

Bayesian MCMC를 이용한 저수량 점 빈도분석:

I. 이론적 배경과 사전분포의 구축

At-site Low Flow Frequency Analysis Using Bayesian MCMC:

I. Theoretical Background and Construction of Prior Distribution

김 상 욱* / 이 길 성**

Kim, Sang Ug / Lee, Kil Seong

Abstract

The low flow analysis is an important part in water resources engineering. Also, the results of low flow frequency analysis can be used for design of reservoir storage, water supply planning and design, waste-load allocation, and maintenance of quantity and quality of water for irrigation and wild life conservation. Especially, for identification of the uncertainty in frequency analysis, the Bayesian approach is applied and compared with conventional methodologies in at-site low flow frequency analysis. In the first manuscript, the theoretical background for the Bayesian MCMC (Bayesian Markov Chain Monte Carlo) method and Metropolis-Hasting algorithm are studied. Two types of the prior distribution, a non-data-based and a data-based prior distributions are developed and compared to perform the Bayesian MCMC method. It can be suggested that the results of a data-based prior distribution is more effective than those of a non-data-based prior distribution. The acceptance rate of the algorithm is computed to assess the effectiveness of the developed algorithm. In the second manuscript, the Bayesian MCMC method using a data-based prior distribution and MLE(Maximum Likelihood Estimation) using a quadratic approximation are performed for the at-site low flow frequency analysis.

keywords : At-site low flow frequency analysis, Uncertainty, Bayesian MCMC, Prior distribution, Metropolis-Hastings algorithm, MLE, Quadratic approximation

요 지

저수분석(low flow analysis)은 수자원공학에서 중요한 분야 중 하나이며, 특히 저수량 빈도분석(low flow frequency analysis)의 결과는 저수(貯水)용량의 설계, 물 수급계획, 오염원의 배치 및 관개와 생태계의 보존을 위한 수량과 수질의 관리에 중요하게 사용된다.

그러므로 본 연구에서는 저수량 빈도분석을 위한 점 빈도분석을 수행하였으며, 특히 빈도분석에 있어서의 불확실성을 탐색하기 위하여 Bayesian 방법을 적용하고 그 결과를 기존에 사용되던 불확실성 탐색방법과 비교하였다. 본 논문의 I 편에서는 Bayesian 방법 중 사전분포(prior distribution)와 우도함수(likelihood function)의 복잡성에 상관없

* 서울대학교 BK21 안전하고 지속가능한 사회기반건설 사업단 박사 후 연구원
Post-Doctor, Seoul National University BK21 SIR Group, Seoul National University, Seoul, 151-744, Korea
(e-mail: plethor1@snu.ac.kr)

** 서울대학교 공과대학 건설·환경공학부 교수
Professor, Dept. of Civil and Environmental Engineering, Seoul National University, Seoul, 151-744, Korea
(e-mail: kilselee@snu.ac.kr)

이 계산이 가능한 Bayesian MCMC(Bayesian Markov Chain Monte Carlo) 방법과 Metropolis-Hastings 알고리즘을 사용하기 위한 여러 과정의 이론적 배경과 Bayesian 방법에서 가장 중요한 요소인 사전분포를 구축하고 이를 비교 및 평가하였다. 고려된 사전분포는 자료에 기반하지 않은 사전분포와 자료에 기반한 사전분포로써 두 사전분포를 이용하여 Metropolis-Hastings 알고리즘을 수행하고 그 결과를 비교하여 저수량 빈도분석에 합리적인 사전분포를 선정하였다. 또한 알고리즘의 수행과정에서 필요한 제안분포(proposal distribution)를 적용하여 그에 따른 알고리즘의 효율성을 채택률(acceptance rate)을 산정하여 검증해 보았다. 사전분포의 분석 결과, 자료에 기반한 사전분포가 자료에 기반하지 않은 사전분포보다 정확성 및 불확실성의 표현에 있어서 우수한 결과를 제시하는 것을 확인할 수 있었고, 채택률을 이용한 알고리즘의 효율성 역시 기존 연구자들이 제시하였던 만족스러운 범위를 가지는 것을 알 수 있었다. 최종적으로 선정된 사전분포는 본 연구의 II편에서 Bayesian MCMC방법의 사전분포로 이용되었으며, 그 결과를 기존 불확실성의 추정방법의 하나인 2차 근사식을 이용한 최우추정(maximum likelihood estimation)방법의 결과와 비교하였다.

핵심용어 : 저수량 점 빈도분석, 불확실성, Bayesian MCMC, Metropolis-Hastings 알고리즘, 사전분포 최우추정방법, 2차 근사법

1. 서 론

저수분석(low flow analysis)은 수공구조물의 설계, 하천환경의 보전 및 생활·공업·농업용수의 안전한 취수를 위한 최소 유량의 보장, 오염원의 배치 등 수량과 수질의 관리에 중요하게 사용된다. 년 평균 유량, 절대 최소 유량 등 저수량을 표현하기 위한 많은 지표들이 존재하지만 특정한 년 최소 유량의 시계열 자료를 이용한 빈도분석의 결과가 저수량을 나타내기 위하여 가장 흔히 사용되어 진다. 저수량 분석을 수행하기 위해서 국내에서는 주로 갈수량에 해당되는 355위 유량(Q_{355})을 이용한 이 값의 10년 평균값(평균 갈수량) 또는 10년 빈도분석 값(기준 갈수량)을 산정하여 사용하고 있으며, 미국 및 영국 등의 국가에서는 7일 지속기간 10년 빈도 유량 (7Q10)을 주로 사용하고 있다. 그러나 수자원장기 종합계획, 댐 건설 장기계획 등의 국내 주요 중장기 계획은 빈도분석의 확정적인(deterministic) 값만을 이용하여 수립되고 있으며, 빈도분석결과의 불확실성을 반영한 확률적인(probabilistic) 값이 이용되는 계획은 찾아보기 힘들다. 이는 불확실성에 대한 인식 부족과 함께 불확실성에 대한 계산방법이 현실을 제대로 반영하지 못함으로써 빈도분석 결과의 정확성에 대한 신뢰도가 낮은 것에 기인한다고 할 수 있다.

빈도분석을 수행하고자 하는 지점에서의 자료의 길이가 충분한 경우에는 점 빈도분석(at-site frequency analysis)을 적용하여 결과를 산정할 수 있다. 점 빈도분석의 첫 번째 과정은 확보된 자료를 합리적으로 표현할 수 있는 확률 분포함수를 선정하는 것이며, 가장 흔히 사용되는 확률 분포함수로는 이모수 또는 삼모수의

Lognormal 분포함수와 Weibull 분포함수라 할 수 있다 (Gumbel, 1958; Gary, 1987; Nathan and MacMahon, 1990). 통계적 적합성 검토를 이용하여 최종적인 확률 분포함수가 선정되면 적률법(method of moment), 최우 추정(maximum likelihood estimation)방법 등과 같은 모수적 방법이나 비모수적 방법을 사용하여 선정된 확률 분포함수의 모수를 추정해야 한다. 추정된 모수를 이용하여 구하고자 하는 재현기간(return period)에 해당되는 분위수(quantile)를 추정할 수 있으며, 최종적으로 특정 지점에 대한 빈도곡선을 작성할 수 있다.

