MNRAS **000**, 000–000 (0000) Preprint 3 September 2019 Compiled using MNRAS LATEX style file v3.0

HORIZON-AGN virtual observatory – 2. Supplementary On-line Material.

I. Davidzon, C. Laigle, P. L. Capak, O. Ilbert, D. C. Masters, S. Hemmati, N. Apostolakos, J. Coupon, S. de la Torre, J. Devriendt, Y. Dubois, D. Kashino, S. Paltani, C. Pichon

May 2019

APPENDIX A: CAVEATS IN THE SOM OF HORIZON-AGN GALAXIES

A1 Resolution and subgrid recipes in HORIZON-AGN

As mentioned in Section 3.3, our modelling of HORIZON-AGN photometry is limited both by the resolution of the simulation and the accuracy of the recipes implemented at the subgrid scale. The spatial resolution of HORIZON-AGN simulation is at best 1 pkpc, and the mass resolution is $\sim 8\times10^7$ M$_\odot$ for dark matter particles. In Dubois et al. (2014) this translates into a lower limit of 10^8 M$_\odot$ in galaxy stellar mass, while we stop at $M = 10^9$ M$_\odot$ to be conservative. Our galaxies are therefore resolved by at least ~ 500 particles. However, in despite of the SFH of each galaxy being well sampled, the scale limit hinder us from accounting for radiative transfer and for the impact of the turbulence on star-formation. It will naturally smooth out clumpiness in the ISM and therefore in the SFH themselves. One can expect that intrinsic SFH and metallicity enrichment histories of our simulated galaxies are less diverse at low than high resolution (see e.g. Lagos et al. 2016, for a study on the limited effect of resolution on the gas mass fraction).

A2 Modelling of the photometry

Beyond these intrinsic limitations of the simulation, we made also several simplifying assumptions while computing galaxy photometry in post-processing. These shortcomings are detailed in Appendix A.5 of Paper I and listed here for completeness. We assume a single and spatially constant IMF and stellar mass-loss prescription as in BC03. Dust distribution is assumed to follow the gas metal distribution, and we take a dust-to-metal ratio constant in time and space. For all these reasons, the simulated photometry naturally presents less variety than the observed one, and the SOM performance that we derive must be considered as optimal estimates.

Besides that, nebular emission is not taken into account when mimicking the COSMOS photometry. Emission lines within the wavelength range of a broad-band filter can boost the measured flux and alter galaxy colours. However, nebular emission should not impair the SOM estimator as flux contamination is proportional to galaxy star formation activity. The resulting SOM would have a different appearance because of the modified colours, but without loosing the capacity to classify galaxy SEDs. Both the 10^9 M$_\odot$ lower limit and the lack of emission lines are tested by means of an empirical simulation built independently of HORIZON-AGN. In this simulation we fit BC03 models to COSMOS2015 galaxies, fixing the redshift to the $z_{\rm phot}$ value provided there. We integrate each best-fit BC03 spectrum within the same filters used in the present work, adding realistic noise to get a photometric catalogue similar to the observed one. The BC03 spectra include UV-optical emission lines (e.g., Lyα, [OII]λ3727, Hβ, [OIII]$\lambda\lambda$4959,5007, Hα) according to Schaerer & de Barros (2009) recipe. Owing to the deep IR data, the COSMOS2015 catalogue includes numerous objects with $M < 10^9$ M$_\odot$ (see fig. 17 in Laigle et al. 2016) and so does this replica. We produce the SOM of the empirical simulation using the empirically simulated galaxies between $M = 10^8$ and 10^{12} M$_\odot$, then we apply our method to calculate their SFR. The resulting values are in good agreement with the intrinsic SFR (Fig. A1). Despite the simplistic approach (e.g., BC03 models assume *tau* or delayed-*tau* SFHs) this test indicates that the SOM estimator should work well also for a survey encompassing a larger stellar mass range and whose galaxy photometry is contaminated by nebular emission.

We also test the i^+ flux normalisation used in the computation of stellar mass and SFR from the SOM (e.g., Fig. 5 in the main text). Since galaxy SEDs in the same cell have similar shape, the normalisation factor (see Eq. 4 in the main text) does not change significantly if another band is used (with the exception of the bluest filters). Fig. A2 shows this is the case for a $K_{\rm s}$ flux normalisation. In the main text we rely on the i^+ because it is has a higher S/N whereas photometric errors in $K_{\rm s}$ would produce additional scatter in the SFR$_{\rm SOM}$ vs SFR$_{\rm sim}$ comparison (cf. Fig. A2 with Fig. 9 in the main text).

