CHANGE DETECTION AND GAUSSIAN PROCESS INFERENCE IN PIECEWISE STATIONARY ENVIRONMENTS UNDER NOISY INPUTS

Tales Imbiriba, Gerald LaMountain, Peng Wu, Deniz Erdoğmuş and Pau Closas

Electrical & Computer Engineering department, Northeastern University, Boston, MA, USA

ABSTRACT

Gaussian Process (GP) inference in dynamical piecewise stationary environments finds application in many modern real-world problems. Here we focus on scenarios in which precise initial training data is available but each subsequent measurement is taken with uncertainty with respect to the location at which it was taken. In such a scenario, the ideal inference methodology should update the GP only if a considerable change occurs. Furthermore, the update strategy should preserve "correct" initial information in unchanged regions. In this contribution, we propose a change detection methodology which operates in conjunction with the inference process. The proposed detection method is based on the Kolmogorov-Smirnov statistics in such way that an exact PDF can be derived under the null hypothesis. Once detection has occurred, the GP can be naively updated by replacing appropriate measurements in the data set by an average computed from the new, noisy samples. Simulations with synthetic data provide a proof of concept of the proposed strategy as well as its robustness to noisy inputs.

Index Terms— Gaussian Process, change detection, uncertain inputs, Kolmogorov-Smirnov test.

1. INTRODUCTION

Gaussian process (GP) regression methods consist of defining stochastic models for functions and performing inference in functional spaces [1]. These methods have been shown to be useful in a wide variety of fields and tasks including regression and classification [1], detection [2], unmixing [3], and Bayesian optimization [4], to name but a few. The basic idea behind GP regression consists of modelling a function $\psi : \mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{x} \mapsto \psi(\boldsymbol{x})$ for which very little information regarding ψ is available other than some smoothness assumption. The GP inference is often performed considering a training set $\mathcal{D} = \{y_k, \boldsymbol{x}_k\}_{k=1}^N$ comprised of paired known noiseless inputs $\boldsymbol{x}_k \in \mathbb{R}^d$ and measured noisy outputs $y_k \in \mathbb{R}$.

In this paper, we are concerned with inference in environments which are piecewise stationary such that the actual process connecting inputs and outputs may present abrupt, localized changes over time. Furthermore we assume that, apart from the initial training set, any new collected sample has uncertainty with respect to the location where it was measured (noisy inputs). This scenario is found in many applications such as indoor positioning and tracking using fingerprinting [5], crowed-sourcing [6, 7], simultaneous localization and mapping (SLAM) [8] and others. In such scenarios, the GP should be adapted only if new arriving samples present a non-neglectable change with respect to the GP learned from the initial training set. Furthermore, if the change is confined to a finite region of the input space, the updating strategy must preserve the original ("correct") data outside of this region. Accomplishing each of these design goals necessitates that one can detect changes (or point changes) within the GP framework.

The problem of detecting abrupt changes in time series has a broad literature under slightly variant taxonomies such as change detection or change point detection; see [9, 10] and references therein. Change point detection of Gaussian processes has also been studied in the literature considering both online and offline inference strategies [11, 12]. In [13] the authors proposed a Bayesian framework to online change point detection. The methodology is built over an *underlying predictive model* (UPM) and a hazard function which allows for recursive inference regarding the "run length", that is, the time since last change point. This work was extended by the consideration of nonparametric UPM using GP time series and autoregressive GPs [11]. In [14] the authors proposed a GP-based online change point detection methodology for periodic time series and using Q-charting computed over the predicted and true time-series new sample. In slightly different paradigm, in [15] a scalable GP model was considered for identifying and characterizing smooth multidimensional changepoints. The proposed methodology consists of defining a change surface model as a mixture of latent functions. The resulting methodology consists in performing joint inference of the mixture coefficients and the latent functions. Recently, the authors in [12] extended a Generalized likelihood ratio test (GLRT) to detect mean shifts over time based on GPs. The method exploits the structure of the covariance matrix and is shown to achieve nearly asymptotically optimal rate in the minimax sense. In all such methods, GP inference was performed to model relevant temporal quantities for unidimensional time-series. This presents challenges when modeling fields acts over large, possibly multidimensional, domains. Thus, different strategies should be sought.