위와 같은 빈도분석과정을 이용하면 특정지점에서의 확정적인(deterministic) 결과 값을 얻을 수는 있지만 빈도분석 결과의 불확실성을 표현할 수 있는 확률적인(probabilistic) 결과를 산정할 수는 없다. 불확실성을 나타내기 위해서 일종의 근사식을 사용하여 대략적인 신뢰구간을 산정하고 이로부터 분석결과의 불확실성을 표현하고자 하는 연구가 진행된 바 있다 (Stedinger, 1983; Chowdhury and Stedinger, 1991; Stedinger et al., 1993, Cohn et al., 2001). 그러나 근사식을 사용한 신뢰구간 산정방법은 확률 분포함수의 모수를 산정함에 있어서 정상성(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 사용한 모수의 불확실성 추정방법을 대신하여 Bayesian 접근방법을 사용한 모수 및 불확실성의 추정이 수행될 수 있다. Bayesian 방법은 근사식을 사용하기 위한 가정 조건이 필요하지 않기 때문에 특히 불확실성을 표현하는 데 있어서 근사식을 사용한 방법

보다 우월할 수 있다. Bayesian 방법을 사용하기 위해서는 사전분포(prior distribution)로부터 사후분포(posterior distribution)를 산정해야 하는데, 이를 계산하기 위해서는 해석적으로 산정되기 어려운 적분항들이 포함되어 있다. 그러므로 과거 초기 연구단계에서는 공액사전분포(conjugated prior distribution)를 적용하여 사후분포를 쉽게 해석적으로 구하여 모수 및 불확실성을 나타내고자 하는 연구가 진행된 바 있다(Vicens et al., 1975; Wood and Rodriguez-Iturbe, 1975a, b). 그러나 공액 사전분포를 사용한 Bayesian 접근 방법의 수자원 공학으로의 적용은 자료를 합리적으로 표현할 수 있는 사전분포를 사용하게 되지 못하는 경우가 많아 Bayesian 방법을 사용하기 위한 원래 목적을 벗어나게 될 수 있다는 논쟁으로 그 간 활발히 적용되지 못하다가 최근 들어 발전한 계산 능력과 전체 탐색법 개념의 알고리즘의 개발로 인하여 90년대 후반부터 다시 빈도 분석 분야, 강우-유출모형의 보정 분야 등에 활발히 사용되고 있다(Coles and Powell, 1996; Madsen and Rosberg, 1997; Kuczera and Parent, 1998; Kuczera, 1999; Zhang and Govindaraju, 2000; Thiemann et al., 2001; Wang, 2001; O'Connell et al., 2002; Vrugt et al., 2003; Kingston et al., 2005; Reis Jr et al., 2005; Reis Jr. and Stedinger, 2005; Kavetski et al., 2006; Seidou et al., 2006; 김, 2007; Lee and Kim, 2007).

그러나 위에서 제시한 저수분석의 중요성에도 불구하고 Bayesian 방법을 사용하여 저수량 빈도분석을 수행하기 위한 일련의 과정들은 연구사례가 상대적으로 적다. 그러므로 본 연구에서는 최근 활발히 사용되고 있는 Metropolis-Hastings 알고리즘(Metropolis et al.)을 적용한 Bayesian Markov Chain Monte Carlo (Bayesian MCMC) 방법을 사용하여 저수량 빈도분석을 수행하고 그 결과를 MLE 이차 근사식을 이용한 결과와 비교하여 두 방법 간의 장단점을 비교하는 연구를 수행하였다. 이를 위하여 연구의 I 편인 본 논문에서는 연구의 수행을 위해 필요한 이론적 배경을 서술하고 Bayesian 방법을 적용하는데 있어서 중요한 사전분포를 구축하는 연구를 수행하였다.

2. Bayesian 방법과 Bayesian MCMC

베이즈의 정리는 A가 먼저 발생하고 그 후에 B가 발생하는 두 개의 사건 A, B가 서로 종속적일 경우 A의 사건에 의해 B 사건의 확률이 달라진다는 것이다. 베이즈의 정리를 수식으로 나타내면 Eq. (1)과 같고, 여기서 각각의 확률 사건을 연속 확률밀도함수(Probability

density function)로 나타내면 베이즈의 정리는 Eq. (2)와 같이 표현될 수 있다.

$$\Pr(B_j|A) = \frac{\Pr(B_j)\Pr(AB_j)}{\sum_{j=1}^n \Pr(AB_j)\Pr(B_j)} \quad (1)$$

$$\pi(\theta|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) = \frac{f(\mathbf{x}_1|\theta) \cdots f(\mathbf{x}_n|\theta)\pi(\theta)}{\int_{\theta} f(\mathbf{x}_1|\theta) \cdots f(\mathbf{x}_n|\theta)\pi(\theta)d\theta} \quad (2)$$

Eq. (2)에서 좌변의 $\pi(\theta|\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ 는 사후분포(Posterior distribution), 우변 분자의 $\pi(\theta)$ 는 사전분포(Prior distribution)라 명명되며, 우변의 분모는 상수로서 주변분포(Marginal distribution)이고, 우변의 분자의 $f(\mathbf{x}_1|\theta) \cdots f(\mathbf{x}_n|\theta)$ 는 발생할 수 있는 모든 가능성을 고려한 우도함수(Likelihood function)이다. 그러므로 Eq. (2)로부터 사후분포는 우도함수와 사전분포의 곱에 비례하게 됨을 알 수 있다. 분석하고자 하는 자료를 나타낼 수 있는 확률밀도함수가 결정되면 이로부터 우도함수를 유도할 수 있고, 적절한 사전분포를 부여함으로써 사후분포로부터 확률밀도함수의 모수를 추출하고 모수의 불확실성을 탐색할 수 있다.

Bayesian 방법을 이용한 모수의 추정에는 모수를 미지의 상수로 간주하는 것이 아니라 미지의 난수(random number)로 간주하게 됨으로써 추정의 관심이 되는 모수의 불확실성의 정도를 확률 모형을 이용하여 표현할 수 있게 된다. 결국 Bayesian 방법을 이용한 모수의 추정은 자료로부터 얻은 모수에 대한 정보와 모수에 대한 과거의 경험 또는 주관적 사전분포를 표현함으로써 보다 정확한 모수의 불확실성에 대한 탐색에 그 목적이 있다고 할 수 있다.

그러나 Bayesian 방법을 이용하여 사후분포를 계산하는 것은 쉬운 작업이 아니다. 특히 Eq. (2)의 우변의 분모의 적분은 결정된 확률밀도함수에 따라 적분이 가능할 수도 있으나, 대부분의 확률밀도함수는 수학적으로 적분하기 어려운 경우가 많다. 또한 확률밀도함수로 표현되는 우도함수와 사전분포를 이용하여 모수의 추출이 가능한 형태의 사후분포를 산정하는 것은 수학적으로 어려운 경우가 많다. 그러므로 Bayesian 방법의 초기 연구단계에서는 위와 같은 Bayesian 방법의 적용에 필요한 계산을 위하여 공액사전분포(conjugated prior distribution)를 이용하는 경우가 많았다.

공액사전분포란 수학적 계산이 가능한 우도함수와 사전분포의 조합으로써 예를 들어 우도함수를 이항분포로 사용하고 사전분포를 Beta분포로 사용하면 사후분포

포가 Beta분포로 계산되어 진다. 그러나 공액사전분포의 사용은 우도함수와 사전분포를 결정하는 데 있어서 많은 한계가 있고, 특히 사전분포를 선정함에 있어서 모수의 과거 경험이나 사용자의 직관과는 무관한 분포를 선정해야 하는 단점이 있다. 이는 Bayesian 방법을 사용하는 데 있어서 가장 중요한 사전분포의 의미와 모순되는 사항이며, 이러한 모순으로 인해 Bayesian 방법은 많은 논란이 있었다.

그러나 최근 계산능력의 하드웨어부분의 발전과 주변 확률분포함수의 적분이 필요 없는 Metropolis-Hastings 알고리즘, Gibbs sampling 알고리즘, 중요도 샘플링(Importance sampling) 알고리즘과 같은 Bayesian MCMC 방법에 입각한 Bayesian 계산방법의 소프트웨어부분의 발전으로 인하여 공액사전분포를 사용하지 않아도 Eq. (2)의 계산이 가능하게 됨으로서 서론에 언급한 연구사례와 같이 최근 들어 수자원공학 분야에서도 다시 활발히 적용되고 있는 실정이다.

