

베이지안 방법을 이용한 우리나라 강수특성(1954-2007)의 변화시점 및 변화유형 분석

김찬수¹ · 서명석^{2,*}

¹공주대학교 응용수학과

²공주대학교 대기과학과

(2009년 2월 24일 접수; 2009년 5월 18일 승인)

Change-point and Change Pattern of Precipitation Characteristics using Bayesian Method over South Korea from 1954 to 2007.

Chansoo Kim¹ and Myoung-Seok Suh^{2,*}

¹Department of Applied Mathematics, Kongju National University

²Department of Atmospheric Science, Kongju National University

(Manuscript received 24 February 2009; in final form 18 May 2009)

Abstract

In this paper, we examine the multiple change-point and change pattern in the 54 years (1954-2007) time series of the annual and the heavy precipitation characteristics (amount, days and intensity) averaged over South Korea. A Bayesian approach is used for detecting of mean and/or variance changes in a sequence of independent univariate normal observations. Using non-informative priors for the parameters, the Bayesian model selection is performed by the posterior probability through the intrinsic Bayes factor of Berger and Pericchi (1996). To investigate the significance of the changes in the precipitation characteristics between before and after the change-point, the posterior probability and 90% highest posterior density credible intervals are examined. The results showed that no significant changes have occurred in the annual precipitation characteristics (amount, days and intensity) and the heavy precipitation intensity. On the other hand, a statistically significant single change has occurred around 1996 or 1997 in the heavy precipitation days and amount. The heavy precipitation amount and days have increased after the change-point but no changes in the variances.

Key words: Annual and heavy precipitation, Bayesian model selection, change-point problem, highest posterior density

1. 서론

총강수량, 강수일수 및 강수강도상에서의 변화는 곧 바로 우리 생활 또는 안전과 직결되기 때문에 많은 연구의 대상이 되어왔다. 최근에는 특히 인위적 요인에 의한 이산화탄소 등 온실기체의 급격한 증가로 지구가 온난화되면서 일본, 인도, 호주 등 많은 지역에서

집중호우의 발생 빈도 및 강도가 크게 증가되었음이 확인되고 있다 (Iwashima and Yamamoto, 1993; Easterling *et al.*, 2000; Haylock and Nicholls, 2000; Meehl *et al.*, 2000; IPCC, 2001; Roy and Balling, 2004). 또한 태풍의 발생빈도와 극한기온의 발생 빈도에서도 변화가 있었음이 보고되고 있다 (Chu and Zhao, 2004; Khaiq *et al.*, 2007). 하지만 중국 북부 지역과 같이 일부 지역에서는 강수량의 감소도 보고되고 있다 (Zhai *et al.*, 2005). 이처럼 강수현상의 변동경향과 그 강도는 지역 및 계절에 따라 다르게 나타나고 있다. 우리나라에서도 최근 장기간의 기상청의 지상관측 자료 분석에서 강수일수에는 변화가 없지만 강수현상의 큰 변동성에도 불구하고 연강수량과 집중호우의 발생빈도 및 강도가 강해

*Corresponding Author: Myoung-Seok Suh, Department of Atmospheric Science, Kongju National University.
Phone: +82-41-850-8533, fax: +82-41-856-8527
E-mail: sms416@kongju.ac.kr

지는 것으로 보고되고 있다 (최영은과 권원태, 2005; 홍기옥 외, 2006; Jung *et al.*, 2002; Ho *et al.*, 2003).

강수특성, 태풍의 발생빈도 및 극한기온의 발생 빈도 등의 변화를 탐지하기 위하여 다양한 분석방법들이 이용되어 왔다 (Perreault *et al.*, 2000a, 2000b; Ho *et al.*, 2003; Chu and Zhao, 2004; Khaiq *et al.*, 2007). 그 중에서 특히 베이지안 분석 방법은 주관적 믿음, 과거 자료, 모델의 모의실험과 새로운 정보 등과 같은 다양한 정보들을 활용하여 불확실성을 유추하는데 합리적 방법을 제공한다. Epstein (1985)은 기상자료에 대한 다양한 베이지안 패러다임과 그 응용성을 소개하였으며 Solow (1988)는 회귀모형에 근거하여 기후변화의 추론에 베이지안 방법을 적용하였다. 최근에는 변동경향과 함께 기후패턴의 급격한 변화 또는 변화시점에 대한 연구들이 수행되고 있다. 특히, Perreault *et al.* (2000a, 2000b)은 정규분포를 따르는 관측 자료들을 이용하여 하나의 변화점을 갖는 네 가지 형태의 모형, 즉, 변화가 없는 모형 (no change model), 평균에만 변화가 있는 모형(mean change), 분산에만 변화가 있는 모형(variance change)과 평균 및 분산 둘 다 변화가 있는 모형 (mean and variance changes)에 대해 베이지안 모형 선택과 추정방법을 소개하였다. 또한 Tapsoba *et al.* (2004)은 서 아프리카에서 강수량의 기상학적 변화특성을 정의하고 설명하기 위해 베이지안 분석 방법을 적용하였다.

변화점 문제에 대한 대부분의 연구는 일반적으로 두 부분으로 이루어진다. 그 첫 번째는 주어진 일련의 관측 자료에 변화점의 유무를 탐지하는 것이다. 즉, 변화가 없는 모형과 변화가 있는 모형에 대해 검정(모형 선택)을 하는 부분이다. 두 번째는 일련의 관측된 자료에서 변화점이 발생되었다는 가정 하에서 모수를 추정하는 부분으로 이루어진다. 김찬수와 서명석 (2008)은 변화유형에 대한 모형의 선택 없이 평균에만 변화를 갖는 모형 (a single mean change model)을 가정하고, 이 모형하에서 무정보 사전분포 (non-informative prior)를 이용하여 변화점에 대한 사후확률 및 변화점 이전과 이후의 평균들에 대해 추정하였다. 따라서 본 연구에서는 지난 54년간(1954-2007) 기상청의 지상 관측 자료를 이용하여 남한평균 및 집중호우 현상 (강수량, 강수일수 및 강수강도)에 대해 평균 및 분산에 변화가 없는 모형 (M_0), 평균에만 변화가 있는 모형 (M_1) 그리고 평균과 분산 모두에 변화가 있는 모형 (M_2)들 중 어느 한 모형을 선택하기 위한 베이지안 모형 선택 문제

를 다루고자 한다. 또한 선택되어진 모형하에서 변화점, 변화점 이전과 이후의 모수 추정과 신뢰도 분석을 하고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 남한평균 및 집중호우(강수량, 강수일수, 강수강도) 시계열 자료의 특성과 변화유형에 대해 간단하게 언급하고, 3장에서는 세 가지 변화유형 모형 중에서 사후확률 이용하여 하나의 모형을 선택하기 위한 베이지안 모형 선택 및 추정 방법을 소개한다. 4장과 5장에서는 남한평균 및 집중호우 강수특성 자료에 베이지안 방법을 적용하여 변화유형별 모형 선택에 대한 모수추정과 최고사후밀도 분석 결과를 제시하고 이에 근거하여 우리나라 강수특성에서의 변화유형에 대한 결론을 도출한다.

2. 자 료

본 연구에서 사용된 자료는 1954년부터 2007년까지 지난 54년간 기상청 소속 14개 지점 (Fig. 1)에서 관측한 일 강수량 자료이다. 분석시점을 1954년으로 한 것은 한국전쟁으로 인하여 서울 등 일부 지점에서 1950-1953 사이에 결측이 있었기 때문이다. 관측지점은 현재 기상청에서 운용중인 77개의 유인 관측소 중 관측기간이 54년 이상이며 상대적으로 결측이 적은 지점들로 선정하였다. 특히 강수현상이 빈번한 우기 (6월, 7월, 8월과 9월) 중 3일 이상 결측이 있는 지점들은 제외하였다. 남한평균 강수량(일수, 강도)은 전국 14개 지점의 연강수량(일수, 강도) 평균을 의미하며 이 때 강수일수는 일강수량이 0.1 mm 이상인 날 수를 의미하고 강수강

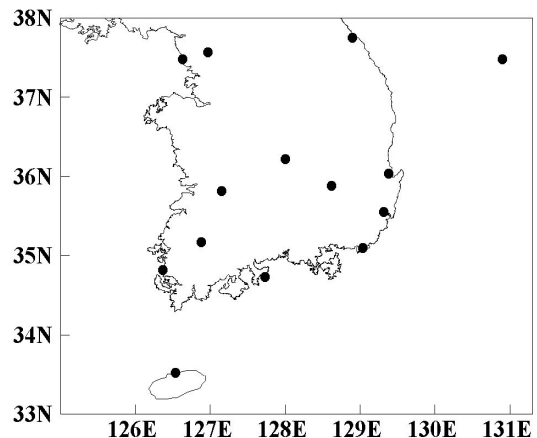


Fig. 1. Location of 14 observing stations used in this study.

