetros del “survey” obtenido mostrándolos por columnas, de manera que cada fila representa cada una de las estrellas sintéticas obtenidas con sus correspondientes parámetros. Dado el volumen de datos de dichos ficheros, se recomienda examinar directamente los mismos (con cualquier editor de texto; en el Anexo E pueden encontrarse los ficheros de resultados finales obtenidos para los tipos de variabilidad simulados.

Existen además un par de ficheros con resultados intermedios (cuyo contenido se describirá más adelante) denominados “*av-xxxx.dat*” y “*gc-xxxx.dat*”, donde “*xxxx*” es la cadena antes descrita. Desde el fichero de configuración “*config.dat*”, puede establecerse que después de cada ejecución se borren o no estos ficheros intermedios. El citado fichero de configuración, permite que puedan modificarse en función de las variantes que se desee obtener cada una de las ejecuciones, y se usa en todos los tipos de estrellas variables sobre las que se trabaja. También hay dos ficheros de código ‘R’ denominados “*plotset1-xxx.R*” y “*plotset2-xxxx.R*” que se utilizan para validar los resultados obtenidos por los ficheros de código antes descritos; su entrada es el fichero “*results-xxxx.dat*” ya descrita. Nuevamente, la cadena “*xxxx*” tiene el mismo significado que en los casos anteriores. Finalmente, hay un fichero ejecutable denominado “*extinc4.exe*” que es llamado desde el fichero “*LMC-SMC-GLX-xxxx.R*” para el cálculo de las extinciones; más adelante se describirá la utilidad de este fichero en detalle.

Los programas ‘R’ contenidos en los ficheros antes descritos, se han ejecutado en el entorno de programación ‘**eclipse Helios**’ con el “plugin” **statET 0.10.0**, que permite una integración total con el editor ‘**eclipse**’ y una cómoda interacción con el lenguaje de programación ‘R’. En un sentido amplio, ‘R’ es un lenguaje de programación con un conjunto de procedimientos implementados que permiten realizar tareas específicas muy diversas, que van desde la aplicación de funciones estadísticas a la generación de gráficos de altas prestaciones. **Todos los ficheros (de código y datos) mencionados en este punto, están disponibles en el anexo E esta memoria** para su examen.

‘R’ es un lenguaje orientado a objetos. Aunque existen algunas interfaces gráficas para ‘R’ como ‘*Rcommander*’, es muy recomendable utilizar ‘R’ como un lenguaje en vez de tratarlo como un programa estadístico convencional. Como entorno de trabajo, ‘R’ ofrece varias ventajas:

- Sus posibilidades gráficas son excelentes.
- Es muy flexible. Los procedimientos (estadísticos o no) estándar se pueden aplicar con solo utilizar el comando apropiado. Además, existen multitud de librerías (a las que se denominan paquetes) programadas por los usuarios de todo el mundo para llevar a cabo procedimientos específicos.
- Aunque R comenzó siendo una herramienta estadística, en la actualidad se utiliza para infinidad de áreas no relacionadas con la estadística, como: análisis matemático, física, química, simulación, etc.
- Es libre en el sentido en que podemos acceder al código escrito por otros usuarios y modificarlo libremente. A pesar de que ‘R’ viene sin garantía alguna, la mayor parte del código de ‘R’, o por lo menos, el código más comúnmente utilizado por los usuarios, ha sido meticulosamente supervisado por estadísticos y académicos de mucho prestigio de todo el mundo (el llamado “*R Coreteam*”).

- Podemos además programar nuestros propios procedimientos y aplicaciones. En la misma página desde la que se puede bajar el programa (<http://www.r-project.org/>), existe abundante documentación sobre cómo utilizarlo.
- Es gratuito

4.3. Parámetros Astrométricos (t , α y δ)

La generación de las muestras sintéticas, exige el conocimiento de los parámetros geométricos de los objetos que se van a modelar, ya que las estrellas sintéticas que forman parte de los “surveys” que se generen, deberán distribuirse geoméricamente tal y como hacen las estrellas reales en dichos objetos: como se ha visto en el punto 3.2, en el espacio de parámetros de estos “surveys” sintéticos aparecen las coordenadas posicionales (tanto ecuatoriales como galácticas) de cada una de las estrellas que lo forman; dichas coordenadas serán utilizadas en la obtención de otros parámetros (como la extinción) que también forman parte del espacio de parámetros indicado. Por otro lado, las estrellas generadas forman parte de diferentes estructuras: vía láctea, LMC ó SMC, y por tanto, también es necesario disponer un modelo de dichos cuerpos.

4.3.1. Gran Nube de Magallanes (LMC)

La Gran Nube de Magallanes es una galaxia enana con forma de disco, satélite de la vía láctea, y como ya se ha indicado miembro del denominado Grupo Local. Sus datos de observación basados principalmente en [Winberg-00], son los indicados en la Tabla 2 mostrada a continuación:

Parámetro	Valor
i	30.7°
θ	151.0°
h	0.35 kpc
λ	1.6 kpc
d	55.2 kpc
D_r	2.3 kpc
D_z	1.64 kpc

Tabla 2: *Parámetros Observacionales de la LMC: inclinación (i), posición (θ), factor de escala vertical (h), factor de escala radial (λ), distancia (d), longitud radial (D_r) y longitud vertical (D_z)*

El modelado de la LMC se realiza utilizando la expresión:

$$\rho(r, z) = D_z \cdot e^{\frac{-|z|}{h}} \times D_r \cdot e^{\frac{-r}{\lambda}}$$

Donde r es la distancia radial de cada punto (estrella) al centro del disco, y z es la altura de cada punto (estrella) sobre el plano horizontal en el que está contenido el disco que representa la LMC. Para cada valor de z y r se compara un número aleatorio con el valor de la función de distribución de z y r , y si es superior, se acepta dicho valor de z o r : de este modo, tanto los valores resultantes de r como los de z siguen una distribución exponencial.

Hay que tener en cuenta sin embargo que los valores de z sobre el plano del disco pueden ser positivos o negativos, por lo que será necesario cambiar el signo de la mitad de ellos que se obtengan durante la generación de las estrellas sintéticas.

Una vez aplicada la expresión anterior para r y para z , transformaremos estas coordenadas en otras x, y, z cartesianas con su origen en el centro del disco. Para que realmente estas nuevas coordenadas se correspondan con las de un disco, a cada valor de r obtenido se le asignan unos valores x, y, z , tales que:

$$r = \sqrt{x^2 + y^2}$$

Esto se hace generando unos valores de ángulo (digamos φ) aleatorios entre 0 y 2π , y haciendo:

$$x = r \times \text{sen} \varphi$$

$$y = r \times \text{cos} \varphi$$

De este modo, llegados a este punto disponemos de las coordenadas x, y, z que recordemos, se corresponden con unos ejes cartesianos con origen el centro del disco que representa la LMC, y cuyos ejes x e y están contenidos en el plano de dicho disco.

El siguiente paso, es transformar las coordenadas x, y, z , a coordenadas terrestres α, δ, t o ascensión recta, declinación y distancia respectivamente; para ello, nos valdremos del apéndice de [Winberg-00], expresiones A3 y A2; según dicho apéndice, las coordenadas x, y, z que se han obtenido más arriba, se corresponden con las coordenadas x', y', z' de las ecuaciones A3 de [Winberg-00]; mediante una simple inversión de matrices, pueden obtenerse las coordenadas x_0, y_0, z_0 ; estas últimas se corresponden con x, y, z pero giradas respecto del eje z un ángulo θ en el sentido contrario a las

agujas del reloj, y respecto del eje x un ángulo i (también denominado inclinación) siendo los valores de θ e i los indicados en la Tabla 2.

Una vez obtenidas las coordenadas x_0, y_0, z_0 hay que pasar a coordenadas terrestres, para lo que es necesario invertir el sistema A2 mostrado en [Winberg-00], obteniendo finalmente α, δ, t . El código que realiza todas las transformaciones descritas, están en el fichero “*LMC-SMC-GLX-cepheids.R*”, como se describió en el punto 4.2 con más detalle.

4.3.2. Pequeña Nube de Magallanes (SMC)

La Pequeña Nube de Magallanes es una galaxia enana irregular cercana a la Vía Láctea, y perteneciente al Grupo Local. Sus datos de observación [Subramanian-11] son los siguientes indicados en la tabla 3. Como se deduce de la ausencia de los valores D_z y h de la tabla anterior, la SMC se modelará como una esfera de radio 1.07 kpc, y por tanto, solo será necesario generar los valores de r de las estrellas sintéticas que la conforman, de manera completamente análoga a como se hacía en el caso de la LMC. Una vez obtenidos los valores de r , se generan unos valores de ángulo φ, ψ aleatorios entre 0 y 2π , y haciendo:

$$x_0 = r \cdot \cos \varphi \cdot \text{sen} \psi$$

$$y_0 = r \cdot \cos \varphi \cdot \cos \psi$$

$$z_0 = r \cdot \text{sen} \varphi$$

Obtenemos x_0, y_0, z_0 ; para pasar de estas últimas coordenadas a α, δ, t , se procede de igual manera que en el caso de la LMC.

Parámetro	Valor
i	0°
θ	0°
λ	1.3 kpc
d	61.0 kpc
D_r	1.07 kpc

Tabla 3: *Parámetros Observacionales de la SMC: inclinación (i), posición (θ), factor de escala radial (λ), distancia (d), radio (D_r).*

4.3.3. La Vía Láctea

Nuestra Galaxia La Vía Láctea, es una galaxia de tipo SBb ó SBbc, con una estructura cuyos elementos más importantes se muestran en la figura 4, y que son:

- Disco: El sol está aproximadamente unos 27 pc por encima del plano galáctico. La distribución de estrellas en la vecindad solar puede aproximarse por dos exponenciales que dan lugar al *disco fino* y al *disco grueso*; ambos discos pueden diferenciarse con facilidad; la metalicidad de las estrellas del disco pueden variar entre -0.7 y -1.0, aunque pueden darse valores extremos de hasta -2.0. En ambos discos predominan las estrellas de población I (jóvenes, y con alto contenidos en metales).
- Bulbo central: tiene un tamaño de unos 3.5 kpc y una relación axial de $\approx 1.0 \text{ kpc}::0.3 \text{ kpc}::0.3 \text{ kpc}$ en la dirección 20° - 35° desde la línea Sol-Centro hacia el primer cuadrante. Hay también una fuerte evidencia de la presencia de una barra en el bulbo; tiene una luminosidad $L_{\odot} = 10^{10}$ (banda K). En el bulbo se han hallado estrellas del tipo RR-Lyrae, lo que implica una edad mayor de 10 Gyr. Rota a una velocidad de 50 km/s/kpc con un periodo de unos 10^8 años, y los valores de metalicidad [Fe/H] pueden varían entre -1.8 y +0.2. Contiene estrellas principalmente de población II.
- Halo Galáctico: es una esfera que se extiende hasta los 100 kpc compuesta principalmente por estrellas de población II (viejas, y pobres en metales). En el halo galáctico se encuentran hasta 130 cúmulos globulares de estrellas, y abundan las enanas blancas.
- También existe una estructura (la más exterior) denominada Corona, que puede llegar hasta más allá de los 100 kpc

En este trabajo se modelarán ambos discos (grueso y fino) y el bulbo galáctico, no considerándose el Halo galáctico ni la Corona. Puede encontrarse una descripción más detallada de los discos finos y gruesos de las galaxias en general en [Yoachim-06]. Los parámetros característicos de los discos son los mostrados en la tabla 4.

Parámetro	Disco Fino	Disco Grueso
i	89.8°	89.8°
θ	121.3°	121.3°
h	0.325 kpc	0.86 kpc
λ	2.8 kpc	3.7 kpc
D_r	15.3 kpc	18.5 kpc
D_z	1.9 kpc	4.3 kpc

Tabla 4: *Parámetros Observacionales de los discos galácticos: inclinación (i), posición (θ), factor de escala vertical (h), factor de escala radial (λ), longitud radial (D_r) y longitud vertical (D_z).*

Los parámetros característicos del bulbo son los mostrados en la tabla 5 que se representa a continuación.

Parámetro	Valor
S_x	2.044
S_y	2.044
S_z	0.681
D	6.13 kpc
d	8.0 kpc

Tabla 5: *Parámetros Observacionales del bulbo galáctico: dispersiones axiales (S_x , S_y , S_z), longitud (D) y distancia (d).*

4.4. Cálculo de Extinciones (A_v) y Magnitudes Absolutas Extinguidas (mv)

Para el cálculo del parámetro A_v , que representa la extinción, es preciso disponer de las coordenadas galácticas lg y bg y de la distancia t . Una vez guardadas las coordenadas galácticas y las distancias de las estrellas sintéticas en una matriz (en esta trabajo dicha matriz se denomina ‘P.survey’) donde se almacenan los parámetros obtenidos, se calcula la extinción correspondiente ejecutando el programa externo “*extinc4.exe*”, que utiliza como

argumento los datos del fichero “gc-xxxx.dat”; con los datos de dicho fichero se calcula la extinción de cada estrella sintética, que el propio programa almacena en el fichero “av-xxxx.dat” ya mencionado en el punto 4.2; los valores de este fichero son leídos luego e incorporados a la matriz de parámetros ‘P.survey’ a través del parámetro **Av**.

Con este parámetro, es ya posible calcular la *magnitud absoluta extinguida* **mv** de cada una de las estrellas sintéticas generadas, tal y como se indicó en el punto 3.1; este valor, se añade también a la matriz ‘P.survey’.

Es importante resaltar que el subíndice del parámetro **Av** hace referencia a que dicha extinción está calculada en la banda V, única para la cual en principio son válidos los resultados del citado programa. El programa “*extinc4.exe*” ha sido escrito por Eduardo B. de Amores, y utiliza un modelo tridimensional descrito en [Amores-05] para calcular el valor de extinción galáctico debido tanto al gas como al polvo interestelar en la banda V, que depende de la distancia al centro galáctico y de la altura al plano galáctico: de ahí la necesidad de disponer de los parámetros **t**, **lg** y **bg**. La extinción también depende del número de brazos de la galaxia que se atraviesan (aumentando en pasos), si bien varía lentamente en las zonas entre dichos brazos.

La estimación del valor de extinción es esencial para el correcto cálculo de distancias, para la corrección de los colores percibidos de los objetos astrofísicos, para el conteo de estrellas y para la realización de modelos de brillo para nuestra galaxia, entre otras aplicaciones.

4.5. Cálculo de otros parámetros de GAIA

GAIA determina varios tipos de parámetros estelares: astrométricos, fotométricos y espectroscópicos de los que es necesario conocer los errores medios que se cometen en su evaluación. En particular, GAIA mide los parámetros **v**, **g**, **SNR**, **sigma** y **rnd.phase**, ya definidos en el punto 3.2.

En el código, se calcula inicialmente **v**: aplicando la expresión indicada en el punto 1.2 correspondiente a la ley periodo-luminosidad, es posible calcular la magnitud aparente **v** en la banda visible en función de la distancia de nuestra estrella sintética.

En este momento, en función de (V-I), **VI** en nuestro espacio de parámetros, se calcula el parámetro **g**, correspondiente a la magnitud de cada estrella sintética en la banda G de GAIA, mediante la expresión dada en [Jordi-10]:

$$C_1 = a + b \times C_2 + c \times C_2^2 + d \times C_2^3$$

Donde: $C_1 = G$, $C_2 = V - I$.

Finalmente, en la última parte de código del fichero “*LMC-SMC-GLX-xxxx.R*”, se calculan los parámetros **SNR**, **sigma** y **rnd.phase**.

La calidad de la medida del espacio de parámetros de GAIA depende principalmente de la relación señal ruido (**SNR** en este trabajo) de la magnitud de la estrella sobre la que se esté midiendo y del número de observaciones, que depende del tiempo que dure la misión y de la posición del satélite en el cielo. Hay otros efectos de segundo orden, como el ruido de fondo del cielo, el ruido instrumental, errores de calibración y el color de las estrellas. La radiación solar es también un elemento importante, pues genera “trampas de electrones” en los elementos CCD de las cámaras que producen distorsión, pero su influencia puede ser controlada. La tabla 5 muestra los errores involucrados. Con todos estos factores, se escribe en el código ‘R’ la función *sigG(g)*; con ella y con el valor de la amplitud **Amp** se calcula **SNR**:

$$SNR = \frac{Amp}{sigG(g)}$$

En [Bruijne-09] se describe el procedimiento de cálculo del error de paralaje estándar para una magnitud dada en banda visible V y un índice de color (V-I), el error de paralaje estándar σ_π [μ as] medio del cielo se calcula así:

$$\begin{aligned} \sigma_\pi[\mu as] &= \sqrt{(6.846 + 435.42z + 9.156z^2 - 0.298z^3 + 0.00702z^4 - 0.0000711z^5 + 7.11 \times 10^{-7} z^6) \times} \\ &\quad \times \sqrt{[0.986 + (1 - 0.986) \times (V - I)]} \\ z &= \left(\frac{4500}{TDI} \right) \times 10^{0.4 \times (G - 15)} \end{aligned}$$

Donde TDI es el número de etapas activas por elemento CCD, y que se obtiene también de la tabla 6. En el código ‘R’, la magnitud σ_π [μ as] se denomina *sigma.parallax*. El error estándar medio de la posición en el cielo σ_0 [μ as] y el error estándar medio de movimiento σ_μ [μ as] se calculan como:

$$\begin{aligned} \sigma_0 &= 0.743 \times \sigma_\pi \\ \sigma_\mu &= 0.526 \times \sigma_\pi \end{aligned}$$

Los errores estándar previstos varían a través del cielo como resultado del proceso de barrido del mismo para realizar la observación; para ello se calcula el valor de la latitud eclíptica *sen(β)*:

$$sen(\beta) = 0.9175 \times sen(\delta) - 0.3987 \times cos(\delta) \times sen(\alpha)$$

Con α , δ las coordenadas ecuatoriales. En el código ‘R’ de este trabajo, *sen(β)* se denomina *sin.beta*.

G_{\min} [mag]	G_{\max} [mag]	Gate Identification	Gate length ("TDI")
5.70	8.84	4	16
8.84	9.59	8	256
9.59	10.34	9	512
10.34	11.10	10	1024
11.10	11.47	11	2048
11.47	11.95	12	2900

Tabla 6: Valores del parámetro TDI. Ver documento GAIA-CA-TN-ESA-JDB-036.

Con ello puede calcularse el parámetro **sigma** como función de *sigma.parallax* y *sin.beta*, tal y como se hace en el código 'R' correspondiente.

El código que implementa estos cálculos se encuentra en el fichero "*LMC-SMC-GLX-xxxx.R*", el orden del cálculo de parámetros es siempre el mismo para todas las familias de estrellas variables simuladas en este trabajo.

4.6. "Survey" de Cefeidas

Con los parámetros observacionales, cálculo de extinciones, magnitud absoluta extinguida, y parámetros astrométricos de GAIA establecidos en los puntos anteriores 4.3, 4.4 y 4.5 para la LMC, SMC y Vía Láctea, estamos ya en condiciones de generar los 'surveys' necesarios para modelar los diferentes tipos de estrellas variables y estudiar el comportamiento de GAIA ante los mismos.

En el caso de las cefeidas, una vez obtenidas las distancias y coordenadas de nuestras estrellas sintéticas, el paso siguiente es el cálculo del resto de parámetros característicos de las mismas, empezando por **logP** (logaritmo del período). Para ello, y en base a [Antonello-02] la distribución del parámetro log P en LMC, SMC y Galaxia, se utiliza una función que genera valores aleatorios para obtener este parámetro partiendo de dos pdf normales truncadas, simulando la distribución bimodal que aparece en dicho artículo para este parámetro. Esta función se denomina en el código 'R' *stars.generator()*, y su parametrización se basa en [Antonello-02].

Existe también la posibilidad, en función de los parámetros de configuración del programa ya descritos, de generar el valor de este parámetro de manera aleatoria, según una distribución uniforme con los mismos límites que los establecidos para las normales truncadas antes mencionadas.

Los valores obtenidos para el parámetro $\log P$ son lo suficientemente buenos para el propósito del “survey” que se está generando, y ‘R’ permite generar los mismos con gran simplicidad y rapidez.

Finalmente, la asignación del valor **MV** se realiza mediante una llamada a la función $\log P.2.MV()$ del fichero “*functions-cepheids.R*”, cuya parametrización se basa en [Sandage-04] para la LMC, [Sandage-08] para la SMC y [Tammann-03] para la Galaxia.

En función del valor (+r/-r) que se le asigne al parámetro ‘M_V’ del fichero “*config.dat*”, la asignación del valor de **MV** en nuestro “Survey” de estrellas sintéticas, puede realizarse por el método antes indicado, o simplemente asignando un valor aleatorio mediante una función uniforme cuyos límites están entre -7.5 y -2.0. El objetivo de asignar un valor aleatorio uniforme es determinar cómo se comporta GAIA ante una asignación “realista” y antes una asignación aleatoria uniforme, determinando como ya se ha indicado en el capítulo anterior los posibles sesgos que aquella sufra.

Posteriormente, para cada estrella sintética se calcula el parámetro **VI** mediante la función ‘ $\log P.2.VI()$ ’, contenidas también en el fichero “*functions-cepheids.R*”; como en el caso de **MV**, mediante el valor (+r/-r) que se asigne al parámetro ‘V_I’ del fichero “*config.dat*”, la asignación de **logP** y **VI** puede hacerse aleatoria y uniformemente, con el mismo objetivo que en el caso del parámetro **MV**; los límites empleados en este caso son 0.2 y 1.1 para **VI**; la parametrización de esta función basa también en [Sandage-04] para la LMC, [Sandage-08] para la SMC y [Tammann-03] para la Galaxia.

Para más detalles, se recomienda estudiar el código directamente tanto del fichero “*functions-cepheids.R*” como del fichero “*LMC-SMC-GLX-cepheids.R*”, cuyo contenido ya se ha descrito en el punto 4.2.

Con las funciones vistas hasta este punto, disponemos ya de los siguientes parámetros: **MV**, **VI**, **logP** y **P** (este último medido en días). Estos cálculos se repiten para la LMC, SMC y para los anillos (grueso y fino) y bulbo galácticos. Para la LMC, el valor de los coeficientes de las distintas correlaciones entre los parámetros antes indicados se obtienen de [Tammann-03] y sobre todo de [Sandage-04]; para la SMC, de [Sandage-04] y mayoritariamente de [Sandage-08]; y para discos y bulbo galáctico de [Tammann-03] y [Sandage-04]. En el caso de la galaxia hay algunos coeficientes de las correlaciones utilizadas para los que no se ha encontrado un valor contrastado por ninguna publicación científica: en ese caso, *se asume para estos valores los mismos que los de la LMC*. Esto es debido a que las nubes de Magallanes, especialmente la LMC, han sido mucho más estudiadas que los elementos galácticos, y que para los objetivos de nuestra simulación, esta asunción es suficiente, dado que el objetivo final del trabajo es analizar los atributos de calidad de las observaciones de GAIA, y no dichas observaciones en sí mismas.