Bayesian MCMC방법이란 마코프 연쇄(Markov chain)과 몬테카를로 적분(Monte carlo integration)을 이용하여 사후분포로부터 모수를 추출하고 통계적 특성치를 계산하는 방법이다. 즉, 마코프 연쇄를 이용하여 모수간의 관계를 구성하고 이를 상당히 큰 수만큼 반복하는 몬테카를로 적분기법을 이용하여 최종적으로 모수의 통계적 특성을 산정하는 방법이다. 여러 가지 Bayesian 계산 방법 중에서 가장 활발히 사용되고 있는 알고리즘은 Metropolis-Hastings 알고리즘으로 기본적인 개념은 Metropolis et al.(1953)에 의하여 만들어 졌으나 최근 들어 개선되어져 활발히 이용되고 있다. 마코프 연쇄는 불변(invariant)인 임의의 함수, f 와 임의의 제안분포(proposal distribution), q 를 이용하여 다음과 같이 정의될 수 있다. 또한 Eq. (3)에서 f 를 구하고자 하는 사후분포, $\pi(\theta|D)$ 가 불변하다는 가정 하에 대체하면 좌변 또는 우변이 큰 부등식 (4)를 얻을 수 있다.

$$q(\theta_{j+1}|\theta_j)f(\theta_j|D) = q(\theta_j|\theta_{j+1})f(\theta_{j+1}|D) \quad (3)$$

$$q(\theta_{j+1}|\theta_j)\pi(\theta_j|D) > q(\theta_j|\theta_{j+1})\pi(\theta_{j+1}|D) \quad (4)$$

Eq. (4)의 부등식을 등식으로 바꿀 수 있는 값은 반드시 존재하는데, 이를 ρ 로 표현하여 고려하면 다음과 같은 등식 (5)를 얻을 수 있다.

$$\rho(\theta_j, \theta_{j+1})q(\theta_{j+1}|\theta_j)\pi(\theta_j|D) = q(\theta_j|\theta_{j+1})\pi(\theta_{j+1}|D) \quad (5)$$

Eq. (5)를 ρ 에 대하여 정리하면 최종적인 Eq. (6)을

얻을 수 있고, ρ 를 채택확률(acceptance probability)이라 한다.

$$\rho(\theta_j, \theta_{j+1}) = \min \left[\frac{q(\theta_j|\theta_{j+1})\pi(\theta_{j+1}|D)}{q(\theta_{j+1}|\theta_j)\pi(\theta_j|D)}, 1 \right] \quad (6)$$

그러므로 채택확률은 마코프연쇄를 구성하는 임의의 두 모수에 대한 제안분포의 값과 사후분포의 값에 의해 계산되어지며, 확률적 개념 하에 1보다 작은 값으로 계산되어진다. 합리적인 채택확률이 계산되고 알고리즘이 효율성을 가지기 위해서는 제안분포가 무작위적인 수를 발생시킬 수 있도록 합리적으로 선정되어야 할 필요가 있다. 채택확률을 이용하여 추정하고자 하는 모수를 구하는 Metropolis-Hastings 알고리즘은 다음과 같은 과정으로 표현할 수 있다.

- Step 1) $j=0$ 에서의 임의의 θ_0 를 선정한다.
- Step 2) 제안분포로부터 제안 모수 θ_* 를 무작위적으로 생성한다.
- Step 3) 식 (6)으로부터 채택확률을 계산한다.
- Step 4) 0과 1사이의 균일분포로부터 무작위수 u 를 생성한다.
- Step 5) 만약 $u < \rho$ 이면, θ_* 를 θ_{j+1} 로 교체하고 반대의 경우에는 θ_j 를 θ_{j+1} 로 교체한다.
(즉 $u > \rho$ 인 경우에는 제안분포로부터 생성된 θ_* 를 사용하지 않는다.)
- Step 6) j 를 1 증가시키고 Step 2)로 돌아간다.
- Step 7) 충분히 큰 수만큼 위 과정을 반복한다.

위의 알고리즘을 이용하여 추정된 모수들은 최종적으로 몬테카를로 적분을 통하여 추정된 모수들의 평균과 같은 통계적 특성치를 근사적으로 얻을 수 있으며, 기존 통계학에서 사용되는 신뢰구간(confidence interval)에 해당되는 신용구간(credible interval)도 얻을 수 있고 이를 이용하여 추정된 모수의 불확실성을 표현할 수 있다. 반복하여 추출된 모수의 개수, n 과 유의수준(significance level), α 에서의 Bayesian MCMC 방법에 의한 신용구간, $100(1-\alpha)\%$ 는 다음 Eq. (7)과 같이 나타낼 수 있다.

$$(\theta_{[(\alpha/2)n]}, \theta_{[(1-\alpha/2)n]}) \quad (7)$$

즉, Eq. (7)에서 $\theta_{[(\alpha/2)n]}$ 와 $\theta_{[(1-\alpha/2)n]}$ 는 추출된 모수들 중에서 $(\alpha/2)n$ 번째와 $(1-\alpha/2)n$ 번째에 해당되는 값이다.

3. 최우추정법(MLE method): 2차 근사식을 이용한 불확실성의 산정

최우추정(Maximum Likelihood Estimation, MLE) 방법은 특정 공간에서의 모집단을 임의의 확률밀도함수로 가정한 후 자료가 추출될 수 있는 확률을 나타내는 우도함수(likelihood function)를 최대가 되도록 하는 확률 밀도함수의 모수를 결정하는 일종의 최적화 방법이라 할 수 있다. 즉 n 개의 자료와 이를 나타낼 수 있는 모수가 포함되어 있는 임의의 확률밀도함수, $f_X(x)$ 가 존재하면 우도함수, L 은 다음과 같이 표현될 수 있다.

$$L(\theta) = \prod_{i=1}^n f_X(x_i | \theta) \quad (8)$$

최우추정방법은 우도함수를 최대화시키도록 하는 θ 를 최적화시키는 방법이므로, 우도함수의 형태에 따른 여러 가지 최적화방법을 고려할 수 있다. 즉, 우도함수가 단일 모드이고 모수의 차원이 크지 않은 경우에는 경사법(gradient method)과 같이 우도함수를 각각의 모수로 편미분한 방정식을 구성한 후 기울기가 0이 되는 모수를 반복법(iterative method)과 같은 방법을 사용하여 구할 수 있다. 그러나 우도함수가 다중 모드이고 모수의 차원이 큰 경우에는 경사법과 같은 최적화 방법은 전역최적해(global optimum)보다는 국부최적해(local optimum)를 찾을 가능성이 많으므로, 최근 들어 사용하고 있는 전역최적화 알고리즘을 사용할 필요가 있다.

우도함수에 따른 최적화방법을 적용하여 추정하고자 하는 모수를 구한 이후에는 모수의 불확실성을 나타내기 위하여 다음 식과 같이 2차 근사식을 이용한 우도비함수(likelihood function ratio)를 이용할 수 있으며, 2차 근사식은 $\theta = \hat{\theta}$ 에서 최대값을 가져야 하므로 다시 임의의 상수 I 를 사용하여 2차식으로 표현될 수 있다.

$$\begin{aligned} \log RL(D| \theta) &= \log \frac{L(D| \theta)}{L(D| \hat{\theta})} \quad (9) \\ &\cong a + b\theta + c\theta^2 \cong -\frac{1}{2}I(\theta - \hat{\theta})^2 \end{aligned}$$

여기서, RL은 상대우도(relative likelihood)를 나타내며 D 는 주어진 자료, $\hat{\theta}$ 는 θ 의 추정치이고 a, b, c 는 상수이다.

