

Multi-index data dimension reduction approach and its applicability in the calculation of indicators of hydrological alteration

Wenhai Zhang, Yangwen Jia, Jinjin Ge, Xiaorong Huang, Guangheng Ni, Jiaming Hou and Hao Wang

ABSTRACT

The length of record (LOR) method is an evaluation method that provides quantitative advice for the amount of computational data required for use of the indicators of hydrological alteration (IHA). The use of multi-index hydrological indicators to reflect river hydrological–ecological characteristics is the essence of the IHA method, while the LOR evaluation result using a single index does not have practical application value in the absence of IHA data volume. In this paper, we expand the LOR method from single index version into multi-index version, apply it to comprehensively analyze the credibility of hydrological alteration (HA) multi-indicators under different data volumes, and explore the relationship between multi-index LOR results and data requirements. Combined with the hydrological–ecological relationship, the practical application criteria of LOR dimension reduction under the condition of multiple HA indicators is given. The results show that the LOR results corresponding to each group of indicators in IHA have different data requirements, so an in-depth understanding of the hydrological–ecological relationship is the key to LOR’s application of IHA data dimension reduction. In addition, we discuss the limitations of the LOR method of multi-index dimension reduction and its application value in IHA calculations.

Key words | data reduction method, data variability, environmental flow, indicators of hydrological alteration, length of record

Wenhai Zhang
Guangheng Ni
Tsinghua University,
Haidian, 100084, Beijing,
China

Wenhai Zhang
Yangwen Jia (corresponding author)
Jinjin Ge
Jiaming Hou
Hao Wang
China Institute of Water Resources and
Hydropower Research,
100038, Beijing,
China
E-mail: jjayw@iwhr.com

Xiaorong Huang
Sichuan University,
No. 24 South Section 1, Yihuan Road, 610065,
Chengdu,
China

INTRODUCTION

Hydrologic regimes play major roles in determining the biotic composition, structure, and function of aquatic, wetland, and riparian ecosystems (Richter *et al.* 1996). The flow stability and flow rate change create habitat conditions for all kinds of organisms (Richter *et al.* 1997). Extreme hydrological conditions and the timing of occurrences of flow determine the integrity of the life cycles of living species and the direction of superiority in the survival competition (Lytle & Poff 2004). Moreover, the magnitude and duration of the pulse flow affect the success and failure of the breeding process (Lake 2008). In addition, hydrological

regimes indirectly affect ecosystems, riparian systems, and wetland ecosystems by altering the water temperature, oxygen content, water chemistry, and sediment transport. Therefore, hydrological alterations are widely considered to be the main drivers of biotic and abiotic conditions in river ecosystems (Poff *et al.* 1997; Scarcella *et al.* 2016). In other words, the indicators of hydrological alteration (IHA) can be used to measure the degree of ecological damage (Richter *et al.* 1997; Poff *et al.* 2006). Among the many existing methods, the IHA method can be used worldwide. This is because the IHA method uses statistical

methods, and analyzes the hydrological indicators associated with biological attributes to quantify how much the flow conditions deviate from natural conditions (Poff & Matthews 2013).

With the enhanced understanding of hydrological alterations and ecological responses, the IHA evaluation has rapidly changed from an ecological impact assessment of a single river hydrological alteration (HA) to the ecological impact assessment of the river network. Increasingly, smaller watersheds are gradually becoming the object of IHA evaluation, leading scientific researchers to increase their data volume demand. That is, scientists and scholars using the IHA evaluation method will face the problem of 'missing data' in the future (Carlisle *et al.* 2010). In some developed countries, almost complete scientific-data-sharing management mechanisms have been implemented. For example, the official website of the United States Geological Survey (USGS) fully discloses long-sequence daily flow data, from which it is relatively easy to obtain basic flow data (Isaacson & Coonrod 2011; Afshari *et al.* 2017). However, owing to monitoring equipment failures and the impacts of natural phenomena, such as earthquakes, floods, or landslides, and data transmission issues, a lack of data from the storage and retrieval process is inevitable (Makhuvha *et al.* 1997; Panu *et al.* 2000).

In some developing countries, the problem of missing data has become further complicated. For example, the rate at which new hydrological monitoring stations are being constructed is not sufficient to meet the increasing data needs of scientists and scholars (Eng *et al.* 2013; Harrigan *et al.* 2018). Moreover, the relations among neighboring countries have been strained for a long time, and cross-border river hydrology data have strict national confidentiality. Even in non-state-confined inland rivers, there are scattered years of hydrological data that have not occurred in the entire publication. When applying IHA calculations, users usually arbitrarily define the amount of data to be used according to their own subjective understanding, or ignore IHA calculation data limits. Two major bottlenecks in the application of IHA thus exist (Giustarini *et al.* 2016). On one hand, the results of the IHA evaluation based on inadequate hydrological data will be seriously questioned. On the other hand, no IHA evaluation can be performed at any time in a basin where any data are missing.

To accelerate the globalization of IHA applications, scientifically reliable results are increasingly being shared by developers and users. Many studies have been conducted to address the above issues. In earlier studies, some scholars used mathematical statistical interpolation to compensate for the missing data. Khalil *et al.* (1999) used a simple interpolation method (such as average interpolation and correlation analysis) to satisfy some data-filling requirements. Wang *et al.* (2015) used the regression relationship between adjacent hydrological stations in the Tarim River Basin in China to compensate for the missing data. All interpolated flow time series that satisfied homogeneity were evaluated by the penalized maximal F-(PMF) uniformity test. Harvey *et al.* (2012) evaluated 15 simple filling techniques, including regression, scaling, and equidistance methods, and found that flow data can improve the accuracy of filling data by performing seasonal grouping or modifying data sets to improve homogeneity. However, in many cases, the relationship between predictors is not a true linear relationship (Govindaraju & Rao 2000; Yang *et al.* 2017).

