Book IV Chapter 2
At-Site Flood Frequency Analysis

George Kuczera

Overview

Evolutionary update of Chapter 10 of ARR(1987):

- Reduced prescription about the choice of flood probability model (GEV, LP3)
- Abandonment of at-site product log-moment fitting of the log-Pearson III (LP3) distribution
- Bayesian fitting methods to make better use of available flood information:
  - gauged and censored flow data plus rating error
  - regional information
  - quantification of uncertainty
- LH moments to assist fitting right tail in difficult cases
- Non-homogeneous probability models to account for long-term climate variability

---

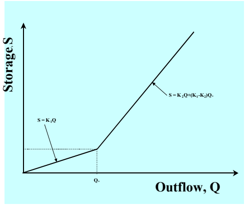
Wider Choice of Flood Probability Model

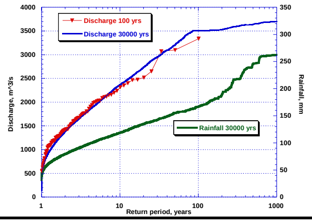
Choice of flood distribution is somewhat arbitrary:

- Both LP3 and Generalized Extreme Value (GEV) distributions fit the bulk of flood data satisfactorily
- GEV has some theoretical justification based on extreme value theory
- Extrapolation well beyond observed flood data not recommended:
  - Even with GEV, big assumption is that observed floods are representative of large flood behaviour
  - Standard probability models are not informed about rainfall-runoff process and stream hydraulics (theory of derived distributions not advanced enough to replace GEV and LP3)
  - ARR will continue to recommend rainfall-runoff/hydraulic based methods in preference to flood frequency methods for ARIs much in excess of observed record

Extrapolation Case study

- A Poisson rectangular pulse rainfall model is used to generate a long record of high resolution rainfall. This is routed through a rainfall-runoff model to generate runoff into the stream system.
- Storage-discharge relationship is bilinear with activation of significant flood terrace storage once a threshold discharge is exceeded.
- Routing model parameters were selected so that major flood terrace storage is activated by floods with an ARI in excess of 100 years.





Why Reject Log Pearson III (LP3) Method of Moments

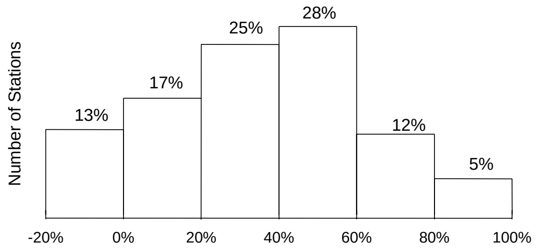
Two major problems with LP3 method-of-moments:

- Relies on log product moments $E[(\log x)^r]$ which are biased and highly variable for small samples. It is noted that USWRC Bulletin 17 recommended using regional skew estimates to combat this.
- Use of log x can overemphasize importance of small x

Monte Carlo simulation illustrates problem with method-of-moments applied to LP3 (Wang,1998):

- For 151 Australian stations assume fitted LP3 is the "true" distribution
- Simulate flood record by sampling from "true" LP3 distribution
- Estimate 50- and 100-year flood by fitting LP3 log-moments and GEV L moments to simulated flood data
- Compute mean-squared-error (MSE) of quantile estimators

Why Reject LP3 Method of Moments



Reduction in MSE of 100-Year event estimate achieved by GEV L moment procedure

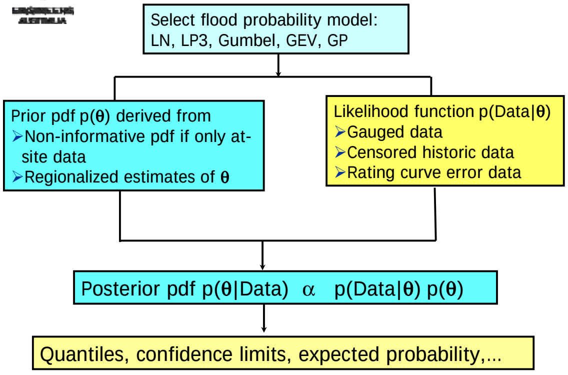
Bayesian Fitting Methods

Bayesian inference provides the complete solution:

