An Alternative Means to Analyze and Communicate Flood Risk

Reza Noubary, Ph.D.[1]
rnoubary@bloomu.edu
Bloomsburg University of Pennsylvania

Roberta Soltz, Ph.D.[2]
robertsoltz@gmail.com

Abstract

Property and business owners as well as their elected or appointed decision-makers in small, rural or isolated communities do not have access to the type of data, maps or interpretive methodologies frequently cited as "best practices" by experts. They do, however, generally record and aggregate data regarding the date and height of flood water on a point-by-point basis. We propose to leverage this data, where available, by using one or more well-known statistical tools that can be sourced by project managers and others to improve understanding, collaboration and decision-making. Application to the data derived from the Susquehanna River in the vicinity of Bloomsburg, Pennsylvania suggest that these tools can succinctly address point-specific issues in a simple, familiar framework to streamline the planning process. Alternative analyses such as these may be valuable to engineers and others with academic, professional and humanitarian interests in developing countries or regions.

Introduction

Periodic floods have a major influence on the scope and quality of lowlands (650 feet elevation or lower) which harbor most of the world's most populous cities. To make matters worse, cities and other urban areas are even more prone to flooding than rural environments due to the relatively greater area covered by pavement and other impermeable materials which limit percolation, hasten runoff and increase risk. Despite this, people are increasingly attracted to urban areas due to the availability of

---

[1] Professor, Department of Mathematics, Computer Science and Statistics, Bloomsburg University, Bloomsburg PA 17815
[2] Research Associate, University of Pittsburgh, Asian Studies Center, 4104 Posvar Hall, 230 S. Bouquet St., Pittsburgh, PA 15260

employment, education, health services and a variety of other cultural and economic factors. In fact, it is estimated that global flood damage could exceed $1 trillion annually by 2050.

On a world-wide basis, government has largely failed to solve this problem. Rapid development continues in flood prone areas throughout the world, and people as well as their investments remain at risk. In many cases, the capacities of existing flood control measures such as dams, sea walls and other barriers are already woefully inadequate given the loss of natural percolation areas that rapid urbanization brings. Even if the funding could be found to build new flood control infrastructure, the costs would be staggering.

It seems to us that the prudent and responsible alternative would be to limit development in flood prone areas to agricultural, park land, or other uses where periodic flooding would not cause significant economic or social impacts.

However, decision-makers are often at a loss to evaluate actual flood risk. Technical studies are often prohibitively expensive and take a very long time to complete, but there are alternatives. Whereas most small, isolated communities do not have access to sophisticated studies, current maps and other best practice tools (see U.S. Army Corps of Engineers *How to Communicate Risk* [http://www.corpsriskanalysisgateway.us/riskcom-toolbox-communicate.cfm]), they do generally collect and aggregate the date and height of floodwater (Brenner et.al. 2013, Arensburg and Hutt 2007)

In this paper, we propose a few alternative methods to evaluate flood risk in smaller or more marginalized communities that do not have updated maps or more sophisticated means to analyze or interpret their records. These alternatives are based on statistical concepts that leverage simple data like water height and date of flood events. The methods proposed involve the analysis of extremes, exceedances, excesses, record heights, and ultimate heights of floodwater.

To demonstrate, methods are applied to data recorded since 1850 from the Susquehanna River at the Town of Bloomsburg in central Pennsylvania. Bloomsburg offers a particularly valuable opportunity to test these tools in that the data have been accurately and meticulously recorded for over 150 years. Further, floods are a significant problem in the area, and they compromise critical infrastructure such as drinking water, sewer, transportation and emergency services.

The River in Bloomsburg floods when the water level exceeds 19 feet. Since 1850, there have been 38 floods with 32.75 feet establishing the highest recorded water level. We hope by using the values derived from these data, we offer a useful and "user friendly" platform to consider flood risk to decision-makers and their constituents in small towns like Bloomsburg and elsewhere in the world especially where flooding can have catastrophic consequences.

Empirical Rule

For the Bloomsburg data, the mean, median and standard deviation are respectively 24.3 ft., 23.4 ft., and 3.4. Also the percentages of floods with sizes to within one, two, and three standard deviations from the mean are respectively 71, 92, and 100 %. These percentages agree with the Empirical rule indicating the normality this data set for. Thus, the probabilities of a future flood greater than 30.9 ft. or 32.1 ft. are respectively 5% and 1%. Also the probability of a record flood (greater than 32.75 ft.) is 0.006. This indicates that there is a small chance that a randomly selected future flood cresting level will exceed the present record.

