



Article

Cluster Analysis on Locally Asymptotically Self-Similar Processes with Known Number of Clusters

Nan Rao ^{1,*} , Qidi Peng ² and Ran Zhao ³ ¹ School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China² Institute of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711, USA; qidi.peng@cgu.edu³ University of Liverpool Management School, University of Liverpool, Liverpool L69 7ZH, UK; ran.zhao@liverpool.ac.uk

* Correspondence: nan.rao@sjtu.edu.cn; Tel.: +86-188-7988-7743

Abstract: We conduct cluster analysis of a class of locally asymptotically self-similar stochastic processes with finite covariance structures, which includes Brownian motion, fractional Brownian motion, and multifractional Brownian motion as paradigmatic examples. Given the true number of clusters, a new covariance-based dissimilarity measure is introduced, based on which we obtain approximately asymptotically consistent algorithms for clustering locally asymptotically self-similar stochastic processes. In the simulation study, clustering data sampled from fractional and multifractional Brownian motions with distinct Hurst parameters illustrates the approximated asymptotic consistency of the proposed algorithms. Clustering global financial markets' equity indexes returns and sovereign CDS spreads provides a successful real world application. Implementations in MATLAB of the proposed algorithms and the simulation study are publicly shared in GitHub.

Keywords: local asymptotic self-similarity; multifractional Brownian motion; stochastic processes clustering; covariance-based dissimilarity measure; approximated asymptotic consistency



Citation: Rao, N.; Peng, Q.; Zhao, R. Cluster Analysis on Locally Asymptotically Self-Similar Processes with Known Number of Clusters. *Fractal Fract.* **2022**, *6*, 222. <https://doi.org/10.3390/fractalfract6040222>

Academic Editor: Sergio Bianchi

Received: 20 February 2022

Accepted: 11 April 2022

Published: 14 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Learning stochastic processes is a fast growing research area in machine learning, as there is a considerable number of machine learning problems that involve the time variable as a component of datasets [1–5]. Due to the nature of the time index, stochastic processes often possess “path features” [6]. This additional information provided by the time index makes performing machine learning on stochastic processes quite different from other types of objects. Several new techniques are developed alongside studying such machine learning problems [7–10]. In this paper, we focus our study on a particular unsupervised learning problem: clustering of stochastic processes. Recently, clustering distribution stationary ergodic time series were motivated by and discussed in [11,12]. Later, Peng et al. [13] developed Khaleghi et al. [12]'s consistent clustering algorithms in order to cluster covariance stationary ergodic (discrete-time or continuous-time) stochastic processes. Please note that a distribution stationary ergodic stochastic process is not necessarily covariance stationary. Therefore, Peng et al. [13] enlarged the class of stochastic processes for which one can apply the consistent clustering algorithms.

The motivation for this paper lies in the fact that assuming the observed data to follow a covariance stationary stochastic process seems still unrealistic. Our paper then tries to overcome this issue through developing a promising algorithm to cluster a more general class of stochastic processes, the so-called locally asymptotically self-similar processes. More precisely, the key path features of the observed stochastic processes are assumed to be known as Assumption 1 (see Section 2). Compared to the stationarity, Assumption 1 is a much weaker constraint, for the reason that in many fields it was proved that the observed paths are sampled from functions (or functionals) of well-known locally

asymptotically self-similar processes. For example, dynamics in financial markets (equity returns and interest rates) can be described based on geometric Brownian motions (gBm); long-term dependent or self-similar phenomena are often modeled by fractional Brownian motions (fBm) [14]. The long-term financial indexes and curves, such as S&P 500, Dow Jones, NASDAQ, interest rates, VIX rates, and currency exchange rates can be modeled using multifractional Brownian motions (mBm) [15–20]. Modeling events using locally asymptotically self-similar processes can be widely found in other fields such as geology, biology, power, and energy [21,22]. Recently, there was growth of investigations on how to test and estimate locally asymptotically self-similar processes, and on how to apply machine learning analysis to such processes [23–30]. However, in as far as we know, there has not yet been a study on clustering locally asymptotically self-similar processes in the literature. Our paper then aims at shedding some light on clustering such processes.

Contrary to the conventional clustering of finite-dimensional data, clustering based on the paths' features of the processes largely removes the noise by capturing the observations' path features. Therefore, a nice dissimilarity measure should be the one that well characterizes the path features. In this context, "nice" refers to the property that the computational complexity and the prediction errors caused by the over-fitting issues are expected to be largely reduced. Moreover, with some path features, consistency of the clustering algorithm [11,12] may be obtained. Among all the stochastic process features, we focus on characterizing the property of ergodicity in this paper. However similar analysis can be made for other patterns of process features such as seasonality, Markov property and martingale property.

Ergodicity [31] is a very typical feature possessed by several well-known processes, which is applied to financial time series analysis. It is tightly related to other process features, such as stationarity, long-term memory, and self-similarity [32,33]. In [12,13], it is shown that both distribution ergodicity and covariance ergodicity lead to obtaining an asymptotically consistent clustering algorithms for clustering processes. In this paper, we take one step further to relax the condition of ergodicity to the "local asymptotic ergodicity" [34] and obtain the so-called "approximately asymptotically consistent algorithms" for clustering processes having such path property. This setting presents such a large class of processes that includes the well-known Lévy processes, some self-similar processes and some multifractional processes [34].

Each clustering stochastic processes problem involves handling data, defining clusters, measuring dissimilarities, and finding groups efficiently; therefore, we organize the paper as follows. Section 2 is devoted to introducing a class of locally asymptotically self-similar processes to which our clustering approaches apply. In Section 3, a covariance-based dissimilarity measure is suggested and in Section 4, the approximately asymptotically consistent algorithms for clustering both offline and online datasets are designed. A simulation study is performed in Section 5, where the algorithms are applied to cluster multifractional Brownian motions (mBm), an excellent representative of the class of locally asymptotically self-similar processes. In Section 6, we perform cluster analysis over a global equity return data. Conventionally, stock returns of countries in the same region are considered to have similar patterns due to the common regional economic factors. However, recent empirical evidence shows that, as financial market globalization increases, global economic clusters switch from "geographical centriods" to "economic development centriods". Our clustering algorithms show how "geography" and "economic development" jointly impact the equity returns of countries or regions. Considering the equities as stochastic processes in the clustering makes the analysis more promising. Finally, Section 7 concludes and provides future prospects.

2. A Class of Locally Asymptotically Self-Similar Processes

Self-similar processes are a class of processes that are invariant in distribution under suitable scaling of time [35–37]. These processes have been used to successfully model

various time-scaling random phenomena observed in high frequency data, especially in the financial data and geological data.

Definition 1 (Self-similar process). A stochastic process $\{Y_t^{(H)}\}_{t \geq 0}$ (here the time indexes set is not necessarily continuous) is self-similar to self-similarity index $H \in (0, 1)$ if, for all $n \in \mathbb{N} := \{1, 2, \dots\}$, all $t_1, \dots, t_n \geq 0$ and all $c > 0$,

$$(Y_{ct_1}^{(H)}, \dots, Y_{ct_n}^{(H)}) \stackrel{\text{law}}{=} (c^H Y_{t_1}^{(H)}, \dots, c^H Y_{t_n}^{(H)}), \tag{1}$$

where $\stackrel{\text{law}}{=}$ denotes the equality in joint probability distribution of two finite-dimensional random vectors.

When $\mathbb{E}|Y_t^{(H)}| < +\infty$ for $t \geq 0$, it follows from (1) that for any $t \geq 0$,

$$\mathbb{E}(Y_t^{(H)}) = c^H \mathbb{E}(Y_{t/c}^{(H)}), \text{ for all } c > 0. \tag{2}$$

Therefore taking $t = 0$ at both hand sides of (2) yields

$$\mathbb{E}(Y_0^{(H)}) = 0. \tag{3}$$

Self-similar processes are generally not distribution stationary but their increment processes can be distribution stationary (any finite subset’s joint distribution is invariant subject to time shift) or covariance stationary (its mean and covariance structure exist and are an invariant subject to time shift). From now on we restrict our setting to stochastically continuous-time self-similar processes only [37]. i.e., a process $\{X_t\}_{t \geq 0}$ is stochastically continuous at $t_0 \geq 0$ if

$$\mathbb{P}(|X(t_0 + h) - X(t_0)| > \varepsilon) \xrightarrow{h \rightarrow 0^+} 0, \text{ for any } \varepsilon > 0.$$

This assumption is weaker than the almost sure continuity. The process $\{X_t\}_{t \geq 0}$ is called (stochastically) continuous time over $[0, +\infty)$ if it is continuous at each $t \geq 0$. For $u > 0$, we call $\{Y(t)\}_t = \{X(t + u) - X(t)\}_t$ the increment process (or simply the increments) of $\{X(t)\}_t$. If a continuous-time self-similar process’ all increment processes are covariance stationary, its covariance structure can be explicitly given below:

Theorem 1. Let $\{X_t^{(H)}\}_{t \geq 0}$ be a self-similar process with index $H \in (0, 1)$ and with covariance stationary increments. Then

$$\mathbb{E}(X_t^{(H)}) = 0, \text{ for all } t \geq 0, \tag{4}$$

and

$$\text{Cov}(X_s^{(H)}, X_t^{(H)}) = \frac{\text{Var}(X_1^{(H)})}{2} (|s|^{2H} + |t|^{2H} - |s - t|^{2H}), \text{ for any } s, t \geq 0. \tag{5}$$

Theorem 1 can be obtained by replacing the distribution stationary increments assumption in Theorem 1.2 in [36] with covariance stationary increments assumption. We briefly provide the proof below.

Proof. We first prove (4). On one hand, using the fact that the increments of $\{X_t^{(H)}\}_t$ are covariance stationary and (3), we obtain

$$\mathbb{E}(X_{mt}^{(H)}) = \mathbb{E}\left(\sum_{k=0}^{m-1} (X_{(k+1)t}^{(H)} - X_{kt}^{(H)}) + X_0^{(H)}\right) = m\mathbb{E}(X_t^{(H)}), \text{ for all } m \in \mathbb{N}. \tag{6}$$

On the other hand, since $\{X_t^{(H)}\}_t$ is self-similar to index H , we have

$$\mathbb{E}(X_{mt}^{(H)}) = m^H \mathbb{E}(X_t^{(H)}), \text{ for all } m \in \mathbb{N}. \tag{7}$$

Putting together (6) and (7) and the fact that $0 < H < 1$, we necessarily have $\mathbb{E}(X_t^{(H)}) = 0$ for all $t \geq 0$. (4) is proved.

For proving (5) we first observe that, for $s, t \geq 0$,

$$\mathbb{E}(X_s^{(H)} X_t^{(H)}) = \frac{1}{2} (\mathbb{E}(X_s^{(H)})^2 + \mathbb{E}(X_t^{(H)})^2 - \mathbb{E}(X_s^{(H)} - X_t^{(H)})^2). \tag{8}$$

Next we can see from the facts that $\{X_t^{(H)}\}_t$ is self-similar to index H ; that its increments are covariance stationary and (4), that

$$\begin{aligned} \mathbb{E}(X_s^{(H)})^2 &= |s|^{2H} \mathbb{E}(X_1^{(H)})^2 = |s|^{2H} \text{Var}(X_1^{(H)}), \text{ for } s \geq 0; \\ \mathbb{E}(X_s^{(H)} - X_t^{(H)})^2 &= \text{Var}(X_{|s-t|}^{(H)}) = |s-t|^{2H} \text{Var}(X_1^{(H)}), \text{ for } s, t \geq 0. \end{aligned} \tag{9}$$

The covariance stationarity yields $\text{Var}(X_1^{(H)}) < +\infty$. (5) then follows from (8) and (9). Theorem 1 is proved. \square

We highlight that, contrary to Theorem 1.2 in [36], the covariance stationary increment process of $\{X_t^{(H)}\}_t$ in Theorem 1 is not necessarily distribution stationary. This fact inspires us to relax the distribution stationarity of the processes to the covariance stationarity in the forthcoming Assumption (A). Below, we introduce a natural extension of self-similar processes, the so-called locally asymptotically self-similar processes [34,38,39].

Definition 2 (Locally asymptotically self-similar process). *A continuous-time stochastic process $\{Z_t^{(H(t))}\}_{t \geq 0}$ with its index $H(\bullet)$ being a continuous function valued in $(0, 1)$, is called locally asymptotically self-similar, if for each $t \geq 0$, there exists a non-degenerate self-similar process $\{Y_u^{(H(t))}\}_{u \geq 0}$ with self-similarity index $H(t)$, such that*

$$\left\{ \frac{Z_{t+\tau u}^{(H(t+\tau u))} - Z_t^{(H(t))}}{\tau^{H(t)}} \right\}_{u \geq 0} \xrightarrow[\tau \rightarrow 0^+]{f.d.d.} \{Y_u^{(H(t))}\}_{u \geq 0} \tag{10}$$

where the convergence $\xrightarrow{f.d.d.}$ is in the sense of all the finite-dimensional distributions.

In (10), $\{Y_u^{(H(t))}\}_u$ is called the *tangent process* of $\{Z_t^{(H(t))}\}_t$ at t [38,39]. Moreover, it is shown (see Theorem 3.8 in [39]) that, if $\{Y_u^{(H(t))}\}_u$ is unique in law, it is then self-similar to index $H(t)$ and it has distribution stationary increments. Then the local asymptotic self-similarity generalizes the conventional self-similarity, in the sense that, any non-degenerate self-similar process with distribution stationary increments is locally asymptotically self-similar and its tangent process is itself. Furthermore, in a weaker sense, it is not difficult to show the following:

Proposition 1. *Let $\{Z_t^{(H)}\}_{t \geq 0}$ be a continuous-time self-similar process with self-similarity index $H \in (0, 1)$ and with covariance stationary increments. Then all its tangent processes share the same mean and covariance function.*

Proof. Since $\{Z_t^{(H)}\}_{t \geq 0}$ is locally asymptotically self-similar, by definition at each $t \geq 0$ there exists a tangent process $\{Y_u^{(H)}\}_{u \geq 0}$ such that

$$\left\{ \frac{Z_{t+\tau u}^{(H)} - Z_t^{(H)}}{\tau^H} \right\}_{u \geq 0} \xrightarrow[\tau \rightarrow 0^+]{f.d.d.} \{Y_u^{(H)}\}_{u \geq 0}. \tag{11}$$

Next we show $\{Y_u^{(H)}\}_{u \geq 0}$'s mean and covariance structure are uniquely determined.

Since $\{Z_t^{(H)}\}_{t \geq 0}$ has covariance stationary increments, for any $u \geq 0, \tau > 0$, define the scaled increments

$$Y_{u,\tau}^{(H)} := \frac{Z_{t+\tau u}^{(H)} - Z_t^{(H)}}{\tau^H}.$$

Again by the fact that $\{Z_t^{(H)}\}_{t \geq 0}$ has covariance stationary increments, using (4) in Theorem 1 we obtain

$$\mathbb{E}(Y_{u,\tau}^{(H)}) = 0, \text{ for all } u \geq 0, \tau > 0, \tag{12}$$

and by (5) in Theorem 1, we have for $u_1, u_2 \geq 0$ and $\tau > 0$,

$$\begin{aligned} \text{Cov}(Y_{u_1,\tau}^{(H)}, Y_{u_2,\tau}^{(H)}) &= \tau^{-2H} \text{Cov}(Z_{t+\tau u_1}^{(H)} - Z_t^{(H)}, Z_{t+\tau u_2}^{(H)} - Z_t^{(H)}) \\ &= \frac{\text{Var}(Z_1^{(H)})}{2} (|u_1|^{2H} + |u_2|^{2H} - |u_1 - u_2|^{2H}), \end{aligned} \tag{13}$$

which does not depend on τ .