Después de los cálculos anteriores, se calcula el valor de la amplitud de la oscilación del brillo de la estrella (parámetro **Amp** del espacio de parámetros del “survey”) de

acuerdo a los datos aportados por [Sandage-04]; estos datos muestran que dicho valor depende fundamentalmente de **MV**. La función por la que se obtiene **Amp** es algo más larga, por lo que para su estudio se recomienda examinarla directamente en el fichero “*functions-cepheids.R*”

Todos los parámetros calculados, así como uno adicional indicando el objeto en el que se encuentra la estrella sintética simulada se van acumulando en una matriz paso a paso, como puede verse en el código contenido en el fichero “*LMC-SMC-GLX-cepheids.R*”. El cálculo del valor de los parámetros descritos hasta este momento es totalmente análogo para LMC, SMC y discos y bulbo galáctico, y para cada uno de estos objetos, se forma la correspondiente matriz con sus valores; adicionalmente, se añade un parámetro más **type** (se ubica en la primera posición), que indica el tipo de objeto al que pertenece la estrella sintética:

- **type** = 1 → LMC.
- **type** = 2 → SMC.
- **type** = 3 → Disco fino.
- **type** = 4 → Disco grueso.
- **type** = 5 → Bulbo Galáctico.

Una vez con las anteriores matrices de cada objeto conformadas, se combinan las de los discos y el bulbo para formar una nueva matriz correspondiente a la galaxia, en la que la contribución de cada objeto individual indicada en [Schneider-06], páginas 46 y siguientes, es la siguiente:

- Disco fino: 82.2 %
- Disco grueso: 4.1 %
- Bulbo galáctico: 13.7 %

Con ello se dispone ya de 3 matrices con los parámetros hasta ahora indicados: ‘P.lmc’, ‘P.smc’ y ‘P.glx’ para LMC, SMC y Galaxia, respectivamente. La matriz ‘P.glx’ ya contiene la mezcla de los distintos elementos galácticos combinados según la proporción antes indicada.

Finalmente, se obtiene la matriz ‘P.survey’, que como ya se ha indicado, es la resultante de colocar de manera consecutiva por columnas las tres matrices parciales anteriores, y en la que se almacenan los valores de los parámetros simulados. Tanto en aquellas matrices como en la resultante ‘P.survey’, los valores de los parámetros están colocados por filas, por una simple cuestión de comodidad.

Posteriormente se calculan las coordenadas galácticas de las estrellas sintéticas obtenidas, transformando **alpha** (ascensión recta) y **delta** (declinación) en **lg** (longitud) y **bg** (latitud) galácticas, formado con estas últimas la matriz ‘g.c’, a la que también se añade

t (distancia); tanto **alpha** como **delta** y **t**, ya están previamente en la matriz ‘P.survey’, como se desprende de los puntos anteriores. Finalmente, la matriz ‘g.c’ se almacena en el fichero “gc-xxxx.dat”, ya mencionado en el punto 4.2, donde la cadena “xxxx” vale en este caso “-cepheids”, de manera que el fichero se denomina “gc-cepheids.dat”, y después se incorporan los valores de **lg** y **bg** a ‘P.survey’. Con las coordenadas galácticas y la distancia **t**, puede ya calcularse el valor de la extinción A_v y de la magnitud absoluta extinguida **mv** según se indicó en el punto 4.4, y a continuación se calculan los parámetros astrométricos de GAIA, como se indica en el punto 4.5.

La matriz ‘P.survey’, debidamente procesada puede presentarse en pantalla, si bien para el número de estrellas calculadas, 10^5 en este caso, es poco práctico; de hecho, el código solo permite presentar los resultados por pantalla para un “survey” de no más de 5.000 estrellas; en cualquier caso se almacena en el fichero de texto “results-cepheids.dat”, y en un objeto ‘R’ en el fichero “Cepheids.Rdata”, fácilmente manipulable con ‘R’, tal y como se indicó en el punto 4.2.

4.7. Validación del “Survey” de Cefeidas

En este apartado se describirá el proceso de validación de los resultados dados por el código R contenido en los ficheros “LMC-SMC-GLX-cepheids.R” y “functions-cepheids.R”; esta validación se lleva a cabo con el código contenido en los ficheros “plotset1-cepheids.R” y “plotset2-cepheids.R”; el código de ambos ficheros se basa en las grandes capacidades gráficas de ‘R’ para realizar dicha validación, mediante la realización de gráficas que puedan ser comparadas con los resultados obtenidos bien en los artículos referenciados a lo largo de esta exposición, o bien en otros suficientemente reconocidos.

Básicamente, el código de los ficheros ‘R’ “plotset1-cepheids.R” y “plotset2-cepheids.R”, leen los datos resultantes de la ejecución del código de los ficheros “LMC-SMC-GLX-cepheids.R” y “functions-cepheids.R” que están contenidos en el fichero de texto “results-cepheids.dat”, y generan las gráficas que muestran las correlaciones existentes entre los distintos parámetros sintetizados para cada una de las estrellas del ‘survey’ obtenido. Este procedimiento no es exclusivo del ‘survey’ de las cefeidas, sino que también se aplica al de las estrellas sintéticas RR-Lyrae y LPV. El objetivo de esta operativa es comparar los resultados gráficos obtenidos con los gráficos de los artículos científicos usados como fuente de los datos de partida para generar los “surveys” que se obtengan.

Los datos del fichero “results-cepheids.dat” se almacenan durante la ejecución del código de validación en la variable ‘data’, y a partir de ahí se trazan las gráficas de las correlaciones existentes en el espacio de parámetros. Se han obtenido 100.000 estrellas sintéticas por cada “survey” (si bien el número de estrellas es un parámetro del programa).

4.7.1. Histograma de periodos

Inicialmente se confecciona un histograma de los valores obtenidos para el parámetro **logP** para todas las cefeidas, de manera que pueda observarse la distribución de dicho parámetros. Para el caso de las cefeidas, dicho histograma se muestra en la figura 15.

Puede observarse que los periodos se distribuyen mayoritariamente entre los valores $\log P = 0.0$ (1.0 días) y $\log P = 1.0$ (10.0 días), según se muestra en la figura 15, que es la variación de periodo generalmente aceptada para este tipo de estrellas variables según [Sandage-08] [Tammann-03] [Sandage-04] (más de 78.954 de un total de 100.000 estrellas que contiene el “survey” de cefeidas obtenido).

Podemos valernos también de ‘R’ para obtener un resumen de la columna correspondiente al logaritmo de los periodos de nuestras estrellas sintéticas:

```
> summary(logP)
  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
-0.4972 0.2696 0.5471 0.5670 0.8233 1.7500
```

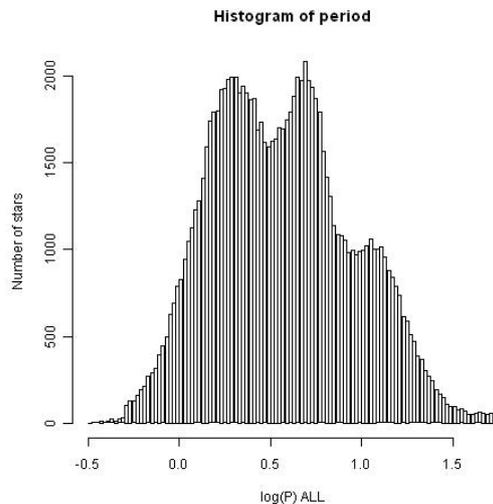


Figura 15: Histograma de periodos en escala logarítmica de todas las Cefeidas simuladas (en LMC, SMC, discos y bulbo).

Que son unos valores razonables para el logaritmo del periodo, si bien puede observarse que se han generado también algunos valores extremos anormales: periodos correspondientes a 0.32 días ($\log P = -0.4970$) y 53.26 días ($\log P = +1.750$), respectivamente, si bien no puede afirmarse que sean valores atípicos, según el criterio indicado en el apéndice C para calcular valores atípicos, que es el que se ha utilizado en este trabajo, ya que se han utilizado funciones truncadas para su generación, y evitar

dichos valores atípicos en la medida de lo posible en las simulaciones realizadas en el código.

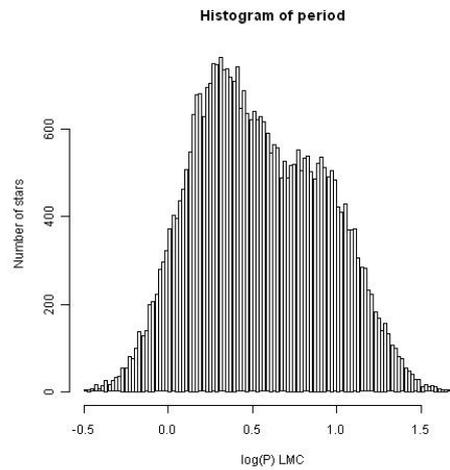


Figura 15a: *Histograma de periodos en escala logarítmica de la Cefeidas simuladas pertenecientes a la LMC.*

Se han obtenido las graficas individuales correspondientes a los histogramas del parámetro $\log P$ de cada uno de los cuerpos celestes simulados: LMC (figura 15a), SMC (figura 15b) y Galaxia (figura 15c); comparados con los gráficos de la figura 2 de [Antonello-02] de la figura 15d, puede observarse que son aproximados a los representados en dicho artículo.

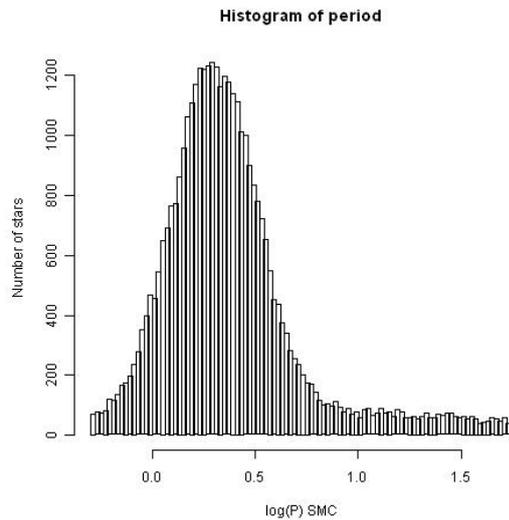


Figura 15b: *Histograma de periodos en escala logarítmica de las Cefeidas simuladas pertenecientes a la SMC.*

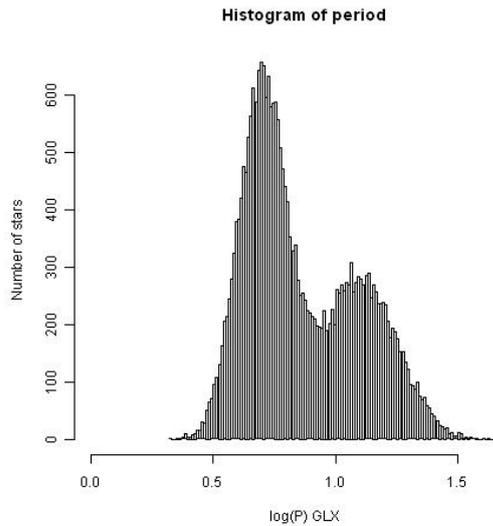


Figura 15c: *Histograma de periodos en escala logarítmica de las Cefeidas simuladas pertenecientes a la Vía Láctea (disco fino + disco grueso + bulbo galáctico).*

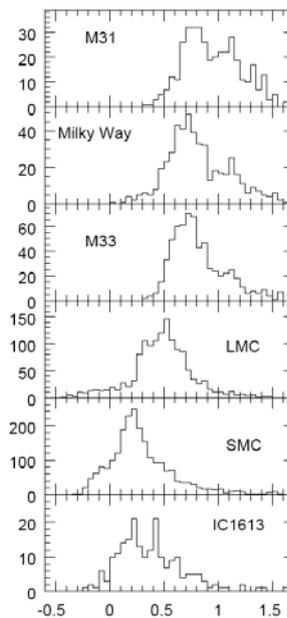


Figura 15d: *Histogramas de logP en Grupo Local; cefeidas reales [Antonello-02]*

4.7.2. Representación de las coordenadas (Coordenadas Ecuatoriales)

La siguiente operación realiza la validación de las coordenadas ecuatoriales obtenidas (ver apéndice B). Cada uno de los elementos simulados, se representa mediante un color en la gráfica de la figura 16:

- LMC:** Rojo.
- SMC:** Naranja.
- Disco Fino:** Azul.

Disco Grueso: Verde.

Bulbo Galáctico: Amarillo.

En la figura 16 pueden verse las coordenadas ecuatoriales de los distintos objetos simulados; en el caso de la galaxia (colores amarillo, verde y azul), puede observarse fácilmente la diferente proporción de cefeidas entre discos y bulbo. También se observan las estrellas correspondientes a la LMC y SMC (colores naranja y rojo).

En la simulación, se ha obtenido el mismo número de estrellas para la Galaxia, LMC y SMC. Nótese que dado que la grafica se obtiene secuencialmente, es decir: primero se dibujan las coordenadas de la galaxia (bulbo y discos), y luego las de la LMC y SMC; se da el hecho de que algunos puntos se superponen sobre los representados antes, de manera que especialmente en el caso de la galaxia (bulbo y discos), pueden verse menos puntos de color amarillo, sin que esto signifique que realmente hay menos estrellas sintéticas en el bulbo (a los que corresponde el color amarillo).

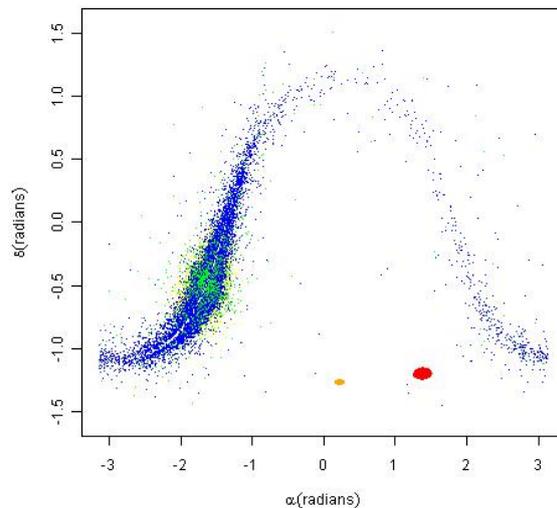


Figura 16: Representación de las Coordenadas Ecuatoriales (α , δ) de las Cefeidas simuladas: en color rojo LMC, naranja SMC, azul Disco Fino, verde Disco Grueso y amarillo bulbo galáctico.

4.7.3. Representación de las coordenadas (Coordenadas Galácticas)

En la figura 17 se muestran las coordenadas galácticas (l_g , b_g) de las estrellas sintéticas obtenidas en la simulación (ver apéndice B). Los colores y su significado son análogos a los indicados en el punto 4.7.2.

Puede apreciarse la posición de LMC y SMC respecto de la galaxia de una manera mucho más intuitiva.

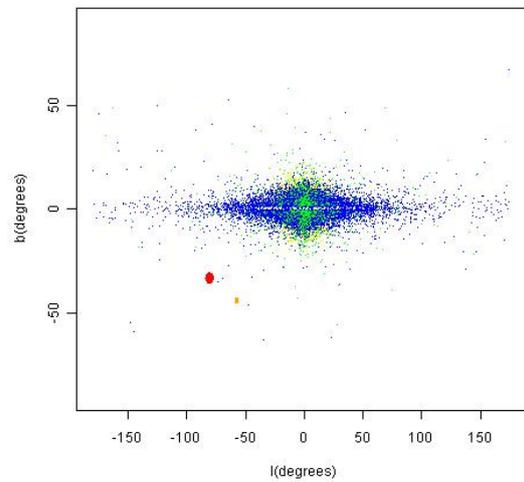


Figura 17: Representación en Coordenadas Galácticas (l , b) de las Cefeidas simuladas: rojo LMC, naranja SMC, azul Disco Fino, verde Disco Grueso y amarillo bulbo galáctico.

4.7.4. Visualización de objetos en Banda G

En la siguiente gráfica (figura 18), se representan aquellas estrellas sintéticas cuya magnitud en banda G es menor que 21, es decir: aquellas que son visibles por GAIA únicamente (recordemos que GAIA realizará un catalogo con los parámetros astrométricos de las estrellas hasta magnitud 20).

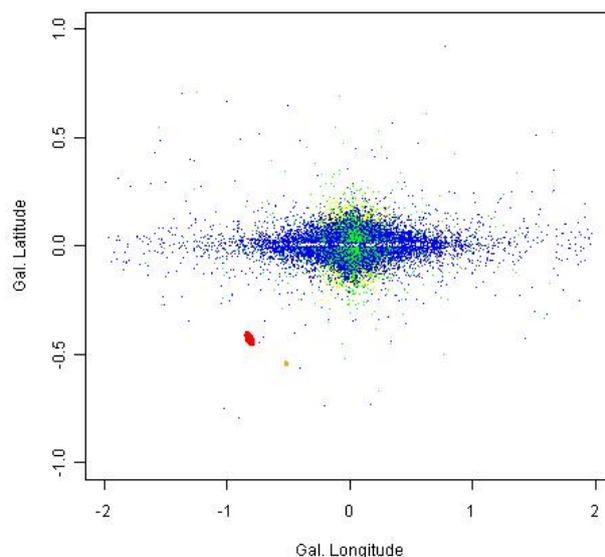


Figura 18: Estrellas Cefeidas simuladas con magnitud < 21 (banda G): Rojo LMC, naranja SMC, azul Disco Fino, verde Disco Grueso y amarillo bulbo galáctico.

Se utilizan los mismos colores para representar cada una de las estrellas en función de su localización: Galaxia, LMC ó SMC. La figura en este caso es similar a las figuras anteriores, ya que se han eliminado del “survey” aquellas estrella con $G > 21$ invisibles para GAIA; analizando los resultados obtenidos, del total de estrellas sintéticas del “survey” (que como se ha indicado son 100.000). Analizando la banda G con más detalle en el “survey” obtenido, se obtienen los siguientes datos:

```
> summary(g)
   Min. 1st Qu.  Median    Mean   3rd Qu.    Max.
 5.725 14.450 16.090 15.510 16.840 21.000
```

4.7.5. Visualización de los valores de Extinción (A_v)

La figura 19 muestra los valores del logaritmo de la extinción ($\log_{10} A_v$) en función de las coordenadas galácticas l_g , b_g (proyectadas). Los valores de extinción se representan en escala de colores (amarillo-naranja-rojo).

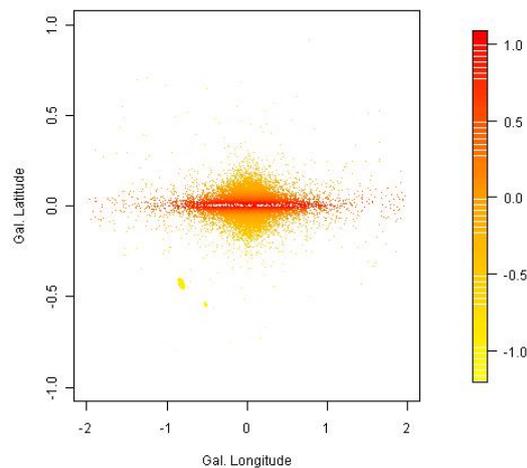


Figura 19: Extinciones de las estrellas Cefeidas simuladas. A mayor extinción, el color de las estrellas representadas es de un rojo más intenso, según la escala mostrada a la derecha.

Los valores de extinción son coherentes: son muy elevados en aquellas estrellas cuyas coordenadas corresponden a una línea de visión que pasa por el plano galáctico, donde están ubicados los planos de los discos de la galaxia, y por tanto donde mayor es la concentración de gas y polvo estelar.

El punto de extinción máxima se corresponde con el centro de la galaxia. En cuanto a los valores de extinción A_v , se muestra a continuación un resumen de sus estadísticos más importantes:

```
> summary(Av)
   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
0.06235 0.08456 0.11010 0.88870 0.49610 12.27000
```

Que trabajando con logaritmos en base 10 es (la gráfica representa $\log_{10}Av$):

```
> summary(log10(Av))
   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
-1.2050 -1.0730 -0.9582 -0.6584 -0.3044  1.0890
```

Operando con ‘R’, y teniendo en cuenta los artículos señalados en el punto 4.4, se puede comprobar que el “survey” sintético obtenido tiene todas las estrellas con su valor Av dentro de valores razonables, por lo que el resultado es coherente con las observaciones y artículos de referencia utilizados.

4.7.6. Diagrama HR

Seguidamente, se estudian los resultados obtenidos para el diagrama HR de las cefeidas, en base al artículo [Tammann-03], que contiene gráficas (ver figura 20b) con las correlaciones de (V-I) en LMC, SMC y Galaxia.

La magnitud (V-I) o VI en el espacio de parámetros del “survey” sintético obtenido, varía entre 0.4 y 1.2 (ver figura 20a), si bien la mayor de las observaciones reales en LMC, SCM y Galaxia se concentran en un intervalo más estrecho de (V-I): entre 0.4 y 0.8. Operando con ‘R’ en las cefeidas sintéticas obtenidas, obtenemos el siguiente resumen:

```
> summary(VI)
   Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
0.06943 0.51480 0.60900 0.61460 0.70970 1.19100
```

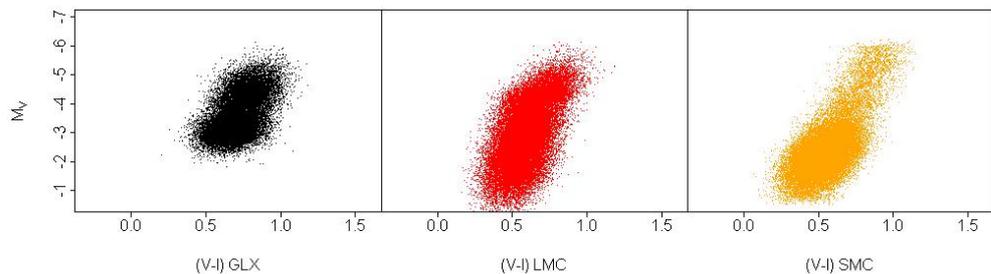


Figura 20a: Diagrama HR correspondiente a la Galaxia (disco fino + disco grueso + bulbo) en color negro; LMC en color rojo y SMC en color naranja.

