BIRTH of the COSMOS Field: Primordial and Evolved Density Reconstructions During Cosmic High Noon

Metin Ata¹*, Francisco-Shu Kitaura²,³, Khee-Gan Lee¹, Brian C. Lemaux⁴, Daichi Kashino⁵, Olga Cucciati⁶, Mónica Hernández-Sánchez²,³ & Oliver Le Fèvre⁷

¹Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), WPI, The University of Tokyo Institutes for Advanced Study (UTIAS), The University of Tokyo, Kashiwa, Chiba, 277-8568, Japan
²Instituto de Astrofísica de Canarias (IAC), Calle Vía Láctea s/n, 38200, La Laguna, Tenerife, Spain
³Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38206, La Laguna, Tenerife, Spain
⁴Department of Physics, University of California, Davis, One Shields Ave., Davis, CA 95616, USA
⁵Department of Physics, ETH Zürich, Wolfgang-Pauli-Straße 27, CH-8093 Zürich, Switzerland
⁶INAF - Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, via Gobetti 93/3, 40129 Bologna, Italy
⁷Aix Marseille Université, CNRS, LAM (Laboratoire d’Astrophysique de Marseille) UMR 7326, 13388 Marseille, France

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
This work presents the first comprehensive study of structure formation at the peak epoch of cosmic star formation over 1.4 ≤ z ≤ 3.6 in the COSMOS field, including the most massive high redshift galaxy proto-clusters at that era. We apply the extended COSMIC BIRTH algorithm to account for a multi-tracer and multi-survey Bayesian analysis at Lagrangian initial cosmic times. Combining the data of five different spectroscopic redshift surveys (zCOSMOS-deep, VUDS, MOSDEF, ZFIRE, and FMOS-COSMOS), we show that the corresponding unbiased primordial density fields can be inferred, if a proper survey completeness computation from the parent photometric catalogs, and a precise treatment of the non-linear and non-local evolution on the light-cone is taken into account, including (i) gravitational matter displacements, (ii) peculiar velocities, and (iii) galaxy bias. The reconstructions reveal a holistic view on the known proto-clusters in the COSMOS field and the growth of the cosmic web towards lower redshifts. The inferred distant dark matter density fields concurrently with other probes like tomographic reconstructions of the intergalactic medium will explore the interplay of gas and dark matter and are ideally suited to study structure formation at high redshifts in the light of upcoming deep surveys.

Key words: cosmology: large-scale structure of Universe – cosmology: theory – galaxies: high-redshift – surveys

1 INTRODUCTION

Our current standard cosmological model predicts a hierarchical clustering of subsequently merging small structures to greater ones (see White & Rees 1978; Fry & Peebles 1978) up to the formation of the largest galaxy super-clusters observed in our present local Universe (e.g. Tully et al. 2014, with a total mass of ~ 10¹² M☉). The formation history and the exploration of the underlying physical phenomena of galaxy clustering over cosmic history remains an important question (see the pioneering work of Kauffmann et al. 1999, and references therein).

Furthermore, progenitors of galaxy clusters and super-clusters and their halos are key probes to understand early structure formation (Cohn & White 2008; Gao et al. 2018) and can be used to constrain a particular dark matter model (Bode et al. 2001). Also, the analysis of galaxy clusters allow us to study galaxy formation, test cosmological parameters (Allen et al. 2011) and constrain non-standard cosmological models (Kravtsov & Borgani 2012; Costanzi Alunno Cerbolini et al. 2013).

In particular, the range of 2 ≤ z ≤ 3 marks the peak epoch of star formation in the Universe, frequently referred to as “Cosmic High Noon” (Somerville & Davé 2015). The processes driving star formation of galaxies and clusters have been studied in previous works, showing a nontrivial relation of star formation and the galaxies’ environmental density (Cooper et al. 2008; Koyama et al. 2013; Kawinwanichakij et al. 2017; Muldrew et al. 2018; Ji et al. 2018). Moreover, the quenching of star formation within massive galaxy clus-
ters has been found (Cooper et al. 2008), suggesting on aver-
age a passive evolution of galaxy clusters during the last \( \approx 10 \) Gyr and studying an environmental dependency (see Lemaux et al. 2012; Belliore et al. 2017; Tomczak et al. 2019; Lemaux et al. 2019). Therefore, an accurate description of the dark matter density distribution at these redshifts will potentially help to understand the interconnection of star formation (see e.g. Behroozi et al. 2013; Madau & Dickinson 2014; Shi et al. 2019) and the location within the cosmic web (e.g. Bond et al. 1996). Detailed analyses of high redshift structures have been accessible via numerical simula-
tions (e.g. Watson et al. 2013; Henriques et al. 2014; Chiang et al. 2017), studying their formation history and the im-
portance of the environment.

While wide field galaxy surveys like SDSS-III Baryon Oscillation Spectroscopic Survey (BOSS) (Alam et al. 2017) and 2MASS (Skutskie et al. 2006) have played an important role in spatially mapping the large-scale structure of galaxies and quasars (see e.g. Ata et al. 2018), these have focused on scales of the baryon acoustic oscillations (BAOs) (Eisenstein et al. 2005; Percival et al. 2007). Some endeav-
ours have been done to map the structure down to megaparsec scales over to smaller footprints in so-called pencil-beams\(^1\), such as ALMA deep fields (Casey et al. 2018) reaching out to \( z \approx 1.5 \). However, no spectroscopic sur-
veys have abundantly mapped the large-scale density dis-
tribution down to megaparsec scales beyond \( z > 1.5 \), al-
though multiple projects aim to push beyond this boundary in the near future, such as DESI (Levi et al. 2013), EUCLID (Amendola et al. 2018), PFS (Takada et al. 2014), 4MOST (de Jong et al. 2012), and MOONS (Cirasuolo et al. 2014). For this reason we are still missing observational validation of quasi-linear structure formation at about \( z \approx 1.5 \), before non-linearities start dominating the emerging cosmic web.

Over nearly two decades, the Cosmic Evolution Survey (COSMOS) (see latest version Laigle et al. 2016) (Capak et al. 2007; Scoville et al. 2007) has been a major ongo-
ing effort to photometrically observe galaxies across a suf-
ciently wide footprint to resolve transverse large-scale struc-
ture, while simultaneously having sufficient depth to probe the entire span of cosmic history from the Local Universe right into (and eventually beyond) the Epoch of Reioniza-
tion (e.g. Scoville et al. 2013). Observations within COSM-
OS have identified large aggregations of galaxies at high redshifts (Yuan et al. 2014), i.e. galaxy proto-clusters, that have been extensively studied over a large redshift range at Cosmic Noon (Diener et al. 2013; 2015; Chiang et al. 2015; Casey et al. 2015; Wang et al. 2016; Lee et al. 2016; Cucciati et al. 2018; Lemaux et al. 2018; Darvish et al. 2020). These overdense structures have been proposed to be possible progenitors of the Coma-like galaxy clusters, potentially assuming total masses of \( \sim (1 - 2) \times 10^{15} M_{\odot} \) at \( z = 0 \) (Lee et al. 2016; Cucciati et al. 2018; Lemaux et al. 2018; Darvish et al. 2020). The COSMOS field is therefore ideally suited to study early structure formation and the evolution of galaxy proto-clusters into the mature structures we observe at the current epoch.

However, some of the aforementioned pioneering stud-

\(^{1}\)Pencil-beam surveys do not statistically resolve the transverse large-scale structures at \( \sim 10 \ h^{-1} \) Mpc scales.
The $z \sim 2.2$–2.8 LATIS (Lyα Tomography IMACS Survey) (Newman et al. 2020), based on the Lyman-$\alpha$ forest tomography technique (Pichon et al. 2001) and thus will provide insights into the relationship of matter clustering and the properties of the IGM. Other future applications include a direct study of galaxy properties as a function of underlying matter density and a full constrained $N$-Body simulation starting from the inferred initial conditions.

This article is structured as follows. In Section 2 we give a brief introduction to the recently developed initial density perturbations reconstructions algorithm COSMIC BIRTH. We discuss the challenges to reconstruct the density perturbations in the COSMOS field, and present the necessary extensions to our algorithm. In Section 3 we give details of the surveys used in this work, how we selected the data and how we constructed the selection functions. In Section 4 we give details of our large-scale bias computation and how we determined the bias evolution with increasing redshift. In Section 5 we describe the application of COSMIC BIRTH to the surveys in the COSMOS field, present the inferred density fields and provide additional diagnostics to validate the COSMIC BIRTH reconstruction algorithm. Finally, we give a summary in Section 6, discuss our results and forecast future efforts based on this work.

Throughout this paper we use a fiducial $\Lambda$CDM flat cosmology with a set of cosmological parameters $\{p_c\}$ given by

$$\{p_c\} = \{\Omega_M = 0.31, \Omega_{\Lambda} = 0.69, \sigma_8 = 0.82, n_s = 0.96, h = 0.68\}$$ (1)

and all distances are given in comoving $h^{-1}$ Mpc units.

## 2 COSMIC BIRTH-ALGORITHM

Density field reconstructions from observed galaxy distributions are an ongoing effort in cosmological science. The aim of this work is to infer the initial density fluctuations $\delta(q)$ at Lagrangian coordinates $q$ given galaxy positions that have been observed in Eulerian redshift-space $x$. These two frames can be connected if the peculiar velocity $v_r(q)$ and displacement $\psi(q)$ fields are known, as

$$q = s - v_r(q) - \psi(q).$$ (2)

Equation 2 shows that the mapping problem is analytically ill-defined, as we need a priori knowledge of the Lagrangian coordinates $q$ at which the displacement field and the peculiar velocity field (see Kitaura et al. 2016) are evaluated. Thus, an iterative solution is required to solve this problem, as proposed in pioneering works (Yahil et al. 1991; Monaco & Efstathiou 1999). This class of iterative mapping schemes were further developed in Kitaura & Angulo (2012); Kitaura et al. (2012b); Kitaura (2013); Heß et al. (2013); Kitaura et al. (2019). One of the key ingredients to achieve higher accuracy on small scales consists of introducing a Bayesian inference framework in the initial conditions reconstruction which takes into account the likelihood of the dark matter tracers. In this way, the typical additional Gaussian smoothing can be avoided and the number counts of objects within the mesh in a given voxel resolution can be correctly treated (Kitaura & Enßlin 2008; Kitaura et al. 2010; Jasche & Kitaura 2010). This allows for high precision on a few Mpc scales (see Kitaura et al. 2012c; Nuza et al. 2014).

