



VICERRECTORÍA DE INVESTIGACIÓN Y POSTGRADO
DIRECCIÓN DE POSTGRADOS Y POSTÍTULOS

FACULTAD DE CIENCIAS
DEPARTAMENTO DE ASTRONOMÍA

**HERRAMIENTAS DE APRENDIZAJE
AUTOMATIZADO PARA LA CLASIFICACIÓN DE
GALAXIAS EN EL CATÁLOGO VVV NIRGC**

Tesis presentada para optar al Grado Académico de Magíster en Astronomía.
--

AUTOR: PAMELA IVONNE MARCHANT CORTÉS

LA SERENA, CHILE, 2 de Julio de 2024

CONSTANCIA

Don

HACE CONSTAR:

Que el trabajo correspondiente a la presente Tesis de Magíster, titulada “TESIS NAME”, ha sido realizada por Doña Pamela Ivonne Marchant Cortés, bajo mi dirección.

Para que conste y en cumplimiento de las normativas vigentes de la Universidad de la Serena, Chile, firmo el presente documento en La Serena, Chile, Julio de 2024

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN ASTRONOMÍA

TÍTULO : HERRAMIENTAS DE APRENDIZAJE AUTOMATIZADO
PARA LA CLASIFICACIÓN DE GALAXIAS EN EL VVV
NIRGC

PRESENTADA POR : PAMELA IVONNE MARCHANT CORTÉS

DIRECTOR DE TESIS : JOSÉ LUIS NILO CASTELLÓN

TRIBUNAL CALIFICADOR

El tribunal de tesis, conformado por:

PRESIDENTE :

MIEMBROS DEL TRIBUNAL: *

*

*

ACUERDAN OTORGARLE LA CALIFICACIÓN DE:

t

La Serena, Chile, **Julio de 2023**

“If you stop to think about it, you’ll have to admit that all the stories in the world consist essentially of twenty-six letters. The letters are always the same, only the arrangement varies. From letters words are formed, from words sentences, from sentences chapters, and from chapters stories.”

“Every real story is a never ending story.”

Michael Ende - The Neverending Story

AGRADECIMIENTOS

Gracias a todas las personas que me han acompañado en este tiempo. Gracias a los fisilocos por los buenos momentos y el apoyo de siempre, por ser grandes amigos y hacer de mi vida universitaria un mejor lugar. Por aquellas largas noches conectados, escuchando mis párrafos y dándome las energías para terminar la tesis.

Gracias a la gente de Córdoba y de la colaboración, por enseñarme como investigar, or esas risas en las reuniones.. A Laura por motivarme a seguir hacia adelante, a Vicky por todos lo correos y las largas horas durante vacaciones para terminar el paper. Gracias a Nilo por el apoyo, enseñanzas y largas discusiones.

Gracias a la gente de Astro. Gracias a la profe Amelia, por escucharme y motivarme, por darme grandes ideas.

Gracias a Felipe, por demostrarme que soy mi propia enemiga y ayudar a superar los obstáculos y creer en mis habilidades.

Gracias a mi familia por siempre creer en mi, es especial a mi mamá, por ser todo en mi vida y estar ahí en cada momento, por siempre motivarme e impulsarme a cumplir mi sueño.

Gracias a Hugo por apoyarme, darme fuerzas y animo en los momentos más difíciles en este duro trayecto.

Gracias y no gracias a mis gatos, que con un simple maullido lograban devolverme las energías.

Resumen

La clasificación es una tarea común en astronomía, y con la creciente disponibilidad de datos, el uso de métodos de aprendizaje automático está en aumento. Sin embargo, la presencia del disco galáctico plantea desafíos significativos para esta tarea. Esta región de baja latitud, conocida como Zona de Evitamiento (ZoA), se caracteriza por su alta extinción interestelar, la densa concentración de estrellas y la limitada disponibilidad de datos extragalácticos.

Aunque el uso de herramientas de aprendizaje automático en la ZoA es limitado, algunos estudios han abordado este desafío. Por ejemplo, Daza-Perilla et al. (2023) aplicó algoritmos en el disco del relevamiento VVV y VVVX, mientras que Zhang et al. (2021) utilizó algoritmos de clasificación automática en el catálogo de 4XMM-DR9, que incluye la ZoA. Sin embargo, al comparar los resultados de este último estudio con la clasificación realizada en el catálogo VVV NIRGC, se observó una discrepancia significativa, con muchas fuentes clasificadas como galaxias cuando en realidad son estrellas o fuentes extendidas galácticas en la región del disco del VVV.

Nuestro trabajo además, investiga las implicaciones del uso de herramientas automáticas, evaluando su confiabilidad y cómo afectan al resultado final. Exploramos la hipótesis de que el área a analizar debería tener condiciones similares a las del conjunto de entrenamiento. En el caso de la ZoA, sus características extremas la diferencian del resto del cielo, lo que dificulta la aplicación directa de modelos de clasificación entrenados en otras regiones. Esta distinción es crucial para la correcta interpretación de la distribución galáctica en el Universo local y el estudio de la Estructura a Gran Escala.

Nuestros resultados subrayan las dificultades en el uso de métodos automáticos para

la clasificación de galaxias en la ZoA y resaltan la importancia de considerar los factores específicos de esta región para mejorar la fiabilidad y precisión de los estudios futuros.

Summary

Classification is a common task in astronomy, and with the increasing availability of data, the use of machine learning methods is on the rise. However, the presence of the galactic disk poses significant challenges for this task. This low-latitude region, known as the Zone of Avoidance (ZoA), is characterized by high interstellar extinction, dense star concentration, and limited availability of extragalactic data.

Although the use of machine learning tools in the ZoA is limited, some studies have addressed this challenge. For example, Daza-Perilla et al. (2023) applied algorithms in the VVV and VVVX disk survey, while Zhang et al. (2021) used automatic classification algorithms in the 4XMM-DR9 catalog, which includes the ZoA. However, when comparing the results of the latter study with the classification performed in the VVV NIRGC catalog, a significant discrepancy was observed, with many sources classified as galaxies when they are actually stars or extended galactic sources in the VVV disk region.

Our work also investigates the implications of using automated tools, evaluating their reliability and how they affect the final result. We explore the hypothesis that the area under analysis should have conditions similar to those of the training set. In the case of the ZoA, its extreme characteristics differentiate it from the rest of the sky, making it difficult to directly apply classification models trained in other regions. This distinction is crucial for the accurate interpretation of the galactic distribution in the local Universe and the study of Large-Scale Structure.

Our results underscore the challenges in using automated methods for galaxy classification in the ZoA and highlight the importance of considering the specific factors of this region to improve the reliability and accuracy of future studies.

Publicaciones científicas

Los resultados de mi tesis han sido publicados en revistas de alto impacto durante los años 2023 y 2024.

En el primer artículo, del cual soy autora principal (Marchant Cortés et al. 2024), presentamos los hallazgos más significativos derivados de mi investigación, en el cual estudiamos y reportamos errores en la clasificación de galaxias usando herramientas de aprendizaje automático.

En el segundo artículo, donde figuro como coautora (Daza-Perilla et al. 2023), publicamos los resultados relativos a la clasificación de galaxias usando técnicas de aprendizaje automático en la Zona de Evitamiento, atendiendo a la importancia del set de entrenamiento.

Galaxies in the zone of avoidance: Misclassifications using machine learning tools

P. Marchant Cortés¹, J. L. Nilo Castellón¹, M. V. Alonso^{2,3}, L. Baravalle^{2,3}, C. Villalon², M. A. Sgró^{2,3},
I. V. Daza-Perilla^{2,3}, M. Soto⁴, F. Milla Castro¹, D. Minniti^{5,6,7}, N. Masetti^{8,5}, C. Valotto^{2,3}, and M. Lares^{2,3}

¹ Departamento Astronomía, Facultad de Ciencias, Universidad de La Serena, Av. Juan Cisternas 1200, La Serena, Chile
e-mail: p.marchantcortes.9@gmail.com

² Instituto de Astronomía Teórica y Experimental (IATE-CONICET), Laprida 854, X5000BGR Córdoba, Argentina

³ Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, X5000BGR Córdoba, Argentina

⁴ Instituto de Investigación en Astronomía y Ciencias Planetarias, Universidad de Atacama, Copayapu 485, Copiapó, Chile

⁵ Instituto de Astrofísica, Facultad de Ciencias Exactas, Universidad Andrés Bello, Av. Fernández Concha 700, Las Condes, Santiago, Chile

⁶ Vatican Observatory, 00120 Vatican City State, Italy

⁷ Departamento de Física, Universidade Federal de Santa Catarina, Trindade 88040-900, Florianópolis, Brazil

⁸ INAF – Osservatorio di Astrofisica e Scienza dello Spazio, Via Piero Gobetti 101, 40129 Bologna, Italy

Received 16 November 2023 / Accepted 28 February 2024

ABSTRACT

Context. Automated methods for classifying extragalactic objects in large surveys offer significant advantages compared to manual approaches in terms of efficiency and consistency. However, the existence of the Galactic disk raises additional concerns. These regions are known for high levels of interstellar extinction, star crowding, and limited data sets and studies.

Aims. In this study, we explore the identification and classification of galaxies in the zone of avoidance (ZoA). In particular, we compare our results in the near-infrared (NIR) with X-ray data.

Methods. We analyzed the appearance of objects in the Galactic disk classified as galaxies using a published machine-learning (ML) algorithm and make a comparison with the visually confirmed galaxies from the VVV NIRGC catalog.

Results. Our analysis, which includes the visual inspection of all sources cataloged as galaxies throughout the Galactic disk using ML techniques reveals significant differences. Only four galaxies were found in both the NIR and X-ray data sets. Several specific regions of interest within the ZoA exhibit a high probability of being galaxies in X-ray data but closely resemble extended Galactic objects. Our results indicate the difficulty in using ML methods for galaxy classification in the ZoA, which is mainly due to the scarcity of information on galaxies behind the Galactic plane in the training set. They also highlight the importance of considering specific factors that are present to improve the reliability and accuracy of future studies in this challenging region.

Key words. catalogs – surveys – infrared: galaxies – X-rays: galaxies

1. Introduction

Astronomy is moving forward as never before, with much of the progress being driven by the unprecedented amount of data produced by large surveys. New tools have begun to play an essential role in analyzing data, such as machine-learning (ML) algorithms, which have increased the efficiency with which we can identify commonalities across large databases and detect faint and complex patterns. These algorithms have become a common tool in astronomy because of the large amount of data coming from all-sky surveys, and have been well tested as classifiers for galaxy morphology (Spindler et al. 2021), young stellar object (YSO) finders (Martín et al. 2019), classifiers of variable stars using light curves (Aguirre et al. 2019), and estimators of photometric redshift (Dainotti et al. 2021), among many others.

All-sky surveys are also increasing in number at different wavelengths. This improves our understanding of the large-scale evolution and structure of the Universe, the formation of stars and galaxies, and the history of the Milky Way (MW). At lower Galactic latitudes, the data from these surveys are scarce. This region is known as the zone of avoidance (ZoA; Shapley 1961), and became more prominent as complete sky surveys increased

in number. It is more critical at $|b| < 10^\circ$, where extragalactic sources and large-scale structure (LSS) behind the MW are obscured by dust and stellar crowding, dimming the sources by more than 25% in the optical and by about 10% in the infrared (IR) wavelengths (Henning et al. 1998). Classifying extragalactic objects within the ZoA is of critical importance. We have the opportunity to minimize this gap and explore the Local Universe in increasing detail. This endeavor is pivotal in defining the cosmography of the nearby Universe, which sheds light on the dynamics of the Local Group, giving us insights into the Universe at larger scales. It will also allow us to decipher various cosmological parameters, including the peculiar velocity of the Local Group, which exhibits a profound discrepancy relative to the cosmic microwave background dipole as seen in Loeb & Narayan (2008). In order to investigate the LSS, it is crucial to obtain a more complete redshift catalog to fill the ZoA gap in the distribution of the largest mass concentrations in the Local Universe, such as the Great Attractor (Kraan-Korteweg et al. 1996), the Perseus-Pisces Supercluster (Ramatsoku et al. 2016), and the Vela Supercluster (Kraan-Korteweg et al. 2017).

The use of near-infrared (NIR) wavelengths together with radio and X-rays has led to a new wave of extragalactic studies



The VVV near-IR galaxy catalogue in a Northern part of the Galactic disc

I. V. Daza-Perilla^{1,2}★, M. A. Sgró^{1,3}★, L. D. Baravalle^{1,3}★, M. V. Alonso^{1,3}, C. Villalon¹,
M. Lares^{1,3}, M. Soto⁴, J. L. Nilo Castellón^{5,6}, C. Valotto^{1,3}, P. Marchant Cortés^{5,6}, D. Minniti^{7,8,9} and
M. Hempel^{7,10}

¹Instituto de Astronomía Teórica y Experimental, CONICET-UNC, Córdoba X5000BGR, Argentina

²Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba (UNC), Córdoba CP:X5000HUA, Argentina

³Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, Córdoba X5000BGR, Argentina

⁴Instituto Multidisciplinario en Investigación y Postgrado (IMIP), Universidad de La Serena., Av. Raúl Bitrán Nachary No 1305, La Serena, 1720236, Chile

⁵Departamento de Astronomía, Universidad de La Serena. Av. Juan Cisternas 1200, La Serena, 1720236, Chile

⁶Instituto de Astronomía y Ciencias Planetarias, Universidad de Atacama, Copayapu 485, Copiapó, 1532297, Universidad de La Serena, Av. Juan Cisternas 1200, La Serena 1720236, Chile

⁷Instituto de astrofísica, Facultad de Ciencias Exactas, Universidad Andrés Bello, Av. Fernández Concha 700, Los Condes, Santiago 7550000, Chile

⁸Vatican Observatory, V00120 Vatican City State, Italy

⁹Departamento de Física, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina 88040 970, Brazil

¹⁰Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany

Accepted 2023 June 8. Received 2023 May 20; in original form 2022 November 9

ABSTRACT

The automated identification of extragalactic objects in large surveys provides reliable and reproducible samples of galaxies in less time than procedures involving human interaction. However, regions near the Galactic disc are more challenging due to the dust extinction. We present the methodology for the automatic classification of galaxies and non-galaxies at low Galactic latitude regions using both images and photometric and morphological near-IR data from the VISTA Variables in the Vía Láctea eXtended (VVVX) survey. Using the VVV NIR Galaxy Catalogue (VVV NIRGC), we analyse by statistical methods the most relevant features for galaxy identification. This catalogue was used to train a convolutional neural network with image data and an XGBoost model with both photometric and morphological data and then to generate a data set of extragalactic candidates. This allows us to derive probability catalogues used to analyse the completeness and purity as a function of the configuration parameters and to explore the best combinations of the models. As a test case, we apply this methodology to the Northern disc region of the VVVX survey, obtaining 172 396 extragalactic candidates with probabilities of being galaxies. We analyse the performance of our methodology in the VVV disc, reaching an F1-score of 0.67, a 65 per cent purity, and a 69 per cent completeness. We present the VVV NIRGC: Northern part of the Galactic disc comprising 1003 new galaxies, with probabilities greater than 0.6 for either model, with visual inspection and with only two previously identified galaxies. In the future, we intend to apply this methodology to other areas of the VVVX survey.

Key words: methods: data analysis – methods: statistical – galaxy: general.

1 INTRODUCTION

The use of statistical tools capable of automatically generating models, deriving catalogues and determining the statistical description of the physical quantities of stellar objects has been favoured by the large amounts of available data. One of these tools is the availability of mathematical models that can automatically learn from data and that, albeit more complex than the classical methods, provide the opportunity to solve various tasks using the large amount of data from surveys (see, e.g. Baron 2019).

The types of problems that can be solved with machine learning approaches include, but are not limited to, solutions via unsupervised models such as the detection of anomalies in the Sloan Digital

Sky Survey (SDSS) quasar spectra (Boroson & Lauer 2010), the dimensionality reduction of infrared spectra for the determination of some physical characteristics of stars (Reis et al. 2018), and the visual inspection of data through the embedding of features into another variables space (Reis et al. 2021). In terms of supervised models for regression or classification, such as the morphological classification of galaxies (Spindler, Geach & Smith 2021), Young Star Object finders (Martón et al. 2019), automated classification of eclipsing binary systems in the VISTA Variables in the Vía Láctea (VVV) survey (Daza-Perilla et al. 2023), Drifting Features: detection and evaluation in the context of automatic RRLs identification in VVV (Cabral et al. 2021), variable star classification across the Galactic bulge and disc with the VVV survey (Molnar et al. 2022) with light curves of stars, classification of galaxies and QSOs (Logan & Fotopoulou 2020), estimation of photometric redshifts (Eriksen et al. 2020), among many others.

* E-mail: vanessa.daza@unc.edu.ar (IVD-P); msgro@unc.edu.ar (MAS); lbaravalle@unc.edu.ar (LDB)

Índice general

Índice general

Índice de figuras

Índice de tablas

1	Introducción	1
1.1	Estructura a Gran Escala del Universo	1
1.2	Zona de Evitamiento	7
1.3	Esfuerzos para resolver los problemas de la Zona de Evitamiento	10
1.4	Aprendizaje Automático en Astronomía	16
1.5	Motivación de esta Tesis	17
2	Datos	18
2.1	Datos Primarios	18
2.1.1	Imágenes y catálogos en infrarrojo cercano	18
2.1.2	Catálogos de rayos X	22
2.2	Datos Secundarios	22
2.3	Herramientas	25
2.3.1	Herramientas de base de datos e imagenes	25
2.3.2	Herramientas de Aprendizaje automatizado	25
3	Comparación de clasificaciones por HAA y métodos no automáticos	31
3.1	Herramientas de aprendizaje automático aplicados a la Zona de Evitamiento	31
3.2	VVV NIRGC	32

3.3	Estadística de la muestra	34
4	Inspección Visual	41
4.1	Descripción de clasificación	42
4.2	Distribución de Energía Espectral y Fotometría	51
5	Zonas de Interés	56
5.1	Zona 1	64
5.2	Zona 2	65
5.3	Zona 3	66
5.4	Zona 4	68
5.5	Zona 5	69
5.6	Aspectos generales de las zonas	70
6	Conclusión	72
7	Discusión	76
7.1	Lo crítico que es clasificar objetos en la ZoA	76
7.2	Ética e Inteligencia Artificial	77
	Bibliografía	80

Índice de figuras

1.1	Similitud entre una neurona y la Estructura a Gran Escala del Universo. (<i>Mark Miller/Virgo Consortium/Visual Complexity</i>)	2
1.2	Figura extraída de Seldner et al. (1977). Se muestra la distribución angular del conteo de galaxias más brillantes que $B \sim 19$ mag en el plano del cielo reconstruido del catálogo de galaxias Lick. Además se observa en la imagen, el número de galaxias observadas en celdas de $10' \times 10'$ a través del hemisferio norte Galáctico, donde el polo norte Galáctico se encuentra en el centro de la imagen y el ecuador Galáctico en los bordes. Es posible notar su distribución no homogénea, con grupos de galaxias representados en blanco y las zonas oscuras representan una menor densidad de galaxias, dando la semejanza de vacío.	3
1.3	Mapa de redshift con la distribución de las galaxias en 6dFGS de Jones et al. (2009) hasta $z = 0.1$. Con cortes en declinación especificados.	5
1.4	Mapa de Anisotropías del Fondo Cósmico de Microondas (CMB) observado por el satélite Planck. Se aprecian las pequeñas fluctuaciones de temperatura, correspondientes a regiones con densidades diferentes (Planck Collaboration et al. 2020)	7
1.5	Distribución de “nebulosas” corregidas por el oscurecimiento galáctico (Hubble 1936). En ella se observa levemente el vacío de galaxias correspondiente a la Zona de Evitamiento.	8
1.6	Figura extraída del trabajo de Kraan-Korteweg & Lahav (2000). Distribución de áreas iguales de Aitoff en coordenadas Galácticas para galaxias con $D \geq 1.3$, codificadas por diámetro, los círculos pequeños representan galaxias con $1.3 \leq D < 2'$, los círculos más grandes $2' \leq D < 3'$ y los círculos grandes representan galaxias con $D \geq 3'$. Las líneas de contorno marcan la absorción en el azul $A_B = 1.0$, determinada por los mapas de extinción de Schlegel et al. (1998). Aquí es posible distinguir la distribución incompleta de las galaxias debido a la ZoA.	9

ÍNDICE DE FIGURAS

1.7	Figura extraída de Kraan-Korteweg & Lahav (2000). Distribución de áreas iguales de galaxias de los catálogos ESO, UGC y MCG codificadas por diámetro corregidos por extinción $D^{\circ} \geq 1'.3$, incluidas las galaxias identificadas en la ZoA en las búsquedas ópticas para niveles de extinción de $A_B \leq 3.^m0$ (contornos). Los diámetros están codificados al igual que en la Fig.1.6, a excepción de las áreas para las que aún no se conocen las posiciones de las galaxias o sus diámetros (correspondientes a las áreas demarcadas). Puede observarse que la ZoA se redujo considerablemente en comparación a la Fig.1.6.	10
1.8	Figura extraída de Macri et al. (2019). Distribución en todo el cielo de las galaxias 2MRS en coordenadas galácticas, utilizando el color para indicar sus desplazamientos al rojo. .	12
1.9	Figura extraída de Baravalle et al. (2018) que muestra la distribución de los objetos de 2MASSX en coordenadas galácticas. El área del relevamiento VVV está delineada con los <i>tiles</i> estudiados representadas por rectángulos.	14
1.10	Figura extraída de Baravalle et al. (2021) que muestra la distribución de galaxias en la región del disco del VVV, superpuesto con los contornos de extinción A_V (1, 3, 5, 7, 9, 11, 13, 15, 20 y 25 mag) derivado de los mapas de extinción de Schlafly & Finkbeiner (2011a). Las galaxias del VVV NIRGC están representadas con círculos rojos, mientras que aquellas que se encuentran en otros surveys, como 2MASSX están representadas con triángulos azules.	15
2.1	Curvas de transmisión para los filtros Z, Y, J, H y K_S del relevamiento VVV en la cámara VIRCAM (curvas de colores etiquetadas), junto a la curva de Eficiencia Cuántica (QE) del detector como una curva superpuesta de color violeta	19
2.2	Construcción de un <i>tile</i> a partir del movimiento del telescopio en 6 diferentes <i>pawprint</i> .	20
2.3	Esquema de los seis grandes relevamientos públicos por el telescopio VISTA. Estos cubren diferentes áreas del cielo en diferentes profundidades con un amplio rango de objetivos científicos.	21
2.4	Esquema de la distribución de <i>tiles</i> para la cobertura del relevamiento VVV en coordenadas Galácticas.	21
2.5	Ejemplo de una Matriz de Confusión con 2 etiquetas	27

ÍNDICE DE FIGURAS

3.1	Histograma acumulativo para la cantidad de datos obtenidos para cada separación de cada cruce realizado con TOPCAT.	34
3.2	Histograma acumulativo con la cantidad de datos duplicados para el cruce entre Z21_disc/GAIA y Z21_disc/VVV	35
3.3	Galaxias en común entre el catálogo VVV NIRGC y la muestra <i>galXray</i> . Las imágenes de las galaxias fueron obtenidas en la banda K_S del relevamiento VVV, con un tamaño de $1' \times 1'$	37
3.4	Distribución de galaxias de Z21 en el disco sur del VVV. las galaxies de la muestra <i>galXray</i> están representadas con puntos rojos. Las galaxias confirmadas visualmente por VVV NIRGC están en naranja	38
3.5	Diagrama de clasificación que ilustra la selección de diversas muestras y submuestras estudiadas, con especial énfasis en la muestra <i>NOMatch</i> , la cual constituye nuestra mayor preocupación en este trabajo.	39
4.1	Modo MULTIGETIMAGE de VSA para obtener imágenes visuales de los relevamientos del telescopio VISTA, desde catálogos con coordenadas ecuatoriales.	42
4.2	Ejemplos para stamps categorizados como “región normalmente abarrotada”, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.816$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.445$	43
4.3	Ejemplos para un campo con una estrella central, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.867$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.445$	44
4.4	Ejemplos para un stamp con la region central vacía, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.826$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.422$	45
4.5	Ejemplos para un grupo de estrellas en el campo, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.826$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.429$	46

ÍNDICE DE FIGURAS

4.6	Ejemplos para regiones con una estrella saturada, ya sea alrededor o en el centro. El panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.967$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.538$	47
4.7	Ejemplos para estructura de Spikes. El panel de arriba corresponde a una fuente con alta probabilidad, $P_x = 0.944$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.387$	48
4.8	Ejemplos para una zona con formación estelar. El panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.969$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.558$	48
4.9	Ejemplos para PUVS, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.544$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.449$	49
4.10	Ejemplos para stamps que muestran una asociación de estrellas masivas. El panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.963$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.49$	49
4.11	Ejemplos para <i>stamp</i> con estructuras de PUVSFR. El panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.953$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.499$	50
4.12	modelamientos de SED para la muestra <i>NOMatch</i> , usando la configuración mostrada en la tabla 4.2	53
4.13	Diagrama mostrando el comportamiendo de los flujos para diferentes submuestras, las galaxias confirmadas visualmente, datos con contraparte en el relevamiento VVV y los <i>NOMatch</i>	54

ÍNDICE DE FIGURAS

5.1	Distribución de las galaxias de la muestra de Zhang et al. (2021) en la región del disco sur del VVV. La muestra <i>galXray</i> está representada por puntos rojos, las galaxias confirmadas visualmente del VVV NIRGC están en naranja, las 4 galaxias en común entre VVV NIRGC y Z21, son diamantes azules y las galaxias de Schröder et al. (2019) por puntos negros. El centro de las sobredensidades reportadas por Soto et al. (2022) están representados por círculos azules semi-transparentes, que denotan el radio de cada sobredensidad amplificado por un factor de 4. Los cuadrados negros marcan las Zonas de Interés estudiadas. Los cuadrados verdes denotan remanentes de supernova compilados por Chandra. Los isocontornos A_V derivados de los mapas de extinción de Schlafly & Finkbeiner (2011a) se superponen en un gradiente gris con niveles de 11, 15, 20, y 25 mag.	57
5.2	Distribución de las galaxias de Zhang et al. (2021) en la región del disco sur del relevamiento VVV (<i>galXray</i>). La muestra <i>galXray</i> esta representada por codificación por colores, con la probabilidad P_x de ser una galaxia. Los cuadrados negros marcan las Zonas de Interés estudiadas. Los isocontornos A_V derivados de los mapas de extinción de Schlafly & Finkbeiner (2011a) se superponen en un gradiente gris con niveles de 11, 15, 20, y 25 mag.	58
5.3	Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z1, incluyendo los objetos de <i>galXray</i> como puntos rojos.	59
5.4	Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z2, incluyendo los objetos de <i>galXray</i> como puntos rojos.	60
5.5	Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z3, incluyendo los objetos de <i>galXray</i> como puntos rojos.	61
5.6	Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z4, incluyendo los objetos de <i>galXray</i> como puntos rojos.	62
5.7	Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z5, incluyendo los objetos de <i>galXray</i> como puntos rojos.	63
5.8	<i>stamps</i> de la zona de interes Z1, en diferentes longitudes de onda con los surveys disponibles.	64
5.9	<i>stamps</i> de la zona de interes Z2, en diferentes longitudes de onda con los surveys disponibles.	66
5.10	<i>stamps</i> de la zona de interes Z3, en diferentes longitudes de onda con los surveys disponibles.	67

5.11	<i>stamps</i> de la zona de interes Z4, en diferentes longitudes de onda con los surveys disponibles.	69
5.12	<i>stamps</i> de la zona de interes Z5, en diferentes longitudes de onda con los surveys disponibles.	70

Índice de tablas

2.1	Longitudes de onda efectivas para el conjunto de filtros VISTA utilizados en las observaciones del relevamiento VVV y la extinción relativa para cada filtro basada en la ley de extinción de Cardelli et al. (1989) (Catelan et al. 2011)	20
3.1	Tabla con con la cantidad de datos del cada cruce, con sus respectiva cantidad de datos duplicados.	36
3.2	tabla con el resultado del cruce de Z21 en disco del relevamiento VVV con Gaia DR3.	40
4.1	Tabla con las clasificaciones de las fuentes Sin contraparte, con el respectivo total de <i>stamps</i> clasificados para cada categoría. La sigla PUVS se refiere a “Photometrically Ultra Variable Star”, mientras que PUVSFR “Photometrically Ultra Variable Star Forming Region”.	42
4.2	Valores de los parámetros utilizados para generar los modelos de SED analizados. Para los valores no descritos en la tabla se usaron los valores por defecto del software.	52
5.1	Estadística de las probabilidades P_x para las distintas zonas y muestras.	71

Capítulo 1

Introducción

1.1. Estructura a Gran Escala del Universo

Al observar nuestro entorno, notamos que existe un orden en todas las cosas que nos rodean. En nuestro planeta existen diminutas estructuras que dan origen a nuevas configuraciones. Por ejemplo, las células se organizan en tejidos, los tejidos forman órganos y, finalmente, los órganos forman a los seres vivos. Este patrón no solo se limita a nuestro planeta, de manera análoga, al dirigir nuestra mirada hacia el Universo que nos rodea, observamos un orden similar. Como habitantes del planeta tierra, orbitamos alrededor del Sol, que a su vez gira en torno al centro galáctico a una distancia aproximada de 8.5 kpc.. Nuestra galaxia es la Vía Láctea (VL), que a su vez es parte de un conjunto de galaxias llamado Grupo Local, que incluye la galaxia Andrómeda (M31), las nubes de Magallanes, la galaxia del triángulo (M33) y entre otras más de 50 galaxias, mayormente enanas (van den Bergh 2000). El Grupo Local, a su vez, forma parte del supercúmulo de Virgo, que abarca más de 100 grupos y cúmulos de galaxias (Einasto et al. 2007).

A medida que aumentamos la escala en la que observamos, nos enfrentamos a sistemas cada vez más organizados y coherentes, lo que aparenta ser el esqueleto de nuestro Universo, pero, ¿Qué son estas estructuras?, ¿Qué representan y como podemos estudiarlas?. A esto le llamamos la Estructura a Gran Escala del Universo (EGE). Para hacernos una idea, podemos notar la figura 1.1, donde apreciamos la similitud que existe entre el ordenamiento de las neuronas observado a través de un microscopio, y lo que resulta de una simulación cosmológica para el estudio de la EGE.

Para poder estudiar la EGE, es necesario conocer los inicios del mapeo de galaxias. Los primeros indicios sobre la concepción de estos objetos, provienen de Immanuel Kant

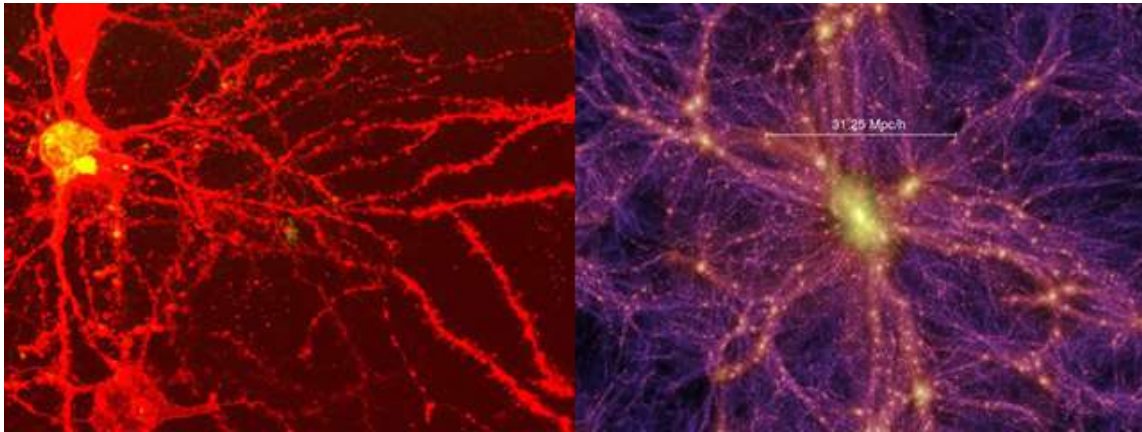


Figura 1.1: Similitud entre una neurona y la Estructura a Gran Escala del Universo. (Mark Miller/Virgo Consortium/Visual Complexity)

(1724–1808), filósofo alemán que motivado por la mecánica newtoniana, postuló que el Sistema Solar se formó a partir de un disco de gas condensado, el cual a su vez, pertenecía a una estructura aún más grande y densa, lo que hoy conocemos como galaxia. Este modelo, sugiere que las estructuras conocidas como galaxias, se asemejaban a los objetos difusos observados en el cielo, a los que llamó “Universos Islas”.

Más tarde, el impactante trabajo de Edwin Hubble en la primera mitad del siglo XX desempeñó un papel crucial en la comprensión de las galaxias. Hubble, en su catálogo de 400 fuentes que denominó “Nebulosas extragalácticas”, clasificó estos objetos entre elípticas y espirales (Hubble 1926). Además, al observar su distribución homogénea en escalas mayores de 100 Megapársecs (Hubble 1934), y culminando con su obra “Realm of the nebulae” (Hubble 1936), abrió la puerta al mapeo de la EGE.

Las mediciones de la EGE experimentaron un gran avance con la introducción de los catálogos de galaxias Lick de Shane & Wirtanen de 1967, de aproximadamente 1 millón de galaxias obtenidas con las placas fotográficas del telescopio refractor de 0.5 m del observatorio Lick. Posteriormente, Seldner et al. (1977) publicaron un mapa detallado mediante conteo de galaxias en celdas angulares en el plano del cielo, basados en el catálogo anterior y revelando así, el primer indicio de las estructuras de hilos, vacíos y grupos en la EGE (ver figura 1.2).

