Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Identifying Irregular Rainfall Patterns Using Persistent Homology

Zabidi Abu Hasan[1], R.U. Gobithaasan[2,3,*], Nur Fariha Syaqina Mohd Zulkepli[3], Mohd Zaharifudin Muhamad Ali[4], Kenjiro T. Miura[5], Paweł Dłotko[6]

1 Institute of Engineering Mathematics, University Malaysia Perlis, 02600 Arau, Perlis, Malaysia
2 Special Interest Group on Modelling & Data Analytics, Faculty of Ocean Engineering Technology & Informatics, University Malaysia Terengganu 21030 Kuala Terengganu, Terengganu, Malaysia
3 School of Mathematical Sciences, Universiti Sains Malaysia, 11800, Penang, Malaysia
4 Water Resources Management and Hydrology Division, 50626 Kuala Lumpur, Malaysia
5 Graduate School of Engineering, Shizuoka University, Hamamatsu, 432-8018 Japan
6 Dioscuri Centre in Topological Data Analysis, Mathematical Institute, Polish Academy of Sciences, Sniadeckich 8, Warsaw, Poland

ARTICLE INFO

Article history:
Received 8 May 2023
Received in revised form 6 September 2023
Accepted 11 November 2023
Available online 11 December 2023

Keywords:
Irregular patterns; Topological data analysis; Persistent homology; Persistence Curves; Flood events

ABSTRACT

Efficient investigation tools are required to elucidate the changes in climatic change caused by various climate processes, variables, and socioeconomic development activities around the world. In this study, we track the changes of daily rainfall at three flood-prone sites in Terengganu between 2012 to 2017. In recent years, topological data analysis (TDA) has been applied in many fields of data analytics to rank, classify, and cluster time series datasets. In this work, we employ Persistent Homology to quantify and identify topological patterns from a rainfall data. A sliding window (SW) approach is used for each 1D rainfall dataset to embed in higher dimensions before computing its Persistence Diagrams (PD). The topological information obtained from PD, namely connected components (H_0) is then retrieved and vectorized in the form of Persistence Curves (Persistence Landscape (PL), Persistence lifetime Curve (PLC), and Persistence Lifetime Entropy (PLE)) to identify unusual rainfall patterns. We employ various types of L_1-norms from these Persistence vectors to identify anomalies in rainfall data which can be used as an early warning flood system. The irregular pattern of Persistence lifetime and Persistence entropy mismatch the actual flood events suggesting that the irregular points may not be as closely related to flood risk. However, PL analysis of the irregular points shows match of about 59% to the flood events. It is expected that other determining factors, for example, land use, cloud cover, and wind information, which can be obtained via satellite gridded data may increase the predictability of flood events thus promotes an effective flood risk management strategies.

1. Introduction

Rainfall patterns play a critical role in many aspects of the environment and human society, including agriculture, water resources, and natural disasters. Rainfall data can be used to make informed decisions about weather forecasting, crop planning, and water conservation efforts.

* *Corresponding author.*
E-mail address: gr@umt.edu.my

https://doi.org/10.37934/araset.35.1.2636

However, rainfall data can be highly variable, both spatially and temporally [1]. Rainfall is influenced by a wide range of meteorological processes, including temperature, humidity, atmospheric pressure, and wind. It is also influenced by climate change, which causes shifts in temperature, precipitation, and other weather patterns. This makes it difficult to identify patterns or trends in the data, and accurately predict future rainfall patterns. In addition, rainfall patterns can vary significantly both within a single location and over time, making it difficult to identify trends or make reliable predictions.

A critical transition in rainfall data refers to a point at which the pattern or distribution of rainfall changes significantly. This can be due to a variety of factors, such as changes in atmospheric conditions, shifts in weather patterns, or the influence of climate change. Critical transitions are abrupt changes in the state of ecosystems, climate, financial systems, or other complex dynamical systems that occur when changing conditions reach a critical or bifurcation threshold [2]. Critical transitions in rainfall data can have significant impacts on local ecosystems and economies, as they can affect the availability of water for agricultural, industrial, and domestic purposes. They can also have indirect impacts on sectors such as energy, transportation, and tourism. Identifying critical transitions in rainfall data can be challenging due to the inherent variability of rainfall patterns and the potential for data quality issues. However, it is important to understand these transitions to better predict and prepare for future changes in rainfall patterns which can be irregular.

Identifying irregular patterns in rainfall data can be important in a variety of contexts because it can help to uncover underlying trends or issues that may not be immediately apparent. For example, in the field of finance, irregular patterns in data can be a warning sign of fraudulent activity or financial instability [3,5]. In healthcare, irregular patterns in data can be an indication of a medical condition or other health issue [6]. In general, being able to identify irregular patterns can help to improve decision making and enable more informed actions to be taken.

