

BH 베이지안 분석을 통한 서울지점 강우자료의 확률적 변화시점 추정**A Probabilistic Estimation of Changing Points of
Seoul Rainfall Using BH Bayesian Analysis**

황 석 환* / 김 중 훈** / 유 철 상*** / 정 성 원****

Hwang, Seok Hwan / Kim, Joong Hoon / Yoo, Chulsang / Jung, Sung Won

Abstract

In this study, occurrences of relative probabilistic changing points between Chukwooki rainfall data (CWK) and modern rain gage data (MRG) were analyzed using Barry and Hartigan (BH) Bayesian changing points estimation method which estimated the changing points by calculation of change probabilities at each point. Since any natural phenomenon cannot be simulated identically and perfectly, a statistical method which can not consider the sequential order has its limitation on prediction of a specific time of occurrence. In this respect, Homogeneity analysis between CWK and MRG was performed through the occurrence investigation of relative probabilistic changing points for four rainfall characteristics of data sets using BH bayesian model which estimate the change point by calculating the relative probabilities in each data points. The results show that statistical characteristics of CWK are not different significantly from MRG, even though considered that there may be little quantitative difference CWK and MRG caused from limitation of measurement accuracy of CWK.

Keywords : bayesian, changing point, rainfall, chukwooki

요 지

본 논문에서는 각각의 시점에서의 변화확률을 산정하여 변화시점을 추정하는 Barry and Hartigan (BH)의 베이지안 변화시점 추정방법 (Bayesian changing points estimation method)을 이용하여 측우기 관측자료계열 (CWK)과 근대우량계 관측자료계열 (MRG)간의 변화에 대한 상대확률적 질점의 발생여부를 분석하였다. 어떠한 자연 현상도 완전히 동일하게 재현되지 않기 때문에 시간적인 순서를 고려하지 못하는 통계적 방법은 구체적인 발생시간을 예측하기 어렵다는 점에서 자료의 변화량 파악은 가능하나 변화시점을 명확히 추정하는데 한계가 있다. 이러한 배경에서, 본 논문에서는 각 시점에서의 변화확률을 산정하여 변화시점을 추정하는 BH 베이지안 방법을 적용하여 CWK와 MRG의 각 강우특성별로 상대확률적인 변화시점 분석을 통하여 CWK와 MRG 간의 동질성 분석을 실시하였다. 분석 결과, CWK의 정성적인(본질적인) 통계적 특성은 MRG와 큰 차이가 없는 것으로 나타났다. 다만, 관측정밀도의 한계로 인한 정량적인 차이가 존재하는 것으로 분석되었다.

핵심용어 : 베이지안, 변화시점, 강우량, 측우기

* 한국건설기술연구원 수자원연구실 수석연구원 (e-mail: sukany@kict.re.kr)

Senior Researcher, Korea Institute of Construction Technology, Gyeonggi-do 411-712, Korea

** 교신저자 · 고려대학교 공과대학 건축·사회환경공학부 교수 (e-mail: jaykim@korea.ac.kr)

Prof., School of Architecture, Civil & Environmental Eng., Korea University, Seoul 136-713, Korea

*** 고려대학교 공과대학 건축·사회환경공학부 교수 (e-mail: envchul@korea.ac.kr)

Prof., School of Architecture, Civil & Environmental Eng., Korea University, Seoul 136-713, Korea

**** 한국건설기술연구원 수자원연구실 연구위원 (e-mail: swjung@kict.re.kr)

Chief Researcher, Korea Institute of Construction Technology, Gyeonggi-do 411-712, Korea

1. 서론

본 논문에서는 각 시점에서의 변화확률을 산정하여 변화시점을 추정하는 베이지안 변화시점 추정방법을 이용하여 측우기 자료와 근대우량계 자료 사이의 변화에 대한 상대확률적 절점의 발생여부를 분석하였고 이를 통해 두 자료계열간의 동질성 분석을 실시하였다. 변화시점을 파악하기 위해 일반적으로 사용되는 방법은 변화시점의 특정위치를 추정하는 형식이다. 그러나 베이지안 과정은 연속시계열에서 확률분포를 산정(연속시계열의 각각의 위치에서 변화확률을 산정)한다. 연속시계열에서의 변화시점은 변화전후 상태에 따라 결정이 되는 상대적인 개념이기 때문에 확률론적 방법으로 변화시점을 결정하는 것은 매우 어렵다. 이러한 관점에서 상대적인 변화시점의 확률을 계산하여 변화시점을 결정하는 베이지안 과정은 변화시점을 보다 합리적으로 추정할 수 있는 방법이다. 측우기 관측 자료와 근대우량계 관측 자료는 연속 시계열자료로서 장기간의 시간적 차이를 보이기 때문에 경년변동을 고려하기 힘든 평균이나 표준편차와 같은 정량적인 기본 통계특성의 크기 비교만으로는 정확히 동질성을 판단하는 것이 매우 힘들다. 그리고 기존의 통계적인 유의수준에서의 측우기 관측 강우량 자료계열과 근대우량계 관측 강우량 자료계열간의 동질성 분석은 해당 강우량 계열의 모집단 분포형에 대한 정확한 추정이 수반되어야만 가설 검정에 대한 신뢰가 가능하고, 동질성을 판단하는 유의수준의 선택에 있어 객관적인 기준을 부여하기 어렵기 때문에 결과에 대한 절대적인 신뢰도를 부여하기 어렵다는 단점이 있다. 따라서 본 연구에서는 베이지안 분석을 이용한 변화시점 추정기법을 이용하여 측우기 관측 강우량 자료계열과 근대우량계 강우량 관측계열간의 강우특성별로 변화시점이 발생하는가를 분석해 이를 토대로 두 자료계열의 동질성에 대한 판단 근거로 이용하고자 한다. 수자원분야에서 베이지안 방법이 적용된 예는, 초기(1970년대)에는 주로 수자원 정책 결정을 위해 필요한 불확실도를 가능해 보는데 이용이 되었고, 이러한 연구로는 Vicens et al. (1975)과 Wood and Rodriguez-Iturbe (1975a, 1975b)의 결과가 있다. 그 후 1990년 이후 컴퓨터의 비약적인 발전으로 인하여 방대한 양의 자료처리가 가능해짐으로써 홍수빈도 분석과 같은 분야에 적용되기 시작하였다. 이러한 연구로는 Madsen and Rosbjerg (1997), Kuczera (1999), O'Connell et al. (2002), Reis Jr. et al. (2005), Reis Jr. and Stedinger (2005), Seidou et al. (2006)이 있다. 국내의 경우는 최근에 김상욱과 이길성 (2008a, 2008b, 2008c)이 저수량 빈도해석을 수행한 바 있고 수위-유량관계곡선의 불확실성을 분석 (2008d)한 바 있다. 이러한 연구들은 주

