

# Local Level Dynamic Random Partition Models for Changepoint Detection

Alice Giampino<sup>\*</sup>, Bernardo Nipoti<sup>†</sup>, Marina Vannucci<sup>‡</sup>, and Michele Guindani<sup>§</sup>

**Abstract.** Motivated by an increasing demand for models that can effectively describe features of complex multivariate time series, e.g. from sensor data in biomechanics, motion analysis, and sports science, we introduce a novel state-space modeling framework where the state equation encodes the evolution of latent partitions of the data over time. Building on the principles of dynamic linear models, our approach develops a random partition model capable of linking data partitions to previous ones over time, using a straightforward Markov structure that accounts for temporal persistence and facilitates changepoint detection. The selection of changepoints involves multiple dependent decisions, and we address this time-dependence by adopting a non-marginal false discovery rate control. This leads to a simple decision rule that ensures more stringent control of the false discovery rate compared to approaches that do not consider dependence. The method is efficiently implemented using a Gibbs sampling algorithm, leading to a straightforward approach compared to existing methods for dependent random partition models. Additionally, we show that the proposed method can be adapted to handle multiview clustering scenarios. Simulation studies and the analysis of a human gesture phase dataset collected through specific sensing technologies show the effectiveness of the method in dynamically clustering multivariate time series and detecting changepoints.

**Keywords:** Bayesian nonparametrics, Bayesian clustering, Chinese restaurant process, Gibbs-type prior, spike-and-slab prior.

## 1 Introduction

Recently, there has been an increased demand for models that can effectively describe features of complex multivariate time-series data. This surge in interest is particularly prominent in fields such as biomechanics, motion analysis, human-computer interaction, and sports science. Here, one of the goals is often to break down continuous human motion or activity into distinct phases and states. The insights derived from such analyses can then be used to improve sports performances and for a deeper study of the complexity of human biomechanics, for example, to understand gait cycles, gesture phases, athletic movements, and even cognitive states. In the analysis of human gesture data,

---

arXiv: [2407.20085](https://arxiv.org/abs/2407.20085)

<sup>\*</sup>Department of Economics, Management and Statistics, University of Milano-Bicocca, Milan, Italy, [alice.giampino@unimib.it](mailto:alice.giampino@unimib.it)

<sup>†</sup>Department of Economics, Management and Statistics, University of Milano-Bicocca, Milan, Italy, [bernardo.nipoti@unimib.it](mailto:bernardo.nipoti@unimib.it)

<sup>‡</sup>Department of Statistics, Rice University, TX, U.S.A., [marina@rice.edu](mailto:marina@rice.edu)

<sup>§</sup>Department of Biostatistics, University of California Los Angeles, CA, U.S.A., [mguindani@ucla.edu](mailto:mguindani@ucla.edu)

information is typically gathered through various sensing technologies, such as motion capture systems, accelerometers, or videos. These technologies yield rich datasets that capture the temporal patterns of human movements. Understanding how the movements change, interact, and unfold over time is essential for capturing the underlying processes driving human motion and behavior. This temporal evolution holds key information about transitions between movement phases, variations in performance, and subtle cues related to coordination and intent, all of which are critical for developing accurate and interpretable models. Such data are commonly collected as multivariate time series, indexed by discrete time points  $t = 1, \dots, T$ , where each time point corresponds to a multivariate observation. We let  $\mathbf{Y}_{1:T} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_T\}$  denote a multivariate time series observed on  $n$  units over  $T$  time points, where each  $\mathbf{Y}_t$  is a  $n$ -dimensional vector, i.e.,  $\mathbf{Y}_t = (Y_{1,t}, Y_{2,t}, \dots, Y_{n,t})$ . To illustrate, consider our application in Section 5.1, where we examine scalar velocity data obtained from acceleration sensor units placed on the hands and wrists of a subject. Dynamic linear models (DLMs) are commonly employed for the analysis of time-series data due to their flexibility and adaptability in handling diverse situations (Petris et al., 2009). They define a class of state-space models and are characterized by a system of two equations: an observation equation, which describes the observed data as a linear combination of latent state variables with noise; and a state equation that describes how latent states evolve over time, thereby tracking the underlying dynamics of the system. We introduce our contribution by referring to a simple yet fundamental DLM, the *local level model*, which describes the observed data as composed of a level component plus random noise,

$$Y_{i,t} = \beta_{i,t} + \varepsilon_{i,t}, \quad (1)$$

where  $\varepsilon_{i,t} \stackrel{\text{iid}}{\sim} \text{N}(0, \tau^2)$ , with  $\text{N}(\mu, \sigma^2)$  denoting the Normal distribution with mean  $\mu$  and variance  $\sigma^2$ . The vector  $\boldsymbol{\beta}_t = (\beta_{1,t}, \dots, \beta_{n,t})$  represents the underlying level or trend of the time series at time  $t$ . In a typical local level model, the evolution of  $\boldsymbol{\beta}_t$  over time is modeled as a random walk, i.e., the level at time  $t$  is predicted to be the same as the level at time  $t - 1$ , plus some random noise. If the variance of the random noise is small, this assumption implies smooth processes over time. Despite its simplicity, the local level model illustrates the fundamental characteristics of many time-series models, and will serve as a basic example throughout.

One of the main goals of our biometric data application is to study how measurements from different units cluster or group over time, as well as how these clusters evolve during an activity, to understand how different body parts cooperate during different stages of a gesture or movement. Furthermore, the detection of changepoints in the clustering of body parts between gesture stages can reveal important insights into the motion and inform gesture detection algorithms. In the Bayesian setting, Bayesian nonparametric (BNP) models are a popular choice for clustering dynamic behavior over time (Quintana et al., 2022). BNP models do not require the upfront specification of the number of clusters; instead, they allow for posterior inference on cluster allocations directly from the data. Existing BNP approaches for clustering time-series data vary in terms of motivation, application, and how time-dependence is introduced. For example, some approaches build on the stick-breaking representation of the Dirichlet Process (Ferguson, 1973; Sethuraman, 1994). In this context, Antoniano-Villalobos

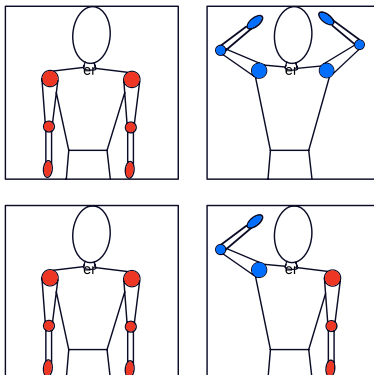


Figure 1: Example of sensors’ clusters at two-time points (left, right). Colors identify the clusters of sensors, defined based on the measurement of their scalar velocity. Top: the arms are clustered together in both images, despite a change in the cluster label. Bottom: the scalar velocity measured across sensors defines two distinct partitions, shown in the left and right images.

and Walker (2016) have developed a stationary Markov model where both the transition and stationary densities are nonparametric infinite mixture models. Nieto-Barajas and Contreras-Cristán (2014) have clustered temporal data while considering several features typical of time-series data (e.g., trends and seasonality). BNP autoregressive models are discussed in Kalli and Griffin (2018), De Iorio et al. (2023), DeYoreo and Kottas (2018), Pavani and Quintana (2025), and Beraha et al. (2023), among others. Alternatively, other authors have explored generalizations of the Pólya urn scheme of Blackwell et al. (1973), see, e.g. Caron et al. (2007), Caron et al. (2017) and Cassese et al. (2019). While our goal is to capture changes in the clustering structure of the data over time (see the motivating example shown in Figure 1), where a *changepoint* corresponds to a shift in the partition of the sensors across successive time points, all the methods mentioned above identify clusters based on the values of associated parameters, e.g. the latent variables  $\beta_{i,t}$  in the illustrative local level model (1). These models can detect changepoints if, for instance, two data points assume a common value at times  $t$  and  $t + 1$  but that value changes significantly between time points. This occurs because model-based clustering with Bayesian nonparametric models essentially relies on specifying a mixture model that depends on a discrete random mixing measure. The probabilistic distribution over random partitions is essentially a by-product of this mixture setup; in essence, clusters in a mixture model are identified based on the values of the process’ atoms. We illustrate this point in Figure 1 in the case of clustering of sensors, where the grouping may not depend on the intensity of the values but rather on the dependence in their movements.

An alternative approach involves directly modeling the sequence of random partitions. In this context, Page et al. (2022) introduced a dependent random partition model (Hartigan, 1990; Barry and Hartigan, 1992) that incorporates auxiliary variables

that determine whether a unit, for which the cluster allocation at time  $t - 1$  has been determined, should be considered for possible cluster reallocation at time  $t$ . As a result, changes in partitions over time are obtained as a by-product of individual clustering assignments. Paganin et al. (2023) proposed a model for a sequence of random partitions over time, considering individual allocation probabilities for each unit rather than the whole partition at each time point. Cremaschi et al. (2023) introduced a Bayesian changepoint model for spatiotemporal data, where partition changes in areal units over time are governed by a spatial association parameter. Their model employs a random partition prior that incorporates spatial features, promoting the co-clustering of nearby areal units based on proximity. Quinlan et al. (2022) presented a method that aims to correlate partitions with the detection of changepoints in multivariate time series in the context of contagion phenomena. However, following Martinez and Mena (2014), they specify only a single unique partition of contiguous clusters for each time series. The focus of their model is on multivariate changepoint detection rather than data partitioning, with changepoints occurring at different time points for each time series. Recently, Corradin et al. (2024) proposed a Bayesian model for clustering time series, where two observations are assigned to the same group if they exhibit structural changes in their behavior at the same time. Their approach assumes that units within the same cluster share changepoints across all time points, with no possibility of shifts of units in different clusters at different time points.

In this paper, we propose viewing the evolution of partitions over time as latent partition-based states within a state-space model. Our approach develops a random partition model (RPM) capable of linking the partition of data points to previous partitions over time, using a straightforward Markov structure to model these changes. The approach builds upon the principles of DLMS but extends them to incorporate latent state equations now operating within the partition context, with time evolution defined by the partitions themselves. Therefore, while in the analysis of time-series data, it is common to identify sudden changes in the observed values of a stochastic process as a changepoint, in this paper, we refer to a changepoint as a change in a latent partition of units. More specifically, we employ a Markov-dependent structure, where the partition at time  $t$ , denoted as  $\pi_t$ , is modeled conditionally on the partition at time  $t - 1$ , accounting for temporal persistence and enabling changepoint detection. The selection of changepoints at each time point  $t$  is driven by a mixture that chooses between one of two partitions. At any given time, the chosen partition can either remain the same as the one at the previous time point or be sampled from a flexible and general random partition model. In the presence of a changepoint, the partition at time  $t$  becomes independent of the partition at time  $t - 1$ . We demonstrate that, based on our construction, the random partitions  $\pi_t$ , for  $t = 1, \dots, T$ , are marginally identically distributed, with the marginal distribution that can be specified so to ensure tractability. Unlike the existing Bayesian nonparametric models cited above, our clustering approach does not directly rely on the values of parameters, such as the  $\beta_{i,t}$ 's in equation (1). Instead, we directly treat dependent random partition allocations as latent structures that drive the dynamics of the observations. In contrast to the dependent random partition model proposed by Page et al. (2022), we jointly consider all the units when identifying changepoints, taking into account the multivariable nature of the dependent partitions.

For posterior inference, our approach relies on the efficiency of a Gibbs sampling algorithm, offering a straightforward implementation compared to existing methods for dependent random partition models. The selection of changepoints involves dependent multiple decisions, and we propose accounting for this time-dependence by adapting the decision-theoretic approach of Chandra and Bhattacharya (2019) in defining the error and non-error terms of the decision loss function. The resulting procedure establishes a simple decision rule that ensures more stringent control of the false discovery rate compared to an approach that does not consider dependence.

Recent methods have been proposed for multiview clustering that allow for dependence between several experimental conditions or features. For example, Franzolini et al. (2023) employ a conditional partially exchangeable model to induce dependence between clustering configurations of the same subjects across different features. Similarly, Dombowsky and Dunson (2025) develop a nonparametric prior that generates dependent random distributions by centering a Dirichlet process on a random product measure, with a single parameter controlling the dependence. While our approach is designed to model the evolution of partitions over time, as in the motivating application shown in Figure 1, it may be adapted to multiview clustering scenarios. In such cases, an analogous latent-state representation of the partitions can be employed to build a hierarchical model that effectively captures the dependence between partitions, describing changes across different groups or experimental conditions. Additionally, we show that the dependence between the partitions generated by the hierarchical model for multiview clustering in Dombowsky and Dunson (2025) coincides with that one of a reparametrization of our prior specification. This is not, however, the focus of this work and we note that moving beyond a two-view setting introduces additional computational challenges, which cannot be addressed by a straightforward adaptation of the algorithm designed for time-series data, and thus require careful treatment to preserve both efficiency and robustness. Finally, we should note that while constructing the prior distribution for the partition bears some resemblance to the use of spike-and-slab priors for variable selection (Tadesse and Vannucci, 2021), dealing with partitions introduces additional complexity in the motivation, modeling and computation.

