

확률강우량 추정을 위한 확률분포함수의 매개변수 추정법에 대한 신뢰성 평가

Reliability Evaluation of Parameter Estimation Methods of Probability Density Function for Estimating Probability Rainfalls

한정우* · 권현한** · 김태웅***

Han, Jeong Woo · Kwon, Hyun-Han · Kim, Tae-Woong

Abstract

Extreme hydrologic events cause serious disaster, such as flood and drought. Many researchers have an effort to estimate design rainfalls or discharges. This study evaluated parameter estimation methods to estimate probability rainfalls with low uncertainty which will be used in design rainfalls. This study collected rainfall data from Incheon, Gangneung, Gwangju, Busan, and Chupungryong gage station, and generated synthetic rainfall data using ARMA model. This study employed the maximum likelihood method and the Bayesian inference method for estimating parameters of the Gumbel and GEV distribution. Using a bootstrap resampling method, this study estimated the confidence intervals of estimated probability rainfalls. Based on the comparison of the confidence intervals, this study recommended a proper parameter estimation method for estimating probability rainfalls which have a low uncertainty.

Key words : probability rainfall, uncertainty, parameter estimation

요 지

최근의 극한 수문사상은 홍수, 가뭄과 같은 심각한 재해를 발생시킨다. 많은 연구자들은 불확실한 미래의 확률강우량 및 유출량의 예측을 위해 많은 노력을 하고 있다. 본 연구에서는 불확실성이 낮은 확률강우량의 산정을 위하여 매개변수 추정법을 평가하였다. 인천, 강릉, 광주, 부산, 추풍령 관측소를 연구 대상 관측소로 선정하여 자료를 수집하였고, ARMA모형을 이용하여 합성강우자료를 구축하였다. 본 연구에서는 극치강우사상에 적합한 것으로 알려진 Gumbel 분포와 GEV 분포모형에 대한 매개변수를 최우도법과 베이지안 추론방법을 사용하여 추정하였으며, Bootstrap 방법을 이용하여 확률강우량의 신뢰구간 길이를 추정하였다. 매개변수 추정 방법별 산정된 확률강우량의 신뢰구간 길이를 비교함으로써 불확실성이 낮은 확률강우량을 산정할 수 있는 매개변수 추정방법을 선정하였다.

핵심용어 : 확률강우량, 불확실성, 매개변수추정

1. 서 론

수공구조물의 설계는 관측강우자료의 관측기간보다 상당히 큰 미래의 재현기간에 해당하는 확률강우량의 산정과 그에 따른 확률홍수량을 산정하여 구조물의 적정규모를 결정하게 된다. 설계홍수량 산정을 위한 강우 분석시에는 극치계열의 연최대치 자료를 주로 사용한다(정중호와 윤용남, 2007). 하지만 이러한 수문 분석과정에는 근본적으로 불확실성이 개입하게 된다.

수문자료의 불확실성은 측정 또는 수집된 자료의 수가 총

분치 못하여 본래의 수문량의 성향을 적절히 반영하지 못하는 점, 수문량의 오차, 수문량의 시간·공간적인 변화의 미반영에 의해 발생한다. 확률분포형을 적용하는 경우에도 선택된 확률분포형이 강우의 특성이나 유출량 등의 수문특성을 완벽하게 재현할 수는 없으므로 확률분포형의 선정과 선택된 확률분포형으로부터 산정된 결과의 불확실성을 내포하게 된다. 또한 확률분포형의 매개변수 산정시 매개변수 추정방법의 차이에 따라 발생하는 확률강우량의 불확실성을 포함하게 된다(홍창선 등, 2001). 하지만 수자원 장기 종합계획, 댐 건설 장기계획 등의 국내 주요 중장기 종합계획은 빈도해석의 확

*한양대학교 대학원 토목공학과 석사과정(E-mail: 00coolguy@hanmail.net)

**전북대학교 토목공학과 조교수

***정회원 · 한양대학교 건설환경시스템공학전공 조교수(교신저자)

정적인(deterministic) 값만을 이용하여 수립되고 있으며, 빈도 분석 결과의 불확실성을 반영한 확률적인(probabilistic) 값이 이용되는 계획은 찾아보기 힘들다(김상욱과 이길성, 2008). 이는 불확실성에 대한 인식 부족과 함께 불확실성에 대한 계산방법이 현실을 제대로 반영하지 못함으로써 빈도분석 결과의 정확성에 대한 신뢰도가 낮은 것에 기인한다. 불확실성을 나타내기 위해서 일종의 근사식을 사용하여 대략적인 신뢰구간을 산정하고 이로부터 분석결과의 불확실성을 표현하고자 한 연구가 진행된 바 있다(Stedinger, 1983; Chowdhury and Stedinger, 1991; Cohn et al., 2001). 그러나 근사식을 사용한 신뢰구간 산정방법은 확률분포함수의 모수를 산정함에 있어서 정규성(normality), 선형성(linearity) 등의 가정이 필요하므로, 불확실성을 산정함에 있어서 비현실적인 값을 산정하거나 과대 추정되는 경우가 있는 것으로 알려져 있다(Reis Jr. and Stedinger, 2005).

근사식을 이용한 매개변수의 불확실성 추정방법을 대신하여 Bayesian 추론방법을 사용한 매개변수 및 불확실성의 추정이 수행될 수 있으며, Bayesian 추론방법은 근사식을 사용하기 위한 가정조건이 필요하지 않으므로 불확실성을 표현하는데 있어 다른 방법보다 우월하다(김상욱과 이길성, 2008). 하지만 Bayesian 추론방법을 사용하여 강우 빈도분석을 실시한 연구사례가 상대적으로 적다. 또한 수자원 분야에서의 Bootstrap 방법을 기초로 해서 연속한 곡선 자료에 대한 신뢰구간 추정기법을 Surtherland et al.(1989)은 제시하였다. 수자원 분야에서 Bootstrap 방법은 기존 난수 발생기법을 이용한 모의기법의 단점을 보완할 목적으로 추계학적 모의 발생에 주로 이용되고 있다(Tasker, 1987). 기존의 원자료를 난수 발생시켜 자료 집단을 생성하는 방법은 자료의 수의 증대는 가져올 수 있으나, 이는 원자료라는 모집단에 국한되어 자료를 발생시켰다. 따라서 단순히 Bootstrap 방법을 이용하여 자료를 발생시키는 것은 자료의 계열성을 반영한 자료의 증가라고 볼 수 없고 자료의 불확실성을 최소화 해주는 방법이라 할 수 없다.

