MNRAS **000**, 1–22 (2019) Preprint 3 June 2019 Compiled using MNRAS LATEX style file v3.0

HORIZON-AGN virtual observatory - 2. Template-free estimates of galaxy properties from colours.

I. Davidzon[1⋆], C. Laigle[2], P. L. Capak[1,3], O. Ilbert[4], D. C. Masters[5], S. Hemmati[5], N. Apostolakos[6], J. Coupon[6], S. de la Torre[4], J. Devriendt[2], Y. Dubois[7], D. Kashino[8], S. Paltani[6], C. Pichon[7,9]

[1] *IPAC, Mail Code 314-6, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA*
[2] *Sub-department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*
[3] *Cosmic Dawn centre (DAWN), Niels Bohr Institute, University of Copenhagen, Juliane Maries vej 30, DK-2100 Copenhagen, Denmark*
[4] *Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France*
[5] *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA*
[6] *Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290, Versoix, Switzerland*
[7] *Sorbonne Universités, CNRS, UMR 7095, Institut dÁstrophysique de Paris, 98 bis bd Arago, 75014 Paris, France*
[8] *Department of Physics, ETH Zürich, Wolfgang-Pauli-strasse 27, CH-8093 Zürich, Switzerland*
[9] *Korea Institute for Advanced Study (KIAS), 85 Hoegiro, Dongdaemun-gu, Seoul, 02455, Republic of Korea*

Accepted –. Received –; in original form –

ABSTRACT

Using the HORIZON-AGN hydrodynamical simulation and self-organising maps (SOMs), we show how to compress the complex, high-dimensional data structure of a simulation into a 2-d grid, which greatly facilitates the analysis of how galaxy observables are connected to intrinsic properties. We first verify the tight correlation between the observed 0.3–5μm broad-band colours of HORIZON-AGN galaxies and their high-resolution spectra. The correlation is found to extend to physical properties such as redshift, stellar mass, and star formation rate (SFR). This direct mapping from colour to physical parameter space is shown to work also after including photometric uncertainties that mimic the COSMOS survey. We then label the SOM grid with a simulated calibration sample, and estimate redshift and SFR for COSMOS-like galaxies up to $z \sim 3$. In comparison to state-of-the-art techniques based on synthetic templates, our method is comparable in performance but less biased at estimating redshifts, and significantly better at predicting SFRs. In particular our "data-driven" approach, in contrast to model libraries, intrinsically allows for the complexity of galaxy formation and can handle sample biases. We advocate that obtaining the calibration for this method should be one of the goals of next-generation galaxy surveys.

Key words: galaxies: formation,evolution – methods: statistical,data analysis

1 INTRODUCTION

One of the most successful techniques to understand galaxy formation is measuring galaxy properties in large-area surveys and comparing the results with cosmological-scale simulations based on theoretical models of galaxy formation. Typically the comparison is done in physical parameter space, so secure estimates of redshift, luminosity (L), stellar mass (M), star formation rate (SFR), must be obtained from observational data. These estimates usually come from the analysis of the spectral energy distribution (SED) or the high-resolution spectrum of galaxies, relying on the correlation between specific wavelengths and physical property: for example H_α emission and star formation (Kennicutt 1998) or $\sim 2\mu$m light and stellar mass (Madau et al. 1998). In the past two decades data from multi-wavelength photometry and spectroscopic surveys have become abundant, so fitting galaxy templates to observed SED and spectra has become the standard method to perform this analysis (among the pioneering studies: Arnouts et al. 1999; Bell & de Jong 2000; Bolzonella et al. 2000; Gavazzi et al. 2002; Pérez-González et al. 2003; Fontana et al. 2004; Gallazzi et al. 2005). Galaxy parameters are usually derived from the maximum-likelihood template (Bolzonella et al. 2000)

⋆ E-mail: iary@caltech.edu

or from the full probability distribution function (PDF) of the template set (Benítez 2000). For quantities like stellar mass or star formation history (SFH), templates are built from stellar population synthesis models (e.g., Bruzual & Charlot 2003; Maraston 2005; Conroy et al. 2009).

The modelling process often introduces several systematic effects that have been shown to severely bias $\boldsymbol{M}$ and SFR estimates in some cases (e.g. Mitchell et al. 2013; Mobasher et al. 2015; Laigle et al. 2019). For instance, the synthetic templates are not guaranteed to have fully realistic features: e.g., their SFH is often an analytically function (as the τ- and inverted-τ models, see Maraston et al. 2010) that does not include either multiple bursts or chemical enrichment. Discrepancy between synthetic templates and real galaxies is also due to the assumptions about their stellar population initial mass function (IMF) and dust attenuation of their stellar light (e.g., Davidzon et al. 2013). Moreover, templates not always take into account nebular emission lines and their impact on the observed broad band photometric colours. Finally, some galaxy types may not even be included in the library (e.g. the old and dusty galaxies at $z \sim 2{-}3$ discussed in Marchesini et al. 2010).

Beyond modelling problems, there are additional systematics introduced by the fitting procedure. The relative abundance of a given SED in the real universe is often not accounted in the synthetic library: most of the fitting codes assume all templates are equally likely. Moreover, the SED (or spectral) fitting algorithms may not treat the template set in an optimal way, using computationally expensive brute-force approaches to explore the entire template library (see Speagle et al. 2016) or introducing systematics when convolving them with instrumental errors (see Cappellari 2017). These uncertainties all propagate into the commonly used statistical descriptors of the galaxy census, such as the galaxy stellar mass function (Ilbert et al. 2013; Grazian et al. 2015) and the specific SFR evolution of star forming galaxies (Santini et al. 2017; Davidzon et al. 2018). Ultimately, all of these uncertainties combine to result in a biased view of galaxy demographics, preventing a clear and straightforward comparison between observations and simulations.

To make such a comparison more robust, significant effort has been devoted to improving template fitting techniques. To date, substantial progress has been achieved in each step of the computation, from the construction of galaxy models with more complex star formation histories (Pacifici et al. 2013) to improved radiative transfer modes including the interstellar medium (da Cunha et al. 2008) and sophisticated Bayesian fitting techniques (Leja et al. 2017; Chevallard & Charlot 2016).

In parallel to this ongoing effort other authors have explored alternate paths, replacing standard template fitting with new techniques based on machine learning (ML). Besides implicitly accounting for biases, a key advantage of ML techniques is indeed their speed, which enables analyses of extremely large data sets. Most of the existing work aims at estimating redshifts (see Salvato et al. 2019 and references therein), with the exception of a few publications where ML has been applied e.g. to recover SFR (Delli Veneri et al. 2019) and specific SFR (Stensbo-Smidt et al. 2017) of $z \sim 0$ galaxies from the Sloan Digital Sky Survey (SDSS). A wider range of galaxy parameters (including stellar mass and metallicity) is estimated in Simet et al. (2019). That study, complementary to ours, use a *supervised* neural network trained with a semi-analytic model simulation.

In the present work we describe a novel technique based on *unsupervised* ML combined with analytic data modelling, to simultaneously provide redshift and SFR estimates for galaxies across a large redshift range ($0 < z \lesssim 3$, spanning about 12 Gyr of universe's life). The ML algorithm adopted here is the self-organising map (SOM, Kohonen 1981), which is an unsupervised manifold learning algorithm used to analyse high-dimensional data (see also Kohonen 2001). Initially popular in engineering research, it soon circulated to many other fields including Astrophysics. Seminal work has used the SOM mainly to classify astronomical objects and their properties, including stellar populations (Hernandez-Pajares & Floris 1994), star vs galaxy separation (Miller & Coe 1996; Maehoenen & Hakala 1995), and morphological types (Molinari & Smareglia 1998). Since it does not require the manifold to have physical meaning, SOM has also been used to classify astronomical publications (Poincot et al. 1998). More recently the SOM has been applied to calibrating redshifts for weak lensing cosmology (Masters et al. 2015), and several recent studies using the SOM will be mentioned throughout this work.

As other ML methods, the SOM starts with a training phase. However, unlike supervised methods, the goal of the training is to create a compressed, lower-dimensional representation of the data rather than estimate an output. In this work we first perform the training on galaxy colours drawn from the Horizon-AGN simulation. We then label the SOM with galaxy properties not learned during the training phase. These labels are drawn from a data model based either on simulations or, as suggested later, *bona fide* galaxies observed as calibration. Since the mapping from the data to the labels is explicit and analytic, control over selection functions, sample biases and the effects of observational noise is retained unlike supervised ML where these factors are part of the learning scheme.

In this paper we focus on galaxy SFR estimates because they are fundamental to constraining galaxy evolution (Madau & Dickinson 2014) along with stellar mass measurements. However, compared to the latter, the SFR estimates from template fitting are much more uncertain: previous work (e.g., Laigle et al. 2019) shows that SFR is more sensitive than stellar mass to SED-fitting assumptions. This sensitivity is inherent to relying on the ultraviolet (UV) continuum as a star formation indicator because it is highly attenuated by dust, geometry dependent, and also sensitive on the details of SFH. Other techniques offer better performance by using either far-infrared (FIR) data ($\geqslant 24\,\mu$m, Le Floc'h et al. 2009; da Cunha et al. 2008) or spectroscopic follow-up (Kennicutt 1998; Kewley et al. 2004) which is expensive and impractical to obtain for every galaxy. Therefore we focus the present study on estimating SFR from the rest-frame UV to near-IR (NIR) photometry, leaving the details of other physical properties to future work.

To test and develop our ML method we use a mock galaxy catalogue of $\sim 8 \times 10^5$ objects extracted from the cosmological hydrodynamical simulation Horizon-AGN (Dubois et al. 2014); the mock catalogue was presented in Laigle et al. (2019, hereafter Paper I) as the first milestone of the Horizon-AGN virtual observatory project. One of the main goals of the project is to bridge the divide between

Table 1. Terminology used throughout this work (see also Sect. 2.2).

Parameter space	M-dimensional space where each dimension is a galaxy feature (a colour in our case).
Colour ($\mathscr{C}$)	Observer's frame colour (e.g., $u-g$).
SOM	Self-organising map connecting the M-d space to a lower dimensional space (N-d).
Grid	Low-dimensional space with rectangular shape ($N=2$) used here to display the SOM.
Cell	Minimum element of the grid, which can contain one or more galaxies.
Weight ($\hat{w}$)	M-d vector for a given cell, connecting it to a point in the parameter space.
Training	Iterative process to define the weights and distribute galaxies into cells.
Mapping	Projecting a new galaxy on the grid, linking it to its best-matching weight/cell.
Calibration	Assign a value (label) to each cell, according to any galaxy property.
$\langle \dots \rangle^{\rm cell}$	Median of a generic quantity computed for galaxies in the same cell.

empirical and theoretical studies by adding observational-like features to simulated galaxy samples. To this purpose, we produced mock catalogues with characteristics similar to the COSMOS survey (Scoville et al. 2007) and *Euclid* (as predicted in Laureijs 2011). By applying our SOM estimator to the COSMOS-like version of the Horizon-AGN galaxies we aim to demonstrate its feasibility for real data sets. Both the simulated data set and the SOM are described in Sect. 2.

Building on this result, in Sect. 4 we introduce the SOM-based estimator of redshift and SFR, and apply it to the COSMOS-like mock catalogue. In Sect. 5 we show that redshift and SFR must be known for only a subset of galaxies in order to "calibrate" the SOM estimator and provide estimates without a synthetic template library. We then discuss possible ways to build such a calibration sample, inspired by previous work proposing a highly complete spectroscopic follow-up in large cosmological surveys (Masters et al. 2015; Hemmati et al. 2019b). To show the improvement of template-free ML with respect to standard fitting techniques, Sect. 5 ends with a comparison between SOM estimates vs the SFR derived in Paper I for the same galaxies by means of the code LePhare (Arnouts et al. 1999; Ilbert et al. 2006). We discuss the results and draw our conclusion in Sect. 6.

Throughout this work we use a flat ΛCDM cosmology with $H_0 = 70.4\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $\Omega_m = 0.272$, $\Omega_\Lambda = 0.728$, and $n_s = 0.967$ (Komatsu et al. 2011, WMAP-7). All magnitudes are in the AB (Oke 1974) system. The IMF is as in Chabrier (2003).

2 DATA AND METHODS

2.1 The Horizon-AGN virtual observatory

The present study relies on a mock galaxy catalogue built from the Horizon-AGN hydrodynamical simulation[1] (Dubois et al. 2014). This catalogue, presented in Laigle et al. (2019), includes 789,354 galaxies extracted from a 1×1 deg^2 light-cone by running the AdaptaHOP halo finder (Aubert et al. 2004) on the stellar particles distribution. Each stellar particle (of mass $\sim 2{\times}10^6\,M_\odot$) is linked to a synthetic simple stellar population (Bruzual & Charlot 2003, hereafter BC03) assuming Chabrier's IMF (Chabrier 2003) and interpolating between the metallicity values available in BC03. The galaxy catalogue is selected in stellar mass ($M_{\rm sim} > 10^9\,M_\odot$) and redshift[2] ($0 < z_{\rm sim} < 4$). The upper limit in redshift is imposed to work with galaxy colours consistently defined across the whole redshift range, i.e. avoiding non-detection in the u band due to the Lyman limit shifting into the band.

Our virtual observatory mimics the optical and near-infrared (NIR) photometry of COSMOS2015 galaxies (Laigle et al. 2016) in 10 broad bands (u, B, V, r, i^{++}, z^+, Y, J, H, $K_{\rm s}$) and 14 medium-band filters (from Subaru/SuprimeCam, see Taniguchi et al. 2007). It also includes the *Spitzer*/IRAC channels centred at 3.6 and 4.5 μm (hereafter [3.6] and [4.5]). In each filter we reproduce the signal-to-noise ratio (S/N) distribution as in the "ultra-deep" stripes of COSMOS2015. Reference 3σ limits (in $3''$ apertures) used in previous work are $K_{\rm s} < 24.7$ and $i^+ < 26.2$ (for a list of sensitivity depths in every filter, see Table 1 in Laigle et al. 2016). After introducing such uncertainties we perturb the original galaxy fluxes accordingly. Attenuation by dust and inter-galactic medium is also taken into account, whereas flux contamination by nebular emission is not implemented. In the following we refer to the attenuated fluxes without photometric errors as *intrinsic*, while *perturbed* photometry is the one that takes into account galaxy S/N.

Further details about the realisation of the Horizon-AGN mock galaxy catalogue can be found in Paper I and Appendix A1. We also use a simpler ("phenomenological") simulation to show that neither the Horizon-AGN limit in stellar mass nor the absence of nebular emission lines affect our main results (Appendix A2).

2.2 The self-organizing maps

In brief, a self-organising map (SOM) represents a high-dimensional data distribution into fewer dimensions (e.g., a 2-D space) through an *unsupervised* neural network that preserves topology (i.e., objects that were multi-dimensional neighbours remain close to each other in the 2-D space).

Assume that the original (compact) space $\mathscr{M}$, with dimensions equal to M, has to be reduced into a space $\mathscr{N}$, which we choose to be bi-dimensional ($N = 2$)[3]. To build the SOM, we create a neural network where each neuron is

[1] http://www.horizon-simulation.org/

[2] cosmological redshifts ($z_{\rm sim}$) in our light-cone include galaxies' peculiar velocity.

[3] Choosing $N = 3$ or higher is possible but it would not offer the same advantage in terms of visualisation.

