Rough-Fuzzy CPD: a gradual change point detection algorithm

Ritwik Bhaduri[1] · Subhrajyoty Roy[2] · Sankar K. Pal[3]

Abstract

Changepoint detection is the problem of finding abrupt or gradual changes in time series data when the distribution of the time series changes significantly. There are many sophisticated statistical algorithms for solving changepoint detection problem, although there is not much work devoted towards gradual changepoints as compared to abrupt ones. Here we present a new approach to solve the changepoint detection problem using the fuzzy rough set theory which is able to detect such gradual changepoints. An expression for the rough-fuzzy estimate of changepoints is derived along with its mathematical properties concerning fast computation. In a statistical hypothesis testing framework, the asymptotic distribution of the proposed statistic on both single and multiple changepoints is derived under the null hypothesis enabling multiple changepoint detection. Extensive simulation studies have been performed to investigate how simple crude statistical measures of disparity can be subjected to improve their efficiency in the estimation of gradual changepoints. Also, the said rough-fuzzy estimate is robust to signal-to-noise ratio, a high degree of fuzziness in true changepoints, and also to hyperparameter values. Simulation studies reveal that the proposed method beats other methods of gradual changepoint detection (including MJPD, HSMUCE, fuzzy methods like FCP, FCMLCP etc) and also popular crisp methods like Binary Segmentation, PELT, and BOCD in detecting gradual changepoints. The applicability of the estimate is demonstrated using multiple real-life datasets including Covid-19. We have developed the python package `roufcp` for broader dissemination of the methods.

Keywords Gradual changepoint detection · Fuzzy set · Rough set · Python package `roufcp`

Ritwik Bhaduri and Subhrajyoty Roy are contributed equally to this work.

✉ Ritwik Bhaduri
    ritwikbhaduri@g.harvard.edu

    Subhrajyoty Roy
    subhrajyoty_r@isical.ac.in

    Sankar K. Pal
    sankar@isical.ac.in

[1] Department of Statistics, Harvard University, 1 Oxford Street, 02138, Massachusetts, Cambridge, USA

[2] Interdisciplinary Statistical Research Unit, Indian Statistical Institute, 203 B.T. Road, 700108, West Bengal, Kolkata, India

[3] Center for Soft Computing Research, Indian Statistical Institute, 203 B.T. Road, 700108, West Bengal, Kolkata, India

1 Introduction

Changepoint Detection (CPD) is the process of detecting significant changes in the data generating process of a time series data. These changes are of two types, abrupt and gradual, though many authors define changepoints as only the abrupt ones. Abrupt changepoints demonstrate very fast and rapid changes in the underlying states of the systems, thus its primary applications include quality control, system reliability engineering, and fault detection and monitoring in industrial plants (Basseville and Nikiforov 1993). Changepoint analysis is also used for time series modeling and signal processing, many studies rely on CPD techniques to segment the study period and then apply different time series methodologies for each homogeneous segment (Lemire 2007). Changepoint detection has successfully proved its practicability in various domains such as finance, sound processing, climatology, network traffic data, etc (see Chen and Gupta 2012 for further references).

In contrast, gradual changepoints do not correspond to rapid discontinuous changes but only smooth continuous

ones, which allow the underlying parameters to vary smoothly between different regimes. In other words, the underlying distribution or the property of the time series remains approximately constant over some time, and then slowly starts to change into another regime followed by another period with constant properties. We refer to the midpoint of the period where the smooth change takes place as a gradual changepoint. In different domains such as linguistics (Sharma and Sankaran 2011), paleoclimatology (Trauth et al. 2018), paleontology (Rose and Bown 1986), and paleobiology (Hunt 2008), studies have shown that the changes in different phenomena are often gradual in nature and not abrupt. For example, the evolution of a language over a few hundred years contains only gradual changes, and it is thus difficult to point out any particular period where the archaic language has evolved into its current state. However, such a changepoint has its use, since without detecting such a changepoint it is extremely difficult to obtain a language model completely, while segmenting the study period into homogeneous regions may allow us to model the language in different periods more clearly. More motivating examples of gradual changepoint can be found in (Vogt and Dette 2015).

It has been established that the changepoint detection algorithms developed to tackle the abrupt changes are inefficient in estimation due to relatively larger variance, in presence of a gradual changepoint (Hušková 1998). One particular problem with these algorithms is that they assume a definitive change in the distribution at a fixed point of time which is not true for a gradual changepoint. In contrast, modeling these points of changes in distribution using fuzzy set theory will allow the presence of a gradual changepoint that is difficult to pinpoint. In other words, the gradual changepoint is characterized by a continuum value in [0,1], unlike in {0,1} as in the case of abrupt changepoint. Also, the incorporation of rough sets in the changepoint detection algorithm would allow us to model the ambiguity about the knowledge of the distribution (as opposed to the probabilistic approach). In this paper, we propose a technique that can incorporate these fuzzy-rough set theory-based modifications in most of the existing abrupt changepoint detection algorithms. As will be shown in this paper, this modification would allow even the simple abrupt CPD algorithms to detect the gradual changepoints with much higher accuracy.

2 Existing literature and our contribution

Analysis of abrupt changepoint detection for quality control and industrial applications dates back to the 1950's Page (1955). Since then, several parametric and nonparametric methods for change detection have been proposed (Basseville and Nikiforov 1993; Brodsky and Darkhovsky 2013). Following Hinkley's idea of likelihood-based single changepoint detection (Hinkley 1970; Scott and Knott 1974) extended it to the detection of multiple changepoints based on Binary Segmentation, which was popularized as a general method to detect abrupt changepoints. Some recent and popular algorithms for CPD include Pruned Exact Linear Time (PELT) Killick et al. (2012), Bayesian Change-point Detection analysis (Turner et al. 2009; Adams and MacKay 2007), Kernel Change-point Detection (Harchaoui et al. 2009), Change-point detection using nonparametric statistical measures (Haynes et al. 2017).

In comparison to the vast literature on the detection of abrupt changepoints (van den Burg and Williams 2020), the investigation of the gradual changepoint is scarce. Although some recent studies (Liang and Xu 2021) have focused on gradual changes in mean, they retain abrupt changes in variance as the defining property of a changepoint. Some early investigations completely in gradual changepoint setup used to measure the performance of control charts under gradual changes (Gan 1992). Few existing parametric methods (Hušková 1998) consider specific forms of smooth changes of mean and perform likelihood-based tests to detect gradual changepoints. On the other hand, nonparametric methods based on p-value (Mallik et al. 2011) and modifications on CUSUM principle (Vogt and Dette 2015) rely on some suitable modifications of existing abrupt changepoint detection algorithms for the gradual change setup. Another approach to gradual changepoint detection is via the application of abrupt CPD techniques in a multiscale setup (Wu and Zhou 2020), of which SMUCE (Frick et al. 2014) and HSMUCE (Pein et al. 2017) are two popular methods. However, there is no extensive empirical investigation of their performances in presence of a gradual changepoint.

Only a few investigations directed towards the detection of gradual changepoints are based on fuzzy techniques. A major portion of them (Chang et al. 2015) uses the fuzzy c-means clustering method or its variants to segment the period of study into homogeneous segments. Other approaches include fuzzy regression algorithms (Chang et al. 2015), and neural networks (Angelo et al. 2011). No approach considers the discernibility issues that are prevalent in the examples with gradual changepoints, for instance, observations of two close time points would naturally come from a similar kind of distribution, converting each of the segments into a rough set instead of a crisp one. The present paper aims to provide a general methodology based on fuzzy rough sets to increase the efficiency of any statistical measure that can be used to detect changepoints.

The novelty of this paper is three-fold. The first aspect is that we mathematically model the uncertainty associated with gradual changepoints as a fuzzy set and the indiscernibility between very close time points using rough set theory. This enables us to devise an algorithm to transform most of the abrupt CPD algorithms into a

gradual CPD algorithm, which is relatively robust to both the low signal-to-noise ratio of the data and the high degree of fuzziness of the actual changepoint. Simulation studies show that our estimates remain fairly accurate over an extensive range of hyperparameter values furthering the robustness of the proposed method. The second aspect of this paper is that we connect the image segmentation problem from the field of Pattern Recognition and the changepoint detection problem from the statistical domain. This connection was discussed by (Chatterjee and Kar 2017), where they formally expressed image segmentation as a special case of the changepoint detection problem. However, our work can be interpreted as going in a reverse direction, by a generalization of methods proposed for image segmentation by (Sen and Pal 2009) for the problem of changepoint detection. The third component of novelty is reducing the computational complexity of the aforementioned image segmentation algorithm by a series of results and fitting the algorithm in the context of the statistical changepoint detection problem. To formally present the proposed algorithm in conjunction with a statistical hypothesis testing framework, we derive the asymptotic distribution of the proposed rough-fuzzy estimator under some suitable regularity conditions. This enables us to test for false positives of the detected changepoints and also allows for multiple changepoint detections. The performance of the proposed method has been illustrated using extensive simulation studies, as well as different real-data applications, including estimation of behavioral changes during the covid-19 period, and detection of changes in language using US presidential speech data. We also observe that our method outperforms several existing changepoint detection methods in different setups. The implementation of the method is made available through a python package `roufcp`, bearing the acronym for the name of the proposed algorithm Rough-Fuzzy CPD.

The rest of the paper is organized as follows: In Section 3, some mathematical concepts are briefly described in connection to the proposed method, which is ultimately described in Section 4.2 in detail. Different ways to speed up the computation of the algorithm are described in Section 4.1. In Section 4.3, we provide a theoretical justification of the proposed method in the light of a hypothesis testing framework with an example. Section 5 provides support for the proposed roufCP method based on extensive simulation studies and also establishes the superiority of the method in comparison to several existing changepoint detection techniques. Finally, Section 6 demonstrates its usefulness in real-life examples, which follows with concluding remarks in Section 7. For brevity of the presentation, all the proofs of the theorems have been provided in the appendix.

3 Mathematical preliminaries

3.1 Problem statement

The mathematical setup of the problem of changepoint detection starts with a multivariate time series signal $\{y_t\}_{t=1}^T$, where each $y_t \in \mathbb{R}^p$. There also exist some fixed points $1 < t_1^* < t_2^* < \dots t_k^* < T$, such that within the interval $[t_{(i-1)}, t_i]$, the time series observations follow the same probability distributional model, but these models differ from one interval to another. In particular,

$$
y_t \sim \begin{cases} F_0(t) & \text{if } 1 \le t < t_1 \\ F_1(t) & \text{if } t_1 \le t < t_2 \\ \dots & \dots \\ F_{(k-1)}(t) & \text{if } t_{(k-1)} \le t < t_k \\ F_k(t) & \text{if } t_k \le t \le T \end{cases}, \tag{1}
$$

and $F_i \ne F_{(i+1)}$ for $i = 0, 1, \dots (k-1)$. A special case follows when the signal is assumed to be piece-wise stationary. The problem of CPD deals with identifying the set of points $\{t_1, t_2, \dots t_k\}$ where the distribution or behavior of the signal changes, from the knowledge of the available time series observations y_ts. The knowledge of the number of changepoints k may or may not be known apriori, in which case it also has to be estimated from the data.

A general framework for existing CPD algorithms is provided in a comprehensive review by (Truong et al. 2020). Most of the existing algorithms involve minimization of a criterion function $V(\mathcal{T}, y)$ to obtain the best set of changepoints $\mathcal{T}$,

$$
V(\mathcal{T}, y) = \sum_{i=1}^k c\left(\{y_t\}_{t_{(i-1)}}^{t_i}\right) + \text{Penalty}(\mathcal{T}). \tag{2}
$$

3.2 Fuzzy set

Fuzzy sets, proposed by (Zadeh 1965), provide a mathematical means of modeling vagueness and imprecise information. They are a generalization of crisp sets or normal sets. A fuzzy set is represented by an ordered pair $\langle X, \mu \rangle$. Denoting $\mathbb{U}$ as the universe of objects, and $X \subseteq \mathbb{U}$, the membership function for the set X can be described as $\mu : \mathbb{U} \to [0, 1]$, indicating the degree of inclusion of an element $x \in \mathbb{U}$ into the set X. Consequently, three cases are possible based on the value of $\mu(x)$:

1. Not included in $\langle \mathbf{X}, \mu \rangle$ if $\mu(x) = 0$.
2. Partially included in $\langle \mathbf{X}, \mu \rangle$ if $0 < \mu(x) < 1$.
3. Fully included in $\langle \mathbf{X}, \mu \rangle$ if $\mu(x) = 1$.

While fuzzy sets have been extensively used in fields like control systems for electronic devices, decision-making support systems in businesses, prediction systems in finance, quantitative pattern analysis for industrial quality assurance, and uncertainty modeling in pattern recognition, image processing and vision (see Pal and Dutta-Majumder 1986; Pathak and Pal 1986 for references), its use in changepoint analysis and statistics, in general, has been very limited. In this paper, we shall model the membership of a data point to a particular distribution as the membership function of a fuzzy set to characterize the ambiguity in the data generation process.

3.3 Rough set

The notion of rough sets, as introduced by (Pawlak 1982), is defined as follows: Let $\mathcal{A} = \langle \mathbb{U}, A \rangle$ be an information system, where $\mathbb{U}$ is the universe of all objects as mentioned before, and A be the set of attributes which are used to identify two elements of $\mathbb{U}$. Let $B \subseteq A$ be a subset of attributes and $X \subseteq \mathbb{U}$. We can approximate the set X using only the information contained in B by constructing lower and upper approximations of X. Let $[x]_B$ denote the equivalence class of object x relative to I_B (equivalence relation induced by the variables in B). In rough set theory, the upper and lower approximations of the set X under I_B are defined below,

- $\underline{B}X$ = B-Lower approximation of X in $\mathbb{U}$ $= \{x \in \mathbb{U} : [x]_B \subseteq X\}$
- $\overline{B}X$ = B-upper approximation of X in $\mathbb{U}$ = $\{x \in \mathbb{U} : [x]_B \cap X \neq \phi\}$

So, the lower approximation of a set X relative to B in $\mathbb{U}$ are the elements in $\mathbb{U}$ which can be certainly classified as elements of X based on B. Intuitively, it is the set of all elements x in $\mathbb{U}$ such that the equivalence class containing x is a subset of the target set X. In other words, the lower approximation is the set of objects that are certainly members of the target set X. On the other hand, the upper approximation is the complete set of objects x such that the equivalence class containing x has a non-empty intersection with X. Intuitively, it is the set of elements in $\mathbb{U}/B$ that cannot be positively (i.e., unambiguously) classified as belonging to the complement of the target set X. In contrast to the lower approximation, the upper approximation is the complete set of objects that are certainly as well as possible members of the target set X.

The pair $\langle \underline{B}X, \overline{B}X \rangle$ denotes the rough representation of the crisp set X with respect to B. This rough representation actually captures the uncertainty in defining X because of the incomplete information provided by the subset of attributes B. Numerical characterization of the roughness of X can be obtained as follows (Pal and Skowron 1999)

$$\rho_B(X) = 1 - \frac{|\underline{B}X|}{|\overline{B}X|} \tag{3}$$

Therefore, $\rho_B = 0$ means, the set X is crisp or exact (with respect to B) and conversely, $\rho_B > 0$ means, X is rough, i.e. ambiguous, with respect to B. This vague definition of X in $\mathbb{U}$ (in terms of lower and upper approximations) signifies incompleteness of knowledge about $\mathbb{U}$.