APPENDIX B: TRAINING AND CALIBRATION GALAXY SAMPLES

B1 Training sample selection

As mentioned in Sect. 4.1, the selection of COSMOS-like galaxies with $S/N > 1.5$ may bias the results. First of all, we note that such a selection corresponds in practice to a magnitude cut $i^+ < 25$ (see Fig. B1, upper panel) with the removal (because of the required u-band detection) of $z > 3.5$ galaxies (Fig. B1, lower panel). Even when working in the narrower range $0 < z < 3.5$, the SOM would still be biased if any galaxy type were missing from the training sample. To investigate this potential issue we consider the SED classes (i.e., the cells) defined in the initial SOM (shown in Fig. 3 of the main text) and verify whether they are well-represented after the S/N cut. We find that most of those classes are still included with a fairly good statistics (Fig. B2). The occupation is larger than 10 galaxies per cell in most of them. Nearly 10 per cent of the cells, predominantly at $z > 3.2$, have less than 5 objects inside; these significant lack of high-z galaxies introduces the $z_{\rm SOM}$ bias shown in Fig. 8 of the main text.

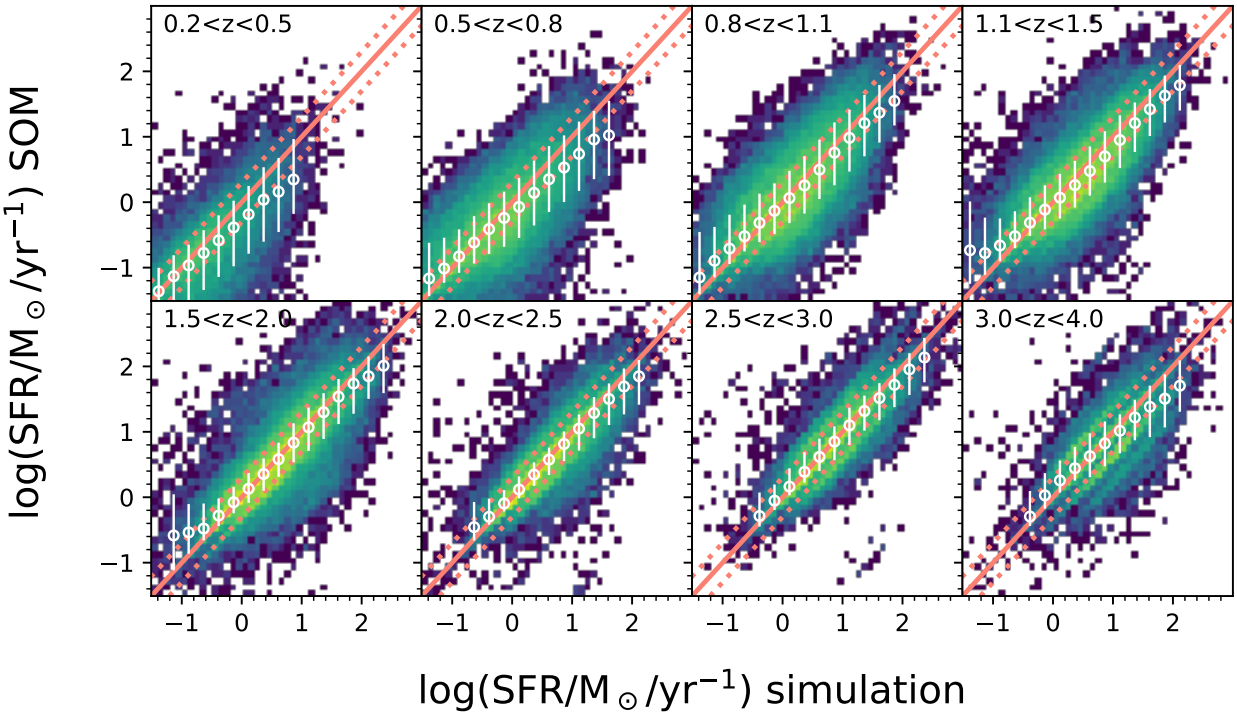

Figure A1. Comparison between the intrinsic SFR and the estimates derived from the SOM for 400,000 mock galaxies with $10^8 < M/M_\odot < 10^{12}$ and $0 < z < 4$. These objects are selected from an empirical simulation that reproduces the observed COSMOS2015 galaxies by means of their best-fit templates. Each panel in the figure show the comparison in a different redshift bin, as indicated. A solid line marks the 1:1 bisector while dotted lines are offset by ±0.15 dex from it. Empty circles show the median in running bins of $\mathrm{SFR_{sim}}$ with error bars derived from the 16th−84th percentile range.