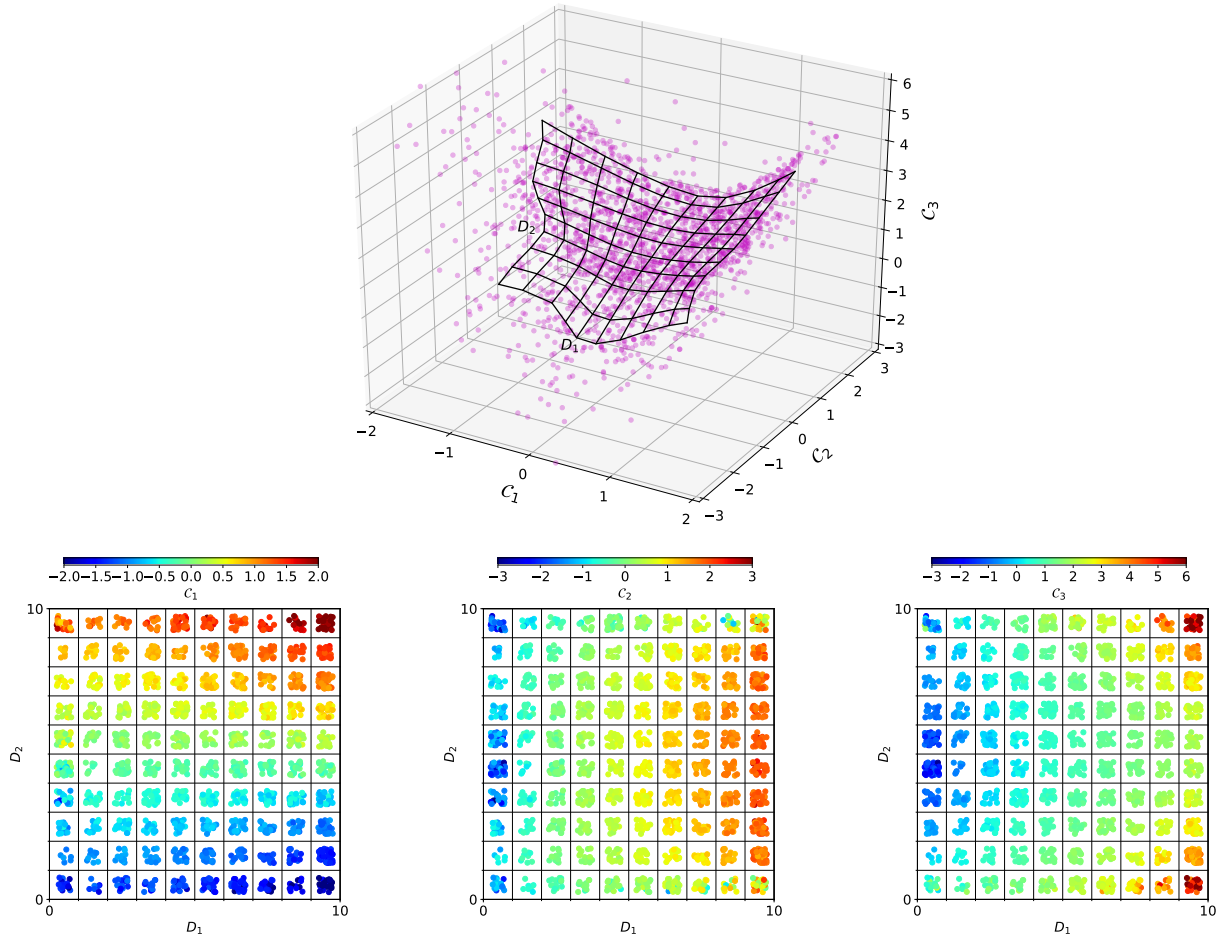

Figure 1. Pedagogical representation of the SOM using an artificial distribution of 2,000 objects in an unspecified 3-D space. *Upper panel:* After setting a 10×10 grid of "cells" the algorithm adapts the grid (black lines) to the manifold of the training objects (magenta points). Here, as well as in the next figures, SOM axes are labelled with conventional names $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$. The grid becomes finer where the density of the training sample is higher, while sparse objects in the outskirt will be linked to the border of the grid. Features $\mathscr{C}_1, \mathscr{C}_2, \mathscr{C}_3$ can be regarded as three galaxy colours to better understand our analysis. *Lower panels:* Layout of the SOM grid in the 2-D space and clustering of the training sample with respect to each feature (in three distinct panels). The 2,000 training objects (small dots) have been allocated into the various cells (each of them is expanded and the position of galaxies is scattered for illustrative purposes). The same 10×10 grid is presented each time with the objects colour-coded according to one of their features ($\mathscr{C}_1, \mathscr{C}_2, \mathscr{C}_3$ from left to right). Similar objects are clustered in the same (or nearby) cell and a smooth transition is observed across the grid. Outliers with extreme characteristics (e.g., $\mathscr{C}_3 > 4$) are pushed to the grid corners. In this example we use a simple distribution of data point for illustrative purposes, and the SOM grid as a consequence is a 3-d surface. In practice the grid can assume more complex high-dimensional configurations. Thanks to this feature the SOM and similar non-linear dimensionality reduction algorithms can accurately map the parameter space of real galaxies (which are a non-linear manifold). This is also a key difference from PCA, which can only assumes a hyper-surface.

associated to a weight vector ($\hat{w}$) of length M. Neurons are ordered in the N-dimensional configuration defined by the user, for instance a rectangular lattice (see a pedagogical example in Fig. 1).

The SOM relies on a *training sample* of objects drawn from $\mathscr{M}$. The neural network explores $\mathscr{M}$ by adapting neurons' weights to the training sample. Such a learning phase proceeds by iteration until the value of each weight gets as close as possible (according to a convergence criterion) to the input data. The first task in the procedure is to find a best-matching unit, i.e., the neuron whose weight $\hat{w}_b$ is the closest from a given data point $\hat{x}$ of the training sample. Then, the weight of any given neuron (including the best-matching unit) is updated during an iterative process:

$$\hat{w}_i(t+1) = \hat{w}_i(t) + \alpha(t)\, \phi(\hat{w}_i, \hat{w}_b, t)\, [\hat{x} - \hat{w}_i(t)]\,. \quad (1)$$

Equation (1) is written for the i-th neuron, updated from step t to $t+1$. The learning coefficient α is a monotonically decreasing function to ensure convergence, while ϕ is a neighbourhood function that modulates the update depending on the distance between the i-th neuron and the best-matching unit: $\phi \propto \exp(-|\hat{w}_i - \hat{w}_b|)$. The procedure is repeated by scanning the other elements of the training sample. It is critical to use a training sample that is representative of the whole space, otherwise the grid of neurons/weights is adjusted to probe only a sub-set of $\mathscr{M}$.

The learning process is unsupervised because it does not require the training sample to be labelled *a priori*. Neurons autonomously organises their weights: hence the name "self-organizing" map. The resulting SOM is a mapping function that connects a point from $\mathscr{N}$ to $\mathscr{M}$ and vice versa. We stress out that the topology is preserved so that in the low-dimensional configuration (the 2-D lattice in our example) two adjacent weights point to nearby regions in $\mathscr{M}$.

In our case the high-dimensional space is the typical baseline of an extragalactic survey, i.e. each dimension is a colour in broad-band filters (e.g., $u-g$, $g-r$, etc.). The training sample is a set of galaxies large enough to span the colour distributions observed in the universe[4]. We choose a rectangular lattice to order the neurons; because of its appearance, and to maintain the same lexicon of previous work, hereafter we will refer to the neurons as "cells" in a 2-D "grid" (Fig. 1)[5]. Each cell is defined by its weight vector, for example $\hat{w}_{ij}$ for the cell with coordinate i,j in the grid. The weight connects its cell to a point in the panchromatic space, i.e. the vector components $(w_{ij,1}; w_{ij,2}; ...; w_{ij,M})$ now represent a set of colours. The terminology used to describe the SOM is summarised in Table 1.

To follow Kohonen's prescriptions – i.e., to find the best-matching units and implement Eq. (1) – the distance between weights and galaxies is computed assuming a Euclidean metrics:

$$d_{ij} = \sqrt{\sum_{m=1}^{M} \left(\mathscr{C}_m - w_{ij,m}\right)^2}, \qquad (2)$$

where the given galaxy is defined by the colour vector $\hat{\mathscr{C}} = (\mathscr{C}_1, \mathscr{C}_2, ..., \mathscr{C}_M)$. The total $i \times j$ number of cells/weights is chosen by the user (see Sect. 3). As a result each SOM cell "contains" one or more galaxies from the training sample, whose colours are similar to the weight vector of that cell. The galaxy-cell association determined during the training phase is performed by the `python` software SOMPY[6]. Before the iterative process, to start with weights that are already close to the galaxy distribution, each weight vector is initialised by setting its colours via principal component analysis (PCA, Chatfield & Collins 1980) of the training sample. A parallelism between these weights and PCA eigenvectors can help understanding the SOM: its weights can be thought as a set of characteristic SEDs that describe the panchromatic space. However, the SOM has important differences from a PCA (see Sect. 3.3). In particular PCA is a linear hyper-surface defined by principle components, so it cannot fully describe a non-linear manifold, which is what we expect the galaxy colour space to be.

Once the SOM is trained, new galaxies can be mapped onto the grid by finding the nearest weight vector to each of them through Eq. (2). Moreover, the grid can be labelled *a posteriori* by looking to another galaxy property not included in the parameter space $\mathscr{M}$. For example, one can consider the redshift distribution of galaxies within a given cell and take the median ($\langle z \rangle^{\rm cell}$) to label that cell. Such explicit labelling step gives our method a key advantage over supervised ML because we keep control of the relationship between features (the broad-band colours) and labels (the redshift in the example above). This means that if we have a model of the bias or errors in our data we can directly account for that. However, it also means that an additional *calibration* phase is required to make the SOM work as an effective galaxy estimator. Indeed, as we will highlight in the following, galaxies clustered together in the colour space also share other (physical) properties.

3 SELF-ORGANISING MAP OF HORIZON-AGN GALAXIES

We apply the procedure summarised in Section 2.2 using Horizon-AGN galaxies as a training sample. The considered features are their broad-band colours $u-B, B-V, V-r, r-i^+, i^+-z^{++}, z^{++}-Y, Y-J, J-H, H-K_{\rm s}, K_{\rm s}-[3.6], [3.6]-[4.5]$. Except for the Subaru intermediate-band filters, which are not included here, this is the same baseline used in Paper I. As the SOM projects that 11-dimensional space into a rectangular grid (likewise Fig. 1) we can explore galaxy physical parameters as a function of their 2-D position, to see whether training objects falling in the same cell have in common other properties besides their broad-band colours. It should be noticed that a galaxy not detected in any of the filters poses a challenge to the SOM as one colour would be ill-defined. This is a common problem in ML methods that will be addressed in Sect. 4.

In the present Section we consider intrinsic colours (i.e., not affected by photometric noise) unless specified otherwise. The results discussed here are instrumental to show the fundamental properties of our method and its full potential in the case of an "ideal" survey. In Sect. 5 we will address the impact of observational uncertainties.

3.1 Generating an *ideal* SOM

Since we decided to adopt a rectangular lattice for the SOM the first step is to decide its size and axial ratio[7]. We set them by iteration, looking at changes in galaxy dispersion as a function of those two quantities (i.e., how tightly clustered are galaxies associated to a given cell). We also check that the number of galaxies per cell is large enough to assure a good sampling in the various region of the parameter space.

First we test the optimal SOM size. Starting from a 20×20 grid, we gradually increase the size by adding ten cells in both dimensions (i.e., maintaining a "square" configuration). Each time we train the SOM with the whole Horizon-AGN catalogue and measure i) the average distance of galaxies from their nearest weight, ii) the number of galaxies associated to each weight. The SOM converges fast with respect to i), so that in grids of $\gtrsim$6,000 cells most of the

[4] Even though we specifically explore the panchromatic space, the SOM can analyse any high-dimensional space defined by a combination of miscellaneous parameters (galaxy morphology, emission line fluxes, etc.).

[5] It should be noticed that in our SOM implementation input data will be normalised in each dimension, re-scaling the distribution to unit variance and centring the mean at zero.

[6] https://github.com/sevamoo/SOMPY

[7] Other 2-D configurations are possible, e.g. a lattice made by hexagonal cells or a spherical projection divided in HEALPIX (see Carrasco Kind & Brunner 2014 for a comparison).

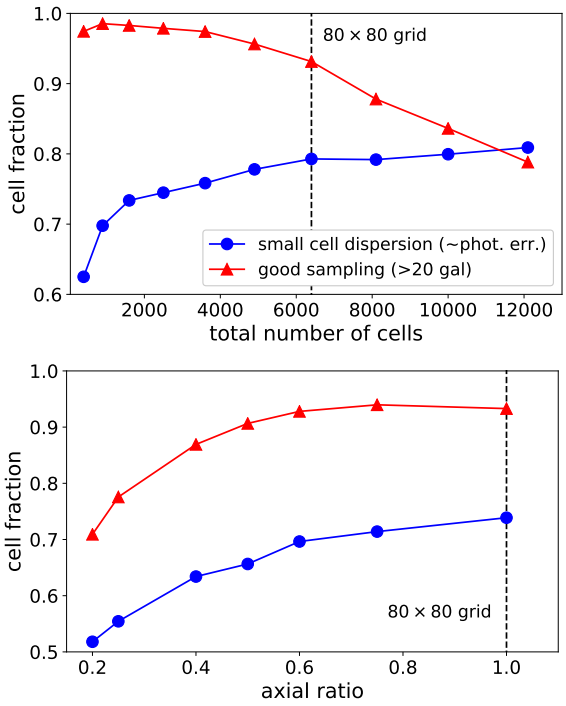

Figure 2. Tests to define size and shape of the SOM grid to optimally describe the galaxy parameter space. Different 2-D configurations are adopted, and each time the SOM algorithm is applied to the same Horizon-AGN sample (see Sect. 3.1). The quality of each configuration is measured by the number of cells with a good sampling (i.e., including >20 galaxies, red triangles) and high clustering (cells where the average distance of galaxies from the weight is comparable with typical photometric errors, blue circles). *Upper panel:* Fraction of cells satisfying both quality criteria while increasing the size of the grid. Only square configurations are considered, starting from a 20×20 cells up to 110×110 (i.e., 12,000 cells in total). The 80×80 configuration shows a good performance before the number of galaxies per cells starts to drop in larger grids. *Lower panel:* The two quality criteria this time are applied to rectangular grids with the same number of cells (6,400) but different aspect ratio. As in the upper panel, red triangles show the fraction of cells populated with a conspicuous number of galaxies while blue circles show the fraction of cells in which galaxies are tightly clustered around the weight. The 80×80 grid gives the best SOM.

galaxies are tightly clustered in their cells, i.e., their distance from the weight in the colour space is smaller then the typical photometric errors in deep surveys like COSMOS (0.01−0.05 mag from optical to IR). On the other hand, the larger the grid of weights, the fewer the number of galaxies associated to each of them. With the 90×90 configuration a significant area of the SOM starts to be under-sampled, with about 15 per cent of the cells defined by less than 20 galaxies each (Fig. 2, upper panel). Therefore we identify a slightly smaller size (80×80 cells) as a good compromise between high resolution and sampling, also considering computational efficiency.

The next step is to define the best geometry for our SOM, namely the ratio between its axes. In the previous test we used only square grids with increasing number of cells, while now we fix the number of cells to 6,400 and modify the aspect of the grid from 1:5 ratio to 1:1 (i.e., the

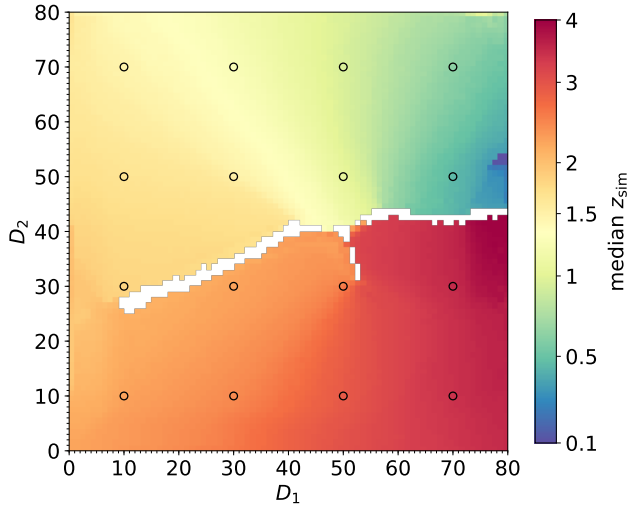

Figure 3. SOM trained with a sample of Horizon-AGN galaxies using their intrinsic colours. The grid is colour-coded according to the median intrinsic redshift of galaxies in the same cell ($\langle z_{sim} \rangle^{cell}$). Black empty circles identify cells whose galaxy spectra are analysed in Fig. 4. The white region is made by empty cells with no galaxy matching their weight. Despite redshift evolution is overall smooth across the map, in the upper-right corner a streak of $\langle z_{sim} \rangle^{cell} \sim 1.5$ cells show the small impact of border effects (see Sect. 3.1).

80×80 configuration) in eight steps. We describe again the quality of each SOM in terms of i) and ii), finding that the best configuration is that with 1:1 axis ratio (Fig. 2, lower panel). In a rectangular grid there are more galaxies not well represented (i.e., far from their best-matching weight) especially when the two sides have very different lengths.

