Panagiotis Papastamoulis[1] / Takanori Furukawa[2] / Norman van Rhijn[2] / Michael Bromley[2] / Elaine Bignell[2] / Magnus Rattray[3]

Bayesian Detection of Piecewise Linear Trends in Replicated Time-Series with Application to Growth Data Modelling

[1] Department of Statistics, School of Information Sciences and Technology, Athens University of Economics and Business, Patision 76, 104 34 Athens, Greece, E-mail: papastamoulis@aueb.gr. https://orcid.org/0000-0001-9468-7613.
[2] Division of Infection, Immunity & Respiratory Medicine, Faculty of Biology, Medicine & Health, University of Manchester, Manchester, UK
[3] Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Michael Smith Building, Oxford Road, Manchester, M13 9PL, UK

Abstract:
We consider the situation where a temporal process is composed of contiguous segments with differing slopes and replicated noise-corrupted time series measurements are observed. The unknown mean of the data generating process is modelled as a piecewise linear function of time with an unknown number of change-points. We develop a Bayesian approach to infer the joint posterior distribution of the number and position of change-points as well as the unknown mean parameters. A-priori, the proposed model uses an overfitting number of mean parameters but, conditionally on a set of change-points, only a subset of them influences the likelihood. An exponentially decreasing prior distribution on the number of change-points gives rise to a posterior distribution concentrating on sparse representations of the underlying sequence. A Metropolis-Hastings Markov chain Monte Carlo (MCMC) sampler is constructed for approximating the posterior distribution. Our method is benchmarked using simulated data and is applied to uncover differences in the dynamics of fungal growth from imaging time course data collected from different strains. The source code is available on CRAN.

Keywords: Change-point detection, Fungal growth data, Markov chain Monte Carlo
DOI: 10.1515/ijb-2018-0052
Received: May 24, 2018; **Revised:** May 17, 2019; **Accepted:** July 6, 2019

1 Introduction

In many applications a non-stationary time series consists of an unknown number of segments. The observed data is described by different statistical generative models within each segment. In such cases, the objective is to identify the number and position of change-points which give rise to different segments of the data as well as to infer all remaining parameters of the underlying statistical model. In principle, there are two approaches for answering these questions: online and offline segmentation [1], which refer to the task of inferring changes during or after the observation process, respectively. In this work, the latter scenario is considered.

We develop a Bayesian method for detecting an unknown number of change-points in the slope of multiple replicated time series. There are many methods for detecting change-points, with the majority of them focused on analysis of univariate time series (reviewed below). Our problem differs from these as we model changes in slope rather than fitting a step function to the mean, and we consider multiple time series with replication. For a given period $t = 1, \ldots, T$ we observe multiple time series which are assumed independent, each one consisting of multiple measurements (replicates). Each time series is assumed to have its own segmentation, which is common among its replicates. Thus, different time series have distinct mean structures in the underlying normal distribution.

Our method is motivated by the need to analyse fungal growth attributes on a massively parallel scale. Specifically, we are interested in identifying mutations which affect fitness in the major human fungal pathogen *Aspergillus fumigatus*. In such studies, growth is characterized by different phases which can be reasonably described by a piecewise linear model, as illustrated in Figure 1. Microbial growth is a complex characteristic

Panagiotis Papastamoulis is the corresponding author.

which is heavily influenced by nutritional, metabolic, proliferative, physiological and genetic factors. Multiple techniques have been developed with which to quantify microbial growth, including direct quantitation of cell counts using flow cytometry or microscopy, colony counts, biomass quantitation, or indirect methods involving light scattering or turbidity measurement in liquid phase cultures, or dye-based methods. Optimisation of data acquisition and analysis has received rather less attention, particularly where the quantitation of growth characteristics in filamentous and aggregative microorganisms, such as *A. fumigatus* or *Streptomyces coelicolor* is complicated by the occurrence of one or several morphological shifts during the mitotic life cycle [2] leading to altered light scattering patterns dependent upon the size and shape of the particulate sample (bacteria or yeast), as well as difference in the index of refraction between the particles and the culture media [3]. In the latter instance an accurate means of defining the number and timing of change-points during growth curve analysis would significantly empower the optimisation of drug discovery screens where inhibitors of microbial growth might be sought; or in optimisation of biotechnological processes where moderation of microbial growth conditions to favour a particular growth phase might boost industrial production of enzymes or metabolites.

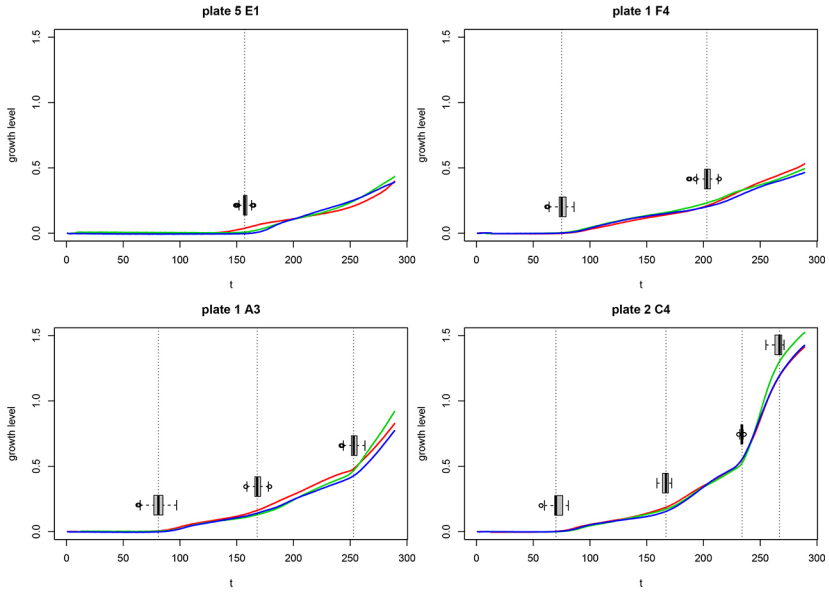


Figure 1: A subset of four growth data series described in Section 4.2, consisting of the growth level of three replicates (red, green and blue), measured every 10 minutes for a series of $T = 289$ time-points. The boxplots display the marginal posterior distribution of each change-point, conditionally on the inferred Maximum A Posteriori number of change-points according to the MCMC sampler detailed in Section 3.1.

In the seminal paper of Green [4], the Reversible Jump MCMC (RJMCMC) algorithm used to detect the number of change-points in coal mining disaster data. Subsequently, the RJMCMC methodology was applied to a variety of change-point detection problems [5–8]. Lavielle and Lebarbier [9] proposed an MCMC sampler to estimate the number of change-points by introducing a latent sequence of independent and identically distributed Bernoulli random variables r_t; $t = 1, \ldots, T$, with T denoting the number of time-points. In this context, $r_j = 1$ indicates that a change occurs at time $t = j$, while $r_j = 0$ means that no change occurs. This approach has the advantage that it can infer the target posterior distribution using an MCMC sampler that operates on random variables of constant dimension, in contrast to the RJMCMC approach. However, it turns out that the estimated marginal posterior probabilities of these artificial binary random variables overestimates the true number of change-points. We illustrate that similar issues arise in our set-up when imposing typical prior assumptions on the number of change-points, such as a Poisson distribution. To overcome this problem, Lavielle and Lebarbier [9] inferred configurations of change-points of high probability by sampling from a modified posterior distribution which is a tempered version of the original target, using a simulated annealing MCMC algorithm. In our set-up, we demonstrate that we are able to accurately infer the number of change-points when using priors that heavily penalize large values of change-points [10].

Chib [11] formulates the change-point model in terms of a latent discrete state variable corresponding to the regime from which a particular observation has been drawn. The posterior distribution for a given number of change-points is then approximated using MCMC sampling, while inference on the number of change-points is carried out by estimating the marginal likelihood of the model using the method in Chib [12]. Fearnhead [13] discusses exact Bayesian inference by assuming that the joint posterior distribution of the parameters is independent across the segments of the time series and also presents an extension that allows the signal to

perform a random walk within each segment. Other Bayesian methods include Dobigeon et al. [14], Hutter [16], Kim and Cheon [17], Rudoy et al. [18], Schütz and Holschneider [19], He [15], Schwaller and Robin [20]. In all of the aforementioned studies, change-points are defined via a step-function in the mean, assuming that segment parameters are independent. This is not the case in our approach, where we are explicitly imposing a continuity assumption between segments which allows us to detect changes in slope. With the exception of Dobigeon et al. [14], Schwaller and Robin [20], all other methods are focusing on univariate time-series.

There are relatively few studies looking specifically at a change-in-slope model [21] [22–25], although they only consider a single time-series. Popular non-Bayesian methods such as binary segmentation [26, 27] do not work for detecting changes-in-slope [22]. Furthermore, standard dynamic programming approaches [28, 29] cannot be directly applied to our problem as discussed by Fearnhead et al. [24].

There is a wide range of non-Bayesian approaches to change-point estimation, see for example Halpern [32], Lu et al. [33], Picard et al. [34], Yildirim [35], Chamroukhi et al. [30], Frick et al. [31]. However, we choose a Bayesian approach for its competency in quantifying uncertainty and flexibility for incorporating prior information.

We construct a Metropolis-Hastings MCMC sampler [36, 37] for jointly inferring the number and position of change-points as well as the related mean parameters by adopting ideas from inference over sparse representations of sequences [10]. An advantage of our approach is that the proposed MCMC algorithm is straightforward to implement since it is based on standard Metropolis–Hastings move types and demands small modelling effort compared to other methods. Although exact integration is possible given the number and locations of change-points, it is time consuming hence we exploit the convenience of the MCMC sampler to approximately sample from the joint posterior distribution of the parameters. The dimensionality of the parameter space is fixed, thus our method avoids the complex step of designing trans–dimensional MCMC transitions as required by RJMCMC methods. Furthermore, we do not have to consider modified versions of the target posterior distribution [9], and there is no requirement for fitting the same model under different number of change-points and approximating the marginal likelihood for model selection.

The rest of the paper is organised as follows. Section 2 introduces the proposed model and the corresponding prior assumptions are presented in Section 2.1. The main MCMC sampler we use is detailed in Section 3.1. Sections 3.2 and 3.3 discuss variance estimation procedures, depending whether the variance is treated as a known parameter or not. The proposed method is illustrated in simulated and real data in Sections 4.1 and 4.2, respectively. The paper concludes in Section 5. More details on the simulation procedure and additional results are given in the Appendix.

