Assessment and ranking flood events in a regulated river using information and complexity measures

Mohamad Basel Al Sawaf*, Kiyosi Kawanisi¹, Cong Xiao¹

Department of Civil and Environmental Engineering, Graduate School of Engineering, Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima 739-8527, Japan

mbasel@hiroshima-u.ac.jp

Abstract. The availability of a robust approach that describe the hidden features of flood events in regulated rivers is of great importance. The key goal of this research is to utilize some of information and complexity measures to assess and rank flood patterns within a regulated river system. To meet this goal, the Metric Entropy ($ME$) as measure of information content and Rényi Complexity ($C_R$) as a quantification for complexity content were employed. To examine the role of river regulation on flood risk control, river stage records of two monitoring stations located at downstream of two different dams were considered in this research. The findings show that information and complexity metrics offer an image of the randomness embedded in dataset and the presence of internal patterns in studied data records. In general, this research shows that natural environmental risks and disasters can be assessed and ranked using a promising physical scheme based on information and complexity measures.

1. Introduction

Understanding different characteristics of rainfall-runoff process is of paramount importance to control the quantity and quality of water resources and mitigate potential flood risks and provide best water management practices.

Past works [1,2] documented that the climate records of the past decades had revealed evolving recurrent extreme weather events and natural disasters over the world, connected to adverse socioeconomic consequences. As a result of climate change, the occurrence of natural environmental disasters such as tropical cyclones, typhoons, droughts and floods had intensified [3].

Japan has a unique location where several disasters occur repeatedly, including volcanic, seismic, typhoons, and floods basically because it is being located in the Ring of Fire. These disasters impair the national sustainable development goals (SDGs). Therefore, further efforts are needed to explore how ecosystems are altered by the countless hydrological processes during usual and severe climates. Vast seasonal rainfall that occurs in the western part of Japan is one of the worst destructive disasters since it comes with landslides and mudflows.

Recently, information-based theories have received an increased concern in hydrological studies to detect the variability in hydrological dataset, including precipitation, temperature, and streamflow [5–7]. In particular, Al Sawaf and Kawanisi [7], investigated the different patterns of a mountainous river in Hiroshima over low and high-frequencies, and demonstrated that river flow has two regime scales. Moreover, they indicated that information and complexity measures can be used to assess flood events with very limited demonstration.
As a result, the purpose of this study is to assess annual flood events variability in a river at two different stations located in Hiroshima prefecture which is mainly affected by the torrential floods that hit west of Japan.

2. Study site and data description
The Gono River which is located in the west region of Japan was chosen for this study. The river originates from Kita-Hiroshima flows through Hiroshima and Shimane prefectures and finally pours into the Sea of Japan. The watershed of the Gono River has a cool temperate climate with four pronounced seasons. Basically, precipitation happens in winter, however, heavy rainfall occurs in the monsoon (June and July) as well as during the typhoon season (August and September).

![Figure 1 Gono River network, its main tributaries, and the studied stations (Yellow disks) along the river channel.](image)

The catchment area of the Gono River is 3963 km². Furthermore, the Gono River has three tributaries, namely, Basin, Saijo, and Kannose Rivers (Fig. 1). The Gono River watershed is monitored by the Japanese ministry of land infrastructure and transport (MLIT) using real-time gauging stations (Fig. 1) that measure the water stage \( (H) \) directly. In this study, Yoshida and Tono stations located at the downstream of the Haji and Haizuka Dams, respectively, were included in the assessment process.

Haji Dam is located at the upstream region of the Gono River, whereas Haizuka Dam is located at the upstream of the Basen River which is one of the main tributaries of the Gono River (Fig. 1).

The studied stations were established by the MLIT and continuously measures water stages. Therefore, the maximum daily water stage of those stations was obtained from the MLIT database from 2002 to 2020. At each stage monitoring station, the local MLIT administrations have defined a set of five levels warning protocols during a flood event.
3. Methodology

To assess and rank the variability of annual flood events at the selected stations, the performed method comprises the following steps.

