

Análisis de datos de mapas de velocidad de galaxias

José Miguel Ladino Méndez^{*}, Omar Moises Asto Rojas^{**} y Jennifer Grisales^{***}

^{*}Universidad Nacional de Colombia

^{**}Universidad Nacional de Ingeniería

^{***}Universidad Industrial de Santander

*Proyecto Final - Ciencia de Datos
LA-CoNGA Physics*

19 de mayo de 2021

Resumen

El presente trabajo muestra los resultados de la construcción de un modelo computacional en base a Python que determina el mapa de velocidad de una galaxia en rotación a través de una función de parámetros cinéticos. Se describen las etapas de la construcción del modelo en donde se realizó un análisis de las varianzas de dichos parámetros y de los modelos sintéticos originados a partir de datos reales. Este proyecto se realiza en el marco de la asignatura Ciencia de Datos del programa LA-CoNGA Physics.

Palabras clave: Parámetros cinéticos, curva de velocidad, análisis de varianzas, modelo sintético.

1. Introducción

El fenómeno de rotación de las galaxias fue descubierto en 1914 a partir de métodos espectroscópicos. No fue sino hasta 1918, que se obtuvo la primera curva de rotación sin embargo, el progreso en las mediciones de la cinemática de galaxias se logró tras el despliegue de grandes telescopios y el desarrollo de técnicas computacionales [1]. Algunos de los desarrollos que han permitido el avance en el estudio de mapas de velocidad radial en líneas de emisión fue por ejemplo el uso de interferometría Fabry-Pérot en las mediciones ópticas.

Los mapas de velocidad radial de galaxias actualmente son obtenidas a través de un gran número de líneas de emisión y absorción. El avance en el desarrollo de técnicas, y el desarrollo de modelos

^{*}e-mail: jmladinom@unal.edu.co

^{**}e-mail: oasto1310@gmail.com

^{***}e-mail: jennifer.grisales@saber.uis.edu.co

astrofísicos han llevado por ejemplo al descubrimiento de la materia oscura, y la sorprendente conclusión de que esta debe ser la influencia gravitatoria dominante en las partes exteriores de las galaxias [2].

El estudio de los mapas de velocidad de galaxias es un enfoque fundamental para comprender la cinemática y extraer información cosmológica como la distribución de masa y de materia oscura. Estudiar los mapas de velocidad implica construir un modelo computacional con algunas propiedades de la galaxia, por ejemplo, su radio, el ángulo de inclinación respecto al plano de visión de la tierra., entre otros

Mediante el presente trabajo buscamos replicar parcialmente la herramienta computacional construida para el estudio de mapas de galaxias *Kinometry*[3] mediante un script de python que construya un modelo de datos reales de mapas de velocidades con el mejor ajuste posible.

Actualmente el software robusto se encuentra disponible en IDL, y una versión muy preliminar en desarrollo en Python. IDL o *Interactive Data Language* es un lenguaje de programación para análisis y visualización de datos, usado por mucho tiempo en áreas del conocimiento como astronomía, medicina, entre otras. Pero encuentra en desuso debido al auge de nuevos lenguajes de programación como Python que por su sencillez, mejoran rápidamente y se optimizan por gracias a los aportes de la comunidad que lo usa.

El trabajo se enfocó en desarrollar un modelo de curvas de velocidad de galaxias para la obtención de sus parámetros cinéticos, utilizando la función de Epinat [4], el desarrollo del modelo se describe en la primera parte de la metodología para posteriormente pasar al análisis de dicho modelo a través de las varianzas de los parámetros cinéticos y modelos sintéticos.

2. Metodología

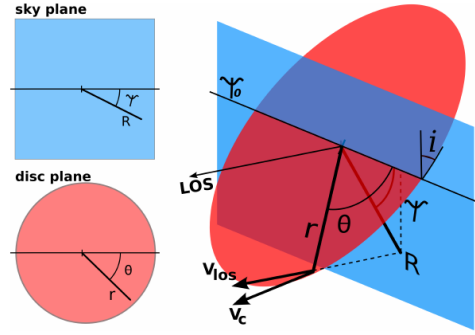
Para construir un software capaz de determinar los parámetros geométricos globales de discos en rotación de galaxias a partir de mapas de velocidad, seguiremos la siguiente metodología:

Construir una galaxia sintética con parámetros del mapa de velocidad definidos, eso se hace considerando una galaxia inclinada respecto al plano del cielo como se muestra en la figura 1,

En donde R corresponde a la medida del radio de la galaxia en el plano del cielo, r corresponde con el radio medio de la galaxia, i es el ángulo de inclinación de la galaxia respecto al plano del cielo de tal forma que,

$$r = R\alpha \quad (1)$$

Donde $\alpha = \sqrt{\cos^2(\Psi) + \sin^2(\Psi)/\cos^2(i)}$. Adicionalmente, la velocidad en la línea de observación v_{los} puede ser expresada como:

Figura 1: Gráfica de r

$$v_{los} = v_s + v_c \frac{\sin(i) \cos(\Psi - \Psi_0)}{\sqrt{\cos^2(\Psi - \Psi_0) + \sin^2(\Psi - \Psi_0) / \cos^2(i)}} \quad (2)$$

En esta expresión, v_s es la velocidad del sistema y v_c es la velocidad circular. De tal forma que para tener un modelo que calcule la velocidad en la línea de observación de la galaxia, es necesario crear datos artificiales que reproduzcan los observables de interés. De tal forma podremos hacer el proceso inverso posteriormente para hacer un fit a partir de datos.

Creemos una malla cuadrada que represente el vector R y Ψ en el plano del cielo. Luego, calculamos sus equivalentes r y α .