In [16] and [17], the authors considered a Kolmogorov-Smirnov (KS) test statistics for measuring the robustness of sequential methods. The resulting test matched a single new observation with K samples from the predictive distribution of the dynamical model. In this contribution, we propose a detection strategy that extends the work presented in [16, 17] by modifying the test statics to exploit the Gaussian characterization provided by the GP. This approach results in a test statistic which is low-complexity and can be computed recursively during a stationary time window. Furthermore, the proposed approach provides an exact probability density function (PDF) for the test statistics under the null hypothesis, allowing the detector to be designed for a given probability of false alarm. Once, for a specific point, a change has been detected we update the GP by replacing the corresponding measurement in the data set by the average of the new points used in the detection process. By operating in such way we avoid the input-uncertainty dissemination across the whole dataset. Simulations with synthetic data and Monte-Carlo runs illustrate the capabilities of the proposed methodology.

This work has been supported by NSF awards CNS-1815349, ECCS-1845833, CBET1804550.

This paper is organized as follows. Section 2 presents the standard GP regression formulation. Section 3 presents the proposed KS detector. Simulations are presented and discussed in Section 4 while the conclusions are presented in Section 5.

2. GAUSSIAN PROCESS REGRESSION

This section briefly presents the standard Gaussian process regression [1]. Given a set of N input-output pairs $\{\boldsymbol{x}_k, y_k\}_{k=1}^N$, $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$, $y \in \mathbb{R}$ related according to an arbitrary model such as

$$y_k = \psi(\boldsymbol{x}_k) + \eta_k \tag{1}$$

with $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$, and $\psi \in \mathcal{H}$ considered to be a function of a reproducing kernel Hilbert space $\mathcal{H}$ defined over a compact set $\mathcal{X}$, GPs assume a Gaussian functional distribution as prior for the function $\psi|\boldsymbol{x}_k \sim \mathcal{N}(0, \kappa(\boldsymbol{x}_k, \boldsymbol{x}_k))$, where κ is a kernel function such that $\kappa(\cdot, \boldsymbol{x}) \in \mathcal{H}$. For a set of input points $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$ the prior distribution for ψ becomes $\boldsymbol{\psi}|\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K})$, where $\boldsymbol{K} \in \mathbb{R}^{N\times N}$ is the Gram matrix with entries $[\boldsymbol{K}]_{ij} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$. For a given set of measurements $\boldsymbol{y} = [y_1, \ldots, y_N]^\top$ associated with the positions $\boldsymbol{X}$, the prior distribution becomes

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{K} + \sigma_\eta^2 \boldsymbol{I}). \tag{2}$$

The predictive distribution allows one to "predict" the value of the function $\psi_\star$ for a new input value $\boldsymbol{x}_\star$. Thus, we have $\psi_\star|\boldsymbol{x}_\star \sim \mathcal{N}(0, \kappa_{\star\star})$, where $\kappa_{\star\star} \triangleq \kappa(\boldsymbol{x}_\star, \boldsymbol{x}_\star)$. Since $\boldsymbol{y}$ and $\psi_\star$ are jointly Gaussian their joint PDF is given by

$$\begin{bmatrix} \boldsymbol{y} \\ \psi_\star \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{K} + \sigma_\eta^2 \boldsymbol{I} & \boldsymbol{\kappa}_\star \\ \boldsymbol{\kappa}_\star^\top & \kappa_{\star\star} \end{bmatrix}\right), \tag{3}$$

