



INAOE

Caracterización del campo ultra profundo de TolTEC a 1.1, 1.4 y 2.0 mm a través de simulaciones numéricas

por

Daniela Espitia Mosquera

Tesis sometida como requisito parcial para
obtener el grado de

**MAESTRA EN CIENCIAS en la especialidad
de ASTROFÍSICA**

en el

**Instituto Nacional de Astrofísica, Óptica y
Electrónica**

Tonantzintla, Puebla

Supervisada por:

Dra. Itziar Aretxaga

Investigadora - INAOE

Dr. Javier Zaragoza Cardiel

Investigador - INAOE

©INAOE 2023

Derechos reservados

El autor otorga al INAOE el permiso de
reproducir y distribuir copias de esta tesis en su
totalidad o en partes



Abstract

The study of submillimeter galaxies (SMGs) is particularly important due to their contribution to the Cosmic Infrared Background (CIB), which in turn represents approximately 50 % of the background radiation emitted by extragalactic objects (Puget et al., 1996). These are galaxies with high star formation rates ($\text{SFR} \gg 100 M_{\odot}/\text{yr}$) at high redshifts ($z > 1$), and 90 % of their infrared light is due to their large amounts of dust. The TolTEC camera is a 3-band imaging polarimeter and has been installed on the Large Millimeter Telescope Alfonso Serrano (LMT). The TolTEC project will undertake a series of 10 Legacy Surveys. One of the first four of these is the Ultra Deep Survey (UDS) of galaxies, which aims to cover an area of 0.8 square degrees at 1.1 mm confusion limit.

This master's thesis presents the results of simulations and the subsequent reduction of synthetic observations (using the TOLTECA and CITLALI packages) to characterize observations of deep fields under TolTEC/LMT conditions. The most recent cosmologically motivated galaxy catalog by Nava-Moreno et al. (2019) is used as input, which includes a simulation of the potentially observable population of Dusty Star Forming Galaxies (DSFGs) with TolTEC. Several simulations were performed during this work. These simulations include the observational effects of sky scanning, atmospheric (both variable and static) cleaning, and detector response.

Thus, from the obtained scientific products, sources are extracted using the PYBDSF (Python Blob Detector and Source Finder) package to build the catalog of detected sources in the three observation bands. By comparing the input and output catalogs, the flux loss, false detection rate, positional uncertainties, radial errors and their expected distribution, the boosting factor, and an analysis of the completeness of the catalogs are estimated.

The results suggest that a raster mapping pattern, with a telescope scan speed of $50''/s$ at altazimuth reference frame, can be an effective strategy for studying fields such as the proposed for UDS. However, due to the operational cost of a single observation, increasing the speed could be a viable option, although with the drawback of losing information from less bright sources in the fields.

The simulations reveal a flux loss due to the TOLTECA simulation process. This affects the boosting results. Additionally, it is observed that artifacts arise in images when variable atmospheres are considered in larger fields, resulting in a higher false detection rate.

The position errors (ϵ) for a raster mapping pattern simulation, with a scanning speed of 50 arcsec/s and sources detected at a $\text{SNR} \geq 3.5$ level are $\sim 0.61_{-0.32}^{+0.64}$ arcsec at 1.1 mm. The results are to the analytical values predicted by Ivison et al. (2007). The simulation considering a Lissajous mapping patter has a radial astrometry problem which results in position errors of $\sim 1.6_{-0.6}^{+1.5}$ arcsec at 1.1 mm band for $\text{SNR} \geq 3$ detection cuts. The completeness analysis of the catalogs shows that for the 1.1 mm band, the catalog is 40% complete at ~ 0.1 mJy ($\sim 4\sigma$) and 74% complete at ~ 0.2 mJy ($\sim 8\sigma$). In the 1.4 mm band, the catalog is 30% complete at ~ 0.08 mJy ($\sim 4\sigma$) and 74% complete at ~ 0.15 mJy ($\sim 8\sigma$). Finally, in the 2.0 mm band, 41% is found at ~ 0.06 mJy ($\sim 4\sigma$) and 0.82% at ~ 0.11 mJy ($\sim 8\sigma$).

These results have contributed to the development of TOLTECA and CITLALI, as they have allowed for the identification and characterization of both their shortcomings and advantages. This is crucial in order to trace the sources of error within the packages and ensure they are ready when real data becomes available.

Resumen

El estudio de la población de galaxias sub-milimétricas (SMGs) es especialmente importante debido a su contribución al Fondo Cósmico Infrarrojo (CIB), que a su vez representa $\sim 50\%$ de la luz de fondo radiada por objetos extragalácticos (Puget et al., 1996). Son galaxias con alta formación estelar ($\gg 100 M_{\odot}/\text{año}$) a altos corrimientos al rojo ($z > 1$) y el 90% de su luz en el infrarrojo se debe a sus grandes cantidades de polvo. En el Gran Telescopio Milimétrico Alfonso Serrano (GTM) ha sido instalada la cámara de continuo TolTEC, la cual cuenta con 7716 detectores y observa de forma simultánea en las bandas de 1.1, 1.4 y 2.0 milímetros. A la fecha se han aprobado 4 censos de legado, entre los cuales se encuentra el Censo Ultra Profundo (UDS, *Ultra Deep Survey*) de galaxias que espera cubrir un área de 0.8 grados cuadrados hasta el límite de confusión del telescopio en la banda de 1.1 mm. La combinación de TolTEC y el GTM permitirán censar el cielo con una resolución de 5 segundos de arco en la banda de 1.1 mm y una velocidad hasta 100 veces mayor de lo que el GTM era capaz de hacer con AzTEC.

Esta tesis de maestría presenta los resultados de la simulación y posterior reducción de observaciones realizadas con TolTEC/GTM, utilizando los paquetes TOLTECA y CITLALI, diseñados específicamente para este instrumento. El objetivo es caracterizar las observaciones de campos profundos en las condiciones de TolTEC/GTM. Se usa como entrada el más reciente catálogo de galaxias de Nava-Moreno et al. (2019), el cual se generó a partir de una simulación cosmológicamente motivada de la población de galaxias polvorientas formadoras de estrellas (DSFGs, *Dusty Star Forming Galaxies*) potencialmente observables con TolTEC. Durante la realización de este trabajo, se llevaron a cabo varias simulaciones, sin embargo, se presentaron los resultados para cuatro de estas. Estas simulaciones incluyen los efectos observacionales de barrido

del cielo, limpieza de la atmósfera (tanto variable como estática) y respuesta de los detectores.

Así, a partir de los productos de ciencia obtenidos, se extraen las fuentes mediante el paquete PYBDSF (*Python Blob Detector and Source Finder*) para construir el catálogo de fuentes detectadas en las tres bandas de observación. Comparando los catálogos de entrada y salida, se estima la recuperación de la densidad de flujo, el porcentaje de fuentes falsas, las incertidumbres de posición, errores radiales y la distribución esperada de los mismos, el factor de *boosting* y un análisis en la completez de los catálogos.

Los resultados sugieren que un patrón de cartografiado tipo *raster*, con una velocidad de telescopio de $50''/s$ y un marco de referencia altacimutal, puede ser una estrategia efectiva para el estudio de campos como el propuesto para el UDS. No obstante, dado al costo operacional de una sola observación, aumentar la velocidad podría ser una opción viable, aunque con el inconveniente de perder información de las fuentes más débiles en los campos.

Se observa una pérdida de flujo en todas las simulaciones, que es debida a un defecto en el paquete de simulación por TOLTECA. Esto afecta los resultados de *boosting*. Además, los resultados indican que en campos muy extensos se presentan artefactos en las imágenes cuando se consideran atmósferas variables, lo que aumenta el porcentaje de detecciones falsas en esos casos.

Los errores de posición, ϵ , revelan que, para la simulación de un patrón de cartografiado *raster* con velocidad de barrido 50 arcsec/s el ϵ medidos para detecciones con $\text{SNR} \geq 3.5$ pueden ser de $\sim 0.61_{-0.32}^{+0.64} \text{ arcsec}$ en la banda de 1.1 mm, comparables con los resultados analíticos propuestos por Ivison et al. (2007). En el caso del cartografiado Lissajous se observa un problema en la astrometría en forma radial, presentándose errores de posición de $\sim 1.5_{-1.5}^{+0.6} \text{ arcsec}$ para cortes en detección de $\text{SNR} \geq 3.5$. El análisis de la completez de los catálogos muestra que, para la banda de 1.1 mm, el catálogo está completo en un 40 % en $\sim 0.1 \text{ mJy}$ (4σ) y 74 % completo en $\sim 0.2 \text{ mJy}$. En la banda de 1.4 mm en $\sim 0.07 \text{ mJy}$ (4σ) el catálogo está completo a un 30 % y a $\sim 0.23 \text{ mJy}$ está 87 % completo. Finalmente en la banda de 2.0 mm el 41 % se encuentra en ~ 0.06 (4σ) y el 90 % a $\sim 0.15 \text{ mJy}$.

Los resultados han contribuido al desarrollo de TOLTECA y CITLALI, al identificar y caracterizar los defectos y las ventajas de ambos. Esto es importante para

rastrear las fuentes de error en los paquetes y asegurarse de que estén listos para cuando se procesen y analicen los datos reales.

Agradecimientos

En principio, quiero agradecer profundamente a mis padres Estella y David por haberme apoyado durante toda mi vida y estar ahí para mí. A mi cuñado José, mi hermana Diana y mi sobrina Naty por ser parte esencial de este proceso, sin mi familia nada de esto hubiese sido posible. También agradezco a Fabi por estar ahí para mí y apoyarme en los momentos más difíciles. Quiero agradecer especialmente a mis asesores la Dra. Itziar Aretxaga y al Dr. Javier Zaragoza por haberme guiado en este camino, por la paciencia y la dedicación que pusieron durante todo este trabajo, por estar siempre dispuestos a resolver mis dudas y ayudarme a cuestionar y desarrollar todas las ideas durante este trabajo, sus consejos me permitieron crecer personal y académicamente. Agradezco a mis compañeros Luisa, Nela, Eli, Marc, Fer, Josue y Edgar por su constante apoyo y amistad invaluable. Agradezco a los doctores Alfredo Montaña, Daniel Rosa y Manuel Zamora por aceptar amablemente ser parte de mi jurado sus preguntas y comentarios ayudaron a mejorar este trabajo.

Quiero agradecer a todo el equipo de TolTEC, por hacerme parte de sus reuniones y permitirme crecer académicamente, especialmente a Grant Wilson, Zhiyuan Ma, y Michael McCrackan por estar dispuestos a resolver amablemente mis dudas.

Agradezco a CONAHCYT por la beca brindada y al INAOE por ofrecerme los espacios de trabajo y a mis profesores por sus enseñanzas. Por último, me agradezco a mí misma por no rendirme y ser persistente en aquello en lo que me apasiona.

Dedicatoria

*A mis padres Luz Estella y David, mi motivación más grande
y a mi monita que sigues iluminando desde el cielo.*

Índice general

Abstract	I
Resumen	III
Dedicatoria	IX
1. Introducción	1
1.1. Galaxias Submilimétricas	2
1.2. Corrección k-negativa	3
1.3. Instrumentos que estudian SMGs	5
1.4. La nueva cámara TolTEC	6
1.5. Objetivo y alcance de este trabajo	9
2. Simulación y reducción de las observaciones	11
2.1. TolTECA	11
2.2. Citlali	17
2.2.1. Estructura general de Citlali	18
2.2.2. Paralelización	19
2.2.3. Clases lali y beammap	20
2.2.4. Reducciones de Ciencia y observaciones de calibración	20
2.2.5. Etapas de la reducción de las simulaciones	22
3. Caracterización de las observaciones simuladas	29
3.1. Productos de ciencia	30
3.1.1. Simulación	31

3.1.2.	Reducción	33
3.1.3.	Eliminación de los picos	35
3.1.4.	Eliminación de la atmósfera	36
3.1.5.	<i>Kernel</i>	37
3.1.6.	Mapas de peso	39
3.1.7.	Co-suma	39
3.1.8.	Elaborador de mapa con el método “Naive”	40
3.2.	Estimación del ruido presente en los campos	45
3.2.1.	Caracterización del ruido en los mapas	46
3.3.	Extracción del catálogo de fuentes a partir de los datos reducidos	48
3.3.1.	Catálogo de fuentes obtenido con PyBDSF	48
3.3.2.	Procesamiento con PyBDSF de los archivos FITS obtenidos con Citlali	50
3.4.	Incertidumbre en la posición	58
4.	Resultados	61
4.1.	Asignación de fuentes	63
4.2.	Detecciones falsas	66
4.3.	Errores de posición	70
4.4.	Análisis de la recuperación de densidades de flujo	77
4.5.	Completez de los catálogos	85
5.	Conclusiones y trabajo futuro	91
A.		95
A.1.	Ejemplo de archivo de configuración YAML del modulo de simulación de TolTECA	95
A.2.	Ejemplo de archivo de configuración YAML de Citlali	96
B.	Palabras claves para el archivo de configuración del simulador de TolTEC	101
C.	Simulaciones adicionales	107
C.1.	Tabla de errores en la fracción de posición para la simulación 1	107

C.2. Simulación 2: <i>raster</i> con atmósfera estática	109
C.3. Simulación 3: velocidad de barrido 3 arcmin/s	125
C.4. Resultados simulación Lissajous	142
Bibliografía	159

Introducción

Uno de los puntos claves en el entendimiento de la formación de galaxias y la evolución del Universo, se encuentra en la realización de censos de galaxias en distintas longitudes de onda. Estos censos en conjunto permiten obtener una visión completa y detallada de la diversidad y las propiedades de las galaxias. Con el instrumento FIRAS (*Far InfraRed Absolute Spectrophotometer*) en el satélite COBE (*Cosmic Background Explorer*) se observó por primera vez el Fondo de Radiación Infrarrojo (CIB, *Cosmic Infrared Background*) (Puget et al., 1996; Fixsen et al., 1998). Sin embargo, al contrastar los resultados con los obtenidos por IRAS (*Infrared Astronomical Satellite*), hecho con galaxias cercanas (Neugebauer et al., 1984), se encontró que el Universo emite una cantidad de radiación IR (infrarroja) comparable con la del óptico y ultravioleta.

Con IRAS se descubrió una población de galaxias locales ($z \leq 0.2$) con $\sim 6\%$ de la emisión IR en el Universo local (Sanders et al., 1988). Estas galaxias se les conoce como Galaxias IR Luminosas, LIRG ($L_{\text{IR}} \sim 10^{11} L_{\odot}$) y ULIRG ($L_{\text{IR}} \sim 10^{12} L_{\odot}$) que presentan tasas de formación estelar para las LIRGs sobre $50 M_{\odot}/\text{año}$ y para las ULIRGs sobre $100 M_{\odot}/\text{año}$ (Sanders et al., 1988). Así, según los resultados de COBE, la mitad de la actividad de formación estelar se presenta a $z \geq 1$, por lo que la mejor predicción posible es que debe de existir una población de galaxias con altas tasas de formación estelar y que están embebidas en grandes cantidades de polvo.

A finales de los 90s, las primeras generaciones de censos extragalácticos en longitudes de onda submilimétricas revelaron la existencia de una nueva población de galaxias polvorientas, ópticamente oscurecidas y con una intensa tasa de formación estelar (Smail et al., 1997; Barger et al., 1998; Hughes et al., 1998). Se encuentran a

altos z (Bian et al., 2016) y dominan la luz de fondo extragaláctica en longitudes de onda desde el lejano infrarrojo hasta el milimétrico (FIR-mm). A estas galaxias se las conoce como Galaxias Submilimétricas (SMG, *Sub-Millimeter Galaxies*), que usualmente se observan a $250 \mu\text{m} - 2 \text{ mm}$, y se encuentran dentro del grupo de galaxias polvorientas formadoras de estrellas (DSFGs, *Dusty Star Forming Galaxies*) (Coppin et al., 2005; Karim et al., 2013; Casey et al., 2014; Zavala et al., 2017).

1.1. Galaxias Submilimétricas

Las SMGs se encuentra principalmente a $z \geq 1$ y pueden entenderse como análogos de los ULIRG (Sanders & Mirabel, 1996) pero a altos z (Bian et al., 2016). Las SMGs presentan una mediana en la distribución de corrimientos al rojo entre $2 < z < 3$ (Chapman et al., 2003; Aretxaga et al., 2003), es decir, cuando el Universo tenía unos 2.5 Gyr. Estas galaxias tienen altas luminosidades en el infrarrojo de $L_{\text{IR}} \geq 10^{12} L_{\odot}$, lo que revela altas tasas de formación estelar (SFR, Star Formation Rate) $\text{SFR} \gg 100 M_{\odot}/\text{yr}$. El 90% de su luz en el IR se debe a la re-radiación de la luz absorbida por el polvo (Smail et al., 2000). Todo esto puede sugerir que las SMGs son progenitoras de las galaxias masivas ($> 10^{12} M_{\odot}$) que observamos hoy en día (Blain et al., 2002; Scott et al., 2008).

El polvo, el mayor de los problemas en observaciones de longitudes de onda ópticas/UV, constituye la fase sólida del medio interestelar (ISM) y no puede ser estudiado directamente. Sin embargo, gracias a su interacción con el campo de radiación, se puede deducir que los granos de polvo son muy pequeños, del orden de nanómetros a micras, y están hechos de silicatos y compuestos de carbono (Draine, 2009).

Dado que las SMGs tienen altas tasas de formación estelar, embebida por enormes cantidades de polvo, las estrellas tipo OB más jóvenes y con temperaturas de decenas de miles de Kelvin, son las principales responsables del campo de radiación en el óptico/UV. Así toda esa cantidad de radiación la absorben los granos de polvo, calentándolos de modo tal que la energía se re-emite como emisión térmica de continuo en el IR (8-1000 μm), esto muestra que la emisión en el IR de una galaxia es un trazador de la población estelar más joven.

Las SMGs además exhiben grandes masas dinámicas y estelares, así como un

agrupamiento y una densidad similar de galaxias (Tacconi et al., 2008; McLure et al., 2013). Además, el rango submilimétrico permite investigar moléculas y compuestos orgánicos presentes en estas galaxias, lo cual proporciona pistas sobre los procesos químicos que pueden tener lugar en las etapas tempranas del Universo.

1.2. Corrección k-negativa

La distribución de la densidad de flujo por longitud de onda o frecuencia emitida por un objeto se conoce como Distribución Espectral de Energía (SED, *Spectral Energy Distribution*). La forma de la SED depende de los mecanismos de emisión del objeto. En las SMGs la emisión térmica del polvo es un mecanismo predominante y se modela como una distribución de cuerpo negro modificado, para la cual la densidad de flujo a una frecuencia ν se calcula como:

$$S_\nu = \left[1 - e^{-\left(\frac{\nu}{\nu_0}\right)^\beta} \right] \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{kT}} - 1}, \quad (1.2.1)$$

donde β es el índice de emisividad espectral con valores típicos de 1.5–2 (Casey et al., 2014), ν_0 es la frecuencia donde la emisión se convierte en ópticamente delgada o igual a la unidad¹ (Draine & Li, 2007), T la temperatura de la fuente emisora (en este caso la temperatura del polvo), h y k son las constantes de Planck y Boltzmann, respectivamente.

Cuando se realizan observaciones en longitudes de onda visibles, se espera que a medida que un objeto cercano se aleja a distancias mayores, su brillo aparente se reduzca considerablemente en comparación con la observación inicial. Sin embargo, en el rango del submm-mm, este efecto intuitivo no se presenta de la misma manera. A medida que incrementa el z , la SED de las galaxias en el marco de referencia del observador se desplazan hacia longitudes de onda (λ) más grandes, al mismo tiempo que su densidad de flujo disminuye. Al observar en el intervalo submm-mm, ese desplazamiento en λ provoca que el pico de la SED se acerque a la λ de observación. Teniendo en cuenta que las SMGs se encuentran a un alto z , el observar la emisión más cerca del pico se contrarresta con la atenuación de la emisión provocada por la

¹También puede ser escrita como $\tau(\nu) = (\nu/\nu_0)^\beta$

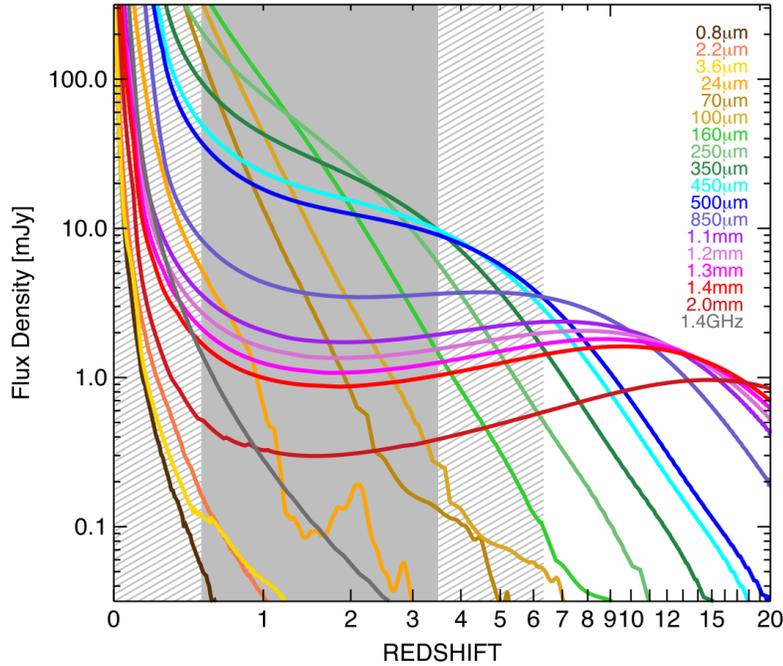


Figura 1.1: Cambio esperado en la densidad de flujo observada con el corrimiento al rojo para una plantilla de una Distribución Espectral de Energía (SED) de una DSFG con una luminosidad fija de $10^{12.5} L_{\odot}$. Se muestran los resultados de observar en diferentes longitudes de onda. Nótese que en rangos submm-mm la densidad de flujo se mantiene relativamente constante en valores de $z \sim 1 - 10$. Tomado de Casey et al. (2014).

mayor distancia a la que se encuentran estas galaxias. La Figura 1.1 se muestra el cambio en la S_{ν} observada con el z para una plantilla SED de una DSFG con una luminosidad fija de $10^{12.5} L_{\odot}$ en varias longitudes de onda que van desde el óptico hasta el radio (1.4 GHz).

Entonces, si se pone una galaxia con la misma luminosidad a corrimientos al rojo $z = 1 - 8$, la densidad de flujo permanece aproximadamente constante e incluso mayores que a corrimientos al rojo algo menores (véase la curva para 2 mm en la Figura 1.1) en estas longitudes de onda. A esto se lo conoce como la corrección k negativa.

En general, la corrección k se aplica al flujo de un objeto desplazado al rojo para trasladarlo a un marco de referencia en reposo (Casey et al., 2014). La terminología ‘negativa’, que está basada en el concepto de magnitudes, se refiere a que la densidad

de flujo aumenta con el z . Así la densidad de flujo S_ν de una SMGs con luminosidad bolométrica infrarroja L_{IR} observada en una frecuencia ν y una SED intrínseca f_ν , se expresa como (Blain et al., 2002):

$$S_\nu = \frac{1+z}{4\pi D_L^2} \frac{f_\nu(1+z)}{\int f_{\nu'} d\nu'} L_{IR} \quad (1.2.2)$$

siendo D_L la distancia de luminosidad a un corrimiento al rojo z .

De hecho, las observaciones submm-mm tienen una corrección *muuy negativa* como se ve en la Figura 1.1. Esta característica representa una oportunidad para hacer accesible el estudio del Universo a alto z .

1.3. Instrumentos que estudian SMGs

Las razones mencionadas previamente, fueron suficientes para impulsar los primeros censos realizados con SCUBA (*Submillimetre Common-User Bolometer Array*) a 850/450 μm . A raíz de esta motivación, se desarrollaron nuevos instrumentos con el fin de llevar a cabo una variedad de censos de las SMGs, por ejemplo, con la cámara de SCUBA a 850/450 μm en el telescopio James Clerk Maxwell (JCMT) de 15 metros (Smail et al., 1997; Barger et al., 1998, 1999; Holland et al., 1999; Hughes et al., 1998; Eales et al., 1999, 2000; Cowie et al., 2002; Scott et al., 2002; Smail et al., 2002; Borys et al., 2003; Serjeant et al., 2003; Webb et al., 2003; Wang et al., 2004; Coppin et al., 2006; Knudsen et al., 2008; Zemcov et al., 2010) y MAMBO (Max-Planck Millimeter Bolometer, Kreysa et al. (1998)) a 1.25 mm, en el telescopio de 30 m del *Institut de radioastronomie millimétrique* (IRAM) (Bertoldi et al., 2007; Greve et al., 2004). Con estos instrumentos se realizaron censos con resolución angular de $\theta_{850\mu\text{m}} = 14.5$ arcsec y $\theta_{1.25\text{ mm}} = 10.7$ arcsec con sensibilidades de $rms_{850\mu\text{m}} = 0.5 - 2.7$ mJy/haz y con áreas de ~ 1300 arcmin² (Ivison, 2001). En esta generación de censos se develó la población de SMGs con $S_{850\mu\text{m}} \gtrsim 5$ mJy. Los centenares de SMGs descubiertas hasta entonces se entendieron como los análogos a altos z de los ULIRGs.

Estos descubrimientos incentivaron el desarrollo de instrumentos y observatorios, alrededor del 2005, que pudiesen hacer censos en el submm-mm cubriendo áreas mayores del cielo. Con ese propósito, se construyó AzTEC (*Astronomical Thermal Emission Camera*, Wilson et al. (2008)) en la banda de 1.1 mm que operó en el JCMT

(Perera et al., 2008; Austermann et al., 2010), en el *Atacama Submillimeter Telescope Experiment* ASTE-10m, (e.g. Aretxaga et al. (2011); Scott et al. (2010); Scott et al. (2012)) y en el Gran Telescopio Milimétrico (GTM) en su configuración de 32 metros (e.g. Zavala et al. (2015)). Otro ejemplo para resaltar es LABOCA (*Large Apex BOlometer CAmera*, Siringo et al. (2009)) que opera a a $870 \mu\text{m}$ en el *Atacama Pathfinder Experiment* (APEX-12m, Güsten et al. (2006); Weiß et al. (2009)).

1.4. La nueva cámara TolTEC

Los avances tecnológicos han permitido construir mejores cámaras para estudiar el Universo en el rango submm-mm. En diciembre del 2021 se instaló en el GTM TolTEC un polarímetro de imagen que observa de forma simultánea en las bandas de 1.1, 1.4 y 2.0 mm. El FWHM para sus tres bandas son 5, 6.3 y 9.5 arcsec, respectivamente (Wilson et al., 2020). TolTEC cuenta con 7716 Detectores de Inductancia Cinética (KIDs, por sus siglas en inglés, *Kinetic Inductance Detector*, Austermann et al. (2018)), distribuidos en 4012, 2532 y 1172 KIDs en las tres bandas de operación². TolTEC es el resultado de una colaboración científica entre la Universidad de Massachusetts, la Universidad Estatal de Arizona, el Instituto Nacional de Astrofísica, Óptica y Electrónica, la Universidad de Cardiff, el Instituto Nacional de Estándares y Tecnología, la Universidad de Northwestern y la Universidad de Michigan.

El Gran Telescopio Milimétrico Alfonso Serrano (GTM, Hughes et al. (2020) es el radio telescopio de plato único que opera a 1 mm más grande y poderoso de su tipo en el mundo. Está diseñado específicamente para observaciones astronómicas en el rango de longitudes de onda de 0.85 a 4 mm. Es un proyecto de colaboración entre México y Estados Unidos de América, siendo el instrumento científico más grande y complejo construido en México. Se encuentra a poco menos de 4600 metros sobre el nivel del mar, en la cima del volcán Sierra Negra. La combinación de un telescopio como GTM (en su configuración de 50 metros) y una cámara de última generación como TolTEC va a permitir una velocidad de cartografiado sobre $10 \text{ deg}^2/\text{mJy}^2/\text{hr}$ (más de 100 veces más rápido que lo que el GTM era capaz de hacer con AzTEC) y una resolución que no ha sido alcanzada en ninguna otra instalación de longitudes de

²Ver: <http://lmtgtm.org/telescope/instrumentation/instruments/toltec/?lang=en>

onda milimétrica de plato simple.

Se planea llevar a cabo una serie de 10 censos de legado cuyos resultados harán públicos con el fin de explotar al máximo sus datos científicos. A continuación se describen los primeros 4 censos aprobados, y cuyas áreas y profundidades esperadas se indican en la Figura 1.2:

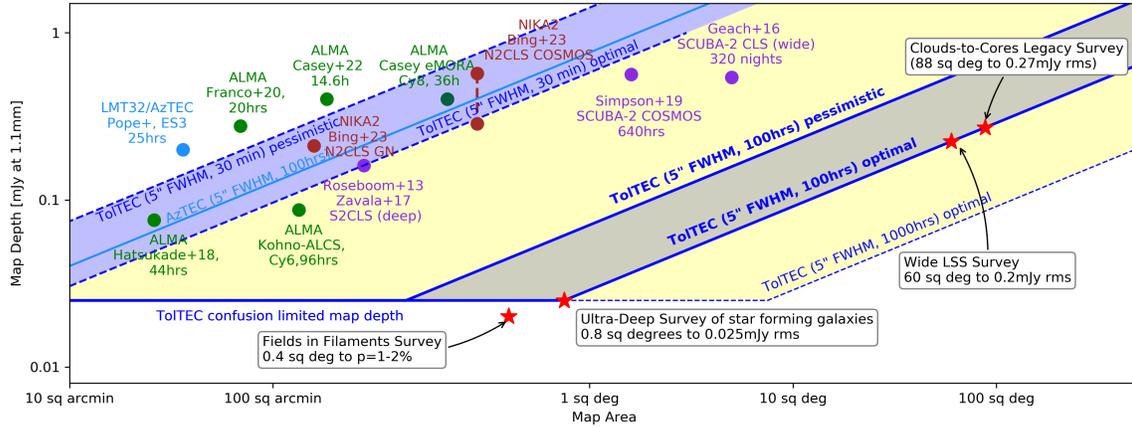


Figura 1.2: Profundidades alcanzadas por distintos censos comparadas a las que se esperan alcanzar con los cuatro censos de legado de TolTEC. (Imagen obtenida en comunicación directa con I. Aretxaga).

- Censo de la estructura a gran escala (LSS):** Será un censo que busca estudiar la distribución y la prevalencia de las ULIRGs y las SMGs en un área entre 40-60 deg² (dependiendo de la velocidad de cartografiado). Se espera alcanzar profundidades en el mapa de 0.25, 0.14 y 0.12 mJy/haz en las bandas 1.1, 1.4 y 2.0 mm, respectivamente.

El enfoque principal de este censo es investigar las galaxias polvorrientas con las mayores tasas de formación estelar, permitiendo así un estudio más completo de la formación y evolución de galaxias en la red cósmica. Además va a permitir ampliar el panorama actual que está limitado por los volúmenes pequeños de observación. Este censo puede otorgar una visión en el FIR de la actividad de los Núcleos Activos de Galaxias (AGNs). También será posible determinar las propiedades de los halos que albergan galaxias polvorrientas a través de mediciones de agrupamiento y correlación angular.

- **Censo Ultraprofundo de Galaxias (UDS):** Este censo espera cubrir un área de 0.8 deg^2 (dependiendo de la velocidad de cartografiado). Su principal objetivo es estudiar la producción de metales y la formación estelar de galaxias en el límite de confusión de 1.1 mm . Se esperan estudiar las LIRGs, las ULIRGs durante los primeros 3 Gyr de tiempo cósmico. Se espera observar campos cosmológicos de legado como COSMOS y CANDELS y contrastar las observaciones con las bases de datos existentes y poder estimar las masas de polvo.
- **Censo de legado de nubes a núcleos:** Es un censo de la función de masa de los núcleos en varias nubes moleculares de la Vía Láctea. Se espera cubrir un área entre $44\text{-}88 \text{ deg}^2$ y se esperan detectar núcleos con masa $M > 0.11 M_{\odot}$ con una profundidad en los mapas de $rms_{1.1 \text{ mm}} = 0.27 \text{ mJy}$. Se pretenden observar 10 nubes moleculares cercanas para identificar y caracterizar los núcleos de cada una de ellas. Combinando estas observaciones con imágenes de *Spitzer* y *Herschel* de las estrellas jóvenes y el gas difuso presente en esas nubes busca entender cómo la presencia de estrellas o las propiedades del gas difuso pueden correlacionarse con las propiedades de los núcleos.
- **Censo de campos magnéticos en filamentos:** Es un censo de campos magnéticos a través de la polarización de la emisión del polvo en regiones de formación estelar. El censo abarca diferentes regiones dentro de las nubes del Cinturón de Gould a escalas de $\sim 0.1 \text{ pc}$.

Como puede verse, las observaciones de las galaxias submilimétricas (SMG) proporcionan un escenario de gran importancia para comprender el Universo. El estudio de las SMG permite investigar los escenarios de formación estelar al analizar la emisión de polvo y comprender los procesos que desencadenan la formación de nuevas estrellas. Además, se puede explorar la evolución de las galaxias, ya que las SMGs son galaxias jóvenes y distantes que posiblemente sean las progenitoras de las galaxias más masivas actuales.

1.5. Objetivo y alcance de este trabajo

Esta tesis se enfoca en distintas simulaciones de observaciones con la cámara de TolTEC, mediante los paquetes de simulación TOLTECA (Ma et al., 2020) y reducción CITLALI (McCrackan et al., 2022) específicamente diseñados para esta cámara. A lo largo del desarrollo de esta tesis ha habido constantes modificaciones a ambos programas por parte del equipo de TolTEC. Por tanto, resulta crucial llevar a cabo una constante caracterización de los resultados obtenidos en las simulaciones, con el objetivo de identificar tanto las mejoras como las deficiencias de los paquetes. Estas caracterizaciones incluyen aspectos como la influencia de la atmósfera y cómo ésta se comporta en los tipos de cartografiado, velocidad de cartografiado, tiempos de exposición, etc. Los resultados de estas y otras simulaciones han contribuido a mejorar aspectos como limpieza de la atmósfera, conversiones erróneas de flujo y la identificación de problemas con el filtro, la precisión en las posiciones de los mapas, entre otros.

Este trabajo tiene como objetivo principal caracterizar las observaciones de campos profundos en las condiciones de TolTEC/GTM. Para ello, se presentan cuatro simulaciones que utilizan como entrada el catálogo de galaxias cosmológicamente motivado más reciente propuesto por Nava-Moreno et al. (2019). Este catálogo sintético contiene la simulación de la población de galaxias polvorientas con formación estelar, que se basa en una simulación de halos de materia oscura (dentro del contexto de N-cuerpos), donde se agregan galaxias y se determinan algunas de sus propiedades físicas (Rodríguez-Puebla et al., 2015, 2017) y las densidades de flujo potencialmente observables con TolTEC/GTM.

En este trabajo, se realizan simulaciones de observaciones del catálogo de (Nava-Moreno et al., 2019) con TOLTECA, y se reducen posteriormente con CITLALI. Con los productos de ciencia, se emplea PYBDSF para extraer y analizar las propiedades de las fuentes detectadas. Las simulaciones incluyen la presencia de una atmósfera variable y estática, así como diferentes tipos de cartografiado, como raster en dos velocidades distintas y uno Lissajous.

Se estudian varios aspectos de los resultados obtenidos, como la eliminación atmosférica, la recuperación/pérdida de flujo, los errores de posición debido a posibles

errores sistemáticos en los paquetes, el *boosting* de las fuentes detectadas, la complez y la tasa de detección falsa. Estos resultados son importantes para identificar las fortalezas y debilidades del paquete y asegurarse de que esté a punto en el momento que se dispongan datos reales y estos se analicen de forma fidedigna.

Simulación y reducción de las observaciones

La cámara TolTEC (Wilson et al., 2020), con 7716 Detectores de Inductancia Cinética (en adelante KIDs, *Kinetic Inductance Detectors*) representa un incremento de un orden de magnitud con respecto a su predecesora, la cámara de bolómetros AzTEC, que contaba con 144 elementos (Wilson et al., 2008), lo que se traduce en una mayor velocidad de cartografiado y volumen de datos. Esto implica la necesidad de incrementar los requerimientos computacionales y de memoria para el manejo de los datos científicos, por lo que, para hacer frente a esos nuevos retos, se ha desarrollado un conjunto de herramientas de software que administran, reducen y visualizan los datos de TolTEC. Así pues, se construye el software TolTECA¹ (dentro del cual está incluido la parte de simulación de observaciones) y el *pipeline* de reducción Citlali². El siguiente capítulo se basa en los desarrollos mostrados por Ma et al. (2020) y McCrackan et al. (2022).

2.1. TolTECA

El software desarrollado para TolTEC incluye el entorno de alto nivel desarrollado en python, TolTECA, que se encarga de la gestión automática de los productos de datos, la elaboración de las simulaciones de los datos en crudo (o sin procesar) de TolTEC utilizando modelos de detectores realistas y un ruido atmosférico. Adicional-

¹Disponible en: <https://github.com/toltec-astro/tolteca>.

²Disponible en: <https://github.com/toltec-astro/citlali>.

mente proporciona una visualización web, construida en el paquete de python, *Plotly Dash*, para hacer diagnósticos rápidos. En lo subsecuente se hará una descripción de las características más importante de TolTECA y Citlali que fueron utilizadas para la realización de este trabajo.

Dentro de TolTECA está incluido el *pipeline* de reducción Citlali, que convierte las señales sin procesar del instrumento en datos listos para ciencia, así como el software infraestructural que gestiona las tareas de reducción y los productos de datos (Ma et al., 2020). Un esquema general de la arquitectura de TolTECA se muestra en la Figura 2.1.

El sub-módulo `tolteca.simu`, dentro de TolTECA, simula la adquisición de datos. Este sub-módulo genera datos sintéticos y archivos de metadatos utilizados como entradas para Citlali. La simulación es importante para caracterizar el proceso de reducción antes de obtener datos reales del cielo. En este apartado se describe el módulo de simulación utilizado en esta tesis, mientras que los detalles de otros paquetes de software de TolTECA se pueden consultar en el artículo de referencia Ma et al. (2020).

En principio, como entradas al simulador se ingresan ya sean archivos FITS (*Flexible Image Transport System*), como por ejemplo mapas de *Herschel* o *Spitzer*, o una tabla con un formato de `astropy` que contenga, por ejemplo, información sobre un catálogo de fuentes puntuales. En este trabajo se usa la simulación cosmológicamente motivada de galaxias polvorientas formadoras de estrellas de Nava-Moreno et al. (2019). También dentro del simulador se consideran modelos de los detectores y sus propiedades de ruido, una tabla de propiedades idealizada, y un patrón de cartografiado para crear las series temporales sin procesar.

El simulador tiene la capacidad de considerar un modelo de atmósfera estático o variable. En el caso de la atmósfera estática, se pueden tener en cuenta diferentes cuartiles, como el 25 %, 50 % o 75 %. Por ejemplo, el valor del 25 % corresponde a las mejores condiciones de observación, donde solamente el 25 % de las noches en el GTM son las mejores³. Para la atmósfera variable, se ha integrado el módulo de simulación atmosférica TOAST⁴ dentro del módulo de TolTECA. TOAST simula bloques tridimensionales de la atmósfera utilizando parámetros meteorológicos y los integra utilizando un formalismo teórico descrito por Church (1995).

³Ver: https://www.lmtobservatory.org/help/user/calculation_toltec.

⁴Ver : <https://toast-cmb.readthedocs.io/en/latest/index.html>.

La simulación atmosférica de TOAST produce ruido correlacionado, el cual varía tanto en tiempo como en espacio. Su implementación en el módulo de TolTECA se calibra a partir de las propiedades de ruido promedio de los mapas de haz para proporcionar una simulación lo más cercana posible a las condiciones esperadas en el sitio del Gran Telescopio Milimétrico Alfonso Serrano (GTM). Con dichas simulaciones se realizan la eliminación del ruido atmosférico y la elaboración de mapas (McCrackan et al., 2022).

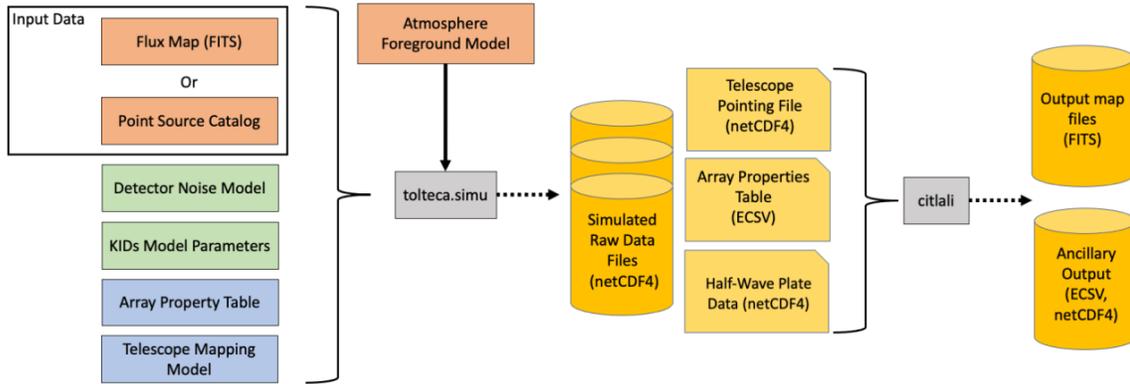


Figura 2.1: Arquitectura general de TolTECA. Obtenido de McCrackan et al. (2022).

Para entender cómo es que funciona la adquisición de los archivos de datos sin procesar es necesario comprender, al menos de forma general, cómo opera la cámara. TolTEC cuenta con 7716 KIDs distribuidos en los 3 arreglos correspondientes a las tres frecuencias de observación (272, 214 y 150 GHz). Cada detector puede ser representado como un inductor dentro de un circuito resonador LC (L: bobina y C: condensador) en miniatura, cuya frecuencia de resonancia cambia como una función de la potencia óptica incidente, de ahí su nombre detectores de inductancia cinética (Austermann et al., 2018).

Para capturar los datos, se utiliza una técnica de cartografía de barrido de cielo debido a la sensibilidad espacial y temporal de los KIDs de TolTEC. El sistema de referencia utilizado para el movimiento de la cámara puede ser Azimut/Elevación o Ascensión recta/Declinación, siguiendo un patrón de cartografiado predefinido que puede ser:

- **Patrón Lissajous:** Es útil para mapas de 4 a 6 minutos de arco en diámetro, y

ofrece buena cobertura y sensibilidad uniforme. No requiere invertir el sentido del barrido, lo que evita grandes aceleraciones y permite una eficiencia de tiempo de observación del 100 %. Se está investigando un nuevo patrón de cartografiado llamado **doblo Lissajous** que permite cubrir escalas mayores de 4 a 6 minutos de arco en los mapas tipo Lissajous.

- **Raster:** El telescopio se mueve a una velocidad constante definida por el usuario y toma una fila completa de datos y gira en el borde de la zona del cielo requerida y desciende una cantidad de pasos definida por el usuario y escanea en dirección opuesta de regreso. Este proceso se realiza hasta que se cubra el tamaño del mapa. Este tipo de cartografiado es el indicado para mapas grandes (mayores a 30 minutos de arco por lado) en donde se puede tomar ventaja de una velocidad de barrido rápida para llevar la señal astronómica a frecuencias temporales más altas que la contaminación atmosférica residual⁵.
- **Rastajous:** Este modo de cartografiado combina uno de tipo *raster* lento con uno más rápido tipo Lissajous. Como el *raster* se realiza lentamente, los tiempos de regreso al final del barrido se mantienen y por tanto el modo Rastajous presenta un 100 % de eficiencia de tiempo de observación. Este modo de cartografiado se creó relativamente tarde en la era de AzTEC y puede llegar a ser útil para adquirir mapas de escala intermedia.

Una visualización gráfica de estos patrones de cartografiado se puede observar en la Figura 2.2. Así mismo, todos estos patrones de cartografiado pueden ser explorados en el planificador de observaciones de TolTEC⁶.

Con cualquiera de los métodos de cartografiado anteriores, los datos de cada KID se guardan en una frecuencia establecida lo que produce una serie temporal individual para cada detector. Una serie temporal es un archivo de datos sin procesar que se compone de las señales de los KIDs, apuntado del telescopio y señales ambientales, las cuales son almacenadas como una función del tiempo (Scott et al., 2008).

⁵Se debe encontrar un balance entre la velocidad de barrido con el tiempo de respuesta del telescopio (~ 5 segundos). La cámara AzTEC fue capaz de realizar cartografiado en velocidades de $200''/m$, pero esto se vio limitado por la velocidad de lectura de los detectores de AzTEC. Como la velocidad de lectura de los detectores de TolTEC es 8 veces mayor que la velocidad de lectura de los de AzTEC, se espera un límite mucho mayor para la velocidad de barrido.

⁶Disponible en https://toltec.lmtgtm.org/toltec_obs_planner.

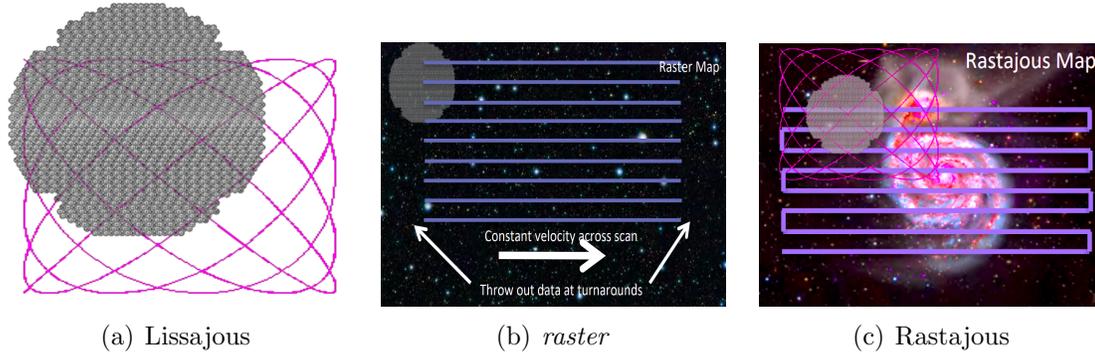


Figura 2.2: Tres estrategias principales de cartografiado disponibles para las observaciones de TolTEC. El esquema aproximadamente circular en color gris representa el arreglo de detectores en una banda, las líneas moradas en (a) representan el patrón de Lissajous, y las líneas azules en (b) representan los barridos por filas en zonas del cielo de interés para un patrón *raster*. La (c) es la representación de la combinación de ambos patrones, Rastajous. Imágenes obtenidas de http://lmtgtm.org/wp-content/uploads/2020/11/Wilson_TolTEC.pdf.

La lectura electrónica en TolTEC usa una aproximación de multiplexación de microondas que admite hasta 1000 detectores o resonadores en cada circuito de lectura, es decir, un máximo de 1000 canales de lectura, a esto se le denomina red. Los detectores de la cámara de TolTEC se dividen en 13 redes donde 7 corresponden al arreglo de 1.1 mm, 4 al de 1.4 mm y 2 al de 2.0 mm (Ma et al., 2020).

El tiempo de muestreo se registra utilizando una señal de paquetes por segundo (PPS) sincronizada con GPS. La multiplexación se utiliza para evitar la superposición de frecuencias de resonancia entre los diferentes KIDs en un mismo arreglo de detectores. Como resultado, TolTEC genera un máximo de 13 archivos de procesamiento de datos ordenados por tiempo (TOD, *Time-Ordered data*) por cada observación. Estos archivos se almacenan en el formato netCDF⁷.

Cada red se lee en paralelo y se crea un archivo de datos sin procesar que contiene las series de datos etiquetadas por tiempo de las señales digitales en fase (I) y en cuadratura (Q) por canal. Esto se utiliza para medir el parámetro de dispersión de la red, representado por el número complejo $S_{21} \equiv I + iQ$. Para caracterizar los detectores, se registra el S_{21} mientras se varía una carga óptica constante a través de

⁷Para más información ver <https://www.loc.gov/preservation/digital/formats/fdd/fdd000332.shtml>.

un rango de frecuencias de prueba f_p centradas en la frecuencia de resonancia f_r de cada detector. El S_{21} medido en función de las frecuencias de prueba sigue el modelo de los KIDs:

$$S_{21}(f_p) = \mathcal{G} \left[\hat{S}_{21}(f_p; f_r, Q_r) \right]$$

donde \hat{S}_{21} es la forma canónica descrita como

$$\hat{S}_{21} = \frac{Q_r}{1 + 2iQ_r \frac{f_p - f_r}{f_r}} \quad (2.1.1)$$

donde Q_r es el factor de calidad del resonador. Introduciendo los siguientes cambios de variables

$$r \equiv \frac{1}{2Q_r} \quad x \equiv \frac{f_p - f_r}{f_r} \quad X \equiv r + ix \quad (2.1.2)$$

la expresión 2.1.1 queda,

$$\hat{S}_{21} = X^{-1} \quad (2.1.3)$$

Por su parte, la expresión \mathcal{G} es una función que está relacionada a la ganancia compleja del circuito de lectura y su expresión parametrizada es:

$$\mathcal{G} \left(\hat{S}_{21}; G, K, M \right) = G\hat{S}_{21} + Kf + M, \quad (2.1.4)$$

siendo G , K y M parámetros complejos. La caracterización del detector implica ajustar un modelo de los KIDs a los datos de frecuencia de barrido y obtener los mejores parámetros del ajuste, como f_r , Q_r , G , K y M . Este proceso se realiza al final de cada barrido mediante el sistema de control de instrumentos (ICS) que utiliza la herramienta `kidscpp` de la librería de Citlali. Los parámetros del modelo se guardan en disco en formato ECSV para su uso posterior.

Tanto el apuntado del telescopio como las series temporales de los detectores se etiquetan en el tiempo y se interpolan linealmente en la frecuencia de muestreo de TolTEC en Citlali. TolTEC también tiene instalado una placa de media onda

de rotación continua (CRHWP, *Continuously Rotating Half-Wave Plate*) como un componente óptico caliente, adicional para las observaciones de polarimetría y reducir el ruido $1/f$ que se ve en los KIDs de nitruro de titanio (Columbro et al., 2019; Essinger-Hileman et al., 2016; Simon et al., 2016). Cuando el CRHWP está en uso (o cuando se simulan observaciones incluyendo polarimetría), se crea un archivo de datos adicional que contiene las posiciones de los ángulos etiquetadas en el tiempo, que será una entrada al *pipeline* de reducción. Esto es un componente muy importante en relación a estudios de polarización, sin embargo, en este trabajo no se realizan simulaciones de polarimetría por lo que no se hará mayor énfasis en este aspecto.

2.2. Citlali

TolTECA implementa un marco de gestión de datos y un submódulo de simulación. La interpretación de la señal de los KIDs, el procesamiento de los TOD y la elaboración de los mapas se realiza con Citlali (McCrackan et al., 2022).

Citlali es el motor de reducción de datos para TolTEC, escrito en C++ y utiliza CMake y Conan para facilitar la compilación multiplataforma. Citlali procesa los datos de TolTEC de manera paralela y confiable, con algoritmos optimizados utilizando la librería Eigen⁸.

Citlali está diseñado para trabajar con información contextual mínima, lo que permite desacoplar la gestión de metadatos del instrumento y la asociación de datos desde el procesamiento estándar de TOD. El programa está completamente paralelizado y no requiere almacenar todos los datos sin procesar en la memoria durante la reducción. Se utiliza la paralelización mediante el código abierto GrPPI (Generic and Reusable Parallel Pattern Programming Interface), desarrollado por del Rio Astorga et al. (2017).

Citlali fue diseñado bajo tres necesidades clave: procesar todos los tipos de datos de TolTEC (ciencia y calibración), realizar reducciones casi en tiempo real de los datos sin procesar y trabajar eficientemente en entornos con memoria limitada o dentro de la configuración de un *cluster* de computación.

⁸Ésta permite almacenar y manipular matrices, y ofrece funciones optimizadas para las Transformadas Rápidas de Fourier (FFT, *Fast Fourier Transform*), convolución de imágenes, interpolación con *splines* y descomposición de valores propios, entre otras. Ver: <http://eigen.tuxfamily.org>

2.2.1. Estructura general de Citlali

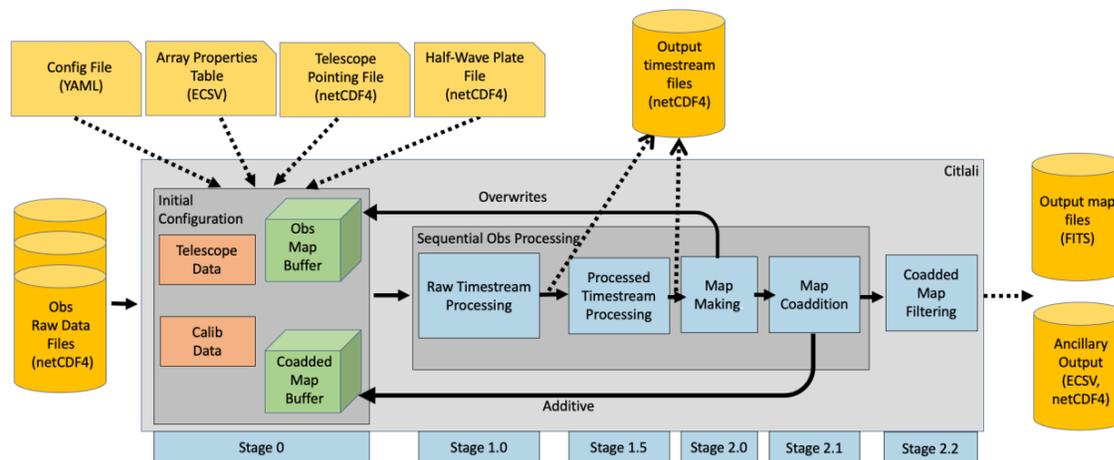


Figura 2.3: Una vista de la estructura general de alto nivel de CITLALI. Se muestran las entradas y salidas. Tomado de McCrackan et al. (2022).

