Plug-and-Play Bayesian Online Change-Point Detection in Gaussian Processes

Anonymous Author(s) Affiliation Address email

Abstract

2	We study the problem of detecting non-stationarity in online time series, when the
1 3	underlying distribution is assumed to be a piecewise 1D-Gaussian process. Draw-
4	ing inspiration from Bayesian online change-point detection methods such as that
5	of Adams and MacKay [2007], we construct a restarted variant of that to specifi-
6	cally deal with arbitrary changes both in mean and variance of 1D-Gaussian pro-
7	cesses. We evaluate our algorithm on both synthetic datasets of varying task diffi-
8	culty and on prevalent real-world data across a variety of fields. Our results com-
9	pare favorably with state-of-the-art, as measured by the detections' F1-SCORE.
10	Code will be provided to ensure easy reproducibility.

11 **1 Introduction**

12 While most statistical models assume that the input data is generated according to some underlying distribution, it is often assumed that the latter does not actively change throughout instances of 13 training or inference. This presents a major setback when tackling real-world problems where the 14 data collection process is often done in a sequential manner and where the parameters of the gener-15 ating distribution, if assumed parametric, are changing continuously. This applies in various ways 16 across a variety of fields in machine learning and statistics: distribution shifts in deep learning (DL), 17 environment non-stationarity in reinforcement learning (RL), to name a few. 18 More precisely, we consider the general setting of a sequential decision-making process, during 19

which a series of abrupt changes, commonly referred to as *change-points* (CP), take place. Change-20 points are sudden shifts in the underlying parameters of the data generating distribution of a given 21 sequence. Detecting these change-points in real-time is of vital importance for the analysis and 22 forecasting of time series, especially in high volatility scenarios, in consumer decision modeling (Xu 23 and Yun [2020]), service provider adaptation to customers, and pricing (Taylor [2018], Kanoria and 24 Oian [2019], Bimpikis et al. [2019], Gurvich et al. [2019]), wireless communication networks (Zhou 25 and Bambos [2015], Zhou et al. [2016]), epidemic networks and control (Nowzari et al. [2016], Kiss 26 et al. [2017]), inventory management (Agrawal and Jia [2019], Huh and Rusmevichientong [2009]), 27 non-stationary multi armed bandits (Alami et al. [2017], Alami and Azizi [2020], Alami [2023, 28 2018], Garivier and Moulines [2011], safety-aware bandits: Alami et al. [2023a]), RL [Alami et al., 29 2023b], and notions of automated quality control (El Mekkaoui et al. [2024]), to name a few. 30

31 Key contributions. We outline our contributions as follows

 We propose a novel variant of the Restarted Bayesian Online Change-Point Detection algorithm (R-BOCPD), that generalizes the modeling scope to the setting where online observation stream is generated from an underlying piecewise stationary 1D-Gaussian process (GP). Our model incorporates changes in either means or variances (or both) of the underlying GP.

• We demonstrate our results experimentally across a wide range of tasks of varying difficulty, on both synthetic and real-world datasets. Our algorithm compares favorably to

Submitted to Workshop on Bayesian Decision-making and Uncertainty, 38th Conference on Neural Information Processing Systems (BDU at NeurIPS 2024). Do not distribute.



Figure 1: Piecewise Stationary Gaussian Process. Starting from change-point c_{ℓ} , the model is assumed to be Gaussian of parameters $(\mu_{\ell}, \sigma_{\ell})$.

state-of-the-art in both online and offline settings, as measured by false-alarm and misde tection rates, detection delay, and finite-time runtime.

• We present concrete promising directions for future work relevant both to theoreticians and practitioners working both on probabilistic modeling under uncertainty and general modeling of time-series data in a variety of fields.

44 **2** Main Results

Given the prevalent use of 1D-Gaussian distributions as statistical priors for a wide range of problems across a variety of domains, we choose to model a piecewise stationary GP, where a set of unknown

47 abrupt change-points $\{c_\ell\}_{\ell=1}^L$ take place such that

$$x_t \sim \mathcal{N}(\mu_\ell, \sigma_\ell) \quad \forall t \in [c_\ell, c_{\ell+1}], \forall \ell \in [1, L]$$
(1)

where L is the unknown total number of change-points throughout [1, t]. We provide Figure 1 for visualization.