The artificial neural network (ANN) technique has gradually replaced the data-filling method of hypothetical variables (simple linear, nonlinear, etc.) by representing the nonlinear mapping between variables (French *et al.* 1992). Khalil *et al.* (2001) applied the concept of a population and ANN to compensate for the peak flow and medium flow of missing data, and proved that it is reasonable to fill gaps in river flow data based on the neural network concept and group value data method. Ilunga & Stephenson (2005) used three different back propagation neural network learning methods to complete the missing data, and verified the difference in accuracy between them and the observed data.

Meanwhile, as researchers strive to improve the accuracy of data, the combination of basic mathematical statistical interpolation methods and hydrological feature conditions is gradually being applied to data filling. Poff *et al.* (2010) and Kennard *et al.* (2010) and others used the method of river classification and grading of characteristics similar to flow rates to establish the baseline conditions for the flow. The hydrological index was used to effectively compensate for the amount of data in non-hydrological stations and to solve the problem of missing IHA data. Hughes & Smakhtin (1996) supplemented the missing data for six sets of catchment areas in southern Africa based on

the use of 1-day flow duration curves for each month of the year. The approach assumes that flows that simultaneously occur at sites in reasonably close proximity to each other correspond to similar percentage points on their respective duration curves. Mobley *et al.* (2012) used the simple drainage area ratio (DAR) method to fill the data based on the assumption that the observed and simulated watersheds had the same physical characteristics. It was proved that the results of the DAR comparison method could well represent extremely low-flow conditions. However, these methods still lack practical physical meaning, and their data accuracy is relatively low.

Other scholars compensated for the missing data through the daily average runoff simulation results by hydrological models. The general hydrological model uses relatively simple mathematical equations to describe the highly correlated processes of water, energy, and vegetation in complex spatial distributions in the watershed (Vrugt *et al.* 2005). The hydrological model with clear physical meaning greatly improves the accuracy of data filling. Shrestha used the distributed variable infiltration capacity (VIC), a distributed hydrological model, to simulate the 1991–2000 long-sequence flow of the Fraser River in British Columbia, Canada. It was found that the VIC model could simulate the annual flow of the IHA index, the median flow rate, the seasonal change of flow rate, and the annual minimum flow rate. However, uncertainties remained in the monthly average flow, pulse flow, and frequency indicators (Shrestha *et al.* 2014). Ryo *et al.* (2015) used the distributed hydrological model (DHM) to produce similar conclusions with respect to the Sagami River basin in Japan.

To increase the ability of the model to portray extreme hydrological events, Kusangaya *et al.* (2018) used downscaled general circulation model (GCM) data and historical climate data, respectively, in hydrological models to compare the results. It was found that GCM data results perform poorly when simulating extremely high and low flows. However, IHA features can be more accurately described if the climatic data are downscaled at finer spatial scales and with adequate spatial detail. Nevertheless, in the establishment of hydrological models, natural geospatial characteristics, such as precipitation, canopy interception, evaporation, seepage, and groundwater flow, must be used (Engeland *et al.* 2001; Todini 2011; Jin *et al.* 2017). These characteristics are actually

types of implicit data. Such data can greatly increase the initial workloads of IHA applications.

To solve the current bottleneck, Timpe & Kaplan (2017) proposed the concept of the length of record (LOR) method. LOR gives the reliability of the IHA calculated by the existing data volume through statistical methods, thereby avoiding collecting too much unnecessary data. Under the assumption of similar area, altitude, and flow, it considers that the reliability calculated by LOR can reflect the IHA calculation data demand within a specific geographical range around a given region. This method requires no addition to the data volume and simplifies the data processing workload for calculating IHA, such as data interpolation or hydrological models. At the same time, the credibility of the existing data is engendered by the fluctuation range and confidence interval of the average value in the IHA results, which reduces the data dimension. However, in that study (Timpe & Kaplan 2017), only the 1-day-max indicator in the IHA was used for the example of the dimensionality evaluation. The corresponding LOR dimensionality reduction effects of the 33 indicators in the IHA have not yet been fully elucidated. Further explanations are needed for the data volume selection and the LOR dimension reduction application method under the IHA multi-indicator.

So far, LOR is the only evaluation method that gives a clear estimate of the amount of data required for IHA calculations through mathematical statistics, as a single indicator cannot comprehensively reflect changes in river hydrological characteristics, and, in the case of identical amounts of data, there may be differences in the sizes of LOR results for different IHA indicators. Therefore, the LOR result of a single indicator cannot provide substantial help for the IHA method application in the choice of data volume. In order to overcome this shortcoming, we attempted to analyze the LOR results of all IHA indicators and discover the mechanism of influence between the hydrological data structure and the LOR evaluation results. On the basis of fully understanding the LOR evaluation results of different IHA indicators, the applicability scope of the LOR method in IHA multi-index applications is proposed to prove the importance of this method in dimension reduction application. The research results can provide effective suggestions for data volume selection in eco-hydrological evaluation.

MATERIALS AND METHODS

Study area and data

The Jinsha River is in the upper reaches of the Yangtze River, China's largest river, with a total length of 3,481 km and a total drainage area of 572,300 km². The runoff and rainfall in the Jinsha River are concentrated in the flood season from June to October, and the runoff from July to September accounts for about 55% of the entire year. The impact of human activities in the basin was relatively late. After 2005, in the mainstream of the Jinsha River, hydropower stations began to be built. The Panzhihua hydrological station is located at the last station of the mainstream of the middle reaches of the Jinsha River and was built in 1965. The daily hydrologic data from 1976 to 2015 used in the article was provided by the Hydrology Yearbook of the Ministry of Water Resources of the People's Republic of China. Panzhihua hydrological station was chosen because the interannual runoff is stable and has complete daily runoff data, which is beneficial to the LOR method calculation.