따라서 본 연구에서는 원자료의 계열 특성을 반영하여 장기간으로 발생시킨 시계열 자료를 Bootstrap 하여 자료수의 부족에서 발생하는 불확실성을 최소화 하고자 하였다. 또한 예측된 확률강우량의 변동성에 대한 불확실성을 반영한 확률강우량을 산정하는 것에 중점을 두어 국내 확률강우량 산정에 주로 사용되는 확률분포형과 매개변수 추정방법의 정확성을 판단하는 것에 연구의 목적을 두었다. 본 연구를 수행하

기 위하여 국내 강우의 지역적 특성을 대표할 수 있고 강우 관측 기록 연한이 30년 이상으로 강우 빈도해석시, 자료수의 부족으로 문제가 없을 것으로 판단된 인천, 강릉, 광주, 추풍령, 부산지역의 강우관측소의 지속기간 24시간 연최대치 강우량 자료를 이용하여 빈도해석을 실시하였으며 신뢰구간을 산정하여 비교분석하였다.

2. 강우자료

본 연구에서는 우리나라의 지역적인 분포를 최대한 고려하기 위하여 1961년부터 2006년까지의 46개년 시 강우자료(hourly rainfall data)를 서울·경기, 충청북도, 전라남도, 강원도, 경상도의 30년 이상의 관측기간을 가진 대표 강우 관측소(그림 1)를 선정한 후, 1961-2006년 기간의 강우자료를 수집하여 지속기간별 연 최대치 자료를 추출하였다.

기상청에서 제공하는 1971-2000년까지 30년의 연평균 강우량에 따르면 인천지점의 연평균 강우량은 1152.3 mm, 강릉지

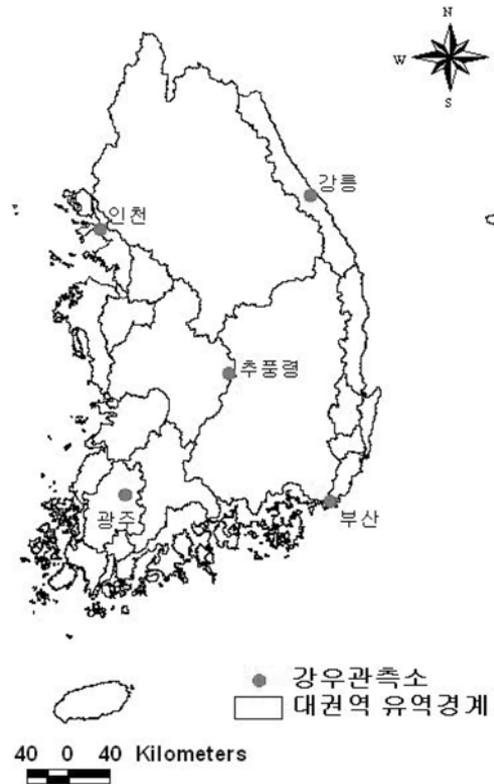


그림 1. 연구대상 강우관측소 위치.

표 1. 대상 강우지점의 지속기간 24시간 연 최대치 강우자료 통계치

| 지점명 | 통계치 | | | |
|-----|--------|----------|-------|-------|
| | 평균(mm) | 표준편차(mm) | 변동계수 | 왜곡도계수 |
| 강릉 | 184.8 | 126.4 | 0.684 | 3.737 |
| 광주 | 147.1 | 61.2 | 0.417 | 1.831 |
| 부산 | 171.4 | 73.3 | 0.428 | 1.490 |
| 인천 | 154.9 | 77.3 | 0.499 | 1.244 |
| 추풍령 | 128.8 | 49.1 | 0.381 | 1.215 |

점은 1401.9 mm, 광주지점은 1367.8 mm, 부산지점은 1491.5 mm 그리고 추풍령지점은 1160.1 mm로 부산지점이 가장 강우량이 많으며, 인천지점이 가장 강우량이 적다. 또한 이상치 검정결과 강릉지점의 강우 자료에 고이상치가 포함되어져 있다. 표 1은 각 지점의 지속기간 24시간 강우자료의 평균(mean), 표준편차(standard deviation), 변동계수(coefficient of variance), 왜곡도계수(coefficient of skewness)를 산정한 결과이다. 강릉지점의 경우 표준편차, 변동계수, 왜곡도계수가 다른 지점보다 확연히 크게 나타나고 있다. 이는 강릉관측소가 가지고 있는 고이상치의 영향을 받은 결과로 판단되며, 고이상치는 2002년 시간당 80 mm와 일최대 870.5 mm의 강우량을 기록하며 강릉지역을 강타한 태풍 루사의 영향을 받은 결과이다. 본 연구에서는 각 지점의 지속기간별 연 최대치 강우자료 중 대표적으로 지속기간 24시간의 강우자료를 이용하여 빈도해석 및 불확실성 분석을 실시하였다.

3. 재현기간별 확률강우량 산정 및 신뢰구간 추정

본 연구에서는 Gumbel 분포와 GEV 분포에 대하여 최우도법과 베이지안 추론법을 사용하여 매개변수를 추정하여 재현기간에 대한 확률강우량을 추정하였다. 또한 ARMA 모형에 의해 장기간의 합성강우량을 발생한 후 발생한 합성강우자료를 무작위 반복 추출하여 자료집단을 만드는 Bootstrap 방법을 이용하여 신뢰구간을 추정하였다. Bootstrap 방법을 통해 발생한 자료에 의해 산정된 재현기간별 확률강우량은 각 재현기간 마다 극한중심정리를 따라 정규분포로 근사된다. 따라서 각 재현기간에 해당하는 95% 신뢰구간의 상한값과 하한값을 산정하였고 신뢰구간의 길이를 산정하여 매개변수별 확률분포형별로 비교하여 불확실성 분석을 실시하였다. 연구의 수행 절차를 그림 2에 나타내었다.

