Streamflow Patterns in a Mountain River at Low and High Frequency Scales and Assessment of Flood events Using Information and Complexity Theory

M. Basel Al Sawaf[1] and Kiyoshi Kawanishi[2]

[1]Hiroshima University
[2]Hiroshima Daigaku - Higashihiroshima Campus

May 5, 2020

Abstract

The selection of a powerful measure to characterize and describe the inputs and outputs of mountainous rivers is of prime importance. Information and complexity metrics have the ability to reveal invaluable information about the hydrological processes that occur within a system. In this work, the hourly streamflow records obtained from five gauging stations for a mountainous river were analyzed to quantify the different patterns and characterize system states at low and high frequencies using increasing aggregation lengths. In addition, we proposed a new extension for the information and complexity theory to be customized for flood assessment. Moreover, we clarified how a pattern (i.e. a word length) by means of information and complexity metrics can be suitably defined. Regarding the low frequency analyses, the information and complexity metrics showed that river discharge has two scaling regimes one of them may describe the river memory characteristics. Furthermore, for high frequency findings, an additional scaling regime that occurs within hourly scales captured by streamflow data obtained by a novel hydroacoustic system, which is one of the novel aspects of our work. Additionally, the power spectral density results match with our findings. This work reveals the performance of information and complexity metrics to be customized for analyzing streamflow patterns at different temporal scales.

Introduction

Precipitation and river discharge are among the most important variables of the global water cycle as they reflect a holistic image about the hydrological processes happing within and over a river basin. Precipitation has enormous spatiotemporal variations which pose a great challenge to maintain sufficient estimates (Ji and Kang, 2015). Streamflow monitoring, on the other hand, is extremely useful to address various water-related applications.

The climate records of the past decades had documented evolving recurrent extreme weather events and natural disasters around the world (Arnell, 1999; Nohara *et al.* , 2006), subjected with negative socioeconomic implications (Wang and Zhang, 2018). Climate change results in a raise in atmospheric temperature coupled with obvious alteration to precipitation patterns (Labat *et al.* , 2004; Gupta*et al.* , 2015). As a consequence of climate change which is evolving apparently in the long run, the frequency of natural disasters such as typhoons, severe tropical cyclones, floods and droughts have been intensified (Hein *et al.* , 2019). Moreover, it was reported by the United Nations International Strategy for Disaster Reduction (UNISDR), between 1998 and 2017, that 91% of all documented disasters in the whole world were induced by extreme weather events including floods, droughts, heatwaves, etc. (Wallemacq and Below, 2017).

In fact, the Japanese Archipelago has a distinctive position where numerous natural disasters happen frequently including seismic, volcanic activities, tsunamis, typhoons, and floods mainly due to being located in the Ring of Fire (Shimokawa *et al.* , 2016). These natural disasters degrade the national sustainable development aspirations and pose additional serious barrier for Japan as it faces multiple challenges fundamentally in terms of population shrinking and declining skilled workforce. Therefore, natural disaster prevention and mitigation framework must be carried out and implemented by a national committee, and hence the role of a multi-disciplinary team is to understand vulnerability and how to minimize and overcome the potential adverse implications (Alcántara-Ayala, 2002). As a result, additional contributions are still required to investigate and address how ecosystems are directly and/or indirectly modified by various hydrological processes during normal and extreme climates.

For hydrological and water resources studies, it is vital to understand streamflow properties induced by extreme climate response during both short-term (few hours) and long-term (several days to several years) and impact of human activities as well. Indeed, diverse parameterization methods have been developed to characterize river discharge patterns and identify changes and complexity in them (Pan *et al.* , 2012; Stosic*et al.* , 2018).

Mountain rivers over the world have a vital role in in maintaining water ecology and conserving biodiversity, as well as, their key functions in flood control (Chen*et al.* , 2019). However, mountain rivers are significantly vulnerable to problems associated with heavy rains during short time. This could be attributed to the fact that stream velocity in mountain regions can vary within a system and subjected to chaotic turbulence (Mihailović *et al.* , 2014). Therefore, investigating the fluctuation and complexity of flow properties for mountain streams will deliver profound understanding about streamflow patterns and their corresponding responses influenced by hydrologic climates and/or human activities.

Considering the future scenarios of streamflow in East Asia, it was projected in the literature that there will be an increase in river discharge by the coming decades (e.g. Arnell, 1999; Nohara et al., 2006). Sato et al., (2012) inferred that at the end of this century river flow will rise due to increases in precipitations. Likewise, Higashino and Stefan (2019) concluded that the annual maximum discharge in Japanese streams are expected to be increased.

Certainly, heavy seasonal rainfall that occur frequently in the western part of Japan is among the worst destructive disasters, since it accompanies by landslides and mudflows. Furthermore, out of all Japanese prefectures, Hiroshima was ranked as the first prefecture in Japan that has the highest number of mountainous slopes ($\sim$32,000) to be susceptible to landslide and mudflow disasters (Tsuchida et al., 2014), followed by Shimane and Yamaguchi prefectures 22,300 and 22,250, respectively. Due to the frequent huge precipitation and associated sudden landslide and mudflow disasters, for the time being, the Ministry of Land Infrastructure and Tourism (MLIT) of Hiroshima prefecture installed a network of real-time water level measurement that collects measurements at multiple sites over Hiroshima's streams, aiming to build up a profound knowledge about the different characteristics related to streamflow response during various rainy events and hence to mitigate the potential risk accompanied with heavy precipitation.

In the recent years, information-based theories have received increased interest in the hydrological studies to detect and address the variability in numerous hydrological variables including precipitation, temperature, and streamflow (e.g. Brunsell, 2010; Elsner and Tsonis, 1993; Koutsoyiannis, 2005; Mishra et al., 2009). Pan et al., (2012) documented the benefits of information-based metrics in their capability to interpret how a model presents patterns of information content and complexity exist in hydrological dataset. Indeed, considerable efforts had shown the applicability of the information and complexity measures to characterize the various patterns in time series analysis. In particular, (Pachepsky *et al.* , 2006, 2016; Pan *et al.* , 2011, 2012), extensively utilized the information and complexity metrics to characterize various soil moisture, streamflow, and rainfall time series using a straightforward symbolic strings approach of 2 characters length per word for system description which is very useful but uncomplicated classification. Nonetheless, there is no work had discussed the importance of considering complex patterns of words to characterize different system states. In other words, how to recommend using short or long length of words to describe different patterns embedded in a hydrological system. In addition, there is almost no work that clearly highlighted the transformation of a system from a state to another especially during short and long terms.