It follows from (11)–(13) that

$$\begin{aligned} \mathbb{E}(Y_u^{(H)}) &= \lim_{\tau \rightarrow 0^+} \mathbb{E}(Y_{u,\tau}^{(H)}) = 0, \text{ for all } u \geq 0; \\ \text{Cov}(Y_{u_1}^{(H)}, Y_{u_2}^{(H)}) &= \lim_{\tau \rightarrow 0^+} \text{Cov}(Y_{u_1,\tau}^{(H)}, Y_{u_2,\tau}^{(H)}) \\ &= \frac{\text{Var}(Z_1^{(H)})}{2} (|u_1|^{2H} + |u_2|^{2H} - |u_1 - u_2|^{2H}), \text{ for } u_1, u_2 \geq 0. \end{aligned} \tag{14}$$

(14) implies that all tangent processes of $\{Z_t^{(H)}\}_{t \geq 0}$ possess zero-mean and equal covariance functions. By the way it is easy to derive from (14) that these tangent processes have covariance stationary increments. Proposition 1 is proved. \square

We remark from Proposition 1 that the tangent processes of $\{Z_t^{(H)}\}_{t \geq 0}$ may not be unique in law, but their finite-dimensional subsets have unique first and second order moments.

Based on the above discussion, throughout this paper, we assume that the observed datasets are sampled from a known number (denoted by κ) of continuous-time processes satisfying the following condition:

Assumption 1. *The processes are locally asymptotically self-similar; their tangent processes' increment processes are autocovariance ergodic.*

In Assumption 1, the autocovariance-ergodicity means that the sample autocovariance functions of the covariance stationary process converges in squared mean to the autocovariance functions of the process, i.e., a zero-mean (that is the case for the tangent processes' increments) continuous-time process $\{X(t)\}_{t \geq 0}$ is autocovariance ergodic if it is covariance stationary and satisfies

$$\frac{1}{T} \int_0^T X(t + \tau)X(t) dt \xrightarrow[T \rightarrow +\infty]{L^2(\mathbb{P})} \mathbb{E}(X(u + \tau)X(u)), \text{ for all } u > 0, \tau \geq 0, \tag{15}$$

where $X_n \xrightarrow[n \rightarrow +\infty]{L^2(\mathbb{P})} X$ denotes the mean squared convergence: $\mathbb{E}|X_n - X|^2 \xrightarrow[n \rightarrow +\infty]{} 0$. Please note that the above convergence (15) yields

$$\frac{1}{n - \tau - 1} \sum_{k=1}^{n-\tau} X(k)X(k + \tau) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}(X(1)X(1 + \tau)), \text{ for } \tau \in \mathbb{N}. \tag{16}$$

Thus, Assumption 1 says that the observed processes’ tangent processes have covariance stationary increments. The typical examples of locally asymptotically self-similar processes satisfying Assumption 1 are fractional Brownian motion (fBm) [40], multifractional Brownian motion (mBm) [41–43] and the generalized multifractional Brownian motion introduced in [44]. Below, we focus our attention on mBm, which is used in our simulation study and real world application (see Sections 5 and 6). MBm is a paradigmatic example of both multifractional stochastic processes and locally asymptotically self-similar processes. It naturally extends the classical fBm by allowing its Hurst parameter to vary with time. The mBm was introduced independently by Peltier and Lévy-Véhel [41] and Benassi et al. [42], using, respectively, integral moving average type representation and harmonizable integral representation of the fBm. These two types of mBms share several core features and their precise connection was studied by Stoev and Taqqu [43], who show that the two types of mBms generally have different correlation structures. The most recent discussion on the definition of mBm is also made in [43], where they define a general class of multifractional Gaussian processes which includes the above two types of mBms as two particular cases. In this paper, we adopt a definition of mBm through the so-called harmonizable integral representation (see (1.3) in [43] or see [42,44]). Please note that our analysis and approaches are valid for all other versions of mBms in the literature.

Definition 3 (Multifractional Brownian motion). *A multifractional Brownian motion $\{W_{H(t)}(t)\}_{t \geq 0}$ is a continuous-time Gaussian process defined by:*

$$W_{H(0)}(0) = 0 \text{ a.s. and } W_{H(t)}(t) = \int_{\mathbb{R}} \frac{e^{it\xi} - 1}{|\xi|^{H(t)+1/2}} d\tilde{W}(\xi), \text{ for } t > 0,$$

where:

- $d\tilde{W}(\xi)$ denotes a complex-valued Gaussian measure (see Proposition 2.1 in [43]) satisfying

$$\int_{\mathbb{R}} \tilde{f}(\xi) d\tilde{W}(\xi) = \int_{\mathbb{R}} f(t) dW(t) \text{ a.s.}$$

for any

$$f \in L^2(\mathbb{R}) := \left\{ \{f : \mathbb{R} \rightarrow \mathbb{R}\} : \int_{\mathbb{R}} |f(u)|^2 du < +\infty \right\},$$

with $\tilde{f}(\xi) := (2\pi)^{-1/2} \int_{\mathbb{R}} e^{i\xi u} f(u) du$ being the Fourier transform of f and $\{W(t)\}_{t \in \mathbb{R}}$ being a standard Brownian motion.

- The Hurst functional parameter $H : [0, +\infty) \rightarrow (0, 1)$ is a Hölder function with exponent $\beta > \sup_{t \in [0, +\infty)} H(t)$. Subject to this constraint the paths of mBm are almost surely continuous functions.

Theorem 4.1 in [43] gives the covariance function of $\{W_{H(t)}(t)\}_{t \geq 0}$: for $s, t \geq 0$,

$$\text{Cov}(W_{H(s)}(s), W_{H(t)}(t)) = D(H(s), H(t))(s^{H(s)+H(t)} + t^{H(s)+H(t)} - |t - s|^{H(s)+H(t)}), \tag{17}$$

where

$$D(s, t) := \frac{\sqrt{\Gamma(2s + 1)\Gamma(2t + 1) \sin(\pi s) \sin(\pi t)}}{2\Gamma(s + t + 1) \sin(\pi(s + t)/2)}.$$

It is known that the pointwise Hölder exponent (pHe) of $\{W_{H(t)}(t)\}_{t \geq 0}$ is almost surely equal to $H(\bullet)$ at each t [42]. Recall that, for a continuous-time nowhere differentiable process $\{Y(t)\}_{t \geq 0}$, its local Hölder regularity can be measured by the pHe ρ_Y defined by: for each $t_0 \geq 0$,

$$\rho_Y(t_0) := \sup \left\{ \alpha \in [0, 1] : \limsup_{\varepsilon \rightarrow 0^+} \frac{|Y(t_0 + \varepsilon) - Y(t_0)|}{|\varepsilon|^\alpha} = 0 \right\}.$$

For a continuous but undifferentiable function, the pHe measures its “local roughness”: the smaller the pHe is at time t , the more “fractals” should be observed around t in the path. When $H(\bullet) \equiv H$ becomes a constant, mBm reduces to an fBm with Hurst parameter equal to H . More generally, it can be seen from [34] that mBm is locally asymptotically self-similar satisfying Assumption 1. Its tangent process at each t is an fBm $\{B^{(H(t))}(u)\}_u$ with index $H(t)$:

$$\left\{ \frac{W_{H(t+\tau u)}(t + \tau u) - W_{H(t)}(t)}{\tau^{H(t)}} \right\}_u \xrightarrow[\tau \rightarrow 0^+]{\text{f.d.d.}} C_{H(t)} \{B^{(H(t))}(u)\}_u, \tag{18}$$

where $C_{H(t)}$ is a deterministic function only depending on $H(t)$. In the literature, some studies are particularly focused on the statistical inference problems around the pHe of processes and their applications. This study is motivated by modeling using locally asymptotically self-similar processes as well. We refer the readers to [45–50].

As one of the most natural extensions of fBm, mBm has, at present, broad applications. Unlike fBm, mBm allows its Hurst parameter H to change with time. This allows us to model different regimes of the stochastic process with one single model. For example, during a financial crisis, asset volatility may rise significantly, while it is much lower in the peaceful period. Likewise, empirical evidence shows that there have been periods of different volatilities in either exchange rates or interest rates. An fBm (or a self-similar process) is unable to capture the above phenomena. This motivates researchers to introduce mBm into finance as an alternative or improvement of fBm.

The assumption of covariance stationarity inspires us to introduce a covariance-based dissimilarity measure between the sample paths, in order to capture the level of differences between the two corresponding covariance stationary processes. Later we show that the assumption of autocovariance-ergodicity is sufficient for the clustering algorithms to be approximately asymptotically consistent.

3. Clustering Stochastic Processes

3.1. Covariance-Based Dissimilarity Measure between Autocovariance Ergodic Processes

Let Z be a process satisfying Assumption 1. Denote by Y its tangent process (see (10)) and denote by X an increment process of Y , i.e., there is some $u \geq 0$ such that $X(t) = Y(t + u) - Y(u)$ for all $t \geq 0$. In Assumption 1, X is autocovariance ergodic. Since we will show that clustering distinct Z s is approximately asymptotically equivalent to clustering the corresponding increment processes X s, then the dissimilarity measures of Z s can be constructed based on those of the autocovariance ergodic processes X s. From (4) we know that the autocovariance process X is zero-mean. Our first main result is then introduction to the following *covariance-based dissimilarity measure* between autocovariance ergodic processes. We refer the readers to [13] for more features of this dissimilarity measure.

Definition 4. *The covariance-based dissimilarity measure d between the discrete-time stochastic processes $X^{(1)}, X^{(2)}$ (in fact $X^{(1)}, X^{(2)}$ denote two covariance structures, each class may contain different process distributions but they share the same autocovariance function) is defined by*

$$d(X^{(1)}, X^{(2)}) := \sum_{m,l=1}^{+\infty} w_m w_l \rho(\text{Cov}(X_{l\dots l+m-1}^{(1)}), \text{Cov}(X_{l\dots l+m-1}^{(2)})), \tag{19}$$

where:

- For any integers $l \geq 1, m \geq 0, X_{l\dots l+m-1}^{(1)}$ is the shortcut notation of the row vector $(X_l^{(1)}, \dots, X_{l+m-1}^{(1)})$.

- The distance ρ between two equal-sized covariance matrixes M_1, M_2 denotes the Frobenius norm of $M_1 - M_2$. Recall that for a matrix $A_{M \times N}$, its Frobenius norm is defined by $\|A_{M \times N}\|_F := \sqrt{\sum_{i=1}^M \sum_{j=1}^N a_{ij}^2}$, where for each $(i, j) \in \{1, \dots, M\} \times \{1, \dots, N\}$, a_{ij} denotes the (i, j) -coefficient of $A_{M \times N}$.
- The sequence of positive weights $\{w_j\}_{j \geq 1}$ should be chosen such that $d(X^{(1)}, X^{(2)}) < +\infty$, i.e., the series on the right-hand side of Equation (19) is convergent. The choice of $\{w_j\}_j$ will be discussed in the forthcoming simulation study in Section 5.

Remark 1. It is important to note that Definition 4 only defines the dissimilarity measures between discrete-time stochastic processes (time series). This is the most common object studied in the literature on clustering stochastic processes. Studying clustering of time series is sufficient for practical applications for at least two reasons. First, a continuous-time path is not observable in the real world. Second, any (stochastically) continuous-time process can be well approximated by its discrete-time paths. In what follows, we will find a mean sample path by a finite-length subsequence of a discretized stochastic process.

Thanks to the autocovariance-ergodicity of the sample processes, the dissimilarity measure d can be estimated by the empirical dissimilarity measure \hat{d} below:

Definition 5. Given two processes' discrete-time sample paths $\mathbf{x}_j = (X_1^{(j)}, \dots, X_{n_j}^{(j)})$ for $j = 1, 2$, let $n = \min\{n_1, n_2\}$, then the empirical covariance-based dissimilarity measure between \mathbf{x}_1 and \mathbf{x}_2 is given by

$$\hat{d}(\mathbf{x}_1, \mathbf{x}_2) := \sum_{m=1}^{m_n} \sum_{l=1}^{n-m+1} w_m w_l \rho(v(X_{l \dots l+m-1}^{(1)}), v(X_{l \dots l+m-1}^{(2)})), \tag{20}$$

where:

- $m_n (\leq n)$ is the largest dimension of the covariance matrix considered by \hat{d} ; in this framework we take $m_n = \lfloor \log n \rfloor$, i.e., the floor number of $\log n$ [12,13].
- For $j = 1, 2, 1 \leq l \leq n$ and $m \leq n - l + 1$, $v(X_{l \dots l+m-1}^{(j)})$ denotes the empirical covariance matrix of the process $X^{(j)}$'s path $(X_l^{(j)}, \dots, X_{l+m-1}^{(j)})$, which is given below:

$$v(X_{l \dots l+m-1}^{(j)}) := \frac{\sum_{i=l}^{n-m+1} (X_i^{(j)} \dots X_{i+m-1}^{(j)})^T (X_i^{(j)} \dots X_{i+m-1}^{(j)})}{n - m - l + 2}, \tag{21}$$

where $(\bullet)^T$ denotes the transpose of a matrix.

Remark 2. Since X is autocovariance ergodic, every empirical covariance matrix $v(X_{l \dots l+m-1})$ is a consistent estimator of the covariance matrix $\text{Cov}(X_{l \dots l+m-1})$ under Frobenius norm and in probability, i.e.,

$$\|v(X_{l \dots l+m-1}) - \text{Cov}(X_{l \dots l+m-1})\|_F \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \text{ for any } l \geq 0. \tag{22}$$

Furthermore, the fact that both d and \hat{d} satisfy the triangle inequalities implies that \hat{d} is a consistent estimator of d . The proof is quite similar to that of Lemma 1 in [13], except that in the former statement the convergence holds in probability. These ergodicity and triangle inequalities are the keys to demonstrate that our algorithms in the next section are approximately asymptotically consistent. We list them in the following paragraphs.

