Extreme data compression for Bayesian model comparison

Heavens, A.F.; Mootoovaloo, A.; Trotta, R.; Sellentin, E.

Extreme data compression for Bayesian model comparison

To cite this article: Alan F. Heavens *et al* JCAP11(2023)048

View the article online for updates and enhancements.

Extreme data compression for Bayesian model comparison

Alan F. Heavens,[a,*] **Arrykrishna Mootoovaloo,**[b] **Roberto Trotta**[a,c,d] **and Elena Sellentin**[e,f]

[a]*Imperial Centre for Inference and Cosmology (ICIC), Department of Physics, Imperial College,*
  Blackett Laboratory, Prince Consort Road, London SW7 2AZ, U.K.
[b]*Department of Physics, University of Oxford, Denys Wilkinson Building,*
  Keble Road, Oxford OX1 3RH, U.K.
[c]*Theoretical and Scientific Data Science,*
  Scuola Internazionale Superiore di Studi Avanzati (SISSA), via Bonomea 265, 34136 Trieste, Italy
[d]*Centro Nazionale "High Performance Computing, Big Data and Quantum Computing", Italy*
[e]*Mathematical Institute, Leiden University, Snellius Gebouw,*
  Niels Bohrweg 1, NL-2333 CA Leiden, The Netherlands
[f]*Leiden Observatory, Leiden University, Oort Gebouw,*
  Niels Bohrweg 2, NL-2333 CA Leiden, The Netherlands

  E-mail: a.heavens@imperial.ac.uk,
  arrykrishna.mootoovaloo@physics.ox.ac.uk, rtrotta@sissa.it,
  sellentin@strw.leidenuniv.nl

ABSTRACT: We develop extreme data compression for use in Bayesian model comparison via the MOPED algorithm, as well as more general score compression. We find that Bayes Factors from data compressed with the MOPED algorithm are identical to those from their uncompressed datasets when the models are linear and the errors Gaussian. In other nonlinear cases, whether nested or not, we find negligible differences in the Bayes Factors, and show this explicitly for the Pantheon-SH0ES supernova dataset. We also investigate the sampling properties of the Bayesian Evidence as a frequentist statistic, and find that extreme data compression reduces the sampling variance of the Evidence, but has no impact on the sampling distribution of Bayes Factors. Since model comparison can be a very computationally-intensive task, MOPED extreme data compression may present significant advantages in computational time.

---

*Corresponding author.

Contents

1 Introduction

Astronomical datasets can be very large, with galaxy catalogues and weak lensing surveys having up to hundreds of millions of objects [e.g., 1–3], and Planck microwave background data [4] containing billions of time-ordered data. Extracting scientific knowledge from such large datasets almost inevitably requires some form of data compression, typically into summary statistics such as point estimates of two-point summaries, either correlation functions or power spectra. Compression without loss of information is generally possible since the datasets are noisy and correlated, and furthermore the information may be captured by a small number of summary statistics, depending on the statistical properties of the underlying data. However, even with compression to two-point summaries, the number of statistics can still be large: for example, cosmic shear surveys such as proposed for the Euclid mission [5] may generate $\sim 10^4$ two-point statistics, and this can present problems for likelihood-based scientific analysis if the covariance of the summaries needs to be simulated, or indeed if their distributions are not known [6]. Both of these considerations motivate the use of more radical data compression techniques. This can also be highly beneficial when the resultant summaries are coupled with techniques such as simulation-based inference (SBI, also known as likelihood-free inference), which is undergoing rapid development thanks in part to novel machine learning techniques [e.g., 7–11]. Typically, the most extreme compression possible without losing information is to reduce the dataset down from its original size to the number of model parameters to be inferred. Such extreme compression can have a number of benefits, in terms of speed of analysis, in reducing the number of simulations required to determine the sampling distribution of the statistics [6], and in making SBI feasible [12, 13], whilst still

allowing for new models to be considered [14]. Heavens et al. [15] derived the optimal data compression algorithm MOPED, for gaussian-distributed data whose mean signal depends on the parameters, showing that extreme data compression can preserve the Fisher matrix, so the compression is, in this sense, lossless. Techniques, based on Karhünen-Loève compression [16] also exist for data whose covariance depends on the parameters, but not the mean, in contrast to MOPED, which applies when the mean depends on the parameters, but not the covariance. However, the former situation does not lead to lossless extreme compression, nor does it generalise neatly to multiple parameters. MOPED, whether in its original Gram-Schmidt orthogonal form, or in the simpler non-orthogonalised version (which we use in this paper) is a special case of using the score function (the gradient of the likelihood with respect to the model parameters) to compress the data, as shown by [17]. Alternative approaches to extreme data compression include information-maximising neural networks (IMNNs) to find maximally-informative summaries, as introduced by [7], and which have in certain cases been shown to be lossless at Fisher level [9]. Typically, the neural network-based methods operate on the fields themselves [e.g. 9–11], whereas the MOPED and score compressions operate on some intermediate, physics-based summaries [e.g. 6, 10, 12, 18].

The main motivation for this paper is to determine whether the qualities of optimal extreme data compression which apply to parameter inference also extend to Bayesian Evidence and Bayes Factors, i.e., to model comparison. The answer to this question is yes, provided that the extreme data compression is performed in each of the models separately, and the compressed data then combined, such that the combined compressed dataset is by construction optimal in all models. This applies to both nested and non-nested models.

With this key result, committed Bayesians may stop reading, but we also take the opportunity to investigate the sampling distribution of the Bayesian Evidence and the related Bayes Factor, to look at their use as frequentist statistics under extreme compression. The motivation for this comes from a suggestion [19] that data compression may help in reducing the sampling variability in the Bayesian Evidence. The answer is that it does, but the sampling variance of the Bayes Factor, which is a ratio of evidences and the relevant quantity for model comparison, remains virtually unchanged. To rephrase this same point, one can use extreme MOPED data compression as effectively as the full dataset for model comparison as well as parameter inference. Since model comparison is typically a computationally expensive exercise, extreme data compression, which often results in faster likelihood evaluation, can be very valuable.

Note that the distinction between the sampling distribution of the Bayesian Evidence and of the Bayes Factor is important, especially since it is tempting to conclude — erroneously — that the sampling distribution of the Bayesian Evidence is sufficiently wide to compromise its usefulness in model selection, but this ignores the fact that much of the variation is common to both models (the data are the same), and cancels out when the ratio of Bayesian Evidence is computed, leaving the Bayes Factor as an effective discriminant, as we show in this paper.

The layout of the paper is as follows. In section 2 we review the MOPED data compression algorithm, and show how it works effectively for model comparison as well as parameter inference, and consider more general score compression. In section 3.1 we explore the frequentist properties of Bayes Factors, showing that the sampling distribution of the Bayes Factor

is unaltered whether the full or compressed data are used. In section 3.2 we apply MOPED compression to the Pantheon+SH0ES dataset [20]) and show that both the cosmological parameters and the Bayes Factor can be determined from just 3 numbers (the MOPED coefficients), as precisely as from the original 1590 supernovae. We conclude in section 4.

