

강우자료의 비동질성 규명을 위한 변동점 분석기법의 상호비교 및 적용

Intercomparison of Change Point Analysis Methods for Identification of Inhomogeneity in Rainfall Series and Applications

이 상 호* / 김 상 욱** / 이 영 섭*** / 성 장 현****

Lee, Sangho / Kim, Sang Ug / Lee, Yeong Seob / Sung, Jang Hyun

Abstract

Change point analysis is a efficient tool to understand the fundamental information in hydro-meteorological data such as rainfall, discharge, temperature etc. Especially, this fundamental information to change points to future rainfall data identified by reasonable detection skills can affect the prediction of flood and drought occurrence because well detected change points provide a key to resolve the non-stationary or inhomogeneous problem by climate change. Therefore, in this study, the comparative study to assess the performance of the 3 change point detection skills, cumulative sum (CUSUM) method, Bayesian change point (BCP) method, and segmentation by dynamic programming (DP) was performed. After assessment of the performance of the proposed detection skills using the 3 types of the synthetic series, the 2 reasonable detection skills were applied to the observed and future rainfall data at the 5 rainfall gauges in South Korea. Finally, it was suggested that BCP (with 0.9 posterior probability) could be best detection skill and DP could be reasonably recommended through the comparative study. Also it was suggested that BCP (with 0.9 posterior probability) and DP detection skills to find some change points could be reasonable at the North-eastern part in South Korea. In future, the results in this study can be efficiently used to resolve the non-stationary problems in hydrological modeling considering inhomogeneity or nonstationarity.

keywords : change point analysis, climate change, inhomogeneity, CUSUM, BCP, DP

요 지

변동점 분석은 강우, 유량, 온도 등의 수문기상자료의 기초적인 정보를 이해함에 있어 유용한 수단으로 활용될 수 있다. 특히 합리적인 방법으로 분석된 미래 강우량에 대한 변동점 분석 결과는 최근 기후변화로 인해 발생하는 홍수나 가뭄과 같은 수문현상을 예측하는 데 있어 문제가 되는 비동질성 또는 비정상성 요인을 해결하는 데 있어 중요하게 활용될 수 있다. 따라서 본 연구에서는 변동성 분석을 위해 이용되는 3가지의 이론(CUSUM, BCP, DP)을 소개함과 함께 3가지 기법들의 특성을 알아보기 위한 상호비교 연구를 수행하였다. 3가지 종류의 합성 모의 강우자료의 생성을 통해 3가지 기법들을 수행한 뒤 그 결과를 특정 성능평가 절차를 거쳐 평가하였으며, 그 중 2가지 기법을 국내 관측 강우자료 및 미래 강우자료의 변동점 분석에 적용하고 그 결과를 제시하였다. 기법 간 비교를 통해 BCP 0.9가 가장 우수한 탐색능력이 있는 것으로 분석되었으며, DP 또한 합리적으로 사용될 수 있을 것으로 판단되었다. 향후 이와 같은 강우자료에 대한 변동점 분석에 관한 연구는 비동질성이나 비정상성을 고려하여 수문모형을 구축하는 연구 등에 있어 효율적으로 사용될 수 있다.

핵심용어 : 변동점 분석, 기후변화, 비동질성, CUSUM, BCP, DP

* 부경대학교 공과대학 토목공학과 교수 (e-mail: peterlee@pknu.ac.kr)

Professor, Department of Civil Engineering, Pukyong National University, Busan, Korea

** 교신저자, 강원대학교 공과대학 토목공학과 조교수 (e-mail: sukim70@kangwon.ac.kr, Tel: 82-33-250-6233)

Corresponding Author, Assistant Professor, Department of Civil Engineering, Kangwon National University, Chuncheon, Korea

*** 강원대학교 공과대학 토목공학과 석사과정 (e-mail: dldudtjq1216@naver.com)

Master Course, Department of Civil Engineering, Kangwon National University, Chuncheon, Korea

**** 국토교통부 영산강홍수통제소 시설연구소 (e-mail: jhsung1@korea.kr)

Researcher, Ministry of Land, Infrastructure and Transport, Yeongsan River Flood Control Office, Korea

1. 서 론

기후변화는 이제 인간의 생활과 밀접한 수문, 수리, 생태 등의 환경을 위협하는 미래의 변화요인으로 인식되었다. 수자원 관련 분야에 있어 기후변화와 관련되어 가장 중요하게 다루어지는 부분은 ‘비정상성(nonstationarity)’ 또는 ‘비동질성(inhomogeneity)’과 관련된 부분이다(용어의 선정에 있어 본 연구에서는 비동질성을 채택하여 사용함). 기후변화와 관련하여 현 시점에서 중요하게 부각되어지는 연구 분야가 수문기상자료의 통계적 특성 변화와 관련된 분야이며, 그 중에서도 관측 및 미래 강우자료를 대상으로 하는 변동점 분석(change point analysis)과 관련된 연구는 미래의 각 지점별 유량을 모의하기 위한 장기유출모형의 구축에 있어 모형을 보정하거나 검증하는 절차에서 비동질성 문제를 해결하기 위해 매우 중요하게 활용될 수 있다.

통계적 가설검정에 바탕을 둔 경향성 분석과 유사한 기법을 활용하는 기존의 방법들과는 달리 최근의 변동점 분석은 변동점의 숫자, 위치, 크기 등을 정량적으로 해석하는 데 있어 유용한 기법들을 제공하고 있다. 이와 같은 변동점 분석기법들은 크게 자료를 순차적으로 하나씩 분석하여 새로 입력된 자료가 기존 자료들과 어떠한 통계적 특성이 있는지를 분석하고 이로부터 변동점의 여부를 결정하는 ‘on-line’ 방법과 자료의 전체를 입력하여 대상자료 전체에 대해 변동점 분석을 수행하는 ‘off-line’ 방법으로 구분되어 질 수 있으며, 변동점을 탐색하는 특정 절차에 따라 비모수적(non-parametric) 방법과 모수적(parametric) 방법들로 구분될 수도 있다. 또한 모수적 절차를 사용하는 부류의 변동점 분석은 모수(parameter) 추정에 있어 frequentist적인 방법 또는 Bayesian적인 방법을 사용하는 지에 따라 세부적으로 구분된다. 또한 통계적 방법과 달리 동적계획법(dynamic programming)을 사용하여 목적함수(objective function)를 최소화하는 최적화 이론을 사용하여 변동점을 탐색하는 기법도 최근 많이 사용되고 있다(Dobigeon and Tourneret, 2007).

전술한 바와 같이 최근 수문기상 자료의 비동질성을 규명하고자 하는 최신의 변동점 분석기법에 대한 연구가 점차 증가하고 있는 상황이나 변동점 분석기법들은 각 기법마다의 접근방식이 모두 달라 같은 수문기상 자료를 대상으로 분석하는 경우에도 다른 결과들을 제시할 확률이 높다. 따라서 자료의 특성과 분석기법의 특성을 먼저 파악함으로써 어떤 변동성 기법을 선정하여 사용하는 것이 가장 과학적으로 변동점을 탐색하였는지를 사전에 결정할 필요가 있다.

그러므로 본 연구에서는 국내에 위치한 5개의 강우관측소의 관측 강우자료와 통계적으로 유사한 동질특성 자

료, 비동질 단일 변동점 자료, 비동질 다중 변동점 자료를 발생시킨 후, 3가지 변동성 분석기법인 CUMulative SUM (CUSUM)기법, Bayesian Change Point (BCP)기법 및 Dynamic Programming (DP)기법을 적용하여 그 결과를 특정 비교절차를 거쳐 상호비교하고 가장 합리적인 결과를 나타내는 분석기법을 제안하기 위한 연구를 수행하였다. 또한 가장 합리적으로 제안된 변동성 분석기법을 5개 강우관측소의 관측 강우자료 및 미래 강우자료에 대해 실제 적용해봄으로써, 5개 강우자료가 과거부터 미래까지 어떠한 변동성을 보일지를 추정해 보았다.

2. 변동점 분석기법의 이론적 배경

2.1 변동점 분석

변동점 분석을 다룬 대부분의 기존 연구들(Hirsch et al., 1982; Burn, 1994; Abdul Aziz and Burn, 2006)은 자료의 동질성을 파악하기 위하여 주로 t-test, Wicoxon test, Mann-Whitney test 등과 같은 통계학적 추론에 이론적 근거를 둔 기법들이 사용되었다. 그러나 이와 같은 통계적 추론에 이론적 바탕을 둔 경향성 분석 기반의 기법들은 동질성이 의심되는 자료의 통계적 특성을 분석하기에 유용한 방법들이긴 하지만, 통계적 검정에 있어 검정을 위해 수립한 귀무가설(null hypothesis)을 채택 또는 기각의 여부만 제시되므로 궁극적으로 변동점 문제에 있어서는 특정 신뢰구간에 대해 동질성이 ‘있다 또는 없다’는 결론만을 제시하게 되므로 변동성 분석결과를 정량적으로 제시하지 못한다는 단점이 있다.