도는 연강수량을 강수일수로 나눈 값이다. 현재 기상청에서 사용하는 집중호우 기준은 $80 \text{ mm } 12 \text{ h}^{-1}$ 이나 여기서는 기존의 기준인 $80 \text{ mm } 24 \text{ h}^{-1}$ 을 적용하였기 때문에 집중호우 일수는 일 강수량이 80 mm 이상인 날 수를 의미한다.

Fig. 2는 1954년부터 2007년까지 지난 54년간 기상청 소속 14개 지점에서 관측한 남한평균 및 집중호우 시의 강수량, 강수일수 및 강수강도에 대한 시계열을

보인 것이다. 남한평균 강수량은 약 887 mm 에서부터 1975 mm 까지 경년변동이 크게 나타나고 있으며 1996년 또는 1997년 이후에 기록적인 남한평균 강수량이 2회 발생하고 있다. 강수일수도 약 86일에서 131일까지 경년변동이 크게 나타나고 있지만 증가경향은 보이지 않고 있다. 남한평균 강수강도와 집중호우 현상 중 강수량 및 강수일수는 다른 연도에 비해 1990년의 중반 부근에서 변화가 있는 특성을 보이고 있다. 집중호우 강

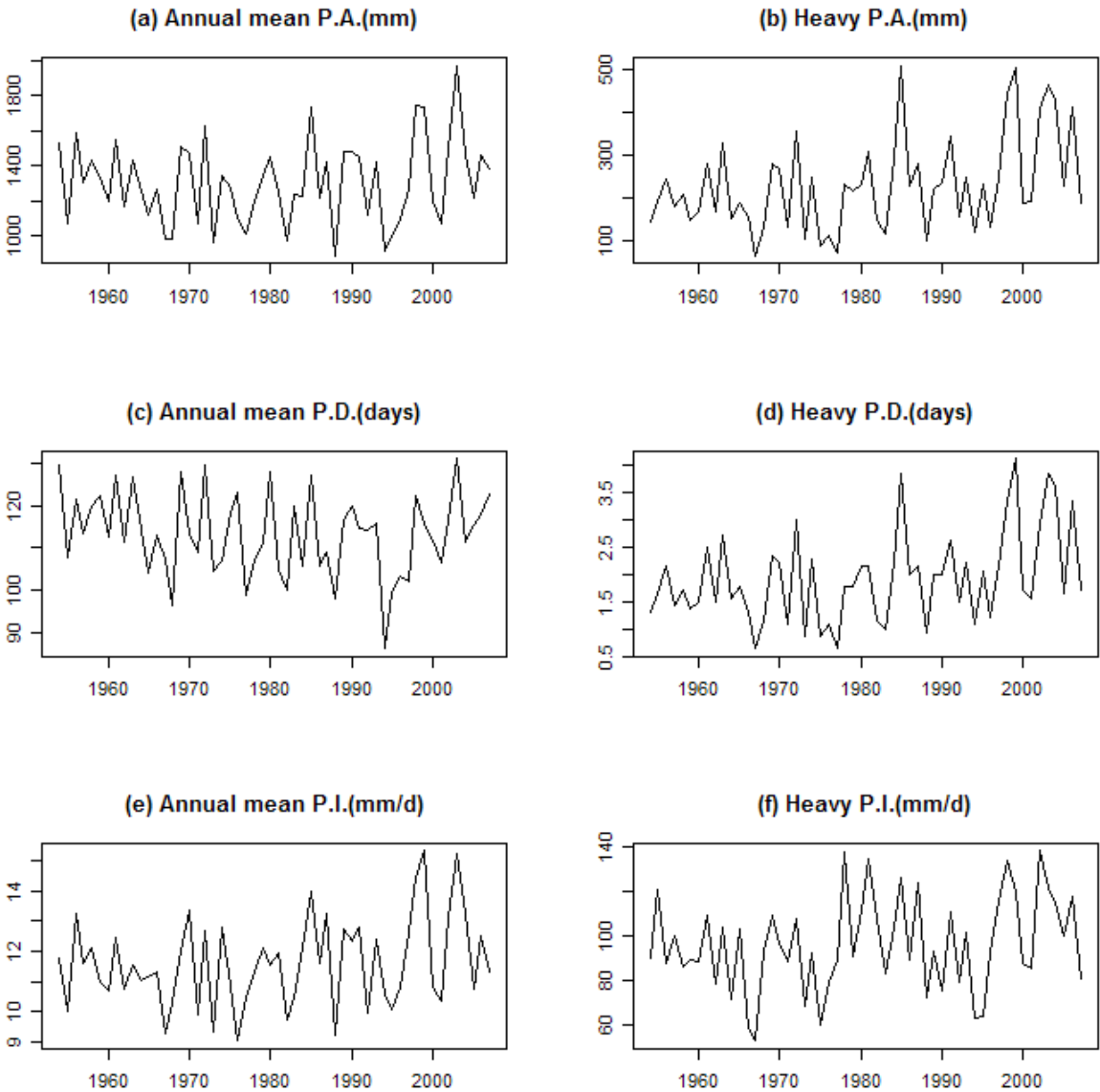


Fig. 2. Times series of the amount, days and intensity of spatially averaged precipitation and heavy precipitation over South Korea from 1954 to 2007. P.A., P.I., and P.D. denote the precipitation amount, precipitation intensity and precipitation days, respectively.

도의 경우에는 1970년대 후반에도 변화가 있는 것으로 보인다. 김찬수와 서명석 (2008)은 최근 30년간 (1976-2005)의 남한평균 강수 및 집중호우에 대한 변화시점 분석에서 집중호우 특성에 1996년을 기점으로 변화가 있음을 보인 바 있다. 자료의 기간을 1954년대로 확장한 본 그림에서도 남한평균 강수강도와 집중호우 현상(강수량, 강수일수)에서 유사한 패턴이 나타나고 있다.

강수특성의 변화 유형은 일반적으로 평균만의 변화, 분산만의 변화 및 평균과 분산의 변화로 나눌 수가 있다. 평균만이 변화될 경우에는 극한적으로 강한 강수 현상(홍수)만 증가(또는 감소)하는 반면 분산만이 변화하거나 평균과 분산 모두에서 변화가 일어날 경우에는 양쪽 극한 현상(홍수와 가뭄)이 모두 증가(감소)하게 되기 때문에 변화시점과 함께 유형도 중요하다. 여기서는 시각적 분석결과와 김찬수와 서명석 (2008)의 연구결과에 근거하여 전 자료를 1996년을 기준으로 두 집단으로 나눈 다음 사전분석의 의미로 다음과 같은 두 가지 검정을 실시하였다. 먼저 두 집단 간에 분산의 동질성을 검정하였고, 이 결과를 이용하여 두 집단의 평균에서 변화가 있었는지를 검정하였다(Table 1). 1996년을 기점으로 두 자료의 분산에서 변화가 있었는지를 검정한 결과, 남한평균 및 집중호우 강수현상에서 p-값들이 유의한 결과를 제시하지 못하고 있음을 알 수 있다. 즉, 우리나라의 남한평균 및 집중호우 강수 특성에서는 분산에 변화가 없었음을 알 수 있으며 이를 근거로 두 자료의 평균에서 차이가 있는지를 검정하였다.

