UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE POSGRADO EN ASTROFÍSICA

INSTITUTO DE ASTRONOMÍA

HACIA UNA DESCRIPCIÓN SEMI-EMPÍRICA DE LA DEMOGRAFÍA MULTI-COMPONENTE Y DISTRIBUCIÓN ESPACIAL DE LAS GALAXIAS

TESIS

QUE PARA OPTAR POR EL GRADO DE: DOCTOR EN CIENCIAS (ASTROFÍSICA)

PRESENTA:

ÁNGEL RUBÉN CALETTE MORÍN

TUTORES:

DR. VLADIMIR ÁVILA REESE
DR. ALDO RODRÍGUEZ PUEBLA

INSTITUTO DE ASTRONOMÍA

CIUDAD UNIVERSITARIA, CD. MX., ENERO 2021

Agradecimientos

A mis tutores, Dr. Vladimir Ávila Reese y Dr. Aldo Rodríguez Puebla quienes siempre me han apoyado y retroalimentado con su conocimiento para mejorar cada día como científico y persona. Estoy infinitamente agradecido.

A mi comité sinodal, Dra. Claudia Lagos, Dr. Jorge Barrera, Dr. Vicente Rodríguez, Dr. José Antonio de Diego y Dr. Miguel Aragón, por su tiempo y sus valiosos comentarios que han contribuido a la culminación de esta tesis.

Resumen

La demografía de la población local de galaxias trazada por sus múltiples componentes (masas estelares, de gas atómico HI, gas molecular H_2, gas neutro total y materia oscura) es el resultado de un complejo proceso evolutivo del campo de perturbaciones primigenio de materia oscura acoplado al gas que disipa, forma estrellas y se retro-alimenta energética y químicamente. Una descripción demográfica multi-componente de la población galáctica local es entonces crucial para entender la evolución de las galaxias. Los catastros de cientos de miles galaxias en el óptico han permitido lograr una descripción estadística completa en función de la masa estelar M_* de las galaxias. No es el caso de las observaciones en radio que trazan las componentes de gas neutro y molecular, HI y H_2 respectivamente. En una primera parte de esta Tesis, se busca subsanar esta situación a través de una extensa compilación de detecciones y no detecciones en radio de HI y H_2 para galaxias de tipos tardíos (LTGs) y tempranos (ETGs). Después de homogeneizar la muestra y corregir por sistemáticos y sesgos, se aplica un análisis de supervivencia para detecciones y límites superiores y así obtener las distribuciones condicionales (DCs) de M_{HI} y M_{H_2} dada la M_*, tanto para galaxias tardías y tempranas como para el total. El primer y segundo momento de estas DCs son la relaciones promedio M_{HI}-M_* y M_{H_2}-M_* y sus dispersiones. En combinación con funciones de masa (FM) estelar meticulosamente calculadas aquí, completas hasta $M_* \approx 3 \times 10^7$ $M_\odot$, se predicen entonces las distribuciones bivariadas de (M_*, M_{HI}) y (M_*, M_{H_2}) cuyas proyecciones dan las FM empíricas de HI y H_2. Con base a la demografía empírica obtenida, se presenta un censo cósmico de estrellas, HI, H_2, medio interestelar y bariones para LTGs, ETGs y todas las galaxias. De aquí, se estiman tiempos característicos de consumo del gas HI y H_2 por formación estelar para las LTGs.

En una segunda parte de la Tesis, se establece la conexión galaxia-halo oscuro a nivel estelar y de gas HI, para LTGs, ETGs y todas las galaxias. Con un potente método estadístico se genera un catálogo de galaxias en los halos y subhalos de una enorme simulación cosmológica de N cuerpos. Por construcción, la población de galaxias de este catálogo sintético presenta la demografía estelar y de HI de las observaciones, tanto para LTGs y ETGs como para todas las galaxias, además de reproducir el acumulamiento espacial en M_* de las galaxias y la fracción de galaxias satélites/centrales en función de M_*. Haciendo uso de una sub-muestra reciente de galaxias proveniente del sondeo SDSS con información de HI, el xGASS, se infieren las correcciones apropiadas para aplicar a nuestras DCs de M_{HI} dada la M_* y así

obtener las DCs correspondientes de galaxias centrales y satélites. Como resultado, se obtienen las correlaciones de M_* y M_{HI} con la masa de los halos, para todas las galaxias, así como para sólo las galaxias centrales y satélites. Esto permite medir en la simulación el acumulamiento espacial en M_{HI} y comparar con las observaciones de catastros ciegos en HI. Se encuentra que los efectos de selección de ALFALFA, el más completo de estos catastros, afectan considerablemente el acumulamiento espacial por HI.

Los resultados de esta Tesis ofrecen una descripción empírica completa de la demografía y distribución espacial de la población local de galaxias, misma que es una antesala a lo que se logrará en la siguiente década con los catastros en radio obtenidos con los radiotelescopios SKA e instrumentos precursores como ASKAP y WSRT. Estos resultados pueden ser usados como comparaciones y calibraciones de modelos semi-analíticos y simulaciones cosmológicas de evolución de galaxias de nueva generación, mismos que son capaces ya de modelar no sólo la componente estelar y de materia oscura de las galaxias, sino que también modelan la componente disipativa bariónica (es decir el gas neutro y molecular del medio interestelar).

Contents

Chapter 1

Introducción

1.1 Antecedentes y Motivación

De acuerdo al paradigma cosmológico actual, conocido como el modelo de Materia Oscura Fría con constante cosmológica Λ (ΛCDM por sus siglas en inglés), $\sim 84\%$ de la densidad de masa no relativista en el Universo se encuentra en forma de materia oscura (Planck Collaboration et al., 2016a), siendo el resto la materia bariónica. El paradigma ΛCDM establece que las galaxias se forman, evolucionan y eventualmente fusionan dentro de estructuras virializadas masivas de materia oscura, referidos en la literatura como *halos de materia oscura*. Estos halos oscuros, de acuerdo al cada vez más establecido paradigma inflacionario, son el producto de la evolución gravitacional de un campo primigenio de perturbaciones en densidad originado de las fluctuaciones del vacío cuántico. Se postula que con el tiempo los halos de materia oscura crecen mediante dos mecanismos impulsados por la gravedad: (1) la incorporación de material difuso o acreción suave y (2) la incorporación de materia mediante fusiones. Dichos procesos establecen las condiciones para la formación de las galaxias en el interior de los halos oscuros (ver reseñas del paradigma cosmológico actual, p. ej. en Baugh, 2006; Avila-Reese, 2007; Longair, 2008; Mo et al., 2010a).

Dentro del paradigma ΛCDM, la formación de galaxias es un proceso complejo y no lineal en el cual la materia oscura debió jugar un papel importante. A medida que las estructuras de materia oscura crecen por acreción suave y fusiones, el gas bariónico inicialmente se encuentran bien mezclado con la materia oscura pero eventualmente se redistribuye hacia una variedad de estructuras más complejas como las que vemos hoy. El gas, originalmente chocado y calentado por el arrastre gravitacional de la materia oscura, se enfría dentro del halo por procesos radiativos, pierde energía y va cayendo hacia el centro del mismo para formar una galaxia. La formación ulterior de estrellas en la galaxia se encuentra regulada por una interacción compleja entre la caída de gas frío y el proceso de calentamiento/expulsión del gas debido a la retroalimentación de las estrellas y sus explosiones, así como eventualmente de un posible Núcleo Galáctico Activo (NGA), proceso que depende del potencial gravitacional dado principalmente

por la *masa de halo*. En halos de baja masa, la retroalimentación estelar, en su mayoría debido a las explosiones de supernova, es capaz no sólo de calentar el medio intereste-lar sino también de expulsar grandes fracciones de gas de la galaxia misma (Dekel & Silk, 1986; Nelson et al., 2015). En halos masivos, por un lado los largos tiempos de enfriamiento del gas calentado por choques en el colapso del halo y, por otro lado, los poderosos procesos energéticos debido a la presencia de agujeros negros supermasivos en fase activa de crecimiento que calientan y/o expulsan el gas, llegan eventualmente a amortiguar la formación estelar de la galaxia y por ende su crecimiento por for-mación estelar in situ (Cattaneo et al., 2009). En este contexto, el paradigma ΛCDM postula de manera natural que las propiedades, distribución espacial y evolución de las galaxias están íntimamente conectadas a la evolución y distribución espacial de los halos oscuros donde residen. En otras palabras, dentro del paradigma ΛCDM, el molde para la formación de galaxias lo provee la materia oscura, aspecto que está a la base de la modelación ab initio de formación y evolución de galaxias, como ser los modelos semi-analíticos y las simulaciones cosmológicas hidrodinámicas; ver reseñas actuales al respecto por ej. en Mo et al., 2010a; Frenk & White, 2012; Somerville & Davé, 2015; Naab & Ostriker, 2017; Lagos et al., 2018.

A pesar de la complejidad y la amplia gama de procesos involucrados en la evolución de las galaxias, es sorprendente que, en general, sigan relaciones bastante regulares en formación estelar, dinámica y estructura con su masa estelar. Lo anterior usualmente se interpreta como un indicativo de que la masa estelar es la propiedad más fundamental de las galaxias. En efecto, la conexión entre la masa estelar de las galaxias y la de los halos de materia oscura, conocida como la relación Masa Estelar-Masa del Halo (MEMH), codifica en gran parte todos aquellos mecanismos que han sido involucrados durante la formación y evolución de las galaxias.

Con base a lo arriba mencionado, es lógico que se hayan desarrollado múltiples en-foques para constreñir la evolución de galaxias conocidos genéricamente como *conexión galaxia-halo* o modelación semi-empírica (v.gr., Cooray & Sheth, 2002; Conroy & Wechsler, 2009; Behroozi et al., 2010; Mo et al., 2010a; Firmani et al., 2010; Moster et al., 2010; Behroozi et al., 2013a; Moster et al., 2013; Rodríguez-Puebla et al., 2017; Wechsler & Tinker, 2018a; Behroozi et al., 2019; Moster et al., 2018). Como veremos más abajo, estos modelos son una verdadera interfase entre lo que es la información observacional de las poblaciones de galaxias y los resultados teóricos de evolución de galaxia ab initio. No obstante, dada la diversidad de las galaxias y la complejidad de sus procesos evolutivos, la descripción de las mismas a un segundo y mayores órdenes de aproximación, requiere de más ingredientes que simplemente una propiedad de los halos (su masa) y una propiedad las galaxias (su masa estelar). A fin de con-
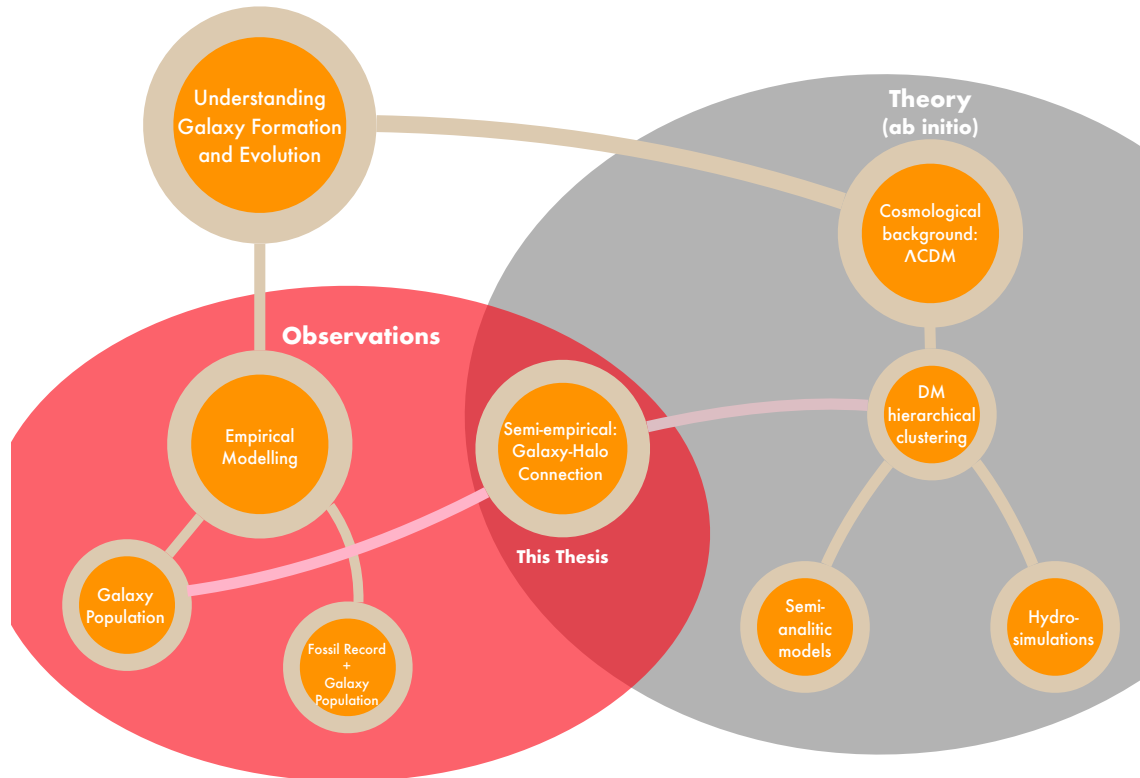
Figure 1.1: Representación esquematica sobre los métodos para modelar la población de galaxias. El enfoque semi-empírico de la conexión galaxia-halo puede ser pensando como la interfaz entre las observaciones y los modelos cosmológicos de formación de estructuras.

siderar esto, se han introducido una gran cantidad de modificaciones y extensiones a los modelos de conexión galaxia-halo. En esta Tesis abordaremos una extensión de dicha conexión para considerar el contenido de gas frío (principalmente hidrógeno atómico y molecular) en las galaxias de bajo corrimiento al rojo. No obstante, para ello, fue necesario antes lograr una descripción homogénea y completa del contenido de gas atómico y molecular de las galaxias locales en función de su masa estelar y morfología.

A continuación se describe someramente los diferentes enfoques para estudiar las propiedades y evolución de galaxias a nivel poblacional, desde los empíricos hasta los teóricos. Esto, con el objetivo de contextualizar el rol que juega la modelación semi-empírica como interfase.

1.1.1 Enfoques para estudiar galaxias a nivel poblacional

La figura 1.1 resume los enfoques para estudiar la formación y evolución de galaxias a nivel poblacional. Por un lado, están los métodos basados sólo en las observaciones

(modelación empírica) y por otro están los métodos teóricos que modelan la evolución de la población de galaxias ab initio, en el contexto cosmológico del ΛCDM arriba descrito e introduciendo la compleja física de los bariones. En la intersección de estos enfoques emergió con fuerza en los últimos años la modelación semi-empírica donde se establece una conexión a nivel estadístico entre la población de galaxias observadas y la población de halos de materia oscura simulados.

Modelación empírica: En las ultimas dos décadas, la astronomía ha entrado en una era de grandes mapeos digitales del cielo a partir de los cuales se pueden inferir de manera homogénea las propiedades fotométricas, parámetros estructurales y morfología para enormes muestras de galaxias; no sólo para poblaciones locales sino que a diferentes corrimientos al rojo, llegando incluso a $z \sim 10$. Estos avances han permitido desarrollar métodos estadísticos puramente empíricos (ver Figura 1.1) basados en la conexión temporal de poblaciones galácticas observadas a diferentes corrimientos al rojo mediante una ecuación de continuidad (Drory & Alvarez, 2008; Peng et al., 2010; Leja et al., 2013, 2015, 2020). Intuitivamente, esta conexión emerge debido a que la evolución de la Función de Masa Estelar de Galaxias brinda información del ensamblaje global de las galaxias (es decir, la cantidad de masa estelar proveniente de la formación estelar in situ y de la obtenida por fusiones, ex situ); mientras que tasa formación estelar establece el ritmo de crecimiento en masa estelar solo in situ. De esta manera, conectando las descripciones estadísticas de las poblaciones de galaxias a diferentes tiempos se logran ciertas inferencias de la evolución promedio de las galaxias. Existe otro tipo de métodos empíricos no basados en la información de "tiempo atrás" sino que en observaciones de la población actual de galaxias con información (principalmente espectral) que permita aplicar el método de registro fósil o arqueológico para reconstruir la historia de formación estelar y crecimiento de masa de las galaxias a través de la síntesis de población estelar (v.gr. Cid Fernandes et al., 2005; Panter et al., 2007, 2008; Ibarra-Medel et al., 2016; Sánchez et al., 2019; Sánchez, 2020). Mientras ambos métodos representan una herramienta valiosa para el estudio detallado del ensamblaje de masa estelar de las galaxias, estos no brindan ninguna información sobre formación de galaxias en el contexto cosmológico.

Modelación Teórica: A partir de primeros principios se sigue la evolución de la población de galaxias formadas en los halos oscuros que crecen en un univeso en expansión. Existen principalmente dos métodos de este tipo:

- **Simulaciones hidrodinámicas:** Calculan las ecuaciones hidrodinámicas, considerando procesos radiativos de calentamiento y enframiento, para el gas inmerso en el campo gravitacional de las estructuras de materia oscura que evolucionan en el régimen no lineal. Debido a su costo computacional, las simu-

laciones hidrodinámicas están limitadas en resolución lo cual impide seguir la amplia gama de procesos y escalas involucrados. En consecuencia, se tienen que usar modelos analíticos de sub-malla que describen la formación estelar, la formación de Agujeros Negros Supermasivos (ANSM), la retroalimentación de las estrellas y los NGAs al medio interestelar y otros. Los modelos sub-malla tienen parámetros libres que deben ser "ajustados" para reproducir las observables que se escojan. Recientes avances han permitido seguir la formación y evolución de las galaxias en volúmenes cosmológicos (v.gr. Hopkins et al., 2014; Vogelsberger et al., 2014; Pillepich et al., 2018; Schaye et al., 2015). Además con técnicas de post-procesado de las simulaciones, se ha podido modelar incluso fases del medio interestelar como ser gas frío de hidrógeno atómico y molecular (ver por ej. Lagos et al., 2015; Diemer et al., 2018, 2019; Popping et al., 2019).

- **Modelos semi-analíticos:** Utiliza recetas fenomenológicas para modelar los procesos físicos relevantes que establecen las propiedades observadas de las galaxias utilizando como esqueleto la formación y evolución de los halos de materia oscura. Estas recetas contienen una gran cantidad de parámetros libres que se ajustan para reproducir las observables que se escojan. Las galaxias en los modelos semi-analíticos son objetos no resueltos caracterizados simplemente por las propiedades integradas de las galaxias, aunque versiones más recientes logran una modelación a nivel de distribuciones radiales azimutalmente promediadas, algo que se había logrado en el pasado con métodos semi-numéricos para galaxias de disco (v.gr. Dalcanton et al., 2007; Avila-Reese et al., 1998; Avila-Reese & Firmani, 2000; Firmani & Avila-Reese, 2000; Dutton et al., 2007). Debido a su versatilidad y relativo bajo costo computacional, los modelos semi-analíticos son una excelente herramienta para generar grandes muestras de galaxias y experimentar diferentes casos (v.gr. White & Frenk, 1991; Baugh, 2006; Somerville & Primack, 1999; Benson, 2012; Somerville & Davé, 2015; Lagos et al., 2018, y muchas más referencias ahí).

A pesar de los logros obtenidos recientemente, ambos métodos tienen limitaciones en introducir los complejos procesos físicos de los bariones y por ende en hacer predicciones de resultados observacionales que son fundamentales para la formulación de una teoría de evolución de galaxias. Por ello, en mayor o menor medida, estos métodos requieren de calibraciones con un conjunto de datos empíricos a fin de hacer predicciones en otros aspectos empíricos. En este sentido, los métodos semi-empíricos se han vuelto claves. Por ejemplo, hoy en día, toda simulación hidrodinámica cosmológica o modelo semi-analítico se suele comparar con la relación MEMH a diferentes épocas

obtenida con los métodos semi-empíricos de conexión galaxia-halo.

Modelación Semi-empírica: Como se mencionó anteriormente, la modelación semi-empírica provee una valiosa interfase entre las observaciones y los modelos teóricos. Gracias al avance en la producción de cajas cosmológicas de N–cuerpos de alta resolución en grandes volúmenes, a partir de las cuales la formación y evolución de los halos de materia oscura se pueden estudiar con gran precisión, así como a los avances en el ensamble de enormes catastros de galaxias observadas, nuevos enfoques semi-empíricos del tipo estadístico han surgido con el propósito de encontrar correlaciones estadísticas entre las propiedades de las galaxias con la de los halos. Esta conexión se logra principalmente mediante dos enfoques estadísticos: (1) la técnica del empate de las abundancias (TEA), que conecta las propiedades de los halos (masa, velocidad circular máxima, historia de formación, etc.) con las propiedades de las galaxias (luminosidad, masa, color, formación estelar, etc.) mediante el empate de sus abundancias (Vale & Ostriker, 2004; Conroy et al., 2006; Conroy & Wechsler, 2009; Behroozi et al., 2010; Rodríguez-Puebla et al., 2012); y (2) el modelo de distribución de la ocupación del halo (DOH) que especifica la función de distribución de probabilidad del número de galaxias de una cierta propiedad dada con las propiedades del halo donde se albergan (Peacock & Smith, 2000; Berlind & Weinberg, 2002; Cooray & Sheth, 2002; Berlind et al., 2003). La TEA se restringe usando principalmente la abundancia de galaxias por su masa estelar y en casos más sofisticados, se separa incluso en galaxias centrales y satélites (v.gr. Yang et al., 2008, 2009a; Rodríguez-Puebla et al., 2012, 2013). Los modelos de DOH se restringen usando principalmente el agrupamiento espacial de galaxias y en algunos casos incluye las funciones de masa estelar condicionales. Ambos enfoques son utilizados principalmente para encontrar la relación MEMH.

Tanto la TEA como el modelo DOH, son dos poderosos y prácticos enfoques estadísticos que proporcionan un entendimiento robusto sobre la conexión galaxia-halo sin conocer previamente los mecanismos físicos detrás de la formación de galaxias. Este enfoque se puede visualizar como la interfaz entre observaciones y el modelo cosmológico (ver Figura 1.1). En consecuencia, este enfoque puede ser utilizado como una herramienta para estudiar la evolución de las propiedades de las galaxias y ultimadamente para restringir los principales mecanismos físicos detrás de la formación de las galaxias. En lo que resta de esta Tesis, nos referiremos a estos enfoques semi-empíricos como simplemente la conexión galaxia-halo[1]

Al restringir la relación entre las galaxias y sus halos, la mayoría de los estudios

---

[1]No obstante, los modelos semi-empríos no toman en cuenta el efecto de los bariones sobre las propiedades de los halos de materia oscura (ver por ejemplo, Beltz-Mohrmann et al., 2020)

supone que la masa estelar o la luminosidad es la principal propiedad de las galaxias que correlaciona con la masa del halo o velocidad circular máxima, es decir, la conexión galaxia-halo es usualmente caracterizada por la función de distribución de probabilidad condicional $\mathcal{P}(g|h)$ donde g se refiere a la masa estelar o luminosidad y h a la masa del halo o su velocidad circular máxima. En realidad, las galaxias tienen diferentes propiedades intrínsecas (color, tasa de formación de estrellas, edad, morfología, etc.), cada una con sus distribuciones y correlaciones con la masa estelar. No obstante, en una primera aproximación al problema de ir más allá de la masa estelar, las galaxias pueden ser acomodadas en dos categorías principales de acuerdo a su propiedad intensiva más relevante, por ejemplo en *galaxias tardías o azules y tempranas o rojas*. Las galaxias tardías son galaxias de disco con poblaciones estelares mayormente de edades intermedias a jóvenes (por ende generalmente azules), con altos contenidos de gas y que aún mantienen una formación activa de estrellas. Las galaxias tempranas son elípticas o lenticulares con poblaciones estelares mayormente viejas (por ende generalmente rojas), con muy bajo contenido de gas y escasa formación estelar. Es claro entonces que estas propiedades no pueden estar determinadas únicamente por la masa del halo de materia oscura en el que residen, sino que, debido a la complejidad del proceso de formación de galaxias, se espera una dependencia con otras propiedades de los halos y el medio ambiente. Dentro del contexto estadístico de la conexión galaxia-halo, lo anterior se puede generalizar al definir la función de distribución de probabilidad condicional conjunta $\mathcal{P}(g_1, ..., g_n | h_1, ..., h_n)$. Aquí g_i representa las propiedades de las galaxias tales como su masa estelar, color, tasa de formación estelar, masa de gas, metalicidad, morfología, etc., mientras que h_i denota las propiedades de los halos de materia oscura, tales como su masa, concentración, tasa de crecimiento, momento angular, etc.

Se han hecho diversos intentos por estudiar la conexión galaxia-halo introduciendo alguna(s) propiedad(es) de las galaxias aparte de la masa estelar. Por ejemplo, en Rodríguez-Puebla et al. (2015) se hicieron inferencias separando a la población de galaxias en azules y rojas de acuerdo a las observaciones y se encontró que la relación MEMH se segrega, es decir es (ligeramente) diferente para un tipo y otro de galaxias. El origen de la conexión galaxia-halo y el establecimiento de las propiedades observables de las galaxias que pueden introducir una segregación en la (estrecha) dispersión de la misma se encuentran entre los problemas sin resolver más importantes en la formación de galaxias. También se han explorado posibles segregaciones de esta relación por propiedades de los halos, tal como su concentración o época característica de formación, algo que se asocia al así llamado "sesgo de ensamblaje" (v.gr. Hearin & Watson, 2013a; Hearin et al., 2016; Zehavi et al., 2019).

Un aspecto de la conexión galaxia-halo que puede enriquecer mucho el entendimiento de las galaxias así como servir de restricciones finas a los modelos semi-analíticos y simulaciones hidrodinámicas, es su extensión para incluir la masa en gas frío (atómico y/o molecular) y por ende lograr así no sólo la relación MEMH, sino que también la relación masa bariónica-masa de halo (la masa bariónica es la suma de masa en estrellas y en gas) y, eventualmente, la relación de masa gaseosa-masa de halo. Lo óptimo es lograr una descripción estadística de funciones multivariadas de la masas estelar, de gas atómico y molecular y del halo (masa virial, que en realidad incluye tanto la masa dominante oscura como la bariónica hasta un radio virial).

1.2 Componente fría gaseosa de las galaxias y extensión de la conexión galaxia-halo para incluir gas

Las galaxias están compuestas tanto de gas frío como de estrellas y las estrellas no siempre son el componente bariónico dominante en una galaxia. De hecho, el metabolismo galáctico que regula en gran parte la evolución de las galaxias se basa en el proceso de acreción de gas frío conducido por el ensamblaje del halo oscuro, su transformación parcial en estrellas y la retroalimentación que sufre por objetos estelares y NGAs. La cantidad resultante de estrellas y gas frío (atómico y molecular, HI y H_2) es entonces un importante indicativo del estado evolutivo de la galaxia. Por otro lado, la suma de masas en estrella y gas es la masa bariónica, $M_{bar} = M_* + M_{gas}$, misma que comparada con la masa total al radio virial, M_{vir}, constituye una pieza clave en el rompecabezas de la conexión galaxia-halo en su contexto cosmológico así como de la formación de galaxias. La razón M_{bar}/M_{vir} (denominada como la fracción bariónica de galaxias) es clave para restringir los modelos de galaxias, particularmente para las galaxias tardías, es decir con morfología dominada por disco (e.g., Firmani & Avila-Reese, 2000; Dutton et al., 2007; Avila-Reese et al., 2008; Dutton et al., 2010).

Mientras que los catastros homogéneos actuales en óptico e infrarrojo han logrado gran profundidad barriendo extensas áreas en el cielo, por lo que abarcan cientos de miles de galaxias locales, los catastros en radio, con los que se mapea la componente de gas atómico HI y de gas molecular H_2, son mucho más limitados en su sensibilidad y cobertura del cielo. Por lo tanto, los catastros ciegos en radio proveen en realidad información poblacional y de correlaciones de galaxias en HI y H_2 limitada y plagada de sesgos. Por ejemplo, es bien sabido que el catastro actual más completo en HI, el Arecibo Legacy Fast ALFA survey (ALFALFA; Giovanelli et al., 2005; Haynes et al.,

2011, 2018), detecta sólo las galaxias más ricas en HI mismas que son típicamente azules y de tipo tardío. En el caso del H_2 la situación es peor aún, debido a las limitaciones observacionales no existen muestreos estadísticos del universo local como ALFALFA para el H_2, sin embargo, se han hecho grandes esfuerzos para tener sondeos que contienen cientos de galaxias tales como xCOLDGASS (Saintonge et al., 2017). En vista de estas dificultades observacionales, no ha sido posible establecer una conexión atendible a nivel de población de galaxias entre sus masas estelares y sus masas en HI y H_2. A fin de estimar correlaciones que implican las masas en HI y H_2, se han elaborado catálogos galácticos incluyendo estas componentes a partir de (1) estudios limitados de seguimientos en radio con cierta profundidad de muestras grandes de galaxias seleccionadas en el óptico/IR o cruzando algunos catastros en radio con catastros (v.gr. Catinella et al., 2010, 2012, 2018; Saintonge et al., 2011; Boselli et al., 2010, 2014a; Papastergis et al., 2012; Kannappan et al., 2013; Stark et al., 2016; van Driel et al., 2016; Butcher et al., 2016); y (2) de inferencias dependientes de modelo basadas, por ejemplo, en las metalicidades de las galaxias o de correlaciones calibradas con propiedades fotométricas (v.gr. Baldry et al., 2008; Zhang et al., 2009; Eckert et al., 2015).

1.2.1 Contribuciones de la Tesis y artículos asociados

Uno de los objetivos que se persiguió en la primera fase de esta Tesis fue justamente lograr la mejor descripción posible de la conexión entre la masa estelar y las masas de HI y H_2 a nivel de la población completa de galaxias en el Universo local. Esto, haciendo uso de la mayor cantidad posible de información observacional disponible, su correcta homogenización, eliminación de posibles sesgos (en particular en lo que concierne a los objetos que no se lograron detectar en radio y para los cuales se reportan límites superiores) y tratamiento estadístico adecuado. Es importante resaltar que en esta tesis hemos optado segregar galaxias por tipo morfológico en vez de color o tasa de fomación estelar primordialmente porque no toda la recopilación observacional utilizada tiene información de estás propiedades. Como resultado de este trabajo, hemos logrado determinar no sólo las relaciones promedio $M_* - M_{HI}$ y $M_* - M_{H_2}$ con sus respectivas dispersiones, separando las galaxias en tempranas y tardías, sino que las distribuciones condicionales completas de M_{HI} y M_{H_2} dada M_*. Estos resultados se reportaron en Calette et al. (2018) y constituyen el Capítulo 2 de esta Tesis. Nuestros resultados fueron usados de inmediato para comparaciones con las predicciones de simulaciones hidrodinámicas cosmológicas, por ejemplo, la simulación TNG-Illustris (Diemer et al., 2019, A.R. Calette es co-autor de este trabajo) así como modelos semi-analíticos, por ejemplo, SHARK (Lagos et al., 2018).

Haciendo uso de las distribuciones condicionales de M_{HI} y M_{H_2} dada la M_* de Calette et al. (2018), se puede usar la Función de Masa Estelar (FME) para mapear entonces las correspondientes Funciones de Masa de HI y de H_2. Para esto se requiere de una FME muy bien determinada, completa hasta las masas más bajas posibles y con sus errores bien evaluados. Como resultado, combinando entonces la FME con las distribuciones condicionales en HI y H_2, se logran funciones bivariadas completas de (M_*, M_{HI}) y (M_*, M_{H_2}) del Universo local, además separando siempre a la población galáctica en dos grupos: galaxias tardías y tempranas. La proyección de estas funciones bivariadas son las Funciones de Masa de HI y H_2. Todos estos resultados fueron reportados en Rodríguez-Puebla et al. (2020, A.R. Calette es segundo autor de este trabajo) y son parte de los Capítulos 3 (determinación completa de la FME desde $M_* \sim 3 \times 10^7$ M$_\odot$hasta $M_* \sim 3 \times 10^{12}$) y Capítulo 4 (Funciones bivariadas, Funciones de Masa y el censo de bariones contenidos en galaxias). Los resultados del Capítulo 3 no son parte medular del proyecto de Tesis pero fue imperioso obtenerlos pues son un ingrediente clave para los cálculos presentados en los capítulos posteriores. Mi contribución a esta parte ha sido a la par del primer autor en lo que refiere a cálculos y desarrollo de códigos pero la conducción del trabajo no fue realizada por mí. El Capítulo 4 es parte estructural del proyecto de Tesis y mi contribución a la parte del artículo relacionado con este capítulo ha sido también a la par del primer autor en cuanto a cálculos y desarrollo de códigos, así como en la conducción del mismo. Particularmente en esta parte he desarrollado la actualización del código en python del capítulo 2 para generar figuras y tablas de 1) la relaciones R_{HI}-M_* y R_{H_2}-M_*, así como su dispersión intrínseca y 2) las funciones condicionales de distribución de R_{HI} and R_{H_2} así como sus momentos (promedio y desviación estándar).

Como se ha mencionado en la Sección anterior, el enfoque semiempírico con el que se logra la conexión galaxia-halo se ha consolidado como una herramienta poderosa para constreñir los procesos astrofísicos y evolutivos de las galaxias y para poner a prueba el paradigma cosmológico subyacente. No obstante la gran mayoría de los estudios dentro de la literatura se han focalizado en restringir la relación MEMH y algunos en la relación $M_{\mathrm{HI}}-M_h$ y unos pocos a la relación $M_{\mathrm{bar}}-M_h$. En los Capítulos 5 y 6 de esta Tesis se busca lograr una conexión galaxia-halo lo más completa posible incluyendo gas atómico, además separando en galaxias con morfologías tempranas y tardías. Para esto, haremos uso de la descripción estadística completa de las galaxias locales por su masa estelar y de gas en HI que logramos en los Capítulos 2-4. Debido a que la conexión galaxia-halo implicará la generación de un catálogo sintético basado en una población de millones de halos oscuros de una simulación cosmológica de N cuerpos, tendremos también la posibilidad de ofrecer una descripción semi-empírica

del acumulamiento espacial de las galaxias en función de M_* y M_{HI}, incluyendo si son centrales o satélites.

1.3 Objetivos

Esta Tesis se enfoca en lograr una descripción estadística completa de la población local de galaxias separada en tardías y tempranas en lo que respecta a sus masas estelares, en HI, H_2, gas frío total, y bariónicas. Esta descripción incluye las relaciones de escala de masa estelar con las fracciones de gas y las distribuciones condicionales completas de estas masas de gas dada la masa estelar, las Funciones de Masa Estelar, de HI, de H_2 y bariónicas, así como las funciones bivariadas de (M_*, M_{HI}) y (M_*, M_{H_2}). Con esta información se buscará extender el análisis de la conexión galaxia-halo, a través del enfoque de la TEA y el sembrado de galaxias sintéticas en una simulación de N cuerpos, para determinar cómo las galaxias tardías y tempranas pueblan sus halos de materia oscura cuando son caracterizadas por su masa estelar y masa de hidrogeno atómico. Los resultados de esta tesis serán claves para restringir los mecanismos más relevantes que gobernaron la formación y evolución de las galaxias. Los objetivos más relevantes propuestos para esta Tesis son:

- Compilar una muestra de galaxias locales, $z \sim 0$, que contenga información sobre gas atómico y molecular, masa estelar y morfología.

- Estudiar los efectos de selección en nuestra muestra y estudiar el impacto de las no detecciones. Desarrollar un catálogo sintético de galaxias para proponer una corrección a las no detecciones.

- Determinar las relaciones HI-M_* y H_2-M_* y sus dispersiones para galaxias tempranas y tardías, así como las distribuciones condicionales completas de M_{HI} y M_{H_2} dado M_*.

- Determinar la FME usando el SDSS DR7 desde galaxias enanas $M_* \sim 3 \times 10^7 M_\odot$ hasta galaxias masivas $M_* \sim 10^{12} M_\odot$ para todas las galaxias y dividido en tempranas y tardías. Lo anterior implica lograr los siguientes sub-objetivos:

  1. Estudiar las correcciones K para galaxias locales.

  2. Desarrollar una metodología para estimar el número de galaxias de baja masa no observadas por efectos del brillo superficial.

  3. Desarrollar una metodología para corregir la FME por fluctuaciones en el conteo debido a la estructura a gran escala del universo.

11

4. Desarrollar una metodología para deconvolucionar por errores a la FME y estudiar el impacto de errores sistemáticos.

- Utilizando la FME determinada y las districiones condicionales de M_{HI} y M_{H_2} dado M_*, determinar las distribuciones bivariadas M_{HI} y M_{H_2} en función de la masa estelar para galaxias tardías y tempranas, así como la población completa.

- Utilizar los resultados de la FME para determinar la función de masa de HI y H_2, gas frío y bariones y a su vez divididas en tardías y tempranas.

- Determinar la densidad cósmica en masa estelar, HI, H_2, gas frío total y en masa de bariones en galaxias tardías y tempranas y para el total.

- Usando la muestra reciente de galaxias locales xGASS (Catinella et al., 2018) y después de aplicarle el tratamiento estadístico adecuado, cuantificar las desviaciones que sufren las distribuciones condicionales de M_{HI} dado M_* por el efecto de ser galaxia central o satélite. Con base a esto, introducir las correcciones pertinentes a nuestras determinaciones de la relación M_{HI}-M_* y su distribución para galaxias centrales y satélites y a su vez divididas en galaxias tardías y tempranas.

- Generar un catálogo sintético de galaxias utilizando simulaciones de $N-$cuerpos y a partir de la relación semi-empírica $M_* - V_{max}$ de nuestra conexión galaxia-halo. Utilizar las fracciones de galaxias tempranas y tardías para asignar morfologías y utilizar las distribución bivariada de HI descrita en el item anterior para asignar gas atómico a las galaxias.

- A partir del catálogo sintético determinar la relaciones $M_{HI} - V_{max}$ y sus dispersiones para galaxias centrales, satélites y divididas en tempranas y tardías. De manera similar para las relaciones $M_{HI} - M_{vir}$.

- Predecir con el catálogo sintético la función de correlación de 2 puntos en función de la masa estelar y en función de la masa en HI. Comprobar si estas funciones están de acuerdo con las observaciones. Explorar los posibles sesgos del catastros ALFALFA en cuanto a la función de correlación de 2 puntos que se miden del mismo.

1.4   Contenido

Esta tesis se divide en 7 Capítulos.

- En el capítulo 2 se presenta y discute la muestra compilada para galaxias con información en HI y H_2. En este capítulo homogeneizamos la muestra a una misma función inicial de masa, cosmología, factor de conversión de CO a luminosidad y tomamos en cuenta efectos de selección. Adicionalmente, tomamos en cuenta los límites superiores reportadas en las observaciones en radio y corregimos por efectos de distancia y sensibilidad. Los datos reportados en este capítulo utiliza el estimador estadístico de Kaplan-Meier para datos censurados. Por último se presenta las relaciones HI-M_* y H_2-M_* y sus distribuciones para galaxias tempranas y tardías a $z = 0$ de la muestra compilada. Con base a catálogos sintéticos de galaxias, mostramos que con la caracterización de la muestra compilada es posible determinar las funciones de masa en HI y H_2 para galaxias tempranas y tardías.

- En el capítulo 3 utilizamos la muestra de galaxias del SDSS DR7 y una submuestra de galaxias cercanas del SDSS DR4 para determinar la función de masa de galaxias desde $M_* \sim 3 \times 10^7 M_\odot$ hasta galaxias masivas $M_* \sim 3 \times 10^{12} M_\odot$. En este capítulo proponemos una serie de mejoras al estimador clásico $1/V_{max}$, en particular a su sensibilidad a la estructura a gran escala del universo, y desarrollamos una metodología para corregir a la función de masa de galaxias por efectos de brillo superficial.

- En el capítulo 4 se discute el formalismo de la función condicional para determinar las funciones de masa de HI, H_2, gas frío y bariónica de las galaxias tempranas y tardías. Se discute también la contribución de las galaxias a la densidad cósmica en masa estelar, HI, H_2, gas frío y de bariones. En este capítulo también se discute el impacto de errores sistemáticos en las funciones de masa y se deconvolucionan por errores sistemáticos. Adicionalmente contrastamos funciones de masa de las componentes mencionadas con nuestro formalismo presentado en este Capítulo contra simulaciones hidrodinámicas de volumenes cosmológicos. Usamos información de la componente estelar, HI, y H_2 de la simulación IllustrisTNG en dos cajas cúbicas que tienen tamaños de 100 y 300 Mpc por lado.

- En el capítulo 5 analizamos el catastro reciente de galaxias con información de contenido de gas HI "extended GASS" (xGASS) con el mismo tratamiento de las no detecciones (límites superiores) y análisis estadístico de supervivencia del Capítulo 2. La particularidad de xGASS es que al ser una muestra tomada del catastro óptico SDSS, cuenta con información sobre si las galaxias son centrales o satélites. De esta manera estudiando cuánto depende de esto el contenido de gas

HI de las galaxias tardías y tempranas. Usando los resultados que inferimos para xGASS, introducimos "correcciones" basadas en el mismo a nuestros resultados de los Capítulos 2 y 4 (las distribuciones condicionales de HI dada la masa estelar) para lograr que los mismos se refieran también a galaxias centrales y satélites, con lo cual incluimos el medio ambiente local a nuestra descripción estadística de la población de galaxias a $z \sim 0$

- En el capítulo 6 generamos un catálogo sintético de galaxias utilizando los catálogos de halos de materia oscura de la simulación Small Multidark Planck (Rodríguez-Puebla et al., 2016). Este catálogo implica la conexión galaxia-halo a nivel estelar y de gas atómico. Para lograr lo anterior, utilizamos la función de masa de galaxias restringida en el capítulo 3 y las distribuciones condicionales de HI y H_2 del capítulo 4, con las correcciones adecuadas encontradas en el capítulo 5 para introducir separaciones en galaxias centrales y satélites, mismas que pueblan halos principales y subhalos, respectivamente. Adicionalmente, utilizamos un catálogo de grupos de galaxias (Yang et al., 2012) para dividir a las galaxias en centrales y satélites, tanto para morfologías tempranas como tardías. Esto con el fin de asignar correctamente las abundancias de galaxias observadas en función de su ambiente y su morfología. Obtendremos una descripción estadística completa de los contenidos de masa estelar y gaseosa de las galaxias en función de los halos o subhalos que habitan. Predeciremos la función de correlación de dos puntos en función de la masa en HI y mostramos que haber introducido el contenido de HI por separado para centrales y satélites fue importante. Se presentará una comparación con la observaciones y exploraremos los sesgos que ellas introducen por los efectos de selección.

- Finalmente, el capítulo 7 enumeramos las conclusiones más importantes alcanzadas en esta tesis y plantemos el trabajo a futuro.

Chapter 2

The HI- and H$_2$-to-stellar mass correlations of late- and early-type galaxies and their consistency with the observational mass functions

ABSTRACT

In this Chapter we compile and carefully homogenize local galaxy samples with available information on stellar, HI and/or H$_2$ masses. After taking into account gas non-detections, we determine the HI- and H$_2$-to-stellar mass relations and their 1σ scatter for late- and early-type galaxies. These relations are fitted to single or double power laws. Late-type galaxies are significantly gas richer than early-type ones, specially at high masses. The H$_2$-to-HI mass ratios as a function of M_* are discussed. We constrain the distribution functions of the HI- and H$_2$-to-stellar mass ratios. We find that they can be described by a Schechter function for late types and a (broken) Schechter + uniform function for early types. Using the observed galaxy stellar mass function and the volume-complete late-to-early-type galaxy ratio as a function of M_*, these distributions are mapped into HI and H$_2$ mass functions. The obtained mass functions are consistent with those inferred from large surveys. The results presented here can be used to constrain models and simulations of galaxy evolution.

2.1 Introduction

Galaxies are complex systems, formed mainly from the cold gas captured by the gravitational potential of dark matter halos and transformed into stars, but also reheated and eventually ejected from the galaxy by feedback processes (see for a recent review Somerville & Davé, 2015). Therefore, the content of gas, stars, and dark matter of galaxies provides key information to understand their evolution and present-day status, as well as to constrain models and simulations of galaxy formation

(see e.g., Zhang et al., 2009; Fu et al., 2010; Lagos et al., 2011; Duffy et al., 2012; Lagos et al., 2015).

Local galaxies fall into two main populations, according to the dominion of the disk or bulge component (late- and early-types, respectively; a strong segregation is also observed by color or star formation rate). The main properties and evolutionary paths of these components are different. Therefore, the present-day stellar, gaseous, and dark matter fractions are expected to be different among late-type/blue/star-forming and early-type/red/passive galaxies of similar masses. The above demands the gas-to-stellar mass relations to be determined separately for each population. Morphology, color and star formation rate correlate among them, though there is a fraction of galaxies that skips the correlations. In any case, when only two broad groups are used to classify galaxies, the segregation in the resulting correlations for each group is expected to be similar for any of these criteria. Here we adopt the morphology as the criterion for classifying galaxies into two broad populations.

With the advent of large homogeneous optical/infrared surveys, the statistical distributions of galaxies, for example the galaxy stellar mass function (GSMF), are very well determined now. In the last years, using these surveys and direct or statistical methods, the relationship between the stellar, M_*, and halo masses has been constrained (e.g., Mandelbaum et al., 2006; Conroy & Wechsler, 2009; More et al., 2011; Behroozi et al., 2010; Moster et al., 2010; Rodríguez-Puebla et al., 2013; Behroozi et al., 2013a; Moster et al., 2013; Zu & Mandelbaum, 2015). Recently, the stellar-to-halo mass relation has been even inferred for (central) galaxies separated into blue and red ones by Rodríguez-Puebla et al. (2015). These authors have found that there is a segregation by color in this relation (see also Mandelbaum et al., 2016). The semi-empirical stellar-to-halo mass relation and its scatter provide key constraints to models and simulations of galaxy evolution. These constraints would be stronger if the relations between the stellar and atomic/molecular gas contents of galaxies are included. With this information, the galaxy baryonic mass function can be also constructed and the baryonic-to-halo mass relation can be inferred, see e.g, Baldry et al. (2008).

While the stellar component is routinely obtained from large galaxy surveys in optical/infrared bands, the information about the cold gas content is much more scarce due to the limits in sensitivity and sky coverage of current radio telescopes. In fact, the few blind HI surveys, obtained with a fixed integration time per pointing, suffer of strong biases, and for H_2 (CO) there are not such surveys. For instance, the HI Parkes All-Sky Survey (HIPASS; Barnes et al., 2001; Meyer et al., 2004) or the Arecibo Legacy Fast ALFA survey (ALFALFA; Giovanelli et al., 2005; Haynes et al., 2011; Huang

16

et al., 2012b), miss galaxies with low gas-to-stellar mass ratios, specially at low stellar masses. Therefore, the HI-to-stellar mass ratios inferred from the crossmatch of these surveys with optical ones should be regarded as an upper limit (see e.g., Baldry et al., 2008; Papastergis et al., 2012; Maddox et al., 2015). In the future, facilities such as the Square Kilometre Array (SKA; Carilli & Rawlings, 2004; Blyth et al., 2015) or precursor instruments such as the Australian SKA Pathfinder (ASKAP; Johnston et al., 2008) and the outfitted Westerbork Synthesis Radio Telescope (WSRT), will bring extragalactic gas studies more in line with optical surveys. Until then, the gas-to-stellar mass relations of galaxies can be constrained: i) from limited studies of radio follow-up observations of large optically-selected galaxy samples or by cross-correlating some radio surveys with optical/infrared surveys (e.g., Catinella et al., 2012; Saintonge et al., 2011; Boselli et al., 2010; Papastergis et al., 2012); and ii) from model-dependent inferences based, for instance, on the observed metallicities of galaxies or from calibrated correlations with photometrical properties (e.g., Baldry et al., 2008; Zhang et al., 2009).

While this thesis does not present new observations, it can be considered as an extension of previous efforts in attempting to determine the HI-, H_2- and cold gas-to-stellar mass correlations of local galaxies over a wide range of stellar masses. Moreover, here we separate galaxies into at least two broad populations, late- and early-type galaxies (hereafter LTGs and ETGs, respectively). These empirical correlations are fundamental benchmarks for models and simulations of galaxy evolution. Our main goal here is to constrain these correlations by using and uniforming large galaxy samples of good quality radio observations with confirmed optical counterparts. Moreover, the well determined local GSMF combined with these correlations can be used to construct the galaxy HI and H_2 mass functions, GHIMF and GH_2MF, respectively. As a test of consistency, we compare these mass functions with those reported in the literature for HI and CO (H_2).

Many of the samples compiled here suffer of incompleteness and selection effects or in many cases the radio observations provide only upper limits to the flux (non detections). To provide reliable determinations of the HI- and H_2-to-stellar mass correlations, for both LTGs and ETGs, here we homogenize as much as possible the data, check them against selection effects that could affect the calibration of the correlations, and take into account the upper limits adequately. We are aware of the limitations of this approach (given the heterogeneous nature of our compiled data, such as selection effects, incompleteness, etc.). Note, however, that in absence of large homogeneous galaxy surveys reporting gas scaling relations over a wide dynamical range and separated into late- and early-type galaxies, the above approach is well

17

supported as well as their fair use.

2.2 Compilation of the Observational Data

The main goal of this section is to present our extensive compilation of observational studies (catalogs, surveys or small samples) that meet the following criteria:

- Include HI and/or H_2 masses from radio observations, and luminosities/stellar masses from optical/infrared observations.

- Provide the galaxy morphological type or a proxy of it.

- Describe the selection criteria of the sample and provide details about the radio observations, flux limits, etc.

- Include individual distances to the sources and corrections for peculiar motions/large-scale structures for the nearby galaxies.

- In the case of non-detections, provide estimates of the upper limits for HI or H_2 masses.

Table 2.1: Observational samples

Sample	Selection	Environment	HI	Detections / Total	H_2	Detections / Total	IMF	Category
UNGC	ETG+LTG	local 11 Mpc	Yes	407 / 418	No	–	diet-Salpeter	Gold
GASS/COLD GASS	ETG+LTG	no selection	Yes	511 / 749	Yes	229 / 360	Chabrier (2003)	Gold
HRS–field	ETG+LTG	no selection	Yes	199 / 224	Yes	101 / 156	Chabrier (2003)	Gold
ATLAS³ᴰ–field	ETG	field	Yes	51 / 151	Yes	55 / 242	Kroupa (2001)	Gold
NFGS	ETG+LTG	no selection	Yes	163 / 189	Yes	27 / 31	Chabrier (2003)	Silver
Stark et al. (2013) compilation*	LTG	no selection	Yes	62/62	Yes	14 / 19	diet-Salpeter	Silver
Leroy+08 THINGS/HERACLES	LTG	nearby	Yes	23 / 23	Yes	18 / 20	Kroupa (2001)	Silver
Dwarfs-Geha+06	LTG	nearby	Yes	88 / 88	No	–	Kroupa et al. (1993)	Silver
ALFALFA dwarf	ETG+LTG	no selection	Yes	57 / 57	No	–	Chabrier (2003)	Silver
ALLSMOG	LTG	field	No	–	Yes	25 / 42	Kroupa (2001)	Silver
Bauermeister et al. (2013) compilation	LTG	field	No	–	Yes	7 / 8	Kroupa (2001)	Silver
ATLAS³ᴰ–Virgo	ETG	Virgo core	Yes	2 / 15	Yes	4 / 21	Kroupa (2001)	Bronze
AMIGA	ETG+LTG	isolated	Yes	229 / 233	Yes	158 / 241	diet-Salpeter	Bronze
HRS–Virgo	ETG+LTG	Virgo core	Yes	55 / 82	Yes	36 / 62	Chabrier (2003)	Bronze
UNAM-KIAS	ETG+LTG	isolated	Yes	352 / 352	No	–	Kroupa (2001)	Bronze
Dwarfs-NSA	LTGs	isolated	Yes	124 / 124	No	–	Chabrier (2003)	Bronze

* From this compilation, we considered only galaxies that were not in GASS, COLD GASS and ATLAS³ᴰ samples.

2.2.1 Systematic Effects on the HI- and H_2-to-stellar mass correlations

To reduce potential systematic effects that can bias how we derive the HI- and H_2-to-stellar mass correlations we homogenize all the compiled observations to a same basis. Following, we discuss some potential sources of bias/segregation and the calibration that we apply to the observations. It is important to stress that for inferring scaling

correlations, as those of the gas fraction as a function of stellar mass, is important to have a statistically representative sample of galaxies *at each mass bin.* Thus, there is no need to have mass limited volume-complete samples (see also subsection 2.5.1). However, a volume-complete sample assures that possible biases on the measure in question due to selection functions in galaxy type, color, environment, surface brightness, etc., are not introduced. The main expected bias in the gas content at a given stellar mass is due to the galaxy type/color, or SFR; this is why we need to separate the samples into at least two broad populations, LTGs and ETGs.

Galaxy type

The gas content of galaxies, at a given M_*, segregates significantly with galaxy morphological type (e.g., Kannappan et al., 2013; Boselli et al., 2014b). Thus, information on morphology is necessary in order to separate galaxies into at least two broad populations, LTGs and ETGs. Besides its physical basis, this separation is important for not introducing biases in the obtained correlations due to selection effects related to the morphology in the different samples used here. For example, some samples are only for late-type or star-forming galaxies, others only for early-type galaxies, etc., therefore by combining them without a separation by morphology would yield correlations that are not statistically representative. We consider ETGs those classified as ellipticals (E), lenticulars (S0), dwarf E, and dwarf spheroidals or with $T < 1$, and LTGs those classified as Spirals (S), Irregulars (Irr), dwarf Irr, and blue compact dwarfs or with $T \geq 1$[1]. The morphological classification criteria used in the different samples are diverse, from individual visual evaluation to automatic classification methods as the one by Huertas-Company et al. (2011). We are aware of the high level of uncertainty introduced by using different morphological classification methods. However, in our case the morphological classification is used for separating galaxies just into two broad groups. Therefore, such an uncertainty is not expected to affect significantly any of our results. It is important to highlight that the terms LTG and ETG are useful only as qualitative descriptors. These descriptors *should not be applied to individual galaxies, but instead to two distinct populations of galaxies in a statistical sense.*

Environment

The gas content of galaxies is expected to depend on environment (e.g., Zwaan et al., 2005; Geha et al., 2012; Jones et al., 2016; Brown et al., 2017). In this study we are

---

[1]T is the numerical value of the de Vaucouleurs morphological classification

not in a position of studying in detail such a dependence, though our separation into LTG and ETG populations partially takes into account this dependence because these populations are segregated by environment (e.g., Dressler, 1980; Kauffmann et al., 2004; Blanton et al., 2005b; Blanton & Moustakas, 2009, and references therein). In any case, in our compilation we include three samples specially selected to contain very isolated galaxies (AMIGA, Geha et al. and Bradford et al. galaxies from SDSS) and one subsample of galaxies from the Virgo Cluster central regions. We will check whether their HI and H_2 mass fractions significantly deviate or not from the mean relations.

Systematic Uncertainties on the Stellar Masses

There are many sources of systematic uncertainty in the inference of the stellar masses related to the choices of: initial mass function (IMF), stellar population synthesis and dust attenuation models, star formation history parametrization, metallicity, filter setup, etc. For inferences from broad-band spectral energy distribution fitting and using a large diversity of methods and assumptions, Pforr et al. (2012) estimate a maximal variation in stellar mass calculations of ~ 0.6 dex. The major contribution to these uncertainties comes from the IMF (see also Marchesini et al., 2009). The IMF can introduce a systematic variation up to ≈ 0.25 dex (see e.g., Conroy, 2013). For local normal galaxies and from UV/optical/IR data (as it is the case of our compiled galaxies), Moustakas et al. (2013) find a mean systematic differences between different mass-to-luminosity estimators (fixed IMF) less than 0.2 dex (see also, , Robotham+2020). We have seen that in most of the samples compiled here, the stellar masses are calculated using roughly similar mass-to-luminosity estimators, but the IMF are not always the same.Therefore, we homogenize the reported stellar masses in the different compiled samples to the mass corresponding to a Chabrier (2003) initial mass function (IMF), and neglect other sources of systematic differences.

Other effects

We also homogenize the distances to the value of $H_0 = 70$ kms^{-1} Mpc^{-1}. In most of the samples compiled here (at least the most relevant ones for our study), *distances were corrected for peculiar motions and large-scale structure effects*. When the authors included helium and metals to their reported HI and H_2 masses, we take care in subtracting these contributions. When we calculate the total cold gas mass, then helium and metals are explicitly taken into account.

Categories

The different HI and H$_2$ samples used in this thesis are wide in diversity, in particular they were obtained with different selection functions, radio telescopes, exposure times, etc. We have divided the different samples into three categories according to the feasibility of each one for determining robust and statistically representative HI- or H$_2$-to-stellar mass correlations for the LTG and ETG populations. We will explore whether the less feasible categories should be included or not for determining these correlations. The three categories are:

1. **Golden:** It includes datasets based on volume-complete (above a given luminosity/mass) samples or on representative galaxies selected from volume-complete samples. The Golden datasets, by construction, are unbiased samples of the galaxy properties distribution.

2. **Silver:** It includes datasets from galaxy samples that are not volume complete but that are attempted to be statistically representative at least for their morphological groups, i.e., these samples do not present obvious or strong selection effects.

3. **Bronze:** This category is for samples selected deliberately by environment, and it will be used to explore the effects of environment on the LTG and ETG HI- or H$_2$-to-stellar mass correlations.

2.2.2 The compiled HI sample

Table 2.2: Number of galaxies with detections and upper limits by morphology

Morphology(%)	Detections(%)	Upper limits(%)	Total
	HI data		
LTG (78%)	1975 (94%)	121 (6%)	2096
ETG (22%)	292 (50%)	288 (50%)	580
	H$_2$ data		
LTG (63%)	533 (75%)	180 (25%)	713
ETG (37%)	124 (29%)	298 (71%)	422

Appendix A presents a summary of the HI samples compiled in this chapter (see also Table 2.1). Table 2.2 lists the total numbers and fractions of compiled galaxies with detection and non detection for each galaxy population. Table 2.3 lists the number of detected and non-detected galaxies for the golden, silver, and bronze categories listed above (§§2.2.1).

Table 2.3: Number of galaxies with detections and upper limits by category

Category (%)	Detections (%)	Upper limits (%)	Total
	HI data		
Golden (58%)	1168 (76%)	374 (24%)	1542
Silver (16%)	391 (94%)	26 (6%)	417
Bronze (26%)	708 (99%)	9 (1%)	717
	H_2 data		
Golden (67%)	385 (51%)	373 (49%)	758
Silver (10%)	91 (76%)	29 (24%)	120
Bronze (23%)	181 (70%)	76 (30%)	257

Figure 2.1 shows the mass ratio $R_{\mathrm{HI}} \equiv M_{HI}/M_*$ vs. M_* for the compiled samples. Note that we have applied some corrections to the reported samples (see above) to homogenize all the data. The upper and bottom left panels of Figure 2.1 show, respectively, the compilations for LTGs and ETGs. The different symbols indicate the source reference of the data and the downward arrows are the corresponding upper limits on the HI-flux for non-detections. We also reproduce the mean and standard deviation in different mass bins as reported in Maddox et al. (2015) for a cross-match of the ALFALFA and SDSS surveys. As mentioned in the Introduction, the ALFALFA survey is biased to high R_{HI} values, specially towards the low mass side. Note that the small ALFALFA subsample of dwarf galaxies by Huang et al. (2012a, dark purple dots) was selected as an attempt to take into account low-HI mass galaxies in the low-mass end.

2.2.3   The compiled H_2 sample

Obtaining H_2 masses directly from observational measurements of molecular hydrogen fluxes in galaxies is not possible. H_2 is a diatomic molecule with identical nuclei, thus, it does not have a permanent dipole moment neither dipolar rotational transitions. H_2 purely rotational quadrupole transitions are the lowest energy transitions in the far infrared (FIR) and weak due to their spontaneous decay lifetimes ($t_{\mathrm{decay}} \sim 100 years$). Other H_2 transitions such as the ortho and para or the lowest vibrational energy transitions need temperatures $T > 100$ K. Therefore, given that H_2 is formed in cold molecular clouds ($T \sim 10 - 30$ K), cold H_2 emission from the ISM of galaxies is practically absent, hence a tracer of the cold H_2 abundance should be used.

The best tracer from the observational point of view is the CO molecule due to its relatively high abundance and its low excitation energy. The H_2 mass is related
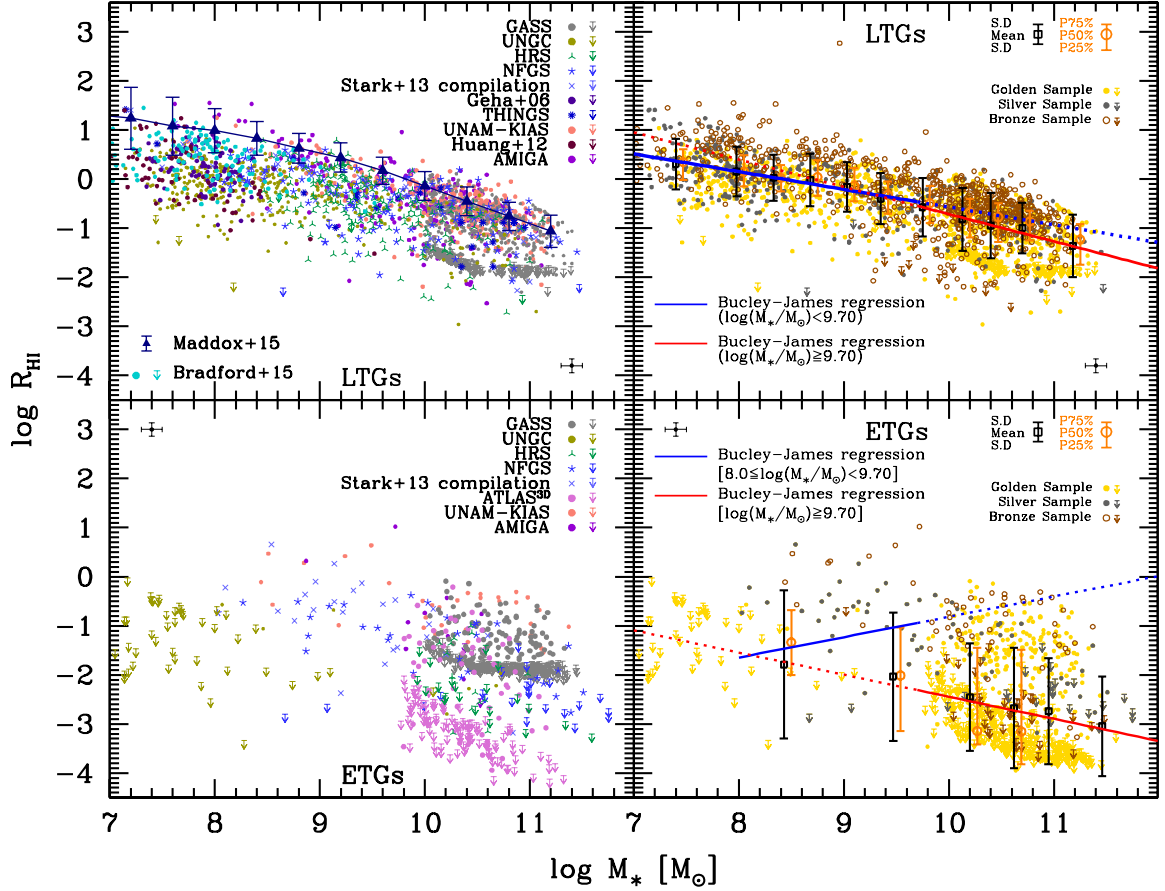
Figure 2.1: Atomic gas-to-stellar mass ratio as a function of M_*. *Upper panels:* Compiled and homogenized data with information on R_{HI} and M_* for LTGs (the different sources are indicated inside the left panel; see Appendix A for the acronyms and authors); downward arrows show the reported upper limits for non detections. The blue triangles with thin error bars are mean values and standard deviations from the v.40 ALFALFA and SDSS cross-match according to Maddox et al. (2015); the ALFALFA galaxies are biased to high values of R_{HI} (see text). Right panel is the same as left one, but with the data separated into three categories: Golden, Silver, and Bronze (yellow, gray, and brown symbols, respectively). The red and blue lines are Buckley-James linear regressions (taking into account non-detections) for the high- and low-mass sides, respectively; the dotted lines show extrapolations from these fits. Squares with error bars are the mean and standard deviation of the data in different mass bins, taking into account non-detections by means of the Kaplan-Meier estimator. Open circles with error bars are the corresponding median and 25-75 percentiles. Estimates of the observational uncertainties are showed in the panel corners (see text). *Lower panels:* Same as in upper panels but for ETGs. In the right panel, we have corrected by distance the galaxies with upper limits from GASS to make them consistent with the distances of the ATLAS[3D] sample (see text), and the upper limits from the latter, where increased by a factor of two to homogenize them to the ALFALFA instrument and signal-to-noise criteria. For the bins where more than 50% of the data are upper limits, the median and percentiles are not calculated.
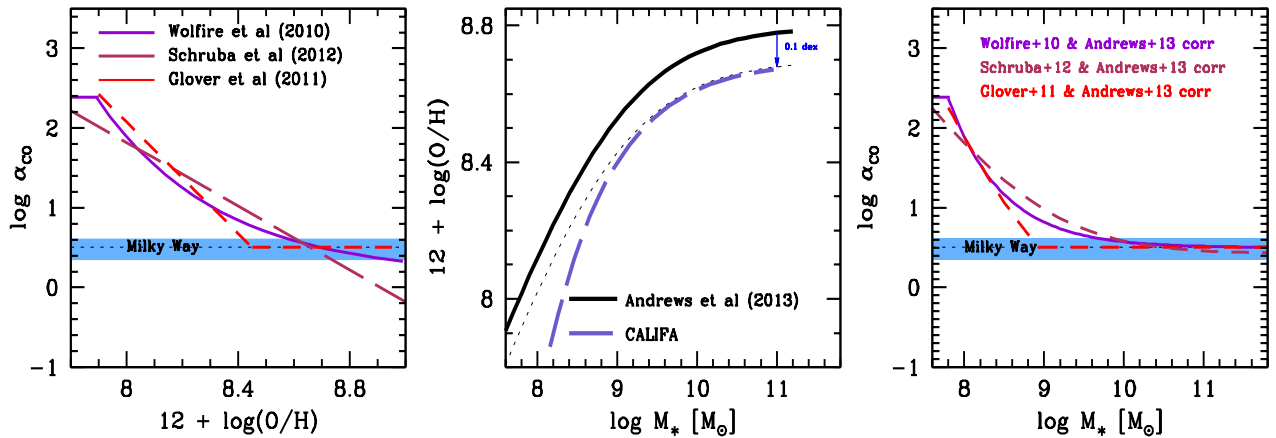
Figure 2.2: *Left panel:* Dependence of the CO-to-H$_2$ factor on gas-phase metallicity as given by physical models (Wolfire et al., 2010; Glover & Mac Low, 2011) calibrated by observations and by a pure empirical approach (Schruba et al., 2012). Observations do not allow to constrain these relations for metallicities lower than $12 + \log_{10}(O/H) \sim 7.9$ *Middle panel:* Dependence of metallicity on mass according to the CALIFA (Sánchez et al., 2013) and SDSS (Andrews & Martini, 2013) surveys. We use an updated relation for CALIFA that includes more galaxies, specially at low masses (S. Sanchez, priv. communication); the masses were corrected from Salpeter to Chabrier IMF. The dotted line is the SDSS relation lowered by 0.1 dex to correct for the aperture effect; notice how well it agrees with the CALIFA relation but it extends to lower masses, so this is the relation we use. *Right panel:* Dependence of the CO-to-H$_2$ factor on mass inferred from the α_{CO}–M_* and $Z - M_*$ dependences plotted in the other panels.

to the CO luminosity through a CO-to-H$_2$ conversion factor: $M_{H_2} = \alpha_{CO}L_{CO}$. This factor has been determined in molecular clouds in the Milky Way (MW), $\alpha_{CO,MW} = 3.2$ (K km s^{-1} pc^{-1})$^{-1}$, with a systematic uncertainty of 30%. It was common to assume that this conversion factor is the same for all galaxies. However, several pieces of evidence show that α_{CO} is not constant, and it depends mainly on the gas-phase metallicity, increasing as the galaxy metallicity decreases (e.g., Boselli et al., 2002; Schruba et al., 2012; Narayanan et al., 2012; Bolatto et al., 2013, and more references therein). As first-order, α_{CO} changes slowly for metallicities larger than $12 + \log_{10}(O/H) \sim 8.4$ (approximately half the solar one) and increases considerably as the metallicity decreases.

In a recent review on the topic, among the several approaches for determining the dependence of α_{CO} on Z in galaxies, Bolatto et al. (2013) recommend to adopt a prescription based on a local physical model for the H$_2$ and CO production and calibrate it with extragalactic observations. In particular, they find that the prescription given in Wolfire et al. (2010), based on photodissociation models with shielding, is the most consistent with the scarce observational data that provides α_{CO} vs. Z in galaxies.

According to Wolfire et al. (2010):

$$\alpha_{\mathrm{CO}} = \alpha_{\mathrm{CO,MW}} \exp\left[\frac{+4.0\Delta A_V}{Z'\bar{A}_{V,MW}}\right] \exp\left[\frac{-4.0\Delta A_V}{\bar{A}_{V,MW}}\right] \qquad (2.2.1)$$

where $\alpha_{\mathrm{CO,MW}} = 3.2$ (in units $M_\odot$ pc^{-2}/K km s^{-1}) is the adopted conversion factor for the Milky Way, $Z' = Z/Z_\odot$ where $\log Z \equiv 12 + \log_{10}(\mathrm{O/H})$, $\Delta A_V \approx 1$, and $\bar{A}_{V,MW}$ is the mean extinction through a giant molecular cloud at Milky Way metallicity $Z_\odot$, with $\bar{A}_{V,MW} \approx 5$ for $\Sigma_{\mathrm{GMC}} \approx 100$ $M_\odot$pc^{-2}. According to Eq. (2.2.1), $\alpha_{\mathrm{CO}} \approx \alpha_{\mathrm{CO,MW}}$ for $Z \gtrsim Z_\odot$. The left panel of Fig. 2.2 shows the Wolfire et al. (2010) relation along with those of Glover & Mac Low (2011) and Schruba et al. (2012).

To relate α_{CO} with stellar mass, we use the mass-metallicity relation for galaxies in the local Universe. Sánchez et al. (2013) and Andrews & Martini (2013) determined the mass-metallicity relation for galaxies using the Calar Alto Legacy Integral Field Area Survey (CALIFA) and SDSS surveys in the stellar mass range $8.4 \leq \log(M_*/M_\odot) \leq 11.2$ and $7.4 \leq \log(M_*/M_\odot) \leq 11.2$, respectively. The work by Sánchez et al. (2013) provides a more reliable estimate of the mass-metallicity relation; recall that the SDSS galaxies are mapped by only one central fiber of fixed aperture, while CALIFA maps the whole galaxies with many integral field units. However, the mass range in the CALIFA sample is limited, while Andrews & Martini (2013) extends to very low masses. We use an updated version of the CALIFA mass-metallicity relation (S. F. Sanchez, priv. communication) and correct M_* to pass from the Salpeter IMF to the Chabrier one used in Andrews & Martini (2013). At the mass range where both studies coincide, they agree modulo a shift in the SDSS relation by $\sim +0.1$ dex in metallicity with respect to the CALIFA one (see the middle panel of Fig. 2.2). This is expected given that CALIFA covers the galaxies up to 2-3 effective radii while SDSS, in most of the cases, covers only the central regions which are typically more metallic than the outer ones (see for a discussion Sánchez et al., 2013). Thus, we use the relation as reported in Andrews & Martini (2013) but lowering it by 0.1 dex. They find that the function proposed by Moustakas et al. (2011) fits well their observational results:

$$12 + \log_{10}(\mathrm{O/H}) = (12 + \log_{10}(\mathrm{O/H})_{\mathrm{asm}})$$

$$-\log_{10}\left(1 + \left(\frac{M_{TO}}{M_*}\right)^\gamma\right), \qquad (2.2.2)$$

with $12 + \log_{10}(\mathrm{O/H})_{\mathrm{asm}} = 8.798$ (we use 8.698, after subtracting 0.1 dex), $M_{TO} = 8.901$, and $\gamma = 0.640$. We are aware that the uncertainties involved in any metallicity-dependent correction remain substantial (Bolatto et al., 2013). Note, however, that

our aim is to introduce and explore at a statistical level a reasonable mass-dependent correction to the CO-to-H_2 factor, which must be better than ignoring it. In any case, we present results both for $\alpha_{CO} = \alpha_{CO,MW}$ and our inferred mass-dependent α_{CO} factor. In fact, the mass-dependent factor is important only for LTGs with $M_* \lesssim 3 \times 10^{10}$ $M_\odot$; for higher masses and for all ETGs, $\alpha_{CO} \approx \alpha_{CO,MW}$[2]

Appendix B presents a description of the CO (H_2) samples that we utilize in this thesis. Table 2.2 lists the number of galaxies with detections and upper limits of the compilation sample in terms of morphology. Table 2.3 lists the number of detections and upper limits for the golden, silver, and bronze categories mentioned above (§§2.2.1).

Figure 2.3 shows the mass ratio $R_{H_2} \equiv M_{H_2}/M_*$ vs. M_* for the compiled samples. Similarly to the R_{HI}– M_* relation, we applied some corrections to observations in order to homogenize our compiled sample and to have this way a more consistent comparison between the different samples. The upper and bottom left panels of Figure 2.3 show, respectively, the compiled datasets for LTGs and ETGs.

2.3   Selection effects

In this Section we check the gas-to-stellar mass correlations from the different compiled samples against possible selection effects. We also introduce, when possible, an homogenization in the upper limits of ETGs. The reader interested only on the main results can skip to Section 2.5.

As seen in Figs. 2.1 and 2.3 there is a significant fraction of galaxies with no detections in radio, for which the authors report an upper limit flux (converted into an HI or H_2 mass). *The non detection of observed galaxies gives information that we cannot ignore,* otherwise a bias towards high gas fractions would be introduced in the gas-to-stellar mass relations to be inferred. To take into account the upper limits in the compiled data, we resort to survival analysis methods for combining censored and uncensored data (i.e., detections and upper limits for non detections; see e.g., Feigelson & Babu, 2012). We will use two methods: the Buckley-James linear regression (Buckley & James, 1979) and the Kaplan-Meier product limit estimator (Kaplan & Meier, 1958a). Both are survival analysis methods commonly applied in Astronomy.[3] The former is useful for obtaining a linear regression from the censored

---

[2]This is well justifyied since massive LTGs have oxygen abundances with typical values larger than $12 + \log_{10}(O/H) \sim 8.7$ while ETG have high metallicities at all masses.

[3]We use the ASURV (Astronomy SURVival analysis) package developed by T. Isobe, M. LaValley and E. Feigelson in 1992 (see also Feigelson & Nelson, 1985), and implemented in the stsdas package (Space Telescope Science Science Data Analysis) in IRAF. In particular, we make use of the *buckleyjames* (Buckley-James linear regression) and *kmestimate* (Kaplan-Meier estimator) routines.
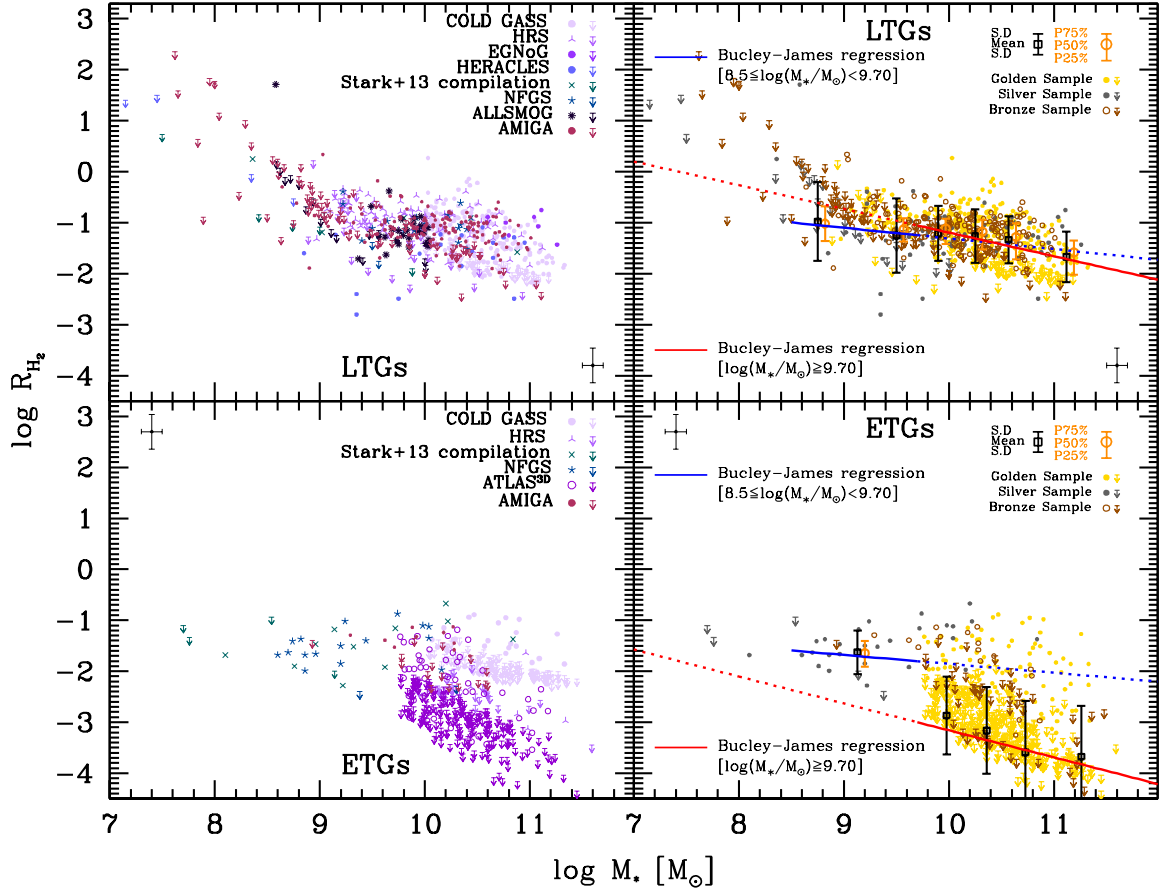
Figure 2.3: Molecular gas-to-stellar mass ratio as a function of M_*. *Upper panels:* Compiled and homogenized data with information on R_{H_2} and M_* for LTGs (see inside the panels for the different sources; see Appendix B for the acronyms and authors); downward arrows show the reported upper limits for non detections. Right panel is the same as left one, but with the data separated into three categories: Golden, Silver, and Bronze (yellow, gray, and brown symbols, respectively). The red and blue lines are Buckley-James linear regressions (taking into account non-detections). The dotted lines show extrapolations from these fits. The green dashed line shows an estimate for the R_{H_2}–M_* relation inferred from combining the empirical SFR–M_{H_2} and SFR–M_* correlations for blue/star-forming galaxies (see text for details). Squares with error bars are the mean and standard deviation of the data in different mass bins, taking into account non-detections by means of the Kaplan-Meier estimator. Open circles with error bars are the corresponding median and 25-75 percentiles. Estimates of the observational/calculation uncertainties are showed in the panel corners (see text). *Lower panels:* The same as in upper panels but for ETGs. In the right panel, we have corrected by distance the galaxies with upper limits from COLD GASS to make them consistent with the distances of the ATLAS[3D] sample (see text). For the bins where more than 50% of the data are upper limits, the median and percentiles are not calculated.

27

and uncensored data. Alternatively, for data that can not be described by a linear relation, we can bin them by mass, use the Kaplan-Meier estimator to calculate the mean, standard deviation,[4] median, and 25-75 percentiles in each stellar mass bin, and fit these results to a function by using conventional methods, e.g., the Levenberg-Marquardt algorithm. For the latter case, the binning in $\log M_*$ is started with a width of ≈ 0.25 dex but if the data is too scarce in the bin, then its width is increased as to have not less than 25% of galaxies than in the most populated bins. Note that, for detection fractions smaller than 50%, the median and percentiles are very uncertain or impossible to be calculated with the Kaplan-Meier estimator (Lee & Wang, 2003), while the mean can be yet estimated for fractions as small as $\sim 20\%$, though with a large uncertainty. In the case of the Bukley-James linear regression, reliable results are guaranteed for detection fractions larger than $70 - 80\%$.

When the fraction of non detections is significant, the inferred correlations could be affected by *selection effects in the upper limits* reported in the different samples. This is the case for ETGs, where a clear systematic segregation between the upper limits of the GALEX Arecibo SDSS Survey (GASS, Catinella et al., 2010, 2012, 2013) and ATLAS[3D] (Serra et al., 2012) or Herschel Reference Survey (HRS, Boselli et al., 2010, 2014a) surveys is observed in the $\log R_{\mathrm{HI}} - \log M_*$ plane (see the gap in the left lower panel of Fig. 2.1), as well as between the CO Legacy Database for GASS (COLD GASS, Saintonge et al., 2011) and ATLAS[3D] (Young et al., 2011) or HRS (Boselli et al., 2014a) surveys in the $\log R_{\mathrm{H_2}} - \log M_*$ plane (see the gap in the left lower panel of Fig. 2.3). The determination of the upper limits depends on distance and instrumental/observational constrains (telescope sensitivity, integration time, spatial coverage, signal-to-noise threshold, etc.). The HI observations of GASS and ATLAS[3D] were carried out with different radio telescopes: the single-dish Arecibo Telescope and the Westerbork Synthesis Radio Telescope (WRST) interferometer array, respectively. Serra et al. (2012) discussed about differences regarding detections between single- and multiple-beam observations. For some galaxies from ATLAS[3D] that they were able to observe also with the Arecibo telescope, they conclude that their upper limits should be increased by a factor of ~ 2 in order to agree with the ALFALFA survey sensitivity and the signal-to-noise threshold they use for declaring non detections in

---

[4]The IRAF package provides actually the standard error of the mean, $SEM = s/\sqrt{n}$, where $s = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}$ is the sample standard deviation, n is the number of observations, and $\bar{x}$ is the sample mean. In fact, s is a biased estimator of the (true) population standard deviation σ. For small samples, the former underestimates the true population standard deviation. A commonly used rule of thumb to correct the bias when the distribution is assumed to be normal, is to introduce the term $n - 1.5$ in the computation of s instead of n. In this case, $s \to \sigma$. Therefore, an approximation to the population standard deviation is $\sigma = (n/\sqrt{n - 1.5}) \times SEM$. This is the expression we use to calculate the reported standard deviations.

their multiple-beam observations. Thus, to homogenize the upper limits, we correct the ATLAS^{3D} upper limits by this factor. In the case of R_{H_2}, the CO observations in the ATLAS^{3D} and COLD GASS samples were taken with the same radio telescope (IRAM).

The GASS (COLD GASS) samples are selected to include galaxies at distances between $\approx$ 109 and 222 Mpc, while the ATLAS^{3D} and HRS surveys include only nearby galaxies, with average distances of 25 and 19 Mpc, respectively. Since the definition of the upper limits *depends on distance*, for the same radio telescope and integration time, more distant galaxies have systematically higher upper limits than closer galaxies. This introduces a clear selection effect. In the case we have information for a sample of galaxies closer than other sample, and under the assumption that both samples are roughly representative of the same local galaxy population, a distance-dependent correction to the upper limits of the non-detected galaxies in the more distant sample should be introduced. Next, we describe our approach to apply such a correction to GASS (COLD GASS) ETG upper limits with respect to the ATLAS^{3D} ETGs.

2.3.1   Corrections to the upper limits of ETGs

Here we have noted that the upper limits reported for the GASS (HI) and COLD GASS (H$_2$) samples in the case of ETGs are significantly larger than those reported for the ATLAS^{3D} or HRS samples. Following Serra et al. (2012), we have corrected the ATLAS^{3D} upper limit values by a factor of two in order to take into account differences between the different telescopes and signal-to-noise thresholds used in this survey and in GASS (see Section 2.8). However, the main reason of the differences in the upper limits among these samples is a selection effect due to the different volumes covered by them. To illustrate this, in the left panel of Fig. 2.4 we plot the histogram of HI masses for ETGs in the $10.10 - 10.65 \log M_*$ bin for GASS (solid black line) and ATLAS^{3D} (dotted black line). Non detections are also included, with values of M_{HI} corresponding to their upper limits. The red lines show the histograms of only detections. The number of GASS ETGs increases as M_{HI} is lower and it has a peak at $\log(M_{HI}/M_\odot)\approx 8.4-9.0$, contributed mainly by the upper limits and consistent with the sensitivity limit of the ALFALFA survey at the distances of the GASS galaxies in the mentioned stellar mass range. For ATLAS^{3D}, with distances much closer than GASS, some ETGs are detected in HI with masses lower than $\log(M_{HI}/M_\odot)=8.4$, but most of them are actually undetected, having upper limits 1–1.5 orders of magnitude lower than in the case of GASS, consistent with the distance differences between both samples. The main difference between the M_{HI} distributions of both samples is in
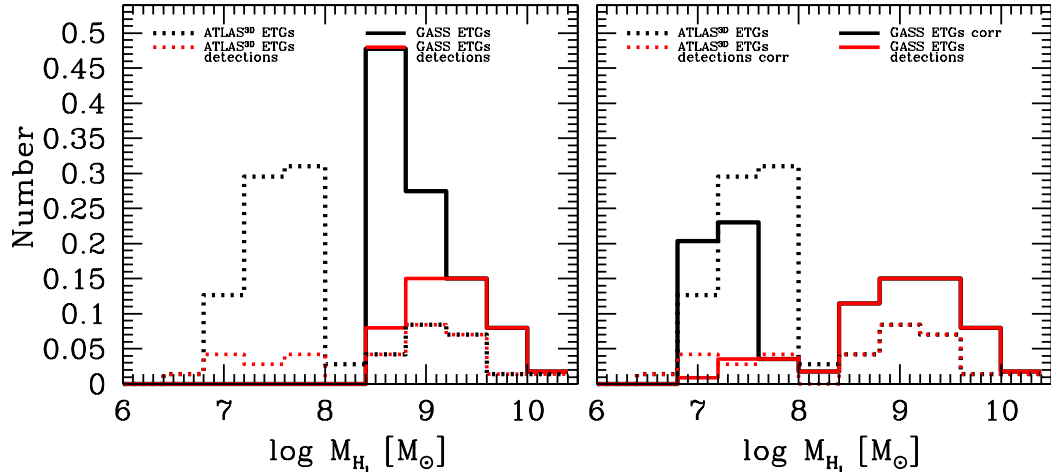
Figure 2.4: *Left panel:* Distributions of HI masses for ETGs in the $10.10 - 10.65$ $\log M_*$ bin for GASS (solid black line) and ATLAS^{3D} (dashed black line). Non detections are also included, with values of M_{HI} corresponding to their upper limits (for ATLAS^{3D}, we use the upper limits increased already by a factor of two as explained in Section 2.8). The red lines show the contribution of detected galaxies. The GASS distribution is clearly limited to much higher upper limits than in ATLAS^{3D}, and this is mainly due to a distance selection effect. *Right panel:* Same as in the left panel but after correcting the upper limits of GASS with respect to the observations of ATLAS^{3D}.

their upper limits, and this is clearly due to a selection effect imposed by the different distance ranges of these samples. Basically, if the undetected GASS ETGs would be at the distances of ATLAS^{3D} ETGs, then probably most of them would not be yet detected in HI, having upper limits lower by 1–1.5 orders of magnitude. Thus, the high values of their upper limits imposed by the volume of GASS, is expected to introduce a bias in the determination of the gas-to-stellar mass correlations of ETGs.

In an attempt to correct for this selection effect in the upper limits, we will assume that the ETGs in the GASS and ATLAS^{3D} (and HRS, too) samples are representative of the same local ETG populations. Then, that the upper limits for the ATLAS^{3D} (or HRS) ETGs are significantly lower than those *of similar stellar mass* galaxies from GASS, is mainly due to the distance differences among these samples. If the GASS ETGs would be as close as those of the ATLAS^{3D} ones, then the upper limit region in the plots of HI-to-stellar mass ratio vs. M_* would be on average lower by a factor equal to the distance ratio to the square. Thus, to homogenize the upper limits in R_{HI} given by the GASS and ATLAS^{3D} samples, we lower the upper limits of the galaxies in the volume-limited sample with more distant galaxies (GASS) by $(D_i/\bar{D}_{\text{ATLAS}^{3D}})^2$, where D_i is the distance of each GASS ETG and $\bar{D}_{\text{ATLAS}^{3D}} = 25$ Mpc is the average distance of the ATLAS^{3D} ETGs. In fact, according to the ATLAS^{3D}
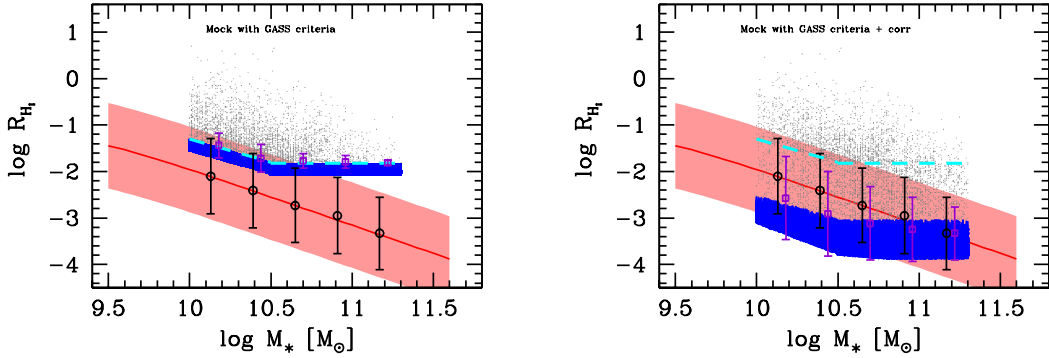
Figure 2.5: *Left panel:* ETGs from our $109 < D < 222$ Mpc volume mock catalog in the R_{HI} vs. M_* plane, following the selection and R_{HI} limits of GASS. All mock ETGs below the GASS R_{HI} limits (dashed line) are assumed as undetected and assigned an R_{HI} value equal to the R_{HI} limit (upper limit; blue arrows). The magenta squares with error bars are the mean and standard deviation calculated in different mass bins with the Kaplan-Meier estimator. The R_{HI}–M_* correlation for ETGs used in the generation of the mock catalog is plotted with the red solid line and shaded area. The circles with error bars are the mean and standard deviation calculated in different mass bins for all the ETGs from the mock catalog. The mock catalog samples very well the input correlation but this is not anymore the case when the R_{HI} limit of GASS is imposed, even if using the Kaplan-Meier estimator to take into account the upper limits. *Right panel:* Same as in left panel but after applying our ATLAS[3D]-based corrections to the upper limits of GASS (see text). The mean and standard deviation in the different mass bins, taking into account the (corrected) upper limits, follow now closely the input correlation.

observations, 25% of ETGs below the upper limit region of GASS were detected (see for an example Fig. 2.4). Therefore, we lower the GASS upper limits as mentioned above for 75% of the galaxies, and for the remaining ones we assign randomly an R_{HI} value between its upper limit and the average upper limit of ATLAS[3D] galaxies at the corresponding stellar mass using a flat distribution. The same procedure is applied to the COLD GASS ETGs for the $R_{\mathrm{H_2}}$ upper limits, where the corresponding $\bar{D}_{\mathrm{ATLAS^{3D}}}$ for COLD GASS is 26 Mpc.

