Current Methods for Flood Frequency Analysis

Beth Faber, PhD, PE
USACE, Hydrologic Engineering Center

Workshop on Non-Stationarity, Hydrologic Frequency Analysis, and Water Management
January 13, 2010
Flood Frequency Analysis

- **Risk management** requires estimates of probability of extreme events with negative consequences.
- Need a relationship between flood magnitude and annual exceedance probability:
  - A probability distribution for the largest flow in any year.
  - Should represent the present and the planning horizon.
- When possible, we infer probability from a record of observations using statistical analysis:
  - Using the past as a representation of the future.
  - When we don’t have flow record, use precipitation distribution and rainfall/runoff modeling to produce flow distribution.
Long Record Gaged Site

- Simple case: gaged series of annual maximum flows
  - largest event each year, as opposed to all events above a threshold (partial duration series)

- Treat these flows as a random, representative sample of realizations from the flood population of interest

- Generally, we assume the sample is IID
  - annual peak flows are random and independent
  - peak flows are identically distributed – homogeneous data set
  - sample is adequately representative of the population

(estimate of the distribution improves with sample size)
Wappinger Creek near Wappingers Falls, NY
USGS 01372500

80 years of annual peak flows
Wappinger Creek near Wappingers Falls, NY
USGS 01372500

80 years of annual peak flows
Empirical vs Theoretical Distribution

• Often fit a known theoretical distribution by estimating distribution parameters from the data set

• Some benefits are:
  – Smooth fit to the data set
  – Compact representation of the data
  – Better description of the potential range of observations than the sample data set
  – A reproducible basis for extrapolating extreme quantiles
Bulletin 17B

• “Guidelines for Determining Flood Flow Frequency” developed by a committee of federal agencies, 1976-1982

• Intended to provide a consistent, objective, reproducible method for flood frequency analysis
  
  — various agencies engaged in flood analysis will produce the same frequency estimates

• Applies to gaged, unregulated annual peak flows

• Recommends use of Log Pearson type III distribution...
Bulletin 17B

- Recommends use of Log Pearson type III distribution
  - estimate parameters from data by method of moments
  - local mean and variance, and skew weighted with regional

B17B includes procedures for:
  - incorporating historical information
  - estimating regional skew coefficient – includes skew map
  - adjusting for low outliers
  - defining sampling error (confidence intervals)

- The Bulletin is undergoing an update to incorporate improved techniques for these procedures
Homogeneity Assumption

- Annual peaks are considered a sample from a single probability distribution – *identically distributed*
  - they must result from the same hydrologic process
    - or multiple processes with consistent frequency of occurrence
  - since we collect data over time, the process must be unchanging – *stationary*
    - or changing in transitory or cyclic manner, for which the cycle is short and repeats in the record
Homogeneity Assumption

• When annual peak floods are not the result of a single process...
  – for example, most annual peaks resulting from snowmelt, some from rare rainfall events
  – if both processes are well represented, might still fit a single overall distribution
  – if few largest are one process and noticeably depart from the range of other process, use Mixed Distribution Analysis
    • develop separate, complete sample for each process,
    • fit a distribution to each, and
    • combine with OR-condition
Homogeneity Assumption

• When the process is not stationary...
  – If the reason for the change can be observed and quantified, adjust the data set to correct earlier events to the current or future condition (remove the trend)
  – Example: urbanization.
    • Gaged flood events are reproduced in a rainfall/runoff model, then re-simulated with current or future urbanization.
    • In some cases, a regional relationship is developed between %-impervious, return period, and a peak factor
  – KEY: the trend must be observable to be removed, and predictable to have an estimate of a future condition
Adjustment for Urbanization
Adjustment for Urbanization
Sampling Error

- The greatest source of error in a frequency analysis is inadequate data for estimating a distribution.
- Due to random occurrence of events, sample might not be representative of the population.
- The skew coefficient in particular is very sensitive to sampling error.

![Bush Kill at Shoemakers, PA](chart)

**Bush Kill at Shoemakers, PA**

- **Annual peak flow**
- **1% chance flow**
- **Skew coefficient**

LP3 fit of increasing sample 1909-2008
Sampling distribution for skew

Monte Carlo experiment

\[ \mu = 3 \]

\[ \sigma = 0.3 \]

\[ \gamma = 1 \]

\[ N = 2000 \]

"Sampling Distribution" of Station Skew

actual skew = 1, sample size 20 to 100
Sampling Error – sample size 20

Wappinger Creek at Wappinger Falls, NY
1929 - 2008

20 year portions of the record

Return Period

Streamflow (cfs)

Exceedance Frequency
Sampling Error – sample size 20

Wappinger Creek at Wappinger Falls, NY
1929 - 2008

Return Period

Streamflow (cfs)

Exceedance Frequency

- fitted LP3 curve
- 20 year records resampled-w/o-replacement

Wappinger Creek at Wappinger Falls, NY
1929 - 2008
Sampling Error – sample size 20

Wappinger Creek at Wappinger Falls, NY
1929 - 2008

Return Period

Streamflow (cfs)

Exceedance Frequency

randomly-generated 20 year records

Wappinger Creek at Wappinger Falls, NY
1929 - 2008
Sampling Error – sample size 60

Wappinger Creek at Wappinger Falls, NY
1929 - 2008

randomly-generated
60 year records
Sampling Error – sample size 100

randomly-generated 100 year records

Wappinger Creek at Wappinger Falls, NY
1929 - 2008
Reducing Sampling Error

• Using a larger data set reduces sampling error and provides better estimate of the frequency relationship
  – narrower confidence interval

• Can create a larger sample of the extremes by including historical events that are the largest of a longer period than the systematic record
  – the fact that a threshold was never exceeded can also be useful, with EMA

• In some cases, paleofloods can give the effect of an even longer record
Historic Flood Information

Wheeling (Ohio) Flood (1907)

"...The worst flood since the memorable 1884 flood now holds sway in the Ohio valley. A new high water record has been established in Pittsburg, and though the mark of '84 was not passed at Wheeling the second flood stage to that destructive water will be attained here this morning. .."

--The Intelligencer, March 15, 1907, p. 1
Paleoflood Data

Paleoflood stage
Non-exceedence level (bound)
Threshold level

(Source: Jarrett 1991, modified from Baker 1987)
Reducing Sampling Error

- Larger data set reduces sampling error and provides a better estimate of the frequency relationship
  - narrower confidence interval

- Larger data set spans more time

- It’s more difficult to create a homogeneous data set over a longer span of time

- Some help comes from regionalization...
Regionalization

• Observations from several independent sites with similar properties can help create a larger sample
• “Substitute space for time”
• More straightforward with precip than flow (Atlas 14)
  – for flow, size and shape of basin affect distribution
• Methods for flow frequency:
  – Index Flood Method: sites vary only by a scale factor
  – Regionalize dimensionless parameter like skew
  – Develop relationship between watershed parameters and flood quantiles (OLS/GLS regression)
Bulletin 17B’s skew map
Regionalization

- Cross-correlation can be a problem – don’t gain as much information from extra sites when they are correlated.

- Generalized Least Squares (GLS) regression accounts for correlation, and for varying errors between data points.
  - provides modest improvement in accuracy
  - gives more realistic estimate of the precision of the estimated parameters
  - separates precision in the regional estimates from the sampling error in the station estimates
Conclusion

• This method of estimating annual frequency of flood events is dependent on the stationarity assumption
  – but, can adjust data to remove recognized trends

• Some techniques to improve the frequency estimate are more sensitive to stationarity

• Regionalization can help improve estimate while spanning less time, but subject to correlation issues

• More current methods and potential future methods will be presented this week