남한평균 강수량 및 강수강도의 경우, 유의수준 5%

에서는 변화가 있었음 알 수 있다. 또한 남한평균 강수 일수의 경우에는 유의수준 5%에서도 유의하지 않다. 집중호우 현상의 경우, 유의수준 5%에서는 모두 유의한 결과를 보이고 있다. 집중호우 일수의 경우 평균과 함께 분산에서도 변화가 있을 수 있지만 유의수준이 높지 않기 때문에 변화시점을 기준으로 분산의 변화보다는 평균에 변화가 있는 다음과 같은 모형을 고려해 볼 수 있다. 남한평균 강수량 및 강수일수와 집중호우 강수강도는 변화가 없는 모형, 남한평균 강수강도와 집중호우 강수량 및 강수일수는 평균에 변화가 있는 모형이 선택될 가능성이 있음을 추론해 볼 수 있다. 따라서 본 연구에서는 세 가지 모형(변화유형), 평균에 변화가 없는 모형(M_0), 평균에만 변화가 있는 모형(M_1)과 평균과 분산 모두에 변화가 있는 모형(M_2) 중에서 하나의 모형을 선택하기 위하여 베이지안 추정기법을 이용하여 모형 선정을 시도하였다. 베이지안 모형 선택의 기본 원리는 각 모형에 대한 사후확률을 계산한 후, 이 중 가장 높은 사후확률을 선택함으로써 이루어진다. 비교하고자 하는 모형이 여러 개인 경우, 각 모형이 선택될 직접적인 확률을 제공하므로 모형 선택 시 유용한 방법을 제공한다.

3. 베이지안 모형 선택

여기서는 정규분포를 따르는 확률표본 $X=(X_1, X_2, \dots, X_n)$ 내에 하나의 변화점이 존재하는지를 검토하기 위해 베이지안 모형 선택 방법을 적용하고자 한다. 남한평균 및 집중호우 강수특성(강수량, 강수일수, 강수강도) 자료에 대한 정규성을 검토하기 위해 정규 확률 플롯

Table 1. Testing results for the differences in the mean and variance after the assumed change-point (1996).

Annual/Heavy precipitation		Period	Mean	Std	Test $H_0: \mu_1 = \mu_2$	Test $H_0: \sigma_1 = \sigma_2$
					p-value	p-value
A.P.A. (mm)	1954-1996	1268.0	215.2	0.017	0.270	
	1997-2007	1458.0	273.7			
A.P.D. (days)	1954-1996	112.69	10.16	0.370	0.408	
	1997-2007	115.67	7.94			
A.P.I. (mm d ⁻¹)	1954-1996	11.348	1.249	0.005	0.109	
	1997-2007	12.710	1.778			
H.P.A. (mm)	1954-1996	204.4	88.2	0.000	0.107	
	1997-2007	335.5	125.8			
H.P.D. (days)	1954-1996	1.730	0.674	0.000	0.078	
	1997-2007	2.773	0.994			
H.P.I. (mm d ⁻¹)	1954-1996	91.22	21.63	0.011	0.757	
	1997-2007	110.24	19.46			

과 Anderson-Darling (1952) 검정을 수행하였다. 대체로 남한평균 강수현상의 경우는 정규성을 만족하였으나, 집중호우의 경우는 정규성을 만족한다고 볼 수 없었다. 따라서 정규성을 만족하도록 자료를 로그 변환한 후 이용하였다. 로그 변환 전과 후에 대한 정규 확률 플롯이 Fig 3에 주어져 있다.

먼저, $x = (x_1, x_2, \dots, x_n)$ 을 확률변수 $X = (X_1, X_2, \dots, X_n)$ 의 일련의 관측치라고 하자. 관측치가 주어진 경우, 임의의 시간에서 하나의 변화점이 있다고 가정하고 연강수 및 집중호우 강수특성(강수량, 강수일수 및 강수강도)에 어떤 변화가 있는지를 검토하기 위해 다음과 같은 세 가지 변화유형의 모형, 즉 변화가 없는 모형 (M_0), 평균에만 변화가 있는 모형(M_1)과 평균 및 분산 모두에 변화가 있는 모형(M_2)을 고려하고자 한다.

모형 M_0 :

$$X_i \sim N(x_i | \mu, \sigma^2), \quad i = 1, 2, \dots, n$$

모형 M_1 :

$$X_i \sim N(x_i | \mu_1, \sigma^2), \quad i = 1, 2, \dots, \tau_1$$

$$X_i \sim N(x_i | \mu_2, \sigma^2), \quad i = \tau_1 + 1, \dots, n$$

모형 M_2 :

$$X_i \sim N(x_i | \mu_1, \sigma_1^2), \quad i = 1, 2, \dots, \tau_2 \tag{1}$$

$$X_i \sim N(x_i | \mu_2, \sigma_2^2), \quad i = \tau_2 + 1, \dots, n$$

여기서 $N(\cdot | \mu, \sigma^2)$ 는 평균이 μ 이고 분산이 σ^2 인 정규분포를 의미하며, $\tau_k, k = 1, 2$ 는 모수이며 모형 M_k 에 대응하는 변화점이다.

본 연구에서는 위에서 언급한 세 가지 모형 중 하나의 모형을 선택하기 위해, 베이즈 인자(Bayes factor) 혹은 사후확률(posterior probability)을 이용한 베이지안기법을 적용하고자 한다. 베이지안 모형 선택에 대한 논의는 Perrault *et al.* (2000b), Tapsoba *et al.* (2004) 과 Son과 Kim (2005) 등에 의해 연구되어졌다. 특히, Son과 Kim (2005)은 다변량 정규분포를 갖는 일련의 관측치에서 무정보 사전분포를 이용하여 평균의 변화, 공분산의 변화 혹은 평균과 공분산의 변화를 선택하기 위해 내재적 베이즈 인자(intrinsic Bayes factor)를 이용한 베이지안 방법을 적용하였다.

베이지안 모형 선택 혹은 가설검정을 수행하기 위해서는 식 (1)에서 모수들에 대한 사전분포가 필요하다. 사전분포를 명시하기 위해서는 전문가의 지식이나

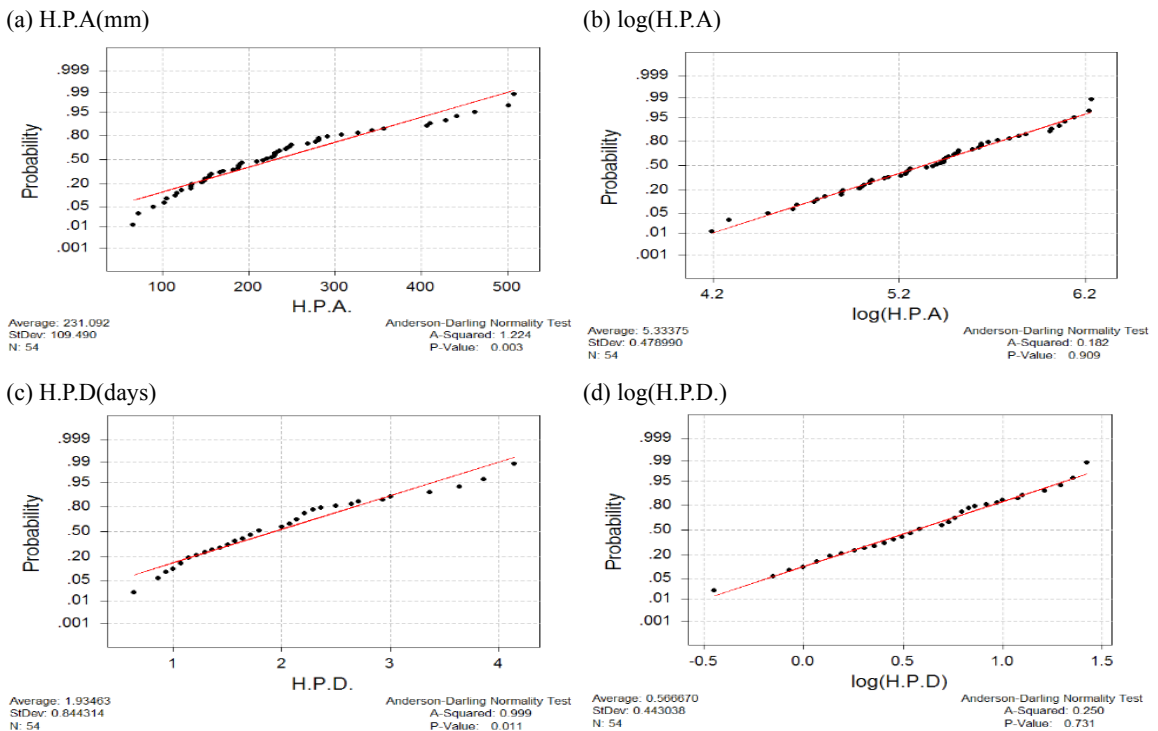


Fig. 3. Probability plot of the heavy precipitation amount and days before and after log transformation.