Parte de los archivos de entrada a Citlali son los archivos de datos sin procesar, apuntado del telescopio y datos del plato de media onda (si se usa polarimetría), todos en formato `netCDF4`, como se observa en el esquema de la Figura 2.3.

Además, para las reducciones de ciencia y algunas observaciones de calibración, Citlali requiere información adicional sobre las posiciones en el cielo de cada KID para elaborar mapas y llevar a cabo la calibración de flujo y el proceso de filtrado post-mapa. Esta información se genera a través de la reducción de los datos del mapa de haz y se guarda en una tabla de propiedades del arreglo (APT) en formato ECSV.

Para realizar la configuración que permite ejecutar Citlali se utiliza un archivo de extensión `YAML`. Las etapas de reducción se pueden activar o desactivar y tienen sus propias opciones de control configurables. Se pueden reducir y cosumar múltiples observaciones. Un ejemplo de archivo de configuración se muestra en el Apéndice A.2.

Como se ve en la Figura 2.3, la reducción de datos con Citlali se divide en tres fases: configuración, procesamiento de la observación y reducción posterior para elaborar el mapa. En la Etapa 0 se verifica la información en el archivo de configuración `YAML`. La APT así como cualquier metadato de calibración fotométrica y astrométrica se almacenan dentro de la clase `citlali::RawObs`. En esta etapa no se utilizan los datos sin procesar. Si se requiere la suma de observaciones, se calculan las dimensiones y se

asigna espacio de memoria. Luego, Citlali crea las carpetas para guardar los FITS de cada observación individual y el mapa final, que es la cosuma de la observaciones.

El procesamiento de la observación se realiza de manera secuencial, iterando sobre las instancias de la clase `citlali::RawObs`. Se asigna la memoria necesaria y se reescribe y se reutiliza en cada iteración para reducir el consumo de memoria. Se lee, interpola y alinea la información de tiempo para los TOD sin procesar, el telescopio y los datos de la CRHWP. Como se ve en la Figura 2.3, el procesamiento se divide en sub-etapas: la etapa 1 realiza la reducción inicial de las series temporales antes de la eliminación de la atmósfera, la etapa 1.5 se encarga de la eliminación atmosférica y los cálculos de los pesos de los detectores (calculado como la inversa de la varianza de las muestras sin etiquetar en los TOD para ese detector), y la etapa 2.0 de la elaboración del mapa.

La etapa 1.0 y 1.5 operan en el dominio del tiempo, pero están separadas debido a la necesidad de realizar la reducción en el mapa de haz de manera iterativa sólo entre las etapas 1.5 y 2.0, no en la etapa 1. La etapa 2.0 crea los mapas bidimensionales a partir de la transformación de los TODs y la etapa 2.1 coloca los mapas por observación en el *buffer* del mapa final. Después de reducir todas las observaciones individuales, se escriben en disco y se comienza la siguiente iteración de `citlali::RawObs`. Los mapas finales se generan solo si todas las observaciones han sido reducidas. En la etapa 2.2, se realiza un filtrado adicional en los mapas para mejorar la identificación de las fuentes y corregir el flujo astrofísico. Como resultado, se obtienen mapas filtrados y sin filtrar que se guardan en disco.

2.2.2. Paralelización

Dada la alta velocidad de muestreo de TolTEC, las observaciones generan una gran cantidad de datos que ocupan un volumen importante de memoria. Además, Citlali utiliza memoria extra para los *buffer* de los mapas de cada observación y el mapa final, lo que representa un desafío en términos de computación y memoria. Para abordar esto, Citlali se diseñó para dividir los datos sin procesar en fragmentos temporales más pequeños e independientes, lo que permite reducirlos de forma secuencial o paralela. Esto evita cuellos de botella, ya que se pueden leer y reducir fragmentos temporales mientras otros están en memoria. La paralelización depende de la velocidad de lectura

del disco en relación con el tiempo de cada fragmento temporal individual.

Los fragmentos temporales se almacenan en las Etapas 1 a través de la clase `citlali::RTCData`, y en las Etapas 1.5 y 2.0 a través de `citlali::PTCData`. Estas clases contienen todos los datos necesarios para reducir los fragmentos del TOD, excepto las coordenadas en el cielo de los detectores. El tamaño del fragmento temporal puede ser especificado en el archivo de configuración para los patrones de cartografiado Lissajous, mientras que para los mapas que utilizan un patrón *raster*, Citlali calcula el tamaño basado en cada fila en ese tipo de cartografiado.

2.2.3. Clases `lali` y `beammmap`

Citlali utiliza diferentes algoritmos de reducción que se combinan en diferentes órdenes o se utilizan de forma iterativa según el objetivo de la observación. Para controlar el flujo de las series temporales y la reducción de mapas, Citlali utiliza dos clases de motores de *pipeline*: `citlali::lali` para las reducciones de observaciones de ciencia, apuntado, foco y astigmatismo, y `citlali::beammmap` para el cartografiado del haz. Estas clases heredan de la clase base `citlali::EngineBase`, que contiene los algoritmos de reducción, los *buffers* de mapas y los metadatos del telescopio y la calibración. Las clases `citlali::RTCData` y `citlali::PTCData` son utilizadas como entradas para las clases de reducción `citlali::RTCProc` y `citlali::PTCProc`, que controlan las etapas 1.0 y 1.5 respectivamente.

2.2.4. Reducciones de Ciencia y observaciones de calibración

Las observaciones de calibración utilizan el motor `citlali::lali`, con modificaciones menores en el marco de referencia de coordenadas y un paso adicional de ajuste a una fuente después de la elaboración del mapa. En las observaciones de calibración con TolTEC, se enfoca el astigmatismo en una fuente puntual brillante para medir y corregir desplazamientos en las posiciones del espejo secundario del GTM y las distorsiones del espejo primario. Se ajustan los mapas de Citlali a una gaussiana bidimensional utilizando la librería `Ceres Solver`, y las amplitudes ajustadas de la fuente se utilizan posteriormente en observaciones futuras. Del mismo modo, las observaciones de apuntado se realizan múltiples veces para corregir errores en el

apuntado del telescopio. Las reducciones de calibración se presentan en una tabla en formato ECSV con los parámetros de la fuente ajustada para cada arreglo de detectores. El diagrama de flujo del *pipeline* de esta clase se muestra en la Figura 2.4.

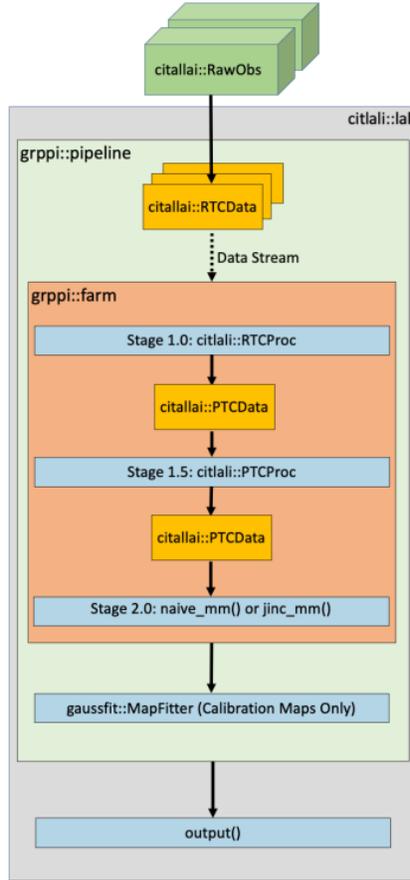


Figura 2.4: Clase motor de *pipeline* citlali:lali

Las reducciones relacionadas al mapa del haz se llevan a cabo mediante un barrido de frecuencia y ajuste de picos de resonancia en los KIDs de TolTEC. La ubicación física de los detectores en cada arreglo debe determinarse, lo cual se logra mediante la observación de una fuente puntual brillante como calibrador. En el mapa del haz, el telescopio sigue un patrón de cartografiado tipo *raster* en coordenadas Azimut-Elevación (Az/El), asegurando que cada detector cruce la fuente en múltiples pasadas.

Para cada detector, se genera un mapa del TOD y se ajusta la fuente a una gaussiana bidimensional para obtener su posición, ancho a mitad de altura (FWHM),

ángulo de rotación y amplitud. Estos parámetros se registran en la APT y se utilizan en observaciones científicas y de calibración posteriores. Las etapas 1.0 a 2.0 se ejecutan en un patrón de transmisión `grppi::pipeline`, que maneja el flujo de datos al leer y devolver cada fragmento temporal desde los datos sin procesar. Los fragmentos temporales se procesan en paralelo utilizando `grppi::farm`, donde un número específico de procesos definidos por el usuario reduce los fragmentos a medida que se reciben del flujo de datos. Cada fragmento se borra de la memoria una vez que su TOD se ha agregado al *buffer* del mapa existente.

En este trabajo, el enfoque se centra en las simulaciones relacionadas con reducciones de datos científicos, por lo que no se profundizará en la clase `citlali::beammap`. Para más detalles, se puede consultar el artículo de McCrackan et al. (2022).

2.2.5. Etapas de la reducción de las simulaciones

El procesamiento de datos por Citlali se lleva a cabo por etapas, mostradas en la Figura 2.5. Estas incluyen la conversión de hasta 7716 lecturas de señal de los KIDs sin procesar en TOD que llevan la señal óptica (Etapa 0), el procesamiento de los TOD y la elaboración de los mapas. Como se indica en la sección 2.1, cada observación de la cámara TolTEC entrega un conjunto de 13 archivos binarios que contienen las series temporales de los KIDs, y un archivo con la información de apuntado del telescopio. Con esto, Citlali empieza su proceso de reducción por etapas:

- **Etapa 0: Procesamiento de los datos de los KIDs**

En esta etapa se transforma el parámetro de dispersión S_{21} (ver sección 2.1) de cada muestra en una unidad proporcional a la intensidad incidente en cada detector, para crear mapas de cielo calibrados en flujo. Esto se logra mediante el cálculo del parámetro de desafinación x utilizando la librería `citlali::kids`. El parámetro x mostrado en la ecuación (2.1.2), mediante la inversión de la ecuación (2.1.4) y (2.1.3). Este parámetro se incluye luego en `grppi::farm` para las clases `lali` y `beammap`.

Las etapas 1-2 (ver Figura 2.5) comprenden el procesamiento de los TOD y la elaboración de los mapas. Esto es implementado en el módulo de la librería central

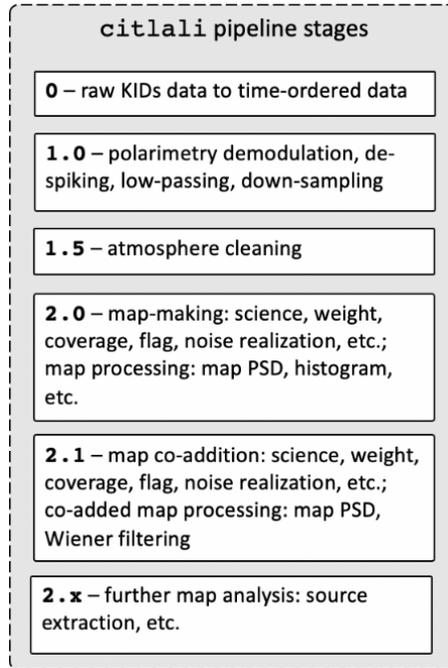


Figura 2.5: Etapas del *pipeline* de reducción de datos de Citlali. Tomado de Ma et al. (2020).

`citlali::lali`. Los algoritmos subyacentes que traducen los TOD en mapas de ciencia finales son re-implementaciones o adaptaciones del *pipeline macana*, desarrollado para la cámara de AzTEC (Perera et al., 2008).

▪ **Etapa 1: Procesamiento de los TOD sin procesar**

- **Despiking:** Durante una jornada de observación, los rayos cósmicos generan picos en los datos temporales de los detectores. Estos picos son significativamente más altos que el promedio de los datos (TOD) y pueden sesgar la media y la varianza utilizadas para los mapas finales. Se identifican los picos al calcular las diferencias entre puntos de datos adyacentes y verificar si superan un umbral basado en desviaciones estándar. Se marca una region del mismo tamaño del filtro pasa-bajo alrededor de cada pico, de modo que no sea usada durante la elaboración del mapa. Este proceso se realiza en cada detector de manera independiente para evitar que esas regiones afecten etapas posteriores.

- ***Pasa-bajo y Pasa-alto:*** Debido a la máxima velocidad de giro del GTM, cualquier señal cuya frecuencia sea mayor que un cierto umbral, no puede tener un origen astrofísico. Para eliminarlas, se aplica un filtro pasa-bajo a los TOD usando una ventana Kaiser-Bessel. Por otra parte, se puede aplicar un filtro pasa-alto para remover cualquier ruido de baja frecuencia.
- ***Reducción del muestreo:*** Dado que la eliminación atmosférica es costosa desde el punto de vista computacional y que TolTEC toma muestras a aproximadamente 488 Hz, sería necesario disponer de un gran volumen de memoria. Para mitigar esta situación, es necesario reducir la muestra o diezmar los TOD por un factor entero, reduciendo efectivamente la tasa de muestreo por el mismo factor.
- ***Fuente sintética:*** Después de procesar los TOD, se genera una serie temporal sintética del mismo tamaño y longitud. Esta serie, conocida como serie temporal del *kernel*, contiene una única fuente puntual en el centro del mapa. Este se procesa del mismo modo que los datos de observación para evaluar los efectos de la eliminación atmosférica en una fuente puntual. Además, se realiza una calibración en flujo multiplicando los TOD por un factor basado en la entrada APT. Posteriormente, se convierten a unidades de densidad de flujo (MJy/sr o mJy/beam) y se aplica una corrección por extinción. El FWHM del *kernel* se utiliza para extraer fuentes, como se discutirá en el Capítulo 3.

■ Etapa 1.5: Reducción de los TOD procesados

- ***Análisis de Componentes Principales (PCA):*** La atmósfera es la principal fuente de ruido en los datos de longitudes de onda milimétricas. Su brillo es mucho mayor que el de las fuentes astrofísicas y varía en tiempo y espacio. Eliminar la señal atmosférica de los TOD es difícil y puede resultar en la pérdida de la emisión extendida captada por los detectores. Se utiliza el análisis de componentes principales (PCA) para suponer que, a escalas temporales cortas, las correlaciones atmosféricas en cada arreglo no cambian significativamente. Los modos propios más grandes corresponden a la señal atmosférica. En cada fragmento temporal, se resta el promedio

de las series temporales y se calcula una matriz de correlación, que se descompone en valores y vectores propios. Los modos propios más grandes se establecen en cero mediante el PCA. La serie temporal sintética también elimina esos modos propios. Los datos identificados en la etapa de eliminación de picos no se incluyen en la matriz de correlación de los detectores y no contribuyen a la limpieza atmosférica.

- **Asignación de pesos a los TOD:** En general, en la mayoría de las observaciones el peso se define como la inversa de la varianza de las muestras que no fueron etiquetadas en los TOD para cada detector. En cada fragmento temporal se calcula el peso de cada detector para ser usados en la etapa 2 que se encarga de la elaboración de los mapas. Aquellas series temporales con pesos que se aparten de la desviación estándar promedio de los detectores, por un margen amplio, se etiquetan para no ser tomados en cuenta en etapas de reducción posteriores.

■ **Etapa 2: Elaboración del mapa:**

En esta etapa, se excluyen los datos etiquetados en el proceso de eliminación de picos, en la APT y en la asignación de pesos a los detectores. Aquí se generan los mapas de señal/flujo, peso/incertidumbre, una fuente puntual sintética y cobertura para cada banda de TolTEC. Se calcula un mapa de relación señal-ruido a partir de los mapas de señal y peso antes de escribir los mapas. Además, se generan mapas de ruido *Jackknife* aplicando el mismo algoritmo utilizado para construir los mapas, multiplicando las muestras ponderadas de los TOD de cada fragmento temporal por un número aleatorio entre ± 1 . Citlali ofrece dos tipos de algoritmos para la generación de mapas.

- **Método con filtrado de Jinc:** En este caso, el TOD se convoluciona con un filtro Jinc truncado. El filtro Jinc tiende a cero en frecuencias espaciales altas, en donde las señales astrofísicas no tienen una contribución, lo que representa la ventaja de quitar el ruido del mapa sin afectar las propiedades de la fuente. Sin embargo, este método es computacionalmente más costoso que el algoritmo descrito a continuación ya que la distancia desde cada píxel del mapa hasta las coordenadas en el cielo de cada detector deben

ser calculadas para cada muestra y multiplicarse por el filtro.

- **Método de Naive:** En este caso, cada mapa es un histograma bidimensional que registra el promedio pesado de todas las muestras de datos que caen dentro de los límites de cada píxel (McCrackan et al., 2022). Así cada muestra contribuye a un solo píxel y no tiene impacto en los píxeles alejados. En esta aproximación se elige una escala angular de píxeles y cada punto de datos en el TOD se agrega en un promedio pesado para cada píxel cuyo límite lo contiene (Ma et al., 2020). A cada punto de datos se le asigna un peso que corresponde a la inversa de la varianza del fragmento temporal del que proviene.

Una de las ventajas de Citlali radica en el hecho de que los mapas se van poblando mediante la paralelización de cada fragmento temporal. De este modo no es necesario reducir toda la serie temporal para posteriormente agregar píxeles al mapa. En esta etapa, se crean de forma simultánea los mapas de señal, cobertura y peso (o incertidumbre). Adicionalmente, para cada mapa de señal se genera un número N , definido por el usuario, de mapas de ruido *Jackknife*.

- **Método de Estimación por Máxima Verosimilitud :** Este tipo de métodos resulta computacionalmente más costoso que los algoritmos implementados en Citlali, descritos anteriormente, pero proporcionan un estimador imparcial de la señal astrofísica. Esto es debido a que operan suponiendo que el flujo de las fuentes astrofísicas no varía ni en espacio ni en tiempo, mientras que la contribución atmosférica sí es variable. Todo esto representa una ventaja dado que permite mejorar la etapa de eliminación atmosférica comparada al algoritmo de PCA mostrado en la etapa 1.5.

Como las series temporales y los metadatos son salidas en las etapas 1 o 1.5 y están parcialmente reducidos y calibrados en flujo, se pueden usar como entradas a aquellos algoritmos de elaboración de mapas con el método de máxima verosimilitud. Algoritmos implementados en *softwares* tales como TOAST (*Time Ordered Astrophysics Scalable*) y el elaborador de mapas

MINKASI⁹ el cual está adaptado a partir del *pipeline* para el Atacama Cosmology Telescope (ACT).

■ Etapa 2.1: Cosuma y filtrado de mapas

- **Cosuma de mapas:** Debido a que muchos proyectos de ciencia requieren de múltiples observaciones separadas en el tiempo, se hace necesario la implementación de una rutina de cosuma estándar luego de que se hayan generado todos los mapas de observación individual. Por su parte, para producir los mapas de ruido cosumados, se seleccionan del conjunto de mapas de ruido, de cada una de las observaciones, para sumarse en cuadratura y ser normalizados mediante los mapas de pesos.
- **Filtro Wiener:** Luego de crear todos los mapas (señal, peso, ruido y *kernel*), Citlali ofrece la opción de aplicar un filtro Wiener optimizado para la recuperación de fuentes puntuales (Scott et al., 2008). Éste utiliza los mapas de fuentes sintéticas cosumadas como plantillas de filtro para corregir por el flujo durante la eliminación atmosférica. El mapa fuente está convolucionado con una modificación rotacionalmente simétrica de la fuente puntual sintética del procesamiento de los TOD del *kernel*. Estas convoluciones se realizan con la transformada rápida de Fourier (FFT), y los mapas finales se re-escalán tomando en cuenta las pérdidas de flujo en la fuente puntual sintética debido al PCA. En el caso de mapas grandes, como las FFTs bidimensionales se deben aplicar iterativamente en las convoluciones, el proceso computacional se vuelve costoso, por este motivo Citlali paraleliza los cálculos de la FFT usando el modulo `Eigen::FFT`. Una de las ventajas de este filtro es que elimina los flujos a grandes escalas, que en general son contribuciones residuales de la atmósfera y adicionalmente minimiza las variaciones entre píxeles a escalas más pequeñas que el tamaño del haz.

Las simulaciones y las reducciones realizadas en este trabajo fueron ejecutadas en dos equipos de supercómputo: el primero de estos, Mextli, se encuentra disponible en

⁹Ver: <https://github.com/sievers/minkasi>.

el INAOE y cuenta con 4 nodos, 1 TB de memoria RAM (256 GB por nodo) y 48 núcleos; el segundo es el equipo de supercómputo de 7 nodos de UMass con una alta velocidad de 100 GbE. Cada uno de los 7 nodos del servidor Dell PowerEdge R640 contiene una placa base de doble-CPU con dos procesadores Intel Xeon Gold 5218 de 16 núcleos, 384 GB de memoria y un adaptador Ethernet Mellanox ConnectX-6 100GB.

Caracterización de las observaciones simuladas

En el capítulo anterior se describen las principales características de los programas de simulación y reducción usados en este trabajo. Este capítulo, se enfoca en mostrar cómo se realizaron en principio las múltiples simulaciones probando distintos patrones de cartografiado, velocidad de telescopio, tiempo de integración, atmósfera variable y estática. Así pues, para cada caso, se muestra cómo se hace la caracterización del mapa simulado final mediante los productos de ciencia entregados luego de todo el proceso de simulación y reducción descrito en el capítulo 2. La Tabla 3.1 muestra el conjunto de archivos de salida que se tienen para cada una de las 3 bandas en las que opera TolTEC. En este capítulo, por simplicidad, se muestran los procedimientos llevados a cabo con imágenes de la banda de 1.1 mm, sin embargo, todos los análisis se realizaron en las 3 bandas que opera TolTEC.

Para la lectura, manipulación, cálculos y análisis sobre los FITS entregados por CITLALI, se realizaron varias rutinas en python3, en especial para modificar los FITS de acuerdo a los requerimientos de entrada para utilizar el código de extracción de fuentes. Este capítulo está organizado así: la sección 3.1 mostrará la forma en cómo se ingresan los parámetros de entrada a la simulación, como sistema de coordenadas, patrones de mapeo, atmósfera estática o variable, velocidad de cartografiado, los cuales fueron discutidos en el capítulo 2, y los respectivos resultados de los productos de ciencia entregados luego de la reducción de esas simulaciones, al igual que la forma en que se hace la representación de píxeles. La sección 3.2 muestra la forma en cómo se utilizan las representaciones de ruido realizadas con el método de *Jackknife* para

caracterizar el ruido presente en las simulaciones, a través del histograma promedio de los histogramas de cada realización. A continuación, la sección 3.3 muestra el procedimiento seguido para extraer las fuentes de los mapas simulados y cuya información se utiliza para caracterizar el campo simulado, como se mostrará en el Capítulo 4.

Nombre de la imagen	# de imágenes	Unidades	Descripción
Mapas de señal: I (op: Q,U)	3(9)	MJy/sr	Mapas de flujo de intensidad total (opcional, los parámetros de Stokes Q y U) para cada uno de los 3 arreglos.
Mapas de peso: I (op: Q,U)	3(9)	(MJy/sr) ⁻²	Mapas de pesos para I (opcional, los parámetros de Stokes Q y U) para cada uno de los 3 arreglos.
Mapas del <i>Kernel</i> I (op: Q,U)	3(9)	Sin unidades	Representaciones filtradas de la PSF o forma del <i>kernel</i>
Mapas de señal a ruido I (op: Q,U)	3(9)	Sin unidades	Resultado de la multiplicación del mapa de señal por la raíz cuadrada del mapa de pesos.
Realizaciones de ruido I (op: Q,U)	3×# mapas usuario	Sin unidades	Mapas I (opcional, los parámetros de Stokes Q y U) de la representación del ruido <i>Jackknife</i> .

Tabla 3.1: Productos de ciencia, dentro del FITS, entregados por CITLALI luego de la reducción de las simulaciones. Entre paréntesis se muestran las opciones para el caso de que se necesiten realizar estudios de polarización.

3.1. Productos de ciencia

Dado que en este trabajo se simulan observaciones con TolTEC de campos ultra profundos, se hace necesario la simulación de múltiples observaciones, en diferentes tiempos, que luego se cosuman resultando en un mapa final con el cual se llevará a cabo el respectivo análisis. Primero, este trabajo usa como entrada al simulador de TolTECA la más reciente simulación de galaxias cosmológicamente motivada de (Nava-Moreno et al., 2019). El catálogo de entrada adoptado consiste en las 33,364

fuentes de una simulación de medio grado cuadrado que toma en consideración el efecto de lentes gravitacionales. El catálogo tiene información, de corrimiento al rojo, ascensión recta ($^{\circ}$), declinación ($^{\circ}$), masa estelar (M_{\odot}), tasas de formación estelar total e infrarroja (M_{\odot}/yr), fracción obscurecida de la tasa de formación estelar f_{obs} , luminosidad infrarroja (L_{\odot}), temperatura de polvo (K) y los flujos (mJy) en las bandas de 500, 850 μm , 1.1, 1.4 y 2.0 mm. Para este trabajo se utiliza la información concerniente a la posición y los flujos para las bandas de 1.1, 1.4 y 2.0 mm.

Para ingresar la información necesaria para realizar las simulaciones hay dos opciones: hacerlo mediante una tabla con formato de `astropy` o mediante una imagen FITS a alta resolución. Durante el desarrollo de esta tesis se encontró que la primera opción tiene la desventaja de ocasionar tiempos de simulación muy prolongados, sin embargo, la segunda reduce considerablemente los tiempos en casi el doble produciendo buenos resultados. Por lo tanto, el catálogo de fuentes puntuales es insertado a una imagen de alta resolución de 0.05 segundos de arco, lo cual resulta en un tamaño total para el FITS de entrada de 64 GB. La imagen de entrada para las 3 bandas, se muestra en la Figura 3.1.

3.1.1. Simulación

El archivo FITS de alta resolución con la información del catálogo sintético de DSFGs se usa como entrada para iniciar la simulación. Un ejemplo de esto se muestra en el apéndice A.1. En principio se debe indicar un nombre de carpeta de salida para los resultados de la simulación (ver capítulo 2), elegir entre una atmósfera estática o variable, frecuencia de cartografiado, tiempos de exposición (sólo válido para el caso *raster*), el tiempo de los fragmentos temporales. En este punto, es importante notar que se realizaron múltiples pruebas con la combinación de estos parámetros de entrada durante la realización de esta tesis con el fin de encontrar la mejor combinación de los mismos. Así por ejemplo, para trabajar con la atmósfera en el simulador, se fueron probando distintos valores numéricos para controlar el número de puntos a usar para abarcar el rango de altitud en el que se está simulando la observación y otros parámetros que controlan la cantidad de *relleno* que se adiciona alrededor del área que se está observando. Una explicación detallada de las variables de entrada para ejecutar el módulo de simulación de TolTECA se puede encontrar en el Apéndice

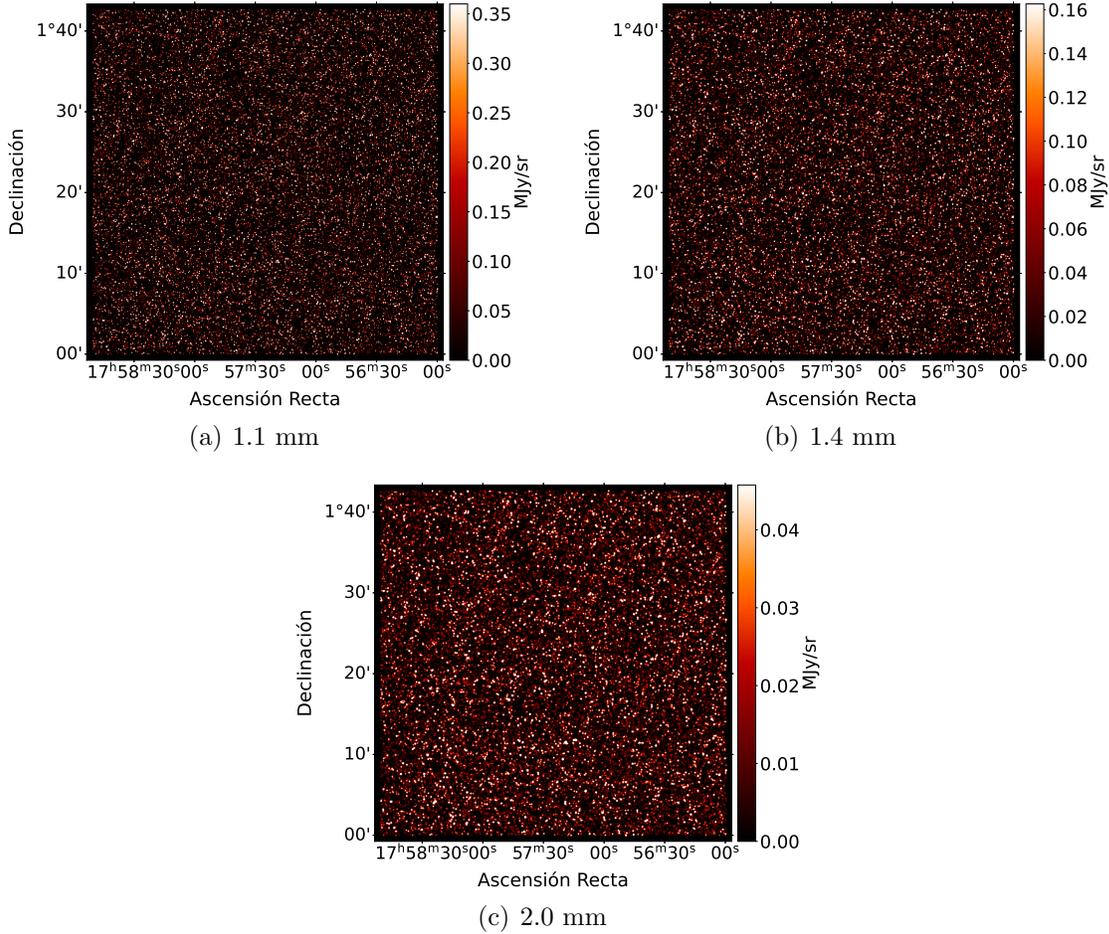


Figura 3.1: Mapa que representa a una alta resolución de 0.05 segundos de arco el catálogo de Nava-Moreno et al. (2019) en las tres bandas que observa TolTEC, el cual se usa como entrada al módulo de simulación de TolTECA.

B.

Como se mencionó, generalmente son necesarias múltiples observaciones en distintos tiempos para tener un mapa final. Por esta razón es necesario realizar el proceso de simulación en forma paralela para ahorrar tiempo de ejecución. Otro aspecto importante es el tipo de marco de referencia con el cual se hará el patrón de cartografiado. En términos generales, si se fija en azimut y elevación el ángulo de barrido, para un barrido *raster* el mapa cambia constantemente el barrido en $\alpha - \delta$ debido a la rotación del cielo. Si se suman múltiples observaciones se obtendría un excelente entrecruza-

miento (*cross-linking*) que suprime el ruido sistemático en los mapas (Scott et al., 2008).

La imagen de entrada, representada en la Figura 3.1, se encuentra centrada en $(\alpha = 269.355^\circ, \delta = 1.355^\circ)$ ¹. Dado gran tamaño de este campo (0.5 deg^2) lo más conveniente es utilizar un tipo de cartografiado *raster* (si es que se quiere cubrir la mayor área posible). Para ello, se define un ancho de 30 minutos de arco en azimut y para cubrir el tamaño en elevación, se definen un número de 60 barridos del telescopio con un espacio entre cada uno de 0.5 minutos de arco, lo que da un total de 30 minutos de arco cubiertos para esta coordenada. También se fija un tiempo de giro para el telescopio como 5 segundos con una velocidad de barrido de 50 segundos de arco por segundo, lo que implica que cada observación tomaría 40 minutos y 55 segundos en completarse (sin tomar en cuenta los tiempos adicionales)².

Para realizar un mapa co-sumado por ejemplo a partir de 12 observaciones en momentos distintos, tal y como se muestra en la Figura 3.2, se deben realizar 12 simulaciones distintas. Como se comentó en el capítulo 2, cada simulación da como resultado un conjunto de 13 archivos binarios que contienen las series temporales de los 12 arreglos de KIDs así como un archivo con la información de apuntado del telescopio, entre otros que controlan el estado de la simulación. Cada conjunto de archivos en cada una de las 12 simulaciones son almacenados para realizar, a continuación, el procedimiento de reducción.

3.1.2. Reducción

Para la reducción, se utiliza el *pipeline* de CITLALI descrito en la sección 2.2. CITLALI está basado inicialmente en el paquete de reducción, Macana, que se utilizó para los datos de AzTEC. CITLALI está basado en los algoritmos descritos por Scott et al. (2008) y Wilson et al. (2008), de modo que muchas de las explicaciones provistas en lo subsecuente provienen de ambas fuentes. Para las simulaciones mostradas en la Figura 3.2, se utiliza una atmósfera variable (que se hace directamente desde el

¹Todas las simulaciones llevadas a cabo en este trabajo se hicieron sobre estas coordenadas como objetivo.

²El tiempo que tarda una observación tipo *raster* está dada por: $\frac{lenght}{speed} \times n_{scans} + t_{giro} \times (n_{scans} - 1)$. Siendo *lenght* el tamaño del barrido, *speed* la velocidad de barrido, n_{scans} el número de barridos, t_{giro} el tiempo de giro del telescopio.

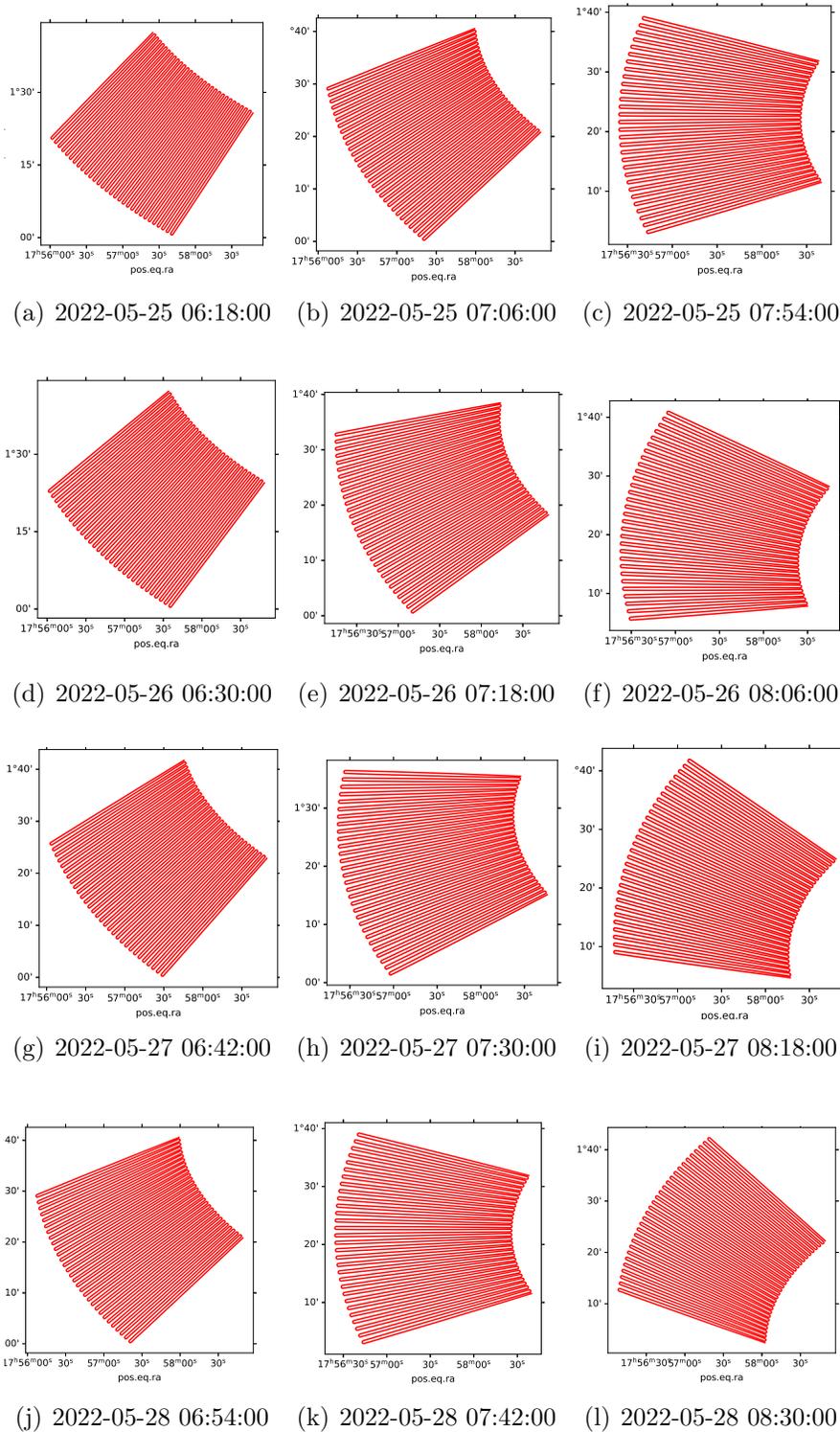


Figura 3.2: Gráficas que muestran el modo de cartografiado tipo *raster* para 12 observaciones realizadas en fechas y horas distintas.

simulador, mediante TOAST). Como se sabe, las fluctuaciones atmosféricas dominan la señal que observan los detectores de TolTEC, por lo que CITLALI intenta eliminar esas fluctuaciones mediante un enfoque de análisis de componentes principales. Esto da como resultado el hecho de que se aplique un filtro espacial a los mapas resultantes que no pueden ser determinados *a priori*³.

En el archivo YAML, que será usado en el proceso de reducción, se define la carpeta de salida de las reducciones, en la cual se almacenarán los productos de ciencia descritos en la Tabla 3.1, para cada una de las observaciones. Adicionalmente, se genera una carpeta que contendrá los productos consumados. Un ejemplo de este archivo se muestra en la sección A.2, en el que se indican el número de nodos a utilizar en la tarea, el tipo de reducción (en este caso de ciencia, pero puede ser mapa de haz, apuntado, etc). La versión del paquete, si se quiere o no agregar algún tipo de corrección a las series temporales o realizar polarimetría, entre otros parámetros de configuración. Muchos de estos parámetros se dejan por defecto, siguiendo las recomendaciones de sus desarrolladores, ya que el paquete se encuentra, de momento, en constante mejora.

3.1.3. Eliminación de los picos

Si bien en esta tesis no se está tratando con observaciones reales, es importante mencionar la relevancia de eliminar datos atípicos en los datos sin procesar de las series temporales. Estas medidas pueden provenir de rayos cósmicos o fallos instrumentales que se identifican como *picos*. La tasa a la cual ocurren se caracteriza para cada instrumento. Los picos en CITLALI se definen como una muestra que está por encima de un valor $n\sigma$ mayor que el valor promedio del fragmento temporal. CITLALI identifica y etiqueta las regiones que cumplen con esta condición, las cuales no se utilizan en los análisis subsecuentes.

La tarea dentro de CITLALI se llama `Despike.h` dentro de la clase `timestream`. En el archivo de reducción se pueden modificar las palabras `sigma` que define el número de desviaciones estándar sobre el cual una muestra se considera un *pico* (en este ejemplo $\sigma = 8$), el `time_constant_sec` que determina el tiempo de decaimiento de un pico,

³Esto no afecta mucho aquellos programas que buscan observar el cielo extragaláctico, sin embargo, sí puede hacerlo para aquellos que buscan medir emisiones extendidas, que están en o por encima de la escala del campo de visión de TolTEC.

o qué tantas muestras alrededor del pico se deben considerar afectadas (en este caso se fija como 0.015 s) y por último el parámetro `window_size` que fija el tamaño de la ventana para hacer el etiquetado si se habilita un filtro pasa-bajo/pasa-alto (fijado en este caso como 32). Los picos se identifican y etiquetan de forma iterativa para no ser tomados en cuenta en el proceso elaboración del mapa. Para cada pico, se marca una región a cada lado: si no se tiene habilitado un filtro pasa-bajo/pasa-alto se usa el `decay_length` que se basa en la tasa de muestreo, el valor del `time_constant_sec`, el valor del pico y se suaviza la región sin marcar; si sí se habilita, se utiliza el tamaño de la ventana de ese filtro. Sin embargo estos datos no pueden ser simplemente descartados ya que el siguiente paso, la eliminación atmosférica, requiere operaciones con matrices, por lo que todas las series temporales deberán tener el mismo número de muestras.

Por lo anterior, es necesario reemplazar esos datos eliminados con valores sintéticos plausibles. Para esto CITLALI sigue los siguientes pasos (basados en los desarrollos de Scott et al. (2008)):

- Realiza una interpolación lineal antes y después de las muestras marcadas a ambos lados de la región. Ésta es la línea base para el detector marcado.
- Se realiza lo mismo para los detectores que no están marcados. Se escala por la capacidad de respuesta y se utiliza para hacer una base lineal de “modelo de cielo” que caracteriza a los detectores buenos.
- Se reemplaza el dato marcado por un valor sintético plausible dado por: $\text{fake_data} = \text{sky_model} + \text{responsivity}(\text{det}) + \text{linear_offset}$.
- Y finalmente se agrega un ruido gaussiano con desviación estándar estimada a partir de los detectores sin marcar.

3.1.4. Eliminación de la atmósfera

La señal proveniente de las fluctuaciones atmosféricas dominan el fondo en las observaciones de la población de galaxias polvorientas (DSFGs). CITLALI elimina esta contribución mediante la técnica de PCA (*Principal Component Analysis*); dado que, en general, las fuentes tenues no están correlacionadas entre los detectores del

arreglo y que por su parte, la atmósfera sí lo está y en todas las escalas de interés, la técnica de PCA es una buena estrategia para eliminarla ya que utiliza el grado de correlaciones para distinguir entre la atmósfera y las fuentes.

Se construye una matriz de covarianza a partir del número de detectores (N_{bolo}) y del número de serie temporal (N_{time}), después de la eliminación de picos para cada barrido y luego se descompone en valores propios. Las amplitudes relativas de los valores propios resultantes representan el grado de correlación de las señales del detector para el modo descrito por el respectivo vector propio. Como las fuentes puntuales débiles y el ruido del detector no se correlacionan entre múltiples detectores, entonces no se hallarán en los modos cuyos valores propios sean grandes. La atmósfera, las fluctuaciones en los detectores entre otras señales de modo común dominan la varianza correlacionada con potencias en modos cuyos valores propios son grandes, por lo que el objetivo es identificar los modos cuyos valores propios sean los más grandes.

Decidir cuántos valores propios eliminar de los datos es una tarea un tanto arbitraria (Scott et al., 2008)⁴. En esta tesis se realizan distintas pruebas quitando una cantidad de valores propios dada por el parámetro `n_eig_to_cut` (en el archivo de reducción) y los resultados serán presentados en el siguiente capítulo para ver qué tanta importancia tiene este valor en la extracción del catálogo de fuentes. La Figura 3.3(a) muestra el mapa de señal reducido cuando no se elimina la atmósfera y en la Figura 3.3(b) cuando sí se hace.

3.1.5. *Kernel*

Si bien, el proceso anterior permite eliminar la fuente de ruido más significativa en las simulaciones, durante todo este proceso de eliminación atmosférica surgen consecuencias para las fuentes puntuales tenues ya que presentan potencia a bajas frecuencias espaciales por lo que, debido al uso de las técnicas PCA, no es posible desacoplar por completo la señal de la atmósfera en esas fuentes y se va a presentar cierta atenuación en el mapa de fuentes resultante (Scott et al., 2008). Por esta razón es necesario trazar todos estos efectos en un perfil de respuesta de una fuente puntual

⁴En este mismo trabajo se encontró empíricamente que eliminar valores propios mayores a 2.5σ de la media (en logaritmos base 10, calculada a partir de la distribución de valores propios), produce el mejor resultado.

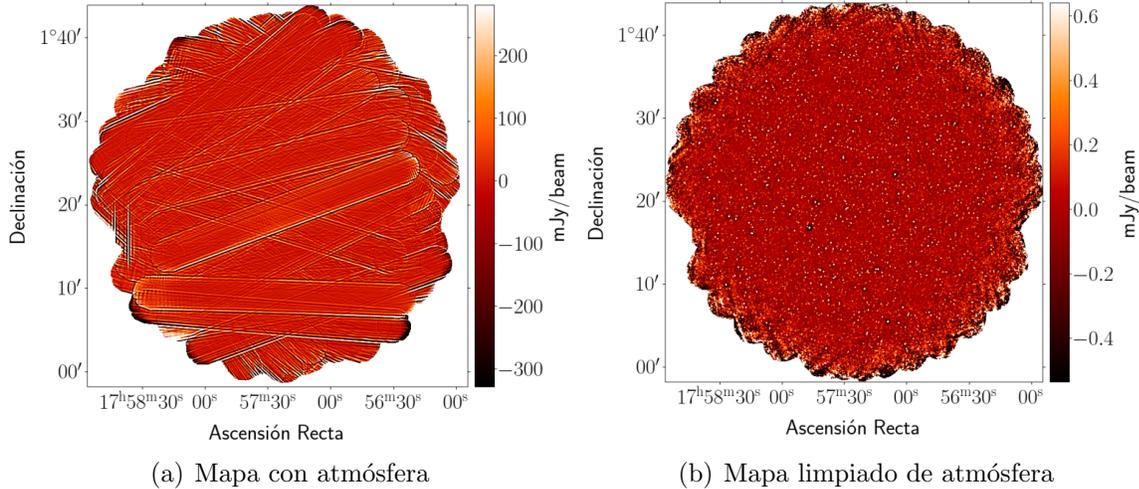


Figura 3.3: Comparación de un mapa a 1.1 mm donde no se elimina la atmósfera (a) vs. uno en donde se elimina (b) utilizando un Análisis de Componentes Principales (PCA), proporcionado por CITLALI. Este ejemplo muestra claramente la inmensa contribución al flujo que da la atmósfera en las observaciones con longitudes de onda milimétricas.

y su amplitud para general el *kernel* de fuente puntual (en adelante sólo *kernel*) que se utiliza posteriormente dentro de CITLALI para el análisis del filtro óptimo del mapa y corregir por atenuación.

Así, por cada observación simulada se crea un *kernel*: **1.** se elimina la atmósfera de cada una de las observaciones guardando el conjunto de valores y vectores propios, **2.** se crea una serie temporal sintética usando señales de fuentes puntuales para hacer una observación simulada de una fuente puntual con un flujo arbitrario fijo y centrada en un campo vacío y sin ruido, **3.** ahora los vectores propios identificados en el paso 1. se proyectan a partir de los datos sintéticos y **4.** el mapa se hace a partir de este dato sintético limpio. Así, la imagen resultante es el *kernel* de fuente puntual y tiene la misma forma y atenuación de una fuente puntual en el mapa de señal limpio para una observación dada. Como ejemplo la Figura 3.4 muestra un acercamiento al *kernel* obtenido de la reducción del mapa mostrado en la Figura 3.3(b). Este es el resultado de la cosuma de los *kernel* de cada una de esas reducciones.

Más adelante lo que se hace es ajustar este *kernel* a una gaussiana bi-dimensional cuyo valor de FWHM será una de las entradas al algoritmo de detección de fuentes

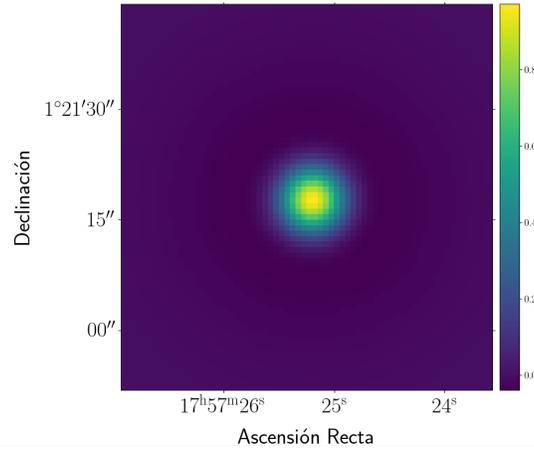


Figura 3.4: *Kernel* obtenido por la reducción de CITLALI.

que se discutirá en la sección 3.3.

3.1.6. Mapas de peso

Para cada una de las observaciones individuales que se simulan se realiza el correspondiente mapa de pesos W sumando en cuadratura los pesos de todas las muestras de los KIDs que contribuyen a un píxel. Por tanto, como el flujo asignado a cada píxel es un promedio pesado de estas muestras, el peso de un píxel será proporcional a $1/\sigma^2$ del flujo estimado. En el archivo de configuración (ver apéndice A.2) se puede elegir entre una estimación de mapa de pesos `approximate` que da un peso constante basado en la sensibilidad y frecuencia de muestreo del detector. Para las simulaciones desarrolladas en esta tesis se elige la opción `full` que toma en cuenta la inversa de la varianza de escaneo del detector.

Es importante notar que, si se tienen varias observaciones, el mapa de pesos (al igual que el del *kernel*) será el resultado de la cosuma de los mapas de peso de cada observación individual.

3.1.7. Co-suma

Como se mencionó antes, es necesario tener en cuenta varias observaciones para co-sumarlas y alcanzar una mayor profundidad en los mapas. Para ello CITLALI ofrece la

posibilidad de sumar distintas observaciones para obtener un mapa final. Esta tarea es controlada dentro de la clase `coadd.h`. En síntesis, primero lee todas las coordenadas de las observaciones y encuentra el máximo y el mínimo global, los que definirán los extremos del mapa cosumado, de modo que todas las observaciones están contenidas.

3.1.8. Elaborador de mapa con el método “Naive”

Como se mencionó previamente, en esta tesis se utiliza la opción para elaborar los mapas con el método “Naive” dado que, a la fecha, es el método que más se ha desarrollado y probado en CITLALI. En principio, se define una malla de píxeles del mapa, en la cual cada píxel es un promedio pesado de las muestras a partir del TOD que cae dentro de ese píxel. Por ejemplo, en la Figura 3.5 se muestra un barrido tipo *raster* que recorre muestra por muestra, detector por detector y encuentra en cuál píxel cae esa muestra particular sin tomar en cuenta la contribución de otros píxeles. La clase `naive_mm` es llamada dentro de la clase `lali.h` (ver capítulo 2).

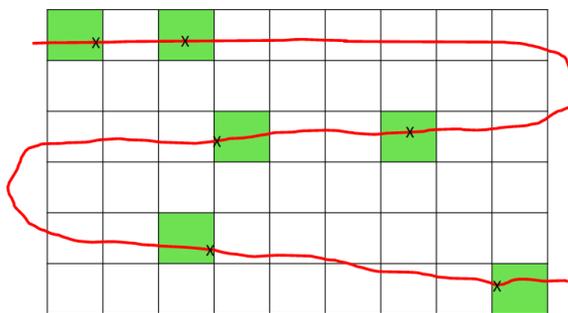


Figura 3.5: Esquema simple que muestra el funcionamiento del elaborador de mapas con el método “Naive”. En rojo se muestra una representación de un mapeo *raster*, en verde el píxel visitado y 'X' las muestras de datos. Obtenido de las presentaciones del *TolTEC Software Telecon CITLALI Talk Series*.

En el archivo de configuración (ver anexo A.2) se puede habilitar esta opción en la sección `mapmaking`. Internamente, la clase busca a través de los detectores (teniendo en cuenta si es una observación de ciencia o de mapa de haz) para definir su índice y posición en el mapa. Calcula el apuntado y busca el vector de píxeles filas/columnas a partir del apuntado para la observación actual y realiza estos mismos pasos para el

mapa co-sumado (si se habilita esa opción) y los mapas de ruido (que se discuten en la sección 3.2). A continuación, se llevan a cabo los siguientes puntos:

- hacer un recorrido a través de las muestras del barrido actual.
- obtener la fila de ese píxel.
- verificar si ese píxel está etiquetado y continuar si no lo está.
- verificar para ver si la posición del píxel está dentro de los límites del mapa. Esto es para cuando el centro del mapa y las dimensiones están sobreescritas de forma manual en el archivo de configuración.
- A continuación, se suman los barridos, peso, *kernel* en el píxel del mapa correspondiente.

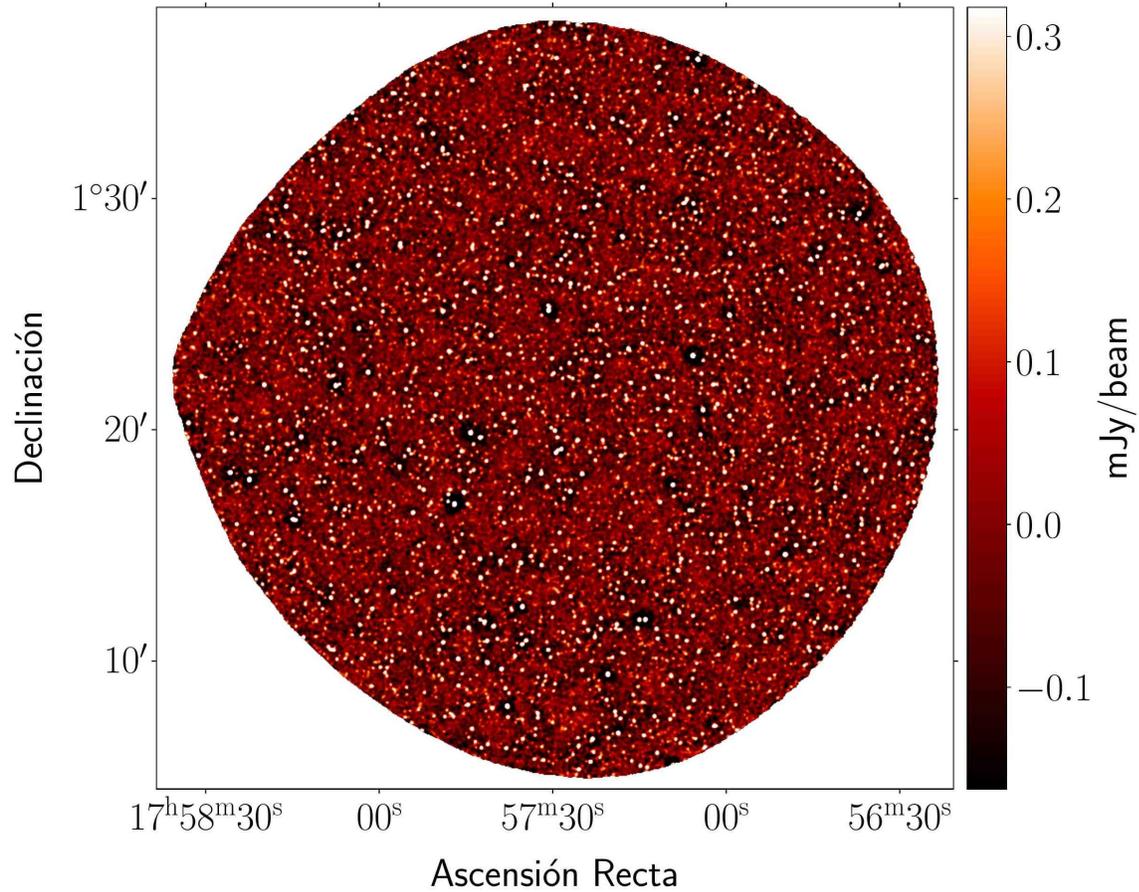
Un tratamiento análogo se sigue para los mapas de ruido.