Remark 3. For every pair of paths

$$\mathbf{x}_1 = (X_1^{(1)}, \dots, X_{n_1}^{(1)}) \quad \text{and} \quad \mathbf{x}_2 = (X_1^{(2)}, \dots, X_{n_2}^{(2)}),$$

sampled from two autocovariance ergodic processes $X^{(1)}$ and $X^{(2)}$, respectively, we have

$$\widehat{d}(\mathbf{x}_1, \mathbf{x}_2) \xrightarrow[n_1, n_2 \rightarrow +\infty]{\mathbb{P}} d(X^{(1)}, X^{(2)}), \tag{23}$$

and

$$\widehat{d}(\mathbf{x}_i, X^{(j)}) \xrightarrow[n_i \rightarrow \infty]{\mathbb{P}} d(X^{(i)}, X^{(j)}), \text{ for } i, j \in \{1, 2\}, \tag{24}$$

where the dissimilarity measure $\widehat{d}(\mathbf{x}_i, X^{(j)})$ between the sample path \mathbf{x}_i and the stochastic process $X^{(j)}$ is defined to be

$$\widehat{d}(\mathbf{x}_i, X^{(j)}) := \widehat{d}(\mathbf{x}_i, (X_1^{(j)}, \dots, X_{n_i}^{(j)})).$$

Remark 4. Thanks to their definitions, the triangle inequalities hold for the covariance-based dissimilarity measure d in (19), as well as for its empirical estimates \widehat{d} in (20). Therefore, for arbitrary processes $X^{(i)}$, $i = 1, 2, 3$ and arbitrary random vectors \mathbf{x}_i , $i = 1, 2, 3$ we have

$$\begin{aligned} d(X^{(1)}, X^{(2)}) &\leq d(X^{(1)}, X^{(3)}) + d(X^{(2)}, X^{(3)}), \\ \widehat{d}(\mathbf{x}_1, \mathbf{x}_2) &\leq \widehat{d}(\mathbf{x}_1, \mathbf{x}_3) + \widehat{d}(\mathbf{x}_2, \mathbf{x}_3), \\ \widehat{d}(\mathbf{x}_1, X^{(1)}) &\leq \widehat{d}(\mathbf{x}_1, X^{(2)}) + d(X^{(1)}, X^{(2)}). \end{aligned}$$

In the next section, we define a proper covariance-based dissimilarity measure between locally asymptotically self-similar processes satisfying Assumption (A), based on the dissimilarity measure d .

3.2. Covariance-Based Dissimilarity Measure between Locally Asymptotically Self-Similar Processes

Now under Assumption (A), we study the asymptotic relationship between the locally asymptotically self-similar process $\{Z_t^{(H(t))}\}_t$ in (10) and its tangent process' increment process. The following result reveals the relationship between local asymptotic self-similarity and covariance stationarity.

Proposition 2. Let $\{Z_t^{(H(t))}\}_{t \geq 0}$ be a locally asymptotically self-similar process satisfying Assumption (A). For each $h > 0$,

$$\left\{ \frac{Z_{t+\tau(u+h)}^{(H(t+\tau(u+h)))} - Z_{t+\tau u}^{(H(t+\tau u))}}{\tau^{H(t)}} \right\}_{u \geq 0} \xrightarrow[\tau \rightarrow 0^+]{f.d.d.} \{X_{u,h}^{(H(t))}\}_{u \geq 0} \tag{25}$$

where $\{X_{u,h}^{(H(t))}\}_{u \geq 0} := \{Y_{u+h}^{(H(t))} - Y_u^{(H(t))}\}_{u \geq 0}$ (see (10)) is an autocovariance ergodic process.

Proof. Let us fix $h > 0$ and pick any finite discrete-time indexes set $T \subset [0, +\infty)$. In Assumption (A), the f.d.d. convergence (10) holds. It then implies

$$\left(\frac{Z_{t+\tau(u+h)}^{(H(t+\tau(u+h)))} - Z_t^{(H(t))}}{\tau^{H(t)}}, \frac{Z_{t+\tau u}^{(H(t+\tau u))} - Z_t^{(H(t))}}{\tau^{H(t)}} \right)_{u \in T} \xrightarrow[\tau \rightarrow 0^+]{\text{law}} (Y_{u+h}^{(H(t))}, Y_u^{(H(t))})_{u \in T}, \tag{26}$$

where we adopt the notation $(a_u, b_u)_{u \in \{u_1, \dots, u_N\}}$ to denote the vector

$$(a_{u_1}, b_{u_1}, a_{u_2}, b_{u_2}, \dots, a_{u_N}, b_{u_N}).$$

It follows from (26) and the continuous mapping theorem that

$$\left(\frac{Z_{t+\tau(u+h)}^{(H(t+\tau(u+h)))} - Z_t^{(H(t))}}{\tau^{H(t)}} - \frac{Z_{t+\tau u}^{(H(t+\tau u))} - Z_t^{(H(t))}}{\tau^{H(t)}} \right)_{u \in T} \xrightarrow[\tau \rightarrow 0^+]{\text{law}} (Y_{u+h}^{(H(t))} - Y_u^{(H(t))})_{u \in T}. \tag{27}$$

(25) then results from (27) and the fact that the choice of T is arbitrary. In Assumption (A), $\{X_{u,h}^{(H(t))}\}_u := \{Y_{u+h}^{(H(t))} - Y_u^{(H(t))}\}_u$ is autocovariance ergodic, hence Proposition 2 is proved. \square

From a statistical point of view, the sequence at the left-hand side of (25) cannot straightforwardly serve to estimate the distribution of the right-hand side $\{X_{u,h}^{(H(t))}\}_u$, because the functional index $H(\bullet)$ at the left-hand side is not observable in practice. To overcome this inconvenience, we note that (25) can be interpreted as follows: when τ is sufficiently small,

$$\left\{ Z_{t+\tau(u+h)}^{(H(t+\tau(u+h)))} - Z_{t+\tau u}^{(H(t+\tau u))} \right\}_{u \in [0, Kh]} \stackrel{\text{f.d.d.}}{\approx} \left\{ \tau^{H(t)} X_{u,h}^{(H(t))} \right\}_{u \in [0, Kh]}' \tag{28}$$

where K is an arbitrary positive integer. Smaller is τ , closer are the two hand sides of (28). However in practice τ has its optimal value, due to the sample size limitation. Therefore, when applied to the real world, (28) says that: given a discrete-time path $Z_{t_1}^{(H(t_1))}, \dots, Z_{t_n}^{(H(t_n))}$ with $t_i = ih\Delta t$ for each $i \in \{1, \dots, n\}$, sampled from a locally asymptotically self-similar process $\{Z_t^{(H(t))}\}_t$, its *localized increment paths* with time index around t_i , i.e.,

$$\mathbf{z}^{(i)} := \left(Z_{t_{i+1}}^{(H(t_{i+1}))} - Z_{t_i}^{(H(t_i))}, \dots, Z_{t_{i+1+K}}^{(H(t_{i+1+K}))} - Z_{t_{i+K}}^{(H(t_{i+K}))} \right), \tag{29}$$

is *approximately* distributed as an autocovariance ergodic increment process of the self-similar process $\{\Delta t^{H(t_i)} X_{u,h}^{(H(t_i))}\}_{u \in [0, Kh]}$. In this example, Δt is the smallest value one can consider for τ in (28). This fact drives us to define the empirical covariance-based dissimilarity measure between two paths of locally asymptotically self-similar processes \mathbf{z}_1 and \mathbf{z}_2 as follows:

$$\widehat{d}^*(\mathbf{z}_1, \mathbf{z}_2) := \frac{1}{L} \sum_{i=1}^L \widehat{d}(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}), \tag{30}$$

where:

- L is chosen from $\{1, \dots, n - K - 1\}$.
- $\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}$ are the localized increment paths defined as in (29). Heuristically speaking, for $i = 1, \dots, n - K - 1$, $\widehat{d}(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)})$ computes the “distance” between the two covariance structures (of the increments of $\{Z_t^{(H(t))}\}_t$) indexed by the time in the neighborhood of t_i , and $\widehat{d}^*(\mathbf{z}_1, \mathbf{z}_2)$ averages the above distances. It is worth noting that the value K describes the “sample size” used to approximate each local distance \widehat{d} . Therefore, its value should be picked neither too large nor too small and it can depend on n . It is suggested that $K \geq 5$ in order that the result of estimating the dissimilarity measure \widehat{d} is acceptable. The largest value one can set for K is $n - 2$ (correspondingly, $L = 1$).

The following observation is straightforward.

Remark 5. Based on the definition (30) and Remark 2, \widehat{d}^* is also a (weakly) consistent estimator of d (see Remark 3) and it also satisfies the triangle inequalities as in Remark 4.

Through employing the dissimilarity measure \widehat{d}^* on locally asymptotically self-similar processes, we obtain the so-called “approximately asymptotically consistent algorithms”, which are introduced in the following section.

4. Approximately Asymptotically Consistent Algorithms

4.1. Offline and Online Algorithms

Please note that the covariance-based dissimilarity measure \widehat{d}^* defined in (30) will aim to cluster covariance structures, not process distributions; therefore the ground truths of

clustering should be those of covariance structures. We thus define the ground truth as follows [13].

Definition 6 (Ground truth of covariance structures). Let $G = \{G_1, \dots, G_\kappa\}$ be a partitioning of \mathbb{N} into κ disjoint sets G_k , $k = 1, \dots, \kappa$, such that the means and covariance structures of \mathbf{x}_i , $i \in \mathbb{N}$ are identical, if and only if $i \in G_k$ for some $k = 1, \dots, \kappa$. Such G is called ground truth of covariance structures. For $N \geq 1$, we denote by $G|_N$ the restriction of G to the first N sequences:

$$G|_N = \{G_k \cap \{1, \dots, N\} : k = 1, \dots, \kappa\}.$$

The processes Z satisfying Assumption 1 are generally not covariance stationary; however, their tangent processes' increments X are covariance stationary. In view of (25) and (28), clustering these processes Z is equivalent to clustering X , based on the covariance structure ground truth of the latter increments. Below we will introduce algorithms aiming to approximate the covariance structure ground truth of X .

Depending on how the information is collected, the processes clustering problems consist of dealing with two separate model settings: the offline and online settings. In the offline setting, the sample size and each path length are time-independent. However, in the online setting, they may both grow with time. As stated in [12], using the offline algorithm in the online setting by simply applying it to the entire data observed at every time step, does not result in an asymptotically consistent algorithm. As a result, we consider clustering offline and online datasets as two approaches and study them separately. Hence the approximated asymptotic consistency will be described in Theorems 2 and 3 below, respectively, for offline and online clustering algorithms. Our offline and online clustering algorithms below are obtained by replacing the dissimilarity measures in Algorithms 1 and 2 in [13] with \hat{d}^* .

For the offline setting, we cluster observed data using Algorithm 1 below. It is a 2-point initialization centroid-based clustering approach. From the distance \hat{d}^* defined in (30), the algorithm picks the farthest two points to be the first two cluster centers (Lines 1–2). Then, iteratively, each following cluster center is chosen to be the point farthest to all the previously assigned cluster centers (Lines 3–5). Finally, the algorithm assigns each remaining point to the nearest cluster (Lines 7–10).

The strategy for clustering online data is presented in Algorithm 2 as follows. At each time t , first update the collection of sample paths $S(t)$ (Lines 1–2); then for each $j = \kappa, \dots, N(t)$, use Algorithm 1 to group the first j paths in $S(t)$ into κ clusters (Lines 6–7); for each cluster the center is selected as the point having the *smallest* index among that cluster, and sort these indexes increasingly (Line 8); calculate the minimum inter-cluster distance γ_j and update the normalization factor η (Lines 9–11). Finally, every sample path in $S(t)$ is assigned to the “nearest” cluster, based on the weighted combination of the distances (given in Line 15) between this observation and the candidate cluster centers obtained at each iteration on j (Lines 14–17).

Algorithm 1: Offline clustering.

Input: *sample paths* $S = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$; *number of clusters* κ .

- 1 $(c_1, c_2) \leftarrow \arg \max_{(i,j) \in \{1, \dots, N\}^2, i < j} \widehat{d}^*(\mathbf{z}_i, \mathbf{z}_j)$;
- 2 $C_1 \leftarrow \{c_1\}; C_2 \leftarrow \{c_2\}$;
- 3 **for** $k = 3, \dots, \kappa$ **do**
- 4 $c_k \leftarrow \arg \max_{i=1, \dots, N} \min_{j=1, \dots, k-1} \widehat{d}^*(\mathbf{z}_i, \mathbf{z}_{c_j})$;
- 5 **end**
- 6 *Assign the remaining points to the nearest centers:*
- 7 **for** $i = 1, \dots, N$ **do**
- 8 $k \leftarrow \arg \min_{k \in \{1, \dots, \kappa\}} \{\widehat{d}^*(\mathbf{z}_i, \mathbf{z}_j) : j \in C_k\}$;
- 9 $C_k \leftarrow C_k \cup \{i\}$;
- 10 **end**

Output: The κ clusters $f(S, \kappa, \widehat{d}^*) = \{C_1, \dots, C_\kappa\}$.

Algorithm 2: Online clustering.

Input: *Sample paths* $\{S(t) = \{\mathbf{z}_1^t, \dots, \mathbf{z}_{N(t)}^t\}\}_t$; *number of clusters* κ ; *a sequence of weights* $\{\beta_j\}_j$.

- 1 **for** $t = 1, \dots, \infty$ **do**
- 2 *Obtain new paths:* $S(t) \leftarrow \{\mathbf{z}_1^t, \dots, \mathbf{z}_{N(t)}^t\}$;
- 3 *Initialize the normalization factor:* $\eta \leftarrow 0$;
- 4 *Initialize the κ clusters:* $C_k(t) \leftarrow \emptyset, k = 1, \dots, \kappa$;
- 5 *Generate $N(t) - \kappa + 1$ candidate cluster centers:*
- 6 **for** $j = \kappa, \dots, N(t)$ **do**
- 7 $\{C_1^j, \dots, C_\kappa^j\} \leftarrow \text{Alg1}(\{\mathbf{z}_1^t, \dots, \mathbf{z}_j^t\}, \kappa, \widehat{d}^*)$;
- 8 $(c_1^j, \dots, c_\kappa^j) \leftarrow \text{sort}(\min \{i \in C_k^j : k = 1, \dots, \kappa\})$; *Choose the smallest index to be the one for cluster center, and sort them increasingly;*
- 9 $\gamma_j \leftarrow \min_{k, k' \in \{1, \dots, \kappa\}, k \neq k'} \widehat{d}^*(\mathbf{z}_{c_k^j}^t, \mathbf{z}_{c_{k'}^j}^t)$;
- 10 $w_j \leftarrow \beta_j$; *w_j is the weight used in \widehat{d}^* ;*
- 11 $\eta \leftarrow \eta + w_j \gamma_j$; *Update the normalization factor;*
- 12 **end**
- 13 *Assign each point to one of the κ clusters:*
- 14 **for** $i = 1, \dots, N(t)$ **do**
- 15 $k \leftarrow \arg \min_{k' \in \{1, \dots, \kappa\}} \frac{1}{\eta} \sum_{j=\kappa}^{N(t)} w_j \gamma_j \widehat{d}^*(\mathbf{z}_i^t, \mathbf{z}_{c_{k'}^j}^t)$;
- 16 $C_k(t) \leftarrow C_k(t) \cup \{i\}$;
- 17 **end**
- 18 **end**

Output: The κ clusters $f(S(t), \kappa, \widehat{d}^*) = \{C_1(t), \dots, C_\kappa(t)\}, t = 1, 2, \dots, \infty$.