2 Model

Let X_{ntr} denote a random variable describing replicate r at time point t for time series n, $n = 1, ..., N$; $t = 1, ..., T$; $r = 1, ..., R$. It is assumed that $\{X_{ntr}; r = 1, \dots, R\}$ is a normally distributed random sample and furthermore that measurements are independent across time, that is:

$$X_{ntr} \sim \mathcal{N}\left(\theta_{nt}, \sigma^2_{nt}\right), \quad \theta_{nt} \in \mathbb{R}, \sigma^2_{nt} > 0 \tag{1}$$

independent (given the model parameters) for $n = 1, ..., N$; $t = 1, ..., T$; $r = 1, ..., R$, where $\mathcal{N}(\cdot, \cdot)$ denotes the normal distribution. At first, we will consider that the variances $\{\sigma^2_{nt}, n = 1, \dots, N; t = 1, \dots, n\}$ are known. In practice, the variance per time-point is estimated at a pre-processing stage as exemplified in Section 3.2. The general case where the variance is treated as a random variable is addressed in Section 3.3. Without any further assumptions, the parameterization of the normal distributions in eq. (1) introduces a large number of mean parameters: a distinct mean parameter $\theta_{nt} \in \Theta = \mathbb{R}$ is assigned to each sample (n) and time-point (t). However, θ_{nt} is shared across replicates (r).

For sample n and an unknown non-negative integer $\ell_n \geqslant 0$, assume that there are $\ell_n + 1$ underlying phases of mean behaviour, identified by the ordered time points

$$1 < \tau_{n1} < \tau_{n2} < ... < \tau_{n\ell_n} < T. \tag{2}$$

Note that both the elements as well as the length of the ordered ℓ_n-tuple $\boldsymbol{\tau}_n = \left(\tau_{n1}, \dots, \tau_{n\ell_n}\right)$ depends on n. Assume that for each phase the mean function is linear in time. For phase $j = 1, \dots, \ell_n + 1$, the piecewise linear mean measurement levels are defined as follows:

$$\mu\left(t; \boldsymbol{\theta}_n, \boldsymbol{\tau}_n\right) = \theta_{n\tau_{n;j-1}} + \frac{\theta_{n\tau_{nj}} - \theta_{n\tau_{n;j-1}}}{\tau_{nj} - \tau_{n;j-1}}\left(t - \tau_{n;j-1}\right), \quad \tau_{n;j-1} \leqslant t \leqslant \tau_{nj}, \tag{3}$$

where we also define $\tau_{n0} := 1$ and $\tau_{n;\ell_n+1} := T$; $\forall n = 1, ..., N$. Note that eq. (3) imposes continuity in the segment means, which is a crucial difference with other approaches (mentioned in the Introduction): the mean for the end of one segment should be equal to the mean at the start of the next segment.

Thus, for sample n, conditionally on $\boldsymbol{\tau}_n$, we can write that

$$X_{ntr}|\boldsymbol{\tau}_n \sim \mathcal{N}\left(\mu\left(t;\boldsymbol{\theta}_n,\boldsymbol{\tau}_n\right),\sigma^2_{nt}\right), \tag{4}$$

independent for $n = 1, ..., N$; $t = 1, ..., T$; $r = 1, ..., R$. Note that given $\boldsymbol{\tau}_n$, the likelihood depends on $\boldsymbol{\theta}$ only through the subset

$$\boldsymbol{\theta}_{\boldsymbol{\tau}_n} \quad := \quad \left\{\theta_{n1}, \theta_{n\tau_{n1}}, \ldots, \theta_{n\tau_{n\ell_n}}, \theta_{nT}\right\}. \tag{5}$$

To be precise, the likelihood is defined as

$$f\left(\boldsymbol{x}_n|\boldsymbol{\theta}_{\boldsymbol{\tau}_n},\boldsymbol{\sigma}^2_n\right) = \left[\prod_{r=1}^{R}\varphi\left(x_{n1r};\theta_{n1},\sigma^2_{n1}\right)\right]\prod_{j=0}^{\ell_n}\prod_{t=\tau_{nj}+1}^{\tau_{n;j+1}}\prod_{r=1}^{R}\varphi\left(x_{ntr};\mu(t;\boldsymbol{\theta}_n,\boldsymbol{\tau}_n),\sigma^2_{nt}\right), \tag{6}$$

where $\varphi(\cdot;\mu,\sigma^2)$ denotes the probability density function of the normal distribution with mean μ and variance σ^2 and $\mu(t;\boldsymbol{\theta}_n,\boldsymbol{\tau}_n)$ defined in eq. (3).

2.1 Prior assumptions

For the mean parameters we assume that

$$\theta_{nt} \sim \mathcal{N}\left(\mu_{0t}, \frac{\sigma^2_{nt}}{\nu_0}\right) \tag{7}$$

independent for $n = 1, ..., N$; $t = 1, ..., T$. The quantities $\nu_0 > 0$ and $\mu_{0t} \in \mathbb{R}$; $t = 1, ..., T$, correspond to fixed hyper-parameters. The following default values are considered: $\mu_{0t} = \frac{1}{NR}\sum_{n=1}^{N}\sum_{r=1}^{R} x_{ntr}$, for $t = 1, ..., T$. It is suggested that the parameter ν_0 should be sufficiently small so that the prior distribution in (7) has large variability around the global mean. This is particularly important in cases where the multiple time series of an experiment exhibit strong heterogeneity. Values between $5 \times 10^{-2} \leqslant \nu_0 \leqslant 5 \times 10^{-1}$ performed reasonably well in our setup. The variance σ^2_{nt} may be fixed (see Section 3.2) or not (see Section 3.3).

In order to specify the prior distribution of locations for a given number of change-points (ℓ_n), we are taking into account the prior assumption that later time-points are more likely to contain changes than earlier time-points. The growth levels of ℓ_n consecutive time-points during the end of the observation period are more likely to break collinearity than earlier stages. For example, all time-series consist of an initial period with very small and almost constant growth level in which no change is expected to occur. On the other hand, the last part of the observation period may exhibit larger heterogeneity. A simple and efficient way to incorporate such a prior information while supporting (with positive probability) all possible configurations with respect to constraint (2) is the following.

Let $g(i;i_1,i_2) = \frac{1}{i_2-i_1+1}\mathbb{I}(i_1 \leqslant i \leqslant i_2)$ denote the probability mass function of the discrete uniform distribution defined over the finite set of integers i such that $i_1 \leqslant i \leqslant i_2$, where $\mathbb{I}(\cdot)$ denotes the indicator function. For $j = 1$ we assume that $\tau_{n1} \sim g(\cdot;2,T-\ell_n)$. For $j \geqslant 2$ and conditionally on the event $(\tau_{n1} = t_1, \ldots, \tau_{n;j-1} = t_{j-1})$, we assume that τ_{nj} follows a (discrete) uniform prior distribution defined over $t_{j-1}+1 \leqslant \tau_{nj} \leqslant T-\ell_n+j-1$. Thus, the prior distribution $f\left(\boldsymbol{\tau}_n|\ell_n > 0\right)$ for a specific realization $(t_1, \ldots, t_{\ell_n})$ of $\boldsymbol{\tau}_n$ is defined as

$$\begin{aligned} f(t_1, \ldots, t_{\ell_n}|\ell_n > 0) &= \mathbb{P}\left(\tau_{n1} = t_1, \ldots, \tau_{n\ell_n} = t_{\ell_n}|\ell_n > 0\right) \\ &= \mathbb{P}(\tau_{n1} = t_1)\prod_{j=2}^{\ell_n}\mathbb{P}(\tau_{nj} = t_j|\tau_{n;j-1} = t_{j-1}, \ldots, \tau_{n1} = t_1) \\ &= g(t_1;2,T-\ell_n)\prod_{j=2}^{\ell_n} g(t_j;t_{j-1}+1,T-\ell_n+j-1) \\ &= \tfrac{1}{T-\ell_n-1}\prod_{j=2}^{\ell_n}\tfrac{1}{T-\ell_n+j-t_{j-1}-1}\mathbb{I}(1 < t_1 < \ldots < t_{\ell_n} < T). \end{aligned} \tag{8}$$

Note that according to eq. (8), $\mathbb{P}(\tau_{n1} = T-\ell_n, \tau_{n2} = T-\ell_n+1, \ldots, \tau_{n\ell_n} = T-1|\ell_n) \approx 1/T$, while $\mathbb{P}(\tau_{n1} = 2, \tau_{n2} = 3, \ldots, \tau_{n\ell_n} = \ell_n+1|\ell_n) \approx 1/T^{\ell_n}$, which satisfies our prior expectation discussed in the previous paragraph. We

assume a-priori independence of $\boldsymbol{\tau}_n$ for $n = 1, \ldots, N$. Obviously, $\boldsymbol{\tau}_n$ makes sense only in the case that the total number of change-points is strictly positive, thus eq. (8) is defined conditionally on the event $\ell_n > 0$.

Finally, the prior distribution of the number of change-points should be defined. Recall that the distribution in eq. (1) assigns a distinct mean parameter per time-point. However, given ℓ_n, only a small subset of θ's will influence the likelihood in eq. (4) and the rest of them will affect the posterior solely due to their contribution to the prior distribution. Our motivation is based on the fact that we are trying to find very minimal models with few change points because it makes the data easier to interpret.

For a given number of change-points (ℓ), the number of possible models is $\binom{T^*}{\ell}$, a number which grows rapidly with ℓ, where $T^* = T - 2$. In order to penalize model complexity we consider prior distributions which favour sparse configurations. A natural way to incorporate such information on our prior distribution is to assume that the prior probability of ℓ change-points is inversely proportional to $\binom{T^*}{\ell}$, that is, the number of models of size ℓ. This approach is also justified by the work of [10], where they consider that the number of change-points follows an *exponentially decreasing* prior distribution.

The prior $\mathbb{P}(\cdot)$ has exponential decrease if, for some constants $C > 0$ and $D < 1$,

$$\mathbb{P}(\ell_n = \ell) \leqslant D\mathbb{P}(\ell_n = \ell - 1), \quad \ell > C\ell_*, \tag{9}$$

where ℓ_* denotes the true value of ℓ_n. In the context of multivariate normal mean models with an underlying sparse true mean vector, it has been shown [10] that asymptotically, priors satisfying (9) lead to posterior distributions that concentrate on the sparse underlying true generative model. Members of the family of the so-called "complexity priors", defined as,

$$f(\ell) = \mathbb{P}(\ell_n = \ell) \propto e^{-a\ell \log(bT^*/\ell)}, \quad a, b > 0; \ell = 0, 1, 2, \ldots \tag{10}$$

have exponential decrease (9) for $b > 1 + e$ [10]. In our applications we consider the choices $b = 3.72$ and $\alpha = 2$ as default values, but we also report various prior sensitivity checks in the Appendix. As noted by Castillo and van der Vaart [10] it holds that $e^{\ell \log(T^*/\ell)} \leqslant \binom{T^*}{\ell} \leqslant e^{\ell \log(eT^*/\ell)}$ implying that (10) is inversely proportional to the number of models of size ℓ. Thus, this choice is suited to the purpose of penalizing model complexity. Note that the right hand side of eq. (10) for $\ell = 0$ should be perceived as the limit $\lim_{\ell \downarrow 0} e^{-a\ell \log(bT^*/\ell)} = 1$.

Finally, we mention that typical prior assumptions on the number of change-points (for example a truncated Poisson or a uniform distribution over a pre-specified set of non-negative integer values) tend to overfit the number of change-points. This behaviour is demonstrated in Section 4.1 using simulated data with a known number of change-points, as well as in Section C of the Appendix using our real dataset.

