



A study on a tendency of parameters for nonstationary distribution using ensemble empirical mode decomposition method

Kim, Hanbeen^a · Kim, Taereem^a · Shin, Hongjoon^b · Heo, Jun-Haeng^{a*}

^aSchool of Civil and Environmental Engineering, Yonsei University

^bUniversity-Industry Foundation, Yonsei University

Paper number: 17-016

Received: 15 February 2017; Revised: 15 March 2017 / 19 March 2017; Accepted: 19 March 2017

Abstract

A lot of nonstationary frequency analyses have been studied in recent years as the nonstationarity occurs in hydrologic time series data. In nonstationary frequency analysis, various forms of probability distributions have been proposed to consider the time-dependent statistical characteristics of nonstationary data, and various methods for parameter estimation also have been studied. In this study, we aim to introduce a parameter estimation method for nonstationary Gumbel distribution using ensemble empirical mode decomposition (EEMD); and to compare the results with the method of maximum likelihood. Annual maximum rainfall data with a trend observed by Korea Meteorological Administration (KMA) was applied. As a result, both EEMD and the method of maximum likelihood selected an appropriate nonstationary Gumbel distribution for linear trend data, while the EEMD selected more appropriate nonstationary Gumbel distribution than the method of maximum likelihood for quadratic trend data.

Keywords: Ensemble empirical mode decomposition, Method of maximum likelihood, Nonstationary Gumbel distribution, Parameter estimation

앙상블 경험적 모드분해법을 활용한 비정상성 확률분포형의 매개변수 추세 분석에 관한 연구

김한빈^a · 김태림^a · 신희준^b · 허준행^{a*}

^a연세대학교 토목환경공학과, ^b연세대학교 산학협력단

요 지

최근 수문자료에서 비정상성 현상들이 관측됨에 따라 비정상성 빈도해석에 관한 연구들이 활발하게 진행되고 있다. 시간에 따라 변화하는 통계적 특성을 고려하기 위하여 다양한 형태의 비정상성 확률분포형이 제시되고 있으며, 비정상성 매개변수를 추정할 수 있는 다양한 방법들이 연구되고 있는 추세이다. 본 연구에서는 앙상블 경험적 모드분해법을 이용한 비정상성 Gumbel 분포형의 매개변수 추정방법을 제시하고 기존에 비정상성 매개변수 추정방법으로 주로 사용되어온 최우도법과 비교해보고자 하였다. 국내 자료의 적용을 위하여 기상청 지점의 다양한 지속기간에 대해 경향성이 나타나는 연 최대치 강우자료를 사용하였다. 적용 결과 선형적 경향성을 나타내는 자료에 대해서는 두 가지 방법 모두 적절한 모형을 선정하였으나, 2차 곡선 형태의 경향성이 존재하는 자료에 대해서는 앙상블 경험적 모드분해법의 경우에만 이러한 경향성을 반영하는 비정상성 Gumbel 모형을 선정하였다.

핵심용어: 앙상블 경험적 모드분해법, 비정상성 빈도해석, 매개변수 추정방법, 최우도법, Gumbel 분포형

*Corresponding Author. Tel: +82-2-2123-2805
E-mail: jhheo@yonsei.ac.kr (J.-H. Heo)

1. 서론

관측된 수문자료에 대하여 적절한 확률분포형을 선정하고 이를 통해 확률수문량을 산정하는 수문빈도해석은 수자원 관리 및 이·치수를 위한 수공구조물의 설계에 있어서 매우 중요한 과정이다. 수문빈도해석은 관측 자료의 통계적 특성이 시간에 따라 변하지 않는 정상성(stationarity) 가정을 바탕으로 수행되어 왔다. 하지만 전 지구적인 기후변화로 인해 극치수문사상의 발생 빈도가 증가하고 있으며(IPCC, 2014) 수문자료에서도 통계적 특성이 시간에 따라 변하는 등의 비정상성 현상이 관측되고 있다(Olsen et al., 1999; Jain and Lall, 2000; Milly et al., 2008). 또한 국내에서도 극한강우의 발생빈도와 강우강도가 증가 등이 관측되고 있다(KMA, 2014). 이에 따라 비정상성을 고려한 빈도해석에 대한 연구가 진행되고 있으며, 특히 비정상성 확률분포형의 적용에 대한 연구가 활발히 진행되고 있다. Leadbetter et al. (1983)은 비정상성 조건에서 기존의 상수 매개변수를 가지는 확률분포형을 적용하는 것은 더 이상 유효하지 않으므로 비정상성 수문자료의 분석을 위한 확률분포형의 개발을 제안하였다. Coles (2001)는 기존 확률분포형의 매개변수를 시간에 대한 함수로 정의한 비정상성 확률분포형을 제시하였으며, 이러한 비정상성 확률분포형은 홍수량, 온도 등과 같은 수문량 뿐만 아니라 강우빈도해석에도 널리 사용되고 있다(Villarini et al., 2010; Katz, 2013; Trambly et al., 2013; Vasiliades et al., 2015).