4.2. Computational Complexity and Consistency of the Algorithms

We describe the computational complexity based on the number of computations of the distance ρ . For Algorithm 1, the 2-point initialization requires $N(N - 1)/2$ times calculations of \widehat{d}^* . From (30) we see that each calculation of \widehat{d}^* consists of $n_{\min} - K - 1$ times computations of \widehat{d} . By (20), \widehat{d} can be obtained through computing $K - \log K + 1$ times distances ρ . Therefore, the total number of computations of ρ is not greater than $N(N - 1)(n_{\min} - K - 1)(K - \log K + 1)/2$. For Algorithm 2, since at each step $j \in \{\kappa, \dots, N - \kappa + 1\}$, Al-

gorithm 1 is run on j observations, the total number of ρ s computations is then less than $(n_{\min} - K - 1)(K - \log K + 1) \sum_{j=\kappa}^{N-\kappa+1} j(j - 1)/2$. The computational complexity is acceptable in practice, and it is quite competitive compared with existing algorithms for clustering stochastic processes.

Now we introduce the notion of approximately asymptotic consistency. Fix a positive integer K . Let $Z^{(1)}, Z^{(2)}$ be two locally asymptotically self-similar processes with respect functional indexes $H_1(\bullet), H_2(\bullet)$. Furthermore, let

$$\{\mathbf{z}_1^{(1)}, \dots, \mathbf{z}_1^{(n-K-1)}\} \quad \text{and} \quad \{\mathbf{z}_2^{(1)}, \dots, \mathbf{z}_2^{(n-K-1)}\}$$

be, respectively, their sample paths $\mathbf{z}_1, \mathbf{z}_2'$ increments, defined as in (29). For $j = 1, 2$, we define the normalized increments by taking the following linear transformation:

$$\mathcal{H}(\mathbf{z}_j^{(i)}) := \frac{\mathbf{z}_j^{(i)}}{\Delta t^{H_j(t_i)}}, \quad \text{for } i = 1, \dots, n - K - 1. \tag{31}$$

Then using (25) we obtain

$$\mathcal{H}(\mathbf{z}_j^{(i)}) \xrightarrow[\Delta t \rightarrow 0]{\text{law}} (X_{0,j}^{(H_j(t_i))}, X_{h,j}^{(H_j(t_i))}, \dots, X_{Kh,j}^{(H_j(t_i))}), \tag{32}$$

where $(X_{0,j}^{(H_j(t_i))}, X_{h,j}^{(H_j(t_i))}, \dots, X_{Kh,j}^{(H_j(t_i))})$ denotes a discrete-time path of the increment of a self-similar process with self-similarity index $H_j(t_i)$. Fix $L \geq 1$. For each empirical dissimilarity measure $\widehat{d}^*(\mathbf{z}_1, \mathbf{z}_2)$, we correspondingly define

$$\widetilde{d}^*(\mathbf{z}_1, \mathbf{z}_2) := \frac{1}{L} \sum_{i=1}^L \widehat{d}(\mathcal{H}(\mathbf{z}_1^{(i)}), \mathcal{H}(\mathbf{z}_2^{(i)})). \tag{33}$$

$\widetilde{d}^*(\mathbf{z}_1, \mathbf{z}_2)$ is another dissimilarity measure between $\mathbf{z}_1, \mathbf{z}_2$, which has a tight relationship to the distance \widehat{d}^* between their tangent processes' increments. Indeed using (32) and the continuous mapping theorem, it is easy to derive the following result.

Proposition 3. For any N independent sample paths $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N$,

$$(\widetilde{d}^*(\mathbf{z}_i, \mathbf{z}_j))_{i,j \in \{1, \dots, N\}, i \neq j} \xrightarrow[\Delta t \rightarrow 0]{\text{law}} (\widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j))_{i,j \in \{1, \dots, N\}, i \neq j} \tag{34}$$

where $\mathbf{x}_1, \mathbf{x}_2$ are the increments of the tangent processes corresponding to $\mathbf{z}_1, \mathbf{z}_2$, respectively.

In particular, when $\Delta t = 1$, $\widetilde{d}^*(\mathbf{z}_1, \mathbf{z}_2) = \widehat{d}^*(\mathbf{z}_1, \mathbf{z}_2)$. In this sense, $\widetilde{d}^*(\mathbf{z}_1, \mathbf{z}_2)$ "approximates" \widehat{d}^* . \widehat{d}^* cannot be observed in practice since the functional indexes of the locally asymptotically self-similar processes are supposed to be unknown. In what follows, \widetilde{d}^* only serves to define the approximate asymptotic consistency. In this concept, "approximate" means the clustering locally asymptotically self-similar processes problem is "approximately" equivalent to the clustering their tangent processes problem.

Now we state the consistency theorems. Through Theorems 2 and 3, below we show that Algorithms 1 and 2 are both approximately asymptotically consistent. Their proofs are inspired by the ones in [12,13]. However different from the consistency theorems in [12,13], the convergences in Theorems 2 and 3 have weaker senses, which are in probability, not almost sure.

Theorem 2. In Assumption 1, Algorithm 1 is approximately asymptotically consistent for clustering the offline sample paths $S = \{\mathbf{z}_1, \dots, \mathbf{z}_N\}$. This means: if \widehat{d}^* is replaced with \widetilde{d}^* in Algorithm 1, the output clusters converge to the covariance structure ground truths of the incre-

ments of the corresponding tangent processes $S' = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ in probability, as $\Delta t \rightarrow 0$ and $n_{\min} := \min\{n_1, \dots, n_N\} \rightarrow +\infty$. More formally,

$$\lim_{n_{\min} \rightarrow +\infty} \lim_{\Delta t \rightarrow 0} \mathbb{P}(f(S, \kappa, \tilde{d}^*) = G_{S'}) = 1, \tag{35}$$

where f is given in Algorithm 1 and $G_{S'}$ denotes the ground truths of the covariance structures that generate the set of paths S' .

Proof. First, we show

$$\mathbb{P}(f(S, \kappa, \tilde{d}^*) = G) \xrightarrow{\Delta t \rightarrow 0} \mathbb{P}(f(S', \kappa, \hat{d}^*) = G), \tag{36}$$

where $G = \{C_1, \dots, C_\kappa\}$ denotes any κ -partition of $\{1, \dots, N\}$. Since if the estimated clusters $f(S, \kappa, \tilde{d}^*) = G$, the samples in S that are generated by the same cluster in G are closer under \tilde{d}^* to each other than to the rest of the samples. Hence we can write

$$\begin{aligned} &\mathbb{P}(f(S, \kappa, \tilde{d}^*) = G) \\ &= \mathbb{P}\left(\bigcup_{\varepsilon > 0} \left(\left\{\max_{\substack{l \in \{1, \dots, \kappa\} \\ i, j \in C_l}} \tilde{d}^*(\mathbf{z}_i, \mathbf{z}_j) < \varepsilon\right\} \cap \left\{\min_{\substack{k, k' \in \{1, \dots, \kappa\}, k \neq k' \\ i \in C_k, j \in C_{k'}}} \tilde{d}^*(\mathbf{z}_i, \mathbf{z}_j) > \varepsilon\right\}\right)\right). \end{aligned} \tag{37}$$

It follows from (37) and (34) that

$$\begin{aligned} &\lim_{\Delta t \rightarrow 0} \mathbb{P}(f(S, \kappa, \tilde{d}^*) = G) \\ &= \mathbb{P}\left(\bigcup_{\varepsilon > 0} \left(\left\{\max_{\substack{l \in \{1, \dots, \kappa\} \\ i, j \in C_l}} \hat{d}^*(\mathbf{x}_i, \mathbf{x}_j) < \varepsilon\right\} \cap \left\{\min_{\substack{i \in C_k, j \in C_{k'} \\ k, k' \in \{1, \dots, \kappa\} \\ k \neq k'}} \hat{d}^*(\mathbf{x}_i, \mathbf{x}_j) > \varepsilon\right\}\right)\right) \\ &= \mathbb{P}(f(S', \kappa, \hat{d}^*) = G), \end{aligned}$$

which proves (36).

Next we show that Algorithm 1 is asymptotically consistent on clustering S' under \hat{d}^* :

$$\mathbb{P}(f(S', \kappa, \hat{d}^*) = G_{S'}) \xrightarrow{n_{\min} \rightarrow \infty} 1. \tag{38}$$

Denote by

$$G_{S'} := \{G_1, \dots, G_\kappa\}.$$

Fix $\delta > 0$. Let n_{\min} denote the shortest path length in S' :

$$n_{\min} := \min\{n_i : i \in \{1, \dots, N\}\}.$$

Denote by δ_{\min} the minimal non-zero dissimilarity measure between the processes with different covariance structures:

$$\delta_{\min} := \min\{d(X^{(k)}, X^{(k')}) : k, k' \in \{1, \dots, \kappa\}, k \neq k'\}. \tag{39}$$

Fix $\varepsilon \in (0, \delta_{\min}/4)$ and let $\delta > 0$ be arbitrarily small. Since there are a finite number N of samples, by Remark 3 there is n_0 such that for $n_{\min} > n_0$ we have

$$\mathbb{P}\left(\max_{\substack{l \in \{1, \dots, \kappa\} \\ i \in G_l \cap \{1, \dots, N\}}} \hat{d}^*(\mathbf{x}_i, X^{(l)}) > \varepsilon\right) < \delta. \tag{40}$$

On one hand, by applying the triangle inequalities (see Remark 5), we obtain

$$\begin{aligned}
 & \max_{\substack{l \in \{1, \dots, \kappa\} \\ i, j \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j) \\
 & \leq \max_{\substack{l \in \{1, \dots, \kappa\} \\ i, j \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, X^{(l)}) + \max_{\substack{l \in \{1, \dots, \kappa\} \\ i, j \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_j, X^{(l)}) \\
 & = 2 \max_{\substack{l \in \{1, \dots, \kappa\} \\ i \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, X^{(l)}). \tag{41}
 \end{aligned}$$

Then by (41) and the fact that $2\varepsilon < \delta_{\min}/2$, the following inclusion holds:

$$\left\{ \max_{\substack{l \in \{1, \dots, \kappa\} \\ i \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, X^{(l)}) \leq \varepsilon \right\} \subset \left\{ \max_{\substack{l \in \{1, \dots, \kappa\} \\ i, j \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j) \leq 2\varepsilon < \frac{\delta_{\min}}{2} \right\}. \tag{42}$$

On the other hand, by applying the triangle inequalities (see Remark 5) and the fact that $2\varepsilon < \delta_{\min}/2$, we also obtain: if

$$\max_{\substack{l \in \{1, \dots, \kappa\} \\ i \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, X^{(l)}) \leq \varepsilon,$$

then

$$\begin{aligned}
 & \min_{\substack{k, k' \in \{1, \dots, \kappa\}, k \neq k' \\ i \in G_k \cap \{1, \dots, N\} \\ j \in G_{k'} \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j) \\
 & \geq \min_{\substack{k, k' \in \{1, \dots, \kappa\}, k \neq k' \\ i \in G_k \cap \{1, \dots, N\} \\ j \in G_{k'} \cap \{1, \dots, N\}}} \left\{ d(X^{(k)}, X^{(k')}) - \widehat{d}^*(\mathbf{x}_i, X^{(k)}) - \widehat{d}^*(\mathbf{x}_j, X^{(k')}) \right\} \\
 & \geq \delta_{\min} - 2\varepsilon > \frac{\delta_{\min}}{2}.
 \end{aligned}$$

Equivalently,

$$\left\{ \max_{\substack{l \in \{1, \dots, \kappa\} \\ i \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, X^{(l)}) \leq \varepsilon \right\} \subset \left\{ \min_{\substack{k, k' \in \{1, \dots, \kappa\}, k \neq k' \\ i \in G_k \cap \{1, \dots, N\} \\ j \in G_{k'} \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j) > \frac{\delta_{\min}}{2} \right\}. \tag{43}$$

It follows from (42), (43) and (40) that for $n_{\min} > n_0$,

$$\begin{aligned}
 & \mathbb{P} \left(\left\{ \max_{\substack{l \in \{1, \dots, \kappa\} \\ i, j \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j) \geq \frac{\delta_{\min}}{2} \right\} \cup \left\{ \min_{\substack{k, k' \in \{1, \dots, \kappa\}, k \neq k' \\ i \in G_k \cap \{1, \dots, N\} \\ j \in G_{k'} \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j) \leq \frac{\delta_{\min}}{2} \right\} \right) \\
 & \leq \mathbb{P} \left(\max_{\substack{l \in \{1, \dots, \kappa\} \\ i \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, X^{(l)}) > \varepsilon \right) < \delta. \tag{44}
 \end{aligned}$$

Please note that (44) is equivalent to

$$\mathbb{P} \left(\left\{ \max_{\substack{l \in \{1, \dots, \kappa\} \\ i, j \in G_l \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j) < \frac{\delta_{\min}}{2} \right\} \cap \left\{ \min_{\substack{k, k' \in \{1, \dots, \kappa\}, k \neq k' \\ i \in G_k \cap \{1, \dots, N\} \\ j \in G_{k'} \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j) > \frac{\delta_{\min}}{2} \right\} \right) \xrightarrow{n_{\min} \rightarrow +\infty} 1.$$

This tells us that the sample paths in S that are generated by the same covariance structures are closer to each other than to the rest of sample paths. Then, by (44), for $n_{\min} > n_0$, each sample path should be “close” enough to its cluster center, i.e.,

$$\mathbb{P}\left(\max_{i \in \{1, \dots, N\}} \min_{k \in \{1, \dots, \kappa-1\}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_{c_k}) \leq \frac{\delta_{\min}}{2}\right) < \delta, \tag{45}$$

where the κ cluster centers’ indexes c_1, \dots, c_κ are determined by Algorithm 1 in the following way:

$$(c_1, c_2) := \arg \max_{i, j \in \{1, \dots, N\}, i < j} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_j),$$

and

$$c_k := \arg \max_{i \in \{1, \dots, N\}} \min_{j \in \{1, \dots, k-1\}} \widehat{d}^*(\mathbf{x}_i, \mathbf{x}_{c_j}), \quad k = 3, \dots, \kappa.$$

These c_1, \dots, c_κ are chosen to index sample paths generated by different process covariance structures. Then by (44), each remaining sample path will be assigned to the cluster center corresponding to the sample path generated by the same process covariance structure. Finally, (38) results from (44) and (45); and (35) is proved by combining (36) and (38). \square

Below we state the consistency theorem concerning the online clustering algorithm.

