Origins Space Telescope: predictions for far-IR spectroscopic surveys

Matteo Bonato[1,2]*, Gianfranco De Zotti[2], David Leisawitz[3], Mattia Negrello[4], Marcella Massardi[1],
Ivano Baronchelli[5], Zhen-Yi Cai[6], Charles M. Bradford[5], Alexandra Pope[7], Eric J. Murphy[8], Lee
Armus[5] and Asantha Cooray[9]

[1]INAF−Istituto di Radioastronomia, and Italian ALMA Regional Centre, Via Gobetti 101, I-40129, Bologna, Italy
[2]INAF, Osservatorio Astronomico di Padova, Vicolo Osservatorio 5, I-35122 Padova, Italy
[3]NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt MD, USA
[4]School of Physics and Astronomy, Cardiff University, The Parade, Cardiff CF24 3AA, UK
[5]California Institute of Technology, Pasadena, CA
[6]CAS Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology
of China, Hefei, Anhui 230026, China
[7]Department of Astronomy, University of Massachusetts Amherst, Amherst, MA
[8]National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903, USA
[9]Department of Physics & Astronomy, University of California, Irvine, CA 92697, USA

Abstract

We illustrate the extraordinary potential of the (far-IR) Origins Survey Spectrometer (OSS) on board
the Origins Space Telescope (OST) to address a variety of open issues on the co-evolution of galaxies
and AGNs. We present predictions for blind surveys, each of 1000 h, with different mapped areas (a
shallow survey covering an area of $10\,\mathrm{deg}^2$ and a deep survey of $1\,\mathrm{deg}^2$) and two different concepts of the
OST/OSS: with a 5.9 m telescope (Concept 2, our reference configuration) and with a 9.1 m telescope
(Concept 1, previous configuration). In 1000 h, surveys with the reference concept will detect from
$\sim 1.9 \times 10^6$ to $\sim 8.7 \times 10^6$ lines from $\sim 4.8 \times 10^5$–2.7×10^6 star-forming galaxies and from $\sim 1.4 \times 10^4$
to $\sim 3.8 \times 10^4$ lines from $\sim 1.3 \times 10^4$–3.5×10^4 AGNs. The shallow survey will detect substantially more
sources than the deep one; the advantage of the latter in pushing detections to lower luminosities/higher
redshifts turns out to be quite limited. The OST/OSS will reach, in the same observing time, line
fluxes more than one order of magnitude fainter than the SPICA/SMI and will cover a much broader
redshift range. In particular it will detect tens of thousands of galaxies at $z \geq 5$, beyond the reach of
that instrument. The polycyclic aromatic hydrocarbons lines are potentially bright enough to allow the
detection of hundreds of thousands of star-forming galaxies up to $z \sim 8.5$, i.e. all the way through the
re-ionization epoch. The proposed surveys will allow us to explore the galaxy–AGN co-evolution up
to $z \sim 5.5 - 6$ with very good statistics. OST Concept 1 does not offer significant advantages for the
scientific goals presented here.

1 INTRODUCTION

In the last fifteen years observations with the
NASA/*Spitzer* and ESA/*Herschel* space telescopes have
driven great progress in our understanding of galaxy
formation and evolution. It is now clear that the bulk of
the super-massive black hole (SMBH) accretion and star
formation (SF) in galaxies occurred in the redshift range
$1 \lesssim z \lesssim 3$ (Merloni & Heinz 2008; Delvecchio et al. 2014;
Madau & Dickinson 2014), and that the galaxy evolution,

unlike dark matter structure formation, developed in a
"top-down" fashion, with the most massive galaxies and
Active Galactic Nuclei (AGNs) forming first ("cosmic
downsizing"). The nearly ubiquitous presence of SMBHs
at the centres of galaxies, and the close relationship be-
tween their masses and the properties of the spheroidal
stellar components (Ferrarese & Ford 2005; Kormendy
& Ho 2013), point to a strong interaction between the
build-up of mass in stars and BH growth. Nonetheless,
the details of these interactions, their impact on cosmic
downsizing, and the physical processes that govern SF

---

*matteo.bonato@inaf.it

in galaxies are still largely unclear.

Since the most active cosmic star-formation phases and the associated AGN growth are dust enshrouded, far-infrared (FIR) and sub-millimeter observations are necessary for their detection and astrophysical characterization. These are major goals of next-generation observatories.

Imaging alone is insufficient to reach these goals. For example, disentangling AGN and SF contributions via photometric SED decomposition proved to be extremely challenging even in the presence of a rich amount of optical, near-infrared and sub-mm (*Herschel*) data (i.e. Delvecchio et al. 2014; Berta et al. 2013). FIR/sub-mm spectroscopy adds key information, as demonstrated by ISO, *Herschel* and *Spitzer*, albeit with limited sensitivity.

A giant leap forward will be made possible by the FIR to sub-mm imaging and spectroscopic observations performed by the Origins Space Telescope (OST)[1]. The OST is an evolving concept for the Far-Infrared Surveyor mission, the subject of one of four science and technology definition studies supported by NASA Headquarters in preparation for the 2020 Astronomy and Astrophysics Decadal Survey. By delivering a three order-of-magnitude gain in sensitivity over previous FIR missions and high angular resolution to mitigate spatial confusion in deep surveys, the OST is being designed to cover large areas of the sky efficiently, enabling searches for rare objects at low and high redshifts.

The Origins Survey Spectrometer (OSS) provides OST's FIR spectroscopic capabilities and enables the execution of wide and deep surveys that will yield large statistical samples of galaxies up to high z. Two design concepts for OST are summarized by Leisawitz et al. (2018), and the mission will be described in detail in a report to the US National Academies' Decadal Survey in 2019. The OSS has six wide-band grating spectrometer modules, which combine to cover the full FIR spectral range (25–590 μm) simultaneously, with sensitivity close to the astronomical background limit, taking advantage of new far-IR detector array technologies. Because each grating module couples to a slit on the sky with of order 100 beams[2], and OST is agile, allowing for scans perpendicular to the slit at up to 60 mas/s, OST will be a powerful spatial-spectral survey machine.

Figure 1 shows the wavelength dependence of the sensitivity of OST with the OSS instrument for two studied telescope concepts. The figure shows $5\,\sigma$ detection limits for surveys of 1,000 h each over areas of 1 deg^2 and 10 deg^2. The wavelength dependence is primarily attributable to varying background intensity, and secondarily to instrumental effects. The OST study team considered an off-axis 9.1 m diameter telescope ("Concept 1") and a 5.9 m diameter on-axis telescope
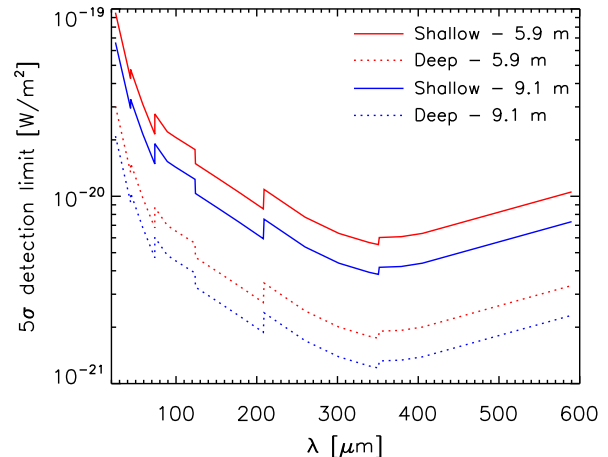

Figure 1. Detection limits of the OSS instrument on OST as a function of wavelength for surveys of 1,000 h each over areas of 1 deg^2 ("deep") and 10 deg^2 ("shallow") with 5.9 m and 9.1 m telescopes.

("Concept 2"). For Concept 2 in survey mode, the mapping speed in units of $[\mathrm{deg}^2(10^{-19}\,\mathrm{W\,m}^{-2})^2/\mathrm{sec}]$, with spectral resolving power $R = 300$, is estimated to be 1.9×10^{-5} (25–44 μm), 6.9×10^{-5} (42–74 μm), 1.2×10^{-4} (71–124 μm), 4.1×10^{-4} (119–208 μm), 1.0×10^{-3} (200–350 μm) and 7.1×10^{-4} (336–589 μm). Mapping speed is proportional to the telescope's light collecting area, and the area ratio of the Concept 1 to the Concept 2 telescope is $52\,\mathrm{m}^2/25\,\mathrm{m}^2 = 2.08$. Thus, the mapping speed for Concept 1 is about twice that of Concept 2. A survey of 1,000 h over an area of 1 deg^2 with a 5.9 m telescope implies observing times of 0.58, 1.65, 4.67, 13.1, 37.1 and 106.1 s per beam at the central wavelengths of the 6 bands listed above, respectively. The observing times per beam at fixed total time are proportional to the square of the wavelength and inversely proportional to the total area and to the square of the telescope size.

The OST will perform unprecedented blind FIR/sub-mm spectroscopic surveys deep and wide enough to provide insight into the physical processes that drive galaxy evolution out to $z \sim 8.5$. In the wavelength interval covered by OSS, atomic and molecular lines are present over a broad range of excitation levels for low- and high-z sources. By studying these lines, OST Guest Observers will be able to derive information about the physical conditions and processes active in dust-enshrouded galaxies and AGNs, and measure changes throughout cosmic history.

The detectable IR spectral lines can come from nuclear activity, SF regions, or both. Discriminating between the two contributors is not trivial. However a key diagnostic stems from the fact that AGNs produce harder radiation and consequently excite metals to higher ionization states than SF regions. Therefore, lines having high-ionization potentials are a powerful tool for recog-

---

[1] http://origins.ipac.caltech.edu/
[2] A descoped version of OSS offers half this number of beams on the sky

nizing nuclear activity (Sturm et al., 2002; Meléndez et al., 2008; Goulding & Alexander, 2009; Weaver et al., 2010), which may be difficult or impossible to identify via optical spectroscopy in the dusty galaxies with intense SF. Ratios between pairs of lines with different ionization potentials and comparable critical densities permit the estimation of the gas ionization state. Ratios between pairs of lines having comparable ionization potentials and different critical densities allow us to estimate the gas density (Spinoglio & Malkan 1992). Rubin (1989) provides a comprehensive discussion of IR density indicators. Sturm et al. (2002) presented diagnostic diagrams involving different ratios of IR lines (e.g., [NeVI]7.63/[NeII]12.81 vs [NeVI]7.63/[OIV]25.89), which can be used to identify composite (star-forming plus AGN) galaxies and discriminate between the two components. Other diagnostics exploiting less luminous lines have been presented by Voit (1992b) and Spinoglio & Malkan (1992). Given the large number of possible combinations of lines, such diagnostic plots are adaptable to various redshift intervals.

Moreover, the OST with OSS will detect molecular emission lines, in particular CO lines, which are tracers of the physical, chemical and dynamical conditions in the dense gas typically associated with photo-dissociation regions (PDRs) and the X-ray dominated regions (XDRs) associated to AGNs. The high-J CO lines, detected by *Herschel* in local galaxies will be observed with the OST up to high z. Such lines, having high ionization potentials, are indicators of AGN activity. They are also useful tracers of the physics of the molecular gas. Sub-mm lines are essentially unaffected by dust extinction and can therefore be used to disentangle the contributions from dust enshrouded AGNs and star formation in high-redshift galaxies. AGNs and star-forming galaxies have different low- to high-J CO luminosity ratios (see e.g. Carilli & Walter 2013) due to the different excitation mechanisms and physical conditions (especially temperature and density) in the molecular gas.

In its high-resolution mode, OSS offers velocity resolution $\lesssim 6.2 \, \mathrm{km \, s^{-1}}$ at $100 \mu m$ (Bradford et al. 2018), and will provide information on powerful molecular outflows, with velocities of thousands $\mathrm{km \, s^{-1}}$. Such outflows are thought to be driven by AGN feedback. As is well known, AGN feedback is a key, but still poorly understood ingredient of most galaxy formation models. A proper study of the outflow detectability requires dedicated, intensive simulations (cf., e.g., González-Alfonso et al., 2017), which are beyond the scope of the present paper.

The OSS on OST will also be able to detect polycyclic aromatic hydrocarbon (PAH) spectral features up to very high redshifts. PAH luminosities are useful indicators of the SF rate (SFR; e.g. Roussel et al., 2001; Förster Schreiber et al., 2004; Peeters et al., 2004; Desai et al., 2007; Shipley et al., 2016). PAHs were detected in a large sample of galaxies at quite high z (343 ultra-

luminous IR galaxies at $0.3 < z < 2.8$; Kirkpatrick et al. 2015), but it is an open question whether PAH features exist in galaxies at higher redshifts. There are indeed indications in the literature of a PAH emission deficit in low-metallicity/low-luminosity star-forming galaxies compared to higher-metallicity and/or higher-luminosity SF galaxies (see e.g. Galametz et al. 2009, Rosenberg et al. 2008, Smith et al. 2007, Madden et al. 2006, O'Halloran et al. 2006, Engelbracht et al. 2005, 2008).

The mass-metallicity relation (Tremonti et al. 2004; Zahid et al. 2013) tells us that the metallicity decreases with decreasing stellar mass, and hence with decreasing luminosity. At high z this effect has a stronger impact since the average metallicity decreases with increasing z (Zahid et al. 2013). The deficit in PAH emission could be due to greater permeability of dust clouds by PAH-destroying or dissociating interstellar radiation (Galliano et al. 2003), as low-metallicity clouds are expected to have lower dust-to-gas ratios. PAHs are potentially detectable by the OST/OSS up to $z \sim 8.5$, allowing us - for the first time - to shed light on these issues and to analyse the properties of the interstellar medium (ISM) in the early Universe.

Some of the FIR lines detectable by the OST/OSS can be used to measure the metallicity of galaxies. Since the metal abundance results from the cumulative star-formation activity and the gas outflow/inflow history in galaxies, the metallicity of stars and gas in galaxies is an important discriminator between various galaxy evolutionary scenarios. One of the main IR metallicity diagnostics is the ([OIII] $51.81 \mu m$+[OIII] $88.36 \mu m$)/[NIII] $57.32 \mu m$ ratio (Nagao et al. 2011, Pereira-Santaella et al. 2017). This diagnostic tool is applicable to star-forming galaxies and AGNs, and it is almost immune to extinction and insensitive to temperature, density, and the hardness of the radiation field. An additional abundance diagnostic is the ([NeII] $12.81 \mu m$ +[NeIII] $15.55 \mu m$)/([SIII] $18.71 \mu m$ +[SIV] $10.49 \mu m$) ratio (Fernández-Ontiveros et al. 2016, 2017). With OSS, OST observers will be able to use these diagnostics to measure the metallicities of hundreds of thousands of galaxies up to z~6.

In this paper, we adopt the Cai et al. (2013) evolutionary model, as upgraded by Bonato et al. (2014b). The model was validated by comparison with a broad array of data: multi-frequency (optical, near-IR, mid-IR, FIR, sub-mm) luminosity functions of galaxies and AGNs at all available redshifts, multi-frequency redshift distributions at several flux density limits, multi-frequency source counts, global and per-redshift slices (Cai et al., 2013; Bonato et al., 2014a)[3]. Appendix A presents comparisons of model predictions with more recent data.

---

[3]Additional comparisons of model predictions with observations can be found at `http://people.sissa.it/~zcai/galaxy_agn/` or `http://staff.ustc.edu.cn/~zcai/galaxy_agn/index.html`.

Although there are newer models in the literature (e.g., Guo et al., 2016; Lacey et al., 2016; Casey et al., 2018), our model is physically grounded and has been tested successfully with the broadest variety of observational data. It also deals self-consistently with the emission of galaxies as a whole, including both the star-formation and the AGN components, with their dependence on galaxy age; this is essential for the purposes of the present paper.

This paper is structured as follows. In Section 2 we present the model. In Section 3 we derive correlations between line and continuum luminosity. In Section 4 we describe our procedure to work out predictions for number counts, IR luminosity functions and redshift distributions of galaxies and AGNs detectable in the OSS wavelength range. In Section 5 we compare the capabilities of the OST 5.9 m and 9.1 m concepts. In Section 6 we investigate in more detail the expected outcome of surveys of different depth with the 5.9 m telescope. In Section 7 we describe the expected scientific impact of the OST/OSS surveys, discuss uncertainties, and compare the OST/OSS performance with that of other FIR/sub-mm instruments. Finally, in Section 8 we summarize our main conclusions.