2 Extreme data compression: Bayesian Evidence with MOPED

2.1 Bayesian Evidence and Bayes Factors for the uncompressed dataset

Let us first consider gaussian-distributed data, with fixed covariance matrix and arbitrary dependence of the mean on the parameters. Data compression with MOPED [15] allows extreme compression to p summary statistics, where p is the number of model parameters, whatever the length of the original data vector. Although the resulting MOPED coefficients are not strictly sufficient statistics except in the case of linear models, they can preserve the entire Fisher matrix, so the likelihood surface near the peak is no broader than when using the full, uncompressed dataset. As a result, the extreme compression can be used for parameter inference without increasing errors, and has been used in many contexts, such as star formation histories of SDSS galaxies [e.g. 21–24], the cosmic microwave background [e.g. 25–27], galaxy power spectrum and bispectrum [e.g. 28, 29], gravitational waves [30], weak lensing [e.g. 12, 31], combined probes [e.g. 32, 33], planetary transits [34], and extended to parameter-dependent covariance matrices [6] and mis-specified models [14]. In this section, we explore how MOPED and related score compression can be used in a new context: model comparison. To begin we review the derivation of the unorthogonalised version of MOPED, following not the original analysis [15], but rather the derivation in [17].

We consider a model with parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_p\}$, data $\boldsymbol{x} \in \mathbb{R}^n$ (where $n \gg p$) generated from a gaussian sampling distribution with covariance matrix $\mathsf{C} \in \mathbb{R}^{n \times n}$ independent of $\boldsymbol{\theta}$ and expectation value of the data $\boldsymbol{\mu}(\boldsymbol{\theta})$. The sampling distribution for the uncompressed dataset is therefore

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \frac{1}{\sqrt{|2\pi\mathsf{C}|}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\mathsf{C}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]. \tag{2.1}$$

Letting $\boldsymbol{X} \equiv \boldsymbol{x} - \boldsymbol{\mu}_*$, where $\boldsymbol{\mu}_*$ is the expectation value of the data at some fiducial set of parameters $\boldsymbol{\theta}_*$, the posterior is

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \frac{1}{\sqrt{|2\pi\mathsf{C}|}} \exp\left[-\frac{1}{2}(\boldsymbol{X}-\tilde{\boldsymbol{\mu}})^T\mathsf{C}^{-1}(\boldsymbol{X}-\tilde{\boldsymbol{\mu}})\right]\pi(\boldsymbol{\theta}) \tag{2.2}$$

for a prior density $\pi(\boldsymbol{\theta})$. Here

$$\tilde{\boldsymbol{\mu}}(\boldsymbol{\theta}) \equiv \boldsymbol{\mu}(\boldsymbol{\theta}) - \boldsymbol{\mu}_*, \tag{2.3}$$

and a linear Taylor expansion around the fiducial point gives

$$\tilde{\boldsymbol{\mu}} \simeq \boldsymbol{\mu}_{,\alpha}\tilde{\boldsymbol{\theta}}_\alpha \quad \text{where} \quad \tilde{\boldsymbol{\theta}} \equiv \boldsymbol{\theta} - \boldsymbol{\theta}_*. \tag{2.4}$$

The comma indicates $\partial/\partial\boldsymbol{\theta}_\alpha$ and the summation convention is used. Hence to this order,

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \frac{1}{\sqrt{|2\pi\mathsf{C}|}} \exp\left(-\frac{1}{2}\boldsymbol{X}^T\mathsf{C}^{-1}\boldsymbol{X}\right)\exp\left(\tilde{\boldsymbol{\mu}}^T\mathsf{C}^{-1}\boldsymbol{X} - \frac{1}{2}\tilde{\boldsymbol{\mu}}^T\mathsf{C}^{-1}\tilde{\boldsymbol{\mu}}\right)\pi(\boldsymbol{\theta}). \tag{2.5}$$

We see that the only coupling of the data $\boldsymbol{x}$ to the parameters is through the $\tilde{\boldsymbol{\mu}}^T \mathsf{C}^{-1} \boldsymbol{X}$ term, and only via the p MOPED coefficients

$$Y_\alpha \equiv \boldsymbol{\mu}_{,\alpha}^T \mathsf{C}^{-1} \boldsymbol{X} \equiv \boldsymbol{b}_\alpha^T (\boldsymbol{x} - \boldsymbol{\mu}_*), \qquad \alpha = 1 \ldots p, \tag{2.6}$$

and we identify the p (unorthogonalised) MOPED weight vectors [15])

$$\boldsymbol{b}_\alpha^T \equiv \boldsymbol{\mu}_{,\alpha}^T \mathsf{C}^{-1}. \tag{2.7}$$

The posterior may then be written

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \frac{1}{\sqrt{|2\pi \mathsf{C}|}} \exp\left(-\frac{1}{2} \boldsymbol{X}^T \mathsf{C}^{-1} \boldsymbol{X}\right) \exp\left[(\boldsymbol{b}_\alpha^T \boldsymbol{X}) \tilde{\theta}_\alpha - \frac{1}{2} (\boldsymbol{b}_\alpha^T \mathsf{C} \boldsymbol{b}_\beta) \tilde{\theta}_\alpha \tilde{\theta}_\beta\right] \pi(\tilde{\boldsymbol{\theta}}), \tag{2.8}$$

or in matrix form

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \frac{1}{\sqrt{|2\pi \mathsf{C}|}} \exp\left(-\frac{1}{2} \boldsymbol{X}^T \mathsf{C}^{-1} \boldsymbol{X}\right) \exp\left(\boldsymbol{Y}^T \tilde{\boldsymbol{\theta}} - \frac{1}{2} \tilde{\boldsymbol{\theta}}^T \Lambda \tilde{\boldsymbol{\theta}}\right) \pi(\tilde{\boldsymbol{\theta}}), \tag{2.9}$$

where $\boldsymbol{Y}$ is a vector of only p elements, consisting of the MOPED coefficients:

$$\boldsymbol{Y} = \mathsf{B}^T \boldsymbol{X} = \mathsf{B}^T (\boldsymbol{x} - \boldsymbol{\mu}_*), \tag{2.10}$$

and $\mathsf{B}^T \in \mathbb{R}^{p \times n}$ is made up of p rows of MOPED vectors $\boldsymbol{b}_\alpha^T$, defined by equation (2.7). In matrix form,

$$\mathsf{B}^T = \Phi^T \mathsf{C}^{-1}, \tag{2.11}$$

where $\Phi^T \in \mathbb{R}^{p \times n}$ is a matrix with rows given by $\boldsymbol{\mu}_{,\alpha}^T$. The symmetric matrix $\Lambda \in \mathbb{R}^{p \times p}$ is

$$\Lambda = \mathsf{B}^T \mathsf{C} \mathsf{B} = \Phi^T \mathsf{C}^{-1} \Phi. \tag{2.12}$$

Note that Heavens et al. [15] applied a Gram-Schmidt orthogonalisation to the MOPED vectors, to decorrelate the elements of $\boldsymbol{Y}$, but the information content is the same if this is not applied, and we omit it as it simplifies the analysis. We have also subtracted $\boldsymbol{\mu}_*$ from $\boldsymbol{x}$ for convenience, and to ensure that a Taylor expansion to linear order is accurate, following [6].