An amalgamation of fuzzy theory and rough theory can be used to characterize systems in which the sets and their elements themselves are ambiguous due to partial knowledge about their attributes, as well as the elements themselves can belong to the set partially. Similar to the lower and upper approximations of a crisp set, the definition of such approximations for a fuzzy set can also be established. While the definition of lower and upper approximations of a set may vary based on the nature of set X (crisp or fuzzy) as well as the nature of the relationship between elements of the universe $\mathbb{U}$, here, we are only concerned with approximations of a fuzzy set X with respect to an indiscernibility relation R which is either an equivalence relation (crisp or fuzzy) or a tolerance relation (crisp or fuzzy). The expressions of upper and lower approximations for both these cases have been derived by Sen and Pal (2009). Since the set approximations obtained using equivalence relations are not always smooth, we will restrict our consideration only to the tolerance relations. The general implicit expressions of these approximations are given below, while a more explicit form of such approximations will be derived later in Eq. 4.1.

A tolerance relation (crisp or fuzzy) is a (crisp or fuzzy) relation that satisfies (crisp or fuzzy) reflexivity and symmetry. Unlike equivalence relations, tolerance relations are not necessarily transitive. When R is a tolerance relation, the space $\langle \mathbb{U}, R \rangle$ is called tolerance approximation space. Associated with a tolerance relation R, there is a membership function S_R of the relation itself, where $S_R(u, v)$ denotes the membership value of the pair (u, v) belonging to the relation R. With the help of these, the lower and upper approximations of any crisp or fuzzy set X are obtained as follows;

$$\begin{aligned}\underline{R}X &= \{(u, \underline{M}(u)) \mid u \in \mathbb{U}\} \\ \overline{R}X &= \{(u, \overline{M}(u)) \mid u \in \mathbb{U}\}\end{aligned} \tag{4}$$

where the approximations of the membership functions are based on the tolerance function S_R of the relation R Sen and Pal (2009);

$$\begin{aligned}\underline{M}(u) &= \inf_{\psi \in \mathbb{U}} \max(\overline{S}_R(u, \psi), \mu_X(\psi)) \\ \overline{M}(u) &= \sup_{\psi \in \mathbb{U}} \min(S_R(u, \psi), \mu_X(\psi))\end{aligned} \tag{5}$$

where μ_X, which takes values in the interval [0, 1], is the membership function associated with set X and

$\overline{S}_R(u, \psi) = 1 - S_R(u, \psi)$. When X is a crisp set, μ_X would take values only from the set $\{0, 1\}$. Similarly, when R is a crisp tolerance function, $S_R(u, \psi)$ would take values only from the set $\{0, 1\}$.

3.4 Entropy measures

Using the definitions of upper and lower approximations of a set, entropy can be defined to quantify the ambiguity in the description of X. While the roughness measure $\rho_B(X)$ as in Eq. 3 gives a measure of ambiguity in the description of X, they are needed to be transformed to properly describe the information gain. In this regard, two types of gain functions as defined by Pal and Pal (1991) are considered.

One choice to use a logarithmic function to measure the gain in incompleteness results in a formula of entropy similar to "gain in information" in Shannon's entropy. The logarithmic entropy measure for quantifying the incompleteness of knowledge about $\mathbb{U}$ with respect to the definability of a set $X \in \mathbb{U}$ is given as

$$H_R^L(X) = -\frac{1}{2}\left(\chi(X) + \chi(X^C)\right) \tag{6}$$

where for any set $D \in \mathbb{U}$, $\chi(D) = \rho_R(D)\log_\beta\left(\frac{\rho_R(D)}{\beta}\right)$. Note that the "gain in incompleteness" term is taken as $\log_\beta\left(\frac{\rho_R(D)}{\beta}\right)$ and for $\beta > 1$ it takes values in $[1, \infty]$.

In contrast, the other kind of entropy measure is defined using the exponential function for the "gain in incompleteness". This class of entropy is derived using $\chi(D) = \rho_R(D)\beta^{\bar{\rho}_R(D)}$ where $\bar{\rho}_R(D) = 1 - \rho_R(D)$. So the expression of entropy becomes

$$H_R^E(X) = -\frac{1}{2}\left(\rho_R(X)\beta^{\bar{\rho}_R(X)} + \rho_R(X^C)\beta^{\bar{\rho}_R(X^C)}\right) \tag{7}$$

Here the gain in incompleteness term is taken as $\beta^{(1-\rho_R)}$ which takes values in $[1, \beta]$ when $\beta > 1$. This class of exponential entropy functions possesses various desirable properties which are not present in the usual Shannon's entropy, as illustrated by Pal and Pal (1991). Some of them are mentioned below.

1. In Shannon's entropy, the gain in information $\log(1/p) \to \infty$ as $p \to 0$ and is undefined for $p = 0$. However, in real life, gain in information from an event, whether highly unlikely or highly probable is expected to be finite. However, the exponential gain function ensures that such a gain in information is bounded between 1 and β.
2. Logarithmic entropy is very sensitive to outliers due to the nature of the log function. In contrast, exponential entropy is more robust to outliers. Particularly, a small probability relative to the other probabilities will bias

the entropy towards the small probability event since the weight function $\log(1/p)$ in logarithmic entropy has an unbounded derivative. In comparison, the weight function $\beta^{(1-p)}$ used in exponential entropy has a bounded derivative.

We have considered $\beta = e$ in the subsequent sections, but any value greater than 1 is suitable.

4 Proposed RoufCP algorithm

4.1 Precomputation

In terms of mathematical setup, we consider $\{y_1, y_2, \dots y_T\}$ as the time series data with $y_t \in \mathbb{R}^p$ for any $t = 1, 2, \dots T$. With the number of changepoints equal to 1, we assume that the data comes from a distribution $\mathcal{F}$ at first and then gradually comes from a different distribution $\mathcal{G}$ after some time. Since the location of the changepoint is ambiguous in nature, this creates the possibility of splitting the set of time points $\mathbb{U} = \{1, 2, \dots, T\}$ into two fuzzy partitions, $\gamma_\mathcal{F} = \langle \mathbb{U}, \mu_\mathcal{F} \rangle$ and $\gamma_\mathcal{G} = \langle \mathbb{U}, \mu_\mathcal{G} \rangle$. Here, $\mu_\mathcal{F}$ and $\mu_\mathcal{G}$ denote the membership function of the respective partitions with observations coming from $\mathcal{F}$ and $\mathcal{G}$ respectively. Clearly, an obvious restriction is that $\mu_\mathcal{F}(t) + \mu_\mathcal{G}(t) = 1$. This is related to the incompleteness in knowledge about $\mathbb{U}$ and can be quantified by the entropy measures described by Sen and Pal (2009).

To formalize this notion of fuzziness, for $t \in \mathbb{U}$, for an estimated changepoint s and bandwidth Δ, we define a fuzzy measure similar to that in Sen and Pal (2009).

$$\mu_{s,\Delta}(t) = \begin{cases} 1 & t \leq s - \Delta \\ 1 - 2\left[\frac{t-(s-\Delta)}{\Delta}\right]^2 & s - \Delta < t \leq s \\ 2\left[\frac{(s+\Delta)-t}{\Delta}\right]^2 & s < t \leq s + \Delta \\ 0 & t > s + \Delta \end{cases} \tag{8}$$

On the basis of this estimated changepoint s, the partitions $\gamma_\mathcal{F}$ and $\gamma_\mathcal{G}$ can be reformalized as $\gamma_s = \gamma_\mathcal{F} = \{(t, \mu_{s,\Delta}(t)) | t \in \mathbb{U}\}$ and $\gamma_s^C = \gamma_\mathcal{G} = \{(t, 1 - \mu_{s,\Delta}(t)) | t \in \mathbb{U}\}$, which captures the fuzzy nature of the two partitions due to the estimation of changepoint by a fixed quantity s.

Following the footsteps of grayness ambiguity, as formulated in Sen and Pal (2009), we consider a tolerance function $S_w(u, v)$ such that,

1. $S_w(u, u) = 1$ for any $u \in \mathbb{U} = \{1, 2, \dots T\}$.
2. $S_w(u, v)$ is a decreasing function in $|u - v|$.
3. $S_w(u, v) = 0$ if $|u - v| \geq 2w$, where w is a chosen window length. This means that sufficiently spaced timepoints can be distinguished quite nicely.

Then, the lower and upper approximations of the fuzzy set γ_s can be constructed as $\underline{\gamma_s} = \left\{ (u, M_{\underline{\gamma_s}}) : u \in \mathbb{U} \right\}$ and $\overline{\gamma_s} = \left\{ (u, M_{\overline{\gamma_s}}) : u \in \mathbb{U} \right\}$, where $M_{\underline{\gamma_s}}$ and $M_{\overline{\gamma_s}}$ are obtained using Eq. 5 with the tolerance function S_w and membership function $\mu_{s,\Delta}(t)$. In a similar way, lower and upper approximations of the fuzzy complement set γ_s^C can also be obtained. However, there is a much easier formula that allows one to obtain expressions of upper and lower approximations under set complementation.

Lemma 1

$$M_{\underline{\gamma_s}}(t) = 1 - M_{\overline{\gamma_s^C}}(t)$$
$$M_{\overline{\gamma_s}}(t) = 1 - M_{\underline{\gamma_s^C}}(t)$$

The proof of Lemma 1 is quite simple and can be found in Appendix A.1. Note that, since the computation of the lower and upper approximations $M_{\underline{\gamma_s}}(t)$ and $M_{\overline{\gamma_s}}(t)$ is independent of the data, it can be pre-computed for the changepoint analysis problems, given the knowledge of the number of timepoints T. However, if T is large, computation of Eq. 5 poses a high memory and computational complexity. However, it is possible to obtain exact expressions of these lower and upper approximations under a very general setup, which greatly reduces both the computational and storage cost complexities of the whole process. For instance, according to Eq. 5 $M_{\underline{\gamma_s}}(t)$ is a minimizer of the function $\max \left(\overline{S}_w(t, \psi), \mu_{s,\Delta}(\psi) \right)$ with respect to ψ. This is shown by the upper envelope curve of $\overline{S}_w(t, \psi)$ and $\mu_{s,\Delta}(\psi)$ in Fig. 1. Thus, it is easy to see that the minimizer would appear at a point t^* where the fuzzy membership function and the tolerance function cross each other, i.e., $\overline{S}_w(t, t^*) = \mu_{s,\Delta}(t^*)$. These observations lead us to the following theorem.

Theorem 1 *If the membership function $\mu_{s,\Delta}(t)$ is given as in Eq. 8 and the tolerance function $S_w(u, v)$ is continuous in its both arguments, then lower and upper approximations of the left partition of a chosen changepoint s are expressed by $\underline{\gamma_s} = \left\{ (u, M_{\underline{\gamma_s}}) : u \in \mathbb{U} \right\}$ and $\overline{\gamma_s} = \left\{ (u, M_{\overline{\gamma_s}}) : u \in \mathbb{U} \right\}$ respectively, where*

$$M_{\underline{\gamma_s}}(t) = \begin{cases} 1 & \text{if } t < (s - 2w - \Delta) \\ 1 - \max_{\{t^* : S_w(t,t^*) + \mu_{s,\Delta}(t^*) = 1\}} S_w(t, t^*) & \text{if } (s - 2w - \Delta) \leq t < (s + \Delta) \\ 0 & \text{if } t \geq (s + \Delta) \end{cases}$$

and

$$M_{\overline{\gamma_s}}(t) = \begin{cases} 1 & \text{if } t < (s - \Delta) \\ \max_{\{t^* : S_w(t,t^*) = \mu_{s,\Delta}(t^*)\}} S_w(t, t^*) & \text{if } (s - \Delta) \leq t < (s + 2w + \Delta) \\ 0 & \text{if } t \geq (s + 2w + \Delta) \end{cases}$$

The derivation of the above expressions has been outlined in Appendix A.2. It is possible to obtain a closed form solution of $S_w(t, t^*) = \mu_{s,\Delta}(t^*)$ and similar equations for some specific tolerance functions. One such specific choice is provided in the following corollary.

Corollary 1 Let the membership function $\mu_{s,\Delta}(t)$ be given as in Eq. 8, and a tolerance function be given as

$$S_w(t, t') = \begin{cases} 0 & \text{if } |t - t'| \geq 2w \\ 2\left[\frac{t' - (t - 2w)}{2w} \right]^2 & \text{if } (t - 2w) < x < (t - w) \\ 1 - 2\left[\frac{x - t}{2w} \right]^2 & \text{if } |t - t'| \leq w \\ 2\left[\frac{(t + 2w) - t'}{2w} \right]^2 & \text{if } (t + w) < x < (t + 2w). \end{cases} \quad (9)$$

Fig. 1 Lower Approximation Curves
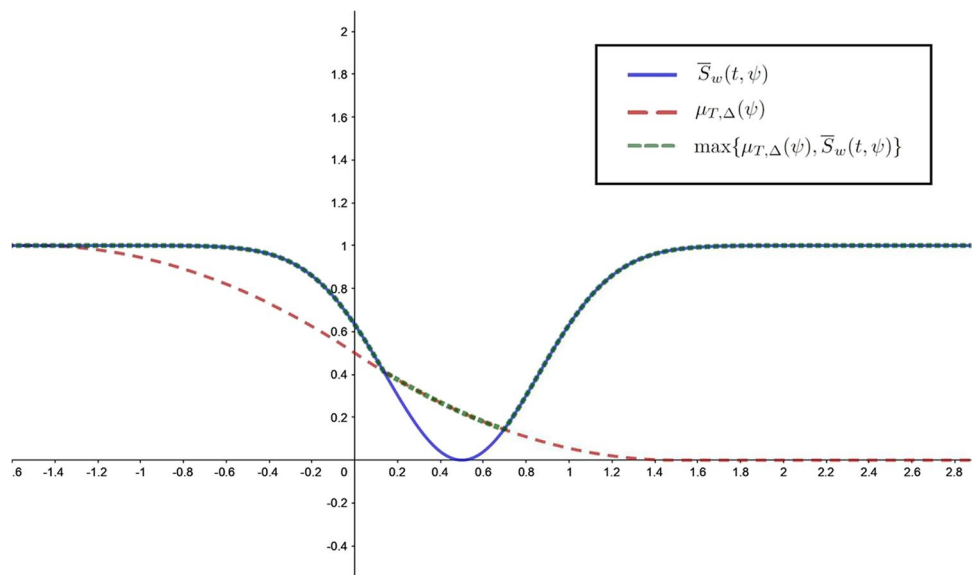
Then the lower and upper approximations of the left and right partitions for a chosen changepoint s can be obtained as $\gamma_s = \left\{ (u, M_{\underline{\gamma_s}}) : u \in \mathbb{U} \right\}$, $\overline{\gamma_s} = \left\{ (u, M_{\overline{\gamma_s}}) : u \in \mathbb{U} \right\}$, $\gamma_s^C = \left\{ (u, M_{\underline{\gamma_s^C}}) : u \in \mathbb{U} \right\}$ and $\overline{\gamma_s^C} = \left\{ (u, M_{\overline{\gamma_s^C}}) : u \in \mathbb{U} \right\}$, where

$$M_{\underline{\gamma_s}}(t) = \begin{cases} 0 & \text{if } t \geq (s + \Delta) \\ 2\left[\frac{(s+\Delta)-t}{2(w+\Delta)} \right]^2 & \text{if } (s-w) \leq t < (s+\Delta) \\ 1 - 2\left[\frac{(t+2w)-(s-\Delta)}{2(w+\Delta)} \right]^2 & \text{if } (s-2w-\Delta) \leq t < (s-w) \\ 1 & \text{if } t < (s-2w-\Delta) \end{cases}$$

and

$$M_{\overline{\gamma_s}}(t) = \begin{cases} 0 & \text{if } t \geq (s + 2w + \Delta) \\ 2\left[\frac{(s+\Delta)-(t-2w)}{2(w+\Delta)} \right]^2 & \text{if } (s+w) \leq t < (s+2w+\Delta) \\ 1 - 2\left[\frac{t-(s-\Delta)}{2(w+\Delta)} \right]^2 & \text{if } (s-\Delta) \leq t < (s+w) \\ 1 & \text{if } t < (s-\Delta), \end{cases}$$

while the approximations for the complementary set can be obtained using Lemma 1.