Throughout this paper, we adopt a flat ΛCDM cosmology with $\Omega_{\mathrm{m}} = 0.31$, $\Omega_{\Lambda} = 0.69$ and $h = H_0/100 \, \mathrm{km \, s^{-1} \, Mpc^{-1}} = 0.67$ (Planck Collaboration XIII, 2016).

2 OUTLINE OF THE MODEL

In the local Universe, spheroids (i.e. ellipticals and bulges of disk galaxies) are characterized by relatively old stellar populations with mass-weighted ages $\gtrsim 8$–9 Gyr, corresponding to formation redshifts $z \gtrsim 1$–1.5, while the disk populations are generally younger. Therefore the progenitors of present day spheroidal galaxies (called proto-spheroids or proto-spheroidal galaxies) are the dominant star forming population at $z \gtrsim 1.5$, whereas the SF at $z \lesssim 1.5$ occurs mainly in galaxy disks.

There is clear evidence of a co-evolution of proto-spheroids and of AGNs hosted by them. Their nuclear activity continues for a relatively short time, of the order of 10^8 yr, after star formation is quenched. At later times, the central super-massive black holes remain mostly inactive, except for occasional "rejuvenations" due to interactions or mergers. The late nuclear activity is mostly associated with S0's and spirals, which contain substantial amounts of interstellar medium that can eventually flow towards the nucleus and be accreted.

To deal with these different evolutionary paths, our reference model (Cai et al., 2013; Bonato et al., 2014b) adopts a "hybrid" approach. It provides a physically grounded description of the redshift-dependent co-evolution of the SFR of spheroidal galaxies and of the accretion rate onto the super-massive black holes at their centers, while the description of the evolution of late-type galaxies and of AGNs associated with them is phenomenological and parametric.

The model considers two sub-populations of late-type galaxies: "warm" (starburst) and "cold" (normal) galaxies. While in the case of spheroids the original model by Cai et al. (2013) dealt simultaneously with the stellar and AGN components, AGNs associated with late-type galaxies were evolved as a separate population. Bonato et al. (2014b) have upgraded the model to allow for co-evolution of late-type galaxies and their AGNs. This was done by exploiting the mean relation between SFR and accretion rate derived by Chen et al. (2013), taking into account the dispersion around it. The relative abundances of type 1 and type 2 AGNs, as a function of luminosity, were taken into account following Hasinger (2008). Bright, optically selected QSOs (for which the Chen et al. 2013 correlation is not applicable) are taken into account by adopting the best fit evolutionary model by Croom et al. (2009) up to $z = 2$ (optical AGNs at higher z are already included in the Cai et al. 2013 model). This approach reproduces the observed bolometric luminosity functions of AGNs at different z (see Bonato et al. 2014b).

3 CORRELATIONS BETWEEN CONTINUUM AND LINE LUMINOSITY

To estimate the number of AGN and galaxy line detections achievable with OST/OSS surveys, we coupled the SFR/IR luminosity functions (for the galaxies) or the bolometric luminosity functions (for the AGNs) of each population, as given by the model, with relationships between line and IR[4] or bolometric luminosities.

We used the calibrations derived by Bonato et al. (2014a,b, 2015, 2017b) for 31 IR fine-structure lines and the 6 polycyclic aromatic hydrocarbon (PAH) bands at 3.3, 6.2, 7.7, 8.6, 11.3 and 12.7 μm, as they will be accessible in the OST/OSS wavelength range.

The 31 IR fine-structure lines are:

- 1 coronal region line: [SiVII] 6.50 μm;
- 8 AGN fine-structure emission lines: [NeVI] 7.65, [ArV] 7.90, [CaV] 11.48, [ArV] 13.09, [MgV] 13.50, [NeV] 14.32, [NeV] 24.31 and [OIV] 25.89 μm;
- 13 stellar/HII region lines: [ArII] 6.98, [ArIII] 8.99, [SIV] 10.49, HI 12.37, [NeII] 12.81, [ClII] 14.38, [NeIII] 15.55, [SIII] 18.71, [ArIII] 21.82, [SIII] 33.48, [OIII] 51.81, [NIII] 57.32 and [NII] 121.9 μm;
- 5 lines from photodissociation regions: [FeII] 17.93, [FeIII] 22.90, [FeII] 25.98, [SiII] 34.82 and

---

[4]SFRs were converted into L_{IR} following Kennicutt & Evans (2012). We neglected the small differences between the calibration based on the Kroupa & Weidner (2003) initial mass function (IMF) used by Kennicutt & Evans (2012) and that based on the Chabrier (2003) IMF used by Cai et al. (2013).
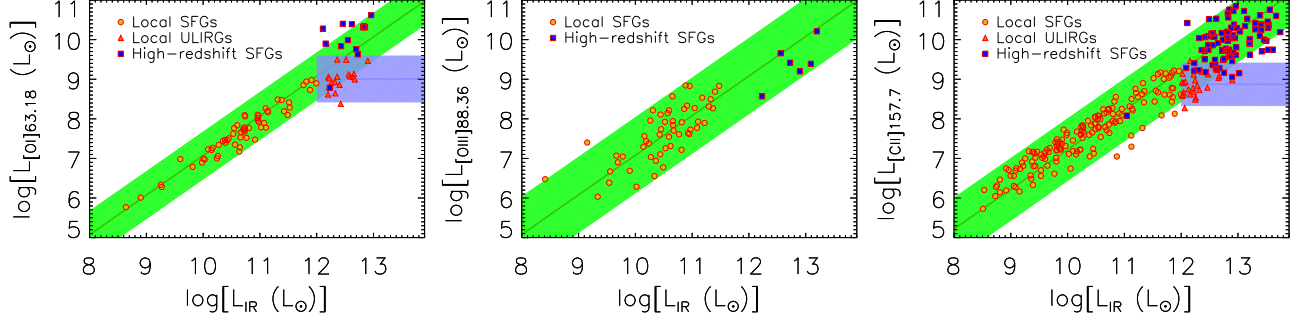
Figure 2. Luminosity of the [OI] 63.18μm (left panel), [OIII] 88.36μm (central panel) and [CII] 157.7μm (right panel) lines, versus continuum IR luminosity. The green bands show the 2σ range around the mean relation $\log(L_\ell/L_\odot) = \log(L_{\rm IR}/L_\odot) + c$ for local SFGs with $L_{\rm IR} < 10^{12}\,L_\odot$ (yellow circles) and high-redshift SFGs (blue squares); the values of $c \equiv \langle\log(L_\ell/L_{\rm IR})\rangle$ are given in Table 1. The azure bands show the 2σ spread around the mean line luminosity for the sample of local ULIRGs ($L_{\rm IR} \geq 10^{12}\,L_\odot$; orange triangles) whose line luminosities appear to be uncorrelated with $L_{\rm IR}$ and are generally lower than expected from the linear relation holding for the other sources. The mean line luminosities $\langle\log(L_\ell/L_\odot)\rangle$ of these objects are given in Table 1 as well. See text for the sources of data points.
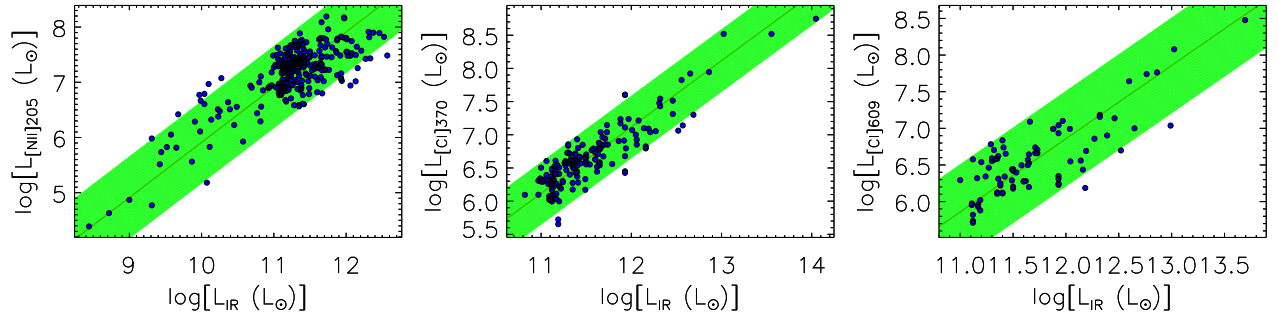
Figure 3. Luminosity of the [NII] 205.2μm (left panel), [CI] 370.4μm (central panel) and [CI] 609.1μm (right panel) lines, versus continuum IR luminosity. The green bands show the 2σ range around the mean linear relation $\log(L_\ell/L_\odot) = \log(L_{\rm IR}/L_\odot) + c$; the values of $c \equiv \langle\log(L_\ell/L_{\rm IR})\rangle$ are given in Table 1. See text for the sources of data points.
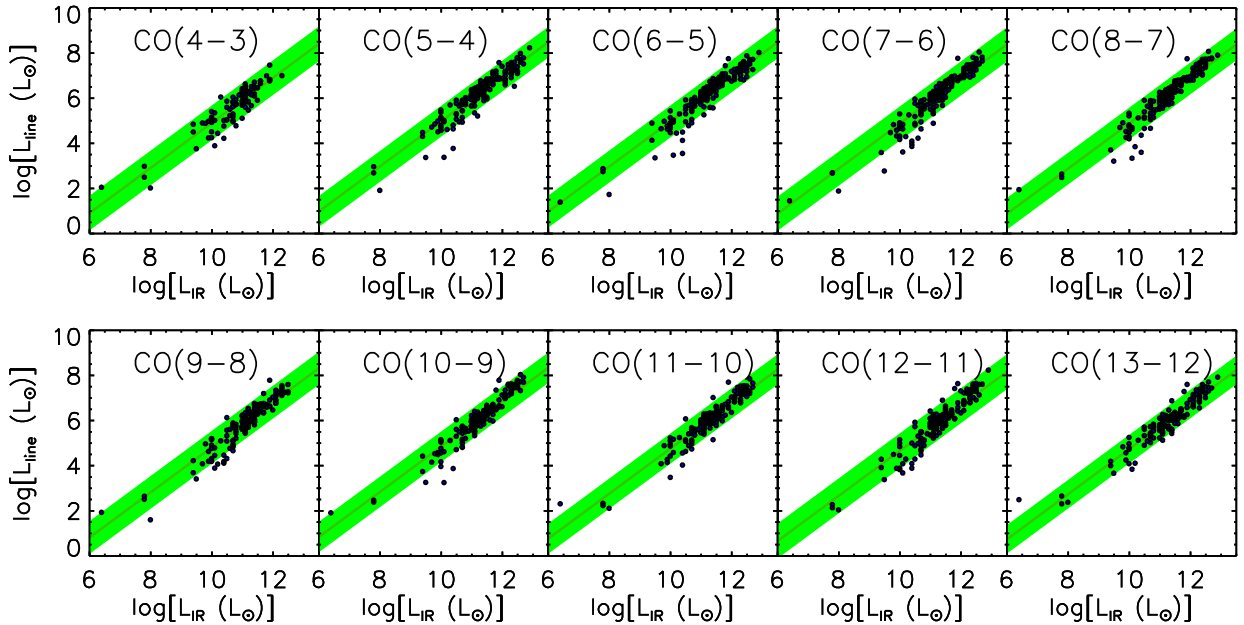
Figure 4. Luminosity of the CO J=4-3 through J=13-12 lines versus continuum IR luminosity. The data points are from Kamenetzky et al. (2016). The green band shows the $2\,\sigma$ range around the mean linear relation $\log(L_\ell/L_\odot) = \log(L_{\mathrm{IR}}/L_\odot) + c$; the values of $c \equiv \langle\log(L_\ell/L_{\mathrm{IR}})\rangle$ are given in Table 1.

[OI] $145.5\,\mu$m;

- 4 molecular hydrogen lines: $H_2\,6.91$, $H_2\,9.66$, $H_2\,12.28$ and $H_2\,17.03\,\mu$m.

We re-calibrated the [OI] 63.18 and [CII] $157.7\,\mu$m photodissociation region lines and the [OIII] $88.36\,\mu$m stellar/HII region line (whose line vs. IR luminosity correlations were given in Bonato et al. 2014a), and added more recent data. For these three lines, in addition to the compilation by Bonato et al. (2014a), we used data taken from: George (2015), for all three lines; Brisbin et al. (2015) and Farrah et al. (2013), for the [OI] 63.18 and [CII] $157.7\,\mu$m lines; and Oteo et al. (2016), Gullberg et al. (2015), Schaerer et al. (2015), Yun et al. (2015) and Magdis et al. (2014), for the [CII] $157.7\,\mu$m line only. We excluded all objects for which there is evidence of a substantial AGN contribution, as was done in our previous work.

The line and continuum measurements of strongly lensed galaxies given by George (2015) were corrected using the gravitational magnifications, μ, estimated by Ferkinhoff et al. (2014) while those by Gullberg et al. (2015) were corrected using the magnification estimates from Hezaveh et al. (2013) and Spilker et al. (2016), available for 17 out of the 20 sources; for the other three sources we used the median value $\mu_{\mathrm{median}} = 7.4$.

For these three lines, Fig. 2 shows the correlations between line and IR luminosity. In this plot, the galaxies are subdivided into three groups: local star-forming galaxies (SFGs; $L_{\mathrm{IR}} < 10^{12}\,L_\odot$, represented by yellow

circles), local Ultra-Luminous IR Galaxies (ULIRGs, $L_{\mathrm{IR}} \geq 10^{12}\,L_\odot$, orange triangles), and high-z SFGs (blue squares).

We warn the reader that the high-z measurements of the three lines mostly refer to strongly lensed galaxies and are therefore affected by the substantial and hard to quantify uncertainty on the correction for magnification, in addition to measurement errors. Not only are estimates of μ for these objects generally poorly constrained by the available data, but they do not necessarily apply to the line emitting gas, which may have a different spatial distribution than the emission used to build the lensing model (differential lensing; Serjeant, 2012). It is, however, reassuring that high-z galaxies do not substantially deviate, on average, from the relations defined by the low-z galaxies.

As noted by Bonato et al. (2014a), luminosities of the [OI] 63.18 and [CII] $157.7\,\mu$m lines in local ULIRGs do not show any significant correlation with L_{IR}. For such objects we adopted a Gaussian distribution of the logarithm of the line luminosity, $\log(L_\ell)$, around its mean value (azure band in Fig. 2). Line luminosities of the other populations (low-z non-ULIRGs and high-z SFGs) are consistent with a direct proportionality between $\log(L_\ell)$ and $\log(L_{\mathrm{IR}})$, i.e. $\langle\log(L_\ell/L_{\mathrm{IR}})\rangle = c \pm \Delta c$, represented by the green bands in Figs. 2, 3 and 4[5]. The mean values and the dispersions are listed in

---

[5]The choice of a direct proportionality between $\log(L_\ell)$ and $\log(L_{\mathrm{IR}})$ for lines associated with SF is supported by the extensive

Table 1 Mean values of the log of line-to-IR (8-1000 μm) continuum luminosities of galaxies, $\langle \log(L_\ell/L_{\mathrm{IR}})\rangle$, and associated dispersions σ. For the PAH 3.3μm, PAH 11.3μm, PAH 12.7μm bands and the H$_2$ 17.03μm, [OI] 63.18μm and [CII] 157.7μm lines, $\langle \log(L_\ell/L_{\mathrm{IR}})\rangle$ has been computed excluding local ULIRGs, for which the line luminosity appears to be uncorrelated with L_{IR}. For the latter objects, the last column gives the mean values of $\log(L_\ell/L_\odot)$ and their dispersions.