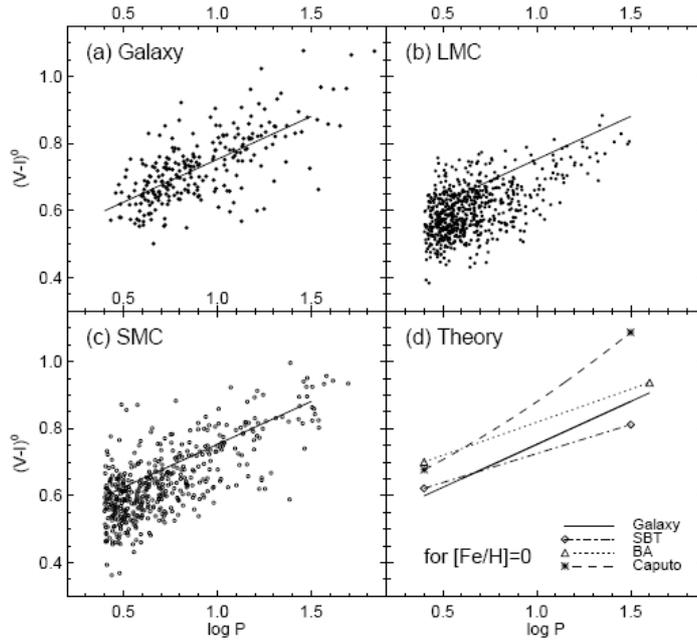


Figura 20b: Distribución teórica de de $(V-I)$ según [Tammann-03]

Donde de además operando, se obtiene que 83.535 cefeidas sintéticas presentan un valor de $0.4 \leq (V - I) \leq 0.8$; por otro lado, aplicando el criterio del apéndice C, entre los valores extremos solo hay 562 valores atípicos de los 100.000 generados, por lo que la correlación utilizada para calcular el “survey” puede considerarse correcta.

4.7.7. Diagrama logP-MV

Este diagrama nos muestra la relación entre el logaritmo del periodo y la magnitud absoluta de la estrella (a 10 pc.). De [Sandage-08], sabemos que de las observaciones realizadas tanto en la LMC, SMC como la Galaxia, cuando el valor del logaritmo del periodo varía entre 0.35 y 1.8, el valor de MV varía entre -2.0 y -7.0 (ver figura 21b).

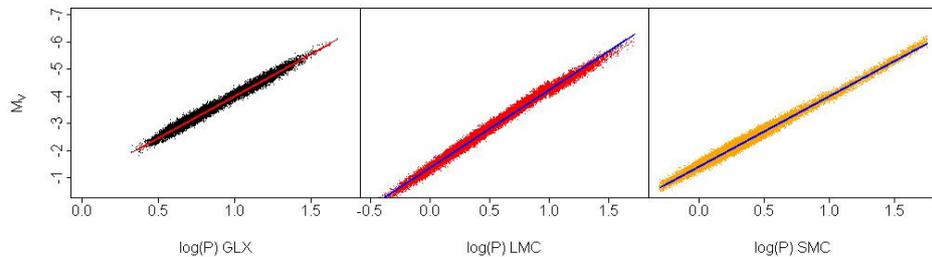


Figura 21a: Diagrama $\log P$ - M_V correspondiente a las Cefeidas simuladas: en negro para la Galaxia (disco fino + disco grueso + bulbo), rojo para la LMC y naranja para la SMC.

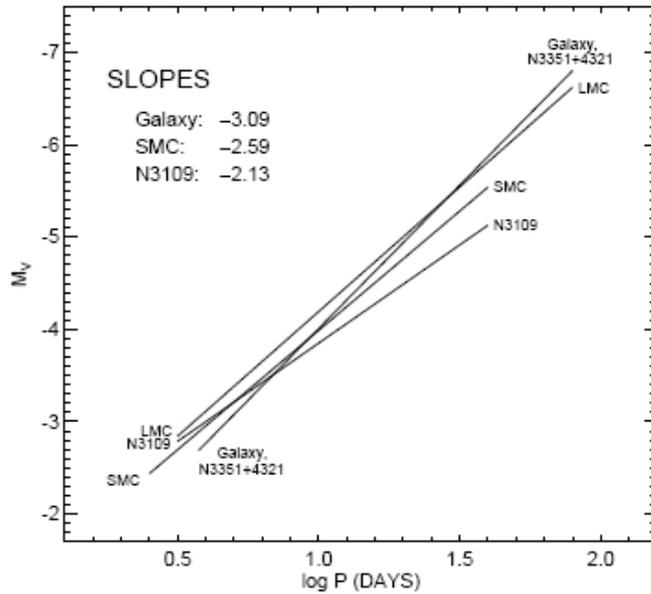


Figura 21b: Distribución teórica de M_v en función de $\log P$ según [Sandage-08] para Galaxia, LMC y SMC. En el gráfico aparecen las pendientes correspondientes a cada objeto (LMC, SMC y galaxia).

Si se comparan estos valores reales de M_v con los obtenidos en la simulación para el parámetro MV , que se muestran gráficamente en la figura 21a para la Galaxia, LMC y SMC respectivamente:

```
> summary(MV)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-6.2210 -3.6090 -2.7980 -2.8970 -2.1100  0.2111
```

Se observa que los valores obtenidos a partir del “survey” sintético son coherentes con las observaciones astronómicas, por lo que pueden validarse los resultados obtenidos.

4.7.8. Diagrama logP-VI

Este diagrama nos muestra la correlación entre el logaritmo del periodo y el color VI; los resultados obtenidos en nuestra simulación se representan en la figura 22.

Los resultados de la correlación obtenida están de acuerdo con los valores observados en [Sandage-08] que se muestran en la figura 21b para $\log P$; aproximadamente en valores de $\log P = 1.0$ aparece una discontinuidad, que también se muestra en la grafica de la figura 22.

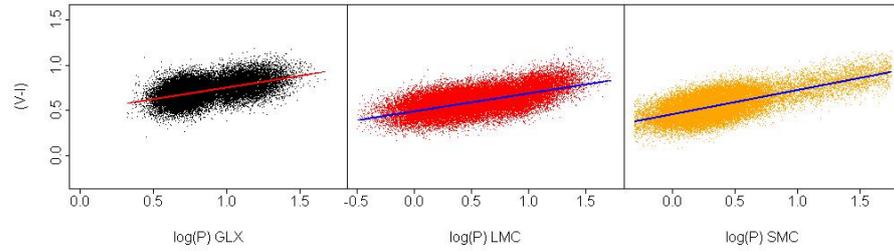


Figura 22: Diagrama logP-VI para la Galaxia (disco Fino + Disco Grueso + Bulbo Galáctico) en color negro, para la LMC en color rojo, y para la SMC en color naranja.

4.7.9. Diagramas PLC

Las siguientes figuras (figura 23a, 23b y 23c) muestran las gráficas tridimensionales (la tercera dimensión es el color) PLC de la Vía Láctea, LMC y SMC donde esta última se representa por una escala de colores amarillo-naranja-rojo. Esta gráfica combina los valores de los tres parámetros que representa, y que ya se han descrito en figuras y diagramas anteriores.

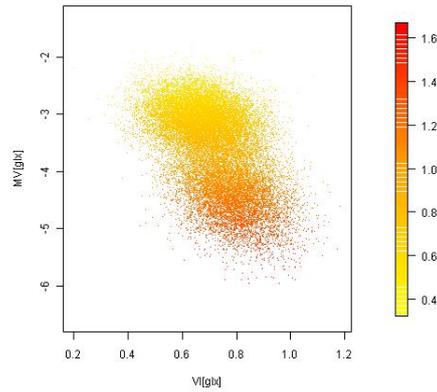


Figura 23a: Diagrama PLC de la Vía Láctea. La escala de color representa el período, en escala logarítmica.

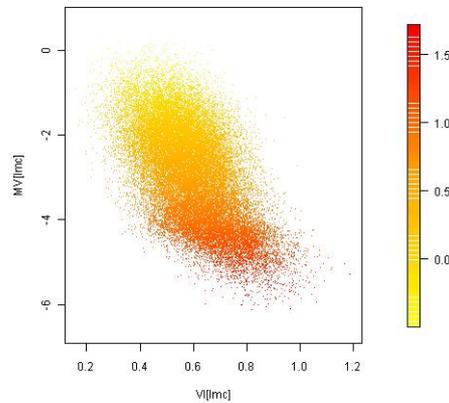


Figura 23b: Diagrama PLC de la LMC. La escala de color representa el período, en escala logarítmica.

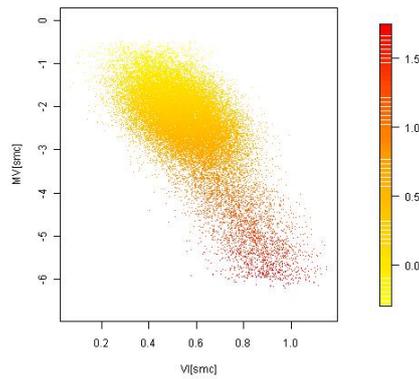


Figura 23c: Diagrama PLC de la SMC. La escala de color representa el período, en escala logarítmica.

4.7.10. Diagrama de Amplitudes

En este punto, se muestra el diagrama de amplitudes de las cefeidas sintetizadas; la gráfica Amplitudes-logaritmo del periodo se muestra en la figura 24. En él puede observarse que la variación máxima de la amplitud está en el intervalo [0.2, 2.0] para LMC, SMC y Galaxia, lo que está de acuerdo con [Sandage-04].

Se muestra también en la figura 24a la curva $\log P-A_B$ que proporciona [Sandage-04], si bien en este caso la amplitud se da en la banda B y no en la V, aunque esta diferencia no cambia de manera significativa la gráfica de amplitudes.

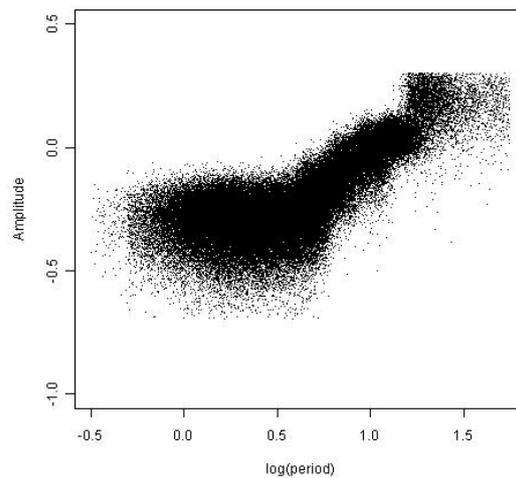


Figura 24: Diagrama de Amplitudes de las Cefeidas sintéticas obtenidas. Se representan las amplitudes en función del periodo de LMC, SMC y de la Galaxia.

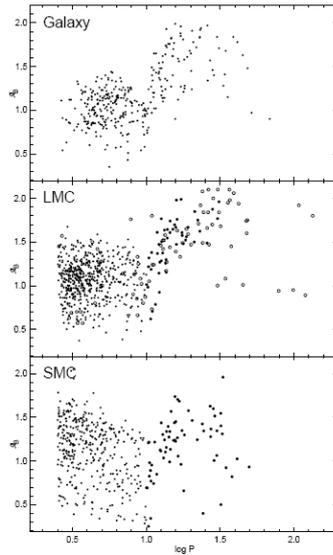


Figura 24a: *Diagrama de Amplitudes para LMC, SMC y Galaxia según [Sandage-04]. Se observa una discontinuidad en la amplitud en las inmediaciones de $\log P = 1.0$*

4.7.11. Diagrama distancia-error de paralaje

La siguiente gráfica (figura 25) representa el error de paralaje en función de la distancia en kpc.

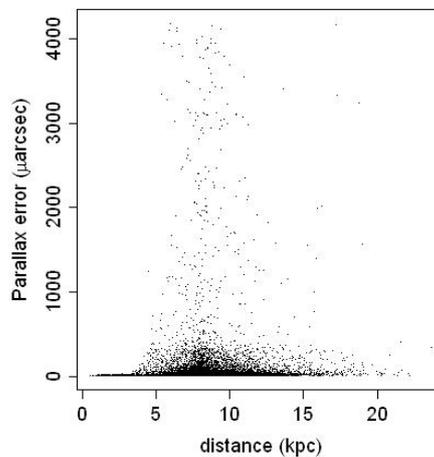


Figura 25: *Diagrama de errores de paralaje de las Cefeidas sintéticas. El mayor error de paralaje aparece en las inmediaciones de $d = 8\text{kpc}$, distancia aproximada del Sol al centro de la Galaxia.*

Este diagrama representa el error de paralaje de las estrellas sintéticas pertenecientes a la Galaxia, y cuyo valor de $G < 21$ (campo visible); se usa la distancia de dicha estrella y el valor obtenida para ella en la simulación de sigma en la banda visible. Aparece un máximo de los errores de paralaje en torno al valor de distancia de 8 kpc. Esta distancia

coincide con la distancia de la tierra al centro de la galaxia, lo que nos indica que los mayores errores de paralaje se producen en al medir los parámetros de las estrellas localizadas en el bulbo galáctico

4.7.12. Diagramas de Lutz-Kelker

En este punto se presentan los diagramas de Lutz-Kelker (figuras 26 y 27) correspondiente a la simulación de cefeidas.

El error de Lutz-Kelker es un error estadístico de medida de las distancias astronómicas, por el cual el paralaje medido de las estrellas es mayor que el realmente existente. Esto provoca que no solo las distancias, sino también las luminosidades medidas, sean de mayor magnitud que las reales: un error en el paralaje de una estrella del 17,5% puede provocar un error en su luminosidad en la banda I (infrarroja) del 30%.

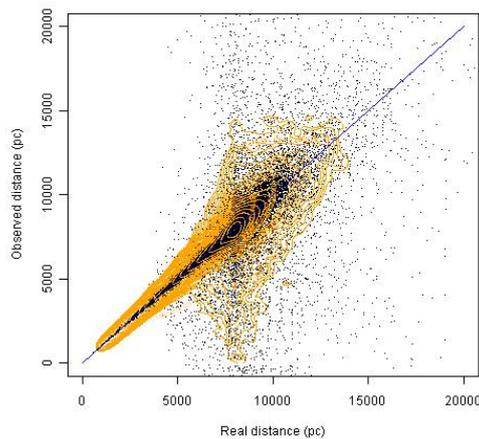


Figura 26: *Diagrama de Lutz-Kelker (I)*. Puede observarse que el mayor error cometido se produce aproximadamente a 8 kpc, distancia del Sol al centro de la Galaxia.

La existencia de este error y la necesidad que ha surgido de corregirlo después de las medidas de alta precisión realizadas por la misión Hipparcos, ha provocado que numerosos investigadores estén trabajando en los mecanismos de corrección del mismo; una explicación detallada de este error así como un método sencillo para corregirlo, puede verse en [Oudmaijer-98].

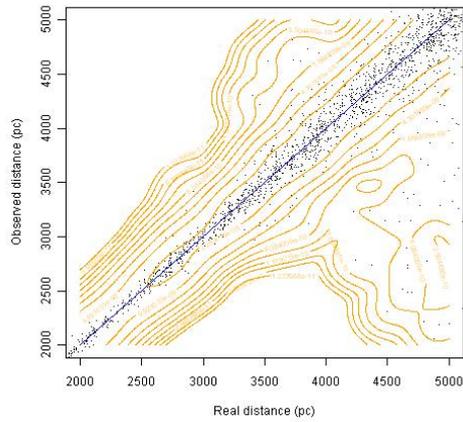


Figura 27: *Diagrama de Lutz-Kelker (II)*. Las curvas representan la estimación de la distancia en dos dimensiones.

Los diagramas anteriores muestran la distancia real (la generada directamente en la simulación) contra la distancia observada, estimada como el valor de una distribución normal con media la distancia real, y varianza el error de paralaje en dos escalas diferentes de distancias (una por cada figura); a partir de ahí, los puntos representan cada una de las estrellas localizadas en función de su posición real y observada, y las curvas representan la estimación de la pdf bidimensional de la distancia.

4.7.13. Histograma de distancias

La figura 28 muestra el histograma de distancias de la simulación de las cefeidas realizada. Su análisis es similar al realizado en el punto 4.7.1

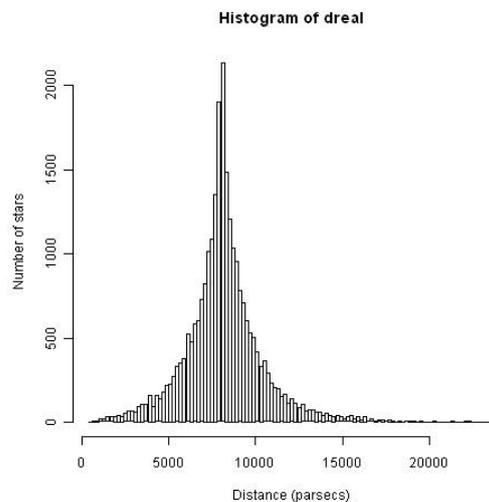


Figura 28: *Histograma de distancias reales*.

4.8. “Survey” de RR Lyrae

Las RR Lyrae se hallan situadas en el diagrama H-R en la rama horizontal, lo que significa que aún poseyendo distintas temperaturas tienen luminosidades similares, formando una línea transversal a lo largo de la banda de inestabilidad. Pulsan en el modo fundamental (RRab), en el primer armónico (RRc) o en ambos (RRb); en este trabajo, solo se considerarán las RR Lyrae del tipo RRab. Las RR Lyrae del tipo RRab tienen curvas de luz asimétricas, con una rama ascendente rápida de períodos de 0.3 a 1.2 días y con amplitudes de 0.5 a 2.0 magnitudes.

En el caso de las RR Lyrae, una vez obtenidas las coordenadas de nuestras estrellas sintéticas (de la misma manera que se hizo con las Cefeidas en el punto 4.6.), el paso siguiente es el cálculo del resto de parámetros característicos de las mismas, empezando por **P** (período en días). Para ello, y en base a [Soszynski-10] para la LMC y SMC y [Collinge-06] para el bulbo galáctico, se aproximan las distribuciones de los períodos con la función *stars.generator()*, en el fichero “*functions-RR_Lyrae.R*”. Como puede verse, las distribuciones del periodo se simulan mediante dos normales truncadas, como en el caso de las Cefeidas.

Existe también la posibilidad, en función de los parámetros de configuración del programa ya descritos, de generar el valor de este parámetro de manera aleatoria, según una distribución uniforme con los mismos límites que los establecidos para las normales truncadas antes mencionadas, tal y como sucedía en el caso de las Cefeidas.

Los valores obtenidos para el parámetro **P** son también en este caso lo suficientemente buenos para el propósito del “survey” que se está generando, y ‘R’ permite generar los mismos con gran simplicidad y rapidez.

Una vez obtenidos los valores de **P**, se obtienen los valores de **MV** mediante la función *m.v()*, también en el fichero “*functions-RR_Lyrae.R*”, y cuyos parámetros numéricos se basan en las tablas 3 y 9 de [Catelan-04]. El cálculo de **MV** depende de la metalicidad *Z*. En astrofísica, la metalicidad describe la abundancia relativa de elementos más pesados que el He en una estrella, independientemente de la posición en la tabla periódica de dichos elementos. La metalicidad de una estrella está relacionada con el momento en el que se formó, lo que permite deducir su edad o zona de la Galaxia donde se formó.

Posteriormente se aborda el cálculo de **VI** mediante la función *vi.pl()* cuya parametrización se ha realizado en base a los datos aportados por [Borissova-09] para la LMC, [Soszynski-10] para la SMC y [Collinge-06] para el bulbo galáctico. Como en el caso de las cefeidas, en función de la parametrización introducida a través del fichero de configuración “*config.dat*”, puede forzarse a un cálculo aleatorio de **VI** según una distribución uniforme. Finalmente, se calcula la amplitud **Amp** mediante la función *a.v()*, cuyos parámetros se han extraído de [Borissova-09] para la LMC, de [Deb-09] para la

SMC y de [Alcock-97] para la Galaxia. El cálculo de **Amp** también depende de la metalicidad Z .

Para más detalles, se recomienda estudiar el código directamente tanto del fichero “*functions-RR_Lyrae.R*” como del fichero “*LMC-SMC-GLX-RR_Lyrae.R*”, cuyo contenido ya se ha descrito en puntos anteriores.

Con los cálculos realizados hasta el momento, se dispone para las RR Lyrae de los parámetros **t**, **alpha**, **delta**, **P**, **logP**, **MV**, **VI** y **Amp** tanto para LMC como para SMC y Galaxia (disco fino, grueso y bulbo).

Como en el caso de las Cefeidas, todos los parámetros calculados, así como uno adicional indicando el objeto en el que se encuentra la estrella sintética simulada se van acumulando en una matriz paso a paso, como puede verse en el código contenido en el fichero “*LMC-SMC-GLX-RR_Lyrae.R*”. El cálculo del valor de los parámetros descritos hasta este momento es totalmente análogo para LMC, SMC y discos y bulbo galáctico, y para cada uno de estos objetos, se forma la correspondiente matriz con sus valores; adicionalmente, se añade un parámetro más **type** (se ubica en la primera posición), que indica el tipo de objeto al que pertenece la estrella sintética:

- **type** = 1 → LMC.
- **type** = 2 → SMC.
- **type** = 3 → Disco fino.
- **type** = 4 → Disco grueso.
- **type** = 5 → Bulbo Galáctico (“glx.I”, $Z=0.006$).
- **type** = 6 → Bulbo Galáctico (“glx.II”, $Z=0.005$).
- **type** = 7 → Bulbo Galáctico (“glx.III”, $Z=0.02$).
- **type** = 8 → Bulbo Galáctico (“glx.IV”, $Z=0.01$).

Obsérvese que se diferencian 4 tipos diferentes de variables RR Lyrae en el bulbo galáctico en función del valor de la metalicidad de las mismas, lo que constituye una diferencia notable respecto al caso de las Cefeidas. Las cadenas de caracteres “glx.I”, “glx.II”, “glx.III” y “glx.IV” diferencian a lo largo del código las estrellas variables del bulbo con las metalicidades $Z=0.006$, $Z=0.005$, $Z=0.02$ y $Z=0.01$ respectivamente. Como ya se ha indicado, las funciones $m.v()$ y $a.v()$ tienen en cuenta las diferentes metalicidades del bulbo.

Una vez con las anteriores matrices de cada objeto conformadas, se combinan las de los discos y el bulbo para formar una nueva matriz correspondiente a la galaxia, en la que la contribución de cada objeto individual indicada en [Schneider-06], como se hizo en el caso de las Cefeidas en el punto 4.6.

Como en el caso de las Cefeidas, se dispone ya para las RR Lyrae de 3 matrices con los parámetros hasta ahora indicados: ‘P.lmc’, ‘P.smc’ y ‘P.glx’ para LMC, SMC y Galaxia, respectivamente. La matriz ‘P.glx’ ya contiene la mezcla de los distintos elementos galácticos combinados según la proporción antes indicada.

Finalmente, se obtiene la matriz ‘P.survey’, que es la resultante de colocar de manera consecutiva por columnas las tres matrices parciales anteriores. Tanto en aquellas matrices como en la resultante ‘P.survey’, los valores de los parámetros están colocados por filas, por una simple cuestión de comodidad.

Posteriormente se calculan las coordenadas galácticas de las estrellas sintéticas obtenidas, de manera totalmente análoga a como se describió para el caso de las Cefeidas, y el valor del parámetro de extinción A_v y la magnitud absoluta extinguida como en las cefeidas, excepto en que se usan los ficheros intermedios “*av-RR_Lyrae.dat*” y “*gc-RR_Lyrae.dat*”. Con ello puede pasarse al valor de los parámetros astrométricos de GAIA según se describió en 4.3.