또한 Meeker and Escobar(1998)는 Eq. (9)에서 θ 의 불확실성을 추정하기 위하여 다음과 같이 정상성(normality)을 가정으로 하는 다음의 근사식을 제안한

바 있다.

$$\frac{\partial \log L}{\partial \theta} \cong -I(\theta - \hat{\theta}) \sim N(0, I) \quad (10)$$

그러므로 Eq. (10)으로부터 $\hat{\theta} \sim N(0, I^{-1})$ 임을 알 수 있고 유의수준 α 에서 다음과 같은 신뢰구간을 유도할 수 있다.

$$\Pr\left[\hat{\theta} - \frac{z_{(1-\alpha/2)}}{\sqrt{I}} < \theta \leq \hat{\theta} + \frac{z_{(1-\alpha/2)}}{\sqrt{I}}\right] \approx 1 - \alpha \quad (11)$$

Bickel and Doksum(1977)은 I 를 자료의 표준오차로부터 추정하였으며, Eq. (11)의 신뢰구간의 제안식은 자료의 수가 많은 경우에는 합리적일 수 있으나, 자료의 수가 적어지면 좋은 결과를 산정할 수 없음을 제안한 바 있다.

4. 확률밀도함수의 선정 및 우도함수

위에서 언급한 바와 같이 저수량 점 빈도분석을 수행하기 위해서는 가장 먼저 확률밀도함수를 선정해야 한다. 저수량 빈도분석을 위한 대상유역의 선정, 자료의 선정 및 특성 등은 본 연구와 병행되어 수행된 동 권의 II편 연구인 저수량 빈도분석의 응용 부분에서 상세히 언급하였으므로, 본 연구에서는 적용 대상 유역인 낙동강 유역에서 Bayesian MCMC방법을 적용하기 위한 확률밀도함수의 선정, 사전분포의 유도, 제안분포의 효율성 검토에 연구의 초점을 맞추어 진행하였다.

본 연구에서 사용되는 확률밀도함수는 Nathan and McMahon(1990), Önöz and Bayazit(2001)이 저수량 빈도분석을 위해 사용한 바 있는 2모수 Weibull 분포로서 확률밀도함수와 확률분포함수(cumulative distribution function)는 각각 다음의 식으로 나타낼 수 있다.

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right] \quad (12)$$

$$F(x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right] \quad (13)$$

여기서, α 는 형상모수(shape parameter)이고 β 는 축척모수(scale parameter)이다.

낙동강 유역의 진동지점에서 얻어진 36년간의 유량 자료를 자연유량으로 환산하고 7일 지속기간 최소유량(7Q)을 각 년마다 계산하여 36개의 7Q 자료를 산정한

Table 1. Results of PPCC Test

| Length of data | Critical value at significance level, 0.05 | Critical value at significance level, 0.01 | PPCC test statistic, \hat{r} |
|----------------|--|--|--------------------------------|
| 36 | 0.9638 | 0.9239 | 0.9486 |

후 PPCC(Probability-Plot Correlation Coefficient) 시험방법을 이용하여 선정된 분포의 적정성을 확인하였다. PPCC 시험방법은 Filliben(1975)에 의해 제안된 방법으로써 정렬된 자료와 그에 따라 추정된 분위수의 상관관계수(\hat{r})를 산정하여 분포의 적정성을 검토하는 방법이다. 계산된 상관관계수, \hat{r} 은 검정하고자 하는 유의수준, α 에서의 한계상관관계수, \hat{r}_α 를 계산하여 비교함으로써 가설을 검정할 수 있다. Table 1은 유의수준 0.05와 0.01에서 산정된 한계상관관계수와 36개의 자료로부터 산정된 상관관계수를 나타낸 것으로써, 산정된 상관관계수는 유의수준 0.01에서의 한계상관관계수 값과 0.05에서의 한계상관관계수 값의 사이에 존재하는 것을 알 수 있다. 즉, 제안된 확률분포인 2모수 Weibull 분포는 유의수준 0.01에서 적정하다고 검정할 수 있었다.

본 연구의 최종적인 목표는 Bayesian MCMC방법과 MLE 방법을 적용하여 저수량 빈도분석을 각각 수행하고 그 결과를 비교 평가함으로써 저수량 빈도분석을 수행하는 데 있어서 두 방법 간의 장점 및 단점을 분석하는 것이며, 이에 대한 결과는 본 연구의 II편을 통하여 언급하였다. 두 가지 방법은 모두 선정된 확률밀도함수를 이용하여 얻어진 우도함수를 필요로 하는 데, 본 연구에서는 2모수 Weibull 분포로부터 얻어진 우도함수를 두 가지 방법에 모두 동일하게 사용하였다. 2모수 Weibull 분포함수를 이용한 우도함수를 나타내면 다음과 같다.

$$L(\mathbf{x}|\alpha, \beta) = \left(\frac{\alpha}{\beta}\right)^n \prod_{i=1}^n \left(\frac{x_i}{\beta}\right)^{\alpha-1} \exp\left[-\sum_{i=1}^n \left(\frac{x_i}{\beta}\right)^\alpha\right] \quad (14)$$

5. 사전분포의 선정 및 비교

확률밀도함수의 선정에 따른 우도함수가 결정되었으므로 Eq. (2)를 2모수 Weibull 분포함수를 이용하여 표현하면 다음의 식과 같다.

$$\pi(\alpha, \beta|\mathbf{x}) = \frac{L(\mathbf{x}|\alpha, \beta)\pi(\alpha, \beta)}{\int_0^\infty \int_0^\infty L(\mathbf{x}|\alpha, \beta)\pi(\alpha, \beta)d\alpha d\beta} \quad (15)$$

여기서, $\pi(\alpha, \beta)$ 는 형상모수와 축척모수의 결합확률밀

도함수로써 두 모수의 특징을 나타내는 사전분포이다. Bayesian 방법을 적용하기 위해서는 적절한 사전분포를 선정하는 것이 가장 중요하고도 어려운 작업이며, 사전분포는 크게 자료에 기반하지 않은 사전분포와 자료에 기반한 사전분포로 구분할 수 있다. 본 연구에서는 자료에 기반하지 않은 사전분포와 자료에 기반한 사전분포를 각각 Eq. (15)에 적용하고 Metropolis-Hastings 알고리즘을 이용하여 두 가지 사전분포를 이용한 추정결과를 비교함으로써 자료에 기반한 사전분포의 저수량 빈도분석에의 이용 가능성을 분석하였다.

5.1 자료에 기반하지 않은 사전분포: Prior I

자료에 기반하지 않은 사전분포 중 대표적인 사전분포는 무정보적 사전분포(non-informative prior distribution)로써 이는 추정하고자 하는 모수에 대한 과거 경험에 대한 정보나 사용자의 주관이 전무한 경우에 사용된다. 무정보적 사전분포로는 Gelman et al.(2004)이 제안한 Normal, Log-normal, Exponential, 균일분포가 사용될 수 있으나, 본 연구에서는 각각의 모수에 대하여 다음과 같은 균일분포만을 적용하고 두 균일분포가 서로 통계적으로 독립적이라 가정 하에 최종적으로 $\pi(\alpha, \beta)$ 를 유도하였다. 여기서 $b=2.5, a=1.5$ 와는 형상모수의 대략적인 범위를 고려하여 임의로 설정하였다.

$$\pi(\alpha) = \frac{1}{b-a} \quad (b=2.5, a=1.5) \quad (16)$$

$$\pi(\beta) = \frac{1}{\beta} \quad (17)$$

$$\therefore \pi(\alpha, \beta) = \pi(\alpha)\pi(\beta) = \frac{1}{\beta} \quad (18)$$

5.2 자료에 기반한 사전분포: Prior II

자료에 기반한 사전분포는 추정하고자 하는 모수의 과거 자료나 사용자의 경험에 기반한 주관적인 판단에 의해 구축될 수 있다. 그러나 사용자의 경험에 기반한 주관적 사전분포는 사용자의 경험을 자료를 이용하여 표현하는데 있어서 정량화가 힘들어 어려운 부분이 있다.

