

극치수문자료의 경향성 분석 개념 및 비정상성 빈도해석

Concept of Trend Analysis of Hydrologic Extreme Variables and Nonstationary Frequency Analysis

이정주* · 권현한** · 김태웅***

Lee, Jeong-Ju · Kwon, Hyun-Han · Kim, Tae-Woong

Abstracts

This study introduced a Bayesian based frequency analysis in which the statistical trend analysis for hydrologic extreme series is incorporated. The proposed model employed Gumbel extreme distribution to characterize extreme events and a fully coupled bayesian frequency model was finally utilized to estimate design rainfalls in Seoul. Posterior distributions of the model parameters in both Gumbel distribution and trend analysis were updated through Markov Chain Monte Carlo Simulation mainly utilizing Gibbs sampler. This study proposed a way to make use of nonstationary frequency model for dynamic risk analysis, and showed an increase of hydrologic risk with time varying probability density functions. The proposed study showed advantage in assessing statistical significance of parameters associated with trend analysis through statistical inference utilizing derived posterior distributions.

Keywords : hydrologic extreme variable, trend analysis, nonstationary, bayesian model

요 지

본 논문에서는 극치수문자료의 경향성 분석 개념을 소개하고 이를 빈도해석과 연계시켜 해석하는 방법론을 제시하고자 Gumbel 극치분포를 기반으로, 시간변화에 의한 수문빈도 특성 변화를 모의할 수 있는 Bayesian 모형을 구성하였다. 사후분포의 매개변수는 깁스표본법에 의한 Markov Chain Monte Carlo Simulation을 통해 추정하였으며, 이를 통해 경향성을 고려한 확률강우량과 불확실성 구간을 추정하였다. 또한 경향성을 고려한 확률강우량이 현재 알려진 확률강우량을 초과할 확률을 통해 동적 위험도 해석과정을 소개하였으며, 현재의 경향성에 대해서 시간에 따라 연속으로 추정된 확률밀도함수를 비교하여 수문학적 위험도가 증가할 수 있음을 모의결과를 통해 확인하였다. 이와 더불어 단순히 경향성의 존재여부를 확인하는데 그치지 않고 사후분포를 통해서 통계적 추론을 수행함으로써 경향성에 대한 통계학적인 유의성을 정량적으로 평가할 수 있었다.

핵심용어 : 극치수문변량, 경향성분석, 비정상성, Bayesian 모형

1. 서 론

최근 지구온난화가 주원인으로 작용하는 기후변화의 가속화로 인해 기후변화 문제는 전 지구적인 주요 현안으로 인식되고 있으며, IPCC(Intergovernmental Panel on Climate Change)와 WMO(World Meteorological Organization)로 대표되는 국제협의기구를 통해 기후변화 문제에 대한 범정부 차원의 조사 및 대책수립과 미래 위험도 저감방안을 모색하고 실천하는 단계에 들어서 있다. 또한, 엘니뇨(El Niño)로 대표되는 기상변동성(climate variability)과 지구온난화로 인한 이상기후현상의 빈번한 발생으로, 이로부터 자유로울 수 없는 미래 수문환경의 변화 양상과 구체적인 대응

수준에 대한 관심과 요구가 급증하고 있다. 기후변화에 의한 수문량의 경향성 증가와, 이상기후에 의한 극치사상의 변동폭 확대 등은 안전을 담보로 하는 치수정책 및 기준의 변화와 안정적 용수 공급을 위해 확보해야 할 수자원의 규모를 결정짓는데 있어서 기본적으로 고려해야 할 사항일 것이다.

국내외적으로 극치사상의 발생빈도 및 양적 증가에 대한 연구는 기후변화와 맞물려 여러 분야에서 활발히 진행되고 있다. 기후변화에 의한 수문환경의 미래 변동성 예측은 시나리오 기반의 GCM(General Circulation Model), RCM(Regional Climate Model)을 이용한 방법과 과거자료의 통계적 해석을 기반으로 한 통계학적 방법으로 나눌 수 있다. 대부분의 통계학적 연구는 극치자료에 대해서 선형회귀분석, Mann-

*정회원 · 전북대학교 토목공학과 박사수료 (E-mail : julee@jbnu.ac.kr)

**정회원 · 교신저자 · 전북대학교 토목공학과 조교수 (E-mail : hkwon@jbnu.ac.kr)

***정회원 · 한양대학교 건설환경공학과 조교수 (E-mail : twkim72@hanyang.ac.kr)

Kendall 경향성 분석 등을 적용하여 경향성의 존재 여부와 경향성의 유의성을 검증하는 것을 목적으로 하고 있다. 그러나 극치수문자료는 일반적인 수문변량과는 다른 통계학적 특성을 가지고 있어 단순히 선형회귀분석과 같은 방법으로는 극치수문량의 경향성을 효과적으로 분석하기 어렵고 이를 수자원계획의 중요한 요소인 빈도개념과 연계시켜 해석하는데 무리가 있다고 지적되고 있다(Katz 등, 2002). 이러한 점에서 본 연구는 극치수문자료의 분포특성을 고려할 수 있는 경향성 분석 방법과 이를 빈도해석과 연계시켜 해석할 수 있는 비정상성 빈도해석(nonstationary frequency analysis) 모형을 소개하고 이를 국내 수문자료에 적용하는데 주안을 두고 있다.

기상은 연속적으로 변화하고 있으며 변화의 증거들은 장기간의 수문기상학적 자료들의 패턴을 분석한 연구들에 의해 증명되고 있다(Knox, 1993). 특히 경년변화를 갖는 ENSO (El Niño-Southern Oscillation)는 수문학적 변량에 가장 두드러진 영향을 주는 기상변동성의 예로서 전 세계적으로 폭넓은 연구가 진행되고 있다(Halpert와 Ropelewski, 1992; Piechota와 Dracup, 1996; Ropelewski와 Halpert, 1996; Ropelewski 등, 1992). 기존 연구들은 기상변동성이 어떠한 방식으로 극한 수문사상의 빈도를 변화시키는지 입증해오고 있으며(Franks와 Kuczera, 2002; Jain과 Lall, 2000, 2001; Milly 등, 2002; Pizaro와 Lall, 2002; Sankarasubramanian과 Lall, 2003), Porparto와 Ridolfi(1998)는 최근 비정상성시계열(nonstationary time series)에서 뚜렷한 변동성이 재현될 수 있음을 설명하였다. 극치수문량의 스펙트럼 분석에서 주기변동성의 존재는 빈도를 유추하는데 있어서 수문변량들을 독립(independent)으로 간주하는 것이 부적절 할 수 있음을 보여주는 하나의 예가 될 수 있다. 이러한 사실들은 한편으로 위험도에 대한 예측 가능성을 시사하고 있으며, 극치사상 분석을 통한 위험도 관리를 보다 효과적으로 할 수 있는 개연성을 부여한다 하겠다.

