




# CONCERTO: Simulating the CO, [CII], and [CI] line emission of galaxies in a 117 deg<sup>2</sup> field and the impact of field-to-field variance

A. Gkogkou (Αθανασία Γκόγκου)<sup>1</sup> , M. Béthermin<sup>1,2</sup> , G. Lagache<sup>1</sup> , M. Van Cuyck<sup>1</sup>, E. Jullo<sup>1</sup>, M. Aravena<sup>3</sup>, A. Beelen<sup>1</sup>, A. Benoit<sup>4,5</sup>, J. Bounmy<sup>6,5</sup>, M. Calvo<sup>4,5</sup>, A. Catalano<sup>6,5</sup>, S. Cora<sup>8,9</sup>, D. Croton<sup>11,12</sup>, S. de la Torre<sup>1</sup>, A. Fasano<sup>1</sup>, A. Ferrara<sup>13</sup>, J. Goupy<sup>4,5</sup>, C. Hoarau<sup>6,5</sup>, W. Hu<sup>1</sup>, T. Ishiyama<sup>15</sup>, K. K. Knudsen<sup>7</sup>, J.-C. Lambert<sup>1</sup>, J. F. Macías-Pérez<sup>6,5</sup>, J. Marpaud<sup>6,5</sup>, G. Mellema<sup>16</sup>, A. Monfardini<sup>4,5</sup>, A. Pallottini<sup>13</sup>, N. Ponthieu<sup>4,5</sup>, F. Prada<sup>10</sup>, Y. Roehlly<sup>1</sup>, L. Vallini<sup>13</sup>, and F. Walter<sup>14</sup>

<sup>1</sup> Aix-Marseille Univ., CNRS, CNES, LAM, Marseille, France  
e-mail: [athanasia.gkogkou@lam.fr](mailto:athanasia.gkogkou@lam.fr)

<sup>2</sup> Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France

<sup>3</sup> Núcleo de Astronomía, Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército 441, Santiago, Chile

<sup>4</sup> Univ. Grenoble Alpes, CNRS, Grenoble INP, Institut Néel, 38000 Grenoble, France

<sup>5</sup> Groupement d'Interet Scientifique KID, 38000 Grenoble and 38400 Saint-Martin d'Hères, France

<sup>6</sup> Univ. Grenoble Alpes, CNRS, LPSC/IN2P3, 38000 Grenoble, France

<sup>7</sup> Department of Space, Earth and Environment, Chalmers University of Technology Onsala Space Observatory, 439 92 Onsala, Sweden

<sup>8</sup> Instituto de Astrofísica de La Plata (CCT La Plata, CONICET, UNLP), Observatorio Astronómico, Paseo del Bosque s/n, B1900FWA La Plata, Argentina

<sup>9</sup> Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Observatorio Astronómico, Paseo del Bosque s/n, B1900FWA La Plata, Argentina

<sup>10</sup> Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía, 18080 Granada, Spain

<sup>11</sup> Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, VIC 3122, Australia

<sup>12</sup> ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Canberra, Australia

<sup>13</sup> Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy

<sup>14</sup> Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

<sup>15</sup> Institute of Management and Information Technologies, Chiba University, 1-33, Yayoi-cho, Inage-ku, Chiba 263-8522, Japan

<sup>16</sup> The Oskar Klein Centre, Department of Astronomy, Stockholm University, AlbaNova 10691, Stockholm, Sweden

Received 4 October 2022 / Accepted 29 November 2022

## ABSTRACT

In the submillimeter regime, spectral line scans and line intensity mapping (LIM) are new promising probes for the cold gas content and star formation rate of galaxies across cosmic time. However, both of these two measurements suffer from field-to-field variance. We study the effect of field-to-field variance on the predicted CO and [CII] power spectra from future LIM experiments such as CONCERTO, as well as on the line luminosity functions (LFs) and the cosmic molecular gas mass density that are currently derived from spectral line scans. We combined a 117 deg<sup>2</sup> dark matter lightcone from the Uchuu cosmological simulation with the simulated infrared dusty extragalactic sky (SIDES) approach. The clustering of the dusty galaxies in the SIDES-Uchuu product is validated by reproducing the cosmic infrared background anisotropies measured by *Herschel* and *Planck*. We find that in order to constrain the CO LF with an uncertainty below 20%, we need survey sizes of at least 0.1 deg<sup>2</sup>. Furthermore, accounting for the field-to-field variance using only the Poisson variance can underestimate the total variance by up to 80%. The lower the luminosity is and the larger the survey size is, the higher the level of underestimate. At  $z < 3$ , the impact of field-to-field variance on the cosmic molecular gas density can be as high as 40% for the 4.6 arcmin<sup>2</sup> field, but drops below 10% for areas larger than 0.2 deg<sup>2</sup>. However, at  $z > 3$  the variance decreases more slowly with survey size and for example drops below 10% for 1 deg<sup>2</sup> fields. Finally, we find that the CO and [CII] LIM power spectra can vary by up to 50% in 1 deg<sup>2</sup> fields. This limits the accuracy of the constraints provided by the first 1 deg<sup>2</sup> surveys. In addition the level of the shot noise power is always dominated by the sources that are just below the detection thresholds, which limits its potential for deriving number densities of faint [CII] emitters. We provide an analytical formula to estimate the field-to-field variance of current or future LIM experiments given their observed frequency and survey size. The underlying code to derive the field-to-field variance and the full SIDES-Uchuu products (catalogs, cubes, and maps) are publicly available.

**Key words.** galaxies: ISM – galaxies: high-redshift – galaxies: luminosity function, mass function – galaxies: star formation – cosmic background radiation – large-scale structure of Universe

## 1. Introduction

The cosmic star formation rate density (SFRD), which is an integral constraint on star formation averaged over the volume of

the observable universe at a given redshift, is a critical measure for theoretical models. Several ultraviolet (UV), optical, and infrared (IR) surveys have aimed to constrain the SFRD up to  $z \sim 10$  (Madau & Dickinson 2014; Bouwens et al. 2015;

Ishigaki et al. 2018; Gruppioni et al. 2020; Khusanova et al. 2021; Fudamoto et al. 2021; Zavala et al. 2021; Wang et al. 2021) revealing that the SFRD rises at early times, peaks at  $z \sim 2$  and drops again toward present day.

The cold molecular gas is tightly linked to the star formation through cosmic time, since it constitutes the fuel of star formation in galaxies (see Carilli & Walter 2013; Tacconi et al. 2020, for a review). It is therefore essential to probe its abundance during different epochs in order to understand the cosmic star formation history (SFH). The molecular hydrogen lines are either too faint to be observed or the UV absorption is very strong. Additionally since they are quadrupole transitions, they require a high temperature to be excited. They thus trace only warm/dense molecular gas (e.g., shock-heated molecular gas), which is a minor fraction of the total amount of the molecular gas. Therefore, the most suitable tracer of the cold molecular gas is the carbon monoxide molecule CO (with  $^{12}\text{C}^{16}\text{O}$  being the most common CO isotope), which is the second most abundant molecule in the Universe and its rotational lines are bright enough to observe even at high redshifts (e.g., Venemans et al. 2017; Decarli et al. 2022).

The method of spectral line scans is a powerful tool to study the evolution of the molecular gas throughout cosmic time. With this method we can search for targets in a given volume without preselecting them, avoiding biasing the sample toward given galaxy types. Several such surveys have aimed to constrain the CO luminosity function (Walter et al. 2014; Decarli et al. 2019, 2020; Riechers et al. 2019), which is the comoving volume density of sources per luminosity, and infer the cosmic molecular gas density at different redshifts. They found that there is an evident evolution of the luminosity functions (LFs) with redshift and, as a consequence, of molecular gas abundance ( $\rho_{\text{H}_2}$ ). The cosmic molecular gas density increases at early times, peaks at  $z \sim 1-3$ , and finally decreases down to the present day, suggesting that there is indeed a coevolution with the SFRD (see also Walter et al. 2020).

Several observational studies have investigated the SFRD at different redshifts, but there is limited knowledge of its spatial distribution at large scales. The fluctuations of the integrated emission of dust in galaxies across cosmic times, which is the cosmic infrared background (CIB), can fill this gap of information, especially at high redshifts (Lagache et al. 2007; Viero et al. 2009; Planck Collaboration XVIII 2011; Amblard et al. 2011; Planck Collaboration XXX 2014). However, it is difficult to break the degeneracies among the different redshift slices especially at  $z > 4$  (Maniyar et al. 2018).

The line intensity mapping (LIM) technique can break these degeneracies. LIM measures the spatial fluctuations of the emission of a given spectral line in multiple frequencies, obtaining a 3D map (see Kovetz et al. 2017, for a review). Therefore, using different lines, it is possible to trace the star formation and gas content in a specific redshift slice. The [CII] line at  $158\mu\text{m}$  is among the brightest far-infrared emission lines and it has been found that there is an empirical correlation between the [CII] emission and the star formation rate (SFR; De Looze et al. 2014; Lagache et al. 2018; Schaerer et al. 2020). Moreover, Zanella et al. (2018) and Vizgan et al. (2022) argue that [CII] is a convenient molecular gas tracer, especially at  $z > 5$ . Theoretical studies also support this picture as shown in Pallottini et al. (2017) and Ferrara et al. (2019) via a numerical simulation and analytical model, respectively. This therefore constitutes [CII] as a great candidate for LIM.

Multiple on-going experiments aim to measure the [CII] fluctuations at high redshifts, such as the Carbon [CII] line in

post-reionization and reionization epoch project (CONCERTO, CONCERTO Collaboration 2020), the instrumentation for the tomographic ionized-carbon intensity mapping experiment (TIME, Crites et al. 2014), and the Prime-Cam spectro-imager (mounted on the Fred Young Submillimeter Telescope telescope - FYST, Stacey et al. 2018). A strong advantage of LIM experiments is that they can observe much larger areas compared to traditional deep spectral scans. For example, compared to GOODS-North/CO Luminosity Density at High Redshift (COLDz,  $0.014\text{ deg}^2$  field), CONCERTO will observe a  $1.4\text{ deg}^2$  field, that is a 100 times larger area.

Nevertheless, the field-to-field variance effect (also referred as cosmic variance in the literature) can still be an obstacle. Rare and randomly distributed bright sources can significantly alter the level of shot noise, while the density fluctuations of faint sources at large scales can impact the clustering component of the power spectra. Therefore the selection of the observed region could introduce significant uncertainties on the power spectra of the matter density or the line intensity fluctuations. It can also affect several other observables, such as the number density of galaxies, the LFs, and all the inferred quantities.

There are studies that investigate the effect of the field-to-field variance on the galaxy number density and other related quantities like the LFs. Moster et al. (2011) used N-body simulations and a recipe for the computation of cosmic variance to quantify its significant excess with respect to the Poisson variance. Similarly, Driver & Robotham (2010) used Sloan Digital Sky Survey (SDSS)  $0.03 < z < 0.1$  data to measure the cosmic variance as a function of survey volume and field aspect ratio. These studies reached a similar conclusion, that any density measurement of normal galaxies is subject to uncertainties derived from the cosmic variance. More specifically, they show that the cosmic variance in the number density of galaxies is  $\sim 70\%$  for  $1\text{ deg}^2$  fields and drops to  $\sim 25\%$  for  $100\text{ deg}^2$  fields. At such low redshift, big areas in the sky correspond to small comoving volumes. The variance effect is thus expected to be weaker at high redshift.

Keenan et al. (2020) used simulated data to investigate the effect of the cosmic variance on the shape of the CO LF and the evolution of the molecular gas mass density with redshift. They found that for a survey size of  $\sim 50\text{ arcmin}^2$  and apparent luminosity of  $\sim 10^{10}\text{ K km s}^{-1}\text{ pc}^2$  the Poisson and cosmic variance uncertainties are equal. Hence, not accounting for cosmic variance leads to underestimating of the total uncertainty by a factor of  $\sim \sqrt{2}$ . Regarding the molecular gas mass density, they point out that the volume required to detect the evolution around the peak of cosmic SFRD should be larger than  $100\text{ arcmin}^2$ .

Although intensity mapping is gaining more and more attention as a technique, there are not enough studies on the effect of the field-to-field variance on the power spectrum. The clustering part of the power spectrum dominates at large scales ( $k \lesssim 1\text{ arcmin}^{-1}$ ) and is linked to the distribution of galaxies in the large-scale structure. The shot noise appears due to the randomly distributed bright galaxies and it is dominant at small scales ( $k \gtrsim 4\text{ arcmin}^{-1}$ ). Estimating the expected variance of these two components is key for scientific interpretation of intensity mapping data.

In this paper, we combined the Simulated Infrared Dusty Extragalactic Sky (SIDES), which is a simulation of the far-infrared and submillimeter sky based on observed empirical relations, with the Uchuu<sup>1</sup> N-body cosmological dark matter -only simulation (Ishiyama et al. 2021). Thanks to the large volume

<sup>1</sup> <http://www.skiesanduniverses.org/Simulations/Uchuu/>

that Uchuu provides alongside its high mass resolution, we can study the field-to-field variance of the power spectra obtained by line intensity maps, as well as its effect on other observables, as luminosity functions.

The paper is organized as follows. In Sect. 2 we briefly describe the Simulated Infrared Dusty Extragalactic Sky (SIDES) approach as well as the building and the properties of the 117 deg<sup>2</sup> simulated lightcone. In Sect. 3 we compare the simulated power spectra with observations of the CIB fluctuations in order to validate our simulation. In Sect. 4 we also compare the simulated LFs with CO and [CII] observations and we investigate the field-to-field variance of the LFs and the cosmic molecular gas density. In Sect. 5 we investigate the contribution to the shot noise. We present a model that can estimate the field-to-field variance of the power spectra given the characteristics of any LIM experiment. Additionally, using this model we make forecasts for CONCERTO and other current/future LIM experiments. Finally, in Sect. 6 we summarize our conclusions.

For the SIDES simulation we assume a [Planck Collaboration XIII \(2016\)](#) cosmology, while for the Uchuu cosmological simulation we assume a [Planck Collaboration VI \(2020\)](#) cosmology.<sup>2</sup> Throughout the paper we use a [Chabrier \(2003\)](#) initial mass function (IMF).

## 2. Simulations

### 2.1. Simulated Infrared Dusty Extragalactic Sky (SIDES)

SIDES is a simulation of the far-infrared (FIR) and submillimeter sky based on observed empirical relations. The initial frameworks ([Sargent et al. 2012](#); [Béthermin et al. 2012](#)) connect the stellar mass with several properties of galaxies (e.g., SFR  $L_{\text{IR}}$ ,  $L_{\text{FIR}}$ ) using a population of normal galaxies (e.g., [Daddi et al. 2007](#); [Schreiber et al. 2015](#)) and starburst galaxies. These are the galaxies above the main sequence of star forming galaxies, which is the tight relation between galaxy SFR and stellar mass. This formalism is extended in [Béthermin et al. \(2013, 2017\)](#) to perform the connection with dark matter halos using abundance matching. Finally, the far-infrared and submillimeter lines are included in [Béthermin et al. \(2022; hereafter B22\)](#). The total size of the simulation presented in B22 is 2 deg<sup>2</sup> and the maximum redshift is 10, although the SIDES model is reliable only up to redshift 7 as also explained in B22. Moreover, in B22 for dark matter, SIDES uses the Bolshoi-Planck cosmological simulation ([Rodríguez-Puebla et al. 2016](#)), which has a volume of (250 h<sup>-1</sup> Mpc)<sup>3</sup> and dark matter particles of  $1.5 \times 10^8 \text{ h}^{-1} M_{\odot}$  mass (halos resolved at  $\geq 100$  particles).

A dark matter lightcone, which encapsulates the information of the position and abundance of the dark matter halos, serves as the starting point for the SIDES simulation. The stellar masses of the galaxies populating the halos are determined using an abundance matching technique (e.g., [Kravtsov et al. 2004](#); [Vale & Ostriker 2004](#)). Subsequently, the generated galaxies are split into passive and star forming with a probability determined by observations ([Davidzon et al. 2018](#)). It is assumed that only the star forming galaxies emit in the FIR and millimeter, so only these type of galaxies need to be assigned with a SFR value. This assumption is supported by the results presented in [Whitaker et al. \(2021\)](#), where it was found that passive galaxies

are extremely faint in the submillimeter regime, so essentially their fluxes do not contribute to the FIR/submm sky.

The stellar mass of a galaxy is highly correlated with its SFR. Therefore, in our model the SFR of the galaxies is defined by their stellar mass. The SFR values of the main sequence and starburst galaxies are drawn accordingly based on the parameterized fit of the SFR- $M_{\text{stellar}}$  relation described in [Schreiber et al. \(2015\)](#). We add a scatter of 0.3 dex to this relation. The drawn SFR value defines the  $L_{\text{IR}}$  of each galaxy ([Kennicutt 1998](#)) and consequently the normalization of its spectral energy distribution (SED). SEDs are selected from the library given in [Magdis et al. \(2012\)](#) and they are updated at  $z > 2$  in [Béthermin et al. \(2015, 2017\)](#). The shape of the SED depends on the galaxy type (main sequence or starburst) and the mean intensity radiation field ( $\langle U \rangle$ ), which is correlated to the temperature of the dust. It is modeled such as  $\langle U \rangle = 1$  corresponds to the local interstellar UV radiation field. Finally, a magnification value ( $\mu$ ) due to the lensing effect is randomly drawn from the distributions presented in [Hezaveh & Holder \(2011\)](#) and [Hilbert et al. \(2007\)](#) for strong and weak lensing, respectively. Therefore, all the sources are either magnified or demagnified following the distributions. This is a simplification since the magnification is not consistent with the underlying dark matter simulation.

SIDES can also simulate the emission of the strongest high-redshift submillimeter lines as [CII], CO and [CI]. The [CII] emission of the galaxies is generated using the  $L_{\text{[CII]}}$ -SFR relation either from [De Looze et al. \(2014, DL14\)](#) or [Lagache et al. \(2018, L18\)](#). The L18 relation predicts a lower [CII] emission at high redshifts. The DL14 relation was derived in the local Universe; however, its validity at high redshift was tested using the follow-up of 118 optically selected galaxies at  $4.4 < z < 5.9$  recently observed with ALMA ([Capak et al. 2015](#); [Le Fèvre et al. 2020](#); [Béthermin et al. 2020](#); [Faisst et al. 2020](#); [Schaerer et al. 2020](#)). It was found that the DL14 relation still holds at these redshifts. [Carniani et al. \(2020\)](#) reached the same result for galaxies at  $z = 6-7$ . A caveat of this recipe is that environmental dependencies are not taken into account.