Spectral line	$\left\langle \log\left(\frac{L_\ell}{L_{\mathrm{IR}}}\right)\right\rangle (\sigma)$	$\langle\log(L_\ell)_{\mathrm{UL}}/L_\odot\rangle (\sigma)$
PAH 3.3μm[4]	-3.11 (0.30)	8.54 (0.23)
PAH 6.2μm[2]	-2.20 (0.36)	-
H$_2$ 6.91μm[3]	-3.97 (0.39)	-
[ArII] 6.98μm[3]	-3.96 (0.32)	-
PAH 7.7μm[2]	-1.64 (0.36)	-
PAH 8.6μm[2]	-2.16 (0.36)	-
[ArIII] 8.99μm[3]	-4.22 (0.69)	-
H$_2$ 9.66μm[2]	-3.96 (0.52)	-
[SIV] 10.49μm[2]	-3.95 (0.69)	-
PAH 11.3μm[1]	-2.29 (0.36)	9.01 (0.28)
H$_2$ 12.28μm[2]	-4.12 (0.54)	-
HI 12.37μm[3]	-4.85 (0.30)	-
PAH 12.7μm[4]	-2.20 (0.33)	9.75 (0.28)
[NeII] 12.81μm[1]	-3.11 (0.45)	-
[ClII] 14.38μm[3]	-5.44 (0.33)	-
[NeIII] 15.55μm[1]	-3.69 (0.47)	-
H$_2$ 17.03μm[1]	-4.04 (0.46)	8.07 (0.34)
[FeII] 17.93μm[3]	-5.18 (0.34)	-
[SIII] 18.71μm[1]	-3.49 (0.48)	-
[ArIII] 21.82μm[3]	-5.40 (0.70)	-
[FeIII] 22.90μm[3]	-6.56 (0.33)	-
[FeII] 25.98μm[3]	-4.29 (0.44)	-
[SIII] 33.48μm[1]	-3.05 (0.31)	-
[SiII] 34.82μm[1]	-2.91 (0.28)	-
[OIII] 51.81μm[1]	-2.84 (0.44)	-
[NIII] 57.32μm[1]	-3.26 (0.16)	-
[OI] 63.18μm	-2.93 (0.30)	9.01 (0.30)
[OIII] 88.36μm	-2.92 (0.48)	-
[NII] 121.90μm[1]	-3.49 (0.36)	-
[OI] 145.52μm[1]	-3.80 (0.43)	-
[CII] 157.7μm	-2.78 (0.40)	8.87 (0.27)
[NII] 205.18μm	-4.09 (0.37)	-
[CI] 370.42μm	-4.88 (0.24)	-
[CI] 609.14μm	-5.14 (0.33)	-
CO(13 − 12)	-5.26 (0.32)	-
CO(12 − 11)	-5.33 (0.39)	-
CO(11 − 10)	-5.23 (0.33)	-
CO(10 − 9)	-5.16 (0.34)	-
CO(9 − 8)	-5.18 (0.36)	-
CO(8 − 7)	-5.12 (0.35)	-
CO(7 − 6)	-5.12 (0.38)	-
CO(6 − 5)	-5.02 (0.35)	-
CO(5 − 4)	-5.02 (0.34)	-
CO(4 − 3)	-5.09 (0.38)	-

[1] Taken from Bonato et al. (2014a)
[2] Taken from Bonato et al. (2014b)
[3] Taken from Bonato et al. (2015)
[4] Taken from Bonato et al. (2017b)

Table 1.

IR luminosities given over rest-frame wavelength ranges different from the one adopted here (8–1000 μm), were corrected using the following relations, given by Stacey et al. (2010) and Graciá-Carpio et al. (2008), respectively:

$$L_{\mathrm{FIR}}(40 - 500\,\mu\mathrm{m}) = 1.5 \times L_{\mathrm{FIR}}(42 - 122\,\mu\mathrm{m}), \quad (1)$$
$$L_{\mathrm{IR}}(8 - 1000\,\mu\mathrm{m}) = 1.3 \times L_{\mathrm{FIR}}(40 - 500\,\mu\mathrm{m}). \quad (2)$$

To this initial set of lines we added the [NII] 205.2 μm stellar/HII region line and the [CI] 370.4 and [CI] 609.1 μm photo-dissociation region lines. We collected data on SFGs from: Lu et al. (2017), Rosenberg et al. (2015) and Walter et al. (2011) for the [CI] 370.4 and [CI] 609.1 μm lines; Lu et al. (2017), Herrera-Camus et al. (2016) and Zhao et al. (2013, 2016) for the [NII] 205.2 μm line. In Fig. 3, we show the correlations between line and IR emission for these lines. The best-fit values for a direct proportionality and the associated dispersions are listed in Table 1.

For AGNs, empirical correlations between line and bolometric luminosities for 11 IR lines in our sample (specifically, the H$_2$ 9.66, [SIV] 10.49, H$_2$ 12.28, [NeII] 12.81, [NeV] 14.32, [NeIII] 15.55, H$_2$ 17.03, [SIII] 18.71, [NeV] 24.31, [OIV] 25.89 and [SIII] 33.48μm lines) were derived by Bonato et al. (2014b).

Bonato et al. (2015) used the IDL Tool for Emission-line Ratio Analysis (ITERA)[6] to calibrate the line-to-bolometric luminosity relations for 16 additional AGN IR lines with insufficient observational data ([SiVII] 6.50, H$_2$ 6.91, [ArII] 6.98, [NeVI] 7.63, [ArV] 7.90, [ArIII] 8.99, [CaV] 11.48, HI 12.37, [ArV] 13.09, [MgV] 13.50, [ClII] 14.38, [FeII] 17.93, [ArIII] 21.82, [FeIII] 22.90, [FeII] 25.98, [SiII] 34.82μm). These calibrations were found to be in excellent agreement with the available data.

We followed the same procedure to calibrate the line-to-bolometric luminosity relations for the remaining 10 AGN atomic spectral lines in our sample. The AGN contributions to the [NIII] 57.32μm, [NII] 121.90/205.18μm, and [CI] 370.42/609.14μm lines turned out to be negligible. The best-fit coefficients of the linear relations $\log(L_\ell/L_\odot) = a \cdot \log(L_{\mathrm{bol}}/L_\odot) + b$ and the associated $1\,\sigma$ dispersions for the lines in our sample are listed in Table 2.

Finally, we derived the line-to-IR luminosity relations for 10 CO lines (J=4-3 through J=13-12), using the measurements of Kamenetzky et al. (2016). For these molecular lines, the data are consistent with both a linear relation and a direct proportionality between $\log(L_{\mathrm{line}})$ and $\log(L_{\mathrm{IR}})$. We chose a direct proportionality to avoid an unnecessary second parameter, as was done for the

---

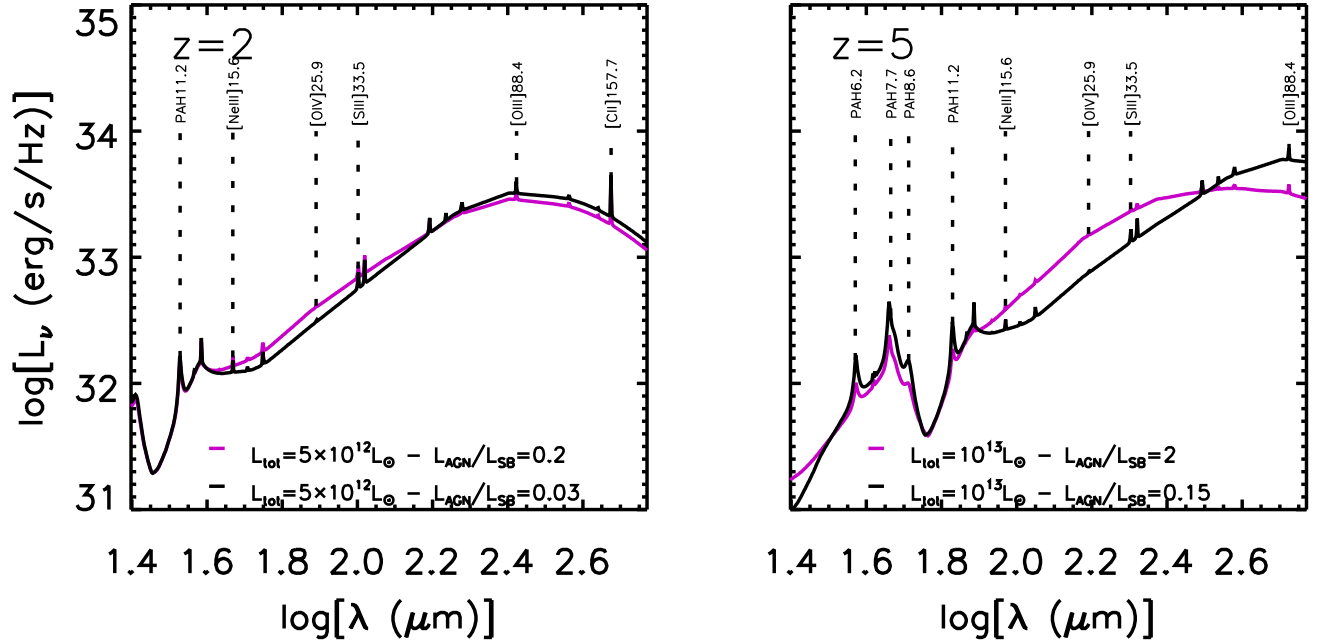simulations (taking into account dust obscuration) carried out by Bonato et al. (2014a)

[6] http://home.strw.leidenuniv.nl/ brent/itera.html

Figure 5. Examples of SEDs with AGN and SF components in the observer frame, over the wavelength range of the OST/OSS (25–590μm). The SEDs are shown for sources at $z = 2$ (left) and $z = 5$ (right) and for the different total source luminosities $L_{\rm tot} = L_{\rm AGN} + L_{\rm IR,SF}$ and $L_{\rm AGN}/L_{\rm IR,SF}$ ratios specified in the inset. All the spectral lines of our sample are included in the SEDs. Some of the brightest lines are labeled.

atomic lines. These relations are shown in Fig. 4, with the corresponding best-fit parameter values listed in Table 1. Note that while the AGN contribution to the high-J CO lines could be significant, the available observational data are not sufficient to support a derivation of the empirical correlation between AGN bolometric and CO line luminosity.

Similarly to what was done in Bonato et al. (2017b), the detection limits for the broad PAH bands were computed assuming that the spectra are degraded to $R = 50$ from the original $R = 300$, resulting in a sensitivity gain by a factor of $\sqrt{6}$.

4 NUMBER COUNTS AND REDSHIFT DISTRIBUTIONS

We computed redshift-dependent line luminosity functions using the Monte Carlo approach described in Bonato et al. (2014a,b). These papers dealt with dust-obscured star formation phases, for which $L_{\rm IR}$ is an excellent estimator of the SFR. The line-$L_{\rm IR}$ relations presented in Section 3 were also based on observations of dusty galaxies, for which the unabsorbed fraction of the UV emission from young stars is small.

However, the OST with OSS will push the observational frontier to lower luminosity, higher redshift galaxies, with higher fractions of unabsorbed UV light, i.e. with smaller fractions of the SFR measured by $L_{\rm IR}$. The following question then arises: Are the relationships derived above for the case "$L_{\rm IR}$ equivalent to star for-
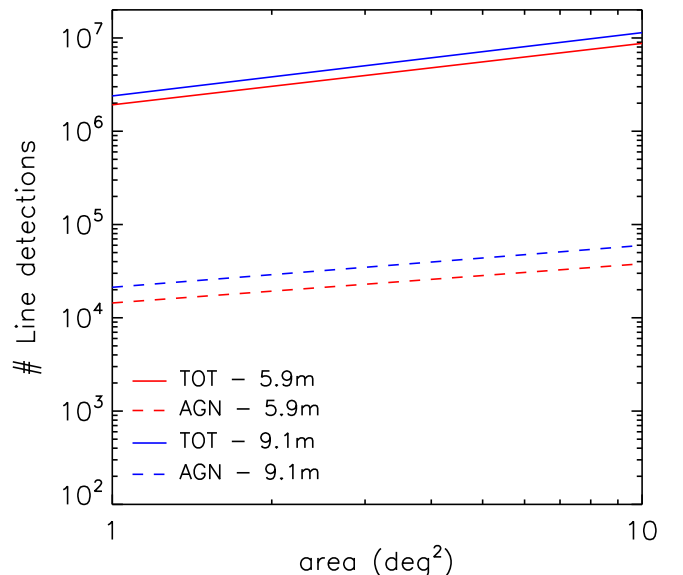

Figure 6. Number of 5σ starburst (solid) and AGN (dashed) line detections as a function of the mapped area, for a survey of 1000 h.

Table 2 Coefficients of the best-fit linear relations between line and AGN bolometric luminosities, $\log(L_\ell/L_\odot) = a \cdot \log(L_{\mathrm{bol}}/L_\odot) + b$, and $1\,\sigma$ dispersions around the mean relations.

Spectral line	a	b	disp (1σ)
[SiVII]6.50μm[2]	0.83	-3.55	0.37
H$_2$6.91μm[2]	0.80	-2.40	0.34
[ArII]6.98μm[2]	0.84	-4.21	0.64
[NeVI]7.63μm[2]	0.79	-1.48	0.42
[ArV]7.90μm[2]	0.87	-3.85	0.32
[ArIII]8.99μm[2]	0.98	-4.15	0.37
H$_2$9.66μm[1]	1.07	-5.32	0.34
[SIV]10.49μm[1]	0.90	-2.96	0.24
[CaV]11.48μm[2]	0.90	-5.12	0.34
H$_2$12.28μm[1]	0.94	-3.88	0.24
HI12.37μm[2]	0.86	-3.84	0.34
[NeII]12.81μm[1]	0.98	-4.06	0.37
[ArV]13.09μm[2]	0.87	-3.85	0.32
[MgV]13.50μm[2]	0.91	-4.01	0.34
[NeV]14.32μm[1]	0.78	-1.61	0.39
[ClII]14.38μm[2]	0.85	-4.83	0.57
[NeIII]15.55μm[1]	0.78	-1.44	0.31
H$_2$17.03μm[1]	1.05	-5.10	0.42
[FeII]17.93μm[2]	0.84	-3.80	0.54
[SIII]18.71μm[1]	0.96	-3.75	0.31
[ArIII]21.82μm[2]	0.98	-5.34	0.36
[FeIII]22.90μm[2]	0.79	-4.85	0.60
[NeV]24.31μm[1]	0.69	-0.50	0.39
[OIV]25.89μm[1]	0.70	-0.04	0.42
[FeII]25.98μm[2]	0.87	-3.71	0.55
[SIII]33.48μm[1]	0.62	0.35	0.30
[SiII]34.82μm[2]	0.89	-3.14	0.52
[OIII]51.81μm	0.92	-2.47	0.32
[OI]63.18μm	0.86	-2.99	0.62
[OIII]88.36μm	0.91	-3.34	0.32
[OI]145.52μm	0.88	-4.32	0.63
[CII]157.70μm	0.90	-4.42	0.57

[1]Taken from Bonato et al. (2014b)
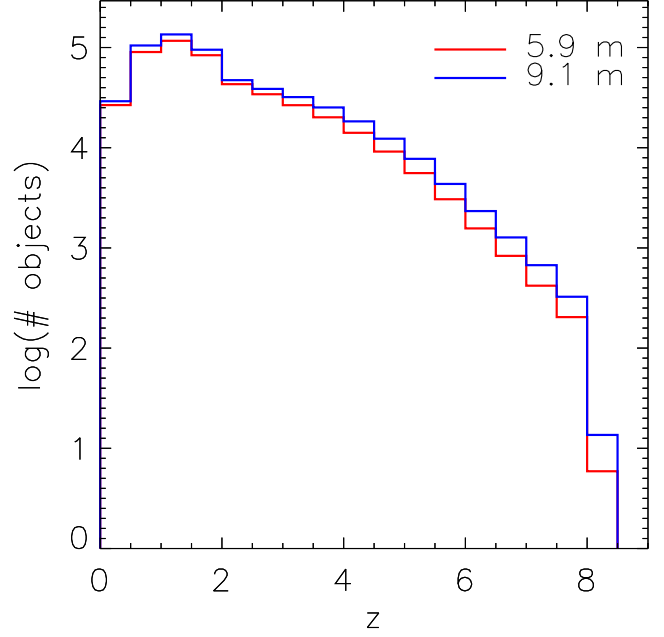[2]Taken from Bonato et al. (2015)

Figure 7. Predicted redshift distributions of galaxies detected with a 5.9 m and a 9.1 m OST spectroscopic survey in 1000 h of observing time, assuming an areal coverage of $1\,\mathrm{deg}^2$.
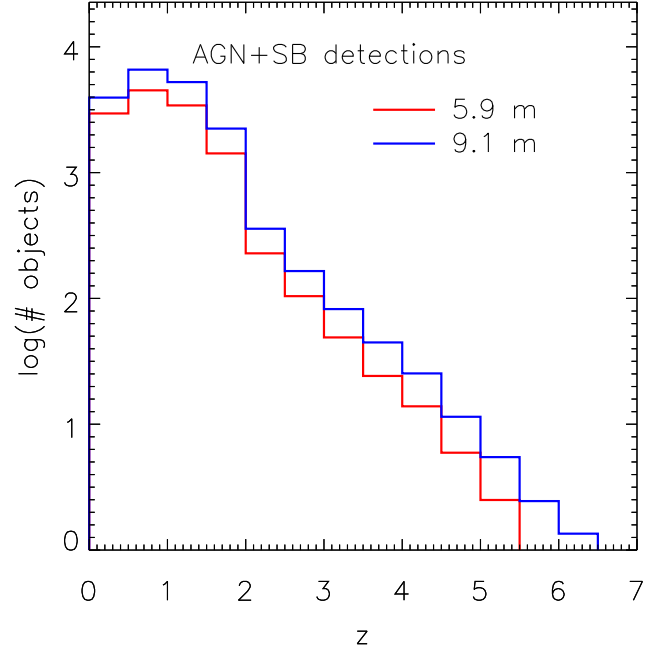

Figure 8. Predicted redshift distributions of galaxies detected simultaneously in at least one SF line and one AGN line with a 5.9 m and a 9.1 m OST spectroscopic survey in 1000 h of observing time, assuming an areal coverage of $1\,\mathrm{deg}^2$. The corresponding AGN redshift distributions are almost identical to these, indicating that whenever the AGN component is detected, the SF component is detected as well.

mation luminosity" to be interpreted as relations of line luminosity with SFR, or with $L_{\rm IR}$?