First, following the symbol strings approach [5], daily stage records were encoded using a Boolean conversion into (S or I). In other words, we mapped each reading value to either S “Secure” if below the threshold stage (secure river site condition), or I “Insecure” otherwise (i.e., insecure river site condition). The threshold stage is predetermined at each monitoring station by the domestic MLIT offices. In this study, the “Designated water level” was selected as a threshold stage. Once the encoded record is determined, a word length \( L \) comprised two characters was defined. Hence, the number of possible different words that could be detected in the studied records was \( 2^L \). As a result, the possible words that could be faced in the encoded dataset were SS, SI, II, and SI (see Fig. 2).

![Figure 2 Boolean Conversion for the river stage records at Tono Station, red dashed lines represent insecure days.](image)

The next step, involved investigating the following three sets of probabilities: i) \( p_{L,i} \), is the state probability of the \( i \)th \( L \) word, where \( i = 1, 2, \ldots , 2^L \); ii) \( p_{L,j|i} \), 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) \( p_{L|i} \) 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 were determined, the information and complexity metrics could be estimated. In this study, we investigated Metric entropy \( (ME) \) as information metric and Rényi Complexity \( (C_R) \) as a quantification for complexity content and explained below.

The Metric Entropy \( (ME) \) is the Shannon entropy \( (E_S) \) divided by the word length \( (L) \); thus, it is a normalization of the Shannon entropy that offers an image of the contained information in a dataset but at the same time it is independent from the word length \( (L) \).

\[
E_S = - \sum_{i=1}^{2^L} p_{L,i} \log_2 p_{L,i} \tag{1}
\]

\[
ME = \frac{E_S}{L} \tag{2}
\]

Rényi Complexity, on the other hand, is a measure of complexity can be defined by the differences of Rényi entropies of conjugated orders [8]. Let the Rényi Complexity \( C_R (\alpha) \) of the order \( \alpha > 1 \) of a distribution of \( L \)-words be:

\[
C_R (\alpha) = - \frac{2}{(\alpha-1) \times L \times \ln(2)} \times (E_R (\frac{1}{\alpha}) - E_R (\alpha)) \tag{3}
\]
the scaling factor yields independency of the word length and eliminates the zero for $\alpha \to 1$. Investigations, $C_R$ is specific in the sense of a measure of complexity only for $\alpha \approx 1$. Thus, the Rényi Complexity be defined as:

$$C_R = \lim_{\alpha \to 1, \alpha > 1} (\alpha)$$  (4)

River stage records were examined year by year to investigate the annual variation of flood patterns. It is important to point out that the above analyses were accomplished using SYMDYN script developed by Wolf [5].

| Year | Yoshida St. | Tono St. |
|------|-------------|----------|
| 2002 | 1           | 1        |
| 2003 | 3           | 4        |
| 2004 | 4           | 4        |
| 2005 | 4           | 1        |
| 2006 | 4           | 4        |
| 2007 | 1           | 3        |
| 2008 | 1           | 3        |
| 2009 | 3           | 4        |
| 2010 | 4           | 4        |
| 2011 | 3           | 4        |
| 2012 | 3           | 4        |
| 2013 | 4           | 4        |
| 2014 | 3           | 4        |
| 2015 | 3           | 3        |
| 2016 | 3           | 4        |
| 2017 | 4           | 4        |
| 2018 | 4           | 4        |
| 2019 | 4           | 1        |
| 2020 | 4           | 4        |
| **Median** | **3** | **4** |

4. Results and discussion

4.1. Ranking and assessment using word number

The number of different words recoded each year gives an important indicator for annual flood pattern. In other words, each word gives an impression about the state of flood per year, for example, if the observed number of different words was 1, it meant that the pattern was SS (i.e., water level at a studied station was less than the selected threshold and no significant floods were detected at that year). Contrary, if the observed number of different words was 4, it meant that the pattern was II, also, it tells that at least one significant flood occurred during that year and progressed at least for two consecutive days. Finally, if the number of different words was three words, it meant that the flood pattern was either
SI or IS and it indicates that at least one significant flood occurred at that year but did not progress for more than one day.