50 Challenge. We aim to detect the change-points online and sequentially from data in real-time, 51 without *à-priori* knowledge on their number, their location, or how often they occur. The latter, 52 indeed, can be used as a plug-and-play prior, as showcased later on in Algorithm 1. In particular, we

indeed, can be used as a plug-and-play prior, as showcased later on in Algorithm 1. In partic
 are interested in designing an algorithm that detects change-points

• *reliably*, with as few misdetections and false-alarms as possible.

• *in real-time*, with as low of a detection delay as possible.

incorporating uncertainty, yielding statistically optimal or near-optimal probabilistic guar antees, allowing for flexibility and control to the decision make.

58 2.1 Change-point Detection as Runlength Inference

We first introduce the notion of *runlength* r_t , which is defined as the number of time steps since the last change-point to the process, given an observed data sequence $\mathbf{x}_{1:t}$ (up to current time step t). Adams and MacKay [2007] introduce an efficient Bayesian approach for handling piecewise stationary processes via computing the posterior distribution over the current runlength r_t . The exact inference on the runlength distribution is done recursively via message-passing as follows

$$p(r_t|\mathbf{x}_{1:t}) \propto \sum_{r_{t-1}} \underbrace{p(r_t|r_{t-1})}_{\text{hazard function}} \underbrace{p(x_t|r_{t-1}, \mathbf{x}_{1:t-1})}_{\text{UPM}} p(r_{t-1}|\mathbf{x}_{1:t-1})$$

65 where the *hazard function* is defined as

$$p(r_t|r_{t-1}) = \begin{cases} H(r_{t-1}) & \text{if } r_t = 0\\ 1 - H(r_{t-1}) & \text{if } r_t = r_{t-1} + 1\\ 0 & \text{otherwise} \end{cases}$$
(2)

where *H* is defined as $H(s) = \frac{\mathbb{P}_{\text{change}}(s+1)}{\sum_{t=s+1}^{\infty} \mathbb{P}_{\text{change}}(t)}$, $\mathbb{P}_{\text{change}}$ denotes the probability distribution over the

67 interval between changepoints, and the *underlying probability model* (UPM) depends on the prob-

ability distribution of x_t . We showcase an illustrating example for our reasoning and algorithmic

69 construction in Appendix A.

Algorithm 1 R-BOCPD for Piecewise Stationary GPs

 $\begin{aligned} \text{Input: } h \in (0,1), \alpha_0 &= 1, \beta_0 = 1, n_0 = 1, \mu_0 = 0 \\ 1: r \leftarrow 1, \nu_{r,r:r} \leftarrow 1, \mu_{r,r:r} \leftarrow \mu_0, \alpha_{r,r:r} \leftarrow \alpha_0, \beta_{r,r:r} \leftarrow \beta_0, n_{r,r:r} \leftarrow n_0. \\ 2: \text{ for } t &= 1, \dots \text{ do} \\ 3: \quad \text{Observe } x_t \sim \mathcal{N} (\mu_t, \nu_t) \\ 4: \quad \text{For each forecaster starting at time } r \text{ and } s \text{ up to time } t, \text{ define} \\ \nu_{r,s:t} \leftarrow \begin{cases} (1-h) \exp(-\ell_{s:t}) \nu_{r,s:t-1} & \text{for } s \in [r,t), \\ h \sum_{i=r}^{t-1} \exp(-\ell_{i:t}) \nu_{r,i:t-1} & \text{for } s = t \end{cases} \\ 5: \quad \text{Estimate the last change-point } \widehat{\tau}_t: \quad \widehat{\tau}_t \leftarrow \operatorname{argmax}_{s \in [r,t]} v_{r,s:t} \\ 6: \quad \text{if } \widehat{\tau}_t = t \text{ then} \end{aligned}$

