Streamflow Record Extension

Nancy A. Barth, PhD
Hydrologist, Dakota Water Science Center
(nabarthur@usgs.gov)

Presentation modified from:
Peter McCarthy and Seth Siefken
Wyoming-Montana Water Science Center
Record Extension

- Bulletin 17C (B17C) Appendix 8 (‘Record Extension with Nearby Sites’) provides
  - discussion
  - guidelines
  - example
B17C Record Extension

- Method:
  - Record augmentation based on cross correlation between flood peaks at a short- and long-term record site
  - Approach obtains unbiased estimates of the mean and variance of the lengthened time series
  - Uses a modified Maintenance of Variance Extension Type 3 (MOVE.3 or MOVE3; Vogel and Stedinger 1985)
Terminology

- Target station = station with the short period of record which is being extended
- Index station = station with the long period of record which is used to augment the target station record
- Concurrent period of record: Overlapping years \((n_1)\) of peak flow records for the target and index stations
Terminology

- Non-concurrent period of record: Period of record (and number of years, \(n_2\)) which the index station is being used to augment the target station.

- Equivalent years of record: The statistical value (equivalence) of the augmented record for use in flood frequency analyses represented in years, \(n_e\). The final analysis will have a record length of \(n_1 + n_e\) years.
Terminology

- Target Station is Suwanee Creek
- Index Station is Etowah River
- Concurrent period: 1985-2004
  - \( n_1 = 20 \) years
- Non-concurrent period: 1892-1984
  - \( n_2 = 93 \) years
Guidelines

- Correlation of concurrent record > 0.80
- Minimum concurrent period of record should be 10 years ($n_1 \geq 10$)
- Basic assumption: Concurrent flow observations at target and index stations have a joint normal probability distribution with skewness of zero (further research is needed as indicated in B17C)
Extension is performed on log-transformed values of streamflow

- $\overline{x}_1$ = index gage sample means for the $n_1$ records of consecutive data
- $\overline{y}_1$ = target gage sample means for the $n_1$ records of consecutive data
- $s_{x_1}$ = index gage sample standard deviation for the $n_1$ records of consecutive data
- $s_{y_1}$ = target gage sample standard deviation for the $n_1$ records of consecutive data
- $\overline{x}_2$ = index gage sample means for the $n_2$ records of extended data
- $s_{x_2}$ = index gage sample standard deviation for the $n_2$ records of consecutive data
Maintenance of Variance Extension (MOVE)

- Simple linear regression between sites produces a biased estimate of the variance
- If the variance of the extended record is important (as in flood frequency analysis) more advance techniques must be used
- MOVE1
  \[ \hat{y}_i = \bar{y}_1 + \frac{s_{y_1}}{s_{x_1}} (x_i - \bar{x}_1) \]
- MOVE2
  \[ \hat{y}_i = \hat{\mu}_y + \frac{\hat{\sigma}_y}{s_x} (x_i - \bar{x}) \]
- MOVE3
  \[ \hat{y}_i = a + b (x_i - \bar{x}_2) \]
Maintenance of Variance Extension (MOVE)

MOVE3 parameters:

\[
a = \frac{(n_1 + n_2)\bar{\mu}_y - n_1 \bar{y}_1}{n_2}
\]

\[
b^2 = \frac{(n_1 + n_2 - 1)s^2_y - (n_1 - 1)s^2_{y_1} - n_1(\bar{y}_1 - \bar{\mu}_y)^2 - n_2(a - \bar{\mu}_y)^2}{(n_2 - 1)s^2_{x_2}}
\]

Where:

\[\bar{\mu}_y\] and \[\bar{\sigma}^2_y\] are unbiased estimators of the mean and variance of the extended data (Matalas and Jacobs 1964):

\[
\bar{\mu}_y = \bar{y}_1 + \frac{n_2}{n_1 + n_2} \left( \hat{\rho} \frac{s_{y_1}}{s_{x_1}} \right) (x_2 - \bar{x}_1)
\]

\[
\bar{\sigma}^2_y = \frac{1}{n_1 + n_2 - 1} \left[ (n_1 - 1)s^2_{y_1} + (n_2 - 1) \left( \hat{\rho} \frac{s_{y_1}}{s_{x_1}} \right)^2 s^2_{x_2} + (n_2 - 1)\alpha^2(1 - \hat{\rho}^2)s^2_{y_1} + \frac{n_1 n_2}{n_1 + n_2} \left( \hat{\rho} \frac{s_{y_1}}{s_{x_1}} \right)^2 (x_2 - \bar{x}_1)^2 \right]
\]

\[
\alpha^2 = \frac{n_2(n_1 - 4)(n_1 - 1)}{(n_2 - 1)(n_1 - 3)(n_1 - 2)}
\]
Uncertainty of Extension