이와 달리 본 연구에서 적용하고자 하는 변동점 분석기법들은 변동점의 위치(location) 및 크기(magnitude)를 정량적으로 제시할 수 있는 기법이라는 점에서 기존의 경향성 분석(trend analysis) 분야의 연구와 차별된다고 볼 수 있다. 다중 변동점 분석(multiple change point analysis) 모형은 n 의 길이를 가지는 특정 관측시계열자료 $[x]_i^{i=1}^n$ 이 변동점들로 인해 구분되어지는 구간을 표현할 수 있는 확률밀도분포함수들(probability density functions), $p_1(x), p_2(x), \dots, p_n(x)$ 에 의해 구분되어 진다는 가정으로부터 수립된다. 따라서 이 모형은 각각의 구간은 난수(random variable)들인 $\mathbf{X}=(X_1, X_2, \dots, X_n)$ 에 대하여 Eq. (1)과 같이 수학적으로 표현될 수 있다.

$$X_i \sim \begin{cases} p_1(x) = p(x|\theta_1), & 1 \leq i \leq \tau_1 \\ p_2(x) = p(x|\theta_2), & \tau_1 < i \leq \tau_2 \\ \vdots \\ p_n(x) = p(x|\theta_n), & \tau_{n-1} < i \leq n \end{cases} \quad (1)$$

여기서, $\theta_1, \dots, \theta_n$ 은 각각의 확률밀도함수에 포함된 모수(parameter)이고, $\tau_1, \dots, \tau_{n-1}$ 은 변동점이며 이들은 모두 특정 기법을 통해 추정되어야 할 대상이다. 따라서 Eq. (1)에서 각각의 모수와 변동점이 합리적으로 추정되면, 특정 시계열자료는 n 개의 구간을 가지는 각각의 동일한 구간들로 구분될 수 있으며, 최종적으로 이와 같이 구분된 자료의 구간에 대한 각각의 평균 등을 계산하여 변동점의 위치 및 변화의 크기를 정량적으로 제시할 수 있다. 특히 Eq. (1)에서 변동점이 한 개만 존재하는 경우 다중 변동점 분석모형은 단일 변동점 분석(single change point analysis)이 된다.

본 연구에서는 on-line적이고 비모수적인 기법에 포함되는 'cumulative sum (CUSUM)'기법, off-line적이고 모수적 기법인 'Bayesian change point (BCP)'기법, 그리고 off-line적이고 시스템이론을 사용하는 '동적계획법(dynamic programming, DP)'을 사용하여 각 기법의 결과를 비교하고 실제 강우자료에 적용된 결과를 제시하였다.

2.2 CUSUM 기법

CUSUM기법은 특정 신호의 변동을 on-line적으로 탐색하는 이론으로 원래는 특정 제조상품의 품질관리를 검사하기 위하여 Page (1954)에 의해 개발되었다. 이 기법은 시계열의 분석을 위해 특정한 확률밀도함수를 사용하지 않기 때문에 모수의 추정과정이 필요없으므로 이론자체가 매우 간단하고 적용이 쉽다는 장점이 있는 반면 자료자체를 하나씩 분석한다는 on-line적 특징으로 인해 자료전체를 대상으로 하는 off-line적 탐색결과에 비해 정확성이 떨어질 수 있다는 단점이 있다. 한편 기상분야에서는 강우관측소 등의 고장이나 특정 원인으로 인한 강우자료의 이상치를 탐색하기 위한 기법의 활용이 많은 편이다(Kampata et al., 2008; Li et al., 2008; Chowdhury and Beecham, 2010; Chelani, 2011; Chu et al., 2012).

CUSUM기법은 새롭게 관측된 측정자료와 기존 자료의 누적된 잔차를 도식화하여 이 누적된 잔차의 변화로부터 새롭게 관측된 측정자료가 기존 자료에 속하게 되는지 새로운 구간에 속하게 되는지를 분석한다. 어떤 연속된 자료인 x_1, x_2, \dots, x_n 이 입력되고 있다고 할 때, 이들의 누적 잔차합을 각각 S_0, S_1, \dots, S_n 으로 표시하면 누적합은 다음과 같은 Eq. (2)로부터 계산할 수 있다.

$$S_t = S_{t-1} + (x_t - \bar{x}), \quad t = 1, 2, 3, \dots, n \quad (2)$$

여기서, 누적 잔차합의 초기값(S_0)는 0으로 입력되며, \bar{x} 는 입력된 자료의 평균이며, CUSUM기법은 이를 도식화하여 결과를 도출한다. 따라서 도식화된 그래프에서 '+'의 일정

한 기울기 값은 이 기간 동안 입력된 자료가 평균보다 크고 변화가 유사한 값으로 입력되었음을 의미하며, '-' 기울기는 그 반대의 경우가 될 것이다. 따라서 이 그래프에서 기울기의 급격한 변화는 기준에 입력된 자료와는 평균에 있어서 매우 다른 양상의 자료가 입력되었음을 의미한다. 이와 같은 분석에 있어서는 급격한 변화가 발생했는지에 대한 신뢰도가 문제가 될 수 있는데, 이를 해결하기 위하여 CUSUM기법에서는 붓트랩(bootstrap)기법을 활용하여 신뢰도(confidence level)를 계산한다. 즉, 붓트랩 기법을 활용한 신뢰도를 계산함으로써 가장 변화가 발생될 수 있는 시점을 추정하고 이 결과를 이용하여 변동점 분석결과에 불확실성을 포함시킨다. CUSUM기법에서는 이를 위하여 붓트랩을 수행하기 이전에 먼저 Eqs. (3)~(5)을 계산한다.

$$S_t^+ = \max[0, x_t + S_{t-1}^+], \quad t = 1, 2, \dots, n \quad (3)$$

$$S_t^- = \min[0, x_t + S_{t-1}^-], \quad t = 1, 2, \dots, n \quad (4)$$

$$\Delta S_t = (S_t^+) - (S_t^-), \quad t = 1, 2, \dots, n \quad (5)$$

이후 붓트랩을 수행하고 Eqs. (3) and (4)를 다시 계산한 뒤, 각각의 붓트랩결과를 이용하여 Eq. (5)에 해당되는 ΔS_t 인 ΔS^j 를 계산하며, 최종적인 신뢰도는 Eq. (6)을 이용한다.

$$CL = 100 \times \frac{R}{N} \quad (6)$$

여기서, R 은 $\Delta S_t > \Delta S^j$ 인 경우에 해당되는 붓트랩결과 수이며, N 은 붓트랩에 의해 발생된 전체자료의 수이다 (Carslaw et al., 2006). 따라서 CL 의 값이 클수록 변동점으로 취급할 수 있는 신뢰도가 높아짐을 알 수 있다.

2.3 BCP 기법

베이저안 추정기법은 연속된 수문기상자료에 있어 평균과 같은 특정 통계특성의 변동과 관련된 변동위치를 확률적으로 분석하는 데 이용될 수 있다. 변동점 분석에 있어 일반적인 이론의 개발 및 보완은 Carlin et al. (1992)에 의해 수행되었으며, 크게 단일 BCP (single Bayesian change point)기법과 다중 BCP (multiple Bayesian change point)기법으로 구분되어 연구가 수행되었다.

단일 BCP기법은 주로 공액 사전분포(conjugated prior distribution)를 사용함으로써 사후확률의 계산에 있어 간단한 면이 있으나, 단일 BCP기법은 단지 1개의 변동점을 찾아내기 때문에 장기 수문기상자료의 경우 먼저 주관적으로 자료의 구간을 분리한 다음 변동점 분석을 통해 변동점을 탐색해야 하는 문제가 있어 사실상 정확한 분석결

과를 보장하기 어려운 측면이 있다. 다중 BCP기법은 주로 Barry and Hartigan (1992, 1993)이 개발한 ‘Product Partition Model (PPM)’을 사용하는 데, 이 모형은 특정 분리구간(partition)의 확률이 ‘사전결합의 곱(product of prior cohesion)’에 비례한다는 가정 하에 유도된다. 또한 Chib (1998)은 이를 위해 Markov과정을 응용한 별도의 다중 BCP기법을 제안하기도 하였고, 계산통계학 분야에서는 각종 응용분야에 대해 몇 가지의 연구(Kim and Cheon, 2010; Fearnhead and Liu, 2011; Fearnhead, 2006; Kim et al., 2009)가 수행되었으나, 수문기상자료에 대해 PPM 모형을 적용한 다중 BCP기법의 활용사례는 국내연구인 Hwang et al. (2010)을 제외하고는 드물다.