The right panel of Fig. 2.4, shows the same histograms as in the left panel but now the upper limits of the GASS sample were corrected as explained above. Observe how close the upper limit distributions of GASS and ATLAS[3D] galaxies are after correcting by the distance selection effect. Further, we use a large mock galaxy catalog to test the procedure applied here to the GASS (or COLD GASS) upper limits for homogenizing them with those of nearby samples such as ATLAS[3D]. The mock catalog is a volume-limited sample (up to 313 Mpc) of 5×10^6 galaxies that sample well the observational GSMF and LTG/ETG fractions as a function of M_* (see Section 2.7). We assign HI masses to each LTG/ETG galaxy by using an *input* R_{HI} distribution for a given M_* (a R_{HI}–M_* relation and its scatter) for LTGs and ETGs. Distances are assigned assuming an isotropic distribution within a sphere of radius of the volume sampled. Note that we ignore any clustering properties of the galaxies. This is a safe assumption as we are only interested on the selection effects introduced by the detection limits of the GASS and ATLAS[3D] samples. Then, we select the ETGs more massive than 10^{10} M$_\odot$ that are in the $109 < D < 222$ Mpc range (the GASS volume), and impose upper limits to the R_{HI} ratio as a function of mass as the one of GASS (see Catinella et al., 2012). Then, we calculate the mean R_{HI} and its standard deviation taking into account the upper limits in mass bins as we did for the observational sample (using the Kaplan-Meier estimator). The question now is whether we recover or not the input R_{HI}–M_* correlation for ETGs.

In the left panel of Fig. 2.5, we plot our input R_{HI}–M_* correlation for ETGs (for this exercise, is described by a double power-law function with the parameters given in Table 2.5 and assuming a lognormal scatter) along with the values from the mock catalog in the $109 < D < 222$ Mpc volume and imposing the sensitivity limit of the GASS sample (dots). All the dots below this limit are plotted as upper limits (blue arrows); they populate the imposed sensitivity limit in the R_{HI} vs M_* diagram. The open circles with error bars are the mean and standard deviation calculated directly from the catalog in $\log M_*$ bins for ETGs in the $109 < D < 222$ Mpc volume, while the magenta squares and error bars are the same means and standard deviations calculated with the Kaplan-Meier estimator for the case of imposing the GASS sensitivity limit.

Thus, after imposing this limit, *the recovered correlation is far from the input one.*

Then, we apply the same corrections we have used for the real GASS data, based on the information from the ATLAS[3D] sample, i.e., the GASS-like imposed upper limits to the mock catalog galaxies were lowered by $D_i^2[\text{Mpc}^2]/25^2\text{Mpc}^2$ in 75% of the cases, and for the remaining, a random detection value for R_{HI} was assigned as explained above. The right panel of Fig. 2.5 shows the result of these corrections along with the mean and standard deviations calculated with the corrected data in the same three mass bins as in the left panel (magenta squares with error bars). Observe that after our corrections, the calculated mean and standard deviation in each mass bin are in better agreement with those corresponding to the mock catalog without any selection, that is, the input R_{HI}–M_* correlation is reasonable well recovered, showing this the necessity of applying the mentioned corrections.

The effect of introducing or not the mentioned above correction to the GASS and COLD GASS upper limits on the determination of the HI- and H_2-to-stellar mass correlations of ETGs are, of course, not so significant as in the experiment shown in Fig. 2.5 because these samples are not the only ones used for that (subsections 2.2.2 and 2.2.3). In Tables 2.4 and 2.5 (cases ETG[ndc]), we present the fitted HI-to-stellar mass correlation for ETGs in the case the upper limits of the GASS sample were not corrected by distance. The double power-law correlation, without the correction, changes slightly at the high-mass end: it would be shallower but with a much larger scatter than when we took into account the correction; the latter is expected due to the strong segregation of the upper limits from COLD GASS and from the less distant ATLAS[3]D and HRS samples. The single power-law would be shallower. Similarly, in these Tables is also present the fitted H_2-to-stellar correlation for ETGs in the case the upper limits of the COLD GASS sample were not corrected by distance. The relations are actually almost the same when taking or not taking into account the correction but the scatter is larger at the high-mass end for the latter case, as expected due to the segregation of the upper limits from COLD GASS and from the less distant ATLAS[3D] and HRS samples.

Note that after our corrections by distance and instruments, the upper limits of the massive ETGs in the GASS/COLD GASS sample are now consistent with those in the ATLAS[3D] (as well as HRS) samples, as seen in the right panels of Figs. 2.1 and 2.3 to be described below, and in Fig. 2.4 in Section 2.3.1. In the case of LTGs, there is no evidence of much lower values of R_{HI} and R_{H_2} than the upper limits given in GASS and COLD GASS for galaxies closer than those in these samples.

2.4 Preliminary results on the R_{HI}–M_* and the $R_{\mathrm{H_2}}$–M_* relations

In the right panels of Figs. 2.1 and 2.3, all the compiled data shown in the left panels are again plotted with dots and arrows for the detections and non detections, respectively. The yellow, dark gray, and brown colors correspond to galaxies from the Golden, Silver, and Bronze categories, respectively (see §§2.2.1). The above mentioned corrections to the upper limits of GASS/COLD GASS and ATLAS$^{\mathrm{3D}}$ ETG samples were applied. Observe that the large gaps in the upper limits between the GASS/COLD GASS and ATLAS$^{\mathrm{3D}}$ (or HRS) samples tend to disappear after the corrections we have applied.

We further group the data in logarithmic mass bins and calculate in each mass bin the mean and standard deviation of $\log R_{\mathrm{HI}}$ and $\log R_{\mathrm{H_2}}$ (black circles with error bars), taking into account the upper limits with the Kaplan-Meier estimator as described above. The orange squares with error bars are for the corresponding medians and 25-75 percentiles, respectively. In some mass bins, the fraction of detections are smaller than 50% for ETGs, therefore, the median and percentiles can not be estimated (see above). However, the mean and standard deviations can be yet calculated, though they are quite uncertain.

As seen in the right panels of Figs. 2.1 and 2.3, the logarithmic mean and median values tend to coincide and the 25-75 percentiles are roughly symmetric in most of the cases. Both facts suggest that the scatter around the mean relations (at least for the LTG population) tend to follow a nearly symmetrical distribution, for instance, a normal distribution in the logarithmic values (for a more detailed analysis of the scatter distributions see section 2.6).

In the following, we check whether each one of the compiled and homogenized samples deviate significantly or not from the mean trends. This could happen due to selection effects in the given sample. For example, we expect systematic deviations in the gas contents for the Bronze samples, because they are selected to contain galaxies in extreme environments. As a first approximation, we apply the Buckle-James linear regression to each one of the compiled individual samples, taking into account this way upper limits. When the data in the given sample are too scarce and/or dominated by non detections, the linear regression is not performed but the data are plotted.
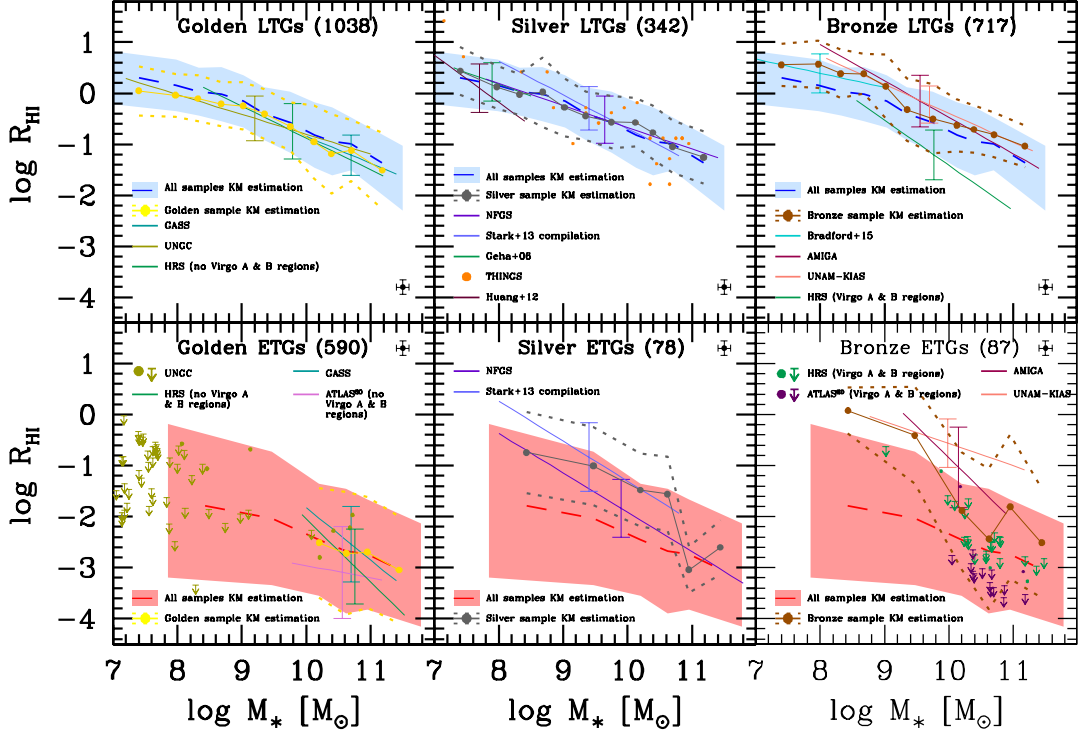
Figure 2.6: Atomic gas-to-stellar mass ratio as a function of M_* for the Golden, Bronze, and Silver LTGs (upper panels) and ETGs (lower panels). The mean and standard deviation in different mass bins, taking into account upper limits by means of the Kaplan-Meier estimator, are plotted for each case (filled circles connected by a dotted line and dotted lines around, respectively). For comparison, the mean and standard deviation (dashed lines and shaded area) from all the LTG (ETG) samples are reproduced in the corresponding upper (lower) panels. For each sample compiled and homogenized from the literature, the Buckley-James linear regression is applied, taking into account upper limits. The lines show the result, covering the range of the given sample; the error bars show the corresponding standard deviations obtained from the regression. When the data are too scarce and dominated by upper limits, the linear regression is not applied but the data are plotted. The number of LTG and ETG objects in each category are indicated in the respective panel.

2.4.1 R_{HI} vs. M_*

In Fig. 2.6, results for $\log R_{\mathrm{HI}}$ vs. $\log M_*$ are shown for LTGs (upper panels) and ETGs (lower panels). From left to right, the regressions for samples in the Golden, Silver, and Bronze categories are plotted. The error bars correspond to the 1σ scatter of the regression. Each line covers the mass range of the corresponding sample. The blue/red dashed lines and shaded regions in each panel correspond to the mean and standard deviation values calculated with the Kaplan-Meier estimator in mass bins for *all* the compiled LTG and ETG samples and previously plotted in Figs. 2.1 and 2.3, respectively. On the other hand, the yellow, gray, and brown dots connected with thin solid lines in each panel are the mean values in each mass bin calculated *only* for the Golden, Silver, and Bronze samples, respectively. The standard deviation are plotted with dotted lines. In the following, we discuss the results shown in Fig. 2.6.

Golden category: For LTGs, the three samples grouped in this category agree well among them in the mass ranges where they overlap; even the 1σ scatter of each sample do not differ significantly among them.[5] Therefore, as expected, these samples provide unbiased information for determining the R_{HI}–M_* relation of LTGs from $\log(M_*/\mathrm{M}_\odot) \approx 7.3$ to 11.4. For ETGs, the deviations of the Golden linear regressions among them and with respect to all galaxies are within their 1σ scatter, which are actually large. If no corrections to the upper limits of the GASS and ATLAS$^{\mathrm{3D}}$ are applied, then the regression to the former would be significantly above than the regression to the latter. Within the large scatter, the three Golden samples of ETGs seem not to be particularly biased, and they cover a mass range from $\log(M_*/\mathrm{M}_\odot) \approx$ 8.5 to 11.5. At smaller masses, the Updated Nearby Galaxy Catalog (UNGC) sample provides mostly only upper limits to R_{HI}.

Silver category: The LTG and ETG samples in this category, as expected, show a more dispersed distribution in their respective R_{HI}–M_* planes than those from the Golden category. However, the deviations of the Silver linear regressions among them and with respect to all the galaxies are within the corresponding 1σ scatter. If any, there is a trend of the Silver samples to have mean R_{HI} values above the mean values of all galaxies in special for ETGs. Since the samples in this category are not from complete volumes, but they were specially constructed for studying HI gas contents, a selection effect towards objects with non-negligible or higher than the mean HI contents can be expected. In any case, the biases are small. Thus, we decide to include the Silver samples to infer the R_{HI}–M_* correlations below in order to increase slightly the statistics (the number of galaxies in this category is actually much lower than in

---

[5]Note also that the 1σ scatter provided by the Buckle-James linear regression is consistent with the standard deviations in the mass bins obtained with the Kaplan-Meier estimator.

the Golden category), specially for ETGs of masses lower than $\log(M_*/\mathrm{M_\odot}) \approx 9.7$ (see Table 2.3).

Bronze category and the effects of environment: The very isolated LTGs (from the UNAM-KIAS and Analysis of the interstellar Medium of Isolated GAlaxies -AMIGA- samples) have HI contents higher than the mean of all the galaxies, specially at lower masses: $\log R_{\mathrm{HI}}$ is $0.1 - 0.2$ dex higher than the average at $M_* \gtrsim 10^{10}$ $\mathrm{M_\odot}$ and these differences increase up to $0.6 - 0.3$ dex for $8 < \log(M_*/\mathrm{M_\odot}) < 9$, though the number of galaxies at these masses is very small. The HI content of the Bradford et al. (2015) isolated dwarf galaxies is also higher than the mean of all the galaxies but not by a factor larger than 0.4 dex. For isolated ETGs, the differences can attain an order of magnitude and are in the limit of the upper standard deviations around the means of all the ETGs. Thus, while isolated LTGs have somewhat higher R_{HI} ratios on average than galaxies in other environments, in the case of isolated ETGs, this difference is very large; isolated ETGs can be almost as gas rich as LTGs. In the Bronze group we have included also galaxies from the central regions of the Virgo Cluster as reported in HRS and ATLAS$^{\mathrm{3D}}$ (only ETGs for the latter). According to Fig. 2.6, the LTGs in this high-density environment are clearly HI deficient with respect to LTGs in less dense environments. For ETGs, the HI content is very low but only slightly lower on average than the HI content of all ETGs. It should be noted that ETGs, in particular the massive ones, tend to be located in high-density environments.

We conclude that the HI content of galaxies is affected by the effects of extreme environments. The most remarkable effect is for ETGs, which in the very isolated environment can be as rich in HI as LTGs. Therefore, we decide to not include galaxies from the Bronze category to determine the R_{HI}–M_* correlation of ETGs. In fact, our compilation in the Golden and Silver categories includes galaxies from a range of environments (for instance, in the largest compiled catalog, UNGC, 58% of the galaxies are members of groups and 42% are field galaxies, see Karachentsev et al., 2014) in such a way that the R_{HI}–M_* correlation determined below should represent an average of different environments. Excluding the Bronze category for the ETG population, we avoid biases due to effects of the most extreme environments. For LTGs, the inclusion of the Bronze category does not introduce significant biases to the R_{HI}–M_* correlation of all galaxies but it helps to improve the statistics. The mean values of R_{HI} in mass bins above $\sim 10^9$ $\mathrm{M_\odot}$ are actually close to the mean values of all the sample (compare the brown solid and blue dashed lines); at lower masses the deviation increases, but the differences are well within the 1σ dispersion.
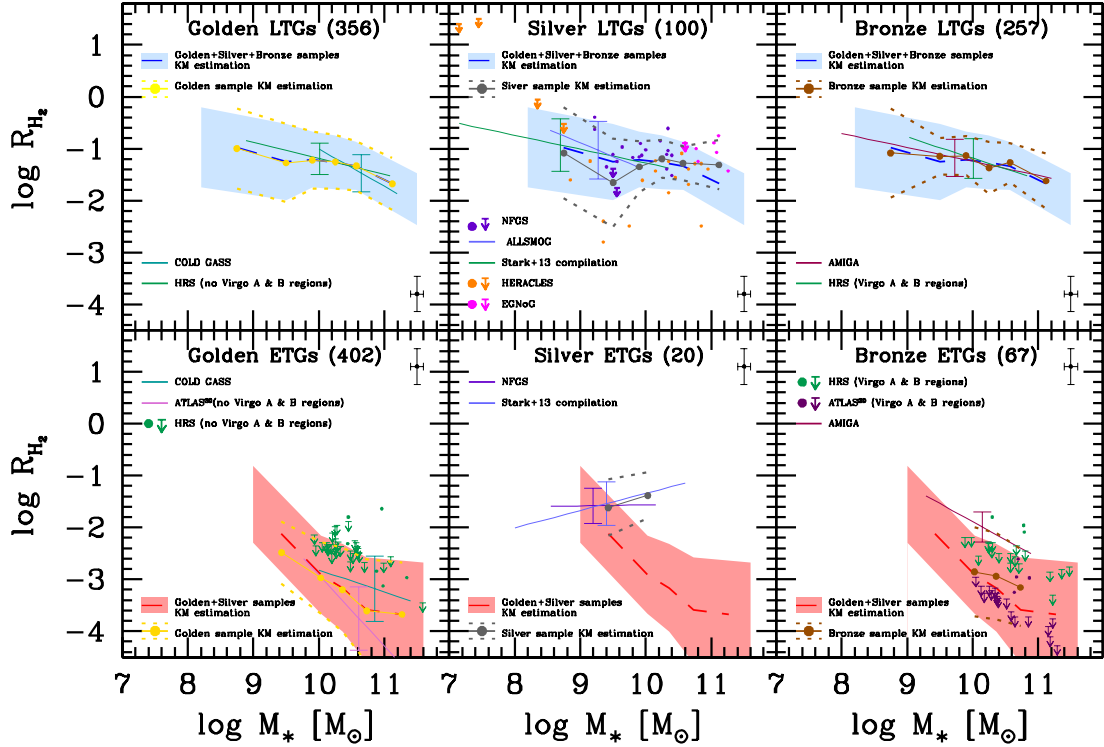
Figure 2.7: Same as Fig. 2.6 but for the molecular gas-to-stellar mass ratio.

2.4.2 R_{H_2} vs. M_*

In Fig. 2.7, we present similar plots as in Fig. 2.6 but for $\log R_{H_2}$ vs. $\log M_*$. The symbol and line codes are the same in both figures. In the following, we discuss the results shown in Fig. 2.7.

Golden category: For LTGs, the two samples grouped in this category agree well among them and with the overall sample, though for masses $< 10^{10}$ M$_\odot$, where the Golden galaxies are only those from the HRS sample, the average R_{H_2} values are slightly larger than those from the overall LTG sample (compare the solid yellow and dashed blue lines), but yet well within the 1σ scatter (shaded area). For ETGs, the deviations of the linear regressions of the Golden samples among them, and with respect to all ETGs, are within the respective 1σ scatters, which are actually large. If no corrections to the upper limits of the GASS and ATLAS^{3D} are applied, then the regression to the former would be significantly above than the regression to the latter. Summarizing, the Golden samples of LTGs and ETGs do not show particular shifts in their respective $R_{H_2}-M_*$ correlations. Therefore, the combination of them are expected to provide reliable information for determining the respective $R_{H_2}-M_*$ correlations; for LTGs, in the $\approx 10^{8.5}-10^{11.5}$ M$_\odot$ mass range, and for ETGs, only for $M_* \gtrsim 10^{10}$ M$_\odot$.

Silver category: The LTG samples present a dispersed distribution in the $\log R_{H_2}-$

logM_* plane but well within the 1σ scatter of the overall sample (shaded area). The mean values in mass bins from samples of the Silver category are in reasonable agreement with the mean values from all the samples (compare the gray solid and blue dashed lines). Therefore, the Silver samples, though scattered and not complete in any sense, seem not to suffer a clear systematic shift in their H_2 content. We include then these samples to infer the R_{H_2}-M_* correlation of LTGs. For ETGs, the two Silver samples provide information for masses below $M_* \sim 10^{10}$ $M_\odot$, and both are consistent with each other. Therefore, we include these samples to infer the ETG R_{H_2}-M_* correlation down to $M_* \sim 10^{8.5}$ $M_\odot$.

Bronze category and the effects of environment: The isolated (from the AMIGA sample) and Virgo central (from the HRS catalog) LTGs have H_2 contents similar to the mean in different mass bins of all the galaxies. If any, the Virgo LTGs have on average slightly higher values of R_{H_2} than the isolated LTGs, specially at masses lower than $M_* \sim 10^{10}$ $M_\odot$, given that the M_{H_2} is denser and more compact than M_{HI} accretion from galaxy interactions could lead to this scenario. Given that LTGs in extreme environments do not segregate from the average R_{H_2} values at different masses of all galaxies, we include them for calculating the $R_{H_2}-M_*$ correlation of LTGs.

For ETGs, the AMIGA isolated galaxies have on average significantly higher values of R_{H_2} than the mean of other galaxies, while those ETGs from the Virgo central regions (from HRS and ATLAS[3D]; mostly upper limits), seem to be on average consistent with the mean of all the galaxies, though the scatter is large. Given the strong deviation of isolated ETGs from the mean trend, we prefer to exclude galaxies from the Bronze category for determining the ETG $R_{H_2}-M_*$ correlation. We conclude that the H_2 content of LTGs is weakly dependent on the environment of galaxies, but in the case of ETGs, very isolated galaxies have systematically higher R_{H_2} values than galaxies in more dense environments.

2.5 The gas-to-stellar mass correlations of the two main galaxy populations

2.5.1 Strategy for constraining the correlations

In spite of the diversity in the compiled samples and their different selection functions, the exploration presented in the previuos Section shows that the HI and H_2 contents as a function of M_* from most of the samples compiled here do not segregate significantly among them. The exception are the Bronze samples for ETGs. Therefore, *the*

Bronze ETGs are excluded from our analysis. The strong segregation is actually by morphology (or color or star formation rate), and this is why we have separated since the beginning the compiled data into two broad galaxy groups, LTGs and ETGs.

To determine gas-to-stellar mass ratios as a function of M_* we need (1) to take into account the upper limits of undetected galaxies in radio, and (2) to evaluate the correlation independently of the number of data points at each mass bin. If we have many data points at some mass bins and only a few ones in other mass bins (as it would happen if we use, for instance, a mass-limited volume complete sample, with much more data points at lower-masses than at large masses), then the overall correlation of $R_{\rm HI}$ or $R_{\rm H_2}$ with M_* will be dominated by the former, giving probably incorrect values of $R_{\rm HI}$ or $R_{\rm H_2}$ at other masses. In view of these two requirements, our strategy to determine the $\log R_{\rm HI}$–$\log M_*$ and $\log R_{\rm H_2}$–$\log M_*$ correlations is as follows:

1. Calculate the logarithmic means and standard deviations (scatter) in stellar mass bins obtained from the compiled data taking into account the non detections (upper limits) by means of the Kaplan-Meier estimator.

2. Get an estimate of the intrinsic standard deviations (scatter), taking into account estimates of the observational errors.

3. Propose a function to describe the relation given by the mean and intrinsic scatter as a function of mass (e.g., a single or double power law).

4. Constrain the parameters of this function by performing a formal fit to the mean and scatter calculated at each mass bin; note that in this case *the fitting gives the same weight to each mass bin, in spite of the number of galaxies in each bin.*

2.5.2 The HI-to-stellar mass correlations

In the upper left panel of Fig. 2.8, along with the data from the Golden, Silver, and Bronze LTG samples, the mean and standard deviation (squares and black error bars) calculated in each mass bin with the Kaplan-Meier method are plotted. In the lower left panel, the same is plotted but for the Golden and Silver ETG samples (recall that the Bronze samples are excluded in this case)[6]. We see that the total standard deviations in $\log R_{\rm HI}$, $\sigma_{\rm dat}$, do not evidence a systematic dependence on mass both for LTGs and ETGs. Then, we can use a constant value for each case. For LTGs, the standard deviations have values around 0.45–0.65 dex with an average of $\sigma_{\rm dat} \approx 0.53$

---

[6]We refer the reader to table 2.1 to see which golden/silver samples have a one-to-one correspondence on HI and H_2 information.
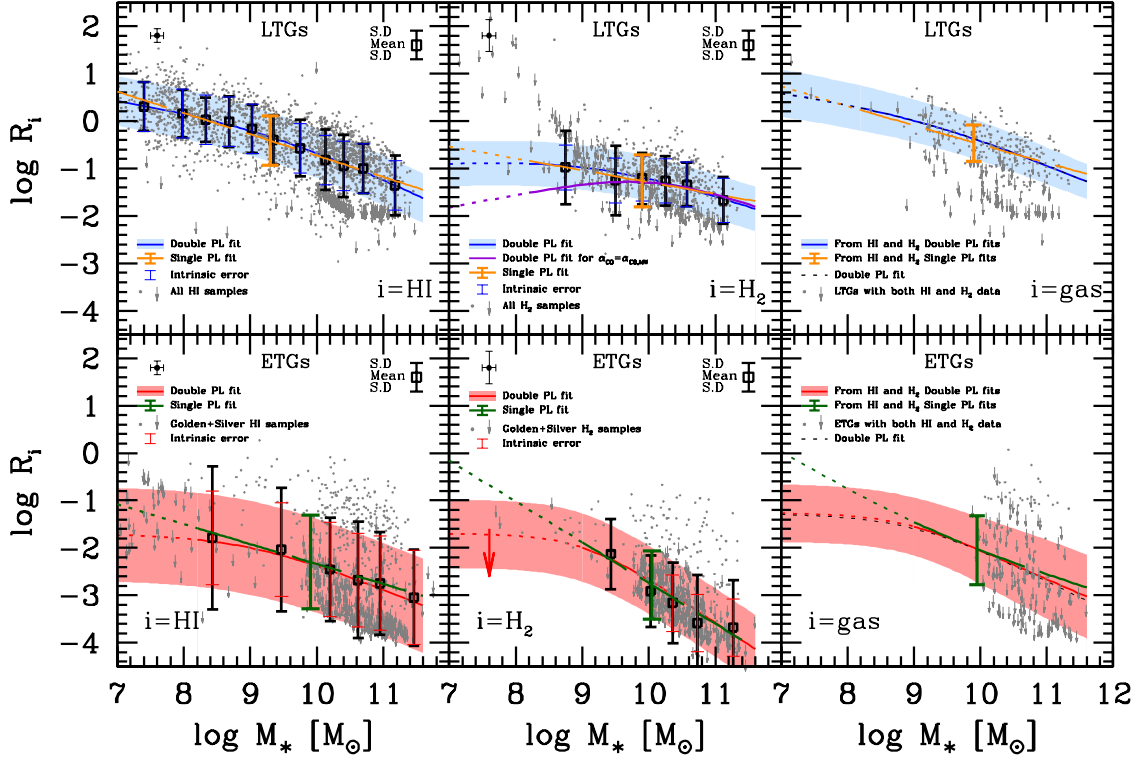
Figure 2.8: *Left panels:* The $R_{HI}–M_*$ correlation for LTGs (upper panel) and ETGs (lower panel). Dots are detections and arrows are upper limits for non detections (for ETGs the Bronze sample were excluded). The squares and error bars are the mean and standard deviation in different mass bins calculated by means of the Kaplan-Meier estimator for censored and uncensored data. The thin error bars correspond to our estimate of the *intrinsic* scatter after taking into account the observational errors (showed in the panel corners). The solid and long-dashed lines in each panel are respectively the best double- and single-power law fits. The shaded areas show the intrinsic scatter; to avoid overcrowding, for the single power-law fit, the intrinsic scatter is plotted only at one point. The dotted lines are extrapolations of the correlations to low masses, where the data are scarce and dominated by upper limits. *Middle panels:* Same as in the left panels but for R_{H_2}. For the ETG population, the double power-law fit was performed with the conservative constrain that below $M_* = 10^9 \, M_\odot$, the low-mass slope is 0. *Right panels:* The $R_{gas}–M_*$ correlations for LTGs and ETGs as calculated from combining the respective double- and single-power law $R_{HI}–M_*$ and $R_{H_2}–M_*$ correlations and taking into account helium and metals (see text). The shaded area and error bar are the (1σ) intrinsic scatter obtained by error propagation of the intrinsic scatter around the corresponding $R_{HI}–M_*$ and $R_{H_2}–M_*$ relations. For completeness, the data from our compilation that have determinations of both HI and H_2 masses are also plotted (the obtained correlations are not fits to these data). Dotted lines are extrapolations of the inferred relations to lower masses. The short dashed lines show the best fits using the double power-law function.

41

Table 2.4: Best fit parameters to the single power law (Eq. 1, $a = b$)

	$\log C'$	a	σ_{dat}	σ_{intr}
	R_{HI}-M_*			
LTG	3.77 ± 0.22	-0.45 ± 0.02	0.53	0.52
ETG	1.88 ± 0.33	-0.42 ± 0.03	1.00	0.99
ETG$^{\mathrm{ndc}}$	1.34 ± 0.46	-0.37 ± 0.05	1.35	1.34
	$R_{\mathrm{H_2}}$-M_*			
LTG	1.21 ± 0.53	-0.25 ± 0.05	0.58	0.47
ETG	5.86 ± 1.45	-0.86 ± 0.14	0.80	0.72
ETG$^{\mathrm{ndc}}$	5.27 ± 1.78	-0.80 ± 0.17	0.95	0.88
	R_{gas}-M_*			
LTG	4.76 ± 0.05	-0.52 ± 0.03	–	0.44
ETG	3.70 ± 0.07	-0.58 ± 0.01	–	0.68

• The suffix "ndc" indicates when for the ETG correlations, no distance correction was applied to the upper limits in the (COLD) GASS samples.

• σ_{dat} and σ_{intr} are in dex.

dex. For ETGs, the standard deviations are much larger and disparate among them than for LTGs (see subsection 2.5.4 below for a discussion on why this could be). We assume an average value of $\sigma_{\mathrm{dat}} = 1$ dex for ETGs.

The *intrinsic* standard deviation (scatter) can be estimated as $\sigma_{\mathrm{intr}}^2 \approx \sigma_{\mathrm{dat}}^2 - \sigma_{\mathrm{err}}^2$ (this is valid for normal distributions), where σ_{err} is the mean statistical error in the $\log R_{\mathrm{HI}}$ determination due to the observational uncertainties. For the stellar mass, the observational errors are typically estimated to be 0.1 dex (see e.g., Conroy, 2013). After homogenizing all the samples to a fixed IMF (Chabrier 2003) we have made the conservative assumption that other sources of systematic errors in the determination of M_* are negligible, see subsection 2.2.1. For the HI mass, a combination of the statistical errors, distance uncertainties, and errors associated with the absolute 21cm flux scale calibration accounts for a total observational error of ≈ 0.1 dex. Therefore, the average error in $\log R_{\mathrm{HI}}$ is ≈ 0.14 dex. Therefore, $\sigma_{\mathrm{intr}} \approx 0.52$ and 0.99 dex for LTGs and ETGs, respectively. These estimates should be taken only as indicative values given the assumptions and rough approximations involved in their calculations. For example, we will see in section 2.6 that the distributions of $\log R_{\mathrm{HI}}$ (detections and non-detections) in different mass bins tend to deviate from a normal distribution, in particular for ETGs

Next, we propose that the HI-to-stellar mass relations can be described by the general function:

$$y(M_*) = \frac{C}{\left(\frac{M_*}{M_*^{\mathrm{tr}}}\right)^a + \left(\frac{M_*}{M_*^{\mathrm{tr}}}\right)^b} \qquad (2.5.1)$$

where $y = R_{\mathrm{HI}}$, C is the normalization factor, a and b are the low- and high-mass

Table 2.5: Best fit parameters to the double power law (Eq. 1, $a \neq b$)

	C	a	b	$\log(M_*^{\mathrm{tr}}/\mathrm{M_\odot})$	σ_{dat}	σ_{intr}
		R_{HI}-M_*				
LTG	0.98 ± 0.06	0.21 ± 0.04	0.67 ± 0.03	9.24 ± 0.04	0.53	0.52
ETG	0.02 ± 0.01	0.00 ± 0.15	0.58 ± 0.03	9.00 ± 0.30	1.00	0.99
ETG$^{\mathrm{ndc}}$	0.02 ± 0.01	0.00 ± 0.55	0.51 ± 0.05	9.00 ± 0.60	1.35	1.34
		$R_{\mathrm{H_2}}$-M_*				
LTG	0.19 ± 0.02	-0.07 ± 0.18	0.47 ± 0.04	9.24 ± 0.12	0.58	0.47
ETG	0.02 ± 0.01	0.00 ± 0.00	0.94 ± 0.15	9.01 ± 0.12	0.80	0.72
ETG$^{\mathrm{ndc}}$	0.02 ± 0.03	0.00 ± 0.00	0.88 ± 0.18	9.01 ± 0.15	0.95	0.88
		R_{gas}-M_*				
LTG	1.69 ± 0.02	0.18 ± 0.01	0.61 ± 0.02	9.20 ± 0.04	–	0.44
ETG	0.05 ± 0.02	0.01 ± 0.03	0.70 ± 0.01	9.02 ± 0.05	–	0.68

- The suffix "ndc" indicates when for the ETG correlations, no distance correction was applied to the upper limits in the (COLD) GASS samples.
- σ_{dat} and σ_{intr} are in dex.

slopes of the function and M_*^{tr} is the transition mass. This function is continuous and differentiable. If $a = b$, then Eq. (2.5.1) describes a single power law or a linear relation in logarithmic scales. In this case, the equation remains as $y(M_*) = C'(M_*/\mathrm{M_\odot})^{-a}$. For $a \neq b$, the function corresponds to a double power law.

We fit the logarithm of function Eq. (2.5.1) to the mean values of $\log R_{\mathrm{HI}}$ as a function of mass (squares in the left panels of Fig. 2.8) with the corresponding (constant) intrinsic standard deviation as estimated above (thin blue/red error bars). For LTGs, the fit is carried out in the range $7.3 \lesssim \log(M_*/\mathrm{M_\odot}) \lesssim 11.2$, while for ETGs in the range $8.5 \lesssim \log(M_*/\mathrm{M_\odot}) \lesssim 11.5$. The Levenberg-Marquardt method is used for the fit (Press et al., 1996). First, we perform the fits to the binned LTG and ETG data using a single power law, i.e., we fix $a = b$. The dashed orange and green lines with an error bar in the left panels of Fig. 2.8 show the results. The fit parameters are given in Table 2.4. We note that *these fits and those of the Buckley-James linear regression for all the data (not binned) in logarithm are very similar.*

Then, we fit to the binned data the logarithm of the double power-law function given in Eq. (2.5.1). The corresponding best-fit parameters are presented in Table 2.5. We note that the fits are almost the same if the total mean standard deviation, σ_{dat}, is used instead of the intrinsic one. The reduced χ^2_{red} are 0.01 and 0.03, respectively. The fits are actually performed to a low number of points (the number of mass bins) with large error bars; this is why the χ^2_{red} are smaller than 1. Note, however, that the error bars are not related to measurement uncertainties but correspond to the population scatter of the data. Therefore, in this case $\chi^2_{\mathrm{red}} < 1$ implies that while the best fit is good, other fits could be also good within the scatter of the correlations. In the case of the single power-law fits, the χ^2_{red} were 0.03 and 0.01, respectively for LTG and ETG.

The double power-law R_{HI}–M_* relations and the estimated intrinsic (1σ) scatter for the LTG (ETG) population are plotted in the left upper (lower) panel of Fig. 2.8 with solid lines and shaded areas, respectively. From the fits, we find for LTGs a

transition mass $M_*^{\mathrm{tr}} = 1.74 \times 10^9$ M$_\odot$, with $R_{\mathrm{HI}} \propto M_*^{-0.21}$ and $M_*^{-0.67}$ at masses much smaller and larger than this, respectively. For ETGs, $M_*^{\mathrm{tr}} = 1 \times 10^9$ M$_\odot$, and $R_{\mathrm{HI}} \propto M_*^{0.0}$ and $M_*^{-0.58}$, at masses much smaller and larger than this, respectively.

Both the double and single power laws describe well the HI-to-stellar mass correlations. However, the former could be more adequate than the latter. In Fig. 2.1 we plot the Buckley-James linear regressions to the R_{HI} vs. M_* data for the low- and high-mass sides (below and above $\log(M_*/\mathrm{M}_\odot) \approx 9.7$; for ETGs the regression is applied only for masses above 10^8 M$_\odot$); the dotted lines show the extrapolation of the fits. The slope at low masses for LTGs, -0.36, is shallower than the one at high masses, -0.55. For ETGs, there is even evidence of a change in the slope sign at low masses. A flattening of the overall (late + early type galaxies) correlation at low masses has been also suggested by Baldry et al. (2008), who have used the empirical mass–metallicity relation coupled with a metallicity-to-gas mass fraction relation (which can be derived from a simple chemical evolution model) to obtain a gas-to-stellar mass correlation in a large mass range. Another evidence that at low masses the R_{HI}–M_* relation flattens comes from the work by Maddox et al. (2015) already mentioned above (see also Huang et al., 2012b). While the sample used by these authors does not allow to infer the R_{HI}–M_* correlation of galaxies due to its bias towards high R_{HI} values (see above), the upper envelope of this correlation can be actually constrained; the high-R_{HI} envelope does not suffer from selection limit effects. As seen for the data from Maddox et al. (2015) reproduced in the left upper panel of our Fig. 2.1, this envelope tends to flatten at $M_* \lesssim 2 \times 10^9$ M$_\odot$,[7] which suggests (but it does not demonstrate) that the mean relation can suffer also such a flattening. Another pieces of evidence in favor of the flattening can be found in Huang et al. (2012a), and more recently in Bradford et al. (2015) for their sample of low-mass galaxies combined with larger-mass galaxies from the ALFALFA survey.

2.5.3 The H$_2$-to-stellar mass correlations

In the upper middle panel of Fig. 2.8, along with the data from the Golden, Silver, and Bronze LTG samples, the mean and standard deviation (error bars) calculated in each mass bin with the Kaplan-Meier method are plotted. In the lower panel, the same is plotted but for the Golden and Silver ETG samples (recall that the Bronze samples are excluded in this case). The poor observational information at stellar

---

[7]In Huang et al. (2014), the $SDSS - GALEX - \alpha.40$ common sample was weighted by V/V_{max} to correct for incompleteness and mimic then the scaling relations derived from a volume-limited sample. However, only galaxies with $M_{HI} \gtrsim 10^{8.2}$ M$_\odot$ are included in their plot of R_{HI} vs. M_* (Fig. 1); at lower masses, the correlation likely continues being biased to high values of R_{HI}. Even that a weak flattening below $M_* \approx 10^9$ M$_\odot$ is observed in their average curve.

masses smaller than $\approx 5 \times 10^8$ M$_\odot$ does not allow us to constrain the correlations at these masses, both for LTG and ETGs. Regarding the total standard deviations, for both LTGs and ETGs, they vary from mass bin to mass bin but without a clear trend. Then we can use a constant value for both cases. For LTGs, the total standard deviations have values around 0.5–0.8 dex with an average of $\sigma_{\rm dat} \approx 0.58$ dex. For ETGs, the average value is roughly 0.8 dex. As in the case of HI (previous subsection), we further estimate indicative values for the *intrinsic* population standard deviations (scatter). For the H$_2$ mass, most of the works used in our compilation report average observational errors of $0.2 - 0.25$ dex. The uncertainty in the $\alpha_{\rm CO}$ parameter has been taken into account, however, it was probably significantly underestimated. In a recent review on the subject, Boselli et al. (2014a) suggest that this uncertainty is actually of the order of 0.3 dex. Thus, considering that the observational errors in the CO flux account for 30% (0.11 dex; e.g., Boselli et al., 2014a), and the uncertainty in the $\alpha_{\rm CO}$ parameter is 0.3 dex, an estimate of the typical error in $\log M_{H_2}$ is 0.32 dex. The estimated error in $\log R_{\rm H_2}$ is then ≈ 0.34 dex, using an error of 0.1 dex in $\log M_*$. Therefore, the estimated mean intrinsic scatters in $\log R_{\rm H_2}$ are $\sigma_{\rm intr} \approx 0.47$ and 0.72 dex for LTGs and ETGs, respectively. Given the assumptions and approximations involved in these estimates, they should be taken with caution. For example, we will see in section 2.6 that the distributions of $\log R_{\rm H_2}$ (detections and non-detections) in different mass bins tend to deviate from a normal distribution, in particular for the ETGs.

We fit the logarithm of function Eq. (2.5.1; $y = R_{\rm H_2}$) to the mean values of $\log R_{\rm H_2}$ as a function of mass (squares in the left panels of Fig. 2.8) with their corresponding scatter as estimated above (thin blue/red error bars), assumed to be the individual standard deviations for the fit. Again, the Levenberg-Marquardt method is used to perform the fit. The fits extend only down to $M_* \approx 5 \times 10^8$ M$_\odot$. First, the fits are performed for a single power law, i.e., we fix $a = b$. The dashed orange and green lines in the middle panels of Fig. 2.8 show the results. The parameters of the fit and their standard deviations are given in Table 2.4. *The fits are very similar to those obtained using the Buckley-James linear regression to the all (not binned) logarithmic data.*

Then, we fit the binned LTG and ETG data to the double power-law function Eq. (2.5.1). In the case of the ETG population, we impose an extra condition to the fit: that the slope of the relation at masses below $\sim 10^9$ M$_\odot$ is flat. The few data at these masses clearly show that $R_{\rm H_2}$ does not increase as M_* is smaller; it is likely that even decreases, so that our assumption of a flat slope is conservative. The corresponding best-fit parameters are presented in Table 2.5. As in the case of the

$R_{\mathrm{HI}} - M_*$ correlations, the reduced χ^2_{red} are smaller than 1 (0.04 and 0.10, respectively), which implies that while the best fits are good, other fits could describe reasonably well the scattered data. In the case of the single power-law fits, χ^2_{red} were 0.04 and 0.07, respectively for LTG and ETG. The double power-law $R_{\mathrm{H_2}} - M_*$ relations and their (1σ) *intrinsic* scatter for the LTG (ETG) population are plotted in the middle upper (lower) panel of Fig. 2.8 with solid lines and shaded areas, respectively. We note that the fits are almost the same if the total mean standard deviation, σ_{dat}, is used instead of the intrinsic one.

From these fits, we find for LTGs, $M_*^{\mathrm{tr}} = 1.74 \times 10^9 \ \mathrm{M_\odot}$, with $R_{\mathrm{H_2}} \propto M_*^{-0.07}$ and $M_*^{-0.47}$ at much smaller and larger masses than this, respectively. [8] For ETGs, $M_*^{\mathrm{tr}} = 1.02 \times 10^9 \ \mathrm{M_\odot}$, with $R_{\mathrm{H_2}} \propto M_*^{0.00}$ and $M_*^{-0.94}$ at much smaller and larger masses than this, respectively. In the middle upper panel of Fig. 2.8, we plot also the best double power-law fit to the $R_{\mathrm{H_2}} - M_*$ correlation of LTGs in the case the α_{CO} factor is assumed constant and equal to the MW value (purple dashed line).

Both the single and double power-law functions describe equally well the $R_{\mathrm{H_2}} - M_*$ correlations for the LTG and ETG population, but there is some evidence of a change of slope at low masses. In Fig. 2.3, the Buckley-James linear regressions to the $R_{\mathrm{H_2}}$ vs. M_* data below and above $\log(M_*/\mathrm{M_\odot}) \approx 9.7$ are plotted (in the former case the regressions are applied for masses only above $10^8 \ \mathrm{M_\odot}$); the dotted lines show the extrapolation of the fits. The slopes in the small mass range at low masses for LTGs/ETGs are shallower than those at high masses. Besides, in the case of ETGs, if the single power-law fit shown in Fig. 2.8 is extrapolated to low masses, ETGs of $M_* \approx 10^7 \ \mathrm{M_\odot}$ would be dominated in mass by $\mathrm{H_2}$ gas. Red/passive dwarf spheroidals are not expected to contain significant fractions of molecular gas. Recently, Accurso et al. (2017) have also reported a flattening in the $\mathrm{H_2}$-to-stellar mass correlation at stellar masses below $\sim 10^{10} \ \mathrm{M_\odot}$.

We refer the reader to table 2.1 in order to see which golden/silver samples have a one-to-one correspondence on HI and $\mathrm{H_2}$ information.

2.5.4 The cold gas-to-stellar mass correlations

Combining the $R_{\mathrm{HI}} - M_*$ and $R_{\mathrm{H_2}} - M_*$ relations presented above, we can obtain now the $R_{\mathrm{gas}} - M_*$ relation, for both the LTG and ETG populations. Here, $R_{\mathrm{gas}} = M_{\mathrm{gas}}/M_* = 1.4(R_{\mathrm{HI}} + R_{\mathrm{H_2}})$, where M_{gas} is the galaxy cold gas mass, including helium and metals (the factor 1.4 accounts for these components). The intrinsic scatter around the gas-to-stellar mass relation can be estimated by propagating the intrinsic scatter around

---

[8]These slopes a very similar to the ones between the sSFR and M_*.

the HI- and H_2-to-stellar mass relations. Under the assumption of null covariance, the logarithmic standard deviation around the composed $\log R_{\rm gas}$–$\log M_*$ relation is given by

$$\sigma_{\rm intr,R_{gas}} = \frac{1}{R_{\rm H_I} + R_{\rm H_2}} \left(R_{\rm H_I}^2 \sigma_{\rm intr,R_{H_I}}^2 + R_{\rm H_2}^2 \sigma_{\rm intr,R_{H_2}}^2 \right)^{\frac{1}{2}} \qquad (2.5.2)$$

The obtained cold gas-to-stellar mass correlations for the LTG and ETG populations are plotted in the right panels of Fig. 2.8. The solid lines and shaded bands (intrinsic scatter given by the error propagation) were obtained from the double power-law correlations, while the solid green lines and the error bars were obtained from the single power-law correlations. For completeness, we plot in Fig. 2.8 also those galaxies from our compilation that have determinations for *both* the HI and H_2 masses. Note that a large fraction of our compilation have not determinations for both quantities at the same time. We fit the results obtained for the singe (double) power-law fits, taking into account the intrinsic scatter, to the logarithm of the single (double) power-law function given in Eq. (2.5.1) with $y = R_{\rm gas}$ and report in Table 2.4 (Table 2.5) the obtained parameters for both the LTGs and ETGs. The fits for the double power-law fit are shown with dotted lines in Fig. 2.8. The standard deviations $\sigma_{\log R_{\rm gas}}$ change slightly with mass; we report an average value for them in Tables 2.4 and 2.5. Both for LTGs and ETGs, the mass at which the $R_{\rm gas}$–M_* correlations change of slope is $M_*^{\rm tr} \approx 1.7 \times 10^9$ M$_\odot$, *that separates dwarfs from low mass galaxies (Bullock & Boylan-Kolchin, 2017)*

According to Fig. 2.8, the LTG and ETG $R_{\rm gas}$–M_* correlations are significantly different among them. The gas content in the former is at all masses larger than in the latter, the difference being maximal at the largest masses. For the LTG population, $M_{\rm gas} \approx M_*$ on average at $\log(M_*/{\rm M}_\odot) \sim 9$, and at lower masses, these galaxies are dominated by cold gas; at stellar masses around 2×10^7 M$_\odot$, $M_{\rm gas}$ is on average three times larger than M_*. For ETGs, there is a hint that at $\sim 10^9$ M$_\odot$, $R_{\rm gas}$ changes from increasing as M_* is smaller to decrease.

2.6 The distributions of the scatter around the gas-to-stellar mass relations

To determine the correlations presented above, we have made use only of the mean and standard deviation of the data in different mass bins. It is also of interest to learn about the scatter distributions around the main relations. Even more, in the next Section we will require the full distributions of $R_{\rm HI}(M_*)$ and $R_{\rm H_2}(M_*)$ in order to generate a mock galaxy catalog through which the HI and H_2 mass functions will

be calculated. The Kaplan-Meier estimator provides information for constructing the probability density function (PDF) at a given stellar mass including the uncensored data. By using these PDFs we explore the distribution of the R_{HI} and $R_{\mathrm{H_2}}$ data (detections + upper limits). Given the heterogeneous nature of our compiled data, these "scatter" distributions should be taken just as a rough approximation. On the other hand, when the uncensored data dominate (this happens in most of the mass bins for the ETG samples), the Kaplan-Meier estimator can not predict very well the distribution of the uncensored data.

Table 2.6: Best fit parameters to the full distributions

	c	d	x_0	$\log(m_*^{\mathrm{tr}}/\mathrm{M_\odot})$	e	f	g	h	i	j
				$P(R_{\mathrm{HI}}\|M_*)$ distributions						
LTG	1.11±0.35	-0.11±0.04	2.45±0.76	8.77±0.45	0.002±0.10	0.61±0.07	–	–	–	–
ETG	-0.42±0.80	-0.02±0.08	2.15±0.55	8.30±0.38	-0.43±1.10	0.52±0.09	-0.22±0.37	0.07±0.04	-1.62±1.08	-0.13±0.11
				$P(R_{\mathrm{H_2}}\|M_*)$ distributions						
LTG	0.70±1.28	-0.07±0.13	0.15±0.03	10.37±0.31	0.19±0.17	0.19±0.16	–	–	–	–
ETG	-0.52±1.19	-0.01±0.11	0.71±0.27	7.90±1.09	0.42±0.50	0.21±0.28	0.24±0.97	0.04±0.09	5.74±3.17	-0.86±0.29

For LTGs the distributions are given by Eq. (2.6.1), while for ETGs, by Eq. (2.6.5).

Late-type galaxies.- Figures 2.9 and 2.10 present the R_{HI} and $R_{\mathrm{H_2}}$ PDFs in different M_* bins for LTGs. Based on the bivariate HI and stellar mass function analysis of Lemonias et al. (2013), who used the GASS sample for (all-type) massive galaxies, we propose that the PDFs of R_{HI} and $R_{\mathrm{H_2}}$ for LTGs can be described by a Schechter (Sch) function (Eq. 2.6.1 below; x denotes either R_{HI} or $R_{\mathrm{H_2}}$). By fitting this function to the R_{HI} data in each stellar mass bin we find that the power-law index α weakly depends on M_* with most of the values being around -0.15 (see also Lemonias et al., 2013), while the break parameter x^* varies with M_*. A similar behavior was found for $R_{\mathrm{H_2}}$ with most of the values of α around -0.10. We then perform for each case (R_{HI} and $R_{\mathrm{H_2}}$) a continuous fit across the range of stellar-mass bins rather than fits within independent bins. The general function proposed to describe the R_{HI} and $R_{\mathrm{H_2}}$ PDFs of LTGs, at a fixed M_* and within the range $\log x \pm d \log x/2$, is:

$$P_{\mathrm{Sch}}(x|M_*) = \frac{\phi^*}{\log e}\left(\frac{x}{x^*}\right)^{\alpha+1} \exp\left(-\frac{x}{x^*}\right), \qquad (2.6.1)$$

and with the normalization condition, $\phi^* = 1/\Gamma(1 + \alpha)$, where Γ is the complete gamma function, which guarantees that the integration over the full space in x is 1. The parameters α and x^* depend on M_*. We propose the following functions for these dependences:

$$\alpha(M_*) = c + d \log M_*, \qquad (2.6.2)$$

Figure 2.9: Distributions (PDFs) of the LTG HI-to-stellar mass ratios in different stellar mass bins (indicated inside the panels). The gray histograms show results from the Kaplan-Meier estimator applied to the data (detections + upper limits), and the solid blue line corresponds to the best fitted number density-weighted distribution within the given mass bin (eq. 2.6.4); the constrained parameters of the mass-dependent PDF (Eq. 2.6.1) are given in Table 2.6. The red dotted line shows the constrained function Eq. (2.6.1) evaluated at the mass corresponding to the logarithmic center of each mass bin. Note that for stellar mass bins $8.85 \leq \log M_* \leq 9.2$ and $10.85 \leq \log M_* \leq 11.5$ the best fit is poor. Nevertheless, the performance on most the stellar mass bins is quite good, thus we do not expect that this two mass bin will affect our results.
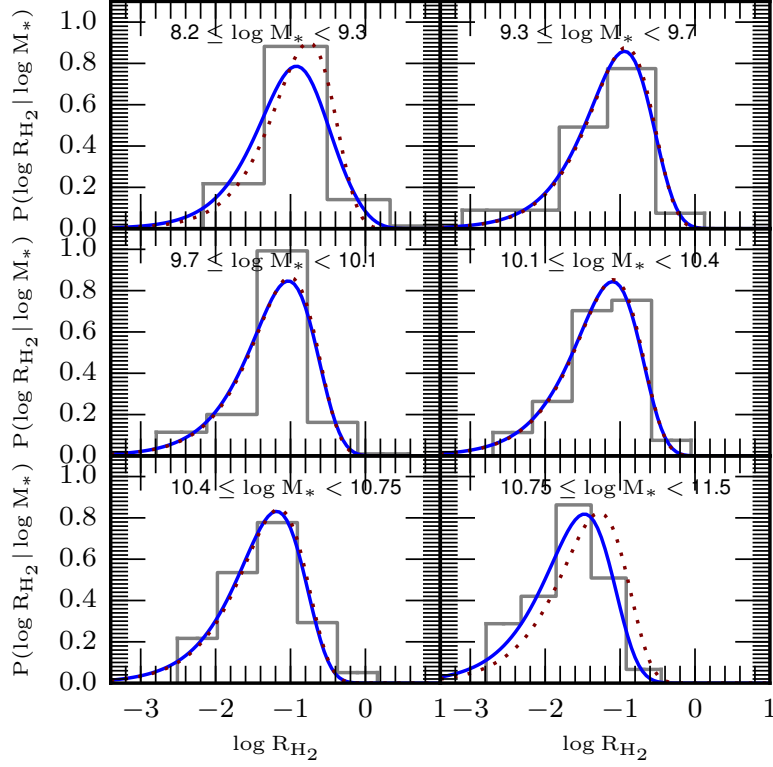
Figure 2.10: Same as Figure 2.9 but for the H_2-to-stellar mass ratios.

and

$$x^*(M_*) = \frac{x_0}{\left(\frac{M_*}{m_{\mathrm{tr}}}\right)^e + \left(\frac{M_*}{m_{\mathrm{tr}}}\right)^f}. \tag{2.6.3}$$

The parameters $c, d, x_0, m_{\mathrm{tr}}, e$, and f are constrained from a *continuous fit across all the mass bins* using a Markov Chain Monte Carlo method following Rodríguez-Puebla et al. (2013). Since the stellar mass bins from the data have a width, for a more precise determination, we convolve the PDF with the GSMF within a given bin. Therefore, the PDF of x averaged within the bin $\Delta M_* = [\mathrm{M}_{*1}, \mathrm{M}_{*2}]$ is:

$$\langle P_{\mathrm{Sch}}(x|\Delta M_*)\rangle = \frac{\int_{\mathrm{M}_{*1}}^{\mathrm{M}_{*2}} P_{\mathrm{Sch}}(x|M_*)\Phi_{late}(M_*)dM_*}{\int_{\mathrm{M}_{*1}}^{\mathrm{M}_{*2}} \Phi_{late}(M_*)dM_*}, \tag{2.6.4}$$

where $\Phi_{late}(M_*)$ is the GSMF for LTGs (see Section 2.7). The constrained parameters are reported in Table 2.6. The obtaiened mass-dependent PDFs are plotted in each one of the panels of Figures 2.9 and 2.10. The solid blue line corresponds to the number density-weighted distribution within the given mass bin (eq. 2.6.4), while the red dotted line is for the function Eq. (2.6.1) evaluated at the mass corresponding to the logarithmic center of each bin. As seen, the Kaplan-Meier PDFs obtained from the data (gray histograms) are well described by the proposed Schechter function averaged within the different mass bins (blue lines), both for R_{HI} and R_{H_2}.
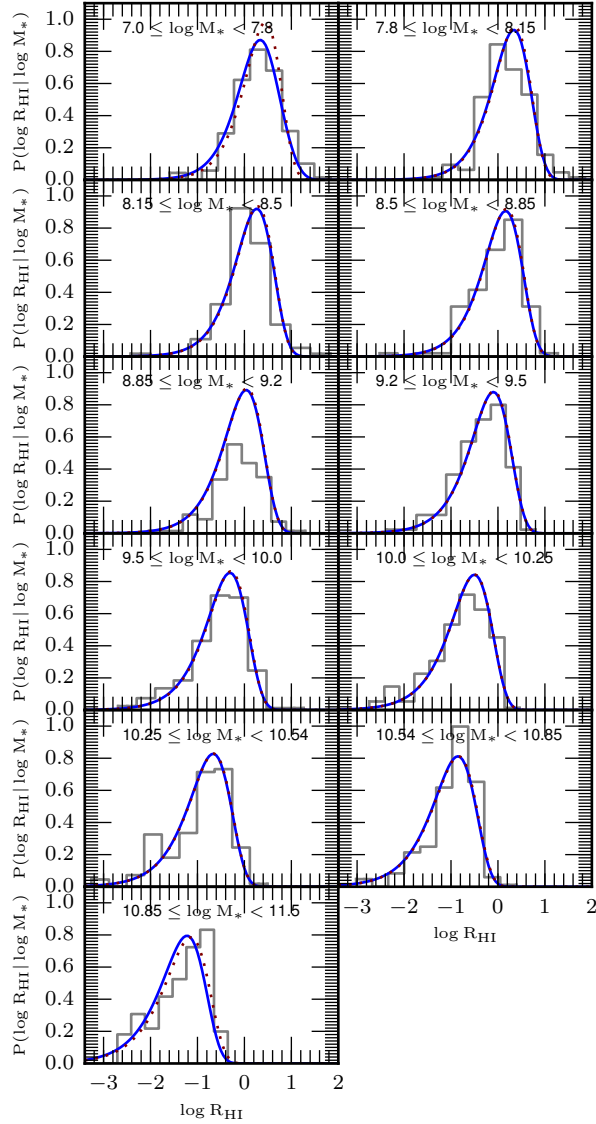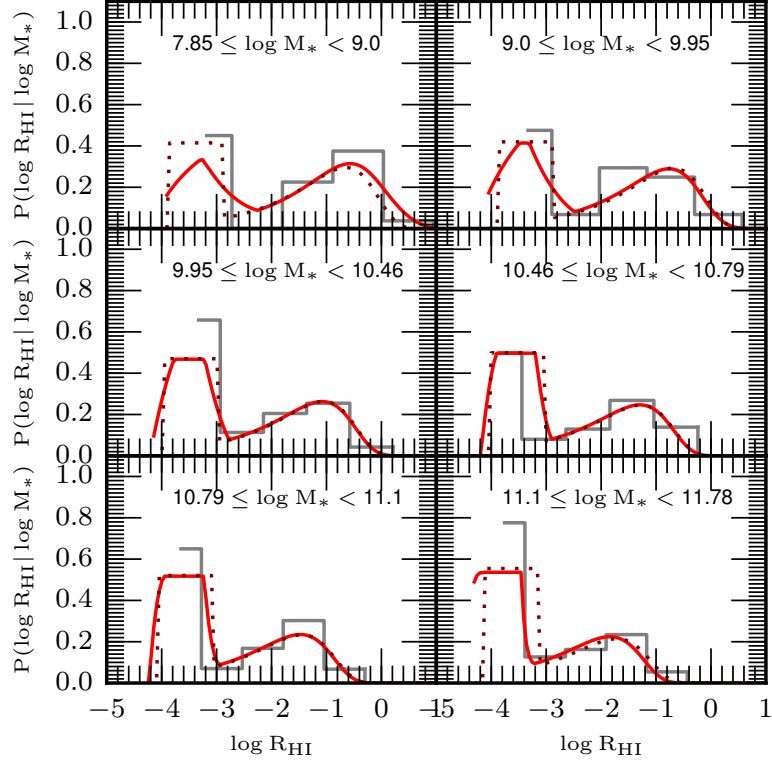
Figure 2.11: Distributions (PDFs) of the ETG HI-to-stellar mass ratios in different stellar mass bins (indicated inside the panels). The gray histograms show results from the Kaplan-Meier estimator applied to the data (detections + upper limits), and the solid ted line corresponds to the best fitted number density-weighted distribution within the given mass bin (eq. 2.6.6); the constrained parameters of the mass-dependent PDF (Eq. 2.6.5) are given in Table 2.6. The red dotted line shows the constrained function Eq. (2.6.5) evaluated at the mass corresponding to the logarithmic center of each mass bin.
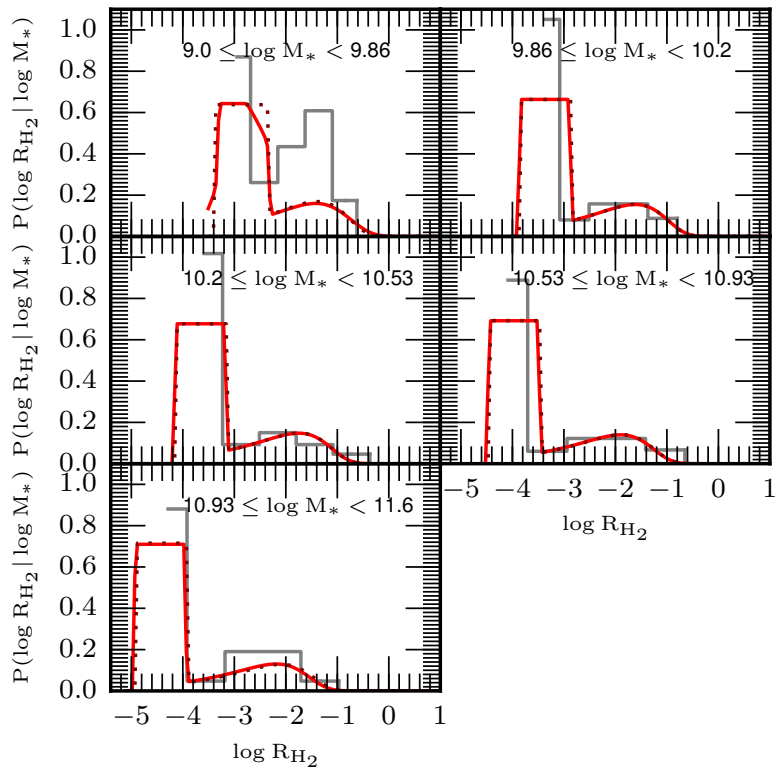
Figure 2.12: Same as Figure 2.11 but for the H$_2$-to-stellar mass ratios.

Early-type galaxies.- We present the R_{HI} and R_{H_2} PDFs for ETGs in Figures 2.11 and 2.12, respectively. The distributions are very extended, implying a large scatter in the $R_{H_2}-M_*$ correlations as discussed in subsections 2.5.2 and 2.5.3.[9] The distributions seem to be bimodal, with a significant fraction of ETGs having gas fractions around a low limit ($\sim 10^{-4}$) and the remaining galaxies with higher gas fractions, following an asymmetrical distribution. The low limit is given by the Kaplan-Meier estimator and it is associated with the reported upper limits of non-detections. We should have in mind that when non-detections dominate, the Kaplan-Meier estimator can not provide a reliable PDF at the low end of the distribution. From a physical point of view, we know that ETGs are in general quiescent galaxies that likely exhausted their cold gas reservoirs and did not accrete more gas. However, yet small amounts of gas can be available from the winds of old/intermediate-age stars. For instance, Sun-like stars can lose $\sim 10^{-4} - 10^{-5}$ of their masses in 1 Gyr; more massive stars, lose higher fractions. A fraction of the ejected material is expected to cool efficiently and ends as HI and/or H_2 gas. On the other hand, those ETGs that have larger fractions of cold gas, could get it by radiative cooling from their hot halos or by accretion from the cosmic web, and/or by accretion from recent mergers (see for a discussion Lagos et al., 2014, and more references therein). The amount of gas acquired depends on the halo mass, the environment, the gas mass of the colliding galaxy, etc. The range of possibilities is large, hence, the scatter around the ETG $R_{HI}-M_*$ and $R_{H_2}-M_*$ relations are expected to be large as semi-analytic models show (Lagos et al., 2014).

To describe the PDFs seen in Figures 2.11 and 2.12, we propose a (broken) Schechter function plus a uniform distribution. The value of R_{HI} or R_{H_2} where the Schechter function breaks and the uniform distribution starts, x_2, seems to depend on M_* (see Figs. 2.11 and 2.12). The lowest values where the distributions end, x_1, are not well determined by the Kaplan-Meier estimator, as mentioned above. To avoid unnecessary sophistication, we just fix x_1 as one tenth of x_2. This implies physical lowest values for R_{HI} and R_{H_2} of $10^{-4 \div -5}$, which are plaussible according to our discussion above. The value of the Schechter parameter α shows a weak dependence on M_* for both HI and H_2. On the other hand, the fraction of galaxies between x_1 and x_2, F, seems to depend on M_*. For the uniform distribution, this fraction is given by $F = P(< x_2|M_*) - P(< x_1|M_*) = \int_{x_1}^{x_2} C d\log x$, where $C = F/(\log x_2 - \log x_1)$; given our assumption of $\log x_2 - \log x_1 = 1$ dex, then $C = F(M_*)$. We parametrize all

---

[9]Given this large scatter, previous works, for small samples of massive galaxies, have suggested that red or early-type galaxies do not follow a defined correlation between M_{HI} and M_* (or luminosity; e.g., Welch et al., 2010; Serra et al., 2012) and between M_{H_2} and M_* (e.g., Saintonge et al., 2011; Lisenfeld et al., 2011; Young et al., 2011).

these dependences on M_* and perform a continuous fit across the range of stellar-mass bins, both for the R_{HI} and $R_{\mathrm{H_2}}$ data. The general function proposed to describe the PDFs of ETGs as a function of M_* within the range $\log x \pm d\log x/2$ is the sum of a Schechter function, $P_{\mathrm{Sch}}(x|M_*)$, and a uniform function in x but dependent on M_*, $C = F(M_*)$:

$$F(M_*) = g + h \log M_*, \quad x_1 \leq x < x_2(M_*), \tag{2.6.5}$$
$$x_2(M_*) = i + j \log M_*,$$
$$P_{\mathrm{Sch}}(x|M_*), \quad\quad x \geq x_2(M_*),$$

where the parameters x^* and α in $P_{\mathrm{Sch}}(x|M_*)$ are described by Eq. (2.6.1) with the normalization condition $\phi^* = (1-F)/\Gamma(1+\alpha)$, and $\log x_1 = \log x_2 - 1$. The parameters x_0, m_{tr}, e, and f of the broken Schechter function and the parameters g, h, i, and j of the uniform distribution are constrained as described for LTGs above, from a continuous fit accross all the mass bins using the number density-weighted PDFs at each stellar mass bin:

$$\langle P_{\mathrm{Sch}}(x|\Delta M_*) + C \rangle =$$

$$\frac{\int_{\mathrm{M}_{*1}}^{\mathrm{M}_{*2}} (P_{\mathrm{Sch}}(x|M_*) + C) \cdot \Phi_{early}(M_*) dM_*}{\int_{\mathrm{M}_{*1}}^{\mathrm{M}_{*2}} \Phi_{early}(M_*) dM_*}, \tag{2.6.6}$$

where $\Phi_{early}(M_*)$ is the GSMF for ETGs (see Section 2.7). The constrained parameters are reported in Table 2.6, both for R_{HI} and $R_{\mathrm{H_2}}$. The obtained mass-dependent distribution function is plotted in each one of the panels of Figures 2.11 and 2.12 The solid red line corresponds to the number density-weighted distribution within the given mass bin (eq. 2.6.6), while the red dotted line is for the proposed broken Schechter + uniform function evaluated at the mass corresponding to the logarithmic center of each bin. As seen, the Kaplan-Meier PDFs obtained from the data (gray histograms) are reasonably well described by the proposed function (eq. 2.6.5) averaged within the different mass bins (red lines), both for R_{HI} and $R_{\mathrm{H_2}}$.

Finally, in Figures 2.13 and 2.14 we reproduce from Figure 2.8 the means and standard deviations obtained with the Kaplan-Meier estimator in different M_* bins (gray dots and error bars) for LTG and ETGs, respectively, and compare them with the means and standard deviations of the general mass-dependent distributions functions given in Equations (2.6.1) and (2.6.5) and constrained with the data (black solid line and the two dotted lines surrounding it). The agreement is rather good in the log-log

R_{HI}–M_* and $R_{\mathrm{H_2}}$–M_* diagrams both for LTGs and ETGs. Black dashed lines are extrapolations of the mean and standard deviation inferences from the distributions mentioned above, assuming they are the same as in the last mass bin with available gas observations. We also plot in these Figures the respective mean double power-law relations determined in subsections 2.5.2 and 2.5.3 (dashed blue or red lines, for LTGs and ETGs respectively; dotted blue or red lines are extrapolations.).

In conclusion, the R_{HI} and $R_{\mathrm{H_2}}$ distributions as a function of M_* described by Equations (2.6.1) and (2.6.5) (with the parameters given in Table 2.6) for LTGs and ETGs, respectively, are fully consistent with the corresponding R_{HI}–M_* and $R_{\mathrm{H_2}}$–M_* correlations determined in subsections 2.5.2 and 2.5.3. Therefore, *Equations (2.6.1) and (2.6.5) provide a consistent description of the* HI- *and* $\mathrm{H_2}$-*stellar mass relations and their scatter distributions, for LTGs and ETGs, respectively.*

2.7 Consistency of the gas-to-stellar mass correlations with the observed galaxy gas mass functions

The HI- and $\mathrm{H_2}$-to-stellar mass relations can be used to map the observed GSMF into the HI and $\mathrm{H_2}$ mass functions (GHIMF and $\mathrm{GH_2MF}$, respectively). This way, we can check whether the correlations we have inferred from observations in subsectiona 2.5.2 and 2.5.3 are consistent or not with the GHIMF and $\mathrm{GH_2MF}$ obtained from HI and CO ($\mathrm{H_2}$) surveys, respectively. In order to carry out this check of consistency, we need a GSMF, on one hand, defined in a large enough volume as to include massive galaxies and to minimize cosmic variance, and on the other hand, complete down to very low masses. As a first approximation to obtain this GSMF, we follow here a procedure similar as in Kravtsov et al. (2014, see their Appendix A). We use the combination of two GSMFs: Bernardi et al. (2013) for the large SDSS volume (complete from $M_* \sim 10^9$ M$_\odot$), and Baldry et al. (2012) for a local small volume but nearly complete down to $M_* \sim 10^7$ M$_\odot$ (GAMA). For high masses, the SDSS-based GMSF presented in Bernardi et al. (2013) is used. These authors have reanalyzed the photometry of the SDSS DR7, taking special care in the background estimate of extended luminous galaxies (see also Simard et al., 2011; He et al., 2013; Mendel et al., 2014; D'Souza et al., 2015; Meert et al., 2016); after this reanalysis, the high-end of the luminosity (mass) function becomes shallower. Their GSMF is well fitted by a Schechter + sub exponential Schechter function. For small masses, the GSMFs determined by Baldry et al. (2012, from the GAMA survey) are used. These authors analyze low
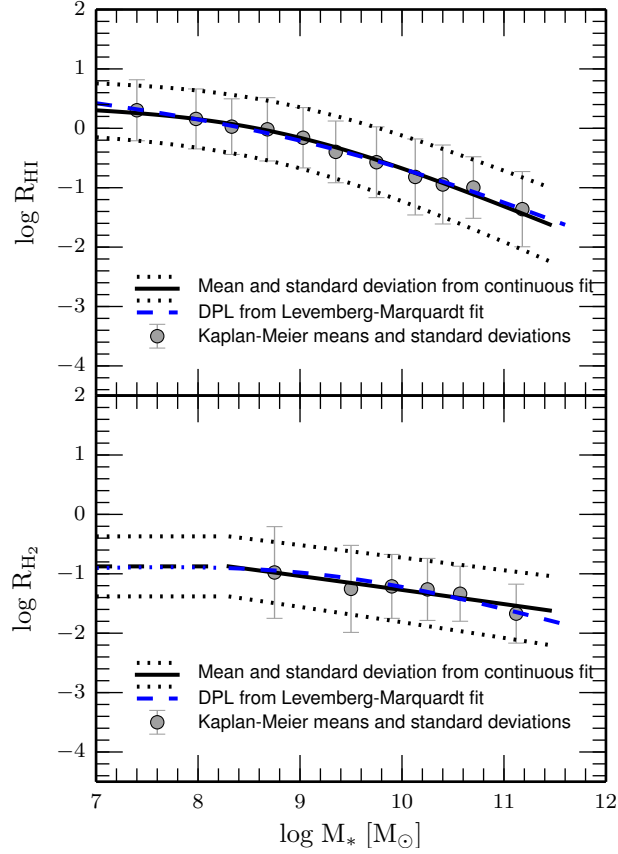
Figure 2.13: Mean and standard deviation as a function of stellar mass (solid and dotted black lines) from the distributions of R_{HI} (upper panel) and R_{H_2} (lower panel) for LTGs as given by Eq. (2.6.1) (see Table 2.6 for the constrained parameters). When the data are insufficient at low masses, the distributions are assumed the same as in the last mass bin (dashed black lines). The gray dots with error bars are the mean and standard deviation obtained with the Kaplan-Meier estimator applied to the data (detections + upper limits) in different mass bins, as shown in Figure 2.8. The double-power law fits to these data as reported in Section 4 are reproduced with the blue dashed lines (the blue dotted lines are extrapolations of these fits).
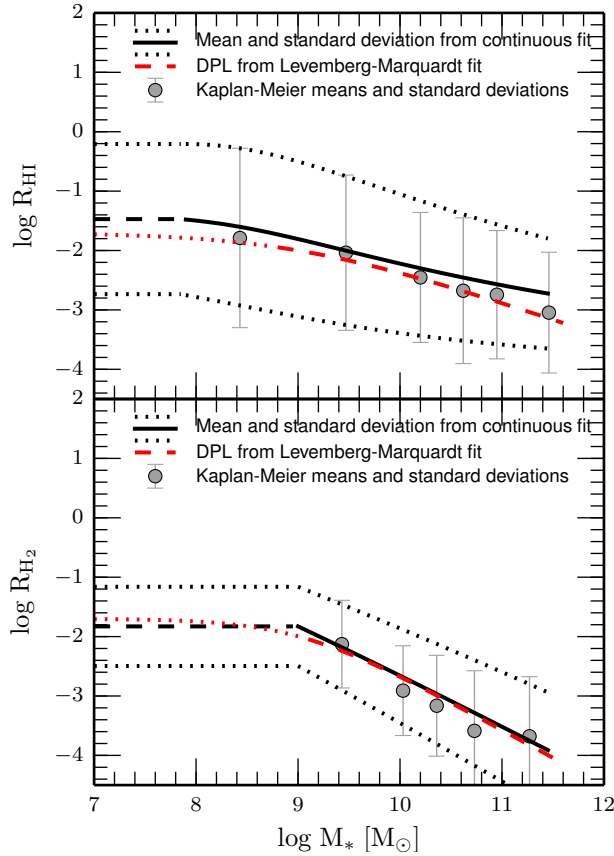
Figure 2.14: Same as in Figure 2.13 but for ETGs.

Table 2.7: GSMF parameters.

α_1	$\log(M_1^*)$ (M$_\odot$)	$\log(\phi_1^*)$ (Mpc^{-3} dex^{-1})	α_2	$\log(M_2^*)$ (M$_\odot$)	$\log(\phi_2^*)$ (Mpc^{-3} dex^{-1})	β
-1.47	9.74	-2.66	0.07	8.84	-2.66	0.37

redshift samples that contain low luminosity galaxies, though a correction for surface brightness incompleteness was not applied. So, their determinations at $M_* \lesssim 10^8$ M$_\odot$ are actually lower limits. This GSMF is well fitted by double Schechter function. Both, high and low masses GSMFs assume Chabrier (2003) IMF to estimate M_*. However, the masses in Bernardi et al. (2013) were calculated by using the Bell et al. (2003) mass-to-luminosity ratios, who employed the PEGASE stellar population synthesis models (Fioc & Rocca-Volmerange, 1997). In Baldry et al. (2012) the masses are calculated using the (Bruzual & Charlot, 2003, BC03) models. Conroy (2013) has shown that the former are systematically larger than the latter by $\approx 0.10 - 0.14$ dex. Therefore, for the Bernardi et al. (2013) GSMF, we dismiss uniformly M_* by 0.12 dex to homogenize the masses to the BC03 population synthesis model.

Thus, we find that a fit to the Baldry et al. (2012) and the Bernardi et al. (2013) GSMF fit corrected by 0.12 dex in mass are combined to obtain a GSMF that spans from $M_* \approx 10^7$ to 10^{12} M$_\odot$. The match of both fits (at the mass where the latter

becomes higher than the former) takes places at $M_* \approx 10^{9.3}$ M$_\odot$. The obtained GSMF is well fitted by the combination of a Schechter function and a sub exponential Schechter function. The respective parameters are given in Table 2.7.

Figure 2.15 presents our combined GSMF (solid line) and some GSMFs reported in the literature: the two used by us (see above), and those from Wright et al. (2017), Papastergis et al. (2012), and Baldry et al. (2008) in small but deep volumes, and D'Souza et al. (2015) in a large volume. We plot both the original data from Bernardi et al. (2013) (pink symbols) and after dismissing M_* by 0.12 dex (blue symbols) to homogenize the stellar masses to the BC03 population synthesis model. There is very good agreement between our combined GSMF and the recent GSMF reported in Wright et al. (2017) for the GAMA data.

Since the GSMF will be used as an interface for constructing the HI and H$_2$ mass functions, it is implicit the assumption that each galaxy with a given stellar mass has its respective HI and H$_2$ content. Hence, the gas mass functions presented below exclude the possibility of galaxies with gas content but not stars, and are equivalent to gas mass functions constructed from optically-selected samples (as in e.g., Baldry et al., 2008; Papastergis et al., 2012). In any case, it seems that the probability of finding only-gas galaxies is very low (Haynes et al., 2011).

We generate a volume complete mock galaxy catalog that samples the empirical GSMF presented above, and that takes into account the empirical volume-complete fraction of ETGs, f_{early}, as a function of stellar mass (the complement is the fraction of LTGs, $f_{\mathrm{late}} = 1 - f_{\mathrm{early}}$). The catalog is constructed as follows:

1. A minimum galaxy stellar mass $M_{*,\mathrm{min}}$ is set ($= 10^7$ M$_\odot$). From this minimum we generate a population of 5×10^6 galaxies that samples the GSMF presented above.

2. Each mock galaxy is assigned either as LTG or ETG. For this, we use the results reported in Moffett et al. (2016a), who visually classified galaxies from the GAMA survey. They consider ETGs those classified as Ellipticals and S0-Sa galaxies. The f_{early} fraction as a function of M_* is calculated as $\Phi_{\mathrm{early}}(M_*)/\Phi_{\mathrm{all}}(M_*)$, with $\Phi_{\mathrm{early}}(M_*) = \Phi_{\mathrm{Ell}}(M_*) + \Phi_{\mathrm{S0-Sa}}(M_*)$, using the fits to the respective GSMFs reported in Moffett et al. (2016a).[10]

3. For each galaxy, R_{HI} is assigned randomly from the conditional probability distribution $P_j(R_{\mathrm{HI}}|M_*)$ that a galaxy of mass M_* and type $j =$ LTG or ETG lies in the $R_{\mathrm{HI}} \pm dR_{\mathrm{HI}}/2$ bin. Then, $M_{HI} = R_{\mathrm{HI}} \times M_*$. The probability distributions for LTGs and ETGs are given by the mass-dependent PDFs presented in Equations (2.6.1) and (2.6.5), respectively (their parameters are given in Table 2.6).

---

[10]Note that Sa galaxies are not included in our definition of ETGs, so that f_{early} is probably overestimated at masses where Sa galaxies are abundant, making that $f_{\mathrm{early}} = 0.5$ at masses lower than the break mass, M^* (see figure 2.16).
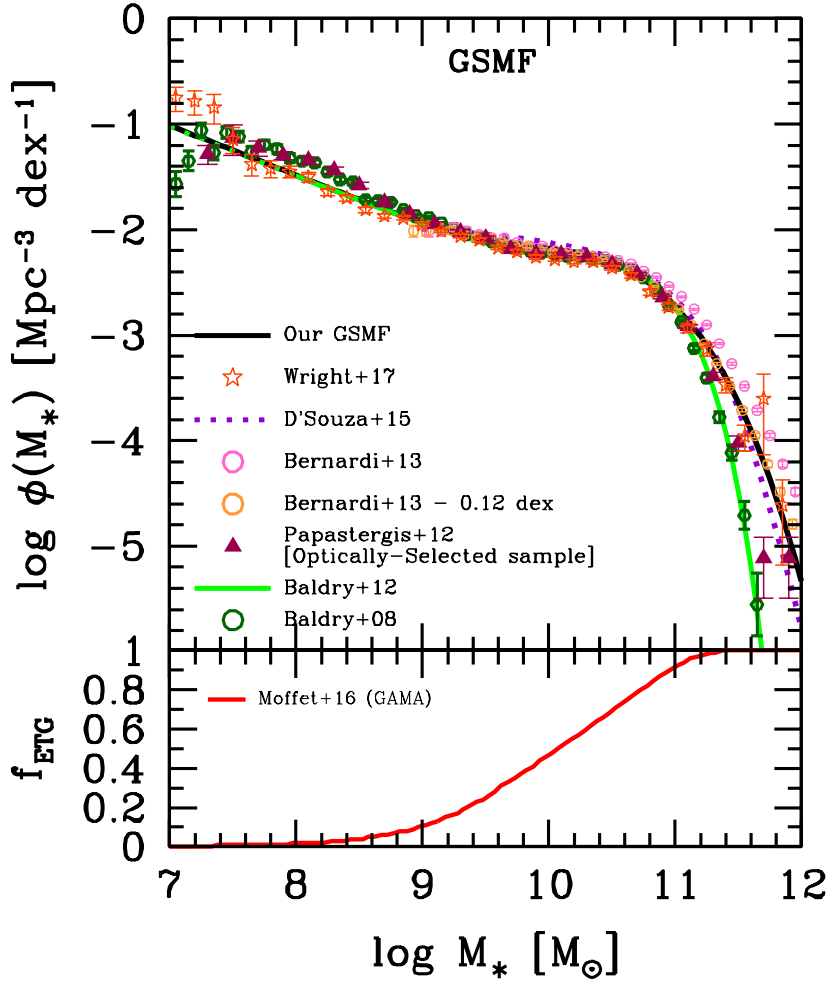
Figure 2.15: Our GSMF obtained from the combination of three observational GSMFs following Kravtsov et al. (2014) (thick solid line): one from the large SDSS DR7 volume but complete only down to $\sim 10^9$ M$_\odot$ (Bernardi et al., 2013, pink open circles with error bars; the orange open circles with error bars are after correcting M_* by 0.12 dex, see text), and two complete down to lower masses but in a very local volume (Wright et al. (2017), Papastergis et al., 2012 and Baldry et al., 2012). We also plot for comparison, the GSMFs reported in Baldry et al. (2008) and D'Souza et al. (2015). The lower panel shows the fraction of ETGs as a function of mass inferred by Moffett et al. (2016a), using GAMA galaxies and their visual morphological classification.
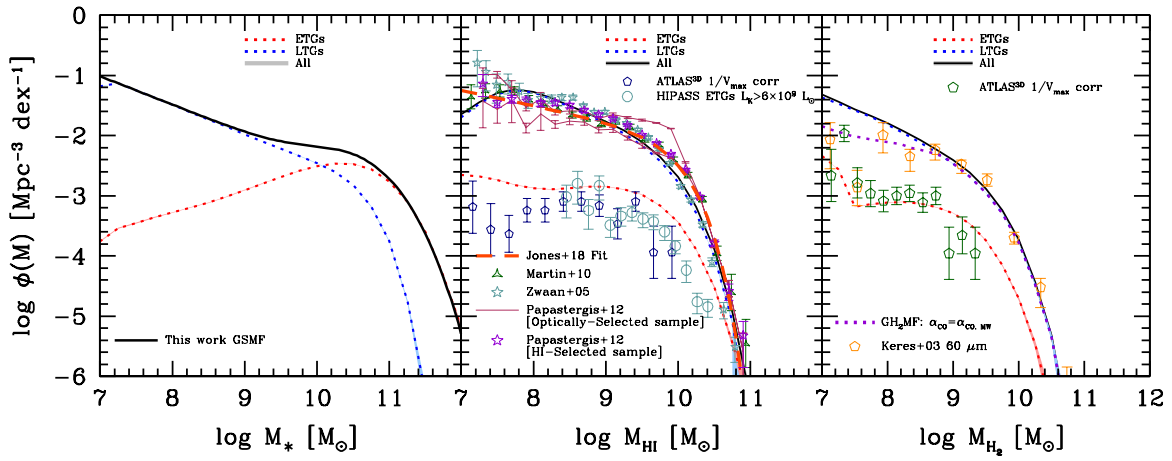
Figure 2.16: *Panel (a):* Total GSMF from the mock catalog that reproduces the empirical GSMF of Fig. 2.15 (solid line). The gray shadow represents the Poisson errors (except for large masses, these errors are thinner than the line thickness). The GSMF from the mock catalog samples very well the empirical GSMF used as input. The blue/red dotted lines and shadows correspond to the LTG/ETG mass function components, using the empirical ETG fraction as a function of M_* shown in Fig. 2.15. *Panel (b):* Same as in panel (a) but for atomic gas, using the mean R_{HI}–M_* relation and its scatter distribution as given in section 2.6. Several observational GHIMF's from blind HI samples, and the ETG GHIMF from ATLAS[3D] and HIPASS surveys are reproduced (see labels inside the panel). *Panel (c):* Same as in panel (a) but for molecular gas. The GH$_2$MF calculated from the Keres et al. (2003) L$_{CO}$ function is reproduced. The dotted purple line is the total GH$_2$MF from the mock catalog when using a R_{H_2}–M_* correlation obtained from our compilation but assuming that $\alpha_{CO}=\alpha_{CO,MW}=$const., as done in Keres et al. (2003).

4. The same procedure as in the previous item is applied to assign $M_{H_2}=R_{H_2}\times M_*$, by using for the $P_j(M_{H_2}|M_*)$ probability distributions the corresponding mass-dependent PDFs for LTGs and ETGs presented in Equations (2.6.1) and (2.6.5), respectively (their parameters are given in Table 2.6).

Our mock galaxy catalog is a volume-complete sample of 5×10^6 galaxies above $M_* = 10^7$ M$_\odot$, corresponding to a co-moving volume of 5.08×10^7 Mpc^3. Since the HI and H$_2$ mass functions are constructed from the GSMF, its mass limit $M_{*,min}$ will propagate in different ways to these mass functions. The co-moving volume in our mock galaxy catalog is big enough as to avoid significant effects from Poisson noise. This noise affects specially the counts of massive galaxies, which are the less abundant objects.

2.7.1 The mock galaxy mass functions

Stellar mass function

The mock GSMF is plotted in panel (a) of Fig. 2.16 along with the Poisson errors given by the thickness of the gray line; except for the highest masses, the Poisson

errors are actually thinner than the line. The mock GSMF is an excellent realization of the empirical GSMF (compare it with Fig. 2.15). We also plot the corresponding contributions to the mock GSMF from the LTG and ETG populations (blue and red dashed lines). As expected, LTGs dominate at low stellar masses and ETGs dominate at high stellar masses. The contribution of both populations is equal ($f_{\rm early} = f_{\rm late} = 0.5$) at $M_*^{\rm cross} = 10^{10.20}$ M$_\odot$ (recall that the fraction $f_{\rm early}$ used here comes from Moffett et al. (2016a), who included Sa galaxies as ETGs; if consider Sa galaxies as LTGs, then $M_*^{\rm cross}$ would likely be higher). In order to predict accurate gas and baryonic mass functions, the present analysis will be further refined in Chapter 4 using results from 3 where several sources of systematic uncertainty in the GSMF measurement and in the definition of the LTG/ETG fractions are taken into account. Our aim here is only to test whether the empirical correlations derived in Section 2.5 are roughly consistent or not with the total HI and H$_2$ empirical mass functions.

HI mass function

In panel (b) of Fig. 2.16, we plot the predicted GHIMF from our mock galaxy catalog using the mean (LTG+ETG) $R_{\rm HI}$–M_* relations and their scatter distributions as given in section 2.6 (black line, the gray shadow shows the Poisson errors). For comparison, we plot also the HI mass functions estimated from the blind HI surveys ALFALFA (Martin et al., 2010; Papastergis et al., 2012, for both their HI- and optically-selected samples; and the latest results from Jones et al., 2018) and HIPASS (Zwaan et al., 2005). At masses larger than $M_{HI} \sim 3 \times 10^{10}$ M$_\odot$, our GHIMF is in vey good agreement with those from the ALFALFA survey but significantly above than the HIPASS one. Martin et al. (2010) argue that the larger volume of ALFALFA survey compared to the HIPASS one, makes ALFALFA more likely to sample the mass function at the highest masses, where objects are very rare. The volume of our mock catalog is even larger than the ALFALFA one. At intermediate masses, $9 \lesssim \log(M_{HI}/{\rm M}_\odot) \lesssim 10.5$, our GHIMF is in reasonable agreement with the observed mass functions but it has in general a slightly less curved shape than these functions. At low masses, $\log(M_{HI}/{\rm M}_\odot) \lesssim 8$, the observed GHIMF's flatten more than our predicted mass function. It could be that the blind surveys start to be incomplete due to sensitivity limits in the radio observations. Note that Papastergis et al. (2012) imposed additional optical requirements to their HI blind sample (see their Section 2.1), which make flatter the low-mass slope. Regarding the optically-selected sample of Papastergis et al. (2012), since it is constructed from a GSMF that starts to be incomplete below $\log(M_*/{\rm M}_\odot) \sim 8$ (see Fig. 2.15), one expects incompleteness in the GHIMF starting at a larger mass in HI. Since our GHIMF is mapped from a volume-

complete GSMF from $M_{*,\min} \approx 10^7$ M$_\odot$, "incompleteness" in M_{HI} is expected to start from the HI masses corresponding to $M_{*,\min} \times P(R_{\mathrm{HI}}|M_{*,\min})$, where the latter is the scatter around the R_{HI}–M_* relation. This shows that our GHIMF can be considered complete from $\log(M_{HI}/\mathrm{M}_\odot) \approx 8$. The slope of the GHIMF around this mass is -1.52, steeper than the slope at the low-mass end of the corresponding GSMF ($\alpha = -1.47$).

In Fig. 2.16 are also plotted the LTG and ETG components of the GHIMF as obtained from our mock catalog. The GHIMF is totally dominated by the contribution of LTGs. Our ETG GHIMF is compared with the ones obtained from observations by using the ATLAS[3D] and HIPASS surveys as reported in Lagos et al. (2014).

H$_2$ mass function

In panel (c) of Fig. 2.16, we plot the predicted GH$_2$MF from our mock galaxy catalog using the mean (LTG+ETG) R_{H_2}–M_* relations and their scatter distributions as given in section 2.6 (black line, the gray shadow shows the Poisson errors). We compute the H$_2$ mass function from the CO luminosity function derived by Keres et al. (2003), who used the small and incomplete FCRAO CO survey (Young et al., 1995) and combined it with the volume-complete FIR survey. We adopt the MW H$_2$-to-CO conversion factor and correct their h parameter to 0.7. Unfortunately, this derivation is highly uncertain since is based on a empirical correlation between the $60\mu m$ and CO luminosities, and the selection effects in both used surveys introduce several biases. The obtained GH$_2$MF is plotted in Fig. 2.16. Our GH$_2$MF decreases faster than the one by Keres et al. (2003) at high masses, roughly agrees with it at intermediate masses, and for masses below $\log(M_{H_2}/\mathrm{M}_\odot) \sim 8.5$, our mass function is steeper. The reason for this latter difference seems to be the mass-dependent CO-to-H$_2$ conversion factor introduced by us (see section 2.2.3). This factor increases as M_* is smaller while in the case of Keres et al. (2003) it is constant. We recalculate the GH$_2$MF by using in the conversion from L_{CO} to M_{H_2} a constant CO-to-H$_2$ factor equal to the MW value, and plot it with the purple dotted line; the mass function at the low-mass side is now in good agreement with that of Keres et al. (2003).