Accordingly, one of the fundamental research questions that we aim to answer is what are the information and hidden hydrological phenomena that can be detected by characterizing streamflow patterns using more complex patterns according to information and complexity theory and how to define the appropriate pattern length that describe the different potential states of a system (dataset, time series, etc.). Therefore, one of the main contributions of the present research is to shed light on streamflow variations in a mountainous river and the nested relationships within its tributaries located at Hiroshima prefecture that has been extremely and repeatedly deteriorated from severe floods. The particular novelty is to examine temporal streamflow patterns at high-frequency scales using real discharge data obtained from both classic and novel hydroacoustic system, also at low-frequency that happen over a basin and sub-basin scales. We also proposed a new extension for the information-based metrics to assess streamflow patterns during flood periods. After describing the monitoring sites in section 2, the methods are given in section 3. Results and discussion are provided in sections 4 and 5, respectively. Eventually, section 6 shows the research conclusions.

Observation site and streamflow dataset description

This research considers the Gōno River which is the largest gravel-bed mountainous river runs through Hiroshima and Shimane prefectures, west of Japan. The watershed of the Gōno River is influenced by cool temperate climate with four obvious seasons: winter (December-February), spring (March-May), summer (June-August), and autumn (September-November). Basically, precipitation happens in winter, however, heavy rainfall occurs in the monsoon (June and July) as well as during typhoon season (August and September). The catchment area of the Gōno River is 3963 km^2 and divided into four sub-watersheds, additionally, the Gōno River has two major tributaries namely, the Basin and Saijo Rivers (Fig. 1). The Gōno River watershed is monitored by the MLIT using four real-time gauging stations (Fig. 1) that measure the water stage (H) directly and discharge (Q_{RC}) indirectly by means of Rating Curve (RC) equations developed empirically using the general quadratic equation as follows (Kawanisi *et al.* , 2016; Higashino and Stefan, 2019):

$Q_{\mathrm{RC}} = (c_1 H + c_2)^2$ (1)

where c_1 and c_2 are constants, which are empirically computed from calibrations with direct discharge measurements accomplished regularly (Kawanisi et al., 2016). More details about the general features of the sub-watersheds and gauging stations are given in Table 1. Remarkably, the water depth at the river sites turn out to be dramatically shallow under low discharge conditions.

Table 1 is here.

Table 1 The general properities of the Gōno River at the studided gauging stations constructed by the MLIT.

Station Name	Sub-basin Area (km^2)	Mean Annual Flowrate (m^3/s)	Designated Maximum Water Level (m)/ corres
Awaya	671	16	4.5/375
Minamihatachiki	680	189	3.5/390
Miyoshi	631	37	3.0/590
Ozekiyama	1981	70	6.0/1000

In this study, the hourly streamflow data records by means of RC method from 2002 to 2017 were obtained from the abovementioned stations and analyzed.

Figure 1 is here.

Fig. 1 The Gono River and its tributary (Basen and Saijo Rivers), the location of the MLIT gauging stations (Green dots) and the position of the two acoustic stations (T1 & T2) of the FAT system (Red dots).

In addition, Kawanisi et al., (2016, 2018) performed long-term streamflow measurements that was located very close to Ozekiyama station (Fig. 1) using the Fluvial Acoustic Tomography (FAT) system. Hence, as a novel feature for this work, the available hourly discharge records measured by FAT from 2016-01 to 2016-06 was used for further comparison with RC records. Past works had deeply discussed the measurement principles and discharge accuracy by means of the FAT in details (Kawanisi *et al.* , 2013, 2016, 2018; Razaz *et al.* , 2013; Bahreinimotlagh *et al.* , 2016; Al Sawaf and Kawanisi, 2019), thus this work does not aim to repeat the previous works, rather it considers the reliable records of streamflow data as observed by the FAT system for further analysis. However, it is vital to point out that discharge measured by the FAT is computed using the main flow equation in open channels as:

$Q_{\mathrm{FAT}} = u \ \times \ A \times \tan\theta$ (2)

where u and A are the cross-sectional averaged velocity and oblique cross-sectional area along transmission line, respectively, and ϑ is the flow angle. As can be seen in Eq (2), unlike the discharge estimated by the RC approach, the discharge computed by FAT comprises both velocity and area (stage) terms.

Methods

Streamflow analysis by information and complexity theoretic measures

Following the symbolic strings approach proposed by Wolf, (1999), the hourly streamflow data was converted into a binary sequence (i.e. 0 or 1). In brief, the performed analyses can be accomplished considering the following steps:

Figure 2 is here.

Fig. 2 Illustration for the symbolic strings method, a) the basic approach, b) illustration for data aggregation using Aggregation Length (AL=2).

The first step is to determine the median value from each streamflow time series. The next step is to map each value of the streamflow time series to 0 if it is below or at the median otherwise it is mapped as 1. Once the corresponding binarized time series is created, we define a window length L ($L \in \mathbb{N}$) (alternatively word length) composed of L consecutive symbols. Thus, the different possible words that can be encountered in a studied system are 2^L . As illustrated in Fig. 2a, if the defined word length is (L =2), the possible words that can be encountered are 00, 01, 10, and 11, and hence, each word describes a state of the system. The next step as stated by Wolf, (1999), is to find the primary ingredients to evaluate information and complexity-based metrics. In other words, this step means to find the three sets of probabilities as: i) $p_{L,i}$: the state probability of the i -th L word where $i = 1, 2, \ldots\ , 2^L$; ii)$p_{L,\ ij}$: is the probability of shifting from the i -th to the j -th L word instantly, where $i = 1, 2, \ldots\ , 2^L$ and $j = 1, 2, \ldots\ , 2^L$; and iii) $p_{L,\ i\to j}$: is the conditional probability for the occurrence of the incidence$j$ -th L word, given that the i -th L word event has been observed before. Once the aforementioned probabilities are determined the information and complexity-based metrics can be estimated.

In the case of this work, we will tackle two information measures as well as two complexity measures, the mean information gain and metric entropy were used to measure the information content in our data. On the other hand, the effective measure of complexity and fluctuation complexity were selected as metrics to quantify the complexity content in our streamflow data, the aforementioned metrics are explained below.