Barry and Hartigan (1992, 1993)은 연속된 시간의 순서에 따라 발생된 관측자료 X_1, X_2, \dots, X_n 에 대하여 각각의 관측치가 분할된 분리구간 i 에 대해 특정 θ_i 를 가지는 각각의 확률밀도분포함수로 표현될 수 있다는 개념과 $0 = i_0 < i_1 < \dots < i_b = n$ 을 만족시키는 각각의 분리구간 내의 확률이 정규분포를 따르는 $\rho = (i_0, i_1, \dots, i_b)$ 로 특정하게 주어질 수 있다는 가정을 사용하여 PPM모형을 개발하였다. PPM모형의 구성인자 중 가장 중요한 요소는 분리구간을 결정하는 확률을 정하는 문제로 Barry and Hartigan (1992, 1993)은 다음과 같은 Eq. (7)을 사용하였다.

$$f(\rho) = K c_{i_0 i_1} c_{i_1 i_2} \dots c_{i_{b-1} i_b} \quad (7)$$

여기서, K 는 모형의 수행에 있어 임의적으로 주어지는 상수이고, c_{ij} 는 발생가능한 각각의 블록 ij 에 대해 특정되는 사전결합 확률값이다. 이에 따라 최종적으로 b 개의 분리구간에서의 확률인 ρ 가 주어졌을 때, ‘블록 사전밀도(block prior density)’라 불리는 $f_{i_{j-1} i_j}(\theta_{i_j})$ 형태의 확률밀도함수를 가지는 $\theta_i (= \theta_{i_1}, \theta_{i_2}, \dots, \theta_{i_b})$ 가 독립적으로 존재할 수 있고 이를 활용하여 사전분포를 구성하고 최종적으로 Eq. (8)과 같은 사후확률분포를 구성하게 된다.

$$f_{ij}(\theta_j | X_{ij}) = \frac{f_{ij}(X_{ij} | \theta_j) f(\theta_j)}{\int f_{ij}(X_{ij} | \theta_j) f(\theta_j) d\theta} \quad (8)$$

PPM모형에서 마지막으로 결정되어야 하는 요소는 c_{ij} 인데, Barry and Hartigan (1993)은 이를 위해 Yao (1984)가 제안한 Eq. (9)와 같은 간단한 식을 사용하였다.

$$c_{ij} = \begin{cases} (1-p)^{j-i-1} p, & j < n \\ (1-p)^{j-i-1}, & j = n \end{cases} \quad (9)$$

여기서, 특정 확률 p 는 0과 1의 범위를 가진다.

Barry and Hartigan (1993)은 Eq. (8)에서 제시된 사후분포

확률을 계산함에 있어 관측자료가 정규분포인 $N(\mu_{ij}, \sigma^2)$ 을 따른다고 보았으며, μ_{ij} 에 대한 사전분포로 $N(\mu_0, \sigma_0^2 / (j-i))$ 를 사용하였으며, $r \in ij \in \rho$ 인 조건에서 μ_r 을 추정하기 위해 Eq. (10)과 같은 간단한 가중조합을 사용하였다.

$$\hat{\mu}_r = (1-\omega) \overline{X_{ij}} + \omega \mu_0 \quad (10)$$

여기서, $\omega = \sigma^2 / (\sigma_0^2 + \sigma^2)$ 으로 계산된다. 사후확률분포에 포함되는 모수(parameter)는 $\mu_0, \sigma^2, p, \omega$ 이며, Barry and Hartigan (1993)은 이들에 대한 사전분포로 Eqs. (11)~(14)를 사용하였다.

$$f(\mu_0) = 1, \quad -\infty \leq \mu_0 \leq \infty \quad (11)$$

$$f(\sigma^2) = 1/\sigma^2, \quad 0 \leq \sigma^2 \leq \infty \quad (12)$$

$$f(p) = 1/p_0, \quad 0 \leq p \leq p_0 \quad (13)$$

$$f(\omega) = 1/\omega_0, \quad 0 \leq \omega \leq \omega_0 \quad (14)$$

여기서, 초모수(hyper-parameter)에 해당되는 p_0, ω_0 는 0과 1 사이에 존재하는 임의의 값을 사용할 수 있으나, 본 연구에서는 Barry and Hartigan (1993)이 사용한 선례를 따라 0.2를 사용하였다. 이와 같은 구성을 통해 궁극적으로 분석되어야 하는 사후분포확률인 Eq. (8)은 쉽게 그 확률을 추출할 수 없는 요소로 구성되어 있으므로, MCMC (Markov Chain Monte Carlo) 기법을 적용해야 하는 데 이를 위해 조건부 확률에 대한 전이확률(transition probability) p 를 Eq. (15)로부터 산정하였다.

$$\frac{p_i}{p_i - 1} = \frac{P(U_i = 1 | \mathbf{X}, U_j, j \neq i)}{P(U_i = 0 | \mathbf{X}, U_j, j \neq i)} = \frac{\int_0^{p_0} p^b (1-p)^{n-b-1} dp \int_0^{\omega_0} \frac{\omega^{b/2}}{(W_1 + B_1 \omega)^{(n-1)/2}} d\omega}{\int_0^{p_0} p^{b-1} (1-p)^{n-b} dp \int_0^{\omega_0} \frac{\omega^{(b-1)/2}}{(W_0 + B_0 \omega)^{(n-1)/2}} d\omega} \quad (15)$$

여기서, b 는 블록의 개수이며, U_i 는 분리구간마다의 확률이고, W_0, W_1 은 각각 $U_i = 0, 1$ 에 해당되는 블록내 제공의 합이고, B_0, B_1 는 $U_i = 0, 1$ 에 해당되는 블록간 제공의 합이다. 본 연구에서는 이와 같은 MCMC기법을 이용한 BCP 기법을 적용하기 위하여 Erdman and Emerson (2007)이 작성한 ‘bcp package’를 활용하였으며, 최종적으로 산정된 사후확률을 이용하여 특정위치에서 발생할 수 있는 변동점의 확률이 제시될 수 있는데, 본 연구에서는 이러한 측면을 고려하여 BCP 0.7 및 BCP 0.9로 명명된 사후확률 기준을 사용하여 결과를 분석하였다.

2.4 DP 기법

Kehagias et al. (2006)은 각 구간의 평균으로부터 자료

사이의 편차를 구간비용(segmentation cost)로 활용하여 이 구간비용을 최소화하도록 하도록 하는 DP기법의 활용 절차를 제안한 바 있으며, Aksoy et al. (2008)은 DP기법을 장기의 수문기상자료에 적용하여 DP기법에 의한 최적화 절차가 수문기상자료의 변동점 분석에 효율적으로 사용될 수 있음을 제안한 바 있다. 또한 Gedikli et al. (2008)은 'AUG'로 명명되는 DP기법과 유사한 'branch-and bound (BB)'라는 최적화 기법을 제안하기도 하였다. 이와 같이 동적계획법(dynamic programming, DP)과 같은 최적화 이론은 시스템의 최적화 분야뿐만 아니라 자료의 구간특성을 최소화하도록 구성한 목적함수(objective function)와 그에 따른 제약조건(constraint)을 이용하여 효율적인 변동점 분석에 활용될 수 있는데, 본 연구에서는 Aksoy et al. (2008)이 제안한 DP기법을 적용하였다.

강우자료와 같은 시계열자료가 $X = [X_1, X_2, \dots, X_T]$ 와 같이 표현될 때, $0 = t_0 < t_1 < \dots < t_K = T$ 를 만족시키며 주어진 자료가 분리되는 시점(즉 시계열의 위치)은 $\mathbf{t} = [t_0, t_1, \dots, t_K]$ 와 같이 구성될 수 있으며, 여기서 t_i 는 변동이 발생하는 특정 시점으로 간주할 수 있다. 즉, 주어진 자료는 변동점에 의해 각각 $[t_0 + 1, \dots, t_1], [t_1 + 1, \dots, t_2], \dots, [t_{K-1}, \dots, t_K]$ 로 구분되며, 이에 따라 2^{T-1} 개의 분리된 구간이 존재할 수 있다. 이와 같은 변동구간을 최적화 이론으로 최적구간을 추정하기 위해서 Eq. (16)과 같은 목적함수를 구성할 수 있다.

$$J(\mathbf{t}) = \sum_{k=1}^K d_{k-1, t_k} \quad (16)$$

여기서, $0 \leq s \leq t \leq T$ 에 대한 $d_{s,t}$ 는 어떤 정규화된 구간별 에러로써 다양한 종류의 함수가 사용될 수 있으나, 본 연구에서는 Eq. (17)과 같이 구간별 평균과 자료에 대한 차이를 제공하여 목적함수를 구성하는 데 활용하였다.