일반적으로 경향성을 파악하는데 많이 이용되는 선형회귀 분석에서 변수의 선택문제는 최적의 모형을 찾는 데 있어서 중요한 부분을 차지한다. 변수 x 와 y 간에 직선회귀모형(linear regression model)을 적합시킬 경우에는 일반적으로 몇 가지 가정이 전제조건을 이룬다(박성현, 1997). 그중에서 극치수문자료의 경향성 해석과 관련하여 정규분포 가정과 등분산(homoscedastic)의 가정을 유념할 필요가 있다.

Katz 등(2002)은 최우도추정(maximum likelihood)에 의한 GEV 분포형 적합 과정에서, 수문학적 변수들의 분포는 Heavy Tail을 가지고 있는 것에 대한 여러 사례들을 언급하였다. 유출량의 경우에 있어서 나타나는 Heavy Tail 특성과 Anderson과 Meerschaert, 1998; Farquharson 등, 1992; Madsen 등, 1997; Morrison과 Smith, 2002) 더불어 극치강수량(시단위, 일단위)의 경우도 그 분포형이 Heavy Tail을 갖는 것으로 알려져 있다(Egozcue와 Ramis, 2001). x , y 평면에서의 최소제곱법 개념으로 단순회귀분석을 통해 극치강수량의 시간변화에 따른 확률분포의 경향성을 분석할 경우 분석자료가 극치분포를 따르게 되기 때문에 정규분포 가정과 등분산의 가정을 충족하지 못하며, 수문학적 빈도해석을 위한 극치분포특성을 충족시키지 못하는 문제가 있다. 극

치강수량 및 홍수 등에 대한 경향성 분석과 관련하여 국내 연구사례를 살펴보면, 이동률 등(2004)은 GCM 모의결과와 마코프 연쇄모형으로 모의된 수문자료를 바탕으로 기뎀지수 등의 변화를 모의하였고, 정대일 등(2008)은 Mann-Kendall 검정을 통해 극치강수량과 홍수량의 경향성을 판별하고, 대수정규 추세모형을 이용하여 설계홍수량을 산정하였다. 김보경 등(2008)은 강수량과 기온자료로부터 추출한 극한지수를 이용하여 극한 기상사상의 통계적 변화 경향을 분석하였으며, 권영문 등(2009)은 강우량의 증가 경향성을 고려한 확률 강우량을 산정하기 위해 극치강수량의 이동평균을 이용한 빈도해석을 통해 경향성을 반영한 사례가 있다.

Reis와 Stedinger(2005)가 제시한 대수정규분포는 홍수량에 대해서 적합한 방법으로 알려져 있으나 우리나라의 경우 빈도해석을 위한 자연유량을 취득하기가 현실적으로 어렵기 때문에 대수정규 추세모형을 홍수량에 적용하는 것은 현실적으로 용이하지 않으며 강수량에 대한 빈도해석이 일반적이다. 이러한 점에서 강수량에 대해서 보편적으로 적용 가능한 분포형 선정 및 이를 이용한 방법론 개발이 필요하다. 수문자료에서 비정상성이란 통계적 특성이 시간에 따라 변화하는 것으로 최근에 나타나는 이상기후, 기후변화, 기상변동 등과 같은 기상학적 원인에 의해서 나타난다. 비정상성 빈도해석은 이용되는 입력 자료에 따라서 다양한 목적으로 활용이 가능하다. 첫째, 극치자료에 대한 변동성을 Simulation을 통해 확률적으로 추론할 수 있다. 즉, 정확한 값을 추정하기 보다는 미래에 대한 변동성을 추론할 때 이용가능하다. 기후변화 시나리오를 활용하여 기후변화에 따른 극치수문량 전망에 적합하다(권현한 등, 2010). 둘째, 단기 예측 모형에 적용이 가능하다. 이러한 경우 사용되는 입력자료는 GCM 단기 예측결과, 해수면온도, 해수면기압 등과 같은 기상상태 변량 등이 활용되게 된다(Sankarasubramanian과 Lall, 2003; Kwon 등, 2008; Kwon 등, 2009).

본 논문에서는 극대치강수량자료의 경향성 평가를 위해서 극치분포를 활용한 경향성 평가 방법을 제시하고자 하며 이를 해석하기 위한 방법론으로 Bayesian 모형을 활용하였다. 본 연구에서 제시하는 방법론은 앞서 나열한 비정상성 빈도해석의 활용을 위한 기본적인 해석을 가능케 하는 역할을 하게 된다. 본 논문에서 Bayesian 모형은 극치분포의 매개변수 및 경향성 분석의 회귀계수 들의 사후분포(posterior distribution)를 유도하는데 이용되게 된다. Bayesian 방법은 통계추론에서 발생할 수 있는 계산 문제를 효율적으로 해결하는 방법으로 빈번히 사용되고 있다. Bayesian 모형을 기반으로 한 극치수문자료의 해석과 관련한 연구로, 권현한과 문영일(2007)은 Bayesian 로지스틱 모형을 이용하여 극치여름강수량에 대해서 태풍 및 기상인자 등의 외부 인자를 고려한 10년 빈도 강수량의 발생확률을 추정하였다. 또한, Kwon 등(2008)은 Montana 지역의 연최대치 홍수량자료 계열에 대해서 계층적 Bayesian 방법을 이용하여 홍수빈도분포의 매개변수를 동역학적으로 추정하였으며, 이에 대한 정량적인 불확실성 분석과 시간에 따라 변동하는 홍수량을 추정하였다.

앞서 언급되었듯이 본 논문은 극치자료의 비정상성 유무를 검증하는데 주안점을 두기 보다는 극치자료의 경향성 분석

과 비정상성해석과정을 연계하는 방법론을 제시하는데 목적을 두고 있음을 밝힌다. 본 논문에서는 첫째, 대상자료 및 Bayesian 모형에 대해서 간략히 설명하고 비정상성 빈도해석 모형의 사후분포 추정 및 해석과정을 설명한다. 둘째, 서울지방의 시단위 강수량자료로부터 추출한 24시간 임의지속시간 연최대강수량을 대상으로 모형의 구성, 통계학적 추론, 결과 활용 등 다양한 관점에서 해석과정을 설명한다. 마지막으로 해석결과를 요약하고 향후 연구방향을 제시하는 단계로 구성되었다.

2. 대상 자료 및 연구 방법

2.1 대상 자료

본 연구에서는 분석을 위한 극치수문사상을 추출하기 위해 서울지점의 시단위 강수량자료를 이용하였다. 시단위 강수량 자료로부터 24시간 지속 이동시간 최대강수량을 추출하여 연최대강수량계열(Annual maxima rainfall) 자료를 구축하였다. 자료계열의 기간은 1961년부터 2008년까지 48년간이다. 본 연구에서는 우선 방법론에 치중하기 위해 서울지점의 강수량만을 이용하였으며, 대상 수문사상에 적합한 확률분포를 판단한 후 분석에 이용하였다.

2.2 Markov Chain Monte Carlo 모의

수리적으로 계산이 불가능하거나 복잡한 적분, 추정 등의 문제에 사용되는 Monte Carlo 기법은 최근에 수리 수문학 분야에서 위험도 및 불확실성을 평가하는 수단으로 널리 이용되고 있다(Kwon과 Moon, 2006). Monte Carlo 기법은 관심대상인 값을 확률변수의 기댓값으로 표현하고 이를 모의를 통하여 추출된 동일한 분포를 따르며 서로 독립인 (independent and identically distributed: iid) 표본들의 표본평균을 이용하여 추정하는 방법이라고 할 수 있다.