3 Inference

3.1 Metropolis–Hastings MCMC Sampler

Assume first that the variances $\boldsymbol{\sigma} = \{\sigma^2_{nt}; n = 1, \ldots, N, t = 1, \ldots, T\}$ are given. This assumption will be relaxed at the end of this section. Observe that conditionally on the vector $\boldsymbol{\sigma}^2$, $\{(\boldsymbol{\theta}_n, \boldsymbol{\ell}_n, \boldsymbol{\tau}_n); n = 1, \ldots, N\}$ are a-posteriori independent. Therefore, the inferential procedure breaks down to N independent tasks. The posterior distribution is written as

$$f\left(\boldsymbol{\theta}, \boldsymbol{\ell}, \boldsymbol{\tau} | \boldsymbol{x}, \boldsymbol{\sigma}^2\right) = \prod_{n=1}^{N} f\left(\boldsymbol{\theta}_n, \ell_n, \boldsymbol{\tau}_n | \boldsymbol{x}_n, \boldsymbol{\sigma}^2_n\right) \tag{11}$$

$$\propto \prod_{n=1}^{N} f\left(\boldsymbol{x}_n | \boldsymbol{\theta}_{\boldsymbol{\tau}_n}, \boldsymbol{\sigma}^2\right) f\left(\boldsymbol{\theta}_n | \boldsymbol{\sigma}^2\right) f(\boldsymbol{\tau}_n | \ell_n) f(\ell_n). \tag{12}$$

Although analytical evaluation of the marginal posterior distribution $f(\boldsymbol{\tau}, \boldsymbol{\ell} | \boldsymbol{x}, \boldsymbol{\sigma}^2) = \int_\Theta f\left(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\ell} | \boldsymbol{x}, \boldsymbol{\sigma}^2\right) \mathrm{d}\boldsymbol{\theta}$ is possible since we use conjugate prior assumptions, the integration cannot be carried out independently within each segment due to the continuity constraint. This fact makes the use of standard dynamic programming approaches not applicable in our set-up and the discrete nature of the sampling problem will make the computation of the involved expressions a time consuming task since the possible combinations of change-points increases rapidly with T. Thus, in order to make inference we sample from the target posterior distribution using a Metropolis-Hastings MCMC sampler that updates $(\boldsymbol{\theta}, \boldsymbol{\ell}, \boldsymbol{\tau})$.

At each step, the state of the chain is updated using four move-types: move 1 updates the number of change-points, move 2 updates the mean parameters by using a random walk proposal centered at the current values, move 3 updates the position of change-points and move 4 updates the subset of mean parameters that are not

allocated to a change-point. In each case the proposed move is accepted according to the usual Metropolis-Hastings acceptance ratio, that is, $R_n = \min\{1, \alpha_n\}$ where

$$\alpha_n = \frac{f\left(\boldsymbol{\theta}'_n, \ell'_n, \boldsymbol{\tau}'_n | \boldsymbol{x}_n, \boldsymbol{\sigma}^2_n\right) \mathrm{P}_{\mathrm{prop}}\left(\{\boldsymbol{\theta}'_n, \ell'_n, \boldsymbol{\tau}'_n\} \to \left\{\boldsymbol{\theta}_n^{(m)}, \ell_n^{(m)}, \boldsymbol{\tau}_n^{(m)}\right\}\right)}{f\left(\boldsymbol{\theta}_n, \ell_n, \boldsymbol{\tau}_n | \boldsymbol{x}_n, \boldsymbol{\sigma}^2_n\right) \mathrm{P}_{\mathrm{prop}}\left(\{\boldsymbol{\theta}_n, \ell_n, \boldsymbol{\tau}_n\} \to \{\boldsymbol{\theta}'_n, \ell'_n, \boldsymbol{\tau}'_n\}\right)}. \tag{13}$$

In eq. (13), $(\boldsymbol{\theta}_n, \ell_n, \boldsymbol{\tau}_n)$ denotes the current state of the n-th chain and $(\boldsymbol{\theta}'_n, \ell'_n, \boldsymbol{\tau}'_n)$ denotes the candidate state. Moreover, we use the notation $\mathrm{P}_{\mathrm{prop}}(x \to y)$ to denote the probability of proposing state y when the current state of the chain is x.

Move 1 This move updates the number of change-points, while keeping the mean parameters constant. We introduce two move types which propose to update the total number of change-points by 1. These move-types are complementary to each other: addition/deletion of a change-point. In the following, $\{\boldsymbol{\tau}_n \cup t\}$ denotes the resulting ordered set when a new change-point t is added to the current configuration $\boldsymbol{\tau}_n$. In a similar fashion, $\{\boldsymbol{\tau}_n \backslash t\}$ denotes the remaining set when a specific member t of $\boldsymbol{\tau}_n$ is removed from the c urrent configuration.

At a given state consisting of ℓ_n change-points, we propose addition/deletion with probabilities $p_a(\ell_n)$ and $p_d(\ell_n) = 1 - p_a(\ell_n)$, respectively. The addition probabilities are defined as

$$p_a(\ell_n) = \begin{cases} 1, & \ell_n = 0 \\ 1/2, & 1 \leqslant \ell_n \leqslant L-1 \\ 0, & \ell_n = L, \end{cases} \tag{14}$$

where L denotes the maximum number of change-points ($L \leqslant T-2$). In case of addition, we propose to add a randomly drawn change-point between two successive ones. The probability of proposing the addition of change-point t_* such that $\tau_{n;j-1} < t_* < \tau_{nj}$; for some $j = 1, \ldots, \ell_n + 1$, is equal to $p_a(\ell_n) \frac{1}{\tau_{nj} - \tau_{n;j-1} - 1} \frac{1}{\ell_n + 1}$. In case that $\tau_{nj} - \tau_{n;j-1} = 1$ the proposed move is immediately rejected. In the reverse move, a the previously added change-point is selected with probability $p_d(\ell_n + 1) \frac{1}{\ell_n + 1}$ and is deleted from $\boldsymbol{\tau}_n \cup t_*$. Thus, the acceptace probability for an addition move is equal to

$$\alpha_a(\ell_n, \boldsymbol{\theta}_n, \boldsymbol{\tau}_n, \boldsymbol{\tau}'_n) := \frac{f\left(\boldsymbol{x}_n | \boldsymbol{\theta}_{\{\boldsymbol{\tau}_n \cup t_*\}}, \boldsymbol{\sigma}^2_n\right) f\left(\{\boldsymbol{\tau}_n \cup t_*\} | \ell_n + 1\right) (1 - p_a(\ell_n + 1))}{f\left(\boldsymbol{x}_n | \boldsymbol{\theta}_{\boldsymbol{\tau}_n}, \boldsymbol{\sigma}^2_n\right) f\left(\boldsymbol{\tau}_n | \ell_n\right) \frac{p_a(\ell_n)}{\tau_{nj} - \tau_{n;j-1} - 1}}, \tag{15}$$

In the case of proposing deletion of a change-point τ_{nj}, the corresponding acceptance ratio term is equal to

$$\alpha_d(\ell_n, \boldsymbol{\theta}_n, \boldsymbol{\tau}_n, \boldsymbol{\tau}'_n) = \frac{1}{\alpha_a\left(\ell_n - 1, \boldsymbol{\theta}_n, \{\boldsymbol{\tau}_n \backslash \tau_{nj}\}, \boldsymbol{\tau}_n\right)}. \tag{16}$$

At this point we underline that using an overfitted set of model parameters (one mean θ_{nt} per time-point $t = 1, \ldots, T$) allows us to use the standard Metropolis-Hastings ratio for proposing additions/deletions of change-points. This would not be true if the number of mean parameters was defined conditionally on ℓ_n: in such a case the Reversible Jump algorithm or integration of mean parameters is required.

Move 2 In this move the update of mean parameters $\boldsymbol{\theta}$ is proposed, while all other parameters remain unchanged. For this purpose a random walk centered at the current values of the chain is used. For subject $n = 1, \ldots, N$, let θ_{nt} denote the current value of the mean parameters at time-point $t = 1, \ldots, T$. Then a new state is proposed according to

$$\theta'_{nt} \sim \mathcal{N}\left(\theta_{nt}, c\sigma^2_{nt}\right),$$

independent for all t and n, for some constant $c > 0$. Recall that σ^2_{nt} denotes the variance of the random sample $(X_{nt1}, \ldots, X_{ntR})$ which is assumed known. Note that the ratio of the proposal distribution for the transitions $\boldsymbol{\theta}_n \to \boldsymbol{\theta}'_n$ and $\boldsymbol{\theta}'_n \to \boldsymbol{\theta}_n$ is 1. Thus, the Metropolis-Hastings acceptance ratio (13) simplifies to

$$\alpha_4(\boldsymbol{\tau}_n, \boldsymbol{\theta}_n, \boldsymbol{\theta}'_n) = \frac{f\left(\boldsymbol{x}_n | \boldsymbol{\theta}'_{\boldsymbol{\tau}_n}, \boldsymbol{\sigma}^2_n\right) f\left(\boldsymbol{\theta}'_n | \boldsymbol{\sigma}^2_n\right)}{f\left(\boldsymbol{x}_n | \boldsymbol{\theta}_{\boldsymbol{\tau}_n}, \boldsymbol{\sigma}^2_n\right) f\left(\boldsymbol{\theta}_n | \boldsymbol{\sigma}^2_n\right)}, \tag{17}$$

Move 3.a The candidate state is generated by using a proposal distribution which will jointly update the change-points $\boldsymbol{\tau}_n$, while the total number of change-points ℓ_n and mean parameters $\boldsymbol{\theta}_n$ are kept constant. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_{\ell_n})$ and

$$\varepsilon_i \sim g(\cdot; -d_1, d_1),$$

where $d_1 > 0$ denotes a pre-specified positive integer and $g(\cdot; -d_1, d_1)$ denotes the discrete uniform distribution over $\{-d_1, -d_1+1, \ldots, d_1-1, d_1\}$. Then, the proposed state is generated as

$$\tau'_{ni} = \tau_{ni} + \varepsilon_i$$

independently for $i = 1, \ldots, \ell_n$, while $\ell'_n = \ell_n$ as well as $\boldsymbol{\theta}'_n = \boldsymbol{\theta}_n$. In this case, the proposal ratio in eq. (13) is equal to 1 so the acceptance ratio is written as the posterior probability ratio.

$$\alpha_1(\ell_n, \boldsymbol{\theta}_n, \boldsymbol{\tau}_n, \boldsymbol{\tau}'_n) := \frac{f\left(\boldsymbol{x}_n | \boldsymbol{\theta}_{\boldsymbol{\tau}'_n}, \boldsymbol{\sigma}_n^2\right) f\left(\boldsymbol{\tau}'_n | \ell_n\right)}{f\left(\boldsymbol{x}_n | \boldsymbol{\theta}_{\boldsymbol{\tau}_n}, \boldsymbol{\sigma}_n^2\right) f\left(\boldsymbol{\tau}_n | \ell_n\right)}, \tag{18}$$

where $\boldsymbol{\theta}_{\boldsymbol{\tau}_n}$ as in (5). A small value for d_1 will be capable of achieving optimal acceptance rates. Thus, this move is oriented towards the local exploration of the posterior surface, given the current state.