로 점빈도해석이나 지역빈도해석에 있어 자료에 기반한 사전분포를 구축하고 모수와 모수의 불확실성을 추정하기 위해 베이지안 기법을 적용하였다. Perreault et al. (2000a, 2000b)은 복잡한 통계모형에 대한 베이지안 추론을 위해 잘 알려진 깃스(Gibbs) 표본추출 방법을 이용하여 대형 수력발전 시스템에 대해 연수력발전량의 변화시점과 변화의 강도를 추정하기 위한 분석을 실시한 바 있다. 그러나 연속자료계열의 각 시점에서의 조건부 변화확률을 산정하여 변화시점을 추정하기 위한 상대확률적 베이지안 변화시점 추정방법 (Barry and Hartigan, 1992)은, 고정시점 변화확률 방법에 비해 시계열 자료의 변화시점을 추출하는데 있어 보다 합리적인 결과를 도출할 수 있음에도 불구하고, 아직까지 수자원 분야에서 적용된 예가 없다. 본 연구에서는 이러한 부분에 착안하여, 그 동안 자료계열 간의 평균적인 통계특성치 위주로 평가되었던 측우기 관측 자료계열과 근대우량계 관측 자료계열간의 동질성에 대해서 발생특성 측면에서 검토를 해보고자 하였다. 서울지점 강우자료계열은 연속시계열이기 때문에 정량적 및 발생적 강우특성에 순차적으로 상대확률적인 변화가 나타나는지에 대한 검토가 필요하고, 이를 통해 두 자료계열 간에 명확하고 지속적인 변화시점이 발생한다면, 두 자료계열의 해당 강우특성 간에는 명확한 측정 혹은 해석적인 차이가 존재할 가능성이 높음을 의미한다. 본 논문에서는 변화시점 분석을 위해 월강우량, 월강우량 대비 최대일강우량의 비, 월강수일수, 월강수일수 대비 월강우량의 비의 4종류의 강우특성 자료계열을 만들었고, 이를 이용해 각 강우특성별로 상대확률적인 변화시점이 나타나는 지를 분석하여 측우기 관측자료계열과 근대우량계 관측자료계열의 각각의 강우특성이 상대확률적으로 동일한 계열인지를 검증하여 보았다.

2. 월별 동질성 분석을 위한 통계치의 설정

본 연구에서 사용한 측우기 관측 강우량 자료는 조희구와 나일성 (1979)과 전종갑과 문병권 (1997)의 연구결과로부터 얻어졌다 (김기욱 등에서 일부 재인용, 2007). 측우기 관측계열을 포함한 서울지점의 강우자료 및 해당 자료계열의 통계특성은 이병설 (1970), 조희구와 나일성 (1979), 정현숙과 임규호 (1994), 전종갑과 문병권 (1997), 황석환 등 (2009)에 자세히 제시되어 있고 본 논문에서는 이를 인용하는 것으로 자료에 대한 소개를 갈음하였다. 다음 Table 1은 측우기 관측계열과 근대우량계 관측계열간의 월별 동질성 분석을 하기 위해 선정된 4개의 통계치로 강우변화 특성을 효과적으로 나타내 줄 수 있다. 강우는 강우량과 같은 양적 특성과 발생횟수와 같은 빈도

Table 1. Statistics Which Are Used to Test the Homogeneity (Detect Changing Points) between CWK and MRG

Statistics	Mathematical Expressions
Monthly rainfall	$Monthly\ rainfall = \sum(daily\ rainfall)$
The ratio of daily maximum rainfall to monthly rainfall	$D_{max}\ ratio = \frac{Maximum\ daily\ rainfall}{Monthly\ rainfall}$
Number of rainy days for each month	$N_{rainy\ days} = Numbers\ of\ rainy\ days$
The ratio of monthly rainfall to numbers of rainy days	$I_{rainy\ days} = \frac{Monthly\ rainfall}{Numbers\ of\ rainy\ days}$

특성 그리고 지속시간에 따른 강도 특성을 복합적으로 가지고 있기 때문에 강우의 변화를 파악하기 위해서는 양적 통계치는 물론 빈도와 강도를 적절히 표현할 수 있는 통계치를 동시에 비교해야만 강우특성의 변화여부에 대한 적절한 판단이 가능하다. *Monthly rainfall*은 각 월의 월강우량이고 $D_{max}\ ratio$ 는 월강우량에 대한 해당 월 일최대 강우량의 평균적인 비율의 변화를 분석하기 위한 통계치이다. $N_{rainy\ days}$ 는 각 월별 평균적인 강우일수를 산정한 통계치이고 $I_{rainy\ days}$ 는 각 월별 강우일수에 대한 월강우량의 비로 강우일에 대한 평균강우강도이다. *Monthly rainfall*은 강우량의 정량적인 변화를 파악하기 위한 기본 통계치이고 $D_{max}\ ratio$ 는 일최대 강우량의 정량적 변화를 분석하여 강우의 규모변화를 파악할 수 있기 때문에 선택하였다. $N_{rainy\ days}$ 는 강우일수의 변화를 정량적으로 나타내므로 이를 분석하면 강우발생 경향의 변화 여부를 쉽게 파악할 수 있고 $I_{rainy\ days}$ 는 강우강도의 변화를 정량적으로 분석 가능하기 때문에 사용하였다.

3. 기본이론

본 논문에서는 Barry and Hartigan (1992, 1993)이 제시한 PPM (product partition model)에 근거한 베이지안 변화시점 분석방법을 이용하였다. 베이지안 방법과 대응하는 대표적인 방법들로는 CBS (circular binary segmentation; Olshen and Venkatraman, 2004)와 BP (break-points; Bai and Perron, 2003) 기법 등이 있다. Bai and Perron의 BP 방법은 여러 조각 (segments)으로 구성된 최적의 분리구간 (partitions)을 결정하기 위해 동적 프로그래밍 알고리즘을 이용한다. Olshen and Venkatraman의 CBS 방법은 binary segmentation (BS; Sen and Srivastava, 1975) 방법의 개선된 형태로, BS가 단일 변화시점 검정에 기반하고 있어 큰 변화구간 사이의 작은 변화구간을 찾는

데는 어려움이 있는데, 이러한 문제를 보완한 방법이다. 이 방법은 임의의 연속계열에서 변화시점이 없다는 귀무가설을 검정하기 위해 우도 비율 (likelihood ratio)을 검정통계량으로 사용하는 추정 알고리즘을 사용하고, 만약 이 검정에서 귀무가설이 기각된다면 해당 연속계열은 다시 분할되고 분할된 조각 (sub-segments)에서 추가로 변화가 나타나지 않을 때까지 반복된다. 그러나 BP와 달리 CBS는 주어진 변화시점 수에 대한 최적의 변화시점의 위치를 찾는 데 어려움이 있다고 알려져 있다 (Barry and Hartigan, 1993).