To normalise the posterior, $p(\boldsymbol{\theta}|\boldsymbol{x}, M) = p(\boldsymbol{x}|\boldsymbol{\theta}, M) \pi(\boldsymbol{\theta})/p(\boldsymbol{x}|M)$, we need the Bayesian Evidence $Z(\boldsymbol{x}) = p(\boldsymbol{x}|M)$, where we have added explicitly the dependence on the choice of model, M. The Evidence is

$$p(\boldsymbol{x}|M) = \frac{1}{\sqrt{|2\pi \mathsf{C}|}} \exp\left(-\frac{1}{2} \boldsymbol{X}^T \mathsf{C}^{-1} \boldsymbol{X}\right) \int \exp\left(\boldsymbol{Y}^T \tilde{\boldsymbol{\theta}} - \frac{1}{2} \tilde{\boldsymbol{\theta}}^T \Lambda \tilde{\boldsymbol{\theta}}\right) \pi(\tilde{\boldsymbol{\theta}}) \, d\tilde{\boldsymbol{\theta}}. \tag{2.13}$$

We see that the data-only prefactor cancels in the normalized posterior:

$$p(\boldsymbol{\theta}|\boldsymbol{x}, M) = \frac{\exp\left(\boldsymbol{Y}^T \tilde{\boldsymbol{\theta}} - \frac{1}{2} \tilde{\boldsymbol{\theta}}^T \Lambda \tilde{\boldsymbol{\theta}}\right) \pi(\tilde{\boldsymbol{\theta}}|M)}{\int \exp\left(\boldsymbol{Y}^T \tilde{\boldsymbol{\theta}} - \frac{1}{2} \tilde{\boldsymbol{\theta}}^T \Lambda \tilde{\boldsymbol{\theta}}\right) \pi(\tilde{\boldsymbol{\theta}}|M) \, d\tilde{\boldsymbol{\theta}}}. \tag{2.14}$$

From this it is apparent that, to the extent that the linear expansion is applicable, we may compute the posterior from the MOPED coefficients alone, without considering $\boldsymbol{x}$ in full — we only need $\mathbf{Y}$. This is a very powerful feature of MOPED compression.

Note that for a Gaussian prior density, $\pi(\tilde{\boldsymbol{\theta}}|M) = \mathcal{N}(\mathbf{0}, \Sigma)$ the integrals may be evaluated analytically (see also [13, 35, 36]) giving

$$Z(\boldsymbol{X}) = \frac{1}{\sqrt{|2\pi \mathsf{C}||\Sigma \Lambda + \mathbb{I}_p|}} \exp\left(-\frac{1}{2} \boldsymbol{X}^T \mathsf{C}^{-1} \boldsymbol{X}\right) \exp\left[\frac{1}{2} \boldsymbol{Y}^T \left(\Lambda + \Sigma^{-1}\right)^{-1} \boldsymbol{Y}\right]. \tag{2.15}$$

2.1.1 Demonstration that the Bayes Factor depends only on MOPED coefficients

Before we derive the general expression for the Bayes Factor in the next section, we motivate our approach by demonstrating that, for nested models, the Bayes Factor from the full dataset depends only on the set of MOPED coefficients defined in the extended model, to linear order in expansion. The posterior odds between models M_0 and M_1 are

$$\frac{p(M_0|\boldsymbol{x})}{p(M_1|\boldsymbol{x})} = \frac{\pi(M_0)p(\boldsymbol{x}|M_0)}{\pi(M_1)p(\boldsymbol{x}|M_1)} = B_{01}(\boldsymbol{x}) \equiv \frac{Z_0(\boldsymbol{x})}{Z_1(\boldsymbol{x})} \tag{2.16}$$

where B_{01} is the Bayes Factor, and the last equality applies if we assume equal prior model probabilities, $\pi(M_0) = \pi(M_1)$. We have defined $Z_i \equiv p(\boldsymbol{x}|M_i)$.

Model M_0, with q free parameters, is said to be 'nested' in model M_1, with $p > q$ parameters, if the first q parameters of Model 1 are common to Model 0, and there is a choice of value for the additional parameters in Model 1 such that the latter reverts to Model 0. In this setting, we choose a fiducial parameter set within the prior support of Model 0 (and, therefore, *a fortiori* in Model 1); in this case, $\boldsymbol{\mu}_*$ is common to both models, and $\boldsymbol{X}$ is the same in both. in this case, the prefactor in the Evidence (equation (2.13)), cancels out in the Bayes Factor, for any choice of prior densities $\pi(\tilde{\boldsymbol{\theta}}_i|M_i)$

$$B_{01}(\boldsymbol{x}) \equiv \frac{Z_0(\boldsymbol{x})}{Z_1(\boldsymbol{x})} = \frac{\int \exp\left(\boldsymbol{Y}^T\tilde{\boldsymbol{\theta}}_0 - \frac{1}{2}\tilde{\boldsymbol{\theta}}_0^T\Lambda_0\tilde{\boldsymbol{\theta}}_0\right)\pi(\tilde{\boldsymbol{\theta}}_0|M_0)\,d\tilde{\boldsymbol{\theta}}_0}{\int \exp\left(\boldsymbol{Y}^T\tilde{\boldsymbol{\theta}}_1 - \frac{1}{2}\tilde{\boldsymbol{\theta}}_1^T\Lambda_1\tilde{\boldsymbol{\theta}}_1\right)\pi(\tilde{\boldsymbol{\theta}}_1|M_1)\,d\tilde{\boldsymbol{\theta}}_1}. \tag{2.17}$$

Note that for the simpler model the extra parameters are set at $\tilde{\boldsymbol{\theta}} = \boldsymbol{0}$, which we accommodate by defining Λ_0 as the top $q \times q$ block of Λ_1.

Just like the posterior for the parameters, the Bayes Factor for the full dataset depends on the data only via $\boldsymbol{Y}$. This motivates adopting MOPED compression for model comparison in the general case, with the potential for a large acceleration of what can be a computationally expensive task for some applications. We develop this in the next section.

It is worth pointing out that the requirements on the accuracy of the Taylor expansion are more stringent for model comparison than for parameter inference, since we are evaluating an integral over the whole parameter space, well beyond the region where the posterior density is appreciable. For some applications this could limit the accuracy of the approach, but for typical cosmological posteriors, their unimodal and relatively simple structure means that the MOPED-derived posteriors are often very close to those obtained from the full dataset, even far from the posterior peak. This is apparent even in the first (astrophysical) example, presented in [15], where the posterior contours are similar between MOPED and full data set at values smaller than the peak by a factor e^{100}, even though this is not guaranteed by the method — the compression does far better than one has any right to expect. We therefore compress the data only once for each model, each at a single fiducial parameter set, and we illustrate later in a cosmological supernova study that the Bayes Factor computation from the compressed data still agrees with the full dataset within sampling errors.

