Gaussian Process Mixture Model: A Covariate-Dependent Bayesian Nonparametric Approach

Guilherme Piantino[†, a]
Hedibert Lopes[, a]

[a]Insper institute of Education and Research, São Paulo, Brazil

Abstract The Gaussian process mixture model (GPMM) is designed to capture regime changes and nonstationary dependence in time-ordered data without imposing a rigid Markovian structure. The model assumes a contiguous segmentation of the time index, assigning a distinct Gaussian process to each regime. To govern this segmentation structure, we employ a Pitman-Yor process adapted as a prior on compositions, which includes the Dirichlet Process as a special case. The estimation procedure utilizes a split-merge MCMC sampler to jointly learn the number and locations of regime boundaries alongside the GP hyperparameters, providing coherent uncertainty quantification. We evaluate the proposed approach through an empirical application to the CBOE Volatility Index (VIX), illustrating its ability to identify meaningful volatility regimes and capture complex nonstationary dynamics in financial time series.

Keywords: Risk measures; Standard deviation.

JEL codes: E3, C41, C43.

1. Introduction

A classical way to represent heterogeneity is through mixture models, which assume that each observation is generated from one of several latent components. In the widely used Gaussian mixture model (GMM), each component is Gaussian and the number of components is fixed (McLachlan and Peel, 2000).

While finite mixtures are effective, fixing the number of components beforehand can be difficult and potentially misleading. Bayesian nonparametric methods address this by placing a prior on an infinite mixture, so that the effective number of components is inferred from the data. A commonly used

How to cite: Piantino, G., Lopes, H. (2025). . *Revista Brasileira de Finanças*, *24*, e202601. https://doi.org/10.xxxx/xxxx
Submitted on January 1, 2026. Published online on March 31, 2026.
Editor in charge: Mr. Editor.
[†]guilhermejlp@al.insper.edu.br

prior is the Dirichlet process (DP), introduced by Ferguson (1973), which induces the Dirichlet process mixture (DPM). Influential work on posterior computation and density estimation includes Escobar and West (1995), with MCMC methodology developed in Neal (2000). In the special case of Gaussian components, the Dirichlet process Gaussian mixture model (DPGMM) provides a flexible, complexity-adaptive alternative to GMMs; see, e.g., Görür and Rasmussen (2010). Related infinite Gaussian mixture formulations have also been developed in the machine learning literature (Rasmussen, 2000).

In time-series applications, partition-based Bayesian nonparametric methods have often been applied in the context of outlier detection, as in Quintana and Iglesias (2003). This is largely because standard DP-mixture formulations rely on exchangeability, an assumption that is inappropriate for ordered data where temporal dependence is intrinsic. To overcome this limitation, researchers have extended these frameworks to allow the random mixing distribution to evolve with covariates. Key developments include the dependent Dirichlet process (MacEachern, 1999, 2000), order-based dependent Dirichlet processes (Griffin and Steel, 2006), and kernel stick-breaking processes (Dunson and Park, 2008).

In contrast, much of the time-series regime literature is built on state-space and switching formulations—including Markov-switching autoregressions (Hamilton, 1989) and hidden Markov models (Rabiner, 1989). Bayesian nonparametric extensions allow an unknown number of regimes while typically retaining a Markovian latent state structure (Fox et al., 2011). Related exchangeable partition probability function (EPPF)–based change-point constructions also often adopt Markovian assumptions; a notable example is Martínez and Mena (2014), who consider discretely observed Ornstein–Uhlenbeck diffusion processes.

Motivated by these limitations, we develop a model for multiple regimes in ordered data that avoids Markovian transition assumptions. Our proposal, the Gaussian process mixture model (GPMM), integrates (i) a nonparametric prior over contiguous segmentations and (ii) a within-regime Gaussian process (GP) likelihood. Standard GPs provide a robust probabilistic framework for nonparametric regression, but the stationarity of their covariance kernels limits their ability to represent regime shifts and abrupt changes. The GPMM addresses this limitation by inferring both the number and locations of regimes directly from the data, representing nonstationarity through a flexible piecewise GP structure.

This connects our work to the broader literature on nonstationary GPs. Nonstationarity is often introduced via spatially varying kernels, such as

process convolution constructions (Higdon et al., 1999) and nonstationary Matérn-type formulations (Paciorek and Schervish, 2004; Katzfuss, 2013), or by modeling input-dependent hyperparameters, including heteroskedastic GP frameworks (Heinonen et al., 2016; Binois et al., 2018).

Another line of work partitions the input space to fit local GPs with distinct hyperparameters, using approaches such as Voronoi tessellations (Kim et al., 2005), treed GPs (Gramacy and Lee, 2008), and local approximate GPs (Gramacy and Apley, 2015). A complementary strategy learns nonlinear transformations (warpings) of the input space to induce stationarity in a latent representation, originating with Sampson and Guttorp (1992) and Schmidt and O'Hagan (2003). Building on this, Damianou and Lawrence (2013) introduced deep Gaussian processes (DGPs), which compose multiple GP layers to learn hierarchical warpings. Recent work continues to develop full Bayesian inference for these models (Sauer et al., 2023).

Our work is most closely related to mixtures and segmentations of GPs. Rasmussen and Ghahramani (2001) proposed infinite mixtures of GP experts using DP-based mixing, while Tresp (2001) studied mixtures of GPs with gating functions. However, mixture-of-experts formulations generally allow non-contiguous assignments and are not tailored to contiguous regime segmentation in ordered data. In the change-point setting, Saatçi et al. (2010) introduced a sequential online Bayesian method in which each segment is modeled by a GP and change-points are governed by a hazard function. In comparison, the GPMM places an EPPF-based prior on ordered compositions, enforcing contiguity by construction, and performs posterior inference over both the segmentation and regime-specific GP hyperparameters using MCMC. This provides posterior uncertainty quantification for both the latent functions and the regime boundaries.

To assess the practical performance of the proposed methodology, we consider both controlled and real-world applications involving time-ordered data with complex nonstationary behavior. First, a simulation study is conducted to evaluate the model's ability to recover latent regime structures and accurately estimate regime-specific dynamics under a known data-generating process. Second, we apply the GPMM to financial time series data, specifically the CBOE Volatility Index (VIX), which exhibits pronounced regime shifts associated with periods of market stress and stability. This empirical application provides a setting in which nonstationarity, abrupt structural changes, and varying dependence patterns are intrinsic, allowing us to demonstrate that the proposed method efficiently recovers meaningful regimes while delivering well-calibrated uncertainty quantification.