$$d_{s,t} = \sum_{\tau=s}^t (x_\tau - \mu_{s,t})^2 \quad (17)$$

여기서, 구간평균 $\mu_{s,t} = \sum_{\tau=s}^t x_\tau / (t-s+1)$ 로 계산된다. 따라서 Eq. (16)을 최소화하는 최적의 변동점인 $\mathbf{t} = [\bar{t}_1, \bar{t}_2, \dots, \bar{t}_K]$ 를 추정할 수 있다. DP기법은 $k = 1, 2, \dots, K$ 에 대한 최적구간을 계산하게 되는데, k개의 구간을 포함하는 최적구간이 존재할 때, 마지막 구간을 $[s+1, t]$ 라 표현하면 처음 k-1 구간은 $[x_1, x_2, \dots, x_s]$ 의 최적구간을 형성하게 되며 이때의 최소 구간비용을 c_s^{k-1} 로 표시할 수 있다. 또한 $[x_1, x_2, \dots, x_t]$ 의 최소 구간비용을 c_t^k 로 표현하면 이들의 관계로부터 Eq. (18)과 같은 전형적인 DP 문제를 나타낼 수 있으며, 이로부터 자료의 구간을 분리할 수 있는 최적의 분리구간을 추정할 수 있다.

$$c_t^k = c_s^{k-1} + d_{s+1, t} \quad (18)$$

3. 기법들의 성능평가를 위한 모의자료의 발생

본 연구에서는 변동점 분석을 위한 CUSUM, BCP, DP 기법의 성능을 평가하기 위하여 동질자료(homogeneous series), 단일 변동점자료(a single shifted series) 및 다중 변동점자료(multiple shifted series)의 3가지 특성을 가지는 합성(synthetic) 강우자료를 발생시켰다. 이 합성모의자료들은 Fig. 1에 나타낸 강원지역에 위치한 철원, 춘천, 원주, 인제 및 대관령 기상관측소로부터 산정된 평균과 분산을 이용하여 이 지역의 관측강우자료와 유사한 통계적 특성을 가지도록 하는 조건을 이용하여 발생되었다. 즉 5개 강우관측소에 대한 최근 10년(2003~2012) 동안의 연평균 강우량의 평균인 1,362 mm과 표준편차 230.4 mm를 이용하여 모의강우자료를 발생시켰다. Table 1에는 본 연구에서 사용된 5개 강우관측소의 기본정보를 수록하였다.

3.1 동질특성 자료의 생성

제안된 3개의 변동점 분석기법들이 변동점이 없는 동질한 자료에 대해 어떠한 결과를 제시하는지를 알아보기 위하여 Eq. (19)와 같은 lag-1 자기상관 모형(autoregressive model)을 적용하여 동질특성을 가진 자료를 발생시켰다.

$$z_i = \phi_1 z_{i-1} + \epsilon_i \quad (19)$$

여기서, ϕ_1 는 자기상관계수(autocorrelation coefficient)이고, ϵ_i 는 평균 0과 분산 $1 - \phi^2$ 을 가지는 정규분포로부터 추출된 임의의 값이다. 본 연구에서는 100년의 길이를 가지는 1,000개의 합성모의 강우자료를 발생시켰으며, Fig.



Fig. 1. Location of the 5 rainfall gauges to generate the 3 types of systhetic rainfall series; (1) Cheolwon, (2) Chuncheon, (3) Wonju, (4) Inje, (5) Daegwanryeong

Table 1. General Information of the 5 Rainfall Gauges

Gauges	Longitude	Latitude	Elevation (m)	Starting year
Cheolwon	127.3	38.1	154.9	1988
Chuncheon	127.7	37.9	76.8	1966
Wonju	127.9	37.3	150.7	1972
Daegwanryeong	128.7	37.7	772.4	1972
Inje	128.2	38.1	198.7	1972

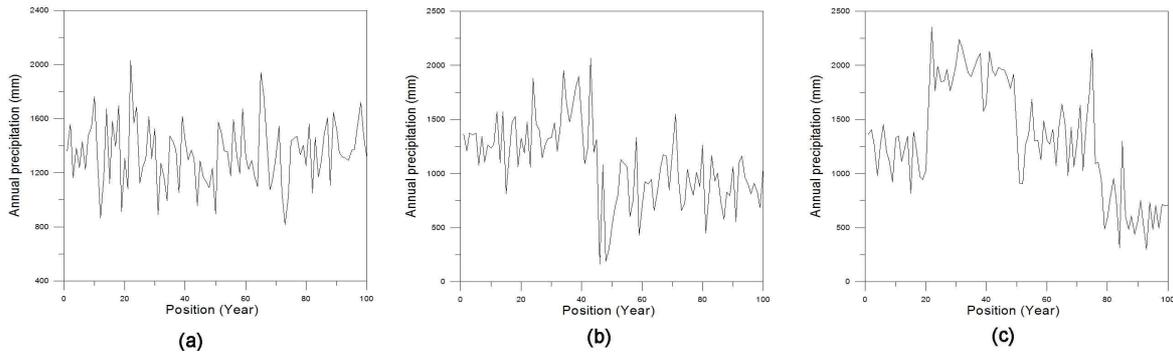


Fig. 2. Examples of Generated Synthetic Time Series of Rainfall Data; (a) Homogeneous Case, (b) Inhomogeneous Case with a Single Shift, (c) Inhomogeneous Case with Multiple Shifts (3 shifts)

2(a)에 그 중 하나를 표시하였다.

3.2 비동질 단일 변동점 자료의 생성

1개의 변동점을 가지는 비동질(inhomogeneous) 자료에 대해 제안된 3개의 변동점 분석기법의 성능을 알아보고 그 결과를 상호비교하기 위하여 Eq. (20)으로부터 단일 변동점 자료를 생성하였다.

$$z_i \begin{cases} z_i^*, & i = 1, \dots, \tau_1 - 1 \\ z_i^* + \lambda_{\tau_1} \sigma, & i = \tau_1, \dots, 100 \end{cases} \quad (20)$$

여기서, z_i^* 는 3.1과 같은 과정을 통해 모의된 동질특성 자료이며, τ_1 은 100년의 길이 중 20년과 80년 사이에서 무작위적으로 선택된 변동점을 나타낸다. 또한 σ 는 z_i^* 로부터 계산된 표준편차이고, λ_{τ_1} 은 Eq. (21)을 통해 무작위적으로 선정된 값이다.

$$\lambda_{\tau_1} = \text{sign}(u - 0.5) \times 3 \times b, \quad u \sim U(0, 1), \quad b \sim B(2, 2) \quad (21)$$

여기서, b 와 λ_{τ_1} 은 -3σ 에서 3σ 사이에서 변동되게 되며, 이를 고려하여 최종적인 단일 변동점 자료를 생성할 수 있다. Fig. 2(b)에는 1,000개의 합성모의 경우자료 중 하나를 나타내었다.

3.3 비동질 다중 변동점 자료의 생성

3.1 및 3.2와 같은 목적으로 3개의 변동점이 존재하는 비

동질 자료를 Eq. (22)를 사용하여 3.2와 유사한 과정을 통해 모의발생시켰다. 또한 본 연구에서는 4개 이상의 변동점을 가지는 자료에 대해서는 별도로 기법간 비교를 수행하지 않았다. Eq. (22)에서 z_i^* 는 3.1과 같은 과정을 통해 모의된 동질특성 자료이며, $\sigma_1, \sigma_2, \sigma_3$ 는 각각 첫째, 둘째, 셋째 구간의 자료를 이용하여 계산된 표준편차이다. 또한 $\lambda_{\tau_1}, \lambda_{\tau_2}, \lambda_{\tau_3}$ 는 Eq. (21)을 이용하여 결정하였으며, 다중 변동점 τ_1, τ_2, τ_3 는 각각 15~25년, 45~55년 및 75~85년의 구간에서 무작위적으로 발생시켰다. Fig. 2(c)에는 발생된 1,000개의 다중 변동점 자료 중 1개를 제시하였다.

$$z_i \begin{cases} z_i^*, & i = 1, \dots, \tau_1 - 1 \\ z_i^* + \lambda_{\tau_1} \sigma_1, & i = \tau_1, \dots, \tau_2 - 1 \\ z_i^* + \lambda_{\tau_2} \sigma_2, & i = \tau_2, \dots, \tau_3 - 1 \\ z_i^* + \lambda_{\tau_3} \sigma_3, & i = \tau_3, \dots, 100 \end{cases} \quad (22)$$

4. 성능평가 방법 및 결과

4.1 변동점 분석기법의 성능평가 방법

변동점이 없는 동질(homogeneous) 자료는 모의된 시계열자료 1,000개 중 변동점이 있는 것으로 탐색된 시계열자료의 개수를 산정하여 각 기법의 성능을 평가할 수 있다. 이는 변동점이 없음에도 불구하고 존재하는 것으로 잘못 탐색된 시계열 자료의 개수를 이용하는 방법으로 전체 시

계열 개수에 대한 비율은 귀무가설이 옳음에도 불구하고 각각 시키는 제1종 오류(type I error)에 해당된다.