Basically, the information entropy proposed by (Shannon, 1948), is a popular measure that quantifies the randomness in a dataset. While the Shannon entropy is given by Eq (3), the Metric Entropy is equal to Shannon entropy (H_S) divided by the word length (L), thus it is a normalization of the Shannon entropy that gives us an image about the contained information in a dataset but at the same time it is independent from the word length (L).

$H_S = - \sum_{i=1}^{n} p_{L,i}$ Log$_2$ $p_{L,i}$(3)

The metric entropy is zero for steady sequence of data, conversely, it increases in a monotonic behavior as the sequence's disorder increases and amounts to its maxima at 1 for evenly distributed random sequences.

Alternatively, the Mean Information Gain (MIG), is a measure of entropy (randomness) that quantifies the amount of information and is defined as the mean amount of information that can be gained about a dataset and given as:

$MIG = - \sum_{i,j=1}^{2^L} p_{L,ij}Log_2$ p$_{L,i\rightarrow j}$(4)

The above equation can be also expressed as the difference of Shannon entropies as:

$MIG = H_S(L+1) - H_S(L)$ (5)

Pachepsky et al., (2016) pointed out that larger values of information gain refer to the greater chance of a system to vary from one state to another.

Instead, the complexity metrics are helpful measures that permit to capture the existence of internal patterns in studied datasets (Pan*et al.* , 2012). The effective measure of complexity (EMC) as stated by Grassberger, (1986) is the least quantity of information has to be amassed required to deliver best possible prediction of the next data element, the effective measure of complexity can be approximated and computed using Eq (6):

$EMC \approx (L+1)$ H$_S(L) -$ L H$_S(L+1)$ (6)

Alternatively, EMC and can be also evaluated using Eq (7) as:

$EMC \approx \sum_{i,j=1}^{2^L} p_{L,ij}$ Log$_2 \frac{p_{L,i\rightarrow j}^{L}}{p_{L,i}}$(7)

Finally, the fluctuation complexity (σ_Γ^2) is one of the most important complexity measures since it defines the fluctuations that occur in a system, i.e. how a system transforms from a pattern to another. The fluctuation complexity is, therefore, a measure for the changes of the net information gain over one or more-time steps. Hence, data that pose a high-level of fluctuation yields larger fluctuation complexity (Bates and Shepard, 1993), the fluctuation complexity is estimated as:

$\sigma_\Gamma^2 = \sum_{i,j=1}^{2^L} p_{L,ij}$ (Log$_2 \frac{p_{L,i}}{p_{L,j}})^2$(8)

Temporal discharge characteristics using information and complexity measures within low-frequency and high- frequency scales

Temporal discharge characterization by means of information and complexity theory was performed at different time domains using growing aggregation lengths (AL). To illustrate, considering a word length of two characters, hence, in the case of AL=1 (basic approach), it means that each hour for a studied streamflow record was substituted directly as one charter to form a part of a word (Fig. 2a). Alternatively, in the case of AL=2, it means that each two successive hours from each studied streamflow record were gathered, averaged, binarized and then substituted as one character to compose a word, for more clarification see Fig. 2b.

In the present research, low-frequency and high-frequency indicate streamflow variations using information and complexity measures over long and short time scales, respectively. In fact, the availability of high-quality of streamflow data (roughly 6 months) measured by means of the FAT close to Ozekiyama gauging station (Fig. 1), offers a unique opportunity to examine the temporal variations of streamflow patterns over short periods. Hence, high-frequency analyses concentrate on the hourly discharge data from January 2016 to the end of June 2016 at Awaya, Minamihatachiki, Miyoshi, Ozekiyama, and FAT stations. Whereas, low-frequency analyses comprise hourly flow data from January 2002 to the end of December 2017 covering Awaya, Minamihatachiki, Miyoshi, and Ozekiyama stations.

For both low and high frequency analyses we used 4 characters-based to describes the different states of system patterns, thus the various potential patterns are $2^L = 2^4 = 16$ probable words.

An extension for the information and complexity measures to examine flood events.

Herein, we provide an extension for the information and complexity theory aiming to assess the annual variation of flood events. According to the abovementioned symbolic strings method in (section 3.1), the Q_{Median} value of each streamflow was used as a threshold to map each streamflow data either to 0 if its equal or less the median and 1 if it is greater than the median. In this approach, the same described procedures were performed, nevertheless, the two changes are i) the maximum daily discharge was taken and investigated instead of hourly discharge, and ii) the threshold discharge values for the observed stations have been changed as presented in Table 1, thus if a maximum daily discharge value is equal or greater than the threshold is converted to 1, otherwise it is assigned as 0. In the case of flood assessment, the word length was set to 2 characters and hence 4 possible words describe the different patterns of the system, additionally, one information and one complexity metrics used in these analyses namely, the metric entropy and effective measure of complexity.

Results

Short term and long-term streamflow observations

Discharge time series for the Gōno River observed at the studied stations are shown in Fig. 3. The Gōno River has a flashy regime, influenced considerably by the intensity of precipitation. Apparently, it can be realized that the discharge records reveal some years with low amount of flow rates. For example, 2002 to 2003, 2007 to 2009, and from 2015 to 2017. On the other hand, it can be distinguished that there is an upward trend in the maximum streamflow records induced by the large amount of precipitated rains. In other words, it is obvious that the maximum streamflow in 2006 peaked at 3300 m^3/s approximately, and the maximum peaks observed at 2010 and 2014 were roughly 3500 and 3600 m^3/s, respectively, Though it was not reported by the MLIT yet, in 2018 heavy rains documented the greatest flood that occurred in Hiroshima prefecture.

Figure 3 is here.

Fig. 3 Discharge time series during the study period (2002-2017) observed at Awaya (red), Minamihatachiki (green), Miyoshi (blue), and Ozekiyama (Black) stations.

Prior to information and complexity results, it is essential to briefly comment on the discharge measurements by FAT (Q_{FAT}) compared to the RC (Q_{RC}) estimates at Ozekiyama observation site. Demonstrated in Fig. 4, a comparison between $Q_{FAT}$and Q_{RC} , it can be seen that both discharge methods show very good agreement. Furthermore, it can be noticed that FAT can capture the fluctuations that take place during very short times scales, this feature in particular, inspired us to profoundly explore this difference. Kawanisi et al., (2018, 2016) investigated the accuracy of river discharge of Q_{FAT} compared to other discharge computation methods (e.g. RC and ADCP) measured at Ozekiyama station and examined the error structure that impair the performance of FAT estimates, it was demonstrated that in the case of streamflow measurements by means of FAT at the Gōno River, the maximum potential error within low-flow circumstances is estimated as 15%.