이에 반해 Markov Chain Monte Carlo(MCMC) 기법은 주어진 다변량 확률분포가 복잡하여 이를 따르는 iid 난수를 얻을 수 없는 경우에 사용가능한 기법으로서 iid 난수 대신 Markov Chain 난수를 추출하여 사용한다. Markov Chain을 통해 난수를 발생시킨다고 해서 정확하게 관심이 되는 확률 분포를 따르지 않지만 이를 일정 시간동안 반복 후에 얻어지는 난수들은 추출을 원하는 분포에 수렴하게 된다. 따라서 MCMC 기법은 복잡한 다변량 확률분포 및 매개변수의 추정을 요하는 문제에서 주로 사용되며 또한 Bayesian 통계 기법에서 사후분포의 추론에 이용될 수 있다. 본 연구에서는 MCMC 기법을 이용하여 종속변수 Y 에 대해서 조건부 분포를 갖는 각 독립변수의 사후분포를 추정하게 된다.

MCMC 기법의 대표적인 방법으로 메트로폴리스-헤스팅 알고리즘(Metropolis-Hastings algorithm)과 깁스표본법(Gibbs sampling) 등이 있으며, 본 연구에서는 깁스표본법을 이용하였다. 깁스표본법은 확률밀도함수로부터 직접 표본을 추출할 수는 없으나 각각의 변수들의 대해서 다른 두 변수들이 주어졌을 때의 조건부 분포가 알려져 있고 이로부터의 표본추출이 가능한 경우에 사용할 수 있다. 깁스표본법의 알고리즘을 간단한 예를 통하여 설명하면 아래와 같다. 2개의 변수를 갖는 다변량 확률밀도함수를 $f(\theta_1, \theta_2)$ 라고 하자.

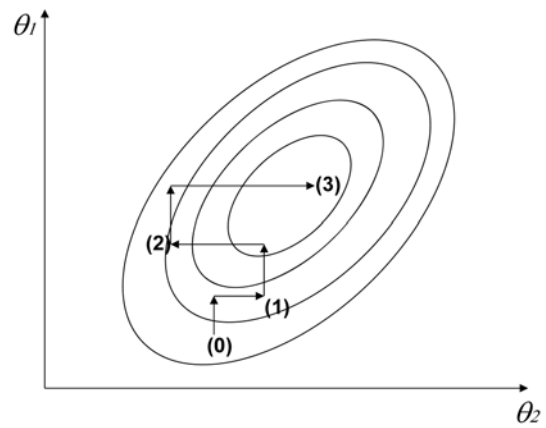


Fig. 1 A Basic scheme for Bayesian Markov Chain Monte Carlo simulation with two parameters using Gibbs sampling technique

[1] 두 변수에 대한 초기 값 $(\theta_1^{(0)}, \theta_2^{(0)})$ 을 부여한다.

[2] 번째 난수 벡터 $(\theta_1^{(i)}, \theta_2^{(i)})$ 가 주어졌을 때 $i+1$ 번째 난수 벡터를 다음과 같은 조건부 분포에서 추출한다.

$$(1) \theta_1^{i+1} \sim f(\theta_1 | \theta_2^{(i)})$$

$$(2) \theta_2^{i+1} \sim f(\theta_2 | \theta_1^{i+1})$$

[3] 위의 과정을 충분히 반복한 후 초기의 일정부분 난수를 제거한 이후의 난수들을 이용한다. 이러한 제거과정을 Burning이라고 하며 Bayesian 해석에서 중요한 구성단계이다 (George와 McCulloch, 1993).

위의 알고리즘에서 보듯이 깁스표본법은 조건부 분포들에서 조건으로 주어지는 변수들의 값을 정확하게 바로 직전의 단계에서 주어진 값들이 사용되게 되며 따라서 조건부 분포에서 추출된 난수들이 안정 상태에 도달하는 것이 주어진 다변량 확률분포를 정확히 따르는 난수가 되는 척도가 되며 깁스표본법을 구현하는데 가장 중요한 부분이 된다. 이러한 표본추출과정을 도식적으로 나타낸 것이 Fig. 1이다. Fig. 1에서 (0)으로 표시된 위치를 초기 가정값이라고 하면, 결합 확률밀도(joint probability distribution)를 최대로 하는 θ_1 을 찾은 후 그 값을 토대로 θ_2 를 찾게 된다(Fig. 1의 (1)번 위치). 이 과정이 한번의 Iteration이 되며, 갱신된 위치를 토대로 앞의 과정을 반복하여 최종적으로 매개변수의 사후분포를 추정하게 된다. 본 연구에서는 10000회의 burning 과정을 거쳐 난수를 안정화시키고 2000회의 샘플링을 하였다.

2.3 경향성을 고려한 비정상성 빈도해석

최근 우리나라를 포함하는 여러 나라에서 극치사상 발생빈도의 증가 양상이 나타나고 있으며, 이를 고려하여 미래에 발생 가능한 극치사상에 대한 분포를 추정하고자 하는 연구가 진행되고 있다. 그러나 방법론적으로 기존 극치분포에서 외부 인자를 고려하는 것은 매개변수 추정 등에 있어서 많은 어려움이 있으며, 특히 증가 및 감소 경향에 대한 불확실성을 동시에 고려하는 것은 더욱 이러한 어려움을 가중시키고 있다. 가장 근본적인 질문은 이러한 구조적인 빈도의 경향성이 존재하는가에 대한 여부라 할 수 있겠다. 지금까지의 연구들에 의하면 온실가스 효과로 인한 지구온난화는 수문 Cycle를 가속화 시키는 방향으로 진행될 가능성이 크다고 지적하고 있으며 이러한 시점에서 극치계열에 대한 경향

성을 고려할 수 있는 모형의 개발은 시기적절하다 하겠다. 본 연구에서는 극대치강수량계열의 비정상성 빈도해석을 위해 우리나라에서 주로 사용되는 Gumbel 분포를 이용하였다. Gumbel 분포의 확률밀도함수와 누적확률밀도함수는 다음과 같다.

$$f(x) = \frac{1}{\sigma} \exp\left[-\frac{(x-\mu)}{\sigma} - \exp\left[-\frac{(x-\mu)}{\sigma}\right]\right] \quad (1)$$

$$F(x) = \exp\left\{\exp\left[-\frac{(x-\mu)}{\sigma}\right]\right\} \quad (2)$$

여기서 μ 는 위치매개변수(location parameter)이고, σ 는 규모매개변수(scale parameter)이다.