De Looze et al. (2014) analyzed the reliability of three of the brightest FIR fine-structure lines, [CII] 157.7 μm [OI] 63.18 and [OIII] 88.36 μm, to trace the obscured plus unobscured SFR (measured by the FIR and by the UV luminosity, respectively) in a sample of low-metallicity dwarf galaxies from the *Herschel* Dwarf Galaxy Survey. Furthermore, they extended the analysis to a large sample of galaxies of various types and metallicities taken from the literature. Their results confirmed that the line luminosity primarily correlates with the SFR.

We could not find in the literature any other study on the relation between FIR lines and SFR that took into account both the UV and the FIR luminosity. In general, the SFR is derived from the FIR luminosity. We have therefore assumed that the conclusions by De Looze et al. (2014) can be extended to all FIR fine structure lines: we do not see any obvious reason why the luminosity of these lines should depend on the fraction of UV radiation absorbed by dust.

The situation is different for the PAH emission because these large molecules may be destroyed by absorption of UV or soft X-ray photons (Voit, 1992a). They more easily survive if they are shielded by dust. A weakening of PAH emission relative to dust emission at low gas-phase metallicity was reported by Engelbracht et al. (2005). Shipley et al. (2016) found that the PAH luminosity correlates linearly with the SFR, as measured by the extinction-corrected Hα luminosity, for gas-phase metallicity $\log Z = 12 + \log(O/H) \geq \log Z_c = 8.55$, but strongly decreases with decreasing Z for $Z \leq Z_c$. We have adopted their empirical correction for $Z \leq Z_c$ galaxies:

$$\log L_{\rm PAH,\lambda}^{\rm corr} = \log L_{\rm PAH,\lambda} + A[\log Z - (\log Z_\odot - \log Z_0)], \quad (3)$$

where $L_{\rm PAH,\lambda}$ is the luminosity of the PAH band at the wavelength λ derived from the IR luminosity, taken as a measure of the SFR (Sect. 3), A is given, for different PAH bands, by Table 4 of Shipley et al. (2016), and $\log Z_0 = \log Z_\odot - 8.55 = 0.14$. For the conversion from $L_{\rm IR}$ to SFR we have adopted the calibration by Kennicutt & Evans (2012):

$$\log({\rm SFR}/M_\odot\,{\rm yr}^{-1}) = \log(L_{\rm IR}/L_\odot) - 9.82. \quad (4)$$

A relation between metallicity, SFR, and stellar mass, $M_\star$, approximately independent of redshift, was found by Hunt et al. (2016):

$$12 + \log(O/H) = -0.14 \log({\rm SFR}) + 0.37 \log(M_\star) + 4.82. \quad (5)$$

Finally, a redshift-dependent relationship between $M_\star$ and SFR was derived by Aversa et al. (2015) using the abundance-matching technique. We have exploited this set of equations to account for the metallicity dependence of the PAH luminosity.

The molecular lines are even more complex. Two main issues are at play: the relationship between SFR and molecular gas density (essentially the Schmidt-Kennicutt law; Kennicutt, 1998), whose shape is still debated, and the uncertain CO-to-H_2 conversion factor (Bolatto et al., 2013). A strong decrease of the CO luminosity to SFR at low metallicity has long been predicted as a consequence of the enhanced photo-dissociation of CO by ultraviolet radiation (e.g., Bolatto et al., 2013). Genzel et al. (2012) and Hunt et al. (2015) found that a direct proportionality between $L_{\rm CO}$ and SFR, consistent with that reported in Section 3, holds for $Z \gtrsim Z_\odot$, while the $L_{\rm CO}$/SFR ratio drops for lower metallicity. We have adopted the empirical relationship presented by Hunt et al. (2015) for $\log Z \lesssim 8.76$ (see their Fig. 5):

$$\log L_{\rm CO}^{\rm corr} = \log L_{\rm CO}({\rm SFR}) + 2.25 \log Z - 19.71, \quad (6)$$

where $L_{\rm CO}({\rm SFR})$ is the relationship derived in Section 3, after converting $L_{\rm IR}$ to SFR using eq. (4). The mean metallicity corresponding to a given SFR was computed as described above.

Clearly, the relationships between the PAH or CO emissions and the SFR (hence our estimates of the corresponding redshift-dependent luminosity functions and number counts) are affected by large uncertainties, especially at high redshifts. FIR spectroscopic surveys conducted with the OST will shed light on these issues.

As in Bonato et al. (2014b), our simulations take into account both the SF and the AGN components (see Sect. 2). We calculated line luminosity functions in 100 redshift bins with a width, Δz, of ~ 0.08. For each redshift bin, at the bin center, we extracted from the model SFR/IR luminosity functions (for galaxies) or from the bolometric luminosity functions (for AGNs), a number of objects, brighter than $\log(L_{\rm IR,min}/L_\odot) = 8.0$, equal to that expected in $1\,{\rm deg}^2$.

The luminosities of SF and AGN components of each source were randomly extracted from the probability distributions that an object at redshift z has a starburst luminosity $L_{\rm IR,SF}$ or an AGN luminosity $L_{\rm AGN}$, given the total luminosity $L_{\rm tot} = L_{\rm IR,SF} + L_{\rm AGN}$. The probability distributions are given by Bonato et al. (2014b).

The line luminosities associated with each component were randomly taken from Gaussian distributions with the mean values and dispersions given in Tables 1 and 2. In this way we get, at the same time, luminosities of *all* the lines of both the SF and AGN components of each simulated object, allowing us to make interesting predictions. For example, we can ask: For how many objects will both components be detectable given the survey sensitivity? How many will be detectable in two or more lines? Figure 5 shows examples of SEDs with AGN and SF components at $z = 2$ and $z = 5$.

We derived line luminosity functions by binning the simulated line luminosities within each redshift bin. We repeated the simulations 1000 times and averaged the
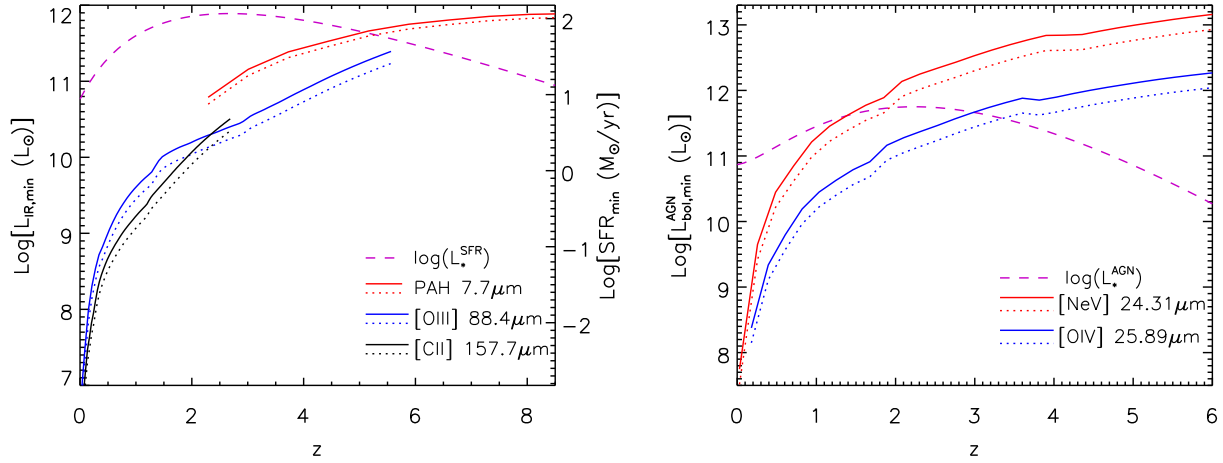
Figure 9. Minimum IR luminosity - SFR (left panel, left and right y-axis respectively) and minimum AGN bolometric luminosity (right panel) detected, as a function of redshift, with a 5.9 m (solid lines) and a 9.1 m (dotted lines) OST survey with an areal coverage of $1\,\mathrm{deg}^2$, through spectroscopic detections of key SF and AGN lines. The dashed purple lines show, for comparison, estimates of the redshift-dependent characteristic luminosity, $L_\star$, of dusty galaxies and AGNs, respectively (see text).
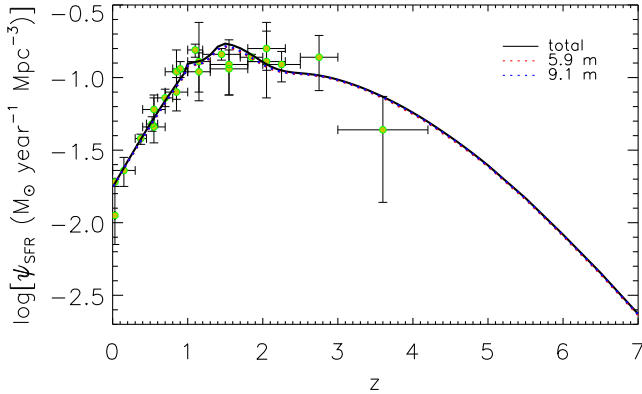

Figure 10. Cosmic SFR density resolved by an OST/OSS deep survey $(1,000\,\mathrm{h}$ over an area of $1\,\mathrm{deg}^2)$ with a 5.9 m (dotted red line) or a 9.1 m (dotted blue dotted) telescope, as a function of z. The black solid line shows the total SFR density yielded by our model. The data points are from Madau & Dickinson (2014).

derived line luminosity functions. Finally, number counts were derived after assigning to each source a redshift drawn at random from a uniform distribution within the redshift bin and computing the corresponding flux. For more details, see Bonato et al. (2014b).

5 COMPARISON BETWEEN THE 5.9 M AND 9.1 M CONCEPTS

The first question we want to address is whether, from the point of view of extragalactic spectroscopic surveys with OST, there is strong scientific motivation to favor the more ambitious Concept 1 (9.1 m telescope), compared to Concept 2 (5.9 m telescope with a central obstruction). With the performances mentioned in

Sect. 1, realistic surveys of different depths could cover areas in the range 1–$10\,\mathrm{deg}^2$.

5.1 Number counts

The survey depth at fixed observing time scales as the square root of the ratio of the effective light collecting areas of the two concepts, i.e. as $(52\,\mathrm{m}^2/25\,\mathrm{m}^2)^{1/2} = 1.44$. To quantify the effect of such a difference in depth, we have computed the number of $5\,\sigma$ line detections in $1000\,\mathrm{h}$ by the OSS with both OST concepts, as a function of the mapped area. The results are shown in Fig. 6.

In all cases, the slopes of the counts are relatively flat at the detection limits, implying that the number of detections in each line increases more with the area than with the depth of the survey. We expect $\sim 1.9 \times 10^6$ to $\sim 8.7 \times 10^6$ line detections for the 5.9 m telescope, depending on the survey area, and $\sim 2.4 \times 10^6$ to $\sim 1.1 \times 10^7$ for the 9.1 m telescope. The number of detections in pure AGN lines ranges from $\sim 1.4 \times 10^4$ to $\sim 3.8 \times 10^4$ for the 5.9 m telescope, depending on the survey area, and from $\sim 2.1 \times 10^4$ to $\sim 6.0 \times 10^4$ for the 9.1 m telescope.

We thus conclude that the decrease of the telescope size from 9.1 m to 5.9 m does not translate into a substantial worsening of the detection statistics, at least in terms of counts.

5.2 Redshift and luminosity distributions

Of course, a comparison of the total number of detections is not enough to answer the question posed at the beginning of this section, since going deeper may allow us to reach poorly populated but interesting regions of

the redshift-luminosity plane.

Figure 7 shows the predicted redshift distributions of sources detectable by OSS in at least one spectral line over an area of $1\,\mathrm{deg}^2$ in 1,000 h of observations for both OST mission concepts. Although the larger telescope surpasses the smaller one in source detections, and does so increasingly at greater redshifts, both telescope sizes allow galaxy detections with good statistics up to $z \simeq 8$, while the statistics are poor in both cases at higher redshifts.

The advantage of a larger telescope is somewhat greater if the objective is to investigate galaxy-AGN co-evolution, which requires that sources be detectable simultaneously in at least one SF line and one AGN line (Fig. 8). At $z \geq 4.5$ the 9.1 m telescope detects about twice as many sources as the 5.9 m telescope, but the statistics are limited in both cases at this redshift.

To investigate the effect of telescope size on the determination of the redshift-dependent IR luminosity functions (hence of the SFR functions) and of AGN bolometric luminosity functions, we computed the minimum L_{IR} and the minimum $L_{\mathrm{bol}}^{\mathrm{AGN}}$ corresponding to the minimum luminosity of some bright lines detectable in a 1,000 h survey of $1\,\mathrm{deg}^2$. To this end, we have used the appropriate value of the mean line-to-IR or bolometric luminosity ratio. The results are shown, as a function of redshift, in Fig. 9.

The luminosities of the chosen SF lines (left panel) are proportional to L_{IR}, so the larger telescope would allow OST observers to reach luminosities lower by a factor of 1.44. In the case of AGN lines, the ratio depends on luminosity so the advantage of the larger telescope slightly increases with z, reaching a factor of $\simeq 1.7$ at $z = 6$ (right panel).

The right-hand scale on the left panel of Fig. 9 shows the SFR corresponding to L_{IR}, as given by eq. (4). Two points should be noted here. First, as discussed in Sect. 4, in the case of the PAH lines eq. (4) breaks down at low metallicities, so the correspondence between L_{IR} derived from the PAH luminosity and the SFR is to be taken with caution, although the effect of metallicity is expected to be minor at the relatively high values of L_{IR} shown here. Second, eq. (4) holds for a Kroupa stellar initial mass function (IMF). Recent ALMA measurements of multiple CO transitions in four strongly lensed sub-mm galaxies at $z \simeq 2 - 3$ (Zhang et al., 2018) have provided evidence for a top-heavy IMF, implying a greater fraction of massive stars. Their preferred IMF would translate into a higher $L_{\mathrm{IR}}/\mathrm{SFR}$ ratio by about 40%.