The article is organized as follows: in Section 2, we present our model and discuss its main properties, including the alternative multiview formulation discussed in Section 2.2. Section 3 addresses posterior inference, including computational methods and decision theory-based changepoint detection. In Section 4 we describe two simulation studies that highlight the key aspects of our model. In Section 5, we present an application to the analysis of human gesture data, and illustrate its performance in two-view clustering using data from a collaborative perinatal project. Finally, concluding remarks and future directions are provided in Section 6. Proofs and additional details on computations and illustrations are available in the Supplementary Material (Giampino et al., 2025).

## 2 Local level dynamic partition model

We describe the key features of the dynamic partition model we propose, a local level dynamic partition model (LLDPM) taking model (1) as reference. More specifically, we

assume that for each unit  $i$ , with  $i = 1, \dots, n$ , the observations are generated from some general likelihood (observation equation) as

$$Y_{i,t} \mid \beta_{i,t} \stackrel{\text{ind}}{\sim} p(y_{i,t} \mid \beta_{i,t}), \quad (2)$$

for  $t = 1, \dots, T$ . Throughout this work, we consider a Gaussian kernel with mean  $\beta_{i,t}$  and variance  $\tau^2$ . However, it is possible to extend this construction to other kernels. The dynamics of the latent state equation are characterized in terms of time-varying partitions of the  $n$  units over time. In order to describe such dynamics, we introduce a dependent RPM that models temporal dependence in terms of sequences of partitions by considering an auxiliary variable determining whether the partition at time  $t - 1$  will be revised for possible reallocation of the units at time  $t$ , for  $t = 1, \dots, T$ . More specifically, we introduce a binary *changepoint* auxiliary variable  $\gamma_t \in \{0, 1\}$ , to detect changes in the partitions of the  $n$  units from time  $t - 1$  to time  $t$ . We use  $|\cdot|$  to indicate the cardinality of a set, and thus denote the number of blocks composing a partition  $\pi_t$  of  $n$  units as  $|\pi_t|$ . Further, we let  $\pi_t = \{C_{1,t}, \dots, C_{|\pi_t|,t}\}$ , where, for each  $j = 1, \dots, |\pi_t|$ , the set  $C_{j,t}$  consists of the indices corresponding to the statistical units assigned to the  $j$ -th cluster according to  $\pi_t$ . Thus, given a partition  $\pi_{t-1}$  at time  $t - 1$ , we assume a *partition-based state equation* characterized as a mixture over two partition models, corresponding to the case of conditionally independent and identical random partitions at subsequent time points. Namely,

$$\begin{aligned} \pi_t \mid \boldsymbol{\pi}_{1:(t-1)}, \boldsymbol{\gamma}_{2:t} &\sim (1 - \gamma_t) \delta_{\pi_{t-1}}(\pi_t) + \gamma_t p^*(\pi_t), \quad t = 2, \dots, T \\ \pi_1 &\sim p^*(\pi_1), \end{aligned} \quad (3)$$

where  $\boldsymbol{\pi}_{1:(t-1)} = (\pi_1, \dots, \pi_{t-1})$  is the vector of previously recorded partitions,  $\boldsymbol{\gamma}_{2:t} = (\gamma_2, \dots, \gamma_t)$  indicates the vector of changepoints, and  $p^*(\cdot)$  is a distribution over the space of partitions of  $n$  units, henceforth referred to as the *base partition distribution*. In addition, we assume the changepoints  $\gamma_t$  in (3) are distributed according to

$$\gamma_t \stackrel{\text{ind}}{\sim} \text{Bern}(\eta_t), \quad t = 2, \dots, T. \quad (4)$$

The probability of a changepoint,  $\eta_t$ , can then be seen as a dependence parameter, with the extreme case  $\eta_t = 0$  leading to almost surely identical partitions at time  $t - 1$  and  $t$ . In the remaining of the work, we will call *partition state model* (PSM) the model for the random partitions  $\boldsymbol{\pi}_{1:T}$  defined through (3) and (4). Alternative partition-based state equations, and thus partition state models, can potentially be defined, to describe more complex forms of temporal dependence. See Section 6 for more discussion. The LLDPM is thus obtained by combining the PSM with (2), where, as standard in Bayesian nonparametric models, we assume that at each time point  $t = 1, \dots, T$ , the values of the parameters  $\beta_{i,t}$ , with  $i = 1, \dots, n$ , coincide within a cluster, although the specific values could differ at different times. If we let  $\boldsymbol{\beta}_t^* = (\beta_{1,t}^*, \dots, \beta_{|\pi_t|,t}^*)$  denote the vector of  $|\pi_t|$  distinct values in  $\boldsymbol{\beta}_t$ , as determined by  $\pi_t$ , such assumption can be formalized as  $\boldsymbol{\beta}_t^* \sim \prod_{j=1}^{|\pi_t|} P_0(\beta_{j,t}^*)$ , for some base probability distribution  $P_0(\cdot)$ .

## 2.1 Prior properties of the PSM

We assume that the base partition distribution  $p^*(\cdot)$  is the one implied by the class of Gibbs-type priors (De Blasi et al., 2013). Distributions over partitions are conveniently described through the exchangeable partition probability function (EPPF), a simple way to define probabilistic partition models based on the number and sizes of blocks, independently of the object labels. De Blasi et al. (2013) show that exchangeable product partition models with probability of each partition depending only on the cardinality of each cluster coincide with the family of Gibbs-type priors. A Gibbs-type EPPF has the following form,

$$\mathbb{P}_\theta [\pi_t = \{C_{1,t}, \dots, C_{|\pi_t|,t}\}] = V_{n,|\pi_t|} \prod_{j=1}^{|\pi_t|} \frac{\Gamma(|C_{j,t}| - \sigma)}{\Gamma(1 - \sigma)}, \quad (5)$$

with  $-\infty \leq \sigma < 1$ , and where, for any  $n \geq 1$  and any  $k = 1, \dots, n$ , the set of non-negative weights  $V_{n,k}$  satisfies the following recursive equation  $V_{n,k} = (n - \sigma k)V_{n+1,k} + V_{n+1,k+1}$ , and is such that  $V_{1,1} = 1$ . Notable elements of the class of Gibbs-type priors are the Dirichlet and the Pitman–Yor processes (Ferguson, 1973; Pitman and Yor, 1997), mixtures of symmetric Dirichlet distributions (Gnedin and Pitman, 2005), the normalized inverse Gaussian processes (Lijoi et al., 2005) and the normalized generalized gamma process (Lijoi et al., 2007). Throughout the paper, we use the Chinese restaurant process (CRP, Pitman, 2006) and its generalization, the two-parameter CRP (2-CRP, Pitman, 2006), as running examples for the base partition distribution  $p^*(\cdot)$ . The CRP, denoted  $p_{\text{CRP}}(\cdot)$ , is parametrized by a concentration parameter  $\theta > 0$  and corresponds to distribution on the partitions induced by the Dirichlet process. The corresponding EPPF is readily obtained by setting  $\sigma = 0$  and  $V_{n,|\pi_t|} = \theta^{|\pi_t|}/(\theta)_n$  in (5), where  $(\theta)_n = \theta(\theta+1) \cdots (\theta+n-1)$  denotes the ascending factorial. The 2-CRP, denoted  $p_\sigma(\cdot)$ , introduces an additional discount parameter  $\sigma \in [0, 1)$ , with  $\theta > -\sigma$ , and coincides with the distribution on the partitions associated with the Pitman–Yor process (Perman et al., 1992; Pitman and Yor, 1997). The corresponding EPPF is obtained by setting  $V_{n,|\pi_t|} = \prod_{j=1}^{|\pi_t|} (\theta + j\sigma)/(\theta + 1)_{n-1}$  in (5). For each specification within the class of Gibbs-type priors, the weights  $V_{n,k}$  can be computed recursively or estimated via Monte Carlo (see Arbel et al., 2017).

We study the temporal dependence of random partitions implied by the proposed PSM. We begin by observing that, once the changepoints  $\gamma_{2:T}$  are marginalized out, the distribution of  $\pi_{1:T}$  implied by the PSM is given by

$$\pi_{1:T} \sim p^*(\pi_1) \prod_{t=2}^T [(1 - \eta_t) \delta_{\pi_{t-1}}(\pi_t) + \eta_t p^*(\pi_t)].$$

Figure 2 shows the average Adjusted Rand Index (ARI) (Rand, 1971) between time-lagged partitions, similarly to Figure 2 in Page et al. (2022), assuming  $T = 10$ , for several values of the temporal dependence parameter,  $\eta_t = \eta \in \{0, 0.05, 0.25, 0.5, 0.75, 0.95\}$ ,  $t = 2, \dots, T$ . For each value of  $\eta$  and for each pair of times  $t_1, t_2 \in \{1, \dots, T\}$ , the average lagged ARI, comparing  $\pi_{t_1}$  and  $\pi_{t_2}$ , is computed over 10,000 partitions simulated from

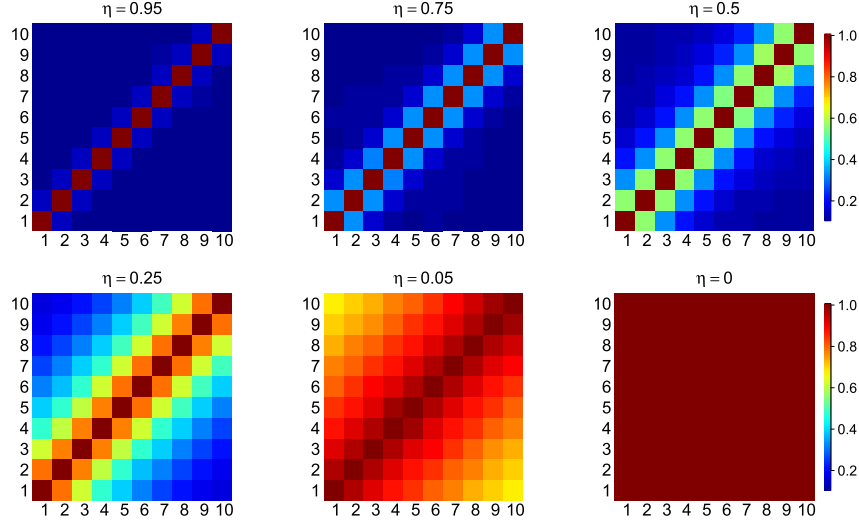


Figure 2: Average lagged ARI for the pairwise comparison of  $T = 10$  random partitions  $\pi_{1:T}$  modeled according to the proposed PSM with  $p^*(\cdot) = p_{\text{CRP}}(\cdot)$ . From top-left panel to bottom-right panel,  $\eta$  ranges in  $\{0.95, 0.75, 0.5, 0.25, 0.05, 0\}$ . For each matrix, the pixel in position  $(i, j)$  refers to the comparison of  $\pi_i$  and  $\pi_j$ . For each value of  $\eta$ , values of the lagged ARI are averaged over a sample of 10,000 partitions. The temporal dependence increases as the temporal dependence parameter  $\eta$  decreases.

the proposed PSM with  $p^*(\cdot) = p_{\text{CRP}}(\cdot)$ . Figure 2 shows increasing temporal dependence of the lagged partitions as  $\eta$  decreases. The dependence across partitions at different times can be investigated more formally by computing the expected Rand Index (ERI). For two random partitions  $\pi_{t_1}$  and  $\pi_{t_2}$ , the Rand Index is a random variable defined as  $R(\pi_{t_1}, \pi_{t_2}) = (A_{t_1, t_2} + B_{t_1, t_2}) / \binom{n}{2}$ , where  $A_{t_1, t_2}$  is the number of pairs of observations that are in the same cluster in both partitions, and  $B_{t_1, t_2}$  the number of pairs that are in different clusters in both partitions. The corresponding ERI,  $\varphi_{t_1, t_2}$ , is then defined as

$$\varphi_{t_1, t_2} = \mathbb{E}[R(\pi_{t_1}, \pi_{t_2})] = \sum_{\pi_{t_1}, \pi_{t_2} \in \mathcal{P}} R(\pi_{t_1}, \pi_{t_2}) p(\pi_{t_1}, \pi_{t_2}),$$

where  $\mathcal{P}$  indicates the space of all possible partitions of  $n$  elements, and  $p(\pi_{t_1}, \pi_{t_2})$  stands for the joint distribution of the random partitions  $\pi_{t_1}$  and  $\pi_{t_2}$ . We obtain the ERI for two random partitions whose joint distribution is defined through the proposed PSM and the base partition distribution  $p^*(\cdot)$  implied by a Gibbs-type prior. We first focus on partitions at subsequent time points, that is  $t_2 = t_1 + 1$ , and then consider the more general case  $t_1 < t_2$ .