국내에서 수행된 강우자료에 대한 비정상성 빈도해석에 관한 연구를 살펴보면, Lee et al. (2010)은 Peaks-Over-Threshold (POT) 자료를 비정상성 Generalized Extreme Value (GEV) 분포형에 적용하였으며, Bayesian 통계기법을 적용하여 비정상성이 반영된 월별 설계강우량을 산정하였고, Kwon and Lee (2011)는 계층적 Bayesian 통계기법을 적용하여 비정상성 Gumbel 분포형의 매개변수의 사후분포를 추정하였으며, 이를 통해 설계강우량의 사후분포 및 불확실성을 예측하였다. Shin et al. (2014)은 비정상성 Gumbel 분포형을 이용해 강우빈도해석을 수행하였으며, 비정상성을 고려한 재현기간을 산정하였다. Jang et al. (2015)은 비정상성 Gumbel 분포형의 규모매개변수를 지수함수, 선형함수, 로그함수의 형태로 가정하여 그 성능을 비교하였으며, 강우 자료를 대상으로 비정상성 빈도해석을 수행하여 적용성을 검토하였다. 또한 Wi et al. (2016)은 다양한 지속기간의 국내 연 최대치 강우자료 및 POT 자료를 비정상성 GEV 및 Generalized Pareto (GP) 분포형에 적용하여 비정상성 강우빈도해석을 수행하였다.

일반적으로 비정상성 확률분포형에 대한 매개변수 추정에는 최우도법이 널리 이용되고 있으며(Coles, 2001; O'Brien and Burn, 2014; Jang et al., 2015; Wi et al., 2016), 이 외에도 다양한 방법들이 연구되고 있다. 수문자료에 내포된 경향성의 형태를 매개변수 추정에 고려하기 위해 Chen et al. (2017)은 비정상성 빈도해석을 위한 매개변수 추정방법으로 가중 최소 자승 추정법(weighted least squares estimation; WLS), 이산 웨이블렛 변환(discrete wavelet transform; DWT), 앙상블 경험적 모드분해법(ensemble empirical mode decomposition; EEMD)을 이용하여 자료로부터 결정되는 추세를 이용한 매개변수 추정을 제안하였고, 세 방법의 결과를 비교하였다. 여기서 적용된 앙상블 경험적 모드분해법은 Huang et al. (1998)에 의해 제안된 경험적 모드분해법이 가진 단점을 보완하기 위하여 Wu and Huang (2009)에 의해 개선된 방법이다. 자료에 내재된 주기와 경향성을 나타내는 요소를 추출할 수 있는 분해분석법으로 비선형, 비정상성인 시계열인 경우 더욱 효과적으로 분해할 수 있다는 장점을 보이며 최근 수문 분야에서 활발하게 적용되고 있다(Huang and Wu, 2008; Breaker and Ruzmaikim, 2011; Lee et al., 2013; Di et al., 2014; Wang et al., 2015). 그 중에서도 자료에 내포된 다양한 주기를 제외한 잔여값에 대한 연구를 살펴보면, Wu et al. (2007)은 주기가 제거된 후 남은 잔여값을 자료의 경향성이라 정의한 후 경험적 모드분해법을 이용하여 비선형, 비정상성인 시계열에서 통계적, 물리적으로 의미있는 경향성의 추출을 제안하였고, Sang et al. (2014)는 수문자료의 경향성 여부를 판단하기 위하여 기존의 Mann-Kendall 방법과 경험적 모드분해법으로부터 추출된 잔여값을 이용하여 경향성 분석을 수행하였다. Kim and Cho (2016)는 해수면 상승의 경향과 가속을 추정하기 위하여 해수면 시계열 자료에 앙상블 경험적 모드분해법을 적용하여 추출된 잔여값에 회귀분석을 수행하여 추세를 분석하는 등 잔여값이 가지는 특성을 이용한 연구가 다양하게 진행되고 있다.

따라서 본 연구에서는 Chen et al. (2017)이 새롭게 제안한 비정상성 빈도해석을 위한 매개변수 추정방법 중 앙상블 경험적 모드분해법을 통해 분해된 잔여값이 가지고 있는 특성을 중점으로 연구를 수행하였으며, 국내 활용 및 기존의 최우도법과의 비교를 위하여 실제 국내 강우자료에 적용하여 비정상성 빈도해석을 수행하고 최적 모형을 선정하여 그 결과를 비교 분석하였다.

2. 방법론

2.1 Gumbel 분포형

Gumbel 분포형은 극치 수문사상에 대한 빈도해석에 널리 사용되며 위치 매개변수(location parameter)와 규모 매개변수(scale parameter)를 가지는 2변수 분포형으로 특히 국내 확률강우량의 산정에 대표적으로 사용되고 있는 분포형이다. 이러한 Gumbel 분포형의 확률밀도함수(probability density function, PDF)와 누가분포함수(cumulative distribution function, CDF)는 Eqs. (1) and (2)와 같다.

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

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

여기서, ϵ 은 위치 매개변수, α 는 규모 매개변수이다.