COSMIC BIRTH (Kitaura et al. 2019) is a Bayesian inference framework, using a nested Gibbs sampling scheme to infer the initial density perturbations $\delta(q)$ on a regular cubic mesh grid with $N_G$ voxels. It maps the galaxy distribution represented by its Cartesian (for a given set of cosmological parameters) Eulerian redshift-space positions $\{q_{\text{obs}}\}$ to Lagrangian real-space coordinates $\{q\}$ expressed as number counts per voxel on a regular mesh: $N_G(q)$. For efficiency the displacement field is computed relying on Augmented Lagrangian perturbation theory (Kitaura & Hess 2013) and so effectively solving Equation 2 (see Section 2.1).

In this way we have a precise description of the action of gravity within $\Lambda$CDM on Mpc scales at redshifts larger than one (Neyrinck 2013). To compute the displacement $\psi(q)$ and velocity fields $v_r(q)$ from the initial density field $\delta(q)$ we apply Hamilton Monte-Carlo (HMC) sampling (see Duane et al. 1987; Jasche & Kitaura 2010; Neal 1993, 2011) with a bias description $B(\delta(q))$ and the galaxy counts on the mesh grid $N_G$. This is key to perform a forward modelling, as we do not obtain the displacement field from the density field defined in Eulerian space, done by inverse approaches (see e.g. Eisenstein et al. 2007). We account for selection effects of the galaxy survey data, i.e. the survey geometry, the angular and the radial selection functions, within a response operator $R$ (see Section 3.2 for details). The COSMIC BIRTH code uses a particularly efficient fourth order leap-frog implementation (see Hernández-Sánchez et al. 2019) to solve the Hamiltonian equations of motion. We recapture the gravity model in Section 2.1 following a description of the probabilistic model in more detail in Section 2.2 and then expand the calculations in Section 2.3 to combine multiple surveys.

### 2.1 Gravity Model

In this section we recapitulate the analytical model used within the COSMIC BIRTH code to compute the gravitational evolution of the cosmic density field.

We rely on augmented Lagrangian Perturbation Theory (ALPT) to simulate structure formation (details can be found in Kitaura & Hess 2013). In this approximation the displacement field $\psi(q)$, mapping a distribution of dark matter particles at initial Lagrangian positions $q$ to the final Eulerian positions $x(z)$ at redshift $z$ ($x(z) = q + \psi(q)$), is split into a long-range $\psi_L(q)$ and a short-range component $\psi_S(q)$, i.e. $\psi(q) = \psi_L(q) + \psi_S(q)$. The long-range component is computed with second order Lagrangian Perturbation Theory (2LPT) $\psi_{2\text{LPT}}$ (for details on 2LPT see Bouchet et al. 1992, 1995; Catelan 1995). The resulting displacement field is convoluted with a kernel $K$: $\psi_L(q) = K(q, r_S) * \psi_{2\text{LPT}}(q)$, given by a Gaussian filter $K(q, r_S) = \exp(-|q|^2/(2r_S^2))$, with $r_S$ being the smoothing radius. The short-range component is modelled with the spherical collapse approximation $\psi_{SC}(q)$ (see Bernardine 1994; Mohayaee et al. 2006; Neyrinck 2013). The resulting ALPT displacement field from combining the long and the short range components given by:

$$\psi_{\text{ALPT}}(q) = K(q, r_S) * \psi_{2\text{LPT}}(q) + (1 - K(q, r_S)) * \psi_{SC}(q)$$ (3)

is used to move a set of homogeneously distributed particles from Lagrangian initial conditions to the Eulerian final ones. We then grid the particles following a clouds-in-cell scheme and phase space mapping (Abel et al. 2012; Hahn et al. 2013) to produce a smooth density field $\delta(r)$. Some improvements
can be obtained preventing voids within larger collapsing regions, which essentially extends these regions towards moderate underdensities (see MUSCLE method in Neyrinck 2016). This approach requires about eight additional convolutions being about twice as expensive, as the approach used here. Moreover, we have checked that the improvement provided by including MUSCLE is not perceptible when using grids with cell resolutions of the order $\sim h^{-1}\text{Mpc}$.

The mapping between Eulerian real space $x(z)$ and redshift space $s(z)$ is given by: $s(z) = r(z) + v_r(z)$, with $r = (\mathbf{r} \cdot \hat{r})/(H_0)$, where $\hat{r}$ is the unit sight line vector, $H$ the Hubble constant, $a$ the scale factor, and $v = v(x)$ the 3-d velocity field interpolated at the position of each halo in Eulerian-space $\mathbf{r}$ using the displacement field $\psi_{\text{ALPT}}(\mathbf{q})$. We split the peculiar velocity field into a coherent $v_{\text{coh}}$ and a (quasi) virialized component $v_\sigma$: $v = v_{\text{coh}} + v_\sigma$. The coherent peculiar velocity field is computed in Lagrangian-space from the linear Gaussian field $\delta^{(1)}(\mathbf{q})$ using the ALPT formulation consistently with the displacement field (see Equation 3):

$$v_{\text{ALPT}}(\mathbf{q}) = \mathcal{K}(\mathbf{q}, r_\ast) * \psi_{\text{LPT}}(\mathbf{q}) + (1 - \mathcal{K}(\mathbf{q}, r_\ast)) * v_{\text{SC}}(\mathbf{q}),$$

with $\psi_{\text{LPT}}(\mathbf{q})$ being the second order and $v_{\text{SC}}(\mathbf{q})$ being the spherical collapse component (for details see Kitaaura et al. 2014). We use the high correlation between the local density field and the velocity dispersion to model the displacement due to (quasi) virialized motions. Effectively, we sample a Gaussian distribution function ($\hat{g}$) with a dispersion (see also Ata et al. 2017) given by $\sigma_r \propto (1 + \delta(r))^S$. Consequently we assume,

$$v_{\sigma}^i = (v_{\sigma} \cdot \hat{r})\hat{r}/(H_0) = \hat{g}(\mathbf{q} \times (1 + \delta(r))^2) \cdot \hat{r}. \quad (5)$$

For the Gaussian streaming model see Reid & White (2011), for non-Gaussian models see e.g. Tinker (2007). In closely virialized systems the kinetic energy approximately equals the gravitational potential and a Keplerian law predicts $\gamma$ close to 0.5, leaving only the proportionality constant $\gamma$ as a free parameter in the model. We leave a detailed investigation of the impact of redshift space distortions for future work.

2.2 Probabilistic Model

We use a Bayesian framework to draw samples of the density field $\delta(q)$ from a posterior probability density $P(\delta(q) \mid N_G(q), B(\delta(q)), R(q))$. The posterior itself is a product of a prior and a likelihood function which we will describe in more detail in the following. We can express the expectation value of galaxies per voxel $A_i = \langle N_G(q_i) \rangle$ for all voxels $i \in \{1, \ldots, N_C\}$ as:

$$A_i = \langle N_G(q_i) \rangle = f_N(q_i) \sum_j R_{ij}(q_i) B_j(\delta(q_i)). \quad (6)$$

with the normalisation factor $f_N(q_i) = \tilde{N} / (B(\delta(q_i)))$ ensuring the right number galaxy density $N$, as given by each survey. We consider the particular case, in which the response function $R$ is limited to the completeness $w_i$ in each cell $i$ as described in Section 3.2: $R_{ij}(q_i) = R_i \delta_i$. We then relate the expected number counts of galaxies per voxel to the actual number counts $N_G$ due to a likelihood model:

$$L(N_G \mid A(q)) = \prod_i \frac{A_i(q)^{N_G(q_i)} \exp(-A_i(q))}{N_G(q_i)!}, \quad (7)$$

where $A(q) = \lambda \left( f_N(q_i) B(\delta(q_i)), R(q) \right)$. Throughout this work we assume a Poisson likelihood (Kitaura & Eifflin 2008; Kitaaura et al. 2010). This is an adequate assumption for tracers with a vanishing small scale clustering, which otherwise become sources of super-Poissonity (e.g. Peebles 1980) that can be modelled with a negative-binomial likelihood (Kitaaura et al. 2014; Neyrinck et al. 2014; Ata et al. 2015). However, since the tracers are mapped to Lagrangian space at very high redshifts (e.g.: $z = 100$), the deviation from Poissonicity becomes negligible.

These conditions are also ideal for a lognormal prior describing the dark matter distribution towards high redshifts (Coles & Jones 1991). Its derivation is precisely based on a comoving framework at initial cosmic times, before shell crossing occurs. Thus, we can define the linearised density as

$$\delta_L(q) = \log(1 + \delta(q)) - \mu, \quad (8)$$

with $\mu = \langle \log(1 + \delta) \rangle - \log\left(\left\langle e^{\delta} \right\rangle \right)$ (Kitaaura et al. 2012a). Since we consider early cosmic times, the overdensity field has little power $|\delta| \ll 1$. Hence, the logarithmic transformation ensures positive densities $\rho$ with $\delta = \rho/\bar{\rho} - 1 = \exp(\delta_L + \mu) - 1$, while we can model the prior $\pi(\delta(q))^3$ for the linear density field $\delta_L$ by a Gaussian distribution with zero mean

$$\pi(\delta(q) \mid C_L(q)) = \frac{1}{\sqrt{(2\pi)^N C_{\text{det}}(C_L(q))}} \exp\left(-\frac{1}{2} \delta^T_L(q) C_L^{-1}(q) \delta_L(q) \right), \quad (9)$$

where $C_L(q) = \left\{\delta_L(q) \delta_L(q)^\dagger\right\}$ is the covariance matrix of the linearised density fields, which depends on the cosmological parameters $\{p_k\}$. Finally, we can express the posterior of $\delta(q)$ through Bayes theorem as:

$$P(\delta(q) \mid N_G(q), B(\delta(q)), R(q), \{p_k\}) \propto \pi(\delta(q) \mid C_L(\{p_k\})) \times L(N_G(q) \mid A(q), R(q)), \quad (10)$$

where the normalisation given by the evidence is not necessary within HMC. Table 1 summarises the main quantities that are sampled and how they are connected to each other.