**Proposition 1.** *Suppose  $t_2 = t_1 + 1$  and  $\eta_{t_1} = \eta_{t_2} = \eta$ . Let  $p^*(\cdot)$  be the distribution over partitions implied by a Gibbs-type prior. Then the ERI for two random partitions*

$\pi_{t_1}$  and  $\pi_{t_2}$  modeled according to a PSM is given by

$$\varphi_{t_1, t_2} = 1 - 2V_{2,2}(1 - V_{2,2})\eta. \quad (6)$$

The ERI is thus a decreasing function of the dependence parameter value  $\eta$ . Moreover, as  $\eta \rightarrow 0$ , the ERI approaches one, its maximum value, indicating that the partitions are identical. Conversely, as  $\eta \rightarrow 1$ , the ERI converges to  $1 - 2V_{2,2}(1 - V_{2,2})$ , which is the ERI of two independent and identically distributed partitions with distribution implied by a Gibbs-type prior. The expression in (6) can be rewritten as a convex linear combination of these two limiting cases, that is  $\varphi_{t_1, t_2} = (1 - \eta) + [1 - 2V_{2,2}(1 - V_{2,2})]\eta$ .

When  $p^*(\cdot) = p_{\text{CRP}}(\cdot)$  or  $p^*(\cdot) = p_\sigma(\cdot)$ , the ERI for two random partitions at subsequent times modeled according to the proposed PSM is obtained by substituting the appropriate specification of  $V_{2,2}$  in Proposition 1, which leads to the following corollary.

**Corollary 1.** *Suppose  $t_2 = t_1 + 1$  and  $\eta_{t_1} = \eta_{t_2} = \eta$ . Then the ERI for two random partitions  $\pi_{t_1}$  and  $\pi_{t_2}$  modeled according to a PSM is given by*

$$\varphi_{t_1, t_2} = 1 - \frac{2\theta}{(\theta + 1)^2}\eta \quad \text{if } p^*(\cdot) = p_{\text{CRP}}(\cdot), \quad (7)$$

$$\varphi_{t_1, t_2} = 1 - \frac{2(\theta + \sigma)(1 - \sigma)}{(\theta + 1)^2}\eta \quad \text{if } p^*(\cdot) = p_\sigma(\cdot). \quad (8)$$

The expression for the ERI in (7) coincides, up to a reparametrization of the dependence parameter  $\eta$ , with an analogous result derived in Dombowsky and Dunson (2025). As expected, the ERI in (7) is recovered by setting  $\sigma = 0$  in (8). Interestingly, the ERI approaches one both when  $\theta \rightarrow -\sigma$  and when  $\theta \rightarrow \infty$ . This behavior is not surprising as, when  $\theta$  approaches  $-\sigma$ , the distributions of both random partitions tend to assign all their mass to the partition with only one block. Similarly, when  $\theta$  increases, the two distributions accumulate mass at the partition composed of  $n$  blocks of size one. In both cases, the two random partitions tend to take the same configuration, regardless of the value taken by the dependence parameter  $\eta$ . Also in this case, (7) and (8) can be written as convex linear combinations of the ERI of two identical random partitions, that is one, and the ERI of two independent and identically distributed random partitions, that is  $\varphi_{t_1, t_2} = (\theta^2 + 1)/(\theta + 1)^2$  if  $p^*(\cdot) = p_{\text{CRP}}(\cdot)$  and  $\varphi_{t_1, t_2} = (\theta^2 + 1)/(\theta + 1)^2 + 2\sigma(\theta + \sigma - 1)/(\theta + 1)^2$  if  $p^*(\cdot) = p_\sigma(\cdot)$ .

The next proposition describes how the dependence between two lagged random partitions modeled through the proposed PSM changes as a function of the lag  $t_2 - t_1$ .

**Proposition 2.** *Suppose  $t_1 < t_2$  and let  $\eta_{t_1} = \eta_{t_2} = \eta$ . Let  $p^*(\cdot)$  be the distribution over partitions implied by a Gibbs-type prior. Then the ERI for two random partitions  $\pi_{t_1}$  and  $\pi_{t_2}$  modeled according to a PSM is given by*

$$\varphi_{t_1, t_2} = 1 - 2V_{2,2}(1 - V_{2,2}) [1 - (1 - \eta)^{t_2 - t_1}].$$

The ERI is thus a decreasing function of the time lag  $t_2 - t_1$ . Furthermore, regardless of the value of the dependence parameter  $\eta$ , the limit of  $\varphi_{t_1, t_2}$  as  $(t_2 - t_1) \rightarrow \infty$

tends towards the ERI of two independent random partitions with identical distribution implied by a Gibbs-type prior. The next corollary specializes the previous result to the cases  $p^*(\cdot)$  being a CRP or a 2-CRP.

**Corollary 2.** *Suppose  $t_1 < t_2$  and let  $\eta_{t_1} = \eta_{t_2} = \eta$ . Then the ERI for two random partitions  $\pi_{t_1}$  and  $\pi_{t_2}$  modeled according to a PSM is given by*

$$\varphi_{t_1, t_2} = 1 - \frac{2\theta}{(\theta + 1)^2} [1 - (1 - \eta)^{t_2 - t_1}] \quad \text{if } p^*(\cdot) = p_{CRP}(\cdot), \quad (9)$$

$$\varphi_{t_1, t_2} = 1 - \frac{2(\theta + \sigma)(1 - \sigma)}{(\theta + 1)^2} [1 - (1 - \eta)^{t_2 - t_1}] \quad \text{if } p^*(\cdot) = p_\sigma(\cdot). \quad (10)$$

We next explore the marginal distribution of the components of  $\pi_{1:T}$ , under the assumption of a PSM for  $\pi_{1:T}$ , with generic base partition distribution  $p^*(\cdot)$ . We prove that at each time  $t = 1, \dots, T$ , the marginal distribution of the random partition  $\pi_t$  is the same as that of the base partition distribution  $p^*(\cdot)$ . Thus, although the PSM explicitly models the evolution of the random partitions  $\pi_{1:T}$  over time, the components of  $\pi_{1:T}$  are marginally identically distributed according to  $p^*(\cdot)$ . Importantly, this result holds regardless of the specific distribution  $p^*(\cdot)$ , which therefore could be specified also outside the class of Gibbs-type priors considered here.

**Proposition 3.** *Let  $\pi_{1:T}$  be modeled according to a PSM. Then, for every  $t = 1, \dots, T$ , the marginal distribution of the random partition  $\pi_t$  coincides with the base partition distribution  $p^*(\cdot)$ .*

## 2.2 Multiview clustering representation of the LLDPM

In many real-world applications, data can be collected from various sources or represented using different feature sets, creating multiple views of the same underlying entities, with each view defined on its own support. In this section, we present an alternative representation of our model that establishes a hierarchical structure in the dependence of partitions, deviating from the assumption of temporally dependent partitions where each partition depends on the previous one. This hierarchical representation effectively describes changes in the partition across different groups or experimental conditions, offering a flexible and interpretable approach to modeling and understanding partition relationships. Such representation exemplifies a multiview clustering model, bearing close connections to the partition introduced by Dombowsky and Dunson (2025) and related to the multiview representation in Franzolini et al. (2023). As noted below (7), the same expected dependence between partitions induced by the hierarchical model for multiview clustering in Dombowsky and Dunson (2025) can be achieved by our model by reparametrizing the dependence parameter  $\eta$ . In their prior specification, dependence is introduced through the concentration parameter of a Dirichlet process, which determines the tendency of data points to form distinct groups. The main difference between our model and theirs lies in the way the dependence is induced. While their models specify a dependence structure on the atoms within a hierarchical model on a product space, our model introduces dependence directly at the level of partitions, providing a more streamlined approach to modeling partition relationships.

Without loss of generality, we begin by examining the case  $T = 2$ , where  $T$  can now be interpreted as the number of different views, groups or experimental conditions, depending on the context of the application. Thus, we focus on the joint distribution of  $\pi_1$  and  $\pi_2$ , as implied by the PSM, which is

$$\boldsymbol{\pi}_{1:2} \sim p^*(\pi_1) [(1 - \eta)\delta_{\pi_1}(\pi_2) + \eta p^*(\pi_2)], \quad (11)$$

where  $\eta$  is used to denote the dependence parameter. This distribution can be obtained as the marginal distribution of a hierarchical random partition model. To see this, we consider a third random partition distributed according to the base partition distribution, namely

$$\tilde{\pi} \sim p^*(\tilde{\pi}). \quad (12)$$

In this alternative representation,  $\tilde{\pi}$  serves as a common parent partition for both  $\pi_1$  and  $\pi_2$ , by assuming the following generative model for  $\pi_1$  and  $\pi_2$ ,

$$\pi_t \mid \tilde{\pi}, \tilde{\gamma}_1, \tilde{\gamma}_2 \stackrel{\text{ind}}{\sim} (1 - \tilde{\gamma}_t) \delta_{\tilde{\pi}}(\pi_t) + \tilde{\gamma}_t p^*(\pi_t), \quad t = 1, 2, \quad (13)$$

where  $\tilde{\gamma}_t \stackrel{\text{iid}}{\sim} \text{Bern}(\tilde{\eta})$ ,  $t = 1, 2$ , controls the whether  $\pi_t$  coincides with the parent partition  $\tilde{\pi}$ . Higher values of  $\tilde{\eta}$  increase the chances of  $\pi_t$  being different from  $\tilde{\pi}$ , allowing for more flexibility and individual variations in the partition structure. Conversely, a lower value of  $\tilde{\eta}$  makes it more likely that  $\pi_t$  will follow the same partition structure as  $\tilde{\pi}$ , enforcing a stronger dependence between the two. Note that, in model (12)–(13), the distribution of  $\pi_2$  is independent of that of  $\pi_1$ , given the partition  $\tilde{\pi}$ . The joint distribution of the random partitions  $\pi_1$  and  $\pi_2$ , the common parent partition  $\tilde{\pi}$ , and the dependence variables  $(\tilde{\gamma}_1, \tilde{\gamma}_2)$ , here denoted by  $p(\boldsymbol{\pi}_{1:2}, \tilde{\pi}, \tilde{\gamma}_1, \tilde{\gamma}_2)$ , can thus be written as

$$p(\boldsymbol{\pi}_{1:2}, \tilde{\pi}, \tilde{\gamma}_1, \tilde{\gamma}_2) \sim p^*(\tilde{\pi}) \left\{ \prod_{t=1}^2 [(1 - \tilde{\gamma}_t) \delta_{\tilde{\pi}}(\pi_t) + \tilde{\gamma}_t p^*(\pi_t)] \tilde{\eta}^{\tilde{\gamma}_t} (1 - \tilde{\eta})^{1 - \tilde{\gamma}_t} \right\}.$$

The distribution of  $\pi_1$  and  $\pi_2$  is obtained by marginalizing the last expression with respect to  $\tilde{\pi}$  and  $(\tilde{\gamma}_1, \tilde{\gamma}_2)$ , which gives (11), provided that  $\tilde{\eta} = 1 - \sqrt{1 - \eta}$ . See Section A.4 in the Supplementary Material (Giampino et al., 2025). Hence, the local level partition model's joint distribution on the partitions can be viewed as a special case of a hierarchical partition model, when considered marginally. This representation highlights the connection between the time-varying formulation of the PSM, as specified through (3)–(4), and an exchangeable model on the partitions, under the assumptions of identical marginal distributions of the partitions,  $p^*(\pi_t)$ , and equal Bernoulli probabilities  $\tilde{\eta}$  across groups. This hierarchical representation can be extended to accommodate  $T \geq 2$  groups, with one parent random partition driving the distribution of  $T$  identically distributed random partitions. Moreover, the assumption of a common Bernoulli probability  $\tilde{\eta}$  across groups can be relaxed to accommodate for group-specific parameters  $\tilde{\eta}_t$ , with  $t = 1, \dots, T$ . Expressing the local level partition model as a hierarchical model provides an alternative viewpoint to appreciate its flexibility in capturing both similarities and dissimilarities among the partitions. The hierarchical structure allows for the sharing of information across groups while still accommodating individual variations in the partition structures.

### 3 Posterior inference

Posterior inference for the parameters of the LLDPM is carried out using a Markov chain Monte Carlo (MCMC) algorithm and, more specifically, a Gibbs sampling scheme. For computational convenience, we marginalize with respect to the local level parameters  $\beta_t$ , thus introducing, for any  $t = 1, \dots, T$ , the conditional distribution of the  $n$ -dimensional vector of observations  $\mathbf{Y}_t$ , given the partition  $\pi_t$ , namely

$$p(\mathbf{Y}_t | \pi_t) = \prod_{j=1}^{|\pi_t|} \int \prod_{i \in C_{j,t}} p(y_{i,t} | \beta_{j,t}^*) P_0(d\beta_{j,t}^*).$$

The algorithm comprises three main steps, including a joint update of the random partition  $\pi_t$  and the changepoint  $\gamma_t$  from their full conditional distribution. The updates also include the dependence parameters  $\eta_t$ , for which we specify a hyperprior distribution as  $\eta_t \stackrel{\text{iid}}{\sim} \text{Beta}(a, b)$ . The steps of the implemented algorithm are described in Section B of the Supplementary Material (Giampino et al., 2025), along with additional details on the posterior distributions and our implementation.