- Because of error in the record extension process, a synthesized peak conveys less information about the target site hydrology than a peak observed at the target site.
- This is accounted for by computing the equivalent years of record ($n_e$) added to the target site by the record extension.
- If correlation $\rho < 1$, then $n_e < n_2$.
- Equivalent years of record in terms of variance can be computed as:

$$n_e = \frac{2}{n_1 - 1 + \frac{n_2}{(n_1 + n_2 - 1)^2(n_1 - 3)}} \left( A\rho^4 + B\rho^2 + C \right) + 1 - n_1$$

A, B, and C are functions of $n_1$ and $n_2$. 
B17C Example

- Target Station is Suwanee Creek
- Index Station is Etowah River
- Concurrent period: 1985-2004
  - \( n_1 = 20 \) years
- Non-concurrent period: 1892-1984
  - \( n_2 = 93 \) years
Evaluate the record

- Does the concurrent period represent a range of peaks at the index station?
B17C Example

- Correlation = 0.85
- MOVE3 fills all years
- Modified MOVE3 only fills $n_e$ years
- Equivalent years of record = 13 (12.54 rounded up)
\[ \hat{y}_i = a + b(x_i - \bar{x}_e) \quad \text{for } i = n_1 + 1, \ldots, n_1 + n_e \]

\[ \bar{x}_e = \frac{1}{n_e} \sum_{i=n_1+1}^{n_1+n_e} x_i \]

\[ s_{xe}^2 = \frac{1}{n_e - 1} \sum_{i=n_1+1}^{n_1+n_e} (x_i - \bar{x}_e)^2 \]

\[ a = \frac{(n_1 + n_e)\hat{\mu}_y - n_1\bar{y}_1}{n_e} \]

\[ b^2 = \frac{(n_1 + n_e - 1)s_y^2 - (n_1 - 1)s_{y1}^2 - n_1(\bar{y}_1 - \hat{\mu}_y)^2 - n_e(a - \hat{\mu}_y)^2}{(n_e - 1)s_{xe}^2} \]

---

**Table 8–2.** MOVE extended record for 13 years (1972 to 1984) for Suwanee Creek at Suwanee, Georgia (station 02334885).

<table>
<thead>
<tr>
<th>Water year</th>
<th>Annual peak streamflow (ft³/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>2,830</td>
</tr>
<tr>
<td>1973</td>
<td>2,010</td>
</tr>
<tr>
<td>1974</td>
<td>2,300</td>
</tr>
<tr>
<td>1975</td>
<td>2,200</td>
</tr>
<tr>
<td>1976</td>
<td>4,310</td>
</tr>
<tr>
<td>1977</td>
<td>4,150</td>
</tr>
<tr>
<td>1978</td>
<td>3,730</td>
</tr>
<tr>
<td>1979</td>
<td>5,180</td>
</tr>
<tr>
<td>1980</td>
<td>3,040</td>
</tr>
<tr>
<td>1981</td>
<td>722</td>
</tr>
<tr>
<td>1982</td>
<td>6,070</td>
</tr>
<tr>
<td>1983</td>
<td>1,460</td>
</tr>
<tr>
<td>1984</td>
<td>2,440</td>
</tr>
</tbody>
</table>
Table 8-2. MOVE extended record for 13 years (1972 to 1984) for Suwanee Creek at Suwanee, Georgia (station 02334885).

<table>
<thead>
<tr>
<th>Water year</th>
<th>Annual peak streamflow (ft³/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>2,830</td>
</tr>
<tr>
<td>1973</td>
<td>2,010</td>
</tr>
<tr>
<td>1974</td>
<td>2,300</td>
</tr>
<tr>
<td>1975</td>
<td>2,200</td>
</tr>
<tr>
<td>1976</td>
<td>4,310</td>
</tr>
<tr>
<td>1977</td>
<td>4,150</td>
</tr>
<tr>
<td>1978</td>
<td>3,730</td>
</tr>
<tr>
<td>1979</td>
<td>5,180</td>
</tr>
<tr>
<td>1980</td>
<td>3,040</td>
</tr>
<tr>
<td>1981</td>
<td>722</td>
</tr>
<tr>
<td>1982</td>
<td>6,070</td>
</tr>
<tr>
<td>1983</td>
<td>1,460</td>
</tr>
<tr>
<td>1984</td>
<td>2,440</td>
</tr>
</tbody>
</table>