2.2.1 Galaxy Bias Description

Finally, we need to specify the connection between the likelihood and the prior through the bias relation $B(\delta)$. In the COSMIC BIRTH framework, non-local bias is described through the displacement. The split-background bias (Kaiser 1984; Bardeen et al. 1986), which is necessary in Eulerian space particularly for massive galaxies (see e.g. Kitaaura et al. 2015), however, becomes negligible when homogenising the galaxy distribution mapping it to Lagrangian

---

\(^2\) Note that the prior is actually a function of the linearised density field $\delta_L(q)$, which is in turn a function of the original density field $\delta(q)$.
space. Hence, we can assume a power-law Lagrangian bias as discussed in Kitaura et al. (2019):

\[ B(\delta(q)) = (1 + \delta(q))^{b(z_q)} f_b(z_q), \]

where \( z_q \) is the redshift at which the Lagrangian coordinates are evaluated (for this study \( z_q = 100 \)), \( b \) the linear large-scale bias and \( f_b(z_q) \) the non-linear correction factor of our bias description (Ata et al. 2017). By using a power-law bias we ensure that the density field is positive, since otherwise any bias less than one can potentially cause negative densities at voxels with \( b \) close to \(-1\). The correction factor \( f_b(z_q) \) can be determined iteratively as presented in Kitaura et al. (2019). The advantage of this bias description is that the only free parameter of our method is reduced to the large-scale bias at Eulerian space, which can be connected to Lagrangian space through passive evolution (Nusser & Davis 1994; Fry 1996):

\[ b(z) = (b(z) - 1) \frac{D(z)}{D(z_q)} + 1, \]

including the linear growth factor \( D(z) \). We will show in Section 4.2 how we describe the large-scale bias evolution for the employed galaxy catalogs.

### 2.3 Multi-tracer Formalism in Lagrangian-space

In this study, we aim at combining the data of five spectroscopic surveys in the COSMOS field. The complexity lies in that they are partially spatially overlapping, and that they are relying on different observing strategies (for more details see Section 3).

Thus merging the surveys into a single catalogue by a pre-processing step, i.e. adding the different catalogues into one data set, represents a difficult task, which in general cannot be accomplished without making a series of simplifying assumptions. Some of the previous pioneering Bayesian inference studies have already applied some kind of multi-tracer treatment, however, in Eulerian space, and without taking into account separate survey masks within the Bayesian scheme (Jasche et al. 2015; Granett et al. 2015; Jasche & Lavaux 2019). Instead, we aim at performing a joint analysis of the different surveys within the Bayesian framework. To achieve this, we consider Lagrangian space, before the action of gravity causes interaction at different scales, and therefore different modes are not yet coupled in Fourier space. Therefore, this approach enables us to consider independent catalogs at very high redshifts (e.g. \( z = 100 \)), and treat each survey through a distinct likelihood component, avoiding mixed likelihood terms. In this way, we are able to combine different catalogues (indexed with superscript \( k \)) taking into account their distinct survey selection functions \( R^k(q) \), number densities \( N^k_C(q) \), and galaxy bias functions \( B^k(\delta(q)) \).

The Eulerian to Lagrangian mapping of COSMIC BIRTH, shown in Equation 2, can lead to a change of the redshift bin of a tracer. Consequently, the bias of this tracer will not coincide with the bias of the redshift bin at its new location. This can be taken into account by keeping track of galaxies staying at a redshift bin, or jumping from one redshift bin to another, which causes a “bias mixing” implemented in the COSMIC BIRTH code (see Section 3 in Kitaura et al. 2019). While this effect is negligible when interpolating the bias within the redshift bins, it has the advantage that a multi-tracer treatment is already implemented in this framework. We can then extend COSMIC BIRTH to perform a full Bayesian multi-tracer & multi-survey analysis to address the challenges of this work following the calculations presented in e.g. Ata et al. (2015, Appendix A) to express the corresponding posterior \( \mathcal{P}^{\text{multi}} \) in Lagrangian coordinates and combine the surveys with their specific likelihoods to construct the combined likelihood \( \mathcal{L}^{\text{multi}} \) by:

\[
\mathcal{L}^{\text{multi}}(N_G(q), | (f_N(q), B(\delta(q)), R(q))) = \prod_k \mathcal{L}^{(k)}(N^k_G(q), | (f^{(k)}_N(q), B^{(k)}(\delta(q)), R^{(k)}(q))),
\]

where index \( k \in [1\ldots N_k] \) denotes the different surveys. Accordingly, we need to generalise Equation 6 and distinguish between the surveys in the reconstructed volume:

\[
A^{(k)} = \left(N^k_G(q)\right) = f^{(k)}_N(q)R^{(k)}(q)\bar{G}^{(k)}(\delta(q)).
\]

In Hamiltonian sampling, one seeks to draw samples of the potential energy term \( U \) of the Hamiltonian, that is linked to the posterior in Equation 10 via

\[ U = - \ln \mathcal{P}. \]
The zCOSMOS-deep survey is the high-redshift component of the spectroscopic galaxy surveys considered in this work. We describe each survey in more detail below and afterwards explain our method to estimate the corresponding survey completeness.

### Surveys in the COSMOS Field

Let us briefly recap the main characteristics of the different spectroscopic galaxy surveys considered in this work.

- **zCOSMOS-deep**
  The zCOSMOS-deep survey is the high-redshift component of the zCOSMOS spectroscopic survey, which covers the central 1 deg$^2$ of the zCOSMOS footprint (Lilly et al. 2007, Lilly et al in prep.) using the VIMOS spectrograph (LeFevre et al. 2003) on the VLT. The targets were chosen from the then-current version of the multi-colour photometric COSMOS catalog (Capak et al. 2007). To isolate galaxies at redshifts of $z > 1.5$, several selection criteria were applied. In particular, a selection in the (U-B)/(V-R) colour-colour plane, called “UBR” selection (Steidel et al. 2004) was combined with the “BzK” selection (Daddi et al. 2004).

For both the “UGR” and “BzK” selections, an additional selection of $22.5 < B_{AB} < 25.0$ and a deep K-band imaging reaching down to $K_{AB} = 23.5$ were applied. Our analysis is done on a tentative version of the zCOSMOS-deep catalog that has been used for the bulk of previous works that employ zCOSMOS-deep data. A refined version of this catalog will be available in the future (Lilly et al. in prep).

- **VUDS**
  The VIMOS Ultra Deep Survey (Le Fèvre et al. 2015), hereafter VUDS, is another spectroscopic survey carried out on the VIMOS spectrograph that was partly operated in the COSMOS field, but considerably deeper than zCOSMOS-deep with up to $\sim 3\times$ larger integration times. VUDS peaks in number density at $z \sim 3$ and was designed to explore multiple questions, including those related to the formation rates of stars and merging of galaxies during the period of time when galaxies were most active. Another major success was to identify and characterize galaxy protoclusters at $z \sim 3$–4 (e.g. Cucciati et al. 2014; Lemaux et al. 2018; Cucciati et al. 2018). The selection is based on photometric redshift selections (Ilbert et al. 2013), with a small fraction of galaxies selected using the Lyman-break technique (Steidel et al. 1996). VUDS and zCOSMOS are based on different versions of COSMOS parent catalog, so the astrometry was not identical for all sources between the two surveys. Because of the different catalogs used, we employed a matching radius of $0.1 - 0.2''$ to identify true duplicates.

- **MOSDEF**
  The MOSFIRE Deep Evolution Field (MOSDEF) Survey (Kriek et al. 2015) is partly taken in the COSMOS field, separated into three redshift intervals at $1.37 \leq z \leq 1.70$, $2.09 \leq z \leq 2.61$, and $2.95 \leq z \leq 3.80$, down to fixed $H_{AB}$ magnitudes of 24.0, 24.5, and 25.0 for each interval. The MOSFIRE spectrograph (McLean et al. 2012) on the Keck-I telescope was used to obtain near infra-red emission line redshifts of the targeted galaxies. The targets were selected from the photometric and grism 3D-HST (Brammer et al. 2012) data, applying photometric redshift and magnitude requirements to the parent catalog and show a spectroscopic success rate of around $\sim 80\%$.

- **ZFIRE**
  The KECK/MOSFIRE Spectroscopic Survey of Galaxies in Rich Environments at $z \sim 2$ (ZFIRE) (Nanayakkara et al. 2016) using the MOSFIRE spectrograph was partly taken in the COSMOS and Hubble Ultra Deep Survey (UDS) field (Beckwith et al. 2006). For the COSMOS field the targets were K-band selected from the photometric parent FourStar Galaxy Evolution Survey (ZF0URG3E) catalog (Straatman et al. 2016), requiring a photometric redshift $2.0 \leq z_{\text{phot}} \leq 2.2$. ZFIRE was designed to observe primarily the $z = 2.095$ galaxy cluster (Yuan et al. 2014).

- **FMOS-COSMOS**
  The FMOS-COSMOS survey (e.g. Silverman et al. 2015; Kashino et al. 2018) (hereafter FMOS) used the Fibre Multi-Object Spectrograph (Kimura et al. 2010) at the Subaru Telescope, observing star forming galaxies at redshifts $z \sim 1.6$ in the near infra-red. The targets were selected from

| Survey     | $N^{(0)}$ | $z$ range     | Parent catalog |
|------------|-----------|---------------|----------------|
| zCOSMOS-deep | 3544     | $1.4 \leq z \leq 3.6$ (*) | COSMOS         |
| VUDS       | 1822     | $1.4 \leq z \leq 3.6$ (*) | COSMOS         |
| MOSDEF     | 401      | $1.4 \leq z \leq 3.6$ (* | 3D-HST         |
| ZFIRE      | 149      | $2.0 \leq z \leq 2.2$ | ZF0URG3E       |
| FMOS-COSMOS| 587      | $1.4 \leq z \leq 1.7$ | COSMOS         |

Table 2. Summary of five surveys used for this study. $N^{(0)}$ resembles the number of galaxies that we use after applying spectroscopic quality criteria and removing duplicates. (*) zCOSMOS-deep, VUDS & MOSDEF exceed the redshift range of our reconstructions. MOSDEF observes the redshift range in intervals of 1.37 $\leq z \leq 1.70$, $2.09 \leq z \leq 2.61$, and $2.95 \leq z \leq 3.80$. In the case of FMOS-COSMOS, about half of the galaxies in the central footprint are used in this work.

Thus, the multi-tracer & multi-survey posterior $p^{\text{multi}}$ writes for a number of $N$ surveys as:

$$-\ln p^{\text{multi}}(\delta(q)|\lambda^{(1)}(q), \lambda^{(2)}(q), \ldots, \lambda^{(N)}(q)) =$$

$$= c - \ln \pi(\delta(q)|C_L(q)) - \ln L^{(1)}(N^{(1)}_G|\lambda^{(1)}(q), R^{(1)}(\delta(q)), R^{(1)}(q))$$

$$- \ln L^{(2)}(N^{(2)}_G|\lambda^{(2)}(q), R^{(2)}(\delta(q)), R^{(2)}(q))$$

$$\ldots$$

$$- \ln L^{(N)}(N^{(N)}_G|\lambda^{(N)}(q), R^{(N)}(\delta(q)), R^{(N)}(q)),$$

where the constant $c$ does not depend on $\delta(q)$.