과거자료 혹은 지역적 특성 등의 정보를 활용하여 사전분포를 유도해야 한다. 그러나 이러한 작업을 수행하는 것은 쉽지가 않다. 또한 각 모형의 모수들에 대해서 알려져 있는 정보가 거의 없기 때문에, 정보가 거의 없는 무정보 사전분포(non-informative prior distribution)를 가정하였다. 무정보 사전분포는 공액사전(conjugate prior)과는 달리 공액사전 분포의 초모수(hyper-parameter)들에 대해 유도를 할 필요가 없기 때문에, 베이즈 인자를 사용하는데 보다 단순하고 자동적인 결과를 제공한다. 모형 M_{k, τ_k} ($k = 1, 2$)를 모형 (1)에서 고정된 변화점 τ_k 를 갖는 모형이라고 하자. 또한, 각 모형 M_0 와 M_{k, τ_k} 에서 $\theta_0 = (\mu, \sigma^2)$, $\theta_1 = (\mu_1, \mu_2, \sigma^2)$, $\theta_2 = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ 라하고, $L(\theta_0|x)$ 와 $L(\theta_k, \tau_k|x)$ 를 모형 M_0 와 M_{k, τ_k} 에서의 우도함수라 하자. 이때, 모형 M_0 에 대한 모형 M_{k, τ_k} 의 베이즈 인자는 다음과 같이 정의된다.

$$B_{k0}(\tau_k) = \frac{m_k(x, \tau_k)}{m_0(x)}, \quad k = 1, 2 \quad (2)$$

여기서

$$m_0(x) = \int_{\Theta} L(\theta_0|x) \pi(\theta_0) d\theta_0, \quad (3)$$

$$m_k(x, \tau_k) = \int_{\Theta_k} L(\theta_k, \tau_k|x) \pi(\theta_k, \tau_k) d\theta_k. \quad (4)$$

Θ_k 는 θ_k 의 모수 공간이고 $\pi_k(\theta_k, \tau_k)$ 는 각 모형의 모수들에 대한 무정보 사전분포이고, $m_0(x)$ 와 $m_k(x, \tau_k)$ 는 각각 M_0 와 M_{k, τ_k} 하에서의 주변 확률밀도함수(marginal probability density function)이다.

무정보 사전분포가 사용되었기 때문에, $B_{k0}(\tau_k)$ 는 임의의 알려지지 않은 상수에 의존하게 되어 베이즈 인자를 계산할 때 문제가 발생된다. 이러한 문제를 해결하기 위해 Berger와 Pericchi (1996)가 제시한 내재적 베이즈 인자(intrinsic Bayes factor: IBF)를 적용하고자 한다. IBF는 데이터 분할 기법으로써 최소 트레이닝 표본(minimal training sample: MTS)를 이용하는 기법이다. MTS는 식 (3)과 (4)에서 주어진 주변 확률밀도함수가 유한(finite)하게 되는 최소 크기의 표본을 의미한다.

$x(l)$ 을 l 번째 뽑힌 MTS라 하면, 모형 M_0 에 대한 고

정된 변화점 τ_k 를 갖는 모형 M_{k, τ_k} 의 IBF는 다음과 같이 정의된다.

$$\begin{aligned} B_{k0}^{IBF}(\tau_k) &= B_{k0}(x, \tau_k) B_{0k}(x(l), \tau_k) \\ &= \frac{m_k(x, \tau_k)}{m_0(x)} \frac{m_0(x(l))}{m_k(x(l), \tau_k)}. \\ B_{k0}^{IBF}(\tau_k) \end{aligned}$$

여기서, $m_k(x(l), \tau_k)$ 는 식 (4)에서 전체 표본 x 대신에 $x(l)$ 을 대입함으로써 정의되어지고, $B_{k0}^{IBF}(\tau_k)$ 는 임의의 상수에 의존하지 않는다. 그러나 IBF는 MTS를 사용하므로 추출된 MTS에 의존하게 된다. 이러한 종속성을 제거하기 위해, 여러 개의 MTS를 사용하여 얻어진 산술평균 IBF(AIBF), 기하평균 IBF(GIBF) 또는 중앙값 IBF(MIBF)를 이용하여 모형선택에 사용할 수 있다. AIBF, GIBF와 MIBF는 다음과 같이 정의된다.

$$\begin{aligned} B_{k0}^{AIBF}(\tau_k) &= B_{k0}(x, \tau_k) \frac{1}{L} \sum_{l=1}^L B_{0k}(x(l), \tau_k), \\ B_{k0}^{GIBF}(\tau_k) &= B_{k0}(x, \tau_k) \left[\prod_{l=1}^L B_{0k}(x(l), \tau_k) \right]^{\frac{1}{L}}, \\ B_{k0}^{MIBF}(\tau_k) &= B_{k0}(x, \tau_k) \text{median}_{1 \leq l \leq L} B_{0k}(x(l), \tau_k). \end{aligned} \quad (5)$$

여기서 $x(l)$ 은 l 번째 MTS는 의미한다.

두 개 이상인 모형들 중에서 하나의 모형을 선택하기 위해서는 베이즈 인자를 이용하는 것 보다는 사후확률을 이용하는 것이 보다 더 유용하다. 따라서 식 (5)의 AIBF를 이용하면 모형 M_{k, τ_k} 의 사후확률은 베이즈 정리에 의해 다음과 같이 계산되어진다.

$$P(M_{k, \tau_k}|x) = \frac{B_{k0}^{AIBF}(\tau_k) P(M_{k, \tau_k})}{P(M_0) + \sum_{\tau_1} B_{10}^{AIBF}(\tau_1) P(M_{1, \tau_1}) + \sum_{\tau_2} B_{20}^{AIBF}(\tau_2) P(M_{2, \tau_2})} \times \quad (6)$$

궁극적으로 모형 M_k 에 대한 사후확률은 식 (6)에서 모수 τ_k 에 대해 주변화함으로써 다음과 같이 얻어진다.

$$P(M_k|x) = \sum_{\tau_k} P(M_{k, \tau_k}|x), \quad k = 1, 2 \quad (7)$$

식 (6)에서 AIBF대신에 GIBF 혹은 MIBF를 대입하면 유사한 결과를 얻을 수 있다. 궁극적으로 식 (7)을 계산

하기 위해서 다음과 같은 무정보 사전분포를 가정하였다.

$$\begin{aligned} \pi_0(\mu, \sigma^2) &= c_0 \frac{1}{\sigma^2}, \\ \pi_1(\mu_1, \mu_2, \sigma^2) &= c_1 \frac{1}{\sigma^2}, \\ \pi_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) &= c_2 \frac{1}{\sigma_1^2} \frac{1}{\sigma_2^2}. \end{aligned} \quad (8)$$

여기서 $c_k=0, 1, 2$ 는 임의의 알려지지 않은 상수이다. 각각의 모형으로부터 우도함수 $L(\theta_i|x), i=0, 1, 2$ 는 다음과 같다.