Theorem 3. *In Assumption 1, Algorithm 2 is approximately asymptotically consistent for clustering the online sample paths $S(t) = \{\mathbf{z}_1^t, \dots, \mathbf{z}_N^t\}$, $t = 1, 2, \dots$. This means: if \widehat{d}^* is replaced with \widetilde{d}^* in Algorithm 2, for any integer $N \geq 1$, the output clusters of the first N paths in $S(t)$*

$$S(t)|_N := \{\mathbf{z}_1^t, \dots, \mathbf{z}_N^t\},$$

converge to the covariance structure ground truths of the increments of the corresponding tangent processes $S'(t)|_N := \{\mathbf{x}_1^t, \dots, \mathbf{x}_N^t\}$ in probability, as $\Delta t \rightarrow 0$ and $t \rightarrow +\infty$. In other words,

$$\lim_{t \rightarrow +\infty} \lim_{\Delta t \rightarrow 0} \mathbb{P}(f(S(t), \kappa, \widetilde{d}^*)|_N = G_{S'(t)}|_N) = 1, \tag{46}$$

where $f(S(t), \kappa, \widetilde{d}^)|_N$ denotes the clustering $f(S(t), \kappa, \widetilde{d}^*)$ restricted to the first N sample paths in $S(t)$. We also recall that $G_{S'(t)}|_N$ is the restriction of $G_{S'(t)}$ to the first N sample paths $\{\mathbf{x}_1^t, \dots, \mathbf{x}_N^t\}$ in $S'(t)$ (see Definition 6).*

Proof. Let us fix $N \geq 1$. First, similar to the derivation of (36) in the proof of Theorem 2, we can obtain

$$\mathbb{P}(f(S(t), \kappa, \widetilde{d}^*)|_N = G_{S'(t)}|_N) \xrightarrow{\Delta t \rightarrow 0} \mathbb{P}(f(S'(t), \kappa, \widehat{d}^*)|_N = G_{S'(t)}|_N). \tag{47}$$

Then it remains to prove

$$\mathbb{P}(f(S'(t), \kappa, \widehat{d}^*)|_N = G_{S'(t)}|_N) \xrightarrow{t \rightarrow +\infty} 1. \tag{48}$$

In what follows we prove (48).

Let $\delta > 0$ be arbitrarily small. Fix $\varepsilon \in (0, \delta_{\min}/4)$, where δ_{\min} is defined as in (39).

Denote by

$$\delta_{\max} := \max \{d(X^{(k)}, X^{(k')}) : k, k' \in \{1, \dots, \kappa\}\}. \tag{49}$$

For $k \in \{1, \dots, \kappa\}$, denote by s_k the index of the first path in $S'(t)$ sampled from $X^{(k)}$, i.e.,

$$s_k := \min \{i \in G_k \cap \{1, \dots, N(t)\}\}. \tag{50}$$

Please note that s_k does not depend on t but only on k . Then denote by

$$m := \max_{k \in \{1, \dots, \kappa\}} s_k. \tag{51}$$

For $j \geq 1$ denote by $S'(t)|_j$ the first j sample paths in $S(t)$. Then from (51) we see that, $m \leq N(t)$ and $S'(t)|_m$ contains paths sampled from all κ distinct processes (covariance structures). Using the fact that $\sum_{j=1}^{+\infty} w_j < +\infty$, we can find a fixed value $J \geq m$ such that

$$\sum_{j=J+1}^{+\infty} w_j \leq \varepsilon. \tag{52}$$

Recall that in the online setting, the i th sample path's length $n_i(t)$ grows with time t . Therefore, by Remark 5, for every $j \in \{1, \dots, J\}$ there exists some $T_1(j) > 0$ such that

$$\sup_{t \geq T_1(j)} \mathbb{P} \left(\max_{\substack{k \in \{1, \dots, \kappa\} \\ i \in G_k \cap \{1, \dots, j\}}} \widehat{d}^*(\mathbf{x}_i^t, X^{(k)}) > \varepsilon \right) < \delta. \tag{53}$$

Since $J \geq m$, by Theorem 2 for every $j \in \{m, \dots, J\}$ there exists $T_2(j) > 0$ such that $\text{Alg1}(S'(t)|_j, \kappa, \widehat{d}^*)$ is asymptotically consistent for all $t \geq T_2(j)$. Since $N(t)$ is increasing as $t \rightarrow +\infty$, there is $T_3 > 0$ such that $N(t) > J$ for $t \geq T_3$. Let

$$T := \max \left\{ \max_{\substack{i \in \{1, 2\} \\ j \in \{1, \dots, J\}}} T_i(j), T_3 \right\}.$$

From Algorithm 2 (Lines 9, 11) we see

$$\eta^t := \sum_{j=1}^{N(t)} w_j \gamma_j^t, \quad \text{with} \quad \gamma_j^t := \min_{\substack{k, k' \in \{1, \dots, \kappa\} \\ k \neq k'}} \widehat{d}^*(\mathbf{x}_{c_k}^t, \mathbf{x}_{c_{k'}}^t). \tag{54}$$

Below we provide upper bounds in probability of η^t and γ_j^t .

Upper bound of γ_j^t : Similar to how (43) is derived, we use the triangle inequalities (Remark 5) and (39) to obtain:

$$\begin{aligned} & \sup_{t \geq T} \mathbb{P} \left(\min_{j \in \{1, \dots, N(t)\}} \gamma_j^t < \frac{\delta_{\min}}{2} \right) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\min_{\substack{j \in \{1, \dots, N(t)\} \\ k, k' \in \{1, \dots, \kappa\} \\ k \neq k'}} (d(X^{(k)}, X^{(k')}) - 2\widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)})) < \frac{\delta_{\min}}{2} \right) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\max_{\substack{j \in \{1, \dots, N(t)\} \\ k \in \{1, \dots, \kappa\}}} \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \frac{\delta_{\min}}{4} \right). \end{aligned} \tag{55}$$

Since the clusters are sorted in the order of appearance of the distinct process covariance structures, we have $\mathbf{x}_{c_k}^t = \mathbf{x}_{s_k}^t$ for all $j \geq m$ and $k \in \{1, \dots, \kappa\}$, where we recall that the index s_k is defined in (50). It follows from (55), the fact that $\varepsilon < \delta_{\min}/4$ and (53) that

$$\sup_{t \geq T} \mathbb{P} \left(\min_{j \in \{1, \dots, N(t)\}} \gamma_j^t < \frac{\delta_{\min}}{2} \right) \leq \sup_{t \geq T} \mathbb{P} \left(\max_{\substack{j \in \{1, \dots, m\} \\ k \in \{1, \dots, \kappa\}}} \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \right) < m\delta. \tag{56}$$

For $j \in \{1, \dots, N(t)\}$, by (54), the triangle inequality, (49) and (56), we have

$$\begin{aligned} & \sup_{t \geq T} \mathbb{P} \left(\max_{j \in \{1, \dots, N(t)\}} \gamma_j^t > \delta_{\max} + 2\varepsilon \right) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\max_{\substack{j \in \{1, \dots, N(t)\} \\ k, k' \in \{1, \dots, \kappa\} \\ k \neq k'}} (d(X^{(k)}, X^{(k')}) + 2\widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)})) > \delta_{\max} + 2\varepsilon \right) \quad (57) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\max_{\substack{j \in \{1, \dots, m\} \\ k \in \{1, \dots, \kappa\}}} \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \right) < m\delta. \end{aligned}$$

Upper bound of η^t : By (56) and the fact that $\sum_{j=1}^{N(t)} w_j \geq w_m$, we have

$$\begin{aligned} & \sup_{t \geq T} \mathbb{P} \left(\eta^t < \frac{w_m \delta_{\min}}{2} \right) \leq \sup_{t \geq T} \mathbb{P} \left(\min_{j \in \{1, \dots, N(t)\}} \gamma_j^t \sum_{j=1}^{N(t)} w_j < \frac{w_m \delta_{\min}}{2} \right) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\min_{j \in \{1, \dots, N(t)\}} \gamma_j^t < \frac{\delta_{\min}}{2} \right) < m\delta. \quad (58) \end{aligned}$$

Recall that $N(t) > J$ for $t \geq T$. Therefore for every $k \in \{1, \dots, \kappa\}$ we can write

$$\begin{aligned} \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) &= \frac{1}{\eta^t} \sum_{j=1}^{m-1} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) \\ &+ \frac{1}{\eta^t} \sum_{j=m}^J w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) + \frac{1}{\eta^t} \sum_{j=J+1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}). \quad (59) \end{aligned}$$

Now we provide upper bounds in probability of the three terms on the right-hand side of (59).

Upper bound of the first term: Using (53) and the fact that $(\eta^t)^{-1} \sum_{j=1}^{m-1} w_j \gamma_j^t \leq 1$, we obtain

$$\begin{aligned} & \sup_{t \geq T} \mathbb{P} \left(\frac{1}{\eta^t} \sum_{j=1}^{m-1} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \right) \leq \sup_{t \geq T} \mathbb{P} \left(\max_{\substack{j \in \{1, \dots, m-1\} \\ k \in \{1, \dots, \kappa\}}} \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \right) \\ & \leq (m-1)\delta. \quad (60) \end{aligned}$$

Upper bound of the second term: Recall that $\mathbf{x}_{c_k}^t = \mathbf{x}_{s_k}^t$ for all $j \in \{m, \dots, J\}$ and $k \in \{1, \dots, \kappa\}$. Therefore, by (53) and the fact that $(\eta^t)^{-1} \sum_{j=m}^J w_j \gamma_j^t \leq 1$, for every $k \in \{1, \dots, \kappa\}$ we have

$$\begin{aligned} & \sup_{t \geq T} \mathbb{P} \left(\frac{1}{\eta^t} \sum_{j=m}^J w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \right) = \sup_{t \geq T} \mathbb{P} \left(\widehat{d}^*(\mathbf{x}_{s_k}^t, X^{(k)}) \frac{1}{\eta^t} \sum_{j=m}^J w_j \gamma_j^t > \varepsilon \right) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\widehat{d}^*(\mathbf{x}_{s_k}^t, X^{(k)}) > \varepsilon \right) < \delta. \quad (61) \end{aligned}$$

Upper bound of the third term: By (52), (58) and (57),

$$\begin{aligned} & \sup_{t \geq T} \mathbb{P} \left(\frac{1}{\eta^t} \sum_{j=J+1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \frac{2\varepsilon^2(\delta_{\max} + 2\varepsilon)}{w_m \delta_{\min}} \right) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\max_{\substack{j \in \{1, \dots, N(t)\} \\ k \in \{1, \dots, \kappa\}}} \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \right) \\ & \quad + \sup_{t \geq T} \mathbb{P} \left(\max_{j \in \{1, \dots, N(t)\}} \gamma_j^t > \delta_{\max} + 2\varepsilon \right) \\ & < 2m\delta. \end{aligned} \tag{62}$$

Combining (59)–(62) we obtain, for $k \in \{1, \dots, \kappa\}$,

$$\sup_{t \geq T} \mathbb{P} \left(\frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \left(2 + \frac{2\varepsilon^2(\delta_{\max} + 2\varepsilon)}{w_m \delta_{\min}} \right) \right) < 3m\delta. \tag{63}$$

Now we explain how to use (63) to prove the asymptotic consistency of Algorithm 2. Consider an index $i \in G_{k'}$ for some $k' \in \{1, \dots, \kappa\}$. On one hand, using the triangle inequalities, we obtain for $k \in \{1, \dots, \kappa\}$, $k \neq k'$,

$$\begin{aligned} \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_i^t, \mathbf{x}_{c_k}^t) & \geq \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t d^*(X^{(k)}, X^{(k')}) \\ & \quad - \left(\frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_i^t, X^{(k')}) - \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) \right) \\ & \geq \delta_{\min} - \widehat{d}^*(\mathbf{x}_i^t, X^{(k')}) + \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}). \end{aligned}$$

Then applying (53) and (63) we obtain

$$\begin{aligned} & \sup_{t \geq T} \mathbb{P} \left(\frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_i^t, \mathbf{x}_{c_k}^t) < \delta_{\min} - \varepsilon \left(3 + \frac{2\varepsilon^2(\delta_{\max} + 2\varepsilon)}{w_m \delta_{\min}} \right) \right) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\widehat{d}^*(\mathbf{x}_i^t, X^{(k')}) + \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \left(3 + \frac{2\varepsilon^2(\delta_{\max} + 2\varepsilon)}{w_m \delta_{\min}} \right) \right) \\ & \leq \sup_{t \geq T} \mathbb{P} \left(\widehat{d}^*(\mathbf{x}_i^t, X^{(k')}) > \varepsilon \right) \\ & \quad + \sup_{t \geq T} \mathbb{P} \left(\frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_k}^t, X^{(k)}) > \varepsilon \left(2 + \frac{2\varepsilon^2(\delta_{\max} + 2\varepsilon)}{w_m \delta_{\min}} \right) \right) \\ & < (3m + 1)\delta. \end{aligned} \tag{64}$$

On the other hand, from (53) we see for any $N \geq 1$,

$$\sup_{t \geq T_1(N)} \mathbb{P} \left(\max_{\substack{k \in \{1, \dots, \kappa\} \\ i \in G_k \cap \{1, \dots, N\}}} \widehat{d}^*(\mathbf{x}_i^t, X^{(k)}) > \varepsilon \right) < \delta. \tag{65}$$

Using again the triangle inequalities, we obtain

$$\begin{aligned}
 & \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_i^t, \mathbf{x}_{c_{j'}^t}^t) \\
 & \leq \left(\frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \right) \widehat{d}^*(\mathbf{x}_i^t, X^{(k')}) + \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_{j'}^t}^t, X^{(k')}) \\
 & \leq \widehat{d}^*(\mathbf{x}_i^t, X^{(k')}) + \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_{j'}^t}^t, X^{(k')}).
 \end{aligned} \tag{66}$$

Let $T' := \max\{T, T_1(N)\}$. It results from (66), (65) and (63) that

$$\begin{aligned}
 & \sup_{t \geq T'} \mathbb{P} \left(\frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_i^t, \mathbf{x}_{c_{j'}^t}^t) > \varepsilon \left(3 + \frac{2\varepsilon^2(\delta_{\max} + 2\varepsilon)}{w_m \delta_{\min}} \right) \right) \\
 & \leq \sup_{t \geq T'} \mathbb{P} \left(\widehat{d}^*(\mathbf{x}_i^t, X^{(k')}) + \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_{j'}^t}^t, X^{(k')}) > \varepsilon \left(3 + \frac{2\varepsilon^2(\delta_{\max} + 2\varepsilon)}{w_m \delta_{\min}} \right) \right) \\
 & \leq \sup_{t \geq T'} \mathbb{P} \left(\widehat{d}^*(\mathbf{x}_i^t, X^{(k')}) > \varepsilon \right) \\
 & \quad + \sup_{t \geq T'} \mathbb{P} \left(\frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_{c_{j'}^t}^t, X^{(k')}) > \varepsilon \left(2 + \frac{2\varepsilon^2(\delta_{\max} + 2\varepsilon)}{w_m \delta_{\min}} \right) \right) \\
 & < (3m + 1)\delta.
 \end{aligned} \tag{67}$$

Since δ and ε can be chosen arbitrarily small, it follows from (64) and (67) that

$$\mathbb{P} \left(\arg \min_{k \in \{1, \dots, K\}} \frac{1}{\eta^t} \sum_{j=1}^{N(t)} w_j \gamma_j^t \widehat{d}^*(\mathbf{x}_i^t, \mathbf{x}_{c_k^t}^t) = k' \right) \xrightarrow{t \rightarrow +\infty} 1, \tag{68}$$

for all $i \in \{1, \dots, N\}$. (48) as well as Theorem 3 is proved. \square

5. Tests on Simulated Data: Clustering Multifractional Brownian Motions

5.1. Efficiency Improvement: \log^* -Transformation

In this section, we show the performance of the proposed clustering approaches (Algorithms 1 and 2) on clustering simulated mBMs given in Definition 3. Recall that mBm is a paradigmatic example of locally asymptotically self-similar processes. Its tangent process is an fBm, which is self-similar. Since the covariance structure of the fBm is nonlinearly dependent on its self-similarity index, we can then apply the so-called \log^* -transformation to the covariance matrixes of its increments, in order to improve the efficiency of the clustering algorithms. More precisely, in our clustering algorithms, we replace in \widehat{d}^* the coefficients of all the covariance matrixes and their estimators with their \log^* -transformations. For $x \in \mathbb{R}$, its \log^* -transformation is defined to be

$$\log^*(x) := \begin{cases} \log x, & \text{if } x > 0; \\ -\log(-x), & \text{if } x < 0; \\ 0, & \text{if } x = 0. \end{cases} \tag{69}$$

By applying such transformation, the observations assigned to any two clusters by the covariance structure ground truths become well separated thus the clustering algorithms become more efficient. For more detail on this efficiency enhancement approach, we refer the reader to Section 3 in [13].