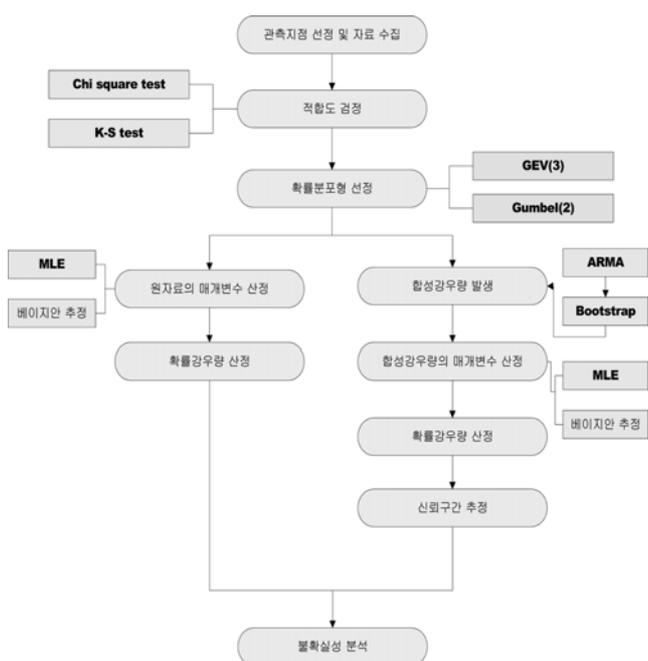


그림 2. 확률강우량 및 불확실성 분석 절차.

3.1 매개변수 추정

매개변수 추정방법은 모멘트법, 최우도법, 확률가중모멘트법이 널리 사용되고 있다. 모멘트법은 이상치가 있는 경우 적용성이 떨어지고 변동성이 큰 단점이 있다. 최우도법과 확률가중모멘트법은 모멘트법에 비해 정확한 추정결과를 내는 방법으로 채택되고 있다. 하지만 확률가중모멘트법의 경우 자료를 오름치순 정렬 후 큰 자료로 갈수록 큰 가중치를 주기 때문에 재현기간이 큰 경우에 모멘트법이나 최우도법보다 지나치게 높은 확률강우량을 산정할 수 있다(정중호와 윤용남, 2007). 따라서 본 연구에서는 최우도법을 적용하였다. 또한 고전적 통계방법의 매개변수 추정방법의 하나인 최우도법과 빈도해석의 결과를 비교하기 위해 분포형별 매개변수의 사후분포를 형성시켜 사후분포의 최빈치에 해당하는 값을 최우추정치로 산정하는 베이지안 추정방법을 사용하였다.

3.1.1 최우도법(Method of Maximum Likelihood)

최우도법은 표본자료가 나올 확률이 최대가 되도록 확률분포모형의 매개변수를 추정하는 방법이다. 주어진 표본자료에 대한 매개변수의 최우추정치(Maximum Likelihood Estimator, MLE)는 표본의 확률 혹은 확률밀도를 최대화하는 매개변수이다. 만약 표본자료가 모집단 $f(x; \theta)$ 에서 취한 것이라면 표본자료의 발생확률은 식 (1)과 같은 우도함수(likelihood function)로 표시할 수 있다.

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta) \quad (1)$$

여기서, $L(\theta)$ 은 매개변수가 θ 인 우도함수이다. $f(x_n; \theta)$ 는 n 번째 표본자료를 표시하는 확률밀도함수이다. 본 연구에서 사용되었던 GEV 분포와 Gumbel 분포의 우도함수는 각각 식 (2)와 (3)과 같다.

$$L(\alpha, \beta, x_0) = \prod_{i=1}^n \frac{1}{\alpha} \left[1 - \beta \left(\frac{x_i - x_0}{\alpha} \right) \right]^{\beta-1} \exp \left[- \left[1 - \beta \left(\frac{x_i - x_0}{\alpha} \right) \right]^{\beta} \right] \quad (2)$$

$$L(\alpha, x_0) = \prod_{i=1}^n \frac{1}{\alpha} \left[- \frac{x_i - x_0}{\alpha} - \exp \left(- \frac{x_i - x_0}{\alpha} \right) \right] \quad (3)$$

여기서, α 는 축척매개변수, β 는 형상매개변수, x_0 는 위치매개변수, x_i 는 관측수문자료이다.

$L(\theta)$ 은 결합확률(joint probability)을 표시하며 이를 최대화하는 매개변수를 추정하는 것이 최우도법이다. 일반적으로 최우도법을 사용하며 가장 효율적인 매개변수 추정치를 얻을 수 있으며, 표본자료의 크기가 충분할 때 매개변수 추정을 위한 다른 방법의 매개변수 추정의 효율성을 비교 평가하는 기준으로 사용되기도 한다. 본 연구에서는 우도함수(식(2)와 (3))를 대수우도함수로 변환한 후 편미분 연립방정식을 이용하여 매개변수를 추정하였다. 표 2에 최우도법에 의해 추정된 매개변수(MLE)를 나타내었다.

