

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS FÍSICAS

DEPARTAMENTO DE FÍSICA DE LA TIERRA Y ASTROFÍSICA



TRABAJO DE FIN DE GRADO

Código de TFG: FTA14

Relaciones estructurales de galaxias remotas a partir del catálogo
CANDELS

Structural Relations of remote galaxies from the CANDELS catalogue

Supervisor: Jesús Gallego Maestro

Rafael Navarro Carrera

Grado en Física

Curso académico 2020-2021

Convocatoria de Junio

Índice

| | |
|----------------------------------------------------------|-----------|
| 1. Resumen | 2 |
| 2. La exploración CANDELS | 3 |
| 3. Caracterización de la Muestra | 4 |
| 4. Relaciones Estructurales | 7 |
| 4.1. Correlación entre parámetros | 8 |
| 4.2. Correlaciones encontradas en la muestra | 9 |
| 4.3. Análisis detallado | 14 |
| 5. Otras Consideraciones | 16 |
| 5.1. Exploración de relaciones no binarias | 16 |
| 5.2. Diagrama color - magnitud | 16 |
| 5.3. Evolución de las relaciones estructurales | 19 |
| 6. Conclusiones Finales | 20 |

1. Resumen

Se caracterizará una muestra de galaxias con objetivo de determinar posibles relaciones estructurales, es decir, relaciones que existen entre las propiedades físicas medidas o derivadas de observaciones astronómicas.

Gracias a diversos proyectos relativamente recientes como CANDELS [Grogin et al., 2011], [Koekemoer et al., 2011]; se ha construido un mapa profundo de diferentes zonas. Se discutirá una muestra de galaxias con redshift comprendido entre $z = 0$ y $z = 0,5$. Se caracterizará la muestra estudiada y se determinarán las posibles relaciones estructurales entre todos los parámetros medidos empleando una matriz de correlación.

El presente proyecto se puede enmarcar dentro de un esquema más amplio, compuesto por otros dos más, los cuales abarcan diferentes rangos de redshift de la muestra de galaxias. Así se puede entender más en profundidad si las relaciones estructurales de las galaxias se han modificado según estas evolucionan, y de qué modo lo han hecho.

Abstract

The underlying principle behind this project is to characterize a galaxy sample having in mind the final objective of finding possible structural relations between them, that is, relationships between different physical properties measured or derived from astronomical observations.

Using CANDELS [Grogin et al., 2011] [Koekemoer et al., 2011], a relatively deep map of different areas of the sky has been obtained. The parameter used to select the galaxy sample is redshift, using galaxies with values that fall between 0 and 0.5, the smallest range of all the catalog.

This project can be conceived as part of one with much broader scope, sharing similar goals with other two that study a galaxy sample of different nature. Allowing to compare galaxies with low, medium and high redshift regimes; thus being able to observe if structural relations have evolved with time, and in what form have they evolved.

As a final result, the sample will be completely characterised, and all of the structural relations found within galaxies will be determined taking advantage of the correlation matrix of all the studied parameters.

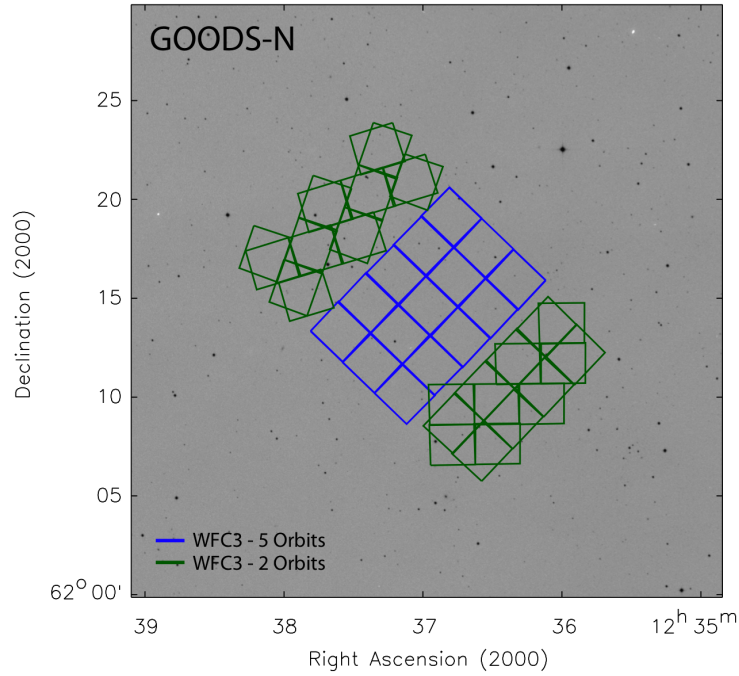


Figura 1: Mosaicos empleados en la zona GOODS N

2. La exploración CANDELS

La construcción de catálogos de objetos astronómicos no es una idea especialmente novedosa ni reciente, pero la realización de uno comprendido por objetos extragalácticos no ha sido posible hasta hace poco. Estas grandes bases de datos en varias longitudes de onda, que estudian el universo a diferentes redshifts, y las cuales alcanzan una profundidad notable permiten estudiar la naturaleza de objetos extragalácticos tanto en el presente como en el pasado.

El estudio CANDELS se trata de un compendio de observaciones realizadas entre el año 2011 y 2013 empleando los instrumentos WFC3 y ACS del HST. Los campos se construyen a base de la combinación de mosaicos durante diferentes órbitas.

En concreto, se hará uso del catálogo CANDELS GOODS-N recientemente compilado [Barro et al., 2019] que combina datos de diferentes instrumentos pertenecientes a Hubble Space Telescope (HST), Gran Telescopio de Canarias (GTC), Spitzer, Herschel, etc. Proporcionando estimaciones de parámetros físicos de las galaxias a partir de observaciones fotométricas y espectroscópicas.

Cabe destacar que la serie de parámetros que se emplearán del catálogo compilado por [Barro et al., 2019] se han obtenido desde el repositorio accesible desde la web del proyecto CANDELS: <https://archive.stsci.edu/missions/hlsp/candels/goods-n/catalogs/v1/>.

El campo GOODS N ha sido ampliamente analizado en gran cantidad de estudios científicos con objeto de estudiar poblaciones galácticas a alto redshift. Es conocido como Hubble GOODS North field (GOODS-N), acrónimo de *Great Observatories Origins Deep Survey* [Giavalisco et al., 2004]. Se encuentra localizado en $R.A. = 189,228621$, $Dec. = +62,238572(J2000)$ y tiene un área de 30 arcmin^2 .

| parámetro | descripción | parámetro | descripción |
|-----------|-------------------|--------------|-------------------------------|
| z[1] | redshift | lsfr[5] | $\log(SFR/(M_{sun}/yr))$ |
| ltau[2] | $\log(\tau/yr)$ | Mu,Mv,Mj[6] | magnitudes absolutas [mag] |
| lage[3] | $\log(edad/yr)$ | Av[7] | extinción en la banda V [mag] |
| lmass[4] | $\log(M/M_{sun})$ | petro_rad[8] | radio petrosiano [px] |

Tabla 1: Parámetros estudiados en la muestra. [1] Redshift (z), definido como $z = \frac{\lambda_{recib} - \lambda_{emit}}{\lambda_{emit}}$, relacionado con la distancia a la que se está observando. [2] Logaritmo de τ en relación a años, indicador de la evolución de la galaxia [Carnall et al., 2019]. [3] Logaritmo de la edad en años. [4] Logaritmo de la masa estelar de la galaxia en relación a la masa solar. [5] Logaritmo de la tasa de formación estelar en masas solares por año, indicador del ritmo al que se forman estrellas en la galaxia. [6] Magnitudes absolutas, indican la luminosidad que emite la galaxia. [7] Radio aparente petrosiano [Petrosian, 1976], empleado para determinar el tamaño de la galaxia. Para estudiar correlaciones será de mayor utilidad el tamaño físico en kpc de las galaxias.

3. Caracterización de la Muestra

Los parámetros que se analizarán se detallan en la Tabla 1.

La muestra de galaxias que comprende el catálogo sigue la distribución que se muestra en la Figura 2, donde se clasifica la muestra según los valores de z o redshift que tienen sus componentes. Se presentan también los estadísticos que permiten comprender las propiedades globales de la muestra en los mismos gráficos.

Como ya se ha mencionado, en este estudio se limitará el rango de redshift a $0 \leq z \leq 0,5$. La distribución resultante se puede apreciar con más detalle en la parte derecha de la Figura 2.