(3)

(7)

7: $r \leftarrow t+1, \nu_{r,r:r} \leftarrow 1, \mu_{r,r:r} \leftarrow \mu_0, \alpha_{r,r:r} \leftarrow \alpha_0, \beta_{r,r:r} \leftarrow \beta_0, n_{r,r:r} \leftarrow n_0.$

8: end if 9: end for

70 2.2 Algorithmic Construction of R-BOCPD

For an algorithmic construction that adheres to our previously outlined design objectives, we draw inspiration from Alami et al. [2020b]'s extension to the seminal *Bayesian Online Change-point Detection* (BOCPD) work of [Fearnhead and Liu, 2007]. Indeed, Alami et al. [2020a] introduce a pruned variant of the latter, which they refer to as the *Restarted Bayesian Online Change-point Detection*, particularly designed to model online changes in the means of univariate Bernoulli-distributed data samples. We propose to extend the R-BOCPD construction to change-points in an online piecewise stationary GP. In particular, we start by introducing a few definitions in the following

78 **Definition 2.1** (Predictor). Given a sequence of observations $\mathbf{x}_{s:t} = (x_s, ..., x_t)$, we define an 79 instance of a predictor as follows

$$\operatorname{PRED}\left(x_{t+1}|\mathbf{x}_{s:t}\right) = \frac{\Gamma\left(\frac{2\alpha_{s:t}+1}{2}\right)}{\Gamma\left(\alpha_{s:t}\right)\sqrt{2\alpha_{s:t}\pi}} \left(1 + \frac{1}{2} \times \frac{\left(x_{t+1} - \mu_{s:t}\right)^{2}}{\frac{\beta_{s:t} \times (n_{s:t}+1)}{n_{s:t}}}\right)^{-\frac{1}{2}}$$
(4)

where PRED $(x|\emptyset) = \frac{\Gamma(\frac{2\alpha_0+1}{2})}{\Gamma(\alpha_0)\sqrt{2\alpha_0\pi}}$ for some chosen $\alpha_0 > 0$ at initialization and incrementing procedure

$$\alpha_{s:t+1} = \alpha_{s:t} + \frac{1}{2} \quad n_{s:t+1} = n_{s:t} + 1 \quad \beta_{s:t+1} = \beta_{s:t} + \frac{n_{s:t} \times (x_{t+1} - \mu_{s:t})^2}{2 \times (n_{s:t} + 1)}$$

- 82 Akin to the our considered construction, instead of dealing with run-length, we introduce the notion
- of forecaster loss as loss incurred by predictors across time
- **Definition 2.2** (Forecaster Loss). Using the predictor, the instantaneous loss of the forecaster s at time t is given by:

$$\ell_{s:t} := -\log \operatorname{PRED}\left(x_t | \mathbf{x}_{s:t-1}\right).$$

Then, let $\widehat{L}_{s:t} := \sum_{s'=s}^{t} \ell_{s':t}$ denotes the cumulative loss incurred by the forecaster *s* from time *s* until time *t* which takes the following form

RF

$$\widehat{L}_{s:t} := \sum_{s'=s}^{t} -\log \operatorname{PRED}\left(x_t | \mathbf{x}_{s':t-1}\right)$$
(5)

⁸⁸ Thus, the forecaster weights update will remain the same (for some parameter $h \in (0, 1)$).

$$\nu_{r,s:t} = \begin{cases} (1-h)\exp\left(-\ell_{s,t}\right)\nu_{r,s:t-1} & \forall s < t, \\ h \times V_{r:t-1} & s = t. \end{cases} \text{ with } V_{r:t-1} := \sum_{i=r}^{t-1}\exp\left(-\ell_{i:t}\right)\nu_{r,i:t-1} \quad (6)$$

⁸⁹ Finally, we keep the same restart procedure as in Alami et al. [2020b], namely

$$\mathbb{E}\mathsf{START}(x_r, ..., x_t) = \mathbb{I}\left[\exists \ s \in (r, t] : \nu_{r, s:t} > \nu_{r, r:t}\right]$$

- where $\mathbb{I}(.)$ is the indicator function. Finally, we describe our algorithm in full pseudo-code in Algo-
- 91 rithm 1.