Persistent homology (PH) is a mathematical tool that can be used to identify topological features in data, such as connected components, loops, and voids [7,8]. These features can be used to identify irregular patterns in data, as they can highlight areas where the data exhibits unusual structure or behaviour. One reason why PH might be used to identify irregular patterns in rainfall data is that it can provide a more comprehensive and robust analysis of the data compared to traditional methods [9-11]. Traditional methods, such as statistical analysis or machine learning algorithms, may be limited in their ability to capture complex patterns or structures in the data. PH, on the other hand, can identify a wide range of topological features, which can allow for a more detailed and nuanced understanding of the data. In addition, PH might be used to identify irregular patterns in rainfall data that is relatively insensitive to noise or perturbations in the data. This can be particularly useful when analyzing rainfall data, as this data is often affected by factors such as instrumentation errors, exposure, and evaporation, which can introduce noise or bias into the data.

Recently, Gidea et al., [3,4] have introduced a method to investigate financial crisis using PH. The method highlighted by them shows that PH is robust to financial data analysis. PH is also able to provide various topological summaries that act to provide the necessary data to detect the financial crisis. In the last decade, there has been a growing body of empirical and theoretical studies, inspired by analysis of abrupt transitions in complex natural systems, devising early warning signals (EWS) in financial markets. Motivated by these studies, in this paper, we investigate whether the application of PH to rainfall time series could help detect a critical transition in extreme events based on the irregular patterns in rainfall data. Therefore, we apply the PH method to investigate the behaviour of rainfall time series in different locations. We compare three different types of topological summaries in vector representations which involve the connected component (H_0) to detect the warning signal.

The arrangement of this paper starts with the introduction of rainfall data and its details in Section 2. In Section 3 we provide a concise and informal review of Persistent Homology (PH) and key concepts employed in this paper. Section 4 describes the topological summaries used in this paper, followed by methodologies. Section 6 presents our findings on rainfall data, which demonstrate that the time series of the norms of persistent lifetime curve, persistence lifetime entropy and persistence landscapes to its variability that can be used as a new EWS for flood events. Finally, section 7 concludes the paper.

2. Rainfall Data

Terengganu is a state that experiences flooding almost every year. Therefore, the rainfall data collected in this state is particularly important for determining the distribution of precipitation and highlighting regions affected by floods. The Malaysian Department of Irrigation and Drainage (DID) manages almost 90 rainfall stations around the state. In this research, we analyzed precipitation information from three stations: Kemaman, Hulu Terengganu, and Besut. The data utilized spans from year 2012 to 2017. The specifics of the data are shown in Table 1. Figure 1 illustrates the rainfall data for three stations that we used in this paper.

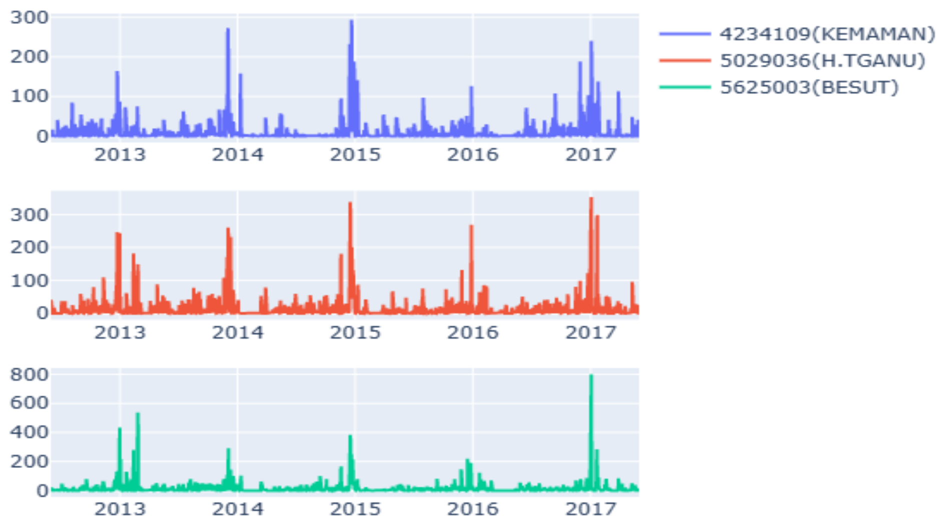


Fig. 1. Daily rainfall data plot for Kemaman, Hulu Terengganu and Besut stations between 2012-2017

Table 1
Details of rainfall stations with data from 2012-2017 (1826 days)