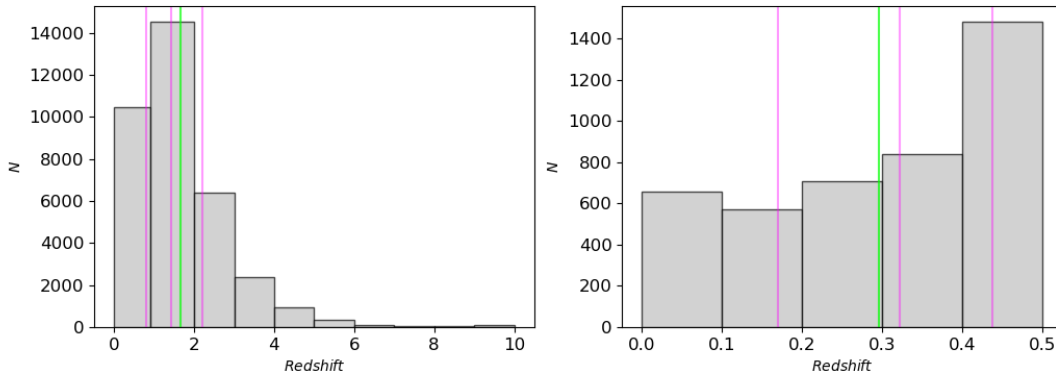


Figura 2: Distribución de la muestra según su redshift (z). La línea verde representa la media aritmética. Las líneas moradas representan los cuartiles.

El rango escogido no es especialmente numeroso, ya que cuenta con 4259 componentes, representando un 12 % de las 35445 galaxias comprendidas en el catálogo. De ahora en adelante, se da por supuesto que todos los análisis posteriores estarán dentro de ese rango de redshift.

A continuación, se procederá a analizar los parámetros estudiados, que se corresponden esencialmente a los mencionados en la Tabla 1. Para ello, se construye un histograma acompañado de media

| parámetro | z | Mu (mag) | Mv (mag) | Mj (mag) | U-V (mag) | petro_rad (px) |
|---------------------------|------|----------|----------|----------|-----------|----------------|
| media | 0.30 | -14.6 | -15.5 | -16.1 | 0.90 | 4.92 |
| desv. std [1] | 0.15 | 8.3 | 8.3 | 8.2 | 0.55 | 1.50 |
| rango. intercuart. [2] | 0.27 | 3.5 | 3.6 | 3.6 | 0.55 | 2.47 |
| min | 0.00 | - | - | - | -4.05 | 3.50 |
| 25 % [3] | 0.17 | -16.0 | -16.9 | -17.4 | 0.58 | 3.50 |
| 50 % [4] | 0.32 | -14.1 | -14.9 | -15.4 | 0.83 | 4.45 |
| 75 % [5] | 0.44 | -12.5 | -13.3 | -13.8 | 1.13 | 5.97 |
| max | 0.50 | 2.0 | 0.6 | -1.6 | 6.26 | 10.73 |

Tabla 2: Estadística para los parámetros observacionales de la muestra. Se toma como incertidumbre la última cifra significativa de las indicadas en la tabla. [1]: Desviación Estándar de los datos para cada variable, se puede entender como la dispersión de los datos. [2]: Rango Intercuartílico, otro parámetro que permite estudiar la dispersión de los datos. [3],[4],[5]: Cuartiles, valores de la variable con probabilidad acumulada equivalente a la mostrada.

| parámetro | z | $\log(M/M_{sun})$ | $\log(SFR/(M_{sun}/yr))$ | Av (mag) [6] | $\log(\tau/yr)$ | $\log(age/yr)$ |
|---------------------------|------|-------------------|--------------------------|--------------|-----------------|----------------|
| mean | 0.30 | 7.31 | -2.20 | 0.26 | 8.78 | 8.91 |
| desv. std [1] | 0.15 | 1.52 | 1.80 | 0.53 | 0.50 | 0.47 |
| rango. intercuart. [2] | 0.27 | 1.54 | 1.38 | 0.20 | 0.20 | 0.50 |
| min | 0.00 | -1.41 | -19.39 | 0.00 | 8.50 | 7.60 |
| 25 % [3] | 0.17 | 6.63 | -2.66 | 0.00 | 8.50 | 8.70 |
| 50 % [4] | 0.32 | 7.35 | -2.00 | 0.00 | 8.50 | 9.00 |
| 75 % [5] | 0.44 | 8.17 | -1.28 | 0.20 | 8.70 | 9.20 |
| max | 0.50 | 12.30 | 2.37 | 4.00 | 10.10 | 10.10 |

Tabla 3: Estadística para los parámetros derivados de la muestra. Se toma como incertidumbre como la última cifra significativa de las indicadas en la tabla. [1],[2],[3],[4],[5]: Consultar en Tabla 2. [6]: Magnitudes extinguidas en la banda V.

(verde) y cuartiles (magenta) en las Figuras 4 y 5. Asimismo, se detallan los principales parámetros estadísticos en las Tablas 2 y 3.

Redshift

La distribución de valores de redshift a lo largo del intervalo seleccionado $0 \leq z \leq 0,5$ se muestra en la Figura 4. Es apreciable una tendencia creciente hacia valores de redshift altos, y un pico en el valor $z = 0,5$.

La tendencia creciente se puede comprobar extendiendo el análisis a valores de z más altos, tal y como se aprecia en la Figura 3. Se puede plantear que esa mayor abundancia puede ser debida a las características del campo elegido, a un efecto de selección que favorece la detección de un mayor número de candidatos a este redshift, u otro efecto desconocido.

Será interesante estudiar otros campos cubiertos por CANDELS para descartar que el fenómeno sea particular del campo seleccionado en este estudio. Se emplea para ello el campo COSMOS

[Nayyeri et al., 2017], catálogo compilado por *Nayyeri et al.* En la Figura 3 se puede apreciar que la distribución de galaxias es diferente a la encontrada en GOODS-N.

Por tanto, la mayor abundancia de galaxias en torno a $z = 0,5$ se trata de una cuestión exclusiva del campo aquí estudiado.

Magnitudes Absolutas y Color

Las magnitudes absolutas siguen una distribución que recuerda a una normal, centrada en un valor promedio. Este resultado es el esperado, ya que las galaxias tienen una luminosidad promedio, si bien existen otras con valores mayores o menores (aunque en menor número).

Al ser el color la diferencia de un par de magnitudes (en este caso U-V), existirá un valor promedio con mayor frecuencia que corresponde a la resta de los valores promedios de sendas luminosidades. La distribución obtenida es, por tanto, razonable.

El resto de distribuciones se pueden interpretar del mismo modo, por lo que algunos no requieren mención individual. A excepción del logaritmo de las edades y del parámetro τ .

Es notable la presencia de valores extremos en el color en los datos de la muestra, que no son físicamente razonables, y que representan una pequeña porción de la muestra. Teniendo en cuenta la Tabla 2, se descartarán valores para el color que escapen al intervalo del 95 % con respecto a la media muestral. Se consideran, a efectos de este análisis, datos sin interés.

Edad y τ

La edad y τ [Carnall et al., 2019] son parámetros de carácter temporal. El segundo de ellos forma parte de los modelos de formación estelar [Carnall et al., 2019].

Las distribuciones de estos parámetros son particulares en tanto que están concentradas en torno a unos valores muy concretos (como se puede apreciar teniendo en cuenta el valor de sus cuartiles).

Esto es razonable ya que la edad de la mayoría de galaxias está localizada en un valor concreto, y son pocas las que se formaron mucho antes o después de ese momento. Además, la obtención de estos parámetros sigue métodos que se basan en la comparación con catálogos, por lo que los resultados tienen una incertidumbre importante, y solo pueden clasificarse en unos valores concretos. Se obtiene que la edad promedio de la muestra es de $t = 0,81 Gyr$. El máximo de edad se encuentra, como es de esperar, por debajo de la edad del universo calculada con la cosmología actual [G. Hinshaw et al., 2013].

Masa

Gran cantidad de distribuciones cuentan con valores que no son físicamente aceptables teniendo en cuenta el tipo de objeto que se está estudiando.