El siguiente paso en la reducción es aplicar un filtro al mapa resultante. Para ello, en este trabajo se utiliza el filtro Wiener implementado dentro de CITLALI y basado en los desarrollos de Perera et al. (2013). Esto se realiza posterior a la elaboración del mapa. Este filtro busca recuperar la pérdida de flujo debido a la limpieza del PCA y utiliza como entrada el *kernel* como una plantilla. Ajustar este filtro al mapa es equivalente a centrar la PSF en cada píxel del mapa y ajustarla.

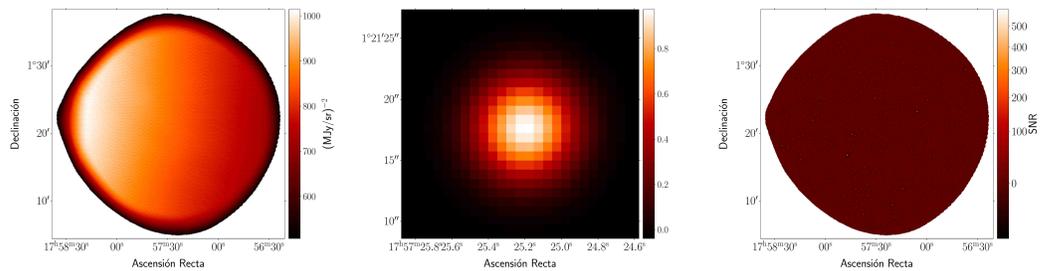
En este trabajo se utiliza la habilitación del filtro Wiener con una plantilla gaussiana. Se hace de este modo dado que la versión del *pipeline* que se utilizó es más estable para este modo de configuración (ver apéndice A.2). Así pues, se crea una plantilla de una gaussiana bi-dimensional con FWHM para las bandas de 1.1, 1.4 y 2.0 mm, de 5, 6.3 y 9.5 segundos de arco, respectivamente.

A modo de ejemplo, los resultados obtenidos de simular y co-sumar 12 observaciones (mostradas en la Figura 3.2) con un método de cartografiado tipo *raster*, incluyendo una atmósfera variable (y un PCA para su eliminación) centrado en las coordenadas ($\alpha = 269.355^\circ$, $\delta = 1.3549^\circ$), al igual que todos los productos de ciencia provistos por la versión v1.2.1-48 de CITLALI y discutidos a lo largo de este capítulo pueden verse (para las bandas 1.1, 1.4 y 2.0 mm) en las Figuras 3.6, 3.7 y 3.8.

Debe tenerse en cuenta que el mapa de señal, al igual que los productos de ciencia, están recortados de modo tal que sólo se representan aquellos píxeles cuyos pesos son mayores al 50 % del peso máximo del mapa, lo cual deja una región aproximadamente



(a) Mapa de señal

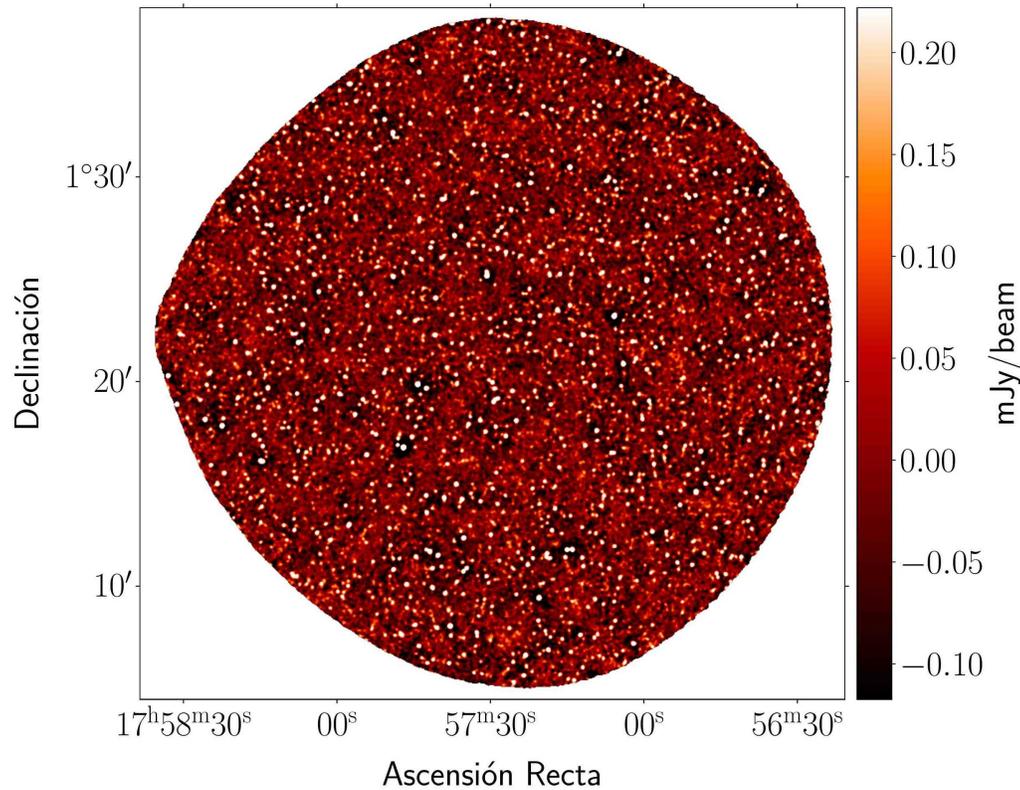


(b) Mapa de peso

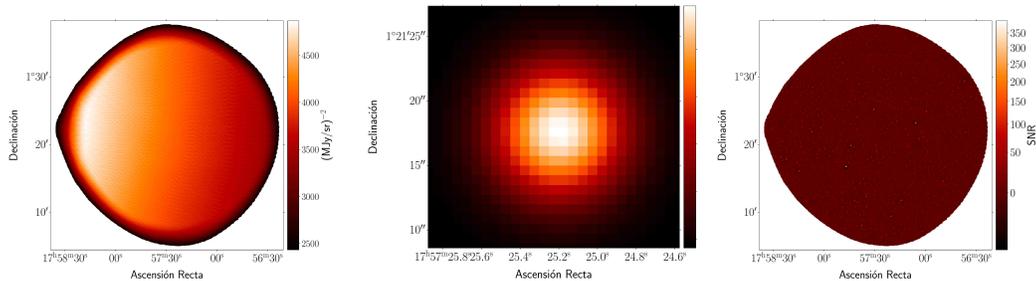
(c) Mapa de *kernel* (Zoom)

(d) Mapa de señal a ruido

Figura 3.6: Ejemplo del resultado de un proceso de simulación de 12 observaciones, con un cartografiado tipo *raster* y considerando un modelo de atmósfera variable. Estos son 4 de los productos de ciencia, para la banda de 1.1 mm, luego del proceso de reducción, se muestran los mapas de señal (a), peso (b), *kernel* (c) y mapa de señal a ruido (d). Todos los resultados fueron simulados y reducidos con el paquete de TolTECA.



(a) Mapa de señal

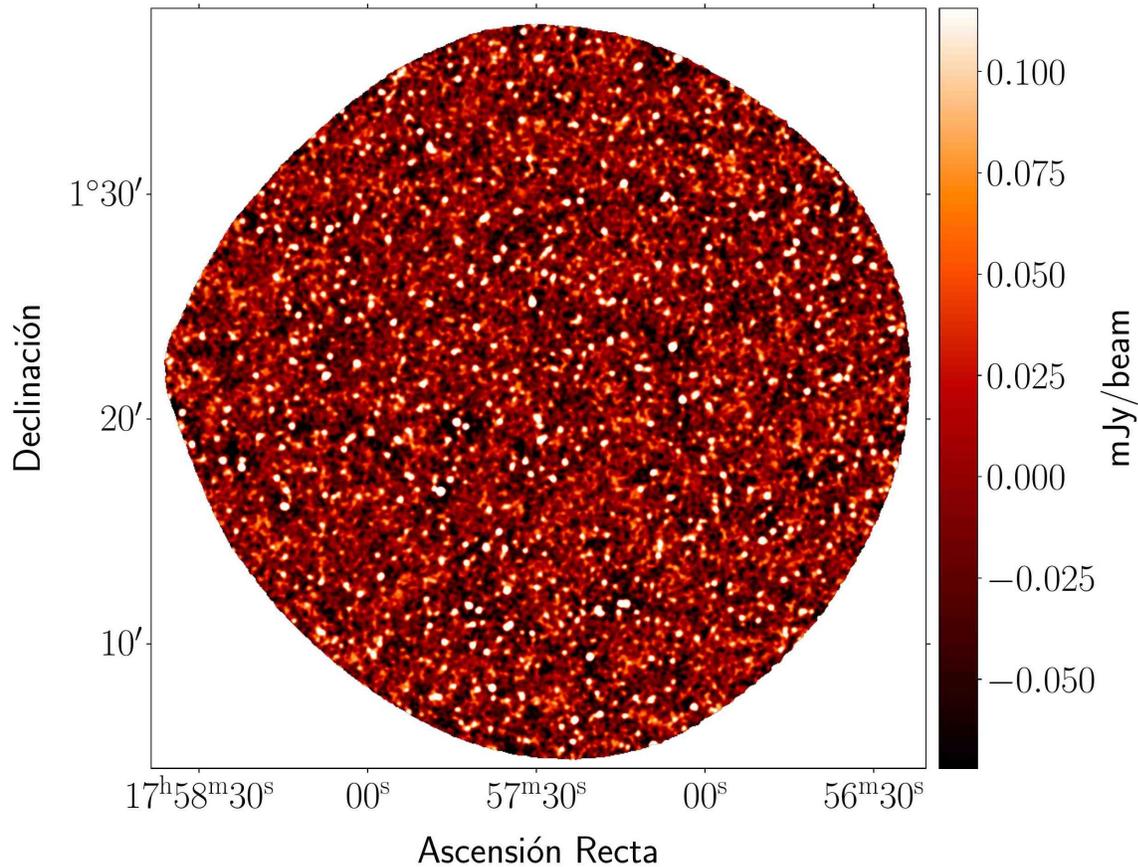


(b) Mapa de peso

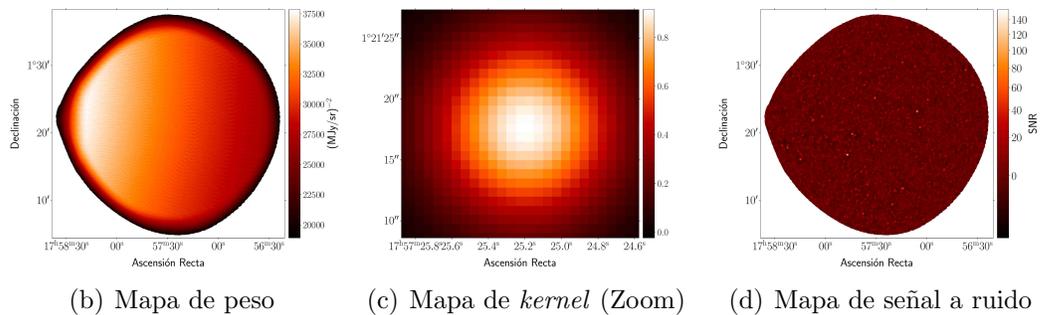
(c) Mapa de *kernel* (Zoom)

(d) Mapa de señal a ruido

Figura 3.7: Ejemplo del resultado de un proceso de simulación de 12 observaciones, con un cartografiado tipo *raster* y considerando un modelo de atmósfera variable. Estos son 4 de los productos de ciencia, para la banda de 1.4 mm, luego del proceso de reducción, se muestran los mapas de señal (a), peso (b), *kernel* (c) y mapa de señal a ruido (d). Todos los resultados fueron simulados y reducidos con el paquete de TolTECA.



(a) Mapa de señal



(b) Mapa de peso

(c) Mapa de *kernel* (Zoom)

(d) Mapa de señal a ruido

Figura 3.8: Ejemplo del resultado de un proceso de simulación de 12 observaciones, con un cartografiado tipo *raster* y considerando un modelo de atmósfera variable. Estos son 4 de los productos de ciencia, para la banda de 2.0 mm, luego del proceso de reducción, se muestran los mapas de señal (a), peso (b), *kernel* (c) y mapa de señal a ruido (d). Todos los resultados fueron simulados y reducidos con el paquete de TolTECA.

circular con un área de ~ 0.22 grados cuadrados. Para poder procesar posteriormente las imágenes, se decide llenar con valores `nan` aquellos píxeles que no cumplan dicha condición. Esto con el propósito de no alterar la estadística para el análisis que permite identificar fuentes. como se discute en la sección 3.3.

3.2. Estimación del ruido presente en los campos

Uno de los pasos requeridos para aplicar el filtro Wiener es el uso de los mapas de ruido. Para estimar el ruido presente en una observación, CITLALI realiza esta tarea mediante la realización de ruido con el método *Jackknife*, multiplicando cada barrido en la serie temporal ya limpia (sin atmósfera, ni picos) por un número aleatorio⁵ ± 1 .

Este proceso se realiza para cada simulación de barrido antes de la elaboración del mapa con el método elegido, en este caso el Naive. Es de notar que, particularmente en CITLALI, no se guardan mapas de ruido por cada observación debido al gran consumo de memoria que esto representa. En su lugar, lo que se hace es cosumar cada mapa de ruido generado por escaneo directamente en el mapa cosumado de ruido correspondiente. La Figura 3.9 muestra un esquema simplificado de esta tarea.

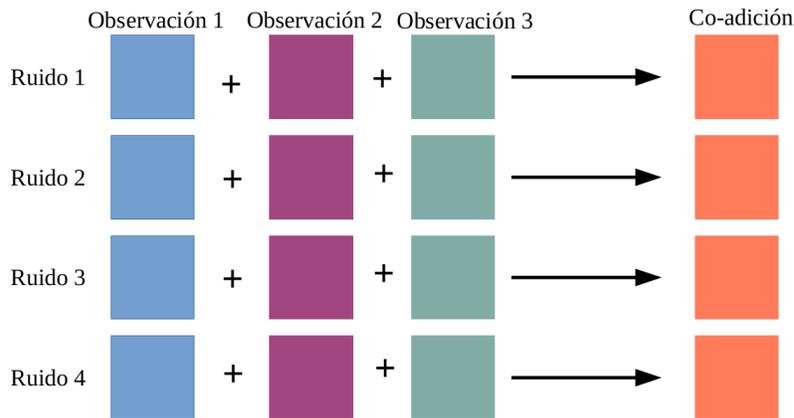


Figura 3.9: Representación esquemática de las realizaciones de ruido en CITLALI. Por cada observación se genera un número N de mapas de ruido que se cosuman. Esto da como resultado un archivo FITS que contiene N realizaciones de ruido cosumadas.

⁵Para ello se utiliza la biblioteca *Boost* de C++, ver: <https://www.boost.org/>.

En esencia, dentro de CITLALI hay un ciclo que busca a través del número de mapas de ruido que ya ha definido el usuario (en este trabajo se utilizan 100 realizaciones de ruido *Jackknife*), si ese píxel está en el mapa y multiplica ese escaneo por un número aleatorio y lo adiciona al mapa.

3.2.1. Caracterización del ruido en los mapas

Teniendo las 100 realizaciones de ruido *Jackknife*, almacenadas dentro de un FITS proporcionado a la salida del proceso de reducción, el siguiente paso es caracterizar ese ruido. Para ello, se realizan histogramas por cada uno de los 100 mapas de ruido. Con esos 100 histogramas se realiza un histograma promedio al que se le ajusta una gaussiana para hallar el valor de desviación estándar. A modo de ejemplo, la Figura 3.10 presenta una muestra de 3 mapas de ruido de un total de 100 utilizados para la simulación (y en general, en las simulaciones subsecuentes, se utiliza el mismo número de mapas de ruido para las estimación del ruido).

En la Figura 3.10 puede verse un exceso de píxeles con flujos positivos en el mapa de señal en contraste a los mapas de ruido, lo cual se debe a la presencia de fuentes tanto brillantes como confusas. Además, existen píxeles con valores de flujo negativo. Esto es debido a que el mapa tiene la característica de tener una media centrada en cero. Las fuentes en el mapa son una versión escalada del *kernel* que, a su vez, presenta algunos lóbulos laterales negativos adyacentes al pico central, lo que contribuye al exceso de señal negativa. Por tanto, la forma de la gráfica es explicada por las fuentes sobre el ruido gaussiano que hace cambiar la distribución de los valores de flujo haciéndola no gaussiana, ampliándola e incluso cambiando el pico a valores negativos.

La Figura 3.11 muestra el histograma de valores de densidad de flujo promedio en el mapa de señal, adicionalmente se representa el histograma promedio a partir de los 100 mapas de ruido, el cual se ajusta una gaussiana centrada en 0 mJy/beam con desviación estándar $\sigma = (24.31 \pm 0.01) \mu\text{Jy}/\text{beam}$. De esta forma se caracterizará el ruido en los resultados que se mostraran en el Capítulo 4.

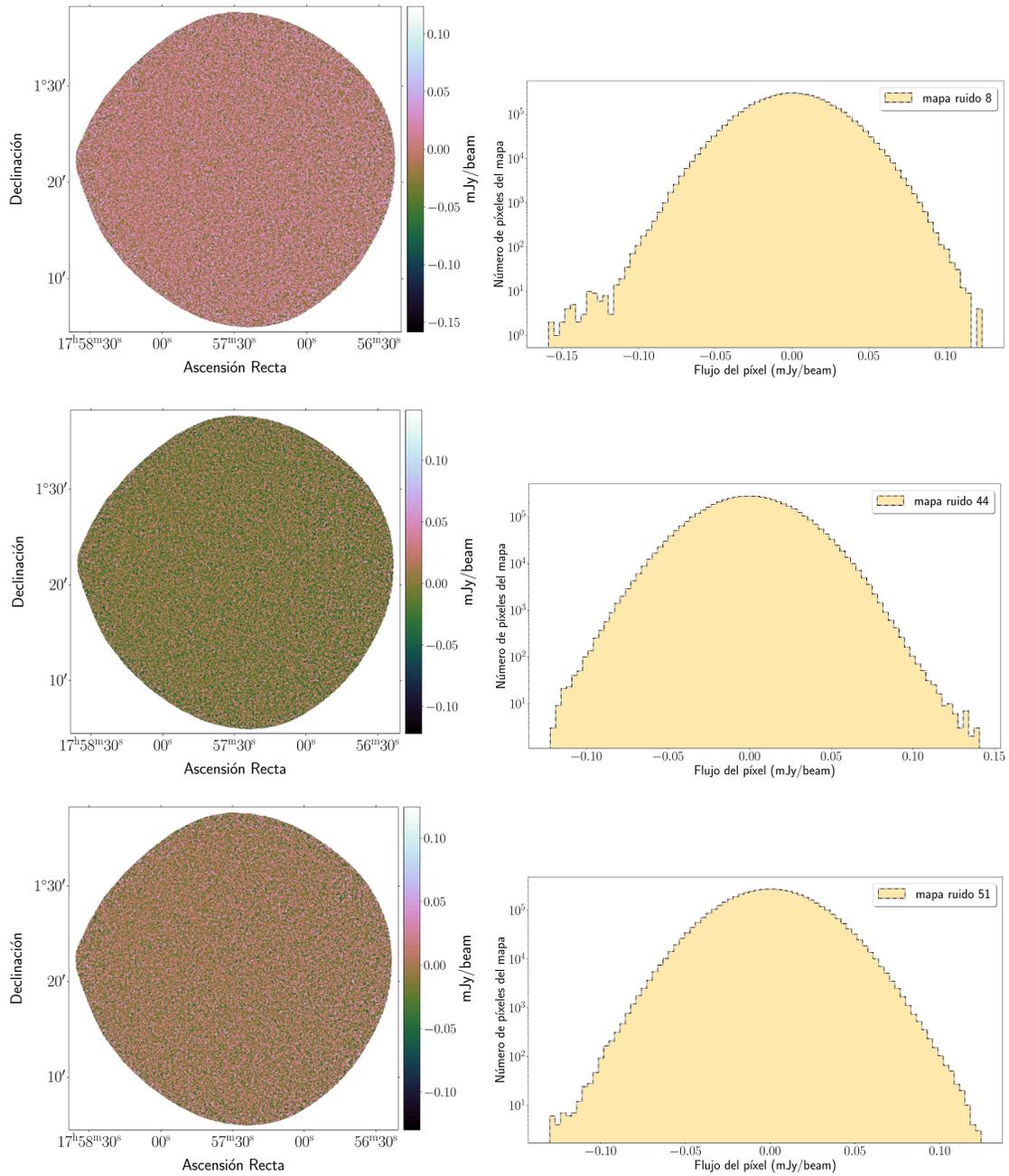


Figura 3.10: Ejemplo de 3 mapas de un total de 100 realizaciones de ruido *Jackknife* para el caso de la simulación en la banda de 1.1 mm.

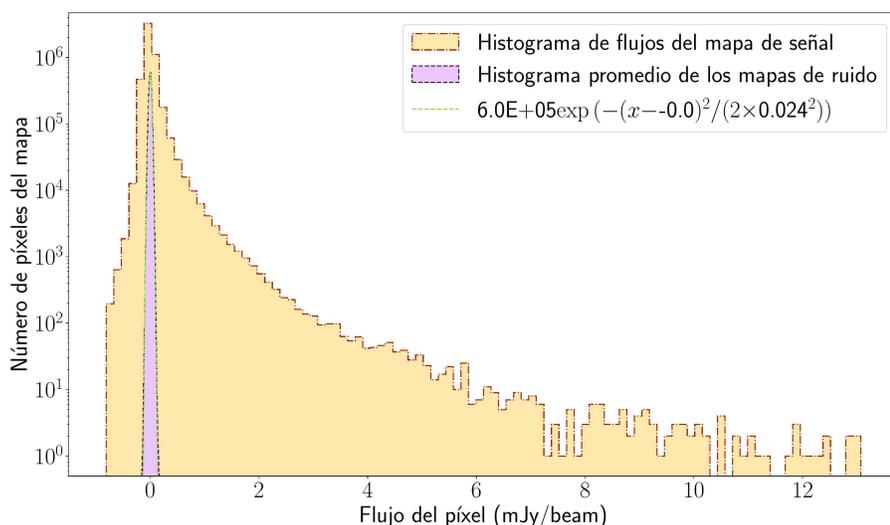


Figura 3.11: Histogramas que muestra en naranja la distribución de píxeles para el mapa de señal y en violeta el histograma promedio de los 100 mapas de ruido, en verde se gráfica el ajuste gaussiano con desviación estándar $\sigma = (24.31 \pm 0.01) \mu\text{Jy}/\text{beam}$.

3.3. Extracción del catálogo de fuentes a partir de los datos reducidos

En esta instancia, ya con las observaciones simuladas y las respectivas reducciones, el siguiente paso es extraer las fuentes que se podrían detectar con TolTEC, para posteriormente estudiar sus propiedades. Caracterizar las ventajas y desventajas de los programas con los cuales se realizará esta tarea es de especial relevancia para cuando se disponga de datos reales. A continuación se explica cómo fue este proceso.

3.3.1. Catálogo de fuentes obtenido con PyBDSF

Con el fin de extraer el catálogo de fuentes de los campos reducidos a partir de las simulaciones, se hace uso del software *Python Blob Detector and Source Finder* (en adelante, PyBDSF), el cual se encuentra disponible en GitHub⁶ (Mohan & Rafferty,

⁶Ver: <https://github.com/lofar-astron/PyBDSF>

2015). Este programa fue diseñado inicialmente para descomponer imágenes de radio-interferometría en fuentes y extraer sus propiedades para usos posteriores⁷. En sí mismo, un software de extracción de fuentes se define como el proceso de ajuste de múltiples gaussianas a los píxeles de las fuentes. Esto tiene sentido dado que una imagen resultado de una observación es la convolución de una gaussiana (ajustada al lóbulo principal de un haz). Este proceso es adecuado dado que la mayoría de las imágenes resultantes de estas observaciones consisten principalmente en fuentes que parecen ser puntuales ó ligeramente extendidas. En este apartado se mostrarán algunas de las principales características del PYBDSF, que fueron utilizadas en esta tesis.

El PYBDSF tiene múltiples capacidades. Puede ser ejecutado sobre imágenes FITS o CASA, incluidos cubos 3-D y 4-D. Con este software es posible leer la imagen reducida con CITLALI en cualquiera de las 3 bandas. Para cada una se realiza inicialmente un cálculo de las estadísticas de la imagen y si los parámetros de entrada se dejan “por defecto”, los mismos se calculan con algoritmos internos. También, se estiman las imágenes promedio y *rms* de fondo; en este punto, si la variación en esas imágenes no es estadísticamente significativa, se toma un valor constante. Se establece un umbral constante para separar el ruido de las fuentes. Este valor puede ser uno fijo o calculado a partir de un algoritmo de Tasa de detección falsa (FDR, *False Detection Rate*).

Posteriormente, PYBDSF identifica las unidades básicas con las cuales opera: estas son llamadas *islas*, que son zonas de emisión continua. Para cada isla se ajustan múltiples gaussianas⁸. En la terminología de búsqueda de fuentes, una isla (o una “mancha”) denota un grupo de 4 píxeles conectados con un brillo por encima del umbral “fusionado” o “agregado”, usualmente entre un rango de $2.5-3\sigma$ con respecto a la imagen de fondo y alrededor de un píxel “semilla”, cuyo brillo está por encima de un umbral de detección, que usualmente es de 5σ (Riggi et al., 2023). El tamaño de la gaussiana se puede fijar al tamaño del haz de restauración, las gaussianas pertenecientes a una isla se agrupan en fuentes discretas. Adicionalmente, se calculan las

⁷PYBDSF se desarrolló para servir a las necesidades del LOFAR (**LO**w **F**requency **AR**ray) operando, principalmente, en los Países Bajos.

⁸Cada isla se puede descomponer en *shapelets*, sin embargo, en este trabajo no se utilizó esa función.

imágenes residuales.

3.3.2. Procesamiento con PyBDSF de los archivos FITS obtenidos con Citlali

A grandes rasgos, la forma en cómo PyBDSF extrae las fuentes puede resumirse en el diagrama mostrado en la Figura 3.12. El programa identifica dentro de la imagen un conjunto de islas de emisión, las cuales se descomponen en Gaussianas, las que posteriormente serán agrupadas, o no, en una fuente siguiendo algunos criterios que serán descritos más adelante.

PyBDSF ofrece distintas tareas, muchas de las cuales fueron usadas durante la realización de esta tesis. Así por ejemplo, para procesar la imagen FITS se ocupa la tarea `process_image`, para mostrar los ajustes `show_fit`, para escribir los catálogos `write_catalog` y para exportar las respectivas imágenes `export_image`. Las implementaciones de estas tareas fueron realizadas dentro de un código en python3.

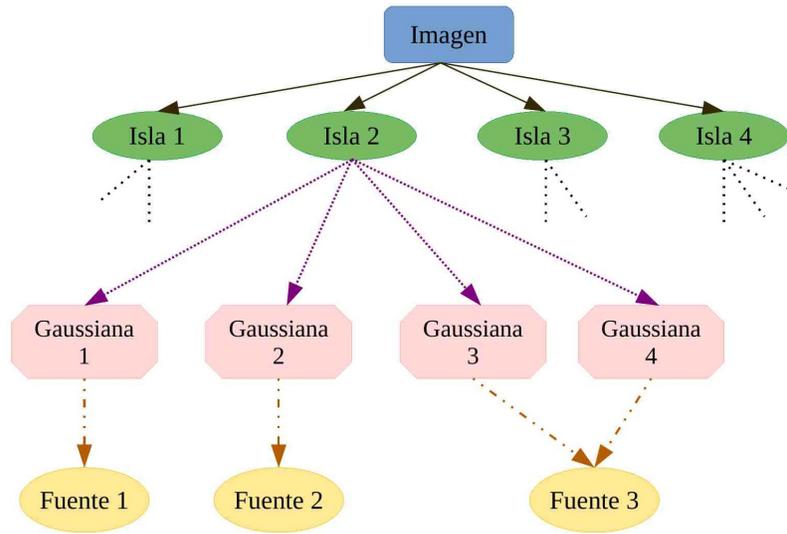
3.3.2.1. Archivos y parámetros de entrada

Se necesita preparar las imágenes FITS para procesarlas con PyBDSF. Esto implica eliminar los píxeles ruidosos que podrían afectar la estadística de la imagen y la extracción del catálogo de fuentes. Como se indica en la sección 3.1, se recortan los mapas de manera que sólo se incluyen los píxeles cuyos valores en el mapa de pesos son mayores al 50% del máximo dentro del mismo⁹. Dado que CITLALI entrega en un mismo FITS todos los productos de ciencia y PyBDSF no está diseñado para separarlos, pues se hace necesario realizar esto último a parte.

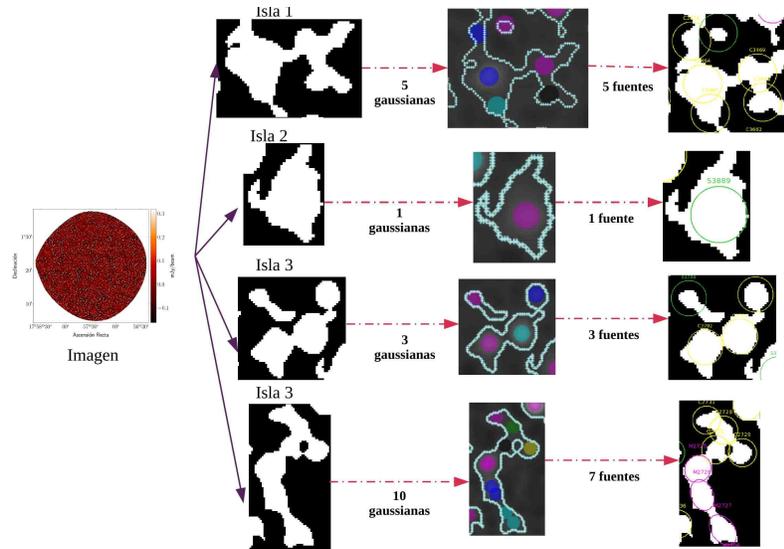
Una vez realizada la separación y recorte de los productos de ciencia, el procesamiento en PyBDSF se inicia con la tarea `process_image`, en este apartado se mencionarán las funciones que fueron utilizadas en este trabajo.

- *Lectura de la imagen:* En este caso, se ingresa el mapa de señal contenido en el FITS reducido con CITLALI para detectar las fuentes. A los píxeles sin información relevante se les asigna el valor de NaN para reducir el tiempo de los cálculos. Se utiliza un código en Python3 para realizar estos cambios y se guarda

⁹Este paso se sigue de acuerdo a lo realizado por Scott et al. (2008).



(a) Esquema jerárquico general de PyBDSF.



(b) Ejemplo del orden de procesos

Figura 3.12: En la parte superior se muestra la jerarquía que sigue un objeto astronómico en PYBDSF y en la parte inferior 4 ejemplos de zonas de emisión contigua, y su proceso hasta formar fuentes.

un nuevo FITS con la información previa del encabezado y la actualización de las unidades (en Jy/beam). Para cada banda, se considera la misma representación en el mapa de cobertura para cortar los mapas de pesos, señal a ruido y

los 100 mapas de ruido con la realización *Jackknife*¹⁰.

Por otro lado, se ajusta una gaussiana bidimensional a cada *kernel* para obtener el FWHM en cada banda de TolTEC. El tamaño teórico del haz para cada una de las 3 bandas es de FWHM = 5, 6.3 y 9.5 segundos de arco. Sin embargo, la resolución angular efectiva tiende a ser mayor debido al filtro Wiener aplicado en las reducciones. Por lo tanto, se utiliza el tamaño del haz estimado a partir del PSF contenido en el *kernel* como el tamaño de haz que se ingresa a PYBDSF para ajustar las fuentes.

- *Estadísticas y umbrales de detección:* calcula la estadística básica de la imagen, como por ejemplo el número de haces por fuente mediante una estimación sensible de un tamaño de caja y de paso. Por defecto, PYBDSF establece un umbral de 5σ para los píxeles 3σ para los límites de las islas (con los parámetros `thresh_pix` y `thresh_isl`), pero estos valores pueden ser ajustados por el usuario. Además, PYBDSF ofrece la opción de calcular los umbrales mediante un método de tasa de detecciones falsas (FDR) descrito por Hopkins et al. (2002).
- *Imágenes promedio y rms:* PYBDSF tiene opciones para calcular de forma iterativa una estimación de la imagen *rms* y promedio sobre el mapa de señal. Esto se hace a través de una pequeña celda que recorre todo el mapa y que incluso puede escalarse cerca de zonas brillantes, en donde es más probable encontrar artefactos. Sin embargo, como CITLALI proporciona un mapa de pesos para cada banda, se estima un mapa *rms* a partir de él y se almacena en un archivo FITS (e.g. Figura 3.13) para ser ingresado a PYBDSF. En cuanto al valor promedio del fondo, se establece en cero debido a que las simulaciones con TolTECA garantizan que sea así¹¹.
- *Identificación de las islas de emisión contigua:* El parámetro `thresh_isl` indica el umbral para el límite de la isla en números de desviación estándar *rms* sobre la media μ . Sí se ingresa un valor alto, se tendrán islas de tamaños más

¹⁰Estos no se utilizan en la instancia de PYBDSF.

¹¹Se exploraron las distintas opciones en PyBDSF, pero los mejores resultados se encontraron utilizando el mapa de pesos para estimar el mapa *rms* y fijando como cero el promedio, en lugar de sólo dejar que el software calculase por sí mismo ambos mapas.

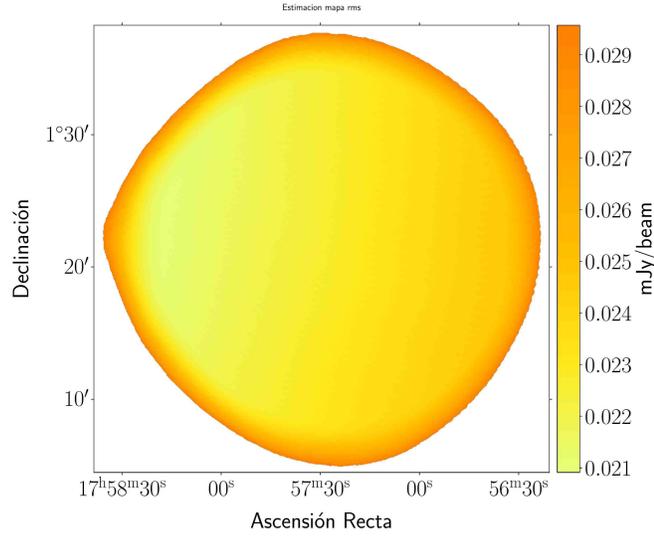


Figura 3.13: Ejemplo de un mapa *rms* en la banda de 1.1 mm, que se usa en PyBDSF. Estimación hecha a través del mapa de pesos provisto por CITLALI, esto es $\sigma = 1/\sqrt{w}$, siendo w el valor del peso en cada píxel.

pequeños y en consecuencia esas regiones serán más pequeñas para realizar el ajuste. Por otro lado, el parámetro $\text{thresh}_{\text{isl}}$ fija el umbral para detectar fuentes. En general $\text{thresh}_{\text{isl}}$ debe ser más pequeño que $\text{thresh}_{\text{pix}}$. Sin embargo, es importante notar que PyBDSF sólo utiliza regiones sobre el valor absoluto, esto es $\text{abs}_{\text{isl}}^{\text{thr}} = \mu + \text{thresh}_{\text{isl}} * \text{rms}$.

Por su parte, el parámetro $\text{thresh}_{\text{pix}}$, análogo al anterior considera islas con picos por encima del umbral absoluto dado como $\text{abs}_{\text{pix}}^{\text{thr}} = \mu + \text{thresh}_{\text{pix}} * \text{rms}$.

Así pues, PyBDSF primero identifica todos los píxeles cuyos valores sean mayores al píxel umbral, luego a partir de cada uno de esos píxeles, todos los 8 píxeles circundantes mayores al umbral límite dado por $\text{thresh}_{\text{isl}}$ se identifican como pertenecientes a una isla. Otro valor que se exploró con PyBDSF fue $\text{maxpix}_{\text{isl}}$ y min_{isl} , que indican el número máximo y mínimo de píxeles con emisión por isla. De otro modo, estos parámetros indican el valor máximo y mínimo de píxeles que debe tener una isla para que sea incluida en los cálculos posteriores. Un ejemplo de las islas ajustadas puede verse en la Figura 3.14.

- *Ajuste de gaussianas a cada isla:* A continuación, los picos identificados en

cada una de las islas serán, por tanto, las suposiciones iniciales para ajustar simultáneamente las gaussianas mediante el algoritmo de Levenberg-Marquardt¹². Así pues, para cada una de las islas que se identifican, se ajustan múltiples gaussianas. En este trabajo se elige la opción de ajustarlas de acuerdo al número de picos que se identificaron en la isla¹³.

Si una de las gaussianas que se ajustan resulta ser una mala solución, se reduce en uno el número de gaussianas y se realiza de nuevo el ajuste hasta que todas las soluciones sean buenas. Después del ajuste en cada isla, se calcula el tamaño deconvolucionado asumiendo el haz teórico. En esta tesis, las fuentes se ajustan al tamaño del FWHM dado por el *kernel* de cada banda, y se realizaron múltiples pruebas para determinar la mejor opción. Además, la identificación de las islas se realiza directamente en el mapa de señal a ruido, pero las mediciones se hacen en el mapa de señal. Los errores en los parámetros ajustados se calculan utilizando una formulación desarrollada por Condon (1997). Esta formulación utiliza ecuaciones de propagación de errores para ajustes gaussianos elípticos bidimensionales en presencia de ruido gaussiano. Un ejemplo de las gaussianas ajustadas a una isla por PYBDSF puede verse en los círculos de colores presentados en la Figura 3.14.

- *Agrupar gaussianas en fuentes:* PYBDSF agrupa las gaussianas ajustadas en cada isla para formar una misma fuente utilizando dos criterios: **1.** ningún píxel en la línea que une los centros de cualquier par de gaussianas tiene un valor inferior al umbral de la isla¹⁴ y **2.** la distancia entre dos centros es menor que la mitad de la suma de sus dos FWHMs a lo largo de la línea que los conecta. Se suman los flujos de las gaussianas identificadas como pertenecientes a una misma fuente para obtener el flujo total. La incertidumbre en el flujo se estima mediante la suma en cuadratura de las incertidumbres en los flujos individuales.

¹²Método utilizado por PYBDSF para resolver el problema de mínimos cuadrados no lineales.

¹³PYBDSF realiza una estimación inicial para los parámetros de estas Gaussianas antes de que se haga el ajuste final. Este método es recomendable si las fuentes no son muy grandes. Otro método en PYBDSF es mediante un mapa de curvatura descrito por Hancock et al. (2012).

¹⁴Que es lo mismo que decir que, la diferencia entre el valor mínimo a lo largo de la línea que une los centros de cualquier par de gaussianas y el valor pico de la gaussiana más pequeña es menor que el producto del umbral de la isla y el valor *rms* de la isla.

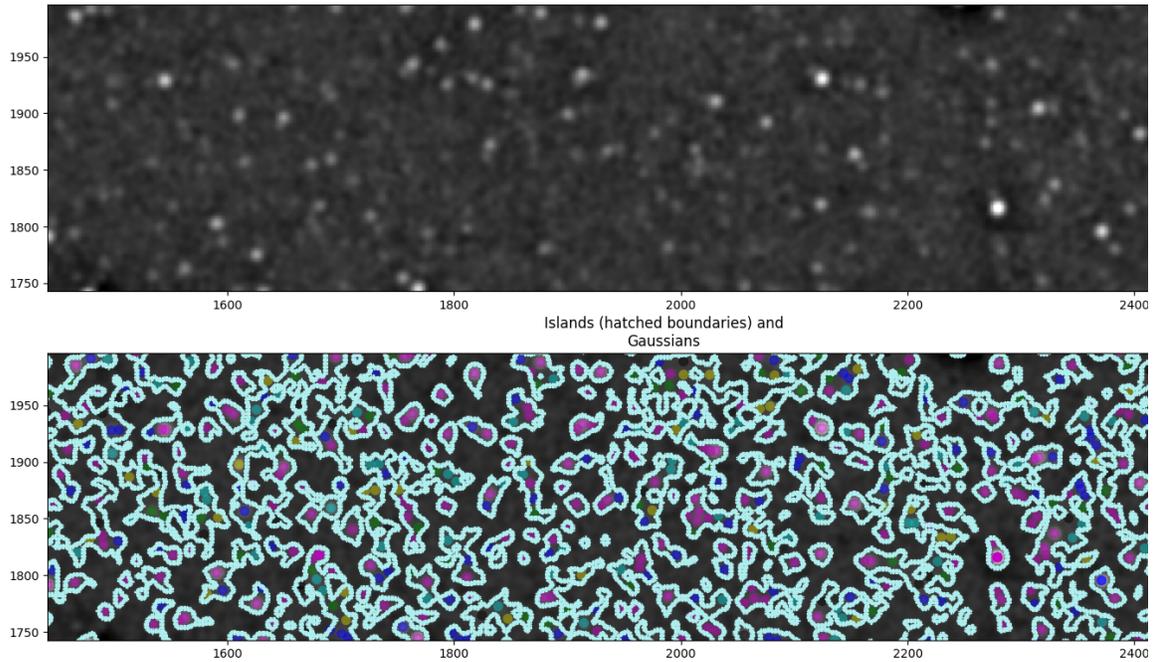


Figura 3.14: (Arriba) imagen sobre la que se va a hacer la identificación de fuentes usando PYBDSF. (Abajo) en la imagen se observan las islas de emisión contigua identificadas por PYBDSF en azul claro. En el interior de las mismas se representan en las fuentes detectadas con círculos de colores.

La posición de la fuente (α, δ) se establece utilizando el centroide de la fuente, que se determina mediante un análisis de momentos, junto con la posición del máximo y el tamaño de la fuente.

En este trabajo PYBDSF se implementa como una librería de *python3*.

3.3.2.2. Información proporcionada en el catálogo extraído

Luego de realizar la ejecución del programa, con las consideraciones mencionadas previamente, se obtiene el catálogo de fuentes, que para esta tesis se elige como una lista de fuentes¹⁵ dado que el objetivo es estudiarlas para realizar comparaciones con las fuentes del catálogo de entrada. El catálogo ya toma en cuenta la corrección en el ángulo de posición y tamaño de la fuente por efectos de proyección de primer orden.

¹⁵Se puede elegir una lista con todas las gaussianas si, por ejemplo, se esta creando un modelo del cielo para ser usado como modelo en la calibración (Mohan & Rafferty, 2015).

Se aplicó el procedimiento descrito en este capítulo a cada archivo FITS (mapas de ciencia y mapas de realizaciones de ruido *Jackknife*), se codificó y ejecutó el algoritmo para obtener el catálogo de salida de PYBDSF para cada una de las tres bandas de observación de TolTEC. El catálogo de PYBDSF contiene información relacionada con el flujo, la morfología y la posición de las fuentes. En este trabajo consideramos los siguientes parámetros:

- 1.- **Source_id**: Número identificador de la fuente.
- 2.- **RA**: Ascensión recta de la fuente, en grados.
- 3.- **E_RA**: Error en la Ascensión recta de la fuente, en grados.
- 4.- **DEC**: Declinación de la fuente, en grados
- 5.- **E_DEC**: Error en la Declinación de la fuente, en grados.
- 6.- **RA_max**: Ascensión recta del máximo de la fuente.
- 7.- **E_RA_max**: Error a 1σ en la Ascensión recta del máximo de la fuente, en grados.
- 8.- **DEC_max**: Declinación del máximo de la fuente, en grados.
- 9.- **E_DEC_max**: Error a 1σ en la declinación del máximo de la fuente, en grados.
- 10.- **Total_flux**: Densidad de flujo total e integrada de la fuente a la frecuencia de referencia, en Jy.
- 11.- **E_Total_flux**: Error a 1σ en la densidad de flujo total de la fuente, en Jy.
- 12.- **Peak_flux**: Densidad de flujo pico por haz de la fuente, en Jy/beam.
- 13.- **E_Peak_flux**: Error a 1σ en la densidad de flujo pico por haz de la fuente, en Jy/beam.
- 14.- **S_Code**: Codificación que define el tipo de estructura que tiene la fuente ajustada, puede ser:
 - **S**: Una sola fuente gaussiana, siendo la única fuente en la isla.

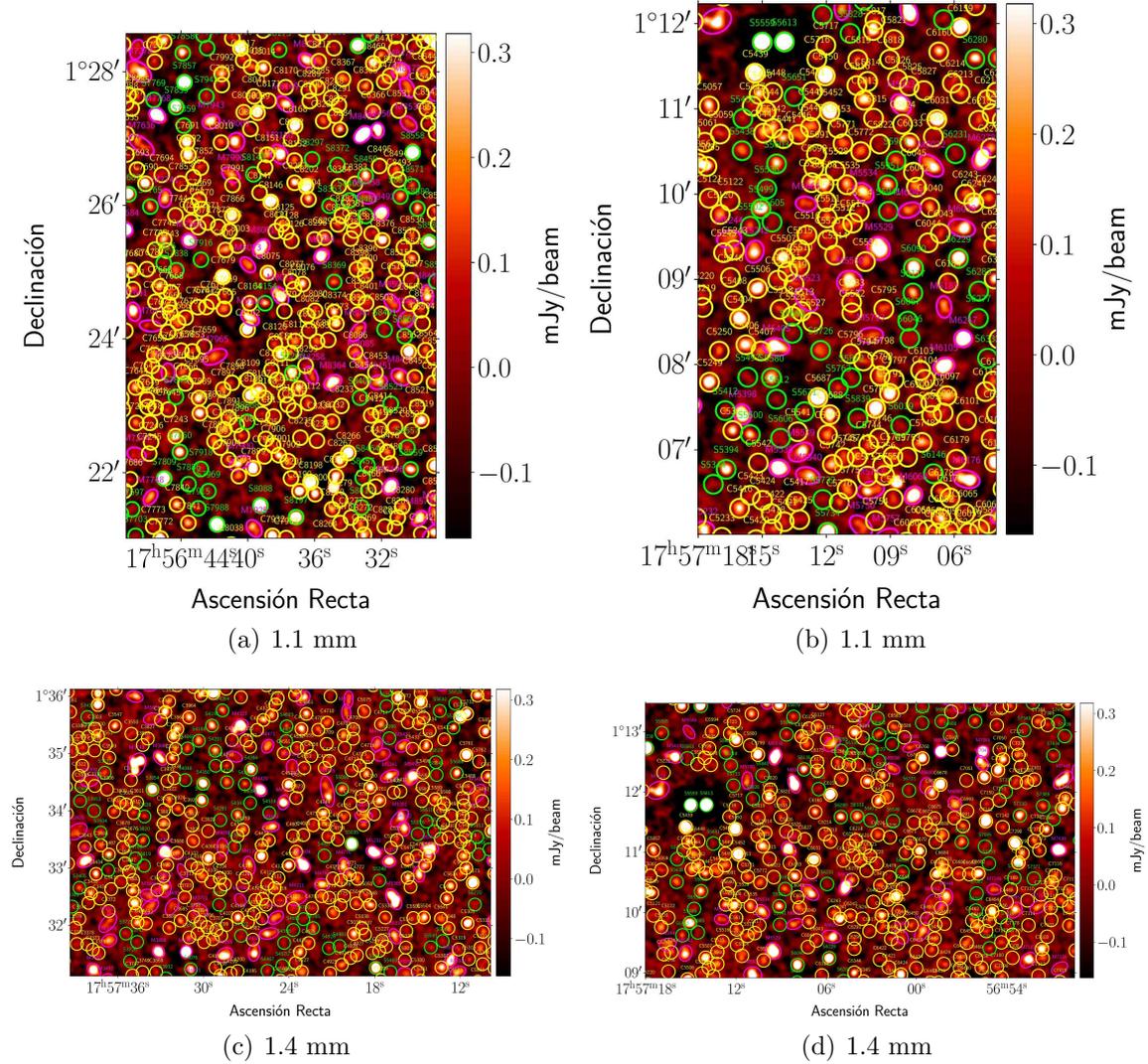


Figura 3.15: Acercamientos en zonas dentro del mapa de señal en la banda de 1.1 mm que muestra las distintas estructuras que tienen las fuentes ajustadas por PYBDSF, en colores verde, amarillo y rosa para las estructuras tipo S, C y M, respectivamente.

- C: Una sola fuente gaussiana en una isla con otras fuentes.
- M: Una fuente multigaussiana.

Las demás columnas de información que ofrece el catálogo pueden consultarse en Mohan & Rafferty (2015). Se presenta en la Figura 3.15 una visualización detallada

de las fuentes detectadas en diferentes zonas del mapa obtenido. El siguiente capítulo se enfocará en el análisis de los resultados obtenidos en el catálogo, incluyendo la comparación de las fuentes de entrada con las detectadas, la evaluación de detecciones falsas, los errores en la posición y análisis de los flujos. Se discutirá también el impacto de estos resultados en futuras observaciones reales.

3.4. Incertidumbre en la posición

Conocer con precisión la posición de las DSFGs es un paso importante para buscar contrapartes en otras longitudes de onda, con el objetivo de poder comprender mejor sus propiedades físicas y su contribución a la formación estelar y su relación con las poblaciones galácticas actuales.

En Ivison et al. (2007) se identificaron contrapartes robustas de SMGs en radio y/o infrarrojo (IR) y, por lo tanto, obtuvieron posiciones precisas para más de dos tercios del Catálogo de Fuentes del Estudio Extragaláctico SCUBA Half-Degree (SHADES, *SCUBA Half Degree Extragalactic Survey*). Esto les permitió realizar una comparación entre las incertidumbres de posición teóricas y las medidas en observaciones reales. En este trabajo se investiga la incertidumbre en la posición de las DSFGs que proviene del ajuste de una gaussiana elíptica a los datos de un mapa, tal como lo entrega PYBDSF. En este ajuste hay 3 parámetros libres: el flujo total A y las coordenadas (α, δ) , por lo que las soluciones propuestas para los errores (Δ) en los tres parámetros es:

$$\Delta A = \sqrt{\frac{1}{\pi}} \frac{h}{\sigma_\theta} \mu' \quad (3.4.1)$$

$$\Delta\alpha = \Delta\delta = \sqrt{\frac{2}{\pi}} \frac{\mu'}{A} h \quad (3.4.2)$$

con h el tamaño del píxel, μ' la dispersión en las mediciones del flujo total y σ_θ el *rms* de un haz con perfil gaussiano. Como $\text{SNR} = A/\Delta A$, usamos (3.4.1) de modo que:

$$A = \text{SNR} \sqrt{\frac{1}{\pi}} \frac{h}{\sigma_\theta} \mu' \quad (3.4.3)$$

reemplazando esto en (3.4.2), queda:

$$\Delta\alpha = \Delta\delta = \frac{\sqrt{2}\sigma_\theta}{\text{SNR}} \quad (3.4.4)$$

como $\sigma_\theta = \text{FWHM}/2\sqrt{2\ln 2}$, así los errores a un 1σ queda como:

$$\Delta\alpha = \Delta\delta = \frac{0.6\text{FWHM}}{\text{SNR}} \quad (3.4.5)$$

Esta expresión solo es válida si se han corregido los efectos por el sesgo de Malquist y de Eddington que dan una sobreestimación en el flujo. De no ser así, la ecuación (3.4.5) debe modificarse como (Peacock, 1998):

$$\Delta\alpha = \Delta\delta = \frac{0.6\text{FWHM}}{\sqrt{\text{SNR}_{\text{app}}^2 - (2\beta + 4)}} \quad (3.4.6)$$

Donde SNR_{app} es la señal a ruido observada aparente y β la pendiente de la curva del número de cuentas integradas ($N(\geq S) \propto S^{-\beta}$), que en este caso se mide directamente de la simulación Nava-Moreno et al. (2019). Ivison et al. (2007) encontró una distribución de distancias radiales representada por:

$$P(r) = r e^{-r^2/2\sigma^2} \quad (3.4.7)$$

siendo r la distancia entre la posición real de la fuente (catálogo de Nava-Moreno et al. (2019)) y la posición medida (extraída en PYBDSF). La ecuación (3.4.7) tiene su máximo en $\Delta\alpha = \Delta\delta = \sigma$. En el capítulo 4.3 se presentan los resultados de este apartado.