where $\boldsymbol{\kappa}_\star \triangleq [\kappa(\boldsymbol{x}_1, \boldsymbol{x}_\star), \ldots, \kappa(\boldsymbol{x}_N, \boldsymbol{x}_\star)]^\top$. Finally the predictive distribution can be obtained by conditioning $\psi_\star$ over the observation and the its respective positions as

$$\psi_\star|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}_\star \sim \mathcal{N}(\mu_{\psi|\boldsymbol{y},\boldsymbol{X},\boldsymbol{x}_\star}, s_{\psi|\boldsymbol{y},\boldsymbol{X},\boldsymbol{x}_\star}) \tag{4}$$

with $\mu_{\psi|\boldsymbol{y},\boldsymbol{X},\boldsymbol{x}_\star} = \boldsymbol{\kappa}_\star^\top \left(\boldsymbol{K} + \sigma_\eta^2 \boldsymbol{I}\right)^{-1} \boldsymbol{y}$, and $s_{\psi|\boldsymbol{y},\boldsymbol{X},\boldsymbol{x}_\star} = \kappa_{\star\star} - \boldsymbol{\kappa}_\star^\top \left(\boldsymbol{K} + \sigma_\eta^2 \boldsymbol{I}\right)^{-1} \boldsymbol{\kappa}_\star$. Using the GP distribution above, the distribution of $p(y_\star)$ can be obtained by replacing (4) in (1)

$$\hat{p}(y_\star) = \mathcal{N}(\mu_{\psi|\boldsymbol{y},\boldsymbol{X},\boldsymbol{x}_\star}, s_{\psi|\boldsymbol{y},\boldsymbol{X},\boldsymbol{x}_\star} + \sigma_\eta^2). \tag{5}$$

The Bayesian framework also provides strategies to estimate free parameters, such as the kernel parameters $\boldsymbol{\theta}$ and the noise power σ_η^2. The classical approach [1] aims at maximizing the marginal likelihood $p(\boldsymbol{y}|\boldsymbol{X}, \sigma_\eta^2, \boldsymbol{\theta})$ with respect to $(\sigma_\eta^2, \boldsymbol{\theta})$.

3. KOLMOGOROV-SMIRNOV BASED DETECTION

In this section we present the detection approach used to determine if data should be replaced in a space region as new data arrives. We assume that the estimated field is stationary during a period T where new noisy observations may arrive for any given input $\boldsymbol{x}$ used in learning the GP. The GP regression provides a distribution characterization at any point $\boldsymbol{x} \in \mathcal{X}$. However, directly using new measurements for updating the GP may introduce new errors such as extra uncertainty due to noisy inputs. This necessitates the institution of a change detection metric, which should present the following desirable characteristics: 1) be computed recursively; 2) be performed using a single new observation; 3) use the Gaussian characterization of field; 4) provide a distribution for the test statistics at least under hypothesis $\mathcal{H}_0$, allowing one to design the detector to a given probability of false alarm (P_{FA}).

Let $y^{(t)}$ be new samples for a given location $\boldsymbol{x}$, with $t = 1, \ldots, L$ being a time index, drawn from a PDF with cumulative distribution function (CDF) $F(\cdot)$, and $G(\cdot)$ be the CDF associated with the GP at $\boldsymbol{x}$. Our hypotheses are given by

$$\begin{cases} \mathcal{H}_0: & F = G \\ \mathcal{H}_1: & F \neq G. \end{cases} \tag{6}$$

Since the distribution F is not readily available we follow the reasoning proposed in [16] and [17], wherein the F was replaced by the single sample estimator $\hat{F}$ such that the empirical CDF for the t-th sample $y^{(t)}$ is given by

$$\hat{F}(\xi) = \begin{cases} 0, & \text{if } \xi < y^{(t)}, \\ 1, & \text{otherwise.} \end{cases} \tag{7}$$

Next, we present the Kolmogorov-Smirnov (KS) statistics that will be used as the building block of the change detector.

