Informed Total-Error-Minimizing Priors:
Interpretable cosmological parameter constraints despite complex nuisance effects

Bernardita Ried Guachalla[1,2,3,4,★], Dylan Britt[1,2,3,4], Daniel Gruen[1,5], and Oliver Friedrich[1,5]

[1] University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstraße 1, 81679 Munich, Germany
[2] Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
[3] Kavli Institute for Particle Astrophysics & Cosmology, 452 Lomita Mall, Stanford, CA 94305, USA
[4] SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA
[5] Excellence Cluster ORIGINS, Boltzmannstraße 2, 85748 Garching, Germany

ABSTRACT

While Bayesian inference techniques are standard in cosmological analyses, it is common to interpret resulting parameter constraints with a frequentist intuition. This intuition can fail, e.g. when marginalizing high-dimensional parameter spaces onto subsets of parameters, because of what has come to be known as *projection effects* or *prior volume effects*. We present the method of Informed Total-Error-Minimizing (ITEM) priors to address this. An ITEM prior is a prior distribution on a set of nuisance parameters, e.g. ones describing astrophysical or calibration systematics, intended to enforce the validity of a frequentist interpretation of the posterior constraints derived for a set of target parameters, e.g. cosmological parameters. Our method works as follows: For a set of plausible nuisance realizations, we generate target parameter posteriors using several different candidate priors for the nuisance parameters. We reject candidate priors that do not accomplish the minimum requirements of bias (of point estimates) and coverage (of confidence regions among a set of noisy realizations of the data) for the target parameters on one or more of the plausible nuisance realizations. Of the priors that survive this cut we select the ITEM prior as the one that minimizes the total error of the marginalized posteriors of the target parameters. As a proof of concept, we apply our method to the Density Split Statistics (DSS) measured in Dark Energy Survey Year 1 data. We demonstrate that the ITEM priors substantially reduce prior volume effects that otherwise arise and allow sharpened yet robust constraints on the parameters of interest.

Key words. Methods: statistical, data analysis, numerical – Cosmology: cosmological parameters

1. Introduction

Present and near-future wide-area surveys of galaxies will provide an unprecedented volume of data over most of the extragalactic sky. Examples of these are the Dark Energy Survey (DES[1]) (Collaboration: et al. 2016, Sevilla-Noarbe et al. 2021), the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST[2]) (LSST Science Collaboration et al. 2009, Ivezić et al. 2019), the Dark Energy Spectroscopic Instrument (DESI) survey[3] (Collaboration et al. 2016), the space mission Euclid[4] (Laureijs et al. 2011), the 4m Multi-Object Spectroscopic Telescope (4MOST)[5] (de Jong et al. 2012), the Kilo-Degree Survey (KiDS)[6] (de Jong et al. 2012), the Hyper Suprime-Cam (HSC)[7] (Aihara et al. 2017), and the Nancy Grace Roman Space Telescope[8] (Eifler et al. 2021). The promise of precision cosmology with these data can only be realized by accurately accounting for nuisance effects – both astrophysical and in the calibration uncertainty of data – with increasingly complex models. Examples of astrophysical effects include intrinsic alignments of galaxies as a nuisance for weak lensing analyses (see Troxel & Ishak 2015 and Lamman et al. 2024 for recent reviews and Secco et al. 2022, Samuroff et al. 2023, Asgari et al. 2021 and Dalal et al. 2023 for the latest analyses), baryonic physics that impact the statistics of the cosmic matter density field such as star formation, radiative cooling and feedback, (e.g. Cui et al. 2014; Velliscig et al. 2014; Mummery et al. 2017: Tröster et al. 2022), galaxy bias and other parameters describing the galaxy-matter connection (see Schmidt 2021 for a recent review, Scherrer & Weinberg 1998, Dekel & Lahav 1999, Baldauf et al. 2011, Heymans et al. 2021, Ishikawa et al. 2021 and Pandey et al. 2022), or systematic effects on the baryon acoustic oscillation signal (Seo & Eisenstein 2007, Ding et al. 2018, Rosell et al. 2021 and Lamman et al. 2023). Calibration-related nuisance effects include measurement biases on galaxy shapes (see Mandelbaum et al. 2019 for a review, MacCrann et al. 2021, Georgiou et al. 2021 and Li et al. 2022), the estimation of redshift distributions of photometric galaxy samples (see Newman & Gruen 2022 for a review and Tanaka et al. 2017, Huang et al. 2017, Hildebrandt et al. 2021, Myles et al. 2021, Cordero et al. 2022), or systematic clustering of galaxies induced by observational effects (e.g. Cawthon et al. 2022, Baleato Lizancos & White 2023).

Cosmological inference from survey data is performed through likelihood analyses. The first step is to predict a model data vector as a function of parameters (both cosmological and nuisance) using a theoretical modeling pipeline. The second step is to infer the posterior probability distribution of the parameters using the observed data vector as a noisy measurement. Finally, one obtains a marginalized posterior for a subset of parameters by projecting out the remaining parameters. The results are summarized in the form of point estimates (best estimate of a parameter value) with their confidence errors. We might desire for these marginalized posteriors to have certain properties:

- *Unbiasedness:* We require the point estimate of each cosmological parameter to have a bias, with respect to the fiducial parameters, smaller than some fraction of the width of the confidence interval (e.g. to be inside of the 1-σ confidence interval or region if the space is 2 or more dimensions). In Sec. 2.3.1 we review which point estimators are currently used in cosmology.
- *Reliability:* In repeated trials (with the same realization of the nuisance, independent noise, and the same fiducial values for the cosmological parameters), we impose that the "true" values of a set of cosmological parameters should be within a derived confidence region of those parameters at least some desired fraction of times. In other words, that the frequentist coverage of independent realizations is consistent with the confidence level (the idea of matching Bayesian and frequentist coverage probabilities has been discussed extensively in Percival et al. 2021).
- *Robustness:* We demand that the previous two requirements hold across a list of plausible nuisance and physical realizations of the fiducial (e.g. under two different fiducial nuisance parameters, we still can recover unbiased and reliable marginal posteriors on the cosmology).
- *Precision:* Among all the possible priors whose resulting confidence regions fulfill the above requirements, we prefer the one which, on average, results in the smallest total, that is, statistical and systematic, uncertainty on the cosmological parameters.

There are two ways in which the treatment of nuisance parameters can cause *systematic* errors and, therefore, violate these statements.

The first one is to assume a nuisance parameter model that is overly simplistic, for instance by fixing a nuisance parameter that indeed needs to be free to accurately describe the observed data. In the left panel of Fig. 1, we present a simple 2D sketch of this situation, in which the assumption that the nuisance parameter $\theta_{\text{Nuisance}} = 0$ leads to an unavoidable bias on the cosmological parameter θ_{Target}[9]. We will call this effect *under-fitting* of the nuisance, following Bernstein (2010). It is expected with coming cosmological surveys that the complexity of the data will increase, and therefore, unknown or not modeled nuisance effects could become under-fitted.

The second is to assume a prior distribution such that the estimated parameters on the marginalized posterior are biased with respect to the "true" values. This systematic error is known as *prior volume effect* or as *projection effect*, and is shown in the right panel from Fig. 1. We emphasize that the prior volume effect is a continuous one, i.e. there are cases in which this error is larger than others, and it is not limited to cases where

distributions are non-symmetric or exclusively non-Gaussian. To avoid confusion, we limit ourselves to the definition of prior volume effect stated above, in which the point estimates calculated from the marginalized posteriors will be substantially offset from the true value being predicted. For real-life examples of the effect, see e.g. Secco et al. (2022); Amon et al. (2022) on intrinsic alignments, Sugiyama et al. (2020); Pandey et al. (2022) on galaxy bias or Sartoris et al. (2020) on galaxy cluster mass profiles. For studies on overcoming this effect in weak lensing analyses see Joachimi et al. (2021), Chintalapati et al. (2022), and Campos et al. (2023).

Another consequence of the prior volume effect is that the *marginalized coverage fraction* (the fraction of marginalized confidence intervals in repeated trials that encompass the true target parameters) may not be consistent with the associated credible level. The Bayesian and frequentist interpretations imply that the full posterior's confidence region does not necessarily have the same the credible region. If this is the case, when marginalizing, these two effects add, and the marginalized coverage fraction is less intuitively characterized. This can be rephrased as follows: the marginalized 1-σ confidence contours obtained with repeated Bayesian analyses in an ensemble of realizations do not necessarily contain the fiducial configuration with a frequency of $\sim 68\%$. The discrepancy may not constitute a problem from a purely Bayesian standpoint since Bayesian statistics do not claim to satisfy frequentist expectations. However, this does not change the fact that one may want to make probabilistic statements about the true value of a physical parameter and that marginalized parameter constraints quoted in cosmological publications are interpreted in a frequentist way by a large fraction (if not the majority) of the cosmology audience. Projection effects in high-dimensional nuisance parameter spaces may hence cause a buildup of wrong intuitions in the inferred parameters. Therefore, in the absence of projections, we would only find the general mismatch between the confidence and credible regions of the marginalized posteriors, and this mismatch can be overcome by using solutions like the one presented in Percival et al. (2021) and Chintalapati et al. (2022).

The situation is further complicated by the fact that a precise model for nuisance effects, with a finite set of parameters and well-motivated Bayesian priors, is often not available. Commonly, at best, some plausible configurations of the nuisance effect are known. These could e.g. be a set of summary statistics measured from a range of plausible hydrodynamical simulations (van Daalen et al. 2019, Walther et al. 2021) or a compilation of different models and parameters that have been found to approximately describe the nuisance, as is the case for intrinsic alignment (Hirata & Seljak 2004; Kiessling et al. 2015; Blazek et al. 2019), the galaxy-matter connection (Szewciw et al. 2022, Voivodic & Barreira 2021, Friedrich et al. 2022, Britt et al. 2024) or in the calibration of redshift distributions (Cordero et al. 2022).

In this paper, our objective is to make statements about the marginalized posteriors of the target parameters that fulfill the above criteria of *unbiasedness*, *reliability*, *robustness*, and *precision*. These are pragmatic objectives, and because they are not consistent with pure Bayesian approaches, one must violate the latter in order to meet the former. We choose to do this by defining the prior on nuisance effects not directly by our prior knowledge of their parameter values, but rather in such a way that our objectives are met. Since the nuisance priors we construct to satisfy these criteria are informed (by knowledge of possible configurations of the nuisance effects) and they minimize the systematic and statistical errors, we call

---

[9] Given our work may be relevant to areas beyond cosmology, we will use the term "target parameter" and "cosmological parameter" interchangeably for the rest of this paper.
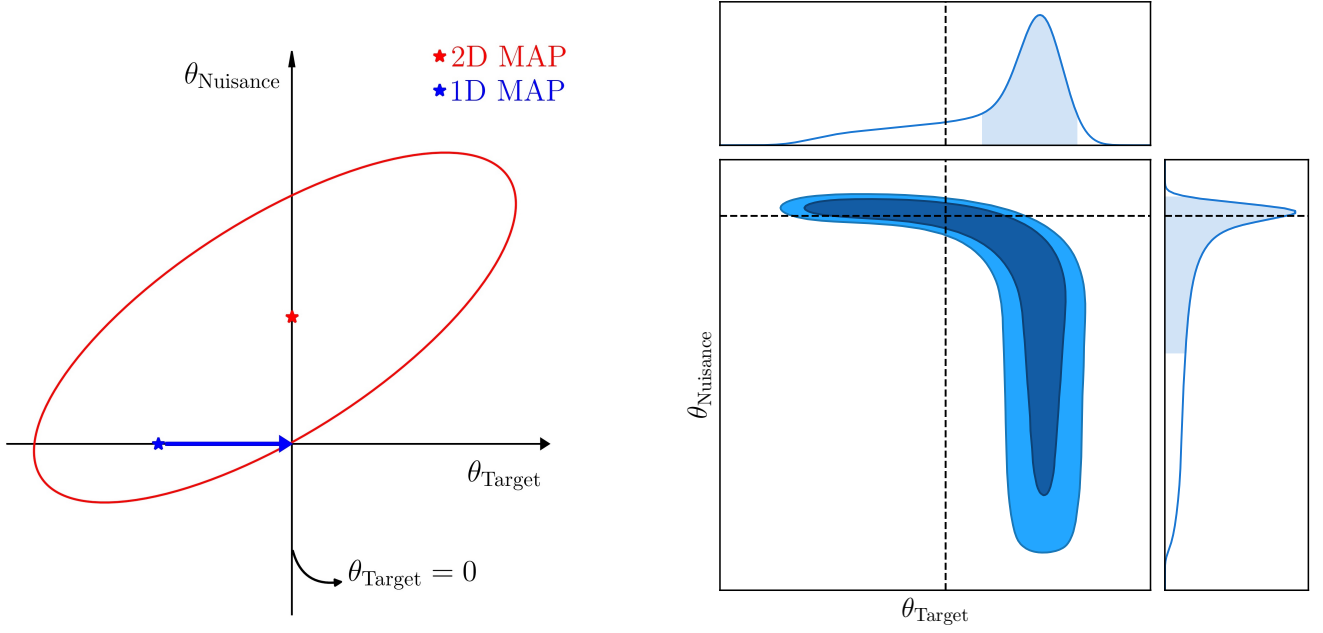
Fig. 1: Sketches on how the treatment of nuisance parameters can cause systematic errors on the target parameter posteriors. For both panels, we assume a 2D space where the parameter $\boldsymbol{\theta} = (\theta_{\text{Target}}, \theta_{\text{Nuisance}})$. *Left:* **Bias due to underfitting a nuisance effect**: In a 2D model, let 0 be the real value of the parameter θ_{Target}, despite the true value of θ_{Nuisance}. When θ_{Nuisance} is a free parameter of the 2D Gaussian model presented in the panel, the true value of the target parameter is recovered with the 2D Maximum a Posteriori (2D MAP) corresponding to the red mark. With a simpler nuisance model that fixes $\theta_{\text{Nuisance}} = 0$ and thus under-fits the nuisance, the 1D MAP (blue mark) of θ_{Target} is biased towards negative values as a result. *Right:* **Prior volume effect**: In this toy illustration, we show the 2D and 1D marginalized posteriors of a distribution with the 1-σ and 2-σ confidence contours with blue and light blue respectively. Such posteriors can, e.g., result from a model that is highly non-Gaussian in the full posterior space. The black dashed line represents the fiducial value, and the intersection of them is inside the 1-σ contour. If we only have access to the 1D marginalized posterior of θ_{Target}, we would hardly estimate the fiducial value with our point estimators. In other words, even if the posterior in the entire parameter space is centered around the correct parameters, the posterior marginalized over the nuisance parameters can be off of the correct target parameters. This bias error is called the *prior volume effect*, and in this paper, we propose a pipeline to overcome it by constraining systematically the priors on the nuisance parameters.