In Fig. 2.16 are also plotted the LTG and ETG components of the GH$_2$MF as obtained from our mock catalog. The GH$_2$MF is totally dominated by the contribution of LTGs. Our ETG GH$_2$MF is compared with the one obtained from observations by using the ATLAS[3D] survey as reported in Lagos et al. (2014).

2.8 Discussion

2.8.1 The H_2-to-HI mass ratio

The global H_2-to-HI mass ratio of a galaxy characterizes its global efficiency of converting atomic into molecular hydrogen. This efficiency is tightly related to the efficiency of large-scale SF in the galaxy (see e.g., Leroy et al., 2008). From the empirical correlations inferred in Section 2.5, we can calculate M_{H_2}/M_{HI} as a function of M_* for both the LTG and ETG populations. We do this by using our double power-law fits to the data. Left panel of Fig. 2.17 presents the obtained M_{H_2}/M_{HI}–M_* relations and their 1σ scatter calculated by propagating the dispersions in the assumption of null covariance. In this sense, the plotted scatter are upper limits, since there is evidence of some (weak) correlation between the HI and H_2 content of galaxies, in particular among those deficient in HI and H_2 (Boselli et al., 2014b). We can plot the same correlations from the mock catalog presented in Section 2.7, which samples the observed GSMF, the LTG and ETG fractions as a function of M_*, and the empirical correlations inferred by us. The middle panels of Fig. 2.17 present what we measure from the mock catalog for LTG (blue), ETG (red), and all galaxies (gray). The lines are the logarithmic means in small mass bins and the shaded regions are the corresponding standard deviations. At low masses, LTGs dominate, so the correlation of all galaxies is practically the one of LTGs. At high masses, ETGs become more important.

According to Fig. 2.17, the molecular-to-atomic mass ratio of LTGs increases with M_*, albeit with a large scatter. On average, M_{H_2}/M_{HI} increases from ≈ 0.1 to ≈ 0.8 for masses ranking from $M_* = 10^8$ $M_\odot$ to 3×10^{11} $M_\odot$. Given that the surface density of LTGs correlates significantly with M_*, one can expect this dependence of M_{H_2}/M_{HI} on M_* at least from two arguments: 1) Disk instabilities, which drive the formation of molecular clouds (e.g., the Toomre criterion Toomre, 1964), are more probable to occur as the disk surface density is higher. 2) The H_2-to-HI mass ratio in galaxies has been shown to be directly related to the hydrostatic gas pressure (Blitz & Rosolowsky, 2006; Krumholz et al., 2009), and this pressure depends on the (gas and stellar) surface density (Elmegreen, 1989). In fact, the physics of H_2 condensation from HI is very complex and it is expected to be driven by local parameters of the ISM (see e.g., Blitz & Rosolowsky, 2006; Krumholz et al., 2009; Obreschkow & Rawlings, 2009). Therefore, the dependence of the H_2-to-HI mass ratio on M_* should be understood as consequence of the correlations of these parameters (their mean values along the galaxy) with M_*, introducing this actually a large scatter in the dependence
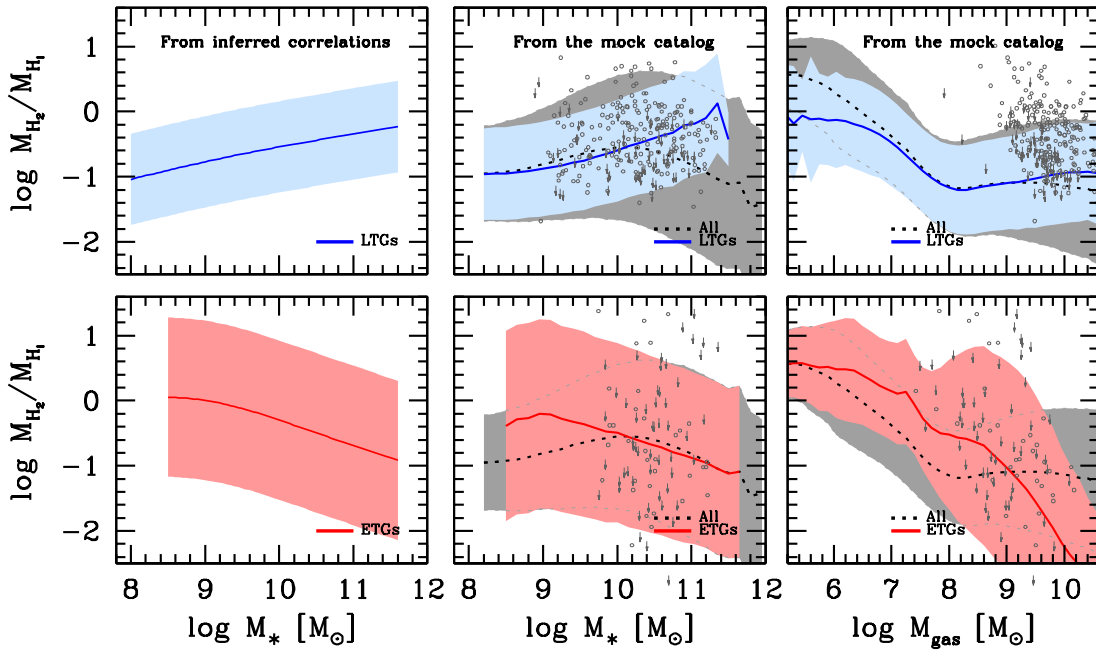
Figure 2.17: *Left panels:* Molecular-to-atomic mass ratio, M_{H_2}/M_{HI}, for LTGs (upper panel) and ETGs (lower panel) inferred from our double power-law fits to the R_{HI}-M_* and R_{H_2}-M_* correlations. The shaded areas are the 1σ scatter obtained by error propagation of the scatter around the R_{HI}-M_* relations. *Middle panels:* Same as in left panels but from our mock catalog generated to sample the empirical GSMF, volume-complete ETG/LTG fractions as a function of mass, and R_{HI}–M_* and R_{H_2}–M_* correlations. The dotted line surrounded by the gray area are the total M_{H_2}/M_{HI} ratio and 1σ scatter as a function of stellar mass. *Right panels:* Molecular-to-atomic mass ratio as a function of the cold gas mass, M_{gas}, from the mock catalog for LTGs (upper panel) and ETGs (lower panel). We plot available detected and undetected cold gas observational data as gray unfilled circles and downward arrows respectively.

of M_{H_2}/M_{HI} on M_*. Indeed, several authors have shown that M_{H_2}/M_{HI} correlates better with the mean gas-phase metallicity or mean stellar surface density than with M_* (e.g., Saintonge et al., 2011; Boselli et al., 2014a).

For ETGs, the trend of the H_2-to-HI mass ratio is inverse to the one of LTGs and with a very large scatter. The ETGs more massive than $\sim 10^{11}$ $M_\odot$ have mean ratios around 0.15 and a 1-σ scatter of $\sim \pm 1$ dex; for intermediate masses, this ratio increases on average, and for ETGs with masses $M_* \sim 10^9$ $M_\odot$, which are actually very rare, their mean H_2-to-HI mass ratios are ~ 1 with the same scatter of $\sim \pm 1$ dex. Even though the gas fraction in ETGs is much smaller than in LTGs at all masses (see Fig. 2.8), the former are also typically more compact than the latter, resulting probably on average in similar or higher gas pressures, and consequently a similar or even higher M_{H_2}/M_{HI} ratios, specially at masses lower than $M_* \approx 10^{10}$ $M_\odot$. In fact, given the large scatter in M_{H_2}/M_{HI} for ETGs, this ratio depends likely on many other internal and external (mergers, environment, etc.) factors that do not correlate significantly with M_*.

Regarding M_{H_2}/M_{HI} vs. $M_{\rm gas}$, for LTGs, which for $M_* > 10^7$ $M_\odot$ have mostly gas masses $> 10^8$ $M_\odot$, there is not any significant dependence, while for ETGs, which are almost inexistent with $M_{\rm gas} \gtrsim 10^9$ $M_\odot$, M_{H_2}/M_{HI} is larger on average for lower values of $M_{\rm gas}$. This can be seen in the right panel of Fig. 2.17, where the mock catalog has been used. Basically, for a given $M_{\rm gas}$, in the mass range $M_{\rm gas} \sim 10^7 - 10^9$ $M_\odot$), ETGs have typically larger H_2-to-HI mass ratios than LTGs. In combination, the H_2-to-HI ratio appears to be larger for lower values of $M_{\rm gas}$. Such a dependence has been reported by Obreschkow & Rawlings (2009) for their compiled sample of galaxies, and predicted by these authors from a physical model.

The dependences of the H_2-to-HI mass ratio on M_*, $M_{\rm gas}$, and morphological type discussed above are in qualitative agreement with several previous observational works, which actually are part of our compilation (Leroy et al., 2008; Obreschkow & Rawlings, 2009; Saintonge et al., 2011; Boselli et al., 2014a; Bothwell et al., 2014). However, our results extend to a larger mass range and separate explicitly the two main populations of galaxies.

2.8.2   The role of environment

There are several pieces of evidence that the atomic gas fraction of galaxies is lower in higher-density environments (e.g., Haynes & Giovanelli, 1984; Gavazzi et al., 2005; Cortese et al., 2011; Catinella et al., 2013; Boselli et al., 2014b). The fact that the ETG population has lower HI gas fractions than the LTG one (Section 2.5), being the former commonly found in higher-density environments, agrees with the mentioned

trends with environment. Thus, due to the morphology-density relation, our determinations of the R_{HI}–M_* (as well as $R_{\mathrm{H_2}}$–M_* and R_{gas}–M_*) correlations for the LTG and ETG populations, account partially for the dependence of these correlations on environment. Moreover, for the very isolated LTGs and for the subsample of LTGs in the Virgo cluster central regions, we confirm higher and lower HI-to-stellar mass ratios than the average of the overall LTG sample, respectively (see subsection 2.4.1). However, this systematic difference with the environment is within the 1σ scatter of the R_{HI}–M_* correlation of LTGs (see Fig. 2.6). Instead, in the case of ETGs, the isolated galaxies have much larger R_{HI} values than the means of all ETGs, above the 1σ scatter; isolated ETGs are almost as HI gas rich as the mean of LTGs.

For molecular gas fraction, the observational results are controversial in the literature. Recent studies seem to incline the controversy to the fact that galaxies in clusters are actually H_2–deficient with respect to similar galaxies in the field, however, the deficiencies are smaller than in the case of HI (Boselli et al., 2014b, and more references therein). Here, for isolated and Virgo-center LTGs, we do not see any systematic segregation of $R_{\mathrm{H_2}}$ from the rest of our compiled LTGs (Fig. 2.7), but in the case of ETGs, the isolated galaxies have on average larger values of $R_{\mathrm{H_2}}$.

In summary, the results from our compilation point out that the HI content of LTGs has a (weak) dependence on environment, mainly due to the fact that at high densities LTGs are HI deficient. Instead, the H_2 content of LTGs seems not to change on average with the environment. In the case of ETGs, those very isolated are significantly more gas rich (both in HI and H_2) than the average among ETGs at a given mass.

An important aspect related to the environment is whether a galaxy is central or satellite. The local environmental effects once a galaxy becomes a satellite inside a halo (ram pressure and viscous stripping, starvation, harassment, tidal interactions) work in the direction of lowering the gas content of the galaxy, likely more as more massive is the halo (Boselli & Gavazzi, 2006; Brown et al., 2017). Part of the scatter in the gas-to-stellar mass correlations are probably due to the external processes produced by these local-environment mechanisms. A result in this direction has been recently shown for the R_{HI}–M_* correlation by Brown et al. (2017). These authors have found that the HI content of satellite galaxies in more massive halos have, on average, lower HI-to-stellar mass ratios at fixed stellar mass and specific SFR. According to their analysis, the systematic environmental suppression of HI content at both fixed stellar mass and fixed specific SFR in satellite galaxies begins in halo masses typical of the group regime ($> 10^{13}$ $M_\odot$), and fast-acting mechanisms such as ram-pressure stripping are suggested to explain their results. In a future work, we will attempt

to characterize the central/satellite nature of our compiled galaxies, as well as to calculate a proxy to their halo masses, in order to study this question.

2.8.3 Comparisons with previous works

In Fig. 2.18 we compare our results with those of previous works. When necessary, the data are corrected to a Chabrier (2003) IMF. Most of the previous determinations of the HI- and H_2-to-stellar mass correlations are not explicitly separated into the two main galaxy populations as done here, and in several cases non detections are assumed to have the values of the upper limits or are not taken into account at all.

In the upper panel, our empirical R_{HI}–M_* correlations for LTGs and ETGs are plotted along with the linear relations given by Stewart et al. (2009) (cyan line, the dashed lines show the 1σ scatter) and Papastergis et al. (2012) (gray line). The former authors used mainly the observational data presented in McGaugh (2005) for disk-dominated galaxies, and the latter authors used samples from Swaters & Balcells (2002), Garnett (2002), Noordermeer et al. (2005), and Zhang et al. (2009), which refer mostly to late-type galaxies. Their fits are slightly above the mean of our LTG R_{HI}–M_* correlation. This is likely because they ignore non-detections. We also plot the logarithmic average values in mass bins reported by Catinella et al. (2013) for GASS (green open circles). Since ETGs progressively dominate in number as the mass increase, our total (density-weighted) R_{HI}–M_* correlation would fall below the one by Catinella et al. (2013), specially at the highest masses. Note that for the data plotted from Catinella et al. (2013), the HI masses of non-detections were set equal to their upper limits. Therefore, the plotted averages are biased to high values of R_{HI}, specially for ETGs which are dominated by non-detections. On the other hand, recall that we have corrected by distance the upper limits of GASS to make them compatible with those of the closer ATLAS^{3D} survey.

More recently, Brown et al. (2015) have used the HI spectral stacking technique for a volume-limited, stellar mass selected sample from the intersection of SDSS DR7, ALFALFA, and $GALEX$ surveys. With this technique the stacked signal of co-added raw spectra of detected and non-detected galaxies (about 80% of the ALFALFA selected sample) is converted into a (lineal) average HI mass. The authors have excluded from their analysis HI-deficient galaxies –typically found within clusters– because of their significant offset to lower gas content. The black dots connected by a dotted line show the logarithm of the average R_{HI} values reported at different stellar mass bins in Brown et al. (2015). Since HI-deficient galaxies –which typically are ETGs– were excluded, then the Brown et al. (2015) correlation should be compared with our correlation for LTGs. Note that with the stacking technique is not possible to obtain

the population scatter in R_{HI} because the reported mean values come from stacked spectra instead from averaging individual values of detections and non detections. However, the stacking can be applied to subsets of galaxies, for example, selected by color. Brown et al. (2015) have divided their sample into three groups by their NUV$-r$ colors: [1,3), [3,5), and [5,8]. The average R_{HI} values at different masses corresponding to the bluest and reddest groups are reproduced in Fig. 2.18 with the blue and red symbols, respectively. Note that the logarithmic mean is lower than the logarithm of the mean. For a lognormal distribution, $\langle \log x \rangle = \log \langle x \rangle - 0.5 \times \sigma^2_{\log x} \ln 10$ (see e.g., Rodríguez-Puebla et al., 2017). Then, for the typical scatter of 0.44 dex corresponding to LTGs, the logarithm of the stacked values of R_{HI} should be lowered by ≈ 0.2 dex to compare formally with our reported values of logarithmic means; this is shown with a black arrow in Fig. 2.18. If the reddest galaxies in the Brown et al. (2015) stacked sample are associated with ETGs (which is true only partially), then for them the correction to a logarithmic mean is of ≈ 1 dex, shown with a red arrow.

Finally, recently van Driel et al. (2016) reported the results from HI observations at the Nancay Radio Telescope (NRT) of 2839 galaxies selected evenly from SDSS. The authors present a Buckley-James linear regression to their data (long-dashed green line in Fig. 2.18), taking into account this way upper limits for non-detections (though their upper limits are quite high given the low sensitivity of NRT). Their fit is for all the sample, that is, they do not separate into LTG/ETG or blue/red groups. In a subsequent paper (Butcher et al., 2016), the authors obtained ~ 4 times more sensitive follow-up HI observations at Arecibo for a fraction of the galaxies that were either not detected or marginally detected; 80% of them were detected with HI masses ~ 0.5 dex lower than the upper limits in van Driel et al. (2016), and the rest, mostly luminous red galaxies, were not detected. If this trend is representative of the rest of the NRT undetected galaxies, Butcher et al. (2016) expect the fit plotted in Fig. 2.18 to be offset toward lower R_{HI} values by about 0.17 dex and even more at the highest masses. This fit is in between a density-weighted fit to our two correlations when taking into account that at high masses the fraction of ETG/red galaxies increases and at low masses LTG/blue galaxies dominate at all.

The lower panel of Fig. 2.18 is similar to the upper panel but for the R_{H_2}–M_* correlations. In the case of the molecular gas content, in the literature there are only a few attempts to determine the relation between M_{H_2} and M_*. In fact, those works that report approximate correlations are included in our compilation: Saintonge et al. (2011) for COLD GASS, and Boselli et al. (2014a) for HRS. The former authors report a linear regression to their binned data assuming H_2 masses for non-detection set equal to their upper limits. The latter authors present a bisector fit using only detected,

Figure 2.18: *Upper panel:* Our empirical HI-to-stellar mass correlations for LTGs and ETGs (blue and red shaded areas, respectively) compared with some previous determinations (see labels inside the panel and details of each determination in the text). These previous determinations are for compilations typically biased to late-type, blue galaxies, and/or do not take into account non detections. The blue and red arrows correspond to estimates of the difference between the logarithm of the mean (the stacking technique provides the equivalent of the mean value) and the logarithmic mean (our determinations are for this case) for standard deviations of 0.52 and 0.99 d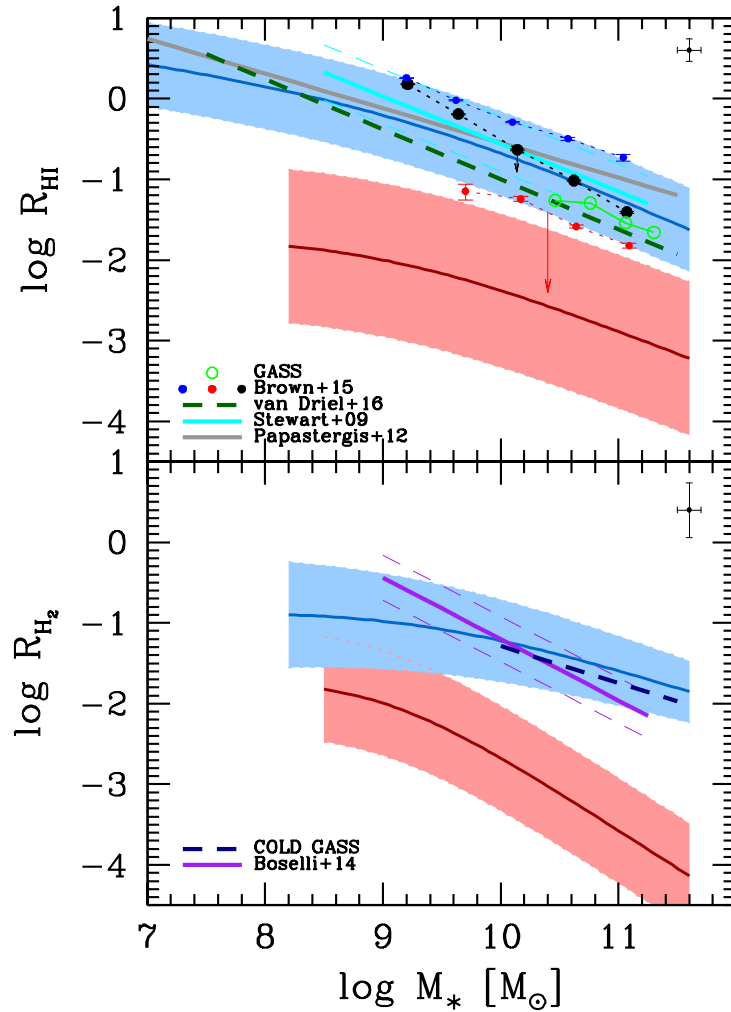ex, respectively (see text for more details). *Lower panel:* Our empirical molecular H_2-to-stellar mass correlations for LTGs and ETGs (blue and red shaded areas, respectively) compared with very rough previous determinations not separated into LTGs and ETGs (see labels inside the panel and details of each determination in the text).

late-type gas-rich galaxies. Therefore, in both cases the reported relations are clearly biased to LTGs and to the side of high R_{H_2} values.

The differences we find between our correlations and those plotted in Fig. 2.18, as discussed above, can be understood on the basis of the different limitations that present each one of the previous works. Having in mind these limitations in each concrete case, we can conclude that the correlations presented here are in rough agreement with previous ones but with respect to them (i) extend the correlations to a larger mass range, (ii) separate explicitly galaxies into their two main populations, and (iii) take into account adequately the non detections.

2.9   Summary and Conclusions

The fraction of stars and atomic and molecular gas in local galaxies is the result of complex astrophysical processes across their evolution. Thus, the observational determination of how these fractions vary as a function of mass provides key information on galaxy evolution at different scales. Before the new generation of radio telescopes, which will bring extragalactic gas studies more in line with optical surveys, the main way to get this kind of information is from studies based on radio follow-up observations of (small) optically-selected galaxy samples. In this work, we have compiled and homogenized from the literature samples with information on M_* and M_{HI} and/or M_{H_2} for galaxies that can be identified belonging to two main operational (in a statistical sense) groups: the LTG and ETG populations. For estimating M_{H_2} from CO observations, we have introduced a mass-dependent CO-to-H_2 conversion factor in agreement with studies that show that this factor is not constant and depends on metallicity (hence, statistically on mass). Results using a constant CO-to-H_2 factor were also presented. Figures 2.1 and 2.3 summarize our compilation in the R_{HI} vs. M_* and R_{H_2} vs. M_* logarithmic diagrams.

Previous to infer the correlations, we have tested how much each one of the compiled samples deviate from the rest and classified them into three categories: (1) samples complete in limited volumes (or selected from them) without selection effects that could affect the calibration of the correlations (Golden), (2) samples that are not complete but are representative of the average galaxy population, without obvious selection effects (Silver), and (3) samples selected by environment (Bronze). We showed that most of the samples, after our homogenization, are suitable to infer the R_{HI}–M_* and R_{H_2}–M_* correlations, except those from the Bronze category in the case of ETGs. These galaxies in extreme environments show significant deviations from the mean trends, and then are not taken into account in our determinations. From

the combination of all the chosen samples, we have calculated the mean, standard deviation, and percentiles of the logarithms of the R_{HI} and $R_{\mathrm{H_2}}$ mass ratios in several stellar mass bins, taking into account non-detected galaxies and their reported upper limits, which are a non-negligible fraction of the data, specially for the ETG population. The accounting of non-detected galaxies and their homogenization among different samples are relevant for determining the gas-to-stellar mass correlations of ETGs.

The mean logarithmic values in mass bins, $\langle \log R_{\mathrm{HI}} \rangle$ and $\langle \log R_{\mathrm{H_2}} \rangle$, with the corresponding (intrinsic) standard deviation calculated by means of the Kaplan-Meier estimator were fitted to the logarithm of single and double power-law functions (Eq. 2.5.1). The parameters of the best fits to these functions, both for LTGs and ETGs, are reported in Tables 2.4 and 2.5, respectively. We highlight the following results from our analysis:

- The R_{HI}–M_* and $R_{\mathrm{H_2}}$–M_* correlations for the LTG and ETG populations, can be described roughly equally well by a single or double power law at masses larger than $\log(M_*/\mathrm{M_\odot}) \gtrsim 9$. For smaller masses, we see some hints of a flattening in these correlations. LTGs have significantly higher HI and H_2 gas fractions than ETGs, the differences increasing at the high- and low-stellar mass ends. For the ETG population, the scatter of the R_{HI}–M_* and $R_{\mathrm{H_2}}$–M_* correlations are much larger than for the LTG one.

- Combining the R_{HI}–M_* and $R_{\mathrm{H_2}}$–M_* correlations and propagating errors, we calculated the cold gas ($M_{\mathrm{gas}}=1.4(M_{HI}+ M_{H_2})$)-to-stellar mass correlations of the LTG and ETG populations. For the former, R_{gas} is around 4 on average at $M_* = 10^7$ $\mathrm{M_\odot}$ and ≈ 1 at $M_* = 1.60 \times 10^9$ $\mathrm{M_\odot}$. At larger masses, R_{gas} continues decreasing sifnificantly. For the ETG population, R_{gas} on average is smaller than 1 even for the smallest galaxies. Galaxies as massive as $M_* = 10^{11}$ $\mathrm{M_\odot}$ have on average R_{gas} ratios smaller than 2.5×10^{-3}. The intrinsic standard deviation of the R_{gas}–M_* correlation of the LTG population is ≈ 0.44 dex while for the ETG one is larger, around 0.68 dex.

- The H_2-to-HI mass ratio implied by our correlations is such that for LTGs, increases on average with M_*, from ≈ 0.1 to 0.8 for masses ranking from $M_* = 10^8$ $\mathrm{M_\odot}$ to 3×10^{11} $\mathrm{M_\odot}$. For ETGs, the trend is the opposite but with large scatter (standard deviation of $\sim \pm 1$ dex). While ETGs have much less gas content than LTGs, the H_2-to-HI mass ratio at intermediate and low masses is higher on average in the former than in the later, and lower at large masses.

- In an attempt to describe the full distributions of R_{HI} and $R_{\mathrm{H_2}}$ as a function of M_* for both the LTG and ETG populations, the respective PDFs from the cen-

sored+uncensored data in different mass bins provided by the Kaplan-Meier estimator were used. For LTGs, we have found that a Schechter function with their parameters depending on M_* offers a good description of the R_{HI} and $R_{\mathrm{H_2}}$ distributions as a function of M_* (Eq. 2.6.1). For ETGs, these distributions look bimodal, with a (broken) Schechter function and a uniform distribution at the low-end side providing an approximate description of them (Eq. 2.6.5). These mass-dependent PDFs offer a full description of the R_{HI}–M_* and $R_{\mathrm{H_2}}$–M_* relations and their scatter distributions for both LTGs and ETGs. Their first and second moments agree very well with our previously determined double power-law correlations (Figures 2.13 and 2.14).

• The mass-dependent distribution functions of R_{HI} and $R_{\mathrm{H_2}}$ were used to map the GSMF into the corresponding HI and H_2 mass functions, both for LTGs and ETGs. We use an empirical GSMF from the combination of GSMFs from a low-z survey and from the overall DR7 sample, following Kravtsov et al. (2014). The fractions of LTGs/ETGs as a function of M_* are calculated from the fitted mass functions of ETGs obtained by Moffett et al. (2016a) using the GAMA survey. The predicted total HI and H_2 mass functions agree with those obtained from empirical determinations in the mass ranges where these determinations are reliable.

Our (marginal) finding of a flattening in the HI- and H_2-to-stellar mass correlations at low masses has been suggested in some previous works (see Section 2.5 for references). For our double power-law fits (Eq. 2.5.1), we find that the transition mass M_*^{tr} is around $1 - 2 \times 10^9$ $\mathrm{M_\odot}$ for both the R_{HI}–M_* and $R_{\mathrm{H_2}}$–M_* correlations and for both the LTG and ETG populations. Interestingly enough, this is the mass that roughly separates normal and dwarf galaxies.

We are aware that our determination of the gas-to-stellar mass relations come from an heterogeneous mix of samples. However, we have shown that there are not significant differences in the R_{HI} and $R_{\mathrm{H_2}}$ values as a function of M_* from volume-limited complete and incomplete samples. Significant differences are observed only for samples selected by environment in the case of ETGs. On the other hand, our correlations for ETGs (and LTGs in the case of molecular gas), are very limited at low masses. They are actually just extrapolations for stellar masses below several 10^8 $\mathrm{M_\odot}$, but we have checked them to be consistent with the very few available determinations (mostly non detections) below these masses.

In spite of the mentioned shortcomings, it is encouraging that the correlations (in fact, the full mass-dependent distributions), when mapped to the HI and H_2 mass functions using the observed GSMF as an interface, are consistent with the mass functions determined from observational radio surveys, at least in the mass ranges where these surveys do not suffer of strong selection, volume, and cosmic

variance effects. Such a self-consistency between the gas-to-stellar correlations and mass functions supports the reliability of our results, which help to pave the way for the next generation of radio telescopes.

The empirical gas-to-stellar mass correlations and their approximate scatter distributions presented in this chapter for the two main populations of galaxies, are useful for understanding global aspects of galaxy evolution as a function of mass. We encourage to use these correlations (or the full mass-dependent PDFs) for comparisons with predictions of models and simulations of galaxy formation and evolution.

Finally, we provide upon request to A. R. Calette a Python-based code that allows to generate plots and electronic tables for both LTGs and ETGs of 1) the $R_{\rm HI}$-M_* and $R_{\rm H_2}$-M_* double power-law relations and their 1σ intrinsic scatters as presented in Fig. 5 and Table 6; and 2) the mass-dependent full $R_{\rm HI}$ and $R_{\rm H_2}$ PDFs as constrained in Section 5, including the first and second moments (mean and standard deviation) of these PDFs.

Chapter 3

The GSMF of all, early- and late-type galaxies

El contenido de este Capítulo constituyó una parte del artículo publicado: Rodriguez-Puebla, A.; Calette, A. R.; Avila-Reese, V.; Rodriguez-Gomez, V.; Huertas-Company, M. 2020, Publications of the Astronomical Society of Australia, 37, 24, article id. e024[1]

ABSTRACT

In this Chapter, we present the GSMF of all, early- and late- type galaxies from $M_* \sim 3 \times 10^7$ to 3×10^{12} $M_\odot$ by combining two spectroscopic samples from the Sloan Digital Sky Survey at the redshift range $0.0033 < z < 0.2$ and Huertas-Company et al. (2011) automated morphological classification. We take into account volume and large-scale structure as well as surface brightness incompleteness corrections. We find that the low-mass end slope of the GSMF, is $\alpha \sim -1.4$, consistent with previous determinations. We estimate the impact of systematics due to mass-to-light ratios and find that our MFs are robust against systematic errors. These results are key ingredients for the results presented in Chapters 4 and 5.

3.1 Introduction

There are many studies that have determined the GSMF in the past, nevertheless, they do not typically report systematic errors or do not deconvolve it from random errors (with a few exceptions, e.g., Bernardi et al., 2010, 2017; Obreschkow et al., 2018) or they are limited in the dynamical range of M_* due to the limited depth of the sample and/or the cosmic variance in the galaxy sample (but see Wright et al., 2017).

In this Chapter we compute the GSMF taking into account what is mentioned above and its decomposition into early- and late-type galaxies. We combine here two

---

[1]Los resultados del artículo relacionados a este Capítulo no son parte medular del proyecto de Tesis pero fue imperioso obtenerlos pues son un ingrediente clave para los cálculos presentados en los siguientes capítulos. Mi contribución al artículo publicado en la parte relacionada a este Capítulo ha sido a la par del primer autor en lo que se refiere a cálculos y desarrollo de códigos pero la conducción del trabajo no fue realizada por mí.

large galaxy samples, the low redshift sample, `low-z`, from the NYU SDSS DR4 (Blanton et al., 2005a,c), and the new photometry pipeline for the SDSS DR7 from Meert et al. (2015, 2016). The `low-z` sample suffers from surface brightness incompleteness, but here we estimate and correct for the fraction of missing galaxies due to this selection effect. As for the SDSS DR7, the new photometry from Meert et al. (2015) shows that galaxy magnitudes were previously underestimated due to sky subtraction problems (see also, Simard et al., 2011); the impact of these new determinations has been studied previously in Bernardi et al. (2017). We extend the Bernardi et al. (2017) analysis by using not only different definitions of galaxy stellar masses but by dividing into two morphology groups, early- and late-type galaxies.

3.2   The Galaxy Samples

To estimate the GSMF over a large dynamical range we use two galaxy samples. Next, we shortly describe the procedure and our determinations.

1) For masses above $M_* = 10^9$ M$_\odot$, we use the SDSS DR7 based on the photometric catalog from Meert et al. (2015) and Meert et al. (2016)[2] at the redshift interval $0.005 < z < 0.2$. Previous studies have concluded that the measurements of the apparent brightnesses based on the standard SDSS pipeline photometry are underestimated due to sky subtraction problems, particularly, in crowded fields (Bernardi et al., 2010; Blanton et al., 2011; Simard et al., 2011; Bernardi et al., 2013; He et al., 2013; Mendel et al., 2014; Kravtsov et al., 2014; Meert et al., 2015; D'Souza et al., 2015; Bernardi et al., 2016; Meert et al., 2016). New determinations of the GSMF based on the new algorithms for obtaining more precise measurements of the sky subtraction, and thus to improve the photometry, have concluded that the bright end of the luminosity/mass function has been systematically underestimated (Bernardi et al., 2017). While there are various groups working in improving the determination of galaxy apparent brightnesses, see references above, Bernardi et al. (2017) showed that all those studies agreed up to 0.1 dex in the GSMF. In this thesis we use the apparent Sérsic r, g, and i band luminosities reported in Meert et al. (2015) and Meert et al. (2016) derived for the SDSS DR7 based on the PyMorph software pipeline (Vikram et al., 2010; Meert et al., 2013). This software has been extensively tested in Meert et al. (2013) and shows that it does not suffer from sky subtraction problems. All magnitudes and colours are K+E corrected at a redshift rest-frame $z = 0$,

---

[2]Available at `http://www.physics.upenn.edu/?ameert/SDSS_PhotDec/`

see sections 3.3.1 and 3.3.2. As described in section 3.3.1, for every galaxy we estimate M_* from five colour-dependent mass-to-light ratios but we define as our fiducial M_* the geometric mean of all the determinations. Using the $1/V_{\mathrm{max}}$ method, we derive six GSMFs based on the mass definitions described above. Consistent with Bernardi et al. (2017), we find that the differences in mass-to-light ratios introduce large discrepancies in the GSMF, especially at the high-mass end. In Figure 3.1 from Section 3.3.1, we find that a shift of $\sim \pm 0.15$ dex in the M_* axis recovers systematic errors in the GSMF due to different mass-to-light ratios.

2) For masses below $M_* = 10^9 \ \mathrm{M}_\odot$, we use the SDSS DR4 NUY-VAGC `low-z` sample,[3] at the redshift interval $0.0033 < z < 0.005$, and ideal to study the low mass/luminosity galaxies (Blanton et al., 2005a,c). As before, all absolute magnitudes and colours were K+E corrected at a redshift rest-frame $z = 0$. Also, we derive M_* from five colour-dependent mass-to-light ratios and, again, we define our fiducial M_* as the geometric mean of all the determinations. We construct the GSMF using the $1/V_{\mathrm{max}}$ method and include missing galaxies due to surface brightness incompleteness, as described in Section 3.3.3. For surface brightness incompleteness we follow closely the methodology described in Blanton et al. (2005c). The latter correction is relevant for the low-mass end. Based on the conclusions from Baldry et al. (2012), we use a simple correction for the low-mass end in order to correct for the local flow model distances from Willick et al. (1997) to the one by Tonry et al. (2000).

3.3 Volume corrections

3.3.1 The SDSS DR7 galaxy sample

In this work we derive the GSMF from a spectroscopic sample of 670,722 galaxies from the SDSS DR7 based on the photometric estimates of the apparent brightnesses in the g, r and i band from Meert et al. (2015) and Meert et al. (2016). In those papers, the authors selected galaxies with extinction-corrected $r-$band Petrosian magnitude between magnitude 14 and 17.77 to derive de Vacouleurs, Sérsic, de Vacoulers+Exponential, and Sérsic+Exponential fits to the observed surface brightness profiles of each galaxies in their SDSS DR7 catalogue. Surface brightness profiles were obtained via the PyMorph pipeline (Vikram et al., 2010; Meert et al., 2013). PyMorph is a python software that uses Sextractor (Bertin & Arnouts, 1996)

---

[3]Available at `http://sdss.physics.nyu.edu/vagc/lowz.html`
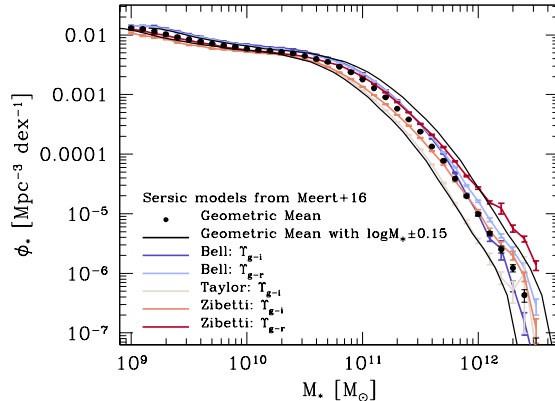
Figure 3.1: The GSMF from our six stellar mass definitions, Equation 3.3.6. Using different stellar masses yield to differences of ~ 0.1 dex at low masses and as high as ~ 1 dex at the high mass-end. In this work we opted to use as our fiducial GSMF as the one derived from the geometric mean of five different stellar masses. The solid lines show a shift of ± 0.15 dex in the stellar mass axis of the our fiducial GSMF, note that it recovers systematics from mass-to-light ratios.

and GALFIT (Peng et al., 2002) to fit both one- an two-components to the seeing convolved surface brightness profiles from the spectroscopic sample of SDSS DR7 galaxies. PyMorph has been extensively tested in Meert et al. (2013, see also, Meert et al., 2015) showning that the algorithm does not suffer from the sky subtraction problems that has been detected in previous studies based on the SDSS, in particular in crowded fields[4].

We estimate the GSMF at the redshift interval between $z = 0.005$ and $z = 0.2$ by using the standard $1/V_{\mathrm{max}}$ method (Schmidt, 1968):

$$\phi_*(M_*) = \frac{1}{\Delta \log M_*} \sum_{i=1}^{N} \frac{\omega_i(\log M_* \pm \Delta \log M_*)}{V_{\mathrm{max},i}}, \qquad (3.3.1)$$

where ω_i is a weight factor correction that depends on the position in the sky for galaxies within the interval $\log M_* \pm \Delta \log M_*/2$, following Bernardi et al. (see also 2010) we assume that $\omega_i = 1.1$; and

$$V_{\mathrm{max},i} = \int_{\Omega} \int_{z_{l,i}}^{z_{u,i}} \frac{d^2 V_c}{dz d\Omega} dz d\Omega. \qquad (3.3.2)$$

---

[4]Recently, various others groups have also improved the determinations of galaxies' surface brightness profiles based on the SDSS by the improving the survey photometry, especially due to sky subtraction problems in crowded fields, (see e.g., Simard et al., 2011; D'Souza et al., 2015, and more reference therein). While in this work we opt to use the photometric catalog from Meert et al. (2015) and Meert et al. (2016), Bernardi et al. (2017) showed that, after a careful comparison, most of those studies agree up to 0.1 dex. Thus, using the photometry derived by other groups will not change significantly our results.

We denote the solid angle of the SDSS DR7 with Ω while V_c refers to the comoving volume enclosed within the redshift interval $[z_{l,i}, z_{u,i}]$. The redshift limits are defined as $z_{l,i} = \max(0.005, z_{\min,i})$ and $z_{u,i} = \min(z_{\max,i}, 0.2)$; where $z_{\min,i}$ and $z_{\max,i}$ are, respectively, the minimum and maximum at which each galaxy can be observed in the SDSS DR7 sample. We estimate $z_{\max,i}$ for every galaxy in the sample by solving iteratively the distance modulus equation:

$$m_{\mathrm{lim,r}} - M_{r,i}^{0.0} = 5 \log D_{L,i}(z_{\max,i}) + 25 \qquad (3.3.3)$$

$$+ K_{gr,i}(z_{\max,i}) - E_{r,i}(z_{\max,i}), \qquad (3.3.4)$$

where $M_{r,i}^{0.0}$ is the Petrosian magnitude K+E-corrected at a rest-frame $z = 0$, $K_{gr,i}(z)$ is the K-correction (see Section 3.3.2) and $E_{r,i} = 1.1z$ (following Dragomir et al., 2018) for the ith galaxy in the sample. For the completeness limits, we use the limiting apparent magnitude in the r-band of $m_{\mathrm{lim,r}} = 17.77$. Similarly, we estimate $z_{\min,i}$ by solving iteratively the distance modulus equation but this time using the limiting apparent magnitude $m_{\mathrm{lim,r}} = 14$.

Errors are estimated using the jackknife technique by diving the galaxy sample into $n = 300$ subsamples of approximately equal size and estimating a $\phi_{*,i}(M_*)$ each time. Thus errors are then given by:

$$\sigma^2 = \frac{n-1}{n} \sum_{i=1}^{n} \left(\phi_{*,i} - \langle \phi_* \rangle\right)^2, \qquad (3.3.5)$$

with $\langle \phi_* \rangle$ as the average of the ensemble.

Stellar masses were derived from several colour-dependent mass-to-light ratios as listed below:

$$M_* = \begin{cases} \Upsilon_r^{\mathrm{B03}}(g-r) \cdot L_r & \text{Bell et al. (2003)} \\ \Upsilon_i^{\mathrm{B03}}(g-i) \cdot L_i & \text{Bell et al. (2003)} \\ \Upsilon_r^{\mathrm{Z09}}(g-r) \cdot L_r & \text{Zibetti et al. (2009)} \\ \Upsilon_i^{\mathrm{Z09}}(g-i) \cdot L_i & \text{Zibetti et al. (2009)} \\ \Upsilon_i^{\mathrm{T11}}(g-i) \cdot L_i & \text{Taylor et al. (2011)} \end{cases} \qquad (3.3.6)$$

and we define our fiducial M_* as the geometrical mean of all the determinations in 3.3.6:

$$M_* = [M_*(\Upsilon_r^{\mathrm{B03}}) \times M_*(\Upsilon_i^{\mathrm{B03}}) \times M_*(\Upsilon_r^{\mathrm{Z09}}) \times \qquad (3.3.7)$$

$$M_*(\Upsilon_i^{\mathrm{Z09}}) \times M_*(\Upsilon_i^{\mathrm{T11}})]^{1/5}$$

Figure 3.2: Colour and redshift dependence of the K-corrections at a rest-frame $z = 0$ for the r (upper left), g (upper right), and i (bottom left) bands from the K-CORRECT algorithm (Blanton & Roweis, 2007a), dashed lines. Solid lines show our best fit models as described in the text.

We use Sérsic apparent magnitudes to derive colours and magnitudes. We apply K+E-corrections at a rest-frame $z = 0$. We use the values reported in Dragomir et al. (2018) for g and i bands given, respectively, by $E_g = 1.3 \times z$ and $E_i = 1.09 \times z$. K-corrections are discuss in Section 3.3.2. We applied a shift of -0.1 dex to the resulting masses from the colour dependent mass-to-light ratios of Bell et al. (2003) to be consistent with the Chabrier (2003) IMF adopted in this thesis.

Figure 3.1 shows the resulting GSMFs as described above. The figure shows that using different recipes for deriving stellar masses yield differences of ~ 0.1 dex at low masses and between $\sim 0.5 - 1$ dex at the high mass-end. This is consistent with the recent study by Bernardi et al. (2017) which showed that differences in mass-to-light ratios introduce discrepancies in the GSMF around ~ 0.5 dex. As a fiducial estimation of the GSMF in this work we opt to utilise the geometric mean of all the masses derived based on the colour dependent mass-to-light ratios listed in Equation (3.3.6), filled circles with error bars. The black solid line shows a shift of ± 0.15 dex in the stellar mass axis of our fiducial GSMF. This systematic is of the same order when confronting SED fitting and colour dependent mass-to-light ratios determinations Robotham et al. (see e.g., 2020) using SED fitting and color with the GAMA sample. Note that these shifts recover most of the differences observed due to systematics in mass-to-light ratios.
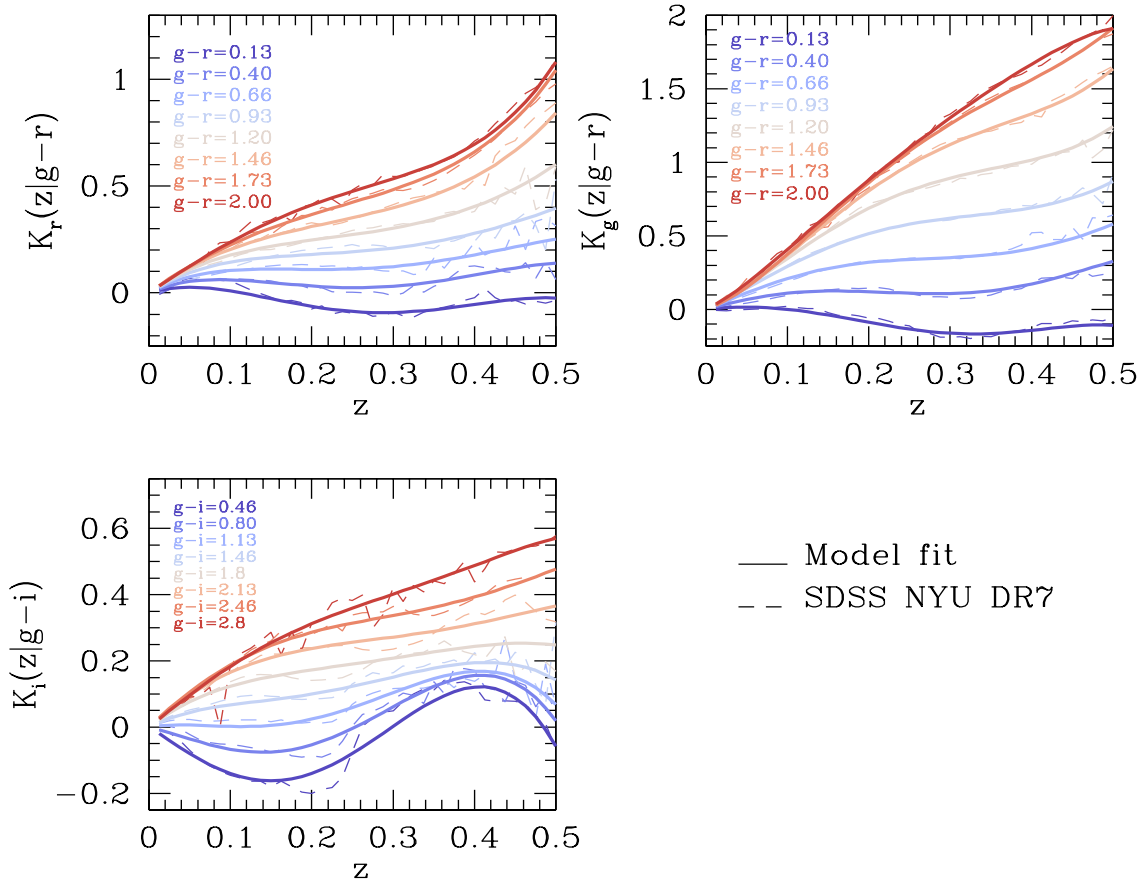
3.3.2   K-corrections

Figure 3.2 shows the colour and redshift dependence of the K-corrections at a rest-frame $z = 0$ for the r (upper left), g (upper right), and i (bottom left) bands, shown as the dashed lines, from the NYU-VAGC SDSS DR7 and calculated from the K-CORRECT algorithm (V4_1_4 Blanton & Roweis, 2007a). In the same figure the solid lines show the best fit according to the following relations:

$$K_j(z|\mathcal{C}) = \mathbf{Z} \mathbf{K}_j^{\mathcal{C}} \mathbf{C} \tag{3.3.8}$$

where j denotes the r, g and i bands while $\mathcal{C}$ denotes the uncorrected $g - r$ and $g - i$ galaxy colours. A similar approach has been done in Chilingarian et al. (2010). The

C and **Z** matrices are respectively given by

$$\mathbf{C} = \begin{pmatrix} 1 \\ C \\ C^2 \\ C^3 \\ C^4 \\ C^5 \end{pmatrix},$$

(3.3.9)

$$\mathbf{Z} = \begin{pmatrix} z & z^2 & z^3 & z^4 & z^5 \end{pmatrix},$$

(3.3.10)

while the $\mathbf{K}_j^{\mathcal{C}}$ matrices for the r, g and i bands are respectively given by equations (3.3.11), (3.3.12) and (3.3.13). Note that our K-corrections are polynomials of degree 5 in both colour and redshift and that in the above set of Equation $K_j(z = 0|\mathcal{C}) = 0$.

$$K_r^{g-r} = \begin{pmatrix} 0.894302 & 2.32866 & -0.787673 & 0.324352 & -0.239774 & 0.0444971 \\ -15.5648 & 1.544 & -2.70992 & 3.42484 & -0.280197 & -0.0221534 \\ 49.7443 & -4.64543 & -8.72852 & 1.14138 & -1.76882 & 0.0702624 \\ -48.9173 & -4.95549 & 2.06966 & 14.5241 & -2.48092 & -0.322153 \\ 3.65716 & 21.3194 & -0.593275 & -6.04982 & -0.157727 & 0.731093 \end{pmatrix} \tag{3.3.11}$$

$$K_g^{g-r} = \begin{pmatrix} 0.0786144 & 4.01535 & -0.883155 & 0.707471 & -2.05303 & 0.793141 \\ -6.81272 & 12.0599 & -10.7157 & 22.086 & -5.46384 & -1.34602 \\ -7.17353 & -52.5682 & -13.5845 & 11.2634 & -6.25812 & 2.61254 \\ 86.1835 & 96.7938 & -72.2792 & -1.44621 & -5.6531 & 9.09575 \\ -106.868 & -23.5461 & 101.815 & -43.5146 & 40.8195 & -21.677 \end{pmatrix} \tag{3.3.12}$$

$$K_i^{g-i} = \begin{pmatrix} -3.01597 & 3.287 & -0.455067 & 0.426496 & -0.242669 & 0.0283777 \\ -1.11123 & -3.04641 & -5.2804 & 2.60911 & 0.134077 & -0.0813698 \\ 68.4078 & -14.6203 & -5.06879 & 0.904234 & -1.82776 & 0.47701 \\ -145.044 & 45.4714 & 8.75605 & 5.9425 & -1.32215 & -0.211679 \\ 59.2903 & -12.387 & -10.8653 & -1.84054 & 0.843326 & -0.0248045 \end{pmatrix} \tag{3.3.13}$$

3.3.3 Galaxy Stellar Mass Function for Low Mass Galaxies

Surface Brightness Correction Completeness

In this thesis we are interested in deriving the GSMF over a wide dynamical mass range, i.e., from dwarf galaxies to massive elliptical galaxies. In Section 3.3.1, we describe that based on the SDSS DR7 galaxy sample we determined the GSMF for galaxies above $M_* \sim 10^9 M_\odot$. In this Section, we determine the GSMF for galaxies above $M_* \sim 10^7 M_\odot$. Deriving the GSMF could be very challenging since the fraction of galaxies of missing galaxies due to surface brightness limits becomes very relevant at the faint-end of the GSMF. Here, we follow a very simple statistical approach in order to quantify the number of galaxies missed due to surface brightness incompleteness limits as described in Blanton et al. (2005c). Our galaxy sample consist of a small volume ($0.0033 < z < 0.05$) carefully constructed to study very low mass/luminosity galaxies from the SDSS NYU-VAGC with a total of 49968 galaxies (Blanton et al., 2005a,c)[5]. Here after we will refer to this galaxy sample as the `low-z` SDSS

Blanton et al. (2005c) estimated that the `low-z` SDSS galaxy sample has a completeness > 70 percent for galaxies in the effective surface brightness range of $18 < \mu_{50,r} < 24$ mag arcsec^{-2} and we consider galaxies only within this range. We assign to each galaxy a weight, $w_{\mu,j}$, which is a function of their central surface brightness and it takes into account the spectroscopic incompleteness ($1/w_{s,j}$), photometric incompleteness ($1/w_{p,j}$), and tiling catalog incompleteness ($1/w_{t,j}$) in the sample. Thus, $w_{\mu,j} = w_{s,j} \times w_{p,j} \times w_{t,j}$. These weights were studied in detail in Blanton et al. (2005c) and provide the correlation between $w_{\mu,j}$ and effective surface brightness, $\mu_{50,r}$, in a tabulated form, see their Table 1. We use cubic spline interpolations of this Table in order to assign a weight $w_{\mu,j}$ to each galaxy in the sample.

The next step in our program is to estimate the number of missed galaxies brighter than $\mu_{50,r} = 24$ mag arcsec^{-2}. To that end, we introduce a model for the distribution of $\mu_{50,r}$ as a function of M_*. We define the fraction of missing galaxies brighter than $\mu_{50,r} = 24$ mag arcsec^{-2} as a function of stellar mass as:

$$f_{\text{loss}}(M_*) = \frac{\sum_j N_{\text{obs}}(\mu_{50,j}|M_*)}{\sum_j N_{\text{real}}(\mu_{50,j}|M_*)}, \qquad (3.3.14)$$

where $N_{\text{real}}(\mu_{50,j}|M_*)$ and $N_{\text{obs}}(\mu_{50,j}|M_*)$ are the real and observed number galaxies with SB between $\mu_{50,j} \pm d\mu_{50,j}$ and stellar masses between $\log M_* \pm d \log M_*/2$, respectively. Thus, our problem reduces to estimating N_{real}. Let us now define $P(\mu_{50,r}|M_*)$

---

[5] `http://sdss.physics.nyu.edu/vagc/lowz.html`
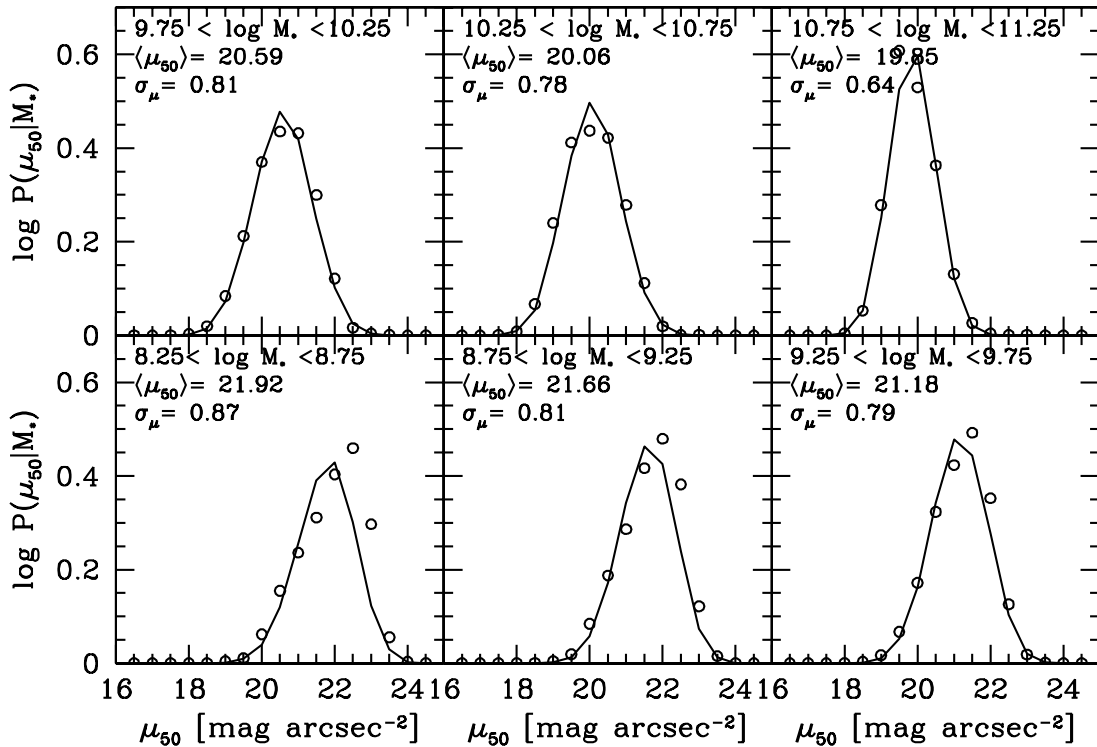
Figure 3.3: Conditional probability distribution $P(\mu_{50,r}|M_*)$ when using stellar mass estimations based on the fits to $g - i$ colors and absolute magnitudes M_i from Zibetti et al. (2009). Empty circles show the resulting distribution from observations of galaxies with Sersic index $n_s \leq 2$ only. Solid lines show the fit to observations when assuming a lognormal model distribution as described in the text.
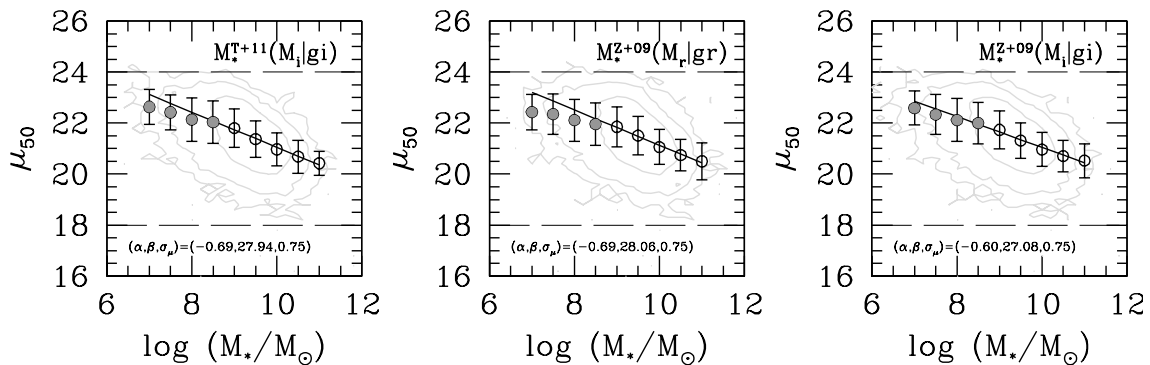
Figure 3.4: SB-to-galaxy stellar mass relation for three different stellar mass estimators used in this thesis. Solid circles show the mean values of μ_{50} that are affected by SB incompleteness while empty circles show that are complete according Blanton et al. (2005c).

84

as the conditional probability distribution of galaxies with SB $\mu_{50,r} \pm d\mu_{50,r}/2$ at a stellar mass bin $\log M_* \pm d \log M_*/2$. We calculate $P(\mu_{50,r}|M_*)$ directly from our galaxy sample by dividing it into stellar mass bins of 0.5 dex. This is done only for galaxies with Sérsic index $n_s \leq 2$ (galaxies with $n_s > 2$ are mostly of high SB, $\mu_{50,r} < 24$ mag arcsec^{-2}, so that it is not necessary to correct them for missing galaxies). For each stellar mass bin, we perform an extra binning of 0.05 dex in SB. We carry out the mentioned binning in M_* and $\mu_{50,r}$ for each of the six different stellar mass estimators described above. As an example, in Figure 3.3 show the distributions $P(\mu_{50,r}|M_*)$ for one of our stellar mass estimators (empty circles). For each M_* bin, we fit $P(\mu_{50,r}|M_*)$ assuming that it is described by a lognormal distribution,

$$P(\mu_{50,r}|M_*) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left[-\frac{(\mu_{50,r} - \langle\mu_{50,r}(M_*)\rangle)^2}{2\sigma_\mu^2} \right], \qquad (3.3.15)$$

where $\langle\mu_{50,r}(M_*)\rangle$ and σ_μ are the mean SB at a given stellar mass and the dispersion around it. We fit these two free parameters for each stellar mass bin. The best fits are plotted with solid lines. This operation is carried out for each of the stellar mass estimators used here. In Figure 3.4, we show an example of the resulting best fits to observations as circles with error bars in the bivariate $(\mu_{50,r}, M_*)$ distribution plane (gray iso-contours) for three of our stellar mass estimators. The dashed lines show our SB magnitudes limits. Similarly to Baldry et al. (2008), we find that the relation between $\langle\mu_{50,r}\rangle$ and $\log M_*$ is linear for galaxies above $M_* \sim 10^9 M_\odot$[6] (filled circles) in Figure 3.4. Departures from this linearity for galaxies below $M_* \sim 10^9 M_\odot$ (filled circles) is an indication that the relation between $\langle\mu_{50,r}\rangle$ and $\log M_*$ is affected by SB incompleteness. We fit the relationship between $\langle\mu_{50,r}\rangle$ and $\log M_*$ for galaxies above $M_* = 10^9 M_\odot$ (where the missing number of low SB galaxies is negligible) as

$$\langle\mu_{50,r}\rangle = \alpha \log M_* + \beta. \qquad (3.3.16)$$

For simplicity, we assume that the dispersion around this relation, σ_μ, is independent of mass and has the same value for all the mass estimators; we assign a value of 0.75 dex, which is close to most of the values determined by fitting Eq. (3.3.15) to the data from our galaxy sample for the three methods of assigning stellar masses. Note that the values of α and β depend on each stellar mass estimator implying that SB corrections are susceptive to systematics due to stellar masses estimators. The next step is to assume that the distribution of real galaxies, N_{real}, can be generated from

---

[6]In fact, Baldry et al. (2008) found that the linearity holds above masses $M_* \sim 10^{8.5} M_\odot$. Here we apply the conservative value of $M_* \sim 10^9 M_\odot$. Nevertheless, we have found that using either Baldry et al. (2008) or our limit, the correction for SB is practically the same.

the probability distribution $P(\mu_{50}|M_*)$ by simply extrapolating equations 3.3.16 and 3.3.15 up to $M_* \sim 10^7$ M$_\odot$. Using the definition of $w_{\mu,j}$ (the SB completeness factor), the observed distribution of galaxies, $N_{\rm obs}$, is thus generated from the probability distribution $P_{\rm obs}(\mu_{50,r}|M_*) = (1/w_\mu) \times P(\mu_{50,r}|M_*)$. The factor of missing galaxies below the SB $\mu_{50} = 24$ mag arcsec^{-2} at a given stellar mass is then

$$f_{\rm loss}(M_*) = \int (1/w_\mu)\, P(\mu_{50,r}|M_*)d\mu_{50,r} \Big/ \int P(\mu_{50,r}|M_*)d\mu_{50,r}. \qquad (3.3.17)$$

Thus, we weight every galaxy in the sample with:

$$w_{\rm SB,j} = w_{\mu,j} \times w_{\rm loss}, \qquad (3.3.18)$$

where

$$w_{\rm loss} = \begin{cases} 1/f_{\rm loss} & {\rm for} \quad n_s \leq 2 \\ 1 & {\rm for} \quad {\rm else.} \end{cases} \qquad (3.3.19)$$

We are now in position to estimate the GSMF corrected by SB incompleteness.

The Dependence of a Stellar Mass Limit Sample with Redshift

In order to calculate the GSMF we start by determining how the apparent magnitude limit of the SDSS transforms into a stellar mass limit. In other words, given that the apparent magnitude limit of the SDSS is $m_{r,\rm lim} = 17.77$ we compute the equivalent in terms of stellar mass, $M_{*,\rm lim}$. Following van den Bosch et al. (2008), we determined the redshift-dependent absolute magnitude limit $M_{r,\rm lim}^{0.0}$ given the apparent magnitude limit from the SDSS $m_{r,\rm lim} = 17.77$

$$M_{r,\rm lim}^{0.0} = m_{r,\rm lim} - 5 \log D_L(z) - 25 - K_{\langle gr \rangle}(z) + E_r(z), \qquad (3.3.20)$$

where D_L, $K_{\langle gr \rangle}$ and E_r are functions described in Section 3.3.1. Note that we have emphasised the use of average colours for the $K-$correction because we are interested in the stellar mass limit for all the galaxies. Thus, the above absolute magnitude limit depends both on redshift and colour (van den Bosch et al., 2008). Using the colour-dependent mass-to-light ratio $\Upsilon_r^{\rm Z09}(g-r)$ from Zibetti et al. (2009), we transform $M_{r,\rm lim}^{0.0}$ into a stellar mass limit

$$M_{*,\rm lim} = -0.84 + 1.654 \times \langle g - r \rangle^{0.0} - 0.4 \times \left(M_{r,\rm lim}^{0.0} - 4.64 \right). \qquad (3.3.21)$$

Finally, we use the mean relationship between colour and stellar mass for blue and red galaxies as well as the fraction of red, $f_{\rm R}$, and blue galaxies, $f_{\rm B}$, to compute the

average colour-stellar mass relationship as

$$\langle g - r \rangle^{0.0} = f_{\mathrm{B}}(g - r)_{\mathrm{B}}^{0.0} + f_{\mathrm{Q}}(g - r)_{\mathrm{Q}}^{0.0}, \tag{3.3.22}$$

where $(g - r)_{\mathrm{B}}^{0.0}$ and $(g - r)_{\mathrm{Q}}^{0.0}$ are the best fit models to the mean colour-stellar mass relationships of blue and red galaxies.

We paused here for a moment and described our method to derive $(g - r)_{\mathrm{B}}^{0.0}$ and $(g - r)_{\mathrm{Q}}^{0.0}$. To do so, we use the SDSS DR7 based on the photometric catalogue from Meert et al. (2016). We choose to use this catalog as contains many more galaxies than the SDSS DR4 and one could derive robust colour distributions. We derived the observed distribution function of galaxy colours as a function of stellar mass, $P_{gr}(g - r|M_*)$, that is the observed distribution of galaxy colours at the range between $(g-r)^{0.0} \pm \Delta(g-r)^{0.0}/2$ and $\log M_* \pm \Delta \log M_*/2$. We divide our space into 20 bins equally spaced for $(g - r)^{0.0}$ between $(g - r)^{0.0} = 0$ and $(g - r)^{0.0} = 1.4$ and into 25 bins equally spaced between $\log M_* = 8.5 - 12$. For galaxy stellar masses we use our fiducial definition from Section 3.3.1.

We assume that the distribution $P_{gr}(g - r|M_*)$ is bimodal and composed of two Gaussian distributions, this is a good approximation as shown by previous studies (e.g., Baldry et al., 2004, 2006). We associate one of the modes of $P_{gr}(g - r|M_*)$ with the distribution of blue galaxies, denoted by $P_{gr,\mathrm{B}}(g - r|M_*)$, while the remaining one with the distribution of red galaxies, denoted by $P_{gr,\mathrm{R}}(g - r|M_*)$. The relation between these distributions is given by:

$$\begin{aligned} P_{gr,\mathrm{B}}(g - r|M_*) =& f_{\mathrm{B}}(M_*)P_{gr,\mathrm{B}}(g - r|M_*) + \\ & f_{\mathrm{R}}(M_*)P_{gr,\mathrm{R}}(g - r|M_*). \end{aligned} \tag{3.3.23}$$

We assume that $P_{gr,j}(g - r|M_*)$, with $j = $ B or R, is a Gaussian distribution given by:

$$\begin{aligned} P_{gr,j}(g - r|M_*) =& \frac{1}{\sqrt{2\pi\sigma_j^2(M_*)}} \times \\ & \exp\left[-\frac{\left((g - r)^{0.0} - (g - r)_j^{0.0}(M_*)\right)^2}{2\sigma_j^2(M_*)} \right], \end{aligned} \tag{3.3.24}$$

where $(g - r)_j^{0.0}(M_*)$, with $j = $ B or R, is the mean colour-stellar mass relationship used in Equation (3.3.22) and $\sigma_j(M_*)$ is the standard deviation that depends on M_*.

The functional forms for $(g-r)_j^{0.0}(M_*)$ and $\sigma_j(M_*)$ are given by

$$(g-r)_j^{0.0}(M_*) = \alpha_j + \beta_j \times \log\left(\frac{M_*}{10^{11}M_\odot}\right), \qquad (3.3.25)$$

and

$$\sigma_j(M_*) = \lambda_j + \kappa_j \times \log(M_*), \qquad (3.3.26)$$

for $j = $ B, R. Finally, for the fraction of red galaxies we assume that

$$f_R(M_*) = \frac{1}{1 + [a + b\,(M_*/M_C)]^\gamma}. \qquad (3.3.27)$$

We performed a χ^2 minimisation procedure to the observed galaxy colour bimodality in order to find the best fitting parameters to the functional forms described above. Our best fitting parameters are: $(\alpha_B, \beta_B, \lambda_B, \kappa_B, \alpha_R, \beta_R, \lambda_R, \kappa_R, a, b, \log M_C, \gamma) = (0.514,$ 0.086, 0.240, -0.015, 0.720, 0.064, -0.068, 0.014, 0.001, 1.390, 10.586, -1.001).

The upper panel of Figure 3.5 shows the dependence of $M_{*,\text{lim}}$ with redshift. The region above $M_{*,\text{lim}}$ is the area above which the NYU-VAGC galaxy sample is a volume-limited sample that is complete in stellar mass. The small grey dots show individual galaxies from the NYU-VAGC sample in the case that stellar masses were determine using the geometric mean of all our stellar mass estimators.

3.3.4 Volume and large scale structure corrections

In a volume-limited sample that is complete in stellar mass, we can derive the GSMF as the number of observed galaxies, N_{gals}, per unit of comoving volume V with stellar masses between $\log M_* \pm \Delta \log M_*/2$, that is: $\phi_*(M_*)\Delta \log M_* = N_{\text{gals}}/V$. Once we determined the dependence of the stellar mass limit with redshift, $M_{*,\text{lim}}$, from the SDSS NYU VAGC sample, we can use the above idea by defining various volume-limited subsamples that are complete in stellar mass. These subsamples were defined by dividing the galaxy redshift range covered by the NYU VAGC, $0.0033 < z < 0.005$, into 20 bins. Therefore, the GSMF for the jth volume-limited subsample at the redshift range $z \pm \Delta z/2$ that is complete in stellar mass can be estimated for the mass bin $\log M_* \pm \Delta \log M_*/2$ as

$$\phi_j(M_*, z)\Delta \log M_* = \frac{N_{\text{gals},j}(M_*, z)}{V(z - \Delta z/2) - V(z + \Delta z/2)}. \qquad (3.3.28)$$

We tested the above methodology with realistic mock galaxy catalogues. To do so, we use the N-body Bolshoi-Planck simulation (Klypin et al., 2016), and halo
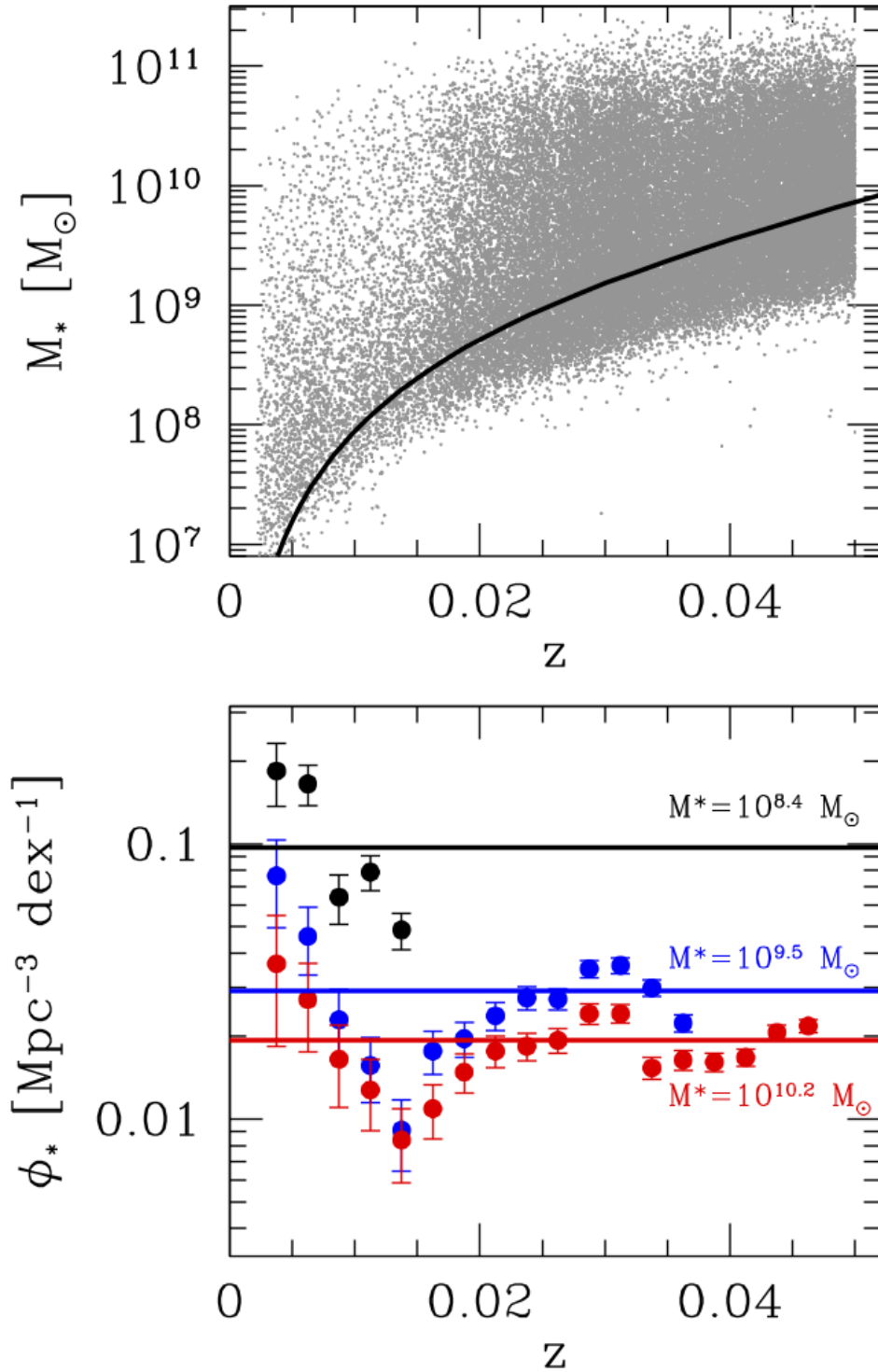
Figure 3.5: *Upper Panel:* Distribution of galaxies in the M_* and redshift plane for the lowz-SDSS galaxy sample, grey dots. The solid lines shows the dependence of the stellar mass completeness limit as a function of redshift for our galaxy sample. *Bottom Panel:* The dependence of the GSMF with redshift for three different stellar masses. Note that the increase and decreases in the amplitudes is due the large scale structure fluctuations. The solid lines show our corrections due to large scales structures as described in the text.

catalogues described in (Rodríguez-Puebla et al., 2016). We use the semi-emperical modelling from (Rodríguez-Puebla et al., 2017) in order to assign galaxies to dark matter halos/subhalos. The galaxies in the catalogue were projected into the redshift space through a lightcone. We use the dependence of the stellar mass limit with redshift described in Section 3.3.3 and include galaxies only within the same redshift range as the NYU VAGC in order to reproduce the observed distribution of galaxies in the M_* and redshift plane for the lowz-SDSS sample. Our results show that the above methodology recovered the original GSMF with differences not larger than $\sim 5\%$. In addition, we have calculated the GSMF using the Stepwise Maximum Likelihood method (Efstathiou et al., 1988) and found very similar similar results (not shown) as those reported based on our methodology.

Finally, we calculate the GSMF corrected by surface brightness incompleteness by

$$\phi_{\mathrm{SB},j}(M_*, z)\Delta \log M_* = \frac{N_{\mathrm{SB,gals},j}(M_*, z)}{V(z - \Delta z/2) - V(z + \Delta z/2)}, \qquad (3.3.29)$$

where

$$N_{\mathrm{SB,gals},j} = \sum_{i=1}^{N_{\mathrm{gals},j}} w_{\mathrm{SB},i}, \qquad (3.3.30)$$

and $w_{\mathrm{SB},i}$ is our SB incompleteness correction given by Equation (3.3.18).

The bottom panel of Figure 3.5 shows $\phi_{\mathrm{SB},j}(M_*, z_j)$ for three different stellar masses, $M_* = 10^{8.4}\mathrm{M}_\odot$, $10^{9.5}\mathrm{M}_\odot$ and $10^{10.2}\mathrm{M}_\odot$. Fluctuations in the amplitude of the GSMF shows that the distributions of galaxies is not uniform across the redshift distribution because of environmental effects arising from large scale structures. In order to minimise the above effect, we compute the weighted mean of the GSMF. In other words, we derive the total GSMF as

$$\langle \phi_{\mathrm{SB}}(M_*)\rangle = \sum_{j=1}^{N=20} \phi_{\mathrm{SB},j}(M_*, z_j) \times w_j \qquad (3.3.31)$$

where $w_j = N_{\mathrm{gals},j}/\sum_j N_{\mathrm{gals},j}$ is the fraction of galaxies at the jth volume-limited subsample centred at the redshift bin $z \pm \Delta z/2$ for the mass bin $\log M_* \pm \Delta \log M_*/2$. The solid line in Figure 3.5 shows the resulting value of $\langle \phi_{\mathrm{SB}}\rangle$ for the masses discussed above.

Figure 3.6 compares the resulting GSMFs when SB corrections are applied $\phi_{\mathrm{SB},*}$ (red filled circles) and when we ignore SB corrections ϕ_* (blue filled circles) for each of the six stellar mass definitions used here. As expected, the SB correction increases the number density of low-mass galaxies. For higher masses than $\sim 3 \times 10^9$ $\mathrm{M}_\odot$, this correction is negligible. For comparison, we reproduce with grey filled circles the

GSMF reported in Baldry et al. (2008), who used also the `low-z` NYU-VAGC sample but for the DR4 as well as the Baldry et al. (2012) from the GAMA survey with the skeletal symbols. In none of them SB corrections were applied. Finally, in all the panels of Figure 3.6 we reproduce the GSMF from the main SDSS DR7 derived in Section 3.3.1. Observe how the GSMFs constructed from the `low-z` NYU-VAGC sample and the ones constructed from the main SDSS DR7 samples match extremely well at $M_* \sim 10^9$ $M_\odot$, but the latter overcomes the former at high stellar masses due to the larger volume covered by the SDSS DR7.
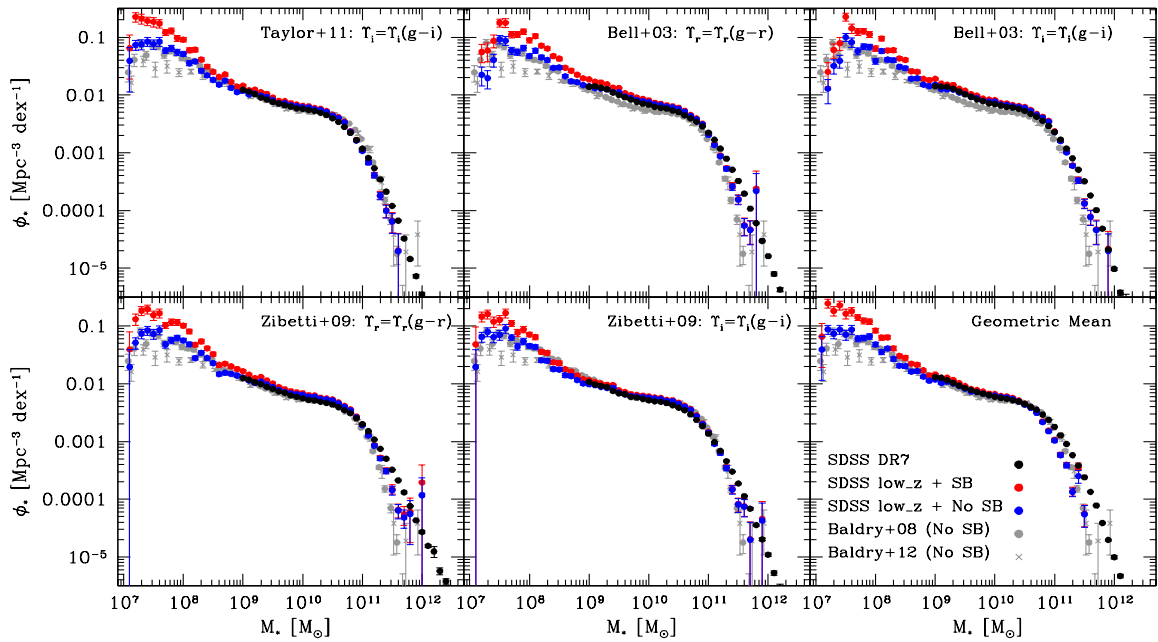
Figure 3.6: Corrected GSMF for surface brightness incompleteness, red filled circles with error bars. Blue filled circles with error bars show the uncorrected GSMF. Grey skeletal symbols and filled circles show the GSMF from Baldry et al. (2012) and Baldry et al. (2008), respectively. The black filled circles show the GSMF from the SDSS DR7 derived from Section 3.3.1. The different panels show the various stellar mass estimators used for this thesis. Note that the at $M_* \sim 10^9$ M$_\odot$ there is a smooth transition between the low-z NYU-VAGC sample and the SDSS DR7 GSMFs.

Finally, we briefly describe our final GSMF. For galaxies below $M_* = 10^9$M$_\odot$, we use the GSMF derived from the `low-z` NYU-VAGC sample, while for galaxies above $M_* = 10^9$M$_\odot$, we use the GSMF from the SDSS DR7 based on the photometric catalog from Meert et al. (2015). We apply a simply correction in our GSMF for passing from the Willick et al. (1997) distance flow model to the Tonry et al. (2000) one. Figure 12 from Baldry et al. (2012) shows that after adjusting the Baldry et al. (2008) GSMF to the Tonry et al. (2000) distances both MFs are in excellent agreement. With that information, we first note that our fiducial (uncorrected) GSMF (bottom right panel from Figure 3.5) is very similar to the Baldry et al. (2008) GSMF, and thus we assume that the impact of correcting by Tonry et al. (2000) distances is equivalent to rescale it to the Baldry et al. (2012) GSMF. Based on the above, we rescale our SB-corrected GSMF as $\phi_{\rm SB,T00} = \langle\phi_{\rm SB}\rangle \times \phi_{\rm B08}/\phi_{\rm B12}$. Recall that our fiducial GSMF uses stellar masses from the geometric mean of all stellar masses described by Equation (3.3.6).

Figure 3.7 shows our final GSMF, $\phi_{\rm SB,T00}$, as the black filled circles with error bars. The filled grey symbols show the GSMF, $\langle\phi_{\rm SB}\rangle$, in which the Willick et al. (1997) model flow is utilised. We also compare to Baldry et al. (2012) and Wright et al. (2017) determinations. Note that after distance and SB corrections our fiducial GSMF is in good agreement with the observed low-mass end slope of the GAMA
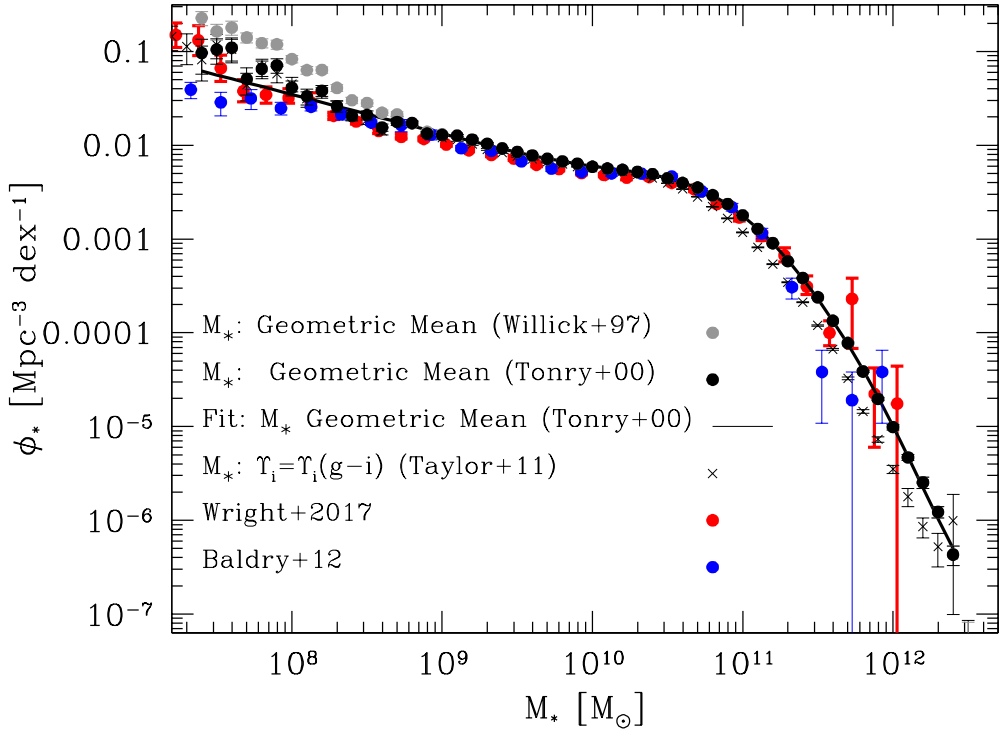
Figure 3.7: Corrected GSMF for SB and for the flow model, filled circles with error bars. This is our fiducial GSMF. The corresponding best fit model is shown with the solid line, see Section 3.5. Grey filled circles shows when using the flow model from Willick et al. (1997). We also show our results when using the mass-to-light ration from Taylor et al. (2011) and compared to Baldry et al. (2012) who used the same mass estimator. Notice that both mass functions are consistent between each other. For completeness we compared to Wright et al. (2017).

survey. For comparison we present our corrected GSMF but when using the Taylor et al. (2011) mass-to-light ratios. Note that in this case our GSMF is consistent with the Baldry et al. (2012) GSMF.

3.4 The GSMF of All Galaxies

Our final GSMF is the result of combining the SDSS NUY-VAGC `low-z` sample, for galaxies with masses $M_* \approx 3 \times 10^7 M_\odot$ to $M_* \approx 10^9 M_\odot$, and the SDSS DR7 sample for galaxies with $M_* \gtrsim 10^9 M_\odot$, based on our fiducial M_* determination. Figure 3.8 presents our final GSMF with the black solid circles and error bars. The black solid line shows the best fit to the data (described below), and the grey shaded area shows a shift in the M_* axis of ± 0.15 dex. As discussed above, in Section 3.3.1 we find that this is a good approximation to the systematic errors in the GSMF due to differences in the mass-to-light ratios. We also note that differences in photometry could introduce a systematic shift of 0.04 between them. This is smaller than the systematics based on different stellar mass estimators from broad band colours (Robotham et al., 2020) In the same figure, we include comparisons to previous works. In order to account
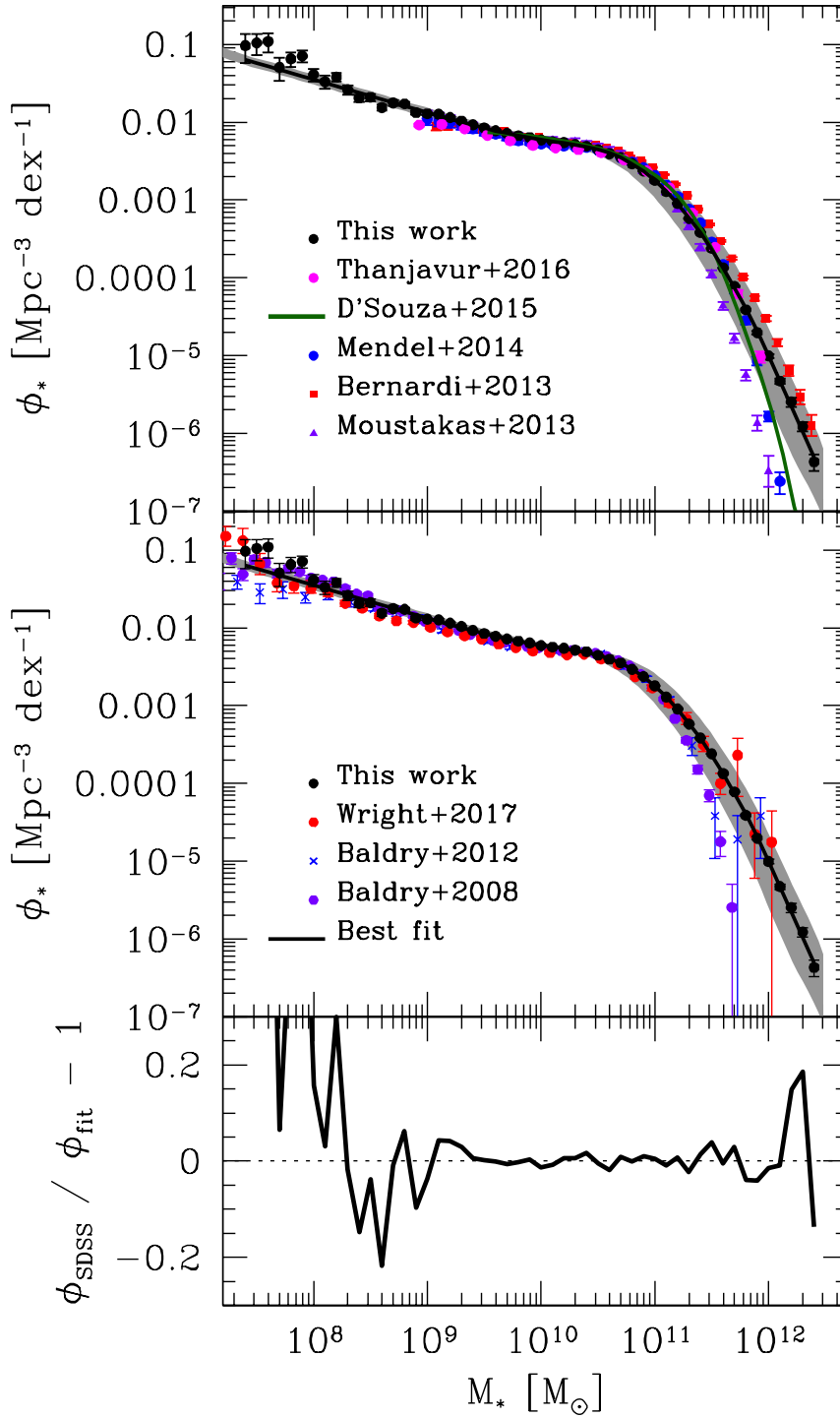
Figure 3.8: Observed GSMF when combining the SDSS NYU-VAGC low-redshift sample and the SDSS DR7 sample, black filled circles with error bars. We reproduce our results in the upper and the middle panels. The best fit model composed of a Schechter function with a sub-exponential slope and a double power-law function is shown as the black solid line. The shaded area shows an estimate of the systematic errors with respect to the best fitting model. The bottom panel shows the residuals for our best fitting model as a function of M_*. We include comparisons to some previous observational determinations of the GSMF: in the upper panel we show determinations that are complete down to $\sim 10^9$ M$_\odot$, mostly based on the SDSS DR7, while in the middle panel we show determinations based on the GAMA survey, which are complete down to $\sim 3-5 \times 10^7$ M$_\odot$, but suffer from cosmic variance at high masses due to the small volume.

for differences in cosmologies, we scale previous studies to our cosmology using the following relations:

$$\phi_{*,\mathrm{us}} = \phi_{*,\mathrm{lit}} \left(\frac{h_{\mathrm{us}}}{h_{\mathrm{lit}}} \right)^3 , \qquad (3.4.1)$$

and

$$M_{*,\mathrm{us}} = M_{*,\mathrm{lit}} \left(\frac{h_{\mathrm{lit}}}{h_{\mathrm{us}}} \right)^2 , \qquad (3.4.2)$$

where $h_{\mathrm{us}} = 0.678$ and h_{lit} is the respective value reported in the literature. Nonetheless, the impact of accounting for different cosmologies is small.