In conclusion, the SOM we will use throughout is made by 80×80 cells. The result is specific for the Horizon-AGN parameter space: the optimal configuration for another training sample may be different. We notice however that the total number of cells is comparable to those used to describe the real COSMOS and CANDELS data sets respectively in Masters et al. (2015, 11,250 cells, but for galaxies up to $z \sim 6$) and Hemmati et al. (2019b, 4,800 up to $z \sim 4$). An alternate method, proposed by Hemmati et al., consists in increasing the grid size until the histogram of each weight vector component matches the distribution of the corresponding colour (see their figure 6). We calculate these histograms and find that indeed they converge when the grid dimension is $\geqslant 80 \times 80$.

3.2 Redshift calibration

After the training phase, we can label each SOM cell according to a given property of the galaxies contained in it. We compute the median redshift $\langle z_{sim} \rangle^{cell}$ (see Table 1) and the relative scatter σ_z defined as $\sigma(z_{sim} - \langle z_{sim} \rangle^{cell})$. The SOM of Horizon-AGN shows the same cell-z relationship found by Geach (2012) in COSMOS, with a smooth redshift evolution as moving across the 2-D space (Fig. 3). The redshift scatter in every cell (not shown in the Figure) is particularly small, with σ_z always between 0.05 and 0.1. The small

scatter is even more remarkable when the σ_z values are divided by a $1+z$ factor, shrinking to 0.01−0.03. This is due to the algorithm's ability of clustering objects with very similar (observer's frame) colours, which correspond to similar redshift. One may expect some redshift interlopers – i.e., objects with a significantly different $z_{\rm sim}$ from the rest of the cell – due to SED degeneracies (Papovich et al. 2001). However, we do not find this in the ideal case discussed here. On the other hand, we do observe boundary effects produced by galaxies with extreme colours, which lie at the limits of the panchromatic manifold (see also the example in Fig. 1). Those galaxies are pushed to the border of the grid, but with a negligible impact on the redshift distribution (the redshift scatter remains modest: $\sigma_z \simeq 0.1$).

In Fig. 3 we also observe a long horizontal stripe of empty cells. No galaxy has been associated to their $\hat{w}$ during the training. The $\langle z_{\rm sim}\rangle^{\rm cell}$ labels explain the physical meaning of this empty region: it is a "caustic" in the parameter space dividing $z \sim 3$ galaxies from those at lower redshift with similar colours. Since we are working with intrinsic photometry, their Lyman vs Balmer break degeneracy (e.g. Stabenau et al. 2008) is fully disentangled.

3.3 High-resolution spectra in the SOM cells

Within the Horizon-AGN virtual observatory broad-band colours are built from high-resolution spectra, which are therefore available for each mock galaxy. These synthetic spectra are built with BC03 models, accounting for the complex SFH and chemical enrichment (Sect. 2.1).

In the SOM we stack galaxy spectra from the same cell, re-normalising them to a fixed i^+ flux to ease the comparison. The fact that the colour-based SOM can efficiently map redshifts (Sect. 3.2) does not necessarily imply that it works as well with spectral features. Given the small amount of spectroscopic information in real data sets, previous work has only proven that broad-band SEDs are well clustered within the grid, with the exception of Rahmani et al. (2018) analysing 142 spectra at $0.5 < z < 1$ and Hemmati et al. (2019b) showing a handful of $z \sim 1$ spectra that have similar shape and are also clustered in nearby cells. We find that this is actually the case for the whole Horizon-AGN (noiseless) sample: in most of the cells galaxy spectra are in excellent agreement (see a few examples in Fig. 4). Regions close to the redshift caustic show more dispersion, mainly because the median stacking is performed in observer's frame and therefore it is affected by the difference between individual redshifts. We check that spectral shapes are even more similar if the comparison is made in rest frame, removing the σ_z scatter.

In addition to the examples shown in Fig. 4 we perform the same stacking analysis in 225 distinct cells evenly distributed across the grid[8], finding that their average inter-quartile dispersion is always $\lesssim$15 per cent (< 8 percent in half of the cells probed). It is worth emphasising the differences between this result and what a PCA classification would give. First of all, PCA provides a basis of eigenvector to be linearly combined, whereas the SOM works also with non-linear transformations. Each weight in the SOM has a clear physical meaning by itself, i.e., it describes a galaxy *phenotype* in the observed frame (Sánchez & Bernstein 2019). On the other hand, in a PCA classification, the 2–4 eigenspectra that usually have the most discriminating power are difficult to interpret. They can be combined to reproduce actual galaxy features, but classes for those resulting spectra are not inherently provided by the PCA and human intervention is required (e.g., defining meaningful regions in a Karhunen-Loe've diagram, Marchetti et al. 2013).

The degree of similarity of the simulated spectra within a given SOM cell also depends on the complexity of their features. Since spectra used in this work are extracted from the Horizon-AGN simulation, their realism and complexity are constrained by the physical models implemented in that simulation. Modelling galaxy evolution on cosmological scales is inevitably done at expenses of resolution. Because Horizon-AGN maximum spatial resolution is at best 1 pkpc, the impact on the inter-stellar medium (ISM) of any process occurring at smaller scale is averaged through the sub-grid recipes. These recipes have been iteratively improved in order to reproduce as well as possible the statistical distribution of integrated galaxy properties throughout cosmic time (in Horizon-AGN such a progress can be tracked through Dubois et al. 2012, 2014; Park et al. 2019). However, they might fail to some extent at reproducing ISM in-homogeneity and clumpiness. As a result, we expect the simulated star-formation histories of our galaxies to be smoother (and their spectra less diverse) than the real ones. As a consequence, the dispersion that we measured within a SOM cell must be considered as a lower limit. A more extended discussion about caveats in our modelling is provided in Appendix A1.

3.4 Exploring other physical parameters in the SOM

The remarkable similarity between spectra in the same cell suggests that those galaxies went through a similar evolutionary path, resulting e.g. in the same mass-to-light ratio at the redshift of observation. Horizon-AGN provides us with galaxy physical parameters such as luminosity (L), stellar mass ($M_{\rm sim}$), and star formation rate (SFR$_{\rm sim}$). Thus, we can use our simulation to test whether the objects that the SOM groups together have other properties in common besides their redshift. Regarding SFR$_{\rm sim}$, this is defined by averaging galaxy SFH over the last 100 Myr, an interval comparable to the timescale of SFR indicators widely used in the literature (Kennicutt & Evans 2012).

We calibrate the SOM to show the typical mass-to-light ratio per cell (Fig. 5, left panel). The procedure is similar to the redshift calibration in Sect. 3.2, this time computing the median $\langle M_{\rm sim}/L_V\rangle^{\rm cell}$ with L_V being the luminosity in the rest-frame V band. In this way we can visualise the variation of this quantity across the grid, which is also an evolution across redshift: more mature galaxies with larger M/L occupy cells with lower $\langle z_{\rm sim}\rangle^{\rm cell}$ (cf. Fig. 3). We also aim at verifying that the scatter inside a given cell is small by calculating the difference between the 84th and 16th percentile in the logarithmic $M_{\rm sim}/L_V$ distribution. We find that in most of the cells this is smaller than 0.2 dex. A few ex-

[8] Those cells have coordinates $D_1 = 5i$ and $D_2 = 5j$, with i and j ranging from 1 to 15.

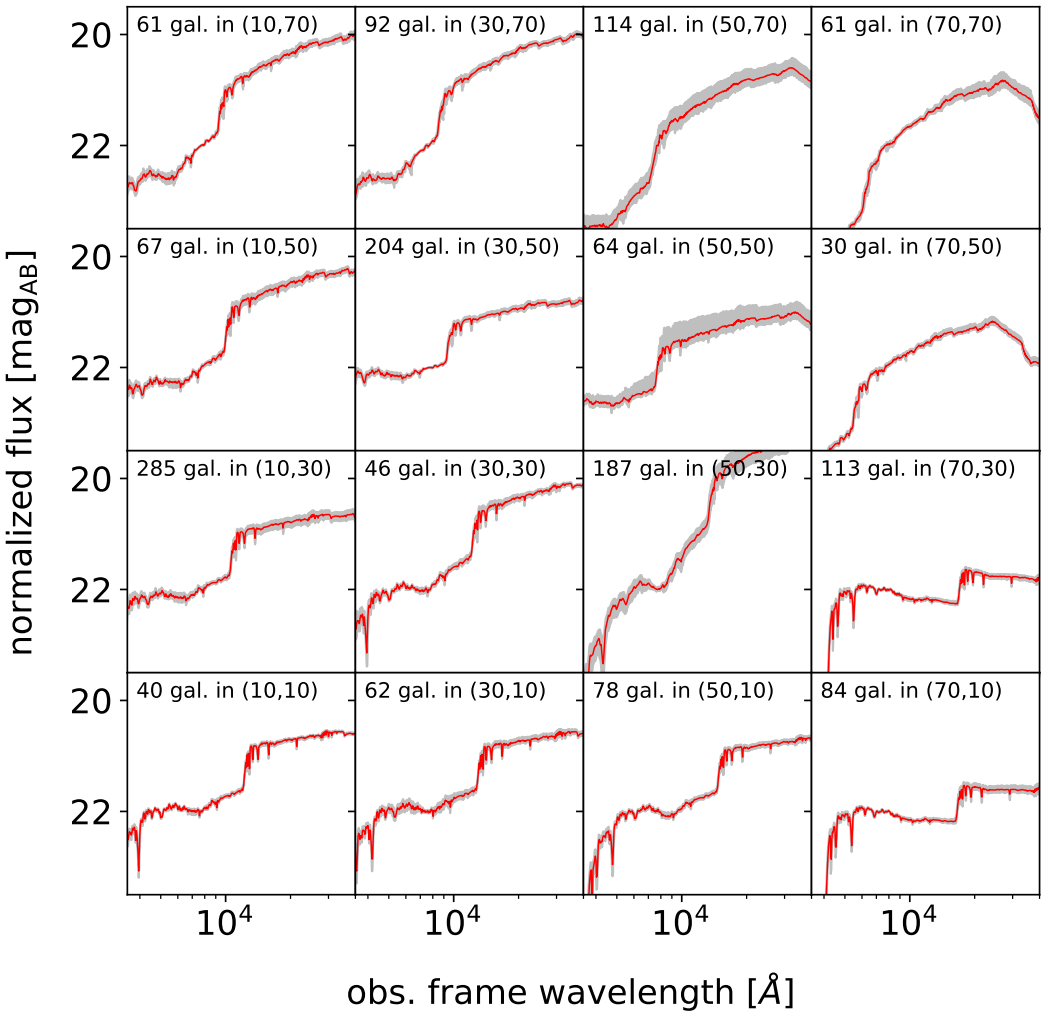

Figure 4. Median stacked spectrum (observer's frame, red line) of galaxies in a given cell with relative inter-quartile dispersion (grey shaded area). Label at the bottom of each panel indicates number of galaxies and cell coordinates (see also empty circles in Fig. 3). Before stacking, spectra have been re-normalised to the arbitrary i^+-band flux of 22 mag. Each galaxy spectrum is redshifted according to its $z_{\rm sim}$ value.

amples of the tight correlation between position in the grid and M/L are shown in the histograms of Fig. 5.

A similar trend can also be observed regarding stellar mass after a scaling factor is applied. This factor is required since we do not train the SOM with information about SED normalisation, as instead other authors do with different ML methods (Bonjean et al. 2019)[9]. This is the same procedure used in Fig. 4 to compare galaxy spectra, which have similar shapes but different magnitude. Therefore, to analyse the intrinsic $M_{\rm sim}$ scatter within a given cell, we first normalise each galaxy to a reference point of $i^+ = 22$ mag. In principle this should be done in rest frame given the fundamental M/L correlation, but since spectra in the same cell are at about the same redshift one can use apparent magnitudes. After such a re-scaling, the $\log(M_{\rm sim}/M_\odot)$ dispersion is smaller than 0.2 dex in most of the cells (see three examples in Fig. 5, and also 6).

The specific SFR (sSFR≡SFR$_{\rm sim}/M_{\rm sim}$) does not need such a normalisation as it is expected to be directly related to M/L. However the scatter is larger than M/L and also stellar mass. In fact, the latter is an integrated quantity strongly connected to the global evolutionary path of the galaxy, whereas the sSFR depends on recent fluctuations in the mass assembly history. The 0.2–0.5 dex dispersion in the SOM (Fig. 5 and 6) reflects such stochasticity; sSFR is also more sensitive than stellar mass to different levels of dust attenuation.

These tight correlations are obtained with a SOM trained with intrinsic colours: larger scatter is expected when galaxy photometry is affected by observational uncertainties. The results shown in the section can be considered as an "asymptotic" limit represented by an ideal galaxy survey with infinite signal-to-noise ratio (S/N). This ideal example can be approximated by the brightest galaxies in an "ultra-deep" survey. In that case the S/N should be high enough to

[9] We prefer working in a pure colour space because an additional dimension with a different dynamical range may not be properly weighed with respect to the others.

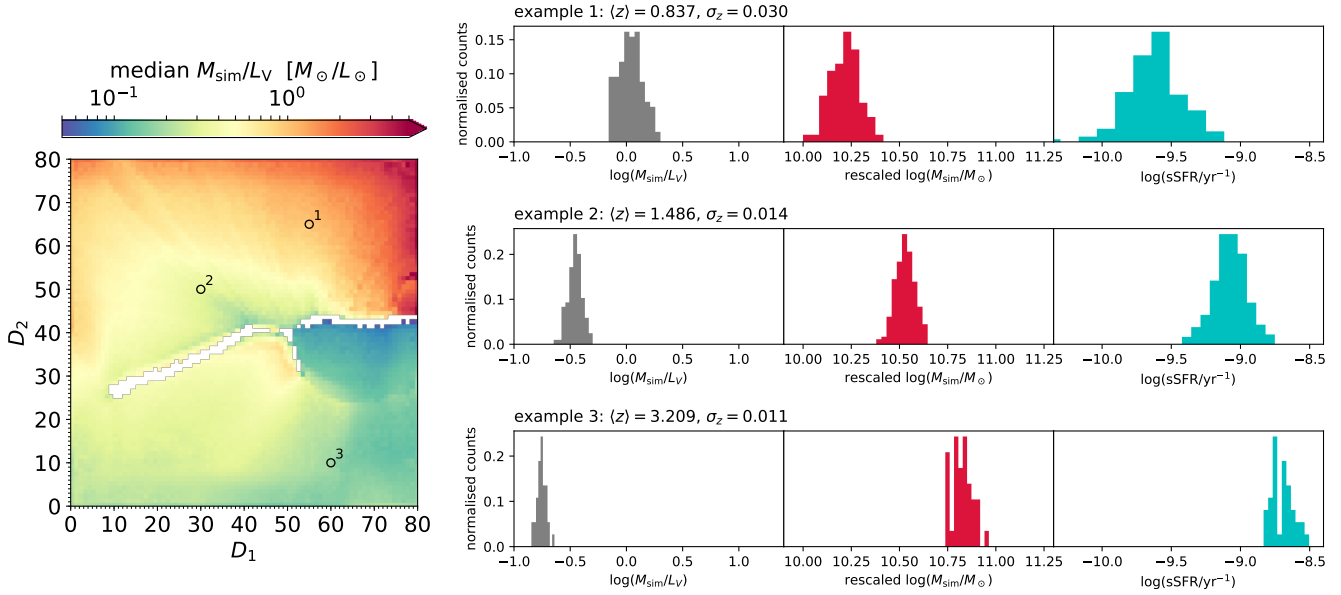

Figure 5. The SOM trained with the intrinsic colours from the Horizon-AGN catalogue is able to group galaxies with similar physical properties together in the same cell. *Left:* The SOM grid (same as Fig. 3) is now labelled according to the median mass-to-light ratio in each cell. Empty circles mark three (randomly extracted) cells used as examples to show the tight correlation between position in the grid and physical properties. *Right:* Each row corresponds to one of the example cells marked in the left panel and shows the logarithmic distribution of M_{sim}/L_V (grey histogram), M_{sim} (red histogram), and sSFR (cyan histogram) in that cell. In the case of the M_{sim} distribution, individual values have been normalised to a common i^+-band flux (corresponding to 22 mag) to show that the dispersion in a given cell is mainly due to the fact that the SOM, being trained with colours only, is not informed about the normalisation of the SED.

enable the SOM to classify galaxies not only with respect to their redshift, but also stellar mass and SFR. Such a clustering ability shall allow, as discussed in the next section, to recover stellar mass and SFR of galaxies with an original technique. In principle, knowing the stellar mass of a "calibration" object (M_{cal}, which can be independently obtained via template fitting) one could get a fairly precise M estimate for other galaxies mapped into the same cell: it would suffice to re-scale M_{cal} by the flux ratio between the known galaxy and the target one: $M \simeq M_{cal} \times (f/f_{cal})$. As suggested by the examples in Fig. 5 in the best-case scenario the uncertainty of such an estimate would be comparable to the typical mass errors of template fitting codes ($\lesssim 0.3$ dex, see e.g. Davidzon et al. 2017) with significant improvement in computational speed (about 10^6 times faster according to Hemmati et al. 2019a).