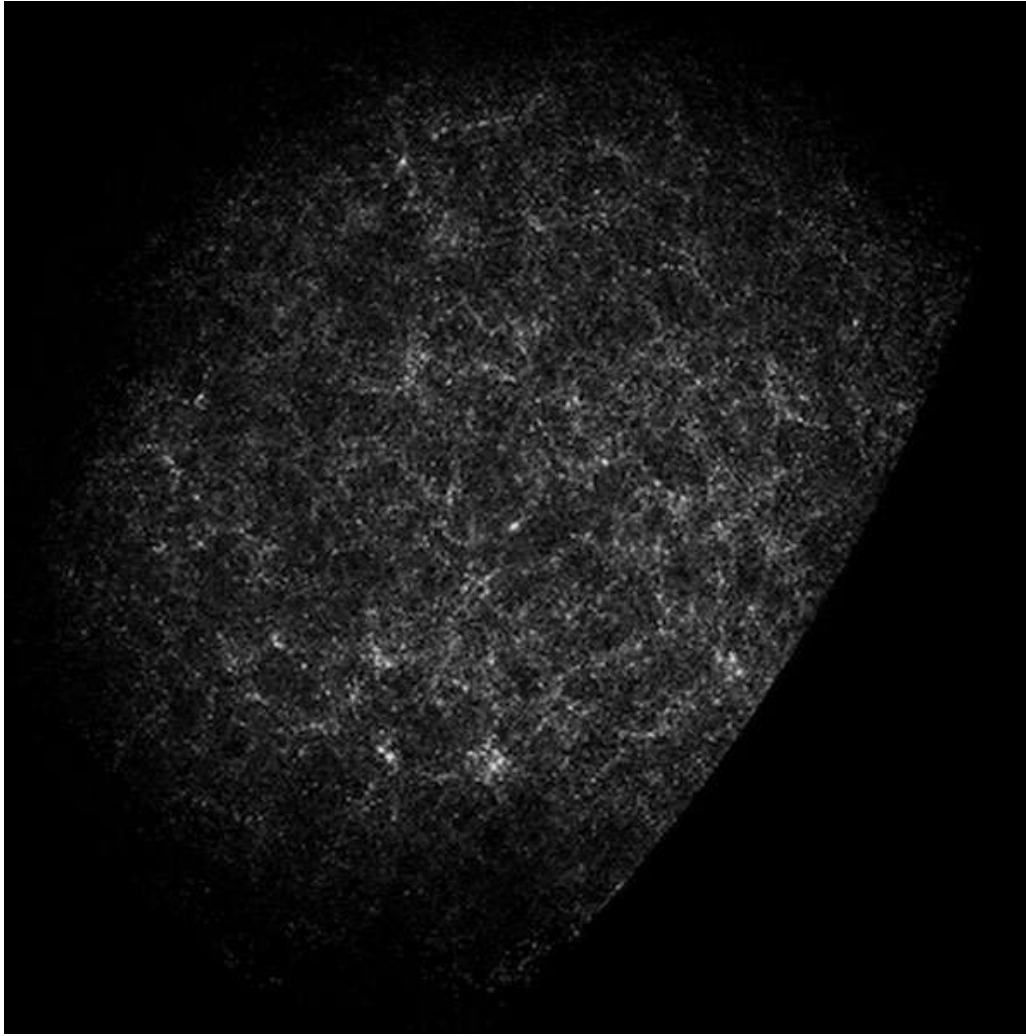


Figura 1.2: Figura extraída de Seldner et al. (1977). Se muestra la distribución angular del conteo de galaxias más brillantes que $B \sim 19$ mag en el plano del cielo reconstruido del catálogo de galaxias Lick. Además se observa en la imagen, el número de galaxias observadas en celdas de $10' \times 10'$ a través del hemisferio norte Galáctico, donde el polo norte Galáctico se encuentra en el centro de la imagen y el ecuador Galáctico en los bordes. Es posible notar su distribución no homogénea, con grupos de galaxias representados en blanco y las zonas oscuras representan una menor densidad de galaxias, dando la semejanza de vacío.

Varios estudios se llevaron a cabo para estudiar la distribución de masa en el Universo (Dirac 1938; Sciama 1953). Estos resultados sentaron las bases para los primeros relevamientos de corrimiento al rojo a gran escala (*redshift surveys*), los cuáles obtuvieron espectros ópticos de galaxias, permitiendo medir así su redshift y analizar su distribución espacial mediante grandes muestras de datos de galaxias. Es posible destacar 2 relevamientos de corrimiento al rojo primordiales, el relevamiento KOS (Kirshner, Oemler, Schechter; Kirshner et al. 1978) y el relevamiento CfA (Center for Astrophysics; Davis et al.

1982). El relevamiento KOS midió los redshift de 164 galaxias (con magnitud $m < 15$) en 8 campos diferentes del cielo, cubriendo un total de 15 grados cuadrados.

El estudio del relevamiento original del CfA, contiene redshifts para 2,400 galaxias con magnitud más brillante que 14.5 a través de los polos galácticos norte y sur, cubriendo un total de 2.7 Estereorradiánes. Con objetivo principal del proyecto, estudiar la cuantificación de la agrupación de galaxias en tres dimensiones, produciendo los primeros mapas tridimensionales de EGE. Posteriormente, el segundo relevamiento del CfA (CfA2) obtuvo espectros para un aproximado de 5,800 galaxias, revelándose la existencia del supercúmulo “La Gran Pared” (conocida en inglés como *The Great Wall*; Geller & Huchra 1989), el cual se extiende por un ancho de $170h^{-1}$ Mpc, confirmando que las galaxias no se distribuyen homogéneamente, sino agrupadas en superficies esféricas, las cuáles conforman vacíos.

Estos relevamientos al corrimiento al rojo han progresado rápidamente en el último tiempo, dando un importante salto con el desarrollo de espectrógrafos multi-objetos, cuya técnica permitió observar numerosas fuentes extragalácticas de manera simultánea. Destacando el “Two Degree Field Galaxy Redshift Survey” (2dFGRS; Colless et al. 2001), con un volumen del orden de $4 \times 10^7 h^3$ Mpc⁻³, magnitud límite de $b_J \leq 19.45$ y redshifts espectroscópicos de más de 200,000 galaxias, el “Sloan Digital Sky Survey” (SDSS; York et al. 2000), con un volumen del orden de $4 \times 10^7 h^3$ Mpc⁻³, magnitud límite de $r \leq 17.77$ y redshifts espectroscópicos de un millón de galaxias, y el “6dF Galaxy Survey” (6dFGS; Jones et al. 2009), con magnitudes límite $K, b_J \leq 12.75, 16.75$ y un volumen efectivo parecido al 2dFGRS.

En la figura 1.3 se encuentra el mapa de redshift con la distribución de galaxias hasta $z = 0.1$ en el relevamiento 6dFGS (Jones et al. 2009). Se puede observar un patrón en la distribución de galaxias, identificando cúmulos de galaxias y la estructura filamental entre estos grupos, haciendo evidente la EGE.

A medida que observamos más lejos en el Universo, observamos también su pasado. En un principio, el Universo consistía en un gas denso y caliente, lleno de partículas como electrones y protones, donde la luz no podía viajar muy lejos (Peebles 1980). A medida que el Universo se expande, y su temperatura baja, la luz comienza a viajar hasta el día de hoy, permitiéndonos observarla como la Radiación de Fondo Cósmico de Microondas (Penzias & Wilson 1965, CMB;). Esta radiación, que se entiende como el remanente de la época de recombinación, se atribuye a la predicción inicial de Alpher et al. (1948). Sin em-

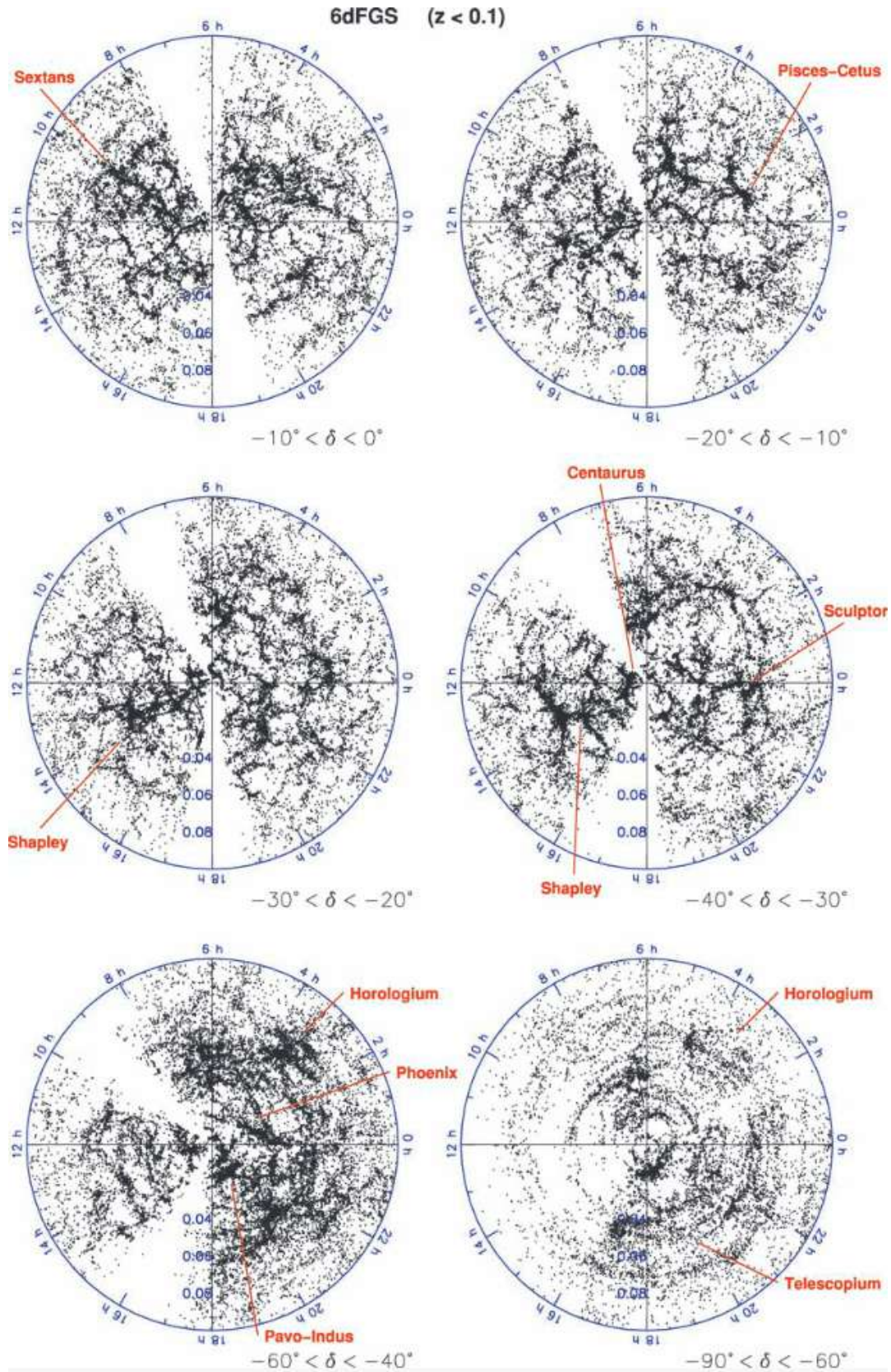


Figura 1.3: Mapa de redshift con la distribución de las galaxias en 6dFGS de Jones et al. (2009) hasta $z = 0.1$. Con cortes en declinación especificados.

bargo, su detección no fue hasta el año 1964, cuando al observar una galaxia, se encontró una señal adicional débil y uniforme en todas las direcciones sin variación temporal, observándose por primera vez el CMB.

La primera misión espacial diseñada específicamente para estudiar el CMB fue el “Cosmic Background Explorer” (Mather 1982), con objetivo principal encontrar la temperatura de esta radiación. En 2001 fue lanzada la misión “Wilkinson Microwave Anisotropy Probe” (WMAP; Bennett et al. 2003) para estudiar con más detalle las fluctuaciones del CMB, sus resultados ayudaron a determinar las proporciones de los principales constituyentes del Universo y así, establecer un modelo estándar cosmológico.

Finalmente, la misión espacial Planck (Tauber et al. 2010; Planck Collaboration et al. 2011) de la Agencia Europea Espacial (ESA), fue lanzada en 2009 para estudiar el CMB en aún más detalles, cubriendo un mayor rango en frecuencia que las misiones anteriores y haciendo posible la separación precisa de los componentes de las longitudes de onda sub-milimétrica y milimétrica en el cielo, incluso separando la emisión de nuestra Galaxia. Con una temperatura de $2.72548 \pm 0.00057 K$ (Fixsen 2009), esta radiación ofrece una ventana para explorar los primeros momentos del Universo y entender sus características fundamentales.

En la figura 1.4, se observa un mapa de las anisotropías en el CMB con el telescopio Planck. Un patrón observable de estas zonas “calientes” y “frías” en el mapa de distribución anisotrópica se conoce como el Dipolo Anisotrópico de la Radiación de Fondo de Microondas, y es altamente creíble que su patrón se debe al efecto Doppler producido por la velocidad peculiar de nuestro Grupo Local (Erdoğdu et al. 2006). Sin embargo, estudios en el movimiento del Grupo Local han llevado a incongruencias con el movimiento dado en el dipolo del CMB (Kraan-Korteweg & Lahav 2000). En el trabajo de Loeb & Narayan (2008), comparan la dinámica del dipolo del CMB con la velocidad peculiar del Grupo Local obtenida del “2MASS Redshift Survey” (2MRS; Huchra et al. 2012), llegando a la conclusión de que esta incongruencia se debe a la falta de galaxias en la zona cercana al disco Galáctico y, que detrás de esta zona, debería existir una galaxia escondida, en masa comparable a Andrómeda ($M \sim 2 \times 10^{12} M_{\odot}$) a una distancia de ~ 1 Mpc, o, un cúmulo de galaxias comparable al cúmulo de Coma ($M \sim 10^{15} M_{\odot}$), a una distancia de ~ 20 Mpc, esto de acuerdo al cálculo del exceso en velocidad hacia el centro Galáctico de $\Delta v \equiv v_{ZOA,x} \sim 120 \text{ km s}^{-1}$.

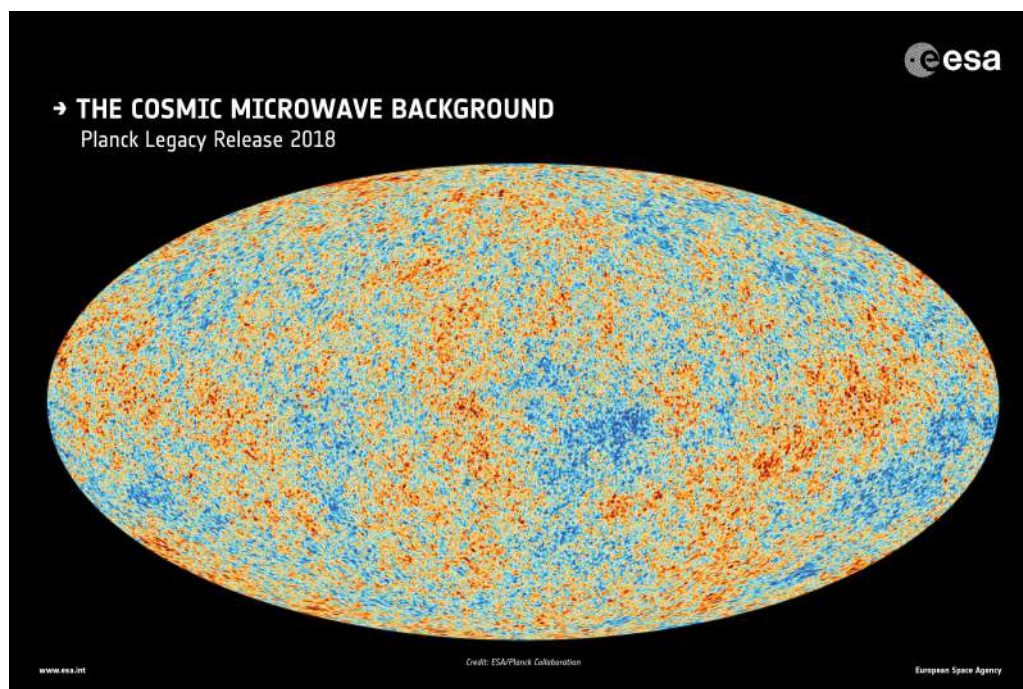


Figura 1.4: Mapa de Anisotropías del Fondo Cósmico de Microondas (CMB) observado por el satélite Planck. Se aprecian las pequeñas fluctuaciones de temperatura, correspondientes a regiones con densidades diferentes (Planck Collaboration et al. 2020)

Esta discrepancia plantea la pregunta ¿Qué es lo que estamos pasando por alto?. Para responder a esto, es fundamental estudiar primero nuestra Galaxia, como interactúa con el Universo local y como afecta a las observaciones extragalácticas (Bilicki & Chodorowski 2010; Bengaly et al. 2017).

1.2. Zona de Evitamiento

La Vía Láctea, una galaxia espiral con una masa de aproximadamente $10^{12} M_{\odot}$ y un radio de alrededor de 30kpc, presenta una estructura compleja y distintiva. Esta estructura incluye el bulbo galáctico, que alberga estrellas ancianas con metalicidades inferiores a las solares ($[Fe/H] = -0.2$; Binney et al. 1991; Blitz & Spergel 1991) y el disco galáctico dividido en disco fino y disco grueso, que contribuyen significativamente a su complejidad. Además, los brazos espirales y el Halo, junto con los cúmulos globulares que albergan poblaciones estelares más antiguas, como las RR Lyrae (Binney et al. 1997), completan este panorama galáctico. Sin embargo, la compleja estructura de la Vía Láctea representa un desafío para el mapeo de la EGE. El polvo y las estrellas de nuestra galaxia oscurecen una franja del cielo en las longitudes de onda visibles, creando un vacío significativo en

los mapas extragalácticos a bajas latitudes galácticas.

Esta región, conocida como la Zona de Evitamiento (ZoA; del inglés *Zone of Avoidance*), se hizo evidente a medida que los catálogos de todo el cielo se expandieron. La primera referencia a esta región se remonta a 1879 con el “General Catalogue of Nebulae”, que identificó una zona con pocas galaxias. Posteriormente, con el “New General Catalogue” (Dreyer 1888, 1895), se logró distinguir estructuras como aglomeraciones de galaxias e incluso partes de la EGE, como los cúmulos de Virgo, Fornax, Perseus, Piscis y Coma (de Vaucouleurs 1981). A principios del siglo XX, con la publicación de “Realm of the Nebulae” (Hubble 1936) y el descubrimiento de la absorción galáctica, la ZoA fue cada vez más estudiada, lo que permitió reducir su tamaño.

La figura 1.5 muestra el mapa de distribución de fuentes extragalácticas y el vacío de fuentes correspondientes a la ZoA.

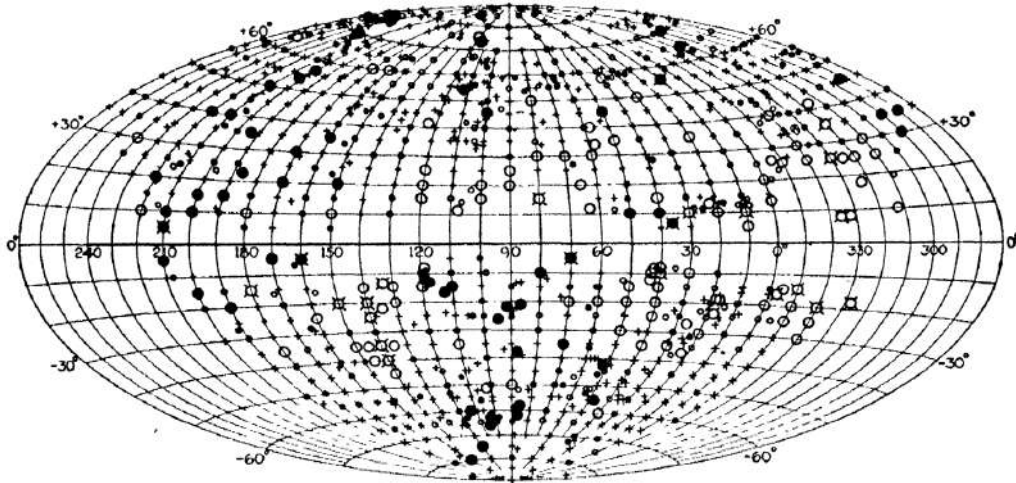


Figura 1.5: Distribución de “nebulosas” corregidas por el oscurecimiento galáctico (Hubble 1936). En ella se observa levemente el vacío de galaxias correspondiente a la Zona de Evitamiento.

Avanzando en el estudio de la EGE, la disponibilidad de nuevos catálogos de galaxias en el cielo ha sido fundamental para comprender esta red cósmica. Catálogos como el “Uppsala General Catalog” (UGC; Nilson 1995) en el hemisferio norte (con declinación $\delta \geq -2.5^\circ$), el “ESO Uppsala Catalog” (Lauberts 1982) para el hemisferio sur ($\delta \leq -17.5^\circ$), y el “Morphological Catalog of Galaxies” (MCG; Vorontsov-Vel’Yaminov & Arkhipova 1963) para la franja entre $-17.5^\circ < \delta < -2.5^\circ$, han proporcionado valiosa información para trazar la distribución de galaxias en coordenadas galácticas.

En la figura 1.6, centrada en el bulbo galáctico, se ilustra la distribución de galaxias codificadas por diámetro, resaltando estructuras relevantes para la dinámica del Universo local, entre estas, se destacan el Supercúmulo local centrado en el cúmulo de Virgo a

$\ell = 284^\circ, b = 74^\circ$, donde ℓ y b son las coordenadas Galácticas, la cadena Perseus-Piscis atravesando la ZoA en $\ell = 95^\circ$ y $\ell = 165^\circ$, la sobredensidad general en la dirección del dipolo del CMB ($\ell = 276^\circ, b = 30^\circ$; (Kogut et al. 1993) y la sobredensidad general de galaxias en la región del Gran Atractor centrada en $\ell = 320^\circ, b = 0^\circ$ (Kolatt et al. 1995), junto con cúmulos como Hydra ($270^\circ, 27^\circ$), Antlia ($273^\circ, 19^\circ$), Centauro ($302^\circ, 22^\circ$) y Pavó ($332^\circ, -24^\circ$).

Sin embargo, a medida que aumenta el número de galaxias conocidas y los mapas extragalácticos de todo el cielo se expanden, la franja correspondiente a la ZoA, ubicada entre $|b| < 10^\circ$, se vuelve cada vez más notable, careciendo de galaxias en los catálogos. Esta escasez se atribuye a la disminución en la luz de fuentes extragalácticas debido al polvo grueso del disco Galáctico, acentuado por el aumento de estrellas de campo cercanas al plano Galáctico. Estos factores obstaculizan la detección de otras fuentes, revelándose con claridad la gran interrogante que representa la ZoA.

La comparación de la ZoA con los mapas de extinción a $100 \mu\text{m}$ de los experimentos DIRBE (Schlegel et al. 1998), demuestra que el área donde el conteo de galaxias decae totalmente coincide casi perfectamente con los contornos de absorción en la banda azul A_B de $1.^m0$ (con $A_B = 4.14 \cdot E(B - V)$) (Cardelli et al. 1989).

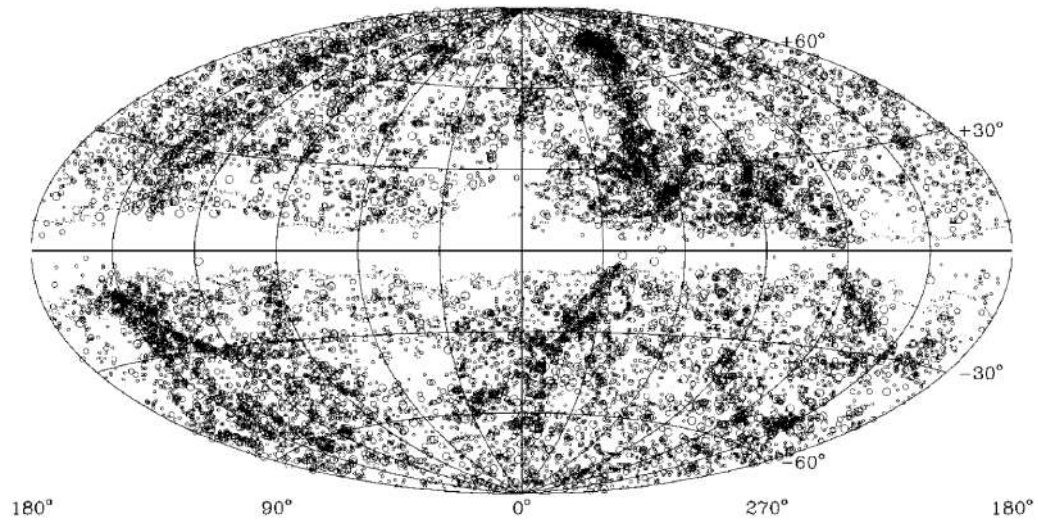


Figura 1.6: Figura extraída del trabajo de Kraan-Korteweg & Lahav (2000). Distribución de áreas iguales de Aitoff en coordenadas Galácticas para galaxias con $D \geq 1.'3$, codificadas por diámetro, los círculos pequeños representan galaxias con $1.'3 \leq D < 2'$, los círculos más grandes $2' \leq D < 3'$ y los círculos grandes representan galaxias con $D \geq 3'$. Las líneas de contorno marcan la absorción en el azul $A_B = 1.^m0$, determinada por los mapas de extinción de Schlegel et al. (1998). Aquí es posible distinguir la distribución incompleta de las galaxias debido a la ZoA.

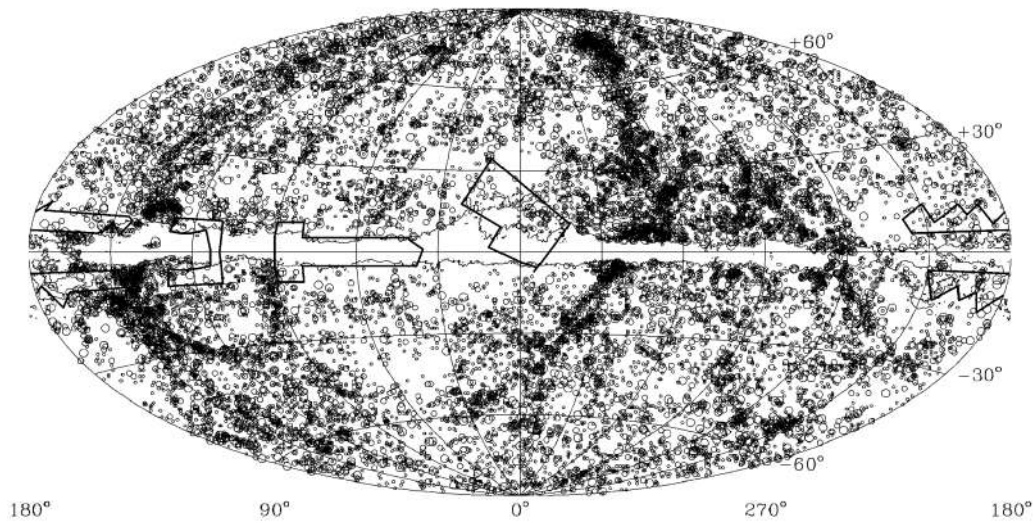


Figura 1.7: Figura extraída de Kraan-Korteweg & Lahav (2000). Distribución de áreas iguales de galaxias de los catálogos ESO, UGC y MCG codificadas por diámetro corregidos por extinción $D^{\circ} \geq 1.3$, incluidas las galaxias identificadas en la ZoA en las búsquedas ópticas para niveles de extinción de $A_B \leq 3.0$ (contornos). Los diámetros están codificados al igual que en la Fig.1.6, a excepción de las áreas para las que aún no se conocen las posiciones de las galaxias o sus diámetros (correspondientes a las áreas demarcadas). Puede observarse que la ZoA se redujo considerablemente en comparación a la Fig.1.6.

Es interesante observar la evolución de la ZoA, al comparar las figuras 1.6 y 1.7. Estas comparaciones muestran cómo las búsquedas profundas de galaxias han reducido significativamente el tamaño de la ZoA, lo que permite trazar de manera más precisa la EGE del Universo cercano.

Además, en la figura 1.7, la densidad de galaxias en la región del Gran Atractor es aún más evidente. Se puede apreciar claramente la conexión de la cadena Perseus-Pisces a través de la Vía Láctea ($\ell = 165^{\circ}$), lo que resalta la importancia de estudiar la ZoA para comprender mejor la distribución y evolución de las galaxias en el Universo cercano.

1.3. Esfuerzos para resolver los problemas de la Zona de Evitamiento

Entre los primeros intentos para detectar galaxias en la ZoA, destaca la investigación liderada por Renée Kraan-Korteweg, quien empleó múltiples instrumentos y diferentes longitudes de onda para explorar esta región. Uno de los primeros estudios relevantes fue el seguimiento de galaxias detectadas en el área de Puppis realizada por Dodd & Brand (1976). En este contexto, Kraan-Korteweg & Huchtmeier (1992) observaron estas galaxias

en el hidrógeno neutro (HI) e identificaron un nuevo cúmulo nunca antes estudiado, cuya densidad es comparable al cúmulo de Virgo.

El tamaño de la ZoA ha sido significativamente reducido gracias a estudios realizados en las longitudes de onda infrarroja (IR), radio y rayos X. Trabajar en el infrarrojo cercano (NIR) ha sido especialmente útil, ya que en esta región del espectro la extinción causada por el polvo y el gas de la Galaxia disminuye. Los fotones provenientes de radiación NIR son aproximadamente diez veces menos afectados por el polvo que los fotones del rango óptico, lo cual es crucial para estudiar galaxias detrás de la Vía Láctea. Se estima que la luz que atraviesa el disco de la Galaxia disminuye en un 25 % en el óptico y solo un 10 % en el NIR (Henning et al. 1998).

Es importante destacar trabajos previos como el “Two Micron All Sky Survey” (2MASS; Skrutskie et al. 2006), que mapeó aproximadamente el 99.998 % del cielo completo en las bandas de NIR J ($1.25\mu\text{m}$), H ($1.65\mu\text{m}$) y K_s ($2.15\mu\text{m}$), con una magnitud límite en $K_s \sim 13$. Este estudio observó regiones con latitudes galácticas $|b| > 8^\circ$ para el bulbo y $|b| > 5^\circ$ para el disco, lo que permitió encontrar, caracterizar y extraer galaxias a través del polvo de la Vía Láctea (Jarrett et al. 2000a).

Por otro lado Jarrett et al. (2000b) llevó a cabo un estudio de confiabilidad que abarcó 7,000 fuentes extendidas detectadas en un área de 1,000 grados cuadrados, incluyendo galaxia y nebulosas. En este trabajo, se empleó una discriminación en planos de color para separar entre estrellas y galaxias, lo que permitió confirmar posteriormente la presencia de 14 galaxias en la ZoA mediante estudios espectroscópicos en HI y ópticos.

Estos primeros resultados prometedores del 2MASS marcaron un hito importante, allanando el camino para que nuevos relevamientos se aventuren en la búsqueda de objetos extragalácticos en la ZoA. Ejemplos de estos esfuerzos incluyen el “UKIRT Infrared Deep Sky Survey” (UKIDSS; Lawrence et al. 2007) realizado con el telescopio del Reino Unido y el “Kilo-degree INfrared Galaxy Survey” (VIKING; Arnaboldi et al. 2007) realizado con el telescopio VISTA. Estos proyectos trabajan en el NIR con una mayor profundidad que el 2MASS.

La importancia de los relevamientos en NIR radica principalmente en su sensibilidad para la detección de galaxias de tipo temprano y en su utilidad como trazadores de cúmulos y grupos masivos. Además, esta banda nos proporciona una mejor estimación de la masa estelar de las galaxias, lo que resulta útil en aplicaciones de la relación Tully-Fisher (Tully & Fisher 1977), y por ende, en la obtención de velocidades peculiares. La figura 1.8 nos permite visualizar cómo estos estudios nos permiten explorar cada vez más a fondo la

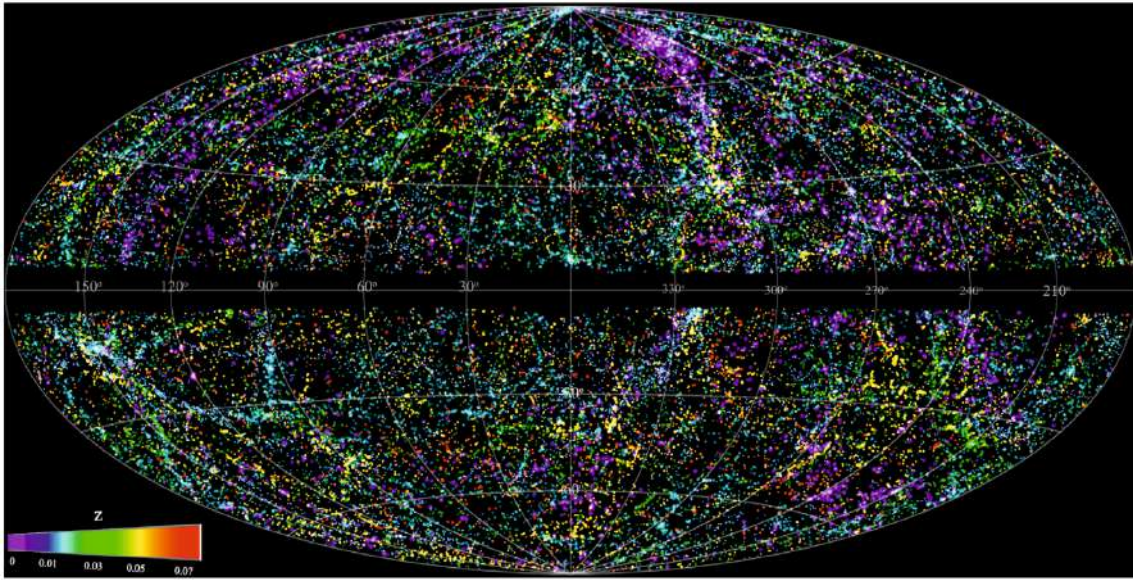


Figura 1.8: Figura extraída de Macri et al. (2019). Distribución en todo el cielo de las galaxias 2MRS en coordenadas galácticas, utilizando el color para indicar sus desplazamientos al rojo.

ZoA, mapeando las galaxias en todo el cielo del 2MRS y reduciendo aún más este vacío de galaxias en comparación con la imagen 1.7. Además, en Macri et al. (2019) se presentan redshifts para las 1,041 galaxias del 2MRS que previamente carecían de esta información, y se destaca que el 90 % de estas están ubicadas entre $|b| = 25^\circ$.

El proyecto “VISTA Variables in the Vía Láctea” (VVV; Minniti et al. 2010) es un relevamiento fotométrico multi-época de la ESO. Se diseñó para estudiar objetos variables y poblaciones estelares en el bulbo ($-10^\circ < \ell < 10^\circ$ y $-10^\circ < b < 5^\circ$) y el disco ($-65^\circ < \ell < -10^\circ$ y $-2^\circ < b < 2^\circ$) de nuestra Galaxia. Sin embargo, este proyecto también ofrece una excelente oportunidad para detectar y estudiar fuentes extragalácticas detrás de la Vía Láctea, como galaxias de campo y cuásares, ya que alcanza una profundidad hasta cuatro magnitudes mayor que el 2MASS. Destacan entre los trabajos importantes el estudio de Amôres et al. (2012), que identificó 204 candidatos a fuentes extendidas en el área del *tile* (1.5° × 1.1°) d003 ($\ell = 298.356^\circ$, $b = -1.650^\circ$), mediante análisis visual de diagramas color-color y su tamaño.

Los trabajos de Coldwell et al. (2014), Galdeano et al. (2021, 2022, 2023) y Duplancic et al. (2024) han dedicado los últimos años a la búsqueda de nuevas galaxias y estructuras detrás del bulbo Galáctico usando los datos del relevamiento del VVV.