Figure 4 is here.

Fig. 4 A Comparison between the discharge performance computed by Q_{FAT} (purple) at Q_{RC}(black) at Ozekiyama observation site.

Streamflow properties during low and high frequency scales using information and complexity measures

Depicted in Fig. 5 the computation results for the information and complexity metrics evaluated for the studied stations during low frequency scales. Apparently, the information metrics findings (Fig. 5 (a, b)) show the presence of two scaling regimes for the examined discharge records. The first scale has a very steep upward trend at short times up to AL=20 hours approximately (zoomed in Fig. 7c), whereas, the second upward trend has a mild slope for longer times. In a similar manner, the complexity contents show a comparable behavior, i.e. in the case of effective measure of complexity (Fig. 5 (c)), a sharp downward trend at short aggregation lengths followed by a moderate slope which is identically opposite to the results of the information measures. On the other hand, the estimate results of the fluctuation complexity (Fig. 5d) have a shape of $\bigwedge$ peaked roughly at AL [?] 10-15 hours (see Fig. 7d), with a gradual descending at long ranges (Fig. 5d).

Figure 5 is here.

Fig. 5 Temporal scales of low frequency streamflow variations according to and complexity measures recorded at each station

Outstandingly, it can be noticed that as the aggregation length increases, the relationship between metric entropy and information complexity follow the Bernoulli distribution as presented in the information-complexity diagram (Fig. 6), peaked at AL [?] 15 hours.

Figure 6 is here.

Fig. 6 Information-Complexity diagram for the streamflow data according to different aggregation length.

To give a conceptual understanding about the high frequency findings, metric entropy and fluctuation complexity according to different aggregation lengths are displayed in Fig. 7 (a, b). Since the investigated period is relatively short (i.e. the hourly data from 2016-01 to 2016-06), the results, will not be sufficiently informative, especially, for the long ranges. However, the comparison manifests some interesting outcomes. First, the computed metric entropy (Fig. 7a), shows that there is a notable variation between the streamflow data records obtained by Ozekiyama, and FAT (since both of these stations are located at the same site). More importantly, the estimated information content by means of FAT has higher values compared to RC, particularly, for AL $\leq$ 4 hours suggesting that there is an additional scaling regime occurs during sub-daily scales (i.e. few hours), and hence the FAT is capable to capture the streamflow fluctuations that occur during hourly scales. Apparently, both high and low frequencies confirm that the information contents (i.e. metric entropy) at small aggregation lengths have consistent slope as streamflow computed by means of RC approach (Fig. 7a, and 7c). Alternatively, the fluctuation complexity estimates (Fig. 7b), demonstrates that there is remarkable difference between FAT and Ozekiyama estimates, therefore, it is advised to consider high resolution streamflow records to accurately investigate the hidden phenomena that cannot be observed by conventional discharge calculation methods.

Figure 7 is here.

Fig. 7 Information and complexity contents for high frequency (a, b) and low frequency (c, d).

Flood assessment by information and complexity metrics

The results of the flood analysis by means of the new customized information and complexity method are revealed in Table 2 and Table 3. In this work, the flood frequency was examined year by year to compare the annual variations of floods. Also, as previously pointed in section 3.3, the word length was set to 2 characters (i.e. 2 successive days), thus 4 possible words are expected.

Table 2 is here.

Table 2 The number of different words observed at each station during flood assesment.

	Word Number	Word Number	Word Number	Word Number
Year	Awaya	Minami Hatachiki	Miyoshi	Ozeki Yama
2002	1	1	1	1
2003	3	3	4	4
2004	4	4	3	4
2005	3	3	3	3
2006	4	4	4	4
2007	1	1	1	1
2008	1	1	1	1
2009	3	4	1	3
2010	4	4	3	4
2011	3	4	4	4
2012	3	3	3	3
2013	3	3	3	4
2014	3	3	3	3
2015	1	1	1	1
2016	3	3	1	3
2017	4	4	3	4

Table 2 conveys information about the number of possible words that appeared each year at each station. As can be understood, during the study period (2002-2017), in 2002, 2007, 2008, and 2015 only one pattern of words was observed (i.e. 00) which means that the maximum discharge in each single day throughout those years was less than the threshold. In contrast, in 2004, 2006, 2010, and 2017, four different words were observed, telling that the occurred floods were greater than the threshold value and high enough so that persisted for at least two consecutive days. Table 3, on the other hand, reports the metric entropy and the effective measure complexity contents. The results in Table 3 shows that in 2002, 2007, 2008, and 2015, the computed metric entropy was 0, which means that there was no entropy (randomness), and hence, there is neither information nor complexity. It can be distinguished that if three different words were observed, it yields that at least one day flood event occurred, correspondingly, the other values of information and complexity contents give additional descriptions about the frequency and amount of the occurred floods. Generally, during the years of this study, three different words were reported (i.e. 00, 01, 10), and in some years 4 days indicating that floods in this mountainous region are an important challenge and hence there are additional actions must be taken to mitigate and minimize flood implications.

Finally, to elucidate the efficiency of this method, let's consider the results recorded at the Ozekiyama station as an example. In one hand, in 2004, it was reported that three flood events were recorded as (11) and another three events as (01) and similarly three events as (10). The estimated metric entropy and effective measure of complexity as can be seen in Table 3 are 0.1 and 0.07, respectively. on the other hand, in 2006, it was documented that two flood events were recorded as (11) and three flood events as (01) also another three events as (10). In this case, the metric entropy and the effective measure of complexity as estimated in Table 3 are 0.09 and 0.03 respectively, which reflect the merit of this approach to sort out the different information about the documented floods.

Table 3 is here.

Table 3 Information and complexity metrics computed at each station for flood assesment approcah.