IHA method

The IHA method is an open-source interface model developed by the [Nature Conservancy \(2009\)](#). It uses daily average flow data and calculates 33 indicators closely related to ecological attributes ([Richter *et al.* 1996](#)), as shown in [Table 1](#). The 33 indices generated by IHA consist of five major categories: (1) magnitude of monthly water conditions; (2) magnitude and duration of annual extreme water conditions; (3) timing of annual extreme conditions; (4) frequency and duration of high and low pulses; and (5) rate and frequency of condition changes ([Magilligan & Nislow 2005](#)). Among them, the characteristics of extreme events are more related to the indicators in Groups 3 and 4, and the average characteristics of traffic are more closely correlated with Group 1 and Group 5 indicators.

LOR analysis

The characterization of natural and altered flow regimes using IHA or other statistical methods requires adequate

flow data. [Richter *et al.* \(1996\)](#) analyzed three rivers with different hydrological conditions and considered that the flow data at the time of calculating IHA would take at least 20 years. However, no one has conducted further research on the reliability of data required by IHA. Given this uncertainty, we modify the definition of data volume requirements in IHA data analysis by Richter. The number of years corresponding to the evaluation result of LOR is used as the lower limit of data volume demand calculated by IHA, and is applied to the demand limit of IHA calculation data volume in similar hydro-geographic spatial feature areas ([Hughes & Smakhtin 1996](#)).

The numerator and denominator in the LOR results are normalized statistics representing confidence level and confidence interval, respectively. First, the LOR data calculation selects the flow station with the least anthropogenic impact and longest record lengths in the desired evaluation area. Then, we determine the IHA indicator that provides the LOR calculation reference. Next, we calculate the target indicator for each year in a data set along with the long-term mean for this indicator. The reference indicator's long-term average values are randomly ordered and grouped into record-length increments ranging from two years to the full LOR. Finally, the mean of each record length increment is calculated for a comparison with the long-term mean. This process is repeated 50,000 times, from which 95%, 90%, and 85% confidence intervals are calculated. Using these statistics, we calculate the LOR required to produce a specified confidence level for a given long-term mean confidence interval for the river in the study ([Table 2](#)).

In this study, we examine as a calculation example the hydrological flow of the Panzhihua hydrological station in the middle areas of the Jinsha River in China from 1976 to 2015. Using the 1-day maximum index as a reference, the 40-year LOR evaluation results at this station are shown in [Figure 1](#). The data volume required to characterize the annual maximum flow within 10% of the long-term mean with 85% confidence is 10 years, and the LOR result is abbreviated as 10/85. [Figure 1](#) illustrates this example, with dashed vertical lines indicating the data volume required for a desired calculation output at the intersection of a dashed horizontal line, representing a given confidence level, and a normal curve, representing a given confidence interval (as CI in [Figure 1](#)). Therefore, the LOR method

Table 1 | Hydrologic parameters used in the indicators of hydrological alteration (IHA) and their grouping definition and abbreviation. Adapted from IHA Manual (Nilsson & Berggren 2000)

IHA statistics group	Hydrologic parameters	Abbreviated name
Group 1: Magnitude of monthly water conditions (12 indices)	Mean value for January	January
	Mean value for February	February
	Mean value for March	March
	Mean value for April	April
	Mean value for May	May
	Mean value for June	June
	Mean value for July	July
	Mean value for August	August
	Mean value for September	September
	Mean value for October	October
	Mean value for October	October
	Mean value for December	December
Group 2: Magnitude and duration of annual extreme water conditions (12 indices)	Annual minima 1-day means	1-day min
	Annual minima 3-day means	3-day min
	Annual minima 7-day means	7-day min
	Annual minima 30-day means	30-day min
	Annual minima 90-day means	90-day min
	Annual maxima 1-day means	1-day max
	Annual maxima 3-day means	3-day max
	Annual maxima 7-day means	7-day max
	Annual maxima 30-day means	30-day max
	Annual maxima 90-day means	90-day max
	Base flow index	Base flow
	Number of zero flow days	Zero flow days
Group 3: Timing of annual extreme water conditions (2 indices)	Julian date of each annual,1-day maximum	Date min
	Julian date of each annual,1-day minimum	Date max
Group 4: Frequency and duration of high and low pulses (4 indices)	Number of low pulses within each water year	Lo pulse #
	Mean or median duration of low pulses (days)	Lo pulse L
	Number of high pulses within each water year	Hi pulse #
	Mean or median duration of high pulses (days)	Hi pulse L
Group 5: Rate and frequency of water condition changes (3 indices)	Rise rates: Mean or median of all positive differences between consecutive daily values	Rise rate
	Rise rates: Mean or median of all negative differences between consecutive daily values	Fall rate
	Number of hydrologic reversals	Reversals

can be deemed an effective method for reducing data by giving the degree of reliability of the IHA result in the absence of the data amount.

RESULTS AND DISCUSSION

Analysis of LOR evaluation results and influencing factors of IHA indicators

The LOR dimensionality reduction results of the IHA hydrological flow outcomes over the past 40 years in the Jinsha River Panzhihua hydrological site are shown in Table 2. It

is observed that the amount of data required to produce a confidence level at 10% of the average flow is two or three times the amount of data required to produce a confidence level at 5% of the average flow for the same CI in Group 1 and Group 2. For example, the amount of data required to produce a 5/95 LOR result for the January indicator is three times the amount required for a 10/95 LOR result. Likewise, more than twice as many years are required for a 5/90 LOR result relative to a 10/90 LOR result for an indicator of the 7-day maximum. In addition, the amount of data required to produce a given LOR result to sustain N-day max in Group 2 is almost twice the amount of data required to sustain N-day min. The date min indicator in Group 3 has