District	Station ID	Station Name	Latitude	Longitude	Maximum rainfall (mm)	Average rainfall (mm)	Standard deviation rainfall (mm)
Kemaman (K)	4234109	JPS. Kemaman	04˚13′ 55″	103˚ 25′ 20″	293.2	7.2	22.1
Hulu Terengganu (H.T)	5029036	Rumah Pam Paya Kemat	05˚ 00′ 30″	102˚ 58′ 10″	352.4	11.1	28.7
Besut (B)	5625003	Paya Peda Besut	05˚ 36′ 00″	102˚ 30′ 55″	799.5	12.4	41.0

3. Persistent Homology

Persistent homology (PH) is a way of studying the topology of a space by tracking how topological features change as the space is filtered [10,12]. The key idea is that a topological feature that appears in the space at some point will typically persist for some range of filters before disappearing. By studying this persistence, we can gain insight into the topological structure of the space. For instance, considering a topological space X and a family of filters F on X,a filter $f \in F$ is a function that assigns to each point $x \in X$ a real number $f(x)$. We assume that the filters are ordered, so that for any two filters f and g, if $f(x) < g(x)$ for all $x \in X$, then f comes before g in the ordering. Given a filter f, we can define the sublevel set of X at f to be the set of all points $x \in X$ such that $f(x) \leq r$, where r is a real number. If we vary r over all real numbers, we obtain a nested sequence of sublevel sets $X_0 \supseteq X_1 \supseteq X_2 \supseteq \cdots \supseteq X_n$ where X_0 is the entire space X and X_{i+1} obtained by taking the sublevel set of X_i at some filter $f \in F$. We can use this nested sequence of sublevel sets to define a sequence of topological spaces $X_0, X_1, X_2, \ldots, X_n$. Each of these spaces is obtained by taking the union of all the sublevel sets in the sequence up to a certain point.

To perform PH, we compute the topological features (e.g. connected components (H_0), loops (H_1), or higher-dimensional holes) of each of these spaces, and record the range of filters for which each feature exists [13]. This range is called the feature's persistence. We can then visualize the persistence of the topological features using a persistence diagram, which is a plot with the filters on the x-axis (birth) and the y-axis (death). Each topological feature is represented by a point in the diagram. The x-coordinate and y-coordinate of the point represents the birth and death filters of the feature, respectively. There are a few different ways to compute persistent homology, depending on the specific form of the filters and the topological features. In this work we employed Vietoris-Rips complexes [8,14] which is the common filter to perform the PH.

4. Topological Summaries

Topological summaries are mathematical tools that allow us to extract key information from complex datasets. These summaries are often used in data analysis, machine learning, and other areas in need of large and complex dataset analysis. One important type of topological summary is the persistent homology, which captures the topological features of a dataset across multiple scales which is visualized using persistent diagram. Another important tool in persistent homology is the persistence life curve, which plots the persistence of each topological feature as a function of scale. Finally, the persistence landscape is a higher-dimensional extension of the persistence diagram that captures more information about the persistence of topological features.

4.1 Persistent Diagram

The most fundamental instrument for topological summaries is the persistent diagram (PD) where we can use it to represent the multiset of birth (b) and death (d) points of the topological features [15]. A multiset is a mathematical structure that is similar to a set but allows for multiple instances of the same element. For $b < \varepsilon < d$, a multiset can be represented as a collection of (b, d), where each (b, d) can appear more than once. The number of times (b, d) it appears in the multiset is known as its multiplicity and denoted as $m(s)$. Let's define a diagonal multiset as

$$\Delta = [(\varepsilon, \varepsilon) | x \in \mathbb{R},\ m(\varepsilon, \varepsilon) = \infty] \quad (1)$$

In general, we define a persistent diagram, dgm as union of two multisets $dgm = \Delta \cup [(b, d) \mid b < d \in \mathbb{R} \cup \{\infty\}]$ [16]. The topological feature observed in this paper is a connected component (H_0) that persist in the dgm. The PD based on the multiset would then encode information about the (b, d) of the connected components as the scale of the space changes. As the scale increases, new connected components may appear, while others may merge or disappear. The dgm would show this by having points representing the connected components appear at the birth time, persist through a range of scales, and then disappear at the death time.