표 2. 최우도법과 베이지안 추론법의 매개변수 추정치

| 지점명 | 분포모형 | 매개변수 추정치 | | | | 신뢰구간 | |
|-----|-----------|----------|-------|----------|-------|------------------------|------------------------|
| | | MLE | | BAYS | | MLE | BAYS |
| 강릉 | Gumbel 분포 | α | 64.6 | α | 65.3 | $46.7 < \alpha < 82.2$ | $53.9 < \alpha < 78.4$ |
| | | x_0 | 141.8 | x_0 | 141.5 | $122.3 < x_0 < 161.4$ | $126.7 < x_0 < 155.5$ |
| | GEV 분포 | α | 50.3 | α | 45.3 | $37.1 < \alpha < 68.0$ | $42.6 < \alpha < 47.9$ |
| | | x_0 | 129.8 | x_0 | 129.6 | $112.6 < x_0 < 146.9$ | $124.2 < x_0 < 135.6$ |
| | | β | 0.4 | β | 0.3 | $0.1 < \beta < 0.7$ | $0.1 < \beta < 0.4$ |
| 광주 | Gumbel 분포 | α | 40.2 | α | 41.2 | $31.8 < \alpha < 48.5$ | $32.9 < \alpha < 51.7$ |
| | | x_0 | 122.0 | x_0 | 121.7 | $109.2 < x_0 < 134.9$ | $111.4 < x_0 < 133.2$ |
| | GEV 분포 | α | 37.7 | α | 35.0 | $29.4 < \alpha < 48.4$ | $36.2 < \alpha < 40.0$ |
| | | x_0 | 118.9 | x_0 | 117.1 | $106.7 < x_0 < 131.3$ | $112.9 < x_0 < 124.4$ |
| | | β | 0.1 | β | 0.2 | $-0.1 < \beta < 0.4$ | $0.0 < \beta < 0.3$ |
| 부산 | Gumbel 분포 | α | 53.5 | α | 52.9 | $39.8 < \alpha < 64.9$ | $43.7 < \alpha < 65.9$ |
| | | x_0 | 139.9 | x_0 | 144.8 | $123.4 < x_0 < 156.4$ | $126.4 < x_0 < 152.3$ |
| | GEV 분포 | α | 50.2 | α | 50.1 | $39.0 < \alpha < 64.7$ | $48.2 < \alpha < 52.0$ |
| | | x_0 | 137.1 | x_0 | 137.4 | $120.5 < x_0 < 153.8$ | $132.1 < x_0 < 143.0$ |
| | | β | 0.1 | β | 0.1 | $-0.1 < \beta < 0.3$ | $0.0 < \beta < 0.2$ |
| 인천 | Gumbel 분포 | α | 51.5 | α | 51.6 | $39.7 < \alpha < 63.2$ | $42.3 < \alpha < 65.1$ |
| | | x_0 | 122.0 | x_0 | 121.8 | $103.2 < x_0 < 140.8$ | $108.3 < x_0 < 134.3$ |
| | GEV 분포 | α | 44.7 | α | 44.4 | $33.9 < \alpha < 58.9$ | $42.5 < \alpha < 46.5$ |
| | | x_0 | 115.2 | x_0 | 114.0 | $100.4 < x_0 < 130.1$ | $107.9 < x_0 < 120.5$ |
| | | β | 0.3 | β | 0.3 | $0.0 < \beta < 0.5$ | $0.2 < \beta < 0.5$ |
| 추풍령 | Gumbel 분포 | α | 35.7 | α | 33.9 | $25.9 < \alpha < 43.1$ | $28.0 < \alpha < 45.1$ |
| | | x_0 | 107.5 | x_0 | 106.7 | $96.1 < x_0 < 119.0$ | $96.7 < x_0 < 118.8$ |
| | GEV 분포 | α | 31.7 | α | 31.1 | $24.1 < \alpha < 41.7$ | $29.3 < \alpha < 33.1$ |
| | | x_0 | 104.4 | x_0 | 108.1 | $93.5 < x_0 < 115.2$ | $99.3 < x_0 < 110.0$ |
| | | β | 0.2 | β | 0.1 | $-0.1 < \beta < 0.5$ | $0.0 < \beta < 0.4$ |

3.1.2 베이지안 추론방법(Bayesian Inference Method)

베이지안 추론방법은 우도원리(likelihood principle)를 사용하여 다른 표집 계획 하에서도 동일한 결론을 도출할 수 있으며, Markov Chain Monte Carlo(MCMC) 방법을 통해 추론하게 되므로 자료의 개수가 적을 경우에도 정확한 결과를 얻을 수 있다는 장점을 가지고 있다(김달호, 2005). 베이지안 추론방법은 기본적으로 두 가지의 항을 가지고 있다. 하나는 $X|\theta \sim f(x|\theta)$ 로 표현되는 우도함수항이며, 다른 하나는 $\theta \sim \pi(\theta)$ 로 표현되는 사전분포항이다. 우도함수항과 사전분포항은 식 (4)와 같은 베이즈 정리(Bayes' Theorem)를 이용하여 추정할 수 있다.

$$p(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} \tag{4}$$

여기서, x 는 관측자료, θ 는 확률분포모형의 매개변수이며, $m(x)$ 는 관측자료 x 의 주변분포이다.

$m(x)$ 는 분포의 우도함수와 사전분포의 곱을 표준화시켜 사후분포의 면적이 1 이하가 되도록 만드는 항($m(x) = \int f(x|\theta)\pi(\theta)d\theta$)이다. 하지만 확률분포모형에 따라 적분항을 수치적으로 계산하는 것은 어렵거나 불가능한 경우가 있다. 이를 수치적인

해법이 아닌 근사방법을 통해 사용하는 방법이 Bayesian MCMC 방법이다. 전통적인 베이지안 모수추정을 위해서는 모수의 주변사후분포(marginal posterior distribution)를 구해야 하나 흔히 모수들의 결합사후분포로부터 모수의 주변사후분포를 구하는 단계에서 적분의 어려움이 있다. 최근 Mote Carlo computing 방법의 개발로 적분항의 풀이를 해결하여 베이지안 추론방법을 발전시켰다(김달호, 2005; 김상욱과 이길성, 2008).

MCMC 방법은 전체조건부 사후분포($m(x)$)로부터 매개변수의 값을 추출하여 적분계산 없이 사후분포($p(\theta|x)$)를 추정해내는 방법이다. MCMC 방법 중 가장 잘 알려진 추출방법은 Gibbs sampler와 Metropolis-Hastings algorithm이다(Reis Jr. and Stedinger, 2005).