극치사상에 나타나는 경향성을 처리하기 위한 방법으로 본 연구에서는 다음과 같은 모형을 제시하고자 한다. 우선, 극치계열 자료 Y 가 앞서 제시한 Gumbel 분포를 따른다고 가정할 때, Gumbel 분포형의 위치매개변수와 규모매개변수는 다음과 같이 시간에 따라 다른 값을 가지도록 구성할 수 있다.

$$Y \sim \text{Gumbel}(\mu(t), \sigma(t)) \quad (3)$$

Bayesian 모형에서, 식 (3)의 모든 매개변수들은 확률분포를 가지게 된다. 즉, $\mu(t)$ 와 $\sigma(t)$ 가 사전분포를 가지게 되며 각 매개변수는 시간의 함수로서 다음의 식 (4)와 식 (5)와 같이 선형관계를 갖는다고 가정한다.

$$p(\mu(t)) \sim N(\alpha_0 + \alpha_1 \times t, \sigma_\alpha^2) \quad (4)$$

$$p(\sigma(t)) \sim N(\beta_0 + \beta_1 \times t, \sigma_\beta^2) \quad (5)$$

여기서, t 는 시간을 의미하며, α 와 β 는 시간과 매개변수 μ 와 σ 의 사전분포를 선형적으로 연계시키는 회귀분석의 매개변수를 나타낸다. Fig. 2는 확률분포의 위치매개변수와 규모매개변수의 변화 조건에 따른 거동을 나타내고 있으며 본 연구에서는 시간에 따라 2개의 매개변수가 동시에 변화한다고 가정하여 통계적 추론(statistical inference)을 실시하였다.

시간에 따른 종속성이 선형이나 비선형(nonlinear)이라는 논란의 여지가 있을 수 있다. 본 연구에서는 이러한 점을 부각하기 보다는 방법론의 전개 및 해석 방법에 대한 개념을 제시하는데 주안점을 두었다. 그러나 자료가 비선형의 경

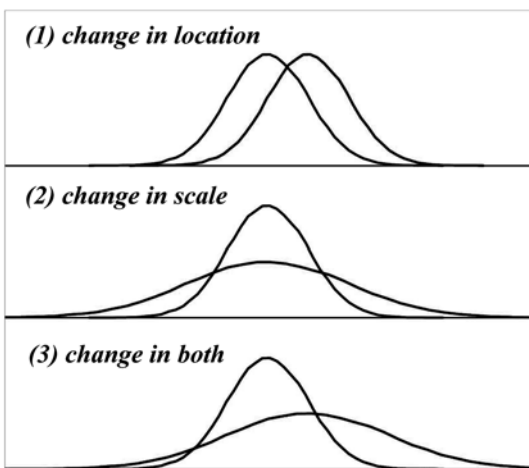


Fig. 2 Changes in location and scale parameters and associated probability density function

향성을 나타낼 경우에도 식 (4)와 식 (5)를 비선형 모형으로 구성함으로써 효과적으로 비선형성을 고려할 수 있다. 본 연구에서는 48년간의 극대치강수량 자료를 대상으로 하며 총 6개의 매개변수의 사후분포를 추정하는 것이 최종 목적이다. 48년의 자료는 6개의 매개변수를 추정하는데 무리가 없으며 4개의 회귀매개변수는 일반적인 경우에 해당하는 정규분포로 가정하였다(Gelman 등, 2004). 정규분포로 가정한 이유는 증가경향과 감소경향에 의한 변화를 모두 고려하기 위해서이며 회귀계수를 정규분포로 가정하더라도, 추정된 결합확률분포(joint distribution)는 Gumbel 분포를 따르게 된다. α 와 β 의 분산 σ_α^2 와 σ_β^2 는 양의 값을 가져야 하므로 Gamma 분포로 가정하였으며 이를 식 (6)~식 (9)와 같이 나타낼 수 있다. 결과적으로 6개의 매개변수에 대한 결합확률(joint probability)은 식 (10)과 같다.

$$\alpha_n \sim N(\mu_{\alpha_n}, \sigma_{\alpha_n}^2) \quad n = 0, 1 \quad (6)$$

$$\beta_n \sim N(\mu_{\beta_n}, \sigma_{\beta_n}^2) \quad n = 0, 1 \quad (7)$$

$$\sigma_n \sim G(k_\alpha, s_\alpha) \quad (8)$$

$$\sigma_\beta \sim G(k_\beta, s_\beta) \quad (9)$$

$$p(\alpha_0, \alpha_1, \beta_0, \beta_1, \sigma_\alpha, \sigma_\beta) \propto 1 \quad (10)$$

Bayes 정리에 의해서 매개변수의 사후분포는 다음과 같이 전개될 수 있다.

$$p(\Theta|y) = \frac{p(\Theta, y)}{p(y)} \propto p(y|\Theta) \cdot p(\Theta) \quad (11)$$

여기서 $\Theta = [\mu(t), \sigma(t), \alpha_0, \alpha_1, \beta_0, \beta_1, \sigma_\alpha, \sigma_\beta]$ 는 전체매개변수들의 집합을 나타내며, $p(\Theta)$ 는 매개변수들의 사전분포를, $p(y|\Theta)$ 는 극대강수량자료 y 의 우도함수(likelihood function)를 의미한다.

$$p(y|\Theta) = \prod_{i=1}^T \text{Gumbel}(y_i | \mu(t), \sigma(t)) \quad (12)$$

여기서, T 는 강수량계열의 자료연수를 나타낸다. 식 (4), 식 (5), 식 (10), 식 (12)에서 정의되는 매개변수들의 사전분포들을 식 (13)에 대입시킴으로써 매개변수들의 사후분포를 식 (13)을 통해 추정할 수 있다. 식 (13)에서는 Gumbel 분포의 위치 및 규모 매개변수와 회귀매개변수들을 위한 비공액사전분포(non-conjugate prior)를 이용하여 여러 매개변수들을 추정한다. 공액사전분포(conjugate prior)는 사전분포와 우도가 공액을 이룬다는 의미로서 사전분포와 사후분포가 동일한 분포를 나타내도록 사전분포를 선택하는 방법으로서 편의성을 기준으로 선택되는 간단한 방법이다. 즉, 우도함수가 이항분포를 따르는 경우 Beta 분포를 사전분포로 사용하게 되면 공액분포가 된다. 본 연구에서는 비공액사전분포를 이용하여 사후분포를 추정하였다.

$$p(y|\Theta) \propto \prod_{i=1}^T \text{Gumbel}(y_i | \mu(t), \sigma(t)) \cdot N(\alpha(t) | \alpha_0 + \alpha_1 t, \sigma_\alpha^2) \cdot N(\sigma(t) | \beta_0 + \beta_1 t, \sigma_\beta^2) \quad (13)$$

식 (13)은 모든 매개변수에 대한 적분을 통해 직접적으로 추정하는 것은 불가능하며, 본 연구에서는 앞서 언급한

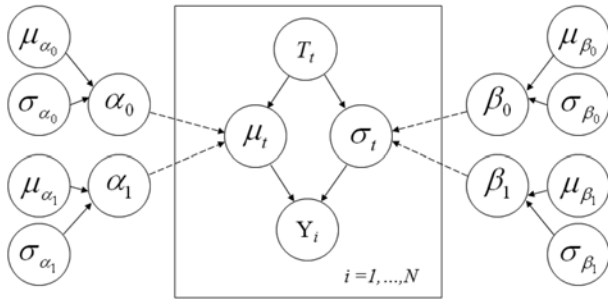


Fig. 3 Bayesian network for nonstationary frequency analysis

MCMC 기법을 도입하여 매개변수들의 사후분포를 추정하게 된다. 본 연구에서는 MCMC 방법 중 깃스표본법을 이용하여 회귀매개변수들을 추정하였으며 모형의 Convergence를 확증하기 위해서 3개의 Chain을 독립적으로 시행하여 Sampling이 효과적으로 혼합(mixing)되도록 하였다. Fig. 3은 본 연구에서 제안하는 Bayesian Network 모형에 대한 개념도를 나타낸다.