Move 3.b This is a similar proposal to Move 3.b, but instead of proposing the simultaneous update of all cut-points, just one entry is modified and the rest remain the same. As in Move 3.a, both the number of change-points as well as the values of mean parameters remain the same. Thus, let i^* denote a randomly drawn index from the set $\{1, \ldots, \ell_n\}$ and

$$\varepsilon \sim g(\cdot; -d_2, d_2),$$

where $d_2 > 0$ denotes a pre-specified positive integer. The proposed state is generated as

$$\tau'_{ni} = \begin{cases} \tau_{ni}, & i \neq i^* \\ \tau_{ni} + \varepsilon, & i = i^*. \end{cases} \tag{19}$$

The Metropolis-Hastings acceptance probability simplifies to eq. (18). In this case, a sufficiently large value for d_2 will propose moves that are more likely to be accepted compared to Move 3.a, since only one entry is changed. Thus, move 3.b will be used as complementary to move 3.a, in order to facilitate the ability of escaping from local modes of the posterior distribution.

Move 4 Let $\boldsymbol{\theta}_{n[-\boldsymbol{\tau}_n]}$ denote the mean of those time-points that do not correspond to change-points for time series n. In this case it can be easily seen that the full conditional distribution of $(\boldsymbol{\theta}_{n[-\boldsymbol{\tau}_n]} | \boldsymbol{\tau}_n, \boldsymbol{x}_n, \ell_n)$ is the prior distribution in eq. (7). Hence, a draw from the prior distribution will perfom a Gibbs update to $\boldsymbol{\theta}_{n[-\boldsymbol{\tau}_n]}$.

Note that both moves 3.a and 3.b are able to propose states that have zero prior probability in eq. (8). Although this is not a frequent event, in this case the proposed state is immediately rejected, since the prior probability ratio is equal to zero.

Recall that the previous MCMC steps are defined conditionally on $\boldsymbol{\sigma}^2$. The next subsection deals with the case where the variance is estimated at a pre-processing stage and plugged into the previously described MCMC sampler. For this purpose, we consider that the variance can be either shared between different time series or not and two estimators are derived. We assume that we have enough data in order to obtain a robust estimate of variance, which is the case in our application. The more general case where the variance is treated as a random variable is discussed in Subsection 3.3. In this case, the variance is updated by the full conditional distribution using a Gibbs sampling step.

3.2 Variance estimation at a pre-processing stage

In this case the variance is considered known and in practice it should be estimated at a pre-processing stage. We use the posterior mean arising from a multivariate normal–inverse gamma model as a point estimate. For this purpose we ignore the piecewise linear parameterization of the mean function and use the same likelihood as in eq. (1) and the same prior assumptions for θ_{nt} as in eq. (7).

We will assume two parameterizations: the full model where the variances are a-priori distributed as:

$$\sigma_{nt}^2 \sim \mathcal{IG}(\alpha_0, \beta_0), \tag{20}$$

independent for $n = 1, \ldots, N$; $t = 1, \ldots, T$, where $\mathcal{IG}(\alpha, \beta)$ denotes the inverse Gamma distribution. The second model parameterization imposes the restriction of common variance across different time series and replicates, that is,

$$\sigma_{nt}^2 = \sigma_t^2 \quad n = 1, \ldots, N. \tag{21}$$

Under (21), a-priori it is assumed that

$$\sigma_t^2 \sim \mathcal{IG}(\alpha_0, \beta_0),$$

independent for $t = 1, ..., T$. The quantities $\alpha_0 > 0$ and $\beta_0 > 0$ correspond to fixed hyper-parameters.

Let us define the function

$$\hat{\beta}_{nt} = \frac{R\nu_0\mu_{0t}^2 + (R+\nu_0)\sum_{r=1}^{R} x_{ntr}^2 - \left(\sum_{r=1}^{R} x_{ntr}\right)^2 - 2\nu_0\mu_{0t}\sum_{r=1}^{R} x_{ntr}}{2(R+\nu_0)},$$

for $t = 1, ..., T$; $n = 1, ..., N$. It easily follows that under (20):

$$\sigma_{nt}^2|\boldsymbol{x} \sim \mathcal{IG}\left(\alpha_0 + R/2, \beta_0 + \beta_{nt}\right),$$

independent for $n = 1, ..., N$; $t = 1, ..., N$. In the case of the restricted parameterization in (21), the corresponding posterior distribution is

$$\sigma_t^2|\boldsymbol{x} \sim \mathcal{IG}\left(\alpha_0 + \frac{NR}{2}, \beta_0 + \sum_{n=1}^{N}\hat{\beta}_{nt}\right),$$

independent for $t = 1, ..., T$. Then, we use the posterior means as the plug-in point estimates of the variance per time-point, that is,

$$\mathbb{E}\left(\sigma_{nt}^2|\boldsymbol{x}\right) = \frac{\beta_0 + \hat{\beta}_{nt}}{\alpha_0 + \frac{R}{2} - 1}, \quad t = 1, \ldots, T; n = 1, \ldots, N \tag{22}$$

$$\mathbb{E}\left(\sigma_t^2|\boldsymbol{x}\right) = \frac{\beta_0 + \sum_{n=1}^{N}\hat{\beta}_{nt}}{\alpha_0 + \frac{NR}{2} - 1}, \quad t = 1, \ldots, T, \tag{23}$$

provided that $\alpha_0 + R/2 > 1$ so that both posterior expectations exist. The default values we use for the constants in eq. (22) and (23) are $\alpha_0 = 1$ and $\beta_0 = 1$, but we also report some prior sensitivity checks in the simulation section. We will use the labels "s1" and "s2" to refer to the MCMC sampler using the plug-in estimates (22) and (23), respectively.

3.3 Updating the variance

Here we discuss the case where the variance in eq. (1) is unknown. In the case where all variances are unrestricted (eq. (20)), the full conditional distributions are:

$$\sigma_{nt}^2|\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\tau} \sim \mathcal{IG}\left[\frac{R+1}{2} + \alpha_0, \frac{1}{2}\sum_{r=1}^{R}\{x_{ntr} - \mu(t; \boldsymbol{\theta}_n, \boldsymbol{\tau}_n)\}^2 + \frac{\nu_0}{2}(\theta_{nt} - \mu_{0t})^2 + \beta_0\right], \tag{24}$$

independent for $n = 1, ..., N$, $t = 1, ..., T$. Note that in this case, the full conditional distribution depends on the piecewise linear mean function $\mu(t; \theta_n, \tau_n)$. Under restriction (21), it follows that:

$$\sigma_t^2|\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\tau} \sim \mathcal{IG}\left[\frac{N(R+1)}{2} + \alpha_0, \frac{1}{2}\sum_{n=1}^{N}\sum_{r=1}^{R}\{x_{ntr} - \mu(t; \boldsymbol{\theta}_n, \boldsymbol{\tau}_n)\}^2 + \frac{\nu_0}{2}\sum_{n=1}^{N}(\theta_{nt} - \mu_{0t})^2 + \beta_0\right], \tag{25}$$

independent for $t = 1, ..., T$.

We will use the labels "s3" and "s4" to refer to the MCMC sampler using the Gibbs steps (20) and (21), respectively. Since the full conditional distribution of σ_{nt} in eq. (20) depends only on the quantities $(\boldsymbol{x}_n, \boldsymbol{\theta}_n, \boldsymbol{\tau}_n)$, the joint posterior distribution factorizes over $n = 1, ..., N$. Thus, we can split the MCMC sampling into N independent samplers, as also done to the case where the variance is fixed. However, this does not hold for eq. (21) where the state of the parameters $\mu(t; \boldsymbol{\theta}_n, \boldsymbol{\tau}_n)$ of all time-series ($n = 1, ..., N$) is required. This imposes a large computational burden since N can be quite large. Therefore, in the examples of the next section we have only applied the first three MCMC samplers (s1, s2 and s3).

4 Results

4.1 Simulation study

We considered simulated datasets of length $T = 1000$ time-points consisting of $N = 1000$ independent multivariate observations, while the number of replicates (R) is equal to $R = 3$. For $n = 1, \ldots, N$, the number of change-points ℓ_n is drawn uniformly at random from the set $\{0, 1, \ldots, 9\}$. A detailed description of the simulation mechanism is given in Section B of the Appendix.

All three samplers were applied using the following set of hyper-parameter values: $\alpha_0 = 0.1$, $\beta_0 = 0.1$, $\nu_0 = 0.005$. The prior distribution on the number of change-points corresponds to the complexity prior distribution in (10), with parameters $\alpha = 2$ and $b = 3.72$. The fixed-variance samplers (s1 and s2) are initialized from a state with 1 change-point with parameters randomly generated from the prior distribution. On the other hand, we observed the sampler s3 remains trapped in areas of low posterior probability when using random starting values. In order to deal with this issue, we initialized sampler s3 using a run of sampler s1 (with 30000 iterations). The inference is based on 50000 iterations following a burn-in period of 20000 iterations.

Our results are benchmarked against the `not` package available in the Comprehensive R Archive Network, which implements the narrowest-over-threshold method of Baranowski et al. [22]. The number of change-points is inferred using a strengthened version of the Bayesian Information Criterion [26, 38]. For the problem of detecting changes in the slope, this approach only considers univariate time-series with constant variance, thus, it is not suitable for multivariate time-series where the variance varies with time (which is the case in our datasets). In order to apply this method we averaged across the replicates of the time-series.

Figure 2 displays the difference of the estimated number of change-points ($\hat{\ell}$) from the true number (ℓ), stratified according to the true number of change-points used to generate the data. We conclude that in all cases the estimates are centered in the true value of the number of change-points. At the lower panel of Figure 2 we calculated the Mean Absolute Error (MAE) of the resulting estimates, with error bars corresponding to the standard error. For smaller number of change-points the MAE is consistently higher for sampler s1. Although the simulation scenario does not assume the same variance per time-series, note that the sampler s2 (which uses the pooled variance estimate in eq. (21)) gives slightly better results.