Table 2 | Length of record (LOR) results for each HA's indicator

IHA statistics group	Indicator	LOR results (years)					
		5/95	5/90	5/85	10/95	10/90	10/85
Group 1: Magnitude of monthly water conditions (12 indices)	January	15	12	10	5	4	3
	February	13	10	8	4	3	2
	March	13	10	8	4	3	2
	April	17	15	13	6	5	4
	May	26	22	19	11	9	7
	June	28	26	23	15	12	10
	July	28	25	23	15	12	10
	August	33	30	28	20	17	14
	September	28	25	23	15	12	10
	October	26	22	19	12	9	7
	November	23	20	17	10	7	6
	December	18	15	13	7	5	4
Group 2: Magnitude and duration of annual extreme water conditions (12 indices)	1-day min	14	12	9	5	4	3
	3-day min	12	9	8	4	3	2
	7-day min	12	9	9	4	3	2
	30-day min	12	9	7	4	3	2
	90-day min	12	10	8	4	3	2
	1-day max	29	26	24	15	12	10
	3-day max	29	26	24	15	12	11
	7-day max	29	26	23	15	12	10
	30-day max	28	26	24	15	12	10
	90-day max	26	24	21	12	9	8
Base flow	20	17	15	7	6	5	
Zero flow day	None	None	None	None	None	None	
Group 3: Timing of	Date min	39	39	38	36	34	32
	Date max	11	9	7	4	3	2
Group 4: Frequency and duration of high and	Lo pulse #	39	38	38	37	34	32
	Lo pulse L	38	38	38	36	35	35
	Hi pulse #	37	37	34	29	26	23
	Hi pulse L	39	38	38	35	34	33
Group 5: Rate and frequency of water	Rise rate	33	32	28	20	17	14
	Fall rate	39	38	37	30	27	25
	Reversals	33	30	27	18	15	13

a higher data volume requirement (nearly 40 years) to produce a given LOR confidence, while the date max indicator shows a lower data volume requirement (less than 10 years) to produce a given LOR confidence. Although the two indicators are in the same group, there is a significant difference in the amount of data required for the same LOR result confidence. In Group 4 and Group 5, the amount of data required to produce a given

LOR result for the reversals indicator is relatively low, as the amount of data required for other indicators is significantly higher (nearly 40 years).

The amount of data required to produce a given LOR result for each indicator of HA showed a positive correlation with the size of the flow during the year. According to the average monthly flow rate of Panzhihua station, the flood season is concentrated from June to October, and floods

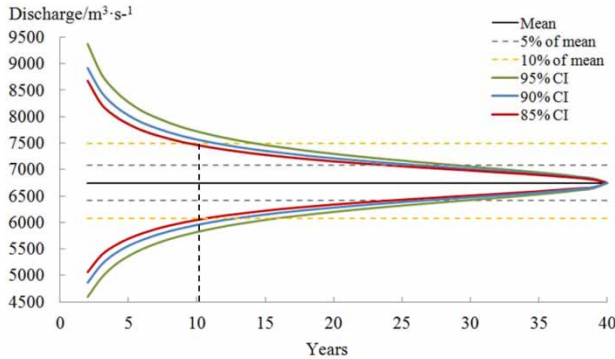


Figure 1 | One-day maximum LOR dimensionality reduction results.

are more frequent from July to September. The indicators in Group 1 of the IHA can reflect the flow change during the year. We can observe from Figure 2 that the greater the monthly average flow, the greater the amount of data that is required to produce a given LOR result. For example, in the flood season in August, the data volume required to produce a 10/95 LOR result will take 20 years; in the dry season in March, the data volume required to produce a 10/95 LOR result will only take four years.

The amount of data required to produce a given LOR result for each indicator in Group 2 has a consistent relationship with the flow rate during the wet and dry seasons. Because the N-day maximum appears in the flood season, the N-day minimum appears in the dry season, and 70% of the water in the whole year appears in the wet season (Proctor et al. 2011). Therefore, it is also stated that the larger the flow rate, the larger the amount of data is required to produce a given LOR result in IHA.

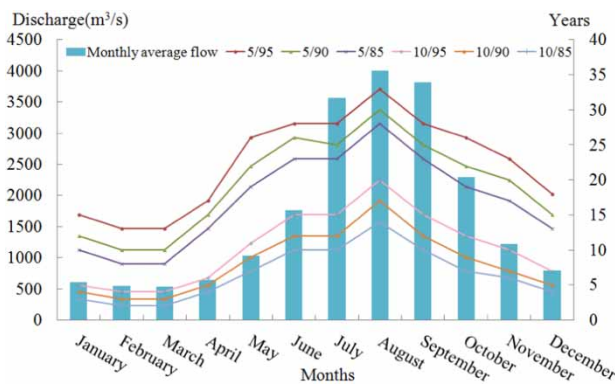


Figure 2 | Correlation between data volume (years) and the monthly average discharge at each LOR result.

On the other hand, regardless of the level of LOR credibility, the data volume requirements for the indicators in Group 3, Group 4, and Group 5 remain relatively high value, which may be related to the uncertainty of the component parameter attributes. The finding also indicates that fewer years cannot guarantee the reliability of the IHA results.

Data variability affects the amount of data reflected by LOR

The coefficient of variation of the hydrological flow data and the size of the mutation intensity will have an impact on the amount of data reflected by the LOR. The LOR corresponding to IHA has its own variation characteristics. The coefficient of variation for the long series of data for each indicator of IHA is shown in the coefficient of variation column in Table 3, and the mutation intensity for the long series of data is shown in the mutation intensity column. We define Mutation intensity = Max/Average, which reflects the deviation intensity of the most catastrophic point. In this paper, the value of Mutation intensity >1.7 (Dimensionless) is used as the boundary of mutation points. In comparing the coefficient of variation of IHA for each group of indicators in Table 3, and combining them with data volumes reflected by the LOR results in Table 2, we find that the larger the coefficient of variation, the greater the number of years that is required for evaluation of the LOR results.