	Metric Entropy	Metric Entropy	Metric Entropy	Metric Entropy	Effective Measure Complexity	Effective
Year	Awaya	Minami Hatachiki	Miyoshi	Ozeki Yama	Awaya	Minami
2002	0	0	0	0	0	0
2003	0.03	0.05	0.04	0.06	0	0
2004	0.08	0.11	0.05	0.1	0.01	0.03
2005	0.03	0.03	0.05	0.03	0	0
2006	0.11	0.09	0.07	0.09	0.03	0.03
2007	0	0	0	0	0	0
2008	0	0	0	0	0	0
2009	0.03	0.06	0	0.05	0	0.02
2010	0.09	0.09	0.05	0.09	0.06	0.06
2011	0.05	0.06	0.08	0.06	0	0.02
2012	0.03	0.03	0.03	0.03	0	0
2013	0.03	0.05	0.03	0.06	0	0
2014	0.05	0.03	0.03	0.03	0	0
2015	0	0	0	0	0	0
2016	0.05	0.05	0	0.05	0	0
2017	0.06	0.09	0.05	0.11	0.016	0.03

Discussion

This study shows the effectiveness of streamflow analysis using the information and complexity theory to detect different discharge patterns that may describe some hydrological processes during low and high frequency scales as well as during flood assessment. Indeed, the selection of suitable metrics to quantify streamflow patterns at different temporal scales is of paramount importance. Interesting issues were emerged from this research and discussed below.

The role of word length and word number in characterizing streamflow patterns

The importance of word length and word number, to the best of our knowledge, was not highlighted sufficiently in the literature. Pachepsky et al., (2016), indicated that investigating the role of word length efficiency to improve the information and complexity metrics would open new avenue for further explorations. Thus, the important question here that we can ask ourselves is what is the recommended word length that should be selected to process the symbolic strings method professionally. In fact, there is no direct rule available to advise an ideal word length that generates the finest results in information and complexity theory. However, since this work investigates the classification and characterization of various streamflow patterns during different scales, we suggested two types of word length (two and four characters). In the case of low and high frequency scales, streamflow patterns are mainly affected by the intensity of various rainfall events, hence, it is vital to examine the influence of different intense events on the discharge data. In this regard, we considered the hydrograph separation method proposed by (Raghunath, 2006) depicted in Fig. 8 to separate the hydrograph into runoff flow and baseflow. As a result, in the case of the Gōno River watershed, the number of N days after peak for the streamflow to get rid of a rainfall inputs are 4 days approximately and hence, we used 4 characters as a word length. Thus, in the case of AL=24 hours, it means that each 24 hours of streamflow records (i.e. 1-day discharge data) were grouped together, and then averaged to form a character, and correspondingly the word pattern is formed from 4 days to describe the state of a system for each 4 successive days.

Figure 8 is here.

Fig. 8 Graphical hydrograph separation method: For the Gōno River catchment $N \approx 4$ days.

In the case of flood assessment, information and complexity measures should be customized in such a way that describe the patterns of floods in terms of occurrence, frequency, etc., moreover, it should be emphasized that it is very important to define the appropriate value for $Q_{threshold}$. In this work, we adapted the corresponding discharge value of the designated maximum water level introduced by the MLIT for each gauging station as a threshold and we used the maximum daily discharge data. According to the MLIT, this level is used as a guide for municipal mayors to issue evacuation warns, also is used as a reference for evacuation decisions by the local residents, etc.

We believe that employing two characters as a word length is, therefore, suitable for the assessment of different future flood scenarios. Increasing the word length to describe flood patterns, is not useful in our opinion, because we assume that having high floods for more than two days means a great natural and national disaster, bearing in mind that we are considering the maximum daily discharge in our analyses. The proposed analyses in this work suggest that the characterizing system patterns by means of information and complexity measures could be customized to be used in different ranges of applications.

In the case of flood assessment, one of the most important applications of considering different word patterns is to propose new contour inundation maps and/or hazard maps for the different discharge gauging stations, to support policy makers to improve their understanding and choose better decisions and alternatives for the related issues and future projects.

Inferences from low and high frequency analyses

Quantifying streamflow patterns by means of information and complexity metrics and addressing different aggregation lengths revealed various interesting behaviors of streamflow during low and high frequencies. Regarding low frequency findings, it can be seen that using different aggregation lengths, the information metrics (metric entropy and mean information gain) for streamflow data recorded at the studied stations have obviously two scaling regimes. The first one with steep slope for shorter AL ranges, and the other one for longer AL ranges. In fact, this finding matches with the results of Al Sawaf et al., (2017) who studied the discharge fluctuations in the Gōno River by means of Detrended Fluctuation Analysis (DFA) and reported of the presence of two scaling regimes of the river discharge fluctuations separated by a crossover time observed around 3-5 days. To compare, it can be noticed in Fig. 5(a & b) the long AL ranges (i.e. AL greater than 20 has a mild slope which is similar to the outcomes of DFA results indicating that this range may reveal the long-memory characteristics of the river flow fluctuations. Of interest, both information and complexity contents evaluated for the studied stations showed similar crossover times detected roughly at AL$\approx$ 20 hours equivalent to 80 hours (see Fig. 7(c&d)). However, one of the challenging tasks in DFA or spectral analyses is to find the crossover time accurately. In the case of these approaches, the crossover time is usually estimated by performing a linear regression fit to the suspected regimes separately, thus, the intersection point of the two fitting lines composes the crossover time. Nevertheless, the findings revealed that crossover times may be estimated from the corresponding aggregation length time where the fluctuation complexity value reaches its peak according to the information-complexity diagram as can be seen in Fig. 6 (Also, refer to the Table 1 in the supplementary materials). In the case of the Ozekiyama station, the crossover time observed at AL= 14 hours equivalent to 56 hours, i.e. the crossover time (56 hours) = aggregation length (14) * word length (4 characters). Therefore, further investigations are still required to interpret and decipher the nested relationships between the information metrics and fractal analysis.

Regarding the high frequency analyses, an interesting phenomenon was observed namely the presence of an extra scaling regime that occurs during sub-daily scales captured by FAT records can be observed at AL$\leq$ 4 equivalent to 16 hours. To verify the existence of this scale, we estimated the power spectrum for the discharge records obtained by both Ozekiyama and FAT stations using the proposed model by Dolgonosov et al., (2008), presented in Fig. 9. As can be distinguished, the spectral analysis shows that both RC and FAT data have two main scaling denoted by S1 for long ranges that are roughly quite similar, and S2 for mid ranges (Fig. 9) with a crossover time around 60 hours which is very near to AL=14 (i.e. 56 hours). Nevertheless, it can also be realized that the presence of a specific slope captured by FAT data namely S3 which is somehow near to AL=4 (i.e. 16 hours). This finding seems to confirm our hypothesis about the existence of an additional scaling regime happens within very short time scales and can be captured by FAT as can be seen in Fig. 7a. However, the slight variation may be confirmed by comparing with another scaling method (fractal analysis) or considering shorter word length for high frequency analysis (e.g. 3 characters per word).