PPM의 이론적인 부분은 Barry and Hartigan (1992, 1993)에 상세히 기술되어 있어, 본 논문에서는 변화시점 추정과 관련된 가장 핵심적인 부분만을 간략히 기술하였다. Barry and Hartigan (BH)의 PPM은 해당 분리구간 안에서는 일정한 평균을 가지는 연속된 블록의 집합으로 구성된 분리구간 (partition) ρ 가 존재한다. 이 모델은 각각의 관측치들이, 독립적인 위치 i 에서의 평균이 μ_i 이고 분산이 σ^2 일 때, $N(\mu_i, \sigma^2)$ 인 독립분포를 가지고 각각이 독립적인 위치 i 에서의 변화시점의 확률을 p 로 가정하였다. μ_{ij} (시점 $i+1$ 에서 j 까지인 블록의 평균)의 사전분포는 $N(\mu_0, \sigma_0^2/(j-i))$ 로 결정된다. 이 식으로부터 블록의 길이에 따라 분산이 변화함을 알 수 있다. 짧은 블록안에서 μ_0 는 큰 편차를 보인다는 것을 알 수 있고, 짧은 지속시간에서는 평균의 변화가 상대적으로 크게 나타난다는 의미가 된다. BH는 임의의 분리구간 ρ 에 평균 μ_r ($r \in ij \in \rho$)에 다음의 추정식을 사용하였고, 이는 Yao (1984)의 식과도 매우 유사하다.

$$\mu_r = (1-w)\bar{X}_{ij} + w\mu_0 \quad (1)$$

여기서 $w = \sigma^2/(\sigma_0^2 + \sigma^2)$ 이다. BH는 네 매개변수와 분리구간 ρ 에 대한 독립 사전확률을 다음과 같이 제시하였고, BH 모델은 결국 이 추정식의 매개변수인 $\mu_0, p, \sigma^2, w = \sigma^2/(\sigma^2 + \sigma_0^2)$ 에 대한 독립 사전확률을 구하는 문제로

수렴하게 된다.

$$f(\mu_0) = 1, \quad -\infty \leq \mu_0 \leq \infty, \quad (2)$$

$$f(\sigma^2) = 1/\sigma^2, \quad 0 \leq \sigma^2 \leq \infty, \quad (3)$$

$$f(p) = 1/p_0, \quad 0 \leq p \leq p_0, \quad (4)$$

$$f(w) = 1/w_0, \quad 0 \leq w \leq w_0, \quad (5)$$

그리고

$$f(\rho) = \frac{1}{p_0} \left[\int_0^{p_0} p^{b-1} (1-p)^{n-b} dp \right] \quad (6)$$

여기서 p_0 와 w_0 는 사전에 선정된 $[0,1]$ 의 값이고, b 는 분리구간 내 블록의 수이다.

이 알고리즘은 분리구간 $\rho = (U_1, U_2, \dots, U_n)$ 으로 시작 되는데, n 은 관측치의 수이고 $U_i = 1$ 은 $i+1$ 위치가 변화 시점임을 의미한다; n 보다 작은 모든 i ($i < n$)에 대하여 U_i 는 0으로, $U_n \equiv 1$ 로 초기화된다. 마코브 연쇄 (Markov Chain; MC)의 각 단계에서, 각각의 위치 i 에서, U_i 의 값은 주어진 자료와 해당 분리구간의 U_i 의 조건분포로부터 추정하고, $i+1$ 위치에서 변화시점의 조건부 확률에 대한 전이확률 p 는 BH가 제시한 다음과 같은 간편 비율법 (simplified ratio)으로 산정할 수 있다.

$$\begin{aligned} \frac{p_i}{1-p_i} &= \frac{P(U_i = 1 \mathbf{X}, U_j, j \neq i)}{P(U_i = 0 \mathbf{X}, U_j, j \neq i)} \\ &= \frac{\left[\int_0^{p_0} p^b (1-p)^{n-b-1} dp \right] \left[\int_0^{w_0} \frac{w^{b/2}}{(W_1 + B_1 w)^{(n-1)/2}} dw \right]}{\left[\int_0^{p_0} p^{b-1} (1-p)^{n-b} dp \right] \left[\int_0^{w_0} \frac{w^{(b-1)/2}}{(W_0 + B_0 w)^{(n-1)/2}} dw \right]} \quad (7) \end{aligned}$$

여기서 W_0, B_0, W_1 과 B_1 은 각각 $U_i = 0$ 과 $U_i = 1$ 일 때 얻어지는 블록내 (within-block)와 블록간 (between-block)의 제곱합이다. 그리고 \mathbf{X} 는 자료계열이다. 조절 (tuning) 매개변수 p_0 와 w_0 는 $[0,1]$ 에서 선택하게 되고 각각의 반복과정을 통하여 사후평균은 현재의 분리구간 상에서 조건부로 갱신된다. 그러나 Eq. (7)의 우측 분모와 분자의 두 적분항 때문에 긴 연속자료에 대한 BH MCMC (Markov Chain Monte Carlo) 알고리즘의 직접 적용시 발산하거나 0으로 수렴하여 수치적으로 불안정할 수 있다. 다행히, 이 적분항은 불완전 베타 적분법 (incomplete beta integrals)에 의해 간략화 할 수 있다. 분리구간의 특정 위치 (주어진 자료와 현재 분리구간)에서의 변화시점에 대한 상대확률 (odds)은 다음과 같이 다시 표현할 수 있다.

$$\begin{aligned} \frac{p_i}{1-p_i} &= \frac{P(U_i = 1 \mathbf{X}, U_j, j \neq i)}{P(U_i = 0 \mathbf{X}, U_j, j \neq i)} \\ &= \left(\frac{W_0}{W_1} \right)^{\frac{n-b-2}{2}} \cdot \left(\frac{B_0}{B_1} \right)^{\frac{b+1}{2}} \cdot \sqrt{\frac{W_1}{B_1}} \quad (8) \\ &\quad \cdot \frac{\int_0^{\frac{B_1 w_0 / W_1}{1 + B_1 w_0 / W_1}} p^{(b+2)/2} (1-p)^{(n-b-3)/2} dp}{\int_0^{\frac{B_0 w_0 / W_0}{1 + B_0 w_0 / W_0}} p^{(b+1)/2} (1-p)^{(n-b-2)/2} dp} \\ &\quad \cdot \frac{\int_0^{p_0} p^b (1-p)^{n-b-1} dp}{\int_0^{p_0} p^{b-1} (1-p)^{n-b} dp} \end{aligned}$$