$$\begin{aligned} L(\theta_0|x) &= L(\mu, \sigma^2|x) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) \end{aligned} \quad (9)$$

$$\begin{aligned} L(\theta_1, \tau_1|x) &= L(\mu_1, \mu_2, \sigma^2, \tau_1|x) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{\tau_1} (x_i - \mu_1)^2 \right. \right. \\ &\quad \left. \left. - \sum_{i=\tau_1+1}^n (x_i - \mu_2)^2 \right] \right) \end{aligned} \quad (10)$$

$$\begin{aligned} L(\theta_2, \tau_2|x) &= L(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \tau_2|x) \\ &= (2\pi)^{-n/2} (\sigma_1^2)^{-\tau_2/2} (\sigma_2^2)^{-(n-\tau_2)/2} \\ &\quad \times \exp\left(-\frac{1}{2\sigma_1^2} \sum_{i=1}^{\tau_2} (x_i - \mu_1)^2 \right. \\ &\quad \left. - \frac{1}{2\sigma_2^2} \sum_{i=\tau_2+1}^n (x_i - \mu_2)^2 \right) \end{aligned} \quad (11)$$

식(8)과 우도함수를 이용하여, 주변밀도함수 $m_0(x), m_1(x, \tau_1)$ 과 $m_2(x, \tau_2)$ 는 다음과 같이 계산되어진다.

$$m_0(x) = c_0 \frac{\Gamma\left(\frac{n-1}{2}\right)}{n^{\frac{1}{2}} \pi^{\frac{n-1}{2}} \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^{\frac{n-1}{2}}}, \quad (12)$$

$$\begin{aligned} m_1(x, \tau_1) &= c_1 \frac{P(\tau_1) \Gamma\left(\frac{n-2}{2}\right)}{\pi^{\frac{n-2}{2}} \tau_1^{\frac{1}{2}} (n-\tau_1)^{\frac{1}{2}}} \\ &\quad \times \frac{1}{\left[\sum_{i=1}^{\tau_1} (x_i - \bar{x}_{\tau_1})^2 + \sum_{i=\tau_1+1}^n (x_i - \bar{x}_{n-\tau_1})^2 \right]^{\frac{n-2}{2}}} \end{aligned} \quad (13)$$

$$\begin{aligned} m_2(x, \tau_2) &= c_2 \frac{P(\tau_2) \Gamma\left(\frac{\tau_2-1}{2}\right) \Gamma\left(\frac{n-\tau_2-1}{2}\right)}{\pi^{\frac{n-2}{2}} \tau_2^{\frac{1}{2}} (n-\tau_2)^{\frac{1}{2}}} \\ &\quad \times \frac{1}{\left[\sum_{i=1}^{\tau_2} (x_i - \bar{x}_{\tau_2})^2 \right]^{\frac{\tau_2-1}{2}} \left[\sum_{i=\tau_2+1}^n (x_i - \bar{x}_{n-\tau_2})^2 \right]^{\frac{n-\tau_2-1}{2}}} \end{aligned} \quad (14)$$

여기서 $\bar{x} = \sum_{i=1}^n x_i/n, \bar{x}_{\tau_k} = \sum_{i=1}^{\tau_k} x_i/\tau_k$ 이고 $\bar{x}_{n-\tau_k} = \sum_{i=\tau_k+1}^n x_i/(n-\tau_k), k=1, 2$ 이고 $P(\tau_k)$ 는 변화점 τ_k 에 대한 사전분포이다.

식(5)에서 각각의 IBF를 계산하기 위해, MTS의 표본 크기를 알고 있어야 한다. 모형 M_0 와 M_{1, τ_1} 는 모형 M_{2, τ_2} 의 특정한 모형이므로 모형 M_{2, τ_2} 의 MTS의 크기만 고려하면 된다. 모형 M_{2, τ_2} 에서 MTS의 크기는 4이다. 왜냐하면, $\tau_2 \geq 2$ 와 $n - \tau_2 \geq 2$ 를 만족하면 모형 M_{2, τ_2} 의 모수들은 추정 가능하고, 식(3)과(4)는 유한한 주변밀도함수를 갖는다. 따라서 MTS를 이용하여, 주변 확률밀도함수 $m_0(x(t))$ 와 $m_k(x(t), \tau_k), k=1, 2$ 는 식(12), (13)과 식(14)에서 x 대신에 $x(t), n$ 대신에 4, τ_k 와 $n - \tau_k$ 대신에 각각 2를 대입하면 쉽게 얻어질 수 있다. 마지막으로 변화점 $\tau_k, k=1, 2$ 에 대한 사전분포 $P(\tau_k)$ 는 MTS에서 주변 확률밀도함수가 유한하도록 다음과 같이 정의할 수 있다.

$$P(\tau_k) = \frac{1}{n-3}, \quad \tau_k=2, 3, \dots, n-2, \quad k=1, 2.$$

따라서 식(7)을 이용하여 각각의 모형 중에서 사후확률이 가장 큰 모형을 선택할 수 있다.

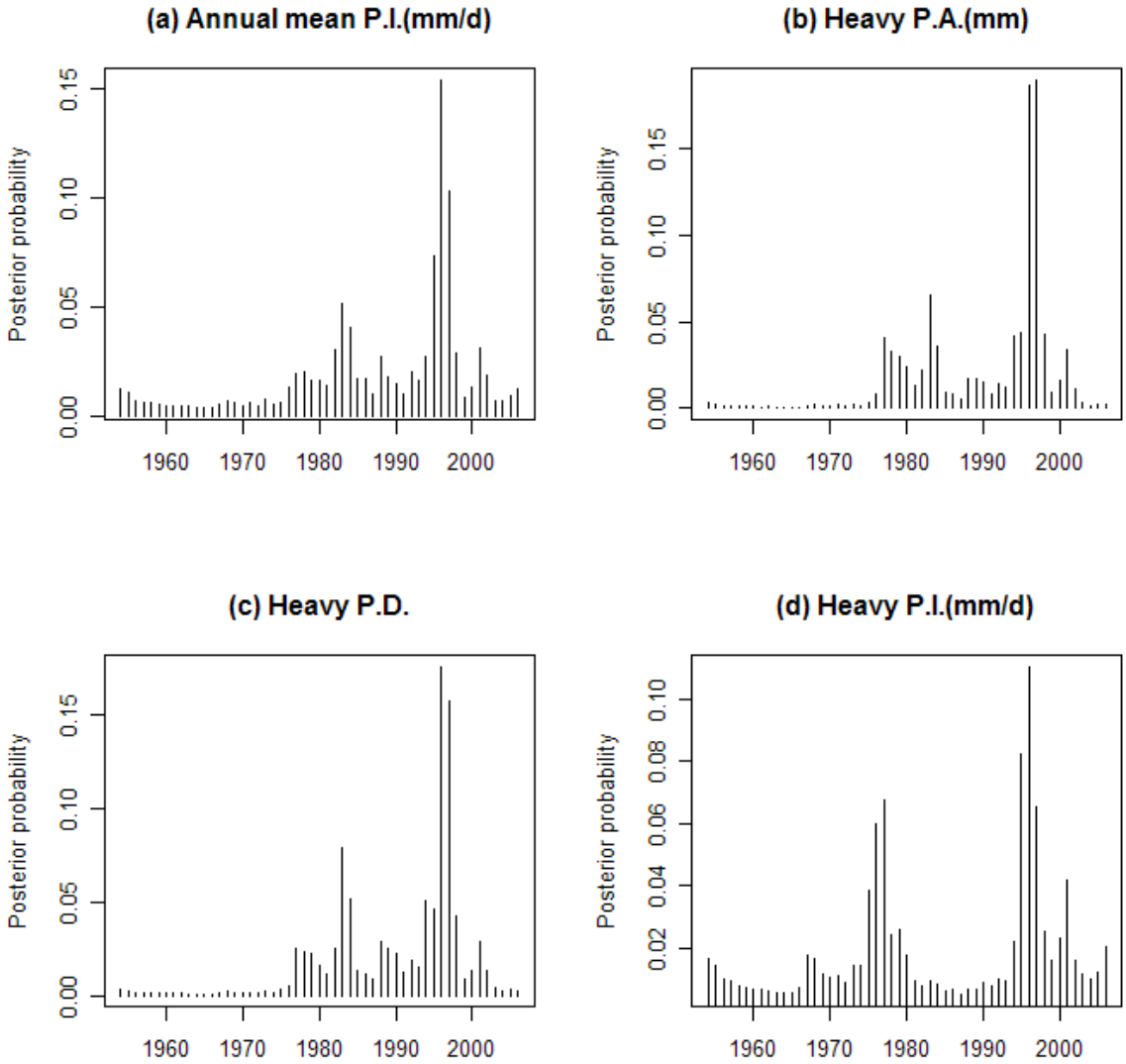


Fig. 4. Marginal posterior probability of the change-point $P(\tau|x)$ for annual precipitation intensity and amount, days and intensity of heavy precipitation.