Another very general result for symmetric tolerance functions is expressed in Theorem 2. It uses the symmetry of the tolerance function to reduce the computation of obtaining the upper (or lower) approximations $M_{\overline{\gamma_s}}$ (or $M_{\underline{\gamma_s}}$) by half. The proof of the theorem has been deferred till Appendix A.3.

Theorem 2 *If the tolerance function $S_w(u,v)$ can be expressed as a function of the absolute difference of its arguments, i.e. $S_w(u,v) = g(|u-v|)$ such that $g(\cdot)$ is symmetric about 0, then;*

$$M_{\underline{\gamma_s}}(t) = 1 - M_{\overline{\gamma_s}}(2s-t) \ \forall t \ : \ \max\{1, 2s-T\} < t < \max\{2s, T\}$$

4.2 The final algorithm

The proposed algorithm Rough-Fuzzy CPD (RoufCP) first calculates a regularity measure based on the input, whose working is similar to that of the usual criterion $V(\mathcal{T}, y)$ in Eq. 2. In other words, for each timepoint t, a two-sample test statistic $R(t)$ is computed to detect the changes in the samples $\{y_{(t-\delta+1)}, y_{(t-\delta+2)}, \ldots y_t\}$ and $\{y_{(t+1)}, y_{(t+2)}, \ldots y_{(t+\delta)}\}$. We expect the regularity measure $R(t)$ to take a higher value when there is no change in distribution and take a lower value when there is a detected change in distribution, hence it could be taken as some suitable transformation of the usual two-sample test statistics used in statistics. For example, one such regularity measure to detect the changes in mean could be based on Hotelling's T^2 test statistic.

$$R(t) = \frac{1}{1 + (\bar{y}_1 - \bar{y}_2)^{\intercal} \Sigma^{-1} (\bar{y}_1 - \bar{y}_2)},$$

where

$$\bar{y}_1 = \delta^{-1} \sum_{t'=(t-\delta+1)}^{t} y_{t'} \quad \bar{y}_2 = \delta^{-1} \sum_{t'=(t+1)}^{(t+\delta)} y_{t'}$$

$$\Sigma = \delta^{-1} \sum_{t'=(t-\delta+1)}^{(t+\delta)} \left(y_{t'} - \frac{1}{2}(\bar{y}_1 + \bar{y}_2) \right) \left(y_{t'} - \frac{1}{2}(\bar{y}_1 + \bar{y}_2) \right)^{\intercal}$$

Combining the pre-computed upper and lower approximations of the two fuzzy partitions γ_s and γ_s^C by the proposed changepoint s, with the regularity measure $R(t)$ obtained from the data, one can create roughness measures for these two partitions as follows:

$$\begin{aligned} \rho_{\Delta,\delta,w}(\gamma_s) &= 1 - \frac{\sum_{t=1}^{T} M_{\underline{\gamma_s}}(t) R(t)}{\sum_{t=1}^{T} M_{\overline{\gamma_s}}(t) R(t)} \\ \rho_{\Delta,\delta,w}(\gamma_s^C) &= 1 - \frac{\sum_{t=1}^{T} M_{\underline{\gamma_s^c}}(t) R(t)}{\sum_{t=1}^{T} M_{\overline{\gamma_s^c}}(t) R(t)} \end{aligned} \quad (10)$$

These roughness measures can be efficiently computed due to the results mentioned in Section 4.1. In particular, to obtain the roughness measures $\rho_{\Delta,\delta,w}(\gamma_s)$ and $\rho_{\Delta,\delta,w}(\gamma_s^C)$ as given in Eq. 10, it is enough to focus the computation on only one of these terms, since the other can be obtained as a byproduct of Lemma 1. Also, Theorem 2 tells us that in the case of symmetric and location invariant tolerance function, only one, either lower or upper, approximation curve is required to be computed, thus reducing the computational complexity to one-fourth of the naive implementation. Finally, Theorem 1 and Corollary 1 can be used to obtain closed-form expressions for the upper (or lower) approximation for the particular choice of tolerance function given in Eq. 9.

Based on these roughness measures, entropy quantifying the ambiguity for the fuzzy partitions of the period for the specifically chosen changepoint s can be expressed using Eqs. 6 or 7, with $\rho_R(X)$ replaced by the roughness measure given in Eq. 10. Thus, we obtain

$$H^E_{\Delta,\delta,w}(s) = \rho_{\Delta,\delta,w}(\gamma_s) e^{(1-\rho_{\Delta,\delta,w}(\gamma_s))} + \rho_{\Delta,\delta,w}(\gamma_s^C) e^{(1-\rho_{\Delta,\delta,w}(\gamma_s^C))} \quad (11)$$

Our proposed method minimizes this entropy (or uncertainty) $H^E_{\Delta,\delta,w}(s)$ to obtain the changepoints.

$$t^* = \min_{t \in \{1,2,\ldots T\}} H^E_{\Delta,\delta,w}(t). \quad (12)$$

This estimated changepoint t^* shown in Eq. 12 will be denoted as the rough-fuzzy CP, and the method will be called Rough-Fuzzy CPD. Note that while the global

Algorithm 1 Rough-Fuzzy
CPD (RoufCP)

Require: A multivariate time series Y_t for $t = 1, 2, \ldots T$ with $Y_t \in \mathbb{R}^p$.
Require: Hyperparameters $\delta, \Delta,$ w and an acceptance threshold α (usually 0.05).

1: **Output:** Estimated changepoints $s_1, s_2, \ldots s_k$.
2: **for** $t = 1$ to T **do**
3:   Create two samples $\mathcal{S}_{1t} = \{Y_{\max\{1, t-\delta+1\}}, \ldots, Y_t\}$ and $\mathcal{S}_{2t} = \{Y_{t+1}, \ldots, Y_{\min\{T, t+\delta\}}\}$
4:   Compute regularity measure $R(t)$ as a similarity between $\mathcal{S}_{1t}$ and $\mathcal{S}_{2t}$
5: **end for**
6: **for** $s = 1$ to T **do**
7:   **for** $t = 1$ to T **do**
8:     Use theorem 1 to compute $M_{\gamma_s}(t)$
9:     Use lemma 2 to compute $M_{\overline{\gamma_s}}(t)$
10:     Use lemma 1 to compute $M_{\underline{\gamma_s^C}}(t)$ and $M_{\overline{\gamma_s^C}}(t)$
11:   **end for**
12:   Compute roughness measure $\rho_{\Delta, \delta, w}(\gamma_s)$ and $\rho_{\Delta, \delta, w}(\gamma_s^C)$ using Eqn (10)
13:   Compute exponential entropy $H_{\Delta, \delta, w}^E(s)$ using Eqn (11)
14: **end for**
15: Find local minima of the sequence $\{H_{\Delta, \delta, w}^E(s) : s = 1, 2, \ldots T\}$ ▷
One can output these local minima as proposed changepoints, however if theoretical expressions for μ and σ as denoted in theorem 3 for the underlying regularity measure $R(t)$ is available, the following testing framework can be used to further enhance its ability.
16: **for** each local minima s_k of $H_{\Delta, \delta, w}^E$ **do**
17:   Compute $H^*(s_k)$ as given in theorem 3 by replacing the infinite sums as sums from $t = 1$ to T
18:   **for** $n = 1$ to T **do**
19:     Use Eqn (14) to compute $A_{s_k}(n)$ by replacing the infinite sums as sums from $t = 1$ to T
20:   **end for**
21:   Compute σ^* for the changepoint s_k as given in theorem 3 using finite sums from $t = 1$ to T
22:   Accept s_k as a changepoint if $a_\delta \left(H_{\Delta, \delta, w}^E(s_k) - H^*(s_k) \right) / \sqrt{\sigma^*} < z_\alpha$, where z_α is the α-th quantile of the standard normal distribution
23: **end for**

---

minimum serves as the estimate of a single changepoint, the local minima of the entropy function, after suitable testing, can be used for multiple changepoint detections. A detailed analysis of how these detected changepoints can be tested using a statistical hypothesis to reduce any false positive detection is described in Section 4.3. The complete algorithm is shown in Algorithm 1.

In contrast to minimizing the entropy, the usual changepoint detection method using the regularity measure $R(t)$ Truong et al. (2020) can be broadly expressed as

$$t^* = \min_{t \in \{1, 2, \ldots T\}} R(t) = \min_{t \in \{1, 2, \ldots T\}} V(\{1, 2, \ldots t\}, \{(t+1), (t+2), \ldots T\}), y). \tag{13}$$

where t^* is the estimated changepoint and $V(\cdot, \cdot)$ is a cost function as shown in Eq. 2, which can be interpreted as a regularity measure. While in this way, the regularity measure

$R(t)$ itself becomes an indicator of the changepoint, it can be greatly enhanced with the help of fuzzy and rough set theory, by constructing the entropy as given in Eq. 11 and then focusing on its minimizer. Hence, the inclusion of fuzzy-rough set theory simply serves the purpose of variance reduction instead of a completely new algorithm. We shall show later that this simple tweak is extremely beneficial if the changes in the time series are smooth. Intuitively, since the underlying distribution of the time series data is unknown, all the information about the changepoint must be gathered in terms of the available observations y_t, which is now summarized only through a single attribute $R(t)$, the regularity measure. Along with this loss of information, as the regularity measure $R(t)$ is computed based on overlapping windows, any information about the locality

of a timepoint will permeate to its neighboring timepoints as well, resulting in incomplete information about the timepoints themselves. This rough resemblance between different time points is modeled by the tolerance relation which eventually leads to the rough set formulation.

4.3 Asymptotic analysis

While the Rough-Fuzzy CPD can be employed and a single changepoint can be detected using Eq. 12, multiple changepoints can be detected by local minima of the curve $H^E_{\Delta,\delta,w}(s)$ as a function of s. However, a reference curve must be computed to select the true changepoints from many local minima. In statistical language, these reference curve is usually computed based on the distribution of the statistic under a suitably chosen null hypothesis, by modifying the problem into a hypothesis testing framework.

Considering the mathematical framework given in Eq. 1, we can formulate the problem of detecting changepoint as a hypothesis testing problem.

$$H_0: \quad F_0 = F_1 = \cdots = F_k = F$$
$$H_1: \quad \text{There is at least one inequality}$$

Since the regularity measure $R(t)$ is an indicator of a possible changepoint, hence under the null hypothesis H_0; $\mathbb{E}(R(t))$ is a constant independent of the time t. Based on this, we obtain the asymptotic null distribution of the proposed statistic under some reasonable assumptions on the asymptotic null distribution of the regularity measure. Clearly, the asymptotic follows when the number of samples for constructing $R(t)$ i.e. δ is tended to infinity, which forces the total number of timepoints T to tend to infinity as well. Thus, in order to talk about asymptotics, we restrict ourselves to an infinite dimensional normed space, which without any loss of generality, can be taken as $l^\infty(\mathbb{N})$, the set of all uniformly bounded functions from $\mathbb{N}$ to $\mathbb{R}$. Also, while the number of timepoints T increases to infinity, the intervals between successive observations tend to zero, in order to make the total period of observation constant pertaining to most practical situations.

Let us denote the regularity measure R_{δ_n} as a $l^\infty(\mathbb{N})$ valued random element defined on some probability measure space $(\Omega, \mathcal{A}, \mathbb{P})$ such that $R_{\delta_n}(t)$ is a $\mathbb{R}$-valued random variable denoting the regularity measure based on the $2\delta_n$ subsamples centered at t for each $t = 1, 2, \ldots \infty$.

Theorem 3 *Assume that under the null hypothesis there is no changepoint, the regularity measure R_{δ_n} has an asymptotic distribution such that for some sequence $a_{\delta_n} \to \infty$,*

$$a_{\delta_n}(R_{\delta_n} - \mu\mathbf{1}(\cdot)) \rightsquigarrow Z$$

where $\mathbf{1} \in l^\infty(\mathbb{N})$ is the identity function $\mathbf{1}(x) = x$, Z is an infinite dimensional Gaussian process with mean function identically equal to 0 and covariance function $\sigma : \mathbb{N} \times \mathbb{N} \to [0, \infty)$. Also, assume the following regularity conditions:

1. $\mu \neq 0$.
2. *For the proposed changepoint s, the series $\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)$ and $\sum_{t=1}^\infty M_{\overline{\gamma_s^C}}(t)$ are convergent.*

Define,

$$b(s) = \frac{\sum_{t=1}^\infty M_{\gamma_s}(t)}{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)}$$

$$\overline{b}(s) = \frac{\sum_{t=1}^\infty M_{\gamma_s^C}(t)}{\sum_{t=1}^\infty M_{\overline{\gamma_s^C}}(t)}$$

Then, the exponential entropy based statistic $H^E_{\Delta,\delta_n,w}(s)$ defined in Eq. 11 has an asymptotic distribution with the same normalizing constant, a_{δ_n} such that as $a_{\delta_n} \to \infty$,

$$a_{\delta_n}(H^E_{\Delta,\delta_n,w}(s) - H^*(s)) \rightsquigarrow Z^*$$

where

$$H^*(s) = (1 - b(s))e^{b(s)} + (1 - \overline{b}(s))e^{\overline{b}(s)}$$

and Z^* is a univariate normally distributed random variable with mean 0 and variance σ^* where;

$$\sigma^* = \sum_{m=1}^\infty \sum_{n=1}^\infty A_s(m)\sigma(m,n)A_s(n)$$

and;

$$A_s(n) = \left[b(s)e^{b(s)} \left\{ \frac{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\gamma_s}(t) - \sum_{t=1}^\infty M_{\gamma_s}(n)M_{\overline{\gamma_s}}(t)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \right\} + \overline{b}(s)e^{\overline{b}(s)} \left\{ \frac{\sum_{t=1}^\infty M_{\overline{\gamma_s^C}}(n)M_{\gamma_s^C}(t) - \sum_{t=1}^\infty M_{\gamma_s^C}(n)M_{\overline{\gamma_s^C}}(t)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s^C}}(t)\right)^2} \right\} \right] \quad (14)$$

provided that the series expression of σ^* is convergent.

For the detailed proof of the above theorem please refer to Appendix A.4. In the above Theorem 3, even if Z is not an infinite-dimensional Gaussian process but follows some other process, Z^* can be modified accordingly to accommodate such changes. In view to the functional delta method described in Römisch (2014), Z^* has the same distribution as $\Psi'_s|_{\mu\mathbf{1}(\cdot)}(Z)$. An immediate extension of Theorem 3 is the analogous result for multiple proposed

changepoints, which can also be accordingly modified for the case when Z does not follow a Gaussian process.