For CO, the fundamental transition is modeled from the  $L_{\text{IR}} - L'_{\text{CO}(1-0)}$  correlation ([Sargent et al. 2014](#)) for the main sequence galaxies, while for the starbursting systems there is an offset of  $-0.46$  dex for  $L_{\text{IR}}$  at a given  $L'_{\text{CO}(1-0)}$ .  $L'_{\text{CO}(1-0)}$  is the observed luminosity and  $L_{\text{IR}}$  is the integrated continuum emission over 8–1000  $\mu\text{m}$ . The flux of the other transitions is computed using a clumpy and diffuse spectral line energy distribution (SLED) template from [Bournaud et al. \(2015\)](#) and following the empirical relation presented in [Daddi et al. \(2015\)](#) which connects the flux ratio of CO(5–4) and CO(2–1) transitions with  $\langle U \rangle$ .

Finally, the two transitions of [CI] also contribute to the millimeter regime. The modeling of the first transition is achieved by taking advantage of the tight correlation between  $L_{\text{[CI]}(1-0)}$  and  $L_{\text{CO}(4-3)}$ . The flux of the second transition is obtained using the correlation between the ratio of the two [CI] transitions and the ratio of CO(7–6) and CO(4–3) fluxes as calibrated and presented in B22.

### 2.2. The Uchuu dark matter simulation

Nowadays cosmological simulations often achieve either high mass resolution or large volumes ([Skillman et al. 2014](#); [Ishiyama et al. 2012](#); [Rodríguez-Puebla et al. 2016](#); [Potter et al. 2017](#); [Cheng et al. 2020](#)). The ultimate goal, however, is to satisfy both requirements simultaneously. Significant progress has been

<sup>2</sup> The discrepancy of the luminosity distance as computed using both cosmologies is less than 0.1%. We thus skipped any recalibration of the used quantities.



**Table 1.** Percentage of correlated and shot noise (SN) power we miss due to the halo mass limit of the cosmological simulation we use.

$z$	[CII] <sub>correlated</sub>		CO(5–4) <sub>correlated</sub>		[CII] <sub>SN</sub>		CO(5–4) <sub>SN</sub>	
	B22 (%)	this work (%)	B22 (%)	this work (%)	B22 (%)	this work (%)	B22 (%)	this work (%)
1	1.53	0.68	3.87	2.94	0.41	0.05	0.48	0.13
2	0.38	0.18	1.16	1.11	0.04	0.0	0.1	0.01
3	2.0	1.33	8.58	7.73	0.02	0.0	0.1	0.03
4	1.67	0.96	6.82	5.66	0.01	0.0	0.06	0.01
5	2.09	1.04	8.53	6.36	0.01	0.0	0.07	0.01
6	1.71	1.05	7.04	6.2	0.01	0.0	0.07	0.01

**Notes.** We give the values for this work (SIDES-Uchuu) and B22 (SIDES-Bolshoi) for both [CII] and CO(5–4) at different redshift.

made by zoom-in cosmological simulations (e.g., Hopkins et al. 2018; Pallottini et al. 2022). The Uchuu collaboration has created a suite of N-body cosmological simulations with various comoving volumes and mass resolutions (Ishiyama et al. 2021). The Uchuu suite consists of four simulations: Uchuu, mini-Uchuu, micro-Uchuu, and shin-Uchuu with comoving volumes of  $2000 h^{-1} \text{Mpc}$ ,  $400 h^{-1} \text{Mpc}$ ,  $100 h^{-1} \text{Mpc}$ ,  $140 h^{-1} \text{Mpc}$ , respectively and mass resolutions of  $3.27 \times 10^8 h^{-1} M_{\odot}$  (halos resolved at  $\geq 40$  particles) for all except for shin-Uchuu which has a resolution of  $8.97 \times 10^5 h^{-1} M_{\odot}$ . For our study, we chose Uchuu, the biggest simulation in size, because it satisfies both requirements of large comoving volume and sufficient mass resolution. With this new simulation in hand, we can thus extensively validate the SIDES simulation and study the field-to-field variance introduced in the intensity mapping observables.

The halo masses of the lightcone we obtained span almost 6 orders of magnitude at  $z = 0$ , while the range gradually decreases to 3–4 orders of magnitude up to  $z = 7$ . After comparing the halo mass functions of Uchuu with the theoretical ones from Despali et al. (2016) at various redshifts, we found that the Uchuu lightcone is complete down to  $\sim 1.3 \times 10^{10} h^{-1} M_{\odot}$  at all redshifts. To investigate the effect of this halo mass limit, we computed the missing fraction of the correlated and shot noise power as a function of the galaxies host halo mass. We followed the same steps as in B22 (Sect. 4.4), where they estimate the contribution of galaxies to the shot noise and clustering of the [CII] power spectrum, as a function of their host halo mass. They do it by computing the sum of the luminosity squared and the product of the luminosity by the linear bias, respectively. Our results are summarized in Table 1. The effect of the halo mass limit is small for both SIDES-Bolshoi and SIDES-Uchuu but it is even smaller in SIDES-Uchuu. This is because at  $z \geq 0.5$  the Bolshoi simulation becomes gradually less complete than Uchuu, even though at  $z = 0$  both simulations share a similar level of completeness.

The cosmological parameters in the Uchuu simulation are consistent with the results from Planck Collaboration VI (2020) and the initial redshift is set to 127. The rest of the initial conditions are generated using the parallel 2LPTic code<sup>3</sup> (Crocce et al. 2006). The GreeM code<sup>4</sup> (Ishiyama et al. 2009, 2012) has been used to track the gravitational evolution while the ROCKSTAR<sup>5</sup> (Behroozi et al. 2013b) (sub)halo finder has been used for the identification of the position and velocity of each

halo and subhalo. Then merger trees were constructed using the CONSISTENT TREES code<sup>6</sup> (Behroozi et al. 2013c).

The outputs of the cosmological simulation are saved as discrete snapshots which represent the evolution stage at different time steps and, hence, different redshifts. We used the different snapshots of the simulation box to create a  $9 \text{ deg} \times 13.6 \text{ deg}$  lightcone within the redshift range  $0 < z < 7$ , which corresponds to a comoving volume of  $7.9 \text{ Gpc}^3$ . The total simulation area we have is  $122 \text{ deg}^2$  ( $9 \text{ deg} \times 13.6 \text{ deg}$ ) but the final exploitable simulation area is  $117 \text{ deg}^2$ , because we cut the total area into smaller square subfields of  $1 \text{ deg}^2$ . For the construction of the lightcone, we remapped the  $(2 \text{ Gpc } h^{-1})^3$  cubical volume of the simulation, following the prescription described in Carlson & White (2010). Taking advantage of the periodicity of the simulated box we broke it into cells, which were then translated by integer offsets to form cuboids. This remapping procedure keeps local structures intact, meaning that the structures inside each cuboid are not cut in half but they rather keep their continuity. We adjusted the remapping parameters in order to get a lightcone long enough ( $z \sim 7$ ). Finally, for each snapshot, we remapped the positions of the halos and subhalos converting from the coordinates of the simulation box to the ones of the newly created lightcone.

### 2.3. From dark matter halos to galaxies: abundance matching

Galaxies form in the gravitational potential wells of the dark matter halos. We thus need to connect the galaxies with their dark matter halos, which can be succeeded by abundance matching. This empirical technique assumes that there is a monotonic relation between some property of the galaxies and some property of the dark matter halos (Kravtsov et al. 2004; Vale & Ostriker 2004; Conroy et al. 2006; Shankar et al. 2006; Behroozi et al. 2013a; Moster et al. 2013).

There are several proxies that can be used to apply the abundance matching method. In the case of galaxies it can be either the stellar mass or the luminosity. In the case of the dark matter halos it could be either the halo mass or the circular velocity. The evolutionary stage of the halo at which we compute these quantities can strongly affect the results. For instance, during a merging episode between two or more dark matter halos, their dark matter content gets stripped faster than their stars. For instance, Niemiec et al. (2019) investigated the stripping timescale of the subhalos and satellite galaxies confirming the theoretically expected trend. Therefore, the usual technique is to

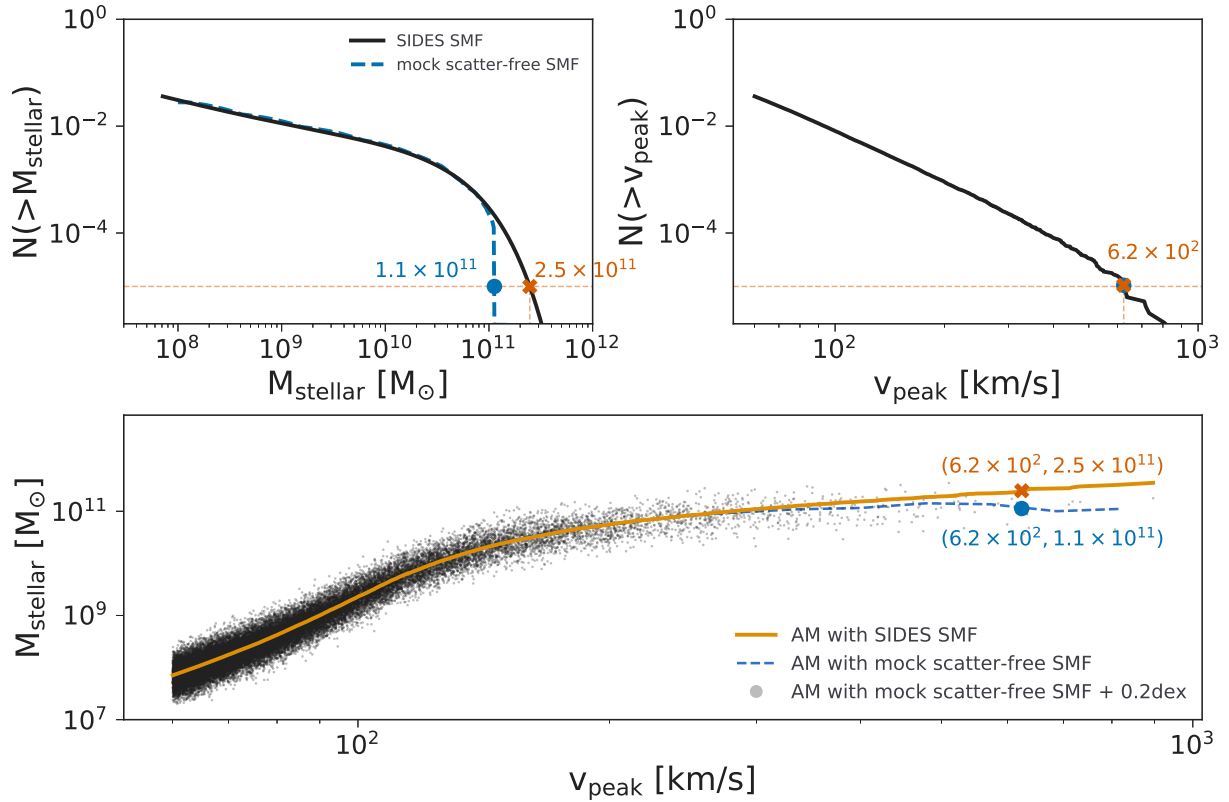
<sup>3</sup> <http://cosmo.nyu.edu/roman/2LPT/>

<sup>4</sup> <http://hpc.imit.chiba-u.jp/~ishiytm/greem/>

<sup>5</sup> <https://bitbucket.org/gfcanford/rockstar/>

<sup>6</sup> <https://bitbucket.org/pbehroozi/consistent-trees/src/main/>





**Fig. 1.** Explanation of our abundance matching (AM) method between dark matter halos and galaxies. In the *top left panel* we show the cumulative SIDES stellar mass function (SMF, black solid line) and the mock scatter-free SMF (dashed blue line, Sect. 2.3) which is obtained after the deconvolution of the SIDES SMF with the 0.2 dex scatter. In the *top right panel* we show the cumulative halo velocity function. In the *bottom panel* we show the resulting relation between the stellar mass and the halo peak velocity when performing direct abundance matching using the SIDES SMF (orange line) and the mock scatter-free SMF (dashed blue line). The black points are the stellar masses assigned to the halos of our simulated catalog using the scatter-free relation (blue dashed line) and adding a 0.2 dex log-normal scatter. All plots are made for the  $0.05 < z < 0.1$  redshift range.

place the  $N$ -th most massive galaxy in the halo with the  $N$ -th highest velocity (or highest halo mass).

The galaxy properties are strongly correlated with the halo properties before any stripping event. This is why the most suitable proxies for matching the abundances of the dark matter halos with those of the galaxies are either the  $M_{\text{peak}}$  or the  $v_{\text{peak}}$  (Behroozi et al. 2019, 2020).  $M_{\text{peak}}$  is the maximum halo mass throughout the entire past merging history of the halo and  $v_{\text{peak}}$  is the maximum circular velocity throughout the entire merging history of the halo.

In Reddick et al. (2013) it is shown that  $v_{\text{peak}}$  gives better results at low redshifts. They reach to this conclusion after comparing the projected two-point correlation function of the SDSS catalog and the two-point correlation function of mock catalogs. The latter are constructed after populating the Bolshoi cosmological simulation (Klypin et al. 2011) with abundance-matched galaxies using several different proxies. It turns out that only the  $v_{\text{peak}}$ -based abundance-matched galaxies can properly recreate the SDSS two-point correlation function.

In order to investigate the effect of the selection of the proxy quantity on our results we used both quantities ( $M_{\text{peak}}$ ,  $v_{\text{peak}}$ ). While the abundance matching quantity can have an impact on the two-point correlation function, as described above, we show in Sect. 3.3 that this choice has no impact on the CIB anisotropies. We also checked and found that the impact of this choice on CO and [CII] intensity mapping power spectra is lower than 10%, which is much less

than the typical uncertainty between different line emission models.

While the observed SMF, which is used as a starting point in SIDES, is already corrected for the scatter between the true and the measured SMF, we still have to take into account that not all the galaxies are exactly on the relation between the proxy and the stellar mass. Reddick et al. (2013) estimates that the scatter around this relation is  $\sim 0.2$  dex. In the presence of scatter, the abundance matching is thus more complex. We used the approach of Behroozi et al. (2010), implemented in the `abundancematching`<sup>7</sup> python module developed by Y. Mao, to deal with the scatter. The code deconvolves the SIDES SMF by the 0.2 dex log-normal scatter to obtain a mock SMF. Then it performs abundance matching using this mock SMF and obtains a proxy- $M_{\text{stellar}}$  relation which is scatter-free. Finally, the stellar mass assigned to each dark matter (sub)halo is the value defined by the scatter-free proxy- $M_{\text{stellar}}$  relation plus the 0.2 dex log-normal scatter around this value.

The different outcomes of abundance matching with or without considering scatter are shown in Fig. 1. The mean of the black points coincides with the resulting  $M_{\text{stellar}} - v_{\text{peak}}$  relation when using the virtual scatter-free SMF. One can notice that at high  $v_{\text{peak}}$  values the mean of the black points (blue dashed line) is below the  $M_{\text{stellar}} - v_{\text{peak}}$  relation resulting after the direct abundance matching between the SIDES SMF and halo peak

<sup>7</sup> <https://github.com/yymao/abundancematching>

velocity function, which ignores the scatter. Compared to the virtual scatter-free SMF, the SIDES “true” SMF has a shallower massive end and implies a steeper relation.

We applied the SIDES method to derive the galaxy properties determined above, from the stellar masses and redshifts. The resulting simulated SIDES-Uchuu catalog is publicly available<sup>8</sup>.

### 3. Validation of the model at large scales using CIB anisotropies

The original SIDES-Bolshoi simulation (B  thermin et al. 2017) is validated by comparing several observables with real data, like the observed continuum number counts from the mid-infrared to the millimeter and the *Herschel* data of CIB anisotropies. The latter aims to validate the ability of SIDES to reproduce the clustering of galaxies at intermediate scales ( $\lesssim 1$  deg). However, the much larger simulation area of SIDES-Uchuu ( $117 \text{ deg}^2$  compared to  $2 \text{ deg}^2$  for SIDES-Bolshoi) offers a more robust validation of the simulation recipes at large scales.

The clustering of SIDES is validated by comparing it with data of individually detected dusty galaxies (Cooray et al. 2010; Maddox et al. 2010). However, measuring the clustering of individual galaxies can be problematic due to the confusion impacting the source extraction at these wavelengths. Because of this difficulty, the measurements disagree with each other depending on the source extraction technique. It is thus impossible to validate a model using the measured two-point correlation function. The CIB power spectrum, though, does not suffer from such limitations. Thus, comparing our model with the CIB power spectra anisotropies offers a more accurate model validation.

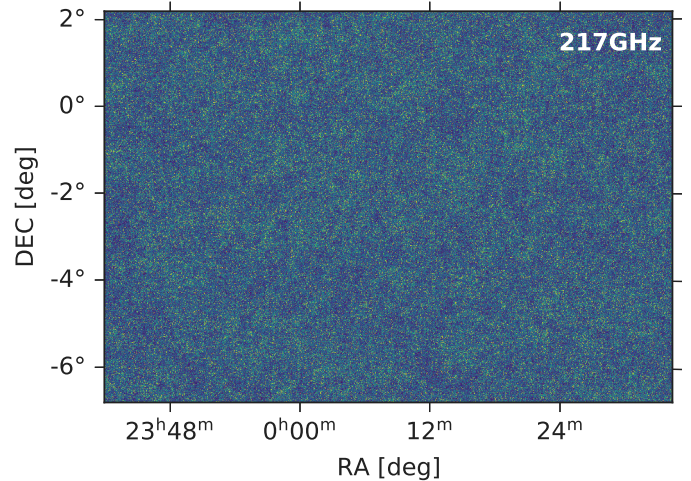
#### 3.1. Simulated maps

The simulated catalogs created by SIDES contain a comprehensive set of information for each galaxy, like its position on the sky, its redshift and its flux in the bandpasses of given experiments. Using the information of the simulated catalogs we can generate maps at different wavelengths, for example, *Spitzer*, *Herschel*, NIKA2, *Planck*-HFI. For more details on the map generator we refer to B22. The generated maps include the emission of the sources at low angular resolution (in the confusion-limited regime). They thus simulate the CIB, which is the cumulative emission of all dusty galaxies along every line of sight.

We generated maps for the *Herschel*/SPIRE three frequency bands, 600, 857, and 1200 GHz (Amblard et al. 2011; Viero et al. 2013; Thacker et al. 2013) as well as for the *Planck* high frequency bands, at 217, 353, 545, and 857 GHz (Planck Collaboration XVIII 2011). In Fig. 2 we show an exemplary simulated map for *Planck* at 217 GHz.