본 연구에서는 자료에 기반한 사전분포를 구축하기

위하여 에르고딕(ergodic)가정을 이용하여 낙동강 유역의 13개 지점의 7Q자료를 이용하여 진동에서의 자료에 기반한 사전분포를 구축하였다. 에르고딕 가정이란 임의 공간에서의 앙상블(ensemble) 평균(또는 공간 평균)과 시간 평균이 같게 된다는 가정으로서 이 가정을 이용하면 Fig. 1에서 진동 지점을 제외한 나머지 13개 지점에서의 통계적 특성치가 진동 지점에서의 13년간의 통계적 특성치와 일치하게 된다는 가정을 수립할 수 있고, 이로부터 진동지점에 대한 자료에 기반한 사전분포를 구축할 수 있다. 그러나 이와 같은 에르고딕 가정이 수문학적 범주에서 만족되려면 진동지점을 포함한 14개 지점의 유량 특성이 균일해야 된다는 가정이 만족되어야 한다. 즉 14개 지점으로 대변되는 낙동강유역이 수문학적으로 동질한 유역(hydro- logical homogeneous region)인지를 평가해 볼 필요가 있는 데, 본 연구에서는 Matlab에서 제공하는 K-means 알고리즘을 이용한 군집분석(cluster analysis) 툴 박스와 14개 지점의 36년간 저수량(7, 8, 9월 유량 제외)만을 이용하여 동질성을 판별해 보았다. 그림 2는 군집분석의 결과를 나타낸 것으로 14개 지점이 하나의 동질한 유역으로 묶여졌을 때, 실루엣(silhouette)값이 모두 0.6을 초과하는 값으로 산정되어 지는 것을 확인할 수 있었으며 이로부터

13개 지점의 유량특성이 동질하다는 것을 확인할 수 있었다.

필요한 사전분포는 형상모수 α 와 축척모수 β 에 대하여 구축되어야 하므로, 먼저 진동지점을 제외한 13개 지점의 7Q 유량을 2모수 Weibull 분포에 적합시킨 후, 13개의 α 추정치와 β 추정치를 각각 모수 λ 를 가지는 지수분포와 형상모수 b , 축척모수 a 를 가지는 2모수 Weibull 분포를 이용하여 다시 적합시켰다. 즉, 에르고딕 가정 하에 공간적으로 흩어져 있는 13개의 α 추정치와 β 추정치를 진동지점에서의 과거 추정치 자료로 이용한 셈이며, α 추정치는 지수분포로 β 추정치는 2모수 Weibull 분포를 이용하여 나타내었다. 각각의 추정 결과를 확률도시(probability plot)를 이용하여 적합성을 검정해 본 결과 각각의 분포에 합리적으로 일치하는 결과 (Fig. 3)를 확인할 수 있었으며, 정량적인 검정을 위하여 유의수준 0.5 %에서의 Kolmogorov-Smirnov 검정을 수행한 결과 두 분포 모두 합리적인 분포로 검정되었다 (Table 2). 최종적으로 추정된 λ 와 형상모수 b , 축척모수 a 의 값은 Table 2와 같다.

위와 같은 과정을 통하여 본 연구에서 사용되어지는 2모수 Weibull 분포의 형상모수 α 와 축척모수 β 에 대한 사전분포의 구축을 나타내면 다음 식과 같다.



Fig. 1. Concept for Construction of Data Based Prior Distribution (Prior II)

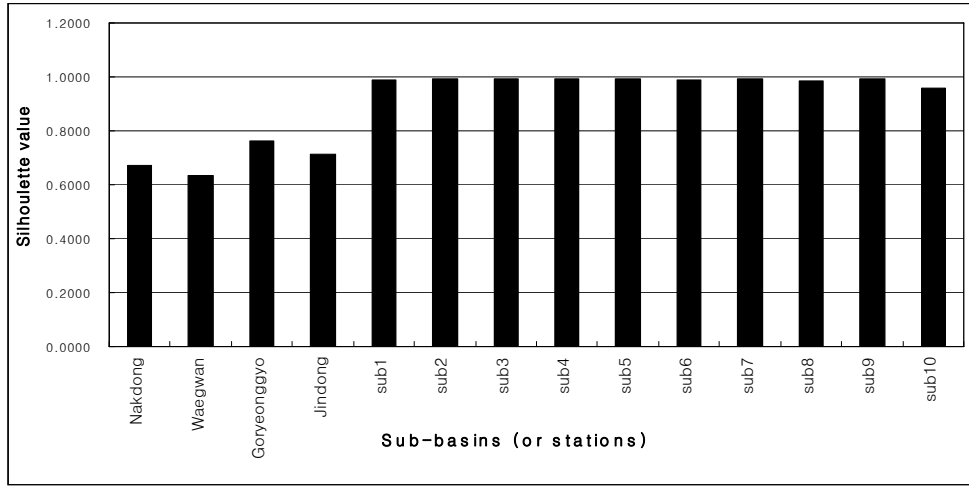


Fig. 2. Results of Cluster Analysis

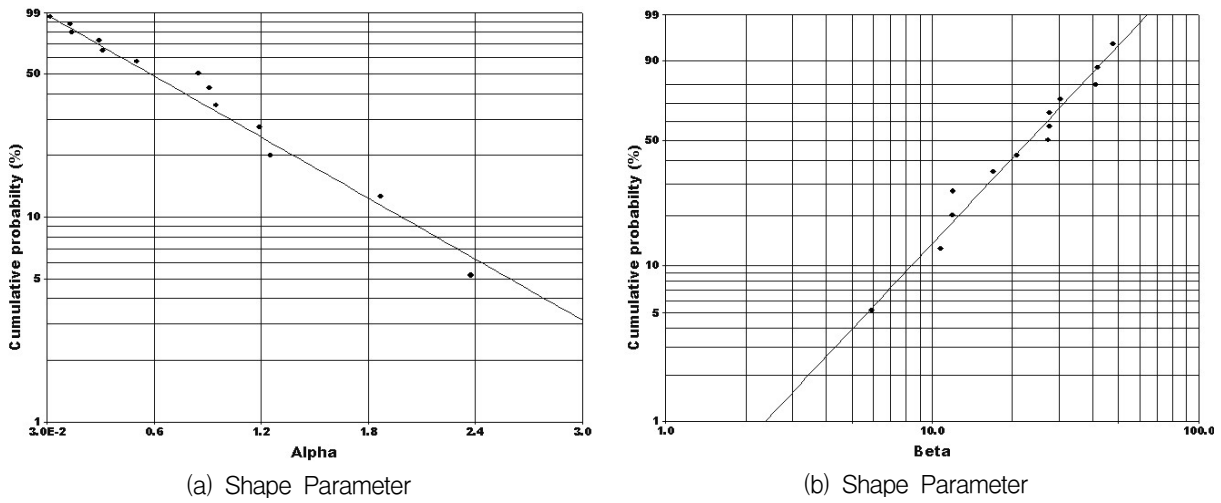


Fig. 3. Evaluation of Agreement using Probability Plot

Table 2. Results of Estimates by MLE and Test Statistics

| Weibull parameters | Estimates by MLE | p-value |
|--|---|---------|
| Shape parameter, α (be assumed as Exponential) | Arrival rate (λ) = 0.9757 | 0.8911 |
| Scale parameter, β (be assumed as Weibull) | Scale (a) = 27.1506 Shape (b) = 1.9189 | 0.9756 |

$$\pi(\alpha) = \exp(-\lambda\alpha) \quad (19)$$

$$\pi(\beta) = \frac{b}{a} \left(\frac{\beta}{a}\right)^{b-1} \exp\left[-\left(\frac{\beta}{a}\right)^b\right] \quad (20)$$

$$\therefore \pi(\alpha, \beta) = \frac{b}{a} \left(\frac{\beta}{a}\right)^{b-1} \exp\left[-\left(\left(\frac{\beta}{a}\right)^b + \lambda\alpha\right)\right] \quad (21)$$

여기서 λ , a , b 는 위에서 각각 추정된 0.98, 27.15, 1.92이다.