본 연구에서는 균일분포를 사전분포로 선정하였고, 마코프 연쇄를 구성하기 위한 방법으로 Gibbs sampler를 사용하였으며, 마코프 연쇄를 구성하며 추출되는 매개변수의 채택율을 높이고 계산의 효율성을 극대화하기 위하여 Metropolis-Hastings algorithm을 이용하여 베이지안 추론을 전개하였다. Gibbs sampler는 다차원의 결합 확률분포가 복잡하여 직접 랜덤포본을 생성하기 어려운 경우 각 변수의 조건부 확률로

부터 랜덤 표본을 반복적으로 생성하면 적절한 조건 하에서 이들의 극한분포가 결합밀도함수가 된다는 사실에 근거하여 난수를 생성하는 방법이다. Gibbs sampling에서 각 변수의 조건부확률로부터 반복적으로 생성된 난수들이 결합확률밀도 함수를 극한 분포로 갖는 Vector Markov Chain을 구성한다. 또한 각 변수의 조건부확률로부터 생성된 랜덤포본의 극한분포는 각 변수의 주변확률분포를 산출한다. Metropolis-Hastings algorithm은 제안분포와 전체조건부 사후분포로부터 매개변수를 표본 추출할 시 전이 확률을 도출하며, Gibbs sampling 시의 전시간상태의 매개변수를 추출해 주는 역할을 한다. MCMC 방법에 의해 각 매개변수별 추정치가 추출된 후 추출된 추정치를 이용하여 평균값을 산정하여 각 매개변수의 최우추정치를 산정하였다. 표 2에 베이지안 추정방법에 의해 추정된 매개변수를 나타내었다.

3.2 적합도 검정

본 연구에서는 실용적이고 신뢰성이 높은 검정 결과를 산출하는 검정과 K-S 검정 방법을 이용하여 각 지점별로 관측자료가 Gumbel 분포형과 GEV 분포형을 따른다는 귀무가설(null hypothesis)을 유의수준 5%에서 적합도 검정을 하였다. 5개 지점 모두 검정값이 한계값보다 작게 산출되어 적합도 검정을 통과하였다.

3.3 추계학적 방법에 의한 합성강우량 발생

본 연구에서는 ARMA 모형과 Bootstrap 방법을 이용하여 자료를 모의 발생시켜 확률강우량의 신뢰구간을 산정하였다. 추계학적 자료발생 방법은 관측자료의 통계학적 특성을 유지하면서 보다 장기간 동안의 자료를 인위적으로 발생시키는 방법이다. 본 연구에서는 자기상관함수(Auto Correlation Function, ACF)와 부분자기상관함수(Partial Auto Correlation Function, PACF)을 이용하여 식(5)의 ARMA모형의 차수를 결정하였다. 본 연구에서는 1000개의 합성강우량을 발생시켰으며, 발생된 초기 값은 원자료의 평균을 사용하기 때문에 이 초기값은 연속적으로 발생하는 강우량의 값에 큰 영향을 미치지(윤용남, 2004), 초기 50개의 강우량은 계열로부터

제거하고 나머지 총 950개의 합성강우자료를 이용하였다.

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + e_t - \alpha_1 e_{t-1} - \alpha_2 e_{t-2} \dots - \alpha_q e_{t-q} \quad (5)$$

여기서, $\beta_p y_{t-p}$ 은 확정론적 성분(deterministic component)이며, $\alpha_q e_{t-q}$ 은 추계학적 성분(stochastic component)을 나타내며 p 와 q 는 지체(lag) 차수를 나타낸다.

전 지점에 대해 자기상관도(Auto Correlation Correlogram, ACC)와 부분자기상관도(Partial Auto Correlation Correlogram, PACC)을 이용하여 ARMA 모형의 차수를 결정하였다. 대표적으로 그림 3(a)와 (b)는 각각 강릉지점의 ACC와 PACC의 결과를 나타내며 그림 3(c)는 강릉지점의 합성 강우량을 나타낸다. 전 지점은 ACC로 부터 MA 0차, PACC로 부터 AR 0차임을 확인하였다. 따라서 ARMA(0,0)을 사용하는 것이 적합하다고 판단되며 백색잡음(white noise)을 발생시켜 확정론적(deterministic)인 성분인 원강우량의 평균에 합하여 합성강우량을 발생시켰다. 합성강우량을 모집단으로 하여 Bootstrap 방법을 통하여 1000개 표본자료를 추출하여 최종적으로 950행 1000열의 자료 행렬을 발생시켰으며 최종 발생된 자료는 확률강우량의 신뢰구간 추정에 사용되었다.

3.4 확률강우량 및 신뢰구간 추정

본 연구에서는 재현기간 10, 20, 50, 80, 100, 200, 500년 등 대부분의 수공구조물 설계에 필요한 확률강우량을 산정하였다. 500년은 일반적인 국내 수공구조물 설계 기준에 벗어나, 최근에 거론되고 있는 이상 강우의 규모를 판단하는데 참고할 수 있도록 하였다(건설교통부, 2000). 표 3에 원자료(raw data)를 이용하여 산정한 각 지점별 재현기간-확률강우량을 나타내었다. 표 3의 결과로부터 각 지점별로 동일 추정법을 기준으로 비교시 GEV 분포가 Gumbel 분포보다 높게 확률강우량을 산정하는 경향을 알 수 있다.