4.2 Persistence Life Curve

Normalized persistence curve (PC) with stable persistence summary was suggested for PD by Chung and Lawson [16]. They provide a generic PC to describe the topological vector. Let dgm be the set of all PD and F be the set of all functions. We denote the Θ be the set of summary statistics or operators that map multi-sets to real numbers. Let a map $P : D \times F \times \Theta \rightarrow \mathbb{R}$ so

$$P(D, \psi, T)(\varepsilon) = T([\psi(dgm; b, d, \varepsilon) | (b, d) \in dgm_{\varepsilon}]) \tag{2}$$

where $P(dgm, \psi, T)(\varepsilon)$ is called the persistence curve of dgm with respect to ψ and T. We use a user-defined function, ψ, while T is a summary statistic or operator that maps multi-sets to real numbers. In this paper, we include two types of PC defined as follows:

a. For lifetime curve (PLC), we let the function $\psi(dgm; b, d, \varepsilon) = d - b$ and $T = \sum$. Then,

$$P(dgm, \psi, \textstyle\sum)(\varepsilon) = \sum([d - b | (b, d) \in dgm_{\varepsilon}]) \tag{3}$$

b. For life entropy curve (PLE), we define the function $\psi(dgm; b, d, \varepsilon) = -\frac{d-b}{L} \log \frac{d-b}{L}$ and $T = \sum$ where $L = \sum_{(b,d) \in dgm}(d - b)$. Then,

$$P(D, \psi, \textstyle\sum)(\varepsilon) = \sum\left(\left[= -\frac{d-b}{L} \log \frac{d-b}{L} \middle| (b, d) \in dgm_{\varepsilon}\right]\right) \tag{4}$$

4.3 Persistence Landscape

Persistence landscape (PL) is a practical and reliable approach for extracting homology summaries that persist throughout time [17]. PL is a set of piecewise linear functions in the 2D space specified by the birth-death coordinates of PD. For each birth-death point $(b, d) \in dgm$, the piecewise function is defined as

$$\Lambda_{(b,d)}(\varepsilon) = \begin{cases} \varepsilon - b & if\ \varepsilon \in \left(b, \frac{b+d}{2}\right) \\ d - \varepsilon & if\ \varepsilon \in \left(\frac{b+d}{2}, d\right) \\ 0 & if\ \varepsilon \notin (b, d) \end{cases} \tag{5}$$

The k-th PL defined by $\lambda_k(\varepsilon) = \max_k\left\{\Lambda_{(b_i,d_i)}(\varepsilon) \mid (b, d) \in dgm\right\}$ consists of finite number of collections off diagonal points, then, the corresponding sequence of functions $\lambda_k: \mathbb{R} \rightarrow [0, +\infty]$ where $k \in N$. Let $\psi(dgm; b, d, \varepsilon) = \min\{\varepsilon - b, d - \varepsilon\}$ and $T = \max_k$ [16]. Then,

$$P(dgm, \psi, \sum)(\varepsilon) \equiv \lambda_k \ . \tag{6}$$

By using tools such as PD, PLC, PLE, and PL we can gain insights into the topological features of a dataset and track their evolution over different scales. These summaries are particularly useful in data analysis and machine learning, where they can help us identify important features and patterns in large datasets.

5. Methodology

In this section, we describe the methods used in our study, which involve phase space reconstruction, L_1-norm, and irregularity detection. Phase space reconstruction is a technique used to transform a time series into a high-dimensional space, which allows us to analyze the dynamics of the system under study. The L_1-norm is a mathematical tool that allows us to measure the magnitude of the difference between two sets of data. It is particularly useful in our study because it can effectively capture the irregularities in the data that we are analyzing. Finally, we will discuss irregularity detection, which is a method used to identify and quantify the irregularities in the data.

5.1 Phase space reconstruction

By using the daily window sliding with constant window $w = 60$, we can divide X into a collection of sliding windows of size 60 as $\{X(60), X(61), \dots, X(1826)\}$ where $X(j) \subset X$ for all $j \in \{60,61, \dots ,1826\}$. For each j, sliding windows at the rainfall day j is denoted as $X(j) = \{x(t) \in \mathbb{R}\}_{t=j-60+1}^{j}$, is a data consisting of 60 days.

To use the topological approach, 1D rainfall data must be embedded into a higher dimension using Taken's embedding [18], which requires two crucial parameters: time-delay (τ) and dimension (d). Let a time series $x: t \to \mathbb{R}$ and a parameter τ, a time delay embedding that is a lift to a time series $\phi: t \to \mathbb{R}^d$ defined by

$$\phi(t) = (x(t), f(t+\tau), \dots, f(t+(d-1)\tau)) \tag{7}$$

In this research, we use time-delay $\tau = 1$ to ensure that each subsequent point in the higher dimensional space is separated by a constant time interval and dimension $d = 3$. A 3D point cloud is created for each time-window to strike a balance between capturing the complexity of the system dynamics and avoiding overfitting. In matrix form, we can illustrate the point cloud data (PCD) in $(60 - d) \times d$ matrix as below:

$$X(j) = \begin{bmatrix} x(j-60+1) & x(j-60+2) & x(j-30+3) \\ \vdots & \vdots & \vdots \\ x(j-2) & x(j-1) & x(j) \end{bmatrix} \tag{8}$$

where $j = 60, \dots ,1826$.