In particular, the coefficient of variation of the date minimum, Lo pulse #, Lo pulse L, and Hi pulse L in the IHA exceeds 0.8, and the amount of data at each LOR credibility level exceeds 30 years. On the other hand, even if there are differences in the variability of the two indicators in the same group, the data volume results of the date minimum and date maximum indicators in Group 3 are 38 years and two years, respectively, for the 10/85 LOR result.

The mutation intensity results are shown in the mutation intensity column in Table 3. By comparing the LOR results of Tables 2 and 3, we can observe that the higher the intensity of the mutation, the greater the amount of data that is required to produce a given LOR result. For example, the mutation intensity in Group 4 exceeds 2.0 while the mutation intensity in Group 1 is less than 1.7. The data

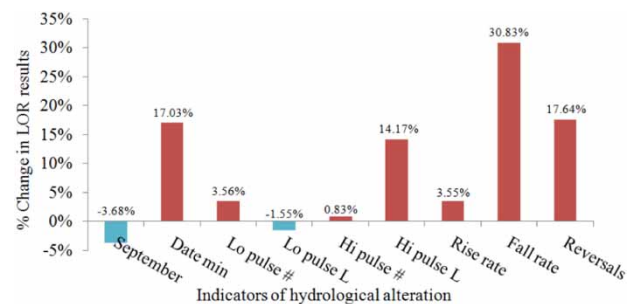
Table 3 | Data variability characteristics of IHA indicators

IHA statistics group	Indicators	Coefficient of variation	Mutation intensity	The number of abnormal points
Group 1: Magnitude of monthly water conditions (12 indices)	January	0.127	1.218	0
	February	0.111	1.192	0
	March	0.108	1.211	0
	April	0.151	1.600	0
	May	0.212	1.436	0
	June	0.253	1.573	0
	July	0.252	1.570	0
	August	0.334	1.861	1
	September	0.257	1.662	0
	October	0.213	1.614	0
	November	0.190	1.370	0
	December	0.150	1.278	0
Group 2: Magnitude and duration of annual extreme water conditions (12 indices)	1-day min	0.121	1.214	0
	3-day min	0.110	1.197	0
	7-day min	0.106	1.191	0
	30-day min	0.104	1.172	0
	90-day min	0.109	1.222	0
	1-day max	0.261	1.675	0
	3-day max	0.262	1.647	0
	7-day max	0.263	1.626	0
	30-day max	0.255	1.566	0
	90-day max	0.218	1.532	0
	Base flow	None	None	None
	zero flow days	0.166	1.468	0
Group 3: Timing of annual extreme water conditions (2 indices)	Date min	0.870	4.040	3
	Date max	0.101	1.171	0
Group 4: Frequency and duration of high and low pulses (4 indices)	Lo pulse #	0.872	4.356	7
	Lo pulse L	0.933	2.750	8
	Hi pulse #	0.537	2.667	4
	Hi pulse L	0.800	3.291	4
Group 5: Rate and frequency of water condition changes (3 indices)	Rise rate	0.331	1.759	2
	Fall rate	0.516	2.996	4
	Reversals	0.305	2.026	3

volume required to produce a given LOR result for Group 4 is 10 years more than that of Group 1.

After removing these high intensity points, the amount of data required to produce a given LOR result among the 33 indicators in IHA was reduced by varying degrees. We define D_{rf} as the amount of data required after the removal of the mutation point, T_{af} as the total amount of data after the removal of the mutation point, D_{ta} as the amount of data required before the removal of the mutation point, and T_{ab} as the total amount of data before the removal of the mutation point. The y-axis percentage value in the figure indicates $D_{rf}/T_{af} - D_{ta}/T_{ab}$, and reflects the degree of change in the amount of data before and after

the removal of the mutation point in Figure 3. Take the data volume at 5/95 LOR credibility as an example; after removing these mutation points, the amount of data will

**Figure 3** | LOR dimensionality reduction after the removal of extreme points.

be significantly reduced for some indicators in Figure 3. For example, although the LOR dimensionality reduction effect after removal of the mutation point improves by 30.83%, fall rate still requires 24 years of long-sequence data. However, the effects of Lo pulse #, Hi pulse #, and the Rise rate indicator reduction of data volume are not significant. This means that Lo pulse #, Hi pulse #, and the Rise rate indicators do not significantly improve the dimensionality reduction effect. This finding shows that, for these indicators, the removal of the mutation point does not further improve the data demand lower limit. In other words, the LOR evaluation itself has a lower limit value, and it is not possible to use just a few years of flow data to make the IHA calculation result extremely reliable.

In addition, for the September index, Lo pulse L showed a slight decrease in the amount of data required to produce a given LOR result after removing the mutation point. This result does not mean that the LOR dimensionality reduction effect of the two indicators has a significant downward trend after the mutation point removal. It is worth mentioning that, when performing LOR calculations, it is necessary to properly remove the abnormal mutation points, and the results after the elimination of errors can become a truly practical reference. However, some physical meanings must be considered when removing the mutation points. For example, after a dam is built, the frequency of rising tides will significantly increase. Therefore, at this time, LOR calculations should retain these mutation points.