Capítulo 4

Resultados

En el capítulo anterior, se detalló la metodología empleada para simular y reducir las observaciones, así como el uso de PyBDSF para extraer el catálogo de fuentes utilizando los productos de ciencia proporcionados por Citlali. En este capítulo, se presentan los resultados de las simulaciones realizadas con el objetivo de evaluar diversos aspectos: la recuperación de la densidad de flujo, el error de posición de las fuentes, el porcentaje de fuentes falsas detectadas mediante cortes en la relación señal a ruido (SNR), el *boosting* y la completez del catálogo extraído por PyBDSF.

A lo largo de la realización de esta tesis, se llevaron a cabo múltiples simulaciones. Sin embargo, en este apartado se presentan los resultados de sólo cuatro de ellas. En síntesis, por cada simulación se tienen 15 productos de ciencia (ver Tabla 3.1), lo que significa que se analizaron 60 productos para obtener los resultados¹. Las características principales de estas simulaciones se detallan en la Tabla 4.1.

En la simulación 1, por ejemplo, se hacen 12 observaciones de $30' \times 30'$ en Alt/Az, lo cual conlleva poco más de 8 horas de tiempo de integración total para el simulador TolTECA. Con ello, el simulador arroja un mapa *rms* (cortado al 50 % del máximo del mapa de pesos) de 0.0209–0.0296 mJy/haz para la banda de 1.1 mm. Sin embargo, estas profundidades parecen demasiado optimistas para la velocidad de cartografiado predicha para TolTEC. Al ingresar el tamaño del mapa ($30' \times 30'$) en la calculadora de tiempos de integración de TolTEC dispuesta en línea², se debería tener un tiempo de integración de poco más de 120 horas tomando en cuenta una atmósfera estática y una velocidad de cartografiado pesimista ($3 \text{ deg}^2/\text{hr}/\text{mJy}^2$ a 1.1 mm). Utilizando

¹Si se incluyen los resultados en crudo se estaría hablando de 120 productos de ciencia.

²Ver: https://toltec.lmtgtm.org/toltec_sensitivity_calculator

Núm. Simulación	Cartografiado (Características)	Núm. Obs.	Tiempo de cada observación	Tiempo de integración	<i>rms</i> 1.1 mm (mJy/haz)	<i>rms</i> 1.4 mm (mJy/haz)	<i>rms</i> 2.0 mm (mJy/haz)
1	<i>raster</i> (vel. 50 arcs/s, atmósfera variable)	12	40 m 55 s	8 h 11 m	(0.0209–0.0296)	(0.0155–0.0219)	(0.0113–0.0160)
2	<i>raster</i> (vel. 50 arcs/s, atmósfera estática)	12	40 m 55 s	8 h 11 m	(0.0387–0.0548)	(0.0267–0.0377)	(0.0178–0.0251)
3	<i>raster</i> (vel. 3 arcmin/s, atmósfera variable)	19	14 m 55 s	4 h 43 m	(0.0446–0.0630)	(0.0329–0.0465)	(0.0241–0.0341)
4	Lissajous ($\omega_x = 5.2$ rad/s, $\omega_y = 4$ rad/s atmósfera variable)	2	10 min	20 min	(0.0443–0.178)	(0.0235–0.105)	(0.0168–0.0752)

Tabla 4.1: Características de las 4 simulaciones con las cuales se realizan los análisis. Se presenta el tipo de cartografiado y sus características de simulación, el número de observaciones, el tiempo de cada una al igual que el tiempo integrado. Las tres últimas columnas presentan los valores máximo y mínimo del mapa *rms* (previamente cortado a aquellos píxeles cuyo valor en el mapa de pesos sea $\geq 50\%$ al valor máximo de dicho mapa) estimado. Todas las reducciones tienen un filtro tipo Wiener. Los tiempos listados en las columnas 4 y 5 no incluyen el tiempo gastado en *overheads*.

condiciones de velocidad de cartografiado optimistas ($12 \text{ deg}^2/\text{hr}/\text{mJy}^2$) se debería estar alcanzando estas profundidades con tiempos de integración de ~ 30 horas. Esta discrepancia fue reportada (comunicación directa con M. McCrackan, desarrollador de Citlali) y se revisará en las próximas versiones de TolTECA y Citlali. Se continuará con el análisis bajo el entendimiento de que estas profundidades corresponden a tiempos de integración superiores a los reportados en la Tabla 4.1.

Dado que en estas simulaciones se hace uso del filtro Wiener, el FWHM efectivo incrementa su valor. De este modo las fuentes se ajustan con una gaussiana con el valor del FWHM del *kernel* (que se obtiene ajustando una gaussiana bi-dimensional sobre la PSF del *kernel*). Además se toma en cuenta el valor final de la PSF para corregir las fuentes extraídas por PyBDSF por el flujo perdido durante el proceso de eliminación atmosférica. En la Tabla 4.2 se muestran los FWHMs correspondientes a cada simulación.

En PyBDSF se ingresa la estimación del mapa *rms* (realizada con el mapa de

Simulación	FWHM		
	1.1 mm (arcs)	1.4 mm (arcs)	2.0 mm (arcs)
1	6.623	8.465	12.820
2	7.091	8.948	13.096
3	6.945	8.733	12.884
4	8.821	9.451	14.011

Tabla 4.2: Anchura a media altura (FWHM) para las tres bandas de TolTEC, resultado del ajuste de una gaussiana bi-dimensional al *kernel* de cada una de las simulaciones (Tabla 4.1).

peso) y se fija como cero el valor promedio del mapa. También se ingresan el mapa de señal y el de relación SNR. A continuación se ejecuta el algoritmo para realizar las mediciones descritas en el Capítulo 3, Sección 3.3. Este proceso entrega un catálogo preliminar y un mapa residual (mapa que resulta de extraer las fuentes identificadas por PYBDSF), entre otros productos. Sin embargo, al analizar el mapa residual se identificó que algunas fuentes no fueron detectadas. Por este motivo, se decide ejecutar nuevamente el algoritmo sobre este mapa residual para así obtener un segundo catálogo. La combinación de ambas ejecuciones da como resultado un catálogo final. En general, PYBDSF realiza detecciones con $\text{SNR} \geq 1$.

En lo subsecuente y por simplicidad, se muestran las etapas de los resultados para la simulación 1 en las tres bandas. Para contrastar se muestran los resultados de la simulación 2 en la banda de 1.1 mm. Las Tablas y gráficos de las otras simulaciones se anexan en las secciones C.2, C.3, y C.4 del apéndice.

4.1. Asignación de fuentes

En este apartado se busca realizar la identificación de contrapartes entre el catálogo de entrada y el de salida. Iniciando con los resultados de la simulación 1, se presentan los mapas de ~ 0.22 grados cuadrados mostrados en las Figuras 3.6, 3.7 y 3.8 para cada una de las 3 bandas. En principio, se debe delimitar la comparación de fuentes de entrada y salida a la huella del mapa. Es decir, las fuentes de entrada que se encuentren en zonas del cielo en donde hay píxeles con valores NaN no se consideran

en el análisis. A continuación, se ordenan las fuentes del mapa de entrada y salida de mayor a menor flujo. Para cada fuente detectada se buscan contrapartes en el catálogo de entrada usando un radio de búsqueda conservativo de un FWHM para las fuentes etiquetadas como tipo S o C (ver sección 3.3.2.2) y de 1.3FWHM^3 para las de tipo M. Dentro de ese radio de búsqueda se selecciona la fuente más brillante en el catálogo de entrada. A estas asignaciones se las denomina como ‘fuentes verdaderas’. Si por el contrario, para una detección, no se encuentra una contraparte en el catálogo de entrada, se les denomina como ‘fuentes falsas’. Su análisis se muestra en la siguiente sección. Los resultados de esta asignación se pueden ver en las Figuras 4.1 y 4.2 en las tres bandas para las densidades de flujo integrado y pico.

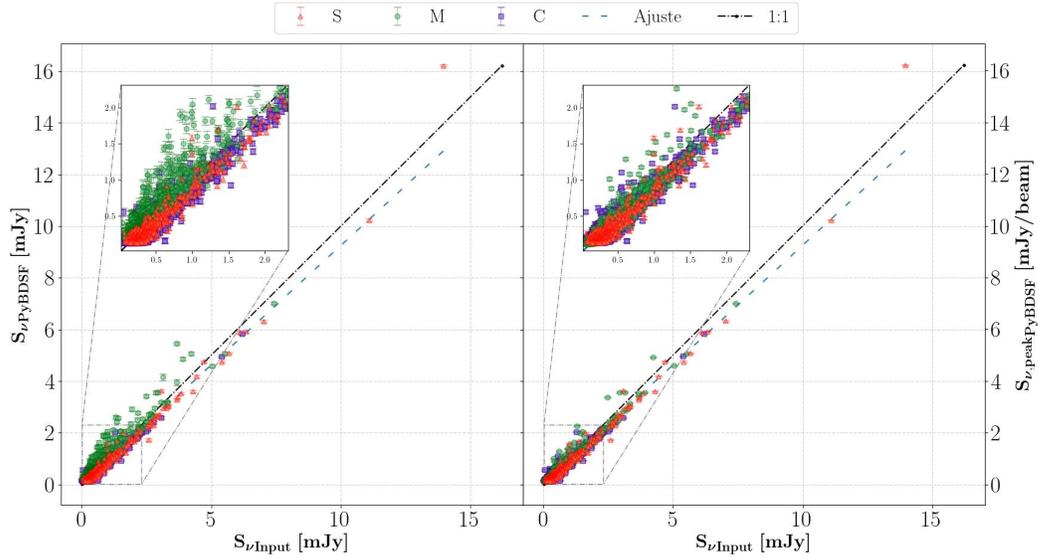


Figura 4.1: Densidades de flujo de salida $S_{\nu,PyBDSF}$ respecto a las de salida $S_{\nu,Input}$ en la banda de **1.1 mm** para cada uno de los tipos de estructuras de fuentes detectados por PyBDSF. Los cuadrados, círculos y triángulos representan las estructuras tipo C, M y S, respectivamente. En la derecha se grafican los flujos integrados totales mientras que la izquierda se muestran los flujos picos. La línea negra indica la relación uno a uno y la línea azul el ajuste a los puntos.

³Este fue el mejor resultado encontrado durante las pruebas realizadas en este trabajo

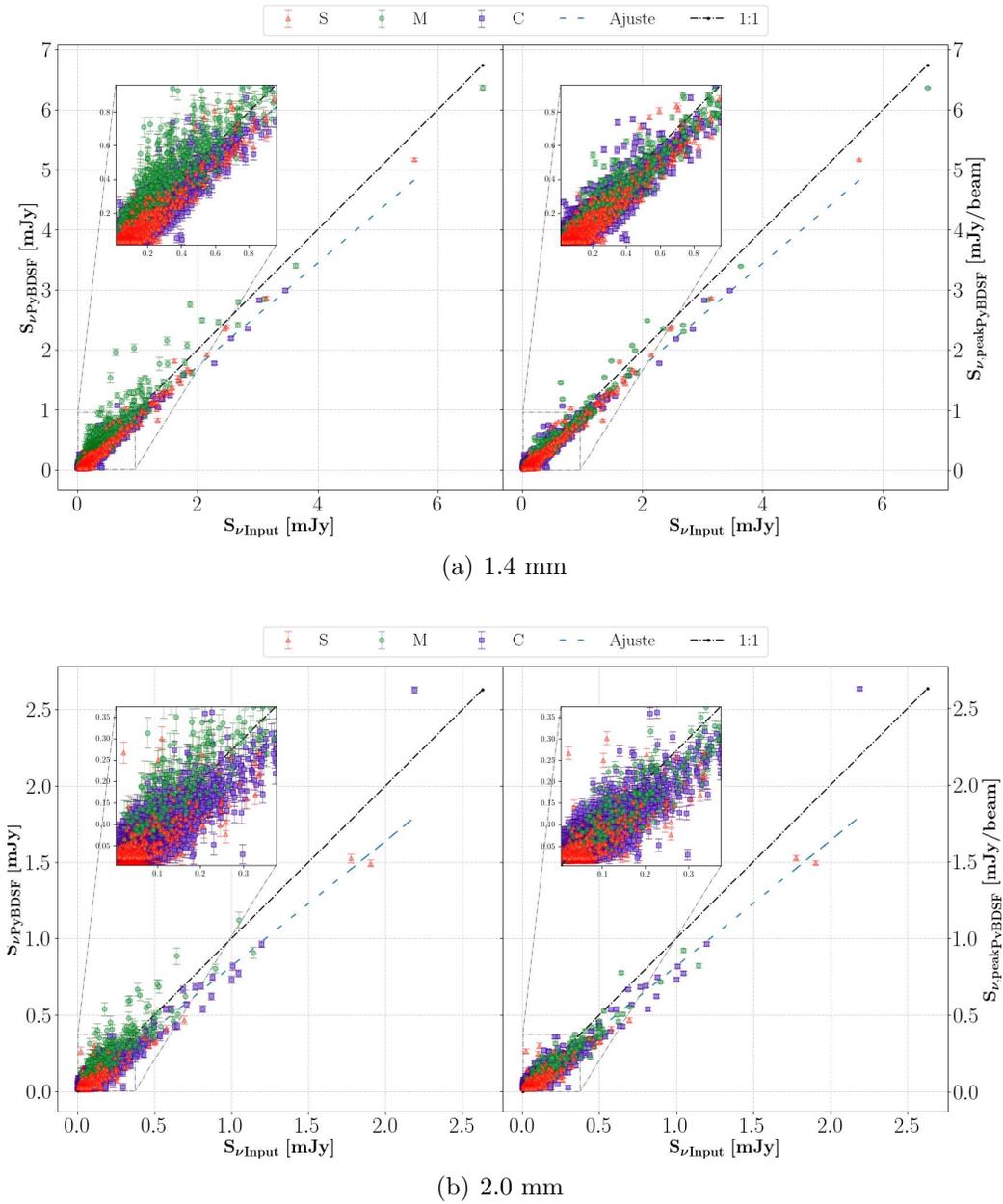


Figura 4.2: Densidades de flujo de salida $S_{\nu,PyBDSF}$ respecto a las de salida $S_{\nu,Input}$ en la banda de **1.4 mm** y **2.0 mm** para cada uno de los tipos de estructuras de fuentes detectados por PyBDSF. Los cuadrados, círculos y triángulos representan las estructuras tipo C, M y S, respectivamente. En la derecha se grafican los flujos integrados totales mientras que la izquierda se muestran los flujos picos. La línea negra indica la relación uno a uno y la línea azul el ajuste a los puntos.

Un resultado que llama la atención es que, a pesar de haberse corregido las fuentes extraídas por la pérdida de flujo en el proceso de eliminación de la atmósfera (PCA), persiste una pérdida de flujo en el mapa de señal. En la sección 4.4 se describe y analiza con más detalle de esta situación. Las fuentes con una estructura tipo ‘M’ tienden a desviarse hacia flujos mayores de la relación esperada de 1:1. Esta discrepancia se debe a que, según la definición de PyBDSF, estas fuentes están compuestas por más de una gaussiana. Por lo tanto, su flujo total es la suma de cada componente, lo que se refleja en un incremento en el flujo.

4.2. Detecciones falsas

La tasa de detecciones falsas se define como el número mayor a $N\sigma$ picos debidos al ruido y que aparecen en zonas del cielo en donde no hay fuentes reales (Scott et al., 2008). Como el ruido puede tener una contribución de señal debido a fuentes débiles no identificadas por estar sobrelapadas unas sobre otras, entonces el mapa está afectado por la presencia de fuentes no detectadas, por lo que utilizar esta definición puede no ser del todo adecuado. Por esta razón, lo que se realiza es la detección de fuentes y luego se realizan cortes en señal a ruido para ver qué tantas de esas fuentes detectadas corresponden al catálogo de Nava-Moreno et al. (2019) y cuántas de estas son detecciones falsas. Las curvas resultantes reflejan el número promedio de puntos que podrían ser detectados de forma errónea.

Los resultados de número de fuentes verdaderas y falsas por cortes en SNR se muestran en la Figura 4.3 y en la Tabla 4.3.

SNR	Fuentes verdaderas \geq SNR	Fuentes falsas \geq SNR (%)
1.1 mm		
1.0	7405	1788 (19.45)
1.5	6801	1142 (14.38)
2.0	6027	573 (8.68)
2.5	5328	264 (4.72)

SNR	Fuentes verdaderas \geq SNR	Fuentes falsas \geq SNR (%)
3.0	4751	105 (2.16)
3.5	4182	31 (0.74)
4.0	3718	12 (0.32)
4.5	3335	3 (0.09)
5.0	3025	1 (0.03)
5.5	2734	1 (0.04)
6.0	2502	0 (0.0)
1.4 mm		
1.0	5442	540 (9.03)
1.5	4856	338 (6.51)
2.0	4194	179 (4.09)
2.5	3685	76 (2.02)
3.0	3227	33 (1.01)
3.5	2805	14 (0.5)
4.0	2463	8 (0.32)
4.5	2172	5 (0.23)
5.0	1930	4 (0.21)
5.5	1732	1 (0.06)
6.0	1576	0 (0.0)
2.0 mm		
1.0	2794	41 (1.45)
1.5	2404	32 (1.31)

SNR	Fuentes verdaderas \geq SNR	Fuentes falsas \geq SNR (%)
2.0	1997	19 (0.94)
2.5	1659	11 (0.66)
3.0	1379	3 (0.22)
3.5	1131	2 (0.18)
4.0	927	1 (0.11)
4.5	797	0 (0.0)
5.0	672	0 (0.0)
5.5	563	0 (0.0)
6.0	493	0 (0.0)

Tabla 4.3: Detecciones verdaderas y falsas por cortes en señal a ruido en las tres bandas. La primera columna muestra los cortes en señal a ruido, la segunda y tercera el número de fuentes verdaderas y falsas por cada corte, y en paréntesis se especifica el porcentaje de detecciones falsas.

Como puede verse, en los resultados de esta simulación, para el corte de $\text{SNR} \geq 3.5$ el número esperado de fuente falsas es 0.7%. Sin embargo, nótese que también hay un porcentaje muchísimo menor de fuentes falsas en cortes superiores. Este resultado se debe a los picos de ruido que sobresalen para ser detectados, así como artefactos en la imagen de señal debido a que en el proceso de eliminación atmosférica aun persisten residuos significativos. Actualmente, es un área activa de trabajo en TolTEC buscar, corregir y mejorar este tipo de resultados en la parte de simulación atmosférica.

Por lo anterior, en la Simulación 2 se considera un modelo de atmósfera estático para probar una situación más favorable en la reducción de ruido. Los detalles se encuentran en el apéndice C.2, para las tres bandas de observación. De momento, se puede mencionar que en esta simulación 2, dado que no se presentan artefactos de la sustracción de la atmósfera, la pérdida de flujo es menor. Adicionalmente, como se observa en la Tabla C.2, a un corte en $\text{SNR} \geq 3$ no se esperan fuentes falsas en

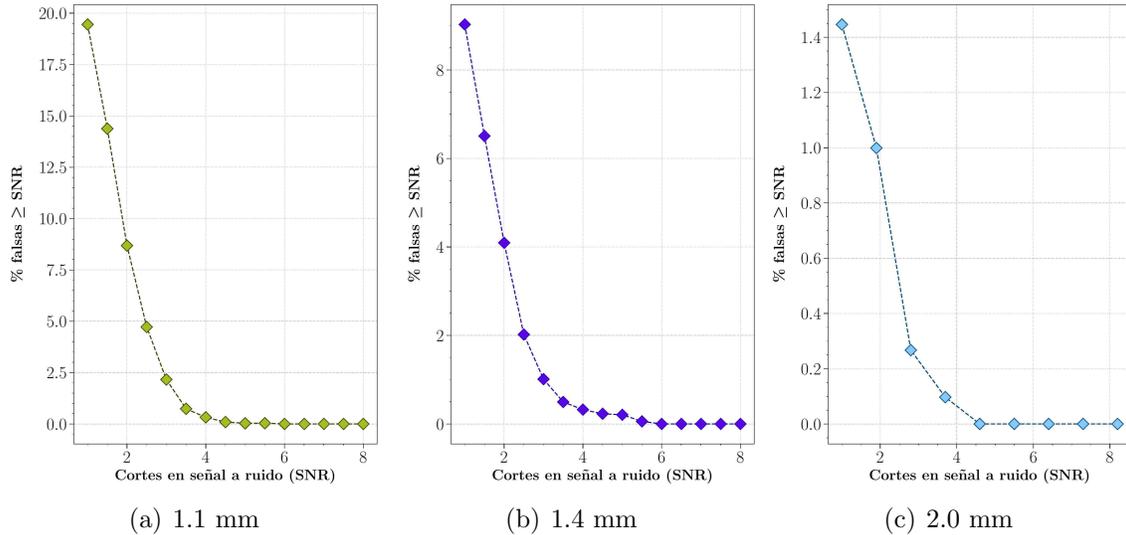


Figura 4.3: Porcentaje de fuentes falsas detectadas sobre umbrales de señal a ruido para la simulación 1, con cartografiado *raster* y atmósfera variable en las tres bandas.

la banda de 1.1 mm y 1.4 mm, y a partir de $\text{SNR} \geq 2$ para la banda de 2.0 mm. Las gráficas de los cortes por SNR, para la simulación 2, se presentan en la Figura 4.4. Este resultado, muestra la importancia que tiene la correcta eliminación de la contribución atmosférica.

Los resultados de las simulaciones 3 y 4 se presentan en las secciones C.3 y C.4. En la simulación 3 puede verse la presencia de artefactos, de modo que, aunque en un porcentaje muy bajo, se presentan fuentes falsas a altas SNR. Se encuentran bajos porcentajes de fuentes falsas ($< 3.3\%$), para una $\text{SNR} \geq 3.5$ en las bandas de 1.1 y 1.4 mm, y para una $\text{SNR} \geq 2$ en la de 2.0 mm (ver Tabla C.5 y Figura C.19). Llama la atención que estas estructuras sólo se presentan en mapas con áreas grandes. Esto se comprobó con varias simulaciones. Como ejemplo puede verse la simulación 4. En este caso, en las bandas de 1.1 y 1.4 mm a un corte de $\text{SNR} \geq 2.5$ no hay presencia de fuentes falsas y a $\text{SNR} \geq 2$ en la de 2.0 mm. El tipo de movimiento que sigue un cartografiado Lissajous asegura una mayor cobertura que uno *raster*, sin embargo no es utilizado en mapas grandes porque no es eficiente.

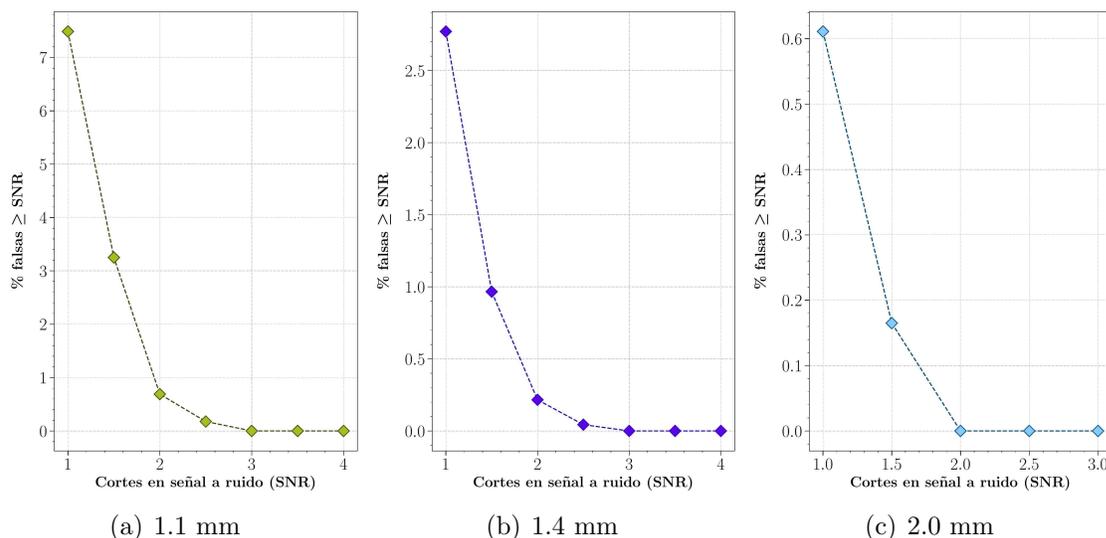


Figura 4.4: Porcentaje de fuentes falsas detectadas sobre umbrales de señal a ruido para la simulación 2, con cartografiado *raster* con atmósfera estática en las tres bandas.

4.3. Errores de posición

Dado que las posiciones de entrada son conocidas, es muy importante caracterizar qué tan precisa es la posición determinada de las fuentes detectadas en el mapa de señal. Esto permitirá realizar correcciones en el simulador TolTECA y ver qué aspectos están fallando.

Para investigar si hay algún sesgo en los errores de posición, se realizan gráficos que permiten ver, de acuerdo a los cortes de señal a ruido, los errores de posición para las fuentes identificadas dentro de radios concéntricos desde el centro del mapa. Esto permite identificar si el proceso de reducción tiene algún problema en la astrometría de la imagen. En la simulación 1 se realizan estas gráficas para fuentes con $\text{SNR} \geq 3.5$ en las tres bandas de observación, como se observa en la Figura 4.5. De estos gráficos puede verse que las fuentes con estructura tipo ‘M’ están en general mucho más desviadas en los errores de posición. Esto se debe a que, como se mencionó, son estructuras que PyBDSF no logra separar, por lo que se asignó un radio de búsqueda mayor, y así mismo el centroide estimado por el programa se desplaza debido a la

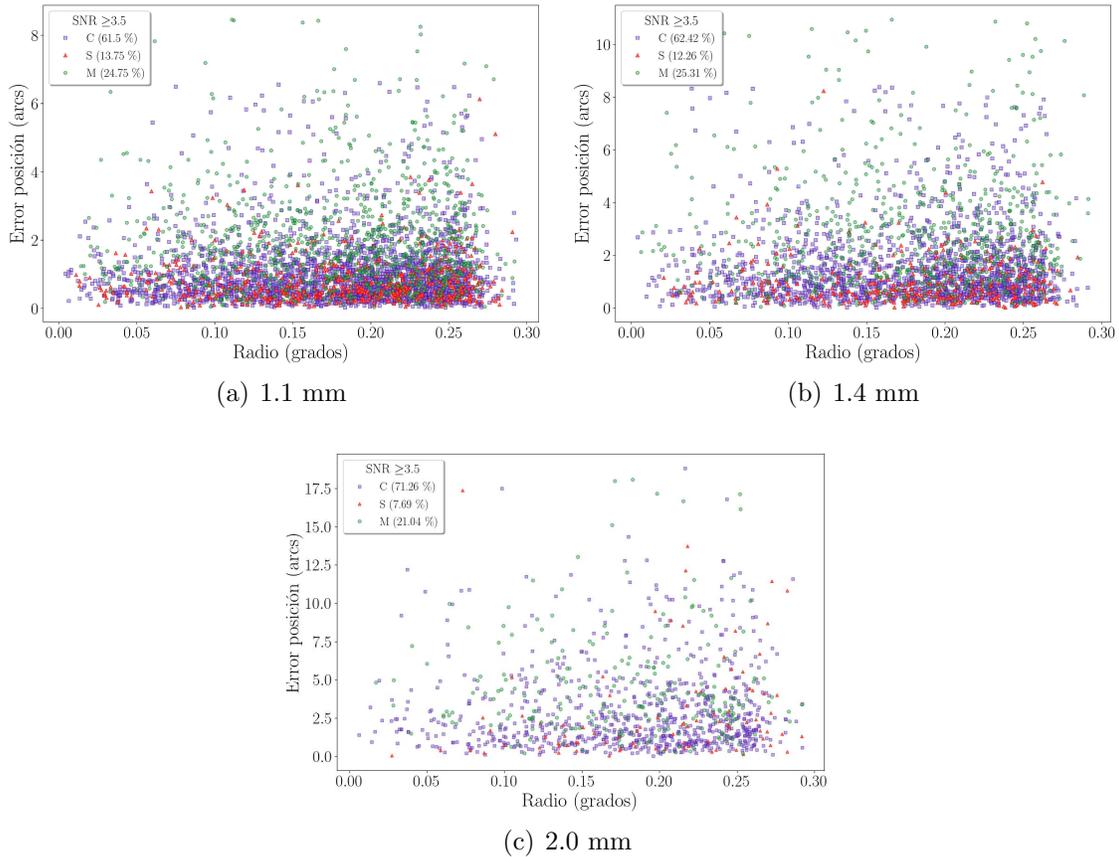


Figura 4.5: Errores de posición en radios concéntricos desde el centro del mapa de la simulación 1 para las tres bandas y cortes en $\text{SNR} \geq 3.5$. Los cuadrados, triángulos y círculos representan las fuentes identificadas como tipo ‘C’, ‘S’ y ‘M’, respectivamente.

contribución de otras fuentes.

Por esta razón, los análisis sobre error de posición se realizan sobre las fuentes tipo ‘S’ y ‘C’ para evitar introducir sesgos en los resultados. Como es de esperarse este tipo de errores aumenta de acuerdo a la longitud de onda, y esto se debe a que el FWHM lo hace también, haciendo más complejo poder resolver las fuentes. Para la simulación 1 no parece haber errores adicionales de posición para diferentes radios concéntricos. Sin embargo, para la simulación 4 (la cual usa un patrón de cartografiado de Lissajous) sí hay un apreciable aumento de error de posición hacia los bordes del mapa (ver apéndices). Sin embargo, los errores de posición incrementan cuando se

usa una atmósfera variable, pero el efecto más notorio se ve en la simulación 4, dado que parece haber un sesgo en las posiciones en δ (ver Figuras C.34 y C.35). Este fue uno de los resultados que se reportaron a los desarrolladores de TolTECA y en los cuales se ha venido trabajando.

Por otro lado, se puede analizar el cociente entre $\alpha_{\text{PyBDSF}}/\alpha_{\text{cat}}$, idealmente debería acercarse a una relación 1:1, tal como se ilustra en la Figura 4.6. En todas las simulaciones (excepto en la Simulación 4) se ve una tendencia constante. Por ejemplo, al calcular los promedios ponderados por intervalos de SNR en la simulación 1 (ver C.1) los valores son ~ 1 . Sin embargo, para un entendimiento más global el análisis se centra en el error de posición radial $\epsilon = \sqrt{\Delta\alpha^2 + \Delta\delta^2}$, como se explicó en la sección 3.4.

El promedio de los errores radiales para cada tipo de estructura por cortes en SNR se muestran en la Tabla 4.4. Como se espera, los errores de posición para las fuentes tipo ‘S’ y ‘C’ son menores y más consistentes que los de tipo ‘M’, que pueden llegar a ser superiores a ~ 3 arcs en la banda de 1.1 mm, y ~ 4 arcs en las de 1.4 y 2.0 mm. En la simulación 1, para las fuentes de tipos ‘S’ y ‘C’ se espera un valor de error de posición radial de $0.61_{-0.32}^{+0.64}$ y $0.77_{-0.43}^{+0.88}$ arcs en la banda de 1.1 mm a un corte de $\text{SNR} \geq 3.5$. Para la banda de 1.4 mm se espera un error de $0.71_{-0.41}^{+0.94}$ y $1.1_{-0.6}^{+1.5}$ arcs a una $\text{SNR} \geq 3.5$ y para la de 2.0 mm con un corte $\text{SNR} \geq 2$ se espera un error de posición de $2.0_{-1.4}^{+2.5}$ y $2.9_{-1.8}^{+4.4}$ arcs. La simulación 3 tiene errores similares (ver Tabla C.6). Sin embargo, en el caso de la simulación 4, dado que el FWHM calculado es mayor respecto a las otras simulaciones (ver Tabla 4.2), para las tres bandas de observación los errores de posición son mayores.

SNR	S		C		M	
	N_s	ϵ_S [arcs]	N_C	ϵ_C [arcs]	N_M	ϵ_M [arcs]
1.1 mm						
≥ 1	1310	$1.0_{-0.6}^{+1.7}$	4949	$1.2_{-0.7}^{+2.0}$	1146	$1.8_{-1.0}^{+1.8}$
≥ 3.5	575	$0.61_{-0.32}^{+0.64}$	2575	$0.77_{-0.43}^{+0.88}$	1035	$1.7_{-0.9}^{+1.6}$
[1, 2)	461	$1.9_{-0.8}^{+2.0}$	907	$2.4_{-1.4}^{+2.5}$	10	$3.0_{-2.2}^{+1.3}$
[2, 3)	207	$1.2_{-0.5}^{+1.4}$	1020	$1.91_{-1.1}^{+2.4}$	49	$3.4_{-1.8}^{+2.5}$

SNR	S		C		M	
	N_s	ϵ_S [arcs]	N_C	ϵ_C [arcs]	N_M	ϵ_M [arcs]
[3, 4)	131	$1.1^{+0.8}_{-0.6}$	798	$1.4^{+1.5}_{-0.8}$	104	$2.5^{+2.7}_{-1.1}$
[4, 6)	145	$0.8^{+1.1}_{-0.5}$	852	$1.1^{+1.0}_{-0.6}$	219	$2.3^{+1.84}_{-1.12}$
[6, 8)	83	$0.72^{+0.68}_{-0.37}$	470	$0.77^{+0.57}_{-0.41}$	203	$1.9^{+1.3}_{-1.0}$
[8, 10)	53	$0.78^{+0.41}_{-0.41}$	251	$0.66^{+0.44}_{-0.40}$	146	$1.7^{+1.5}_{-0.7}$
[10, 20)	110	$0.53^{+0.29}_{-0.26}$	467	$0.50^{+0.40}_{-0.23}$	273	$1.3^{+1.2}_{-0.7}$
≥ 20	120	$0.37^{+0.13}_{-0.12}$	184	$0.37^{+0.15}_{-0.15}$	142	$1.0^{+1.2}_{-0.7}$
1.4 mm						
≥ 1	905	$1.5^{+2.6}_{-1.0}$	3726	$1.8^{+3.0}_{-1.2}$	811	$2.5^{+2.7}_{-1.4}$
≥ 2.5	451	$0.85^{+1.15}_{-0.51}$	2450	$1.3^{+2.0}_{-0.8}$	784	$2.4^{+2.7}_{-1.4}$
≥ 3.5	344	$0.71^{+0.94}_{-0.41}$	1751	$1.1^{+1.5}_{-0.6}$	710	$2.3^{+2.3}_{-1.3}$
[1, 2)	381	$2.8^{+3.0}_{-1.6}$	858	$3.4^{+3.1}_{-1.9}$	9	$4.6^{+2.4}_{-1.2}$
[2, 3)	128	$1.7^{+2.0}_{-0.9}$	788	$2.7^{+3.3}_{-1.5}$	51	$4.4^{+3.1}_{-1.9}$
[3, 4)	98	$1.4^{+1.4}_{-0.8}$	593	$2.0^{+2.4}_{-1.1}$	73	$4.2^{+3.7}_{-2.3}$
[4, 6)	89	$0.9^{+1.1}_{-0.5}$	626	$1.4^{+1.7}_{-0.8}$	172	$3.1^{+2.7}_{-1.6}$
[6, 8)	55	$0.97^{+0.81}_{-0.51}$	342	$1.0^{+1.0}_{-0.6}$	115	$2.9^{+1.4}_{-1.6}$
[8, 10)	33	$0.62^{+0.61}_{-0.22}$	162	$0.82^{+0.71}_{-0.45}$	114	$2.4^{+1.9}_{-1.3}$
≥ 10	121	$0.44^{+0.34}_{-0.22}$	357	$0.56^{+0.48}_{-0.28}$	277	$1.5^{+1.7}_{-0.9}$
2.0 mm						
≥ 1	350	$3.3^{+5.2}_{-2.1}$	2161	$3.5^{+4.8}_{-2.3}$	283	$4.1^{+4.5}_{-2.2}$
≥ 2	153	$2.0^{+2.5}_{-1.4}$	1564	$2.9^{+4.4}_{-1.8}$	280	$4.1^{+4.5}_{-2.2}$
[2, 3)	49	$3.0^{+3.2}_{-1.2}$	544	$5.0^{+4.1}_{-3.0}$	25	$8.0^{+8.9}_{-3.0}$
[3, 4)	34	$1.8^{+3.3}_{-1.0}$	383	$3.4^{+3.8}_{-1.8}$	35	$5.9^{+5.0}_{-2.3}$

SNR	S		C		M	
	N_s	ϵ_S [arcs]	N_C	ϵ_C [arcs]	N_M	ϵ_M [arcs]
[4, 6)	35	$2.2^{+1.7}_{-1.7}$	334	$2.4^{+3.5}_{-1.3}$	65	$4.5^{+5.0}_{-2.1}$
[6, 8)	15	$0.9^{+1.3}_{-0.4}$	133	$1.7^{+1.6}_{-0.9}$	51	$3.4^{+2.0}_{-1.5}$
≥ 8	20	$0.8^{+1.5}_{-0.5}$	168	$0.9^{+1.2}_{-0.4}$	104	$3.0^{+3.1}_{-1.5}$

Tabla 4.4: Estimaciones del error de posición radial por cortes de señal a ruido para las tres bandas, donde ‘SNR’ indica el umbral o rango de señal a ruido considerado, las letras S, C y M indican el tipo de estructura de fuente identificada, ‘N’ el número de fuentes y ϵ la estimación del error junto a sus incertidumbres asociadas a un intervalo de confianza del 68 %.

Las Figuras 4.7(a), 4.7(b) y 4.8, presentan los histogramas de errores de posiciones de las fuentes para las tres bandas. En estas gráficas se representan los *offsets* en α y δ , así como el resultado conjunto de los errores en α y δ . A este último se le ajusta una gaussiana, que para el caso de la banda de 1.1 mm da $\sigma = (0.54 \pm 0.02)$ arcs. En la banda de 1.4 mm se tiene un resultado de $\sigma = (0.77 \pm 0.02)$ arcs y en la de 2.0 mm $\sigma = (1.419 \pm 0.004)$ arcs. Estos resultados muestran que los radios de búsqueda usados en la simulación 1 (ver Tabla 4.2) corresponden a $\sim 5\sigma$.

Así mismo, dentro de estas Figuras (4.7(a), 4.7(b) y 4.8) se presenta el histograma de errores radiales ϵ , junto con el ajuste a los datos, según la ecuación de Ivison et al. (2007) para el tamaño del haz sintético. Adicionalmente se representa el ajuste con esta ecuación dejando el valor del tamaño del haz como parámetro libre para cada banda. Como puede verse, en todos los casos la distribución analítica de Ivison et al. (2007) con el tamaño de haz sintético es consistente y comparable con la obtenida dejando el tamaño del haz libre. Sin embargo, a pesar de que las SNR fueron corregidas por el sesgo de Malmquist (ver ecuación 3.4.6), se presenta un exceso de probabilidad a SNR grandes. Esto puede deberse a que en esta tesis se trabaja con densidades de flujo muy pequeños en un campo muy confuso. Adicionalmente, la distribución ajustada en el histograma de errores conjuntos no está centrada en cero mientras que en Ivison et al. (2007) sí lo está.

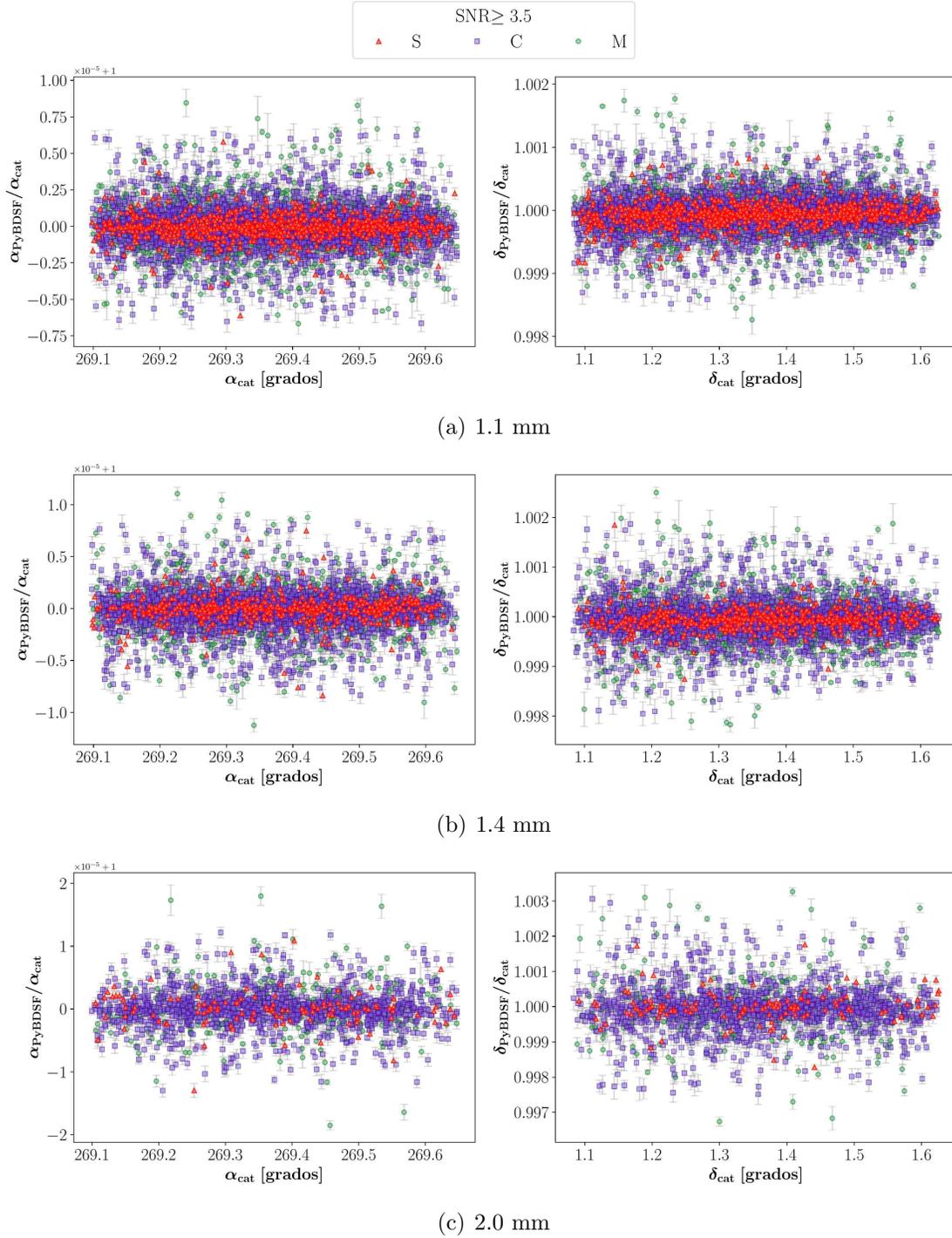
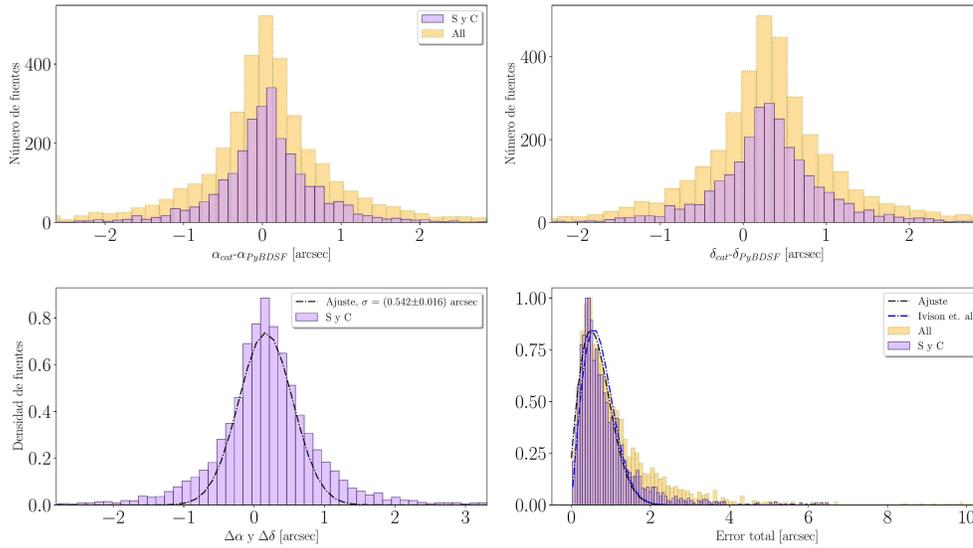
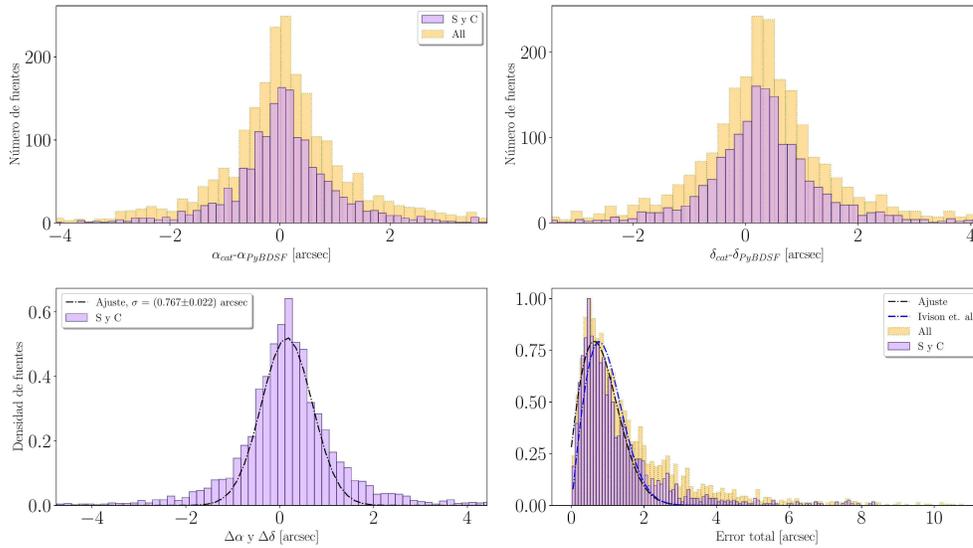


Figura 4.6: Cocientes de las posiciones en α , δ respecto a las de entrada para cada una de las tres bandas en la simulación 1. Los puntos rojos, violetas y verdes corresponden a las fuentes tipo ‘S’, ‘C’ y ‘M’.



(a) 1.1 mm



(b) 1.4 mm

Figura 4.7: Histogramas de errores de posición para las fuentes detectadas en las bandas de 1.1 mm y 1.4 mm para la simulación 1. La primera fila presenta el error en α y δ . La gráfica inferior izquierda presenta el resultado del histograma conjunto de los errores en α y δ y la gráfica inferior derecha, el error de posición radial, donde la línea discontinua negra muestra el ajuste de la ecuación (3.4.7) dejando el tamaño del haz como parámetro libre y la línea azul el ajuste fijando el tamaño del haz al valor sintético.

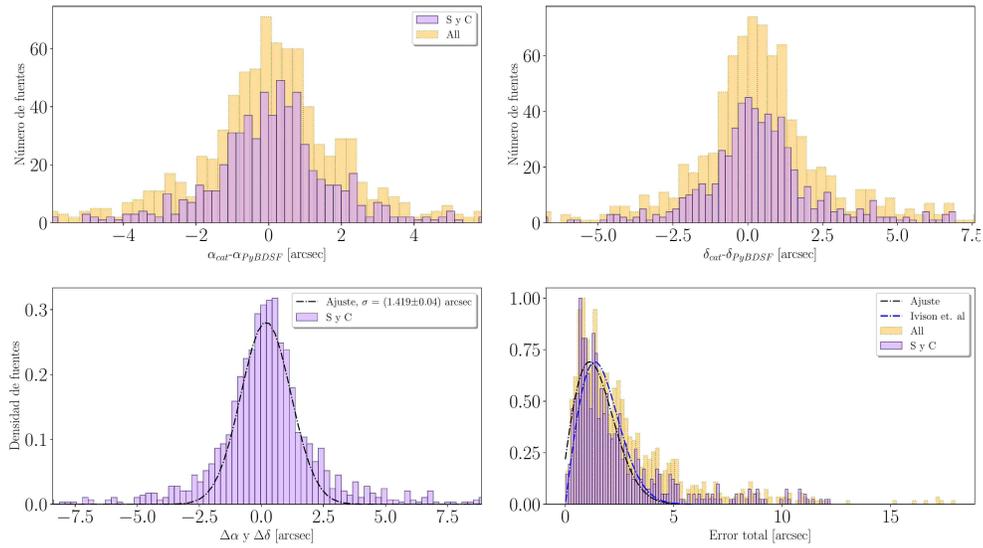


Figura 4.8: Histogramas de errores de posición para las fuentes detectadas en la banda de 2.0 mm para la simulación 1. La primera fila presenta el error en α y δ . La gráfica inferior izquierda presenta el resultado del histograma conjunto de los errores en α y δ y la gráfica inferior derecha, el error de posición radial, donde la línea discontinua negra muestra el ajuste de la ecuación (3.4.7) dejando el tamaño del haz como parámetro libre y la línea azul el ajuste fijando el tamaño del haz al valor sintético.

4.4. Análisis de la recuperación de densidades de flujo

En el desarrollo de TolTECA y Citlali, uno de los puntos clave abordados ha sido la correcta recuperación de la densidad de flujo. A la fecha de elaboración de este trabajo, las simulaciones presentadas muestran evidencia de una ligera pérdida de densidad de flujo. Para caracterizar esta situación, se lleva a cabo una comparación entre la densidad de flujo de entrada y los medidos con el programa PyBDSF. Como se mencionó previamente, la densidad de flujo asignada a cada tipo de fuente (S, C y M) corresponde a la fuente más brillante dentro de su radio de búsqueda.

El resultado de comparar la densidad de flujo de entrada del catálogo de Nava-Moreno et al. (2019) y las densidades de flujo obtenidos mediante el procedimiento de extracción de fuentes se presenta en las Figuras 4.1, 4.2(a) y 4.2(b). Para las

fuentes que tienen densidades de flujo más débiles no se espera una relación 1:1 en densidad de flujo debido al efecto de *boosting*. Este efecto ocasiona que en las fuentes detectadas con bajas SNR, se midan densidades de flujo sistemáticamente mayores que sus densidades de flujo de entrada a medida que el número de fuentes aumenta hacia flujos más débiles (Hogg & Turner, 1998). Por lo tanto, el siguiente paso consiste en caracterizar este efecto mediante el cálculo del cociente entre la densidad de flujo medido y la densidad de flujo de entrada. Los resultados se presentan en la Figura 4.9 para las tres bandas. En este caso, se calcula el factor de *boosting* ($S_{\nu,Out}/S_{\nu,Inp}$) para cada *bin* de SNR, mediante un promedio ponderado. Las barras de error representan el intervalo de confianza al 68 %. Llama la atención que el *boosting* muestre un valor inferior a 1, ya que a baja SNR el *boosting* se espera represente un incremento de la densidad de flujo detectada. Esto se explica por el hecho de que la SNR utilizada en la gráfica es la derivada de las medidas de PyBDSF, que presenta una ligera pérdida de flujo. Esta situación se ilustra de manera más clara al analizar los histogramas de cociente de densidades de flujo por cada tipo de fuente, tal como se muestra en la Figura 4.10. Allí se representan las fuentes con una $SNR \geq 3.5$, y puede observarse que caracterizan una media ponderada del cociente de densidades de flujo en la banda de 1.1 mm de $0.93^{+0.29}_{-0.21}$, en la de 1.4 mm de $0.90^{+0.39}_{-0.19}$ y en la de 2.0 mm de $0.92^{+0.65}_{-0.18}$. Esto implica una pérdida de flujo⁴ para cada banda de aproximadamente el 8 %, 11.9 % y 17 % en la simulación 1.

Dado que las densidades de flujo a la salida experimentan esa pérdida, se realiza el mismo procedimiento pero considerando las densidades de flujo de entrada, es decir, aquellas fuentes que fueron asignadas a las detecciones de PyBDSF. Las Tablas 4.5, 4.6 y 4.7 muestran el cociente medio ponderado $S_{\nu,PyBDSF}/S_{\nu,Input}$ para cada banda, en función de rangos de densidades de flujo de entrada. Para ver mejor el efecto del *boosting*, se presentan los cocientes de densidad de flujo promedio por intervalos de densidades de flujo de entrada en las Figuras 4.11, 4.12 y 4.13. Las barras de error indican el intervalo de confianza al 68 %. Así mismo, la gráfica 4.14 muestra esta relación en escala logarítmica. Para señales a ruido muy bajas ($SNR < 3.5$), puede existir un incremento en la señal debido a las fuentes contaminantes de ~ 1.5 , ~ 2.5 , y ~ 3.5 a 1.1, 1.4 y 2.0 mm, respectivamente.

⁴Este fue un problema reportado a los desarrolladores de TolTEC y Citlali y actualmente se está trabajando en ello.