### 3 SURVEY DATA & COMPLETENESS
Within the COSMOS field, several spectroscopic surveys have been undertaken (see e.g. Hasinger et al. 2018, for a summary), focusing mainly on star forming galaxies at high redshifts. This work uses data from five different surveys, summarised in Table 2. We describe each survey in more detail below and afterwards explain our method to estimate the corresponding survey completenesses.
the COSMOS photometric catalog within a redshift range of $1.4 \leq z_{\text{phot}} \leq 1.7$, additionally applying limits on the Ultra-VISTA K-band magnitude limit and the $H\alpha$ flux predicted from SED fitting. FMOS objects are pre-selected with secure photometric redshift using the filters available in the COSMOS photometric survey. In this analysis we utilize FMOS observations in the range of $149.8 \leq \text{R.A.} \leq 150.4$ and $1.8 \leq \text{DEC} \leq 2.5$.

In summary, the surveys apply different colour, magnitude and photometric redshift pre-selection cuts to the parent photometric catalogs to efficiently select high redshift targets for spectroscopy. In the case of galaxies that were spectroscopically observed in more than one of these surveys, we kept the one with better redshift quality, based on the signal-to-noise ratio of the observed spectra (see Section 3.2.1). From the redshift distributions of the surveys shown in Figure 1 we can see that the number density of galaxies peaks at $2 \leq z \leq 2.5$, mainly contributed by the zCOSMOS-deep survey.

Figure 1. Radial distribution functions for the five considered surveys shown as function of redshift with a bin width of $\Delta z = 0.16$. The two grey dashed vertical lines signpost the considered redshift range $1.4 \leq z \leq 3.6$ in this work.

The footprints (galaxy distribution in angular coordinates) of each survey are shown in Figure 2. From this figure we can recognise that the edges of zCOSMOS-deep, represented in blue, have been observed with only a single pointing, thus having a lower number density, while in the central area the pointings are overlapping, allowing for a denser targeting. Regarding the VUDS survey, represented in orange, Figure 2 shows the gaps in between the quadrants of an individual VIMOS pointing, whereas within one quadrant the target sampling is very homogeneous. Focusing on the MOSDEF and ZFIRE surveys depicted in green and red, respectively, one can see that they cover smaller areas in the central part of the COSMOS field. The FMOS survey, represented in magenta, shows a sparser targeting compared to the rest of the surveys considered in this study.

Figure 2. Survey footprints: Spatial distribution of the observed galaxies in the right ascension (R.A.)-declination (DEC) plane. We follow the same colour coding as in Figure 1. At a redshift of $z = 2.5$, the maximum separation $\Delta \text{DEC} = 0.91^\circ$ corresponds to a comoving transverse distance of $d_{\text{DEC}} = 64.8h^{-1}\text{Mpc}$, while for $\Delta \text{R.A.} = 1.02^\circ$ the maximum separation is $d_{\Delta \text{R.A.}} = 70.2h^{-1}\text{Mpc}$ assuming the cosmological parameters in Section 1.

3.2 Survey Completeness Estimation

The response function $R$ (see Table 1, Section 2) represents the accuracy with which each voxel $i$ has been observed. In a Bayesian analysis, the likelihood function should therefore account for the uncertainty of the galaxy number counts $N_G(q)$ as a function of the completeness, shown in Equation 6. The response function is calculated from an angular and a radial component, $R_\alpha$ and $R_r$, respectively, as they can be independently calculated:

$$R = R_\alpha \cdot R_r. \quad (17)$$

It is practical to consider them separately, since the angular part is not subject to redshift space distortions, contrary to the radial part. Following Kitaura et al. (2019), we compute the angular response operator in Eulerian space $R_\alpha(s) = R_\alpha(r)$ once, and in each Gibbs sampling iteration map it to Lagrangian space $R_\alpha(q)$ through the displacement field (see Table 1). The radial response function can be trivially computed from the distribution of large-scale tracers in Lagrangian real space coordinates $(q)$ in each Gibbs sampling iteration, and multiplied according to Equation 17. Let us thus focus in detail on the computation of the angular completeness.

3.2.1 Angular completeness

In the following we adopt the terminology of the VIMOS related surveys (e.g. VVDS,VIPERS) (see e.g. Ilbert et al.
The ratio of the spectroscopic parents (III) to the photometric targets (II) is called Spectroscopic Sampling Rate (SSR). Many factors have an influence on the SSR, which we call Colour Sampling Rate (CSR). To compute the angular completeness, we need to distinguish between four selection stages, starting from the photometric parent catalog and yielding the final spectroscopic survey, shown in Figure 3. In the first stage, observers apply photometric selection criteria on top of the parent photometric catalog (I) to select galaxies with certain properties (e.g. star forming, redshift range etc). The resulting catalog consists of photometric targets (II), leading to a second stage, as shown in Figure 3. The transition from the photometric parent (I) to the photometric target catalog (II) is called Colour Sampling Rate (CSR) and accounts for the colour-colour, colour-magnitude and photometric redshift selections. A fraction of the photometric targets (II) is chosen for spectroscopy, which we call Spectroscopic Parent (III). The ratio of (II) and (III) is called Target Sampling Rate (TSR), regardless of the quality of the spectra, $\text{TSR} = N_{\text{III}}/N_{\text{II}}$. To this end, we construct a mesh grid $80 \times 80$ cells over the R.A.-DEC plane (Figure 2), resulting in a resolution of about 0.9 arcmin per cell. We note that the reconstructions are computed with a comoving resolution of $d_k = 2\, h^{-1}\, \text{Mpc}$ (see Section 5). This corresponds to a angular aperture of $\theta = 1.5'$ at a redshift of $z = 3.6$ and thus coarser than the TSR resolution. We set the value of the selection function in between the VUDS quadrants and outside the borders of the surveys to zero.

Figure 3. Flowchart describing the process to obtain the final spectroscopic galaxies, starting from the parent photometric survey. The arrow symbol on the left represents the name of the survey. The descriptive text on the right denotes the selection/operation that is undertaken upon the survey, leading to the next stage. We use an adapted terminology of the VVDS/VIMOS collaboration: Stage (I) to (II) is called Colour Sampling Rate (CSR), (II) to (III) Target Sampling Rate (TSR), and finally (III) to (IV) Spectroscopic Success Rate (SSR).

To compute the angular completeness, we need to differentiate between four selection stages, starting from the photometric parent catalog and yielding the final spectroscopic survey, shown in Figure 3. In the first stage, observers apply photometric selection criteria on top of the parent photometric catalog (I) to select galaxies with certain properties (e.g. star forming, redshift range etc). The resulting catalog consists of photometric targets (II), leading to a second stage, as shown in Figure 3. The transition from the photometric parent (I) to the photometric target catalog (II) is called Colour Sampling Rate (CSR) and accounts for the colour-colour, colour-magnitude and photometric redshift selections. A fraction of the photometric targets (II) is chosen for spectroscopy, which we call Spectroscopic Parent (III). The ratio of (II) and (III) is called Target Sampling Rate (TSR), regardless of the quality of the spectra, $\text{TSR} = N_{\text{III}}/N_{\text{II}}$. To this end, we construct a mesh grid $80 \times 80$ cells over the R.A.-DEC plane (Figure 2), resulting in a resolution of about 0.9 arcmin per cell. We note that the reconstructions are computed with a comoving resolution of $d_k = 2\, h^{-1}\, \text{Mpc}$ (see Section 5). This corresponds to a angular aperture of $\theta = 1.5'$ at a redshift of $z = 3.6$ and thus coarser than the TSR resolution. We set the value of the selection function in between the VUDS quadrants and outside the borders of the surveys to zero.

Let us describe the different selection steps in more detail:

- **Colour Sampling Rate (CSR)**

One can make the robust assumption that the photometric pre-selections applied on the parent photometric catalog (I) are constant over the footprint of each survey. In such a case, the CSR will only result in an overall normalization factor of the number density of galaxies (see Pezzotta et al. 2017), but not influence the angular dependent clustering. Therefore, we can safely absorb this factor into the radial selection factor.

- **Target Sampling Rate (TSR)**

First, we reproduce the photometric pre-selection criteria. Then we build the ratio of the photometric targets $N_{\text{II}}$ with the number of the spectroscopic parents $N_{\text{III}}$, regardless of the quality of the spectra, $\text{TSR} = N_{\text{III}}/N_{\text{II}}$. To this end, we construct a mesh grid $80 \times 80$ cells over the R.A.-DEC plane (Figure 2), resulting in a resolution of about 0.9 arcmin per cell. We note that the reconstructions are computed with a comoving resolution of $d_k = 2\, h^{-1}\, \text{Mpc}$ (see Section 5). This corresponds to a angular aperture of $\theta = 1.5'$ at a redshift of $z = 3.6$ and thus coarser than the TSR resolution. We set the value of the selection function in between the VUDS quadrants and outside the borders of the surveys to zero.

- **Spectroscopic Sampling Rate (SSR)**

Finally, we select a subset $N_{\text{IV}}$ of galaxies from the spectroscopic parent sample, that have high redshift accuracies, as described follows. For the zCOSMOS-deep and VUDS surveys we demand redshift flags of $\geq 2$. For the MOSDEF survey we apply quality flags of $\geq 3$, which corresponds to a redshift confidence of $> 95\%$ and a minimum signal-to-noise ratio of $2 \leq S/N \leq 3$. For the ZFIRE survey we apply quality flag of $2$, which corresponds to a $S/N \geq 5$ and $|z_{\text{spec}} - z_{\text{phot}}| \leq 0.2$. For the FMOS survey we demand $3 \leq S/N \leq 5$, which translates into a redshift quality flag of $\geq 2$. The SSR is then calculated as $\text{SSR} = N_{\text{III}}/N_{\text{IV}}$.

The resulting angular selection masks

$$w_k^A = \text{TSR}^k \times \text{SSR}^k,$$

in the R.A.-DEC plane are shown in Figure 4 for each survey $k$. In the case of the zCOSMOS-deep survey the higher targeting rate in the center of the survey footprint, caused by several overlapping pointings with the VIMOS spectrograph, can be appreciated. For the VUDS survey on the contrary, each pointing is unique, and hence, the areas do not overlap. This results in thin stripes between the quadrants of the pointings in the footprint, showing the inter-CCD gaps in the VIMOS focal plane.

Once we have computed the angular mask, we need to project it into three dimensions, as required by the COSMIC BIRTH code.