In the upper panel of Figure 3.8, we reproduce the GSMFs from previous determinations with stellar mass completeness above $M_* \sim 10^9$ M$_\odot$. The violet triangles with error bars are the determinations from Moustakas et al. (2013), who used a spectroscopic sample of SDSS DR7 galaxies from the NYU-VAGC with redshifts $0.01 < z < 0.2$ combined with observations from GALEX. The red squares with error bars are the estimation obtained in Bernardi et al. (2013) from a sample of SDSS DR7 galaxies with photometry based on the PyMorph software pipeline at $z \sim 0.1$. Here we reproduce their result based on Sérsic luminosities. Additionally, we compute the GSMF using the stellar mass estimates from Sérsic photometry from Mendel et al. (2014) who used the Simard et al. (2011) SDSS DR7 sample of g and r band photometry and extended to u, i and z bands, blue filled circles with error bars. We show the best fitting model from D'Souza et al. (2015), who estimated the GSMF by stacking images of galaxies with similar stellar masses and concentrations to correct Model magnitudes from the SDSS DR7, dark green solid line. Finally, we compare our result to Thanjavur et al. (2016), who derived the GSMF using the analysis from Mendel et al. (2014).

Our GSMF agrees well with previous determinations at the $\sim 10^{9.3} - 10^{11}$ M$_\odot$ range. At the high mass end, it is shallower than previous determinations (e.g., Moustakas et al., 2013) except to Bernardi et al. (2013), who use Sérsic photometry from the SDSS DR7. As extensively discussed in Bernardi et al. (2017), there are two systematic effects that could lead to differences when comparing to previous determinations from the literature; assumptions on mass-to-light ratios and estimations of galaxy surface brightness. In the case of Moustakas et al. (2013) and D'Souza et al. (2015), who used cmodel and Model magnitudes, the comparison is not obvious due to systematic effects in both mass-to-light ratios and photometry (Bernardi et al., 2017). In the case of Mendel et al. (2014) and Bernardi et al. (2013), effects on photometry are not the dominant ones but mass-to-light ratios. Nonetheless, those differences are within the expected systematic effect, especially at the massive-end, (Bernardi et al., 2017, see also Figure 3.1). We therefore conclude that when compar-

ing to other previous determinations, the differences that we observe are consistent with the differences expected from systematic effects. Indeed, Figure 3.8 shows that most of the previous determinations are within our region of systematic errors. Thus, hereafter we will assume that our shift of ± 0.15 dex in the M_* axis approximately captures systematics not only from stellar population models but also from photometry.

The middle panel of Figure 3.8 presents comparisons to some previous determinations from deep but small-volume samples. The purple dots with error bars are from Baldry et al. (2008), who used the SDSS NYU-VAGC `low-z` sample but did not include missing galaxies due to surface brightness incompleteness. In addition, we compare to Baldry et al. (2012), who used the GAMA survey for galaxies at $z < 0.06$, and complete down to $r = 19.4$ mag for two thirds of the galaxy sample and to $r = 19.8$ for one third of the sample. Finally, we reproduce the observed GSMF from Wright et al. (2017), who also used the GAMA survey to estimate the GSMF.

At low masses our results are in excellent agreement with the GAMA GSMFs. This is encouraging since the GAMA survey does not suffer from surface brightness incompleteness, at least within the stellar mass range that we are comparing our results. This is an indication that the surface brightness corrections described in Section 3.3.3 are able to recover the slope of the GSMF at low masses. Consistent with the values reported in Baldry et al. (2012) and Wright et al. (2017), we find that the faint-end slope of the GSMF is $\alpha \approx -1.4$, below we describe in more detail the fitting model for the GSMF. The above is also in good agreement with Sedgwick et al. (2019) who recently determined the low mass-end of the GSMF by identifying low surface brightness galaxies based on data of core-collapse supernovae. The authors used the IAC Stripe 82 legacy project (Fliri & Trujillo, 2016) and the SDSS-II Supernovae Survey (Frieman et al., 2008).

At the massive end we notice, however, some apparent tension between our and the GAMA results. Effects due to cosmic variance (due to the small redshift and angular coverage of the GAMA sample) could explain those differences as well as systematics in the mass-to-light ratios. Indeed, we see that some of the data are within the systematic errors. In addition, note that Figure 3.7 from Section 3.3.3 shows that using the mass-to-light ratios from Taylor et al. (2011), utilised in the Baldry et al. (2012) GSMF, tend to underestimate the high-mass end of the GSMF.

Table 3.1: Best fitting parameters for the GSMF (Eqs. 3.5.1-3.5.3)

$\log \phi_S^*$ $\left[\mathrm{Mpc}^{-3}\mathrm{dex}^{-1}\right]$	α_S	β	$\mathcal{M}_D = \mathcal{M}_S$ $[\mathrm{M}_\odot]$
-3.019 ± 0.067	-1.418 ± 0.025	0.660 ± 0.011	10.897 ± 0.036

$\log \phi_D^*$ $\left[\mathrm{Mpc}^{-3}\mathrm{dex}^{-1}\right]$	α_D	δ	γ
-2.267 ± 0.120	-0.207 ± 0.169	3.660 ± 0.347	1.236 ± 0.080

3.5 Best Fitting Model to the GSMF

To provide an analytic form to our GSMF we choose to use a function composed of a Schechter function with a sub-exponential decreasing slope and a double power-law function. Note that the resulting high-mass end of our GSMF is shallower than an exponential function, and, thus, better fitted to a power-law (see also Tempel et al., 2014). The Schechter sub-exponential function is given by:

$$\phi_{*,\mathrm{S}}(M_*) = \phi_S^* \ln 10 \left(\frac{M_*}{\mathcal{M}_S}\right)^{1+\alpha_S} \exp\left[-\left(\frac{M_*}{\mathcal{M}_S}\right)^\beta\right], \qquad (3.5.1)$$

where ϕ_S^* is the normalisation parameter in units of Mpc^{-3} dex^{-1}, α is the slope at the low-mass end, $\mathcal{M}_S$ is the characteristic mass, and β is the parameter that controls the slope at the massive end; note that $\beta = 1$ corresponds to a Schechter function. The double power-law function is given by:

$$\phi_{*,\mathrm{D}}(M_*) = \phi_D^* \ln 10 \left(\frac{M_*}{\mathcal{M}_D}\right)^{1+\alpha_D} \left[1 + \left(\frac{M_*}{\mathcal{M}_D}\right)^\gamma\right]^{\frac{\delta-\alpha_D}{\gamma}}, \qquad (3.5.2)$$

where ϕ_D^* is the normalization parameter in units of Mpc^{-3} dex^{-1}, α and δ control the slope at low and high masses, respectively, while γ determines the speed of the transition between the low and high mass regimes; and $\mathcal{M}_D$ is the characteristic mass of the transition. Finally, the analytic form for fitting the observed GSMF is given by

$$\phi_{*,\mathrm{model}}(M_*) = \phi_{*,\mathrm{S}}(M_*) + \phi_{*,\mathrm{D}}(M_*), \qquad (3.5.3)$$

where we assumed that $\mathcal{M}_S = \mathcal{M}_D$.

We find the best fit parameters $\vec{p}_{\mathrm{GSMF}} = (\phi_S^*, \alpha_S, \mathcal{M}_S, \beta, \phi_D^*, \alpha_D, \delta, \gamma)$, that maximize the likelihood function $\mathcal{L} \propto \exp(-\chi^2/2)$ by using the Markov chain Monte Carlo (MCMC) method algorithm described in Rodríguez-Puebla et al. (2013). Here

$$\chi^2 = \sum_{i=1}^{N_{\mathrm{obs}}} \left(\frac{\phi_{*,\mathrm{SDSS}}^i - \phi_{*,\mathrm{model}}^i}{\sigma_{\mathrm{SDSS}}^i}\right)^2, \qquad (3.5.4)$$

with $N_{\rm obs}$ as the number of observational data points of the GSMF each with an ith value of $\phi^i_{*,\rm SDSS}$ and an error of $\sigma^i_{\rm SDSS}$. The ith value of our model is given by $\phi^i_{*,\rm model}$.

We sample the best-fit parameters by running a set of ten chains with 1×10^5 MCMC models each. Table 3.1 lists the best fit parameters. For our best fitting model we find that $\chi^2 = 85.42$ from a number of $N_{\rm obs} = 50$ observational data points. Our model consist of $N_{\rm p} = 8$ free parameters, thus the reduced χ^2 is $\chi^2/{\rm d.o.f.} = 2.03$. The upper and middle panels of Figure 3.8 show our best fitting model as the black solid line and the bottom panel shows the residuals as a function of M_*. Our best fitting model has an error of $\sim 2\%$ at the range $M_* \sim 2 \times 10^9 - 5 \times 10^{11} \rm M_\odot$ and an error lower than $\sim 10\%$ at the mass range $M_* \sim 7 \times 10^8 - 1 \times 10^{12} \rm M_\odot$. For lower masses errors can be up to $\sim 20\%$.

A valid question is how much we improve the analytic prescription when using a Schechter sub-exponential plus a double power-law function model confronted to a double Schechter function model, commonly employed by previous authors (see e.g., Baldry et al., 2012; Wright et al., 2017). We have explored this possibility but assuming a Schechter function, $\beta = 1$ in Equation (3.5.1), and a Schechter sub-exponential function, that is, we are adding a extra degree of freedom due to the shallow decay at the high mass-end. Based on this alternative, we repeat our fitting procedure but this time finding that $\chi^2 = 662.817$ from a number of $N_{\rm obs} = 50$ observational data points. Now, our model consist of $N_{\rm p} = 6$ free parameters resulting in a reduced χ^2 of $\chi^2/{\rm d.o.f.} = 15.06$. This is worse than when combining Schechter sub-exponential and double power-law functions. Thus, hereafter we will consider only the Schechter sub-exponential model.

3.6 The GSMFs of Early- and Late-Type Galaxies

Our main goal for this thesis is to construct bivariate distributions as well as mass function based on the observed gas mass conditional probability distribution functions and the GSMFs of early- and late-type galaxies. In this section, we determine the GSMF of early- and late-type galaxies from the SDSS DR7 spectroscopic sample with the public automated morphological galaxy classification by Huertas-Company et al. (2011).[7] The morphological classification in Huertas-Company et al. (2011) was determined based on support vector machine algorithms. Here we use their tabulated probabilities for each SDSS galaxy as being classified as an early type, $P(E)$. For masses below the completeness of the SDSS DR7 sample, we use an extrapolation of the observed fraction of early type galaxies. We will come back to this point later in

---

[7]http://gepicom04.obspm.fr/sdss_morphology/Morphology_2010.html
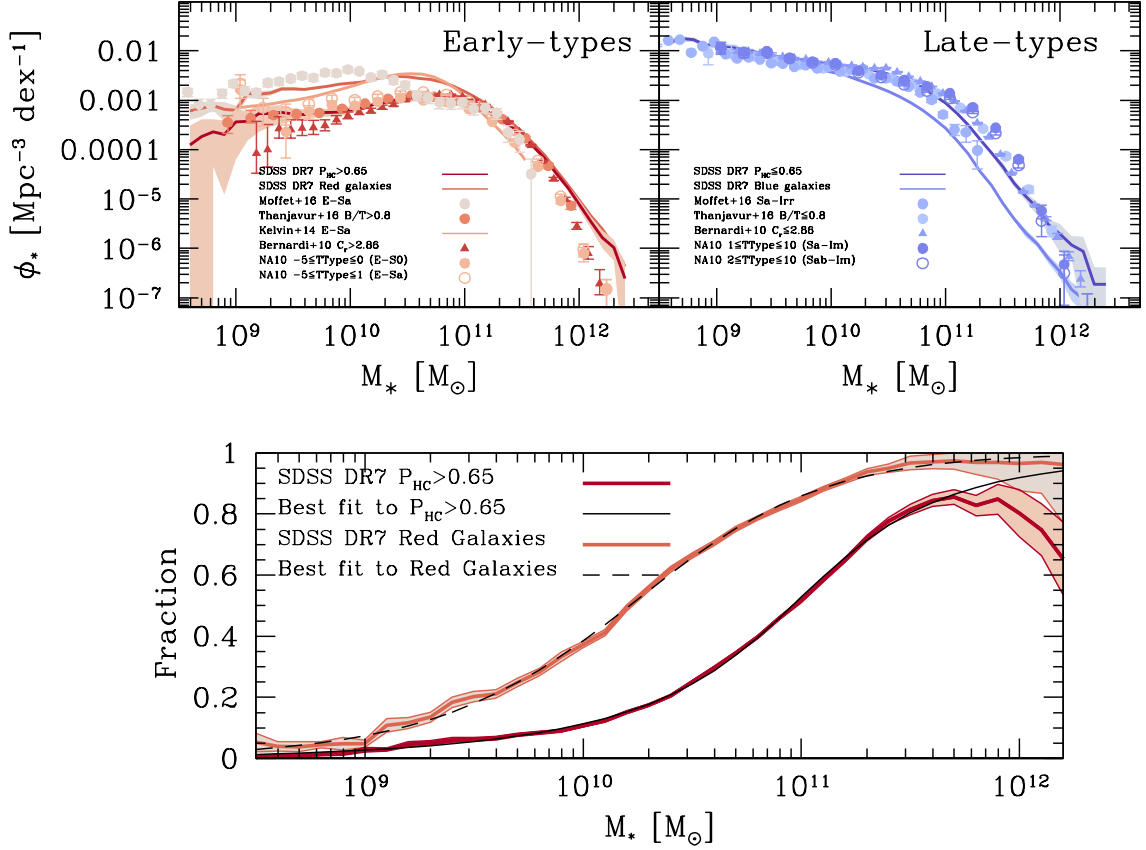
Figure 3.9: SDSS DR7 GSMFs for early- and late-type galaxies, left and right upper panels, respectively. Early- (late-)type galaxies are defined as those with $P(E) > 0.65$ ($P(E) \leq 0.65$) from the tabulated probabilities of Huertas-Company et al. (2011). This is equivalent to morphological types that comprises E and S0 galaxies or $T \leq 0$ (Sa to Irr galaxies or $T > 0$). We compare to various previous determinations from the literature as indicated by the legends, see also the text for details. Our determinations are in general in good agreement with previous determinations from SDSS spectroscopic samples, while a tension is evident with determinations of Moffet et al. (2016) for the GAMA survey. We also present our resulting GSMFs for blue and red galaxies. These GSMFs follow closely those by morphology from the GAMA survey. The bottom panel shows our number density-weighted fractions of early-type and red galaxies as a function of M_*. Their corresponding best fit models (Eq. E.0.1) are shown with solid and dashed lines, respectively.

this section.

From a catalog of galaxies with visual morphological classification (UNAM-KIAS; Hernández-Toledo et al., 2010) we find that galaxies with types $T \leq 0$ are mostly those with $P(E) > 0.65$, and those with $P(E) \leq 0.65$ correspond mostly to $T > 0$; here T is the Fukugita et al. (2007) notation.[8] Based on the above, we consider as early-type galaxies those with a probability $P(E) > 0.65$ while late-type galaxies those with $P(E) \leq 0.65$. We checked that our morphology definition between early- and late-type galaxies is consistent with the morphological classification based on the concentration parameter $c = R_{90}/R_{50}$. That is, the division between early- and late-types is approximately at $c = 2.85$ (see below and also, Hyde & Bernardi, 2009; Bernardi et al., 2010).

We calculate the SDSS DR7 GSMF of early- and late-type galaxies using the $1/V_{\mathrm{max}}$ method described in Section 3.3.1. Figure 3.9 shows the corresponding GSMFs of early and late types in the upper left and right panels, respectively. For comparison we show the GSMFs for red and blue galaxies based on a color cut limit in the $(g - r)^{0.0} - M_*$ diagram. In this diagram, we find that a rough division criteria from blue to red galaxies is given by the color limit of $(g - r)^{0.0} = 0.66$.[9] In the same figure, we compare our results to different determinations from the literature as we describe below. All the data in this figure have been renormalised to our cosmology.

Recently, Moffett et al. (2016a) visually classified morphologies in the GAMA survey, and reported the GSMF for different morphologies. Here we reproduce their GSMF from E to Sa galaxies as early types, and the complement as late types. Contrary to our definition, Sa galaxies are included in the early-type group; this is because the authors report S0 and Sa galaxies as one morphology group. As shown in Figure 3.9, the GSMF of early-type galaxies from Moffett et al. (2016a) results in an overabundance of low-mass galaxies compared to other studies. We reproduce the results from Thanjavur et al. (2016) with bulge-to-total ratios of $B/T > 0.8$ as early types, and $B/T \leq 0.8$ as late types. Thanjavur et al. (2016) used the bulge-to-disc decomposition from Simard et al. (2011) SDSS DR7 spectroscopic sample, and stellar masses derived from Mendel et al. (2014). We also include results from Kelvin et al.

---

[8]Huertas-Company et al. (2011) define as elliptical galaxies those objects with $T \leq 0$, S0s as $T = 1$, Sabs as $2 < T < 4$ and, Scd as $4 \leq T < 7$ based on the Fukugita et al. (2007) morphology classification. Huertas-Company et al. (2011) included elliptical galaxies and S0s as early-type galaxies which corresponds to galaxies with types $T \leq 1$ in the Fukugita et al. (2007) notation, and $T \leq 0$ when using the Nair & Abraham (2010a) notation, see below. In the de Vacouleurs notation this is equivalently to $T = 0$.

[9]While this is just a rough division line, we used it as a practical method for decomposing the GSMF into two main groups. Notice that in Section 3.3.3 we apply a more sophisticated method to derive the distribution of blue and red galaxies. Additionally, we checked that both methods give similar results.

(2014). Similarly to Moffett et al. (2016a), Kelvin et al. (2014) visually classified morphologies in the GAMA survey. We again use their GSMF from E to Sa galaxies for early types since the authors combined S0-Sa galaxies as in Moffett et al. (2016a). The filled triangles with error bars show the GSMF from Bernardi et al. (2010) for galaxies with concentration parameter $c > 2.86$ for early types, and $c \leq 2.86$ for late types.[10] Finally, using the Nair & Abraham (2010a) morphology catalog, who visually classified 14,034 spectroscopic galaxies from the SDSS DR4, we derive the GSMF for early-type galaxies.[11] We utilise their morphological notation and define early-type galaxies as those objects with $-5 \leq T \leq 0$ (E-S0s), equivalent to $T \leq 1$ in the Fukugita et al. (2007) notation. We additionally derive the GSMF with morphologies between $-5 \leq T \leq 1$ in the Nair & Abraham (2010a) notation, which include Sa galaxies.

In general our results agree with previous determinations, especially with those from the SDSS spectroscopic samples. In contrast, the GSMF of early-type galaxies from the visual classification of the GAMA survey are systematically above our results at the low-mass end, $M_* \lesssim 2 \times 10^{10} \mathrm{M}_\odot$, but closer to our classification based on galaxy colour. While it is not clear the reason of the differences outlined above (the inclusion of Sa galaxies as early-types, environment, etc.), in Chapter 4 we will discuss the impact of using galaxy colour instead of morphology when deriving the HI, H_2, cold gas, and baryonic MFs separated into two main galaxy populations.

Finally, the bottom panel of Figure 3.9 shows the resulting fraction of early-type galaxies as a function of stellar mass, $f_E(M_*)$. In addition, we show the fraction of red galaxies when using our $g - r$ colour cut limit, $f_r(M_*)$. We find that the fraction at which early-type galaxies is 50% is at $M_* \sim 10^{11}\mathrm{M}_\odot$, while at $M_* \sim 1.6 \times 10^{10}\mathrm{M}_\odot$ and $M_* \sim 8 \times 10^{11}\mathrm{M}_\odot$ the fractions are 16% and 84%, respectively. For red galaxies, the fraction of 50% is at $M_* \sim 10^{10}\mathrm{M}_\odot$, while at $M_* \sim 3 \times 10^9\mathrm{M}_\odot$ and $M_* \sim 10^{11}\mathrm{M}_\odot$ the fractions are 16% and 84%, respectively. Note that the characteristic mass at which the fraction of early-type galaxies is 50% is a factor of ~ 10 larger than for red galaxies. In general, $f_E(M_*)$ rises slower than the fraction $f_r(M_*)$. In the same figure we present the best fit model to the data. After exploring different functions, we find that two *sigmoid* functions accurately describe the functionality of $f_E(M_*)$ or $f_r(M_*)$:

$$f_k(M_*) = \frac{1 - A}{1 + e^{-\gamma_1(x_{C,1} + x_{0,1})}} + \frac{A}{1 + e^{-\gamma_2(x_{C,2} - x_{0,2})}}, \qquad (3.6.1)$$

---

[10]Figure 5 in Bernardi et al. (2010) shows that using $c = 2.85$ separates galaxies into earlier and later morphologies. While this selection criteria is not perfect, their Figure 18 shows that using the above concentration is very similar to the E+S0 GSMF based on the Fukugita et al. (2007) sample.

[11]We construct volume-limited samples that are complete in M_* and compute the GSMF as described in Section 3.3.3. In this case we slightly modified Eq. (3.3.21) by shifting our stellar mass limit by 0.4 dex, that is, $\log M_{*,\mathrm{lim,NA10}}(z) = \log M_{*,\mathrm{lim}}(z) + 0.4$.

Table 3.2: Best fit parameters to the fraction of early-type and red galaxies

Sample	A	γ_1	$\log \mathcal{M}_{C,1}$ [M$_\odot$]	$x_{0,1}$	γ_2	$\log \mathcal{M}_{C,2}$ [M$_\odot$]	$x_{0,2}$
$P_{\mathrm{HC}} > 0.65$	0.46	3.75	11.09	0.09	1.51	10.38	0.462
Red galaxies	0.21	2.44	10.66	0.36	1.81	9.68	0.070

where $k = E$ or r, $x_{C,i} = M_*/\mathcal{M}_{C,i}$, with $i = 1, 2$. The best-fit parameters for the two fractions are listed in Table 3.2.

To derive the analytic model for the GSMF of early- and late-type galaxies we use the best fit model to our GSMF, Section 3.5, and the best fit model for $f_E(M_*)$. For masses below 5×10^8 M$_\odot$ we extrapolate $f_E(M_*)$. This is an acceptable approximation since as seen in Fig. 3.9, the fraction $f_E(M_*)$ tends to ~ 0 below $M_* = 10^9$ M$_\odot$. Recall that our main goal in this chapter is to derive the MFs for HI, H$_2$, cold gas, and baryons by combining the observed HI and H$_2$ CPDFs with the GSMF, both for early- and late-type galaxies, over a large mass range. Thus, at this point we are in a position to determine these MFs.

3.7 Summary and Conclusions

In this chapter we determined the GSMF from the SDSS DR7 based on the photometric catalog from Meert et al. (2015) and Meert et al. (2016) for masses above $M_* = 10^9$ M$_\odot$. For masses down to $\sim 3 \times 10^7$M$_\odot$, we used the `low-z` SDSS DR4 (Blanton et al., 2005a,c), and corrected it from surface brightness incompleteness and fluctuations due to large scale structures. We determined also the fractions of early- and late-type galaxies by using the SDSS DR7 morphological classification of Huertas-Company et al. (2011) and therefore mass functions for these two populations.

Stellar masses were derived from five colour-dependent mass-to-light ratios. We used as our fiducial definition the geometric mean of these five stellar masses derived for each galaxy. We also determined the impact of systematic errors in M_* due to mass-to-light ratio uncertainties in our MFs.

We find that the low-mass slope of our GSMF, corrected for surface brightness incompleteness, is $\alpha \approx -1.4$, consistent with recent determinations based on the deeper surveys such as GAMA (Wright et al., 2017), and estimations based on the search of low surface brightness galaxies from core-collapse supernovae (Sedgwick et al., 2019). The slope for the high mass-end is shallower than previous determination most likely as the result of the new photometric catalog employed in this Chapter (Meert et al., 2015). Similar results have been reported in Bernardi et al. (2017).

From these results we highlight the following conclusions:

- The total GSMF is well fitted by a function composed of a sub-exponential Schechter function and a double power-law function. This fitting model has an error of less than $\sim 2\%$ in the mass range $2 \times 10^9 - 5 \times 10^{11}$ M$_\odot$. At the smallest and largest masses, the deviations increase to values above $\sim 20\%$. In contrast, the commonly employed double Schechter function model performs considerably worse.

- Systematic errors due to stellar population synthesis models, that affect results on mass-to-light ratios, introduce a systematic effect on the normalisation of the GSMF, especially at the massive-end. We find differences between $\sim 0.5 - 1$ dex, consistent with the result discussed in Bernardi et al. (2017).

Chapter 4

The bivariate gas–stellar mass distributions and the mass functions of early- and late-type galaxies at $z \sim 0$

ABSTRACT

In this Chapter we report the bivariate HI- and H_2-stellar mass distributions of local galaxies in addition of an inventory of galaxy mass functions, MFs, for HI, H_2, cold gas, and baryonic mass, separately into early- and late-type galaxies. The latter is determined using the HI and H_2 conditional distributions and the GSMF reported in Chapters 2 and 3 respectively. The obtained HIMFs agree with radio blind surveys. Similarly, the H_2 MFs are consistent with CO follow-up optically-selected samples. We deconvolve our MFs from random errors to obtain the intrinsic MFs. Using the MFs, we calculate cosmic density parameters of all the baryonic components. Baryons locked inside galaxies represent 5.4% of the universal baryon content, while ∼96% of the HI and H_2 mass inside galaxies reside in late-type morphologies. Our results imply cosmic depletion times of H_2 and total neutral H in late-type galaxies of -1.3 and 7.2 Gyr, respectively, which shows that late type galaxies are on average inefficient in converting H_2 into stars and in transforming HI gas into H_2. Our results provide a fully self-consistent empirical description of galaxy demographics in terms of the bivariate gas–stellar mass distribution and their projections, the MFs.

---

[1]Este Capítulo es parte estructural del proyecto de Tesis. Mi contribución a la parte del artículo relacionado con este Capítulo ha sido a la par del primer autor en lo que refiere a cálculos y desarrollo de códigos, así como en la conducción del mismo. Particularmente he desarrollado la actualización del código en python del capítulo 2 para generar figuras y tablas de 1) la relaciones R_{HI}-M_* y R_{H_2}-M_*, así como su dispersión intrínseca y 2) las funciones condicionales de distribución de R_{HI} and R_{H_2} así como sus momentos (promedio y desviación).

4.1 Introduction

The determination of the matter-energy content of the Universe is one of the most important achievements from the recent advances in observational cosmology (e.g., Planck Collaboration et al., 2016a, 2018). Current determinations are fully consistent with the spatially-flat Λ Cold Dark Matter (ΛCDM) cosmology, with a present-day matter-energy content dominated by the cosmological constant, $\Omega_\Lambda = 0.689$, and contributions of cold dark matter and baryon matter of $\Omega_{\mathrm{cdm}} \approx 0.262$ and $\Omega_{\mathrm{bar}} = 0.049$, respectively (for a value of the normalized Hubble constant of $h = 0.674$, Planck Collaboration et al., 2018; Aver et al., 2015; Cooke et al., 2018). Therefore, the universal baryon mass density fraction is $f_{\mathrm{bar,U}} \equiv \Omega_{\mathrm{bar}}/\Omega_{\mathrm{m}} = 0.158$, where $\Omega_{\mathrm{m}} = \Omega_{\mathrm{cdm}} + \Omega_{\mathrm{bar}}$. How much of these baryons, and their different components, are locked inside galaxies? This Chapter addresses this question by quantifying the contribution from stars, atomic and molecular gas in galaxies of different masses and morphological types.

According to the current paradigm of structure formation, non-baryonic dark matter played a major role in the evolution of the non-linear structures that we see today. Particularly, galaxies are believed to form and evolve within extended dark matter haloes, where multiple physical mechanisms are responsible for self-regulating star formation and thus setting up their observed properties (for reviews see, Mo et al., 2010a; Frenk & White, 2012; Somerville & Davé, 2015). As dark matter structures and galaxies evolve, baryons are redistributed from an initial smooth distribution to a more complex variety of structures. Of primordial importance for galaxy evolution is the amount of neutral hydrogen available for the formation of stars. Gas radiative cooling within the haloes regulates the inflow of cold gas to galaxies. The subsequent formation of stars is regulated by a complex interaction between cold gas inflows and the gas heating/outflows produced by stars, a process that on long timescales will depend on halo mass. In low-mass halos, the stellar feedback, mostly form Supernova (SN) explosions, is able not only to heat the interstellar medium (ISM) but also to expel large gas fractions from the galaxy. In high-mass haloes, the long cooling time of shock-heated gas and the powerful feedback from rapidly accreting supermassive black holes that heats and/or expels the gas tend to suppress the star formation. Thus, it is not surprising that the expected fraction of baryons inside galaxies will differ from the universal baryon fraction, $f_{\mathrm{bar,U}}$. Therefore, constraining the fraction of baryons and their different components in galaxies (mainly stars, atomic and molecular gas), is essential to constrain the processes that have taken place during the evolution of the galaxies.

One of the main properties of galaxies are their stellar masses M_*. Indeed, the abundance of galaxies as a function of M_* provides important clues regarding the evolution of the galaxy population (e.g., Peng et al., 2010, 2012; Yang et al., 2012; Rodríguez-Puebla et al., 2017). Over the last two decades, there has been a remarkable progress in assembling large galaxy samples from multi-wavelength sky surveys that have led to robust determinations of the galaxy stellar mass function (GSMF; for recent discussions, and compilations of observations up to high redshifts see, Conselice et al., 2016; Rodríguez-Puebla et al., 2017). While there have been similar efforts in assembling galaxy samples for atomic gas mass, M_{HI}, based on radio blind observations (e.g., Zwaan et al., 2003; Meyer et al., 2004; Koribalski et al., 2004; Kovac et al., 2005; Martin et al., 2010; Haynes et al., 2011; Hoppmann et al., 2015; Haynes et al., 2018) or from follow-up subsamples based on optical/infrared surveys (e.g., Springob et al., 2005; van Driel et al., 2016), these are relatively shallow and/or in small volumes compared to the optical/infrared sky surveys, as well as strongly affected by selection effects. Therefore, the demographical analysis of M_{HI} is challenging especially when determining the low- and high-mass ends of the HI mass function, HI MF (for a more detailed discussion, see Jones et al., 2018)[2], as well as other statistics like the HI two-point correlation functions[3]. The situation is not that different and even more challenging for the molecular gas as there are not blind galaxy samples in H_2. Nonetheless, there are some notable efforts to use optically-selected samples combined with small and shallow CO surveys to indirectly derive, from the (uncertain) CO-to-H_2 mass conversion factor, the galaxy mass function in H_2, H_2 MF (e.g., Keres et al., 2003; Lagos et al., 2014; Saintonge et al., 2017; Andreani et al., 2018). Unfortunately, these CO surveys are also subject to incompleteness and selection effects or subject to a large fraction of galaxies with upper limits reported due to flux detection limits.

As mentioned above, galaxy formation is a non-linear and complex process. Remarkably, well-defined correlations (usually power-laws) are, however, found from the observations. Among these are the correlations between the HI and stellar mass, M_{HI}–M_*, and the H_2 and stellar mass, M_{H2}–M_*. While both correlations present large scatter, when divided into early- and late-type galaxies they tend to follow different and tighter correlations (e.g., Calette et al., 2018, and references therein). This is not surprising given that the formation histories of early- and late-type galaxies were different. Thus, understanding the contribution of these two populations to the abundance of galaxies traced by HI and H_2 provides further key constrains to galaxy formation theory models.

---

[2]As we will discuss in Section 4.5, studying the very low-mass end of the HI MF is beyond the scope of this work.

[3]Two point correlation functions will be discussed in Chapter 6

In a recent work, Calette et al. (2018, hereafter Chapter 2) were able to determine empirically not only the mean M_{HI}–M_* and M_{H_2}–M_* relations and their scatters for early/late-type galaxies but also the full conditional probability distribution functions of M_{HI} and M_{H_2} given M_*, hereafter HI-CPDF and H_2-CPDF, respectively. In the present Chapter, we combine the empirical CPDFs with the $z = 0$ GSMF to derive the bivariate gas-to-stellar mass distributions and the MFs for the HI, H_2, cold gas, and baryon components, for all galaxies as well as for early and late types. Thus, the present Chapter represents a natural continuation of Chapter 2, introducing besides some updates. These updates include new constraints on the best fitting parameters to the observed CPDFs from Chapter 2.

The results reported in this Chapter integrate the HI- and H_2-CPDFs and the GSMF presented in Chapter 3 to offer a full statistical description of the local galaxy demographics traced by the stellar, HI, H_2, total cold gas, and baryon mass components. This statistical description of the local galaxy demographics is much more complete than the typically employed GSMF for constraining models and simulations of galaxy formation. The new generation of semi-analytic models (e.g., Croton et al., 2016; Lagos et al., 2018; Henriques et al., 2019; Yung et al., 2019) and cosmological Hydrodynamics simulations (e.g., Hirschmann et al., 2014; Vogelsberger et al., 2014; Schaye et al., 2015; Pillepich et al., 2018; Davé et al., 2019), and their post-processing outcomes, are now able to predict stellar, HI, and H_2 masses for large galaxy populations in cosmological boxes (see e.g., Lagos et al., 2015; Diemer et al., 2018, 2019; Popping et al., 2019). The empirically-based results presented here are optimal for comparing with these predictions as well as for calibrating theoretical models of galaxy evolution (see e.g. Romeo, 2020). The results to be presented in this Chapter are the basis for further studies as the inference of the galaxy-halo connection extended to HI, H_2, cold gas, and baryon masses.

4.2 Modelling the bivariate distributions and MFs from the conditional distribution functions

In this Section we describe the statistical method for deriving the HI and H_2 mass functions, MFs (as well as the total cold gas and baryon MFs), from the GSMF and the respective correlations of M_{HI} and M_{H_2} with M_*, or more generally, the respective full mass conditional distribution functions, CPDFs. In general, our approach allows us to calculate bivariate distribution functions of the HI or H_2 mass and the stellar mass. One can imagine that our methodology is equivalent to an optically-selected

volume-limited sample that it is complete in stellar mass, with HI and H_2 gas masses determined for every galaxy in the sample, and for which any MF can be determined. When information about morphology is available, the CPDFs are useful for deriving the corresponding MFs into different morphological components. Here, we consider that the galaxy population is divided into two main morphological groups: early- and late-type galaxies. Following Chapter 2, our definition of early-type galaxies includes morphological types that comprises E and S0 galaxies or equivalently $T \leq 0$ from the Nair & Abraham (2010a) morphology classification. Late-type galaxies are just the complement, from Sa to Irr. Below we briefly describe the basic ingredients for calculating the MFs:

- **Conditional Distribution Functions (CPDFs)**: For a fixed morphology, a galaxy of mass M_* has the chance of having either a HI or H_2 mass described by their corresponding CPDFs. We denote the CPDFs of early- and late-types by $P_E(M_j|M_*)$ and $P_L(M_j|M_*)$, respectively, where $j =$ HI or H_2. The HI-CPDF and H2-CPDF contain information about all the moments of the HI- and H_2-to-stellar mass correlations. We use the observed HI-CPDF and H2-CPDF from section 2.6. In Section 4.2.2, we describe the functional forms for the CPDFs proposed in Chapter 2.

- **Galaxy Stellar Mass Function**: The GSMF is an important input since it allows us to project the CPDFs into their corresponding MFs. We derived the GSMF for all galaxies, as well for the early- and late-type, based on the SDSS. Details regarding our methodology to compute the observed GSMF over ~ 5 decades in M_*, as well as its decomposition into early- and late-type galaxies are presented in Chapter 3.

4.2.1 Generalities

As discussed above, a CPDF, $P_j(M_j|M_*)$, determines the chances that a galaxy of mass M_* possess a specific galaxy property M_j, with $j =$ HI, H_2, cold gas or baryonic mass. Note that the units of P_j is per $M_\odot$. The relation between the distribution P_j in bins per $M_\odot$ to dex^{-1}, $\mathcal{P}_j$, is given by

$$\mathcal{P}_j(M_j|M_*) = P_j(M_j|M_*) \times \frac{M_j}{\log e}. \qquad (4.2.1)$$

The advantage of using $P_j(M_j|M_*)$ is that it contains information about *all the moments* of the distribution, in particular the mean $M_j - M_*$ relation and its standard deviation.

The *joint* distribution function of M_* and M_j, hereafter referred as the *bivariate* distribution function, is defined as:

$$\Phi(M_j, M_*) = \frac{d^2 N(M_j|M_*)}{V d\log M_j d\log M_*} = \mathcal{P}_j(M_j|M_*)\phi_*(M_*), \qquad (4.2.2)$$

where $d^2 N$ is the bivariate number of galaxies within the mass range $\log M_* \pm d\log M_*/2$ and $\log M_j \pm d\log M_j/2$ in a given volume V, and $\phi_*(M_*)$ is the GSMF in units of $\mathrm{Mpc}^{-3}\mathrm{dex}^{-1}$. The integration (marginalisation) of $\Phi(M_j, M_*)$ over M_* results in the total MF for M_j, $\phi_j(M_j)$, that is,

$$\phi_j(M_j) = \int \Phi(M_j, M_*)d\log M_* =$$
$$\int \mathcal{P}_j(M_j|M_*)\phi_*(M_*)d\log M_*. \qquad (4.2.3)$$

The above equation shows how the CPDFs are projected into a number density function via the GSMF. Note that integration of $\Phi(M_j, M_*)$ over M_j gives the total GSMF[4].

As discussed previously, when studying the properties of galaxies it is useful to separate them into, at least, two morphological components such as early types, or spheroid-dominated galaxies, and late types, or disk-dominated galaxies. Thus, the total GSMF can be formally represented as the contribution of these two types

$$\phi_*(M_*) = \phi_{*,E}(M_*) + \phi_{*,L}(M_*), \qquad (4.2.4)$$

denoted respectively by $\phi_{*,E}$, and $\phi_{*,L}$. In terms of the fraction of early- and late-type galaxies (f_E and f_L), their corresponding galaxy stellar MFs are given respectively by $\phi_{*,E} = f_E \times \phi_*$, and $\phi_{*,L} = f_L \times \phi_*$, with $f_E + f_L = 1$.

Early- and late-type galaxies are different in their HI- and H_2-to-stellar mass distributions. Thereby, Equation (4.2.3) can be generalised in terms of the distribution $\mathcal{P}_{i,j}(M_j|M_*)$, where the subscripts indicate $i =$ early or late type, and $j =$ HI, H_2, cold gas or baryonic mass. Then, the generalisation of Equation (4.2.3) to galaxies

---

[4]In the literature there are different methods to determine multivariate joint distributions, one example is the copula approach. A copula is function that joint multivariate cumulative distribution functions to their corresponding marginal distributions. They are useful to model the dependence between random variables based on uniform marginals. According to the Sklar's theorem, any multivariate joint distribution is totally defined given the marginal distributions and a copula describing the structure between random variables. More details on the copula approach and the application to the galaxy luminosity function the reader is referred to Takeuchi (2010) and Takeuchi et al. (2013). Here we use the CPDFs formalism for two reasons: 1) the input data that we use are characterised on that format, see Chapter 2 and below; and 2) Our goal is to determine the mass functions using the CPDFs.

with morphological type i and mass component j is:

$$\phi_{j,i}(M_j) = \int f_i(M_*)\mathcal{P}_{i,j}(M_j|M_*)\phi_*(M_*)d\log M_*. \tag{4.2.5}$$

Finally, the *total* CPDFs are calculated from the respective conditional distributions of early- and late-type galaxies as:

$$\begin{aligned}
\mathcal{P}_j(M_j|M_*) =& f_E(M_*) \times \mathcal{P}_{E,j}(M_j|M_*)+ \\
& f_L(M_*) \times \mathcal{P}_{L,j}(M_J|M_*),
\end{aligned} \tag{4.2.6}$$

with $j = \mathrm{HI}, \mathrm{H_2}$, cold gas or baryonic mass.

4.2.2 The HI and $\mathrm{H_2}$ Conditional Distribution Functions

As shown in Equations (4.2.2) and (4.2.5), the conditional or bivariate distribution functions are useful to statistically determine the MFs. Evidently, in the case of atomic and molecular gas, we are assuming that for every galaxy that is optically selected, there must exist HI and $\mathrm{H_2}$ counter parts. The discussion on the possible existence of pure HI or $\mathrm{H_2}$ galaxies, those that will not be observed in optically selected samples but rather in radio blind surveys, is out of the scope of this work. Note that if they exist, the chance of observing those galaxies is very low over the mass ranges that we will derive the MFs. For example, in the case of pure HI galaxies, the ALFALFA survey has found $\sim 1.5\%$ of HI sources that were not clearly associated to an optical counterpart. Of those, $\sim 75\%$ are likely tidal in origin (Haynes et al., 2011). Thus, $\sim 0.4\%$ of HI source observed in the ALFALFA survey are purely gaseous galaxies candidates, most of them at the mass range $10^7 < M_{\mathrm{HI}}/\mathrm{M_\odot} < 10^{10}$ (Cannon et al., 2015). As we will show, our completeness limit for the HI MF is $M_{\mathrm{HI}} \sim 10^8 \mathrm{M_\odot}$. The above fraction, could be considered as an upper limit as some of these sources have already detected optical counterparts revealing unusual high HI mass-to-light ratios (Cannon et al., 2015). Thus we conclude that our results are unlikely to be affected by excluding pure gas galaxies in our analysis.

The Calette et al. (2018) HI and $\mathrm{H_2}$ Conditional Distribution Functions

Here we use the results from Chapter 2 (Calette et al., 2018) who determined the HI- and $\mathrm{H_2}$-to-stellar mass ratio distributions (CPDFs) as a function of M_* from a large compilation of optically-selected samples with radio observations. Next, we briefly describe the steps taken in Chapter 2 to derive the HI and $\mathrm{H_2}$ CPDFs. The reader is referred to that Chapter for details.

The compiled data described in Chapter 2 consist of a set of incomplete and inhomogeneous samples. We first homogenised all these samples to a common IMF, cosmology, radio telescope configuration and sensitivity, and CO-to-luminosity conversion factor. Then, we selected only those samples without obvious biases due to selection effects such as environment. Radio non detections, reported in the literature as upper limits, are an important source of uncertainty when deriving distributions or correlations. In Chapter 2 we *included* non detections to derive the HI and H_2 CPDFs. Below we briefly describe the treatment that we employed for radio non detections.

In our compiled samples most of radio non detections are early-type galaxies representing a non negligible fraction of intermediate and massive galaxies, which are (typically) gas poor. An important fraction of those galaxies are from the GASS (Catinella et al., 2013) and the COLD-GASS (Saintonge et al., 2011) surveys at distance of $109 < D/\mathrm{Mpc} < 222$. Compared to other more nearby samples of intermediate and massive early-type galaxies with measurements of HI and H_2 mass, such as the ATLAS 3D (Serra et al., 2012) at $\bar{D} \sim 25\mathrm{Mpc}$, we noted that the upper limits of the GASS/COLD-GASS samples are $\sim 1 - 2$ orders of magnitude larger than nearby samples (Chapter 2). The above lead us to first introduce a correction for the upper limits of the GASS/COLD-GASS surveys by a distance effect. Recall that radio non detections or upper limits depend not only on the sensitivity of the radio telescope or integration time but also on the distance to the object. In Chapter 2 we corrected the upper limits of the GASS/COLD-GASS samples by a distance effect by using nearby samples such as the ATLAS 3D survey. Briefly, our correction consists in using the distances and upper limits from nearby samples to estimate the upper limits in the GASS and COLD-GASS as if these two samples were at the same distance as the nearby ones. We validated our procedure by using a mock galaxy survey by applying similar distance-sensitivity effects as GASS/COLD-GASS surveys, for details see Chapter 2. For late-type galaxies, notice that most of them are detected in radio due to their large fractions of gas and it is not necessary to introduce the above corrections. Next, we describe the treatment of the upper-limits to derive the HI and H_2 CPDFs.

In our analysis from Chapter 2 we *included* upper limits, or left-censored data, by using the Kaplan & Meier (1958b) non-parametric estimator. This estimator provides a reconstruction of information lost by censoring. Feigelson & Nelson (1985) adapted this estimator for astronomical samples. We used the ASURV package based on Feigelson & Nelson (1985) to derive the HI and H_2 CPDFs from our compiled samples. We have also applied the censoring Buckley & James (1979) regression method to derive the relationship and standard deviations between the HI- and H_2-

to-stellar mass ratio and M_*. We note that the regression results are consistent with the (logarithmic) mean and standard deviation values obtained from the CPDFs based on the Kaplan & Meier (1958b) estimator.

The functional forms of the HI and H$_2$ Conditional Distribution Functions

For the HI and H$_2$ CPDFs of late-type galaxies, in Chapter 2 we found that they are described by a Schechter function. In the case of early-type galaxies, the CPDFs are better described by a (broken) Schechter function plus a uniform distribution at the low$-\mathcal{R}_j$ values. Following, we describe in more detail these functional forms.

We begin by introducing the following Schechter-type probability distribution function for the HI- or H$_2$-to-stellar mass ratios, $\mathcal{R}_j = M_j/M_*$, in the range $\log \mathcal{R}_j \pm d \log \mathcal{R}_j/2$:

$$\mathcal{S}_{i,j}(\mathcal{R}_j) = \frac{\ln(10)}{\mathcal{N}_{i,j}} \left(\frac{\mathcal{R}_j}{\mathcal{R}_{i,j}^*} \right)^{\alpha_{i,j}+1} \exp \left(-\frac{\mathcal{R}_j}{\mathcal{R}_{i,j}^*} \right), \tag{4.2.7}$$

where the morphology is represented with i = early or late type, and the galaxy property is represented with j = HI or H$_2$. The parameters are: the characteristic gas-to-stellar mass ratio, $\mathcal{R}_{i,j}^*$, the normalisation parameter, $\mathcal{N}_{i,j}$, which constrains the probability to be between zero and one,[5] and the power-law slope $\alpha_{i,j}$ for the part of the distribution of galaxies with low gas-to-stellar mass ratio.

- Late-type Galaxies:

For late-type galaxies, that is $i = L$, in Chapter 2 we found that the HI-CPDF and H2-CPDF is described by the Schechter-type distribution function given by Eq. (4.2.7) with the parameters $\alpha_{L,j}$ and $\mathcal{R}_{L,j}^*$ functions of M_* as follows:

$$\alpha_{L,j} = \alpha_{0;L,j} \log M_* + \alpha_{1;L,j}, \tag{4.2.8}$$

and

$$\mathcal{R}_{L,j}^* = \frac{\mathcal{R}_{0;L,j}^*}{\left(\frac{M_*}{\mathcal{M}_{L,j}^*} \right)^{\beta_{L,j}} + \left(\frac{M_*}{\mathcal{M}_{L,j}^*} \right)^{\gamma_{L,j}}}. \tag{4.2.9}$$

Consider that $\mathcal{S}_{L,j}(\log \mathcal{R}_j)d\log \mathcal{R}_j = \mathcal{S}_{L,j}(\log M_j - \log M_*)d\log(M_j/M_*)$. By definition M_* is fixed, thus the HI and H$_2$ CPDFs of late-type galaxies are given by:

$$\mathcal{P}_{L,j}(M_j|M_*)d\log M_j = \mathcal{S}_{L,j}(\log M_j - \log M_*)d\log M_j. \tag{4.2.10}$$

---

[5]For $\alpha_{i,j} > -1$ then $\mathcal{N}_{i,j} = \Gamma(1 + \alpha_{i,j})$, with $\Gamma(x)$ the complete gamma function. In general $\mathcal{N} \propto \int_{-\infty}^{\infty} x^\alpha \exp(-x)dx$.

The above explicitly shows that the integration over conditional distribution functions can also be interpret as convolutions in Equation (4.2.3).

- Early-type Galaxies:

In the case of early-type galaxies, $i = E$, we showed in Chapter 2 that both the HI-CPDF and H2-CPDF are described as the sum of two distribution functions; the Schechter-type distribution function, $\mathcal{S}_{E,j}$, and a uniform function, $\mathcal{U}_{0,j}$,

$$\mathcal{E}_j(\mathcal{R}_j) = \begin{cases} \mathcal{U}_{0,j} & \mathcal{R}_{0,j} \leq \mathcal{R}_j < \mathcal{R}_{1,j} \\ A \times \mathcal{S}_{E,j}(\mathcal{R}_j) & \mathcal{R}_{1,j} \leq \mathcal{R}_j \end{cases}, \tag{4.2.11}$$

where $\mathcal{R}_{0,j} = \mathcal{R}_{1,j}/10$,[6] and $\log \mathcal{R}_{1,j} = r_{0,j} \log M_* + r_{1,j}$, while the uniform distribution is given by

$$\mathcal{U}_{0,j}(M_*) = \frac{p_{0,j} \log M_* + p_{1,j}}{\Delta}, \tag{4.2.12}$$

and

$$A = (1 - \mathcal{U}_{0,j} \times \Delta) \times \frac{\mathcal{N}_{i,j}}{\eta_{i,j}(\mathcal{R}_{1,j})}, \tag{4.2.13}$$

where in Chapter 2 we assumed that $\Delta = \log 10 = 1$ dex, the symbol $\eta_{i,j}(\mathcal{R}_{1,j})$ takes into account the fraction of galaxies in the Schechter-type mode for galaxies with gas ratio above $\mathcal{R}_{1,j}$.[7] The HI-CPDF and H_2-CPDF of early-type galaxies are:

$$\mathcal{P}_{E,j}(M_j|M_*)d\log M_j = \mathcal{E}_j(\log M_j - \log M_*)d\log M_j. \tag{4.2.14}$$

Constraints on the best fitting parameters

In Chapter 2 the best fit parameters for late-type galaxies, Equations (4.2.8)-(4.2.9), and for early-type galaxies, Equations (4.2.11)-(4.2.13), were constrained using the observed HI- and H_2-CPDFs on various stellar mass bins. Computing CPDFs over M_* bins requires of the GSMF in addition of the fraction of early/late-type galaxies (see Section 4.3.1 for more details). Since we are using slightly different inputs, namely the GSMF and the fractions of early/late-type galaxies, than in Chapter 2, we prefer to perform our own fits to the same data, for consistency. The results are presented in Section 4.3.1. The differences with the parameters reported in Chapter 2 are actually small.

---

[6]As discussed in Chapter 2, the observed data imply that the HI- and H_2-to-M_* ratios will not be lower than $\sim 10^{-4} - 10^{-5}$. This seems plausible since even for galaxies that transformed all their gas into stars, the gas mass recycled to the ISM by stellar evolution could provide the above minimal floor for the gas mass ratios.

[7]Similarly to late-types, in the case that $\alpha_{i,j} > -1$ then $\eta_{i,j}(\mathcal{R}_{1,j}) = \gamma(1 + \alpha_{i,j}, \mathcal{R}_{1,j})$, with $\gamma(x,a)$ as the incomplete gamma function. In general $\eta_{i,j}(a) \propto \int_a^\infty x^\alpha \exp(-x)dx$.

4.2.3 The Cold Gas and Baryonic Conditional Distribution Functions

Once we have constructed the HI-CPDF and H_2-CPDF we can now define the conditional distributions for the cold gas and baryon masses, M_{gas} and M_{bar}.

The total cold gas content in a galaxy is composed of HI, H_2, helium, and metals; helium and metals account for roughly 30% of the cold gas, $M_{\text{He}} + M_{\text{Z}} \approx 0.3 M_{\text{gas}}$. Therefore, $M_{\text{gas}} = M_{\text{HI}} + M_{H_2} + M_{\text{He}} + M_{\text{Z}} = 1.4 \times (M_{\text{HI}} + M_{H_2})$. For simplicity, let M_{HI} and M_{H_2} be two independent random variables. Section 4.5 discusses the validity of this assumption. Then, M_{gas} is a random variable with the conditional distribution function:

$$
\begin{aligned}
P_{\text{gas}}(M_{\text{gas}}|M_*) &= \frac{1}{1.4} \int P_{\text{HI}}\left(0.71 M_{\text{gas}} - M_{H_2}|M_*\right) \times \\
&\qquad P_{H_2}(M_{H_2}|M_*)dM_{H_2}, \\
&= \frac{1}{1.4} \int P_{\text{HI}}(M_{\text{HI}}|M_*) \times \\
&\qquad P_{H_2}(0.71 M_{\text{gas}} - M_{\text{HI}}|M_*)dM_{\text{HI}},
\end{aligned}
\tag{4.2.15}
$$

or after some algebra, the same distribution function but per bin in log space is:

$$
\begin{aligned}
\mathcal{P}_{\text{gas}}(M_{\text{gas}}|M_*) &= \int \frac{\mathcal{P}_{\text{HI}}(0.71 M_{\text{gas}} - M_{H_2}|M_*)}{1 - 1.4\ M_{H_2}/M_{\text{gas}}} \times \\
&\qquad \mathcal{P}_{H_2}(M_{H_2}|M_*)d\log M_{H_2}, \\
&= \int \mathcal{P}_{\text{HI}}(M_{\text{HI}}|M_*) \times \\
&\qquad \frac{\mathcal{P}_{H_2}(0.71 M_{\text{gas}} - M_{\text{HI}}|M_*)}{1 - 1.4\ M_{\text{HI}}/M_{\text{gas}}} d\log M_{\text{HI}}.
\end{aligned}
\tag{4.2.16}
$$

For the baryonic conditional distribution functions, we again assume that M_{gas} and M_* are two independent random variables. Thus $M_{\text{bar}} = M_{\text{gas}} + M_*$ is a random variable with a distribution function given by

$$
\begin{aligned}
P_{\text{bar}}(M_{\text{bar}}|M_*) &= \int P_{\text{gas}}(M_{\text{bar}} - \mathcal{M}_*|M_*) \times \\
&\qquad \delta(\mathcal{M}_* - M_*)d\mathcal{M}_* \\
&= P_{\text{gas}}(M_{\text{bar}} - M_*|M_*),
\end{aligned}
\tag{4.2.17}
$$

where P_{gas} is the conditional distribution function for gas, Equation (4.2.15), and the

Dirac-δ function appears explicitly for the M_* term. Similarly as above, we find that

$$\mathcal{P}_{\mathrm{bar}}(M_{\mathrm{bar}}|M_*) = \frac{\mathcal{P}_{\mathrm{gas}}(M_{\mathrm{bar}} - M_*|M_*)}{1 - M_*/M_{\mathrm{bar}}}. \qquad (4.2.18)$$

Finally, we derive the gas and baryon MFs using Equations (4.2.5), (4.2.16) and (4.2.18), the last two valid for early- and late-type galaxies.

4.3 Results

In this Section we present our fits to the HI- and H_2-CPDFs for early- and late-type galaxies from Chapter 2, the corresponding correlations (first and second moments), the bivariate HI- and H_2-stellar mass distributions, and the HI and H_2 MFs. We also present the total cold gas and baryonic MFs. We will show that our empirically-inferred HI and H_2 MFs agree with direct determinations from *blind* or optically/infrared (selected) radio galaxy surveys. Previous works related to our approach are, e.g., Obreschkow & Rawlings (2009); Lemonias et al. (2013); and Butcher et al. (2018).

For those interested in using our results, we provide a PYTHON code containing all the necessary information to reproduce the results presented here, for details see Section 4.7.

4.3.1 The HI and H_2 Conditional Distribution Functions

Section 4.2.2 describes the functional forms for the HI- and H_2-CPDFs of early- and late-type galaxies proposed in Chapter 2. Here, by using the determinations of the CPDFs for early- and late-type galaxies from Chapter 2, we find the best fit parameters of the proposed functional forms: a Schechter-type function and a Schechter-type + Uniform function, respectively (see Section 4.2.2). While Chapter 2 reported their corresponding best fit parameters, here we opt for an update based on our own determinations of the GSMFs, for consistency. There are two reasons for doing this: *i)* When fitting a CPDF that is determined over stellar mass bins, one should take into account contributions to this CPDF from the different masses. Weighting the conditional distributions by the GSMF takes care on that, see equations 6 and 8 of Section 5 from Chapter 2; and *ii)* Chapter 2 used the fraction of bulge-dominated galaxies from Moffett et al. (2016b) as a proxy to the fraction of early-type galaxies. As discussed in Section 3.6, the results from Moffett et al. (2016a), and thus Moffett et al. (2016b), overestimate the fraction of early-type galaxies compared to the SDSS morphological catalogues. The above could be due to the inclusion of Sa galaxies into the group of early-types. We used the above to argue in favor of our derived fraction of early-type galaxies based on the automated morphological classification from Huertas-Company et al. (2011).

Following Chapter 2, we use the Bayesian approach described previously through a MCMC method applied jointly to all the data (the CPDFs in different M_* bins) to find the best fit parameters of the proposed functions. These are listed in Table
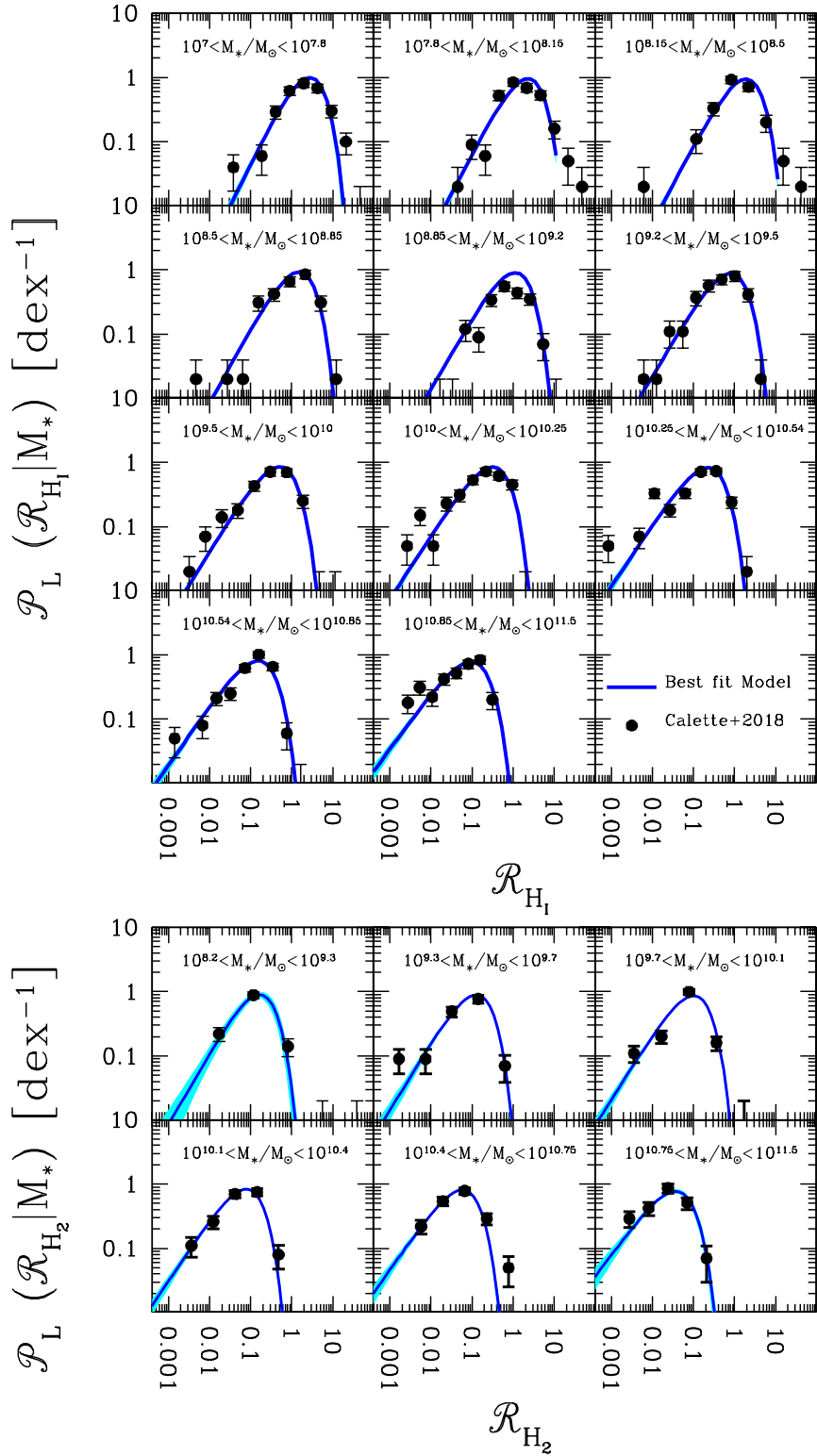
Figure 4.1: HI and H_2 mass CPDFs for late-type galaxies. The results for the compilation sample from Chapter 2 are shown as filled circles with error bars. Note that the above results include non-detections since the authors used the Kaplan & Meier (1958b) estimator for uncensored data in their analysis. Our best fitting models as well as their $1-\sigma$ scatter are shown as the solid blue lines and shaded cyan regions.
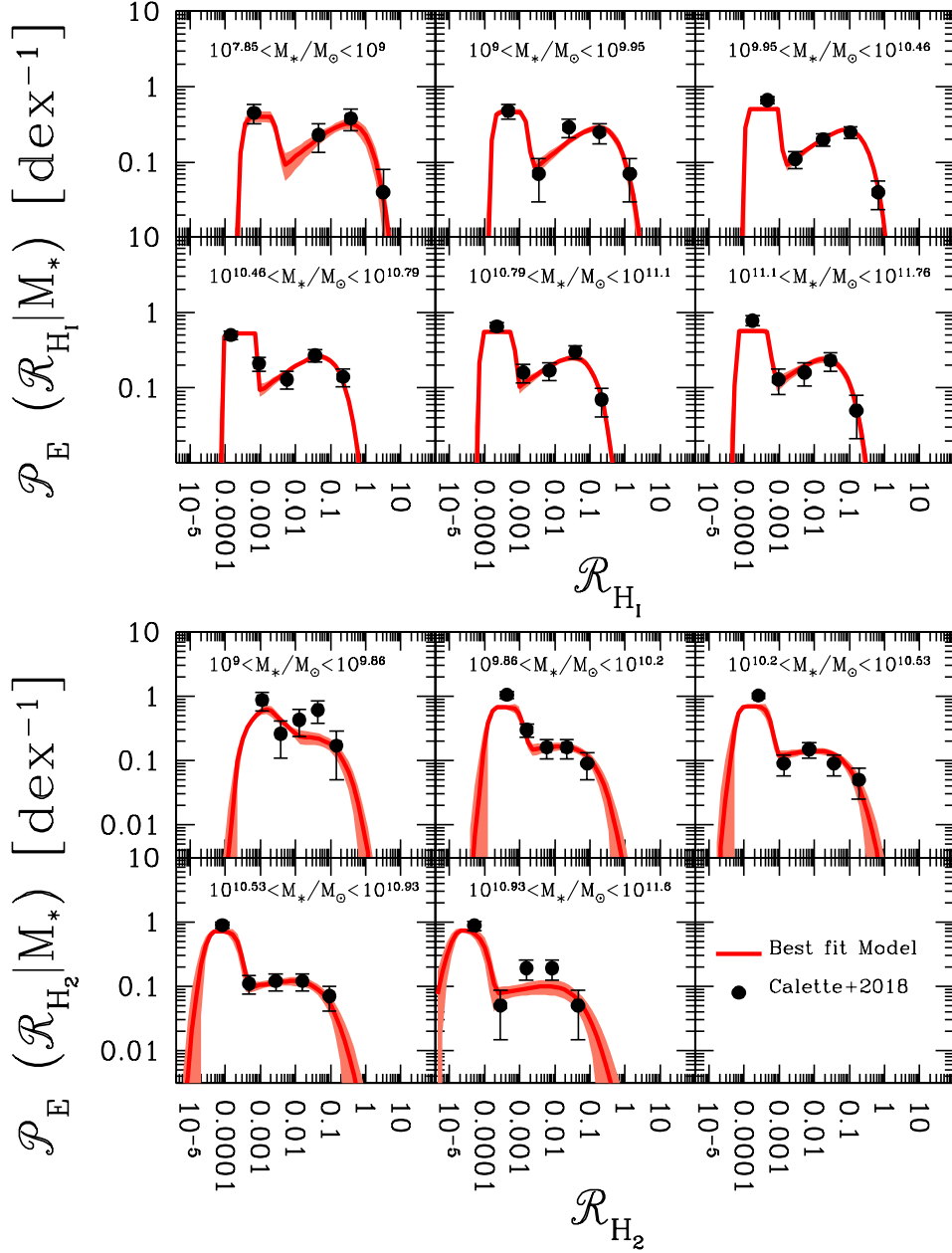
Figure 4.2: Same as Figure 4.1 but for early-type galaxies. Note that the CPDFs of early-type galaxies reported in Chapter 2 account for upper limits corrected by distance selection effects when necessary and the treated with the Kaplan & Meier (1958b) estimator, see Section 4.2.2. Our best fitting models as well as their $1 - \sigma$ scatter are shown as the solid red lines and shaded red regions.
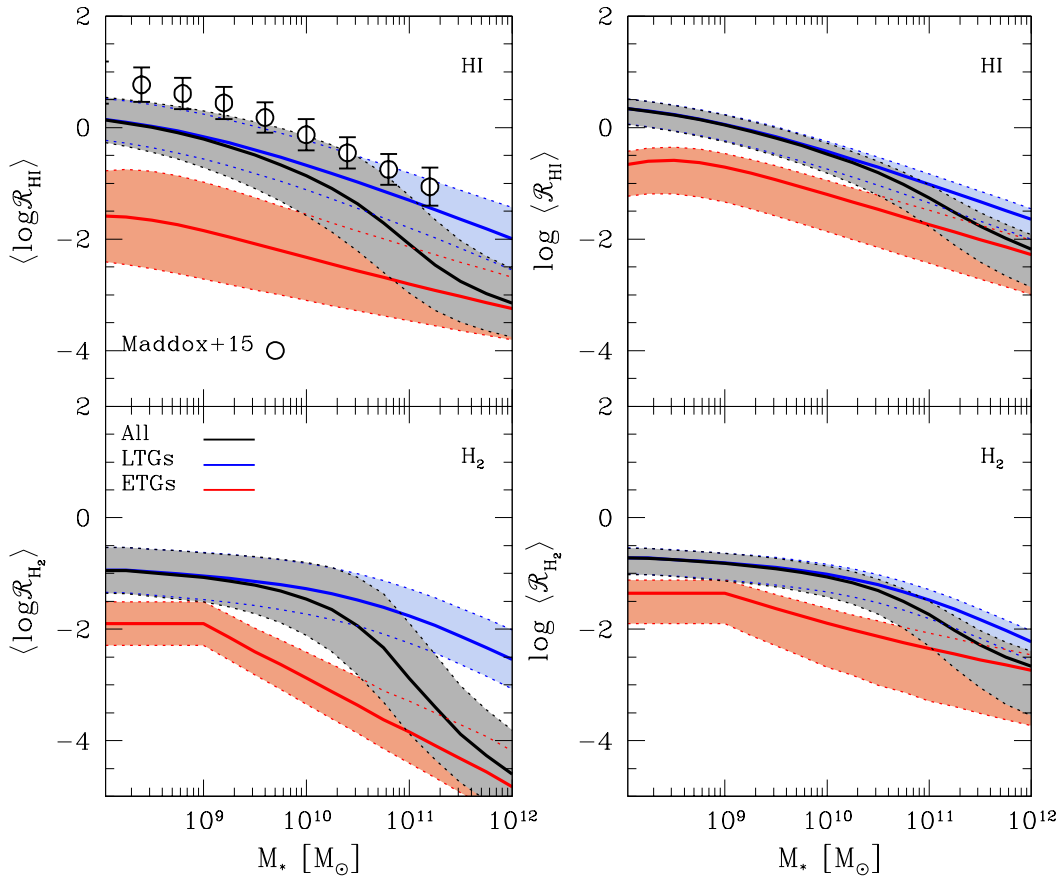
Figure 4.3: Logarithmic (left panels) and arithmetic (right panels) averaged mass ratios $\mathcal{R}_j$ as a function of M_* from our analysis, with $j = \mathrm{HI}, \mathrm{H_2}$. Blue and red lines are for early- and late-type galaxies, respectively, while the black lines correspond to all galaxies. The shaded areas show the respective standard deviations. Notice that $\log\langle\mathcal{R}_j(M_*)\rangle \geq \langle\log\mathcal{R}_j(M_*)\rangle$ and the dispersion reduces for the arithmetic mean. The open circles with error bars in the upper left panel correspond to the data from ALFALFA galaxies with SDSS spectral and stellar mass counterparts according to Maddox et al. (2015).

Table 4.1: Best fit parameters of the HI and H_2 mass CPDFs for late- and early-type galaxies

Late-Type Galaxies (Eqs. 4.2.8–4.2.10)										
Component	$\alpha_{0;l,j}$	$\alpha_{1;l,j}$	$\mathcal{R}^*_{0;l,j}$	$\mathcal{M}^*_{l,j}$	$\beta_{l,j}$	$\gamma_{l,j}$				
HI	-0.127±0.036	1.279±0.345	2.598±0.745	8.646±0.399	-0.018±0.108	0.577±0.053				
H_2	-0.085±0.120	0.830±1.213	0.122±0.037	10.595±0.301	0.841±0.195	0.063±0.089				
Early-Type Galaxies (Eqs. 4.2.11–4.2.14)										
Component	$\alpha_{0;e,j}$	$\alpha_{1;e,j}$	$\mathcal{R}^*_{0;e,j}$	$\mathcal{M}^*_{e,j}$	$\beta_{e,j}$	$\gamma_{e,j}$	$p_{0,j}$	$p_{1,j}$	$r_{0,j}$	$r_{1,j}$
HI	-0.052±0.067	-0.074±0.6840	1.573±0.533	8.354±0.258	-0.820±0.272	0.468±0.077	0.060±0.032	-0.113±0.338	-0.259±0.015	-0.310±0.160
H_2	0.059±0.069	-1.491±0.725	0.674±0.229	8.182±0.317	-0.686±0.412	0.375±0.156	0.017±0.074	0.515±0.785	-1.084±0.074	7.980±0.724

4.1. Figure 4.1 shows our best fitting models for late-type galaxies compared to the CPDFs from Chapter 2. Figure 4.2 shows the same but for early-type galaxies. We notice that our best fit parameters are very similar to those determined in Chapter 2.

4.3.2 The HI- and H_2-to-stellar mass correlations

Next, we explore the resulting first and second moments from our best fitting models to the observed HI- and H_2-CPDFs, shown in Figure 4.3. The left panels of the figure present the logarithmic mean $\langle \log \mathcal{R}_j \rangle$ and its corresponding standard deviation, $\sigma_{\log \mathcal{R}_j}$, $j =$ HI or H_2, as a function of M_* for early- and late-type galaxies as well as for all the galaxies. At low masses the correlation of all galaxies approaches the one of late-type galaxies while at high-mass end it approaches early types. The above trends are just the consequence of the observed fraction of early/late types. Figure 4.3 shows that early- and late-type galaxies follow different $\langle \log \mathcal{R}_j \rangle - M_*$ correlations. Therefore, due to the strong bimodality of these correlations conclusions based on some subset of galaxies as representative of all galaxies would lead to incorrect results.

In the literature, sometimes the gas-to-stellar mass relations are reported using the arithmetic mean (though the results are plotted in logarithmic diagrams). The right panel of Figure 4.3 shows $\log \langle \mathcal{R}_j \rangle$ vs. M_* from our empirical CPDFs. As is clearly seen, there are notable differences when computing different ways of averaging the distributions: *i)* $\log \langle \mathcal{R}_j(M_*) \rangle > \langle \log \mathcal{R}_j(M_*) \rangle$, being larger the difference for the early-type galaxies;[8] and *ii)* the standard deviations from the arithmetic mean is smaller than from the logarithmic mean.

In the left upper panel of Fig. 4.3 we reproduce the results from Maddox et al. (2015) for the ALFALFA galaxies with SDSS spectral and stellar mass counterparts. It is clear that the ALFALFA survey is biased towards galaxies with high HI-to–M_* ratios. In other words, the ALFALFA survey mainly detects galaxies in the upper envelope of the full distribution of $\mathcal{R}_{\mathrm{HI}}$ (see also Maddox et al., 2015) and is dominated mostly by late types (see also e.g., Haynes et al., 2011).

---

[8]In the case of the arithmetic mean, the contribution of low values, even if they dominate in number, could be in some cases significantly lower than higher values. Then, for the arithmetic mean the contribution of low $\mathcal{R}_j$ values is minimised contrary to the logarithmic mean of $\mathcal{R}_j$.
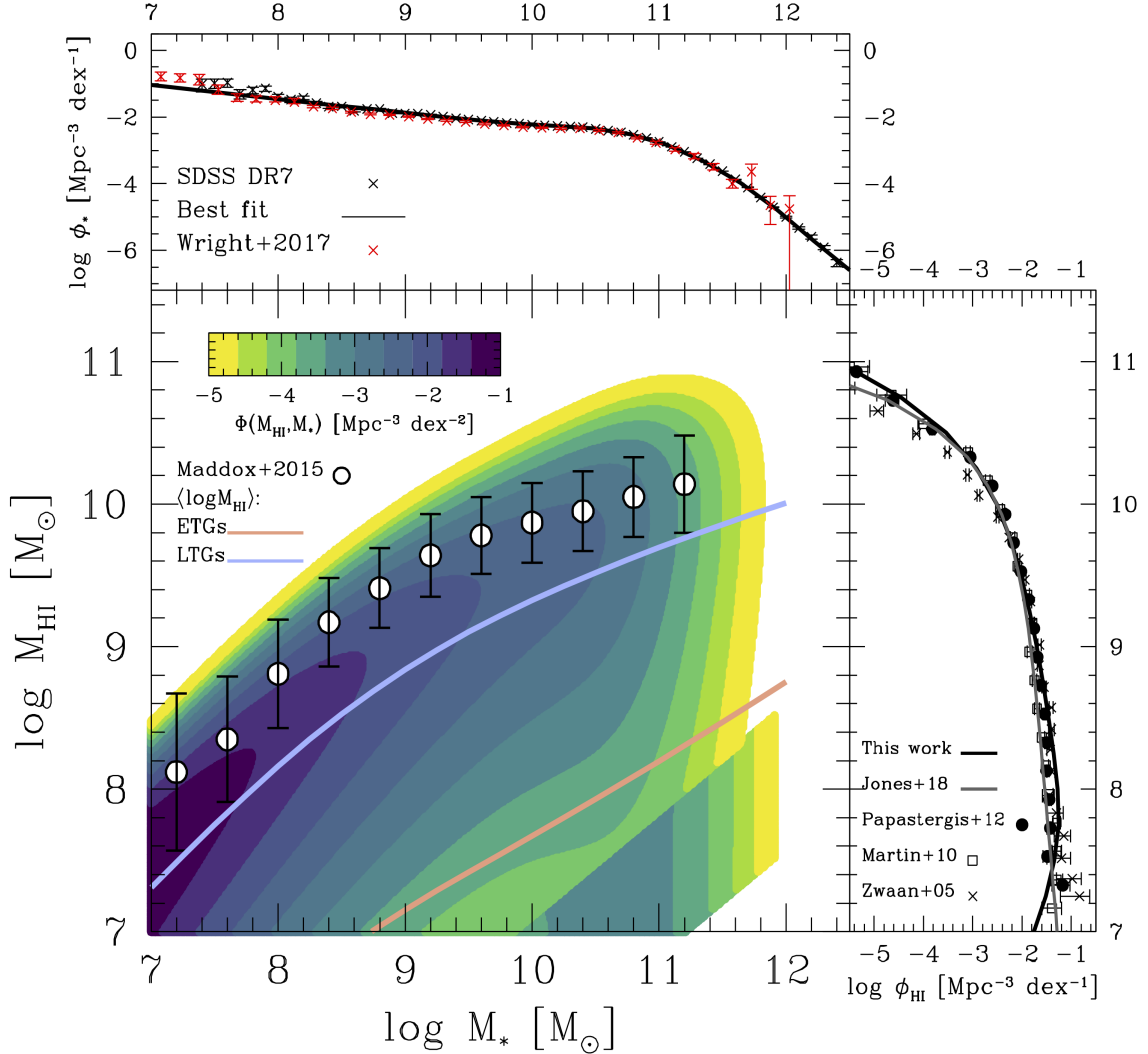
Figure 4.4: Atomic gas–stellar mass bivariate (joint) distribution function. The color code shows various number density levels as indicated by the legends. Due to the rising slope of the MFs at low masses most of the galaxies are located at small HI and stellar masses.Note that the discontinuity seen at the low-HI and high-stellar masses is due to the assumption of an uniform function for the lowest values of gas-to-stellar mass ratios of early-type galaxies where non-detections piled up. Recall that in our analysis non detections (upper limits) are included by using the non-parametric estimator Kaplan & Meier (1958b) for censored data in Chapter 2. The solid lines show the mean $\langle \log M_{HI} \rangle$ as a function of M_*, both for early- and late-type galaxies. The upper panel shows the GSMF which is the result of integrating the bivariate distribution function along the M_{HI} axis, while the bottom right panel shows the same but for the HI MF which results from integrating along the M_* axis. We compare to some previous observational determinations of the MFs and the relationship between M_{HI} and M_* derived in Maddox et al. (2015) for the ALFALFA survey with SDSS spectral and stellar mass counterparts.

Figure 4.5: Same as Figure 4.4 but for molecular gas. Note that while there are more non-detections for H_2 observations these are mostly from early-type galaxies that represent only a small fraction overall in the H_2 mass bivariate distribution function. We also compared to previous determinations from Keres et al. (2003).

4.3.3 The Bivariate Mass Distribution Functions

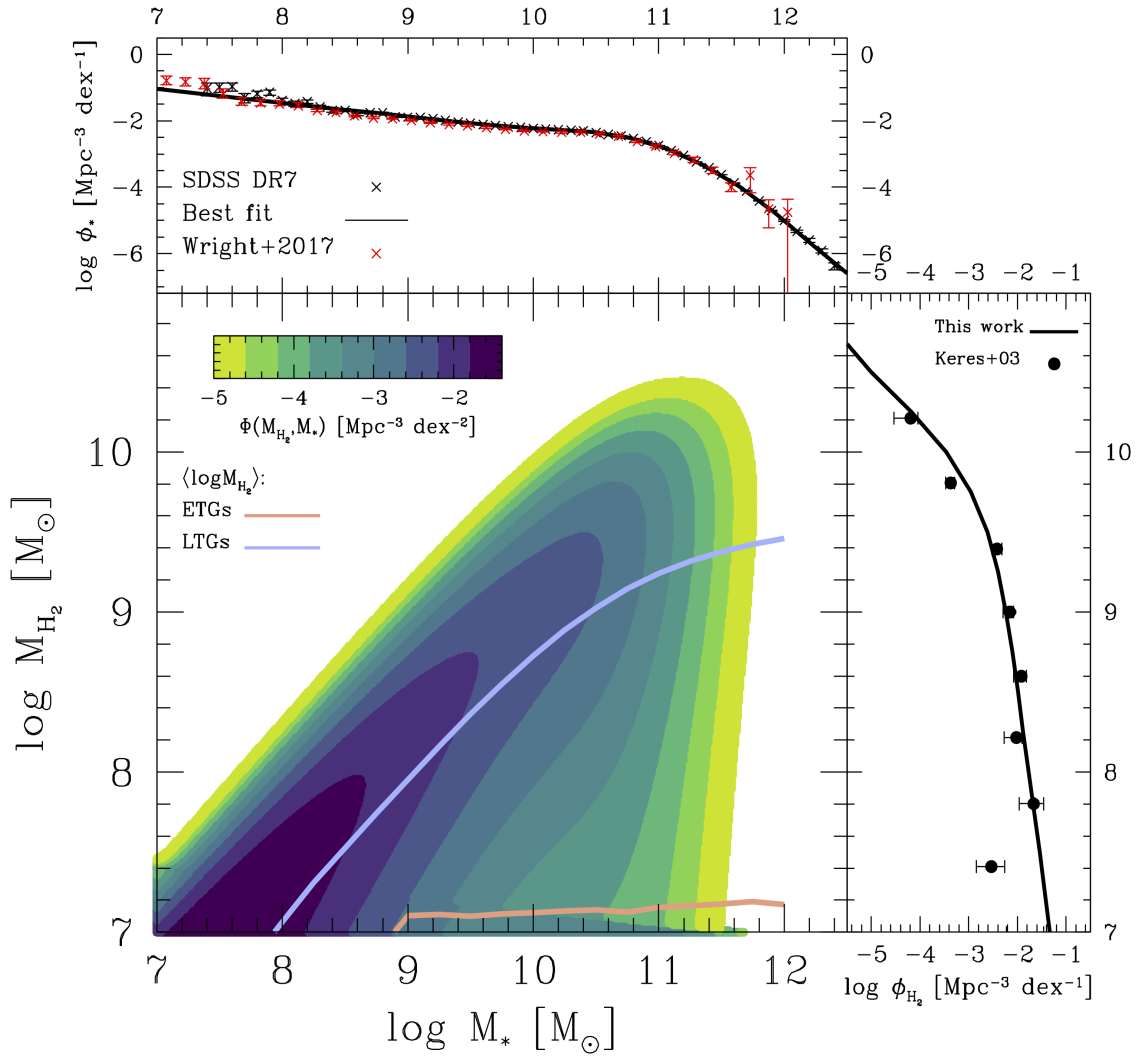Figure 4.4 shows the resulting bivariate stellar-HI mass distribution function, $\Phi(M_{\mathrm{HI}}, M_*)$, see Equation (4.2.2). The color code shows various number density levels for $\Phi(M_{\mathrm{HI}}, M_*)$. Notice that $\Phi(M_{\mathrm{HI}}, M_*)$ is for all galaxies, that is, it includes the contribution from early- and late-type galaxies. The discontinuity in the number density isocontours at the bottom right of the diagrams is related to contributions from the non-detections from early types. Recall that for the CPDFs of early types we assumed an uniform function (or top-hat) for the lowest values of the gas-to-stellar mass ratios $\mathcal{R}_{\mathrm{HI}}$, where the non-detection piled up, [9] see Section 4.2.2 and Figure 4.2. In the bottom right and upper panels of the same figure we present respectively our measurements of the HI MF and GSMF with the solid black lines. We compare the HI MF with *blind* HI galaxy surveys based on ALFALFA (Jones et al., 2018; Papastergis et al., 2012; Martin et al., 2010) and on HI Parkes All Skye Survey HIPASS (Zwaan et al., 2005). While in the next subsection we discuss in more detail the comparison with previous works, for the moment we note that our total HI MF is in good agreement with the above direct observations. In the case of ALFALFA this is a revealing result given the strong selection bias of this survey towards HI-rich and late-type galaxies as seen Figure 4.4 (open circles reproduce the results from Maddox et al., 2015, see also the discussion of the previous subsection and Figure 4.3). As we will discuss in the next section, the above reflects that the total HI MF is dominated by late-type galaxies.

Figure 4.4 explicitly shows the contribution of galaxies of different stellar masses to the HI MF. Particularly we observe that the low mass-end of the HI MF is composed mainly by low M_* galaxies but there is also a non-negligible contribution from a population of high M_* galaxies. Most of these high M_* galaxies are early-type (quenched) galaxies for which there is a significant fraction of non-detections ($\sim 55\%$). In Chapter 2 we included non-detections for the determination of the HI-CPDF based on methods of censored data (Kaplan & Meier, 1958b). Nonetheless, the contribution of non-detections is only marginal because the fraction of early-type galaxies at those masses is low, see the bottom panel of Figure 3.9. In addition, Figure 4.4 shows that the completeness limit in the HI MF, due to our stellar mass limit of $M_* = 10^7 \mathrm{M}_{\odot}$, is at $M_{\mathrm{HI}} \sim 10^8 \mathrm{M}_{\odot}$ (see below), which excludes a large region of galaxies with non-detections.

Similarly to Figure 4.4, Figure 4.5 presents the resulting bivariate stellar–H_2 mass distribution function for all galaxies and the mean $\langle \log M_{\mathrm{H2}} \rangle$ for early- and late-type galaxies. The resulting total H_2 MF is shown with the solid line in the bottom right

---

[9]Note that the top-hat is not the result of applying the Kaplan & Meier (1958b) estimator as we a posteriori redistributed the lowest values of $\mathcal{R}_{\mathrm{HI}}$ (including upper limits) into a uniform function.

panel and compared to the Keres et al. (2003) H_2 MF based on the CO luminosity function. At the low-mass end there is a substantial population of galaxies with non-detections, roughly $\sim 78\%$, which are mostly early-type galaxies. As above, non-detections have been included in the statistical analysis of the H2-CPDFs, for details see Chapter 2. Nevertheless, from the contour density level their contribution is marginal. Finally, we can conclude that our H_2 MF is complete for $M_{\mathrm{H2}} \gtrsim 10^7 \mathrm{M}_\odot$.

4.3.4 The Mass Functions

Next, we discuss in detail the MFs presented above. In particular, we focus on the determinations separately for early- and late-type galaxies based on the morphology classification described in Section 3.6. The various panels in Figure 4.6 present the MFs for atomic and molecular gas, total cold gas, and baryons, as indicated by the labels. In all the panels, the MFs for late-type galaxies are shown as blue filled circles with error bars, while for early-type galaxies are shown as red circles with error bars. These symbols are for a stellar mass limit of $M_* > 10^7 \mathrm{M}_\odot$. Instead, we use blue/red open circles when the MFs is incomplete. We also calculate the MFs in the hypothetical case of a mass limit of $M_* = 0$ shown with the blue/red dashed lines. Note that the total MFs for HI and cold gas are not plotted. This is because in these cases the MFs of late-type galaxies are hardly distinguishable from the total one at all masses. We also omit the total H_2 MF. This is because it is hardly distinguishable from the MF of late-type galaxies at $M_{\mathrm{H2}} \lesssim 2 \times 10^{10} \mathrm{M}_\odot$ while for larger masses it is indistinguishable from the MF of the early-type galaxies. In the case of the baryon masses, the total MF is plotted with black filled circles.

We note that our determinations for the MFs are the result of the convolution between random errors and the intrinsic MFs, similarly as it happens with the direct observational determinations of MFs. In Section 4.5.1 we discuss the impact from random errors and present the intrinsic MFs, after deconvolving by these errors.

The HI Mass Function

We compare our results with previous direct observational determinations of the *total* HI MFs. According to our results, late-type galaxies dominate the HI MF, even at the highest masses, so that it is practically equivalent to the total HI MF. In Figure 4.6 we reproduce the best fit to the HI MF from Jones et al. (2018) based on the final catalogue of the blind HI ALFALFA Survey, dark grey solid line. The violet solid line shows the best fit model reported in Butcher et al. (2018) to the HI MF from the Nançay Interstellar Baryons Legacy Extragalactic Survey (NIBLES), which
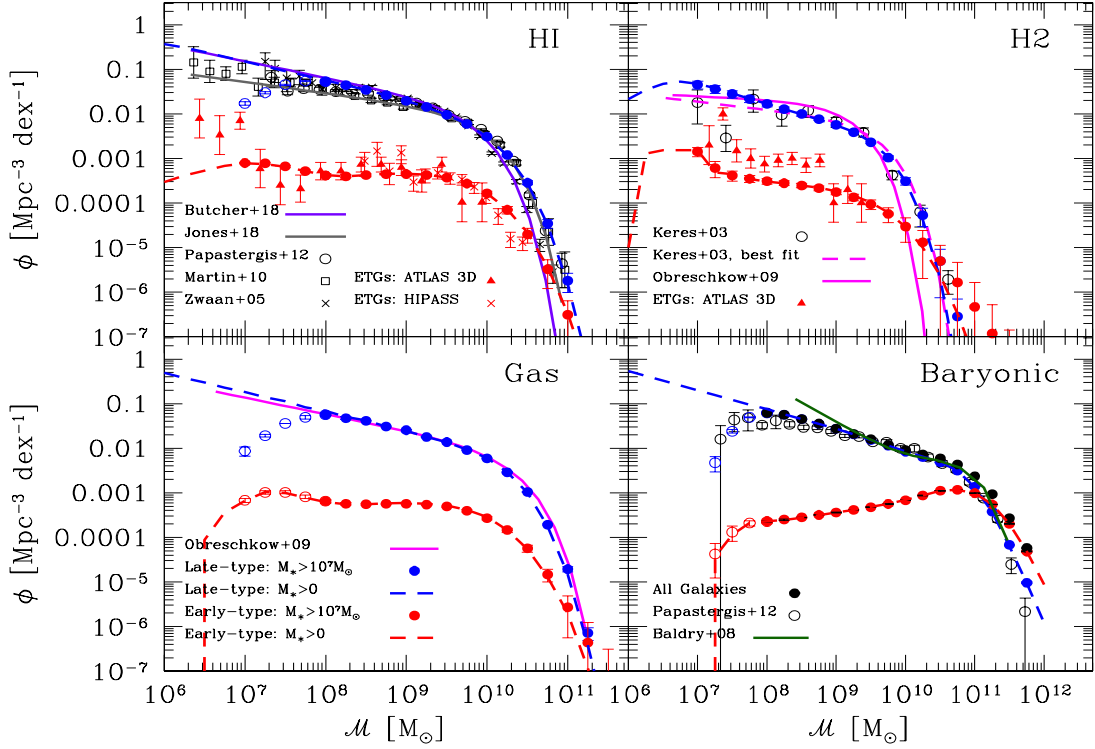
Figure 4.6: Results on the galaxy MFs of early- and late-type galaxies for atomic gas, left upper panel, molecular gas, right upper panel, cold gas, bottom left, and baryons, bottom right panel. In all the panels, late-type galaxies are shown as the blue circles with error bars while early-type galaxies are shown as the red circles with error bars, when using a stellar mass limit of $M_* = 10^7 M_\odot$. Filled blue/red circles indicate when the MFs are complete, while open circles clearly show that the MFs became incomplete. The dashed lines are for MFs when using a stellar mass limit of $M_* = 0$. The total MFs for HI and cold gas are not shown because they are practically indistinguishable from the respective MFs of late-type galaxies. Our results are in good agreement with observational determinations of the total MFs. For only early-type galaxies, we compare our results with those from the ATLAS 3D sample (Serra et al., 2012; Young et al., 2011, for HI and H_2 respectively as red triangles). While we observe some tension we suspect that selection effects are more likely to artificially increase the amplitude of their MF at low masses.

is a project that complemented recent and/or ongoing large blind HI surveys. Results from Papastergis et al. (2012) and Martin et al. (2010) based on the 40 per cent sample of the ALFALFA Survey are shown respectively with black open circles and squares. The skeletal symbols with error bars show the results from Zwaan et al. (2005), who used the HI Parkes All Sky Survey HIPASS. Note that our HI MF for late-type galaxies, which dominates the total MF, is in good agreement with direct inferences from HI blind surveys, particularly those based on ALFALFA as discussed in Figure 4.4. As for the HI MF of early-type galaxies, we plot the determinations from the ATLAS 3D (Serra et al., 2012) and HIPASS (Lagos et al., 2014) samples shown as the red solid triangles and skeletal symbols, respectively.

Our resulting HI MFs are in good agreement with direct determinations from radio observations. This is particularly true for HI masses above the completeness limit corresponding to our M_* limit for the GSMF. These masses are $M_{\rm HI} \sim 10^8 {\rm M}_\odot$ for late-type galaxies, and $M_{\rm HI} \sim 10^7 {\rm M}_\odot$ for early-type galaxies. Even when extrapolating to a limit mass of $M_* = 0$, at the low-mass end we find a good agreement with direct determinations, though the early-type galaxies from the ATLAS 3D sample present a higher amplitude for masses below $10^7 {\rm M}_\odot$. However, those extrapolations should be taken with caution as it is not clear whether we can extrapolate our bivariate distribution functions to such low masses. In conclusion, we consider that the remarkable consistency between *our* HI *MFs and radio blind surveys above* $M_{\rm HI} \sim 10^8 M_\odot$ *is reassuring* and validates the HI-CPDFs determined in Chapter 2. Recall that the observational data used in that Chapter were derived from various heterogeneous samples, affected by many selection effects, including those related to the non radio detections. Therefore, the agreement between the HI MF with that of the blind radio observations is far from trivial, unless adequate corrections are introduced and the data are adequately analysed from the statistical point of view.