Finalmente para el cálculo de la velocidad en la línea de observación v_{los} , primero realizamos el cálculo de la velocidad circular en la galaxia v_c que se obtiene mediante la ecuación 3, dando como resultado el mapa de la figura 4:

$$v_c = V_t \frac{((r/R_0)^a)}{(1 + (r/R_0)^g)} \quad (3)$$

Donde finalmente, a través de la ecuación 2 podemos obtener el mapa de velocidad 5

A través de la secuencia lógica anterior y considerando las proyecciones geométricas relevantes hemos construido un modelo en python que consta de una función 'ModeloVlos' que permite construir un mapa de velocidad a partir de los parámetros:

- x_0 : x coordinate of the galaxy center
- y_0 : y coordinate of the galaxy center
- v_{sys} : velocity of the galaxy center
- i : zenith angle of inclination in the galaxy
- ϕ_{i0} : Angle for the projected disk

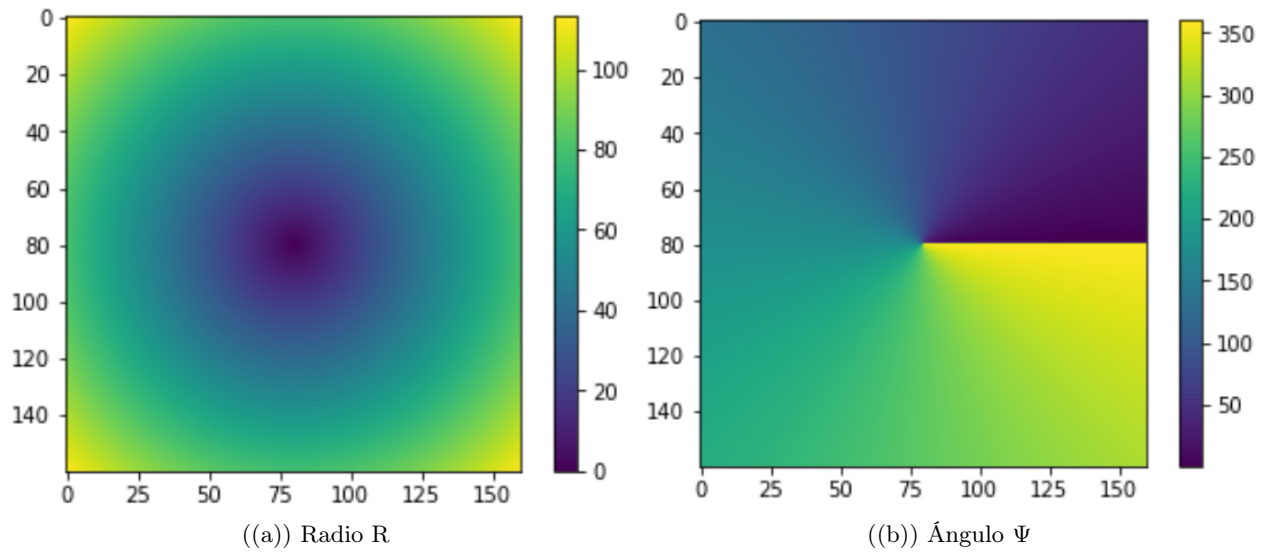


Figura 2: Matrices R y Ψ sintéticas que reproducen la observación directa de una galaxia en el plano del cielo.

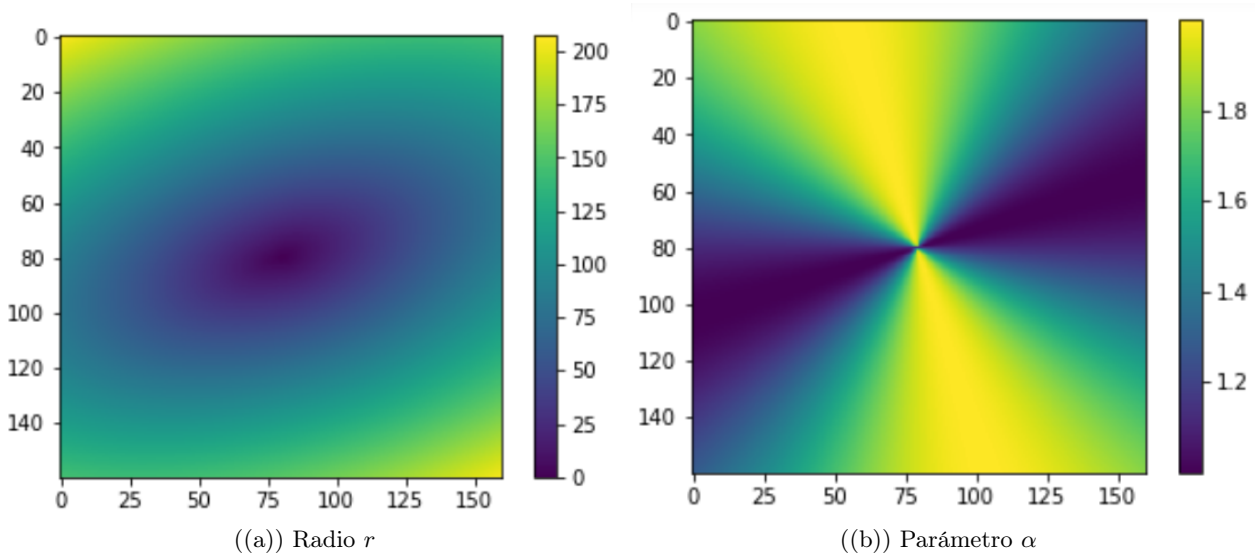
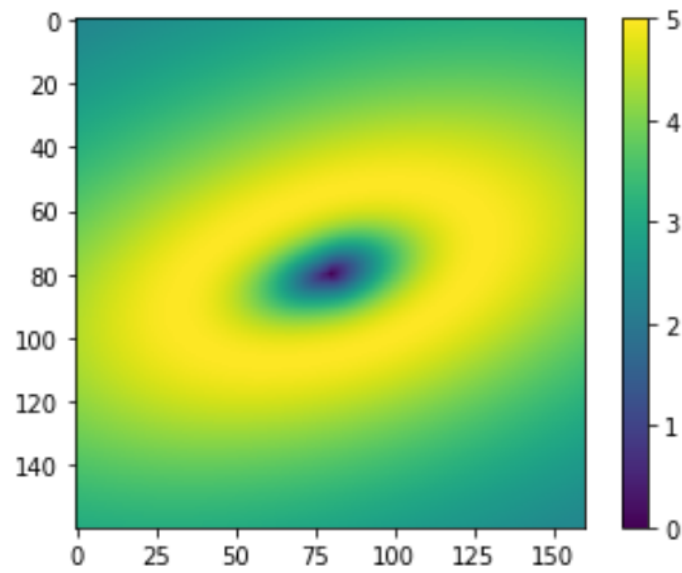
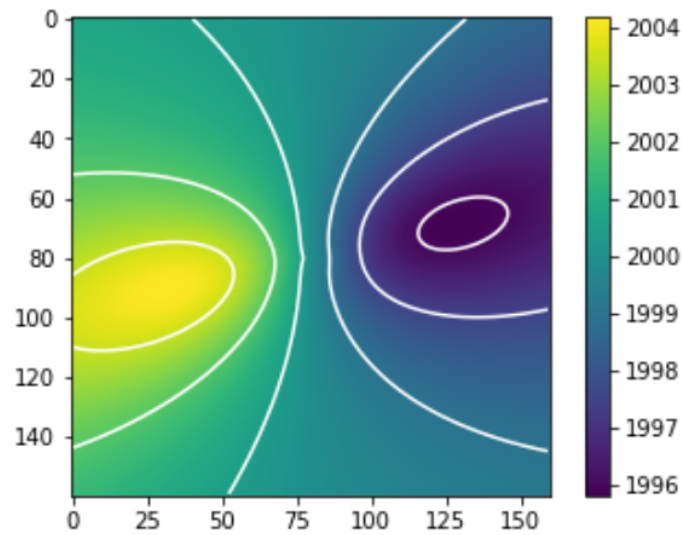