2.2 Bayesian Evidence and Bayes Factor with MOPED compression for non-nested models

We now develop the formalism for using MOPED extreme data compression for computing Bayes Factors for non-nested models, with the nested case requiring only a minimal adjustment. In order for the Bayes Factor to give the correct Bayesian posterior odds ratio (for equal model priors), the data entering in each of the Bayesian Evidence calculations have to be the same, or else the $p(\text{data})$ denominator in Bayes theorem will not cancel in the ratio. Since the MOPED vectors defined within each model are optimal for parameter inference in their respective models, we construct a compressed dataset from the union of the MOPED coefficients from both (or, more generally, all) models. In this way the augmented compressed dataset will be optimal for parameter inference in both models, since we will not lose information by adding additional data beyond what is required. Of course we must take care to include correctly the correlations between all the compressed data. We thus compare the analysis with the full dataset with the data compressed to

$$\boldsymbol{Y}^T = \begin{pmatrix} \boldsymbol{Y}_0^T \\ \boldsymbol{Y}_1^T \end{pmatrix} = \begin{pmatrix} \mathsf{B}_0^T(\boldsymbol{x} - \boldsymbol{\mu}_{*0}) \\ \mathsf{B}_1^T(\boldsymbol{x} - \boldsymbol{\mu}_{*1}) \end{pmatrix} \in \mathbb{R}^{p+q}, \tag{2.18}$$

where $\boldsymbol{\mu}_{*0}$ and $\boldsymbol{\mu}_{*1}$ are fiducial means in the two models. In the nested case, these would be chosen to be equal, and since $\boldsymbol{Y}_0$ is already contained in $\boldsymbol{Y}_1$, it is discarded, with $\boldsymbol{Y}_1$ alone being retained. Here we assume non-nested models when both $\boldsymbol{Y}_0$ and $\boldsymbol{Y}_1$ are present, and specialize to the nested case later.

The Bayesian Evidence for each model, $Z_i(\boldsymbol{Y})$ is obtained by very similar arguments to the preceding analysis. For Gaussian data $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathsf{C})$, $\boldsymbol{Y} \sim \mathcal{N}(\bar{Y}, \Lambda)$ is also Gaussian-distributed, with mean

$$\bar{\boldsymbol{Y}}(\tilde{\boldsymbol{\theta}}) = \begin{pmatrix} \mathsf{B}_0^T(\boldsymbol{\mu} - \boldsymbol{\mu}_{*0}) \\ \mathsf{B}_1^T(\boldsymbol{\mu} - \boldsymbol{\mu}_{*1}) \end{pmatrix}, \tag{2.19}$$

and covariance $\Lambda = \Phi^T \mathsf{C}^{-1} \Phi \in \mathbb{R}^{(p+q) \times (p+q)}$, where in block form $\Phi = (\Phi_0, \Phi_1)$. The Bayesian Evidence for model i is then

$$Z_i(\boldsymbol{Y}) = \frac{1}{\sqrt{|2\pi\Lambda|}} \int \exp\left[-\frac{1}{2}\left(\boldsymbol{Y} - \bar{\boldsymbol{Y}}\right)^T \Lambda^{-1} \left(\boldsymbol{Y} - \bar{\boldsymbol{Y}}\right)\right] \pi(\tilde{\boldsymbol{\theta}}_i | M_i)\, d\tilde{\boldsymbol{\theta}}_i. \tag{2.20}$$

With a Taylor expansion for $\boldsymbol{\mu}$ as before (equation (2.4)), $\bar{\boldsymbol{Y}} = \mathsf{B}^T \Phi \tilde{\boldsymbol{\theta}} \equiv \Lambda \tilde{\boldsymbol{\theta}}$, we find

$$Z_i(\boldsymbol{Y}) = \frac{1}{\sqrt{|2\pi\Lambda|}} \exp\left[-\frac{1}{2}(\boldsymbol{Y} - \bar{\boldsymbol{Y}}_{*i})^T \Lambda^{-1}(\boldsymbol{Y} - \bar{\boldsymbol{Y}}_{*i})\right]$$
$$\times \int \exp\left[(\boldsymbol{Y} - \bar{\boldsymbol{Y}}_{*i})^T \Lambda^{-1} \mathsf{A}_i \tilde{\boldsymbol{\theta}}_i - \frac{1}{2}\tilde{\boldsymbol{\theta}}_i^T \mathsf{A}_i^T \Lambda^{-1} \mathsf{A}_i \tilde{\boldsymbol{\theta}}_i\right] \pi(\tilde{\boldsymbol{\theta}}_i | M_i) d\tilde{\boldsymbol{\theta}}_i, \tag{2.21}$$

where $\mathsf{A}_i \equiv \Phi^T \mathsf{C}^{-1} \Phi_i$ and

$$\bar{\boldsymbol{Y}}_{*i} = \begin{pmatrix} \mathsf{B}_0^T(\boldsymbol{\mu}_{*i} - \boldsymbol{\mu}_{*0}) \\ \mathsf{B}_1^T(\boldsymbol{\mu}_{*i} - \boldsymbol{\mu}_{*1}) \end{pmatrix}, \tag{2.22}$$

which will have zeros in one or other block depending on whether $i = 0$ or 1.

2.2.1 Nested models with MOPED compression

There is a significant simplification for nested models, where we dispense with the top blocks of $\boldsymbol{Y}$ and Φ^T, and take $\boldsymbol{\mu}_{*0} = \boldsymbol{\mu}_{*1}$. As a consequence $\bar{\boldsymbol{Y}}_{*i} = \boldsymbol{0}$. A_i can be replaced by Λ (which reduces to Λ_1), since it always comes in combination with $\tilde{\boldsymbol{\theta}}_i$, and we extend the $\tilde{\boldsymbol{\theta}}_0$ vector to include zeros from the extended parameter space. With these simplifications, the Evidence becomes

$$Z_i(\boldsymbol{Y}) = \frac{1}{\sqrt{|2\pi\Lambda|}} \exp\left(-\frac{1}{2}\boldsymbol{Y}^T\Lambda^{-1}\boldsymbol{Y}\right) \int \exp\left(\boldsymbol{Y}^T\tilde{\boldsymbol{\theta}}_i - \frac{1}{2}\tilde{\boldsymbol{\theta}}_i^T\Lambda\tilde{\boldsymbol{\theta}}_i\right) \pi(\tilde{\boldsymbol{\theta}}_i|M_i)d\tilde{\boldsymbol{\theta}}_i. \quad (2.23)$$

This is very similar to the Evidence computed from the full data set (equation (2.13)), but with a different data-dependent prefactor to the integral.

When we set the prior on the extended model parameters to Dirac delta functions in Model 0, Λ effectively becomes Λ_0, and when we cancel the common prefactors, we recover equation (2.17), and find that the Bayes Factor for the MOPED compressed data for nested models is exactly the same as the Bayes Factor for the original dataset. In other words, to the extent that a linear Taylor expansion is accurate, no information is lost in performing Bayesian model comparison using MOPED coefficients rather than the full dataset.