them *Informed Total-Error-Minimizing (ITEM) priors*. How we construct these priors has two important consequences for the interpretation of the resulting parameter constraints:

- In the presence of limited information about the nuisance effect, the considered realizations can be used to tune the priors of the nuisance parameters to ensure certain properties of the resulting constraints on the target parameters. Hence, we do not assume the ITEM prior directly represents information about the nuisance effect or our knowledge about it. It is merely a tool for interpretable inference of the target parameters. As a consequence, we relinquish the ability to derive meaningful posterior constraints on the nuisance parameters.
- The coverage fraction and bias of any procedure to derive parameter constraints in a set of repeated experiments will be sensitive to the "true" values of cosmological and nuisance parameters that underlie those experiments. Given a set of plausible realizations of cosmology and nuisances, we can therefore only ensure a *minimum coverage* and a *maximum bias* of the constraints derived from ITEM priors.

This paper is structured as follows: We start by describing the general methodology for obtaining ITEM priors in Sec. 2. In Sec. 3, we investigate the performance of the method in a test case: cosmological parameter constraints marginalized over models for non-Poissonian shot-noise in Density Split Statistics

(Friedrich et al. 2018; Gruen et al. 2018). We explore the change in simulated and DES Y1 results of these statistics when using ITEM priors defined from a simplified set of plausible nuisance realizations. Finally, discussion and conclusions are given in Sec. 4.

2. Methodology

Assume that some observational data can be described in a data vector $\hat{\boldsymbol{\xi}}$ of N_d data points. Let $\boldsymbol{\xi}[\boldsymbol{\theta}]$ be a model for this data vector that depends on a parameter vector $\boldsymbol{\theta}$ of N_p parameters, and let $\mathbf{C}$ be the covariance matrix of $\hat{\boldsymbol{\xi}}$. In many cases, it is reasonable to assume that the likelihood $\mathcal{L}$ of finding $\hat{\boldsymbol{\xi}}$ given the parameters $\boldsymbol{\theta}$ is a multivariate Gaussian distribution:

$$\ln \mathcal{L}(\hat{\boldsymbol{\xi}}|\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i,j} (\hat{\xi}_i - \xi_i[\boldsymbol{\theta}]) \mathrm{C}_{ij}^{-1} (\hat{\xi}_j - \xi_j[\boldsymbol{\theta}]) + C \, , \tag{1}$$

where C_{ij}^{-1} are the elements of the inverse of the covariance matrix $\mathbf{C}$, and C is a constant as long as the covariance does not depend on the model parameters. We derive the posterior distribution $\mathcal{P}(\hat{\boldsymbol{\xi}}|\boldsymbol{\theta})$ using Bayes' theorem:

$$\mathcal{P}(\boldsymbol{\theta}|\hat{\boldsymbol{\xi}}) \propto \mathcal{L}(\hat{\boldsymbol{\xi}}|\boldsymbol{\theta})\Pi(\boldsymbol{\theta}) \, , \tag{2}$$

with $\Pi(\boldsymbol{\theta})$ being a prior probability distribution incorporating *a priori* knowledge or assumptions on $\boldsymbol{\theta}$.
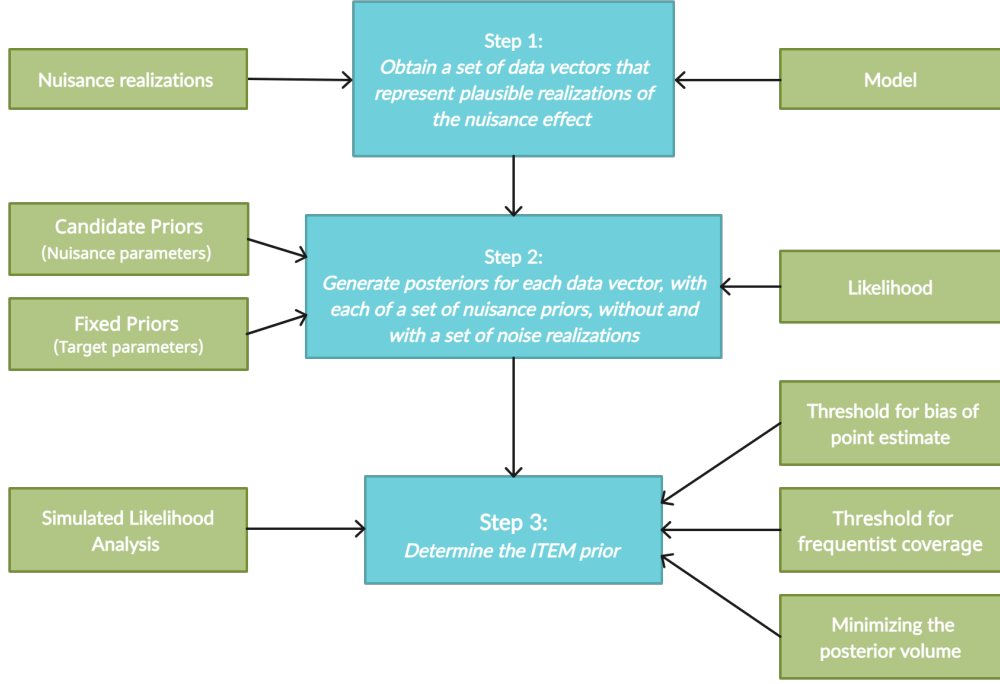
Fig. 2: Basic steps and concepts present in our ITEM prior methodology. In light blue, we summarize the steps of our pipeline, and in green, we highlight the ingredients of the methodology.

In many situations, parameters $\boldsymbol{\theta}$ can be divided into two types: target parameters $\boldsymbol{\theta}_T$ and nuisance parameters $\boldsymbol{\theta}_N$, as done in Fig. 1 (θ_{Target} and θ_{Nuisance} respectively). This distinction is subjective and depends on which parameters one would like to make well-founded statements about (i.e., *robust*, *valid*, *reliable*, and/or *precise* estimates). Also, this division is based on which parameters are needed to describe the complexities of the experiment but are not considered intrinsically of interest. For example, in a cosmological experiment, the target parameters are usually those describing the studied cosmological model, such as the Λ-Cold Dark Matter (ΛCDM) and its variations. This could, e.g., include the present-day density parameters (Ω_x, with x representing a specific content of the universe), the Hubble constant (H_0), and the equation of state parameter of dark energy (w). On the other hand, the nuisance parameters account for observational and astrophysical effects, or uncertainties in the modeling of the data vector.

Recent cosmological analyses have chosen an approximate nuisance model with a limited number of free parameters for all these complex effects. These either assume a wide flat prior distribution over the nuisance parameters or an informative prior distribution set by simulation and/or calibration measurements (Aghanim et al. 2020; Asgari et al. 2021; Sugiyama et al. 2022; Abbott et al. 2022). As we explained in Sec. 1, precarious modeling of nuisance parameters can cause systematic errors, leading to the following consequences:

– *Limited modeling of the nuisance effects*: The assumed model $\boldsymbol{\xi}[\boldsymbol{\theta}]$ may not sufficiently describe the nuisance effects. For example, there could be an oversight of one underlying source of the nuisance, the model of a nuisance effect may be truncated at too low an order, or it can be entirely neglected. This results in an intrinsic bias on the inferred target parameters, as would, e.g., be revealed by performing

the analysis on simulated data that include the full nuisance effect.
– *Prior-volume effects*: The prior distribution $\Pi(\boldsymbol{\theta}_N)$ over one or multiple nuisance parameters could result in biases in the marginalized posterior when projecting to lower dimensions.

In this work, we introduce a procedure that determines a prior distribution on the nuisance parameters which ensures the criteria presented in Sec. 1 for marginalized posteriors of the target parameters. We will call the resulting priors *Informed Total-Error-Minimizing* priors, or ITEM priors. In Sec. 2.1, we describe how to obtain different data vectors given a model and plausible realizations of the nuisance effect. In Sec. 2.2, we generate posteriors for each data vector given three ingredients: the likelihood of the model, a set of fixed prior distributions for the target parameters, and a collection of nuisance candidate priors. In Sec. 2.3 we determine the ITEM prior after requiring the posteriors to account for specific criteria based on maximum biases, a minimum coverage fraction, and the minimization of the posterior uncertainty. The steps are summarized in Fig. 2.

2.1. Step 1: Obtain a set of data vectors that represent plausible realizations of the nuisance effect

Assume that at fixed target (e.g. cosmological) parameters $\boldsymbol{\theta}_T$, the range of plausible realizations of a nuisance effect is well represented by a set of n different resulting data vectors:

$$\hat{\boldsymbol{\xi}}^i = \boldsymbol{\xi}[\boldsymbol{\theta}_T, \text{nuisance realization } i], \text{ for } i = 1, ..., n. \quad (3)$$

These could, e.g. be based on a model for the nuisance effect with different nuisance parameter values. Alternatively, the $\hat{\boldsymbol{\xi}}^i$ could result from a set of simulations with different assumptions that produce realizations of the nuisance effect (e.g. the impact of baryonic physics on the power spectrum from a variety of
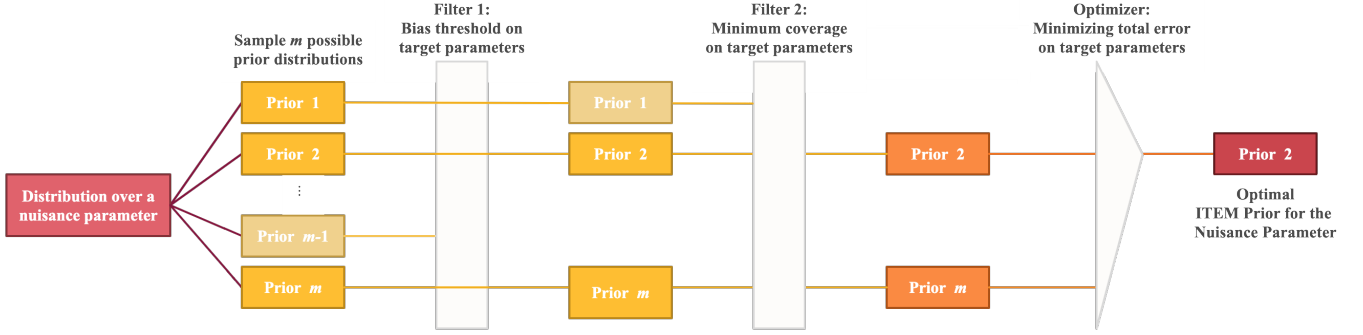
Fig. 3: Flowchart of the procedure to obtain the ITEM prior. First, a set of m candidate prior distributions is proposed for the nuisance parameters. Then, the first criterion filters some of the priors by requiring that the bias on the target parameters in a simulated likelihood analysis be less than x^* times its posterior's statistical uncertainty (see Eq. 6). Second, we require that the remaining priors result in a minimum coverage y^* (see Eq. 8) over the target quantities. This leads to $m' \leq m$ candidate priors, which are finally optimized by minimizing the target parameters' uncertainty (see Sec. 2.3.3).

hydrodynamical simulations, as done in Chisari et al. 2018 and Huang et al. 2019). In some other cases, direct measurements of the nuisance effect may be used, and e.g. a bootstrapping of those measurements could yield a set of possible realizations. For the sake of notation, let us assume that the nuisance realizations are given in terms of particular nuisance parameter combinations θ_N^i of the considered nuisance model: $\hat{\xi}^i = \hat{\xi}[\theta_T, \theta_N^i]$. Note, however, that this is no assumption that a model to be used in the analysis shares those same nuisance parameters. For the following tests to be meaningful, we require these data vectors to contain no noise, or noise that is negligible compared to the covariance considered in the likelihood analysis.