이 수식은 자료계열의 길이에 상관없이 BH 과정이 적용될 수 있도록 수치적으로 안정한 항들로 구성되어 있다. 결국, 임의의 분리구간 ρ 와 자료계열 \mathbf{X} 에 대하여 Eq. (1)과 Eq. (7)로부터 MCMC 수행을 통해 사후평균을 추정하게 된다. 본 논문에서는 BH PPM 방법의 적용을 위해 bcp package (Erdman and Emerson, 2007b)를 사용하였고, 조절 매개변수 p_0 와 w_0 는 비교적 적합하다고 알려진 0.2 (Barry and Hartigan, 1993; Yao, 1994)를 초기치로 사용하였다. 그리고 본 논문에서 제시한 Figs. 1~4에서 첫 번째는 자료 각각의 위치에서의 사후평균을 나타내고, 두 번째는 각각의 위치에서 MCMC 수행결과가 변화시점으로 판정될 확률 (변화에 관한 사후확률)을 나타낸다.

4. 분석결과

BH 베이지안 방법에 의한 변화시점 분석결과는 다음 Figs. 1~4와 같다. 우선 강우특성에 대한 전반적인 변화시점 분석을 위해 월강우량, 월강우량 대비 최대일강우량의 비, 월강우일수, 월강우일수 대비 월강우량의 비 (강우일에 대한 일강우강도)의 계열을 만들었다. 분석 목적은 측우기 관측 강우량자료계열 (1777~1907년)과 근대우량계 관측 강우량자료계열 (1908~2006년)간에 각 강우특성별로 상대확률적인 변화시점이 나타나는지를 분석하여 측우기 자료계열과 근대우량계 자료계열의 연속시간적인 동질성 여부를 판단해 보고, 더 나아가 상대확률적 크기를 토대로 그 양적 차이의 정도를 가늠해 보고자 함이다. 측우기 강우량에 있어 상대적으로 월강우량이 커서 자료의 신뢰도가 비교적 높다고 추정되는 6월에서 9월까지의 월별 강우특성 자료계열을 만들어 이를 분석한 결과는 Figs. 1~4와 같다. 그림의 좌측은 원자료계열 (MOO)에 대

한 분석결과이고 우측은 측우기 최소 관측정밀도인 약 2 mm를 고려하여 근대우량계 관측계열에서 2 mm 이하의 일강우량을 제외한 계열(M20)에 대한 분석결과이다. M20의 경우는 관측정밀도의 차이에서 오는 강우발생적인 차이를 최소화하여 유사한 조건에서 비교하려는 목적이다.

4.1 월별 변화시점 해석기준

결과의 기술에 앞서 본 논문에서 빈번히 사용되는 사전 확률(prior probability)과 사후확률(posterior probability)에 대한 설명을 간략히 할 필요가 있다. 두 용어는 베이저스 이론(Bayes's theorem)에서 사용되는 용어로서 베이저스 이론은 새로운 정보에 의해 사건 A의 발생확률을 알 때, 이를 근거로 B가 발생할 조건부 확률 $P(B|A)$ 를 구하는 방법이다. 여기서 사전확률은 새로운 정보가 주어지기 이전에 B가 발생할 확률 $P(B)$ 를 의미하고 사후확률은 새로운 정보가 주어진 이후에 사상 B가 발생할 확률 $P(B|A)$ 를 의미한다. 사후평균(posterior mean)에 대해서도 동일한 개념으로 해석하면 된다. 분석결과에서 제시한 변화확률의 해석에 있어 사후평균이 변화시점 전후로 장기적인 지속성이 없는 경우 이를 단기평균변화에 의한 일시적 사후변화확률의 변화로 보았고 이 변화시점을 "변동시점"으로 해석하였다. 그리고 사후평균이 변화시점 전후로 장기적인 지속성이 있는 경우 이를 장기평균변화에 의한 장기적 사후확률의 변화로 보았고 이 변화시점을 "변화시점"으로 해석하였다.

4.2 월별 변화시점 분석결과

6월의 경우는 Fig. 1과 같다. *Monthly rainfall*의 경우 변화에 대한 사후확률이 높은 시점이 여럿 존재하나 변화시점전후로 지속성이 없어(변동시점) 사후평균이나 사후확률에서 명확한 변화시점을 찾기는 힘들다. $D_{max} ratio$ 는 전체적으로 변화가 없음을 알 수 있다. $N_{rainy days}$ 의 경우는 M00의 경우 1908년을 전후로 뚜렷한 변화를 보이고 있다. M20을 기준으로 볼 때 $N_{rainy days}$ (사후평균의 경우)는 예전이나 근대이후 크게 차이를 보이지 않으나 약간 증가한 것을 볼 수 있다. $I_{rainy days}$ 는 명확한 변화시점을 찾기 어렵다. 6월의 경우를 종합해 보면, 관측최소단위의 영향으로 인한 강우일수의 차이를 감안하면 강우의 정량적인 차이는 시점 각각에서 확률적으로 명확하지 않음을 알 수 있다.

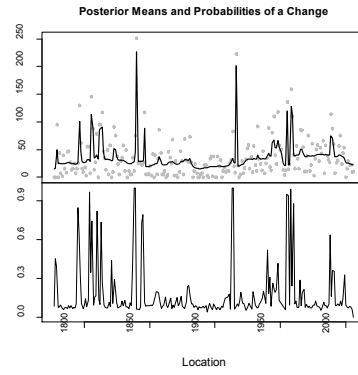
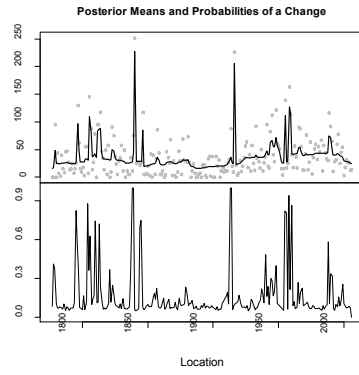
7월의 경우는 Fig. 2와 같다. *Monthly rainfall*의 경우 M00과 M20 모두 1908년을 전후로 뚜렷한 변화시점을 보이고 있지 않다. 이는, 7월의 경우, 상대적으로 강우량이나 강우빈도가 커서 2 mm 이하의 강우는 월강우량의 차이에 거의 영향을 미치지 못함을 의미한다. 더불어 강우