The H_2 Mass Function

In the upper right panel of Figure 4.6 we present the results for the H_2 MF. The H_2 MF is largely dominated by late-type galaxies below $M_{H_2} \sim 2 \times 10^{10} {\rm M}_\odot$, but for larger masses, early-type galaxies are more abundant. In the same panel we reproduce the total H_2 MF from Keres et al. (2003), who used a CO luminosity function from FIR and $B-$band limited galaxy samples and adopting a constant CO-to-H_2 conversion factor of $\alpha_{\rm CO} = 4.76$, open black circles with error bars. The dashed line shows the best fit to a Schechter function derived in Obreschkow & Rawlings (2009). Additionally, we show the results from the ATLAS 3D sample for early-type galaxies (Lagos et al., 2014) with the filled triangles. The magenta solid line shows the

results from Obreschkow & Rawlings (2009), who derived the H_2 MF by introducing a phenomenological model for the H_2-to-HI mass ratio that depends on the galaxy morphological type and its total cold gas mass.

When comparing to the H_2 MF from Keres et al. (2003) we observe a good agreement with our results. At the low mass end, though, the Keres et al. (2003) MF seems to be slightly shallower than ours. It is not clear the origin for this discrepancy. One possibility is due the constant α_{CO} factor employed by the authors. Based on previous empirical studies, Chapter 2 showed that ignoring the dependence of α_{CO} with M_*[10] flattens the low-mass end of the H_2 MF, consistent with the results from Keres et al. (2003). Another possibility is an effect of the incompleteness of the CO luminosity function. As for Obreschkow & Rawlings (2009), our results are consistent for masses below $M_{H_2} \sim 3 \times 10^9 M_\odot$. For larger masses, the Obreschkow & Rawlings (2009) MF falls much stepper than ours. Similar to our analysis, Obreschkow & Rawlings (2009) used the relationships between galaxy properties to derive their MF. As mentioned above, their phenomenological model employed the dependence of the H_2-to-HI mass ratio with morphology and cold gas mass. While the above differences could be arguably referred to the use of different estimators for the H_2 gas masses, it could be also an indication that random errors are larger when using only one galaxy parameter. Recall that in this work we are using M_*. In Section 4.5.1, we will show that after deconvolving from random errors, our intrinsic H_2 MF becomes steeper at the high-mass end, and more consistent with the derivation from Obreschkow & Rawlings (2009). Nonetheless, it is difficult to conclude the origin of the above differences given the different nature of the models employed in both studies.

As for early-type galaxies, our results are consistent with those from the ATLAS 3D (Lagos et al., 2014) at the high mass end but they lie slightly below at the low mass end. It is unclear the reason of the above discrepancy for low-M_{H_2} early-type galaxies, though large-scale and environmental selection effects could boost the inferences of the MF when using the $1/V_{max}$ estimator, see for example, Section 3.3.3 and Baldry et al. (2008). Recall that in the case of HI, the ATLAS 3D also presents slightly enhancement at the low-mass end of the HI MF. Thus, selection effects are more likely to artificially increase the amplitude of the MF for low-mass galaxies in the ATLAS 3D sample.

---

[10]In Section 2.2.3, we have constrained the CO-to-H_2 conversion factor to be mass dependent: $\log(\alpha_{CO}) = 0.15 + 0.35[1 + 0.1(3 \times 10^{10}/M_*)^{0.64}]$ down to $M_* = 10^8$ $M_\odot$ and for lower masses the value of α_{CO} remains constant. Therefore, α_{CO} increases as M_* decreases saturating to a value of ≈ 250 for $M_* < 10^8$ $M_\odot$. This is due to the empirical dependences of α_{CO} on the gas-phase metallicity, and the dependence of the latter with M_*

The Gas and Baryonic Mass Functions

Similarly to the HI MF, the cold gas MF is completely dominated by late-type galaxies, even at the high-mass end. In the bottom-left panel of Figure 4.6 we compare our results with the phenomenological determination from Obreschkow & Rawlings (2009, pink solid line). These authors combined their inference of the H_2 MF with the HI MF from Zwaan et al. (2005) to derive the gas MF. Despite the differences mentioned above for the Obreschkow & Rawlings (2009) H_2 MF, their total cold gas MF is in excellent agreement. This is not surprising as it is just reflecting that HI is much more abundant than H_2.

Finally, we show our results for the baryonic MFs in the bottom right panel of Figure 4.6. The baryonic MF is very similar to the GSMF at the high-mass end but the at low-mass end is steeper as the contribution of cold gas becomes more relevant. On the other hand, late-type galaxies dominate the baryonic MF for $M_{\rm bar} < 10^{11}$ $M_\odot$, while at the high-mass end, early-type galaxies are more abundant than the late-type ones. We reproduce with the green solid line the baryonic MF from Baldry et al. (2008). These authors combined the GSMF from the `low-z` survey of the SDSS DR4, the same galaxy survey used here for low masses, with a closed-box model and the mass-metallicity relation to derive cold gas masses for their baryonic MF. The open black circles show the MF from Papastergis et al. (2012), who defined baryonic mass as $M_{\rm bar} = 1.4 \times M_{\rm HI} + M_*$. We notice that these previous baryonic MF determinations are in good agreement with our results at the mass range $\sim 2 \times 10^9 - 2 \times 10^{11}$ $M_\odot$, while for lower an higher masses there are some differences, which are easy to understand.

The MF from Baldry et al. (2008) is steeper than our MF at low masses. This might be a consequence of the fact that the Baldry et al. (2008) GSMF itself is steeper compared to other determinations, e.g., Baldry et al. (2012). As these authors discuss, the disagreement between the Baldry et al. (2008) and Baldry et al. (2012) GSMFs is just the result of different flow models for distances, which affect significantly nearby low-mass galaxies. Recall that our GSMF has been corrected to be consistent with the flow model of Tonry et al. (2000). Additionally, the gas masses in Baldry et al. (2008) were actually obtained from a close-box model constrained with the empirical mass-metallicity relation. The combination of these two assumptions are likely the result of a steep baryonic MF at low masses, which differs from our results and those of Papastergis et al. (2012).

Regarding the high mass end, our baryonic MF falls shallower than those of Papastergis et al. (2012) and Baldry et al. (2008). This is because our GSMF is shallower. As discussed in Section 3 there are two main systematic effects that could lead to different GSMFs, mass-to-light ratios and the determination of galaxy surface brightness

Table 4.2: Cosmic density of HI, H_2, gas, stars and baryons for all, LTGs and ETGs. The fraction of each component is denoted as $f_j = \Omega_j/\Omega_{\mathrm{bar,U}}$ with $\Omega_{\mathrm{bar,U}} = 0.048$.

	$\Omega_{H_2}/10^{-4}$	f_{H_2}	$\Omega_{HI}/10^{-4}$	f_{HI}	$\Omega_{\mathrm{gas}}/10^{-4}$	f_{gas}	$\Omega_*/10^{-4}$	f_*	$\Omega_{\mathrm{bar}}/10^{-4}$	f_{bar}
All	0.86 ± 0.05	0.18%	4.24 ± 0.10	0.88%	6.85 ± 0.92	1.43%	20.40 ± 0.08	4.25%	26.01 ± 0.13	5.42%
LTG	0.82 ± 0.04	0.17%	4.09 ± 0.10	0.85%	6.59 ± 0.89	1.37%	13.20 ± 0.05	2.75%	18.25 ± 0.12	3.80%
ETG	0.04 ± 0.01	$\sim 0.01\%$	0.15 ± 0.02	0.03%	0.21 ± 0.03	0.04%	7.21 ± 0.03	1.50%	7.76 ± 0.37	1.62%

(especially due to sky subtraction problems). Both effects are likely to affect the high-mass end of the baryonic MF. In addition, due to the small volumes of the surveys used in Baldry et al. (2008) and Papastergis et al. (2012), cosmic variance enhances the differences. Moreover, Papastergis et al. (2012) determinations are biased towards LTGs, missing largely the population of ETGs which are dominant at the high-mass end, thus, there are differences with our total GBMF but a very good agreement with our LTGs GBMF.

4.4 Cosmic density parameters and relevant timescales

4.4.1 Cosmic density parameters

The cosmic density parameter measures the average mass density of some matter species with respect to the critical density, ρ_c. Here, we determine the mass density in stars, HI, H_2, cold gas, and baryons that are locked inside galaxies by using their MFs. The differential mass density function $d\rho_j(M_j)$ for some galaxy mass component M_j in the mass range $\log M_j \pm d\log M_j/2$ is: $d\rho_j(M_j) = M_j \times \phi_j(M_j)d\log M_j$, where ϕ_j is in units of $\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$. Thus the cosmic mass density is given by:

$$\rho_j = \int_{-\infty}^{\infty} d\rho_j(M_j) \tag{4.4.1}$$

with the cosmic density parameter

$$\Omega_j = \frac{\rho_j}{\rho_c}, \tag{4.4.2}$$

where the critical density is

$$\rho_c = 2.775 \times 10^{11} h^{-1} \mathrm{M}_\odot/(h^{-1}\mathrm{Mpc})^3 = 1.2756 \times 10^{11} \mathrm{M}_\odot/\mathrm{Mpc}^3 h_{67.8}^2.{}^{[11]}$$

The limits of integration in Equation (4.4.1) reflect that we are considering all the spectrum of masses from galaxies. In reality, this is not possible, due to completeness

---

[11]We use this symbol to emphasised that $H_0 = 67.8$ km s^{-1} Mpc^{-1} in our cosmology.

limits in galaxy samples. Here, we consider the following mass limits for all our components: $M_{\rm inf} = 10^7$ $M_\odot$ and $M_{\rm upper} = 10^{12.6}$ $M_\odot$. We notice that using values smaller than $M_{\rm inf}$ and/or larger than $M_{\rm upper}$ do not substantially change our results. This is because the multiplicity functions, $\propto M_j \times \phi_j(M_j)$, have a maximum around the knee of the MFs.

Figure 4.7 shows the different values of each Ω_j corresponding to all galaxies and separately for early- and late-type galaxies, listed in Table 4.2. The Ω_j values are presented as the fractions in per cents with respect to the universal matter density ($\Omega_{\rm m} = 0.307$, left axis) and the universal baryonic density ($\Omega_{\rm bar,U} = 0.048$, right axis). To estimate errors in our cosmic density parameters, we use all our MCMC models for the HI-CPDF and H2-CPDF in addition of all our MCMC models to the fit of the GSMFs. We found that the largest uncertainties arise from the uncertainties in the CPDFs.

In the past, there have been some assessments of the cosmic density parameters at $z \sim 0$. Usually, these studies do not report cosmic density parameters for different populations and for different components at the same time. As mentioned in the Introduction, it is important to distinguish between different populations given that late- and early-type galaxies follow different formation histories. Studies close to ours are the ones by Fukugita et al. (1998); Fukugita & Peebles (2004), and Read & Trentham (2005a). Below we present and compare our results with many previous determinations from the literature.

• HI *cosmic density:* The atomic hydrogen in late-type galaxies is ~ 27 times larger than in early-type galaxies, which means that $\sim 96\%$ of HI mass is in late-type galaxies. Previously, Zwaan et al. (2005); Read & Trentham (2005a); Martin et al. (2010); Braun (2012); Delhaize et al. (2013); Hoppmann et al. (2015); Butcher et al. (2018), and Jones et al. (2018) have derived the HI cosmic density parameter by using either blind HI galaxy surveys (HIPASS and ALFALFA) or indirect techniques. The mean value from these determinations is $\Omega_{\rm HI} = 4.2 \times 10^{-4}$ with a lower bound of $\Omega_{\rm HI}^- = 3 \times 10^{-4}$ and an upper bound of $\Omega_{\rm HI}^+ = 6.2 \times 10^{-4}$, shown as the grey box in Figure 4.7[12]. Recently, using a spectral stacking technique and from WSRT observations of 1895 galaxies crossed with the SDSS, Hu et al. (2019) found $\Omega_{\rm HI} = (4.15 \pm 0.26) \times 10^{-4}$. Our determined value is in good agreement with these previous determinations, in particular with the latter one.

• H_2 *cosmic density:* The molecular hydrogen cosmic density in late-type galaxies is ~ 21 times larger than in early-type galaxies. This implies that $\sim 95\%$ of H_2 mass is in late-type galaxies. Using the CO surveys from Young et al. (1995), Keres et al.

---

[12]All the values for the papers listed above have been renormalised to a units of $h_{67.8}^{-1}$.
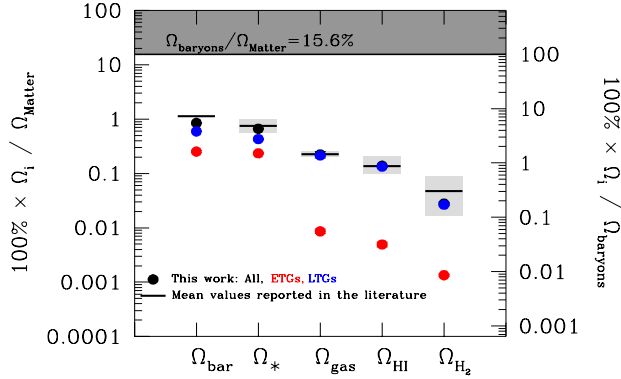
Figure 4.7: Density parameter Ω of HI, H_2, cold gas, and baryonic mass locked in all galaxies as well as in early- and late-type galaxies (coloured filled circles; the errors are smaller than the circle size). The Ω parameter values are reported as fractions in per cents of the universal matter (left axis) and baryonic (right axis) densities. The gray boxes show the range of values from previous determinations and the horizontal lines correspond to the mean of these values.

(2003) determined that $\Omega_{H2} = (1.64 \pm 0.63) \times 10^{-4}$, while from the observations in CO from Maeda et al. (2017) they report $\Omega_{H2} = 0.51 \times 10^{-4}$. Obreschkow & Rawlings (2009) used a phenomenological model to derive $\Omega_{H2} = (1.01 \pm 0.39) \times 10^{-4}$. Read & Trentham (2005a) find that $\Omega_{H2} = 2.68 \times 10^{-4}$. The above ranges of values are shown with grey box in Figure 4.7. As can be seen, our results are consistent with the range of determinations described above, especially with the results from Obreschkow & Rawlings (2009).

- *Cold gas cosmic density:* Most of the cold gas is located in late-type galaxies, $\sim 96\%$. Keres et al. (2003) found that $\Omega_{gas} = (6.34 \pm 1.62) \times 10^{-4}$, which includes the resulting abundance of HI from Zwaan et al. (1997). Obreschkow & Rawlings (2009) used their best phenomenological model to the H_2-to-HI ratio with the HIPASS results from the Zwaan et al. (2005) sample to derive $\Omega_{gas} = (6.49 \pm 1.18) \times 10^{-4}$, while using the values for HI and H_2 masses from Read & Trentham (2005a), we calculate that $\Omega_{gas} = 7.95 \times 10^{-4}$ after correcting from helium and heavier metals. Our value of $\Omega_{gas} = (6.85 \pm 0.92) \times 10^{-4}$ is consistent with the above results.

- *Stellar cosmic density:* The stellar cosmic density in late-type galaxies we derive from the SDSS is approximately ~ 1.8 larger than in early-type galaxies. Thus, $\sim 64\%$ of the mass in stars at $z \sim 0.1$ is in late-type galaxies. From the compilation by Madau & Dickinson (2014), the stellar cosmic density lies within $\Omega_* = (28.06 - 17.71) \times 10^{-4}$, while the derivations from Wright et al. (2017) and Baldry et al. (2008) are respectively $\Omega_* = 17.14 \times 10^{-4}$ and $\Omega_* = 29.73 \times 10^{-4}$. Our result for the cosmic density for all galaxies, $\Omega_* = (20.20 \pm 0.08) \times 10^{-4}$, is consistent with the above values.

• *Baryonic cosmic density:* Finally, we find that there is a factor of ~ 2.4 more baryons in late-type galaxies than in early-types, and thus, $\sim 71\%$ of the baryons are in late-type galaxies. Read & Trentham (2005a) found that $\Omega_{\rm bar} = 35 \times 10^{-4}$ which is a factor of ~ 1.3 larger than our results. We find a cosmic density parameters ratio of $\Omega_*/\Omega_{\rm bar} \approx 1.3$. Finally, our baryon density parameter is $\approx 5.4\%$ of the universal baryon fraction, $f_{\rm bar,U} = 0.156$, or equivalently ~ 18 times lower than $f_{\rm bar,U}$. Most of the baryons are definitively not locked inside galaxies.

Cosmic timescales

We are now in a position to derive some relevant cosmic timescales, such as the mean galaxy depletion times. We focus on late-type galaxies because most of the star formation occurs in those galaxies. To do so, we use the observed cosmic star formation rate (CSFR) at $z \sim 0.1$. According to Madau & Dickinson (2014), who derived the CSFR from far-UV and IR rest-frame luminosities, the CSFR is $\dot{\rho}_* \sim 90 \times 10^{-4}$ $M_\odot$ yr^{-1} Mpc^{-3} after correcting to a Chabrier (2003) IMF. Unfortunately, the authors report the CSFR for all galaxies but not divided into different groups. The recent study by Sánchez et al. (2019), based on the fossil record analysis of a sample of more than 4×10^4 galaxies from the SDSS MaNGA survey, report similar values to the the total CSFR of $\dot{\rho}_* = 114.82 \pm 67.61 \times 10^{-4}$ $M_\odot$ yr^{-1} Mpc^{-3} or $\dot{\Omega}_* = \dot{\rho}_*/\rho_{\rm crit} = (9 \pm 5) \times 10^{-14}$ yr^{-1} corrected to a Chabrier (2003) IMF. The authors also derived the CSFRs for star-forming galaxies; $\dot{\Omega}_{*,\rm SFG} = (6.5 \pm 3.8) \times 10^{-14}$ yr^{-1}. In the following we use their value for star-forming galaxies as a representative determination for late-type morphologies, that is, $\dot{\Omega}_{*,\rm L} \approx \dot{\Omega}_{*,\rm SFG}$.

We begin our discussion by estimating the mean molecular hydrogen depletion time of late-type galaxies, $\bar{t}_{\rm dep,L}(H_2) = \Omega_{H_2,L}/\dot{\Omega}_{*,\rm L}$. The H_2 depletion timescale is defined as the time at which a galaxy (or a molecular cloud) would consume its H_2 gas reservoir by forming stars at the current SFR. From our cosmic density parameters we find that $\bar{t}_{\rm dep,L}(H_2) \approx 1.3$ Gyrs. This is consistent with the mean depletion time $\bar{t}_{\rm dep}(H_2) = 0.96$ Gyr reported in Saintonge et al. (2017) for star-forming galaxies in a volume complete sample. Note, however, that for local individual galaxies the molecular depletion time could vary from ~ 0.9 to 3 Gyrs (e.g., Kennicutt, 1998; Bigiel et al., 2008; Leroy et al., 2008, 2013). Also, we estimate the mean total cold gas depletion time of late-type galaxies, $\bar{t}_{\rm dep,L}(\rm gas) = \Omega_{\rm gas,L}/\dot{\Omega}_{*,\rm L}$, and find $\bar{t}_{\rm dep,L}(\rm gas) \approx 10.14$ Gyrs, that is, ~ 8 times larger than for the molecular gas component. The values we find for these two timescales are consistent with the proposal that, on average, for local late-type galaxies, *i)* the global conversion of molecular gas into stars is inefficient (recall that the H_2 depletion times of observed local star-forming regions are actually

40-500 Myr, e.g., Lada et al., 2010, 2012); and *ii)* the global conversion of atomic to molecular hydrogen gas is also inefficient, or equivalently, the molecular cloud formation efficiency is low. Thus, the mean star formation efficiency, SFE, of local late-type/star-forming galaxies is low despite the fact that they contain a considerable amount of interstellar gas; according to Table 4.2, on average approximately 36% of the baryons in these galaxies are in form of cold gas.

According to Leroy et al. (2008), the SFE of a galaxy is the inverse of the neutral H gas depletion time, that is, the time required for current star formation to consume the neutral H reservoir. The SFE can be estimated as the product of the net efficiency of converting H_2 gas into stars, $SFE(H_2) = 1/t_{dep}(H_2)$, and the net efficiency of molecular cloud formation given by the mass fraction of H_2 with respect to the total neutral H mass in the galaxy, i.e., $M_{H_2}/(M_{HI} + M_{H_2})$. Thus, using our estimations of the cosmic parameters for late-type galaxies, we calculate the cosmic (mean) SFE of late-type galaxies as:

$$SFE_L(H) = SFE_L(H_2) \times \frac{\Omega_{H_2,L}}{\Omega_{HI,L} + \Omega_{H_2,L}} = \tag{4.4.3}$$

$$\frac{1}{\bar{t}_{dep,L}(H_2)} \times \frac{\Omega_{H_2,L}}{\Omega_{gas,L}/1.4} = 1.4 \times \frac{\dot{\Omega}_{*,SF}}{\Omega_{gas,L}} =$$

$$\frac{1.4}{\bar{t}_{dep,L}(gas)} = 1.38 \times 10^{-10} yr^{-1}.$$

The inverse of this efficiency is the cosmic neutral H depletion time, $\bar{t}_{dep,L}(H) \approx 7.25$ Gyrs. Note that the relationship between the SFE based on the neutral H gas reservoir and the SFE based on the total cold gas reservoir is $SFE_L(H) = 1.4 \times SFE_L(gas)$, or equivalently, $\bar{t}_{dep,L}(H) = \bar{t}_{dep,L}(gas)/1.4$. The factor 1.4 takes into account He and metals.

We calculate also the cosmic SF timescale of late-type galaxies, which is given by the inverse of the cosmic specific SFR, $\bar{t}_{SF,L} = \Omega_{*,L}/\dot{\Omega}_{*,L} \approx 20.3$ Gyrs; this is a factor of ~ 1.5 larger than the present age of the Universe. The cosmic SF timescale can be understood as the time required for the current cosmic SFR density to double the current cosmic stellar mass content. Interestingly enough, the ratio $\bar{t}_{dep,L}(H)/\bar{t}_{SF,L} = (\Omega_{H2} + \Omega_{HI})/\Omega_* = 0.36$, that is, the gas reservoir of late-type galaxies has not yet been dramatically consumed by star formation. Including He and metals in the gas reservoir, this ratio increases to ~ 0.5.
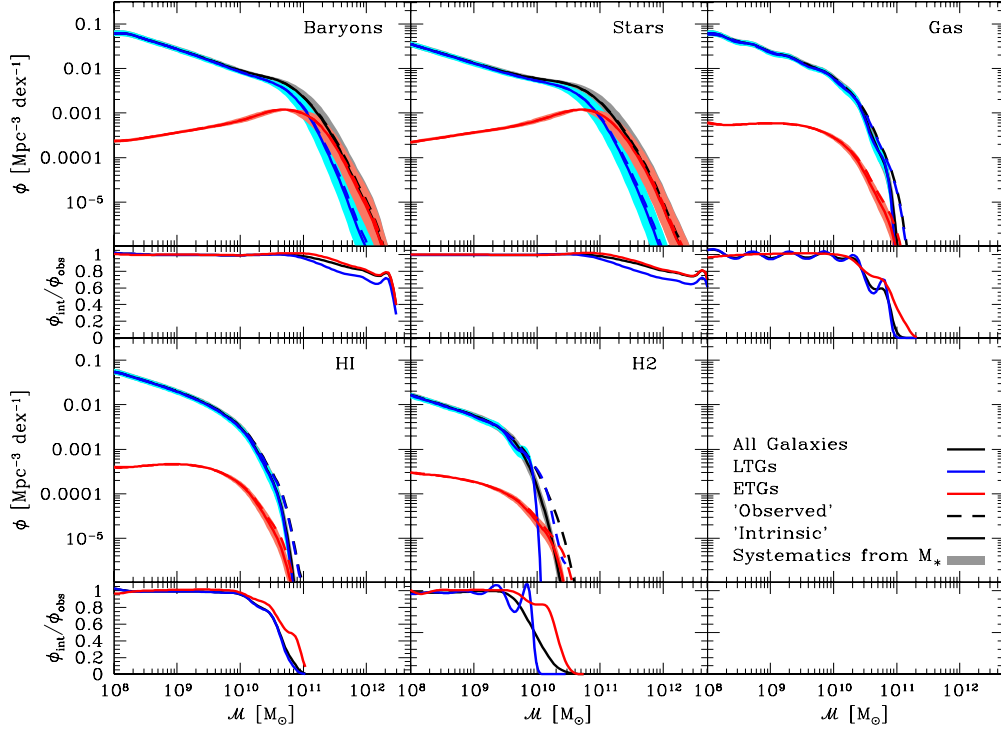
Figure 4.8: Impact of random and systematics errors in the baryonic, stellar, cold gas, atomic and molecular gas MFs for all and separately for early- and late-type galaxies. The dashed lines show the "observational" MF from Section 4.3.4 while the solid lines show the MF after deconvolving from random errors, i.e., the intrinsic MFs. Systematic errors are shown with the shaded areas. While the impact of random errors affects notably the total cold, atomic, and molecular gas MFs, the impact of systematic uncertainty on M_* is apparently marginal on them at low mass end. However, the systematic uncertainties on M_* are noticeable in the stellar and baryonic MFs.

4.5 Discussion

In this Chapter we employed a statistical approach that allows to project the observed HI- and H_2-CPDFs into their corresponding MFs, when using the GSMF as an interface or pivotal function. Additionally, the cold gas and baryon MFs are obtained from the above. Our empirical approach makes use of the following observational data as input:

1. The local GSMF over a large dynamical range and separated into early- and late-type galaxies.

2. The observed CPDFs of HI and H_2 as a function of M_*, both for early- and late-type galaxies.

As a result, our approach provides a fully self-consistent and complete empirical description of the demographics of the local population of early- and late-type galaxies for a broad mass range. Furthermore, by construction, our MFs are derived separately for early- and late-type galaxies. As discussed in Section 4.3.4, our HI and H_2 MFs are actually consistent with several previous determinations from radio *blind* or optically/infrared (selected) galaxy samples. Actually, the above level of agreement is not trivial due to the chain of assumptions and corrections for the data sets we used here and in Chapter 2, and it reinforces the robustness of the observational information employed. Note, however, that the above agreement is only valid above our completeness limit for the GSMF of $M_* = 10^7 M_\odot$, which corresponds to a completeness limit of $M_{HI} \sim 10^8 M_\odot$ and $M_{H_2} \sim 10^7 M_\odot$ respectively for the HI and H_2 MFs. In that regard, we are unable to constrain the very low mass end of the HI and H_2 MFs.

Below, we highlight aspects that we consider are relevant for the success of our empirical approach:

- The HI and H_2 CPDFs for early- and late-type galaxies. We used the CPDFs from Chapter 2, where we derived the CPDFs from a compilation of many incomplete and inhomogeneous samples, carefully homogenised to a common IMF, cosmology, CO-to-luminosity conversion factor, and accounting for selection biases.

- The effect from upper limits in radio surveys. In addition to the above mentioned homogenisation and corrections, it was important to take into account the upper limits reported in the original sources, when radio detections were not achieved. The fraction of non-detections in the compilation from Chapter 2 was non negligible, especially for early-type galaxies. Non-detections were corrected

by distance/sensitivity effects. Instead of ignoring radio non-detection or using the upper limits as the true values, as is commonly done in the literature, we derived the CPDFs by including them in our statistical analysis based on the non-parametric Kaplan-Meier estimator for censored data.

Next, we briefly discuss below some potential caveats on our approach. Over the next subsections we will discuss them in more detail and show that they do not affect our main conclusions.

- The assumption that the HI and H_2 masses are two independent random variables. In reality this is not true; for example, Obreschkow & Rawlings (2009) showed that the H_2-to-HI mass ratio depends on the morphological type. Note, however, that this was partially taken into account in our approach in a statistical sense. Recall that we use the observed mass CPDFs separately for early- and late-type galaxies, that is, the dependence with morphology is roughly included, as shown in Chapter 2, see Figure 14 from that Chapter and Fig. 4.3.

- Differences on the mass-to-light ratios. Figure 3.1 in Section 3.3.1 shows that the different mass-to-light ratios used to estimate M_* lead to different GSMFs, with differences up to 0.5–1 dex in number densities at the high-mass end (see also Bernardi et al., 2017). While we choose to use the geometric mean over the five mass-to-light ratios described in Section 3.3.1, one could naturally question that the agreement of our HI and H_2 MFs with the observed ones is relative because using a different GSMF could result in different MFs. In Section 4.5.1 we explore and quantify the impact of systematics from varying mass-to-light ratios, and show that its effect is marginal in the obtained HI and H_2 MFs.

- Random errors from stellar mass estimates. Inevitably, random errors propagate to our MFs resulting in a Eddington (1940) bias effect. Thus, the comparison with the results based on radio surveys is not trivial as they do not suffer of an Eddington (1940) bias effect due to M_* errors. Nonetheless, measurements of the HI and H_2 masses are also subject to random errors. In subsection 4.5.1 we deconvolve our MFs with the random errors, not only as a method to compare with results from radio surveys but also for obtaining the intrinsic MFs to be used to constrain the predictions from galaxy formation models.

- The morphological classification from the SDSS DR7. To derive our GSMF separated into early- and late-type galaxies we used the morphological classification based on the Huertas-Company et al. (2011) vector machine analysis of

the SDSS DR7. As shown in Figure 3.9, we find that the obtained early/late-type GSMFs using this classification are consistent with other determinations for the SDSS but disagree with those based on the visual classification from the GAMA survey. While we explore in detail this effect in Appendix C, we do not include it as one of the main source of uncertainty.

We conclude this section by emphasising the robustness of the MFs derived when combining observational gas-to-stellar mass correlations from small data sets with the GSMF, (see also, Lemonias et al., 2013; Butcher et al., 2018). While this is an indirect method to study the demographics of the galaxy distribution, it is a valid and valuable approach that gives results that are comparable to direct observations and generalise them into a full bivariate distribution.

4.5.1   The Impact of Random and Systematics Errors

When deriving stellar, HI and H_2 masses, there are two sources of errors that will inevitably propagate over the MFs: the random and systematic errors. In this Section, we discuss the impact of both sources of errors on our results.

Random Errors

The estimation of masses from both photometric and radio observations, are subject to random errors. Here, we determine their impact on our resulting MFs. For simplicity, we assume that random errors follow lognormal distributions with a constant dispersion and independent of galaxy morphology. For the stellar masses, we assume a dispersion of $\sigma = 0.1$ dex following Behroozi et al. (2010), Mendel et al. (2014), and Rodríguez-Puebla et al. (2017). For HI masses, $\sigma = 0.14$ dex, and for H_2 masses, $\sigma = 0.22$ dex, following Calette et al. (2018, and more references therein). As for the gas and baryonic masses, we assume errors of respectively $\sigma = 0.14$ dex and $\sigma = 0.1$ dex as they are dominated by HI and M_* components, respectively, especially at high masses, where random errors have a larger impact. Thus, our "observational" MFs[13] are the result of the convolution of the distribution of random errors and the respective intrinsic MFs. That is, our "observational" MFs are given by $\phi_{\mathrm{obs}} = \mathcal{G} * \phi_{\mathrm{int}}$, where the symbol $*$ denotes the convolution operation, $\mathcal{G}$ is the distribution of random errors, and ϕ_{int} is the intrinsic MF. For more details, the reader is referred to Appendix D. There we describe our numerical algorithm for deconvolving the intrinsic MF, ϕ_{int}.

---

[13]In the preceding sections we omit to use the term "observational" MFs to avoid confusion about our methodology. Here we use this term to refer that our determinations, similar to direct measurements of the MFs from galaxy surveys, suffer from random errors.

In Figure 4.8 we reproduce with dashed lines the "observational" MFs derived in Section 4.3.4. Their corresponding intrinsic MFs are shown with solid lines. In the same figure, we present the ratios $\phi_{\rm obs}/\phi_{\rm int}$ to show the effect of the deconvolution. The effect of deconvolving from random errors is small at low-intermediate masses but it increases at the massive-end since the MFs are steeper (Eddington, 1940). This is simply because the convolution depends on the logarithmic slope of the intrinsic MF (e.g., Cattaneo et al., 2008); the steeper the slope the larger the effect on the MFs. This is also the reason why we observe a lower impact in the baryonic and stellar MFs compared to the HI, H_2, and cold gas MFs; the latter fall steeply at the high-mass end. For example, the intrinsic HI MF is a factor of ~ 4 lower than the "observational" one at $M_{\rm HI} \sim 6 \times 10^{10}$ $M_\odot$, while the intrinsic H_2 MF is an order of magnitude lower than the "observational" MF at $M_{\rm H2} \sim 2 \times 10^{10}$ $M_\odot$. The intrinsic gas MF is an order of magnitude lower than the "observational" MF at $M_{\rm gas} \sim 10^{11}$ $M_\odot$. Note that for the HI, H_2, and cold gas MFs the impact of random errors is more noticeable in late-type galaxies than in early-type ones.

Systematic Errors

In addition to random errors, systematic errors have an impact when determining the MFs. The IMF is one of the most important sources of systematic errors for the GSMF. In this work we assumed an universal IMF given by the Chabrier (2003) function. While there is much debate on the IMF (see e.g., Bastian et al., 2010; Conroy et al., 2013; Bernardi et al., 2018), exploring the different alternatives is beyond the scope of this Thesis.

The stellar masses are calculated typically using colour-dependent mass-to-light ratios based on results from stellar population synthesis (SPS) models (for a recent review see Conroy, 2013). Thus, the calculated stellar masses depend on the used SPS model. This introduces a systematic uncertainty in M_*. Indeed, systematics in M_* from SPS can be as large as ~ 0.25 dex, see e.g, Pérez-González et al. (2008); Muzzin et al. (2009); Moustakas et al. (2013); Rodríguez-Puebla et al. (2017) and references therein. Recently, Bernardi et al. (2017) showed that systematics from SPS introduces errors that are as large as ~ 0.5 dex in the normalisation of the GSMF at the high mass-end. In Section 3.3.1 we found similar differences by using various recipes of colour dependent mass-to-light ratios. While in this work we calculate five different stellar masses for every galaxy, and decided to use the geometric mean of the five as our fiducial definition of M_*, the above inevitable introduces the question of *which stellar mass definition shall we use when deriving our MFs*. Additionally, Bernardi et al. (2017) determined that systematics in photometry are of ~ 0.1 dex. In order

to quantify the impact of stellar populations in our MFs, in Section 3.3.1 we noted that a constant shift of $\sim \pm 0.15$ dex in the stellar mass axis reproduces systematic errors in the GSMF. In addition, Figure 3.8 shows that the same shift in the stellar mass axis could also explain differences from photometry. Thus, hereafter, we will use a shift of ± 0.15 dex in the stellar mass axis as our fiducial model for systematic errors in the GSMF. Note that we are assuming that this shift will be independent of morphology and we are ignoring systematic errors in the atomic and molecular gas components.

Figure 4.8 shows the impact of systematic errors from SPS models and photometry as the shaded areas around their corresponding ϕ_{int} (solid lines). The effects of systematics is non-negligible at the massive-end of the stellar and baryonic MFs; we observe differences up to ~ 0.6 dex in their normalisations. This is approximately the same both for early- and late-type galaxies. The impact of systematic errors in the gas, HI and H_2 MFs is marginal; we notice a shift in their normalisations of ~ 0.07 at their low-mass ends but they increases respectively to ~ 0.4, ~ 0.4 and ~ 0.3 dex at their massive ends. The above is due to the steeper slopes observed at the high-mass end from these MFs. In conclusion, the impact of systematic uncertainties in M_* is only marginal for the derived HI, H_2, and cold gas MFs, making our results robust against this source of uncertainties.

As mentioned in the Introduction, the value of deriving robust MFs is that they can be used as key tools for constraining the processes that govern the evolution of the galaxies. However, using direct measurements from observations to constrain galaxy formation models is not trivial due to random and systematic errors, as discussed here. We end this section by emphasising the importance of deconvolving from random errors and understanding the impact of systematic errors when reporting results on galaxy demographics.

4.6 Applications: Comparison to Hydrodynamical Simulations

This Chapter has been devoted to develop a self-consistent approach for determining various levels on the demographics of galaxies traced by their different baryonic components and morphological classification, early- and late-type galaxies. This approach is supported by two key observables, separately determined for early- and late-type galaxies; *i)* the gas-stellar mass conditional distributions functions, and *ii)* the galaxy stellar MFs. Based on this, we can calculate any moment that charac-

terise the gas-stellar mass relationships, gas-stellar mass bivariate distributions as well as MFs (see for example, Figs. 4.4 and 4.5). Thus, our aim in this Section is to show a simple application on the potential of our approach by using IllustrisTNG Magneto-hydrodynamical simulation predictions.

4.6.1 The Illustris TNG Magneto-hydrodynamical simulation

The Next GenerationIllustris (Pillepich et al., 2018, hereafter IllustrisTNG) is a series of magneto-hydrodynamical simulations that model the coupled evolution of galaxies and dark matter halos within the ΛCDM paradigm. IllustrisTNG is the successor of the Illustris simulaton (Vogelsberger et al., 2014) using an updated galaxy formation model (Springel, 2010, ,AREPO) including new physics and refinements to the original Illustris model.

The galaxy formation physics considered in IllustrisTNG includes:

- Radiative gas cooling modulated by a time-variable ultraviolet bacground.

- Star formation regulated by a subgrid model for the insterestellar meduim.

- Feedback by supernova explosions.

- Metal Enrichment.

- Growth of supermassive black holes.

- AGN feedback.

These simulations are run in three boxes with cubic volumes of 50, 100 and 300 Mpc side length referred as TNG50, TNG100 and TNG300 respectively. Here we use TNG100 and TNG300 boxes.

Atomic and molecular contents in IllustrisTNG simulation

Current large cosmological Hydrodynamics simulations are not able to resolve densities at which molecular hydrogen forms, therefore the atomic-to-molecular transition must be post-processed. Diemer et al. (2018) developed a post-processing framework largely based on Lagos et al. (2015) to estimate the abundance atomic and molecular hydrogen for IllustrisTNG simulation based on the use of empirical, simulated-based and theoretical models for the atomic-to-molecular transition. Such models are:

1. Empirical models: Leroy et al. (2008).

2. Simulation-based models: Gnedin & Draine (2014) and Gnedin & Kravtsov (2011b).

3. Analytic models: Krumholz (2013b) and Sternberg et al. (2014).

Galaxies were selected to be complete in both stellar and cold gas mass by setting a minimum limit in these quantities. Mass limit for TNG100 is 2×10^8 $M_\odot$ for both, cold gas and stellar mass. TNG300 stellar mass cut is 5×10^{10} $M_\odot$ and 5×10^9 for cold gas.

Morphological classification

To make a fair comparison with IllustrisTNG, one would like to have not only stellar and cold gas contents but also morphological information on galaxies. Moreover these morphologies must be comparable with observations.

In a recent work Rodriguez-Gomez et al. (2019) generated synthetic images for IllustrisTNG simulated galaxies using SKIRT radiative tranfer code in the stellar mass range $9.8 < \log M_* < 11.3$. These authors obtained 27000 images designed to match the Pan-STARRS observational sample.

Here, we use three morphology estimators applied to IllustrisTNG synthetic images for the stellar mass component: (1) based on concentration index C (Bershady et al., 2000; Conselice et al., 2003), (2) Sérsic (1963) index and (3) probability of being ETG, P_{HC}, from the analysis presented in Huertas-Company et al. (2019) to the images generated for the IllustrisTNG based on a Convolutional Neural Network, CNN, trained on optical morphologies from the SDSS (Domínguez Sánchez et al., 2018). We do not use these estimators for HI and H_2 mass components due to very low number of galaxies in Diemer et al. (2018).

An important caveat is that when training on observations, simulations do not reproduce realistic morphologies. For this reason, the comparison with the results using Huertas-Company et al. (2019) should be taken with care. Moreover, the disagreement with our results might be in large part due to the above.

4.6.2 Confronting IllustrisTNG MFs to our empirical determinations

Figure 4.9 presents a comparison between our deconvolved MFs and the results from the magneto-hydrodynamic cosmological simulation IllustrisTNG (Marinacci et al., 2018; Naiman et al., 2018; Nelson et al., 2018; Pillepich et al., 2018; Springel et al., 2018) for the stellar, HI and H_2 components. The left panel of Figure 4.9 shows the
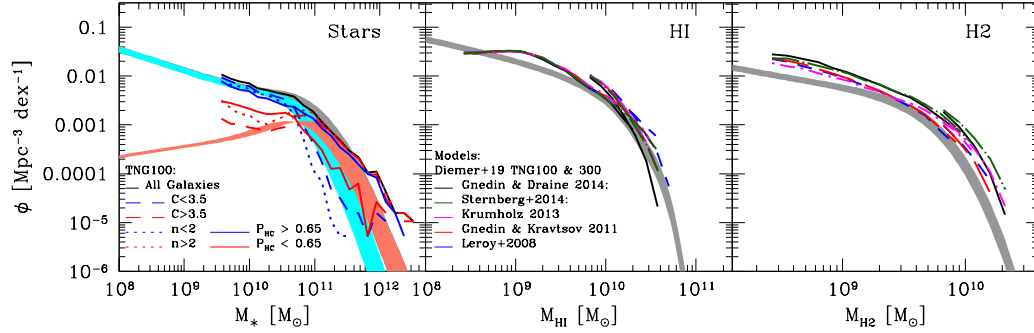
Figure 4.9: Comparison of our deconvolved MFs to the results from IllustrisTNG. *Left panel:* The GSMF of early- and late-types compared to the IllustrisTNG 100 MFs. Morphological classifications are derived based on the synthetic images generated in Rodriguez-Gomez et al. (2019) according to the concentration index C (Bershady et al., 2000; Conselice et al., 2003), long dashed lines, and Sérsic (1963) index, dotted lines. Results based on the analysis of the optical morphologies from the IllustrisTNG synthetic images Huertas-Company et al. (2019) are shown with the solid lines. Our results are plotted as cyan, red and gray shaded regions for LTGs, ETGs and total GSMFs respectively. We find a better agreement with IllustrigTNG results based on the concentration and Sérsic index. *Middle panel:* The models developed in Diemer et al. (2018) for the HI applied to the IllustrisTNG 100 and 300 simulations in Diemer et al. (2019) are shown with the various lines. The black solid line shows the model based on Gnedin & Draine (2014), the green long dashed-dotted lines show the model based on Sternberg et al. (2014), while the magenta short dashed-dotted and red dashed lines show the models from Krumholz (2013a) and Gnedin & Kravtsov (2011a), respectively. Finally, the model from Leroy et al. (2008) is shown with the blue dashed lines. We plot our total deconvolved HI-MF with gray shaded areas. *Right panel:* Similar to the middle panel but for H_2. The models applied to the IllustrisTNG 100 and 300 simulations in Diemer et al. (2019). We plot our total deconvolved H_2-MF with gray shaded areas.

comparison with our total, early- and late-type deconvolved GSMFs, shaded areas represent systematics in our measurements. The long dashed lines show the morphological classification according to the concentration index C for the IllustrisTNG. The dotted lines show the same but for the Sérsic (1963) index. The solid lines show result using Huertas-Company et al. (2019). For consistency, we define early-type galaxies using $P_{\mathrm{HC}} > 0.65$. Our results show that the total GSMF from the IllustrisTNG is very consistent within the *systematic errors* only at the high-mass end. It is important to note here that showing systematics errors could help to alleviate some potential tensions between observations and models. As for the morphological classification, we notice that the results based on the concentration and Sérsic index are more consistent with our results than the results based on the CNNs. The middle and right panels compares respectively our HI and H_2 MFs to the models developed in Diemer et al. (2018) for the HI and H_2 and applied to the IllustrisTNG 100 and 300 simulations in Diemer et al. (2019), see the labels and the caption for the models shown in the figure. The different models from Diemer et al. (2019) are very similar between each other and in a good agreement for the HI-MF. This is not the case for the H_2-MF, the normalization of H_2-MFs in the simulations are above ours by a factor of $\sim 0.4 - 0.6$ dex at low and high-mass end respectively.

4.7 Summary and Conclusions

We present a self-consistent empirical approach that unifies local galaxy gas-to-stellar mass correlations and the MFs of galaxies traced by their different baryonic components. We make available a PYTHON code that displays tables and figures with all the relevant statistical distributions and correlations discussed in this Chapter.[14] Next, we summarise our main results which can be used for comparing with theoretical predictions or as input for modeling galaxy mock catalogs:

- *Conditional probability distribution functions (CPDFs):* Section 4.2.2 presents the functional forms for the HI and H_2 mass conditional distributions given M_* (the CPDFs), which are described by Equations (4.2.7)-(4.2.12). Our best-fit parameters to the empirical information presented in Chapter 2 are listed in Table 4.1, while Figures 4.1 and 4.2 show the data with their corresponding best fits in various stellar mass bins. Theoretical predictions for the HI, H_2 and cold gas CPDFs can be confronted with our empirically constrained distributions, for all galaxies as well as for early and late types in case the morphological

---

[14]`https://github.com/arcalette/Python-code-to-generate-Rodriguez-Puebla-2020-results`

classifications are available. If these predictions are limited in stellar and/or gas masses, then our (analytical) HI-CPDFs and H_2-CPDFs and their moments can be easily calculated over the same mass ranges as the theoretical predictions for a comparison. The HI- and H_2-CPDFs combined with the GSMF allowed us to calculate the respective bivariate mass distributions for all the galaxy population as plotted in Figures 4.4 and 4.5.

- *Moments of the CPDFs:* The (analytical) HI- and H_2-CPDFs contain the information about any moment of the distributions. Figure 4.3 (see also Figs. 4.4 and 4.5) shows the $\langle \log M_j \rangle$-$\log M_*$ relationships, with $j = $ HI, H_2, for early- and late-type galaxies as well as for all galaxies. In addition we present these relationships using the arithmetic mean, $\langle M_j \rangle$. As expected, these relationships lie above from those calculated with the logarithmic mean, $\langle \log M_j \rangle$. Moreover standard deviations can vary significantly if they are computed with respect to the arithmetic or logarithmic mean, which also depends on the shape of the distributions. Other statistical measures that can be used to characterise the population distributions are medians and percentiles, for example. As mentioned above, any statistical quantity can be computed with our CPDFs and confronted with both theoretical and/or observational results.

- *Calculated Mass Functions:* Section 4.3.4 presents the results of calculating with our approach the MFs for atomic, molecular, cold gas and baryons for early- and late-type galaxies, as well as for all galaxies. As discussed in Section 4.5.1, random errors in mass determinations artificially decrease the slope of the "observational" MFs, an effect that affects especially the high-mass end, and that would lead to incorrect conclusions when comparing to theoretical predictions. Figure 4.8 presents our MFs deconvolved from random errors, that is, the *intrinsic* MFs, for different baryon matter components, and separately for early- and late-type galaxies. In the same Section, we studied the effects on the MFs from systematic errors in M_*, also shown in Figure 4.8. In Section 3.3.1 and Figure 3.8 we showed explicitly that systematic errors in the GSMF due to mass-to-luminosity ratios and photometric uncertainties are well represented by a shift in the M_*-axis of ± 0.15 dex. The effect of random errors in the baryonic MF is of the same order while for the gas MFs the propagated systematic errors in M_* have a negligible effect. Note that our MFs are complete only above a given mass limit, $\sim 3 \times 10^7$ $M_\odot$ for the GSMF, $\sim 10^7$ $M_\odot$ for the H_2 MF, and $\sim 10^8$ $M_\odot$ for the HI, cold gas, and baryonic MFs.

From the results summarised above we highlight the following conclusions:

- The HI, H_2, and cold gas MFs are mostly dominated by late-type galaxies. In general, we notice that our HI MF is in good agreement with previous determinations from blind surveys. Similarly the H_2 MF is consistent with previous determinations based on CO follow-up optically-selected samples. When we compare to the HIPASS and ATLAS 3D surveys for early-type galaxies, our HI MF is consistent with those observations. However, our H_2 MF for early-type galaxies is in tension at the low-mass side with the MF derived from the ATLAS 3D survey.

- Our "observational" MFs were deconvolved from random errors to obtain the intrinsic MFs. The effect of random errors is small at the low-mass end but larger at the high-mass end of our MFs. This is because the convolution depends on the logarithmic slope of the intrinsic MFs. Because the baryonic and stellar MFs are shallower at the massive-end the effects are relatively small, but the atomic, molecular and cold gas MFs have steeper slopes resulting in a larger effect.

- While for the stellar (and hence baryonic) MF systematic errors due to mass-to-light ratio uncertainties introduce a non-negligible effect, especially at the high-mass end, for the atomic, molecular and gas MFs the effects of systematics are small. We thus conclude that our determinations for the gas MFs are robust against systematic errors in the the M_* determination.

- We determined the $z \sim 0$ cosmic densities of HI, H_2, cold gas, stars and baryons locked in galaxies calculated from the respective MFs. Our results are in good agreement with previous determinations from different local censuses. Most of the atomic and molecular H gas is in late-type galaxies, $\sim 96\%$ of the mass density, while this fraction decreases to $\sim 70\%$ and $\sim 65\%$ for baryons and stars. We find that the fraction of HI and H_2 in galaxies with respect to the universal baryon fraction is respectively $\sim 1\%$ and $\sim 0.2\%$ while the respective fractions for mass in stars is $\sim 4\%$. Baryons in galaxies (the ionised and hot gas were not included) are $\sim 5.4\%$ of the universal baryon fraction.

- Based on the values reported in the literature for the local CSFR of star-forming (late-type) galaxies, we estimated the cosmic H_2 and total gas depletion times of late-type galaxies. These timescales, $t_{\mathrm{dep}}(H_2) \approx 1.3$ Gyr and $\bar{t}_{\mathrm{dep,L}}(\mathrm{gas}) \approx 10.14$ Gyr, respectively, imply that galaxies, on average, are inefficient in transforming their molecular gas into stars, and are inefficient to transform their atomic gas into molecular gas. The depletion time for the total neutral hydrogen is

145

$\bar{t}_{\mathrm{dep,L}}(\mathrm{H}) = \bar{t}_{\mathrm{dep,L}}(\mathrm{gas})/1.4 \sim 7.25$ Gyrs. On the other hand, the average cosmic SF timescale (the inverse of the cosmic sSFR) is $\bar{t}_{\mathrm{SF,L}} \approx 20.3$ Gyrs, which implies that the ratio $\bar{t}_{\mathrm{dep,L}}(\mathrm{H})/\bar{t}_{\mathrm{SF,L}} = 0.38$. This shows that the gas reservoir of late-type galaxies has not yet been dramatically consumed by star formation.

Here, we provided a statistical description for calculating any moment to characterise the gas-to-stellar mass correlations, the HI- and H_2-stellar mass bivariate distributions as well as all the respective MFs. One of our motivations for this work is to provide the community with a full self-consistent phenomenological description of the local galaxy population for various properties and divided into the two main morphological types in order to be confronted with theoretical results, both from semi-analytical models and cosmological hydrodynamical simulations. The next generation of sensitive radio telescopes will be able to survey large samples of extragalactic sources in HI and H_2 gas, something that is a common practice with current optical surveys. Thus, robust and unbiased bivariate distributions and MFs of HI and H_2 gas over large mass ranges will be routinely derived in the future along with the relationships of the gas contents with their optical/IR properties. Preparatory to that, and to pave the road to these surveys, studies based on radio follow-up observations of (relative small) optically-selected galaxy samples provide valuable information that can be used for the gas demographics of galaxies. In this work, we have exploited the results from many of these studies, and by means of the conditional (or bivariate) approach we were able to derive the abundances of local galaxies as traced by different baryonic components and separated into the two main groups of galaxies, early and late types.

The present work is the second paper of a series. In Chapter 2, we derived the CPDFs of HI and H_2 as a function of M_*, separately for early- and late-type galaxies, for an extensive compilation and homogenisation of radio data from the literature. In the present Chapter, we made extensive use of these data. In Chapter 6, we will use the MFs derived here to extend the galaxy-halo connection for different baryonic components, and we will show that not only the HI and H_2 MFs derived here are in good agreement with radio blind or optically-selected surveys but also with the observed galaxy spatial clustering as a function of HI gas mass.

Chapter 5

H I conditional distributions of central and satellite galaxies

ABSTRACT

In this Chapter we characterise the conditional distributions of the H I gas-to-stellar mass ratio, $R_{\mathrm{H\,I}}$, given the stellar mass, M_*, of local galaxies from $M_* \sim 10^7$ to 10^{12} $\mathrm{M_\odot}$ separated into centrals and satellites as well as into late- and early-type galaxies (LTGs and ETGs, respectively). To do so, we use 1) the homogeneous "eXtended GALEX Arecibo SDSS Survey", `xGASS`(Catinella et al., 2018), by re-estimating their detection limits and taking into account properly the remaining upper limits in our statistical analysis; and 2) the results from the large compilation of H I data reported in Chapter 2. We use the conditional distributions combined with the Galaxy Stellar Mass Function presented in Chapter 4 and 3 respectively to infer the bivariate $M_{\mathrm{H\,I}}$ and M_* distribution of all galaxies as well of the late-/early-type and central/satellite subsamples and their combinations. Satellites are on average less H I gas-rich than centrals at low and intermediate masses, more for the ETGs than for the LTGs, but at $M_* > 3 - 5 \times 10^{10}$ $\mathrm{M_\odot}$ the differences are negligible. The differences in the H I gas content are much larger when galaxies are separated into LTGs and ETGs than when the separation is into centrals and satellites. Our empirical H I Mass Function is strongly dominated by central galaxies at all masses. The empirically constrained bivariate $M_{\mathrm{H\,I}}$ and M_* distributions presented here can be used to compare and constrain theoretical predictions as well as to generate galaxy mock catalogs.

5.1 Introduction

The evolution of galaxies implies the interplay of many complex processes. Among them: gas cooling within dark matter haloes, transformation of the cool atomic hydrogen (H I) gas into cold dense molecular hydrogen (H_2) clouds, the formation of stars in the densest regions of these clouds, and the ulterior feedback that the stars and their explosions exert on the interstellar medium (see e.g., Mo et al., 2010b). Therefore, the amounts of H I and H_2 gas with respect to the stellar mass, morpho-

logical type, colors, and other galaxy properties, are crucial for understanding the evolutionary stage of local galaxies (Lagos et al., 2011, 2014). It is also well-known that the environment, in particular whether a galaxy is central or satellite, plays a role in the evolution of galaxies, so that information on the gas fractions of galaxies as a function of environment is also relevant (e.g., Stevens et al., 2019, and more references therein).

Although H I gas is the dominant component in the interstellar medium of local galaxies, its detection is not easy because of its weak 21-cm emission line. Great efforts have been made to build large H I surveys as the H I Parkes All-Sky Survey (HIPASS; Meyer et al., 2004) and Arecibo Fast Legacy ALFA Survey (ALFALFA; Giovanelli et al., 2005; Haynes et al., 2011). However, these blind radio surveys are not yet as deep and do not cover such large areas as the optical/infrared extragalactic surveys, introducing this several sample selection effects. Thus, the inferred H I gas scaling relations, as well as other correlations and H I spatial distributions, result biased (c.f. Meyer et al., 2007; Haynes et al., 2011; Huang et al., 2012c; Papastergis et al., 2013; Maddox et al., 2015; Guo et al., 2017; Calette et al., 2018). As an attempt to overcome the strong selection effects of the blind H I radio surveys, "well-controlled" H I samples were constructed by means of radio follow-up observations of optically selected galaxy samples or by cross-correlating blind radio surveys with UV/optical/infrared surveys (e.g., Wei et al., 2010a; Catinella et al., 2013, 2018; Papastergis et al., 2012; Kannappan et al., 2013; Boselli et al., 2014a; Eckert et al., 2015; Stark et al., 2016; van Driel et al., 2016; Masters et al., 2019). These samples were designed for a variety of scientific goals, in such a way that they are diverse and heterogeneous, covering different mass ranges, distances, and H I flux detection limits, and commonly they are far from completeness in stellar mass.

In Chapter 2 (and with updates in Chapter 4) we undertook the task of compiling and homogenizing from the literature as much as possible galaxy samples in the spirit of the ones listed above (including most of them) with the additional requirement of information on the galaxy morphology because the H I gas content of galaxies strongly depends on morphology, hence it is more appropriate to analyze them separately. We have taken into account the reported upper limits for the radio non-detections, and after homogenizing and correcting some of them for distance limitations, we have applied a survival analysis to determine gas correlations. As a result, we were able to constrain not only the mean $M_{\rm H I}$–M_* relation for late- and early-type galaxies (LTG and ETG, respectively) down to $M_* \sim 3 \times 10^7$ M$_\odot$, but the respective full conditional probability density distribution functions (PDFs) of $M_{\rm H I}$ given M_*, $P(M_{\rm H I}|M_*)$. From these PDFs, one can calculate any moment of the distributions,

in particular the standard deviation and percentiles around the mean relation.

In Chapter 4 we used the well-constrained Galaxy Stellar Mass Function (GSMF) for all, late- and early-type galaxies down to $\sim 3 \times 10^7$ M$_\odot$ presented there, and combined them with the H I conditional PDFs given M_* to generate the bivariate or joint M_* and $M_{\rm HI}$ distribution function. By projecting this bivariate distribution into the H I axe, we obtained the H I MFs, for LTGs and ETGs, as well as for all galaxies. We have shown that our empirical H I MF (corresponding to a volume-limited sample complete above $M_* \sim 3 \times 10^7$ M$_\odot$) agrees well with those measured from blind radio surveys. This shows that the H I mass function is much less affected by the selection effects of the blind radio surveys than the H I gas scaling relations.

In Chapter 2 we showed that the conditional PDFs of the H I-to-stellar mass ratio, $R_{\rm HI} \equiv M_{\rm HI}/M_*$, given M_* can be well described by a Schechter-type function for LTGs (see also Lemonias et al., 2013) and a (broken) Schechter-type function plus a top-hat function for ETGs, having the latter significantly lower values of H I gas content than the former (see Figure 5.1 below). These distributions and the main relations calculated from them do not make a difference between central and satellite galaxies. Though it is not clear whether the H I gas fraction of galaxies correlates directly or not with the large-scale environment (see for a discussion Chapter 2, and the references therein), at the level of central and satellite galaxies, the latter seem to have lower H I gas contents at a given stellar mass than the former (e.g., Stark et al., 2016; Brown et al., 2016, but see Lu et al., 2020).

In this Chapter we will introduce a correction to our empirical H I conditional PDFs for LTGs and ETGs in such a way that they can be separated into central and satellite galaxies. For this, we will use the recent H I observational survey xGASS (eXtended GALEX Arecibo SDSS Survey Catinella et al., 2013, 2018). xGASS is an homogeneously constructed H I, UV, and optical galaxy sample with well defined limits in $R_{\rm HI}$, M_*, and volume. Since this survey was constructed from SDSS, most of the galaxies can be separated into centrals and satellites making use of the Yang et al. (2007, 2012) halo-based group definition applied to SDSS. Thus, we will be able to explore the differences in H I content between centrals and satellites. We also will explore how much may deviate the H I-to-stellar mass relation and the full $M_{\rm HI}$ distributions estimated from xGASS from our empirical determinations in Chapter 2. In any case, have in mind that for these determinations, one of the relevant samples we have used for $M_* > 10^{10}$ M$_\odot$ was GASS. The xGASS survey is just the GASS survey with an extension down to 10^9 M$_\odot$.

The Chapter is organised as follows. In section 5.2 we describe the xGASS survey and our processing. Section 5.3 presents the results of our statistical analysis of
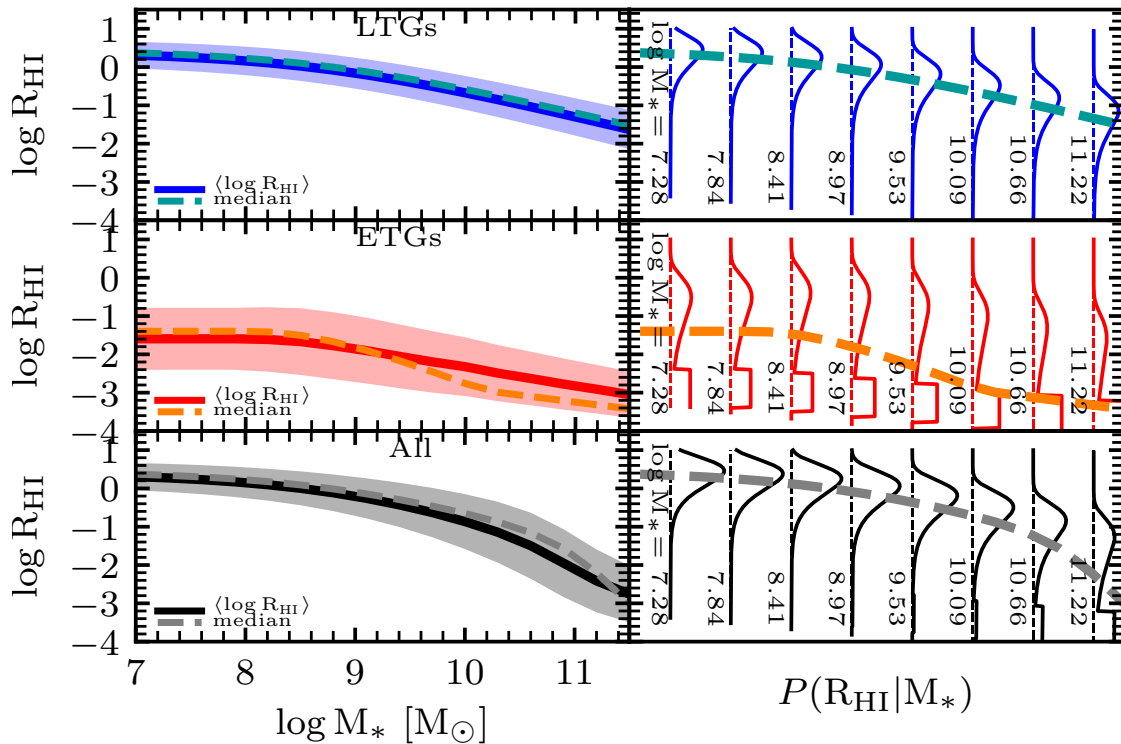
Figure 5.1: First and second moments and the full H I gas conditional distributions as a function of M_* from Chapter 2 and updated in Chapter 4. *Left panels:* Logarithmic mean R_{HI}–M_* relation, $R_{HI} \equiv M_{HI}/M_*$, and its standard deviation (thick solid line and shaded area, respectively) for late-type, early-type, and all galaxies, from top to bottom. The dashed lines show again the relations but using medians instead of logarithmic means. *Right panels:* The respective full R_{HI} conditional PDFs at the stellar masses indicated in the panels. The PDFs are plotted in linear scale.

xGASS. We present the H I-to-stellar mass relations for LTGs and ETGs separated into centrals and satellites, as well as, the respective full H I conditional distributions and the join fit of analytic functions to them. In Section 5.4 we use the xGASS H I conditional distributions to separate the distributions constrained in Chapters 2 and 4 into centrals and satellites. Section 5.5 is devoted to discuss our results. Finally, in Section 5.6 we present a summary of the Chapter and the conclusions.

5.2   Analysis of the xGASS survey

Before discussing the xGASS survey, in Figure 5.1 we present a summary of the results from Chapter 2 and updated in Chapter 4. The R_{HI} conditional PDFs given M_*, $P(R_{HI}|M_*)$, for LTGs and ETGs are shown in the right upper and medium panels, respectively. For LTGs, the PDFs are described by a Schechter-like function, while for ETGs, a Schechter-like + top-hat function describes better the PDFs. The left panels show the corresponding logarithmic mean values and standard deviations (first and second moments of the R_{HI} distributions) with the solid lines surrounded by the shaded regions, respectively. The dashed lines are the corresponding medians also

reproduced in the right panels. While for LTGs, both the mean and median R_{HI}–M_* relations are similar, for ETGs, they differ, specially at the high-mass side. The bottom panels show the resulting R_{HI} conditional PDFs for all galaxies as well as the respective first and second moments. We infer the R_{HI} conditional distribution for all galaxies by using the fractions of ETGs as a function of M_* from the Sloan Digital Sky Survey (SDSS) based on the Huertas-Company et al. (2011) morphological classification, corrected for volume completeness (see Chapter 3 for details). As mentioned in the Introduction, in Chapters 2 and 3 we do not separate the R_{HI} conditional PDFs into centrals and satellites. The aim of this Chapter is to use the xGASS sample for calculating the ratios of central and satellite to total R_{HI} conditional PDFs for both late- and early-type galaxies, and to use these results to attain the separation into centrals and satellites from our empirical LTG and ETG R_{HI} PDFs.

The xGASS (Catinella et al., 2018) is an R_{HI}-limited census of 1179 galaxies selected by redshift and M_* in the ranges $0.01 \leq z \leq 0.05$ and $10^9 \mathrm{M}_\odot \leq M_* \leq 10^{11.5}\mathrm{M}_\odot$, respectively. The sample galaxies were drawn from the intersection of SDSS DR7 (Abazajian et al., 2009), GALEX (Martin et al., 2005) and projected ALFALFA footprints (Haynes et al., 2011). The xGASS consists of two samples : (1) GASS (Catinella et al., 2010, 2012, 2013), a sample of galaxies with $M_* > 10^{10}$ M$_\odot$ and redshift $0.025 \leq z \leq 0.05$, and (2) low-mass extension of GASS (hereafter low-GASS Catinella et al., 2018), a sample of galaxies with stellar masses in the range $10^9 \mathrm{M}_\odot \leq M_* \leq 10^{10.2}\mathrm{M}_\odot$ and redshift $0.01 \leq z \leq 0.02$. Both samples were constructed in such a way that the stellar mass distribution of the targets is roughly flat. The xGASS survey is the most complete H I observational study of a local optically-based representative galaxy sample to date.

In the xGASS sample the H I mass is obtained from the H I observations of ALFALFA $\alpha.40$ or the Cornell H I digital archive (Springob et al., 2005). For galaxies with no H I information, observations were performed using Arecibo Radio Telescope with the strategy of observing the targets until detected or until a limit of a few percent in R_{HI} ratio is reached. Such detection limits for each sample are:

- GASS: $R_{\mathrm{HI}} > 0.015$ for galaxies with $M_* > 10^{10.5}\mathrm{M}_\odot$ and a constant H I mass limit of $M_{\mathrm{HI}} = 10^{8.7}$ M$_\odot$ for galaxies with lower stellar masses.

- low-mass extension of GASS (hereafter GASS-low): $R_{\mathrm{HI}} > 0.02$ for galaxies with $M_* > 10^{9.7}\mathrm{M}_\odot$ and constant H I mass limit of $M_{\mathrm{HI}} = 10^8$ M$_\odot$ for lower mass galaxies.

The detection limits in R_{HI} are established mainly by the telescope sensitivity, integration time, and the redshift range of the surveys.

5.2.1 Morphology and central/satellite designations for xGASS galaxies

At fixed M_*, the gas content in galaxies varies significantly with morphology (e.g., Kannappan et al., 2013; Boselli et al., 2014b; Calette et al., 2018). Thus, we introduce a morphological characterization for xGASS galaxies.

Here, we use the Huertas-Company et al. (2011) automated morphological classification for $\sim$700 000 galaxies from the SDSS DR7 spectroscopic sample, where each galaxy has a probability of being elliptical, S0, Sab and Scd by means of support vector machines (SVM) method and the Fukugita et al. (2007) sample as a training set. On the other hand, Meert et al. (2015) calibrated Huertas-Company et al. (2011) probabilities to T-types using a simple linear model given by,

$$T = -4.6 \cdot P(Ell) - 2.4 \cdot P(S0) + 2.5 \cdot P(Sab) + 6.1 \cdot P(Scd) \qquad (5.2.1)$$

The latter was constrained using the visual classification of Nair & Abraham (2010b) by a linear regression. Hence, we assign T-type to xGASS galaxies with eq. (5.2.1) and the probability classification from Huertas-Company et al. (2011). Of the 1179 galaxies in the xGASS sample we find that 1150 are in the Huertas-Company et al. (2011) morphology catalog.

We separate xGASS galaxies into two broad morphological groups: LTGs and ETGs. We consider ETGs as those galaxies with $T < 0.5$ and LTGs as those with $T \geq 0.5$ following Meert et al. (2015). The above corresponds respectively to $S0$ or earlier and Sab or later morphologies, see their equation (8) for details.

To segregate galaxies into centrals and satellites we use the xGASS flag env_code_B defined as[1]:

$$\text{env\_code\_B} = \begin{cases} 0: & \text{satellite} \\ 1: & \text{isolated central} \\ 2: & \text{group central} \\ -1: & \text{not in group catalog} \end{cases}$$

We consider centrals those galaxies with env_code_B=1 or 2. The term isolated central does not imply what typically is known in the literature as an isolated environment but it refers to the presence of only one galaxy within the halo. Satellites are those with env_code_B=0. As described in Janowiecki et al. (2017), for determining whether a galaxy is central or satellite in xGASS, the authors used the Yang et al. (2007) halo-

---

[1]xGASS data description: `https://xgass.icrar.org/assets/data/xGASS_representative_sample.readme`

based group catalog updated to the SDSS DR7. For `xGASS`, the 'modelB' group catalog was adopted, and cases of "galaxy shredding" and false pairs have been resolved by visual inspection (see details in Janowiecki et al., 2017).

Fortunately, only a small fraction of `xGASS` galaxies, 2%, are not in the Yang et al. (2007) 'modelB' catalog or suffer of galaxy shredding and false pairs. Approximately 30% of `xGASS` galaxies are classified as satellites in groups, $\sim 50\%$ as isolated centrals, and $\sim 20\%$ as centrals (the most massive member) in groups.

The central/satellite designation adopted for the `xGASS` survey has been used in several works for studying the effects of environment on the gas content of galaxies (e.g., Janowiecki et al., 2017, 2020; Stevens et al., 2019; Cortese et al., 2020; Watts et al., 2020). Nevertheless, it should be stressed that galaxy group finders like the Yang et al. (2005, 2007) halo-based may suffer of membership allocation and central/satellite designation errors. In §§5.5.2, we discuss on this caveat and how it can affect the results obtained in this Chapter.

The final sample of `xGASS` galaxies with morphology and central/satellite classifications accounts for 1134 objects.

In panels (a) and (d) of Figure 5.2, we present these galaxies in the $R_{\mathrm{H\,I}}$–M_* plane separated into LTGs and ETGs, respectively. In each panel central and satellite galaxies are plotted with open circles and crosses, respectively, and upper limits are shown with downward arrows. As seen, the number of galaxies with upper limits is significant: 55% for ETGs and 17% for LTGs. The dot-dashed and dashed lines show the imposed detection limit in the `GASS` and `GASS`-low samples, respectively. Most of the upper limits pile up close to these lines. However, since galaxies are at different distances the distribution of the upper limits is somewhat scattered. In the same panels, we reproduce the logarithmic means of LTGs and ETGs obtained in Chapters 2 and 4. Their corresponding $R_{\mathrm{H\,I}}$ conditional distributions at different stellar masses are shown respectively in the panels (b) and (e). In these panels, we also reproduce the `xGASS` detection limits. Clearly the empirical distribution of $R_{\mathrm{H\,I}}$ is truncated by the `xGASS` detection limits. This truncation is particularly abrupt for ETGs, which are above the first moments of the empirical $R_{\mathrm{H\,I}}$ PDFs (the red solid line in panel d).

An upper limit in H I mass is reported when a galaxy in a given survey has not been detected in the 21-cm line for the defined integration time and above a given signal-to-noise ratio. The H I mass upper limit is calculated using the respective H I flux detection limit and the distance to the galaxy, $M_{\mathrm{HI}}^{\mathrm{u.l.}} \propto D(z)^2$. When inferring any correlation or probability distribution from $M_{\mathrm{H\,I}}$, it is *mandatory* to account for upper limits. In §§5.2.3 we describe the survival analysis we follow to do so. In addition, it is important to note that the `xGASS` upper limits are quite high, specially for ETGs.
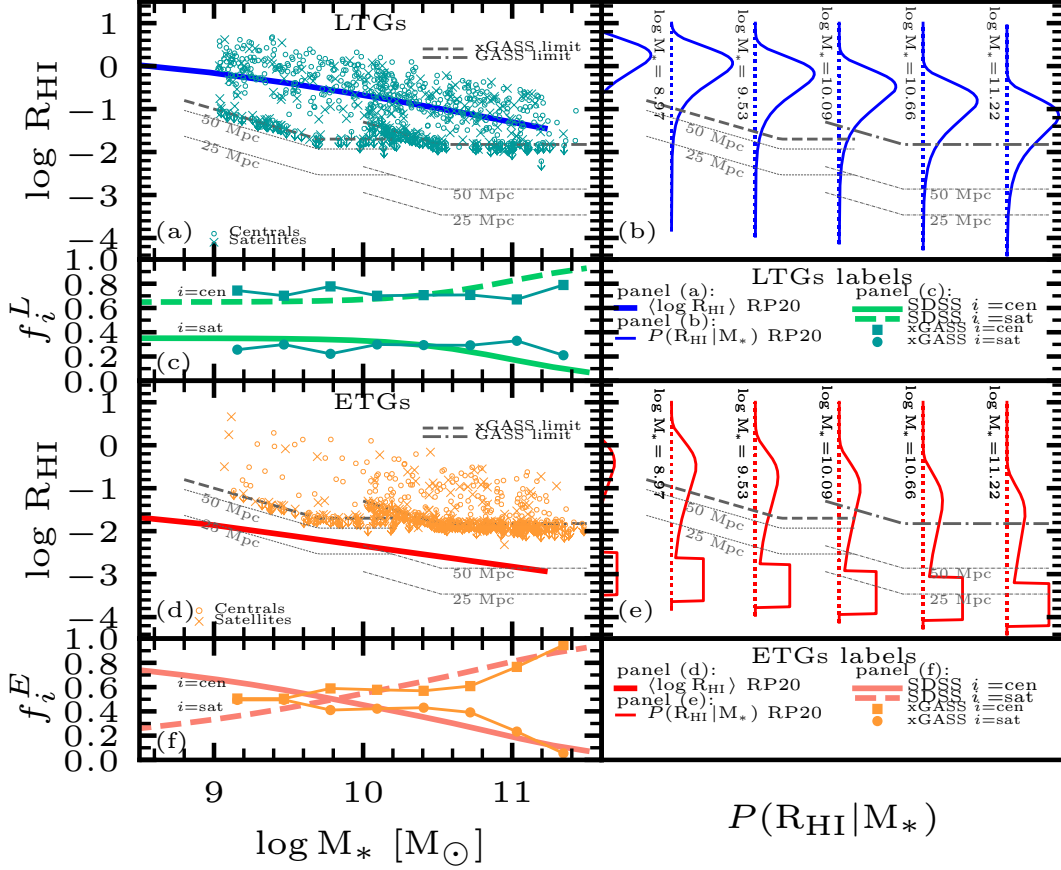
Figure 5.2: Presentation of the `xGASS` sample. *Panel (a):* LTGs in the $\log R_{\mathrm{HI}}$-$\log M_*$ diagram, with centrals and satellites plotted as empty circles and crosses, respectively. The downward arrows indicate the reported upper limits for non detected galaxies in radio. Dot-dashed and dashed lines show the imposed limit detection in the `GASS` and the low-`GASS` samples, respectively. We reproduce the logarithmic mean of LTGs obtained in Chapters 2 and 4 with blue solid line. *Panel (b):* Chapter 4 LTGs R_{HI} conditional distributions at different stellar masses. *Panel (c):* Fraction of LTGs that are satellites, circles, or centrals, squares, as a function of M_*. The respective fractions as inferred from SDSS DR7 using Huertas-Company et al. (2011) morphological classification and Yang et al. (2012) central/satellite division are plotted with the dashed and solid lines, respectively. *Panel (d):* Same as panel (a) but for ETGs. *Panel (e):* Same as panel (b) but for ETGs. *Panel (f):* Same as panel (e) but for ETGs.

This is due to the large distances in this survey, in particular for `GASS`. In fact, many radio observations of closer galaxies detect H I below the `xGASS` detection limits. On the other hand, the H I detection limits of other closer galaxy samples, after taking into account the differences in the observational and instrumental settings, result in much lower values than those from `xGASS`, in particular for the `GASS` sample. Thus, the upper limits from `xGASS` are strongly biased by a distance selection effect. Following Chapter 2, in §§5.2.2 we attempt to correct for this bias in the upper limits.

Panels (c) and (f) of Figure 5.2 present respectively the `xGASS` fraction of LTGs and ETGs that are satellites, circles, or centrals, squares, as a function of M_*. In the same panels, the solid lines correspond to fits to the satellite fractions for LTGs and ETGs from the Yang et al. (2012) SDSS DR7 galaxy group catalog (the dashed lines are the respective central fractions and they are by definition the complements of the solid lines; see Appendix E). At this point, it is important to ask ourselves if `xGASS` suffers of selection effects that could bias the sample by morphology (for the morphological classification adopted here, i.e., Huertas-Company et al., 2011) or by environment. A bias in the morphology is not relevant when the inferred $R_{\mathrm{H\,I}}$–M_* relations and $R_{\mathrm{H\,I}}$ distributions are for LTGs or ETGs separately. However, the possible bias is expected to affect the relations and distributions for all, central, and satellite galaxies when averaging among LTGs and ETGs.

In Figure E.1 in Appendix E, we compare the fractions of satellites and ETGs from `xGASS` as a function of M_* with those measured from SDSS DR7 (panels a and d, respectively). As seen, the `xGASS` fraction of satellites as a function of M_* roughly agree with that from the SDSS DR7 (the fraction of centrals is the complement). However, this is not the case for the fraction of ETGs (the fraction of LTGs is the complement): `xGASS` selects systematically a higher fraction of ETGs than SDSS up to $M_* \sim 10^{11}$ M$_\odot$. Obviously, the differences remain when considering only central or satellite galaxies, but they are larger for satellites (compare panels b and c), and for $M_* \gtrsim 10^{11}$ M$_\odot$, the difference even inverts. Note that the flat distribution in mass of `xGASS` is not an issue in Figure 5.2 given that the comparisons between fractions are at a given M_*.

For the inferences in Section 5.3 of the $R_{\mathrm{H\,I}}$–M_* relations and $R_{\mathrm{H\,I}}$ distributions given M_* corresponding to all galaxies (LTGs + ETGs), to all centrals (LTGs + ETGs) and to all satellites (LTGs + ETGs), we introduce weights for the `xGASS` galaxies in order to be consistent with the observed fractions of centrals and satellites that are ETGs as a function of M_* from the SDSS DR7. The weighting procedure is described in Appendix E.

5.2.2 Correcting the H I upper limits

In Chapter 2 it was found that the reported $M_{\rm HI}$ upper limits for ETGs in the `GASS` sample are above by at least 1.5 dex than those from the ETG ATLAS$^{\rm 3D}$ survey (Cappellari et al., 2011; Serra et al., 2012, and also significantly above than those from the Herschel Reference Survey, HRS, Boselli et al., 2010, 2014a).[2] The reason for this is that `GASS` galaxies are at larger distances ($100 \leq D[{\rm Mpc}] \leq 200$, with a median of 165 Mpc) than the ETG from ATLAS$^{\rm 3D}$ (median of 25 Mpc). The larger the distance the harder is to detect H I in low-$M_{\rm HI}$ galaxies, hence the H I detection limits are higher. In an attempt to homogenize the values of the upper limits in $R_{\rm HI}$ of these two surveys, in Chapter 2 we decreased the reported ETG upper limits from `GASS` by $(D_i(z)/\overline{D}_{\rm ATLAS3D})^2$, being D_i the luminosity distance of each `GASS` ETG and $\overline{D}_{\rm ATLAS3D}$ the median luminosity distance of ATLAS$^{\rm 3D}$. Moreover, we considered the fraction of galaxies in ATLAS$^{\rm 3D}$ with H I *detections* below the `GASS` $R_{\rm HI}$ detection limit ($\sim 25\%$) by assigning $R_{\rm HI}$ values (detections) to 25% of the non-detected `GASS` ETGs; the $R_{\rm HI}$ values were drawn from a uniform distribution in between the `GASS` and ATLAS$^{\rm 3D}$ $R_{\rm HI}$ detection limits (corrected to the same integration time and signal-to-noise ratio used in ALFALFA). We used mock catalogs to validate this methodology. Performing a similar analysis to `GASS`-low will require information of a survey like the ATLAS$^{\rm 3D}$. Unfortunately, this survey extends only down to stellar masses slightly smaller than $\sim 10^{10}$ M$_\odot$.

In the case of LTGs, since they are typically gas rich, most of them are detected in `GASS` in spite of their relatively shallow H I detection limit. On the other hand, for LTGs there is not a closer and homogeneous sample similar to ATLAS$^{\rm 3D}$. Thus, in Chapter 2, we did not attempt to correct the upper limits of LTGs from `GASS` by the distance effect. For `GASS`-low, the fraction of radio detected LTGs from closer samples below the `GASS`-low detection limit is slightly larger than in `GASS`. The overall fraction of upper limits for LTGs in `xGASS` is 17%. Following the above argument for ETGs, it would be desirable to attempt to correct by distance the upper limits of LTGs, too.

As mentioned above, there are not close samples, as ATLAS$^{\rm 3D}$, with more or less defined detection limits in $R_{\rm HI}$ for $M_* < 10^{10}$ M$_\odot$, both for early- and late-type galaxies. However, we can use the empirically constrained $R_{\rm HI}$ distributions in Chapter 2 (and updated in Chapter 4) to estimate corrections for the `GASS`-low upper limits due to their bias by distance. Even more, to obtain homogeneous corrections, we decided to use these empirically-constructed distributions to (re)estimate corrections also for the `GASS` ETGs as well as for LTGs. For `GASS` ETGs, the corrections to be

---

[2]The 1.5 dex was obtained after taking into account the instrumental sensitivity and the integration time differences between both surveys.

obtained here are actually very similar to those as in Chapter 2.

Following the discussion above, in Appendix F we describe in detail our procedure to correct the upper limits of ETGs and LTGs for xGASS.

5.2.3   Statistical analysis including H I upper limits

In order to estimate from xGASS the R_{HI}–M_* relations separated into central and satellite galaxies or, even more, the full R_{HI} conditional PDFs given M_*, as in Chapter 2, the upper limits should be taken into account. In observational Astrophysics, it is often that we are interested on particular astronomical objects (e.g. stars, galaxies, etc) and in order to create samples to study them, we set a selection criteria based on a property, P_1, to construct such observational samples (for example stellar mass or luminosity). But there are situations when we are also interested in another property, P_2 (for example H I content). Nevertheless, due to instrumental limitations we cannot always measure the property P_2 in all objects, instead we assign upper limits or "censored" data values. In such situation it is necessary to build a parent sample based on a well studied property P_1 and then examine for the property of interest P_2 from property P_1. The above description is exactly the case for the xGASS sample, in which $P_1 = M_*$ and $P_2 = M_{\mathrm{HI}}$.

To use both detections and upper limits from xGASS, in this work we rely on Kaplan-Meier (KM) non-parametric estimator (Kaplan & Meier, 1958a) specifically developed for the analysis of censored data in clinical research, but properly adapted to astronomical data by Feigelson & Nelson (1985). For a given sample, the KM estimator allows us to obtain the cumulative distribution function (CDF) when including the censored data. We construct the R_{HI} CDFs at different stellar mass bins. After the corrections applied to the ETG upper limits (see above), the minimum R_{HI} values (censored data) used in the KM estimator are around $-3.0 < \log R_{\mathrm{HI}} < -3.5$, and the CDFs at these values start with fractions typically of 0.3–0.4. This means that around 30-40% of ETGs are R_{HI} upper limits. As mentioned in the footnote of Appendix F.1, in Chapter 2 we assigned real values (detections) to these galaxies by assuming they follow a top-hat function of width ~ 1 dex below the minimum upper limit value of the given mass bin. Our main argument was that even quiescent ETGs should have H I gas fractions larger than a few 10^{-5}, taking into account the stellar mass loss and some minimum amount of cosmic accretion.

To compute the CDFs with the KM estimator we group the xGASS data in stellar mass bins of width $\Delta \log M_* = 0.31$ dex at the interval $9 \leq \log M_* \leq 11.5$.
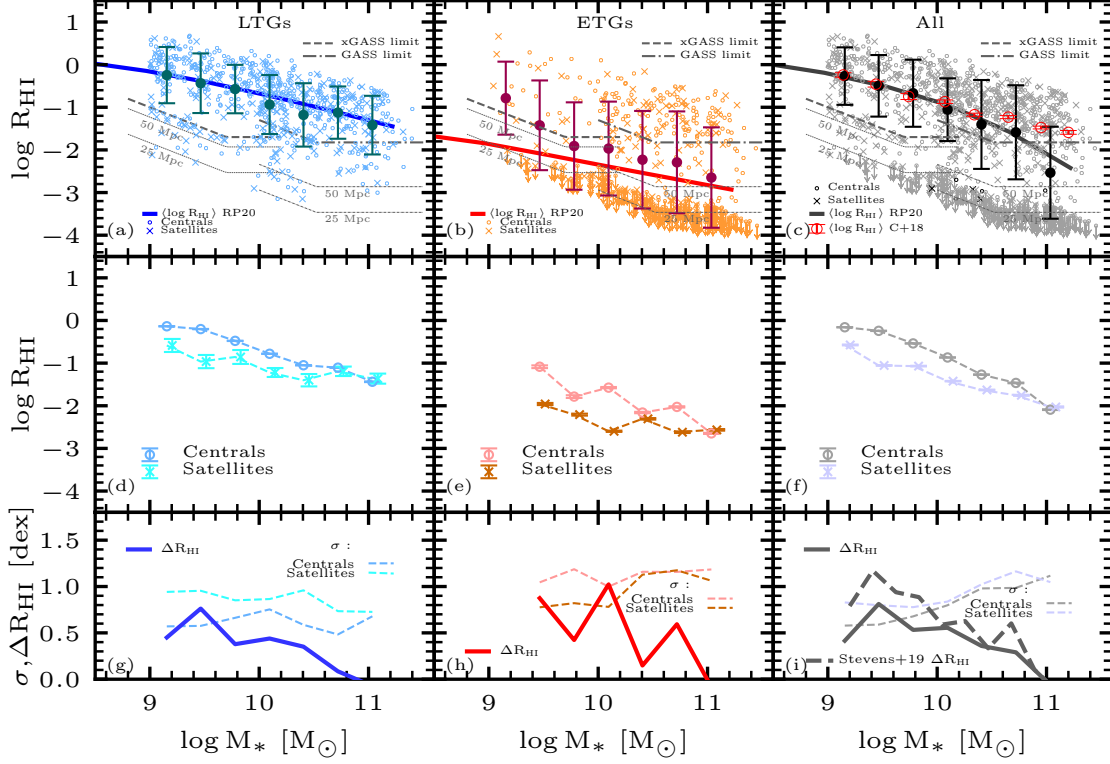
Figure 5.3: *Upper panels:* `xGASS` galaxies in the $\log R_{\mathrm{H\,I}} - \log M_*$ diagram, as in Figure 5.2 but after corrections applied to the upper limits (see text). The dots with error bars are the logarithmic means and standard deviations in M_* bins obtained with the KM estimator for taking into account upper limits. The solid lines show the mean $R_{\mathrm{H\,I}}$–M_* relations from Chapter 4. In panel (c), the red circles are the logarithmic means as reported in Catinella et al. (2018). *Medium panels:* Logarithmic means and their error on the mean in M_* bins obtained with the KM estimator for the subpopulations of central (open circles with error bars) and satellite (crosses with error bars) galaxies, for late-type, early-type, and all galaxies from left to right. *Lower panels:* Second moments of the $\log R_{\mathrm{H\,I}}$ distributions from the KM estimator for the subpopulations of central and satellite galaxies (dashed lines) showed in the upper panels. The solid lines are the relative differences between the means of central and satellite subpopulations showed in the medium panels. The long-dashed line in panel (i) corresponds to the relative differences between the medians of centrals and satellites as reported for `xGASS` in Stevens et al. (2019).

5.3 Results from xGASS

5.3.1 Correlations for all, central, and satellite galaxies

In the upper panels of Figure 5.3 we plot again the xGASS data as in Figure 5.2 but after applying to the upper limits the procedure described in §§5.2.2; we added a third panel showing the data for the whole sample. For each M_* bin of width $\Delta \log M_* = 0.31$ dex, we use the procedure based on the KM estimator described in §§5.2.3 to calculate the mean logarithmic value of R_{HI} and the standard deviation at each bin. The results are plotted with circles and error bars. For comparison, the thick solid line in each panel is the respective logarithmic mean relation as obtained in Chapter 4 and also reproduced in Figures 5.1 and 5.2 above. For LTGs, xGASS is in very good agreement with our empirical relation from Chapter 4. In the case of ETGs, the averages of xGASS galaxies (after correcting the upper limits by the distance bias) are slightly above than the corresponding relation from Chapter 4 but within the standard deviations. Note that these upper limits lie now around the GASS and low-GASS detection limits shifted to the distance of 25 Mpc.

In the right panel of Figure 5.3, corresponding to all galaxies, we reproduce the logarithmic mean R_{HI} values reported by Catinella et al. (2018), red open circles. These authors calculated the means (i) setting the H I mass of non-detections to their upper limit values (this leads to overestimate the mean), and (ii) applying weights to correct for the stellar mass bias of the sample, that is, to make the sample mass complete in volume. Regarding the latter item, it is not expected to be relevant for the means calculated in small mass bins since the weights are roughly the same for similar masses. As seen, at low and intermediate masses our means roughly agree with those from Catinella et al. (2018) but at the largest masses, where ETGs dominate, our means are lower than those reported by these authors. The above is due to the special treatment we applied to adequately include the upper limits of ETGs. Recall that we also weighted xGASS galaxies by morphology and environment to agree with the SDSS DR7 fractions as a function of M_*, see §§5.2.1. The weights correct mainly the excess of ETGs in xGASS with respect to SDSS up to $M_* \sim 10^{11}$ M$_\odot$ and the defect at larger masses (the latter specially applies for satellites), see Figure E.1. Therefore, the average values plotted in Figure 5.3 for all galaxies are weighted towards LTGs up to $M_* \sim 10^{11}$ M$_\odot$ and against them at higher masses.