- Works for any probability model
- Can efficiently use gauged flow data, censored historic data and information on rating errors
- Censor outliers to improve fit
- Provides optimal quantiles, accurate expected probability estimates and accurate confidence limits
- Efficient combination of at-site and regional information and transfer of data from nearby gauges

But requires Monte Carlo sampling → need computer!

Bayesian Fitting



Select flood probability model:
LN, LP3, Gumbel, GEV, GP

Prior pdf p(**θ**) derived from
- Non-informative pdf if only at-site data
- Regionalized estimates of **θ**

Likelihood function p(Data|**θ**)
- Gauged data
- Censored historic data
- Rating curve error data

Posterior pdf p(**θ**|Data) α p(Data|**θ**) p(**θ**)

Quantiles, confidence limits, expected probability,...

Expected Parameter Quantiles

- 1 in Y AEP flood quantile q_Y $P(q > q_Y \mid \theta, M) = \frac{1}{Y} = \int_{q_Y}^{\infty} p(q \mid \theta, M)\, dq$
- Quadratic loss function $L[\hat{q}(D), q(\theta)] = \alpha[\hat{q}(D) - q(\theta)]^2$ penalizes under and overdesign equally
- Expected value of quantile minimizes quadratic loss

$$E[q_Y \mid D, M] = \int_{\theta} q_Y(\theta, M)\, p(\theta \mid D, M)\, d\theta$$

- Approximate expected value of quantile using expected parameter quantile

$$E[q_Y|D,M] \approx q_Y[E(\theta|D,M)]$$

Expected Probability Quantiles

- Design flood distribution integrates out parameter uncertainty using total probability theorem

$$p(q \mid D,M)=\int_{\theta} p(q \mid \theta,M)\,p(\theta \mid D,M)\,d\theta$$

- Consider the 1 in Y AEP flood quantile q_Y

$$P(q > q_Y \mid D,M)=\int_{\theta}\left(\int_{q_Y}^{\infty} p(q \mid \theta,M)\,dq\right)p(\theta \mid D,M)\,d\theta$$

$$=\int_{\theta} P(q > q_Y \mid \theta,M)\;p(\theta \mid D,M)\,d\theta$$

is the expected probability of Q exceeding q_Y

Fitting LP 3 Probability Model to Gauged Data

Hunter river at Singleton: 31 years of gauged peaks

76.19	171.7	218.0	668.2	1373.	124.0	276.0	894.8
1373.	279.9	202.4	4049.	2321.	2534.	3313.	1231.
1390.	12515	1098.	447.4	478.5	180.3	164.2	229.3
2123.	965.5	2749.	48.98	76.45	911.7	925.9	

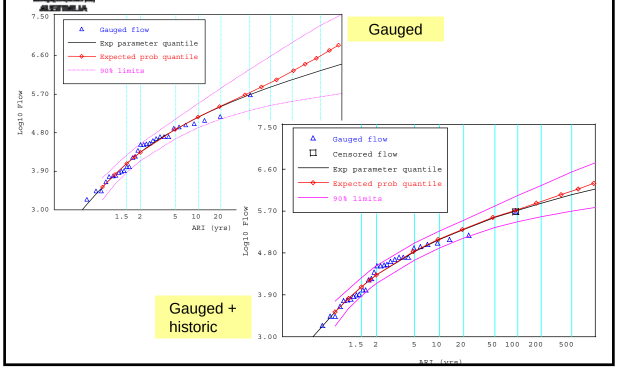
The gauged record spanned 1938 to 1969.

The biggest flood in that record occurred in 1955.