3.1. The one-sample Kolmogorov-Smirnov statistic

Let $\hat{F}$ and $\hat{G}$ be the empirical distributions with 1 and K samples, respectively. Then, the KS statistic is given by

$$D(K) = \sup_y |\hat{F}^{[1]}(y) - \hat{G}^{[K]}(y)| \,. \tag{8}$$

In [16], the authors showed that for any $K > 1$ odd, $D(K)|\mathcal{H}_0$ is distributed according to a discrete uniform distribution confined in the interval $[0.5, 1]$, with support $S = \{(K+1)/2K, (K+3)/2K, \ldots, K\}$. One should note that as $K \to \infty$, $\hat{G}^{[K]} \to G$, the discrete support S tends to the continuous real line segment $[0.5, 1]$. Thus, $D(K) \to D(\infty) \triangleq D$, with

$$D = \sup_y |\hat{F}^{[1]}(y) - G(y)| \tag{9}$$

It is thus reasonable to approximate the distribution of $D|\mathcal{H}_0$ as a uniform continuous distribution with support $S_c = [0.5, 1]$.

3.2. KS-Detector

Assuming that an arbitrary, usually small number of samples $y^{(t)}$ may arrive in a given interval T, we can combine the individual statistics D_t to form a new KS-based test statistic. In [17], considering the mean KS statistics showed good performance for measuring the robustness of sequential methods. Following the same reasoning, we have:

$$\bar{D} = \frac{1}{L} \sum_{t=1}^{L} D_t \,, \tag{10}$$

which leads to the detector:

$$\bar{D} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \tau \,, \tag{11}$$

with $\tau \in (0, 1)$ being the detection threshold.

The detection statistic in (10) is a linear combination of L independent uniformly distributed random variables D_t and its distribution has being approximated as a Gaussian PDF by virtue of the central limit theorem. Here, however, due to the constraint that L might

be very small we recur to a more exact solution. For this, we use the result that the sum of i.i.d. random variables distributed according to $\mathcal{U}_{[0,1]}$ is distributed according to an Irwin-Hall distribution [18, 19].

We propose to modify the test statistics (10) in order to use this result. First, note that since $D_t \sim \mathcal{U}_{[0.5,1]}$, we have that $\widetilde{D}_t = 2D_t - 1 \sim \mathcal{U}_{[0,1]}$ and the test statistics can be reformulated as

$$\widetilde{D} = \frac{1}{L}\sum_{t=1}^{L} \widetilde{D}_t \ , \tag{12}$$

leading to the following detector

$$\widetilde{D} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} \tau. \tag{13}$$

Under $\mathcal{H}_0$, we have that $L\widetilde{D}$ is distributed according to a Irwin-Hall distribution whose CDF is given by:

$$\Theta_{L\widetilde{D}}(y) = \frac{1}{L!}\sum_{q=0}^{\lfloor y \rfloor}(-1)^q \binom{L}{q}(y-q)^L. \tag{14}$$

For a complete characterization of a detector, the distribution under $\mathcal{H}_1$ is also required. However, this distribution depends on sensible information regarding the actual change happening to the modeled function ψ. Such information is rarely available and, thus, we will resort to empirical performance curves as will be discussed later.

The final missing piece in designing the detector is a strategy for selecting the detection threshold τ. Typically one might select this value in order to obtain a low probability of false-positives, that is, a low P_{FA}. The P_{FA} is given by

$$P_{\text{FA}} = P(\widetilde{D} > \tau | \mathcal{H}_0). \tag{15}$$

Thus, by multiplying (13) by L in both sides, we can use (14) to determine τ for a given P_{FA}:

$$\tau = \frac{1}{L}\Theta^{-1}_{L\widetilde{D}}(1 - P_{\text{FA}}) \ . \tag{16}$$

4. SIMULATION RESULTS

4.1. Toy Problem

To perform all simulations we consider the following toy example where each observed noisy data point is given by

$$y_k^{(t)} = f(x_k) - \varphi(x_k)u(t-1) + n_k^{(t)} \tag{17}$$

where $t = 0, 1, \ldots$ is the time index of the observations,

$$f(x_k) = \cos(x_k)\exp(-0.3x_k) \tag{18}$$

is the underlying generative function and $n_k^{(t)}$ is additive noise distributed as $\mathcal{N}(0, \sigma_n^2)$. Finally, φ is a function modeling an arbitrary localized field change and $u(t)$ is the Heaviside step function.