량의 양적차이도 예전이나 근대이후 별 차이를 보이지 않고 있다. 단, 1960년을 기준으로 사후확률이 크게 나타나는 점은 주목할 만한 결과이다. $D_{max} ratio$ 는 변화양상을 보이지 않고 있다. $N_{rainy days}$ 는 M00의 경우 1908년을 전후로 변화양상을 보이거나 M20의 경우는 변화양상을 보이고 있지 않다. 따라서 7월의 경우 예전과 근대이후 $N_{rainy days}$ 에 변화가 있다고 보기는 힘들다. $I_{rainy days}$ 는 뚜렷한 변화시점을 찾기 힘들다. 7월의 경우를 종합해 보면, 7월의 경우도 관측최소단위의 영향으로 인한 강우일수의 차이를 감안하면 강우의 정량적인 차이는 미소하여 변화여부가 명확하지 않음을 알 수 있다. 그러나 최근 7월에 강우량이 집중되고 있다는 점을 고려할 때, 1960년 이후 강우량의 변화가 감지되고 있음은 주목할 만하다. Lamb (1995)도 그의 저서에서 1960년대 중반 이후 북반구의 10도 간격의 경위도 대기압 또는 온도의 평균치로부터 산정되는 변동지수(Index of variability)가 급격히 상승하고 있음을 언급하고 있다.

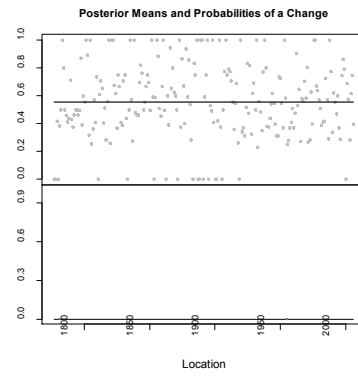
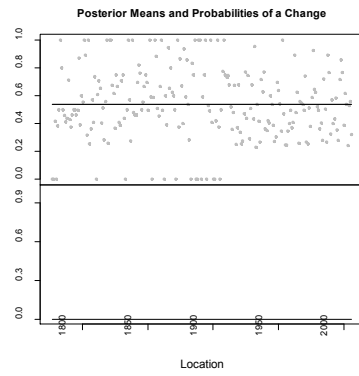
8월의 경우는 Fig. 3과 같다. *Monthly rainfall*의 경우 7월과 마찬가지로 M00과 M20 모두 1908년을 전후로 뚜렷한 변화시점을 보이고 있지 않다. 8월의 경우도 상대적으로 강우량이나 빈도가 커서 2 mm 이하의 강우는 월강우량의 차이에 거의 영향을 미치지 못함을 의미한다. $D_{max} ratio$ 도 변화시점을 보이지 않는다. $N_{rainy days}$ 는 M00의 경우 1900년 초반을 전후로 변화양상을 보이거나 M20의 경우는 변화시점을 보이지 않는다. 따라서 8월의 $N_{rainy days}$ 의 변화는 과거와 근대이후에 차이가 있다고 보기 힘들고 M00에서의 차이는 관측의 정밀도 차이에서 기인한 것으로 판단된다. $I_{rainy days}$ 도 1908년을 전후로 뚜렷한 변화시점을 보이지 않는다. 단, M00의 사후평균이 근대이후 평균적으로 약간 감소한 경향을 보이거나 이는 측우기 기록에서 2 mm 이하 강우의 누락으로 인한 강우일수의 감소에 따라 측우기 기록의 $I_{rainy days}$ 가 상대적으로 크게 산정된 결과로 해석할 수 있다.

9월의 경우는 Fig. 4와 같다. *Monthly rainfall*의 경우 M00과 M20 모두 변화시점을 보이지 않는다. $D_{max} ratio$ 도 변화시점을 보이지 않는다. $N_{rainy days}$ 의 경우 M00의 경우는 1908년을 전후로 변화양상을 보이거나 M20의 경우는 변화양상을 보이지 않는다. 이 또한 측우기 관측 정밀도에 의한 차이로 설명이 가능하고 9월의 경우 $N_{rainy days}$ 의 변화는 특별히 포착되지 않는다. $I_{rainy days}$ 는 M00의 경우 1900년 전후로 변화양상(감소)을 보이거나 이는 앞서 언급한 바와 같이 관측정밀도 차이에 의한 월강우일수의 차이에서 기인한 결과로 볼 수 있다. 이는 7월과 8월의 결과에서도 9월만큼 뚜렷하지는 않으나 확인할 수 있다. M20의 경우는 변화시점을 보이지 않고 있다.

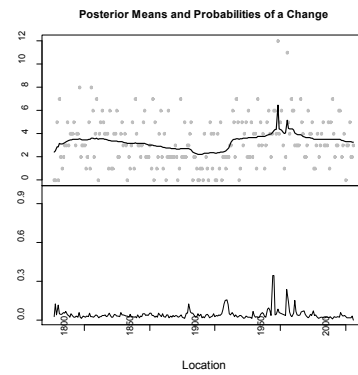
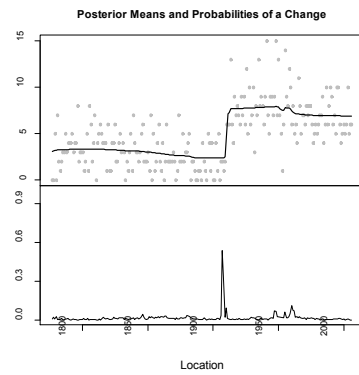
Monthly rainfall



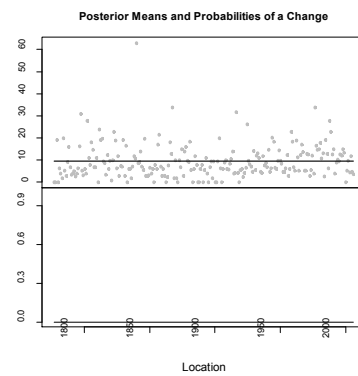
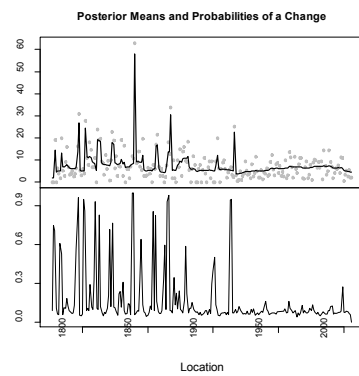
D_{\max} ratio



$N_{\text{rainy days}}$



$I_{\text{rainy days}}$

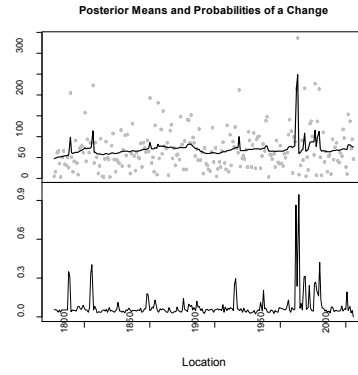
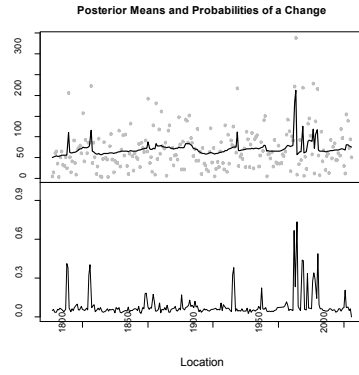