In principle, it is possible to consider more than one realization of the target parameters as well (e.g., more than one cosmology for the same nuisance parametrization). This extra layer of complexity can be added to the pipeline, but as a first pass, we will limit ourselves to a fixed target parameter vector for simplicity, yet suggest doing this in future works.

2.2. Step 2: Generate posteriors for each data vector, with each of a set of nuisance priors, with and without a set of noise realizations

We choose our ITEM prior among the *space of nuisance priors*, i.e. the space of the parameters that describe a family of considered prior distributions $\Pi(\theta_N)$. The details of that family may depend on: the preferred functional form of the priors, on consistency relations that are *a priori* known, and on practical considerations like limits in the computational power. As mentioned before, the distinction between target and nuisance is subjective, and these labels may change between iterations of the pipeline (i.e., a prior on a target parameter can be fixed for one analysis, but it can be studied and optimized as a nuisance in another).

For example, we may search for an ITEM prior among a set of uniform prior distributions for one particular nuisance parameter, $\Pi(\theta_{\mathrm{Nuisance}}) \sim \mathcal{U}(a, b)$, where a and b are the lower and upper bounds, respectively. In that case, we propose to sample uniformly and independently over both a and b, subject to the consistency relation that $a < b$. In practice, we can e.g. sample m_a lower bounds ($a_i \in [a_1, a_2, ..., a_{m_a}]$) and m_b upper bounds ($b_j \in [b_1, b_2, ..., b_{m_b}]$). Besides requiring that $a_i < b_j$ for all tested pairs one could demand, if there was

some reason to, that other requirements are met, e.g. a minimum and maximum value for the prior width $b_j - a_i$, or that some nuisance parameter values are contained within the considered priors. When combining all these configurations, we will have up to $m_a \cdot m_b$ possible priors with different widths and midpoints.

This idea can be extended to many dimensions, and the combination of all possible priors that could be used will lead to a total number of priors we will denote by m. Therefore, for any given measurement of the data $\hat{\xi}$, this will result in $n \cdot m$ different posteriors on the full set of parameters θ.

$$\mathcal{P}_j(\theta|\hat{\xi}^i) \propto \mathcal{L}(\hat{\xi}^i|\theta)\Pi_j(\theta_N) , \qquad (4)$$

for $i = 1, ..., n$ and $j = 1, ..., m$.

In the next steps of our pipeline, we generate such sets of posteriors for different data vector realizations to determine which priors meet the criteria we outlined in Sec. 1. To obtain all these posteriors, instead of running a large number of Monte-Carlo-Markov-Chains (MCMCs) for the $n \cdot m$ total posteriors, we propose to do an importance sampling (Siegmund 1976, Hesterberg 2003) over the posterior derived from the widest prior (i.e., the prior on the nuisance that encloses all the other priors on its prior volume). We refer to Sec. 3.4.2 from Campos et al. (2023) for a summary on importance sampling. Additionally, we generate a collection of n_{noise} realizations by adding multi-variate Gaussian noise (according to our fiducial covariance) to each noise-less $\hat{\xi}^i$.

While we have focused here on a family of uniform priors, it is straightforward to generalize our pipeline to e.g. Gaussian prior distributions, or other functional forms that are commonly used in cosmological analyses.

2.3. Step 3: Determine the ITEM prior

Our goal is to find a prior on the nuisance parameters that returns robust, valid, reliable, and precise posteriors for the target parameters. To determine the ITEM prior, we apply filters to the considered family of nuisance prior distributions. The first and second ones are to set an upper limit for the bias and a lower limit for the coverage fraction of the posterior respectively. The third and final one is an optimization that minimizes the uncertainty on the target parameters. Fig. 3 synthesizes these steps in a diagram.

2.3.1. Criterion 1: Maximum bias for the point estimate

Constraints on target parameters are commonly reported as confidence intervals around a point estimate for those parameters. For example, that point estimate may be the maximum a posteriori probability (MAP) estimator, the maximum of a marginalized posterior on the target parameters, or the 1D marginalized mean (that we will refer as mean for simplicity). To devise a measure for the bias with respect to the true values of the target parameters let us return to the exercise that was already mentioned in Sec. 1: Consider a noise-free realization ξ of a data vector, and a data vector model $\xi[\boldsymbol{\theta}]$, which we assume to model the data perfectly, and the fiducial parameters $\boldsymbol{\theta}$ are known. We use this model to run an MCMC around ξ as if the latter was a noisy measurement of the data. If the model $\xi[\boldsymbol{\theta}]$ is non-linear in the nuisance parameters, then e.g. the MAP of the marginalized posterior for the target parameters obtained from that MCMC can be biased with respect to to the true parameters (even though ξ is noiseless and the model $\xi[\boldsymbol{\theta}]$ is assumed to be perfect).

In Sec. 2.1 we have obtained a set of realizations $\hat{\boldsymbol{\xi}}^i$ (with $i = 1, \ldots, n$) of the data, corresponding to n different realizations of the considered nuisance effects. For each marginalized posterior from $\hat{\boldsymbol{\xi}}^i$ we can obtain a point estimator. The difference between these estimates and the true target parameters underlying our data vectors will then act as a measure of how much projection effects bias our inference.

If $\boldsymbol{\theta}_T$ is the true target vector and $\hat{\boldsymbol{\theta}}_T^i$ is the vector estimate from the nuisance realization i, we then use the Mahalanobis distance (Mahalanobis 1936; Etherington 2019) to quantify the bias between the two, i.e.

$$x_i = \sqrt{(\boldsymbol{\theta}_T - \hat{\boldsymbol{\theta}}_T^i)^\top \mathbf{C}_i^{-1} (\boldsymbol{\theta}_T - \hat{\boldsymbol{\theta}}_T^i)} , \qquad (5)$$

where $\mathbf{C}_i^{-1}$ is the inverse covariance matrix of the parameter posterior, which we directly estimate from the chain for the i-th nuisance realization. Note that Eq. 5 is related to the χ-square test.

The quantity x_i measures how much the point estimate deviates from the truth compared to the overall extension and orientation of the posterior constraints. The first filter we apply to our family of candidate priors demands all x_i to be smaller than a maximum threshold, which we denote x^*. This means that we calculate x_i for each realization of the nuisance and we then determine the maximum:

$$x = \max_i \{x_i\} \qquad (6)$$

of those Mahalanobis distances. All candidate priors that do not meet the requirement

$$x \leq x^* \qquad (7)$$

will be excluded. In other words, we demand that the prior results in less than x^* bias for any realization of the nuisance with respect to the target parameter uncertainties. That leaves us with an equal or smaller number of candidate priors, depending on the considered threshold quantity x^*. For example, in the DES Y3 2-point function analyses, this threshold was chosen as $0.3\sigma_{2D}$ (with the subscript "2D" standing for the 68% credible interval) for the joint, marginalized posterior of S_8 and Ω_m (Krause et al. 2021).

2.3.2. Criterion 2: Minimum frequentist coverage

We now apply a second filter to the candidate priors: a minimum coverage probability for the 1-σ confidence region of the marginalized posterior of the target parameters (i.e. the smallest sub-volume of target parameter space that still encloses about 68% of the posterior). This coverage probability measures how often the true target parameters would be found within that confidence region in independent repeated trials. In a Bayesian analysis, it can in general not be expected that this coverage matches the confidence level of the considered region (in our case 68%). This has e.g. been extensively discussed in Percival et al. (2021). The presence of projection effects does impact this additionally when studying the marginalized posteriors.

To measure the coverage probability, we proceed as follows for a given prior: For each of the n_{noise} noise realizations per data vector $\hat{\boldsymbol{\xi}}^i$ described in Sec. 2.2, we determine the 1-σ confidence region (e.g., the contour in the target parameter space or the 1D confidence limit of each parameter). The fraction of these confidence regions that enclose the true target parameters is an estimate for the coverage probability. We will denote with y_i the fraction corresponding to the i-th nuisance realization. In a spirit similar to our previous filter we consider the minimum coverage per nuisance realization,

$$y = \min_i \{y_i\} , \qquad (8)$$

and then require for each nuisance realizations to have a coverage of at least y^*,

$$y \geq y^*, \qquad (9)$$

in order for a candidate nuisance prior to pass our coverage criterion. In other words, we will demand that a prior result in a coverage of at least y^*.

It is worth mentioning that the value of y^* is not necessarily the probability associated to a confidence level of 1-σ (i.e. $\sim 68\%$) because the coverage fraction will scatter around that value following a binomial distribution. As an example, for 100 independent noisy realizations, the probability of having z out of the 100 realizations covering the true parameters will be $P(z) = \text{Bin}(z; n = 100, p = 0.68)$. Therefore, ensuring a minimum coverage such that the cumulative binomial distribution at that value is statistically meaningful should be enough. Using our same previous example, a coverage of 62 or more out of 100 realizations will happen with 90% probability. The larger the number of nuisance realizations used to derive an ITEM prior, the smaller the coverage threshold should be chosen to avoid false positives of low coverage. The smaller the number of nuisance realizations used to derive an ITEM prior, the higher the coverage threshold should be chosen to avoid false positives of low coverage.

In the presence of a bias on the target parameters, however, there will be a mismatch between the coverage probability and the confidence level of the posterior. If an otherwise accurate multivariate Gaussian posterior is not, on average, centered on the true parameter values, any of its confidence regions will have a lower coverage than nominally expected. To account for this effect in the coverage test without double-penalizing bias, and later to find the prior that minimizes posterior volume in a total error sense, we propose to appropriately raise the confidence level to be used. For our implementation of this, we will make the assumption of a Gaussian posterior. For an unbiased posterior, this means that among many realizations of the noisy data vector, the log-posterior value of the true parameters is a χ^2 distributed

random variable with the number of degrees of freedom equal to the number of target parameters. In the case of a biased analysis, the log-posterior of the true parameters instead follows a non-central χ^2 distribution.

To calculate the required confidence level, we proceed as follows:

- calculate for which value the cumulative distribution function of a non-central χ^2-distribution representing the biased marginalized posterior takes a value of 68%,
- at that value, evaluate the cumulative distribution function of a central χ^2-distribution (with the same number of degrees of freedom). This will always be higher than 68%.

It is this higher confidence level that we will use to calculate the enlarged, "total error" 1-σ confidence region for the noisy realizations of our data vectors ξ_i and to measure the coverage probabilities.

The χ^2-distribution with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. Its cumulative distribution function (CDF) is:

$$\mathrm{CDF}(\chi^2; k) = \frac{\gamma(\frac{k}{2}, \frac{\chi^2}{2})}{\Gamma(\frac{k}{2})}, \tag{10}$$

where γ is the lower incomplete gamma function and Γ is the gamma function. In Fig. 4, we plot with a continuous line the CDF of a χ^2-distribution with two degrees of freedom ($k = 2$) for visualization purposes.

The non-central χ^2-distribution with k degrees of freedom is an extension of the central χ^2-distribution and it has a non-centrality parameter λ_i defined as:

$$\lambda_i = (\boldsymbol{\theta}_T - \hat{\boldsymbol{\theta}}_T^i)^\top \mathbf{C}_i^{-1} (\boldsymbol{\theta}_T - \hat{\boldsymbol{\theta}}_T^i), \tag{11}$$

where $\boldsymbol{\theta}_T$ is the true target parameter vector, $\hat{\boldsymbol{\theta}}_T^i$ is the target parameter estimate for nuisance realization i, and $\mathbf{C}_i^{-1}$ is the associated inverse covariance matrix. We point out that $\lambda_i = x_i^2$ from Eq. 5. In Fig. 4 we show the CDF of two non-central χ^2-distributions with two degrees of freedom, but different noncentrality parameters. For $\mathrm{CDF}(\chi^2; k = 2, \lambda = 1, 2) = 0.68$, it is clear that $\mathrm{CDF}(\chi^2; k = 2) > 0.68$. This augmented cumulative probability accounts for the overall bias, including the one induced by projection effects.

2.3.3. Criterion 3: Minimizing the posterior volume

The candidate priors that have passed our previous two filters follow our definitions of *unbiasedness*, *reliability*, and *robustness*. The ITEM prior corresponds to the one among them that yields the smallest posterior uncertainty on the target parameters. This final selection step is necessary because it is in fact easier to meet our previous criteria with wide posteriors that result from unconstrained nuisance priors.

We quantify the target parameter random error by approximating the volume that the 1-σ confidence contours (68%) enclose. For a Gaussian posterior of any dimensionality, the volume would be given in terms of the parameter covariance matrix $\mathbf{C}_i$ as (see Huterer & Turner 2001 for a derivation):

$$V_i = \pi \det(\mathbf{C}_i)^{1/2}, \tag{12}$$

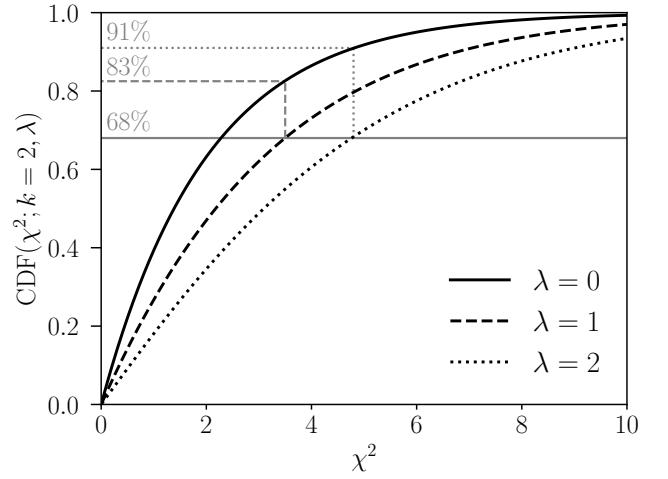and for simplicity we will use that formula also for our (potentially non-Gaussian) posteriors.