The same argument used for stellar mass applies to SFR, which is not shown in Fig. 5 as it will be thoroughly discussed in the following. A flux re-normalisation will be necessary as in the case of stellar mass. In the present analysis we use the i^+ band for the flux scaling factor since it has one of the deepest sensitivity limits. The K_s band is a better proxy for stellar mass but in COSMOS2015 (and therefore in our COSMOS-like sample) its 3σ limit ~1.5 mag shallower than i^+ and would result in a larger scatter of the results (but still preserving the properties of the SOM, see Appendix A2).

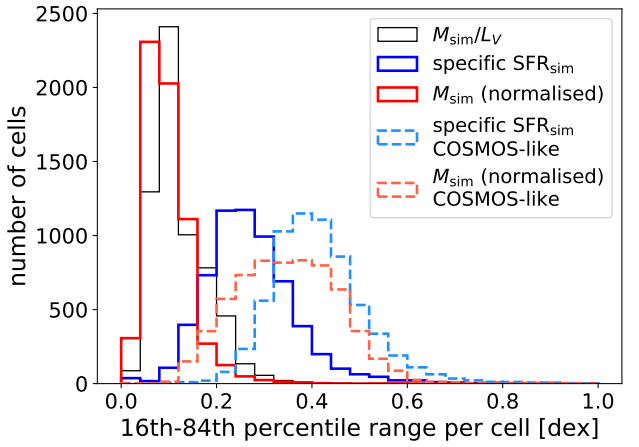

Figure 6. Logarithmic dispersion of various physical properties in each cell of the SOM, defining such a scatter as the difference between 84th and 16th percentile of the logarithmic mass-to-light ratio (thin black histogram), sSFR (blue), stellar mass (red). For sSFR and stellar mass, solid lines show the results when intrinsic colours are used (see Sect. 3.4) while dashed-line histograms are the case in which galaxy photometry is perturbed with COSMOS-like errors (Sect. 4.1). Since the SOM is trained only with colours, an SED normalisation has to be applied for stellar masses; we used the i^+-band fluxes to this purpose, re-scaling galaxies to $i^+ = 22$ as in Fig. 4.

4 A NOVEL ESTIMATOR OF GALAXY SFR

We concluded Sect. 3 by suggesting that the SOM can be used to recover galaxy physical parameters emperically using the same framework described in Masters et al. (2015) for photometric redshift computation. In this section we discuss how to calibrate the SOM in order to derive SFRs from broad-band colours in a fast but accurate way. We focus on SFR since standard SED fitting shows its limit when deriving this quantity (see Paper I). One can obtain robust SFR estimates based on IR imaging or spectroscopy, but this is generally possible for a small fraction of galaxies (which indeed we can use as a calibration sample for our method). A SOM-based estimator for stellar masses should also be feasible, but it would likely be based on template fitting, at least for calibration purposes, given the current state of the art in estimating stellar mass[10]. For template based approaches, we refer the reader to (Hemmati et al. 2019a) for an overview of how to effectively use the SOM. Here, our goal is to empirically calibrate properties based on direct indicators rather than depending on synthetic SED libraries.

With this goal in mind we devise a method that can be applied to real data, requiring some adjustment to the SOM. We replace the ideal (noiseless) photometry used in the previous section with a catalogue that mimics the COSMOS survey, including errors and selection functions. In other words, apparent magnitudes and colours of Horizon-AGN galaxies are now perturbed with observational-like errors and selection effects. As a consequence we use a different training sample, selecting only galaxies above a given S/N threshold. So, the observational-like training sample does not include objects that are undetected in some band. In Sect. 4.1 we show that, after the S/N selection, photometric errors do not impair the relationship between SOM cells and galaxy properties shown above.

In Sect. 4.2 we explain the details of our methods. In that context we modify the way to label SOM cells. In fact, in previous sections we made calibrated versions of the SOM by labelling its cells with median values of either redshift or other physical parameters, i.e., under the assumption of knowing them for the whole sample. Hereafter we assume to know the SFR of a small subset of 6,400 galaxies (i.e., one object per cell) and use them to label the grid. At the moment we do not make particular assumptions on how such a calibration sample is built; this is discussed later in Sect. 5 where we also compare to SFR estimates from template fitting.

4.1 The COSMOS-like SOM

The observational uncertainties to perturb the Horizon-AGN photometry are statistical errors affecting apparent magnitudes so that our mock galaxy catalogue resembles the quality of COSMOS data (see Paper I). For this reason in the following we refer to the (noisy) Horizon-AGN sample also with the term "COSMOS-like", in contrast to the previous version with intrinsic photometry (Sect. 3). We do not model confusion noise and contamination by saturated stars; this kind of issues shall be addressed in future work after providing the Horizon-AGN virtual observatory with simulated images.

The training phase of the implemented algorithm does not account for cases of non-detection (e.g., when the "observed" flux is smaller than the flux error). Therefore, while working with perturbed photometry, we limit the analysis to a galaxy sub-sample with $S/N > 1.5$ in each broad-band filter. A statistically correct treatment of lower and upper limits in the input colours would require an improved SOM algorithm that is beyond the goal of the present work. We note that also template fitting codes often neglect such a treatment (as highlighted in Sawicki 2012). The S/N preselection will restrict the analysis to $z \lesssim 3.5$ because galaxies at higher redshift are u-band drop-outs (Steidel et al. 1996) with $S/N \ll 1$ in that filter. Given the sensitivity of our catalogue the S/N threshold roughly corresponds to a flux-limited survey with a cut at $i^+ < 25$ (see Appendix A3). Besides the removal of $z > 3.5$ galaxies from the S/N-selected sample there are other caveats, which are listed in Appendix A3. None of them affects the analysis between $z = 0.2$ and $z \sim 3$, but there is a "boundary effect" at the lowest and highest redshifts of the range (see discussion below).

After this preliminary test we use the perturbed photometry of Horizon-AGN galaxies with $S/N > 1.5$ to produce a new SOM; the multi-dimensional space is the same as in Sect. 3 (11 broad-band colours). Figure 7 shows the resulting redshift map limited to $z_{\rm sim} \lesssim 3.5$. The redshift evolution across the 80×80 grid is similar to the ideal SOM, although some details are smeared out because of photometric errors. For example the gap between low- and high-z regions is now filled by scattered galaxies (cf. Fig. 3). We quantify redshift dispersion in each cell through the normalised median absolute deviation (NMAD, Hoaglin et al. 1983) and the outlier fraction. The former is defined as $1.48\times{\rm median}(|\Delta z|)/(1+\langle z\rangle^{\rm cell})$, where $\Delta z \equiv z_{\rm sim}-\langle z\rangle^{\rm cell}$ (Fig. 7, middle panel). The latter is the fraction of objects in a given cell having $|\Delta z|/(1+\langle z\rangle^{\rm cell}) > 0.15$ (Fig. 7, lower panel). These metrics are similar to those used to describe the quality of $z_{\rm phot}$ estimates (e.g. Ilbert et al. 2013). The NMAD is on average < 0.047, with only 343 cells above 0.1 (mainly in the formerly empty region of the z caustic). The outlier fraction is overall small, being less than 10 per cent in 5,228 cells (namely 82 per cent of the grid). On the other hand systematics at the borders are more evident than before.

Besides redshift, Horizon-AGN galaxies remain well clustered also with respect to M and sSFR (see Fig. 6). With perturbed photometry ${\rm sSFR}_{\rm sim}$ is slightly less constrained than in the ideal SOM: e.g., now there are no cells with scatter <0.2 dex. On the other hand only ~10 per cent of the cells exceed 0.5 dex dispersion in sSFR. These are cells hosting low-sSFR galaxies. The results are comparable to other classification methods (e.g. the NUV $-$ r vs $r-K$ diagram, Arnouts et al. 2013). The typical scatter in $\log(M_{\rm sim}/M_\odot)$ within one cell is 0.3–0.4 dex, much larger than before because not only the input colours but also the i^+-band rescaling now is done with perturbed fluxes. However, we note that it is of the same order of statistical errors of M measurements for observed galaxies (e.g., Davidzon et al. 2017), a further indication that the SOM estimator can work also in the observed universe.

[10] Other options are available in some case, e.g. dynamical masses from spectroscopy (Courteau et al. 2014).

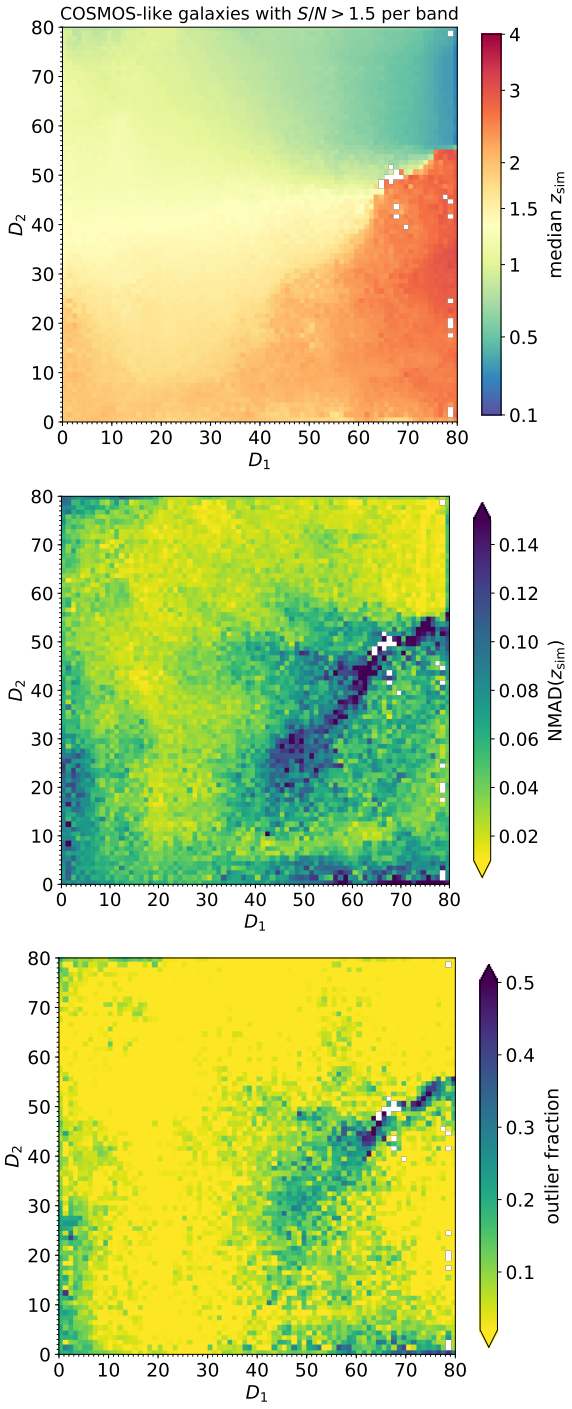

Figure 7. SOM algorithm applied to Horizon-AGN colour space, including photometric errors. A sub-sample of galaxies with robust colours ($S/N > 2$ in all the photometirc pass-bands) is used to train and label the SOM. *Upper panel:* Cells are colour-coded according to their median redshift (compare to Fig. 3). The redshift range is now $0 < z < 3.2$ because of the S/N limit imposed in the u band (see text). For the same reason 13 cells sparsely distributed across the grid are empty (coloured in white). Other 9 empty cells identify the redshift "caustic", the rest of which is filled by scattered objects. *Middle panel:* normalised median absolute deviation (NMAD). *Lower panel:* fraction of redshift outliers in each cell.

4.2 Galaxy redshift and SFR measurements

Encouraged by the previous tests, we proceed in the implementation of the SOM estimator. First of all we need a reference sample to label the SOM grid with both redshift and SFR values. Therefore we assume to "observe" one galaxy per cell to obtain an estimate of their redshift and SFR. These galaxies belong to the *calibration sample* and their "measured" properties are dubbed $z_{\rm cal}$ and $\mathrm{SFR}_{\rm cal}$. For the moment we do not make stringent requirements about how $z_{\rm cal}$ and $\mathrm{SFR}_{\rm cal}$ are measured: they may come e.g. from a spectroscopic survey, but not necessarily (e.g., Arnouts et al. 2013). We only make the assumptions that these are *bona fide* galaxies with reliable z and SFR, and they cover the entire 80×80 grid (i.e., there is one calibration galaxy per cell). Each calibration galaxy is randomly targeted among those in the given cell. For sake of simplicity, we do not model observational uncertainty so the $z_{\rm cal}$ and $\mathrm{SFR}_{\rm cal}$ values correspond to $z_{\rm sim}$ and $\mathrm{SFR}_{\rm sim}$ of the given galaxy. In Sect. 5.1 we will discuss which kind of survey might provide such a calibration sample, modifying $z_{\rm cal}$ and $\mathrm{SFR}_{\rm cal}$ accordingly.

The other galaxies in the SOM, not used for the calibration, will get an estimate of redshift and SFR from the SOM through the procedure described here. The method takes into account not only the best-matching cell in which galaxies lie but also the nearby ones. This choice is motivated by the impact of colour uncertainties: Even though a COSMOS-like galaxy is associated to a specific cell after the training phase, weights of other cells are still compatible within the error bars with its colours and there is a non-negligible probability that one of them is in better agreement with the intrinsic features of the galaxy.

For each entry of the mock catalogue our algorithm includes the following steps:

(i) consider $N_{\rm c}$ cells: the best-matching unit in which the Horizon-AGN galaxy reside and the nearest $N_{\rm c} - 1$ cells;
(ii) calculate the distance between the galaxy and each of those cells, with a modified version of Eq. (2) that takes into account photometric errors:

$$d_i = \sqrt{\sum_n \left(\mathscr{C}_n - w_{i,n}\right)^2 / \Delta\mathscr{C}_n^2}\,, \tag{3}$$

where $\Delta\mathscr{C}_n$ is the 1σ uncertainty for the n-th colour and i is one of the $N_{\rm c}$ cells;
(iii) take the $z_{\rm cal}$ and $\mathrm{SFR}_{\rm cal}$ labels of the $N_{\rm c}$ cells;
(iv) compute their distance-weighted mean $z_{\rm SOM}$ and $\mathrm{SFR}_{\rm SOM}$.

In particular the resulting SFR is re-normalised (as done for stellar masses in Sect. 3.4):

$$\mathrm{SFR}_{\rm SOM} = \frac{\sum_{i\in N_{\rm nc}} (\mathrm{SFR}_{\rm cal}/d_i^2) \times (f/f_{\rm cal,i})}{\sum_{i\in N_{\rm c}} 1/d_i^2}\,, \tag{4}$$

where $f/f_{\rm cal,i}$ is the flux ratio in a reference band (we choose i^+) between the given photometric galaxy and the *bona fide* one that labels the i-th cell. The square distance from the i-th cell is also used in the weighted mean to compute $z_{\rm SOM}$.

We set $N_{\rm c} = 10$ as the scatter generated by photometric errors typically involves the first surrounding cells. We verify that including more distant neighbours do not alter the results owing to the $1/d^2$ factor in Eq. (4). A more ac-

curate, object-by-object determination of N_c could be done defining the 11-dimensional ellipsoid enclosing the 68 per cent confidence limit in all dimensions jointly. However the colours' covariance matrix is necessary for such a task (Press et al. 1992) and that is not available in Horizon-AGN, to be consistent with the real COSMOS catalogue. To tackle this limitation, (Hemmati et al. 2019a) suggest an empirical method based on mapping Monte Carlo extraction of SEDs within their photometric errors to the SOM. A more rigorous Bayesian approach can be found in Carrasco Kind & Brunner (2014) and Buchs et al. (2019, see also Bonnett 2015 for neural network redshifts). We postpone to future work a thorough analysis of the redshift PDF via SOM.