2.2 최우도법을 이용한 비정상성 Gumbel 분포형의 매개변수 추정

비정상성 Gumbel 분포형은 매개변수를 시간(t)에 대한 함수로 가정하여 나타낼 수 있으며, 매개변수 추정방법으로는 일반적으로 최우도법(method of maximum likelihood; ML)이 이용된다. 최우도법은 주어진 자료들의 발생결합확률 또는 확률분포 매개변수의 함수로 정의되는 우도함수를 결정하고, 이 우도함수를 최대화하는 매개변수를 최종적으로 산정하는 매개변수 추정방법이다. 본 연구에서는 위치 매개변수와 규모 매개변수를 시간에 대한 1차식(linear form)과 2차식

Table 1. Nonstationary Gumbel distributions for ML

Distribution	Number of parameters	Location parameter ($\epsilon(t)$)	Scale parameter ($\alpha(t)$)
ML-GUM (1,0)	3	$\epsilon_0 + \epsilon_1 t$	α_0
ML-GUM (0,1)	3	ϵ_0	$\alpha_0 + \alpha_1 t$
ML-GUM (1,1)	4	$\epsilon_0 + \epsilon_1 t$	$\alpha_0 + \alpha_1 t$
ML-GUM (2,0)	4	$\epsilon_0 + \epsilon_1 t + \epsilon_2 t^2$	α_0
ML-GUM (0,2)	4	ϵ_0	$\alpha_0 + \alpha_1 t + \alpha_2 t^2$
ML-GUM (2,1)	5	$\epsilon_0 + \epsilon_1 t + \epsilon_2 t^2$	$\alpha_0 + \alpha_1 t$
ML-GUM (1,2)	5	$\epsilon_0 + \epsilon_1 t$	$\alpha_0 + \alpha_1 t + \alpha_2 t^2$
ML-GUM (2,2)	6	$\epsilon_0 + \epsilon_1 t + \epsilon_2 t^2$	$\alpha_0 + \alpha_1 t + \alpha_2 t^2$

(quadratic form)으로 나타내어 비정상성을 고려하였으며 각 모형을 Table 1과 같이 제시하였다.

N 개의 관측자료(x_1, \dots, x_N)에 대한 비정상성 Gumbel 분포형의 대수우도함수($\ell(\theta)$)는 Eq. (3)과 같다.

$$\begin{aligned} \ell(\theta) &= \sum_{t=1}^N \log(f_{Gumbel}(x_t; \epsilon(t), \alpha(t))) \quad (3) \\ &= \sum_{t=1}^N \log\left[\frac{1}{\alpha(t)} \exp\left\{-\frac{x_t - \epsilon(t)}{\alpha(t)} - \exp\left\{-\frac{x_t - \epsilon(t)}{\alpha(t)}\right\}\right\}\right] \\ &= -N \log(\alpha(t)) - \sum_{t=1}^N \frac{x_t - \epsilon(t)}{\alpha(t)} - \sum_{t=1}^N \exp\left\{-\frac{x_t - \epsilon(t)}{\alpha(t)}\right\} \end{aligned}$$

여기서, θ 는 비정상성 Gumbel 분포형의 매개변수의 벡터, $f_{Gumbel}(x_t; \epsilon(t), \alpha(t))$ 는 매개변수 $\epsilon(t)$ 와 $\alpha(t)$ 를 가지는 비정상성 Gumbel 분포형의 확률밀도함수를 나타낸다. $\epsilon(t)$ 와 $\alpha(t)$ 는 Table 1에서 제시하는 형태를 사용함으로써 다양한 비정상성 Gumbel 분포형에 대한 대수우도함수($\ell(\theta)$)를 구할 수 있으며, 대수우도함수($\ell(\theta)$)를 최대화하는 매개변수 $\epsilon(t)$ 와 $\alpha(t)$ 를 Newton-Raphson 방법, Broyden-Fletcher-Goldfarb-Shanno (BFGS) 방법 등과 같은 최적화기법을 통해 추정할 수 있다. 본 연구에서는 1, 2차 도함수의 유도없이 최적화가 가능하여 비정상성 빈도해석의 매개변수추정에 널리 사용되는 BFGS 방법을 이용하여 비정상성 Gumbel 분포형의 매개변수를 추정하였다.

2.3 앙상블 경험적 모드분해법을 이용한 비정상성 Gumbel 분포형의 매개변수 추정

Huang et al. (1998)에 의해 개발된 경험적 모드분해법(empirical mode decomposition; EMD)은 체거름 알고리즘(sifting algorithm)을 통해 주어진 자료를 일련의 고유한 내재 모드함수(intrinsic mode function; IMF_i)와 잔여값(Residue)으로 분해하는 분석법이다. 따라서 자료에 내포된 다양한 주기를 분해한 후 추출되는 자료가 가진 고유의 장기적 경향성을 알 수 있다(Wu et al., 2007). EMD를 통해 분해된 내재모드함수는 전체 자료 계열에서 극값(최댓값 및 최솟값)의 개수와 부호변화점(zero crossing)의 개수는 같거나 최소 1이어야 하고, 어느 지점에서든 상위선과 하위선의 평균은 0이어야 한다는 두 가지 조건을 만족해야 한다. 주어진 시계열 자료를 $y(t)$, $t = 1, 2, 3, \dots, n$ 이라 할 경우 자료를 분해하는 체거름 과정은 아래와 같다(Huang et al., 1998; Wu and Huang, 2004; Kim et al., 2015).

