

# 절단된 홍수 자료에 대한 확률가중적률 추정량

## Expected Probability Weighted Moment Estimator for Censored Flood Data

전종준\*, 김영오\*\*, 김용대\*\*\*, 박준형\*\*\*\*

Jeon Jong-June, Kim, Young-Oh, Kim Yongdai, Park Junehyeong

### 요 지

미래의 연별 최대 강수량 예측의 정확성을 향상시키는데 역사적 자료가 도움이 된다는 많은 연구 결과가 있었다. 관측의 오차와 자료의 손실로 역사자료를 이용한 강수 예측 방법은 절단자료의 분석을 중심으로 연구되었다. 대표적인 역사자료의 이용방법으로 조건부 적률을 이용한 B17B [Interagency Committee in Water Data, 1982], 조건부적률과 적률 관계식을 이용한 Expected Moment Algorithm(EMA) [Cohn et al.:1997], 조건부 확률가중적률을 이용한 Partial Probability Weighted Moment (PPWM)[Wang ; 1991] 방법이 있다.

본 연구에서는 역사적 자료를 반영하는 방법에 있어 B17B와 EMA의 관계를 밝히고 그러한 관계가 PPWM에 동일하게 적용할 수 있음을 보였다. 우리는 B17B와 EMA의 관계를 적률방정식으로 표현하였고 PPWM에서 확률가중 적률 방정식을 정의함으로써 PPWM을 확장하였다. 본 연구에서 제안한 새로운 역사 자료를 이용한 강수예측 방법론을 Expected Probability Weighted Moment (EPWM) 방법이라고 부르고 그 예측 방법의 성능을 다른 예측방법과 시뮬레이션 결과를 통해 비교하였다. 역사 자료 방법론의 비교는 Generalized Extreme Value (GEV) 분포를 이용하여 이루어졌으며, 각 방법론은 GEV분포의 형태모수(shape parameter)따라 다른 특성을 나타낸다는 것을 보였다. 뿐만 아니라 여기서 제안한 EPWM 방법은 대부분의 경우에 좋은 추정량을 준다는 것을 보였다.

**핵심용어 : 절단자료, GEV 분포, probability weighted moment, 역사 자료, 홍수 빈도 분석**

### 1. 서론

홍수량 빈도 분석에 역사적 자료를 이용하여 신뢰도를 높일 수 있다는 연구결과가 오래 전부터 있었다. 최근 계측자료와는 달리 홍수량의 역사적 자료는 일정 수준 이상에서만 기록되거나 혹은 일정 수준을 초과하였다는 정보만을 가지고 있는 경우가 많다[Calenda et al., 2005]. 이런 역사적 홍수 자료의 특성으로 홍수 빈도 예측에 최근의 계측자료와 과거의 절단자료의 정보를 함께 이용하려는 연구가 활발하게 진행되었다[Wang, 1990; Frances et al., 1994; Hosking, 1995; Cohn et al., 1997; Reis, 2005; Boni et al., 2007; Bocchiola and Rosso, 2009]. 특히 홍수량 추정에 모멘트 방법과 모수모형을 결합하여 기존에 제안되었던 B17B의 성능을 개선시킨 EMA[Cohn et al., 1997]는 역사적 자료를 최근 실측자료와 결합하는 새로운 방법을 제시하였다. 본 연구에서는 우선 B17B에서 EMA의 아이디어가 파생된 수리적인 관계를 밝히고 그 관계를 이용하여 확률가중적률방법[Wang ; 1991]과 모수모형을 결합하는 방법을 제안할 것이다. 그리고 제안된 모형과 다른 모형과의 성능을 비교할 것이다.

\* 서울대학교 자연과학대학 통계학과 박사과정 · E-mail: jjjeon@gmail.com  
\*\* 정회원 · 서울대학교 건설환경공학부 부교수 · E-mail: yokim05@snu.ac.kr  
\*\*\* 서울대학교 자연과학대학 통계학과 부교수 · E-mail: ydkim0903@gmail.com  
\*\*\*\* 서울대학교 자연과학대학 건설환경공학부 석사과정 · E-mail: sai0259@snu.ac.kr

## 2. B17B와 EMA

본 연구에서는 자료의 절단 임계치( $h$ )가 알려진 경우를 가정한다. 역사적 자료의 수를  $N_H$ , 계측자료를  $N_S$  하 하고  $N(=N_H+N_S)$ 은 총 자료 수를 나타낸다. 위 첨자로 사용되는 부등호는 임계치보다 낮거나 높은 자료의 수를 나타낸다. 예를 들어  $N_H^>$  는 임계치보다 큰 역사적 자료의 수를 나타낸다.  $X$ 는 홍수량을 나타내는 확률변수고  $x$ 는 실현된 확률변수의 값으로 관측된 홍수량을 의미한다. 첨자의 표기는 자료의 수에 정의한 형식을 따른다.