3. 적 용

본 연구에서는 서울지방 24시간 연최대강수량을 대상으로 모형의 적합성을 평가하였다. 앞서 언급했듯이 자료의 경향성 및 비정상성 빈도해석을 위해서 Gumbel 분포를 이용하여 연구를 진행하게 되며 Gumbel 분포가 서울지방 강수량에 적합한지를 평가하기 위해서 시각적인 검토를 우선적으로 실시하였다. Fig. 4는 정규분포 및 Gumbel 확률지에 서

울강수량을 도시한 결과를 나타내며 정규분포의 경우 자료 계열이 이론적인 정규분포의 Quantile(직선) 값을 크게 벗어나고 있는 반면 Gumbel 분포의 경우 Upper 및 Lower Tail부분의 거동을 잘 묘사하고 있다. 또한 경험적인 누기확률밀도함수와 Gumbel 분포형의 누기확률밀도함수가 유사한 거동을 하고 있음을 Fig. 4에서 확인할 수 있다. 따라서 Gumbel 분포형이 서울지방 연최대강수량의 분포특성을 표현하는데 무리가 없을 것으로 판단되며 극치자료계열의 선형성 및 비정상성 빈도해석 연구를 진행하였다.

선형회귀분석의 경우 자료계열을 정규분포로 가정하기 때문에 극치강수량의 경향성과 빈도해석을 효과적으로 연계해석하기 어렵다는 점을 앞서 언급한바 있다. 정규분포의 경우 종의 모양(bell-shape)을 가지고 있고 Skewness가 0이기 때문에 회귀분석 결과에 대한 불확실성 또한 정규분포를 가지게 된다. 극치자료계열의 경향성을 분석할 때, 극대치강수량 Y 가 $Gumbel(\mu(t), \sigma(t))$ 을 따라서 분포된다면 기존 회귀분석결과에 비해서 절편과 기울기가 작아지게 됨을 직감적으로 알 수 있다. Fig. 5는 자료의 분포형을 정규분포와 Gumbel 분포로 가정했을 때 각 분포형이 적합되는 과정을 개념적으로 표현한 것으로서 각 분포형의 Tail 거동은 매우 다르게 나타나게 된다. 이런 경우 같은 Quantile 값이라도 정규분포의 경우 Gumbel 분포보다 매우 높은 빈도의 값으로 해석되게 된다.

이는 Gumbel분포와 같은 극치분포의 경우 양의 Skewness를 나타내기 때문에 Upper Tail이 두꺼운 특성을 가지며 분포의 중앙값(Median)을 연결시켜 경향성을 나타낼 경우 선형

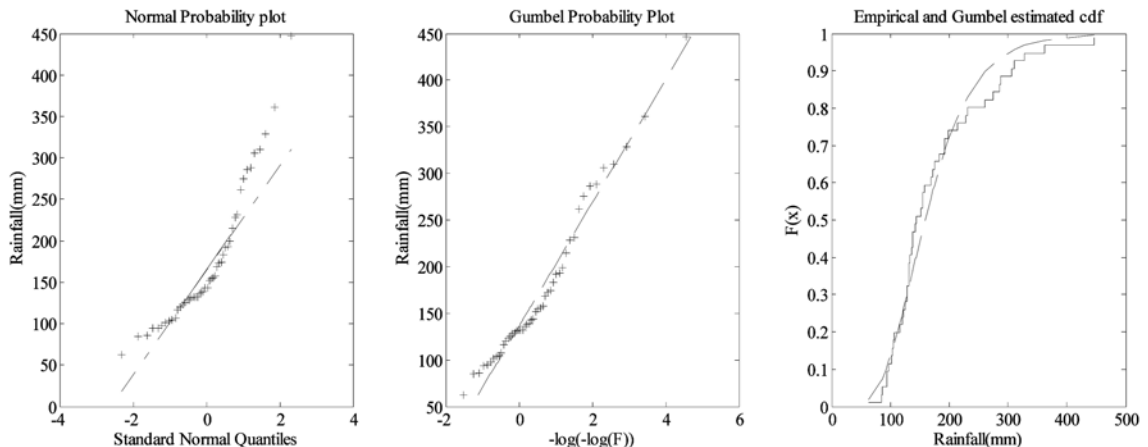


Fig. 4 Visual inspection of tail behavior through probability plot. (a) Normal probability plot for annual maxima, (b) Gumbel probability plot for annual maxima, (c) Empirical CDF versus Theoretical Gumbel CDF

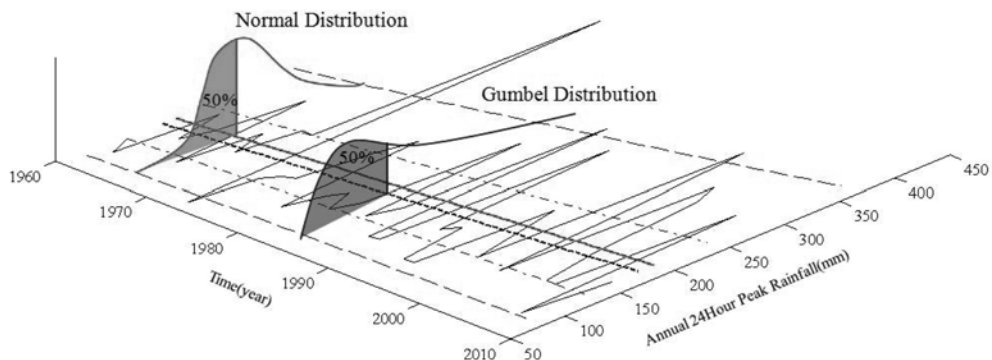


Fig. 5 Concept on Bayesian regression analysis with Gumbel distribution. A regression scheme with Normal distribution is also tabulated for a comparison of the Bayesian regression

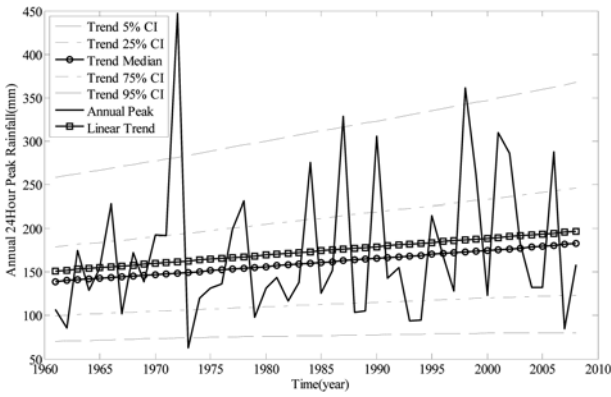


Fig. 6 Comparison of trend analysis of extreme rainfall time series between Normal distribution and Gumbel distribution