2.2.2 Gaussian prior

With a gaussian prior on the parameters, the Bayesian Evidence (equation (2.21)) for the different models (nested or otherwise) reduces, after application of the Woodbury formula and the matrix determinant lemma, to a very simple expression

$$Z_i(\boldsymbol{Y}) = \frac{1}{\sqrt{|2\pi Q_i|}} \exp\left[-\frac{1}{2}(\boldsymbol{Y} - \bar{\boldsymbol{Y}}_{*i})^T Q_i^{-1}(\boldsymbol{Y} - \bar{\boldsymbol{Y}}_{*i})\right] \quad (2.24)$$

where

$$Q_i = \Lambda + \mathsf{A}_i^T\Sigma_i\mathsf{A}_i. \quad (2.25)$$

2.3 Score compression

We can generalise the result further to arbitrary likelihood functions, using the score compression of Alsing & Wandelt [17], of which MOPED is a special case. Here they expand the log likelihood as

$$\mathcal{L} = \mathcal{L}_* + \tilde{\boldsymbol{\theta}}^T\boldsymbol{\nabla}\mathcal{L}_* - \frac{1}{2}\tilde{\boldsymbol{\theta}}^T\boldsymbol{J}_*\tilde{\boldsymbol{\theta}} \quad (2.26)$$

where $\boldsymbol{J} \equiv -\boldsymbol{\nabla}\boldsymbol{\nabla}^T\mathcal{L}$, and $*$ indicates that quantities are evaluated at the fiducial parameters.

$\mathcal{L}_*$ depends on the data, but we see again that the $\mathcal{L}_*$ (which is a source of variability in the Bayesian Evidence) cancels in the Bayes Factor if the models are nested and $\boldsymbol{\theta}_*$ lies in the domain of both M_0 and M_1. The only coupling between data and parameters at linear order is through the score function

$$\boldsymbol{t} \equiv \boldsymbol{\nabla}\mathcal{L}_*. \quad (2.27)$$

$\boldsymbol{J}$ is strictly constant in the Gaussian case if C is independent of $\boldsymbol{\theta}$, and is given by the Fisher matrix

$$\langle\boldsymbol{J}\rangle = \boldsymbol{F} = \boldsymbol{\nabla}\boldsymbol{\mu}^T\mathsf{C}\boldsymbol{\nabla}^T\boldsymbol{\mu}. \quad (2.28)$$

If C is parameter-dependent, the expression is complicated, and couples data and parameters (see [17], equation (15)), and replacing $\boldsymbol{J}$ by $\boldsymbol{F}$ would be an approximation, but possibly a good one.

We see that the preceding arguments hold in this case as well, that for nested models, the $\mathcal{L}_*$ term cancels out in both the posterior and also the Bayes Factor, if the fiducial parameters are common to both (all) models. As a result, general score compression, using the trick of concatenating the scores from all models considered, can effectively be used for model comparison, with the additional approximation (2.28).

3 Example applications

3.1 Distribution of Bayes Factors

Bayesians may wish to turn away now, as we consider the frequentist properties of Bayesian Evidence and Bayes Factors, as investigated by Jenkins & Peacock [37] and Joachimi et al. [19]. These papers note that the Bayesian Evidence and the Bayes Factor are noisy statistics, and consider their sampling variability and its impact on the outcome of model comparison. Note that we remain of the Bayesian view that the Bayes Factor can indicate a preference for one model over another, but not falsify models in absolute terms, and we are not seeking frequentist methods to do this, which typically rely on tail probabilities of $p(\text{data}|M)$ as opposed to posterior model probabilities, $p(M|\text{data})$, cf. [38–40].

To the committed Bayesian, the sampling distribution is not relevant, since the Bayesian view is that there is one set of data on which the inference is based, and it does not matter what other realisations of the data might have yielded. Nevertheless, Z is a statistic, and its sampling distribution could be of some interest. There is a contribution to the variability of Z from the data-only prefactor in equation (2.13) in the gaussian case (and we come to a similar conclusion in the more general score compression case: see section 2.3). The log of the prefactor is distributed as a χ^2 distribution with n degrees of freedom with a variance of $2n$. However, this main source of variability (the prefactor) cancels in nested models when the Bayes Factor is computed, so the quantity on which model comparison depends is not subject to this source of large variability.

This behaviour is apparent in figure B.1 of [19], where the Bayes Factor is very stable even though Z_0 and Z_1 are highly variable, which would invite the naïve interpretation that the sampling variance of the Evidence may hinder model comparison. However, even if the sampling distribution of $\ln Z_0$ and $\ln Z_1$ is large in comparison with $\ln B_{01}(\boldsymbol{X})$, the Bayes Factor favours one model consistently, because Z_0 and Z_1 are not independent, as they use the same data realisation.

We illustrate the behaviour of a non-nested case with the following toy non-linear models:

- $M_0: \ x = E\sin(Ft); \quad \boldsymbol{\theta}_0 = (E, F)$

- $M_1: \ x = at^2 + bt; \quad \boldsymbol{\theta}_1 = (a, b)$

where t has 20 values uniformly spaced between 0 and π (an example is shown in figure 1).

The data are generated with $E = 2.5$ and $F = 1.0$, and the fiducial values are $E_* = 3.0$, $F_* = 1.2, a_* = -1.5, b_* = 4.7$). The distribution of Bayesian Evidence values for each model
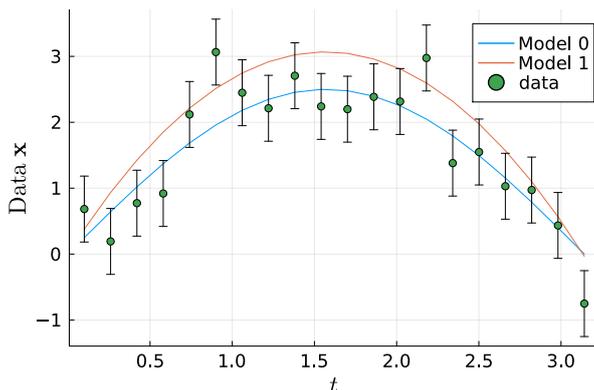
Figure 1. Example dataset, with noisy data drawn from the mean curve of Model 0. Parameters for Model 1 are arbitrary.
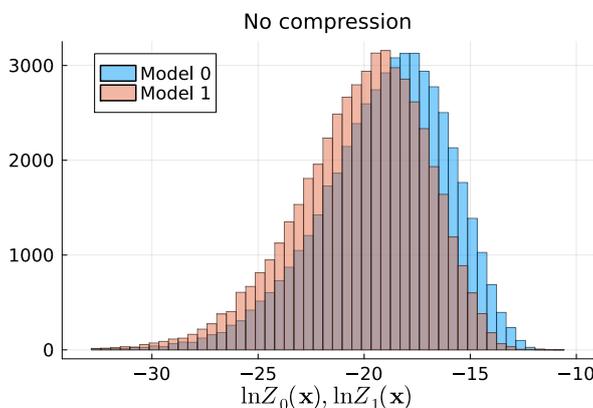
Figure 2. The distribution of Bayesian Evidence for 50000 noise realisations of non-nested toy models M_0 and M_1 with details given in the text, obtained from the full dataset. The normalisation of the y axis is arbitrary.

from 4×10^4 realisations is shown in figure 2 for the full dataset, and from MOPED-compressed data in figure 3. Notice that the considerable sample variance exceeds the difference of the mean Evidence values. As suggested by [19], the width of the sampling distribution of the compressed data Evidence is reduced compared to the full dataset.