모멘트는 조건부 모멘트와 확률로서 표시되는데 이는 B17B 추정량을 이해하는 중요한 사실이다.

$$E(X^r) = P(X \leq h)E(X^r|X \leq h) + P(X > h)E(X^r|X > h)$$

B17B는 아래에 정의된 두개의 조건부 확률과 두 개의 조건부 모멘트 추정량을 이용하여 모멘트를 추정하고 그것을 이용하여 사전에 정의된 LP3, GEV, GPD 분포모형의 모수를 찾는 방법이다.

$$\begin{aligned} \hat{P}(X \leq h) &= \frac{N^<}{N} \\ \hat{P}(X > h) &= \frac{N^>}{N} \\ \hat{E}(X^r | X \leq h) &= \frac{\sum (x_S^<)^r}{N_S^<} \\ \hat{E}(X^r | X > h) &= \frac{\sum [(x_H^>)^r + (x_S^>)^r]}{N^>} \end{aligned}$$

임계치 이하의 역사적 자료는 그 사건의 개수만 알 수 있기 때문에 임계치 이하의 조건부 모멘트의 추정에 역사적 자료가 직접적으로 사용되지 않았다. 역사적 자료는 확률추정과 임계치 이상의 모멘트의 추정에 사용된다. 만약 분석할 데이터의 분포를 먼저 가정하고 임계치 이하의 조건부 모멘트를 관측된 실측자료와 분포모형에서 유도되는 조건부 확률의 가중평균을 사용한다고 해보자. 이러한 방법은 가정한 분포모형의 분포모형과 관측된 자료의 정보를 같이 사용한다는 의미로 볼 수 있다. 여기서 사용되는 가중치는 임계치 이하의 역사적 자료 수와 실측지 자료의 수의 비로 정의하면 B17B 추정량으로부터 유도되는 새로운 추정치가 정확히 EMA방법이 제안하는 모형방정식과 일치함을 밝힐 수 있다. 즉, EMA 방법은 사전에 정의된 분포모형이 있을 때, 임계치 이하의 조건부 모멘트의 추정에 분포모형의 성질을 이용하여 정보를 반영하는 방법이다.

$$\hat{E}_\theta X^r I(X \leq h) = L \frac{\sum (X_S^<)^r}{N_S^<} + (1-L)E_\theta X^r I(X \leq h), \quad L = \frac{N_S^<}{N^<}$$

자료의 무한히 많고 추정된 모수가 사실이라면 위의 식은 성립한다. 즉, 자료의 수가 무한히 많을 때 EMA는 모멘트와 주어진 확률 모형이 만족시켜야 하는 조건을 이용하여 추정의 신뢰도를 높이고자 한 방법이다. 다음 장에서는 B17B에서 EMA로 확장의 방법을 이용하여 기존에 제안되었던 PPWM 방법을 확장하겠다. 그리고 여기서 제안된 새로운 방법은 EPWM 방법이라 부르겠다.

## 3. PPWM과 EPWM

PPWM은 조건부 확률가중모멘트를 이용하여 사전에 정해진 분포의 모수를 찾는 방법이다. 확률가중모멘트는 임계점을 기준으로 두 개의 적분으로 표현된다.

$$EXF(X)^r = EXF(X)^r I(X \leq h) + EXF(X)^r I(X > h)$$

Wang[1990]은 위 우변의 두 항에 대한 불편 추정치를 제안하였다.  $EXF(X)^r I(X \leq h)$ 에 대한 불편추정치를 EMA의 확장방법과 동일한 방법으로 수정하면 다음과 같은 추정량이 나온다.

$$\begin{aligned} E_{\theta}XF(X)^r &= \frac{N_H^<}{N^<} E_{\theta}XF(X)^r I(X \leq h) \\ &+ \frac{N_S^<}{N^<} \left[ \frac{1}{N_S} \sum_{i=r+1}^{N_S} \frac{(i-1)(i-2)\cdots(i-r)}{(N_S-1)(N_S-2)\cdots(N_S-r)} x_{(i)}^* \right. \\ &\left. + \frac{1}{N} \sum_{i=r+1}^N \frac{(i-1)(i-2)\cdots(i-r)}{(N-1)(N-2)\cdots(N-r)} x_{(i)}^{**} \right] \end{aligned}$$

단  $x_{(i)}^*$ 는  $x_s$ 의 순서통계량이고  $x_{(i)}^{**}$ 은  $x$ 의 순서통계량이다. 이 추정량을 통해 얻어지는 확률가중모멘트의 추정치로 모수에 대한 추정량을 얻고 다시 위 추정량을 새로 계산한다. 이 같은 작업을 반복하여 최종적으로 수렴한 모수를 확률모형의 추정량으로 사용한다. 우리는 이 알고리즘을 EPWM 방법이라고 부르겠다.