Fig. 4: Cumulative distribution functions of two degrees of freedom central and non-central χ^2-distributions. The non-central χ^2-distribution plotted shows that for the same significance threshold, the central χ^2-distribution will always have a larger CDF.

Let m' be the number of candidate priors that passed the filters listed in Sec. 2.3.1 and 2.3.2. Let $V_{i,j}$ be the volume enclosed by the 1-σ ellipsoid of the i-th posterior distributions at the j-th prior, $j = 1, ..., m'$. Then, we propose to consider the maximum of these volumes across all data vectors:

$$V_j^{\max} = \max_i (V_{i,j}). \tag{13}$$

Similarly as done in Sec. 2.3.2, we account for the bias when optimizing the posterior uncertainty by considering the volume enclosed by the extended central χ^2-distribution. This allows us to minimize the total error, including random and systematic contributions. The re-scaled volume includes a factor equivalent to (see App. A):

$$\tilde{V}_{i,j} \propto V_{i,j} \cdot \mathrm{CDF}^{-1}(0.68; k = N_T, \lambda_i)^{N_T/2}. \tag{14}$$

We choose the prior that has the minimum volume $\tilde{V}_{\mathrm{ITEM}} = \min_j (\tilde{V}_j^{\max})$ as the ITEM prior.

3. ITEM Priors For Density Split Statistics

In this section, we examine the performance of the ITEM prior when analyzing a higher-order statistic of the matter density field called Density Split Statistics. We do this as a proof of concept under simplified assumptions, particularly on the nuisance realization deemed plausible. Readers are referred to Britt et al. 2024 for more careful analysis of potential nuisance effects relevant for those statistics.

3.1. Density Split Statistics

We test our ITEM prior methodology in considering as the relevant nuisance effect the galaxy-matter connection models employed for the Density Split Statistics (DSS, Friedrich et al. 2018; Gruen et al. 2018 for the Dark Energy Survey Year 1, DES Y1, analysis). The data vector of DSS is a compressed version of the joint PDF of projected matter density and galaxy count. These studies consider two ways of generalizing a linear,

Table 1: Parameters used in the simulated likelihood analyses for the α model. From left to right, we list the fiducial values and prior ranges (using $\mathcal{U}$ to denote a uniform prior) used to simulate the synthetic data for the two cases of stochasticity. The nuisance realizations and the priors are chosen to be either the same or the derived results from Friedrich et al. 2018; Gruen et al. 2018. In our work, the cosmological parameters would be the target ones, while the stochasticity parameters would be the nuisance ones. Ω_m, S_8 and b have fixed priors and unchanged values for the parameter vectors.

	Parameter Vector 1 *Non-Stochasticity*	Parameter Vector 2 *Buzzard Stochasticity*	DSS Original Prior Distribution
Target Parameters			
Cosmological Parameters			
Ω_m	0.286	0.286	$\mathcal{U}[0.1,0.9]$
σ_8	0.820	0.820	$\mathcal{U}[0.2,1.6]$
Nuisance Parameters			
Tracer Galaxies			
b	1.54	1.54	$\mathcal{U}[0.8,2.5]$
Stochasticity			
α_0	1.00	1.26	$\mathcal{U}[0.1,3.0]$
α_1	0.00	0.29	$\mathcal{U}[-1.0,4.0]$

deterministic galaxy-matter connection, which is their most relevant nuisance effect. The first one accounts for an extension of the Poissonian shot-noise in the distribution of galaxy counts by adding two free parameters: α_0 and α_1. For simplicity, we will call this the α-model. The second one more simply parametrizes galaxy stochasticity via one correlation coefficient r. We will refer to this as the r-model. For a derivation and explanation of these models, see App. B.

3.2. Simulated likelihood analysis with wide priors

In this subsection, we present results of simulated likelihood analyses of DSS with the wide prior distributions used in the original DES Y1 analysis (Friedrich et al. 2018; Gruen et al. 2018).

Our study focuses on the five quantities listed in Table 1. The cosmological parameters are Ω_m[10] and σ_8[11]. The parameters describing the galaxy-matter connection are the linear bias b[12], and the stochasticity parameters α_0 and α_1. These two last parameters were introduced in the DSS to accurately model the galaxy-matter connection. However, the study did not introduce a physically motivated prior distribution of these parameters. For a more detailed explanation on the α parameters, including their physical meaning and origin, we refer to Friedrich et al. (2022).

As a derived parameter, we also consider S_8[13]:

$$S_8 = \sigma_8 \sqrt{\Omega_m/0.3} \,. \tag{15}$$

Both Ω_m and S_8 are the target parameters (as presented in, e.g. Krause et al. 2021), while α_0 and α_1 are the nuisance

parameters whose priors we are optimizing. The prior on b remains unchanged, since we find its joint posterior with cosmological parameters to be symmetric and well constrained, at least in the case without modeling the excess skewness of the matter density distribution ΔS_3 as an additional free parameter, which is the only case we consider here.

The first step of the ITEM prior method is the use of different realizations of the nuisance. We limit our analysis by considering two realizations of the stochasticity listed in the second and third columns of Table 1. The first one corresponds to a configuration in which there is no stochasticity in the distribution of galaxies, i.e. where the shot-noise of galaxies follows a Poisson distribution, which is often assumed to be correct at sufficiently large scales (but which has been shown to fail at exactly those scales, see e.g. Friedrich et al. 2018; MacCrann et al. 2018). That is equivalent to setting $\alpha_0 = 1.0$ and $\alpha_1 = 0.0$, since these two values turn Eq. B.5 into a Poisson distribution. We call this the *Non-stochasticity* case. The second one corresponds to the best-fit shot-noise parameters found by Friedrich et al. (2018) on REDMAGIC mock catalogs (Rozo et al. 2016), constructed given realistic DES Y1-like survey simulations called Buzzard-v1.1 (DeRose et al. 2019). These Buzzard mock galaxy catalogs (Wechsler et al. 2022) have been used extensively in DES analyses (DeRose et al. 2022). The values for the stochasticity parameters are $\alpha_0 = 1.26$ and $\alpha_1 = 0.29$. We call this the *Buzzard stochasticity* case.

These two realizations should be considered as a simple test case for the methodology. This work, as mentioned previously, emphasizes that the resulting ITEM prior is limited because of the few, and potentially unrealistic nuisance realizations considered. There is a wide range of plausible stochasticity configurations that could be taken into account (see e.g. the findings of Friedrich et al. 2022). In a companion paper (Britt et al. 2024) we consider diverse realizations of halo occupation distributions (HODs) to derive such a set of stochasticity configurations.

For our simulated likelihood analysis, we obtain noiseless data vectors for each nuisance realization from the model prediction (Friedrich et al. 2018) for the Buzzard-measured values of the stochasticity parameters. Additionally, we used the priors presented in the last column of Table 1, which are the ones used originally in the DES Y1 analysis of DSS (Gruen

---

[10] The present-day energy density of matter in units of the critical energy density.
[11] The present-day linear root-mean-square amplitude of relative matter density fluctuations in spheres of radius 8 h^{-1} Mpc.
[12] The square root of the ratio of the galaxy and matter auto-correlation functions on large scales.
[13] The parameter S_8 has been widely used because it removes degeneracies between Ω_m and σ_8 present in analysis of cosmic structure. Some analyses have found there to be a tension between the S_8 measured in the late universe (Aghanim et al. 2020, Heymans et al. 2021, Lemos et al. 2021 and Dalal et al. 2023) and from the CMB (Raveri & Hu 2019, Asgari et al. 2021 and Anchordoqui et al. 2021), rendering it of particular interest to have an interpretable posterior for.
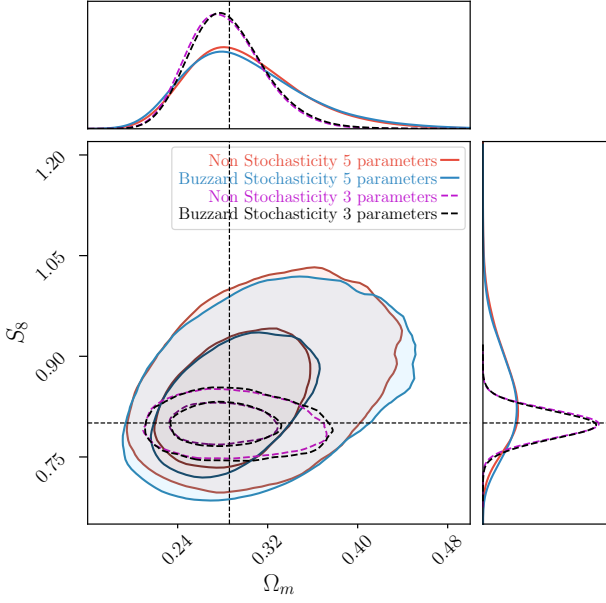
Fig. 5: Marginalized (Ω_m, S_8) posteriors with 1-σ and 2-σ contours from simulated DSS α model of synthetic, noiseless data vectors using the fiducial values and priors from Table 1. The red and blue contours show the case with 5 degrees of freedom, while the purple and black contours show the case with 3 degrees of freedom, with $\alpha_0 = [1.0, 1.26]$ and $\alpha_1 = [0.0, 0.29]$ respectively.

et al. 2018). To obtain the corresponding posterior, we run MCMCs using the EMCEE algorithm (Foreman-Mackey et al. 2013), leaving the stochasticity parameters free with the original broad priors from the last column in Table 1. We point to Abbott et al. (2023) for a discussion of the use of posterior samplers. We also run two additional MCMCs fixing the stochasticity parameters to the values listed in Table 1 for comparison purposes. In Fig. 5, we show the four estimated marginalized posteriors of Ω_m and S_8. We find that in the cases in which we considered free stochasticity parameters, the posterior returns much larger uncertainties over both cosmological parameters, with a somewhat shifted mean.

Subsequently, we fit a 2D Gaussian distribution centered at the 2D marginalized MAP and with the chains' parameter covariance matrix, and evaluate the off-centering of the chains from the true cosmological parameters the model was evaluated for when generating the data vector. Fig. 6 illustrates the biases for each chain obtained with the original wide priors on stochasticity parameters. Both cases are biased: the bias of the Buzzard stochasticity case (in blue) reaches a value of $0.42\sigma_{2D}$, more than the defined DES threshold of $0.3\sigma_{2D}$. The bias, in both cases, is systematically pointing towards lower and higher values on Ω_m and S_8, respectively.

In an additional case study, we assume an alternative parametrization of galaxy stochasticity (the r model, introduced in App. B) as a basis, yet still use the data vectors obtained previously with the α model. The r model is simpler than the α model because it only considers one correlation parameter r for the effect of stochasticity. This test checks whether the conditions demanded by the ITEM formalism can be met even when a model does not produce the data, which happens often with real data, and how much information the ITEM prior can recover in such a case.
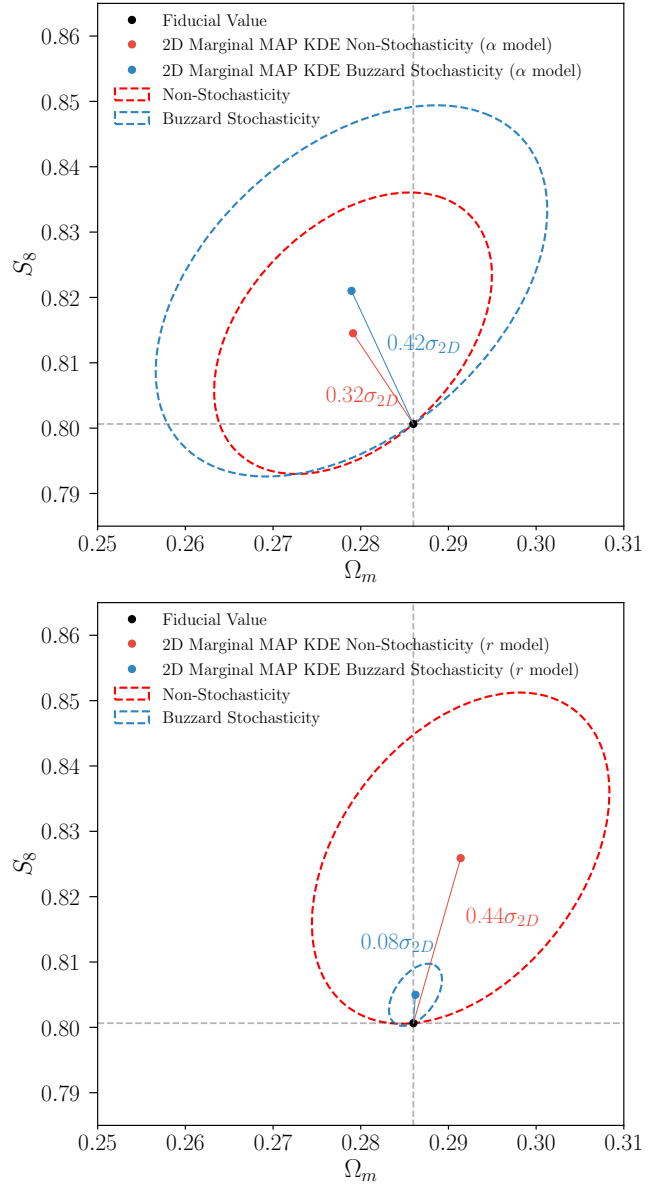

Fig. 6: Parameter biases in simulated likelihood analyses for the α model (top) and the r model (bottom): in red and blue, we show the ellipses for the 2D marginalized constraints of the non-stochasticity and the Buzzard stochasticity configurations, respectively. These figures follow Fig. 4 from Krause et al. (2021). For the α model, these are centered on their corresponding 2D marginal MAP. Because of prior volume effects, the marginalized constraints are not centered on the input cosmology. For the r model, there is a systematic bias in the non-stochasticity case that is explained by the sharp upper bound from the prior on r.