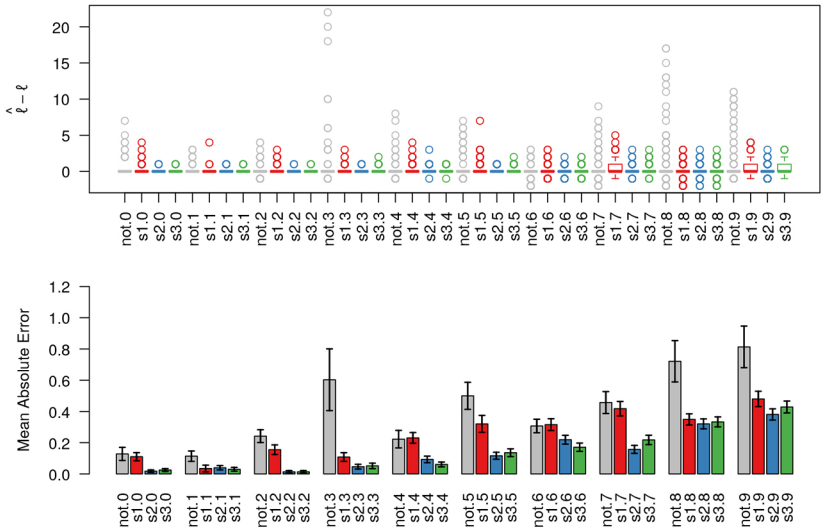


Figure 2: Benchmarking the estimation of the number of change-points arising from the narrowest-over-threshold ("not") method of [22] and our MCMC sampler with fixed different variance (s1), fixed shared variance (s2) and unknown different variance (s3) per time-series, under the complexity prior distribution. The number after the name of each sampler indicates the true number of change-points and for each value 100 synthetic datasets were simulated. Each time-series consists of three replicates and 1000 time-points.

We conclude that in many cases the method of [22] tends to overfit the number of change-points, resulting to worse performance than our method. Clearly, benchmarking against this simpler approach should not be perceived as a fair comparison between the two methods, but as a means of highlighting the usefulness of our modelling.

Figure 3 displays the output of the MCMC sampler s1 (lower panel) and s2 (upper panel) for a time-series where the true number of change-points equals to 8. The output of sampler s3 is almost identical to the lower panel of Figure 3. The posterior distribution of the number of change-points is shown at the left panels, considering both the complexity prior distribution (gray trace) as well as a Poisson(1) distribution truncated on the

set {0, 1, ..., 30}. Observe that the first choice quickly converges to a state with 8 change-points (that is, the true number). This is not the case under the Poisson prior distribution, which supports larger values than the true number of change-points. This behaviour is typical in other simulated datasets we tried, therefore, all results in the remaining sections are based on the complexity prior. The right panels display the posterior distribution of change-point locations (under the complexity prior) and it evident that the method is able to accurately infer all change-point locations.

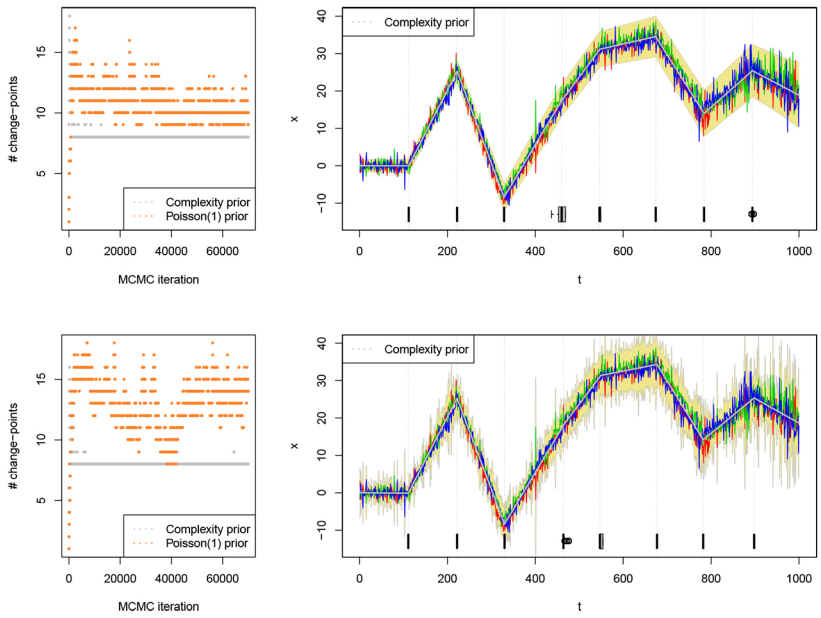


Figure 3: Example of a simulated time-series with 8 change-points. Replicates are shown in red, blue and green color. Left: MCMC trace of the sampled number of change-points considering both the complexity and a truncated Poisson(1) prior distributions (every 20th iteration is displayed). Right: output of the MCMC sampler conditionally on the estimated MAP of changepoints (which is equal to 8) according to the complexity prior. The upper and lower panels correspond to the MCMC sampler using the variance estimates arising from eq. (23) (sampler s2) and (22) (sampler s1), respectively. The gray lines correspond to the posterior mean estimates of the piecewise linear mean function and the boxplots display the posterior marginal distribution of each change-point. Vertical dotted lines correspond to the central line of each boxplot (median). The coloured outer regions correspond to two estimated standard deviations from the mean.

Additional results based on synthetic data are provided in Section B of the Appendix, including prior sensitivity checks as well as alternative simulation scenarios. Further comparisons against `not` are provided in Section D of the Appendix.

4.2 Phase detection in parallel time-series analysis of fungal growth

The filamentous fungal pathogen *Aspergillus fumigatus* is a major pathogen of the human lung causing more deaths per annum than tuberculosis or malaria [39]. A time series study of fungal growth was performed in liquid culture by analysing, in parallel, the growth characteristics of 411 independent transcription factor gene deletion mutants. The mutant strains were cultivated in a microtiter plate containing 200 μL of a fungal culture medium and incubated at 37 °C. Optical density (at 600 nm) was measured at 10 minute intervals for a total period of 48 hours. The growth analysis was performed on three separate occasions.

The observed data consists of $N \times R \times T$ growth levels for $R = 3$ replicates of $N = 411$ objects (mutants) measured every 10 minutes for $T = 289$ time-points. Figure 1 displays the observed time series for four mutants. Visual inspection reveals that describing growth with a piecewise linear mean function with an unknown number of segments is a reasonable assumption for the observed data. Regarding the fixed hyper-parameter values, we considered that $\alpha = 2$ (eq. (10)) and $\nu_0 = 10^{-1}$ (eq. (7) and (23)). After estimating the variance per time-point using the estimator in (23), the MCMC sampler s2 ran for $m = 50000$ iterations, following a burn-in period of 20000.

The boxplots in Figure 1 correspond to the estimate of the marginal posterior distribution of each change-point for specific subset of four mutants, conditionally on the mode of the posterior distribution of the number of change-points. Figure 4 displays the averaged profile per mutant (mean of three replicates) coloured according to the most probable number of change-points for each of $N = 411$ subjects. We conclude that the majority of the mutants (343) consist of three growth phases. It is clear that mutants with a smaller number of change-points

are also the more slowly growing mutants, which is reasonable since these mutants most likely have not been able to reach the later growth phases in the time of the experiment. In particular the method inferred 35 mutants with only 2 growth phases and slow growth behaviour while 12 mutants have a single phase and very slow growth behaviour. Finally, 21 mutants consist of 4 growth phases and some of them exhibit a faster growth rate at later observation stages ($t > 220$).

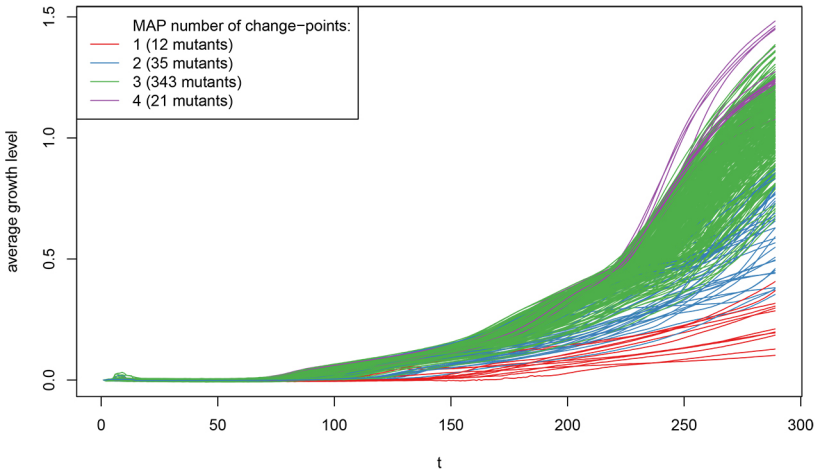


Figure 4: Visualization of the growth dataset with respect to the estimated MAP number of phases. For plotting convenience, each curve corresponds to the average growth time series across the three replicates.

Amongst 12 fungal mutants identified as having a single change-point during growth curve analysis, and therefore exhibiting severely retarded growth kinetics, seven mutants had previously been characterised [40–45]. Without exception previously characterised mutants had been reported as having various morphological defects, and these morphogenesis mutants were correctly identified in our analysis. The remaining five mutants have, until now, remained uncharacterised and therefore provide promising candidates to investigate further in order to establish their roles in fungal morphogenesis. Detailed phenotypical characterisations of the identified mutants will be described elsewhere.

a

We have also considered a Poisson(1) prior distribution on the number of change-points. As already demonstrated in Section 4.1, in this case the sampler selects a larger number of change-points which are less interpretable. The reader is referred to Section C of the Appendix.

5 Discussion

A method for inferring the number of change-points in the underlying piecewise linear mean function of replicated time-series has been presented. A crucial characteristic of the model is that each time-point may have its own mean, an assumption which introduces an overfitting number of parameters. The method is able to penalize overfitting models by using an exponentially decreasing prior distribution [10] on the number of change-points and it was demon strated that this approach leads to a posterior distribution that can accurately recover the underlying sparse structure of the model.

We considered that the variance may be fixed or treated as unknown. In the first case we used a plug-in estimate arising from a pre-processing stage of the data. Moreover two different variance parameterizations were introduced, depending on whether the variance of each time-point is shared between different time-series or not. We have also discussed the case where the variance is treated as unknown, updating its state by an extra Gibbs sampling step. According to our simulations, the samplers with fixed variance (s1 and s2) are quite competitive with the sampler s3 which also updates the variance.

There are many interesting extensions of our research. For example, one could assume more general models between replicates, such as a multivariate normal distribution with full covariance matrix and/or replicate-dependent means, or even models that are not necessarily normal. The core mechanism of the proposed MCMC sampler will be the same in these situations and it would be interesting to investigate whether the method can produce robust results in such settings. In our setup we observed that our sampler does not face any convergence issues and quickly reaches to a state where the number of change-points reflects the underlying structure of the model. In the previously mentioned generalizations however, it might be beneficial to seek ways of improving the mixing and accelerating convergence by e.g. embedding our sampler to parallel-tempering

schemes. Another interesting extension of our research is to explore the usage of data transformations for stabilizing the variance of time series along time.

In our biological application, we found that all of the slow-growing mutant strains identified by our method, and which had previously been characterised in the literature, were known to play roles in fungal morphogenesis. Further experiments are planned to explore how the growth dynamics of the mutants considered here changes under different environmental conditions. Our simple change-point method provides a useful low-dimensional model of the growth dynamics to explore gene-environment effects on the growth phenotype.