Ecological standard for multi-index LOR dimension reduction evaluation application

The amount of data required to produce a given LOR result differs between the 33 indicators in IHA. Therefore, when the maximum volume of all 33 indicators is selected as the standard data amount for better LOR results, the required data volume is still close to the total amount of data, which has no practical significance. In fact, the maximum amount of data required to produce a 5/95 LOR result using all 33 indicators in IHA is 39 years, only reducing the amount by one year. Even producing a lower 10/85 LOR result using all 33 indicators in IHA, the amount of data required is 35 years. However, it is not necessary to

use the maximum amount of data available as the lower limit of the data volume when producing a given LOR result, as the maximum amount required varies among the 33 indicators in IHA. Instead, we should focus on the hydrological–ecological response that we must evaluate. If the maximum amount of data required to produce a given LOR result among all the indicators in IHA that affect the key hydrological indicators of the biological life history process as the amount of IHA data, then IHA data dimensionality reduction will no longer be difficult. The application of multi-indicator LOR dimension reduction can achieve wide recognition for the reliability of IHA evaluation results in the absence of data, because such data volume evaluation results are not only the evaluation of applied mathematical statistics, but also carry sufficient hydrological, ecological, and physical significance.

According to the ecological–hydrological response relationships that must be analyzed by IHA, we review the relevant literature and summarize a series of results obtained through experiments and observations (Barbour et al. 1999; Greig et al. 2007; Turkoglu 2010; Murgulet et al. 2016; Mwedzi et al. 2017; Xia et al. 2017). Additionally, we propose the LOR dimensionality reduction reference indicators with ecological and hydrological significance, as shown in Table 4. Some ecological and hydrological relationships are obviously only closely related to the indicators in a certain group. For example, if water temperature, sediment transport, and animal shelter area are related to Group 1, then all indicators in Group 1, or several of them, can be used as reference indicators for LOR. Such reference indicators are recorded as Key in Table 4. Taking the water temperature at Panzhihua station as an example, even the data volume required to produce a highly confident 5/90 LOR result can be reduced by 25%. Some ecological–hydrological relationships may have a greater relationship with individual indicators in two or three groups. For example, changes in the invertebrate community are importantly associated with Groups 3 and 4, while changes in the aquatic plant community are associated with Groups 2, 4, and 5 in IHA. The amount of data reflected by the LOR results of these indicators can then be collectively used as data constraints, and are recorded as Key in Table 4.

Table 4 | Referential ecological standard under multi-indicator LOR dimension reduction application

No.	Factors or Situation	Group 1: Magnitude of monthly water conditions (12 indices)	Group 2: Magnitude and duration of annual extreme water conditions (12 indices)	Group 3: Timing of annual extreme water conditions (2 indices)	Group 4: Frequency and duration of high and low pulses (4 indices)	Group 5: Rate and frequency of water condition changes (3 indices)
1	Temperature	Key				
2	Nutrients change		Key		Key	
3	Dissolved oxygen	Key				
4	Sediment transport	Key				
5	Invertebrates			Key	Key	
6	Zooplankton	Key				
7	Aquatic plants		Key		Key	Key
8	Diffusion of pollutants		Key			
9	Shelter area	Key				
10	Drought resistance		Key			Key

Study limitations and application

This study has several limitations. We generally believe that changes in the average volume of the flow become the driving force affecting the habitat, and that the reliability parameter value calculated by the LOR is based on the data sample mean as the evaluation criterion. In the case of identical amounts of data, the LOR result of the reference indicator of the data variation amplitude equalization feature will have higher credibility. Therefore, we also need to further understand the hydrological indicators, so that the LOR fitting results curve reflects a close ecological response relationship.

In addition, the influencing factors of LOR results are obviously positively correlated with the flow. This is because Panzhihua station, used in this paper, is located in a high-altitude area where human activity is relatively low, with continuous daily flow and relatively long data record length (>40 years). However, if it is replaced by other regions, or is impacted by other factors such as the relatively large impact of human activities, lower elevations of hydrological stations, or a larger watershed area, the correlation between LOR results and monthly average flow will require more hydrological sample data to justify.

Nonetheless, the rational application of the LOR dimension reduction method does not mean that the workload of data statistics and collection can be reduced in future work.

When using the IHA method to evaluate the ecological-hydrological relationship of a particular fish habitat, the reference indicators for ecologically significant LOR data reduction will vary across fish species. In the Yunnan region of China (Pan *et al.* 2015), owing to hydropower development and construction, fish reproduction mainly depends on proliferation and release. Therefore, the indices of Groups 3, 4, and 5, which are closely related to the breeding season, are not used as the main reference for LOR. In the United States, Missouri (Westhoff *et al.* 2016) temperature mainly affects the life history of the smallmouth bass, and Group 1 will be more reasonable as a reference index for LOR dimensionality reduction calculation. Therefore, different organisms in different regions have different key influencing factors and are subject to human influences (Paus *et al.* 2016). Our future workload should thus be placed on strengthening the statistics of key biological influencing factors. Table 4 presents only general statistics of some of the dimensionality reduction criteria; the presentation is not yet perfect. The LOR method will achieve its practical value only by means of a large number of statistics and experiments, which are used to continuously supplement and improve the priorities of key hydrological indicators for important periods of various organisms. This practical value will be achieved by refining the detailed list of biological species, geographic characteristics, and data reduction criteria as much as possible, after which the

multi-index LOR dimension reduction method will achieve its due value.

CONCLUSIONS

In this paper, we discuss in detail the results of multi-indicator evaluation of the LOR method and its application in data reduction of IHA calculations. Through the LOR reliability analysis of all the indicators in IHA, we found that the amounts of data required to produce a given LOR result are significantly different among the 33 indicators in IHA, and that the amount of data required has a consistent relationship with the size of average monthly flow and the variability of hydrological data. If the selection of data used is arbitrary when producing LOR results using the indicators in IHA, there will be no scientific basis or no actual data dimensionality reduction effect. By determining the objects of IHA analysis, the various key hydrological indicators affecting the target biological activities are used as reference indicators for LOR, and the multi-indicator LOR evaluation results achieve the practicality of IHA data dimensionality reduction. Continuous improvement of the relationship between biological activities and key hydrological indicators will become the development basis and future research direction of multi-index LOR data reduction.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude for the financial support of the National Basic Research Program of China (973 Program) (2015CB452701).