5.2. Simulation Methodology

We select Wood–Chan’s method [51,52] to simulate mBm, and use the implementation (MATLAB) of Wood–Chan’s method in *FracLab* (version 2.2) by INRIA in our simulation study (<https://project.inria.fr/fraclab/download/overview/>, accessed on 1 January 2022). This method outputs independent sample paths of the following form:

$$\left\{ W_{H(i/n)}\left(\frac{i}{n}\right) \right\}_{i=0,1,\dots,n'} \tag{70}$$

where $n \geq 1$ is an input integer. Now we would select $w_j = 1/(j^2(j+1)^2)$ so that d (see (19)) is a convergent series (well-defined). To show that this choice is reasonable, we consider the stochastic process

$$\left\{ W_{H(i\Delta)}(i\Delta) \right\}_{i=0,1,\dots'} \tag{71}$$

where $\Delta > 0$ is some given mesh. For each $t_0 = 0, \Delta, 2\Delta, \dots$, the increments of the tangent process (see the right-hand side of (28))

$$\left\{ C_{H(t_0)} n^{-H(t_0)} B^{(H(t_0))}(i\Delta) \right\}_{i=0,1,\dots}$$

is given by: for $i = 0, 1, \dots$,

$$X^{(H(t_0))}(i\Delta) = C_{H(t_0)} n^{-H(t_0)} (B^{(H(t_0))}((i+1)\Delta) - B^{(H(t_0))}(i\Delta)).$$

As increments of fBm, $X^{(H(t_0))}(\bullet)$ is autocovariance ergodic. Moreover for $i, j = 0, 1, \dots$, we have, using the definition of \log^* (see (69)), the covariance function of fBm (see (5)) and the fact that $\sup_{s \geq 0} H(s) \leq 1$,

$$\begin{aligned} & \log^* (\text{Cov}(X^{(H(t_0))}(i\Delta), X^{(H(t_0))}(j\Delta))) \\ &= \log^* \left(\frac{C_{H(t_0)}^2 \Delta^{2H(t_0)}}{2} (|i-j-1|^{2H(t_0)} + |i-j+1|^{2H(t_0)} - 2|i-j|^{2H(t_0)}) \right) \tag{72} \\ &= \mathcal{O}(\log(|i-j|+1)), \text{ as } |i-j| \rightarrow +\infty. \end{aligned}$$

From the definition of d in (19) and (72) we can see that, by taking $w_j = 1/(j^2(j+1)^2)$ and using (72), for any $t_0, t'_0 = 0, \Delta, 2\Delta, \dots$,

$$d(X^{(H(t_0))}, X^{(H(t'_0))}) = \mathcal{O} \left(\sum_{l,m=1}^{+\infty} \frac{\sum_{i,j=l}^{l+m-1} \log(|i-j|+1)}{l^2(l+1)^2 m^2(m+1)^2} \right),$$

where

$$\begin{aligned} & \sum_{l,m=1}^{+\infty} \frac{\sum_{i,j=l}^{l+m-1} \log(|i-j|+1)}{l^2(l+1)^2 m^2(m+1)^2} = 2 \sum_{l,m=1}^{+\infty} \frac{\sum_{k=1}^{m-1} (m-k) \log(k+1)}{l^2(l+1)^2 m^2(m+1)^2} \\ & \leq 2 \sum_{l,m=1}^{+\infty} \frac{\log m}{l^2(l+1)^2(m+1)^2} < +\infty. \end{aligned}$$

Therefore, we showed that the choice $w_j = 1/(j^2(j+1)^2)$ makes d well-defined.

5.3. Synthetic Datasets

To construct a collection of the mBm paths with distinct functional indexes $H(\bullet)$, we set the function form of $H(\bullet)$ in each of the predetermined clusters. Two functional forms of $H(\bullet)$ are selected for synthetic data study:

- **Case 1 (Monotonic function):** The general form is taken to be

$$H(t) = 0.5 + h \cdot t/Q, \quad t \in [0, Q], \tag{73}$$

where $Q > 0$ is a fixed integer and different values of h correspond to different clusters. We then predetermine five clusters with various h s to separate different clusters. In this study, we set $Q = 100, h_1 = -0.4, h_2 = -0.2, h_3 = 0, h_4 = 0.2$ and $h_5 = 0.4$. The trajectories of the five functional forms of $H(\bullet)$ in different clusters are illustrated in the left graph of Figure 1.

- **Case 2 (Periodic function):** The general form is taken to be

$$H(t) = 0.5 + h \cdot \sin(\pi t/Q), \quad t \in [0, Q], \tag{74}$$

where different values of h lead to different clusters. Specifically, we take $Q = 100, h_1 = 0.4, h_2 = 0.2, h_3 = 0, h_4 = -0.2$ and $h_5 = -0.4$. The trajectories of the corresponding five functional forms of $H(\bullet)$ are illustrated in the left graph of Figure 2.

- **Case 3 (Small turbulence on H):** We proceed to examine if the proposed algorithm has the capacity of distinguishing processes with very similar behaviors. To this end we take consider clustering an fBm with Hurst parameter H_f , and an mBm with the Hurst functional parameter

$$H(t) = H_f + 0.1 \cdot \sin(\pi t/Q), \quad t \in [0, Q], \tag{75}$$

We perform the tests using two values of H_f : (i) $H_f = 0.2$ and (ii) $H_f = 0.8$. In both cases, the index $H(t)$ of the corresponding mBm is regarded to be H_f plus some noise.

We demonstrate the approximated asymptotic consistency of the proposed algorithms by conducting both offline and online clustering analysis. We denote the number of observed data points in each time series by $n(t)$, and denote the number of time series paths by $N(t)$.

In the offline setting, the number of observed paths does not depend on time t ; however, the lengths do. To construct offline datasets, we perform the following:

1. For $i = 1, \dots, 5$, simulate 20 mBm paths in group i (corresponding to h_i), each path is with length of 305. Then the total number of paths $N = 100$. To be more explicit we denote by

$$S := \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,305} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,305} \\ \vdots & \vdots & \ddots & \vdots \\ x_{100,1} & x_{100,2} & \cdots & x_{100,305} \end{pmatrix}, \tag{76}$$

where each row is an mBm discrete-time path. For $i = 1, \dots, 5$, the data from the i th group are given as:

$$S^{(i)} := \begin{pmatrix} x_{20(i-1)+1,1} & x_{20(i-1)+1,2} & \cdots & x_{20(i-1)+1,305} \\ \vdots & \vdots & \ddots & \vdots \\ x_{20i,1} & x_{20i,2} & \cdots & x_{20i,305} \end{pmatrix}. \tag{77}$$

2. At each $t = 1, \dots, 100$, we observe the first $n(t) = 3t + 5$ values of each path, i.e.,

$$S_{\text{offline}}(t) := \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,3t+5} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,3t+5} \\ \vdots & \vdots & \ddots & \vdots \\ x_{100,1} & x_{100,2} & \cdots & x_{100,3t+5} \end{pmatrix}.$$

The online dataset does not require observed paths to be of equal length, and can be regarded as an extension of the offline case. Introducing the online dataset aims at mimicking the situation where new time series are observed as time goes. In our simulation study, we use the following method to construct online datasets:

1. For $i = 1, \dots, 5$, simulate 20 mBm paths in group i (corresponding to h_i), each path is with length of 305 (see (76) and (77)).
2. At each $t = 1, \dots, 100$ and $i = 1, \dots, 5$, we observe the following dataset in the i th group:

$$S_{\text{online}}^{(i)}(t) := \begin{pmatrix} \tilde{x}_{1,1} & \tilde{x}_{1,2} & \cdots & \cdots & \cdots & \cdots & \tilde{x}_{1,n_1(t)} \\ \tilde{x}_{2,1} & \tilde{x}_{2,2} & \cdots & \cdots & \cdots & \tilde{x}_{2,n_2(t)} & \\ \vdots & \vdots & \ddots & \vdots & & & \\ \tilde{x}_{N_i(t),1} & \tilde{x}_{N_i(t),2} & \cdots & \tilde{x}_{N_i(t),n_{N_i(t)}(t)} & & & \end{pmatrix},$$

where

- $\tilde{x}_{k,l}$ s are the (k, l) -coefficients in $S^{(i)}$ given in (77).
- $N_i(t) := 6 + \lfloor (t-1)/10 \rfloor$ denotes the number of paths in the i th group. Here $\lfloor \bullet \rfloor$ denotes the floor number. That is, starting from 6 paths in each group, 1 new path will be added into each group as t increases by 10.
- $n_l(t) := 3(t - (l-6)^+)^+ + 5$, with $(\bullet)^+ := \max(\bullet, 0)$. This means each path observes three new values as t increases by 1.

Since at each time t , the covariance structure ground truth being known, we can then evaluate the clustering performance in terms of the so-called “misclassification rate”, which is calculated based on averaging the proportion of mis-clustered paths in each scenario. For more detail on this notion of misclassification rate we refer the readers to Section 4 in [13].

5.4. Experimental Results

We demonstrate the asymptotic consistency of our clustering algorithms by computing the misclassification rates using simulated offline and online datasets. Below we summarize the simulation study results.

Case 1 (Monotonic function):

When $H(\bullet)$ s are chosen to be give monotonic functionals of the form (73) (see the left graph in Figure 1), the right graph in Figure 1 illustrates the behavior of the misclassification rates corresponding to Algorithm 1 applied to offline data setting (solid line), and Algorithm 2 applied to online data setting (dashed line). From this result we observe the following:

- (1) Both algorithms attempt to be consistent in their circumstances, as the time t increases, in the sense that the corresponding misclassification rates decrease to 0.
- (2) Clustering mBms are asymptotically equivalent to clustering their tangent processes' increments.
- (3) The online algorithm seems to have an overall better performance: its misclassification rates are 5–10% lower than that of offline algorithm. The reason may be that at early time steps the differences among the $H(\bullet)$ s are not significant. Unlike the offline clustering algorithm, the online one is flexible enough to catch these small differences.

Case 2 (Periodic function):

The same converging behaviors are found in the case of periodic functional form of $H(\bullet)$ as specified in (74). Their trajectories are illustrated in the left graph of Figure 2. The clustering performance shown in the right graph of Figure 2 indicates the following:

- (1) Both misclassification rates of the clustering algorithms have generally a declining trend as time increases.

- (2) As the differences among the periodic function $H(\bullet)$ s values go up and down, the misclassification rates go down and up accordingly.
- (3) The online clustering algorithm has an overall worse performance than the offline one. This may be because starting from $t = 20$ the differences among $H(\bullet)$ s become significantly large. In this situation, the offline clustering algorithm can better catch these differences, since it has a larger sample size (20 paths in each group) than the online one.

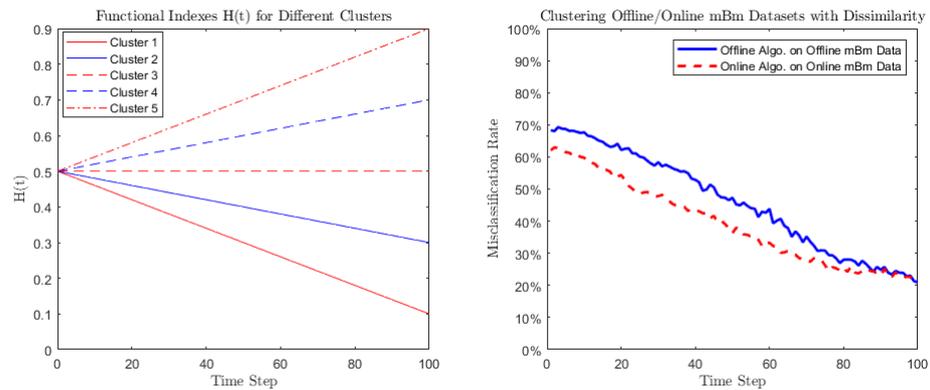


Figure 1. The functional form of $H(\bullet)$ follows Equation (73): $H(t) = 0.5 + h_i \cdot t/100$ with $t = 0, 1, \dots, 100$. The left graph plots $H(\bullet)$ correspond to 5 different clusters. The right graph illustrates the misclassification rates output by (i) offline algorithm on offline dataset (solid line) and (ii) online algorithm on online dataset (dashed line). Both algorithms are performed based on the \log^* -transformed covariance-based dissimilarity measure.

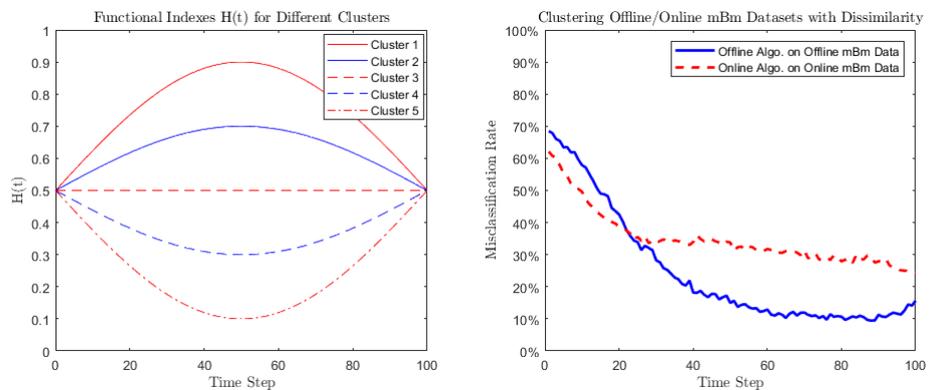


Figure 2. The functional form of $H(\bullet)$ follows Equation (74): $H(t) = 0.5 + h_i \cdot \sin(\pi t/100)$ with $t = 0, 1, \dots, 100$. The left graph plots $H(\bullet)$ corresponding to 5 different clusters. The right graph illustrates the misclassification rates output by (i) offline algorithm on offline dataset (solid line) and (ii) online algorithm on online dataset (dashed line). Both algorithms are performed based on the \log^* -transformed covariance-based dissimilarity measure.

Case 3 (Small turbulence on H):

We continue the simulation analysis using the case where we intend to distinguish the classical fBm process with the mBm process with similar function H . We take two values of H_f (as a constant) for the fBm: (i) $H_f = 0.2$ and (ii) $H_f = 0.8$. In both cases, the corresponding mBm has the H function slightly different with H_f . The functional form of H for mBm is specified in Equation (75).