In particular, we project each R.A.-DEC bin value of $w_k^A$ into our cubical mesh grid on which we perform the reconstructions. As the resolution of $w_k^A$ is higher compared to the reconstruction mesh grid, we average over each sight at a voxel $i$ and thus obtain the angular completeness $R_i^k$ for all $k$ surveys and cell $i$. Figure 5 shows the projection.
of the VUDS angular selection mask into the reconstructed volume. The data preparation and details of the coordinate system are explained in Section 5.1. The top panel of Figure 5 shows the stripes and the corresponding gaps in between the pointings of the survey in the Eulerian frame $R(5)$. It shows the stripes and the corresponding gaps in between. The top panel of Figure 5 shows the stripes and the corresponding gaps in between. This means, that unobserved angular regions in Eulerian space, might have been effectively partially observed in Lagrangian space. As expected, we can see that the deformation is stronger towards lower redshifts, where the growth of structures is more evolved.

4 LARGE-SCALE GALAXY BIAS

The COSMIC BIRTH algorithm accounts for stochastic and non-linear Lagrangian bias, while non-local Eulerian bias is modelled through the displacement field connecting Eulerian to Lagrangian space (see Section 2 and for further details in Kitaura et al. 2019). The only free parameter is the large-scale Eulerian bias (Kaiser 1984), which needs to be determined from observations or simulations. We rely on detailed bias studies of highly star forming galaxies due to the nature of the surveys considered in this study (see Section 4.1).

4.1 Bias Studies in Simulations and Observations

The various galaxy properties are in general correlated with their clustering behaviour, and hence, are indicators of how they trace the underlying dark matter density field. More massive and luminous galaxies, such as luminous red galaxies (LRGs) for instance (see e.g. Alam et al. 2017) show a strong clustering, tracing mainly the peaks of the density field (Kitaura et al. 2015). These galaxies are passively evolving, showing low stellar formation activity and old stellar populations.

Apart from LRGs, star forming galaxies can be identified with photometric techniques (see e.g. Daddi et al. 2004) and emission line spectroscopy, frequently called emission line galaxies (ELGs) in literature. The galaxy bias of highly star forming galaxies, such as [OII], [OIII], Hα detected galaxies (see e.g. Delubac et al. 2017; Kaasinen et al. 2017), UV emitting Lyman-break galaxies (Kollmeier et al. 2003) and Lyman-α emitters (LAEs, see e.g. Kennicutt (1998)) have been extensively studied in the literature.

Figure 4. Final angular completeness mask $w^k$ for all five surveys computed from TSR and SSR in the R.A.-DEC plane normalized to unity.

Unlike LRGs that trace only the densest peaks of the density field, ELGs can populate the density field at nearly the whole range of overdensities. Therefore, their bias is close to unity at low redshifts (Favole et al. 2016). However, in this work we consider higher redshifts, when the dark matter field was less evolved, displaying weaker density perturbations. Hence, these types of galaxies display an increasing bias towards high redshifts (Okada et al. 2016; Guo et al. 2019).

Studies based on numerical simulations have shown the bias as a function of redshift and star formation rate, comparing the galaxy and matter density fields (see Table 1 in Chiang et al. 2013). Within the VUDS survey a bias measurement for the redshift range of $2 \leq z \leq 5$ has been accomplished in Durkalec et al. (2015, 2018) using two-point clustering analysis. Similar analysis has been performed for the FMOS-COSMOS survey, studying the projected correlation function and from there estimating the bias at a median redshift of $\bar{z} = 1.58$ (Kashino et al. 2017).

4.2 Large-Scale Bias Evolution

In this section, we describe how the galaxy bias evolution is modelled. If we assume that the galaxy population of a survey is exactly identical across the light-cone, we can apply a passive time evolution bias model, shown in Equation 12 (Nusser & Davis 1994; Fry 1996), to obtain the large-scale bias. This is in fact the approach of COSMIC BIRTH, mapping one galaxy population from Eulerian to Lagrangian frame by applying the corresponding displacement field (see Section 2). In that case the mapped galaxy population remains the same and we can estimate the large-scale bias of these galaxies given different underlying density perturbations. However, this is not the case when galaxy formation and evolution have to be taken into account. If we want to determine the large-scale bias of one galaxy population corresponding to one redshift range based on another galaxy
population from a different redshift range, passive evolution cannot be applied. In such a case the redshift evolution of the galaxy populations must be modelled. This can be achieved by introducing a correlation coefficient $r$ between the dark matter and the galaxy density field as a function of redshift (Tegmark & Peebles 1998). As the selection criteria of the surveys do not significantly change with observed distance, we expect a moderate enhancement of the galaxy bias towards higher redshifts beyond the one described by passive evolution. Thus, we write:

$$b(z_2) = \left(1 - \frac{D(z_2)}{D(z_1)}\right)^2 - 2r(z_1) \left(1 - \frac{D(z_2)}{D(z_1)}\right) b(z_1) + b^2(z_1)$$

(19)

for two redshifts $z_1$ and $z_2$ with $z_1 < z_2$, a correlation coefficient $r(z_1)$, and the linear growth function $D(z)$ given by

$$D(z) = \frac{H(z)}{H_0} \int_z^\infty \frac{dz'}{H(z')} \int_0^{\infty} \frac{dz''}{H(z'')} \cdot \int_0^{\infty} \frac{dz'''}{H(z''')}.$$  

(20)

normalized to unity at redshift zero $D(z = 0) = 1$. The evolution of the bias and the correlation coefficient are coupled as:

$$r(z_1) = \left(1 - \frac{D(z_1)}{D(z_2)}\right) + r(z_2) \left(\frac{D(z_2)}{D(z_1)}\right) b(z_2) \left| b(z_1) \right|.$$  

(21)

A perfect correlation of $r(z_2) = 1$ at an earlier redshift $z_2$ will remain like that for all times, whereas $r(z_1)$ always tends towards 1, regardless of its initial value, according to Equation 21. In the case of a perfect correlation $r(z_2) = 1$, Equation 19 equals Equation 12 (described in Section 2) and no change of the galaxy population is expected along redshift. However, $r(z_2) < 1$ implies a varying correlation coefficient, effectively describing a cosmic evolution of the galaxy distribution, which maybe caused by galaxy formation or evolution (Tegmark & Peebles 1998).

---

**Figure 5.** Single slice plot of the projected window function $R_s$ for the VUDS survey into the reconstructed volume, normalized to unity, showing 0 for non observed regions and 1 for maximum completeness. The top panel represents the window calculated directly from the observations in Eulerian space $s$, showing the effects of the gaps in between the VUDS quadrants (compare Figure 4), emphasizing the importance of an accurate selection function handling for this work. The bottom panel presents the same slice in Lagrangian space $q$. The whole declination angle on the Y-axis corresponds to $200 h^{-1} $ Mpc in comoving distance.

**Figure 6.** Large-scale galaxy bias $b(z)$ as a function of redshift $z$ assuming Equation 19. The blue dashed line and the corresponding grey error band show the result for $r = r(z = 1.5) = 0.8$ and $b(z = 1.5) = 1.88\pm0.33$. The orange dashed line shows the outcome for $b(z = 1.5) = 1.88$ and $r = r(z = 1.5) = 1.0$. The circles denote previous bias measurements whereas the red triangles show values derived from numerical simulations.
4.3 Large-Scale Bias Computation

In the following we explain our strategy to compute the large-scale bias, as required to perform the dark matter reconstructions throughout the redshift range of 1.4 ≤ z ≤ 3.6.

Given that the galaxy populations from the 5 considered surveys are similar, we assume as a null hypothesis, that galaxies share the same bias at a given redshift, and that their bias passively evolves through Equation 12 (equivalent to a perfect correlation r(z) = 1 in Equation 19). Then, we choose a narrow redshift range for which the deviation from passive evolution is expected to be negligible. In particular we select the range 1.4 ≤ z ≤ 1.8 embedded in a rectangular volume that extends from 2875 h⁻¹ Mpc ≤ d_L ≤ 3387 h⁻¹ Mpc in line-of-sight distance. We perform a series of COSMIC BIRTH runs with the data set in this volume with a varying bias at z = 1.5 from 1 to 2.5 (the general set-up of the COSMIC BIRTH reconstructions is described in Section 5.1). From these runs we find that the reconstructed primordial matter density shows unbiased power spectra w.r.t. the theoretical linear one for b ≈ 1.8 ± 0.3 at redshift z = 1.5 being conservative (and the rest of bias values given by passive evolution).

This is in good agreement with the findings of Cochrane et al. (2017, see Table 4, where they obtain b = 1.78±0.08 at z = 1.48), and also roughly compatible with the clustering analysis of the FMOS-COSMOS survey (Kashino et al. 2017, see Table 3, finding b = 2.44±0.32 at median redshift z = 1.59) within the estimated uncertainties.

The passive evolution model, based on our low redshift bias measurement, evolves to higher redshifts according to the blue solid line including the large associated uncertainties represented by the shaded light blue area in Figure 6. We can see from this Figure that there is a clear trend in the observational data points to favour higher bias values, than the ones predicted by passive evolution towards high redshifts, although all data points are compatible with this model within the large uncertainties. On the other hand, a series of COSMIC BIRTH runs disfavour 2σ deviations from the upper bias limits quoted in the literature, as they lead to unreasonable biased dark matter reconstructions (see Appendix A). We note, that the selection function can lead to an excess of power in the two point statistics on large scales, and thereby higher bias values can be inferred (see e.g. Thomas et al. 2011). We therefore investigate the theoretical predictions for the bias evolution in simulations.

As a result, we find that the galaxy samples considered in this work cover the stellar mass range of M_L = 10^{9.3} M_☉ to M_L = 10^{10.5} M_☉ peaking at M_L = 10^{9.8} M_☉ (Lamaux et al in prep.). According to this finding, we obtain the data points represented in red upwards pointing triangles in Figure 6 (see Table 1 in Chiang et al. 2013). We find that these simulation based data are in agreement with the observational measurements, however favour slightly lower bias values. In the spirit of being conservative, and avoid bias ranges which can be affected by selection effects, we include the simulation (Chiang et al. 2013) and observational (Cochrane et al. 2017; Kashino et al. 2017; Durkalec et al. 2015, 2018) data points in a least squares fit. The resulting bias evolution model is represented in solid red with the uncertainty given by the light red shaded area in Figure 6.