Como en el caso de las Cefeidas, la matriz ‘P.survey’, debidamente procesada puede presentarse en pantalla, si bien como ya se ha indicado, para el número de estrellas calculadas, 10^5 en este caso, es poco práctico; el código solo permite presentar los resultados por pantalla para un “survey” de no más de 5.000 estrellas; en cualquier caso se almacena en el fichero de texto “*results-RR_Lyrae.dat*”, y en un objeto ‘R’ en el fichero “*RR_Lyrae.Rdata*”, fácilmente manipulable con ‘R’, tal y como se indicó en el punto 4.2.

4.9. Validación del “survey” de RR Lyrae

En este apartado se describirá el proceso de validación de los resultados dados por el código ‘R’ contenido en los ficheros “*LMC-SMC-GLX-RR_Lyrae.R*” y “*functions-RR_Lyrae.R*”; esta validación se lleva a cabo con el código contenido en los ficheros “*plotset1-RR_Lyrae.R*” y “*plotset2-RR_Lyrae.R*”; el código de ambos ficheros se basa en las grandes capacidades gráficas de ‘R’ para realizar dicha validación, mediante la realización de gráficas que puedan ser comparadas con los resultados obtenidos bien en los artículos referenciados a lo largo de esta exposición, o bien en otros suficientemente reconocidos.

Básicamente, el código de los ficheros ‘R’ “*plotset1-RR_Lyrae.R*” y “*plotset2-RR_Lyrae.R*”, leen los datos resultantes de la ejecución del código de los ficheros “*LMC-SMC-GLX-RR_Lyrae.R*” y “*functions-RR_Lyrae.R*” que están contenidos en el fichero de texto “*results-RR_Lyrae.dat*”, y generan las gráficas que muestran las correlaciones existentes entre los distintos parámetros sintetizados para cada una de las estrellas del ‘survey’ obtenido. Este procedimiento es el mismo que ya se ha descrito para las Cefeidas.

Los datos del fichero “*results-RR_Lyrae.dat*” se almacenan durante la ejecución del código de validación en la variable ‘*data*’, y a partir de ahí se trazan las gráficas de las correlaciones existentes en el espacio de parámetros. Como en el caso de las Cefeidas, se han obtenido 100.000 estrellas sintéticas en este “survey” (si bien el número de estrellas es un parámetro variable del programa).

4.9.1. Histograma de periodos

Inicialmente se confecciona un histograma de los valores obtenidos para el parámetro **P** para todas las RR Lyrae en modo fundamental (RRab), de manera que pueda observarse la distribución de dicho parámetro. Para el caso de las RRab, dicho histograma se muestra en la figura 29.

Puede observarse que los periodos se distribuyen mayoritariamente entre los valores $P = 0.3$ días y $P = 0.8$ días, según se muestra en la mencionada figura 19, que es la variación del periodo generalmente aceptada para este tipo de estrellas variables según [Soszynski-10] [Collinge-06] (99.812 de un total de 100.000 estrellas que contiene el “survey” de RRab obtenido).

A diferencia de las Cefeidas, el periodo en el caso de las RR Lyrae se mide en días, y no en el logaritmo de días, dado el corto valor de periodo que presenta esta clase de variables.

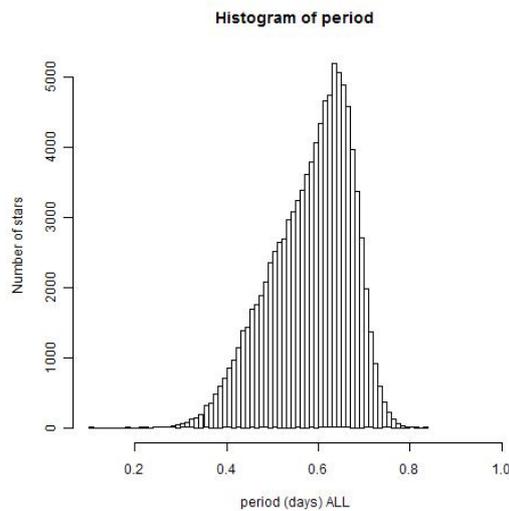


Figura 29: Histograma del periodo P (en días) de las variables RR Lyrae simuladas (LMC + SMC + Disco fino + Disco Grueso + Bulbo galáctico). Se han simulado un total de 10^5 estrellas sintéticas RR Lyrae (RRab).

Podemos valernos también de ‘R’ para obtener un resumen de la columna correspondiente a los periodos de nuestras estrellas sintéticas:

```
> summary(P)
Min.   1st Qu.  Median   Mean   3rd Qu.  Max.
0.1068 0.5238  0.5981  0.5829  0.6505  0.8336
```

Que son unos valores razonables para el valor del periodo, si bien puede observarse que se han generado también algunos valores extremos anormales: periodos menores de 0.34425 días o mayores de 0.8345 días respectivamente, según el criterio indicado en el apéndice C para calcular valores atípicos: según dicho algoritmo, solo 668 estrellas sintéticas están fuera de dichos límites, ya que se han utilizado funciones truncadas para su generación, y evitar dichos valores atípicos en la medida de lo posible en las simulaciones realizadas en el código.

Se han obtenido las graficas individuales correspondientes a los histogramas del parámetro **P** de cada uno de los cuerpos celestes simulados: LMC (figura 29a) y SMC (figura 29b), y Galaxia (figura 29c) que se muestran a continuación; dichos gráficos, comparados con los gráficos de [Soszynski-10] y [Collinge-06] de las figuras 29d y 29e, puede observarse que son aproximados a los representados en dichos artículos.

Obsérvese que a diferencia de lo que ocurría en las Cefeidas, la escala de periodos en todas las figuras anteriores están medidas en días, y que los resultados están también de acuerdo con la figura 5 del punto 3.3.

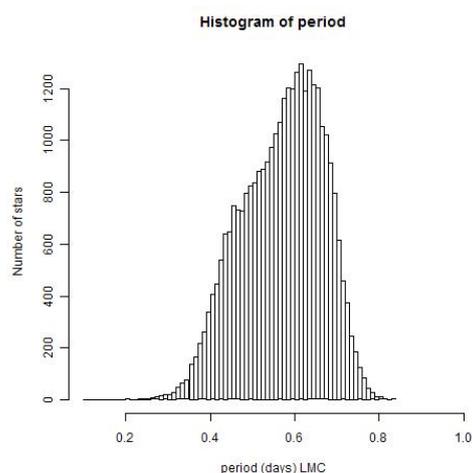


Figura 29a: *Histograma del periodo P de las estrellas sintéticas correspondientes a la LMC.*

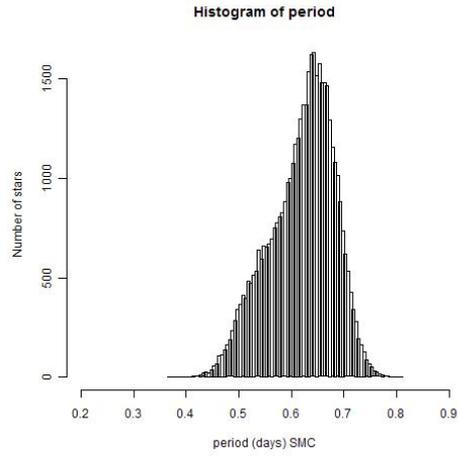


Figura 29b: *Histograma del periodo P de las estrellas sintéticas correspondientes a la SMC.*

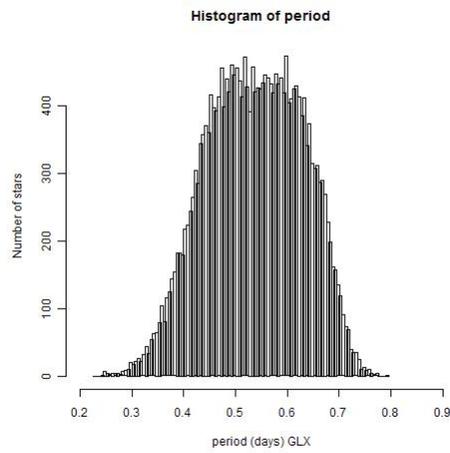


Figura 29c: *Histograma del periodo P de las estrellas sintéticas correspondientes a la Galaxia.*

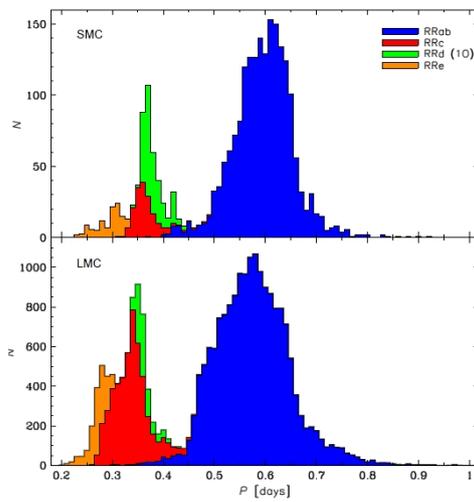


Figura 29d: *Histograma del periodo P (días) en LMC, SMC [Soszynski-10].*

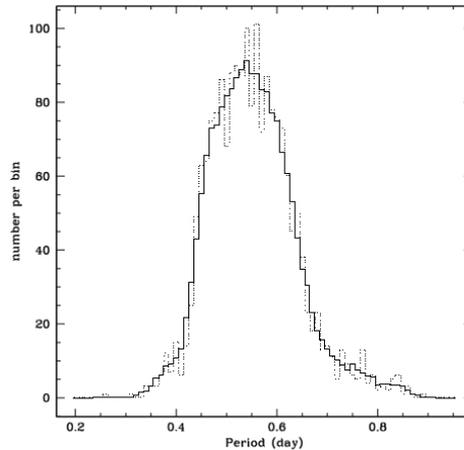


Figura 29e: *Histograma del periodo P (días) en la Galaxia [Collinge-06]*

4.9.2. Representación de las coordenadas (Coordenadas Ecuatoriales)

El proceso de cálculo de las coordenadas ecuatoriales, y los resultados son análogos a los obtenidos para las Cefeidas: ver punto 4.7.2. Dado que en ambos casos responden a la misma distribución, la gráfica obtenida para las RR Lyrae es muy similar a la obtenida para las Cefeidas, por lo que no se representa aquí otra vez.

4.9.3. Representación de las coordenadas (Coordenadas Galácticas)

Como en el caso anterior, el proceso de cálculo de las coordenadas galácticas, y los resultados son análogos a los de las Cefeidas: ver punto 4.7.3. Dado que en ambos casos responden a la misma distribución, la gráfica obtenida para las RR Lyrae es muy similar a la obtenida para las Cefeidas, por lo que no se representa aquí otra vez.

4.9.4. Visualización de objetos en Banda G

Procediendo de la misma manera que se hizo para las Cefeidas, se obtienen unos resultados para las RR Lyrae análogos a los que obtuvieron para las Cefeidas en el punto 4.7.4, por lo que no se incluye aquí la gráfica, al no aportar información adicional a la obtenida para las Cefeidas sintéticas. Analizando la banda G con más detalle en el “survey” obtenido, se obtienen los siguientes datos:

```
> summary(g)
Min.   1st Qu. Median   Mean   3rd Qu.  Max.
8.023  19.120  19.400  18.910  19.580  21.000
```

4.9.5. Visualización de los valores de Extinción (A_v)

También en este caso tanto la gráfica obtenida de los valores de extinción para las estrellas sintéticas RR Lyrae así como las conclusiones que se pueden extraer de la misma son en todo similar a las obtenidas para las Cefeidas sintéticas del punto 4.7.5, por lo que no se incluye la citada gráfica al no añadir información adicional.

Los valores de extinción son de nuevo coherentes: son muy elevados en aquellas estrellas cuyas coordenadas corresponden a una línea de visión que pasa por el plano galáctico, donde están ubicados los planos de los discos de la galaxia, y por tanto donde mayor es la concentración estelar. En cuanto a los valores de extinción A_v , se muestra a continuación un resumen de sus estadísticos más importantes:

```
> summary(Av)
  Min. 1st Qu.  Median    Mean   3rd Qu.    Max.
0.06404 0.08438 0.10870 0.59330 0.11740 7.48600
```

Que trabajando con logaritmos (la gráfica representa $\log_{10}A_v$) da:

```
> summary(log10(Av))
  Min. 1st Qu.  Median    Mean   3rd Qu.    Max.
-1.1940 -1.0740 -0.9637 -0.7203 -0.9304  0.8742
```

Operando con ‘R’, y teniendo en cuenta los artículos y estudios referenciados en el punto 4.4, se puede comprobar que el “survey” sintético obtenido para las RR Lyrae tiene todas las estrellas con su valor A_v dentro de límites razonables, por lo que el resultado es coherente con las observaciones y artículos de referencia utilizados.

4.9.6. Diagrama HR

Seguidamente, se estudian los resultados obtenidos para el diagrama HR de las RR Lyrae sintéticas, en base a las tablas del artículo [Catelan-04]. Operando con ‘R’ en las RRab sintéticas obtenidas, obtenemos el siguiente resumen:

```
> summary(VI)
  Min. 1st Qu.  Median    Mean   3rd Qu.    Max.
0.1481  0.4994  0.5462  0.5569  0.6131  0.9893
```

La figura 30, muestra los resultados obtenidos para la Galaxia, LMC y SMC respectivamente.

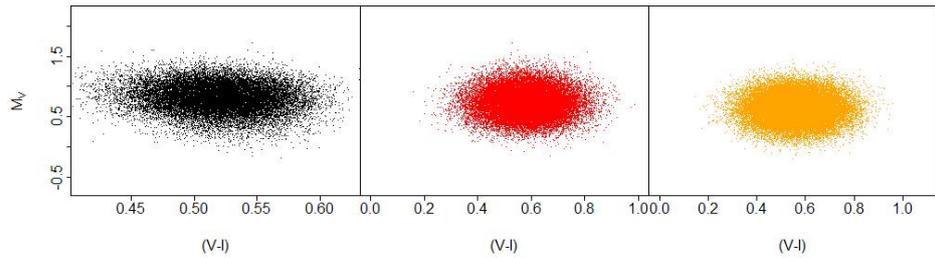


Figura 30: Diagrama HR para las RR Lyrae sintéticas: negro galaxia (Disco fino + Disco grueso + Bulbo Galáctico); rojo LMC y naranja SMC.

4.9.7. Diagrama logP-MV

Este diagrama nos muestra la relación entre el logaritmo del periodo y la magnitud absoluta de la estrella (a 10 pc.), y se muestra en la figura 31; el resumen de los valores MV obtenidos es el siguiente:

```
> summary(MV)
  Min.   1st Qu.  Median    Mean   3rd Qu.    Max.
-0.2009  0.5459  0.6982  0.6999  0.8524  1.7240
```

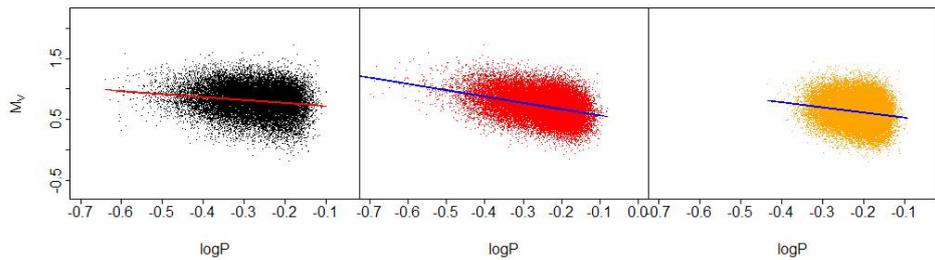


Figura 31: Diagrama logP-MV para las estrellas sintéticas RR Lyrae: el color negro se corresponde a las estrellas de la galaxia (Disco fino + Disco grueso + Bulbo galáctico), el rojo a las de la LMC y el naranja a las estrellas de la SMC.

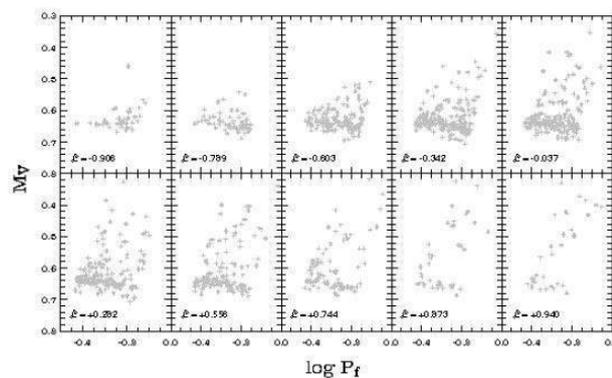


Figura 31a: Distribución de M_v según para las variables RR Lyrae [Catelan-04]

Comparados con la figura 31a obtenida de [Catelan-04], vemos que se obtienen valores algo más dispersos que los indicados en dicho artículo; no obstante, calculando los valores atípicos según el apéndice C, los

límites para los mismos son $[-0.37, 1.7719]$; no habiendo ningún valor de M_V para las 10^5 estrellas sintéticas obtenidas fuera de estos límites, se pueden aceptar como válidos los mismos.

4.9.8. Diagrama logP-VI

Este diagrama nos muestra la correlación entre el logaritmo del periodo y el color VI; los resultados obtenidos en nuestra simulación se representan en la figura 32. Los resultados de la correlación obtenida están de acuerdo con los valores observados en [Collinge-06] para el caso de la galaxia que se muestran en la figura 32a.

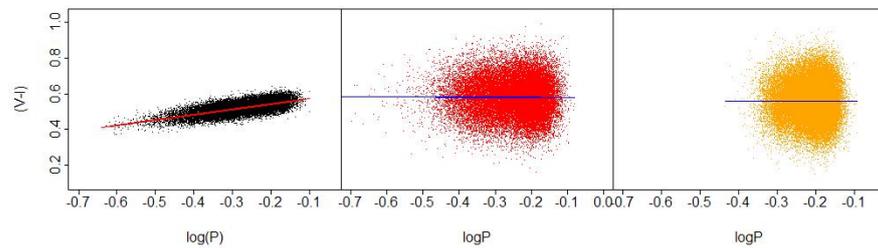


Figura 32: Diagrama logP-VI para las estrellas RR Lyrae sintéticas: negra para la galaxia (Disco fino + Disco grueso + bulbo galáctico); rojo para LMC y naranja para SMC.

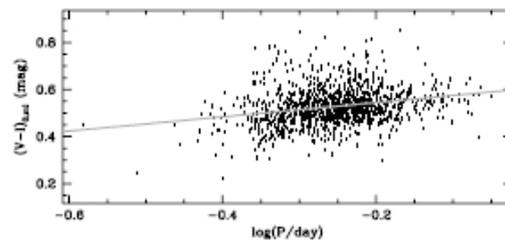


Figura 32a: Diagrama logP-VI para las estrellas variables RR Lyrae de la Galaxia, obtenido de [Collinge-06].

4.9.9. Diagramas PLC

Las siguientes figuras (figura 33a, 33b y 33c) muestran las gráficas tridimensionales (la tercera dimensión es el color) PLC de la Vía Láctea, LMC y SMC donde esta última se representa por una escala de colores amarillo-naranja-rojo.

Esta gráfica combina los valores de los tres parámetros que representa, y que ya se han descrito en figuras y diagramas anteriores.

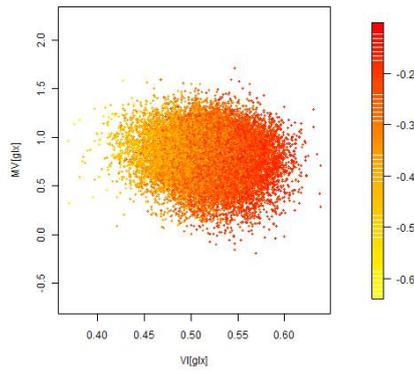


Figura 33a: Diagrama de la Vía Láctea PLC de la Galaxia: en escala de colores se representa $\log P$.

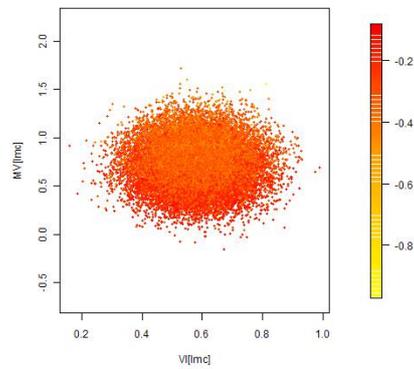


Figura 33b: Diagrama de la Vía Láctea PLC de la LMC: en escala de colores se representa $\log P$.

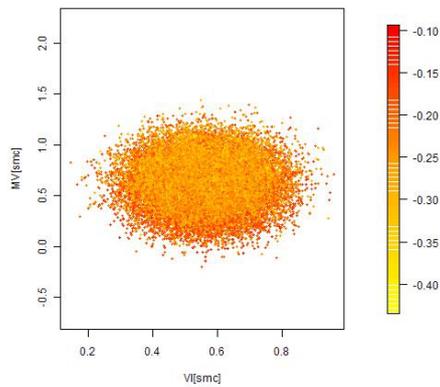


Figura 33c: Diagrama de la Vía Láctea PLC de la SMC: en escala de colores se representa $\log P$.

4.9.10. Diagrama de Amplitudes

En este punto, se muestra el diagrama de amplitudes de las estrellas RR Lyrae sintetizadas; la gráfica Amplitudes-logaritmo del periodo se muestra en la figura 34. En él puede observarse que la variación máxima de la amplitud está en el intervalo $[-2.0, 1.0]$ para LMC, SMC y Galaxia, lo que está de acuerdo con [Borissova-09] (figura 34a), [Deb-09] (figuras

34b y 34c) y [Alcock-97] (figura 34d). De [Deb-09] se han combinado las dos figuras indicadas para obtener la relación **logP-Amp** deseada.

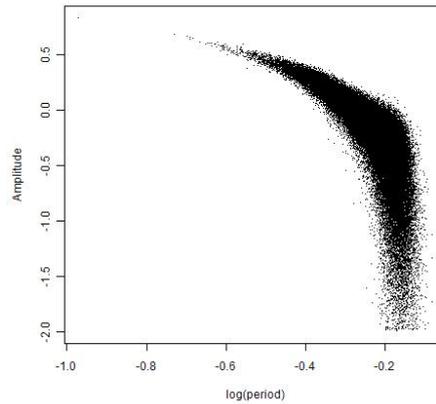


Figura 34: Diagrama de Amplitudes de las estrellas RR Lyrae sintéticas; están representadas conjuntamente las de la Vía Láctea, LMC y SMC.