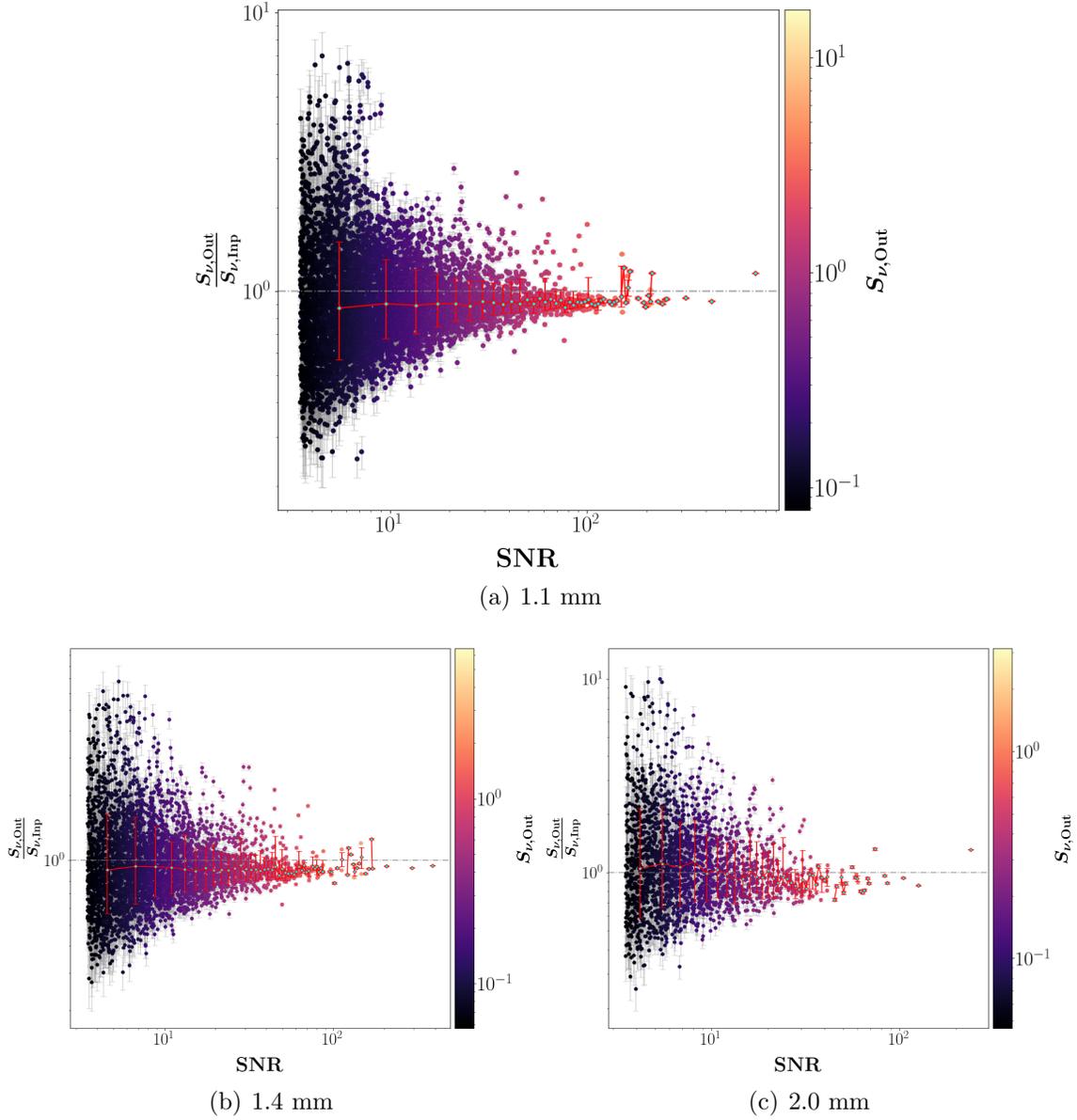


Figura 4.9: Factor de *boosting* estimado de la simulación, que se calcula como la razón entre la densidad de flujo medida por PyBDSF $S_{\nu,Out}$ y el flujo de entrada del catálogo de Nava-Moreno et al. (2019). La línea roja representa el promedio ponderado del factor de *boosting* para cada intervalo de señal a ruido medido (*bin*). Nótese que el valor del cociente inferior a 1 se debe a la pérdida de densidad de flujo experimentada en las etapas de simulación de la observación. Las barras de error indican el intervalo de confianza del 68%

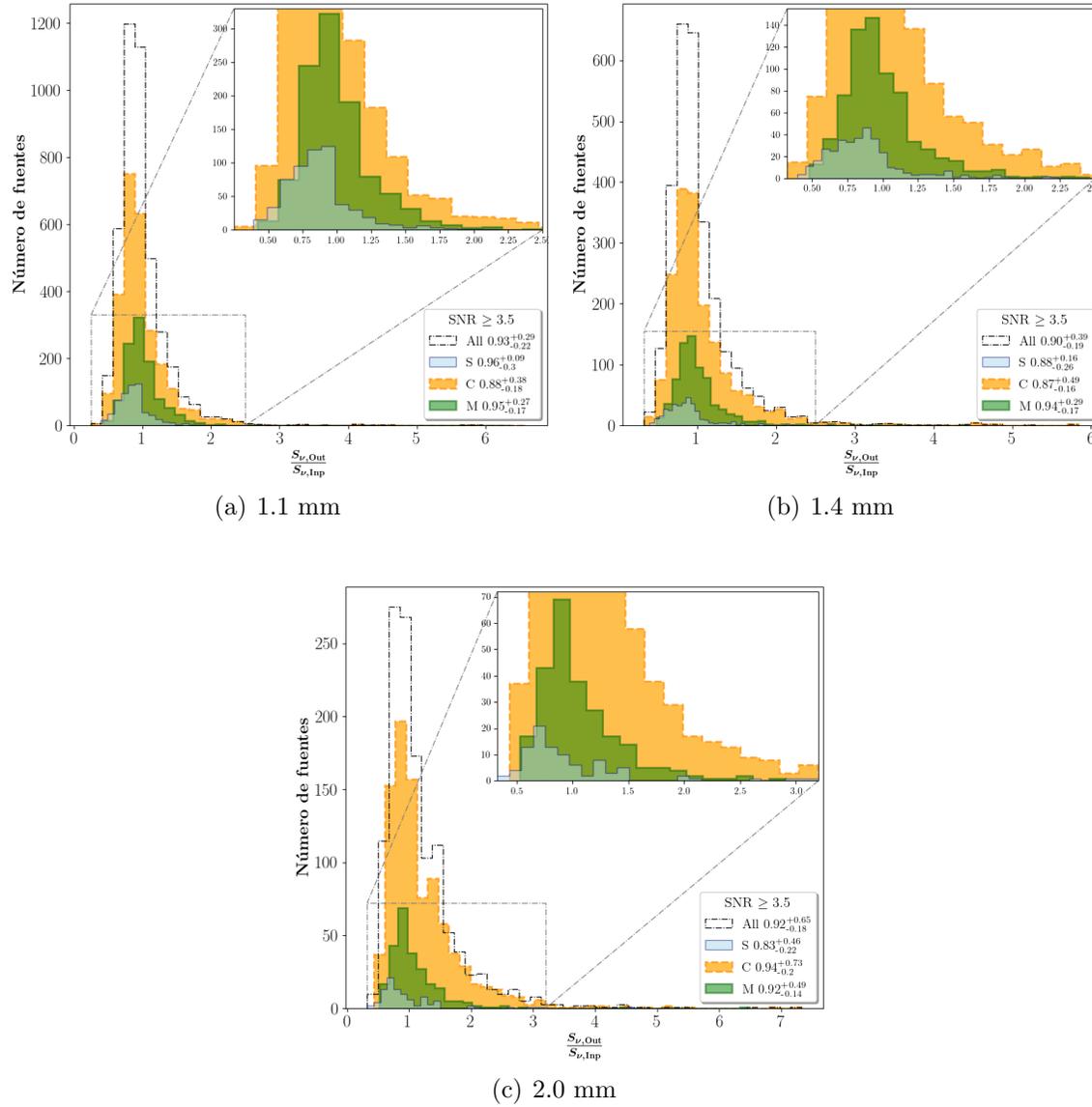


Figura 4.10: Histogramas del cociente de densidades de flujo de salida y entrada para las fuentes detectadas. Se muestran en azul, naranja y verde los histogramas para las fuentes tipo ‘S’, ‘C’ y ‘M’ y en línea punteada el histograma conjunto.

En general, las simulaciones que incorporan una atmósfera variable suelen mostrar un mayor efecto de *boosting*, posiblemente debido a la presencia de artefactos que no se pudieron eliminar mediante el análisis de componentes principales (PCA). Por ejemplo, en la simulación 3 (consultar Figuras C.24), se observan factores promedio máximos de aumento de ~ 3 , ~ 4.5 , y ~ 8 a 1.1, 1.4 y 2.0 mm, respectivamente, debido a la presencia de artefactos en dicha simulación. Un caso similar se presenta en la simulación 4 (ver Figura C.38). En contraste, en la simulación 2 (Figura C.10), que utiliza una atmósfera estática sin artefactos, los factores de aumento son de aproximadamente ~ 1.6 , ~ 1.7 , y ~ 3 en cada banda. Esto demuestra la importancia de eliminar correctamente los efectos atmosféricos. e

$S_{\nu, \text{Inp}}$ [mJy]	$\frac{S_{\nu, \text{Out}}}{S_{\nu, \text{Inp}}}$		
	S	C	S y C
[0.02, 1.41)	0.834 ± 0.002	0.85 ± 0.001	0.846 ± 0.001
[1.41, 2.81)	0.859 ± 0.002	0.87 ± 0.002	0.865 ± 0.001
[2.81, 4.2)	0.909 ± 0.002	0.818 ± 0.008	0.903 ± 0.002
[4.2, 5.59)	0.896 ± 0.002	0.894 ± 0.004	0.896 ± 0.002
[5.59, 6.99)	0.905 ± 0.002	0.913 ± 0.004	0.907 ± 0.002
[6.99, 8.38)	0.876 ± 0.004	–	0.876 ± 0.004
≥ 8.38	1.045 ± 0.001	–	1.013 ± 0.001

Tabla 4.5: Cocientes de densidades de flujo de salida y entrada para la banda de **1.1 mm**. La primera columna muestra los intervalos de valores de densidades de flujo del catálogo de Nava-Moreno et al. (2019), la segunda y tercera columna el cociente medio ponderado de densidades de flujo recuperada para las fuentes tipo S y C respectivamente y la última columna el cociente medio combinado de ambas. Los símbolos ‘–’ indican los intervalos de densidades de flujo en donde no se encontraron ese tipo de fuentes.

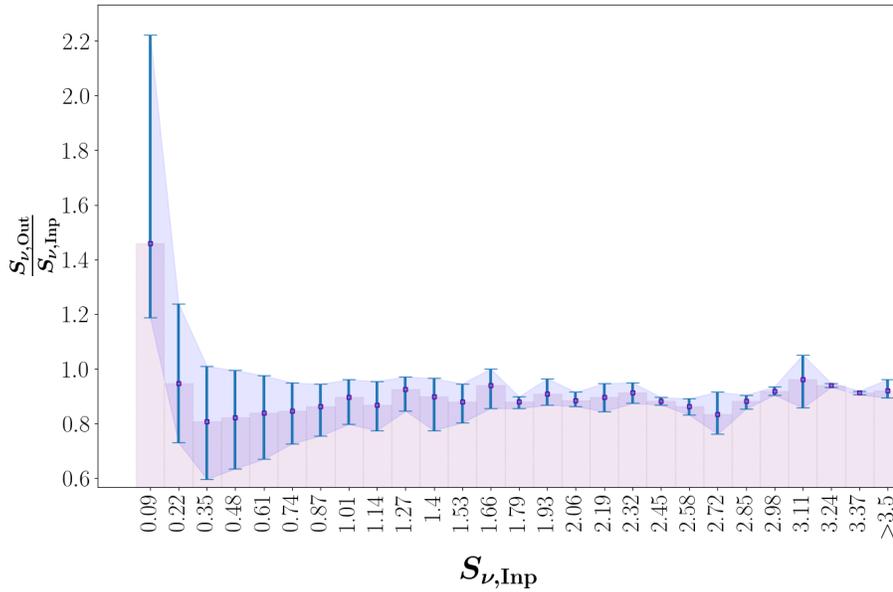


Figura 4.11: Cociente medio ponderado de densidades de flujo a **1.1 mm** por rangos de densidades de flujo de entrada. Las barras de error indican el intervalo de confianza del 68 %.

Rangos por flujo de entrada [mJy]	$\frac{S_{\nu, \text{Out}}}{S_{\nu, \text{Inp}}}$		
	S	C	S y C
[0.01, 0.76)	0.759 ± 0.003	0.836 ± 0.001	0.819 ± 0.001
[0.76, 1.51)	0.849 ± 0.003	0.841 ± 0.002	0.845 ± 0.002
[1.51, 2.25)	0.891 ± 0.003	0.835 ± 0.008	0.883 ± 0.003
[2.25, 3.0)	0.946 ± 0.005	0.812 ± 0.004	0.864 ± 0.003
[3.0, 3.75)	0.897 ± 0.006	0.879 ± 0.004	0.886 ± 0.003
≥ 3.75	0.906 ± 0.003	–	0.906 ± 0.003

Tabla 4.6: Cocientes de densidades de flujo de salida y entrada para la banda de **1.4 mm**. La primera columna muestra los intervalos de valores de densidades de flujo del catálogo de Nava-Moreno et al. (2019), la segunda y tercera columna el cociente medio ponderado de densidades de flujo recuperada para las fuentes tipo S y C respectivamente y la última columna el cociente medio combinado de ambas. Los símbolos ‘–’ indican los intervalos de densidad de flujo en donde no se encontraron ese tipo de fuentes.

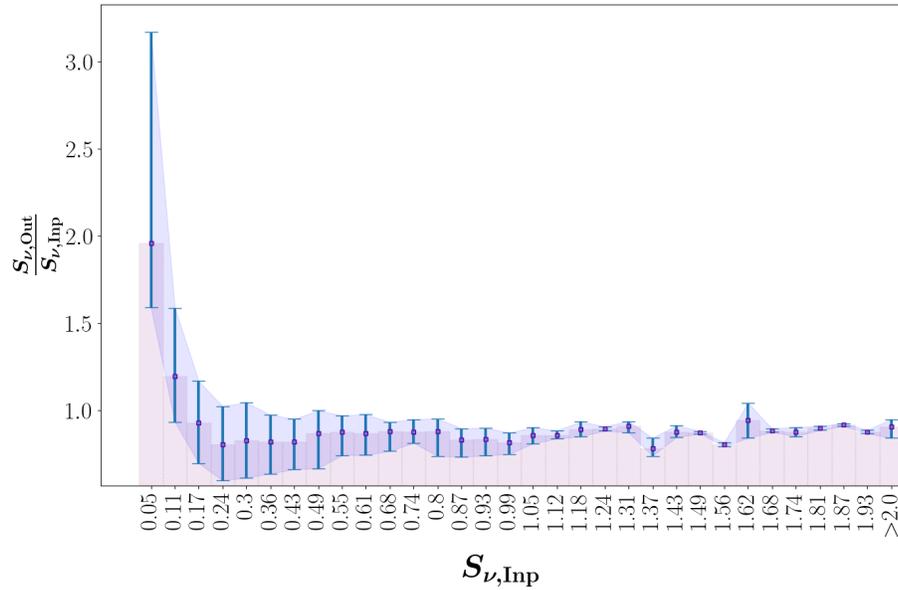


Figura 4.12: Cociente medio ponderado de densidades de flujo a **1.4 mm** por rangos de densidades de flujo de entrada. Las barras de error indican el intervalo de confianza del 68 %.

Rangos por flujo de entrada [mJy]	$\frac{S_{\nu,Out}}{S_{\nu,Inp}}$		
	S	C	S y C
[0.02, 0.26)	0.749 ± 0.013	0.892 ± 0.004	0.878 ± 0.004
[0.26, 0.5)	0.677 ± 0.009	0.742 ± 0.004	0.731 ± 0.004
[0.5, 0.74)	0.682 ± 0.013	0.758 ± 0.009	0.734 ± 0.007
[0.74, 0.98)	–	0.800 ± 0.007	0.800 ± 0.007
[0.98, 1.22)	–	0.776 ± 0.006	0.776 ± 0.006
≥ 1.22	0.812 ± 0.005	1.202 ± 0.005	1.005 ± 0.004

Tabla 4.7: Cocientes de densidades de flujo de salida y entrada para la banda de **2.0 mm**. La primera columna muestra los intervalos de valores de densidades de flujo del catálogo de Nava-Moreno et al. (2019), la segunda y tercera columna el cociente medio ponderado de densidades de flujo recuperada para las fuentes tipo S y C respectivamente y la última columna el cociente medio combinado de ambas. Los símbolos ‘–’ indican los intervalos de densidad de flujo en donde no se encontraron ese tipo de fuentes.

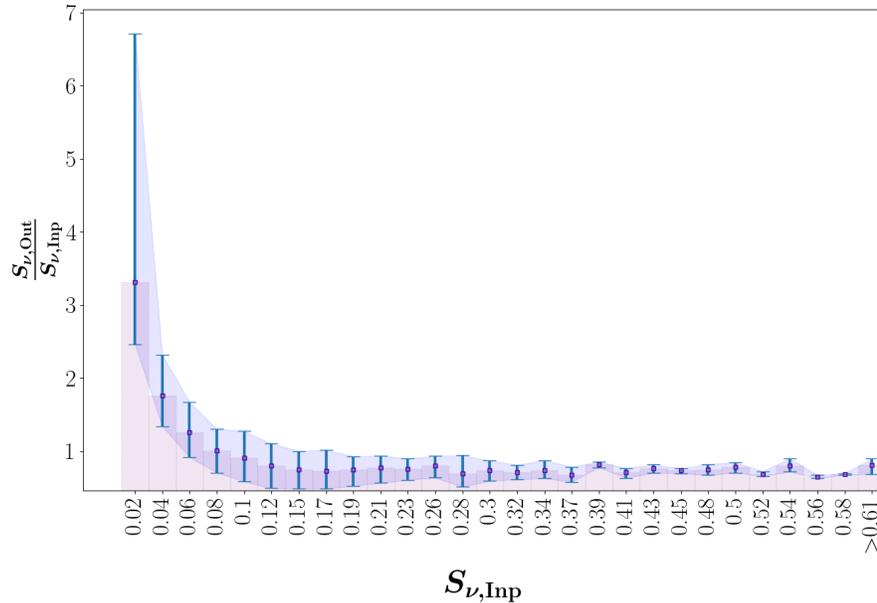


Figura 4.13: Cociente medio ponderado de densidades de flujo a **2.0 mm** por rangos de densidades de flujo de entrada. Las barras de error indican el intervalo de confianza del 68 %.

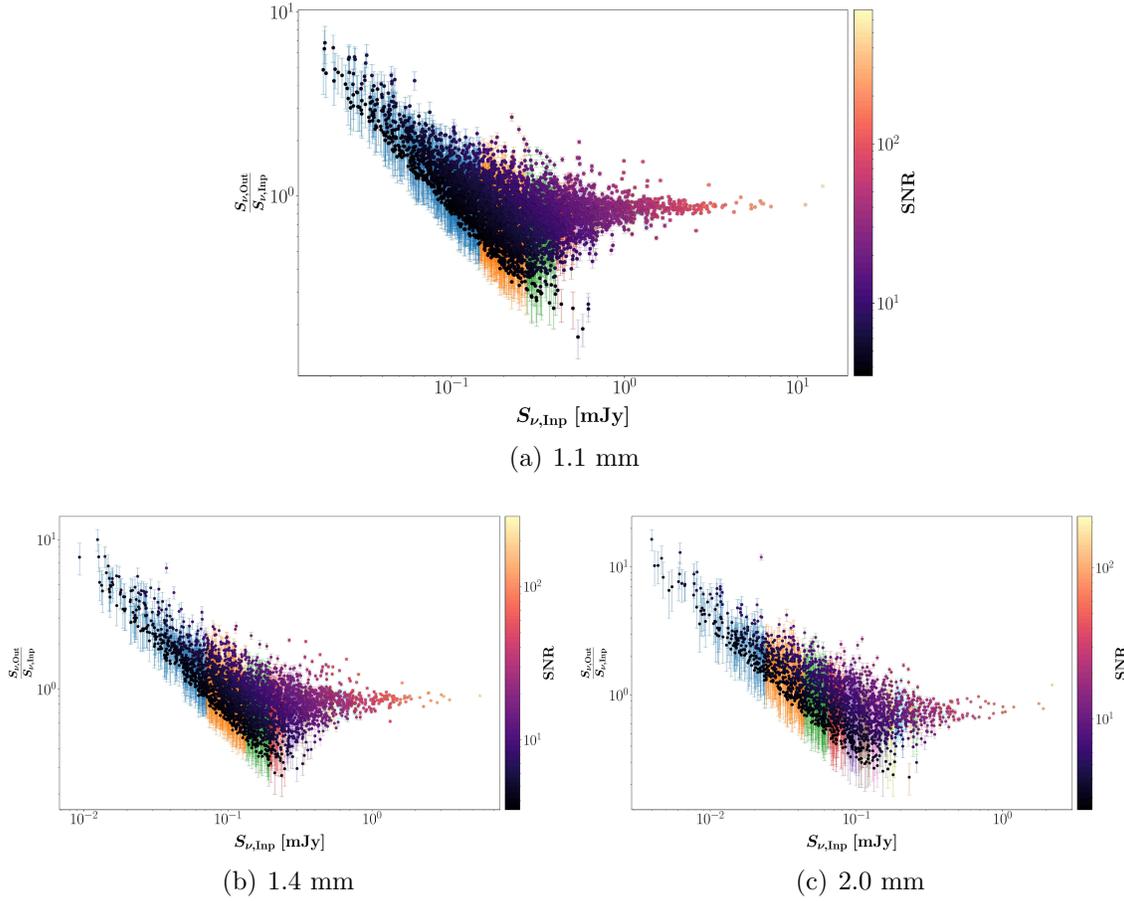


Figura 4.14: Cociente de densidad de flujo de salida respecto al de entrada ($S_{\nu,Out}/S_{\nu,Inp}$) en escala logarítmica. Las barras de error indican un intervalo de confianza del 68 %, mientras que los colores muestran la señal a ruido local en cada medida.

4.5. Completez de los catálogos

Para cuantificar la eficiencia de detección de fuentes es necesario realizar simulaciones en las cuales se evalúe la capacidad que tiene el método para identificarlas, a esto se le conoce como completez (Scott et al., 2008). Usualmente lo que se hace es introducir fuentes en posiciones aleatorias, una por una, en un mapa de señal que esté escasamente poblado (Casey et al., 2014). Luego se corre el algoritmo estándar de extracción de fuentes y se analiza si son detectadas o no. En Scott et al. (2008) se

introducen 1000 fuentes por cada *bin* de densidad de flujo.

Dado que estamos ante un mapa altamente confundido, lo que se hace en esta tesis es buscar cuántas fuentes del catálogo de Nava-Moreno et al. (2019) que están dentro de la huella de observación fueron asignadas a las fuentes obtenidas en el proceso de extracción de PyBDSF. Los resultados para la simulación 1 se presentan en la Figura 4.15 junto al ajuste esperado descrito por Bing et al. (2023):

$$C(x) = \frac{1 + \operatorname{erf}\left(\frac{x-x_0}{y_0}\right)}{2} \quad (4.5.1)$$

siendo x_0 y y_0 parámetros libres y x el rango de densidades de flujo de entrada. Así pues, estos resultados sugieren que para la banda de 1.1 mm el catálogo está completo en un 40 % en ~ 0.1 mJy (4σ) y 74 % completo en ~ 0.2 mJy. En la banda de 1.4 mm en ~ 0.07 mJy (4σ) el catálogo está completo a un 30 % y a ~ 0.23 mJy está 87 % completo. Finalmente en la banda de 2.0 mm el 41 % se encuentra en ~ 0.06 (4σ) y el 90 % a ~ 0.15 mJy. Los resultados para las simulaciones 2 y 3 (ver Figuras C.14 y C.28) indican resultados similares, tomando en cuenta el *rms* presentado en cada caso (ver Tabla 4.1). La simulación 4 presenta valores mucho peores. Sin embargo, cabe

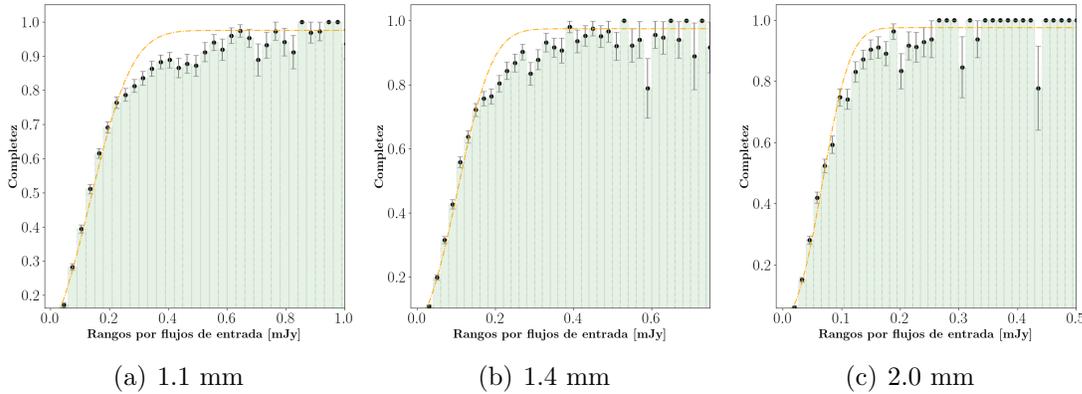


Figura 4.15: Completez obtenida para la simulación 1. Los puntos en la gráfica junto con las barras de error estimadas muestran las fuentes que han sido identificadas y asociadas con las fuentes de entrada en cada *bin* de densidad de flujo. Además, se ha trazado una línea discontinua de color naranja que representa la función de mejor ajuste de los datos con la ecuación (4.5.1).

resaltar que ésta sólo simula 20 minutos en tiempo de integración, de modo que es de esperarse que no se tenga una estadística tan alta como en las otras simulaciones. Las barras de error, en estas gráficas corresponden a un intervalo de confianza del 68% usando una aproximación gaussiana a una distribución binomial⁵

Se analiza a continuación casos que afectan a la completez del catálogo. La Tabla 4.8 presenta el número de fuentes de entrada del catálogo de Nava-Moreno et al. (2019), dentro de la huella del mapa de la simulación 1, para cortes de densidades de flujo de entrada, el número de fuentes detectadas con PyBDSF y el número de fuentes de entrada que fueron asignadas como verdaderas por encima de ese umbral.

$S_{\nu, \text{umbral}}$ (mJy)	N_{cat}	N_{asig} ($S_{\nu} \geq S_{\nu, \text{umbral}}$)	N_{PyBDSF}
banda 1.1 mm			
0.025	14545	7366	7400
0.05	12287	7020	7155
0.075	10178	6529	6234
0.1	8562	5999	5352
0.125	7185	5414	4612
0.15	6071	4835	3920
0.175	5140	4269	3435
0.2	4453	3802	3008
0.225	3853	3356	2701
0.25	3385	2996	2393
banda 1.4 mm			
0.02	13308	5343	5442
0.04	9887	4972	5088
0.06	7501	4498	4187
0.08	5731	3939	3488
0.1	4498	3412	2908
0.12	3635	2930	2424

⁵Consultar https://www.statsmodels.org/dev/generated/statsmodels.stats.proportion.proportion_confint.html.

$S_{\nu, \text{umbral}}$ (mJy)	N_{cat}	N_{PyBDSF} ($S_{\nu} \geq S_{\nu, \text{umbral}}$)	N_{asig}
0.14	3013	2533	2069
0.16	2534	2187	1794
0.18	2168	1910	1548
0.2	1854	1670	1354
banda 2.0 mm			
0.013	10159	2710	2794
0.026	6024	2455	2679
0.039	3895	2133	2163
0.052	2769	1816	1725
0.065	2054	1516	1376
0.078	1573	1264	1090
0.091	1271	1085	882
0.104	1037	910	732
0.117	863	781	591
0.13	739	678	505

Tabla 4.8: Número de fuentes recuperadas sobre umbrales de densidades de flujo de entrada. La primera columna muestra el corte por densidades de flujo de entrada S_{ν} (mJy). La segunda columna, el número de fuentes de entrada sobre ese umbral. La tercera columna, el número de fuentes medidas a la salida sobre ese umbral y la última columna, el número de fuentes que fueron asignadas sobre ese umbral.

A modo de ejemplo, nótese que en el corte sobre 0.1 mJy a 1.1 mm, a la entrada se tienen 8,562 fuentes, se logran recuperar 5,999 por encima de este umbral de densidad de flujo (esto es un $\sim 70\%$ del catálogo de entrada). Sin embargo, el programa detecta 5,352 fuentes por encima ese umbral (esto es un $\sim 62\%$ del catálogo de entrada). La completez se mide sobre fuentes recuperadas sin importar a qué densidad de flujo estas correspondan.

A este efecto se le suma otro que se evidenció al revisar los mapas de señal: hay muchas fuentes que se mezclan entre sí. Para ello se ilustran con dos ejemplos provistos en la Figura 4.16. Cómo se explicó, se elige la fuente más brillante dentro del radio

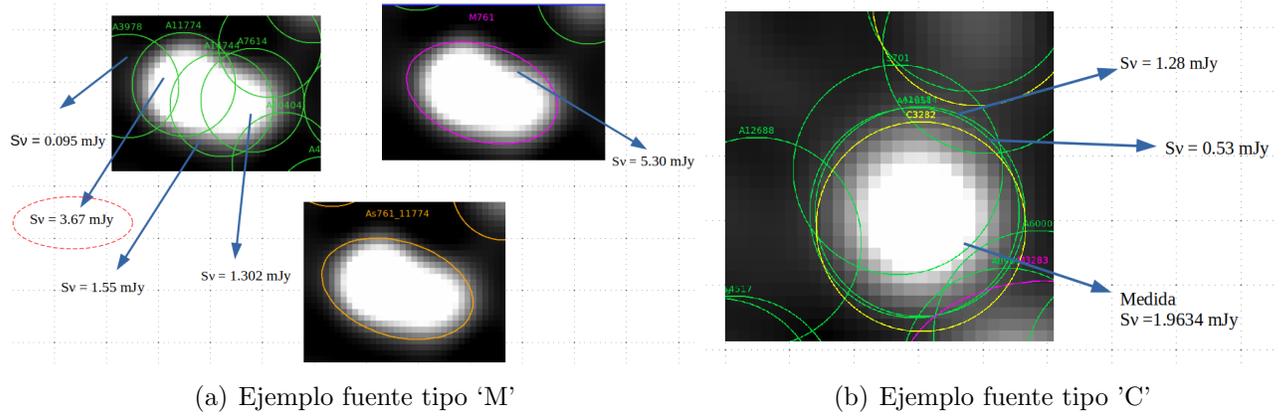


Figura 4.16: Ejemplos de fuentes detectadas con densidades de flujo altas. Se muestra la composición de una fuente tipo 'M' y una de tipo C.

de búsqueda para hacer la correspondencia entre fuentes de entrada y salida. En el caso de las fuentes tipo 'M' el programa no logra descomponer las fuentes de entrada que contribuyen a la isla, de modo que la densidad de flujo de una fuente M está integrada por las densidades de flujo de sus componentes. Muchas fuentes de tipo 'M' están compuestas por múltiples fuentes brillantes que no se pueden aislar una de la otra y por lo tanto, aunque las densidades de flujo de entrada son grandes respecto al *rms* del mapa, la completez para esa densidad de flujo no alcanza el 100%. El mismo caso se da para las fuentes tipo 'C' donde una o más fuentes caen dentro de un haz.

Conclusiones y trabajo futuro

En esta tesis, se presentaron los resultados de cuatro simulaciones de observaciones realizadas con TolTEC y el Gran Telescopio Milimétrico en las bandas de 1.1, 1.4 y 2.0 mm, tal como se describe en el capítulo 2. Estas simulaciones se llevaron a cabo utilizando los paquetes de TolTECA (sección 2.1) y posteriormente se redujeron mediante el flujo de trabajo (*pipeline*) Citlali (sección 2.2), los cuales fueron diseñados específicamente para las observaciones con este instrumento. Tres de estas simulaciones incluyeron la presencia de una atmósfera variable y otra que considera una atmósfera estática que representa las mejores condiciones posibles de observación. Se presentan 3 simulaciones con patrones de cartografiado *raster*, dos de estas con una velocidad de telescopio de 50 arcs/s y otra con 3 arcmin/s cubriendo un área de $\sim 0.22 \text{ deg}^2$, y una última simulación con patrón Lissajous $\sim 0.023 \text{ deg}^2$ y velocidades angulares de telescopio $\omega_x = 5.2 \text{ rad/s}$ y $\omega_y = 4 \text{ rad/s}$.

Se aplicó la técnica de análisis por componentes principales (PCA) a todas las simulaciones con el fin de eliminar la influencia de la atmósfera, seguido de un proceso de filtrado. Los resultados mostraron la presencia de artefactos causados por la presencia de la atmósfera en mapas con áreas extensas. Este hallazgo se confirmó en distintas simulaciones a diferentes velocidades y mediante diferentes procesos de PCA. Adicionalmente, se identificó que el filtro Wiener aumenta la anchura a mitad de altura (FWHM) esperada para cada banda, de modo que se presentaron algunas estructuras que no son usuales en el *kernel*. Esto provoca aumentos considerables en el valor del FWHM con el cual se ajustan las gaussianas para extraer las fuentes. Este efecto se vio especialmente pronunciado en el mapa con cartografiado Lissajous¹.

¹Es importante mencionar que el equipo de TolTEC está actualmente trabajando en la resolución

Considerando los productos de ciencia por cada simulación y a su vez en cada banda, se llevaron a cabo una serie de pasos para extraer el catálogo de fuentes utilizando el paquete PyBDSF (ver Capítulo 3). Las fuentes extraídas se clasificaron según el tipo de estructura (S: fuente gaussiana única en la isla; C: fuente gaussiana con otras en la isla; M: fuente múltiple). Se asignaron las fuentes del catálogo de entrada dentro de radios de búsqueda conservativos, mostrando que en general ese radio incluye $\sim 99\%$ de probabilidad de asignar correctamente las DSFGs de Nava-Moreno et al. (2019) que produce ese flujo a la salida. En todos los resultados, se observa una consistencia en la pérdida de flujo de las fuentes a la salida. Esta pérdida de flujo se debe a un defecto del paquete de simulación de TolTECA, en el cual se ha estado trabajando. La pérdida de flujo es de alrededor del 3%, 8% y del 7.9%, para 1.1, 1.4 y 2.0 mm, respectivamente y en el caso de las mejores condiciones de observación.

Esta pérdida de flujo inherente en todas las simulaciones afecta los resultados de *boosting* ya que, si se analiza desde el punto de vista de los flujos de entrada, se ve que a flujos débiles hay un incremento sistemático en el flujo medido a la salida, pero si se observan las SNR medidas no se ve tan evidente este efecto. El *boosting* es mucho más pronunciado en el caso de las simulaciones con atmósfera variable, donde se observan más artefactos, como el caso de la simulación con velocidad de telescopio de 3 arcmin/s. En este último el *boosting* puede llegar a ser de ~ 3 , ~ 4.5 , y ~ 8 veces en cada banda, respectivamente, debido a la fuerte presencia de artefactos. En contraste, en el caso de la simulación con una atmósfera estática, los factores de aumento son de aproximadamente ~ 1.6 , ~ 1.7 , y ~ 3 en cada banda. Esto demuestra la importancia de eliminar correctamente los efectos atmosféricos. Este resultado ha influido en la investigación del equipo de TolTEC encargado de la simulación atmosférica.

El porcentaje de fuentes falsas esperado para la simulación con *raster* y atmósfera variable, es cercano a cero para cortes de $\text{SNR} \geq 3.5$ en la banda de 1.1 mm y 1.4 mm y $\text{SNR} \geq 2.5$ para la de 2.0 mm. En ausencia de artefactos, no se esperan fuentes falsas a un corte en $\text{SNR} \geq 3$ en la banda de 1.1 mm y 1.4 mm, y a partir de $\text{SNR} \geq 2$ para la banda de 2.0 mm.

Al comparar los resultados, se muestra que un patrón de cartografiado tipo *raster*,

de este y otros problemas identificados

con una velocidad de telescopio de 50 arcsec/s y un marco de referencia altacimutal, puede ser una estrategia efectiva para el estudio de campos como el propuesto para UDS. No obstante, dado al costo operacional de una sola observación, aumentar la velocidad podría ser una opción viable, aunque con el inconveniente de perder información de fuentes más débiles en los campos.

Los errores de posición, ϵ , para la simulación de un patrón de cartografiado *raster* con velocidad de barrido de 50 arcsec/s fueron de $\sim 0.61_{-0.32}^{+0.64}$ arcsec en la banda de 1.1 mm para fuentes detectadas con $\text{SNR} \geq 3.5$. Estos valores son comparables con los resultados analíticos propuestos por Ivison et al. (2007). Sin embargo, en el caso del cartografiado Lissajous se observa un sesgo en la astrometría de forma radial presentándose errores de posición de alrededor de $\sim 1.9''$. Los resultados para la simulación 1 (raster a 50 arcsec/s) muestran que para la banda de 1.1 mm el catálogo está completo en un 40% en ~ 0.1 mJy (4σ) y 74% completo en ~ 0.2 mJy. En la banda de 1.4 mm en ~ 0.07 mJy (4σ) el catálogo está completo a un 30% y a ~ 0.23 mJy está 87% completo. Finalmente en la banda de 2.0 mm el 41% se encuentra en ~ 0.06 (4σ) y el 90% a ~ 0.15 mJy. La simulación 4 (Lissajous) presenta valores mucho peores. Sin embargo, cabe resaltar que ésta sólo simula 20 minutos en tiempo de integración, de modo que es de esperarse que no se tenga una estadística tan robusta como las otras simulaciones.

Los resultados proporcionados por las simulaciones han permitido explorar la mejor combinación de parámetros de entrada en PyBDSF, lo cual permite obtener un catálogo robusto. El catálogo de salida contiene información sobre las posiciones detectadas, densidades de flujo total y pico (lo cual permitirá calcular luminosidades y tasas de formación estelar, SFR), estructura de las fuentes ajustadas, entre otra información. Estos resultados han aportado en el desarrollo de TolTECA y Citlali, ya que han permitido identificar y caracterizar las falencias y las ventajas de ambos. Esto es crucial para poder rastrear las fuentes de error dentro de los paquetes y asegurarse de que estén a punto para el momento en que se dispongan datos reales.

Trabajo a futuro

Será necesario investigar cómo separar las fuentes de estructura tipo M en sus respectivas componentes y asignar correctamente las densidades de flujo a cada una.

De igual forma se espera probar la extracción de fuente mediante el método de mapa de curvatura descrito por Hancock et al. (2012) que decide cuántas componentes gaussianas se pueden ajustar a una isla. También se espera realizar una correspondencia entre los catálogos extraídos a 1.1, 1.4 y 2.0 mm a fin de poder analizar cuántas fuentes pueden ser detectadas en las tres bandas y cuales no. Con esto se harán las respectivas estimaciones de la Distribución Espectral de Energía (SEDs) y tasas de formación estelar (SFR)

También se espera trabajar con datos reales, para lo cual se empezará en la extracción de fuentes del campo Extended Groth Strip (EGS) de TolTEC y otros adquiridos durante las labores de comisionado. El EGS ha sido objeto de observaciones profundas en múltiples longitudes de onda que abarcan todo el rango desde radio hasta rayos X (Davis et al., 2007), lo que permite importantes estudios sobre la naturaleza de las DSFG y otras poblaciones de alto z . Las observaciones de TolTEC de 1.1, 1.4 y 2.0 mm en el campo EGS como muestra principal de estudio permitirán medir el número de cuentas de fuentes milimétricas y se centrarán en la población de galaxias infrarrojas luminosas (LIRGs, $L_{\text{IR}} = 10^{11} - 10^{12} L_{\odot}$), en asociación con el Telescopio Espacial James Webb (JWST). Dado que el EGS también es objetivo del censo profundo JWST/CEERS, la posición de las fuentes a 1.1 mm determinará la contraparte óptico/NIR/MIR en las imágenes del JWST (precisión astrométrica de $\sim 1''$). El uso de estos datos auxiliares permitirá realizar determinaciones de evolución morfológica a través del z , masa, oscurecimiento óptico, corrimiento al rojo fotométrico y excitación, además de estudiar las propiedades del ruido de confusión. Se espera trabajar en eso durante el proyecto de doctorado.

Ejemplos de archivos de configuración

A.1. Ejemplo de archivo de configuración YAML del módulo de simulación de TolTECA

```
1 # vim: et ts=2 sts=2 sw=2
2 ---
3 simu:
4   jobkey: 6h30min
5   instrument:
6     name: toltec
7     polarized: false
8   sources:
9     - data_exts:
10       - array_name: a1100
11         extname: 0_flux_a1100
12       - array_name: a1400
13         extname: 1_flux_a1400
14       - array_name: a2000
15         extname: 2_flux_a2000
16     filepath: rendered_squareCat_50mas.fits
17     type: image
18   - type: toltec_power_loading
19     atm_model_name: am_q25
20     atm_cache_dir: null
```

```

21  obs_params:
22    f_smp_mapping: 12Hz #12.2Hz
23    f_smp_probing: 122Hz #488Hz
24    t_exp: null
25  perf_params:
26    chunk_len: 10s
27    mapping_eval_interp_len: 0.05s
28    aplm_eval_interp_alt_step: 1.6 arcmin
29  mapping:
30    ref_frame: altaz
31    target: 269.355003876d 1.3549934376818d
32    t0: 2022-05-25T06:30:00 # 2022-05-25T06:18:00
33    type: raster
34    length: 30 arcmin
35    space: 0.5 arcmin
36    n_scans: 60
37    # The scan speed.
38    speed: 50 arcsec/s
39    t_turnaround: 5s
40    rot: 0 deg
41  plots:
42  - type: visibility
43  - type: mapping

```

A.2. Ejemplo de archivo de configuración YAML de Citlali

```

1  # vim: et ts=2 sts=2 sw=2
2  ---
3
4  reduce:
5    jobkey: all_redu_case0_neig_3
6    inputs:
7      - path: all_simu_case0
8        select: 'obsnum ==
9          [1,101,201,301,401,501,601,701,801,901,2,102]'
10   steps:
11   - name: citlali

```

```
11 version: "~=1.2" # requires compatibility with v1.0.0
12 config:
13   image_frame_params:
14     pixel_size: 0.75 arcsec
15   low_level:
16     kids:
17       fitter:
18         modelspec: gainlintrend
19         weight_window:
20           fwhm_Hz: 1.5e4
21           type: lorentz
22       solver:
23     runtime:
24       parallel_policy: omp # seq, omp, tbb
25       n_threads: 7 # number of threads to parallelize over
26       output_dir: null
27       use_subdir: true
28       reduction_type: science #beammap, pointing, etc
29     meta:
30       version: tolteca...
31     interface_sync_offset: #dict for each toltec[0..12]+hwp+lmt/
#match keys with file interface key
32     - toltec0: 0
33     - toltec1: 0
34     - toltec2: 0
35     - toltec3: 0
36     - toltec4: 0
37     - toltec5: 0
38     - toltec6: 0
39     - toltec7: 0
40     - toltec8: 0
41     - toltec9: 0
42     - toltec10: 0
43     - toltec11: 0
44     - toltec12: 0
45     - hwp: 0
46   timestream:
47     offset: 0.0
48   polarimetry:
```

```
49     enabled: false
50 output:
51     enabled: false
52     format: netcdf
53     chunk_type: rtc #ptc, both
54 chunking:
55     length_sec: 10.0
56 weighting:
57     type: full #approximate
58 despiking:
59     enabled: true
60     min_spike_sigma: 8.0
61     time_constant_sec: 0.015
62     window_size: 32
63 filter:
64     enabled: true
65     a_gibbs: 50.0
66     freq_high_Hz: 16.0
67     freq_low_Hz: 0.0
68     n_terms: 32
69 downsample:
70     enabled: true
71     factor: 1
72 kernel:
73     enabled: true
74     filepath: null
75     type: internal_gaussian #image, internal_airy
76     image_ext_name: signal
77 clean:
78     enabled: true
79     cut_std: 0.0
80     grouping: array_name
81     n_eig_to_cut: 3
82 mapmaking:
83     enabled: true
84     grouping: array_name #array, detector, polarization, nw
85     method: naive #jinc
86     pixel_axes: icrs #icrs, lb
87     pixel_size_arcsec: 0.75
```

```
88     crval1_J2000: 0 #absolute pointing_x
89     crval2_J2000: 0 #absolute pointing_y
90     crpix1: 0
91     crpix2: 0
92     cunit: MJy/Sr # "mJy/beam" # mJy/beam
93     x_size_pix: 0 #null or value
94     y_size_pix: 0 #null or value
95     beammap:
96         iter_tolerance: 0.2
97         iter_max: 1
98     source_fitting:
99         enabled: false
100        model: gaussian
101        bounding_box_arcsec: 10
102        initial_guess: brightest_pixel #table
103     coadd:
104         enabled: true
105         cov_cut: 0.9
106         noise_maps:
107             enabled: true
108             n_noise_maps: 100
109         filtering:
110             enabled: true
111             type: wiener_filter
112     wiener_filter:
113         gaussian_template: true
114         gaussian_template_fwhm_arcsec:
115             a1100: 5
116             a1400: 6.3
117             a2000: 9.5
118         lowpass_only: false
119         highpass_only: false
120         normalize_error: false
```


Palabras claves para el archivo de configuración del simulador de TolTEC

A continuación, se muestra en la Tabla B.1 las principales palabras claves utilizadas por el módulo de simulación de observaciones de TolTECA

key	Descripción
SimuConfig : Configuración del diccionario para el simulador	
jobkey	Identificador único del trabajo
instrument	Diccionario que contiene la configuración del instrumento
mapping	Configuración de la trayectoria de cartografiado del simulador
obs_params	Diccionario que contiene los parámetros de observación
sources	Lista que contiene las fuentes de entrada para la simulación
perf_params	Diccionario que contiene los parámetros relacionados con el rendimiento
plots	Lista que contiene la configuración para graficar
exports	Lista que contiene la configuración para exportar

Continúa en la página siguiente.

key	Descripción
plot_only	Realiza las gráficas definidas en <code>plots</code>
export_only	Exporta la configuración de la simulación como se define en <code>exports</code>
ObsParamsConfig: Parámetros relacionados a la observación	
t_exp	La duración de la observación a simular
f_smp_mapping	Frecuencia de muestreo para evaluar en los modelos de cartografiado
f_smp_probing	Frecuencia de muestreo para evaluar las señales del detector
PerfParamsConfig: Parámetros relacionados con la optimización del rendimiento	
chunk_len	Tamaño del <code>chunk</code> para dividir la simulación y reducir el consumo de memoria (valor por defecto: 10 segundos)
catalog_model_render_pixel_size	Tamaño del píxel para renderizar el modelo de catálogo fuente (valor por defecto: 0.5 arcsec)
mapping_eval_interp_len	Tamaño de interpolación para acelerar la evaluación del mapeo (valor por defecto: None)
mapping_erfa_interp_len	Tamaño de interpolación para acelerar la transformación de coordenadas de AltAZ a ICRS (valor por defecto: 300 segundos)
aplm_eval_interp_alt_step	Paso de interpolación para la altitud para acelerar la evaluación del modelo de carga del arreglo (array power loading model) (valor por defecto: 2.0 arcmin)
pre_eval_sky_bbox_padding_size	Tamaño del relleno para agregar al <code>sky bbox</code> para los cálculos previos a la evaluación (valor por defecto: 4.0 arcmin)

Continúa en la página siguiente.

key	Descripción
pre_eval_t_grid_size	Tamaño de la cuadrícula de tiempo usada para los cálculos previos a la evaluación (valor por defecto: 200).
anim_frame_rate	velocidad de fotogramas para graficar la animación (valor por defecto: 1.0 Hz)

MappingConfig: Sub esquemas para los items de configuración registrados

type	Clase de configuración
lmt_tcs	LmtTcsMappingConfig
raster	RasterMappingConfig
lissajous	LissajousMappingConfig
double_lissajous	Double_lissajousMappingConfig
rastajous	RastajousMappingConfig

LmtTcsMappingConfig: Configuración de la clase para el modelo de trayectoria creado desde el archivo TCS de GTM

key	Descripción
filepath	La ruta al archivo tel.nc

RasterMappingConfig: Configuración para el modo de cartografiado raster

target	El nombre del objetivo o las coordenadas en el cielo ^a
target_frame	El marco en qué se especifica la coordenada del objetivo ^a
ref_frame	El marco de referencia en el cual se hará el patrón de cartografiado ^a
t0	Tiempo en el que inicia la observación ^a
length	Longitud del barrido
space	Espacio entre barridos ^b
n_scans	Número de barridos ^b

Continúa en la página siguiente.

key	Descripción
rot	Ángulo de rotación con respecto a la dirección +longitud ^a
speed	Velocidad de barrido ^b
t_turnaround	Tiempo en dar la vuelta después de cada barrido ^b
LissajousMappingConfig: Configuración para el modo de cartografiado Lissajous ^b	
x_length	Tamaño del patrón en la dirección x
y_length	Tamaño del patrón en la dirección y
x_omega	Frecuencia angular en la dirección x
y_omega	Frecuencia angular en la dirección y
delta	Ángulo de fase de x con respecto a y
Double lissajousMappingConfig: Configuración para el modo de cartografiado doble Lissajous	
x_length_0	Tamaño del patrón principal en la dirección x
y_length_0	Tamaño del patrón principal en la dirección y
x_omega_0	Frecuencia angular en x para el patrón principal
y_omega_0	Frecuencia angular en y para el patrón principal
delta_0	Ángulo de fase de x con respecto a y para el patrón principal
x_length_1	Tamaño del patrón principal en la dirección x
y_length_1	Tamaño del patrón principal en la dirección y
x_omega_1	Frecuencia angular en x para el patrón menor
y_omega_1	Frecuencia angular en y para el patrón menor
delta_1	Ángulo de fase de x con respecto a y para el patrón menor
delta	Ángulo de fase del patrón principal y con respecto al patrón menor y
RastajousMappingConfig: Configuración para el modo de cartografiado Rastajous	

Continúa en la página siguiente.

key	Descripción
-----	-------------

Nota: Dado que este modo de cartografiado es una combinación entre un *raster* y un doble Lissajous, las palabras de configuración descritas para cada caso aplican de igual modo para éste.

InstrumentConfig: Sub esquemas para los items de configuración registrados

nombre	Clase de configuración
toltec	ToltecObsSimulatorConfig

ToltecObsSimulatorConfig: Configuración de la clase para la simulación de la observación con TolTEC

key	Descripción
-----	-------------

polarized	Booleano, que establece realizar o no una simulación de señal polarizada
hwp	Diccionario que contiene la configuración del HWP
calobj_index	La ruta del archivo del objeto de calibración
array_prop_table	Tabla de propiedades del arreglo a utilizar en lugar de la que se proporciona en el <code>calobj_index</code>

SourcesConfig: Sub esquemas para los items de configuración registrados

type	Clase de configuración
image	ImageSourceConfig
point_source_catalog	PointSourceCatalogSourceConfig
toltec_power_loading	ToltecPowerLoadingModelConfig

ImageSourceConfig: Configuración de la clase para la imagen del modelo fuente creado a partir del archivo de imagen FITS

filepath	La ruta al archivo de imagen FITS
data_exts	Las asignaciones de extensiones FITS a etiquetas de elementos de datos

Continúa en la página siguiente.

key	Descripción
PointSourceCatalogSourceConfig: Configuración de la clase para el catálogo de fuentes puntuales creado a partir de archivo de catálogo	
filepath	La ruta al archivo del catálogo
name_col	Columna para el nombre de la fuente
pos_cols	Columna para las coordenadas ['ra', 'dec']
data_cols	Las asignaciones para las columnas a las etiquetas de los datos
ToltecPowerLoadingModelConfig: Configuración de la clase para calcular el <i>detector power loading</i> de TolTEC	
atm_model_name	Modelo de atmósfera a usar
tel_surface_rms	Superficie del telescopio RMS
det_noise_factor	Factor de ruido del detector
atm_model_params	Ajustes del modelo de atmósfera de toast
atm_cache_dir	Directorio para almacenar los datos del modelo atmosférico

^a Aplica para los otros modos de cartografiado.

^b Aplica para el modo de cartografiado Rastajous.

Tabla B.1: Descripción de los parámetros de entrada usados en el archivo de configuración para la simulación de una observación utilizando el módulo de simulación de TolTECA.