In order to model a smooth and continuous field change over space we consider the following bump-shaped function:

$$\varphi(x) = \begin{cases} \xi \exp\left(1 - \frac{1}{r-(x-c)^2}\right), & \text{if} (x-c)^2 < r \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

where $\xi > 0$ controls the maximum function amplitude, r is the radius of the bump, and c is the point were the function is centered.

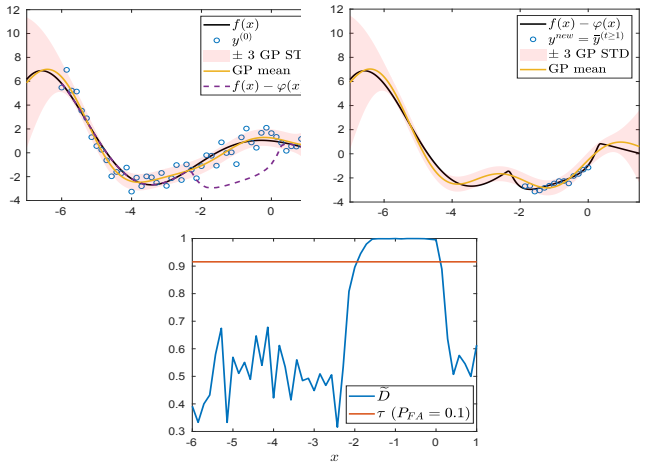


Fig. 1: The toy example before any detection (top left), after the detection and replacement of points (top right), and the detection statistics using $L = 3$ new samples (bottom).

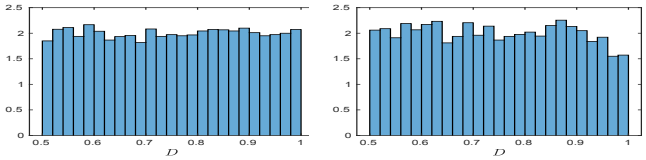


Fig. 2: Histomgrams of D under $\mathcal{H}_0$ using the true distribution $p(y)$ (left) and the distribution $\hat{p}(y)$ obtained using the GP (right).

For the simulations we consider different values of σ_n^2 in order to obtain SNRs (signal-to-noise ratios) of 5, 10 and 15 dB, in addition to different values for ξ and different maximum values of t, that is, $L = 3$ and $L = 10$. Here the SNR was defined as $\text{SNR} = 10\log_{10}\left(\mathbb{E}[|f(x_k) - \mathbb{E}[f(x_k)]|^2]/\sigma_n^2\right)$.

Fig. 1 presents an example of the devised strategy for a 10 dB SNR and $\xi = 3\sigma_n$. In the top left panel the initial function $f(x)$, the initial noisy observations ($y_k^{(0)}$), the GP fit, and the change that will occur in the function after training (dashed line) are presented. In the top right panel the function after the change happened and the GP fit incorporating the samples detected as $\mathcal{H}_1$ (blue circles are the average of the L samples for each position) are presented. The bottom panel presents the test statistics $\widetilde{D}$ and the threshold τ computed for a $P_{\text{FA}} = 0.1$. It is evident that with as few as 3 samples per location, the detection approach was able to identify the change in the function with a high degree of precision.