Corollary 2 Under the same assumptions and conditions of Theorem 3, the vector of exponential entropy based statistic $H^E_{\Delta,\delta_n,w}(s)$ for multiple proposed changepoints $s_1, s_2, \ldots s_k$ has the following asymptotic distribution as $a_{\delta_n} \to \infty$;

$$a_{\delta_n}\begin{bmatrix} H^E_{\Delta,\delta_n,w}(s_1) - H^*(s_1) \\ H^E_{\Delta,\delta_n,w}(s_2) - H^*(s_2) \\ \ldots \\ H^E_{\Delta,\delta_n,w}(s_k) - H^*(s_k) \end{bmatrix} \rightsquigarrow \mathcal{N}_k(\mathbf{0}_k, \Sigma^*_{k\times k})$$

where $\mathbf{0}_k$ is the k dimensional null vector, and the entries of the $k \times k$ dispersion matrix are given as;

$$\left(\Sigma^*_{k\times k}\right)_{(i,j)} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} A_{s_i}(m)\sigma(m,n)A_{s_j}(n) \qquad i,j = 1, 2, \ldots k$$

In particular, if $|s_i - s_j| > (4w + 2\Delta)$, then $\left(\Sigma^*_{k\times k}\right)_{(i,j)} = 0$.

The conditions mentioned in Theorem 3 hold for a multitude of regularity measures. In particular, for almost any usual two-sample test statistic, a suitable transformation can be performed to create a regularity measure satisfying the conditions. To illustrate the applicability of Theorem 3, we consider a simple example here. Take the two sample-based U-statistic

$$u_{\delta_n}(t) = \frac{1}{\delta_n^2}\sum_{i=1}^{\delta_n}\sum_{j=1}^{\delta_n} h(y_{t-i}, y_{t+j}) \tag{15}$$

where $h(y_{t-i}, y_{t+j}) = (y_{t-i} - y_{t+j})^2$. We assume that $y_1, y_2, \ldots y_n$ follow a normal distribution $N(\mu, \sigma^2)$. Under the null hypothesis that there is no changepoint, $y_{t-i} - y_{t+j}$ follows a normal distribution $N(0, 2\sigma^2)$. An application of two sample U-statistics theorem Serfling (2009) can be employed to show that

$$\sqrt{2\delta_n}(u_{\delta_n}(t) - 2\sigma^2) \rightsquigarrow N(0, 8\sigma^4)$$

Generalizing this for all timepoints t yields

$$\sqrt{2\delta_n}(u_{\delta_n} - 2\sigma^2\mathbf{1}(\cdot)) \rightsquigarrow Z$$

where Z is an infinite dimensional Gaussian process with mean function identically equal to 0 and covariance function $\sigma(t, t+s) = 4\max\{(2-\lambda), 0\}\sigma^4$ for any s and t satisfying $\lim_{\delta_n \to \infty} s/\delta_n = \lambda$. Also, one can easily verify that

$$\frac{1}{\delta_n^2}\sum_{i=1}^{\delta_n}\sum_{j=1}^{\delta_n}(y_{t-i} - y_{t+j})^2 = s^2_{t,-} + s^2_{t,+} + (\bar{y}_{t,-} - \bar{y}_{t,+})^2$$

where $\bar{y}_{t,-}, s^2_{t,-}$ are respectively the sample mean and sample variance of $\{y_{t-\delta_n}, y_{t-\delta_n+1}, \ldots y_{t-1}\}$ and $\bar{y}_{t,+}, s^2_{t,+}$ are respectively the sample mean and sample variance of $\{y_{t+1}, y_{t+2}, \ldots y_{t+\delta_n}\}$. This shows that the U-statistic considered in Eq. 15 is a suitable transformation of the mean difference statistic used for changepoint detection.

5 Simulation studies

In this section, we analyze the numerical performance of our proposed changepoint detection method. The section is organized as follows. First, we quantify how much Rough-Fuzzy CPD improves upon the usual changepoint detection methods (hereon, refers to as the base method) based on Eq. 13, using the same regularity measures. Then, we check the performance of our method under different parameters of the underlying true changepoint. Specifically, we address the effect of the fuzziness of the true change point and signal-to-noise ratio of the data on the performance of Rough-Fuzzy CPD as compared to the base methods. Following this, we provide an extensive comparison between the performances of the proposed method with various existing CPD methods. A variety of methods including standard methods for abrupt changepoint detection, multiscale methods, and different fuzzy clustering-based techniques have been chosen for this purpose. Finally, we perform a sensitivity analysis to study the dependence of the performance of our proposed method with varying choices of hyperparameters. All the following simulations and real data applications have been performed by the python package roufcp developed by us.

5.1 Improvement upon base method

To demonstrate how incorporating fuzzy rough set theory into the usual CPD methods improves its performance, we consider two variants of the same CPD methods. In the first variant, Eq. 13 is used to obtain the estimate of the changepoint based on the regularity measure $R(t)$, whereas, in the second variant, the regularity measure is combined with roughness measures to estimate the changepoint using Eq. 12. To compare the performances, we consider three simulation setups of the form

$$y_t = \mu(t) + \epsilon_t$$

where $\mu(t)$ is the mean function dependent on time, while ϵ_t denotes a standard Gaussian white noise process. Three different choices of the mean curve $\mu(t)$ are assumed. As shown in Fig. 2a, scenario (S1) considers an abrupt changepoint in the mean function which corresponds to the usual definition of a changepoint. On the other hand, scenarios (S2) and (S3)

consider gradual changes in the mean function, denoting a piecewise linear continuous jump and smooth jump respectively. Figures 2b and c show the mean functions and one simulated data for these scenarios.

To establish the comparison, we consider three test statistics that are well used in a two-sample setup to detect the changes in the mean. The parametric statistics as the square of the two sample t-test statistic, nonparametric Kolmogorov-Smirnov test statistic, and Augmented Dickey-Fuller test statistic among unit root tests were chosen. In each case, the regularity measure $R(t)$ is obtained by taking the reciprocal of the test statistics, to make sure $R(t)$ has a higher value in absence of changepoints and a lower value in presence of changepoint. Both Eqs. 13 and 12 were used to obtain the estimates of the changepoints as mentioned before, and the process was repeated for 200 Monte Carlo resamples and used to obtain error measures. In applying Rough-Fuzzy CPD, $\delta = 50$ was chosen to compute the regularity measure, while different values of w and Δ were used in the experiment to see their effects.

Table 1 summarizes the result for all three simulation setups described above. We begin analyzing the results with the changepoint of type ($S1$), where a discrete jump in the mean function has occurred. Note that, the hyperparameters w and Δ control the amount of fuzziness and roughness to

incorporate when detecting changepoints. Clearly, $w = \Delta = 0$ would entail just a transformation of $R(t)$ as the exponential entropy, and both the estimating Eqs. 13 and 12 would yield the same changepoint. Thus, in general, increasing w and Δ would increase RMSE (root mean square error) if the true model has a discrete jump change in the mean function. Indeed, for $w = \Delta = 5$, for the KS test, we have a 74% decrease in MSE while with higher values of w and Δ, we see higher RMSE for Rough-Fuzzy CPD compared to the base methods using t-test and KS-test. For ADF test statistic, however, the proposed method outperforms the base method across all hyperparameter values, achieving more than 70% reduction in MSE. Hence, we see it is important to use appropriate values of hyperparameters. Overall, for discrete jump type changepoints, Rough-Fuzzy CPD performs poorly, giving higher MSE than the base model for t-test and Kolmogorov Smirnov tests, and higher values of w and Δ will decrease the estimation accuracy. This is because the fundamental assumption of the proposed method viz. "change is fuzzy in nature and not abrupt" is violated in this case. Now, we look at the other 2 types of changepoints where the change is not abrupt and occurs gradually over some time.

From Table 1 we observe the performance of Rough-Fuzzy CPD for situation ($S2$) with continuous change in mean function. For both the t-test statistic and the
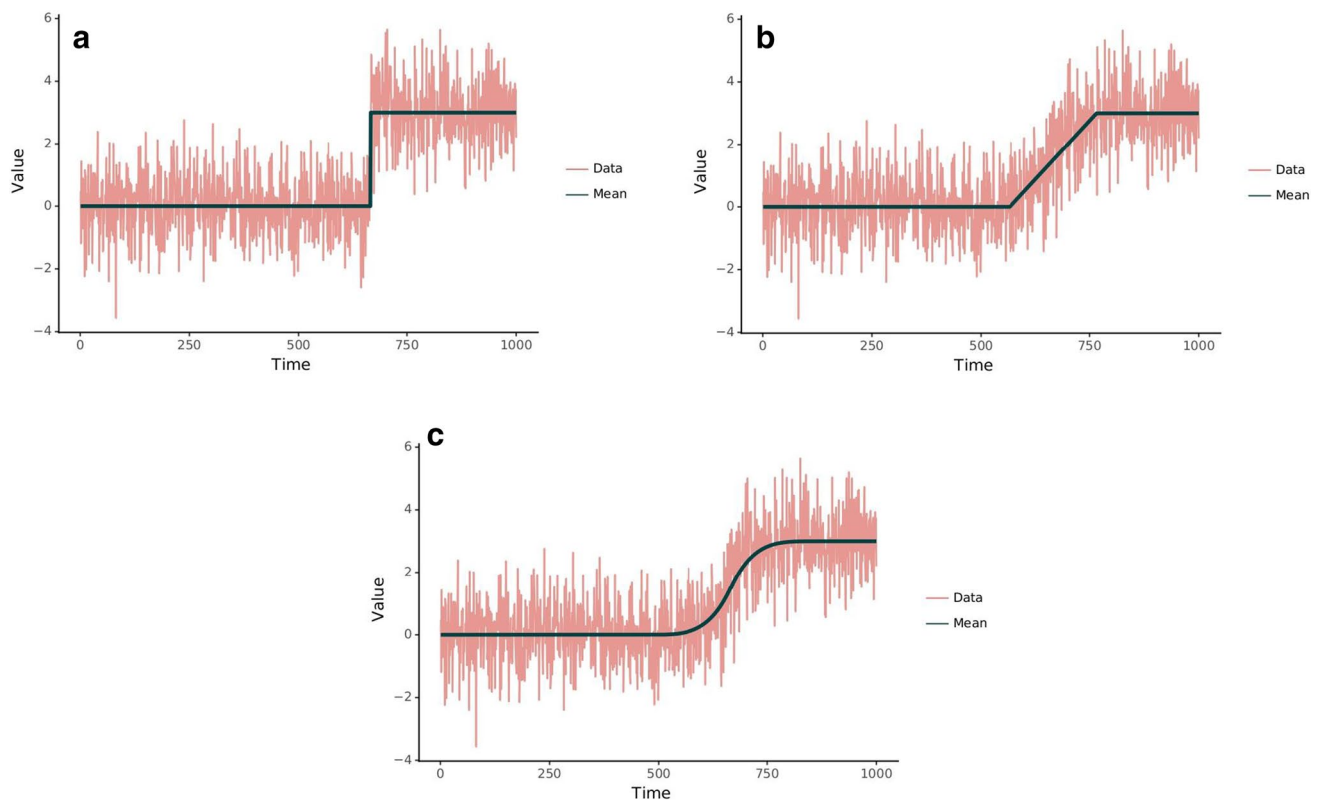

Fig. 2 Different Scenarios of changepoints with Changes in mean of the data

Table 1 Comparison of Rough-Fuzzy CPD (proposed) vs usual (base) methods based on cost minimization for 3 types of mean shift changepoints

Simulation setup	Base test statistic	Parameters for proposed method	RMSE		Relative Decrease in MSE
			Proposed	Base	
Discrete Jump in Mean (S1)	t-test	$w = \Delta = 5$	1.591	1.752	17.53%
		$w = \Delta = 25$	4.961		-701.81%
		$w = \Delta = 50$	6.567		-1436.26%
		$w = \Delta = 100$	21.446		-14883.88%
	KS-test	$w = \Delta = 5$	2.001	3.928	74.08%
		$w = \Delta = 25$	4.926		-57.27%
		$w = \Delta = 50$	7.404		-255.29%
		$w = \Delta = 100$	25.533		-4125.33%
	ADF-test	$w = \Delta = 5$	38.115	106.684	87.23%
		$w = \Delta = 25$	41.395		84.94%
		$w = \Delta = 50$	43.843		83.11%
		$w = \Delta = 100$	55.305		73.12%
Continuous Jump in Mean (S2)	t-test	$w = \Delta = 25$	25.232	35.501	49.48%
		$w = \Delta = 50$	11.326		89.82%
		$w = \Delta = 100$	17.419		75.92%
	KS-test	$w = \Delta = 25$	26.881	79.723	88.63%
		$w = \Delta = 50$	15.071		96.42%
		$w = \Delta = 100$	23.778		91.10%
	ADF-test	$w = \Delta = 25$	207.811	347.614	64.26%
		$w = \Delta = 50$	197.768		67.63%
		$w = \Delta = 100$	189.827		70.18%
Smooth Jump in Mean (S2)	t-test	$w = \Delta = 25$	12.242	18.001	53.75%
		$w = \Delta = 50$	9.596		71.58%
		$w = \Delta = 100$	18.768		-8.70%
	KS-test	$w = \Delta = 25$	11.640	37.714	90.47%
		$w = \Delta = 50$	10.127		92.78%
		$w = \Delta = 100$	22.936		63.01%
	ADF-test	$w = \Delta = 25$	197.282	286.67	52.64%
		$w = \Delta = 50$	187.246		57.33%
		$w = \Delta = 100$	181.6		59.87%

Kolmogorov-Smirnov statistic, the RMSE is much lower in Rough-Fuzzy CPD relative to the base method based on the regularity measure $R(t)$. However, with an increase in w and Δ, the RMSE reduces first and then increases, possibly suggesting the existence of optimal hyperparameters in between. Because of the specific parametric setup, the Augmented Dickey-Fuller (ADF) test and nonparametric Kolmogorov Smirnov test generally perform worse than the parametric t-test. However, in practical applications when the data generating processes are not known, ADF and KS test statistics might perform better in conjunction with our proposed improvement. An interesting phenomenon occurs when we use the ADF test statistic-based regularity measure. As shown in Figure 3, the ADF test usually outputs two changepoints around the true changepoint at 666; but as the fuzziness is incorporated into the model by increasing w and Δ, the bimodal distribution gradually becomes unimodal.

However, there possibly remains a negative bias in the estimation, as indicated by the mode of the density curve of the Rough-Fuzzy CPD for $w = \Delta = 50$.