	DATASET							
Algorithm	VERY EASY		Easy		MEDIUM		HARD	
	F1-Score \uparrow	$Delay \downarrow$	F1-Score \uparrow	$Delay \downarrow$	F1-Score \uparrow	$Delay \downarrow$	$F1-Score \uparrow$	$Delay \downarrow$
BINSEG [Scott and Knott, 1974] PELT [Killick et al., 2012]	0.87 1.0	_	1.0 1.0	_	0.8 0.909	_	0.6 0.67	_
CUSUM ¹ [Basseville and Nikiforov, 1993] GLRT [Keshavarz et al., 2018]	0.632 1.0	0.444 0.009	 1.0	0.14	0.889	0.339	0.174	0.879
R-BOCPD (OURS)	1.0	0.01	1.0	0.016	1.0	0.038	0.9	0.498

Table 1: Benchmark results for various algorithms across different synthetic datasets of varying difficulty. We report the F1-SCORE and DELAY for each method. Delay is normalized by the size of the stationary periods for each detected change-point for appropriate unified performance assessment across sequences of different lengths. We highlight here that the first family of methods is offline, hence naturally would get much smaller delays. The second group of methods is online, akin to our proposed method. For the interest of comparison within our considered setting, we only measure detection delays for online methods.

2.3 Empirical Results 92

Measuring Performance. We design a diverse synthetic task suite and also evaluate on prevalent 93 open-source real-world time-series data (https://github.com/alan-turing-institute/TCPD). We evalu-94 ate our algorithm on two key metrics: F1-Score and detection delay. We normalize the latter by 95 the length of its corresponding stationary period, i.e for change-point c_{ℓ} , delay d_{ℓ} is normalized by 96 97

 $c_{\ell+1} - c_{\ell}$.

ALGORITHM	DATASET				
	JFK PASSENGERS	CO ₂ CANADA	BUSINV	BITCOIN	
BINSEG [Scott and Knott, 1974] PELT [Killick et al., 2012]	1.0 0.5	0.67 0.67	0.24 0.20	0.43 0.43	
GLRT [Keshavarz et al., 2018]	1.0	0.8	0.57	0.58	
RBOCPDMS [Knoblauch et al., 2018]	—	_	0.27		
GPTS-CP [Saatçi et al., 2010]	—	—	0.62	—	
ADAGA [Caldarelli et al., 2022]	—	—	0.77	—	
R-BOCPD (OURS)	1.0	1.0	0.8	0.8	

Table 2: Benchmark results for various algorithms across different real-world datasets belonging to a variety of domains. We mainly compare in terms of F1-SCORE. We were unable to reproduce RBOCPDMS, GPTS-CP and ADAGA at the time of submission. Instead, we reproduce their exact setting for the BUSINV dataset.

Runtime. Given the online nature of our proposed algorithm, we analyze to what extent our algorithm (and others) allow for smooth inference in real-time. We see that our runtime compares favorably with state-of-theart, which makes it especially useful for modeling longcontext sequences.

Algorithm	Runtime (ms) \downarrow
BINSEG [Scott and Knott, 1974] PELT [Killick et al., 2012]	2.82666 15.3721
GLRT [Keshavarz et al., 2018]	3.0112
R-BOCPD (OURS)	0.7097

Table 3: Runtime (in milliseconds) per iteration. The first family of methods is offline, while the second one is online, similarly to R-BOCPD.

Discussion & Future Work. Our work sheds light on online change-point detection when the un-99 derlying distribution of the streaming data is a 1D-Gaussian. Being both quite relevant in practice 100 and theoretically interpretable, this lays the ground for a variety of possible extensions and future 101 work. Among these we list- establishing potential theoretical optimality guarantees, scaling to mix-102 tures of detectors and building hybrid algorithms. We defer a detailed discussion of the latter to 103 Appendix **B**. 104

¹CUSUM primarily detects shifts in the mean of a process and does not account for changes in the covariance structure. To fairly showcase how it performs, we design the "Very Easy" setting where only changes in the mean of underlying process occur.