6 Software and data availability

Our algorithm is available as an `R` package [46] at the Comprehensive R Archive Network [47]. Scripts to reproduce real and simulated data analysis are available online at `https://github.com/mqbssppe/growthPhaseMCMC`.

Acknowledgements

The authors would like to acknowledge the assistance given by IT services and use of the Computational Shared Facility of the University of Manchester. The suggestions of two anonymous reviewers helped to improve the findings of this study.

Funding

Research was funded by MRC award MR/M02010X/1(Funder Id: http://dx.doi.org/10.13039/501100000265).

Conflicts of interest: The authors have declared no conflict of interest.

A Details of the MCMC sampler

The following pseudo-code summarizes the workflow of the proposed Metropolis–Hastings MCMC sampler for samplers s1 and s2.

1. Pre-processing step to estimate the variance according to eq. (22) (for sampler s1) or (23) (for sampler s2).

2. For $n = 1, \ldots, N$

- **a.** Give some initial values $\boldsymbol{\ell}_n^{(0)}, \boldsymbol{\theta}_n^{(0)}, \boldsymbol{\tau}_n^{(0)}$
- **b.** For $m = 1, \ldots, M$
  - **i. Move 1:** Propose addition or deletion of a change-point and accept the candidate state $\ell'_n, \boldsymbol{\tau}'_n$ according to eq. (15) or (16), respectively. In case of acceptance set $\left(\ell_n^{(m)}, \boldsymbol{\tau}_n^{(m)}\right) = (\ell'_n, \boldsymbol{\tau}'_n)$, otherwise $\left(\ell_n^{(m)}, \boldsymbol{\tau}_n^{(m)}\right) = \left(\ell_n^{(m-1)}, \boldsymbol{\tau}_n^{(m-1)}\right)$.
  - **ii. Move 2:** Propose to update the mean parameters and accept the candidate state $\boldsymbol{\theta}'_n$ according to eq. (17). In case of acceptance set $\boldsymbol{\theta}_n^{(m)} = \boldsymbol{\theta}'_n$, otherwise $\boldsymbol{\theta}_n^{(m)} = \boldsymbol{\theta}_n^{(m-1)}$.
  - **iii.** Propose to update the change-points: with probability 0.5 choose **Move 3.a**, otherwise choose **Move 3.b**. Accept the candidate state $\boldsymbol{\tau}'_n$ according to (18). In case of acceptance set $\boldsymbol{\tau}_n^{(m)} = \boldsymbol{\tau}'_n$.
  - **iv. Move 4:** Update the mean parameters that are not allocated to a change-point by sampling from the prior distribution.

In the case that the variance is unknown,the sampler is augmented by an extra Gibbs sampling step, which will update the variances using the full conditional distributions descibed in Section 3.3 of the paper.

In the presented applications we considered that the total number of MCMC iterations is equal to $M = 70000$, while the first 20000 are discarded as burn-in period. The parameters of the MCMC sampler are defines as follows: $c = 0.05$ (Move 2), $d_1 = 1$ (Move 3.a), $d_2 = T/20$ (Move 3.b). The MCMC sampler is initialized from a state with one randomly selected change-point. The parameters $\boldsymbol{\theta}_n$ are initialized from the mean of the multivariate normal distribution corresponding to the posterior distribution arising from eq. (7) and (20), for $n = 1, \ldots, N$.

B Simulation study details

We considered simulated datasets of length $T = 1000$ time-points consisting of $N = 1000$ independent multivariate observations. Three simulated datasets of dimensionality $N \times R \times T$ were genera ted according to (4) considering that the number of replicates (R) is equal to $R = 3$ or 6. For $n = 1, ..., N$, the number of change-points ℓ_n is drawn uniformly at random from the set $\{0, 1, ..., 9\}$. Given $\ell_n > 0$, the "global position" of the j-th change-point is generated as

$$\tau_j = \frac{T}{(\ell_n + 1)} j + y_j, \quad j = 1, \dots, \ell_n \tag{26}$$

where y_j denote independent draws from the Binomial(100,0.5) distribution, $j = 1, \dots, \ell_n; n = 1, ..., N$. According to this scheme note that $\mathbb{E}(\tau_{j+1} - \tau_j) = \frac{T}{(\ell_n+1)}$ and $\mathrm{Var}(\tau_{j+1} - \tau_j) = 50$.

Recall that our model assumes that the replicates are iid random variables for a given time-point t and time-series n. However, we would rather benchmark our method in case that this assumption is not necessarily true. In order to introduce extra noise between replicates we assumed that the true change-point for each replicate has some variation around the time-point in eq. (26). Therefore, it was considered that replicate r changes its mean function at time

$$\tau'_{jr} = \tau_j + dz_{jr} \tag{27}$$

where $d \sim \mathcal{DU}\{-1, 1\}$ and $z \sim \mathcal{P}(2)$, independent for r, j, where $\mathcal{P}(\cdot)$ denotes the Poisson distribution. The expression level of each subject $n = 1, ..., N$ was initialized at time-point $t = 1$ by a baseline mean equal to zero for the first phase. In case that the number of change-points for subject n is positive ($\ell_n > 0$), the expression level for each phase was assumed to correspond to a linear function by randomly generating the slope S_j for each consecutive phase according to $S_j = \omega_j |Y_j|$ where $Y_j \sim \mathcal{N}(0, 0.3^2)$, independent for $j = 1, \dots, \ell_n$. ω_j denotes a discrete random variable with values in $\{-1, 1\}$, distributed according to a Markov process with $\mathbb{P}(\omega_j = 1 | \omega_{j-1} = -1) = \mathbb{P}(\omega_j = -1 | \omega_{j-1} = 1) = p$, where $p = 0.8$.

Each replicate was allowed to have some variation around the endpoints of each phase according to the $\mathcal{N}(0, \sigma^2_{\text{rep}})$ distribution, where

$$\sigma^2_{\text{rep}} = 1 \,. \tag{28}$$

B.1 Simulation study 1: different variance per time-series

Regarding the observation variance σ^2_{nt} it is assumed that

$$\sigma^2_{nt} \sim \mathcal{G}\left(1, -\frac{0.9}{T-1} t + \frac{T - 0.1}{T - 1}\right),$$

independent for $t = 1, ..., T$ and $n = 1, ..., N$. Note that $\{\sigma^2_{nt}; n = 1, \dots, N\}$ is a random sample of size N from a Gamma distribution. In order to estimate the variance per time point and time series both estimators (22) and (23) are considered.

Figure 5 illustrates a random subset of 10 simulated time-series with a number of change-points ranging in $0, 1, ..., 9$. Note that there is strong heterogeneity between time-series and that there are instances where replicates deviate from the iid assumption. We will refer to the specific data generation procedure with the term "noisy scenario". Note that when $d = 0$ in eq. (27) and $\sigma^2_{\text{rep}} \to 0$ in eq. (28) the simulation scenario does not introduce any extra noise between replicates and the data is generated by the assumptions imposed by the proposed model. Note that the results of this simulation procedure are shown in Figure 2 of the main paper.

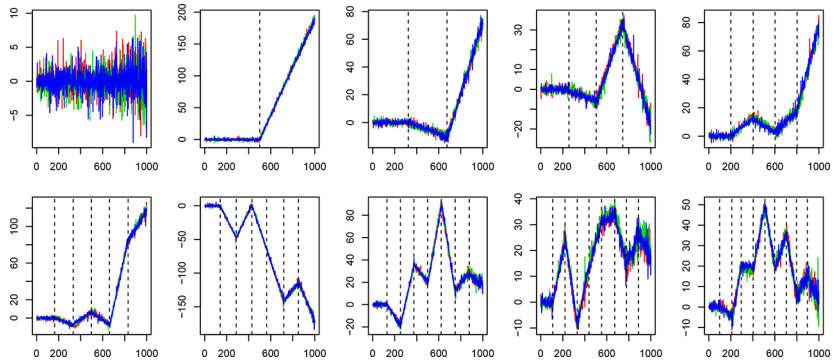
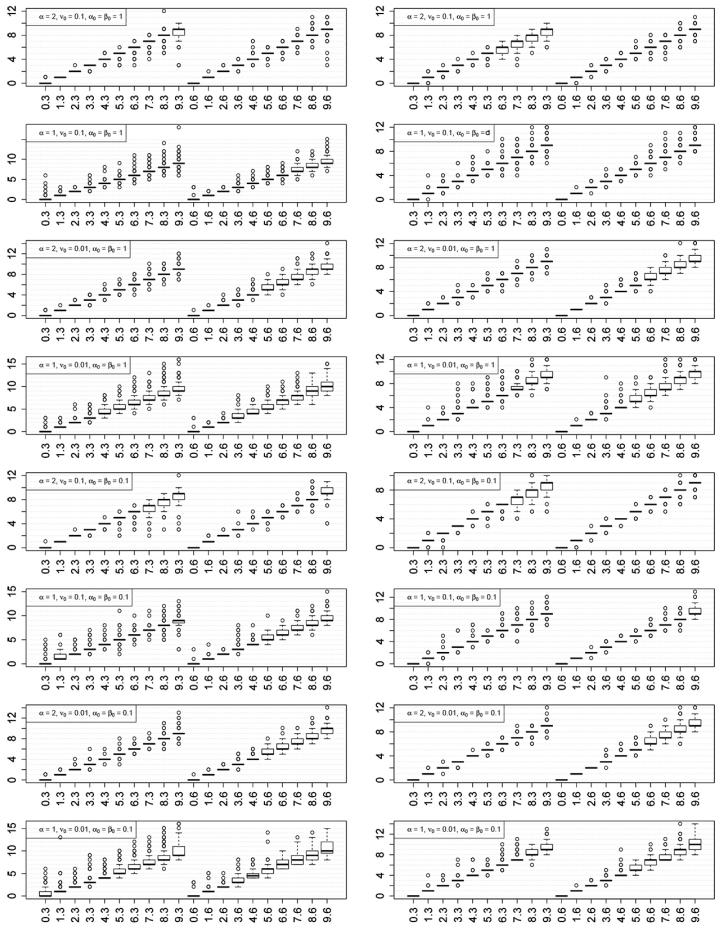


Figure 5: An example of 10 subjects from the simulated time-series with $T = 1000$ time-points with 3 replicates (blue, red, green) and number of change-points equal to $0, 1, \ldots, 9$ (indicated by dashed vertical lines).

Next we perform some prior sensitivity checks by considering different combinations of the hyper-parameters. Figure 6 shows the selected number of change-points using the approximate MAP estimate from the MCMC sample. Prior-sensitivity checks are performed by considering that $\alpha \in \{1, 2\}$ in eq. (10), $\nu_0 \in \{0.01, 0.1\}$ in eq. (7) and $(\alpha_0, \beta_0) \in \{(0.1, 0.1), (1, 1)\}$ in eqs. (22)–(23). The results are stratified with respect to the true number of change-points used to generate each time series and we conclude that it is accurately estimated in most cases. Recall that the parameter α controls how fast is the exponential decrease in the prior distribution of the number of change-points, hence larger values of α yield heavier penalties for complex models. This behaviour is reflected in Figure 6 where we observe that $\alpha = 1$ tends to produce larger MAP estimates than $\alpha = 2$. Observe also that the results are reasonably robust with respect to the parameter ν_0.