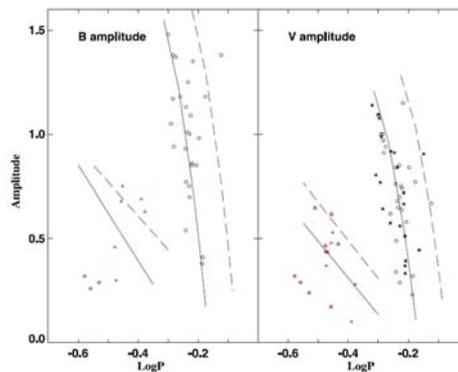


Figura 34a: Diagrama de Amplitudes para la LMC obtenidas de [Borissova-09]. Solo las estrellas representadas en color negro en la banda V son RRab, por lo que en este trabajo solo se considera la gráfica correspondiente a estas.

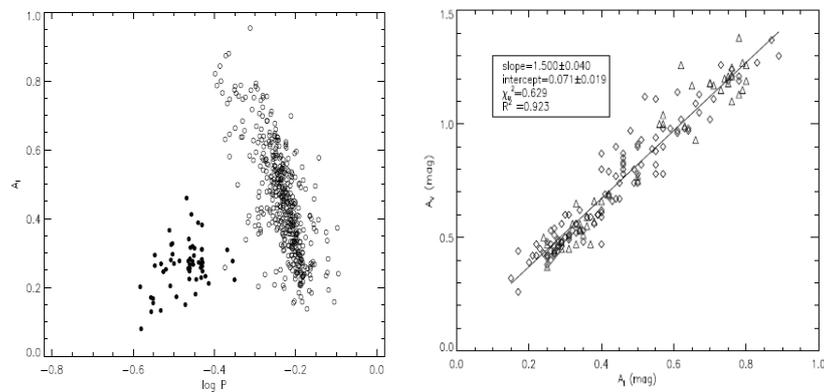


Figura 34b: Diagrama de Amplitudes $\log P-A_I$ (izquierda) y A_I-A_V (derecha) obtenidos de [Deb-09] de varios cúmulos estelares usados como referencia para la SMC. En el diagrama de la izquierda, los círculos no rellenos pertenecen a las RR Lyrae RRab, que son las únicas que se consideran en este trabajo.

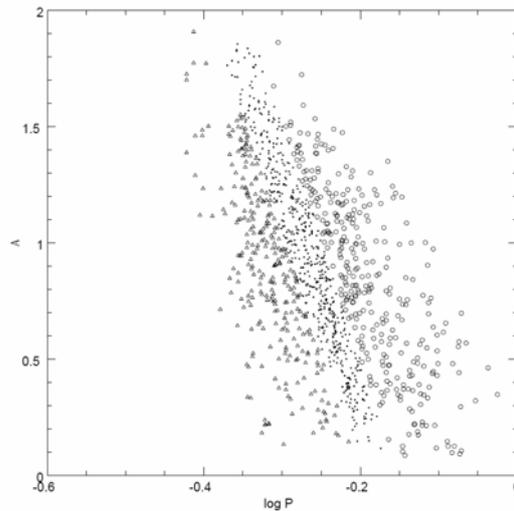


Figura 34c: Diagrama de Amplitudes de las RR Lyrae del núcleo de la Vía Láctea obtenido de [Alcock-97]. Los distintos tipos de puntos representan la metalicidad: círculos metalicidad baja, puntos sólidos media y triángulos alta metalicidad.

Aunque la figura 34c corresponde a RR Lyrae del bulbo galáctico, se ha asumido en este trabajo que dicha curva es válida también para el resto de la galaxia (disco fino y disco grueso).

4.9.11. Diagrama distancia-error de paralaje

También para las RR Lyrae se ha hecho la grafica distancia – error de paralaje; su aspecto e interpretación es en todo análoga a la misma gráfica obtenida para las Cefeidas (ver punto 4.7.11), por lo que no se incluye aquí al no aportar información adicional.

4.9.12. Diagramas de Lutz-Kelker

En el caso del los diagramas de Lutz-Kelker encontramos que su interpretación y aspecto es análogo al obtenido en el caso de las Cefeidas (ver punto 4.7.12), por lo que no se incluye aquí al no aportar nueva información.

4.9.13. Histograma de distancias

La grafica obtenida es de nuevo análoga a la obtenida en el caso de las Cefeidas (ver punto 4.7.13 para más información).

4.10. “Survey” de LPV (“Long Period Variable”)

Las estrellas de largo periodo, es un tipo de estrella variable que como ya se ha indicado en el punto 3.3 presenta variaciones en su brillo que en general abarcan periodos temporales de meses o incluso años (aunque también las hay cuyo período es de días). Son estrellas gigantes muy brillantes, desde la clase espectral F. La mayoría son gigantes rojas o gigantes AGB (“Asymptotic Giant Branch”) de masa pequeña o media, de clase espectral M, S ó C, y habitualmente de color naranja o rojo. Una característica específica de estas estrellas recientemente descubierta, es que se distribuyen en series que pueden identificarse fácilmente sobre un diagrama $K_s - \log P$, esta peculiaridad es debida a la compleja naturaleza de sus curvas de luz. Durante mucho tiempo la existencia de series en las LPV pasó desapercibida, ya que es necesario realizar observaciones durante largos periodos de tiempo (años) para caracterizar completamente las curvas de luz de estas estrellas. No se han encontrado artículos recientes que clasifiquen las estrellas de la Vía Láctea (discos, bulbo) en series, por lo que se ha asumido que su distribución -en series- es la misma que la que se da para la LMC en [Soszynski-09], por ser este el artículo de referencia para este tipo de estrellas variables.

Por lo que respecta a la obtención en este trabajo del “survey” de las LPV, el proceso en general es en todo similar a los descritos hasta ahora en el punto 4.6 para las Cefeidas, y en el punto 4.8 para las RR Lyrae, si bien tiene la particularidad de añadir dos parámetros adicionales: **serie**, que indica la serie a la que pertenece cada estrella sintetizada y **K_s** , que es el índice de color de cada estrella sintetizada.

En el caso de las LPV, una vez obtenidas las coordenadas de nuestras estrellas sintéticas (de manera exactamente igual a como se llevó a cabo con Cefeidas y RR Lyrae), el paso siguiente es el cálculo del resto de parámetros característicos de las mismas, empezando por la terna (**K_s** , **logP**, **serie**), donde el período se expresa en días en escala logarítmica. Para ello, y en base a [Soszynski-09] para la LMC y Vía Láctea y [Yta-03] para la SMC, dispone de la función *stars.generator()* en el fichero “*functions-Long_Period.R*”, que genera dicha terna para cada estrella sintetizada. Un vector con las ternas antes mencionadas, se almacena en un objeto denominado *LMC* (para la LMC), *SMC* (para la SMC) y *GLX.xxx* (para la Vía Láctea: discos en los que “xxx” puede valer “thin”, “thick” o bulbo, en los que “xxx” vale “bulge”). Posteriormente se asigna el valor de los vectores columna de la matriz antes indicada conteniendo K_s , logP y de la serie a los vectores *S.xxx*, *K_s.xxx* y *logP* que almacenan la serie, K_s y el logaritmo del período, respectivamente. El sufijo “xxx” puede tomar los valores: “lmc”, “smc”, “thin”, “thick” o “bulge”. Existe también la posibilidad, en función de los parámetros de configuración del programa ya descritos, de generar el valor K_s de manera aleatoria, según una distribución uniforme.

La distribución de estas estrellas en series (en 5 series principales: A, B, C1, C2 y D) representa una dificultad adicional, por lo que para su generación (ver punto 4.11.2) se ha usado la técnica de cópulas estadísticas descrita en el apéndice D de este documento.

Una vez que se dispone de la terna de parámetros antes indicados, así como de las coordenadas (distancia **d**, y coordenadas ecuatoriales **alpha** y **delta**), es posible calcular el parámetro M_k , mediante la expresión ya conocida del punto 1.2

$$M_x = m_x + 5 - 5 \times \log d$$

Donde “x” toma el valor “k”, banda en la se trabaja en este tipo de variables de forma habitual; la función $m.k()$ del fichero “*functions-Long_Period.R*” es la que realiza este cálculo de M_k ; este parámetros en el código se denomina **MK**. Con M_k calculado se calcula M_v , parámetro que en el código se denomina **MV** y que pertenece al espacio de parámetros mediante la función $m.v()$ del fichero antes indicado. Esta última función realiza el cálculo:

$$M_v = M_k + (V - K)$$

Es decir: pasa la magnitud M de la banda K a la banda V (visible); a su vez, $(V-K)$ según [Valenti-04] puede expresarse en función de M_k según se indica en las gráficas de dicho artículo; la curva de la gráfica se aproxima por un modelo lineal del tipo:

$$(V - K) = b_0 + b_3 \times M_k^3$$

En definitiva, se tiene que:

$$M_v = b_0 + M_k + b_3 \times M_k^3$$

Que resulta ser la expresión que ejecuta la función $m.v()$ antes indicada. Con el parámetro **MV** calculado, se procede al cálculo del parámetro **VI**, utilizando las correlaciones de [Soszynski-09] para LMC y SMC y de [Yoshifusa-03] para la Vía Láctea a través de la función $vi.pl()$ del fichero “*functions-Long_Period.R*”. Finalmente se calcula el parámetro **Amp** con la función $a.v()$, también incluida en el fichero “*functions-Long_Period.R*”, aproximando dicho valor por una distribución normal que emule los gráficos de amplitudes de [Groenewegen-04].

El resto de parámetros se calculan de manera análoga a como se describió en el punto 3.1. En el caso de las LPV, hay, como en las Cefeidas, 5 tipos de variables dependiendo de su ubicación:

- **type** = 1 → LMC.
- **type** = 2 → SMC.
- **type** = 3 → Disco fino.
- **type** = 4 → Disco grueso.
- **type** = 5 → Bulbo Galáctico.

La matriz ‘P.survey’ obtenida, debidamente procesada, puede presentarse en pantalla si bien para el número de estrellas calculadas, 10^5 en este caso, es poco práctico; de hecho,

el código solo permite presentar los resultados por pantalla para un “survey” de no más de 5.000 estrellas; en cualquier caso se almacena en el fichero de texto “*results-Long_Period.dat*”, y en un objeto ‘R’ en el fichero “*Long_Period.Rdata*”, fácilmente manipulable con ‘R’, tal y como se indicó en el punto 4.2.

4.11. Validación del “survey” de LPV

Como en el caso de las Cefeidas y RR Lyrae, en este punto se describirá el proceso de validación de los resultados dados por el código ‘R’ contenido en los ficheros “*LMC-SMC-GLX-Long_Period*” y “*functions-Long_Period.R*”; también como en el caso de las Cefeidas y RR Lyrae esta validación se lleva a cabo con el código contenido en los ficheros “*plotset1-Long_Period*” y “*plotset2-Long_Period*”.

Como ya se ha indicado con las restantes variables sintéticas, el código de los ficheros ‘R’ “*plotset1-Long_Period*” y “*plotset2-Long_Period*”, lee los datos resultantes de la ejecución del código de los ficheros “*LMC-SMC-GLX-Long_Period*” y “*functions-Long_Period*” que están contenidos en el fichero de texto “*results-Long_Period*”, y generan las graficas que muestran las correlaciones existentes entre los distintos parámetros sintetizados para cada una de las estrellas del ‘survey’ obtenido. Este procedimiento es el mismo que ya se ha descrito para las Cefeidas y RR Lyrae.

Los datos del fichero “*results-Long_Period*” se almacenan durante la ejecución del código de validación en la variable ‘*data*’, y a partir de ahí se trazan las gráficas de las correlaciones existentes en el espacio de parámetros. Como en el caso de las Cefeidas y RR Lyrae, se han obtenido 100.000 estrellas sintéticas para este “survey” (si bien el número de estrellas es un parámetro variable del programa).

4.11.1. Histograma de periodos

Inicialmente se confecciona un histograma de los valores obtenidos para el parámetro **logP** para todas las LPV, de manera que pueda observarse la distribución de dicho parámetros. Para el caso de las cefeidas, dicho histograma se muestra en la figura 35.

Puede observarse que los periodos se distribuyen mayoritariamente entre los valores $\log P = 0.25$ (1.77 días) y $\log P = 3.5$ (3.162,2 días), según se muestra en la figura 35, que es la variación de periodo generalmente aceptada para este tipo de estrellas variables según [Sonszynski-09] (más de 99.965 de un total de 100.000 estrellas que contiene el “survey” de cefeidas obtenido).

Podemos valernos también de ‘R’ para obtener un resumen de la columna correspondiente al logaritmo de los periodos de nuestras estrellas sintéticas:

```
> summary(logP)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.1393 1.4840  2.2690  2.1740  2.8840  3.5910
```

Que son unos valores razonables para el logaritmo del periodo.

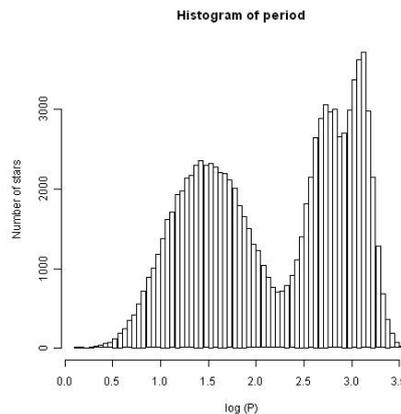


Figura 35: *Histograma de periodos en escala logarítmica de todas las LPV simuladas (en LMC, SMC, discos y bulbo).*

4.11.2. Histograma de color (K_s) y series

Como ya se ha indicado en el punto 4.10, en este tipo de estrellas aparecen dos parámetros nuevos: K_s y *serie*. Se han simulado dichos parámetros en base a [Soszynski-09] para la LMC, e [Yta-03] para la SMC. En el caso de la Vía Láctea, dado que no hay datos, se ha asumido que dicha gráfica es la misma que la de la LMC según [Soszynski-09]; la figura 36 muestra la grafica obtenida para las estrellas LPV simuladas.

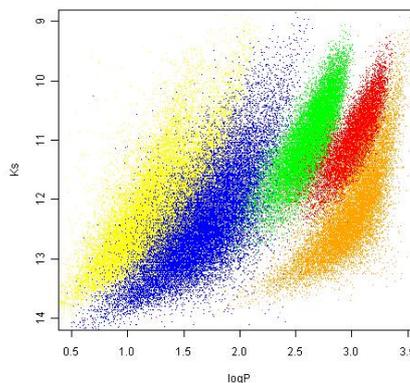


Figura 36: Gráfica K_s – logP de todas las LPV sintéticas. En color amarillo serie A; en color azul serie B; en color verde serie C1, en color rojo serie C2, en color naranja serie D. Simulación mediante cópulas; 100.000 estrellas sintéticas.

La figura 37 muestra el histograma de K_s para las LPV simuladas; K_s se distribuye mayoritariamente entre 9.5 y 14. Ambas figuras están de acuerdo con [Soszynski-09] (figura 37a) y [Yta-03] (figura 37b).

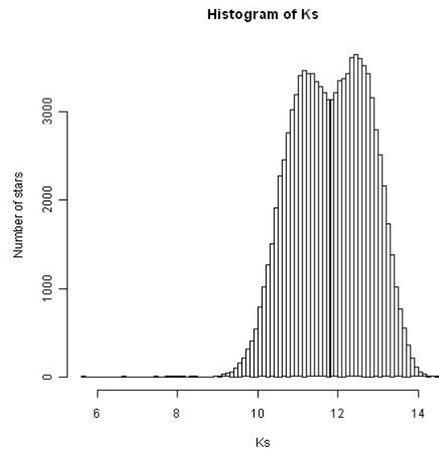


Figura 37: Histograma de todos los valores de K_s de las variables LPV simuladas.

Un resumen del parámetro K_s sintetizado en el “survey” de LPV se muestra a continuación:

```
> summary(Ks)
  Min. 1st Qu.  Median  Mean   3rd Qu.  Max.
5.622 11.090  11.840 11.820 12.570 14.680
```

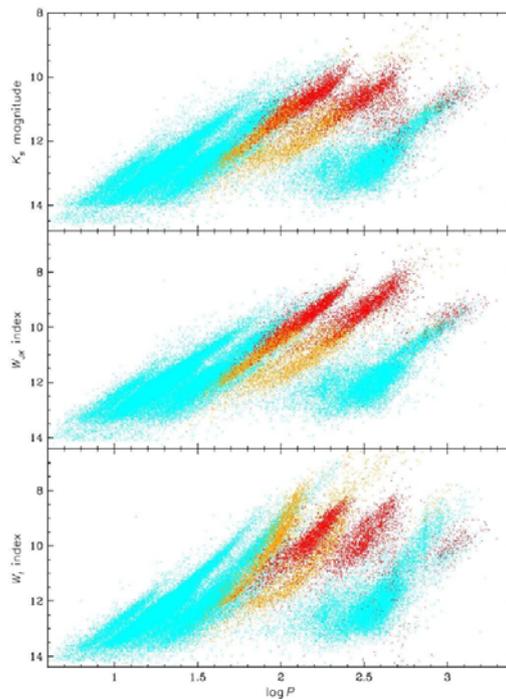


Figura 37a: distribución de K_s en función de $\log P$ para LMC (figura superior). Pueden verse las distintas series: en azul serie A y B; en rojo serie C1; en rojo-anaranjado serie C2 y en azul serie D. Extraído de [Soszynski-09].

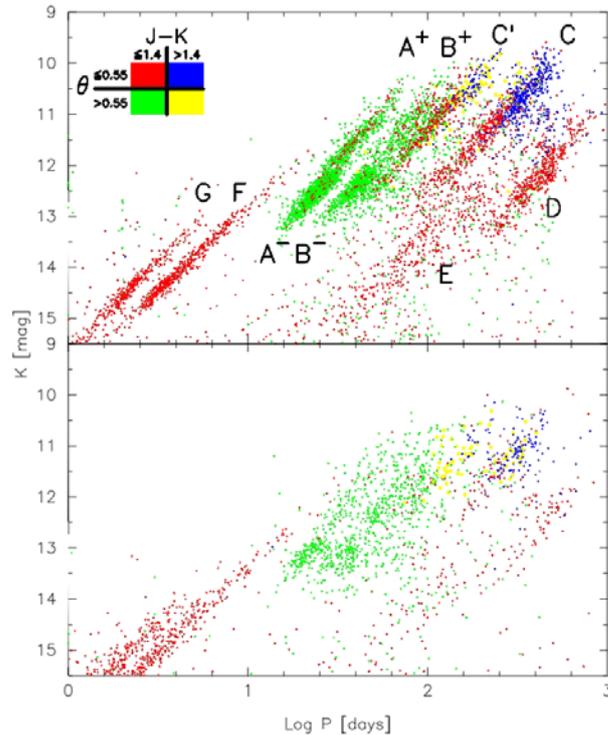


Figura 37b: distribución de K_s en función de $\log P$ para LMC (figura superior) y SMC (figura inferior). Pueden verse las distintas series: en rojo (izquierda) series G y F (no considerada en este trabajo); en verde serie A y B; en rojo-amarillo serie C1 (C'); en rojo-azul serie C2 (C); y en rojo (abajo a la derecha) serie D. Extraído de [Yta-03].

4.11.3. Representación de las coordenadas (Coordenadas Ecuatoriales)

El proceso de cálculo de las coordenadas ecuatoriales, y los resultados son análogos a los obtenidos para las Cefeidas y RR Lyrae: ver punto 4.7.2. Dado que en ambos casos responden a la misma distribución, la gráfica obtenida para las LPV es muy similar a la obtenida para las Cefeidas y RR Lyrae, por lo que no se representa aquí otra vez.

4.11.4. Representación de las coordenadas (Coordenadas Galácticas)

Como en el caso anterior, el proceso de cálculo de las coordenadas galácticas, y los resultados son análogos a los de las Cefeidas y RR Lyrae: ver punto 4.7.3. Dado que en ambos casos responden a la misma distribución, la gráfica obtenida para las LPV es muy similar a la obtenida para las Cefeidas y RR Lyrae, por lo que no se representa aquí otra vez.

4.11.5. Visualización de objetos en Banda G

Procediendo de la misma manera que se hizo para las Cefeidas y RR Lyrae, se obtienen unos resultados para las LPV análogos a los que obtuvieron para las Cefeidas en el punto 4.7.4, por lo que no se incluye aquí la gráfica, al no aportar información adicional a la obtenida para las Cefeidas sintéticas. Analizando la banda G con más detalle en el “survey” obtenido, se obtienen los siguientes datos:

```
> summary(g)
  Min. 1st Qu. Median Mean 3rd Qu.  Max.
11.77  17.19  17.56  17.57  18.23  21.00
```

4.11.6. Visualización de los valores de Extinción (A_v)

También en este caso tanto la gráfica obtenida de los valores de extinción para las variables sintéticas LPV así como las conclusiones que se pueden extraer de la misma son en todo similar a las obtenidas para las Cefeidas sintéticas del punto 4.7.5, y para las variables RR Lyrae, por lo que no se incluye la citada gráfica al no añadir información adicional.

Los valores de extinción son de nuevo coherentes: son muy elevados en aquellas estrellas cuyas coordenadas corresponden a una línea de visión que pasa por el plano galáctico, donde están ubicados los planos de los discos de la galaxia, y por tanto donde mayor es la concentración estelar. En cuanto a los valores de extinción A_v , se muestra a continuación un resumen de sus estadísticos más importantes:

```
> summary(Av)
  Min. 1st Qu. Median Mean 3rd Qu.  Max.
0.0000 0.0000 0.0000 0.5766 0.0000 8.6390
```

Obsérvese que hay valores de la extinción nulos: esto es así porque en el “survey” de las LPV las extinciones de las variables de la LMC y SMC se hecho nula.

Operando con ‘R’, y teniendo en cuenta los artículos y estudios referenciados en el punto 4.6, se puede comprobar que el “survey” sintético obtenido para las RR Lyrae tiene todas las estrellas con su valor A_v dentro de límites razonables (teniendo en cuenta que dicho valor de A_v es nulo para las variables de la LMC y SMC), por lo que el resultado es coherente con las observaciones y artículos de referencia utilizados.

4.11.7. Diagrama HR

Seguidamente, se estudian los resultados obtenidos para el diagrama HR de las LPV sintéticas, en base a los artículos [Soszynski-09] y [Yta-03]. Operando con ‘R’ en las LPV sintéticas obtenidas, obtenemos los siguientes resúmenes:

```
> summary(VI)
  Min.   1st Qu.  Median   Mean   3rd Qu.   Max.
0.2037  0.4973  0.5895  0.5818  0.6689  0.9166
```

```
> summary(MV)
  Min.   1st Qu.  Median   Mean   3rd Qu.   Max.
-1.4650 -1.2950 -0.8308 -0.6295 -0.1683  5.8520
```

La figura 38, muestra los resultados obtenidos para la Galaxia, LMC y SMC respectivamente.