4. 결 과

지난 54년간 우리나라 강수에서 남한평균 및 집중호우시의 강수량, 강수일수, 강수강도의 자료를 이용하여 변화점이 존재하는지를 검토하고, 변화점이 존재한다면 어떤 유형의 변화가 발생하였는지를 분석하고자 한다.

세 가지 변화모형 M_0, M_1 과 M_2 중에서 어떤 모형(변화 유형)이 타당성이 있는지를 검토하기 위해 무정보 사전 분포를 가정하였고, AIBF, GIBF와 MIBF를 이용하여 최적의 적합모형을 식 (7)을 사용하여 판별하였다. MTS는 임의의 변화점 τ_k 를 기준으로 변화점 이전과 이후 양쪽에

서 각각 두 개의 MTS를 단순임의 추출을 통해 선택하였고, 모형에 대한 사전분포는 동일하게 가정하였다.

Table 2에 사후확률을 이용한 모형 선택의 결과가 제시되어 있다. 남한평균 강수현상에 있어, 강수량과 강수일수는 AIBF, MIBF와 GIBF의 베이즈 인자를 이용한 모든 사후확률이 모형 M_0 을 선호하고 있음을 볼 수 있다. 즉, 남한평균 강수량 및 강수일수는 변화가 없는 모형을 선호하고 있다. 반면에 남한평균 강수강도의 경우, AIBF를 이용한 사후확률에서는 모형 M_1 의 사후확률이 모형 M_0 의 사후확률 보다 약간 크게 나왔지만, MIBF와 GIBF를 이용한 사후확률은 모형 M_0 의

사후확률이 크다는 것을 알 수 있다. 이러한 결과는 변화시점을 기준으로 평균에 변화가 있음을 암시하고 있으나 그 차이가 크게 나타난다고는 볼 수 없다. 이러한 결과는 90% 최고사후밀도(highest posterior density: HPD) 신뢰구간 등의 추후 분석에서도 유사하게 나타나고 있다. 집중호우 현상에서, 강수량과 강수일수는 모형 M_1 의 사후확률이 다른 모형의 사후확률 보다 확연히 크게 나타나기 때문에, 평균에 변화가 있는 모형 M_1 이 적합하다는 것을 알 수 있다. 그러나 집중호우강도의 경우 AIBF, MIBF와 GIBF 모두 모형 M_0 을 선호하고 있어 변화가 없는 모형이 선택되어진다.

세 모형 중에서 평균에만 변화가 있는 모형 M_1 이 적어도 1회 이상 선택된 남한평균 강수강도, 집중호우시의 강수량 및 강수일수, 그리고 집중호우 강도에 대해 변화 시점, 90% HPD, 변화시점을 기준으로 이전과 이후의 두 평균 차에 대한 사후확률의 분포는 김찬수와 서명석 (2008)의 결과를 이용하여 분석하였다. 먼저, 남한평균 강수강도 및 집중호우 현상의 변화시점 τ 에 대한 주변사후확률 $P(\tau|x)$ 가 Fig. 4에 주어져 있다. 남한평균 강수강도와 집중호우 현상의 변화점에 대한 베이시안 추정엔 변화점의 사후확률이 가장 크게 갖는 시점을 선택하면 된다. 먼저, 집중호우의 강수량 및 강수일수는 1996년 혹은 1997년을 기점으로 평균에 변

화가 있음을 알 수 있다. 남한평균 및 집중호우의 강수강도는 모형 선택에서 변화가 없는 모형이 선호되었기 때문에 비록 1996년에 변화가 있음을 알 수 있지만 그 변화가 크지 않음을 추론할 수 있다. 다중 변화점(multiple change-points)들이 존재하는지를 검토하기 위해 단순히 변화점 이전과 이후의 두 집단으로 나누어서 앞에서 언급한 방법을 반복적으로 적용하였다. 즉, 집중호우 강수량과 강수일수 자료를 1954년부터 변화점까지 한집단으로, 변화점에서 2007년까지 또 하나의 집단인 두 개의 부집단(subgroup)으로 나눈 다음에 각각의 부집단에 대해 변화점이 존재하는지를 동일한 방법으로 검토하였다. 그 중에서 자료 기간이 상대적으로 긴 변화점 이전 자료, 즉, 1954년부터 변화시점까지의 부집단에 대해 추가의 변화점이 존재하는지에 대한 모형 선택에 대해 조사한 결과 추가의 변화점이 있는 모형은 선택되지 않았다.

Fig. 5에서 남한평균 강수강도 및 집중호우(강수량, 빈도수 및 강수강도) 특성의 변화점 이전과 이후의 모평균 μ_1 과 μ_2 에 대한 주변 사후확률밀도 함수가 각각 주어져 있다. 검은 실선은 변화점 이전의 주변사후확률밀도함수 $\pi(\mu_1|x)$ 이고 점선은 변화점 이후의 주변사후확률밀도함수 $\pi(\mu_2|x)$ 이다. 만약에 변화점을 기점으로 평균에 변화가 매우 크다면 두 사후확률밀도함수는 서

Table 2. Posterior probabilities of each model for the precipitation characteristics according to the Bayes factor calculated by AIBF, MIBF and GIBF, respectively.

Annual/Heavy precipitation Bayes Factor	Selected model	Posterior probability		
		$P(M_0 x)$	$P(M_1 x)$	$P(M_2 x)$
A.P.A.(mm)	AIBF M_0	0.5774156	0.2953532	0.1272312
	MIBF M_0	0.5950869	0.3022562	0.1026569
	GIBF M_0	0.6536319	0.2572194	0.0891486
A.P.D.(days)	AIBF M_0	0.5604926	0.3681409	0.0713665
	MIBF M_0	0.6585685	0.2644173	0.0770141
	GIBF M_0	0.6784105	0.2610729	0.0605165
A.P.I.(mm d ⁻¹)	AIBF M_1	0.3670212	0.3970232	0.2359557
	MIBF M_0	0.4005909	0.3708801	0.2285290
	GIBF M_0	0.4671583	0.3448671	0.1879746
H.P.A.(mm)	AIBF M_1	0.1196030	0.7766733	0.1037238
	MIBF M_1	0.1546884	0.7331210	0.1121907
	GIBF M_1	0.1814827	0.7061607	0.1123566
H.P.D.(days)	AIBF M_1	0.1549016	0.7312202	0.1138782
	MIBF M_1	0.1684553	0.7327572	0.0987875
	GIBF M_1	0.2334299	0.7665701	0.0
H.P.I.(mm d ⁻¹)	AIBF M_0	0.5178669	0.3752859	0.1068472
	MIBF M_0	0.5657178	0.3625038	0.0717784
	GIBF M_0	0.6314142	0.3137898	0.0547960