- (1) 주어진 시계열의 극값(국소 최댓값 및 최솟값)을 식별

한 후 삼차 스플라인(cubic spline) 보간법을 이용하여 각각 상위선(upper envelope, $y_u(t)$)과 하위선(lower envelope, $y_l(t)$)을 구한다.

- (2) 상위선과 하위선의 평균선(mean envelope, $y_m(t) = (y_u(t) + y_l(t))/2$)을 구한다.
- (3) 원자료계열 $y(t)$ 에서 평균선을 공제한 후($y(t) - y_m(t)$)의 자료계열이 위에서 언급된 내재모드함수의 조건을 만족할 경우 이를 추출된 내재모드함수 IMF_i 라 정의한다.
- (4) (3)에서 추출된 내재모드함수를 제거한 후의 시계열을 원자료계열 $y(t)$ 로 재설정하고, 남은 자료계열이 단조함수 또는 하나의 극값만 존재하여 더 이상 새로운 계열이 추출되지 않을 때까지 (1)-(3)의 과정을 반복한다.
- (5) 최종적으로 원자료계열 $y(t)$ 는 Eq. (4)와 같이 분해된 내재모드함수(IMF_i)와 잔여값($Residue$)의 합으로 나타낼 수 있다.

$$y(t) = \sum_{i=1}^N IMF_i + Residue \quad (4)$$

여기서 N 은 추출된 내재모드함수의 개수이다.

하지만 EMD를 통해 자료를 분해할 경우 mode-mixing의 문제가 발생되므로 이를 해결하기 위해 주어진 자료에 인위적으로 백색잡음을 추가한 후 자료를 분해하고 추후 잡음을 제거하는 앙상블 경험적 모드분해법(ensemble empirical mode decomposition; EEMD)이 개발되었다(Wu and Huang, 2009). 본 연구에서는 EEMD를 이용하여 자료 분해를 수행하였으며, 이를 통해 추출된 잔여값은 자료에 내재된 장기 경향성(long-term trend)을 나타낸다(Wu et al. 2007).

EEMD를 이용한 매개변수 추정을 위해 본 연구에서는 Chen et al. (2017)이 제안한 1, 2차 모멘트의 성질을 이용하여 원자료인 연최대지 강우자료를 분해했을 때의 잔여값을 평균잔여값(mean-residue)이라 정의하고, 원자료와 평균잔여값의 잔차를 제공하는 시계열(squared deviation from mean-residue)을 EEMD를 이용해 분해했을 때의 잔여값을 분산잔여값(variance-residue)이라 정의하였다. 비정상성을 고려하기 위해 평균잔여값과 분산잔여값을 각각 1차식(linear form) 및 2차식(quadratic form)으로 정의하였으며 회귀분석을 통해 각 계수를 산정하였다. 산정된 평균잔여값($\mu(t)$)과 분산잔여값($v(t)$)을 모멘트법 기반의 Gumbel 분포형 매개변수 추정식인 Eqs. (5) and (6)에 대입하여 위치 매개변수($\epsilon(t)$) 및 규모 매개변수($\alpha(t)$)를 추정하였다.

Table 2. Nonstationary Gumbel distributions for EEMD

Distribution	Number of parameters	Mean-residue ($\mu(t)$)	Variance-residue ($v(t)$)
EEMD-GUM (1,0)	3	$\mu_0 + \mu_1 t$	v_0
EEMD-GUM (0,1)	3	μ_0	$v_0 + v_1 t$
EEMD-GUM (1,1)	4	$\mu_0 + \mu_1 t$	$v_0 + v_1 t$
EEMD-GUM (2,0)	4	$\mu_0 + \mu_1 t + \mu_2 t^2$	v_0
EEMD-GUM (0,2)	4	μ_0	$v_0 + v_1 t + v_2 t^2$
EEMD-GUM (2,1)	5	$\mu_0 + \mu_1 t + \mu_2 t^2$	$v_0 + v_1 t$
EEMD-GUM (1,2)	5	$\mu_0 + \mu_1 t$	$v_0 + v_1 t + v_2 t^2$
EEMD-GUM (2,2)	6	$\mu_0 + \mu_1 t + \mu_2 t^2$	$v_0 + v_1 t + v_2 t^2$

$$\epsilon(t) = \mu(t) - 0.5772 \sqrt{v(t)} \quad (5)$$

$$\alpha(t) = 0.7797 \sqrt{v(t)} \quad (6)$$

EEMD를 통해 분해된 평균잔여값과 분산잔여값에 대한 식의 형태에 따라 정의된 비정상성 Gumbel 분포형을 Table 2와 같이 제시하였다.

