

Bayesian MCMC를 이용한 저수량 점 빈도분석:

II. 빈도분석의 적용 및 결과의 평가

At-site Low Flow Frequency Analysis Using Bayesian MCMC:

II. Application and Comparative Studies

김상욱*, 이길성**, 김경태***

Sang Ug Kim, Kil Seong Lee, Kyungtae Kim

요 지

본 연구에서는 Bayesian MCMC 방법과 2차 근사식을 이용한 최우추정(Maximum Likelihood Estimation, MLE)방법을 이용하여 낙동강 유역의 본류지점인 낙동, 왜관, 고령교, 진동지점에 대한 점 빈도분석을 수행하고 그 결과로써 불확실성을 포함한 빈도곡선을 작성하였다. 통계적 실험을 통한 두 가지 추정방법의 분석을 위하여 먼저 자료의 길이가 100인 8개의 합성 유량자료 셋을 생성하여 비교 연구를 수행하였으며, 이를 자료길이 36인 실측 유량자료의 추정결과와 비교하였다.

Bayesian MCMC 방법에 의한 평균값과 2차 근사식을 이용한 최우추정방법에 의한 모드에서의 2모수 Weibull 분포의 모수 추정값은 비슷한 결과를 보였으나, 불확실성을 나타내는 하한값과 상한값의 차이는 Bayesian MCMC 방법이 2차 근사식을 이용한 최우추정방법보다 불확실성을 감소시켜 나타내는 것을 알 수 있었다. 또한 실측 유량자료를 이용한 결과, 2차 근사식을 이용한 최우추정방법의 경우 자료의 길이가 감소됨에 따라 불확실성의 범위가 합성 유량자료를 사용한 경우에 비해 상대적으로 증가되지만, Bayesian MCMC 방법의 경우에는 자료의 길이에 대한 영향이 거의 없다는 결론을 얻을 수 있었다. 그러므로 저수량 빈도분석을 수행하기 위해 충분한 자료를 확보할 수 없는 국내의 상황을 감안할 때, 위와 같은 결론으로부터 Bayesian MCMC 방법이 불확실성을 표현하는데 있어서 2차 근사식을 이용한 최우추정방법에 비해 합리적일 수 있다는 결론을 얻을 수 있었다.

핵심용어 : 불확실성, Bayesian MCMC, 2차 근사를 이용한 최우추정방법, 2모수 Weibull 확률 분포, 합성 유량자료, 실측 유량자료, 빈도곡선

1. 서 론

정량적으로 저수량을 분석하기 위해서 여러 가지 지표들이 사용될 수 있는데, 대표적인 것들이 MAR(Mean annual runoff), MDF(Mean daily flow), MF(Median flow), AMF(Absolute minimum flow) 등이다. 또한 위와 같은 개략적인 지표들 외에도 유황곡선의 작성이나 빈도분석의 결과를 이용해서도 여러 가지 저수량을 분석할 수 있는 결과를 얻을 수 있다. 현재까지 저수분석(Low flow analysis)에서 가장 많이 사용되는 값으로 미국에서는 빈도분석을 이용한 7일 지속기간 10년 빈도유량(7Q10)이고 국내에서는 355위 유량의 10년 빈도에 해당하는 유량(기준갈수량)이라 할 수 있다. 두 값을 산정하기 위해서는 빈도분석을 수행해야 하는데, 빈도분석은 주어진 자료를 나타낼 수 있는 확률분포함수의 선정, 모수의 추정, 추정된 모수를 이용한 재현기간별 분위수(Quantile) 산정의 절차를 통하여 통상적으로 산정된다. 즉 홍수량이나 저수량의 발생빈도에 해당되는 유량을 통계적으로 분석함으로써 필요한 재현기간에 해당되는 유량을 얻을 수 있는 방법이라 할 수 있다. 모수의 추정이 완료되면 주어진 재현기간에 대한 분위수를 추정하게 되며, 추정된 분위수에 대한 신뢰구간을 산정하고 재현기간에 따른 빈도곡선을 작성함으로써 원하는 재현기간에 따른 저수량 또