Exceedances

The Theory of Exceedances deals with the number of times a specified threshold is exceeded. Assuming independent and identically distributed events (floods) we may wish to determine the probability of r exceedances in the next n occurrences. To apply this to Bloomsburg data, we note that since 1850 in there have been 38 floods exceeding 19.8ft with the largest 32.7ft. So, the mean expected value, variance and standard deviation of the number of exceedances of 32.7ft during the next 10 and 100 years would respectively be:

10/173 = 0.058, (10) (172) (183)/ (173)2(174) = 0.06, and 0.246

100/173 = 0.58, (100) (172) (273)/ (173)2(174) = 0.902, and 0.95

These estimates have relatively large standard deviations due to a small sample size.

Records

This theory deals with values that are strictly greater than or less than all previous values. Usually the first value is counted as a "record." A value is a record (upper record or record high) if it exceeds or is superior to all previous values.

To predict the future records, we have developed a simple method utilizing the following results of the theory of records for an independent and identically distributed sequence of observations (McDonnell 2013).

(a) If there is an initial sequence of n_1 observations and a batch of n_2 future observations, then the probability that this additional batch contains a new record is $n_2/(n_1+n_2)$.

(b) For large n, $P_{r,n}$ the probability that a series of length n contains exactly r records is:

$$P_{r,n} \sim \frac{1}{(r-1)!n}(\ln(n)+\gamma)^{r-1}$$

For the Bloomsburg data, we note that there have been 38 floods exceeding 19.8ft. This gives a rate of 38/166 per year. Thus, the probability estimate of a record flood (as in (a), above) during the next 10 years is:

10/176 = 0.057 ~ 0.06.

Also, using (b), the probability of 3 records in 38 observations is 0.003. With the above rate, we expect two floods in the next 10 years and the probability of 3 records in 40 observations is 0.002930091. Hence, P (no record in next 10 years) = P (3 records in 40 observations) / (3 records in the 38 observations) = P (3 records in 40 observations) / P (3 records in 38 observations) = 0.94. This gives the probability of a record in the next 10 years as 0.06 or 6%.

Excesses

In this approach, the probabilities of future large floods are calculated by developing models for the upper tail of the distribution for height of floodwaters. Because values above an appropriate threshold carry more information about the future large floods, this approach is reasonable. Here, one usually assumes that the tail of the distribution for flood sizes belongs to a given parametric family and then proceeds to do inference using excesses, that is, the floods greater than some predetermined value. It has been shown that the most appropriate model for tail is the so-called Pareto Distribution that includes models for short, medium, and long tails. We applied this approach to the Bloomsburg data and found that the best fit is a short tail based on the largest four floods as follows:

$$\bar{G}_l(x) = 1 - (1 + 0.24327x)^{-0.15764} \qquad 0 < x < 4.1106.$$

This led to the following upper bound for floods in Bloomsburg:

$$V_{\max} = 32.7 \text{ ft.}$$

This is virtually the same as the largest flood in Bloomsburg that occurred in 2011.

Ultimate Flood

In terms of predictive value, we can avoid large standard errors and provide a confidence interval for the upper bound based on the most recent large floods or record floods. Let Y and $Y_1, Y_2, \dots, Y_n$ represent flood size for a given region where:

$$Y_1 \le Y_2 \le \ldots \le Y_n$$

Assuming that the distribution function *F(y)* has a lower endpoint and certain conditions are satisfied, a level (1 – *p*) confidence interval for the maximum of *Y* is (De Haan 1981):

$$\{Y_1 + (Y_2 - Y_1)\big/[(1-p)^{-k} - 1],\ Y_1\}$$

De Haan (1981) has also shown that

$$\frac{ln\ m(n)}{ln[(y_{m(n)} - y_3)/(y_3 - y_2)]}$$

is a good estimate for *k*.

To apply this result, we need to choose an integer *m (n)* depending on *n* such that $m(n) \to \infty$ and $m(n)/n \to 0$ as $n \to \infty$. It is shown that the following choice works well even for the worst case:

$$m(n) = \sqrt{eT_r} + \sqrt{t_r} = \sqrt{2.718282T_r} + \sqrt{t_r}$$

For the Bloomsburg data, T_r = 107 and t_r = 1 and *m(n)* = 18. Since y_1 = 32.75 ft., y_2 = 32.70ft, y_3 = 31.2ft, y_{18} = 23.5ft. We have k = 1.767 leading to a 90% one-sided prediction interval for the upper bound as (32.75ft, 33.0ft).