2. The model

We observe a time-ordered sequence $\mathbf{y} = (Y_1, \ldots, Y_n)$ at fixed inputs X_t. Dependence among the observations is introduced through an ordered partition of the index set into K contiguous blocks, each of which is interpreted as a distinct regime. We represent the segmentation by its composition (vector of block lengths):

$$\rho = (n_1, \ldots, n_K), \qquad n_k \in \{1,2,\ldots\}, \qquad \sum_{k=1}^{K} n_k = n. \tag{1}$$

Given ρ, we define the cumulative endpoints

$$i_0 = 0, \qquad i_k = \sum_{j=1}^{k} n_j, \quad k = 1, \ldots, K,$$

so that $0 = i_0 < i_1 < \cdots < i_K = n$. The internal regime boundaries are $\{i_1, \ldots, i_{K-1}\}$, and the corresponding contiguous blocks are specified by

$$S_k = \{i_{k-1} + 1, \ldots, i_k\}, \qquad n_k = |S_k|. \tag{2}$$

Each block S_k is associated with block-specific GP hyperparameters $\boldsymbol{\phi}_k \in \Phi$. We define the time-indexed hyperparameter process $\{\boldsymbol{\phi}_t\}_{t=1}^{n}$ by

$$\boldsymbol{\phi}_t = \boldsymbol{\phi}_k, \qquad t \in S_k,$$

so that $\boldsymbol{\phi}_t$ is piecewise constant over time and can change only at the internal boundaries.

Conditional on ρ, we collect the block-specific hyperparameters as $\boldsymbol{\phi}_{1:K} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)$ and assume

$$p(\boldsymbol{\phi}_{1:K} \mid \rho) = \prod_{k=1}^{K} p(\boldsymbol{\phi}_k),$$

so that, *a priori*, the block hyperparameters are independent and identically distributed given the partition.

For each block k, let $f_k(\cdot)$ denote the latent function with GP prior governed by $\boldsymbol{\phi}_k$, and define the latent evaluations

$$f_t = f_k(X_t), \qquad t \in S_k,$$

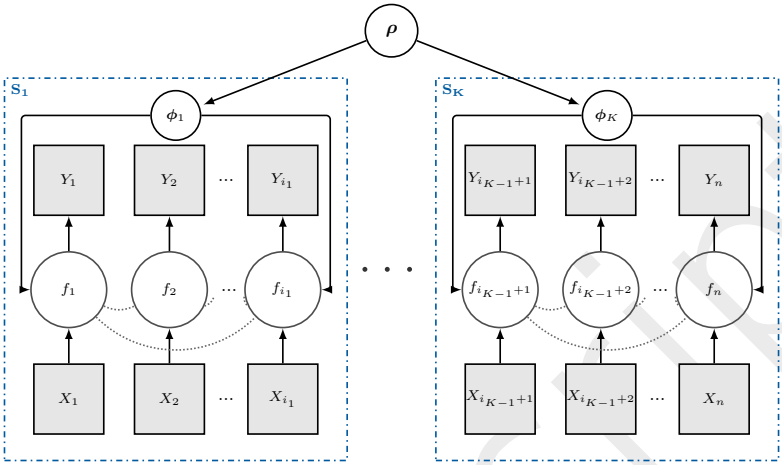


Figure 1
Graphical representation for the GPMM.

with global latent vector $\mathbf{f} = (f_1, \ldots, f_n)$. Observations are conditionally independent given $\mathbf{f}$, with Y_t depending only on f_t through a Gaussian noise model as detailed below.

As shown in Figure 1,[1] this hierarchical structure illustrates how dependence flows from the partition ρ, which determines the contiguous segments S_k, to the regime-specific hyperparameters ϕ_k, which characterize the non-Markovian GP dependence. These parameters, combined with the inputs X_t, govern the latent function evaluations f_t, which in turn generate the observed sequence $\mathbf{y}$.

3. Prior on the segmentation

The prior distribution on the segmentation ρ is specified via a Gibbs-type prior on random partitions, specifically the Pitman–Yor process (PYP), which includes the DP as a special case. Rather than fixing K, this formulation allows the number and location of regimes to be inferred directly from the data, without sacrificing computational efficiency.

We formulate our model within the framework of Product Partition Models

[1] Additional details on the GP component of the diagram are available in (Rasmussen and Williams, 2006; Gramacy, 2020).

(PPMs), which offer a natural representation of Gibbs-type priors. Under this formulation, the probability of a partition factorizes as the product of cohesion functions over its blocks. When the cohesion depends only on block cardinality, the resulting partition distribution corresponds to a Gibbs-type prior (De Blasi et al., 2015). We adopt this construction while restricting the support to contiguous intervals. Because cohesion depends solely on block size, the model induces a coherent joint distribution over the number and lengths of regimes that is consistent with the underlying Gibbs-type EPPF.

A Gibbs–type prior is characterized by an EPPF of the form

$$p_K^{(n)}(n_1,\ldots,n_K) = V_{n,K}\prod_{k=1}^{K}(1-\sigma)_{n_k-1},$$

where $\sigma < 1$ is a discount parameter, $(1-\sigma)_{n_k-1}$ denotes the rising factorial $(1-\sigma)(2-\sigma)\cdots(n_k-1-\sigma)$, and the nonnegative weights $V_{n,K}$ satisfy the forward recursion

$$V_{n,K} = (n-K\sigma)\,V_{n+1,K} + V_{n+1,K+1}, \qquad V_{1,1} = 1.$$

Different choices of $(\sigma, V_{n,K})$ recover familiar Bayesian nonparametric priors. In particular:

- **Dirichlet process.** For $\sigma = 0$,

$$V_{n,K} = \frac{\theta^K}{(\theta)_n},$$

  which implies logarithmic growth of K.

- **Pitman–Yor process.** For $\sigma \in (0,1)$,

$$V_{n,K} = \frac{(\theta+\sigma)_{K-1,\sigma}}{(\theta+1)_{n-1}},$$

  which implies polynomial growth of K.

For $\sigma \geqslant 0$, Gibbs–type priors are nonparametric and provide explicit control over the growth of the number of blocks through the discount parameter σ.

We opt to work with the PYP instead of a simpler DP because the DP is the limiting case where $\sigma \to 0$; estimating σ allows the data to determine which kind of prior is used, providing a parsimonious yet adaptable prior for

segmentation. Specifically, this distinction determines the partition structure: while the DP implies logarithmic growth in the number of blocks and favors homogeneous segment lengths, the PYP allows for polynomial growth and induces a heavy-tailed distribution where many short segments can coexist with a few long ones.

3.1 Exchangeable random order distributions and set compositions

Gibbs–type priors are typically defined for unordered exchangeable partitions through their EPPF. In our setting, however, temporal order must be preserved. We therefore consider compositions of n, which are in one-to-one correspondence with ordered partitions of $[n]$ into contiguous blocks. In particular, each contiguous segmentation is uniquely represented by the block-length sequence, where the length of the kth block equals the difference between successive change points.