회귀분석과 다른 결과를 가지게 된다. Fig. 6은 자료계열을 정규분포로 가정하는 경향성과 Gumbel 분포로 가정했을 경우 경향성을 비교한 결과로서 사각형과 함께 표시된 선으로 나타나는 정규분포 경향선이 원과 함께 표시된 선으로 나타나는 Gumbel 분포 경향선보다 큰 값을 나타내고 있음을 확인할 수 있다. 중앙값이외에 5%, 25%, 75%, 95%의 불확실성 구간(confidence interval; CI)을 사후분포로부터 추정하여 함께 도시하였다. Fig. 6에서 보는바와 같이 불확실성 구간이 중앙값을 중심으로 좌우 대칭이 아닌 비대칭으로 Upper Tail이 Lower Tail 보다 두껍게 나타나는 것을 알 수 있으며 이는 Gumbel 분포형으로 극대강수량계열의 경향성을 나타낸 결과라 할 수 있다.

Bayesian 모형의 장점은 모든 매개변수가 사후분포로 귀결되기 때문에 통계적인 추론이 자연적으로 가능하게 되며 이를 통해 추정된 매개변수의 유의성 여부를 정량적으로 해석할 수 있다. Fig. 7은 식 (4)와 식 (5)에서 나타낸 매개변수들의 사후분포를 핵밀도함수로 추정한 결과를 나타낸다. 그림에서 α_0 와 β_0 의 경우 모두 양의 값을 가지고 작은 분

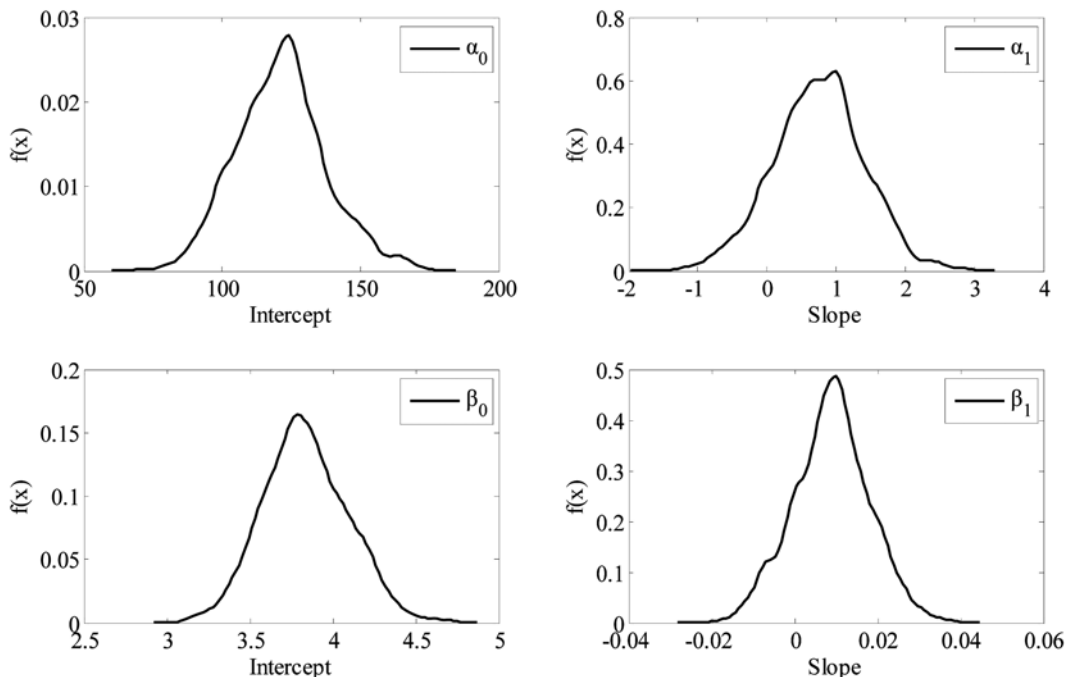


Fig. 7 Model parameters of trend model in nonstationary frequency analysis and associated uncertainties. Here, α_0 and β_0 indicate intercepts while α_1 and β_1 represent slopes in location and scale parameter of Gumbel distribution

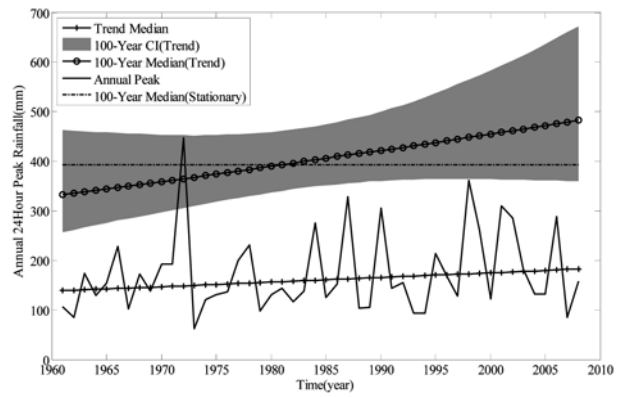


Fig. 8 Comparison of trend analysis of extreme rainfall time series between Normal distribution and Gumbel distribution

산을 가지는 사후분포를 나타내고 있어 통계적으로 유의한 특성을 가지고 있으나 기울기인 α_1 과 β_1 은 상대적으로 작은 확률어기는 하지만 0을 중심으로 양의 값과 음의 값이 동시에 나타나고 있어 통계적으로 유의성이 떨어진다 하겠다. Fig. 7의 사후분포로부터 α_1 과 β_1 이 양의 기울기를 가질 확률을 정량적으로 적분해보면 각각 0.87, 0.83으로 유의수준 10%에서 모두 기각되는 결과를 확인할 수 있다. 이는 서울 지방의 극대강수량계열의 경향성이 그리 크지 않기 때문이며, 유의한 경향성을 갖는 지점의 자료를 이용할 경우 사후분포의 거동은 다른 양상을 보일 것이다.

본 연구의 주요 목적중 하나는 이러한 극치자료의 경향성을 빈도해석과 연계시키는 것이다. 이를 위해서 식 (13)을 통해 추정된 사후분포의 매개변수를 활용하게 되며 원하는 재현기간(T)에 해당하는 확률수문량은 다음 식 (14)에 나타낸 Gumbel 분포의 Quantile 함수로부터 추정이 가능하다.