Changepoint detection inherently involves making multiple comparisons, since the decisions are temporally dependent. To address this multi-comparison problem, we use a compound decision-theoretic approach to detect the presence of a changepoint, which is based on a loss function that takes simultaneously into account the sequence of decisions, and it is defined as a linear combination of measures of the false positive and true positive (or, alternatively, false negative) decisions (Sun and Cai, 2007). In a Bayesian context, Müller et al. (2004) and Müller et al. (2006) have demonstrated that, under the assumption of both independent hypotheses and independent (marginal) loss functions, the optimal approach for minimizing the resulting posterior expected loss involves thresholding the posterior probabilities of the changepoint ( $\text{PPC}_t$ ),  $\Pr(\gamma_t = 1 | \mathbf{Y}_{1:T})$ , for  $t = 2, \dots, T$ , as estimated from the MCMC output. However, such a procedure does not inherently control for a given false discovery rate (FDR) unless such control is explicitly accounted for. That is, we need to determine the optimal threshold in order to control the FDR at a specific desired level, say  $\zeta$ . More in detail, we consider the Bayesian FDR (BFDR, Newton et al., 2004),

$$\text{BFDR}_m(h) = \frac{\sum_{t=2}^T (1 - \text{PPC}_t) \mathbb{1}_{\{\text{PPC}_t > h\}}}{\max\left(\sum_{t=2}^T \mathbb{1}_{\{\text{PPC}_t > h\}}, 1\right)}, \quad (14)$$

where  $h$  is the chosen threshold and  $\mathbb{1}_{\{\text{PPC}_t > h\}}$  is an indicator function such that  $\mathbb{1}_{\{\text{PPC}_t > h\}} = 1$  if  $\text{PPC}_t > h$ , and 0 otherwise. The optimal threshold  $h^*$  corresponds to the minimum value of  $h$  that ensures that the  $\text{BFDR}_m$  is less than  $\zeta$ . In formulas,  $h^* = \min\{h : \text{BFDR}_m(h) \leq \zeta\}$ . The previous testing procedure can be classified as a *marginal* approach (hence, the subscript in  $\text{BFDR}_m$ ) since it fails to consider existing dependence either among hypotheses or in the decisions themselves. Sun et al. (2015) extended this framework to the spatial setting, explicitly accounting for dependence among hypotheses, as induced by a spatial model. More recently, Chandra and Bhattacharya (2019) introduced *non-marginal* loss functions and non-marginal decision rules

that directly account for dependence in the decision-making process. Their procedure incorporates additional information about dependence among tests into the definition of error and non-error terms for subgroups of hypotheses. Specifically, the approach penalizes each hypothesis based on incorrect decisions related to other dependent tests, thus defining a compound loss where decisions about dependent tests are interrelated. We adapt their framework to our setting, and the resulting changepoint detection procedure is presented in Section C of the Supplementary Material (Giampino et al., 2025).

## 4 Simulation studies

We present two simulation studies to illustrate the performance of our LLDPM under different data-generating mechanisms. We consider two versions of the LLDPM, defined by the choice of base partition distribution. In the first version, denoted LLDPM<sub>0</sub>, we set  $p^*(\cdot) = p_{\text{CRP}}(\cdot)$ , corresponding to a CRP. In the second version, denoted LLDPM<sub>0,25</sub>, we use a 2-CRP and set  $p^*(\cdot) = p_{\sigma}(\cdot)$  with a discount parameter  $\sigma = 0.25$ . Such choice of  $\sigma$  is arbitrary and simply intended to demonstrate the applicability of our method when the base partition distribution differs from the CRP. While other choices are possible, additional investigations, not reported here, suggest that the method is rather robust to such variations. We then investigate the models ability to accurately detect changepoints in both independent and autocorrelated data scenarios, while also recovering the time-specific latent cluster structure. For comparison, we evaluate our models against six alternatives: (1) the Dependent Random Partition Model (DRPM) introduced by Page et al. (2022), which, to our knowledge, is the only model-based approach that introduces time dependence directly through partitions; (2) the Linear Dependent Dirichlet Process (LDDP) described by De Iorio et al. (2004), which incorporates time explicitly within the atoms of the dependent process; (3) the Weighted Dependent Dirichlet Process (WDDP) by Quintana et al. (2022), which integrates time into the weights of the Dirichlet process; (4) the Griffiths–Milne Dependent Dirichlet Process (GMDDP) introduced by Lijoi et al. (2014) and implemented in the R package `BNPmix` (Corradin et al., 2021), which defines multivariate vectors of dependent and identically distributed Dirichlet processes; (5) the Product Partition Model (PPM) implemented in the R package `ppmSuite` (Page and Quinlan, 2022), a Bayesian model where the set of changepoint indicators between time series are correlated; (6) the Autoregressive Dirichlet Process (AR1-DP) introduced by De Iorio et al. (2023), which adopts an autoregressive prior to allow for temporal dynamic clustering.

It is important to emphasize that among the six alternatives considered, only the DRPM is explicitly designed to model the evolution of random partitions, as is the case for our proposed model. The remaining models were not devised for this purpose. Specifically, the PPM is designed for multivariate time series and detects changepoints as shifts in the mean of the process rather than changes in partitions. The LDDP, WDDP, and GMDDP models are adapted for changepoint detection by organizing the data as a single vector and incorporating time as a covariate, following the approach of Page et al. (2022). The AR1-DP considers a sequence of autoregressive random probability measures over time, but does not explicitly incorporate changepoints detection. To identify changepoints when using models that are not specifically designed to detect

changes in the partition evolution over time, we rely on a similarity-based criterion that compares cluster configurations between consecutive time points. Given that recording a changepoint any time the two partitions are not identical would lead to an overidentification of changepoints, our approach allows for a more flexible assessment of partition similarity. This criterion is applied to the competitor models DRPM, LDDP, WDDP, GMDDP, and AR1-DP. Specifically and only for these methods, a changepoint is identified when the ARI between consecutive partitions falls below 85%, in combination with the same FDR correction adopted for our LLDPM. This adjustment leads to improved performance for these methods. See Section D.1 of the Supplementary Material (Giampino et al., 2025), for details. For the PPM model, we adopt a different strategy and rely on the estimated individual changepoints, as the model does not account for time-specific partitions.

We evaluate the performance of the above methods on two key tasks: accurately estimating the true partition of subjects at each time point and correctly identifying changepoints in partition evolution over time. For the first task, we use the ARI to compare the true partition with the one estimated by minimizing the lower bound of the posterior expected Variation of Information (VI, Wade and Ghahramani, 2018). For the second task, we use six standard metrics, namely specificity, accuracy, recall, precision, F1-score, and area under the curve (AUC), as defined in Section D.1 of the Supplementary Material (Giampino et al., 2025). For the PPM model, these metrics are first computed at the individual changepoint level and then averaged across the entire sample. Finally, due to its autoregressive structure, the AR1-DP is applied exclusively in the autoregressive data scenario of Section 4.2.

## 4.1 Simulations with independent data

We begin by simulating independent data characterized by similar partition structures over different intervals of time, trying to mimic the data-generating process characterizing the LLDPM. We evaluate performance across 50 replicated datasets and different numbers of observations, specifically  $n = \{20, 50, 100\}$ , over  $T = 100$  time points. In each scenario, we consider data partitions characterized by eight changepoints. The time axis is divided into nine consecutive blocks of varying lengths, each representing a segment of the 100 total time points. Within each block, the partition of the units remains fixed and corresponds to one of three configurations. Two of these configurations consist of three clusters and differ only in the assignment of units to clusters, while the third configuration includes just two clusters. These configurations alternate across blocks. At each time point, given the time-specific partition defined by the active configuration, observations for each cluster are drawn from a Normal distribution. The variance is fixed at 0.01 and shared across all clusters, while the mean for each cluster is independently sampled from a Normal distribution with arbitrarily chosen parameters. An example of the generated data is shown in Figure D.1 in the Supplementary Material (Giampino et al., 2025).

To analyze these data, the specification of the two versions of our LLDPM is completed by assuming  $\theta = 1$  and that the base probability distribution  $P_0(\cdot)$  is Normal

$n$	Measure	LLDPM <sub>0</sub>	LLDPM <sub>0.25</sub>	DRPM	LDDP	WDDP	GMDDP	PPM
20	specificity	0.99 (0.01)	0.97 (0.03)	0.99 (0.01)	0.91 (0.17)	0.47 (0.04)	<b>1.00</b> (0.00)	0.77 (0.01)
	accuracy	0.99 (0.01)	0.97 (0.03)	0.96 (0.01)	0.92 (0.16)	0.52 (0.03)	<b>1.00</b> (0.00)	0.72 (0.01)
	recall	<b>1.00</b> (0.00)	<b>1.00</b> (0.00)	0.63 (0.14)	<b>1.00</b> (0.00)	1.00 (0.02)	<b>1.00</b> (0.00)	0.23 (0.01)
	precision	0.90 (0.09)	0.76 (0.16)	0.89 (0.12)	0.81 (0.32)	0.14 (0.01)	<b>1.00</b> (0.00)	0.08 (0.00)
	F1	0.95 (0.05)	0.85 (0.11)	0.73 (0.12)	0.85 (0.28)	0.25 (0.01)	<b>1.00</b> (0.00)	0.13 (0.01)
	AUC	0.99 (0.01)	0.98 (0.01)	0.80 (0.08)	0.85 (0.21)	0.73 (0.05)	<b>1.00</b> (0.00)	0.50 (0.01)
50	specificity	0.96 (0.03)	0.91 (0.08)	0.99 (0.01)	0.44 (0.40)	0.29 (0.10)	<b>1.00</b> (0.03)	0.66 (0.00)
	accuracy	0.97 (0.02)	0.91 (0.07)	0.96 (0.36)	0.49 (0.36)	0.35 (0.09)	<b>1.00</b> (0.02)	0.63 (0.00)
	recall	<b>1.00</b> (0.00)	<b>1.00</b> (0.00)	0.67 (0.15)	<b>1.00</b> (0.00)	1.00 (0.02)	<b>1.00</b> (0.00)	0.34 (0.00)
	precision	0.73 (0.15)	0.56 (0.23)	0.88 (0.11)	0.34 (0.38)	0.11 (0.01)	<b>0.99</b> (0.09)	0.08 (0.00)
	F1	0.84 (0.10)	0.70 (0.18)	0.74 (0.10)	0.41 (0.35)	0.20 (0.02)	<b>0.99</b> (0.07)	0.14 (0.00)
	AUC	0.98 (0.01)	0.91 (0.14)	0.83 (0.07)	0.68 (0.20)	0.64 (0.05)	<b>1.00</b> (0.01)	0.50 (0.00)
100	specificity	0.95 (0.03)	0.83 (0.10)	<b>0.99</b> (0.01)	0.38 (0.36)	0.23 (0.04)	0.96 (0.11)	0.58 (0.00)
	accuracy	0.95 (0.03)	0.85 (0.09)	<b>0.96</b> (0.02)	0.43 (0.33)	0.28 (0.04)	0.28 (0.10)	0.56 (0.00)
	recall	<b>1.00</b> (0.00)	<b>1.00</b> (0.00)	0.60 (0.15)	<b>1.00</b> (0.00)	0.88 (0.03)	<b>1.00</b> (0.00)	0.44 (0.00)
	precision	0.66 (0.17)	0.40 (0.16)	0.81 (0.15)	0.26 (0.32)	0.09 (0.01)	<b>0.92</b> (0.24)	0.08 (0.00)
	F1	0.78 (0.12)	0.55 (0.16)	0.68 (0.13)	0.34 (0.30)	0.16 (0.01)	<b>0.93</b> (0.20)	0.14 (0.00)
	AUC	0.97 (0.02)	0.83 (0.17)	0.79 (0.08)	0.67 (0.18)	0.55 (0.02)	<b>0.98</b> (0.06)	0.52 (0.00)

Table 1: Section 4.1. Performance measures for changepoint detection with independent data, for the LLDPM under two specifications and five competing models. The values correspond to the average (standard deviation) over 50 replicates.

with mean zero and variance  $\zeta^2$ . We further specify the following hyperprior distributions:  $\tau^2 \sim \text{Inv-Gamma}(15, 3)$  and  $\zeta^2 \sim \text{Inv-Gamma}(15, 3)$ . Additionally, we set  $\eta_t \stackrel{\text{iid}}{\sim} \text{Beta}(0.1, 0.9)$ , which places greater probability on the absence of a changepoint at each time point. When fitting the DRPM model, we set the parameters as suggested in Page et al. (2022), except for the variances, which we fixed as in our model. The temporal dependence parameter in the DRPM is modeled as the analogous parameters  $\eta_t$  in the LLDPM. Given the different parameterizations of the two models, this corresponds to a Beta(0.9, 0.1) distribution. Finally, for all the other considered models, we used the same variances as in our model and specified the default hyperprior parameter values.