En el caso particular de la $\log(M/M_{sun})$, que hace referencia a la masa de las galaxias, valores

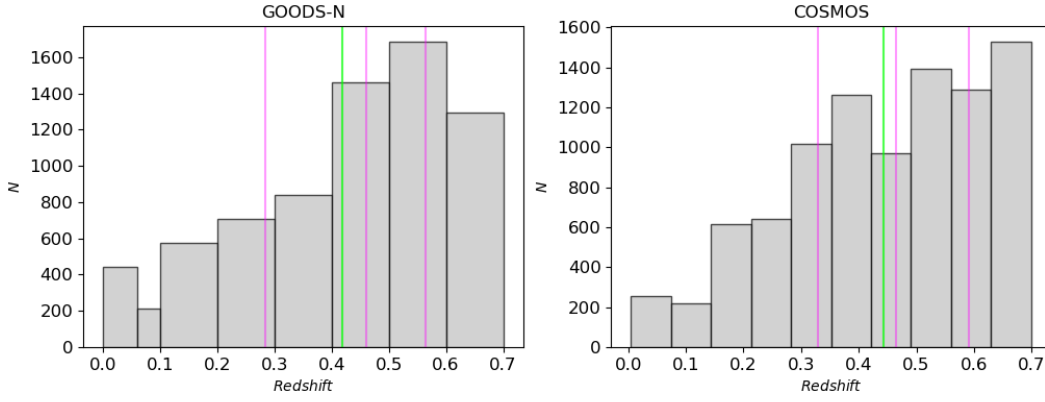


Figura 3: Izquierda: Distribución de z extendida hasta $z=0.7$ donde se aprecia el máximo en torno al valor $z=0.5$. Derecha: Distribución de z para el campo de COSMOS, donde se aprecia la diferencia con la obtenida para el campo GOODS-N

como $\log(M/M_{sun}) = 0$ ó $\log(M/M_{sun}) = 2$ son inverosímiles ya que corresponderían a galaxias con masa equivalente a la del Sol, o cien veces mayor a esta.

La galaxia menos masiva conocida hasta el momento es *Segue 2* [Belokurov et al., 2009], la cual tiene una masa aproximada de $M = 6 \times 10^5 M_{sun}$, por lo que $\log(M/M_{sun}) = 6$. Esta cota no es aplicable al catálogo aquí empleado, ya que en buena medida la luminosidad está relacionada con la masa en las galaxias [Pskovskii, 1965]. Por tanto, es de esperar que galaxias más lejanas sufran efecto de selección, y por tanto solo se observen las de una masa considerablemente mayor que *Segue 2*.

En [Barro et al., 2019], se afirma que el catálogo cubre por completo galaxias con $\log(M/M_{sun}) = 10$ y $A_v < 2$, lo que confirma la necesidad de imponer una cota inferior razonable. Por tanto, se establecerá, con el criterio arriba citado, una cota inferior de $\log(M/M_{sun}) = 6$ para realizar los prosiguientes estudios.

4. Relaciones Estructurales

En primer lugar se realizará un estudio preliminar con los datos filtrados acorde a las cotas ya mencionadas, para poder buscar posibles relaciones estructurales entre parámetros. Se separan los parámetros entre los observados y derivados, aunque será conveniente considerarlos conjuntamente más tarde para observar posibles relaciones entre unos y otros.

Se tratará entonces de determinar las relaciones estructurales que relacionen diferentes variables entre sí. Después se estudiarán con mayor énfasis las relaciones con el parámetro $\log(M/M_{sun})$, de gran relevancia en la caracterización de la muestra en nuestro estudio.

Finalmente, resultará de utilidad buscar relaciones más allá de las existentes entre parejas de parámetros y con $\log(M/M_{sun})$, haciendo uso de diferentes técnicas de representación de datos.

4.1. Correlación entre parámetros

Teniendo en mente que el objetivo final es encontrar posibles relaciones estructurales entre parámetros de la muestra, resulta conveniente realizar un mapa visual que permita discriminar de una forma sencilla.

Para ello resultará útil construir una serie de figuras empleando herramientas como Python (Python Software Foundation, <https://www.python.org/>) y el paquete estadístico Pandas (NumFOCUS, <https://pandas.pydata.org/>). Se hace uso del siguiente código implementado en Python 3.8.5:

```
plt.matshow(corr_obs, cmap='coolwarm')
plt.xticks(range(ds_obs.select_dtypes(['number']).shape[1]), ds_obs.
            select_dtypes(['number']).columns, fontsize=12, rotation=90)
plt.yticks(range(ds_obs.select_dtypes(['number']).shape[1]), ds_obs.
            select_dtypes(['number']).columns, fontsize=12, rotation=0)
cb = plt.colorbar()
plt.figtext(0.5, 0.03, 'Parámetros Observacionales', fontsize = 12,
            horizontalalignment = "center")
plt.savefig('figuras/correl_obs')
```

De ese modo es posible obtener las Figuras 7 y 6, las cuales se pueden interpretar como mapas bidimensionales en los que se lee la correlación entre cada par de parámetros estudiados. La diagonal tiene $r = 1$ donde r es el coeficiente de correlación, ya que relaciona una variable consigo misma.

Las relaciones que se observan entre las tres magnitudes absolutas M_U , M_V y M_J ; así como entre las variables $la2t$ y $ltau$ no son significativas, ya que las magnitudes absolutas indican cuan brillante es el objeto (y por tanto están directamente relacionadas para los objetos estudiados); y el par restante de variables también está relacionada a través de su definición:

$$la2t = \log(t/\tau) = \log(age) - \log(\tau)$$

Una vez descartadas las correlaciones no significativas, el mapa puede indicar visualmente las parejas que están correlacionadas en mayor medida, aunque hay que tener en cuenta que muchas de ellas contienen datos fuera del rango esperado en el caso de que no haya sido posible realizar la medida correspondiente (ya que se asigna por defecto un valor de -99), por lo que es conveniente revisarlas posteriormente con más detenimiento.

En principio, a la vista de la figura 7 es de esperar que exista relación entre $\log(M/M_{sun})$ y las magnitudes M_V , M_U , M_J , así como una correlación más ligera con $\log(SFR/(M_{sun}/yr))$, z y $\log(age/yr)$.

Tal y como se aprecia en las Figuras 8 y 9, los parámetros que más se correlacionan con $\log(M/M_{sun})$ son M_V y $\log(SFR/(M_{sun}/yr))$. Los relacionados con la edad presentan un aspecto diferente, ya que toman únicamente valores discretos, tal y como se explicó anteriormente.

A la vista de las figuras no se aprecia una relación clara entre $\log(M/M_{sun})$ y radio o redshift, el aspecto de la distribución es una nube dispersa sin una correlación bien definida.

4.2. Correlaciones encontradas en la muestra

Las correlaciones presentadas a lo largo de las figuras aportan información sobre las posibles relaciones estructurales presentes dentro de la muestra de galaxias. De entre todas ellas se seleccionan las físicamente relevantes, es decir, las que muestran un grado de correlación significativo, y que no están causadas por efectos de selección.

Este es el caso de la correlación entre z y M_V , ya que está causada por un efecto de selección: a distancias lejanas cada vez se detectan fuentes más brillantes (o más masivas). Este efecto se conoce como *Malmquist Bias* [Malmquist, 1922].

Las relaciones más significativas son, esencialmente, las siguientes:

- Logaritmo de la Masa-SFR: en la Figura 8 se observa una buena correlación entre la masa de las galaxias y la tasa de formación estelar, con un coeficiente $r = 0,64$. Esta correlación se puede entender relacionando la mayor masa de la galaxia con una mejor capacidad de esta para conservar gas, y por tanto, dando lugar a mayor tasa de formación estelar, tal y como se observa en la correlación obtenida.
- Logaritmo de la Masa-Magnitud banda V: clara correlación, tal y como se aprecia en la Figura 9, con $r = 0,95$. Lo que nos dice que la masa está estrechamente relacionada con la luminosidad de las galaxias, hecho bien conocido y explorado [Pskovskii, 1965].
- Logaritmo de la Masa- Logaritmo de la Edad: correlación que se muestra en la Figura 8, con $r = 0,35$. Sin embargo la forma peculiar de $\log(age/yr)$ dificulta la correcta interpretación de las posibles correlaciones. Este fenómeno se puede explicar mediante el fenómeno de *down-sizing* [Cowie et al., 1996], el cual propone que los procesos de formación estelar aparecieron antes en los sistemas de mayor masa, para propagarse después a galaxias menos masivas.
- Magnitud banda V-SFR: correlación con $r = 0,75$. Explicable debido a la existente entre el logaritmo de la Masa y SFR.
- Magnitud banda V-Tamaño: galaxias con tamaño aparente mayor se agrupan en torno a valores mayores de M_v . En concreto, cuenta con un coeficiente $r = 0,47$. Se trata de un tamaño aparente, por lo que la interpretación no es clara. Cabría esperar una relación entre tamaño físico y magnitud absoluta, que se explorará más adelante.
- Redshift-SFR: galaxias con mayor redshift tienen mayor SFR. Esta correlación tiene un $r = 0,45$. Se puede entender ya que es un hecho conocido que la tasa de formación estelar fue más intensa en el pasado, como se discutirá más adelante.