Por otro lado, en la zona del disco del relevamiento del VVV, Baravalle et al. (2018) presentó un método morfo-fotométrico basado en SExtractor+PSFEx, (Bertin & Arnouts 1996; Bertin 2011) para identificar y caracterizar fuentes extragalácticas, como una apro-

ximación automática al proceso realizado por Amôres et al. (2012). Los resultados mostraron una mejora sustancial tanto en el número de detecciones como en la calidad de la fotometría. Este método fue probado en los *tiles* d010 ($\ell = 308.569^\circ$, $b = -1.649^\circ$) y d115 ($\ell = 295.438^\circ$, $b = 1.627^\circ$, ver fig. 1.9), donde se confirmaron un total de 345 y 185 fuentes extragalácticas respectivamente, con sus parámetros fotométricos.

En Baravalle et al. (2019), se reportó el primer cúmulo de galaxias confirmado en la ZoA, denominado VVV-J144321.06–611753.9, a $z = 0.234$ en el *tile* d015 ($\ell = 315.836^\circ$, $b = -1.650^\circ$). Este cúmulo está situado en una región de 30×30 minutos de arco al cuadrado centrada en la galaxia más brillante. La extinción media en K_s se estima en 0.73 ± 0.01 mag, revelando un total de 25 galaxias en este cúmulo. Además, se determinaron los colores medios corregidos por extinción de las galaxias, con valores de $(H - K_s) = 0.34 \pm 0.05$ mag, $(J - H) = 0.57 \pm 0.08$ mag y $(J - K_s) = 0.87 \pm 0.06$ mag. Otros parámetros importantes incluyen el radio a mitad de luminosidad media $R_{1/2} = 1.59 \pm 0.''16$, $C = 3.01 \pm 0.08$ y el índice de Sérsic medio $n = 4.63 \pm 0.39$. Estos resultados fueron respaldados por un seguimiento espectroscópico utilizando el telescopio infrarrojo de GEMINI Sur, con el instrumento FLAMINGOS-2, confirmando tanto los parámetros fotométricos como el redshift, validando la naturaleza temprana de estas galaxias dentro de un cúmulo con características comunes.

Más recientemente, en Baravalle et al. (2021), se presentó el “VVV NIR Galaxy Catalogue” (VVV NIRGC) que incluye 5,563 galaxias confirmadas visualmente en todo el disco del VVV (ver figura 1.10 para la distribución de galaxias). Este catálogo proporciona propiedades fotométricas y morfológicas para una muestra significativa de galaxias. El mismo procedimiento descrito en Baravalle et al. (2018) fue utilizado para calcular estas propiedades, y además, se realizó una transformación fotométrica detallada en el trabajo de Soto et al. (2013) para traducir las magnitudes del relevamiento VVV al sistema fotométrico de 2MASS. De esta forma se estimó la densidad estelar definida como el número de estrellas con $(K_s)_0 < 15$ mag en un área de 1 grado cuadrado, obteniendo un valor medio de $\log(N/\text{grados}^2) = 4.933 \pm 0.154$, valor significativamente más alto que el reportado por datos del 2MASS.

Finalmente, en Soto et al. (2022), se identificaron sobredensidades de galaxias en la ZoA utilizando los datos de VVV NIRGC y se caracterizaron los agrupamientos de galaxias mediante tres métodos: las teselaciones de Voronoi (Voronoi 1908) que consiste en celdas que dividen una región según la cantidad de puntos que la región tenga, Minimum spanning tree (MST) que se ha usado en la detección de sobredensidades en diferentes objetos astronómicos (Campana et al. 2008) y el ordenamiento de puntos para identificar

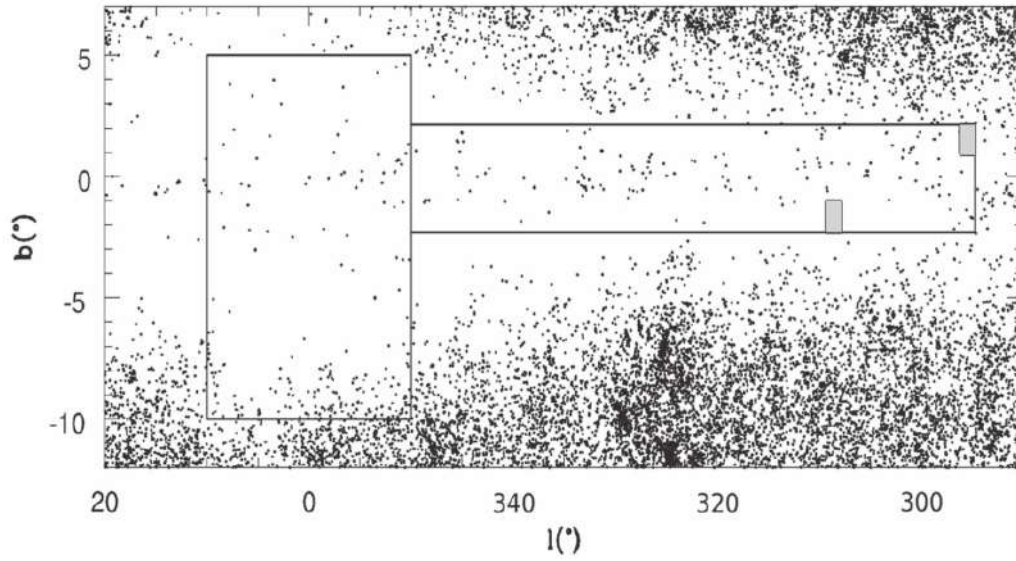


Figura 1.9: Figura extraída de Baravalle et al. (2018) que muestra la distribución de los objetos de 2MASSX en coordenadas galácticas. El área del relevamiento VVV está delineada con los *tiles* estudiados representadas por rectángulos.

estas estructuras (OPTICS; Ordering Points To Identify the Clustering Structure; Ankerst et al. 1999), además de identificar grupos compactos de galaxias mediante pertenencia, aislamiento, compacidad y límites de flujo. Se reportaron 57 sistemas utilizando teselaciones de Voronoi, 56 con MST, 98 con OPTICS y 81 grupos compactos. Estos resultados representan un avance importante en el entendimiento de la EGE y su dinámica en la región local del Universo.

En el rango espectral de radio, la Vía Láctea se destaca por ser transparente a la radiación de la línea de 21-cm del hidrógeno neutro, lo que facilita la detección de galaxias de fondo ricas en H I. Un relevamiento significativo en este sentido fue realizado por McIntyre et al. (2015) con el “Arecibo L-band Feed Array Zone of Avoidance” (ALFA ZOA), cubriendo un área de 15 grados cuadrados centrada en $\ell = 192^\circ$, $b = -2^\circ$, donde se identificaron 61 galaxias. Además, Ramatsoku et al. (2016) llevó a cabo un estudio de imagen de la línea 21-cm H I en el filamento del Supercluster Perseus-Piscis que cruza la ZoA, encontrando 211 galaxias en un área de 9.6° , con 62 % de ellas teniendo contrapartes en NIR en el UKIDSS. Por otro lado, Schröder et al. (2019) presentó un catálogo de galaxias en la parte norte de la ZoA utilizando el radiotelescopio Effelsberg de 100m, identificando 170 detecciones en $|b| < 6^\circ$, de las cuales 67 habían sido registradas en H I, 46 tenían contrapartes en el óptico o NIR, y 29 no tenían con contraparte en ninguna otra longitud

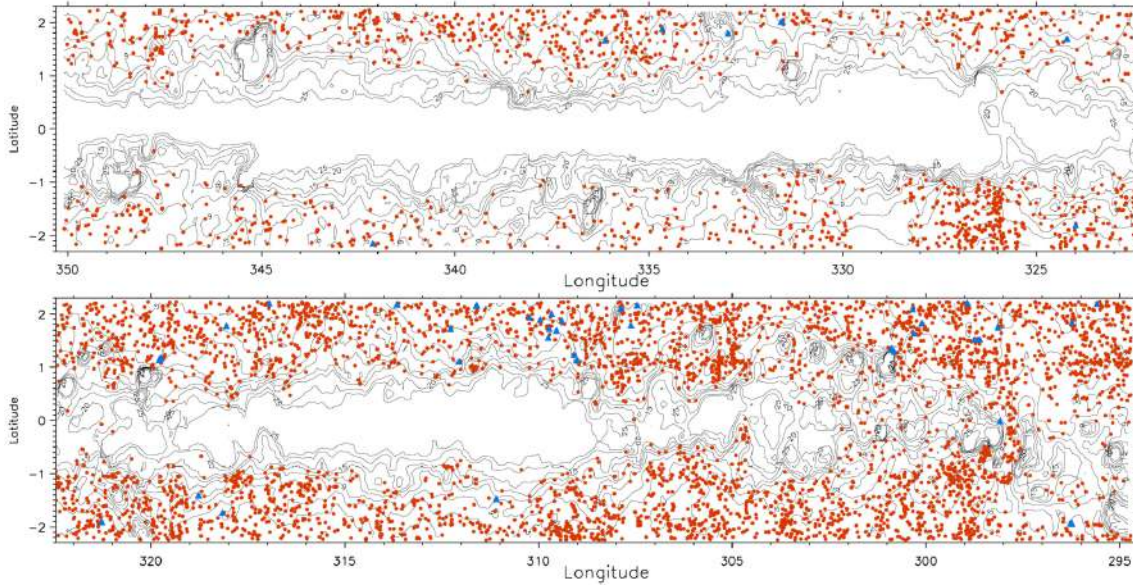


Figura 1.10: Figura extraída de Baravalle et al. (2021) que muestra la distribución de galaxias en la región del disco del VVV, superpuesta con los contornos de extinción A_V (1, 3, 5, 7, 9, 11, 13, 15, 20 y 25 mag) derivado de los mapas de extinción de Schlafly & Finkbeiner (2011a). Las galaxias del VVV NIRGC están representadas con círculos rojos, mientras que aquellas que se encuentran en otros surveys, como 2MASSX están representadas con triángulos azules.

de onda aparte de HI. Recientemente, Kollatschny et al. (2021) reportaron un cúmulo de galaxias a $z = 0.163 \pm 0.003$ detrás de la Vía Láctea, confirmando que las fuentes de radio pueden ser indicadores útiles de la presencia de cúmulos de galaxias en la ZoA.

Los rayos X también son una herramienta valiosa para estudiar fuentes extendidas en la ZoA, ya que la Vía Láctea es transparente a la emisión dura de rayos X. Los cúmulos de galaxias masivos son especialmente fuertes emisores de rayos X, lo que ayuda en el trazamiento de la EGE (Lopes de Oliveira et al. 2006). En particular, el plano Galáctico alberga algunos de los cúmulos más luminosos en rayos X, como los cúmulos de Perseo, Ofiuco, Triangulum Australis y PKS 0745 – 191, que son particularmente destacados en el cielo. Cuatro cúmulos luminosos de rayos X a $z < 0.2$ en todo el cielo de 2 – 10 keV se encuentran en $|b| < 10^\circ$. Además, es bien sabido que los cúmulos de galaxias son estructuras colapsadas gravitacionalmente que albergan más de $10^{14} M_\odot$ de gas calentado difuso entre $10^7 - 10^8$ K, y proyectos como ROSAT (Truemper 1982, 1992), All Sky Suvey (RASS; Voges et al. 1996) y XMM-Newton (Jansen et al. 2001) detectan fácilmente sistemas ricos, incluso detrás del Vía Láctea. A pesar de estas condiciones favorables en rayos X, hay relativamente pocos estudios en el plano Galáctico, con ejemplos notables como el catálogo “Clusters in the Zone of Avoidance” (CIZA, Ebeling et al. 2002; Kocevski et al. 2007),

que representa la primera búsqueda sistemática de un cúmulo de galaxias luminoso en rayos X detrás de la Vía Láctea, y el “Nuclear Spectroscopic Telescope Array” (NuSTAR) serendipitous survey (Harrison et al. 2013), que, aunque no se enfoca específicamente en la búsqueda de fuentes extragalácticas detrás de la Vía Láctea, ha contribuido al campo.

1.4. Aprendizaje Automático en Astronomía

La astronomía está experimentando un rápido crecimiento en el volumen y complejidad de los datos debido a el aumento de telescopios y la mejora de sus capacidades observacionales. Es por esto, que las nuevas herramientas desempeñan un rol esencial en el análisis de los datos, y recurrir a enseñar a computadoras para realizar este análisis se vuelve más importante que nunca.

Los algoritmos de aprendizaje automático (del inglés, *machine-learning*) poseen una alta eficiencia en la identificación de patrones en grandes bases de datos y la detección de señales débiles y complejas. Esta capacidad los convierte en herramientas poderosas para combatir problemas estadísticos y científicos en la astronomía contemporánea.

El reino del aprendizaje automático se divide en 2 grandes familias de algoritmos; el aprendizaje supervisado y el no-supervisado. En el aprendizaje no-supervisado, podemos nombrar a k -means y Gaussian Mixture Model (GMM; Dempster et al. 1977), que son algoritmos que requieren una definición a priori de la cantidad de grupos esperados; algoritmos basados en densidad de datos, como las teselaciones de Voronoi y DBSCAN (Ester et al. 1996) que requieren sólo una definición del número mínimo de clusters y la distancia entre los puntos, o, aquellos que incluyen clasificación de fuentes y estimación de parámetros físicos con mapas de auto-organización (Masters et al. 2017, 2019; Hemmati et al. 2019; Logan & Fotopoulou 2020).

En la otra rama, el aprendizaje supervisado requiere un set de entrenamiento utilizado para aprender las correlaciones subyacentes entre la entrada (*input*) presentado y el objetivo, diseñado para estimar las propiedades o etiquetas proporcionadas por el usuario, donde se debe tener un conocimiento perfecto de las etiquetas o las propiedades de referencia utilizadas como salidas (*outputs*) en el set de entrenamiento. En esta familia, uno puede observar, por ejemplo, las Redes Neuronales Artificiales (ANN; Artificial Neural Networks) que han sido adoptadas en su habilidad para clasificar fuentes puntuales versus fuentes extendidas en varios trabajos anteriores (Storrie-Lombardi et al. 1992; Lahav et al. 1995), con una aplicación muy notable del parámetro de estelaridad (CLASS_STAR) en

SEXTRACTOR, y su mejora en Soumagnac et al. (2015) para magnitudes más débiles. Otro aprendizaje supervisado popular, es la exploración de árboles de decisión, como el *Random Forest* y *Support Vector Machines* (e.g. para la clasificación de variables cataclísticas en grandes surveys espectroscópicos, o detectando candidatos a cuásar de datos fotométricos de multi-longitud de onda,

1.5. Motivación de esta Tesis

La principal motivación es evaluar la confiabilidad de las Herramientas de Aprendizaje Automatizado (de ahora en adelante, HAA) en la astronomía. Para esto, analizaremos el desempeño y la confiabilidad de estas herramientas en un caso de ambiente extremo : la Zona de Evitamiento. La ZoA es un laboratorio ideal para este tipo de estudios, pues en una región del cielo muy acotada ($|b| < 10^\circ$), es posible encontrar los mas diversos tipos de fuentes astronomicas. Las condiciones de esta región permiten generar todo tipo de información, desde fuentes perfectamente puntuales hasta otras que pueden confundir al mas avezado experto.

Para lograr identificar y analizar las problemáticas inherentes a la aplicación de HAA en la astronomía, definimos los siguientes objetivos:

- Evaluar la confiabilidad de las HAA en astronomía.
- Estudiar las posibles problemáticas y beneficios derivados del uso de HAA en astronomía.
- Estudiar el uso específico de HAA para caracterizar y clasificar estructuras en la ZoA

Como caso de estudio, analizaremos particularmente el trabajo de Zhang et al. (2021), donde aplican un algoritmo de clasificación en un relevamiento que abarca todo el cielo, incluyendo datos extragalácticos en la ZoA.

Capítulo 2

Datos

El presente análisis se apoya en dos conjuntos de datos. El conjunto principal incluye información en NIR y en rayos X, mientras que el conjunto secundario abarca catálogos que incorporan datos en el rango espectral radio, infrarrojo medio, NIR y óptico. La elección del conjunto de datos principal se fundamenta en la hipótesis de que tanto en rayos X como en NIR, la luz atraviesa la Vía Láctea, permitiéndonos así estudiar fuentes extragalácticas.

2.1. Datos Primarios

2.1.1. Imágenes y catálogos en infrarrojo cercano

El telescopio VISTA (Visible and Infrared Survey Telescope for Astronomy; Emerson et al. 2004) es un telescopio de 4.1 metros de apertura principal, perteneciente al Observatorio Europeo Austral (ESO) e instalado en el Observatorio del Cerro Paranal. Comenzó sus operaciones en diciembre de 2009 y fue diseñado para llevar a cabo relevamientos de gran escala en el cielo del hemisferio sur.

Inicialmente, el principal instrumento instalado en VISTA es la cámara de campo amplio VIRCAM (the Vista InfraRed CAMera; Emerson et al. 2006), con una resolución angular de 0.35 segundos de arco y un campo de visión que ofrece desde 1.1 a 1.5 grados de diámetro. VIRCAM cuenta con 5 filtros de banda ancha, Z ($0.87\text{ }\mu\text{m}$), Y ($1.02\text{ }\mu\text{m}$), J ($1.25\text{ }\mu\text{m}$), H ($1.64\text{ }\mu\text{m}$) y K_s ($2.14\text{ }\mu\text{m}$) y un filtro de banda estrecha a $1.18\text{ }\mu\text{m}$. Podemos observar las curvas de transmisión de cada filtro en la figura 2.1, junto a la curva de Eficiencia Cuántica del detector. La luz de galaxias en IR es dominada por poblaciones estelares viejas y rojas.

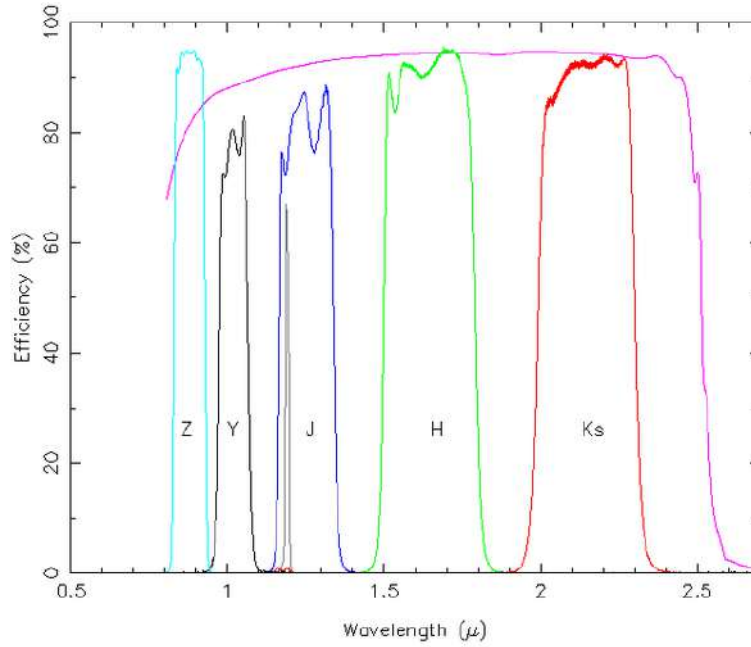


Figura 2.1: Curvas de transmisión para los filtros Z, Y, J, H y K_s del relevamiento VVV en la cámara VIRCAM (curvas de colores etiquetadas), junto a la curva de Eficiencia Cuántica (QE) del detector como una curva superpuesta de color violeta

VIRCAM cuenta con un arreglo de 16 detectores de 2048×2048 cada uno, quedando con un total de 67 megapíxeles otorgando 0.6 grados cuadrados de área de píxeles activa y una escala de píxel de 0.34 segundosdearco/píxel. Una imagen de 1.5×1.1 grados cuadrados es construida en base a la superposición de 6 exposiciones (conocidas como *pawprints*), construidas al seguir un patrón norte–norte–oeste–sur–sur. Este patrón de observación se ejecuta con la finalidad de llenar los vacíos entre los sensores de la cámara. La figura 2.2 muestra el movimiento de los *pawprints* para la creación de un área uniforme conocida como *tile*.

VVV

VISTA es un telescopio que trabaja en longitudes de onda infrarrojas y en su etapa original condujo cinco relevamientos principales, que han sido propuestos y seleccionados bajo un determinado objetivo científico (en la figura 2.3 se muestra la cobertura espacial de estos relevamientos), entre ellos se encuentra el relevamiento VVV (VISTA Variables in the Vía Láctea; Minniti et al. 2010), el cual es un relevamiento público de la ESO diseñado para estudiar la población estelar de las regiones centrales de nuestra Galaxia, gracias a la calidad de los datos obtenidos por VISTA+VIRCAM. Este relevamiento cubre

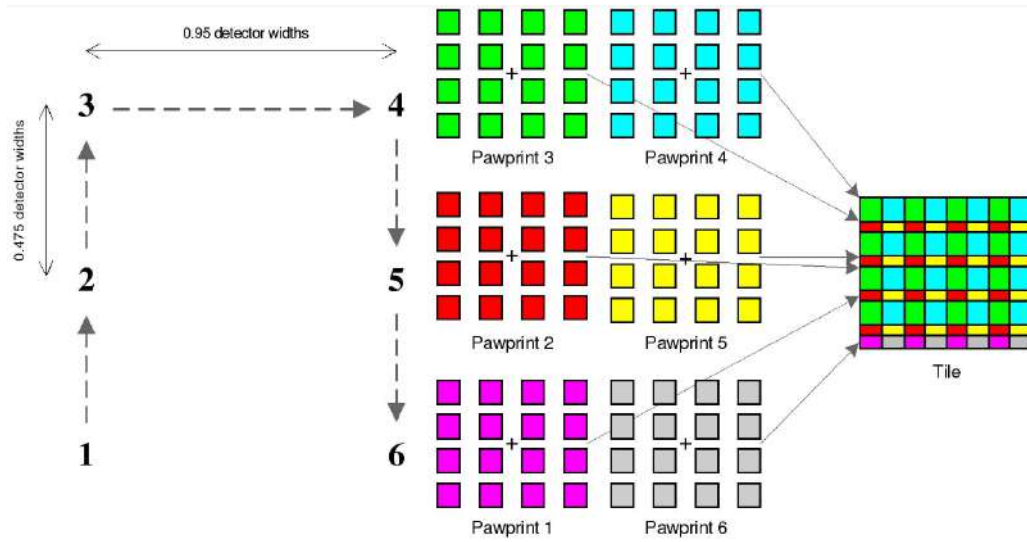


Figura 2.2: Construcción de un *tile* a partir del movimiento del telescopio en 6 diferentes *pawprint*

300 grados cuadrados del bulbo Galáctico ($-10^\circ \leq \ell \leq 10^\circ$ y $-10^\circ \leq b \leq 5^\circ$) y 220 grados cuadrados del disco ($295^\circ \leq \ell \leq 350^\circ$ y $-2^\circ \leq b \leq 2^\circ$), en la figura 2.4 se observa la cobertura del relevamiento en el bulbo galáctico y parte del disco fino. Cada *tile* es identificado por el prefijo “b” del 201 al 396 para los que pertenecen al bulbo y con el prefijo “d” del 001 al 152 para los que cubren el disco. En la tabla 2.1 se encuentran los valores fotométricos característicos para VVV, como el ancho máximo a media altura que presenta un determinado *peak* en transmisión (FWHM). En el caso de las imágenes astronómicas, corresponde a una estimación de la calidad de una imagen que mide cuánto el telescopio y la atmósfera han “difuminado” una fuente puntual, en este caso una estrella en varios píxeles del CCD. Además, la tabla muestra las magnitudes límites para cada banda de VIRCAM usada en VVV.

Banda	λ_{eff} (μm)	A_X/A_V	$A_X/E(B - V)$	FWHM	Magnitud límite (mag)
Z	0.878	0.499	1.542	0.12	21.3
Y	1.021	0.390	1.206	0.10	20.7
J	1.254	0.280	0.866	0.18	20.2
H	1.646	0.184	0.567	0.30	19.3
K_s	2.149	0.118	0.364	0.30	18.2

Tabla 2.1: Longitudes de onda efectivas para el conjunto de filtros VISTA utilizados en las observaciones del relevamiento VVV y la extinción relativa para cada filtro basada en la ley de extinción de Cardelli et al. (1989) (Catelan et al. 2011)

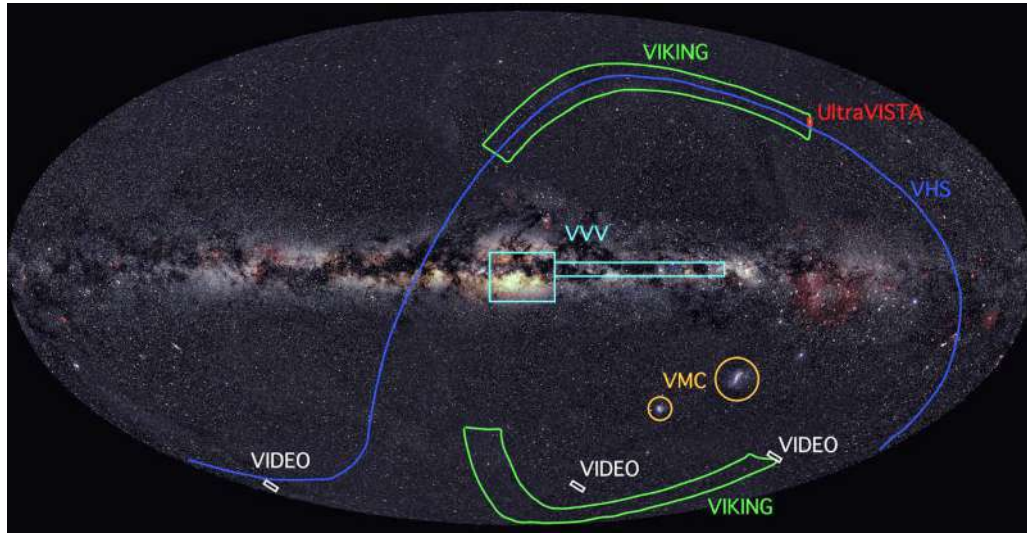


Figura 2.3: Esquema de los seis grandes relevamientos públicos por el telescopio VISTA. Estos cubren diferentes áreas del cielo en diferentes profundidades con un amplio rango de objetivos científicos.

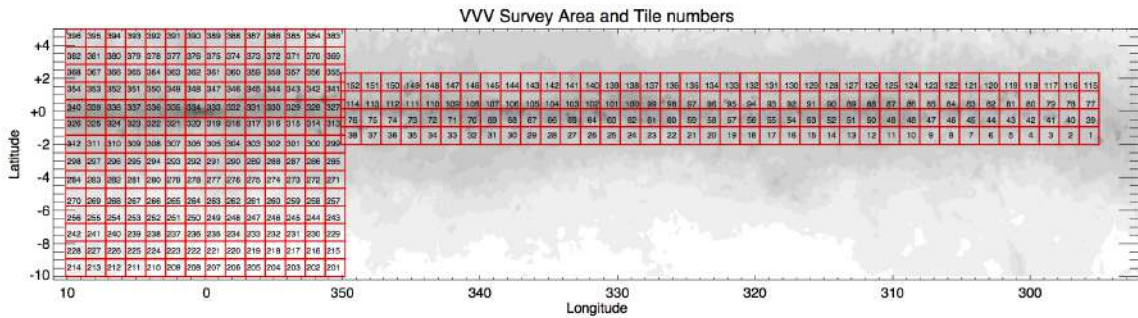


Figura 2.4: Esquema de la distribución de *tiles* para la cobertura del relevamiento VVV en coordenadas Galácticas.

Las imágenes obtenidas por VISTA+VIRCAM son transferidas al *Cambridge Astronomy Survey Unit* (CASU), unidad encargada del procesamiento y reducción de los datos (Irwin et al. 2004). Esto incluye la substracción de *dark*, correcciones de anomalías, aplanar el campo y substracción inicial del cielo.

Los catálogos generados, proveen imágenes pre-procesadas, donde se incluyen objetos detectados con sus respectivas posiciones, flujos y algunas medidas obtenidas utilizando diferentes aperturas. Cada objeto incluye además, un parámetro denominado “*flag*”, que indica la clasificación morfológica más probable, donde el valor $flag=-1$ corresponde a objetos estelares; $flag=+1$ a no estelares; $flag=-2$ a objetos compactos; $flag=-7$ a objetos con píxeles dañados; $flag=-9$ a objetos saturados; y $flag=0$ a ruido. Todos los detalles al respecto pueden encontrarse en (Saito et al. 2012).

Los flujos obtenidos en los catálogos fotométricos generados por CASU en las diferentes aperturas, son convertidos a magnitudes con la siguiente ecuación:

$$mag = ZP - 2.5 \log(flux/exptime) - apcor - percorr \quad (2.1)$$

Con ZP el punto cero, $apcor$ la corrección por apertura y $percorr$, la corrección basada en la oscuridad media del cielo. Todos estos valores y características fotométricas pueden encontrarse en los *header* de cada observación, disponibles en la base de datos de CASU¹.

Para el cálculo de la magnitud AB, se siguió el procedimiento de Hewett et al. (2006), usando el espectro de Vega de Bohlin & Gilliland (2004) y asumiendo una magnitud cero para todos los filtros. Para cada filtro, usaron el flujo constante (f_v), que da el mismo valor del flujo integrado en la banda que Vega. Por definición, el sistema de magnitud AB sigue la ecuación:

$$AB_v = -2.5 \log(f_v) - 48.60 \quad (2.2)$$

Con f_v se encuentra en unidades de ergs/s/cm²/Hz.

2.1.2. Catálogos de rayos X

Nuestro trabajo hace uso de datos provistos por el telescopio X-ray Multi-Mirror Mission (XMM-Newton; Jansen et al. 2001) de la ESA, lanzado en Diciembre de 1999. XMM-Newton está equipado con tres instrumentos, EPIC, OM y RGS. Con un campo de visión de 30 minutos de arco de diámetro, este telescopio proporciona observaciones en las bandas de energía blanda (0.2 - 2.0 keV), media (2.0 - 4.5 keV) y dura (4.5 - 12.0 keV), con resoluciones aproximadas de 6.0, 11.0 y 30.0 segundos de arco, respectivamente. Además, ofrece capacidades espectroscópica en el rango de energía de 0.15 a 15 keV.

2.2. Datos Secundarios

Adicionalmente, el trabajo utiliza datos provenientes de distintos observatorios en tierra y en el espacio, para complementar la información pancromática de ciertas fuentes.

¹<http://casu.ast.cam.ac.uk/vistasp/>

Radio

El Sydney University Molonglo Sky Survey (SUMSS; Bock et al. 1999) es un estudio de imágenes de radio en el cielo, enfocado a áreas con declinaciones $\delta < -30^\circ$, este relevamiento representa un hito en la radioastronomía debido a su capacidad para cartografiar el cielo en longitudes de onda de radio. SUMSS se encuentra ubicado en el Telescopio de Síntesis del Observatorio Molonglo (MOST), operando a un frecuencia central de 843 MHz y proporcionando una resolución angular efectiva de alrededor de 45 segundos de arco. Gracias a su alta resolución, ha sido capaz de detectar y caracterizar una amplia gama de fuentes de radio en el Universo. Estas observaciones han revelado la estructura de la radioemisión galáctica, brindando información sobre la formación de galaxias, cúmulos de galaxias y otros fenómenos astrofísicos a gran escala.

Infrarrojo Lejano

El Wide-field Infrared Survey Explorer (WISE; Wright et al. 2010) es relevamiento que lleva a cabo un estudio completo del cielo en infrarrojo medio y lejano, utilizando fotometría simultánea en cuatro filtros, W1 ($3.4\mu\text{m}$), W2 ($4.6\mu\text{m}$), W3 ($12\mu\text{m}$) and W4 ($22\mu\text{m}$), con una resolución aproximada de 6.1, 6.4, 6.5, y 12.0 segundos de arco, respectivamente. Esta misión ha permitido la identificación de galaxias con formación estelar intensa, galaxias con núcleos activos y asteroides en el Sistema Solar. Su capacidad para penetrar el polvo interestelar ha contribuido en la comprensión de la evolución estelar y galáctica.

Por otro lado, el Telescopio Espacial Spitzer (Werner et al. 2004) ha sido un pilar en la observación infrarroja, gracias a su Infrared Array Camera (IRAC; Fazio et al. 2004), que opera en las bandas de 3.6 (IRAC1), 4.5 (IRAC2), 5.8 (IRAC3), 8.0 (IRAC4) μm , con una resolución angular aproximada de 2.7, 2.8, 5.1 y 6.4 segundos de arco, respectivamente. Las observaciones de Spitzer han proporcionado una ventana sin igual al Universo oculto detrás del polvo interestelar, brindando información crucial para la comprensión de la evolución estelar y galáctica.

Infrarrojo Cercano

El Two Micron All-Sky Survey (2MASS; Skrutskie et al. 2006) es un relevamiento en infrarrojo que mapeó todo el cielo en las bandas de infrarrojo cercano, evitando muchos de los sesgos de observación que afectaron a relevamientos anteriores. Se llevó a cabo utilizando un par de cámaras y telescopios ecuatoriales cassegrain de 1.3 metros, idénticos

y totalmente dedicados, ubicados en el Observatorio Whipple en Mt. Hopkins, al sur de Tucson, Arizona, y en el Observatorio Interamericano Cerro Tololo, cerca de La Serena, Chile. El 2MASS observó el cielo en un sistema fotométrico en tres bandas infrarrojas: J ($1.25\ \mu\text{m}$), H ($1.65\ \mu\text{m}$) y K_s ($2.17\ \mu\text{m}$), y con una resolución angular de ~ 2.0 segundos de arco.

Los efectos de la extinción interestelar se reducen a diez veces en relación a la banda B, y la distribución de energía espectral (SED; del inglés *Spectral Energy Distribution*) de la mayoría de las galaxias alcanzan su punto máximo en longitudes de onda del NIR. Además, las luminosidades de la banda K son una excelente aproximación para calcular la masa bariónica, ya que la relación Masa-Luminosidad estelar es bastante constante en todos los tipos de galaxias en esta longitud de onda (*e.g.* dentro de un factor de dos; Bell & de Jong 2001). Esto hace que el NIR sea la región espectral elegida para mapear la distribución de la materia bariónica en el Universo cercano.

Óptico

El DECam Plane Survey (DECaPS; Schlafly et al. 2018) ha permitido un mapeo detallado del cielo visible en múltiples bandas ópticas mediante el uso de la Cámara de Energía Oscura (DECam; Flaugher et al. 2015), instalada en el Observatorio Interamericano Cerro Tololo. Con una resolución angular de aproximadamente 0.27 segundos de arco por píxel, DECaPS ha proporcionado imágenes de alta resolución con una amplia cobertura. Estas imágenes han sido valiosas en el estudio de la distribución tridimensional de cúmulos estelares, la identificación de estrellas variables en nuestra Galaxia y la caracterización de poblaciones estelares en diversas regiones.