An examination of historic records indicates that during the ungauged period 1820 to 1937 there was only one flood that exceeded the 1955 flood – this flood occurred in 1820.

Over the ungauged period 1820 to 1937 there was one flood above and 117 floods below the threshold discharge corresponding to the 1955 flood.

Bayesian Fitting: Gauged+Historic



Gauged

Gauged + historic

Bayesian Fitting: Gauged+Historic

1 in Y AEP	Quantile q_Y	Quantile confidence limits		Expected AEP for q_Y
		5% limit	95% limit	
10	3888	2222	8287	1/9.9
50	12729	5537	52157	1/43
100	19548	7372	108144	1/74
500	47366	11799	549773	1/210

Gauged

1 in Y AEP	Quantile q_Y	Quantile confidence limits		Expected AEP for q_Y
		5% limit	95% limit	
10	3281	2188	5006	1/9.6
50	9351	5786	16416	1/48
100	13535	7751	27100	1/93
500	28615	12790	87327	1/363

Gauged + historic

Bayesian Fitting: Using Regional Prior Information

Prior information on parameters from regional analysis is particularly important for 3-parameter flood models such as LP3 or GEV

Example: Hunter river at Singleton using LP3

Prior info on log skew: mean 0.0, standard dev 0.35

	No prior information		With prior information	
LP3 parameter	Mean	Std deviation	Mean	Std deviation
m	6.426	0.236	6.423	0.236
$\log_e s$	0.350	0.127	0.322	0.127
g	0.146	0.643	0.028	0.305

	No prior information			With prior information		
1 in Y	Quantile q_Y	Quantile confidence limits		Quantile q_Y	Quantile confidence limits	
AEP		5% limit	95% limit		5% limit	95% limit
10	3,888	2,222	8,287	3,619	2,184	6,933
50	12,729	5,537	52,157	10,677	5,294	29,104
100	19,548	7,372	108,144	15,670	7,050	50,274

LH Moments

LH moments are linear combinations of ranked observations

$L_1(h) = E[X_{h+1:h+1}]$ ……………………………………location

$L_2(h) = 0.5\ E[X_{h+2:h+2} - X_{h+1:h+2}]$………………………..…..scale

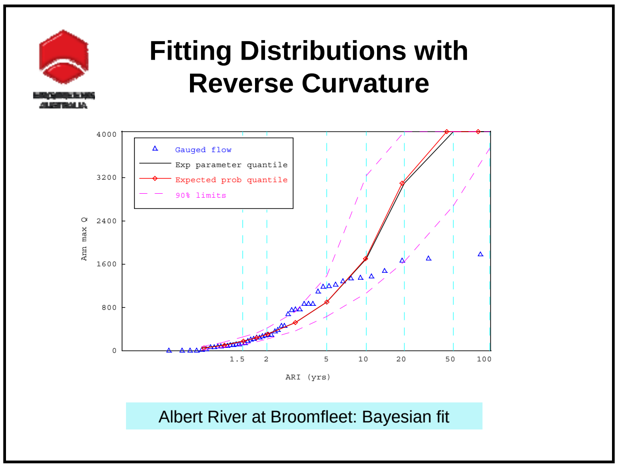
$L_3(h) = 0.333\ E[X_{h+3:h+3} - 2X_{h+2:h+3} + X_{h+1:h+3}]$………..shape

where $X_{i:n}$ is the i^{th} rank (smallest to largest) in n samples

As h is increased LH moments increasingly focus on the right tail of the flood distribution