(a) M00

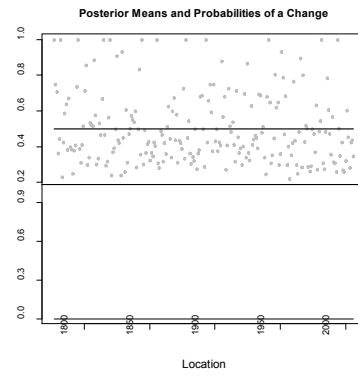
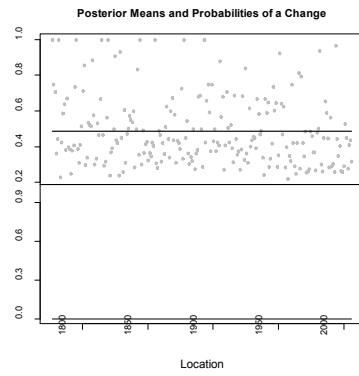
(b) M20

Fig. 1. Estimated Posterior Means and Posterior Probabilities of Each Statistical Rainfall Characteristics using BH Bayesian Changing Points Estimation Method for JUN

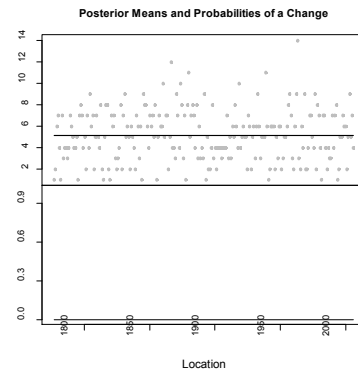
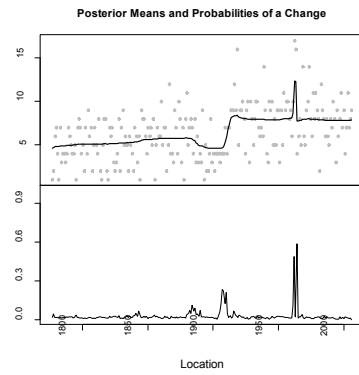
Monthly rainfall



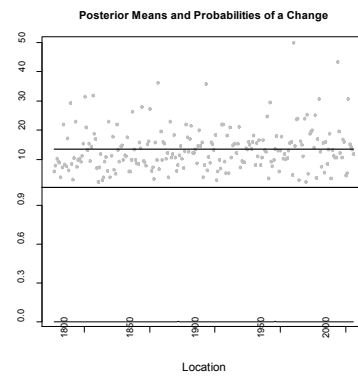
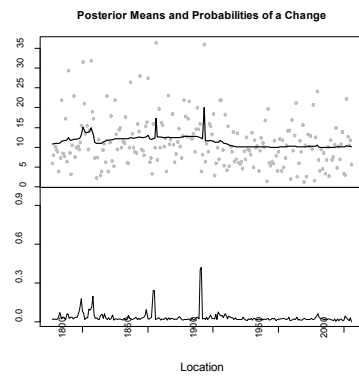
D_{\max} ratio



$N_{\text{rainy days}}$



$I_{\text{rainy days}}$

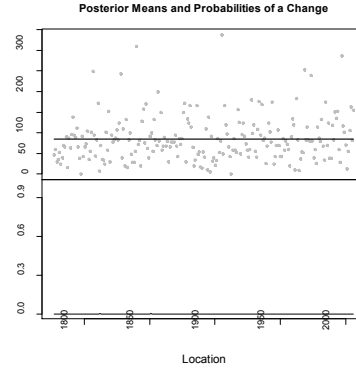
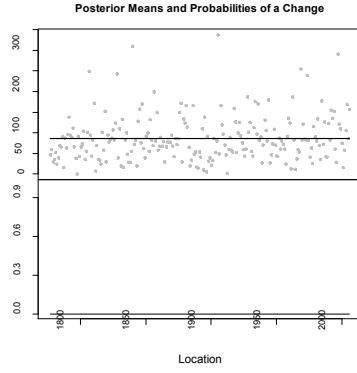


(a) M00

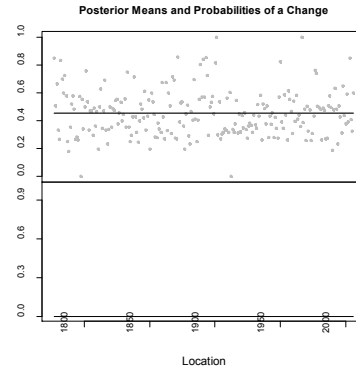
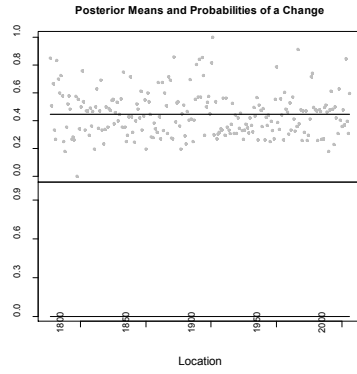
(b) M20

Fig. 2. Estimated Posterior Means and Posterior Probabilities of Each Statistical Rainfall Characteristics using BH Bayesian Changing Points Estimation Method for JUL

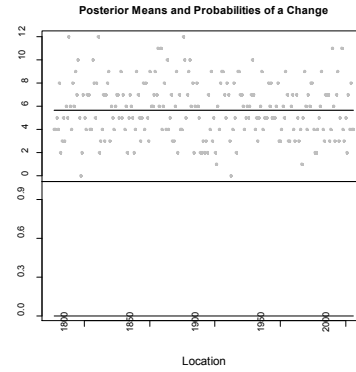
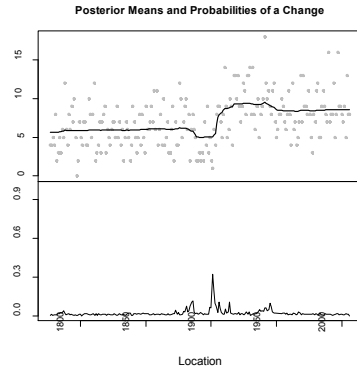
Monthly rainfall