2.4 최적 확률분포형 선정

기존의 정상성 확률분포형을 이용한 지점빈도해석에서는 최적 분포형 선정을 위한 적합도 검정 방법으로 χ^2 -검정, Kolmogorov-Smirnov (KS)검정, Cramer von Mises (CVM) 검정, Probability Plot Correlation Coefficient (PPCC) 검정 등을 사용한다. 하지만 비정상성 확률분포형을 적용하는 경우, 확률분포형이 시간에 따라 계속 변화하므로 기존의 적합도 검정 방법을 적용하는데 어려움이 있다. Akaike (1973)에 의해 개발된 모형 평가 방법인 AIC (Akaike's Information Criterion)는 두 모형간의 차이를 나타내는 Kullback-Leibler information (Kullback and Leibler, 1951)과 매개변수의 추정에 사용되는 대수우도함수의 관계를 이용하여 Eq. (7)과 같이 정의된다.

$$AIC = -2\ell(\theta) + 2k \quad (7)$$

여기서, $\ell(\theta)$ 은 대수우도함수, k 는 모형의 매개변수이며, AIC 값의 절대적인 크기에 관계없이 그 값이 가장 작은 확률분포형을 최적 분포형으로 선정한다. 비정상성 확률분포형의 경우에도 $\ell(\theta)$ 이 산정된 경우 간단하게 AIC값을 구할 수 있으므로, 비정상성 빈도해석에서의 최적 분포형 선정 방법으로 널

리 이용되고 있다(Laio et al., 2009; Villarini et al., 2009, 2010; Katz, 2013).

3. 적용 및 결과

3.1 대상 지점

본 연구에서는 자료기간이 30년 이상인 국내 기상청 64개 지점의 다양한 지속기간에 대한 연 최대치 강우자료를 대상으로 Mann-Kendall 경향성 분석을 수행하였으며, 1%의 유의 수준에서 경향성이 있는 것으로 나타나는 자료들을 대상으로 연구를 진행하였다. Fig. 1은 경향성이 나타나는 자료가 포함된 강우관측지점을 지도상에 나타낸 것이며, 지점에 대한 자료 길이 및 지속기간 별 경향성 유무는 Table 3에 나타내었다.

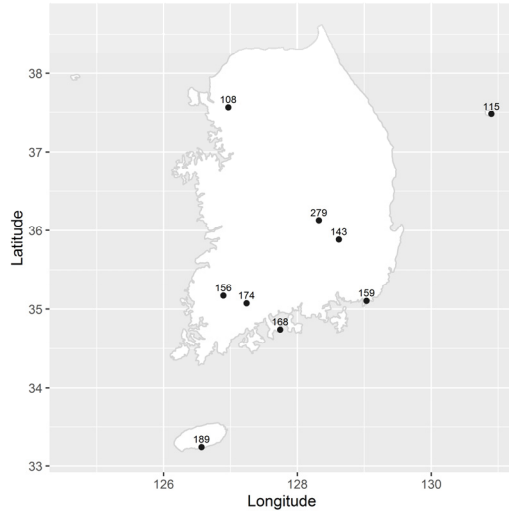


Fig. 1. Location of sites used in this study

3.2 앙상블 경험적 모드분해법

EEMD를 이용하여 경향성이 나타나는 자료가 포함된 강우관측지점의 강우 자료를 분해한 결과, 자료에 내재된 변동성이 평균적으로 짧은 주기에서 긴 주기 순으로 추출된 후 최종적으로 단조함수 또는 하나의 극값만 존재하는 형태의 잔여값이 추출되었다. 각 지점별로 4개 또는 5개의 내재모드함수로 분해되었으며(Table 4), Fig. 2는 서울(108) 지점의 지속기간 9시간 자료를 EEMD로 분해한 결과를 대표적으로 나타낸 것이다.

3.3 비정상성 Gumbel 분포형 선정

경향성이 나타나는 자료에 대하여 EEMD와 최우도법을 이용해 앞서 제시된 Tables 1 and 2의 8가지 비정상성 Gumbel 분포형에 대한 매개변수와 AIC값을 산정하였으며, 기존의 정상성 빈도해석 방법과의 비교를 위하여 정상성 Gumbel 분포형(GUM(0,0))에 대한 AIC값을 산정하였다. Table 5는 경향성이 나타나는 강우자료(서울(108)-9, 12시간, 울릉도

Table 4. The number of IMF at each site

Site number	Site name	Data period	The number of IMF
108	Seoul	1935-2015	5
115	Ulleungdo	1944-2015	5
143	Daegu	1916-2015	5
156	Gwangju	1939-2015	5
159	Busan	1943-2015	5
168	Yoesu	1943-2015	5
174	Suncheon	1972-2015	4
189	Seogwipo	1961-2015	4
279	Gumi	1972-2015	4