We emphasise that the entire procedure takes less than 30 minutes of wall clock time, whereas to process the same 371,168 galaxies LePhare needs more than 100 hours (without considering the computational time to estimate redshifts in the first run). In Fig. 8 we provide a comparison between the true redshifts of Horizon-AGN galaxies and those derived either through the SOM (upper panel) or LePhare (z_{phot}, lower panel). The Figure shows 371,168 COSMOS-like galaxies from $z = 0$ to ~3.5, namely the S/N-selected sample with the exception of the *bona fide* galaxies used for calibration. Overall, z_{SOM} are in decent agreement with z_{sim}, despite the significant scatter. NMAD and outlier fraction are computed with respect to $\Delta z \equiv z_{SOM} - z_{sim}$, being respectively 0.044 and 6.1 per cent. Galaxies at $z_{sim} < 0.2$ and $z_{sim} > 3.2$ are the most problematic as they suffer from SOM boundary effects: at those redshifts, i.e. the extremes of the distribution, there are too few galaxies to train a distinct cell. For instance there are only 910 galaxies with $z_{sim} < 0.2$, spread across 23 cells; in each cell they represent 5–12 per cent of the objects because they are classified together with a much larger number of $0.2 < z_{sim} < 0.5$ galaxies.

The NMAD and outlier fraction we find are both larger than those computed in Masters et al. (2019) for the COSMOS z_{SOM}, but their method slightly differs from ours as they use a deeper sample for calibration and then map (typically brighter) spectroscopic galaxies on the SOM. Here we compute the NMAD and outlier fractions with a sample that goes fainter, which explains the slightly worse results. In Paper I we discussed the caveats of using such a sub-set of galaxies to assess SED fitting quality of the parent photometric sample, as it is a limited (and sometimes biased) representation of the entire population. The comparison of Fig. 8 does not have this caveat because the same Horizon-AGN galaxies are used in both cases. We find that LePhare NMAD (0.024) and outlier fraction (1.5 per cent) are significantly smaller. It is worth noticing that those estimates take advantage of a more advanced Bayesian framework during the SED fitting: a galaxy z_{phot} is defined as the median of the PDF(z) resulting from LePhare template library. More complex applications of the SOM (Masters et al. 2019; Sánchez & Bernstein 2019) yield to redshift estimates comparable in precision to LePhare. What is nonetheless surprising is the significantly better performance of our SOM in terms of redshift bias (as pointed out also in Masters et al. 2019) defined as the mean of $(z_{SOM} - z_{sim})/(1 + z_{sim})$. This can be deduced from Fig. 8 by observing the scatter in the two panels: the one of z_{SOM} vs z_{sim} is larger but more symmetric. After removing outliers, the z_{SOM} bias is -0.001, a factor 10 smaller than the bias resulting from LePhare.

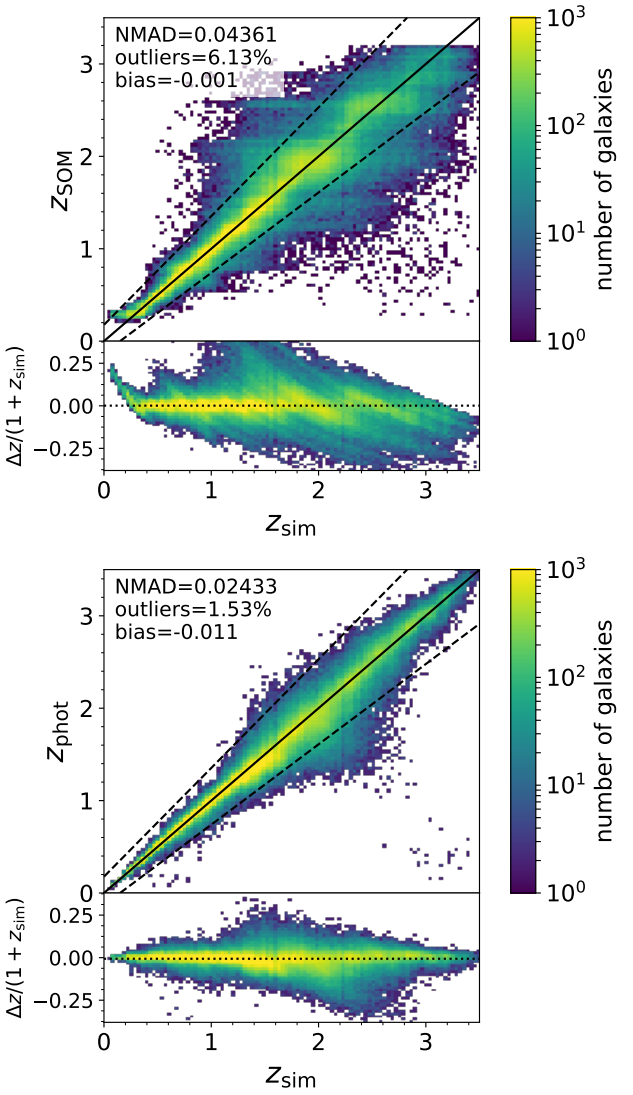

Figure 8. Upper panel: photometric redshifts derived from the SOM estimator (z_{SOM}) compared to the intrinsic redshifts (z_{sim}), for the 371,168 COSMOS-like training sample not used for calibration. A solid line shows the 1:1 bisector while dashed lines mark the $\pm 0.15(1+z)$ threshold used to compute the outlier fraction. Metrics describing the quality of results are quoted in the upper-left corner. The bottom of the figure shows the scatter $\Delta z/(1+z_{sim})$ as a function of redshift. Lower panel: for the same galaxy sample, z_{sim} values are compared to estimates from standard SED fitting (LePhare code). Symbols are the same as in the upper panel.

Along with z_{SOM} this method provides SFR estimates at the same time. Fig. 9 shows how they compare to the intrinsic SFR_{sim} in different redshift bins[11]. The correlation between the two is very tight for $SFR_{sim} > 1\,M_\odot\,yr^{-1}$, whereas the method overestimates lower levels of star formation (at least at $z > 0.8$). There is also a systematic underestimation for the most star-forming galaxies, but that offset is always

[11] See Fig. A4 for an alternate version of the calculation using K_s fluxes for the re-scaling in Eq. (4).

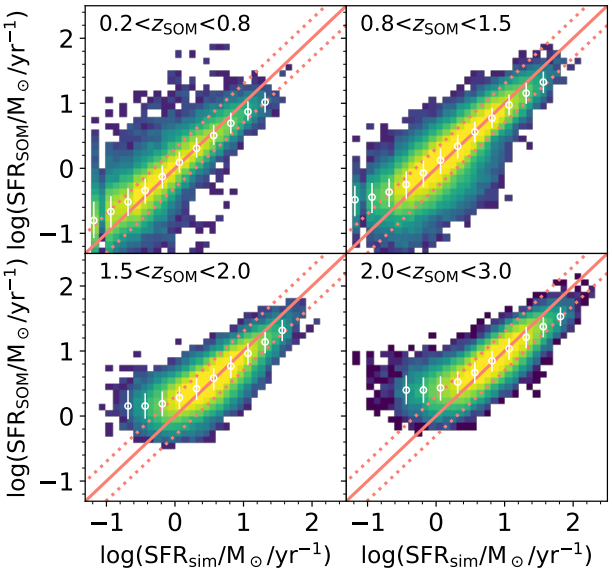

Figure 9. Comparison between intrinsic star formation rates (SFR_{sim}) and estimates obtained through the SOM (SFR_{SOM}) for the same Horizon-AGN galaxy sample shown in Fig. 8 (using also the same colour map for object density). Each panel shows a different redshift interval: galaxies are binned according to their z_{SOM}, which is provided by our method at the same time (see Sect. 4.2). Empty circles are median values in running bins of SFR, with error bars computed as the difference between 84th and 16th percentile. A solid line indicates the 1:1 bisector and dotted lines the ±0.15 dex offset from it.

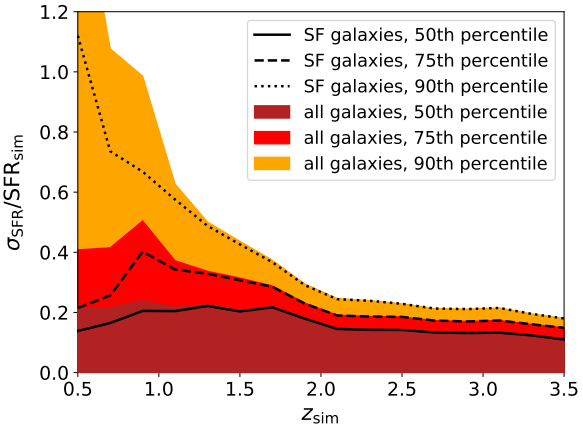

Figure 10. Fractional error of SOM-derived SFR estimates (σ_{SFR}/SFR_{sim}) due to the SFR_{cal} stochasticity, i.e. the random selection of the *bona fide* galaxies. To compute σ_{SFR} we repeat our SOM calibration procedure (Sect. 4.2) 100 times with a different SFR_{cal} basis (i.e., randomly re-extracting the 6,400 *bona fide* galaxies) and derive SFR_{SOM} every time. Shaded areas (from dark red to orange) show the fractional error embedding 50, 75, and 90 percent of the COSMOS-like sample as a function of redshift. Solid lines show the same quantities computed for star forming galaxies only ($sSFR_{sim} > 10^{-10}\,yr^{-1}$).

smaller than 0.15 dex. The two systematics have different explanations. The trend in the low-SFR_{sim} regime is due to z_{SOM} outliers: because of SED degeneracy, a few galaxies[12] already in the red sequence are miss-classified in cells mostly occupied by dusty star forming galaxies at higher z. This bias should diminish in deeper surveys, as they better disentangle redshift degeneracies. On the opposite hand, the bending of the SFR_{SOM} vs SFR_{sim} relation for the most star forming objects is a consequence of the intrinsic SFR distribution inside those cells. Galaxies with the highest activity are in the tail of such distribution so it is unlikely that one of them is selected for calibration. The reference SFR_{cal} extracted in those cells is usually 0.1–0.2 dex lower than the maximum, explaining the underestimate shown in Fig. 9.

We also investigate the statistical error due to the way to assemble the calibration sample. Depending on the selected *bona fide* galaxy, the SFR_{cal} label in a given cell may change significantly. A Monte Carlo simulation can quantify this uncertainty. We produce 100 calibration samples of the COSMOS-like SOM, each time randomly extracting a different set of *bona fide* galaxies. The standard deviation of the SFR_{SOM} estimates (σ_{SFR}) within the 100 realisations is calculated as a function of redshift. At $z_{sim} > 2$, most of the galaxies have σ_{SFR}/SFR_{sim} smaller than 30 per cent (Fig. 10). At lower redshift, ~70 per cent of the sample still shows such a small uncertainty, while the other galaxies are more sensitive to the SFR_{cal} extraction (although their errors do not exceed a factor ~2). The latter ones mainly include quiescent or post-starburst galaxies in cells with a large spread in SFR (see Fig. 10). With the same Monte Carlo we also quantify the z_{SOM} error which is always below 4 per cent.

[12] Note that the spatial density map used in Fig. 9 is in logarithmic scale.

5 APPLICATION OF THE SOM ESTIMATOR TO PRESENT AND FUTURE SURVEYS

So far we have applied the new method without discussing how to build its calibration sample. The *bona fide* galaxies can be observed with present or future facilities to collect robust measurements for their redshift and SFR. We model two possible surveys to this purpose in Sect. 5.1 and then discuss their selection effect (Sect. 5.2). For each "pseudo-survey" we focus on the observational uncertainties expected in SFR_{cal} since z_{cal} errors have a negligible impact on the SFR_{SOM} estimates. Eventually, we compare these results (more realistic than Sect. 4.2) to standard template fitting (see Sect. 5.3).

5.1 How to build the SFR calibration sample?

We discuss the realisation of a calibration sample within the Horizon-AGN framework to be consistent with the rest of our analysis. Thanks to the wealth of observations in the COSMOS field a similar attempt can also be made using real data, although with some limitations; we postpone this test to future work. We propose two alternate calibrations for our method:

C1) a spectroscopic follow-up targeting one galaxy per cell, to derive SFR_{cal} from their Hα flux;

C2) a combination of UV and IR imaging covering a portion of the field, providing SFR for several galaxies per cell via energy balance equation.

The SFR indicator adopted in the former scenario usually follows Kennicutt (1998):

$$SFR(H\alpha) = 5.4 \times 10^{-42} \frac{L_{H\alpha}}{\mathrm{erg\,s^{-1}}}\ M_\odot\ \mathrm{yr}^{-1}, \quad (5)$$

in which the original coefficient (7.9×10^{-42}) has been converted to Chabrier's IMF. $L_{H\alpha}$, namely the luminosity of the Hα line, must be corrected for dust absorption. This correction can be done e.g. by using the Balmer decrement:

$$E(B-V) = \frac{E(H\beta - H\alpha)}{k(H\beta) - k(H\alpha)}. \quad (6)$$

The numerator on the right-hand side of Eq. (6) is the colour excess due to dust reddening (see equation 2 in Moustakas et al. 2006) while the denominator comes from an attenuation function $k(\lambda)$, e.g. the one in Cardelli et al. (1989). For the C2 version there are different approaches in the literature to derive SFRs from UV+IR luminosity. The one used in Arnouts et al. (2013) is based on the formula

$$SFR(\mathrm{NUV, IR}) = 8.6 \times 10^{-11} \frac{L_{\mathrm{IR}} + 2.3 L_{\mathrm{NUV}}}{\mathrm{erg\,s^{-1}}}\ M_\odot\ \mathrm{yr}^{-1}, \quad (7)$$

where L_{IR} is the total IR luminosity (8–1000 μm) and L_{NUV} is the monochromatic luminosity in the near-UV rest frame filter. The IR luminosity accounts for the new-born stars enshrouded by dust that do not contribute to the NUV term. Different dust-correction calibrations have been proposed for Eq. (7), also depending on the observations used as a proxy for L_{IR} (see Hao et al. 2011). Further details about these SFR indicators can be found in Kennicutt & Evans (2012) and references therein.

Given its advantages in terms of computational speed, the proposed ML method represents and effective tool for the future large-area surveys. In this perspective, we design each calibration sample to be assembled by means of next-generation facilities. For instance the spectroscopic survey required for C1 could be carried out in the optical with the 4-metre Multi-Object Spectroscopic Telescope (4MOST, de Jong et al. 2019) and up to 1.8 μm with the Multi-Object Optical and Near-infrared Spectrograph (MOONS, Taylor et al. 2018). 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. Thanks to their unparalleled multiplexing (placing 10^2 to 10^3 fibers simultaneously) it should be feasible to observe the several hundreds of *bona fide* galaxies needed to label the SOM cells, providing not only SFR measurements but also spectroscopic redshifts (z_{spec}). The large field of view, especially of 4MOST, is also beneficial, since in this case targets are spread across 1 deg^2 area. However these are not the only options: for instance the Prime Focus Spectrograph at Subaru (PFS, Takada et al. 2014) will have similar capabilities. In principle the *James Webb* Space Telescope (JWST) could also be considered for near-IR spectroscopy, although it is not optimised for surveying a large number of targets distant from each other (see the discussion in Davidzon et al. 2018).

To realise the C2 sample, one option could be FIR observations from the proposed SPICA observatory[13] or Origins[14] mission, both expected to launch in the 2030s. We can imagine to using these telescopes to scan ~0.1 deg^2 of our field, probing the wavelength range between 20 and 230 μm. This would result in robust L_{IR} estimates up to $z \sim 3$ (Gruppioni et al. 2017; Kaneda et al. 2017). To complete Eq. (7) with rest-frame NUV luminosity one can assume to rely on GALEX data at $z < 0.5$ (see Arnouts et al. 2013) and deep u and B photometry at higher redshift. In the future those data could be superseded by higher-resolution photometry from CASTOR[15] and from the Large Synoptic Survey Telescope (LSST, LSST Science Collaboration et al. 2009). We stress out that all these next-generation facilities are expected to observe COSMOS as one of their calibration deep fields (Capak et al. 2019), so it is fair to assume that our COSMOS-like simulation can benefit from them.