Simulaciones adicionales

C.1. Tabla de errores en la fracción de posición para la simulación 1

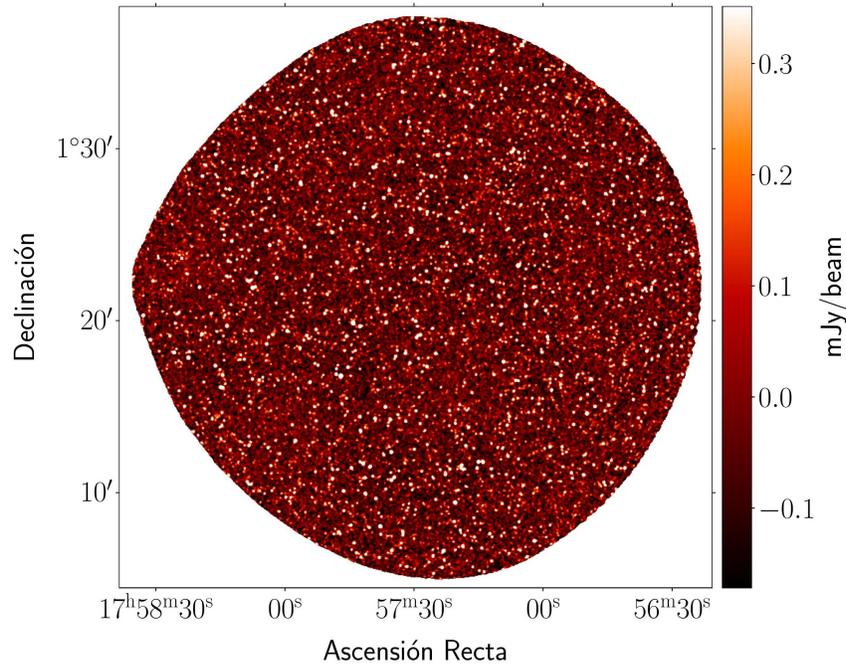
SNR	N	$\alpha_f = \frac{\alpha_{\text{cat}}}{\alpha_{\text{PyBDSF}}}$	$\delta_f = \frac{\delta_{\text{cat}}}{\delta_{\text{PyBDSF}}}$	$\epsilon = \sqrt{\alpha_d^2 + \delta_d^2}$ (arcs)
1.1 mm				
≥ 1	7405	$1.0 \pm 1.4 \times 10^{-9}$	$0.99 \pm 3.2 \times 10^{-7}$	$1.2_{-0.8}^{+1.9}$
≥ 3.5	4182	$1.0 \pm 1.4 \times 10^{-9}$	$0.999 \pm 3.2 \times 10^{-7}$	$0.89_{-0.53}^{+1.20}$
[1, 2)	1378	$1.0 \pm 2.8 \times 10^{-8}$	$1.0 \pm 6.7 \times 10^{-6}$	$2.2_{-1.2}^{+2.5}$
[2, 3)	1276	$1.0 \pm 1.9 \times 10^{-8}$	$1.0 \pm 4.4 \times 10^{-6}$	$1.8_{-1.0}^{+2.3}$
[3, 4)	1033	$1.0 \pm 1.5 \times 10^{-8}$	$0.99 \pm 3.5 \times 10^{-6}$	$1.4_{-0.7}^{+1.7}$
[4, 6)	1216	$1.0 \pm 9.8 \times 10^{-9}$	$0.99 \pm 2.3 \times 10^{-6}$	$1.2_{-0.6}^{+1.3}$
[6, 8)	756	$1.0 \pm 8.8 \times 10^{-9}$	$0.99 \pm 2.0 \times 10^{-6}$	$0.95_{-0.53}^{+1.10}$
[8, 10)	450	$1.0 \pm 8.8 \times 10^{-9}$	$0.99 \pm 1.9 \times 10^{-6}$	$0.87_{-0.51}^{+0.99}$
[10, 20)	850	$1.0 \pm 3.9 \times 10^{-9}$	$0.99 \pm 8.7 \times 10^{-7}$	$0.63_{-0.33}^{+0.79}$
≥ 20	446	$1.0 \pm 1.6 \times 10^{-9}$	$0.99 \pm 3.7 \times 10^{-7}$	$0.42_{-0.16}^{+0.67}$
1.4 mm				

SNR	N	$\alpha_f = \frac{\alpha_{\text{cat}}}{\alpha_{\text{PyBDSF}}}$	$\delta_f = \frac{\delta_{\text{cat}}}{\delta_{\text{PyBDSF}}}$	$\epsilon = \sqrt{\alpha_d^2 + \delta_d^2}$ (arcs)
≥ 1	5442	$1.0 \pm 2.7 \times 10^{-9}$	$0.99 \pm 6.0 \times 10^{-7}$	$1.9_{-1.2}^{+2.9}$
≥ 2.5	3685	$1.0 \pm 2.7 \times 10^{-9}$	$0.99 \pm 6.0 \times 10^{-7}$	$1.4_{-0.92}^{+2.2}$
≥ 3.5	2805	$1.0 \pm 2.7 \times 10^{-9}$	$0.99 \pm 6.1 \times 10^{-7}$	$1.2_{-0.8}^{+1.8}$
[1, 2)	1248	$1.0 \pm 3.8 \times 10^{-8}$	$1.0 \pm 9.1 \times 10^{-5}$	$3.1_{-1.8}^{+3.2}$
[2, 3)	967	$1.0 \pm 2.7 \times 10^{-8}$	$0.99 \pm 6.5 \times 10^{-6}$	$2.6_{-1.5}^{+3.3}$
[3, 4)	764	$1.0 \pm 2.2 \times 10^{-8}$	$1.0 \pm 5.2 \times 10^{-6}$	$2.0_{-1.2}^{+2.7}$
[4, 6)	887	$1.0 \pm 1.5 \times 10^{-8}$	$0.99 \pm 3.4 \times 10^{-6}$	$1.6_{-0.9}^{+2.1}$
[6, 8)	512	$1.0 \pm 1.4 \times 10^{-8}$	$0.99 \pm 3.1 \times 10^{-6}$	$1.3_{-0.7}^{+1.6}$
[8, 10)	308	$1.0 \pm 1.3 \times 10^{-8}$	$0.99 \pm 3.0 \times 10^{-6}$	$1.1_{-0.6}^{+1.6}$
≥ 10	755	$1.0 \pm 2.9 \times 10^{-9}$	$0.99 \pm 6.5 \times 10^{-7}$	$0.7_{-0.4}^{+1.2}$
2.0 mm				
≥ 1	2794	$1.0 \pm 9.5 \times 10^{-9}$	$0.99 \pm 2.2 \times 10^{-6}$	$3.6_{-2.3}^{+4.8}$
≥ 2	1997	$1.0 \pm 9.8 \times 10^{-9}$	$0.99 \pm 2.2 \times 10^{-6}$	$3.0_{-1.9}^{+4.3}$
[2, 3)	618	$1.0 \pm 5.2 \times 10^{-8}$	$1.0 \pm 1.2 \times 10^{-5}$	$4.9_{-2.9}^{+4.2}$
[3, 4)	452	$1.0 \pm 4.4 \times 10^{-8}$	$1.0 \pm 1.0 \times 10^{-5}$	$3.4_{-2.0}^{+3.9}$
[4, 6)	434	$1.0 \pm 3.2 \times 10^{-8}$	$0.99 \pm 7.5 \times 10^{-6}$	$2.6_{-1.4}^{+3.7}$
[6, 8)	199	$1.0 \pm 3.3 \times 10^{-8}$	$0.99 \pm 7.6 \times 10^{-6}$	$2.0_{-1.2}^{+2.1}$
≥ 8	292	$1.0 \pm 1.1 \times 10^{-8}$	$0.99 \pm 2.6 \times 10^{-6}$	$1.4_{-0.9}^{+2.6}$

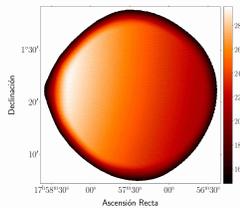
Tabla C.1: Errores en la fracción para las posiciones de entrada vs salida para las fuentes verdaderas de acuerdo al umbral en SNR establecido en la primera columna. La primera columna muestra el rango de señal a ruido, la segunda el número de fuentes dentro de ese rango, la tercera y cuarta columna muestran la fracción en las posiciones de ascensión recta (α) y declinación (δ) junto a su error y la última el error en la posición radial de la fuente junto a sus errores en un intervalo de confianza del 68 %.

C.2. Simulación 2: *raster* con atmósfera estática

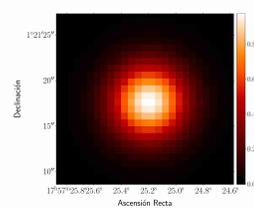
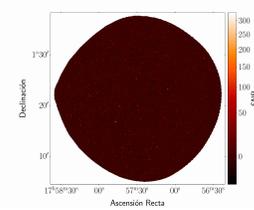
Resultados para una simulación de observación con cartografiado *raster* a una velocidad de telescopio de 50 arcs/s y un modelo de atmósfera estática. Los resultados de la simulación corresponden a ~ 8 horas de tiempo de integración. Se obtiene un mapa con un area de 0.22 grados cuadrados y un *rms* entre (0.0387–0.0548) mJy/haz.



(a) Mapa de señal



(b) Mapa de peso

(c) Mapa de *kernel* (Zoom)

(d) Mapa de señal a ruido

Figura C.1: Productos de ciencia para la simulación con un cartografiado *raster* y atmósfera estática en la banda de 1.1 mm.

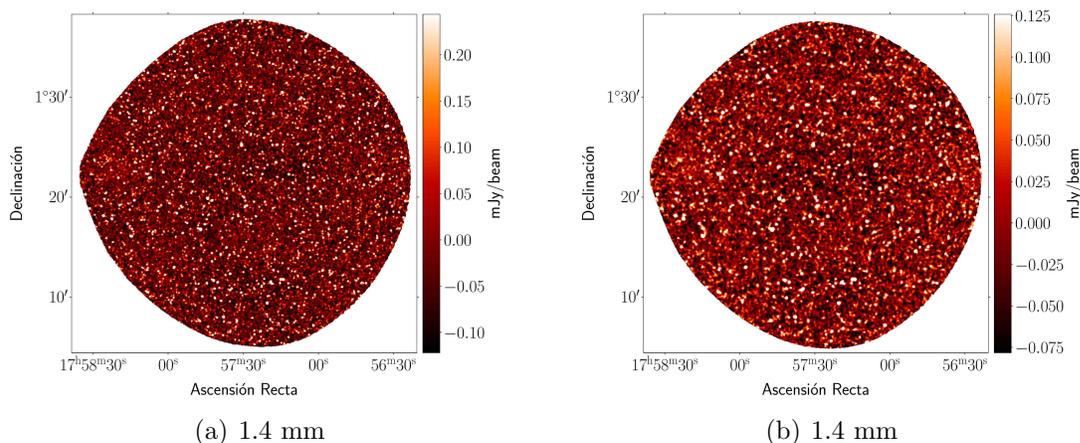


Figura C.2: Mapas de señal para la simulación con cartografiado *raster* y atmósfera estática.

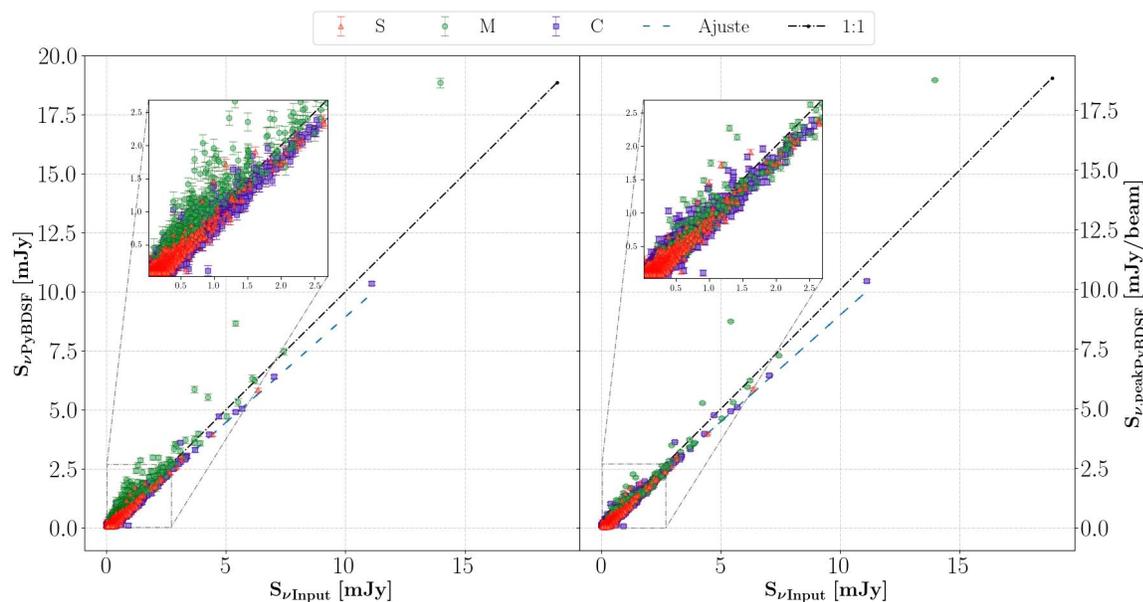
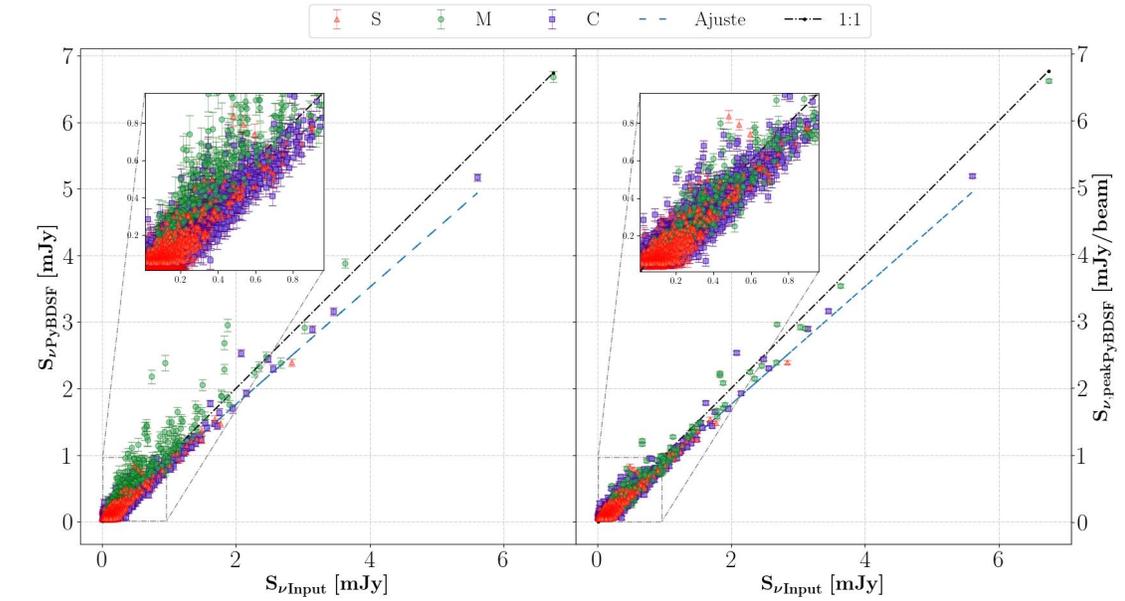
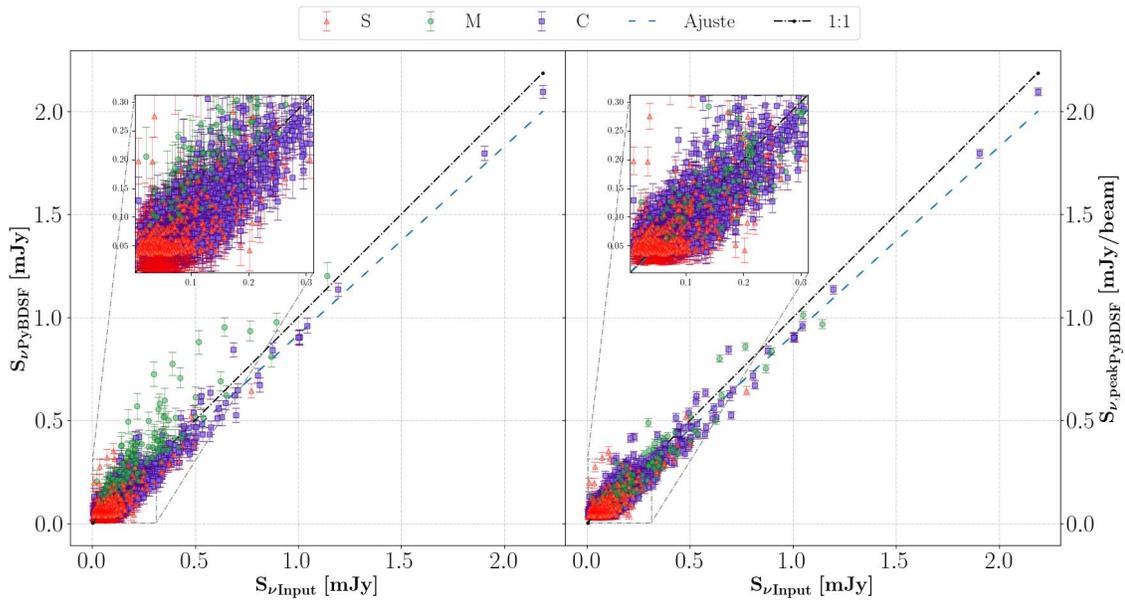


Figura C.3: Densidades de flujo de salida $S_{\nu,PyBDSF}$ respecto a las de salida $S_{\nu,Input}$ en la banda de **1.1 mm** para cada uno de los tipos de estructuras de fuentes detectados por PyBDSF. Los cuadrados, círculos y triángulos representan las estructuras tipo C, M y S, respectivamente. En la derecha se grafican los flujos integrados totales mientras que la izquierda se muestran los flujos picos. La línea negra indica la relación uno a uno y la línea azul el ajuste a los puntos.



(a) 1.4 mm



(b) 2.0 mm

Figura C.4: Densidades de flujo de salida $S_{\nu,PyBDSF}$ respecto a las de salida $S_{\nu,Input}$ en la banda de **1.4 mm** y **2.0 mm** para cada uno de los tipos de estructuras de fuentes detectados por PyBDSF. Los cuadrados, círculos y triángulos representan las estructuras tipo C, M y S, respectivamente. En la derecha se grafican los flujos integrados totales mientras que la izquierda se muestran los flujos picos. La línea negra indica la relación uno a uno y la línea azul el ajuste a los puntos.

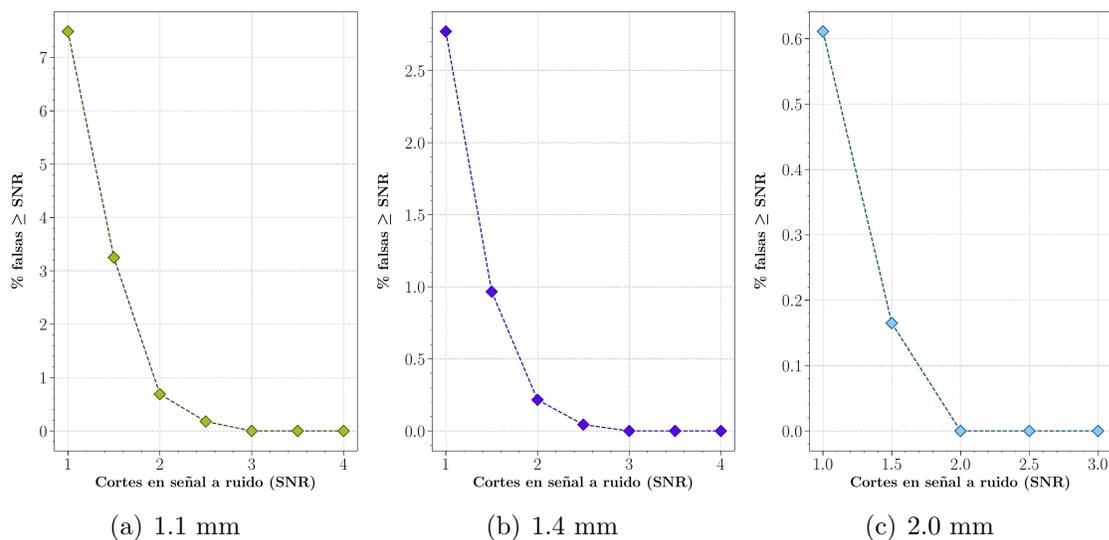


Figura C.5: Porcentaje de fuentes falsas detectadas sobre umbrales de señal a ruido para la simulación con cartografiado *raster* con atmósfera estática en las tres bandas.

SNR	Fuentes verdaderas \geq SNR	Fuentes falsas \geq SNR (%)
1.1 mm		
1.0	5931	480 (7.49)
1.5	5119	172 (3.25)
2.0	4175	29 (0.69)
2.5	3394	6 (0.18)
3.0	2805	0 (0.0)
3.5	2380	0 (0.0)
4.0	2050	0 (0.0)
1.4 mm		
1.0	4210	120 (2.77)
1.5	3485	34 (0.97)

SNR	Fuentes verdaderas	Fuentes falsas
	\geq SNR	\geq SNR (%)
2.0	2771	6 (0.22)
2.5	2291	1 (0.04)
3.0	1894	0 (0.0)
3.5	1614	0 (0.0)
4.0	1360	0 (0.0)
2.0 mm		
1.0	2114	13 (0.61)
1.5	1815	3 (0.17)
2.0	1413	0 (0.0)
2.5	1096	0 (0.0)
3.0	857	0 (0.0)

Tabla C.2: Detecciones verdaderas y falsas, con la simulación usando un patrón de cartografiado *raster* y atmósfera estática, por cortes en señal a ruido en las tres bandas. La primera columna muestra los cortes en señal a ruido, la segunda y tercera el número de fuentes verdaderas y falsas por cada corte, así en paréntesis se especifica el porcentaje de detecciones falsas.

SNR	S		C		M	
	N_s	ϵ_S [arcs]	N_C	ϵ_C [arcs]	N_M	ϵ_M [arcs]
1.1 mm						
≥ 1	1381	$1.5^{+2.2}_{-1.0}$	4067	$1.2^{+1.7}_{-0.7}$	483	$1.8^{+1.5}_{-1.0}$
≥ 1.5	910	$1.2^{+1.7}_{-0.7}$	3726	$1.1^{+1.5}_{-0.7}$	483	$1.8^{+1.5}_{-1.0}$
≥ 2	637	$1.0^{+1.3}_{-0.6}$	3065	$1.0^{+1.3}_{-0.6}$	473	$1.8^{+1.5}_{-1.0}$
≥ 3	382	$0.72^{+0.91}_{-0.41}$	1980	$0.77^{+0.89}_{-0.43}$	443	$1.8^{+1.5}_{-1.0}$

SNR	S		C		M	
	N_s	ϵ_C	N_C	ϵ_C	N_M	ϵ_M
[2, 3)	255	$1.6^{+1.5}_{-0.8}$	1085	$1.5^{+1.8}_{-0.8}$	30	$1.4^{+3.6}_{-0.6}$
[3, 4)	124	$1.2^{+1.2}_{-0.7}$	583	$1.3^{+1.2}_{-0.7}$	48	$3.0^{+1.2}_{-1.1}$
≥ 4	258	$0.58^{+0.56}_{-0.31}$	1397	$0.63^{+0.67}_{-0.34}$	395	$1.7^{+1.3}_{-0.9}$
1.4 mm						
≥ 1	1003	$2.2^{+3.5}_{-1.4}$	2808	$1.5^{+2.4}_{-0.9}$	399	$2.6^{+2.2}_{-1.4}$
≥ 1.5	575	$1.6^{+2.3}_{-1.0}$	2511	$1.4^{+2.1}_{-0.8}$	399	$2.6^{+2.2}_{-1.4}$
≥ 3	218	$1.0^{+1.0}_{-0.6}$	1324	$1.0^{+1.22}_{-0.6}$	352	$2.5^{+2.2}_{-1.4}$
[2, 3)	174	$2.2^{+2.2}_{-1.4}$	667	$1.9^{+2.5}_{-1.0}$	36	$4.1^{+2.4}_{-1.9}$
[3, 4)	79	$1.6^{+1.4}_{-0.9}$	426	$1.5^{+1.6}_{-0.8}$	29	$3.1^{+2.6}_{-1.4}$
≥ 4	139	$0.73^{+0.92}_{-0.37}$	898	$0.84^{+0.84}_{-0.45}$	323	$2.4^{+1.9}_{-1.3}$
2.0 mm						
≥ 1	420	$4.0^{+5.2}_{-2.5}$	1549	$3.0^{+4.8}_{-1.8}$	145	$4.4^{+4.1}_{-2.0}$
≥ 1.5	245	$3.2^{+4.2}_{-2.0}$	1425	$2.8^{+4.6}_{-1.7}$	145	$4.4^{+4.1}_{-2.0}$
≥ 3	79	$2.1^{+3.3}_{-1.2}$	655	$1.9^{+2.6}_{-1.1}$	123	$4.3^{+3.6}_{-1.7}$
[2, 3)	70	$2.9^{+2.8}_{-1.6}$	467	$3.8^{+4.5}_{-2.0}$	19	$6.5^{+3.5}_{-4.3}$
[3, 4)	35	$2.7^{+2.9}_{-1.6}$	241	$2.9^{+3.7}_{-1.7}$	20	$6.2^{+1.2}_{-2.4}$
≥ 4	44	$1.4^{+2.2}_{-0.6}$	414	$1.5^{+2.0}_{-0.8}$	103	$4.1^{+3.8}_{-1.8}$

Tabla C.3: Estimaciones del error de posición por cortes de señal a ruido para las tres bandas en la simulación 2, donde 'SNR' indica el umbral o rango de señal a ruido considerado, las letras S, C y M indican el tipo de estructura de fuente identificada, 'N' el número de fuentes y ϵ la estimación del error junto a sus incertidumbres asociadas a un intervalo de confianza del 68%.

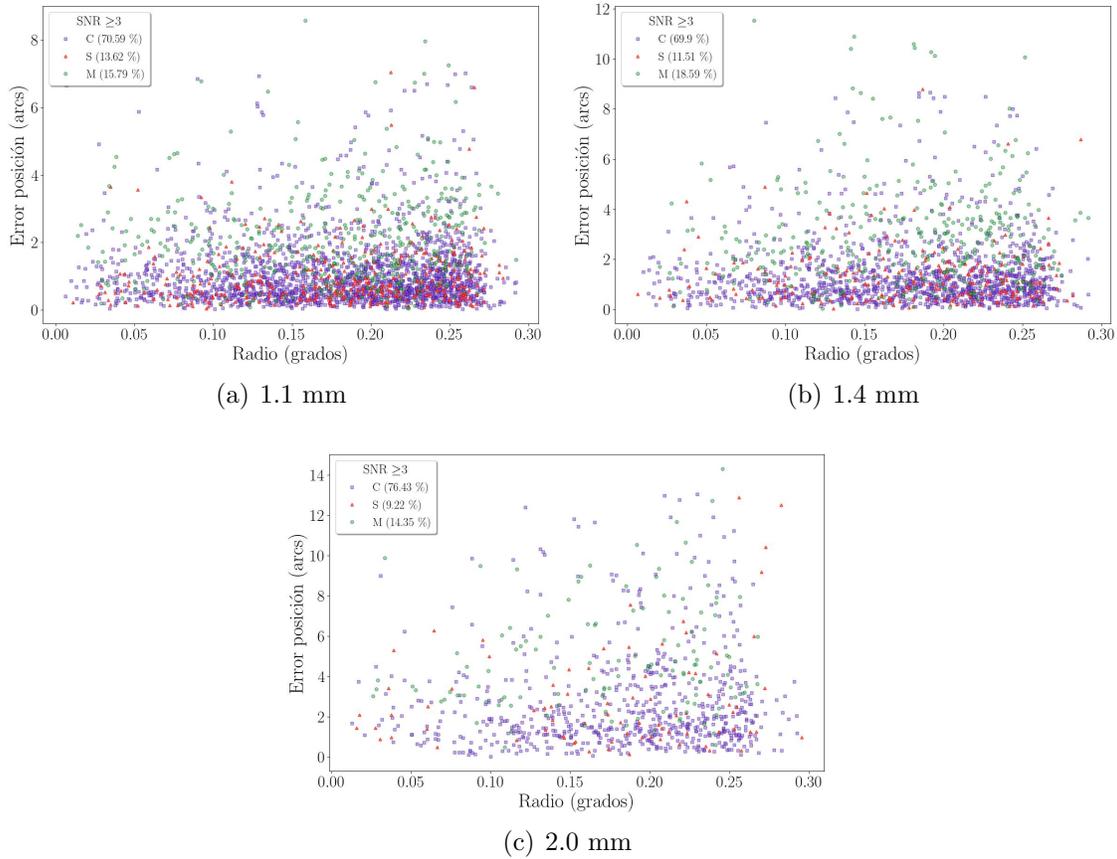


Figura C.6: Errores de posición en radios concéntricos desde el centro del mapa de la simulación 2 para las tres bandas y cortes en $\text{SNR} \geq 3$. Los cuadrados, triángulos y círculos representan las fuentes identificadas como tipo 'C', 'S' y 'M', respectivamente.

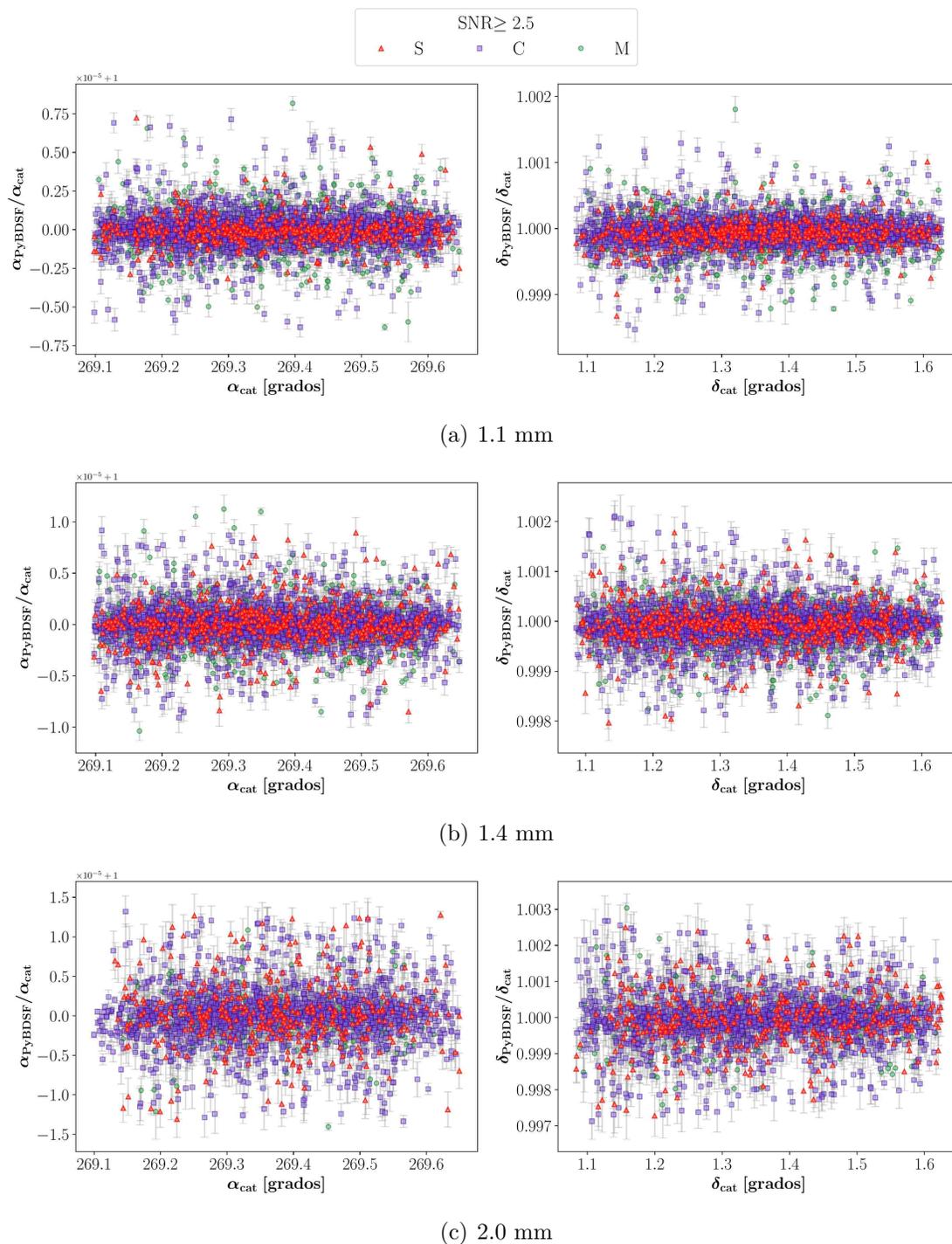
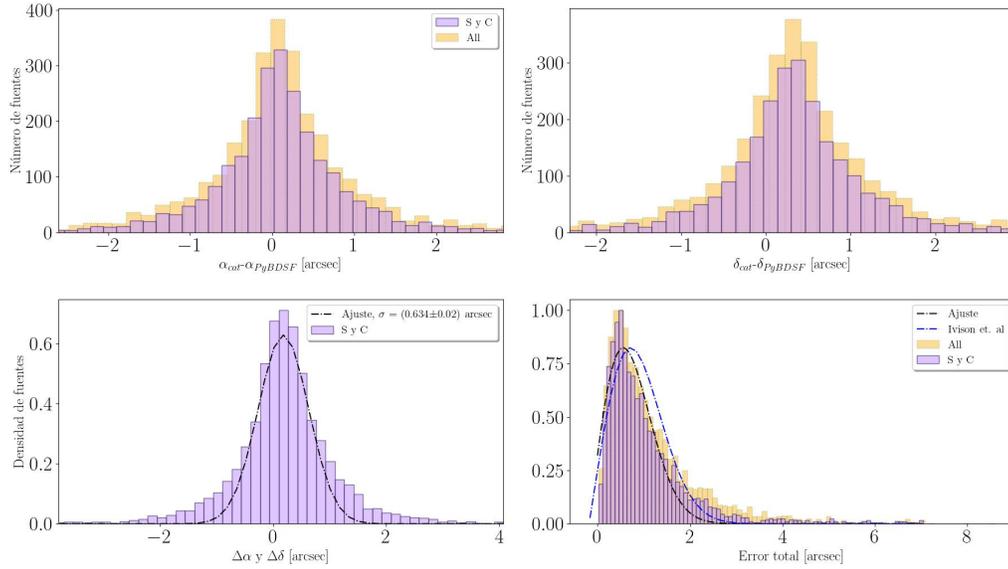
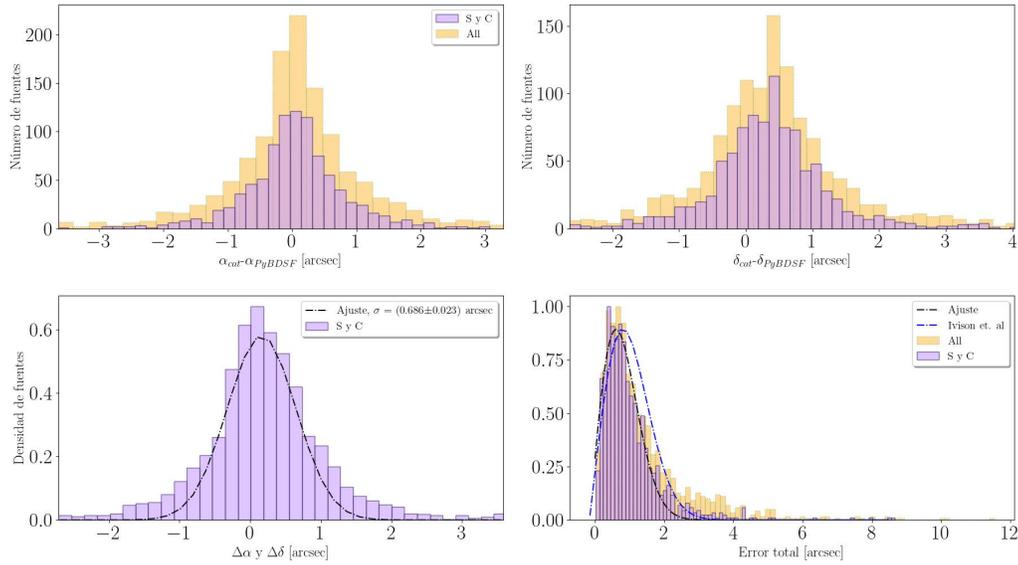


Figura C.7: Cocientes de las posiciones en α , δ respecto a las de entrada para cada una de las tres bandas en la simulación 2. Los puntos rojos, violetas y verdes corresponden a las fuentes tipo 'S', 'C' y 'M'.

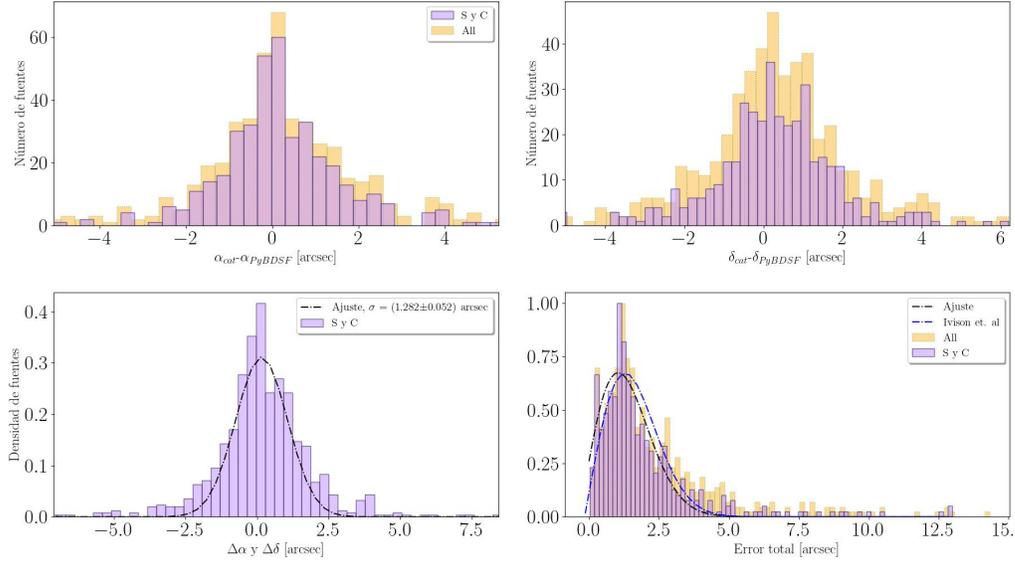


(a) 1.1 mm



(b) 1.4 mm

Figura C.8: Histogramas de errores de posición para las fuentes detectadas en las bandas de 1.1 mm y 1.4 mm para la simulación 2. La primera fila presenta el error en α y δ . La gráfica inferior izquierda presenta el resultado del histograma conjunto de los errores en α y δ y la gráfica inferior derecha, el error de posición radial, donde la línea discontinua negra muestra el ajuste de la ecuación (3.4.7) dejando el tamaño del haz como parámetro libre y la línea azul el ajuste fijando el tamaño del haz al valor sintético.



(a) 2.0 mm

Figura C.9: Histogramas de errores de posición para las fuentes detectadas en la banda de 2.0 mm para la simulación 2. La primera fila presenta el error en α y δ . La gráfica inferior izquierda presenta el resultado del histograma conjunto de los errores en α y δ y la gráfica inferior derecha, el error de posición radial, donde la línea discontinua negra muestra el ajuste de la ecuación (3.4.7) dejando el tamaño del haz como parámetro libre y la línea azul el ajuste fijando el tamaño del haz al valor sintético.

$S_{\nu, \text{umbral}}$ (mJy)	N_{cat}	N_{PyBDSF}	N_{asig}
	$(S_{\nu} \geq S_{\nu, \text{umbral}})$		
banda 1.1 mm			
0.04	13279	5865	5931
0.08	9787	5533	5884
0.12	7436	5052	4905
0.16	5697	4401	3953
0.2	4453	3729	3168
0.24	3559	3134	2603
0.28	2922	2632	2203

$S_{\nu,\text{umbral}}$ (mJy)	N_{cat}	N_{PyBDSF}	N_{asig}
$(S_{\nu} \geq S_{\nu,\text{umbral}})$			
banda 1.1 mm			
0.32	2448	2247	1850
0.36	2105	1951	1611
banda 1.4 mm			
0.03	11481	4115	4210
0.06	7512	3835	4047
0.09	5089	3353	3097
0.12	3639	2795	2440
0.15	2760	2300	1952
0.18	2169	1888	1595
0.21	1727	1544	1314
0.24	1447	1320	1102
0.27	1203	1111	923
banda 2.0 mm			
0.02	7636	2052	2114
0.04	3783	1799	2066
0.06	2306	1480	1601
0.08	1527	1177	1174
0.1	1088	895	867
0.12	829	728	686
0.14	646	580	534
0.16	520	475	435

Tabla C.4: Número de fuentes recuperadas sobre umbrales de densidades de flujo de entrada en la simulación 2. La primera columna muestra el corte por densidades de flujo de entrada S_{ν} (mJy). La segunda columna, el número de fuentes de entrada sobre ese umbral. La tercera columna, el número de fuentes medidas a la salida sobre ese umbral y la última columna, el número de fuentes que fueron asignadas sobre ese umbral.

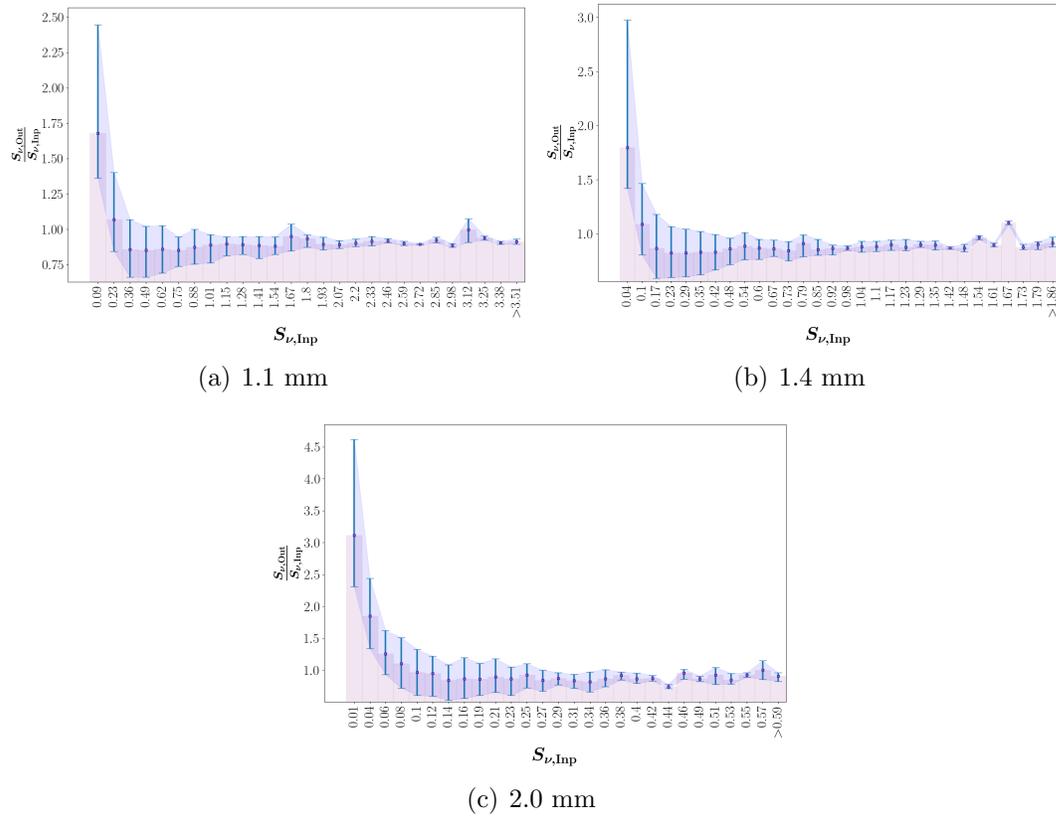


Figura C.10: Cociente medio ponderado de densidades de flujo por rangos de densidades de flujos de entrada en la simulación 2. Las barras de error indican el intervalo de confianza del 68 %.

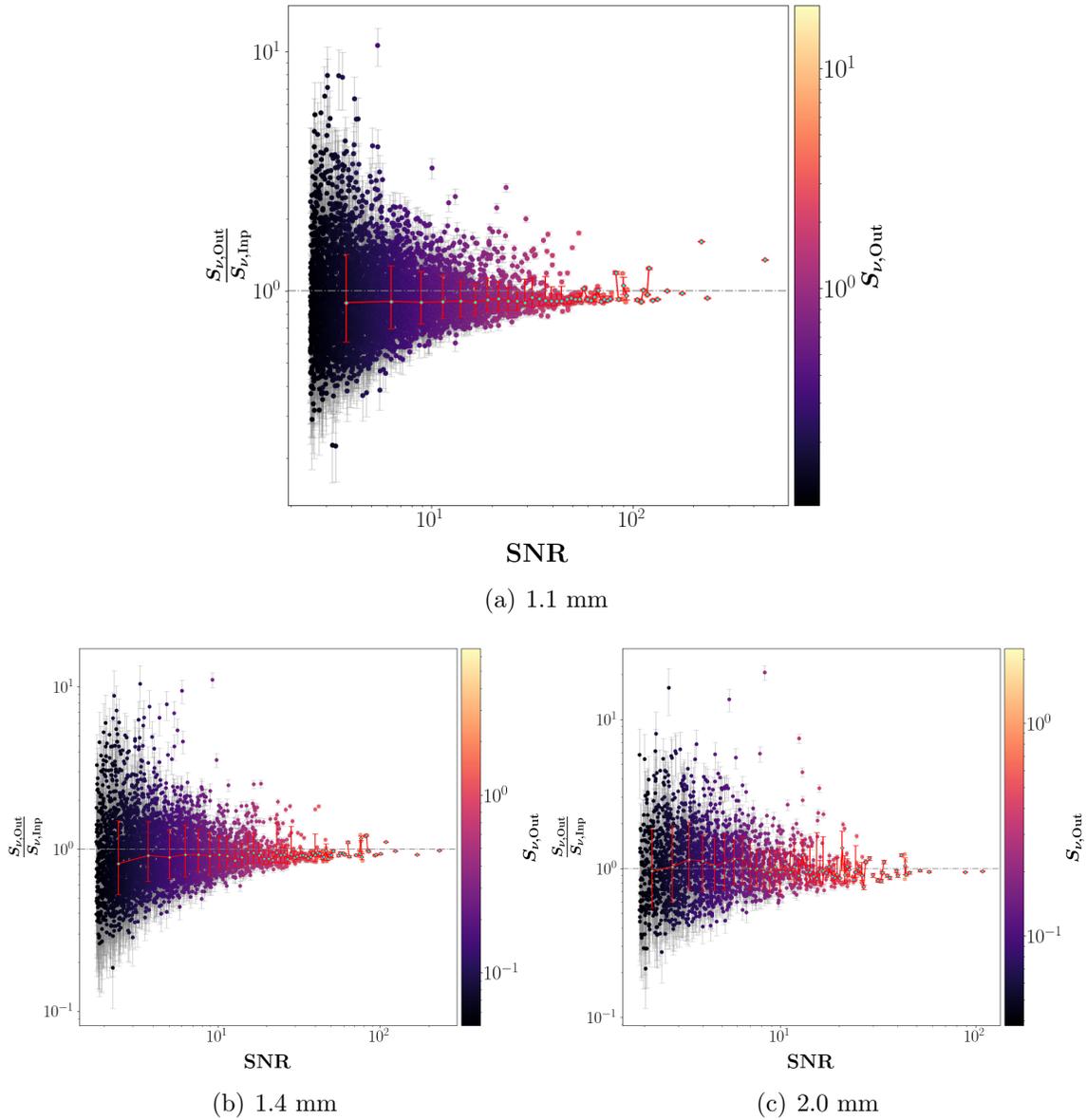
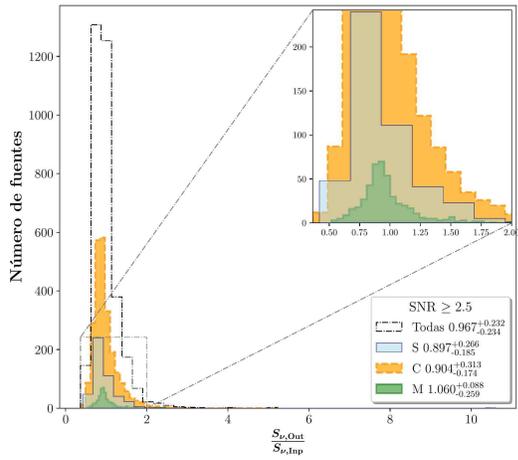
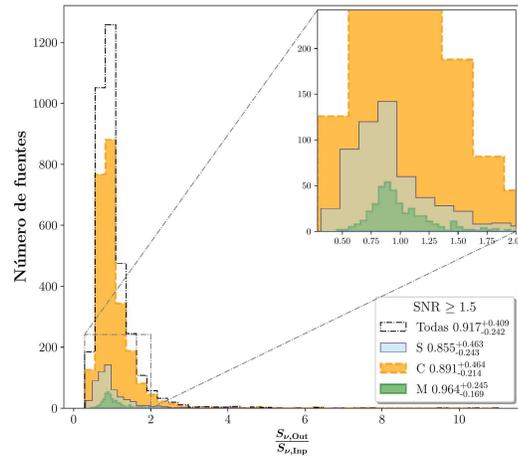


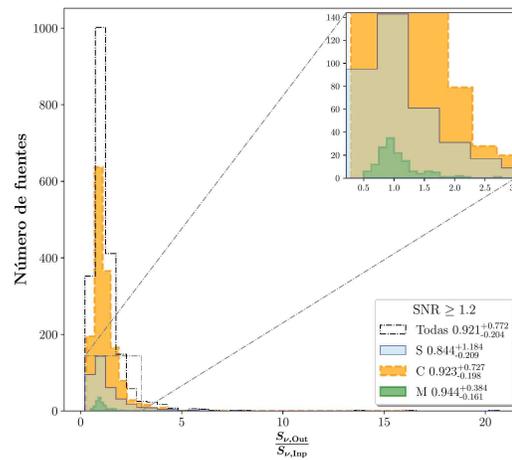
Figura C.11: Factor de *boosting* estimado de la simulación 2, que se calcula como la razón entre la densidad de flujo medida por PyBDSF $S_{\nu,Out}$ y el flujo de entrada del catálogo de Nava-Moreno et al. (2019). La línea roja representa el promedio ponderado del factor de *boosting* para cada intervalo de señal a ruido medido (*bin*). Nótese que el valor del cociente inferior a 1 se debe a la pérdida de flujo experimentada en las etapas de simulación de la observación. Las barras de error indican el intervalo de confianza del 68 %



(a) 1.1 mm



(b) 1.4 mm



(c) 2.0 mm

Figura C.12: Histogramas del cociente de densidades de flujo de salida y entrada para las fuentes detectadas. Se muestran en azul, naranja y verde los histogramas para las fuentes tipo 'S', 'C' y 'M' y en línea punteada el histograma conjunto.

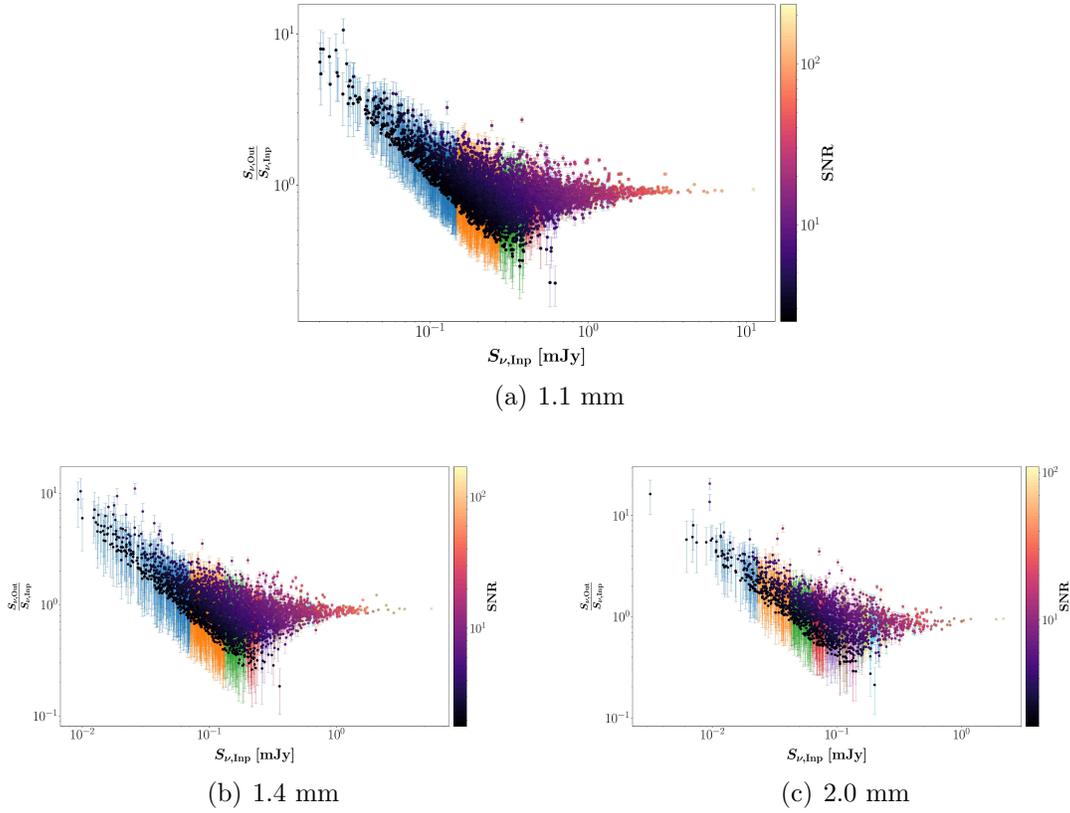


Figura C.13: Cociente de densidad de flujo de salida respecto al de entrada ($S_{\nu,\text{Out}}/S_{\nu,\text{Inp}}$) en escala logarítmica para la simulación 2. Las barras de error indican un intervalo de confianza del 68 %, mientras que los colores muestran la señal a ruido local en cada medida.

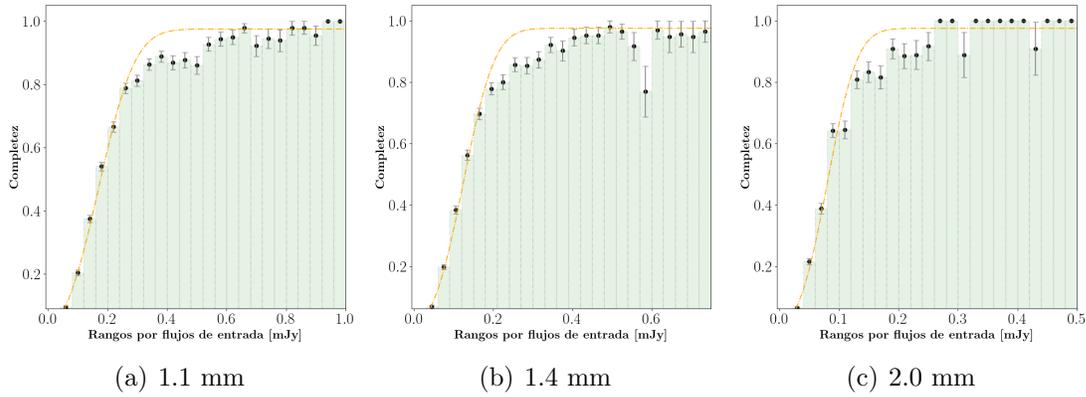
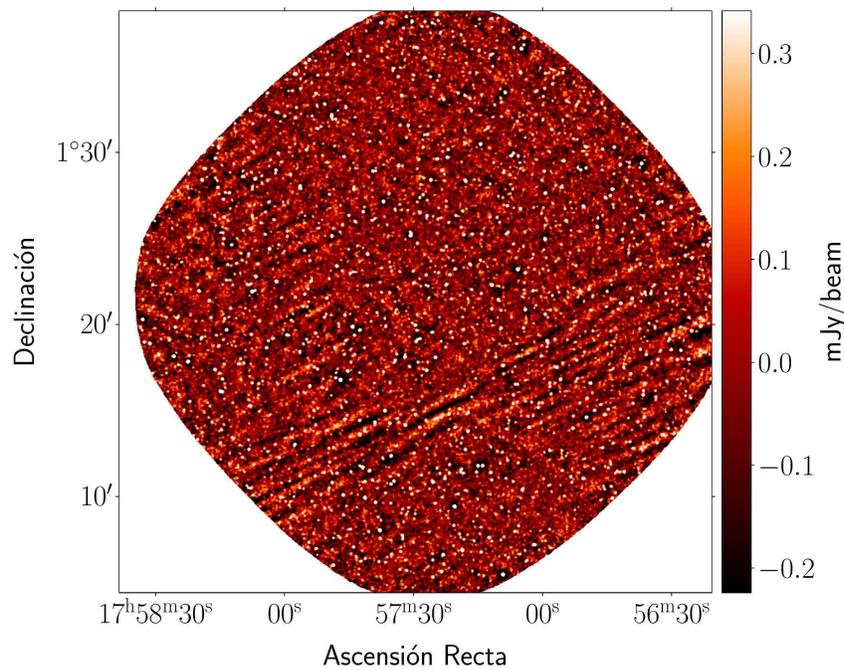


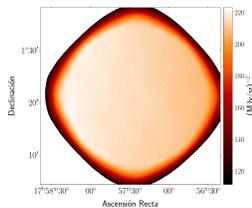
Figura C.14: Completez obtenida para la simulación 2. Los puntos en la gráfica junto con las barras de error estimadas muestran las fuentes que han sido identificadas y asociadas con las fuentes de entrada en cada *bin* representado en el gráfico. Además, se ha trazado una línea discontinua de color naranja que representa la función de mejor ajuste de los datos con la ecuación (4.5.1).