The last remaining question is why there is an additional regime that was captured by means of FAT. Though it needs further exploration to clearly describe this phenomenon, it can be said that the FAT system measures the discharge according to the fundamental discharge equation as given in Eq (2). Unlike discharge estimated by means of the RC method, the velocity and area (depth) terms are embedded directly in streamflow estimates and hence FAT can clearly show the high resolution of discharge estimate.

Figure 9 is here.

Fig. 9 Power spectra results for $Q_{Ozekiyama}$ (black) and Q_{FAT} (purple) during (2016/01 to 2016/06).

Conclusions

Describing and characterizing the inputs and outputs discharge time series is of greatest importance. This paper utilizes the information and complexity theory due to its ability to deliver useful information about the hydrological processes that occur within a system. The hourly discharge records obtained from five gauging stations for a mountainous river were analyzed to quantify the different patterns and characterize system states at low and high frequencies using increasing aggregation lengths. Furthermore, new extension for the information and complexity theory to be customized for flood assessment was proposed. Several interesting issues was revealed and learned from this research are summarized below.

Firstly, considering the analyses by means of the information and complexity theory, it is vital to define an appropriate word length professionally since word pattern play an important role in describing system states and the hidden regimes and structures. Secondly for low frequency analysis, by means of increasing aggregation lengths, it was detected the presence of two scaling regimes one for the short aggregation lengths and the other one for long ranges that may reflect the long memory characteristics of river flow fluctuations. Alternatively, in the case of high frequency analyses, it was confirmed that river fluctuations have extra sub daily (hourly) regime that was captured by streamflow data obtained by FAT. Moreover, the power spectral density analysis confirms our findings. In conclusion, this work reveals the proficiency of information and complexity metrics to be customized for streamflow analyses.

Acknowledgments

This study was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI grant number JP17H03313.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Data Availability Statement

Streamflow data as computed by Q_{RC} was provided by the Ministry of Land Infrastructure Transport (MLIT), Japan. Whereas Q_{FAT} of this study are available from the corresponding author, upon reasonable request.

References

Alcántara-Ayala I. 2002. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. *Geomorphology* DOI: 10.1016/S0169-555X(02)00083-1

Arnell NW. 1999. Climate change and global water resources. In *Global Environmental Change* . DOI: 10.1016/S0959-3780(99)00017-5

Bahreinimotlagh M, Kawanisi K, Danial MMMM, Al Sawaf MBMB, Kagami J. 2016. Application of shallow-water acoustic tomography to measure flow direction and river discharge. *Flow Measurement and Instrumentation* **51** : 30–39 DOI: 10.1016/j.flowmeasinst.2016.08.010

Bates JE, Shepard HK. 1993. Measuring complexity using information fluctuation. *Physics Letters A* DOI: 10.1016/0375-9601(93)90232-O

Chen Q, Guo B, Zhao C, Zhang J. 2019. A comprehensive ecological management approach for northern mountain rivers in China. *Chemosphere* **234** : 25–33 DOI: 10.1016/j.chemosphere.2019.06.042

Dolgonosov BM, Korchagin KA, Kirpichnikova N V. 2008. Modeling of annual oscillations and 1/f-noise of daily river discharges. *Journal of Hydrology* **357** (3–4): 174–187 DOI: 10.1016/j.jhydrol.2008.04.023

Grassberger P. 1986. Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics* **25** (9): 907–938 DOI: 10.1007/BF00668821

Gupta SC, Kessler AC, Brown MK, Zvomuya F. 2015. Climate and agricultural land use change impacts on streamflow in the upper midwestern United States. *Water Resources Research* DOI: 10.1002/2015WR017323

Hein W, Wilson C, Lee B, Rajapaksa D, de Moel H, Athukorala W, Managi S. 2019. Climate change and natural disasters: Government mitigation activities and public property demand response. *Land Use Policy* **82** : 436–443 DOI: https://doi.org/10.1016/j.landusepol.2018.12.026

Higashino M, Stefan HG. 2019. Variability and change of precipitation and flood discharge in a Japanese river basin. *Journal of Hydrology: Regional Studies* **21** : 68–79 DOI: 10.1016/j.ejrh.2018.12.003

Ji Z, Kang S. 2015. Evaluation of extreme climate events using a regional climate model for China. *International Journal of Climatology* DOI: 10.1002/joc.4024

Kawanisi K, Bahrainimotlagh M, Al Sawaf MBMB, Razaz M, M.B. MB, Razaz M. 2016. High-frequency streamflow acquisition and bed level/flow angle estimates in a mountainous river using shallow-water acoustic tomography. *Hydrological Processes* **30** (13): 2247–2254 DOI: 10.1002/hyp.10796

Kawanisi K, Razaz M, Yano J, Ishikawa K. 2013. Continuous monitoring of a dam flush in a shallow river using two crossing ultrasonic transmission lines. *Measurement Science and Technology* **24** (5): 055303 DOI: 10.1088/0957-0233/24/5/055303

Kawanisi K, Al Sawaf MB, Danial MM. 2018a. Automated Real-Time Streamflow Acquisition in a Mountainous River Using Acoustic Tomography. *Journal of Hydrologic Engineering* **23** (2): 04017059 DOI: 10.1061/(ASCE)HE.1943-5584.0001604

Kawanisi K, Al Sawaf MB, Danial MM. 2018b. Automated real-time stream flow acquisition in a mountainous river using acoustic tomography. *Journal of Hydrologic Engineering* **23** (2) DOI: 10.1061/(ASCE)HE.1943-5584.0001604

Labat D, Goddéris Y, Probst JL, Guyot JL. 2004. Evidence for global runoff increase related to climate warming. *Advances in Water Resources* DOI: 10.1016/j.advwatres.2004.02.020

Mihailović DT, Nikolić-Dorić E, Drešković N, Mimić G. 2014. Complexity analysis of the turbulent environmental fluid flow time series. *Physica A: Statistical Mechanics and its Applications* DOI: 10.1016/j.physa.2013.09.062

Nohara D, Kitoh A, Hosaka M, Oki T. 2006. Impact of climate change on river discharge projected by multimodel ensemble. *Journal of Hydrometeorology* DOI: 10.1175/JHM531.1

Pachepsky Y, Guber A, Jacques D, Simunek J, Van Genuchten MT, Nicholson T, Cady R. 2006. Information content and complexity of simulated soil water fluxes. *Geoderma* DOI: 10.1016/j.geoderma.2006.03.003

Pachepsky YA, Martinez G, Pan F, Wagener T, Nicholson T. 2016. Evaluating Hydrological Model Performance using Information Theory-based Metrics. *Hydrology and Earth System Sciences Discussions* DOI: 10.5194/hess-2016-46

Pan F, Pachepsky YA, Guber AK, Hill RL. 2011. Information and complexity measures applied to observed and simulated soil moisture time series. *Hydrological Sciences Journal* DOI: 10.1080/02626667.2011.595374

Pan F, Pachepsky YA, Guber AK, McPherson BJ, Hill RL. 2012. Scale effects on information theory-based measures applied to streamflow patterns in two rural watersheds. *Journal of Hydrology* DOI: 10.1016/j.jhydrol.2011.10.018

Raghunath HM. 2006. *Hydrology : Principles, Analysis, Design* . NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS: New Delhi.