Gaged peak discharge
Note: horizontal bars represent perception thresholds
Peak flow analyses

B17C
\( g = -0.455 \) (log space)
\( Q_{1\%AEP} = 5879 \text{ cfs} \)

B17C Modified MOVE3
\( g = 0.205 \) (log space)
\( Q_{1\%AEP} = 7091 \text{ cfs} \)
Additional Considerations

- Do the target and index stations have similar basin characteristics, hydrology and flood mechanisms?
  - Look specifically at the basin sizes, small basins frequently have different flood mechanisms from large basins
  - What are the mechanisms of the peak flows (rainfall type, snowpack, orographic and lake effects)
- Are the target and index stations nested basins?
Example Revisited

- Correlation = 0.71 for 1985-2018
- Why the big change in correlation?
Are the basins hydrologically similar?
Are the basins hydrologically similar?

- Evaluate the timing of peaks

<table>
<thead>
<tr>
<th>Index Gage</th>
<th>Target Gage</th>
<th>Difference in days of peaks</th>
</tr>
</thead>
<tbody>
<tr>
<td>5030</td>
<td>1440</td>
<td>179</td>
</tr>
<tr>
<td>3090</td>
<td>385</td>
<td>13</td>
</tr>
<tr>
<td>12200</td>
<td>2150</td>
<td>95</td>
</tr>
<tr>
<td>9340</td>
<td>948</td>
<td>0</td>
</tr>
<tr>
<td>9080</td>
<td>1220</td>
<td>0</td>
</tr>
<tr>
<td>27100</td>
<td>3760</td>
<td>0</td>
</tr>
<tr>
<td>5940</td>
<td>1320</td>
<td>0</td>
</tr>
<tr>
<td>7660</td>
<td>695</td>
<td>128</td>
</tr>
<tr>
<td>10900</td>
<td>2540</td>
<td>45</td>
</tr>
<tr>
<td>9420</td>
<td>1190</td>
<td>51</td>
</tr>
<tr>
<td>10500</td>
<td>2650</td>
<td>10.46</td>
</tr>
<tr>
<td>19500</td>
<td>4350</td>
<td>12.04</td>
</tr>
<tr>
<td>11300</td>
<td>2360</td>
<td>10.12</td>
</tr>
<tr>
<td>15000</td>
<td>2900</td>
<td>17.32</td>
</tr>
<tr>
<td>5530</td>
<td>815</td>
<td>7.39</td>
</tr>
<tr>
<td>8900</td>
<td>882</td>
<td>7.53</td>
</tr>
<tr>
<td>9270</td>
<td>2090</td>
<td>9.79</td>
</tr>
<tr>
<td>7100</td>
<td>1260</td>
<td>8.47</td>
</tr>
<tr>
<td>13000</td>
<td>2940</td>
<td>10.78</td>
</tr>
<tr>
<td>15300</td>
<td>3270</td>
<td>11.1</td>
</tr>
<tr>
<td>9000</td>
<td>2560</td>
<td>10.35</td>
</tr>
<tr>
<td>4670</td>
<td>1210</td>
<td>8.39</td>
</tr>
<tr>
<td>9930</td>
<td>1530</td>
<td>8.96</td>
</tr>
<tr>
<td>6000</td>
<td>1780</td>
<td>9.35</td>
</tr>
<tr>
<td>16200</td>
<td>7870</td>
<td>14.3</td>
</tr>
<tr>
<td>12100</td>
<td>3390</td>
<td>11.22</td>
</tr>
<tr>
<td>9700</td>
<td>1070</td>
<td>8.04</td>
</tr>
<tr>
<td>5060</td>
<td>708</td>
<td>6.54</td>
</tr>
<tr>
<td>15500</td>
<td>4700</td>
<td>11.75</td>
</tr>
<tr>
<td>14100</td>
<td>2390</td>
<td>9.99</td>
</tr>
<tr>
<td>12700</td>
<td>1010</td>
<td>7.87</td>
</tr>
<tr>
<td>12200</td>
<td>3010</td>
<td>10.56</td>
</tr>
<tr>
<td>3220</td>
<td>2050</td>
<td>9.62</td>
</tr>
<tr>
<td>9940</td>
<td>2000</td>
<td>9.35</td>
</tr>
</tbody>
</table>
B17C
\(g = -0.109\) (log space)
\(Q_{1\%AEP} = 7631\) cfs
Demonstration of Tools