To obtain a prior on ordered compositions from an EPPF, we follow the exchangeable random order construction in (Pitman, 2006). Starting from an exchangeable partition of $[n]$ with EPPF $p_K^{(n)}$, we randomize the order of its K blocks by applying a uniformly distributed random permutation of $[K]$. The induced probability mass function on compositions is

$$\Pr(\rho) \;=\; \binom{n}{n_1,\ldots,n_K} \frac{1}{K!}\, p_K^{(n)}(n_1,\ldots,n_K). \tag{3}$$

The multinomial coefficient $\binom{n}{n_1,\ldots,n_K}$ counts label assignments to K ordered blocks of these sizes, and $1/K!$ reflects the uniform random reordering of blocks.

By strictly reweighting the combinatorial order, this construction preserves the marginal distribution of K while ensuring position neutrality (i.e., conditional on block sizes, all orders are equally likely). This is advantageous because it avoids encoding time-asymmetry and the prior governs only the number and length of segments, while the data likelihood drives the specific placement of changes.

We take the underlying EPPF in (3) to be the sampling formula for the PYP with discount $\sigma \in [0,1)$ and strength $\theta > -\sigma$. Substituting the PYP EPPF into (3) yields the prior on ordered compositions:

$$\Pr(\rho) \;=\; \frac{n!}{K!\prod_{k=1}^{K} n_k!}\, \frac{\prod_{i=1}^{K-1}(\theta + i\sigma)}{(\theta+1)_{n-1}} \prod_{k=1}^{K} (1-\sigma)_{n_k-1}. \tag{4}$$

This constitutes the composition prior used in our implementation. The functional dependence on block sizes stems from the Pitman–Yor EPPF, multiplied

by the combinatorial factor.

Although the underlying EPPF is exchangeable, we use (4) solely to map a prior on partitions to a prior on compositions. Given a composition ρ, the blocks correspond deterministically to the contiguous intervals in (2). Contiguity is therefore enforced by construction.

4. Gaussian process likelihood within blocks

Having specified the prior on the regime boundaries, we now turn to the within-regime dependence structure. Conditional on the partition and the induced contiguous blocks, we model the observations within each regime using independent GPs. Each block k is associated with a latent function f_k with a GP prior

$$f_k \mid \boldsymbol{\phi}_k \sim \mathcal{GP}\big(0, C_{\boldsymbol{\phi}_k}(\cdot,\cdot)\big), \tag{5}$$

and observations within the block follow

$$Y_t \mid f_k(X_t), \sigma_{\mathrm{n}}^2 \sim \mathcal{N}\big(f_k(X_t), \sigma_{\mathrm{n}}^2\big), \qquad t \in S_k, \tag{6}$$

where the observation noise variance σ_{n}^2 is shared across all blocks and $C_{\boldsymbol{\phi}_k}(\cdot,\cdot)$ denotes a covariance (kernel) function. To prevent the model from explaining regime shifts as noise changes, we share the observation-noise variance across regimes. This way, regimes differ only through their GP hyperparameters.[2]

Our implementation adopts a squared exponential (RBF) kernel, which induces smooth latent functions without assuming parametric Markov transition dynamics:

$$C_{\boldsymbol{\phi}_k}(x,x') = s_k^2 \exp\left(-\frac{(x-x')^2}{2\ell_k^2}\right), \tag{7}$$

with block-specific length-scale ℓ_k and signal variance s_k^2, so that $\boldsymbol{\phi}_k = (\ell_k, s_k^2)$.

Conditional on ρ and σ_{n}^2, the block-specific GP components are independent *a priori*; in particular, the collection $\{(\boldsymbol{\phi}_k, f_k)\}_{k=1}^{K}$ is independent given $(\rho, \sigma_{\mathrm{n}}^2)$, with priors placed on the kernel hyperparameters $\boldsymbol{\phi}_k$.

Integrating out f_k yields the marginal likelihood for each block,

$$p(\mathbf{y}_{S_k} \mid \mathbf{x}_{S_k}, \boldsymbol{\phi}_k, \sigma_{\mathrm{n}}^2) = \mathcal{N}\Big(\mathbf{y}_{S_k}; \mathbf{0}, K_k(\boldsymbol{\phi}_k) + \sigma_{\mathrm{n}}^2 I_{|S_k|}\Big), \tag{8}$$

[2] A heteroskedastic extension with block-specific observation noise variances is straightforward.

where

$$K_k(\boldsymbol{\phi}_k) = \left[C_{\boldsymbol{\phi}_k}(X_i, X_j)\right]_{i,j \in S_k} \tag{9}$$

is the $|S_k| \times |S_k|$ covariance matrix induced by the kernel evaluated on the inputs in block S_k, and $I_{|S_k|}$ denotes the identity matrix of dimension $|S_k|$. For standard kernel functions such as the squared exponential, the matrix $K_k(\boldsymbol{\phi}_k) + \sigma_{\mathrm{n}}^2 I_{|S_k|}$ is positive definite, ensuring the likelihood is well-defined.

Assuming conditional independence across blocks given the inputs $\mathbf{x}$, the partition ρ, the block-specific hyperparameters $\{\boldsymbol{\phi}_k\}_{k=1}^K$, and the shared noise variance σ_{n}^2, the full likelihood factorizes as

$$p\Big(\mathbf{y} \mid \mathbf{x}, \rho, \{\boldsymbol{\phi}_k\}_{k=1}^K, \sigma_{\mathrm{n}}^2\Big) = \prod_{k=1}^{K} p(\mathbf{y}_{S_k} \mid \mathbf{x}_{S_k}, \boldsymbol{\phi}_k, \sigma_{\mathrm{n}}^2). \tag{10}$$

5. Posterior Computation

For posterior inference under the GPMM, we assign a prior $p(\rho)$ to the partition and independent priors $p(\boldsymbol{\phi}_k)$ to the regime-specific kernel hyperparameters, while treating the observation noise variance σ_{n}^2 as fixed.

Combining the partition prior with the block-wise GP likelihood yields the joint posterior distribution

$$p(\rho, \boldsymbol{\phi}_{1:K} \mid \mathbf{y}, \sigma_{\mathrm{n}}^2) \propto p\Big(\mathbf{y} \mid \mathbf{x}, \rho, \boldsymbol{\phi}_{1:K}, \sigma_{\mathrm{n}}^2\Big)\, p(\rho) \prod_{k=1}^{K} p(\boldsymbol{\phi}_k), \tag{11}$$

where the likelihood function is given by Equation (10).