Figure 4 shows that, although the Evidence values differ markedly for the full and compressed datasets, the Bayes Factors are almost identical, with deviations being due to the linear Taylor expansion not being perfect in the nonlinear case.

Our conclusion from this study is that MOPED compression reduces the variance of the sampling distribution of the Bayesian Evidence, so it would appear advantageous to use such a data compression. However, the Evidence values from an individual noise realisation are highly correlated, as they are based on the same data, and the log of the Bayes Factor has a much smaller variance than the variance of the individual $\ln Z$ values. We find that the uncompressed data give the same Bayes Factor as the MOPED compressed data (inasmuch as the neglect of higher-order derivatives of $\boldsymbol{\mu}$ is valid). As a result MOPED compression can be used for accurate model comparison, with in some cases a large increase in speed of
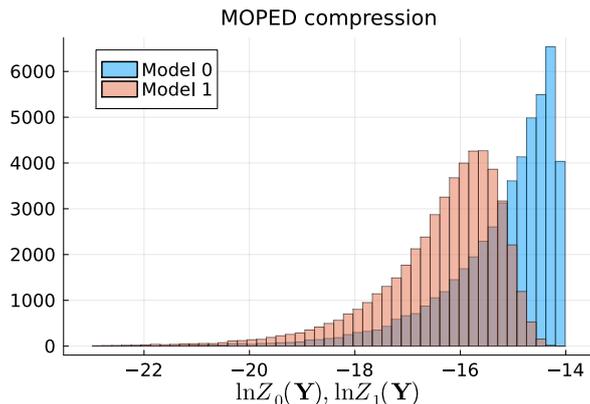
Figure 3. As figure 2, but for a compressed dataset consisting of the four MOPED coefficients, two from each model.
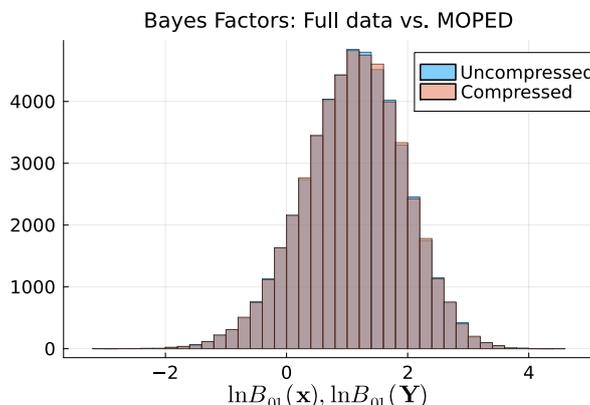
Figure 4. Comparison of Bayes Factor sampling distribution for the nonlinear, non-nested toy models M_0 and M_1 with details given in the text, from the full dataset and the MOPED-compressed dataset. The distributions overlay almost perfectly.

computation. We also note that with a single full likelihood evaluation and one compressed likelihood evaluation at the fiducial parameters, the full dataset Evidence can be computed, if it were ever needed, from the compressed Evidence:

$$Z(\boldsymbol{X}) = \frac{\sqrt{|2\pi\Lambda|}}{\sqrt{|2\pi\mathsf{C}|}} \frac{\exp(-\frac{1}{2}\boldsymbol{X}^T\mathsf{C}^{-1}\boldsymbol{X})}{\exp(-\frac{1}{2}\boldsymbol{Y}^T\Lambda^{-1}\boldsymbol{Y})} Z(\boldsymbol{Y}).$$

3.2 Cosmological example — the Pantheon+SH0ES dataset

In this section, we re-analyse the Pantheon+SH0ES data presented in [20]. This consists of 1701 supernovae, with Cepheid calibration. We use the distance moduli from the catalogue[1] (with a fiducial SNIa magnitude determined from SH0ES 2021 Cepheid host distances), and the supplied covariance matrix, including systematics, to define a simple gaussian likelihood function for 1590 supernovae with $z > 0.01$. While not nearly as sophisticated as a Bayesian hierarchical model [e.g. 41], this simplified setup serves to demonstrate parameter inference

---

[1]https://github.com/PantheonPlusSH0ES.

and Bayesian Evidence comparisons for the full and MOPED-compressed datasets. The two models being compared are:

- a flat universe, parameterised by $\boldsymbol{\theta}_0 = (\Omega_{\mathrm{m}}, h)$;

- a curved universe, which allows Ω_Λ to vary away from $1 - \Omega_{\mathrm{m}}$, with parameters $\boldsymbol{\theta}_1 = (\Omega_{\mathrm{m}}, h, \Omega_\Lambda)$,

where h is the Hubble-Lemaître constant in units of $100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and Ω_{m} is the matter density parameter, and Ω_Λ the vacuum energy density parameter. Priors on all varying parameters are taken to be uniform over the range $[0, 1]$. We choose a fiducial model with $\Omega_{\mathrm{m}} = 0.3$, $h = 0.7$ and $\Omega_\Lambda = 0.7$. The compression concatenates the MOPED coefficients from both models, but since they are nested, this is identical to using the MOPED coefficients from the extended, 3-parameter model, yielding

$$\mathbf{Y} = (1304.68,\ 23857.17,\ -1341.94). \tag{3.1}$$

All the cosmological information about Ω_{m}, h and Ω_Λ is contained in just these three numbers. The covariance matrix of these coefficients is

$$\mathrm{Cov}(\boldsymbol{Y}) = \begin{pmatrix} 3120.0 & 35028.9 & -2632.9 \\ 35028.9 & 668580.5 & -36364.1 \\ -2632.9 & -36364.1 & 2559.2 \end{pmatrix}. \tag{3.2}$$

We use the nested sampling code dynesty[2] [42] with 5000 live points, and no linear approximation is made for the parameter dependence of the mean distance modulus in the likelihood function. We find virtually identical posteriors for full and compressed data, shown in figure 5. The natural log of the Bayesian Evidence for the full dataset is 838.88 ± 0.035 for the flat model, and 837.96 ± 0.037 for the extended model. For the MOPED-compressed data, the log Bayesian Evidence values are -22.35 and -23.29, with the same errors. Hence the Bayes Factor are

$$\begin{array}{ll} -0.92 \pm 0.05; & \text{Full dataset (from 1590 numbers)} \\ -0.94 \pm 0.05; & \text{MOPED compression (from 3 numbers),} \end{array} \tag{3.3}$$

with in both cases the flat model being slightly favoured by the data. The error is an estimate from the values of about 0.036 for each, since we do not know the extent to which the errors are correlated.