Si bien estas últimas tres relaciones son estadísticamente menos significativas, debido a que el valor de r es menor al de las mostradas en primer lugar, a la vista de las figuras obtenidas todavía son apreciables y dignas de mención.

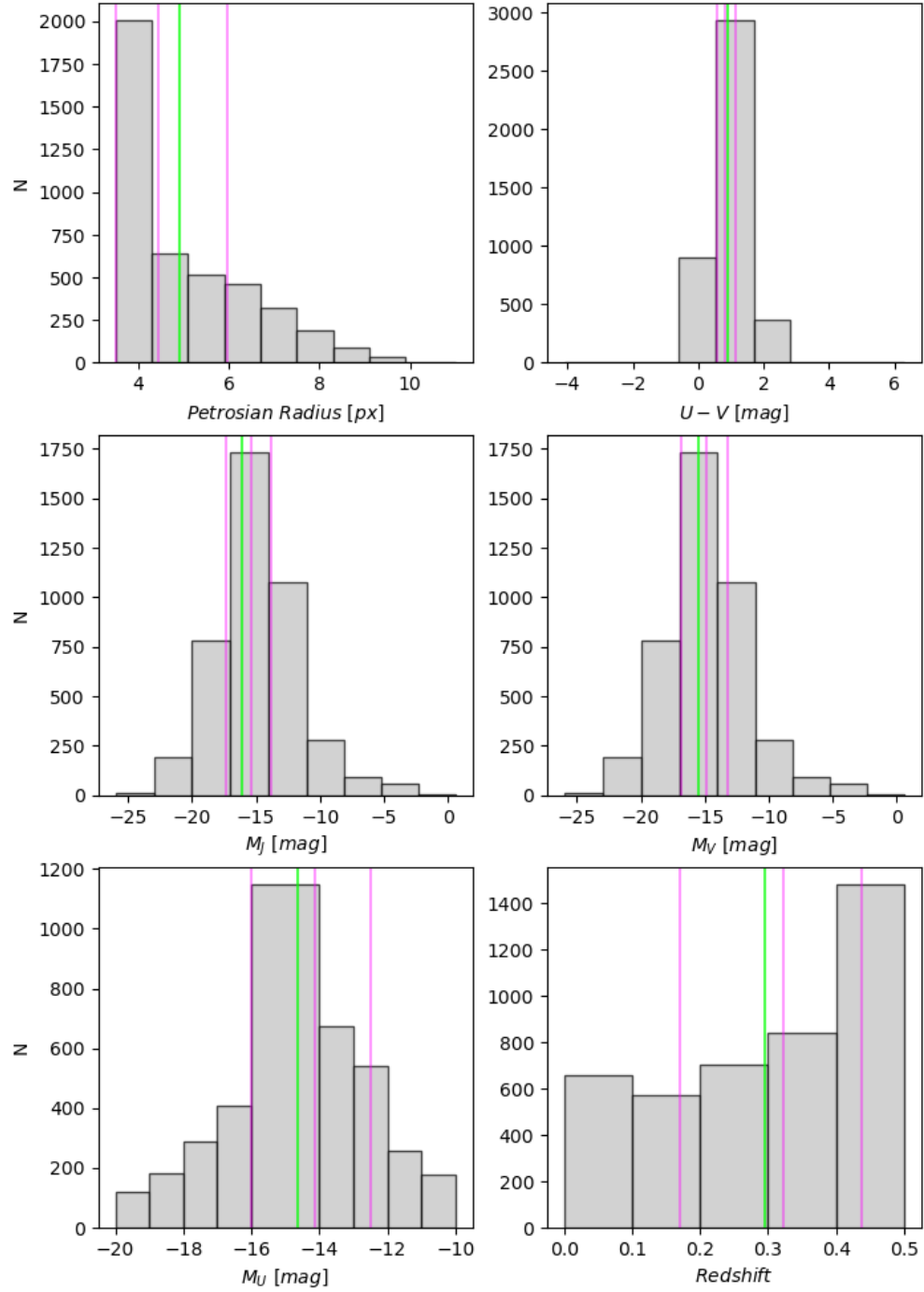


Figura 4: Distribución de los parámetros observados. La línea verde representa la media aritmética. Las líneas moradas representan los cuartiles.

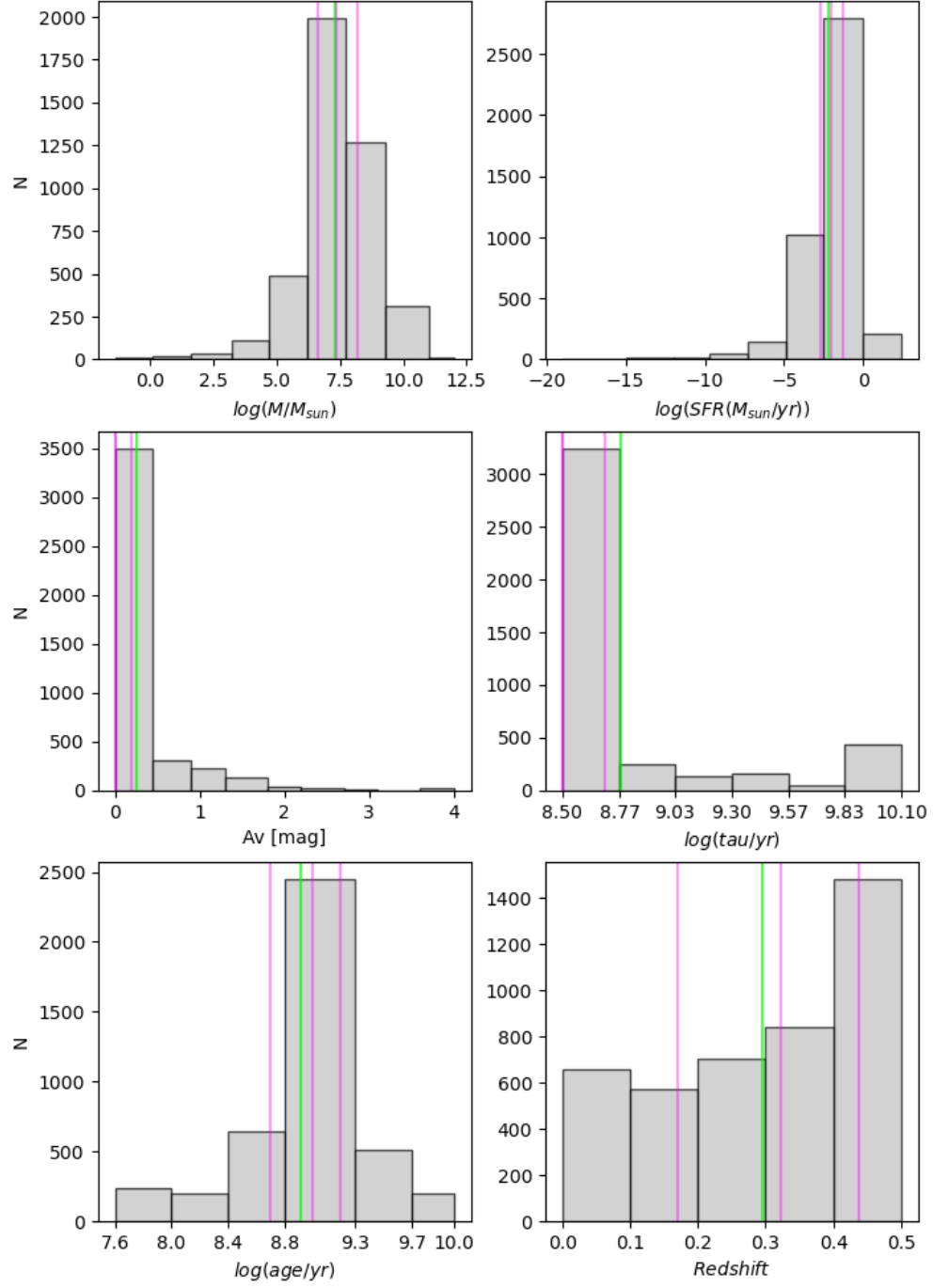


Figura 5: Distribución de los parámetros derivados. La línea verde representa la media aritmética. Las líneas moradas representan los cuartiles.