6. 제안분포의 선정과 Metropolis-Hastings 알고리즘의 효율성 검증

위의 이론적 배경에서 언급한 바와 같이 Metropolis-Hastings 알고리즘을 수행하기 위해서는 적절한 제안분포(proposal distribution)가 필요하며, 최종적으로 채택률(acceptance rate)을 계산하여 적절한 제안분포인지를 검증해 보아야 한다. Chip and Greenberg(1995)는 Metropolis-Hastings 알고리즘을 수행하기 위한 5가지의 적절한 사전분포를 제안한 바 있는데, 본 연구에서

는 5개의 제안분포 중 Eq. (22)와 같이 이변량 정규분포(bivariate normal distribution)를 사용한 확률보행(random walk) 형태의 제안분포를 사용하였다.

$$q(\theta_{j+1}|\theta_j) = q_1(|\theta_{j+1} - \theta_j|) \quad (22)$$

여기서, q_1 은 이변량 정규분포이며, 알고리즘 내에서 추정하고자 하는 모수가 적절한 값을 가지도록 정규분포의 분산은 수동적인 방법을 통하여 산정하였다. 이변량 정규분포를 사용하면 채택확률을 산정하는 식에서 $q(\theta_j|\theta_{j+1}) = q(\theta_{j+1}|\theta_j)$ 가 되어 채택확률의 산정이 사후 분포에만 관련되므로 알고리즘의 효율성이 좋은 것으로 Chip and Greenberg(1995)가 제안한 바 있으며 본 연구에서도 이들의 결론에 의해 이변량 정규분포를 사용하였고, 채택률을 산정하여 제안분포의 적정성을 확인하였다.

위의 과정을 통하여 우도함수, 2가지 종류의 사전분포, 제안분포가 모두 구축되었으므로 Metropolis-Hasting 알고리즘을 사용한 Bayesian MCMC 방법에 의해 추정하고자 하는 모수의 추정이 가능하다. 본 연구에서는 알고리즘의 최대 반복횟수를 100,000번으로 설정하였으며, 최초 샘플링된 1,000개의 추정치는 모형

의 안정성을 위하여 평균계산 시 제외하였으며 (Burn-in=1,000), Burn-in 값은 모의된 모수가 변동하지 않는 구간을 주관적으로 선정하여 산정하였다. 그림 4는 Bayesian MCMC방법에 의해 추정된 99,000개의 2모수 Weibull 분포의 형상모수 α 와 축척모수 β 를 나타낸 것이다.

알고리즘의 안정성을 검증하기 위해서는 위에서 언급한 바와 같이 채택률을 계산할 필요가 있다. 채택률이란 알고리즘의 반복회수에 대한 추정된 모수의 채택된 비율을 의미하며, 채택률이 낮은 경우에는 모수의 새로운 값이 추정되지 않으므로 제안분포 등을 새로 고려해야 하고 반대로 채택률이 너무 높은 경우에는 모수를 탐색하는데 있어서 과도한 시간이 필요하므로 반복 횟수를 증가시키게 되어 계산시간이 너무 길어지는 비효용성을 가지게 된다. 적절한 채택률은 알고리즘의 효율성과 크게 관계있으며 Robert et al.(1994)은 약 0.45를 Gamerman(1997)은 0.2에서 0.5사이를 적절한 채택률로 제시한 바 있다. Fig. 4의 추정결과로부터 채택률을 산정해 본 결과, 구성된 알고리즘의 채택률은 0.42로써 기존 연구자들이 제시한 적절한 채택률과 유사한 것을 확인할 수 있었다.

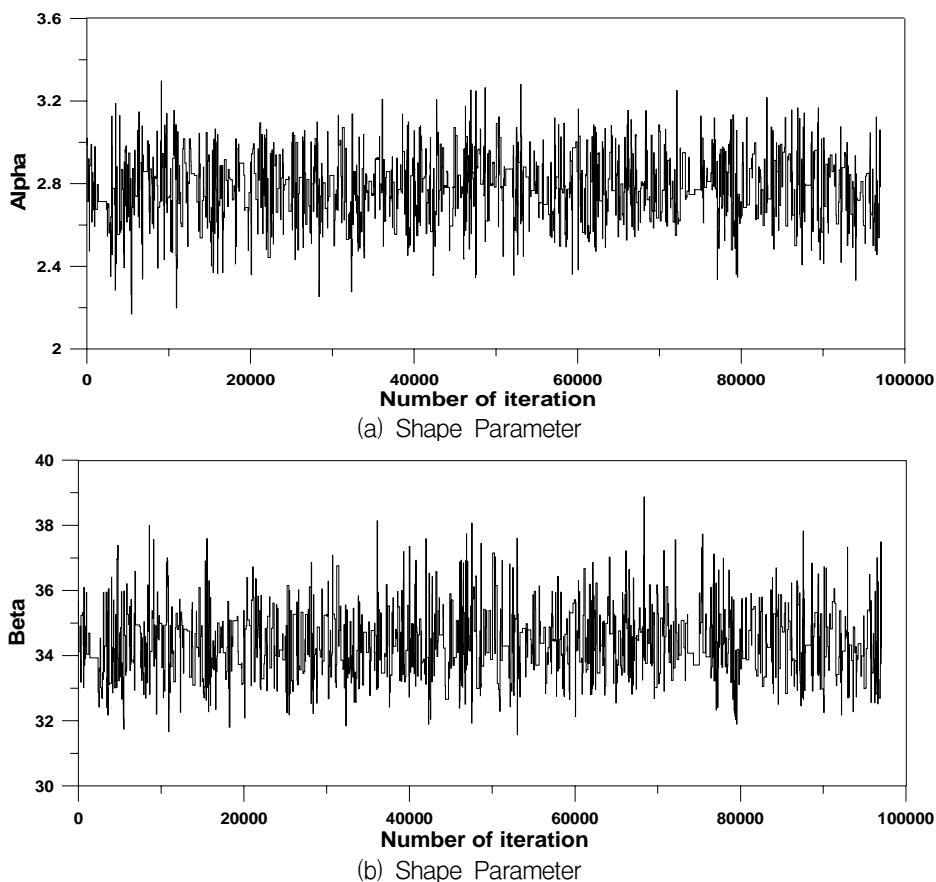


Fig. 4. Series of Parameters in Metropolis-Hastings Algorithm

7. Prior I 과 Prior II의 적합성 비교

본 절에서는 본 연구의 최종 목표인 Bayesian MCMC의 결과와 MLE 2차 근사방법을 사용한 결과를 비교함으로써 두 방법 간의 장, 단점을 분석하기 위한 사전 준비단계로서 위에서 제안된 두 가지의 사전분포 중 우수한 사전분포를 선정하였다. 우수한 사전분포를 선정은 사전에 알고 있는 참값(true parameter)로부터 필요한 자료를 생성한 후, 이 자료를 다시 각각의 사전분포를 이용한 Bayesian MCMC방법에 적용하여 최종결과를 비교하는 통계적 실험을 통해 수행될 수 있다. 본 연구에서는 이와 같은 통계적 실험을 위하여 진동지점의 36개 7Q값에 대해 최우추정법을 이용하여 추정된 $\alpha = 2.8371$ 과 $\beta = 34.4203$ 을 참값(true parameter)로 사용하였다. 이 값을 이용하여 2모수 Weibull 분포로부터 100개의 모의자료를 생성하였으며, 이용된 36개 자료에 대한 통계치와 모의된 100개의 자료에 대한 통계치를 비교하면 Table 3과 같다. 이 표로부터 모의된 자료가 실측자료의 통계치와 유사하게 모의된 것을 확인할 수 있으며, 모의된 100개의 자료를 이용함에 있어서 큰 문제가 없음을 확인하였으며 모의자료의 통계치는 자료의 길이가 100인 100개의 앙상블을 모의한 후 이들의 통계치들을 산술평균하여 나타낸 것이다.