#### 3.2. Power spectrum estimate from the model and comparison with observations

We computed the auto power spectrum of each of the maps mentioned in Sect. 3.1 as well as the cross power spectra for all the combinations of bands that have observational constraints. We compared the model with CIB observational data from *Planck* and *Herschel* in order to assess how well SIDES can reproduce the clustering of the galaxies. For the power spectra computation we used the *powspec*<sup>9</sup> python package, which computes



**Fig. 2.** Simulated  $13.6^\circ \times 9^\circ$  CIB *Planck* map at 217 GHz. Sources with flux higher than 225 mJy have been removed to mimic the analysis of CIB anisotropies of Planck Collaboration XXX (2014).

the square of the Fourier transform of the map and averages it in k-scale.

The randomly distributed bright sources cause a significant enhancement of the Poisson component (shot noise) in the observed power spectrum. Hence each experiment masks the bright sources by applying certain flux cuts, which depend on the frequency. The CIB power spectrum data we use were observed in various filters using different flux cuts. Thus in order to consistently compare our model with the observations we masked the bright sources in the same way. For simplicity, we excluded from the simulated catalogs all the sources with fluxes above the given flux cut and then generated the maps, instead of applying a mask on the already generated map. The discrepancy between the two methods in the case of a diffuse field like the CIB emission, is negligible (more details about the effect of masking in Van Cuyck et al. in prep.). For *Planck* the flux cuts are 225, 315, 350, and 710 mJy at 217, 353, 545, and 857 GHz, respectively (Planck Collaboration XXX 2014) while for *Herschel*/SPIRE we used two sets of simulated data, that follow the two different masking techniques used by the corresponding observational data sets. Viero et al. (2013) mask all the sources above 300 mJy in the analyzed band (method m1), while Viero et al. (2019) mask sources brighter than 300 mJy at 1200 GHz whatever the analyzed band (method m2).

Subsequently, in order to compare the model with the data, we had to apply the appropriate color corrections. We multiplied our model with a factor to convert from the theoretical SED flux to the measured flux that follows the  $\nu I_\nu = \text{constant}$  convention (Lagache et al. 2020, Appendix A). The color correction factors for all the bands are given in Table 2.

*Planck* used a planet model to calibrate the 545 GHz and 857 GHz channels. However, for 545 GHz the planet calibrations agreed with the CMB dipole calibrations within 1.5%. Given the relative calibration errors between the two filters, the absolute calibration for 857 GHz is given to be within 2.5% of the CMB calibration (Planck Collaboration Int. XLVIII 2016). The absolute calibration uncertainty of *Herschel* for 600 and 857 GHz is 8% (Viero et al. 2013). Therefore we consider *Planck*’s highest calibration accuracy as the reference. We thus correct the *Herschel*/SPIRE data by dividing with the 1.047 and 1.003 recalibration factors for 600 and 857 GHz, respectively, as given in Bertincourt et al. (2016). In the end, when comparing the model

<sup>8</sup> <https://cesamsi.lam.fr/instance/sides/home>

<sup>9</sup> Public code by Alexandre Beelen hosted at <https://zenodo.org/record/4507624#.YTiIfyZr3mE>

**Table 2.** Color corrections factors (cc) used to multiply the SIDES model before comparing with observations:  $C_{\ell, \nu, \nu'}^{\text{data}} = cc_{\nu} \times cc_{\nu'} \times C_{\ell, \nu, \nu'}^{\text{model}}$ .

Band	$cc_{\nu}$	Ref.
<i>Planck</i> 217 GHz	1.119	a
<i>Planck</i> 353 GHz	1.097	a
<i>Planck</i> 545 GHz	1.068	a
<i>Planck</i> 857 GHz	0.995	a
SPIRE 600 GHz	0.9739	b
SPIRE 857 GHz	0.9887	b
SPIRE 1200 GHz	0.9880	b
SPT 150 GHz	1.1411	b
SPT 220 GHz	1.0059	b

**References.** a: [Planck Collaboration XXX \(2014\)](#) b: [Lagache et al. \(2020\)](#).

with the *Herschel* data, we added on top of the SIDES model the *Planck* absolute calibration uncertainty mentioned above, for 545 and 857 GHz. There is no recalibration factor for the 1200 GHz band in the literature to correct the data. We have thus added to the SIDES model the 8% given uncertainty of *Herschel*.

### 3.3. The impact of the abundance matching proxy selection

We followed the abundance matching procedure as explained in Sect. 2.3, using both  $v_{\text{peak}}$  and  $M_{\text{peak}}$  as matching quantities. Subsequently, we created one *Planck* map for each quantity and computed their power spectra, respectively.

In Fig. 3 we show the SIDES models for  $v_{\text{peak}}$  and  $M_{\text{peak}}$ . In all the different filters the two models almost coincide with each other. They agree at better than 5%. This proves that the choice between these two quantities has a negligible effect, at least in the context of this work. Following the argument presented in [Reddick et al. \(2013\)](#) and taking into account the low impact of this choice, we chose the  $v_{\text{peak}}$  as proxy. All the results presented hereafter are obtained using this quantity.

### 3.4. Comparison with the data

The comparison between the *Planck* data and the SIDES model is shown in Fig. 3. The black dots represent the measurements of the CIB anisotropy power spectrum ([Planck Collaboration XXX 2014](#)) and the blue lines the model of the power spectrum resulting from the SIDES simulation for the corresponding frequency bands. We also included the absolute calibration uncertainty as a shaded area on top of the model line, but it is not visible since it is almost negligible (1.5% and 2.5% for 545 and 857 GHz, respectively). In each panel we show the data and model of the cross power spectra between different bands where we can see that there is an overall agreement. In the auto and cross power spectra that involve the 217 GHz band, the disagreement between the model and the CIB data is  $\leq 20\%$ . This is also the case for the rest of the auto and cross power spectra, except for the lowest  $\ell$  data points and the  $857 \times 857$  GHz case where the discrepancy is up to a factor of 2.

The comparison between the *Herschel*/SPIRE data ([Viero et al. 2013, 2019](#)) and the SIDES model is shown in Fig. 4. The model is able to adequately predict the power spectrum, except at 600 GHz where there is a systematic offset. The level of disagreement in the cases of  $600 \times 1200$  GHz,  $857 \times 1200$  GHz, and  $1200 \times 1200$  GHz is lower than 10% in both the large and small

scales. However, the discrepancy between the model and the data for  $600 \times 600$  GHz,  $600 \times 857$  GHz, and  $857 \times 857$  GHz is 35%, 25%, and 16%, respectively. From that we can conclude that the model at 1200 GHz can reproduce the data to high accuracy while the lower the frequency the larger the discrepancy.

Previous studies in the literature that measured and modeled the CIB anisotropies have spotted inconsistencies between the data and the fitted models as well as between the different experiments. [Lagache et al. \(2020\)](#) compare the shot noise power of the CIB anisotropies and the model-based predicted values both for *Planck* and *Herschel*. The expected behavior is to obtain a lower shot noise level when masking deeper. However, the *Herschel*/SPIRE measurements at the 600, 857, and 1200 GHz bands are incompatible with this expected behavior. The shot noise derived with a higher flux limit is lower than that derived with a lower flux limit, and the model used for the prediction is not systematically higher or lower than the measurements, preventing any robust conclusion. Moreover, [Maniyar et al. \(2021\)](#) compare the measurements of the CIB power spectra taken from *Planck* and *Herschel* as well as cross power spectra measurements between bands of the same experiment. The power spectra of *Planck* and *Herschel* agree well on the large angular scales but they are clearly different on small angular scales. Since SIDES agrees with *Planck* but not with *Herschel* at the same frequencies, it is possible that the discrepancy between SIDES and *Herschel* comes from measurement systematics.

We also investigated the agreement between our model and South Pole Telescope (SPT) power spectra as well as SPT  $\times$  SPIRE cross power spectra. The results are discussed in detail in Appendix A.

The overall agreement between the SIDES model and the observational data taken from both the *Planck* and *Herschel* instruments, especially on the large scales, shows that the SIDES simulation can realistically reproduce the clustering of galaxies as revealed from their continuum emission. Hence, it is a reliable tool to use in order to simulate the CONCERTO ([CONCERTO Collaboration 2020](#)) observations and develop the needed tools to post process and interpret the observational data.

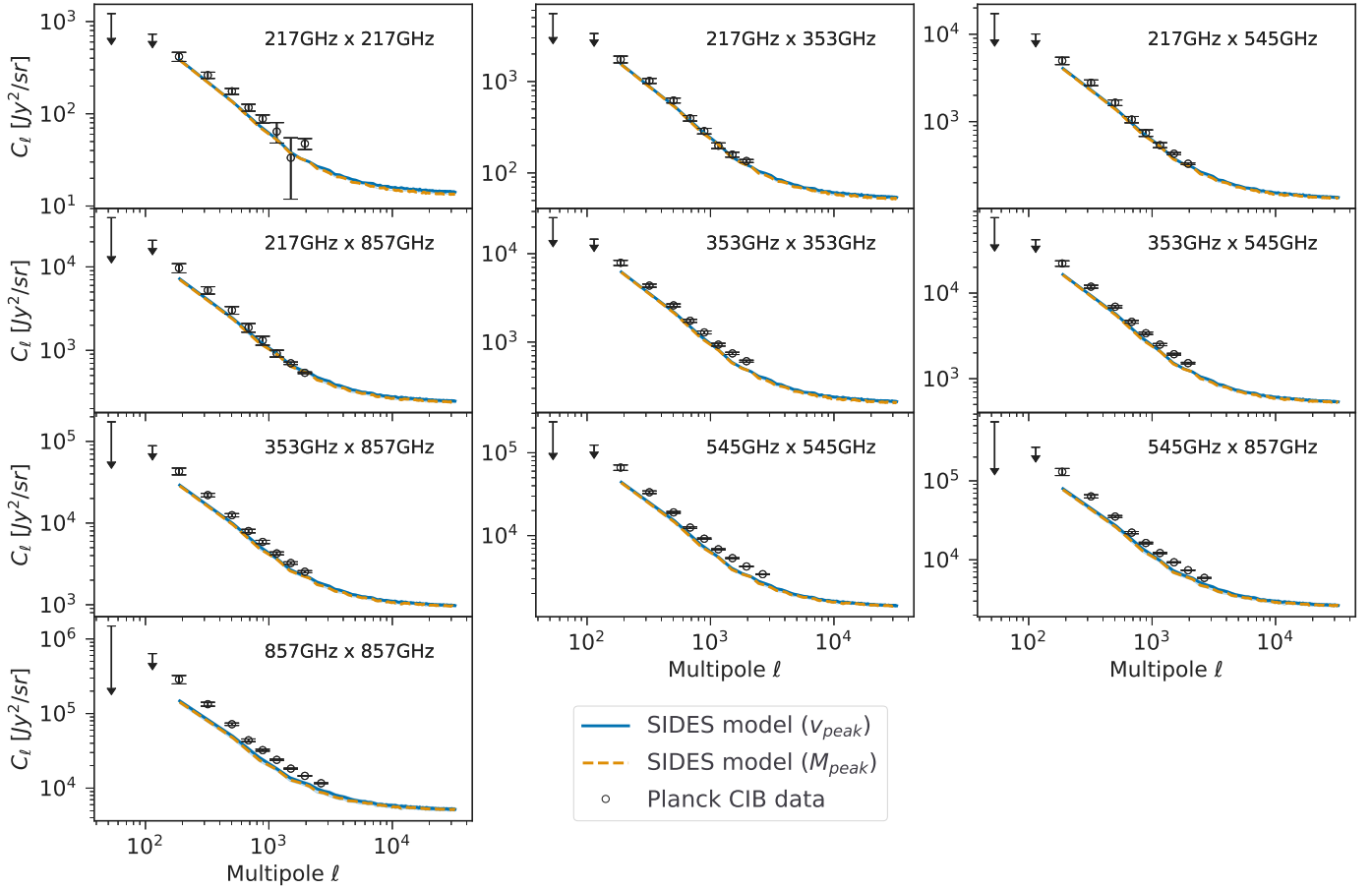
## 4. Line luminosity functions

The observed LF provides important constraints to test our simulation. In Sect. 4.1, we compare the LFs from our model with observational data. This allows us to validate the empirical relations embedded in SIDES-Uchuu to generate the emission of the different spectral lines. In Sect. 4.2 we investigate the field-to-field variance introduced in the LF and its dependence on survey size, while in Sect. 4.3 we study how this variance propagates through to the molecular gas density. Finally, in Sect. 4.5 we present a public tool for the computation of the LFs and the molecular gas variance based on our simulation.

### 4.1. Validation of the spectral line recipes introduced in SIDES

In order to validate the newly added recipes for the spectral lines in SIDES (B22) we compared the CO and [CII] LFs with observational data provided by recent surveys. This test has already been carried out in B22 for a single simulated field of  $2 \text{ deg}^2$ . In this work we extended the validation to our simulated field of  $117 \text{ deg}^2$  total size, thanks to the Uchuu cosmological simulation. This allowed us to test the validity of our simulation





**Fig. 3.** Comparison of SIDES and measured CIB power spectra for *Planck*. The solid blue lines represent the SIDES power spectra for the *Planck* bandpass, using  $v_{\text{peak}}$  as proxy for the abundance matching. The dashed orange line is the power spectrum obtained using the  $M_{\text{peak}}$  as proxy. The black points are the observational data from *Planck* (Planck Collaboration XXX 2014).

at a higher precision and test if the field-to-field variance could explain some discrepancies.

#### 4.1.1. CO luminosity functions

Millimeter interferometers provide rich constraints on various CO transitions. We used data from the ALMA spectroscopic survey in the *Hubble* Ultra Deep Field (ASPECS), which spans a  $4.6 \text{ arcmin}^2$  region and covers band 3 (84–115 GHz) and 6 (212–272 GHz) (Decarli et al. 2019, 2020). The band 3 and band 6 windows correspond to different redshift ranges for each CO line. This offers a wide variety of constraints on the various CO transitions and redshifts.

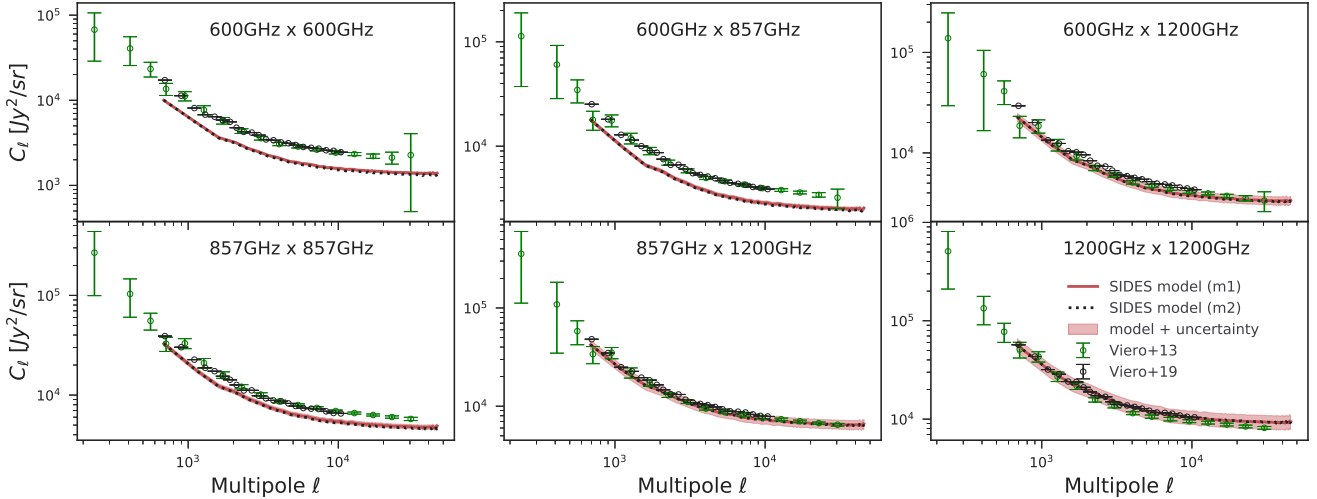
We cut the simulated catalogs into subfields of  $4.6 \text{ arcmin}^2$  which is the size of the ASPECS survey. We then computed individually for each catalog the number of galaxies per comoving volume using the same redshift ranges as ASPECS (Decarli et al. 2019, 2020). We compared all the resulting CO LFs with the ASPECS data and the LF of the total simulated field.

We show in Fig. 5 the LF constructed from the entire Uchuu field and the median of the multiple LFs from the  $4.6 \text{ arcmin}^{-1}$  Uchuu subfields together with the ASPECS data. The median LF sharply drops to zero faster than the LF of the entire Uchuu in all redshift cases. The bright sources are less common and thus the majority of the subfields have zero sources above a certain luminosity bin. The last nonzero luminosity bin is denoted with

an arrow in Fig. 5. However, this is not the case for the entire Uchuu field where there are more bright sources.

The LF from entire Uchuu field and the SIDES-Bolshoi LF agree very well at the faint-end and the knee. There are only some discrepancies ( $\lesssim 20\%$ ) at the bright-end of the LF at some redshift slices (e.g.,  $2.01 < z < 3.11$ ,  $0.27 < z < 0.63$ ,  $0.7 < z < 1.17$ ). We computed the field-to-field variance, constructing multiple LFs from  $2 \text{ deg}^2$  Uchuu subfields and found that the SIDES-Bolshoi LF (red line in Fig. 5) still lies inside the  $2\sigma$  range. We could thus conclude that the offset between the two LFs is just a statistical fluctuation.