Table 3. Site information and results of Mann-Kendall trend test

Site number	Site name	Data period	Mann-Kendall trend test									
			1HR	2HR	3HR	6HR	9HR	12HR	15HR	18HR	24HR	48HR
108	Seoul	1935-2015	NT	NT	NT	NT	T	T	T	T	NT	T
115	Ulleungdo	1944-2015	T	T	T	T	T	T	T	T	T	T
143	Daegu	1916-2015	NT	NT	NT	NT	NT	T	T	T	T	T
156	Gwangju	1939-2015	NT	T	NT	NT	NT	NT	NT	NT	NT	T
159	Busan	1943-2015	NT	NT	NT	NT	NT	NT	NT	NT	NT	T
168	Yoesu	1943-2015	T	T	T	T	T	T	NT	NT	NT	NT
174	Suncheon	1972-2015	T	T	NT	NT	NT	NT	NT	NT	NT	NT
189	Seogwipo	1961-2015	NT	NT	NT	NT	NT	NT	T	NT	NT	NT
279	Gumi	1972-2015	NT	NT	NT	NT	NT	NT	NT	NT	NT	T

※ T: Trend, NT: No trend

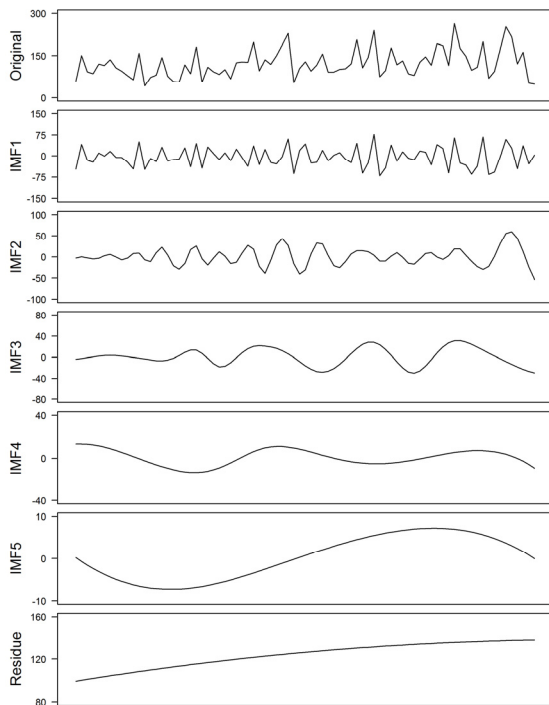


Fig. 2. Result of EEMD at Seoul (108) site (Duration: 9HR)

(115)-6시간, 대구(143)-12시간, 구미(279)-48시간)에 대한 정상성 및 비정상성 Gumbel 분포형의 AIC값 산정 결과를 나타낸 것이며, 각 비정상성 Gumbel 분포형의 AIC값 중 가장 작은 값을 Bold로 표기하였다. 경향성이 나타나는 자료에 대하여 빈도해석을 수행한 경우, GUM(0,0)을 적용했을 때보다 더 낮은 AIC값을 가지는 비정상성 Gumbel 분포형이 존재함을 Table 5를 통해 확인할 수 있다. 동일한 비정상성 Gumbel 분포형에 대하여 매개변수 추정 방법에 따른 AIC값을 비교해보면 모든 경우에서 최우도법이 EEMD보다 낮은 AIC값을 가

지는 것을 확인할 수 있다. 매개변수 추정방법 별로 선정된 모형을 비교해보면, 108(9), 115(6), 143(12) 지점의 경우 EEMD와 최우도법 모두 동일한 분포형을 최적 분포형으로 선정하였다. Fig. 3은 위의 지점에 대한 연 최대강우자료와 강우자료의 분산값에 대한 시계열 및 이를 EEMD로 분해한 평균잔여값과 분산잔여값을 나타낸 그래프로, 점선은 원자료, 실선은 잔여값을 나타낸다. 이를 통해 잔여값이 시계열 자료의 추세를 잘 나타냄을 확인할 수 있다. 108(9) 지점의 경우, 평균잔여값과 분산잔여값 모두 1차(선형) 경향성을 나타내고 있으며, EEMD와 최우도법 모두 GUM(1,1)을 최적 분포형으로 선정하였다. 115(6)과 143(12)의 경우 평균잔여값은 1차(선형) 경향성을 보이고 분산잔여값에서는 경향성이 존재하지 않는 것으로 나타났으며, EEMD와 최우도법 모두 GUM(1,0)을 최적 분포형으로 선정하였다.

대부분의 지점에서 위의 결과와 마찬가지로 EEMD와 최우도법 모두 동일한 분포형을 선정하였으나, Table 5의 108(12), 279(48) 지점의 경우 서로 다른 분포형을 선정하였다. Fig. 4는 108(12), 279(48) 지점에 대한 자료의 시계열과 잔여값을 나타낸 그래프이다. 두 지점의 평균잔여값은 2차(곡선) 경향성을 보이며, 분산잔여값은 1차(선형) 경향성을 보임을 확인할 수 있다. 그러나 최우도법을 통한 최적 분포형 선정 결과 두 지점 모두 평균과 분산에 1차(선형) 경향성이 고려된 GUM(1,1)이 선정된 반면, EEMD를 통한 최적 분포형 선정 결과 평균에는 2차(곡선) 경향성, 분산에는 1차(선형) 경향성이 고려된 GUM(2,1)이 선정되었다. 따라서 자료의 시계열이 2차(곡선) 경향성을 가지는 경우, 최우도법을 통해 선정된 모형의 AIC값이 가장 낮음에도 불구하고 EEMD 방법이 자료의 2차(곡선) 경향성을 잘 반영하여 최적 분포형을 추정하는 것을 확인하였다.