According to this bias study, we have found some moderate evidence (given the hitherto large uncertainties due to the small volumes covered by high redshift galaxy surveys) for a cosmic evolution of the galaxy bias beyond passive evolution, hinting towards ongoing galaxy formation and merging processes at these redshifts. A coefficient of r(z = 1.5) = 0.6 corresponds to r(z = 3.6) = 0.42 (see Equation 21). This apparently tiny variation, reduces the tension with observationally constrained biases at redshift z > 1.5. However, a proper verification of a deviation from passive evolution requires a deeper study, extending the runs we performed at low redshift to higher ones. The available data at this stage might not be sufficient to make stronger claims and we leave a more detailed investigation of the bias evolution to a forthcoming work.

The resulting large-scale bias calculations presented in this section can be fed into the COSMIC BIRTH code to produce unbiased dark matter reconstructions, as shown in Section 5.

5 COSMIC BIRTH APPLIED TO THE COSMOS FIELD

In this section we present the application of COSMIC BIRTH code (see Kitaura et al. 2019, and Section 2) to the spectroscopic surveys in the COSMOS field (see Section 3). Previous pioneering COSMOS density field estimates were based on photometric redshifts (Kovac et al. 2010) and used tessellations of the observed galaxy fields (Scoville et al. 2013; Darvish et al. 2015), but did not account for the selection function (Amara et al. 2012; Smolčić et al. 2017), or focused on individual high density peaks (e.g. Wang et al. 2016). Therefore, this work represents the first comprehensive study to address all the above mentioned issues, taking into account structure formation, selection functions, redshift-dependent bias descriptions and redshift-space distortions within a forward Bayesian analysis.

5.1 Data Preparation and Setup

The COSMIC BIRTH code in its current version performs calculations on cubical regular meshes in comoving Cartesian coordinates and uses the corresponding galaxy positions in redshift-space on the light-cone with their corresponding bias obtained from galaxy catalogues as input source.

First we assume in this study a ΛCDM model with cosmological parameters defined in Equation 1. Then we select the input catalogues comprehending 5 different redshift surveys, as described in Section 3.1. The large-scale bias as a function of redshift is obtained from the data itself and further constrained according to some previous studies, as explained in Section 4.3. From this the connection to Lagrangian bias including a nonlinear and nonlocal treatment is internally computed, as explained in Section 2.2.1. With the given cosmology we can translate the angular and redshift coordinates for each galaxy into comoving Cartesian coordinates (x, y, z). The corresponding geometry and angular completeness to each survey is computed, as explained in detail in Section 3.2.1. While the COSMIC BIRTH code does not assume the plane parallel approximation at any step, this approximation is nearly fulfilled given the large distances to the galaxies of the considered redshift range and the narrow angular coverage (see Figure 1). We take advantage of that
for visualisation purposes, and choose a coordinate transformation so that the centre of the eCOSMOS-deep survey is aligned to DEC = 0° and R.A. = 180°, which makes the Y-axis approximately coincide with the declination angle (for angles close to zero, as in this case) and the X-axis with the redshift $z$ (see e.g. Figures 5,11). The final results are presented in the original coordinate system. The reconstructions comprise the galaxies of the five mentioned surveys within a redshift range of $1.4 \leq z \leq 3.6$, which translates into a comoving distance of $d_{\text{com}} = 1898 \, h^{-1} \, \text{Mpc}$ along the line-of-sight. For computational reasons we have split the reconstruction into four cubical volumes making sure that a large enough volume (in terms of mode-coupling) is taken in each case of $512 \, h^{-1} \, \text{Mpc}$ side length (Sorce et al. 2016). This permitted us to achieve a mesh grid resolution resolution of $2 \, h^{-1} \, \text{Mpc}$ using meshes of $256^3$ voxels. We placed the four volumes successively along the line-of-sight considering an overlapping region of $50 \, h^{-1} \, \text{Mpc}$ on each side (acknowledging that the density correlation function drops at scales lower than $20 \, h^{-1} \, \text{Mpc}$). Additionally, we did not place galaxies in a buffer zone of $25 h^{-1} \, \text{Mpc}$ at the edges of each volume, and took this into account in the completeness. Thus, the full reconstructions extends from $2875 \, h^{-1} \, \text{Mpc}$ to $4773 \, h^{-1} \, \text{Mpc}$ in comoving line-of-sight distance. We note, that COSMIC BIRTH code takes light-cone evolution into account within each reconstructed volume. In particular, we choose 8 redshift bias for each sub-volume (see Figure 4 in Kitaura et al. 2019).

### 5.2 Numerical Tests and Convergence Behaviour

Our numerical tests showed that the COSMIC BIRTH reconstructions are insensitive in terms of the power spectrum to deviations from up to 30% from our best large-scale bias evolution estimate shown in Figure 6. We note, that this already excludes the high end of bias values allowed within $2\sigma$ confidence levels by the observations (see discussion in Section 4.3). The convergence behaviour of the power spectra from the Lagrangian density fields (for the best large-scale bias model), starting from a perfectly homogeneous density field $\delta(q) = 0$ is shown in Figure 7. The reconstructions show unbiased power spectra already after the 40th iteration. Up to this iteration we consider the chain to be in the burn-in phase and not representing the target distribution of which we aim at drawing samples from. Furthermore, the top panel of Figure 8 shows the mean power spectrum of the initial density field $\delta(q)$ at $z = 100$ averaged over 6000 samples, illustrating the $1\sigma$ standard deviation as a grey band. The bottom panel shows the corresponding ratios with the theoretical linear $\Lambda$CDM prediction. The reconstructed density fields show unbiased power spectra over all scales, confirming an accurate bias treatment. The effect of a mishandled bias in our study is shown in Appendix A. We note that sampling the large-scale bias from data sets like the ones considered here with such a low data volume fraction is going to be very inaccurate, although the Bayesian framework allows for it (see Granett et al. 2015; Jasche & Lavaux 2017).

Once the cosmology has been chosen, and the data input defined, there is only one free parameter in the COSMIC BIRTH code, which is the large-scale bias of the different populations. As shown in Section 4, there is a confidence region for the large-scale bias as a function of redshift for the galaxies considered in this study. Given the the low volume filling fraction of the data region with respect to the entire volume, which needs to be considered to keep the mode-coupling effects from large-scale modes low (see e.g. Sato et al. 2009; Takada & Hu 2013), the question arises whether the reconstructions are sensitive to bias. To validate the independence of the HMC samples, we calculate the correlation length for the inferred density field. The correlation length $\xi(\delta_i)^n$ for a particular density voxel $\delta_i$ at an iteration distance $n$ over
Figure 9. Correlation length \( C(\delta_j) \) of 800 randomly chosen density field voxels \( \delta_j \) over a sample of \( N = 4000 \) HMC realization with an iteration length \( n \in [0...200] \). The colour code shows the correlation length for a particular \( \delta_j \) while the black solid line represents the average over all density voxels. We consider \( C(\delta_j)_n < 0.1 \) to be uncorrelated, which our HMC chain drops below after \( n = 40 \) iterations. Thus, each 30th HMC realization is an independent sample.

\[
C(\delta_j)_n = \frac{1}{N - n} \sum_{i=0}^{N-n} \frac{(\delta_j^i - \langle \delta_j \rangle)(\delta_j^{i+n} - \langle \delta_j \rangle)}{\sqrt{\sigma^2(\delta_j)}} ,
\]

where \( \langle \delta_j \rangle = \frac{1}{N} \sum_i \delta_j^i \) is the mean of the density voxel \( \delta_j \), and \( \sigma^2(\delta_j) = \frac{1}{N} \sum_i (\delta_j^i - \langle \delta_j \rangle)^2 \) the corresponding variance. We show the correlation length in Figure 9 for 800 randomly chosen density voxels \( \delta_j \), with \( j \in [1...800] \) in the data region of our reconstructed volume. This demonstrates that we draw independent about each ~40th iteration.

To quantify the Lagrangian to Eulerian mapping using ALPT within COSMIC BIRTH, we show the density distribution of the initial and final density fields in Figure 10. On the same scale on the x-axis the density field value \( \delta \) is plotted against the normalized probability distribution. We can see that the density field in Lagrangian space follows closely a Gaussian distribution with mean peaked at zero, and very small variance and skewness, \( \mu \approx 0, \sigma^2 = 0.001, s = 0.001 \). The histogram for the Eulerian density field also has a mean value close to zero, \( \mu \approx 0 \), however shows a variance of \( \sigma^2 = 0.95 \), and a skewness of \( s = 5.373 \). This is induced by gravity over cosmic time scales and is in excellent agreement to previous findings studied in detail in Neyrinck (2013), comparing the displacement fields of second order Lagrangian perturbation theory (2LPT) and N-body simulations at different redshifts. Especially at redshifts \( z \geq 1 \) the 2LPT displacements and the N-body results are in very good agreement and thus represent a very reasonable choice for this study (even more so, since we use ALPT).

5.3 Density Inference Results

The resulting reconstructions corresponding to our best large-scale bias evolution estimation are shown in Figure 11 as slice plots. On the X-axis we show the line-of-sight distance\(^4\) as redshift \( z \), and on the Y-axis we show the declination (DEC). The slices are shown with a thickness of 6 \( h^{-1} \) Mpc (averaging over three neighbouring cells). The corresponding convergence behaviour, power spectra, and matter statistics was discussed in Section 5.2 and shown in Figures 7–10.