Despite the significant uncertainty introduced in the LF by the small ASPECS survey size, the SIDES-Uchuu LFs are still within the  $1\sigma$  confidence level of the observations. However, at the higher CO transitions ( $J_{\text{upper}} \geq 6$ ) the model appears to be systematically lower as also found in B22 for SIDES-Bolshoi. Considering the large Uchuu area, we can rule out that this trend is caused by an underdensity of the SIDES-Bolshoi simulation. This shift at high- $J$  could be explained by the contamination of the ASPECS measurements by interlopers. This might be also related to the physics of the high- $J$  CO emission. Toward high- $z$  the interstellar medium of galaxies evolves, and in general it gets more turbulent and overall warmer. There are hints of a shift of the peak of the CO SLED toward higher  $J$  (Vallini et al. 2018). We take this effect into account by linking the CO(5–4)/CO(2–1) ratio to  $\langle U \rangle$ , as mentioned in Sect. 2.1. But this correlation is probably too weak to boost the high- $J$  CO emission. The shift



**Fig. 4.** Comparison of the SIDES power spectrum model with *Herschel*/SPIRE data. The solid red line corresponds to the SIDES model where we have followed the masking technique of Viero et al. (2013), noted as m1. On top of the SIDES models we illustrate the corresponding systematic uncertainty (i.e., absolute calibration) as described in Sect. 3.2 (red shaded area). The dotted black line is the SIDES model following the masking technique of Viero et al. (2019), noted as m2. The green points are the corresponding observational data presented in Viero et al. (2013) while in black the data from Viero et al. (2019).

could also be due to the correlation between the luminosity bins, since ASPECS uses a sliding binning. Therefore, in the case of an over or underdensity the LF value at all the bins would be higher or lower, respectively. We also find that this effect also stands for nonoverlapping bins (see Sect. 4.4). However, the fact that the systematic shift of the model is always toward lower values is intriguing and makes it difficult to conclude.

#### 4.1.2. [CII] luminosity functions

Measuring the [CII] LF is more challenging compared to CO. The [CII] sources observed in the submm/mm atmospheric window reside at higher redshifts, therefore their counterparts can be very faint and difficult to observe. It is thus difficult to identify [CII], when a single line is detected. On top of that the number density of [CII] sources is rather small and larger fields are required. However, the ALPINE survey (Le Fèvre et al. 2020; Béthermin et al. 2020; Faisst et al. 2020) managed to target with ALMA the [CII] lines in  $118 < z < 5.9$  normal star forming galaxies and put further constraints on the [CII] LF (Loiacono et al. 2021; Yan et al. 2020).

For the computation of the [CII] LFs we followed the same steps as in CO. However, in this case we could not use the field size of the survey because ALPINE targeted selected galaxies from both the cosmic evolution survey (COSMOS) field and the *Chandra* deep field south (CDFS). This ALPINE sample is biased toward UV selected galaxies with well-defined spectroscopic redshifts (see Faisst et al. 2020, for a discussion on the biases). Most of the ALPINE galaxies (89%) are in the COSMOS field and they dominate the statistics. We thus chose to use the COSMOS size in our estimate. We computed the [CII] LFs at  $4.4 < z < 4.6$  and  $5.3 < z < 5.9$ , which are the redshift ranges observed by ALPINE.

We show in Fig. 6 a comparison of the SIDES-derived [CII] LFs with the ALPINE data at  $z \sim 4.5$  and  $z \sim 5.5$ . In both cases we have also included the resulting LF of the SIDES-Bolshoi (B22) for comparison. Similarly to the CO LFs, the median LF drops to zero due to the lower number of bright sources included in the  $2 \text{ deg}^2$ -sized subfields compared to the total Uchuu field.

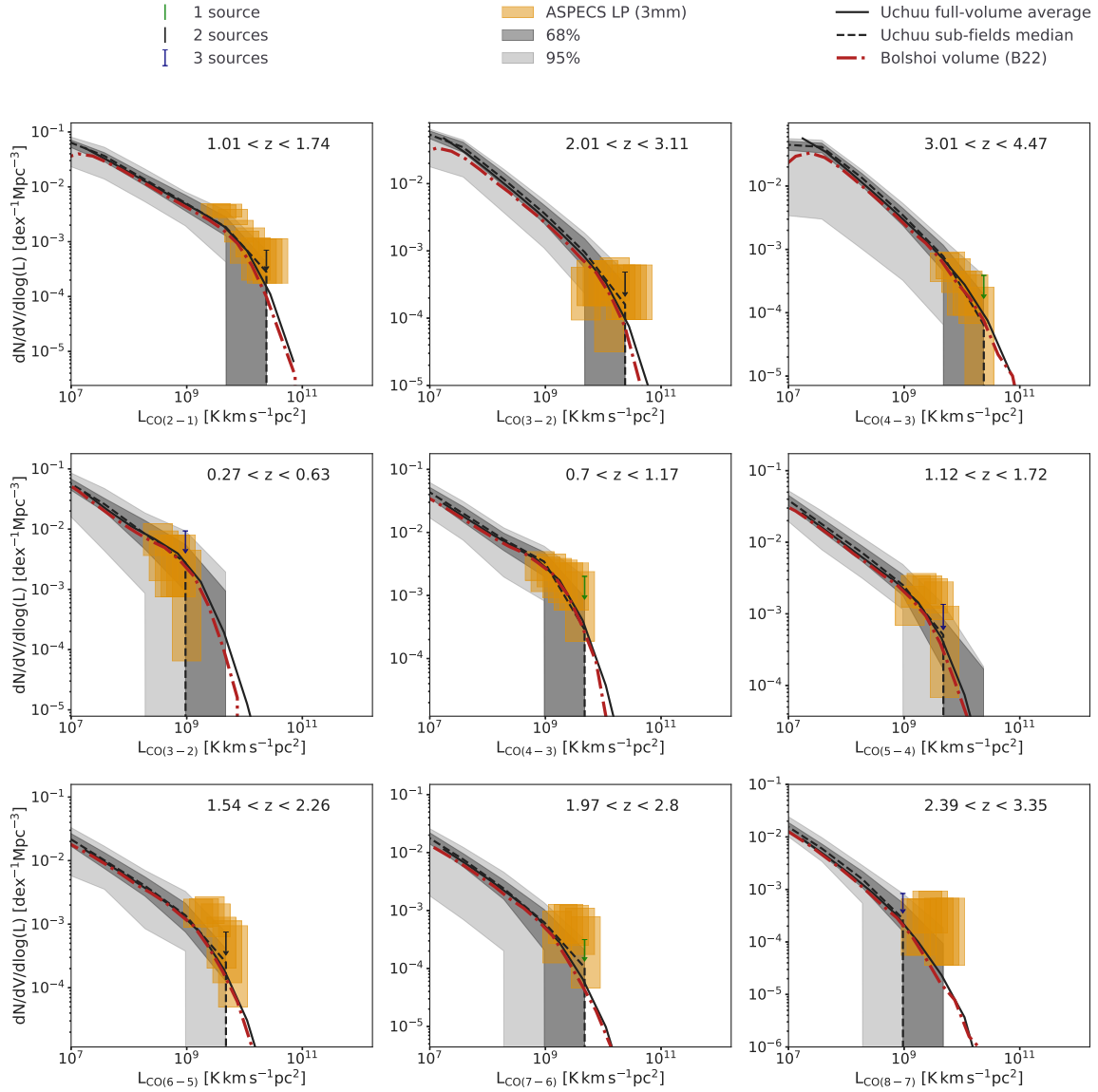
The total Uchuu LF and Bolshoi LF perfectly agree with each other up to  $\sim 10^{10} L_{\odot}$  for both redshift slices, while at the bright end the agreement remains better than  $2 \sigma$  of the field-to-field variance (shaded area). Similarly to the CO LFs and because of the much larger volume, the Uchuu LF goes to higher luminosities while Bolshoi is too small to contain such bright sources.

Similarly to the SIDES-Bolshoi LFs, the SIDES-Uchuu LFs agree at  $\sim 1 \sigma$  with the ALPINE measurements at  $z \sim 4.5$ , except for the highest luminosity point. It is most probable that this is a statistical fluctuation since this point is derived from just two ALPINE objects (Yan et al. 2020). At  $z \sim 5.5$  there is an overall agreement between  $10^8$  and  $10^9 L_{\odot}$ , while the highest and the lowest luminosity points are lower than the simulation by  $2 \sigma$ . In the case of the faintest point, it could be explained by the incompleteness of detection of the ALPINE survey at lower luminosities, as discussed in Yan et al. (2020). The highest luminosity point may be affected by small number statistics at the bright end or a bias of the ALPINE sample against the most dusty and potentially [CII]-luminous galaxies. Finally, our simulation agrees at  $\sim 1 \sigma$  with the Loiacono et al. (2021) measurement, which was taken from blind ALPINE detections and even though it is less precise (a few objects), it is less sensitive to assumptions or systematic effects.

The Yan et al. (2020) LF measurements serve as a lower limit because they are obtained using a UV-selected sample. The survey targets only sources bright enough in the UV, possibly excluding bright [CII] sources that are faint in the UV. On the other hand, the LF measurements of Loiacono et al. (2021) are obtained using the serendipitous sources of the ALPINE survey, that are strongly affected by the clustering. Their data point is therefore considered an upper limit. Our simulation is between the lower and upper limits offered by these measurements.

#### 4.2. Variance of line luminosity functions

Comparing the LFs of Figs. 5 (gray areas) and 6 (blue and green area) we see, as expected, that smaller survey sizes lead to more uncertain LFs. It is common practice to estimate the errors of the LF using only Poisson statistics. However, clustering can lead to



**Fig. 5.** Comparison of the CO LF resulting from the SIDES simulation with ASPECS observational data. Each LF is created using sources from 117 different 4.6 arcmin<sup>2</sup>-sized simulated subfields. This field size is chosen to match the size of ASPECS. The gray shaded areas with different transparencies correspond to the 16th–84th and 5th–95th percentile confidence intervals. The black solid line is the LF of the entire volume of the Uchuu simulation, and the dashed line is the median of the multiple LFs computed from all the subfields. The red line shows the resulting LF presented in B22 where the Bolshoi-Planck cosmological simulation was used, while the arrows show the last luminosity bin of the SIDES-Uchuu LFs that contains at least one source. The different colors stand for a different number of sources.

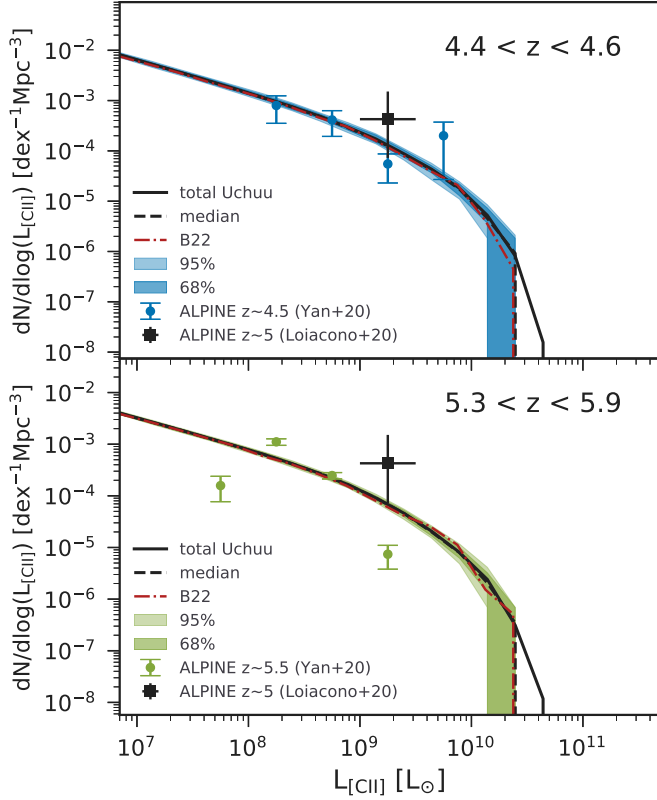
much larger uncertainties. Thanks to the large volume offered by the Uchuu cosmological simulation we can properly study the contribution of the clustering in the field-to-field variance.

We cut the total exploitable Uchuu field (117 deg<sup>2</sup>) in multiple smaller subfields of the requested size. We constructed 12 sets of LFs from 12 different subfield sizes, respectively. The selected sizes are 0.0003, 0.0013, 0.0069, 0.0312, 0.125, 0.25, 0.5, 1, 2, 4, 8, and 16 deg<sup>2</sup>. For all the fields of size below 1 deg<sup>2</sup> we could create a large number of subfields given the total Uchuu field. However, in order to save computational time we only used 117 out of the total subfields, by randomly selecting a smaller region out of each 1 deg<sup>2</sup> subfield. For the fields with sizes of 1, 2, 4, 8, and 16 deg<sup>2</sup> we created 117, 54, 24, 12, and 6 subfields. Above 8 deg<sup>2</sup> we had a low number of fields and thus poorer statistics. Furthermore, the subfields with size  $\geq 1$  deg<sup>2</sup> were contiguous, and thus not fully independent which could

lead to an underestimation of the variance because of the large-scale modes. Finally, for each subfield size we computed the mean value ( $\mu_{LF}$ ) and standard deviation ( $\sigma_{LF}$ ) of each luminosity bin. We used the mean and not the median value because, as discussed above, the mean value does not drop to zero as fast as the median. We could thus define the relative variance ( $\sigma_{LF}/\mu_{LF}$ ) for a wider luminosity range.

The observed field-to-field variance in the LFs can be modeled as the combination of two components. The first one is the Poisson component, which is caused by the fluctuations of the number of sources in a given volume. The LF is the number of sources ( $N$ ) per luminosity bin per unit of comoving volume. The mean value and standard deviation of the LFs from various subfields will thus be proportional to the Poisson mean and standard deviation, respectively, that is,  $\mu_{LF} \propto N$  and  $\sigma_{LF} \propto \sqrt{N}$ . By dividing these two quantities, the luminosity bin size and





**Fig. 6.** Comparison of the SIDES [CII] LFs with ALPINE observations. *Top panel:* ALPINE data at  $z \sim 4.5$ , *bottom panel:* ALPINE data at  $z \sim 5.5$ . In both panels, the black solid line corresponds to the total Uchuu field LF and the dashed line corresponds to the median LF of 54  $2 \text{ deg}^2$  SIDES-Uchuu subfields. The shaded areas are the 16th–84th and 5th–95th percentile intervals while the red line is the SIDES-Bolshoi LF (B22). The blue and green points are the data from Yan et al. (2020), while the black square is the data point from Loiacono et al. (2021).

comoving volume cancel out, and so  $\sigma_{\text{LF}}/\mu_{\text{LF}} = 1/\sqrt{N}$ . The larger the volume the higher the number of sources leading to lower relative variance. The second component is caused by the clustering, which practically describes the excess of probability to find a source next to another compared to the Poisson distribution. The clustering is linked to the distribution of the dark matter halos since galaxies are formed inside the halos. The total relative LF variance, which is the quadratic sum of the two, and can be modeled by

$$\frac{\sigma_{\text{LF}}}{\mu_{\text{LF}}} = \sqrt{y + \frac{1}{N}}, \quad (1)$$

where  $y$  is a term depending on the geometry of the field and the angular correlation function. It can be computed using (Blake & Wall 2002; Béthermin et al. 2010):

$$y = \frac{\int_{\text{field}} \int_{\text{field}} w(\theta) d\Omega_1 d\Omega_2}{\Omega^2}, \quad (2)$$

where  $w(\theta)$  is the angular two-point auto correlation function of the sources and  $\Omega$  the angular size of the field. The  $y$  term and thus the contribution of clustering to the relative uncertainties is proportional to the amplitude of the correlation. Also, it depends on the size of the field for a given shape. Indeed,  $y$  is the average value of the correlation function for two points

**Table 3.** Best-fit parameters of Eq. (5).

$L_{\text{bin}}$ ( $\text{K km s}^{-1} \text{ pc}^2$ )	$A'$	$\gamma$
$10^7$	0.003	1.67
$2.8 \times 10^7$	0.004	1.71
$7.7 \times 10^7$	0.005	1.67
$2.2 \times 10^8$	0.005	1.77
$6 \times 10^8$	0.006	1.71
$1.7 \times 10^9$	0.006	1.82
$4.6 \times 10^9$	0.01	1.68
$1.3 \times 10^{10}$	0.009	2.01
$3.6 \times 10^{10}$	0.01	1.28

selected randomly in the field. Since usually  $w(\theta)$  decreases with increasing  $\theta$ ,  $y$  will be smaller for wider fields. Both the relative Poisson ( $1/N$ ) and clustering ( $y$ ) uncertainties decrease with increasing field size but in a different way. There is thus a competition for the most dominant component at each field size. In order to investigate and compare these different trends, we modeled the variation of the relative uncertainties only as a function of the field size  $\Omega$ . The number of sources is linked to the survey size through the number density of galaxies  $\rho$ :

$$N = \rho\Omega. \quad (3)$$

The dependence of  $y$  on the survey size is complex but in the case of a power law  $w(\theta) = A\theta^{1-\gamma}$ , Blake & Wall (2002) demonstrate that:

$$y \propto \Omega^{(1-\gamma)/2}. \quad (4)$$

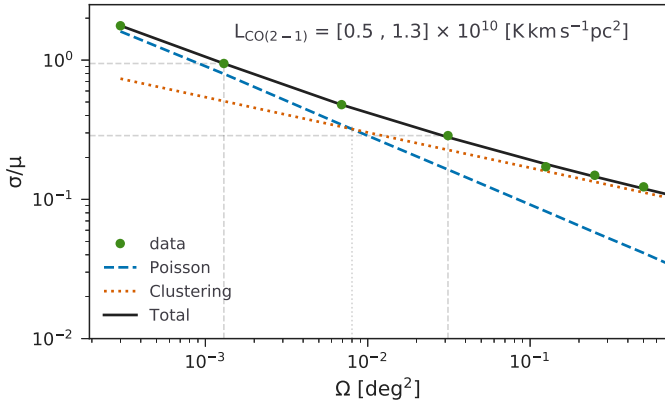
The factor  $A$  depends on the flux and wavelength (Béthermin et al. 2010). It is hence different for each luminosity and CO transition at a given redshift for simplicity. We thus considered it a constant for each luminosity and CO transition. Therefore, combining Eqs. (1) and (2), we get:

$$\frac{\sigma_{\text{LF}}}{\mu_{\text{LF}}} = \sqrt{A'\Omega^{(1-\gamma)/2} + (\rho\Omega)^{-1}}, \quad (5)$$

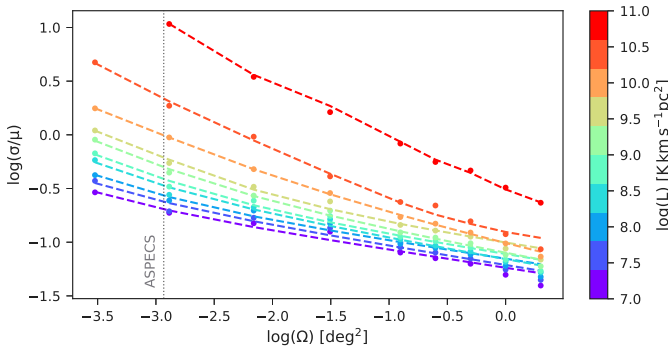
where  $A'$  is another constant derived from the integral in Eq. (2) and proportional to  $A$ . The resulting best-fit  $A'$  and  $\gamma$  values are summarized in Table 3.