Figure 5 and the Bayes Factors of eq. (3.3) show the strength of MOPED for both parameter inference and now for model comparison as well. The latter, along with the sampling distributions of the Bayes Factor in the compressed case (figure 4) are the main new results of this paper.
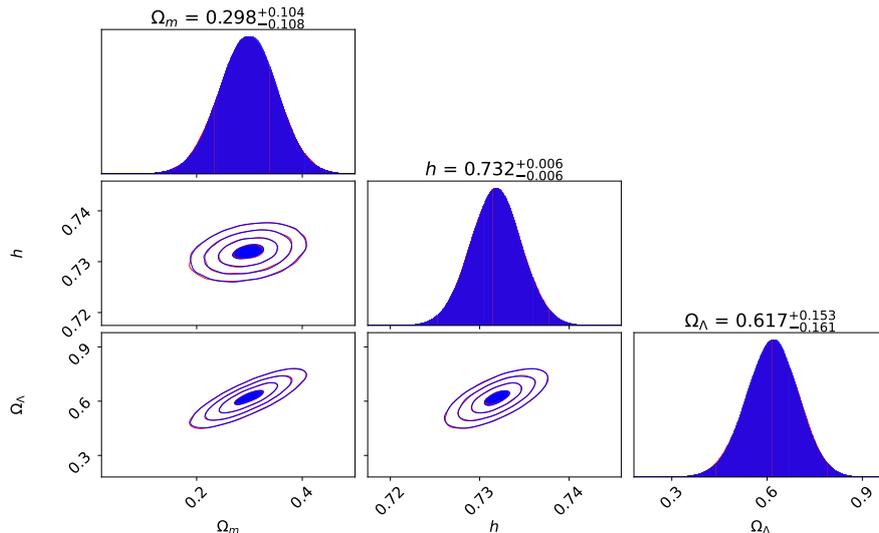
Figure 5. Marginal posteriors for cosmological parameters $(\Omega_{\mathrm{m}}, h, \Omega_{\Lambda})$, using the full dataset of size 1590 (red) and only the three MOPED-compressed coefficients (blue). The posteriors are almost identical, resulting in a purple shade in the 1D marginals.

4   Discussion and conclusions

In this paper we have shown how the extreme data compression algorithm MOPED [15] can be used to compute Bayes Factors with essentially no loss of accuracy, thus extending its power from parameter inference to now include model comparison. The key to this is to perform MOPED compression in all the model spaces, and to concatenate the MOPED coefficients into a slightly larger compressed data space, of dimension equal to the sum of the number of parameters if the models are not nested. This is still an extreme compression, if the number of data far exceed the number of parameters, which is typically the case with cosmological survey data. In the nested case the MOPED coefficients come from the more complex model. Many of the advantages extend to other forms of score compression when MOPED is not applicable.

The computation of Bayes Factors is generally expensive, requiring integrals over the parameter space or sophisticated nested sampling techniques, so where MOPED leads to much faster likelihood evaluations from the massively reduced dataset, the process may be significantly accelerated.

Finally, we have investigated the sampling distribution of the Bayesian Evidence and of Bayes Factors, for uncompressed and MOPED compressed data. We found that while extreme data compression reduces the variance of the Evidence for each individual model, as previously claimed, the sampling distribution for the Bayes Factor remains unaffected by extreme compression of linear models, and virtually identical for nonlinear models. Thus treating the Bayes Factor as a frequentist statistic, it is just as effective to compute it from the compressed dataset as the full set. In a cosmological application, we have shown that, assuming a simple gaussian data model for the Pantheon+SH0ES dataset [20], the Bayes Factor for the flat vs non-flat model can be computed from only 3 compressed numbers, rather than the full dataset of 1590 supernovae, with no loss of accuracy.

Acknowledgments

Data availability. No new data are used in this publication, except for those generated randomly by the processes described fully within.

References

[1] K. Kuijken et al., *The fourth data release of the Kilo-Degree Survey: ugri imaging and nine-band optical-IR photometry over 1000 square degrees*, *Astron. Astrophys.* **625** (2019) A2 [`arXiv:1902.11265`] [ɪɴSPIRE].

[2] DES collaboration, *Dark energy survey year 3 results: cosmological constraints from the analysis of cosmic shear in harmonic space*, *Mon. Not. Roy. Astron. Soc.* **515** (2022) 1942 [`arXiv:2203.07128`] [ɪɴSPIRE].

[3] R. Dalal et al., *Hyper Suprime-Cam Year 3 Results: Cosmology from Cosmic Shear Power Spectra*, `arXiv:2304.00701` [ɪɴSPIRE].

[4] PLANCK collaboration, *Planck 2018 results. I. Overview and the cosmological legacy of Planck*, *Astron. Astrophys.* **641** (2020) A1 [`arXiv:1807.06205`] [ɪɴSPIRE].

[5] EUCLID collaboration, *Euclid Definition Study Report*, `arXiv:1110.3193` [ɪɴSPIRE].

[6] A. Heavens, E. Sellentin, D. de Mijolla and A. Vianello, *Massive data compression for parameter-dependent covariance matrices*, *Mon. Not. Roy. Astron. Soc.* **472** (2017) 4244 [`arXiv:1707.06529`] [ɪɴSPIRE].

[7] T. Charnock, G. Lavaux and B.D. Wandelt, *Automatic physical inference with information maximizing neural networks*, *Phys. Rev. D* **97** (2018) 083004 [`arXiv:1802.03537`] [ɪɴSPIRE].

[8] F. Gerardi, S.M. Feeney and J. Alsing, *Unbiased likelihood-free inference of the Hubble constant from light standard sirens*, `arXiv:2104.02728` [`DOI:10.1103/PhysRevD.104.083531`] [ɪɴSPIRE].

[9] T.L. Makinen, T. Charnock, J. Alsing and B.D. Wandelt, *Lossless, scalable implicit likelihood inference for cosmological fields*, *JCAP* **11** (2021) 049 [*Erratum ibid.* **04** (2023) E02] [`arXiv:2107.07405`] [ɪɴSPIRE].

[10] N. Jeffrey, J. Alsing and F. Lanusse, *Likelihood-free inference with neural compression of DES SV weak lensing map statistics*, *Mon. Not. Roy. Astron. Soc.* **501** (2021) 954 [`arXiv:2009.08459`] [ɪɴSPIRE].

[11] T.L. Makinen et al., *The Cosmic Graph: Optimal Information Extraction from Large-Scale Structure using Catalogues*, arXiv:2207.05202 [DOI:10.21105/astro.2207.05202] [INSPIRE].

[12] K. Lin, M. von Wietersheim-Kramsta, B. Joachimi and S. Feeney, *A simulation-based inference pipeline for cosmic shear with the Kilo-Degree Survey*, arXiv:2212.04521 [INSPIRE].