모의된 100개의 자료를 이용하고 같은 우도함수, 같

은 제안분포, 두 가지 종류의 사전분포인 Prior I 과 Prior II를 사용하여 Metropolis-Hastings 알고리즘을 수행하였으며, 2.5 %, 평균, 97.5 %에 해당되는 각각의 모수의 추정결과들을 Table 4에 나타내었다.

Table 4를 보면 평균값에 있어서 Prior II가 Prior I 보다 참값을 추정하는 데 있어서 보다 정확하게 추정하였음을 알 수 있다. 또한 불확실성을 표현할 수 있는 97.5 %와 2.5 %의 차이는 α 가 Prior I, Prior II의 경우 각각 1.1061, 0.6613 이고 β 가 각각 10.882와 4.1224 로써 Prior II를 사용한 경우 불확실성 측면에서 많은 감소가 있었음을 알 수 있다. 추정된 모수의 2.5 %, 평균값, 97.5 %에 해당되는 모수를 이용하여 2모수 Weibull 확률밀도함수를 나타내면 Fig. 5와 같고 이 그림으로부터도 Prior II가 Prior I 보다 우수하다는 결론을 얻을 수 있었다.

그러므로, Bayesian MCMC방법을 이용하여 저수량 점 빈도분석을 수행하고자 하는 경우, 자료에 기반하지 않은 무정보적 사전분포보다는 자료에 기반한 사전분포를 사용하는 것이 Bayesian MCMC를 수행하는 데 있어서 합리적일 수 있다는 결론을 얻을 수 있었으며, 본 연구의 II편에서는 자료에 기반한 사전분포를 이용한 Bayesian MCMC 방법과 MLE 2차 근사방법의 추정결과를 비교함으로써 저수량 빈도분석에 있어서 불확실성을 보다 정확하게 분석하고자 하는 연구를 수행하였다.

Table 3. Statistical Properties of Each Data (at Jindong Station)

| Types of data | Mean | Standard deviation | Coefficient of variation | Coefficient of skewness | Coefficient of kurtosis |
|---|---------|--------------------|--------------------------|-------------------------|-------------------------|
| Synthetic data (100 years) with $\alpha = 2.8371$, $\beta = 34.4203$ | 32.6943 | 13.3770 | 0.4092 | 0.3235 | 2.5272 |
| Gauged data (36 years) | 32.8461 | 12.8604 | 0.3915 | 0.3788 | 2.8204 |

Table 4. Results of Bayesian MCMC with Prior I and Prior II

| Parameters | Prior I (Using non-data-based-prior) | | | Prior II (Using data-based-prior) | | |
|----------------------|--------------------------------------|---------|---------|-----------------------------------|---------|---------|
| | 2.5 % | Mean | 97.5 % | 2.5 % | Mean | 97.5 % |
| True alpha (=2.8371) | 1.9041 | 2.5012 | 3.0102 | 2.4798 | 2.7832 | 3.1411 |
| True beta (=34.4203) | 31.2311 | 36.7625 | 42.1131 | 32.8653 | 34.2711 | 36.9877 |

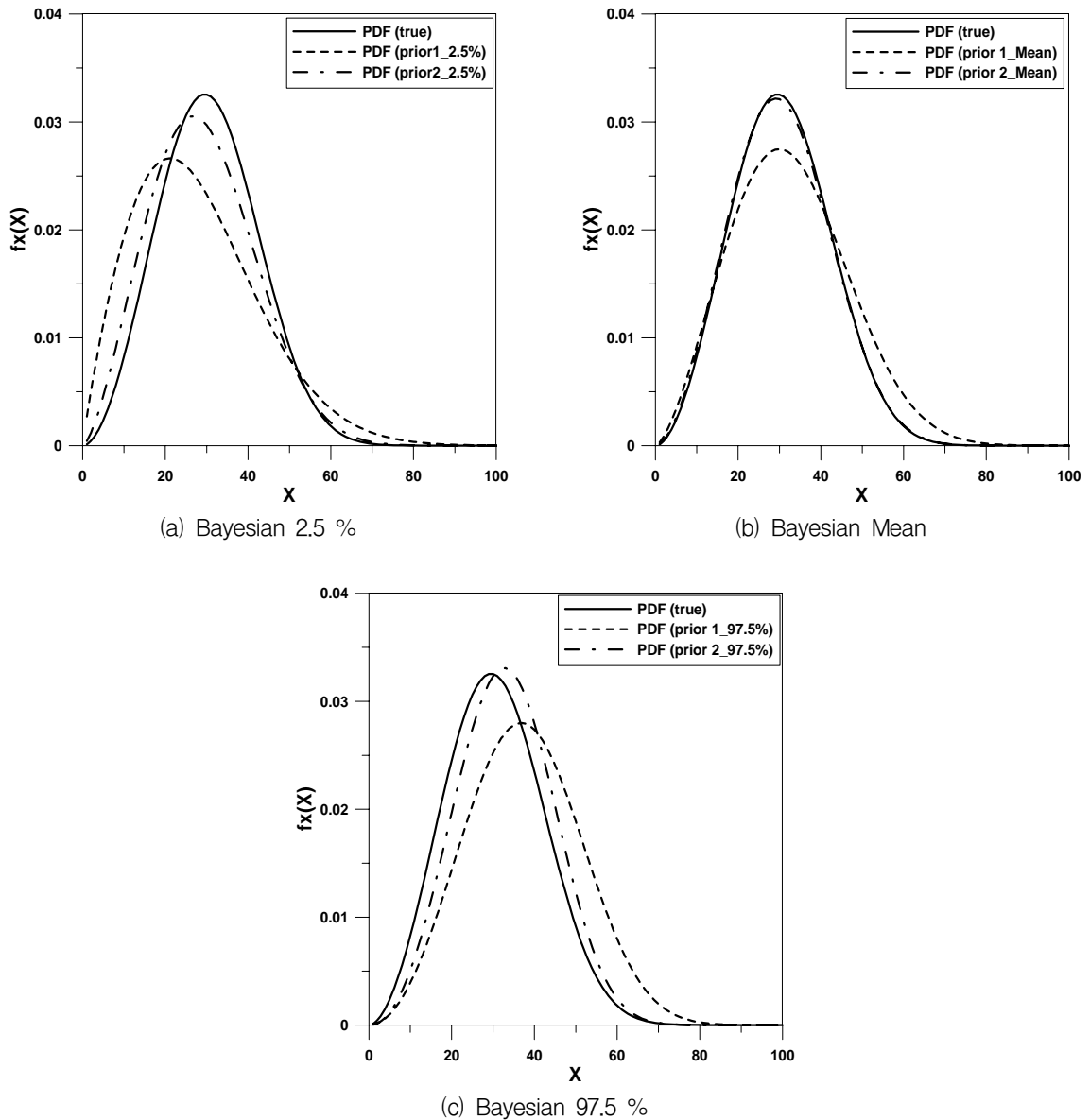


Fig. 5. Comparison with PDFs of Weibull Distribution

8. 결론 및 향후 연구

불확실성을 고려한 점 빈도분석의 수행 연구의 I 편에 해당되는 본 연구에서는 Bayesian MCMC 방법과 MLE 2차 근사방법을 이용하여 저수량 점 빈도분석에서 추정될 수 있는 불확실성의 표현을 위한 여러 가지 이론적 배경을 서술하였다. 또한 Bayesian MCMC 방법을 적용하는 데 있어서 가장 중요한 요소인 사전분포를 선정함에 있어서 자료에 기반하지 않은 사전분포와 에르고딕 가정을 이용한 자료에 기반한 사전분포를 구축하고, 두 가지 사전분포를 통계적 실험을 통하여 비교함으로써 자료에 기반한 사전분포가 평균값의 추정과 불확실성 측면에서 보다 나은 결과를 도출함을 확인하

였다. 그러므로 일반적으로 Bayesian 방법을 적용하는 경우에 계산을 간편하게 하기 위하여 공액사전분포를 사용하는 등 과거의 경험을 합리적으로 나타내지 못하는 사전분포를 선정하여 사용하는 것은 최종 결과가 현실을 제대로 반영하지 못 할 요소가 많으므로 가능하면 자료에 기반한 사전분포를 구축하여 사용하는 것이 합리적이라는 결론을 얻을 수 있었다. 본 연구의 II 편에서는 I 편에서 제시한 이론들과 선정된 사전분포를 사용하여 낙동강 유역에서의 저수량 점 빈도분석을 수행하였다. 또한 I 편에서 상세히 서술하지 못한 저수량 자료에 대한 한계성 부분과 적용방법을 추가적으로 설명하였다.