Turning to the situation ($S3$) with a smooth change in mean function, the results are found to be similar to those in scenario $S2$. Here also Rough-Fuzzy CPD has reduced the MSE by more than 50% in most cases. However, we see that for $w = \Delta = 100$ the efficacy of the model is greatly reduced and it even performs slightly worse than the t-test-based method. Again, as mentioned earlier, it is of utmost importance to choose the values of the hyperparameters w and Δ carefully to obtain accurate estimates of the change-point even in situations where the mean function is gradual and the underlying assumption of the method is not violated. However, as shown by Table 1, Rough-Fuzzy CPD generally obtains a higher reduction in MSE for $S2$ (continuous jump) and $S3$ (smooth jump) compared to $S1$ (discrete jump). So,

Fig. 3 Base method along with proposed Rough-Fuzzy CPD based on the regularity measure with Augmented Dickey Fuller statistic for continuous jump changepoint (*S2*)
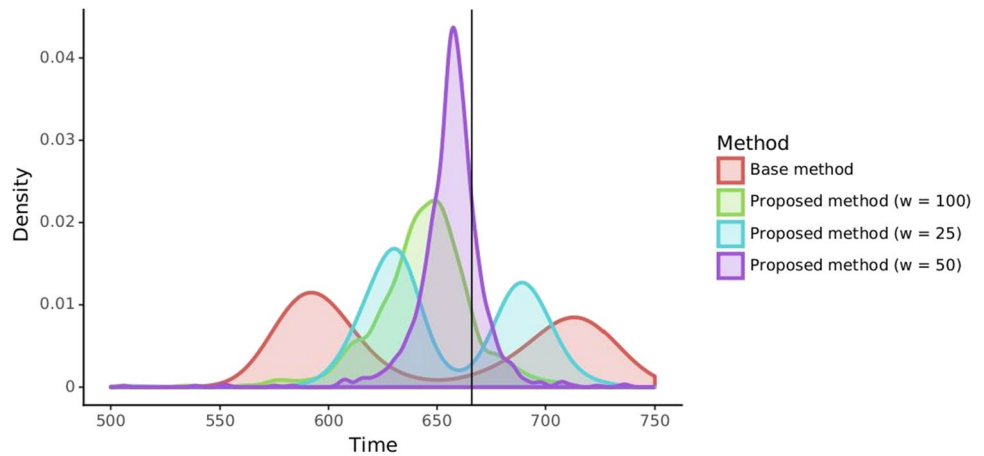
with a greater degree of ambiguity in the changepoint, the relative performance of our model increases as expected. We shall illustrate this phenomenon in greater detail and more rigorously in Section 5.2.

5.2 Effect of fuzziness of the true changepoint and SNR of the data

In our earlier simulations, scenarios (*S2*) and (*S3*) depict such situations where a gradual change has occurred. To check how Rough-Fuzzy CPD performs under different levels of graduality or "fuzziness" in change, we consider different cases of scenario *S2* where the mean function is defined as follows;

$$\mu_{s,\Delta}(t) = \begin{cases} 0 & t \leq s_0 - \mathcal{F} \\ \frac{t-(s_0-\mathcal{F})}{\mathcal{F}} & s_0 - \mathcal{F} < t \leq s_0 + \mathcal{F} \\ 2 & t > s_0 + \mathcal{F} \end{cases}$$

where s_0 is the true changepoint and the parameter $\mathcal{F}$ defines the degree of graduality or "fuzziness" of the changepoint. To see the effect of $\mathcal{F}$ on the estimate of Rough-Fuzzy CPD, we consider 1000 resamples of scenario *S2* with true changepoint at $s_0 = 666$, for 15 different values of $\mathcal{F}$ ranging from 10 to 150. The MSE for each level of fuzziness $\mathcal{F}$ is calculated using a Monte Carlo method, for the estimated changepoint obtained by Rough-Fuzzy CPD as well as the underlying base method with the Kolmogorov-Smirnov statistic.

Figure 4 shows the RMSE of estimated changepoints by Rough-Fuzzy CPD and its base counterpart for different values of fuzziness $\mathcal{F}$, along with the relative decrease obtained by Rough-Fuzzy CPD. As shown in Fig. 4a, a higher degree of fuzziness leads to higher error in estimation for both methods, though Rough-Fuzzy CPD is not as severely affected as the base method. Figure 4b depicts that the relative performance gain by the proposed rough-fuzzy

improvement is only possible when the fuzziness in the true changepoint crosses a certain threshold, about $\mathcal{F} = 31$ for our specific setup. However, as the fuzziness increases, due to the extremely sensitive performance of the base method, Rough-Fuzzy CPD could achieve nearly $95 - 99\%$ performance gain in terms of MSE.

While detecting fuzzy changepoints is a hard problem, doing so in noisy data with a low Signal-to-Noise Ratio (SNR) is even harder. Signal-to-Noise Ration (SNR) is defined as the ratio of $\mathbb{E}(S^2)$ and $\mathbb{E}(N^2)$ where N is the noise component of the data and S is the true signal component of the data, which mainly comprises the mean function in the time series observations. To check the performance of Rough-Fuzzy CPD under different SNRs, we keep noise N the same as the standard Gaussian white noise and vary the value of signal S by changing the size of the jump in the mean function. To illustrate its effect, we consider scenario (*S2*) with continuous linear change in mean function, with different jump sizes ranging $1/5, 1/4, \ldots 1/2, 1, 2, \ldots 10$. For each such setup, we use the Kolmogorov-Smirnov test statistic to estimate the changepoints using the proposed method and the base method for 1000 resamples, and Monte Carlo estimate of MSE is obtained. In each resample, the true changepoint value is kept fixed at 666. Figure 5 shows the variation of RMSE of the estimated changepoint for the proposed Rough-Fuzzy CPD as well as the base method. We observe that Rough-Fuzzy CPD has obtained lower RMSE than the base method for all values of SNR. While SNR increases, the predictive capabilities of both methods increase. However, it is important to note that while the base method performs very badly for lower values of SNR (< 1), the performance of Rough-Fuzzy CPD using the same regularity measure is not so severely affected. In fact, throughout the low and high values of SNR, the relative improvement of MSE achieved by Rough-Fuzzy CPD remains fairly uniform ranging from $88 - 97\%$, with the highest reduction achieved when SNR is 2 or 3.

Fig. 4 Performance of proposed Rough-Fuzzy CPD and base method (KS-test)for different values of fuzziness of true changepoint
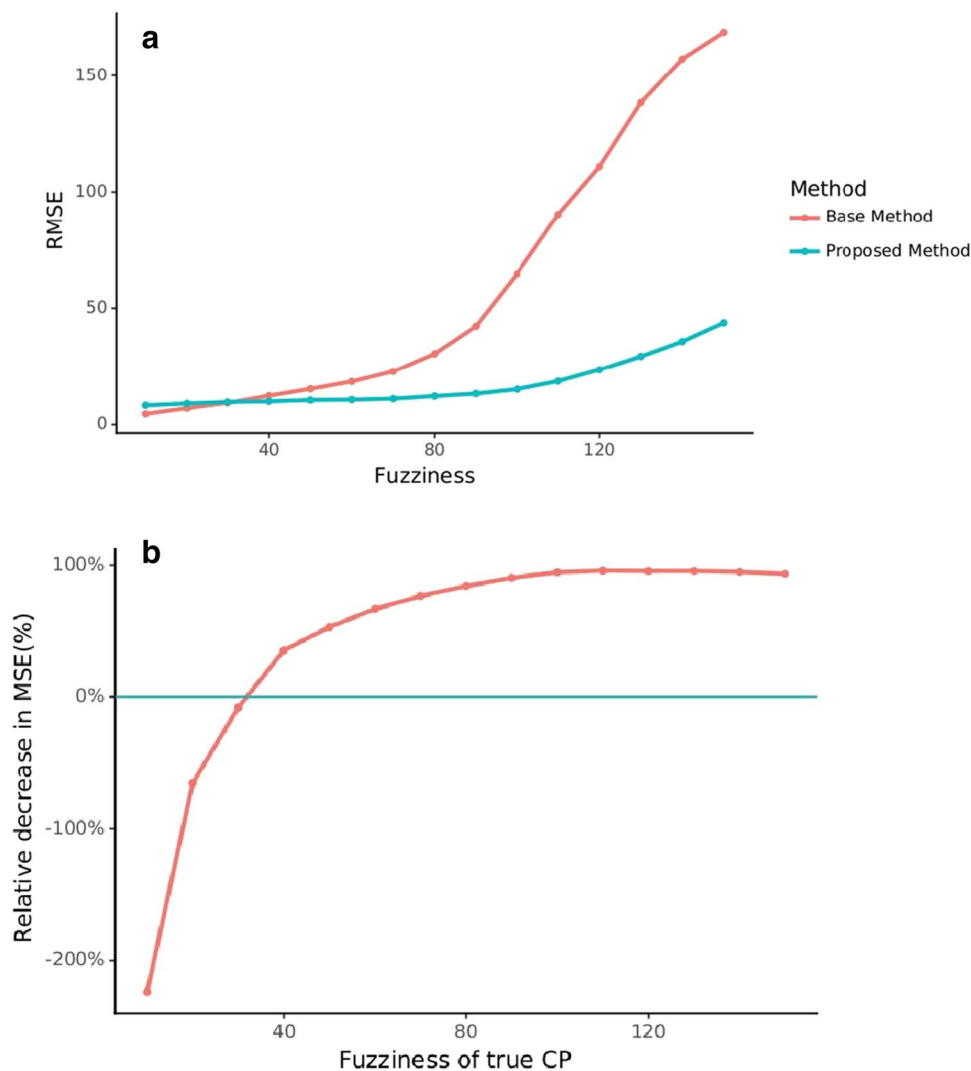

Fig. 5 Performance of proposed Rough-Fuzzy CPD and base method (KS-test)for different values of signal-to-noise ratio
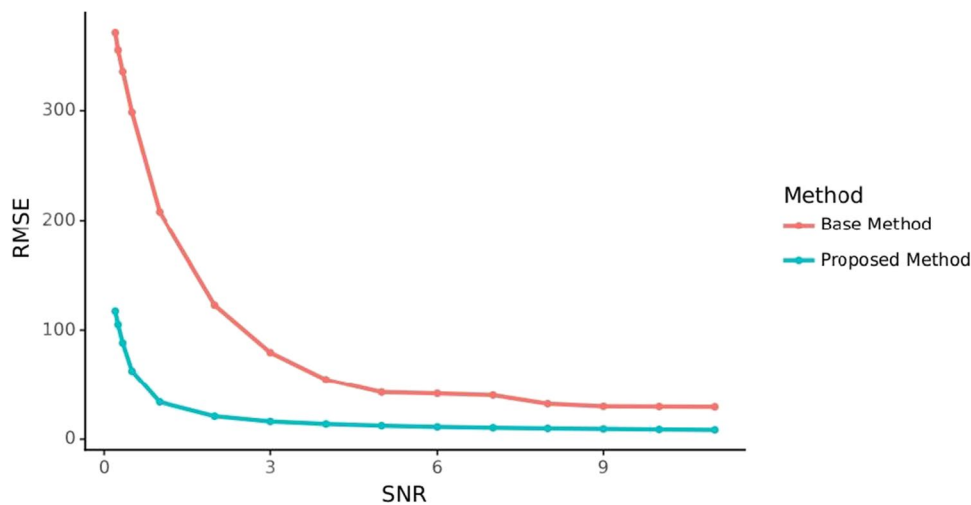
Table 2 Performance comparison of the proposed Rough-Fuzzy CPD (or RoufCP) with existing fuzzy changepoint detection methods under the simulation setups

Mean Setup	Error Setup	RoufCP	BS	PELT	BOCD	FCP	FSCP	FCML	SMUCE	HSMUCE	MJPD
Cont. Jump (S2)	Gaussian	18.46	20.27	32.47	**16.84**	211.66	27.49	50.37	46.93	39.44	37.73
	Gamma	**19.68**	20.56	31.9	71.34	209.6	27.56	54.41	335.56	62.32	45.94
	t_5	26.92	**23.44**	30.35	109.13	250.05	30.51	55.07	315.73	35.46	38.24
	AR	16.64	16.17	32.93	**11.91**	167.28	25.46	50.22	39.81	37.54	37.4
	MA	**22.58**	24.85	30.71	31.97	241.33	29.97	51.25	277.67	63.31	38.13
	ARMA	**8.94**	14.18	33.89	32.42	43.53	18.55	58.8	55.5	43.42	37.37
Smooth Jump (S3)	Gaussian	13.66	14.11	22.31	**11.17**	131.85	19.27	37.83	30.16	22.54	33.31
	Gamma	**13.51**	13.58	21.67	65.21	126.75	19.08	43.41	322.71	39.66	41.22
	t_5	18.14	**16.4**	21.57	111.84	187.78	22.33	43.75	315.77	20.63	34.03
	AR	11.02	11.33	22.82	**8.71**	70.27	17.2	35.33	19.44	18.61	32.88
	MA	**15.6**	16.98	21.8	20.83	178.21	21.42	38.05	267.03	44.97	33.74
	ARMA	**7.03**	8.46	22.75	17.54	10.33	11.79	42.46	53.74	32.81	32.81

Values indicate RMSE in estimating the true changepoint. The lowest values in each row are colored red

5.3 Comparison with existing methods

To compare the performance of the Rough-Fuzzy CPD with existing changepoint detection algorithms, we have chosen different changepoint detection methods. The chosen methods can be broadly classified into 3 main categories - state-of-the-art abrupt changepoint detection algorithms, fuzzy changepoint detection algorithms, and finally some other gradual changepoint detection algorithms. For the abrupt changepoint detection algorithms, we have chosen Binary Segmentation Scott and Knott (1974), Pruned Exact Linear Time (PELT) Killick et al. (2012) and Bayesian Online changepoint Detection (Adams and MacKay 2007). To compare the performance of the Rough-Fuzzy CPD with existing fuzzy changepoint detection algorithms, we have chosen three such methods, namely the FCP algorithm Chang et al. (2015) for regression models, Fuzzy shift changepoint (FSCP) algorithm Lu et al. (2016) and Fuzzy classification maximum likelihood changepoint (FCMLCP) algorithm Lu and Chang (2016). Finally, we have considered 3 gradual changepoint detection methods which do not use fuzzy methods. They are Multiscale Jump Point Detection (MJPD) Wu and Zhou (2020), Simultaneous Multiscale Changepoint Estimator (SMUCE) Frick et al. (2014) and the Heterogeneous Simultaneous Multiscale Changepoint Estimator (H-SMUCE) Pein et al. (2017). To compare the performances between Rough-Fuzzy CPD and these existing methods, we consider Rough-Fuzzy CPD with the tuning parameters $w = \Delta = 50$ in combination with the standard

t-test statistic. While the choice of parameters is subjective, sensitivity analysis (see Section) for w and Δ show that the performance of the proposed method is quite robust to the choice of these hyperparameters.

Since we are only dealing with gradual changepoint detection, we consider the mean functions from scenarios (S2) and (S3) for comparison. To perform a more comprehensive comparison, we consider 6 different models for the errors ϵ_t.

(Gaussian) $\epsilon_t \sim$ iid $\mathcal{N}(0, 1)$
(Gamma) $\epsilon_t \sim$ iid Gamma(θ, k) with $\theta = k = 1$
($t-$distribution) $\epsilon_t \sim$ iid t_5.
(AR) $\epsilon_t = \sqrt{0.91}X_t$ where $X_t - 0.3X_{t-1} = \eta_t$ with η_t being standard Gaussian white noise process.
(MA) $\epsilon_t = X_t/\sqrt{1.25}$ where $X_t = \eta_t + 0.5\eta_{t-1}$ with η_t being standard Gaussian white noise process.
(ARMA) Combining the AR and MA models above, we consider ϵ_t generated from an ARMA(1,1) process with long run variance 1, i.e. $\epsilon_t = X_t/2.142857$ where $X_t - 0.3X_{t-1} = \eta_t + 0.5\eta_{t-1}$ and η_t's are iid $\mathcal{N}(0, 1)$.