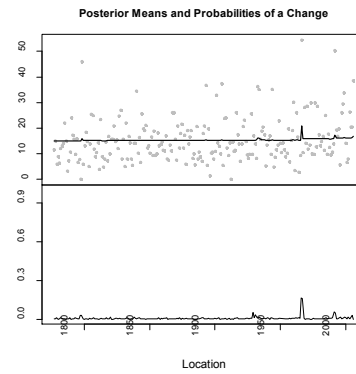
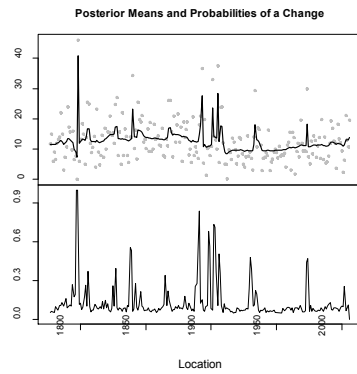
D_{\max} ratio



$N_{rainy\ days}$



$I_{rainy\ days}$

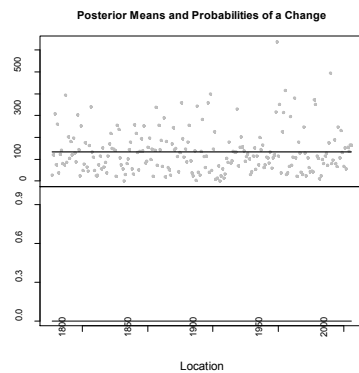
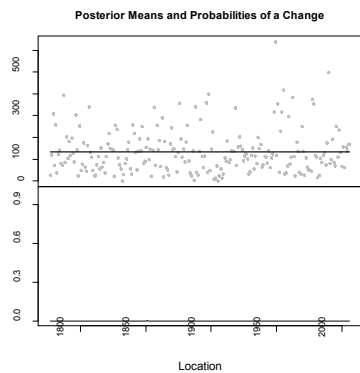


(a) M00

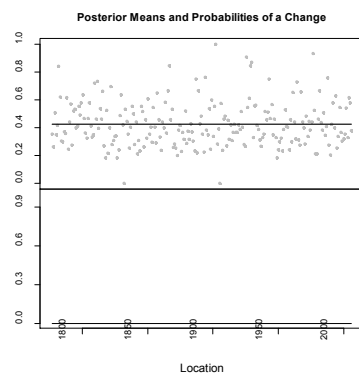
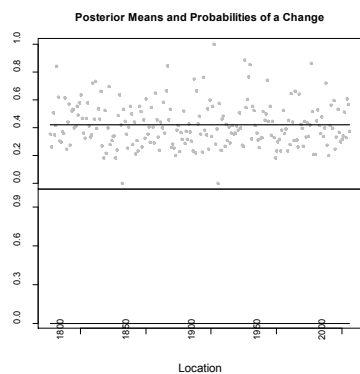
(b) M20

Fig. 3. Estimated Posterior Means and Posterior Probabilities of Each Statistical Rainfall Characteristics using BH Bayesian Changing Points Estimation Method for AUG

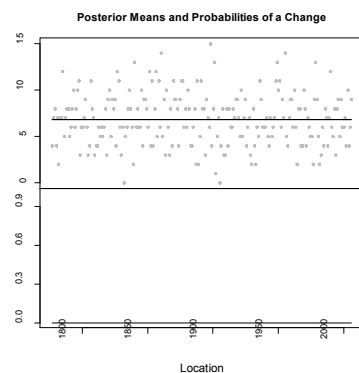
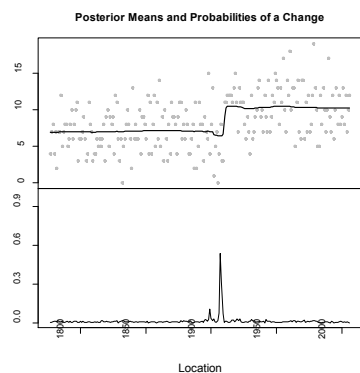
Monthly rainfall



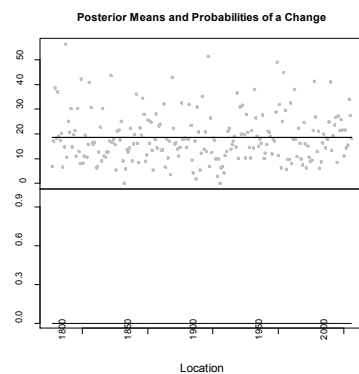
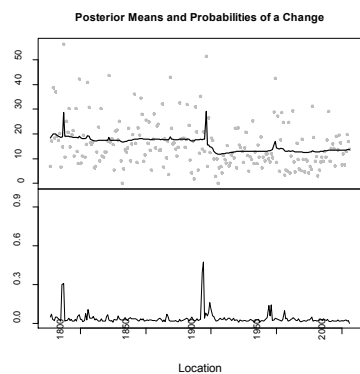
D_{\max} ratio



$N_{rainy\ days}$



$I_{rainy\ days}$



(a) M00

(b) M20

Fig. 4. Estimated Posterior Means and Posterior Probabilities of Each Statistical Rainfall Characteristics using BH Bayesian Changing Points Estimation Method for SEP

5. 결 론

어떠한 자연 현상도 완전히 동일하게 재현되지 않기 때문에 시간적인 순서를 고려하지 못하는 통계적 방법은 구체적인 발생시간을 예측하기 어렵다는 점에서 자료의 변화량의 파악은 가능하나 변화시점을 명확히 파악하는데 한계가 있다. 이러한 배경에서, 본 논문에서는 각각의 시점에서의 변화확률을 산정하여 변화시점을 추정하는 베이시안 변화시점 추정방법(Bayesian changing points estimation method)을 이용하여 측우기 관측자료계열(CWK, 1777~1907년)과 근대우량계 관측자료계열(MRG, 1908~2006년) 사이의 변화에 대한 상대확률적 절점의 발생여부를 분석하였다. 상대적으로 월강우량이 커서 자료의 신뢰도가 비교적 높다고 추정되는 6월에서 9월까지의 결과에 대해 분석한 결론은 다음과 같다.