* 정희원 · 서울대학교 BK21 SIR 사업단 박사후 연구원 · E-mail : plethor1@snu.ac.kr

** 정희원 · 서울대학교 공과대학 건설환경공학부 교수 · E-mail : kilselee7@snu.ac.kr

*** 서울대학교 공과대학 건설환경공학부 석사과정 · E-mail : greattea@snu.ac.kr

는 홍수량을 구하여 사용할 수 있게 된다. 그러나 국내 대부분의 주요 중장기 계획은 빈도분석의 확정적인(Deterministic) 값을 이용하여 수립되고 있으며, 빈도분석결과의 불확실성을 반영한 확률적인(Probabilistic) 값이 이용되는 계획은 찾아보기 힘들다. 즉 대부분의 연구는 확률분포의 선정과 그에 따른 모수의 추정에만 초점이 맞추어져 있고 추정된 모수의 불확실성에 따르는 빈도유량을 산정하는 연구는 많지 않은데, 이는 불확실성에 대한 인식 부족과 함께 불확실성에 대한 계산방법이 현실을 제대로 반영하지 못함으로써 빈도분석결과의 정확성에 대한 신뢰도가 낮은 것에 기인한다. 특히 본 연구의 I 편의 연구동향에서 언급한 바와 같이 빈도유량의 불확실성에 대한 연구는 홍수량을 대상으로 하는 경우에는 진행된 사례가 있으나 저수량 빈도분석에 있어서 불확실성을 고려한 연구는 찾아보기 힘들다. 그러므로 본 연구에서는 불확실성을 계산하는데 있어서 기존에 사용된 정상성(Normality), 선형성(Linearity) 등의 가정조건을 사용하지 않는 Bayesian 방법론을 사용하여 저수량 빈도분석을 수행함으로써 기존 방법과 Bayesian 방법을 비교평가하고 개선점을 제시하고자 하였다.

2. 대상구역의 선정 및 자료의 선정

본 연구의 수행을 위하여 선정된 구역은 낙동강 유역(그림 1)으로서 본류상의 4개의 수위관측지점인 낙동, 왜관, 고령교, 진동지점의 실측 유량자료와 10개의 중권역의 유역 유출자료를 이용하였다. 점 빈도분석을 수행하기 위해서는 일반적으로 자료의 개수가 30개 이상이 되어야 함을 추천하고 있는 데(Rao and Hamed, 2000), 본 연구에서 사용되어지는 각 유역의 말단에 위치되어 있는 대부분의 수위관측소의 경우에는 자료의 기간이 약 20년 미만으로 점 빈도분석이 직접 수행되기에는 부족한 면이 있으므로 각 소유역별 저수량 빈도분석을 위해서는 점 빈도분석 외에 지역빈도분석을 수행하는 것이 타당할 수 있다. 그러므로 본 연구에서는 본류상에 위치한 낙동, 왜관, 고령교, 진동지점만을 대상으로 하여 1966년부터 2001년까지의 36개년에 해당하는 일 유량자료를 수집하고 점 빈도분석을 수행하였다. 저수량 빈도분석을 수행하고자 하는 경우 가장 먼저 산정해야 하는 것은 자연유량(Natural flow)이다. 그러나 국내 축적된 자료를 이용하여 임의 지점에서 자연유량을 산정하는 것은 필요 자료의 부족으로 인하여 올바른 자연유량이 산정되기 매우 어려운 실정이다. 이러한 문제를 해결하기 위해서 건교부와 한국수자원공사(2006)는 낙동강 유역의 유역조사사업을 수행하면서 낙동강 유역 각 중권역의 36년간 자연유량 자료를 PRMS(Precipitation-Runoff Modeling System)모형을 이용하여 모의한 바 있다. 모형을 이용한 자연유량 자료는 강우자료와 유역특성을 이용하여 원하는 기간만큼의 유량자료를 수문학적 모형을 통하여 모의할 수 있다는 장점이 있어 자연유량을 산정하기 힘든 국내 실정에 대한 대안으로서 사용될 수 있는 방법이라 할 수 있다.

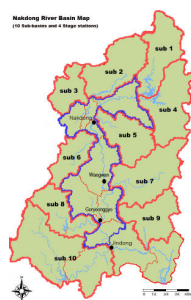
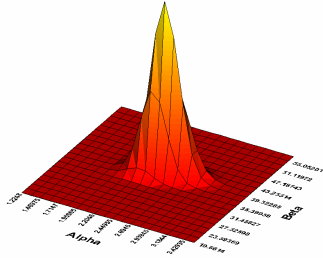


Figure 1. Nakdong river basin map

3. 2차 근사를 이용한 최우추정방법과 Bayesian MCMC 방법의 수행

본 연구에서는 2모수 Weibull 분포를 사용하고, 그 모수의 값으로써 8개의 다른 값을 사용하여 자료의 길이가 100인 7일 지속기간 최소유량(7Q)을 발생시킨 후, 각각의 자료 셋을 이용하여 Bayesian MCMC 방법과 2차 근사식을 이용한 최우추정방법을 이용하여 다시 자료의 모수의 2.5%, 평균, 97.5%에 해당되는 모수를 추정하였다. 본 연구의 I 편에서 언급한 바와 같이 최우추정(Maximum likelihood estimation, MLE)방법은 우도함수를 최대화하는 모수를 결정하는 최적화 문제이다. 이를 위해서는 먼저 우도함수의 형태 및 특성을 파