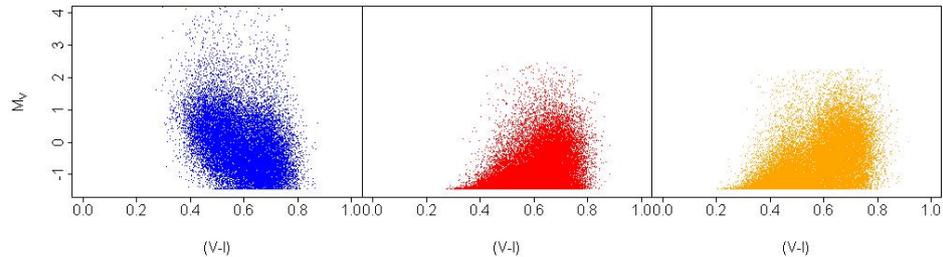


Figura 38: Diagrama HR para las LPV sintéticas: azul galaxia (Disco fino + Disco grueso + Bulbo Galáctico); rojo LMC y naranja SMC.

Puede observarse como por debajo de $M_v = -1.465$ no aparece ninguna de las estrellas sintéticas del “survey” sintético obtenido.

4.11.8. Diagrama logP-MV

Este diagrama nos muestra la relación entre el logaritmo del periodo y la magnitud absoluta de la estrella (a 10 pc.), y se muestra en la figura 39.

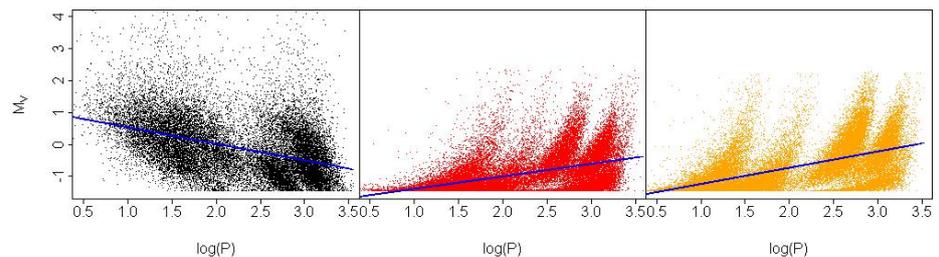


Figura 39: Diagrama logP-MV para las estrellas sintéticas LPV: el color negro se corresponde a las estrellas de la galaxia (Disco fino + Disco grueso + Bulbo galáctico), el rojo a las de la LMC y el naranja a las estrellas de la SMC

En el punto 4.10 se ha descrito extensamente el procedimiento para obtener M_v , y las fuentes que justifican dicho procedimiento.

4.11.9. Diagrama logP-VI

Este diagrama nos muestra la correlación entre el logaritmo del periodo y el color VI; los resultados obtenidos en nuestra simulación se representan en la figura 40.

Los resultados de la correlación obtenida están de acuerdo con los valores observados en [Collinge-06] para el caso de la galaxia que se muestran en la figura 32a.

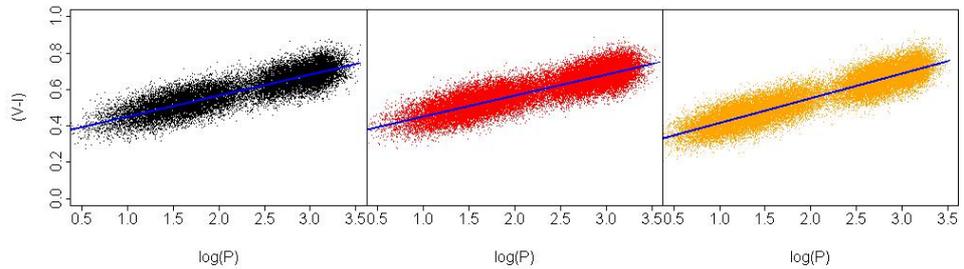


Figura 40: Diagrama logP-VI para las estrellas LPV sintéticas: negra para la galaxia (Disco fino + Disco grueso + bulbo galáctico); rojo para LMC y naranja para SMC.

4.11.10. Diagramas PLC

Las siguientes figuras (figura 41a, 41b y 41c) muestran las gráficas tridimensionales (la tercera dimensión es el color) PLC de la Vía Láctea, LMC y SMC donde esta última se representa por una escala de colores amarillo-naranja-rojo, para las variables LPV sintéticas del “survey” obtenido.

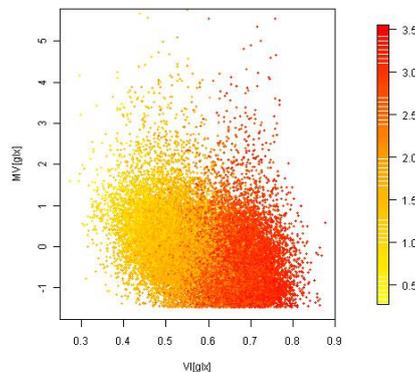


Figura 41a: Diagrama de la Vía Láctea PLC de la Galaxia para las LPV: en escala de colores se representa log P.

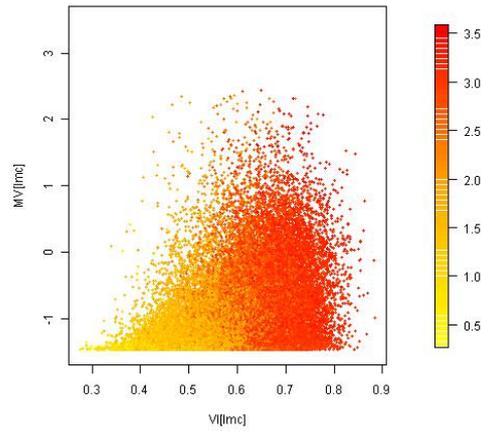


Figura 41b: Diagrama de la Vía Láctea PLC de la LMC para las LPV: en escala de colores se representa $\log P$.

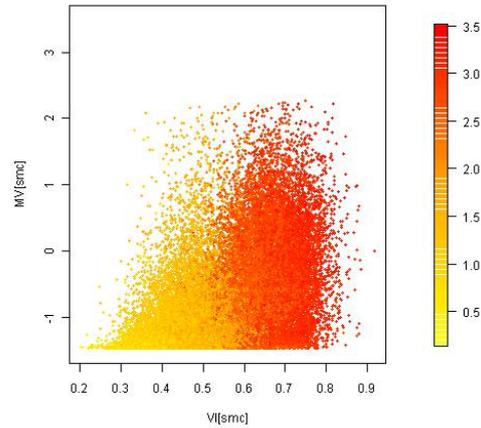


Figura 41c: Diagrama de la Vía Láctea PLC de la SMC para las LPV: en escala de colores se representa $\log P$.

Esta gráfica combina los valores de los tres parámetros que representa, y que ya se han descrito en figuras y diagramas anteriores.

4.11.11. Diagrama de Amplitudes

En este punto, se muestra el diagrama de amplitudes de las estrellas LPV sintetizadas; la gráfica Amplitudes-logaritmo del periodo se muestra en la figura 42. En él puede observarse que la variación máxima de la amplitud está en el intervalo $[-1.5, 0.5]$ para LMC, SMC y Vía láctea, lo que está de acuerdo con [Soszynski-09] (figura 42a). De [Groenewegen-04], se ha usado la relación $\frac{A_l}{A_v} = 0.49$ para obtener A_v partiendo de A_l .

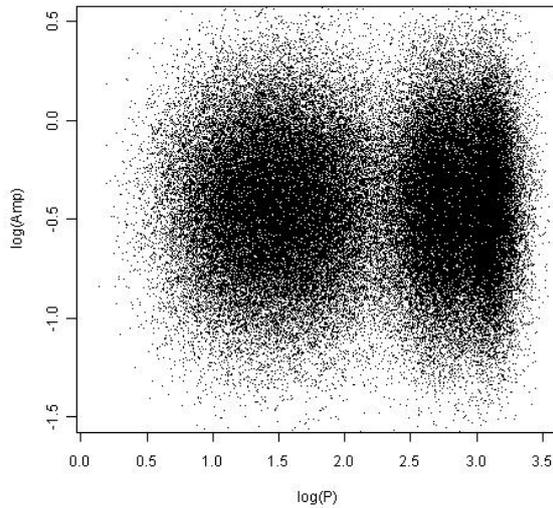


Figura 42: *Diagrama de Amplitudes de las estrellas LPV sintéticas; están representadas conjuntamente las de la Vía Láctea, LMC y SMC.*

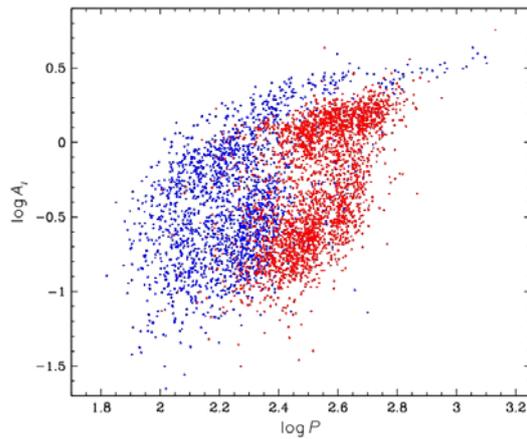


Figura 42a: *Diagrama de Amplitudes de las RR Lyrae del núcleo de la Vía Láctea obtenido de [Soszynski-09]. Los puntos azules representan las variables ricas en Oxígeno; los puntos rojos, las variables ricas en Carbono.*

4.11.12. Diagrama distancia-error de paralaje

También para las LPV se ha hecho la gráfica distancia – error de paralaje; su aspecto e interpretación es en todo análoga a la misma gráfica obtenida para las Cefeidas (ver punto 4.7.11) y RR Lyrae, por lo que no se incluye aquí al no aportar información adicional.

4.11.13. Diagramas de Lutz-Kelker

En el caso de los diagramas de Lutz-Kelker encontramos que su interpretación y aspecto es análogo al obtenido en el caso de las Cefeidas (ver punto 4.7.12) y RR Lyrae, por lo que no se incluye aquí al no aportar nueva información.

4.11.14. Histograma de distancias

La gráfica obtenida es de nuevo análoga a la obtenida en el caso de las Cefeidas (ver punto 4.7.13 para más información) y las RR Lyrae.

Capítulo 5: *CONSIDERACIONES FINALES*

5.1 Introducción

En este último capítulo, se van a obtener las conclusiones obtenidas de la realización de este trabajo, y se van a indicar los trabajos futuros que quedan aún por realizar. Ya en el capítulo 4 se ha descrito la implementación de una de las subtareas en las que se ha dividido la tarea central del SE objeto del trabajo; también se ha realizado en el mismo capítulo la validación de dicha implementación, en base siempre al conocimiento extraído del experto astrofísico y/o de sus indicaciones.

Se describirán las dificultades encontradas para la realización del SE deseado, y sobre todo aquellas relacionadas con la captura del conocimiento, así como los trabajos futuros que aún restan para la conclusión del SE, y posibles mejoras que podrían introducirse.

Es importante señalar que el diseño de un SE para analizar la calidad de GAIA es en sí mismo uno de los objetivos principales de la misión, ya que la misma como se ha indicado anteriormente durará 5 años con una posible extensión de 2 años más: cada 6 meses se repetirán los análisis sobre las observaciones realizadas, y por otra parte la rotación del personal investigador que trabajará sobre los resultados de GAIA será elevada: contratos que terminan, causas personales, etc. Dotar al proceso de análisis de un elevado grado de automatización permitirá tener procesos de análisis estables, homogéneos y bien definidos, independientes de los factores antes indicados.

5.2 Resultados Obtenidos

Al margen de la descomposición en tareas y la descomposición inferencial ya descritas en el capítulo 3, se han obtenido los “surveys” sintéticos para Cefeidas, RR Lyrae y LPV. Los ficheros con los resultados se incluyen en el anexo D en el que está disponible el código; ya en capítulo 4 se describió en detalle la estructura del código y nombre de todos los ficheros, incluidos los de resultados. Dado el volumen de datos, se recomienda el examen directo de dichos ficheros de resultados para ver en detalle la estructura del mismo, que por otro lado es una estructura plana en un fichero de texto.

Para la obtención de las muestras sintéticas, se han utilizado las correlaciones de BB.DD completas, en el sentido en que dichas BB.DD no estén sesgadas, por ejemplo por no incluir las estrellas más débiles de un determinado tipo de variabilidad. Los catálogos que no tienen sesgos, se denominan *catálogos completos*; este tipo de catálogos se caracteriza por tener un elevado número de estrellas, y unos sistemas de observación avanzados muy sensibles que les permita alcanzar un elevado grado de calidad observacional.

El “survey” que mejor se adapta a las características antes mencionadas, es OGLE, si bien es imposible evitar los *sesgos* de manera absoluta: por ejemplo OGLE puede perder alguna RR Lyrae, que son más débiles que las Cefeidas.

Se han utilizado otros “surveys”, como Hipparcos, MASS, etc., aunque son BB.DD más sesgados que OGLE. Existen otros “surveys” que por estar orientados a un tipo de estrella específica presentan un sesgo grande, pero son muy útiles para generar estrellas sintéticas del tipo al que están orientados.

5.3 Trabajos Pendientes

Evidentemente, el trabajo pendiente más importante, es proceder a la implementación de las restantes subtareas, e integrarlas en un todo único estructurado que realice la función deseada que constituya el SE en cuestión.

Para alcanzar este objetivo, es necesario que otras unidades de trabajo de GAIA completen algunas de estas tareas: algunas de las subtareas descritas en el capítulo 3 pueden ser realizadas por otros equipos; en este caso, dichos equipos podrían verse desde nuestro SE como agentes que realizan un trabajo determinado. En este caso, podría ser conveniente modular la técnica de captura y modelado de conocimiento utilizando la metodología multiagente MAS-CommonKADS, que se adapta mejor a entornos colaborativos (o distribuidos) de conocimiento. En esta metodología son muy importantes la fase de conceptualización y la de análisis; el propósito de la fase de conceptualización es comprender el problema, identificar los actores activos y pasivos, además de elaborar los requisitos que permitan formular unos objetivos para generar la solución. En la fase de análisis se definen los modelos de agente, tareas, coordinación, comunicación, organización y experiencia (conocimientos).

Desde el punto de vista estrictamente funcional, sería preciso ampliar los “surveys” simulados con otros adicionales que nos permitan afinar más la evaluación de calidad de CU7. En este trabajo, se han utilizado las variables Cefeidas, RR Lyrae y LPV; la elección de estas variabilidades no es caprichosa: se trata de tres tipos de variabilidad bien conocida, que presentan amplitudes elevadas y periodos relativamente bien caracterizados, pero existen otras variabilidades que deberían de ser incorporadas al estudio de calidad de CU7:

- Estrellas variables binarias eclipsantes, un tipo específico de estrella variable que en realidad está formado por un par de estrellas de diferente luminosidad relativamente cercanas (tanto que desde la tierra se ven como una única estrella), y en la que uno de los componentes del par eclipsa al otro de manera periódica debido a la rotación orbital de dicho componente: existen Bases de Datos reales de dicho tipo de estrellas de calidad suficiente como para ser tenido en cuenta en los trabajos de QA. Estas estrellas pueden presentar

cualquier frecuencia y amplitud, por lo que pueden permitir explorar los límites de sensibilidad de GAIA.

- Estrellas variables δ -Scuti: Se trata de estrellas variables menos conocidas, en el sentido de que su mecanismo de variabilidad no está aún totalmente comprendido por los astrofísicos. Presentan unas frecuencias muy elevadas (o cortos periodos de solamente horas) y pueden tener cualquier valor de amplitud.
- Estrellas variables γ -Doradus: Estrellas variables muy interesantes desde el punto de vista astrofísico, pero también con un mecanismo de variabilidad poco conocido. Presentan periodos del orden de 1 ó 2 días, y una amplitud relativamente baja.

Por otro lado, sería conveniente localizar algún estudio que permitiera la inclusión en el espacio de parámetros de trabajo de algunos de los parámetros que permita analizar CU7 con una mayor calidad, como por ejemplo temperatura (obtenida con el BP de GAIA, ver punto 1.2) y gravedad (obtenida con el RP de GAIA, ver punto 1.2). Estos parámetros son analizados por otra unidad de DPAC, en concreto CU8, y tiene interés analizar si la inclusión en los parámetros simulados de los “surveys” sintéticos obtenidos de la influencia de los mismos puede afinar los estudios de calidad realizados.

Finalmente, es esencial la inclusión de la descomposición de Fourier de las frecuencias de variabilidad de los objetos analizados en los “surveys” sintéticos, parámetro no incluido en este trabajo, y quede de ser incluido por su incidencia en los resultados obtenidos.

Por otra parte, puede ser conveniente resolver la ausencia de valores de algunos parámetros, pues como ya se ha indicado en capítulos anteriores pues como ya se ha indicado, en algunos de los objetos que se han simulado (discos y núcleo galáctico), no existen análisis previos que permitan aplicar correlaciones bien estudiadas; esto como ya se ha indicado es debido a que se han estudiado con mayor intensidad la LMC y SMC que nuestra propia galaxia: en algunos casos esta ausencia de datos es notoria, como en el caso de las variables LPV, para las que hay escasos datos acerca de las mismas. Aunque el experto puede dar siempre pautas para reducir el impacto de dichos valores ausentes, un análisis basado en observaciones siempre es más adecuado.

Desde el punto de vista estadístico, y en lo que se refiere a la tarea de evaluación de igualdad de pdf multivariantes, debería de analizarse algún método alternativo al más conocido T^2 de Hotelling, ya que este último asume que la distribución multivariante bajo análisis es normal; asumir que las pdf obtenidas del procesamiento de varios de los parámetros que constituyen los “surveys” (reales o sintéticos) utilizados, es en general una restricción fuerte.

La principal dificultad que se ha encontrado para la confección de este trabajo, ha sido sin duda la captura del conocimiento del experto, como por otra parte era de prever, ya que en todo proceso de modelado del conocimiento, la captura del mismo es la fase más

comprometida y difícil: las causas de de esta dificultad son conocidas, y bien descritas en [Musen-93].

5.4 Resumen final

A lo largo de esta memoria, se ha presentado el diseño de un SE experto que puede ser utilizado como una herramienta permita evaluar la calidad de los datos obtenidos por la misión espacial GAIA, que son de un volumen enorme. Se han utilizado las técnicas de captura de conocimiento y modelado del mismo según una metodología bien establecida, describiendo el dominio en el que debe de trabajar dicho sistema experto, especificando los objetos del mismo y las relaciones establecidas entre ellos, así como entendiendo las técnicas de análisis necesarias en este dominio:

- Astrofísica (conocimientos cualitativos) en cuanto a modelos de variables.
- Métodos estadísticos.
- Análisis de Series temporales.

Y por supuesto, programación en 'R'. Se ha procedido también a la implementación de una de las subtareas en las que se ha descompuesto la tarea del SE.

En estas condiciones, este trabajo constituye un intento serio de aplicar SE en el dominio de la astrofísica, en el que existen muy pocas aplicaciones de SBC en general, y de SE en particular, a pesar de ser uno de los dominios en los que a priori más fácilmente sería aplicar técnicas de IA/IC, debido a la gran cantidad de datos de manejan actualmente los astrofísicos, que quedan almacenadas en Bases de Datos, sin que en muchos casos sean conocidas por los propios astrofísicos, en parte por falta de una ontología que permita facilitar la comunicación entre los expertos de área.

Finalmente, se han descrito algunas mejoras que pueden incidir positivamente en la calidad de dicho SE, siempre dentro de las restricciones impuestas por GAIA.

Madrid, Agosto de 2.012

APENDICE A

a. Antecedentes de este trabajo

Este Trabajo Fin de Máster, se encuadra dentro de las actividades que la U.N.E.D está realizando para esta misión espacial, dentro del paquete de trabajo GWP-M-721 del CU7 o Unidad de Coordinación 7. Puede encontrarse una descripción muy detallada de la organización establecida para el análisis de los datos enviados por dicha misión en el apéndice A.

En particular, el autor de este documento ya colaboró en al año 2.008 en el proyecto mediante la elaboración de un PFC en la UNED (Proyecto Fin de Carrera denominado): **“Desarrollo de software de análisis estadístico de 'surveys' de variabilidad para la misión espacial Gaia”** [García-08], el cual:

- Crea la estructura general de desarrollo indicado por el DPAC.
- Simula los procesos FAR (“False Alarm Rate”), consistentes en averiguar la probabilidad de que los vectores obtenidos correspondan o no a los objetos de estudio, y TF (“Transfer Function”), que determina cual es la probabilidad de observar un objeto de este tipo. Ambos procesos sirven para eliminar el sesgo de la muestra, y se ejecutan varias veces sobres los datos.
- Estima la función de densidad de las muestras mediante el uso de Kernel fijo sobre una malla n-dimensional equiespacialmente distribuida.
- Implementa la reducción de la dimensionalidad de los datos con el fin de poder representarlos.

Posteriormente, Enrique Atienza, también desde la UNED continuó el trabajo iniciado en el PFC antes mencionado a través de su propio PFC en 2.010: **“Optimización de los Estimadores de Densidad de Probabilidad e Implementación de Test Estadísticos Multivariantes para la misión espacial ‘GAIA’”** [Atienza-10], en el que se establecieron los siguientes objetivos:

- Cambiar la malla n-dimensional equiespacialmente distribuida por un mapa auto-organizativo. En este proceso se parte de una malla ortogonal de puntos y a continuación se entrena con datos de entrada con el fin de que ésta se adapte, progresivamente, a la distribución de la muestra. Esta malla optimizará la posterior estimación de densidad.
- La estimación de densidad a través de Kernel adaptativo, desarrollada de una manera óptima, puede llevar a mejoras significativas sobre los anchos de banda fijos o los procedimientos no adaptativos. Sin embargo, procedimientos no óptimos, o bien la característica particular de los datos, pueden desembocar en resultados muy inferiores. En la práctica, deben emplearse diversas metodologías. Por consiguiente se pretende estimar la densidad mediante el

uso de Kernel adaptativos de diversos tipos en los puntos de la malla, ya entrenada mediante el mapa auto-organizativo.

- El anterior proyecto usaba Kernel de matrices diagonales, las cuales siempre están orientados a los ejes de coordenadas, esto no resulta óptimo para muestras no paramétricas de muy distintos tipos. Por tanto, se implementarán Kernel de matrices completas con el fin de que el elipsoide del hiperespacio (el Kernel en sí mismo) se oriente arbitrariamente sobre la distribución de los datos.
- Por último, se crearon ficheros de test para todas las clases y con la estructura necesaria para realizar los test científicos (realizados por el DPAC en el mes de noviembre de 2009). Estos test tuvieron como objetivo comparar las distintas propuestas de Kernel adaptativos, contrastarlas con los Kernel fijos, estimar los tiempos de proceso necesarios y optimizar los parámetros de tiempo real.

Además, en paralelo se realizan otros trabajos, que implican las siguientes tareas:

- Escribir un “Taker” que calcule media, varianza y rangos de una muestra.
- Test multivariantes (Hotelling T^2 y rangos).
- Etiquetado de datos atípicos (“outliers”).
- Normalización de las PDF estimadas.
- Promediado de las PDF estimadas.
- Test científicos de regresión multivariante con cefeidas.
- Optimizar los parámetros de algunas de las clases implementadas.
- Implementación de la reducción dimensional en hiper mayas n-dimensionales irregulares.
- Integrar algunas clases implementadas en código de SOM.