확률강우량의 신뢰구간 추정을 위해 재현기간별 확률강우량의 결과값이 정규분포를 생성할 수 있도록 Bootstrap 방법을 이용하였다. 원자료를 이용하여 추정된 확률강우량과 발생

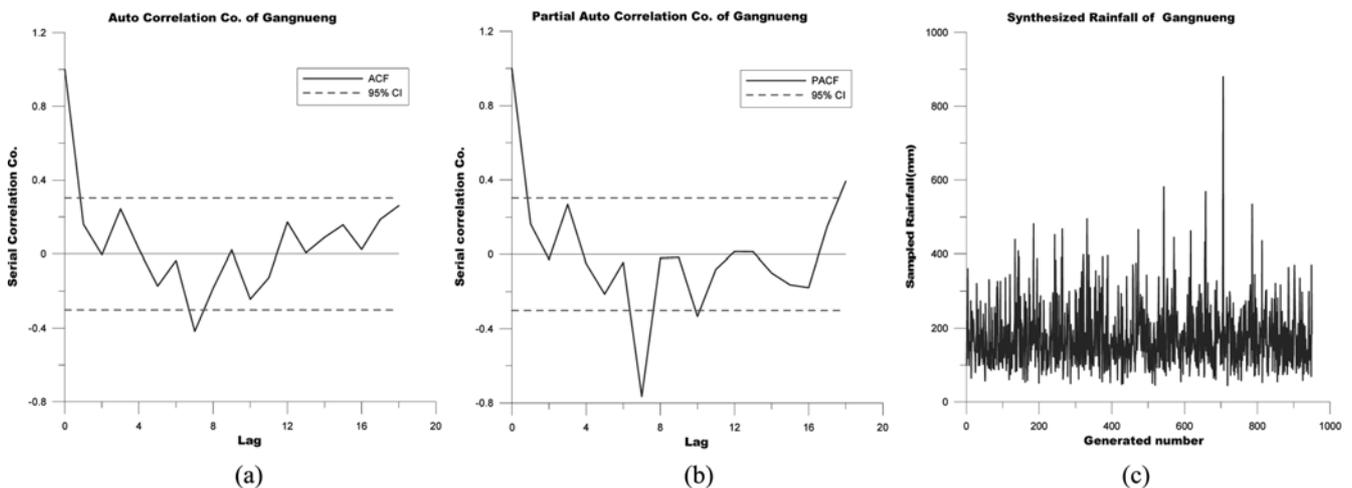


그림 3. 강릉지점의 ACC, PACC 및 합성강우량.

표 3. 원자료의 재현기간별 확률강우량(mm)

| 지점명 | 재현기간 (year) | Gumbel 분포 | | GEV 분포 | |
|-----|----------------|-----------|---------|--------|---------|
| | | 최우도법 | 베이지안추정법 | 최우도법 | 베이지안추정법 |
| 강릉 | 10 | 287.1 | 288.5 | 307.7 | 280.9 |
| | 20 | 333.5 | 335.5 | 404.6 | 358.4 |
| | 50 | 393.7 | 396.4 | 576.2 | 490.3 |
| | 80 | 424.3 | 427.3 | 689.4 | 574.7 |
| | 100 | 438.8 | 442.0 | 750.5 | 619.5 |
| | 200 | 483.7 | 487.4 | 976.2 | 781.5 |
| | 500 | 542.9 | 547.4 | 1380.6 | 1060.9 |
| 광주 | 10 | 212.4 | 214.4 | 219.2 | 215.6 |
| | 20 | 241.3 | 244.0 | 258.7 | 257.2 |
| | 50 | 278.7 | 282.4 | 316.3 | 320.3 |
| | 80 | 297.8 | 301.9 | 348.7 | 356.9 |
| | 100 | 306.8 | 311.2 | 364.8 | 375.5 |
| | 200 | 334.7 | 339.8 | 418.2 | 438.4 |
| | 500 | 371.6 | 377.6 | 497.2 | 535.2 |
| 부산 | 10 | 260.2 | 263.9 | 263.6 | 262.1 |
| | 20 | 298.7 | 301.9 | 310.3 | 307.4 |
| | 50 | 348.5 | 351.3 | 375.8 | 370.4 |
| | 80 | 373.8 | 376.3 | 411.5 | 404.5 |
| | 100 | 385.8 | 388.2 | 429.0 | 421.2 |
| | 200 | 423.0 | 425.0 | 485.7 | 474.8 |
| | 500 | 472.0 | 473.6 | 566.7 | 550.9 |
| 인천 | 10 | 237.8 | 237.9 | 252.8 | 252.8 |
| | 20 | 274.8 | 275.1 | 317.1 | 318.9 |
| | 50 | 322.8 | 323.1 | 420.9 | 426.8 |
| | 80 | 347.1 | 347.6 | 484.3 | 493.3 |
| | 100 | 358.7 | 359.1 | 517.3 | 528.0 |
| | 200 | 394.5 | 395.0 | 632.8 | 650.4 |
| | 500 | 441.7 | 442.4 | 821.8 | 852.8 |
| 추풍령 | 10 | 184.7 | 183.1 | 192.1 | 187.7 |
| | 20 | 210.4 | 207.6 | 228.4 | 217.7 |
| | 50 | 243.7 | 239.3 | 282.9 | 260.2 |
| | 80 | 260.6 | 255.4 | 314.2 | 283.6 |
| | 100 | 268.6 | 263.0 | 330.0 | 295.1 |
| | 200 | 293.5 | 286.7 | 383.2 | 332.7 |
| | 500 | 326.3 | 317.9 | 464.2 | 387.0 |

자료를 이용하여 추정된 95% 신뢰구간(confidential band)의 길이를 표 4에 나타내었다. 모든 지점에서 재현기간의 증가에 따라 신뢰구간 길이도 증가하는 결과를 나타낸다. 또한 같은 분포형에서 매개변수 추정방법에 따른 신뢰구간 길이를 비교했을 경우 베이지안 추론방법이 MLE 보다 현저히 좁은 신뢰구간을 산출하였다. 이는 우도함수와 이를 이용하여 유도한 완전 조건부 사후분포와 제안분포를 이용하여 Gibbs sampling 할 경우 Metropolis-Hastings algorithm을 이용하여 채택율을 높여 추정값들의 표준편차를 줄이며 매개변수의 추

정치를 추정하는 것에 기인한다. 동일한 매개변수 추정방법별로 분포형에 따른 신뢰구간은 Gumbel 분포가 GEV 분포보다 신뢰구간이 좁게 나타난다. 이는 GEV 분포가 형상매개변수의 작은 차이에도 확률강우량의 차이가 많이 나며, 형상매개변수는 매개변수 추정시 다른 매개변수에 비하여 추정 편차가 큰 경향을 보였다. 이로 인하여 Gumbel 분포가 좁은 신뢰구간이 산출되었다.