We implemented the Gibbs sampling scheme detailed in Section B.1 of the Supplementary Material (Giampino et al., 2025), running the algorithm for 10,000 iterations and discarding the first 5,000 as burn-in. We randomly selected a subset of replications and visually inspected the trace plots of the time-specific number of clusters and the partition entropy. These diagnostics revealed no signs of convergence issues. Examples of these trace plots are provided in Section D.2 of the Supplementary Material (Giampino et al., 2025). Subsequently, we calculated the optimal partition for each time point based on the estimated posterior similarity matrix. To determine whether a specific time point  $t$  should be considered a changepoint, we employed the Bayesian False Discovery Rate method (Newton et al., 2004; Müller et al., 2006). As explained in Section C of the Supplementary Material (Giampino et al., 2025), we implemented a penalized version of FDR with a control level set at 0.01/3.

Table 1 summarizes the performance of the seven models considered, including our LLDPM<sub>0</sub> and LLDPM<sub>0.25</sub>, and models 1–5 introduced at the beginning of Section 4, across varying data dimensions  $n$ . Notably, both versions of our model exhibit strong performance, consistently outperforming LDDP, WDDP, and PPM. All changepoint de-

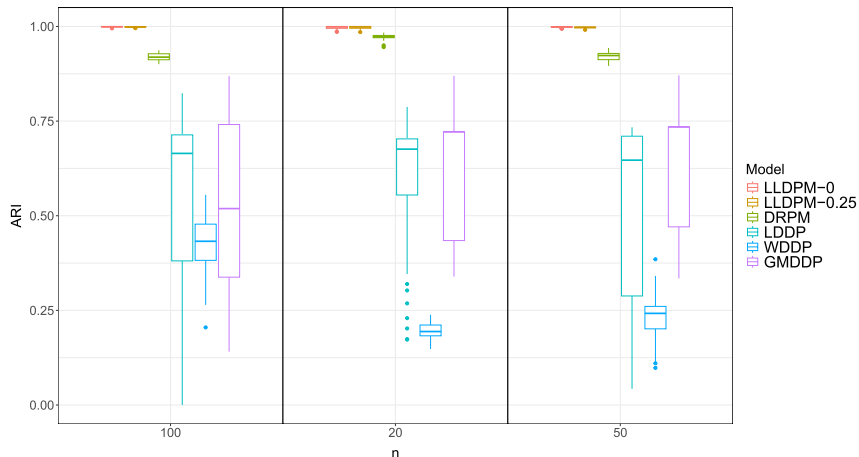


Figure 3: Section 4.1. Boxplots of the average ARI, to evaluate the clustering performance when analyzing independent data, with  $n = \{20, 50, 100\}$ , for two versions of our LLDPM model and four alternatives. The PPM is not included in the comparison as the model does not account for time-specific partitions of the units. Results are based on 50 replicated datasets and averaged over 100 time points. For each dataset and each method, at each time point, the true partition is compared with the partition estimated by minimizing the lower bound of the posterior expected VI.

tection metrics highlight the robustness of LLDPM, displaying strong and stable results regardless of the underlying base process specification, whether based on a CRP or a 2-CRP with  $\sigma = 0.25$ . Further remarks are needed for the remaining two competitors, GMDDP and DRPM. While GMDDP performs well in detecting changepoints, as reflected in the evaluation metrics, it struggles to accurately recover the true time-specific latent cluster structures, as illustrated in Figure 3. Consequently, GMDDP may not be well-suited for applications where both changepoint detection and cluster analysis are of interest. Also DRPM shows strong results in changepoint detection, achieving specificity levels comparable to those of LLDPM across all settings. However, it suffers from lower recall, indicating a tendency to miss many true changepoints. In contrast, LLDPM excels in both tasks, consistently delivering high changepoint detection accuracy and near-perfect ARI scores for latent cluster recovery, see Figure 3. Finally, we observe that as the sample size increases, performance tends to decline across all models, due to the greater likelihood of subjects switching clusters.

To evaluate the performance of our model for sample sizes larger than those considered in Table 1, we used both versions of the LLDPM to analyze 50 replicated datasets with  $n = 500$  observations over  $T = 100$  time points. Due to memory limitations and prohibitively long runtimes, the competing models could not be fitted. In contrast, for our models, we adopted a blockwise Gibbs sampling strategy to manage the computational load, splitting the sampling into smaller blocks of iterations to prevent memory and performance issues. Both  $\text{LLDPM}_0$  and  $\text{LLDPM}_{0.25}$  successfully identified the

changepoints, achieving perfect recall. However, their performance in correctly identifying periods with no-change point declined, leading to an increase in false-positives. Specifically,  $\text{LLDPM}_0$  achieved over 70% specificity and accuracy, with an AUC close to 80%. However, its F1 score dropped to around 40% due to lower precision.  $\text{LLDPM}_{0.25}$  showed a more polarized behavior with specificity and accuracy around 40%, an AUC of 70%, and an F1 score of about 25%. Despite these differences in changepoint detection metrics, both models achieved an average ARI of approximately 99.95%, indicating a near-perfect recovery of the time-specific latent cluster structures.

## 4.2 Simulations with autoregressive data

In a second simulation study, we aim to assess the performance of the two versions of our LLDPM in scenarios where the data are generated from a process characterized by an autoregressive structure of order one, or AR(1). Specifically, we consider datasets with  $T = 30$  time points and  $n = 20$  units, generated from the model

$$Y_{i,t} = \lambda Y_{i,t-1} + \tilde{\beta}_{i,t} + \tilde{\varepsilon}_{i,t}, \quad (15)$$

where the error terms  $\tilde{\varepsilon}_{i,t}$  are assumed to be independent and identically distributed from a Normal with mean zero and variance one. At each time point  $t = 1, \dots, T$ , units are partitioned based on the distinct values taken by  $\tilde{\beta}_{i,t}$ . The number of distinct values, and thus the number of clusters, varies between one and two over time. Specifically, there are two clusters when  $t$  is divisible by five or nine, and one cluster otherwise. When  $t$  is divisible by five, the two clusters are of equal size; when  $t$  is divisible by nine, one cluster is more than twice the size of the other. This setup gives rise to three distinct configurations that alternate across time points, with the allocation of units to clusters updated whenever the configuration changes. The observations are thus simulated from model (15), with the values of  $\tilde{\beta}_{i,t}$  at each time point assigned to arbitrarily fixed, cluster-specific values. An example of such data is shown in Figure D.4 in the Supplementary Material (Giampino et al., 2025). When fitting the two versions of our model, unlike in Section 4.1, we set  $\theta$  based on the expected prior number of clusters,  $\mathbb{E}[|\pi_t|]$ , for each time point  $t = 1, \dots, T$ . For the LLDPM with specification  $p^*(\cdot) = p_{\text{CRP}}(\cdot)$ , this is given by  $\mathbb{E}[|\pi_t|] = \sum_{i=1}^n \theta/\theta + i - 1$ , while for the LLDPM with  $p^*(\cdot) = p_\sigma(\cdot)$  and  $\sigma \in (0, 1)$ , by  $\mathbb{E}[|\pi_t|] = [(\theta + \sigma)_n / (\theta + 1)_{n-1} - \theta] / \sigma$  (Pitman, 2006). By setting  $\mathbb{E}[|\pi_t|] = 2$  for all  $t = 1, \dots, T$ , we obtain  $\theta = 0.32$  for  $\text{LLDPM}_0$  and  $\theta = -0.07$  for  $\text{LLDPM}_{0.25}$ . Moreover, we let the autoregressive parameter  $\lambda$  vary in  $\{0.25, 0.5, 0.75, 0.9\}$  to see if increasing the dependence across times affect the performances of the model. For the other parameters, we use the same specifications as in the simulation study of Section 4.1. As before, we run the MCMC for 10,000 iterations, discarding the first half as burn-in.

The results of our analysis, based on 50 replicates, are presented in Table 2. Both versions of the LLDPM consistently outperform most other models in detecting changepoints across all considered metrics. In this second study, we observe that GMDDP exhibits reduced effectiveness compared to the previous simulation. Among the alternatives, LDDP stands out for its strong ability to identify changepoints, although its performance declines in terms of specificity, precision, and accuracy as  $\lambda$  increases. In

$\lambda$	Measure	LLDPM <sub>0</sub>	LLDPM <sub>0.25</sub>	DRPM	LDDP	WDDP	GMDDP	PPM	ARI-DP
0.25	specificity	0.91 (0.08)	0.95 (0.06)	0.98 (0.04)	<b>1.00</b> (0.00)	0.82 (0.05)	0.87 (0.10)	0.83 (0.02)	0.64 (0.15)
	accuracy	0.91 (0.06)	0.94 (0.04)	0.52 (0.04)	<b>1.00</b> (0.02)	0.52 (0.03)	0.65 (0.07)	0.49 (0.00)	0.83 (0.07)
	recall	0.92 (0.07)	0.94 (0.06)	0.13 (0.07)	1.00 (0.04)	0.25 (0.05)	0.45 (0.07)	0.21 (0.02)	<b>1.00</b> (0.00)
	precision	0.93 (0.07)	0.96 (0.05)	0.91 (0.16)	<b>1.00</b> (0.00)	0.62 (0.09)	0.81 (0.12)	0.61 (0.01)	0.77 (0.08)
	F1	0.92 (0.06)	0.95 (0.04)	0.21 (0.11)	<b>1.00</b> (0.02)	0.35 (0.06)	0.58 (0.08)	0.31 (0.02)	0.87 (0.05)
	AUC	0.91 (0.06)	0.94 (0.04)	0.54 (0.04)	<b>1.00</b> (0.02)	0.53 (0.03)	0.66 (0.07)	0.52 (0.00)	0.82 (0.08)
0.5	specificity	0.88 (0.10)	0.94 (0.07)	0.96 (0.18)	<b>0.99</b> (0.02)	0.80 (0.06)	0.67 (0.16)	0.82 (0.02)	0.55 (0.14)
	accuracy	0.89 (0.07)	0.93 (0.05)	0.51 (0.04)	<b>1.00</b> (0.01)	0.51 (0.04)	0.58 (0.08)	0.49 (0.00)	0.79 (0.07)
	recall	0.90 (0.08)	0.93 (0.07)	0.13 (0.13)	<b>1.00</b> (0.00)	0.25 (0.06)	0.51 (0.07)	0.22 (0.03)	<b>1.00</b> (0.00)
	precision	0.90 (0.07)	0.95 (0.06)	0.84 (0.21)	<b>0.99</b> (0.02)	0.59 (0.08)	0.65 (0.12)	0.60 (0.01)	0.72 (0.07)
	F1	0.90 (0.06)	0.94 (0.05)	0.21 (0.11)	<b>1.00</b> (0.01)	0.35 (0.07)	0.57 (0.07)	0.32 (0.02)	0.84 (0.04)
	AUC	0.89 (0.07)	0.93 (0.05)	0.54 (0.04)	<b>1.00</b> (0.01)	0.52 (0.04)	0.60 (0.07)	0.52 (0.00)	0.78 (0.07)
0.75	specificity	0.86 (0.12)	0.92 (0.08)	0.94 (0.07)	<b>0.96</b> (0.08)	0.77 (0.07)	0.45 (0.25)	0.83 (0.02)	0.34 (0.14)
	accuracy	0.87 (0.06)	0.92 (0.06)	0.52 (0.04)	<b>0.98</b> (0.04)	0.49 (0.04)	0.66 (0.16)	0.49 (0.00)	0.69 (0.07)
	recall	0.88 (0.08)	0.91 (0.08)	0.16 (0.08)	<b>1.00</b> (0.00)	0.24 (0.06)	0.84 (0.14)	0.22 (0.02)	1.00 (0.01)
	precision	0.89 (0.09)	0.93 (0.07)	0.79 (0.17)	<b>0.97</b> (0.06)	0.55 (0.09)	0.65 (0.14)	0.61 (0.01)	0.64 (0.05)
	F1	0.88 (0.06)	0.92 (0.06)	0.25 (0.10)	<b>0.98</b> (0.03)	0.33 (0.07)	0.73 (0.12)	0.32 (0.02)	0.78 (0.04)
	AUC	0.87 (0.06)	0.92 (0.06)	0.53 (0.04)	<b>0.98</b> (0.04)	0.51 (0.04)	0.65 (0.16)	0.52 (0.00)	0.67 (0.07)
0.9	specificity	0.85 (0.10)	0.90 (0.10)	<b>0.95</b> (0.10)	0.49 (0.27)	0.77 (0.09)	0.37 (0.31)	0.81 (0.03)	0.22 (0.10)
	accuracy	0.87 (0.06)	<b>0.90</b> (0.06)	0.51 (0.04)	0.75 (0.13)	0.49 (0.04)	0.65 (0.16)	0.49 (0.01)	0.62 (0.05)
	recall	0.88 (0.06)	0.90 (0.07)	0.13 (0.08)	<b>0.98</b> (0.04)	0.24 (0.07)	0.91 (0.19)	0.23 (0.03)	0.97 (0.04)
	precision	0.89 (0.08)	<b>0.91</b> (0.07)	0.79 (0.17)	0.70 (0.12)	0.55 (0.12)	0.63 (0.18)	0.60 (0.01)	0.59 (0.03)
	F1	0.88 (0.07)	<b>0.90</b> (0.05)	0.22 (0.10)	0.81 (0.08)	0.33 (0.08)	0.75 (0.12)	0.33 (0.03)	0.73 (0.03)
	AUC	0.86 (0.06)	<b>0.90</b> (0.06)	0.52 (0.05)	0.73 (0.14)	0.50 (0.04)	0.65 (0.15)	0.52 (0.00)	0.60 (0.05)