Figura 3: Datos sintéticos de las matrices r y α a partir de los datos observados (artificiales) de la figura 2

- V_t : Transversal velocity
- R_0 : Observed radius

Figura 4: Perfil de velocidad v_c Figura 5: Perfil de velocidad v_{los}

- a : Adimensional fit parameter
- g : Adimensional fit parameter

Con el objetivo de realizar un ajuste a los mapas de velocidad reales o sintéticos utilizando el

modelo de mapa de velocidad propuesto (figura 5), se procede a implementar el método de mínimos cuadrados, para minimizar las discrepancias entre los valores de velocidad de cada pixel entre el mapa de velocidad real o sintético y el mapa de velocidad modelo. La función a usar es *leastsq* de *Scipy*, la cuál se aplicará a la función *error* definida entre la diferencia de valores de velocidad de los datos reales o sintéticos y el modelo, para determinar los valores de los 9 parámetros del modelo que mejor se adapten al mapa de velocidad real o sintético. Como lo muestra la siguiente sección de resultados, también se ha realizado una optimización de los valores de los parámetros que puede encontrar la función *leastsq*, haciendo uso de la matriz de covarianza *Cov* con la cuál se puede calcular un estimado de la varianza *Var* para cada parámetro entre los datos analizados en la función error, así:

$$Var = Tr(Cov) \left(\frac{z_{fit} - z}{n - p} \right), \quad (4)$$

donde $Tr(Cov)$ es la suma de los elementos de la diagonal de la matriz de covarianza de cada parámetro, z_{fit} es los valores de velocidad usando los parámetros ajustados, z son los valores de velocidad del conjunto de datos a analizar real o sintético, n es el número de datos en z y p es el número de parámetros a ajustar.

3. Análisis y Resultados

3.1. Primer modelo

Para mostrar la ejecución del software que modela los mapas de velocidad de las galaxias, a continuación se muestra la aplicación del modelo mediante mínimos cuadrados, en el cual se encuentra la tupla de 9 parámetros que mejor se ajusta a los datos de velocidad reales o sintéticos introducidos.

El primer conjunto de datos de velocidad a analizar son simulados y se encuentran en el archivo "Dataset .txt", como se aprecia en la figura 6, el primer recuadro corresponde al mapa de velocidades galáctico proveniente de la aplicación del modelo, el segundo proviene de los datos simulados del archivo y el tercer recuadro describe el residuo resultante de la diferencia entre lo obtenido en el modelo y los datos simulados.

Como lo muestran las barras de color en la figura 6, el modelo identifica satisfactoriamente la orientación de la velocidad de rotación de la galaxia, de tal manera que por un lado en la parte negra la galaxia tendrá la misma velocidad que del lado contrario (zona amarilla) pero en dirección contraria. Adicionalmente, en el recuadro de la resta es importante recalcar que mientras más cercano al pixel (0) este cada pixel, el mapa de velocidad del modelo tendrá un comportamiento más similar al real o sintético.

El segundo conjunto de datos de velocidad a analizar son de mediciones reales, provenientes del archivo "Dataset.fits", , como se ve en la figura 7, el modelo nuevamente se aproxima satisfactoriamente mostrando su correspondiente mapa de velocidad de residuos en el tercer recuadro.