The marginal posterior distribution of the partition is formally defined by integrating out the block hyperparameters:

$$p(\rho \mid \mathbf{y}, \sigma_{\mathrm{n}}^2) = \int p(\rho, \boldsymbol{\phi}_{1:K} \mid \mathbf{y}, \sigma_{\mathrm{n}}^2)\, d\boldsymbol{\phi}_{1:K}. \tag{12}$$

In practice, this integral is analytically intractable. Therefore, rather than evaluating the marginalization in closed form, we perform inference jointly for ρ and $\boldsymbol{\phi}_{1:K}$, and obtain posterior summaries for ρ via Monte Carlo marginalization using draws from the joint posterior in (11).

5.1 Split–merge RJMCMC for posterior inference

Standard MCMC moves that adjust individual change points often exhibit poor mixing and high computational cost. We therefore adopt a split–merge

strategy that proposes global modifications to the partition structure in a single step. This approach parallels split–merge samplers for mixture models (Jain and Neal, 2004) and their extensions to ordered partitions (Martínez and Mena, 2014), enabling the chain to escape local modes more effectively.

Our sampler targets the joint posterior $p(\rho, \boldsymbol{\phi}_{1:K} \mid \mathbf{y}, \sigma_n^2)$ using a split–merge Metropolis–within–Gibbs scheme[3]. Structural split and merge moves that change the number of segments K are implemented within a reversible jump MCMC (RJMCMC) framework. At each iteration, we alternate between (i) structural updates of the composition ρ via split, merge, or boundary-shuffle moves, and (ii) random-walk Metropolis updates of the regime-specific kernel hyperparameters $\boldsymbol{\phi}_k$ on the log scale. The latent functions are analytically marginalized, so all likelihood evaluations rely on blockwise GP marginal likelihoods.

We enforce a minimum block size $m_{\min}$. Assuming ρ is the current composition, we define

$$m(\rho) = \#\{j \in \{1, \ldots, K\} : n_j \geqslant 2m_{\min}\},$$

$$v(n) = n - 2m_{\min} + 1,$$

where $m(\rho)$ is the number of blocks eligible for splitting under the minimum size constraint and $v(n)$ is the number of admissible split locations within a block of length n. When $K > 1$, we attempt a split with probability q and otherwise attempt a merge; when $K = 1$, only splits are proposed.

Split. We select uniformly an index j from the set $\{j : n_j \geqslant 2m_{\min}\}$ and then choose a split location $r \sim \text{Unif}\{m_{\min}, \ldots, n_j - m_{\min}\}$, producing two child lengths $(r, n_j - r)$. Let the parent block have hyperparameters $\boldsymbol{\phi}_j = (\ell_j, s_j^2)$. Child hyperparameters are proposed via multiplicative log-normal perturbations,

$$\ell_{j,L} = \ell_j \exp(\varepsilon_{\ell,L}), \quad \ell_{j,R} = \ell_j \exp(\varepsilon_{\ell,R}),$$

$$s_{j,L}^2 = s_j^2 \exp(\varepsilon_{s,L}), \quad s_{j,R}^2 = s_j^2 \exp(\varepsilon_{s,R}),$$

with independent $\varepsilon_{\cdot,\cdot} \sim \mathcal{N}(0, 0.05^2)$ on the log scale.

[3] Source code demonstrating the algorithm's implementation is available at https://github.com/gpiantino/gpmm.

Merge. We select uniformly an adjacent pair $(j, j+1)$ and merge them into a single block of length $n_j + n_{j+1}$. Let (ℓ_j, s_j^2) and (ℓ_{j+1}, s_{j+1}^2) denote the corresponding hyperparameters and define weights $w_j = n_j/(n_j + n_{j+1})$ and $w_{j+1} = 1 - w_j$. The merged hyperparameters are set by the length-weighted geometric mean,

$$\ell^\star = \exp\{w_j \log \ell_j + w_{j+1} \log \ell_{j+1}\},$$

$$(s^2)^\star = \exp\{w_j \log s_j^2 + w_{j+1} \log s_{j+1}^2\}.$$

Shuffle. After the split/merge step, if $K > 1$ we optionally propose a boundary shuffle: we choose uniformly an adjacent pair $(i, i+1)$, set $T = n_i + n_{i+1}$, and resample the boundary $r \sim \mathrm{Unif}\{m_{\min}, \ldots, T - m_{\min}\}$, replacing (n_i, n_{i+1}) by $(r, T-r)$ while keeping hyperparameters fixed.

Acceptance probabilities. Up to a normalizing constant, the unnormalized target density is

$$\pi(\rho, \boldsymbol{\phi}_{1:K}) \;\propto\; p(\rho) \prod_{k=1}^{K} p(\boldsymbol{\phi}_k)\, p(\mathbf{y}_{S_k} \mid \boldsymbol{\phi}_k, \sigma_{\mathrm{n}}^2). \tag{13}$$

For a proposal from $x = (\rho, \boldsymbol{\phi})$ to $x' = (\rho', \boldsymbol{\phi}')$, we accept with probability

$$\alpha(x \to x') = \min\left\{1,\, \frac{\pi(x')}{\pi(x)}\, \frac{q(x \mid x')}{q(x' \mid x)}\, |J|\right\},$$

where $|J|$ is the Jacobian determinant of the reversible-jump, dimension-matching transformation.

For split/merge moves, the discrete proposal probabilities:

$$q_{\mathrm{split}}(\rho' \mid \rho) = \frac{1}{m(\rho)} \cdot \frac{1}{v(n_j)}, \qquad q_{\mathrm{merge}}(\rho' \mid \rho) = \frac{1}{K-1},$$

with reverse probabilities defined analogously. In a split, child hyperparameters are proposed by independent Gaussian perturbations on the log scale, which implies the Jacobian $|J_{\mathrm{split}}| = \ell_{j,L} \ell_{j,R} s_{j,L}^2 s_{j,R}^2$. For merge moves, the Jacobian is the reciprocal under the reverse mapping, i.e., $|J| = 1/|J_{\mathrm{split}}|$. For shuffle moves, which are symmetric, $|J| = 1$.

Hyperparameter updates. Conditional on the current composition, we select one block k uniformly at random and update its kernel hyperparameters by applying a random-walk Metropolis step to either $\log \ell_k$ or $\log s_k^2$, chosen with equal probability. Each proposal is accepted or rejected according to the resulting posterior ratio.

When treating the Pitman–Yor parameters as random, we additionally update (σ,θ) within the MCMC. The discount parameter σ is updated via a Gaussian random walk on the logit scale, while the strength parameter θ is updated via a Gaussian random walk on the log scale, with both moves accepted or rejected using a Metropolis step based on the posterior ratio.

5.2 Posterior Mixture of Gaussian Processes

To fully quantify uncertainty, we approximate the posterior predictive distribution via Monte Carlo integration, averaging over the MCMC draws of the partition and GP hyperparameters. This approach accounts for both the structural uncertainty of the regimes and the functional uncertainty within them.