- Excel Spreadsheet by Beth Faber
  - MOVE3 record extension
  - B17C (modified MOVE3) record extension

- Excel Macro by Peter McCarthy
  - MOVE3 record extension
    - Allows multiple index gages
    - Tests index gages for highest correlation
  - B17C (modified MOVE3) record extension
    - Tests index gages for highest correlation
Demonstration of Tools

- R-code/function by Seth Siefken
  - MOVE3 record extension
    - Allows multiple index gages
    - Tests index gages for highest correlation
    - Writes watstore files for PEAKFQ consumption
    - Provides metrics for adjusting confidence intervals when running a peak flow analysis using MOVE3 record extension method (methods described in SIR 2018-5046)
  - Developing B17C (modified MOVE3) method
R-code/function by Seth Siefken

MOV3 Record Extension with PFFREX

Seth Siefken
2021-06-17

MOV3 is a technique for extending the record at a streamgage based on information at a nearby, hydrologically related site. The original technique is described by Vogel and Stedinger (1985) and Bulletin 17C (England and others, 2019) describes a slightly modified approach for use with the Expected Moments Algorithm. PFFREX implements both the original and Bulletin 17C methods. Peak flow input data is provided in the WATSTORE format (downloadable from NWIS). PFFREX performs MOV3 analysis for user-specified input parameters, reads the WATSTORE file and outputs results in the WATSTORE format along with additional diagnostic information.

Full MOV3 Record Extension

The original MOV3 method described by Vogel and Stedinger (1985) uses a modified linear regression technique to extend the record at a streamgage of interest (target site) using information from a nearby, hydrologically related streamgage (index site) with a record of data overlapping that of the target site. The number of years of overlapping data between the two sites is denoted $n_1$, and the number of additional years of data at index site is denoted $n_2$. For the full MOV3 record extension, data from the index site is used to synthesize peak flow values at the target site for all $n_2$ years of additional record, so that the extended record length at the target site is $n_1 + n_2$.

The period of overlap ($n_1$) must be long enough to establish a viable statistical relationship between the sites. Bulletin 17C recommends at least 10 years of overlap. The correlation between the two sites must be sufficiently large to perform a valid analysis. Bulletin 17C recommends a correlation coefficient of at least 0.8. The MOV3 tools in PFFREX screen all possible index sites against the input criteria for overlap to select the best index sites. If possible, the tool will use multiple index sites to synthesize peak flows at the target sites, using the index site with the best correlation for each missing year to reconstruct the peak flow values.

Because there is uncertainty in the relationship between the peaks at the target and index sites that are synthesized at the target site using MOV3 does not contribute the same amount of uncertainty to the overall distribution at the target site as a peak recorded at the target site. As a result, the confidence in the frequency analysis using a full MOV3 record extension will require adjustment to reflect the uncertainty in the synthesized peaks. Sando and McCarthy (2018) describe the necessary adjustments to the confidence interval of the peak flows. The following example code outlines how to perform a full record extension MOV3 analysis:

```r
library(PFFREX)
dir.create("./MOVE3_results") #Create directory for output files
inputWATSTORE <- system.file("extdata", "MOVE3_Bulletin17C_WATSTORE.txt", package="PFFREX") #WATSTORE data file containing peak-flow data for target and index sites
outputFolder <- "/./MOVE3_results" #Folder for MOVE3 results
targetSite <- c("02334885") #Site number of target site for which record will be extended
minOverlap <- 10 #Minimum number of years of overlap required between target and index sites. Bulletin 17C recommends at least 10.
minCorrel <- 0.8 #Minimum Pearson correlation coefficient between target and index sites.
excludeCodes <- c("3", "4", "8", "A") #Codes in WATSTORE file which will cause peaks to be excluded from MOV3 analysis

#Run MOV3 analysis, setting Bulletin17C = TRUE will result in only ne peaks being synthesized
MOVE3FromW6(inputWATSTORE, outputFolder, targetSite, yrmin = minOverlap, rhomin = minCorrel, excludeCodes = excludeCodes, Bulletin17C = TRUE)
```
R-code/function by Seth Siefken

- Show files and results

![File list with details](image-url)
References

Bulletin 17C, Appendix 8

- Modified MOVE3 accounting for equivalent years of record
- Introduces MOVE3

Vogel and Stedinger, “Minimum Variance Streamflow Record Augmentation Procedures” (1985)


- MOVE1 and MOVE2 comparison

Matalas and Jacobs, “A Correlation Procedure for Augmenting Hydrologic Data” (1964)

- Unbiased estimators of mean and standard deviation for extended record
Questions?