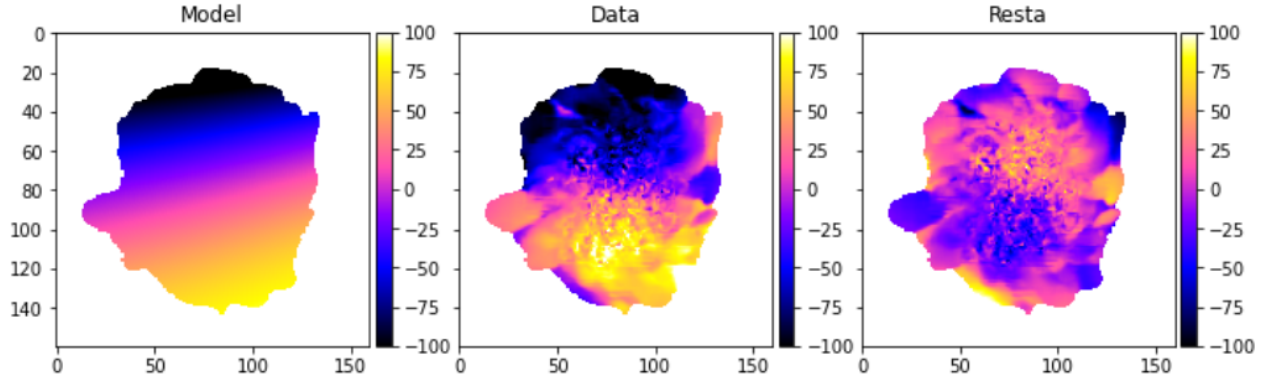


Figura 6: Aplicación del modelo a datos simulados de "Dataset .txt", usando como valores iniciales en *leastsq*: $x_0 = 0$, $y_0 = 0$, $v_{sys} = 1$, $i = 1^\circ$, $\phi_0 = 1^\circ$, $v_t = 1$, $R_0 = 1$, $a = 1$, $g = 1$.

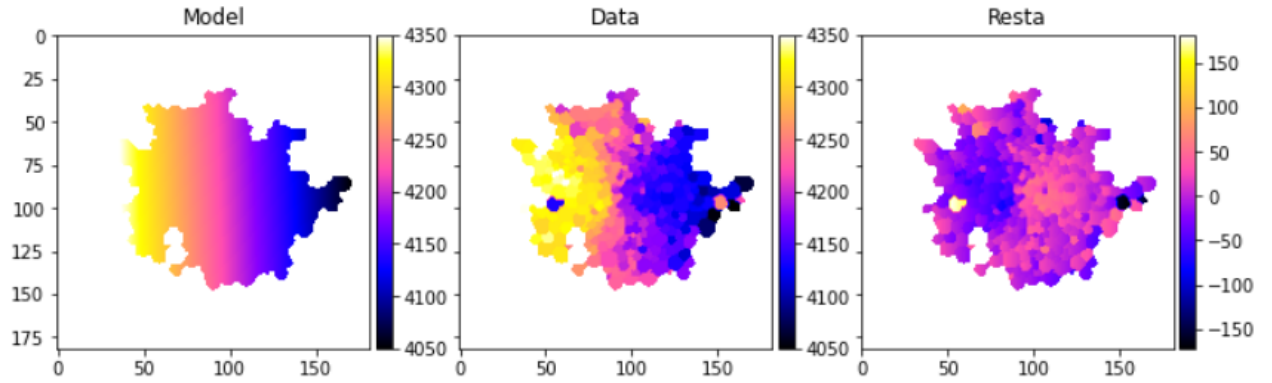


Figura 7: Aplicación del modelo a datos reales de "Dataset.fits", usando como valores iniciales en *leastsq*: $x_0 = 0$, $y_0 = 0$, $v_{sys} = 1$, $i = 1^\circ$, $\phi_0 = 1^\circ$, $v_t = 1^\circ$, $R_0 = 1$, $a = 1$, $g = 1$.

Como se puede apreciar, en la aplicación de los dos modelos anteriores la aproximación del mapa de velocidades mantiene un patrón de cambio general en todo el mapa sin detallar los cambios abruptos locales, además de que el modelo se ajusta de una manera lineal con respecto a la simetría de los datos reales (horizontalmente en el primer ajuste y verticalmente en el segundo).

3.2. Modelo optimizado a partir de las varianzas de los parámetros

Por otro lado, para optimizar estos dos anteriores resultados de la aplicación del modelo, se modifica el software analizando las varianzas de la generación aleatoria de tuplas posibles de parámetros en función de una varianza entre tuplas aleatorias que definirá la generación de las posibles tuplas de parámetros iniciales y el número de tuplas de parámetros a generar, dónde finalmente se devolverá en el modelo, aquel mapa de velocidad con la tupla de parámetros que tengan las varianzas más pequeñas obtenidas. En la figura 8, se muestra la aplicación del modelo

optimizado para el primer conjunto de datos simulados:

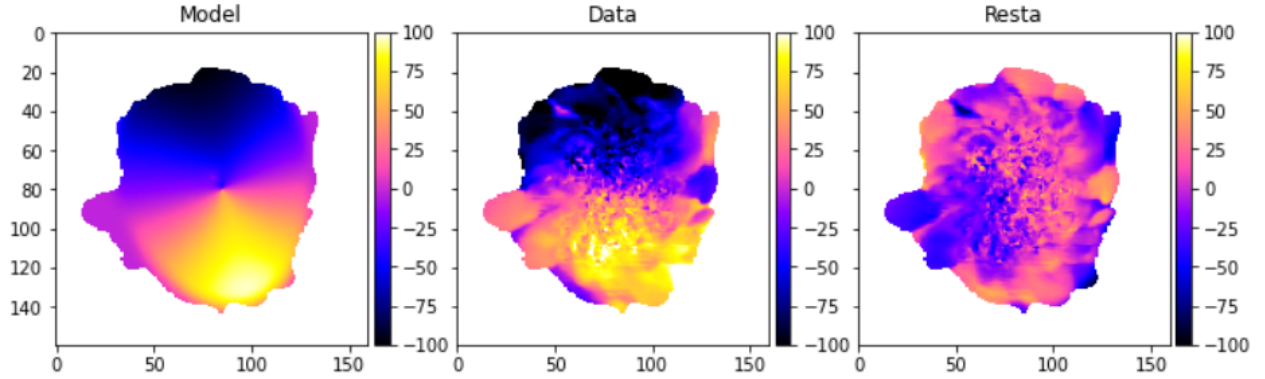


Figura 8: Aplicación del modelo de parámetros optimizados en "Dataset.txt", usando como valores iniciales en *leastsq*: $x_0 = 0$, $y_0 = 0$, $v_{sys} = 1$, $i = 1^\circ$, $\phi_0 = 1^\circ$, $v_t = 1$, $R_0 = 1$, $a = 1$, $g = 1$.