따라서 매개변수 산정방법 또는 분포형 별로 추정된 신뢰구간을 비교하여, 좁은 신뢰구간을 산출하는 분포형별 매개변

표 4. 재현기간별 신뢰구간 길이(mm)

| 지점명 | 재현기간 (year) | Gumbel 분포 | | GEV 분포 | |
|-----|----------------|-----------|---------|---------|---------|
| | | 최우도법 | 베이지안추정법 | 최우도 | 베이지안추정법 |
| 강릉 | 10 | 10.232 | 2.906 | 20.0422 | 1.843 |
| | 20 | 20.132 | 3.682 | 32.186 | 2.453 |
| | 50 | 35.415 | 4.707 | 41.358 | 3.619 |
| | 80 | 39.739 | 5.215 | 62.478 | 4.712 |
| | 100 | 40.852 | 5.457 | 79.669 | 5.181 |
| | 200 | 45.305 | 6.232 | 92.145 | 5.622 |
| | 500 | 47.754 | 7.244 | 157.018 | 10.719 |
| 광주 | 10 | 17.436 | 2.106 | 15.210 | 7.211 |
| | 20 | 23.247 | 3.152 | 22.825 | 7.519 |
| | 50 | 29.055 | 3.573 | 30.428 | 7.595 |
| | 80 | 30.131 | 4.203 | 38.035 | 7.636 |
| | 100 | 31.215 | 4.628 | 39.219 | 7.643 |
| | 200 | 32.010 | 4.839 | 53.240 | 7.720 |
| | 500 | 32.077 | 5.254 | 68.458 | 11.380 |
| 부산 | 10 | 27.608 | 2.203 | 25.026 | 2.696 |
| | 20 | 29.446 | 3.672 | 40.034 | 4.488 |
| | 50 | 34.040 | 5.149 | 65.056 | 6.275 |
| | 80 | 34.968 | 5.873 | 89.073 | 7.176 |
| | 100 | 35.880 | 6.610 | 100.081 | 8.061 |
| | 200 | 38.649 | 7.341 | 130.105 | 8.960 |
| | 500 | 46.205 | 13.219 | 180.158 | 17.032 |
| 인천 | 10 | 19.075 | 2.034 | 8.546 | 3.483 |
| | 20 | 23.305 | 3.241 | 10.211 | 4.871 |
| | 50 | 28.848 | 4.060 | 50.154 | 5.262 |
| | 80 | 32.095 | 5.184 | 80.441 | 5.569 |
| | 100 | 33.478 | 5.597 | 100.759 | 6.953 |
| | 200 | 37.665 | 7.132 | 110.218 | 9.037 |
| | 500 | 43.445 | 10.638 | 180.272 | 13.212 |
| 추풍령 | 10 | 13.215 | 1.560 | 8.732 | 3.214 |
| | 20 | 13.991 | 2.349 | 11.645 | 4.001 |
| | 50 | 17.493 | 3.128 | 17.464 | 4.887 |
| | 80 | 21.700 | 3.593 | 23.285 | 10.002 |
| | 100 | 22.877 | 3.903 | 28.527 | 10.403 |
| | 200 | 25.172 | 4.371 | 29.108 | 11.606 |
| | 500 | 26.044 | 4.680 | 39.460 | 12.107 |

수 추정법이 통계적 유의성이 높은 확률강우량을 산정한다는 근거로 확률강우량 추정치의 정확도를 비교하였다. 신뢰구간의 불확실성은 변동계수를 이용하여 정량화하였다. 전 대상지점에 대하여 재현기간별 변동계수를 산정하고 표 5에 산정결과를 나타내었다. 표 5의 변동계수의 평균을 통해, 추정방법을 기준으로 변동계수를 비교하였을 경우 베이지안 추정방법이 최우도법 보다 낮은 변동계수를 가지며 분포형을 기준으로 비교할 경우 Gumbel 분포가 GEV 분포보다 낮은 변동계수를 나타낸다.

4. 결 론

본 연구에서는 확률강우량의 불확실성 분석을 통하여 매개변수추정 방법별, 분포형별로 신뢰성이 높은 결과를 추정하는 매개변수 추정방법 및 확률분포형을 평가하고자 하였다. 매개변수 추정방법으로는 고전적인 통계방법 중의 하나로 근사식을 이용하는 최우도법과 추정치의 불확실성을 최소화할 수 있다고 알려진 베이지안 추론 방법을 사용하여 비교하였다. 또한 확률분포형은 극치강우사상에 널리 사용되고 있는