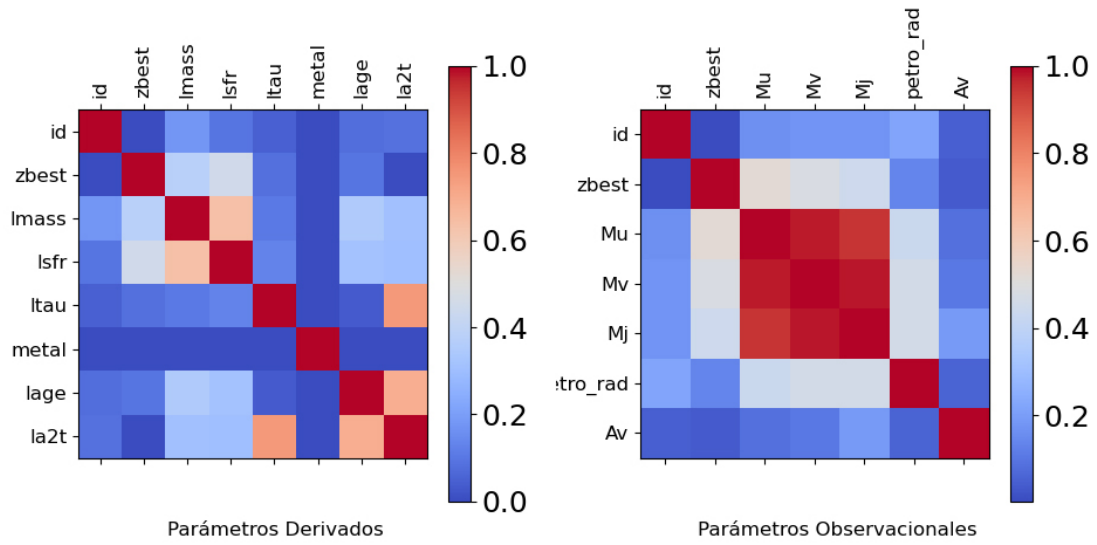


Figura 6: Matriz de Correlación entre parámetros observacionales y derivados. Más saturación de color (rojo o azul) indica correlación (positiva o negativa) más intensa entre los datos.

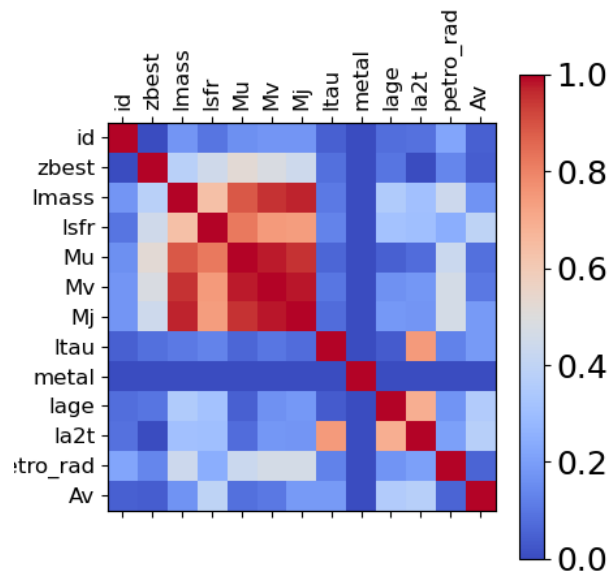


Figura 7: Matriz de Correlación entre parámetros. Un color más rojo indica una correlación mayor.

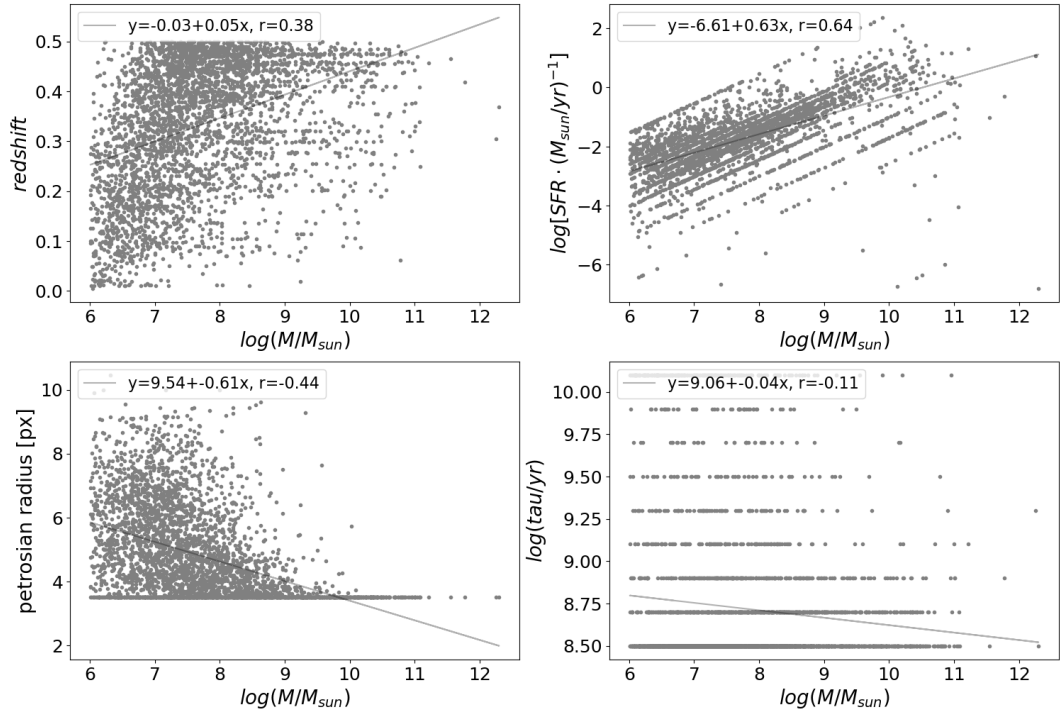


Figura 8: Relaciones de parámetros con $\log(M/M_{sun})$. Se incluye el parámetro de correlación así como la ecuación de la recta que mejor se ajusta a los datos.

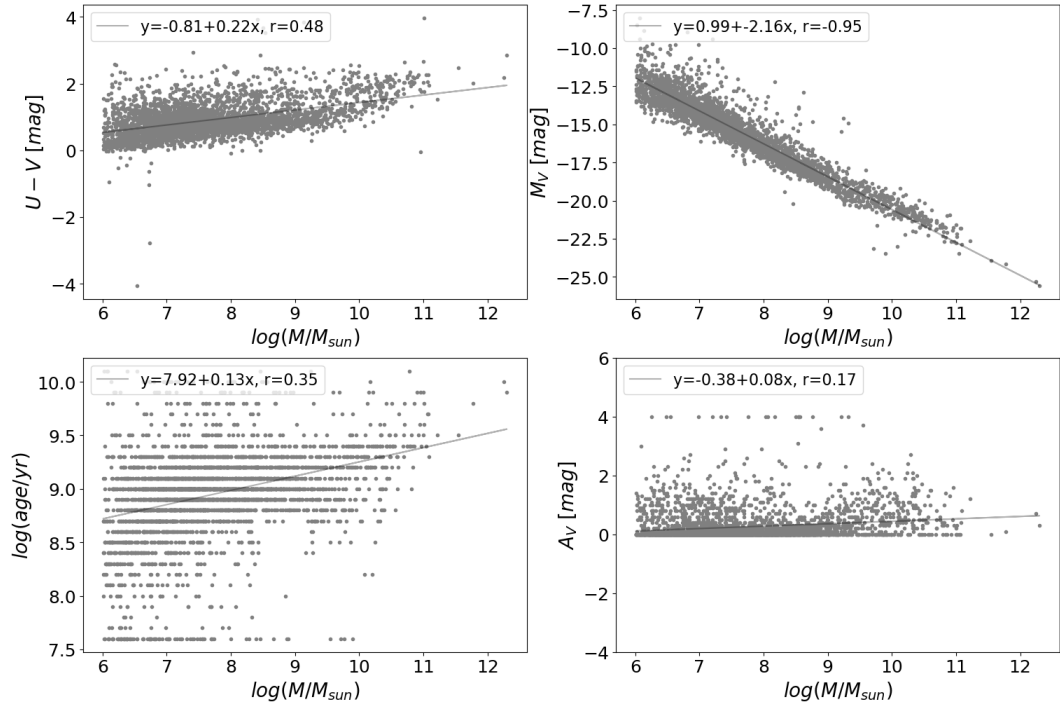


Figura 9: Relaciones de parámetros con $\log(M/M_{sun})$. Se incluye el parámetro de correlación así como la ecuación de la recta que mejor se ajusta a los datos.