El modelo optimizado obtenido a simple vista, genera un mapa de velocidades galáctico el cual a diferencia del anterior que solo es a partir de una tupla de parámetros iniciales (el de la figura 6), presenta un ajuste radial y polar con respecto al centro del mapa de velocidades de datos simulados, generando un mapa más confiable en este sentido, sin embargo los cambios abruptos locales no se detectan con mayor detalle. En la figura 8, se aprecian estos mismos efectos del modelo optimizado para el conjunto de datos del mapa de velocidades real.

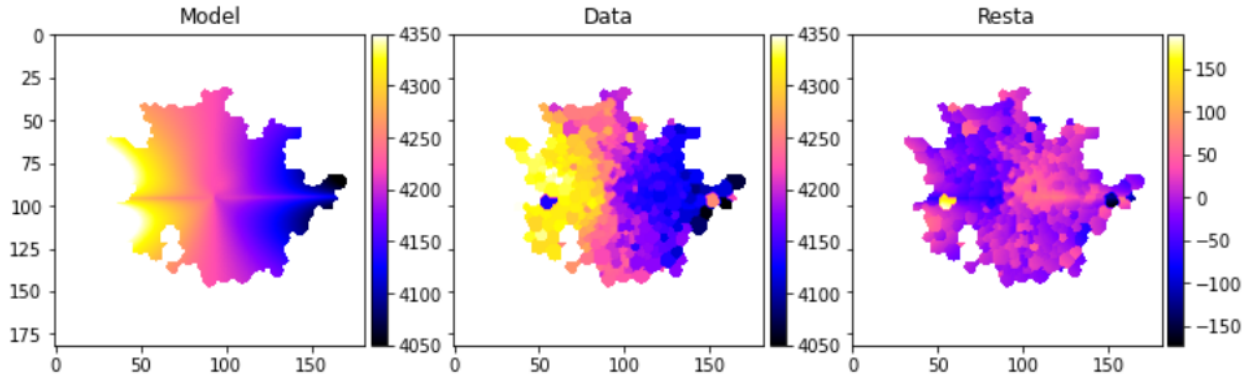


Figura 9: Aplicación del modelo de parámetros optimizados en "Dataset.fits", usando como valores iniciales en *leastsq*: $x_0 = 0$, $y_0 = 0$, $v_{sys} = 1$, $i = 1^\circ$, $\phi_0 = 1^\circ$, $v_t = 1$, $R_0 = 1$, $a = 1$, $g = 1$.

3.3. Análisis del primer modelo y del modelo optimizado a partir de las varianzas

En orden de estudiar el residuo de los modelos, como lo muestran las figuras 10 y 11, se puede identificar cuán buena es la aproximación del modelo del software realizando una gráfica de la distribución de restas en forma de histograma.

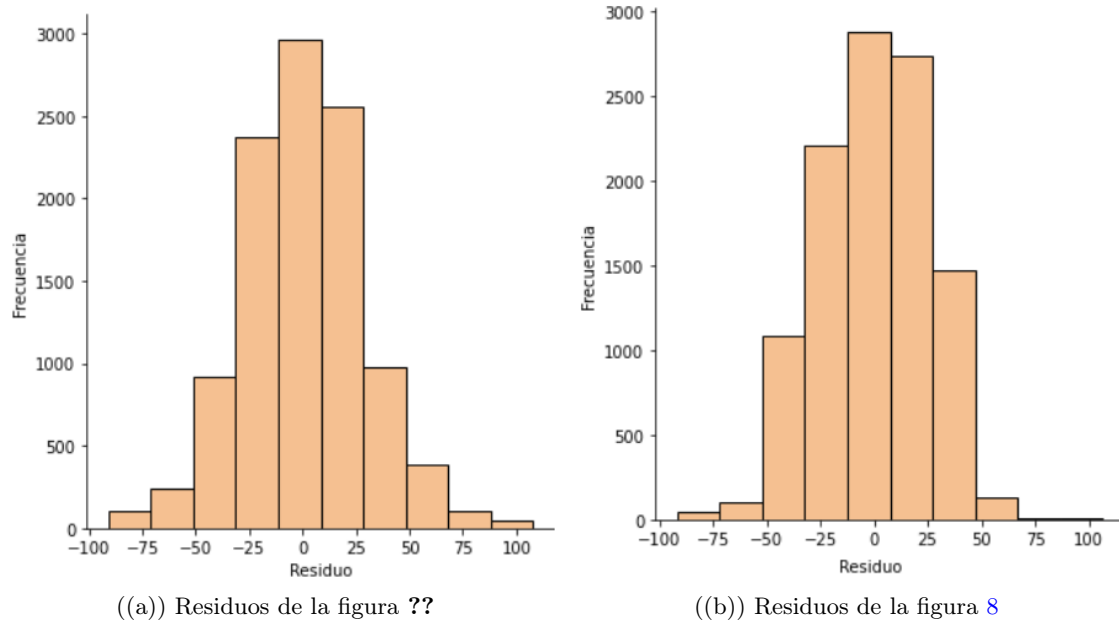


Figura 10: Histogramas de las restas de los mapas de velocidad para el conjunto de datos simulados

Evidentemente, los ajustes realizados tienen un comportamiento de una distribución normal centrada en el cero, tanto para el mapa de velocidad de la galaxia de los datos simulados como la de datos reales.