Table 2: Section 4.2. Performance measures for changepoint detection with AR(1) data for two versions of our LLDPM model and six alternatives. Values correspond to the average (standard error) over 50 replicates.

contrast, both versions of the proposed LLDPM demonstrate robustness, maintaining high performance across all metrics and scenarios, regardless of the degree of temporal dependence in the simulated data. This consistency highlights the reliability of the LLDPM across a wide range of settings. Furthermore, the LDDP falls short in accurately recovering the time-specific latent cluster structure. As shown in Figure 4, both versions of the LLDPM consistently achieve ARI indices above 0.7 across datasets generated with varying autoregressive coefficients, clearly outperforming the other methods. Overall, when considering both changepoint detection and cluster structure recovery, our model provides reliable and robust inference across a wide range of scenarios.

## 5 Applications

We apply our LLDPM model, using the specification  $p^*(\cdot) = p_{\text{CRP}}(\cdot)$ , to two real-world datasets: first, the Gesture Phase Segmentation data (Madeo et al., 2013), and second, data from the Collaborative Perinatal Project (CPP, Longnecker et al., 2001). Gesture Phase Segmentation focuses on identifying and classifying different phases of hand movements, which is an important task in applications such as sign language recognition, human-computer interaction, and motion analysis. In contrast, CPP is a large-scale epidemiological study that investigates prenatal and perinatal factors affecting pregnancy outcomes. These two applications underscore the versatility of our model across different domains: the first, in Section 6, demonstrates LLDPM’s effectiveness in changepoint detection for time-series data, while the second, presented in Section 5.2, illustrates its applicability in a multiview analysis setting.

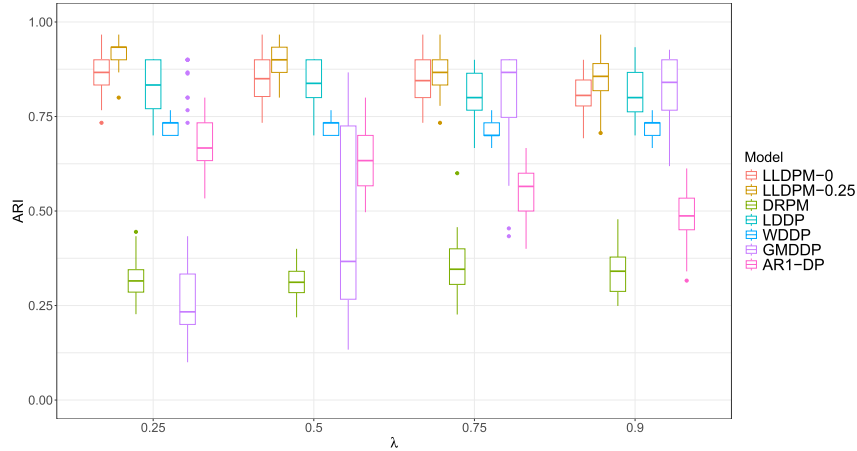


Figure 4: Section 4.2. Boxplots of the average ARI, to evaluate clustering performance when analyzing AR(1) data, comparing two versions of our model with five alternatives. The PPM is excluded from the comparison as it does not account for time-specific partitions of the units. The analysis considers different values of the autoregressive coefficient  $\lambda = \{0.25, 0.5, 0.75, 0.9\}$ . Results are based on 50 replicated datasets and averaged over 30 time points. For each dataset and each method, at each time point, the true partition is compared with the partition estimated by minimizing the lower bound of the posterior expected VI.

## 5.1 Gesture Phase Segmentation data

We present an analysis of video-recorded data for human gesture segmentation. The goal is to segment videos into distinct phases exhibiting different motion patterns, e.g., to identify time lapses of the video that need to be removed from a clip (Parvathy et al., 2021). More specifically, we employ the Gesture Phase Segmentation dataset, publicly accessible for download at the following URL: <https://archive.ics.uci.edu/ml/datasets/gesture+phase+segmentation>. While Madeo et al. (2013) analyzed this dataset using Support Vector Machines (SVM) to segment, and classify gesture data streams, they did not consider partitioning gesture phases into their components and investigating the temporal dependence between partitions. Our goal, in contrast, is to detect the latent cluster structure at each frame while capturing changepoints, using our proposed LLDPM random partition model with time dependence.

The dataset contains sensor data recordings of users recounting comic book stories facing an Xbox Microsoft Kinect<sup>TM</sup> sensor. The dataset provides scalar velocity and acceleration values over four sensors, placed on the left hand, right hand, left wrist, and right wrist, leading to  $n = 8$  sensor measurements at regular time intervals (frames). These values were obtained by normalizing hand and wrist positions relative to the head and spine position using a fixed displacement offset of 3 to measure velocity. This offset is a constant adjustment applied to account for the spatial relationship between

the sensors and the torso, which helps standardize the measurements and accurately calculate the velocity. Our analysis is based on a processed version of the data, which we prepare following the approach outlined by Hadj-Amar et al. (2024). The preprocessing consists of four main stages. First, we apply a two-point moving average filter to smooth the time series. Second, to reduce the data volume while preserving essential patterns, we downsample the series by selecting every fifth data point, following the approach suggested by Romanuke (2021). This strategy allows us to handle longer time lags while minimizing the number of parameters and reducing computational complexity. Third, we transform the data using a square root transformation, which Hadj-Amar et al. (2024) found to be effective in fitting a Gaussian distribution to the data. Finally, we standardize the time series to ensure uniformity and comparability across the dataset. The processed data is illustrated in Figure 5. For additional context, see Figure E.5 in the Supplementary Material (Giampino et al., 2025), which plots the data along with details on the video phases previously identified by Madeo et al. (2013): D (rest position, from the Portuguese “descanso”), P (preparation), S (stroke), H (hold), and R (retraction).

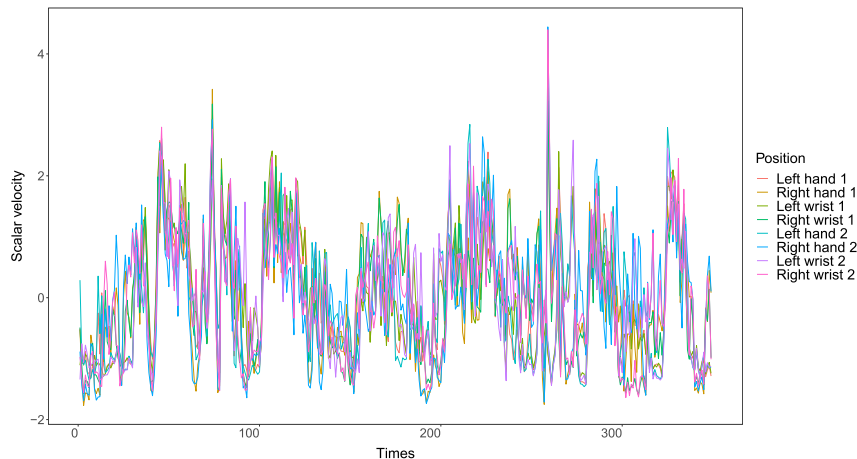


Figure 5: Section 5.1, Human Gesture data. Scalar velocity of the left and right hand and the wrists after preprocessing ( $T = 349$ ).

Here, we illustrate our model’s performance in analyzing this data. Gesture phase segmentation poses several challenges. Firstly, it is inherently subjective, with no definitive starting point for each phase, leading to varying segmentations by different experts for the same video. Additionally, similar patterns may occur across different activities; for example, stationary hand gestures may be seen in both the rest and hold phases identified by Madeo et al. (2013). Furthermore, the data may include nuisance movements, such as touching glasses while speaking, which can cause fluctuations in sensor-recorded scalar velocity (Madeo et al., 2013). To implement the LLDPM, we consider an a priori expected number of clusters equal to two at each time point, corresponding to the specification  $\theta = 0.49$  for the concentration parameter of the CRP. The prior probability

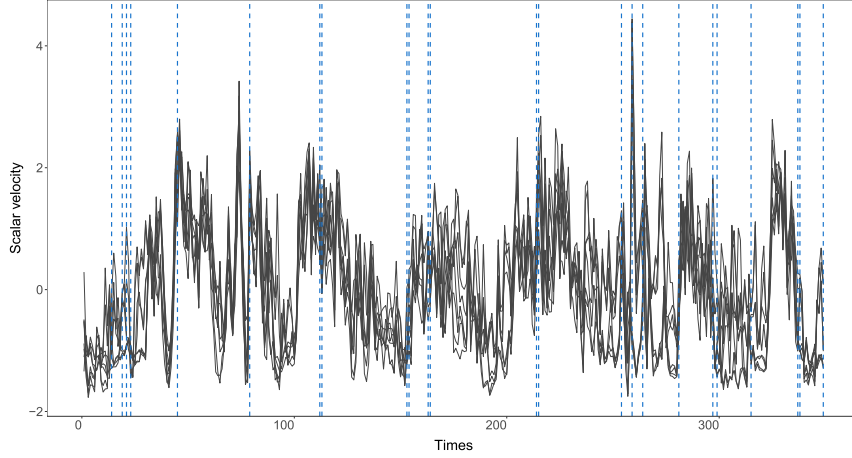


Figure 6: Section 5.1, Human Gesture data. Changepoints (dashed line) identified with the LLDPM.

of a changepoint is assumed  $\eta_t \stackrel{\text{iid}}{\sim} \text{Beta}(0.1, 0.9)$ , suggesting that the model assumes a relatively low prior probability of a changepoint at each time point. For the distribution  $P_0(\cdot)$  of the coefficients  $\beta_{i,t}$ , we assume a Normal distribution with a mean zero and variance  $\zeta^2 = 1$ . Finally, following Corradin et al. (2021), we set the variance of the kernel,  $\tau^2$ , to follow an Inverse-Gamma distribution with a scale parameter equal to the sample variance of the standardized data, i.e.,  $\tau^2 \sim \text{Inv-Gamma}(2, 1)$ . We run our Gibbs sampling algorithm for 10,000 iterations, with the first half discarded as burn-in.

The estimated changepoints are displayed in Figure 6 and appear spread throughout the time series, possibly indicating that the person wearing the sensors is gesticulating while recalling the comic story. When compared with the segmentation in Madeo et al. (2013), the changepoints identified by the LLDPM primarily occur during the activity phase, which correspond to the preparation, retraction, and stroke phases in the video. Additionally, most of these changepoints are observed during transitions between different phases, with a notable increase in frequency during the reading phases. These changepoints likely reflect movements related to the subject’s gesticulation and body language while narrating, as well as shifts in hand and wrist positions during the assigned activity. Figure 7 illustrates the clustering of the eight sensors within a specific time window, chosen for representation purposes. The estimated clustering is obtained by minimizing the lower bound of the posterior expected VI. The numbers on the plot correspond to the sensor measurements. Notably, a distinct pattern allocation emerges, where all the sensors are grouped into a single primary cluster. Additionally, a secondary pattern takes shape, forming two clusters: one containing all the sensors from the left arm (odd numbers) and the other with sensors from the right arm (even numbers) grouped together.