El SuperCOSMOS H-alpha Survey (SHS; Parker et al. 2005) ha ampliado nuestra comprensión del hemisferio sur en el espectro óptico mediante observaciones en múltiples bandas espectrales, con una resolución angular típica de alrededor de 1.2 segundos de arco por píxel. SHS ha obtenido imágenes de alta calidad que abarcan desde el cielo austral hasta el ecuador celeste, permitiendo la identificación y caracterización de una amplia variedad de objetos; desde cúmulos abiertos y galaxias, hasta estrellas variables y asteroides. Las observaciones de SHS han enriquecido nuestra base de datos sobre poblaciones estelares en la Vía Láctea y han proporcionado una visión holística de la estructura del Universo en esta región.

La misión espacial Gaia (Gaia Collaboration et al. 2016) de la Agencia Espacial Europea (ESA) ha establecido nuevos estándares en la astrometría de precisión. Con su telesco-

pio con diseño astrométrico, Gaia ha obtenido mediciones con una resolución angular de alrededor de 0.0017 segundos de arco por píxel en su cámara de ciencia y una resolución de alrededor de 0.025 segundos de arco por píxel en su instrumento astrométrico. Esta alta precisión ha permitido la construcción de un catálogo tridimensional de más de mil millones de estrellas en la Vía Láctea y alrededores.

2.3. Herramientas

En esta sección describimos las herramientas utilizadas para obtener los catálogos e imágenes necesarios para lograr los objetivos de esta tesis.

2.3.1. Herramientas de base de datos e imágenes

Las herramientas usadas en la tesis para obtener catálogos e imágenes consisten en *Aladin Sky Atlas Desktop* (Bonnarel et al. 2000), que proporciona un atlas del cielo interactivo que permite visualizar imágenes astronómicas y relevamientos completos. *Aladin* también permite acceder a la base de datos de accediendo a la base de datos Simbad (*Set of Identifications, Measurements and Bibliography for Astronomical Data*) y VizieR para completar con información adicional, todo parte del servicio de CDS (Strasbourg astronomical Data Center²). También se utilizó el software TOPCAT (Taylor 2005), el cuál es una herramienta interactiva de visualización y edición gráfica para trabajar con datos tabulados. TOPCAT permite conectarse también a las bases de datos de CDS.

Además de estas herramientas, se utilizó el servicio VISTA Science Archive³ (VSA; Cross et al. 2012), donde se almacenan los productos de datos de imágenes y catálogos generados por VIRCAM en el Telescopio VISTA. También se empleó la calculadora de extinción galáctica de NASA/IPAC Extragalactic Database⁴ que permite calcular la extinción galáctica en función de las coordenadas astronómicas

2.3.2. Herramientas de Aprendizaje automatizado

En esta tesis estudiamos los resultados de la aplicación de diferentes algoritmos de aprendizaje automatico realizados por Zhang et al. (2021), de forma que en esta sección

²<https://cds.unistra.fr/>

³<http://vsa.roe.ac.uk/>

⁴https://ned.ipac.caltech.edu/extinction_calculator?in_csys=Galactic&in_equinox=J2000.0&obs_epoch=2000.0&lon=320.36868879452&lat=-1.2153122253902

describimos las principales características de estas herramientas, además del significado de las métricas.

Aún así, es importante aclarar que durante el desarrollo de esta tesis, no se aplicaron HAA. En contraste, nos dedicamos a estudiar las diferentes dificultades que surgen al momentos de clasificar fuentes, en especial en las zonas extremas como la ZoA.

Evaluar el rendimiento de un modelo entrenado es igualmente crucial tanto como la preparación de datos y el entrenamiento del modelo en sí. La utilización de diversas métricas para la evaluación del desempeño permite obtener una comprensión más sólida y precisa de la capacidad predictiva del modelo, incluso antes de implementarlo en la producción y aplicarlo a datos inéditos. La omisión de una evaluación exhaustiva del modelo de aprendizaje automático mediante diferentes métricas y la dependencia exclusiva en la precisión pueden ocasionar problemas al desplegar el modelo en datos no observados previamente, resultando en predicciones insatisfactorias. Esta problemática emerge porque, en ciertos casos, los modelos no aprenden verdaderamente, sino que memorizan patrones específicos. Como consecuencia, su habilidad para generalizar en contextos desconocidos se ve comprometida.

Las métricas de evaluación, diseñadas para valorar el desempeño de un modelo de aprendizaje automático, abarcan:

Matriz de confusión

Una matriz de confusión es una representación en forma de matriz que resume los resultados de las predicciones realizadas en una prueba binaria. Usualmente se emplea para describir el rendimiento de un modelo de clasificación al evaluarlo en un conjunto de datos de prueba cuyos valores reales se conocen previamente.

Para el caso de una matriz de confusión que involucra dos etiquetas de clase, existen cuatro posibles resultados de las predicciones, los cuales se clasifican según su concordancia con los valores reales:

- Verdadero Positivo (TP): Se presenta cuando el modelo predice correctamente que una instancia pertenece a la clase positiva, y en efecto, esta instancia es de la clase positiva.
- Verdadero Negativo (TN): Ocurre cuando el modelo predice de manera correcta que una instancia es de la clase negativa, y esta instancia realmente pertenece a dicha clase.

- Falso Positivo (FP): Se produce cuando el modelo predice erróneamente que una instancia pertenece a la clase positiva, pero en realidad es de la clase negativa. Esto corresponde a un error de tipo I.
- Falso Negativo (FN): Sucede cuando el modelo predice incorrectamente que una instancia es de la clase negativa, cuando en verdad pertenece a la clase positiva. Esto se denomina error de tipo II.

Estos elementos de la matriz de confusión proporcionan una visión detallada de cómo el modelo clasifica las instancias en función de sus verdaderas clases, permitiendo una evaluación más completa del desempeño del modelo en distintas métricas y contextos.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figura 2.5: Ejemplo de una Matriz de Confusión con 2 etiquetas

Accuracy

También llamado “Exactitud”, es una métrica comúnmente utilizada para evaluar problemas de clasificación. Representa el número de predicciones correctas en relación con el total de predicciones realizadas. Matemáticamente, se expresa como:

$$accuracy = \frac{TP + TN}{TOTAL}, \quad (2.3)$$

Donde TP es el número de verdaderos positivos, TN es el número de verdaderos negativos y $TOTAL$ es el total de instancias en la prueba. En términos más simples, el *accuracy* se puede entender con la pregunta: *¿Con qué frecuencia es correcto el clasificador?*.

Si bien el *accuracy* es una métrica útil, puede tener limitaciones en casos de desbalance de clases, es decir, cuando una clase tiene una representación mucho mayor que la otra.

Por ejemplo, si el 95 % de las instancias pertenecen a la Clase A y solo el 5 % a la Clase B, un clasificador que simplemente prediga la Clase A en todas las ocasiones obtendrá un alto *accuracy* del 95 %. Sin embargo, este alto valor de *accuracy* no necesariamente refleja un buen desempeño del modelo, ya que no está capturando las instancias de la Clase B de manera efectiva. En estos casos, otras métricas como la precisión, el recall y la puntuación-F1 son más adecuadas para evaluar el rendimiento del modelo.

Precision

En esta métrica, se suman todas las predicciones positivas vs las TP. Cuando la clasificación de clase está desequilibrada, la medida de *accuracy* por sí solo puede llevar a resultados no fiables, para evitar esto se utilizan más métricas en conjunto. Con la precisión es posible medir la calidad del modelo a clasificar mediante la ec.2.4

$$precision = \frac{TP}{TP + FP}, \quad (2.4)$$

Se puede entender al preguntar *¿Qué fracción de objetos clasificados como positivos son positivos?*

Recall

También llamada “Exhaustividad” o “Sensibilidad”, es una métrica importante en problemas de clasificación, especialmente cuando una de las clases es más relevante que la otra. Representa la proporción de instancias positivas que el modelo clasifica correctamente con respecto al total de instancias positivas reales. Matemáticamente, se expresa por la ec.2.5:

$$recall = \frac{TP}{TP + FN}, \quad (2.5)$$

El recall responde a la pregunta: *“De todas las instancias reales positivas, ¿Qué fracción predijo correctamente el modelo?”*.

El recall es especialmente útil en situaciones en las que la detección de instancias positivas es crucial y no queremos pasar por alto ninguna. Sin embargo, es importante considerar que, a medida que el recall aumenta, la posibilidad de tener más falsos positivos también aumenta. Por lo tanto, existe un compromiso entre recall y precisión, y es necesario encontrar el equilibrio adecuado según el contexto y los objetivos del problema de clasificación.

F1-score

Es una métrica que combina tanto la precisión como el recall en una sola medida, proporcionando una forma de evaluar el rendimiento del modelo teniendo en cuenta tanto los falsos positivos como los falsos negativos. Es especialmente útil en situaciones en las que el desequilibrio entre las clases es significativo, alcanzando su mejor valor en $F1\text{-score} = 1$, y el peor en un 0. Por lo tanto, el resultado de esta cantidad considera tanto la calidad como la cantidad de las clasificaciones positivas realizadas por el modelo. Se define por la ecuación:

$$F1 = \frac{2 \times \textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}}, \quad (2.6)$$

El $F1\text{-score}$ proporciona un balance entre la precisión y el recall, lo que significa que penaliza tanto los falsos positivos como los falsos negativos. Si un problema tiene clases desequilibradas y es importante evitar tanto los falsos positivos como los falsos negativos, el F1-score puede ser una métrica más apropiada para evaluar el rendimiento del modelo que la precisión o el recall por separado.

Además, es importante de entender los diferentes algoritmos de aprendizaje automático no-supervizado utilizados en Zhang et al. (2021), que corresponde a:

Random Forest

El algoritmo de *Random forest* es un ensamble de aprendizaje automático en donde se combinan árboles de decisión durante el entrenamiento, de modo que cada “árbol en el bosque” se construye utilizando una muestra *bootstrap*⁵ (Efron 1979) del conjunto de datos y un subconjunto aleatorio de características. Cada árbol depende de los valores de un vector aleatorio muestreado de forma independiente y con la misma distribución. Los parámetros de un bosque aleatorio son las variables y umbrales utilizados para dividir cada nodo durante el entrenamiento (Breiman 2001). La predicción final se obtiene combinando las predicciones de cada árbol individual. Por lo general, esto se hace contando cuántas veces se predice cada resultado para las tareas de clasificación. Se reconoce que el *Random Forest* es robusto, capaz de manejar conjuntos de datos de alta dimensión sin tender al sobreajuste de datos.

bootstrap es un método que estima la distribución de muestreo al tomar múltiples

⁵*Bootstrap* es un método estadístico que estima la distribución de la muestra, al tomar múltiples muestras con reemplazo de una sola muestra aleatoria. Estas nuevas muestras se denominan “muestras repetidas”. Cada una de las muestras tienen el mismo tamaño que la muestra original

muestras con reemplazo de una sola muestra aleatoria. Estas nuevas muestras se denominan muestras repetidas. Cada muestra tiene el mismo tamaño que la muestra original.

Rotation Forest

Rotation Forest (Rodríguez et al. 2006; Kuncheva & Rodríguez 2007) es una técnica de aprendizaje automático que combina las fortalezas de los árboles de decisión. Opera dividiendo el conjunto de datos en subconjuntos, aplicando una transformación Análisis de Componentes Principales PCA por sus siglas en inglés, (Pearson F.R.S. 1901) a cada subconjunto, y luego entrenando árboles de decisión en estos datos transformados. El objetivo principal es reducir la correlación entre los árboles de decisión individuales, mejorando así su poder predictivo general. *Rotatio Forest* emplea la aleatorización al utilizar un subconjunto aleatorio de características y un subconjunto aleatorio de muestras de datos para cada árbol en el conjunto, lo que ayuda a combatir el sobreajuste, mejorando la generalización.

LogitBoost

LogitBoost (Friedman et al. 2000) es un algoritmo de impulso diseñado principalmente para tareas de clasificación binaria. A diferencia de los métodos de impulso tradicionales como *AdaBoost* (Schapire 2013), que se centran en minimizar los errores de clasificación, *LogitBoost* minimiza la función de pérdida logística. Funciona entrenando secuencialmente una serie de clasificadores débiles, donde cada nuevo clasificador se entrena para corregir los errores cometidos por los anteriores. *LogitBoost* asigna diferentes pesos a las muestras de entrenamiento para priorizar los ejemplos más desafiantes, lo que lo hace particularmente efectivo al tratar con conjuntos de datos desequilibrados. Este algoritmo se basa en el aumento de gradiente y puede optimizar una amplia gama de funciones de pérdida diferenciables.

Capítulo 3

Comparación de clasificaciones por HAA y métodos no automáticos

3.1. Herramientas de aprendizaje automático aplicados a la Zona de Evitamiento

La ZoA presenta un desafío en la clasificación de objetos extragalácticos mediante HAA, por esto, existen pocos trabajos en esta zona aplicando métodos automáticos. Esta situación se debe principalmente a la calidad de los datos que existen, que no permiten un buen análisis. Dentro de los estudios, podemos destacar el trabajo de Vavilova et al. (2018), en donde generan una distribución de galaxias sintéticas para así recuperar la distribución espacial de estas a bajas latitudes. A este algoritmo le llamaron “*algorithm of darning the ZoA*”. Singularmente, esta herramienta sirve para hacer comparaciones entre datos reales de la ZoA y estas simulaciones. De esta manera, podemos construirnos una idea del potencial de los datos de futuros telescopios. Otro trabajo es el de Jones et al. (2019) donde usan herramientas de aprendizaje profundo para evaluar la eficiencia y el *accuracy* de su algoritmo en la identificación automática de galaxias en la ZoA usando datos NIR de los telescopios VISTA y UKIDSS.

Por último, en el reciente trabajo de Daza-Perilla et al. (2023), donde aplican HAA en el disco del relevamiento VVV usando la clasificación ya realizada para el VVV NIRGC, se aplicaron métodos de aprendizaje automático supervisado y no-supervisado para separar galaxias de no-galaxias en el disco norte del relevamiento VVVX ($10^\circ < \ell < 20^\circ$, $-4.5^\circ < b < +4.5^\circ$). Esta elección realizada debido a la similitud en las regiones del disco

sur del relevamiento VVV y del disco norte del relevamiento del VVVX, principalmente por los niveles de extinción. Así, a partir de una muestra de 5,509 galaxias (las cuales habían sido inspeccionadas visualmente) del catálogo VVV NIRGC y 74,238 no-galaxias, se generaron 2 muestras, una basada en imágenes (denominada IS) y la otra en su información fotométrica (llamada PS). Estos 2 enfoques fueron por: (i) la posible dificultad de adquirir ambos tipos de información para una fuente concreta; (ii) la doble confirmación de la clasificación de modelos supervisados de aprendizaje automático entrenados con información de imágenes e información de objetos físicos; (iii) el costo computacional en la clasificación; y (iv) la estimación de la calidad de los resultados cuando se utilizan datos de imágenes, fotometría o ambos.

El trabajo de Zhang et al. (2021) (Z21 de aquí en adelante) es el único trabajo existente respecto a la aplicación de HAA para la clasificación de elementos, que considera la Zona de Evitamiento. Los autores utilizan como base el catálogo 4XMM-DR9 (Webb et al. 2020), compuesto por 550,124 fuentes en 1,152 grados cuadrados en todo el cielo para clasificar los objetos, el catálogo final cuenta con 43,956 estrellas, 191,528 galaxias y 314,640 cuásares (QSO) usando rayos X.

3.2. VVV NIRGC

El Catálogo de Galaxias en el Infrarrojo Cercano del VVV (VVV NIRGC; Baravalle et al. 2021), es el más profundo y completo catálogo de fuentes extragalácticas en la ZoA hecho hasta ahora. El catálogo original esta compuesto por 5,563 galaxias, y es hasta ahora el que ha llegado a menor latitud y longitud galáctica (entre $295^\circ < \ell < 350^\circ$ y $2^\circ < b < 2^\circ$). Gracias al poder del telescopio VISTA, el catálogo VVV NIRGC dispone de galaxias tan débiles como $K_s \sim 17$, en regiones con extinciones $A_{K_s} = 1.423$.

Considerando el potencial de los rayos X para detectar galaxias a traves de la Vía Láctea, estudiamos el comportamiento del catálogo de Z21 en el área de cobertura del VVV NIRGC, encontrando que este presenta mas de 4000 galaxias que el VVV NIRGC. **¿Es posible que no fuesen detectadas estas galaxias por un telescopio tan potente como el VISTA? ¿Es posible que las HAA hayan generado la más grande base de datos de la ZoA hasta ahora descubierta?**

Al analizar la muestra de Z21 descubrimos que los autores diseñaron una estrategia de clasificacion que es crítico analizar en detalle.

1. Realizan un cruce entre los 550,123 datos del 4XMM con datos fotométricos provenientes del Sloan Digital Sky Survey (Sloan Digital Sky Survey; York et al. 2000) DR12 y el relevamiento WISE (WISE; Wright et al. 2010), generando así las muestras *XMM-SDSS* y *XMM-WISE*, además de una muestra combinada *XMM-WISE-SDSS*.
2. A continuación, realizan un cruce entre la muestra combinada *XMM-WISE-SDSS* con SDSS DR16 y LAMOST DR5 (Large Sky Area Multi-object Fiber Spectroscopic Telescope; Cui et al. 2012; Luo et al. 2015) para obtener su clasificación espectral conocida, necesaria para el entrenamiento del algoritmo de clasificación.
3. Obtienen cuatro muestras M_x , M_{xo} , M_{xi} y M_{xio} , dependiendo de su combinación de datos, sólo rayos X, rayos X-óptico, rayos X-infrarrojo y rayos X-infrarrojo-óptico.
4. Mediante el software WEKA (Waikato Environment for Knowledge Analysis; Frank et al. 2005), aplicaron los tres algoritmos de aprendizaje automático, mencionados anteriormente sobre cada muestra, seleccionando la combinación algoritmo-muestra con el mayor *accuracy* como resultado final. Los algoritmos seleccionados fueron *Rotation Forest*, con un *accuracy* del 77.80 % para M_x , *Random Forest* con un 89.42% para M_{xi} y *LogitBoost* con un *accuracy* del 92.82 % y 94.26 % para M_{xo} y M_{xio} , respectivamente.
5. La clasificación final de los objetos se basa en el valor de la probabilidad P_x , P_{xi} , P_{xo} y P_{xio} para ser clasificado como estrella, galaxia o QSO.

Examinando los datos de Z21, encontramos que existen 15,423 objetos de Z21 en la región cubierta por relevamiento VVV en el disco de Vía Láctea. Estos objetos se encuentran divididos en 1,666 estrellas (10.80 % del total), 4,031 QSOs (20.14 %) y 9,726 galaxias (63.06 %). Entre ellos, menos del 30 % de los objetos clasificados como galaxias poseen contraparte en el infrarrojo, mientras que no existen galaxias en la zona con información en el óptico. También agregar que de la sub-muestra de M_{xi} en el disco sur, el 82.79 % están clasificadas como estrellas con un $P_{xi} = 1$. Dado que no se encuentran galaxias que posean contraparte en el óptico, y un número bajo de ellas tienen contraparte en el infrarrojo, decidimos utilizar la clasificación considerando solo rayos X. De ahora en adelante nos referiremos a la muestra de 9,726 galaxias en el disco del relevamiento VVV como la muestra *galXray*. Esta muestra posee una mediana de probabilidad de 0.664. Además

podemos notar un desbalance importante entre la cantidad de estrellas clasificadas y las galaxias. Nuestro objetivo principal en este trabajo es detectar estas galaxias con emisión en rayos X, en las bandas del infrarrojo cercano (NIR) del relevamiento VVV.

3.3. Estadística de la muestra

Antes de proceder con el análisis detallado de las fuentes, llevamos a cabo un estudio del radio de cruce óptimo para los diferentes catálogos. Utilizando `topcat`, cruzamos la muestra *galXray* con distintos catálogos infrarrojos y ópticos con detecciones en el área de estudio. Estos catálogos fueron 2MASS, Gaia EDR3, VVV DR2, GLIMPSE y AllWISE.

Realizamos cruces utilizando radios de: $x_n = 0.1 \cdot n$ para n de 1 a 15, y $x_n = 2.0 + 0.5 \cdot (n - 1)$ con n de 1 a 13, para cada catálogo, donde x_n corresponde al radio de cruce en segundos de arco. Posteriormente, generamos histogramas acumulativos normalizados para cada conjunto de datos (ver figura 3.1). Destaca que, al aumentar el radio de cruce con AllWISE, la cantidad de fuentes cruzadas continuó incrementando. En contraste, observamos que para VVV, GLIMPSE, Gaia y 2MASS, la cantidad de fuentes cruzadas aumenta hasta cierto límite, identificando un límite superior adecuado entre 1.3 a 3 segundos de arco en el caso de VVV.

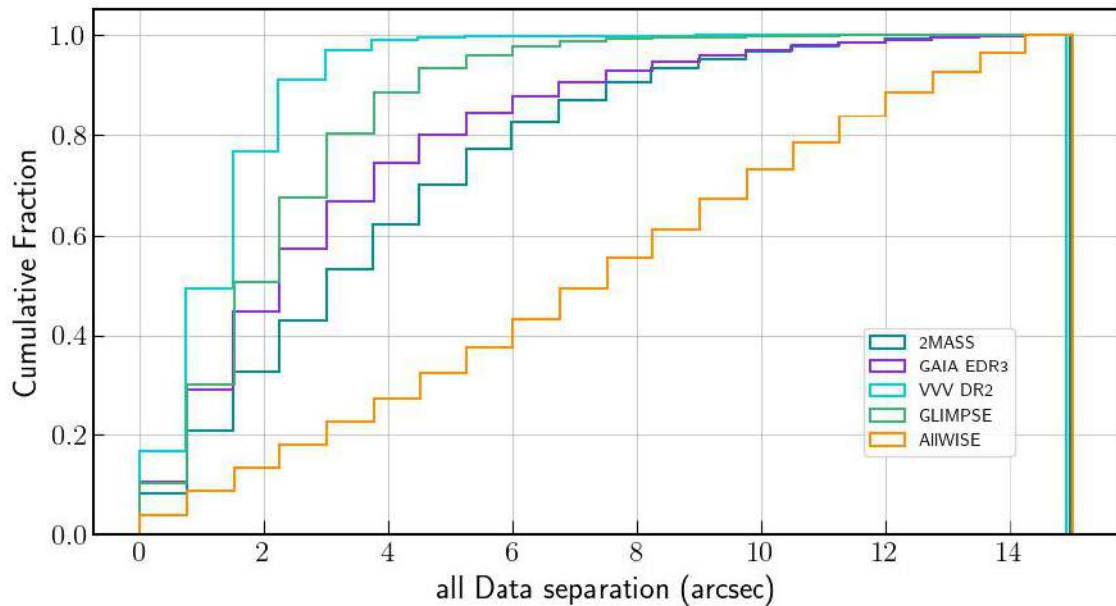


Figura 3.1: Histograma acumulativo para la cantidad de datos obtenidos para cada separación de cada cruce realizado con `topcat`.

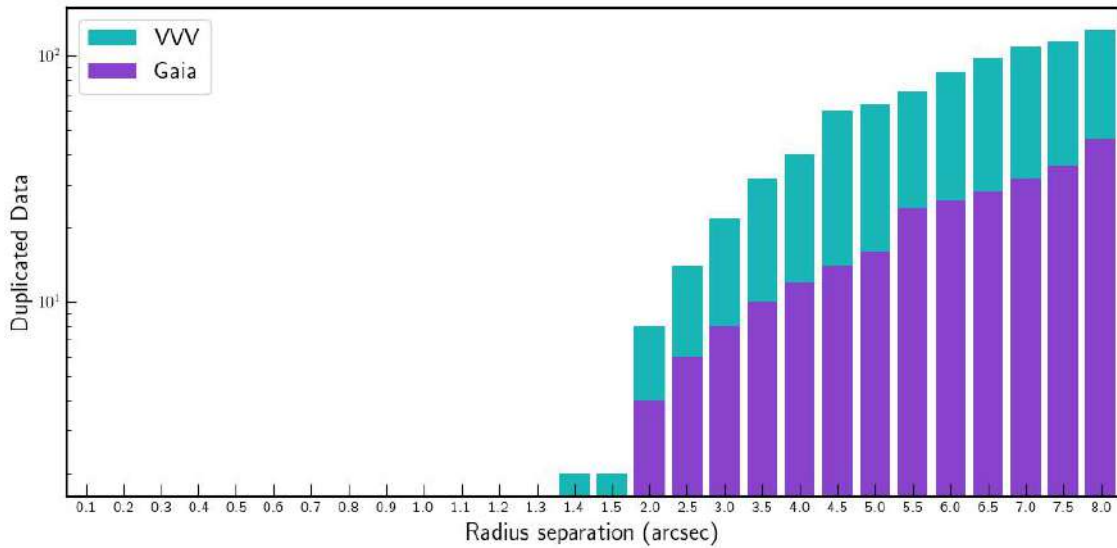


Figura 3.2: Histograma acumulativo con la cantidad de datos duplicados para el cruce entre Z21_disc/GAIA y Z21_disc/VVV

Posteriormente, se analizó la cantidad de datos duplicados para cada radio de cruce y catálogo. Esto corresponde a datos con el mismo ID del catálogo 4XMM-DR9, que se encuentran a una distancia dentro del radio de cruce para cada fuente. Se observó que no hubo duplicados en los cruces con 2MASS, GLIMPSE y ALLWISE. En cambio, para los cruces con VVV, los duplicados comenzaron a aparecer después de 1.3 segundos de arco, y en el caso de Gaia, a partir de 2.0 segundos de arco (ver figura 3.2). Es relevante señalar que para los relevamientos con mayor resolución y precisión, se identificaron duplicados a radios mucho más bajos, mientras que en el caso de ALLWISE, con una resolución inferior, no se encontraron datos duplicados. Podemos encontrar un resumen del total de datos en cada cruce, con sus respectivos duplicados en la tabla 3.1.

Utilizaremos un radio de cruce de 1.3 segundos de arco, elegido para mantener el radio con la mayor cantidad de cruce evitando a la vez, duplicados. En la literatura, es común encontrar cruces para fuentes extragalácticas con unos pocos arcosegundos (Marrese et al. 2019). Además, al tratarse de objetos extragalácticos, la separación angular entre la fuente en distintos catálogos puede ser menor en comparación con las fuentes de naturaleza galáctica.

Encontrado el radio de cruce óptimo para el cruce de datos del telescopio XMM-Newton con el relevamiento VVV, procedimos a comparar la muestra *galXray* con el catálogo VVV NIRGC. Este cruce resultó en cuatro galaxias en común, con posiciones inferiores a 2 segundos de arco. En contraste, el VVV NIRGC comparte 45 galaxias con

n	Radio	Total de cruce	Duplicados	n	Radio	Total de cruce	Duplicados
1	0.1	165	0	1	2.0	29,382	8
2	0.2	653	0	2	2.5	35,200	14
3	0.3	1,366	0	3	3.0	38,717	22
4	0.4	2,340	0	4	3.5	40,578	32
5	0.5	3,577	0	5	4.0	41,453	40
6	0.6	4,945	0	6	4.5	41,914	60
7	0.7	6,432	0	7	5.0	42,110	64
8	0.8	8,054	0	8	5.5	42,199	72
9	0.9	9,749	0	9	6.0	42,245	86
10	1.0	11,562	0	10	6.5	42,272	98
11	1.1	13,472	0	11	7.0	42,293	110
12	1.2	15,314	0	12	7.5	42,304	116
13	1.3	17,235	0	13	8.0	42,317	128
14	1.4	19,172	2	14	15.0	42,418	288
15	1.5	21,084	2				

Tabla 3.1: Tabla con con la cantidad de datos del cada cruce, con sus respectiva cantidad de datos duplicados.

otros autores (Schröder et al. 2007; Williams et al. 2014; Said et al. 2016; Schröder et al. 2019, ver Baravalle et al. (2018), sección 2.2). Por otro lado, la muestra *galXray* solo tiene una galaxia en común con Schröder et al. (2007), que corresponde a la galaxia DZOA 4653–11 (J134736.00–603703.8) con una velocidad radial de CMB de 4041 ± 86 km/s (Radburn-Smith et al. 2006) y una probabilidad P_x de 0.552 de ser galaxia por Z21. La figura 3.3 muestra las cuatro galaxias en común con el VVV NIRGC en la banda K_s del VVV, con un tamaño de $1' \times 1'$. En los paneles superiores, se observa la naturaleza de tipo temprano de estas galaxias, con P_x de 0.575 y 0.561, mientras que los paneles inferiores revelan su estructura de bulbo de galaxias de tipo tardío, con P_x de 0.765 y 0.778, respectivamente.

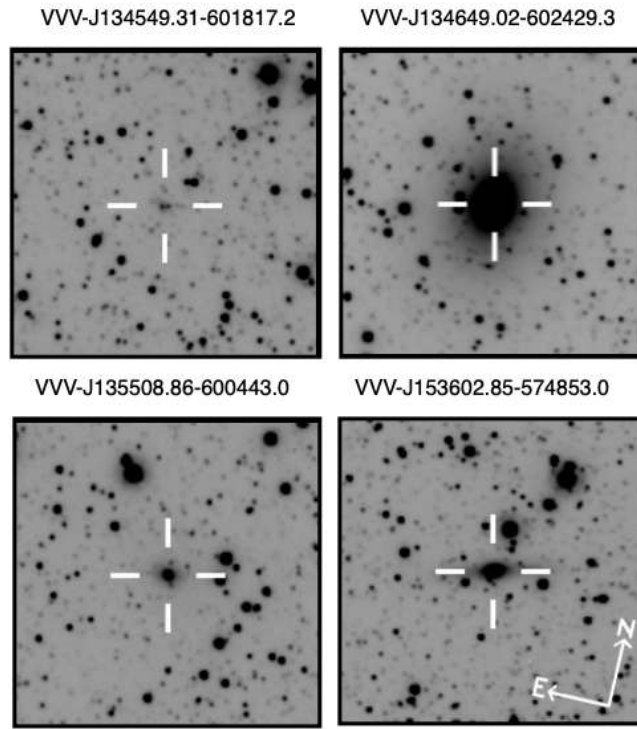


Figura 3.3: Galaxias en común entre el catálogo VVV NIRGC y la muestra *galXray*. Las imágenes de las galaxias fueron obtenidas en la banda K_S del relevamiento VVV, con un tamaño de $1' \times 1'$.

Además de esto, se llevó a cabo un cruce entre la muestra *galXray* y la salida original de SExtractor+PSFEx usada para generar el catálogo del VVV NIRGC, con el objetivo de examinar la morfología de las fuentes. Se identificaron 3,229 fuentes en común, las cuáles fueron inspeccionadas visualmente para Baravalle et al. (2021) (que representa el 33.29 % del total de *galXray*). De estas fuentes, 3,225 son de naturaleza galáctica, mientras que las 4 restantes corresponden a las galaxias mostradas en la figura 3.3. De las fuentes galácticas, 3,183 son fuentes puntuales y 64 son extendidas, principalmente constituidas por nubes de gas.

La figura 3.4 muestra la distribución de galaxias dentro del disco sur del relevamiento VVV correspondientes a nuestra muestra *galXray*, así como las cuatro galaxias en común con el catálogo VVV NIRGC y las galaxias confirmadas visualmente del mismo catálogo. Además, se muestran los contornos de extinción A_V derivados de los mapas de extinción de Schlafly & Finkbeiner (2011a). En general, la muestra *galXray* esta distribuida de manera uniforme en las regiones internas del disco y muestra altas extinciones a latitudes galácticas muy bajas. Por otro lado, el VVV NIRGC se encuentra en latitudes más altas, y no existe superposición entre los diferentes relevamientos debido a las altas extinciones

interestelares y las diferentes longitudes de onda de cada instrumento.

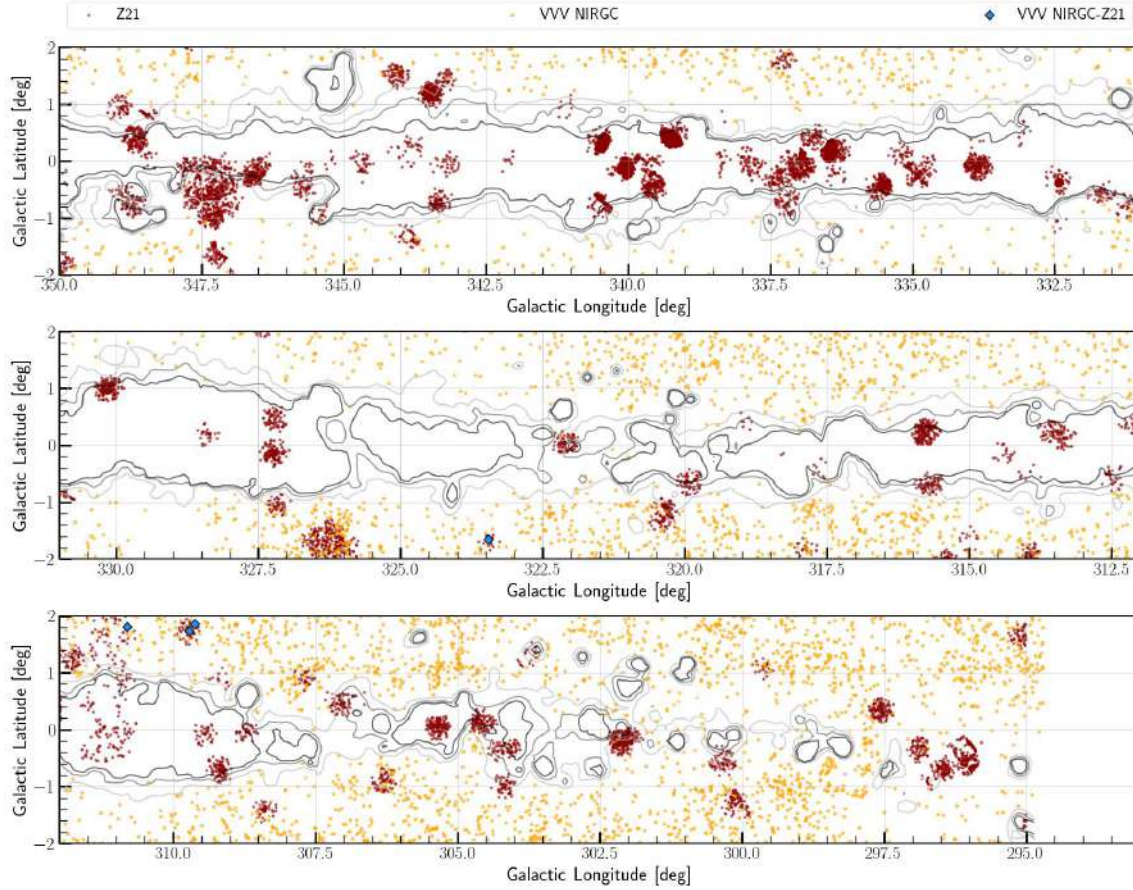


Figura 3.4: Distribución de galaxias de Z21 en el disco sur del VVV. las galaxias de la muestra *galXray* están representadas con puntos rojos. Las galaxias confirmadas visualmente por VVV NIRGC están en naranja

Finalmente, las 6,497 fuentes restantes fuera del cruce con el catálogo de SExtractor+PSFEx y que representan el 66.80 % de la muestra *galXray*, fueron sometidas a un cruce con todos los catálogos disponibles en la opción “*Sky crossmatch against remote table from the CDS Vizier or SIMBAD services*” de TOPCAT. Así, identificamos 732 fuentes en el catálogo del relevamiento VVV DR2. Sin embargo, al examinarlas visualmente, comprobamos que se tratan de objetos brillantes, específicamente estrellas, que no fueron incluidos en la salida de SExtractor+PSFEx debido a límites de brillo establecidos en el código. En el caso de las 5,765 fuentes restantes, el cruce no arrojó ningún resultado importante. Además, estas fuentes carecen de contrapartes en todas las longitudes de onda. También se buscó en diferentes catálogos accesibles en internet como NOIRLab Source

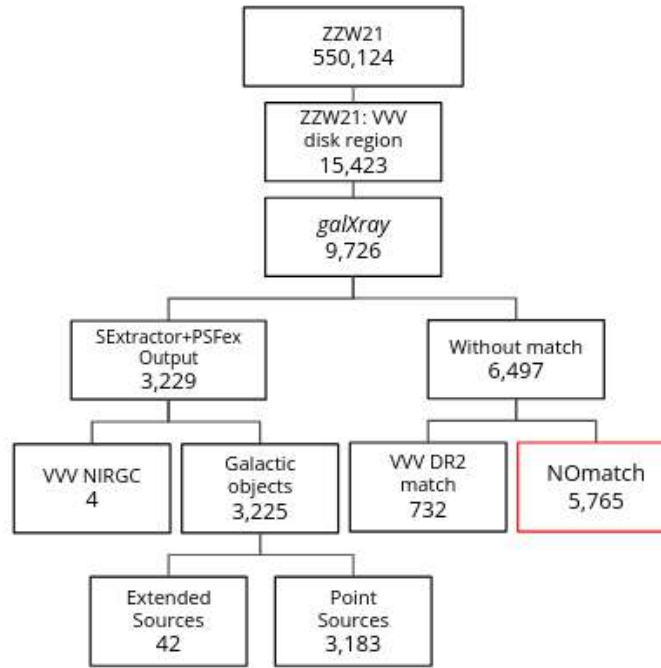


Figura 3.5: Diagrama de clasificación que ilustra la selección de diversas muestras y submuestras estudiadas, con especial énfasis en la muestra *NOMatch*, la cual constituye nuestra mayor preocupación en este trabajo.