Let $\{\Theta^{(m)}\}_{m=1}^{M}$ denote M samples from the posterior (after burn-in and thinning), where each $\Theta^{(m)} = (\rho^{(m)}, \{\boldsymbol{\phi}_k^{(m)}\}_{k=1}^{K^{(m)}}, \sigma^{(m)}, \theta^{(m)})$ includes a partition $\rho^{(m)}$ with $K^{(m)}$ blocks and the corresponding block-specific hyperparameters $\boldsymbol{\phi}_k^{(m)} = (\ell_k^{(m)}, s_k^{2(m)})$, along with the Pitman–Yor parameters $\sigma^{(m)}$ and $\theta^{(m)}$.

For a given MCMC sample m, the predictive distribution at time t (which belongs to some block k under partition $\rho^{(m)}$) is Gaussian, conditional on that specific segmentation and parameters:

$$p(Y_t \mid \Theta^{(m)}, \mathbf{y}) = \mathcal{N}\left(Y_t \mid \mu_t^{(m)}, v_t^{(m)}\right). \tag{14}$$

The posterior mean $\mu_t^{(m)}$ and variance $v_t^{(m)}$ are computed using standard GP conditioning on the observations within block k (as determined by $\rho^{(m)}$):

$$\mu_t^{(m)} = (\mathbf{k}_*^{(m)})^\top (K_k^{(m)} + \sigma_{\mathrm{n}}^2 I)^{-1} \mathbf{y}_{S_k}, \tag{15}$$

$$v_t^{(m)} = k_{**}^{(m)} - (\mathbf{k}_*^{(m)})^\top (K_k^{(m)} + \sigma_{\mathrm{n}}^2 I)^{-1} \mathbf{k}_*^{(m)} + \sigma_{\mathrm{n}}^2, \tag{16}$$

where $K_k^{(m)}$ is the kernel matrix for block k evaluated at the training inputs, $\mathbf{k}_*^{(m)}$ is the cross-covariance vector between the test point X_t and the training inputs in block k, and $k_{**}^{(m)}$ is the prior variance at X_t, all computed using the hyperparameters $\boldsymbol{\phi}_k^{(m)} = (\ell_k^{(m)}, s_k^{2(m)})$ from the m-th MCMC sample.

The full posterior predictive density is obtained by marginalizing over the posterior distribution of Θ, resulting in a mixture of Gaussians:

$$p(Y_t \mid \mathbf{y}) \approx \frac{1}{M} \sum_{m=1}^{M} \mathcal{N}\left(Y_t \mid \mu_t^{(m)}, v_t^{(m)}\right). \tag{17}$$

To obtain pointwise summaries[4], we draw samples from this mixture and compute empirical statistics. For each MCMC iteration m, we sample

$$\widetilde{Y}_t^{(m)} \sim \mathcal{N}\left(\mu_t^{(m)}, v_t^{(m)}\right). \tag{18}$$

The posterior predictive mean is then estimated as the average of the component means:

$$\widehat{Y}_t = \frac{1}{M} \sum_{m=1}^{M} \mu_t^{(m)}, \tag{19}$$

while credible intervals are estimated from the empirical quantiles of the simulated draws $\{\widetilde{Y}_t^{(m)}\}_{m=1}^{M}$. For example, a 95% credible interval is given by $[\widetilde{Y}_t^{(0.025)}, \widetilde{Y}_t^{(0.975)}]$, where $\widetilde{Y}_t^{(q)}$ denotes the q-th quantile of the draws.

This approach naturally propagates uncertainty from both the partition structure (which blocks exist and where boundaries lie) and the GP hyperparameters (the smoothness and amplitude within each regime), providing a fully Bayesian characterization of predictive uncertainty. The mixture distribution accounts for scenarios where the model is uncertain about regime boundaries, appropriately widening credible intervals in regions of structural ambiguity.

6. Experimental Results

We evaluate the proposed model through two applications. First, a simulation study validates the recovery of ground-truth segmentations and hyperparameters. Second, we apply the method to the CBOE Volatility Index (VIX) to demonstrate its effectiveness in capturing latent regime shifts within real-world financial time series.

6.1 Setup

Prior Specification. We assign independent log-normal priors to the block-specific kernel hyperparameters:

$$\log \ell_k \sim \mathcal{N}(\log 10, 0.6^2), \qquad \log s_k^2 \sim \mathcal{N}(\log 1, 0.8^2).$$

[4] We estimate in-sample, but standard GP out-of-sample prediction applies directly at any x^*.

For the segmentation, we use the PYP in (4) with discount $\sigma \in [0,1)$ and strength $\theta > 0$, treating both as unknown and updating them via Metropolis–Hastings moves during MCMC. We place a Gamma hyperprior on θ,

$$\theta \sim \text{Gamma}(a,b),$$

using the shape–rate parameterization (so $\mathbb{E}[\theta] = a/b$). Specifically, we set $a = 5$ and $b = 0.5$. No explicit hyperprior term is imposed on σ in the implementation; σ is updated over $(0,1)$ via a random-walk Metropolis step on the logit scale. The observation noise variance is treated as fixed in all experiments, set to $\sigma_n^2 = 0.1^2$.

Inference settings. We set the minimum segment length to $m_{\min} = 3$. Kernel hyperparameters are updated via random-walk Metropolis on the log scale (SD 0.15). Pitman–Yor parameters are updated similarly: σ on the logit scale (SD 0.35) and θ on the log scale (SD 0.4).

Likelihoods empoly Cholesky decompositions with a 10^{-8} diagonal jitter (fallback to 10^{-6}). To minimize computational cost, we update only the likelihoods of blocks affected by local moves, refreshing the full set every 200 iterations to prevent numerical drift. Posterior predictive summaries are computed by subsampling up to 800 retained draws.

Chains are thinned every 5 steps. We run 15,000 iterations (burn-in 7,500) for the synthetic experiment and 20,000 (burn-in 10,000) for the VIX dataset. Convergence diagnostics (Raftery and Lewis, 1992) indicate partition size stability after approximately 93 and 392 iterations, respectively. Although the chains exhibit the autocorrelation typical of partition samplers, these run lengths were sufficient to reliably identify posterior change points.

6.2 Simulation Study

To evaluate regime identification, we generate a synthetic time series ($n = 300$) by concatenating three independent GP segments. We employ a Squared Exponential kernel with length-scales $(\ell_1,\ell_2,\ell_3) = (30.0, 2.0, 10.0)$ and fixed noise $\sigma_n^2 = 0.1^2$. This configuration produces a sequence of smooth, high-frequency, and intermediate dynamics. True change-points at $t = 100$ and $t = 200$ serve as ground truth.