The Poisson term  $(\rho\Omega)^{-1}$  was directly computed from the number of sources in the simulated catalog (Eq. (3)), where  $\rho$  was directly computed from our catalog. For the clustering term,  $A'\Omega^{(1-\gamma)/2}$ , we fit to our simulated data the parameterized function of Eq. (5) to get the parameters  $A'$  and  $\gamma$  that define the contribution of the clustering.

We show as an example in Fig. 7 the total variance of the CO(2–1) LF for the luminosity bin  $L_{\text{CO}(2-1)} = [0.5, 1.3] \times 10^{10} [\text{K km m}^{-1} \text{ pc}^2]$  as a function of survey size at  $z = 1.2\text{--}1.6$ . As expected, the total relative uncertainty increases when the survey size decreases, by a factor of 3 from  $0.0312 \text{ deg}^2$  ( $112 \text{ arcmin}^2$ ) to  $0.0013 \text{ deg}^2$  ( $4.6 \text{ arcmin}^2$ ), while we would have expected a factor of 5 for pure Poisson behavior. We also show the decomposition of the total variance in the Poisson and the clustering component. The larger the survey size the more significant the contribution of the clustering component becomes. Equality is reached for a field size of  $0.008 \text{ deg}^2$  ( $28.8 \text{ arcmin}^2$ ).



**Fig. 7.** Total relative variance and its decomposition into the Poisson and clustering components for the luminosity bin  $L_{\text{CO}(2-1)} = [0.5, 3.6] \times 10^{10} \text{ [K km s}^{-1} \text{ pc}^2]$ .



**Fig. 8.** Dependence of the total relative variance of the CO(2-1) LF on the survey size. In different colors we show the different luminosity bins of 0.4 dex size. The points are computed from the SIDES simulations and the dashed lines correspond to the model described by Eq. (5).

Beyond this size, the clustering of the galaxies is responsible for larger uncertainties than what is expected from the Poisson law.

The relative contribution of the two components varies depending on the choice of the luminosity bin, since it affects both the clustering and the source density. This effect is shown in Appendix B (Fig. B.1), where we show an equivalent of Fig. 7 for various luminosity bins. In Fig. 8 we show the total relative uncertainty as computed directly from the simulation and modeled using Eq. (5) as a function of the survey size, for different luminosity bins which are color coded accordingly.

We can finally remark that the relative uncertainty of the low luminosity bins decreases slower with survey size than in the high ones. Even though the constraints in these bins are better because of the larger number of sources, the precision of the constraints is less tightly dependant on the field size than the bright luminosity bins. Therefore, in order to measure the faint-end of the LF it would be more efficient to use multiple independent small fields rather than one large field.

#### 4.3. Variance of the molecular gas mass density

One of the main factors regulating the evolution of galaxies is their star formation and its evolution with time. As the molecular gas constitutes the fuel for the star formation, measuring the abundance of the molecular gas through cosmic time is receiving considerable interest (e.g., Bisigello et al. 2022). The molecular

gas density ( $\rho_{\text{H}_2}$ ) is usually computed using the first moment of the observed LF ( $L'_{\text{CO}(1-0)}$ ) after assuming a conversion factor ( $\alpha_{\text{CO}}$ ),

$$\rho_{\text{H}_2} = \alpha_{\text{CO}} \times \int_0^\infty L'_{\text{CO}} \Phi(L'_{\text{CO}}) d \log(L'_{\text{CO}}). \quad (6)$$

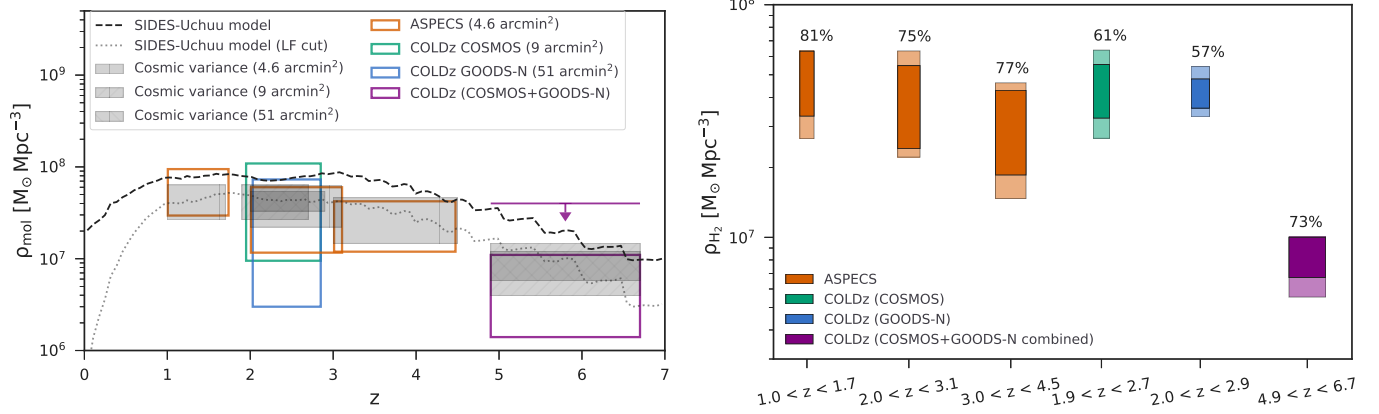
The significant intrinsic variance in the LFs (Sect. 4.2) is thus propagated to the molecular gas mass density estimation. Usually, a fixed value is assumed for the  $\alpha_{\text{CO}}$  conversion factor and the conversion from the luminosity of a given transition to  $L'_{\text{CO}(1-0)}$ . We use  $\alpha_{\text{CO}} = 3.6 M_\odot$ . In this paper, we ignore the impact of the conversion factor on the uncertainties and we focus instead on the impact of the field-to-field variance on the precision of the molecular gas density.

We first show in Fig. 9 on the left the evolution of  $\rho_{\text{H}_2}$  with redshift as computed with SIDES when integrating both the full range of the LFs and only down to a luminosity cut. We additionally compare with measurements from ASPECS (Decarli et al. 2020) and COLDz (Riechers et al. 2019). Both surveys use a luminosity cut at  $\sim 5 \times 10^9 \text{ K km s}^{-1} \text{ pc}^2$ . The SIDES model without the luminosity cut is systematically higher than the measurements. In contrast, the model that includes the luminosity cut agrees very well with the observations. Integrating the LFs down to  $\sim 5 \times 10^9 \text{ K km s}^{-1} \text{ pc}^2$  leads to an underestimation of the molecular gas density. For instance, at  $z = 0.5, 2$  and  $5$  the discrepancy factor is 8.2, 1.9, and 1.7, respectively.

We note that adopting smaller ( $\leq 117 \text{ deg}^2$ ) Uchuu fields introduces significant uncertainties. We estimated this variance for several redshift ranges by cutting the simulated area into multiple subfields with size equal to the corresponding survey that we compared our simulation with. However, for the combined COLDz field (COSMOS and GOODS-N, purple square in Fig. 9) on the left we selected a  $9 \text{ arcmin}^2$ -sized SIDES subfield (COLDz COSMOS size) and a  $51 \text{ arcmin}^2$ -sized subfield (COLDz GOODS-N size) without them overlapping. We created 117 realizations and computed the LF for each one of them. We obtained the molecular gas density value from the CO LF of each subfield using Eq. (6). We then computed the 5th and 95th percentile confidence levels of  $\rho_{\text{H}_2}$  at various redshifts, which are presented in Fig. 9 on the left as gray shaded rectangular areas. Overall, there is an excellent agreement between the interval predicted by SIDES for various realizations of the cosmic variance and the observational data.

All the CO surveys account for the field-to-field variance using the Poissonian uncertainties. However, they neglect the additional variance due to the clustering of the galaxies. In our study we differentiated between the two variance components. We thus show in Fig. 9 on the right the total field-to-field variance with the fractional contribution of the Poisson variance indicated for each redshift slice. The difference between the total and Poisson variances reveals how the clustering acts as an extra source of variance. It is significant ( $>20\%$ ) especially at  $z > 2$  and it should be taken into account by observational surveys.

The variance strongly depends on the survey size, as discussed in Sect. 4.2. Therefore, we would like to address the question of what is the minimum survey size that will allow us to distinguish the evolution of the molecular gas from its intrinsic variance. For that purpose we cut the total simulated catalog in smaller ones of different sizes, varying from  $0.00028 \text{ deg}^2$  ( $1 \text{ arcmin}^2$ ) to  $9 \text{ deg}^2$  and computed the molecular gas mass density (Eq. (6)) at a fixed redshift bin. We performed this analysis for various redshift bins centered at 1.4, 2.6, 3.8, 4.9, and 6 (taking  $\delta z = 0.4$ ). The results are presented in Fig. 10, where we also compare our work with the results from Keenan et al. (2020).



**Fig. 9.** Field-to-field variance of the molecular gas mass density,  $\rho_{\text{H}_2}$ , at several redshift slices where observational data are available. *Left:* evolution of  $\rho_{\text{H}_2}$  with redshift. The empty boxes are the observational data from ASPECS (Decarli et al. 2020) and COLDz (Riechers et al. 2019). We also include the upper limit offered by COLDz assuming that all the galaxies could be CO(2–1) emitters at high  $z$ . The black dashed line is the SIDES model without a luminosity cut and the black dotted line is the SIDES model with a luminosity cut at  $\sim 5 \times 10^9 \text{ K km s}^{-1} \text{ pc}^2$ . The steps in both lines are caused by the selected redshift grid that the cosmological simulation snapshots were taken. The shaded gray areas correspond to the  $\rho_{\text{H}_2}$  field-to-field variance which has been computed for the same redshift ranges and sizes as the corresponding surveys. *Right:* field-to-field variance in the molecular gas density estimates, while the dark-colored areas show the Poisson-only variance. The contribution of the latter to the total variance is given as a percentage on top of each rectangle. The different colors indicate the different surveys, hence different size and redshift slice.

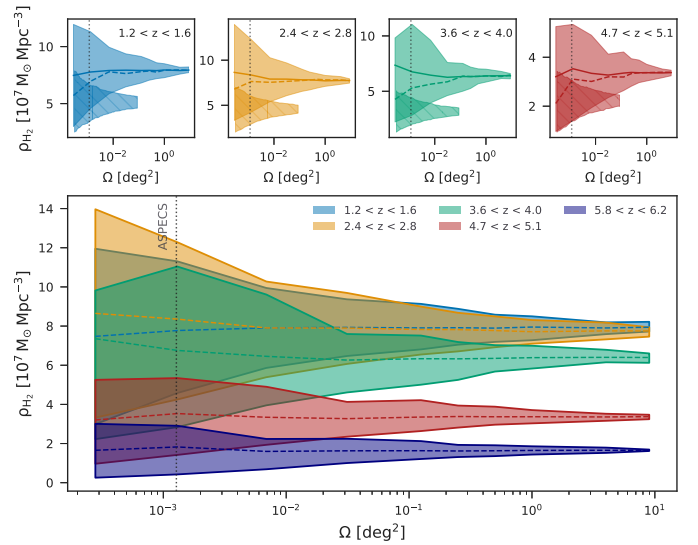
There is a shift between the mean  $\rho_{\text{H}_2}$  of our model and the one from Keenan et al. (2020). This could be explained by their use of a cut off when computing the integral of the first moment of the LF ( $L'_{\text{CO}} = 10^9 \text{ K km s}^{-1} \text{ pc}^2$ ) to obtain the molecular gas density. On the contrary, we keep the whole luminosity range given in our simulated catalogs, leading to systematically higher  $\rho_{\text{H}_2}$  values. This difference can already introduce a discrepancy of the level of  $\sim 20\%$ . Furthermore, the two approaches adopt different cosmological simulations, scaling relations to assign CO luminosities to galaxies, and level of scatter on the scaling relations. In spite of this divergence there is an agreement between the two models at the level of  $1\sigma$  for survey sizes below  $\sim 40 \text{ arcmin}^2$  ( $\sim 10^{-1} \text{ deg}^2$ ).

The bottom panel of Fig. 10 shows how the constraints in the various redshift bins improve with the field size. We need at least  $1 \text{ deg}^2$  surveys to avoid having an overlap between the  $2.4 < z < 2.8$  and the  $3.6 < z < 4.0$   $1\sigma$  confidence regions and thus start to have hints of an evolution. This is about an order of magnitude larger field than what Keenan et al. (2020) suggested, indicating that they may have underestimated the level of the total variance at these redshift ranges. In order to be able to probe the evolution between  $3.6 < z < 4$  and  $4.7 < z < 5.1$ , we need at least  $2 \times 10^{-2} \text{ deg}^2$  surveys. This value agrees with Keenan et al. (2020).

#### 4.4. Correlation among the luminosity function bins

The independence of the luminosity bins is a common assumption for the fit of the LF. We used the Uchuu simulation to test its validity. In Fig. 11 we show 10 randomly drawn LFs for three different survey sizes. In the case of the smaller survey size, we can see that the values of the low luminosity bins do not affect the brighter bins. They are rather randomly varying and the luminosity bins can be considered independent. This behavior is different for larger survey sizes and this is more clear for the  $1 \text{ deg}^2$  case, especially for luminosity bins below  $10^{10} \text{ K km s}^{-1} \text{ pc}^2$ . When one luminosity bin is high all the other bins tend to be high as well and vice versa. This reveals a high level of correlation between the different luminosity bins.

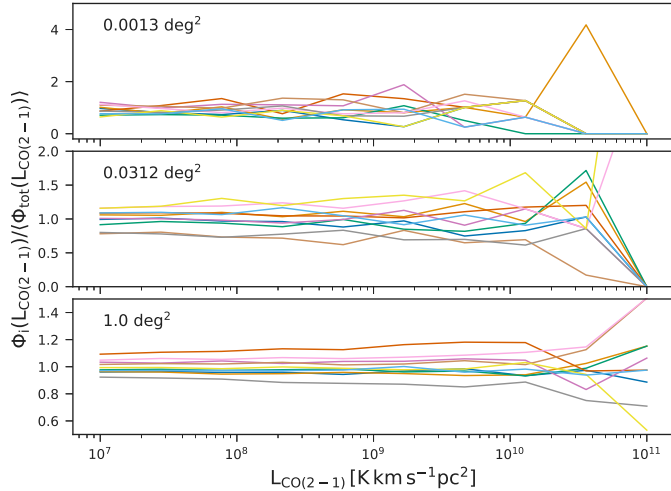
In order to better visualize the correlation level we computed the Pearson correlation matrix between luminosity bins



**Fig. 10.** *Top:* cosmic variance in the  $\rho_{\text{H}_2}$  values as a function of the survey size for different redshift slices. The solid and dashed lines are the mean and median values, respectively. We have also included the corresponding results from Keenan et al. (2020) (the shaded and hatched areas). *Bottom:* evolution with redshift of the variance in the molecular gas mass density as a function of survey sizes, where it is easier to spot what is the proper survey size that will allow to probe the evolution of the molecular gas density.

and show it as a 2D map in Fig. 12. The larger the survey size the more correlated the luminosity bins. On top of that the faint luminosity bins are more correlated compared to the bright bins, at all survey sizes. For a given size, the faint sources are more abundant compared to the very bright ones, the number of faint sources will thus be high if there is an overdensity or low if there is an underdensity. Therefore, the amplitude of the LF at all the adjacent low luminosity bins is expected to similarly vary from field to field, this is why they appear more correlated. However, the bright sources are rare and their number density





**Fig. 11.** Ten randomly drawn LFs for three different survey sizes, of  $0.0013 \text{ deg}^2$  ( $4.6 \text{ arcmin}^2$ ),  $0.0312 \text{ deg}^2$  ( $112 \text{ arcmin}^2$ ), and  $1 \text{ deg}^2$  from top to bottom, respectively. Each LF is normalized by dividing by the total Uchuu LF.

could differently vary from one field to another resulting to less correlated bins.

The confirmation of this expected effect is important for all the surveys that use the observed LFs to obtain other physical quantities like the molecular gas density. Bigger survey sizes can offer smaller error bars but the luminosity bins are more correlated with each other. Therefore, one should take into account that the luminosity bins are not independent when fitting the LF.

#### 4.5. A tool to estimate the variance of line luminosity functions and molecular gas densities

To help the user investigate the SIDES-Uchuu catalogs<sup>10</sup>, we offer several simple scripts gathered in a jupyter notebook to derive the most useful quantities. A first set of scripts computes the LF of the requested transition in a specific redshift slice and for a given survey size. The second part of the output is the expected field-to-field variance for the same survey parameters. This part of the script assumes that the fields are squares. Finally, the code can compute the Pearson correlation matrix between luminosity bins.

The second set of scripts compute the field-to-field variance on the mass density of the molecular gas ( $\rho_{\text{H}_2}$ ). The input parameters are similar to the previous set of scripts and it returns  $\rho_{\text{H}_2}$  as well as the 5%, 16%, 84%, and 95% confidence levels. The different options available for the user are documented in detail in the notebook.

## 5. Power spectra variance

The total signal of an intensity map can be decomposed into fluctuations at different scales. The power spectrum of an intensity map describes how much the fluctuations of each scale contributes to the final map. By studying the CO and [CII] power spectra we can obtain important physical information such as, the SFRD at high redshifts and the clustering of galaxies. The power spectrum consists of the clustering and the shot noise component. The former describes the correlation of the sources

at large scales and the latter is the flat component caused by the randomly distributed sources at all scales.

The power spectra obtained by line intensity mapping experiments can significantly vary from one field to another. This could be caused either by several bright sources in the field of the survey, or by the selection of the field itself. Observing an over or underdensity of sources could affect the level of the power spectrum at all scales. This effect has not been studied before, but thanks to the Uchuu simulation we could study and quantify the intrinsic field-to-field variance of such power spectra.

Similarly to our study of the LF variance (Sect. 4.2), we generated cubes corresponding to 117 different  $1 \text{ deg}^2$  subfields, and computed the power spectrum for each. In Fig. 13 we show the resulting power spectra for both the CO and [CII] lines, as well as the corresponding subfields that were used to obtain them. The level of the power spectra can vary by a factor of 2–10 depending on the line, the observing frequency and the scale. Therefore, trying to fit a highly uncertain power spectrum could strongly affect the inference of physical quantities like the clustering of the galaxies and the SFRD.