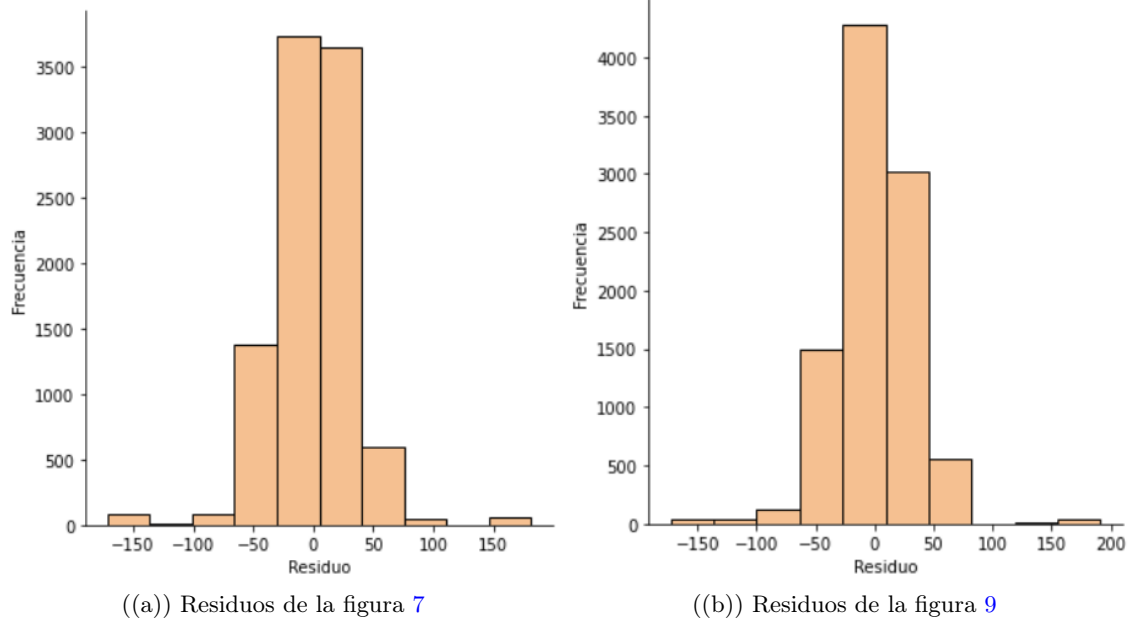


Figura 11: Histogramas de las restas de los mapas de velocidad para el conjunto de datos reales

La escogencia de los parámetros iniciales del mapa de velocidades del modelo presenta resultados favorables a simple vista, sin embargo no deja de ser ambiguo cuán bueno ha sido el ajuste. Para garantizar que los parámetros encontrados en el modelo optimizado a partir de varias tuplas de parámetros se ajustan mejor que con el modelo inicial solo con una tupla de parámetros, se introduce la variable llamada como residuo total s , la cual viene definida como

$$s = \sum_{i=1}^n |res_i|, \quad (5)$$

es decir, s es la suma de los valores absolutos de res_i , el valor del residuo para cada pixel i . Así, si $s_1 < s_2$ significa entonces, que el modelo que genera un residuo total s_1 va a ajustar mejor los parámetros al conjunto de datos reales o sintéticos a analizar. Teniendo en cuenta esto, s toma los siguientes valores para los 4 modelos anteriormente aplicados:

	Modelos	s
Conjunto de datos simulados	figura 10(a)	235910.31
	figura 10(b)	238128.94
Conjunto de datos reales	figura 11(a)	218374.98
	figura 11(b)	230365.08

Cuadro 1: Valores de los residuos totales s

En el cuadro 1, se hace evidente entonces que tanto en el conjunto de datos simulados como en el de datos reales, el modelo optimizado a partir de varias tuplas de parámetros da un mejor resultado de s , brindando un mejor ajuste a los parámetros encontrados.

3.4. Análisis del modelo a través de la generación de modelos sintéticos

Para determinar la confiabilidad de el modelo estudiado se trabajó con modelos sintéticos, estos son generados en función a parámetros de velocidad de galaxias verdaderas, una vez obtenidos estos modelos se procedió a perturbarlos con una muestra representativa de restos obtenidas a partir del ajuste de datos reales obtenidos de G. Bekiaris [5] y de los datos de varios estudios de galaxias de PerotFabry.

A continuación se procederá a describir los pasos realizados:

3.4.1. Obtención de los parámetros cinéticos de galaxias reales

Para la obtención de los parámetros se acudió al trabajo de B. Epinat [6] y G. Bekiaris [5], en el último paper se encontró una tabla con los principales parámetros cinéticos de galaxias, además de una tabla en donde mostraban el espacio de parámetros mostrados.

Los datos obtenidos de la fuente y completados con la consideración del espacio de parámetros, se reunieron en los histogramas mostrados en el anexo 15.

Analizando los datos obtenidos, se tuvo que ajustar los parámetros reales que se tenía (PA, i, V_{sys}) con sus respectivas matrices de velocidad obtenidas de la fuente de datos para así obtener los parámetros faltantes (a, g, R_0). Estos últimos parámetros obtenidos se muestran en las figuras: 13(a), 13(b) y 13(c).