We find homogeneously distributed initial cosmic density fields (see top panel of Figure 11 for individual reconstructions of \( 1 + \delta(q) \)). This is further demonstrated in the statistical matter distribution shown in the upper panel of Figure 10) with negligible skewness and kurtosis values, as expected for a Gaussian density field. Furthermore, the top panel of Figure 11 shows no redshift evolution. This is in contrast with the second upper panel, in which the Eulerian density for the corresponding reconstruction \( 1 + \delta(s) \) is shown. In fact, the corresponding matter statistics shown in the lower panel of Figure 10 shows a highly non-Gaussian distribution. Further inspection of the second upper panel of Figure 11, shows an increase of the matter fluctuations towards low redshifts, particularly enhanced through the appearance of large cosmic voids, depicted in dark blue. The corresponding ensemble averages (over 6000 Gibbs-sampling iterations) to the upper panels are shown in the middle panels, (\( 1 + \delta(q) \)) and (\( 1 + \delta(s) \)), respectively. In particular, the average over the ensemble of Eulerian density fields (i.e., the expected dark matter field from a Bayesian calculation) shows a high correlation with the galaxy field, as it should happen, when

\[
\delta(q) \approx 1.1 \delta(s)
\]

\(^4\) We note that the line-of-sight distance corresponds to the redshift of the observations only for the light-cones. For the initial density fields \( \delta(q) \) at \( z = 100 \), X-axis only shows a distance measure.
Figure 11. COSMIC BIRTH reconstructed density field $1 + \delta$ shown as slices in line-of-sight and declination coordinates with thickness of $6 h^{-1}$ Mpc. Phase-space mapping with the corresponding tetra-hedra tessellation has been performed as mentioned, in Section 2. The whole declination angle on the $Y$-axis corresponds to $200 h^{-1}$ Mpc in comoving coordinates. From top to bottom the two upper slice plots show firstly a single realization the Lagrangian initial density field $1 + \delta(q)$ at $z = 100$ and secondly a light-cone realization at Eulerian redshift-space $1 + \delta(s)$ with the individual galaxy positions plotted on top. The two middle panels represent the mean distributions of the inferred initial Lagrangian $\langle 1 + \delta(q) \rangle$ at $z = 100$ and final Eulerian $\langle 1 + \delta(s) \rangle$ density fields averaged over 6000 HMC realizations, respectively, where on top of the Eulerian density field the galaxy positions are plotted with black dots. Finally we show in the bottom panel the mean divided by its variance $\mu/\sigma^2$ of the Eulerian density field with the galaxy positions plotted on top.

the initial cosmic density field is accurately recovered. One can also appreciate the vanishing fluctuations in regions with low completeness, in concordance with a Bayesian analysis (see for a comparison, as an example the completeness of VUDS depicted in Figure 5). The bottom panel shows the signal to noise ratio obtained through the ratio of the mean density over its variance $\mu/\sigma^2$. As expected, we find a higher signal to noise ratio in the data region. Interestingly, cosmic voids are particularly prominent in this measure.

To further assess the matter distribution in the reconstructions, we show tomographic slice plots for various R.A. ranges (grey vertical stripes) in Figure 12, showing the survey footprints on the left-hand side and the corresponding density reconstructions next to it on the right. This permits
us to evaluate the extension of the proto-clusters and cosmic voids. A more detailed study will be presented in a forthcoming publication. A comparison between the Lagrangian and Eulerian ensemble average reconstructions (middle panels in Figure 11) shows higher density fluctuations in Eulerian space, and small displacements of the center of mass of the proto-clusters towards high redshifts. Focusing on the lowest redshift regions one can observe stronger variations of the shape of the proto-clusters.

To further investigate this, we zoom into a prominent proto-cluster region with $2.28 \leq z \leq 2.51$ and $1.6^\circ \leq \text{DEC} \leq 2.9^\circ$ (see Figure 13). This comparison shows three high density proto-clusters at $z \sim 2.45$, $z \sim 2.4$ and $z \sim 2.3$ growing throughout the cosmological timescale starting from $z = 100$ down to their light-cone redshifts. At first glance, the proto-clusters seem to grow in place, with negligible displacements, in accordance with linear perturbation theory, describing the growth of perturbation fixed in comoving frame between the initial redshift $z_q$ to the final redshift $z_f$ with $\delta(r, z_f) = \delta(r, z_q) D(z_f)/D(z_q)$, where $D(z)$ is the linear growth rate. However, a more detailed inspection reveals a complex non-spherical accumulation of mass when comparing the middle and the right panels. This can be well appreciated when comparing to the Gaussian smoothed galaxy field shown on the left panel. We also find from this comparison how the two proto-clusters on the left, which appear...
as separate entities in the left panel, are actually connected, most likely having a dark matter bridge in between their respective galaxy distributions. It is also interesting to see the accumulation of matter through the action of gravity from a ring-like shape cloud to a spherical overdensity region in the region centred at $z \sim 2.38$ and $\text{DEC} \sim 1.9^\circ$, comparing the Lagrangian and Eulerian density fields. The Bayesian analysis taking the completeness into account, also shows us which regions are more likely to be true cosmic voids. In particular, the right panel shows a deep void towards higher declination angles in the redshift range $2.3 < z < 2.4$, which feeds the overdensity peaks on its left and right. This is in agreement with the cosmic void initially discovered from the three dimensional Ly$\alpha$ forest tomography (Krolewski et al. 2018). Another void region is forming directly under the proto-clusters at $z \sim 2.45$ and $z \sim 2.4$. We will provide a more detailed analysis of the various structures in the reconstructed density field in a subsequent paper.

### 6 CONCLUSIONS AND DISCUSSION

In this work we presented for the first time a comprehensive multi-survey reconstruction effort of the primordial and evolved density fields with the COSMIC BIRTH algorithm performed in the COSMOS field during the epoch of Cosmic Noon ($1.4 \leq z \leq 3.6$). To our knowledge, this is the ever attempted large-scale structure analysis of its kind at high redshift, probing the quasi-linear regime of gravitational structure formation, before non-linear shell-crossing started dominating the emerging cosmic web.

We combined the data from five spectroscopic galaxy surveys in the COSMOS field, which have partially overlapping footprints, and do not share a common observing strategy. Therefore, we had to make a special effort in estimating the angular selection function for each survey, based on the the individual targeting strategies for the parent photometric catalogs, which was missing for these surveys.

Also, we applied for the first time a multi-tracer and multi-survey likelihood formalism in Lagrangian space within a Bayesian inference framework. This allowed us to combine surveys with different selection functions, galaxy bias, number densities, and redshift ranges, connected through the underlying dark matter distribution. Although earlier works have presented reconstructions using multiple galaxy populations, they however shared a unique survey geometry. Therefore, our new method is more general and has a wider range of possible applications. Since we expect a correlation between the spatial distribution of the galaxy catalogues, a covariance term would in principle be necessary when different observations are combined within one joint likelihood analysis. We bypass this problem by firstly mapping all tracers to Lagrangian coordinates. This allows us to assume Poisson likelihoods, since only in the homogeneous epoch before gravity coupled separated spatial regions, identical and independently distributed large-scale structure tracers can be assumed.

Despite of the large number of surveys, the COSMIC BIRTH code showed an efficient performance, converging within about 40 iterations, and showing low correlation lengths of about 30 iterations. The resulting reconstructions reveal for the first time a holistic view on the matter density field and its primordial fluctuations jointly inferred from five spectroscopic surveys.

We also revised the bias of star forming galaxies towards high redshifts, finding some moderate evidence for a stronger evolution than the one described by passive evolution.

The inferred density fields have a large number of potential applications. In particular, we have found several high density regions across the whole reconstructed volume. We successfully reconstructed a number of observationally known proto-cluster regions previously reported by Cucciati et al. (2014); Chiang et al. (2015); Diener et al. (2015); Casey et al. (2015); Lee et al. (2016); Wang et al. (2016); Nanayakkara et al. (2016); Darvish et al. (2020). However, these previous proto-cluster studies were typically performed using individual galaxy surveys directly on the observed galaxy positions in redshift-space. Major improvements have been done combining two surveys and statistically sampling the redshifts in two-dimensional slices including a Voronoi tessellation to estimate the galaxy density field (Cucciati et al. 2018). Still, the mass estimates are prone to projection effects due to their peculiar velocities (Kaiser 1984, 1987), and also assuming a velocity dispersion measure which is not valid for non-virialised objects at high redshifts. In a subsequent paper, we will carry out a more detailed analysis of these structures. Since we have reconstructed the initial conditions of the COSMOS volume, we will be able to run
constrained N-body simulations based on the inferred initial conditions, enabling us to study the full cosmic evolution of the proto-clusters in detail (Ata et al. in prep.). This will, for example, permit us to direct model the late-time properties of the galaxy clusters that will coalesce from the proto-clusters observed in the COSMOS surveys. Furthermore, the Bayesian formalism will permit us to directly quantify the uncertainty of late-time cluster properties.

In addition to studying the large-scale structure evolution, there are many potential applications for the contemporary density field derived in COSMOS. Our density map could be used to directly address the question of galaxy evolution in the context of environment, without having to use contrived statistics (e.g., counts in cylinders, N-nearest neighbours) to define the galaxy environment (Cooper et al. 2008; Koyama et al. 2013; Kawinwanichakij et al. 2017; Muldrew et al. 2018; Ji et al. 2018). More remarkably, by treating the observed COSMOS galaxies as tracer particles in the constrained N-body simulations based on these reconstruction, we will also be able to track them to their eventual \( z = 0 \) environments and thus link them directly with well-studied trends in the Local Universe (e.g., Kauffmann et al. 2004).

Another clear application is to compare the dark matter reconstructions from galaxy tracers with reconstructions obtained from hydrogen Lyman-\( \alpha \) forest tomography observations in the CLAMATO survey (Lee et al. in prep.). This will allow to directly test the fluctuating Gunn-Peterson approximation which posits a monotonic relationship between Lyman-\( \alpha \) forest absorption and the underlying density field. While the CLAMATO data is reconstructed moderate-resolution, low-S/N LBG spectra, there also exist several high-resolution absorption spectra that have been observed of bright quasars in the field. On these high-resolution spectra, we will aim also to carry out an analysis of the line widths and column densities as a function of the underlying density field. The CLAMATO data could be used to directly address the question of galaxy environments and thus link them directly with well-studied trends in the Local Universe (e.g., Kauffmann et al. 2004).

The method applied in this study is also interesting for the Galaxy Evolution component of the planned Subaru PFS Subaru Strategy Program, which will target ~12 – 15 deg of deep spectroscopy over three continuous fields, i.e. an order-of-magnitude larger area than COSMOS. The galaxy number densities are also very comparable, or better, than those reconstructed in this paper. For the lower-redshift NIR-selected sample at \( z \sim 1 - 1.5 \), the PFS will obtain spectra for galaxies at a number density of \( n \approx 3 \times 10^{-3} h^3 \text{Mpc}^{-3} \), slightly better than the FMOS-COSMOS sample which has \( n \approx 2 \times 10^{-3} h^3 \text{Mpc}^{-3} \) at \( z \sim 1 - 1.7 \). This should enable direct cosmic web analysis from the galaxy reconstructions alone. Closer to the peak of Cosmic Noon at \( z \sim 2 - 3 \), an optically-selected LBG sample is planned to yield number densities equivalent to VUDS or zCOSMOS-deep (\( n \approx 3 \times 10^{-3} h^3 \text{Mpc}^{-3} \)). While this is less than the combined survey sample used in this paper, it should still be sufficient to identify and characterise proto-clusters within the volume.

As wide-field spectroscopic surveys probe ever deeper into cosmic history, the utility of density reconstruction techniques will become increasingly important to fully exploit the rich scientific possibilities that will open up.