Next, we will analyse the distributions of the test statistics and the performance of the detectors for exact and uncertain inputs. Since it is not possible to obtain the test statistics distribution under $\mathcal{H}_1$, we measure the detection performance using empirical *receiver operating characteristic* (ROC) curves. The empirical ROCs can be obtained by computing empirical probabilities of false alarm and detection from the available data [20]. All ROCs displayed in this paper were averaged over 100 Monte Carlo runs.

4.2. Test statistics distribution

In Section 3, we showed how the one-sample KS statistic distribution under $\mathcal{H}_0$ can be approximated as an uniform distribution $\mathcal{U}_{[0.5,1]}$. To confirm this hypothesis we performed the test in a scenario where

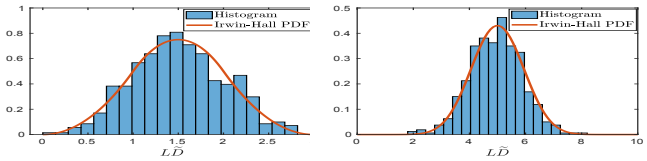


Fig. 3: Histomgrams of $L\widetilde{D}$ under $\mathcal{H}_0$ using the distribution $\hat{p}(y)$ obtained using the GP using $L = 3$ (left) and $L = 10$ (right).

no change was made to the underlying function (*i.e.*, $\varphi(x) = 0$, $\forall x$). Fig. 2 presents the histograms performed using 2000 samples for the one-sample test statistics D. In the left panel we considered G to be the CDF associated with the true $p(y)$ distribution while in the right panel we considered G to be the CDF associated with Gaussian process estimated PDF $\hat{p}(y)$ (see, Eq. (5)). By comparing both histograms one can see that the uniform assumption is maintained even when replacing G by $\hat{G}$ given by the GP.

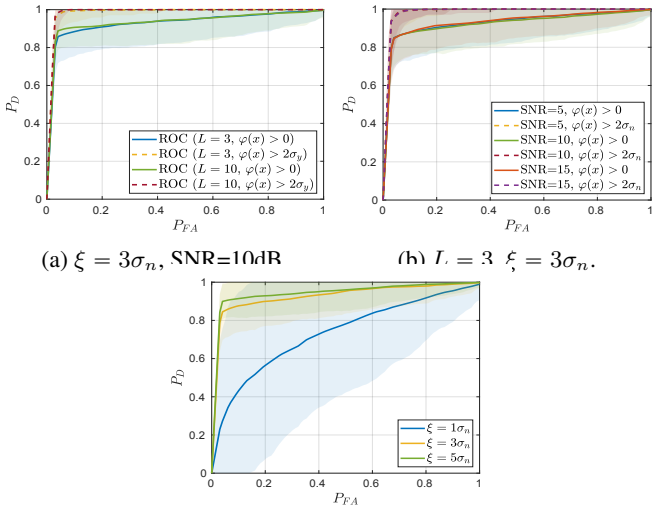
We also analyze the distribution of $L\widetilde{D}$, the histograms for which, accompanied by the plot of the analytucal Irwin-Hall PDF, are displayed in Fig. 3 for $L = 3$ (left) and $L = 10$ (right). In both panels we used $G \triangleq \hat{G}$ obtained using the GP. By analyzing Figure 3 we can notice the clear agreement between the Irwin-Hall PDF and the histograms which indicates that our assumptions are reasonable.

4.3. Detection performance analysis for exact inputs

In this section, we analyze the performance of the proposed detection and estimation methodology using new data observed for exact inputs. We performed 100 Monte Carlo runs for each simulation and plotted the average empirical ROC curves, see Fig. 4, for different L (panel a), SNR (panel b) and maximum bump function amplitude ξ (panel c). In all subplots the solid lines represent ROC curves for which detection errors include all miss-detected points for which $\varphi(x) \neq 0$. For the ROCs with dashed lines, we considered detection errors only the miss-detections occurring for points x for which $\varphi(x) > 2\sigma_n$. This is reasonable since $\varphi(x)$ is smooth and goes to zero in the edges. The plots indicates that the proposed detector is not very sensitive to the number of new samples presenting high detection probabilities even for $L = 3$.