[13] F. Leclercq, *Simulation-based inference of Bayesian hierarchical models while checking for model misspecification*, arXiv:2209.11057 [DOI:10.48550/ARXIV.2209.11057].

[14] A. Heavens, E. Sellentin and A. Jaffe, *Extreme data compression while searching for new physics*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 3440 [arXiv:2006.06706] [INSPIRE].

[15] A. Heavens, R. Jimenez and O. Lahav, *Massive lossless data compression and multiple parameter estimation from galaxy spectra*, *Mon. Not. Roy. Astron. Soc.* **317** (2000) 965 [astro-ph/9911102] [INSPIRE].

[16] M. Tegmark, A. Taylor and A. Heavens, *Karhunen-Loève eigenvalue problems in cosmology: How should we tackle large data sets?*, *Astrophys. J.* **480** (1997) 22 [astro-ph/9603021] [INSPIRE].

[17] J. Alsing and B. Wandelt, *Generalized massive optimal data compression*, *Mon. Not. Roy. Astron. Soc.* **476** (2018) L60 [arXiv:1712.00012] [INSPIRE].

[18] G. Jung et al., *Quijote-PNG: Quasi-maximum Likelihood Estimation of Primordial Non-Gaussianity in the Nonlinear Dark Matter Density Field*, *Astrophys. J.* **940** (2022) 71 [arXiv:2206.01624] [INSPIRE].

[19] B. Joachimi, F. Köhlinger, W. Handley and P. Lemos, *When tension is just a fluctuation: How noisy data affect model comparison*, *Astron. Astrophys.* **647** (2021) L5 [arXiv:2102.09547] [INSPIRE].

[20] D. Scolnic et al., *The Pantheon+ Analysis: The Full Data Set and Light-curve Release*, *Astrophys. J.* **938** (2022) 113 [arXiv:2112.03863] [INSPIRE].

[21] A. Heavens, B. Panter, R. Jimenez and J. Dunlop, *The Complete star formation history of the universe*, *Nature* **428** (2004) 625 [astro-ph/0403293] [INSPIRE].

[22] B. Panter, A.F. Heavens and R. Jimenez, *Star formation and metallicity history of the SDSS galaxy survey: Unlocking the fossil record*, *Mon. Not. Roy. Astron. Soc.* **343** (2003) 1145 [astro-ph/0211546] [INSPIRE].

[23] B. Panter, R. Jimenez, A.F. Heavens and S. Charlot, *The star formation histories of galaxies in the Sloan Digital Sky Survey*, *Mon. Not. Roy. Astron. Soc.* **378** (2007) 1550 [astro-ph/0608531] [INSPIRE].

[24] B. Panter, R. Jimenez, A.F. Heavens and S. Charlot, *The Cosmic Evolution of Metallicity from the SDSS Fossil Record*, *Mon. Not. Roy. Astron. Soc.* **391** (2008) 1117 [arXiv:0804.3091] [INSPIRE].

[25] S. Gupta and A.F. Heavens, *Fast parameter estimation from the CMB power spectrum*, *Mon. Not. Roy. Astron. Soc.* **334** (2002) 167 [astro-ph/0108315] [INSPIRE].

[26] A. Zablocki and S. Dodelson, *Extreme data compression for the CMB*, *Phys. Rev. D* **93** (2016) 083525 [arXiv:1512.00072] [INSPIRE].

[27] H. Prince and J. Dunkley, *Data compression in cosmology: A compressed likelihood for Planck data*, *Phys. Rev. D* **100** (2019) 083502 [arXiv:1909.05869] [INSPIRE].

[28] Y. Lai, C. Howlett and T.M. Davis, *Faster cosmological analysis with power spectrum without simulations*, arXiv:2306.00388 [INSPIRE].

[29] D. Gualdi et al., *Enhancing BOSS bispectrum cosmological constraints with maximal compression*, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 3713 [`arXiv:1806.02853`] [INSPIRE].

[30] P. Graff, M. Hobson and A. Lasenby, *An investigation into the Multiple Optimised Parameter Estimation and Data compression algorithm*, *Mon. Not. Roy. Astron. Soc.* **413** (2011) L66 [`arXiv:1010.5907`] [INSPIRE].

[31] L. Thiele, G.A. Marques, J. Liu and M. Shirasaki, *Cosmological constraints from HSC Y1 lensing convergence PDF*, `arXiv:2304.05928` [INSPIRE].

[32] R. Ruggeri and C. Blake, *Compressing combined probes: redshift weights for joint lensing and clustering analyses*, *Mon. Not. Roy. Astron. Soc.* **498** (2020) 2948 [`arXiv:2006.11297`] [INSPIRE].

[33] L. Wenzl et al., *Cosmology with the Roman Space Telescope — Synergies with CMB lensing*, *Mon. Not. Roy. Astron. Soc.* **512** (2022) 5311 [`arXiv:2112.07681`] [INSPIRE].

[34] P. Protopapas, R. Jimenez and C. Alcock, *Fast identification of transits from light-curves*, *Mon. Not. Roy. Astron. Soc.* **362** (2005) 460 [`astro-ph/0502301`] [INSPIRE].

[35] R. Trotta, *Bayes in the sky: Bayesian inference and model selection in cosmology*, *Contemp. Phys.* **49** (2008) 71 [`arXiv:0803.4089`] [INSPIRE].

[36] F. Leclercq, W. Enzi, J. Jasche and A. Heavens, *Primordial power spectrum and cosmology from black-box galaxy surveys*, *Mon. Not. Roy. Astron. Soc.* **490** (2019) 4237 [`arXiv:1902.10149`] [INSPIRE].

[37] C.R. Jenkins and J.A. Peacock, *The power of Bayesian evidence in astronomy*, *Mon. Not. Roy. Astron. Soc.* **413** (2011) 2895 [`arXiv:1101.4822`] [INSPIRE].

[38] R.L. Wasserstein and N.A. Lazar, *The ASA statement on p-values: Context, process, and purpose*, *Am. Stat.* **70** (2016) 129.

[39] H. Koo, R.E. Keeley, A. Shafieloo and B. L'Huillier, *Bayesian vs frequentist: comparing Bayesian model selection with a frequentist approach using the iterative smoothing method*, *JCAP* **03** (2022) 047 [`arXiv:2110.10977`] [INSPIRE].

[40] R.E. Keeley and A. Shafieloo, *On the distribution of Bayesian evidence*, *Mon. Not. Roy. Astron. Soc.* **515** (2022) 293 [`arXiv:2111.04231`] [INSPIRE].

[41] H. Shariff, X. Jiao, R. Trotta and D.A. van Dyk, *BAHAMAS: New Analysis of Type Ia Supernovae Reveals Inconsistencies with Standard Cosmology*, *Astrophys. J.* **827** (2016) 1 [`arXiv:1510.05954`] [INSPIRE].

[42] J.S. Speagle, *dynesty: a dynamic nested sampling package for estimating Bayesian posteriors and evidences*, *Mon. Not. Roy. Astron. Soc.* **493** (2020) 3132 [`arXiv:1904.02180`] [INSPIRE].