(a) Using variance estimator (22) (b) Using variance estimator (23)

Figure 6: Estimation of the number of change-points on synthetic datasets generated under different variance per time series. The variance was estimated according to (a): the different variance estimator (22) and (b): the same variance estimator (23). Different combinations of prior parameters $(\alpha, \nu_0, \alpha_0, \beta_0)$ were used to the MCMC sampler. Each pair of numbers in the horizontal axis displays the true number of change-points (first entry) and number of replicates (second entry).

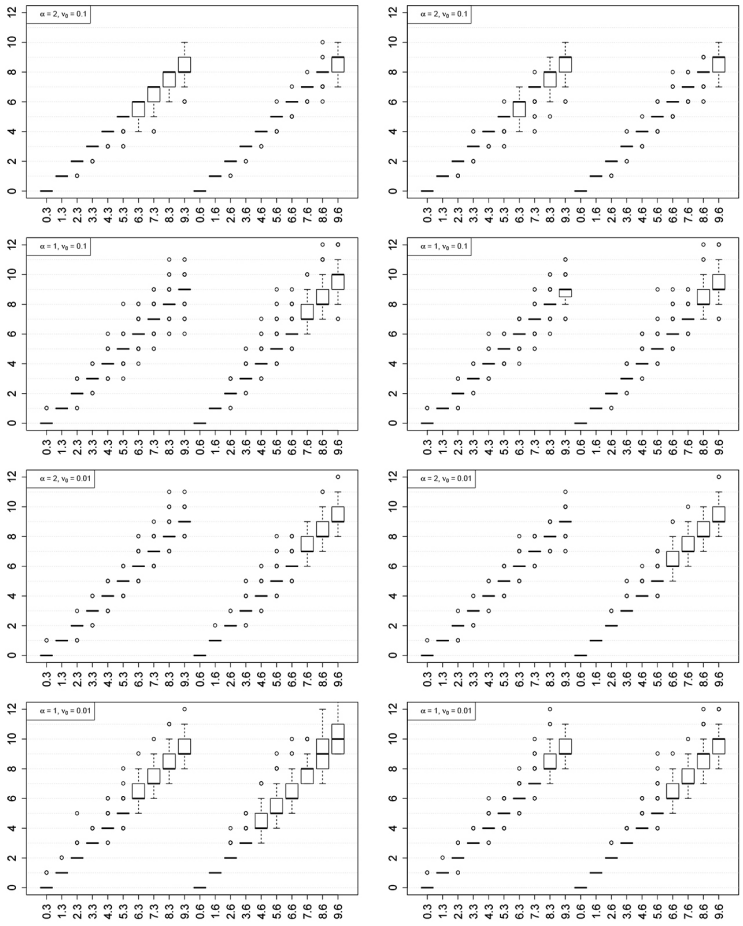
B.2 Simulation study 2: same variance per time-series

In this section we replicate the "noisy" simulation scenario of Section B but now the variance is restricted to be the same between time series. More specifically, it is assumed that $\sigma^2_{nt} = \sigma^2_t$ for all $n = 1, \ldots, N$ and that

$$\sigma^2_t \sim \mathcal{G}\left(1, -\frac{0.9}{T-1}t + \frac{T-0.1}{T-1}\right),$$

independent for $t = 1, \ldots, T$. Furthermore, we also consider the "exact" simulation scenario where no additional noise is introduced between replicates and expression levels. In order to estimate the variance per time point and time series we used the estimator in (23). Then, the MCMC sampler was run with the same number of iterations and burn-in period as previously (70000 and 20000, respectively).

The estimate of the number of change-points are shown in Figure 7(a) and Figure 7(b) for the noisy and exact simulation scenarios. We should note that the results are improved when replicating the analysis based on simulated datasets which are generated exactly by the assumed model, as illustrated in Figure 7(b), particularly when $\alpha = 1$ and $\nu_0 = 0.01$.



(a) Noisy Scenario (b) Exact Scenario

Figure 7: Benchmarking the estimation of the number of change-points on synthetic datasets according to the "noisy" and "exact" simulation scenarios. Different combinations of prior parameters ν_0 in eq. (7) and α in eq. (10) were used to the MCMC sampler. Each pair of numbers in the horizontal axis displays the true number of change-points (first entry) and number of replicates (second entry).

C Benchmarking against the Poisson distribution on the real dataset

Recall that in Section 4.1 we demonstrated, using simulated data, that the estimated number of change-points tends to overfit when using a Poisson instead the complexity prior distribution. Here, we illustrate that this is also the case on our real dataset. We consider again a Poisson(1) distribution, truncated on the set $\{0, 1, \ldots, 30\}$ and use the same four time-series of the real dataset, depicted in Figure 1 of the main paper. As shown in Figure 8, the posterior distribution of the number of change-points under the Poisson prior distribution supports much larger values than the complexity prior. The estimated posterior mode of the number of change-points correspond to 5 for "plate 5 E1" (instead of 1 under the complexity prior), 5 for "plate 1 F4" (instead of 2), 6 for "plate 1 A3" (instead of 3) and 7 for "plate 2 C4" (instead of 4). We conclude that in this case the sampler assigns additional change-points to intermediate observation periods, compared to the ones selected under the complexity prior. These additional change-points identify very small changes in the slope of the time-series and they do not contribute much in the interpretation of growth characteristics.

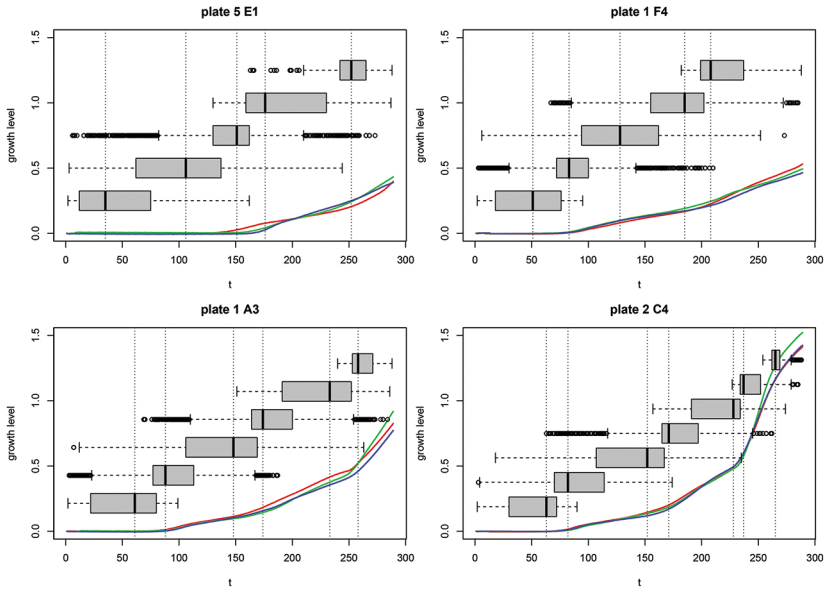


Figure 8: Change-point locations conditionally on the MAP number of change-points according to a Poisson(1) prior distribution, truncated on the set $\{0, 1, \ldots, 30\}$ for four time-series of our real dataset.

In order to further inspect the differences between the results arising from the two different prior assumptions, let us define the following random variables:

$$z_{nt} = \begin{cases} 1, & \text{if } t \in \{\tau_1, \ldots, \tau_{\ell_n}\} \\ 0, & \text{otherwise} \end{cases} \tag{29}$$

for $t = 1, \ldots, T, n = 1, \ldots, N$. Note that z_{nt} is a binary random variable with 1 denoting the event that a change-point is assigned at time-point t, $t = 1, \ldots, T$, for time-series $n = 1, \ldots, N$. The posterior probability $P(z_{nt} = 1 | \boldsymbol{x}, \ell_n)$ can be estimated directly by averaging across the MCMC output. Figure 9 illustrates the estimates of these posterior probabilities for a subset of four time-series. The peaks correspond to the locations of sampled change-points across the MCMC run. Notice that a blue peak is always accompanied by a red one, which means that under the Poisson distribution the set of inferred change-points actually contains the change-points locations selected from the complexity prior. However, the reverse is not necessarily true, especially for the first time series ("plate 5 E1"). Finally, notice that under the complexity prior distribution all intermediate time-points between two peaks are assigned zero posterior probabilities of containing change-points. This is not the case for the Poisson prior distribution, where all time-points contain a change-point with strictly positive posterior probabilities, which is due to the presence of the additional change-points.

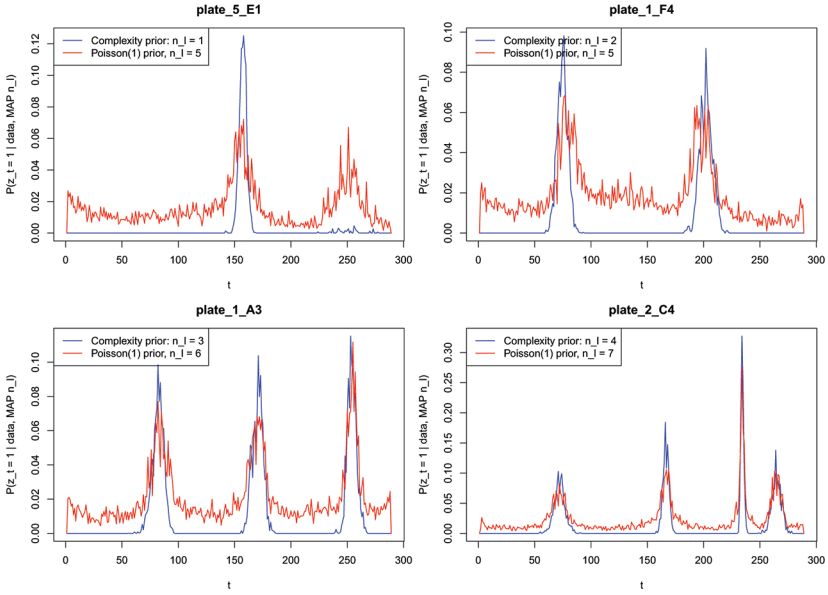


Figure 9: Estimated posterior probabilities of the binary state variables z_{nt}, $t = 1, \ldots, T$ in eq. 29 for four time-series of the real dataset, conditionally on the most probable number of change-points (n_ℓ), according to the complexity and Poisson(1) prior distributions on the number of change-points. The results are averaged across 20 independent chains.

D Additional benchmarking against the narrowest-over-threshold method

In this section we provide some additional results regarding synthetic and true datasets when using the `not` package of Baranowski et al. [22]. We consider one of our simulated datasets as well as one time-series of our real data, which illustrate the typical performance of each method in our data. As already mentioned, `not` deals with univariate time-series so in order to apply this method we averaged across the replicates of the time-series. The results are illustrated in Figure 10. The first column consists of a simulated dataset (using the same generating mechanism described in Section B, where the number of change-points is equal to 3 and our proposed method is able to succesfully detect it (shown in first row of Figure 10). On the other hand, the method of Baranowski et al. [22] overfits the number of change-points. A similar pattern is illustrated for the real dataset shown in the second column of Figure 10, where the narrowest-over-threshold method infers a much larger number of change-points compared to our approach, which is not realistic.