Table 5. AIC values for stationary and nonstationary Gumbel distributions

Model	Site number (duration)									
	108 (9)		108 (12)		115 (6)		143 (12)		279 (48)	
	EEMD	ML	EEMD	ML	EEMD	ML	EEMD	ML	EEMD	ML
GUM (0,0)	854.65	854.07	869.33	869.27	688.12	680.41	983.17	979.01	461.50	458.48
GUM (1,0)	853.25	852.68	869.51	868.42	674.48	670.15	979.97	973.36	461.52	457.47
GUM (0,1)	857.17	853.84	870.00	868.09	686.65	679.55	985.15	980.90	465.83	455.77
GUM (1,1)	849.82	847.75	867.85	862.20	675.19	671.76	981.95	974.62	462.43	447.99
GUM (2,0)	853.60	851.25	869.38	867.10	677.29	672.02	982.94	973.91	461.64	457.31
GUM (0,2)	861.70	854.84	875.11	869.30	698.18	681.13	986.98	982.77	469.04	457.73
GUM (2,1)	850.73	848.98	867.78	863.43	677.89	675.18	984.91	975.42	460.38	449.66
GUM (1,2)	853.29	847.98	872.10	862.78	681.59	673.73	984.32	975.65	465.09	449.99
GUM (2,2)	853.95	849.56	870.99	864.51	686.17	675.51	987.26	976.90	461.99	451.56