Catalog¹ y SkyView². A lo largo de este trabajo, nos referimos a estas 5,765 fuentes como la muestra *NOMatch*.

La figura 3.5 resume el procedimiento adoptado para seleccionar la muestra final y así identificar las diferentes fuentes que forman parte de las muestras *galXray* y *NOMatch* de galaxias, motivos principales en este estudio de Z21.

Con el objetivo de depurar la muestra de estrellas, realizamos un cruce entre los 15,423 objetos en la región del disco del relevamiento VVV y el catálogo de Gaia DR3 *Astrophysical Params* (Creevey et al. 2023). Sin embargo, a pesar de los resultados obtenidos, decidimos descartar esta limpieza de estrellas y trabajar con la muestra original de 15,423 objetos.

La tabla 3.2 resume los resultados de este cruce. 1,953 *galXray*, 1,706 QSO y 862 estrellas. Al analizar la probabilidad de que una fuente sea un QSO o una galaxia, observamos que el 99.027 % de las *galXray* con coincidencia en Gaia tienen una probabilidad $P_{\text{star}} > 99\%$ para ser una estrella, sin encontrarse *galXray* con $P_{\text{gals}} > 50\%$ o $P_{\text{qso}} > 50\%$. Respecto a las estrellas clasificadas por Z21 con coincidencia en Gaia, el 94.199 % tie-

¹<https://datalab.noirlab.edu/nscdr2/index.php>

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

ne $P_{\text{star}} > 99\%$ y solo una estrella tiene $P_{\text{gals}} > 50\%$ para ser considerada una galaxia. En cuanto a los QSOs con coincidencia en Gaia, encontramos que el 97.831 % tiene $P_{\text{star}} > 99\%$, mientras que 3 QSOs tienen $P_{\text{gals}} > 50\%$ y solo 6 $P_{\text{qso}} > 50\%$. Este cruce nos permitió depurar la muestra a 11,008 fuentes sin confirmación, lo que representa el 71.37 % de las fuentes clasificadas por Z21 en la región de estudio que no se encuentran en el catálogo de Gaia DR3.

Muestra (Cruce)	Probabilidad de Gaia	Cantidad del fuentes correspondiente
Estrellas (862)	$P_{\text{star}} > 99\%$	94.199 %
	$P_{\text{gal}} > 50\%$	1
	$P_{\text{qso}} > 50\%$	0
<i>galXray</i> (1,953)	$P_{\text{star}} > 99\%$	99.027 %
	$P_{\text{gal}} > 50\%$	0
	$P_{\text{qso}} > 50\%$	0
QSO (1,706)	$P_{\text{star}} > 99\%$	97.831 %
	$P_{\text{gal}} > 50\%$	3
	$P_{\text{qso}} > 50\%$	6

Tabla 3.2: tabla con el resultado del cruce de Z21 en disco del relevamiento VVV con Gaia DR3.

Capítulo 4

Inspección Visual

Debido a las anomalías de los resultados decidimos realizar una inspección visual de todas las fuentes de la muestra *NOMatch*. Usamos el modo `MULTIGETIMAGE` de VSA (ver fig 4.1) para descargar las imágenes del catálogo ingresado en formato ASCII en coordenadas ecuatoriales de cada objeto, en las cinco bandas del relevamiento VVV. Para esta inspección se ingresó un tamaño de imagen de 0.5×0.5 minutos de arco, a la que llamaremos *stamp*. La inspección visual implicó observar las diferencias en la luminosidad superficial de las fuentes en las cinco bandas, procedimiento que se describe detalladamente en Baravalle et al. (2018) para la identificación y clasificación de galaxias. Este método señala que los objetos estelares muestran una luminosidad superficial comparable en todas las bandas del VVV, mientras que las fuentes extendidas poseen una mayor luminosidad superficial en longitudes de onda más largas (J, H y K_s), pero son débiles o apenas visibles en longitudes de onda más cortas (Z e Y). En primera instancia, realizamos una inspección visual superficial de cada fuente, identificando ciertas similitudes en los diferentes *stamps*. Posteriormente, se llevó a cabo una inspección visual más detallada para cada una de las fuentes en *NOMatch*, logrando clasificarlas en 10 categorías no excluyentes entre sí, considerando tanto la fuente central como el entorno en el *stamp* y permitiendo que una fuente pueda tener múltiples características. Finalmente, se realizó una nueva inspección visual para clasificar cada fuente con su característica más representativas. Las categorías definidas son listadas en la tabla 4.1 junto al número de *stamps* clasificados en cada una. El detalle de cada categoría será explicado en la subsección siguiente. El propósito de esta clasificación fue validar la clasificación de galaxias propuesta por Z21.

Select the programme/survey you wish to use: VVV: VISTA Variables in the Via Lactea

Database release to use: VVDR5

Source waveband	Z - ☒ Y - ☒ J - ☒ H - ☒ Ks - ☒	
ID column	no	optional, used to specify the frame to be used
Size (arcmin)	0.5	max 12 arcmin
Email:	pamela.marchant@userena.cl	
confirm Email:	pamela.marchant@userena.cl	
File of coordinates:	Browse... sin_contraparte_5.cat	max 500 lines, max total area 500 sq. arcmin e.g. 500 1x1 or 20 5x5 images
cross hair:	yes	mark the position on the jpeg image?
Mode:	standard	

Submit Reset

Figura 4.1: Modo MULTIGETIMAGE de VSA para obtener imágenes visuales de los relevamientos del telescopio VISTA, desde catálogos con coordenadas ecuatoriales.

Clasificación en NIR	# de Fuentes	% correspondiente al total
Región normalmente abarrotada	3,243	56.25 %
Estrella central	1,008	17.48 %
Región central vacía	971	16.84 %
Grupo estelar	266	4.65 %
Estrella brillante saturada	153	2.61 %
<i>Spikes</i>	48	0.83 %
Normal SFR	29	0.50 %
PUVS	22	0.38 %
Asociación de estrellas masivas	13	0.23 %
PUVSFR	12	0.21 %
Casos especiales	5	0.03 %

Tabla 4.1: Tabla con las clasificaciones de las fuentes Sin contraparte, con el respectivo total de *stamps* clasificados para cada categoría. La sigla PUVS se refiere a “Photometrically Ultra Variable Star”, mientras que PUVSFR “Photometrically Ultra Variable Star Forming Region”.

4.1. Descripción de clasificación

Aquí proporcionaremos una descripción de las diversas características encontradas en las imágenes NIR de la región del disco sur, ordenadas según el número de casos.

1. Región Normalmente Abarrotada

Esta categoría abarca la mayoría de los casos, representando un 56.25 % del total de los casos analizados. Se caracteriza por la presencia de estrellas distribuidas por todo el campo, con variaciones mínimas entre cada filtro. Si bien es posible observar ciertos niveles de extinción en las bandas mas azules, no se percibe un cambio drástico en el número de estrellas detectadas al llegar a la banda K_s . No se detectan estructuras extendidas en esta clasificación (ver figura 4.2).

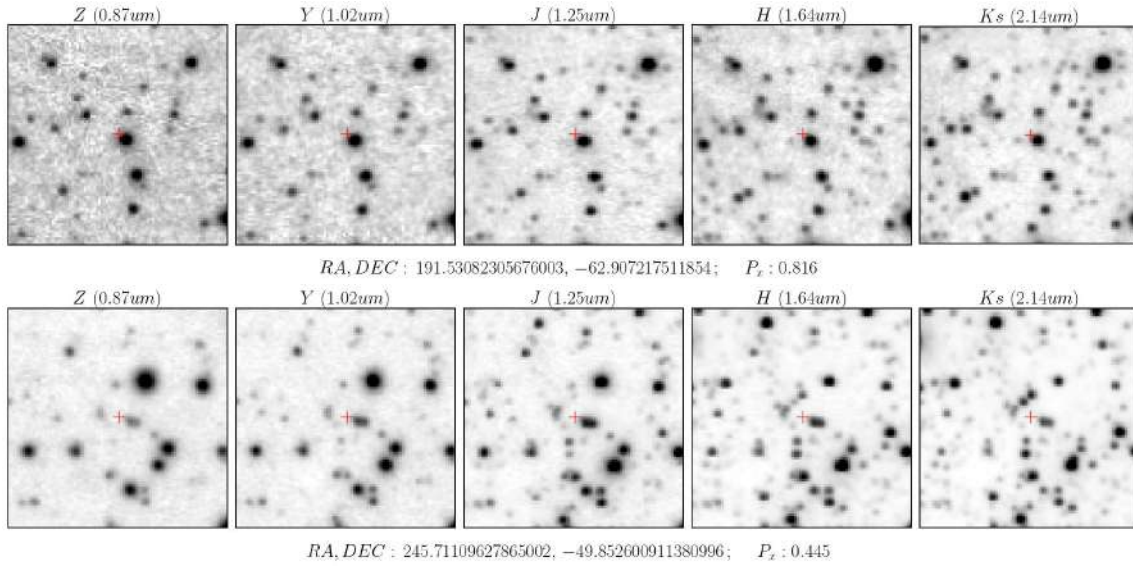


Figura 4.2: Ejemplos para stamps categorizados como “región normalmente abarrotada”, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.816$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.445$.

2. Estrella Central

Este caso, que comprende un 17.48 % implica la presencia de una estrella central en el campo de estudio, ya sea pequeña, tenue o grande. Esta categoría no es excluyente, por lo que es posible observar casos con estrellas saturadas, *spikes* o grupos estelares. En la figura 4.3, se muestra una estrella débil en el centro, con una alta probabilidad de ser una galaxia en rayos X, mientras que en el panel inferior se presenta una estrella central más grande y brillante, con una baja probabilidad (P_x) de ser una galaxia en rayos X.

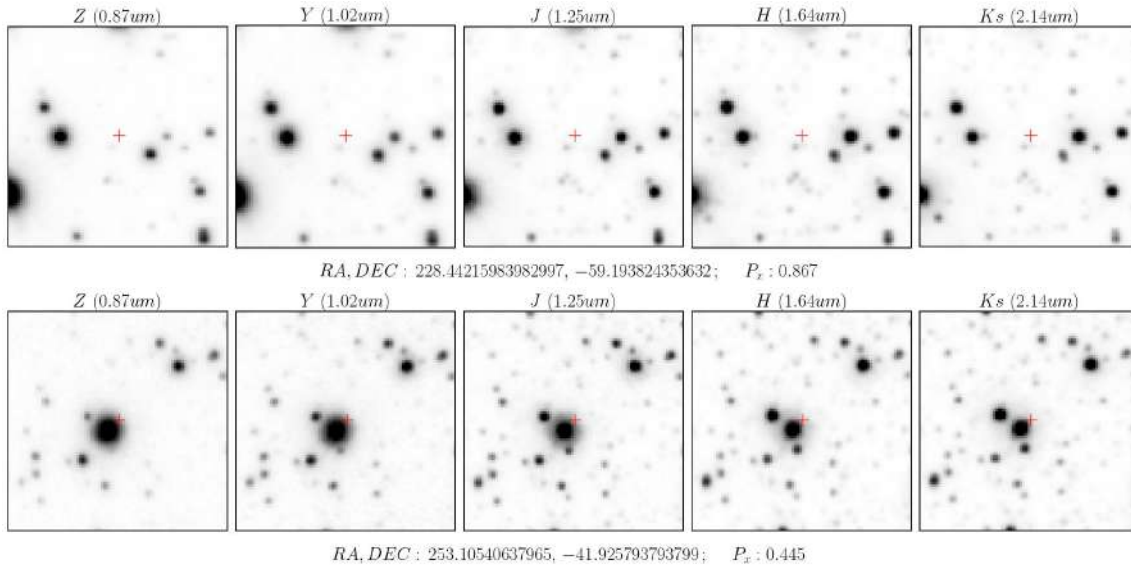


Figura 4.3: Ejemplos para un campo con una estrella central, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.867$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.445$.

3. Región Central Vacía

Constituyendo el 16.84 % de las fuentes clasificadas, esta categoría representa una región “aparentemente vacía” en el centro del cuadro, donde no se observan estrellas cercanas ni cambios que indiquen la presencia de una galaxia a través del polvo en el NIR. Es importante destacar que el hecho de que la fuente se clasifique como centro vacío, no significa necesariamente que esté vacía, sino que no se puede detectar ningún objeto en ese punto con el telescopio VISTA. Un ejemplo de esta categoría se puede encontrar en la figura 4.4.

4. Grupo Estelar

Esta categoría muestra un conjunto o asociación de estrellas, cercanas a la parte central del *stamp*, pero muy juntas entre si. Constituye el 4.65 % de los casos observados. En algunas bandas, este conjunto de estrellas puede ser fácilmente confundido con una fuente extragaláctica debido a la proximidad entre las esestrellas y los niveles de extinción estelar. Ejemplos de este caso se ven reflejados en la figura 4.5.

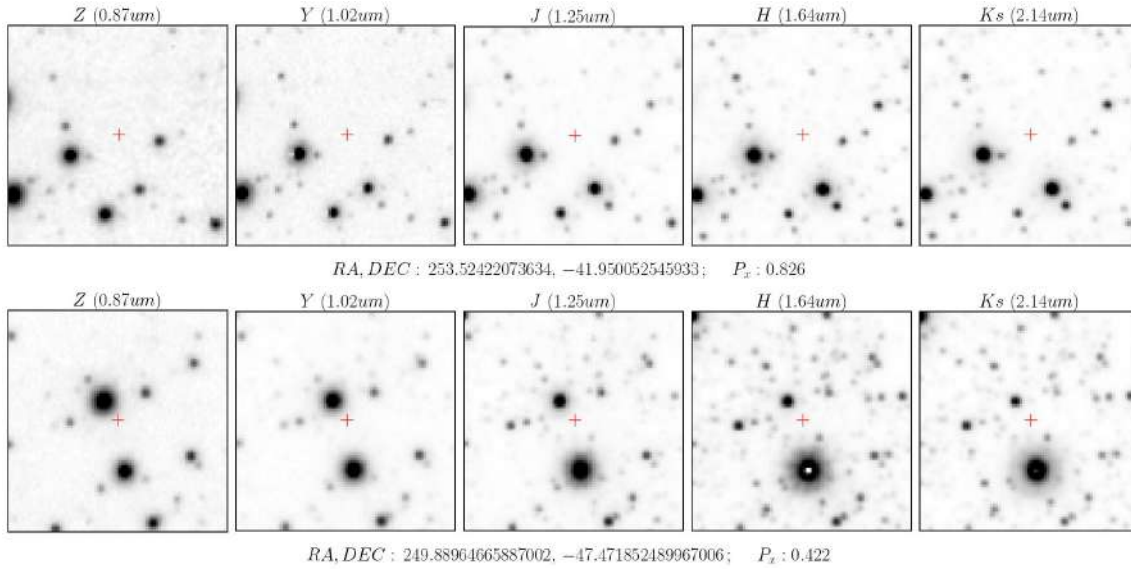


Figura 4.4: Ejemplos para un stamp con la region central vacía, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.826$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.422$.

5. Estrella Brillante Saturada

En este tipo de clasificación, se observa la presencia de una estrella extremadamente brillante en alguna región del *stamp*, la cual está saturada en una o varias bandas. Esta estrella se distingue por ser significativamente más grande que las demás fuentes y exhibir píxeles saturados en su centro, observados como un punto blanco. Por lo general, este tipo de clasificación va acompañada de *spikes* presentes en la categoría siguiente (ver figura 4.6).

6. Spikes

Este tipo de clasificación se refiere a aquellas *stamps* que muestran una estrella masiva cercana, la cual produce un patrón de difracción de la luz en forma de *spikes*. Esta estructura es claramente discernible en la imagen 4.7. Para clasificar esta categoría, se consideró la presencia de este patrón exclusivamente.

7. Región de Formación Estelar Normal

En este caso se observa una región de formación estelar en la mayoría de las bandas. Su variación no es tan pronunciada como en el caso de los PUVSFR (definida del acrónimo del inglés para “Estrella altamente variable fotométricamente”) y es posible distinguir

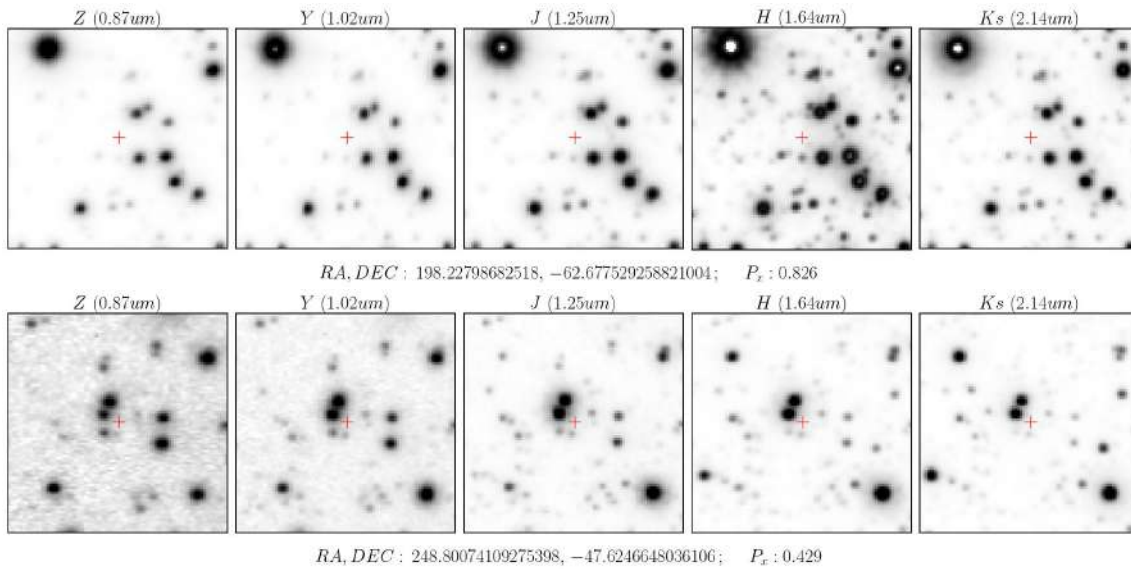


Figura 4.5: Ejemplos para un grupo de estrellas en el campo, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.826$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.429$.

otros tipos de estructuras, como grupos de estrellas, estrellas saturadas o *spikes* débiles provenientes de estrellas fuera de la *stamp*. Este caso representa el 50 % del total de la muestra. (ver figura 4.8)

8. Photometrically Ultra Variable Star

Definimos el acrónimo PUVS, que hace referencia a la presencia de una estrella cercana al centro que experimenta variaciones significativas en su brillo entre diferentes bandas. A pesar que a veces corresponden a estrellas variables, estos objetos a menudo son erróneamente identificados como fuentes extendidas. En la figura 4.9, se pueden observar estrellas de alto brillo superficial cerca del centro del *stamp* en ambas imágenes, especialmente en longitudes de onda más largas.

9. Asociación de Estrellas Masivas

Este caso corresponde al 0.23 % del total de datos, con solo 13 fuentes clasificadas exclusivamente de esta manera. Se caracteriza por la presencia de una región con varias estrellas masivas, donde la mayoría de las estrellas están saturadas y muestran estructuras de *spikes* en un área muy cercana. En este caso particular, las coordenadas de los *stamps* están a sólo unos pocos segundos de arco de distancia, lo que sugiere que pertenece a una

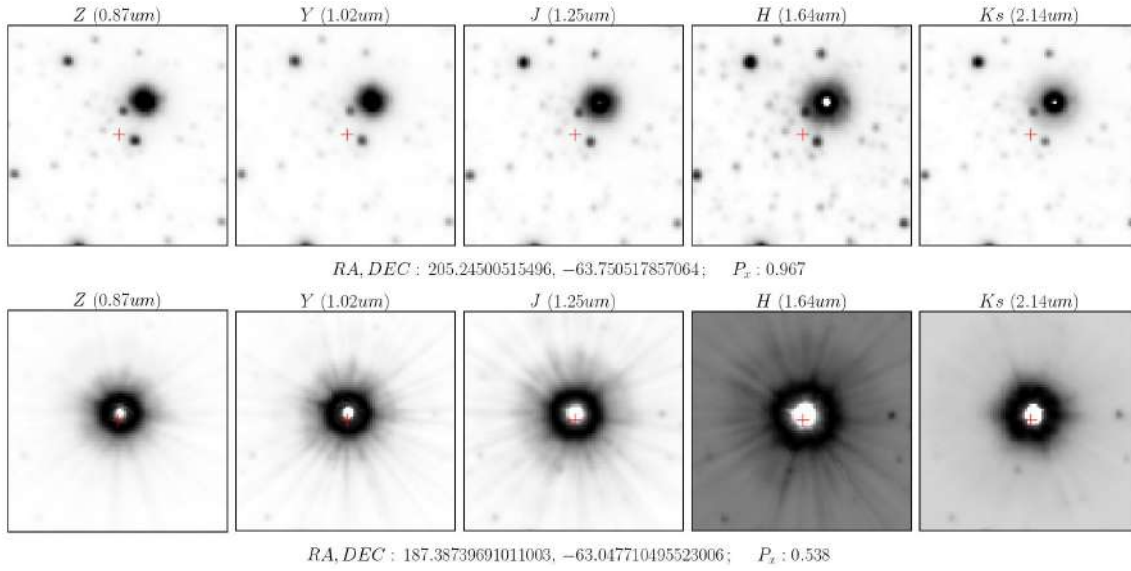


Figura 4.6: Ejemplos para regiones con una estrella saturada, ya sea alrededor o en el centro. El panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.967$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.538$.

zona de alta formación estelar, con fuerte emisión en rayos X (Ver figura 4.10).

10. Photometrically Ultra Variable Star Forming Region

La categoría PUVSFR, que se define como “Región de formación estelar muy variable fotométricamente”, corresponde a los casos en los que no es fácilmente apreciable una zona de gas alrededor de las estrellas o del *stamp* en las longitudes de onda más cortas. Sin embargo, al observar en longitudes de onda más largas (J , H y K_S), notamos un aumento significativo en su brillo, llegando a abarcar gran parte del *stamp* e incluso a saturar la estrella acompañante (ver figura 4.11, panel superior). En el caso del panel inferior de la imagen, el cambio significativo entre las bandas de longitudes de onda corta y las longitudes de onda larga se debe principalmente a los altos niveles de extinción en la zona, siendo posible observar tal nivel de detalle en las bandas H y K_S .

Es importante tener en cuenta en la clasificación, que es posible que las fuentes detectadas por XMM-Newton no sean detectadas por el telescopio VISTA, debido a las severas condiciones del disco galáctico con alta extinción estelar. Por lo que en el caso de las regiones centrales vacías, y región normalmente abarrotada, la presencia de una galaxia no observable en el relevamiento VVV puede ser posible. Sin embargo, el **resultado más importante de esta sección, es que no se encontró ninguna galaxia durante la inspección visual.**

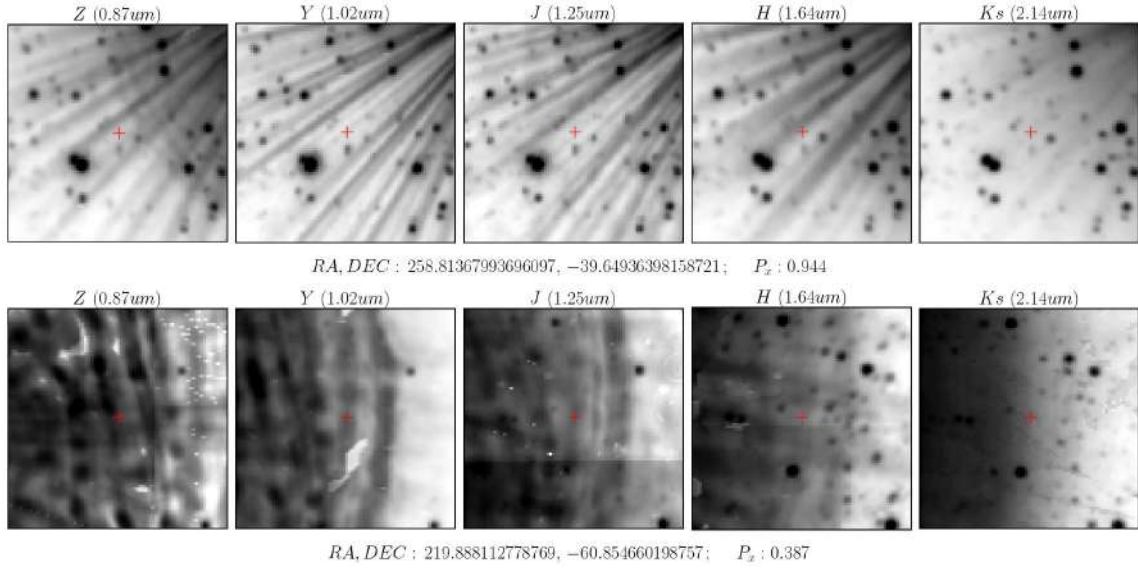


Figura 4.7: Ejemplos para estructura de Spikes. El panel de arriba corresponde a una fuente con alta probabilidad, $P_x = 0.944$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.387$.

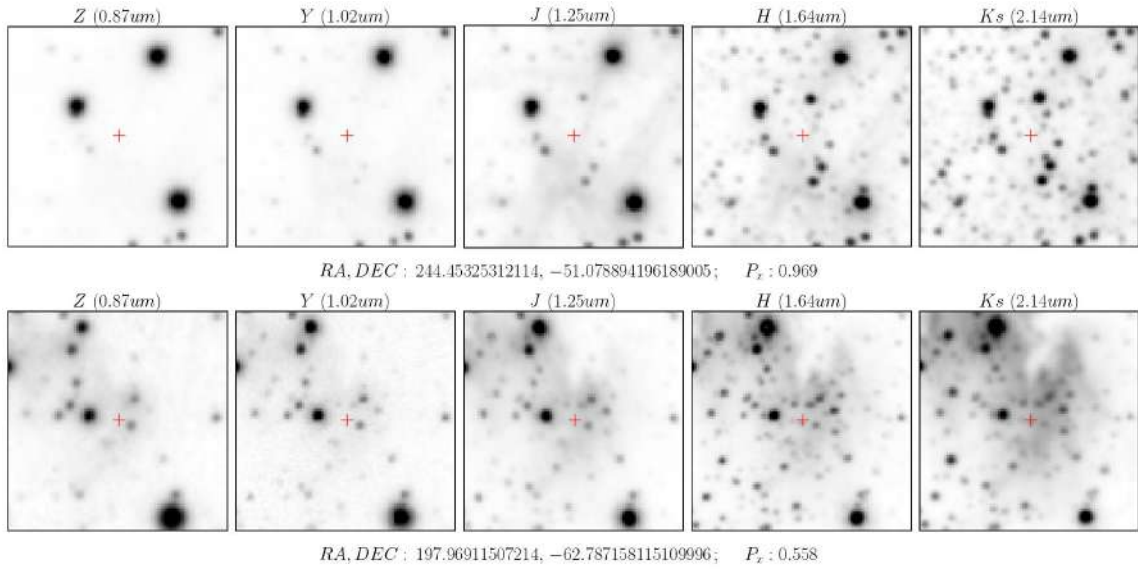


Figura 4.8: Ejemplos para una zona con formación estelar. El panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.969$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.558$.

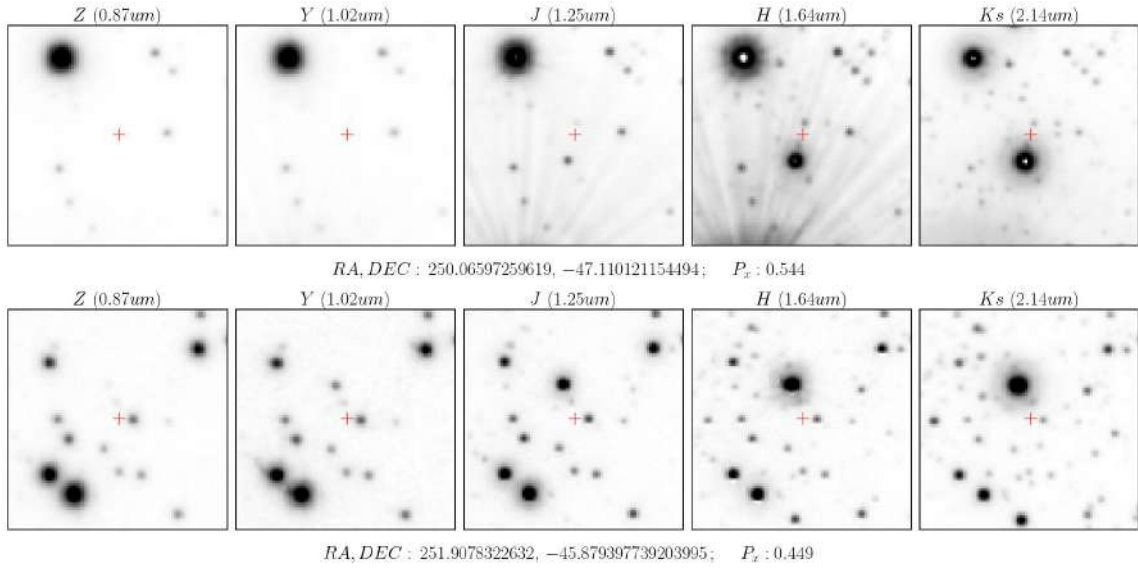


Figura 4.9: Ejemplos para PUVS, el panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.544$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.449$.

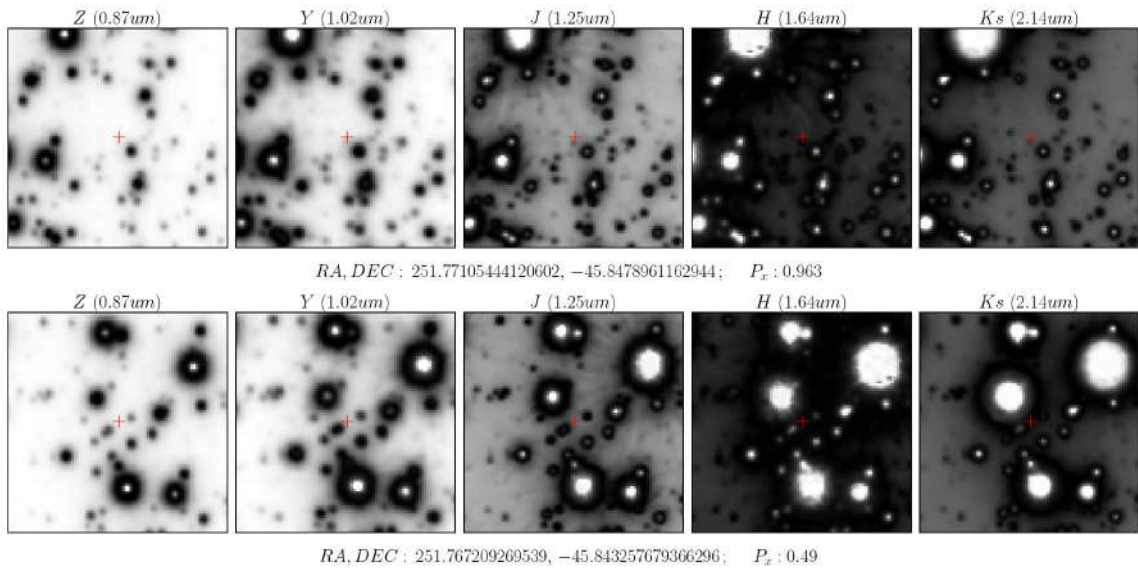


Figura 4.10: Ejemplos para stamps que muestran una asociación de estrellas masivas. El panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.963$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.49$.