4.4. Detection performance analysis for uncertain inputs

In this section we examine the deterioration of the proposed detector when location errors exist in the new samples $y_k^{(t)}$, $t > 0$. For this, we modify the model as $y_k^{(t)} = f(x_k+\varepsilon_k^{(t)})-\varphi(x_k+\varepsilon_k^{(t)})u(t-1)+n_k^{(t)}$ where $\varepsilon_k^{(t)} \sim \mathcal{N}(0, \sigma_x^2)$, and σ_x^2 is the variance of the location uncertainty. It is important to discuss the notion of large and small σ_x values. Since we are considering additive Gaussian noise, one should recall that most (99.7%) of the realizations of the random variable ε_k are expected to belong to the interval $[-3\sigma_x, +3\sigma_x]$. Thus, the notion of large, or small, values of σ_x depends on the variation of the function $f(x)$ in such intervals around x. In the case of the toy example, we can see that $\sigma_x = 0.1$ can be taken to be a moderate value, while $\sigma_x = 0.5$ is very large. This can be seen if we consider $f(2+\varepsilon)$. If $\sigma_x = 0.1$ then $f(2+\varepsilon) \in [-0.07, -0.33]$ which is a considerable variation and not by any means negligible. However, if $\sigma_x = 0.5$, then $f(2+\varepsilon) \in [-0.33, 0.75]$, the range of which is considerably larger than the measurement noise power $\sigma_n^2 = 0.65$ used to provide an SNR of 10 dB. This situation is even



(a) $\xi = 3\sigma_n$, SNR=10dB (b) $L = 3$, $\xi = 3\sigma_n$.

(c) $L = 3$, SNR=10dB

Fig. 4: ROC curves for different values of L, different SNR, and different ξ. Dashed ROCs where built by labeling as $\mathcal{H}_1$ only points for which $\varphi(x) > 2\sigma_n$.

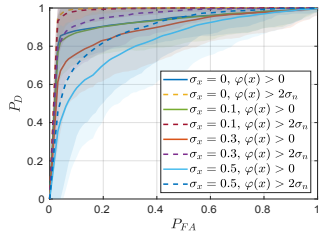


Fig. 5: ROCs for different σ_x values.

worse if we consider $x = -3$ due to the larger slope of f in this point: $f(-3+\varepsilon) \in [-0.05, 2.72]$ and $[-2.68, 6.86]$ for $\sigma_x = 0.1$ and 0.5, respectively.

We performed 100 MC runs of the proposed algorithm and computed the ROCs, displayed in Fig. 5, for different values of σ_x, $L = 3$ and σ_n adjusted to provide a SNR of 30 dB. It can be seen that the proposed detector performs well under moderate input noise power ($\sigma_x = 0.1$). When σ_x increases to 0.3 and 0.5 a clear degradation is noticeable. However, larger function changes ($\phi(x) > 2\sigma$) can still be detected with relatively high probability of detection. For instance, for a $P_{FA} = 0.1$, we have $P_D \approx 0.84$ and $P_D \approx 0.7$ for $\sigma_x = 0.3$ and $\sigma_x = 0.5$, respectively.

5. CONCLUSIONS

In this paper we proposed a KS-based change point detection method with Gaussian processes. The ability of discriminating which new data to include in the GP is important in order to i) keep the database within a bounded dimension, and ii) avoid systematic replacement of the noiseless initial dataset. The proposed methodology allowed the construction of a recursive detection statistics with exact distribution under $\mathcal{H}_0$. Once the detection is performed the GP can be adapted by simply replacing measurements for the respective locations. Simulation clearly shows the accuracy of the proposed detector when no input noise is present and the robustness of the detector under the presence of location errors.