Table 1 conveys information about the number of different words recorded each year at both Yoshida and Tono stations. Obviously, it can be noticed that in the case of Yoshdia Station, the following years 2002, 2007, and 2008 reported only one number of different words (i.e., SS). In other words, the daily water level recorded at this station during the mentioned years was less than the investigated threshold.

In general, the median number of different words recorded at Tono was greater than that observed at Yoshida Station.

### 4.2. Ranking and assessment using information and complexity measures

Assessment and ranking using information and complexity metrics is very useful since it provides further understanding regarding the randomness and internal patterns of embedded flood events.

#### Table 2. Number of different words per year observed at Yoshida and Tono Stations.

| Year | Yoshdia St. | Tono St. |
|------|-------------|----------|
|      | ME  C_k  Overall score | Ranking | ME  C_k  Overall score | Ranking |
| 2002 | 0  0   0 | 11 | 0  0   0 | 11 |
| 2003 | 0.05 0.61 0.33 | 8 | 0.08 0.96 0.52 | 8 |
| 2004 | 0.14 1.37 0.76 | 2 | 0.19 1.64 0.92 | 1 |
| 2005 | 0.07 0.91 0.49 | 7 | 0  0   0 | 11 |
| 2006 | 0.15 1.43 0.79 | 1 | 0.12 1.27 0.70 | 3 |
| 2007 | 0  0   0 | 11 | 0.03 0.4  0.22 | 10 |
| 2008 | 0  0   0 | 11 | 0.03 0.39 0.22 | 10 |
| 2009 | 0.03 0.4  0.22 | 9 | 0.10 1.14 0.62 | 5 |
| 2010 | 0.09 1.05 0.57 | 6 | 0.16 1.48 0.82 | 2 |
| 2011 | 0.03 0.4  0.22 | 9 | 0.11 1.19 0.65 | 4 |
| 2012 | 0.03 0.39 0.21 | 10 | 0.08 0.96 0.52 | 8 |
| 2013 | 0.09 1.07 0.58 | 5 | 0.11 1.19 0.65 | 4 |
| 2014 | 0.03 0.4  0.22 | 9 | 0.09 1.07 0.58 | 6 |
| 2015 | 0.03 0.4  0.22 | 9 | 0.03 0.4  0.22 | 10 |
| 2016 | 0.05 0.61 0.33 | 8 | 0.10 1.14 0.62 | 5 |
| 2017 | 0.09 1.07 0.58 | 5 | 0.06 0.80 0.43 | 9 |
| 2018 | 0.10 1.11 0.61 | 4 | 0.09 1.05 0.57 | 7 |
| 2019 | 0.07 0.91 0.49 | 7 | 0  0   0 | 11 |
| 2020 | 0.11 1.19 0.65 | 3 | 0.06 0.80 0.43 | 9 |

First, the ME is a measure that quantifies the randomness in a dataset and represents an image of the contained information in a dataset, while simultaneously being independent of word length. On the other hand, complexity metrics are helpful measures that capture the existence of internal patterns in studied datasets, i.e., how a system transforms from one pattern to another. Hence, data that contain a high level of fluctuation have larger C_k value.

Table 2 compares the values of ME (information) and C_k (complexity) as observed at Yoshida and Tono Stations. It can be seen that in the case of one word observed per year (e.g., 2002, 2007, and 2008...
in Table 1 in the case of Yoshida St. and in 2002, 2005, and 2019 in the case of Tono St.) the metric entropy and Rényi Complexity values were 0 in Table 2. That is to say, the stage within all days of these years were less than the specified threshold and hence no disorder pattern could be noticed. Interestingly, in the case of Yoshida Station, it can be verified that the highest values of information and complexity were recorded in 2006 and 2004, respectively. However, in the case of Tono Station, it can be verified that the highest values of information and complexity were recorded in 2004 and 2010, respectively, suggesting that several patterns can be observed at Tono station due to different patterns of water release from Haizuka Dam (upstream of Tono Station).