The original DSS prior used on r follows a uniform distribution: $\Pi(r) \sim \mathcal{U}[0.0, 1.0]$, where $r = 1$ recovers the non-stochasticity case (see App. B). The sharp upper bound results in a projection effect as seen in Fig. 7, with posteriors biased towards large values of S_8, particularly in the non-stochastic case. We also illustrate the bias for both chains in the bottom panel of Fig. 6 with the original wide prior on r.
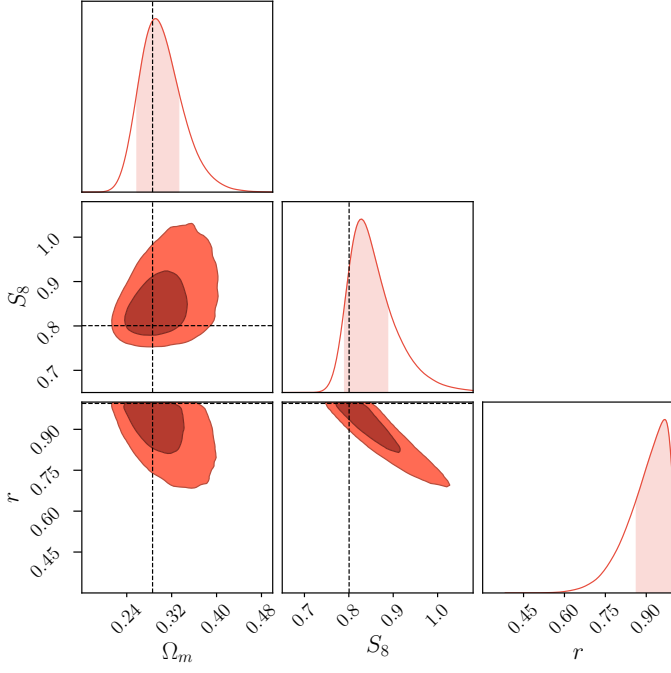
Fig. 7: Marginalized posteriors (Ω_m, S_8, r) from the simulated DSS r model using a non-stochastic parameter vector from the α model. The fiducial values are represented by the dashed lines, hence for $r_{\text{fiducial}} = \max[r] = 1.0$, it is in the right edge of the plot. Because $r > 1$ is unphysical, projection effects cause a systematic offset between the mean or peak of the 1D marginalized posteriors, particularly in r and S_8, and their input values.

3.3. ITEM priors on the Density Split Statistics

In this subsection, we detail the procedure to obtain ITEM priors for two nuisance models: first in Sec. 3.3.1, a simple approach considering the α model as a baseline, and second in Sec. 3.3.2, a case when the data vectors are simulated with the α model, but analyzed with the r model, leading to a systematic error.

3.3.1. ITEM prior for the α model

In the original DSS work, the nuisance parameters α_0 and α_1 had uniform priors. We make the same choice here and only vary the limits of these priors.[14]

We start by sampling the candidate priors as done in the example at Sec. 2.2: we shrink the bounds of the priors ($\min[\alpha_0]$, $\max[\alpha_0]$, $\min[\alpha_1]$ and $\max[\alpha_1]$) with a step of 0.1 units per cut until these reach the closest rounded fiducial value of the parameter vector:

$$\min[\alpha_0] \in \{0.1, 0.2, ..., 0.9\}$$
$$\max[\alpha_0] \in \{1.4, 1.5, ..., 3.0\}$$

$$\min[\alpha_1] \in \{-1.0, -0.9, ..., -0.1\}$$
$$\max[\alpha_1] \in \{0.3, 0.4, ..., 4.0\} \ .$$

The final number of candidate priors corresponds to 58,140 when combining each different bound. Next, we apply an importance sampling to the original 5D wide posteriors

---

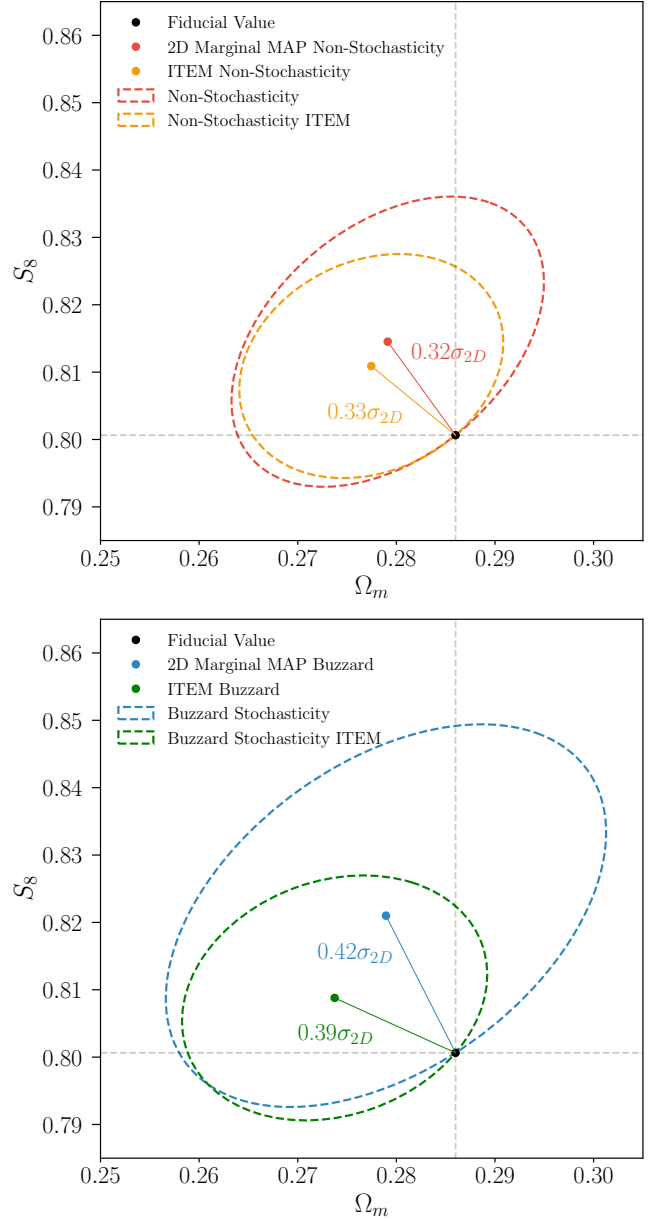[14] We note that one could also vary, e.g., the mean and width of a Gaussian prior.

Fig. 8: Parameter biases of the data vectors with the original and ITEM priors: the red and blue ellipses show wide contours (same from Fig. 6), while the yellow and green ones obtained after enforcing the $0.4\sigma_{2D}$ bias threshold and 1D coverage threshold for the 2D marginalized constraints. Both are centered in their respective 2D marginalized MAP. Due to parameter volume effects, the marginalized constraints from the baseline prior analysis are not centered on the input cosmology. *Top:* Simulated likelihood analysis for the data vector without shot-noise ($\alpha_0 = 1.0$, $\alpha_1 = 0.0$). *Bottom:* Simulated likelihood analysis for the data vector with Buzzard stochasticity ($\alpha_0 = 1.26$, $\alpha_1 = 0.29$). The dashed horizontal and vertical lines indicate the fiducial parameter values.

described in Sec. 3.2 and follow the three steps described in Sec. 2 to obtain the ITEM priors:

First, we estimate, for each candidate prior, the maximum Mahalanobis distance x from Eq. 6 for the different posteriors. We set the upper threshold to $0.4\sigma_{2D}$ for the first filter of the ITEM priors. A total of 1,980 priors had a bias of less than

Table 2: The original and ITEM priors ranges for the α model with a maximum bias of $0.4\sigma_{2D}$ and a minimum coverage of 56% for both Ω_m and S_8. The bias and error coverage quantities from both nuisance realizations are illustrated for each prior as *non-stochasticity/Buzzard stochasticity* cases respectively. In the final three columns, we include the average error of the final estimate for Ω_m and S_8 from both nuisance realizations and the 2D approximated, normalized and re-scaled volume of the 2D marginalized posterior $\tilde{V}$ defined in Sec. 2.3.3.

Parameter	Prior	Prior Width	Bias [σ_{2D}]	Ω_m Coverage	S_8 Coverage	$\overline{\sigma}_{\Omega_m}$	$\overline{\sigma}_{S_8}$	$\tilde{V}/\tilde{V}_{\text{Original}}$
Original Priors								
α_0	$\mathcal{U}[0.1, 3.0]$	2.9						
			0.32/0.42	76%/76%	72%/52%	0.052	0.068	1.00
α_1	$\mathcal{U}[-1.0, 4.0]$	5.0						
ITEM Prior								
α_0	$\mathcal{U}[0.6, 2.0]$	1.4						
			0.33/0.39	80%/64%	68%/56%	0.040	0.049	0.58
α_1	$\mathcal{U}[-1.0, 1.1]$	2.1						

Fig. 9: The marginalized posterior distributions for the chains from the α and r models from DSS. These have been plotted with the widened 1-σ and 2-σ confidence levels to account for the bias as presented in Sec. 2.3.2. In the top and bottom panels, we illustrate the non-stochasticity and the Buzzard stochasticity respectively. The larger contours correspond to the posterior simulated using the original DSS α model priors. The narrow continuous contours illustrate the one with the ITEM prior on the α model, and the discontinuous contours for the r model.

$0.4\sigma_{2D}$ (~3% of the original set, indicating that prior volume related biases are prevalent in this inference problem). On the upper panel from Fig. C.2, we show their distribution on the prior space. These mainly preferred lower values for the upper prior bound on α_1, indicating that the original asymmetric prior caused a prior volume related bias.

Second, we run 50 noisy simulated realizations, 25 for each nuisance realization, following the criterion from Sec. 2.3.2. These noisy realizations were simulated by adding multivariate Gaussian noise with the covariance matrix from Gruen et al. (2018) to the noiseless data vectors. We choose the threshold $z = 14$ such that the cumulative binomial distribution for independent 25 realizations is at least 90% (this corresponds to a $14/25 = 56\%$ coverage). Only a small fraction of priors should thus be excluded due to chance fluctuations in the number

of contours that contain the input parameter values. In this step, we considered the systematic bias as explained in Sec. 2.3.2. A total of 1,868 out of the 1,980 remaining candidate priors qualified under the 1D coverage requirement for both nuisance realizations. In Fig. C.3 we show the distribution of the coverages among the tested priors.

The final step corresponds to the minimization of the error from Sec. 2.3.3. The resulting ITEM priors for α_0 and α_1 are listed in the second row of Table 2. We first note that the prior widths were reduced to ~48% and 42% their original ones respectively. In particular, the prior in α_0 remained centered around 1.3, while all upper bounds such that $\max[\alpha_1] > 1.1$ were rejected. The biases obtained with the original and ITEM priors are displayed in Fig. 8. While the bias in Ω_m slightly increased, the one on S_8 decreased as expected from the imposed

threshold. The coverages for both nuisance realizations did not change more than ±12% with respect to the coverages from the original priors. In the last three columns of Table 2, we present the average standard deviations and the statistical uncertainty of the target parameters. These were reduced, corresponding to an improvement that is formally equivalent to an experiment surveying more than three times the volume. Finally, the marginalized distributions are shown in Fig. 9.

Note that in a proper implementation of the ITEM prior program one should consider an exhaustive list of possible nuisance realizations, instead of just the two realizations considered in our example. In the companion study Britt et al. 2024, we include shot-noise realizations generated from a range of plausible HOD descriptions to obtain fiducial stochasticity parameters and derive a more realistic result than the one presented here. The nuisance realizations derived there will be used in a more thorough application of the ITEM prior method in the future.

Due to recent discussions on the impact of using different point estimators in cosmological analyses (e.g. Amon et al. 2022), in App. C we explore the ITEM priors obtained when, instead of using the marginalized MAP, we use the MAP and means as point estimators.