5.2 L_1-norm

In this study, we implemented 0-dimensional topological features, called connected components, on PCDs. For each PCD, the PH was applied to calculate all persistent pairs and all pairs obtained will be vectorized in a PLC, PLE and PL. Let $P = P(dgm, \psi, \sum)(\varepsilon)$ be the vector form, for each persistence summary, L_1-norm will be calculated as follows:

$$\|P\| = \sum_{i=1}^{n} P_i \,. \tag{9}$$

As a result, we obtain L_1-norm time series denoted as $Y = \{y(j) \in \mathbb{R}\}_{j=60}^{1826}$ where $y(j)$ is a L_1-norm value at the rainfall $j = 60, \dots, 1826$. For PL, we use the average of five strip denoted as,

$$Y_{PL} = \frac{1}{5}\sum_{k=1}^{5}\|P\|_k. \tag{10}$$

5.3 Irregular Detection

To find the changes of the irregular patterns in the rainfall data using the topological information derived as Y, we compute the differentiated value (DV) for two days [4]. Then, after computing the DV, we applied sliding window for identifying the local threshold of the irregular patterns. The sliding window is set to 10% of the number data obtained. For each window we are going to compute the mean and standard deviation which becomes our local thresholds. Then we compare with the DV. The irregular pattern will appear outside the band where we use the following band formula:

$$band(u) = \bar{y}_u \pm C\sigma_u \tag{11}$$

where $C \in \mathbb{R}$ is a constant, $\bar{y}_u$ and $C\sigma_u$ are the mean and standard deviation for window u. To test the effectiveness of the irregular patterns obtained, we let $C = 5$ and compared it to the actual flood dates reported by DID between June 2012 and May 2017 to assess its efficacy.

6. Result and Discussions

This section summarises the outcomes of our phase space reconstruction, persistence diagrams, and persistence landscapes produced by applying PH to the daily precipitation data from stations in Terengganu. We propose the notion of sliding windows since our purpose is to examine data based on the variation of daily topological information. With $\tau = 1$, the length of the sliding window employed in this work is 60 days. We then use Takens' embedding with $\tau = 1$ and $d = 3$ to a time series for each sliding window of precipitation data. We then acquire three-dimensional point cloud data from the phase space reconstruction. We extracted and tested for H_0 and H_1 topological information, thus we let $d = 3$. However, user may increase the dimension if higher order of homology is needed.

Each window is marked by the end date of the data it contains; hence, the output for each window is derived only from the data of prior dates. This time-ordered series of sliding windows is subjected to H_0 to examine the daily transition of topological properties. There are three distinct time series findings, including L_1-norm of the PLC, PLE and PL.

The PH analysis of the rainfall data identified several irregular patterns in the data. The points highlighted in Table 2 represent the irregular patterns of the topological summary. The PH may be indicative of a complex or dynamic system, as the presence of irregular points suggests that there are deviations from regular or expected patterns in the data. The irregular points could be the result of external factors or influences that are not reflected in the rest of the data. For example, if the data represents rainfall patterns, the irregular points could be caused by unusual weather events such as floods [19]. The irregular points could be caused by errors or uncertainties in the data collection process. It is important to carefully consider and investigate the potential sources of such errors when analyzing the data. The irregular points could represent opportunities for further study or investigation. For example, if the data represents rainfall patterns, the irregular points could be indicative of areas that are particularly susceptible to drought or flood, and further analysis could help to identify the underlying causes of these patterns.

We retrieved the flood start date provided by DID for each of the three sites between June 2012 and May 2017 to match the beginning of the irregularity points in the rainfall data. Based on the observation and comparison of irregular points in PH and flood events, it has been determined that an alert signal should be implemented. We will provide a signal based on the difference between the irregular point and flood events. The total number of days before or after the flood, with 35 days being the maximum ever recorded is also given. Additionally, it is possible for alert signal to be sent out more than once. Therefore, the lowest value will be chosen and inserted into the table. The word "No" will be shown if the irregular value does not match with the flood events and "false" for signal outside the given range. Table 3 contains the results of the analysis.

Table 2
Irregular patterns detection from station Kemaman, Hulu Terengganu and Besut using three type of persistence representations (PLC, PLE and PL).