표 5. 재현기간별 변동계수

| 지점명 | 분포형 | 추정법 | 재현기간별 변동계수 | | | | | | | 평균 |
|-----|-----|------|------------|-------|-------|-------|-------|-------|-------|-------|
| | | | 10년 | 20년 | 50년 | 80년 | 100년 | 200년 | 500년 | |
| 강릉 | GUM | MLE | 0.051 | 0.046 | 0.056 | 0.058 | 0.059 | 0.060 | 0.061 | 0.056 |
| | | BAYS | 0.007 | 0.011 | 0.011 | 0.012 | 0.013 | 0.014 | 0.014 | 0.012 |
| | GEV | MLE | 0.016 | 0.047 | 0.055 | 0.072 | 0.079 | 0.086 | 0.088 | 0.063 |
| | | BAYS | 0.005 | 0.012 | 0.015 | 0.014 | 0.015 | 0.015 | 0.016 | 0.013 |
| 광주 | GUM | MLE | 0.059 | 0.061 | 0.061 | 0.060 | 0.060 | 0.060 | 0.061 | 0.060 |
| | | BAYS | 0.013 | 0.012 | 0.015 | 0.014 | 0.013 | 0.013 | 0.014 | 0.013 |
| | GEV | MLE | 0.049 | 0.045 | 0.049 | 0.052 | 0.055 | 0.061 | 0.070 | 0.054 |
| | | BAYS | 0.023 | 0.021 | 0.021 | 0.019 | 0.020 | 0.021 | 0.022 | 0.021 |
| 부산 | GUM | MLE | 0.069 | 0.063 | 0.061 | 0.063 | 0.063 | 0.064 | 0.065 | 0.064 |
| | | BAYS | 0.005 | 0.007 | 0.008 | 0.009 | 0.010 | 0.011 | 0.012 | 0.009 |
| | GEV | MLE | 0.065 | 0.078 | 0.097 | 0.117 | 0.130 | 0.146 | 0.167 | 0.114 |
| | | BAYS | 0.007 | 0.008 | 0.009 | 0.011 | 0.012 | 0.013 | 0.014 | 0.011 |
| 인천 | GUM | MLE | 0.077 | 0.066 | 0.061 | 0.061 | 0.060 | 0.063 | 0.066 | 0.065 |
| | | BAYS | 0.005 | 0.008 | 0.009 | 0.009 | 0.010 | 0.010 | 0.012 | 0.009 |
| | GEV | MLE | 0.088 | 0.060 | 0.063 | 0.069 | 0.077 | 0.087 | 0.097 | 0.077 |
| | | BAYS | 0.006 | 0.012 | 0.009 | 0.010 | 0.012 | 0.013 | 0.013 | 0.011 |
| 추풍령 | GUM | MLE | 0.037 | 0.042 | 0.042 | 0.046 | 0.048 | 0.049 | 0.049 | 0.045 |
| | | BAYS | 0.012 | 0.011 | 0.013 | 0.016 | 0.016 | 0.016 | 0.017 | 0.014 |
| | GEV | MLE | 0.029 | 0.031 | 0.041 | 0.047 | 0.052 | 0.053 | 0.056 | 0.044 |
| | | BAYS | 0.011 | 0.012 | 0.013 | 0.013 | 0.016 | 0.017 | 0.018 | 0.014 |

Gumbel 분포형과 GEV 분포형을 선정하여 분석을 실시하였다. Bootstrap 방법을 이용하여 연속된 곡선의 신뢰구간 추정 방법을 이용하여 신뢰성 분석을 실시한 결과 다음과 같은 결론을 도출하였다.

- 1) 최우도법과 베이지안 추론방법에 의해 추정된 매개변수 추정치의 비교 결과, 매개변수 추정치의 차이는 크지 않았으나 유의수준 5%에서의 추정치의 신뢰구간은 베이지안 추론방법이 좁게 나타나는 것을 확인하였다.
- 2) 신뢰구간 길이를 산정하여 불확실성을 분석한 결과 베이지안 추론방법을 선택하여 확률강우량을 산정한 분포 모형이 최우도법을 사용한 모형보다 현저하게 좁은 신뢰구간을 산출하는 것을 확인할 수 있었다. 이는 베이지안 추론방법의 추정치는 우도함수에 근거한 완전 조건부 사후분포와 완전 조건부 사후분포의 목적 함수로 추정된 제안 분포사이에서 채택율을 높이며 매개변수의 추정치를 추출함으로써 추정치들의 표준편차가 작은 결과를 산출하는데 기인한다.
- 3) 매개변수 추정방법을 기준으로 비교 시 Gumbel 분포형이 GEV 분포형 보다 좁은 신뢰구간을 산출함을 확인하였다. GEV 분포의 확률강우량은 형상매개변수에 민감하게 반응하고, 형상매개변수의 변동폭이 다른 매개변수의 변동폭보다 크게 나타났다. 따라서, 형상매개변수를 가지는 GEV 분포가 Gumbel 분포 보다 넓은 신뢰구간을 가진다.

감사의 글

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국학술진흥재단(또는 한국과학재단)의 지원을 받아 수행된 연구(No. 2009-0069460)입니다.

참고문헌

건설교통부 (2000) 1999년도 수자원관리기법개발 연구조사 보고서, 제 1권 한국 확률강우량도 작성, pp. 77-114.

김달호 (2005) R과 WinBUGS를 이용한 베이지안 통계학. 자유아카데미.

김상욱, 이길성 (2008) Bayesian MCMC를 이용한 저수량 점빈도분석: II. 적용과 비교분석. 한국수자원학회 논문집, 한국수자원학회, 제41권, 제1호, pp. 46-63.

윤용남 (2004) 공업수문학. 청문각.

정중호, 윤용남 (2007) 수자원설계실무. 구미서관.

홍창선, 원석연, 안재현, 안원식 (2001) 확률강우량 산정방법의 신뢰도 분석. 한국도시방재학회논문집, 한국도시방재학회, 제1권, 제3호, pp. 111-122.

Chowdhury, J.U. and Stedinger, J.R. (1991) Confidence interval for design flood with estimated skew coefficient. *Journal of Hydraulic Engineering*, Vol. 117, No. 7, pp. 811-931.

Cohn, T.A., Lane, W.L. and Stedinger, J.R. (2001) Confidence intervals for expected moments algorithm flood quantile estimates. *Water Resources Research*, Vol. 37, No. 6, pp. 1695-1706.

Reis Jr., D.S. and Stedinger, J.R. (2005) Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrol-*

ogy, Vol. 313, No. 1, pp. 97-116.
Stedinger, J.R. (1983) Confidence intervals for design events. *Journal of Hydraulic Engineering*, Vol. 109, No. 1, pp. 13-27.
Surtherland, D.H., Olshen, R.A., Biden, E.N. and Wyatt, M.P. (1988) *The Development of Mature Walking*, London Mackeith Press.
Tasker, G.D. (1987) Comparison of methods for estimating low

flow characteristic of streams. *Water Resources Bull.*, Vol. 23, No. 6, pp. 1077-1083.

© 논문접수일 : 09년 06월 26일
© 심사의뢰일 : 09년 06월 30일
© 심사완료일 : 09년 12월 02일