$$X(t) = \mu(t) - \sigma(t) \cdot \ln(-\ln(1 - 1/T)) \quad (14)$$

Fig. 8은 경향성을 고려하여 추정된 100년 빈도 강수량과

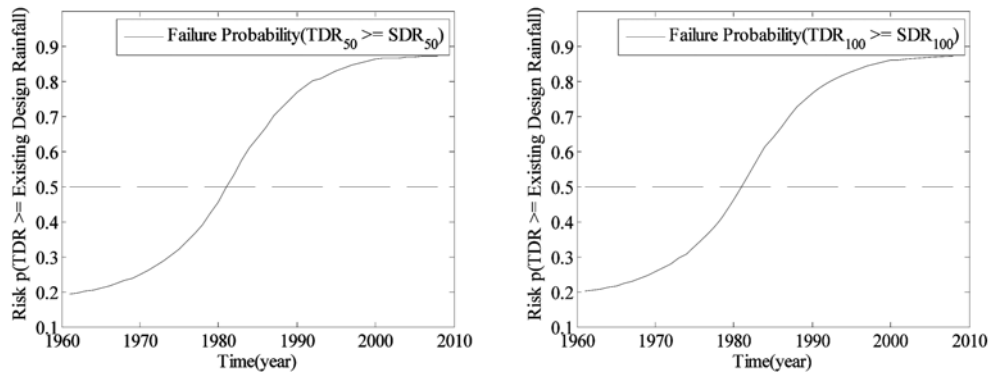


Fig. 9 Failure probability $p(\text{nonstationary Design Rainfall} > \text{Stationary Design Rainfall})$ estimated by posterior distribution

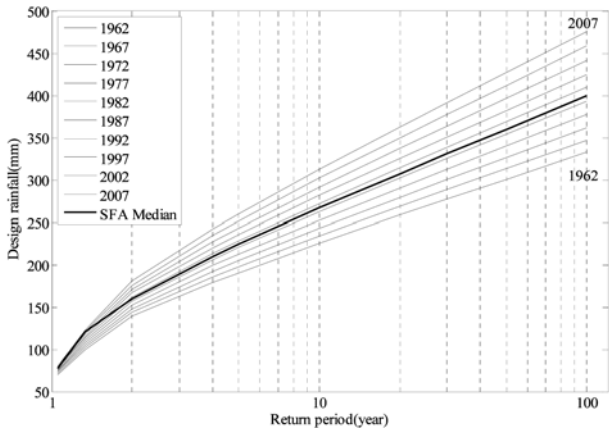


Fig. 10 Design rainfall estimates using different parameter sets (location and scale parameters) according to each year

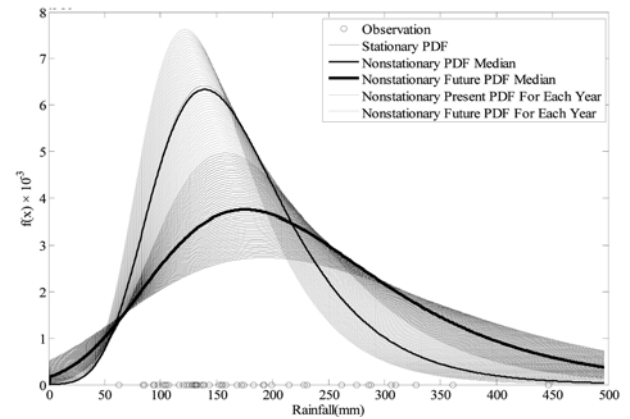


Fig. 11 Evolution of probability density function derived from the nonstationary frequency analysis. The blue and Black solid line show PDF's median values from PDF ensembles

기존 정상성가정을 기반으로 추정된 100년 빈도 강수량을 비교한 그림이며, 비정상성 빈도해석으로부터 추정된 100년 빈도 강수량의 불확실성 구간을 사후분포로부터 추정하여 나타내었다. 경향성이 존재한다고 가정하면 1982년을 전후로 현재의 100년 빈도 강수량의 빈도가 100년 빈도 이하로 낮아지고 있음을 확인할 수 있다. 이러한 사항은 동적 위험도 관리(dynamic risk management) 관점에서 매우 중요한 요소로서, 비정상성 빈도해석을 통한 결과와 불확실성 구간 정보를 이용한 다양한 결과활용 방법 중 재해보험, 구조물의 위험도 관리 등에 있어서 주요 정보로 활용되는 비선형적 초과확률에 대한 접근을 가능하게 해준다.

Fig. 9는 비정상성 빈도해석 결과활용에 대한 예시로서 동적 위험도를 도시한 것이다. 극대강수량의 경향성이 존재한다는 가정아래 추정된 50년/100년 빈도 강수량이 현재 우리

가 알고 있는 50년/100년 빈도 강수량을 초과할 수 있는 확률을 시간에 따라 도시한 결과이다. 실선으로 표시된 초과 확률이 1982년 이후로 50%를 초과하고, 이후 증가하는 것을 확인할 수 있다.

Fig. 10은 경향성이 고려된 재현기간별 확률강수량을 도시한 그림이다. 연도별로 추정된 위치 및 규모매개변수를 이용하여 산정한 빈도곡선을 5년 간격으로 도시하였다. 1982년 이후로 정상성 빈도해석 값을 초과하여 빈도곡선이 그려지는 확인할 수 있다.

마지막으로 Fig. 11은 정상성 빈도해석과정에서 추정된 Gumbel 확률밀도함수와 시간에 따라 연속적으로 추정된 확률밀도함수를 비교한 것이다. Fig. 11에서는 현재 극치강수량의 확률밀도함수뿐만 아니라 지금의 경향성이 미래에도 유

Table 1. Posterior density of parameters and associated uncertainty bounds. Note that nonstationary μ and σ are aggregated over time

Parameters		Mean	SD	2.50%	median	97.50%
Stationary	μ	139.10	8.40	122.60	139.20	155.80
Nonstationary	μ	139.55	11.76	119.90	139.65	157.45
Stationary	σ	57.34	7.35	44.63	56.87	73.60
Nonstationary	σ	58.08	9.67	44.46	57.57	76.12
α_0		121.30	16.28	90.95	121.50	155.30
α_1		0.75	0.66		0.75	2.00
β_0		3.84	0.26	3.36	3.82	4.35
β_1		0.01	0.01		0.01	0.03

지된다는 가정으로 미래의 확률밀도함수를 추정하여 비교하였다. 도시한 바와 같이 확률분포의 Upper Tail이 경향성의 영향으로 점증적으로 두꺼워지는 효과를 확인할 수 있다. 경향성을 고려한 비정상성해석결과와 매개변수들의 사후분포에 대한 정보는 Table 1에서 확인할 수 있다.

4. 결 론

본 논문에서는 극치수문자료의 경향성 분석 개념을 소개하고 이를 빈도해석과 연계시켜 해석하는 방법론을 제시하고자 Gumbel 극치분포를 기반으로, 시간변화에 의한 수문빈도 특성 변화를 모의할 수 있는 Bayesian 모형을 구성하였다. 사후분포의 매개변수는 깃스표본법에 의한 Markov Chain Monte Carlo Simulation을 통해 추정하였으며, 이를 통해 경향성을 고려한 확률강우량과 불확실성 구간을 추정하였다. 또한 경향성을 고려한 확률강우량이 현재 알려진 확률강우량을 초과할 확률을 통해 동적 위험도 해석과정을 소개하였으며, 현재의 경향성에 대해서 시간에 따라 연속으로 추정된 확률밀도함수를 비교하여 수문학적 위험도가 증가할 수 있음을 모의결과를 통해 확인하였다. 본 연구를 통해 제시된 결과를 요약하면 다음과 같다.