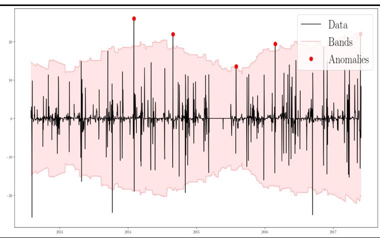
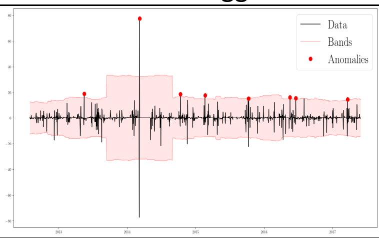
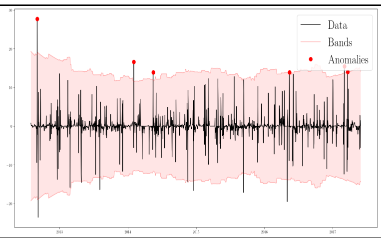
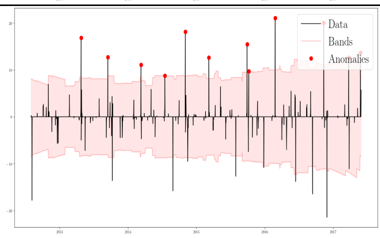
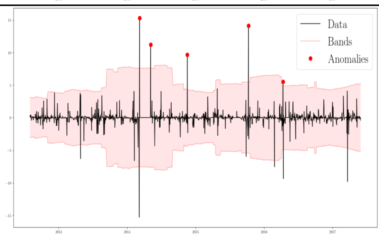
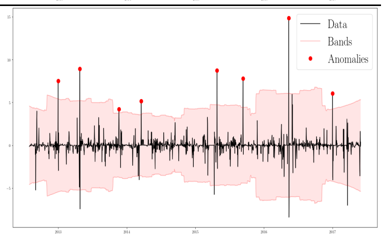
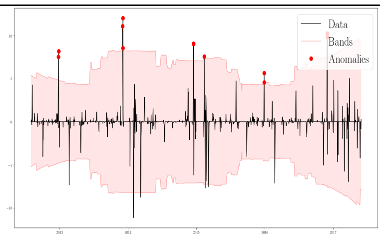
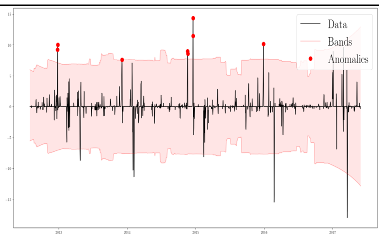
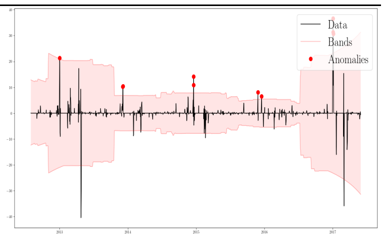
Type	Kemaman	Hulu Terengganu	Besut
PLC			
PLE			
PL			

Table 3
Details of signal (days) using irregular value.

District	Actual flood events	PLC	PLE	PL
KEMAMAN	25/12/2012	No	No	**25/12/2012 (first day)**
	2/12/2013	No	15/9/2013 (false)	3/12/2013 (1 day late)
	9/1/2014	1/2/2014 (after 23 days)	11/3/2014 (false)	**5/12/2013 (early 34 days)**
	19/12/2014	29/8/2014 (false)	17/7/2014 (false)	**16/12/2014 (early 3 days)**
	24/12/2014	No	3/11/2014 (early 36 days)	**17/12/2014 (early 7 days)**
	None		9/3/2015 (false)	13/2/2015 (false)
	None	1/8/2015 (false)	29/9/2015 (false)	29/12/2015 (false)
	None		8/10/2015 (false)	30/12/2015 (false)
	None	26/2/2016 (false)	26/2/2016 (false)	30/11/2016 (false)
	None		12/11/2016 (false)	1/12/2016 (false)
	3/1/2017	27/5/2017 (false)	25/3/2017 (false)	No
	21/1/2017	No	27/5/2017 (false)	No
HULU TERENGGANU	24/12/2012	No	12/10/2012 (false)	25/12/2012 (1 day late)
	31/12/2012	No	25/12/2012 (early 5 days)	**26/12/2012 (early 5 days)**
	2/12/2013	27/5/2013	19/8/2013 (false)	4/12/2013 (2 days late)
	18/12/2014	10/3/2014 (false)	18/11/2014 (early 30 days)	**18/12/2014 (first day)**
	30/12/2014	27/6/2014 (false)	No	**19/12/2014 (early 11 days)**
	None	22/3/2015	N/A	29/12/2015 (false)
	31/12/2016	No	No	No
	3/1/2017	24/4/2017 (false)	No	No
	22/1/2017	No	No	No
BESUT	24/12/2012	3/9/2012 (false)	No	No
	31/12/2012	No	31/12/2012 (first day)	No
	None	N/A	26/4/2013 (false)	1/1/2013 (false)
	None	N/A	20/11/2013 (false)	5/12/2013 (false)
	None	N/A	N/A	6/12/2013 (false)
	18/12/2014	2/2/2014 (false)	19/3/2014 (false)	**18/12/2014 (first day)**
	27/12/2014	16/5/2014 (false)	No	**19/12/2014 (early 8 days)**
	None	N/A	27/4/2015 (false)	N/A
	None	N/A	12/9/2015 (false)	N/A
	31/12/2016	14/5/2016 (false)	13/5/2016 (false)	No signal
	3/1/2017	4/3/2017 (false)	1/1/2017 (early 2 days)	**2/1/2017 (early 1 day)**
	21/1/2017	22/3/2017 (false)	No	**5/1/2017 (early 16 days)**

