Time-Series Data Mining of Minimum Design Height for River Bridge Deck Using Seasonal Trend Decomposition

Haiyan Xie and Xi Hu
Department of Technology, Illinois State University, Turner Hall 5100, Normal, IL 61790-5100, USA
Email: hxie@ilstu.edu;

Abstract. Using the traditional calculation of minimum design height (MDH) could not provide the effective decision support for engineers under complex situations. This research proposes a time-series data mining system for the decomposition and analysis of seasonal trends of hydrologic fluctuations, using stage-discharge curves, Log-Pearson frequency curves, and plotting positions. The research purpose is to provide the decision support based on precise time series for bridge design and feasibility analysis. Prior to conducting the gage-height design for bridges, this study investigates the construction site of the bridge at Kinston Mines of Illinois River in the United States and processes the correlation tests between gage heights and hydrologic discharge. In addition, the study simulates and verifies the system outcomes using Mann-Kendall Mutation, stability tests, and non-linear regressions. The results show the correlation coefficient of 0.9719 for the discharge trends from 1994 to 2017. In addition, the study selects the best-fitted theoretical frequency curve based on the most appropriate convergence to the empirical frequency curve. Overall, this research achieves satisfactory results for the MDH design and has substantial influence for engineers and designers to meet design requirements and save project costs. The data mining method sheds light upon data analysis for engineering design.

1. Introduction
The minimum design height (MDH) of bridge deck is always an inevitable factor in the feasibility study of a new river bridge [1, 2]. The traditional methods to calculate the MDH are usually based on the statistical means of multiple observations on hydraulic data [3-7]. However, the current practice of bridge maintenance requires the considerations of the hydrological fluctuations of rivers [5, 6]. Particularly, hydrologic discharge is the volume of water flowing past a given point in the stream in a given period of time which is an important indicator for river changes [5, 6, 8]. It is named after stream-flow by fluvial hydrologists and plays a critical role in the study of hydrogeological cycles. The most common form of discharge is water flow derived from net precipitation and melting glaciers [8]. Bridge design and maintenance should consider gage height or stage, which is the water height in the stream above a reference point [5, 6, 8, 9]. Empirically, there is a quantitative relation between gage height and discharge in a specific cross-section of river, which can be manifested by Stage-Discharge Curve [10]. However, in the real world, transport systems connect geographical isolation by highway, railway, and airlines. Hence, it’s undoubted that bridges, which connect two points over two sides of rivers or valleys at specific heights, are significant components of the transportation systems. In the planning and design phases of river bridges, engineers and planners should always consider flooding argumentation in order
to make sure that this newly-constructed bridge meet the specific flooding control requirement. This ensures that it will not be submerged by a once-in-a-century flood [10, 11]. Based on a design flood, bridge engineers should define the flood discharge for the particular project according to historical time-series data. Then they the graph the stage-discharge curve correspondingly to find the conforming gage height. In addition, when deciding on the minimum vertical height from the lowest point of the bridge center and surface water, engineers and planners should take into account the navigable clearance height if this waterway is eligible for river transportation [12]. For example, on Illinois River in the United States, the minimum design height of bridge deck consists of designed flood level and navigable clearance height [3].

Researchers noticed the knowledge gap between the traditional methods of bridge design and the current needs of cost control on maintenances and repair. Engineers, designers and planners need to have improved understanding on how to estimate and predict the possible river flow changes. For example, Amirreza and Arash stated briefly that the relationship between stage and discharge was one of the methods used to measure the continuous river flow [5]. Researchers proposed ideas for the estimation of stage-discharge curve in open channels [5, 6]. For instance, the model of Ahmadi et al. introduced the rating curve taking advantage of the discharge information at a referenced water level [6]. Another method was suggested by Fread DL using numerical solutions of the dynamic, unsteady, non-uniform flow equations [10, 11]. Recent studies proposed inverse methods based on Manning’s equation to estimate the discharge from remotely sensed observations of water surface elevation, width and slope [12, 13]. Researchers discussed that useful discharge estimates could be retrieved for certain types of river reaches using satellite images. The estimates provided information on the river widths and the log-linear relationships of at-a-station hydraulic geometries (AHG) [14]. The log-Pearson type III distribution is the upgraded transformation of Pearson type III distribution and currently proposed to do flood-frequency analysis by U.S. federal agencies as described by B17.IACWD-1982 [15]. The method of moment’s estimator in log space might be suitable to estimate the parameters of the distribution [7]. Nevertheless, researchers devoted much time and effort to update the methods to the accuracy of stage-discharge curve and estimation of distribution parameters. Furthermore, few scholars gathered those two methods in practice though both are functional tools for flooding control or flood argumentation [12, 13]. In this study, the authors consider using the combination of these two technical means to determine the minimum design river bridge height and explicitly explain the feasibilities.