REFERENCES

- Afshari, S., Fekete, B. M., Dingman, S. L., Devineni, N., Bjerklie, D. M. & Khanbilvardi, R. M. 2017 [Statistical filtering of river survey and streamflow data for improving at-a-station hydraulic geometry relations](#). *Journal of Hydrology* **547**, 443–454.
- Barbour, M. T., Gerritsen, J., Snyder, B. D. & Stribling, J. B. 1999 *Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macro Invertebrates and Fish*, 2nd ed. EPA 841-8-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC.
- Carlisle, D. M., Falcone, J., Wolock, D. M., Meador, M. R. & Norris, R. H. 2010 [Predicting the natural flow regime: models for assessing hydrological alteration in streams](#). *River Research & Applications* **26** (2), 118–136.
- Eng, K., Carlisle, D. M., Wolock, D. M. & Falcone, J. A. 2015 [Predicting the likelihood of altered streamflows at ungauged rivers across the conterminous United States](#). *River Research & Application* **29** (6), 781–791.
- Engeland, K., Gottschalk, L. & Tallaksen, L. 2001 [Estimation of regional parameters in a macro scale hydrological model](#). *Hydrology Research* **32** (3), 161–180.
- French, N., Krajewsky, F. & Cuykendall, R. 1992 [Rainfall forecasting in space and time using neural network](#). *Journal of Hydrology* **137** (1–4), 1–31.
- Giustarini, L., Parisot, O. & Ghoniem, M. 2016 [A user-driven case-based reasoning tool for infilling missing values in daily mean river flow records](#). *Environmental Modelling & Software* **82** (C), 308–320.
- Govindaraju, R. S. & Rao, A. R. 2000 [Artificial neural networks in hydrology](#). *Journal of Hydrologic Engineering* **5** (2), 124–137.
- Greig, S. M., Sear, D. A. & Carling, P. A. 2007 [A review of factors influencing the availability of dissolved oxygen to incubating salmonid embryos](#). *Hydrological Processes* **21** (3), 323–334.
- Harrigan, S., Hannaford, J., Muchan, K. & Marsh, T. J. 2018 [Designation and trend analysis of the updated UK Benchmark Network of river flow stations: the UKBN2 dataset](#). *Hydrology Research* **49** (2), 552–567.
- Harvey, C. L., Dixon, H. & Hannaford, J. 2012 [An appraisal of the performance of data-infilling methods for application to daily mean river flow records in the UK](#). *Hydrology Research* **43** (5), 618–636.
- Hughes, D. A. & Smakhtin, V. 1996 [Daily flow time series patching or extension: a spatial interpolation approach based on flow duration curves](#). *International Association of Scientific Hydrology Bulletin* **41** (6), 851–871.
- Ilunga, M. & Stephenson, D. 2005 [Infilling streamflow data using feed-forward back-propagation \(BP\) artificial neural networks: application of standard BP and pseudo Mac Laurin power series BP techniques](#). *Water SA* **31** (2), 491–503.
- Isaacson, K. & Coonrod, J. 2011 [USGS streamflow data and modeling sand-bed rivers](#). *Journal of Hydraulic Engineering* **137** (8), 847–851.
- Jin, Y., Liu, J., Lin, L., Wang, A. & Chen, X. 2017 [Exploring hydrologically similar catchments in terms of the physical characteristics of upstream regions](#). *Hydrology Research* **49** (5), 1467–1483.
- Kennard, M. J., Pusey, B. J., Olden, J. D., Mackay, S. J., Stein, J. L. & Marsh, N. 2010 [Classification of natural flow regimes in Australia to support environmental flow management](#). *Freshwater Biology* **55** (1), 171–193.
- Khalil, M. F., Elshorbagy, K. A., Kassab, S. Z. & Fahmy, R. I. 1999 [Effect of air injection method on the performance of an air](#)