Y varios más.

b. Estructura del DPAC

Para hacer frente al reto que supone el procesar la ingente cantidad de datos que la misión GAIA va a recoger, la comunidad científica junto con la ESA, han establecido una infraestructura que abarca sistemas de procesado de datos específicos tratando de evitar la duplicación de los trabajos y estableciendo los roles necesarios para realizarlos, y que asume la responsabilidad de generar y publicar todos los resultados intermedios y finales de la misión científica GAIA.

Los principios subyacentes del procesado de datos para la misión GAIA han sido desarrollados por la comunidad científica, y sus piezas constituyentes han sido simuladas con éxito en simulaciones más o menos complejas. Durante esta fase de definición y simulación, se ha tratado de identificar los aspectos críticos que implicaba el procesado de los datos que GAIA recogerá (volumen de datos, calibraciones, potencia computacional y

otros muchos), para anticipar los riesgos inherentes a una misión del calibre de GAIA, sin precedentes en la astrofísica.

Como resultado de estos trabajos preliminares, la comunidad científica se ha organizado e integrado con el objetivo de unir sus fuerzas y evitar trabajos improductivos, creando un consorcio específico para el análisis de los datos obtenidos con la misión GAIA: el **DPAC**, siglas correspondientes a la descripción inglesa “Data Processing and Analysis Consortium”. De manera resumida, DPAC es una comunidad científica cooperativa Europea que incluye el Centro Científico de Operaciones de GAIA ó SOC (“Science Operations Centre”), y un consejo ejecutivo. DPAC está compuesto por unos 300 individuos pertenecientes a la comunidad científica repartidos en más de 15 países, e incluyendo 6 grandes centros de procesamiento de datos (ver tabla 3). Este consorcio ha estimado de la manera más precisa posible los recursos necesarios para el tratamiento de los datos de GAIA, y ha puesto bajo el control de una sola organización el material, los recursos humanos y los presupuestos adecuados hasta la finalización del tratamiento de los datos generados por la misión GAIA, final estimado alrededor del año 2.020. El pasado año 2.006 el DPAC propuso a la ESA un sistema de tratamiento de datos capaz de gestionar el volumen y la complejidad de los datos obtenidos a lo largo de la misión.

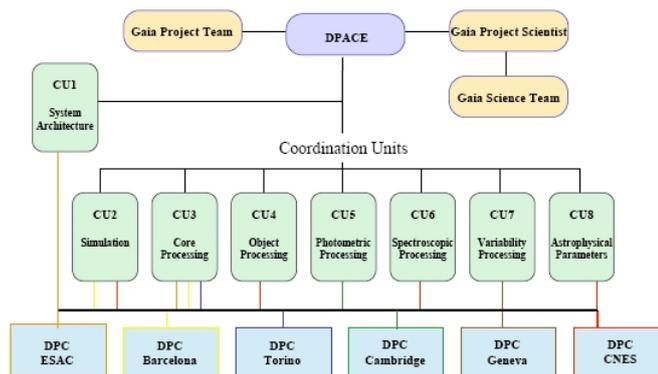


Figura 1: Estructura organizativa del DPAC, con los CU y DPC. Puede verse CU7 y su localización dentro del mismo.

Como puede verse en la figura 1, el consorcio DPAC está organizado como un grupo de 8 unidades de coordinación o CU (“Coordination Units”) con responsabilidad sobre una tarea específica de procesamiento de datos, y un comité ejecutivo denominado DPACE (“Data Processing and Analysis Consortium Executive”), los CU son los elementos básicos de DPAC. Se ha tratado de definir de la manera más precisa posible las interfases entre diferentes CU, así como la tarea responsabilidad de cada uno de los 8 CU, y sus atribuciones (ver tabla 2), que de manera resumida son:

- Definir tareas concretas de proceso de datos y responsables para cada una de las mismas.
- Establecer prioridades en los trabajos a realizar.
- Diseñar, optimizar, ensayar e implementar algoritmos de procesamiento de datos.

- Verificar la calidad de los trabajos y productos realizados.

CU / Filiación	Localización	Responsable	Responsabilidad
CU1 / ESAC	Madrid	W. O'Mullane	Arquitectura sistema
CU2 / UB	Barcelona	X. Luri	Simulación
CU3 / ARI	Heidelberg	U. Bastian	Procesado básico
CU4 / ULB	Bruselas	D. Pourbaix	Procesado de objetos
CU5 / IOAC	Cambridge	Van Leeuwen	Procesado optométrico
CU6 / OBSPM	Paris	D. Katz	Procesado de espectros
CU7 / Obs. Ginebra	Ginebra	L. Eyer	Procesado variabilidad
CU8 / MPIA	Heidelberg	C. Bailer Jones	Parámetros físicos

Tabla 7: Localización y Responsables de cada una de las CU del DPAC

Cada CU dispone de un responsable, que está ayudado por uno o dos asistentes y cuando es necesario, también por un responsable técnico. Este equipo directivo de cada CU, es responsable de adquirir y gestionar los recursos de dicho CU, así como las actividades del mismo.

El consorcio DPAC como tal, está coordinado a su vez al más alto nivel por el comité ejecutivo denominado DPACE. Este comité es responsable de temas no específicos o no relacionados directamente con la gestión directa de los CU, como la definición de los mecanismos de interacción e intercambio de datos entre los CU y entre éstos y el resto de la comunidad científica. Actualmente este Comité ejecutivo está compuesto por un presidente, un secretario, por los responsables de cada CU, y por un representante del CNES (“Centre National d’Etudes Spatiales”). Además de los 8 CU mencionados, que están ya operativos, existe un noveno de futura creación, que será responsable de desarrollar las herramientas necesarias para que la comunidad científica pueda acceder al Catálogo de GAIA.

DPC	Localización	Responsable	CU asociado
DPC-B	Barcelona	S. Girona	CU2, CU3
DPC-C	CNES	X. Passot	CU2, CU4, CU6, CU8
DPC-E	ESAC	J. Hoar	CU1, CU3
DPC-G	Ginebra	M. Beck	CU7
DPC-I	Cambridge	Van Leeuwen	CU5
DPC-T	Turin	A. Volpicelli	CU3

Tabla 8: Tabla donde se muestran los centros de procesos de datos de GAIA para cada uno de los CU, así como su localización y responsable asignado.

APENDICE B

a. Sistemas de Coordenadas Astronómicas

En este trabajo, tienen gran importancia los sistemas de coordenadas, pues dentro del espacio de parámetros de los “Surveys” que se van a obtener, aparecen las coordenadas ecuatoriales y las galácticas de las estrellas sintéticas; ambos sistemas de coordenadas (junto con otros que no trataremos aquí) son ampliamente utilizados en astrofísica de posición para determinar la ubicación exacta de los objetos celestes. El problema que trata de resolver la astronomía de posición consiste fundamentalmente en determinar las posiciones que ocupan los astros en el firmamento, independientemente de las grandes distancias a las que se hallan.

Dada la pequeña longitud del radio terrestre en comparación con tales distancias, es asumible identificar la Tierra con un punto del espacio, con lo cual, las posiciones en que se encuentran los distintos observadores, que en realidad varían con su localización geográfica, pueden considerarse coincidentes. Dicho punto se toma como centro de la esfera celeste, sobre cuya superficie se sitúan las posiciones aparentes de los astros. De este modo, aunque la visión del observador sea incapaz de apreciar las distancias relativas que les separa de los astros, distingue los puntos de la intersección de la superficie esférica con las visuales dirigidas hacia los distintos cuerpos celestes y puede calcular su distancia angular, arco de círculo máximo comprendido entre dos astros.

Para situar sobre la esfera celeste estas posiciones aparentes es preciso elegir el sistema determinado de coordenadas esféricas:

- Un eje llamado eje fundamental o de referencia.
- Un círculo máximo, determinado por la intersección de la esfera celeste con un plano perpendicular al eje y que se denomina círculo fundamental.
- Dos coordenadas, una de las cuales se mide sobre el círculo fundamental y la otra sobre los semicírculos máximos que pasan por los puntos de intersección de la esfera celeste con el eje.
- Un sentido de medición de los arcos: directo cuando se efectúa en sentido contrario al recorrido por las agujas del reloj, y llama retrogrado, cuando ocurre lo contrario.

b. Coordenadas horizontales

Son aquellas que están referidas al horizonte del observador, tal y como se muestra en la figura 2. En este sistema, el origen de coordenadas es un

punto de la superficie terrestre, es decir, es un sistema topocéntrico cuyo eje fundamental es la vertical del lugar, línea que sigue la dirección de la plomada. Su punto de intersección de la esfera celeste situado encima del observador se llama **cenit**, y el opuesto, **nadir**. El círculo fundamental es el horizonte del lugar. Los círculos menores paralelos al mismo se denominan **almicantarates** y los semicírculos máximos que pasan por el cenit, el nadir y un astro determinado reciben el nombre de **círculos verticales**, o **vertical** del astro.

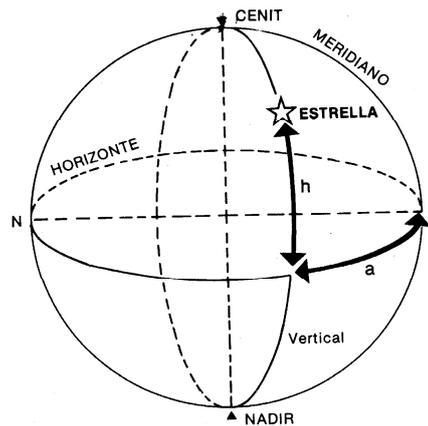


Figura 2: *Coordenadas Horizontales. Puede verse el cenit y el nadir, puntos opuestos de la esfera celeste, así como las coordenadas a (acimut) y h (altura).*

Las coordenadas horizontales son: la **altura** o **altitud** y el **acimut**. La altitud es la altura del astro sobre el horizonte (arco de semicírculo vertical comprendido entre el horizonte del lugar y el centro del astro); se mide de 0° a 90° a partir del horizonte, y tiene signo positivo para los astros situados por encima de este (y, por consiguiente, visibles) y negativos para los situados debajo del horizonte; se representa por la letra **h**.

Con frecuencia se emplea, en lugar de la altura, la distancia cenital, que es el arco semicírculo vertical comprendido entre el cenit y el centro del astro. Se representa por **Z** y se relaciona con la altura por la ecuación:

$$h = 90^\circ - Z$$

El acimut es el arco del horizonte medido en sentido retrógrado desde el punto Sur hasta la vertical del astro. Su valor va de 0° a 360° y se representa por la letra **A**. El acimut al igual que la altura nos la da la sombra del gnomon.

Los instrumentos utilizados para la determinación de las coordenadas horizontales son el teodolito (que permite determinar las dos coordenadas simultáneamente), y el sextante.

En el sistema de coordenadas horizontales, la altitud y el acimut de los astros cambia constantemente por efecto de la rotación terrestre y también según el horizonte del observador, por lo que habrá que tener presente que los valores obtenidos estarán en función del tiempo y del lugar.

c. Coordenadas ecuatoriales absolutas

Existen dos tipos de coordenadas ecuatoriales: las locales (a veces también llamadas horarias) y las ecuatoriales absolutas, estas últimas están entre las que se usan en este trabajo. Las coordenadas ecuatoriales absolutas son aquellas que están referidas al ecuador celeste; también se las llama coordenadas celestes o uranográficas. Este tipo de coordenadas surgió por los inconvenientes que presentan la utilización de coordenadas ecuatoriales locales (ya que estas últimas varían con el tiempo), y por la necesidad de obtener unas coordenadas que sean constantes.

El eje fundamental, así como el resto de los demás elementos que determinan este sistema de coordenadas, se muestran en la figura 3.

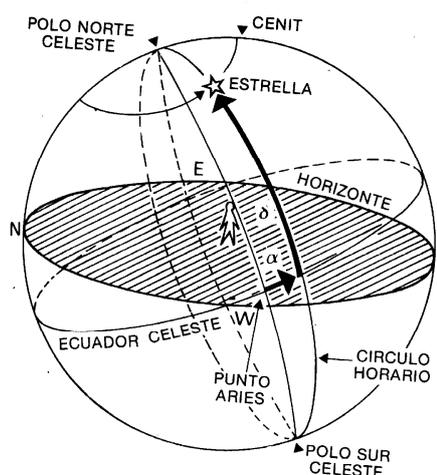


Figura 3: Coordenadas Ecuatoriales. Se muestran los círculos más representativos y las coordenadas (α , δ) de la estrella de interés.

Las coordenadas ecuatoriales absolutas son la **declinación** y la **ascensión recta**. La declinación es el arco del círculo horario comprendido entre el ecuador celeste y el centro del astro, medido de 0° a 90° a partir del ecuador; su valor es positivo cuando corresponde a un astro situado en el hemisferio boreal, y negativo en caso contrario; se representa por δ (**delta** en este trabajo).

La ascensión recta es el arco de ecuador celeste medio en sentido directo a partir del punto Aries (γ) (ver figuras 3 y 4) hasta el meridiano que

contiene al astro; varía de 0 horas a 24 horas y antiguamente se representaba por AR pero actualmente se representa con α (**alpha** en este trabajo). La ascensión recta está relacionada con el ángulo horario por la ecuación fundamental de la Astronomía de posición:

$$T = \alpha + H$$

Donde T es la hora sidérea (ver figura 4).

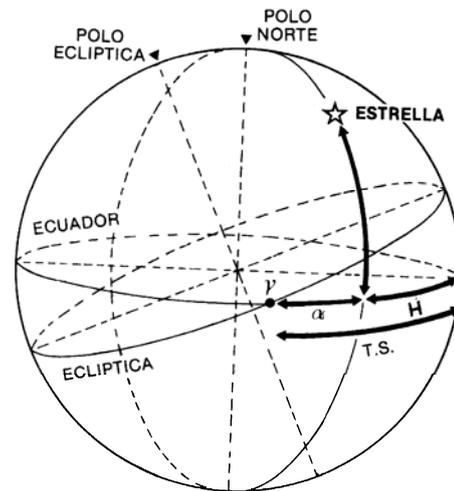


Figura 4: *Coordenadas Horizontales y hora sidérea T y punto γ . Se muestran los círculos y puntos más relevantes de estas coordenadas.*

Dado que el punto Aries (γ) es común para todos los observadores, las coordenadas ecuatoriales son universales y permanecen constantes en el tiempo.

Para determinar este tipo de coordenadas se emplea también, el anteojo meridiano, basándose en el cálculo de la hora de la culminación del astro, pues cuando este pasa por el meridiano del lugar su ángulo horario es nulo y la fórmula fundamental:

$$T = \alpha + H$$

Queda entonces reducida a: $T = \alpha$; es decir: la ascensión recta es igual a hora sidérea cuando culmina el astro.

d. **Coordenadas Galácticas**

Este sistema también será utilizado en este trabajo, junto con el sistema de coordenadas ecuatoriales absolutas descrito en b. El plano fundamental de este sistema de coordenadas es el **plano galáctico**, como se indica en la figura 5. Este plano galáctico se define con el ángulo diedro que forma

dicho plano con el ecuador celeste (que es de $I = 62.87^\circ$) y con la ascensión recta de uno de los dos puntos donde se interseca, $\alpha_l = 18h 51'$. Las dos coordenadas galácticas son:

Latitud galáctica (b): Es el ángulo que forma la visual del astro con el plano galáctico. Este valor varía entre 0° y 90° en dirección a P_g y entre 0° y -90° en dirección a P'_g (P_g es el Polo Norte Galáctico, y P'_g es el Polo Sur Galáctico).

Longitud galáctica (l): Es el ángulo diedro que forma el círculo máximo que pasa por P_g y el astro con el círculo máximo que pasa por P_g y el origen de longitud. Este origen se define como un punto situado a una distancia angular $\nu = 32.88^\circ$ del punto α_l recorrido en sentido del movimiento diurno aparente sobre el plano galáctico, y apunta hacia el centro de la galaxia. Este valor varía entre 0° y 360° contándose en el sentido contrario al movimiento diurno aparente.

Este sistema es totalmente universal y es el que se emplea generalmente para problemas de mecánica galáctica. En este trabajo, a las coordenadas (**l**, **b**) se las denomina (**lg**, **bg**), y forman parte del espacio de parámetros.

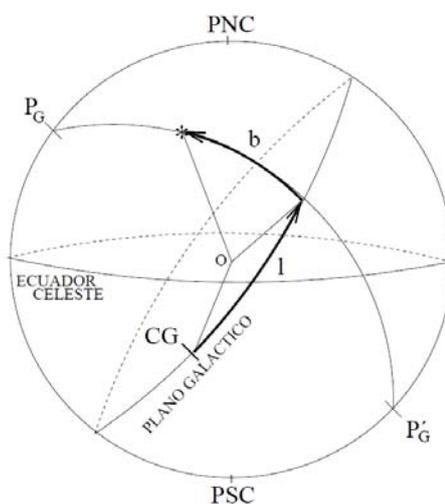


Figura 5: *Coordenadas Galácticas. Obsérvese el plano galáctico, el ecuador terrestre, el polo norte galáctico y las coordenadas l y b del punto de interés.*

e. Transformación de ecuatoriales a galácticas (y viceversa)

Dado que en este trabajo se utilizan las coordenadas ecuatoriales y las galácticas, y dado que se obtienen las segundas a partir de las primeras, se mostrará el paso de unas a otras en este apartado; para ello, es útil la figura 6.

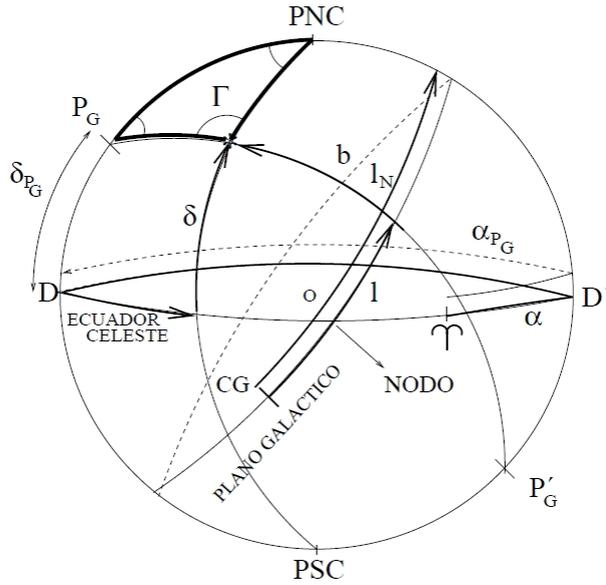


Figura 6: Transformación de *Coordenadas Galácticas a Ecuatoriales*

La figura 6 muestra la relación entre las coordenadas galácticas (l, b) y las ecuatoriales (α, δ). Denominando $\alpha_{P_G} = 12h 51.4'$ a la Ascensión Recta del Polo Norte Galáctico, y $\delta_{P_G} = 28^\circ 8'$ a la Declinación del Polo Norte Galáctico (situado en la Constelación de La Cabellera de Berenice), y l_n la longitud del Polo Norte Celeste (o PNC). Puesto que el origen de coordenadas de l está muy próximo al Centro galáctico (o CG), y este está a unos 33° del nodo, esto es: el punto donde el plano galáctico cruza de sur a norte el ecuador celeste, de la figura 6, donde está exactamente la mitad del nodo D-D':

$$l_n = 33 + 90 = 123.$$

Del triángulo esférico marcado en la figura 6 se tiene que:

Lado $90 - b$: Angulo $\alpha - \alpha_{P_G}$; lado $90 - \delta$: Angulo $l_n - l$; lado $90 - \delta_{P_G}$: Angulo Γ

Por tanto, aplicando el teorema del seno:

$$\frac{\text{sen}(\alpha - \alpha_{P_G})}{\text{sen}(90 - b)} = \frac{\text{sen}(l_n - l)}{\text{sen}(90 - \delta)}$$

Y como $\text{sen}(90 - x) = \cos(x)$, tendremos que:

$$\cos \delta \times \text{sen}(\alpha - \alpha_{P_G}) = \text{sen}(l_n - l) \times \cos b$$

Aplicando el teorema del coseno:

$$\cos(90 - \delta) = \cos(90 - \delta_{P_G}) \cos(90 - b) + \operatorname{sen}(90 - \delta_{P_G}) \operatorname{sen}(90 - b) \cos(l_n - l)$$

Es decir:

$$\operatorname{sen} \delta = \operatorname{sen} \delta_{P_G} \operatorname{sen} b + \cos \delta_{P_G} \cos b \cos(l_n - l)$$

Aplicando el teorema del coseno al otro lado:

$$\cos(90 - b) = \cos(90 - \delta_{P_G}) \cos(90 - \delta) + \operatorname{sen}(90 - \delta_{P_G}) \operatorname{sen}(90 - \delta) \cos(\alpha - \alpha_{P_G})$$

O también:

$$\operatorname{sen} b = \operatorname{sen} \delta_{P_G} \operatorname{sen} \delta + \cos \delta_{P_G} \cos \delta \cos(\alpha - \alpha_{P_G})$$

Por lo tanto, de la ecuación anterior tendremos que:

$$b = \operatorname{sen}^{-1}(\operatorname{sen} \delta_{P_G} \operatorname{sen} \delta + \cos \delta_{P_G} \cos \delta \cos(\alpha - \alpha_{P_G}))$$

Aplicando el teorema del seno por el coseno (el ángulo Γ no nos interesa en estos cálculos):

$$\begin{aligned} \cos(l_n - l) \operatorname{sen}(90 - b) &= -\cos(\alpha - \alpha_{P_G}) \operatorname{sen}(90 - \delta) \cos(90 - \delta_{P_G}) + \cos(90 - \delta) \operatorname{sen}(90 - \delta_{P_G}) \\ \cos(\alpha - \alpha_{P_G}) \operatorname{sen}(90 - \delta) &= -\cos(l_n - l) \operatorname{sen}(90 - b) \cos(90 - \delta_{P_G}) + \cos(90 - b) \operatorname{sen}(90 - \delta_{P_G}) \end{aligned}$$

Que también podemos expresar como:

$$\begin{aligned} \cos(l_n - l) \cos b &= \operatorname{sen} \delta \cos \delta_{P_G} - \cos \delta \operatorname{sen} \delta_{P_G} \cos(\alpha - \alpha_{P_G}) \\ \cos(\alpha - \alpha_{P_G}) \cos \delta &= \operatorname{sen} b \cos \delta_{P_G} - \cos b \operatorname{sen} \delta_{P_G} \cos(l_n - l) \end{aligned}$$

Operando las expresiones anteriores podemos obtener:

$$l = l_n - \operatorname{tg}^{-1} \left(\frac{\cos \delta \operatorname{sen}(\alpha - \alpha_{P_G})}{\operatorname{sen} \delta \cos \delta_{P_G} - \cos \delta \operatorname{sen} \delta_{P_G} \cos(\alpha - \alpha_{P_G})} \right) = l_n - \operatorname{tg}^{-1} \left(\frac{p}{q} \right)$$

Es importante tener en cuenta que si $p \times q < 0$ y $q < 0 \Rightarrow \alpha = \alpha + 180$; si $p \times q < 0$ y $q > 0 \Rightarrow \alpha = \alpha + 360$; y finalmente, si $p + q < 0$ entonces $\alpha = \alpha + 180$.