진행된 연구는 확률분포의 모수를 추정하는데 있어

서 발생하는 불확실성만을 대상으로 하였다. 그러나 향후 연구로써 1% 유의수준만을 만족시켰던 2모수 Weibull 분포보다 더욱 적절한 분포를 고려하여 연구를 수행하고 결과를 비교함으로써 분포의 선정에 따라 발생하는 빈도분석에 있어서의 불확실성도 함께 고려할 수 있으리라 판단된다.

감사의 글

본 연구는 21세기 프런티어 연구개발 사업인 수자원의 지속적 확보기술개발 사업단(과제번호 1-7-3)의 서울대학교 공학연구소를 통한 연구비 지원(30%)과 서울대학교 BK21 안전하고 지속가능한 사회기반건설사업단의 연구비 지원(70%)에 의해 수행되었습니다. 연구비 지원에 심심한 감사의 뜻을 표합니다.

참고 문헌

김상욱 (2007). Low flow frequency analysis using Bayesian approach. 박사학위논문, 서울대학교.

Bickel, P.J., and Doksum, K.A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, Inc., San Francisco, CA.

Chip, S., and Greenberg, E. (1995). "Understanding the Metropolis-Hastings algorithm." *Journal of the American Statistical Association*, Vol. 49, No. 4, pp. 327-335.

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.

Coles, S.G., and Powell, E.A. (1996). "Bayesian methods in extreme value modeling: A review and new developments." *International Statistical Review*, Vol. 64, No. 1, pp. 119-136.

Filliben, J.J. (1975). "The probability plot correlation coefficient test for normality." *Technometrics*, Vol. 17, No. 1, pp. 111-117.

Gamerman, D. (1997). *Markov Chain Monte Carlo-Stochastic Simulation for Bayesian*

Inference. Chapman & Hall, London, UK.

Gary, D.T. (1987). "A comparison of methods for estimating low flow characteristics of streams." *Water Resources Bulletin*, Vol. 23, No. 6, pp. 1077-1083.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis, Second Edition*. Chapman & Hall/CRC, New York, NY.

Gumbel, E.J. (1958). *Statistics of Extremes*. Columbia University Press, New York, NY.

Kavetski, D., Kuczera, G., and Fanks, S.W. (2006). "Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory." *Water Resources Research*, Vol. 42, W03407.

Kingston, G.B., Lambert, M.F., and Maier, H.R. (2005). "Bayesian training of artificial neural networks used for water resources modeling." *Water Resources Research*, Vol. 41, W12409, DOI: 10.1029/2005WR004152.

Kuczera, G. (1999). "Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference." *Water Resources Research*, Vol. 35, No. 5, pp. 1551-1557.

Kuczera, G., and Parent E. (1998). "Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm." *Journal of Hydrology*, Vol. 211, pp. 69-85.

Lee, K.S., and Kim, S.U. (2007). "Identification of uncertainty in low flow frequency analysis using Bayesian MCMC method." *Hydrological Processes*, In press(on-line published).

Madsen, H., and Rojsberg, H.D. (1997). "Generalized least squares and empirical Bayes estimation in regional partial duration series index flood modeling." *Water Resources Research*, Vol. 33, No. 4, pp. 771-781.

Meeker, W.Q., and Escobar, L.A. (1998). *Statistical Methods for Reliability Data*. John Wiley and Sons, New York, NY.

Metropolis, N., Rosenbluth, A.W., Teller A.H., and Teller E. (1953). "Equations of state calculations by fast computing machines." *Journal of Chemical Physics*, Vol. 21, pp. 1087-1092.

- Nathan, R.J., and McMahon, T.A. (1990). "Practical aspects of low flow frequency analysis." *Water Resources Research*, Vol. 26, No. 9, pp. 2135-2141.
- O'Connell, D.R.H., Ostenaar, D.A., Levish, D.R., Klinger, and R.E. (2002). "Bayesian flood frequency analysis with paleohydrologic bound data." *Water Resources Research*, Vol. 38, No. 5, pp. 1-14.
- Önöz, B., and Bayazit, M. (2001). "Power distribution for low streamflows." *Journal of Hydrologic Engineering*, Vol. 6, No. 5, pp. 429-435.
- Reis Jr., D.S., and Stedinger, J.R. (2005). "Bayesian MCMC flood frequency analysis with historical information." *Journal of Hydrology*, Vol. 313, pp. 97-116.
- Reis Jr., D.S., Stedinger, J.R., and Martins, E.S. (2005). "Bayesian generalized least squares regression with application to log Pearson type III regional skew estimation." *Water Resources Research*, Vol. 41, W10419.
- Roberts, G.O., Gelman, A., and Gilks, W.R. (1994). *Weak Convergence and Optimal Scaling of Random Walk Metropolis-Hastings Algorithms*. Technical Report, University of Cambridge.
- Seidou, O., Ouarda, T.B.M.J., Barbet, M., Bruneau, P., and Bobee, B. (2006). "A parametric Bayesian combination of local and regional information in flood frequency analysis." *Water Resources Research*, Vol. 42, W11408, DOI: 10.1029/2005WR004397.
- Stedinger, J.R. (1983). "Confidence intervals for design events." *Journal of Hydraulic Engineering*, Vol. 109, No. 1, pp. 13-27.
- Stedinger, J.R., Vogel, R.M., and Foufoula-Georgiou, E. (1993). *Frequency Analysis of Extreme Events*. In *Handbook of Hydrology*, Maidment, D.(eds). McGraw-Hill, New York, Chapter 18.
- Thiemann, M., Trosset, M., Gupta, H.V., and Sorooshian, S. (2001). "Bayesian recursive parameter estimation for hydrologic models." *Water Resources Research*, Vol. 37, No. 10, pp. 2521-2535.
- Vicens, G.J., Rodriguez-Iturbe, I., and Schaake Jr, J.C. (1975). "A Bayesian framework for the use of regional information in hydrology." *Water Resources Research*, Vol. 11, No. 3, pp. 405-414.
- Vrugt, J.A., Gupta, H.V., Bouten, W., and Sorooshian, S. (2003). "Shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters." *Water Resources Research*, Vol. 39, No. 8, SWC 1-16.
- Wang, Q.J. (2001). "A Bayesian joint probability approach for flood record augmentation." *Water Resources Research*, Vol. 37, No. 6, pp. 1707-1712.
- Wood, E.F., and Rodriguez-Iturbe, I. (1975a). "Bayesian inference and decision making for extreme hydrologic events." *Water Resources Research*, Vol. 11, No. 4, pp. 533-542.
- Wood, E.F., and Rodriguez-Iturbe, I. (1975b). A Bayesian approach to analyze uncertainty among flood frequency models. *Water Resources Research*, Vol. 11, No. 6, pp. 839-843.
- Zhang, B., and Govindaraju, R.S. (2000). "Prediction of watershed runoff using Bayesian concepts and modular neural networks." *Water Resources Research*, Vol. 3, pp. 753-762.

(논문번호:07-106/접수:2007.10.10/심사완료:2007.12.21)