For comparison, we analyzed the same data with four of the alternative models considered in this work. The identified changepoints are displayed in Figures E.6, E.7,

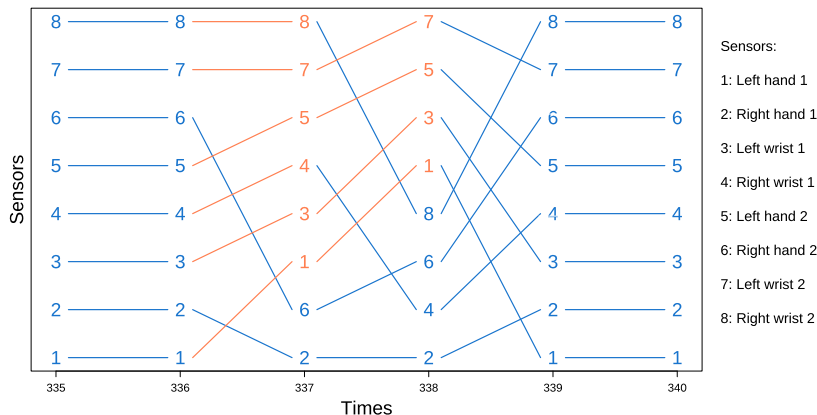


Figure 7: Section 5.1, Human Gesture Data. Time-specific clusters obtained by minimizing the lower bound of the posterior expected VI, from time 335 to time 340. Numbers refer to the measurements of scalar velocity (1-4) and acceleration (5-8) in sensor units placed on the hands and wrists of a subject, whereas colors indicate clusters.

E.8 and E.9 in the Supplementary Material (Giampino et al., 2025). It is notable that partition-based models appear to identify fewer changepoints. For instance, the DRPM detected only 15 changepoints (4.3% of the total time points), primarily at the beginning of the time series, and the LLDPM identified 24 changepoints (6.9%). In contrast, the other alternative models, namely the LDDP, WDDP and GMDDP, identified approximately 14% of the changepoints, with 48, 50 and 51 changepoints, respectively, out of 349 time points. These observations instill confidence in the performance of the LLDPM, which thus might represent a reliable modeling option, particularly in real-world applications where the ground truth is unknown.

## 5.2 Collaborative Perinatal Project data

The CPP data were collected to investigate the potential effects of Dichlorodiphenyl-dichloroethylene (DDE) exposure on pregnancy outcomes. In addition to maternal serum DDE concentration, the dataset includes key variables for each enrolled pregnant woman, such as hospital of admission, smoking status, gestational age (in days), and newborn birth weight (in grams). It has been widely used in epidemiological research to examine the impact of prenatal environmental exposures on neonatal health, offering valuable insights for public health policies and maternal care strategies (see, e.g., Klebanoff, 2009). The original study by Longnecker et al. (2001) investigated the relationship between DDE exposure and two primary outcomes: preterm delivery and small-for-gestational-age births.

In our analysis, we adopt the multiview framework outlined in Section 2.2, applying the LLDPM model to both gestational age and newborn birth weight. This approach treats the two measurements as distinct views of the same subjects. We focus on a subset

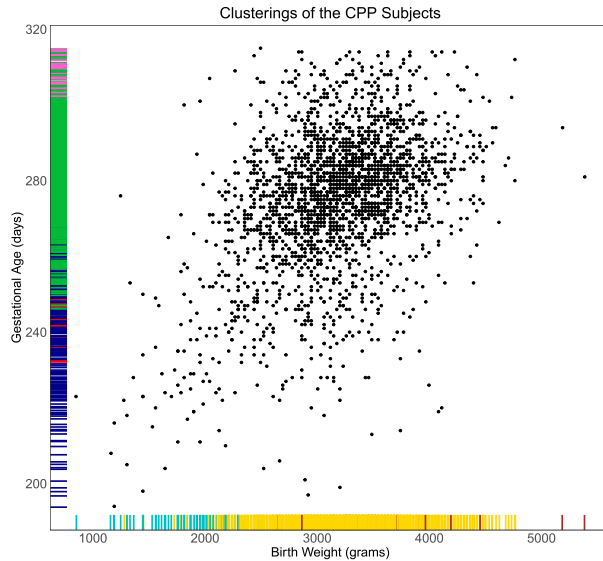


Figure 8: Section 5.2, CPP data. Scatterplot of gestational age (days) and birth weight (grams), with colors indicating cluster assignment.

of the CPP dataset, available in the `BNPmix` package (Corradin et al., 2021), which includes data from 2,312 women. The analysis is first conducted on the full dataset and then stratified by hospital. In all cases, we assume a prior number of cluster equal to two and obtain  $\theta = 0.12$  when  $n = 2,312$ . The prior probability of a change between the partitions of the two views is assumed  $\eta \stackrel{\text{iid}}{\sim} \text{Beta}(0.1, 0.9)$ . For the distribution  $P_0(\cdot)$  of the coefficients  $\beta_{i,j}$ , we assume a Normal distribution with a mean zero and variance  $\zeta^2 = 1$ , and we place an inverse-gamma prior on the variance parameter,  $\tau^2 \sim \text{Inv-Gamma}(2, 1)$ . We run our Gibbs sampling algorithm for 10,000 iterations, with the first half discarded as burn-in. Figure 8 shows the scatterplot with the inferred view-specific clustering of the subjects displayed on the axis, estimated by minimizing the lower bound of the posterior expected VI (Wade and Ghahramani, 2018). The clusters identified from the first view (gestational age) and the second view (birth weight) exhibit patterns similar to those found in the analysis by Dombowsky and Dunson (2025), whose model is specifically designed for multiview clustering. Gestational age can be divided into four clusters, each representing a different duration range of the pregnancy in days. For example, the pink cluster corresponds to longer pregnancies, the blue cluster to shorter ones, and the red cluster to durations around 240 days. Similarly, birth weight is categorized into three clusters: one primarily includes newborns with high birth weight, another captures those with lower birth weights, and the largest cluster corresponds to average-weight newborns. This discrepancy in the number and structure of clusters across the two views corresponds to what is defined as a changepoint when the framework is applied to model the evolution of partitions over time. Importantly, although the two inferred partitions differ, they reveal a common pattern: clusters linked

to higher birth weights tend to align with those associated with longer gestational periods. This reflects a biologically plausible relationship: as gestational age increases, so does the expected birth weight.

To examine variability across hospitals, we perform the same analysis after stratifying the data by hospital. Table 3 reports, for each hospital, the sample size  $n$  and the value of  $\theta$ , which is specified based on the assumption that the prior number of clusters is two for both views. Additionally, it summarizes the results of the LLDPM across hospitals, showing that the estimated probability  $\hat{\eta}$  of obtaining different partitions between the two views varies by hospital, ranging from 0.76 to 1. This variability underscores the model’s ability to adapt when applied to subsets of different sizes.

Hospital	1	2	3	4	5	6	7	8	9	10	11	12
$n$	481	124	150	77	205	154	141	141	117	384	151	188
$\theta$	0.32	0.20	0.19	0.22	0.18	0.19	0.19	0.19	0.20	0.16	0.19	0.18
$\hat{\eta}$	1.00	0.87	1.00	0.77	1.00	0.92	0.99	1.00	1.00	1.00	1.00	1.00

Table 3: Section 5.2, CPP data stratified by hospital. For each of the 12 hospitals in the dataset, we report the sample size  $n$ , the value of  $\theta$ , and the estimated probability  $\hat{\eta}$  of a change between the latent partitions characterizing the two views: newborn birth weight and gestational age.

## 6 Discussion

We introduced a new approach for modeling partitions with temporal dependence that relies on the definition of partition-based state equations in a state-space model, thus extending the dynamic linear model widely used in multivariate time series analysis. By characterizing the evolution of the local level equation across partitions and detecting changepoints, our model introduces dependence only when supported by the data. This is achieved by incorporating a mixture representation, reminiscent of the spike-and-slab prior commonly used in variable selection, to model the dependence between partitions. Here, a changepoint is defined as a shift in the partition of units between two consecutive time points. If this shift in cluster allocations coincides with a change in the process mean, our model can detect the changepoint in the traditional sense. However, if the process mean changes while the partition of units remains unchanged, the model will not identify a changepoint. This behavior reflects the model’s design: it is tailored to detect changes in clustering structure rather than in the underlying process itself. We have illustrated the model’s performance using synthetic data, which highlighted its accuracy in recovering cluster structures and identifying changepoints. In an application to human gesture data, our model yields interpretable results by identifying most changepoints during periods of activity.

Drawing upon the framework presented in Section 2.2, our model can be extended to analyze multivariate data across multiple sources, where dependence arise beyond temporal factors, e.g., varying experimental conditions. Importantly, also in this extended context, the random partitions maintain a marginal identical distribution, e.g. the distribution of a CRP. Although extending our algorithm for partition changepoint detection

to multiview scenarios is not straightforward, an application to epidemiological data displays its effectiveness in a two-view clustering context. Another potential extension involves applications where multiple variables are observed across multiple subjects. Our model can be adapted to accommodate this added complexity, provided the data follow an experimental design in which changepoints are expected to be shared across subjects and the subjects are exchangeable. This structure does not apply to the human gesture dataset described in Section 5.1, where, although data are available for multiple individuals, each subject recounts a different comic story. As a result, the time series reflect distinct underlying processes, and changepoints are not shared across subjects. A related example is offered by Ricci et al. (2024), who recently proposed a biclustering approach that groups subjects based on similarities in temporal trends observed during a task-based functional magnetic resonance imaging (fMRI) experiment.

Another exciting avenue for future research is the development of dependent random partition state-space models capable of modeling latent state equations that encode more complex dynamics. This extension could encompass higher-order time dependence, allowing for exploring more complex temporal relationships and patterns in data. In order to ensure smoother transitions over time, a hierarchical model formulation could shrink the base partition distribution  $p^*(\cdot)$  to the partition estimated at the previous time or to a fixed “anchor” partition. However, care will be needed to ensure computational feasibility and interpretability of the latent variables  $\gamma_t$ . See, e.g., Paganin et al. (2021) or Dahl et al. (2025) for details on recently proposed anchored partitioned models. On a related note, while we have assumed that the changepoint selection indicators  $\gamma_t$  are independent across time, they could instead be modeled as dependent on time-varying covariates, such as respiratory data or other measurements of expended effort in the analysis of human gesture data. Along similar lines, alternative strategies could be explored to enhance flexibility in modeling partition similarity over time. For instance, the changepoint indicator  $\gamma_t$  could be modeled as a multinomial latent variable distinguishing between no change from the previous partition, a transition to a new partition, or a reversion to a previously observed partition. This approach could also build on the framework proposed by Dahl et al. (2025). Alternatively, the mixture formulation in (3), describing the conditional distribution of  $\pi_t$ , could be recast as a linear combination between a base process  $p^*(\cdot)$  and a more flexible dependent process in lieu of point mass at  $\pi_{t-1}$ .

Finally, we note that our model is designed to efficiently handle large  $T$  through the introduction of the auxiliary changepoint variables  $\gamma_t$ . When no changepoint is detected, the partition remains unchanged across consecutive time points, allowing the algorithm to quickly proceed to the next iteration and significantly improve computational efficiency. Beyond changepoint detection, our framework opens the door to extensions in spatial and spatio-temporal domains. In purely spatial contexts, a multiview representation can be introduced, where spatial counterparts of the auxiliary changepoint variables detect changes in partitions across regions. This reconceptualizes the role of a changepoint as a shift in spatial structure, with clear relevance to biomedical tissue and transcriptomics data, where understanding spatial dependence among neighboring cells is critical. For spatio-temporal applications, the changepoint indicator  $\gamma_{s,t}$  could be defined to vary simultaneously across time  $t$  and spatial location  $s$ , with spatial covariates

incorporated to capture region specific dynamics and interactions. Allowing these auxiliary changepoint variables to be modeled as spatio-temporally dependent would enable changepoint probabilities to evolve with both spatial and temporal covariates. Such developments would substantially expand the scope of our framework, offering novel tools for analyzing complex datasets where clustering structures evolve across both space and time.

### Acknowledgments

The authors are grateful to the Associate Editor and two anonymous Referees for their valuable comments and suggestions, which greatly improved the paper.

### Funding

Alice Giampino and Bernardo Nipoti acknowledge support of MUR – Prin 2022 – Grant no. 2022CLTYP4, funded by the European Union – Next Generation EU.

## Supplementary Material

Supplementary Material for “Local Level Dynamic Random Partition Models for Changepoint Detection” (DOI: [10.1214/25-BA1560SUPP](https://doi.org/10.1214/25-BA1560SUPP); .pdf). The Supplementary Material contains the proofs of the results in Section 2, additional details on posterior computation, an extended description of the changepoint detection procedure, and further information on the simulation studies and applications.