악하여 최우추정에 필요한 알고리즘을 검토할 필요가 있다. 선정된 확률밀도함수는 본 연구의 I 편에서 사용되었던 2모수 Weibull 분포(식 1)이다. 또한 이에 대한 우도함수를 다시 나타내면 다음 식과 같고, 이를 각각의 모수의 범위 내에서 그림으로 나타내면 그림 2와 같다. 아래 식에서 α 는 형상모수(Shape parameter)이고 β 는 축척모수(Scale parameter)이다. 또한 식 3은 Bayesian MCMC를 수행하기 위한 사전분포이다.



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

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

$$\pi(\alpha, \beta) = \frac{b}{a} \left(\frac{\beta}{a}\right)^{b-1} \exp\left[-\left(\left(\frac{\beta}{a}\right)^b + \lambda\alpha\right)\right] \quad (3)$$

Figure 2. Surface plot of likelihood function with shape parameter = 2.8371 and scale parameter = 34.4203 (여기서, λ, a, b 는 각각 추정된 0.98, 27.15, 1.92)

4. 합성유량자료를 이용한 결과의 분석

위에서 구축된 모형과 통계적 실험을 위하여 생성된 8개의 합성 유량자료 셋으로부터 얻어진 자료길이 100인 8개 셋의 7Q유량을 이용하여 Metropolis-Hastings 알고리즘을 100,000번 반복하여 사후분포로부터 모수를 샘플링하였다. 단, 알고리즘의 안정성을 위하여 최초 샘플링된 1,000개는 무시하고 99,000개의 추정치만을 이용하여 통계적 특성치를 산정하였다. 또한 Bayesian MCMC 방법의 비교 대안으로 최우추정법을 통해 얻어진 모수의 추정값과 이에 대한 신뢰구간을 산정하였다. 두 방법에 의한 결과값들을 비교해보면, 확정적인 값만을 필요로 하는 경우에는 최우추정방법보다 적용 절차가 복잡한 Bayesian MCMC 방법을 사용하는 것이 큰 장점을 가지지 못함을 알 수 있다. 그러나 확실적인 개념이 필요한 경우, 즉 불확실성을 고려한 모수의 추정은 두 가지 방법에 의한 결과로부터 Bayesian MCMC 방법이 우수하다는 결론을 얻을 수 있다. 즉 Bayesian MCMC 방법을 사용하는 경우 모수의 불확실성을 나타내는 범위가 감소되어져 나타나는 것을 알 수 있다(그림 3과 그림4).

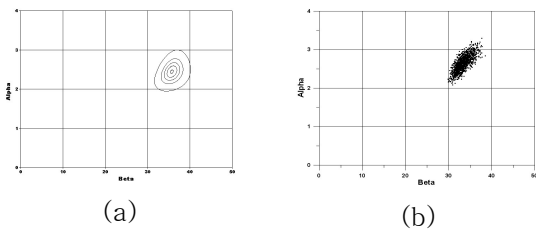


Figure 3. Contour plots of MLE and Bayesian MCMC(Case: set 7)
(a) MLE, (b) Bayesian MCMC

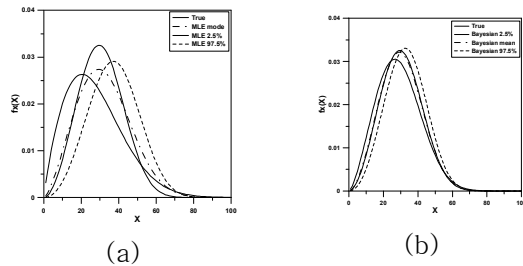


Figure 4. Probability density function with MLE and Bayesian MCMC
(a) MLE, (b) Bayesian MCMC

5. 실측유량자료를 이용한 결과의 분석

낙동, 왜관, 고령교, 진동의 4개 지점의 실측유량을 이용하여 4절에서 수행한 것과 같은 과정을 다시 수행하였다. Bickel and Doksum(1977)은 2차 근사식을 이용한 신뢰구간식을 제안하면서 자료의 길이가 감소되는 경우 신뢰구간에 의해 추정된 하한값과 상한값의 정확성이 감소되어 불확실성 측면에서 과대추정될 수 있다는 결론을 제시한 바 있는데, 실측유량을 이용하여 최우추정방법을 수행한 결과는 이와 일치되는 결과를 보여주었다. 즉, 합성유량자료보다 기간이 짧으므로 불확실성이 크게 나타나는 것을 알 수 있었다. 이에 비해 Bayesian MCMC 방법의 결과는 자료의 길이가 감소되어도 2차 근사식을 이용한 최우추정방법보다 불확실성