APENDICE C

a. Determinación de valores atípicos

Se definen como valores atípicos moderados a aquellos valores z de la muestra que cumplen cualquiera de las dos condiciones siguientes:

$$z < q_1 - \frac{3}{2} \times (q_3 - q_1)$$

$$z > q_3 + \frac{3}{2} \times (q_3 - q_1)$$

Se definen como valores atípicos severos a aquellos valores z de la muestra que cumplen cualquiera de las dos condiciones siguientes:

$$z < q_1 - 3.0 \times (q_3 - q_1)$$

$$z > q_3 + 3.0 \times (q_3 - q_1)$$

Siendo q_1 el primer cuartil, y q_3 el tercer cuartil. Los cuartiles resultan de dividir la distribución en cuatro partes iguales. Así, el primer cuartil q_1 deja por debajo de él una cuarta parte de observaciones y tres cuartas partes por encima, mientras que el tercer cuartil q_3 deja por debajo de este valor las tres cuartas partes de las observaciones y una cuarta parte de estas por encima. Los cuartiles se calculan de la siguiente forma:

- Se ordenan los datos de menor a mayor, es decir en orden creciente o ascendente.
- Para calcular el primer cuartil:

$$q_1 = \frac{1}{4}(n+1)$$

Si el dato cae entre dos enteros, se redondea hacia arriba, y ese resultado será el que ubique la posición de q_1 en la lista ordenada de datos.

- Para calcular el tercer cuartil:

$$q_3 = \frac{3}{4}(n+1)$$

Si el dato cae entre dos enteros, se redondea hacia arriba y ese resultado será el que ubique la posición de q_3 en la lista ordenada de datos.

- En 'R' los cuartiles se calculan con la función:

quantile(x, ...)

Solo se considerarán como valores atípicos en este trabajo, aquellos que sean severos según el criterio anterior.

APENDICE D

a. Teoría estadística de Cópulas

En este apéndice se presentan las bases teóricas necesarias para entender el funcionamiento de la distribución cópula bivalente. Aunque la cópula es una técnica multivalente, nos limitaremos en este apéndice al caso bivalente, al ser el que se utiliza en este trabajo.

Analizaremos pues la función de distribución conjunta asociada a un determinado fenómeno de la naturaleza, que afecta a un par de variables aleatorias cuya distribución individual es independiente de la distribución conjunta asociada al citado fenómeno. Para una información más general, se puede recurrir a [Embrechts-03] ó [Levi-01], si bien en cualquier tratado de estadística Multivalente pueden encontrarse referencias útiles.

Una cópula, C , es una función de distribución multivalente cuyas leyes marginales se distribuyen uniformemente entre $[0, 1]$. Para el objeto de nuestro trabajo bivalente, $C(u, v) = p[U \leq u, V \leq v]$ es una función definida $[0, 1]^2 \rightarrow [0, 1]$ que verifica las siguientes tres propiedades:

- $C(u, v)$ es una función creciente para cada una de sus componentes.
- $C(u, 1) = u$ y $C(1, v) = v$.
- $\forall a_1 < a_2, \forall b_1 < b_2, C(a_1, b_1) + C(a_2, b_2) - C(a_1, b_2) - C(a_2, b_1) \geq 0$

Teorema de Sklar: Sea una función de distribución bidimensional cuyas marginales son F_x y F_y . Entonces existe una cópula $C / \forall (x, y) \in [-\infty, \infty]^2$ tal que:

$$F(x, y) = C(F_x(x), F_y(y))$$

Si las distribuciones marginales son continuas, la cópula es única. Por tanto, a partir de las cópulas, es posible crear distribuciones bivariantes con distribuciones marginales definidas. De esta forma, si C es una cópula y F_x y F_y son dos distribuciones marginales, $C(F_x(x), F_y(y))$ es una distribución bivalente.

De la definición del *Teorema de Sklar*, se deduce que las funciones de distribución marginales univariantes pueden ser independientes de la de la cópula.

La aplicabilidad de este Teorema a nuestro trabajo, se fundamenta en la premisa de que la estructura del fenómeno de la naturaleza cuya ocurrencia afecta a varios parámetros a la vez (en nuestro caso en las LPV los parámetros K_s y $\log P$), es independiente de la función de distribución seguida por cada parámetro individualmente.

Función de Distribución Inversa: Si F es una función de distribución, entonces su función inversa generalizada, es toda función $F^{(-1)}$ definida en $[0, 1]$ tal que:

- Si $t \in \text{Im}(F)$ y $x \in [-\infty, \infty]$ entonces $F^{(-1)}(t) = x$ y $F(x) = t$. Por tanto, $\forall t \in \text{Im}(F), F(F^{(-1)}(t))=t$.
- Si $t \notin \text{Im}(F)$ entonces:

$$F^{(-1)}(t) = \inf \{x / F(x) \geq t\} = \sup \{x / F(x) \leq t\}$$
- Si F es estrictamente creciente, tiene una única función inversa $F^{(-1)}$.

Corolario del Teorema de Sklar: Se definen F, C, F_x y F_y como en los enunciados anteriores, y F_x^{-1}, F_y^{-1} como las respectivas funciones inversas generalizadas de F_x, F_y ; entonces $\forall (u, v) \in [0, 1]^2$:

$$C(u, v) = F(F_x^{-1}(x), F_y^{-1}(y))$$

Densidad de una Cópula: Se sabe que, si existe, la densidad *pdf* de una función de distribución F , se define como:

$$f(x, y) = \frac{\partial F(x, y)}{\partial x \partial y}$$

La función de densidad de una cópula, simbolizada por c es:

$$c(u, v) = \frac{\partial C(u, v)}{\partial u \partial v}$$

A partir de $c(u, v)$, la densidad f de la función de distribución F puede obtenerse como:

$$f(x, y) = c(F_x(x), F_y(y))f_x(x)f_y(y)$$

Distribución Condicionada de una Cópula: Sea $C_1(u, v)$ la derivada de $C(u, v)$ respecto de u :

$$\frac{\partial C(u, v)}{\partial u} = C_1(u, v)$$

Si la distribución conjunta de X e Y es $F(x, y) = C(F_x(x), F_y(y))$, entonces la distribución condicionada $Y/X = x$ es:

$$F_{Y/X=x}(y) = C_1(F_x(x), F_y(y))$$

Cópula de Supervivencia: Sea $S(x) = P(X > x)$. La función de supervivencia conjunta $S(x, y) = P(X > x, Y > y)$ no es $1 - F(x, y)$ como podría pensarse, sino:

$$S(x, y) = 1 - F_x(x) - F_y(y) + F(x, y)$$

Por tanto, en el caso de una cópula, donde sabemos que $C(u, v) = P[U \leq u, V \leq v]$, la función de supervivencia será:

$$C_s(u, v) = p[U > u, V > v] = 1 - u - v + C(u, v)$$

Obteniendo por tanto que: $S(x, y) = C_s(F_x(x), F_y(y))$.

Concordancia y Discordancia: Sean ahora (x, y) y (x', y') dos observaciones aleatorias de un vector (X, Y) , entonces (x, y) e (x', y') son concordantes si $(x - x')(y - y') > 0$ y discordantes en caso contrario.

Coefficiente de correlación No Lineal de Kendall: Una de las principales ventajas de las cópulas, es que permiten analizar coeficientes de correlación no lineales. Habitualmente se ha utilizado como coeficiente de correlación no lineal, el denominado coeficiente de correlación de Kendall, que muestra una relación directa con el parámetro a que aparece en la función de distribución multivariante cópula.

Sea (X, Y) un par de variables aleatorias. La medida de concordancia τ de Kendall se define como la diferencia entre las probabilidades de concordancia y discordancia de dos observaciones distintas, (x, y) y (x', y') , de ese par aleatorio:

$$\tau(x, y) = p[(x - x')(y - y') > 0] - p[(x - x')(y - y') < 0]$$

Y que cumple las siguientes propiedades:

- $-1 \leq \tau \leq 1$

- Si X e Y son concordantes, entonces $\tau = 1$
- Si X e Y no son concordantes, entonces $\tau = -1$
- Si X e Y son independientes, entonces $\tau = 0$. La implicación inversa, no tiene porque ser cierta en general; es decir: si $\tau = 0$, las variables X e Y no tienen por qué ser independientes.
- Si f y g son dos funciones estrictamente crecientes, entonces $\tau[f(X), g(Y)] = \tau(X, Y)$; es decir: el coeficiente de correlación de Kendall es invariante ante una transformación estrictamente creciente de las variables aleatorias, ya que sólo depende del rango de cada observación.
- τ no depende de la cópula de (X, Y) .

A continuación se van a presentar las cópulas utilizadas en este trabajo: la cópula normal (o Gaussiana), y la cópula t-Student, pertenecientes a la familia de las cópulas elípticas. Las cópulas elípticas son las que se encuentran asociadas a variables aleatorias cuya función de distribución multivariante sea simétrica, por lo que las curvas de nivel de las variables aleatorias que tengan este tipo de cópulas forman elipses.

Cópula Normal (Gaussiana): La función Gaussiana bivalente que pertenece a la cópula Gaussiana tiene la forma:

$$C_{\rho}(u, v) = \Phi_2(\Phi^{-1}(u), \Phi^{-1}(v))$$

Donde $(u, v) \in [0, 1]^2$, $\Phi_2(\dots, \dots)$ es la función de distribución conjunta de una Gaussiana bivalente con media $(0, 0)^T$ y matriz de covarianza $\mathbf{R}_{(2 \times 2)}$; $\Phi^{-1}(\cdot)$ es la función inversa de la distribución acumulada Gaussiana estándar y ϕ es la función de densidad de la distribución normal. La función de densidad de la cópula es:

$$c_{\rho}(u, v) = \frac{\phi_2(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u)) \times \phi(\Phi^{-1}(v))}$$

El uso de la cópula Gaussiana bivalente es interesante ya que codifica la dependencia en la misma forma en que la distribución normal bivalente lo hace usando el parámetro de dependencia ρ , con la diferencia de que se calcula para variables aleatorias con cualesquiera marginales arbitrarias.

Cópula t-Student: Sea X un vector aleatorio con distribución t de Student sesgada $X \sim st_D(v, \mu, \Sigma, \gamma)$:

$$C_{v,R}^t(u, v) = t_{v,R}^n(t_v^{-1}(u), t_v^{-1}(v))$$

Donde $\mathbf{R}_{(2 \times 2)}$ es la matriz de covarianza, $t_{v,R}^n(\cdot)$ es la función de distribución bivalente y $t_v^{-1}(\cdot)$ es la función inversa asociada a la distribución t univariante con v grados de libertad.

Se puede apreciar que las ecuaciones anteriores se encuentran definidas en términos de las funciones de distribución inversas de ϕ^{-1} y t_v^{-1} , estas se evalúan en las variables $u_i = F_i(u_i)$ con $i = 1, 2$ donde F_i corresponde a una función de distribución arbitraria, de aquí que las variables aleatorias cuya estructura de dependencia este descrita por una copula Gaussiana o una t de Student puede describir comportamientos marginales diferentes. Las distribuciones derivadas de esta manera son conocidas como meta-distribuciones y pueden ser pensadas como una generalización de las distribuciones multivariantes tradicionales. Por ejemplo, una distribución meta-Gaussiana sería aquella cuyo comportamiento multivariante es descrito por una copula Gaussiana, pero al menos una de las distribuciones marginales sea diferente de la distribución normal.

En este trabajo se han usado como funciones marginales la distribución normal o Gaussiana, y la distribución Weibull. Para más detalles, ver código del fichero “*functions-Long_Period.R*”

ANEXO E

a. Código R

Este anexo contiene TODO el código y ficheros de datos y resultados mencionados en el punto 4.2 en CD (adjunto).

Abreviaturas y Acrónimos

BBDD	Base de Datos
BE	Bias Estimation
BP	Blue Photometer
CCD	Charge Coupled Device
CD	Compact Disc
CU	Coordination Unit
DPAC	Data Process and Analysis Consortium
DPACE	Data Process and Analysis Consortium executive
ESA	European Space Agency
FAR	False Alarm Rate
HR	Hertzprung-Rusell
IA	Inteligencia Artificial
IC	Ingeniería del Conocimiento
IS	Ingeniería del Software
LMC	Large Magellanic Cloud
LPV	Long Period Variables
OGLE	Orbital Gravitational Lensing Experiment
OO	Orientado a Objetos
PFC	Proyecto Fin de Carrera
PLC	Periodo Magnitud Color
QA	Quality Assessment
RP	Red Photometer
RVS	Radial Velocity Spectrometer
SBC	Sistemas Basados en el Conocimiento
SE	Sistema Experto

SMC	Small Magellanic Cloud
SNR	Signal to Noise Ratio
SOM	Self Organized Maps
TF	Transfer Function
UML	Unified Modeling Language
UNED	Universidad Nacional de Educación a Distancia
WP	Working Process

Bibliografía

[Alcock-97]	The RR Lyrae Population of the Galactic Bulge from the MACHO Database: Mean Colors & Magnitudes. C. Alcock, R. A. Allsman, D. R. Alves, T. S. Alxelrod, A. C. Becker, A. Basu, L. Baskett, D. P. Bennett, K. H. Cook, K. C. Freeman, K. Griest, J. A. Gueru, M. J. Lehner, S. L. Marshall, D. Minniti, B. A. Peterson, M. R. Pratt, A. W. Rodgers, C. W. Stubbs, W. Sutherland, T. Vandehei and D. L. Welch. Submitted to The Astrophysical Journal. arXiv:astro-ph/9707311v1 28 Jul 1.997
[Amores-05]	Models for Interstellar Extinction in the Galaxy. E. B. Amores and J. R. Lepine The astronomical Journal, 130:659-673, 2.005 August
[Antonello-02]	Variable stars in nearby Galaxies. VI Frecuency-Period distribution of Cepheids in IC 1613 and other galaxies of the local group. E. Antonello, D. Fugazza and L. Mantegazza. Astronomy & Astrophysics 29 March 2.002
[Atienza-10]	Optimización de Estimadores de Funciones de Densidad de Probabilidad e implementación de Test Estadísticos Multivariantes para la Misión Espacial GAIA Enrique Atiénzar, Madrid 2.010 (UNED)
[Borissova-09]	Properties of the RR Lyrae stars in the inner regions of the Large Magellanic Cloud. III. Near-Infrared study. J. Borissova, M. Rejkuba, D. Minniti, M. Catelan and V. D. Ivanov. Astronomy & Astrophysics 502, 505-514 (2.009)
[Brams-89]	Teoría de Juegos y Seguridad Nacional. Steven J. Brams y D. Marc Kilgour. Ministerio de Defensa. Gobierno de España, 1.989.
[Bruijne-09]	GAIA Astrometric Performance: Summer-2.009 status. Jos de Bruijne. GAIA-CA-TN-ESA-JDB-055-01.1 rev 01. 8 dec 2.009.
[Catelan-04]	The RR Lyrae Period-Luminositi Relation. I. Theoretical Calibration. M. catelan, Barton J. Pritzl and Horace A Smith. ApJ Supplement Series, in Press. arXiv:astro-ph/0406067v1 2 Jun 2.004
[Collinge-06]	Catalog of Fundamental-Mode RR Lyrae Stars in the Galactic Bulge from the OGLE. Matthew J. Collinge, Takahiro Sumi and Daniel Frabryky. The Astronomical Journal, 651:197-210 2.006 November 1.
[Deb-09]	Physical parameters of the Small Magellanic Cloud. RR Lyrae and the distance sacale. Sukanta Deb, Harinder P. Singh. arXiv:0910.4806v1 [astro-ph.SR] 26 Oct 2.009
[Embrechts-03]	Using Copulae to bound the Value-at-risk for functions of dependent risk. P. Embrechts, P. Hoeing, A. Juri. Finance & Stocastics 7(2), 145-167, 2.003
[Garcia-08]	Desarrollo de Software sw Análisis estadístico de "surveys" de variabilidad para la Misión espacial GAIA Fernando García Sedano, Madrid 2.008 (UNED)
[Groenewegen-04]	Long Period Variables in the Magellanic Clouds: OGLE + 2 MASS + DENIS M. A. T. Groenewegen. Astronomy & Astrophysics 425, 595--613 (2.004)
[Ita-03]	Variable stars in the Magellanic Clouds: Results from OGLE and SIRIUS. Yoshifusa Ita, Toshihiko Tanabe, Noriyuki Matsunaga, Yasushi Nakagima, Chie Nagashima, Takahiro Nagayama, Daisuke Kato, Mikio Kurita, Tetsuya Nagata, Shuji Sato, Motohide Tamura, Hidehiko, Nakaya and oshikazu Nakada. R. Astron. Soc. 000, 1-9 (2.002)
[Jonhston-97]	An Expert System Approach to Astronomical Data Analysis. Mark D. Jonhston. Space Telescope Science Institute, 3.700 San Martin Drive. Baltimore, MD 21218
[Jordi-06]	Gaia photometers: recommended design. Jordi, C., Hog, E., Brown, A., deBruijne, J. GAIA-CH-TN-UB-CJ-038-1. April, 26 2.006.
[Jordi-10]	GAIA broad band photometry. C. Jordi, M. Gebran, J.M. Carrasco, J. de Bruijne, H. Voss, C. Fabricius, J. Knude, A. Vallenari, A. Mora arXiv:10080815v2 [astrp-ph.IM] 14 Oct. 2.010

[Lafler-64]	An RR Lyrae Star Survey with the Iick 20-inch Astrograph. II. The Calculation of RR Lyrae periods by Electronic Computer. J. Lafler and T. D. Kinman. American Astronomical Society, July 1.964
[Levi-01]	Etude des dependences entre les branches Auto et Incendie avec la theorie des copulas. Charles M. Levi. Guy Carpenter - Universite Louis Pasterur. Strasbourg, 2.001
[Musen-93]	An Overview of Knowledges Adquisition. M. A. Musen Editors: J. M. David, J. P. Krivine and R. Simmons. Second Generation of Expert Systems, pp. 405-427. Springer Verlag, 1.993
[Oudmaijer-98]	The Lutz-kelker bias in trigonometric parallaxes. Rene D. Uodmaijer, Martin A. T. Gronewegen and Hans Schrijver. R. Astron. Soc. 000, 000-000, Feb 2.008. arXiv:astro-ph/9801093v1 12 Jan 1.998.
[Palma-00]	Ingenieria del Conocimiento. De la Extracción al Modelado de Conocimiento. J. T. Palma, E. Paniagua, F. Martín, R. Marin. Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, Valencia (España) pp. 46-72
[Sanchez-05]	Clasificación Automatizada de Estrellas variables: Algoritmos para la determinación de periodos y aplicación a datos simulados. Néstor Sanchez, Raul Velasquez, Sandra Ocando, Maoly Carrasquero, Noreida Guerrero and Mailyn Uribe. CIENCIA 13(1), 44-55, 2.005.
[Sandage-04]	New Period-Luminosity and Period-Color relations of classical Cepheids- II. Cepheids in LMC. A. Sandage, G. A. Tammann, and B. Reindl. arXiv:astro-ph/0402424v2 6 Jul 2.004
[Sandage-08]	New Period-Luminosity and Period-Color relations of classical Cepheids- III. Cepheids in SMC. A. Sandage, G. A. Tammann, and B. Reindl. Astronomy & Astrophysics, Oct 2.008. arXiv:0810.1780v1 [astro-ph] 9 Oct 2.008.
[Schneider-06]	Extragalactic Astronomy and Cosmology. An introduction. Prof. Dr. Peter Schneider Springer Berlin Heidelberg NY. ISBN-13978-3-540-33174-2
[Schreiber-99]	Knowledge Engineering and Management. The CommonKADS Methodology. G. Schreiber, H. Akkermans, A. Anjewierden, R. de Hoog, N. Shadbot, W. Van de Velde, B. Wielinga. MIT Press. Cambridge, Massachusetts, 1.999 ISBN 0-262-19300-0.
[Soszynski-09]	The optical Gravitational Lensing Experiment. The OGLE III Catalog of variable stars. III. RR Lyrae stars in the Large magellanic Cloud. I. Soszynski, A. Udalski, M. K. Szymanski, M. Kubiak, G Pietrzynski, L. Wzyzkowski, O. Szcwcyk, K. Ulaczyk, and R. Polaski. Acta Astronomica, Vol 59 (2.009) pp. 1-18 arXiv:0903.2482v2 [astro-ph.SR] 3 Apr 2.009
[Soszynski-09]	The optical Gravitational Lensing Experiment. The OGLE III Catalog of variable stars. IV. Long Period Variables in the Large Magellanic Cloud. I. Soszynski, A. Udalski, M. K. Szymanski, M. Kubiak, G Pietrzynski, L. Wzyzkowski, O. Szcwcyk, K. Ulaczyk, and R. Polaski. Acta Astronomica, Vol 59 (2.009) pp. 239-253 arXiv:0910.1354v1 [astro-ph.SR] 7 Oct 2.009
[Soszynski-10]	Properties of the RR Lyrae stars in the inner regions of the Small Magellanic Cloud. The OGLE Catalog of variable stars. IX. RR Lyrae stars in the Small Magellanic Cloud. I. Soszynski, A. Udalski, M. K. Szymanski, M. Kubiak, G. Pietrynski, L. Wzyzkowski, K. Ulaczyk and R. Poleski. Acta Astronomica, Vol 60 (2.010) pp. 0-0.
[Subramanian-11]	The 3D structure of the Small Magellanic Cloud. Smitha Subramanian, Annapurni Subramanian. arXiv: 1109.3980v1 [astro-ph.CO] 19 Sept 2.011
[Tammann-03]	New Period-Luminosity and Period-Color relations of classical Cepheids- I. Cepheids in the Galaxy. G. A. Tammann, A. Sandage, and B. Reindl. Astronomy & Astrophysics, March 2.003. arXiv:astro-ph/0303378v1 17 Mar 2.003.
[Valenti-04]	The Red Giant Branch in Near-Infrared Colour-Magnitude Diagrams. I: The calibration of photometric indices. E. Valenti, F. R. Ferraro, L. Origlia. R. Astron. Soc. 000, 1-13 (2.003). arXiv:astro-ph/0403563v1 23 Mar 2.004

[Winberg-00]	Structure of the Large Magellanic Cloud from 2MASS. Martin D. Winberg, Sergei Nikolaev. Department of Physics & Astronomy, University of Massachusetts, Amherst MA 01003-4525 The Astrophysical Journal, 548: 712-726 2.001 February 20.
[Yoachim-06]	Structural Parameters of Thin and Thick Disks in Edge-on Disk Galaxies. Peter Yoachim and Julianne J. Dalcanton. Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195. The Astronomical Journal, 131:226-249 2.006 January