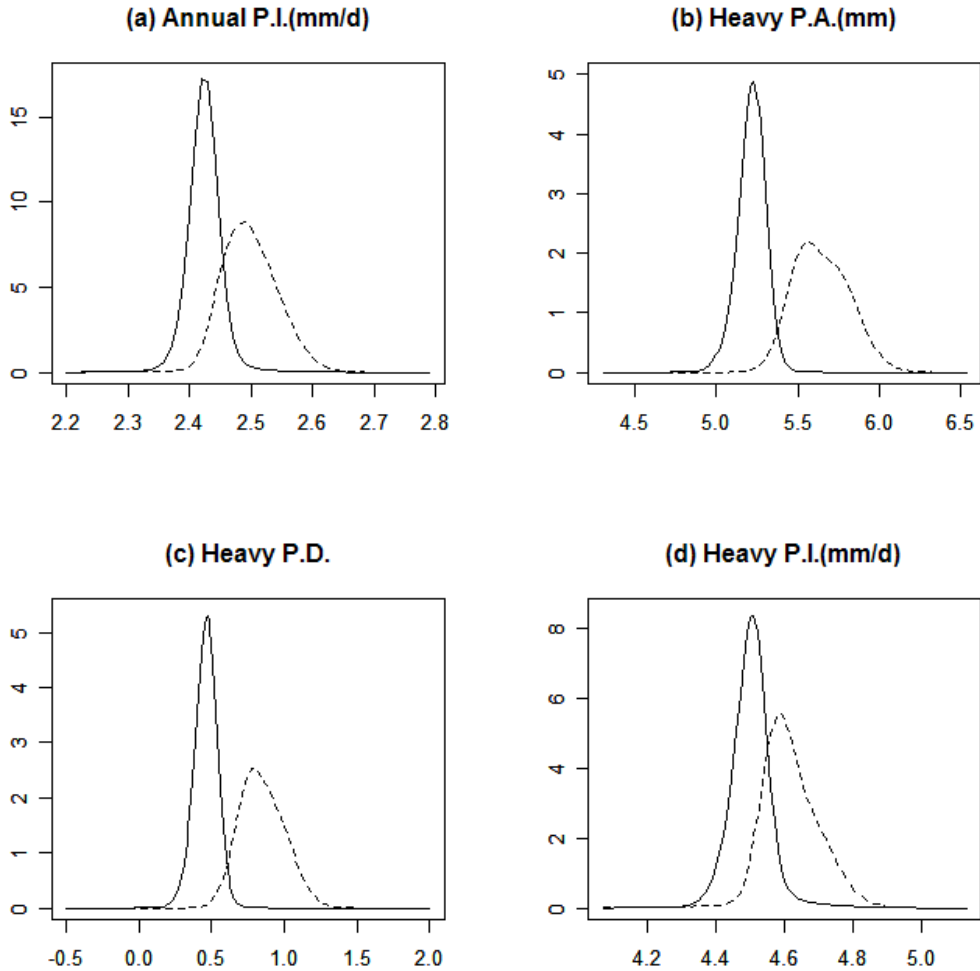


Fig. 5. Marginal posterior density functions $\pi(\mu_i|x)$, $i=1,2$ for the annual and log-transformed heavy precipitation characteristics. The solid line is a marginal posterior density $\pi(\mu_1|x)$ before the change-point and the dotted line is that of $\pi(\mu_2|x)$ after the change-point, respectively.

로 겹치는 부분이 거의 없게 되지만, 변화가 작다면 상당 부분 겹치게 될 것이다. Fig. 5에서 보는 것과 같이, 집중호우 강수량과 강수일수 경우, 변화점 이전과 이후의 두 주변사후분포 사이에 겹치는 부분이 적다는 것을 알 수 있고 이것은 변화점 이전과 이후의 평균에서 변화가 크다는 것을 알 수 있다. 반면에 집중호우 강수강도의 경우 두 확률분포는 많은 부분이 중첩되어 있어 평균의 변화가 적다는 것을 알 수 있다. 또한 남한 평균 강수강도의 경우도 상당 부분 중첩되어 있어, 평균에 변화는 있지만 그 크기가 작기 때문에 모형 선택에서 M_1 보다는 M_0 모형이 선택될 가능성이 많다는 것을 알 수 있다. 이러한 결과들은 앞에서의 모형 선택 결과와 동일한 결과를 제시하고 있다.

Fig. 6에서는 변화점 이전과 이후의 두 평균차 $\delta = \mu_2 - \mu_1$ 에 대한 주변사후분포 $\pi(\delta|x)$ 와 90% HPD 신뢰영역이 주어져 있다. 집중호우시의 강수량과 일수의 경우, 변화점 이전과 이후의 두 평균 차에 대한 90% HPD 신뢰구간이 0을 포함하고 있지 않은 반면에 집중호우 강도는 90% HPD 신뢰영역이 0을 포함하고 있어 모형 선택에서와 같이 동일한 결과를 제시하고 있다. 남한평균 강수강도의 경우 90% HPD 신뢰구간의 하한이 0을 포함하지는 않지만, 거의 0에 가까운 값을 가지고 있어 변화의 경향은 있지만 크지 않다는 것을 제시한다. 또한 집중호우의 현상 중 강수량과 강수일수의 경우, 사후확률(Table 3)의 p-값이 유의한 결과(<0.05)를 주고 있으므로 변화점 이전 보다는 이후에 평균이

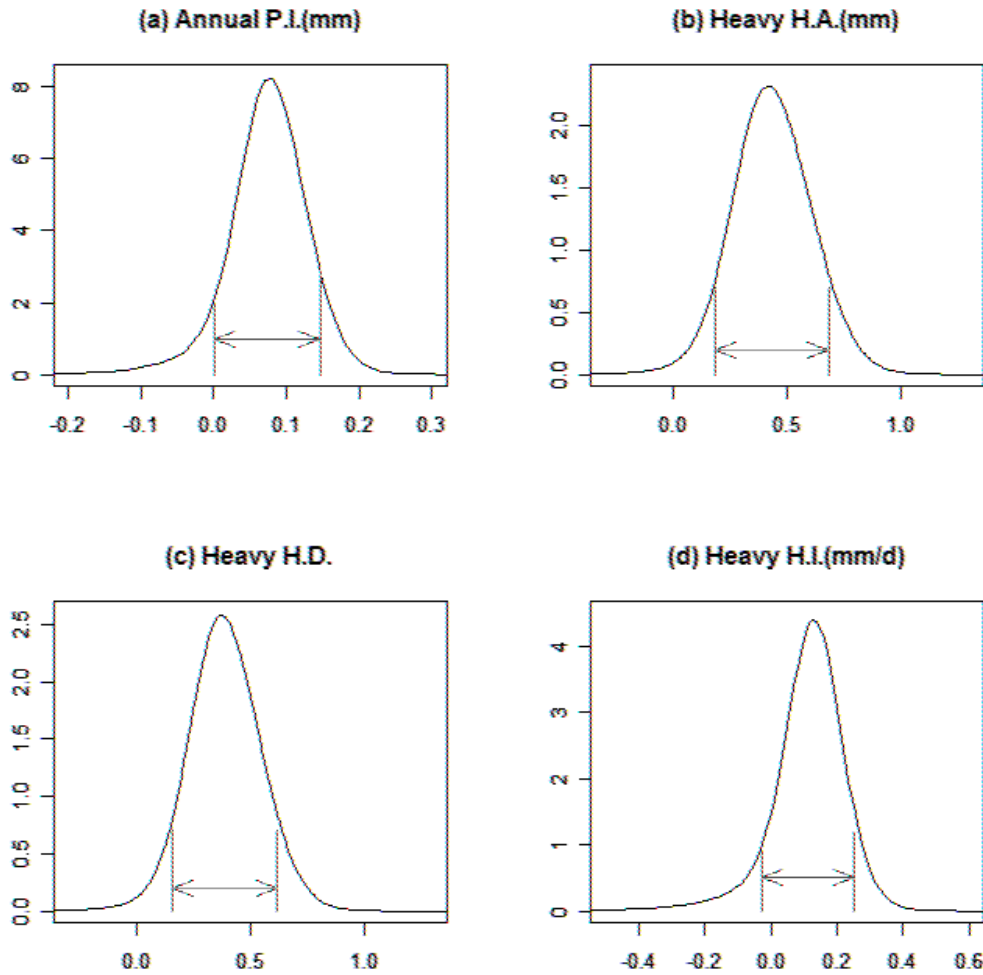


Fig. 6. Marginal posterior density functions $\pi(\delta = \mu_2 - \mu_1 | x)$ and 90% HPD credible intervals for the annual and log-transformed heavy precipitation characteristics. The arrow means 90% HPD credible intervals, respectively.

증가하고 있음을 알 수 있다.