이 크게 증가되지 않는 것을 알 수 있다. 이로부터 Bickel and Doksum(1977)이 제시한 결과와 같은 결과를 얻을 수 있었으며, 특히 국내에서 저수량 빈도분석을 수행하고자 하는 경우에는 대부분의 지점에서 자료의 길이가 짧은 한계성이 존재하므로 2차 근사식을 이용한 최우추정방법보다 Bayesian MCMC 방법을 사용하는 것이 저수량 빈도분석에서 존재하는 불확실성을 표현하는 데 있어서 합리적일 수 있다는 결론을 얻을 수 있었다. 위에서 제시한 실측 유량자료에 대한 Bayesian MCMC 방법과 2차 근사식을 이용한 최우추정방법을 낙동, 왜관, 고령교 지점에도 같은 과정으로 적용하였다. 단, 사전분포는 진동지점을 대상으로 하는 자료에 기반한 분포만을 타 지점에도 사용하였으며, 각 지점에서의 자료에 기반한 사전분포를 구축하여 점 빈도분석을 Bayesian MCMC 방법으로 수행하는 과정은 추가적인 연구를 통하여 수행하고 그 결과를 비교함으로써 또 다른 의미 있는 결론을 얻을 수 있을 것으로 판단된다. 최종적으로 두 가지 추정방법에 의해 얻어진 모수의 2.5 %, 평균값(또는 모드값), 97.5 %에 해당되는 모수를 이용하여 낙동, 왜관, 고령교, 진동지점에서의 1년부터 100년의 재현기간에 해당되는 분위수를 추정하였다. 그림 5에는 각각 낙동, 왜관, 고령교, 진동지점에 대하여 최종적으로 구축된 불확실성을 포함한 빈도곡선을 나타내며 모든 지점에서 실선으로 나타낸 그림이 Bayesian MCMC 방법이고 점선으로 나타낸 그림이 2차 근사식을 이용한 최우추정방법에 의한 결과로서 각각 아래로부터 2.5 %, 평균, 97.5 %를 나타낸다. 4개의 모든 지점에서 평균값에 대한 빈도곡선의 차이는 두 방법에 있어서 큰 차이를 나타내지 않지만, 불확실성 측면에서는 Bayesian MCMC 방법이 불확실성을 감소시켜 나타낼 수 있음을 확인할 수 있었다.

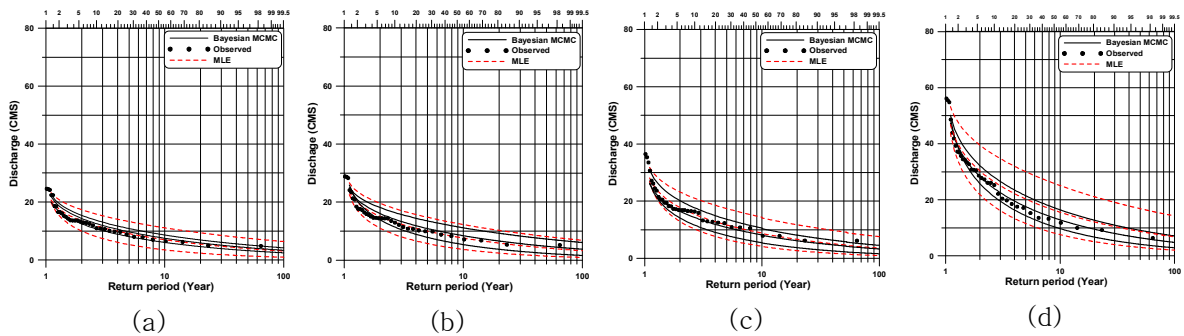


Figure 5. Low flow frequency curves considering uncertainty

6. 결론

본 연구는 Bayesian 방법을 이용하여 저수량 빈도분석을 수행하기 위한 연구로써 I 편에서 수행된 이론적 배경과 Bayesian MCMC 방법과 2차 근사식을 이용한 최우추정방법을 이용하기 위해 구축된 자료에 기반한 사전분포, 제한분포, 신뢰구간의 산정식 등을 이용하여 낙동강 유역의 본류지점인 낙동, 왜관, 고령교, 진동지점에 대한 점 빈도분석을 수행하고 그 결과로써 불확실성을 포함한 빈도곡선을 작성하였다.

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.

참고문헌

- 건교부, 한국수자원공사 (2006). **낙동강 유역조사사업 보고서**
- Bickel, P.J., and Doksum, K.A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day, Inc., San Francisco, CA.
- Rao, A.R., and Hamed, K.H. (2000). *Flood Frequency Analysis*. CRC Press, Boca Raton, FL.