#### 4. GEV분포에서 EPWM의 성능비교

모수가 세 개인 GEV 분포의 형태는 다음과 같다.

$$F(x; \xi, \alpha, \kappa) = \exp \left\{ - \left[ 1 - \kappa \left( \frac{x - \xi}{\alpha} \right) \right]^{1/\kappa} \right\}$$

$\xi$ 는 위치모수,  $\alpha$ 는 크기모수,  $\kappa$ 는 형태모수이다. 1000번의 몬테카를로 시뮬레이션을 통해 나온 분위수의 상대적 위험도(RMSE)와 추정량의 위험도(MSE)를 비교하였다. 계측자료의 수, 역사자료의 길이,  $\kappa$ 의 참값에 따라 시뮬레이션을 수행하였다. 그 중에 형태모수인  $\kappa$ 의 크기에 따른 추정량의 성질이 흥미롭다.

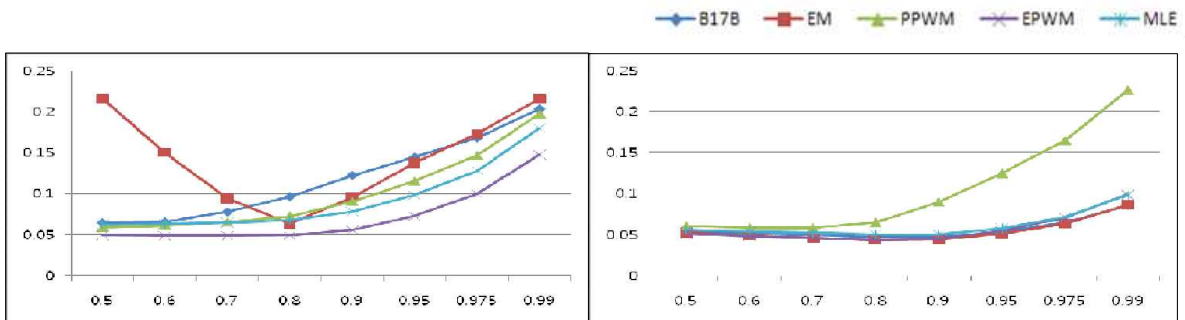


그림 9  $\kappa = -0.3, N_S = 50, N_H = 5N_S$

그림 10  $\kappa = -0.1, N_S = 50, N_H = 5N_S$

Parameters		$\xi$			$\alpha$			$\kappa$			
		$\kappa$	-0.3	-0.2	-0.1	-0.3	-0.2	-0.1	-0.3	-0.2	-0.1
MSE	B17B		77.934	37.299	34.234	186.931	32.541	13.922	0.015	0.007	0.004
	EM		1008.813	143.803	38.652	463.512	94.558	28.316	0.021	0.010	0.006
	PPWM		34.325	37.652	42.536	32.843	26.513	23.001	0.009	0.014	0.026

	EPWM	27.445	28.877	31.509	19.783	16.608	14.722	0.006	0.005	0.005
	MLE	37.920	35.349	34.272	19.851	16.387	14.687	0.006	0.005	0.005
Bias	B17B	-1.700	0.121	0.378	10.916	4.062	0.983	0.112	0.060	0.025
	EM	-19.485	-6.634	-1.775	18.158	7.953	3.037	0.132	0.075	0.035
	PPWM	0.234	0.341	0.531	-0.200	-0.263	-0.368	0.011	0.021	0.039
	EPWM	0.073	-0.044	-0.159	-0.204	-0.440	-0.587	0.002	-0.005	-0.008
	MLE	2.727	2.273	1.845	1.614	1.056	0.627	-0.006	-0.003	0.001
Sd	B17B	8.663	6.106	5.839	8.232	4.006	3.599	0.053	0.058	0.059
	EM	25.083	9.990	5.958	11.568	5.595	4.370	0.058	0.063	0.066
	PPWM	5.854	6.127	6.500	5.727	5.142	4.782	0.094	0.116	0.156
	EPWM	5.238	5.374	5.611	4.443	4.052	3.792	0.080	0.072	0.073
	MLE	5.521	5.494	5.556	4.153	3.908	3.781	0.077	0.073	0.070