4.3. Análisis detallado

Si bien a través de los métodos mostrados anteriormente es posible delimitar las correlaciones más relevantes existentes en la muestra, así como dar una explicación plausible a las que tienen origen físico y rechazar las que provienen de efectos de selección; es conveniente realizar un análisis más pormenorizado y profundo de cada una de ellas.

Hasta el momento se han empleado exclusivamente métodos estadísticos paramétricos, y por tanto se está aceptando implícitamente la hipótesis de que los datos han de seguir una distribución estadística conocida. Tampoco se ha comprobado la fiabilidad de los resultados: la significación estadística de estos.

Por tanto, resulta interesante implementar métodos no paramétricos, tal y como se describe en [Javier Gorgas García, 2014]. Y realizar tests de contraste de hipótesis para conocer la significación y validez de las conclusiones obtenidas.

La metodología será la siguiente:

- Obtener los coeficientes de correlación de Kendall y Spearman
- Contrastar las correlaciones para una significación del 5 %
- Ajustar las rectas usando el estimador de Theil-Sen

Además, aplicar el método de Theil-Sen para realizar el ajuste es ventajoso al ser más robusto frente a posibles 'outliers', los cuales producen un ajuste de peor calidad empleando el equivalente método paramétrico de mínimos cuadrados.

A la vista de los resultados mostrados en la Tabla 4, todas las correlaciones son significativas para un nivel de confianza del 95 %, y el valor de los coeficientes de correlación es del mismo orden que los obtenidos anteriormente.

En el caso del ajuste lineal, los parámetros m y n tampoco son notablemente distintos de los obtenidos con el ajuste por mínimos cuadrados, sin embargo, es de esperar que se vean menos afectados en el caso de existir valores extremos.

Tal y como se aprecia en la Figura 10, es posible obtener de nuevo la representación visual de la correlación entre parámetros, tal y como se hizo anteriormente. En la Figura 10 se pueden apreciar las mismas estructuras que en la Figura 7, su equivalente paramétrico. Únicamente existe diferencia en los valores numéricos para cada pareja de datos, pero las correlaciones relevantes se mantienen invariantes.

Todo esto sugiere que los métodos paramétricos, en este caso, han arrojado resultados válidos. Sin embargo, es preferible emplear los no paramétricos siempre que sea posible debido a que evitan realizar hipótesis innecesarias sobre la población, y debido también a su mayor robustez.

| correlación | $\log(M/M_{sun}) - \log(SFR/(M_{sun}/yr))$ | $\log(M/M_{sun}) - M_v$ | $\log(M/M_{sun}) - \log(age)$ | $z - \log(SFR/(M_{sun}/yr))$ |
|------------------|--------------------------------------------|-------------------------|-------------------------------|------------------------------|
| r - pea [1] | 0.64 | 0.95 | 0.35 | 0.45 |
| significativo | S | S | S | S |
| τ - ken [2] | 0.51 | 0.82 | 0.28 | 0.32 |
| significativo | S | S | S | S |
| r - spe [3] | 0.67 | 0.95 | 0.38 | 0.46 |
| significativo | S | S | S | S |
| m [4] | 0.73 | -2.19 | 0.12 | 3.34 |
| n | -7.26 | 1.14 | 8.05 | -3.01 |

Tabla 4: Estadística no Paramétrica. [1]: coeficiente de correlación de Pearson, método paramétrico para estimar correlación, junto a la significación de la correlación para el 5 %. Se indica como: S - existe correlación significativa, N - no existe correlación significativa; [2]: tau de Kendall, método no paramétrico para estimar correlación, junto a la significación de la correlación para el 5 %; [3]: r de Spearman, método no paramétrico para estimar correlación, junto a la significación de la correlación para el 5 %; [4]: pendiente (m) y ordenada en el origen (n) obtenidos mediante el estimador de Theil-Sen.

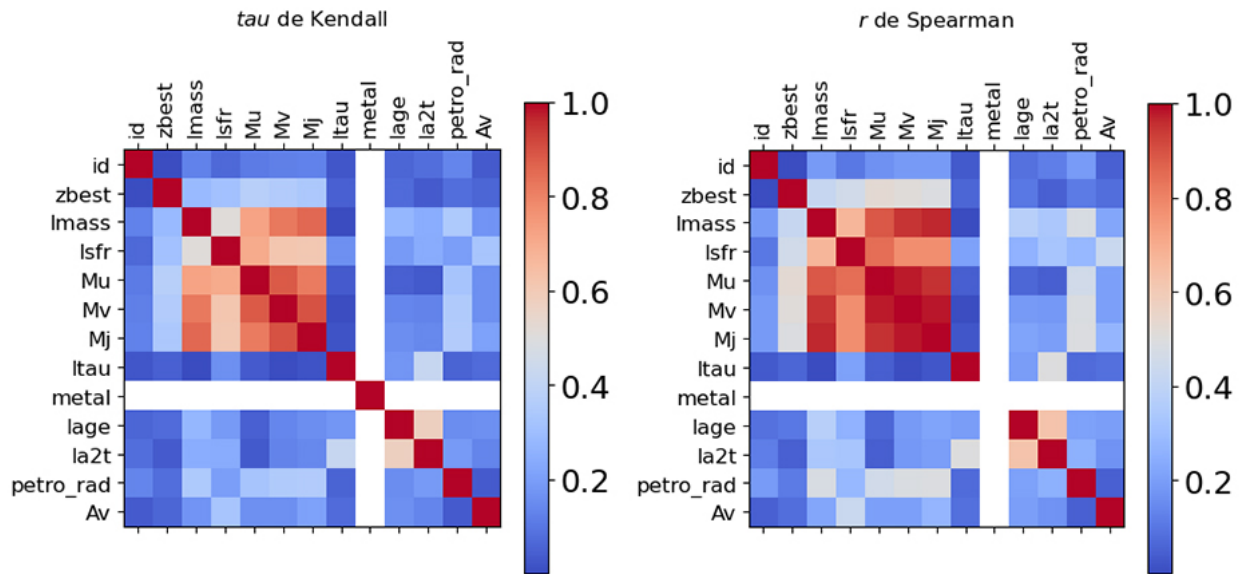


Figura 10: Valor absoluto del parámetro τ de Kendall y r de Spearman para cada uno de los parámetros estudiados. La interpretación de los mismos es distinta [Javier Gorgas García, 2014], pero ambos indican el grado de relación entre los datos.

5. Otras Consideraciones

5.1. Exploración de relaciones no binarias

Es también interesante buscar otras correlaciones existentes entre parejas de parámetros, tarea que será más sencilla haciendo uso de las Figuras 6 y 7. Resulta útil definir el tamaño físico de las galaxias en función de su tamaño aparente (radio petrosiano), distancia (determinada a partir del redshift empleando la cosmología [G. Hinshaw et al., 2013], y escala de placa del detector según consta en [Barro et al., 2019].

Se estudiará pues, la relación entre redshift z , logaritmo de la tasa de formación estelar $\log(SFR/(M_{sun}/yr))$ y tamaño. También la existente entre el logaritmo de la masa $\log(M/M_{sun})$, magnitudes M_U, M_V, M_J y tamaño. Para ello se hará uso de gráficos que permitan estudiar las anteriores correlaciones, incluyendo además un tercer componente de color.

En la Figura 11 se aprecia cómo por lo general las galaxias con tasas de formación estelar mayores tienen un tamaño promedio situado entre los 2.5 kpc y 5 kpc. Y cómo las galaxias a mayor redshift son las de mayor tamaño. Estudiando en mayor profundidad la muestra, se puede concluir que existe posible un sesgo experimental. El tamaño de la mayor fuente es de $10px$, y la distribución de tamaños se puede observar en la Figura 4. Por tanto, se detectan únicamente galaxias con tamaños aparentes pequeños, por lo que quedan excluidas las galaxias cercanas de gran tamaño.

Otro efecto a tener en cuenta es que el volumen observado crece según z crece, por lo que se un número mayor de galaxias a z elevado que a z bajo. Además, las galaxias muy masivas son en promedio menos abundantes en el universo [Schechter, 1976]. Todo ello conduce a que dentro del catálogo no existan galaxias de gran tamaño relativamente próximas. Nuevamente se puede apreciar en la Figura 13, donde queda claro que las galaxias de tamaños físicos mayores únicamente aparecen a distancias grandes. Aquí se precia cómo tampoco aparecen galaxias pequeñas a distancias grandes, de nuevo por sesgo experimental.