Table 2 shows the RMSE in estimating the true changepoint. We see that for the continuous jump in mean (setup S2), the proposed method, RoufCP, performs best in 3 out of the 6 cases while for the smooth jump in mean (setup S3), roufCP performs best in 3 out of the 6 cases. We observe that for both (S2) and (S3), Roufcp has the lowest

Fig. 6 Heatmap showing RMSE in estimating changepoint in scenarios $S2$ and $S3$ for different values of w and Δ
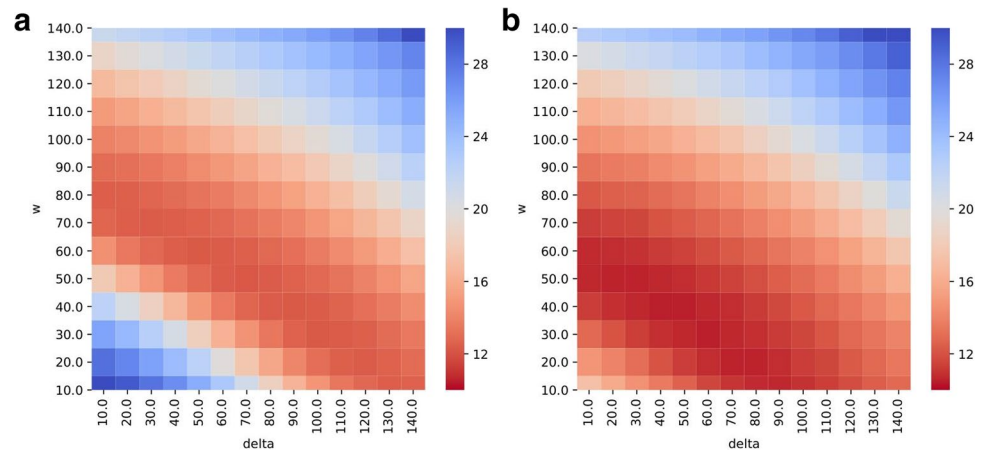
RMSE for Gamma, MA, and ARMA models. Furthermore, the proposed RoufCP method achieves the second lowest RMSE value for Gaussian and t-distributed errors under both setups ($S2$) and ($S3$), and for autoregressive errors for setup ($S3$).

5.4 Sensitivity analysis

The proposed method Rough-Fuzzy CPD depends on 3 main factors viz. the regularity measure $R(t)$, the tunable hyperparameters w and Δ. w denotes the degree of roughness of the tolerance function with higher values indicating greater roughness. Similarly, the parameter Δ determines the fuzziness of the membership function with higher values corresponding to greater fuzziness. While we do not have control over the regularity measure $R(t)$ as it is exogenous to Rough-Fuzzy CPD, we need to choose the values of w and Δ judiciously to obtain better estimates. To understand the effect of parameters w and Δ we perform some sensitivity analysis.

We consider two special scenarios to study the effect of hyperparameters. Figure 6a shows the variation of RMSE for scenario ($S2$) with continuous jump in mean function and Fig. 6b shows the variation of RMSE for scenario ($S3$) with smooth jump in mean function with standard Gaussian white noise errors. In both the cases, we vary the ordered pair $\langle w, \Delta \rangle$ in $\{10, 20, \ldots, 140\} \times \{10, 20, \ldots, 140\}$ and estimate the changepoints using RoufCP. It is evident from Fig. 6 that there exists a wide range of optimal values of w and Δ, and too high or too low values of both these parameters together are detrimental to the efficacy of the model. The difference in the range of values of RMSE between Fig. a and b can be attributed to the difference in the mean curve of the data. As shown by using a shared color scale, Rough-Fuzzy CPD achieves a lower RMSE for scenario ($S3$) than in scenario ($S2$) with their corresponding optimal hyperparameter values. One possible reason
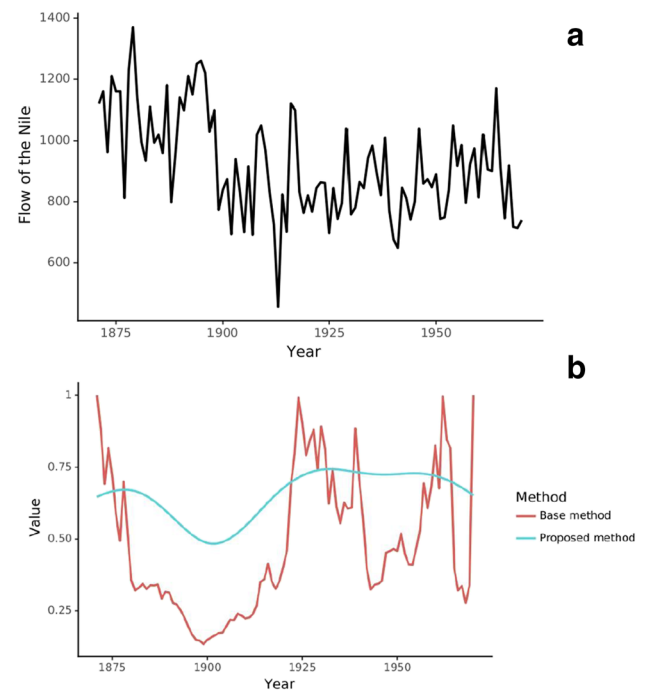

Fig. 7 Performance of proposed Rough-Fuzzy CPD and base method (KS-test) on Nile dataset

could be the resemblance of the chosen membership function for Rough-Fuzzy CPD with the change in the mean function for the scenario ($S3$).

6 Application to real data

We consider three real data sets to show the performance of our method Rough-Fuzzy CPD over the usual regularity measure-based methods. Two popular benchmark datasets in changepoint analysis, namely "Flow of the River Nile" data and "Seatbelts" data regarding monthly Road Casualties in

Fig. 8 Performance of proposed
Rough-Fuzzy CPD and base
method (KS-test) on Seatbelts
dataset
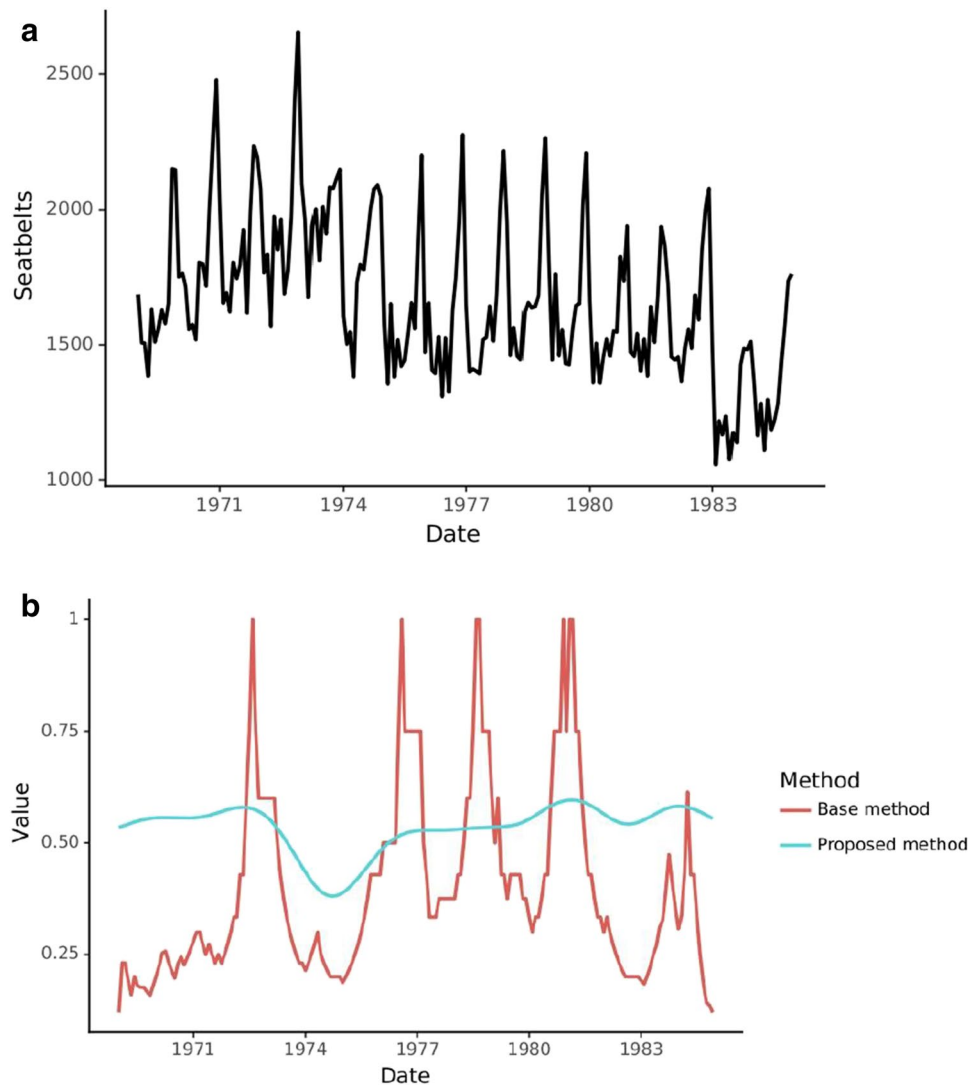
Great Britain 1969 − 84, collected from `datasets` package Durbin et al. (2001) in R, provide the first example. For the second illustration, changepoints are estimated from the time-varying reproduction rate from the coronavirus disease of 2019 (COVID-19) incidence dataset. The final example tries to analyze the gradual changes in the language in the speeches of the presidents of the United States of America. The last example illustrates the applicability of the proposed method Rough-Fuzzy CPD for multivariate datasets.

6.1 Benchmark datasets

The very well-known "Flow of the river Nile" dataset comprises measurements of the annual flow of the Nile river at Aswan from 1871 to 1970. The data has a possible change-point near 1898 which is associated with the Fashoda incident (Bickley and Bates 1984) in the same year.

The result for the Nile dataset is shown in Figure 7, where we use a transformation of the t-test statistic as a regularity measure. As shown in Fig. 7a, there are multiple minima detected by the base method, while the proposed rough-fuzzy improvement smooths out these minima, and the false positives are automatically removed. The estimated changepoint by our Rough-Fuzzy CPD turns out to be at 1902, which coincides with the completion of Zifta Barrage and Assiut Barrage Wegmann (1922) and is close to the commonly believed changepoint at 1898.

Another popular benchmark dataset "Seatbelts" consists of the number of drivers in Great Britain wearing seatbelts during the period January, 1969 to December, 1984. There are possibly two evident changepoints present in the data, the first one corresponds to the start of seat belt legislation movements from 1972 when the seatbelts were enforced compulsory in newly manufactured cars, and the second

Fig. 9 Changepoints in estimated R_t for India
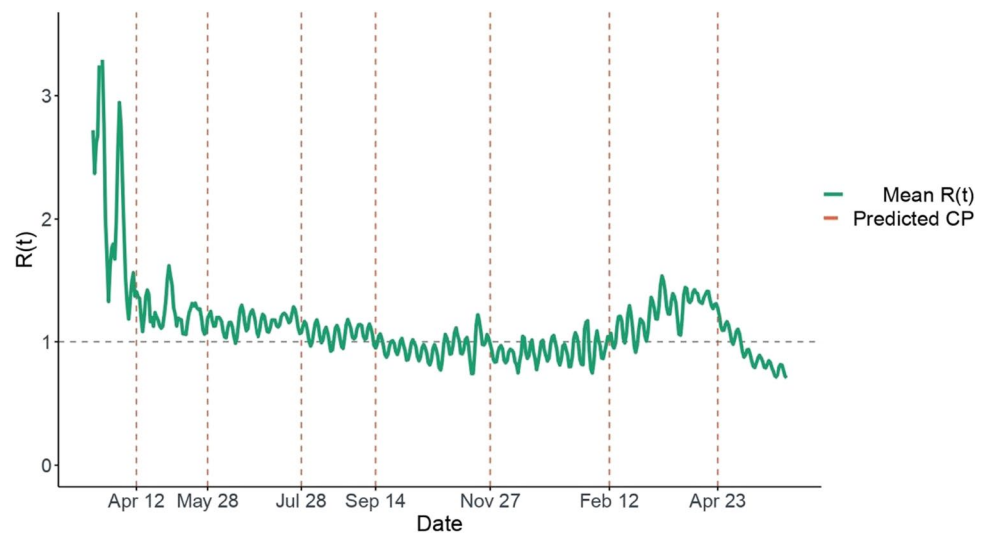
one corresponds to the compulsory enforcement of seat belt wearing while driving in 1983 van den Burg and Williams (2020).

Figure 8 explains the performance of Rough-Fuzzy CPD in detecting changepoints for Seatbelts data. Similar to Nile data, the regularity measure based on the nonparametric Kolmogorov Smirnov test tends to have too many local minima, each of which can be possibly thought of as an estimate of changepoint. However, in conjunction with the rough-fuzzy improvement, the prominent local minima appear in 1975 and 1983, both of which are close to the true changepoints.

6.2 COVID-19 data

The coronavirus disease of 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus. First detected in Wuhan, China in December 2019, it has spread across the globe and has infected nearly 200 Million cases by 4th August 2021. There has been a multitude of epidemiological models trying to estimate the spread of COVID-19 in various countries and states (Purkayastha et al. 2020). However, many such models fail to predict the case counts accurately due to the frequently changing public health measures aimed at simultaneously controlling the spread of the disease and preventing economic crises. This makes the identification of changepoints in the epidemic data of particular importance. Another use of identifying changepoints in the spread of COVID-19 is inference regarding the efficacy of non-pharmaceutical interventions.

Such work has been done for some countries like USA Zhang et al. (2020) and Germany Dehning et al. (2020). To identify changepoints, Dehning et al. (2020) calculated the time-varying rate of transmission $\beta(t)$ in the

usual Susceptible Infected Recovered (SIR) model Dehning et al. (2020) and detected changepoints in $\beta(t)$ through the course of the pandemic. For the SIR model, the transmission rate is defined as the rate of transmission of infection from an infected individual to a susceptible individual Dehning et al. (2020).

While the SIR model and its parameters are relatively straightforward to understand and easy to estimate, the dynamics of transmission of COVID-19 are much more complex due to the presence of factors like undetected asymptomatic transmissions, latency period, incubation period, delay in reporting, under-reporting, different transmission rates for asymptomatic and symptomatic individuals, quarantined and hospitalized individuals, and erroneous testing resulting in false negatives, to name a few Zimmermann et al. (2021). So, $\beta(t)$ alone might not be a suitable representative of the spread of the disease.

So, we look at another related parameter which is the basic reproduction number R_0. While the SIR model assumes constant R_0 we estimate the time-varying reproduction number R_t. R_t can be defined as the expected number of cases directly generated by one case in the population at time t Purkayastha et al. (2020). We estimate R_t using the `EpiEstim` package in R based on the incidence curve from 15th March, 2020 to 1st June, 2021 in India. The required data on the number of confirmed cases of COVID-19 has been collected from an API made by the volunteer-driven covid19india group. We then apply Rough-Fuzzy CPD on the estimated R_t to identify the changepoints. Note that modeling the changepoints in a fuzzy manner, instead of crisp, is logical and appropriate here as interventions and measures rolled out by the government are impossible to be implemented throughout a vast country like India instantaneously.