Another important detail is the target selection strategy. The C1 campaign is a "blind" follow-up that randomly selects (at least) one photometric galaxy per SOM cell. With the next-generation spectrographs mentioned above it should be feasible probing the whole SOM space in a limited number of nights: for example 4MOST will have ~2,400 science fibers (which can be placed within a 4 deg^2 field of view) vs the 6,400 cells to be targeted. On the other hand, the photometric data provided in C2 must be complemented by reliable redshifts to estimate galaxy rest-frame luminosity. In the assumption of using SPICA, this shall result from its FIR high sensitivity grating spectrometer. Moreover, we argue that this would not be an issue even in case of a pure photometric follow-up. In fact, following the parallelism between our simulated lightcone and the actual COSMOS one, we can assume that there is an extended number of z_{spec} already covering our field. In this regard COSMOS has a unique z_{spec} catalogue thanks to the ongoing survey for a Complete Calibration of the Color-Redshift Relation (C3R2, Masters et al. 2017). This programme is systematically collecting spectroscopic information in each cell of the COSMOS SOM (already covering ~75 per cent of it, Masters et al. 2019). Besides C3R2, other z_{spec} surveys are expected in the near future (e.g. with PFS) since COSMOS is among the "calibration fields" of several future missions. For example we can suppose that the Horizon-AGN lightcone, like the real COSMOS field, is within the *Euclid* mission footprint, so *Euclid* grism redshifts will also be available. *Euclid* will collect spatially resolved Hα fluxes from $z = 0.9$ to 1.8 (down to $0.5-3 \times 10^{-16}$ erg cm^{-2} s^{-1}, Pozzetti et al. 2016) that can be used e.g. for aperture correction calibration of the multi-slit instruments.

[13] SPace Infrared telescope for Cosmology and Astrophysics, https://spica-mission.org/
[14] https://asd.gsfc.nasa.gov/firs/
[15] CASTOR is the Cosmological Advanced Survey Telescope for Optical and ultraviolet Research proposed by the Canadian Space Agency (Côte et al. 2012). This satellite could launch as early as 2027, surveying the UV with a ×30 better resolution than GALEX and a ×100 larger field of view than HST.

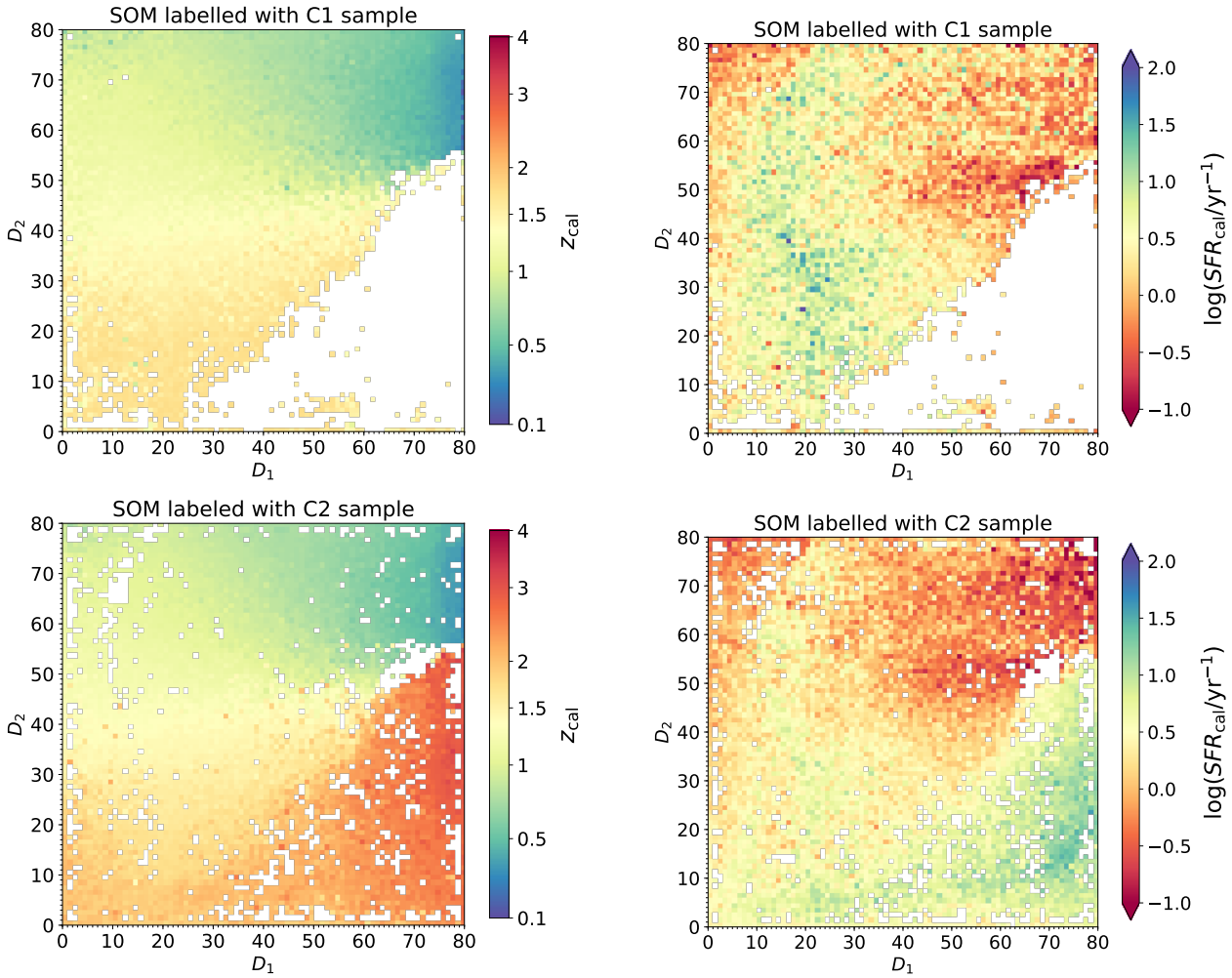

Figure 11. Horizon-AGN SOM trained with COSMOS-like colours and calibrated to work as a redshift and SFR estimator. *Upper panels:* the SOM is labelled according to the redshift (z_{cal}, *left panel*) and SFR (SFR_{cal}, *right*) of 4,749 "spectroscopic" galaxies coming from a pseudo-survey of calibration (C1). White pixels in the colour map are empty cells not covered by the C1 sample. *Lower panels:* in this case the SOM is labelled according to z_{cal} and SFR_{cal} values (*left* and *right* panel, respectively) coming from an alternate calibration sample (C2). The 37,780 galaxies in this sample come from an area of 0.1 deg^2 within the Horizon-AGN lightcone. Their SFR_{cal} is assumed to be measured from their UV and IR luminosity, that in principle could be obtained with a deep pencil-beam imaging survey.

5.2 Observational bias in the calibration samples

The pseudo-surveys described above would suffer from different selection effects. We model some of them to make each calibration sample more realistic and eventually get the SFR_{SOM} of Horizon-AGN galaxies closer to the estimates expected from observations. As stated above, we focus on SFR measurements because typical spectroscopic redshift errors are sub-dominant in the present analysis (therefore we assume $z_{spec} \equiv z_{sim}$ in the following).

For C1 we select spectroscopic targets among the brightest galaxies in each cell, i.e., objects brighter than $\langle i^+ \rangle^{cell}$ to enhance the probability of a $S/N > 10$ line detection in reasonable exposure time. We assume that the C1 survey can detect galaxies with Hα flux $F_{H\alpha} > 2{\times}10^{-17}$ erg cm^{-2} s^{-1} and $z < 1.7$. The redshift limit is due to the hypothesis that the reddest filter available for spectroscopy is in the H band. The $F_{H\alpha}$ threshold comes from a rough approximation of the sensitivity of future instruments resulting in $S/N > 10$. In practice, since nebular emission is yet to be added to their spectra, we predict $F_{H\alpha}$ of Horizon-AGN galaxies from their SFR_{sim} and $E(B-V)_{sim}$ by inverting Kennicutt's relation (Eq. 5 and 6).

With these constraints we can label 4,749 cells using C1 *bona fide* galaxies. For each of them we obtain a logarithmic SFR_{cal} estimate by perturbing the original $\log(SFR_{sim}/M_\odot/yr^{-1})$ with random Gaussian noise. The Gaussian standard deviation is set to $\sigma = 0.18$ dex from comparison to state-of-the-art surveys (e.g. FMOS-COSMOS, Kashino et al. 2019). The dominant source of error is the aperture correction, which in future surveys may be mitigated by larger slits[16] 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

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

stellar extinction (see Kashino et al. 2013). The expected impact of these biases is discussed in Sect. 5.3. In Fig. 11 the area of the SOM covered by the C1 sample is highlighted with their $z_{\rm cal}$ (upper-left panel) and $\rm SFR_{cal}$ values (upper right). The latter illustrates the solidity of manifold learning: despite the stochastic $\rm SFR_{cal}$ fluctuations due to the random selection of these galaxies, the SOM labels still show a clear evolutionary pattern. In addition, magnitude and redshift distributions are shown in Appendix A3.

In the case of calibration C2, a $19' \times 19'$ patch of the Horizon-AGN area is "observed" in UV and FIR. The patch is choosen to be the bottom-right corner of the lightcone, but results are not affeted if its location changes. 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. For sake of simplicity we do not attempt to reconstruct galaxy $L_{\rm UV}$ and $L_{\rm IR}$ in order to directly apply Eq. (7). Instead we use the median $\rm SFR_{sim}$ of C2 galaxies in the same cell, and perturb it with Gaussian noise ($\sigma = 0.1$ dex, see Ilbert et al. 2015) to produce the final $\rm SFR_{cal}$. This is expected to be a good proxy of the UV+IR estimator, whose timescale is similar to the 100 Myr interval used to define $\rm SFR_{sim}$ in Horizon-AGN (it is also comparable with Hα-derived measurements, see Kashino et al. 2019).

In a traditional analysis a key concern in the design of C2 would be cosmic variance uncertainty, hereafter referred as "sample variance"[17]. 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 specific types of objects may be over-represented and dominate the measured density.

Buchs et al. (2019) does an extensive analysis of how sample variance would affect mappings like the one we propose, 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 this method. However, we are limited by the box size of Horizon-AGN and the $19' \times 19'$ sub-region we have seletced 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 show 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). Sample variance can be appreciated also in the redshift and magnitude distributions shown in Appendix A3. We also note that the SFR map from C2 calibration (Fig. 11, lower-right panel) differs from C1 (Fig. 11, upper-right) because in the latter the $F_{H\alpha}$ threshold favourites the targeting of more actively star-forming galaxies.

[17] 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.

5.3 SFR estimates and comparison with template fitting

Once the SOM has been labelled either with C1 or C2 *bona fide* galaxies (Fig. 11) we apply the procedure described in Sect. 4.2 to assign a $\rm SFR_{SOM}$ estimate (Eq. 4) to each photometric galaxy. Some distinctions should be noticed with respect to the test shown in Sect. 4.2, where the calibration has been done with one random galaxy per cell and no errors in their $\rm SFR_{cal}$. Here the C1 calibration implies a narrower redshift range (we restrict the analysis at $z < 1.5$) while C2 differs from the original test because $\rm SFR_{cal}$ is calculated by stacking several objects per cell instead of considering a single galaxy.

The outcomes can be compared to that obtained in Paper I by using LePhare. We do not show the $z_{\rm SOM}$ vs $z_{\rm sim}$ comparison as the trend is similar to Fig. 8 (upper panel). Despite the additional uncertainties introduced in Sect. 5.2, the figure of merit does not change. For the C2 calibration, which covers the same redshift range of Fig. 8, $\rm NMAD(\mathit{z}_{SOM})$ and outlier fraction remain 0.043 and 6 per cent. In the following we focus on the $\rm SFR_{SOM}$ results, which show remarkable improvement with respect to template fitting estimates ($\rm SFR_{phot}$).

We remind that the estimate of physical properties via template fitting involves a two-step procedure. First, to find their $z_{\rm phot}$, LePhare fits galaxy SEDs with a composite set of templates (described in Laigle et al. 2016); then, after fixing the redshift of each galaxy to $z_{\rm phot}$, the code calculates the SFR (along with stellar mass and other physical quantities) by means of another SED library, this time made from BC03 models. As the mock galaxies has COSMOS-like features, the configuration of LePhare for Horizon-AGN is the same used for the COSMOS2015 catalogue; more details about LePhare input parameters and results can be found in Paper I.

The upper panels of Fig. 12 shows the comparison between $\rm SFR_{phot}$ and $\rm SFR_{sim}$ in different bins from $z_{\rm phot} = 0.2$ to 3. Under- and over-estimates are clearly visible, generating a two parallel sequences in the distribution. Such a bimodality is due to SED fitting degenarcies. By comparing the performance of LePhare in a (COSMOS-like) dusty vs dust-free Universe, Paper I isolated the major role of dust attenuation in driving this bimodality. In particular, the choice of the extinction curves used to build the library of templates is pivotal. This remains true even when the redshift is fixed to its intrinsic value instead of $z_{\rm phot}$. Inadequate extinction models or $E(B-V)$ values may cause indeed an overestimation or an underestimation of the SFR, as shown in Appendix B of Paper I.

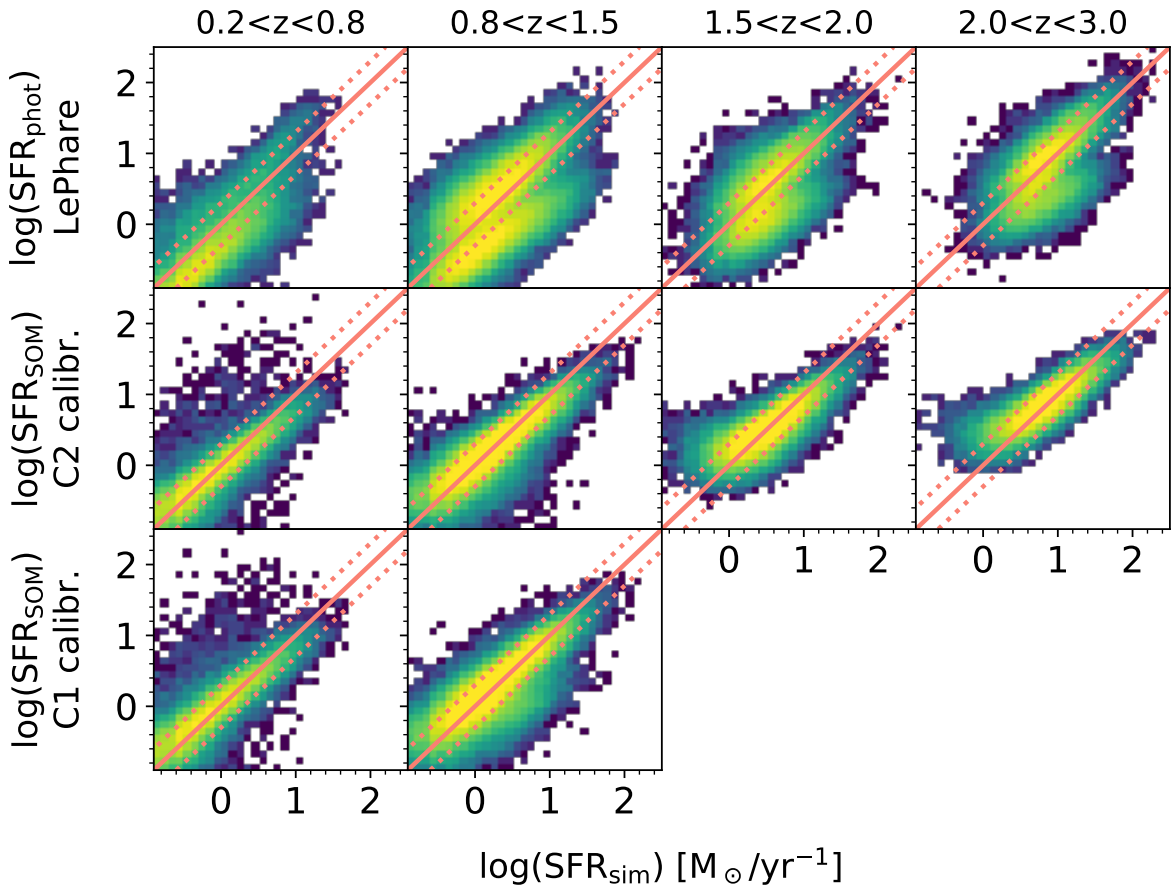

Figure 12. Comparison between intrinsic SFR_{sim} vs estimates obtained in from different methods: standard template fitting using the code LePhare (*upper row of panels*); using the SOM method with the calibration described as version C2 in the text (*middle panels*); SOM method with a different calibration sample, referred as C1 in the text (*lower panels*). Galaxies are binned in the different redshift bins according to either their z_{phot} from LePhare or z_{SOM} in the case of the new method proposed here. Version C1 is limited to the first two z bins because of the way the calibration sample is constructed. In each panel the solid line indicates the 1:1 relationship and the dotted lines a ±0.15 dex offset from it.