변동점이 1개 존재한 비동질(inhomogeneous) 단일 변동점 자료는 탐색된 변동점의 개수로 인한 오류와 변동점의 위치로 인한 오류로 분석기법의 성능을 평가할 수 있다. 본 연구에서는 비동질 단일 변동점 자료의 성능을 평가하기 위하여 분석결과를 ‘탐색 정확(correctly identified)’, ‘탐색 일부 정확(well identified)’ 및 ‘탐색 부정확(poorly identified)’로 나누어 제시하였다. 여기서 ‘탐색 정확’은 탐색된 변동점의 개수가 ‘1’이면서, 변동점의 위치도 모의자료와 일치되는 경우를 의미하며, ‘탐색 일부정확’은 탐색된 변동점의 개수는 ‘1’이나 변동점 위치가 모의자료로부터 ±2년 이내인 경우를 의미한다. 탐색된 변동점의 개수가 ‘1’이 아니거나 변동점의 위치가 ±2년을 벗어나는 경우는 ‘탐색 부정확’으로 구분하였다. 변동점이 3개인 비동질 다중 변동점 자료는 변동점의 개수와 Beaulieu et al. (2009)이 제안한 Eq. (23)의 ‘C’값을 사용하여 기법들의 성능을 평가하였다.

$$C = \begin{cases} \frac{1}{nd} \sum_{i=1}^{nd} (p_i^d - p_i)^2, & nr = nd \\ \frac{1}{nr} \left(\sum_{i=1}^{nd} (p_i^d - p_i)^2 + |nr - nd|(n-1)^2 \right), & nr > nd \\ \frac{1}{nd} \left(\sum_{j=1}^{nr} (p_j^d - p_j)^2 + |nr - nd|(n-1)^2 \right), & nr < nd \end{cases} \quad (23)$$

Eq. (23)에서 p^d 와 p 는 각각 특정기법에 의해 탐색된 변동점의 위치 및 모의자료에 존재하는 실제 변동점의 위치이며, nd 와 nr 은 각각 탐색된 변동점의 개수 및 실제 변동점의 개수(본 연구에서는 ‘3’)를 의미한다.

또한 본 연구에서는 Domonkos (2011)가 변동점 탐색기법의 효율성을 분석하기 위하여 제안한 3가지의 지표를 추가적으로 사용하여, 3개 기법의 성능을 알아보았다. 제안된 3가지의 지표는 ‘탐색능력(power of detection, P_w)’, ‘탐색오류비율(false alarm rate, F_a)’ 및 ‘탐색 기능(detection skill, E_d)’로 다음과 같이 Eqs. (24)~(26)을 통해 산정되었다.

$$P_w = S_R / S \quad (24)$$

$$F_a = S_F / (S_R + S_F) \quad (25)$$

$$E_d = (S_R - S_F) / S \quad (26)$$

여기서, S_R 은 올바르게 탐색된 시계열의 개수이고, S_F 는 잘못 탐색된 시계열의 개수이며, S 는 전체 시계열 개수이다. 단, 이 식에서 올바르게 탐색된 기준과 잘못 탐색된 기준을 설정할 필요가 있는데, 이는 아래 4.5절에서 다시 설명하였다.

4.2 동질특성 자료에 대한 성능평가 결과 및 변동점 숫자만을 이용한 결과

Table 2는 동질특성 자료에 대한 결과를 제1종 오류 비율로 환산하여 나타낸 것이며, Fig. 3과 Table 3은 동질특성 자료, 비동질 단일 변동점 자료 및 비동질 다중 변동점 자료에 대한 분석결과를 제시한 것이다.

Table 3을 보면, CUSUM, BCP 0.9 및 DP기법은 모두 약 93%가 넘는 시계열을 모두 변동점이 없는 것으로 탐색하여 만족할만한 결과를 제시한 반면 BCP 0.7의 경우는 81.2%에 그치고 있는 것을 알 수 있으며, 특히 BCP 0.9 기법과 DP기법은 각각 98.2%와 97.1%를 보여 우수한 결과를 나타냈음을 알 수 있다. 변동점 탐색 숫자만을 이용한 결과를 Table 3에서 살펴보면, 비동질 단일 변동점 자료에 대해서는 CUSUM, BCP 0.7, BCP 0.9 및 DP기법의 결과가 각각 82.8%, 70.8%, 89.7%, 79.6%로 BCP 0.9기법이 가장 우수하였으나, DP기법은 CUSUM기법에 비해 다소 성능이 낮은 것으로 나타났다. 또한 비동질 다중 변동점 자료에 대해서는 4개 기법의 결과가 각각 75.5%, 59.9%, 79.8%, 84.5%로 나타나 BCP 0.9기법보다 DP기법이 우수한 것으로 나타났으며,

Table 2. Type I Error to Homogeneous Series by 4 Detection Skills

Error	CUSUM	BCP 0.7	BCP 0.9	DP
Type I error (%)	6.7	18.8	1.8	2.9

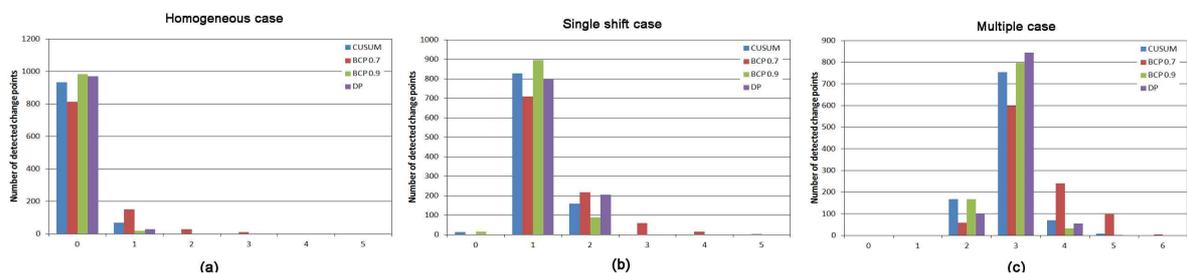


Fig. 3. Diagrams of the Number of Detected Change Points

Table 3. Results of the Number of Series with or Without Detected Change Points

Change points (True)	Change points (Detected)	CUSUM		BCP 0.7		BCP 0.9		DP	
		Number	%	Number	%	Number	%	Number	%
0	0	933	93.3	812	81.2	982	98.2	971	97.1
	1	67	6.7	151	15.1	18	1.8	29	2.9
	2	0	0	27	2.7	0	0	0	0
	3	0	0	10	1	0	0	0	0
	4	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0
1	0	14	1.4	0	0	16	1.6	0	0
	1	828	82.8	708	70.8	897	89.7	796	79.6
	2	158	15.8	218	21.8	87	8.7	204	20.4
	3	0	0	58	5.8	0	0	0	0
	4	0	0	15	1.5	0	0	0	0
	5	0	0	1	0.1	0	0	0	0
3	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0
	2	167	16.7	58	5.8	168	16.8	101	10.1
	3	755	75.5	599	59.9	798	79.8	845	84.5
	4	70	7	241	24.1	33	3.3	54	5.4
	5	8	0.8	98	9.8	1	0.1	0	0
	6	0	0	4	0.4	0	0	0	0

CUSUM기법과 BCP 0.7기법의 결과가 비슷하게 좋지 않은 성능을 보인 것으로 판단되었다. 그러나 변동점 분석의 목적은 변동점의 탐색에 국한되는 것이 아니라 궁극적으로 탐색된 변동점의 위치를 이용하여 통계적 특성(여기서는 평균)이 유사한 구간을 분리하고 각 구간에서 산정된 강우량 크기를 비교하는 것이 매우 중요하므로 다음 절에서 변동점의 위치를 함께 고려하여 다시 비교하였다.

4.3 변동점 위치를 고려한 비동질 단일 변동점 자료에 대한 분석

이 절에서는 모의된 1,000개의 비동질 단일 변동점 자료를 대상으로 분석결과를 성능평가 방법에서 제시한 바와 같이 ‘탐색 정확(correctly identified)’, ‘탐색 일부정확(well identified)’ 및 ‘탐색 부정확(poorly identified)’로 나누어 분석하였다. Table 4는 4개 기법에 대한 각각의 분석 결과를 나타내었는데, 일부 결과에 있어 변동점 숫자만을 이용한 경우와 다른 결과가 얻어졌다. 먼저 ‘탐색 정확’이 경우 CUSUM, BCP 0.7, BCP 0.9 및 DP기법의 결과는 각각 53.5%, 72.3%, 77.8%, 65.2%로 나타나 BCP 0.9가 가장 우수하다는 것은 동일한 결론이나 CUSUM과 DP의 경우 변동점의 위치를 정확하게 탐색하는 숫자가 상당히 감소하게 됨을 알 수 있으며, 특히 CUSUM의 경우는 탐색속

자만을 이용한 경우보다 탐색능력이 상당히 감소하여 나타났음을 알 수 있다. 또한 ‘탐색 일부정확’의 경우에서도 4개 기법의 결과가 79.5%, 88.3%, 90.2%, 87.6%로 나타나 탐색능력의 경향이 ‘탐색 정확’에서와 유사하게 나타났음을 알 수 있다(여기서 비율계산을 위해 사용된 분모는 시계열 개수인 1,000이 아니라, Table 3에서 변동점이 1로 탐색된 시계열의 개수 828 (CUSUM), 708 (BCP 0.7), 897 (BCP 0.9), and 796 (DP)이어야 한다). 따라서 BCP기법은 다소 사후확률 기준이 낮아지더라도 탐색위치에 있어서는 이와 관련없이 정확한 위치를 탐색할 수 있다는 사실을 알 수 있었다. 또한 DP기법의 경우 ‘탐색 일부정확’이 상당히 개선되어 ±2년 내에 있는 변동점의 위치는 BCP 0.7과 유사한 수준까지 상당히 잘 탐색하고 있음을 알 수 있었다.