Figure 9 shows the estimated changepoints along with the estimates of R_t. We see that our method detects 7

Fig. 10 Analysis of change-point detection of US President speeches data
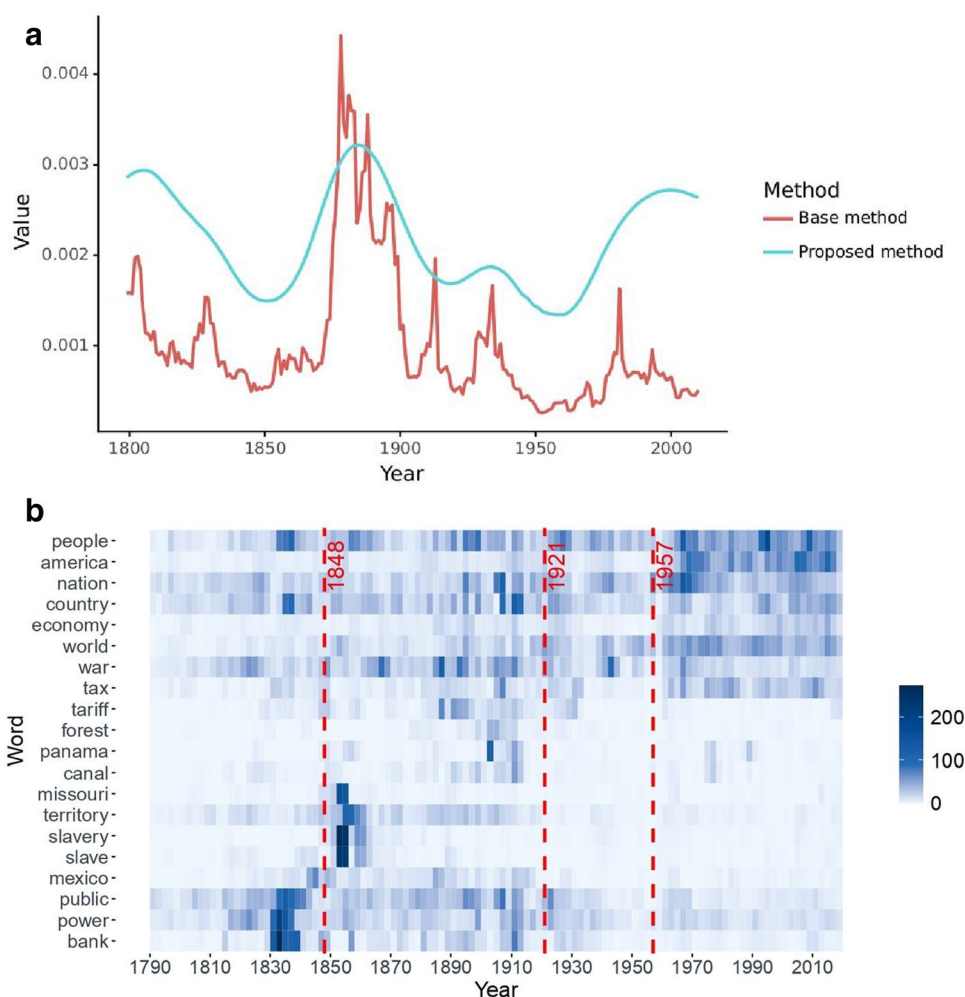
changepoints on 12th April, 28th May, 28th July, 14th September and 27th November in 2020 and 12th February and 23rd April in 2021. The first changepoint (on 12th April, 2020) corresponds to the initial elbow in the estimate of R_t and it is very close to the end of the 1st Lockdown in India (on 14th April, 2020). This shows that the effect of the Lockdown was not immediate and it took the entire 1st Lockdown to bring down the value of R_t substantially and check the unrestricted spread of the disease. Further, the changepoint detected on 12th February, 2021, is very close to 10th February, 2021 which is regarded as the start of the 2nd wave in India Covid-19 in India (2021). Also, the changepoint detected on 12th April 2021, marks the date after which there is a substantial decrease in the value of R_t. In fact, immediately after this date, the value of R_t comes below 1 which is followed by a sharp decline in new cases. Among the other detected changepoints, those on 12th April, 28th May, 28th July and 27th November are very close to the date of commencement of Lockdown 2, Unlock 1, Unlock 3 and Unlock 7 respectively. There are no detected changepoints near the other lockdowns and unlock phases. This implies that the effects of Unlock

1, Unlock 3, and Unlock 7 were much more pronounced on R_t than in the other nationwide intervention phases in India.

6.3 US President Speech Data

From the presidency of George Washington (April 30, 1789) to Joe Biden (current), 46 presidents of the United States have delivered countless speeches during their presidency to address contemporary political, economical, and social issues. Among them, more than a thousand speeches have been recorded in the UVA Miller Center database (Presidential Speeches 2021). Starting from George Washington's first inaugural speech "Fellow citizens of the senate and the house of representatives, among the vicissitudes incident to life no event could have filled me..." to Franklin D. Roosevelt's "iffy" or Donald Trump's "covfefe", the English language spoken by the US presidents have undergone countless changes. While some of the sophisticated archaic words have transformed into more natural colloquial synonyms, the topics addressed by the presidents also varied over

time with relevance to contemporary political, economical, and social issues. This linguistic shift is obviously gradual in nature, hence the proposed RoufCP method can be used to segment the time from 1789 to 2021 into several homogeneous regions, where the languages are similar and are expected to have simpler homogeneous linguistic models. One important thing to note here is that the presidential speech dataset poses a particular challenge in identifying changepoints owing to its high dimensionality because of the huge number of words in the corpus. Such a challenge was not present in the other datasets we considered earlier.

Figure 10b shows the gradual shift in usage of different words. A darker color indicates a higher incidence of a word in a particular period. The words have been chosen to represent different topics that presidents across different periods addressed in their speeches. The topics range from slavery, and war to domestic and international politics, economy, etc. The proposed RoufCP detects 3 changepoints in the years 1848, 1921, and 1957 respectively. Now, in Fig. 10b, we can see that before 1848 higher usage of words like "public", "power", "bank" etc. while immediately after 1848 speeches of presidents included words like "slave", "Missouri" etc. in higher frequencies which corresponds to the movement against slavery and the eventual abolition of slavery in 1865. After that, the second changepoint in 1921 marks a shift in politics and a decrease in usage of words like "Panama canal", "Mexico", and "territory". Finally, we can see that after the last changepoint in 1957, there is a stark increase in the frequency of words like "war", "tax", "America", and "economy", which is related to various events like the cold war and space race. In this way, using RoufCP, we can detect the major changepoints in US politics and linguistics of the English language from the Presidential speech dataset.

7 Conclusion

Though gradual changepoints are present in various time series data, they are usually overlooked. Our study aims to contribute in this direction by enriching the existing methodologies using the principles of the Fuzzy Rough set theory. The biggest strength of our approach is that it is independent of the base method, and thus any changepoint detection algorithm expressed as shown in Fig. 13 can be subjected to improvement by our proposal. Here, we have presented only 3 cases of base methods for computing the regularity measure - parametric two-sample test, nonparametric two-sample tests, and stationarity tests (unit root test). However, our choices are not limited to the aforementioned cases. Any method for detecting changepoint which provides any type of anomaly score

or regularity score can be used in our framework with suitable transformation. Thus rough-fuzzy CPD allows one to utilize the rich collection of methods existing for crisp changepoint detection for the problem of fuzzy changepoint detection.

As various simulations show, even for very simple regularity measures, like t-test statistic and Kolmogorov-Smirnov test statistic, combining them with the rough-fuzzy CPD increases their efficiency by far. In all cases, except for the discrete jump models (where the assumption of fuzziness in changepoint is violated), rough-fuzzy CPD reduces the MSE in detecting the changepoints. In comparison to existing fuzzy and crisp changepoint detection algorithms, rough-fuzzy CPD turns out to be more efficient in the estimation of continuous and smooth changes in the mean. Furthermore, under a gradual change in the mean function, our method outperforms several existing changepoint detection methods by achieving smaller MSE, for different error distributions.

The asymptotic distribution mentioned in Theorems 3 and 2 connects the rough-fuzzy CPD to a hypothesis testing framework, allowing one to output the statistical significance of the estimated changepoints as well. Since Theorem 2 allows us to derive the joint asymptotic distribution of the proposed entropy at multiple time points, we can rule out the false positives and apply the rough-fuzzy CPD for multiple changepoint detection problems.

From a computational point of view, this paper addresses the issue of obtaining closed-form expressions for upper and lower approximations of the rough fuzzy partitions, described by an estimated changepoint, thus allowing the overall algorithm to be much faster. In general, the time and space complexity of the overall algorithm will be dominated by the cost of computing the regularity measures.

Our approach of using fuzzy and rough set theory has definite advantages which are reflected in the robustness of the rough-fuzzy CPD. Under changes in the signal-to-noise ratio, this method gives a steady improvement over the regularity measure-based changepoint detection. We observe that for a diverse range of SNR, the improvement in accuracy of estimates obtained by the rough-fuzzy CPD over that of the base model remains fairly uniform ranging from 88% to 96%. On the other hand, increasing the fuzziness $\mathcal{F}$ in the continuous change in mean function increases the efficiency of the rough-fuzzy CPD in comparison to the base statistic used. As fuzziness increases, a rapid increment in the efficiency in estimation can be observed. Also, rough-fuzzy CPD rarely outputs an outlying or spurious false positive estimate of changepoint, precisely because the entropy curve $H^E_{\Delta,\delta,w}(s)$ possesses more smoothness properties than its regularity measure $R(t)$ counterpart. Even in the cases when the base model predicts a bimodal distribution

of estimated changepoints, our model shrinks the two modes towards the true changepoint, as illustrated in Fig. 3. Finally, through sensitivity analysis, we have shown that the rough-fuzzy CPD performs reasonably well on a wide range of hyperparameter values.

On the flip side, there are some obvious limitations. The rough-fuzzy CPD may not perform well if the assumption of fuzziness in the true changepoint is violated (i.e., there is a discrete jump discontinuity in the mean function), or if the regularity measure $R(t)$ is a bad indicator for the type of changepoint that we are trying to detect. Another limitation could be the choice of the hyperparameters w, Δ, and δ.

A future possibility for extension of this investigation may consider building a probabilistic view of the detected changepoint, which should enable one to provide a confidence interval for the estimate of the changepoint, which is more meaningful in terms of gradual change. Further, rough-fuzzy CPD can be viewed as an extension of an image segmentation algorithm Sen and Pal (2009) into the domain of changepoint detection. A future endeavor in this direction could be to generalize a family of image segmentation algorithms to fit perfectly in the context of a changepoint detection problem.

8 Software

For broader dissemination of our work, we have developed the python package `roufcp` which is available at pypi.org/project/roufcp/. All codes for the package has been open sourced and are made available at a github respository github.com/subroy13/roufcp.

Appendix A: Section title of first appendix

A.1: Proof of Lemma 1

Starting with the lower approximation;

$$
\begin{aligned}
M_{\underline{\gamma}_s}(t) &= \inf_{\psi \in \mathbb{U}} \max \left(\overline{S}_w(t,\psi), \mu_{s,\Delta}(\psi) \right) \\
&= \inf_{\psi \in \mathbb{U}} \left[1 - \min \left(1 - \overline{S}_w(t,\psi), 1 - \mu_{s,\Delta}(\psi) \right) \right] \\
&= \inf_{\psi \in \mathbb{U}} \left[1 - \min \left(S_w(t,\psi), \overline{\mu}_{s,\Delta}(\psi) \right) \right] \\
&= 1 - \sup_{\psi \in \mathbb{U}} \min \left(S_w(t,\psi), \overline{\mu}_{s,\Delta}(\psi) \right) \\
&= 1 - M_{\overline{\gamma}_s^c}(t)
\end{aligned}
$$

The other equality proceeds in an exactly similar way.

A.2: Proof of Theorem 1

To establish an explicit formula for lower and upper approximations, we first consider the situation when $M_{\underline{\gamma}_s}(t) = 0$, which happens if and only if there exists a $\psi \in \mathbb{U}$ such that, both $\overline{S}_w(t,\psi)$ and $\mu_{s,\Delta}(\psi)$ are equal to 0. While the complementary tolerance function $\overline{S}_w$ is 0, if and only if S_w is 1, i.e. the two arguments satisfy $t = \psi$. On the other hand, $\mu_{s,\Delta}(\psi) = 0$ if and only if $\psi \geq (s + \Delta)$, combining this with $t = \psi$ yields, $t \leq (s + \Delta)$.

On the other extreme, $M_{\underline{\gamma}_s}(t) = 1$, if and only if, either $S_w(t,\psi) = 0$ or $\mu_{\Delta,s}(\psi) = 1$. The former happens when $|t - \psi| \geq 2w$, and the latter happens if $\psi \leq (s - \Delta)$. Thus, for any $t \leq (s - 2w - \Delta)$, $M_{\underline{\gamma}_s}(t) = 1$.

By symmetry, $M_{\overline{\gamma}_s}(t) = 0$ if $t \geq (s + 2w + \Delta)$ and $M_{\overline{\gamma}_s}(t) = 1$ if $t \leq (s - \Delta)$.

To see the behaviour of $M_{\underline{\gamma}_s}(t)$ when $(s - 2w - \Delta) \leq t \leq (s + \Delta)$, we note that $M_{\underline{\gamma}_s}(t)$ is a minimizer of $\max \left(\overline{S}_w(t,\psi), \mu_{s,\Delta}(\psi) \right)$ with respect to ψ. Since both of these functions $\overline{S}_w(t,\psi)$ and $\mu_{s,\Delta}(\psi)$ are continuous, the minimax value must be attained at a point t^* where the fuzzy membership function and the tolerance function crosses each other, i.e., $\overline{S}_w(t,t^*) = \mu_{s,\Delta}(t^*)$.

A.3: Proof of Theorem 2

We can rewrite Theorem 1, as follows,

$$
M_{\overline{\gamma}_s}(t) = \begin{cases} 1 & \text{if } t < (s - \Delta) \\ \max\limits_{\{t^* : S_w(t,t^*) = \mu_{s,\Delta}(t^*)\}} \mu_{s,\Delta}(t^*) & \text{if } (s - \Delta) \leq t < (s + 2w + \Delta) \\ 0 & \text{if } t \geq (s + 2w + \Delta) \end{cases}
$$

To prove this lemma, we consider the 3 cases separately.

1. $t < s - 2w - \Delta$
2. $t \geq s + \Delta$
3. $t \in [s - 2w - \Delta, \ s + \Delta)$

1. *Case 1:* $t < s - 2w - \Delta$ Here, $M_{\underline{\gamma}_s}(t) = 1$ and $(2s - T) > (s + \Delta + 2w)$. Thus, $M_{\overline{\gamma}_s}(2s - t) = 0 = 1 - M_{\underline{\gamma}_s}(t)$ which proves Lemma 2 for this case.
2. *Case 2:* $t \geq s + \Delta$ Here, $M_{\underline{\gamma}_s}(t) = 0$ and $(2s - T) \leq (s - \Delta)$. Again, symmetric to the previous case, $M_{\overline{\gamma}_s}(2s - t) = 1 = 1 - M_{\underline{\gamma}_s}(t)$ which proves Lemma 2 for case 2 also.
3. *Case 3:* $(s - \Delta) \leq t < (s + 2w + \Delta)$ In this case, $(2s - T) \in (s - \Delta, \ s + \Delta + 2w]$. Hence,

$$M_{\overline{\gamma_s}}(2s-t) = \max_{\{t^*:S_w(2s-t,t^*)=\mu_{s,\Delta}(t^*)\}} \mu_{s,\Delta}(t^*)$$

$$= \max_{\{t^*:S_w(2s-t,t^*)=1-\mu_{s,\Delta}(2s-t^*)\}} \left(1-\mu_{s,\Delta}(2s-t^*)\right)$$

Since, $\mu_{s,\Delta}(t^*) = 1-\mu_{s,\Delta}(2s-t^*)$

$$= \max_{\{t^*:S_w(2s-t,t^*)+\mu_{s,\Delta}(t^*)=1\}} S_w(2s-t,t^*)$$

$$= \max_{\{z^*:S_w(2s-t,2s-z^*)+\mu_{s,\Delta}(z^*)=1\}} S_w(2s-t,2s-z^*)$$

Putting $z^* = 2s - t^*$

$$= \max_{\{z^*:S_w(t,z^*)+\mu_{s,\Delta}(z^*)=1\}} S_w(t,z^*)$$

Since, $S_w(\cdot,\cdot)$ is location invariant and

$g(\cdot)$ is symmetric about 0

$$= 1 - M_{\underline{\gamma_s}}(t)$$

So, Lemma 2 also holds for case 3, proving that the result is true in general.