The medium panels of Figure 5.3 show the logarithmic mean values of R_{HI} and their errors of the mean, this time for central and satellite galaxies separately. Centrals have on average slightly higher H I gas fractions than the average. For satellites, the
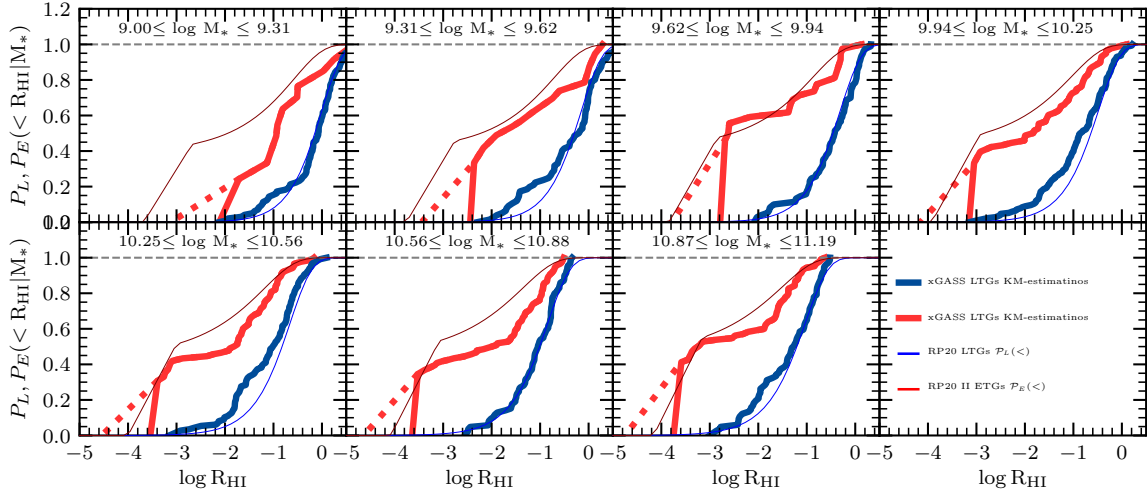
Figure 5.4: Cumulative histograms of LTG (blue lines) and ETG (red lines) H I conditional distributions (CDFs) at different M_* bins from the processed `xGASS` sample. For comparison, fits to the respective CDFs from Chapter 2 are shown with thin lines. For a correct comparison, these fits were averaged within the mass ranges of the bins.

differences are more pronounced especially towards lower stellar masses. Overall, centrals have higher H I gas contents than satellites, in particular at lower masses.

In the lower panels, we plot the corresponding logarithmic standard deviations for centrals and satellites at each mass bin for LTG, ETGs, and all galaxies. The population of ETGs presents larger scatters around the $R_{H I}$–M_* relations for centrals and satellites than LTGs. In each of the lower panels of Figure 5.3, we plot also the relative differences between the corresponding central and satellite means, $\Delta R_{H I} \equiv \langle \log R_{H I,cen} \rangle - \langle \log R_{H I,sat} \rangle$ (thick solid lines), plotted in the medium panels. As seen, these differences tend to be smaller than the corresponding standard deviations, both for LTGs and ETGs, specially at larger masses. The above implies that the subsamples of central and satellite galaxies tend to be mixed among them within their standard deviations. Nonetheless, on average, satellite galaxies have lower H I gas contents than centrals, specially at low masses. Finally, in the panel (i) corresponding to all galaxies, we reproduce the relative differences between the central and satellite medians reported in Stevens et al. (2019) for `xGASS` (long-dashed line). In spite that they measure medians and we logarithmic means and that they set non-detections to their upper limit values, the agreement is reasonable.

5.3.2 Conditional H I distributions for all, central, and satellite galaxies

Firstly, in Figure 5.4 we compare the $R_{\mathrm{H\,I}}$ conditional CDFs of late- and early-type galaxies from the processed `xGASS` sample (thick solid lines) with those inferred empirically in Chapters 2 and 4 (thin solid lines). The fits were averaged within the width of the M_* bin. The cumulative distributions for `xGASS` ETGs start at fractions around $0.3 - 0.4$. These are the fractions of the remaining upper limits after our corrections, see §§5.2.2 and 5.2.3. If we proceed as in Chapter 2, then to the upper limits of ETGs should be assigned $R_{\mathrm{H\,I}}$ values following a top-hat function of width ~ 1 dex below the lowest upper limit value in each mass bin (see above). This is shown in Figure 5.4 with the dotted lines. The `xGASS` H I conditional CDFs for LTGs agree well with those found in Chapter 2. For ETGs, the CDFs from `xGASS` tend to be somewhat shifted to higher $R_{\mathrm{H\,I}}$ values than those determined previously in Chapter 2 from an extensive data compilation from the literature.[3] These differences are seen also in the respective logarithmic mean values plotted in Figure 5.3.

Figure 5.5 presents the $R_{\mathrm{H\,I}}$ conditional CDFs in different M_* bins calculated as described in §§5.2.3 for the whole `xGASS` sample (black lines), and for only centrals (dark grey lines) and satellites (light gray lines), that is, $P^i(> R_{\mathrm{H\,I}}|M_*)$, where i refers to all, central or satellite, respectively. The lower M_*, the larger the difference in the distributions between central and satellite galaxies, with the latter being shifted to the low-$R_{\mathrm{H\,I}}$ side. Recall that for calculating these distributions, the `xGASS` sample has been weighted by morphology and environment to agree with the SDSS DR7 fractions as a function of M_*, see §§5.2.1 and Appendix E. The main bias of `xGASS` galaxies is actually by morphology; the bias by environment is small and mainly due to the former.

Figures 5.6 and 5.7 are as Figure 5.5 but now for LTGs and ETGs from `xGASS`, respectively, that is, these figures show the $R_{\mathrm{H\,I}}$ conditional CDFs $P^i_j(> R_{\mathrm{H\,I}}|M_*)$, where i refers to all, centrals or satellites, and j refers to LTG or ETG.

---

[3]Recall that in Chapter 2, for inferring the H I conditional distributions, (i) were used not only the `GASS` survey but other samples, and (ii) for converting to detections a fraction of ETG `GASS` upper limits, a uniform $R_{\mathrm{H\,I}}$ distribution was used while here the empirical $R_{\mathrm{H\,I}}$ distributions for ETGs constrained in Chapter 2 were used, see §§F.1.
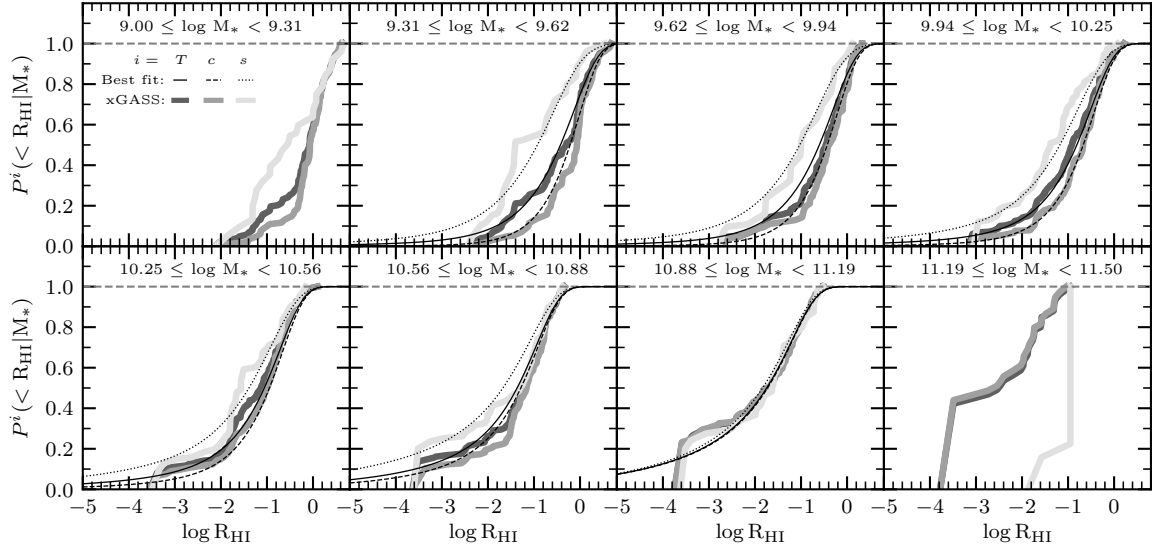
Figure 5.5: Cumulative histograms of the H I conditional distributions (CDFs) at different M_* bins from the processed `xGASS` sample of all galaxies and for only centrals and satellites, see color notation in the first panel. The solid, dashed, and dotted lines are our best joint fits to the different subpopulations shown in this Figure and in Figures 5.6 and 5.7, see text.
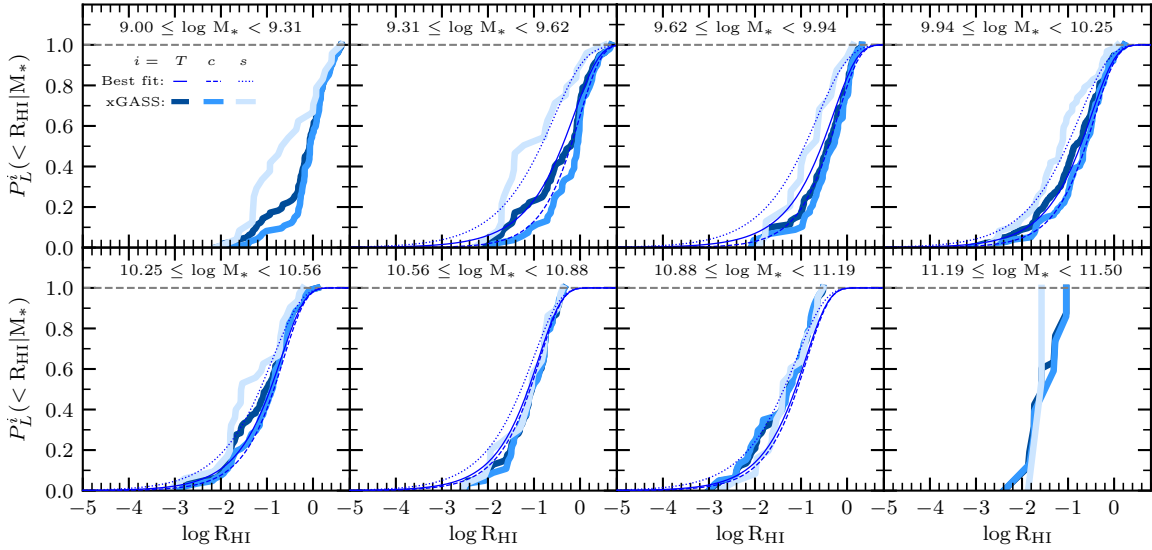

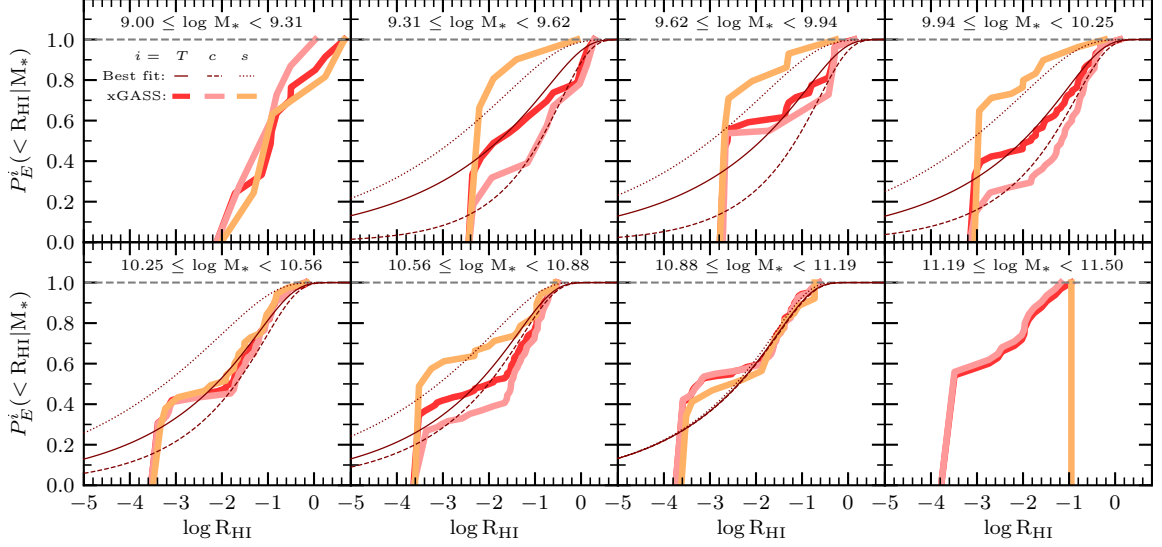**Figure 5.6:** Same as Figure 5.5 but for the subsample of LTGs.

Figure 5.7: Same as Figure 5.5 but for the subsample of ETGs.

5.3.3 Corrections from xGASS to calculate H I distributions for centrals and satellites

We would like to obtain from the xGASS analysis presented above a way to estimate the H I conditional CDFs of central and satellite galaxies when only the average CDFs (among centrals and satellites) are known. If $\mathcal{P}_j(> R_{\mathrm{H\,I}}|M_*)$, $j =$LTG or ETG, are the H I conditional CDFs from Chapter 4, then the corresponding CDFs for central and satellites can be calculated as:

$$\mathcal{P}_j^i(> R_{\mathrm{H\,I}}|M_*) = \left[\frac{P_j^i(> R_{\mathrm{H\,I}}|M_*)}{P_j(> R_{\mathrm{H\,I}}|M_*)}\right]_{\mathrm{xGASS}} \times \mathcal{P}_j(> R_{\mathrm{H\,I}}|M_*), \qquad (5.3.1)$$

where i refers to either central or satellite galaxy, and the sub-index xGASS refers to analytic fits to the H I CDFs constrained above. Thus, our goal now is to (i) perform a continuous analytic fit to the different xGASS H I CDFs given M_* entering in Eq. (6.2.6), and (ii) to be able to extrapolate the fits to lower stellar masses than those of the xGASS sample.

The H I conditional CDFs from the processed xGASS data presented in Figures 5.5–5.7 are for the whole sample as well as for different subsamples. In many cases, the numbers of objects in a given M_* bin, specially for subsamples containing ETGs and satellites, are low. Then, the CDFs are poorly defined and may suffer of strong sample variance. In view of this, performing fits independently to each CDF is not viable. Besides, it is important that the fitted functions describing the CDFs obey by construction the law of total probability. According to this law applied to our context, the relation of the total conditional probability distribution of $R_{\mathrm{H\,I}}$ given M_*,

$P_T(< R_{\rm H\,I}|M_*)$, with, for example, two subsamples A and B, with their respective conditional probability distributions $P_A(< R_{\rm H\,I}|M_*)$ and $P_B(< R_{\rm H\,I}|M_*)$, is given by:

$$P_T(< R_{\rm H\,I}|M_*) = P_A(M_*)P_A(< R_{\rm H\,I}|M_*) + P_B(M_*)P_B(< R_{\rm H\,I}|M_*), \qquad (5.3.2)$$

where $P_A(M_*)$ and $P_B(M_*)$ are the marginalized probability distributions of these subsamples. In our case, the marginalized probabilities are the fractions of galaxies in the samples A and B as a function of stellar mass, $\phi_A(M_*)/\phi_T(M_*)$ and $\phi_B(M_*)/\phi_T(M_*)$, respectively. In Appendix G we present the different equations that should be obeyed according to the law of total probability for the whole sample of galaxies and different subsamples of LTGs/ETGs, centrals/satellites, and their combinations. In these "probability conservation" equations enter different fractions of subsamples (the marginalized probability distributions) as a function of M_*. In Appendix E we obtain analytic fits to these fractions using SDSS data. As discussed in §§ 5.2.1, the fractions of ETGs (centrals or satellites) as a function of M_* in `xGASS` are different to those from SDSS. This is why we decided to weight the `xGASS` sample to agree with the SDSS DR7 fractions. Having done this, we can use then the SDSS fractions in the mentioned above "probability conservation" equations.

Based on the considerations discussed above, we implement the following strategy for obtaining the fits to the $R_{\rm H\,I}$ conditional CDFs of the whole `xGASS` sample as well as of different subsamples:

1. Propose parametric functions that describe the $R_{\rm H\,I}$ conditional CDFs given M_* of the following four galaxy subsamples: all LTGs, all ETGs, central LTGs, and central ETGs.

2. Calculate the $R_{\rm H\,I}$ conditional CDFs given M_* for: the whole sample of galaxies, and the four subsamples of centrals, satellites, satellite LTGs, and satellite ETGs, from the CDFs of the previous item by means of the equations implying the law of total probability (see Appendix G).

3. Implement a continuous joint fitting procedure to the $R_{\rm H\,I}$ conditional CDFs given M_* of the whole sample and the different subsamples mentioned above as obtained from `xGASS` after our processing (Figs. 5.5–5.7) in order to constrain the parameters of the functions mentioned in the first item.

4. Extrapolate the constrained fits to lower stellar mass bins, that is, for $M_* < 10^9$ $M_\odot$, the mass limit of `xGASS`.

For item (1), we propose a generic function for the four subsets of H I CDFs, the

164

Table 5.1: Best fit parameters to four sets of H I CDFs

CDFs	a	b	c	e
LTGs	0.005 ± 0.09	0.53 ± 0.09	0.79 ± 0.18	0.67 ± 0.12
LTGs Centrals	-0.21 ± 0.15	0.71 ± 0.15	0.67 ± 0.12	0.60 ± 0.11
ETGs	0.07 ± 0.05	0.22 ± 0.07	0.86 ± 0.09	0.65 ± 0.09
ETGs Centrals	-0.004 ± 0.11	0.31 ± 0.15	1.09 ± 0.13	0.75 ± 0.12

incomplete gamma function:[4]

$$\mathcal{P}(< x | M_*) = \frac{1}{\Gamma(\alpha)} \int_0 e^{-x} x^{\alpha-1} dx \qquad (\alpha > 0), \qquad (5.3.3)$$

where Γ is the gamma function, $x \equiv R_{\mathrm{H\,I}}/\mathrm{R}_0$, and the parameters α and R_0 depend on M_*. We parametrize these dependencies as:

$$\alpha(M_*) = a(\log M_* - 10) + b \qquad (5.3.4)$$

where a and b are the slope and normalization of the power law, respectively, and

$$\mathrm{R}_0(M_*) = \frac{c}{\left(\frac{M_*}{M_{\mathrm{tr}}}\right)^d + \left(\frac{M_*}{M_{\mathrm{tr}}}\right)^e} \qquad (5.3.5)$$

here c is a normalization coefficient, M_{tr} is the transition mass where the double power law changes its slope, d and e are the slopes for the low- and high-mass sides. In fact, for the mass range of `xGASS` galaxies, a single power law is enough for $R_0(M_*)$. However, since we will extrapolate the fits to `xGASS` $R_{\mathrm{H\,I}}$ CDFs to lower masses, the second power law is necessary. We have found that the values of d and M_{tr} can be fixed, that is, they are not left as free parameters. These values were constrained in Chapter 4 from the H I CDFs of LTGs and ETGs for the compilation and processing presented in Chapter 2 in a large M_* range; we fix these parameters to those values constrained therein: $d = -0.018$ and $\log(M_{\mathrm{tr}}/\mathrm{M}_\odot) = 8.646$ for LTGs; $d = -0.820$ and $\log(M_{\mathrm{tr}}/\mathrm{M}_\odot) = 8.354$ in the case of ETGs. Thus, for the functions Eq. (5.3.3–5.3.5) remain four free parameters, a, b, c, and e. The above function Eq. (5.3.3) is proposed to describe each one of the four subsamples of CDFs mentioned in item (i). Therefore, we have 16 free parameters in all.

We constrain the 16 free parameters by jointly fitting the 9 sets of $R_{\mathrm{H\,I}}$ conditional

---

[4]We have shown in Chapter 2 that the H I conditional PDFs given M_* can be described by Schechter-like functions. Thus, it is reasonable to propose the incomplete gamma function for describing the respective cumulative PDFs. On the other hand, given the low numbers and non-regular variations in the $R_{\mathrm{H\,I}}$ CDFs with mass of some subsamples from `xGASS`, it does not make much sense to search for functions with more parameters.

CDFs from `xGASS` mentioned in items (i) and (ii) above, and plotted in Figures 5.5–5.7. To do so we use a Monte Carlo Markov Chain method described in detail in Rodríguez-Puebla et al. (2013). We did not use the information from the largest and lowest stellar mass bins in all the cases because the data in these bins are scarce and the corresponding CDFs are poorly determined. In Table 5.1 we present the best constrained values for the 16 free parameters. With these values, the four `xGASS` $R_{\mathrm{H\,I}}$ conditional CDFs mentioned in item (i) above are fully described. By using the equations from Appendix G, the other five $R_{\mathrm{H\,I}}$ CDFs mentioned in item (ii) above result also described. Thus, any `xGASS` H I conditional CDF given M_* is described analytically by the fits, in particular those CDFs in the brackets in Eq. (6.2.6).

The obtained best fits from the continuous joint fitting procedure are shown in Figures 5.5–5.7 with thin solid, dashed, and dotted lines. The fits capture the main trends of the different $R_{\mathrm{H\,I}}$ conditional CDFs with $R_{\mathrm{H\,I}}$ and M_*. For some mass bins of ETGs (Fig. 5.7), the fits depart from the data. However, note that the CDFs in these cases move away from a systematic trend with mass. Recall that the fits are designed for capturing the continuous trends for the all, late-, and early-type samples *jointly*. While the whole exercise can be probably improved by proposing functions with more parameters, we prefer to keep it as simple as possible since the complexity (and possible uncertainties) of the data describing the whole sample and the different subsamples makes statistically non significant a mere quantitative improvement in the fits.

Finally, the fitting functions to the `xGASS` $R_{\mathrm{H\,I}}$ conditional CDFs given M_* can be extrapolated to masses lower than 10^9 M$_\odot$. For this, we need the fractions and subfractions as a function of M_* entering in the equations of "probability conservation" presented in Appendix G. We use the fits of these fractions and subfractions to the SDSS data presented in Appendix E, and extrapolate them to lower masses.

5.4 The bivariate $M_{\mathrm{H\,I}}$ and M_* distributions of central and satellite galaxies

We are now in position to apply the `xGASS`-based "corrections" (and their extrapolations to lower masses) to the $R_{\mathrm{H\,I}}$ conditional CDFs of LTGs and ETGs from Chapter 4 to obtain the corresponding CDFs for central and satellite galaxies, see Eq. (6.2.6). From these CDFs as a function of M_* we can calculate any statistical estimator, for example the first and second moments, that is, the $R_{\mathrm{H\,I}}$–M_* relations and their scatters.

Following, we extend the results showed in Chapter 4 regarding the joint or bivariate M_* and R_{HI} distribution for all galaxies but now separating them into centrals and satellites. As discussed in that Chapter, by combining the M_{HI} (or R_{HI}) conditional PDFs given M_* and the galaxy stellar mass function (GSMF), $\phi_*(M_*)$, the bivariate distribution function, $\Phi(R_{\mathrm{HI}}, M_*)$, can be calculated. This function is defined as the bivariate number of galaxies within the mass ranges $\log M_* \pm d\log M_*/2$ and $\log R_{\mathrm{HI}} \pm d\log R_{\mathrm{HI}}/2$ in a given volume V and it has units of $\mathrm{dex}^{-2}\ \mathrm{Mpc}^{-3}$.

In the left panels of Figure 5.8, from top to bottom, we show the bivariate M_* and R_{HI} distributions for all, late-type, and early-type galaxies, respectively. The colored isocountours correspond to different intervals of bivariate number densities, $\Phi(R_{\mathrm{HI}}, M_*)$, as indicated in the palette. For constructing these bivariate distributions we have used the R_{HI} conditional PDFs given M_* for LTGs and ETGs, the GSMF and the fractions of LTGs and ETGs as a function of M_* reported in Chapter 4. The left columns, from top to bottom, show the bivariate distributions for all, late-type, and early-type galaxies. The black, blue, and cherry solid lines are the respective logarithmic mean relations, $\langle \log R_{\mathrm{HI}} \rangle$-$\log M_*$. As extensively discussed in Chapter 2 and 4, since LTGs dominate in number density at low masses, the $\langle \log R_{\mathrm{HI}} \rangle$-$\log M_*$ relation of all galaxies is similar to the one of LTGs up to $M_* \sim 10^{10}\ \mathrm{M}_\odot$. At larger masses, the fraction of ETGs, which have much lower H I gas contents (compare the medium and bottom left panels of Figure 5.8), increase and then the relation of all galaxies strongly falls to be finally similar to the one of ETGs at $M_* \gtrsim 10^{11.7}\ \mathrm{M}_\odot$. Note that the R_{HI} distribution for ETGs is non-regular, with a second concentration of galaxies at very low values of R_{HI}. The above is due to the top-hat component of the R_{HI} conditional PDFs (see Figure 5.1).

The new results from this Chapter are the bivariate distributions for the galaxies separated into *central and satellites*, both for the LTG and ETG subsamples as well as for the total galaxy population. In each one of the left panels of Figure 5.8, the up and down lines with respect to the logarithmic mean relations show the corresponding logarithmic mean relations for the central and satellite subsamples, respectively.

The medium and right panels present in detail the bivariate M_* and R_{HI} distribution of the respective central and satellite subsamples with their respective logarithmic mean relations (same solid lines as in the left panels). The dashed and dotted lines in these panels show the corresponding arithmetic mean relations, $\langle R_{\mathrm{HI}} \rangle$-$M_*$, and the relations using the median of R_{HI}, respectively.

For LTGs, satellites have on average lower H I gas contents than centrals. In particular, at the low-M_* side, the H I gas-rich galaxies are all centrals (there are not gas-rich satellites). On the other hand, the gas-poor low-mass LTGs are mostly
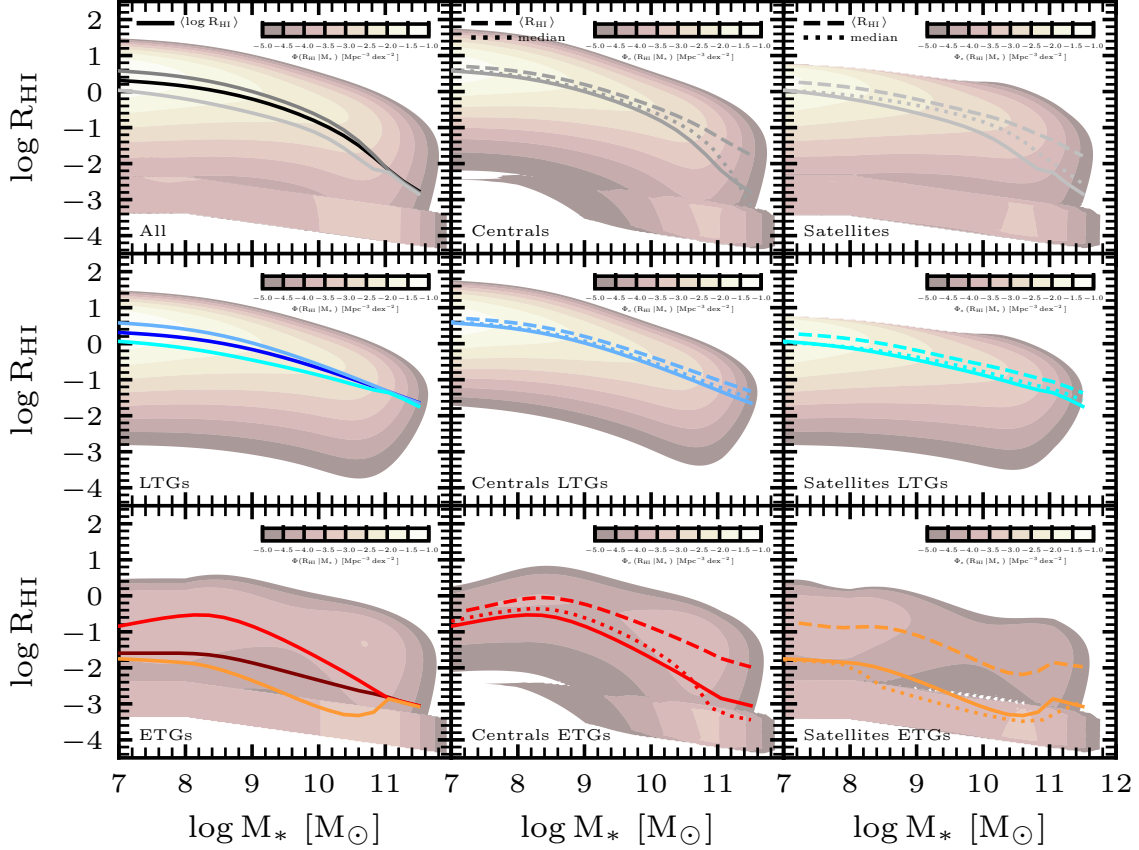
Figure 5.8: Empirical bivariate R_{HI} and M_* distributions, $\Phi(R_{\mathrm{HI}}|M_*)$. *Upper panels:* From left to right, the distributions for all, central, and satellite galaxies. The solid black, dark gray and light gray lines in the right panel compare the logarithmic means, $\langle \log R_{\mathrm{HI}} \rangle (M_*)$, of all, central, and satellite galaxies, respectively. Satellite galaxies have lower H I gas contents than centrals. The two last lines are reproduced in the medium and right panels, respectively. In these two last panels are also shown the arithmetic means ($\langle R_{\mathrm{HI}} \rangle (M_*)$, dashed line) and the medians (dotted line). *Medium panels:* As the upper panels but now for only LTGs. *Lower panels:* As the upper panels but now for only ETGs. The distribution for ETGs is hihgly bimodal. Hence, the different statistical estimators differ significantly among them.

satellites. At the high-M_* side central and satellite LTGs have approximately similar R_{HI} gas distributions.

For ETGs, the differences in the R_{HI} distribution between centrals and satellites are more significant than for LTGs. At the low-M_* side, among the ETGs, satellites are much more common than centrals. The H I gas contents of these satellite ETGs is strongly bimodal, with a subpopulation of galaxies with R_{HI} values close to those of the central ETGs and another subpopulation with very low R_{HI} values. For the central low-mass ETGs ($\lesssim 10^{10}$ M$_\odot$), there is a small fraction of them with relatively high values of R_{HI}. They probably correspond to the so-called blue ETGs, some of which are even star-forming (Lacerna et al., 2016, 2020). The blue/star-forming ETGs are typically very isolated galaxies and they indeed are expected to have relatively high gas fractions. At the high-M_* side, among the ETGs, centrals are more common than satellites. The differences in the R_{HI} distributions of the ETG centrals and satellites become small.

5.4.1 The H I mass functions

As shown in Chapter 4, the integration (marginalisation) of the bivariate M_{HI} and M_* distribution over M_* results in the H IMF. The top block of Figure 5.9 presents the above distribution, $\Phi(M_{\mathrm{HI}}|M_*)$, for all galaxies and the projected H IMF (right rotated panel). We also plot the logarithmic mean of M_{HI} as a function of M_* for all, central, and satellite galaxies, as well as the decomposition of the H IMF into centrals and satellites. For completeness, the GSMFs of all, central, and satellite galaxies are plotted in the upper panel; these functions are actually input in our approach along with the H I conditional PDFs given M_*.

In Chapter 4 it was shown that our empirical H IMF agrees well with those measured from the blind radio surveys ALFALFA and HIPASS, down to the completeness of our inference, $M_{\mathrm{HI}} \sim 10^8$ M$_\odot$ imposed by the completeness limit of the input GSMF, $M_* \sim 10^7$ M$_\odot$. As seen in Fig. 5.9, the H IMF is dominated by central galaxies at all masses. The fraction of centrals (satellites) is $\sim 90\%$ ($\sim 10\%$) or more (less) for $M_{\mathrm{HI}} \gtrsim 10^9M_\odot$, and for masses down to $\sim 10^8$ M$_\odot$, the fraction decreases down to $\sim 70\%$ (increases up to $\sim 30\%$). The differences in abundance between central and satellites are larger for M_{HI} than for M_*. In the medium and bottom blocks of Fig. 5.9 we present the bivariate distributions and their projections, the H IMF and GSMF, as in the top block, but for the subsamples of LTGs and ETGs. Since LTGs dominante in abundance, their mass functions are similar to those of the whole galaxy population.
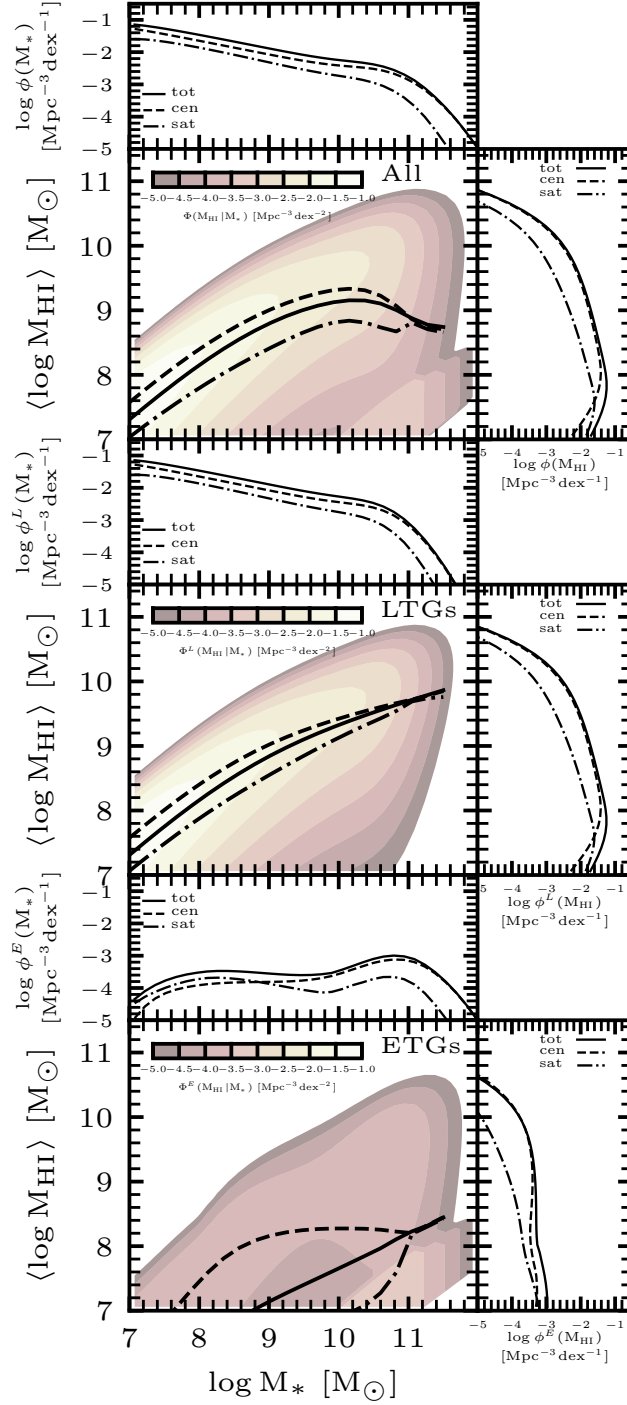
Figure 5.9: *Upper block of panels:* Bivariate M_{HI} and M_* distribution of all galaxies along with its logarithmic mean relation (solid line) as well as the relations for only central (dashed line) and satellite (dot-dashed line) galaxies. The projection of the bivariate distributions when integrating over M_* are the H I MF for all, central and satellite galaxies, shown in the right panel. The top panel shows the respective GSMFs, which are actually the observational input in our approach, along with the respective H I conditional PDFs. *Medium block of panels:* As the upper block but only for LTGs. *Lower block of panels:* As the upper block but only for ETGs.

5.5 Discussion

5.5.1 On the H I gas fraction of central and satellite galaxies

There are several pieces of evidence that the H I gas fraction of galaxies tends to be lower in higher-density environments (e.g., Haynes & Giovanelli, 1984; Gavazzi et al., 2005; Cortese et al., 2011; Catinella et al., 2013; Rasmussen et al., 2012; Boselli et al., 2014b). Studies of the H I gas content of member galaxies within clusters have shown that galaxies in most massive clusters are H I deficient, especially toward the center (e.g., Haynes & Giovanelli, 1984; Bravo-Alfaro et al., 2000; Solanes et al., 2001; Serra et al., 2012; Rasmussen et al., 2012; Taylor et al., 2012; Gavazzi et al., 2013). However, the above can be mainly due to the morphology-density relation, that is, ETGs, which are *intrinsically* gas-poorer than LTGs because exhausted their gas efficiently and early, are more abundant in the higher density regions of groups and clusters. On the other hand, the H I gas content in very isolated LTGs is on average higher than in cluster LTGs, however, the differences tend to be within the 1σ scatter, see Chapter 2 and references therein. The differences between these two opposite environments are larger for ETGs.

Other authors, rather than exploring environmental effects in specific clusters or for very isolated galaxies, used statistical samples to study the effects of the cluster/group mass and richness on the H I gas content of galaxies, mainly the satellite ones (e.g, Hess & Wilcots, 2013; Yoon & Rosenberg, 2015; Stark et al., 2016; Brown et al., 2017; Lu et al., 2020). Once a galaxy becomes a satellite inside a halo, the local environmental effects (ram pressure and viscous stripping, starvation, harassment, tidal interactions) work in the direction of lowering the gas content of the galaxy, more efficiently as more massive and rich is the halo (see for reviews e.g., Stark et al., 2016; Stevens et al., 2019, and references therein). It is worth of mention that in simulations (Wright et al., 2019) it was found that what matters most is not the halo mass, but the ratio between the satellite galaxy mass to the halo mass.

By means of the H I statistical stacking technique applied to an overlap between the ALFALFA survey and the SDSS Yang et al. (2007) halo-based group catalog, Brown et al. (2017) have found that the satellites in more massive halos have on average lower H I content at fixed M_* and specific SFR. According to their analysis, the systematic environmental suppression of H I content at both fixed M_* and fixed specific SFR in satellites begins in halo masses typical of the group regime ($> 10^{13}$ $M_\odot$), and fast-acting mechanisms such as ram-pressure stripping are suggested to explain their results. Stark et al. (2016) use RESOLVE, a volume-limited multi-

Table 5.2: Two sample KS test to R_{HI} conditional PDFs of centrals and satellites for late-, early- and all type galaxy populations.

$\log M_*$	p_L	p_E	p_A
7.45	$1.98{\times}10^{-10}$	0.002	$1.12{\times}10^{-11}$
7.9	$1.12{\times}10^{-11}$	$1.02{\times}10^{-5}$	$1.12{\times}10^{-11}$
8.35	$4.81{\times}10^{-11}$	$2.76{\times}10^{-5}$	$1.26{\times}10^{-7}$
8.8	$1.78{\times}10^{-4}$	$1.23{\times}10^{-6}$	$3.76{\times}10^{-8}$
9.25	$1.02{\times}10^{-5}$	0.002	$2.76{\times}10^{-5}$
9.7	$2.76{\times}10^{-5}$	$1.23{\times}10^{-6}$	$9.7{\times}10^{-4}$
10.15	0.002	$9.7{\times}10^{-4}$	0.032
10.6	0.095	0.040	0.056
11.05	0.240	0.950	0.680

wavelength census of ~ 1500 local galaxies, to study the H I-to-stellar mass ratio, R_{HI}, of satellite galaxies as a function of the halo (group) mass. They find that at fixed M_*, satellites have decreasing R_{HI} values with increasing halo mass starting at $M_h \sim 10^{12}$ M$_\odot$. The analogous relationship for centrals is uncertain and due to the poor overlap in stellar masses between centrals and satellites in the selected halo mass bins, it is not clear how different are the R_{HI} values of centrals and satellites at a fixed M_*. Their results for satellites suggest the presence of starvation and/or stripping mechanisms associated with halo gas heating in intermediate-mass groups.

The question that we address in this Section is how different is the H I gas content, at fixed stellar mass, between centrals and satellites separated explicitly into late- and early-type galaxies. In Section 5.3, we presented the respective results for the `xGASS` survey. Notice that upper limits were corrected for the distance bias, discussed in detail Section 5.2.2, and included into our survival statistical analysis as described in Section 5.2.3. Figure 5.3 shows that the $\langle \log R_{\mathrm{HI}} \rangle - \log M_*$ relations of centrals and satellites are different, specially for ETGs. At a fixed M_*, satellites have on average lower H I gas contents than centrals, the differences increasing as M_* decreases. For LTGs, these differences at $\sim 10^9$ M$_\odot$ are of ~ 0.6 dex, decreasing apparently to 0 at masses $\sim 10^{11}$ M$_\odot$. For ETGs, the differences are of ~ 1 dex at masses lower than $\sim 10^{10}$ M$_\odot$. However, it should be noted that the scatters (standard deviations) around the mean relations of centrals and satellites are large and the differences between the corresponding relations of both populations are lower than their standard deviations.

By using the `xGASS` measurements to the H I conditional distributions, in Section
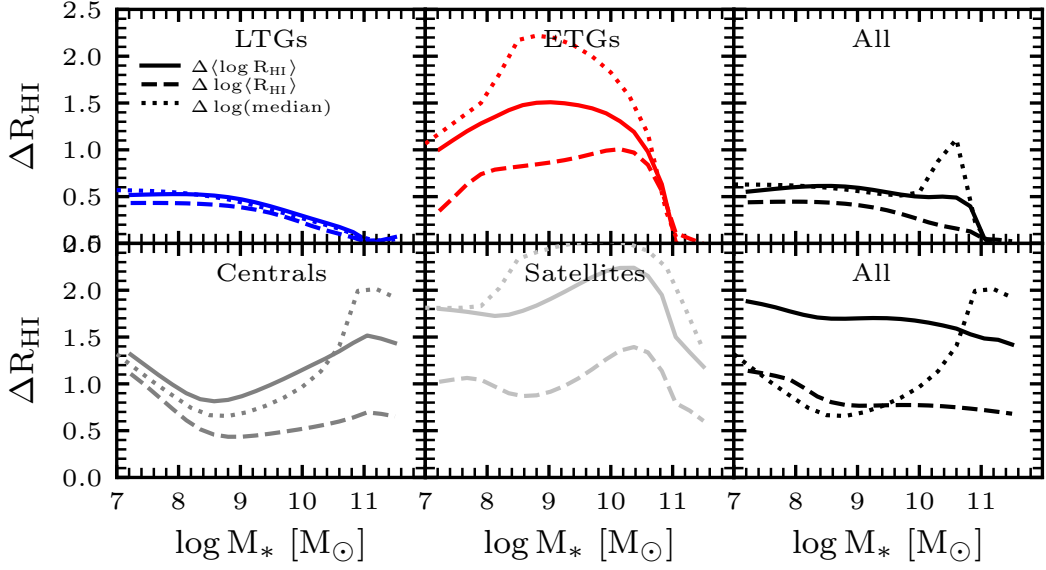
Figure 5.10: *Upper panels:* Relative differences between the logarithmic means (solid line), arithmetic means (dashed line), and medians (dotted line) of centrals and satellites as a function of M_*. From left to right, these relative differences are shown for the LTG, ETG, and whole galaxy populations. In all the cases, centrals have higher H I gas contents than satellites, but at the largest masses, these differences become very small. *Lower panels:* As the upper panels but in this case the relative differences are between LTGs and ETGs for the central, satellite and whole galaxy population (from left to right, respectively). LTGs have much higher H I gas contents than ETGs at all masses.

5.3.3 we constrained a set of proposed functions that allowed us to project the R_{HI} conditional PDFs for LTGs and ETGs presented in Chapter 4 into their corresponding distributions of centrals and satellites. The obtained joint (bivariate) distributions of R_{HI} and M_* are shown in Figure 5.8 along with their respective relations using different statistical estimators. As discussed in Section 5.4, the bivariate distributions of centrals and satellites are different both for LTGs and ETGs, and consequently for all galaxies. The differences depend on mass and for ETGs they are not easy to quantify by statistical estimators due their non-regular distribution of R_{HI}.

To dig deeper into the differences in the H I distributions of central and satellite galaxies, we apply a two-sample Kolgomorov-Smirnov test to the obtained R_{HI} conditional PDFs given M_* of centrals and satellites for late-type, early-type and all galaxies. The respective $p-$values are reported in Table 5.2. Quantitatively, the central and satellite H I distributions are different at the 95% or higher level ($p < 0.05$) for masses smaller than $M_* \sim 3 \times 10^{10}$ M$_\odot$ in all the cases. For larger masses, the differences are apparently weaker and both centrals and satellites are consistent with being drawn from the same distribution in their H I gas content.

In the upper panels of Figure 5.10, the relative differences in $\langle \log R_{\mathrm{H\,I}} \rangle$ (solid lines) and median $R_{\mathrm{H\,I}}$ (dotted lines) between centrals and satellites as a function of mass are shown for late-type, early-type, and all galaxies, from left to right, respectively. We show also the respective differences but for the arithmetic mean, $\langle R_{\mathrm{H\,I}} \rangle$. For LTGs, the relative difference between centrals and satellites is negligible at masses around 10^{11} $\mathrm{M}_\odot$ and it increases up to ~ 0.55 dex at $M_* \sim 5 \times 10^8$ $\mathrm{M}_\odot$, remaining similar at lower masses. The relative differences for the arithmetic mean are slightly smaller than for the logarithmic mean or the medians. For ETGs, the relative differences between centrals and satellites are larger than for LTGs. Since for ETGs, both centrals and satellites, the $R_{\mathrm{H\,I}}$ conditional distributions given M_* are non-regular, the statistical estimators (geometric or arithmetic mean and medians) significantly differ among them, and consequently, also differ the relative differences among these estimators for centrals and satellites. Our results suggest that the relative difference in $\langle \log R_{\mathrm{H\,I}} \rangle$ is negligible for $M_* > 10^{11}$ $\mathrm{M}_\odot$, but at lower masses, satellites are much H I gas-poor than centrals, by ~ 1.2 dex at masses around $3 \times 10^8 - 5 \times 10^9$ $\mathrm{M}_\odot$. As for the relative difference in the medians, is larger than in the logarithmic means, specially in the mass range $3 \times 10^8 - 10^{10}$ $\mathrm{M}_\odot$. For the arithmetic means, the relative difference is significantly lower at all masses. The arithmetic means minimize the contribution of galaxies with very low $R_{\mathrm{H\,I}}$ values, which in the case of ETGs, as already discussed, are dominant. For the combined population of late- and early-type galaxies, the relative differences in $\langle \log R_{\mathrm{H\,I}} \rangle$ between centrals and satellites are $0.4 - 0.6$ dex for $M_* < 5 \times 10^{10}$ $\mathrm{M}_\odot$. The differences are slightly larger for the medians and smaller for the arithmetic means.

Finally, from Figure 5.8 we note that the H I distributions differ much more when they are separated into late- and early-types (left panels) than when the separation is into centrals and satellites (top panels). The lower panels of Figure 5.10 show the relative differences in $\langle \log R_{\mathrm{H\,I}} \rangle$ (solid lines) and median $R_{\mathrm{H\,I}}$ (dotted lines) between LTGs and ETGs as a function of mass for central, satellite and all galaxies, from left to right, respectively. We also show the respective differences but for the arithmetic mean, $\langle R_{\mathrm{H\,I}} \rangle$, dahed lines. The relative differences in the lower panels are much larger than in lower panels. Overall, the above can be interpreted as that the present-day H I gas content of galaxies depends more on their internal nature, that is, whether they are of late or early type morphology, than on external conditions associated to whether the galaxy is central or satellite. Nevertheless, this claim is only valid at a very general level. As mentioned above, there are pieces of evidence that the H I gas content of satellite galaxies at a fixed stellar mass can be lower in massive haloes than in less massive ones.
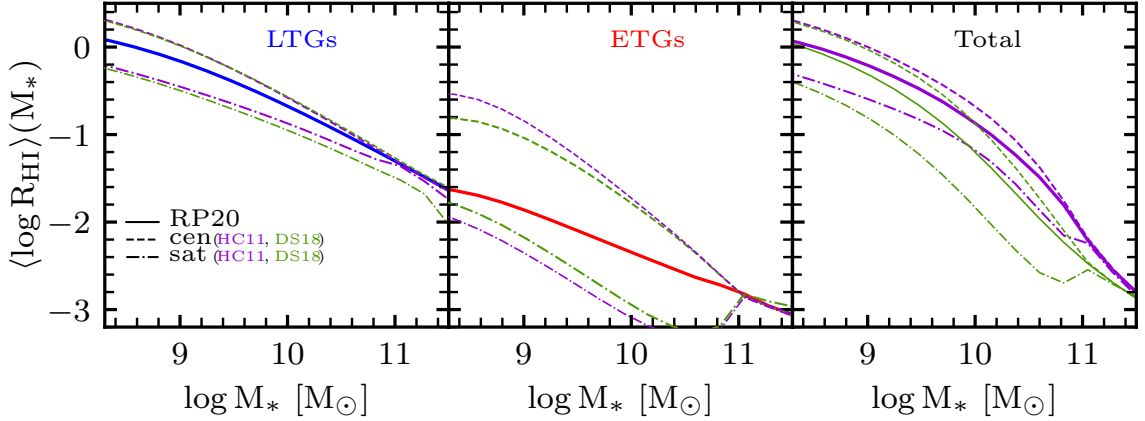
Figure 5.11: Logarithmic mean R_{HI}–M_* relations for LTGs, ETGs and all galaxies (solid lines) and their respective decomposition into central (short-dashed lines) and satellite (dot-dashed lines) using the Huertas-Company et al. (2011, magenta; see also Fig. 5.8) and Domínguez Sánchez et al. (2018, green) morphological classifications.

5.5.2 Caveats

Effects of different morphological classifications

The results presented here depend partially on the adopted criteria to morphologically classify galaxies as late- or early types. According to the above, we have separated the `xGASS` sample into LTGs and ETGs, and estimated the different fractions and subfractions as a function of M_* (Appendix E) required for our fitting procedure, by using the automatic morphological classification of Huertas-Company et al. (2011) implemented for SDSS galaxies. Next, we explore how much our results are affected by using an alternative morphological classification. Domínguez Sánchez et al. (2018) applied an automatic classification method to determine the morphology of the SDSS galaxies. We use their results to separate the `xGASS` sample into LTGs and ETGs, by employing the same morphological division criterion as we did in the case of the Huertas-Company et al. (2011) classification. Recall that elliptical and S0 galaxies were defined as ETGS, and from Sa to later types as LTGs.

The Domínguez Sánchez et al. (2018) morphological classification finds more ETGs than the one from Huertas-Company et al. (2011) at all masses, see Appendix E. As a consequence, the fractions of the different subpopulations change in `xGASS`, changing also the H I conditional CDFs corresponding to these subpopulations We have repeated the whole analysis presented in §5.3 but for the Domínguez Sánchez et al. (2018) morphological classification, and obtained different functions for the $[P_i^j(> R_{\mathrm{HI}}|M_*)/P_i(> R_{\mathrm{HI}}|M_*)]_{\mathrm{xGASS}}$ ratios appearing in Eq. (6.2.6). By using these new functions, we calculated the corresponding H I CDFs of central and satellite

galaxies for the LTG and ETG populations.

Figure 5.11 compares the resulting mean $\langle \log R_{\mathrm{H\,I}} \rangle$-$\log M_*$ relations of centrals and satellites for the LTG and ETG populations from the Huertas-Company et al. (2011) and the Domínguez Sánchez et al. (2018) morphological classifications. For the former, the line code is the same as in Figure 5.8. As seen, the differences in the $\langle \log R_{\mathrm{H\,I}} \rangle$-$\log M_*$ relations of centrals and satellites introduced by the uncertainty in morphological classification are negligible for LTGs. The differences for ETGs using Domínguez Sánchez et al. (2018) are from ~ 0.35 to 0.05 dex larger at masses $\sim 2 \times 10^8$ and $M_* \sim 2 \times 10^{10}$ M$_\odot$, respectively. At larger masses, the differences are negligible. The total relations shown in the right panel are the weighted averages of LTGs and ETGs. The different weights are given by the corresponding fractions as a function of M_*, and they depend on the morphological classification scheme, see Appendix E. This is why the total $\langle \log R_{\mathrm{H\,I}} \rangle$-$\log M_*$ relation is different when using one or another morphological classification. Since for the Domínguez Sánchez et al. (2018) classification the fraction of ETGs is larger than for the Huertas-Company et al. (2011) classification, and because ETGs are H I gas poorer than LTGs, the mean $\langle \log R_{\mathrm{H\,I}} \rangle$-$\log M_*$ relation in the former case is below than in the latter case.

In conclusion, variations in the morphological classification affect our inferences of the difference between the H I gas content of centrals and satellites but not significantly and only for ETGs. If we adopt the Domínguez Sánchez et al. (2018) morphological classification instead of Huertas-Company et al. (2011), these differences for ETGs in `xGASS` will be smaller by up to ~ 0.3 dex at small masses than those reported here.

Effects of membership and central/satellite designation errors

In this Chapter, we have used the `xGASS` survey for modeling the differences in the H I gas content between central and satellites galaxies. As mentioned in §§5.2.1, in `xGASS` the central/satellite assignation is made using the SDSS group catalog from Yang et al. (2005, 2007). This group catalog, as others, may suffer of membership allocation and central/satellite designation errors. For example, Campbell et al. (2015) estimated that, in a group catalog constructed using the Yang et al. (2005) group finder in a galaxy mock sample, the fraction of satellites that are truly satellites in the mock (purity) is around 70%, while for centrals, the purity decreases from $\sim 95\%$ at low group masses, $\sim 10^{12}$ h^{-1}M$_\odot$, to below 60% at masses $\gtrsim 10^{14}$ h^{-1}M$_\odot$. On the other hand, the fraction of satellites in haloes that are correctly assigned to groups (completeness) is $\sim 80\%$ independent of the halo mass, while for centrals the completeness decreases from $\sim 90\%$ at low halo masses to $\sim 60\%$ at the largest masses.

The main source of confusion for centrals at large group masses is the central inversion problem, when the most luminous or massive galaxy is a satellite rather than the true central (van den Bosch et al., 2005; Skibba et al., 2011).

A conclusion from Campbell et al. (2015) is that the errors in membership and central/satellite designation in group catalogs tend to equalize the properties of distinct galaxy populations, for example, satellite versus central galaxies. Thus, the differences in the H I gas fractions between centrals and satellites inferred with `xGASS` (see Section 5.3) are expected to be larger. The above implies also that the differences in the overall H I distributions of central and satellites reported in Section 5.4 could be larger.

5.6   Summary and Conclusions

We have analyzed the multi-wavelength `xGASS` survey (Catinella et al., 2018), applying the same procedure as in Chapter 2 to correct the upper limits by distance bias and to treat them with a survival analysis in order to infer statistical distributions of the H I gas content of galaxies.

We have found that for LTGs, the $R_{\mathrm{H\,I}}$–M_* relation and the full $R_{\mathrm{H\,I}}$ conditional distributions as a function of M_* from `xGASS` agree very well with those empirically determined in Chapter 2 for a larger mass range (Figure 5.3). For ETGs, the $R_{\mathrm{H\,I}}$ distributions of `xGASS` galaxies imply slightly higher values of $R_{\mathrm{H\,I}}$ than our previous determinations. For `xGASS` LTGs, centrals are on average more H I gas-rich than satellites of the same stellar mass, the differences vanishing for $\log(M_*/\mathrm{M_\odot}) > 10.8$, while at the lowest masses, $9.0 < \log(M_*/\mathrm{M_\odot}) \lesssim 9.7$, these average differences attain values of 0.5-0.7 dex. For ETGs, the differences between centrals and satellites are larger than for LTGs. However, in both cases, the 1-σ scatter around the $R_{\mathrm{H\,I}}$-M_* relations of centrals and satellites are larger than the relative differences of their means.

By means of a continuous fitting procedure to the processed `xGASS` data, we determined a set of functions that allowed us to project our empirical H I conditional cumulative distributions (CDFs) given M_* of both LTGs and ETGs into the components, in each case, corresponding to central and satellite galaxies. In other words, `xGASS` provides the information to estimate the H I conditional distributions of centrals and satellites from the overall H I conditional distributions for both late- and early-type galaxies. Our fitting procedure allowed us to extrapolate the above mentioned functions to stellar masses lower than those of the `xGASS` survey. By combining the $R_{\mathrm{H\,I}}$ conditional distributions given M_* with the corresponding GSMFs, the bi-

variate M_* and $R_{\mathrm{H\,I}}$ distribution functions, $\Phi(M_*, R_{\mathrm{H\,I}})$, for late-type, early-type, and all galaxies, separated into centrals and satellites, were calculated (Figure 5.8). The main results obtained from this exercise are as follows.

- For LTGs, satellites have on average less H I gas content than centrals. Up to $M_* \sim 10^9$ $M_\odot$, the relative differences are ~ 0.5 dex and all the gas-rich dwarf LTGs are centrals. For larger masses, the relative difference decreases up to the point that at $M_* \gtrsim 3 \times 10^{10}$ $M_\odot$, the differences become statistically not significant.

- Since the bivariate distribution is regular for LTGs, even for centrals and satellites separately, the $R_{\mathrm{H\,I}}$–M_* relations calculated with different statistical estimators are roughly similar.

- For ETGs, the bivariate distributions for centrals and satellites differ more than for LTGs, satellites being on average less devoid of H I than centrals up to $M_* \sim 5 \times 10^{10}$ $M_\odot$. However, the $R_{\mathrm{H\,I}}$ distributions of satellite ETGs are strongly bimodal, with a fraction of them having $R_{\mathrm{H\,I}}$ values close to those of the central ETGs and another fraction with very low $R_{\mathrm{H\,I}}$ values. At masses larger than 5×10^{10} $M_\odot$, central ETGs are already more abundant than satellite ETGs but both have statistically similar H I gas contents.

- Since the bivariate distributions for ETGs, both centrals and satellites, are non-regular, the $R_{\mathrm{H\,I}}$–M_* relations calculated with different statistical estimators are different. In particular, the relation based on arithmetic means, $\langle R_{\mathrm{H\,I}} \rangle$, is significantly above than the relations based on logarihtmic means or medians.

- The projection of the bivariate distribution when integrating it over M_* is the H I MF, which as shown in Chapter 4, agrees well with those ones measured in blind radio surveys. We show here that the H I MF is completely dominated by central galaxies at all masses, both for LTGs and ETGs.

Overall, our results show that the differences in the bivariate $R_{\mathrm{H\,I}}$ and M_* distribution for local galaxies are significantly more evident between late- and early-type samples than between central and satellite samples. The above suggests that the H I gas content of galaxies depends more on their internal nature, that is, whether they are of late or early type morphology, than on external conditions associated to whether the galaxy is central or satellite.

In this Chapter we presented a full statistical description of the H I gas content of local galaxies as a function of their stellar mass and separated into late and early

types, as well as into central and satellites. These results can be used for comparisons with theoretical predictions of galaxy evolution, and for adding the H I gas component in approaches aimed to model the local galaxy population. In particular, our results can be used to establish the M_*-$M_{\mathrm{H\,I}}$-M_h connection in the outcome of large N-body cosmological simulations, where complete mock galaxy catalogs can be generated. In Chapter 6, we present results of this connection including predictions on the spatial clustering of galaxies using both their stellar and H I masses.

Chapter 6

The Galaxy-Halo Connection extended to gas components

ABSTRACT

In this Chapter we extend the local stellar Galaxy-(Sub)Halo connection to the atomic hydrogen (H I) component by seeding semi-empirically galaxies into a large N-body dark matter (DM) simulation. The main input to construct the mock galaxy catalog are: our constrained stellar mass-to-halo circular velocity (M_*–$V_{\max}$) relation, assuming a scatter independent of any galaxy property, and the empirical $M_{\rm H\,I}$ conditional probability distributions given M_* for central and satellite galaxies. The galaxy stellar and H I mass functions, and the stellar mass spatial clustering from our mock catalog agree by construction with those from observations. We predict a logarithmic mean $M_{\rm H\,I}$–$M_{\rm DM}$ relation that increases with mass up to $M_{\rm DM} \sim 10^{12}$ M$_\odot$, attaining a maximum of $\langle \log(M_{\rm H\,I}/{\rm M}_\odot) \rangle \sim$ 9.2. At higher (sub)halo masses, $\langle \log(M_{\rm H\,I}) \rangle$ decreases slightly with $M_{\rm DM}$, that is, the $M_{\rm H\,I}$–$M_{\rm DM}$ relation is not a monotonic increasing function, and the scatter around this relation is large. The bivariate $M_{\rm H\,I}$ and $M_{\rm DM}$ distribution is broad and clearly bimodal for $M_{\rm DM} \gtrsim 10^{12}$ M$_\odot$, something inherited from the input H I conditional distributions given M_*. We also present the distribution of the total (central + satellites) H I gas mass within a halo as a function of halo mass. The H I projected two-point correlation functions increase weakly in amplitude as the $M_{\rm H\,I}$ threshold increases. These functions have higher amplitude than the ones measured in the blind H I ALFALFA survey but we show that this is mainly due to the selection effects of this survey. We discuss the implications of our results in the light of predictions from semi-analytical models and hydrodynamics simulations of galaxy evolution.

6.1   Introduction

According to the current scenario of cosmic structure formation, galaxies formed and evolved within Cold Dark Matter (CDM) halos from accreted and cooled gas within them. Depending on the different astrophysical processes and on the halo gravitational potential, a fraction of the accreted gas is deposited into the galaxy, part of this gas is transformed into stars while other part is expelled by feedback processes or returned to the interstellar medium due to short-lived stellar populations. Therefore, the cold gas

content that present-day galaxies have, given their halo and stellar masses (M_h and M_*, respectively), provides important information to constrain the different processes of galaxy evolution (see e.g., Crain et al., 2017, and more references therein). The cold (neutral) gas in the interstellar medium of galaxies is composed by atomic hydrogen (H I), molecular hydrogen (H_2), helium, and metals, being the dominant in mass the H I component (e.g., Fukugita & Peebles, 2004; Read & Trentham, 2005b; Rodríguez-Puebla et al., 2020).

Much work has been done in studying the H I content of galaxies in the nearby Universe. Among the largest efforts in this direction, we highlight the completion of the blind radio surveys H I Parkes All-Sky Survey (HIPASS; Meyer et al., 2004) and Arecibo Fast Legacy ALFA Survey (ALFALFA; Giovanelli et al., 2005; Haynes et al., 2011, 2018). From these surveys it was determined the local H I mass function (MF; Zwaan et al., 2005; Martin et al., 2010; Papastergis et al., 2012; Jones et al., 2018). While the total H I MF seems robust against the sample selection effects (down to the masses allowed by the 21-centimeter flux limit of the used radiotelescopes), this is not the case for the H I-to-stellar mass relation (see e.g., Maddox et al., 2015) and the projected two-point correlation function (2PCF) at different H I mass bins. The shallow flux limits, H I line width thresholds, and the small volumes of current H I surveys bias the samples to H I-rich and blue galaxies (e.g., Baldry et al., 2008; Haynes et al., 2011; Papastergis et al., 2012; Maddox et al., 2015; Chauhan et al., 2019). Moreover, this bias is mass dependent (see e.g., Huang et al., 2012c; Maddox et al., 2015). As a result, the H I 2PCFs measured, for instance from the H I-selected ALFALFA survey, are expected to be lower than for all galaxies (Huang et al., 2012c). Indeed, the measurements show that for H I-selected galaxies, the clustering amplitude is low and, depending on how the corrections to the sample selection effects are applied, different dependencies of clustering on $M_{\mathrm{H\,I}}$ are found (Basilakos et al., 2007; Meyer et al., 2007; Martin et al., 2012; Papastergis et al., 2013; Guo et al., 2017). By means of a subhalo abundance matching model (SHAM), Guo et al. (2017) attempted to reproduce the clustering measurements of the ALFALFA sample by selecting only halos that formed relatively late. This result is interpreted as an intrinsic assembly bias effect on the H I gas content of galaxies. However, recent semi-analytical models of galaxy evolution do not find assembly age to be strongly correlated with the H I content of galaxies (Spinelli et al., 2020; Chauhan et al., 2020).

In the future, facilities such as the the Five-hundred-meter Aperture Spherical radio Telescope (FAST Nan et al., 2011; Li et al., 2013), the Square Kilometre Array (SKA Carilli, 2014; Blyth et al., 2015), or precursor instruments such as the Australian SKA Pathfinder (ASKAP Johnston et al., 2008) and the outfitted Westerbork

Synthesis Radio Telescope (WSRT), will bring extragalactic H I studies more in line with optical/infrared surveys. The blind H I surveys will be then much deeper and for larger volumes than the current ones, helping this to reduce the strong sample selection effects. By now, the H I-to-stellar mass and other correlations have to be constrained from radio follow-up observations of optically-selected galaxy samples or by cross-correlating some radio surveys with optical/infrared surveys (e.g., Catinella et al., 2013, 2018; Wei et al., 2010b; Saintonge et al., 2011; Papastergis et al., 2012; Kannappan et al., 2013; Boselli et al., 2014a; Eckert et al., 2015; Stark et al., 2016).

In Chapter 2 we undertook the task of compiling and homogenizing from the literature as much as possible galaxy samples, in the spirit of the ones listed above (including most of them), with the additional requirement of information on the galaxy morphology because the HI gas content of galaxies strongly depends on morphology. For determining the H I mass distributions, we have taken into account the reported upper limits for the radio non-detections, and after correcting them for possible biases, we have applied a survival analysis. As a result, we were able to constrain not only the mean $M_{\mathrm{H\,I}}$–M_* and $M_{\mathrm{H_2}}$–M_* relations and their standard deviations for late- and early-type galaxies (LTG and ETG, respectively) down to $M_* \sim 3 \times 10^7$ $\mathrm{M_\odot}$, but the respective full conditional probability density distribution functions (PDFs) of the H I- and $\mathrm{H_2}$-to-stellar mass ratios for a given M_*. An important step in Chapter 2 was to analyze separately late- and early-type galaxies given their very different gas distributions.

In Chapter 3 we derived a well-constrained Galaxy Stellar Mass Function (GSMF) for all, late- and early-type galaxies down to $\sim 3 \times 10^7$ $\mathrm{M_\odot}$, and combined them with the H I and $\mathrm{H_2}$ conditional PDFs to generate the bivariate or joint $(M_*,M_{\mathrm{H\,I}})$ and $(M_*,M_{\mathrm{H_2}})$ distribution functions in Chapter 4. By projecting these bivariate distributions into the H I and $\mathrm{H_2}$ axes, we obtained the H I and $\mathrm{H_2}$ MFs, for LTGs and ETGs, as well as for all galaxies. In Chapter 4 it was shown that our empirical H I and $\mathrm{H_2}$ MFs, both for all and for only ETGs, agree well with those measured from blind radio surveys or optically-selected radio samples corrected for volume incompleteness. In particular, we have shown that the H I MF from blind radio surveys, like HIPASS and ALFALFA, are not affected by selection effects. We concluded that our empirical statistical description of the local galaxy population regarding the stellar and H I and $\mathrm{H_2}$ gas contents integrates and extends well a large body of observational information.

6.1.1 The H I–dark matter connection

The bivariate M_* and $M_{\mathrm{H\,I}}$ distribution of the local galaxy population described above can be used to extend the empirical galaxy-halo connection to the H I gaseous compo-

nent of galaxies. There were some previous attempts to set the galaxy-halo connection for $M_{\rm HI}$, by means of Halo Occupation Distribution (HOD) and SHAM models (e.g., Papastergis et al., 2012, 2013; Guo et al., 2017; Padmanabhan et al., 2017; Paul et al., 2018; Obuljen et al., 2019), or from physically motivated models that employed results from semi-empirical models (see e.g., Popping et al., 2015). The results are quite diverse, both regarding the H I-to-halo mass relation and its scatter, and the H I galaxy clustering. A potential shortcoming in applying the SHAM is that the key assumption of this technique -the existence of a monotonic and tight relation between $M_{\rm HI}$ and M_h (or any other halo property)- is probably not obeyed. On the other hand, the use of the measured H I 2PCFs in the HOD models could lead to incorrect H I-to-halo mass relations given the strong biases the measured H I 2PCFs suffer due to the sample selection effects mentioned above. There are also some recent studies aimed at determining the H I gas content of halos selected from optical galaxy group catalogs and by measuring the H I stacked spectra of the entire groups (Guo et al., 2020) or by using the measured H I masses of the galaxy members from blind H I surveys (Ai & Zhu, 2018; Tramonte & Ma, 2020; Lu et al., 2020).

The results from the semi-empirical studies mentioned above may be affected by several selection and confusion effects in the observational data that they use. On the other hand, they determine the total galaxy H I mass within haloes but not the central (satellite) galaxy H I mass in haloes (subhaloes). Thus, there is still a lack of full and self-consistent empirical determinations of the galaxy stellar-H I-(sub)halo mass connection, which is crucial for constraining theoretical models or for comparing with predictions from semi-analytic models and hydrodynamics simulations (e.g., Kim et al., 2017; Baugh et al., 2019; Villaescusa-Navarro et al., 2018; Spinelli et al., 2020; Chauhan et al., 2020).

In this Chapter, in order to establish the above mentioned self-consistent connection, we take a different approach with respect to previous ones. First, we perform a non-parametric SHAM in a large N-body cosmological simulation. The SHAM is applied to the maximum circular velocity of halos, $V_{\rm max}$, and peak circular velocity of subhalos, $V_{\rm peak}$ in order to reproduce the *stellar* galaxy clustering. The SHAM result is used to assign stellar masses to the halos and subhalos in the simulation, and given M_*, the H I mass is drawn from the empirical H I-to-stellar mass distributions from Chapter 4. For the latter step, the H I-to-stellar mass distributions of LTGs and ETGs are allowed to differ between centrals and satellites by using the functions constrained in Chapter 5 based on the recent extended GALEX Arecibo SDSS Arecibo survey (xGASS Catinella et al., 2018), where galaxies were separated into centrals and satellites. In this way, we construct an empirically-based mock halo-galaxy catalog

that allows us to predict the H I-to-halo mass relation for central and satellite galaxies as well as the H I spatial clustering.

The outline of this Chapter is as follows. In section 6.2 we present our methodology and the inputs needed to construct a mock catalog with galaxies (halos) information. In section 6.3 we show our results from the mock catalog generated such as the extended Galaxy-Halo connection, along with some consistency tests. We discuss these results in section 6.4. Finally we give our conclusions and future work in section 6.5.

6.2   The Galaxy-Halo connection

This Section describes our procedure to attain the galaxy-halo connection by including both the empirical stellar and H I mass distributions, and to obtain also predictions on the spatial clustering traced by H I mass. The procedure implies the construction of a mock galaxy catalog based on the outcome of a large cosmological N-body simulation. Figure 6.1 presents an illustrative scheme of all the ingredients of our procedure, and it refers to the subsections where the details can be consulted. In short, we begin by specifying the halo catalog of the N-body simulation that we employed, Section 6.2.1; then we describe how we assign stellar masses to every halo and subhalo in the simulation, Section 6.2.2, and their corresponding late- or early-type morphologies, Section 6.2.3. Once we have specified the above properties for our mock galaxies, we assign the H I mass from empirical distributions, Section 6.2.4. Finally, every (sub)halo in the simulation ends with an assigned stellar and H I mass. As a result, the full distribution of $M_{\mathrm{H\,I}}$ and V_{max} or M_{DM} is obtained and the H I 2PCF can be measured. We can also apply to the mock catalog cuts to emulate selection effects in the observational surveys for comparing with them, Section 6.2.5.

6.2.1   The simulation

We use the snapshot $z = 0$ of the cosmological $N-$body `SmallMultiDark-Planck` (`SMDP`) simulation (Klypin et al., 2016) for halo properties and the spatial distribution of galaxies. The `SMDP` is a simulation with 3840^3 particles and mass resolution of $9.63 \times 10^7 \mathrm{M}_\odot/h$ in a box of $L = 400 h^{-1}$ Mpc on a side. The adopted cosmology in the `SMDP` simulation is based on the flat Λ CDM model with $\Omega_\Lambda = 0.693$, $\Omega_{\mathrm{M}} = 0.307$, $\Omega_{\mathrm{B}} = 0.048$, $h = 0.678$, and $\sigma_8 = 0.829$, compatible with *Planck*15 results (Planck Collaboration et al., 2016b). Dark matter halos/subhalos were identified using the Rockstar halo finder (Behroozi et al., 2013b). Halo masses were defined using
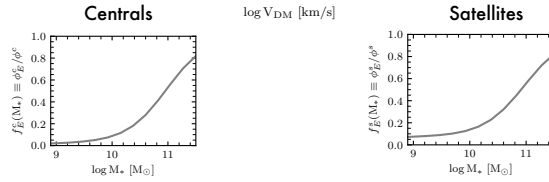
1. DARK MATTER (SUB)HALO CATALOG. Extract a halo or Subhalo from snapshot $z = 0$ of the SMDP simulation, Section 6.2.1.
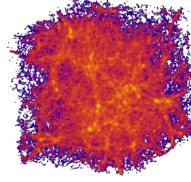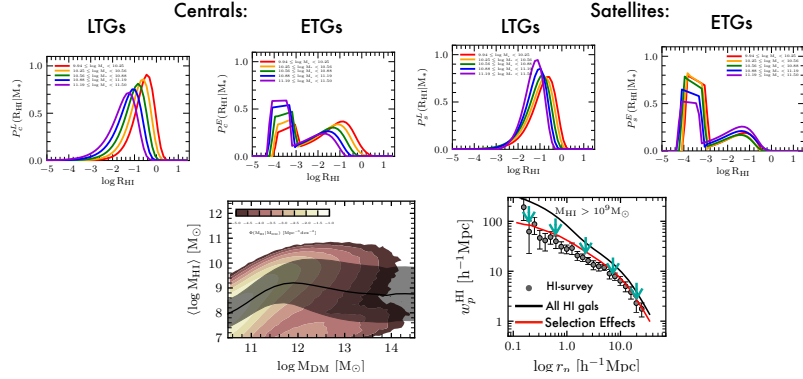
2. STELLAR MASS. Assign a stellar mass using halo Vmax from SHAM, Section 6.2.2.
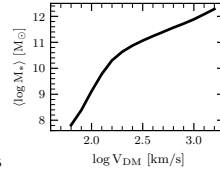
3. MORPHOLOGY. Choose an early- or late-type morphology based on the SDSS Yang+2012 galaxy group and the Huertas-Company morphologies, Section 6.2.3.

4. HI MASS. Assign a HI mass using the central/satellite conditional probability distributions of early- and late-type galaxies from Calette+2020, Section 6.2.4.

5. HI-HALO CONNECTION. Selection Effects, Section 6.2.5.

Figure 6.1: Schematic figure for the 5 steps taken to implement our mock galaxy catalogue: 1) we use the halo catalogues of the SMDP simulation; 2) stellar masses are assigned using the $V_{\rm DM}$-M_* relation assuming a total scatter of $\sigma = 0.15$ dex using Chapter 3 GSMF; 3) an early- or late-type morphology is assigned based on the observed abundance in the SDSS DR7 galaxy groups catalogues (Yang et al., 2012) and the Huertas-Company et al. (2011) morphologies; 4) H I masses are assigned depending on whether the galaxy is central or satellite and whether the galaxy is central or satellite based on Chapter 5 PDFs, and 5) our H I-halo connection is totally defined. Selection effects that emulate those of the observational surveys can be introduced in the mock catalog.

spherical overdensities according to the redshift-dependent virial overdensity $\Delta_{\mathrm{vir}}(z)$ given by the spherical collapse model; for the selected cosmological model, $\Delta_{\mathrm{vir}} = 333$ at $z = 0$. As discussed in the next subsection, for subhalos[1] we used the highest V_{max} reached along the main progenitor branch of the subhalos. The halo/subhalo catalogs are entirely downloadable[2] and their statistical properties were presented in detail in Rodríguez-Puebla et al. (2016).

6.2.2 Stellar masses

For every halo and subhalo in the SMDP simulation a galaxy stellar mass is assigned via a *non-parametric* SHAM technique. SHAM is a simple statistical approach for connecting a halo or subhalo property (M_h, V_{max}, etc.) to that of a galaxy property (luminosity, M_*, etc.). In its simple form, SHAM assumes that there is a unique monotonic relation between the selected halo/subhalo and galaxy properties. The above clearly implies that central and satellite galaxies have identical galaxy-halo relationships. In the past, several authors have tested the effects of SHAM assumptions on the resulting galaxy clustering (see e.g., Yang et al., 2009a; Neistein et al., 2011; Rodríguez-Puebla et al., 2012, 2013; Campbell et al., 2018). These studies have shown that using identical relationships between centrals and satellites leads to inconsistent results as compared to the observed spatial clustering of galaxies. In particular, Rodríguez-Puebla et al. (2012, see also Rodríguez-Puebla et al., 2013) suggested that these relationships should be different in order not only to reproduce the observed clustering of galaxies but also the observed abundance of galaxies in groups of different halo masses. Nonetheless, other studies have found that the assumptions from SHAM can still be valid when instead of halo/subhalo mass are used, for instance, V_{max} or the highest V_{max} reached along the main progenitor branch of the halo/subhalo's merger tree, V_{peak} (Reddick et al., 2013a; Hearin et al., 2013; Campbell et al., 2018; Dragomir et al., 2018; Wechsler & Tinker, 2018b, and more references therein). Following Dragomir et al. (2018), here we assume that the halo property that correlates best with M_* and reproduces the observed galaxy spatial clustering is:

$$V_{\mathrm{DM}} = \begin{cases} V_{\mathrm{max}} & \text{for distinct halos} \\ V_{\mathrm{peak}} & \text{for subhalos} \end{cases}. \qquad (6.2.1)$$

Typically, authors use SHAM in its most simple form, that is, the cumulative halo and subhalo velocity function and the cumulative GSMF are matched in order to

---

[1]Subhalos are halos that merged with a larger halo and have survived as bound entities inside the distinct halo.

[2]http://hipacc.ucsc.edu/Bolshoi/MergerTrees.html

determine the M_*-$V_{\rm DM}$ relation. The above assumes that there is no scatter around the mean M_*-$V_{\rm DM}$ relation. In reality, it is expected that this relation has some scatter since the properties of the galaxies might be determined by different halo properties and/or some environmental factors. For example, in the case of the M_*–M_h relation, analysis of large galaxy group catalogs (Yang et al., 2009b; Reddick et al., 2013b), the kinematics of satellite galaxies (More et al., 2011), galaxy clustering (Shankar et al., 2014; Rodríguez-Puebla et al., 2015), as well as galaxy clustering combined with galaxy lensing (Zu & Mandelbaum, 2015), it was found that this dispersion is of the order of $\sigma \sim 0.15-0.20$ dex. To take this into account, SHAM should be modified to include a dispersion around the M_*-$V_{\rm DM}$ relationship. Then, in this Chapter we use a more general procedure for SHAM as described below.

We define $\mathcal{H}(M_*|V_{\rm DM})$ as the conditional probability distribution function that a halo with velocity in the range $\log V_{\rm DM} \pm d\log V_{\rm DM}/2$ hosts a galaxy with stellar mass in the range $\log M_* \pm d\log M_*/2$. We assume that $\mathcal{H}(M_*|V_{\rm DM})$ is a lognormal function,

$$\mathcal{H}(M_*|V_{\rm DM}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log M_* - \mu_M)^2}{2\sigma^2}\right], \qquad (6.2.2)$$

where $\mu_M = \langle \log M_*(V_{\rm DM}) \rangle$ is the mean logarithmic stellar mass-to-$V_{\rm DM}$ relationship and σ is the scatter, assumed to be independent of $V_{\rm DM}$.[3] Then the connection between the stellar mass and halo $V_{\rm DM}$ is given by the following equation:

$$\phi_{\rm g}(M_*) = \int \mathcal{H}(M_*|V_{\rm DM})\phi_V(V_{\rm DM})d\log V_{\rm DM}, \qquad (6.2.3)$$

where

$$\phi_V(V_{\rm DM}) = \phi_{\rm h}(V_{\rm max}) + \phi_{\rm sub}(V_{\rm peak}), \qquad (6.2.4)$$

is the (sub)halo $V_{\rm DM}$ function and the units for $\phi_{\rm g}, \phi_V, \phi_{\rm h}$ and $\phi_{\rm sub}$ are in Mpc^{-3}dex^{-1}.

In this Chapter we use the GSMF, $\phi_{\rm g}(M_*)$, reported in Chapter 3, which ranges from dwarf galaxies, $M_* \sim 3 \times 10^7 {\rm M}_\odot$, to massive galaxies at the centres of galaxy clusters, $M_* \sim 3 \times 10^{12}$. We note that the the GSMF from Chapter 3 was corrected for surface brightness incompleteness. For the total scatter around the M_*-$V_{\rm DM}$ relationship,[4] we assume $\sigma = 0.15$ dex in order to be consistent with the results discussed above for the M_*–M_h relation (we notice that the scatter around the M_*–$V_{\rm max}$ relation is expected to be lower than the one around the M_*–M_h relation). We derive

---

[3]Along this Chapter log is for logarithm base 10, and the width of the lognormal distribution, σ, is in dex. The calligraphic face is used for this case, that is, $\mathcal{H}(x|y) = H(x|y) \times \frac{x}{\log e}$.

[4]Note that this scatter is composed by the physical (intrinsic) scatter and random errors in the stellar mass determination. Estimates of random errors are of the order of 0.1 dex (e.g., Tinker et al., 2017, Chapter 3). The above implies that our assumed intrinsic scatter is then 0.11 dex.
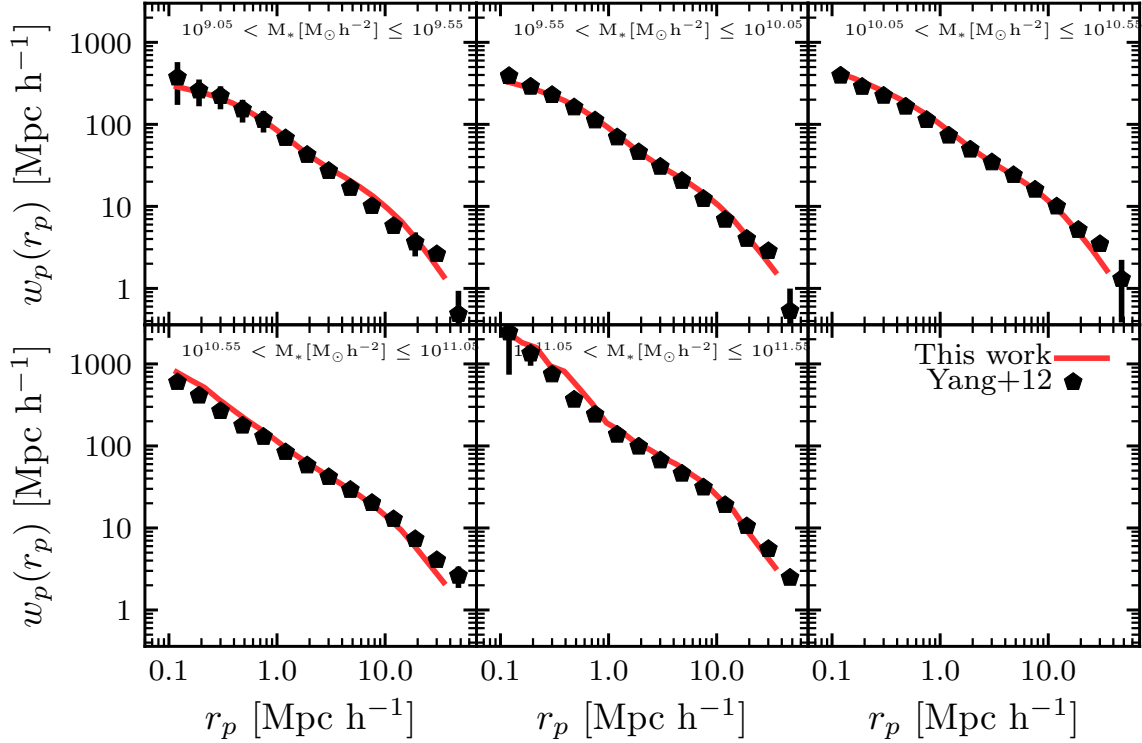
Figure 6.2: Stellar 2PCF in different stellar mass bins measured directly from our mock galaxy catalog. The red solid lines are our results and black pentagons are Yang et al. (2012) inferences from SDSS observations for a given stellar mass bin, indicated at the top of each panel.

a non-parametric mean M_*-V_{DM} relationship by numerically deconvolving Equation (6.2.3) as described in Appendix D.

The second panel of Figure 6.1 shows our obtained SHAM relationship between M_* and V_{DM}; see also Figure 6.5 below. Similarly to previous authors, our relationship increases steeply with V_{DM} for velocities below $V_{\mathrm{DM}} \sim 160$ km/s and it is shallower at higher velocities. Finally, Figure 6.2 shows that, as expected, our resulting mock galaxy catalogue reproduces well the observed SDSS DR7 projected 2PCFs at different stellar mass bins from Yang et al. (2012). For the 2PCFs, we assumed the plane parallel approximation with an observer along the z direction. This is estimated by means of the Landy & Szalay (1993) estimator and we integrate over the line of sight from $r_\pi = 0$ to $r_\pi = 40 \ h^{-1}$Mpc.

6.2.3 Morphologies

In this subsection we describe how we assign early and late-type morphologies to the galaxies in our mock catalogue. Notice, however, we assume that the scatter around the M_*–V_{DM} relation is indepedent of morphology, that is, the $P(M_*|V_{\mathrm{DM}})$ distribution is assumed to be independent of morphology. The reason for this is that

in the literature there is not yet a consensus on whether the galaxy-halo connection depends on galaxy properties such as morphology, color or star-formation rate. For example, some works find that at a fixed *halo mass* the stellar masses of blue/star-forming galaxies are on average larger (e.g., More et al., 2011; Rodríguez-Puebla et al., 2015) than red/quenched galaxies, while other authors have found opposite results (e.g., Moster et al., 2018). Others simply do not find significant differences (Zu & Mandelbaum, 2016; Behroozi et al., 2019). Nonetheless, all of them agree that at fixed *stellar mass*, red/quenched galaxies are more likely to reside in more massive halos than blue/star-forming galaxies, consistent with weak lensing studies (e.g., Mandelbaum et al., 2006, 2016). Since color/star formation rate correlate with galaxy morphology, similar trends are expected when morphology is used instead.

Assuming that $P(M_*|V_{\rm DM})$ is independent of galaxy properties like morphology is actually irrelevant for our results since we are interested here in predictions of the galaxy-halo connection for the *whole* galaxy population. However, the idea of introducing galaxy morphology is just a necessary step in order to assign H I masses in our mock catalog as it will become clearer later below.

Morphologies are assigned based on the observational fractions of ETGs for centrals, $f_E^{\rm cen}$, and satellites, $f_E^{\rm sat}$, as a function of stellar mass. To do so, we first generate a random number $U_{\mathcal{T}}$ uniformly distributed between 0 and 1. Then, for a given galaxy with mass M_*, the morphology $\mathcal{T}$ is assigned as

$$
\mathcal{T} = \begin{cases} \text{ETG} & \text{if } U_{\mathcal{T}} > f_E^{\rm i}(M_*) \\ \text{LTG} & \text{otherwise} \end{cases} , \tag{6.2.5}
$$

where the index i refers either to a central or satellite galaxy. The fractions of ETGs for centrals and satellites were derived in Chapter 5 for SDSS DR7 based on the galaxy group catalogue from Yang et al. (2012), the photometric catalogues of Meert et al. (2015) and Meert et al. (2016), morphologies from Huertas-Company et al. (2011), and the volume corrections from Chapter 3.

6.2.4   H I mass

At this point, every halo and subhalo in the `SMDP` simulation at the snapshot $z = 0$ has been assigned with a stellar mass M_* and galaxy morphology $\mathcal{T}$ depending on whether the galaxy is central or satellite. We use now these two galaxy properties to assign H I masses.

In Chapter 2 we found that the H I conditional PDFs given M_* for LTGs are well described by a Schechter function while ETGs are well described by a Schechter func-

tion plus a top-hat function. The parameters for the H I distributions were updated in Chapter 4. Note that these two distributions do *not* separate between centrals and satellites. Sampling H I masses using these two distributions could be problematic as it will be ignoring any environmental processes in satellite galaxies such as gas stripping. At the same time, the H I 2PCFs will be overestimated, specially at small distances, that is, at the 1-halo term. To overcome this problem, here we use the prescription presented in Chapter 5 to obtain the H I distributions for LTGs and ETGs separated into central and satellite galaxies.

Briefly, in Chapter 5 we used the `xGASS` survey (Catinella et al., 2018), which is a homogeneous sample that extends `GASS` (Catinella et al., 2010, 2013) down to $M_* \sim 10^9$ M$_\odot$. The `xGASS` detection limits in H I are such that the H I ratios, $R_{\mathrm{HI}} \equiv M_{\mathrm{HI}}/M_*$, are above 0.015 for massive galaxies and above 0.02 for the less massive ones. For galaxies non detected in H I, the upper limits of M_{HI} are reported. Galaxies are classified as satellites or centrals using the SDSS DR7 Group Yang et al. (2007) catalogue and corrected for shredding (Janowiecki et al., 2017). To assign the morphology $\mathcal{T}$, we use the Huertas-Company et al. (2011) classification, mentioned above. In Chapter 5 we followed Chapter 2 and applied to the `xGASS` sample the upper-limit corrections and the statistical procedures described there. Thus, we obtained the H I conditional distributions given M_* for LTGs and ETGs, separated into centrals and satellites. Then, we applied a continuous and joint fit to all these distributions; for more details we refer to the reader to Chapter 5.

The fits to the H I distributions for all LTGs and ETGs from `xGASS`, that is not separating into centrals and satellites, are roughly consistent with those from Chapter 2. However, there are three reasons why we cannot use directly the H I distributions determined from `xGASS` to separate into centrals and satellites: 1) `xGASS` extends only down to $M_* \sim 10^9$ M$_\odot$; 2) the `GASS` survey is a just a subsample of all the compiled data set in Chapter 2, that is, the latter involves more objects; and 3) we want to be consistent with the distributions reported in Chapter 2 and updated in Chapter 4. To consider the above and at the same time holding the differences in the H I distributions of central and satellite found from `xGASS`, in Chapter 5 we computed the H I distributions separated into centrals and satellites as follows:

$$\mathcal{P}_i^j(> R_{\mathrm{HI}}|M_*) = \frac{\mathcal{P}_i^j(> R_{\mathrm{HI}}|M_*)_{\mathrm{xGASS}}}{\mathcal{P}_i(> R_{\mathrm{HI}}|M_*)_{\mathrm{xGASS}}} \times \mathcal{P}_i(> R_{\mathrm{HI}}|M_*), \qquad (6.2.6)$$

where i refers to either LTG or ETG, j to either central or satellite galaxy, and the sub-index `xGASS` refers to the distributions as constrained in Chapter 5. The distributions $\mathcal{P}_i$ correspond to the ones reported in Chapter 4.
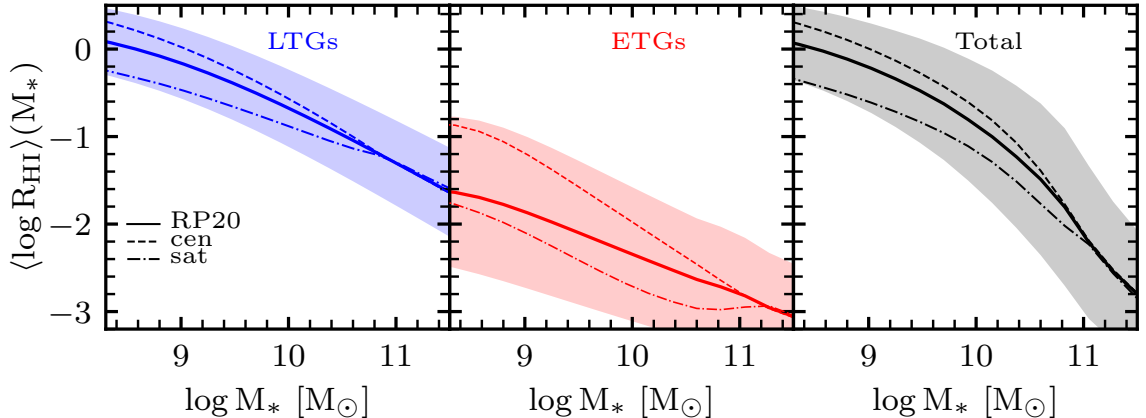
Figure 6.3: Empirical mean R_{HI}–M_* relation and its scatter (solid line surrounded by a shaded region) for LTG, ETGs, and all galaxies, from left to right. The logarithmic means and standard deviations are calculated from the HI conditional distributions reported in Chapter 4. The corresponding mean R_{HI}–M_* relations of central and satellite galaxies, calculated after applying the `xGASS`-based corrections as described in the text and in Chapter 5, are shown in each panel with dashed and dot-dashed lines.