The dashed purple lines in Fig. 9 show, for comparison, estimates of the characteristic luminosity, $L_\star$, as a function of redshift, for dusty galaxies and AGNs (left- and right-hand panel, respectively). The plotted values of $L_\star(z)$ refer to the modified Schechter function defined by eq. (9) of Aversa et al. (2015). In the case of dusty galaxies, $L_{\mathrm{IR,min}} < L_\star$ up to $z \simeq 5.9$ or $z \simeq 6.2$ for the

5.9 m and the 9 m telescope, respectively. This OST/OSS will then provide a good description of the evolution of the bulk of the IR luminosity function, hence of the IR luminosity density, up to these redshifts. In the AGN case, $L_{\mathrm{bol,min}} < L_\star$ up to $z \simeq 3$ or $z \simeq 3.3$ for the two telescope sizes.

Finally, Fig. 10 shows the fractions of the SFR density resolved by the deep spectroscopic surveys as a function of z for the two telescope sizes. The SFR density is almost fully resolved in both cases ($\simeq 97\%$ with the 5.9 m concept, $\simeq 98\%$ with the 9.1 m telescope).

5.3 Confusion

The telescope in both OST concepts is diffraction limited at $30\,\mu\mathrm{m}$. Throughout most of the OSS wavelength range, a larger telescope provides better angular resolution, and will be less prone to spatial confusion. Is source confusion an issue, especially for the 5.9 m telescope? Since spectroscopy adds a third dimension, confusion is in general far less significant than would be the case for a continuum survey. In spectroscopy, confusion arises if lines from different sources along the line of sight can show up by chance in the same spectral and spatial resolution element.

Figure 11 shows the total integral number counts of lines per beam[7] and per spectral resolution element at different wavelengths within the OST/OSS spectral coverage, for the 5.9 m and 9.1 m telescope concepts. The line detection limits for a deep survey of $1\,\mathrm{deg}^2$ are represented by the vertical dashed lines. Even in the worst case (longest wavelength and smaller telescope size), the source density per resolution element is $< 1/22$, i.e. the confusion is only marginal, and observations at shorter wavelengths, where the detection limits are well above the confusion limit, will help to resolve confusion in three (spatial-spectral) dimensions. Thus, line confusion is never an issue for the 9.1 m telescope and realistic exposure times. The 5.9 m telescope option can hit the confusion limits only at the longest wavelengths, and only for very deep surveys ($\geq 1,000\,\mathrm{h}$ over $1\,\mathrm{deg}^2$).

6 SURVEYS WITH THE 5.9 M OST

The analysis presented in the previous section has not exposed any scientifically compelling advantage of the 9.1 m telescope over the 5.9 m telescope. In this section we focus on the latter option and compare in more detail the outcome of surveys covering different areas at fixed total observing time, considering the cases of a wide and shallow ($10\,\mathrm{deg}^2$ area) and a deep ($1\,\mathrm{deg}^2$) survey, each conducted in an observing time of 1,000 h.

In Fig. 12 we display the redshift distributions of galax-

---

[7]Each beam has θ_{FWHM} of $1.13\lambda/D$ and a solid angle of $\theta_{\mathrm{FWHM}}^2 \times \pi/(4\ln 2)$
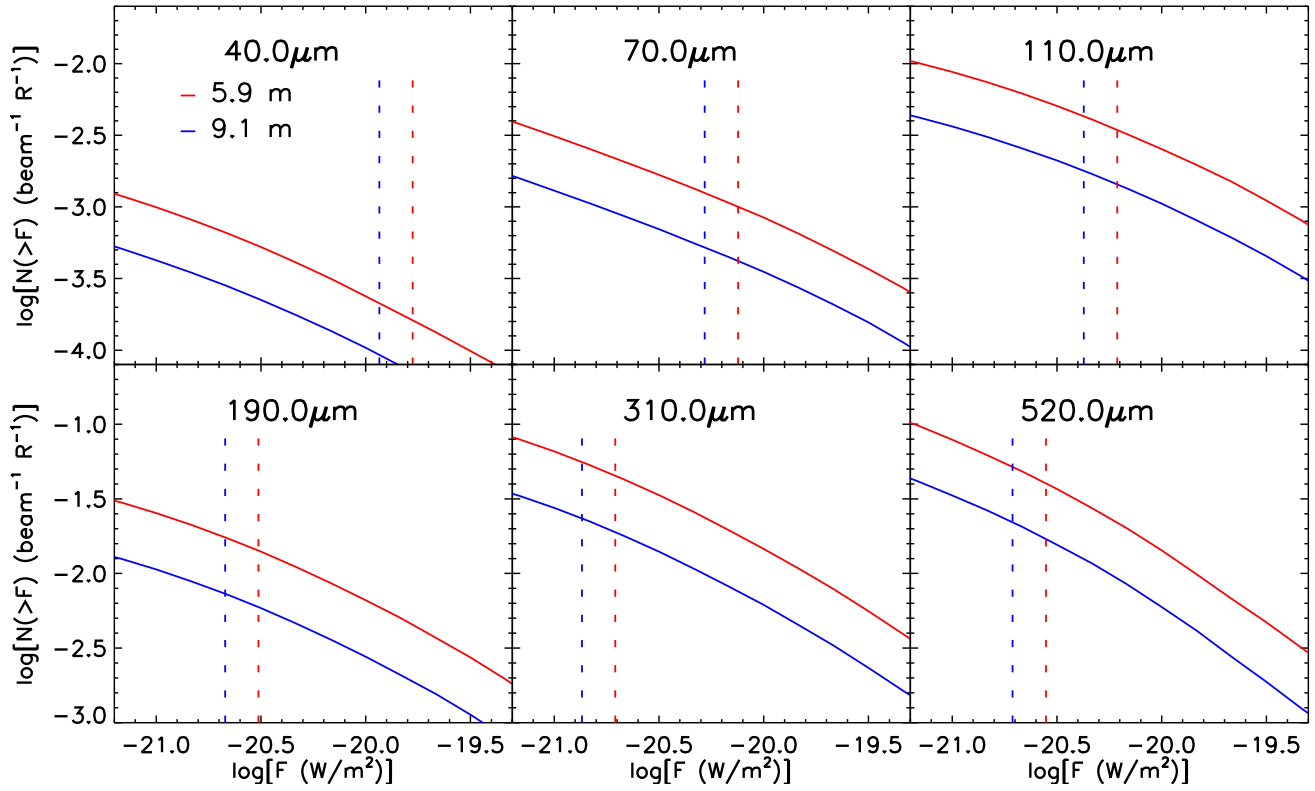
Figure 11. Total integral line number counts per beam and per spectral resolution element at different wavelengths within the OST/OSS spectral coverage for the 5.9 m (red) and the 9.1 m (blue) telescopes. The vertical dashed lines represent the detection limits for surveys of 1,000 h covering $1 \deg^2$.
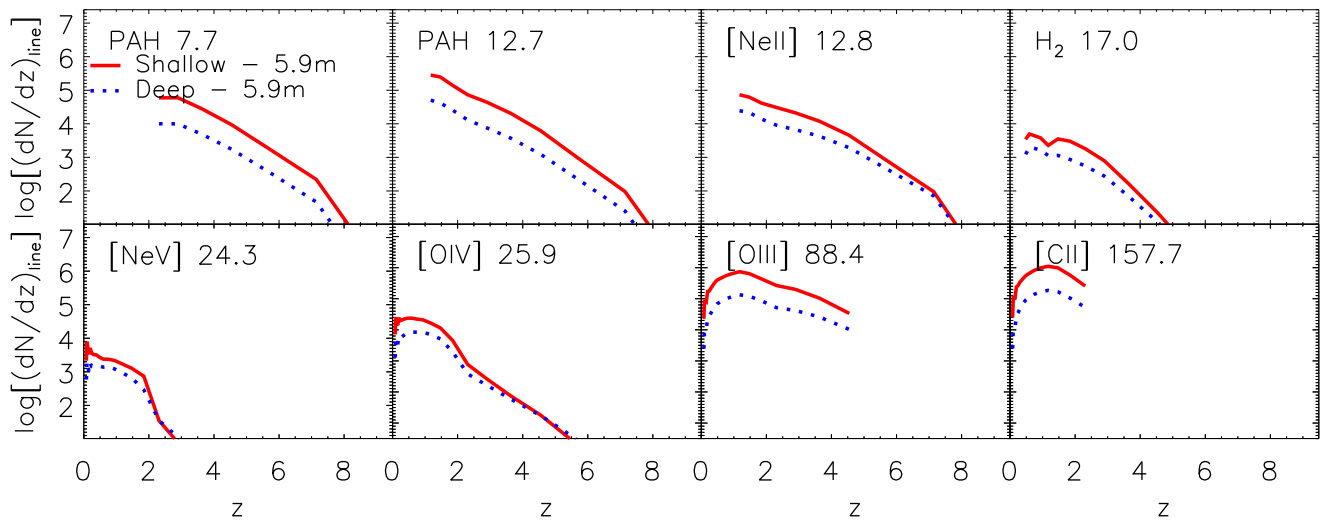

Figure 12. Predicted redshift distributions of galaxies detected in the three brightest AGN lines ([NeV]14.32, [NeV]24.31, and [OIV]-25.89 μm) and in selected lines indicating star formation, namely those with relatively high detection rates. The predicted distributions pertain to shallow (solid red lines, $10 \deg^2$ areal coverage) and deep (dotted blue lines, $1 \deg^2$ areal coverage) surveys, each conducted in 1000 h of observing time with a 5.9 m telescope.
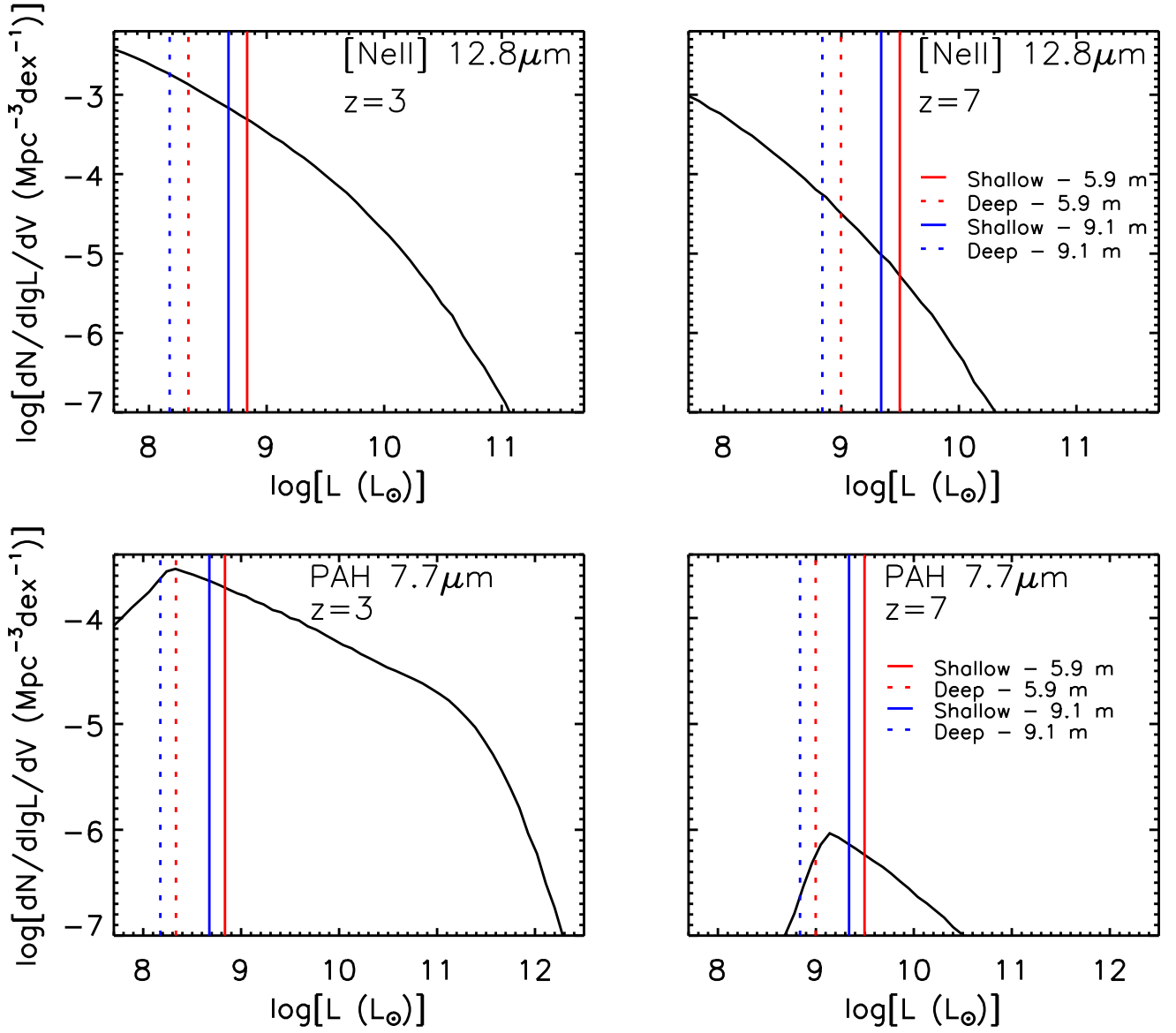
Figure 13. Luminosity functions of the fine structure line [NeII] 12.8 μm (upper panels) and the PAH 7.7 μm line at $z = 3$ and $z = 7$. The decline of the faint end of the PAH luminosity function is due to the effect of decreasing metallicity.

ies observed in the three brightest AGN lines ([NeV]-14.32, [NeV]24.31, and [OIV]25.89 μm) and in some of the most often detected star formation indicators. Obviously, only the shortest wavelength lines can be detected up to very high redshifts. The figure clearly shows that the OST is so sensitive that even the shallow survey can reach the highest redshifts. Again, the number of detections increases more with increasing survey area than with increasing depth. Hence, the shallow survey yields better statistics than the deep one, except at the highest redshifts, and only for certain lines. Even in the most favourable cases, the deep survey offers only minor advantages from a statistical perspective.

A factor of 10 decrease in the surveyed area at fixed total observing time translates into a decrease in the minimum detectable line luminosity at given z by a factor of $10^{1/2} \simeq 3.16$. In the case of the fine structure lines, such as [NeII] 12.8 μm (upper panels of Fig. 13), which are estimators of the total (dust obscured and unobscured) SFR, going deeper simply means extending the luminosity function to fainter levels. In the case of the [NeII] 12.8 μm line, the ratios of numbers of detections *per unit area* between the deep and the shallow surveys are $\simeq 3.2$ and 7.5 at $z = 3$ and $z = 7$, respectively, for a 5.9 m telescope. These ratios do not compensate for the factor of 10 difference in survey area, so the total number of detections is larger for the wide, shallow survey.

In the case of PAH lines the minimum detectable luminosity is in the range where their emission is substantially reduced due to a decrease in metallicity, as discussed in Sect. 4, and a corresponding dip is seen at the low-luminosity end of the PAH luminosity function. Hence, the increase in the number of PAH detections *per unit area* with decreasing detection limit is rather modest. For example, the ratio of the number of detections of the 7.7 μm line between the deep and shallow surveys is $\simeq 1.7$ at $z = 3$ and $\simeq 2.1$ at $z = 6$ (lower panels of Fig. 13). The increase in the total number of detections per unit area is also small since PAHs are the dominant contributors. On the other hand, this luminosity range carries information about the poorly known dependence of $L_{\rm IR}$ on metallicity and on the impact of varying metallicity on the PAH emission.

According to our model, the brightest IR AGN line, [OIV]25.89μm, will be detected in $\sim$ 33,000 sources by the shallow survey and in $\sim$ 12,000 sources by the deep one, reaching, in both cases, $z \sim 5.5$. The shallow survey will yield about 750,000 PAH detections up to $z \sim 8.5$, the deep survey about 135,000 up to $z \sim 7.5$. The counter-intuitive decrease in the maximum redshift with increasing survey depth, at fixed observing time, is due to the increasing scarcity of bright PAH lines at high z, a consequence of decreasing metallicity: large-area surveys are needed to find rare sources.