6. REFERENCES

[1] C. K. I. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2, MIT press Cambridge, MA, 2006.

[2] T. Imbiriba, J. C. M. Bermudez, J.-Y. Tourneret, and C. Richard, "Detection of nonlinear mixtures using Gaussian processes: Application to hyperspectral imaging," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 7949–7953.

[3] T. Imbiriba, J. C. M. Bermudez, C. Richard, and J.-Y. Tourneret, "Nonparametric detection of nonlinearly mixed pixels and endmember estimation in hyperspectral images," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1136–1151, 2016.

[4] P. I. Frazier, "A tutorial on Bayesian optimization," *arXiv preprint arXiv:1807.02811*, 2018.

[5] D. Dardari, P. Closas, and P. M. Djurić, "Indoor tracking: Theory, methods, and technologies," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1263–1278, April 2015.

[6] Eva Arias-de Reyna, Pau Closas, Davide Dardari, and Petar M Djuric, "Crowd-Based Learning of Spatial Fields for the Internet of Things: From Harvesting of Data to Inference," *IEEE Signal Processing Magazine*, vol. 35, no. 5, pp. 130–139, 2018.

[7] I. Santos, J. J. Murillo-Fuentes, and P. M. Djurić, "Recursive Estimation of Dynamic RSS Fields Based on Crowdsourcing and Gaussian Processes," *IEEE Transactions on Signal Processing*, vol. 67, no. 5, pp. 1152–1162, 2019.

[8] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, "Simultaneous localization and mapping: A survey of current trends in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, Sep. 2017.

[9] Dengsheng Lu, P Mausel, E Brondizio, and Emilio Moran, "Change detection techniques," *International journal of remote sensing*, vol. 25, no. 12, pp. 2365–2401, 2004.

[10] Samaneh Aminikhanghahi and Diane J Cook, "A survey of methods for time series change point detection," *Knowledge and information systems*, vol. 51, no. 2, pp. 339–367, 2017.

[11] Yunus Saatçi, Ryan D Turner, and Carl Edward Rasmussen, "Gaussian process change point models," in *ICML*, 2010, pp. 927–934.

[12] Hossein Keshavarz, Clayton Scott, and XuanLong Nguyen, "Optimal change point detection in Gaussian processes," *Journal of Statistical Planning and Inference*, vol. 193, pp. 151–178, 2018.

[13] Ryan Prescott Adams and David JC MacKay, "Bayesian online changepoint detection," *Technical report, University of Cambridge, Cambridge*, 2007.

[14] Varun Chandola and Ranga Raju Vatsavai, "A Gaussian process based online change detection algorithm for monitoring periodic time series," in *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 2011, pp. 95–106.

[15] William Herlands, Andrew Wilson, Hannes Nickisch, Seth Flaxman, Daniel Neill, Wilbert Van Panhuis, and Eric Xing, "Scalable Gaussian processes for characterizing multidimensional change surfaces," in *Artificial Intelligence and Statistics*, 2016, pp. 1013–1021.

[16] P. M. Djurić and J. Míguez, "Assessment of nonlinear dynamic models by Kolmogorov–Smirnov statistics," *IEEE transactions on signal processing*, vol. 58, no. 10, pp. 5069–5079, 2010.

[17] P. M. Djurić, M. F. Bugallo, P. Closas, and J. Míguez, "Measuring the robustness of sequential methods," in *2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. IEEE, 2009, pp. 29–32.

[18] N. Johnston, S. Kotz, and N. Balakrishnan, "Continuous univariate distributions (volume 2)," 1994.

[19] James E Marengo, David L Farnsworth, and Lucas Stefanic, "A Geometric Derivation of the Irwin-Hall Distribution," *International Journal of Mathematics and Mathematical Sciences*, vol. 2017, 2017.

[20] S. M. Kay, *Fundamentals of statistical signal processing: detection theory*, Prentice-hall, 1998.