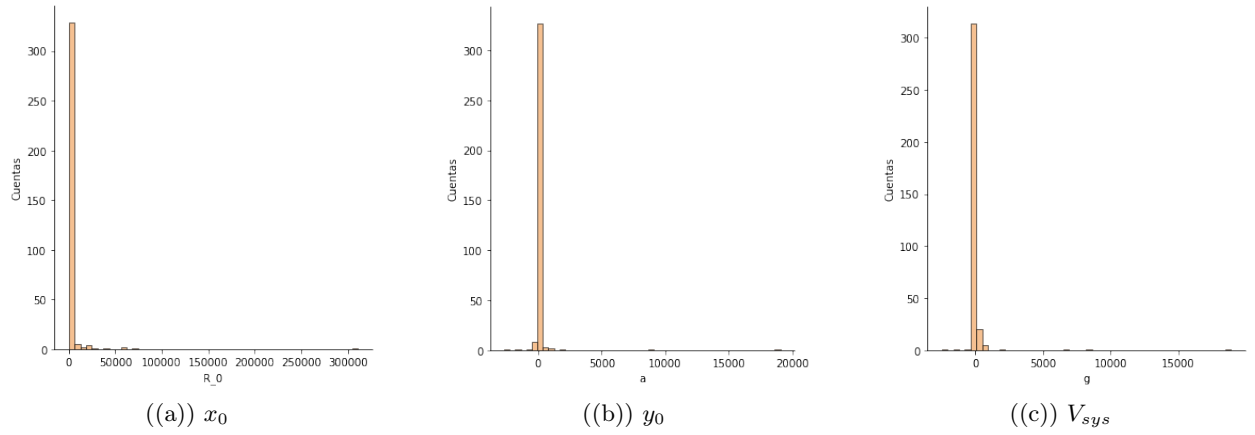
Los datos obtenidos después del ajuste, tienen mejor sentido en relación al perfil de velocidad de la función de Epinat usado en el trabajo de Epinat[4]. El perfil de velocidad se puede ver en la figura 13

3.4.2. Obtención de los modelos sintéticos

Una vez obtenido el conjunto de parámetros cinéticos reales, se porcedió a generar 1000 modelos sintéticos. Se tomó en consideración la elección aleatoria de los parámetros para así obtener matrices

Parameter	Units	Range (min)	Range (max)
$N_{\text{data,xy}}$	Pixels	32	96
PSF_{FWHM}	Pixels	4.0 (fixed)	4.0 (fixed)
LSF_{FWHM}	km s^{-1}	20	60
V_{sys}	km s^{-1}	0 (fixed)	0 (fixed)
x_0	Pixels	$0.4 \times N_{\text{data,xy}}$	$0.6 \times N_{\text{data,xy}}$
y_0	Pixels	$0.4 \times N_{\text{data,xy}}$	$0.6 \times N_{\text{data,xy}}$
PA	Degrees	0	45
i	Degrees	5	85
r_t	Pixels	$0.5 \times \text{PSF}_{\text{FWHM}}$	$2.0 \times \text{PSF}_{\text{FWHM}}$
r_0	Pixels	$1.0 \times \text{PSF}_{\text{FWHM}}$	$3.0 \times \text{PSF}_{\text{FWHM}}$
V_t	km s^{-1}	50	300
σ_{model}	km s^{-1}	20	100

Figura 12: Espacio de parámetros usados para generar los datos [5]



de velocidades lo más cercanas a la realidad.

3.4.3. Obtención de la muestra de restos

Para la obtención de la muestra de los restos se acudió a [PerotFabry](#), en ahí se descargó un conjunto de matrices de velocidades para ajustarlas con el modelo realizado. Una vez ajustadas las matrices, todos los restos se agruparon en una muestra de restos.

3.4.4. Obtención de modelos perturbados y análisis

Una vez obtenida la muestra de restos y los modelos sintéticos se procede a tomar un modelo sintético y sumarle un error de la muestra de errores para obtener un modelo perturbado. Una vez obtenido el modelo perturbado se ajusta con la función de Epinat estudiada, los nuevos parámetros

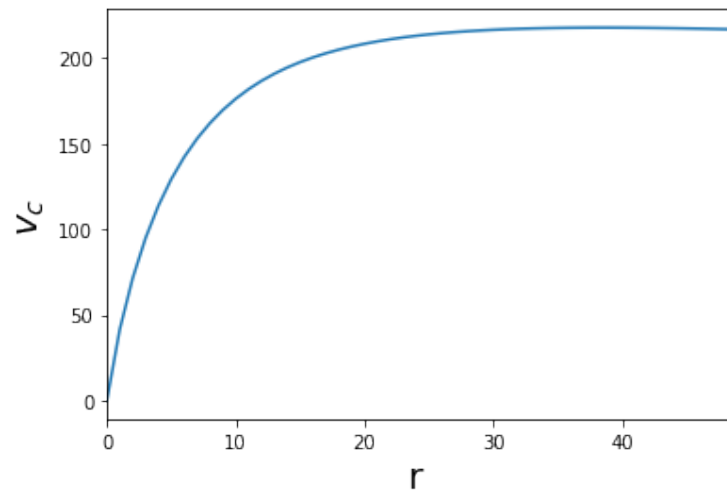


Figura 13: Perfil de velocidad de la función de Epinat[4]

ajustados se igualan con los parámetros del modelo sintético sin perturbar. Se nota que si hay una diferencia es debido a la suma del resto, hay veces que se obtiene los mismos valores de parámetros y es porque hay restos que son muy pequeños. Todas las diferencias de parámetros obtenidas se suman y se dividen entre el total de las veces que los restos característicos fueron tomados, para así obtener una incertidumbre global relacionada a los generación de modelos sintéticos obtenidos a partir de datos reales. Las incertidumbres se pueden ver en la siguiente tabla¹⁴.

Parámetros	Incertidumbre
Δx_0	0.01302091
Δy_0	0.08733882
ΔV_{sys}	0.05042536
Δi	1.64693889
ΔPA	0.24508438
ΔV_t	0.4941802
ΔR_0	1.33315411
Δa	0.00404823
Δg	0.05857176

Figura 14: Tabla de incertidumbres de cada parámetro.

4. Conclusiones

- Se desarrollo un modelo de velocidades de galaxias en PYTHON que obtiene los parámetros cinéticos de estas además el uso del script desarrollado es sencillo y puede ser obtenido fácilmente en el siguiente [repositorio](#).
- El análisis de la determinación de parámetros a partir de las varianzas, optimiza el software satisfactoriamente, brindando un mejor ajuste de los mapas de galaxias simulados y reales en función del número de tuplas de parámetros iniciales a considerar y su dispersión.
- En comparación con el análisis a través de las varianzas de los parámetros, en los modelos sintéticos toman en cuenta datos reales de diferentes galaxias.
- La incertidumbre de los parámetros a través del análisis de modelos sintéticos depende de la cantidad de galaxias estudiadas.