The [CII] 157.70μm line, the strongest line emitted by the cool gas in galaxies ($<10^4$ K; see Carilli & Walter 2013), is detectable up to $z \sim 3.2$ by the two surveys. We predict that the line will be seen in about 1.6 million galaxies by the shallow survey and in $\sim 280,000$ galaxies by the deep survey. Moreover, the shallow and deep surveys will yield $\sim 12,300$ and $\sim 4,500$ detections, respectively, up to $z \sim 5$, in the four warm H_2 lines, which are good diagnostics of the turbulent gas (upper right-hand panel of Fig. 12).

The conclusion that there is no need for exposures longer than those of the shallow survey to reach the highest redshifts is confirmed by Fig. 14, which illustrates the global redshift distributions of sources detectable with the two notional surveys. The shallow survey will detect $\sim 2.7 \times 10^6$ galaxies, while the deep survey will detect $\sim 4.8 \times 10^5$. The peak of the distributions is at $z \sim 1.5$, close to the peak of SF and BH accretion activity. The shallow and deep surveys will detect about 31,000 and about 12,000 galaxies at $z > 5$, respectively.

For the fixed amount of observing time considered, the deep survey does not offer any significant advantage over the shallow survey, even if we are looking for sources detected simultaneously in at least one SF line and one AGN line. Figure 15 shows that the deep survey detects more such objects only at $z > 5$, but even there the difference is small and the statistics are poor. The total number of detections of such high-z sources are of $\sim 3.4 \times 10^4$ and $\sim 1.3 \times 10^4$ for the shallow and deep surveys, respectively.

Next consider the two metallicity diagnostics described in Sect. 1, involving the following spectral lines: 1) [NeII] 12.81, [NeIII] 15.55, [SIII] 18.71, and [SIV] 10.49μm; or 2) [OIII] 51.81, [OIII] 88.36, and [NIII] 57.32μm. Figure 16 shows the redshift distributions of sources detectable by the shallow and deep surveys conducted with a 5.9 m telescope in 1000 h of observing time, for which these two metallicity diagnostics can be built thanks to the simultaneous detection of the two sets of lines.

The deep survey is advantageous for the first metallicity diagnostic at $z > 4$, although the total number of detections is larger for the shallow survey ($\sim 2,300$ vs $\sim 1,400$). In the case of the second metallicity diagnostic the shallow survey is better at all redshifts: the total number of detections is $\simeq 3.6 \times 10^5$ with $\sim 1,000$ galaxies at $z > 5$, versus $\simeq 1.0 \times 10^5$ with ~ 600 galaxies at $z > 5$ for the deep survey.

7 DISCUSSION

7.1 Galaxy-AGN co-evolution

As shown in Fig. 8, we predict the simultaneous detection of SF and AGN lines for a large sample (a few tens of thousands) of galaxies. The brightest IR AGN line - [OIV]25.89μm - will allow us to estimate the bolometric luminosity of the AGN component of the sources up to
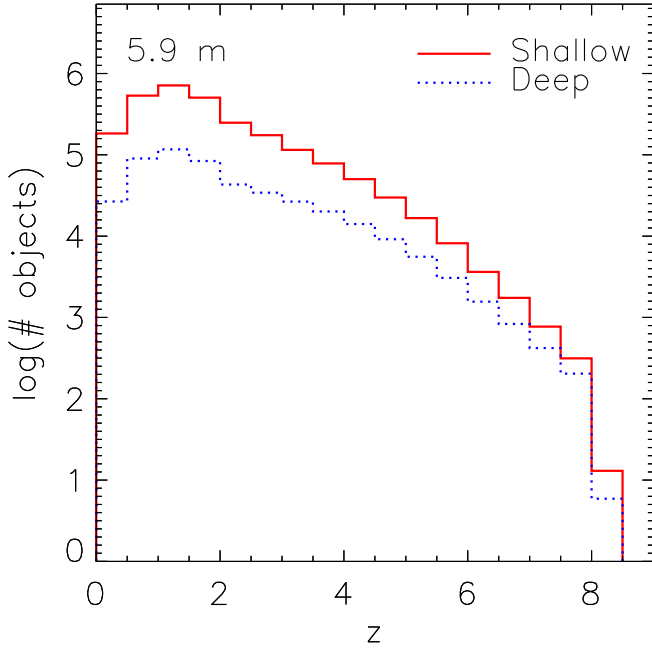
Figure 14. Predicted global redshift distributions of galaxies detected in spectral lines by the shallow ($10\,\mathrm{deg}^2$) and deep ($1\,\mathrm{deg}^2$) surveys with a 5.9 m telescope. Each survey has a total amount of observing time equal to 1000 h.
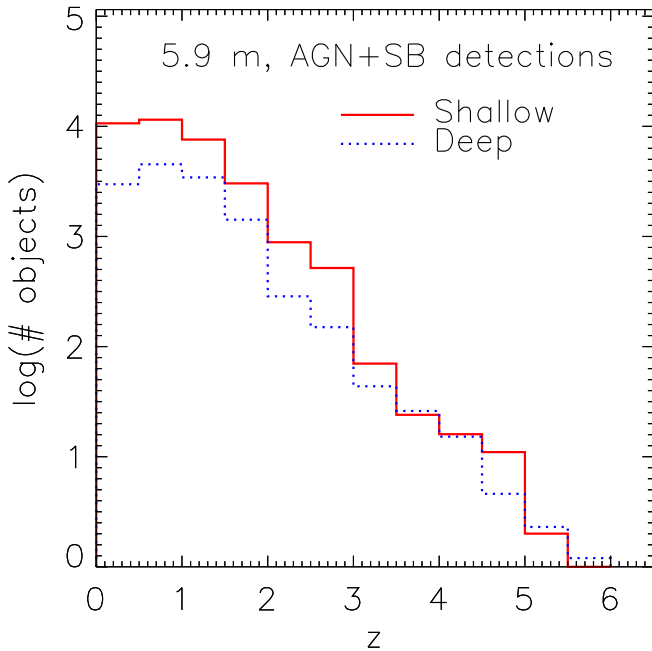

Figure 15. Predicted redshift distributions of galaxies detected simultaneously in at least one SF line and one AGN line by the deep ($1\,\mathrm{deg}^2$) and shallow ($10\,\mathrm{deg}^2$) surveys with a 5.9 m telescope. Both surveys have a total amount of observing time equal to 1000 h. As in the case of Fig. 8, these distributions are almost identical to the redshift distributions of AGNs detected by the two surveys.

$z \sim 5.5 - 6$, while starburst lines provide estimates of the IR luminosity, and hence the SFR up to $z \sim 8 - 8.5$ (Fig. 12). In other words, for the first time OST will enable observers to explore the physical processes driving galaxy-AGN (co-)evolution throughout most of cosmic history, with excellent statistics.

Figure 17 illustrates the extent to which galaxy-AGN co-evolution can be tested with OSS spectroscopy at three redshifts ($z = 2$, 3 and 5) around and beyond the peak of SF and BH accretion activity. The figure shows, as a function of area mapped, the minimum ratio of the AGN bolometric luminosity, L_{AGN}, to the IR luminosity, $L_{\mathrm{IR}} \equiv L_{\mathrm{IR,SF}}$, of the host galaxy for which OST can detect both the [OIV] 25.89μm and the [OIII] 88.36μm lines. As shown above, these lines are good indicators of the AGN and SF luminosities, respectively. The curves refer to five values of the total bolometric luminosity, $L_{\mathrm{tot}} = L_{\mathrm{AGN}} + L_{\mathrm{IR,SF}}$, of the source[8], ranging from 5×10^{11} to $5 \times 10^{13}\, L_{\odot}$[9].

During the dust-obscured intense SF phase, the BH is expected to accrete at, or perhaps slightly above, the Eddington limit. At the end of this phase, its mass should, on average, be close to that given by the correlations with host galaxy properties. The corresponding AGN bolometric luminosity, L_{AGN}, can be about an order of magnitude greater than the IR luminosity of the host galaxy, $L_{\mathrm{IR,SF}}$. For a source with $L_{\mathrm{tot}} = 10^{13}L_{\odot}$, the deep survey ($1\,\mathrm{deg}^2$) could detect both AGN and SF lines down to $L_{\mathrm{AGN}}/L_{\mathrm{IR,SF}} \simeq 0.017$ at z=2, $\simeq 0.05$ at $z = 3$ and $\simeq 0.15$ at $z = 5$.

For Eddington-limited accretion, the corresponding BH masses are $\simeq 5.2 \times 10^6\, M_{\odot}$, $\simeq 1.5 \times 10^7\, M_{\odot}$, and $\simeq 4.1 \times 10^7\, M_{\odot}$ at redshifts 2, 3, and 5, respectively. The final total luminosity of these objects is AGN dominated; using the Cai et al. (2013) Eddington ratios, we find that the final BH masses are $3.1 \times 10^8\, M_{\odot}$, $2.6 \times 10^8\, M_{\odot}$ and $1.4 \times 10^8\, M_{\odot}$, respectively. Thus the OST/OSS surveys can detect AGNs when their BH masses are factors ~ 60 (at $z \simeq 2$), ~ 18 (at $z \simeq 3$), and ~ 3 (at $z \simeq 5$) lower than the final values, for a galaxy with a total IR luminosity of $10^{13}\, L_{\odot}$. The probed BH mass growth range will be larger for the rare more luminous sources. Thus, the [OIV]25.89μm line, which is not much affected by dust obscuration, is an efficient tool for taking a census of faint AGNs inside dusty star-forming galaxies.

Existing information on co-evolution is very limited and based on a combination of X-ray and far-IR/sub-mm data. X-ray observations are currently the most efficient way to reliably identify AGNs. However, X-ray emission associated with stellar processes can reach luminosities of $\simeq 10^{42}\,\mathrm{erg\,s^{-1}}$ (cf., e.g., Lehmer et al., 2016), and this makes the identification of lower luminosity AGNs

---

[8] During the dust-enshrouded active SF phase, $L_{\mathrm{IR,SF}}$ is approximately the bolometric luminosity of the galaxy.

[9] Note that the density of these (ultra-)luminous IR galaxies quickly decreases with increasing redshift.
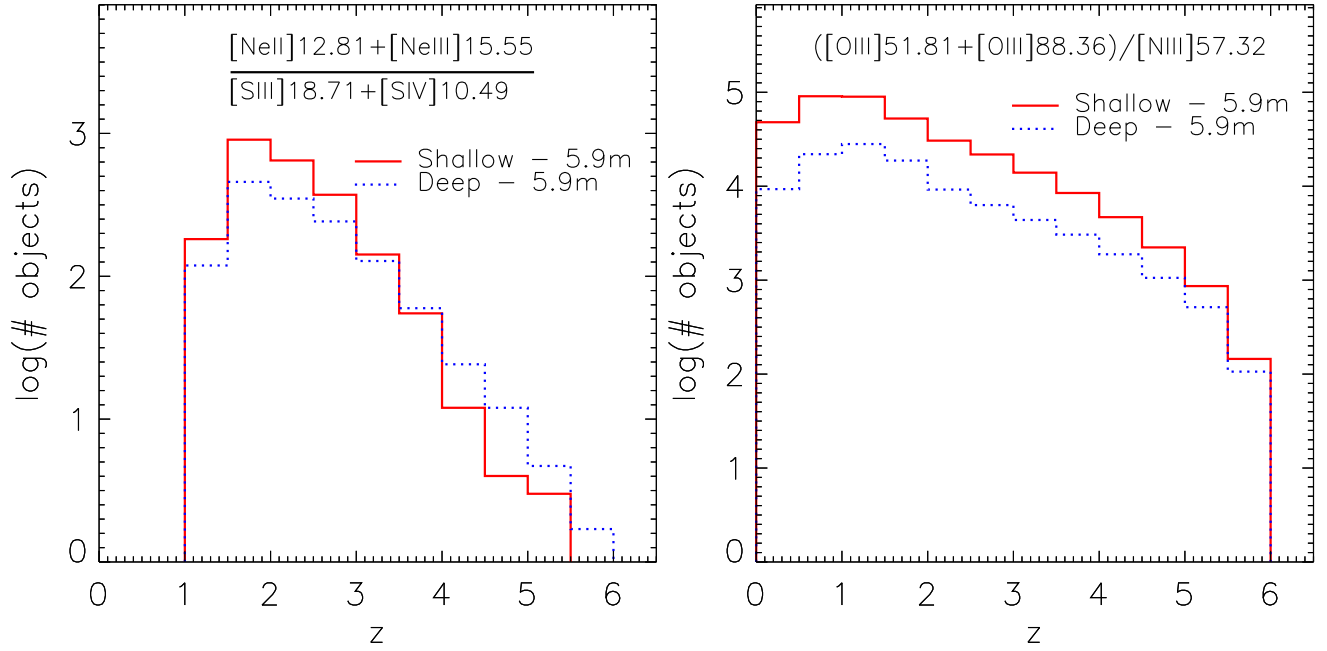
Figure 16. Left panel: predicted redshift distributions of galaxies detected simultaneously in the [NeII] 12.81, [NeIII] 15.55, [SIII] 18.71, and [SIV] 10.49μm lines by the shallow (10 deg^2) and deep (1 deg^2) surveys with a 5.9 m telescope. Both surveys have a total amount of observing time equal to 1000 h. Right panel: the same, but for the simultaneous detection of the [OIII] 51.81, [OIII] 88.36, and [NIII] 57.32μm lines.
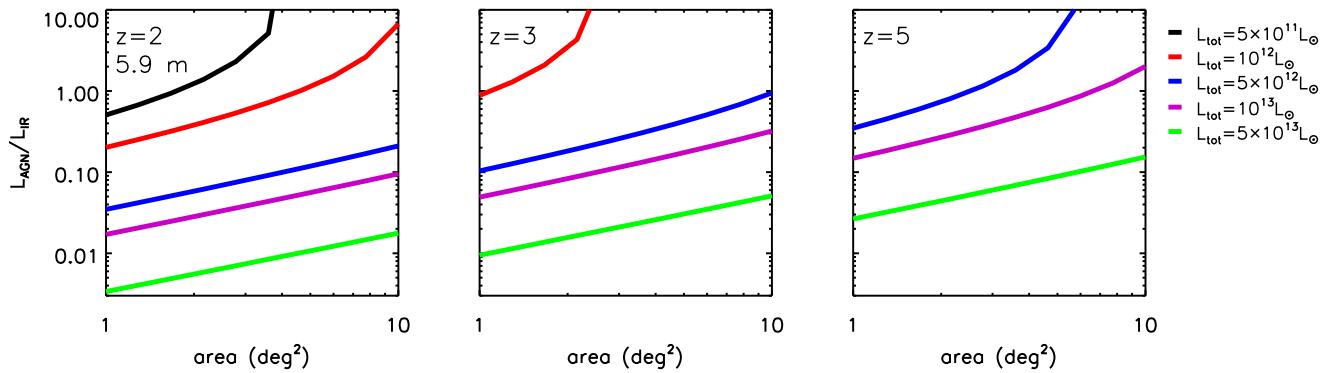

Figure 17. Minimum ratios of AGN bolometric luminosities to SF IR luminosity $L_{\rm IR} \equiv L_{\rm IR,SF}$ for which lines associated with both components (the [OIV] 25.89μm and the [OIII] 88.36μm lines, respectively) are detectable with the OST/OSS at $z = 2$ (left), $z = 3$ (centre), and $z = 5$ (right), as a function of the mapped area for five values of $L_{\rm tot} = L_{\rm AGN} + L_{\rm IR,SF}$.

challenging or impossible. OST observations will not have this limitation; OST will be able to detect less luminous AGNs in the local Universe.

For example, using an X-ray bolometric correction of 20, as appropriate at the luminosities considered (Lusso et al., 2012), we find that an OST (5.9 m) deep survey with OSS can detect AGN line signatures via the [OIV]25.89μm line up to a factor of $\sim$12 fainter at $z \simeq 0.2$. The redshift distribution of galaxies detected simultaneously in at least one SF line and one AGN line (Fig. 15) implies that the OST will detect the AGN component of about 3,500 galaxies at $z \leq 0.2$.