Figure 2 shows the posterior inference. The model accurately recovers the latent structure, exactly identifying the first change-point ($t = 100$). The second inferred boundary (grey dashed line) slightly precedes the ground truth ($t = 200$), as the signal between $t \approx 192$ and 200 follows the third-regime shape

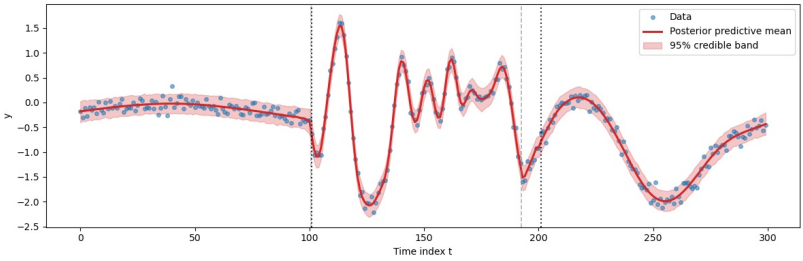


Figure 2

Posterior inference on simulated data. Blue dots: observations; Red line: posterior mean; Shaded area: 95% credible band. Vertical black dotted lines mark true change-points ($t = 100, 200$); the grey dashed line marks the inferred boundary.

almost perfectly and lacks the high-frequency oscillations of the second regime. The predictive mean (red line) captures the distinct local dynamics, while the tight 95% credible bands indicate confident, regime-specific inference.

To demonstrate robustness, we conducted 10 independent Monte Carlo runs, observing high stability across all metrics. The model achieved an average RMSE of 0.090 ± 0.003, CRPS of 0.051 ± 0.002, and NLPD of -0.956 ± 0.026, confirming precise signal recovery and well-calibrated uncertainty. Segmentation accuracy was similarly strong, with an average change-point error of just 1.95 ± 1.52 indices. Notably, the Pitman-Yor discount parameter σ converged toward zero (posterior mean 0.0036, 95% CI $[0.0000, 0.0490]$), effectively reducing the prior to a Dirichlet Process. This behavior is consistent with the simulated data's uniform regime lengths ($n = 100$), which render the power-law flexibility of a positive discount unnecessary.

6.3 VIX Data

In our second application, we analyze the VIX. The dataset comprises daily closing values from January 1, 2015, to December 31, 2023, sourced from Yahoo Finance, with $n = 2{,}264$ observations. We index observations by consecutive trading days, ensuring the model's time-scale parameters represent trading units rather than calendar time.

The VIX reflects the 30-day implied volatility of S&P 500 options and serves as a proxy for financial stress and risk aversion. This timeframe captures diverse market conditions, ranging from low-volatility phases to sudden spikes and prolonged crises, making it ideal for testing the model's ability to detect latent structural changes. We benchmark our analysis against six major market events likely to mark volatility regime changes.

China 'Black Monday' / US flash crash (2015-08-24): China slowdown fears triggered a disordered U.S. open, liquidity strains, and ETF dislocations.

Brexit vote result (2016-06-24): Surprise "Leave" vote caused a sharp Sterling sell-off, global equity declines, and safe-haven flight.

'Volmageddon' (2018-02-05): Implied volatility spike amplified by forced deleveraging in short-volatility products.

COVID panic peak (2020-03-16): Peak pandemic stress fueled by lockdowns and funding strains.

Russia invades Ukraine (2022-02-24): Invasion triggered geopolitical repricing; risk assets sold off while commodities surged.

US bank turmoil (2023-03-10): SVB failure sparked solvency fears, driving volatility in bank equities, credit spreads, and rates.

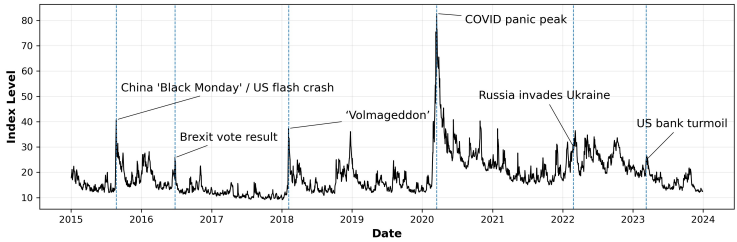
These events are not imposed as structural constraints but serve as economic benchmarks for interpreting the posterior segmentation. We model the VIX as a univariate sequence partitioned into contiguous regimes with homogeneous dynamics. In the implementation used here, we work with the log-transformed series $y_t = \log(\text{VIX}_t)$ and use integer time indices as inputs, i.e., $X_t = t - 1$ for $t = 1, \ldots, n$.

Posterior change-point probabilities are estimated at each time t as the proportion of MCMC samples initiating a new regime. As shown in Figure 3b, peaks in this trajectory pinpoint the moments where the data strongly support a structural shift.

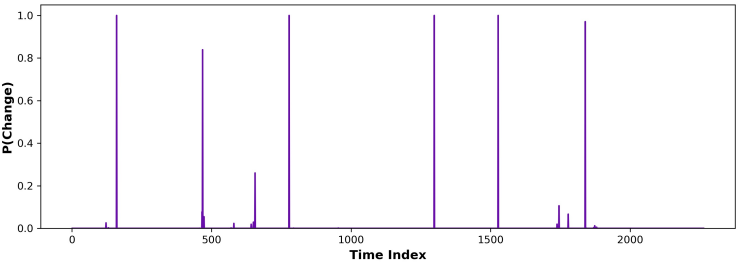
Comparing with Figure 3a, we can see that the model successfully identifies structural changes corresponding to China's "Black Monday, "Volmageddon,' and ' the COVID-19 panic.

To obtain an interpretable summary of the latent structure, we compute a Maximum A Posteriori (MAP) estimate by conditioning on the modal segmentation (the most frequently sampled composition in the MCMC output). For this fixed partition, we average the kernel hyperparameters across all corresponding draws and subsequently compute the posterior mean for each segment using standard GP conditioning.

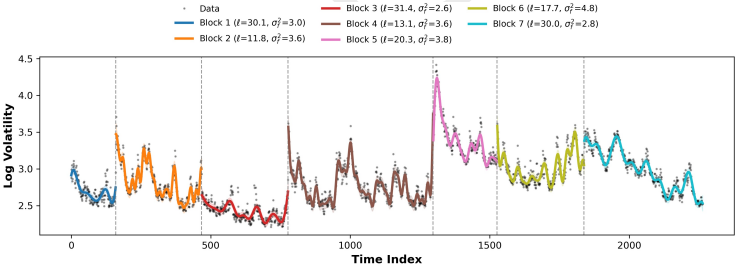
As shown in Figure 3c, block 3 represents a period of relative calm and low volatility, consistent with the absence of major external shocks. Subsequently, Block 6 marks a stabilization phase emerging from the high-stress pandemic era, leading into the final segment (Block 7), which identifies a distinct downward shift in volatility.