3.3.2. ITEM prior for r model with a different basis model

In this subsection, we investigate how the ITEM priors can help in cases where the assumed nuisance models are not the generators of the data. We use the same two-parameter vectors from Sec. 3.3.1 (the non-stochasticity and the Buzzard stochasticity) and follow the same procedure. However, we use the r model presented in App. B.2 to obtain posteriors. The parameters considered are Ω_m, σ_8, b and r. We use the same prior distribution over Ω_m, σ_8 and b, and change the prior on r. We then obtain two posteriors using the non-stochasticity and the Buzzard realization of the nuisance in the simulated data. Next, we sample different priors for r: we changed the lower bound min[r] and kept max[r] fixed to 1.0:

$$\min[r] \in \{0.00, 0.01, ..., 0.99\} .$$

The total number of priors is 99. Then, we repeat the three steps of the ITEM prior pipeline.

In this case there is a larger systematic bias due to the sharp upper bound on r and the nuisance realization from a different model. In Fig. 10 we show the resulting bias for different lower bounds on min[r]. Because of the systematic error present in the data vector, there is a clear trade-off. On one side, for the non-stochasticity case (red line in Fig. 10), priors with a higher lower bound on r (from min[r] = 0.9 and above) will be substantially less biased. This is expected as priors with higher min[r] will exclusively select values close to $r = 1$, near the fiducial values of the parameter vector (we recall that $\alpha_0 = 1.0$, $\alpha_1 = 0.0$ is algebraically equivalent to $r = 1$ as shown in App. B). On the other side, the Buzzard stochasticity case (blue line in Fig. 10) becomes highly biased for min[r] > 0.9. Those priors on r restrict the parameter space substantially to a "more" Poissonian configuration of the galaxy matter connection (see App. B), increasing their bias drastically. To account for the systematic bias, we choose a flexible bias threshold of $0.5\sigma_{2D}$, which reduces the number of candidate priors to 68.

Subsequently, and as we did in Sec. 3.3.1, we run 50 noisy simulated realizations, 25 for each nuisance realization using the same data vectors. We require the noisy realizations to have
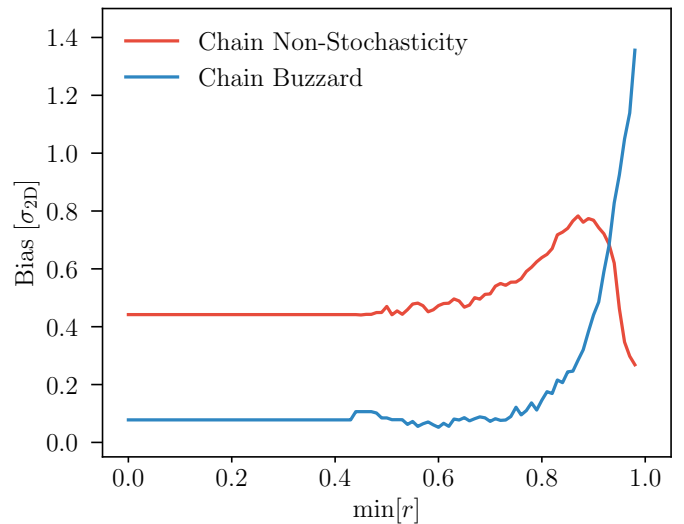
Fig. 10: Biases of the non-stochasticity and Buzzard chains for different lower bounds of $\Pi(r)$. After a min[r] ∼ 0.9, we see a trade-off between the two stochasticity cases.

at least a 56% 1D coverage for both Ω_m and S_8 out of the 25 simulated realizations. This second filter did not reduce the number of candidate priors because all of them fulfilled the requirement.

Finally, we determine the ITEM prior by optimizing the total error from Sec. 2.3.3. In Table 3 we summarize the resulting prior, with the corresponding biases, coverage, and uncertainties. We find similar statistics but using a prior substantially less wide (reduction of 68%). For comparison purposes, in Fig. 9 we show the marginalized posterior distributions.

4. Discussion and Conclusion

In the era of precision cosmology, observations require increasingly complex models that account for astrophysical and observational systematic effects. Prior distributions are assigned to the nuisance parameters modeling these phenomena based on either limits in which they are plausible, ranges of simulation results, or actual calibration uncertainty. However, often the analyses using such priors yield biased or overly uncertain marginalized constraints on cosmological parameters due to so-called projection or prior volume effects. In this paper, we have presented a procedure to overcome projection effects through a pragmatic choice of nuisance parameter priors based on a series of simulated likelihood analyses. We named this the Informed Total-Error-Minimizing Priors (ITEM) method.

The ITEM approach performs these simulated likelihood tests in order to find a prior that, when applied, leads to a posterior with some desired statistical properties and interpretability. We divide the parameters of a model into two types: the target parameters (the ones of interest) and the nuisance parameters (the ones to optimize the prior for). Then, our approach finds an equilibrium between the under-fitting and overly free treatment of nuisance effects. This is done by enforcing three requirements on the target parameter posteriors found when applying the nuisance parameter prior in question in the analysis:

– the biases on the target parameters need to be lower than a certain fraction of their statistical uncertainties;

Table 3: The original and ITEM priors range for the r model with a maximum bias of $0.5\sigma_{2D}$ and a minimum coverage of 56% for both Ω_m and S_8. As in Table 2, the bias and coverage quantities from both nuisance realizations are separated with a backslash.

Parameter	Prior	Prior Width	Bias $[\sigma_{2D}]$	Ω_m Coverage	S_8 Coverage	$\overline{\sigma}_{\Omega_m}$	$\overline{\sigma}_{S_8}$	$\tilde{V}/\tilde{V}_{\text{Original}}$
Original Prior								
r	$\mathcal{U}[0.00, 1.00]$	1.00	0.44/0.08	84%/76%	56%/56%	0.039	0.059	1.00
ITEM Prior								
r	$\mathcal{U}[0.68, 1.00]$	0.32	0.50/0.09	84%/76%	56%/56%	0.037	0.053	0.91

- in the presence of noise, we need to still cover the original fiducial parameter values a certain fraction of times in some desired confidence region;
- these criteria both being met, we prefer nuisance parameter priors that lead to small target parameter posterior volumes.

We test the ITEM prior pipeline using two different nuisance realizations from the density split statistics analysis framework (Friedrich et al. 2018; Gruen et al. 2018). We simulate two realistic data vectors, where one does not include stochasticity between galaxy counts and the underlying matter density contrast, whereas the other considers a specific stochasticity found in an N-body simulation based mock catalog. For both cases, we apply the above criteria for limiting bias, ensuring coverage, and minimizing uncertainties, to a list of candidate priors. The result from applying the ITEM prior on the constraining power on target parameters corresponds to an improvement equivalent to an experiment surveying more than three times the cosmic volume. Additionally, we study the case in which the model used to treat the nuisance effect is not the same as the one used to generate the data, a common situation in practice. The range of the resulting ITEM prior is less than half of the original prior's size, resulting in a smaller total uncertainty of the target parameters with a similar level of bias and coverage as obtained with the original prior.

One of the main limitations of this method is that its results may depend on the fiducial values we assign to the parameter of interest, and formally requires a complete set of plausible realizations of the nuisance effect. In this first study, we fixed the target parameters and chose just two nuisance realizations, however, one can extend our method to explore a variety of target and nuisance realizations in future works that apply it to actual inference.

To summarize, in this work we aimed to explore an alternative approach to assigning prior distributions to nuisance parameters. The method presented is flexible, as it allows us to define requirements and propose priors, given some set of plausible nuisance realizations, for a reliable and information-maximizing statistical inference.

Data Availability

The Dark Energy Survey Year 1 data (Density Split Statistics) used in this article is publicly available at https://des.ncsa.illinois.edu/releases/y1a1/density.

References

Abbott, T., Aguena, M., Alarcon, A., et al. 2022, Physical Review D, 105
Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2023, DES Y3 + KiDS-1000: Consistent cosmology combining cosmic shear surveys
Aghanim, N., Akrami, Y., Ashdown, M., et al. 2020, Astronomy & Astrophysics, 641, A6
Aihara, H., Arimoto, N., Armstrong, R., et al. 2017, Publications of the Astronomical Society of Japan, 70
Amon, A., Gruen, D., Troxel, M., et al. 2022, Physical Review D, 105
Anchordoqui, L. A., Di Valentino, E., Pan, S., & Yang, W. 2021, Journal of High Energy Astrophysics, 32, 28–64
Asgari, M., Lin, C.-A., Joachimi, B., et al. 2021, Astronomy & Astrophysics, 645, A104
Baldauf, T., Seljak, U., Senatore, L., & Zaldarriaga, M. 2011, Journal of Cosmology and Astroparticle Physics, 2011, 031–031
Baleato Lizancos, A. & White, M. 2023, Journal of Cosmology and Astroparticle Physics, 2023, 044
Bernstein, G. M. 2010, Monthly Notices of the Royal Astronomical Society, 406, 2793–2804
Blazek, J. A., MacCrann, N., Troxel, M., & Fang, X. 2019, Physical Review D, 100
Britt, D., Gruen, D., Friedrich, O., Yuan, S., & Guachalla, B. R. 2024, Bounds on galaxy stochasticity from halo occupation distribution modeling
Campos, A., Samuroff, S., & Mandelbaum, R. 2023, Monthly Notices of the Royal Astronomical Society, 525, 1885
Cawthon, R., Elvin-Poole, J., Porredon, A., et al. 2022, Monthly Notices of the Royal Astronomical Society, 513, 5517
Chintalapati, P., Gutierrez, G., & Wang, M. 2022, Physical Review D, 105
Chisari, N. E., Richardson, M. L. A., Devriendt, J., et al. 2018, MNRAS, 480, 3962
Collaboration, D., Aghamousa, A., Aguilar, J., et al. 2016, The DESI Experiment Part I: Science,Targeting, and Survey Design
Collaboration:, D. E. S., Abbott, T., Abdalla, F. B., et al. 2016, Monthly Notices of the Royal Astronomical Society, 460, 1270
Cordero, J. P., Harrison, I., Rollins, R. P., et al. 2022, Monthly Notices of the Royal Astronomical Society, 511, 2170
Cui, W., Borgani, S., & Murante, G. 2014, Monthly Notices of the Royal Astronomical Society, 441, 1769–1782
Dalal, R., Li, X., Nicola, A., et al. 2023, Phys. Rev. D, 108, 123519
de Jong, R. S., Bellido-Tirado, O., Chiappini, C., et al. 2012, Ground-based and Airborne Instrumentation for Astronomy IV
Dekel, A. & Lahav, O. 1999, The Astrophysical Journal, 520, 24–34

DeRose, J., Wechsler, R. H., Becker, M. R., et al. 2019, The Buzzard Flock: Dark Energy Survey Synthetic Sky Catalogs

DeRose, J., Wechsler, R. H., Becker, M. R., et al. 2022, Phys. Rev. D, 105, 123520

Ding, Z., Seo, H.-J., Vlah, Z., et al. 2018, Monthly Notices of the Royal Astronomical Society

Eifler, T., Miyatake, H., Krause, E., et al. 2021, MNRAS, 507, 1746

Etherington, T. R. 2019, PeerJ, 7, e6678

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publications of the Astronomical Society of the Pacific, 125, 306

Friedrich, O., Gruen, D., DeRose, J., et al. 2018, Phys. Rev. D, 98, 023508

Friedrich, O., Halder, A., Boyle, A., et al. 2022, Monthly Notices of the Royal Astronomical Society, 510, 5069–5087

Georgiou, C., Hoekstra, H., Kuijken, K., et al. 2021, Astronomy & Astrophysics, 647, A185

Gruen, D., Friedrich, O., Krause, E., et al. 2018, Phys. Rev. D, 98, 023507

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357

Hesterberg, T. C. 2003, Advances in Importance Sampling

Heymans, C., Tröster, T., Asgari, M., et al. 2021, Astronomy & Astrophysics, 646, A140

Hildebrandt, H., van den Busch, J. L., Wright, A. H., et al. 2021, Astronomy & Astrophysics, 647, A124

Hinton, S. R. 2016, The Journal of Open Source Software, 1, 00045

Hirata, C. M. & Seljak, U. c. v. 2004, Phys. Rev. D, 70, 063526

Huang, H.-J., Eifler, T., Mandelbaum, R., & Dodelson, S. 2019, Monthly Notices of the Royal Astronomical Society, 488, 1652

Huang, S., Leauthaud, A., Murata, R., et al. 2017, Publications of the Astronomical Society of Japan, 70

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90

Huterer, D. & Turner, M. S. 2001, Physical Review D, 64

Ishikawa, S., Okumura, T., Oguri, M., & Lin, S.-C. 2021, The Astrophysical Journal, 922, 23

Ivezić, v., Kahn, S. M., Tyson, J. A., et al. 2019, The Astrophysical Journal, 873, 111

Joachimi, B., Lin, C.-A., Asgari, M., et al. 2021, Astronomy &; Astrophysics, 646, A129

Kiessling, A., Cacciato, M., Joachimi, B., et al. 2015, Space Science Reviews, 193, 67–136

Krause, E., Fang, X., Pandey, S., et al. 2021, Dark Energy Survey Year 3 Results: Multi-Probe Modeling Strategy and Validation

Lamman, C., Eisenstein, D., Aguilar, J. N., et al. 2023, Monthly Notices of the Royal Astronomical Society, 522, 117