2. Literature Review

In hydrology, a rating curve is a graph of discharge versus stage for a given spot on the stream at a gauge station [9]. Generally, the rating curve is plotted as stage on Y-axis versus discharge on X-axis [8]. At a specific gauge station, the stream discharge is measured, and the stage can also be acquired by water level gauge. Theoretically, the curve slope diminishes gradually with a level off to 0 [8, 9].

Log-Pearson III distribution (LP3) is the updated transformation of Pearson III distribution (P3) [13, 14, 16]. In term of P3, there are shape and location parameters α, β and ε with the probability density function in Equation (1):

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left( \frac{x-\mu}{\beta} \right)^{\alpha-1} e^{-\beta(x-\mu)}$$

where $\alpha > 0$, $(x-\mu)/\beta > 0$, $\Gamma(\alpha)$ complete gamma function.

The density function can be confirmed if these three parameters defined. The parameters of the P3 distribution are functions of the first three population moments ($\bar{x}, \sigma_x$ and $y_x$): $\alpha = \frac{4}{y_x^2}$; $\beta = \frac{\sigma_x y_x}{\sqrt{2}}$; $\epsilon = \bar{x}(1 - 2 \frac{\sigma_x}{y_x})$, in which, $\bar{x}$, $\sigma_x$ and $y_x$ are sample mean, the variance coefficient and skewness coefficient(Cs) respectively. Similarly, the LP3 distribution has probability density function:

$$f(q) = \frac{1}{|\beta|q\Gamma(\alpha)} \left( \frac{\ln q - \mu}{\beta} \right)^{\alpha-1} e^{-\left( \frac{\ln q - \mu}{\beta} \right)}$$

Where, $\ln q = x$ has a P3 distribution with parameters $\alpha$, $\beta$ and $\epsilon$. The density function is defined for $\alpha > 0$, and $0 < q < \exp [\epsilon]$ for $y_x < 0$, and $\exp [\epsilon] \leq q$ for $y_x > 0$. When $y_x = 0$, the PL3 distribution
converges to a LN distribution.

However, the differences between LP3 and P3 are not limited to the density functions. As far as variance coefficient (Cv) or \( \sigma_x \) for LP3, it should be calculated as follow:

\[
\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(K_i - 1)^2}{n-1}}
\]

(3)

In which, \( K_i \) is modulus coefficient; \( n \) is sample size. In term of skewness coefficient \( \sigma_x \) (Cs), it can be estimated by Cv. For instance, \( Cs = (2 - 3) \cdot Cv \). In hydrological computation, when a specific frequency \( P \) given, the random response variable \( x_p \) is acquired through an analysis of frequency density curve. In practice, the complicated process of calculation has been simplified in order to make it more efficient. When conducting frequency computation, based on the calculated Cv value, a hypothetical Cs should be given in order to find the \( \phi \) with a specific frequency value of \( P \). Then, values of \( \phi \) and Cv can be substituted into formula 5 to get the according \( \ln x_p \).

\[
\phi = \frac{\ln x - \ln \bar{x}}{C_v \ln x}
\]

(4)

\[
\ln x_p = \ln \bar{x} \cdot (1 + C_v \cdot \phi)
\]

(5)

where, \( \phi \) is standardized value, named coefficient of deviation mean.

3. Method Design

In this research, the following steps are used for the calculation and analysis of time series of the seasonal trends of hydrological data. The steps can help to calculate the minimum design height of river bridge deck minus navigable clearance.