5.2. Diagrama color - magnitud

A partir de los datos disponibles es posible obtener algunas relaciones relevantes para los objetos extragalácticos, como puede ser el diagrama color-magnitud mostrado en la Figura 14 [Iskra Strateva et al., 2001]. En él se aprecia levemente una bimodalidad en la distribución (como se puede apreciar en las líneas de contorno y la forma de la distribución). La bimodalidad existente se puede entender como que las galaxias se distribuyen según sus colores en la *blue cloud* y *red sequence*.

Existe una diferenciación entre la población estelar de galaxias más rojas (generalmente elípticas), las cuales tienen baja tasa de formación estelar y estrellas evolucionadas. Y de las más azules (mayoritariamente espirales), las cuales mantienen una tasa de formación estelar mayor y una población más joven, de ahí su color azul.

De haberse comparado este diagrama para diferentes valores de redshift z , debería apreciarse cómo el número de galaxias que pueblan la nube roja crece según decrece el valor del redshift al que se observa, ya que la población estelar es de mayor edad; y hay mayor prevalencia de galaxias de tipos tempranos (elípticas y lenticulares).

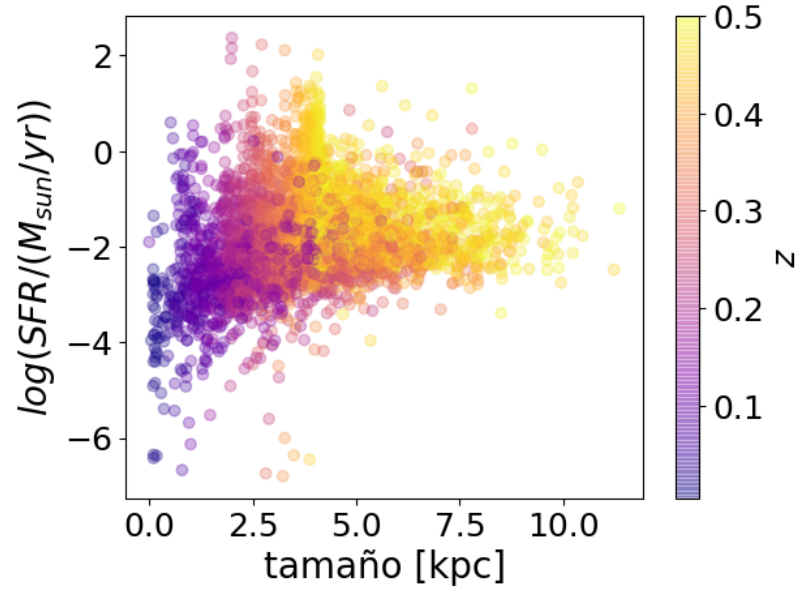


Figura 11: Se incluye un tercer parámetro codificado en forma de color.

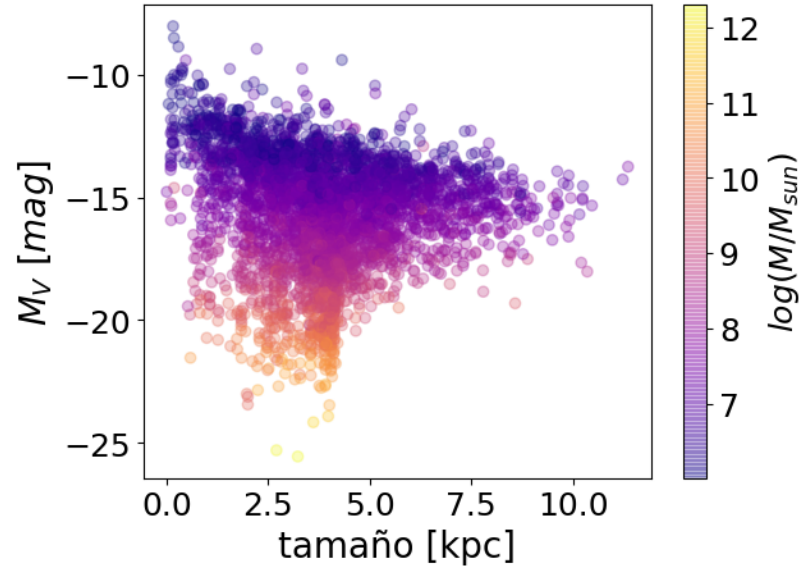


Figura 12: Ídem a la Figura 11.

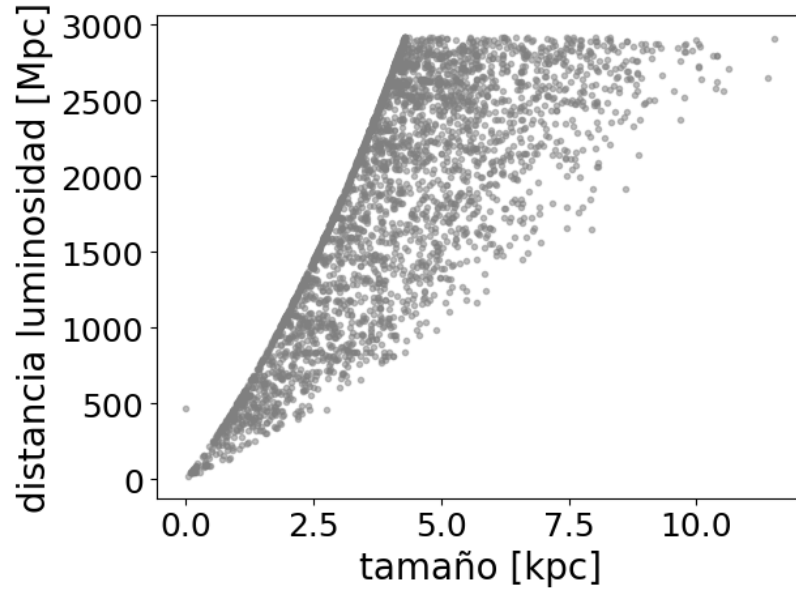


Figura 13: Se muestra la relación entre el tamaño físico y la distancia de luminosidad bajo cosmología WMAP9 [G. Hinshaw et al., 2013].

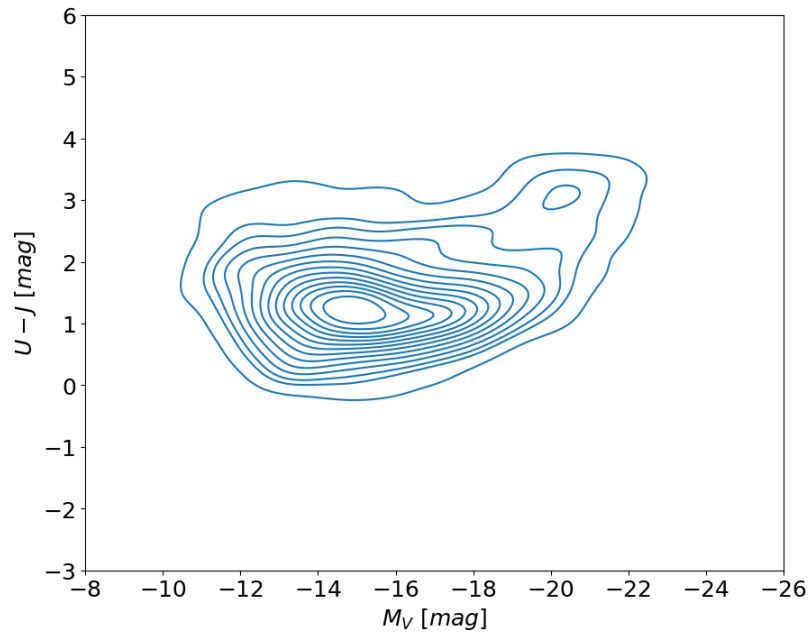


Figura 14: Diagrama color magnitud de las galaxias de la muestra.

5.3. Evolución de las relaciones estructurales

Otro de los objetivos es conocer cómo ha variado la distribución de galaxias y sus relaciones estructurales a lo largo de la historia del universo. Para ello se estudian los intervalos en z siguientes: $\{0, 0.5\}$; $\{1, 1.5\}$; $\{1.5, 3\}$. El primero de ellos ha sido el tratado en detalle a lo largo de estas páginas, por lo que se procederá a comparar algunos de los resultados aquí obtenidos con los obtenidos para diferentes valores de redshift.

Cosmic Dawn: $z=1.5$ a $z=3$

El número de galaxias de la muestra es grande en comparación al encontrado para el rango $\{0, 0.5\}$, con 16339 elementos entre $\{1.5, 3\}$. Este hecho puede explicarse por el aumento de volumen de muestreo con la distancia, así como la ausencia de *mergers* entre galaxias que dan lugar a sistemas de mayor masa, pero reducen su abundancia.