### 5.1. Contribution to the shot noise

The shot noise component of the power spectrum is defined as

$$P_{\text{shot}} = \int_0^{S_o} S^2 \frac{dN}{dS} dS = \int_0^{S_o} S^3 \frac{dN}{dS} d \ln(S), \quad (7)$$

where  $S_o$  is the flux above which the sources are masked or removed in a given survey. In order to investigate which sources contribute the most to the Poisson noise, we focused on the integrand in Eq. (7). In Fig. 14 we show the quantity  $S^3 \frac{dN}{dS} d \ln(S)$  as a function of flux. We also show in the same plot, as reference, the level of the shot noise as computed from Eq. (7).

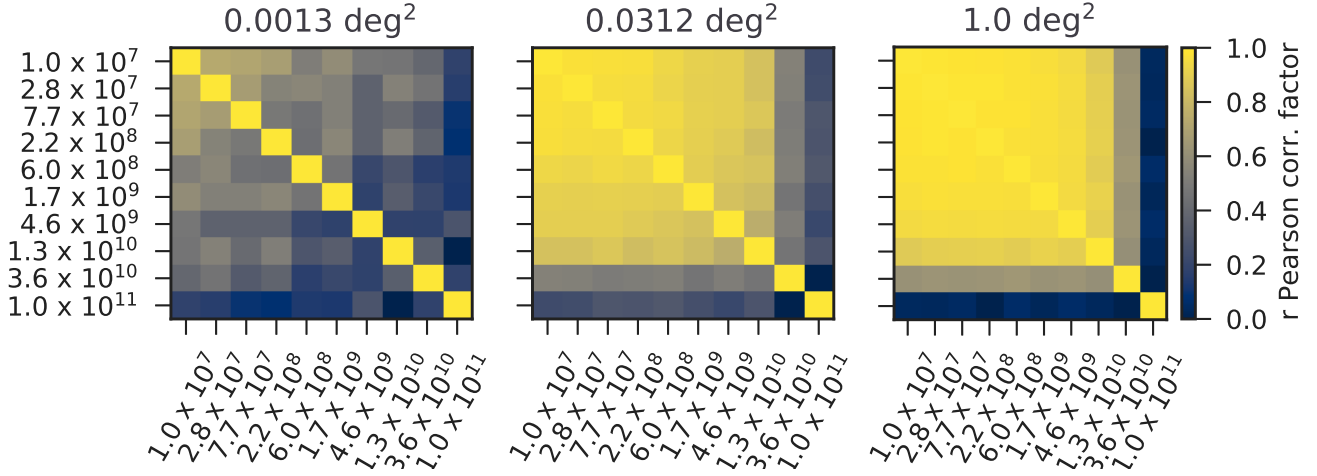
We see from Fig. 14 that for both CO and [CII] the brightest sources are the ones that contribute the most to the power of the shot noise and this trend is independent of the observing frequency. We also see that all the curves converge to the level of the shot noise as computed from the integral, which is consistent with the fact that all the shot noise is produced by the highest  $d \ln S$  bins. The behavior shown in Fig. 14 is the opposite with respect to the CIB case, where the faint sources are the most contributing ones (Lagache et al. 1999). This means for CO and [CII] that even if we apply flux cuts (to remove the brightest sources) the most contributing sources to the shot noise level will always be the ones that exist right below the flux cut.

It will be challenging for the first-generation intensity mapping surveys to detect individual bright sources. But even if they detect some, since the shot noise level is very sensitive to the flux cut, applying an incomplete masking could lead to an incorrect interpretation of the shot noise. Finally, the interpretation of the shot noise will offer information about the bright sources just below the detection threshold. LIM experiments will offer statistical information about these sources, but classical surveys are more suitable for studying them (e.g., Yue & Ferrara 2019).

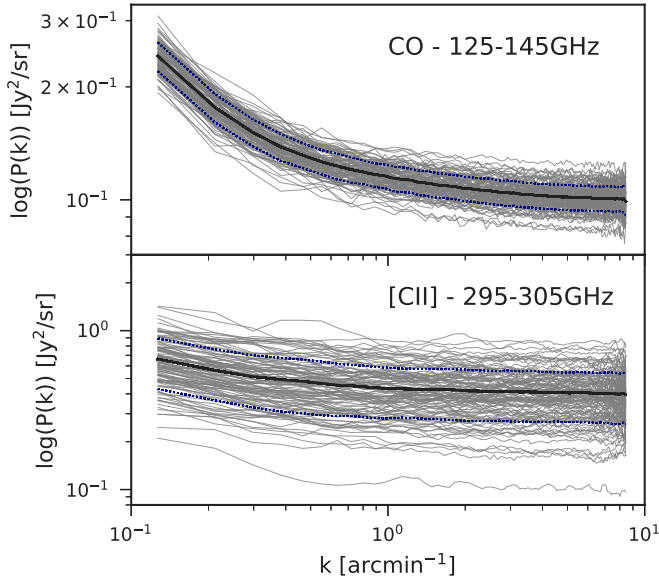
### 5.2. Power spectrum variance model

The variance of the power spectra (Fig. 13) depends on the observed line, the survey size ( $\Omega$ ), the observing frequency ( $\nu$ ), and the number of frequency channels ( $\Delta\nu$ ) averaged to obtain the final power spectrum of each subfield. For instance, CONCERTO observes from 130 to 310 GHz with a spectral

<sup>10</sup> <https://cesamsi.lam.fr/instance/sides/home>



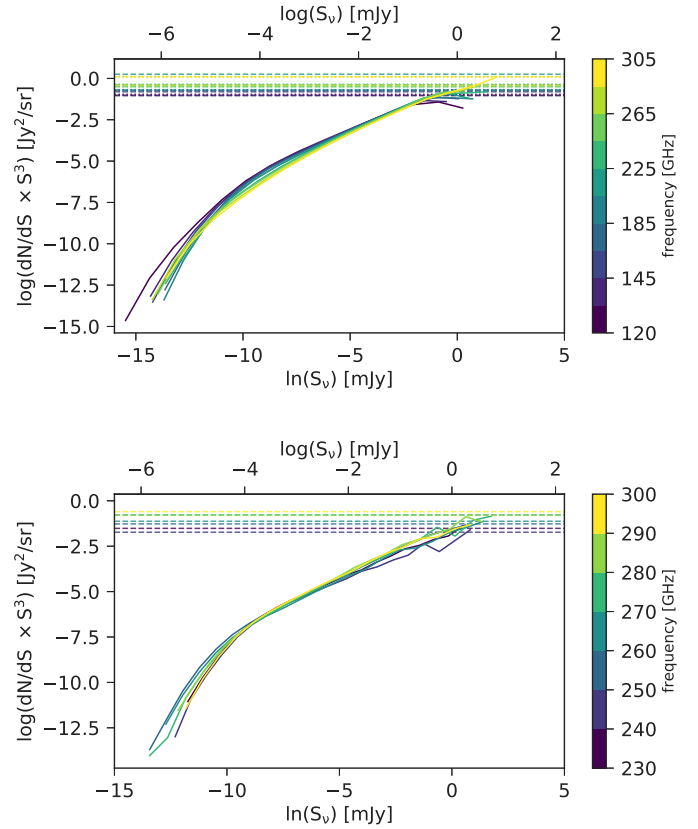
**Fig. 12.** Pearson correlation matrix among the luminosity bins used to construct the LFs. Each panel corresponds to different sizes, from left to right:  $0.0013 \text{ deg}^2$  ( $4.6 \text{ arcmin}^2$ ),  $0.0312 \text{ deg}^2$ , and  $1.0 \text{ deg}^2$ . The units of the luminosity bins are  $\log(L/[K^{-1} \text{ km s}^{-1} \text{ pc}^2])$ .



**Fig. 13.** Power spectra of 117  $1 \text{ deg}^2$ -sized Uchuu subfields for the CO and [CII] lines, averaged over the frequency range shown in each plot. *Top:* CO power spectra. *Bottom:* [CII] power spectra. In both panels, the solid black line is the mean power spectrum at all  $k$ , while the dotted blue lines denote the  $1\sigma$  range.

resolution of up to  $1.3 \text{ GHz}$ , so if we average the power spectra of the first 15 channels, meaning  $\Delta\nu = 19.5 \text{ GHz}$ , the resulting power spectrum will be at the center of  $130\text{--}149.5 \text{ GHz}$ .

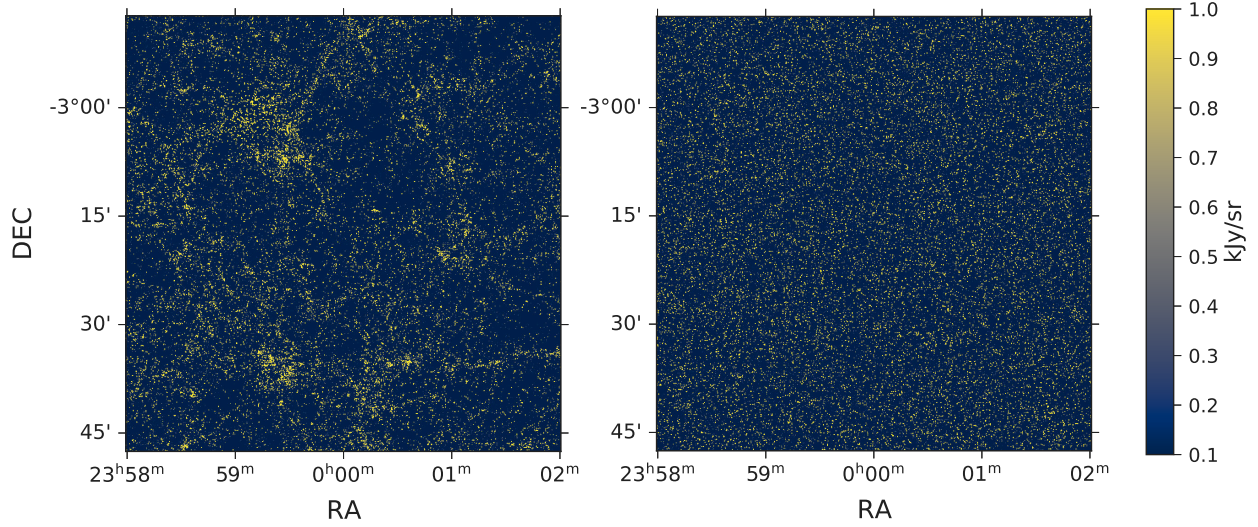
We aimed to measure the variance as a function of these various parameters. However, the variance on the power spectra could be very different for the shot-noise and the clustering components. We thus needed to compute each component separately in each subfield. We created a new set of cubes with the same size and sources as the original ones but with their (RA, DEC) coordinates randomly shuffled, erasing in this way any clustering information embedded in the original cubes (defined as shuffled cubes hereafter). The followed procedure is visualized in Fig. 15. This allowed us to measure the shot-noise level. We then subtracted this quantity from the total power spectrum to obtain the clustering component and then computed the variance for each  $k$ . We found that the variance is not scale dependent. We thus aver-



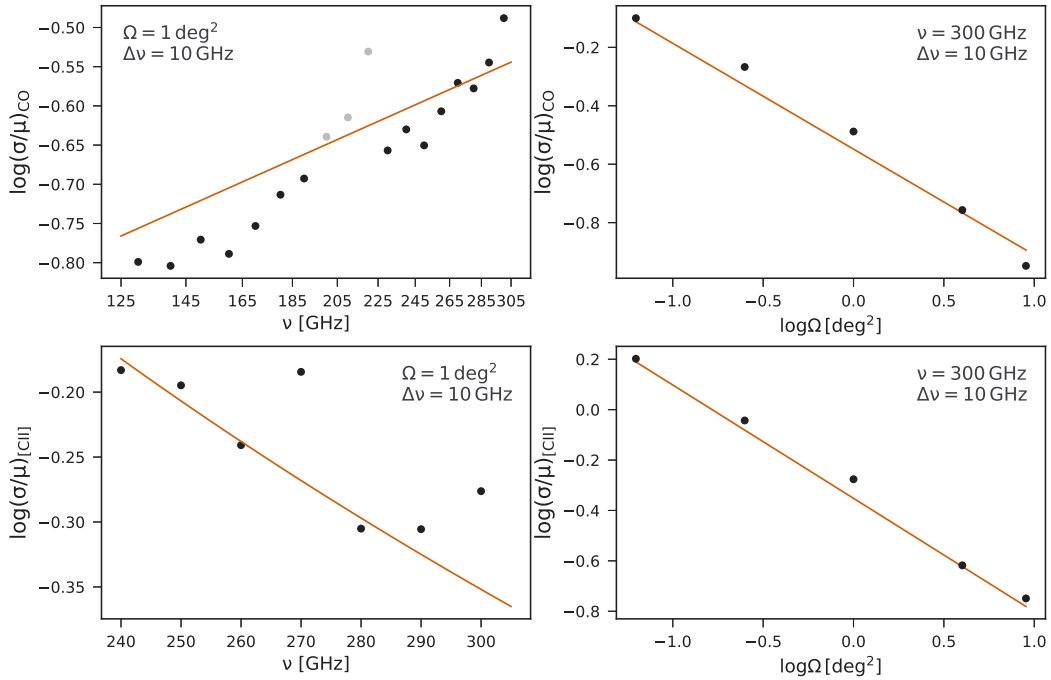
**Fig. 14.** Source flux contribution to the CO (*top*) and [CII] (*bottom*) shot noise power spectrum (see Eq. (7)). The different color lines correspond to different frequency slices with the CONCERTO frequency resolution ( $1.5 \text{ GHz}$ ) within the whole CONCERTO observing frequency range. The dashed lines are the expected shot noise level as computed by Eq. (7).

aged the variance at the various scales to obtain the final variance of the clustering component.

We subsequently aimed to obtain a model that would enclose all the information of field-to-field variance dependence. We first created a set of simulated data that our model would be able to fit. For this purpose, we created five data sets corresponding to



**Fig. 15.** Visualization of the 275 GHz slice of an original CO cube on the left and the shuffled cube *on the right*. The shuffled cube contains the same sources (same redshift and luminosity) as the one *on the left* but is randomly distributed.



**Fig. 16.** Dependence of the clustering power spectrum variance ( $0.13 < k < 1 \text{ arcmin}^{-1}$ ) on the observing frequency ( $\nu$ ) and the survey size ( $\Omega$ ). *Top row*: the black points are the CO data while the high transparency points are the ones excluded due to the CO(2–1) contamination as explained in Sect. 5.2. The solid orange line is the best-fit model. *Bottom row*: the black points are the [CII] data and the orange solid line is the best-fit model.

five different field sizes ( $0.0625 \text{ deg}^2$ ,  $0.25 \text{ deg}^2$ ,  $1 \text{ deg}^2$ ,  $4 \text{ deg}^2$ , and  $9 \text{ deg}^2$ ). In each case we cut the simulation in subfields of the corresponding size and computed the power spectra for the whole range of frequencies (125–305 GHz) and frequency step ( $\Delta\nu = 5, 10, 20 \text{ GHz}$ ). A subset of the simulated data are shown as black points in Fig. 16. We finally fitted all the simulated data for a given line and a given component (shot-noise or clustering) by the following parametric form:

$$\frac{\sigma}{\mu} = c \left( \frac{\nu}{\nu_o} \right)^\alpha \left( \frac{\Delta\nu}{\Delta\nu_o} \right)^\beta \left( \frac{\Omega}{\Omega_o} \right)^\gamma, \quad (8)$$

where  $c, \alpha, \beta$ , and  $\gamma$  are parameters determining the scaling, while we set the normalization to  $\nu_o = 100 \text{ GHz}$ ,  $\Delta\nu_o = 5 \text{ GHz}$ , and  $\Omega_o = 1 \text{ deg}^2$ .

We observe an excess of variance at around 230 GHz for CO. This is caused by low-redshift bright sources seen in CO(2–1) as discussed in B22. This excess of the CO power spectra in this frequency range complicates the process of searching for a universal model that could fit all the given data. We therefore chose to exclude the CO data points between 210 and 230 GHz (light-colored points in Fig. 16).

We provide two sets of best-fit parameters for each line: one for the clustering component and one for the shot noise



**Table 4.** Best-fit parameters of Eq. (8) for the CO and [CII] lines.

	CO		[CII]	
	Poisson	Clustering	Poisson	Clustering
$c$	0.1	0.2	5.43	6.22
$\alpha$	0.71	0.58	-2.1	-2.4
$\beta$	-0.52	-0.39	-0.62	-0.61
$\gamma$	-0.31	-0.37	-0.45	-0.46

component. All the parameters are summarized in Table 4, while in Fig. 16 we show the best-fit model and how the relative uncertainty ( $\frac{\sigma}{\mu}$ ) depends on the survey size and frequency. We can thus use it to estimate the expected relative variance of the observed power spectra.

The  $\alpha$  slope is positive for CO and negative for [CII] revealing the opposite dependence on frequency. In the case of CO the higher the observing frequency the higher the relative variance, while in the case of the [CII] line the trend is the exact opposite. The [CII] trend is not surprising, since the lower frequency probes very high redshifts, where the number of objects are very small and star formation tends to appear in the most overdense regions (e.g., Behroozi et al. 2013a; Béthermin et al. 2013). The CO intensity fluctuations are the result of multiple rotational lines and are more difficult to interpret.

The  $\beta$  slope for the CO Poisson component is  $-0.52$ . It is very close to the  $-0.5$  expected if the shot noise in the various frequency slices was independent (B22, Appendix D). The values obtained for [CII] and for the CO clustered component are slightly different, which suggests that there may be significant correlations between the power spectra of the various frequency slices.

The  $\gamma$  parameter for both components of CO and [CII] is systematically higher than the expected value in a pure Poisson case ( $-0.5$ ). The variance thus decreases with the field size slower than what is expected in the absence of any clustering. This is similar to what has been found for the LFs (Sect. 4.2). It is not surprising for the clustering component, but is less intuitive for the Poisson one. However, these results make sense if we consider that the variance of the LFs, from which the Poisson component derives, exceeds the Poisson term.

### 5.3. Correlation between the Poisson and the clustering components

The dependence of the variance of the clustering power spectrum with the field size is expected to deviate from the expected behavior of a purely Poisson case. We see a similar trend for the variance of the Poisson component. This could mean that both components are similarly dependent on the local over or under-density. Of course, we do not expect a perfect correlation, since the shot noise is mainly caused by the bright sources while the sources around the knee of the LFs are those that maximally contribute to the clustered component (e.g., B22). However, as we find in Sect. 4.4, the various luminosity bins are correlated and we can thus expect a significant effect. In order to quantify it, we measured the level of correlation between the clustering and shot noise component among the various Uchuu subfields.