5. 요약 및 결론

본 연구에서는 1954년부터 2007년까지 지난 54년간의 기상청 관측 강수량 자료를 이용하여 남한평균 및 집중호우시의 강수특성(강수량, 강수일수, 강수강도)의 변화시점 및 변화유형에 대해 분석하였다. 변화점 문제에 대한 대부분의 연구는 일반적으로 주어진 일련의 관측 자료에 변화점 유무를 탐지하는 부분과 일련의 관측된 자료에서 변화점이 발생되었다는 가정 하에서 모수를 추정하는 부분으로 이루어진다. 본 연구에서는 남한평균 및 집중호우 특성 (강수량, 강수일

Table 3. Posterior probability for $\delta = \mu_2 - \mu_1$

Selected precipitation characteristics	Posterior probability
	$P(\delta < 0 x)$
A.P.I.(mmd ⁻¹)	0.080
H.P.A.(mm)	0.011
H.P.D.(days)	0.012
H.P.I.(mmd ⁻¹)	0.122

수, 강수강도)에서 변화점이 없는 모형(M_0), 평균에만 변화가 있는 모형(M_1) 그리고 평균과 분산 모두에 변화(M_2)가 있는 모형을 가정하여 각각에 대해 모수추정과 신뢰도 분석을 베이시안 방법을 적용하였다. 분석과정에서 집중호우의 경우 정규성을 만족시키기 위하여 로그변환을 하였으며 무정보 사전분포에서 베이즈

인자 계산은 Berger and Pericchi (1996)의 3가지 방법(산출평균, 기하평균 및 중앙값)을 이용하였다. 또한 변화시점 전후의 평균 차에 대한 90% 최고사후밀도와 사후확률분포는 김찬수와 서명석 (2008)의 방법을 이용하였다.

지난 54년간의 우리나라 강수특성의 경우 집중호우시의 강수량과 강수일수에서는 모수추정 및 90% 최고사후밀도 분석 모두에서 1996년을 전후로 평균에서만 변화가 있었음이 확인되었다. 하지만 남한평균 강수특성(강수량, 강수일수 및 강수강도)과 집중호우시의 강수강도에서는 모수추정 및 90% 최고사후밀도 분석에서 평균 및 분산 모두 변화점이 전혀 없거나 변화의 강도가 매우 약한 것으로 나타났다. 남한평균 강수일수에서는 변화점이 전혀 없으나 집중호우 일수에서는 집중호우 빈도수가 증가하는 변화점이 확실히 나타나는 것으로 보아 우리나라에서 최근 연강수량이 크게 증가한 것은 주로 강수강도가 강한 집중호우가 자주 발생한 것이 주원인으로 볼 수가 있다. 이러한 결과는 1996년을 전후로 남한평균 강수강도와 집중호우 특성에 변화가 있었다는 김찬수와 서명석(2008)의 연구와는 일치한다고 볼 수 있다. 다만 변화강도가 크지 않았던 남한평균 강수강도와 집중호우 강도에서는 변화가 유의하지 않은 것으로 나타났다. Kim *et al.* (2006)의 Fig. 2와 본 연구의 Fig. 4(d)에서도 보인 바와 같이 1996년 전후와 함께 1970년대 말에도 집중호우 강도에서 변화의 신호가 약하게 나타나고 있다. 하지만 본 연구결과는 1970년대 말을 전후로 집중호우 강수량에 큰 증가가 있었다는 Ho *et al.* (2003)와 Kim *et al.* (2006)의 연구결과와는 변화 시점에서 큰 차이를 보이고 있다. 유사한 자료를 이용하였음에도 본 연구와 Kim *et al.* (2006) 연구 결과에서 변화시기가 다르게 선정된 원인에는 연구방법의 상이함 외에도 Kim *et al.* (2006)이 연 강수량이 아니고 태풍이 주로 영향을 주는 8, 9월의 자료만을 사용한 점, 사용 지점수가 12개인 점 그리고 집중호우의 정의를 100 mm d⁻¹ 로 한 점들이 복합적으로 작용한 것으로 본다.

감사의 글

이 연구는 기상청 기상지진기술개발사업(CATER 2009-3114)의 지원으로 수행되었습니다. 유익한 지적을 통해 논문의 질을 향상시켜 주신 심사위원님들께 감사드립니다.

참고문헌

- 김찬수, 서명석, 2008: 우리나라에서 최근 (1976-2005) 강수의 변화 시점. *대기*, **18**, 111-120.
- 최영은, 권원태, 2005: 지리학에서의 기후변화 연구. *한국기상학회지*, **41**, 249-262.
- 홍기옥, 서명석, 나득균, 2006: 최근 30년간(1976-2005) 우리나라 강수의 시·공간변동과 지리환경. *한국지구과학회지*, **27**, 433-449.
- Anderson, T. W., and D. A. Darling, 1952: Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Math. Statist.*, **23**, 193-212.
- Berger, J. O., and L. R. Pericchi, 1996: The intrinsic Bayes factor for model selection and prediction. *J. Amer. Statist. Assoc.*, **91**, 109-122.
- Chu, P. S., and X. Zhao, 2004: Bayesian change-point analysis of tropical cyclone activity: The central North Pacific case. *J. Climate*, **17**, 4893-4901.
- Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns, 2000: Climate extremes: Observations, modeling, and impacts. *Science*, **289**, DOI 10.1126/science.289.5487.206
- Epstein, E. S., 1985: Statistical inference and prediction in climatology: *A Bayesian Approach*, Meteor. Monograph. **20**, American Meteor. Soc., 199pp.
- Haylock, M., and N. Nicholls, 2000: Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910-1998. *Int. J. Climatol.*, **20**, 1533-1541.
- Ho, C.-H., J.-Y. Lee, M.-H. Ahn, and H.-S. Lee, 2003: A sudden change in summer rainfall characteristics in Korea during the late 1970s. *Int. J. Climatol.*, **23**, 117-128.
- IPCC, 2001, Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Albritton DL, Meira Filho LG (eds). Cambridge University
- Iwashima, T., and R. Yamamoto, 1993: A statistical analysis of the extreme events. Long-term trend of heavy daily precipitation. *J. of Meteor. Soc. of Japan*, **71**, 37-640.
- Jung, H.-S., Y. Choi, J.-H. Oh, and G.-H. Lim, 2002: Recent trends in temperature and precipitation over South Korea. *Int. J. Climatol.*, **22**, 1327-1337.
- Khaliq, M. N., T. B. M. J. Ouarda, A. St-Hilaire, and P. Gachon, 2006: Bayesian change-pint analysis of heat spell occurrences in Montreal, Canada. *Int. J. Climatol.*, **27**, 805-818.
- Kim, J.-H., C.-H. Ho, M.-H. Lee, J.-H. Jeong, and D. Chen, 2006: Large increase in heavy rainfall associated with tropical cyclone landfalls in Korea after the late 1970. *Geophys. Res. Lett.*, **33**, doi:10.1029/2006GL027430.
- Meehl, G. A., F. Zwiers, J. Evans, T. Knutson L. Mearns, and P. Whetton, 2000: Trends in extreme weather and climate events. Issues related to modeling extremes in projections

- of future climate change. *Bull. Amer. Meteor. Soc.*, **81**, 427-436.
- Perreault, L., J. Bernier, B. Bobée, and E. Parent, 2000a: Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *J. Hydrology*, **235**, 221-241.
- Perreault, L., J. Bernier, B. Bobée, and E. Parent, 2000b: Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *J. Hydrology*, **235**, 242-263.
- Roy, S. S., and R. C. Jr. Balling, 2004: Trends in extreme daily precipitation Indices in India. *Int. J. Climatol.*, **24**, 457-466.
- Solow, A. R., 1988: A Bayesian approach to statistical inference about climate change. *J. Climate*, **1**, 512-521.
- Son, Y. S., and S. W. Kim, 2005: Bayesian single change point detection in a sequence of multivariate normal observations. *Statistics*, **39**, 373-387.
- Tapsoba, D., M. Hache, L. Perreault, and B. Bobée, B., 2004: Bayesian rainfall variability analysis in West Africa along cross sections in space-time grid boxes. *J. Climate*, **17**, 1069-1082.
- Zhai, P., X. Zhang, H. Wan, and X. Pan, 2005: Trends in total precipitation and frequency of daily precipitation extremes over China. *J. Climate*, **18**, 1096-1108.