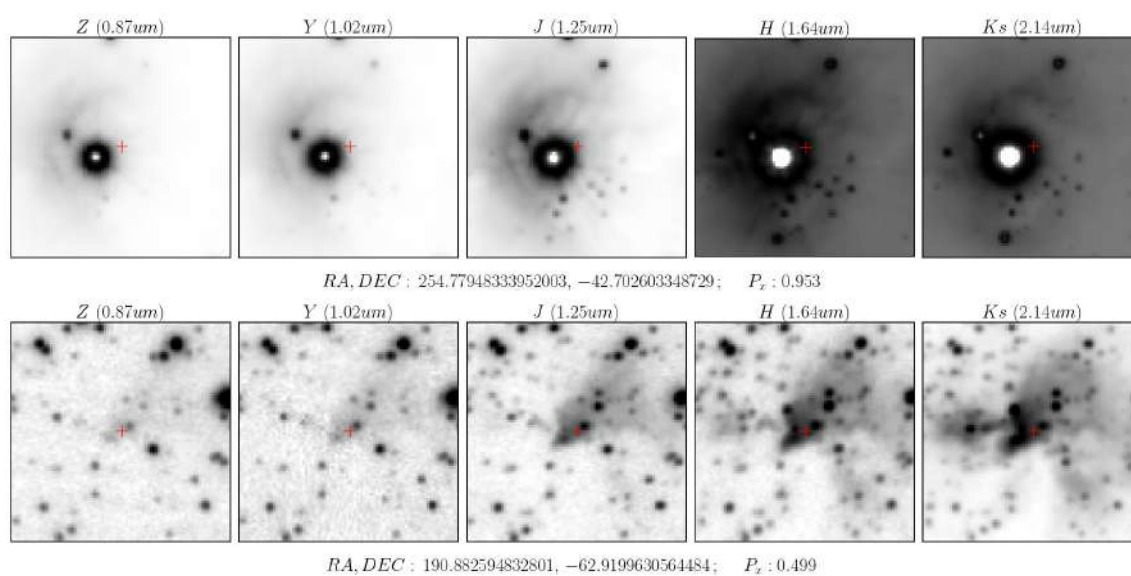


Figura 4.11: Ejemplos para *stamp* con estructuras de PUVSFR. El panel de arriba corresponde a una fuente con alta probabilidad $P_x = 0.953$ para ser galaxia, mientras que el panel de abajo corresponde a uno con bajo $P_x = 0.499$.

4.2. Distribución de Energía Espectral y Fotometría

En vista de no haber encontrado galaxias al inspeccionar cada fuente de la muestra *NOMatch*, decidimos estudiar el comportamiento de los flujos de algunas fuentes. Esto mediante el modelamiento de SEDs y el análisis en la distribución de los flujo de algunas fuentes.

Modelamiento de SEDs

En esta parte abordaremos el proceso de modelado de la “Distribución de Energía Espectral” (SED) como una herramienta fundamental en nuestro estudio para comprender las características y comportamientos de las fuentes observadas en el disco sur. El modelado de SEDs es una herramienta que permite analizar la radiación emitida por las fuentes a lo largo de un amplio rango de espectro electromagnético. Al considerar diferentes parámetros y condiciones, podemos simular y comparar cómo se espera que se comporte la radiación de diversas fuentes bajo diferentes escenarios. En este contexto, empleamos la herramienta de X-Cigale (Yang et al. 2020), un extensión del código Cigale (Boquien et al. 2019), que nos permite generar SEDs sintéticos desde los rayos X hasta IR lejano y explorar así, cómo varían los datos en respuesta a cambios en los parámetros físicos.

Con el objetivo de modelar parámetros para galaxias cercanas, seleccionamos redshifts entre $z = 0.05$ y 0.4 , considerando los redshifts esperados para galaxias observadas a través del disco de la Vía Láctea en las bandas J , H y K_s del telescopio VISTA, además de ser valores de redshifts esperados según trabajos como Galdeano et al. (2021, 2023); Baravalle et al. (2019). Para construir el componente galáctico, utilizamos un modelo de Historia de Formación Estelar (SFH; *Star Forming History*) de doble exponencial. La elección de la exponencial doble en lugar de la exponencial simple se basa en investigaciones previas que han demostrado un mejor ajuste en casos de SFH más complejas, a menudo asociadas con galaxias enanas de baja masa y galaxias del Grupo Local (Salim et al. 2016; Smith et al. 2021). Para modelar la componente estelar del espectro, utilizamos la biblioteca de Bruzual & Charlot (2003) para poblaciones estelares individuales (SSP). Adoptamos una Función de Masa Inicial (IMF) de Chabrier (2003) y consideramos metalicidades de 0.02 (metalicidad solar), 0.008 y 0.05. Además atenuamos la emisión estelar utilizando la ley de atenuación de Calzetti et al. (2000). En este trabajo, hemos probado una serie de valores indicados en la Tabla 4.2, de acuerdo con las condiciones requeridas para cada paráme-

tro, según los trabajos de Mountrichas et al. (2021); Yang et al. (2022). Para aquellos no especificados, se usaron los valores por defectos del código.

Tabla 4.2: Valores de los parámetros utilizados para generar los modelos de SED analizados. Para los valores no descritos en la tabla se usaron los valores por defecto del software.

Parameter	Model/values
Star formation history: Double exponential	
age (Myr)	7000
τ_{burst}	0
Normalize	False
Simple Stellar Population: Bruzual & Charlot (2003)	
imf	1: Chabrier (2003)
metallicity	0.008, 0.02, 0.05
Galactic dust extinction	
Dust attenuation law	Calzetti et al. (2000)
Amplitude of the UV bump	3.0 ^a
X-ray Module	
α_{ox}	-1.5
Restframe parameters	
D4000	True
Redshift module	
redshift	0.05, 0.1, 0.2, 0.3, 0.4

^a Valor para la Vía Láctea

La idea principal de este modelado es reproducir las observaciones utilizando los datos del XMM-Newton y VVV. Esto nos permitirá investigar si las fuentes que muestran emisión en rayos X también presentan emisión observable a través de la Vía Láctea. Además, podremos determinar si estas galaxias deberían ser detectables y observables con VVV. La figura 4.12, muestra nuestra estimación de los SED para distintos redshift de una galaxia típica esperada en la muestra, incluyendo la región de rayos X. Notamos que no se detecta emisión en el canal F8, correspondiente a la emisión total del telescopio XMM-Newton (0.2 – 12 keV). Esta metodología no nos permite obtener una estimación de la emisión de ambas bandas. Al cambiar algunos valores de los parámetros del código, este resultado se mantiene, dándonos a entender un problema en el manejo de los datos entregados al código.

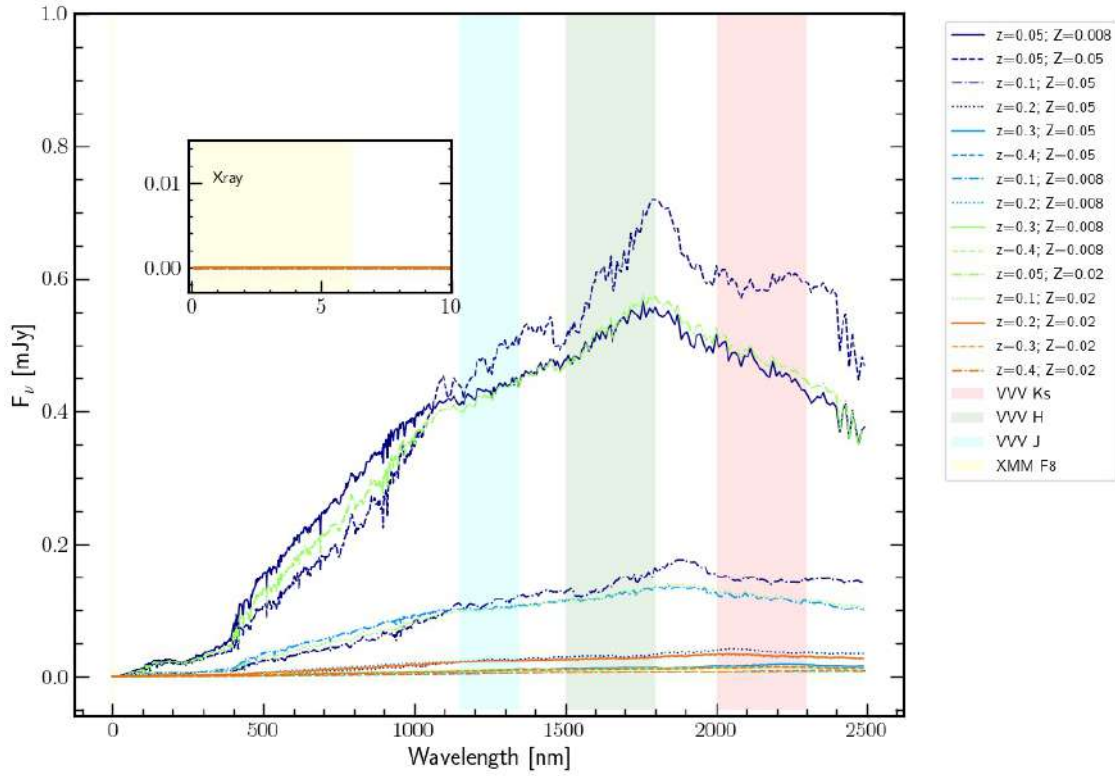


Figura 4.12: modelamientos de SED para la muestra *NMatch*, usando la configuración mostrada en la tabla 4.2

Fotometría

También revisamos el análisis fotométrico para obtener una comparación más detallada de las fuentes y su interacción con el entorno. Para ello, bajamos el catálogo completo del 4XMM-DR9 a través de VizieR ¹. Este proceso nos permitió obtener el flujo correspondiente a cada fuente mediante un cruce entre las tablas y datos recopilados en nuestro estudio. Posteriormente, separamos los datos en dos submuestras, la primera con contraparte en la muestra de salida de SExtractor+PSFEX y la muestra *NMatch*. Denominaremos las submuestras como galaxias confirmadas, *nan* y *no_nan*, estos dos últimos debido a la presencia o ausencia de columnas con valores "nan".^{en} la visualización de las tablas. Para llevar a cabo el análisis, consideramos tres tipos de muestras: 1. las cuatro galaxias confirmadas visualmente del catálogo VVV NIRGC, la submuestra de *galXray* con contraparte en la salida de SExtractor+PSFEX (*no_nan*) y la submuestra *NMatch* (*nan*). En la figura 4.13, presentamos un diagrama con las diferentes bandas relevantes del estudio

¹<https://vizier.cds.unistra.fr/viz-bin/VizieR-3?-source=IX/59/xmm4dr9s>

en el *eje-x*, que incluyen XMM (rayos X duros), XMM (rayos X suaves), VVV-*J*, VVV-*H* y VVV-*K_S*. En el *eje-y*, los flujos se representan como $\log(F)$, [mW m^{-2}]. La barra de color indica la extinción galáctica conforme al mapa de extinción de polvo elaborado por Schlafly & Finkbeiner (2011b), donde los puntos más claros indican niveles más altos de extinción.

Es importante destacar la distinción entre los rayos X duros (< 2 keV) y blandos (> 2 keV), debido a la atenuación que los últimos experimentan por parte del disco galáctico. Esto hace que la detección de fuentes extragalácticas a través del disco sea más probable mediante los rayos X duros.

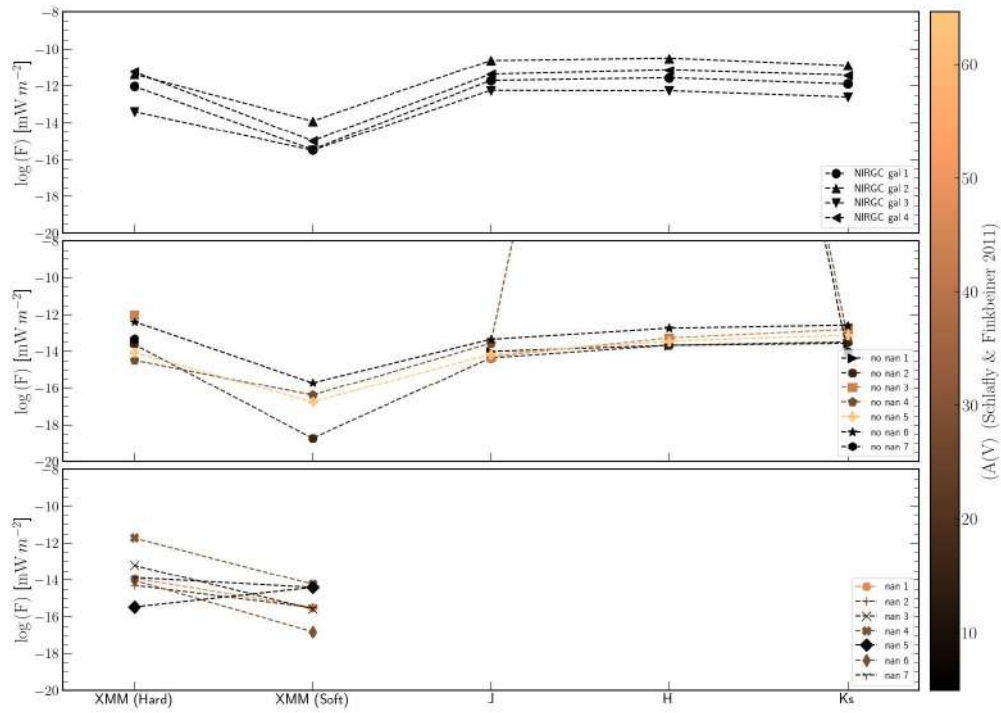


Figura 4.13: Diagrama mostrando el comportamiento de los flujos para diferentes submuestras, las galaxias confirmadas visualmente, datos con contraparte en el relevamiento VVV y los *NOmatch*.

Al observar los datos de Z21 y galaxias del VVV NIRGC en la figura 4.13, se aprecia un mayor flujo en los rayos X duros en comparación con los rayos X suaves, con una diferencia de aproximadamente tres órdenes de magnitud. Esto sugiere una similitud en las emisiones de las galaxias confirmadas y los datos *no nan*, como se observa en la curva superior y central del gráfico. Sin embargo, en el caso de las galaxias, los flujos son ligeramente mayores ($\log(F_{\text{nirgc}}) \gtrsim -12$ versus $\log(F_{\text{nn}}) < -12$), recordando su distribución

en zonas con extinción galáctica no tan altas.

No fue posible extraer información relevante a la clasificación de las fuentes `nan` y `no_nan`. Sin embargo, podemos confirmar la presencia de fuentes con emisión en rayos X similar a las encontradas para las cuatro galaxias del catálogo del VVV NIRGC.

Capítulo 5

Zonas de Interés

En este capítulo, profundizamos el estudio de la distribución de algunas fuentes de la muestra de *galXray*, basadas en la distribución observada en el diagrama de dispersión de la figura 5.1. Además de las fuentes de rayos X, la figura muestra el VVV NIRGC completo y las galaxias de Schröder et al. (2019). Seleccionamos cinco “Zonas de Interés” distribuidas a lo largo del plano galáctico, numeradas del Z1 al Z5 en orden descendente por longitud galáctica. Estas zonas exhiben características únicas que varían desde alta extinción hasta estructuras importantes en el estudio de la EGE. Es fundamental también considerar la distribución de las probabilidades para ser galaxias, proporcionada por Z21 a lo largo del disco galáctico, como se muestra en la figura 5.2. En este mapa, la probabilidad P_x se ve reflejada por la barra de color, representando el amarillo el más alto P_x y los tonos morados, una probabilidad más baja. Las zonas de interés se distinguen por su color amarillo, con la excepción de la Z3, seleccionada por su latitud galáctica cercana a 0° y su ubicación cerca del puente de Norma (Radburn-Smith et al. 2006).

Para comparar la distribución de las fuentes de Z21 en estas zonas, utilizamos imágenes de XMM-Newton en los canales eb4, eb3 y eb2 (0.5 – 1.0, 1.0 – 2.0 y 2.0 – 4.5 keV, respectivamente) disponibles en Aladin. Estas imágenes revelaron estructuras extendidas galácticas en las zonas Z1, Z2, Z3 y Z5, sugiriendo la presencia de remanentes de supernovas, mientras que la Z4 parece indicar una estructura de gas no uniforme, como analizaremos más adelante.

Además agregamos las galaxias de *galXray* sobre las imágenes de cada zona del canal eb3 (ver figuras 5.3, 5.4, 5.5, 5.6 y 5.7, respectivamente). Estas figuras muestran una distribución no homogénea de los datos de *galXray*, lo que sugiere una correlación con la estructura del gas caliente en cada zona.

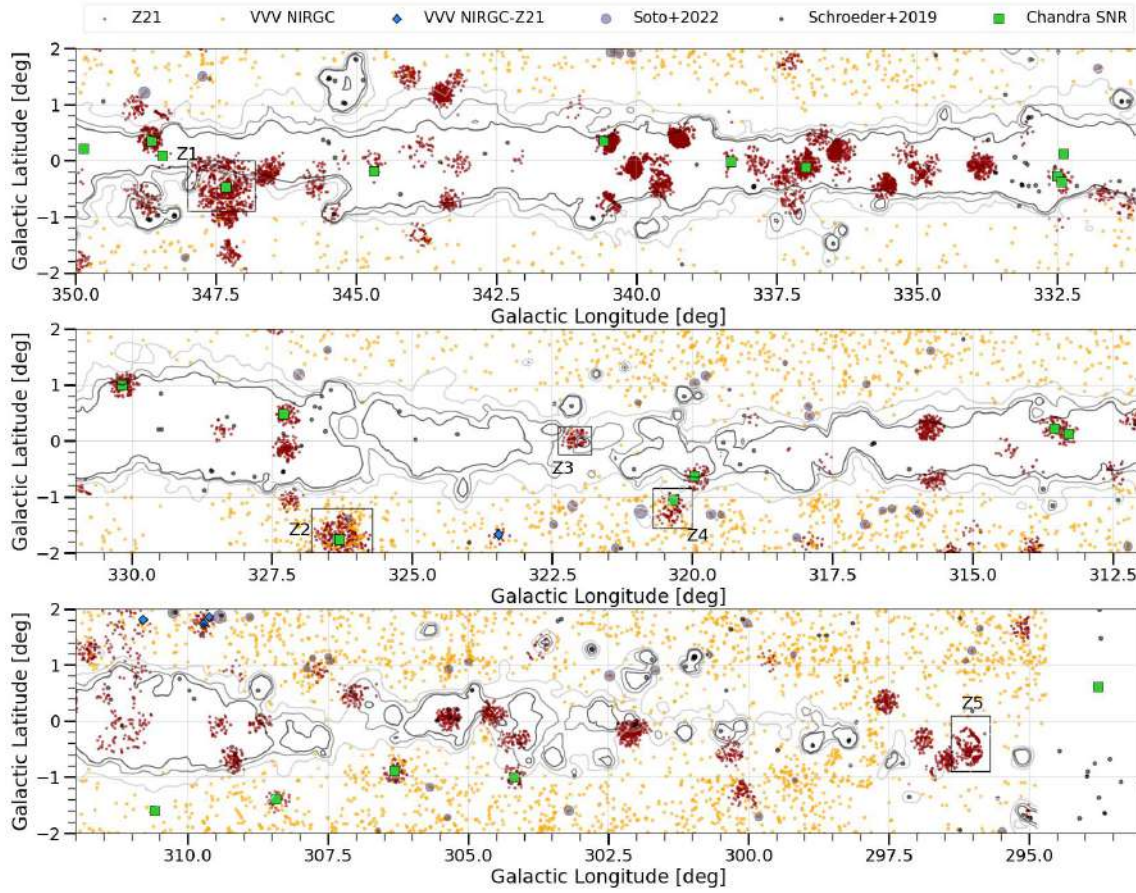


Figura 5.1: Distribución de las galaxias de la muestra de Zhang et al. (2021) en la región del disco sur del VVV. La muestra *galXray* está representada por puntos rojos, las galaxias confirmadas visualmente del VVV NIRGC están en naranja, las 4 galaxias en común entre VVV NIRGC y Z21, son diamantes azules y las galaxias de Schröder et al. (2019) por puntos negros. El centro de las sobredensidades reportadas por Soto et al. (2022) están representados por círculos azules semi-transparentes, que denotan el radio de cada sobredensidad amplificado por un factor de 4. Los cuadrados negros marcan las Zonas de Interés estudiadas. Los cuadrados verdes denotan remanentes de supernova compilados por Chandra. Los isocontornos A_V derivados de los mapas de extinción de Schlafly & Finkbeiner (2011a) se superponen en un gradiente gris con niveles de 11, 15, 20, y 25 mag.

Para un análisis más detallado, inspeccionamos las imágenes de cada zona en diferentes longitudes de onda, desde radio hasta los rayos X, utilizando las bandas de SUMSS (843.0 MHz), AllWISE W4 (22.0 μm), W3 (12.0 μm), SPITZER irac 4 (8.0 μm), irac 3 (5.8 μm), 2MASS K (2.16 μm), VVV H (1.5 μm), DECaPS Y (1009.0 nm), z (926.0 nm), i (784.0 nm), r (642.0 nm), g (473.0 nm), SHS (590.0 nm), XMM eb2 (0.5 – 1.0 keV), eb3 (1.0 – 2.0 keV) y eb4 (2.0 – 4.5 keV).

Las figuras 5.8, 5.9, 5.10, 5.11, 5.12 muestran las zonas de interés en las diferentes

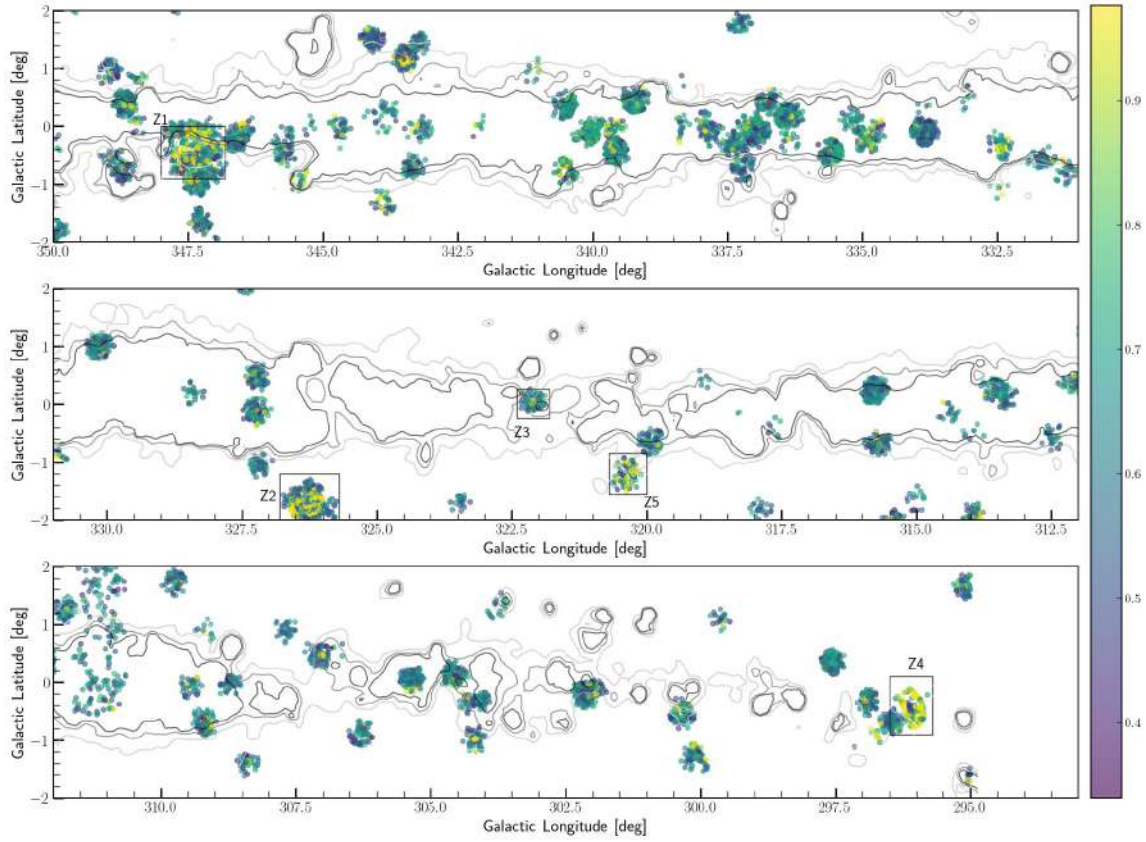


Figura 5.2: Distribución de las galaxias de Zhang et al. (2021) en la región del disco sur del relevamiento VVV (*galXray*). La muestra *galXray* esta representada por codificación por colores, con la probabilidad P_x de ser una galaxia. Los cuadrados negros marcan las Zonas de Interés estudiadas. Los isocontornos A_V derivados de los mapas de extinción de Schlafly & Finkbeiner (2011a) se superponen en un gradiente gris con niveles de 11, 15, 20, y 25 mag.

longitudes de onda, permitiendo entender las condiciones del campo en los diferentes rangos espectrales, destacando la profundidad y calidad alcanzada por el telescopio VISTA. En este caso, utilizamos la banda K_s del telescopio 2MASS, debido a que no se encuentra disponible esta banda para el telescopio VISTA en ALADIN.

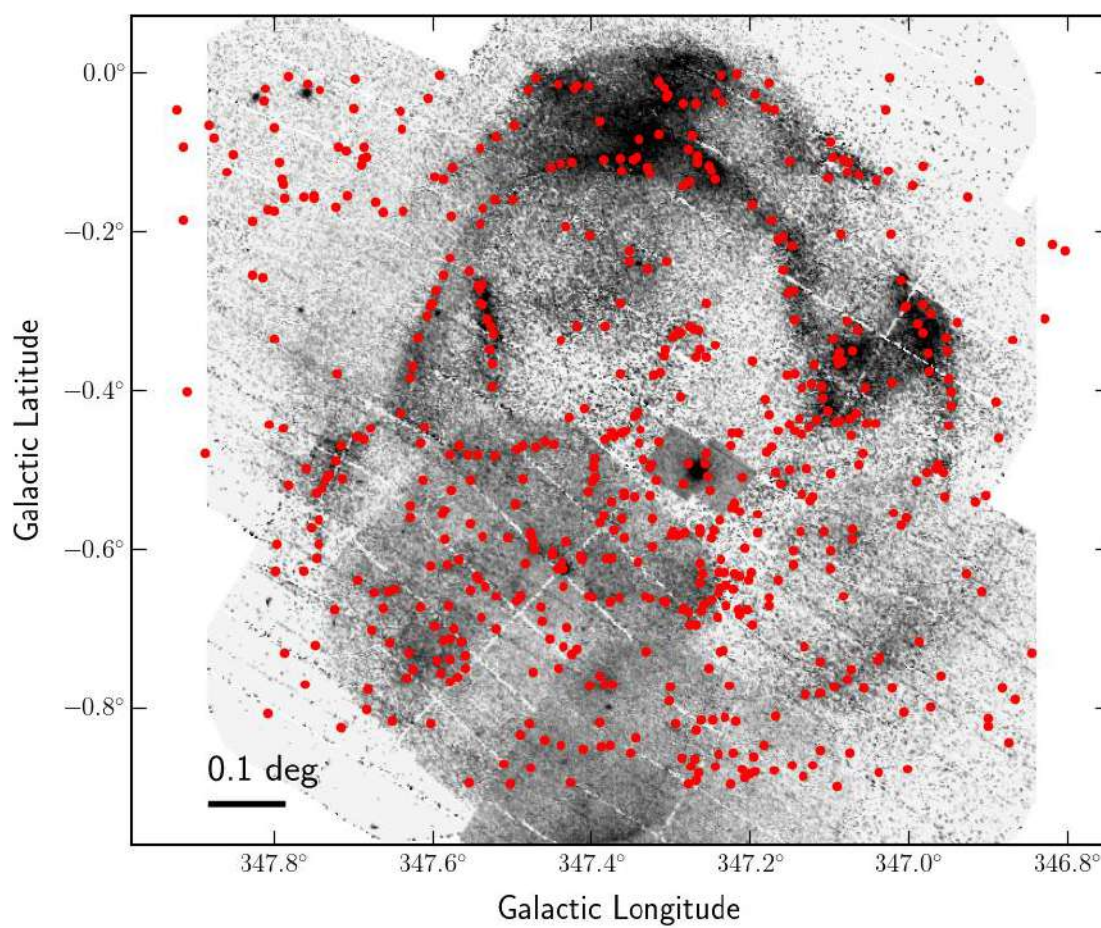


Figura 5.3: Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z1, incluyendo los objetos de *galXray* como puntos rojos.

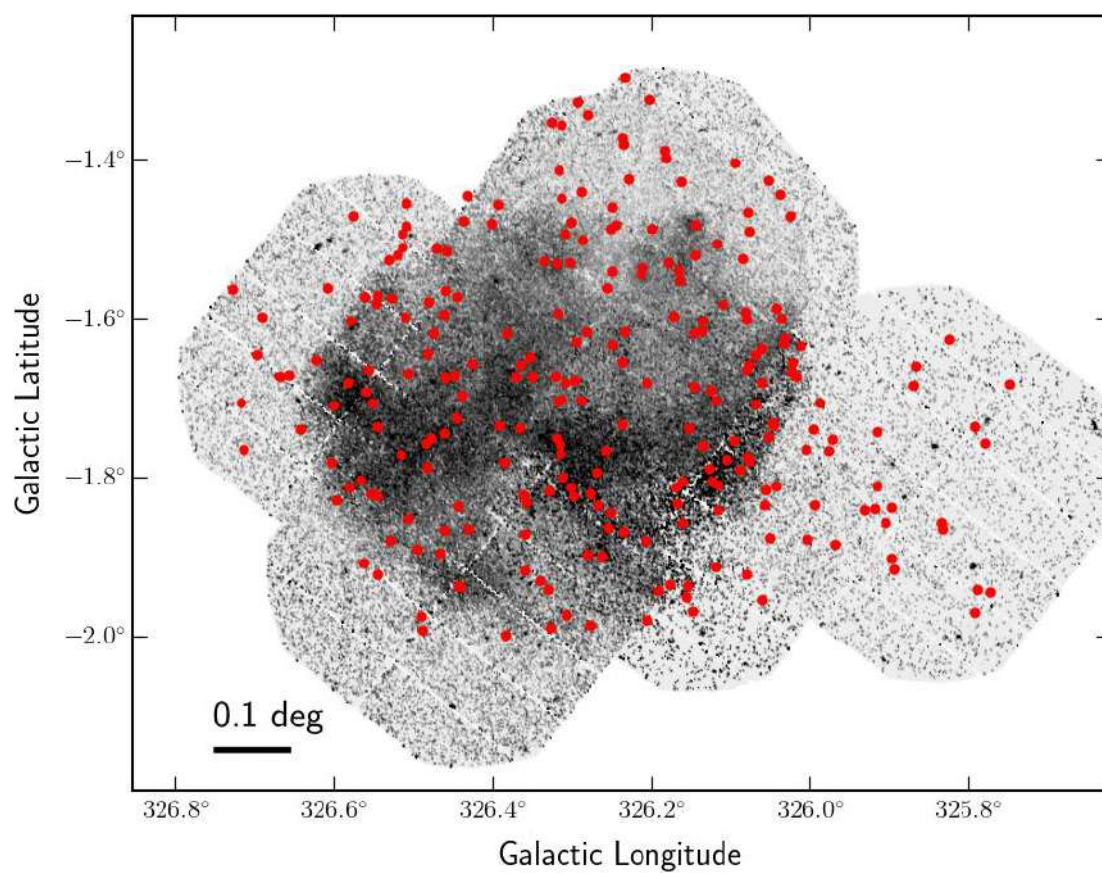


Figura 5.4: Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z2, incluyendo los objetos de *galXray* como puntos rojos.

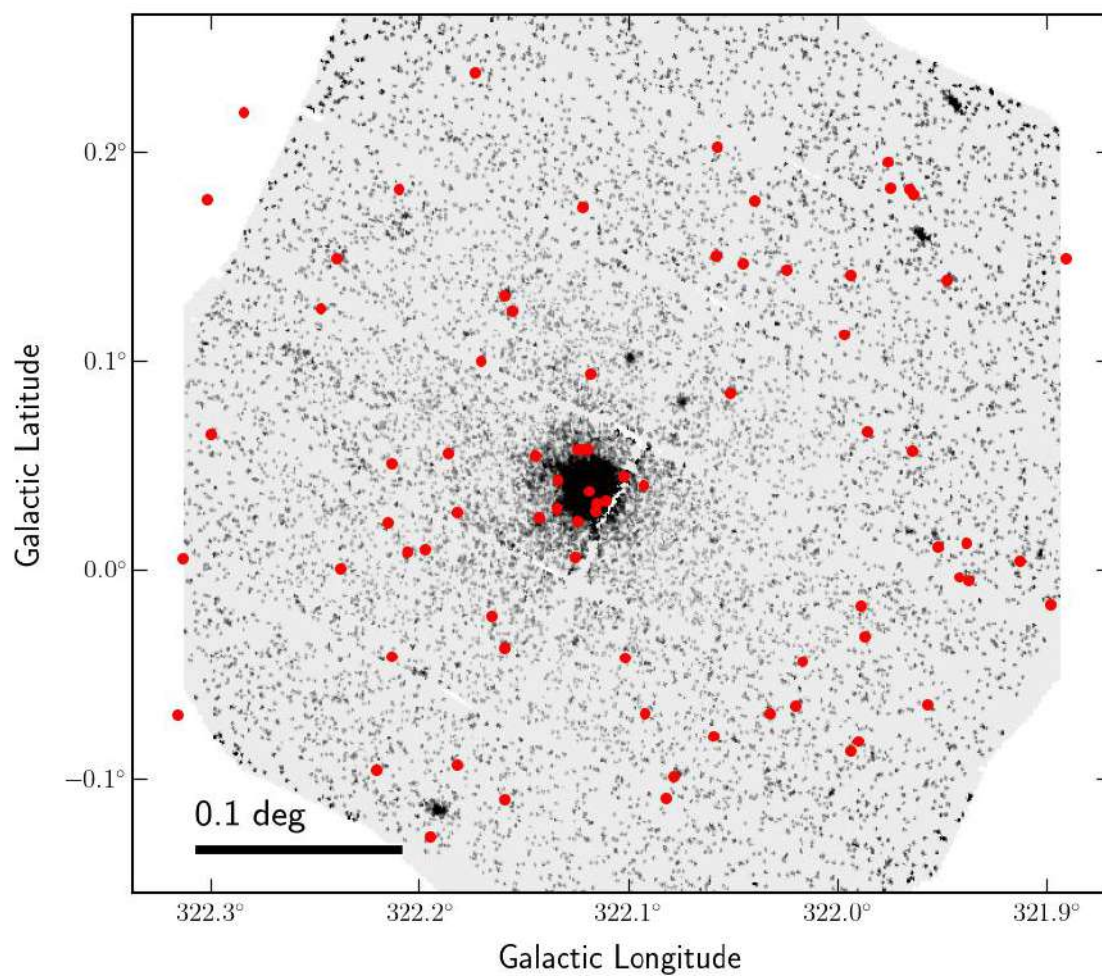


Figura 5.5: Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z3, incluyendo los objetos de *galXray* como puntos rojos.