In Fig. 17 we show how the amplitude of the clustering and the shot noise are correlated. When the clustering is high in a given subfield, the shot noise level tends to be high. The Pearson cor-

relation coefficient varies between 0.41 and 0.67. As a result, the total variance of the power spectrum is neither the quadratic sum of the two components nor their simple sum. Therefore one should be cautious when combining these two sources of uncertainty.

### 5.4. Consequences for line intensity mapping experiments

Using our model we forecast the expected uncertainties on  $P(k)_{\text{clust}}$  associated with field-to-field variance for CONCERTO and other current or upcoming experiments. TIME will cover a rectangular  $0.01 \text{ deg}^2$ -sized field and will observe in the 183–326 GHz frequency range. The South Pole Telescope Summer-time Line Intensity Mapper (SPT-SLIM, Karkare et al. 2022) survey will cover  $1 \text{ deg}^2$  at 120–180 GHz. Finally, the CO Mapping Array Project (COMAP, Cleary et al. 2022) and the Prime-Cam on FYST (CCAT-Prime Collaboration 2023) will survey  $4 \text{ deg}^2$  patches at 26–34 GHz and 210–420 GHz, respectively. The properties of the LIM experiments we consider are summarized in Table 5.

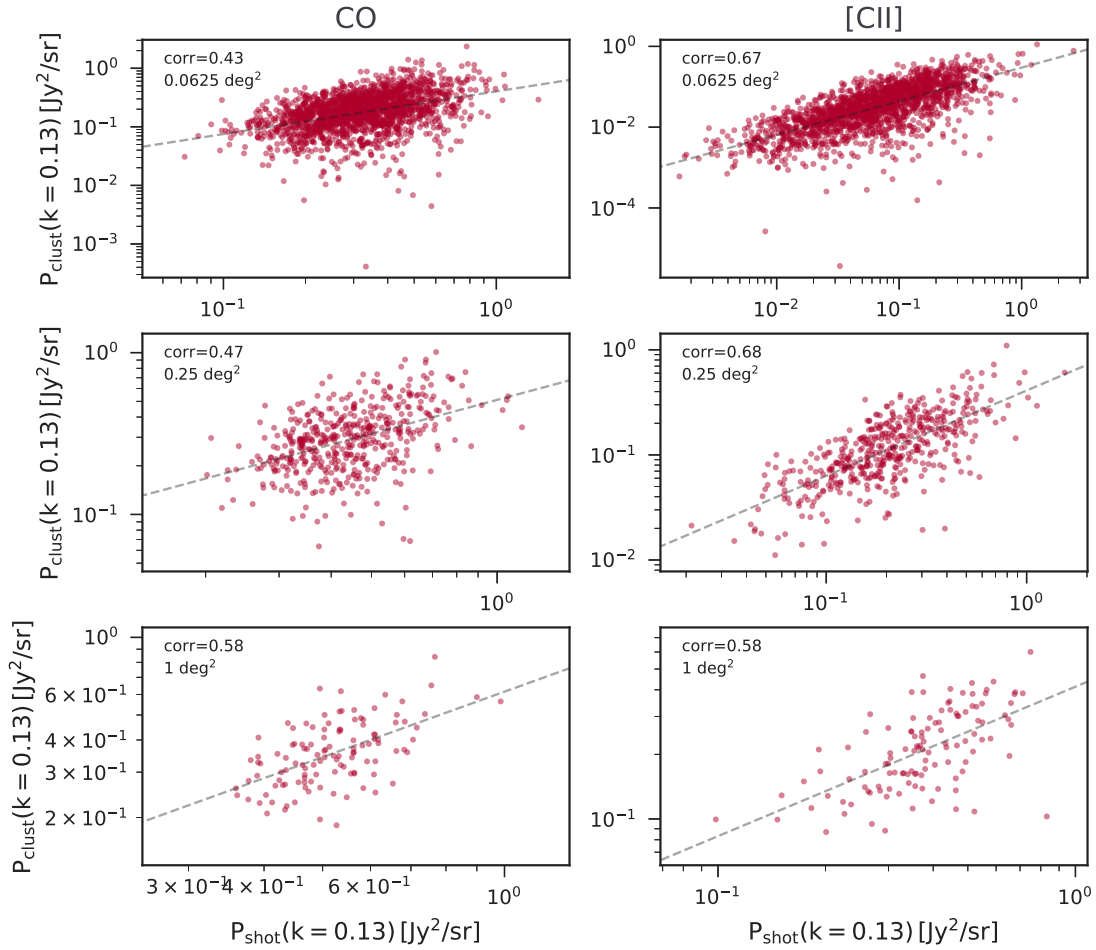
Given the field size and the frequency range of each experiment we obtained the relative uncertainty of both the CO and [CII] power spectra caused by the field-to-field variance. The results are shown in Fig. 18. We used different line styles for the different number of averaged channels ( $\Delta\nu$ ). Averaging more frequency channels results in lower uncertainties. For instance, the  $\Delta\nu = 20 \text{ GHz}$  curves are below the  $\Delta\nu = 5 \text{ GHz}$  ones.

For frequencies below 280 GHz ( $z \gtrsim 6$  for [CII]) CONCERTO will not robustly constrain the [CII] clustering power spectrum since the field-to-field uncertainties exceed 50%. However, at higher frequencies ( $z \lesssim 6$ ) these uncertainties are smaller and it should be possible to provide meaningful constraints if the signal is detected. On the other hand, CONCERTO will be able to get a good estimate of the CO clustering power spectrum below  $\sim 250 \text{ GHz}$  where the field-to-field uncertainties are smaller than 20%. However, the interpretation of the shot noise for both [CII] and CO will be difficult due to the high field-to-field variance, the correlation with the clustering component, and because it is mainly produced by rare bright sources (Sect. 5.1).

Regarding our forecast for the other [CII] LIM experiments, for TIME it is quite pessimistic with an uncertainty higher than 100% at all frequencies. However, our model was designed for square field areas which is very different than the TIME strategy ( $1.3 \times 0.007 \text{ deg}^2$ ). It is thus possible that we have overestimated its field-to-field uncertainty. The wide field area of the FYST/Prime-Cam experiment could offer the most promising results with a field-to-field uncertainty below 20% even up to  $z = 6$ . Even though at high redshift the uncertainties increase, this still remains the most accurate amongst the experiments. Finally, SPT-LIM will provide good constraints on CO ( $\sim 20\%$  field-to-field variance), which is the main goal of the experiment. However, because of its low frequency coverage, the [CII] power spectrum will be highly uncertain ( $>100\%$ ). But [CII] is not a science goal of this experiment.

Regarding the CO power spectrum, the COMAP (Pathfinder and EoR) experiment will be negligibly affected by the field-to-field variance with field-to-field uncertainties well below 20% for its whole frequency range. However, these results come from an extrapolation of our model, which could include some extra uncertainties we have not considered. SPT-LIM and FYST/Prime-Cam will also recover the CO power spectra with small uncertainty ( $\lesssim 20\%$ ).

According to the power spectrum formalism, the clustering component is proportional to the square of both the



**Fig. 17.** Correlation between the power spectra of the clustering and the shot noise from various subfields for the frequency range 295–305 GHz. The left column is for the CO power spectra for sizes from top to the bottom: 0.0625, 0.25, and 1 deg<sup>2</sup>. The right column is for the [CII] power spectra for the same sizes as the CO. On the top left of each panel we show the correlation value, while the dashed line shows the corresponding linear line.

total emission of the galaxies and their linear clustering bias (Kovetz et al. 2017). Therefore the uncertainty we have obtained here should be reduced by a factor of  $\sim 2$  when propagated to the integrated line emission and subsequently to the SFRD or the molecular gas density history. However, this assumes that we will be able to break efficiently the degeneracies between emission and clustering. Considering the high uncertainty when measuring the molecular gas density evolution from current surveys (left Fig. 9), the current generation of LIM experiments will still offer more accurate estimates. Future experiments with wider fields and higher sensitivities will offer even more accurate constraints.

## 6. Conclusion

Combining the SIDES model with the Uchuu dark matter simulation, which offers a large simulated volume and high mass resolution at the same time, allows us to test the SIDES model with a much higher accuracy than in B22. We compared the SIDES model with the power spectra of the CIB fluctuations in various *Planck* and *Herschel*/SPIRE bands. The overall good agreement confirmed the ability of SIDES to realistically reproduce the clustering of galaxies at large scales. As an additional validation test, we compared the line LFs predicted by SIDES with observational data from surveys like ASPECS and ALPINE. All

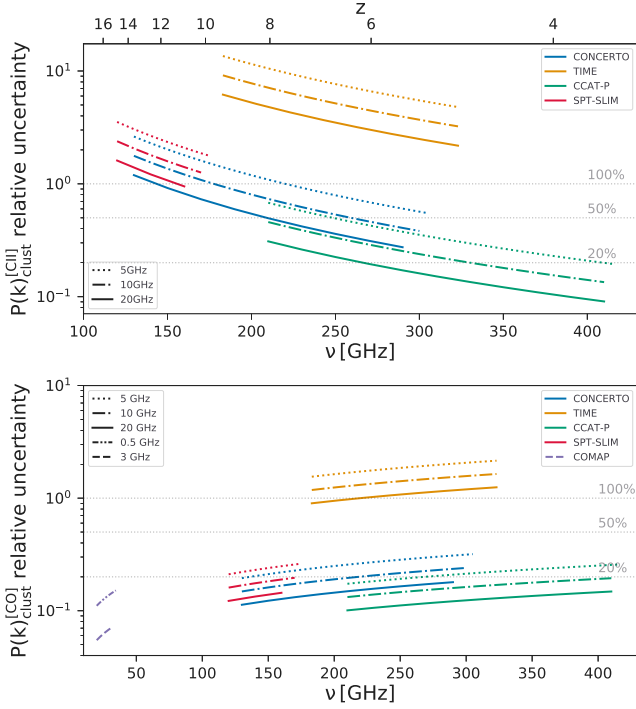
the SIDES LFs from multiple different subfields lay inside the  $1\sigma$  confidence level of the observations confirming the validity of the newly added recipes for the lines (CO and [CII], B22).

Regarding the line LFs we find that for low luminosity bins and large survey sizes, the field-to-field variance is larger than the Poisson uncertainty. This excess is caused by the galaxy clustering. We modeled the contribution of the Poisson and the clustering variance to determine precisely for each field size the luminosity range where the clustering is comparable to the Poisson term and needs to be taken into account. Finally, we show that the luminosity bins tend to be correlated with each other, especially for fields larger than 0.0312 deg<sup>2</sup>, which are dominated by the clustering rather than the Poisson variance.

We also studied the effect of the field-to-field variance on the cosmic molecular gas density measurements as a consequence of the CO LF uncertainties. For small survey sizes (e.g., ASPECS) the field-to-field uncertainty is large ( $\sim 50\%$ ). Comparing our findings with other similar studies, we find that the total variance at certain sizes (e.g., 1 deg<sup>2</sup>) could be up to an order of magnitude higher. We show that in order to differentiate between the  $\rho_{\text{H}_2}$  variation due to the field-to-field variance and the actual evolution at  $z < 4$  we need survey sizes on the order of 1 deg<sup>2</sup>, while at  $z > 4$  the survey size should be above  $2 \times 10^{-2}$  deg<sup>2</sup>. We see this steep change because the  $\rho_{\text{H}_2}$  evolution at  $z \gtrsim 4$  is faster than at  $z \lesssim 4$ . However, in order for spectral scans to reduce

**Table 5.** Characteristics of current and future line intensity mapping experiments.

	TIME	SPT-LIM	CONCERTO	COMAP	CCAT-P
Field size [deg <sup>2</sup> ]	0.0091	1	1.4	4	4
Targeted line(s)	[CII], CO	CO	[CII], CO	CO(1–0), CO(2–1)	[CII]
Frequency range [GHz]	183–326	120–180	130–310	26–34	210–420

**Fig. 18.** Level of relative uncertainty of the clustering component (the standard deviation of  $P(k)_{\text{clust}}$  divided by the average) of the [CII] power spectra (*top panel*) and the CO (*bottom panel*) for various LIM experiments (see Table 5). The different line styles correspond to different number of averaged channels ( $\Delta\nu$ ).

the field-to-field variance, it is not only the covered area on the sky that matters but also the observing frequency range (redshift range) they use, since the goal is to succeed as large comoving volumes as possible.

Both the clustering and the shot noise components of the LIM power spectra significantly vary from one field to another. This can strongly affect the inference of physical quantities, like the line LF and subsequently the SFRD. Moreover, we show that the brightest nonmasked sources, which are usually just below the detection limit, are the ones that contribute the most to the shot noise level. This means that: 1) the shot noise level is sensitive to the flux cut one chooses when masking the brightest sources, 2) the study of the shot noise level will target the sub-detection threshold sources, which can be more efficiently studied by classical surveys.

Additionally, we show that there is a significant level of correlation between the clustering and the shot noise power, which increases with the field size and is higher for the [CII] than CO line. This could affect the interpretation of the clustering power leading to over or underestimate of the physical quantities deriving from it. Additionally, because of this correlation, there is no analytical formula that describes how the clustering and Poisson variance should be combined to get the total variance.

The field-to-field variance of the power spectra (both clustering and shot noise) depends on the survey size, the observed frequency as well as on the frequency slice used to compute the 2D power spectra. We provided two models, one for CO and one for [CII] (separately for the clustering and the shot noise part), that estimate the level of the expected variance given the above mentioned quantities. This is a useful tool for CONCERTO and any other intensity mapping experiment that aims to optimize its configuration.

By using our variance estimation model we forecast how the field-to-field variance will limit the precision of the constraints provided by some current and near-future LIM experiments. The significant field-to-field variance of both the clustering and shot-noise components of the power spectra along with the correlation between them provides a quite pessimistic forecast for CONCERTO. At low frequencies the [CII] power spectrum exceeds the level of 50% uncertainty, while for higher frequencies it is slightly smaller but still well above 20%. On the other hand, COMAP and FYST/Prime-Cam, designed to cover a bigger field (4 deg<sup>2</sup>), seem to be more promising for the CO and [CII] power spectra accurate modeling.

The Uchuu extension of the SIDES code and its products are publicly available<sup>11</sup>. We also offer some complementary tools to help the community explore the products of our simulation and adapt them to their needs.

**Acknowledgements.** This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 788212) and from the Excellence Initiative of Aix-Marseille University-A\*Midex, a French “Investissements d’Avenir” programme. The Uchuu simulations were carried out on Aterui II supercomputer at Center for Computational Astrophysics, CfCA, of National Astronomical Observatory of Japan, and the K computer at the RIKEN Advanced Institute for Computational Science. The Uchuu data effort has made use of the skun@IAA\_RedIRIS and skun6@IAA computer facilities managed by the IAA-CSIC in Spain (MICINN EU-Feder grant EQC2018-004366-P). T.I. has been supported by IAAR Research Support Program in Chiba University Japan, MEXT/JSPS KAKENHI (Grant Number JP19KK0344 and JP21H01122), MEXT as “Program for Promoting Researches on the Supercomputer Fugaku” (JPMXP1020200109), and JICFuS. AF, AP, LV acknowledge support from the ERC Advanced Grant INTERSTELLAR H2020/740120 (PI: Ferrara). MA acknowledges support from FONDECYT grant 1211951, ‘ANID + PCI + INSTITUTO MAX PLANCK DE ASTRONOMIA MPG 190030’, ‘ANID + PCI + REDES 190194’, and ANID BASAL project FB210003. SAC acknowledges funding from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, PIP-2876), Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación (Agencia I+D+i, PICT-2018-3743), and Universidad Nacional de La Plata (G11-150), Argentina. GM acknowledges support by Swedish Research Council grant 2020-04691.

## References

- Amblard, A., Cooray, A., Serra, P., et al. 2011, *Nature*, **470**, 510
- 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., & Wu, H.-Y. 2013b, *ApJ*, **762**, 109
- Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., et al. 2013c, *ApJ*, **763**, 18