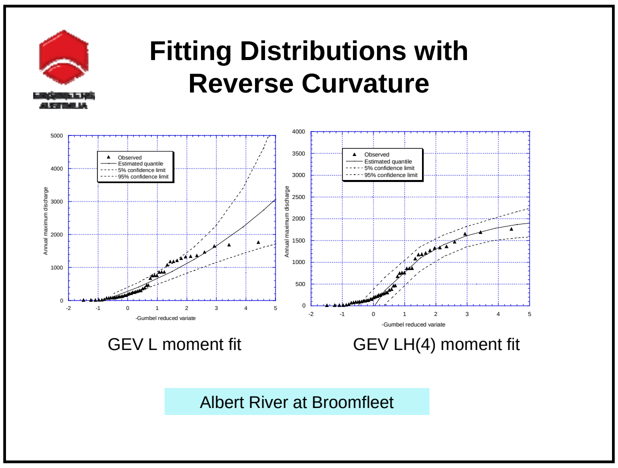
When h=0 get L moments:

- Almost unbiased and normally distributed, which is particularly useful in regionalisation studies
- For at-site estimation more accurate estimators are available (eg for GEV MOM and GMLE are superior)



Fitting Distributions with Reverse Curvature

Albert River at Broomfleet: Bayesian fit

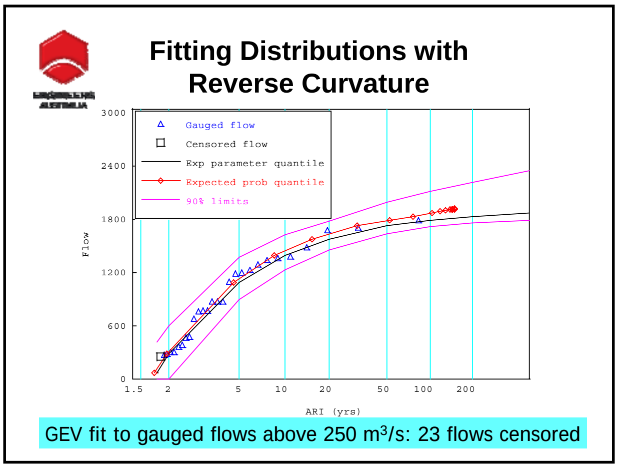


Fitting Distributions with Reverse Curvature

GEV L moment fit

GEV LH(4) moment fit

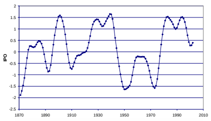
Albert River at Broomfleet



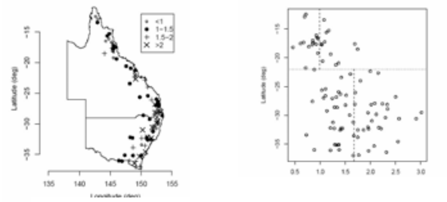
Fitting Distributions with Reverse Curvature

GEV fit to gauged flows above 250 m^3/s: 23 flows censored

Multi-Decadal Climate Variability and Flood Risk

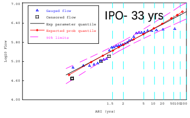


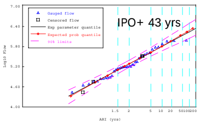
Flood risk appears to vary with Inter Pacific Oscillation (IPO)

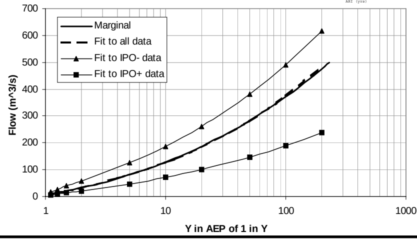


Ratio of 10-year IPO- and IPO+ quantiles (Micevski et al., 2005)

Non-Homogeneous Record: Clarence river







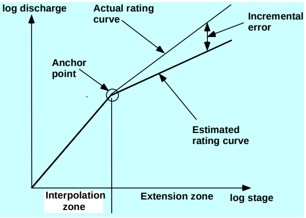
Multi-Decadal Climate Variability and Flood Risk

Implications:

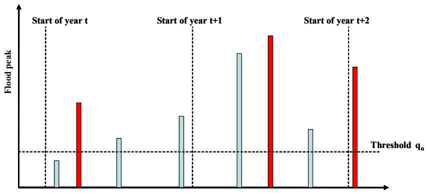
- Eastern NSW and S-E Queensland flood risk affected by IPO variation
- Standard flood frequency analysis:
  - OK for long records
  - Serious bias in flood risk possible if short records sample from one IPO epoch
- If an IPO epoch is undersampled need to extend record or use regional estimator

Rating Curve Error

- Big floods almost always much bigger than largest gauged flow
- Rating curves need to be extrapolated → Can introduce large and correlated rating error with coefficient of variation ~30%
- Not very much is known about these errors → Concern is that we are fitting to systematic rating curve errors



Peak-Over-Threshold Series



Annual maximum series throws away useful information

e.g. in year t+1 the second biggest flood > biggest flood in year t

→ Form POT (peak-over-threshold) series to include all <u>independent</u> flood events above a threshold

What is the distribution of time T between <u>independent</u> flood peaks with peak value exceeding q?

Peak-Over-Threshold Series

Let w be the maximum value in the POT time series of duration T

w = max $\{q_1,\dots,q_n\}$

Can show the expected time between consecutive peaks exceeding w is

$$T_P(w)=\frac{1}{\text{Expected number of peaks}>w\text{ per unit time}}=\frac{1}{\nu P(Q>w)}$$

where ν is average number of events exceeding threshold per unit time and P(Q>w) is probability of a peak exceeding w in any event

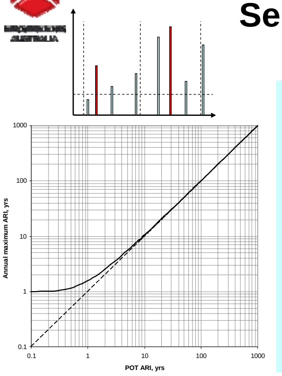
The expected number of years between annual maximum events with magnitude in excess of w is

$$T_A(w)=\frac{1}{1-\exp\left(-\frac{1}{T_P(w)}\right)}$$

Annual maximum ARI

POT average recurrence interval (ARI)

Peak-Over-Threshold Series



(i) T_A > 10 years

Use of annual maximum series is generally preferred because it yields virtually identical answers to POT analysis and annual maximum series are easier to extract and define. This series is generally used in design, as low AEPs in this range are generally required for estimation of a design flood for a structure.

(ii) T_A < 10 years

Use of POT series is generally preferred because all events are of interest in this range, whether they are the highest in the particular year of record or not. The annual maximum series may omit many events of interest.

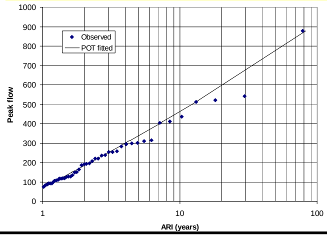
Peak-Over-Threshold Series

A GP distribution with distribution function $P(Q \le q \mid \theta) = 1 - \left(1 - \frac{\kappa(q - q_*)}{\beta}\right)^{\frac{1}{\kappa}}$ was fitted to the 47 peaks in a 47-year POT series for Styx river yielding the following estimates: $\beta = 148.7$, $\kappa = -0.024$ and $q_* = 73.999$.

The average recurrence interval between two peaks exceeding a magnitude of w is

$$T_P(w) = \frac{1}{\nu P(Q > w)} = \frac{1}{\nu\left[1 - \frac{\kappa(w - q_*)}{\beta}\right]^{1/\kappa}}$$

where ν is the average number of flood peaks above the threshold q_* per year and is equal to 1.



Conclusions

- Revision of ARR flood frequency procedures is evolutionary and exploits new research developments
- Significant changes:
  - Choice of distribution eg GEV, LP3
  - Improved accuracy in design flood estimates
  - Efficient use of pre-gauged data
  - Improved fitting capabilities in difficult cases
  - Accurate quantification of uncertainty
  - Efficient use of at-site and regional data
  - Recognition that flood risk varies on multi-decadal scales in some parts of Australia