The same photometric galaxies[18] are reported in the middle and lower panels of Fig. 12 comparing intrinsic SFRs with the values computed through the SOM (C2 and C1 version respectively). The performance of the SOM is significantly better than LePhare. The SFR_{SOM} distribution does not show the same bimodality observed for the SFR_{phot} estimates because the SOM fitting is based only on *observed* SEDs, which naturally include the "correct" dust attenuation law and $E(B-V)$ range. In LePhare a grid of templates is built without strong observational priors so the library is affected by artificial degeneracies.

Also our method is model dependent, not for the SOM itself but because of the SFR_{cal} labels. These measurement requires some theoretical prescription (e.g., about dust attenuation or IMF). However we argue that the required assumptions, for either C1 or C2, introduce a milder bias than template fitting. For instance the C1 calibration requires the choice of a dust extinction curve (Eq. 6) but the difference between models is small: e.g., $k(H\beta) - k(H\alpha) = 1.07, 1.27$ for Cardelli et al. (1989) and Calzetti et al. (2000) respectively. On the contrary LePhare templates are constrained also by data at bluer r.f. wavelengths where the dust extinction curve plays a more important role (Ilbert et al. 2009); for the same models in the example above: $k(2000) - k(3000) = 3.18$ or 1.95.

With respect to the C2 sample one may notice that the energy balance equation is also implemented in template fitting codes (e.g. MAGPHYS, da Cunha et al. 2008) sometimes in very elaborated ways including also the AGN contribution (e.g. SED3FIT, Berta et al. 2013). In fact one may also use one of those codes instead of Eq. 7 but not for the other Horizon-AGN galaxies that do not belong to C2. Moreover, even if the whole galaxy sample were observed in UV and FIR, codes like MAGPHYS are extremly expensive in terms of computational time and run only after fixing the redshift. This 2-step fitting procedure, which is of widespread use in the literature, raises several issue (e.g., propagation of z uncertainties, Grazian et al. 2015; Davidzon et al. 2017). On the other hand, the SOM does not require such a procedure providing z and SFR estimates simultaneously.

Fig. 13 proposes the same comparison of Fig. 12

[18] All the *bona fide* galaxies used for calibration have been excluded from the comparison.

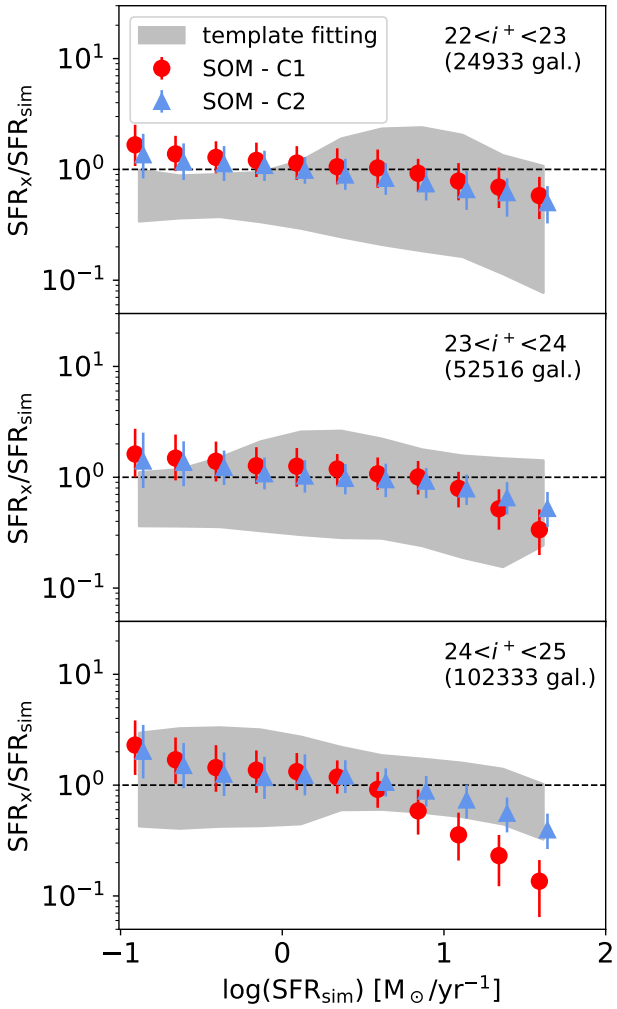

Figure 13. Median offset between intrinsic SFR and SOM-based estimates, for Horizon-AGN galaxies with COSMOS-like photometry (each panel showing a different i^+-band magnitude range). The SFR_{SOM} estimates are derived wither using the SOM calibration C1 (red circles) or C2 (blue triangles); symbols are horizontally shifted of −0.03 and +0.03 dex respectively for sake of clarity. Error bars delimit the 16th-84th percentile range of each SFR_{SOM}/SFR_{sim} distribution. Estimates from LePhare are also included, but given the bimodality of SFR_{phot}/SFR_{sim} (see Fig. 9, upper panels) The figure shows only the intervale between 16th and 84th percentile in running bins of SFR_{sim}. In each panel a dotted line marks the 1:1 ratio.

in a different flavour, i.e. showing the median ratio SFR_{SOM}/SFR_{sim} in three bins of apparent magnitude from $i^+ = 22$ to 25. At $i^+ < 23$ (Fig. 13, middle panel) there is an excellent agreement of both C1 and C2 estimates with the intrinsic SFR. Such a trend is still observed at $23 < i^+ < 24$ even though the most star forming galaxies start to be systematically underestimated by more than a factor 2. This is a border effect inherent to the SOM analysis already discussed in Sect. 4.2 and having a larger sample that allowed for a more representative SOM would mitigate this effect (Buchs et al. 2019). The discrepancy becomes more accentuated at $24 < i^+ < 25$ (Fig. 13, lower panel) especially for the C1 calibration that by construction relies on *bona fide* galaxies systematically brighter than the average. This could also be accounted for by better sampling these cells (see Masters et al. 2019). Concerning LePhare, under- and over-estimates would compensate each other resulting in a misleading $SFR_{phot}/SFR_{sim} \simeq 1$. Therefore Fig. 13 does not show the median of the SFR_{phot} distribution but only the interval between the 16th and 84th percentile. Such a dispersion is significantly larger than the ML estimates in all the magnitude bins.

6 SUMMARY AND CONCLUSION

Compared to the large number of studies measuring galaxy redshifts with ML techniques, little progress has been made concerning other physical parameters. In spite of that, ML methods will be pivotal in the near future to derive stellar mass, star formation history, and other galaxy properties in extremely large data sets from surveys such as *Euclid* and LSST. In addition to their unprecedented speed, these algorithms (particularly the unsupervised ML methods) may lead to a "new paradigm" in which human intervention (i.e., the application of interpretative models) starts after galaxy classification and demographics have been decided by the machine. However, results may be affected by new kinds of systematics introduced, e.g., during the data reduction process or the training set selection. A thorough investigation of ML performance and the role of its "observational priors" is thus imperative before such high expectations may be deemed justified.

With this perspective, we have explored advantages and limitations of the SOM as a galaxy parameter estimation tool independent of model templates. We chose the SOM because it is an unsupervised dimensionality reduction algorithm able not only to learn the complex structure of data but also to project it in a lower-dimensional space (2-d in our analysis) still preserving its "topological" features. It should be clarified that our goal is not advocating for the SOM to replace standard template fitting: the two *complementary* approaches should be used in synergy, the same way semi-analytic and hydrodynamical simulations have contributed to inform each other and together improve our understanding of galaxy evolution.

We tested the SOM with a mock galaxy catalogue (presented for the first time in Laigle et al. 2019, Paper I) derived from the Horizon-AGN hydrodynamical simulation. Galaxies cover 1 deg^2 area and a redshift range $0 < z < 4$, with a photometric baseline similar to state-of-the-art surveys (broad-band filters from u to [4.5μm]). The SOM has been trained using as input only galaxy colours, to be an analogue of "classical" SED fitting codes like LePhare or EAZY (Brammer et al. 2008). In principle other (e.g., morphological) features may be trained for, but this is left to future work. After classifying the mock galaxies in about 6,400 different classes (called "cells" in our jargon) we explored the connections between the class/cell a galaxy belongs to and its physical properties. Then, we calibrated *a posteriori* the SOM by labelling each cell with a value of z and SFR, so that other galaxies in the same cell may have a proxy of their own redshift and star formation activity. Eventually we used the calibrated SOM to estimate the SFR for a sub-sample

of about 375,000 mock galaxies between $z \sim 0$ and 3. Our findings are summarised in the following.

- The SOM is an effective tool to visualise the characteristics of a complex, *non-linear* manifold as the Horizon-AGN lightcone. Galaxies are organised in a (human-readable) 2-d grid without smearing out the features of their original parameter space. Moreover this is computational inexpensive, suggesting a convenient way to describe and inspect the properties of extremely large simulations (Mitra et al. 2015).
- Since in our case the parameter space is made by observer's frame colours, the SOM works like an SED fitting algorithm without a pre-compiled library of templates: the SOM adapts its cells/weights to the data so that galaxies with similar colours (i.e., similar SEDs) end in the same cell. We find that also the high-resolution spectra turn out to be nicely classified, in spite of using only broad-band photometry for training.
- We confirm that objects in the same cell have similar redshift (as shown in Masters et al. 2015 in the observed universe) but we also find that their M/L and sSFR is similar, with typical scatter of 0.15 and 0.3 dex respectively. Also M and SFR, after taking into account a normalisation factor, are well correlated to the cell where a galaxy lies. After including photometric uncertainties (modelled after the COSMOS2015 survey) and rejecting objects with $S/N < 1.5$ in any band, we trained again the SOM: the correlation between galaxy properties and cells was still present, although with larger scatter. This indicates that our analysis can be reproduced in real optical-NIR surveys (provided a sufficient depth of the observations).
- We have measured the redshift of COSMOS-like ($S/N > 1.5$) galaxies through the SOM, finding a fairly good agreement with intrinsic $z_{\rm sim}$ but a larger scatter than template fitting: The $z_{\rm SOM}$ vs $z_{\rm sim}$ comparison results into NMAD = 0.044 and about 6 per cent of catastrophic errors, whereas with LePhare the metrics is 0.024 and 1.5 per cent. On the other hand the redshift bias is significantly smaller in the SOM case (-0.001, compared to -0.011 in LePhare). We considered such a result sufficiently good for our purposes so we did not attempt to improve the redshift estimator (as done e.g. in Buchs et al. 2019).
- Leveraging these SOM properties, we have developed a new SFR estimator for photometric galaxies. We have assumed that a small fraction of them (10 per cent or less) has been followed-up to serve as a calibration sub-sample, providing labels ($z_{\rm cal}$, $\mathrm{SFR}_{\rm cal}$) to the SOM cells. We have discussed possible follow-up strategies with optical-NIR spectroscopy or with UV+FIR instruments, and the possible bias introduced by either of them. After accounting for such uncertainties we have compared the $\mathrm{SFR}_{\rm SOM}$ of COSMOS-like galaxies with their intrinsic $\mathrm{SFR}_{\rm sim}$. Overall the dispersion (defined as the range between 16th and 84th percentile in logarithmic bins of SFR) is ± 0.2 dex, with a small systematic offset (median $\log(\mathrm{SFR}_{\rm SOM}/\mathrm{SFR}_{\rm sim}) \simeq 0.02 - 0.04$ dex). The most active galaxies are an exception, being significantly underestimated because they are in cells whose majority of objects (including the calibration ones) are less star forming.
- LePhare $\mathrm{SFR}_{\rm phot}$ estimates are also available in Horizon-AGN and we have compared them to the new indicator. The latter performs remarkably better: $\mathrm{SFR}_{\rm SOM}$ are more precise but also significantly less biased, as they do not rely on a template library that introduces *artificial degeneracies* in the SED fitting (as investigated in Paper I).

The suboptimal performance of the SOM as a redshift machine found in this analysis is partly due to the fact that we have not entirely followed Masters et al. (2015) prescriptions, e.g. we did not use cell occupation as a prior nor we distinguished a deeper calibration sample from the rest of the survey (or use a large survey of multiple fields). We note that the comparison is not straightforward since Masters et al., working with observed galaxies, are forced to use a spectroscopic sub-sample that is biased to some extent (see Paper I). However, as highlighted in Masters et al., on that spectroscopic sample their $z_{\rm SOM}$ figure of merit is better than LePhare.

On the other hand, the better performance of our SOM method to compute SFR does not imply that it is bias-free: some systematics may be introduced while selecting the calibration galaxies and measuring their $\mathrm{SFR}_{\rm cal}$. We argue however that model assumptions in the $\mathrm{SFR}_{\rm cal}$ calculation are generally less severe than those involved in the construction of libraries from stellar population synthesis models, with a coarse grid of $E(B-V)$ values, simplistic SFHs, fixed stellar metallicity, and other limitations to which SFR is sensitive (Papovich et al. 2001, see also discussion in Paper I). Moreover if the sub-sample used for calibration turns out to be strongly biased it can be replaced by a better one without re-classifying the target galaxies, while any improvement in the template library of LePhare would require to run again that (computationally expensive) code over the whole catalogue. It should also be emphasised that estimates of redshift and physical parameters are provided simultaneously – a unique advantage of the SOM method that in future developments shall allow for a better treatment of z error propagation.

We aim at transferring our method to the real COSMOS catalogue in the next paper of this series, even though data available in that field may be able to calibrate the SFR only in a limited portion of the SOM. Nonetheless, this effort can result in an original comparison between different estimators. For example, one could derive SFR from radio continuum stacking (as in Karim et al. 2011) vs UV+IR luminosity (as in Ilbert et al. 2015) for galaxies in the same cells, easily identifying the region of the parameter space where the indicators disagree.

The present work is also intended to provide a new science case for upcoming large-area surveys. The SOM requires an accurate calibration sample relatively modest in size (but large enough to limit sample variance effects, Buchs et al. 2019)) and then billions of galaxies (e.g. from the 15,000 deg^2 of *Euclid*) can be efficiently mapped to get a proxy for their redshift *and* physical properties. This is particularly true for the *Euclid* Deep Fields, which will have a photometric baseline similar to the one assumed here thanks to the complementary surveys in optical (Hawaii two-O, PI: D. Sanders) and MIR (*Euclid*/WFIRST *Spitzer* Legacy Survey, PI: P. Capak). We also mentioned the contribution that 4MOST, MOONS, and PSF may provide to calibrating the galaxy colour space, owing to their unprecedented multi-

plexing. Our case study also supports the concept of a deep surveying of COSMOS with CASTOR, SPICA, and Origins, to continue its use as a reference field for the coming decades.