Razaz M, Kawanisi K, Nistor I, Sharifi S. 2013. An acoustic travel time method for continuous velocity monitoring in shallow tidal streams. *Water Resources Research* **49** (8): 4885–4899 DOI: 10.1002/wrcr.20375

Sato Y, Kojiri T, Michihiro Y, Suzuki Y, Nakakita E. 2012. Estimates of Climate Change Impact on River Discharge in Japan Based on a Super-High-Resolution Climate Model. *Terrestrial, Atmospheric and Oceanic Sciences* **23** (5): 527 DOI: 10.3319/TAO.2012.05.03.02(WMH)

Al Sawaf MBMB, Kawanisi K. 2019. Novel high-frequency acoustic monitoring of streamflow-turbidity dynamics in a gravel-bed river during artificial dam flush. *CATENA* **172** : 738–752 DOI: 10.1016/j.catena.2018.09.033

Al Sawaf MBMB, Kawanisi K, Kagami J, Bahreinimotlagh M, Danial MMMM. 2017. Scaling characteristics of mountainous river flow fluctuations determined using a shallow-water acoustic tomography system. *Physica A: Statistical Mechanics and its Applications* **484** : 11–20 DOI: 10.1016/j.physa.2017.04.168

Shannon CE. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* DOI: 10.1002/j.1538-7305.1948.tb01338.x

Shimokawa S, Fukahori H, Gao W. 2016. Wide-area Disaster Prevention of Storm or Flood Damage and its Improvement by Using Urban Planning Information System. *Procedia - Social and Behavioral Sciences* DOI: 10.1016/j.sbspro.2015.12.065

Stosic T, Stosic B, Singh VP. 2018. q-triplet for Brazos River discharge: The edge of chaos? *Physica A: Statistical Mechanics and its Applications* **495** : 137–142 DOI: 10.1016/j.physa.2017.12.061

Tsuchida T, Kano S, Nakagawa S, Kaibori M, Nakai S, Kitayama N. 2014. Landslide and mudflow disaster in disposal site of surplus soil at Higashi-Hiroshima due to heavy rainfall in 2009. *Soils and Foundations* DOI: 10.1016/j.sandf.2014.06.005

Wallemacq P, Below R. 2017. Economic Losses, Poverty & Disasters 1998-2017. *United Nations Office for Disaster Risk Reduction (UNDRR)*

Wang K, Zhang A. 2018. Climate change, natural disasters and adaptation investments: Inter- and intra-port competition and cooperation. *Transportation Research Part B: Methodological* DOI: 10.1016/j.trb.2018.08.003

Wolf F. 1999. Berechnung von Information und Komplexität in Zeitreihen-Analyse des Wasserhaushaltes von bewaldeten Einzugsgebieten.*Bayreuther Inst. für Terrestrische Ökosystemforschung.***65** : 1–164

Figure legends

Fig. 1 The Gono River and its tributary (Basen and Saijo Rivers), the location of the MLIT gauging stations (Green dots) and the position of the two acoustic stations (T1 & T2) of the FAT system (Red dots).

Fig. 2 Illustration for the symbolic strings method, a) the basic approach, b) illustration for data aggregation using Aggregation Length (AL=2).

Fig. 3 Discharge time series during the study period (2002-2017) observed at Awaya (red), Minamihatachiki (green), Miyoshi (blue), and Ozekiyama (Black) stations.

Fig. 4 A Comparison between the discharge performance computed by QFAT (purple) at QRC (black) at Ozekiyama observation site.

Fig. 5 Temporal scales of low frequency streamflow variations according to and complexity measures recorded at each station.

Fig. 6 Information-Complexity diagram for the streamflow data according to different aggregation length.

Fig. 7 Information and complexity contents for high frequency (a, b) and low frequency (c, d).

Fig. 8 Graphical hydrograph separation method: For the Gōno River catchment N [?] 4 days.

Fig. 9 Power spectra results for $Q_{Ozekiyama}$ (black) and Q_{FAT} (purple) during (2016/01 to 2016/06).

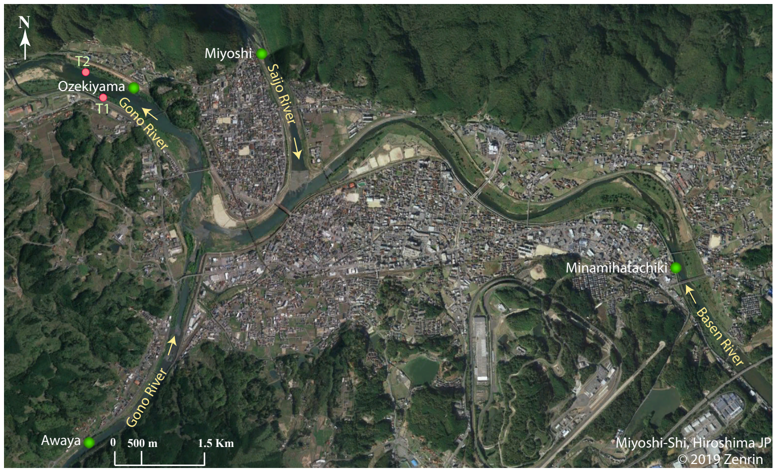

Fig.1 The Gono River and its tributary (Basen and Saijo Rivers), the location of the MLIT gauging stations (Green dots) and the position of the two acoustic stations (T1 & T2) of the FAT system (Red dots).