- Rank the observed values of discharge data from high to low first, calculate the empirical frequency of each discharge item and graph empirical curve thereafter.
- Define hydrological frequency distribution curve (Log-Pearson III Frequency Curve).
- Work out the values of \( \ln \bar{x} \), Cv and Cs by using Central Moment Method and suppose them as the subjects for first attempt. Due to large error comes out when calculating Cs by Central Moment, it is generally estimated by 2 - 3 times of Cv value.
- The identified \( \ln \bar{x} \), Cv and Cs help to calculate \( \phi \). Plot \( \ln \bar{x} \) as y-axis versus \( P \) as x-axis, Log-Pearson frequency curve is graphed.
- Change the Cs then find the decent Log-Pearson frequency curve which fits the empirical curve well. Then, take advantage of those parameters of this well-fitted curve as population statistics for practical applying.
- Calculate the design discharge value with a designated frequency.

4. Data Collection and Analysis

All the data used in this study is derived from USGS. A gauge station on the stream at Kinston Mines recorded monthly data of stage and discharge of Illinois River from October 1993 to July 2018. Due to malfunction of stage measurement instrument or tele-transmission failure, especially before 2002, there were some deletions of monthly data of gauge height. But the stage data is sufficient for this study because the lost data only stands for a small proportion.

Prior to graph stage-discharge curve, it’s necessary to test the correlation between gage height and discharge to ensure that there is a certain relation between these two factors. The monthly average data of gage height and discharge is used to do correlation test; the result of correlation test indicates that the correlation coefficient is 0.9719.

Mann-Kendall (M-K) test (Mann 1945; Kendall 1975), a non-parametric way, in which is widely utilized to evaluate the statistically significant trends in hydrological and climatological time series. So as to make sure whether there are extreme values of discharge data and if the time series discharge data are acceptable within a specific interval, M-K test is conducted. With a significant level \( \alpha = 0.05 \) and \( U_{0.05} = \pm 1.96 \), figure 1 (a) shows that the discharge of Illinois River at Kinston Mines has a distinct
uptrend according to both UFk and UBk curves. In addition, there are three UFk excess values which are assigned with the year 2011, 2016 and 2017 respectively. The difference between UFk and upper limit \( U_{0.05} = 1.96 \) is most significant. Hence, it’s affirmative that there is at least one monthly average extreme discharge data in 2017 and 2017 should be selected as typical year for subsequent definition of design water level.

Due to some lost data points of gage height, of which some assigned discharge data is not paired, so these discharge data points were not undertaken in drawing the stage-discharge curve. In the end, the data points of gage height and discharge used for drawing following curve (as shown in figure 1 (b) and (c)) were one-to-one correspondence (the data will be available upon request). Based on the observed values of discharge, the data is ranked from high to low and serial numbers \( i (i=1, 2, \ldots 24) \) are the newly-ranked data. Figure 1(d) shows the empirical frequency calculated by following formula:

\[
P_i = \left( \frac{i}{n+1} \right) \times 100\%
\]

In this research, three different attempts of plotting position are carried out. The results are shown in figures 2 (a) – (f) respectively. Figures 2(g), 2(h) and 3 combine the outcomes of those three attempts to analyse which curve with a set of \( C_v \) and \( C_s \) is best-fitted and shrinks towards the empirical frequency curve. The outcome of first attempt is the most ideal, so the \( C_v \) and \( C_s \) are the estimated parameters of

\[
\ln x_p = \ln \bar{x} \cdot (1 + C_v \cdot \varnothing)
\]

Furthermore, this curve will be used to do flood forecasting at Kinston Mines of Illinois River. The \( \ln \bar{x} \) is a constant value of 10.49, calculated from the average of annual maximum discharge values within 24 years.

![Discharge Curves](image)
5. Conclusions

This study discusses the minimum design height of a river bridge. In terms of the correlation test between gage height and discharge, the correlation coefficient is significantly closed to 1. In other words, it means the stage-discharge curve is drawn based on reliable data points. The M-K test was well applied to verify the annual discharge variation trend. Precisely, the discharge of Illinois River at Kinston Mines has an upward tendency from 1994 to 2017, which means the according gage height will also goes up to some extent. Hence, when doing estimation of design gage height at Kinston, Mines, the design discharge should be amplified slightly greater. Compared Pearson Frequency Curve with Log-Pearson Frequency Curve, the principles of each other is largely identical but with minor differences. Log-Pearson Frequency Curve is satisfyingly used to do plotting positions.