4.4 변동점 위치를 고려한 비동질 다중 변동점 자료에 대한 분석

여기서는 변동점 위치를 고려하여 비동질 다중 변동점 자료에 대한 분석을 다시 수행하였다. 다중 변동점 자료에 대한 탐색 위치를 고려한 분석에 있어서는 앞서 제시한 C값을 사용하였으며, 4가지 기법에 대해 산정된 C값의 기술통계치를 Table 5에 제시하였다. 제시된 4개 기법인 CUSUM, BCP 0.7, BCP 0.9 및 DP기법에 대한 C값은

각각 2,287, 2,100, 2,048, 2,109로 BCP 0.9기법이 가장 우수한 탐색을 수행했음을 알 수 있었으며, CUSUM기법이 가장 좋지 않은 결과를 보였다. 또한 BCP 0.7기법과 DP기법은 매우 유사한 결과를 보이고 있음을 알 수 있었으나, DP의 경우 표준편차가 BCP 0.7보다 크게 나타나 BCP 0.7보다 좋지 않은 성능을 보였음을 알 수 있었다.

4.5 3가지 지표를 활용한 분석결과

이 절에서는 앞서 제시한 3가지의 분석지표인 ‘탐색능력 (power of detection, P_w)’, ‘탐색오류비율(false alarm rate, F_a)’ 및 ‘탐색 기능(detection skill, E_d)’에 따른 분석결과를 제시하였다. 비동질 단일 변동점 자료는 Table 4에 제시된 각 기법별 ‘탐색 일부정확’의 결과를 이용하여 Eqs. (24)~(26)의 S_R 을 계산하였다. 다중 변동점 분석 자료의 경우는 S_R 을 계산함에 있어 Table 4에 해당되는 값을 다시 계산할 필요가 있는데, 이를 위해 먼저 Table 6의 Eq. (23)에 의해

계산된 결과 중 $nr \neq nd$ 인 경우에 해당되는 4분위수를 제시하였다. 여기서 25%에 해당되는 C 값이 의미하는 바는 이 값보다 적은 C 값으로 계산된 시계열의 경우 변동점의 개수 및 위치를 상당히 잘 탐색하였음을 의미한다고 볼 수 있다. 따라서 본 연구에서는 비동질 다중 변동점 자료에 대한 S_R 을 계산함에 있어 $nr = nd$ 로 분석된 시계열의 개수와 $nr \neq nd$ 인 경우 C 값의 25%보다 적은 C 값을 가지는 시계열의 개수를 더하여 S_R 을 계산하는 기준으로 활용하였다.

Table 7 및 Fig. 4는 분석결과로서 3가지 지표 중 E_d 는 P_w 와 F_a 를 함께 고려하여 만들어진 지표이므로 이 값이 높을수록 탐색기술 능력이 우수한 것으로 볼 수 있다. 먼저 비동질 단일 변동점 자료의 경우 4개의 분석기법에 따른 E_d 는 각각 31.6, 25.0, 61.8, 39.4로 계산되어 BCP 0.9기법과 DP기법이 가장 우수하였음을 알 수 있었으며, 비동질 다중 변동점 자료의 경우 E_d 는 각각 -31.0, -45.0, 37.0, 19.2로 산정되어 같은 결론이 얻어짐을 알 수 있다.

Table 4. Results of the Number of Series with a Single Change Point by 4 Detection Skills

Item	CUSUM		BCP 0.7		BCP 0.9		DP	
	Number	%	Number	%	Number	%	Number	%
Corrected identified	443	53.5	512	72.3	698	77.8	519	65.2
Well identified	658	79.5	625	88.3	809	90.2	697	87.6
Poorly identified	109	13.2	83	11.7	88	9.8	99	12.4

Table 5. Descriptive Statistics of C by 4 Detection Skills

Statistics	CUSUM	BCP 0.7	BCP 0.9	DP
Mean	2,287	2,100	2,048	2,109
Std.	2,897	2,648	2,514	3,048
MIN	0	0	0	0
MAX	8,105	8,004	7,854	8,288

Table 6. Quartiles of C Values by 4 Detection Skills When $nr \neq nd$

Quartile	CUSUM	BCP 0.7	BCP 0.9	DP
25%	5147.1	4911.0	4998.0	5312.0
Median	6003.0	5714.0	5489.0	6258.0
75%	6307.5	6101.0	6001.0	6684.0

Table 7. The number of Series to Calculate the 3 Efficiency Indicators

Indicators	Single shifted case				Multiple shifted case			
	CUSUM	BCP 0.7	BCP 0.9	DP	CUSUM	BCP 0.7	BCP 0.9	DP
S	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
S_R	658	625	809	697	345 (51)	275 (74)	685 (87)	596 (48)
S_F	342	375	191	303	655	725	315	404
P_w	65.8	62.5	80.9	69.7	34.5	27.5	68.5	59.6
F_a	34.2	37.5	19.1	30.3	65.5	72.5	31.5	40.4
E_d	31.6	25.0	61.8	39.4	-31.0	-45.0	37.0	19.2

5. 변동점 분석기법의 실제 강우자료의 적용 및 결과

이 절에서는 BCP 0.9기법과 DP기법을 활용하여 두 가지의 실제 강우자료에 대한 변동점을 분석하고 각 관측소별 분석특성을 제시하였다. 첫 번째 실제 강우자료는 각각의 관측소의 관측 개시일로부터 2012년까지 관측된 일강우자료이며, 두 번째 실제 강우자료는 2013년부터 2099년까지의 미래 일강우자료이다. 우리나라 기상청은 RCP (Representative Concentration Pathways) 4.5와 8.5에 따른 국내 주요 관측지점에 대한 미래 강우자료를 GCM (General Circulation Model)인 HadGEM2-AO를 사용하

여 산정한 뒤, RCM (Regional Circulation Model)인 HadGEM3-RA모형을 사용하여 산정한 바 있다. 본 연구에서는 기상청이 예측한 RCP 4.5와 8.5에 따른 자료를 5개 강우관측지점에 대해 재산정한 두 가지의 미래 강우자료를 사용하여 변동점 분석을 시행하였다(Fig. 5).

특히 이와 같은 분석에 있어 관측 자료와 미래 자료의 경계점인 2013년 인근에서는 두 자료의 통계적 특성에 따라 인위적인 변동점이 존재하는 문제가 발생할 수 있다. 이와 같은 문제를 해결하기 위해서는 분위사상법(quantile mapping) 등을 사용하여 특정 자료를 보정하거나 이와 유사한 통계적 방법을 통해 자료를 보정한 뒤 변동점 분

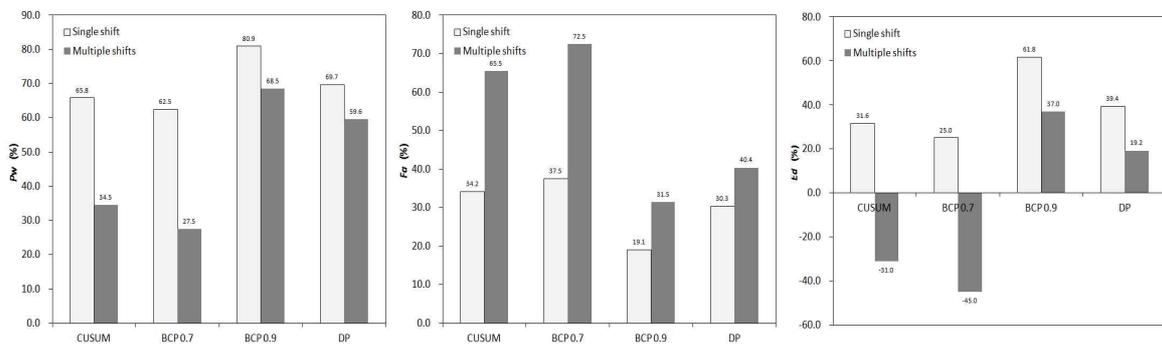


Fig. 4. Results of the 3 Efficiency Indicators (single shift & multiple shifts)