ACKNOWLEDGEMENTS

The authors thank Bahram Mobasher and Karl Glazebrook for their help with the MOSDEF and ZFIRE surveys, respectively. MA thanks Jiaxin Han, Ben Granett and Masahiro Takada for helpful discussions and the hospitality at the IAC. FSK thanks for the support from the grants RYC2015-18693, SEV-2015-0548 and AYA2017-89891-P. KGL acknowledges support from JSPS KAKENHI Grant Number JP19K14755.

REFERENCES

Abel T., Hahn O., Kaehler R., 2012, MNRAS, 427, 61
Alam S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 470, 2617
Allen S. W., Evrard A. E., Mantz A. B., 2011, Annual Review of Astronomy and Astrophysics, 49, 409
Amara A., et al., 2012, MNRAS, 424, 553
Amendola L., et al., 2018, Living Reviews in Relativity, 21
Ata M., Kitaura F.-S., Müller V., 2015, MNRAS, 446, 4250
Ata M., et al., 2017, Monthly Notices of the Royal Astronomical Society, 467, 3993
Ata M., et al., 2018, Monthly Notices of the Royal Astronomical Society, 473, 4773
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, The Astrophysical Journal, 304, 15
Beckwith S. V. W., et al., 2006, The Astronomical Journal, 132, 1729
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Belfiore F., et al., 2017, Monthly Notices of the Royal Astronomical Society, 466, 2570
Bernetarde F., 1994, The Astrophysical Journal, 427, 51
Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93
Bond J. R., Komlan L., Pogosyan D., 1996, Nature, 380, 603
Bouchet F. R., Juszkiewicz R., Colombi S., Pellat R., 1992, ApJ, 394, L5
Bouchet F. R., Colombi S., Hivon E., Juszkiewicz R., 1995, A&A, 296, 575
Brammer G. B., et al., 2012, ApJS, 200, 13
Capak P., et al., 2007, The Astrophysical Journal Supplement Series, 172, 99
Casey C. M., et al., 2015, Astrophysical Journal Letters, 808, L33
Casey C. M., Hodge J., Zavala J. A., Spilker J., da Cunha E., Staguhn J., Finelstein S. L., Drew P., 2018, ApJ, 862, 78
Catelan P., 1995, MNRAS, 276, 115
Chiang Y. K., Overzier R., Gebhardt K., 2013, Astrophysical Journal, 779, 127
Chiang Y. K., et al., 2015, Astrophysical Journal, 808, 37
Chiang Y.-K., Overzier R. A., Gebhardt K., Henriques B., 2017, The Astrophysical Journal, 844, L23
Cirasuolo M., et al., 2014, MOONS: the Multi-Object Optical and Near-infrared Spectrograph for the VLT. p. 91470N, doi:10.1111/12.2056012
Cochrane R. K., Best P. N., Sobral D., Smail I., Wake D. A., Stott J. P., Geach J. E., 2017, MNRAS, 469, 2913
Cohn J. D., White M., 2008, Mon. Not. R. Astron. Soc, 385, 2025
Coles P., Jones B., 1991, MNRAS, 248, 1
Cooper M. C., Tremonti C. A., Newman J. A., Zabludoff A. I., Geach J. E., 2017, MNRAS, 469, 2913
Cooper M. C., Tremonti C. A., Newman J. A., Zabludoff A. I., 2013, Monthly Notices of the Royal Astronomical Society, 430, 245
Costanzi Alunno Cerbolini M., Sartoris B., Xia J. Q., Biviano A., Borgani S., Viel M., 2013, Journal of Cosmology and Astroparticle Physics, 2013
Cucciati O., et al., 2014, Astronomy and Astrophysics, 570
Cucciati O., et al., 2018, Astronomy and Astrophysics, 619

MNRAS 000, 1–19 (2020)
BIRTH of the COSMOS Field

Nanayakkara T., et al., 2016, ApJ, 826, 21
Neal R. M., 1993, in Technical Report CRG-TR-93-1. No. September. Dept. of Computer Science, University of Toronto
Neal R. M., 2011, Handbook of Markov Chain Monte Carlo, pp 113–162
Newman A. B., et al., 2020, ApJ, 891, 147
Neyrinck M. C., 2013, Monthly Notices of the Royal Astronomical Society, 428, 141
Neyrinck M. C., 2016, MNRAS, 455, L11
Neyrinck M. C., Aragón-calvo M. A., Jeong D., Wang X., 2014, Monthly Notices of the Royal Astronomical Society, 441, 646
Nusser A., Davis M., 1994, The Astrophysical Journal, 421, L1
Numa S. E., Kitaura F. S., Heß S., Libeskind N. I., Muller V., 2014, Monthly Notices of the Royal Astronomical Society, 445, 988
Okada H., et al., 2016, Publications of the Astronomical Society of Japan, 68, 47
Patrick Bos E. G., Kitaura F. S., van de Weygaert R., 2019, Monthly Notices of the Royal Astronomical Society, 488, 2573
Peebles P. J. E. P. J. E., 1980, The large-scale structure of the universe. Princeton University Press
Percival W. J., Cole S., Eisenstein D. J., Nichol R. C., Peacock J. A., Pope A. C., Szalay A. S., 2007, Monthly Notices of the Royal Astronomical Society, 381, 1053
Pezzotta N., et al., 2017, Astronomy & Astrophysics, 604, A33
Pichon C., Vergely J. L., Rollinde E., Colombi S., Petitjean P., 2001, Monthly Notices of the Royal Astronomical Society, 326, 597
Porqueres N., Jasche J., Lavaux G., Enßlin T., 2019, Astronomy & Astrophysics, 630, A151
Reid B. A., White M., 2011, MNRAS, 417, 1913
Rorai A., Carswell R. F., Haehnelt M. G., Becker G. D., Bolton J. S., Murphy M. T., 2018, MNRAS, 474, 2871
Sato M., Hamana T., Takahashi R., Takada M., Yoshida N., Matsubara T., Sugiyama N., 2009, Astrophysical Journal, 701, 945
Scodoglio M., et al., 2018, A&A, 609, A84
Scoville N., et al., 2007, ApJS, 172, 1
Scoville N., et al., 2013, May The Astrophysical Journal Supplement Series, 206, 2013
Shi K., et al., 2019, The Astrophysical Journal, 879, 9
Silverman J. D., et al., 2015, Astrophysical Journal, Supplement Series, 220, 12
Skrutskie M. F., et al., 2006, The Astronomical Journal, 131, 1163
Smaili N., et al., 2017, Astronomy and Astrophysics, 597
Somerville R. S., Davé R., 2015, ARA&A, 53, 51
Sorce J. G., et al., 2016, MNRAS, 455, 2078
Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, ApJ, 462, L17
Steidel C. C., Shapley A. E., Pettini M., Adelberger K. L., Erb D. K., Reddy N. A., Hunt M. P., 2004, The Astrophysical Journal, 604, 534
Straatman C. M. S., et al., 2016, ApJ, 830, 51
Takada M., Hu W., 2013, Physical Review D - Particles, Fields, Gravitation and Cosmology, 87, 123504
Takada M., et al., 2014, Publications of the Astronomical Society of Japan, 66
Tegmark M., Peebles P. J. E., 1998, The Astrophysical Journal, 500, L79
Thomas S. A., Abdalla F. B., Lahav O., 2011, Phys. Rev. Lett., 106, 241301
Tinker J. L., 2007, MNRAS, 374, 477
Tomczak A. R., et al., 2019, Monthly Notices of the Royal Astronomical Society, 484, 4695
Tully R. B., Courtois H., Hoffman Y., Pomarède D., 2014, Nature, 513, 71
Wang H., Mo H. J., Yang X., Van Den Bosch F. C., 2013, Astrophysical Journal, 772, 63
Wang H., Mo H. J., Yang X., Jing Y. P., Lin W. P., 2014, Astrophysical Journal, 794, 94
Wang T., et al., 2016, The Astrophysical Journal, 828
Watson W. A., Iliev I. T., D’Alloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, Monthly Notices of the Royal Astronomical Society, 433, 1230
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Yahil A., Strauss M. A., Davis M., Huchra J. P., 1991, The Astrophysical Journal, 372, 380
Yuan T., et al., 2014, ApJ, 795, L20
Zarouhi S., Hoffman Y., Fisher K. B., Lahav O., 1995, The Astrophysical Journal, 449, 446
Zucca E., et al., 2009, A&A, 508, 1217
de Jong R. S., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, p. 84600T, doi:10.1117/12.956239

APPENDIX A: IMPACT OF AN INACCURATE BIAS TREATMENT

Even though the data region within the reconstructed mesh grid only occupies a small fraction, wrong galaxy bias estimates can have a strong impact on the accuracy of the reconstructions, as we demonstrate here. In particular, we perform a COSMIC BIRTH run on a sub-volume covering the redshift range of $1.8 < z < 2.2$ with the data described in Section 3. We assume a large scale bias $\sim 60\%$ systematically higher than the one from our regular runs, as presented in Figure 6 and discussed in Section 2.2.1. This bias exceeds the mean observational measurements in that redshift range by $2\sigma$, and therefore represents a conservative upper bias limit (see the measurement by Kashino et al. 2017, extrapolated to higher redshifts with passive evolution in Figure 6).

The results for this test are represented in Figure A1. On the top we show the inferred density field $\delta(x)$ with the galaxy positions plotted on top represented by black dots. Although the reconstructed structures match the galaxy positions, a deeper inspection shows that the data region around DEC = 2.2 is under-weighted, due to the overestimated bias value and artificially high density structures at the edge of the data region are formed. In the bottom plot we show the mean power spectrum (red dashed) and the $1\sigma$ standard deviation (grey band) of the inferred initial density fields for 1000 samples after convergence including the ratio of the inferred power spectra with the theoretical $\Lambda$CDM prediction below. Considering the low volume filling fraction, unbiased power spectra with the theoretical one are expected. However, a clear excess of power can be seen at the lowest modes, excluding the proposed bias values in this test. A thorough analysis of the bias, will be presented in a forthcoming paper.

This paper has been typeset from a TeX/LaTeX file prepared by the author.
Figure A1. Results from COSMIC BIRTH runs with a bias exceeding the mean measurement given by the literature by $2\sigma$ in the redshift range $1.8 < z < 2.2$. Top: Slice plot of the light-cone density field $\delta(s)$ with the galaxy positions plotted on top represented by black dots. Bottom: Power spectra of the corresponding initial density fields $\delta(q)$. We show the mean (blue dashed line) and the standard deviation (grey band) for 1000 HMC iterations, including the ratio with the theoretical $\Lambda$CDM power spectrum in the panel below.