Table 3 shows the first day record of 22 flood events that occurred in the past in Kemaman, Hulu Terengganu and Besut. Overall, the irregular pattern of PL has the best performance to match the irregular pattern with actual flood events compared to the other patterns. The total number of alerts that can be found is 13 signals, which is equal to 59%. This suggests that there is some level of correlation between the PL vector and flood, but the correlation may not be as strong as desired. This may indicate that the irregular points identified by PL are not the only factors influencing flood risk. On the other hand, PLC and PLE irregular patterns do not match with the actual flood events. The mismatch between the PLC, PLE and actual flood events suggests that the irregular points identified by persistent homology may not be closely related to the flood events. However, the 59% match in the persistence landscape analysis indicates that the irregular points do show some association with the flood events, although it does not capture all the factors that contribute to flood risk. Another reason is the irregular points identified by PH are related to other aspect of the data or variables, such as the dynamics of the weather systems that create the rainfall patterns rather than the floods themselves. This could indicate that while the irregular points do not directly cause flood events, they may be an indicative of conditions that are favorable for flood events to occur.

We test only for the connected component (H_0) because it is a relatively simple topological feature to compute and understand, making them a useful starting point for rainfall data analysis tasks. While H_0 is useful for identifying regular patterns, such as connected regions in the rainfall data, other homology groups, such as $H_1$and H_2, can also be used to identify more complex or irregular patterns. H_1 can be used to identify loops or cycles in the data, which may indicate patterns such as repeating structures or cyclical behavior. H_2, on the other hand, can be used to identify voids or holes in the data, which may indicate patterns such as missing or hidden structures. Furthermore, the use of multiple homology groups can provide an enhanced comprehensive understanding of the data topology and can give insights into different types of patterns that may not be visible from the H_0 alone.

7. Conclusion

This study examines the application of PH to determine the irregular rainfall pattern and its relationship to actual flood events. The proposed method was evaluated using rainfall data from three distinct stations in Terengganu. Three topological summaries from PH were utilized to identify the signal of topological features based on the date of rainfall, and the differentiated value (DV) of L_1-norm was then applied to identify the irregular points.

In general, PL could be utilized as a preprocessing step to examine rainfall data in the hydrological area. The signal of topological features derived using PL shows that PH can produce a signal based on the DV. The comparison of three distinct types of topological summaries via PH were carried out where PL analysis of the irregular points showed match of about 59% to the flood events. On the other hand, the mismatch between the PLC and PLE of the irregular points and the pattern of flood events suggest that the irregular points may not be as closely related to flood risk as initially thought and that other factors may also be important in determining flood risk. Further research may be necessary to fully understand the relationship between the topological feature of rainfall and flood events and to develop effective flood risk management strategies.

In conclusion, there are other factors influencing flood risk that are not reflected in the employed PD vectorization effectively. Land use and topography also play a role in determining flood risk, and these factors may not be captured by the vectorized PD of the rainfall data. Further research may be necessary to extract other vectorization method which include H_1 and H_2 of PD to identify irregular points, flood risk, and other factors that contribute to flood events. It may also be useful to consider other analysis methods such as critical slowing down theory (mean, variance, autocorrelation, and power spectral) that can provide better understanding of the system.

Acknowledgments
This work was supported by the Ministry of Education Malaysia grant FRGS/1/2019/STG06/UMT/02/2. The authors also acknowledge Department of Irrigation and Drainage Malaysia for providing the rainfall and water level data.