C.3. Simulación 3: velocidad de barrido 3 arcmin/s

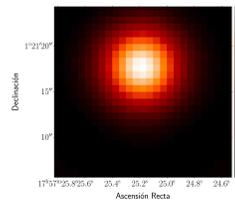
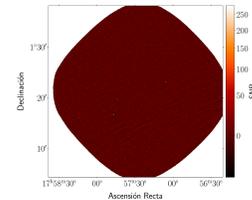
Resultados para la simulación de observación con cartografiado raster a una velocidad de telescopio de 3 arcmin/s y un modelo de atmósfera variable. Los resultados de la simulación corresponden a ~ 19 horas de tiempo de integración. Se obtiene un mapa *rms* entre (0.0446, 0.0630) mJy/haz en la banda de 1.1 mm.



(a) Mapa de señal



(b) Mapa de peso

(c) Mapa de *kernel* (Zoom)

(d) Mapa de señal a ruido

Figura C.15: Productos de ciencia para la simulación con un cartografiado *raster* y velocidad de telescopio 3 arcmin/s en la banda de 1.1 mm y considerando un modelo de atmósfera variable.

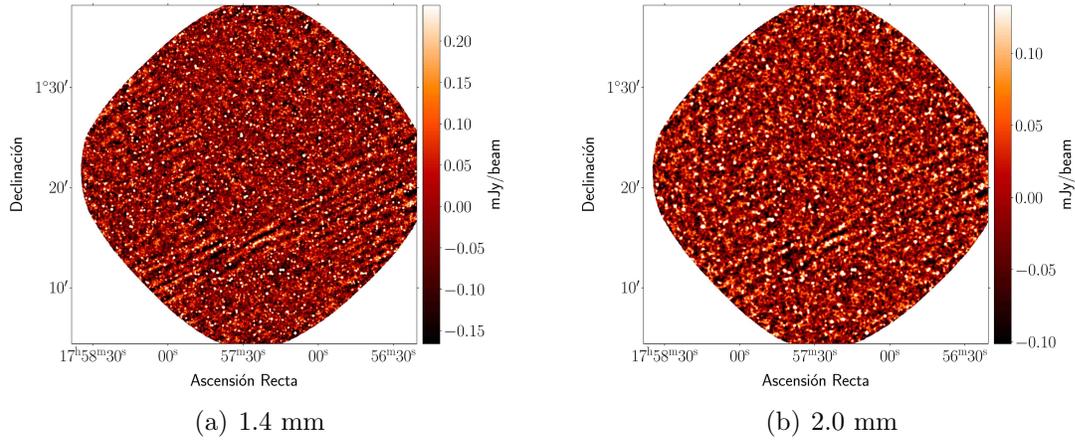


Figura C.16: Mapas de señal para la simulación con cartografiado *raster* y velocidad de barrido de telescopio 3 arcmin/s.

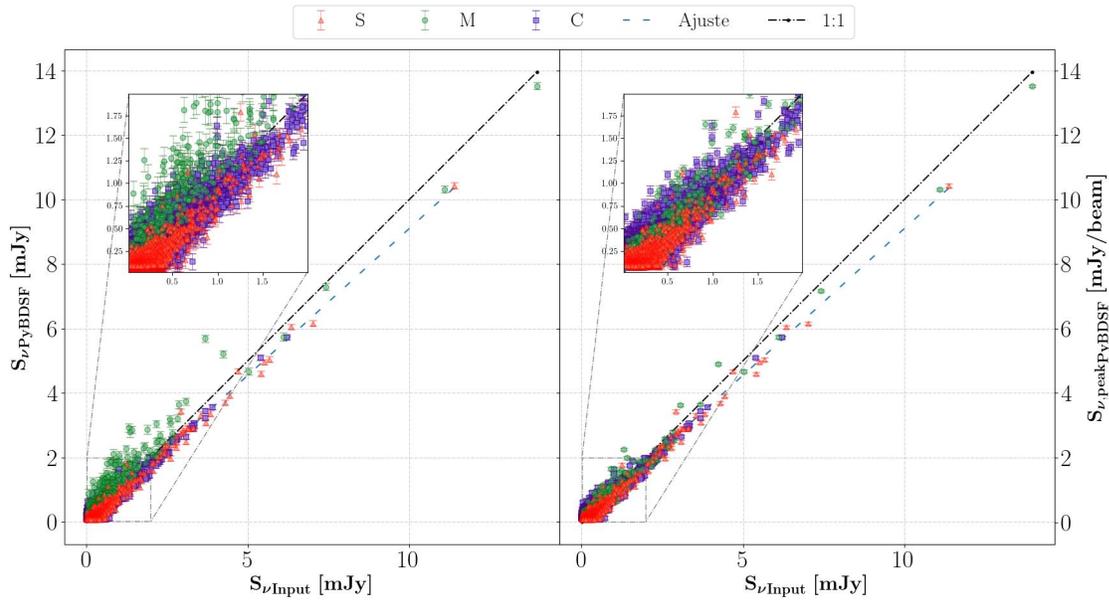
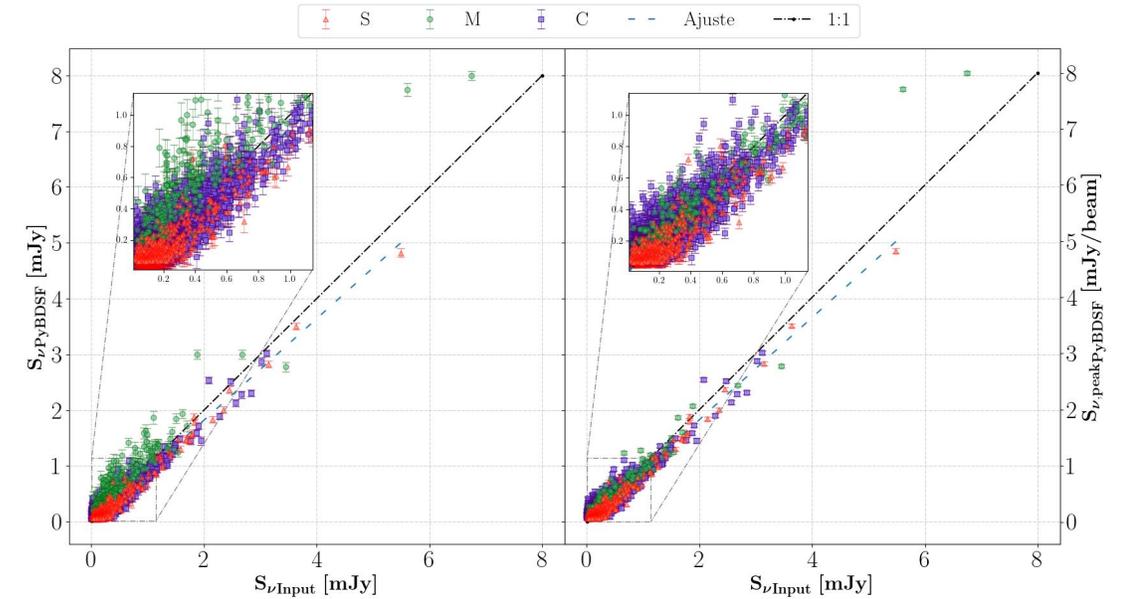
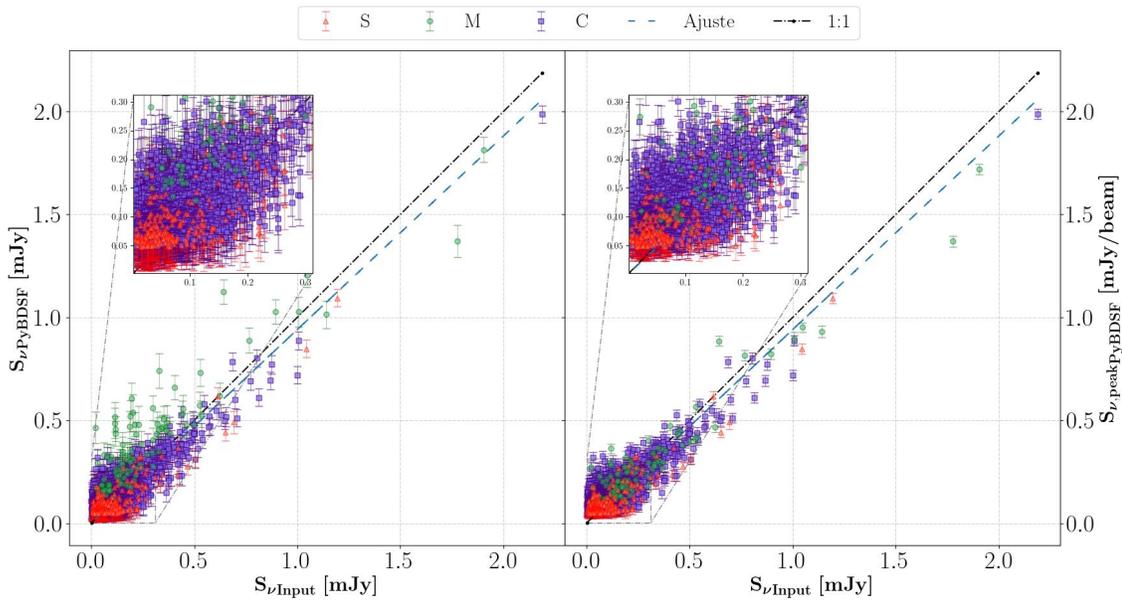


Figura C.17: Densidades de flujo de salida $S_{\nu,PyBDSF}$ respecto a las de salida $S_{\nu,Input}$ en la banda de **1.1 mm** para cada uno de los tipos de estructuras de fuentes detectados por PyBDSF. Los cuadrados, círculos y triángulos representan las estructuras tipo C, M y S, respectivamente. En la derecha se grafican los flujos integrados totales mientras que la izquierda se muestran los flujos picos. La línea negra indica la relación uno a uno y la línea azul el ajuste a los puntos.



(a) 1.4 mm



(b) 2.0 mm

Figura C.18: Densidades de flujo de salida $S_{\nu,PyBDSF}$ respecto a las de salida $S_{\nu,Input}$ en la banda de **1.4 mm** y **2.0 mm** para cada uno de los tipos de estructuras de fuentes detectados por PyBDSF. Los cuadrados, círculos y triángulos representan las estructuras tipo C, M y S, respectivamente. En la derecha se grafican los flujos integrados totales mientras que la izquierda se muestran los flujos picos. La línea negra indica la relación uno a uno y la línea azul el ajuste a los puntos.

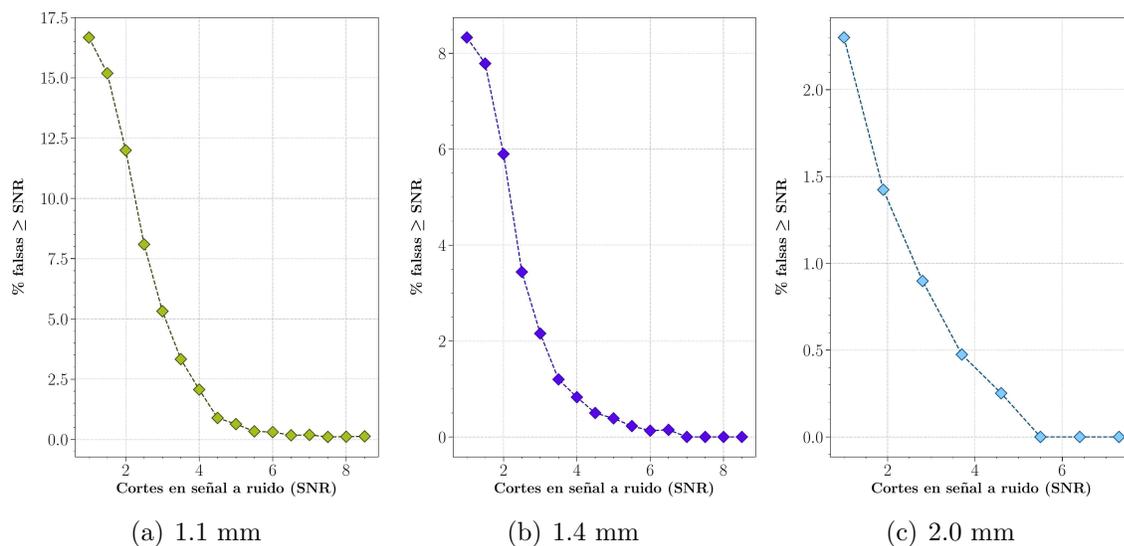


Figura C.19: Porcentaje de fuentes falsas detectadas sobre umbrales de señal a ruido para la simulación con un patrón *raster* y con velocidad de barrido del telescopio 3 arcmin/s.

SNR	Fuentes verdaderas \geq SNR	Fuentes falsas \geq SNR (%)
1.1 mm		
1.0	6147	1230 (16.67)
1.5	5696	1020 (15.19)
2.0	4909	669 (11.99)
2.5	4078	359 (8.09)
3.0	3365	189 (5.32)
3.5	2846	98 (3.33)
4.0	2368	50 (2.07)
4.5	2004	18 (0.89)
5.0	1719	11 (0.64)

SNR	Fuentes verdaderas \geq SNR	Fuentes falsas \geq SNR (%)
5.5	1493	5 (0.33)
6.0	1308	4 (0.3)
6.5	1165	2 (0.17)
7.0	1060	2 (0.19)
1.4 mm		
1.0	4017	365 (8.33)
1.5	3778	319 (7.79)
2.0	3222	202 (5.9)
2.5	2610	93 (3.44)
3.0	2131	47 (2.16)
3.5	1729	21 (1.2)
4.0	1435	12 (0.83)
4.5	1196	6 (0.5)
5.0	1029	4 (0.39)
5.5	880	2 (0.23)
6.0	767	1 (0.13)
6.5	675	1 (0.15)
7.0	590	0 (0.0)
2.0 mm		
1	2293	54 (2.3)
2	1633	23 (1.39)
3	889	7 (0.78)

SNR	Fuentes verdaderas	Fuentes falsas
	\geq SNR	\geq SNR (%)
4	529	2 (0.38)
5	320	0 (0.0)
6	207	0 (0.0)
6.5	165	0 (0.0)
7.0	141	0 (0.0)

Tabla C.5: Detecciones verdaderas y falsas, con la simulación usando un patrón *raster* y velocidad de telescopio 3 arcmin/s, por cortes en señal a ruido en las tres bandas. La primera columna muestra los cortes en señal a ruido, la segunda y tercera el número de fuentes verdaderas y falsas por cada corte, así en paréntesis se especifica el porcentaje de detecciones falsas.

SNR	S		C		M	
	N_s	ϵ_C	N_C	ϵ_C	N_M	ϵ_M
1.1 mm						
≥ 3.5	316	$0.8^{+1.6}_{-0.4}$	2219	$0.9^{+1.7}_{-0.5}$	311	$1.8^{+1.4}_{-1.1}$
[2, 3)	153	$2.2^{+2.2}_{-1.3}$	1363	$2.1^{+2.6}_{-1.2}$	28	$1.1^{+1.8}_{-0.4}$
[3, 4)	101	$2.1^{+2.1}_{-1.2}$	859	$1.7^{+2.4}_{-1.0}$	37	$1.5^{+2.2}_{-0.79}$
[4, 6)	102	$1.3^{+1.1}_{-0.71}$	885	$1.4^{+1.9}_{-0.8}$	73	$1.9^{+2.1}_{-0.93}$
[6, 8)	46	$0.8^{+1.5}_{-0.4}$	317	$0.9^{+1.0}_{-0.4}$	61	$2.1^{+1.3}_{-1.1}$
≥ 8	121	$0.41^{+0.30}_{-0.19}$	602	$0.53^{+0.48}_{-0.27}$	161	$1.6^{+1.2}_{-1.0}$
1.4 mm						
≥ 1	489	$2.0^{+3.3}_{-1.4}$	3253	$2.0^{+3.0}_{-1.2}$	275	$2.7^{+2.1}_{-1.3}$
≥ 2.5	213	$1.1^{+2.2}_{-0.7}$	2131	$1.6^{+2.5}_{-1.0}$	266	$2.7^{+2.1}_{-1.3}$
≥ 3.5	149	$0.7^{+1.1}_{-0.4}$	1346	$1.3^{+1.8}_{-0.8}$	234	$2.7^{+2.1}_{-1.3}$

SNR	S		C		M	
	N_s	ϵ_S [arcs]	N_C	ϵ_C [arcs]	N_M	ϵ_M [arcs]
[2, 3]	112	$3.0^{+3.4}_{-1.8}$	949	$2.6^{+3.0}_{-1.5}$	30	$2.7^{+4.0}_{-1.8}$
[3, 4]	44	$1.6^{+2.5}_{-0.6}$	630	$2.2^{+2.9}_{-1.3}$	22	$3.1^{+1.5}_{-1.3}$
[4, 6]	44	$1.4^{+1.6}_{-0.7}$	548	$1.6^{+1.9}_{-0.9}$	76	$3.0^{+3.0}_{-1.3}$
[6, 8]	31	$0.54^{+0.53}_{-0.18}$	250	$1.1^{+1.2}_{-0.7}$	45	$3.0^{+1.5}_{-1.1}$
[8, 10]	11	$0.66^{+0.59}_{-0.30}$	113	$0.78^{+0.73}_{-0.46}$	35	$2.4^{+1.7}_{-1.0}$
≥ 10	45	$0.48^{+0.37}_{-0.26}$	171	$0.54^{+0.53}_{-0.23}$	66	$1.8^{+1.8}_{-0.7}$
2.0 mm						
≥ 1	276	$5.5^{+5.7}_{-3.7}$	1913	$4.0^{+5.0}_{-2.5}$	104	$5.2^{+5.3}_{-2.5}$
≥ 2	95	$3.1^{+6.5}_{-2.0}$	1435	$3.6^{+4.7}_{-2.2}$	103	$5.2^{+5.1}_{-2.5}$
≥ 3.5	26	$1.3^{+1.0}_{-0.9}$	359	$1.8^{+2.1}_{-1.1}$	66	$4.2^{+3.5}_{-1.9}$
[2, 3]	52	$5.0^{+5.6}_{-2.7}$	677	$4.9^{+4.6}_{-2.8}$	15	$7.0^{+7.4}_{-2.2}$
[3, 4]	15	$2.2^{+5.3}_{-0.8}$	326	$4.0^{+4.7}_{-2.3}$	19	$6.5^{+4.7}_{-3.7}$
[4, 6]	14	$1.6^{+1.9}_{-0.5}$	281	$2.5^{+3.0}_{-1.3}$	27	$5.9^{+4.5}_{-3.1}$
[6, 8]	8	$1.4^{+0.8}_{-0.8}$	87	$1.6^{+1.3}_{-0.8}$	19	$4.5^{+3.1}_{-1.2}$
≥ 8	6	$0.37^{+0.76}_{-0.26}$	64	$0.89^{+0.74}_{-0.50}$	23	$2.8^{+2.2}_{-1.3}$

Tabla C.6: Estimaciones del error de posición por cortes de señal a ruido para las tres bandas en la simulación 3, donde ‘SNR’ indica el umbral o rango de señal a ruido considerado, las letras S, C y M indican el tipo de estructura de fuente identificada, ‘N’ el número de fuentes y ϵ la estimación del error junto a sus incertidumbres asociadas a un intervalo de confianza del 68 %.

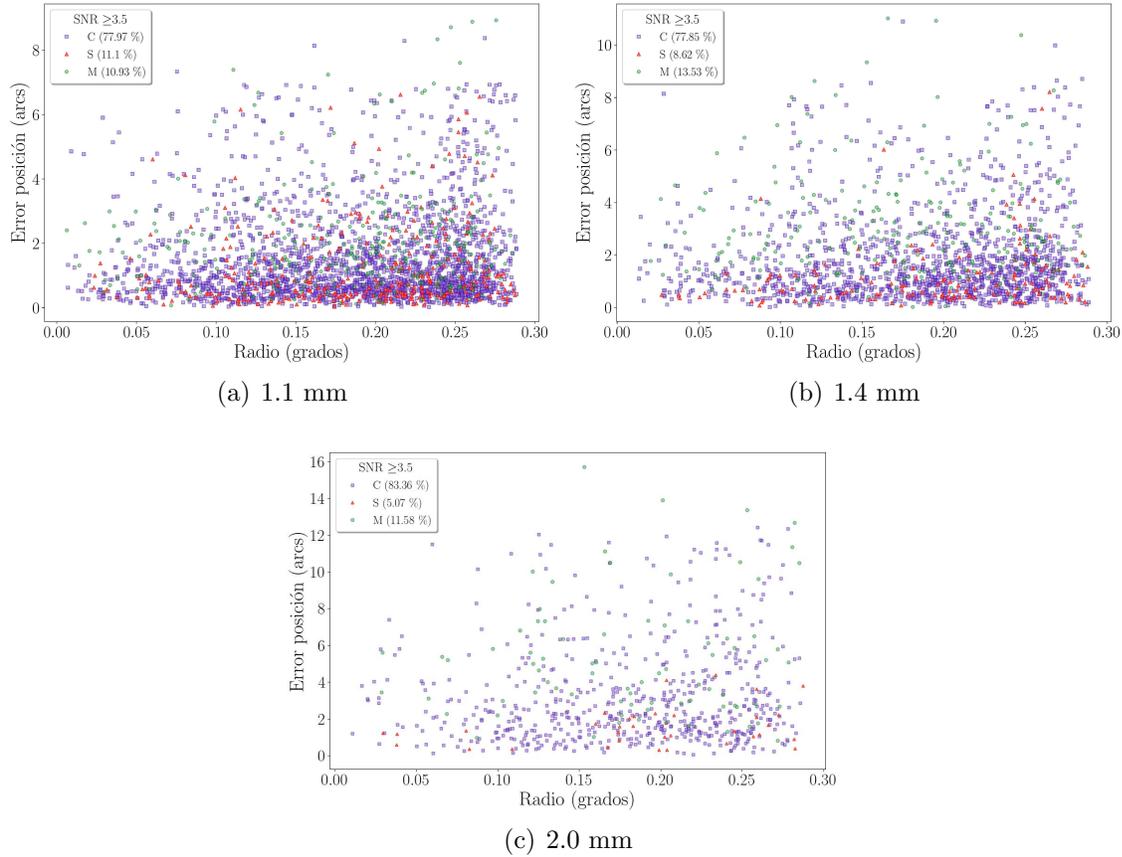


Figura C.20: Errores de posición en radios concéntricos desde el centro del mapa de la simulación 3 para las tres bandas y cortes en $\text{SNR} \geq 3.5$. Los cuadrados, triángulos y círculos representan las fuentes identificadas como tipo 'C', 'S' y 'M', respectivamente.

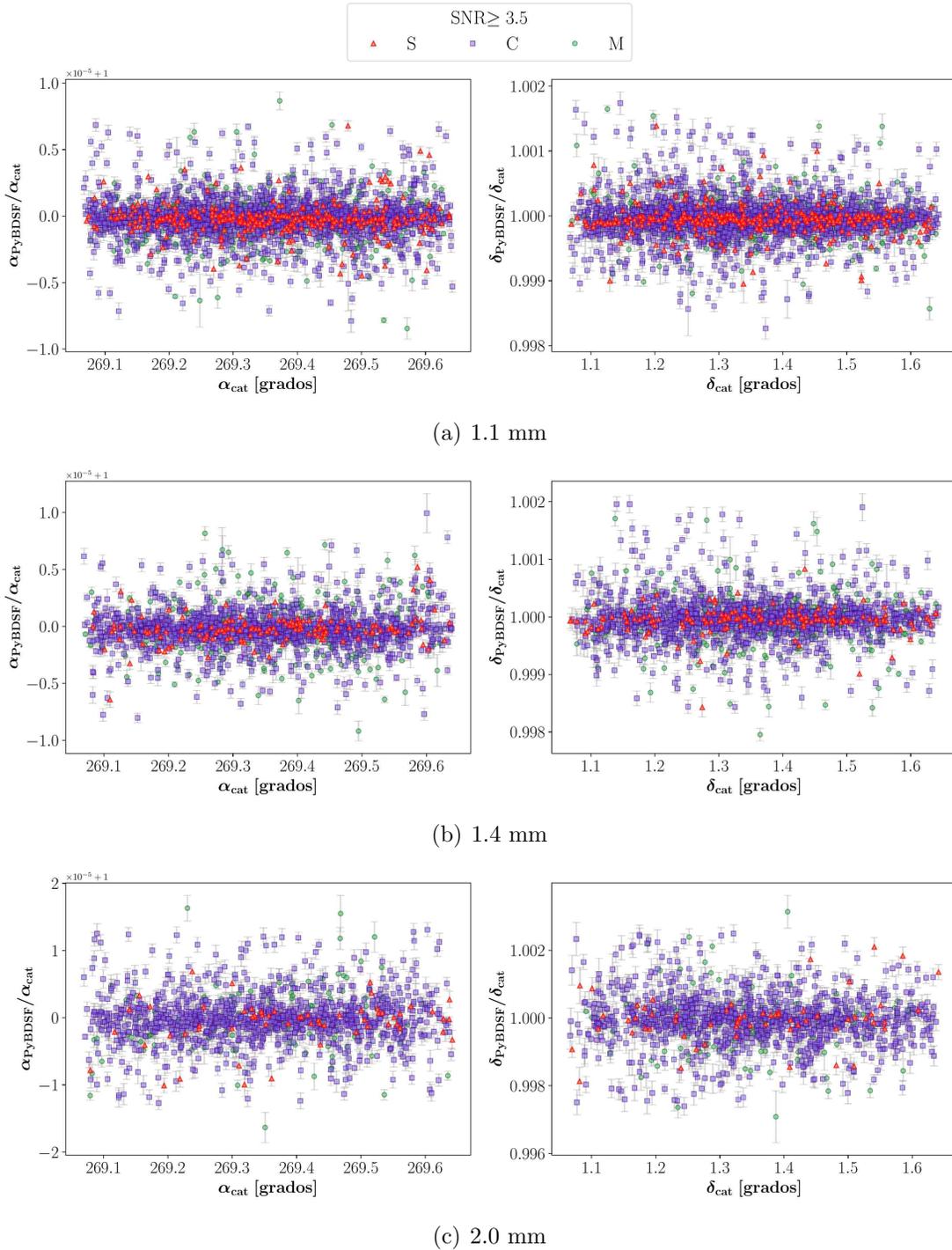
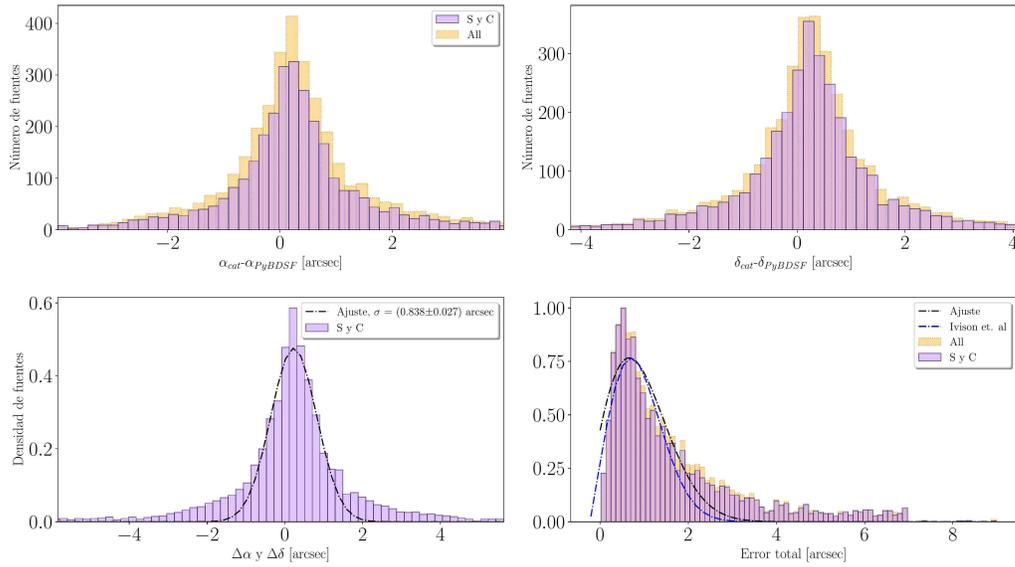
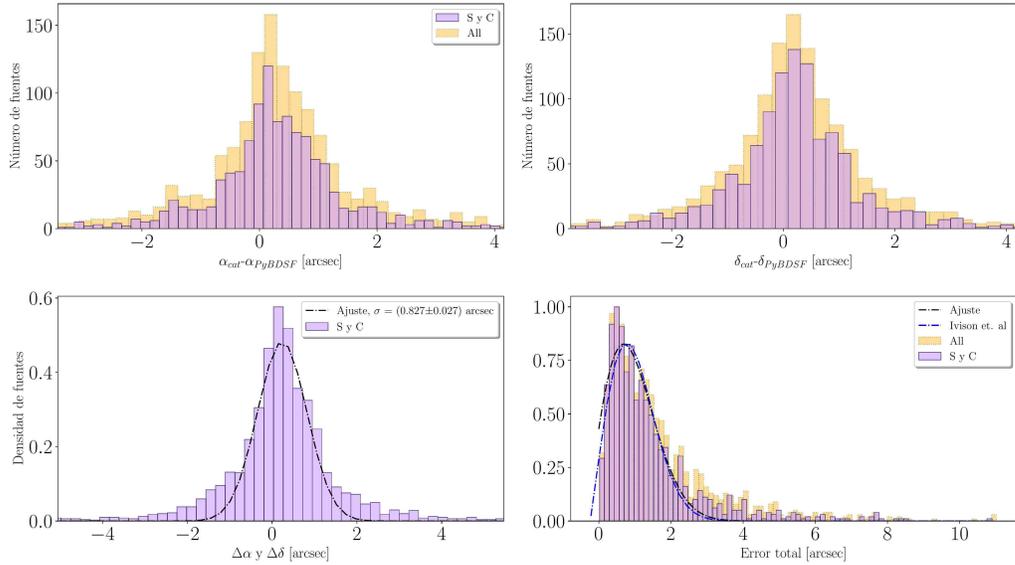


Figura C.21: Cocientes de las posiciones en α , δ respecto a las de entrada para cada una de las tres bandas en la simulación 3. Los puntos rojos, violetas y verdes corresponden a las fuentes tipo ‘S’, ‘C’ y ‘M’.



(a) 1.1 mm



(b) 1.4 mm

Figura C.22: Histogramas de errores de posición para las fuentes detectadas en las bandas de 1.1 mm y 1.4 mm para la simulación 3. La primera fila presenta el error en α y δ . La gráfica inferior izquierda presenta el resultado del histograma conjunto de los errores en α y δ y la gráfica inferior derecha, el error de posición radial, donde la línea discontinua negra muestra el ajuste de la ecuación (3.4.7) dejando el tamaño del haz como parámetro libre y la línea azul el ajuste fijando el tamaño del haz al valor sintético.

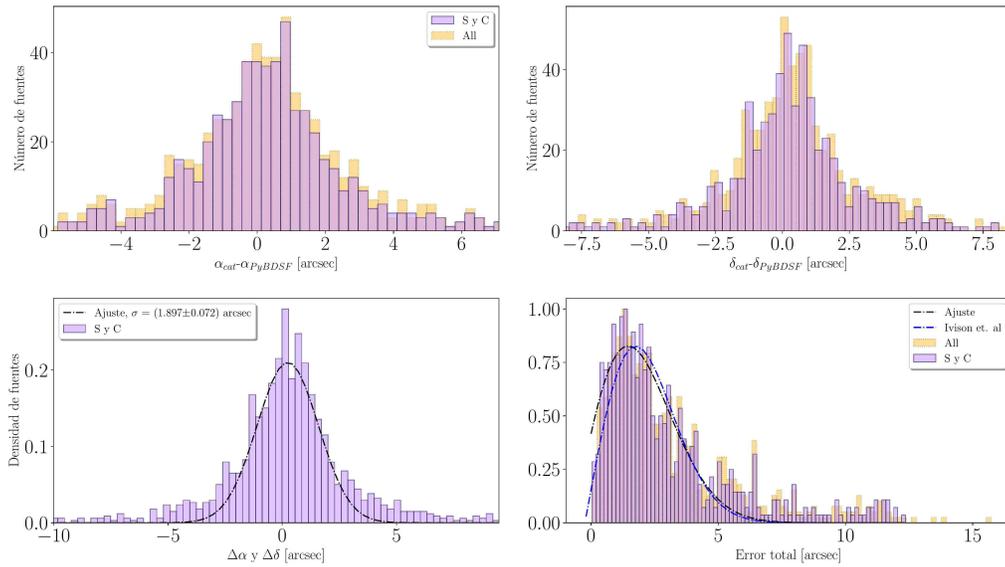


Figura C.23: Histogramas de errores de posición para las fuentes detectadas en la banda de 2.0 mm para la simulación 3. La primera fila presenta el error en α y δ . La gráfica inferior izquierda presenta el resultado del histograma conjunto de los errores en α y δ y la gráfica inferior derecha, el error de posición radial, donde la línea discontinua negra muestra el ajuste de la ecuación (3.4.7) dejando el tamaño del haz como parámetro libre y la línea azul el ajuste fijando el tamaño del haz al valor sintético.

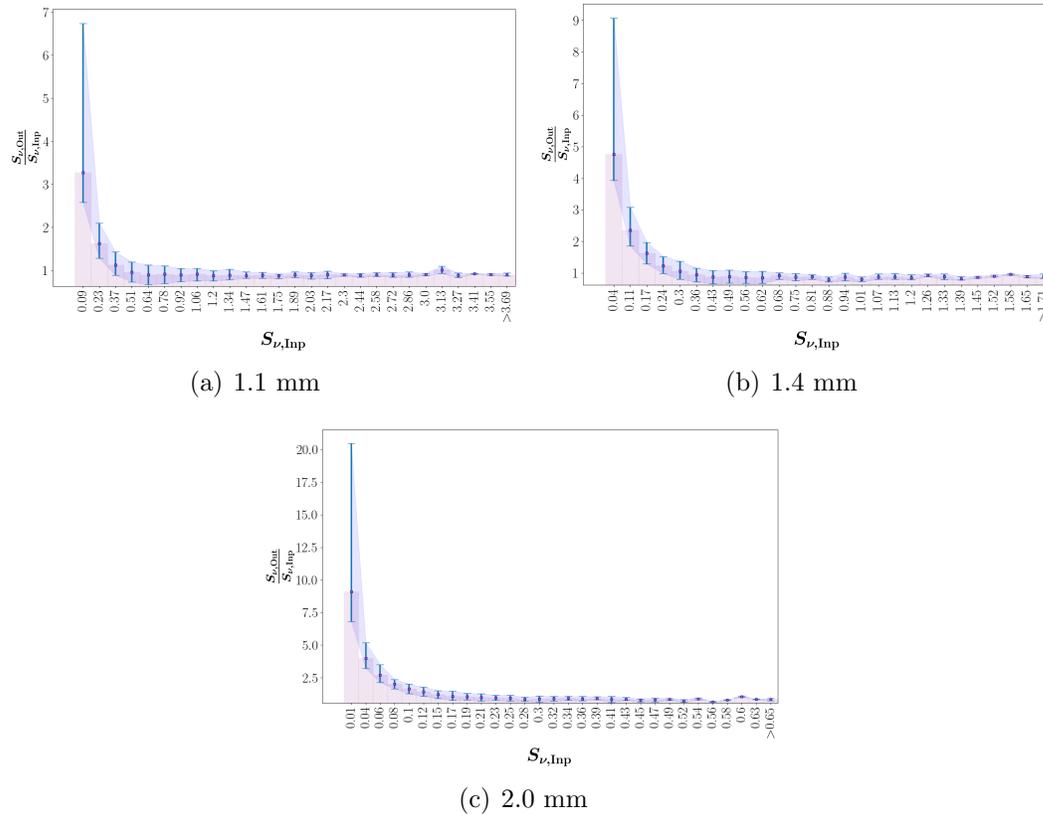


Figura C.24: Cociente medio ponderado de densidades de flujo por rangos de densidades de flujos de entrada en la simulación 3. Las barras de error indican el intervalo de confianza del 68 %.

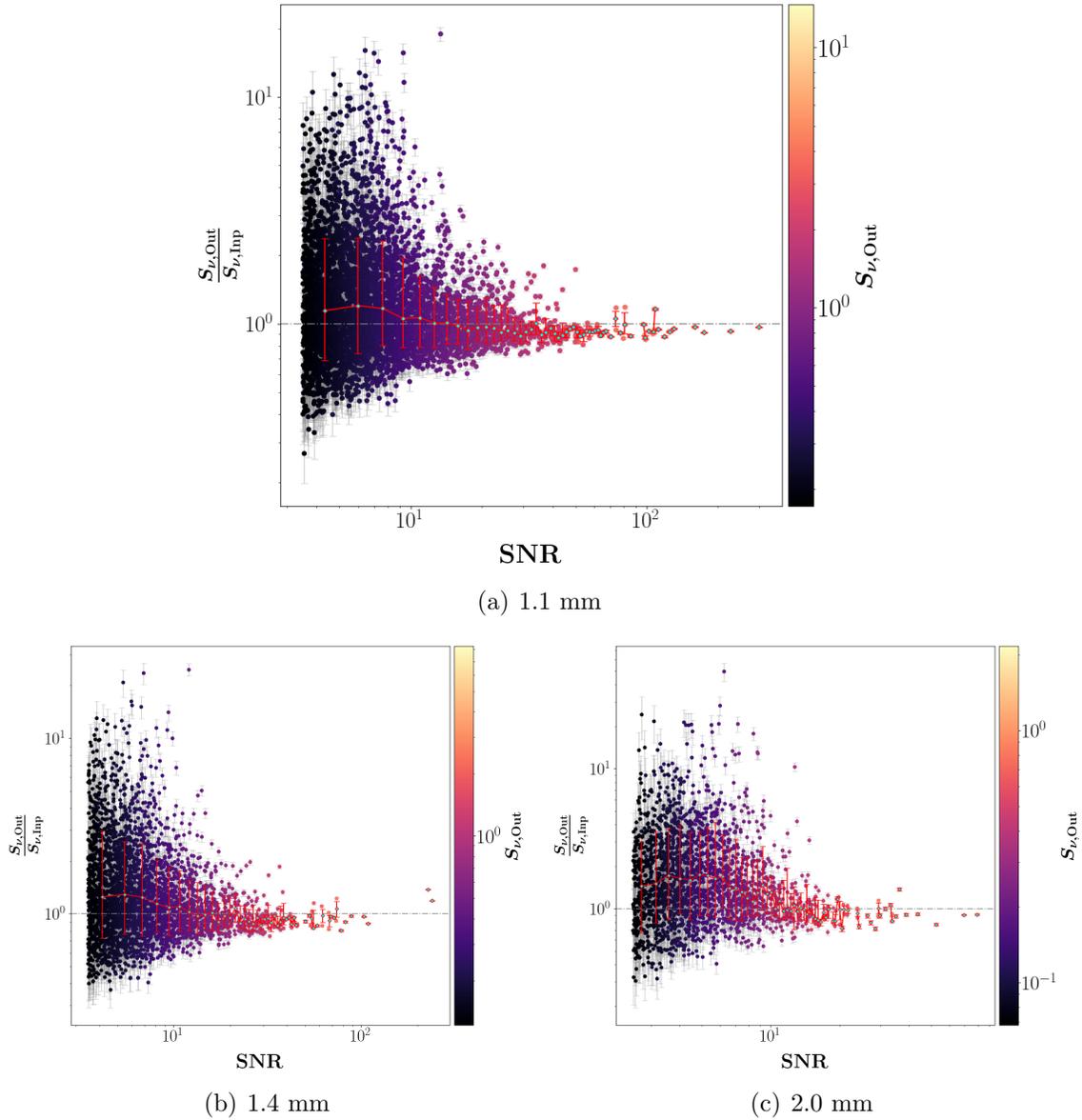
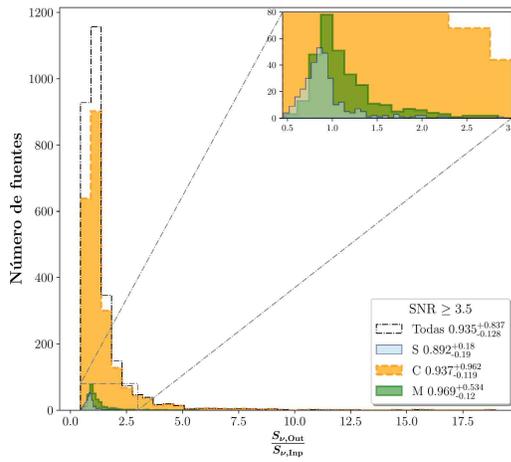
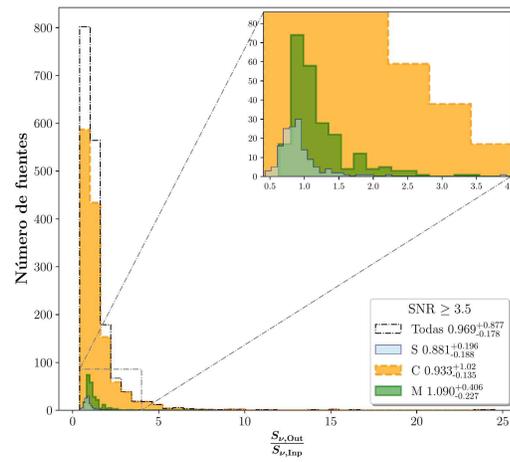


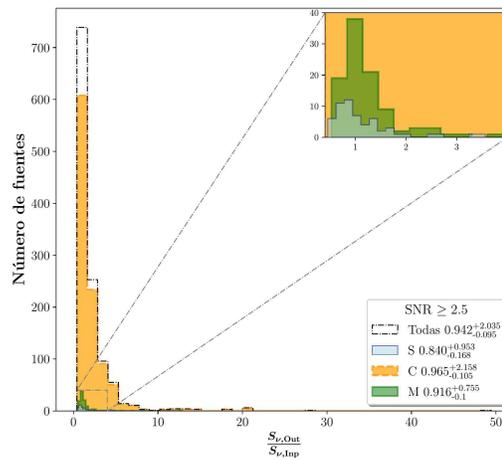
Figura C.25: Factor de *boosting* estimado de la simulación 3, que se calcula como la razón entre la densidad de flujo medida por PyBDSF $S_{\nu,Out}$ y el flujo de entrada del catálogo de Nava-Moreno et al. (2019). La línea roja representa el promedio ponderado del factor de *boosting* para cada intervalo de señal a ruido medido (*bin*). Nótese que el valor del cociente inferior a 1 se debe a la pérdida de flujo experimentada en las etapas de simulación de la observación. Las barras de error indican el intervalo de confianza del 68 %



(a) 1.1 mm



(b) 1.4 mm



(c) 2.0 mm

Figura C.26: Histogramas del cociente de densidades de flujo de salida y entrada para las fuentes detectadas. Se muestran en azul, naranja y verde los histogramas para las fuentes tipo 'S', 'C' y 'M' y en línea punteada el histograma conjunto.

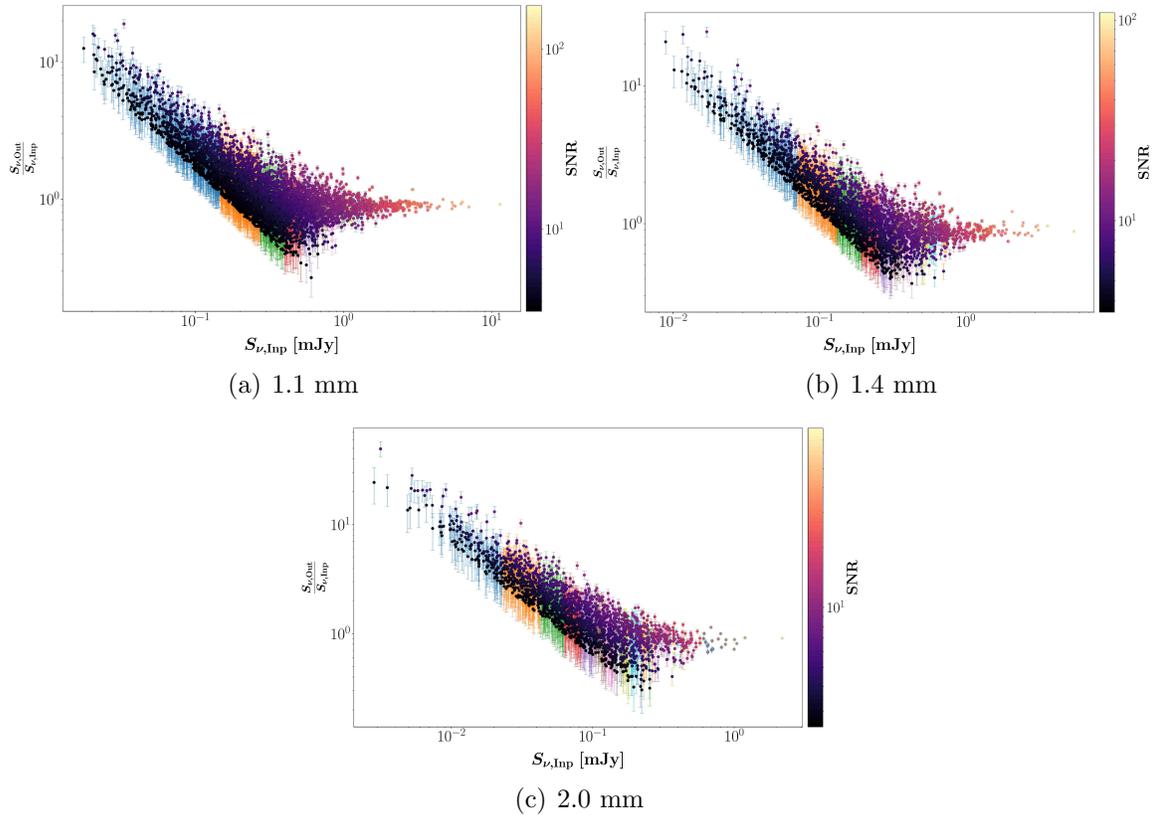


Figura C.27: Cociente de densidad de flujo de salida respecto al de entrada ($S_{\nu,Out}/S_{\nu,Inp}$) en escala logarítmica para la simulación 3. Las barras de error indican un intervalo de confianza del 68 %, mientras que los colores muestran la señal a ruido local en cada medida.

$S_{\nu,\text{umbral}}$ (mJy)	N_{cat}	N_{PyBDSF}	N_{asig}
	$(S_{\nu} \geq S_{\nu,\text{umbral}})$		
banda 1.1 mm			
0.054	12592	5839	6147
0.108	8541	5147	5911
0.161	5943	4335	4933
0.215	4277	3524	3835
0.269	3260	2856	3027
0.323	2555	2328	2389
0.377	2073	1933	1904
0.43	1700	1603	1555
0.484	1442	1380	1311
0.538	1242	1195	1131
banda 1.4 mm			
0.04	10469	3748	4017
0.079	6087	3194	3896
0.119	3861	2609	3236
0.159	2701	2110	2493
0.198	1971	1677	1866
0.238	1535	1371	1433
0.278	1210	1121	1135
0.318	990	936	927
0.357	824	785	766
0.397	706	680	638
banda 2.0 mm			
0.029	5651	1953	2293
0.058	2522	1431	2195
0.087	1419	1063	1629
0.116	924	785	1087
0.146	644	584	766

$S_{\nu, \text{umbral}}$ (mJy)	N_{cat}	N_{PyBDSF} ($S_{\nu} \geq S_{\nu, \text{umbral}}$)	N_{asig}
banda 1.1 mm			
0.175	474	446	538
0.204	347	336	387
0.233	261	256	275
0.262	206	203	208
0.291	172	170	155

Tabla C.7: Número de fuentes recuperadas sobre umbrales de densidades de flujo de entrada en la simulación 3. La primera columna muestra el corte por densidades de flujo de entrada S_{ν} (mJy). La segunda columna, el número de fuentes de entrada sobre ese umbral. La tercera columna, el número de fuentes medidas a la salida sobre ese umbral y la última columna, el número de fuentes que fueron asignadas sobre ese umbral.

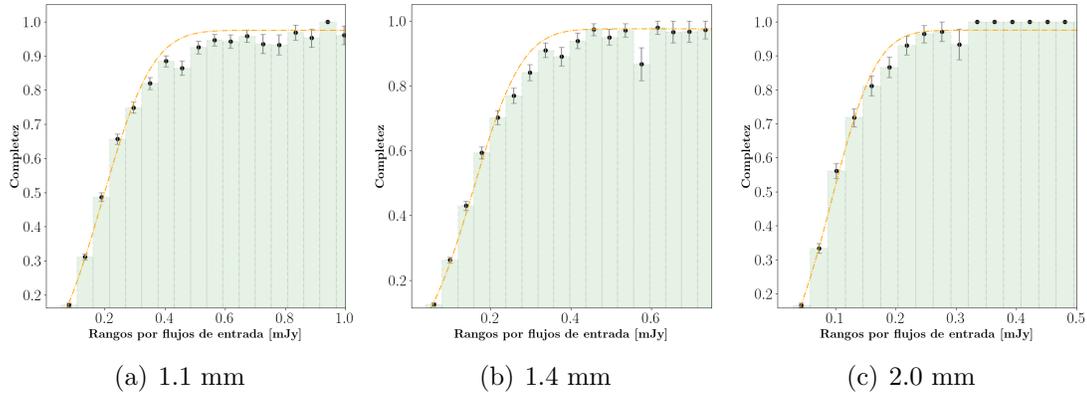
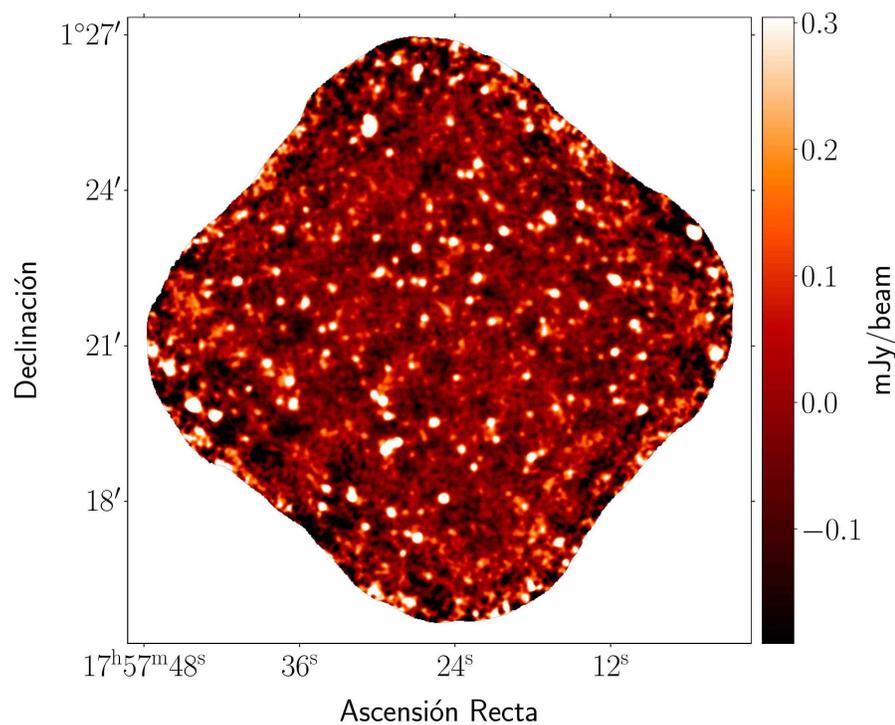


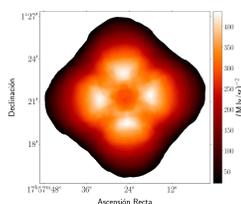
Figura C.28: Completez obtenida para la simulación 3. Los puntos en la gráfica junto con las barras de error estimadas muestran las fuentes que han sido identificadas y asociadas con las fuentes de entrada en cada *bin* representado en el gráfico. Además, se ha trazado una línea discontinua de color naranja que representa la función de mejor ajuste de los datos con la ecuación (4.5.1)

C.4. Resultados simulación Lissajous

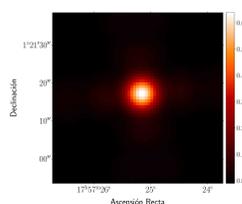
Resultados para una simulación de observación con cartografiado Lissajous y un modelo de atmósfera variable. Los resultados de la simulación corresponden a 20 minutos de tiempo de integración. Se obtiene un mapa *rms* entre (0.0443–0.178) mJy/haz en la banda de 1.1 mm.