The ratio of the HI distributions from `xGASS` in Equation (6.2.6), should be considered as the factor needed to project the total HI distributions of LTGs and ETGs into their corresponding distributions of centrals and satellite galaxies. The above warrants that our total distributions are consistent with the ones reported in Chapter 4. Since `xGASS` extends only down to $M_* \sim 10^9 \, M_\odot$, for lower masses in Chapter 5 we extrapolated the fits obtained for the ratios in Eq. (6.2.6) using also extrapolations from observations for the fractions of the different subpopulations required in the fits, see for details Chapter 5. Figure 6.3 shows the mean $\langle \log M_{HI}(M_*) \rangle$ relations and the standard deviations around them as calculated from the distributions from Chapter 4 for LTGs and ETGs (solid lines) as well as the mean relations for only central and satellite galaxies in each case (dashed and dot-dashed lines, respectively).

To assign HI masses to our mock central or satellite galaxies, we first generate a random number U_{HI} uniformly distributed between 0 and 1. Then, for a given galaxy with mass M_* and morphology $\mathcal{T}$, the HI mass is given by solving the following equation for R_{HI}

$$U_{HI} = \mathcal{P}_i^j(> R_{HI}|M_*), \tag{6.2.7}$$

where i refers to morphology $\mathcal{T}$ (ETG or LTG), and j to either central or satellite galaxy. Since the HI masses are assigned following robust empirical HI conditional PDFs given M_* and the mock galaxy catalog reproduces the observed GSMF by construction (see above), then one expects that the catalog should reproduce also the observed HI mass function (see Chapter 4). Figure 6.4 confirms this expectation.
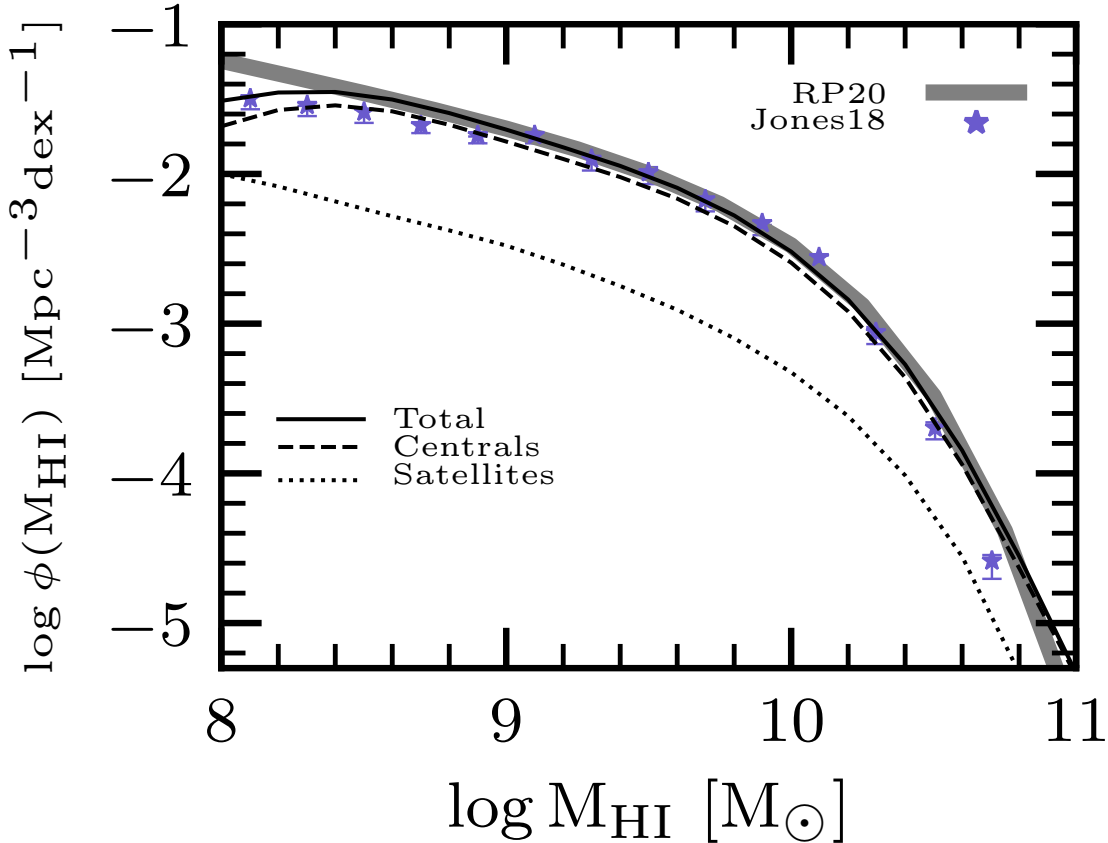
Figure 6.4: H I mass function measured directly from our mock catalog (black solid line). The dashed and dotted lines correspond to only centrals and satellites, respectively. For comparison, we also plot Jones et al. (2018) inferences (purple stars) using ALFALFA 100% (α.100), and our empirical inference of the *intrinsic* H I mass function from Chapter 4 (thick solid line).

In this figure are shown also the H I mass functions of central and satellite galaxies separately. The former dominate by far the H I gas content of galaxies, specially at lower masses.

Finally, our mock galaxy catalog contains three properties both for central and satellite galaxies: stellar mass, M_*, morphology, $\mathcal{T}$, and H I mass, M_{HI}. In other words, we have implicitly determined the multivariate conditional distribution function $\mathcal{P}(M_{\mathrm{HI}}, \mathcal{T}, M_*|V_{\mathrm{DM}})$ that describes the connection between haloes and galaxies. Due to the assumptions used in this Chapter the above multivariate distribution is $\mathcal{P}(M_{\mathrm{HI}}, \mathcal{T}, M_*|V_{\mathrm{DM}}) = \mathcal{P}(M_{\mathrm{HI}}, \mathcal{T}|M_*)\mathcal{P}(M_*|V_{\mathrm{DM}})$.

6.2.5 Selection effects

The goal of this Chapter is to extend the galaxy-(sub)halo connection to the H I component by means of the methodology described above. To check the robustness

of our predicted results on the H I-(sub)halo connection it will be key to compare the observed spatial clustering of galaxies as traced by their H I mass with the clustering measured in our mock catalog.

To date, the largest survey in H I, where the projected 2PCF has been measured for several $M_{\rm HI}$ thresholds, is the ALFALFA blind radio survey (Papastergis et al., 2013; Guo et al., 2017). However, a direct comparison between the resulting clustering of our mock galaxy catalog and the ALFALFA survey is not trivial. To do so, we will need first to determine the selection function of the ALFALFA survey and then apply the same criteria to our mock galaxy catalogue. As discussed in the Introduction, ALFALFA is affected by several sample selection effects (which depend on many aspects such as instrumental limits, depth, integration time, etc.). These effects introduce strong biases in the surveyed galaxy population.

Here, in order to emulate the selection function of the ALFALFA survey, we use the $R_{\rm HI}$-M_* correlation reported in Maddox et al. (2015, see also Huang et al., 2012c). We assume that this correlation is lognormal distributed, and we interpolate the median and the dispersion as a function of M_* as tabulated in their Table 1. We define this distribution as $\mathcal{S}(R_{\rm HI}|M_*)$. Thus, for a given galaxy from our mock catalog with properties M_* and $M_{\rm HI}$, we generate a random number U_S uniformly distributed between 0 and 1, and we determine whether the galaxy will be observed by the ALFALFA survey if

$$\mathcal{S}(> R_{\rm HI}|M_*) \geq U_S, \tag{6.2.8}$$

otherwise it is not observed by ALFALFA. Note that the above is independent of the galaxy morphology or whether the galaxy is central or satellite. This criterion will be applied to our catalog when comparing the H I spatial clustering with ALFALFA in §§6.3.2.

6.3   Results

The previous Section describes how we generate a mock halo-galaxy catalog in the `SMDP` N-body simulation. By construction the mock catalog follows the observed total GSMF as well as the empirical H I mass conditional PDFs of galaxies given their M_*. It is expected that the mock galaxy catalog reproduces the observed stellar projected 2PCFs and the H I mass function. Indeed, in Figures 6.2 and 6.4 we showed that this is the case. Next, we present our results on the galaxy-halo connection including H I mass measured from our mock catalogue.

In the upper panels of Figure 6.5, we show the mean logarithmic M_*–$V_{\rm DM}$ and
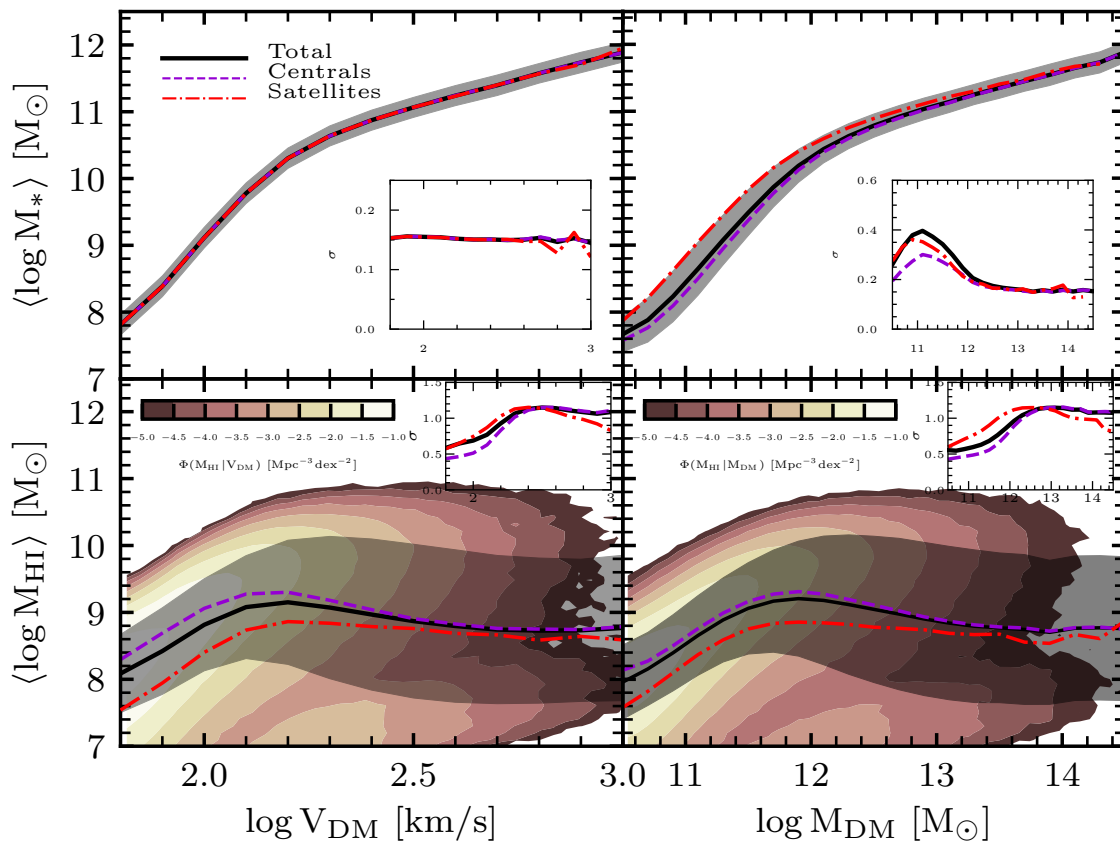
Figure 6.5: Stellar and H I galaxy-(sub)halo connection from the semi-empirical mock galaxy catalog. *Upper panels:* Mean logarithmic M_*–V_{DM} and M_*–M_h relations, solid black line, and their standard deviations, shaded region. The dashed and dashed-dotted lines correspond to the means of central and satellite galaxies. The insets show the scatter (SD) of each case. The scatter around the M_*–V_{DM} relation was assumed the same for central and satellite galaxies and independent of any galaxy property. *Lower panels:* As the upper panels but for $M_{\mathrm{H\,I}}$. Notice that unlike the M_*–M_{DM} relation, the $M_{\mathrm{H\,I}}$–M_* relation is not monotonic increasing and it has a huge scatter, specially at large masses. The colored isocountours correspond to the bivariate (joint) distributions, $\Phi(R_{\mathrm{H\,I}}|V_{\max})$ and $\Phi(R_{\mathrm{H\,I}}|M_{\mathrm{DM}})$ in the left and right diagrams, respectively. Note that the distributions are bimodal, something inherited from the input $M_{\mathrm{H\,I}}$ conditional distributions of LTGs and ETGs given M_*.

M_*–$M_{\rm DM}$ relations[5] for all galaxies as well as the respective standard deviations. Recall that the M_*–$V_{\rm DM}$ relation was assumed to be the same for central and satellite galaxies. However, in the case of the M_*–$M_{\rm DM}$ relation, right upper panel, satellite galaxies segregate from centrals towards lower subhalo masses for a given stellar masses (see e.g., Rodríguez-Puebla et al., 2012). As was shown by these authors (see also Rodríguez-Puebla et al., 2013), SHAM reproduces the observed spatial clustering of galaxies *only* when the stellar-to-(sub)halo relations of centrals and satellites are different in the direction shown in Fig. 6.5. In our SHAM, this segregation emerges because $V_{\rm DM}$ is the halo property assumed to correlate better with M_* instead of M_h, and, as discussed in §§6.2.2, this form of SHAM is expected to generate results consistent with the spatial clustering of galaxies (see also, Reddick et al., 2013b, and Dragomir et al., 2018). Finally, note that the scatter around the M_*–$M_{\rm DM}$ relation increased with respect to the M_*–$V_{\rm DM}$ relation, specially at low masses, where the segregation between centrals and satellites is larger.

Once we have shown that our mock galaxy catalog is quite realistic, we can proceed to explore the predictions from this catalog for *(i)* the galaxy-halo connection for the H I gas content (§§6.3.1), and *(ii)* the H I projected 2PCFs for different $M_{\rm H I}$ thresholds (§§6.3.2). These are the main goals of this Chapter.

6.3.1   The galaxy H I-(sub)halo connection

The lower panels of Figure 6.5 show our resulting total $M_{\rm H I}$–$V_{\rm DM}$ and $M_{\rm H I}$–$M_{\rm DM}$ relations. In this figure, we plot the mean $\langle \log M_{\rm H I} \rangle$ and their corresponding standard deviation for all galaxies (black solid line and shaded region) as well as the mean relation for central (violet dashed line) and satellite (red dashed-dotted line) galaxies.

On average, $M_{\rm H I}$ for all galaxies increases as $M_{\rm H I} \propto V_{\rm DM}^{3.5}$ or $M_{\rm H I} \propto M_{\rm DM}^{1.0}$ up to a peak at $V_{\rm DM} \approx 160$ km/s or $M_h \approx 10^{12}$ M$_\odot$, respectively; these trends are dominated by central galaxies, while satellite galaxies follow roughly similar trends but with lower values of $M_{\rm H I}$. For large values both for $V_{\rm DM}$ and $M_{\rm DM}$, the H I gas content of (sub)haloes poorly depend on their scale. If any, $\langle \log M_{\rm H I} \rangle$ slightly decreases while the scatter strongly increases. The decrease of $\langle \log M_{\rm H I} \rangle$ is mainly due to the central galaxies, while for satellites the mean relation keeps constant. At the largest velocities or masses, $\langle \log M_{\rm H I} \rangle$ is constant with $V_{\rm DM}$ or $M_{\rm DM}$. Our results show that *the $M_{\rm H I}$–$V_{\rm DM}$ and $M_{\rm H I}$–$M_{\rm DM}$ relations are neither monotonically increasing nor tight.* In §§6.4.1, we discuss the implications of the above for the traditional form of SHAM, and in §§6.4.3 we compare our results with those from semi-analytic models

---

[5]Note that here $M_{\rm DM}$ is the virial mass M_h in the case of distinct haloes but for subhaloes this is the highest M_h along the main progenitor branch of their merger tree.

and hydrodynamics simulations, some of which have predicted that the $M_{\rm HI}$–$M_{\rm DM}$ relation is non monotonic. In particular, the change of behavior of this relation at $M_h \sim 10^{12}$ is linked to the onset of AGN feedback (Kim et al., 2017; Baugh et al., 2019; Chauhan et al., 2020).

The H I–(sub)halo connection found here can be understood as the combination of the stellar mass conditional distribution given $V_{\rm DM}$, $P(M_*|V_{\rm DM})$, and the empirical H I mass conditional distribution given M_*, $P(M_{\rm HI}|M_*)$. The former was assumed to be *independent of galaxy morphology and environment*, while the latter strongly segregates by galaxy morphology and weakly by environment (defined by their central or satellite nature, see Fig. 6.3. By analyzing the two mentioned above distributions, we can interpret the results shown in Figure 6.5. For $V_{\rm DM} \lesssim 160$ km/s or $M_{\rm DM} \lesssim 10^{12}$ M$_\odot$, the corresponding stellar masses are below $M_* \sim 10^{10}$ M$_\odot$. The galaxy population at these masses is completely dominated by LTGs, see Figure 6.12 below or Figure E.1 in Appendix E. For these galaxies, as seen in Figure 6.3, on average $R_{\rm HI} \propto M_*^{-0.4}$, that is $M_{\rm HI} \propto M_*^{0.6}$. On the other hand, from Figure 6.5, we see that $M_* \propto V_{\rm DM}^{6.4}$ at low velocities. Therefore, roughly $M_{\rm HI} \propto V_{\rm DM}^{3.8}$, which is close to what it is seen in Fig. 6.5 for $V_{\rm DM} \lesssim 160$ km/s. For $M_* > 10^{10}$ M$_\odot$, on average $R_{\rm HI}$ for all galaxies strongly decreases with M_* due to the increasing fraction of ETGs as M_* increases, which have much lower values of $R_{\rm HI}$ than LTGs, see Fig. 4.3 in Chapter 4. From this figure, roughly $R_{\rm HI} \propto M_*^{-1.2}$ or $M_{\rm HI} \propto M_*^{-0.2}$, and since for large velocities, roughly $M_* \propto V_{\rm DM}^{1.7}$ (Fig. 6.5), then $M_{\rm HI} \propto V_{\rm DM}^{-0.3}$, in rough agreement with what it is seen in Figure 6.5 for $V_{\rm DM} > 160$ km/s. For the largest velocities, the total $R_{\rm HI}$–M_* relation flattens (see Fig. 4.3 in Chapter 4), hence the total $M_{\rm HI}$–$V_{\rm DM}$ relation flattens, too.

The scatter distribution around the total $M_{\rm HI}$–$V_{\rm DM}$ relation is large. For $V_{\rm DM} \lesssim 160$ km/s, the standard deviations increase from ~ 0.5 to 0.9 dex. For central galaxies, the scatter is lower, while the opposite applies for satellites. For larger velocities, the standard deviation increases up to 1.2 dex and at the largest $V_{\rm DM}$ values it slightly decreases. A large scatter is expected due to the broadness of the total $M_{\rm HI}$ conditional PDFs at the stellar masses corresponding to these velocities. The total PDFs are broad because of (i) the strong difference between LTGs and ETGs in the distributions of H I masses given M_*, see Fig. 6.3, and (ii) the large scatter around the mean $R_{\rm HI}$–M_* relations of LTGs and ETGs, as seen in Fig. 6.3. An expected consequence of the former is that the scatter around the mean $M_{\rm HI}$–$V_{\rm DM}$ relation at the $\sim 160 - 400$ km/s range should be significantly segregated by morphology, with LTGs lying above the mean relation and ETGs below it. Elsewhere we will present these results for the LTG and ETG populations separately, and under different assumptions
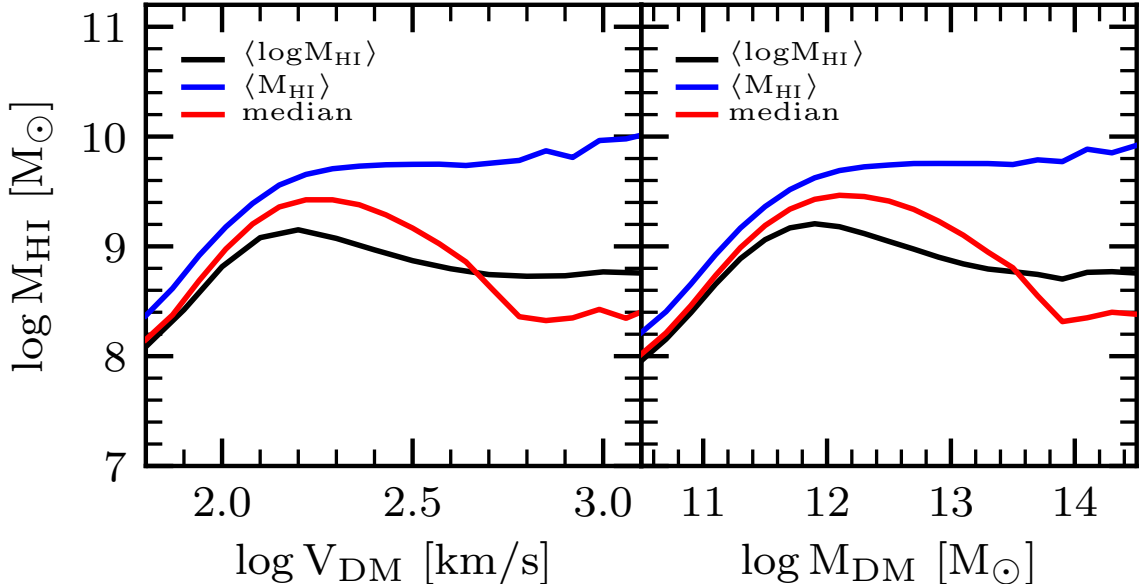
Figure 6.6: The galaxy H I-to-halo connection using three different statistical esti-
mators: logarithmic mean, arithmetic mean, and median. The relations are different
because the join distributions are asymmetric and even bimodal (see Fig. 6.5).

regarding the covariance of the scatter around the $V_{\rm DM}$-M_* relation with morphology
or color. According to Fig. 6.5, around the $M_{\rm HI}$–$V_{\rm DM}$ relation there is also a (weak)
segregation by central and satellite galaxies. Up to ~ 400 km/s, the former lies on
average slightly above the mean $M_{\rm HI}$–$V_{\rm DM}$ relation and the latter below it. However,
since this segregation is small, it is not expected to be source of the large scatter
around the $M_{\rm HI}$–$V_{\rm DM}$ relation.

It is worth of mentioning that the scatter around the $M_{\rm HI}$–$V_{\rm DM}$ relation is not only
large but also with a highly asymmetric distribution, and with even a bimodality
at $V_{\rm DM} > 160$ km/s as seen in Figure 6.5. As a result of the broad asymmetric
distribution, different statistical estimators taken as representative of the population
will significantly differ among them. Figure 6.6 shows our predicted galaxy H I-halo
connection using the geometrical mean $\langle \log M_{\rm HI} \rangle$ (as in Fig. 6.5, black line), the
arithmetic mean $\langle M_{\rm HI} \rangle$ (blue line), and the median (red line). Up to $V_{\rm DM} \sim 160$
km/s, where the scatter is relatively small and symmetric, the differences among the
three statistical estimators are relatively small. For higher halo velocities, since the
distribution of $M_{\rm HI}$ presents a long tail at its low-velocity side, the arithmetic mean
results much higher than the logarithmic mean and the median. The median differs
from the geometric mean due to the increasing relevance at high velocities of the
second peak in the $M_{\rm HI}$ distribution at its low-velocity end.

The explanations given above about the shape of the mean $M_{\rm HI}$–$V_{\rm DM}$ relation and
its scatter apply also to the shape of the mean $M_{\rm HI}$–$M_{\rm DM}$ relation and its scatter
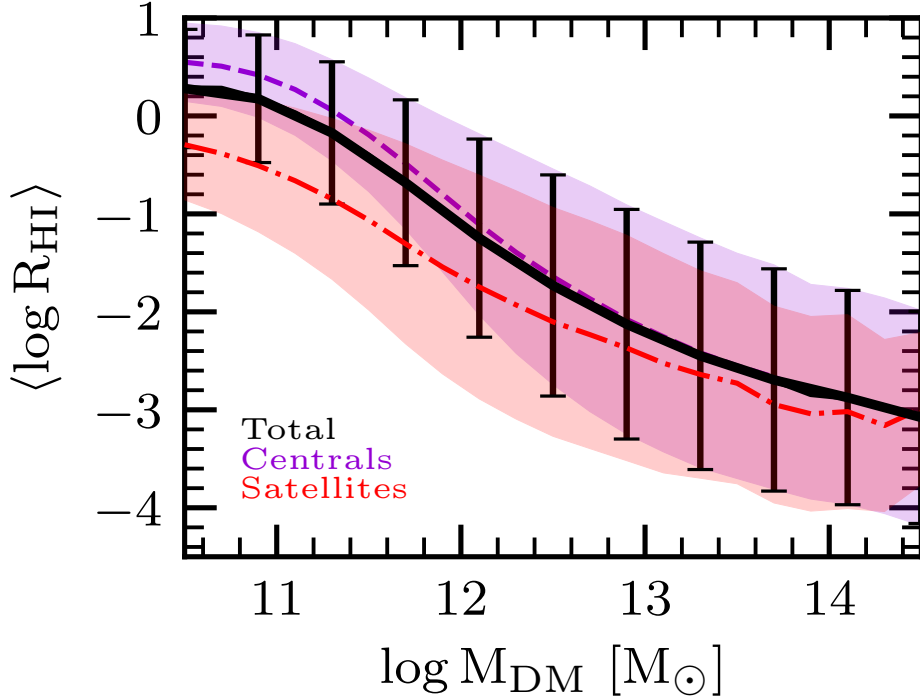
Figure 6.7: Logarithmic mean of the M_{HI}-to-M_* ratio as as function of (sub)halo mass M_{DM} for all galaxies (black solid line) as well as for only central (violet dashed line) and satellite (red dashed-dotted line) galaxies. The standard deviations for the central and satellite galaxy relations are shown with the shaded regions.

(lower right panel of Fig. 6.5). The difference is that in the latter case, instead of $P(M_*|V_{\mathrm{DM}})$, we have $P(M_*|M_{\mathrm{DM}})$, which segregates by environment (see upper right panel of Fig. 6.5). The segregation of the M_*-M_{DM} relation between centrals and satellites produces that the segregation of the M_{HI}–M_{DM} relation between centrals and satellites be weaker than in the M_{HI}–V_{DM} relation due to a compensation effect. Observations show that satellites have on average smaller H I masses at a given M_* than centrals (see Fig. 6.3, and Chapter 5) but at a given M_{DM}, satellites have on average larger M_* than centrals. Hence, the difference in M_{HI} at a given M_{DM} between satellites and centrals becomes smaller than in the case of a given V_{DM}. Distinct haloes contain on average more H I gas than subhaloes of similar masses in the mass range of $\sim 10^{11} - 10^{13}$ M$_\odot$, but the differences are small. In Figure 6.7 we plot the measured total R_{HI}–M_{DM} relation along with the relations corresponding to central and satellite galaxies. The shadow areas around them correspond to the standard deviations. Note that the scatter around the mean H I-to-halo mass relation of satellites galaxies is larger than for the central ones, specially at lower masses.

The total galactic HI gas content inside distinct haloes

Based on optical galaxy group catalogs for determining halo (group) masses, some authors have recently presented observational estimates of the *total* H I mass within these haloes, that is, the sum of H I masses of the central and satellite galaxies in the halo, $M_{\rm HI}^{\rm tot} = M_{\rm HI}^{\rm cen} + M_{\rm HI}^{\rm sats}$ (Ai & Zhu, 2018; Obuljen et al., 2019; Guo et al., 2020; Tramonte & Ma, 2020; Lu et al., 2020). The upper panel of Figure 6.8 shows the mean $\langle M_{\rm HI}^{\rm tot} \rangle$ and standard deviation[6] as a function of M_h as measured from our mock catalog. The mean for central galaxies, $\langle M_{\rm HI}^{\rm cen} \rangle$, and the sum of all satellite H I masses within a given distinct halo, $\langle M_{\rm HI}^{\rm sats} \rangle$, are also plotted with dashed and dot-dashed lines, respectively. The two sets of lines are for two minimum H I masses used to account for galaxies within the distinct halo, $M_{\rm HI,min} = 10^7$ M$_\odot$, black lines, and $M_{\rm HI,min} = 10^8$, red lines.

Figure 6.8 shows that the dependence of $M_{\rm HI}^{\rm tot}$ on M_h can not be described by a double power law. In the $10^{12} \lesssim M_h/{\rm M}_\odot \lesssim 10^{13}$ regime the relation flattens. This is because for haloes up to $\sim 2 \times 10^{12}$ M$_\odot$, central galaxies dominate by far the H I gas content in the halos; for them, the relation flattens at these masses. For larger masses, the contribution of satellites increases and at $M_h \sim 10^{13}$ M$_\odot$ the total H I mass of satellites is already equal to the H I mass of centrals of the same halo mass. At $M_h > 10^{14}$ M$_\odot$ the H I gas of satellites is on average ~ 1 dex larger than in the central galaxy. In the lower panel of Figure 6.8 we show the same as in the upper panel but for the stellar mass. Two minimum stellar masses were used to sum up satellites, 10^7 and 10^8 M$_\odot$. The differences between both cases are negligible. Unlike $M_{\rm HI}$, the dependence of M_* on M_h can be described by a double-power law. This is mainly because the dependence of M_* on M_h for centrals does not flattens or decreases as in the case of $M_{\rm HI}$. Elsewhere we will study in detail the conditional stellar and H I mass functions of haloes in our empirical model.

We compare our semi-empirical inferences with previous ones in Figure 6.9. In this Figure we show with isocountours the full join ($M_{\rm HI}^{\rm tot}$, M_h) distribution. The total H I mass estimates using H I stacked spectra for galaxy groups by Guo et al. (2020) agree with our findings up to $M_h \sim 10^{13}$ M$_\odot$ (we have transformed their halo masses defined at the $200\rho_m$ radius to our virial masses). For larger halo masses, their total H I masses are lower than our determinations. As seen in Figure 6.9 the above is mainly due to the lower H I masses they have from satellites, $M_{\rm HI}^{\rm sats}$, while the agreement with the H I mass of centrals at all halo masses is encouraging. The low values of $M_{\rm HI}^{\rm sats}$ may be due to the SDSS group catalogue not sampling the high halo mass end with enough

---

[6]Here we present arithmetic means because it is the way results are reported in in order to compare directly to observational works (see e.g. Guo et al., 2020).
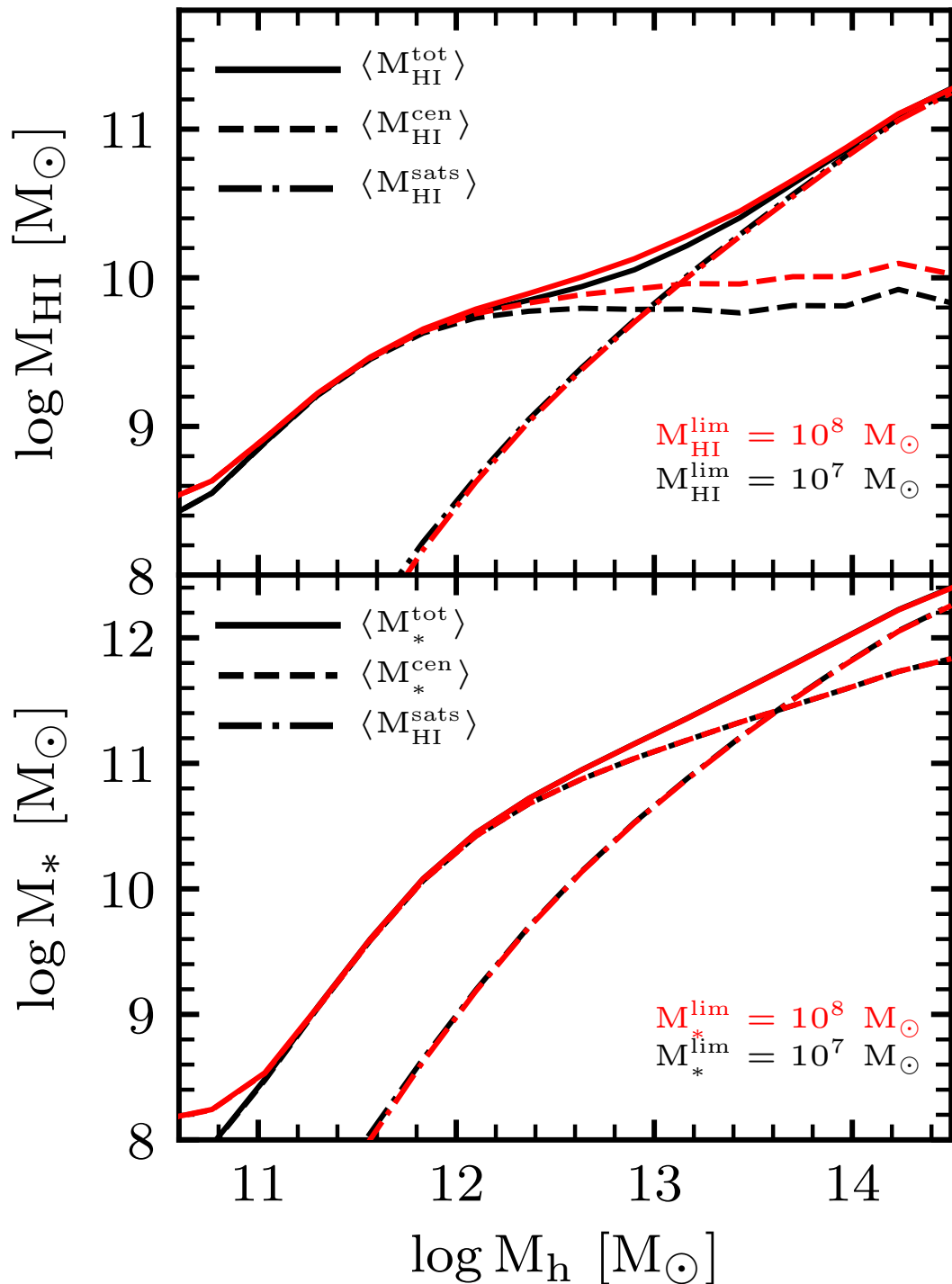
Figure 6.8: Mean arithmetic *total* H I mass (top panel) and *total* stellar mass (bottom panel) within a halo as a function of its virial mass (solid lines). The dashed and dot-dashed lines show the respective H I and stellar masses contained only in the central galaxy and the sum of all satellites within the halo, respectively. Results for two threshold H I/stellar masses are shown (see labels in the panels).
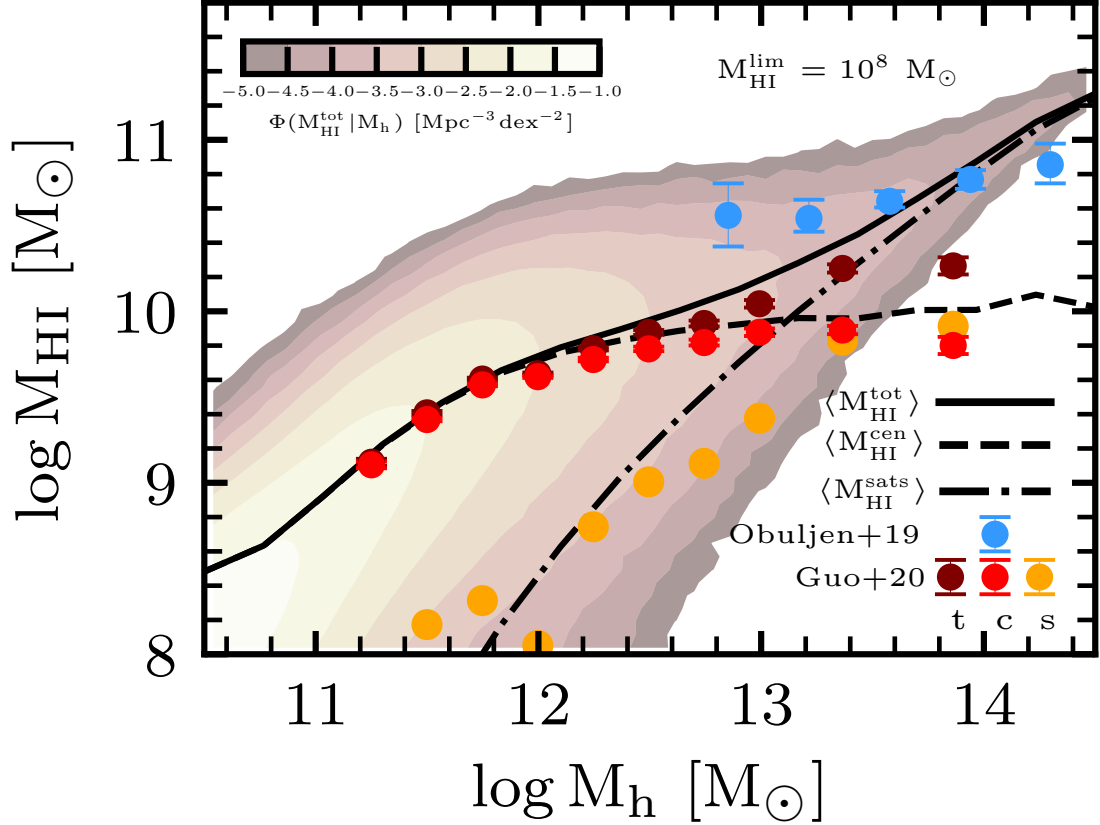
Figure 6.9: Same as top panel of Figure 6.8, but now we compare our results for $M_{\rm HI}^{\rm lim} = 10^8$ $M_\odot$ with Guo et al. (2020) and Obuljen et al. (2019) observational inferences. Note that Guo et al. (2020) inferred not only the total H I content within haloes but also the contributions from the central. The isocountours correspond to the full bivariate distribution.

statistics, due to the too low 21-cm line fluxes of many satellites as to contribute to the H I stacked spectra above the allowed instrumental signal-to-noise ratio or due to the use of a too narrow window in the allowed H I velocity in such a way that not all satellites are captured (Dr. Claudia Lagos, priv. communication). On the other hand, the determinations for massive groups in Obuljen et al. (2019), roughly agree with our results at the high halo mass range. Unlike Guo et al. (2020), Obuljen et al. (2019) do not directly measure the H I content of haloes, but instead use empirical relations to derive it. To determine $M_{\rm HI}^{\rm tot}$ as a function of group (halo) mass they integrate the ALFALFA galaxy H I mass functions at different group masses.

6.3.2   The H I galaxy spatial clustering

We have discussed already that when implementing the SHAM for $V_{\rm DM}$ instead of $M_{\rm DM}$, the projected 2PCFs, $\omega_p^*(r)$, from the SDSS DR7 galaxies in stellar mass bins are recovered, see Figure 6.2. Next, we present our predictions for the projected
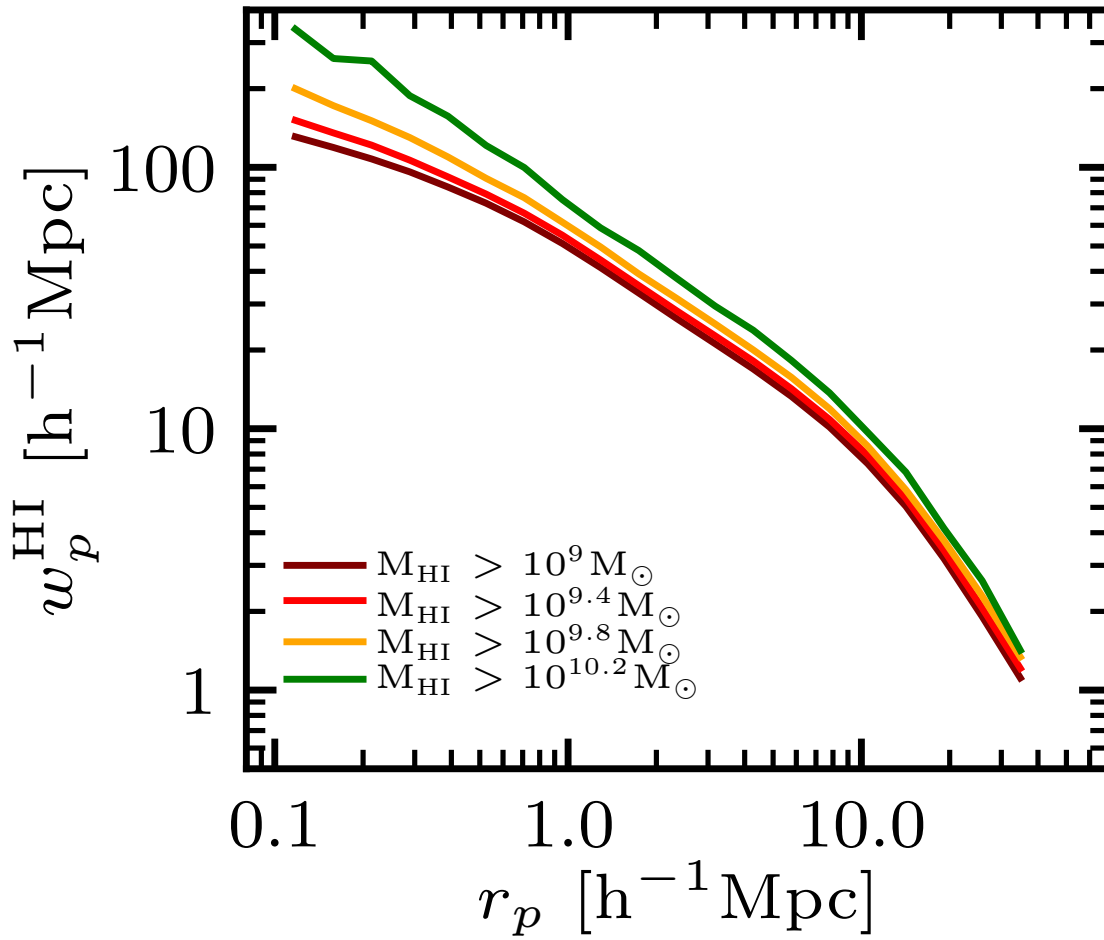
Figure 6.10: Projected H I 2PCFs from the galaxy mock catalog measured above four M_{HI} thresholds as indicated by the labels.

2PCFs as a function of M_{HI}.

Figure 6.10 shows the resulting H I projected 2PCFs, $\omega_p^{\mathrm{HI}}(r)$, for various M_{HI} thresholds from our mock galaxy catalog. In general, the amplitude of $\omega_p^{\mathrm{HI}}(r)$ is almost independent of the M_{HI} threshold. In more detail, however, we notice that the amplitude of the one-halo term increases with M_{HI} while the two-halo term is almost independent of M_{HI}, and in consequence, the resulting projected 2PCFS are not power laws. The above can be broadly understood as the result of the differences between the $M_{\mathrm{HI}} - V_{\mathrm{DM}}$ (or $M_{\mathrm{HI}} - M_{\mathrm{DM}}$) relations of centrals and satellites and the flattening with similar values of the mean $\langle \log M_{\mathrm{HI}} \rangle$ for both relations at the halo high-mass. Note that in low mass thresholds apparently central galaxies contribute more to $\omega_p^{\mathrm{HI}}(r)$ than satellites but for high mass thresholds, centrals and satellites contribute equally to $\omega_p^{\mathrm{HI}}(r)$. This behaviour of $\omega_p^{\mathrm{HI}}(r)$ is very different to the one of $\omega_p^*(r)$. In particular, $\omega_p^*(r)$ depends strongly on mass.

There are other features related to the H I projected 2PCFs that are worth of mentioning. We begin by emphasising the fact that the 2PCFs mass thresholds used in Fig. 6.10 are most of the time above the mean $M_{\mathrm{HI}} - V_{\mathrm{DM}}$ and $M_{\mathrm{HI}} - M_{\mathrm{DM}}$ relations, see Fig. 6.5. In other words, the larger the H I mass threshold, the larger the distance from the mean $\langle \log M_{\mathrm{HI}} \rangle$ values. The above is interesting for various reasons. First, increasing the H I mass threshold implies that we are sampling *rare* H I masses for a given V_{DM} or M_{DM}. Second, similar mass thresholds have been used in the past to derive the H I galaxy-halo connection (see e.g., Guo et al., 2017). Recall that HOD (and related) models assume that galaxy properties, in this case H I mass, are *totally* determined by a halo property as V_{max} or M_h. The lower panels of Figure 6.5 show that this is not the case. In view of the above, it is thus relevant to ask whether HOD (and similar) models are the *appropriate* tools to constrain the *real* H I-to-halo mass relation. We will come back to this point later in this Chapter.

Figure 6.11 shows again the H I projected 2PCFs, $\omega_p^{\mathrm{HI}}(r)$, for different M_{HI} thresholds as indicated in the panels. The thick red lines correspond again to the measurements from our mock catalog. One of the key aspects in our mock catalog is that we allowed the H I mass conditional distributions for a given M_*, both for LTGs and ETGs, to be different for centrals and satellites as observations show, see Section 6.2.4. The dashed black lines in Figure 6.11 show the resulting $\omega_p^{\mathrm{HI}}(r)$ functions from our mock catalog when assuming that the H I distributions are the same for centrals and satellites, that is, the ratio in Eq. (6.2.6) is set equal to 1. The fact that satellite galaxies have lower H I gas fractions than centrals clearly works in the direction of lowering $\omega_p^{\mathrm{HI}}(r)$ at small distances, at the 1-halo term.

In Figure 6.11, we compare our results to the corresponding $\omega_p^{\mathrm{HI}}(r)$ functions from
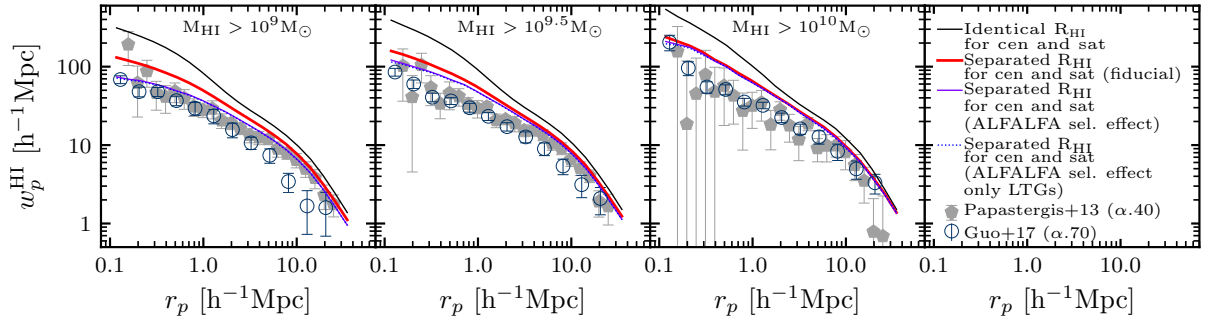
Figure 6.11: Projected H I 2PCFs measured from different M_{HI} thresholds in the galaxy mock catalog. The solid red line is our prediction for all local galaxies. This is the H I spatial clustering expected from a large-volume and deep H I blind survey. The solid black line shows the projected 2PCFs from our approach but assuming the same H I gas content for central and satellite galaxies. The violet solid line corresponds to the projected 2PCFs measured in the catalog after emulating the ALFALFA survey selection effect (see text), while the blue dotted line is when the condition of not taking into account ETGs is also included. The symbols with error bars correspond to observational measurements from ALFALFA, see labels in the right panel.

the ALFALFA α.40 survey (Haynes et al., 2011) as measured by Papastergis et al. (2013, pentagons) and from the ALFALFA α.70 survey as measured by Guo et al. (2017, open circles). Beyond the differences between both studies, it is evident that our predicted H I 2PCFs are above of both studies. At this point, it is important to recall that the ALFALFA survey suffers from selection effects, see Introduction. The main reason for these effects is related to the low detection limit in radio flux and its dependence on the H I line width (Haynes et al., 2011; Huang et al., 2012c; Chauhan et al., 2019). As a result, the R_{HI}–M_* relation of ALFALFA is biased as it observes only H I-rich galaxies, avoiding galaxies with intermediate to low values of R_{HI} (Huang et al., 2012c; Maddox et al., 2015, see Figure 4.4 of Chapter 4 to see the above). H I-poor galaxies tend to be earlier types and reside in denser environments, hence, they are expected to be more clustered than H I-rich galaxies. Therefore, the observed low clustering of ALFALFA observations as compared to our predictions is most likely due to the lack of H I-poor highly-clustered galaxies in the ALFALFA survey. The $\omega_p^{\mathrm{HI}}(r)$ functions from our mock galaxy catalogue showed in Figure 6.11 are predictions to what will be measured in future deep and large H I surveys, not affected by significant selection effects.

Section 6.2.5 describes a simple way to emulate the non-trivial selection effects of the ALFALFA survey. The violet lines in Figure 6.11 show the resulting $\omega_p^{\mathrm{HI}}(r)$ functions from our mock catalog after applying this ALFALFA-like selection. The functions approximate to those reported from ALFALFA, in special for the low M_{HI} threshold. Nonetheless, they still have lower amplitudes than our predictions.
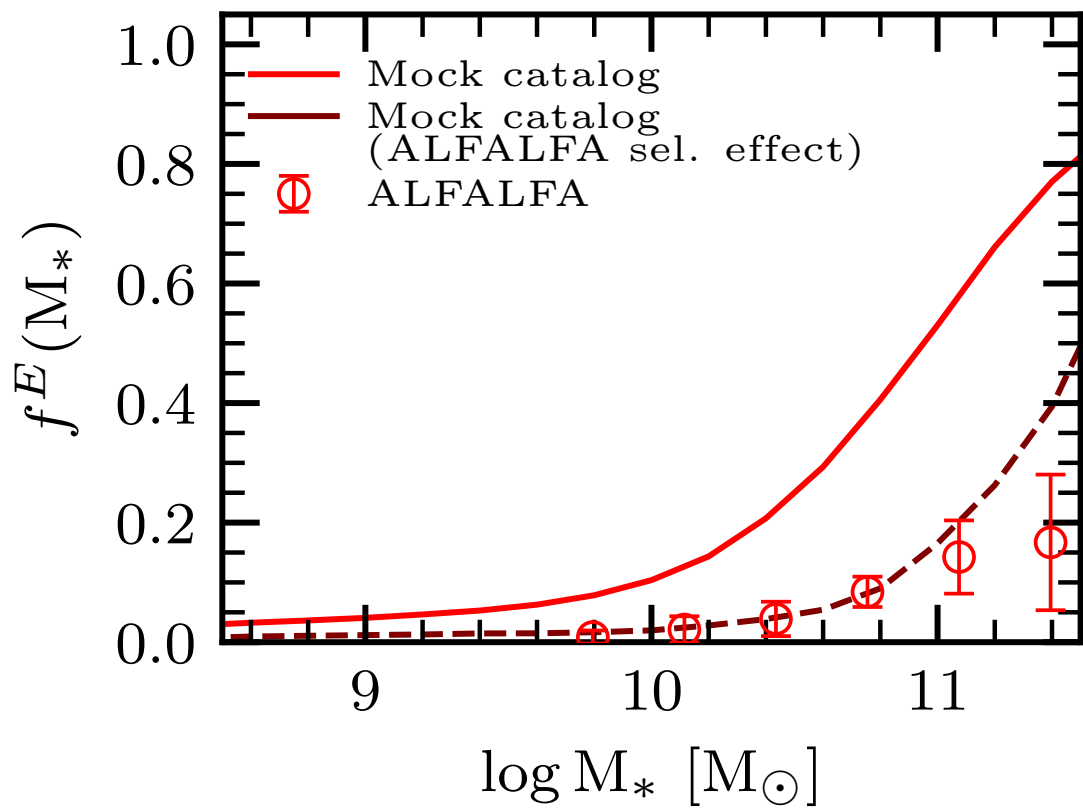
Figure 6.12: Fraction of ETGs as a function of M_* as measured in the complete galaxy mock catalog (solid line) and after imposing the ALFALFA-like selection (dashed line). The corresponding fractions measured in the ALFALFA survey in many stellar mass bins are showed with circles and error bars.

As it is well known, optically selected and H I-selected galaxies do not fully overlap, such that H I-selected surveys miss the most massive, gas-poor galaxies, which typically are ETGs. As mentioned in the Introduction (see for references therein), ALFALFA indeed is strongly biased against gas-rich, blue galaxies. Thus, by imposing the ALFALFA-like selection effect in our mock catalog, the fraction of ETGs is expected to decrease. Indeed this happens, but when comparing our fraction of ETGs as a function of M_* with ALFALFA, the latter shows lower fractions. In Figure 6.12 we present this comparison. As seen, the fraction of ETGs, $f^E(M_*)$, in the ALFALFA survey is very low even at high masses. This is because massive ETGs have both low H I content and large W_{50} equivalent widths both of which act against detection in this and other surveys (see Obreschkow et al., 2013; Chauhan et al., 2019). In addition to the ALFALFA-like selection effect we assume that ALFALFA observes LTGs only. The resulting measured $\omega_p^{\rm HI}(r)$ are plotted in Figure 6.11 with blue dotted lines. They are almost indistinguishable from the previous case.

Our analysis shows that while the selection effects introduced by the shallowness of a blind H I survey, like ALFALFA, is not crucial for measuring the H I mass function (our H I mass function is consistent with ALFALFA, see Fig. 6.4), these selection effects become critical for the H I spatial clustering (as well as for the H I-to-stellar mass correlation, see Chapters 2 and 4). In §§6.4.2 we discuss on the caveats of our approach and how, by overcoming them, we could improve our comparison with the galaxy H I clustering from ALFALFA.

6.4 Discussion

6.4.1 H I mass is not determined by the halo scale: Implications for SHAM and HOD

The assumption that the halo mass/maximum circular velocity determines the properties of the galaxies is central in galaxy-halo models such as the HOD and SHAM. Previous works have shown that, at a good approximation, this is the case for galaxy properties such as luminosity and stellar mass (for a discussion and references see Dragomir et al., 2018). As discussed in Section 6.3.2, when using $M_{\rm HI}$ the central assumption of SHAM or HOD models is clearly infringed as can be seen in the lower panels of Fig. 6.5. Thus expecting to derive a $M_{\rm HI}$–$V_{\rm DM}$ (or $M_{\rm HI}$–$M_{\rm DM}$) relationship from these models is fruitless. In particular, it is clear that, by construction, SHAM will tend to predict an increasingly monotonic mean relation between $M_{\rm HI}$ and $V_{\rm DM}$ (or $M_{\rm DM}$); a relation that will struggle to reproduce the observed clustering from H I-

surveys as discussed in §§6.3.2. In this section we elaborate more this argument by deriving some $M_{\rm HI}$–$V_{\rm DM}$ relations from SHAM and studying their resulting clustering in H I mass.

Figure 6.13 presents three different $M_{\rm HI}$–$V_{\rm DM}$ relations obtained by means of SHAM (see section 6.2.2) using the HI mass function from Chapter 3. We assume that the scatter around the mean $M_{\rm HI}$-$V_{\rm DM}$ is lognormally distributed and use three different values for the scatter: $\sigma =$0.15 (blue line), 0.40 (green line), and 0.60 (cyan line) dex. The black solid line and the shaded area reproduce our predicted mean $M_{\rm HI}$–$V_{\rm DM}$ relation and its scatter as showed in Fig. 6.5. It is clear that SHAM predicts that the average relations increase monotonically with $V_{\rm DM}$. Notice, however, that by increasing the value of the scatter, the SHAM results approach to our $M_{\rm HI}$–$V_{\rm DM}$ relation. Nonetheless, by construction, SHAM will always introduce a tight correlation between $M_{\rm HI}$ and $V_{\rm DM}$ or M_h regardless of the scatter assumed.

Figure 6.14 shows the resulting H I clustering based on our experiments from SHAM. We note that there are two potential flaws that are evident for SHAM. Firstly, the fact that SHAM does not separate between centrals and satellites. As discussed in Section 6.2.4 this assumption is problematic as it is similar to ignore environmental effects for satellite galaxies, so it is not surprising that the one halo-term is overestimated. Secondly, since SHAM introduces a monotonic correlation between $M_{\rm HI}$ and $V_{\rm DM}$, the amplitude of the two-halo term will always increase with $M_{\rm HI}$, thus overestimating the two-halo term. Note that as decreasing the scatter, the tighter and stronger is the correlation with $V_{\rm DM}$, and thus the larger the amplitude of the 2PCFs.

The left panel of Figure 6.15 is as Figure 6.13 but for the halo mass $M_{\rm DM}$. In the case of the relations obtained by means of traditional SHAM, for passing from $V_{\rm DM}$ to M_h, we used the tight relation between these two quantities as measured in N-body simulations (e.g., Rodríguez-Puebla et al., 2016). We reproduce in this plot also a previous determination of the H I-to-halo mass relation with SHAM using the observed H I mass function, Padmanabhan & Kulkarni (2017, green solid line). The SHAM implemented by these authors did not include scatter; this is why their H I-to-halo mass relation looks similar to our case with small scatter, blue line. As seen, this relation is far from what we actually found with our semi-empirical model.

In the past, some authors used in different ways the H I spatial clustering information provided by current blind H I surveys for constraining the galaxy H I-halo connection by means of HOD or SHAM models (e.g., Padmanabhan et al., 2017; Guo et al., 2017; Obuljen et al., 2019). As discussed in §§6.3.2, the H I projected 2PCFs are measured for galaxies with masses larger than a given $M_{\rm HI}$ threshold, and the values of this threshold are typically above the mean H I-to-halo mass relation as
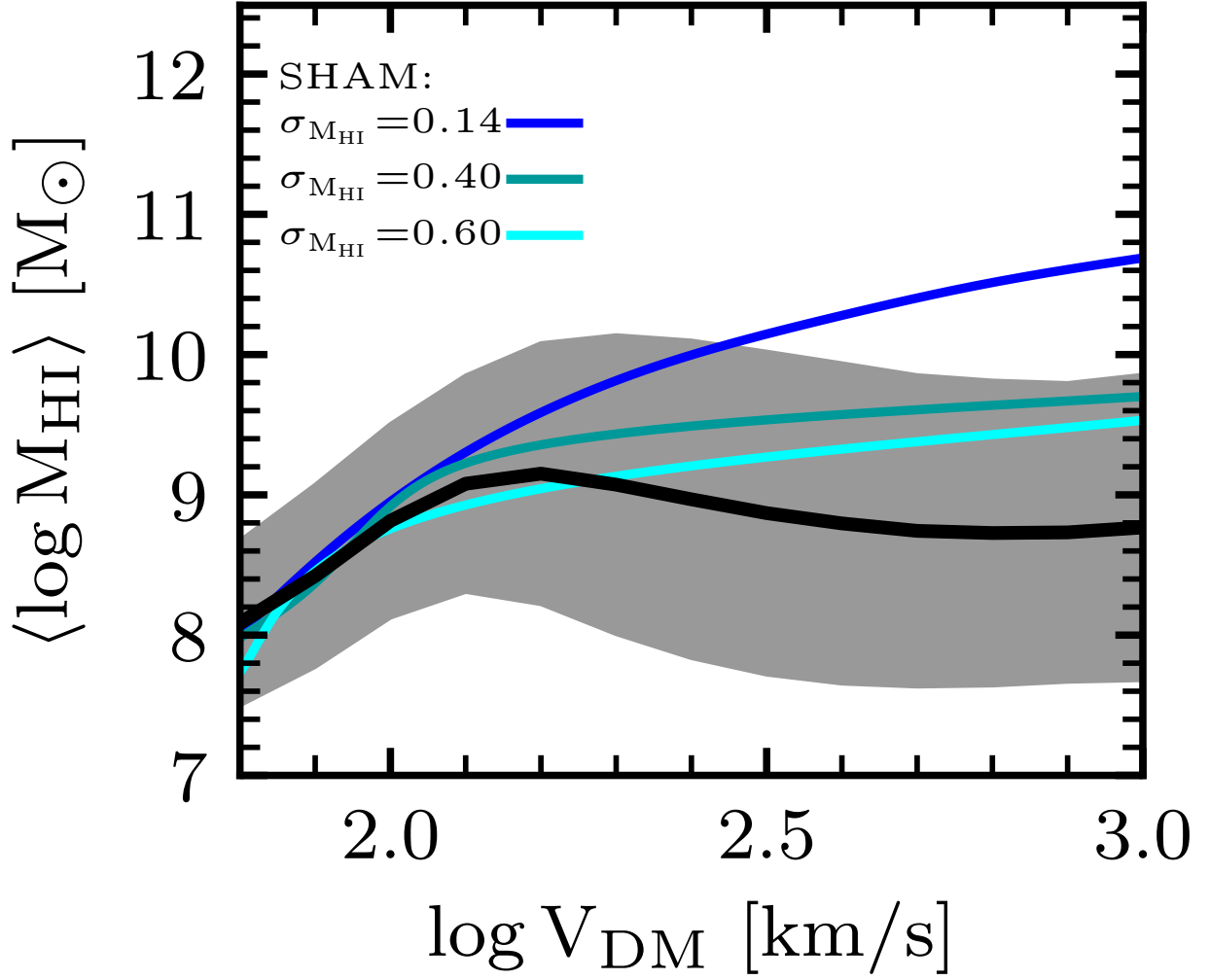
Figure 6.13: SHAM results using the observed H I Mass Function for three different scatters assumed for the M_{HI}–V_{DM} relation (see labels). The black solid line and shaded area are our result as shown in Figure 6.5.
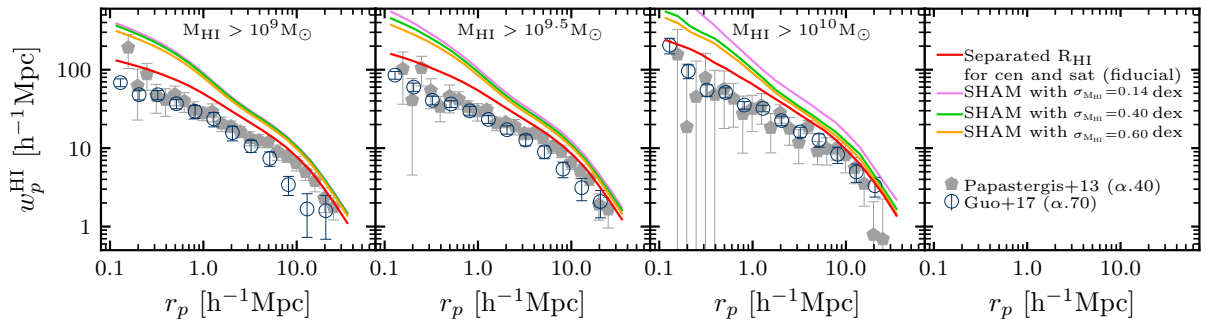

Figure 6.14: As Figure 6.11 but using the different SHAM models shown in Fig. 6.13. The solid line is as in Figure 6.11 and it corresponds to the prediction from our galaxy H I-(sub)halo connection.
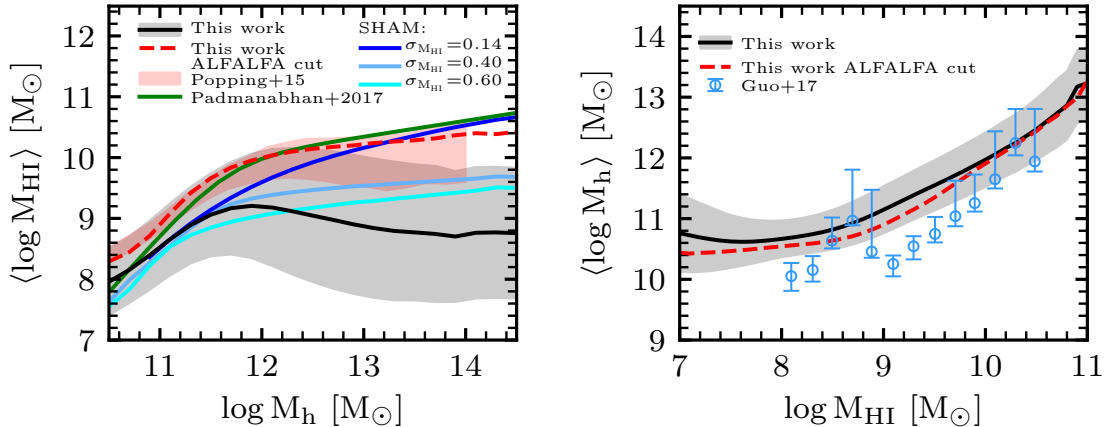
Figure 6.15: *Left panel:* Logarithmic mean M_{HI}–M_{DM} relation and its scatter as shown in Fig. 6.5 (black solid line and gray shaded area) and the resulting relation when applying the ALFALFA-like selection criterion (dashed red line), which selects H I-rich galaxies. The results from the SHAM models presented in Fig. 6.13 are also shown here. The green solid line is for the SHAM model (without scatter) from Padmanabhan et al. (2017), and the pink shaded area is for the model for star-forming galaxies presented in Popping et al. (2015). *Right panel:* The inverse mean M_{HI}–M_{DM} relation. Line symbols are as in the left panel. The circles with error bars correspond to result from a SHAM model extended to introduce assembly bias and constrained with the ALFALFA H I clustering (Guo et al., 2020).

determined here, implying that the measured 2PCFs correspond actually to a *biased* galaxy population. On the other hand, it is well known that H I blind surveys are strongly biased to gas-rich, blue galaxies. Actually, both shortcomings are related because the minimum H I mass threshold is determined by the H I mass completeness limit of the survey, and the latter is partially related to the H I survey sensitivity, which introduces a bias against gas-poor, red galaxies (typically ETGs).

In Figure 6.15 we show the M_{HI}–M_{DM} relation (left panel) and its inverse, the M_{DM}–M_{HI} relation (right panel), as measured from our catalog when imposing the ALFALFA-like selection, red dashed lines. Recall that with this selection, the $\omega_p^{\mathrm{HI}}(r)$ function with a threshold of $M_{\mathrm{HI}} = 10^9$ M$_\odot$ resulted close to the ALFALFA measures (Fig. 6.11). For comparison, the respective relations along with their standard deviations as measured from our galaxy mock catalog without imposing any selection criterion, are also plotted in Figure 6.15. An implication of this Figure and Figure 6.11 is that by (roughly) reproducing the ALFALFA H I spatial clustering, the galaxy H I-halo connection results biased to high M_{HI} values at a given halo mass or to low halo masses at a given M_{HI}. In the latter case, the difference would be larger if the ALFALFA H I clustering would be better reproduced by assuming that the scatter around the M_*–V_{DM} relation depends on morphology (see §§6.4.2 below).

In the right panel of Figure 6.15, the results from Guo et al. (2017) are reproduced.

These authors constrained their extended SHAM model with their measures of $\omega_p^{\mathrm{HI}}(r)$ from ALFALFA. They found that their simple SHAM model predicts much higher H I clustering than the measurements from ALFALFA. Then, they extended the model by introducing additional halo parameters, and found that the ALFALFA clustering is reproduced in a model that puts H I-rich galaxies into halos that formed late (assembly bias). We notice that their SHAM refers only to distinct haloes populated by central galaxies. By comparing to our results, we can conclude that introducing the assembly bias as a way to help to reproduce the H I clustering of ALFALFA it implies a M_{DM}–M_{HI} relation biased to low mass haloes.

Finally, the left panel of Figure 6.15 presents the results based on the more physically motivated inferences of H I mass from Popping et al. (2015) for star-forming galaxies. In order to infer H I gas mass, the authors used the star formation histories from SHAM and the inverted star formation rate–surface density relations to infer galaxy HI masses. It is interesting how our M_{HI}–M_{DM} relation for gas rich galaxies (due to the imposed ALFALFA-like selection, dashed line) is similar to the one of those authors.

6.4.2   Caveats and the galaxy H I clustering of ALFALFA

Morphological classification

In this Chapter we use the observed fractions of ETG for centrals and satellites based on the automatic morphological classification from Huertas-Company et al. (2011). As discussed in Section 6.2.3, we introduce galaxy morphology just as a necessary step in order to assign H I masses to our galaxies. Thus, a potential concern is the effects of using alternative morphological classifications in our methodology. To study the above, we use next an alternative morphological classification for SDSS galaxies.

Based also on SDSS, Domínguez Sánchez et al. (2018) presented a new automatic morphological classification of galaxies. As shown in Figures E.1 and E.2 of Appendix E the Domínguez Sánchez et al. (2018) classification implies a higher fraction of both central and satellite ETGs than Huertas-Company et al. (2011) up to $M_* \sim 10^{11}\ \mathrm{M_\odot}$. When we use the fractions from Domínguez Sánchez et al. (2018) to assign H I masses, we note that, as expected, some of the results for the H I-halo connection changed. Figure 6.16 shows the logarithmic mean M_{HI}–M_{DM} relation for all galaxies and for centrals and satellites using both the Huertas-Company et al. (2011) (same line code as in Fig. 6.5) and Domínguez Sánchez et al. (2018, the lower extreme of the shades areas) morphological classifications. For the latter classification, the logarithmic mean values of M_{HI} result lower than for the former classification, in particular for subhaloes

(satellite galaxies). This is expected because, as mentioned above, Domínguez Sánchez et al. (2018) classify a larger fraction of ETGs, specially in the case of satellites, than Huertas-Company et al. (2011), and ETGs contain less H I than LTGs of the same stellar mass. The shaded areas in this figure can be considered as an uncertainty in our inferences due to the uncertainty in the galaxy morphological classification. Similarly, in Figure 6.17, we plot the H I projected 2PCFs from our mock catalog using both morphological classifications. Inferences using Huertas-Company et al. (2011) and Domínguez Sánchez et al. (2018) are plotted with solid red and dashed cherry lines. The differences seen are actually smaller than those found when applying the ALFALFA selection, and hence they would be second order effect. We conclude that our inferences may depend on the assumed morphological classification –as well as on the separation criterion– for defining LTG and ETGs but the main trends and conclusions are unaffected by this.

The comparison to the ALFALFA H I clustering

In Section 6.3.2, we showed that our mock galaxy catalogue tends to overpredict the observed H I clustering from the ALFALFA survey, even after imposing an ALFALFA-like selection effect and in the extreme case of assuming that galaxies in this survey consist only of late-types. Thus a potential concern is on our assumptions in assigning morphology to our galaxies seeded in the simulation haloes. As explained in Section 6.2.3, we assumed that the scatter around the M_*–V_{DM} relation is independent of galaxy morphology. We are aware that this assumption implies that clustering properties of ETGs and LTGs are almost identical, at least for the two-halo term. Thus, by construction one expects that our mock catalogue will not recover the well known clustering properties of ETGs and LTGs as a function of stellar mass. This shortcoming in combination with the ALFALFA selection effects may work in the direction of overpredicting the H I clustering. There are at least three options that are straight forward to solve the above:

1. The scatter around the M_*–V_{DM} relation depends on morphology.- In this assumption, ETGs and LTGs occupy haloes of different masses. In particular, if there is a strong correlation between halo and morphology in which haloes of ETGs are more massive than those of LTGs at fixed M_*, then the clustering of ETGs will be stronger due to the halo bias (see e.g., Rodríguez-Puebla et al., 2015).

2. Assembly bias.- Halo masses of ETGs and LTGs could be identical but their properties are different. If galaxy morphology strongly correlates with halo's
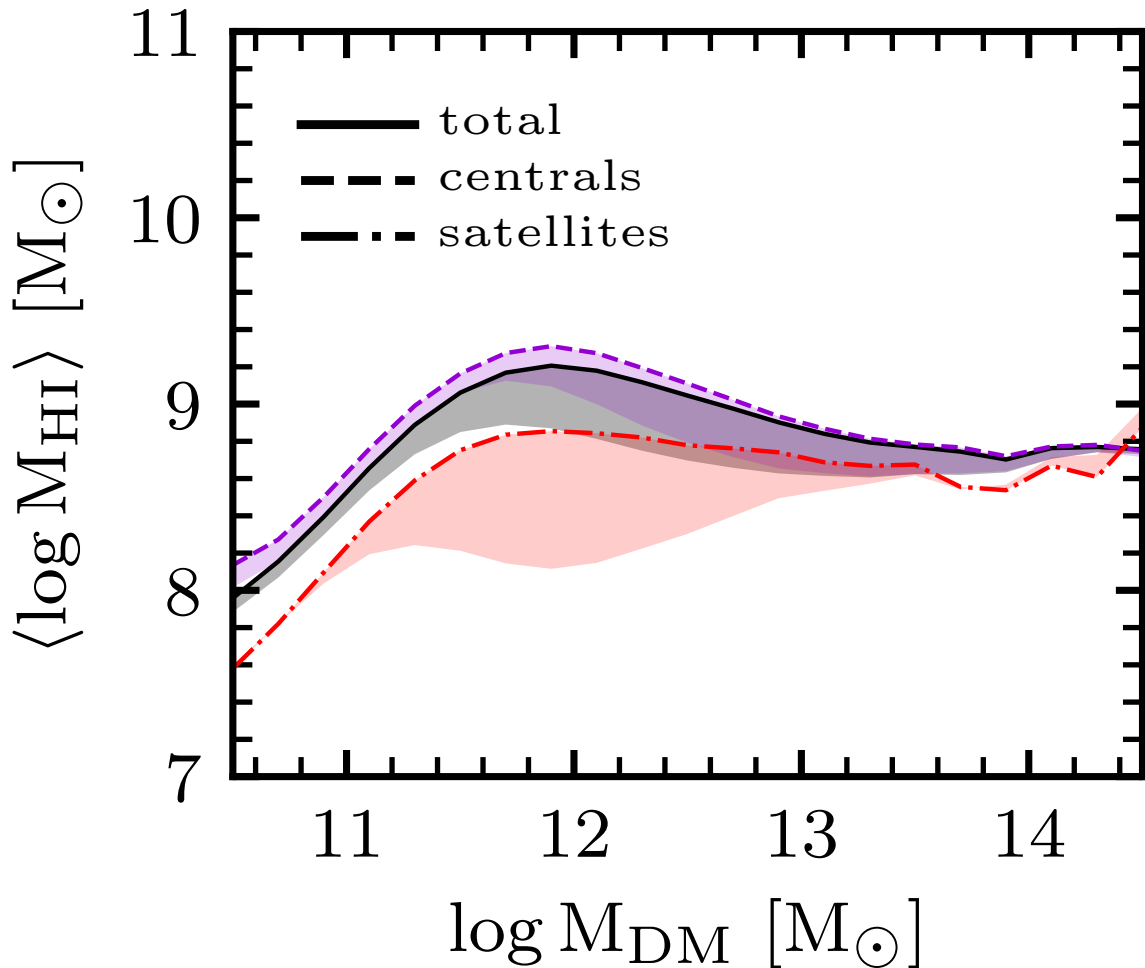
Figure 6.16: The lines show the logarithmic mean M_{HI}–M_{DM} relations for all (black), central (violet), and satellite (red) galaxies using the Huertas-Company et al. (2011) morphological classification for defining LTGs and ETGs, as shown in Figure 6.5. The lower extremes of the shaded areas show the same relations but when using the Domínguez Sánchez et al. (2018) morphological classification. The latter relations are below the former ones, specially for satellites. The shaded area can be considered as the uncertainty in our method due to differences in the morphological classification.

Figure 6.17: Projected H I 2PCF for $M_{HI} > 10^9$ M$_\odot$ for our fiducial model using the Huertas-Company et al. (2011, red solid line) and the Domínguez Sánchez et al. (2018, dashed line) morphological classifications. The observations are as in Figure 6.11

assembly time, such as formation time, then the clustering of ETGs will be stronger due to the halo assembly bias (see e.g., Hearin & Watson, 2013b).
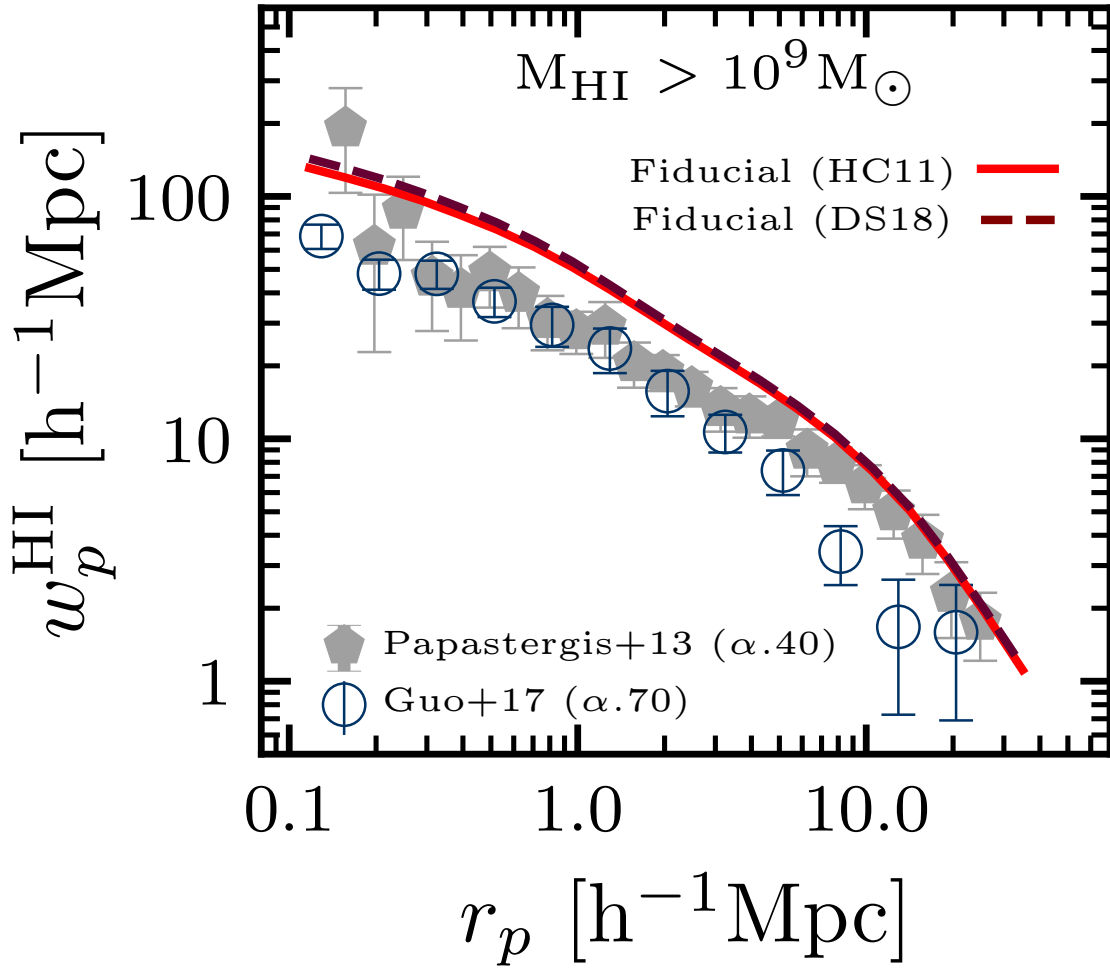
3. A combination of both.

Studying the above options is beyond the scope of this Thesis, for which we have focused to study galaxy population as *whole*. While we will study in more detail the dependence of the galaxy-halo connection with galaxy morphology in a forthcoming work, this is actually irrelevant as it is obvious that any dependence will be wash-out (averaged) when studying the whole population.

Finally, the $\omega_p^{\mathrm{HI}}(r)$ functions measured in future large and deep H I surveys, which will not be affected by significant selection effects, are expected to be similar to those reported in Section 6.3.2.

6.4.3   Comparisons to theoretical predictions

New generation semi-analyitic models (SAM) and cosmological hidrodynamics simulations, after post-processing, are able to predict the total galaxy H I gas content within the virial radius of haloes. In Figure 6.18 we show the *median* of the total H I mass, $M_{\mathrm{HI}}^{\mathrm{tot}}$, a as function of M_h for several SAMs (Kim et al., 2017; Baugh et al., 2019; Spinelli et al., 2020; Chauhan et al., 2020) and the Illusttris-TNG100 simulation (Stevens et al., 2019; Chauhan et al., 2020). We have homogenized the halo masses to the virial mass. Our semi-empirical inference for the median $M_{\mathrm{HI}}^{\mathrm{tot}}$–$M_h$ relation is showed with the red thick line.

Our results show a relatively smooth increasing of $M_{\mathrm{HI}}^{\mathrm{tot}}$ with M_h in the $3 \times 10^{11} - 10^{13}$ M$_\odot$ mass range, similar to what predict the GAEA SAM (Spinelli et al., 2020) and the Illustris-TNG100 simulation, and in tension with other SAMs: an old version of GALFORM Kim et al. (2017), its new version (Baugh et al., 2019), and SHARK (Chauhan et al., 2020). The decrease of $M_{\mathrm{HI}}^{\mathrm{tot}}$ predicted by these models at the halo masses where the central dominates in the total H I content of haloes, is associated mainly to the AGN feedback. The feedback, mainly mechanical, keeps the gas in the halo hot, preventing this the accretion of cold gas to the galaxy. The major difference is with GALFORM (Baugh et al., 2019) despite that we find an excellent agreement regarding their total M_* in haloes, see lower panel of Figure 6.18. The AGN feedback strength affects more drastically the H I gas content of their galaxies than the stellar masses. Thus, the H I gas fraction of galaxies seems to be an important constrain for the AGN feedback in models and simulations, along with the high-mass end of the GSMF.
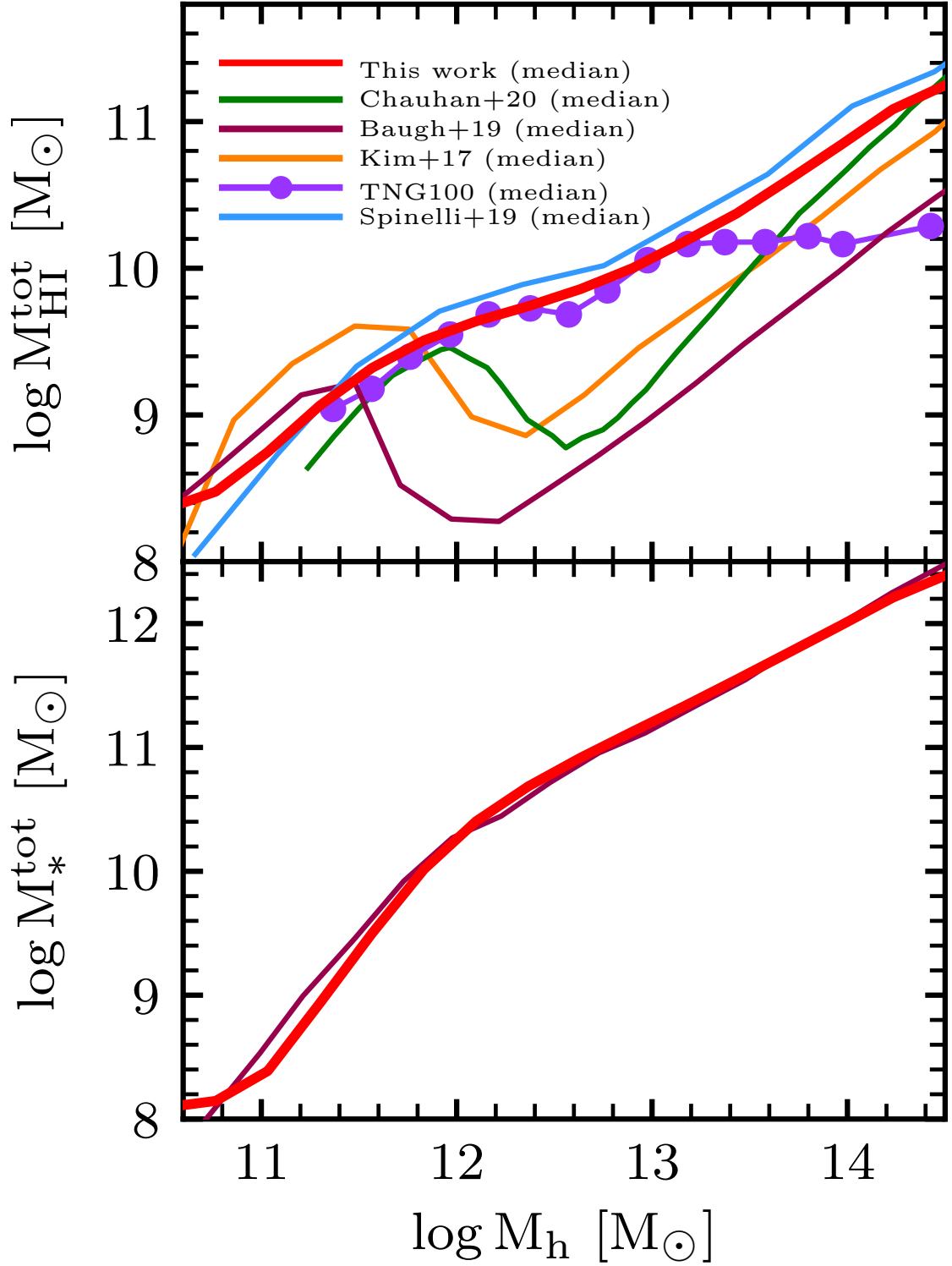
Figure 6.18: *Upper panel:* Median total H I mass inside haloes of mass M_h as empirically determined here (red line) compared to several predictions from SAMs and hydrodynamics simulations. *Lower panel:* As upper panel but for the total stellar mass compared to the SAM results from Baugh et al. (2019).

6.5 Summary and Conclusions

In this Chapter we have presented a semi-empirical approach to link the H I gas content of galaxies given their stellar masses to dark matter (sub)haloes in the `SMDP` N-body cosmological simulation. Galaxies are initially linked to dark matter (sub)haloes via the M_*–V_{DM} relation[7] derived from SHAM. We assume that galaxies are lognormally distributed around the M_*–V_{DM} relation with a dispersion of 0.15 dex. To every galaxy in the catalog we assign either an early- or late-type morphology based on the observed fractions of ETGs for centrals and satellites from the SDSS Yang et al. (2012) group catalog and the Huertas-Company et al. (2011) morphologies. Finally, we assign a H I mass using the central/satellite M_{HI} conditional PDFs given M_* of ETGs and LTGs derived from observations in Chapters 2 and 5. Thus, for every halo or subhalo in the `SMDP` simulation we have assigned a stellar mass, galaxy morphology, and H I mass. We emphasis that the morphology assignation is a *necessary* step for sampling the M_{HI} distributions. In this Chapter, we are interested in studying the galaxy stellar-H I-halo connection for the galaxy population as a *whole* and not in its segregation by morphology. Our main results are as follows:

- The value of $\langle \log M_{\mathrm{HI}} \rangle$ for the whole population as a function of V_{DM} (M_{DM}) monotonically increases up to $V_{\mathrm{DM}} \sim 160$ km/s ($M_{\mathrm{DM}} \sim 10^{12}$ M$_\odot$) where it reaches a maximum value of $\langle \log(M_{\mathrm{HI}}/\mathrm{M}_\odot) \rangle \sim 9.2$, Fig. 6.5. At higher (sub)halo velocities (masses) it decreases only slightly. This is in contrast to $\langle \log M_* \rangle$ which always increases monotonically as a function of V_{max} (M_h).

- The scatter around $\langle \log M_{\mathrm{HI}} \rangle$ increases with V_{max} (M_{DM}). At low (sub)halo velocities (masses), $V_{\mathrm{max}} \sim 80$km/s ($M_{\mathrm{DM}} \sim 10^{11}M_\odot$) this is ~ 0.5 dex. At $V_{\mathrm{DM}} \gtrsim 160$ km/s ($M_{\mathrm{DM}} \gtrsim 10^{12}M_\odot$) it increases rapidly reaching then a maximum value of ≈ 1.2 dex.

- In general, the M_{HI} conditional PDFs as a function of V_{max} (M_{DM}) are broad and highly asymmetric with a long tail towards lower values of M_{HI}. For $V_{\mathrm{max}} \gtrsim 160$ km/s ($M_{\mathrm{DM}} \gtrsim 10^{12}M_\odot$), the PDFs are bimodal with a second peak appearing at lower values of M_{HI}, a behaviour simply inherited from the input H I conditional PDFs as function of M_*. As a result, different statistical estimators as logarithmic mean, arithmetic mean, and median of M_{HI} as a function of V_{max} (M_{DM}) differ among them (Fig. 6.6).

---

[7]Here we assumed that V_{DM} corresponds to the maximum circular velocity for distinct haloes, whereas for subhaloes it is the peak maximum circular velocity reached along the halo's main progenitor branch, see Eq. (6.2.1).

- There are not significant differences in $\langle \log M_{\rm HI} \rangle$ as a function of $V_{\rm DM}$ or $M_{\rm DM}$ between centrals and satellites. On average, satellites have slightly lower values of H I mass at a given $V_{\rm DM}$ or $M_{\rm DM}$ than centrals (Fig. 6.5). The scatter is broader for satellites than for centrals.

- The H I projected 2PCFs from our mock catalog increase only slowly in amplitude as the H I mass threshold increases (Fig. 6.10).

- Assuming identical H I mass conditional PDFs for centrals and satellites results in 2PCFs that have higher amplitudes, specially at the one-halo term, than in our more realistic fiducial model, where satellites have lower H I gas fractions than centrals, see Figs. 6.7 and 6.10.

- Our predicted H I projected 2PCFs have higher amplitudes than the ones measured in the blind H I ALFALFA survey, see Fig. 6.11. When emulating AL-FALFA selection effects in our catalog, which selects gas rich galaxies, the projected 2PCFs are similar, but yet slightly above than ALFALFA.

- SHAM is unable to reproduce realistic 2PCFs (Fig. 6.14) as it assumes identical $\langle \log M_{\rm HI}(V_{\rm DM}) \rangle$ relations for centrals and satellites and predicts that it monotonically increasing.

- The total galactic (central + satellites) H I gas content in distinct haloes, $M_{\rm HI}^{\rm tot}$, strongly depends on M_h and is completely dominated by centrals up to $\sim 10^{12}$ $M_\odot$, while at higher masses the contribution from satellites increases; at $M_h \sim 10^{13}$ $M_\odot$ both contributions are roughly equal and a $M_h = 10^{14}$ $M_\odot$ the H I in satellites is larger by > 1 dex (Fig. 6.8). The mean $M_{\rm HI}^{\rm tot}$–M_h relation does not present a dip at the masses where AGN feedback is expected to be important.

The results presented here offer new constraints on the empirical galaxy-(sub)halo connection both for stellar and H I masses. These results are relevant for calibrating and testing the predictions from new generation SAM and hydrodynamics simulations of galaxy evolution within the ΛCDM cosmology. We have presented some preliminary comparisons with some of these theoretical predictions and discussed that our semi-empirical results can be relevant to constrain the strength of the AGN feedback.

Our results on the galaxy H I-(sub)halo connection show that $M_{\rm HI}$ is not totally determined by the halo properties $V_{\rm max}$ or $M_{\rm DM}$. The above has deeply implications as it infringes the central assumptions in the galaxy-halo connection models such as the HOD and SHAM. Thus, alternative tools, like the one used here, should be used in order to derive a realistic H I-to-halo mass relation. We also found that the

H I conditional PDFs as a function of V_{max} or M_{DM} are highly asymmetric and even bimodal. In view of the above, to characterize the dependence of $M_{\mathrm{H\,I}}$ on M_{DM} at large masses is better to use the median of $M_{\mathrm{H\,I}}$ as a function of M_{DM} (or V_{DM}). In this case, it is seen evidence of a weak decrease of $M_{\mathrm{H\,I}}$ with M_{DM} for (sub)haloes larger than $M_{\mathrm{DM}} \sim 10^{12}$ $\mathrm{M}_{\odot}$, suggesting that the galaxies in these (sub)haloes have exhausted most of their gas reservoir or did not accreted it a long time ago or even ejected their gas, e.g. by AGN feedback.