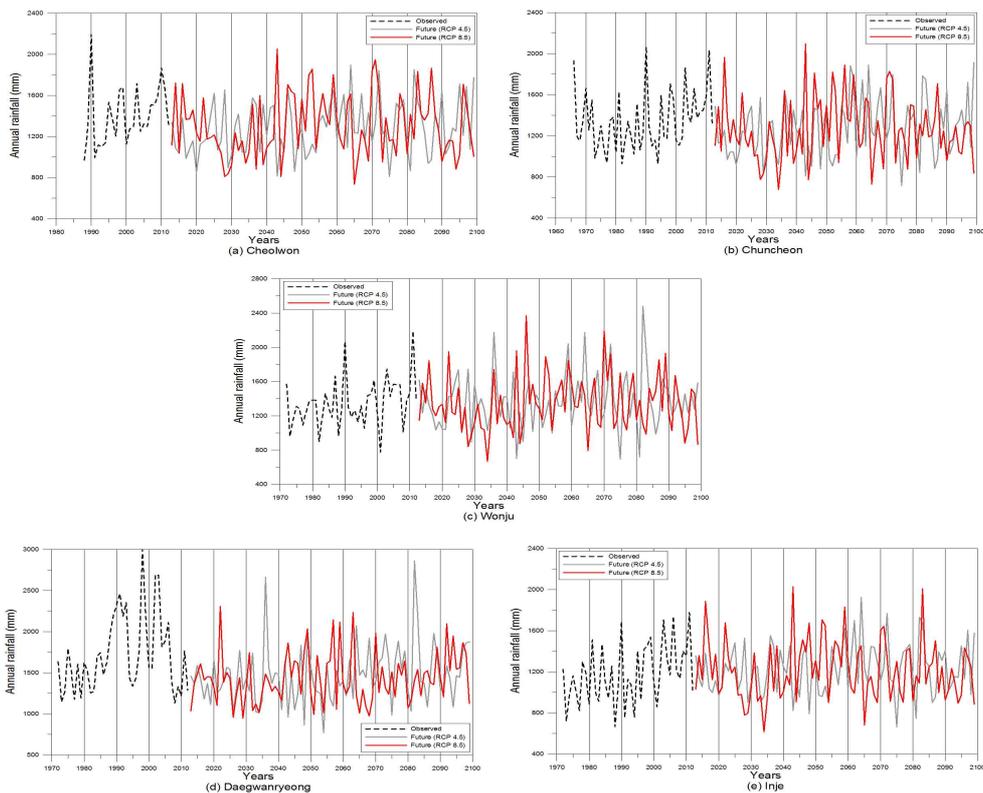


Fig. 5. Real Rainfall Series at 5 Rainfall Gauges

석을 시행할 수 있으나, 이에 관한 부분은 변동점 분석에 있어 본 연구의 범위를 벗어나는 측면이 있어 추가적인 연구가 필요한 부분이라 판단된다. 따라서 본 연구에서는 경계점에서의 문제를 피하기 위하여 과거 및 미래강우자료를 분리하여 변동점 분석을 시행하였다. Table 8은 각각 구성된 실제 강우자료의 기술통계치를 보여주는 표로써, 대관령 강우관측소의 경우 다른 4개의 관측지점보다 매우 큰 평균과 표준편차를 보이고 대부분의 경우 왜곡도가 0과 1 사이에 존재하는 반면 대관령 관측소는 1보다 큰 값을 가지고 있어 타 관측소보다 오른쪽으로 치우친 분포를 가지고 있음을 알 수 있다.

Tables 9 and 10에는 관측 강우자료 및 미래 RCP 4.5 & 8.5 강우자료에 대한 변동점 분석결과를 요약하여 제시하였으며, 탐색된 변동점을 이용하여 분리된 구간과 각각의 분리구간에서의 연평균강우량을 함께 나타내었다. 철원 강우관측소의 경우 과거자료에 대해서는 두 기법 모두 변동점을 찾지 못하였으나, 미래 자료에 대해서는 BCP 0.9가 1개씩의 변동점을 찾아내었다. 춘천 강우관측소의 경우는 과거자료에 대해서 두 기법 모두 변동점을 찾지 못하였고, 미래 자료에 대해서는 두 기법 모두 유사한 시기의 변동점을 1개씩 탐색하였음을 알 수 있다. 원주 강우관측소는 두 기법 모두 과거 자료에 대한 변동점을 탐색하지 못한 반면 BCP 0.9는 1개씩의 변동점을 탐색하였고, 대관령 관측소의 경우는 두 기법 모두 과거자료에 대해서 2개의 변동점을 탐색하였으나 미래자료에 대해서는 BCP 0.9가 RCP 4.5 자료에 대해서만 1개의 변동점을 탐색하였다. 마지막으로 인제 강우관측소의 경우는 두 기법의 결

과가 모든 자료에 대해 일치하고 있음(과거자료 변동점 2개, 미래자료 변동점 없음)을 알 수 있다. 이러한 분석결과를 보면, 철원, 춘천, 원주 관측소에서 관측된 과거 관측자료에 대해서는 두 기법 모두 변동점이 없는 것으로 탐색된 것을 알 수 있으며, 대관령과 인제 관측소에서 관측된 과거 관측자료에 대해서는 두 기법 모두 2개씩의 변동점을 탐색하였음을 알 수 있다. 미래 자료의 경우 BCP 0.9는 철원(RCP 4.5 & 8.5), 원주(RCP 4.5 & 8.5), 대관령(RCP 4.5)에서 1개씩의 변동점을 탐색하였으나 DP는 변동점을 탐색하지 못하여 DP기법의 미래자료에 대한 탐색능력이 BCP 0.9에 비해 다소 낮게 나타나는 특징을 알 수 있었다. 특히 BCP 0.9에서는 탐색되지 않았으나 DP는 탐색한 변동점이 나타나지 않아 우리나라 북동부 지역에 위치한 강우관측소의 변동점 분석을 위해서는 BCP 0.9기법이 보다 적절할 것으로 판단된다.

6. 결론 및 향후과제

본 연구에서는 최근 활발히 적용되고 있는 CUSUM기법, BCP기법 및 DP기법에 관한 연구동향 및 이론적 배경을 소개하고, 이들의 적용상의 특성 및 결과를 상호비교하는 연구를 수행하였다. 분석기법 간의 상호비교를 객관적으로 알아보기 위하여 먼저 강원지역에 위치하고 있는 철원, 춘천, 원주, 대관령, 인제 강우관측소로부터 추출된 강우자료의 통계치를 산정한 다음 이를 이용하여 100년의 길이를 가지는 변동점이 없는 동질특성 자료, 비동질 단일 변동점 자료, 비동질 다중 변동점 자료 1,000개를 생성

Table 8. Descriptive Statistical Characteristics of the Real Rainfall Data at 5 Rainfall Gauges

Cases	Gauges	Mean	Std.	Skewness	Kurtosis
Observed	Cheolwon (1988-2012)	1398.9	290.4	0.842	4.053
	Chuncheon (1966-2012)	1341.5	286.0	0.796	3.300
	Wonju (1972-2012)	1347.5	279.6	0.752	4.829
	Daegwanryeong (1972-2012)	1723.7	470.6	1.054	3.392
	Inje (1972-2012)	1180.6	290.2	0.290	2.611
Future (RCP 4.5)	Cheolwon	1282.5	258.8	0.349	2.527
	Chuncheon	1272.4	278.2	0.400	2.639
	Wonju	1367.6	318.7	0.741	4.649
	Daegwanryeong	1469.8	351.7	1.145	5.993
	Inje	1227.4	246.2	0.314	3.032
Future (RCP 8.5)	Cheolwon	1298.7	303.9	0.434	2.455
	Chuncheon	1261.5	304.2	0.547	2.942
	Wonju	1345.2	328.3	0.648	3.397
	Daegwanryeong	1434.7	309.0	1.154	2.455
	Inje	1212.2	289.7	0.653	3.336

Table 9. Result of change point analysis by BCP 0.9

Gauges	Data	Change points (Year)	Homogeneous period (Years)	Mean rainfall (mm)
Cheol-won	Obs.	Not found	1988-2012	1399
	RCP 4.5	2047	2013-2047	1237
			2048-2099	1313
	RCP 8.5	2043	2013-2043	1223
			2044-2099	1340
Chun-cheon	Obs.	Not found	1966-2012	1341
	RCP 4.5	2058	2013-2058	1209
			2059-2099	1344
	RCP 8.5	2048	2013-2048	1220
			2049-2099	1291
Wonju	Obs.	Not found	1972-2012	1348
	RCP 4.5	2059	2013-2059	1332
			2060-2099	1409
	RCP 8.5	2046	2013-2046	1327
			2047-2099	1366
Daegwan-ryeong	Obs.	1984, 2000	1972-1984	1441
			1985-2000	1942
			2001-2012	1740
	RCP 4.5	2064	2013-2064	1395
			2065-2099	1582
RCP 8.5	Not found	2013-2099	1435	
Inje	Obs.	1995, 2004	1972-1995	1083
			1996-2004	1289
			2005-2012	1352
	RCP 4.5	Not found	2013-2099	1227
	RCP 8.5	Not found	2013-2099	1212