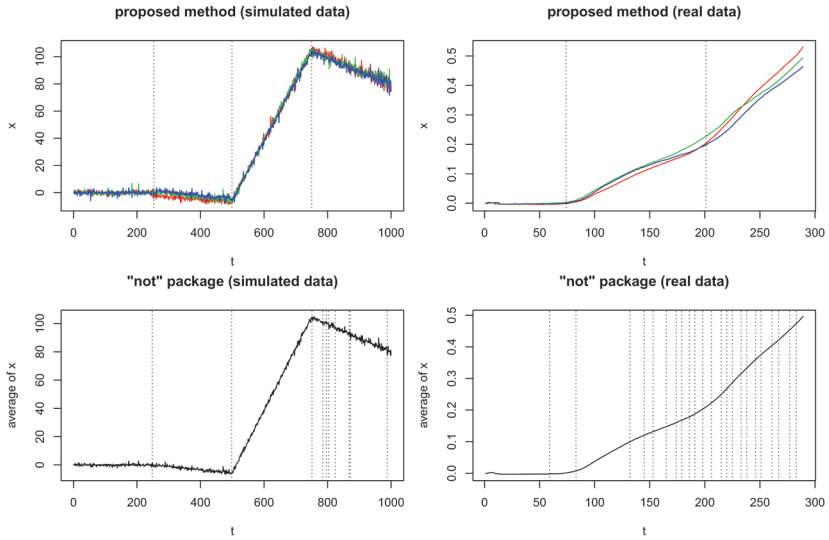


Figure 10: Comparison with the narrowest-over-threshold method as implemented in the `not` R package. Inferred change-points are indicated by dotted lines.

References

[1] Basseville M, Nikiforov IV. *Detection of abrupt changes: theory and application* Vol. 104. Englewood Cliffs: Prentice Hall, 1993

[2] Fischer M, Sawers RG. A universally applicable and rapid method for measuring the growth of Streptomyces and other filamentous microorganisms by methylene blue adsorption-desorption. Appl Environ Microbiol. 2013;79:4499–502.

[3] Stevenson K, McVey AF, Clark IB, Swain PS, Pilizota T. General calibration of microbial growth in microplate readers. Sci Rep. 2016;6:38828.

[4] Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 1995;82:711–32.

[5] Johnson TD, Elashoff RM, Harkema SJ. A Bayesian change-point analysis of electromyographic data: detecting muscle activation patterns and associated applications. Biostatistics. 2003;4:143. DOI: 10.1093/biostatistics/4.1.143.

[6] Punskaya E, Andrieu C, Doucet A, Fitzgerald WJ. Bayesian curve fitting using MCMC with applications to signal segmentation. IEEE Trans Signal Process. 2002;50:747–58.

[7] Tai YC, Kvale MN, Witte JS. Segmentation and estimation for SNP microarrays: A Bayesian multiple change-point approach. Biometrics. 2010;66:675–83. DOI: 10.1111/j.1541-0420.2009.01328.x.

[8] Zhao X, Chu P-S. Bayesian changepoint analysis for extreme events (typhoons, heavy rainfall, and heat waves): An RJMCMC approach. J Clim. 2010;23:1034–46.

[9] Lavielle M, Lebarbier E. An application of MCMC methods for the multiple change-points problem. Signal Process. 2001;81:39–53.

[10] Castillo I, van der Vaart A. Needles and straw in a haystack: Posterior concentration for possibly sparse sequences. Ann Stat. 2012;40:2069–101. DOI: 10.1214/12-AOS1029.

[11] Chib S. Estimation and comparison of multiple change-point models. J Econometrics. 1998;86:221–41. http://www.sciencedirect.com/science/article/pii/S0304407697001152.

[12] Chib S. Marginal likelihood from the Gibbs output. J Am Stat Assoc. 1995;90:1313–21.

[13] Fearnhead P. Exact and efficient Bayesian inference for multiple changepoint problems. Stat Comput. 2006;16:203–13.

[14] Dobigeon N, Tourneret J-Y, Scargle JD. Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model. IEEE Trans Signal Process. 2007;55:414–23.

[15] He C. Bayesian multiple change-point estimation for exponential distribution with truncated and censored data. Commun Stat - Theo Methods. 2017;46:5827–39. DOI: 10.1080/03610926.2016.1161797.

[16] Hutter M. Exact Bayesian regression of piecewise constant functions. Bayesian Anal. 2007;2:635–64. DOI: 10.1214/07-BA225.
[17] Kim J, Cheon S. Bayesian multiple change-point estimation with annealing stochastic approximation Monte Carlo. Comput Stat. 2010;25:215–39. DOI: 10.1007/s00180-009-0172-x.
[18] Rudoy D, Yuen SG, Howe RD, Wolfe PJ. Bayesian change-point analysis for atomic force microscopy and soft material indentation. J R Stat Soc: Ser C (Appl Stat). 2010;59:573–93. DOI: 10.1111/j.1467-9876.2010.00715.x.
[19] Schütz N, Holschneider M. Detection of trend changes in time series using Bayesian inference. Phys. Rev. E. 2011;84:021120. DOI: 10.1103/PhysRevE.84.021120.
[20] Schwaller L, Robin S. Exact bayesian inference for off-line change-point detection in tree-structured graphical models. Stat Comput. 2017;27:1331–45. DOI: 10.1007/s11222-016-9689-3.
[21] Stephens DA. Bayesian Retrospective Multiple-Changepoint Identification. Applied Statistics. 1994;43:159–159. DOI: 10.2307/2986119.
[22] Baranowski R, Chen Y, Fryzlewicz P. Narrowest-over-threshold detection of multiple change-points and change-point-like features, 2016. arXiv preprint arXiv:1609.00293.
[23] Cahill N, Rahmstorf S, Parnell AC. Change points of global temperature. Environ Res Lett. 2015;10:084002. http://stacks.iop.org/1748-9326/10/i=8/a=084002.
[24] Fearnhead P, Maidstone R, Letchford A. Detecting changes in slope with an lo penalty. J Comput Graphical Stat. 2018;0:1–11. DOI: 10.1080/10618600.2018.1512868.
[25] Schroeder AL, Fryzlewicz P. Adaptive trend estimation in financial time series via multiscale change-point-induced basis recovery. Stat Its interface. 2013;6:449–61.
[26] Fryzlewicz P. Wild binary segmentation for multiple change-point detection. Ann Stat. 2014;42:2243–81. DOI: 10.1214/14-AOS1245
[27] Scott AJ, Knott M. A cluster analysis method for grouping means in the analysis of variance. Biometrics. 1974;30:507–12. http://www.jstor.org/stable/2529204.
[28] Jackson B, Scargle JD, Barnes D, Arabhi S, Alt A, Gioumousis P, et al. An algorithm for optimal partitioning of data on an interval. IEEE Signal Process Lett. 2005;12:105–8.
[29] Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. J Am Stat Assoc. 2012;107:1590–8.
[30] Chamroukhi F, Mohammed S, Trabelsi D, Oukhellou L, Amirat Y. Joint segmentation of multivariate time series with hidden process regression for human activity recognition. Neurocomputing. 2013;120:633–644. http://www.sciencedirect.com/science/article/pii/S0925231213004086, image Feature Detection and Description.
[31] Frick K, Munk A, Sieling H. Multiscale change point inference. J R Stat Soc: Ser B (Stat Method). 2014;76:495–580. DOI: 10.1111/rssb.12047.
[32] Halpern AL. Multiple-changepoint testing for an alternating segments model of a binary sequence. Biometrics. 2000;56:903–8. DOI: 10.1111/j.0006-341X.2000.00903.x.
[33] Lu Q, Lund R, Lee TC. An MDL approach to the climate segmentation problem. Ann Appl Stat. 2010;4:299–319. DOI: 10.1214/09-AOAS289.
[34] Picard F, Lebarbier E, Budinská E, Robin S. Joint segmentation of multivariate Gaussian processes using mixed linear models. Comput Stat Data Anal. 2011;55:1160–70. http://www.sciencedirect.com/science/article/pii/S0167947310003580.
[35] Yildirim S, Singh SS, Doucet A. An online expectation–maximization algorithm for changepoint models. J Comput Graphical Stat. 2013;22:906–26. DOI: 10.1080/10618600.2012.674653.
[36] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970;57:97–109. http://www.jstor.org/stable/2334940.
[37] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys. 1953;21:1087–92.
[38] Liu J, Wu S, Zidek JV. On segmented multivariate regression. Statistica Sinica. 1997;7:497–525
[39] Brown GD, Denning DW, Gow NA, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. Sci Transl Med. 2012;4:165rv13–165rv13.
[40] Amich J, Schafferer L, Haas H, Krappmann S. Regulation of sulphur assimilation is essential for virulence and affects iron homeostasis of the human-pathogenic mould Aspergillus fumigatus. PLOS Pathog. 2013;9:1–24. DOI: 10.1371/journal.ppat.1003573.
[41] Bertuzzi M, Schrettl M, Alcazar-Fuoli L, Cairns TC, Muñoz A, Walker LA, et al. The pH-responsive PacC transcription factor of Aspergillus fumigatus governs epithelial entry and tissue invasion during pulmonary aspergillosis. PLoS Pathog. 2014;10:e1004413.
[42] Dinamarco TM, Almeida RS, de Castro PA, Brown NA, dos Reis TF, Ramalho LN, et al. Molecular characterization of the putative transcription factor SebA involved in virulence in Aspergillus fumigatus. Eukaryotic Cell 2012;11:518–31.
[43] Gsaller F, Hortschansky P, Furukawa T, Carr PD, Rash B, Capilla J, et al. Sterol biosynthesis and azole tolerance is governed by the opposing actions of SrbA and the CCAAT binding complex. PLOS Pathog. 2016;12:1–22. DOI: 10.1371/journal.ppat.1005775.
[44] Lee M-K, Kwon N-J, Lee I-S, Jung S, Kim S-C, Yu J-H. Negative regulation and developmental competence in Aspergillus. Sci Rep. 2016;6:28874.
[45] Willger SD, Puttikamonkul S, Kim K-H, Burritt JB, Grahl N, Metzler LJ, et al. A sterol-regulatory element binding protein is required for cell polarity, hypoxia adaptation, azole drug resistance, and virulence in Aspergillus fumigatus. PLOS Pathog. 2008;4:1–18. DOI: 10.1371/journal.ppat.1000200.
[46] Papastamoulis P. beast: Bayesian Estimation of Change-Points in the Slope of Multivariate Time-Series, 2017. http://CRAN.R-project.org/package=beast, r package version 1.0.
[47] R Development Core Team. R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2008. http://www.R-project.org, ISBN 3-900051-07-0.