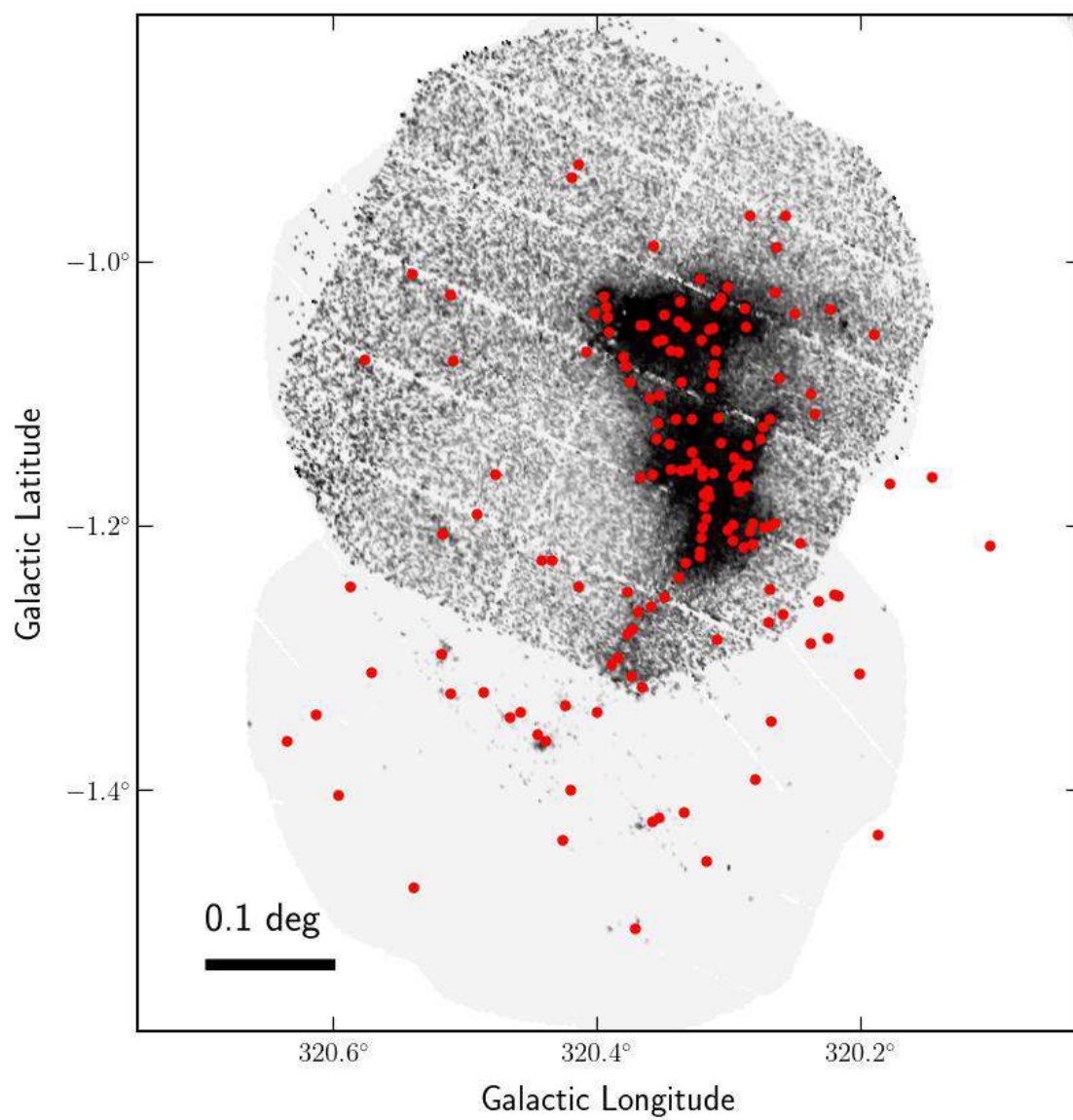


Figura 5.6: Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z4, incluyendo los objetos de *galXray* como puntos rojos.

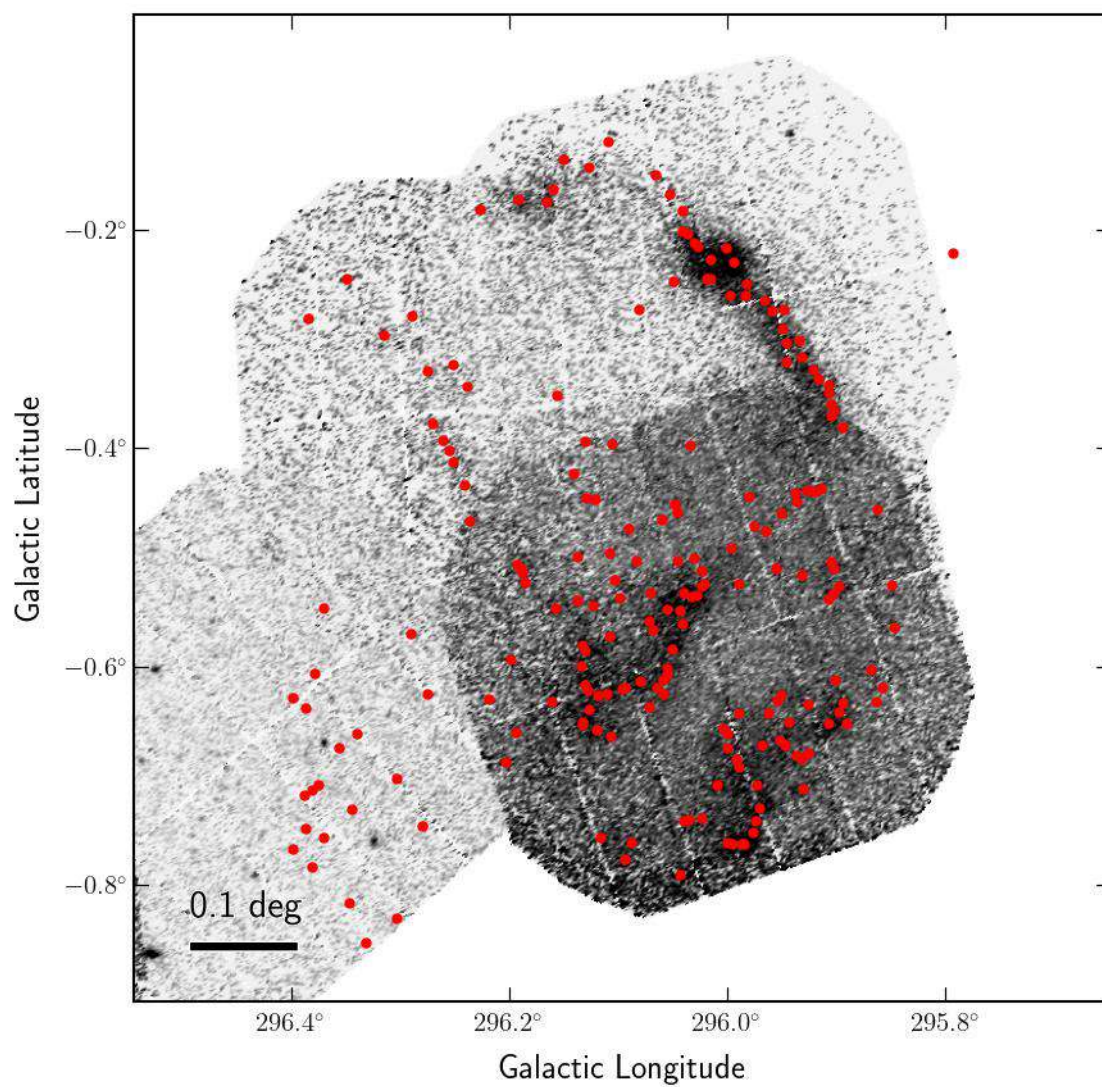


Figura 5.7: Imagen del canal eb3 del telescopio XMM-Newton de la zona de interés Z5, incluyendo los objetos de *galXray* como puntos rojos.

5.1. Zona 1

En las diferentes longitudes de onda para la Z1, como se muestra en la imagen 5.8, podemos observar que la emisión en rayos X detectada por XMM también se observa ligeramente en SUMSS y algunas bandas ópticas, como DECaPS *g* e *i*. Además se nota una distribución simétrica en H- α en SHS.

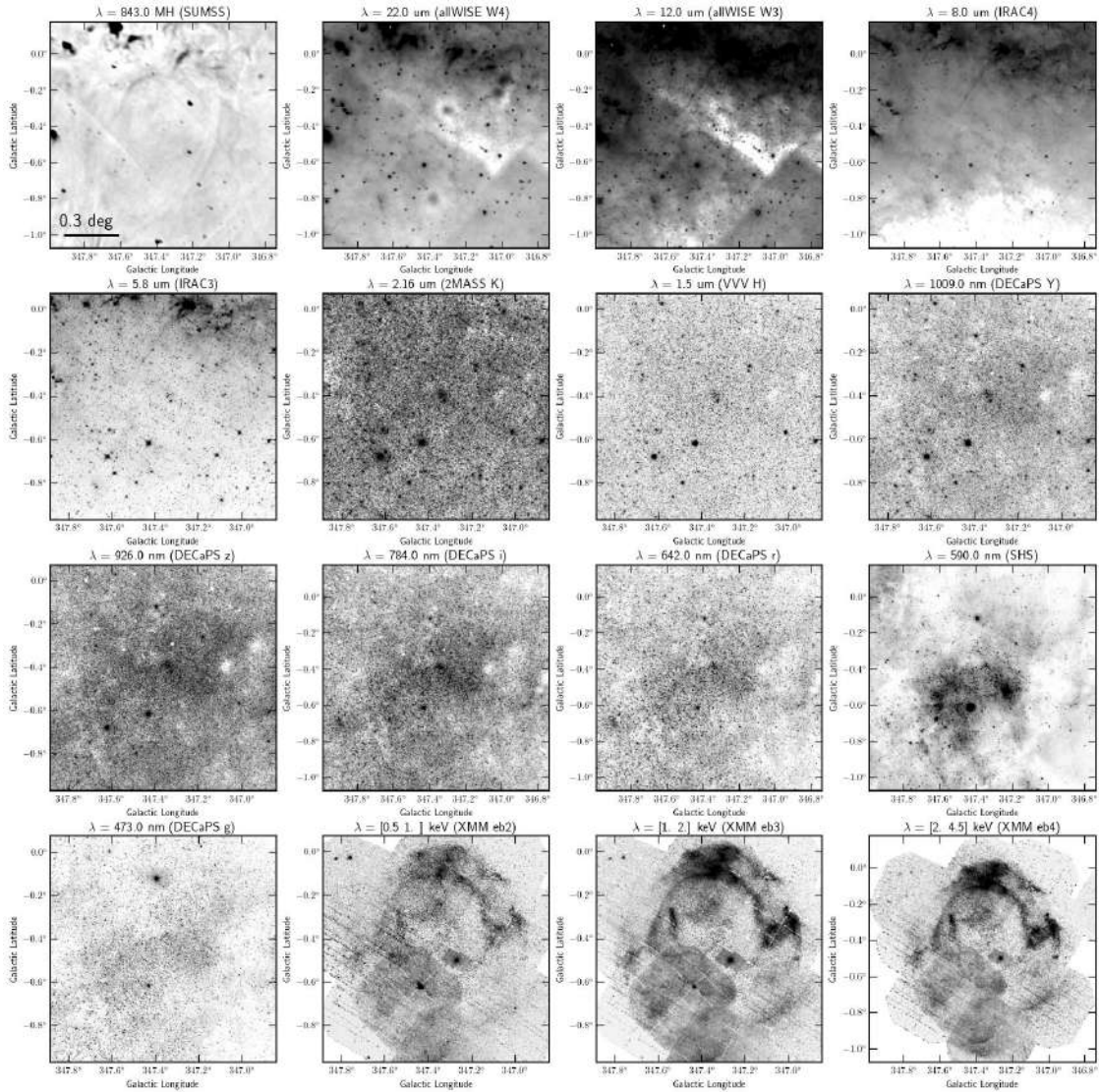


Figura 5.8: *stamps* de la zona de interés Z1, en diferentes longitudes de onda con los surveys disponibles.

En la Zona Z1, que está centrada en α : 17h13m21.31s, δ : -39°41'32.05'', es posible observar una **estructura circular simétrica**, que se asemeja a una esfera, rodeando una fuente de rayos X puntual central. Podemos notar que las 562 fuentes de *galXray* en esta

zona (como puntos rojos en la figura 5.3), siguen la estructura de la emisión de rayos X, con una mayor densidad en las zonas con mayor emisión (zonas más oscuras). De acuerdo a estudio previos (Slane et al. 1999; Vasquez et al. 2005; Cassam-Chenaï et al. 2004; Tateishi et al. 2021), esta región corresponde a un Remanente de Supernova, de tipo *shell*, caracterizada por radiación no-térmica, SNR RX J1713.7–3946. su fuente puntual central de rayos X corresponde a 1WGA J1713.4–3949, la cuál tiene una razón de flujo de rayos-X-a-óptico, coherente con lo que cabría esperar de una estrella de neutrones.

En el trabajo de Slane et al. (1999), se estudia también la posibilidad de que corresponda a una fuente extragaláctica. Sin embargo, estudios más recientes muestran que es altamente probable que la fuente puntual corresponda a reliquia compacta de la supernova progenitora del remanente en la categoría de supernovas de tipo II (Cassam-Chenaï et al. 2004). La otra fuente puntual que se observa en el remanente de supernova, ubicada a 12 arcos segundos de la fuente puntual central en rayos X, corresponde a la estrella Supergigante roja HD 155603. Los objetos clasificados como galaxias por Z21 pertenecientes a la Z1 delimitan claramente la estructura de remanente de supernova, con una probabilidad media de ser galaxia de 0.712

5.2. Zona 2

Al inspeccionar los *stamps* en diferentes longitudes de onda correspondiente a la Z2 (α : $15^h52^m26.793^s$, δ : $-56^\circ11'29.343''$) en la figura 5.9, se puede apreciar la **estructura circular de gas** mostrada en los distintos canales de rayos X en la figura 5.4. Además, se observa una estructura de tipo **burbuja** en radio, con un diámetro de aproximadamente 36 minutos de arco. En la banda del telescopio SHS, se aprecia una leve emisión que se superpone con el borde de la esfera en radio. Sin embargo, en las demás bandas no se observan otras estructuras significativas. En el caso de las imágenes del telescopio SPITZER, la resolución no es suficiente para resolver la Z3 a estas escalas. La estructura de la Z2, presenta una forma similar a la de un remanente de supernova, como se observa en la Z1. Las galaxias de esta zona tienen una mediana de P_x de 0.649. No se encontró información adicional sobre este objeto en la literatura, además de encontrar las estrellas UCAC4 170-140217, 2MASS J15523109-5551020 y CD-56 6060.

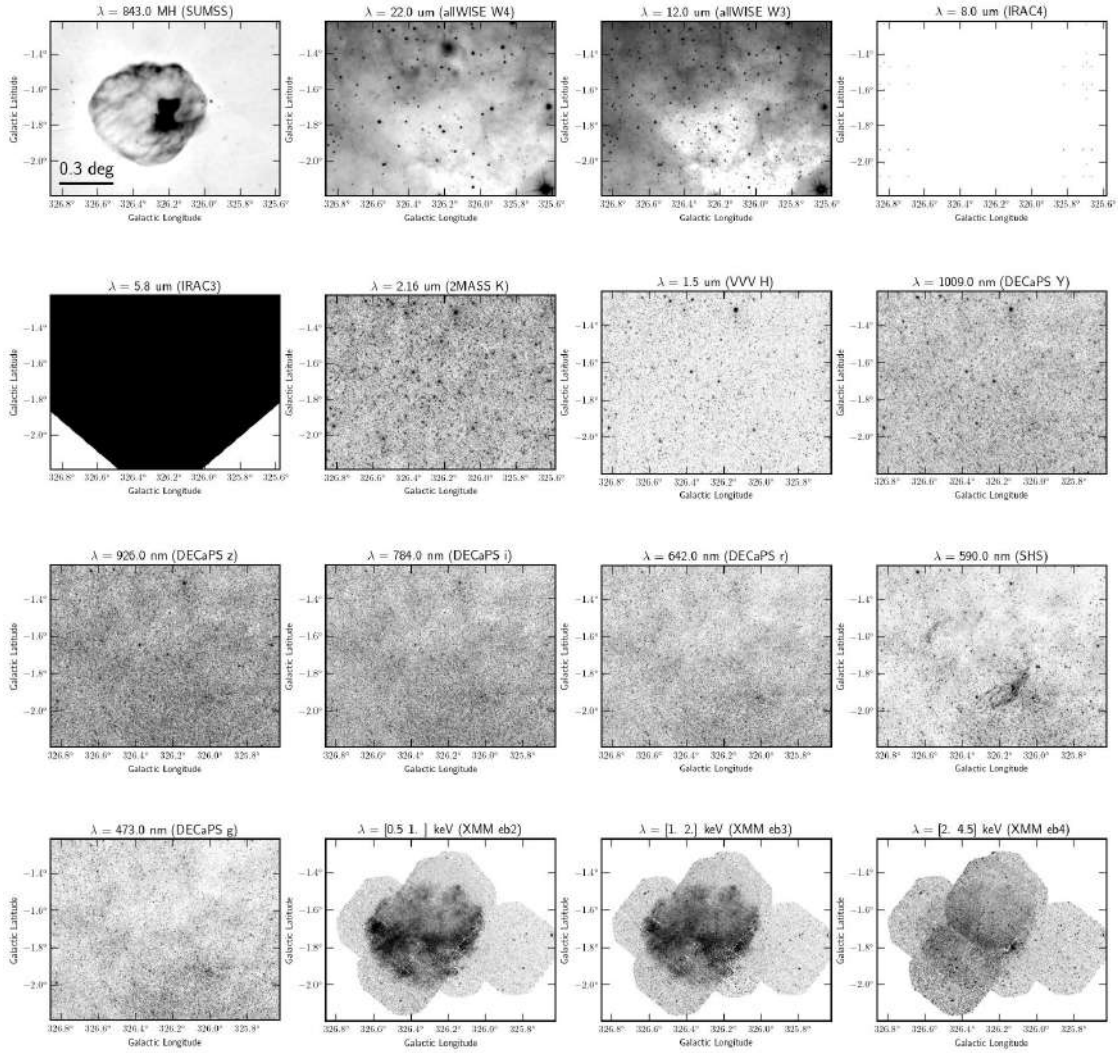


Figura 5.9: *stamps* de la zona de interés Z2, en diferentes longitudes de onda con los surveys disponibles.

5.3. Zona 3

Podemos observar en la figura 5.10 la **emisión central de rayos X** en los tres canales del telescopio XMM, así como una **estructura esférica** que rodea una parte central en radio. Sin embargo, no se aprecia alguna estructura o fuente extendida en los demás *stamps* de la zona.

Centrada en α : 15h20m34.605s, δ : $-57^{\circ}07'56.599''$, la emisión central mostrada en la figura 5.5 corresponde a la estrella de neutrones binaria de rayos X **Circinus X-1**, ubicada dentro de un remanente de supernova. Esta fuente es conocida como una de las más brillantes de rayos X del cielo. Heinz et al. (2013) estudió el remanente de supernova natal de

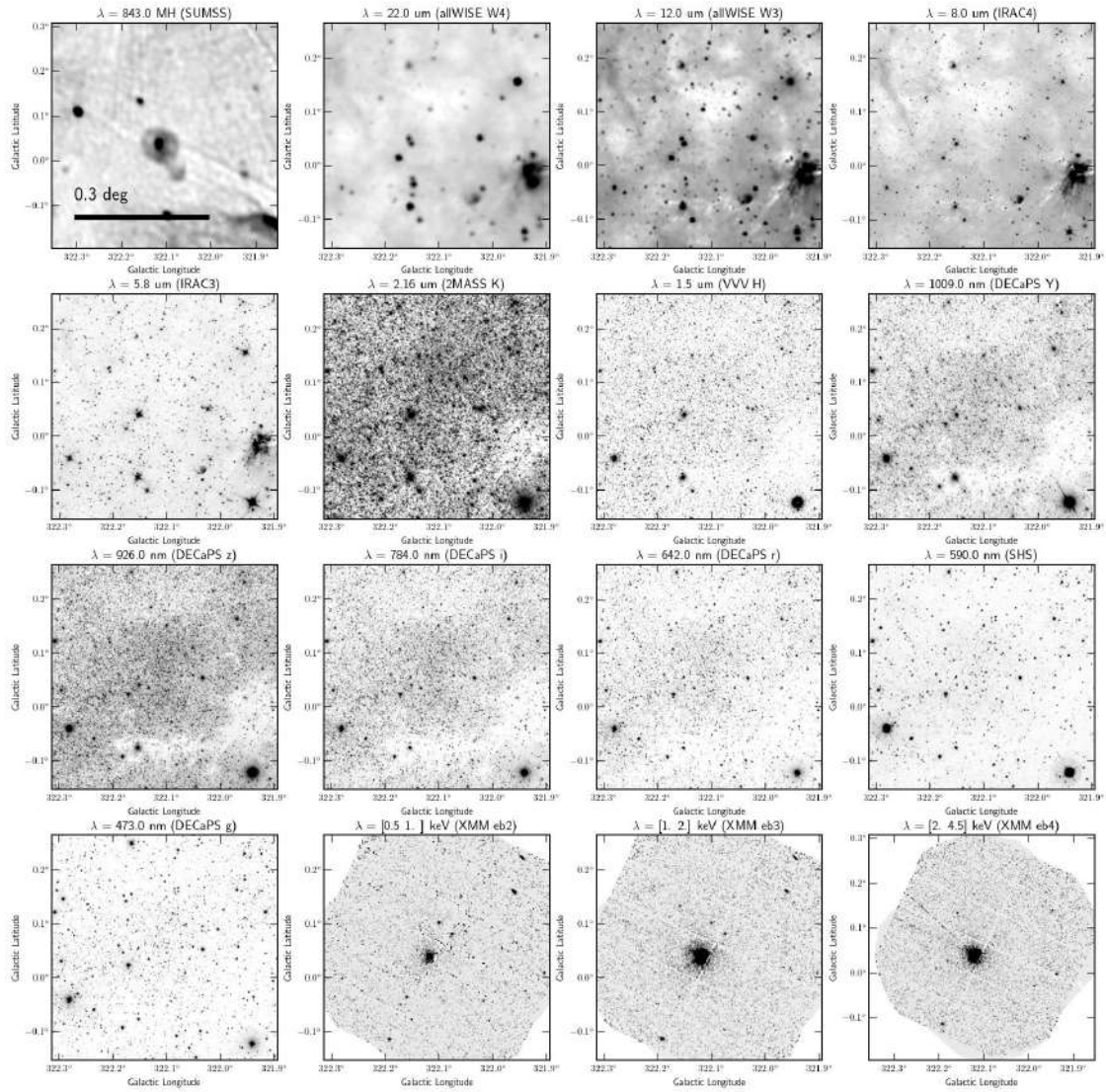


Figura 5.10: *stamps* de la zona de interés Z3, en diferentes longitudes de onda con los surveys disponibles.

la estrella de neutrones Circinus X-1, comparando la emisión en rayos X con la emisión en radio. En el *stamp* de SUMSS de la figura 5.10, es posible observar en el centro el remanente de supernova y el jet de la fuente binaria que se extiende en dirección diagonal hacia abajo-derecha. Los jets de Circinus X-1 han sido objeto de estudio a lo largo de los años (Sadeh et al. 1979; Phillips et al. 2007; Johnston et al. 2016; Coriat et al. 2019). Las galaxias de la Z3 delinean claramente la estructura de remanente de supernova al igual que en Z1, con una probabilidad media de ser galaxia de 0.674

5.4. Zona 4

En esta zona centrada en α : 15h14m43.7601s, δ : -59°09'40.265'', en los canales del telescopio XMM de la imagen 5.11, se observa una **estructura irregular** en la emisión de rayos X, mientras que en radio, notamos una emisión fuerte. Estas dos emisiones parecen corresponder al mismo objeto, lo que podría indicar una zona de formación estelar. No se logra distinguir ninguna estructura destacada en los demás *stamps*.

Este objeto conocido como "La mano Cósmica"(del inglés "*The Cosmic Hand*"), es el remanente de supernova (SNR) RCW 89 o MSH 15-52, con la estrella de púlsar joven PSR B1509-58 (Caswell et al. 1981). Las partículas relativistas y los campos producidos por este púlsar alimentan la nebulosa de viento de púlsar (*pulsar wind nebula*; PWN) brillante y extensa en rayos X, generando la forma que le da su nombre. RCW 89, observada por primera vez por Mills et al. (1961) a una distancia de 5.2 kpc, ha sido estudiada en diferentes longitudes de onda, debido a que el púlsar posee uno de los campos magnéticos más extremos, además de una gran extensión (Aharonian et al. 2005; Gaensler & Slane 2006; An et al. 2014; Romani et al. 2023). Con una probabilidad media de ser galaxia de 0.744.

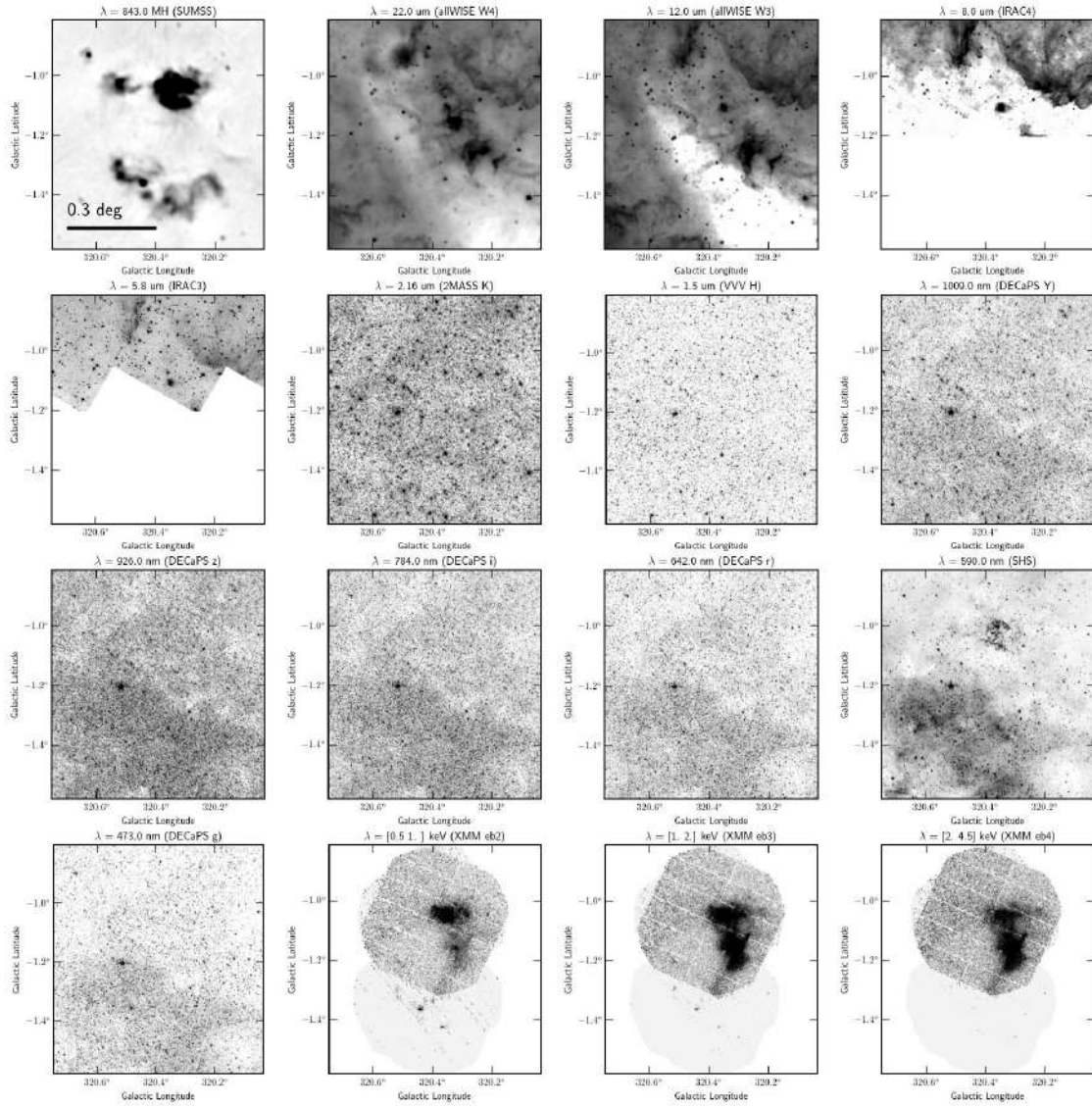


Figura 5.11: *stamps* de la zona de interés Z4, en diferentes longitudes de onda con los surveys disponibles.

5.5. Zona 5

En la Z5, centrada en α : 11h51m50.437s, δ : -62°35'19.304" (ver figura 5.12), se observa una emisión en forma **circular** superpuesta con otra emisión en rayos X, también observable levemente en radio. Sin embargo, no se nota mayor emisión en las demás bandas. Con una probabilidad media de ser galaxia de 0.966, no se encontró mayor información en la literatura sobre esta zona, además de existir un posible nova en la zona más brillante en rayos X (Nova Cen 2008; Hughes et al. 2010).

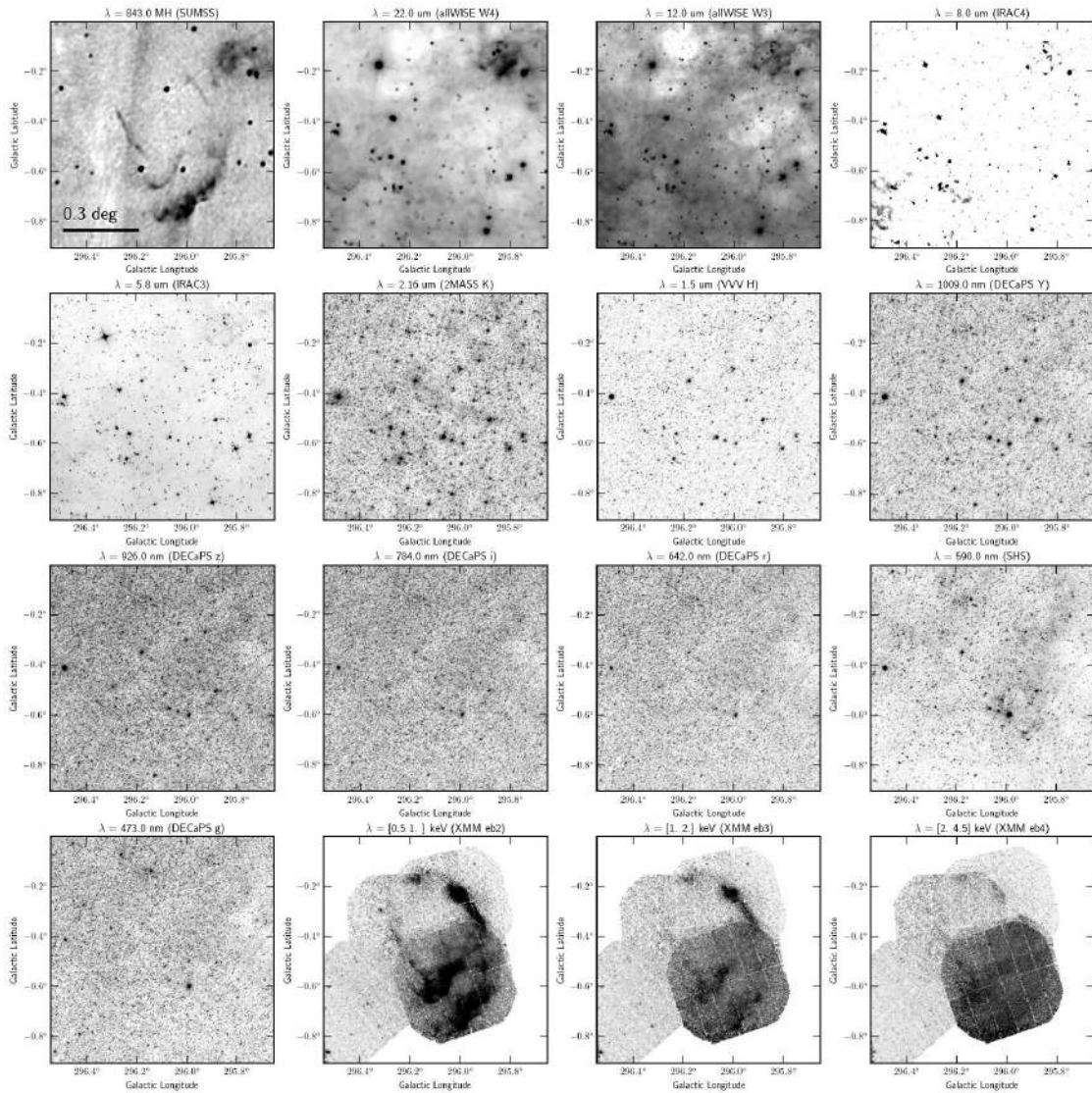


Figura 5.12: *stamps* de la zona de interés Z5, en diferentes longitudes de onda con los surveys disponibles.

5.6. Aspectos generales de las zonas

Además, se encontró que las zonas de interés Z1, Z2 y Z5 pertenecen al catálogo de remanentes de supernova del telescopio Chandra (Chandra Catalog of Galactic Supernova Remnants; Green 2019)¹. Varias sobredensidades de la muestra *galXray* están también catalogadas como SNR en Chandra.

En los remanentes de supernova ocurren procesos de emisión a partir de los choques, generados principalmente por la rotación de su estrella de neutrones central, interacción

¹https://hea-www.harvard.edu/ChandraSNR/snrcat_gal.html

de partículas y calentamiento del medio, que podemos observar de forma de radiación de sincrotron, Bremsstrahlung no térmico y térmico, emisión de dos fotones, emisión libre-ligada (Vink 2012).

Es interesante además, analizar la estadística de las diferentes muestras. La tabla 5.1 resume la estadística de la probabilidad P_x de que las fuentes sean galaxias para cada una de las zonas de interés, incluyendo la muestra de *galXray* y toda la muestra de Z21. La tabla muestra la identificación y el número de datos contenidos en las columnas (1) y (2), así como los cuartiles en (3), (4) y (5). Las fuentes de Z2 y Z3 presentan una mediana parecida a la muestra completa. La Z1 presenta altos niveles de probabilidad para ser galaxia, siendo que hay varios estudios que demuestran ser un remanente de supernova. La mediana de Z5 es la más alta, siendo un valor muy diferente comparado con las otras regiones. Además Z5 es la única zona de interés que se encuentra en una zona de baja extinción, en comparación con los demás datos. Además, resulta interesante que la distribución de los cuartiles de la muestra de Z21 con la de *galXray* no muestre cambios significativos entre los resultados, a pesar de los diferentes niveles de extinción de cada muestra.

Muestra de galaxias	# objetos	Q1	Q2	Q3
Z21	191,528	0.538	0.657	0.804
<i>galXray</i>	9,726	0.552	0.664	0.743
Z1	562	0.591	0.712	0.953
Z2	244	0.539	0.649	0.958
Z3	78	0.552	0.674	0.712
Z4	158	0.566	0.744	0.963
Z5	203	0.704	0.966	0.972

Tabla 5.1: Estadística de las probabilidades P_x para las distintas zonas y muestras.

Capítulo 6

Conclusión

En respuesta al aumento de la cantidad de datos en astronomía, el uso de herramientas de aprendizaje automático aparece como una solución cada vez más recurrente. La aplicación de estos algoritmos para la clasificación de objetos extragalácticos en regiones de baja latitud es un desafío particularmente complejo debido a los altos niveles de extinción presentes y, en consecuencia, el reducido número de galaxias en comparación a estrellas en estas zonas.