(a) Log VIX series with major market events (2015–2023).



(b) Posterior change-point probabilities $\widehat{p}_t$.



(c) MAP segmentation of VIX

Figure 3

Analysis of the VIX dataset. Panel (a) shows the raw log-volatility series annotated with key financial events. Panel (b) highlights the specific time points where the probability of a regime change is high. Panel (c) MAP segmentation with hyperparameters averaged over MCMC draws matching the modal partition.

To assess partition stability, we compute the Posterior Similarity Matrix (PSM), where entry (i, j) is the posterior probability that observations i and j belong to the same regime. Figure 4 shows a clear block-diagonal structure, indicating strong agreement on most regime boundaries. The main exception

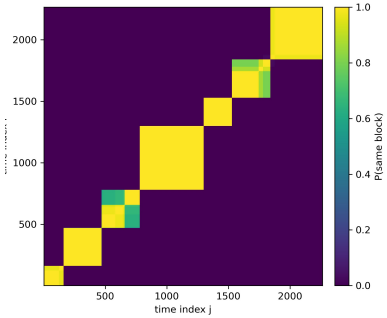


Figure 4
Posterior similarity matrix for the VIX time series.

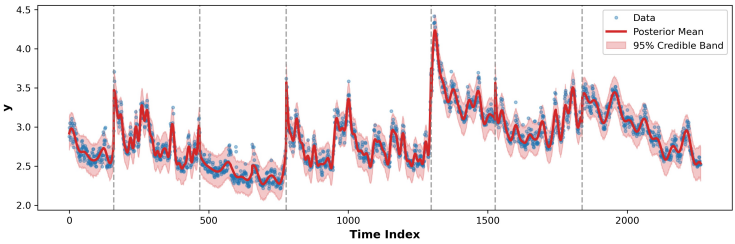


Figure 5
Posterior predictive mixture under the GPMM.

is Block 3, where a diffuse pattern suggests uncertainty between two closely related low-volatility regimes (one with a lower volatility level) and a single continuous phase. This is consistent with the change-point probability plot, which spreads posterior mass across the transition window rather than concentrating at a single index. A similar effect appears in Block 6, around the Russian invasion of Ukraine, where elevated change-point uncertainty indicates a potential short transitional regime.

Finally, to quantify how well the model learned the historical regimes, we evaluated the in-sample predictive performance using the same metrics employed in the simulation study. Results confirm strong predictive performance, with an RMSE of 0.074, CRPS of 0.043, and NLPD of -1.09. These metrics indicate precise tracking of log-volatility and robust probabilistic calibration, matching or exceeding the simulation results.

Regarding model structure, the discount parameter σ concentrated near zero (mean 0.0015, 95% CI $[0.0000, 0.0153]$), effectively collapses the prior to

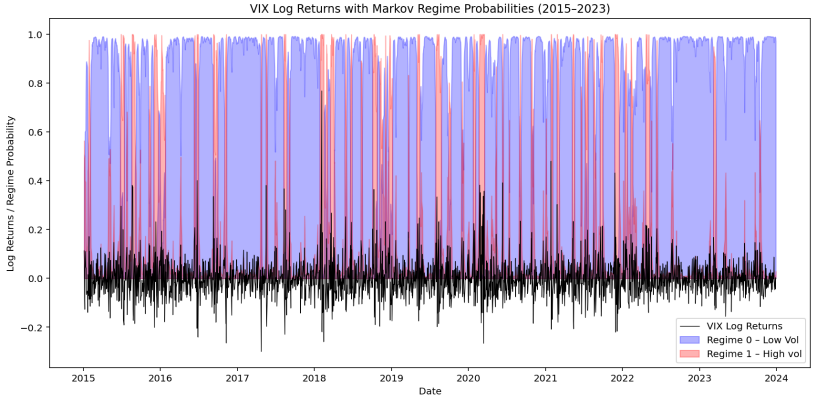


Figure 6

Markov Switching Model Results. The black line represents daily log returns. The blue and red shaded areas indicate the smoothed probability of the market being in the low volatility regime (Regime 0) and high volatility regime (Regime 1).

a DP. This structural shift is theoretically significant because the PYP typically generates a "power-law" effect characterized by a few massive clusters and many outliers. In contrast, the DP favors more homogeneously sized segments, indicating that volatility states are more balanced in their temporal persistence than a general PYP would imply.

To compare the results with a benchmark model, we run a Markov switching model on the VIX index data in the same period, with two regimes and switching variance. The model identifies two distinct states of VIX behavior: a low volatility regime (Regime 0), characterized by near-zero mean log returns and reduced variance; and a high volatility regime (Regime 1), characterized by elevated variance and sharp log return spikes, reflecting sudden increases in market fear.

Figure 6 presents the smoothed regime probabilities estimated by the two-state Markov switching model applied to the VIX index. To satisfy the model's stationarity requirement, we analyze log returns rather than raw VIX levels. This transformation captures daily percentage changes, allowing the model to detect structural shifts in volatility regimes. The black line represents these daily VIX log returns, while the shaded areas indicate the smoothed probability of the market being in each of the two identified regimes at each point in time. The blue corresponds to Regime 0 (Low Volatility), reflecting periods in which VIX log returns exhibit near-zero mean and low variance. The red shading corresponds to Regime 1 (High Volatility), capturing episodes of elevated variance and sharp return spikes associated with sudden surges in market fear.

As is visually evident, the low-volatility state (Regime 0) dominates the sample, while the model identifies brief but intense bursts of high volatility (Regime 1) that align closely with the previously listed market events. Sharp spikes in the probability of the high-volatility regime correspond to exogenous shocks, including the August 2015 China "Black Monday," the June 2016 Brexit vote, the February 2018 "Volmageddon," the March 2020 peak of the COVID-19 panic, the February 2022 Russian invasion of Ukraine, and the March 2023 U.S. banking turmoil. Smoothed regime probabilities confirm that transitions into these high-stress states are abrupt, whereas subsequent normalization is more gradual.

As the results illustrate, the Markov switching model reduces VIX dynamics to a binary classification between high and low volatility states, which reveals a fundamental limitation of the approach. In contrast, the GPMM captures a richer and more heterogeneous regime structure, identifying not only the volatility level of each segment but also its shape and local dependence pattern through regime-specific GP hyperparameters. Furthermore, the GPMM is not constrained to a fixed number of regimes, allowing for a more nuanced segmentation that goes beyond a simple high/low dichotomy. Finally, unlike the Markov switching model, the GPMM provides full uncertainty quantification over both the regime boundaries and the within-regime dynamics, offering a more interpretable characterization of structural change in financial time series.