Alternative methods to detect faint AGNs include sub-arcsec mid-IR imaging to isolate the AGN component from the emission due to SF (currently possible only for nearby AGNs; e.g., Asmus et al., 2015), and mid-IR photometry in the wavelength range 3–30 μm (e.g., Assef et al., 2013). The latter approach exploits the fact that the AGN emission is "hotter" than that due to star formation to decompose the observed spectral energy distribution, separating the two components; it works primarily for intrinsically luminous but heavily obscured AGNs.

Current studies of the connection between AGN activity and SF have been carried out using either the X-ray emission of star-forming galaxies (e.g., Delvecchio et al., 2015) or the SF properties of X-ray selected AGNs (e.g., Stanley et al., 2015). In both cases, caveats apply to the interpretation of the results. X-ray detections were found in only a small fraction of the star-forming galaxies selected from *Herschel* surveys. For example, only about 10% of galaxies in the Delvecchio et al. (2015) sample were detected in X-rays. For the overwhelming majority of galaxies, only average AGN accretion rates (or bolometric luminosities) could be determined by stacking. The X-ray selected sample by Stanley et al. (2015) has the opposite problem: only upper limits to the IR luminosity due to SF could be placed for 75.4% of the sources. In both cases, a large fraction of objects have only photometric redshift estimates. Spectroscopic redshifts are available for 53% of the Stanley et al. (2015) sample, while for the one by Delvecchio et al. (2015) the fraction of spectroscopic redshifts is $\simeq$ 60% at $z \leq 1.5$, but drops to 20% at $z \simeq 2$. For comparison, the proposed OST/OSS surveys will detect both SF and AGN spectral signatures for tens of thousands of sources up to $z \sim 5.5 - 6$.

7.2 Evolution of gas metallicity

Detection of the bright PAH bands up to z$\sim$8.5 will allow us, for the first time, to follow their entire history up to the re-ionization epoch, and therefore to explore the properties of the ISM and to address many open questions regarding their formation, destruction and dependence on metallicity.
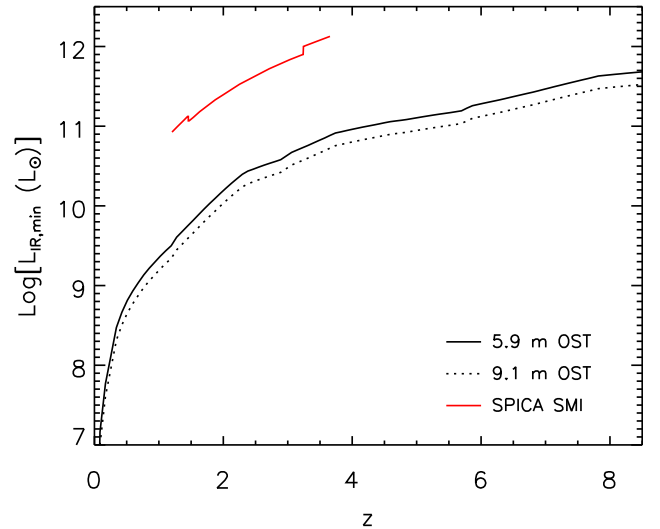

Figure 18. Comparison of the minimum L_{IR} reachable with SPICA/SMI and OST/OSS, as a function of redshift. The upper solid (red) line shows the estimate by Kaneda et al. (2017, lower panel of their Fig. 4) for galaxies with 100% SF detected via the 6.6, 7.7, 8.6 and 11.3 μm PAH features in their "deep survey" (565.5 h of observing time, without overheads, over an area of 1 deg^2). The lower solid (black) line and the dotted line show the corresponding minimum L_{IR} for OST/OSS surveys with a 5.9 m or a 9.1 m telescope, respectively, over the same area and in the same observing time. The difference between the SPICA/SMI and the OST/OSS limiting L_{IR} is actually bigger than shown in this figure because the OST/OSS estimate takes into account the decrease of PAH luminosities with decreasing gas metallicity, while the SPICA/SMI estimate does not.

Most galaxies will be detected by the OST/OSS in at least 2 lines ($\sim$60% of the galaxies observed with the 5.9 m OST shallow survey, $\sim$70% with the deep survey). Therefore, we will have robust spectroscopic redshift measurements. Figure 16 shows that the OST will allow us to measure metallicity in galaxies up to z$\sim$6.

The results presented in this paper pertain to blind surveys. Obviously, the OST/OSS could go deeper by targeting known high-z galaxies, but blind, unbiased surveys are needed to obtain unbiased censuses of the SFR density and the BH accretion rate density.

7.3 Caveats

Although our reference model was validated against a wide variety of data, it necessarily relies on a number of simplifying assumptions and, as any other model, cannot match the enormous variety of individual galaxy and AGN properties.

Recent observations (Dowell et al., 2014; Asboth et al., 2016; Ivison et al., 2016) have yielded estimates of the space density of distant ($z \gtrsim 4$) dusty star-forming galaxies well in excess of model predictions. As shown in the Appendix, the model matches quite well the redshift

distribution of such sources estimated by Ivison et al. (2016), but underestimates by a factor of $\simeq 2.5$ their (highly uncertain) total counts.

The interpretation and even the reality of the excess are still unclear (Béthermin et al., 2017; Donevski et al., 2018). If real, the excess would imply that the OST will detect several times more high-z galaxies. Note that most redshift estimates used to search for high-z galaxies are photometric, based on observations in a few sub-mm bands, and as such are quite uncertain. The OST/OSS will provide fully spectroscopic redshift distributions. These powerful tools can be used to discriminate among models that perform similarly well on higher-level statistics, such as source counts (see, e.g., Fig. 9 of Ivison et al., 2016).

Recent observations of the four strongest CO emitters at $z \simeq 2 - 3$ selected from the literature (Zhang et al., 2018) have provided evidence of an IMF with a higher fraction of massive stars than the Chabrier IMF adopted for our reference model. To what extent would our predictions be affected if this is a general property of high-z star-forming galaxies? The answer is: the effect would be only marginal. This is because our line luminosity functions are derived from IR luminosity functions, which are observationally determined up to $z \simeq 4$ (Gruppioni et al., 2013) and are well reproduced by the model. If we change the IMF, we must also modify other parameters to preserve agreement with the data.

Further uncertainties are associated with the relationships between line and IR or bolometric luminosity. As pointed out in Sect. 4, in several cases there are not enough data to accurately determine the slope of the relation. Data themselves must be taken with caution since, although we have done the selection as carefully as we could, it is possible that some line luminosities attributed to star formation are contaminated by AGN contributions and vice-versa. In Sect. 4, we also noted that data on high-z galaxies frequently refer to strongly lensed galaxies and are therefore prone to errors in the corrections applied for gravitational magnification and differential lensing.

In consideration of the aforementioned caveats, the present results should be taken with a grain of salt. They are a sort of best guess based on state-of-the art understanding.

7.4 Comparison to other instruments

From Fig. 6 we can see that 1000 h of observing time with both OST concepts will yield from a few to several tens of thousands of AGN detections, and from a few to $\simeq 10$ millions of SF detections, depending on the extent of the surveyed area. The detected sources will have broad redshift distributions, extending beyond $z = 8$ in the case of SF galaxies (Figs. 7 and 14) and up to $z \simeq 5.5$ in the case of AGNs (Figs. 8 and 15).

Blind spectroscopic surveys of similar areas (1–10 deg^2) with the low-resolution Mid-infrared Instrument (SMI) on board the SPace IR telescope for Cosmology and Astrophysics (SPICA) have been discussed by Bonato et al. (2015) and, more recently, by Kaneda et al. (2017). These surveys will exploit the large field of view (12 arcmin×10 arcmin), high speed, low-resolution ($R \sim 50 - 120$) spectroscopic and photometric camera, covering the wavelength range $17 - 36\,\mu$m, and are focussed on the detection of PAH features.

According to Kaneda et al. (2017), the planned "wide" survey (covering 10 deg^2 with an observing time of 435 h, excluding overheads) will detect $\simeq 81,600$ galaxies, while the "deep" survey (1 deg^2 with an observing time of 565.5 h, excluding overheads) will detect $\simeq 22,200$ galaxies. Both surveys will attain redshifts somewhat above $z = 4$, bracketing the peak of the cosmic SFR density.

We note that the numbers of SPICA/SMI detections are probably overestimated because they do not take into account the decrease of the PAH luminosity with decreasing metallicity [eq. (3)]. Since our calculations take this effect into account, the comparison with the estimates by Kaneda et al. (2017) is not straightforward. Figure 18 shows that, ignoring this difference, the OST/OSS goes more than one order of magnitude deeper than the SPICA/SMI. Also the OST/OSS can reach L_{IR} below $10^{12}\,L_\odot$ (i.e. SFRs below $\simeq 100 M_\odot\,\text{yr}^{-1}$) up to $z \simeq 8$ while the SPICA/SMI is limited to $z \lesssim 4$.

The SPICA Far Infrared Instrument (SAFARI; Roelfsema et al., 2018) includes a grating spectrograph with low ($R = 300$) and medium ($R \sim 2,000 - 11,000$) resolution observing modes covering the $34 - 230\,\mu$m wavelength range. However, it has a rather narrow FoV and was not designed as a survey instrument. It is expected to detect, in 1000 h of observing time, from some to a few tens of thousands of galaxies and, like the SMI, it will be limited to $z < 5$ (Bonato et al. 2014a,b; Gruppioni et al. 2016). For comparison, the OST/OSS surveys with a 5.9 m telescope will detect a few tens of thousands of galaxies at $z \geq 5$.

Ground-based (sub-)mm telescopes, such as the Atacama Large (sub-)Millimeter Array (ALMA), the NOrthern Extended Millimeter Array (NOEMA), and the Large Millimeter Telescope (LMT) are allowing us to extend the detection of long-wavelength far-IR and sub-mm lines to higher redshifts. Further substantial progress is expected with new instruments, such as CCAT-prime (Stacey et al., 2018) and the Atacama Large-Aperture Submm/mm Telescope (AtLAST; Bertoldi, 2018).

8 SUMMARY

The unprecedented sensitivity of the Origins Space Telescope with the OSS instrument will allow us to address a variety of open issues concerning the evolution of galaxies and AGNs. To illustrate its potential we estimated

the counts of 53 IR spectral lines (9 of them are typical AGN lines). This was done by coupling relationships between line and IR continuum luminosity (or bolometric luminosity in the case of AGNs) with the SFR and luminosity functions by Cai et al. (2013), which accurately match the observational data. We have adopted the relationships found by Bonato et al. (2014a,b, 2015, 2017b) for the SF contribution to the emission of 22 IR fine-structure lines and 6 PAH lines, and for the AGN emission of 27 IR lines. For the SF contribution of 3 lines, the relationships were recalibrated using new data, and new relationships were determined for 10 CO lines, for the SF contribution to 3 lines, and for the AGN contribution to 5 lines. *We took into account the dependence of the starburst lines on gas metallicity.*

We have compared the outcomes of blind surveys with two concepts for the OST, a 5.9 m telescope (Concept 2), and 9.1 m telescope (Concept 1). Our analysis found no compelling scientific case for the larger telescope.

We then investigated in more detail the performance of the 5.9 m telescope option for different surveyed areas, at fixed observing time. The number of line (and source) detections steadily increases with increasing area. Focussing on two extreme cases – surveys of 1 and $10\,\mathrm{deg}^2$, each with an observing time of $1,000\,\mathrm{h}$ – we find that both can detect emission lines associated with star formation (allowing the estimate of the SFR) up to $z \simeq 8.5$ (i.e., all the way through the reionization epoch), and AGN lines (allowing an estimate of the bolometric luminosity) up to $z \simeq 5.5$.

The larger area survey will detect $\sim 8.7 \times 10^6$ lines from $\sim 2.7 \times 10^6$ star-forming galaxies and $\sim 3.8 \times 10^4$ lines from 3.5×10^4 AGNs. It will generally provide better statistics, except at the highest redshifts where, however, the advantage of the deeper survey is modest and the statistics are poor anyway. The deeper survey will do significantly better only for some metallicity diagnostics at $z > 4$, and to investigate the poorly known dependence of the PAH emission (and, more generally, of L_{IR}) on metallicity.

Both surveys will almost completely resolve the cosmic SFR density. The OST with OSS will provide estimates of the IR luminosity function down to below the characteristic luminosity $L_\star$ up to $z \simeq 6$. Thus it will allow us to reconstruct essentially the full SF history, complementing the information from ultra-deep surveys at near-IR wavelengths, which miss heavily dust-enshrouded galaxies. It will also allow us to learn about the evolution of metallicity and the physical conditions in the interstellar medium up to very high redshifts. Note the key role played by the large OST/OSS field of view: we are dealing here with rare sources that cannot be picked up in statistically significant numbers by surveys with the James Webb Space Telescope, which can hardly be expected to carry out deep surveys over areas much larger than $100\,\mathrm{arcmin}^2$ (e.g., Finkelstein et al., 2015).

The model also gave us the probability distributions that an object at redshift z has a starburst IR luminosity $L_{\mathrm{IR,SF}}$ or an AGN bolometric luminosity L_{AGN}, given the total luminosity $L_{\mathrm{tot}} = L_{\mathrm{IR,SF}} + L_{\mathrm{AGN}}$. Our approach to deriving the line luminosity functions allowed us to obtain, at the same time, luminosities of all the lines of both the SF and the AGN components of each simulated object. We could thus estimate for how many objects both components will be detectable, given the survey sensitivity and the number of objects that will be detectable in two or more lines.

This allowed us to investigate the OST/OSS potential to provide insight into galaxy-AGN co-evolution. We found that the OST will be able to probe BH mass growth by large factors, especially, but not only, at moderate redshifts ($z \lesssim 3$) and for luminous sources (star-formation plus AGN luminosity $\gtrsim 10^{13}\,L_\odot$). Spectroscopic surveys with OST will thus be a powerful tool to get a census of faint AGNs inside dusty star-forming galaxies over a broad redshift range.

Compared to SPICA/SMI, which is also expected to carry out blind surveys similar to those discussed here, the OST/OSS will reach, in the same amount of observing time, fluxes more than a factor of 10 fainter and will detect galaxies with L_{IR} below $10^{12}\,L_\odot$ up to $z \simeq 8$, while the SPICA/SMI is limited to $z \lesssim 4$.

On the basis of the existing simulations (Bonaldi et al. 2019) and our estimates, we expect that the OST/OSS will be able to measure redshifts of a substantial fraction of the galaxies detected by the Square Kilometre Array (SKA[10]) deep surveys. The SKA will provide an independent estimate of the SFR in these galaxies, and will also not be affected by obscuration (see Mancuso et al. 2015, e.g. their Fig. 12), but SKA measurements will be susceptible to contamination by nuclear radio activity. Together, OST and SKA will provide a more comprehensive view of the galaxy/AGN evolutionary scenario over a broad frequency range.

9 ACKNOWLEDGEMENTS

---

[10] www.skatelescope.org

support from INAF under PRIN SKA/CTA FORECaST. Z.Y.C. is supported by the National Science Foundation of China (grant No. 11503024). M.N. acknowledges financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 707601.

A MODEL VERSUS DATA

As mentioned in Sect. 1, our reference model was validated against a broad variety of multifrequency data available until 2014. Many additional data have accumulated since then. We have therefore decided to check the model's consistency with these more recent data. Figure 19 shows that the differential counts yielded by the model are in excellent agreement with data from SCUBA 2 and ALMA surveys. As illustrated by the figure, the model includes contributions from different populations (normal and star-bursting late-type galaxies, proto-spheroidal galaxies) and takes into account the effect of gravitational lensing. The contribution of radio sources to the counts is from Tucci et al. (2011).