To get sensible understanding regarding the information and complexity patterns observed at Yoshida and Tono Stations, Table 3 shows the number of days that in which water level at those stations were greater than the selected threshold. As can be seen, Yoshida station in 2020 shares the same scores with Tono station in 2011. Outstandingly, in 2020, flood patterns observed in Yoshida station were two single-day flood events and two events persisted for two-consecutive days. Alternatively, in 2011, Tono station showed the same annual pattern (i.e., two single-day flood events and two events persisted for two-consecutive days), and hence, their $ME$ and $CR$ were similar. In the same manner, Yoshida and Tono Stations shared the same information and complexity values in 2015 and the same annual flood pattern (i.e., one single-day flood event).

### Table 3. Annual flood patterns and quantification of information and complexity scores.

| Year | Yoshida St. | Tono St. | | | |
|------|-------------|----------|-------------|-------------|---------------|
|      | $ME$ | $CR$ | Day(s) of water level greater than threshold | $ME$ | $CR$ | Day(s) of water level greater than threshold |
| 2011 | 0.03 | 0.4 | Single flood in one day | 0.11 | 1.19 | Two flood events each of them within single day |
| 2015 | 0.03 | 0.4 | Single flood in one day | 0.03 | 0.4 | Single flood in one day |
| 2020 | 0.11 | 1.19 | Two flood events each of them within single day | 0.06 | 0.8 | Single flood within 2 successive days |

#### 4.3. Annual ranking of flood events

Overall ranking of flood events was performed using the following empirical formulae as:

\[
\text{Overall score} = \frac{ME + CR}{2}
\]

The highest overall score indicates the higher occurrence of floods observed per year as it reflects the high number of disordered and complex patterns. The overall score is given it Table 2, the annual events ranked then starting by the severest year. In general, it seems that 2004 and 2006 were among the top affected years by floods for both studied stations.
5. Conclusions
Understanding different characteristics of floods and their hidden patterns is of paramount importance to control the quantity and quality of water resources and mitigate potential flood risks and provide best water management practices.

In this study, we assessed flood patterns using the daily records of water stages obtained from two stations located at the downstream of dams. The main purpose was to investigate the hidden patterns of floods using information and complexity measures namely the Metric Entropy as measure for information content and Rényi Complexity. The findings show that information metric offers an image of the randomness embedded in the dataset whereas complexity indicates the presence of internal patterns in studied datasets. In general, this research shows that natural environmental risks and disasters can be assessed and ranked using a promising physical scheme based on information and complexity measures.

Acknowledgments
This research was supported by JSPS KAKENHI grant number JP21K14253.

References
[1] Arnell N W 1999 Climate change and global water resources Global Environmental Change
[2] Nohara D, Kitoh A, Hosaka M and Oki T 2006 Impact of climate change on river discharge projected by multimodel ensemble J. Hydrometeorol.
[3] Hein W, Wilson C, Lee B, Rajapaksa D, de Moel H, Athukorala W and Managi S 2019 Climate change and natural disasters: Government mitigation activities and public property demand response Land use policy 82 436–43
[4] Pachepsky Y A, Martinez G, Pan F, Wagener T and Nicholson T 2016 Evaluating Hydrological Model Performance using Information Theory-based Metrics Hydrol. Earth Syst. Sci. Discuss.
[5] Wolf F 1999 Berechnung von Information und Komplexität in Zeitreihen-Analyse des Wasserhaushaltes von bewaldeten Einzugsgebieten Bayreuther Inst. für Terr. Ökosystemforschung. 65 1–164
[6] Hauhs M and Lange H 2008 Classification of Runoff in Headwater Catchments: A Physical Problem? Geogr. Compass 2 235–54
[7] Al Sawaf M B and Kawanisi K 2020 Assessment of mountain river streamflow patterns and flood events using information and complexity measures J. Hydrol. 590 125508
[8] Wolf J and Prandle D 1999 Some observations of wave-current interaction Coast. Eng. 37 471–85