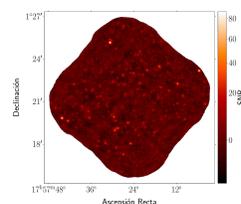
(a) Mapa de señal



(b) Mapa de peso



(c) Mapa de *kernel* (Zoom)



(d) Mapa de señal a ruido

Figura C.29: Productos de ciencia para la simulación con un cartografiado Lissajous en la banda de 1.1 mm y considerando un modelo de atmósfera variable.

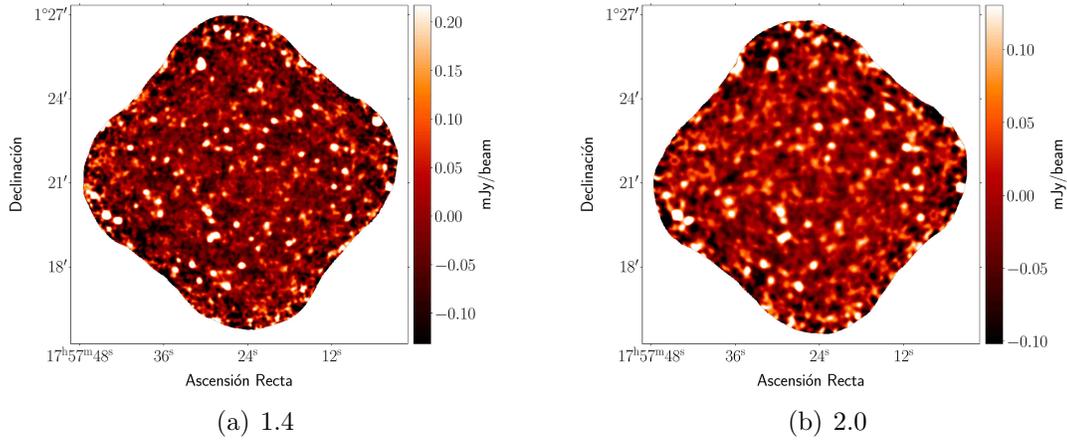


Figura C.30: Mapas de ciencia para la simulación con cartografiado cartografiado Lissajous.

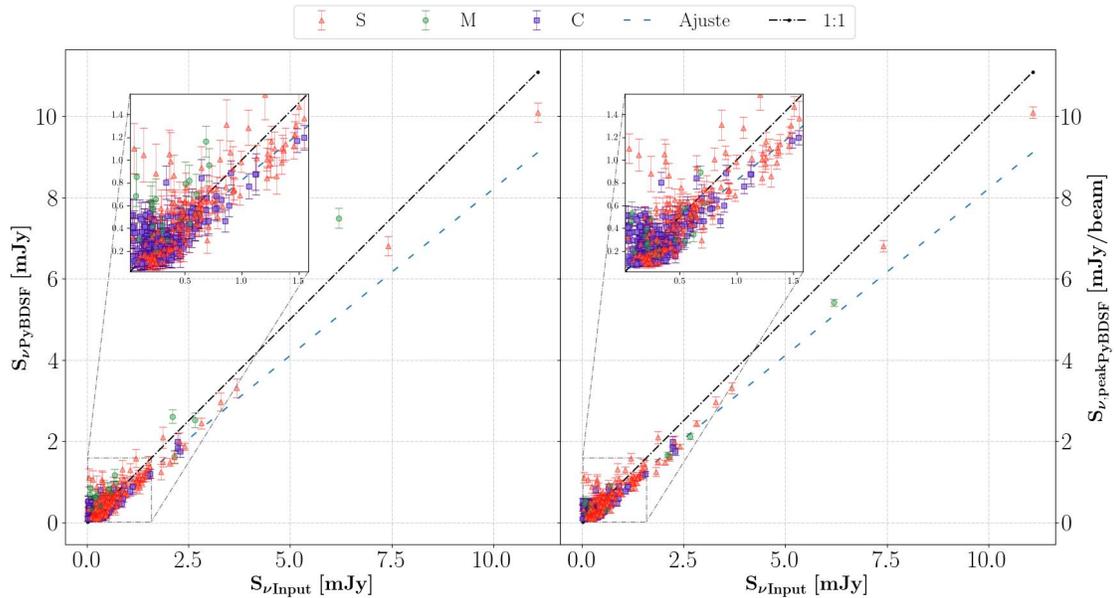


Figura C.31: Densidades de flujo de salida $S_{\nu, PyBDSF}$ respecto a las de salida $S_{\nu, Input}$ en la banda de **1.1 mm** para cada uno de los tipos de estructuras de fuentes detectados por PyBDSF. Los cuadrados, círculos y triángulos representan las estructuras tipo C, M y S, respectivamente. En la derecha se grafican los flujos integrados totales mientras que la izquierda se muestran los flujos picos. La línea negra indica la relación uno a uno y la línea azul el ajuste a los puntos.

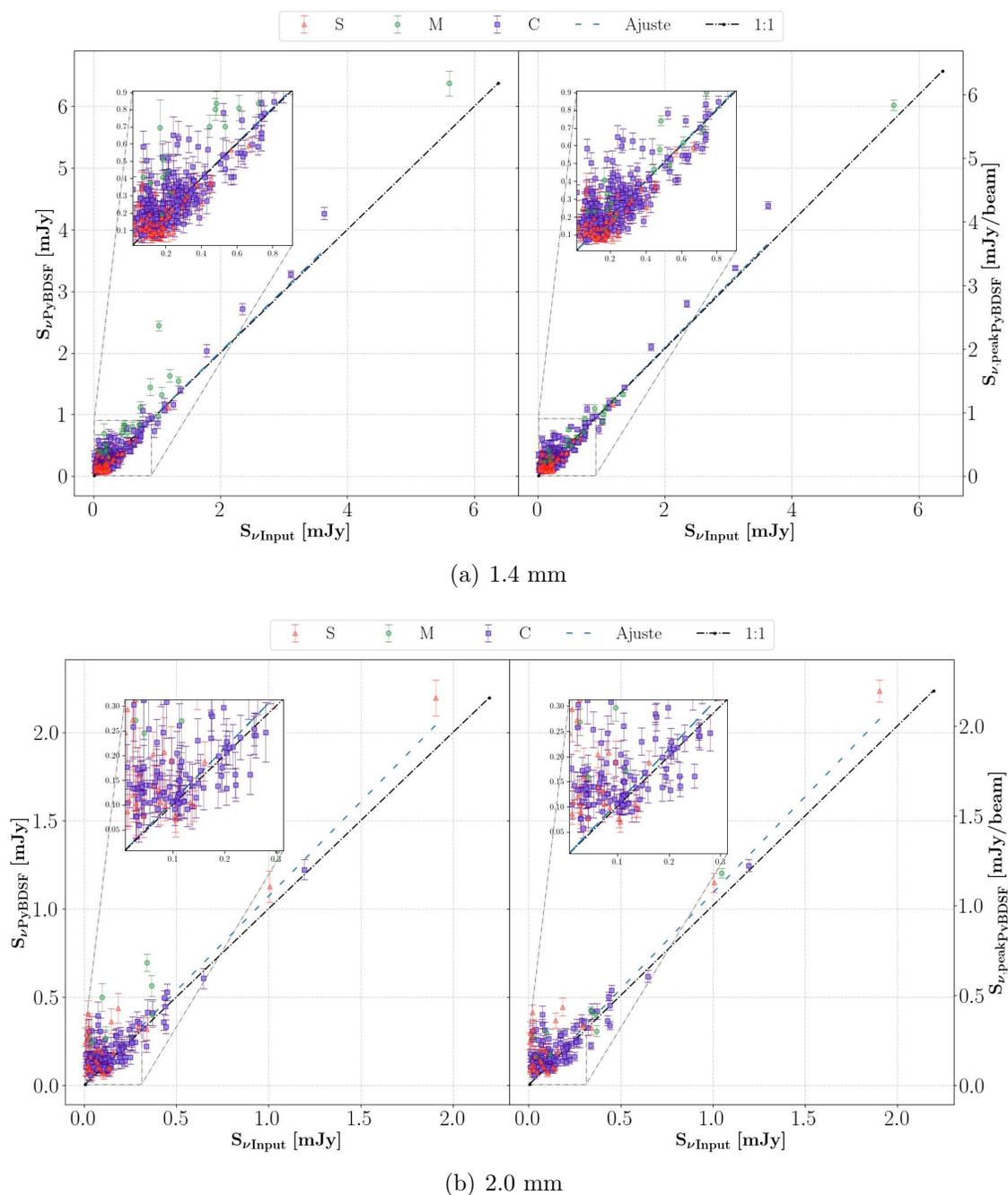


Figura C.32: Densidades de flujo de salida $S_{\nu,PyBDSF}$ respecto a las de salida $S_{\nu,Input}$ en la banda de **1.4 mm** y **2.0 mm** para cada uno de los tipos de estructuras de fuentes detectados por PyBDSF. Los cuadrados, círculos y triángulos representan las estructuras tipo C, M y S, respectivamente. En la derecha se grafican los flujos integrados totales mientras que la izquierda se muestran los flujos picos. La línea negra indica la relación uno a uno y la línea azul el ajuste a los puntos.

SNR	Fuentes verdaderas \geq SNR	Fuentes falsas \geq SNR (%)
1.1 mm		
1.0	445	16 (3.47)
1.5	443	16 (3.49)
2.0	386	13 (3.26)
2.5	325	6 (1.81)
3.0	254	0 (0.0)
3.5	192	0 (0.0)
1.4 mm		
1.5	304	5 (1.62)
2.0	275	4 (1.43)
2.5	234	1 (0.43)
3.0	203	0 (0.0)
3.5	170	0 (0.0)
4.0	144	0 (0.0)
2.0 mm		
1.5	150	1 (0.66)
2.0	143	1 (0.69)
2.5	128	0 (0.0)
3.0	100	0 (0.0)

Tabla C.8: Detecciones verdaderas y falsas, con la simulación usando un patrón de cartografiado Lissajous, por cortes en señal a ruido en las tres bandas. La primera columna muestra los cortes en señal a ruido, la segunda y tercera el número de fuentes verdaderas y falsas por cada corte, así en paréntesis se especifica el porcentaje de detecciones falsas.

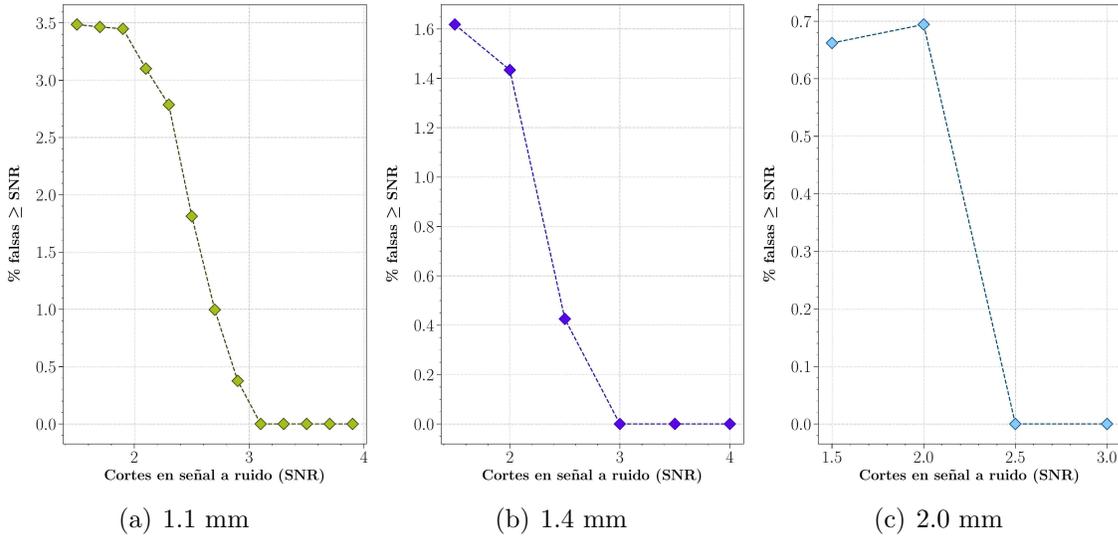


Figura C.33: Porcentaje de fuentes falsas detectadas sobre umbrales de señal a ruido para la simulación con cartografiado Lissajous en las tres bandas.

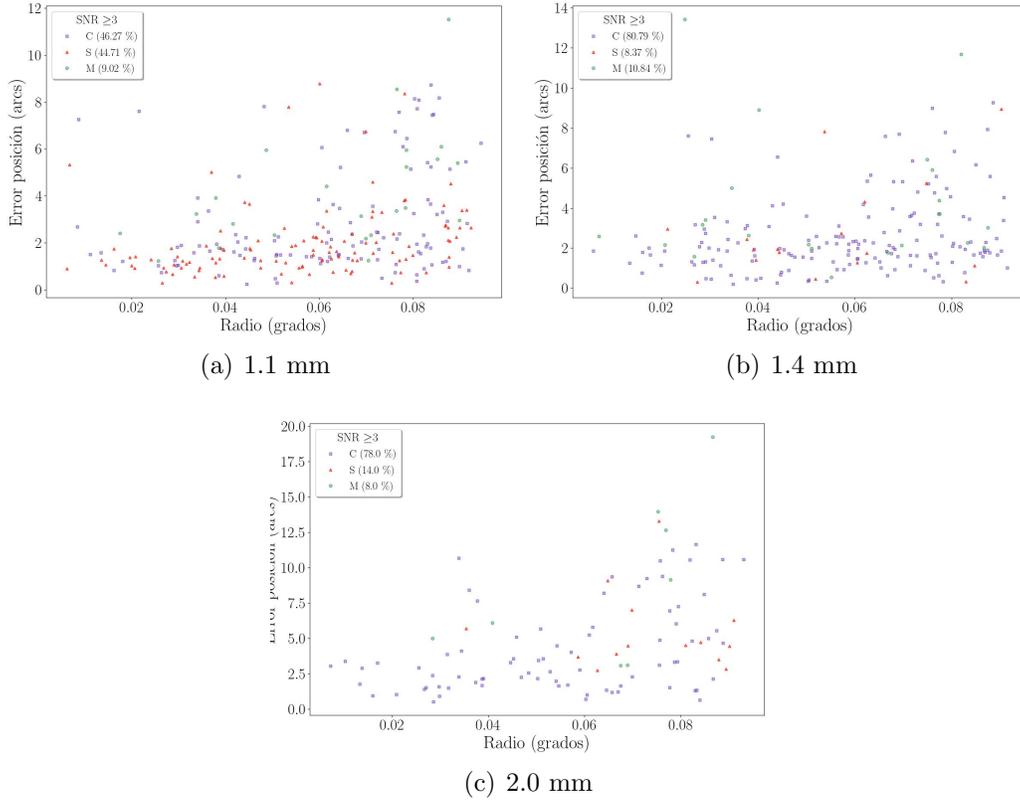
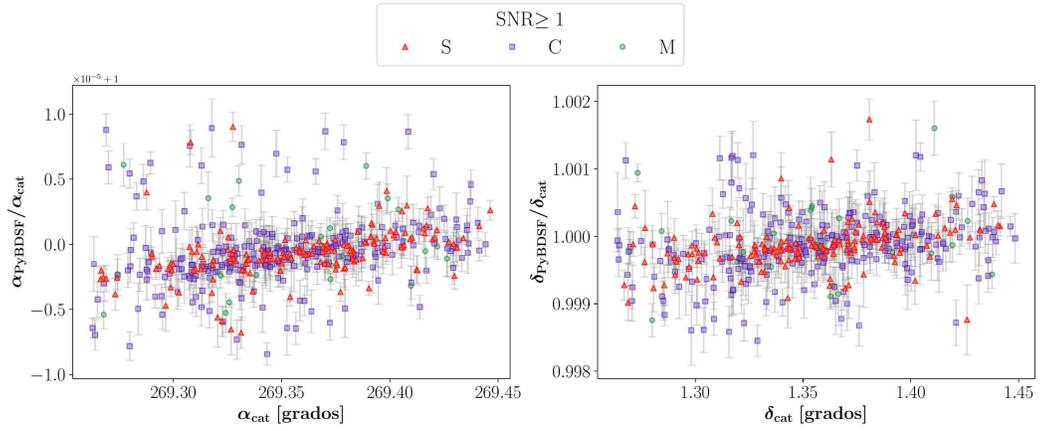


Figura C.34: Errores de posición en radios concéntricos desde el centro del mapa de la simulación 4 para las tres bandas. Los cuadrados, triángulos y círculos representan las fuentes identificadas como tipo 'C', 'S' y 'M', respectivamente.

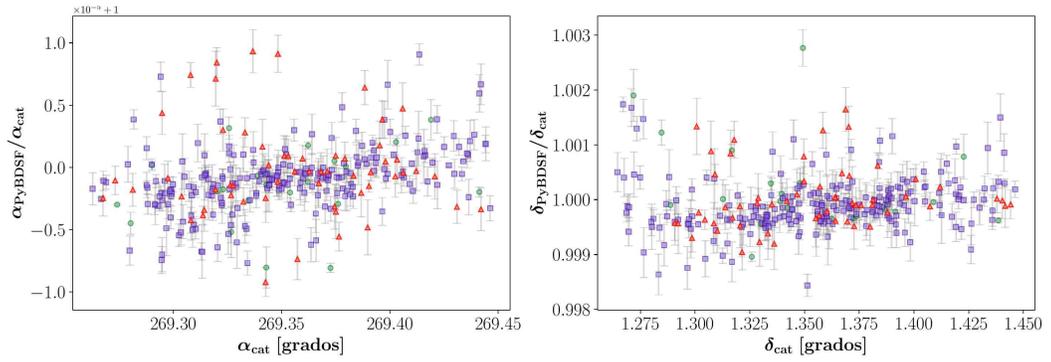
SNR	S		C		M	
	N_s	ϵ_S [arcs]	N_C	ϵ_C [arcs]	N_M	ϵ_M [arcs]
1.1 mm						
≥ 1	158	$1.7^{+1.6}_{-0.8}$	262	$2.3^{+3.0}_{-1.1}$	26	$3.4^{+2.5}_{-1.2}$
≥ 1.5	157	$1.7^{+1.7}_{-0.8}$	261	$2.3^{+3.0}_{-1.1}$	25	$3.4^{+2.6}_{-1.2}$
≥ 3	114	$1.6^{+1.5}_{-0.7}$	118	$2.0^{+3.5}_{-0.8}$	23	$3.4^{+2.6}_{-1.1}$
[2, 3)	22	$2.2^{+1.5}_{-0.9}$	107	$2.6^{+2.8}_{-1.3}$	3	$4.2^{+5.7}_{-2.3}$
≥ 3.5	115	$1.5^{+1.5}_{-0.7}$	128	$2.0^{+3.2}_{-0.8}$	23	$3.4^{+2.6}_{-1.1}$

SNR	S		C		M	
	N_s	ϵ_S [arcs]	N_C	ϵ_C [arcs]	N_M	ϵ_M [arcs]
[3, 4)	30	$1.9^{+2.6}_{-0.9}$	71	$2.3^{+3.8}_{-1.1}$	9	$5.4^{+2.4}_{-2.9}$
[4, 5)	24	$1.5^{+1.9}_{-0.8}$	26	$1.7^{+2.0}_{-0.5}$	7	$3.5^{+1.8}_{-1.2}$
≥ 5	60	$1.4^{+0.8}_{-0.5}$	21	$1.8^{+1.1}_{-1.0}$	7	$2.8^{+0.4}_{-0.5}$
1.4 mm						
≥ 1.5	57	$3.0^{+4.2}_{-1.6}$	225	$2.3^{+2.8}_{-1.1}$	22	$2.8^{+3.4}_{-0.8}$
[2, 3)	21	$3.4^{+1.8}_{-1.9}$	51	$3.0^{+3.3}_{-1.0}$	–	–
≥ 3.5	9	$1.8^{+0.6}_{-1.1}$	110	$1.8^{+1.0}_{-0.8}$	18	$2.4^{+1.6}_{-0.5}$
[3, 4)	8	$2.9^{+4.7}_{-1.6}$	48	$3.2^{+2.4}_{-1.8}$	3	$12.0^{+1.2}_{-3.6}$
≥ 4	9	$1.7^{+0.6}_{-1.1}$	116	$1.8^{+1.2}_{-0.8}$	19	$2.6^{+1.7}_{-0.6}$
2.0 mm						
≥ 1.5	30	$4.6^{+5.5}_{-1.8}$	112	$3.4^{+5.6}_{-1.9}$	8	$7.6^{+6.2}_{-4.3}$
[1.5, 4.5)	21	$5.7^{+6.6}_{-3.2}$	74	$3.8^{+5.9}_{-2.1}$	3	$12.6^{+4.5}_{-2.4}$
≥ 3.5	9	$3.9^{+0.6}_{-0.9}$	42	$2.5^{+2.5}_{-1.2}$	5	$5.0^{+3.9}_{-1.9}$
≥ 4.5	9	$3.9^{+0.6}_{-0.9}$	38	$2.2^{+3.2}_{-0.9}$	5	$5.0^{+3.9}_{-1.9}$

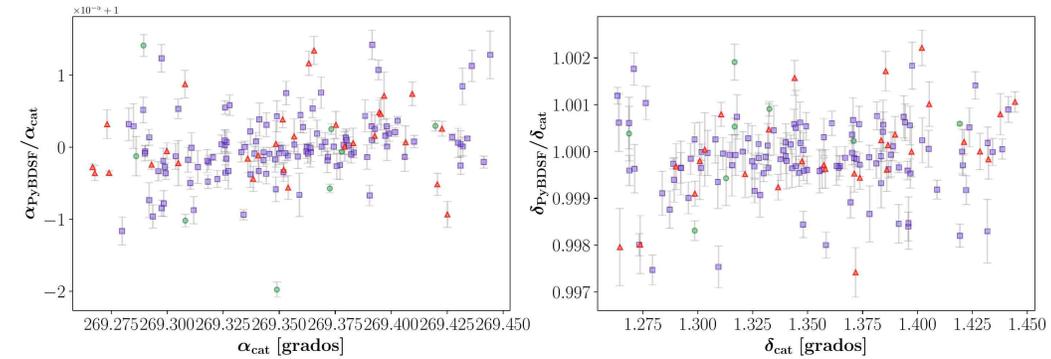
Tabla C.9: Estimaciones del error de posición por cortes de señal a ruido para las tres bandas en la simulación 4, donde 'SNR' indica el umbral o rango de señal a ruido considerado, las letras S, C y M indican el tipo de estructura de fuente identificada, 'N' el número de fuentes y ϵ la estimación del error junto a sus incertidumbres asociadas a un intervalo de confianza del 68%.



(a) 1.1 mm



(b) 1.4 mm



(c) 2.0 mm

Figura C.35: Cocientes de las posiciones en α , δ respecto a las de entrada para cada una de las tres bandas en la simulación 4. Los puntos rojos, violetas y verdes corresponden a las fuentes tipo 'S', 'C' y 'M'.

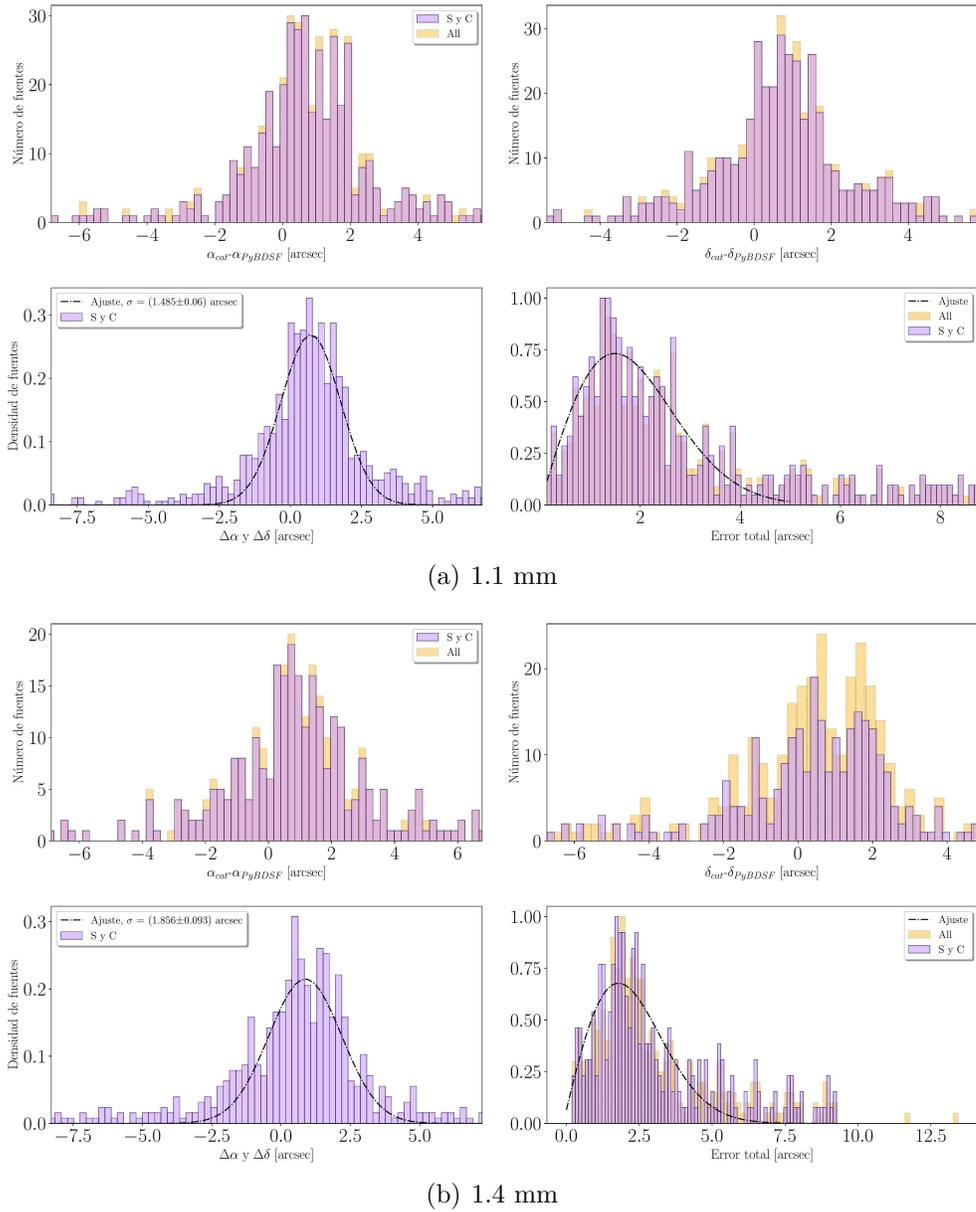
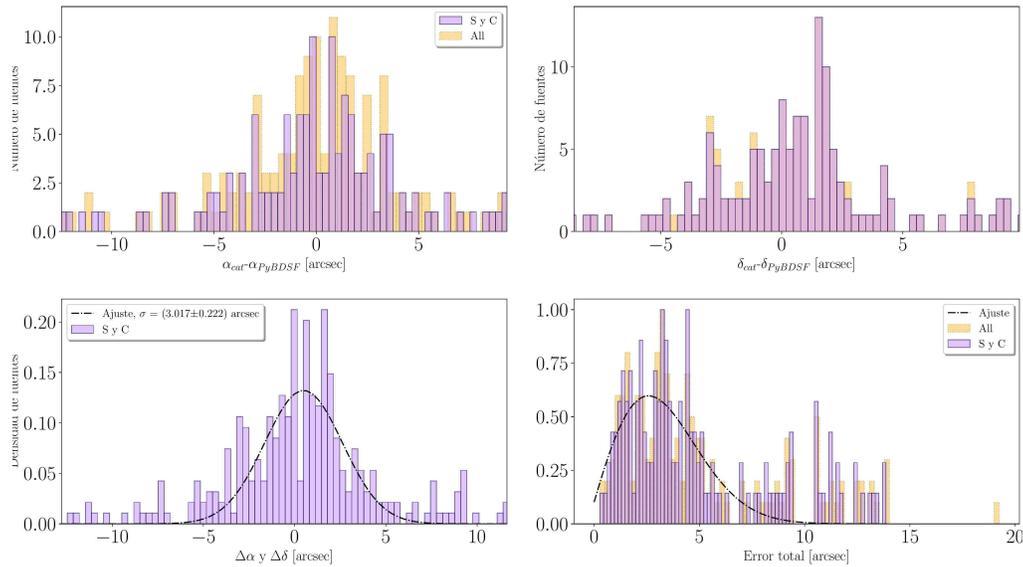


Figura C.36: Histogramas de errores de posición para las fuentes detectadas en las bandas de 1.1 mm y 1.4 mm para la simulación 4. La primera fila presenta el error en α y δ . La gráfica inferior izquierda presenta el resultado del histograma conjunto de los errores en α y δ y la gráfica inferior derecha, el error de posición radial, donde la línea discontinua negra muestra el ajuste de la ecuación (3.4.7) dejando el tamaño del haz como parámetro libre y la línea azul el ajuste fijando el tamaño del haz al valor sintético.



(a) 1.4 mm

Figura C.37: Histogramas de errores de posición para las fuentes detectadas en las bandas de 2.0 mm para la simulación 4. La primera fila presenta el error en α y δ . La gráfica inferior izquierda presenta el resultado del histograma conjunto de los errores en α y δ y la gráfica inferior derecha, el error de posición radial, donde la línea discontinua negra muestra el ajuste de la ecuación (3.4.7) dejando el tamaño del haz como parámetro libre y la línea azul el ajuste fijando el tamaño del haz al valor sintético.

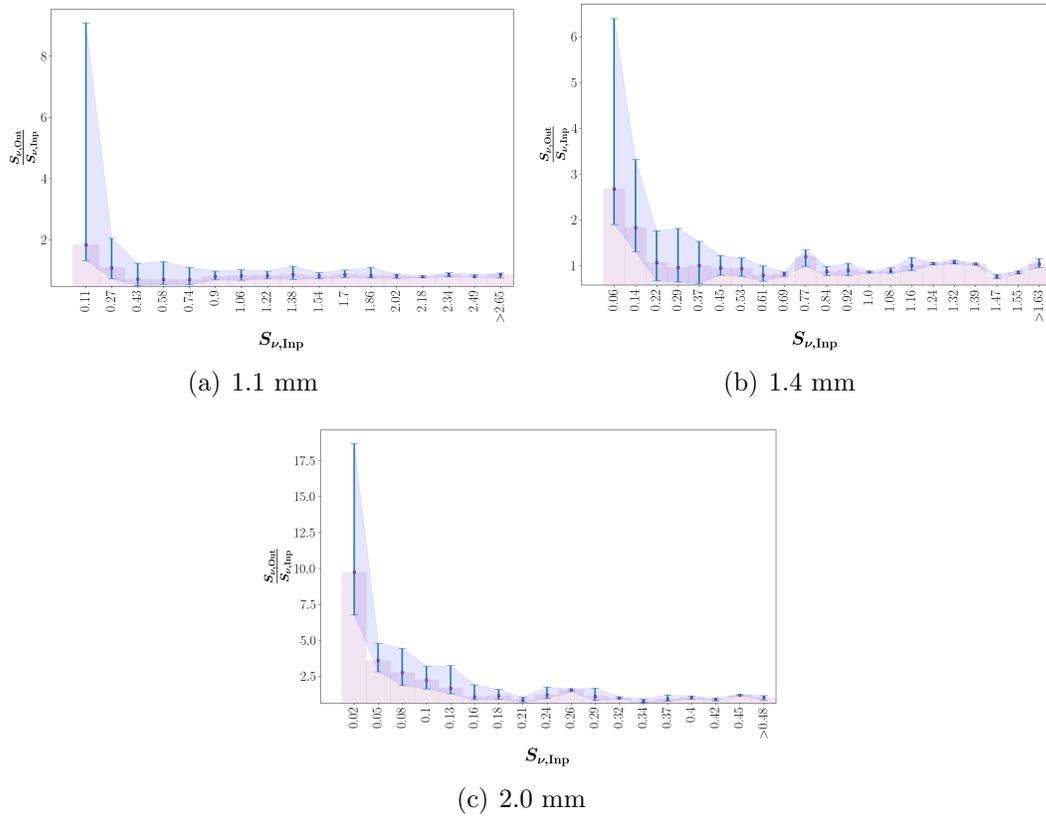


Figura C.38: Cociente medio ponderado de densidades de flujo por rangos de densidades de flujos de entrada en la simulación 4. Las barras de error indican el intervalo de confianza del 68 %.

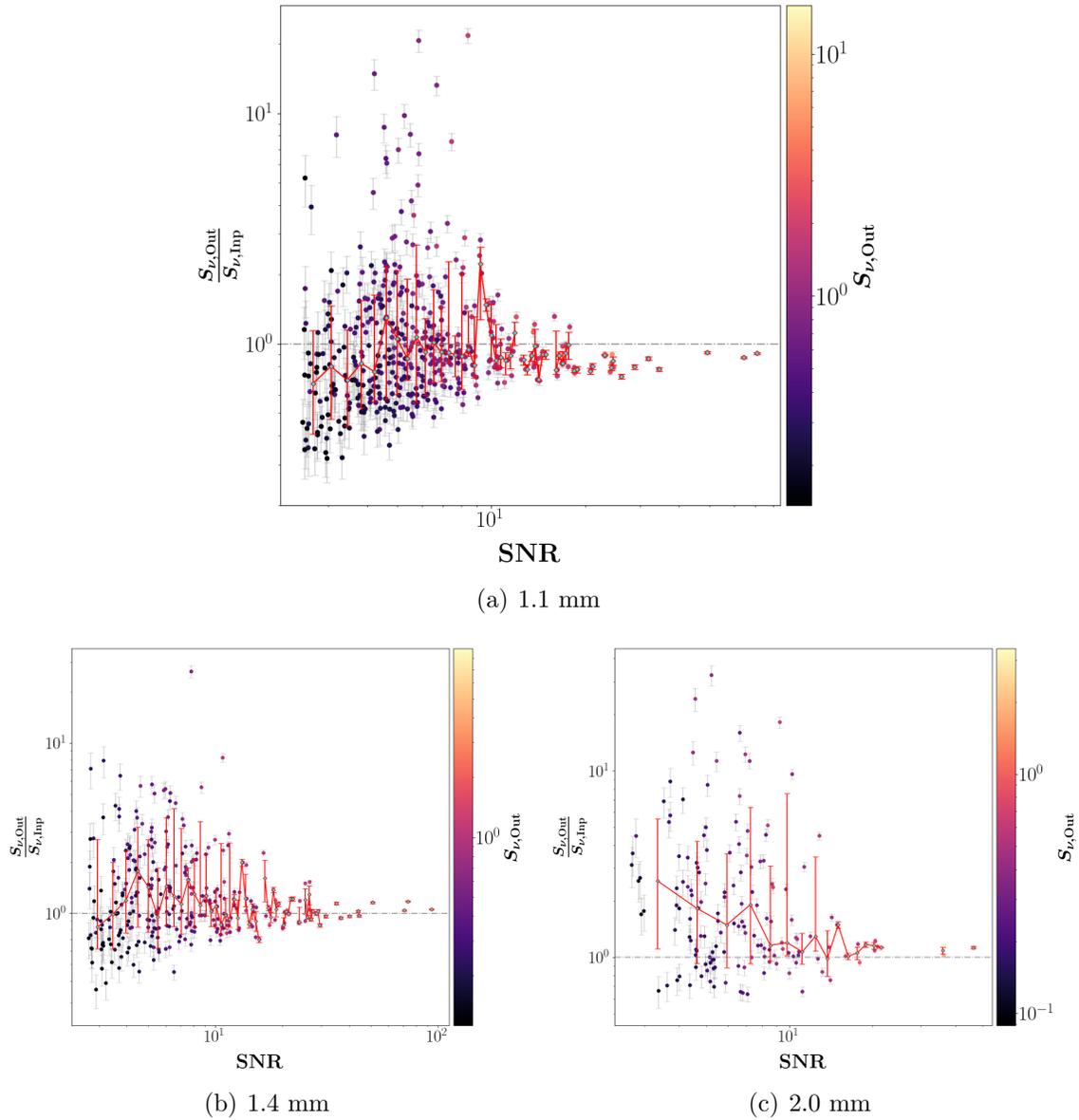
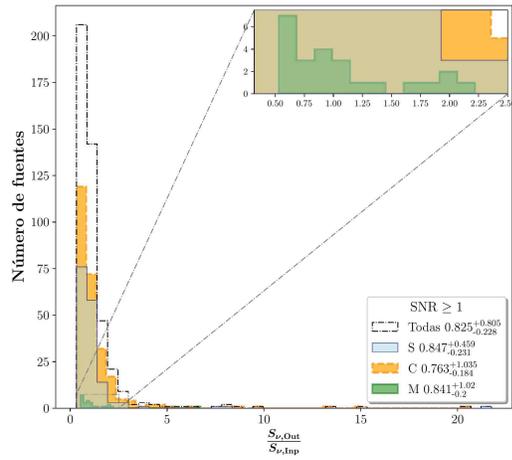
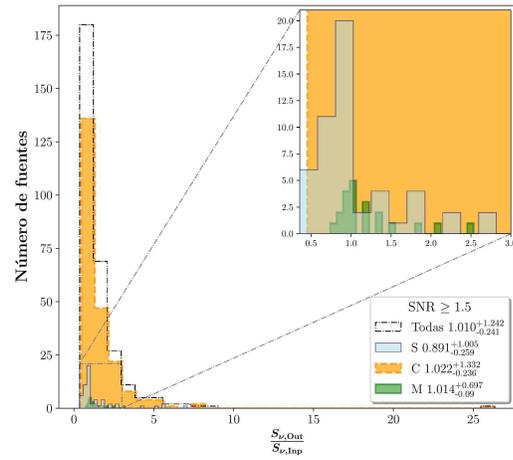


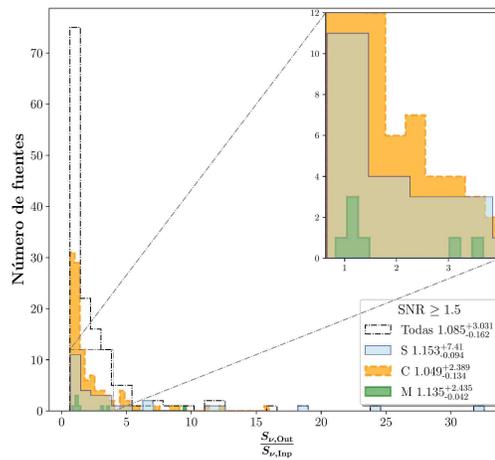
Figura C.39: Factor de *boosting* estimado de la simulación 4, que se calcula como la relación entre la densidad de flujo medida por PyBDSF $S_{\nu,Out}$ y el flujo de entrada del catálogo de Nava-Moreno et al. (2019). La línea roja representa el promedio ponderado del factor de *boosting* para cada intervalo de señal a ruido medido (*bin*). Nótese que el valor del cociente inferior a 1 se debe a la pérdida de flujo experimentada en las etapas de simulación de la observación. Las barras de error indican el intervalo de confianza del 68 %



(a) 1.1 mm



(b) 1.4 mm



(c) 2.0 mm

Figura C.40: Histogramas del cociente de densidades de flujo de salida y entrada para las fuentes detectadas. Se muestran en azul, naranja y verde los histogramas para las fuentes tipo 'S', 'C' y 'M' y en línea punteada el histograma conjunto.

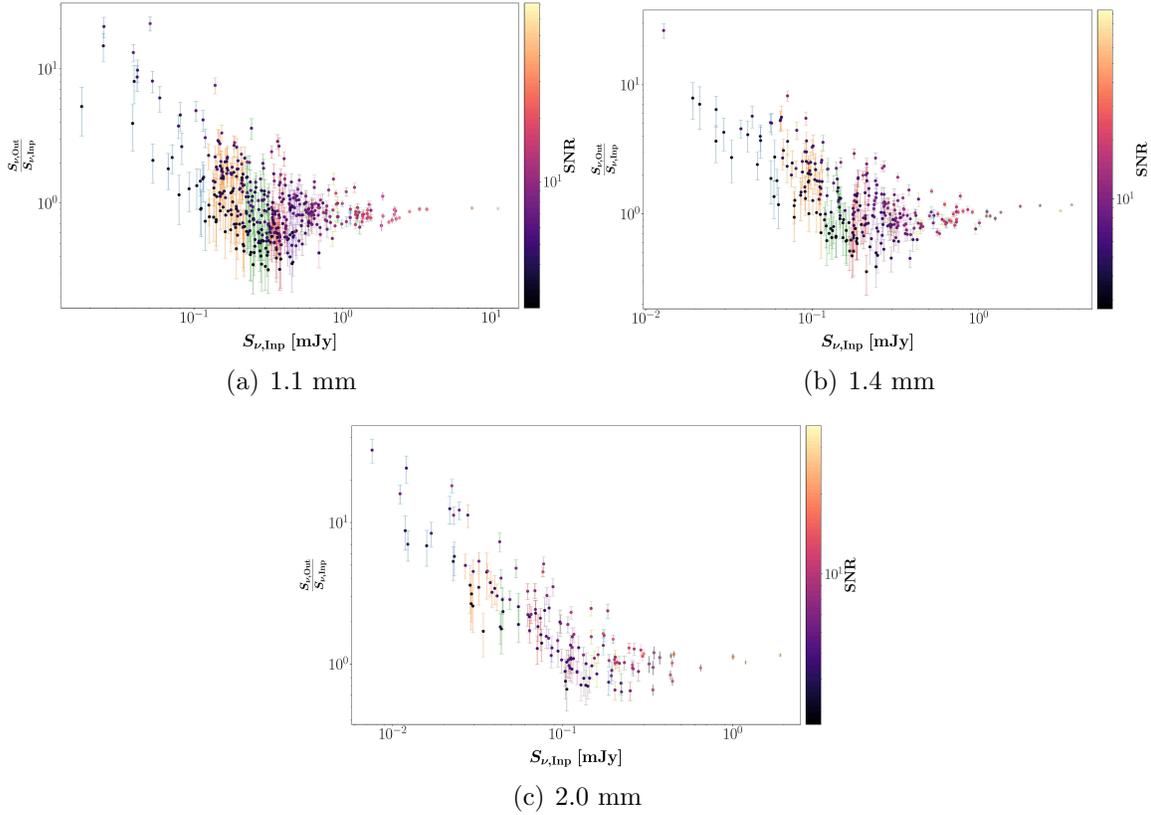


Figura C.41: Cociente de densidad de flujo de salida respecto al de entrada ($S_{\nu,\text{Out}}/S_{\nu,\text{Inp}}$) en escala logarítmica para la simulación 4. Las barras de error indican un intervalo de confianza del 68 %, mientras que los colores muestran la señal a ruido local en cada medida.

$S_{\nu,\text{umbral}}$ (mJy)	N_{cat}	N_{PyBDSF}	N_{asig}
	$(S_{\nu} \geq S_{\nu,\text{umbral}})$		
banda 1.1 mm			
0.06	1187	433	446
0.12	788	413	420
0.18	522	360	350
0.24	359	296	289
0.3	280	244	227
0.36	222	196	188

$S_{\nu,\text{umbral}}$ (mJy)	N_{cat}	N_{PyBDSF} ($S_{\nu} \geq S_{\nu,\text{umbral}}$)	N_{asig}
0.42	181	160	163
0.48	140	130	137
0.54	119	110	103
0.6	102	95	87
banda 1.4 mm			
0.05	906	291	304
0.1	457	247	282
0.15	282	199	237
0.2	197	159	188
0.25	138	121	156
0.3	104	95	119
0.35	80	76	93
banda 2.0 mm			
0.05	296	112	150
0.1	109	80	135
0.15	56	49	85
0.2	40	37	60
0.25	28	26	42
0.3	20	18	31
0.35	13	13	19
0.4	10	10	13
0.45	6	6	7
0.5	5	5	6

Tabla C.10: Número de fuentes recuperadas sobre umbrales de densidades de flujo de entrada en la simulación 4. La primera columna muestra el corte por densidades de flujo de entrada S_{ν} (mJy). La segunda columna, el número de fuentes de entrada sobre ese umbral. La tercera columna, el número de fuentes medidas a la salida sobre ese umbral y la última columna, el número de fuentes que fueron asignadas sobre ese umbral.

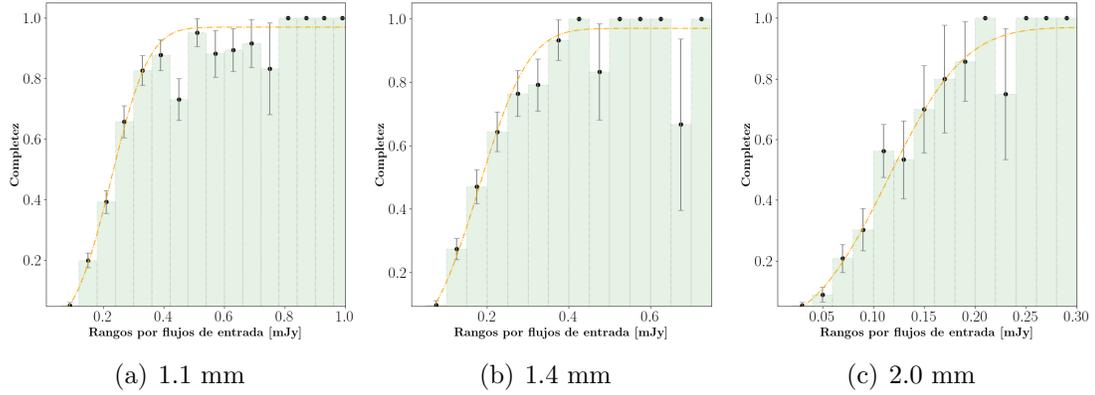


Figura C.42: Completez obtenida para la simulación 4. Los puntos en la gráfica junto con las barras de error estimadas muestran las fuentes que han sido identificadas y asociadas con las fuentes de entrada en cada *bin* representado en el gráfico. Además, se ha trazado una línea discontinua de color naranja que representa la función de mejor ajuste de los datos con la ecuación (4.5.1).

Bibliografía

- Aretxaga I., Hughes D. H., Chapin E. L., Gaztañaga E., Dunlop J. S., Ivison R. J., 2003, MNRAS, 342, 759
- Aretxaga I., et al., 2011, MNRAS, 415, 3831
- Austermann J. E., et al., 2010, MNRAS, 401, 160
- Austermann J. E., et al., 2018, Journal of Low Temperature Physics, 193, 120
- Barger A. J., Cowie L. L., Sanders D. B., Fulton E., Taniguchi Y., Sato Y., Kawara K., Okuda H., 1998, Nature, 394, 248
- Barger A. J., Cowie L. L., Sanders D. B., 1999, ApJ, 518, L5
- Bertoldi F., et al., 2007, ApJS, 172, 132
- Bian F., Kewley L. J., Dopita M. A., Juneau S., 2016, ApJ, 822, 62
- Bing L., et al., 2023, arXiv e-prints, p. arXiv:2305.07054
- Blain A. W., Smail I., Ivison R. J., Kneib J. P., Frayer D. T., 2002, Phys. Rep., 369, 111
- Borys C., Chapman S., Halpern M., Scott D., 2003, MNRAS, 344, 385
- Casey C. M., Narayanan D., Cooray A., 2014, Phys. Rep., 541, 45
- Chapman S. C., Blain A. W., Ivison R. J., Smail I. R., 2003, Nature, 422, 695
- Columbro F., et al., 2019, Astronomische Nachrichten, 340, 83

- Condon J. J., 1997, *PASP*, 109, 166
- Coppin K., Halpern M., Scott D., Borys C., Chapman S., 2005, *MNRAS*, 357, 1022
- Coppin K., et al., 2006, *MNRAS*, 372, 1621
- Cowie L. L., Barger A. J., Kneib J. P., 2002, *AJ*, 123, 2197
- Davis M., et al., 2007, *ApJ*, 660, L1
- Draine B. T., 2009, in Henning T., Grün E., Steinacker J., eds, *Astronomical Society of the Pacific Conference Series Vol. 414, Cosmic Dust - Near and Far*. p. 453 ([arXiv:0903.1658](https://arxiv.org/abs/0903.1658)), doi:10.48550/arXiv.0903.1658
- Draine B. T., Li A., 2007, *ApJ*, 657, 810
- Eales S., Lilly S., Gear W., Dunne L., Bond J. R., Hammer F., Le Fèvre O., Crampton D., 1999, *ApJ*, 515, 518
- Eales S., Lilly S., Webb T., Dunne L., Gear W., Clements D., Yun M., 2000, *AJ*, 120, 2244
- Essinger-Hileman T., et al., 2016, *Review of Scientific Instruments*, 87, 094503
- Fixsen D. J., Dwek E., Mather J. C., Bennett C. L., Shafer R. A., 1998, *ApJ*, 508, 123
- Greve T. R., Ivison R. J., Bertoldi F., Stevens J. A., Dunlop J. S., Lutz D., Carilli C. L., 2004, *MNRAS*, 354, 779
- Güsten R., Nyman L. Å., Schilke P., Menten K., Cesarsky C., Booth R., 2006, *A&A*, 454, L13
- Hancock P. J., Murphy T., Gaensler B. M., Hopkins A., Curran J. R., 2012, *MNRAS*, 422, 1812
- Hogg D. W., Turner E. L., 1998, *PASP*, 110, 727
- Holland W. S., et al., 1999, *MNRAS*, 303, 659

- Hopkins A. M., Miller C. J., Connolly A. J., Genovese C., Nichol R. C., Wasserman L., 2002, *AJ*, 123, 1086
- Hughes D. H., et al., 1998, *Nature*, 394, 241
- Hughes D. H., et al., 2020, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. p. 1144522, doi:10.1117/12.2561893
- Iverson R., 2001, in Tran Thanh Van J., Mellier Y., Moniez M., eds, *Cosmological Physics with Gravitational Lensing*. p. 233 (arXiv:astro-ph/0007162), doi:10.48550/arXiv.astro-ph/0007162
- Iverson R. J., et al., 2007, *MNRAS*, 380, 199
- Karim A., et al., 2013, *MNRAS*, 432, 2
- Knudsen K. K., van der Werf P. P., Kneib J. P., 2008, *MNRAS*, 384, 1611
- Kreysa E., et al., 1998, in Phillips T. G., ed., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 3357, Advanced Technology MMW, Radio, and Terahertz Telescopes*. pp 319–325, doi:10.1117/12.317367
- Ma Z., et al., 2020, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. p. 114522O, doi:10.1117/12.2560735
- McCrackan M., et al., 2022, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. p. 121891H, doi:10.1117/12.2629095
- McLure R. J., et al., 2013, *MNRAS*, 432, 2696
- Mohan N., Rafferty D., 2015, *PyBDSF: Python Blob Detection and Source Finder, Astrophysics Source Code Library*, record ascl:1502.007 (ascl:1502.007)
- Nava-Moreno N. A., Montaña-Barbano A., Aretxaga I., 2019, *Tesis de maestría, Instituto Nacional de Astrofísica, Óptica y Electrónica*
- Neugebauer G., et al., 1984, *ApJ*, 278, L1
- Peacock J. A., 1998, *Cosmological Physics*. Cambridge University Press, doi:10.1017/CBO9780511804533

- Perera T. A., et al., 2008, MNRAS, 391, 1227
- Perera T., Wilson G., Scott K. S., Austermann J., Mancera A., 2013, in American Astronomical Society Meeting Abstracts. p. 402.07
- Puget J. L., Abergel A., Bernard J. P., Boulanger F., Burton W. B., Desert F. X., Hartmann D., 1996, A&A, 308, L5
- Riggi S., et al., 2023, Astronomy and Computing, 42, 100682
- Rodríguez-Puebla A., Primack J. R., Behroozi P., Faber S. M., 2015, arXiv e-prints, p. arXiv:1508.04842
- Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, MNRAS, 470, 651
- Sanders D. B., Mirabel I. F., 1996, ARA&A, 34, 749
- Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, ApJ, 325, 74
- Scott S. E., et al., 2002, MNRAS, 331, 817
- Scott K. S., et al., 2008, MNRAS, 385, 2225
- Scott K. S., et al., 2010, MNRAS, 405, 2260
- Scott K. S., et al., 2012, Monthly Notices of the Royal Astronomical Society, 423, 575
- Serjeant S., et al., 2003, MNRAS, 344, 887
- Simon S., et al., 2016, Journal of Low Temperature Physics, 184, 534
- Siringo G., et al., 2009, A&A, 497, 945
- Smail I., Ivison R. J., Blain A. W., 1997, ApJ, 490, L5
- Smail I., Ivison R. J., Owen F. N., Blain A. W., Kneib J. P., 2000, ApJ, 528, 612
- Smail I., Ivison R. J., Blain A. W., Kneib J. P., 2002, MNRAS, 331, 495

- Tacconi L. J., et al., 2008, *ApJ*, 680, 246
- Wang W. H., Cowie L. L., Barger A. J., 2004, *ApJ*, 613, 655
- Webb T. M., et al., 2003, *ApJ*, 587, 41
- Weiß A., et al., 2009, *ApJ*, 707, 1201
- Wilson G. W., et al., 2008, *MNRAS*, 386, 807
- Wilson G. W., et al., 2020, in Zmuidzinas J., Gao J.-R., eds, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 11453, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy X*. p. 1145302, doi:10.1117/12.2562331
- Zavala J. A., et al., 2015, *MNRAS*, 452, 1140
- Zavala J. A., et al., 2017, *MNRAS*, 464, 3369
- Zemcov M., Blain A., Halpern M., Levenson L., 2010, *ApJ*, 721, 424
- del Rio Astorga D., Dolz M. F., Fernández J., García J. D., 2017, *Concurrency and Computation: Practice and Experience*, 29, e4175