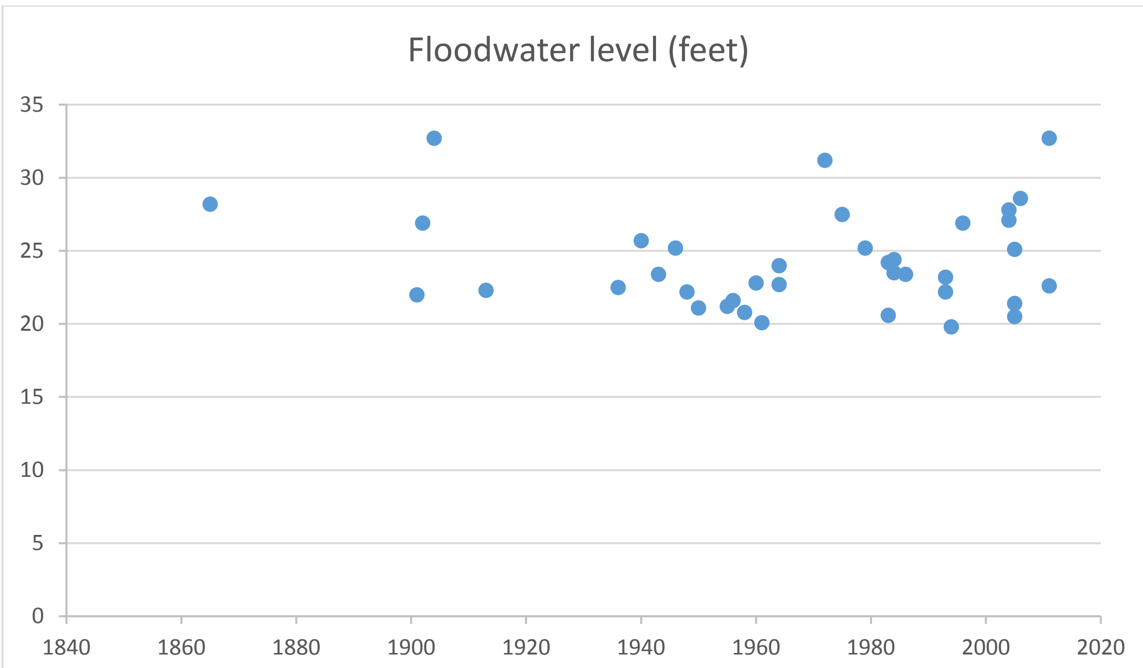
Summary

The application of alternative statistical methodology demonstrated in this paper for the Susquehanna/Bloomsburg area is a useful complement to best practice methods to evaluate flood risk in that:

1. They can be used to communicate point-specific risk, and answer the question most commonly asked -- "*How likely is it that this particular area will be inundated in any given year?"*

2. Risk can be communicated using a familiar, simple scale such as "*On a scale of one-to-ten, this area scores 8 or has an 80% chance of flooding in any given year.*"
3. Threat can be evaluated quickly with relevant data and without immediate reference to maps which may not be available.
4. Such analyses can serve as an important touchstone between decision-makers, developers and others early and effectively in the flood control process.

Instances of flooding (water level >19.0 feet) at a point of reference in the vicinity of Bloomsburg, PA.



References

1. Arensburg, A. and Hutt, S. Challenges and Issues for Water Management in Northwest China. IDEAS, July 9, 2007.

2. Brenner, R., Keung, C., Rosenblum, B., Soltz, R., Wolfe, S. (2013). "China's Loess Plateau-A Region of Heterogeneous Environmental Communities. *Proc.,* GASI, Rome, Italy.
3. Davis, R. and S. Resnick, Tail estimates motivated by extreme-value theory. *Annals of Statistics.* 12 (1984) 1467-1487.
4. DeHaan, L. Estimation of the minimum of a function using order statistics. *Journal of the American Statistical Association.* 76 (1981) 467- 469.
5. Department of the Army, Office of the Chief Engineer. EM 1110-2-1411, March 1965.
6. Federal Emergency Management Agency [FEMA], Region 10. The 100-year Flood Myth. Pages 1-6, undated training document.
7. Hill, B.M. A Simple general approach to inference about the tail of a distribution. *Annals of Statistics.* 3 (1975) 1163-1174.
8. McDonnell, T. Global Flood Damage. *Mother Jones.* August 18, 2013
9. Pickands, J. Statistical inference using extreme order statistics, *Annals of Statistics.* 3 (1975) 131-199.
10. Smith, R.L. Forecasting records by maximum likelihood, *J. of the Amer. Stat. Assoc*. 83 (1988) 331-338.