105 **References**

- R. P. Adams and D. J. MacKay. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*, 2007.
- S. Agrawal and R. Jia. Learning in structured mdps with convex cost functions: Improved regret
 bounds for inventory management. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 743–744, 2019.
- R. Alami. Thompson Sampling for the non-Stationary Corrupt Multi-Armed Bandit. In *The 14th European Workshop on Reinforcement Learning*, volume 14, Lille, France, Oct. 2018. URL https://hal.science/hal-01963539.
- R. Alami. Bayesian change-point detection for bandit feedback in non-stationary environments. In
 E. Khan and M. Gonen, editors, *Proceedings of The 14th Asian Conference on Machine Learning*,
 volume 189 of *Proceedings of Machine Learning Research*, pages 17–31. PMLR, 12–14 Dec
 2023. URL https://proceedings.mlr.press/v189/alami23a.html.
- R. Alami and O. Azizi. TS-GLR: an Adaptive Thompson Sampling for the Switching Multi-Armed
 Bandit Problem. In *NeurIPS 2020 challenges of real world reinforcement learning workshop*,
 Virtual, Canada, Dec. 2020. URL https://hal.science/hal-03628791.
- R. Alami, O. Maillard, and R. Féraud. Memory Bandits: a Bayesian approach for the Switching
 Bandit Problem. In *NIPS 2017 31st Conference on Neural Information Processing Systems*,
 Long Beach, United States, Dec. 2017. URL https://hal.science/hal-01811697.
- R. Alami, O. Maillard, and R. Féraud. Restarted bayesian online change-point detector achieves optimal detection delay. In *International conference on machine learning*, pages 211–221. PMLR, 2020a.
- R. Alami, O. Maillard, and R. Feraud. Restarted Bayesian online change-point detector achieves optimal detection delay. In H. D. III and A. Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 211–221. PMLR, 13–18 Jul 2020b. URL https://proceedings.mlr.press/v119/alami20a.html.
- R. Alami, M. Mahfoud, and M. Achab. A risk-averse framework for non-stationary stochastic multi armed bandits, 2023a. URL https://arxiv.org/abs/2310.19821.
- R. Alami, M. Mahfoud, and E. Moulines. Restarted bayesian online change-point detection for non stationary markov decision processes, 2023b. URL https://arxiv.org/abs/2304.00232.
- M. Basseville and I. V. Nikiforov. *Detection of abrupt changes: theory and application*. Prentice Hall, 1993.
- K. Bimpikis, O. Candogan, and D. Saban. Spatial pricing in ride-sharing networks. *Operations Research*, 67(3):744–769, 2019.
- E. Caldarelli, P. Wenk, S. Bauer, and A. Krause. Adaptive Gaussian process change point detection.
 In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 2542–2571. PMLR, 17–23 Jul 2022. URL https://proceedings.
 mlr.press/v162/caldarelli22a.html.
- S. El Mekkaoui, G. Boukachab, L. Benabbou, and A. Berrado. Deep learning based vessel arrivals
 monitoring via autoregressive statistical control charts. *WMU Journal of Maritime Affairs*, pages
 1–18, 2024.
- P. Fearnhead and Z. Liu. On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):589–605, 2007.
- A. Garivier and E. Moulines. On upper-confidence bound policies for switching bandit problems.
 In Algorithmic Learning Theory: 22nd International Conference, ALT 2011, Espoo, Finland, October 5-7, 2011. Proceedings 22, pages 174–188. Springer, 2011.