A.4: Proof of Theorem 3

We first fix the parameters Δ, w and fix the proposed change-point s. Let us define, $\rho_{\Delta,\delta_n,w}(\gamma_s)$ as a function from $l^\infty(\mathbb{N})$ to $\mathbb{R}$, denoted by $\rho^{(s)}$;

$$\rho^{(s)}(R_{\delta_n}) = 1 - \frac{\sum_{t=1}^\infty M_{\underline{\gamma_s}}(t)R(t)}{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)R(t)} \tag{16}$$

We would first compute the Fréchet derivative of this at $\mu\mathbf{1}(\cdot)$.

$$\left| \rho^{(s)}(\mu\mathbf{1}+h) - \rho^{(s)}(\mu\mathbf{1}) - \frac{\sum_{n=1}^\infty \left[\sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) - \sum_{t=1}^\infty M_{\underline{\gamma_s}}(n)M_{\overline{\gamma_s}}(t)\right]h(n)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \right|$$

$$= \left| \frac{\sum_{n=1}^\infty \left[\sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) - \sum_{t=1}^\infty M_{\underline{\gamma_s}}(n)M_{\overline{\gamma_s}}(t)\right]h(n)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2 + \left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)h(t)\right)} \right.$$
$$\left. - \frac{\sum_{n=1}^\infty \left[\sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) - \sum_{t=1}^\infty M_{\underline{\gamma_s}}(n)M_{\overline{\gamma_s}}(t)\right]h(n)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \right|$$

$$= \left| \frac{\sum_{n=1}^\infty \left[\sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) - \sum_{t=1}^\infty M_{\underline{\gamma_s}}(n)M_{\overline{\gamma_s}}(t)\right]h(n)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \right| \left| \frac{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)h(t)}{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)(\mu+h(t))} \right|$$

where $|\cdot|$ represents the usual absolute value norm on real numbers. Now, since $\mu \neq 0$, without loss of generality assume that, $\mu > \epsilon > 0$ for some small non-negative ϵ. With $\|h\|_\infty \to 0$, we can thus make,

$$\left| \frac{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)h(t)}{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)(\mu+h(t))} \right| \leq \left| \frac{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)h(t)}{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)(\epsilon/2)} \right| \leq B\|h\|_\infty$$

for sufficiently small $\|h\|_\infty$ and some finite real number B. On the other hand,

$$\left| \frac{\sum_{n=1}^\infty \left[\sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) - \sum_{t=1}^\infty M_{\underline{\gamma_s}}(n)M_{\overline{\gamma_s}}(t)\right]h(n)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \right|$$

$$\leq \sum_{n=1}^\infty \left| \frac{\left[\sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) - \sum_{t=1}^\infty M_{\underline{\gamma_s}}(n)M_{\overline{\gamma_s}}(t)\right]h(n)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \right| \quad \text{by triangle inequality}$$

$$\leq \frac{2\|h\|_\infty}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \sum_{n=1}^\infty \left| \sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) \right| \quad \text{since, } |h(n)| \leq \|h\|_\infty$$

$$\leq \frac{2\|h\|_\infty}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \sum_{n=1}^\infty \sum_{t=1}^\infty \left| M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) \right|$$

$$\leq \frac{2\|h\|_\infty}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \left(\sum_{n=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2 \quad (\because M_{\underline{\gamma_s}}(t) \leq M_{\overline{\gamma_s}}(t) \text{ and both are nonnegative})$$

$$= \frac{2}{\mu}\|h\|_\infty$$

Thus,

$$\frac{1}{\|h\|_\infty}\left| \rho^{(s)}(\mu\mathbf{1}+h) - \rho^{(s)}(\mu\mathbf{1}) - \rho^{(s)'}(\mu\mathbf{1})(h) \right| < \frac{2B}{\mu}\|h\|_\infty$$

which goes to 0 as $\|h\|_\infty \to 0$, where $\rho^{(s)'}(\mu\mathbf{1}) : l^\infty(\mathbb{N}) \to \mathbb{R}$ is;

$$\rho^{(s)'}(\mu\mathbf{1})(h) = \sum_{n=1}^\infty \left[\frac{\sum_{t=1}^\infty M_{\overline{\gamma_s}}(n)M_{\underline{\gamma_s}}(t) - \sum_{t=1}^\infty M_{\underline{\gamma_s}}(n)M_{\overline{\gamma_s}}(t)}{\mu\left(\sum_{t=1}^\infty M_{\overline{\gamma_s}}(t)\right)^2} \right]h(n) \tag{17}$$

To see that the derivative given in Eq. 17 is well defined, note that

$$\left| \rho^{(s)'}(\mu\mathbf{1})(h) \right| \leq \frac{2}{\mu}\|h\|_\infty,$$

which shows that the infinite series is convergent as $h \in l^\infty(\mathbb{N})$.

This shows that the function $\rho^{(s)}$ given in Eq. 16 is Fréchet differentiable (see de Miranda and Fichmann 2005 for definition) at $\mu\mathbf{1}$ and the derivative is given by Eq. 17. In a very similar way, the roughness measure corresponding to the complementement set γ_s^C can also be shown to be Fréchet differentiable at $\mu\mathbf{1}$ as a function of $\{R(t) : t = 0, 1, 2, \dots\}$.

Now, in order to obtain the exponential entropy as shown in Eq. 11, we consider the function $g : [0, 1]^2 \to \mathbb{R}$ defined as;

$$g(x, y) = xe^{(1-x)} + ye^{(1-y)}$$

Clearly, g is Fréchet differentiable at every point of the domain. Finally, a chain rule can be applied on the composition of g and the roughness measures to show that, Ψ_s is Fréchet differentiable at $\mu\mathbf{1}$ where $\Psi_s : l^\infty(\mathbb{N}) \to \mathbb{R}$ is such that $\Psi_s(R_{\delta_n}) = H^E_{\Delta,\delta_n,w}(s)$.

Since Fréchet differentiability implies Hadamard differentiability (see discussion followed by Definition 2.1 of Shao 1993), and outputs a linear operator as the derivative, Ψ_s is also Hadamard differentiable with the derivative given by;

$$\Psi'_s\big|_{\mu\mathbf{1}(\cdot)}(h) = \sum_{n=1}^{\infty} A_s(n)h(n)$$

where the function $A(n)$ is as given in Eq. 14. Clearly, this is a linear and continuous map.

The proof of the result now follows from an infinite dimensional generalization of the delta method (Theorem 1 and Section 1.4 of Römisch 2014). Based on the discussion above, we see that due to the assumption, Z is a Gaussian process, and also the Hadamard derivative of the function Ψ_s is linear. Thus, $\Psi'_s(Z)$ is a normally distributed random variable with the mean 0 and variance σ^*, as given in the statement of the theorem.

References

Adams RP, MacKay DJ (2007) Bayesian online changepoint detection. arXiv:0710.3742

Angelo DMF, Palhares RM, Takahashi RH, Loschi RH, Baccarini LM, Caminhas WM (2011) Incipient fault detection in induction machine stator-winding using a fuzzy-bayesian change point detection approach. Applied Soft Computing 11(1), 179–192

Basseville M, Nikiforov IV, et al (1993) Detection of Abrupt Changes: Theory and Application vol. 104. Prentice hall Englewood Cliffs

Bickley LSMDF, Bates D (1984) The Fashoda Incident of 1898: Encounter on the Nile. Oxford University Press. https://books.google.co.in/books?id=pk2aAAAAIAAJ. Accessed: 2021-06-16

Brodsky E, Darkhovsky BS (2013) Nonparametric Methods in Change Point Problems vol. 243. Springer

Chang S-T, Lu K-P, Yang M-S (2015) Fuzzy change-point algorithms for regression models. IEEE Transactions on Fuzzy Systems 23(6):2343–2357

Chatterjee RK, Kar A (2017) Global image thresholding based on change-point detection. In: Proceedings of the International MultiConference of Engineers and Computer Scientists 2017, pp. 434–438

Chen J, Gupta AK (2012) Parametric statistical change point analysis: with applications to genetics, medicine, and finance

Covid-19 in India (2021) Why second coronavirus wave is devastating. https://www.bbc.com/news/world-asia-india-56811315. Accessed: 2021-04-22

de Miranda JCS, Fichmann L (2005) A generalization of the concept of differentiability. Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo 6(4):397–427

Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, Priesemann V (2020) Inferring change points in the spread of covid-19 reveals the effectiveness of interventions. Science 369(6500). https://doi.org/10.1126/science.abb9789

Durbin J, James Durbin SJK, Koopman SJ (2001) Time series analysis by state space methods. Oxford Statistical Science Series. Clarendon Press. https://books.google.co.in/books?id=XRCu5iSz_HwC. Accessed: 2021-06-16

Frick K, Munk A, Sieling H (2014) Multiscale change point inference. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76(3):495–580. https://doi.org/10.1111/rssb.12047

Gan F (1992) Cusum control charts under linear drift. Journal of the Royal Statistical Society: Series D (The Statistician) 41(1):71–84

Harchaoui Z, Moulines E, Bach FR (2009) Kernel change-point analysis. In: Advances in Neural Information Processing Systems, pp. 609–616

Haynes K, Fearnhead P, Eckley IA (2017) A computationally efficient nonparametric approach for changepoint detection. Statistics and Computing 27(5):1293–1305

Hinkley DV (1970) Inference about the change-point in a sequence of random variables. Oxford University Press

Hunt G (2008) Gradual or pulsed evolution: when should punctuational explanations be preferred? Paleobiology 34(3):360–377

Hušková M (1998) Estimators in the location model with gradual changes. Comment Math Univ Carol 39(1):147–157

Killick R, Fearnhead P, Eckley IA (2012) Optimal detection of change-points with a linear computational cost. Journal of the American Statistical Association 107(500):1590–1598

Lemire D (2007) A better alternative to piecewise linear time series segmentation. In: Proceedings of the 2007 SIAM International Conference on Data Mining, pp. 545–550. SIAM

Liang W, Xu L (2021) Gradual variance change point detection with a smoothly changing mean trend. Stat 10(1):327

Lu K-P, Chang S-T, Yang M-S (2016) Change-point detection for shifts in control charts using fuzzy shift change-point algorithms. Computers & Industrial Engineering 93:12–27

Lu K-P, Chang S-T (2016) Detecting change-points for shifts in mean and variance using fuzzy classification maximum likelihood change-point algorithms. J Comput Appl Math 308:447–463

Mallik A, Sen B, Banerjee M, Michailidis G (2011) Threshold estimation based on a p-value framework in dose-response and regression settings. Biometrika 98(4):887–900

Page E (1955) A test for a change in a parameter occurring at an unknown point. Biometrika 42(3/4):523–527

Pal SK, Dutta-Majumder DK (1986) Fuzzy Mathematical Approach to Pattern Recognition. Halsted Press

Pal NR, Pal SK (1991) Entropy: a new definition and its applications. IEEE Trans Syst Man Cybern 21(5):1260–1270

Pal SK, Skowron A (1999) Rough-fuzzy Hybridization: a New Trend in Decision Making. Springer

Pathak A, Pal SK (1986) Fuzzy grammars in syntactic recognition of skeletal maturity from x-rays. IEEE Trans Syst Man Cybern 16(5):657–667

Pawlak Z (1982) Rough sets. International Journal of Computer and Information Sciences

Pein F, Sieling H, Munk A (2017) Heterogeneous change point inference. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79(4):1207–1227. https://doi.org/10.1111/rssb.12202

Presidential Speeches (2021) Miller Center. https://millercenter.org/the-presidency/presidential-speeches. Accessed: 2021-06-16

Purkayastha S, Bhattacharyya R, Bhaduri R, Kundu R, Gu X, Salvatore M, Mishra S, Mukherjee B (2020) A comparison of five epidemiological models for transmission of sars-cov-2 in india. https://doi.org/10.1101/2020.09.19.20198010

Römisch W (2014) Delta method, infinite dimensional. Statistics Reference Online, Wiley StatsRef

Rose KC, Bown TM (1986) Gradual evolution and species discrimination in the fossil record. Rocky Mountain Geology 24(special_paper_3), 119–130

Scott AJ, Knott M (1974) A cluster analysis method for grouping means in the analysis of variance. Biometrics 30(3):507–512

Sen D, Pal SK (2009) Generalized rough sets, entropy, and image ambiguity measures. Trans. Sys. Man Cyber. Part B 39(1):117–128. https://doi.org/10.1109/TSMCB.2008.2005527

Serfling RJ (2009) Approximation Theorems of Mathematical Statistics vol. 162. John Wiley & Sons

Shao J (1993) Differentiability of statistical functionals and consistency of the jackknife. Ann Stat 21(1):61–75

Sharma D, Sankaran L (2011) Cognitive and social forces in dialect shift: Gradual change in london asian speech. Language variation and Change

Trauth M, Foerster V, Junginger A, Asrat A, Lamb H, Schaebitz F (2018) Abrupt or gradual? change point analysis of the late pleistocene-holocene chew bahir record from southern ethiopia. Quaternary Research 90(2):321–330

Truong C, Oudre L, Vayatis N (2020) Selective review of offline change point detection methods. Signal Processing 167:107299

Turner R, Saatci Y, Rasmussen CE (2009) Adaptive sequential bayesian change point detection. In: Temporal Segmentation Workshop at NIPS, pp. 1–4

van den Burg GJ, Williams CK (2020) An evaluation of change point detection algorithms. arXiv:2003.06222

Vogt M, Dette H et al (2015) Detecting gradual changes in locally stationary processes. Ann Stat 43(2):713–740

Wegmann E (1922) The design and construction of dams: including masonry, earth, rock-fill, timber, and steel structures, also the principal types of movable dams. John Wiley & Sons, Incorporated. https://books.google.co.in/books?id=Lc91swEACAAJ. Accessed: 2021-06-16

Wu W, Zhou Z (2020) Multiscale jump testing and estimation under complex temporal dynamics

Zadeh LA (1965) Fuzzy sets. Information and control 8(3):338–353

Zhang S, Xu Z, Peng H (2020) Change point modeling of covid-19 data in the united states. Society of Statistics, Computer and Applications (SSCA)

Zimmermann L, Bhattacharya S, Purkayastha S, Kundu R, Bhaduri R, Ghosh P, Mukherjee B (2021) Sars-cov-2 infection fatality rates in india: Systematic review, meta-analysis and model-based estimation. Studies in Microeconomics 9(2):137-179. https://doi.org/10.1177/23210222211054324