## References

- Antoniano-Villalobos, I. and Walker, S. G. (2016). “A nonparametric model for stationary time series.” *Journal of Time Series Analysis*, 37(1): 126–142. [MR3439535](#). doi: <https://doi.org/10.1111/jtsa.12146>. 2
- Arbel, J., Favaro, S., Nipoti, B., and Teh, Y. W. (2017). “Bayesian nonparametric inference for discovery probabilities: Credible intervals and large sample asymptotics.” *Statistica Sinica*, 839–858. [MR3675033](#). 7
- Barry, D. and Hartigan, J. A. (1992). “Product partition models for change point problems.” *The Annals of Statistics*, 260–279. [MR1150343](#). doi: <https://doi.org/10.1214/aos/1176348521>. 3
- Beraha, M., Guglielmi, A., Quintana, F. A., De Iorio, M., Eriksson, J. G., and Yap, F. (2023). “Childhood obesity in Singapore: A Bayesian nonparametric approach.” *Statistical Modelling*, 1471082X231185892. 3
- Blackwell, D., MacQueen, J. B., et al. (1973). “Ferguson distributions via Pólya urn schemes.” *The Annals of Statistics*, 1(2): 353–355. [MR0362614](#). 3
- Caron, F., Davy, M., and Doucet, A. (2007). “Generalized Pólya Urn for Time-Varying Dirichlet Process Mixtures.” In *Proceedings of the Twenty-Third Conference on Un-*

- certainty in Artificial Intelligence*, UAI'07, 33-40. Arlington, Virginia, USA: AUAI Press. 3
- Caron, F., Neiswanger, W., Wood, F., Doucet, A., and Davy, M. (2017). “Generalized Pólya urn for time-varying Pitman-Yor processes.” *Journal of Machine Learning Research*, 18(27). MR3634894. 3
- Cassese, A., Zhu, W., Guindani, M., and Vannucci, M. (2019). “A Bayesian Nonparametric Spiked Process Prior for Dynamic Model Selection.” *Bayesian Analysis*, 14(2): 553 – 572. MR3934097. doi: <https://doi.org/10.1214/18-BA1116>. 3
- Chandra, N. K. and Bhattacharya, S. (2019). “Non-marginal decisions: A novel Bayesian multiple testing procedure.” *Electronic Journal of Statistics*, 13(1): 1535–1570. MR3911692. doi: <https://doi.org/10.1214/19-EJS1535>. 5, 12
- Corradin, R., Canale, A., and Nipoti, B. (2021). “BNPmix: An R package for Bayesian nonparametric modeling via Pitman-Yor mixtures.” *Journal of Statistical Software*, 100: 1–33. 13, 21, 23
- Corradin, R., Danese, L., KhudaBukhsh, W. R., and Ongaro, A. (2024). “Model-based clustering of time-dependent observations with common structural changes.” *arXiv preprint arXiv:2410.09552*. 4
- Cremaschi, A., Cadonna, A., Guglielmi, A., and Quintana, F. (2023). “A change-point random partition model for large spatio-temporal datasets.” *arXiv preprint arXiv:2312.12396*. 4
- Dahl, D. B., Warr, R. L., and Jensen, T. P. (2025). “Dependent random partitions by shrinking toward an anchor.” *Journal of the American Statistical Association*, 1–13. 25
- De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2013). “Are Gibbs-type priors the most natural generalization of the Dirichlet process?” *IEEE transactions on pattern analysis and machine intelligence*, 37(2): 212–229. 7
- De Iorio, M., Favaro, S., Guglielmi, A., and Ye, L. (2023). “Bayesian nonparametric mixture modeling for temporal dynamics of gender stereotypes.” *The Annals of Applied Statistics*, 17(3): 2256–2278. MR4637666. doi: <https://doi.org/10.1214/22-aos1717>. 3, 13
- De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). “An ANOVA model for dependent random measures.” *Journal of the American Statistical Association*, 99(465): 205–215. MR2054299. doi: <https://doi.org/10.1198/016214504000000205>. 13
- DeYoreo, M. and Kottas, A. (2018). “Modeling for dynamic ordinal regression relationships: An application to estimating maturity of rockfish in California.” *Journal of the American Statistical Association*, 113(521): 68–80. MR3803440. doi: <https://doi.org/10.1080/01621459.2017.1328357>. 3
- Dombowsky, A. and Dunson, D. B. (2025). “Product centred Dirichlet processes for

- Bayesian multiview clustering.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, qkaf021. 5, 9, 10, 23
- Ferguson, T. S. (1973). “A Bayesian analysis of some nonparametric problems.” *The Annals of Statistics*, 209–230. MR0350949. 2, 7
- Franzolini, B., De Iorio, M., and Eriksson, J. (2023). “Conditional partial exchangeability: a probabilistic framework for multi-view clustering.” *arXiv preprint arXiv:2307.01152*. 5, 10
- Giampino, A., Nipoti, B., Vannucci, M., and Guindani, M. (2025). “Supplementary Material for “Local Level Dynamic Random Partition Models for Changepoint Detection”.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/25-BA1560SUPP>. 5, 11, 12, 13, 14, 15, 17, 20, 22
- Gnedin, A. V. and Pitman, J. (2005). “Exchangeable Gibbs partitions and Stirling triangles.” *Записки научных семинаров ПОМИ*, 325(0): 83–102. MR2160320. doi: <https://doi.org/10.1007/s10958-006-0335-z>. 7
- Hadj-Amar, B., Jewson, J., and Vannucci, M. (2024). “Bayesian Sparse Vector Autoregressive Switching Models with Application to Human Gesture Phase Segmentation.” *Annals of Applied Statistics*, to appear. MR4782500. doi: <https://doi.org/10.1214/24-aos1892>. 20
- Hartigan, J. A. (1990). “Partition models.” *Communications in statistics-Theory and methods*, 19(8): 2745–2756. MR1088047. doi: <https://doi.org/10.1080/03610929008830345>. 3
- Kalli, M. and Griffin, J. E. (2018). “Bayesian nonparametric vector autoregressive models.” *Journal of Econometrics*, 203(2): 267–282. MR3770826. doi: <https://doi.org/10.1016/j.jeconom.2017.11.009>. 3
- Klebanoff, M. A. (2009). “The collaborative perinatal project: a 50-year retrospective.” *Paediatric and perinatal epidemiology*, 23(1): 2. 22
- Lijoi, A., Mena, R. H., and Prünster, I. (2005). “Hierarchical mixture modeling with normalized inverse-Gaussian priors.” *Journal of the American Statistical Association*, 100(472): 1278–1291. MR2236441. doi: <https://doi.org/10.1198/016214505000000132>. 7
- Lijoi, A., Mena, R. H., and Prünster, I. (2007). “Controlling the reinforcement in Bayesian non-parametric mixture models.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(4): 715–740. MR2370077. doi: <https://doi.org/10.1111/j.1467-9868.2007.00609.x>. 7
- Lijoi, A., Nipoti, B., and Prünster, I. (2014). “Bayesian inference with dependent normalized completely random measures.” *Bernoulli*. MR3217444. doi: <https://doi.org/10.3150/13-BEJ521>. 13
- Longnecker, M. P., Klebanoff, M. A., Zhou, H., and Brock, J. W. (2001). “Association between maternal serum concentration of the DDT metabolite DDE and preterm and small-for-gestational-age babies at birth.” *The Lancet*, 358(9276): 110–114. 18, 22

- Madeo, R. C., Lima, C. A., and Peres, S. M. (2013). “Gesture unit segmentation using support vector machines: segmenting gestures from rest positions.” In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 46–52. [18](#), [19](#), [20](#), [21](#)
- Martinez, A. F. and Mena, R. H. (2014). “On a nonparametric change point detection model in Markovian regimes.” *Bayesian Analysis*. [MR3293958](#). doi: <https://doi.org/10.1214/14-BA878>. [4](#)
- Müller, P., Parmigiani, G., and Rice, K. (2006). “FDR and Bayesian multiple comparisons rules.” *Proc. Valencia / ISBA 8<sup>th</sup> World Meeting on Bayesian Statistics*. [MR2433200](#). [12](#), [15](#)
- Müller, P., Parmigiani, G., Robert, C., and Rousseau, J. (2004). “Optimal sample size for multiple testing: the case of gene expression microarrays.” *Journal of the American Statistical Association*, 99(468): 990–1001. [MR2109489](#). doi: <https://doi.org/10.1198/016214504000001646>. [12](#)
- Newton, M. A., Noueir, A., Sarkar, D., and Ahlquist, P. (2004). “Detecting differential gene expression with a semiparametric hierarchical mixture method.” *Biostatistics*, 4(1): 155–176. [12](#), [15](#)
- Nieto-Barajas, L. E. and Contreras-Cristán, A. (2014). “A Bayesian nonparametric approach for time series clustering.” *Bayesian Analysis*, 9(1): 147–170. [MR3188303](#). doi: <https://doi.org/10.1214/13-BA852>. [3](#)
- Paganin, S., Herring, A. H., Olshan, A. F., Dunson, D. B., and Study, N. B. D. P. (2021). “Centered Partition Processes: Informative Priors for Clustering (with Discussion).” *Bayesian Analysis*, 16(1): 301–370. [MR4255332](#). doi: <https://doi.org/10.1214/20-ba1197>. [25](#)
- Paganin, S., Page, G. L., and Quintana, F. A. (2023). “Informed Random Partition Models with Temporal Dependence.” *arXiv preprint arXiv:2311.14502*. [4](#)
- Page, G. and Quinlan, J. (2022). “ppmSuite: A Collection of Models that Employ a Product Partition Distribution as a Prior on Partitions.” *R package version 0.2*, 4. [13](#)
- Page, G. L., Quintana, F. A., and Dahl, D. B. (2022). “Dependent modeling of temporal sequences of random partitions.” *Journal of Computational and Graphical Statistics*, 31(2): 614–627. [MR4425090](#). doi: <https://doi.org/10.1080/10618600.2021.1987255>. [3](#), [4](#), [7](#), [13](#), [15](#)
- Parvathy, P., Subramaniam, K., Prasanna Venkatesan, G., Karthikaikumar, P., Varghese, J., and Jayasankar, T. (2021). “Development of hand gesture recognition system using machine learning.” *Journal of Ambient Intelligence and Humanized Computing*, 12: 6793–6800. [19](#)
- Pavani, J. and Quintana, F. A. (2025). “A Bayesian Multivariate Model With Temporal Dependence on Random Partition of Areal Data for Mosquito-Borne Diseases.” *Statistics in Medicine*, 44(3-4): e10325. [MR4860478](#). doi: <https://doi.org/10.1002/sim.10325>. [3](#)

- Perman, M., Pitman, J., and Yor, M. (1992). “Size-biased sampling of Poisson point processes and excursions.” *Probability Theory and Related Fields*, 92(1): 21–39. MR1156448. doi: <https://doi.org/10.1007/BF01205234>. 7
- Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic linear models with R*. Springer Science & Business Media. MR2730074. doi: <https://doi.org/10.1007/b135794>. 2
- Pitman, J. (2006). *Combinatorial stochastic processes: Ecole d’été de probabilités de saint-flour xxxii-2002*. Springer Science & Business Media. MR2245368. 7, 17
- Pitman, J. and Yor, M. (1997). “The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator.” *The Annals of Probability*, 855–900. MR1434129. doi: <https://doi.org/10.1214/aop/1024404422>. 7
- Quinlan, J. J., Page, G. L., and Castro, L. M. (2022). “Joint random partition models for multivariate change point analysis.” *Bayesian Analysis*, 1(1): 1–28. MR4692541. doi: <https://doi.org/10.1214/22-ba1344>. 4
- Quintana, F. A., Müller, P., Jara, A., and MacEachern, S. N. (2022). “The dependent Dirichlet process and related models.” *Statistical Science*, 37(1): 24–41. MR4371095. doi: <https://doi.org/10.1214/20-sts819>. 2, 13
- Rand, W. M. (1971). “Objective criteria for the evaluation of clustering methods.” *Journal of the American Statistical Association*, 66(336): 846–850. 7
- Ricci, F. Z., Sudderth, E. B., Lee, J., Peters, M. A., Vannucci, M., and Guindani, M. (2024). “Bayesian temporal biclustering with applications to multi-subject neuroscience studies.” *arXiv preprint arXiv:2406.17131*. 25
- Romanuke, V. (2021). “Time series smoothing improving forecasting.” *Applied Computer Systems*, 26(1): 60–70. 20
- Sethuraman, J. (1994). “A constructive definition of Dirichlet priors.” *Statistica Sinica*, 639–650. MR1309433. 2
- Sun, W. and Cai, T. T. (2007). “Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control.” *Journal of the American Statistical Association*, 102: 901–912. MR2411657. doi: <https://doi.org/10.1198/016214507000000545>. 12
- Sun, W., Reich, B. J., Cai, T. T., Guindani, M., and Schwartzman, A. (2015). “False Discovery Control in Large-Scale Spatial Multiple Testing.” *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 77: 59–80. MR3299399. doi: <https://doi.org/10.1111/rssb.12064>. 12
- Tadesse, M. G. and Vannucci, M. (2021). “Handbook of Bayesian variable selection.” *Chapman & Hall/CRC*. MR4945873. 5
- Wade, S. and Ghahramani, Z. (2018). “Bayesian cluster analysis: Point estimation and credible balls (with discussion).” MR3807860. doi: <https://doi.org/10.1214/17-BA1073>. 14, 23