표 1 모수추정 결과에 대한 성능 비교

GEV분포에서 모멘트 법은  $\kappa$ 가  $-1/3$  보다 큰 경우에만 추정이 가능하다. 그 이유는 모수 추정에 필요한 3차 모멘트가  $\kappa$ 가  $-1/3$  보다 큰 경우에만 존재하기 때문이다. 반면 확률가중적률은  $\kappa$ 가  $-1$ 보다 큰 경우면 항상 존재하므로 실제 추정할 수 있는 모형의 공간이 모멘트법에 비해 크다. 시뮬레이션 결과에도 알 수 있듯이  $\kappa$ 가  $-1/3$ 과 가까운  $-0.3$ 인 경우에도 모멘트법에 기반한 방법이 모두 좋지 않음을 확인할 수 있다. 그리고 흥미로운 점은 모수 모형을 결합한 EPWM 방법이 Wang[1990]이 제안한 PPWM보다 항상 좋다는 것이다. 모수 모형의 결합이 가중확률모멘트의 추정치의 편이의 변화 없이 분산을 줄여준다는 것에서 그 이유를 찾을 수 있다.

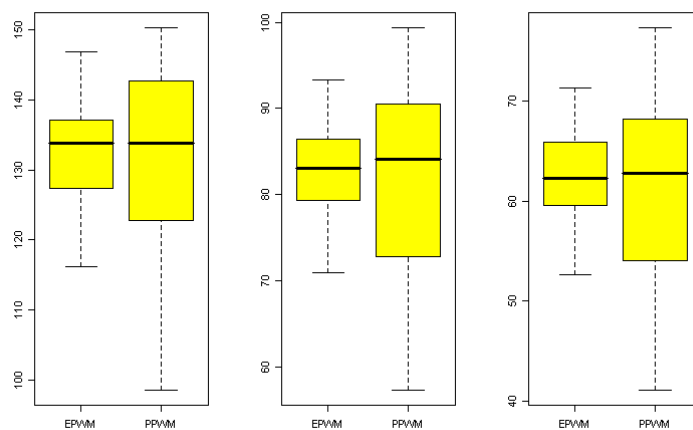


그림 15 The 1<sup>st</sup>PWM, The 2<sup>nd</sup> PWM, The 3<sup>rd</sup> PWM

## 5. 결론

본 연구는 기존에 개발된 B17B와 EMA의 관계를 규명하였고, 그 관계를 통해 확률가중모멘트를 이용한 PPWM 방법을 자연스럽게 확장하였다. GEV 분포 하에서 한편 기존에 알려진 절단자료의 분석방법들과 제안된 EPWM 방법을 비교해본 결과 많은 경우 EPWM 방법이 모수의 추정이나 분위수의 추정에 더 우수함을 시뮬레이션을 통해 살펴보았다. 이는 모멘트 방법에서 나타나

는 GEV 분포에서의 문제점을 EPWM이 어느 정도 해결해 줄 수 있고 실제 홍수데이터 분석에 더 효과적으로 사용될 수 있다는 것을 의미한다.

## 참 고 문 헌

- 1 Bocchiola, D.,R. Rosso (2009), Use of a derived distribution approach for flood prediction in poorly gauged basins: A case study in Italy, *Advances in Water Resources*, 32:8, 1284-1296.
- 2 Boni, G., L. Ferraris, F. Giannoni, G. Roth, R. Rudari (2007), Flood probability analysis for un-gauged watersheds by means of a simple distributed hydrologic model, *Advances in Water Resources*, 30:10, 2135-2144.
- 3 Calenda,G., C.P.Mancini, and E.Volpi, Distribution of the extreme peak floods of the Tiber River from the XV century, *Advances in Water Resources*, 28:6, 615-625.
- 4 Cohn, T. A., W. L. Lane, and W. G. Bairer (1997),An algorithm for computing moments-based flood quantile estimates when historical flood information is available, *Water Resour. Res.*, 33:9, 2089-2096.
- 5 Frances, F., J. Salas, and D. Boes (1994),Flood frequency analysis with systematic and historical or paleoflood data based two-parameter general extreme value models. *Water Resour. Res.*, 30:6, 1653-1664.
- 6 Hosking, J. (1995), In *Recent Advances in Life-Testing and Reliability*, Boca Raton, Florida, CRC Press.
- 7 Interagency Committee in Water Data (1982), Guidelines for determining flood flow frequency, Bull. 17b, Interagency Comm. on Water Data, Washington, D.C.
- 8 Reis, D. S. Jr., and J. R. Stedinger (2005), Bayesian MCMC flood frequency analysis with historical information, *J. Hydro.*, 313, 97-116.
- 9 Wang, Q. J. (1990), Unbiased estimation probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the gev distribution, *J. Hydro.*, 120, 115-124.