첫째, 본 연구에서는 극치수문자료의 분포특성을 검토하고 일반적인 회귀분석을 통한 경향성 분석 방법의 문제점을 평가하였으며, 극치분포를 활용한 회귀분석 방법을 제안하였다.

둘째, 극치분포를 활용한 경향성 검토 결과를 수문빈도해석 결과로 연계시켜 해석할 수 있는 Bayesian 방법론을 소개하고 시간에 따른 수문학적 빈도의 변화 양상을 모의할 수 있는 기법을 제시하였다.

셋째, 단순히 경향성의 존재여부를 확인하는데 그치지 않고 사후분포를 통해서 통계적 추론을 수행함으로써 경향성에 대한 통계학적인 유의성을 정량적으로 평가할 수 있었다.

넷째, 본 연구에서 제시된 방법론을 통해서 유도된 사후분포 및 통계적 추론 결과를 바탕으로 다양한 동적위험도 정보를 유도하는 과정을 소개하였다.

참고문헌

권영문, 박진원, 김태웅(2009) 강우량의 증가 경향성을 고려한 목표연도 확률강우량 산정, **대한토목학회논문집**, 대한토목학회, 제29권, 제2B호, pp. 131-139.

권현한, 문영일(2007) 기상정보 및 태풍특성을 고려한 계절 강수량의 확률론적 모형 구축, **대한토목학회논문집**, 대한토목학회, 제27권, 제1B호, pp. 45-52.

권현한, 박래건, 최병규, 박세훈(2010) Climate change scenario generation and uncertainty assessment: Multi variables and potential hydrological impacts, 2010년 **한국수자원학회 학술발표회 논문집**, 한국수자원학회, pp. 268-272.

김보경, 김병식, 김형수(2008) 극한지수를 이용한 극한 기상사상의 변화 분석, **대한토목학회논문집**, 대한토목학회, 제28권, 제1B호, pp. 41-53.

박성현(1997) **회귀분석**, 민영사.

이동률, 김웅태, 유철상(2004) 기후변화가 기상학적 가뭄과 홍수에 미치는 영향, **한국수자원학회논문집**, 한국수자원학회, 제37권, 제4호, pp. 315-328.

정대일, 제리 스테던저, 성장현, 김영오(2008) 기후 변화를 고려

한 홍수 위험도 평가, **대한토목학회논문집**, 대한토목학회, 제28권, 제1B호, pp. 55-64.

Anderson, P.L. and Meerschaert, M.M. (1998) Modeling river flows with heavy tails, *Water Resources Research*, Vol. 34, No. 9, pp. 2271-2280.

Egozcue, J.J. and Ramis, C. (2001) Bayesian hazard analysis of heavy precipitation in eastern Spain, *International Journal of Climatology*, Vol. 21, No. 10, pp. 1263-1279.

Farquharson, F.A.K., Meigh, J.R., and Sutcliffe, J.V. (1992) Regional flood frequency analysis in arid and semi-arid areas, *Journal of Hydrology*, Vol. 138, pp. 487-501.

Franks, S.W. and Kuczera, G. (2002) Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales, *Water Resources Research*, Vol. 38, No. 5, DOI: 10.1029/2001WR000232.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004) *Bayesian data analysis*, Boca Raton, Fla., Chapman & Hall/CRC., ISBN 1-58488-388-X.

George, E.I. and McCulloch, R.E. (1993) Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 881-889.

Halpert, M.S. and Ropelewski, C.F. (1992) Surface-Temperature Patterns Associated with the Southern Oscillation, *Journal of Climate*, Vol. 5, No. 6, pp. 577-593.

Jain, S. and Lall, U. (2000) Magnitude and timing of annual maximum floods: Trends and large-scale climatic associations for the Blacksmith Fork River, Utah, *Water Resources Research*, Vol. 36, No. 12, pp. 3641-3651.

Jain, S. and Lall, U. (2001) Floods in a changing climate: Does the past represent the future?, *Water Resources Research*, Vol. 37, No. 12, pp. 3193-3205.

Katz, R.W., Parlange, M.B., and Naveau, P. (2002) Statistics of extremes in hydrology, *Advances in Water Resources*, Vol. 25, No. 8-12, pp. 1287-1304.

Knox, J. (1993) Large increase in flood magnitude in response to modest changes in climate, *Nature*, Vol. 361, pp. 430-432.

Kwon, H.H., Brown, C., and Lall, U. (2008) Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling, *Geophysical Research Letters*, AGU, Vol. 35, DOI: 10.1029/2007GL032220.

Kwon, H.H., Brown, C., Xu, K.Q., and Lall, U. (2009) Seasonal and annual maximum streamflow forecasting using climate information: application to the Three Gorges Dam in the Yangtze River basin, China, *Hydrological Science Journal*, Vol. 54, No. 3, pp. 582-595.

Kwon, H.H. and Moon, Y.I. (2006) Improvement of overtopping risk evaluations using probabilistic concepts for existing dams, *Stochastic Environmental Research and Risk Assessment*, Vol. 20, No. 4, pp. 223-237.

Madsen, H., Rasmussen, P.F., and Rosbjerg, D. (1997) Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. at-site modeling. *Water Resources Research*, Vol. 33, pp. 747-757.

Milly, P.C.D., Wetherald, R.T., Dunne, K.A., and Delworth, T.L. (2002) Increasing risk of great floods in a changing climate, *Nature*, Vol. 415, pp. 514-517.

Morrison, J.E. and Smith, J.A. (2002) Stochastic modeling of flood peaks using the generalized extreme value distribution, *Water Resources Research*, Vol. 38, No. 12, DOI: 10.1029/2001WR000502.

Piechota, T.C. and Dracup, J.A. (1996) Drought and regional hydrologic variation in the United States: Associations with the El Nino Southern Oscillation, *Water Resources Research*, Vol. 32, No. 5, pp. 1359-1373.

Pizaro, G. and Lall, U. (2002) El Niño and Floods in the US West: What can we expect?, *Eos Trans.*, AGU, Vol. 83 No. 32, pp.

349-352.

- Porparto, A. and Ridolfi, L. (1998) Influence of weak trends on exceedance probability, *Stochastic Hydrology and Hydraulics*, Vol. 12, No. 1, pp. 1-15.
- Reid, D.S. and Stedinger, J.R. (2005) Bayesian MCMC flood frequency analysis with historical information, *Journal of Hydrology*, Vol. 313, No. 1-2, pp. 97-116.
- Ropelewski, C.F. and Halpert, M.S. (1996) Quantifying Southern Oscillation - Precipitation relationships, *Journal of Climate*, Vol. 9, No. 5, pp. 1043-1059.
- Ropelewski, C.F., Halpert, M.S., and Wang, X.L. (1992) Observed tropospheric biennial variability and its relationship to the Southern Oscillation, *Journal of Climate*, Vol. 5, No. 6, pp. 594-614.
- Sankarasubramanian, A. and Lall, U. (2003) Flood quantiles in a changing climate: Seasonal forecasts and causal relations, *Water Resources Research*, Vol. 39, No. 5, DOI: 10.1029/2002WR001593.

(접수일: 2010.2.24/심사일: 2010.6.10/심사완료일: 2010.7.1)