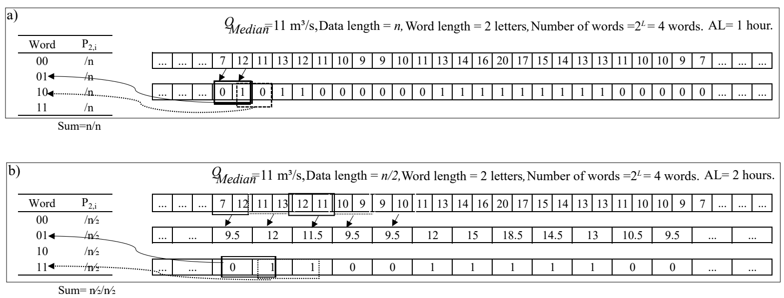

Fig. 2 Illustration for the symbolic strings method, a) the basic approach, b) illustration for data aggregation using Aggregation Length (AL=2).

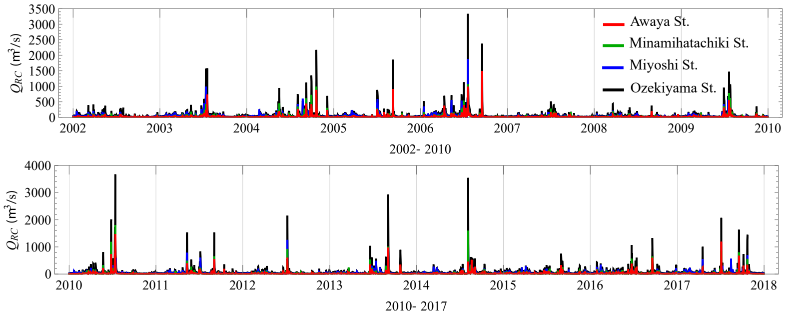

Fig. 3 Discharge time series during the study period (2002-2017) observed at Awaya (red), Minamihatachiki (green), Miyoshi (blue), and Ozekiyama (Black) stations.

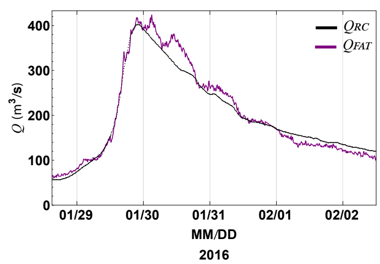

Fig. 4 A Comparison between the discharge performance computed by Q_{FAT} (purple) at Q_{RC} (black) at Ozekiyama observation site.

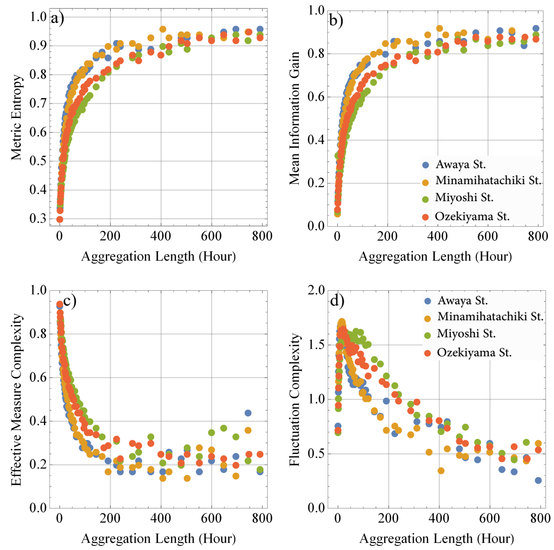

Fig. 5 Temporal scales of low frequency streamflow variations according to and complexity measures recorded at each station

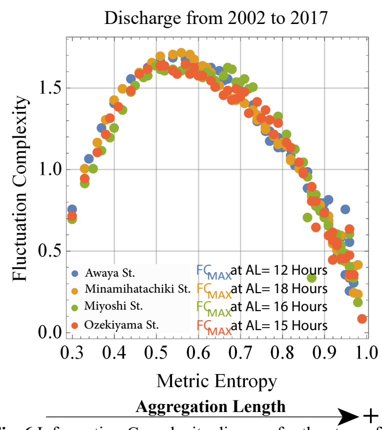

Fig. 6 Information-Complexity diagram for the streamflow data according to different aggregation length.

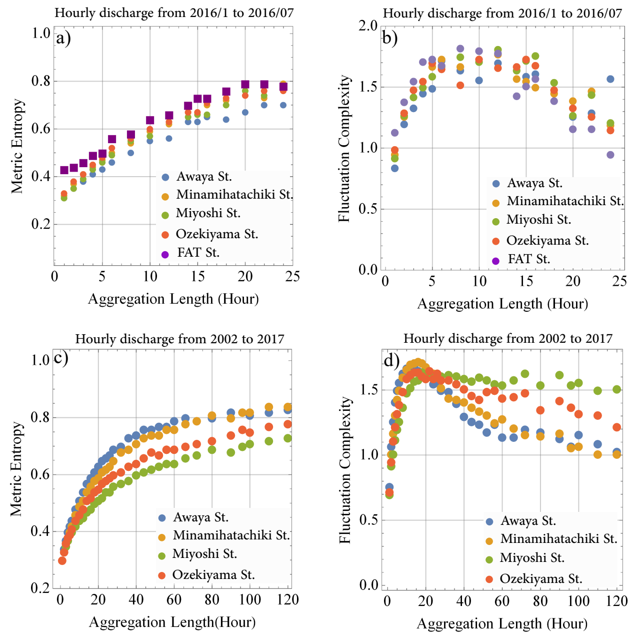

Fig. 7 Information and complexity contents for high frequency (a, b) and low frequency (c, d).

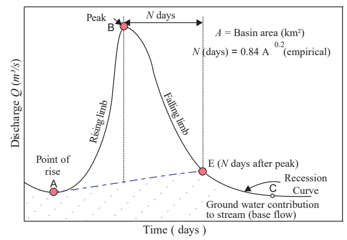

Fig. 8 Graphical hydrograph separation method: For the Gōno River catchment $N \approx 4$ days.

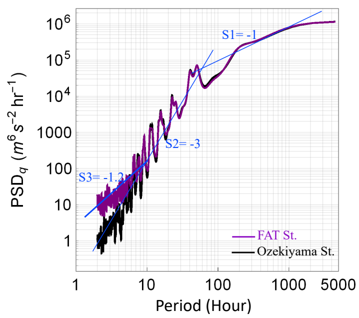

Fig. 9 Power spectra results for *QOzekiyama* (black) and *QFAT* (purple) during (2016/01 to 2016/06).