The research shows that design discharge with a designated return period can be deduced by the theoretical frequency curve, and then the design gage height is worked out through stage-discharge curve. The design discharge with a 10-year return period corresponds to a 23-ft gage height at Kinston Mines of Illinois River. Furthermore, when considering the minimum height of a hypothetical river bridge, it can be defined by navigable clearance and design gage height.

This hypothetical study is for the minimum design height of river bridge deck and for the first time, shows the possibility to combine the function of stage-discharge curve, Log-Pearson III Frequency Curve
Curve or Pearson III frequency curve. In general, this study is an exploration to connect hydrological system and transport, which inspires more researchers to step forwards an interdisciplinary domain. There are still several limitations in this study. For instance, the study subject is merely a hypothetical object, so there is no data material collected about navigation system, such as the ship height. In addition, in the process of plotting positions, it’s inevitable to generate some personal errors due to the curves were delineated manually. In future research, the authors plan to collect navigable clearance data to complete the study.

Acknowledgments
Authors wishing to acknowledge the financial support from the Sustainability Grant of Illinois State University (10400-01-51000000).

References
[1] Cheng Y-M, Leu S-S. Integrating data mining with KJ method to classify bridge construction defects. Expert Systems with Applications. 2011;38(6):7143-50.
[2] Wong K-Y. Design of a structural health monitoring system for long-span bridges. Structure and Infrastructure Engineering. 2007;3(2):169-85.
[3] IDOT. Standard Specifications for Road and Bridge Construction. Illinois Department of Transportation. http://www.idot.illinois.gov/Assets/uploads/files/Doing-Business/Manuals-Guides-&-Handbooks/Highways/Construction/Standard-Specifications/Standard%20Specifications%20for%20Road%20and%20Bridge%20Construction%202016.pdf. 2016. p. 1225.
[4] AASHTO L. AASHTO LRFD bridge design specifications. Transportation (Amst) American Association of State Highway and Transportation Officials, Inc: Washington, DC. 2007.
[5] Kavousizadeh A, Ahmadi A. High-performance approach for estimating stage-discharge curves in the open channels. Journal of hydrology. 2018; 565:197-213.
[6] Maghrebi MF, Ahmadi A. Stage-discharge prediction in natural rivers using an innovative approach. Journal of hydrology. 2017; 545:172-81.
[7] Haan CT. Statistical methods in hydrology: The Iowa State University Press; 1977.
[8] Kennedy E. Discharge ratings at gaging stations: US Geological Survey Techniques of Water-Resources Investigations, book 3, chap. A10. 1984;59.
[9] Herschy RW. Hydrometry: principles and practices. 1999.
[10] Fread D. A dynamic model of stage-discharge relations affected by changing discharge. 1973.
[11] Fread D. COMPUTATION OF STAGE-DISCHARGE RELATIONSHIPS AFFECTED BY UNSTEADY FLOW 1. JAWRA Journal of the American Water Resources Association. 1975;11(2):213-28.
[12] Durand M, Neal J, Rodriguez E, Andreadis KM, Smith LC, Yoon Y. Estimating reach-averaged discharge for the River Severn from measurements of river water surface elevation and slope. Journal of Hydrology. 2014; 511:92-104.
[13] Garambois P-A, Monnier J. Inference of effective river properties from remotely sensed observations of water surface. Advances in Water Resources. 2015; 79:103-20.
[14] Gleason CJ, Smith LC. Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. Proceedings of the National Academy of Sciences. 2014;111(13):4788-91.
[15] Pierson TC. Hyperconcentrated flow—transitional process between water flow and debris flow. Debris-flow hazards and related phenomena: Springer; 2005. p. 159-202.
[16] Turban E, Sharda R, Delen D, Efraim T. Decision support and business intelligence systems: Pearson; 2014.