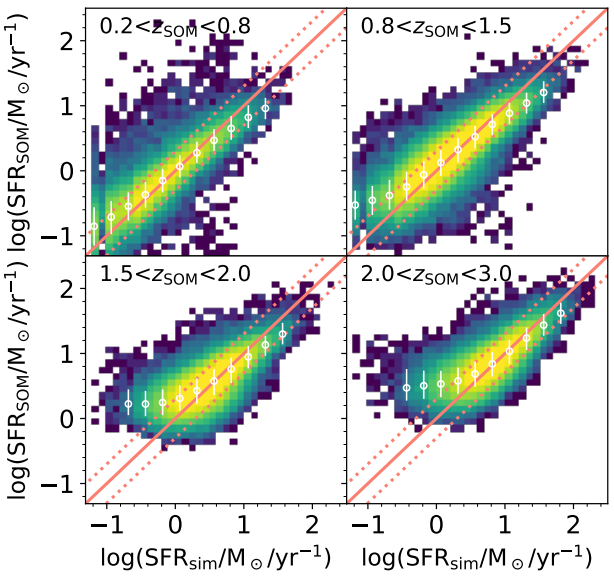

Figure A2. $\mathrm{SFR_{SOM}}$ vs $\mathrm{SFR_{sim}}$ comparison, similar to Fig. 9 (see main text) but normalising the SEDs as a function of their K_s flux instead of i^+.

B2 Calibration sample C1

Another step in the SOM method is to label its cells using a sample of *bona fide* galaxies. Sect. 5 presented two possible ways of building such a sample with future telescopes.

The calibration sample C1 relies on the unparalleled multiplexing that 4MOST and MOONS will offer (placing 10^2 to 10^3 fibers simultaneously) in order to observe the several hundreds of *bona fide* galaxies needed to label the SOM cells. Both spectrographs will start operations in 2020−2022, at the ESO telescopes VISTA and VLT respectively. Together they shall enable for Hα detection in thousands of galaxies between $z = 0$ and ∼1.7, which is the highest redshift that MOONS can reach before H_α moves outside its wavelength window (which ends at 1.8 μm). 4MOST will observe H_α emission up to $z \sim 0.4$, i.e. one sixth of the SOM map (see the redshift map in Fig. 7). According to its expected sensitivity, 4MOST should provide $S/N > 6$ spectra for galaxies brighter than 20 mag, in 2h of exposure with "grey sky" conditions[1]. The spectral resolution ranges from ∼4,000 in the bluest band, up to 7,700. The MOONS Consortium predicts that the sensitivity of their instrument shall reach $S/N = 5$ for a line flux of 0.6×10^{-17} erg s^{-1} cm^{-1}, in 1h exposure[2].

Concerning the "observational bias" introduced by C1, a major concern is the survey strategy of selecting targets among the brightest galaxies in each cell, i.e., objects brighter than $\langle i^+ \rangle^\mathrm{cell}$ to enhance the probability of a $S/N > 10$ line detection. The dominant source of error is the aperture correction, which in future surveys

[1] Sensitivity per resolution element, converted from the S/N per Å provided by the 4MOST team at https://www.4most.eu/cms/facility/capabilities/. Sensitivity for emission lines is not provided.

[2] Predicted S/N per resolution element including background subtraction, assuming an emission line with FWHM=200 km s^{-1} (see https://vltmoons.org/ and Taylor et al. 2018)