The clustering results are presented in Figure 3. We find general convergence performance in both cases (for $H_f = 0.2, 0.8$). More specifically, the online algorithm outperforms the offline algorithm for both datasets. The misclassification rate is lower than 20% when $t \geq 40$ for the online algorithm results. In both cases, as the difference between the two

Hurst parameters goes up and down, we see the corresponding misclassification rates go down and up, which is in line with the consistency of our clustering algorithms. When $H_f = 0.8$, the percent changes between the Hurst parameters of fBm and mBm are smaller than those for $H_f = 0.2$, therefore the performance of clustering presented in the right graph is slightly worse. In conclusion, the proposed algorithms demonstrate a robust convergence feature when clustering fBm and mBm with similar Hurst parameters.

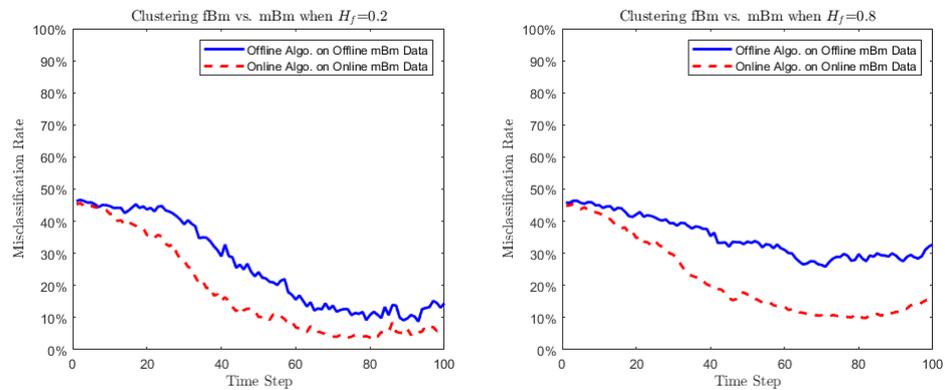


Figure 3. The clustering performance using fBm with H_f and mBm with $H(\bullet)$ given in Equation (74): $H(t) = H_f + 0.1 \cdot \sin(\pi t/100)$ with $t = 0, 1, \dots, 100$. The left graph is the case where $H_f = 0.2$, and the right graph is the case where $H_f = 0.8$. Both graphs illustrate the misclassification rates output by (i) offline algorithm on offline dataset (solid line) and (ii) online algorithm on online dataset (dashed line). Both algorithms are performed based on the \log^* -transformed covariance-based dissimilarity measure.

Please note that for each pair of paths with length $n(t)$, we took $K = n(t) - 2$ and $L = 1$ in \hat{d}^* ; however, any other value of K could be taken. We have provided easily readable and editable MATLAB codes of the proposed algorithms and simulation study replications. All the codes used in this section can be found publicly online (https://github.com/researchcoding/clustering_locally_asymptotically_self_similar_processes/, accessed on 1 January 2022).

5.5. Comparison to Traditional Approaches Designed for Clustering Finite-Dimensional Vectors

In this section, we use the benchmark methods to cluster the simulated dataset proposed in Section 5.3 and compare them to our novel approach. For the specification of the benchmark method, we select K -means and hierarchical clustering with Euclidean distance, which are commonly used in clustering analysis of non-stochastic vectors or processes. First of all, performing both approaches is more time-consuming than our clustering algorithms. More importantly, the misclassification rates from these two traditional algorithms are plotted in Figures 4 and 5.

As opposed to the performance shown in Figures 1 and 2, the benchmark algorithms do not converge when clustering locally asymptotically self-similar processes. Though the numbers of clusters are clearly predetermined in both algorithms, none of the benchmark algorithm demonstrates declining misclassification rate with the increased number of observations. Figure 4 shows the diverging performance of the clustering algorithm with an increased then flat misclassification over time. Figure 5 shows almost flat misclassification rates. None of the results shows a misclassification rate lower than 60%.

From the simulation analysis, we conclude that the conventional approaches have failed to cluster the locally asymptotically self-similar processes (with different functional forms of $H(\bullet)$) in this particular case. The proposed methods in this paper are tailored to cluster the locally asymptotically self-similar stochastic processes with known number of clusters.

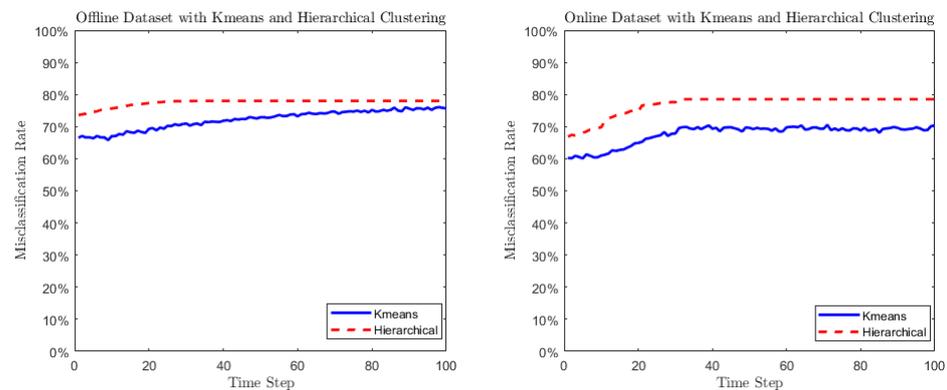


Figure 4. The functional form of $H(\bullet)$ follows Equation (73): $H(t) = 0.5 + h_i \cdot t/100$ with $t = 0, 1, \dots, 100$. The left graph plots the misclassification rates output by K -means and hierarchical clustering methods using the offline dataset. The right graph shows the misclassification rates using the online dataset.

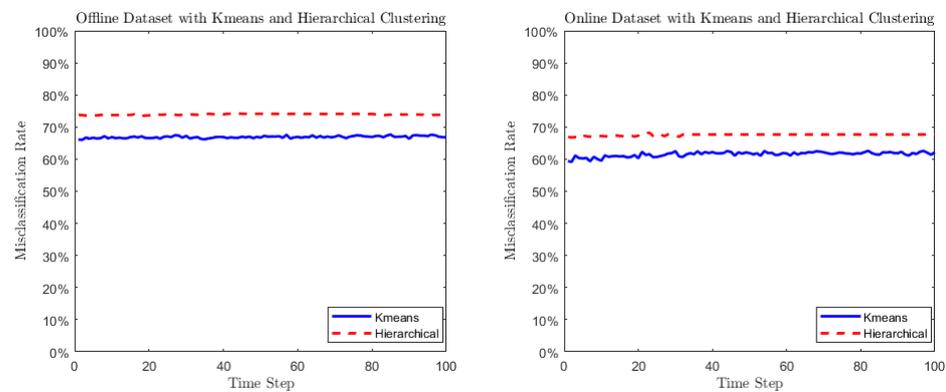


Figure 5. The functional form of $H(\bullet)$ follows Equation (74): $H(t) = 0.5 + h_i \cdot \sin(\pi t/100)$ with $t = 0, 1, \dots, 100$. The left graph plots the misclassification rates output by K -means and hierarchical clustering methods using the offline dataset. The right graph shows the misclassification rates using the online dataset.

6. Real World Application: Clustering Global Financial Markets

6.1. Motivation

In this section, we apply the proposed clustering algorithm on the real world datasets. We select a global equity return data to perform the clustering analysis. Conventionally, stock returns of countries in the same region are considered to have similar patterns due to the common regional economic factors. In other words, economic entities in close geographical distance have potentially more trades and other economic ties. Therefore, these countries are influenced by similar economic and finance factors. However, recent empirical evidence regarding the financial markets show that globalization has broken the geographical barrier and creates new economic factors in contrast to the geographical ones. As a result, global economic clusters switch from “geographical centriods” to “economic development centriods”. Emerging markets demonstrate increasingly similar financial market patterns and correlations [53], whereas developed economic entities share increasing financial market similarity [54].

Bianchi and Pianese [15] pioneered the stock returns modeling using locally asymptotically self-similar processes (e.g., mBm process). Bianchi et al. [18] and Peng and Zhao [27] provide empirical support on this time-varying self-similar feature of the financial time series. Given this fact, the financial market index is a perfect example of stochastic processes, where our proposed clustering algorithms are applicable. We examine the connection of global financial markets by answering as to whether economic entities are well-behaved and clustered by geographical distribution or by development level.

6.2. Data and Methodology

In empirical cluster analysis, we use two asset classes from global financial markets: the country-level stock index and the sovereign credit default swap spread. The *stock index return* captures the upside (growth) characteristics, and the *sovereign CDS spread* captures the downside (credit risk) characteristics of underlying economic entities. We provide a more detailed description below.

- **Equity indexes returns:** We cluster the global stock indexes based on their empirical time-varying covariance structure. We use Algorithms 1 and 2 as the clustering approach. We select the index constituents of MSCI ACWI (All Country World Index) as the underlying stochastic processes in the datasets for clustering analysis. Each of the indexes is a realized path representing the historical monthly returns of the underlying economic entities. MSCI ACWI is the leading global equity market index and covers more than 85% of the market capitalization of the global stock market (As of December 2018, as reported on <https://www.msci.com/acwi>, accessed on 1 January 2022).
- **Sovereign CDS spreads:** We cluster the sovereign credit default swap (CDS) spreads of global economic entities. The sovereign CDS is an insurance-like financial product that provides default protection of treasury bonds for the economic entity (e.g., the government). The CDS spread reflects the cost to insurer of the exposure on a sovereign entity's default. We select a five-year sovereign CDS spread as the indicator of sovereign credit risk, as the five-year product has the best liquidity on the CDS market. We overlap the sample of economic entities between the stock and CDS datasets, and the same set of underlying economics entities are present in the clustering analysis. Our CDS data source is Bloomberg.

The extant literature [15,18,27] shows that these types of financial time series demonstrate a “long memory” path feature. These stochastic processes are naturally modeled as locally asymptotically self-similar processes such as fBm and mBm. In the fBm and mBm cases, the proposed algorithms are applicable to cluster the stock returns and CDS spreads. Similar to Section 5, we cluster the increments of the indexes returns with the log*-transformed dissimilarity measure. The data sample consists of stock indexes and sovereign CDS spreads including 23 developed economic entities and 24 emerging markets. The detailed constituents are present in Table 1 from the data source at <https://www.msci.com/acwi>, accessed on 1 January 2022. The geographical regions include the Americas, EMEA (Europe, Middle East and Africa), the Pacific, and Asia.

We construct both offline and online datasets using monthly returns of stock indexes and monthly spreads of sovereign CDS. The offline dataset of the monthly return of stock index data starts in June 2005, includes the financial crisis period in September 2007, and ends in November 2019. We include the global financial market crisis to present the systematic risk contagion effect in our clustering analysis. The online dataset starts on January 1989, which covers the 1997 Asian financial crisis, 2003 dot-com bubble and the 2007 subprime mortgage crisis, and ends in November 2019. The offline dataset contains 47 observed paths, and each has 174 time points. In the online setting, the longest stochastic path has 371 observations, and the shortest time series have 174 observations. The online dataset begins with 33 economic entities and ends with 47 economic entities.

As for monthly sovereign CDS spreads, we remove the Netherlands, Qatar, Singapore, Greece and United Arab Emirates from the sample due to insufficient observations. We end up with 42 economic entities with CDS spread in the clustering analysis. The offline dataset starts in June 2005 and ends in November 2019, and each economic entity has 174 observations.

6.3. Clustering Results

We compare the clustering results using offline and online datasets with the number of clusters based on (i) geographical regions (four groups: Americas, Europe and the Middle East, the Pacific, and Asia) and (ii) development level (two groups: emerging markets and

developed markets). The clustering factor (geographical region or development level) with the lower misclassification rate represents the most relevant partition of the economics entities. We examine whether the geographical region or development level differentiates between the financial market's behaviors the most.

Table 1. The categories of major stock and sovereign CDS markets in the MSCI ACWI (All Country World Index). There are 23 developed economic entities and 24 emerging countries or areas. The geographical regions contain Americas, EMEA (Europe, Middle East and Africa), Pacific and Asia. Markets with * have missing sovereign CDS data in our data sample.

Developed Markets			Emerging Markets		
Americas	Europe & Middle East	Pacific	Americas	Europe & Middle East & Africa	Asia
Canada	Austria	Australia	Brazil	Czech Republic	China (Mainland)
USA	Belgium	Hong Kong	Chile	Greece *	India
	Denmark	Japan	Colombia	Hungary	Indonesia
	Finland	New Zealand	Mexico	Poland	Korea
	France	Singapore *	Peru	Russia	Malaysia
	Germany			Turkey	Pakistan
	Ireland			Egypt	Philippines
	Israel			South Africa	Taiwan
	Italy			Qatar *	Thailand
	The Netherlands *			United Arab Emirates *	
	Norway				
	Portugal				
	Spain				
	Sweden				
	Switzerland				
	United Kingdom				

Source: MSCI ACWI (All Country World Index) market allocation. <https://www.msci.com/acwi>, accessed on 1 January 2022.

Table 2 shows that the misclassification rates for clustering using development levels are significantly and consistently lower than that of the geographical region. The empirical result is robust, using offline or online settings and using stock indexes or sovereign CDS spreads. The clustering performance tends to support the dominance of the development level in comparison to the geographical region, in terms of characterizing the financial market features in recent decades. The best results (lowest misclassification rate) are (i) clustering the offline dataset using offline algorithm, and (ii) clustering the online dataset using an online algorithm, when the cluster factor is using the development level (emerging markets vs. developed markets). There are 30% to 50% decreases on misclassification rates when clustering using development level compared to when clustering using region.

Table 3 presents misclassified economic entities in the cluster analysis, when we use the development level as the clustering factor. For global stock indexes, the misclassification often occurs in the case where developed economic entities is clustered into the emerging market group. Stock index returns from Austria, Finland, and Portugal are clustered in the emerging market group in both offline and online algorithms. The misclassification within the developed market group demonstrates a more random pattern. For sovereign CDS spreads, the misclassification occurs in the cases where emerging economic entities are clustered into the developed market group. Sovereign CDS spreads of Chile, China (Mainland), Czech Republic, Korea, Malaysia, Mexico, Poland, and Thailand are consistently misclassified into the developed market group. A plausible reason is that these economic entities have low sovereign credit risk compared to their peers in the most recent years.

Table 2. The misclassification rates of clustering algorithms on different datasets. The clustering facts are geographical region and development level, respectively. Panel A presents the results from clustering global stock indexes, and Panel B presents the results from clustering sovereign CDS spreads.

Panel A		Offline Algorithm		Online Algorithm	
Stock Returns	Regions	Emerging/Developed	Regions	Emerging/Developed	
offline dataset	61.70%	29.79%	55.32%	36.17%	
online dataset	53.19%	44.68%	51.06%	14.89%	
Panel B		Offline Algorithm		Online Algorithm	
CDS Spreads	Regions	Emerging/Developed	Regions	Emerging/Developed	
offline dataset	64.29%	28.57%	71.43%	26.19%	
online dataset	54.76%	47.62%	59.52%	26.19%	

In both stock and CDS cases, we show that clustering financial time series by development level outperforms geographical region in our sample. This empirical evidence supports that economic globalization is breaking the geographic barrier and enhances the comovement of financial market performance by economics development status.