ACKNOWLEDGEMENTS

ID thanks Stefano Andreon, Sirio Belli, Micol Bolzonella, Keerthana Jegatheesan, Chris Hayward, and Lucia Pozzetti for useful discussions. CL is supported by a Beecroft Fellowship. OI acknowledges the funding of the French Agence Nationale de la Recherche for the project "SAGACE". This research was supported in part by the National Science Foundation under Grant No. NSF PHY-1748958 and by the NASA ROSES grant 12-EUCLID12-0004. The analysis presented in this work relied on the HPC resources of CINES (Jade) under the allocation 2013047012 and c2014047012 made by GENCI and on the Horizon and CANDIDE clusters hosted by Institut d'Astrophysique de Paris. We warmly thank S. Rouberol for maintaining these clusters on which the simulation was post-processed. This research is part of ERC grant 670193 and HORIZON-UK *and is also partly supported by the Centre National d'Etudes Spatiales (CNES). Several* `python` *packages were used, including* ASTROPY *(Astropy Collaboration et al. 2013; Price-Whelan et al. 2018) and* SOMPY *(main contributors: Vahid Moosavi, Sebastian Packmann, Iván Vállas).*

REFERENCES

Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540
Arnouts S., et al., 2013, A&A, 558, A67
Astropy Collaboration et al., 2013, A&A, 558, A33
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bell E. F., de Jong R. S., 2000, MNRAS, 312, 497
Benítez N., 2000, ApJ, 536, 571
Berta S., et al., 2013, A&A, 551, A100
Bolzonella M., Miralles J.-M., Pelló R., 2000, A&A, 363, 476
Bonjean V., Aghanim N., Salomé P., Beelen A., Douspis M., Soubrié E., 2019, A&A, 622, A137
Bonnett C., 2015, MNRAS, 449, 1043
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buchs R., et al., 2019, arXiv e-prints,
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Capak P., et al., 2019, arXiv e-prints,
Cappellari M., 2017, MNRAS, 466, 798
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Carrasco Kind M., Brunner R. J., 2014, MNRAS, 438, 3409
Chabrier G., 2003, PASP, 115, 763
Chatfield C., Collins A. J., 1980, Introduction to Multivariate Analysis, 1 edn. Chapman & Hall, London, doi:10.1007/978-1-4899-3184-9
Chevallard J., Charlot S., 2016, MNRAS, 462, 1415
Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486
Côte P., et al., 2012, in Space Telescopes and Instrumentation 2012: Optical, Infrared, and Millimeter Wave. p. 844215, doi:10.1117/12.926198
Courteau S., et al., 2014, Reviews of Modern Physics, 86, 47
Davidzon I., Bolzonella M., Coupon J., Ilbert O., Arnouts S., et al. 2013, A&A, 558, A23
Davidzon I., et al., 2017, A&A, 605, A70
Davidzon I., Ilbert O., Faisst A. L., Sparre M., Capak P. L., 2018, ApJ, 852, 107
Delli Veneri M., Cavuoti S., Brescia M., Longo G., Riccio G., 2019, MNRAS, 486, 1377
Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, MNRAS, 420, 2662
Dubois Y., et al., 2014, MNRAS, 444, 1453
Fontana A., et al., 2004, A&A, 424, 23
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Gavazzi G., Bonfanti C., Sanvito G., Boselli A., Scodeggio M., 2002, ApJ, 576, 135
Geach J. E., 2012, MNRAS, 419, 2633
Grazian A., et al., 2015, A&A, 575, A96
Gruppioni C., et al., 2017, Publications of the Astronomical Society of Australia, 34, e055
Hao C.-N., Kennicutt R. C., Johnson B. D., Calzetti D., Dale D. A., Moustakas J., 2011, ApJ, 741, 124
Hemmati S., et al., 2019a, arXiv e-prints,
Hemmati S., et al., 2019b, ApJ, accepted, arXiv:1808.10458
Hernandez-Pajares M., Floris J., 1994, MNRAS, 268, 444
Hoaglin D. C., Mosteller F., Tukey J. W., 1983, Understanding robust and exploratory data anlysis
Ilbert O., et al., 2006, A&A, 457, 841
Ilbert O., et al., 2009, ApJ, 690, 1236
Ilbert O., McCracken H. J., Le Fevre O., Capak P., Dunlop J., et al. 2013, A&A, 556, A55
Ilbert O., et al., 2015, A&A, 579, A2
Kaneda H., et al., 2017, Publications of the Astronomical Society of Australia, 34, e059
Karim A., et al., 2011, ApJ, 730, 61
Kashino D., et al., 2013, ApJ, 777, L8
Kashino D., et al., 2019, ApJS, 241, 10
Kennicutt Jr. R. C., 1998, ApJ, 498, 541
Kennicutt R. C., Evans N. J., 2012, ARA&A, 50, 531
Kewley L. J., Geller M. J., Jansen R. A., 2004, AJ, 127, 2002
Kohonen T., 1981, Technical Report Report TKK-F-A461, Hierarchical Ordering of Vectorial Data in a Self-Organizing Process. Helsinki University of Technology, Espoo, Finland
Kohonen T., 2001, Self-Organizing Maps, 3 edn. Springer Series in Information Sciences Vol. 30, Springer-Verlag, Berlin, doi:10.1007/978-3-642-97610-0
Komatsu E., Smith K. M., Dunkley J., et al. 2011, ApJS, 192, 18
LSST Science Collaboration et al., 2009, preprint, (`arXiv:0912.0201`)
Lagos C. d. P., et al., 2016, MNRAS, 459, 2632
Laigle C., et al., 2016, ApJS, 224, 24
Laigle C., et al., 2019, MNRAS, 486, 5104
Laureijs R., 2011, pp 1–116
Le Floc'h E., et al., 2009, ApJ, 703, 222
Leja J., Johnson B. D., Conroy C., van Dokkum P. G., Byler N., 2017, ApJ, 837, 170
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106
Maehoenen P. H., Hakala P. J., 1995, ApJ, 452, L77
Maraston C., 2005, MNRAS, 362, 799
Maraston C., Pforr J., Renzini A., Daddi E., Dickinson M., Cimatti A., Tonini C., 2010, MNRAS, 407, 830
Marchesini D., et al., 2010, ApJ, 725, 1277
Marchetti A., et al., 2013, MNRAS, 428, 1424
Masters D., et al., 2015, ApJ, 813, 53
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
Miller A. S., Coe M. J., 1996, MNRAS, 279, 293
Mitchell P. D., Lacey C. G., Baugh C. M., Cole S., 2013, MNRAS, 435, 87
Mitra S., Davé R., Finlator K., 2015, MNRAS, 452, 1184
Mobasher B., et al., 2015, ApJ, 808, 101
Molinari E., Smareglia R., 1998, A&A, 330, 447

Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, ApJ, 731, 113
Moustakas J., Kennicutt Jr. R. C., Tremonti C. A., 2006, ApJ, 642, 775
Oke J. B., 1974, ApJS, 27, 21
Pacifici C., Kassin S. A., Weiner B., Charlot S., Gardner J. P., 2013, ApJ, 762, L15
Papovich C., Dickinson M., Ferguson H. C., 2001, ApJ, 559, 620
Park M.-J., et al., 2019, arXiv e-prints,
Pérez-González P. G., Gil de Paz A., Zamorano J., Gallego J., Alonso-Herrero A., Aragón-Salamanca A., 2003, MNRAS, 338, 508
Poincot P., Lesteven S., Murtagh F., 1998, A&AS, 130, 183
Pozzetti L., et al., 2016, A&A, 590, A3
Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., 1992, Numerical recipes in C. The art of scientific computing
Price-Whelan A. M., et al., 2018, AJ, 156, 123
Rahmani S., Teimoorinia H., Barmby P., 2018, MNRAS, 478, 4416
Salvato M., Ilbert O., Hoyle B., 2019, Nature Astronomy, 3, 212
Sánchez C., Bernstein G. M., 2019, MNRAS, 483, 2801
Santini P., et al., 2017, ApJ, 847, 76
Sawicki M., 2012, PASP, 124, 1208
Schaerer D., de Barros S., 2009, A&A, 502, 423
Scoville N., et al., 2007, ApJS, 172, 38
Simet M., Chartab Soltani N., Lu Y., Mobasher B., 2019, arXiv e-prints, p. arXiv:1905.08996
Speagle J. S., Capak P. L., Eisenstein D. J., Masters D. C., Steinhardt C. L., 2016, MNRAS, 461, 3432
Stabenau H. F., Connolly A., Jain B., 2008, MNRAS, 387, 1215
Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, Astrophysical Journal Letters, 462, L17
Stensbo-Smidt K., Gieseke F., Igel C., Zirm A., Steenstrup Pedersen K., 2017, MNRAS, 464, 2577
Takada M., et al., 2014, PASJ, 66, R1
Taniguchi Y., et al., 2007, ApJS, 172, 9
Taylor W., et al., 2018, in Ground-based and Airborne Instrumentation for Astronomy VII. p. 107021G, doi:10.1117/12.2313403
da Cunha E., Charlot S., Elbaz D., 2008, MNRAS, 388, 1595
de Jong R. S., et al., 2019, The Messenger, 175, 3

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 $\sim 8{\times}10^7\,\mathrm{M}_\odot$ for dark matter particles. In Dubois et al. (2014) this translates into a lower limit of $10^8\,\mathrm{M}_\odot$ in galaxy stellar mass, while we stop at $M = 10^9\,\mathrm{M}_\odot$ 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 $10^9\,\mathrm{M}_\odot$ 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_{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 $M < 10^9\,\mathrm{M}_\odot$ (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 $10^{12}\,\mathrm{M}_\odot$, 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). Since galaxy SEDs in the same cell have similar shape, the normalisation factor (see Eq. 4) does not change significantly if another band is used (with the exception of the bluest filters). Fig. A4 shows this is the case for a K_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_s would produce additional scatter in the $\mathrm{SFR_{SOM}}$ vs $\mathrm{SFR_{sim}}$ comparison (cf. Fig. A4 with Fig. Fig. 9).

A3 Training sample selection

As mentioned in Sect. 4.1, the selection of COSMOS-like galaxies with $S/N > 1.5$ may also bias the results. First of all, we note that such a selection corresponds in practice to a magnitude cut $i^+ < 25$ (see Fig. A2, upper panel) with the

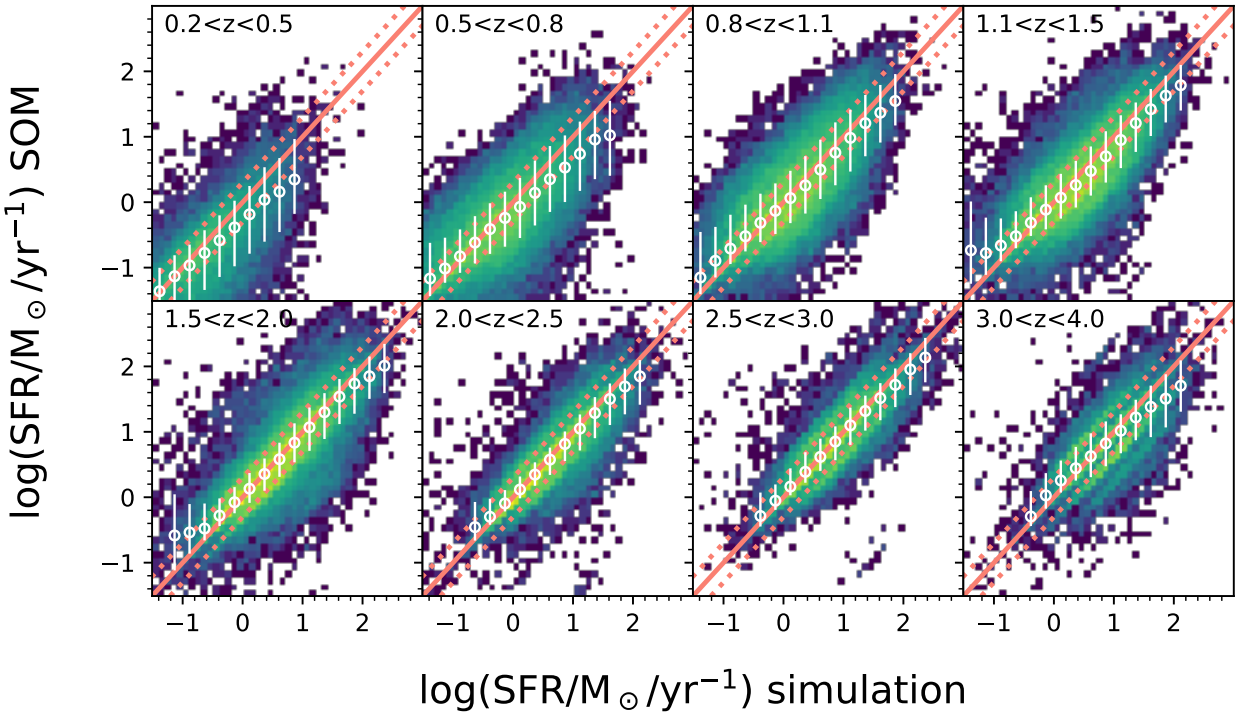

Figure A1. Comparison between the intrinsic SFR and the estimates derived from the SOM for 400,000 mock galaxies with $10^8 < M/M_\odot < 10^{12}$ and $0 < z < 4$. These objects are selected from an empirical simulation that reproduces the observed COSMOS2015 galaxies by means of their best-fit templates. 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.

removal (because of the required *u*-band detection) of $z > 3.5$ galaxies (Fig. A2, 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) 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. A3). 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_{SOM} bias shown in Fig. 8.

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

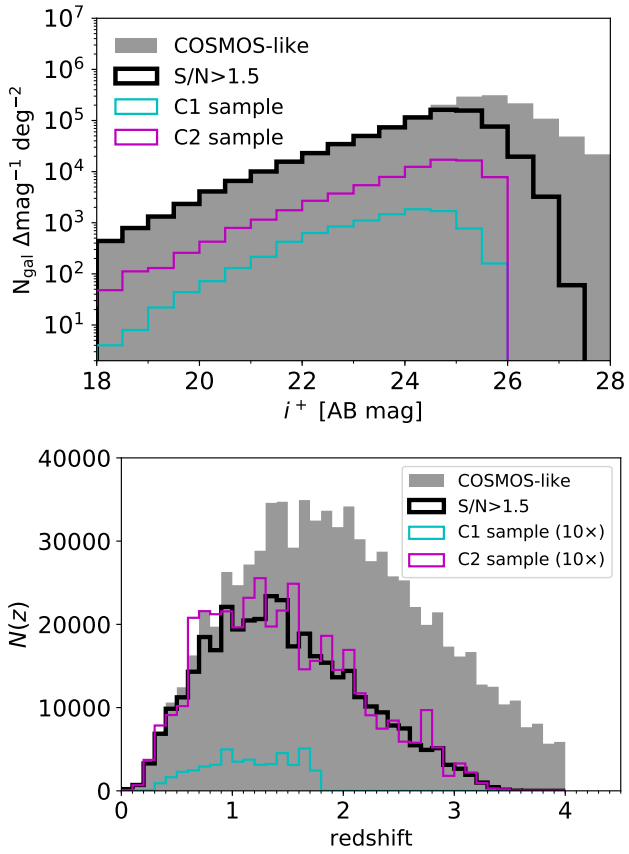

Figure A2. Magnitude and redshift distribution (*upper* and *lower* panel respectively) of Horizon-AGN galaxies with COSMOS-like uncertainties in their photometry. Both panels show the distribution of the parent sample (filled histogram), the S/N-selected sample (black line, see Section 4.1), and the 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.

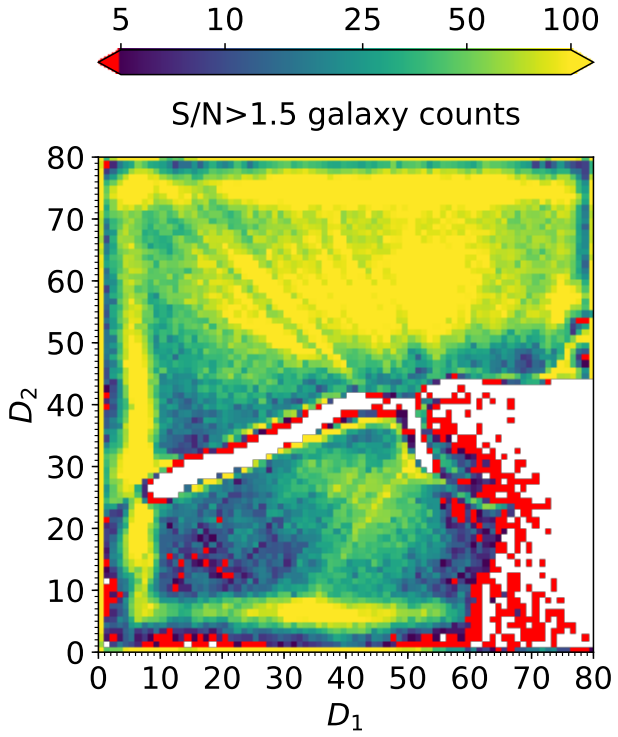

Figure A3. The Horizon-AGN SOM (same as Fig. 3) colour-coded according to the number of galaxies per cell with $S/N > 1.5$ in each band (white pixels indicate empty cells).

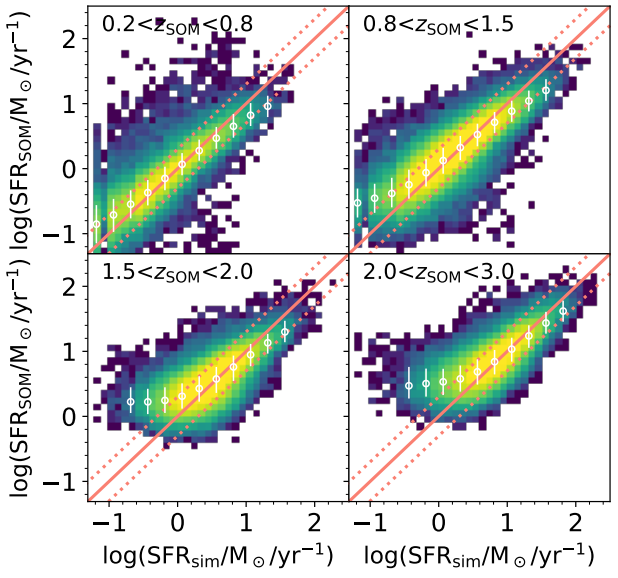

Figure A4. SFR_{SOM} vs SFR_{sim} comparison, simlar to Fig. 9 but normalising the SEDs as a function of their K_s flux instead of i^+.