Lamman, C., Tsaprazi, E., Shi, J., et al. 2024, The Open Journal of Astrophysics, 7

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, Euclid Definition Study Report

Lemos, P., Raveri, M., Campos, A., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 6179–6194

Li, X., Miyatake, H., Luo, W., et al. 2022, Publications of the Astronomical Society of Japan, 74, 421–459

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints [arXiv:0912.0201]

MacCrann, N., Becker, M. R., McCullough, J., et al. 2021, Monthly Notices of the Royal Astronomical Society, 509, 3371–3394

MacCrann, N., DeRose, J., Wechsler, R. H., et al. 2018, Monthly Notices of the Royal Astronomical Society, 480, 4614–4635

Mahalanobis, P. C. 1936, Proceedings of the National Institute of Sciences (Calcutta), 2, 49

Mandelbaum, R., Blazek, J., Chisari, N. E., et al. 2019, BAAS, 51, 363

Mummery, B. O., McCarthy, I. G., Bird, S., & Schaye, J. 2017, Monthly Notices of the Royal Astronomical Society, 471, 227–242

Myles, J., Alarcon, A., Amon, A., et al. 2021, Monthly Notices of the Royal Astronomical Society, 505, 4249

Newman, J. A. & Gruen, D. 2022, Annual Review of Astronomy and Astrophysics, 60, 363

pandas development team, T. 2020, pandas-dev/pandas: Pandas

Pandey, S., Krause, E., DeRose, J., et al. 2022, Physical Review D, 106

Percival, W. J., Friedrich, O., Sellentin, E., & Heavens, A. 2021, Monthly Notices of the Royal Astronomical Society, 510, 3207–3221

Raveri, M. & Hu, W. 2019, Physical Review D, 99

Rosell, A. C., Rodriguez-Monroy, M., Crocce, M., et al. 2021, Monthly Notices of the Royal Astronomical Society, 509, 778–799

Rozo, E., Rykoff, E. S., Abate, A., et al. 2016, Monthly Notices of the Royal Astronomical Society, 461, 1431–1450

Samuroff, S., Mandelbaum, R., Blazek, J., et al. 2023, Monthly Notices of the Royal Astronomical Society, 524, 2195

Sartoris, B., Biviano, A., Rosati, P., et al. 2020, Astronomy & Astrophysics, 637, A34

Scherrer, R. J. & Weinberg, D. H. 1998, The Astrophysical Journal, 504, 607–611

Schmidt, F. 2021, Journal of Cosmology and Astroparticle Physics, 2021, 033

Secco, L., Samuroff, S., Krause, E., et al. 2022, Physical Review D, 105

Seo, H. & Eisenstein, D. J. 2007, The Astrophysical Journal, 665, 14–24

Sevilla-Noarbe, I., Bechtol, K., Carrasco Kind, M., et al. 2021, The Astrophysical Journal Supplement Series, 254, 24

Siegmund, D. 1976, The Annals of Statistics, 4, 673–684

Sugiyama, S., Takada, M., Kobayashi, Y., et al. 2020, Physical Review D, 102

Sugiyama, S., Takada, M., Miyatake, H., et al. 2022, Physical Review D, 105

Szewciw, A. O., Beltz-Mohrmann, G. D., Berlind, A. A., & Sinha, M. 2022, The Astrophysical Journal, 926, 15

Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2017, Publications of the Astronomical Society of Japan, 70

Tröster, T., Mead, A. J., Heymans, C., et al. 2022, A&A, 660, A27

Troxel, M. & Ishak, M. 2015, Physics Reports, 558, 1–59

van Daalen, M. P., McCarthy, I. G., & Schaye, J. 2019, Monthly Notices of the Royal Astronomical Society, 491, 2424–2446

Velliscig, M., van Daalen, M. P., Schaye, J., et al. 2014, Monthly Notices of the Royal Astronomical Society, 442, 2641–2658

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261

Voivodic, R. & Barreira, A. 2021, Journal of Cosmology and Astroparticle Physics, 2021, 069

Walther, M., Armengaud, E., Ravoux, C., et al. 2021, Journal of Cosmology and Astroparticle Physics, 2021, 059

Wechsler, R. H., DeRose, J., Busha, M. T., et al. 2022, The Astrophysical Journal, 931, 145

Appendix A: Volume of re-scaled biased confidence interval

Here we give a derivation for Eq. 14, the re-scaling factor of the confidence interval volume when accounting for a bias by applying the change in confidence level according to the prescription given in Sec. 2.3.2.

The volume of the n-dimensional ellipsoid is given by:

$$V_n = \frac{2}{n} \frac{\pi^{n/2}}{\Gamma(n/2)} a_1 \cdot a_2 \cdot ... \cdot a_n , \qquad (A.1)$$

where a_i is the i-th semi-axis.

Assume the ellipsoid under consideration is the one defined by the target parameter covariance matrix. What the re-scaling does it to change the length of the semi-axes by a constant factor, namely by the ratio of the inverse CDF of the non-central χ^2 distribution with non-centrality parameter equal to the bias $\lambda = x^2$ defined in Eq. B.2, to the inverse CDF of the central χ^2 distribution.

$$
\begin{aligned}
V_n^{\text{extended}} &= \frac{2}{n} \frac{\pi^{n/2}}{\Gamma(n/2)} a_1 \cdot a_2 \cdot ... \cdot a_n \cdot \\
&\quad \left(\frac{\text{CDF}^{-1}(0.68; k = n, \lambda)}{\text{CDF}^{-1}(0.68; k = n, 0)} \right)^{n/2} \\
&= V_n \left(\frac{\text{CDF}^{-1}(0.68; k = n, \lambda)}{\text{CDF}^{-1}(0.68; k = n, 0)} \right)^{n/2}
\end{aligned}
\qquad (A.2)
$$

Since the denominator does not depend on the bias λ of the respective prior, it is left out from Eq. 14.

Appendix B: Summary of the DSS and its Stochasticity models

The spatial distribution of galaxies and matter can be described by their respective density contrast fields

$$\delta(x) = \frac{\rho(x) - \overline{\rho}}{\overline{\rho}}, \qquad (B.1)$$

where $\rho(x)$ is the density at position x and $\overline{\rho}$ is the mean density, of either matter (m) or galaxies (g).

The distribution of matter δ_m can not be directly observed, but it is traced by galaxies δ_g distributed in the sky. On large scales, the relation between galaxies and matter density contrasts can be approximated as linear

$$\delta_g \approx b \cdot \delta_m, \qquad (B.2)$$

with b being the linear galaxy bias.

We start by considering a distribution of galaxies in a fixed range around a redshift z_f, that we call *foreground galaxy population*. The 2D density field of the position of galaxies is obtained by applying a circular top-hat filter: at redshift z_f, we would count how many of the galaxies are inside of it. With this information, the sky can be divided into quintiles by ranked spatial number density. Next, we add another set of galaxies located at a higher redshift z_b ($z_f < z_b$) that we call the *background galaxy population*.

The light emitted by the background galaxies will be subject to gravitational shear when passing through the tidal gravitational field of the foreground. The effect will be stronger in zones where the density of foreground galaxies is higher.

To evaluate this, we measure the tangential shear that the background sources will suffer in each quintile of foreground density. The observed lensed galaxies will trace the distribution and number density of the source galaxies. This information is synthesized in a data vector of both the distribution of foreground galaxies and the lensing signals.

If galaxy counts and the matter density were perfectly correlated, then a split of the sky by galaxy density would be identical to a split by the matter density. In a more realistic scenario, however, there is scatter of galaxy count in a volume at fixed matter density. Reasons for this include the discrete sampling of the galaxy density field by individual galaxies, higher order galaxy bias terms that are averaged over in projection, intrinsic stochasticity in three dimensions, or observational systematics.

It is often assumed that the distribution of galaxy counts N_g traces a smooth field $\delta_g = (1 + b\delta_m)\overline{N}_g$, related to the matter overdensity δ_m in that region by a linear bias b, is given by Poisson shot-noise,

$$P(N_g|\delta_m) = \text{Poisson}[N_g, (1 + b\delta_m)\overline{N}_g], \qquad (B.3)$$

where $\overline{N}_g$ is the mean count of galaxies within the volume of such a region[15]. This means that the variance of N_g for fixed δ_g satisfies

$$\frac{\text{Var}[N_g|\delta_g]}{\langle N_g|\delta_g \rangle} = 1. \qquad (B.4)$$

When assuming a deterministic relationship between δ_m and δ_g, a consequence is that the variance of N_g for fixed δ_m is also

$$\frac{\text{Var}[N_g|\delta_m]}{\langle N_g|\delta_m \rangle} = 1. \qquad (B.5)$$

If the strong assumption of linear, deterministic bias is not fulfilled, DSS analysis needs to consider more degrees of freedom in Eq. B.5 to account for the effect of stochasticity.

Appendix B.1: α model: Parametric model for non-Poissonianity

As a way of generalizing the galaxy-matter connection, it is possible to add two stochasticity parameters, α_0 and α_1, to obtain an equation for a generalized Poisson distribution,

$$\frac{\text{Var}[N_g|\delta_m]}{\langle N_g|\delta_m \rangle} = \alpha_0 + \alpha_1 \delta_m, \qquad (B.6)$$

in which, for $\alpha_0 = 1.0$ and $\alpha_1 = 0.0$, we recover Eq. B.5.

The original prior distributions for the α parameters assigned in the original work from Friedrich et al. (2018); Gruen et al. (2018) are described in Table 1. The only actual constraint on stochasticity is that $\alpha_0 > 0$, but the boundary $0.1 < \alpha_0$ was originally motivated with numerical arguments. See App. E from Friedrich et al. (2018) for more details on the derivation for the other boundaries.

---

[15] Here, Poisson$[k, \lambda]$ denotes the probability of drawing a count k from a Poisson distribution with mean λ.
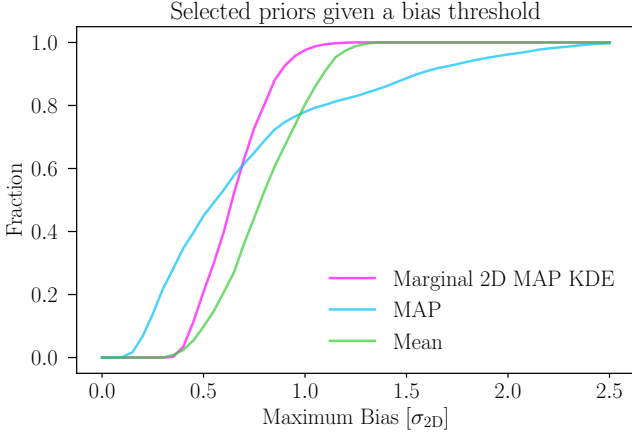
Fig. C.1: Fraction of the total number of priors from Sec. 3.3.1 that have a maximum bias smaller than σ_{2D}. In magenta, light blue, and light green we show the results for the 2D marginal MAP, the MAP, and the mean estimators respectively.

Appendix B.2: r model: correlation $r \neq 1$ between galaxy density and matter density

The second approach used in DSS is the introduction of a free parameter to the baseline linear bias model: a correlation coefficient r between the random fields δ_g and δ_m (described in Sec. IV of Friedrich et al. 2018). This new degree of freedom captures a possible stochasticity. For $r = 1$ the random fields δ_g and δ_m would be perfectly correlated.

The covariance of δ_g and δ_m can be parametrized by the correlation coefficient:

$$r = \frac{\langle \delta_g \delta_m \rangle}{\sqrt{\langle \delta_g^2 \rangle \langle \delta_m^2 \rangle}} \tag{B.7}$$

This can lead to a δ_m dependence of the ratio in Eq. B.5 and the general model for the galaxy count:

$$\mathrm{Var}[N_g] = \overline{N}_g + \overline{N}_g^2 b^2 \mathrm{Var}[\delta_m], \tag{B.8}$$

$$\mathrm{Cov}[N_g, \delta_m] = \overline{N}_g b r \mathrm{Var}[\delta_m]. \tag{B.9}$$

Mathematically, the possible values that r could take are $-1 \leq r \leq 1$. For galaxies one expects a positive correlation to matter density, so the original prior for r followed a uniform distribution $\Pi(r) \sim \mathcal{U}[0.0, 1.0]$.

Appendix C: ITEM priors with the MAP and Mean estimators

In this appendix, we show a complementary ITEM prior analysis using the Maximum A Posteriori (MAP) and mean as point estimators on the α model and contrast their performance with respect to the marginal MAP estimator used in the main text of this paper. We use specific colors for each point estimator, as shown in Fig. C.1.