- lift pump. *International Journal of Heat & Fluid Flow* **20** (6), 598–604.
- Khalil, M., Panu, U. S. & Lennox, W. C. 2001 **Groups and neural networks based streamflow data infilling procedures**. *Journal of Hydrology* **241** (3), 153–176.
- Kusangaya, S., Warburton, M. L. & Garderen, E. A. V. 2018 **Evaluation of uncertainty in capturing the spatial variability and magnitudes of extreme hydrological events for the uMngeni catchment, South Africa**. *Journal of Hydrology* **557**, 931–946.
- Lake, P. S. 2008 **Flow-Generated Disturbances and Ecological Responses: Floods and Droughts**. In: *Hydroecology and Ecohydrology: Past, Present and Future* (P. J. Wood, D. M. Hannah & J. R. Sadler, eds). John Wiley & Sons, Chichester, pp. 75–92.
- Lytle, D. A. & Poff, N. L. 2004 **Adaptation to natural flow regimes**. *Trends in Ecology & Evolution* **19** (2), 94–100.
- Magilligan, F. J. & Nislow, K. H. 2005 **Changes in hydrologic regime by dams**. *Geomorphology* **71** (1), 61–78.
- Makhuvha, T., Pegram, G., Sparks, R. & Zucchini, W. 1997 **Patching rainfall data using regression methods. 1. Best subset selection, EM and pseudo-EM methods: theory**. *Journal of Hydrology* **198** (1–4), 289–307.
- Mobley, J. T., Culver, T. B. & Burgholzer, R. W. 2012 **Environmental flow components for measuring hydrologic model fit during low flow events**. *Journal of Hydrologic Engineering* **17** (12), 1325–1332.
- Murgulet, D., Murgulet, V. & Spalt, N. 2016 **Impact of hydrological alterations on river-groundwater exchange and water quality in a semi-arid area: Nueces River, Texas**. *Science of the Total Environment* **572**, 595–607.
- Mwedzi, T., Zimunya, T. G. & Bere, T. 2017 **Disentangling and ranking the influence of multiple stressors on macroinvertebrate communities in a tropical river system**. *International Review of Hydrobiology* **102** (5–6), 103–113.
- Nilsson, C. & Berggren, K. 2000 **Alterations of riparian ecosystems caused by river regulation: dam operations have caused global-scale ecological changes in riparian ecosystems. How to protect river environments and human needs of rivers remains one of the most important questions of our time**. *AIBS Bulletin* **50** (9), 783–792.
- Pan, L., Li, W. T., Xie, J. J., Que, Y. F., Zhao, N. & Yang, Z. 2015 **Length-weight relationships of eleven fish species from the middle reaches of Jinsha River, Southwest China**. *Journal of Applied Ichthyology* **31** (3), 549–551.
- Panu, U. S., Khalil, M. & Elshorbagy, A. 2000 **Streamflow data infilling techniques based on concepts of groups and neural networks**. In: *Artificial Neural Networks in Hydrology*. *Water, Science and Technology Library*, Vol. 36 (R. S. Govindaraju & A. R. Rao, eds). Springer, Dordrecht, The Netherlands, pp. 235–258.
- Paus, K. H., Muthanna, T. M. & Braskerud, B. C. 2016 **The hydrological performance of bioretention cells in regions with cold climates: seasonal variation and implications for design**. *Hydrology Research* **47** (2), 291–304.
- Poff, L. R. & Matthews, J. H. 2013 **Environmental flows in the Anthropocene: past progress and future prospects**. *Current Opinion in Environmental Sustainability* **5** (6), 667–675.
- Poff, L. R., Allan, J. D., Bain, M. B., Karr, J. R., Prestegard, K. L. & Richter, B. D. 1997 **The natural flow regime**. *Bioscience* **47** (11), 769–784.
- Poff, L. R., Bledsoe, B. P. & Cuhacyan, C. O. 2006 **Hydrologic variation with land use across the contiguous United States: geomorphic and ecological consequences for stream ecosystems**. *Geomorphology* **79** (3), 264–285.
- Poff, N. L., Richter, B. D., Arthington, A. H., Bunn, S. E., Naiman, R. J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B. P., Freeman, M. C., Henriksen, J., Jacobson, R. B., Kennen, J. G., Merritt, D. M., O’Keeffe, J. H., Olden, J. D., Rogers, K., Tharme, R. E. & Warner, A. 2010 **The ecological limits of hydrologic alteration (ELOHA): a new framework for developing regional environmental flow standards**. *Freshwater Biology* **55**, 147–170.
- Proctor, E. N., Wu, J., Hynes, M. J., Dosch, J. & Simeone, D. M. 2011 **Inhibition of the notch signaling pathway targets cancer stem cells in pancreatic ductal adenocarcinoma**. *Journal of Surgical Research* **165** (2), 303–305.
- Richter, B. D., Baumgartner, J. V., Powell, J. & Braun, D. P. 1996 **A method for assessing hydrologic alteration within ecosystems**. *Conservation Biology* **10** (4), 1163–1174.
- Richter, B., Baumgartner, J. & Wigington, R. 1997 **How much water does a river need?** *Freshwater Biology* **37** (1), 231–249.
- Ryo, M., Iwasaki, Y., Yoshimura, C. & Oliver, C. S. V. 2015 **Evaluation of spatial pattern of altered flow regimes on a river network using a distributed hydrological model**. *Plos One* **10** (7), e0133833.
- Scarcella, G., Grati, F., Bolognini, L., Domenichetti, F., Malaspina, S. & Manoukian, S. 2016 **Time-series analyses of fish abundance from an artificial reef and a reference area in the central-adriatic sea**. *Journal of Applied Ichthyology* **31** (S3), 74–85.
- Shrestha, R. R., Peters, D. L. & Schnorbus, M. A. 2014 **Evaluating the ability of a hydrologic model to replicate hydro-ecologically relevant indicators**. *Hydrological Processes* **28** (14), 4294–4310.
- The Nature Conservancy 2009 **Indicators of hydrological alteration version 7.1 User’s Manual**. <https://www.conservationgateway.org/Files/Pages/indicators-hydrologic-alt.aspx47.aspx>
- Timpe, K. & Kaplan, D. 2017 **The changing hydrology of a dammed Amazon**. *Science Advance* **3** (11), e1700611.
- Todini, E. 2011 **History and perspectives of hydrological catchment modelling**. *Hydrology Research* **42** (2–3), 73–85.
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W. & Verstraten, J. M. 2005 **Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation**. *Water Resources Research* **41** (1), 143–148.
- Turkoglu, M. 2010 **Winter Bloom of coccolithophore *Emiliania huxleyi* and environmental conditions in the Dardanelles**. *Hydrology Research* **41** (2), 104–114.

- Wang, H., Chen, Y., Deng, H. & Pan, Y. 2015 [Detecting changes in extreme streamflow in the Tarim River, Northwest China](#). *Quaternary International* **380**, 149–158.
- Westhoff, J. T., Paukert, C., Ettinger-Dietzel, S., Dodd, H. & Siepker, M. 2016 [Behavioural thermoregulation and bioenergetics of riverine smallmouth bass associated with ambient cold-period thermal refuge](#). *Ecology of Freshwater Fish* **25** (1), 72–85.
- Xia, S., Yu, L. & Chen, B. 2017 [Effect of water level fluctuations on wintering goose abundance in Poyang Lake wetlands of China](#). *Chinese Geographical Science* **26** (2), 1–11.
- Yang, T., Asanjan, A. A., Welles, E., Gao, X., Sorooshian, S. & Liu, X. 2017 [Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information](#). *Water Resources Research* **53** (4), 2786–2812.

First received 20 May 2018; accepted in revised form 3 September 2018. Available online 3 October 2018