Se encuentra que la mayor parte de objetos con tasa de formación estelar se sitúan en torno a $\log(M/M_{sun}) = 8$ y $\log(M/M_{sun}) = 10$. Son numerosas las galaxias con tasa de formación estelar elevada y masa pequeña o mediana. Este hecho se puede apreciar en la Figura 15.

Cosmic Noon: $z=1$ a $z=1.5$

Vuelve a existir un claro aumento en el número de galaxias con 7429 elementos entre $\{1, 1.5\}$. Se aprecia un corrimiento del color hacia el azul, y una tendencia general hacia mayores tasas de formación estelar. Esto lleva a pensar que en comparación con el presente, en ese momento existió mayor actividad de formación estelar, y la población era por lo general más joven.

Si bien los procesos de formación estelar en las galaxias elípticas suelen ser rápidos, también existen sistemas donde la SFR se ha mantenido en el tiempo. En concreto, la curva que relaciona la tasas de formación estelar con la masa varía con la obtenida para diferentes redshifts, por lo que la secuencia principal de galaxias con formación estelar ha variado en tanto que los sistemas forman hoy en día menos estrellas.

Este efecto se puede relacionar con el *downsizing* [Cowie et al., 1996], en el que la actividad se va propagando de sistemas más masivos a otros de menor masa. Unificar este hecho observacional con el modelo jerárquico de evolución galáctica escapa a este trabajo. De nuevo, esto es claro a la vista de la Figura 15.

Universo Local: $z=0$ a $z=0.5$

Este periodo corresponde al aquí estudiado, y ha sido el que se toma como referencia para comparar las diferentes propiedades que exhibe la muestra de galaxias a diferentes redshifts.

Por lo general, se puede apreciar la existencia de galaxias de masa mayor, hecho en concordancia con el modelo de universo jerárquico [Shaun Cole et al., 2000]). También se aprecia cómo los procesos de formación estelar se han desplazado a galaxias menos masivas.

También es notable cómo el color de esta muestra es el más rojo (por la mayor presencia de población estelar evolucionada), así como un notable decrecimiento en el número de galaxias de la muestra (aunque este efecto está principalmente causado por efectos observacionales).

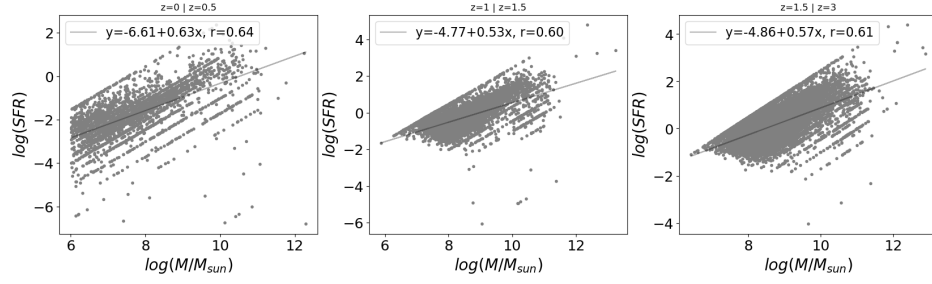


Figura 15: Comparación de la relación logaritmo de masa estelar y tasa de formación estelar. Se ilustra el efecto de downsizing, ya que es evidente cómo a redshifts menores aparece mayor SFR en objetos de menor masa. Sin embargo esta afirmación habría de respaldarse con mayores evidencias, ya que bien podría ser causada por efecto de selección, al no detectarse objetos poco masivos a redshifts grandes.

6. Conclusiones Finales

Aunque el catálogo Candles tiene un gran potencial, y constituye una muestra realmente rica, en este trabajo únicamente se ha realizado un análisis superficial de la muestra. Una vez caracterizada la muestra y entendidas las relaciones que existen entre los miembros del estudio, se ha podido reducir un catálogo de miles de componentes y tener una idea clara de cómo se distribuye la muestra en cada uno de los parámetros, y las relaciones que existen entre ellos.

Al ser un catálogo puramente observacional, aparecen efectos de selección de manera natural, por lo que es de vital importancia hacer una correcta interpretación física de cada una de las relaciones encontradas. Ya que lo que a priori parece una clara correlación puede no deberse más que a un efecto observacional, como sucede en el caso de la luminosidad y z [Malmquist, 1922].

También es notable la aparente discretización de valores para parámetros como la edad, τ y otros. Esto se debe al método empleado para derivar estos parámetros a partir de otros observacionales. En este caso, se ha realizado comparando con catálogos previamente construidos (pudiendo ser estos empíricos o teóricos). De cualquier modo, solo cubren el espacio de parámetros para ciertos valores concretos, de ahí el efecto observado al representar los datos.

Como conclusión, todo el análisis realizado me ha permitido profundizar en varios campos impartidos a lo largo del presente curso académico como astrofísica extragaláctica o estadística y análisis de datos. Resultando por tanto provechoso y sumamente interesante como una aplicación práctica de los métodos allí aprendidos.

Personalmente, he profundizado en el análisis de datos con Python (desarrollando más de 800 líneas de código que permiten realizar el análisis y obtener las figuras de manera directa), en la búsqueda de bibliografía de referencia; y he comprendido la importancia de dar una explicación física a los fenómenos observados, descartando los que no tienen validez. También ha sido útil como una primera aproximación a la redacción de un texto de carácter científico, tanto en cuanto a su contenido como a su continente.

Como complemento al trabajo realizado, y para facilitar la continuación del mismo, se presenta el material necesario en la web <https://sites.google.com/view/explore-candles/home>, hosteada en el servicio *google sites*. Aquí se podrá acceder al repositorio del código que se ha hecho disponible públicamente en *github*, así como recuperar los resultados y gráficos obtenidos en el trabajo.

Referencias

- [Barro et al., 2019] Barro, G. et al. (2019). The candels/shards multi-wavelength catalog in goods-n: photometry, photometric redshifts, stellar masses, emission line fluxes and star formation rates. *arXiv*.
- [Belokurov et al., 2009] Belokurov, V. et al. (2009). Segue 2: A prototype of the population of satellites of satellites.
- [Carnall et al., 2019] Carnall, A. C. et al. (2019). How to measure galaxy star formation histories. i. parametric models. *The Astrophysical Journal*, 873:44.
- [Cowie et al., 1996] Cowie, L. L. et al. (1996). New insight on galaxy formation and evolution from keck spectroscopy of the hawaii deep fields.
- [G. Hinshaw et al., 2013] G. Hinshaw, D. Larson, E. K. et al. (2013). Nine-year wilkinson microwave anisotropy probe (wmap) observations: Cosmological parameter results.
- [Giavalisco et al., 2004] Giavalisco, M. et al. (2004). The great observatories origins deep survey: Initial results from optical and near-infrared imaging 1.
- [Grogin et al., 2011] Grogin, N. A. et al. (2011). Candels: The cosmic assembly near-infrared deep extragalactic legacy survey.
- [Iskra Strateva et al., 2001] Iskra Strateva, Zeljko Ivezić, G. R. K. et al. (2001). Color separation of galaxy types in the sloan digital sky survey imaging data.
- [Javier Gorgas García, 2014] Javier Gorgas García, Nicolás Cardiel López, J. Z. C. (2014). *Estadística Básica para Estudiantes de Ciencias*.
- [Koekemoer et al., 2011] Koekemoer, A. M. et al. (2011). Candels: The cosmic assembly near-infrared deep extragalactic legacy survey - the hubble space telescope observations, imaging data products and mosaics.
- [Malmquist, 1922] Malmquist, K. G. (1922). *Arkiv for Matematik, Astronomi och Fysik*, 23.
- [Nayyeri et al., 2017] Nayyeri, H. et al. (2017). Candels multi-wavelength catalogs: Source identification and photometry in the candels cosmos survey field. *The Astrophysical Journal Supplement Series*, 228:7.
- [Petrosian, 1976] Petrosian, V. (1976). Surface brightness and evolution of galaxies. *The Astrophysical Journal*, 209:1–5.
- [Pskovskii, 1965] Pskovskii, Y. P. (1965). *Soviet Astronomy*, 9:253–258.
- [Schechter, 1976] Schechter, P. (1976). An analytic expression for the luminosity function for galaxies.
- [Shaun Cole et al., 2000] Shaun Cole, Cedric G. Lacey, C. M. B. et al. (2000). Hierarchical galaxy formation.