En este contexto, nos enfocamos en el trabajo llevado a cabo por Zhang et al. (2021), Z21, quienes emplearon algoritmos de aprendizaje automático para clasificar objetos en el relevamiento de cielo completo 4XMM-DR9 entre estrellas, galaxias y QSOs. Basándose en datos de rayos X del telescopio XMM-Newton, generaron cuatro muestras diferentes basadas en su contraparte en el NIR provenientes de AllWISE y en el óptico desde SDSS. Además, durante la fase de entrenamiento se incorporaron datos de otro rango de longitud de onda, lo que introdujo emisiones con señales potencialmente distinta.

Aunque presentaron un catálogo de galaxias de todo el cielo, es importante tener en cuenta que los algoritmos pueden tener limitaciones para detectar galaxias fuera del rango de características utilizadas en el entrenamiento. Por lo tanto, en nuestro trabajo comparamos las galaxias identificadas por estos autores en el rango de rayos X, en una región de baja latitud galáctica, con objetos detectados en el régimen del NIR mediante SExtractor+PSFEx. Verificar con un conjunto de datos diferente, como el relevamiento del VVV en el NIR, es fundamental, ya que el uso de diversas fuentes de datos puede conducir a clasificaciones erróneas, especialmente en latitudes galácticas más bajas.

La clasificación propuesta por Z21 considerando únicamente los rayos X, empleó el algoritmo *Rotation Forest* con un *accuracy* de 77.80 %. Sin embargo, al aplicar el mis-

mo algoritmo de clasificación a datos con características diferentes, como en el caso de la ZoA, pueden surgir clasificaciones erróneas, especialmente en latitudes galácticas más bajas. En estos casos, es crucial considerar la transferencia de datos o el aprendizaje por transferencia, un enfoque estratégico que permite que los conocimientos adquiridos previamente a partir de un conjunto de datos mejoren el rendimiento del algoritmo en otro set de datos, mitigando así los desafíos asociados con las características de señal y ruido variables.

En el caso Z21, el algoritmo fue entrenado usando datos del SDSS, que consisten principalmente en galaxias brillantes y de gran tamaño. Dado que no se encontraron galaxias ópticas en la región estudiada, es posible que el algoritmo clasifique objetos con emisión de rayos X como galaxias, pero esto solo puede ser válido cuando son brillantes y no están oscurecidos por el polvo galáctico. Por lo tanto, incluso los objetos con emisión de rayos X que se asemejan a las galaxias del SDSS deben tratarse con precaución. En las muestras a través del disco galáctico, ciertos objetos pueden clasificarse erróneamente como galaxias, cuando en realidad son estrellas combinadas u otras estructuras galácticas extendidas.

Del 4XMM-DR9 en todo el cielo, se identificaron 15,423 objetos de Z21 dentro del área del disco sur del VVV. De estos, Z21 clasificó 1,666 estrellas, 4,031 QSOs y 9,726 galaxias. En esta misma región, Baravalle et al. (2021) generaron el catálogo VVV NIRGC con 5,563 galaxias confirmadas visualmente. Sin embargo, la coincidencia entre estas dos muestras resultó en solo cuatro galaxias en común, además de una galaxia reportada por Schröder et al. (2019).

Los 5,765 objetos de la muestra *NOMatch* fueron inspeccionados visualmente en este trabajo. Esta etapa enfrentó desafíos particulares, ya que implicaba la inspección de regiones con alta extinción interestelar y aglomeraciones estelares. Es evidente que la mayoría de las diez características resumidas en la tabla 4.1 corresponden a regiones galácticas normalmente abarrotadas con varias estrellas distribuidas en todo el área. Cerca del 90.57 % de los casos presentaban alguna “Región normalmente abarrotada”, “Estrella central” o “Región central vacía”. Menos del 1 % mostró indicios de regiones de formación estelar. En el 22.13 % de los casos observamos estrellas centrales, como en los casos de “Estrella central” y “Asociación estelar”. Entre los casos de “Estrella brillante saturada” y “*spikes*” suman el 3.44 %, indicando que la emisión de rayos X podría provenir de alguna de estas fuentes. Los casos de “Región normalmente abarrotada” y “Región central vacía” representan el 73 % de todos los casos. La ausencia de galaxias en el régimen del NIR no necesariamente implica una falta de galaxias en estas últimas regiones, sino que sugiere que puede ser difícil detectarlas, especialmente en áreas con mayor extinción. Las gala-

xias pueden ser detectadas por el telescopio XMM-Newton pero permanecen indetectables en las bandas del NIR. Por lo tanto, el resultado más significativo es que **el 99.95 % de las fuentes clasificadas como galaxias por Z21 en el área de estudio, no son galaxias.**

Además, utilizamos los resultados de relevamientos en diferentes longitudes de onda para realizar una inspección pancromática visual. Definimos cinco “Zonas de Interés” basadas en la distribución de la muestra *galXray*. Se compararon imágenes de cada área en diferentes longitudes de onda, desde radio hasta rayos X (figuras 5.8 a 5.12). Por otro lado, mostramos las altas probabilidades de que estas zonas contengan galaxias, como se observa en la figura 5.2 y la tabla 5.1. Basándonos en estas imágenes podríamos concluir que estas zonas de interés corresponden más a la emisión de estructuras galácticas extendidas en lugar de galaxias individuales. Las fuentes clasificadas como galaxias por Z21 que pertenecen a las zonas de interés son parte de estructuras galácticas, como remanentes de supernovas o regiones de formación estelar. Se necesitan más estudios en estas regiones para comprender completamente su naturaleza y características.

Los conjuntos de datos desbalanceados son comunes en el aprendizaje automático y el reconocimiento de patrones (Lemaître et al. 2017), lo que puede comprometer el proceso de aprendizaje. La mayoría de los algoritmos de aprendizaje automático convencionales esperan una distribución equilibrada de clases o un costo de clasificación errónea (He & Garcia 2009). Este problema puede ser especialmente crítico al intentar distinguir entre galaxias y no galaxias en bajas latitudes galácticas, donde los números de estrellas y asociaciones son extremadamente importantes. Los clasificadores convencionales a menudo priorizan la minimización en la tasa de error general, lo que puede resultar en un sesgo hacia la clase mayoritaria y, por lo tanto, en una clasificación inexacta. La métrica más común utilizada en la clasificación es la precisión, que mide la proporción de predicciones correctas con respecto al número total de muestras de entrada (Blagec et al. 2020). Sin embargo, en la ZoA, donde se enfrentan conjuntos de datos no balanceados, como lo es al clasificar galaxias y no galaxias, el *accuracy* por sí solo puede no ser la métrica óptima. En Daza-Perilla et al. (2023), donde se clasifican galaxias de no-galaxias en el disco galáctico, utilizamos la métrica de *F1-score* como una mejor opción, ya que considera tanto la cantidad como la calidad de la clasificación. En Z21 no abordan este problema, aplicando métodos de clasificación automática sobre los datos, además de usar el *accuracy* para evaluar la eficiencia de los algoritmos.

Respondiendo a las preguntas mencionadas anteriormente, **¿Es posible que no fuesen**

detectadas estas galaxias por un telescopio tan potente como el telescopio VISTA?, Es posible, hay que considerar que tanto en las categorías de *Región normalmente abarrotada* y *Región central vacía*, puede existir una galaxia que ni el telescopio VISTA ni el XMM–Newton sean capaces de resolver. Sin embargo, la evidencia encontrada durante esta Tesis apunta a que las fuentes clasificadas como galaxias por Z21 no son galaxias. **¿Es posible que las HAA hayan generado la más grande base de datos de la ZoA hasta ahora descubierta?** No, al inspeccionar visualmente las fuentes, tanto individualmente como en su conjunto, encontramos que las detecciones clasificadas como galaxias por Z21, corresponden a fuentes galácticas, que pueden ser estrellas o remanentes de supernova.

Nuestro trabajo resalta la importancia de tener un conjunto de entrenamiento representativo, en especial al trabajar en zonas complejas como la ZoA. Un conjunto de entrenamiento adecuado, asegura una clasificación precisa y confiable, mejorando la pureza de la clase positiva en particular y los resultados en general. Esto es de suma importancia cuando se mapea la EGE en bajas latitudes galácticas. De esta manera, es posible confirmar que la característica más importante con respecto a la mala clasificación en nuestra zona de estudio por Z21, corresponde al set de entrenamiento, utilizando SDSS y LAMOST en la ZoA, donde no poseen contrapartes extragalácticas a bajas latitudes.

Capítulo 7

Discusión

En el trabajo de Z21, se presenta un algoritmo de clasificación diseñado para un catálogo que abarca todo el cielo. Sin embargo, al comparar los resultados de Z21 con otros trabajos en la zona, se observan discrepancias notables en la clasificación de fuentes extragalácticas en bajas latitudes galácticas. En primer lugar, Z21 emplea un algoritmo de aprendizaje automático simple para toda la región del cielo, sin tener en cuenta la variación en las características de las zonas, como en el caso de la ZoA. Además, la decisión basada únicamente en el *accuracy* para determinar el mejor modelo de clasificación puede ser problemática cuando se enfrenta a muestras desbalanceadas. En este caso, el desbalance evidente en la muestra refleja la naturaleza intrínseca de la observación a través del disco galáctico, donde la cantidad de estrellas supera considerablemente a la de galaxias.

7.1. Lo crítico que es clasificar objetos en la ZoA

Con el fin de resolver la dinámica del grupo local, resulta crucial desarrollar metodologías para definir la cosmografía del Universo Cercano. La importancia de mapear la ZoA radica en la posibilidad de estimar parámetros cosmológicos como la velocidad peculiar Δv del grupo Local, que posee una gran discrepancia con respecto al dipolo del CMB y 2MRS, y también verificar la coherencia del modelo LCDM (Loeb & Narayan 2008). Una de las posibles explicaciones de estos fenómenos es la existencia de una estructura extragaláctica masiva que estaría oscurecida por nuestra Galaxia. Esta brecha ha sido significativamente reducida con ayuda del catálogo 2MASS Extended Sources (Macri et al. 2019), siendo acotada a $|b| < 5^\circ$. Aún así, existen diversos estudios para analizar esta discrepancia y descubrir otros comportamientos anómalos en las anisotropías del dipolo en la Estructura a Gran Escala, sin encontrar evidencia (Bengaly et al. 2017).

El trabajo de (Jones et al. 2019), donde se aplican aprendizaje profundo para identificar galaxias en la ZoA, con el objetivo de estudiar la eficacia de estas herramientas en esta zona. Utilizan el programa *Galyxi Vysion*, desarrollado para la identificación de galaxias en la ZoA, con el que obtienen un *accuracy* del 91.75 %. Aunque estos resultados representan un gran éxito, considerando los desafíos en obtener datos extragalácticos de buena calidad en la ZoA, los autores mencionan que si bien, tienen buenos resultados, la inspección visual siempre será crucial.

Sin mebargo, el trabajo de Daza-Perilla et al. (2023) es un buen ejemplo de como una buena selección en el set de entrenamiento, junto con las métricas y técnicas utilizadas puede resultar en una buena clasificación. En este trabajo combinan dos enfoques diferentes, aplicando HAA en cada una de ellas. Un enfoque basado en imágenes (*Image-based Sample; IS*) y el otro de su información fotométrica (*Photometry-basd Sample; PS*). Entrenan su algoritmo a partir de la clasificación realizada por Baravalle et al. (2021) en el área del disco norte del relevamiento VVV, con un total de 5,509 galaxias y 74,238 no-galaxias, usando técnicas de balanceo de datos.

El aumento en la cantidad de trabajos dedicados a esta zona, junto con la disponibilidad de datos con mejor resolución y profundidad, no solo mejora nuestra comprensión de los principios cosmológicos fundamentales, sino que también destaca la importancia de contar con un set de entrenamiento representativo de la muestra considerada. Un set adecuado garantiza una clasificación precisa y fiable, mejorando la completitud y pureza de la muestra, lo que es de suma importancia a la hora de mapear la EGE en la ZoA. Esta mejora en la clasificación beneficia diversos campos de la astronomía, como el estudio del dipolo del CMB, la velocidad peculiar de las galaxias del Grupo Local, la isotropía del Universo Local y la comprensión del Gran Atractor.

7.2. Ética e Inteligencia Artificial

Como Stephen Hawking dijo es su libro “*Brief Answers to the Big Questions*: “... el advenimiento de la IA superinteligente sería lo mejor o lo peor que le haya pasado a la humanidad.” ... “Nuestro futuro es una carrera entre el creciente poder de nuestra tecnología y la sabiduría con la que la utilizamos. Asegurémonos de que gane la sabiduría.” (Hawking 2018).

A pesar de que herramientas como estas son usadas desde antes del año 2000, la aparición de herramientas de Inteligencia Artificial de uso cotidiano hizo necesaria la elabora-

ción de la primera norma mundial sobre la ética de la IA de la UNESCO en noviembre de 2021. Esto enfocado a la protección de los derechos humanos y dignidad, fundamentados en la transparencia y equidad. El uso de las IAs en ciencia se encuentra muy presente, sin embargo aún no sabemos como hacer un buen uso de ellas (Sutrop 2019).¹

Estudios de la confiabilidad en los resultados de los algoritmos

A pesar de que las herramientas de clasificación automática son ampliamente utilizados en ciencia, el estudio sobre su confiabilidad solo se ha planteado en los vehículos inteligentes y en el área de salud, específicamente, en el uso de robots para operaciones.

Para poder entender primero si un algoritmo es confiable, hay que entender que la misma palabra de confiabilidad no esta muy clara y su análisis conceptual suele distinguir entre confianza y seguridad, podemos definir su diferencia mediante dos reacciones. La ruptura de la confianza por parte de otra persona nos hace sentir traicionados, mientras que, si simplemente hemos sentido seguridad de una acción o algo (por ejemplo, que un auto encienda) y no se cumplen nuestras expectativas, simplemente nos sentimos decepcionados o frustrados. La mayoría de los filósofos que han escrito sobre la confianza comparten este punto de vista, según el cual la confianza y la seguridad pueden diferenciarse por nuestras reacciones. En lo que difieren es en la respuesta a la pregunta: ¿qué implica la confianza? (Sutrop 2019).

En este sentido, es crucial destacar la labor de la Dra. Genevera Allen, cuyo enfoque se centra en el desarrollo de herramientas estadísticas de aprendizaje automático para ayudar a los investigadores a realizar descubrimientos basados en datos reproducibles, es decir, a dar sentido a los grandes datos. Desde la perspectiva estadística, los científicos deben continuar cuestionando la precisión y la reproducibilidad de los descubrimientos científicos realizados mediante HAA hasta que desarrollen nuevos sistemas computacionales capaces de autoevaluarse.

Allen plantea la siguiente pregunta: ¿Podemos confiar realmente en los descubrimientos que se están realizando actualmente mediante técnicas de aprendizaje automático aplicadas a grandes conjunto de datos?. Su respuesta es “*No sin verificación*”, y señala que se está trabajando en sistemas de aprendizaje automático de próxima generación que evaluarán la incertidumbre y la reproducibilidad de sus propias predicciones. El aprendizaje automatizado es una rama de la estadística y la informática que se ocupa de crear sistemas informáticos capaces de aprender de los datos en lugar de seguir instrucciones explícitas.

¹<https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>

Alles destaca también que gran parte de la atención en el campo de las HAA se ha centrado en el desarrollo de modelos predictivos que permitan a la máquina hacer predicciones sobre datos futuros en función de su comprensión de los datos estudiados. Sin embargo, señala que el problema es que *“Muchas de estas técnicas están diseñadas para hacer siempre una predicción”, “Nunca regresan con ‘No sé’ o ‘No descubrí nada’, porque no están hechos para hacerlo”*.²

En este sentido, es importante nunca olvidar que somos nosotros los que les estamos enseñando a los algoritmos a clasificar o modelar lo que queremos. Entendiendo que, si esperamos una clasificación entre galaxias y estrellas dentro de una zona, debemos esperar que el algoritmo clasifique ciertos objetos como galaxias o estrellas, siendo que no pertenecen a ningún grupo. En este caso, es importante evaluar la realización de una inspección visual para verificar si la técnica utilizada responde de manera esperada para los objetivos científicos, entregando resultados comparables a lo realizado de manera normal, por una persona que aplica no solo su conocimiento científico, si no que también le razonamiento y así criticar y evaluar adecuadamente sus resultados.

²<https://www.eurekalert.org/news-releases/611930>

Bibliografía

- Aharonian, F., Akhperjanian, A. G., Aye, K. M., et al. 2005, *Astron. Astrph.*, 435, L17
- Alpher, R. A., Herman, R., & Gamow, G. 1948, *Physical Review*, 74, 1198
- Amôres, E. B., Sodré, L., Minniti, D., et al. 2012, *Astron. J.*, 144, 127
- An, H., Madsen, K. K., Reynolds, S. P., et al. 2014, *Astrophys. J.*, 793, 90
- Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. 1999, in *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, SIGMOD '99* (New York, NY, USA: Association for Computing Machinery), 49–60
- Arnaboldi, M., Neeser, M. J., Parker, L. C., et al. 2007, *The Messenger*, 127, 28
- Baravalle, L. D., Alonso, M. V., Minniti, D., et al. 2021, *Mon. Not. R. Astron. Soc.*, 502, 601
- Baravalle, L. D., Alonso, M. V., Nilo Castellón, J. L., Beamín, J. C., & Minniti, D. 2018, *Astron. J.*, 155, 46
- Baravalle, L. D., Nilo Castellón, J. L., Alonso, M. V., et al. 2019, *Astrophys. J.*, 874, 46
- Bell, E. F. & de Jong, R. S. 2001, *Astrophys. J.*, 550, 212
- Bengaly, C. A. P., J., Bernui, A., Alcaniz, J. S., Xavier, H. S., & Novaes, C. P. 2017, *Mon. Not. R. Astron. Soc.*, 464, 768
- Bennett, C. L., Halpern, M., Hinshaw, G., et al. 2003, *Astrophys. J. Suppl.*, 148, 1
- Bertin, E. 2011, in *Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX*, ed. I. N. Evans, A. Accomazzi, D. J. Mink, & A. H. Rots, 435

- Bertin, E. & Arnouts, S. 1996, *Astron. Astroph. Supplement (Journal)*, 117, 393
- Bilicki, M. & Chodorowski, M. J. 2010, *Mon. Not. R. Astron. Soc.*, 406, 1358
- Binney, J., Gerhard, O., & Spergel, D. 1997, *Mon. Not. R. Astron. Soc.*, 288, 365
- Binney, J., Gerhard, O. E., Stark, A. A., Bally, J., & Uchida, K. I. 1991, *Mon. Not. R. Astron. Soc.*, 252, 210
- Blagec, K., Dorffner, G., Moradi, M., & Samwald, M. 2020, arXiv e-prints, arXiv:2008.02577
- Blitz, L. & Spergel, D. N. 1991, *Astrophys. J.*, 379, 631
- Bock, D. C. J., Large, M. I., & Sadler, E. M. 1999, *Astron. J.*, 117, 1578
- Bohlin, R. C. & Gilliland, R. L. 2004, *Astron. J.*, 127, 3508
- Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, *Astron. Astroph. Supplement (Journal)*, 143, 33
- Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, *Astron. Astrph.*, 622, A103
- Breiman, L. 2001, *Machine learning*, 45, 5
- Bruzual, G. & Charlot, S. 2003, *Mon. Not. R. Astron. Soc.*, 344, 1000
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *Astrophys. J.*, 533, 682
- Campana, R., Massaro, E., Gasparrini, D., Cutini, S., & Tramacere, A. 2008, *Mon. Not. R. Astron. Soc.*, 383, 1166
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, *Astrophys. J.*, 345, 245
- Cassam-Chenaï, G., Decourchelle, A., Ballet, J., et al. 2004, *Astron. Astrph.*, 427, 199
- Caswell, J. L., Milne, D. K., & Wellington, K. J. 1981, *Mon. Not. R. Astron. Soc.*, 195, 89
- Catelan, M., Minniti, D., Lucas, P. W., et al. 2011, in *RR Lyrae Stars, Metal-Poor Stars, and the Galaxy*, ed. A. McWilliam, Vol. 5, 145
- Chabrier, G. 2003, *Pub. Astron. Soc. Pacific*, 115, 763
- Coldwell, G., Alonso, S., Duplancic, F., et al. 2014, *Astron. Astrph.*, 569, A49

- Colless, M., Dalton, G., Maddox, S., et al. 2001, *Mon. Not. R. Astron. Soc.*, 328, 1039
- Coriat, M., Fender, R. P., Tasse, C., et al. 2019, *Mon. Not. R. Astron. Soc.*, 484, 1672
- Creevey, O. L., Sordo, R., Pailler, F., et al. 2023, *Astron. Astrph.*, 674, A26
- Cross, N. J. G., Collins, R. S., Mann, R. G., et al. 2012, *Astron. Astrph.*, 548, A119
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *Research in Astronomy and Astrophysics*, 12, 1197
- Davis, M., Huchra, J., Latham, D. W., & Tonry, J. 1982, *Astrophys. J.*, 253, 423
- Daza-Perilla, I. V., Sgró, M. A., Baravalle, L. D., et al. 2023, *Mon. Not. R. Astron. Soc.*[[arXiv]2306.07141]
- de Vaucouleurs, G. 1981, *Bulletin of the Astronomical Society of India*, 9, 1
- Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1
- Dirac, P. A. M. 1938, *Proceedings of the Royal Society of London Series A*, 165, 199
- Dodd, R. J. & Brand, P. W. J. L. 1976, *Astron. Astroph. Supplement (Journal)*, 25, 519
- Dreyer, J. L. E. 1888, , 49, 1
- Dreyer, J. L. E. 1895, , 51, 185
- Duplancic, F., Alonso, S., Coldwell, G., et al. 2024, *Astron. Astrph.*, 682, A153
- Ebeling, H., Mullis, C. R., & Tully, R. B. 2002, *Astrophys. J.*, 580, 774
- Efron, B. 1979, *The Annals of Statistics*, 7, 1
- Einasto, M., Saar, E., Liivamägi, L. J., et al. 2007, *Astron. Astrph.*, 476, 697
- Emerson, J., McPherson, A., & Sutherland, W. 2006, *The Messenger*, 126, 41
- Emerson, J. P., Sutherland, W. J., McPherson, A. M., et al. 2004, *The Messenger*, 117, 27
- Erdoğdu, P., Huchra, J. P., Lahav, O., et al. 2006, *Mon. Not. R. Astron. Soc.*, 368, 1515

- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)
- Fazio, G. G., Hora, J. L., Allen, L. E., et al. 2004, *Astrophys. J. Suppl.*, 154, 10
- Fixsen, D. J. 2009, *Astrophys. J.*, 707, 916
- Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *Astron. J.*, 150, 150
- Frank, E., Hall, M., Holmes, G., et al. 2005, *Weka*, ed. O. Maimon & L. Rokach (Boston, MA: Springer US), 1305–1314
- Friedman, J., Hastie, T., & Tibshirani, R. 2000, *The Annals of Statistics*, 28, 337
- Gaensler, B. M. & Slane, P. O. 2006, *Annual Review of Astronomy & Astrophysics*, 44, 17
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *Astron. Astrph.*, 595, A1
- Galdeano, D., Coldwell, G., Duplancic, F., et al. 2022, *Astron. Astrph.*, 663, A158
- Galdeano, D., Ferrero, G. A., Coldwell, G., et al. 2023, *Astron. Astrph.*, 669, A7
- Galdeano, D., Pereyra, L., Duplancic, F., et al. 2021, *Astron. Astrph.*, 646, A146
- Geller, M. J. & Huchra, J. P. 1989, *Science*, 246, 897
- Green, D. A. 2019, *Journal of Astrophysics and Astronomy*, 40, 36
- Harrison, F. A., Craig, W. W., Christensen, F. E., et al. 2013, *Astrophys. J.*, 770, 103
- Hawking, S. 2018, *Brief answers to the big questions* (Bantam)
- He, H. & Garcia, E. A. 2009, *IEEE Transactions on Knowledge and Data Engineering*, 21, 1263
- Heinz, S., Sell, P., Fender, R. P., et al. 2013, *Astrophys. J.*, 779, 171
- Hemmati, S., Capak, P., Pourrahmani, M., et al. 2019, *Astrophys. J. Let.*, 881, L14
- Henning, P. A., Kraan-Korteweg, R. C., Rivers, A. J., et al. 1998, *Astron. J.*, 115, 584
- Hewett, P. C., Warren, S. J., Leggett, S. K., & Hodgkin, S. T. 2006, *Mon. Not. R. Astron. Soc.*, 367, 454

- Hubble, E. 1934, *Astrophys. J.*, 79, 8
- Hubble, E. P. 1926, *Astrophys. J.*, 64, 321
- Hubble, E. P. 1936, *Realm of the Nebulae*
- Huchra, J. P., Macri, L. M., Masters, K. L., et al. 2012, *Astrophys. J. Suppl.*, 199, 26
- Hughes, J. P., Slane, P., Posselt, B., et al. 2010, *The Astronomer's Telegram*, 2771, 1
- Irwin, M. J., Lewis, J., Hodgkin, S., et al. 2004, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 5493, *Optimizing Scientific Return for Astronomy through Information Technologies*, ed. P. J. Quinn & A. Bridger, 411–422
- Jansen, F., Lumb, D., Altieri, B., et al. 2001, *Astron. Astrph.*, 365, L1
- Jarrett, T. H., Chester, T., Cutri, R., et al. 2000a, *Astron. J.*, 120, 298
- Jarrett, T. H., Chester, T., Cutri, R., et al. 2000b, *Astron. J.*, 119, 2498
- Johnston, H. M., Soria, R., & Gibson, J. 2016, *Mon. Not. R. Astron. Soc.*, 456, 347
- Jones, D., Schroeder, A., & Nitschke, G. 2019, *arXiv e-prints*, arXiv:1903.07461
- Jones, D. H., Read, M. A., Saunders, W., et al. 2009, *Mon. Not. R. Astron. Soc.*, 399, 683
- Kirshner, R. P., Oemler, A., J., & Schechter, P. L. 1978, *Astron. J.*, 83, 1549
- Kocevski, D. D., Ebeling, H., Mullis, C. R., & Tully, R. B. 2007, *Astrophys. J.*, 662, 224
- Kogut, A., Lineweaver, C., Smoot, G. F., et al. 1993, *Astrophys. J.*, 419, 1
- Kolatt, T., Dekel, A., & Lahav, O. 1995, *Mon. Not. R. Astron. Soc.*, 275, 797
- Kollatschny, W., Meusinger, H., Hoefft, M., et al. 2021, *Astron. Astrph.*, 652, A24
- Kraan-Korteweg, R. C. & Huchtmeier, W. K. 1992, *Astron. Astrph.*, 266, 150
- Kraan-Korteweg, R. C. & Lahav, O. 2000, , 10, 211
- Kuncheva, L. I. & Rodríguez, J. J. 2007, in *Multiple Classifier Systems: 7th International Workshop, MCS 2007, Prague, Czech Republic, May 23-25, 2007. Proceedings 7*, Springer, 459–468

- Lahav, O., Naim, A., Buta, R. J., et al. 1995, *Science*, 267, 859
- Lauberts, A. 1982, ESO/Uppsala survey of the ESO(B) atlas
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *Mon. Not. R. Astron. Soc.*, 379, 1599
- Lemaître, G., Nogueira, F., & Aridas, C. K. 2017, *Journal of Machine Learning Research*, 18, 1
- Loeb, A. & Narayan, R. 2008, *Mon. Not. R. Astron. Soc.*, 386, 2221
- Logan, C. H. A. & Fotopoulou, S. 2020, *Astron. Astrph.*, 633, A154
- Lopes de Oliveira, R., Lima Neto, G. B., Mendes de Oliveira, C., Janot-Pacheco, E., & Motch, C. 2006, *Astron. Astrph.*, 459, 415
- Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, *Research in Astronomy and Astrophysics*, 15, 1095
- Macri, L. M., Kraan-Korteweg, R. C., Lambert, T., et al. 2019, *Astrophys. J. Suppl.*, 245, 6
- Marchant Cortés, P., Nilo Castellón, J. L., Alonso, M. V., et al. 2024, arXiv e-prints, arXiv:2403.03098
- Marrese, P. M., Marinoni, S., Fabrizio, M., & Altavilla, G. 2019, *Astron. Astrph.*, 621, A144
- Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2017, *Astrophys. J.*, 841, 111
- Masters, D. C., Stern, D. K., Cohen, J. G., et al. 2019, *Astrophys. J.*, 877, 81
- Mather, J. C. 1982, *Optical Engineering*, 21, 769
- McIntyre, T. P., Henning, P. A., Minchin, R. F., Momjian, E., & Butcher, Z. 2015, *Astron. J.*, 150, 28
- Mills, B. Y., Slee, O. B., & Hill, E. R. 1961, *Australian Journal of Physics*, 14, 497
- Minniti, D., Lucas, P. W., Emerson, J. P., et al. 2010, , 15, 433
- Mountrichas, G., Buat, V., Yang, G., et al. 2021, *Astron. Astrph.*, 646, A29

- Nilson, P. 1995, VizieR Online Data Catalog, VII/26D
- Parker, Q. A., Phillipps, S., Pierce, M. J., et al. 2005, *Mon. Not. R. Astron. Soc.*, 362, 689
- Pearson F.R.S., K. 1901, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2, 559
- Peebles, P. J. E. 1980, *The large-scale structure of the universe*
- Penzias, A. A. & Wilson, R. W. 1965, *Astrophys. J.*, 142, 419
- Phillips, C. J., Deller, A., Amy, S. W., et al. 2007, *Mon. Not. R. Astron. Soc.*, 380, L11
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, *Astron. Astrph.*, 536, A1
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, *Astron. Astrph.*, 641, A1
- Radburn-Smith, D. J., Lucey, J. R., Woudt, P. A., Kraan-Korteweg, R. C., & Watson, F. G. 2006, *Mon. Not. R. Astron. Soc.*, 369, 1131
- Ramatsoku, M., Verheijen, M. A. W., Kraan-Korteweg, R. C., et al. 2016, *Mon. Not. R. Astron. Soc.*, 460, 923
- Rodriguez, J., Kuncheva, L., & Alonso, C. 2006, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28, 1619
- Romani, R. W., Wong, J., Di Lalla, N., et al. 2023, *Astrophys. J.*, 957, 23
- Sadeh, D., Meidav, M., Wood, K., et al. 1979, *Nature*, 278, 436
- Said, K., Kraan-Korteweg, R. C., Jarrett, T. H., Staveley-Smith, L., & Williams, W. L. 2016, 462, 3386
- Saito, R. K., Minniti, D., Dias, B., et al. 2012, *Astron. Astrph.*, 544, A147
- Salim, S., Lee, J. C., Janowiecki, S., et al. 2016, *Astrophys. J. Suppl.*, 227, 2
- Schapire, R. E. 2013, in *Empirical inference* (Springer), 37–52
- Schlafly, E. F. & Finkbeiner, D. P. 2011a, *Astrophys. J.*, 737, 103
- Schlafly, E. F. & Finkbeiner, D. P. 2011b, *Astrophys. J.*, 737, 103
- Schlafly, E. F., Green, G. M., Lang, D., et al. 2018, *Astrophys. J. Suppl.*, 234, 39

- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *Astrophys. J.*, 500, 525
- Schröder, A. C., Flöer, L., Winkel, B., & Kerp, J. 2019, *Mon. Not. R. Astron. Soc.*, 489, 2907
- Schröder, A. C., Mamon, G. A., Kraan-Korteweg, R. C., & Woudt, P. A. 2007, *Astron. Astrph.*, 466, 481
- Sciama, D. W. 1953, *Mon. Not. R. Astron. Soc.*, 113, 34
- Seldner, M., Siebers, B., Groth, E. J., & Peebles, P. J. E. 1977, *Astron. J.*, 82, 249
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *Astron. J.*, 131, 1163
- Slane, P., Gaensler, B. M., Dame, T. M., et al. 1999, *Astrophys. J.*, 525, 357
- Smith, M. V., van Zee, L., Salim, S., et al. 2021, *Mon. Not. R. Astron. Soc.*, 505, 3998
- Soto, M., Barbá, R., Gunthardt, G., et al. 2013, *Astron. Astrph.*, 552, A101
- Soto, M., Sgró, M. A., Baravalle, L. D., et al. 2022, *Mon. Not. R. Astron. Soc.*, 513, 2747
- Soumagnac, M. T., Abdalla, F. B., Lahav, O., et al. 2015, *Mon. Not. R. Astron. Soc.*, 450, 666
- Storrie-Lombardi, M. C., Lahav, O., Sodre, L., J., & Storrie-Lombardi, L. J. 1992, *Mon. Not. R. Astron. Soc.*, 259, 8P
- Sutrop, M. 2019, *Trames. Journal of the Humanities and Social Sciences*, 23, 499
- Tateishi, D., Katsuda, S., Terada, Y., et al. 2021, *Astrophys. J.*, 923, 187
- Tauber, J. A., Mandolesi, N., Puget, J. L., et al. 2010, *Astron. Astrph.*, 520, A1
- Taylor, M. B. 2005, in *Astronomical Society of the Pacific Conference Series*, Vol. 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert, 29
- Truemper, J. 1982, *Advances in Space Research*, 2, 241
- Truemper, J. 1992, , 33, 165
- Tully, R. B. & Fisher, J. R. 1977, *Astron. Astrph.*, 54, 661

- van den Bergh, S. 2000, The Galaxies of the Local Group
- Vasquez, J., Cappa, C., & McClure-Griffiths, N. M. 2005, *Mon. Not. R. Astron. Soc.*, 362, 681
- Vavilova, I. B., Elyiv, A. A., & Vasylenko, M. Y. 2018, *Russian Radio Physics and Radio Astronomy*, 23, 244
- Vink, J. 2012, , 20, 49
- Voges, W., Boller, T., Dennerl, K., et al. 1996, in *Roentgenstrahlung from the Universe*, ed. H. U. Zimmermann, J.ümper@, & H. Yorke, 637–640
- Voronoi, G. 1908, *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 1908, 198
- Vorontsov-Vel’Yaminov, B. A. & Arkhipova, V. P. 1963, *Trudy Gosudarstvennogo Astronomicheskogo Instituta*, 33, 1
- Webb, N. A., Coriat, M., Traulsen, I., et al. 2020, *Astron. Astrph.*, 641, A136
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, *Astrophys. J. Suppl.*, 154, 1
- Williams, W. L., Kraan-Korteweg, R. C., & Woudt, P. A. 2014, 443, 41
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *Astron. J.*, 140, 1868
- Yang, G., Boquien, M., Brandt, W. N., et al. 2022, *Astrophys. J.*, 927, 192
- Yang, G., Boquien, M., Buat, V., et al. 2020, *Mon. Not. R. Astron. Soc.*, 491, 740
- York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, *Astron. J.*, 120, 1579
- Zhang, Y., Zhao, Y., & Wu, X.-B. 2021, *Mon. Not. R. Astron. Soc.*, 503, 5263