<sup>11</sup> <https://cesamsi.lam.fr/instance/sides/home>

- Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, *MNRAS*, **488**, 3143
- Behroozi, P., Conroy, C., Wechsler, R. H., et al. 2020, *MNRAS*, **499**, 5702
- Bertincourt, B., Lagache, G., Martin, P. G., et al. 2016, *A&A*, **588**, A107
- Béthermin, M., Dole, H., Beelen, A., & Aussel, H. 2010, *A&A*, **512**, A78
- Béthermin, M., Daddi, E., Magdis, G., et al. 2012, *ApJ*, **757**, L23
- Béthermin, M., Wang, L., Doré, O., et al. 2013, *A&A*, **557**, A66
- Béthermin, M., Daddi, E., Magdis, G., et al. 2015, *A&A*, **573**, A113
- Béthermin, M., Wu, H.-Y., Lagache, G., et al. 2017, *A&A*, **607**, A89
- Béthermin, M., Fudamoto, Y., Ginolfi, M., et al. 2020, *A&A*, **643**, A2
- Béthermin, M., Gkogkou, A., Van Cuyck, M., et al. 2022, *A&A*, **667**, A156
- Bisigello, L., Vallini, L., Gruppioni, C., et al. 2022, *A&A*, **666**, A193
- Blake, C., & Wall, J. 2002, *MNRAS*, **337**, 993
- Bournaud, F., Daddi, E., Weiss, A., et al. 2015, *A&A*, **575**, A56
- Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, *ApJ*, **803**, 34
- Capak, P. L., Carilli, C., Jones, G., et al. 2015, *Nature*, **522**, 455
- Carilli, C., & Walter, F. 2013, *ARA&A*, **51**, 105
- Carlson, J., & White, M. 2010, *ApJS*, **190**, 311
- Carniani, S., Ferrara, A., Maiolino, R., et al. 2020, *MNRAS*, **499**, 5136
- CCAT-Prime Collaboration, (Aravena, M., et al.) 2023, *ApJS*, **264**, 7
- Chabrier, G. 2003, *PASP*, **115**, 763
- Cheng, Y.-T., Chang, S.-C., & Bock, J. J. 2020, *ApJ*, **901**, 142
- Cleary, K. A., Borowska, J., Breyse, P. C., et al. 2022, *ApJ*, **933**, 182
- Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, *ApJ*, **647**, 201
- CONCERTO Collaboration (Ade, P., et al.) 2020, *A&A*, **642**, A60
- Cooray, A., Amblard, A., Wang, L., et al. 2010, *A&A*, **518**, L22
- Crites, A. T., Bock, J. J., Bradford, C. M., et al. 2014, in *SPIE Astronomical Telescopes + Instrumentation*, eds. W. S., Holland, & J., Zmuidzinas (Montréal, Quebec, Canada), 91531W
- Croce, M., Pueblas, S., & Scocimarro, R. 2006, *MNRAS*, **373**, 369
- Daddi, E., Dickinson, M., Morrison, G., et al. 2007, *ApJ*, **670**, 156
- Daddi, E., Dannerbauer, H., Liu, D., et al. 2015, *A&A*, **577**, A46
- Davidzon, I., Ilbert, O., Faisst, A. L., Sparre, M., & Capak, P. L. 2018, *ApJ*, **852**, 107
- Decarli, R., Walter, F., González-López, J., et al. 2019, *ApJ*, **882**, 138
- Decarli, R., Aravena, M., Boogaard, L., et al. 2020, *ApJ*, **902**, 110
- Decarli, R., Pensabene, A., Venemans, B., et al. 2022, *A&A*, **662**, A60
- De Looze, I., Cormier, D., Lebouteiller, V., et al. 2014, *A&A*, **568**, A62
- Despali, G., Giocoli, C., Angulo, R. E., et al. 2016, *MNRAS*, **456**, 2486
- Driver, S. P., & Robotham, A. S. G. 2010, *MNRAS*, **407**, 10
- Faisst, A. L., Schaerer, D., Lemaux, B. C., et al. 2020, *ApJS*, **247**, 61
- Ferrara, A., Vallini, L., Pallottini, A., et al. 2019, *MNRAS*, **489**, 1
- Fudamoto, Y., Oesch, P. A., Schouws, S., et al. 2021, *Nature*, **597**, 489
- Gruppioni, C., Béthermin, M., Loiacono, F., et al. 2020, *A&A*, **643**, A8
- Hall, N. R., Keisler, R., Knox, L., et al. 2010, *ApJ*, **718**, 632
- Hezaveh, Y. D., & Holder, G. P. 2011, *ApJ*, **734**, 52
- Hilbert, S., White, S. D. M., Hartlap, J., & Schneider, P. 2007, *MNRAS*, **382**, 121
- Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, *MNRAS*, **480**, 800
- Ishigaki, M., Kawamata, R., Ouchi, M., et al. 2018, *ApJ*, **854**, 73
- Ishiyama, T., Fukushige, T., & Makino, J. 2009, *PASJ*, **61**, 1319
- Ishiyama, T., Nitadori, K., & Makino, J. 2012, ArXiv e-prints [arXiv:1211.4406]
- Ishiyama, T., Prada, F., Klypin, A. A., et al. 2021, *MNRAS*, **506**, 4210
- Karkare, K. S., Anderson, A. J., Barry, P. S., et al. 2022, *J. Low Temp. Phys.*, **209**, 758
- Keenan, R. P., Marrone, D. P., & Keating, G. K. 2020, *ApJ*, **904**, 127
- Kennicutt, R. C. 1998, *ARA&A*, **36**, 189
- Khusanova, Y., Béthermin, M., Le Fèvre, O., et al. 2021, *A&A*, **649**, A152
- Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, *ApJ*, **740**, 102
- Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, ArXiv e-prints [arXiv:1709.09066]
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., et al. 2004, *ApJ*, **609**, 35
- Lagache, G., Puget, J.-L., & Gispert, R. 1999, *Ap&SS*, **269**, 263
- Lagache, G., Bavouzet, N., Fernandez-Conde, N., et al. 2007, *ApJ*, **665**, L89
- Lagache, G., Cousin, M., & Chatzikos, M. 2018, *A&A*, **609**, A130
- Lagache, G., Béthermin, M., Montier, L., Serra, P., & Tucci, M. 2020, *A&A*, **642**, A232
- Le Fèvre, O., Béthermin, M., Faisst, A., et al. 2020, *A&A*, **643**, A1
- Loiacono, F., Decarli, R., Gruppioni, C., et al. 2021, *A&A*, **646**, A76
- Madau, P., & Dickinson, M. 2014, *ARA&A*, **52**, 415
- Maddox, S. J., Dunne, L., Rigby, E., et al. 2010, *A&A*, **518**, L11
- Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, *ApJ*, **760**, 6
- Maniyar, A., Béthermin, M., & Lagache, G. 2018, *A&A*, **614**, A39
- Maniyar, A., Béthermin, M., & Lagache, G. 2021, *A&A*, **645**, A40
- Moster, B. P., Somerville, R. S., Newman, J. A., & Rix, H.-W. 2011, *ApJ*, **731**, 13
- Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, **428**, 3121
- Niemiec, A., Jullo, E., Giocoli, C., Limousin, M., & Jauzac, M. 2019, *MNRAS*, **487**, 653
- Pallottini, A., Ferrara, A., Bovino, S., et al. 2017, *MNRAS*, **471**, 4128
- Pallottini, A., Ferrara, A., Gallerani, S., et al. 2022, *MNRAS*, **513**, 5621
- Planck Collaboration XVIII. 2011, *A&A*, **536**, A18
- Planck Collaboration XXX. 2014, *A&A*, **571**, A30
- Planck Collaboration XIII. 2016, *A&A*, **594**, A13
- Planck Collaboration VI. 2020, *A&A*, **641**, A6
- Planck Collaboration Int. XLVIII. 2016, *A&A*, **596**, A109
- Potter, D., Stadel, J., & Teysier, R. 2017, *Comput. Astrophys.*, **4**, 2
- Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, *ApJ*, **771**, 30
- Riechers, D. A., Pavesi, R., Sharon, C. E., et al. 2019, *ApJ*, **872**, 7
- Rodríguez-Puebla, A., Behroozi, P., Primack, J., et al. 2016, *MNRAS*, **462**, 893
- Sargent, M. T., Béthermin, M., Daddi, E., & Elbaz, D. 2012, *ApJ*, **747**, L31
- Sargent, M. T., Daddi, E., Béthermin, M., et al. 2014, *ApJ*, **793**, 19
- Schaerer, D., Ginolfi, M., Béthermin, M., et al. 2020, *A&A*, **643**, A3
- Schreiber, C., Pannella, M., Elbaz, D., et al. 2015, *A&A*, **575**, A74
- Shankar, F., Lapi, A., Salucci, P., De Zotti, G., & Danese, L. 2006, *ApJ*, **643**, 14
- Skillman, S.W., Warren, M.S., Turk, M.J., et al. 2014, ArXiv e-prints [arXiv:1407.2600]
- Stacey, G. J., Aravena, M., Basu, K., et al. 2018, in *Ground-based and Airborne Telescopes VII*, eds. H. K. Marshall, & J. Spyromilio, 10700 (Bellingham: International Society for Optics and Photonics (SPIE)), 107001M
- Tacconi, L. J., Genzel, R., & Sternberg, A. 2020, *ARA&A*, **58**, 157
- Thacker, C., Cooray, A., Smidt, J., et al. 2013, *ApJ*, **768**, 58
- Vale, A., & Ostriker, J. P. 2004, *MNRAS*, **353**, 189
- Vallini, L., Pallottini, A., Ferrara, A., et al. 2018, *MNRAS*, **473**, 271
- Venemans, B. P., Walter, F., Decarli, R., et al. 2017, *ApJ*, **845**, 154
- Viero, M. P., Ade, P. A. R., Bock, J. J., et al. 2009, *ApJ*, **707**, 1766
- Viero, M. P., Wang, L., Zemcov, M., et al. 2013, *ApJ*, **772**, 77
- Viero, M. P., Reichardt, C. L., Benson, B. A., et al. 2019, *ApJ*, **881**, 96
- Vizgan, D., Greve, T. R., Olsen, K. P., et al. 2022, *ApJ*, **929**, 92
- Walter, F., Decarli, R., Sargent, M., et al. 2014, *ApJ*, **782**, 79
- Walter, F., Carilli, C., Neeleman, M., et al. 2020, *ApJ*, **902**, 111
- Wang, L., Gao, F., Best, P. N., et al. 2021, *A&A*, **648**, A8
- Whitaker, K. E., Williams, C. C., Mowla, L., et al. 2021, *Nature*, **597**, 485
- Yan, L., Sajina, A., Loiacono, F., et al. 2020, *ApJ*, **905**, 147
- Yue, B., & Ferrara, A. 2019, *MNRAS*, **490**, 1928
- Zanella, A., Daddi, E., Magdis, G., et al. 2018, *MNRAS*, **481**, 1976
- Zavala, J. A., Casey, C. M., Manning, S. M., et al. 2021, *ApJ*, **909**, 165



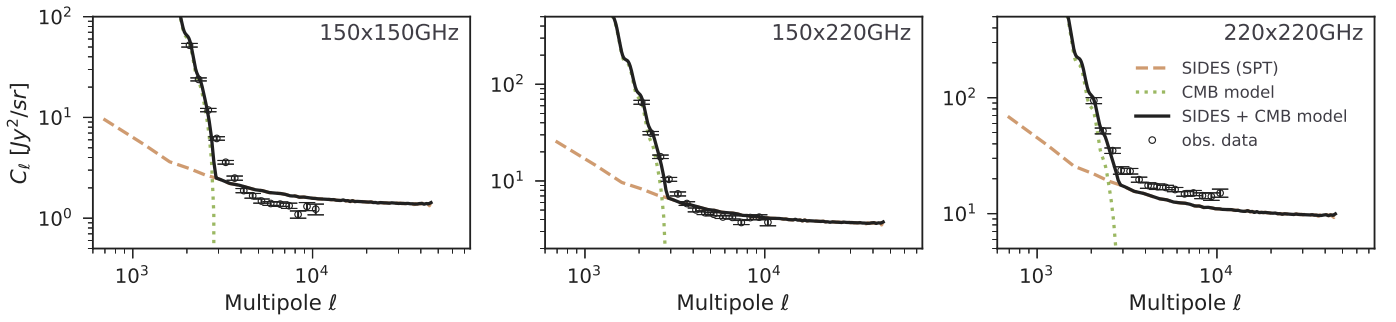
## Appendix A: Comparing SIDES with SPT and SPTxSPIRE power spectra

We generated simulated maps for the South Pole Telescope (SPT) frequency bands at 150 and 220 GHz (Hall et al. 2010) and compared the SIDES power spectra with both the SPT and SPTxSPIRE power spectra. The SPT data are contaminated with the cosmic microwave background (CMB) emission, especially at low frequency. In order to account for this we add a CMB model<sup>12</sup> on top of the SIDES model when comparing with the data (Fig. A.1). The masking procedure we follow for the SPIRE bands is the one explained in Sect. 3.2, while for SPT we mask all the sources with flux above 6.4 mJy at 150 GHz, as done in Viero et al. (2019). The applied color corrections are summarized in Table 2.

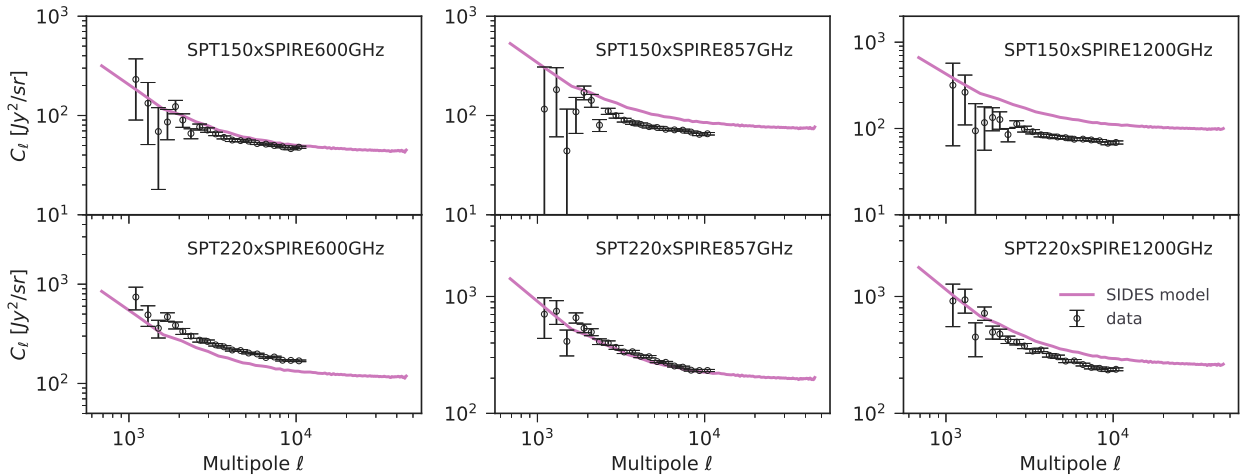
The sum of the SIDES and CMB model can reproduce the SPT measurements below  $\ell = 3 \times 10^3$ . However, at higher multipoles the SIDES model agrees with the data only

for the SPT150x220 cross power spectrum. It fails to reproduce the SPT150x150 and SPT220x220 data. The discrepancy between SPT220x220 data and SIDES is on the order of  $\sim 20\%$ , while SIDES can reproduce the *Planck*217x217 data at  $\sim 10\%$ , although at much larger scales (Sect. 3.4).

Additionally, we see significant discrepancies between SPTxSPIRE measurements and our model in Fig. A.2. SIDES is systematically above the data, except for SPT220xSPIRE600 where it is below by  $\sim 20\%$ , and for SPT220xSPIRE857 where there is a good agreement between the two, despite the fact that both SPT220x220 (Fig. A.1) and SPIRE857x857 (Fig. 4) measurements are below the model. Even though the SPIRE data can be overall reproduced by the SIDES model as shown in Sect. 3.4, the discrepancies in the SPTxSPIRE case can even exceed 40%. We thus suspect that the inconsistencies between the measurements and the model may come from these specific CIB measurements published in Viero et al. (2019).



**Fig. A.1.** Comparison of the SPT auto and cross power spectra of the 150 and 220 GHz bands (black points, Viero et al. 2019) and the SIDES model (brown dashed line). The green dotted line model is the contribution of the CMB in the data. The sum of the SIDES and CMB model is shown with the black solid line.



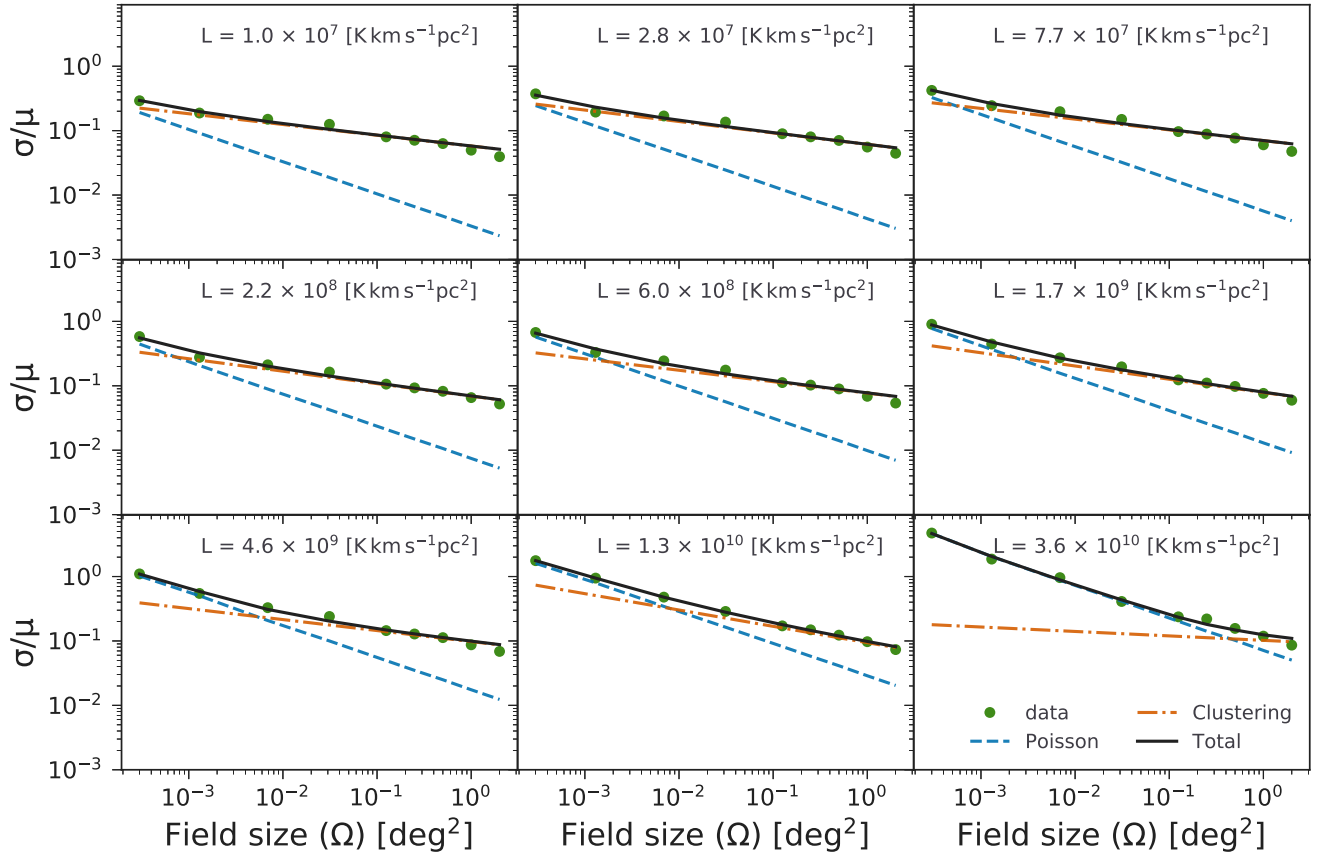
**Fig. A.2.** Comparison of the SPTxSPIRE data (black points, Viero et al. 2019) with the SIDES model (purple solid line).

<sup>12</sup> <http://pla.esac.esa.int/pla/#home>

## Appendix B: Luminosity function variance with survey size: Poisson and clustering contribution

In Fig. B.1 we show the decomposition of the total variance in the CO LFs into the Poisson and clustering components as a function of the survey size, for different luminosity bins. At the

highest luminosities ( $3.6 \times 10^{10} L_{\odot}$ ), the Poisson component is the dominant source of variance up to  $1 \text{ deg}^2$ , after which the clustering starts to dominate. For lower luminosities, this transition appears at lower field sizes. For very low luminosities ( $< 10^8 L_{\odot}$ ), even  $0.001 \text{ deg}^2$  fields are dominated by the clustering component.



**Fig. B.1.** Decomposition of the total variance of the CO LFs into the Poisson and clustering components. The variance is shown as a function of the survey size for different luminosity bins in order to visualize at which luminosities and sizes the clustering component excessively contributes to the total variance.