As mentioned in sub-section 7.3, recent estimates have yielded space densities of $z \gtrsim 4$ dusty star forming galaxies well in excess of model predictions. To check how our model behaves in this respect, we have compared the predicted redshift distribution with the estimate by Ivison et al. (2016), based on a much larger sample than those previously available. To this end, we simulated the selection criteria applied by Ivison et al. (2016): $S_{500\mu m} \geq 30$ mJy and ultra-red colours in the *Herschel*/SPIRE bands, (i.e., $S_{500\mu m}$ larger than both $S_{350\mu m}$ and $S_{250\mu m}$). Moreover, we took into account the uncertainty in their photometric redshifts, $\Delta z/(1 + z) = 0.14$. To the estimate of the total number of galaxies satisfying their selection criteria, Ivison et al. (2016) applied a large correction for incompleteness (a factor of 36.0 ± 8.2); the result is therefore quite uncertain. Thus, we have chosen to compare model predictions with the redshift distribution normalized to the number of objects. As illustrated by Fig. 20, we find good agreement with the shape of the distribution, although the normalization is low by a factor of $\simeq 2.5$.

Furthermore, the model, coupled with the well established correlation between radio and far-IR emission, successfully reproduces the radio luminosity functions of SF galaxies, which is observationally determined up to $z \simeq 5$ (Bonato et al., 2017a).

REFERENCES

Asboth V., et al., 2016, MNRAS, 462, 1989

Asmus D., Gandhi P., Hönig S. F., Smette A., Duschl W. J., 2015, MNRAS, 454, 766

Assef R. J., et al., 2013, ApJ, 772, 26

Aversa R., Lapi A., de Zotti G., Shankar F., Danese L., 2015, ApJ, 810, 74

Berta S., et al., 2013, A&A, 551, A100

Bertoldi F., 2018, in Atacama Large-Aperture Submm/mm Telescope (AtLAST). p. 3, doi:10.5281/zenodo.1158842

Béthermin M., et al., 2017, A&A, 607, A89

Bolatto A. D., Wolfire M., Leroy A. K., 2013, ARA&A, 51, 207

Bonaldi A., Bonato M., Galluzzi V., Harrison I., Massardi M., Kay S., De Zotti G., Brown M. L., 2019, MNRAS, 482, 2

Bonato M., et al., 2014a, MNRAS, 438, 2547

Bonato M., et al., 2014b, MNRAS, 444, 3446

Bonato M., et al., 2015, MNRAS, 452, 356

Bonato M., et al., 2017a, MNRAS, 469, 1912

Bonato M., et al., 2017b, ApJ, 836, 171

Bradford C. M., et al., 2018, in Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave. p. 1069818, doi:10.1117/12.2314049

Brisbin D., Ferkinhoff C., Nikola T., Parshley S., Stacey G. J., Spoon H., Hailey-Dunsheath S., Verma A., 2015, ApJ, 799, 13

Cai Z.-Y., et al., 2013, ApJ, 768, 21

Carilli C. L., Walter F., 2013, ARA&A, 51, 105

Carniani S., et al., 2015, A&A, 584, A78

Casey C. M., et al., 2018, ApJ, 862, 77

Chabrier G., 2003, PASP, 115, 763

Chen C.-T. J., et al., 2013, ApJ, 773, 3

Coppin K., et al., 2006, MNRAS, 372, 1621

Croom S. M., et al., 2009, MNRAS, 399, 1755

De Looze I., et al., 2014, A&A, 571, A69

Delvecchio I., et al., 2014, MNRAS, 439, 2736

Delvecchio I., et al., 2015, MNRAS, 449, 373

Desai V., et al., 2007, ApJ, 669, 810

Donevski D., et al., 2018, A&A, 614, A33

Dowell C. D., et al., 2014, ApJ, 780, 75

Engelbracht C. W., Gordon K. D., Rieke G. H., Werner M. W., et al. 2005, ApJ, 628, L29

Engelbracht C. W., Rieke G. H., Gordon K. D., Smith J.-D. T., Werner M. W., et al. 2008, ApJ, 678, 804

Farrah D., et al., 2013, ApJ, 776, 38

Ferkinhoff C., et al., 2014, ApJ, 780, 142

Fernández-Ontiveros J. A., Spinoglio L., Pereira-Santaella M., Malkan M. A., Andreani P., Dasyra K. M., 2016, ApJS, 226, 19

Fernández-Ontiveros J. A., et al., 2017, PASA, 34, e053

Ferrarese L., Ford H., 2005, Space Sci. Rev., 116, 523

Finkelstein S. L., Dunlop J., Le Fevre O., Wilkins S., 2015, preprint, (arXiv:1512.04530)

Förster Schreiber N. M., Roussel H., Sauvage M., Charmandaris V., 2004, A&A, 419, 501
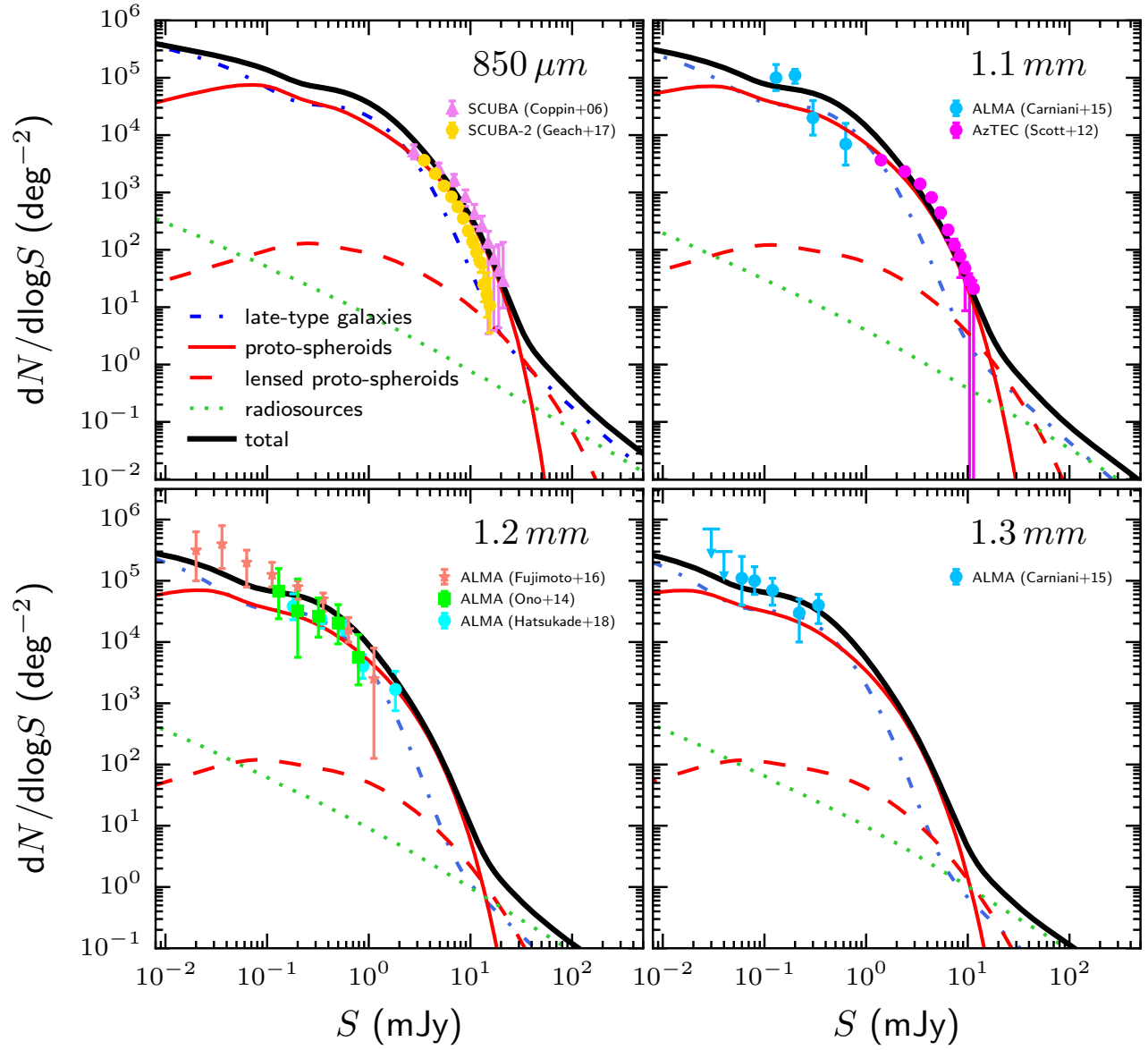
Fujimoto S., Ouchi M., Ono Y., Shibuya T., Ishigaki M.,

Figure 19. Comparison of the predictions of the Cai et al. (2013) model with differential counts from recent SCUBA 2 and ALMA surveys. The data are from Coppin et al. (2006), Scott et al. (2012), Ono et al. (2014), Carniani et al. (2015), Fujimoto et al. (2016) and Hatsukade et al. (2018).
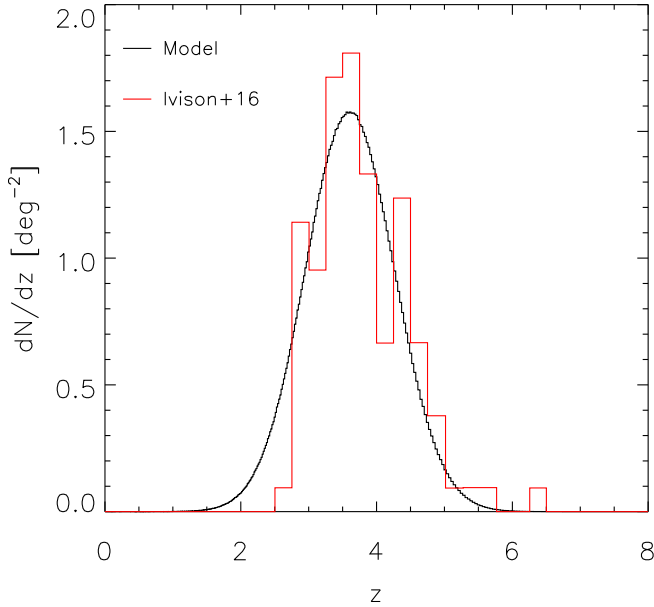
Figure 20. Comparison between the Ivison et al. (2016) redshift distribution and the theoretical one derived from the Cai et al. (2013) model.

Nagai H., Momose R., 2016, ApJS, 222, 1

Galametz M., Madden S., Galliano F., Hony S., Schuller F., et al. 2009, A&A, 508, 645

Galliano F., Madden S. C., Jones A. P., Wilson C. D., Bernard J.-P., Le Peintre F., 2003, A&A, 407, 159

Genzel R., et al., 2012, ApJ, 746, 69

George R. D., 2015, PhD thesis, University of Edinburgh

González-Alfonso E., et al., 2017, PASA, 34, e054

Goulding A. D., Alexander D. M., 2009, MNRAS, 398, 1165

Graciá-Carpio J., García-Burillo S., Planesas P., Fuente A., Usero A., 2008, A&A, 479, 703

Gruppioni C., et al., 2013, MNRAS, 432, 23

Gruppioni C., et al., 2016, MNRAS, 458, 4297

Gullberg B., et al., 2015, MNRAS, 449, 2883

Guo Q., et al., 2016, MNRAS, 461, 3457

Hasinger G., 2008, A&A, 490, 905

Hatsukade B., et al., 2018, PASJ,

Herrera-Camus R., et al., 2016, ApJ, 826, 175

Hezaveh Y. D., et al., 2013, ApJ, 767, 132

Hunt L. K., et al., 2015, A&A, 583, A114

Hunt L., Dayal P., Magrini L., Ferrara A., 2016, MNRAS, 463, 2002

Ivison R. J., et al., 2016, ApJ, 832, 78

Kamenetzky J., Rangwala N., Glenn J., Maloney P. R., Conley A., 2016, ApJ, 829, 93

Kaneda H., et al., 2017, PASA, 34, e059

Kennicutt Jr. R. C., 1998, ApJ, 498, 541

Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531

Kirkpatrick A., Pope A., Sajina A., Roebuck E., Yan

L., Armus L., Díaz-Santos T., Stierwalt S., 2015, ApJ, 814, 9

Kormendy J., Ho L. C., 2013, ARA&A, 51, 511

Kroupa P., Weidner C., 2003, ApJ, 598, 1076

Lacey C. G., et al., 2016, MNRAS, 462, 3854

Lehmer B. D., et al., 2016, ApJ, 825, 7

Leisawitz D., et al., 2018, in Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave. p. 1069815, doi:10.1117/12.2313823

Lu N., et al., 2017, ApJS, 230, 1

Lusso E., et al., 2012, MNRAS, 425, 623

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Madden S. C., Galliano F., Jones A. P., Sauvage M., 2006, A&A, 446, 877

Magdis G. E., et al., 2014, ApJ, 796, 63

Mancuso C., et al., 2015, ApJ, 810, 72

Meléndez M., et al., 2008, ApJ, 682, 94

Merloni A., Heinz S., 2008, MNRAS, 388, 1011

Nagao T., Maiolino R., Marconi A., Matsuhara H., 2011, A&A, 526, A149

O'Halloran B., Satyapal S., Dudik R. P., 2006, ApJ, 641, 795

Ono Y., Ouchi M., Kurono Y., Momose R., 2014, ApJ, 795, 5

Oteo I., et al., 2016, ApJ, 827, 34

Peeters E., Spoon H. W. W., Tielens A. G. G. M., 2004, ApJ, 613, 986

Pereira-Santaella M., Rigopoulou D., Farrah D., Lebouteiller V., Li J., 2017, MNRAS, 470, 1218

Planck Collaboration XIII 2016, A&A, 594, A13

Roelfsema P. R., et al., 2018, PASA, 35, e030

Rosenberg J. L., Wu Y., Le Floc'h E., Charmandaris V., Ashby M. L. N., Houck J. R., Salzer J. J., Willner S. P., 2008, ApJ, 674, 814

Rosenberg M. J. F., et al., 2015, ApJ, 801, 72

Roussel H., Sauvage M., Vigroux L., Bosma A., 2001, A&A, 372, 427

Rubin R. H., 1989, ApJS, 69, 897

Schaerer D., et al., 2015, A&A, 576, L2

Scott K. S., et al., 2012, MNRAS, 423, 575

Serjeant S., 2012, MNRAS, 424, 2429

Shipley H. V., Papovich C., Rieke G. H., Brown M. J. I., Moustakas J., 2016, ApJ, 818, 60

Smith J. D. T., Draine B. T., Dale D. A., Moustakas J., Kennicutt Jr. R. C., Helou G., et al. 2007, ApJ, 656, 770

Spilker J. S., et al., 2016, ApJ, 826, 112

Spinoglio L., Malkan M. A., 1992, ApJ, 399, 504

Stacey G. J., Hailey-Dunsheath S., Ferkinhoff C., Nikola T., Parshley S. C., Benford D. J., Staguhn J. G., Fiolet N., 2010, ApJ, 724, 957

Stacey G. J., et al., 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series.

p. 107001M, doi:10.1117/12.2314031

Stanley F., Harrison C. M., Alexander D. M., Swinbank A. M., Aird J. A., Del Moro A., Hickox R. C., Mullaney J. R., 2015, MNRAS, 453, 591

Sturm E., Lutz D., Verma A., Netzer H., Sternberg A., Moorwood A. F. M., Oliva E., Genzel R., 2002, A&A, 393, 821

Tremonti C. A., et al., 2004, ApJ, 613, 898

Tucci M., Toffolatti L., de Zotti G., Martínez-González E., 2011, A&A, 533, A57

Voit G. M., 1992a, MNRAS, 258, 841

Voit G. M., 1992b, ApJ, 399, 495

Walter F., Weiß A., Downes D., Decarli R., Henkel C., 2011, ApJ, 730, 18

Weaver K. A., et al., 2010, ApJ, 716, 1151

Yun M. S., et al., 2015, MNRAS, 454, 3485

Zahid H. J., Geller M. J., Kewley L. J., Hwang H. S., Fabricant D. G., Kurtz M. J., 2013, ApJ, 771, L19

Zhang Z.-Y., Romano D., Ivison R. J., Papadopoulos P. P., Matteucci F., 2018, Nature, 558, 260

Zhao Y., et al., 2013, ApJ, 765, L13

Zhao Y., et al., 2016, ApJ, 819, 69