The question of which point estimator to use for cosmological analysis is still in debate. DES used both the marginal MAP and the mean for presenting their summary cosmological statistics (Krause et al. 2021; Amon et al. 2022; Secco et al. 2022). Both the MAP and the 2D marginal MAP
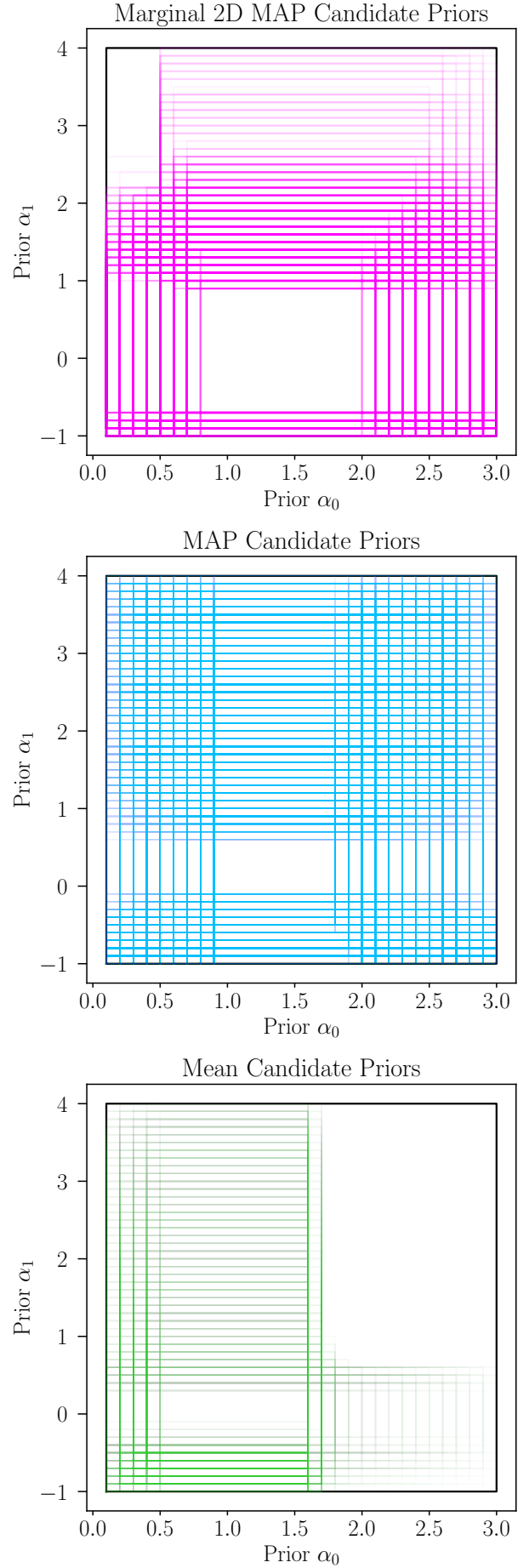
Fig. C.2: Priors that passed the first filter for each point estimator, or in other words, had a bias less than $0.4\sigma_{2D}$.
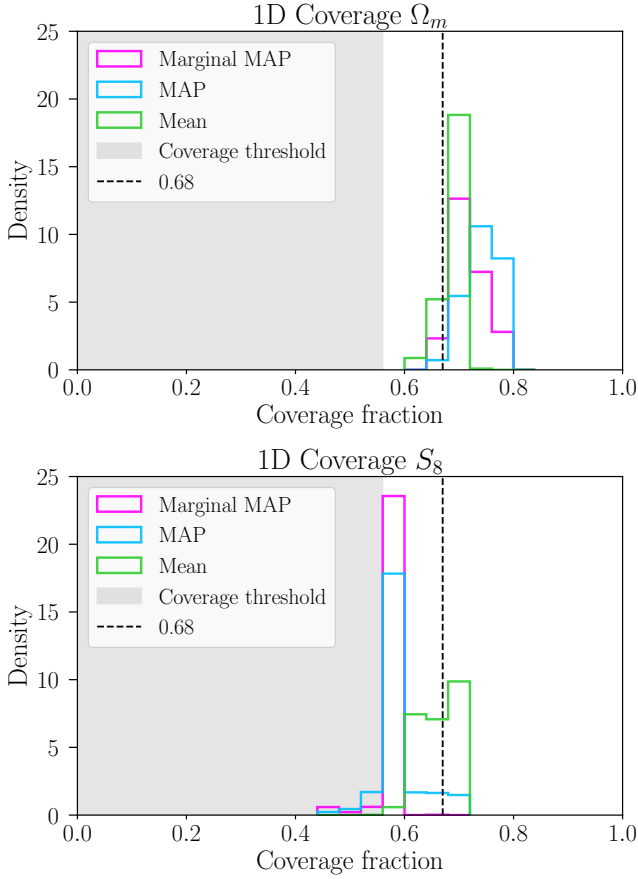
Fig. C.3: 1D Coverages for the candidate priors that passed the filter 1 for the different point estimators. On the upper panel we show the 1D coverage on Ω_m, while on the lower panel, we display the 1D coverage on S_8. Histograms are normalized to contrast their distribution along the vertical axis.

are non-trivial to estimate and sensitive to the algorithm used, as studied in Abbott et al. (2023).

In our study, the MAP was obtained using the maximum value output from CHAINCONSUMER (Hinton 2016), which Abbott et al. (2023) found to have a Gaussian kernel density estimation that leads to ~10% larger errors. On the other hand, the 2D marginal MAP was estimated using CHAINCONSUMER's Gaussian kernel density estimate (KDE) to smooth the marginalized posteriors, and therefore, estimate the marginal MAP. The estimate of this point is affected by the scale, therefore, after studying the impact of the bandpass, it was decided to keep the default values presented by CHAINCONSUMER in their API (kde = 1.0 in v0.33.0).

We use the same candidate priors as in Sec. 3.3.1, but we first studied the σ_{2D} biases for each model realization, but now using additional point estimators.

In Fig. C.1, we show the fraction of priors as a function of the maximum bias of the two stochasticity realizations. The MAP curve, in blue, resulted in both the lowest and largest maximum biases possible due to the prior volume effect present on the whole dimensional space. In magenta we have the 2D marginal MAP curve. This is the point estimator that returned the most constrained values of biases (all are covered with ~1.0 σ_{2D}). The advantage of using the 2D marginal MAP rather than the MAP as the main statistics in this context is that the 2D marginal MAP will not return values as biased as the ones calculated using

Table C.1: Number of priors passing each filter given a point estimator.

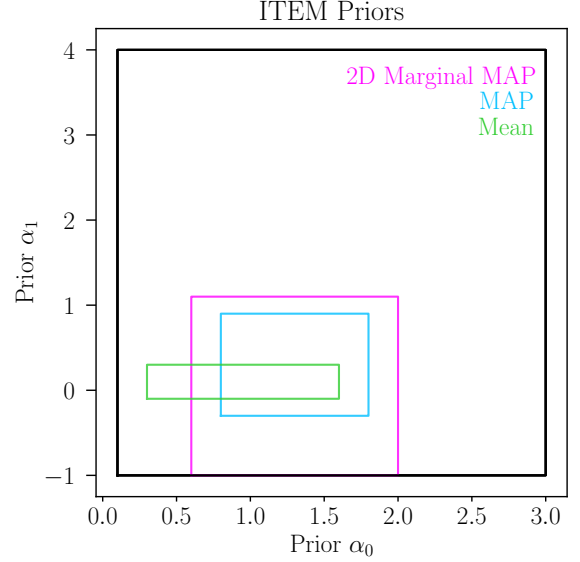| Point estimator	Filter 1	Filter 2
2D marginal MAP	1,980	1,868
MAP	20,160	18,232
Mean	1,401	1,400

Fig. C.4: Distribution of the ITEM priors for the different tested point estimators.

the MAP in the presence of projection effects. Finally, in green, the mean estimator had the worst bias due to volume effects happening when marginalizing the 1D posteriors.

Following Sec. 3.3.1, we used a bias of $0.4\sigma_{2D}$ for each estimator. The number of candidate priors that passed those bias thresholds are listed in Table C.1, and are shown in Fig. C.2.

The next step in determining the ITEM priors is to study the fraction of priors that have a minimum frequentist coverage. The 1D coverage distributions are visible in Fig. C.3. Following Sec. 3.3.1, we required the candidate priors to have the same coverage fractions for the noisy realizations, i.e. at least a 56% (14 out of 25 realizations) coverage for Ω_m and S_8. This lower limit on coverages decreased the number of priors as shown in the last column of Table C.1.

The last step of the ITEM prior implementation is to minimize the total uncertainty of the posterior. In Fig. C.4, we show the resulting priors on both stochasticity parameters α_0 and α_1, which differ depending on the point estimator.

The ITEM priors for both α_0 and α_1 have substantially smaller widths with respect to the one from the original prior (see Table 2), especially the case of the mean. In particular, the upper limit on the parameter α_1 is reduced to less than half of the original prior width in all cases. In Table C.2, we display the main statistics for the ITEM priors obtained using different point estimators as a reference, and, in Fig. C.5 we show their marginalized posteriors.

All the biases for both nuisance result in similar values and we obtained consistent coverages despite the use of varying different priors. The $\overline{\sigma}$ uncertainties for the ITEM priors and the total uncertainty $\tilde{V}/\tilde{V}_{\text{Original}}$ decreased consistently with the prior volume reduction. The marginal posteriors from Fig. C.5 are also

Table C.2: The ITEM priors for the α model for different point estimators.

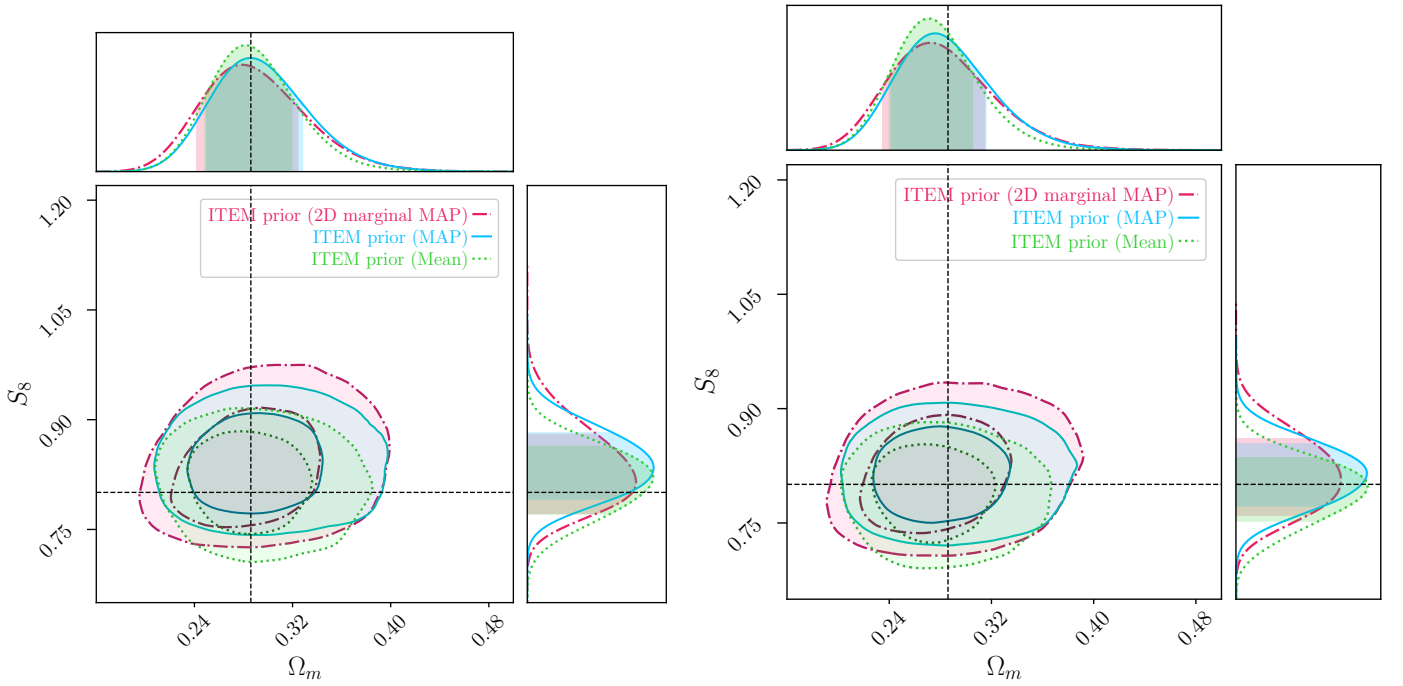| Parameter	Prior	Prior Width	Bias [σ_{2D}]	Ω_m Coverage	S_8 Coverage	$\overline{\sigma}_{\Omega_m}$	$\overline{\sigma}_{S_8}$	$\tilde{V}/\tilde{V}_{\text{Original}}$
2D MAP								
α_0	$\mathcal{U}[0.6, 2.0]$	1.4						
			0.33/0.39	80%/64%	68%/56%	0.040	0.049	0.58
α_1	$\mathcal{U}[-1.0, 1.1]$	2.1						
MAP								
α_0	$\mathcal{U}[0.8, 1.8]$	1.0						
			0.30/0.40	80%/80%	80%/56%	0.037	0.040	0.52
α_1	$\mathcal{U}[-0.3, 0.9]$	1.2						
Mean								
α_0	$\mathcal{U}[0.3, 1.6]$	1.3						
			0.35/0.40	68%/72%	80%/64%	0.033	0.040	0.34
α_1	$\mathcal{U}[-0.1, 0.3]$	0.4						

Fig. C.5: Resulting marginalized posteriors for the ITEM priors obtained with different point estimators. Similar to Fig. 9, the 1-σ and 2-σ confidence levels had been widened to account for the bias of each case.

in agreement with what we would expect from divergent priors: the three of them are different.