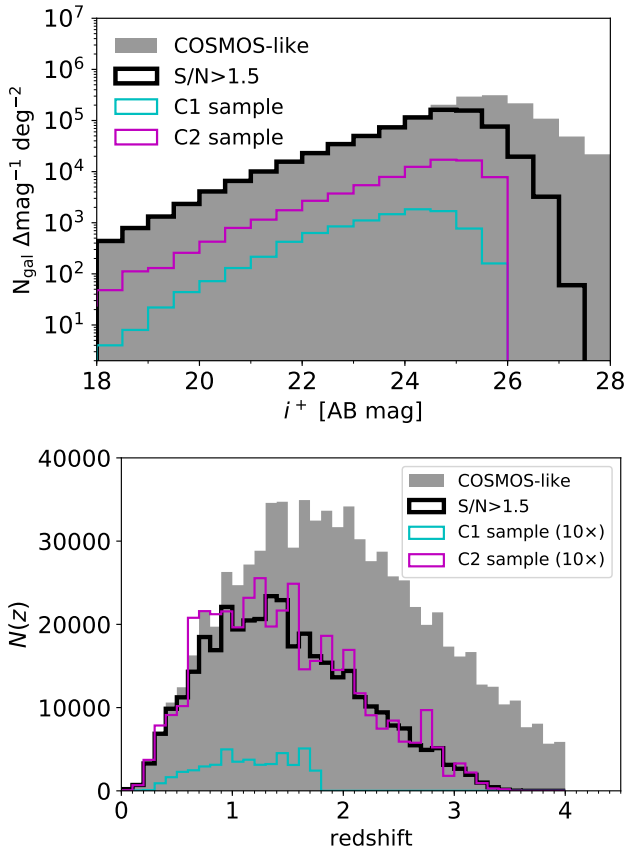

Figure B1. Magnitude and redshift distribution (*upper* and *lower* panel respectively) of Horizon-AGN galaxies with COSMOS-like uncertainties in their photometry. Both panels show the distribution of the parent sample (filled histogram), the S/N-selected sample (black line, see Sect. 4.1), and the objects extracted to label the SOM in the C1 (cyan) and C2 (magenta) scenarios (see Sect. 5.1 for more detail about the two pseudo-surveys). In the bottom panel, number counts of C1 and C2 sub-samples are boosted by a factor 10 to improve readability.

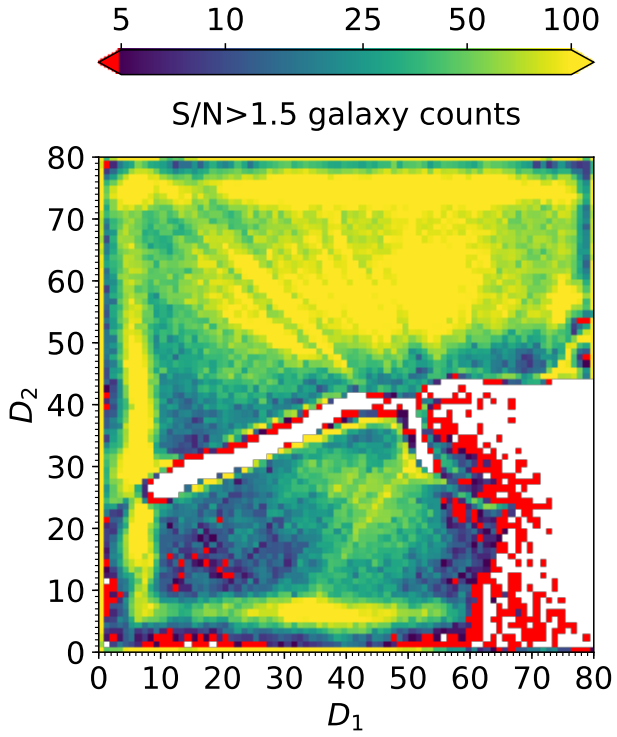

Figure B2. The Horizon-AGN SOM (same as Fig. 3) colour-coded according to the number of galaxies per cell with $S/N > 1.5$ in each band (white pixels indicate empty cells).

may be mitigated by larger slits[3] or by overlapping with grism or IFU spectroscopy (HST, *Euclid*, JWST) that would allow for an object-by-object correction. We do not attempt to model systematic uncertainties related e.g. to the dust prescription (see Eq. 6) or to the conversion factor between nebular and stellar extinction (see Kashino et al. 2013). Fig. 11 illustrates the solidity of manifold learning: despite the stochastic SFR_{cal} fluctuations due to the random selection of these galaxies, the SOM labels still show a clear evolutionary pattern.

B3 Calibration sample C2

In the case of C2, a $19' \times 19'$ patch of UV and FIR pseudo-observations is chosen to be the S-W corner of the Horizon-AGN lightcone and supposed to be observed in UV and FIR. Final results are preserved if we change the location of the patch. Compared to the previous calibration, this one has the advantage that more *bona fide* galaxies are observed in each cell, averaging the intrinsic variance in their star formation activity. Moreover, the redshift range is larger (see B1, bottom panel).