Table 10. Result of Change Point Analysis by DP

Gauges	Data	Change points (Year)	Homogeneous period (Years)	Mean rainfall (mm)
Cheol-won	Obs.	Not found	1988-2012	1399
	RCP 4.5	Not found	2013-2099	1282
	RCP 8.5	Not found	2013-2099	1299
Chun-cheon	Obs.	Not found	1966-2012	1342
	RCP 4.5	2058	2013-2058	1209
			2059-2099	1344
	RCP 8.5	2043	2013-2043	1203
2044-2099			1294	
Wonju	Obs.	Not found	1972-2012	1348
	RCP 4.5	Not found	2013-2099	1368
	RCP 8.5	Not found	2013-2099	1345
Daegwan-ryeong	Obs.	1984, 2000	1972-1984	1441
			1985-2000	1942
			2001-2012	1740
	RCP 4.5	Not found	2013-2099	1470
RCP 8.5	Not found	2013-2099	1435	
Inje	Obs.	1995, 2004	1972-1995	1083
			1996-2004	1289
			2005-2012	1352
	RCP 4.5	Not found	2013-2099	1227
	RCP 8.5	Not found	2013-2099	1212

하였다. 생성된 모의강우자료들은 이미 변동점에 관한 모든 정보를 알고 있는 자료들이므로 이 자료들에 4개(BCP의 경우 0.9 및 0.7 사용)의 변동점 분석기법을 적용하고 그 결과들을 케이스에 따라 분류하여 적절한 상호비교 절차에 따라 비교하였다. 상호 비교결과 동질특성 자료의 경우는 BCP 0.9기법과 DP기법이 가장 우수한 결과를 제시하는 것으로 분석되었다.

또한 본 연구에서는 비교연구를 통해 제시된 BCP 0.9 기법과 DP기법을 5개 강우관측소에서 얻어진 실제 강우 자료에 적용함으로써 강원지역에 위치한 5개 강우관측소 지점의 변동점을 분석하였다. 특히 실제 강우자료를 적용함에 있어 본 연구에서는 관측자료 뿐만 아니라 기상청에서 모의한 RCP 4.5 및 RCP 8.5에 따르는 2099년까지의 미래 강우자료에 대한 변동점 분석도 함께 수행함으로써 미래 강우자료에 대한 변동점 정보도 알아보았다. 실제 강우자료를 사용한 분석결과 각각의 관측소 및 2개의 분석기법마다 조금씩 다른 결과를 보였으나, 우리나라 강원권에 위치한 강우관측소의 변동점 분석을 위해서는 BCP 0.9기법이 보다 적절할 것으로 판단되었다.

본 연구를 진행함에 있어 평균에 대한 변동점 분석이론을 평균과 분산을 함께 고려할 수 있는 이론으로 확장시킬 필요가 있다는 점과 과거 자료와 미래 자료를 결합시키는 경우 경계점에서의 처리를 위한 별도의 연구를 수행할 필요가 있다는 점이 제시되었으며, 이와 같은 향후 연구를 통해 보다 적절한 변동점 분석 결과를 제시할 수 있을 것으로 판단되었다. 이와 같은 강우자료에 대한 변동성 분석결과는 향후 특정지점에서의 미래 유출량을 모의하기 위한 장기 유출모형의 구축에 필요한 모형보정 등에 활용될 수 있을 것으로 판단되며, 특정지점에서의 유량자료에 적용하면 특정원인 때문에 발생하는 유량의 변동정보를 위치 및 크기 등의 정량적 정보로 확인할 수 있다는 유용성이 있을 것으로 판단된다.

감사의 글

본 연구는 2012년도 강원대학교 학술연구조성비(120120502) 및 국토교통부 물관리연구사업의 연구비지원(11기술혁신C06)에 의해 수행되었습니다. 논문의 심사를 통해 좋은 의견을 제시해주신 심사위원들께 감사드립니다.

References

Abdul Aziz, O.I., and Burn, D.H.T (2006). "Trends and

variability in the hydrological regime of the Mackenzie River Basin." *Journal of hydrology*, Vol. 319, pp. 282-294.

Aksoy, H., Gedikli, A., Unal, N.E., and Kehagias, A. (2008). "Fast segmentation algorithms for long hydrometeorological time series." *Hydrological Processes*, Vol. 22, No. 23, pp. 4600-4608.

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.

Beaulieu, C., Seidou, O., Ouarda, B.M.J., and Zhang, X. (2009). "Intercomparison of homogenization techniques for precipitation data continued: Comparison of two recent Bayesian change point models." *Water Resources Research*, Vol. 45, W08410.

Burn, D.H. (1994). "Hydrologic effects of climatic change in West Central Canada." *Journal of Hydrology*, Vol. 160, pp. 53-70.

Carlin, B.P., Gelfand, A.E., and Smith, A.F.M. (1992). "Hierarchical Bayesian analysis of changepoint problems." *Applied Statistics*, Vol. 41, No. 2, pp. 389-405.

Carslaw, D.C., Ropkins, K., and Bell, M.C. (2006). "Change-point detection of gaseous and particulate traffic-related pollutants at a roadside location." *Environ. Sci. Technol.* Vol. 40, No. 22, pp. 6912-6918.

Chelani, A.B. (2011). "Change detection using CUSUM and modified CUSUM method in air pollutant concentrations at traffic site in Delhi." *Stochastic Environmental Research and Risk Assessment*, Vol. 25, pp. 827-834.

Chib, S. (1998). "Estimation and comparison of multiple change-point models." *Journal of Econometrics*, Vol. 86, pp. 221-241.

Chowdhury, R.K., and Beecham, S. (2010). "Australian rainfall trends and their relation to the southern oscillation index." *Hydrological Processes*, Vol. 24, No. 4, pp. 504-514.

Chu, H.J., Pan, T.Y., and Liou, J.J. (2012). "Change-point detection of long-duration extreme precipitation and the effect on hydrologic design: a case study of south

- Taiwan." *Stochastic Environmental Research and Risk Assessment*, Vol. 26, pp. 1123-1130.
- Dobigeon, N., and Tourmeret, J.Y. (2007) "Joint segmentation of wind speed and direction using a hierarchical model." *Comput. Statist. Data Anal.*, Vol. 51, pp. 5603-5621.
- Domonkos, P. (2011). "Efficiency evaluation for detecting inhomogeneities by objective homogenization methods." *Theoretical and Applied Climatology*, Vol. 105, No. 3-4, pp. 455-467.
- Erdman, C., and Emerson, J.C. (2007). "bcp: an R package for performing a Bayesian analysis of change point problems." *Journal of Statistical Software*, Vol. 23, No. 3, pp. 1-13.
- Fearnhead, P. (2006). "Exact and efficient Bayesian inference for multiple change point problems." *Statistics and Computing*, Vol. 16, pp. 203-213.
- Fearnhead, P., and Liu, Z. (2011). "Efficient Bayesian analysis of multiple changepoint models with dependence across segments." *Statistics and Computing*, Vol. 21, pp. 217-229.
- Gedikli, A., Aksoy, H., and Unal, N.E. (2008). "Segmentation algorithm for long time series analysis." *Stochastic Environmental Research and Risk Assessment*, Vol. 22, pp. 291-302.
- Hirsch, R.M., Slack, J.R., and Smith, R.A. (1982). "Techniques of trend analysis for monthly water quality data." *Water Resources Research*, Vol. 18, pp. 107-121.
- Hwang, S.H., Kim, J.H., Yoo, C., and Jung, S.W. (2010). "A Probabilistic estimation of changing points of Seoul rainfall using BH Bayesian analysis." *Journal of Korea Water Resources Association*, Vol. 43, No. 7, pp. 645-655 (In Korean).
- Kampata, J.M., Parida, B.P., and Moalafhi, D.B. (2008). "Trend analysis of rainfall in the headstreams of the Zambezi river basin in Zambia." *Phys. Chem. Earth*, Vol. 33, pp. 621-625.
- Kehagias, A., Nidelkou, E., and Petridis, V. (2006). "A dynamic programming segmentation procedure for hydrological and environmental time series." *Stochastic Environmental Research and Risk Assessment*, Vol. 20, pp. 77-94.
- Kim, J., and Cheon, S. (2010). "Bayesian multiple change-point estimation with annealing stochastic approximation Monte Carlo." *Computational Statistics*, Vol. 25, pp. 215-239.
- Kim, C., Suh, M.S., and Hong, K.O. (2009). "Bayesian changepoint analysis of the annual maximum of daily and subdaily precipitation over South Korea." *Journal of Climate*, Vol. 15, pp. 6741-6757.
- Li, Z.L., Xu, Z.X., Li, J.Y., and Li, Z.J. (2008). "Shift trend and step changes for runoff time series in the Shiyang River basin, northwest China." *Hydrological Processes*, Vol. 22, No. 23, pp. 4639-4646.
- Page, E.S. (1954). "Continuous inspection scheme." *Biometrika*, Vol. 41, pp. 100-115.
- Yao, Y.C. (1984). "Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches." *The Annals of Statistics*, Vol. 12, No. 4, pp. 1434-1447.

논문번호: 14-003	접수: 2014.01.02
수정일자: 2014.02.06/03.12	심사완료: 2014.03.12