- 153 I. Gurvich, M. Lariviere, and A. Moreno. Operations in the on-demand economy: Staffing services
- with self-scheduling capacity. *Sharing Economy: Making Supply Meet Demand*, pages 249–278, 2019.
- W. T. Huh and P. Rusmevichientong. A nonparametric asymptotic analysis of inventory planning
 with censored demand. *Mathematics of Operations Research*, 34(1):103–123, 2009.
- Y. Kanoria and P. Qian. Blind dynamic resource allocation in closed networks via mirror backpressure. *arXiv preprint arXiv:1903.02764*, 2019.
- H. Keshavarz, C. Scott, and X. Nguyen. Optimal change point detection in gaussian processes.
 Journal of Statistical Planning and Inference, 193:151–178, 2018.
- R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598, 2012.
- I. Z. Kiss, J. C. Miller, P. L. Simon, et al. Mathematics of epidemics on networks. *Cham: Springer*, 598:31, 2017.
- J. Knoblauch, J. E. Jewson, and T. Damoulas. Doubly robust bayesian inference for non-stationary
 streaming data with beta -divergences. *Advances in Neural Information Processing Systems*, 31,
 2018.
- C. Nowzari, V. M. Preciado, and G. J. Pappas. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems Magazine*, 36(1):26–46, 2016.
- Y. Saatçi, R. D. Turner, and C. E. Rasmussen. Gaussian process change point models. In *Proceedings* of the 27th International Conference on Machine Learning (ICML-10), pages 927–934, 2010.
- A. J. Scott and M. Knott. A cluster analysis method for grouping means in the analysis of variance.
 Biometrics, 30(3):507–512, 1974.
- T. A. Taylor. On-demand service platforms. *Manufacturing & Service Operations Management*, 20 (4):704–720, 2018.
- G. J. J. van den Burg and C. K. I. Williams. An evaluation of change point detection algorithms,
 2022. URL https://arxiv.org/abs/2003.06222.
- K. Xu and S.-Y. Yun. Reinforcement with fading memories. *Mathematics of Operations Research*, 45(4):1258–1288, 2020.
- Z. Zhou and N. Bambos. Wireless communications games in fixed and random environments. In
 2015 54th IEEE Conference on Decision and Control (CDC), pages 1637–1642. IEEE, 2015.
- 183 Z. Zhou, P. Glynn, and N. Bambos. Repeated games for power control in wireless communications:
- Equilibrium and regret. In 2016 IEEE 55th Conference on Decision and Control (CDC), pages 3603–3610. IEEE, 2016.

Supplementary Material

188			
189			

191 Table of Contents 193 A Learning through "Runlength" Inference 8 194 B Future Work 8

198 A Learning through "Runlength" Inference

A simple example of the aforementioned approach would be to use a constant hazard function htaking some value in (0, 1). Given that, In the sense that $p(r_t = 0|r_{t-1})$ is independent of r_{t-1} and is constant, giving rise, à priori, to geometric inter-arrival times for change points ($\mathbb{P}_{change}(s+1) = h(1-h)^s$). Thus, the recursive runlength distribution computation becomes:

$$p(r_t \neq 0 | \mathbf{x}_{1:t}) \propto (1-h) p(x_t | r_{t-1}, \mathbf{x}_{1:t-1}) p(r_{t-1} | \mathbf{x}_{1:t-1})$$
$$p(r_t = 0 | \mathbf{x}_{1:t}) \propto h \sum_{r_{t-1}} p(x_t | r_{t-1}, \mathbf{x}_{1:t-1}) p(r_{t-1} | \mathbf{x}_{1:t-1})$$

203 **B** Future Work

204 We list a few directions for future work, as outlined in the discussion.

- Establishing potential theoretical optimality guarantees (for instance in terms of detection delay and false-alarm rate, akin to that in Alami et al. [2023b] in the case where the underlying distribution is piecewise multinomial), given the particular algorithm construction we adopt.
- Scaling to online probabilistic (posterior-weighted) mixtures of forecasters, which would reduce the sensitivity to the prior choice of *h* in Algorithm 1 and would potentially allow to systematically increase the confidence predictions given an increase in resources.
- Given the increasing availability of ground-truth annotators such as that of van den Burg and Williams [2022], this would allow to potentially build hybrid algorithms on top where parts of the algorithm can be learned/tuned from past data.

215