Table 3. The misclassification cases when using offline algorithm on offline dataset and online algorithm on online dataset. Panel A reports the mis-categorized economics entities in the stock index clustering case, and Panel B reports the mis-categorized economics entities in the sovereign CDS spreads clustering case. The algorithm clusters the dataset into two groups: emerging market group and developed market group. The mis-categorized outcome are reported, where (i) entities from developed markets incorrectly clusters in emerging market, or (ii) vice versa.

Panel A: Equity Indexes Returns			
Group 1 (Emerging Markets)		Group 2 (Developed Markets)	
Incorrect-Offline	Incorrect-Online	Incorrect-Offline	Incorrect-Online
Austria	Austria	Korea	Czech Republic
Finland	Finland	Chile	Qatar
Germany	Portugal	Philippines	Peru
Ireland		Malaysia	South Africa
Italy		Mexico	
Norway			
Portugal			
Spain			
New Zealand			
Panel B: Sovereign CDS Spreads			
Group 1 (Emerging Markets)		Group 2 (Developed Markets)	
Incorrect-Offline	Incorrect-Online	Incorrect-Offline	Incorrect-Online
Ireland	Ireland	Chile	Chile
Italy	Portugal	China (Mainland)	China (Mainland)
Portugal		Czech Republic	Czech Republic
Spain		Korea	Hungary
		Malaysia	Korea
		Mexico	Malaysia
		Poland	Mexico
		Thailand	Poland
			Thailand

7. Conclusions and Future Prospects

In this paper, we introduce the problem of clustering locally asymptotically self-similar processes. A novel covariance-based dissimilarity measure is proposed to obtain approximately asymptotically consistent clustering algorithms for both offline and online settings. We showed that the recommended algorithms are competitive for at least three reasons:

- (1) Given their flexibility, our algorithms are applicable to clustering any distribution stationary ergodic processes with finite variances, any autocovariance ergodic processes, and locally asymptotically self-similar processes whose tangent processes have autocovariance ergodic increments. Multifractional Brownian motion (mBm) is an excellent representative of the latter class of processes.
- (2) Our algorithms are efficient enough in terms of their computational complexity. A simulation study is performed on clustering mBm. The results show that both offline and online algorithms are approximately asymptotically consistent.
- (3) Our algorithms are successfully applied to cluster the real world financial time series (equity returns and sovereign CDS spreads) via the development level and via regions. The outcomes are self-consistent with the financial markets behavior and they reveal the level of impact between the economic development and regions on equity returns.

Finally, we list the following open problems which could be left for future research.

- (1) The clustering framework proposed in our paper only focuses on the cases where the true number of clusters κ is known. The problem for which κ is supposed to be unknown remains open.
- (2) If we drop the Gaussianity assumption, the class of stationary incremental self-similar processes becomes much larger. This will yield an introduction to a more general class of locally asymptotically self-similar processes, whose autocovariances do not exist. This class includes linear multifractional stable motion [55,56] as a paradigmatic example. Cluster analysis of such stable processes will no doubt lead to a wide range of applications, especially when the process distributions exhibit heavy-tailed phenomena. Neither the distribution dissimilarity measure introduced in [12] nor the covariance-based dissimilarity measures used in this paper would work in this case, hence new techniques are required to cluster such processes, such as considering replacing the covariances with covariations [35] or symmetric covariations [57] in the dissimilarity measures.

Author Contributions: Conceptualization and Methodology: N.R. and Q.P.; Formal analysis and investigation: N.R., Q.P. and R.Z.; Writing: N.R., Q.P. and R.Z.; Data collection and analysis: R.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in the simulation study are available at https://github.com/researchcoding/clustering_locally_asymptotically_self_similar_processes/, accessed on 1 January 2022. The datasets used in the real world application are available at <https://www.msci.com/acwi>, accessed on 1 January 2022.

Acknowledgments: The authors would like to express their special thanks to the assistant editor Finley Shen and the three anonymous reviewers for their careful reading of the manuscript and many insightful feedback and suggestions, which have led to a significant improvement of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

fBm	fractional Brownian motion
mBm	multifractional Brownian motion
gBm	geometric Brownian motion

References

- Cotofrei, P. Statistical temporal rules. In Proceedings of the 15th Conference on Computational Statistics—Short Communications and Posters, Berlin, Germany, 24–28 August 2002.
- Harms, S.K.; Deogun, J.; Tadesse, T. Discovering sequential association rules with constraints and time lags in multiple sequences. In *International Symposium on Methodologies for Intelligent Systems*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 432–441.
- Jin, X.; Lu, Y.; Shi, C. Distribution discovery: Local analysis of temporal rules. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 469–480.
- Jin, X.; Wang, L.; Lu, Y.; Shi, C. Indexing and mining of the local patterns in sequence database. In *International Conference on Intelligent Data Engineering and Automated Learning*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 68–73.
- Keogh, E.; Kasetty, S. On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Min. Knowl. Discov.* **2003**, *7*, 349–371. [[CrossRef](#)]
- Lin, J.; Keogh, E.; Lonardi, S.; Patel, P. Finding motifs in time series. In Proceedings of the 2nd Workshop on Temporal Data Mining, Edmonton, AB, Canada, 23–26 July 2002; pp. 53–68.
- Li, C.S.; Yu, P.S.; Castelli, V. MALM: A framework for mining sequence database at multiple abstraction levels. In Proceedings of the Seventh International Conference on Information and Knowledge Management, Bethesda, MA, USA, 2–7 November 1998; pp. 267–272.
- Bradley, P.S.; Fayyad, U.M. Refining Initial Points for K-Means Clustering. *ICML* **1998**, *98*, 91–99.
- Keogh, E.; Chakrabarti, K.; Pazzani, M.; Mehrotra, S. Dimensionality reduction for fast similarity search in large time series databases. *Knowl. Inf. Syst.* **2001**, *3*, 263–286. [[CrossRef](#)]
- Java, A.; Perlman, E.S. Predictive Mining of Time Series Data. *Bull. Am. Astron. Soc.* **2002**, *34*, 741.
- Khaleghi, A.; Ryabko, D.; Mary, J.; Preux, P. Online clustering of processes. In Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, La Palma, Canary Islands, 21–23 April 2012; pp. 601–609.
- Khaleghi, A.; Ryabko, D.; Mary, J.; Preux, P. Consistent algorithms for clustering time series. *J. Mach. Learn. Res.* **2016**, *17*, 1–32.
- Peng, Q.; Rao, N.; Zhao, R. Covariance-based dissimilarity measures applied to clustering wide-sense stationary ergodic processes. *Mach. Learn.* **2019**, *108*, 2159–2195. [[CrossRef](#)]
- Comte, F.; Renault, E. Long memory in continuous-time stochastic volatility models. *Math. Financ.* **1998**, *8*, 291–323. [[CrossRef](#)]
- Bianchi, S.; Pianese, A. Multifractional properties of stock indices decomposed by filtering their pointwise Hölder regularity. *Int. J. Theor. Appl. Financ.* **2008**, *11*, 567–595. [[CrossRef](#)]
- Bianchi, S.; Pantanella, A.; Pianese, A. Modeling and simulation of currency exchange rates using multifractional process with random exponent. *Int. J. Model. Optim.* **2012**, *2*, 309–314. [[CrossRef](#)]
- Bertrand, P.R.; Hamdouni, A.; Khadhraoui, S. Modelling NASDAQ series by sparse multifractional Brownian motion. *Methodol. Comput. Appl. Probab.* **2012**, *14*, 107–124. [[CrossRef](#)]
- Bianchi, S.; Pantanella, A.; Pianese, A. Modeling stock prices by multifractional Brownian motion: An improved estimation of the pointwise regularity. *Quant. Financ.* **2013**, *13*, 1317–1330. [[CrossRef](#)]
- Bianchi, S.; Frezza, M. Fractal stock markets: International evidence of dynamical (in) efficiency. *Chaos Interdiscip. J. Nonlinear Sci.* **2017**, *27*, 071102. [[CrossRef](#)] [[PubMed](#)]
- Pianese, A.; Bianchi, S.; Palazzo, A.M. Fast and unbiased estimator of the time-dependent Hurst exponent. *Chaos Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 031102. [[CrossRef](#)] [[PubMed](#)]
- Marquez-Lago, T.T.; Leier, A.; Burrage, K. Anomalous diffusion and multifractional Brownian motion: Simulating molecular crowding and physical obstacles in systems biology. *IET Syst. Biol.* **2012**, *6*, 134–142. [[CrossRef](#)] [[PubMed](#)]
- Song, W.; Cattani, C.; Chi, C.H. Multifractional Brownian motion and quantum-behaved particle swarm optimization for short term power load forecasting: An integrated approach. *Energy* **2020**, *194*. [[CrossRef](#)]
- Sikora, G. Statistical test for fractional Brownian motion based on detrending moving average algorithm. *Chaos Solitons Fractals* **2018**, *116*, 54–62. [[CrossRef](#)]
- Balcerek, M.; Burnecki, K. Testing of fractional Brownian motion in a noisy environment. *Chaos Solitons Fractals* **2020**, *140*, 110097. [[CrossRef](#)]
- Balcerek, M.; Burnecki, K. Testing of multifractional Brownian motion. *Entropy* **2020**, *22*, 1403. [[CrossRef](#)]
- Jin, S.; Peng, Q.; Schellhorn, H. Estimation of the pointwise Hölder exponent of hidden multifractional Brownian motion using wavelet coefficients. *Stat. Inference Stoch. Process.* **2018**, *21*, 113–140. [[CrossRef](#)]
- Peng, Q.; Zhao, R. A general class of multifractional processes and stock price informativeness. *Chaos Solitons Fractals* **2018**, *115*, 248–267. [[CrossRef](#)]

28. Vu, H.T.; Richard, F.J. Statistical tests of heterogeneity for anisotropic multifractional Brownian fields. *Stoch. Process. Their Appl.* **2020**, *130*, 4667–4692. [[CrossRef](#)]
29. Bicego, M.; Trudda, A. 2D shape classification using multifractional Brownian motion. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 906–916.
30. Kirichenko, L.; Radivilova, T.; Bulakh, V. Classification of fractal time series using recurrence plots. In *Proceedings of the 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 9–12 October 2018; pp. 719–724.
31. Krengel, U. *Ergodic Theorems*; de Gruyter Studies in Mathematics; de Gruyter: Berlin, Germany, 1985.
32. Grazzini, J. Analysis of the emergent properties: Stationarity and ergodicity. *J. Artif. Soc. Soc. Simul.* **2012**, *15*, 7. [[CrossRef](#)]
33. Samorodnitsky, G. Extreme value theory, ergodic theory and the boundary between short memory and long memory for stationary stable processes. *Ann. Probab.* **2004**, *32*, 1438–1468. [[CrossRef](#)]
34. Boufoussi, B.; Dozzi, M.; Guerbaz, R. Path properties of a class of locally asymptotically self-similar processes. *Electron. J. Probab.* **2008**, *13*, 898–921. [[CrossRef](#)]
35. Samorodnitsky, G.; Taqqu, M.S. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*; Chapman & Hall: New York, NY, USA, 1994.
36. Embrechts, P.; Maejima, M. An introduction to the theory of self-similar stochastic processes. *Int. J. Mod. Phys. B* **2000**, *14*, 1399–1420. [[CrossRef](#)]
37. Embrechts, P.; Maejima, M. *Selfsimilar Processes*; Princeton Series in Applied Mathematics; Princeton University Press: Princeton, NJ, USA, 2002.
38. Falconer, K. Tangent fields and the local structure of random fields. *J. Theor. Probab.* **2002**, *15*, 731–750. [[CrossRef](#)]
39. Falconer, K. The local structure of random processes. *J. Lond. Math. Soc.* **2003**, *67*, 657–672. [[CrossRef](#)]
40. Mandelbrot, B.B.; Van Ness, J.W. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **1968**, *10*, 422–437. [[CrossRef](#)]
41. Peltier, R.F.; Lévy-Véhel, J. *Multifractional Brownian Motion: Definition and Preliminary Results*; Technical Report 2645; Institut National de Recherche en Informatique et en Automatique, INRIA: Rocquencourbe, France, 1995.
42. Benassi, A.; Jaffard, S.; Roux, D. Elliptic Gaussian random processes. *Revista Matemática Iberoamericana* **1997**, *13*, 19–90. [[CrossRef](#)]
43. Stoev, S.A.; Taqqu, M.S. How rich is the class of multifractional Brownian motions? *Stoch. Process. Their Appl.* **2006**, *116*, 200–221. [[CrossRef](#)]
44. Ayache, A.; Véhel, J.L. On the identification of the pointwise Hölder exponent of the generalized multifractional Brownian motion. *Stoch. Process. Their Appl.* **2004**, *111*, 119–156. [[CrossRef](#)]
45. Ayache, A.; Taqqu, M.S. Multifractional processes with random exponent. *Publicacions Matemàtiques* **2005**, *49*, 459–486. [[CrossRef](#)]
46. Bianchi, S.; Pantanella, A. Pointwise regularity exponents and well-behaved residuals in stock markets. *Int. J. Trade Econ. Financ.* **2011**, *2*, 52–60. [[CrossRef](#)]
47. Cadoni, M.; Melis, R.; Trudda, A. Financial crisis: A new measure for risk of pension fund portfolios. *PLoS ONE* **2015**, *10*, e0129471. [[CrossRef](#)] [[PubMed](#)]
48. Frezza, M. Modeling the time-changing dependence in stock markets. *Chaos Solitons Fractals* **2012**, *45*, 1510–1520. [[CrossRef](#)]
49. Frezza, M.; Bianchi, S.; Pianese, A. Forecasting Value-at-Risk in turbulent stock markets via the local regularity of the price process. *Comput. Manag. Sci.* **2022**, *19*, 99–132. [[CrossRef](#)]
50. Garcin, M. Fractal analysis of the multifractality of foreign exchange rates. *Math. Methods Econ. Financ.* **2020**, *13–14*, 49–73.
51. Wood, A.T.; Chan, G. Simulation of stationary Gaussian processes in $[0, 1]^d$. *J. Comput. Graph. Stat.* **1994**, *3*, 409–432. [[CrossRef](#)]
52. Chan, G.; Wood, A.T. Simulation of multifractional Brownian motion. In *COMPSTAT*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 233–238.
53. Demirer, R.; Omay, T.; Yuksel, A.; Yuksel, A. Global risk aversion and emerging market return comovements. *Econ. Lett.* **2018**, *173*, 118–121. [[CrossRef](#)]
54. Ang, A.; Longstaff, F.A. Systemic sovereign credit risk: Lessons from the US and Europe. *J. Monet. Econ.* **2013**, *60*, 493–510. [[CrossRef](#)]
55. Stoev, S.; Taqqu, M.S. Stochastic properties of the linear multifractional stable motion. *Adv. Appl. Probab.* **2004**, *36*, 1085–1115. [[CrossRef](#)]
56. Stoev, S.; Taqqu, M.S. Path properties of the linear multifractional stable motion. *Fractals* **2005**, *13*, 157–178. [[CrossRef](#)]
57. Ding, Y.; Peng, Q. Series representation of jointly SaS distribution via a new type of symmetric covariations. *Commun. Math. Stat.* **2021**, *9*, 203–238. [[CrossRef](#)]