As the SFR_{cal} values come from UV+IR energy balance (Eq. 7 in the main text) precise redshifts for the *bona fide* galaxies should be gathered to compute rest frame luminosities. This apparently implies an additional follow-up to get spectroscopic redshifts (see discussion in Sect. 5.1). However, following the parallelism between the Horizon-AGN lightcone and COSMOS2015, we can assume that in our field there is already an extended number of them. In the real COSMOS field these z_{spec} come from two decades of spectroscopic campaigns, most notably the ongoing survey for a Complete Calibration of the Color-Redshift Relation (C3R2, Masters et al. 2017). The C3R2 programme is systematically collecting spectroscopic redshifts in each cell of the COSMOS SOM (already covering ~75 per cent of it, Masters et al. 2019) building on previous surveys as zCOSMOS (Lilly et al. 2007), VUDS (Le Fevre et al. 2015), Keck/DEIMOS (Hasinger et al. 2018).

In a traditional analysis a key concern in the design of C2 would be cosmic variance uncertainty, hereafter referred as "sample variance"[4]. This is due to the fact that the necessary UV and FIR photometry is taken from deep pencil-beam surveys covering only 0.1 deg^2 (10% of the total area of the Horizon-AGN lightcone). The relatively small area means that specific types of objects may

[3] Although this is not the case of either 4MOST or MOONS, which will have fibers with a similar diameter of Subaru FMOS.

[4] Although the term less popular in the literature (see Moster et al. 2011) *sample variance* is more advisable to indicate the uncertainty due to large-scale clustering which changes the observed density of a given type of object. Strictly speaking, *cosmic variance* describes the intrinsic limitation of cosmological experiments, which cannot be reproduced in other observable universes.

be over-represented and dominate the measured density. Given the quality of our results, sample variance does not seem to dramatically impair our method. Nonetheless, it might bias the SOM in a subtle way. An assessment of sample variance uncertainties for a given survey design is always recommended, but difficult to pursue because ideally it requires several independent realisations of the survey itself.

Buchs et al. (2019) does an extensive analysis of how sample variance would affect mapping functions like the one proposed here, and how to correct for it. Their analysis relies on multiple realisations of a Gpc-scale numerical simulation (although limited to $z < 1.5$). They find that sample variance is a major contributor to the error budget, accounting for 3–5 times the shot-noise uncertainty when a calibration sample covers an area of 1 deg^2 (see their figure 4). However, if the parent photometric data set is observed over a large enough area, the variance in the calibration field can be determined and corrected for. So, in principle sample variance is much less of an issue with the SOM method. However, we are limited by the box size of Horizon-AGN and the $19' \times 19'$ sub-region we have selected still corresponds to a small cosmic volume in which some galaxy types may be under-represented and others over-abundant respect to the rest of the sample. For example if a dark matter over-density is present along the line of sight of the pseudo-survey, then a higher fraction of *bona fide* galaxies at that redshift will be "red and dead" and perhaps cells of star-forming objects will not be filled. The bottom panels of Fig. 11 (see main text) shows this concept: even though C2 galaxies span the whole redshift range, a few isolated cells (or small groups of cells) are empty. On the other hand, C1 galaxies are dispersed in a much larger volume and are likely to live in a broader variety of environments (besides the fact they fill a contiguous area of the SOM by construction). A certain degree of sample variance can be appreciated also in the redshift and magnitude distributions shown in Fig. B1. We also note that sample variance may be misidentified with other survey selection effects: for instance the SFR map from C2 calibration differs from C1 not for the smaller area, but mainly because in the latter the "spectroscopic" $F_{H\alpha}$ threshold favourites the targeting of more actively star-forming galaxies.

REFERENCES

Buchs R., et al., 2019, arXiv e-prints,
Dubois Y., et al., 2014, MNRAS, 444, 1453
Hasinger G., et al., 2018, ApJ, 858, 77
Kashino D., et al., 2013, ApJ, 777, L8
Lagos C. d. P., et al., 2016, MNRAS, 459, 2632
Laigle C., et al., 2016, ApJS, 224, 24
Le Fevre O., et al., 2015, A&A, 576, A79
Lilly S. J., et al., 2007, ApJS, 172, 70
Masters D. C., Stern D. K., Cohen J. G., Capak P. L., Rhodes J. D., Castander F. J., Paltani S., 2017, ApJ, 841, 111
Masters D. C., et al., 2019, ApJ, in press
Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, ApJ, 731, 113
Schaerer D., de Barros S., 2009, A&A, 502, 423
Taylor W., et al., 2018, in Ground-based and Airborne Instrumentation for Astronomy VII. p. 107021G, doi:10.1117/12.2313403

This paper has been typeset from a TEX/LATEX file prepared by the author.