References

[1] Muhammad, N. S., J. Abdullah, and P. Y. Julien. "Characteristics of rainfall in peninsular Malaysia." In *Journal of Physics: Conference Series*, vol. 1529, no. 5, p. 052014. IOP Publishing, 2020. https://doi.org/10.1088/1742-6596/1529/5/052014
[2] Scheffer, Marten, Jordi Bascompte, William A. Brock, Victor Brovkin, Stephen R. Carpenter, Vasilis Dakos, Hermann Held, Egbert H. Van Nes, Max Rietkerk, and George Sugihara. "Early-warning signals for critical transitions." *Nature* 461, no. 7260 (2009): 53-59. https://doi.org/10.1038/nature08227
[3] Gidea, Marian, and Yuri Katz. "Topological data analysis of financial time series: Landscapes of crashes." *Physica A: Statistical Mechanics and its Applications* 491 (2018): 820-834. https://doi.org/10.1016/j.physa.2017.09.028
[4] Gidea, Marian, Daniel Goldsmith, Yuri Katz, Pablo Roldan, and Yonah Shmalo. "Topological recognition of critical transitions in time series of cryptocurrencies." *Physica A: Statistical mechanics and its applications* 548 (2020): 123843. https://doi.org/10.1016/j.physa.2019.123843
[5] Ismail, Mohd Sabri, Saiful Izzuan Hussain, and Mohd Salmi Md Noorani. "Detecting early warning signals of major financial crashes in bitcoin using persistent homology." *IEEE Access* 8 (2020): 202042-202057. https://doi.org/10.1109/ACCESS.2020.3036370
[6] Perea, Jose A., and John Harer. "Sliding windows and persistence: An application of topological methods to signal analysis." *Foundations of Computational Mathematics* 15 (2015): 799-838. https://doi.org/10.1007/s10208-014-9206-z
[7] Zomorodian, Afra, and Gunnar Carlsson. "Computing persistent homology." In *Proceedings of the twentieth annual symposium on Computational geometry*, pp. 347-356. 2004. https://doi.org/10.1145/997817.997870
[8] H. Edelsbrunner and J. Harer, "Computational Topology," *Open Probl. Topol. II*, pp. 493–545, 2010https://doi.org/10.1016/B978-044452208-5/50049-1
[9] J. A. Perea, "an Application of Topological Methods To Signal," vol. 1, no. 919, pp. 1–34, 2013.
[10] Myers, Audun, Elizabeth Munch, and Firas A. Khasawneh. "Persistent homology of complex networks for dynamic state detection." *Physical Review E* 100, no. 2 (2019): 022314. https://doi.org/10.1103/PhysRevE.100.022314
[11] Gobithaasan, R. U., Zabidi Abu Hasan, Krithana Devi Selvarajh, Khai-Sam Wong, Shukri Mamat, Mohd Zaharifudin Muhamad Ali, Kenjiro T. Miura, and Pawe Dotko. "Clustering Selected Terengganu's Rainfall Stations Based on Persistent Homology." *Thai Journal of Mathematics* (2022): 197-211.
[12] Munkres, James R. *Elements of algebraic topology*. CRC press, 2018. https://doi.org/10.1201/9780429493911
[13] Zomorodian, Afra, and Gunnar Carlsson. "Computing persistent homology." In *Proceedings of the twentieth annual symposium on Computational geometry*, pp. 347-356. 2004. https://doi.org/10.1145/997817.997870
[14] Vietoris, Leopold. "Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen." *Mathematische Annalen* 97, no. 1 (1927): 454-472. https://doi.org/10.1007/BF01447877
[15] Edelsbrunner, Letscher, and Zomorodian. "Topological persistence and simplification." *Discrete & Computational Geometry* 28 (2002): 511-533. https://doi.org/10.1007/s00454-002-2885-2
[16] Chung, Yu-Min, and Austin Lawson. "Persistence curves: A canonical framework for summarizing persistence diagrams." *Advances in Computational Mathematics* 48, no. 1 (2022): 6. https://doi.org/10.1007/s10444-021-09893-4
[17] Bubenik, Peter. "Statistical topological data analysis using persistence landscapes." *J. Mach. Learn. Res.* 16, no. 1 (2015): 77-102.
[18] Takens, Floris. "Detecting strange attractors in turbulence." In *Dynamical Systems and Turbulence, Warwick 1980: proceedings of a symposium held at the University of Warwick 1979/80*, pp. 366-381. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. https://doi.org/10.1007/BFb0091924
[19] Sanyal, Joy, and Xi Xi Lu. "Application of remote sensing in flood management with special reference to monsoon Asia: a review." *Natural Hazards* 33 (2004): 283-301. https://doi.org/10.1023/B:NHAZ.0000037035.65105.95