7. Conclusions

Empirical results on synthetic and VIX data confirm GPMM effectiveness in recovering interpretable regimes and providing calibrated predictive uncertainty. The data also suggests that VIX regimes do not follow a heavy-tailed size distribution for the regimes; instead, the model naturally adopts the DP framework, with moderate size, homogeneous regimes.

In comparison with other models, unlike standard state-space or Markovian approaches, the GPMM conditions on the entire segment history to model latent dynamics, enabling it to capture long-range correlations and volatility persistence without requiring data differencing. This allows the framework to handle non-stationary financial series while preserving essential structural information. Furthermore, the batch MCMC approach provides stronger global coherence than sequential online methods. By leveraging the full dataset, the model avoids the "hindsight" limitations of greedy estimation, ensuring that regime boundaries are refined using all available evidence rather than reacting only to local observations.

When compared specifically to the Markov switching benchmark, the GPMM offers several notable advantages. First, it does not require transforming the series to achieve stationarity, the GPMM operates directly on the log-level series, preserving the structural content of the original data. Second, rather than constraining the number of regimes a priori, the GPMM infers the number and location of regime boundaries from the data, yielding a richer and more heterogeneous segmentation that goes beyond a binary high/low volatility classification. Third, each regime is characterized not only by its volatility level but also by its shape and local dependence structure through regime-specific GP hyperparameters, enabling a more nuanced description of within-regime dynamics. Finally, the GPMM provides full uncertainty quantification over both the regime boundaries and the latent functions, offering a probabilistic characterization of structural change that the Markov switching framework cannot deliver.

References

Binois, M., Gramacy, R. B. and Ludkovski, M. (2018). Practical heteroskedastic gaussian process modeling for large simulation experiments, *Journal of Computational and Graphical Statistics*, *27*(4), 808–821.

Damianou, A. and Lawrence, N. D. (2013). Deep gaussian processes, *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 207–215.

De Blasi, P., Lijoi, A., Prünster, I. and Rigon, T. (2015). Are gibbs-type priors the most natural generalization of the dirichlet process?, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(2), 212–229.

Dunson, D. B. and Park, J.-H. (2008). Kernel stick-breaking processes, *Biometrika*, *95*(2), 307–323.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures, *Journal of the American Statistical Association*, *90*(430), 577–588.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems, *The Annals of Statistics*, *1*(2), 209–230.

Fox, E. B., Sudderth, E. B., Jordan, M. I. and Willsky, A. S. (2011). Bayesian nonparametric inference of switching dynamic linear models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(8), 1569–1585.

Görür, D. and Rasmussen, C. E. (2010). Dirichlet process gaussian mixture models: Choice of the base distribution, *Proceedings of the 27th International Conference on Machine Learning (ICML)*.

Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*, Chapman and Hall/CRC, Boca Raton, FL.

Gramacy, R. B. and Apley, D. W. (2015). Local gaussian process approximation for large computer experiments, *Journal of Computational and Graphical Statistics*, *24*(2), 561–578.

Gramacy, R. B. and Lee, H. K. H. (2008). Bayesian treed gaussian process models with an application to computer modeling, *Journal of the American Statistical Association*, *103*(483), 1119–1130.

Griffin, J. E. and Steel, M. F. J. (2006). Order-based dependent dirichlet processes, *Journal of the American Statistical Association*, *101*(473), 179–194.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, *57*(2), 357–384.

Heinonen, M., Mannerström, H., Rousu, J., Kaski, S. and Lähdesmäki, H. (2016). Non-stationary gaussian process regression with hamiltonian monte carlo, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 732–740.

Higdon, D., Swall, J. and Kern, J. (1999). Non-stationary spatial modeling, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds), *Bayesian Statistics 6*, Oxford University Press, pp. 761–768.

Jain, S. and Neal, R. M. (2004). A split–merge markov chain monte carlo procedure for the dirichlet process mixture model, *Journal of Computational and Graphical Statistics*, *13*(1), 158–182.

Katzfuss, M. (2013). Bayesian nonstationary spatial modeling for very large datasets, *Environmetrics*, *24*(3), 189–200.

Kim, H. M., Mallick, B. K. and Holmes, C. C. (2005). Analyzing nonstationary spatial data using piecewise gaussian processes, *Journal of the American Statistical Association*, *100*(470), 653–668.

MacEachern, S. N. (1999). Dependent nonparametric processes, *Proceedings of the Section on Bayesian Statistical Science*, American Statistical Association, pp. 50–55.

MacEachern, S. N. (2000). Dependent dirichlet processes, *Technical report*, The Ohio State University, Department of Statistics.

Martínez, J. and Mena, R. H. (2014). A bayesian nonparametric approach to multiple change-point problems in discretely observed diffusions, *Bayesian Analysis*, *9*(4), 851–880.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*, Wiley, New York.

Neal, R. M. (2000). Markov chain sampling methods for dirichlet process mixture models, *Journal of Computational and Graphical Statistics*, *9*(2), 249–265.

Paciorek, C. J. and Schervish, M. J. (2004). Nonstationary covariance functions for gaussian process regression, *in* S. Thrun, L. K. Saul and B. Schölkopf (eds), *Advances in Neural Information Processing Systems 16*, MIT Press, pp. 273–280.

Pitman, J. (2006). *Combinatorial Stochastic Processes*, Vol. 1875 of *Lecture Notes in Mathematics*, Springer, Berlin.

Quintana, F. A. and Iglesias, P. L. (2003). Bayesian clustering and product partition models for outlier detection in time series, *Journal of Time Series Analysis*, *24*(6), 665–679.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition, *Proceedings of the IEEE*, *77*(2), 257–286.

Raftery, A. E. and Lewis, S. (1992). How many iterations in the Gibbs sampler?, *Bayesian Statistics*, *4*, 763–773.

Rasmussen, C. E. (2000). The infinite gaussian mixture model, *in* S. A. Solla, T. K. Leen and K.-R. Müller (eds), *Advances in Neural Information Processing Systems 12*, MIT Press, pp. 554–560.

Rasmussen, C. E. and Ghahramani, Z. (2001). Infinite mixtures of gaussian process experts, *Advances in Neural Information Processing Systems 14*, MIT Press, pp. 881–888.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA.

Saatçi, Y., Turner, R. D. and Rasmussen, C. E. (2010). Gaussian process change point models, *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pp. 927–934.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of non-stationary spatial covariance structure, *Journal of the American Statistical Association*, *87*(417), 108–119.

Sauer, A., Renganathan, A., Gramacy, R. B. and Higdon, D. (2023). Active learning for deep gaussian process surrogates, *Technometrics*, *65*(1), 4–18.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations, *Journal of the Royal Statistical Society: Series B*, *65*(3), 743–758.

Tresp, V. (2001). Mixtures of gaussian processes, *Advances in Neural Information Processing Systems 13*, MIT Press, pp. 654–660.

,