On the other hand, we presented predictions on the H I spatial clustering for the whole population of local galaxies. Currently, blind H I surveys, like ALFALFA, are shallow and introduce strong selection effects. We have shown that while these selection effects are not crucial for measuring the H I mass function (see Fig. 6.4), they become critical for the H I spatial clustering (as well as for the H I-to-stellar mass correlation, see Chapters 2 and 4). In this sense, we showed that using the ALFALFA H I spatial clustering for constraining H I galaxy-halo connection may lead to incorrect results. The results on the H I spatial clustering presented here can be use for testing theoretical predictions as well as for estimates in the designing of large and deep H I surveys that will be completed with forthcoming radio telescopes such as SKA or the Pathfinder instruments.

Finally, the fact that our mock catalog predicts (slightly) higher H I clustering than ALFALFA even after imposing the selection effects of ALFALFA, suggests that the scatter around the M_{*}–V_{DM} relation should depend on morphology in such a way that the clustering properties of LTGs and ETGs resulted different, as observations show. In a forthcoming work we will explore the effects of introducing the above mentioned dependence of the scatter on morphology. In any case, we highlight that the results presented here for the *whole* galaxy population are valid whether there is a dependence of the scatter around the M_{*}–V_{DM} relation on morphology or not.

Chapter 7

Concluding remarks and outlook

In this thesis we have developed a fully self-consistent semi-empirical framework for multiple baryonic tracers allowing the description of the demographics and the spatial distribution of galaxies, as well as providing a realistic characterization of the galaxy-halo connection. In order to achieve this goal, the research conducted for this thesis was broadly divided into two main parts.

The first part of this thesis consisted in determining the observed HI- and H_2-to-stellar mass ratio distributions of early- and late-type galaxies (ETGs and LTGs) from a large compilation of optically-selected samples with radio observations, Chapter 2. Deriving these distributions directly from our compiled data was impossible since it consisted in a set of many incomplete and inhomogenous samples with a non negligible fraction of radio non detection. Firstly, we homogenized to a common IMF, cosmology, CO-to-luminosity conversion factor and account for selection biases such as environment. Secondly, when radio detections were not possible, many of our compiled samples reported them as upper limits. Here, we used upper limits to derive the ratio distributions but after handling them properly. We first noted that radio non detections depend not only on the sensitivity of the radio telescope and integration time but also on the distance to the object. Based on a realistic mock galaxy survey, all our radio non detections were corrected to have the same distances. The observed HI- and H_2 ratio distributions of ETGs and LTGs were finally derived by using the non-parametric Kaplan-Meier estimator for censored data to reconstruct the information lost by radio non detections. The above allows us to conclude that the HI- and H_2 ratio distributions of ETGs and LTGs are well behaved as a function of stellar mass but they are quite broad and asymmetric. Thus, when studying the ratio distribution it *does* matter on whether ones reports the median, arithmetic mean or geometric mean as a function of stellar mass.

In Chapters 3-4, we developed a fully self-consistent and empirical approach that unifies the local gas distributions and the galaxy mass functions (MFs) traced by their different baryonic components. Roughly, this approach consisted of projecting the previously constrained HI- and H_2 gas distributions into their corresponding mass functions (MFs) by using the galaxy stellar mass function (GSMF) as a pivotal

function. The GSMFs were determined for five orders of magnitude in stellar mass, $M_* \sim 3 \times 10^7 - 3 \times 10^{12}$ M$_\odot$, by combining two spectroscopic samples from the SDSS within the redshift range $0.0033 < z < 0.2$. We developed a simple model to correct our GSMF from surface brightness incompleteness and effects from large scale structures that mainly affects low-mass galaxies. This approach, allows us to derive an inventory of galaxy MFs for HI, H$_2$, cold gas, and baryonic mass, separately into ETGs and LTGs, as well as their corresponding cosmic densities. We also characterized systematic errors and deconvolved from random errors to our inventory of galaxy MFs. Thus, these results represents a fully self-consistent empirical description of the galaxy demographics that is ideal to compare or to constrain galaxy formation models.

In Chapter 5, we analyzed the multi-wavelength `xGASS` survey (Catinella et al., 2018), following the same procedure as in Chapter 2 to correct the upper limits by distance bias and adequately treat them by means of survival analysis to model the differences in the H I gas content between all, late and early galaxies, as well as into centrals and satellites. We compute $R_{\rm HI}$–M_* relations and the full $R_{\rm HI}$ conditional distributions as function of M_*, for these populations. We determined a set of functions to project our empirical H I conditional cumulative distributions as a function M_* of both LTGs and ETGs into the components, in each case, corresponding to central and satellite galaxies, by performing continuous fitting to the processed `xGASS` data. By combining these $R_{\rm HI}$ conditional distributions with the corresponding GSMFs, we calculate the bivariate distribution functions of M_* and $R_{\rm HI}$, for late-type, early-type, and all galaxies, separated into centrals and satellites.

In the second part of this thesis work, Chapter 6, we developed a more complete framework to derive the galaxy-halo connection by assigning stellar masses and the H I gas content of galaxies to dark matter (sub)haloes in the large N-body cosmological simulation SmallMultiDark-Planck (Klypin et al., 2016). In brief, we use (Sub)Halo Abundance Matching, SHAM, to connect galaxy stellar masses to the maximum circular velocity of halos, $V_{\rm DM}$. Then, we use the observed fractions of ETGs for centrals and satellites from a large group galaxy catalog from the SDSS DR7 (Yang et al., 2012) to assign either an early- or late-type morphology to every galaxy in the simulation. Based on these fractions and the $M_{\rm HI}$ conditional PDFs given M_* of ETGs and LTGs for centrals and satellites derived in Chapters 4 and 6, we assign H I masses to every galaxy in the simulation. The above mock galaxy catalogue allows us to obtain no only the galaxy HI-stellar-halo connection but also to compute the spatial clustering of galaxies as a function of stellar mass and H I mass. In particular, in this part of the thesis we were interested in studying the H I-halo connection and the clustering

of the H I mass. Our mock shows that on average, the H I mass is not determined by halo mass (M_h) or V_{DM}, which would imply that galaxy-halo connection models cannot be used to infer the H I-halo connection but other models, like the one proposed in this thesis, should be considered. Also, in this part we provide predictions on the projected two-point correlation function (2PCF) for the whole galaxy population that are expected to be observed by future large and deed radio survey.

Next, we highlight the main conclusions reached all along of this thesis:

- We derived the observed HI- and H_2-to-stellar mass ratio distributions for a large and homogenized compilation of local ETGs and LTGs over a broad stellar mass range. Both for HI and H_2, the distributions of LTGs are well described by Schechter function while for ETGs the distributions are better characterized by a broken Schechter function plus a uniform distribution.

- Our observed ratio distributions confirm the well known results that, on average, at a fixed stellar mass, LTGs have larger H I and H_2 masses than ETGs. We find that these differences increase respectively at the high- and low-stellar mass ends.

- For both ETGs and LTGs, we find that the mean HI- and H_2-to-stellar mass ratio relations are well described by power laws. However, we note some hints of a flattening for galaxies with masses lower than than $\log(M_*/\mathrm{M}_\odot) \sim 9$. This flattening may suggest that the way dwarf galaxies acquired, converted and retained their gas in the ISM, and subsequently transformed it into stars, is different to low galaxies.

- On average, LTGs have a larger fraction of H I than H_2 gas mass. However, in the case of ETGs, on average, both fractions are roughly equal. In addition, we find that the mean depletion time for molecular gas in LTGs is ~ 8 times larger than the cold gas depletion time, this may suggest that LTGs galaxies are inefficient in converting HI gas mass into H2 and in transforming H2 into stars.

- We computed the total GSMF at $z \sim 0$ and find that the slope of the low-mass end is $\alpha \sim -1.4$, consistent with recent determinations based on deeper surveys. The slope of the high-mass end is shallower than previous determination as the result of the photometric catalogue employed in this thesis.

- We find that the total GSMF is well fitted by a function composed of a sub-exponential Schechter function and a double power law function. Based on our

Bayesian fitting approach, we discard the commonly model of a double Schechter function.

- We find that our H I MFs for all, ETGs and LTGs are in good agreement with previous determinations from blind surveys. Similarly the H_2 MFs are consistent with previous determinations based on CO follow-up optically-selected samples. We find that the H I, H_2, and cold gas MFs are mostly dominated by late-type galaxies.

- We deconvolved our MFs from random errors to obtain the intrinsic ones.

- Our determinations for the gas MFs are robust against systematic errors in mass-to-light ratios.

- The $z \sim 0$ cosmic densities of H I, H_2, cold gas, stars and baryons were determined from our MFs. We find that the total density of baryons within galaxies (the ionised and hot gas were not included) is $\sim 5.4\%$ of the universal baryon fraction.

- We develop a full statistical description of the H I gas content of local galaxies as a function of their stellar mass, segregating into late and early types, as well as into central and satellites.

- We calculated bivariate distribution functions of M_* and R_{HI} for late-type, early-type, and all galaxies, separated into centrals and satellites.

- Our bivariate distribution functions determinations suggest that the H I gas content of galaxies depends more on whether they are of late or early type morphology, than conditions associated to whether a galaxy is central or satellite

- The value of $\langle \log M_{\mathrm{HI}} \rangle$ does not increase monotonically as a function of $V_{\max}$ or M_h, unlike $\langle \log M_* \rangle$ which always increases monotonically as a function of $V_{\max}$ (M_h).

- The H I conditional PDFs as a function of $V_{\max}$ or M_{DM} are too broad and highly asymmetric with long tails toward lower values of M_{HI}.

- In general, there are not significant differences between centrals and satellites for the $\langle \log M_{\mathrm{HI}} \rangle$ as a function of $V_{\max}$ or M_{DM}. Nonetheless, on average, centrals have larger fractions of gas than satellites.

- Our predicted H I projected 2PCFs have higher amplitudes than the ones measured from the H I ALFALFA survey. When selecting only gas rich galaxies in

our mock catalogue, that is, emulating the selection effects in the ALFALFA survey, the resulting projected 2PCFs are similar to ALFALFA yet slightly above.

- The total galactic H I gas content in distinct haloes strongly depends on M_h. Below $\sim 10^{12} M_\odot$ this relation is dominated by central galaxies but above is dominated the total contribution of satellite within the halos.

The results presented in this thesis offer a fully self-consistent semi-empirical framework for the demographics of galaxies traced by H I and stellar mass and a realistic characterization of the galaxy-halo connection. In addition we want to highlight that our results could be used extensively in several ways and below we discuss two broad and straight forward applications for the future:

1. Galaxy formation theory: Our results offers an ideal set of results that can be used not only for comparison but for constraining the new generation of semi-analytic models and hydrodynamics simulations of galaxy evolution within the ΛCDM cosmology.

2. Observations: Current blind H I surveys are relatively shallow with strong selection effects. Our framework offers an ideal tool to understand such selection effects. Moreover, they can be used for designing large and deep H I surveys to be completed with forthcoming radio telescopes as SKA or the Pathfinder instruments.

Finally, as for the galaxy-halo connection, several modifications could add more realism to the framework considered here.

- As discussed in Chapter 6, instead of assuming that galaxy morphology is independent of the scatter in the M_*–$V_{\max}$ relation, use a model by correlating galaxy morphology with this scatter. In the future, we will combine the information from the clustering properties of optical and H I surveys in order to constrain this relationship. Moreover, the above will allows us to study the effects of assembly bias.

    In addition to all the information already available we will:

- Add luminosities galaxy colours, star formation rates. Our framework divides galaxies into two main broad galaxy distributions, however, galaxies have different photometrical properties and can also be divided into blue and red galaxies or star-forming and quenched galaxies. While morpholgies and colors/star formation correlates the mapping beteween these properties it is not one-to-one.

223

Thus adding galaxy colors will help to add more realism to our mock catalogue and ease the comparison to theoretical models. Exploiting of the information from multi-wavelength surveys will be key for this part.

- Add galaxy sizes. For Star-forming galaxies, galaxy sizes, star formation rates and their total content of cold gas correlate through the well known relations between star formation and cold gas surface densities. We will use this information in order to constrain the galaxy-halo size relationship.

- Extrapolate to higher redshifts. In this Thesis we provide results based on local measurements but it would be ideal to extend our semi-empirical framework to higher redshifts.

Appendix A

The compiled galaxy samples with HI information

A.1 Golden category

Updated Nearby Galaxy Catalog (UNGC; Karachentsev et al., 2013, 2014): It is the most representative and homogeneous sample of galaxies (869, most of them of low masses) in the Local Volume, located within 11 Mpc or with corrected radial velocities $V_{LG} < 600$ Km s^{-1}. The authors mention that the sample is complete to $M_B \sim -11$ mag, spanning all morphologies. However, we take a more conservative limit, having in mind that at low luminosities the fraction of hardly-to-detect low surface brightness (LSB) galaxies strongly increases. Karachentsev et al. (2013) report the mean $B-$band surface brightness (SB) within the Holmberg isophote, $\bar{\mu}_{B,26}$ for the UNGC galaxies. The SB decreases on average as lower is the luminosity. For LTGs, the distribution of SBs appears to be incomplete from $M_B \approx -13.5$ mag, in such a way that most of the galaxies could be lost at lower luminosities. This is in agreement with the completeness limit suggested by Klypin et al. (2015) for UNGC, based on the turnover that suffers the luminosity function constructed by them at this luminosity. In view of these arguments, we consider complete the UNGC sample for LTGs, only from $M_B \approx -13.5$ mag ($M_* \approx 10^{7.2-7.4}$ M$_\odot$); the few LTGs below this limit are of high SB and are expected then to contain less gas than the average. Since ETGs are of higher SBs than LTGs, the SB distribution for the small fraction of them seems not to be affected even at the lowest observed luminosities, $M_B \sim -11$ mag. There are 561 galaxies with available HI data (for details regarding the data sources on HI fluxes, see Table 3 from Karachentsev et al., 2013); 90 of them do not obey our completeness limit. We estimate stellar masses from the reported K-band luminosities and $B - K$ colors as in Avila-Reese et al. (2008), who calculated the mass-to-light ratios for HSB and LSB galaxies following Bell et al. (2003) and Verheijen (1997),

respectively. The obtained masses (assuming a diet Salpeter IMF) were corrected to the Chabrier IMF. To separate HSB and LSB galaxies we use the reported $\bar{\mu}_{B,26}$, and transform it to a central surface brightness, $\mu_{0,B}$ assuming an exponential disk. Thus, the criterion $\mu_{B,0} > 22.5$ mag/arcsec^2 for selecting LSB galaxies corresponds to $\bar{\mu}_{B,26} > 24.6$ mag/arcsec^2. Karachentsev et al. (2013) apply corrections for peculiar motions in the determination of the distances of all the galaxies.

GALEX Arecibo SDSS Survey (GASS; Catinella et al., 2013): It is an optically-selected sub-sample of 760 galaxies more massive than 10^{10} M$_\odot$ taken from a parent SDSS DR6 sample volume limited in the redshift range $0.025 < z < 0.05$ and cross-matched with the ALFALFA and GALEX surveys. The HI information comes from follow-up observations carried out with the Arecibo 305 m telescope and detections taken from the ALFALFA survey or the Cornell HI digital archive. The R_{HI} limit of the sample is well controlled: 0.015 for $\log(M_*/\mathrm{M}_\odot) > 10.5$ and up to 0.05 for smaller masses. There are 473 detections and 287 non detections; for the latter, upper limits are provided. For the morphological type, we use the Huertas-Company et al. (2011) automatic classification applied to the SDSS DR7. These authors, first of all, provide for each galaxy the probability of being early type, PE, i.e., E or S0. We have tested this probability in a catalog of galaxies with careful visual morphological classification (UNAM-KIAS, see below; Hernández-Toledo et al., 2010) and found that galaxies of types $T \leq 1$ are mostly those with $PE > 0.65$, and those with $PE \leq 0.65$ correspond mostly to $T > 1$.[1] Thus, we consider here as ETGs those with $PE > 0.65$, and the complement are LTGs. We find a good correlation between the ETGs and LTGs this way defined with those defined using the concentration parameter $c = R_{90}/R_{50}$ to characterize the galaxy type, with the value of $c = 2.85$ for separating the LTG population from the ETG one (for the latter, it is asked additionally to obey the color criterion $NUV - r > 5$, Deng, 2013). The stellar masses in Catinella et al. (2013) were calculated from the spectral energy distribution (SED) of the SDSS galaxies (Salim et al., 2007) and assuming a Chabrier (2003) IMF.

Herschel Reference Survey – field galaxies (HRS; Boselli et al., 2010, 2014a,c,b): It is a $K-$band volume limited ($15 \leq D/\mathrm{Mpc} \leq 25$) sample of 323 galaxies complete to $K_s = -12$ and -8.7 mag for LTGs and ETGs, respectively. The authors collected and homogenized from the literature HI data for 315 galaxies, and CO data for most of them. The morphological type was taken from NED or, if not available, from their own classification. Stellar masses are derived from i-band luminosities and $g - i$ colors (from Cortese et al., 2012) by using stellar mass-to-light

---

[1]Huertas-Company et al. (2011) define as ETGs those with $T \leq 1$, but the T index in their case is from the Fukugita et al. (2007) notation, which assigns $T = 1$ to lenticulars instead of $T = 0$ as in the usual de Vacouleurs notation.

ratios as given in Zibetti et al. (2009), and assuming a Chabrier IMF. The distances were corrected for the peculiar motions and presence of clusters. The sample includes objects in environments of different density, from the core of the Virgo cluster, to loose groups and fairly isolated systems. To match the Golden category, we exclude the numerous galaxies from the Virgo Cluster center (regions A and B), which bias the sample to high densities.

ATLAS^{3D} HI sample – field ETGs (Serra et al., 2012): ATLAS^{3D} is a sample of 166 local ETGs observed in detail with integral field unities (IFUs; Cappellari et al., 2011). The distance range of the sample is in between 10 and 47 Mpc; the sample includes 39 galaxies (24% out of the galaxies) from the Virgo Cluster. For the Golden category, we exclude those ETGs in the Virgo core. The sample is not complete, but after excluding the large number of Virgo core galaxies, it is expected to be representative of the local population of ETGs since the galaxies were selected from a complete volume-limited parent sample. The masses range from $\approx 10^{9.8}$ to $10^{11.3}$ M$_\odot$; more massive galaxies are not found typically in small volumes. We estimate stellar masses using the $\log(M_*) = \log(0.5) + \log(L_K)$, where L_K is the K-band luminosity inferred from the K-band absolute magnitude. The HI observations were carried out in the Westerbork Synthesis Radio Telescope (Serra et al., 2012). They use ALFALFA spectra to determine M_{HI} upper limits using one resolution element and find that M_{HI} limit is a factor ~ 2 above the HI mass limit obtained with their data. The R_{HI} limit detection increases with mass on average by more than 1.5 orders of magnitude, attaining values a slow as $\sim 10^{-4}$ for the most massive systems. Because of the ATLAS^{3D} galaxies are nearby, the upper limits are much lower than in the case of the GASS galaxies in the same mass range.

A.2 Silver category

Nearby Field Galaxy Survey (NFGS; Jansen et al., 2000b,a; Wei et al., 2010a; Kannappan et al., 2013, see more references therein): It is a broadly representative sample of 198 local galaxies spanning stellar masses $M_* \sim 10^8 - 10^{12}$ M$_\odot$ and all the morphological types. Morphological classification was obtained from Jansen et al. (2000a). The sample is not complete in volume; galaxies span distances from 2 to 306 Mpc. Distances were derived from the Virgo centric flow corrected velocities with respect to the centroid of the Local Group. Stellar masses were estimated using a variant of the code described in Kannappan & Gawiser (2007) and improved in Kannappan et al. (2009), which fits the SED and integrated spectrum of a galaxy with a suite of stellar populations models. Both the diet Salpeter and the Chabrier

(2003) IMFs were used. The single-dish HI fluxes for most of the galaxies were taken from the HyperLeda database (Paturel et al., 2003) or were obtained by the authors with the Green Bank Telescope (GBT) Spectrometer. The sample provides strong upper limits up to $R_{HI} \sim 0.1$; all galaxies with larger ratios are detected (139, and the rest have only upper limits).

Stark et al. (2013) compilation: These authors compiled and homogenized from the literature 323 galaxies with available HI, CO, and multi-band imaging data. Most of the compiled galaxies are from the GASS, NFGS and ATLAS^{3D} surveys described above. We use here only those galaxies that are not in these surveys (67 galaxies). The authors use morphological type to separate galaxies into two groups, coincident with our morphology criterion for ETGs and LTGs. In their compilation are included some blue compact dwarfs (BCDs). We exclude those BCDs classified as early types. The stellar masses were calculated following Kannappan et al. (2013). The optical and NIR information required for this calculation were taken from SDSS DR8 (for those galaxies outside the SDSS footprint, the $BVRI$ photometry from the SINGS sample is used) and 2MASS, respectively.

Leroy et al. (2008) THINGS sample: It is a sample of selected 23 nearby, star-forming galaxies, which we associate with LTGs; 11 are dwarf, H$_I$-dominated galaxies and 12 are large well-defined spiral galaxies. The HI information of the galaxies comes from "The HI Nearby Galaxy Survey" (THINGS, Walter et al., 2008) and it was obtained with the NRAO Very Large Array (VLA). The stellar masses are calculated from 3.6 μm information taken from the Spitzer Infrared Nearby Galaxies Survey (SINGS Kennicutt et al., 2003). To convert the 3.6 μm intensity to surface stellar mass density, they use a K-to-3.6 μm calibration and adopt a fixed $K-$band mass-to-light ratio, $\Upsilon_*^K = 0.5 M_\odot / L_\odot$, assuming a Kroupa (2001) IMF; M_* is calculated from integrating the surface stellar mass density.

Dwarf LTGs (Geha et al., 2006): It is a sample of 101 dwarf galaxies, 88 out of them with HI measurements and being of late types. Galaxies with absolute magnitudes $M_r - 5\log_{10}(h_{70}) > -16$ were selected from the low-luminosity spectroscopy catalog of Blanton et al. (2005c), based on the SDDS. Distances are estimated based on a model of the local velocity field (Willick et al., 1997). Possible selection effects related to the Blanton et al. (2005c) catalog are that it does not span the full range of environments (there are not clusters), and LSB dwarfs are missed. Stellar masses are based on the optical SDSS i-band magnitude and $g - r$ colors using the mass-to-light ratios of Bell et al. (2003). The M_{HI} masses were obtained by Geha et al. (2006) from the H$_I$ integrated fluxes measured with the Arecibo 305 m telescope and the GBT.

ALFALFA dwarf sample (Huang et al., 2012a): It consists of 176 low HI mass

dwarf galaxies from the ALFALFA survey. The galaxies were selected to have $M_{HI} < 10^{7.7}$ M$_\odot$ and HI line widths < 80 km s^{-1} (*s-com* sample). This sample is not complete in a volume-limited sense but it probes the extreme low HI mass tail of the ALFALFA survey. Stellar masses are obtained through SED fitting following Salim et al. (2007), assuming a Chabrier (2003) IMF. Only 57 out of the 176 galaxies have stellar mass determination. These galaxies have HI detections and high gas fractions, they are dwarf irregulars.

A.3   Bronze category

UNAM-KIAS catalog of isolated galaxies (Hernández-Toledo et al., 2010): It is a magnitude-limited sample ($m_r > 15.2$ mag) of galaxies from the SDSS DR5 that obey strict isolation criteria; it is composed of 1520 galaxies spanning all morphological types. The morphological classification was carried out by the authors. We have searched HI information for these galaxies in HyperLeda (the 21-cm line magnitudes corrected for self-absorption, m_{21}^c). The HI masses are calculated as $M_{H_I}[M_\odot] = 2.356 \times 10^5 \cdot d_L^2 \cdot F_{21}$, where $F_{21}[\text{Jy Kms}^{-1}] = 10^{0.4(17.40-m_{21}^c)}$ and d_L is the luminosity distance to the galaxy in Mpc. For the HI non-detections, we have searched rms noise limits in the Digital archive of HI 21 centimeter line spectra of optically selected galaxies (Springob et al., 2005), finding data only for 7 galaxies. Non-detected HI upper mass limits are estimated as $M_{H_I}^{\lim}[M_\odot] = 1.5 \cdot \text{rms} \cdot \delta W$, where δW is the full width of the HI line obtained from the Tully-Fisher relation of Avila-Reese et al. (2008) ($\delta W = 2V_m$ is assumed). For LTGs (ETGs), we find 272 (24) detections and 7 (0) non-detections. Stellar masses are taken from the group catalog of Yang et al. (2007), where the Bell et al. (2003) mass-to-light ratios for a Kroupa (2001) IMF were used.

Analysis of the interstellar Medium of Isolated GAlaxies (AMIGA; Lisenfeld et al., 2011): It is a redshift-limited sample ($1500 \leq v_{rec}$ [km s^{-1}] ≤ 5000) consisting of 273 isolated galaxies with reported multi-band imaging and CO data. We perform the same procedure described above for the UNAM-KIAS sample to estimate detected and non-detected HI masses. For LTGs (ETGs) galaxies, we find 203 (11) detections. Only 4 non-detections were found, all for ETGs. The stellar masses were calculated as described above for the UNGC sample. Morphologies were obtained using higher resolution images from SDSS or their own images.

Low-mass Isolated galaxies (Bradford et al., 2015): It is a sample of 148 isolated low-mass galaxies ($7 \leq \log(M_*/\text{M}_\odot) \leq 9.5$) drawn from the SDSS NSA catalog (see Geha et al., 2012). Isolated galaxies are defined as those without massive hosts

(at least 0.5 dex more massive than the given galaxy) at projected distances less than 1.5 Mpc. HI measurements were obtained using the 305 m Arecibo and the 100 m Greenbank telescopes. Stellar masses are calculated in the NSA catalog using the kcorrect software of Blanton & Roweis (2007b) using the SDSS and GALEX photometric bands and assuming a Chabrier 2003 IMF. For the morphology, we use the Huertas-Company et al. (2011) automatic classification, following the same procedure described above for the GASS survey, finding classification for 128 out of the 148 galaxies; all of them are of late type. Indeed, according to Geha et al. (2012) all the isolated low-mass galaxies in the local Universe are star forming (late-type) objects.

Herschel Reference Survey – Virgo galaxies: This is the same HRS sample described above but taking into account only galaxies from the Virgo Cluster central regions A and B (59). Therefore, this sample is biased to contain galaxies in a very high density environment.

ATLAS^{3D} HI sample – Virgo core ETGs: This is the same ATLAS^{3D} sample described above but taking into account account only the Virgo core ETGs (15). Therefore, this sample is biased to contain ETGs in a very high density environment.

Appendix B

The compiled galaxy samples with CO (H$_2$) information

B.1 Golden category

Herschel Reference Survey (HRS)– field galaxies: It is the same sample described in §§A.1 (excluding Virgo Cluster core), with 155 galaxies with available CO information (101 detections and 54 non detections). The authors either used compiled CO observations from the literature or they carried out their own observations with the National Radio Astronomy Observatory (NRAO) Kitt Peak 12 m telescope (Boselli et al., 2014a). A MW constant or H-band luminosity-dependent (Boselli et al., 2002) CO-to-H$_2$ conversion factor is applied to calculate M_{H_2}.

CO Legacy Legacy Database for GASS (COLD GASS; Saintonge et al., 2011): This is a program aimed at observing CO(1-0) line fluxes at the IRAM 30 m telescope for galaxies from the GASS survey described in §§A.1. From the CO fluxes, the total CO luminosities, and hence the H$_2$ masses, were calculated for 349 galaxies. The authors apply the MW constant CO-to-H$_2$ conversion factor.

ATLAS^{3D} H$_2$ sample – field ETGs (Young et al., 2011): This is the same sample described in §§A.1 (excluding Virgo Cluster core) but with observations in CO using the IRAM 30 m Radio Telescope. The sample amounts for 243 ETGs with CO observations. The authors use the constant MW CO-to-H$_2$ conversion factor.

B.2 Silver category

Stark et al. (2013) compilation: It corresponds to the same compiled galaxy sample described in §§A.2. The authors observed 35 galaxies of the NFGS with the IRAM 30 m and the ARO 12 m telescopes to measure the CO $(J \rightarrow 2 - 1)$ (IRAM) and $(J \rightarrow 1 - 0)$ (IRAM & ARO) lines. For the other galaxies, the H$_2$ information

from previous works was used. Stark et al. (2013) use the MW constant CO-to-H_2 factor for estimating M_{H_2}.

Leroy et al. (2008) HERACLES sample: It is the same sample described in §§A.2. The H_2 information for the 23 galaxies (LTGs) comes from the CO $J \to 2-1$ maps from the HERA CO-Line Extragalactic Survey (HERACLES Leroy et al., 2008, CO $J \to 2-1$ is related to CO $J \to 1-0$ by assuming the ratio $I_{\text{CO}}(2 \to 1)/I_{\text{CO}}(1 \to 0) = 0.8$), and CO $J \to 1-0$ maps from the Berkeley-Illinois-Maryland Association (BIMA) Survey of Nearby Galaxies (BIMA SONG Helfer et al., 2003). The MW constant CO-to-H_2 conversion factor was used.

APEX Low-redshift Legacy Survey for MOlecular Gas: (ALLSMOG; Bothwell et al., 2014) Using the APEX telescope, the $CO(2 \to 1)$ emission line was measured to trace H_2 in 42 late-type galaxies of masses 8.5 $<\log(M_*/\text{M}_\odot)<$ 10, in the redshift range $0.01 < z < 0.03$ and with metallicities $12 + \log(O/H) > 8.5$. Morphological classification was taken from NED. The stellar masses are derived based on SED fitting (Kauffmann et al., 2003) using the SDSS DR7 optical data. To obtain the $CO(1 \to 0)$ line luminosities, the $CO(2 \to 1)$ emission line is assumed to be fully thermalized. A MW constant or metallicity-dependent (Wolfire et al., 2010) CO-to-H_2 conversion factor were applied to infer the H_2 masses.

Bauermeister et al. (2013) compilation: We take from this literature compilation 8 galaxies in the low-redshift range $0.05 \le z \le 0.1$. All of them are star forming and we associate them to LTGs. Their stellar masses are in the range $4 \times 10^{10} M_\odot \le M_* \le 1.6 \times 10^{11} M_\odot$ and they were calculated by fitting SDSS *ugriz* photometry to a grid of models spanning a wide range of star formation histories. The H_2 masses are obtained by the authors from CO $J \to 1-0$ intensity maps with CARMA, using a MW constant CO-to-H_2 conversion factor.

B.3   Bronze category

Analysis of the interstellar Medium of Isolated GAlaxies (AMIGA; Lisenfeld et al., 2011): This is the same sample described in §§A.3. The authors carried out their own observations of CO($J \to 1-0$) with the IRAM 30 m or the 14 m FCRAO telescopes for 189 galaxies and 87 more were compiled from the literature. An aperture correction is applied to the CO data. A MW constant CO-to-H_2 conversion factor is used to compute M_{H_2}.

Herschel Reference Survey – Virgo core: This is the same HRS sample described above but taking into account only the Virgo Cluster core regions A and B galaxies (62). Therefore, this sample is biased to contain galaxies in a very high

density environment.

ATLAS^{3D} **H$_2$ sample – Virgo core ETGs**: This is the same ATLAS^{3D} sample described above but taking into account account only the Virgo core ETGs (21). Therefore, this sample is biased to contain ETGs in a very high density environment.

Appendix C

The impact of galaxy classification: the criteria for separating the galaxy population into two groups

For our goal of projecting gas scaling correlations (more precisely, the gas CPDFs) into gas MFs separately for early- and late-type galaxies, the derivation of the fraction of early-type galaxies as a function of M_*, $f_E(M_*)$, was an important step. As discussed in Section 3.6, based on the morphological classification from Huertas-Company et al. (2011) we found the GSMFs of early- and late-type galaxies that are in good agreement with the results based on the visual classification from Nair & Abraham (2010a), and with the classification based on concentration utilised in Bernardi et al. (2010). In contrast, we found that our GSMFs of early- and late-type galaxies are in tension when comparing to those from the GAMA survey with their visual morphological classification, but interestingly enough, they agree with the GAMA GSMFs when we use a $g - r$ color criterion to separate our galaxies into the two populations. Recall that for the GAMA classification, Sa galaxies are included into their early-type group since their visual classifications combines S0 and Sa galaxies (Kelvin et al., 2014; Moffett et al., 2016a), contrary to our definition; see Section 3.6 for more details. Thus, understanding the impact of using different criteria to separate the galaxy population into two main groups is of great importance in our study. Following, we study the impact of using galaxy color instead of morphology in order to give a rough idea of what would it be the result of using a very different proxy to galaxy morphology (a one close to the GAMA survey, for instance).

The lower panel of Figure 3.9 presented the fractions of early-type galaxies as well as of red galaxies as a function of M_*. The fraction of red galaxies is clearly larger than the one of early-type galaxies at all masses. Based on the SDSS DR7 sample described in Chapter 3, we found that the great majority of the galaxies that are classified as early-type are actually red; the fraction of early-type galaxies with blue colours has a maximum at $M_* \sim 8 \times 10^{10} M_\odot$ representing only $\sim 5\%$ of the population. In contrast,
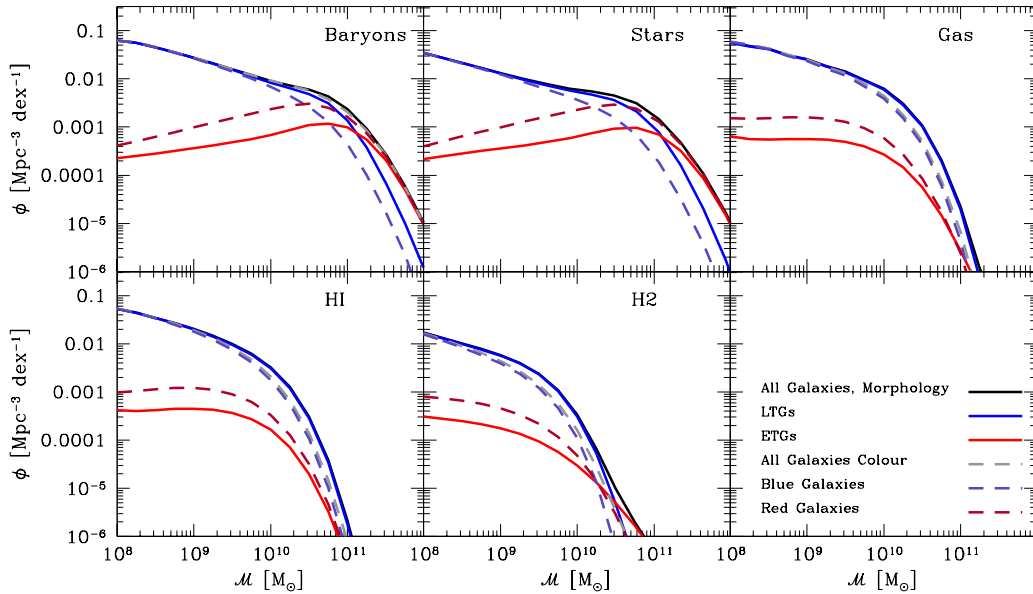
Figure C.1: Impact on the MFs due to the use of two different criteria for the division of the galaxy population. The solid lines show the original MF from Section 4.3.4, based on galaxy morphology, while the dashed lines show the results when using galaxy colour. The classification according to galaxy colours results in a overabundance of red galaxies compared to early-types, especially at intermediate and low masses.

the fraction of red galaxies classified as late-types is larger than $\sim 10\%$ at practically all masses, and it peaks at $M_* \sim 2 \times 10^{10} M_\odot$ with a contribution of $\sim 50\%$ (we also observe a second peak at the massive-end $M_* \sim 4 \times 10^{11} M_\odot$). Similar results have been reported in previous studies (see e.g., Masters et al., 2010b). Additionally, note that we ignored the effects of reddening due to extinction from the galaxy inclination, which would misclassify galaxies based on their colours (see e.g., Masters et al., 2010a). Therefore, from the physical point of view, the separation of the galaxy population by colour is, perhaps, not as "clean" or reliable as morphology.

Figure C.1 presents the resulting MFs when using the fraction of red galaxies, $f_r(M_*)$, as a proxy for early-type galaxies, dashed lines. The solid lines reproduce the results from Fig. 4.6, where our morphology-based fraction, $f_E(M_*)$, was used. Notice that for HI, H_2, and cold gas mass, not only the MFs of blue and red galaxies are different to their morphological counterparts but also the total MFs. The above can be understood in the following terms. Using the fraction of red galaxies as a proxy of early-type galaxies results in a large fraction of red galaxies misclassified as late-types as discussed above. However, the above has a lager impact for early-type galaxies with low to intermediate masses than at high masses, while for late-type

235

galaxies, the major impact is from intermediate to high masses. As a consequence, on one hand, the resulting HI and H_2 MFs see an increase of early-type galaxies at their low-mass ends. Interesting enough, the use of f_r instead of f_E would produce HI and H_2 MF of early-type galaxies in better agreement with the inferences of the ATLAS 3D sample. On the other hand, lowering the fraction of late-type galaxies at intermediate-high masses, which have significantly larger gas fractions than early-type galaxies, affects the projected total HI and H_2 MFs, and they would be in tension with direct observations, especially with the HI MF from the ALFALFA and HIPASS surveys.

Finally, we emphasise that the above does not imply that using galaxy colours will lead to incorrect inferences of the gas MFs but that combining two different criteria for dividing the galaxy population will lead to a very different results that, perhaps, will be in tension with the observations. Thus, the success of our determinations is in part that we are using data sets that are consistent between each other in that regards the morphological separation into two galaxy subpopulations.

Appendix D

Deconvolution Algorithm

Individual mass estimates are subject to random errors. Thus, every MF that is inferred from observations through indirect estimations of any type of mass (we will denote this as ϕ_{obs}) is the result of the random errors over the intrinsic mass, (it will be denoted by ϕ_{int}). Formally, we can represent the observed ϕ_{obs} as the convolution of ϕ_{int}:

$$\phi_{\text{obs}}(M) = \int \mathcal{G}(\log M - \log x)\phi_{\text{int}}(x)d\log x. \qquad (\text{D.0.1})$$

We will assume that random errors have a lognormal distribution, denoted by $\mathcal{G}(\log M - \log x)$:

$$\mathcal{G}(\log M - \log x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\log^2\left(\frac{M}{x}\right)\right], \qquad (\text{D.0.2})$$

where σ are the 1-σ statistical fluctuations, in either directions, in the inferred galaxy masses. Note that in Equation (D.0.1) the units for ϕ_{obs} and ϕ_{int} are in $\text{Mpc}^{-3}\text{dex}^{-1}$.

The basic idea of our algorithm is simple. We start by defining the following relation:

$$\phi_{\text{int}}^j(M) = \phi_{\text{int}}^{j-1}(M) \int \mathcal{G}(\log M - \log x)\frac{\phi_{\text{obs}}}{\phi_{\text{conv}}^{j-1}}(x)d\log x, \qquad (\text{D.0.3})$$

where

$$\phi_{\text{conv}}^{j-1}(x) = \int \mathcal{G}(\log x - \log y)\phi_{\text{int}}^{j-1}(y)d\log y, \qquad (\text{D.0.4})$$

with ϕ_{int}^{j-1} denoting the jth iterated intrinsic MF. Note that as ϕ_{conv}^{j-1} approaches to ϕ_{obs} the above equation converges to the maximum likelihood solution for ϕ_{int}^{j-1}, in other words, we have found the numerical solution to the intrinsic MF, ϕ_{int}. The zero-th iteration is defined as convolution of the observed MF with the lognormal distribution $\mathcal{G}$ described above:

$$\phi_{\text{int}}^0(M_*) = \int \mathcal{G}(\log M_* - \log x)\phi_{\text{obs}}(x)d\log x. \qquad (\text{D.0.5})$$

We declare that the ϕ_{int}^j has converged when the parameter $\Delta \leq 7\%$, defined as the relative error between the observed MF and the j-th iterated intrinsic MF convolved

with the random error distribution:

$$\Delta = \frac{100\%}{N} \sum_i \left| 1 - \frac{\int \mathcal{G}(\log M_* - \log x)\phi^j_{\text{int},i} d\log x}{\phi_{\text{obs},i}} \right|. \qquad \text{(D.0.6)}$$

The summation in the above definition goes over all the tabulated values of individual reports of the observed MF ϕ_{obs}. By trial and error we found that the value of $\Delta = 7\%$ is a compromise between accuracy and efficiency. Typically, $\Delta = 7\%$ was reached in less than 10 iterations.

Appendix E

Correcting `xGASS` to the morphology and environment distributions of SDSS

In this Appendix, we begin by defining the fractions of the different galaxy sub-samples required to perform the joint analytic fitting to the `xGASS` HI conditional CDFs (§§5.3.3). Then, we compare these fractions from `xGASS` to those from SDSS DR7 and their fits. Finally, we explain our procedure for weighting `xGASS` galaxies in order they reproduced the fractions of ETGs and of satellites from SDSS. Following, in the definition of the different fractions, for simplicity, we omit the dependence on M_*:

(i) *Fraction of ETGs/LTGs:*
Defined as the ratio of ETGs and total mass functions, $f^E \equiv \phi^E/\phi$. The fraction of LTGs, fL, is the complement, this is $f_L = 1 - f^E$.

(ii) *Fraction of centrals/satellites*:
Defined as the ratio of centrals and total mass functions, $f^c \equiv \phi^c/\phi$. The fraction of satellites, f^s, is the complement $f^s = 1 - f^c$.

(iii) *Fraction of centrals ETGs/LTGs*:
Defined as $f_E^c \equiv \phi_E^c/\phi^c$, the ratio of ETGs that are centrals and centrals mass functions. f_L^c is the complement of f_E^c, this is $f_L^c = 1 - f_E^c$.

(iv) *Fraction of satellites ETGs/LTGs*:
Defined as $f_E^s \equiv \phi_E^s/\phi^c$, the ratio of ETGs that are satellites and centrals mass functions. f_L^s is the complement of f_E^s, this is $f_L^s = 1 - f_E^s$.

(v) *Fraction of LTGs that are satellites/centrals*:
Defined as $f_j^L \equiv \phi_j^L/\phi^L$, with $j = c, s$, the ratio of LTGs that are satellites/centrals and LTGs mass functions.

(vi) *Fraction of ETGs that are satellites/centrals*:

Defined as $f_j^E \equiv \phi_j^E/\phi^E$, with $j = c, s$, the ratio of ETGs that are satellites/centrals and ETGs mass functions.

Figure E.1 shows most of the fractions defined above for xGASS, salmon squares connected with solid lines. From left to right, the upper panels show the fractions of satellites for all galaxies, and the fraction of ETGs for the samples of central and satellite galaxies. The lower panels show the fractions of ETGs for all galaxies, and the fraction of satellites for the samples of LTGs and ETGs. In these panels, the data from SDSS DR7 are also plotted (circles). We use Meert et al. (2015) photometry and an average stellar mass from five different mass-to-luminosity prescriptions, updated galaxy group catalogs from Yang et al. (2007), and the Huertas-Company et al. (2011) morphological classification (see Chapter 3 for details and for the corrections applied to obtain a volume-limited sample). As seen in panels (b) and (c), the fractions of ETGs in the central and satellite subsamples are systematically larger up to $M_* \sim 10^{11}$ $M_\odot$ for xGASS than for SDSS; at larger masses, the difference inverts for the subsample of satellite galaxies. In the insets of these panels, we plot the ratios of the respective fractions of SDSS and xGASS.

As mentioned in §§5.2.1, to infer from xGASS the R_{HI}–M_* relations and R_{HI} distributions given M_* corresponding to all galaxies, as well as to the subsamples of central and satellite galaxies, the biases of xGASS with respect to SDSS in morphology and environment should be corrected. For this, we adopt a methodology similar as the used in Catinella et al. (2018) for recovering a volume-limited sample. When we compute the above mentioned R_{HI}–M_* relations or the whole R_{HI} distributions given M_*, we apply weights to the xGASS galaxies to recover the SDSS fractions of ETGs for the central and satellite subsamples. Weights are simply obtained as the ratios shown in the insets of panels (b) and (c) of Figure E.1. This automatically recovers also the overall SDSS fraction of ETGs and the overall fraction of satellites. In any case, note that the relevant bias of the xGASS sample with respect to SDSS is by morphology; the bias in selecting central/satellite galaxies is small and mainly due to the former.

For the above procedure and for extrapolating the fits to the R_{HI} conditional CDFs from xGASS to masses lower than $M_* = 10^9$ $M_\odot$, analytical fits to the SDSS fractions are required. The fits are performed to the overall fraction of satellites (panel a) and the fractions of ETGs for the central and satellite subsamples, panels (b) and (c), respectively. We perform MCMC multiparametric fits to the SDSS fractions $f_E^c(M_*)$ and $f_E^s(M_*)$ with a composition of two analytic Sigmoid functions following the procedure described in Rodríguez-Puebla et al. (2013). The fitted analytic function is given by,

$$f_E^j(M_*) = \frac{1 - A}{1 + e^{-\gamma_1(x_{C,1} + x_{0,1})}} + \frac{A}{1 + e^{-\gamma_2(x_{C,2} - x_{0,2})}}, \tag{E.0.1}$$
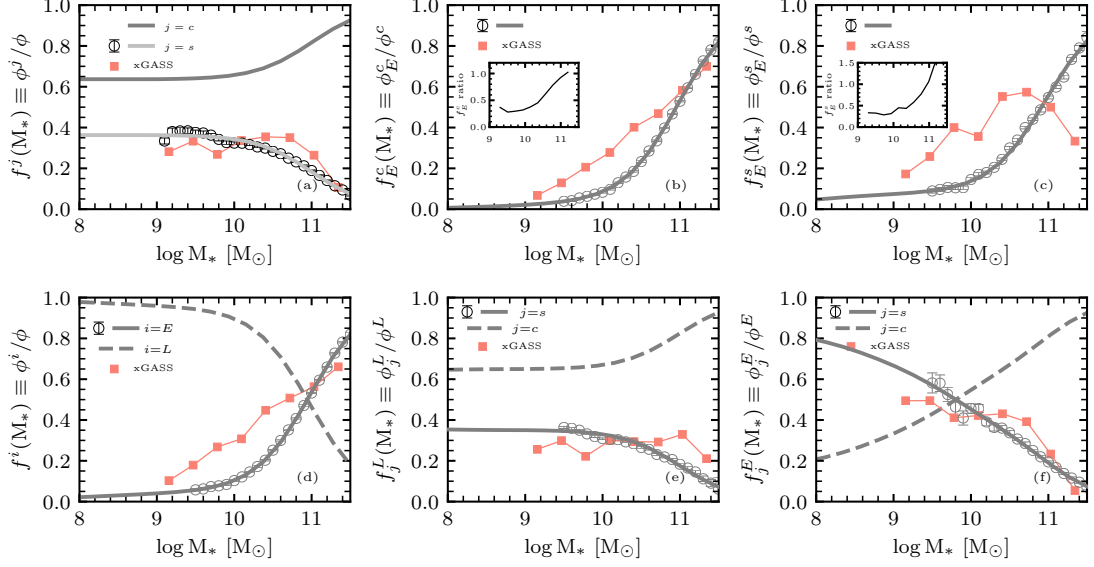
Figure E.1: Different fractions of subsamples of galaxies calculated from the SDSS using the Yang et al. (2012) group catalog for defining centrals and satellites and the Huertas-Company et al. (HC11 2011) morphological classifications. The fractions corresponding to the xGASS sample are shown with filled salmon squares connected by solid lines. The solid lines are fits to the SDSS data; the dashed lines show the respective complementary fractions. The insets in panels (b) and (c) are the ratios of the SDSS to xGASS fractions.

where $j = c$ or s, $x_{C,i} = M_*/\mathcal{M}_{C,i}$, with $i = 1, 2$. For the overall fraction $f^s(M_*)$, we use an analytic function composed of a Sigmoid and constant function given by,

$$f^s(M_*) = 1 - \left[A \cdot \frac{1}{1 + e^{-\gamma(x_C - x_0)}} + H \right], \tag{E.0.2}$$

where H is the constant function. Here, the Sigmoid normalization factor is defined as $A \equiv 1 - H$.

The obtained fits are shown in Figure E.1 with the solid gray lines. The fractions in the lower panels were calculated from the fractions of the upper panels. The dashed gray lines in all the panels are just the respective complementary fractions.

Finally, in §§5.5.2 we explore the effects on our results of using a morphological classification different to the used here. The explored alternative classification was the one by (Domínguez Sánchez et al., 2018). In Figure E.2 we show the same results as in Figure E.1 but using the (Domínguez Sánchez et al., 2018) morphologies for both the xGASS and SDSS galaxies. It is interesting that in this case, the differences in the fraction of ETGs among the xGASS and SDSS samples are less than when using the Huertas-Company et al. (2011) classification.
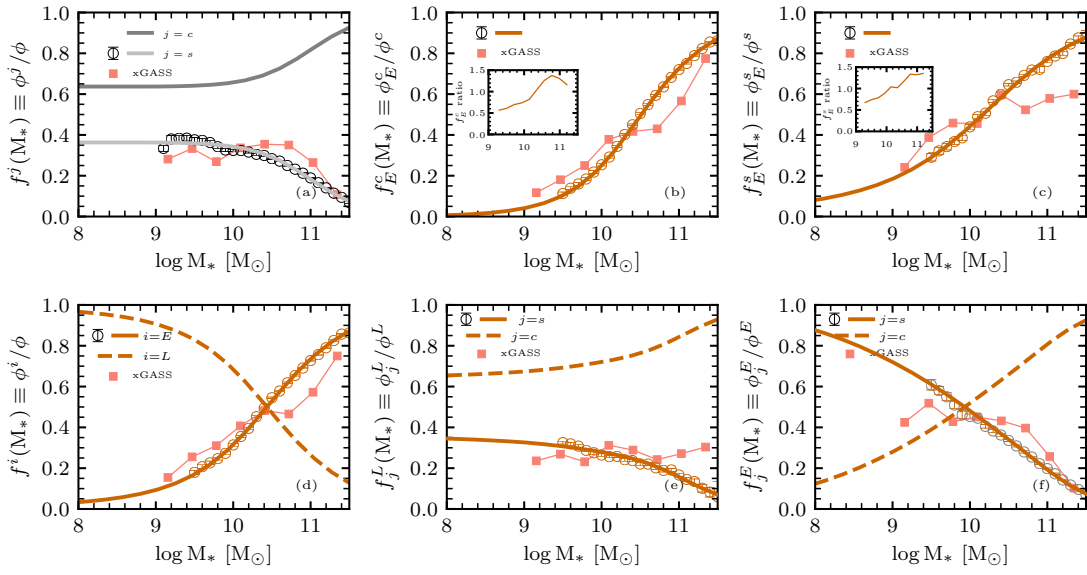
Figure E.2: Same as figure E.1 but using the Domínguez Sánchez et al. (2018) morphological classification.

Appendix F

Procedure for correcting the upper limits of xGASS

F.1 Upper limits of ETGs

By comparing the left and right panels of Figure 5.2, we note that the xGASS detection limits truncate significantly the HI conditional PDFs of ETGs. Having in mind these distributions, we ask ourselves how the non-detected galaxies in xGASS would be seen in the R_{HI}–M_* plane if they were observed with the same instrument, observational setup, and allowed signal-to-noise ratio but at lower distances than those of xGASS galaxies. The thin dash-dotted and dotted lines in Figure 5.2 labeled with 50 Mpc and 25 Mpc show the shift that the GASS and GASS-low detection limits would have at these distances, assuming median distances of 165 Mpc for the former, and 65 Mpc for the latter.

In Chapter 2, based on the ATLAS$^{\mathrm{3D}}$ results, we homogenized the GASS upper limits assuming that galaxies were observed at 25 Mpc. From this assumption, we then estimated the fraction of upper limits in GASS that would have been detected in between the GASS detection limit and this limit shifted to 25 Mpc (as mentioned above, from ATLAS$^{\mathrm{3D}}$ we estimated this fraction to be 25%). For the remaining 75% fraction, we re-estimated their upper limits. As seen, even for such small distance, yet a significant fraction of GASS ETGs would remain as non-detected but their now homogenized upper limits result much lower than the original ones. These upper limits along with those from other ETG samples compiled in Chapter 2, pile up around low values of R_{HI}. The larger the mass, the smaller these values. From the performed continuous fit to the observed R_{HI} distributions in M_* bins, the R_{HI} values where the upper limits pile up were constrained by the function $\mathcal{R}_1(M_*)$, see Eq. (4.2.11) in Chapter 4. The values of $\mathcal{R}_1(M_*)$ correspond roughly to those where the top-hat functions start in the conditional PDFs for ETGs shown in Figure 5.1 or 5.2.

The fraction of galaxies in the top-hat functions correspond to the fractions of non detections.[1] As expected, for $M_* \gtrsim 10^{10}$ M$_\odot$, the values of $\mathcal{R}_1(M_*)$ are close to the upper limits of ATLAS$^{\text{3D}}$. However, have in mind that in Chapter 2 we included other galaxy samples besides `GASS` and ATLAS$^{\text{3D}}$.

Based on the analysis of Chapter 2, we proceed here as follows in order to correct the `xGASS` upper limits of ETGs by the distance bias:

1. From the empirical ETG R_{HI} conditional PDFs reported in Chapter 4, calculate the fraction of galaxies that lie in each stellar mass bin in between the `GASS` and ATLAS$^{\text{3D}}$ R_{HI} detection limits (as done in Chapter 2), and in between the R_{HI} detection limit of the of low-`xGASS` and $\mathcal{R}_1(M_*)$ (recall that in ATLAS$^{\text{3D}}$ there are not low-mass galaxies).

2. Convert to detections a fraction of the `xGASS` upper limits at each M_* bin equal to the respective fraction as calculated in item (i). To assign R_{HI} values to these detections, pick randomly them from the empirical ETG R_{HI} conditional PDFs in the R_{HI} ranges determined in the previous item (in Chapter 2, for `GASS` galaxies, a uniform distribution was assumed).

3. Lower the upper limits by a factor $(D_i(z)/25Mpc)^2$ of the remainder galaxies, where $D_i(z)$ is in Mpc. This is equivalent to say that these galaxies, with similar observational setups and signal-to-noise ratios as used in `xGASS` and `GASS`, will remain undetected in HI at the distance of 25 Mpc, but their upper limits are re-calculated accordingly to this distance.

It is worth of mentioning that for $M_* > 10^{10}$ M$_\odot$, the values of the fractions calculated in item (i) are around $30 - 40\%$, larger than the $\sim 25 - 30\%$ fraction of galaxies detected by ATLAS$^{\text{3D}}$ in between the detection limit of this survey and the one of `GASS` (see Chapter 2).

F.2 Upper limits of LTGs

From Figure 5.2 we see that the `xGASS` detection limits affect only the low-end of our empirical HI conditional PDFs of LTGs. The fraction of LTGs with upper limits that pile up around these limits is relatively small. Note that if these galaxies were closer,

---

[1] To estimate the R_{HI} distributions of ETGs, in Chapter 2 we assumed that the real values (detections) of R_{HI} should be up to ~ 1 dex below the upper limit values after corrections and survival analysis, following a uniform distribution. This is why the R_{HI} conditional PDFs shown in Figure 5.1 have a top-hat distribution of ~ 1 dex width at their low-R_{HI} ends; see Chapter 2 for arguments in favor of this assumption and for a discussion.

then they likely would have been detected in HI, as is the case for galaxies from the closer HRS sample, see Chapter 2. Thus, we convert the upper limit of a given LTG to a *detection* with the $R_{\rm HI}$ value randomly picked out from the tail of the empirical $R_{\rm HI}$ conditional PDF given M_* from Chapter 4.

Appendix G

Conservation equations

As discussed in section 5.3.3, performing fits to xGASS CDFs must obey the law of total probability. Here, we present the "probability conservation equations" in order to satisfy such requirement for the whole set of galaxies, different subsets of LTGs/ETGs, centrals/satellites, and their combinations.

First, to describe the HI conditional CDFs of all LTGs and ETGs, and central LTGs and ETGs (four sets of CDFs) we propose the analytic incomplete gamma function given by Eq.(5.3.3) for each one of these populations.

The remaining five sets of HI CDFs to be used also for the fitting procedure are described by the below listed five equations that obey the law of total probability, and that allow us to use the above mentioned four sets of CDFs for calculating these five sets of CDFs. Such equations require information on different fractions of populations and subpopulations of galaxies as a function of M_*. In Appendix E we discuss how we estimate these fractions. For simplicity, we do not show the dependence of these fractions on M_* in the following equations:

- *HI CDFs of the whole sample:*

$$P^T(< R_{\mathrm{HI}}|M_*) = f^L \cdot P^L(< R_{\mathrm{HI}}|M_*) + f^E \cdot P^E(< R_{\mathrm{HI}}|M_*) \qquad (\mathrm{G.0.1})$$

where f^E and f^L are the fractions of ETGs and LTGs, respectively; $f^E + f^L = 1$.

- *HI CDFs of the subsample of centrals:*

$$P^c(< R_{\mathrm{HI}}|M_*) = f_L^c \cdot P_L^c(< R_{\mathrm{HI}}|M_*) + f_E^c \cdot P_E^c(< R_{\mathrm{HI}}|M_*) \qquad (\mathrm{G.0.2})$$

where f_E^c and f_L^c are the fractions of centrals that are ETGs and LTGs, respectively; $f_E^c + f_L^c = f^c$

- *HI CDFs of the subsample of satellites*

$$P^s(< R_{\mathrm{HI}}|M_*) = \frac{1}{f^s} \left[P^T(< R_{\mathrm{HI}}|M_*) - f^c \cdot P^c(< R_{\mathrm{HI}}|M_*) \right] \qquad (\mathrm{G.0.3})$$

where $P^T(< R_\mathrm{HI}|M_*)$ and $P^c(< R_\mathrm{HI}|M_*)$ are the total and centrals CDFs given by eqs. (G.0.1) and (G.0.2) respectively. f^c and f^s are the fraction of centrals and satellites, $f^c + f^s = 1$.

- *HI CDFs of the subsample of satellites that are LTGs*

$$P_L^s(< R_\mathrm{HI}|M_*) = \frac{1}{f_s^L} \left[P_L^T(< R_\mathrm{HI}|M_*) - f_c^L \cdot P_L^c(< R_\mathrm{HI}|M_*) \right] \qquad (G.0.4)$$

where $P_L^T(< R_\mathrm{HI}|M_*)$ and $P_L^c(< R_\mathrm{HI}|M_*)$ are the total LTGs and LTGs centrals CDFs analytic fits given by eq.(5.3.3) respectively. f_c^L and f_s^L are the fractions of LTGs that are centrals and satellites, $f_c^L + f_s^L = f^L$

- *HI CDFs of the subsample of satellites that are ETGs*

$$P_E^s(< R_\mathrm{HI}|M_*) = \frac{1}{f_s^E} \left[P_E^T(< R_\mathrm{HI}|M_*) - f_c^E \cdot P_E^c(< R_\mathrm{HI}|M_*) \right] \qquad (G.0.5)$$

where $P_E^T(< R_\mathrm{HI}|M_*)$ and $P_E^c(< R_\mathrm{HI}|M_*)$ are the ETGs and ETG centrals CDFs analytic fits given by eq.(5.3.3), respectively. f_c^E and f_s^E are the fractions of ETGs that are centrals and satellites, $f_c^E + f_s^E = f^E$.

List of Publications

Referred journals:

[1] Papers published in referred journals.

Bibliography

Abazajian K. N., et al., 2009, ApJS, 182, 543

Accurso G., et al., 2017, MNRAS, 470, 4750

Ai M., Zhu M., 2018, ApJ, 862, 48

Andreani P., Boselli A., Ciesla L., Vio R., Cortese L., Buat V., Miyamoto Y., 2018, A&A, 617, A33

Andrews B. H., Martini P., 2013, ApJ, 765, 140

Aver E., Olive K. A., Skillman E. D., 2015, Journal of Cosmology and Astro-Particle Physics, 2015, 011

Avila-Reese V., 2007, Astrophysics and Space Science Proceedings, 2, 115

Avila-Reese V., Firmani C., 2000, RMxAA, 36, 23

Avila-Reese V., Firmani C., Hernández X., 1998, ApJ, 505, 37

Avila-Reese V., Zavala J., Firmani C., Hernández-Toledo H. M., 2008, AJ, 136, 1340

Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681

Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, MNRAS, 373, 469

Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945

Baldry I. K., et al., 2012, MNRAS, 421, 621

Barnes D. G., et al., 2001, MNRAS, 322, 486

Basilakos S., Plionis M., Kovač K., Voglis N., 2007, MNRAS, 378, 301

Bastian N., Covey K. R., Meyer M. R., 2010, ARA&A, 48, 339

Bauermeister A., et al., 2013, ApJ, 768, 132

Baugh C. M., 2006, Reports on Progress in Physics, 69, 3101

Baugh C. M., et al., 2019, MNRAS, 483, 4922

Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379

Behroozi P. S., Wechsler R. H., Conroy C., 2013a, ApJ, 770, 57

Behroozi P. S., Wechsler R. H., Conroy C., 2013b, ApJ, 770, 57

Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143

Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289

Beltz-Mohrmann G. D., Berlind A. A., Szewciw A. O., 2020, MNRAS, 491, 5771

Benson A. J., 2012, NA, 17, 175

Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587

Berlind A. A., et al., 2003, ApJ, 593, 1

Bernardi M., Shankar F., Hyde J. B., Mei S., Marulli F., Sheth R. K., 2010, MNRAS, 404, 2087

Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697

Bernardi M., Meert A., Sheth R. K., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2016, MNRAS, 455, 4122

Bernardi M., Meert A., Sheth R. K., Fischer J.-L., Huertas-Company M., Maraston C., Shankar F., Vikram V., 2017, MNRAS, 467, 2217

Bernardi M., Sheth R. K., Dominguez-Sanchez H., Fischer J.-L., Chae K.-H., Huertas-Company M., Shankar F., 2018, MNRAS, 477, 2560

Bershady M. A., Jangren A., Conselice C. J., 2000, AJ, 119, 2645

Bertin E., Arnouts S., 1996, A&AS, 117, 393

Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2846

Blanton M. R., Moustakas J., 2009, ARA&A, 47, 159

Blanton M. R., Roweis S., 2007a, AJ, 133, 734

Blanton M. R., Roweis S., 2007b, AJ, 133, 734

Blanton M. R., et al., 2005a, AJ, 129, 2562

Blanton M. R., Eisenstein D., Hogg D. W., Schlegel D. J., Brinkmann J., 2005b, ApJ, 629, 143

Blanton M. R., Lupton R. H., Schlegel D. J., Strauss M. A., Brinkmann J., Fukugita M., Loveday J., 2005c, ApJ, 631, 208

Blanton M. R., Kazin E., Muna D., Weaver B. A., Price-Whelan A., 2011, AJ, 142, 31

Blitz L., Rosolowsky E., 2006, ApJ, 650, 933

Blyth S., et al., 2015, Advancing Astrophysics with the Square Kilometre Array (AASKA14), p. 128

Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207

Boselli A., Gavazzi G., 2006, PASP, 118, 517

Boselli A., Lequeux J., Gavazzi G., 2002, A&A, 384, 33

Boselli A., et al., 2010, PASP, 122, 261

Boselli A., Cortese L., Boquien M., 2014a, A&A, 564, A65

Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Lagos C., Saintonge A., 2014b, A&A, 564, A66

Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Gavazzi G., Lagos C., Saintonge A., 2014c, A&A, 564, A67

Bothwell M. S., et al., 2014, MNRAS, 445, 2599

Bradford J. D., Geha M. C., Blanton M. R., 2015, ApJ, 809, 146

Braun R., 2012, ApJ, 749, 87

Bravo-Alfaro H., Cayatte V., van Gorkom J. H., Balkowski C., 2000, AJ, 119, 580

Brown T., Catinella B., Cortese L., Kilborn V., Haynes M. P., Giovanelli R., 2015, MNRAS, 452, 2479

Brown G. M., Johnston K. G., Hoare M. G., Lumsden S. L., 2016, MNRAS, 463, 2839

Brown T., et al., 2017, MNRAS, 466, 1275

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Buckley J., James I., 1979, Biometrika, 66, 429

Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343

Butcher Z., Schneider S., van Driel W., Lehnert M. D., Minchin R., 2016, A&A, 596, A60

Butcher Z., Schneider S., van Driel W., Lehnert M. D., 2018, A&A, 619, A89

Calette A. R., Avila-Reese V., Rodríguez-Puebla A., Hernández-Toledo H., Papastergis E., 2018, RMxAA, 54, 443

Campbell D., van den Bosch F. C., Hearin A., Padmanabhan N., Berlind A., Mo H. J., Tinker J., Yang X., 2015, MNRAS, 452, 444

Campbell D., van den Bosch F. C., Padmanabhan N., Mao Y.-Y., Zentner A. R., Lange J. U., Jiang F., Villarreal A., 2018, MNRAS, 477, 359

Cannon J. M., et al., 2015, AJ, 149, 72

Cappellari M., et al., 2011, MNRAS, 413, 813

Carilli C. L., 2014, preprint, (arXiv:1408.5317)

Carilli C. L., Rawlings S., 2004, NAR, 48, 979

Catinella B., et al., 2010, MNRAS, 403, 683

Catinella B., et al., 2012, A&A, 544, A65

Catinella B., et al., 2013, MNRAS, 436, 34

Catinella B., et al., 2018, MNRAS, 476, 875

Cattaneo A., Dekel A., Faber S. M., Guiderdoni B., 2008, MNRAS, 389, 567

Cattaneo A., et al., 2009, Natur, 460, 213

Chabrier G., 2003, PASP, 115, 763

Chauhan G., Lagos C. d. P., Obreschkow D., Power C., Oman K., Elahi P. J., 2019, MNRAS, 488, 5898

Chauhan G., Lagos C. d. P., Stevens A. R. H., Obreschkow D., Power C., Meyer M., 2020, MNRAS, 498, 44

Chilingarian I. V., Melchior A.-L., Zolotukhin I. Y., 2010, MNRAS, 405, 1409

Cid Fernandes R., Mateus A., Sodré L., Stasińska G., Gomes J. M., 2005, MNRAS, 358, 363

Conroy C., 2013, ARA&A, 51, 393

Conroy C., Wechsler R. H., 2009, ApJ, 696, 620

Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201

Conroy C., Dutton A. A., Graves G. J., Mendel J. T., van Dokkum P. G., 2013, ApJ, 776, L26

Conselice C. J., Bershady M. A., Dickinson M., Papovich C., 2003, AJ, 126, 1183

Conselice C. J., Wilkinson A., Duncan K., Mortlock A., 2016, ApJ, 830, 83

Cooke R. J., Pettini M., Steidel C. C., 2018, ApJ, 855, 102

Cooray A., Sheth R., 2002, PhR, 372, 1

Cortese L., Catinella B., Boissier S., Boselli A., Heinis S., 2011, MNRAS, 415, 1797

Cortese L., et al., 2012, A&A, 544, A101

Cortese L., Catinella B., Cook R. H. W., Janowiecki S., 2020, MNRAS, 494, L42

Crain R. A., et al., 2017, MNRAS, 464, 4204

Croton D. J., et al., 2016, ApJS, 222, 22

D'Souza R., Vegetti S., Kauffmann G., 2015, MNRAS, 454, 4027

Dalcanton J. J., Seth A. C., Yoachim P., 2007, Astrophysics and Space Science Proceedings, 3, 29

Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827

Dekel A., Silk J., 1986, ApJ, 303, 39

Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, MNRAS, 433, 1398

Deng X.-F., 2013, Research in Astronomy and Astrophysics, 13, 651

Diemer B., et al., 2018, ApJS, 238, 33

Diemer B., et al., 2019, arXiv e-prints,

Domínguez Sánchez H., Huertas-Company M., Bernardi M., Tuccillo D., Fischer J. L., 2018, MNRAS, 476, 3661

Dragomir R., Rodríguez-Puebla A., Primack J. R., Lee C. T., 2018, MNRAS, 476, 741

Dressler A., 1980, ApJ, 236, 351

Drory N., Alvarez M., 2008, ApJ, 680, 41

Duffy A. R., Kay S. T., Battye R. A., Booth C. M., Dalla Vecchia C., Schaye J., 2012, MNRAS, 420, 2799

Dutton A. A., van den Bosch F. C., Dekel A., Courteau S., 2007, ApJ, 654, 27

Dutton A. A., Conroy C., van den Bosch F. C., Prada F., More S., 2010, MNRAS, 407, 2

Eckert K. D., Kannappan S. J., Stark D. V., Moffett A. J., Norris M. A., Snyder E. M., Hoversten E. A., 2015, ApJ, 810, 166

Eddington Sir A. S., 1940, MNRAS, 100, 354

Efstathiou G., Ellis R. S., Peterson B. A., 1988, MNRAS, 232, 431

Elmegreen B. G., 1989, ApJ, 338, 178

Feigelson E. D., Babu G. J., 2012, Modern Statistical Methods for Astronomy: With R Applications. Cambridge

Feigelson E. D., Nelson P. I., 1985, ApJ, 293, 192

Fioc M., Rocca-Volmerange B., 1997, A&A, 326, 950

Firmani C., Avila-Reese V., 2000, MNRAS, 315, 457

Firmani C., Avila-Reese V., Rodríguez-Puebla A., 2010, MNRAS, 404, 1100

Fliri J., Trujillo I., 2016, MNRAS, 456, 1359

Frenk C. S., White S. D. M., 2012, Annalen der Physik, 524, 507

Frieman J. A., et al., 2008, AJ, 135, 338

Fu J., Guo Q., Kauffmann G., Krumholz M. R., 2010, MNRAS, 409, 515

Fukugita M., Peebles P. J. E., 2004, ApJ, 616, 643

Fukugita M., Hogan C. J., Peebles P. J. E., 1998, ApJ, 503, 518

Fukugita M., et al., 2007, AJ, 134, 579

Garnett D. R., 2002, ApJ, 581, 1019

Gavazzi G., Boselli A., van Driel W., O'Neil K., 2005, A&A, 429, 439

Gavazzi G., Fumagalli M., Fossati M., Galardo V., Grossetti F., Boselli A., Giovanelli R., Haynes M. P., 2013, A&A, 553, A89

Geha M., Blanton M. R., Masjedi M., West A. A., 2006, ApJ, 653, 240

Geha M., Blanton M. R., Yan R., Tinker J. L., 2012, ApJ, 757, 85

Giovanelli R., et al., 2005, AJ, 130, 2598

Glover S. C. O., Mac Low M.-M., 2011, MNRAS, 412, 337

Gnedin N. Y., Draine B. T., 2014, ApJ, 795, 37

Gnedin N. Y., Kravtsov A. V., 2011a, ApJ, 728, 88

Gnedin N. Y., Kravtsov A. V., 2011b, ApJ, 728, 88

Guo H., Li C., Zheng Z., Mo H. J., Jing Y. P., Zu Y., Lim S. H., Xu H., 2017, ApJ, 846, 61

Guo H., Jones M. G., Haynes M. P., Fu J., 2020, ApJ, 894, 92

Haynes M. P., Giovanelli R., 1984, AJ, 89, 758

Haynes M. P., et al., 2011, AJ, 142, 170

Haynes M. P., et al., 2018, ApJ, 861, 49

He Y. Q., Xia X. Y., Hao C. N., Jing Y. P., Mao S., Li C., 2013, ApJ, 773, 37

Hearin A. P., Watson D. F., 2013a, MNRAS, 435, 1313

Hearin A. P., Watson D. F., 2013b, MNRAS, 435, 1313

Hearin A. P., Zentner A. R., Berlind A. A., Newman J. A., 2013, MNRAS, 433, 659

Hearin A. P., Zentner A. R., van den Bosch F. C., Campbell D., Tollerud E., 2016, MNRAS, 460, 2552

Helfer T. T., Thornley M. D., Regan M. W., Wong T., Sheth K., Vogel S. N., Blitz L., Bock D. C.-J., 2003, ApJS, 145, 259

Henriques B. M. B., White S. D. M., Lilly S. J., Bell E. F., Bluck A. F. L., Terrazas B. A., 2019, MNRAS, 485, 3446

Hernández-Toledo H. M., Vázquez-Mata J. A., Martínez-Vázquez L. A., Choi Y.-Y., Park C., 2010, AJ, 139, 2525

Hess K. M., Wilcots E. M., 2013, AJ, 146, 124

Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, MNRAS, 442, 2304

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, MNRAS, 445, 581

Hoppmann L., Staveley-Smith L., Freudling W., Zwaan M. A., Minchin R. F., Calabretta M. R., 2015, MNRAS, 452, 3726

Hu W., et al., 2019, MNRAS, 489, 1619

Huang S., Haynes M. P., Giovanelli R., Brinchmann J., Stierwalt S., Neff S. G., 2012a, AJ, 143, 133

Huang S., Haynes M. P., Giovanelli R., Brinchmann J., 2012b, ApJ, 756, 113

Huang S., Haynes M. P., Giovanelli R., Brinchmann J., 2012c, ApJ, 756, 113

Huang S., et al., 2014, ApJ, 793, 40

Huertas-Company M., Aguerri J. A. L., Bernardi M., Mei S., Sánchez Almeida J., 2011, A&A, 525, A157

Huertas-Company M., et al., 2019, MNRAS, 489, 1859

Hyde J. B., Bernardi M., 2009, MNRAS, 396, 1171

Ibarra-Medel H. J., et al., 2016, MNRAS, 463, 2799

Janowiecki S., Catinella B., Cortese L., Saintonge A., Brown T., Wang J., 2017, MNRAS, 466, 4795

Janowiecki S., Catinella B., Cortese L., Saintonge A., Wang J., 2020, MNRAS, 493, 1982

Jansen R. A., Franx M., Fabricant D., Caldwell N., 2000a, ApJS, 126, 271

Jansen R. A., Fabricant D., Franx M., Caldwell N., 2000b, ApJS, 126, 331

Johnston S., et al., 2008, Experimental Astronomy, 22, 151

Jones M. G., Papastergis E., Haynes M. P., Giovanelli R., 2016, MNRAS, 457, 4393

Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, MNRAS,

Kannappan S. J., Gawiser E., 2007, ApJ, 657, L5

Kannappan S. J., Guie J. M., Baker A. J., 2009, AJ, 138, 579

Kannappan S. J., et al., 2013, ApJ, 777, 42

Kaplan E. L., Meier P., 1958a, Journal of the American Statistical Association, 53, pp. 457

Kaplan E. L., Meier P., 1958b, Journal of the American Statistical Association, 53, pp. 457

Karachentsev I. D., Makarov D. I., Kaisina E. I., 2013, AJ, 145, 101

Karachentsev I. D., Kaisina E. I., Makarov D. I., 2014, AJ, 147, 13

Kauffmann G., et al., 2003, MNRAS, 341, 54

Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713

Kelvin L. S., et al., 2014, MNRAS, 444, 1647

Kennicutt Jr. R. C., 1998, ApJ, 498, 541

Kennicutt Jr. R. C., et al., 2003, PASP, 115, 928

Keres D., Yun M. S., Young J. S., 2003, ApJ, 582, 659

Kim H.-S., Wyithe J. S. B., Baugh C. M., Lagos C. d. P., Power C., Park J., 2017, MNRAS, 465, 111

Klypin A., Karachentsev I., Makarov D., Nasonova O., 2015, MNRAS, 454, 1798

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Koribalski B. S., et al., 2004, AJ, 128, 16

Kovac K., Oosterloo T. A., van der Hulst J. M., 2005, in Jerjen H., Binggeli B., eds, IAU Colloq. 198: Near-fields cosmology with dwarf elliptical galaxies. pp 351–354 (`arXiv:astro-ph/0508072`), doi:10.1017/S1743921305004047

Kravtsov A., Vikhlinin A., Meshscheryakov A., 2014, preprint, (`arXiv:1401.7329`)

Kroupa P., 2001, MNRAS, 322, 231

Kroupa P., Tout C. A., Gilmore G., 1993, MNRAS, 262, 545

Krumholz M. R., 2013a, MNRAS, 436, 2747

Krumholz M. R., 2013b, MNRAS, 436, 2747

Krumholz M. R., McKee C. F., Tumlinson J., 2009, ApJ, 693, 216

Lacerna I., Hernández-Toledo H. M., Avila-Reese V., Abonza-Sane J., del Olmo A., 2016, A&A, 588, A79

Lacerna I., Ibarra-Medel H., Avila-Reese V., Hernández-Toledo H. M., Vázquez-Mata J. A., Sánchez S. F., 2020, arXiv e-prints, p. arXiv:2001.05506

Lada C. J., Lombardi M., Alves J. F., 2010, ApJ, 724, 687

Lada C. J., Forbrich J., Lombardi M., Alves J. F., 2012, ApJ, 745, 190

Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649

Lagos C. d. P., Davis T. A., Lacey C. G., Zwaan M. A., Baugh C. M., Gonzalez-Perez V., Padilla N. D., 2014, MNRAS, 443, 1002

Lagos C. d. P., et al., 2015, MNRAS, 452, 3815

Lagos C. d. P., Tobar R. J., Robotham A. S. G., Obreschkow D., Mitchell P. D., Power C., Elahi P. J., 2018, MNRAS, 481, 3573

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Lee E. T., Wang J. W., 2003, Statistical Methods for Survival Data Analysis. Wiley

Leja J., van Dokkum P., Franx M., 2013, ApJ, 766, 33

Leja J., van Dokkum P. G., Franx M., Whitaker K. E., 2015, ApJ, 798, 115

Leja J., Speagle J. S., Johnson B. D., Conroy C., van Dokkum P., Franx M., 2020, ApJ, 893, 111

Lemonias J. J., Schiminovich D., Catinella B., Heckman T. M., Moran S. M., 2013, ApJ, 776, 74

Leroy A. K., Walter F., Brinks E., Bigiel F., de Blok W. J. G., Madore B., Thornley M. D., 2008, AJ, 136, 2782

Leroy A. K., et al., 2013, AJ, 146, 19

Li D., Nan R., Pan Z., 2013, in van Leeuwen J., ed., Vol. 291, Neutron Stars and Pulsars: Challenges and Opportunities after 80 years. pp 325–330 (`arXiv:1210.5785`), doi:10.1017/S1743921312024015

Lisenfeld U., et al., 2011, A&A, 534, A102

Longair M. S., 2008, Galaxy Formation

Lu Y., Yang X., Liu C., Guo H., Xu H., Katsianis A., Wang Z., 2020, arXiv e-prints, p. arXiv:2008.09804

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Maddox N., Hess K. M., Obreschkow D., Jarvis M. J., Blyth S.-L., 2015, MNRAS, 447, 1610

Maeda F., Ohta K., Seko A., 2017, ApJ, 835, 120

Mandelbaum R., Seljak U., Kauffmann G., Hirata C. M., Brinkmann J., 2006, MNRAS, 368, 715

Mandelbaum R., Wang W., Zu Y., White S., Henriques B., More S., 2016, MNRAS, 457, 3200

Marchesini D., van Dokkum P. G., Förster Schreiber N. M., Franx M., Labbé I., Wuyts S., 2009, ApJ, 701, 1765

Marinacci F., et al., 2018, MNRAS, 480, 5113

Martin D. C., et al., 2005, ApJ, 619, L1

Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, ApJ, 723, 1359

Martin A. M., Giovanelli R., Haynes M. P., Guzzo L., 2012, ApJ, 750, 38

Masters K. L., et al., 2010a, MNRAS, 404, 792

Masters K. L., et al., 2010b, MNRAS, 405, 783

Masters K. L., et al., 2019, MNRAS, 488, 3396

McGaugh S. S., 2005, ApJ, 632, 859

Meert A., Vikram V., Bernardi M., 2013, MNRAS, 433, 1344

Meert A., Vikram V., Bernardi M., 2015, MNRAS, 446, 3943

Meert A., Vikram V., Bernardi M., 2016, MNRAS, 455, 2440

Mendel J. T., Simard L., Palmer M., Ellison S. L., Patton D. R., 2014, ApJS, 210, 3

Meyer M. J., et al., 2004, MNRAS, 350, 1195

Meyer M. J., Zwaan M. A., Webster R. L., Brown M. J. I., Staveley-Smith L., 2007,
ApJ, 654, 702

Mo H., van den Bosch F. C., White S., 2010a, Galaxy Formation and Evolution

Mo H., van den Bosch F. C., White S., 2010b, Galaxy Formation and Evolution

Moffett A. J., et al., 2016a, MNRAS, 457, 1308

Moffett A. J., et al., 2016b, MNRAS, 462, 4336

More S., van den Bosch F. C., Cacciato M., Skibba R., Mo H. J., Yang X., 2011,
MNRAS, 410, 210

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V.,
Naab T., Oser L., 2010, ApJ, 710, 903

Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121

Moster B. P., Naab T., White S. D. M., 2018, MNRAS, 477, 1822

Moustakas J., et al., 2011, preprint, (arXiv:1112.3300)

Moustakas J., et al., 2013, ApJ, 767, 50

Muzzin A., Marchesini D., van Dokkum P. G., Labbé I., Kriek M., Franx M., 2009,
ApJ, 701, 1839

Naab T., Ostriker J. P., 2017, ARA&A, 55, 59

Naiman J. P., et al., 2018, MNRAS, 477, 1206

Nair P. B., Abraham R. G., 2010a, ApJS, 186, 427

Nair P. B., Abraham R. G., 2010b, ApJS, 186, 427

Nan R., et al., 2011, International Journal of Modern Physics D, 20, 989

Narayanan D., Krumholz M. R., Ostriker E. C., Hernquist L., 2012, MNRAS, 421, 3127

Neistein E., Li C., Khochfar S., Weinmann S. M., Shankar F., Boylan-Kolchin M., 2011, MNRAS, 416, 1486

Nelson D., Genel S., Vogelsberger M., Springel V., Sijacki D., Torrey P., Hernquist L., 2015, MNRAS, 448, 59

Nelson D., et al., 2018, MNRAS, 475, 624

Noordermeer E., van der Hulst J. M., Sancisi R., Swaters R. A., van Albada T. S., 2005, A&A, 442, 137

Obreschkow D., Rawlings S., 2009, MNRAS, 394, 1857

Obreschkow D., Ma X., Meyer M., Power C., Zwaan M., Staveley-Smith L., Drinkwater M. J., 2013, ApJ, 766, 137

Obreschkow D., Murray S. G., Robotham A. S. G., Westmeier T., 2018, MNRAS, 474, 5500

Obuljen A., Alonso D., Villaescusa-Navarro F., Yoon I., Jones M., 2019, MNRAS, 486, 5124

Padmanabhan H., Kulkarni G., 2017, MNRAS, 470, 340

Padmanabhan H., Refregier A., Amara A., 2017, MNRAS, 469, 2323

Panter B., Jimenez R., Heavens A. F., Charlot S., 2007, MNRAS, 378, 1550

Panter B., Jimenez R., Heavens A. F., Charlot S., 2008, MNRAS, 391, 1117

Papastergis E., Cattaneo A., Huang S., Giovanelli R., Haynes M. P., 2012, ApJ, 759, 138

Papastergis E., Giovanelli R., Haynes M. P., Rodríguez-Puebla A., Jones M. G., 2013, ApJ, 776, 43

Paturel G., Petit C., Prugniel P., Theureau G., Rousseau J., Brouty M., Dubois P., Cambresy L., 2003, VizieR Online Data Catalog, 7237, 0

Paul N., Choudhury T. R., Paranjape A., 2018, MNRAS, 479, 1627

Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

Peng Y.-j., et al., 2010, ApJ, 721, 193

Peng Y.-j., Lilly S. J., Renzini A., Carollo M., 2012, ApJ, 757, 4

Pérez-González P. G., et al., 2008, ApJ, 675, 234

Pforr J., Maraston C., Tonini C., 2012, MNRAS, 422, 3285

Pillepich A., et al., 2018, MNRAS, 473, 4077

Planck Collaboration et al., 2016a, A&A, 594, A13

Planck Collaboration et al., 2016b, A&A, 594, A13

Planck Collaboration et al., 2018, arXiv e-prints, p. arXiv:1807.06209

Popping G., Behroozi P. S., Peeples M. S., 2015, MNRAS, 449, 477

Popping G., et al., 2019, arXiv e-prints,

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1996, Numerical Recipes in Fortran 90 (2Nd Ed.): The Art of Parallel Scientific Computing. Cambridge University Press, New York, NY, USA

Rasmussen J., Mulchaey J. S., Bai L., Ponman T. J., Raychaudhury S., Dariush A., 2012, ApJ, 757, 122

Read J. I., Trentham N., 2005a, Philosophical Transactions of the Royal Society of London Series A, 363

Read J. I., Trentham N., 2005b, Philosophical Transactions of the Royal Society of London Series A, 363, 2693

Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013a, ApJ, 771, 30

Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013b, ApJ, 771, 30

Robotham A. S. G., Bellstedt S., Lagos C. d. P., Thorne J. E., Davies L. J., Driver S. P., Bravo M., 2020, MNRAS, 495, 905

Rodriguez-Gomez V., et al., 2019, MNRAS, 483, 4140

Rodríguez-Puebla A., Drory N., Avila-Reese V., 2012, ApJ, 756, 2

Rodríguez-Puebla A., Avila-Reese V., Drory N., 2013, ApJ, 767, 92

Rodríguez-Puebla A., Avila-Reese V., Yang X., Foucaud S., Drory N., Jing Y. P., 2015, ApJ, 799, 130

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS, 462, 893

Rodríguez-Puebla A., Primack J. R., Avila-Reese V., Faber S. M., 2017, MNRAS, 470, 651

Rodríguez-Puebla A., Calette A. R., Avila-Reese V., Rodriguez-Gomez V., Huertas-Company M., 2020, PASA, 37, e024

Romeo A. B., 2020, MNRAS, 491, 4843

Saintonge A., et al., 2011, MNRAS, 415, 32

Saintonge A., et al., 2017, ApJS, 233, 22

Salim S., et al., 2007, ApJS, 173, 267

Sánchez S. F., 2020, ARA&A, 58, annurev

Sánchez S. F., et al., 2013, A&A, 554, A58

Sánchez S. F., et al., 2019, MNRAS, 482, 1557

Schaye J., et al., 2015, MNRAS, 446, 521

Schmidt M., 1968, ApJ, 151, 393

Schruba A., et al., 2012, AJ, 143, 138

Sedgwick T. M., Baldry I. K., James P. A., Kelvin L. S., 2019, MNRAS, 484, 5278

Serra P., et al., 2012, MNRAS, 422, 1835

Sérsic J. L., 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Shankar F., et al., 2014, ApJ, 797, L27

Simard L., Mendel J. T., Patton D. R., Ellison S. L., McConnachie A. W., 2011, ApJS, 196, 11

Skibba R. A., van den Bosch F. C., Yang X., More S., Mo H., Fontanot F., 2011, MNRAS, 410, 417

Solanes J. M., Manrique A., García-Gómez C., González-Casado G., Giovanelli R., Haynes M. P., 2001, ApJ, 548, 97

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Spinelli M., Zoldan A., De Lucia G., Xie L., Viel M., 2020, MNRAS, 493, 5434

Springel V., 2010, ARA&A, 48, 391

Springel V., et al., 2018, MNRAS, 475, 676

Springob C. M., Haynes M. P., Giovanelli R., Kent B. R., 2005, ApJS, 160, 149

Stark D. V., Kannappan S. J., Wei L. H., Baker A. J., Leroy A. K., Eckert K. D., Vogel S. N., 2013, ApJ, 769, 82

Stark D. V., et al., 2016, ApJ, 832, 126

Sternberg A., Le Petit F., Roueff E., Le Bourlot J., 2014, ApJ, 790, 10

Stevens A. R. H., et al., 2019, MNRAS, 483, 5334

Stewart K. R., Bullock J. S., Wechsler R. H., Maller A. H., 2009, ApJ, 702, 307

Swaters R. A., Balcells M., 2002, A&A, 390, 863

Takeuchi T. T., 2010, MNRAS, 406, 1830

Takeuchi T. T., Sakurai A., Yuan F.-T., Buat V., Burgarella D., 2013, Earth, Planets, and Space, 65, 281

Taylor E. N., et al., 2011, MNRAS, 418, 1587

Taylor R., Davies J. I., Auld R., Minchin R. F., 2012, MNRAS, 423, 787

Tempel E., et al., 2014, A&A, 566, A1

Thanjavur K., Simard L., Bluck A. F. L., Mendel T., 2016, MNRAS, 459, 44

Tinker J. L., et al., 2017, ApJ, 839, 121

Tonry J. L., Blakeslee J. P., Ajhar E. A., Dressler A., 2000, ApJ, 530, 625

Toomre A., 1964, ApJ, 139, 1217

Tramonte D., Ma Y.-Z., 2020, MNRAS,

Vale A., Ostriker J. P., 2004, MNRAS, 353, 189

Verheijen M. A. W., 1997, PhD thesis, PhD thesis, Univ. Groningen, The Netherlands , (1997)

Vikram V., Wadadekar Y., Kembhavi A. K., Vijayagovindan G. V., 2010, MNRAS, 409, 1379

Villaescusa-Navarro F., et al., 2018, ApJ, 866, 135

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt Jr. R. C., Thornley M. D., Leroy A., 2008, AJ, 136, 2563

Watts A. B., Catinella B., Cortese L., Power C., 2020, MNRAS, 492, 3672

Wechsler R. H., Tinker J. L., 2018a, ARA&A, 56, 435

Wechsler R. H., Tinker J. L., 2018b, ARA&A, 56, 435

Wei L. H., Kannappan S. J., Vogel S. N., Baker A. J., 2010a, ApJ, 708, 841

Wei L. H., Kannappan S. J., Vogel S. N., Baker A. J., 2010b, ApJ, 708, 841

Welch G. A., Sage L. J., Young L. M., 2010, ApJ, 725, 100

White S. D. M., Frenk C. S., 1991, ApJ, 379, 52

Willick J. A., Strauss M. A., Dekel A., Kolatt T., 1997, ApJ, 486, 629

Wolfire M. G., Hollenbach D., McKee C. F., 2010, ApJ, 716, 1191

Wright A. H., et al., 2017, MNRAS, 470, 283

Wright R. J., Lagos C. d. P., Davies L. J. M., Power C., Trayford J. W., Wong O. I., 2019, MNRAS, 487, 3740

Yang X., Mo H. J., van den Bosch F. C., Jing Y. P., 2005, MNRAS, 356, 1293

Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153

Yang X., Mo H. J., van den Bosch F. C., 2008, ApJ, 676, 248

Yang X., Mo H. J., van den Bosch F. C., 2009a, ApJ, 695, 900

Yang X., Mo H. J., van den Bosch F. C., 2009b, ApJ, 695, 900

Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, ApJ, 752, 41

Yoon I., Rosenberg J. L., 2015, ApJ, 812, 4

Young J. S., et al., 1995, ApJS, 98, 219

Young L. M., et al., 2011, MNRAS, 414, 940

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019, MNRAS, 483, 2983

Zehavi I., Kerby S. E., Contreras S., Jiménez E., Padilla N., Baugh C. M., 2019, ApJ, 887, 17

Zhang W., Li C., Kauffmann G., Zou H., Catinella B., Shen S., Guo Q., Chang R., 2009, MNRAS, 397, 1243

Zibetti S., Charlot S., Rix H.-W., 2009, MNRAS, 400, 1181

Zu Y., Mandelbaum R., 2015, MNRAS, 454, 1161

Zu Y., Mandelbaum R., 2016, MNRAS, 457, 4360

Zwaan M. A., Briggs F. H., Sprayberry D., Sorar E., 1997, ApJ, 490, 173

Zwaan M. A., et al., 2003, AJ, 125, 2842

Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, MNRAS, 359, L30

van Driel W., et al., 2016, A&A, 595, A118

van den Bosch F. C., Tormen G., Giocoli C., 2005, MNRAS, 359, 1029

van den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, MNRAS, 387, 79