Monthly rainfall 은, M00과 M20 모두, 모든 달에서 두자료계열의 경계(1907년과 1908년)를 전후로 지속성 있는 뚜렷한 변화시점은 나타나지 않았다. D_{max} ratio는 6월, 7월, 8월, 9월에서 M00과 M20 모두 두자료계열의 경계(1907년과 1908년)를 전후로 지속성 있는 뚜렷한 변화시점이 나타나지 않았다. $N_{rainy\ days}$ 는 M00의 경우 모든 달의 두자료계열의 경계에서 지속성 있는 뚜렷한 변화시점이 나타났다. 특히 6월은 7월, 8월, 9월에 비해 보다 명확한 변화양상을 보였다. M20의 경우 6월은 두자료계열의 경계에서 약한 변화양상을 보였다. 특히 6월의 경우는 상대적으로 변화시점을 전후로 지속성이 있었고 1900년을 전후로 사후평균이 감소한 경향을 보였다. 그 외 7월에서 9월까지의 두자료계열의 경계에서 변화시점이 나타나지 않았다. $I_{rainy\ days}$ 는 M00과 M20 모두 6월은 두자료계열의 경계에서 변화시점이 나타나지 않았다. M00의 7월, 8월, 9월은 두자료계열의 경계에서 약한 변화양상이 나타나고 있고 변화시점 전후로 지속성도 있었다. 특히 9월은 상대적으로 뚜렷한 변화시점을 보였다. 그러나 M20의 7월, 8월, 9월, 두자료계열의 경계에서 변화시점이 나타나지 않았다. 따라서 $I_{rainy\ days}$ 는 강우특성 자체의 변화라기 보다는 측우기 관측자료계열과 근대우량계 관측자료계열의 관측 최소정밀도 차이에서 기인한 강우일수 산정의 기준 차이로부터 발생한 결과로 해석하는 것이 적절할 것으로 판단된다.

이러한 결과를 종합해 보면, 측우기 관측자료계열의 본질적인 특성은 근대우량계 관측자료계열과 큰 차이가 없다고 판단된다. 단, 관측정밀도 한계와 관측방식의 차이에서 기인한 강우특성의 정량적 차이가 일부 존재한다고 판단되고, 차이의 한계는 일강우량 기준 2mm 이하로 추정된다. 2mm의 차이를 감안하더라도 강우일수의 경우는 6

월에서 상대적으로 뚜렷한 변화 가능성을 보였다. 이는 6월의 경우 강우일수가 실제 변화하였을 가능성(월별 강우특성이 변화하였을 가능성)이 높음을 의미한다. 더불어, 최근 전 세계적으로 온도와 강우량의 변화가 보고되고 있는 1960년 전후 시점에서, 서울지점 강우량의 변화도 명확하진 않지만 감지되었다는 점은 주목할 만하다. 본 연구는 PPM(product partition model)에 근거한 베이시안 변화시점 분석방법을 이용하여 측우기 강우량의 신뢰도를 객관적으로 제고(提高)함으로써 측우기 강우량이 근대이전 세계 최고(最高)의 강우량 자료임을 다시 한 번 과학적으로 입증하였고, 이를 통해 가용한 강우량 자료를 100년 이상 확장할 수 있는 근거를 추가로 제시하였다는 점에서 의의가 있다고 본다.

참고문헌

- 김기욱, 유철상, 박민규, 김현준 (2007). “강우빈도해석에서의 측우기자료의 유용성 평가.” **한국수자원학회논문집**, 한국수자원학회, 제40권, 제11호, pp. 851-859.
- 김상욱, 이길성 (2008a). “Bayesian MCMC를 이용한 저수량 점 빈도해석: I. 이론적 배경과 사전분포의 구축.” **한국수자원학회논문집**, 한국수자원학회, 제41권, 제1호, pp. 35-47.
- 김상욱, 이길성 (2008b). “Bayesian MCMC를 이용한 저수량 점 빈도해석: II. 적용과 비교분석.” **한국수자원학회논문집**, 한국수자원학회, 제41권, 제1호, pp. 49-63.
- 김상욱, 이길성 (2008c). “Bayesian 다중회귀분석을 이용한 저수량(Low flow) 지역 빈도분석.” **한국수자원학회논문집**, 한국수자원학회, 제41권, 제3호, pp. 325-340.
- 김상욱, 이길성 (2008d). “베이시안 회귀분석을 이용한 수위-유량관계곡선의 불확실성 분석.” **한국수자원학회논문집**, 한국수자원학회, 제41권, 제9호, pp. 943-958.
- 이병설 (1970). “서울의 연강수량 및 하기강수량의 Normality에 관한 연구.” **한국기상학회지**, 한국기상학회, 제5권, pp. 11-14.
- 진종갑, 문병권 (1997). “측우기 강우량 자료의 복원과 분석.” **한국기상학회지**, 한국기상학회, 제33권, 제4호, pp. 691-707.
- 정현숙, 임규호 (1994). “서울 지역 월강수량 강수일의 관계, 1770-1907.” **한국기상학회지**, 한국기상학회, 제30권, 제4호, pp. 487-505.
- 조희구, 나일성 (1979). “18세기 한국의 기후변동: 강우량을 중심으로.” **동방학지**, 제22권, pp. 83-103.
- 황석환, 김중훈, 유철상, 유도근 (2009). “서울지점 강우자료의 정량적 동질성 분석.” **한국방재학회논문집**, 한국

- 방재학회, 제9권, 제4호, pp. 29-35.
- Barry, D., and Hartigan, J.A. (1992). "Product partition models for change point problems." *The annals of Statistics*, Vol. 20, No. 1, pp. 260-279.
- Barry, D., and Hartigan, J.A. (1993). "A bayesian analysis for change point problems." *Journal of the American Statistical Association*, Vol. 88, No. 421, pp. 309-319.
- Erdman, C., and Emerson, J.W. (2007b). "bcp: A Package for Performing a Bayesian Analysis of Change Point Problems." R package version 1.8.4, URL <http://CRAN.R-project.org/>.
- Kuczera, G. (1999). "Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference." *Water Resources Research*, Vol. 35, No. 5, pp. 1551-1557.
- 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.
- O'Connell, D.R.H., Ostenaar, D.A., Levish, D.R., and Klinger, R.E. (2002). "Bayesian flood frequency analysis with paleohydrologic bound data." *Water Resources Research*, Vol. 38, Issue 5, p. 1058.
- Perreault, L., Bernier, J., Bobée, B., and Parent, E. (2000a). "Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited." *Journal of Hydrology*, Vol. 235, pp. 221-241.
- Perreault, L., Bernier, J., Bobée, B., and Parent, E. (2000b). "Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting." *Journal of Hydrology*, Vol. 235, pp. 242-263.
- 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 long Pearson type III regional skew estimation." *Water Resources Research*, Vol. 41, W10419.
- 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.
- Vicens, G.J., Rodriguez-Itrube, 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.
- Wood, E.F., and Rodriguez-Itrube, 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-Itrube, I. (1975b). "Bayesian Approach to Analyze Uncertainty Among Flood Frequency Models." *Water Resources Research*, Vol. 11, No. 6, pp. 839-843.
- Yao, Y.C. (1984). "Estimation of a noisy discrete-time step function: bayes and empirical bayes approaches." *Annals of Statistics*, Vol. 12, pp. 1434-1447.

논문번호: 09-038	접수: 2009.04.01
수정일자: 2010.03.23/06.30	심사완료: 2010.06.30