5. Anexo

5.1. Histogramas de datos

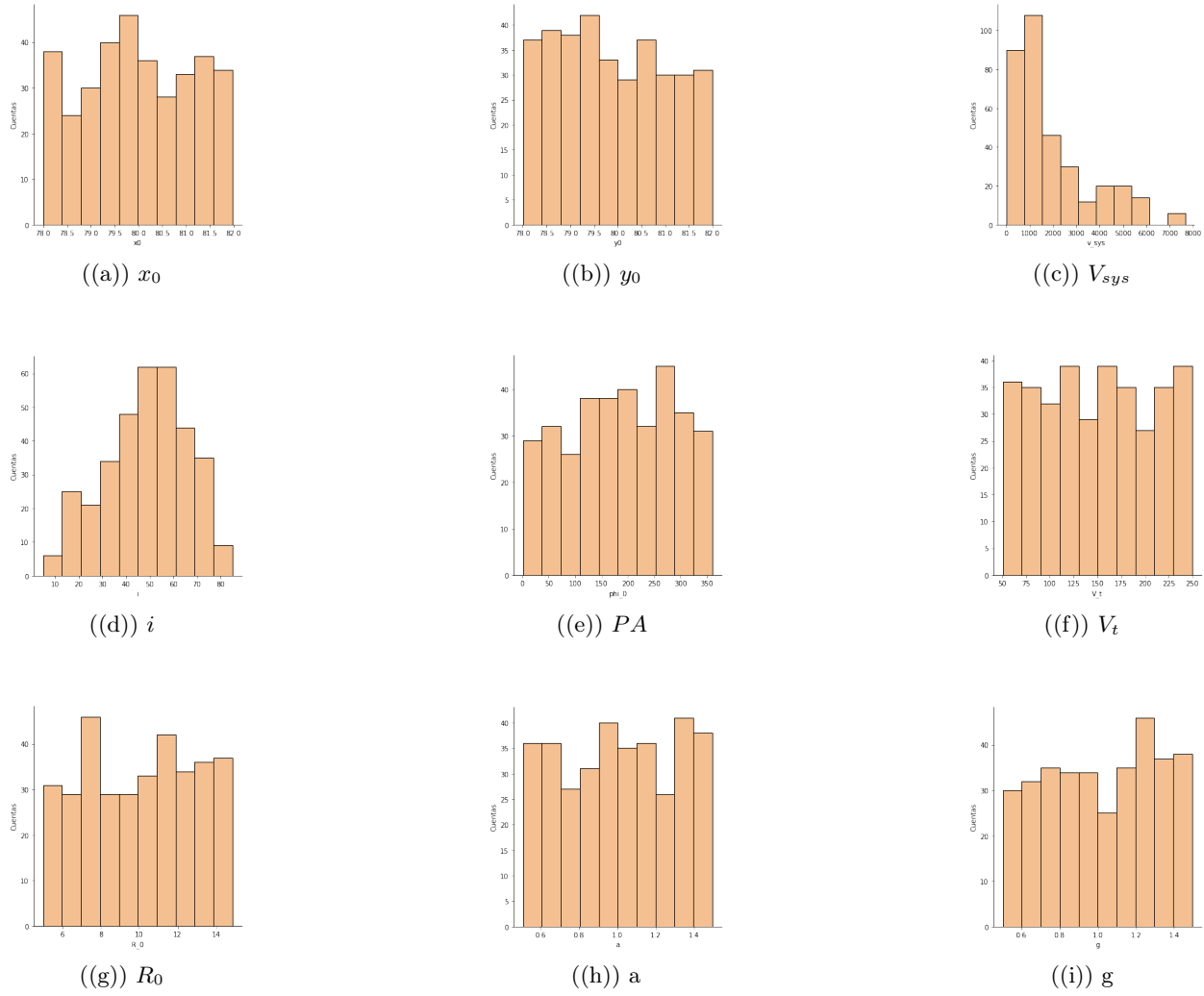


Figura 15: Parámetros cinéticos; V_{sys} , PA , i : fueron obtenidos deñ paper de G. Bekiaris [5]

Referencias

- [1] J.E. Beckman, A. Zurita, and Juan Vega Beltrán. Kinematic measurements of gas and stars in spiral galaxies. *Lecture Notes and Essays in Astrophysics*, 1:43–62, 08 2004.
- [2] P. J. Teuben. Velocity Fields of Disk Galaxies. In E. Athanassoula, A. Bosma, and R. Mujica, editors, *Disks of Galaxies: Kinematics, Dynamics and Perturbations*, volume 275 of *Astronomical Society of the Pacific Conference Series*, pages 217–228, December 2002.
- [3] D. Krajnović, M. Cappellari, P. T. de Zeeuw, and Y. Copin. Kinemetry: a generalization of photometry to the higher moments of the line-of-sight velocity distribution. *MNRAS*.
- [4] B Epinat, P Amram, and M Marcelin. Ghasp: an $\text{H}\alpha$ kinematic survey of 203 spiral and irregular galaxies—vii. revisiting the analysis of $\text{H}\alpha$ data cubes for 97 galaxies. *Monthly Notices of the Royal Astronomical Society*, 390(2):466–504, 2008.
- [5] Georgios Bekiaris, Karl Glazebrook, Christopher J Fluke, and Roberto Abraham. Kinematic modelling of disc galaxies using graphics processing units. *Monthly Notices of the Royal Astronomical Society*, 455(1):754–784, 2016.
- [6] B Epinat, P Amram, C Balkowski, and M Marcelin. Evidence for strong dynamical evolution in disc galaxies through the last 11 gyr. ghasp viii—a local reference sample of rotating disc galaxies for high-redshift studies. *Monthly Notices of the Royal Astronomical Society*, 401(4):2113–2147, 2010.