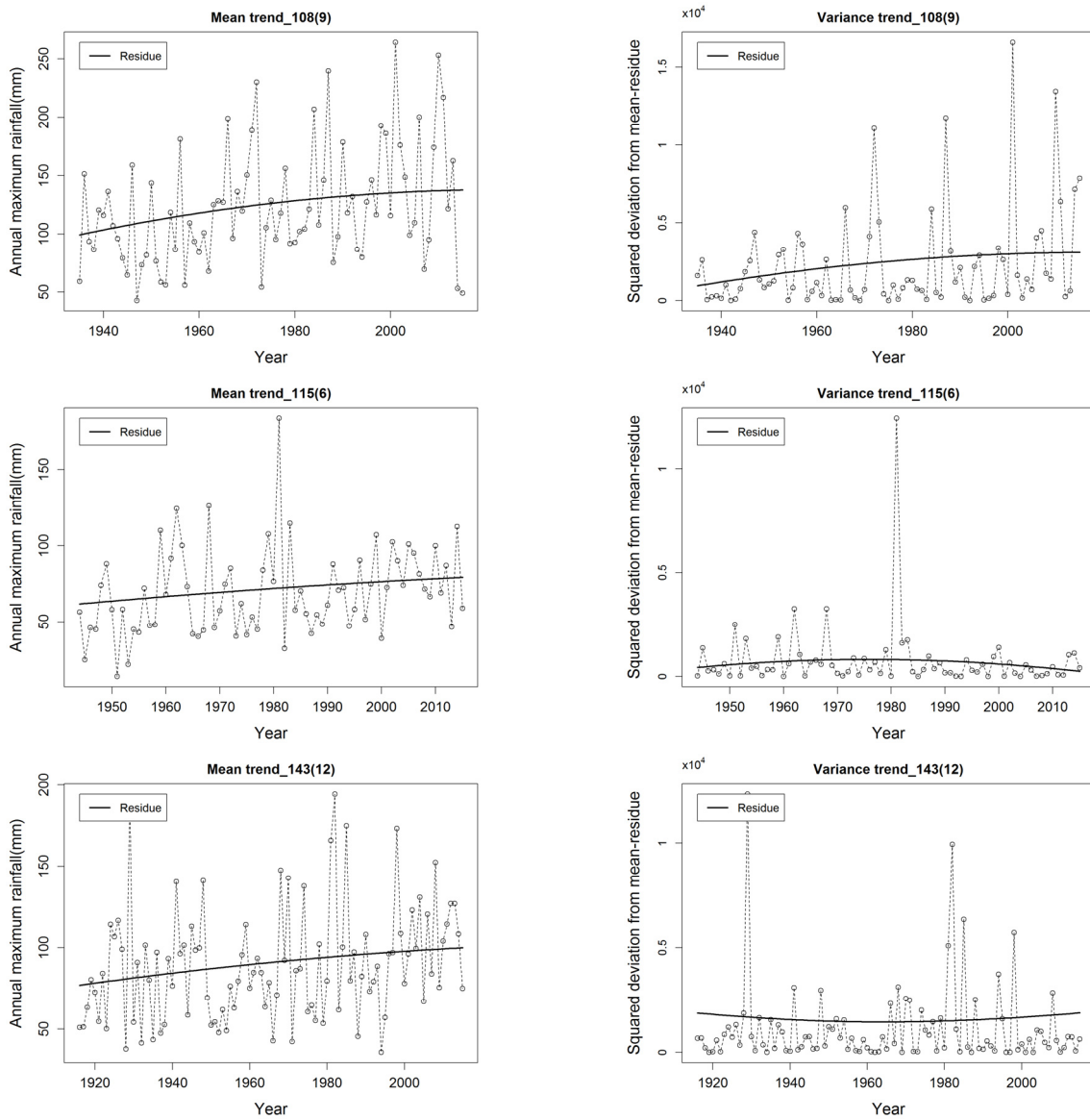


Fig. 3. Time series data and mean- and variance-residues (in case that the EEMD and ML methods have the same results)

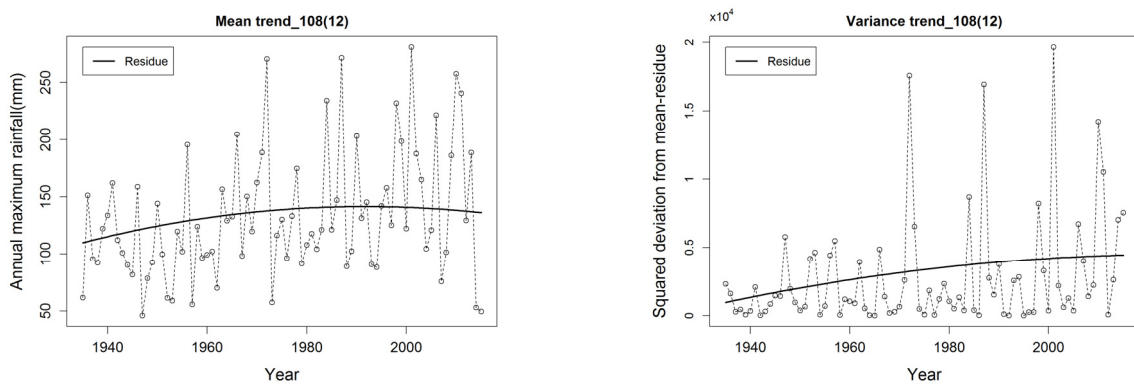


Fig. 4. Time series data and mean- and variance- residues (in case that the EEMD and ML methods have the different results)

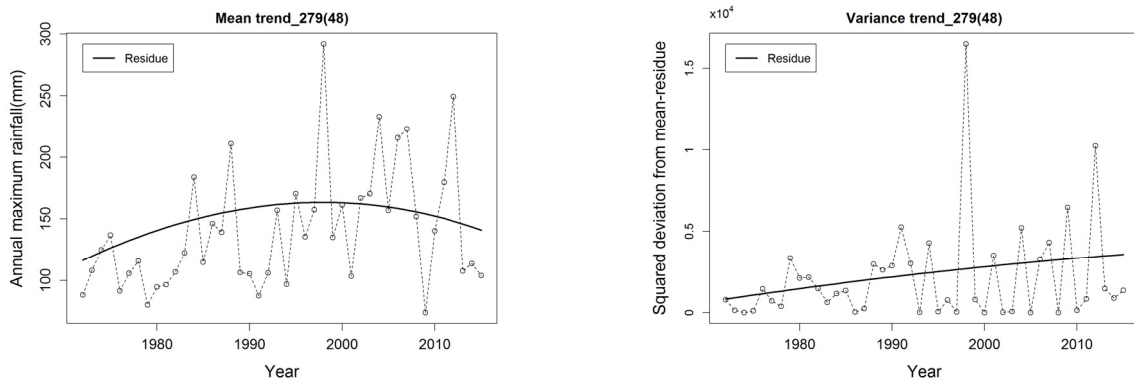


Fig. 4. Time series data and mean- and variance- residues (in case that the EEMD and ML methods have the different results)(Continue)

4. 결론

본 연구에서는 경향성이 존재하는 연 최대치 강우자료를 앙상블 경험적 모드분해법으로 분해한 결과로부터 다양한 비정상성 Gumbel 분포형에 대한 매개변수를 추정하고 최적 분포형을 선정하였으며, 기존의 최우도법을 이용한 비정상성 빈도해석 결과와 비교하여 다음과 같은 결론을 얻었다.

- 1) 앙상블 경험적 모드분해법을 통해 분해된 평균잔여값과 분산잔여값은 각각 자료의 평균과 분산의 경향을 나타내며, 이를 통해 해당 자료의 평균 및 분산에 대한 추세를 시간에 대한 함수로 정의하여 비정상성 확률분포형의 매개변수 추정에 적용한 결과, 자료의 추세를 적절하게 반영할 수 있음을 확인하였다.
- 2) 평균 또는 분산에 1차(선형) 경향성이 존재하는 자료에 대해서 앙상블 경험적 모드분해법을 이용해 비정상성 빈도 해석을 수행한 경우, 위치 매개변수 또는 규모 매개변수를 선형적인 함수로 정의한 비정상성 Gumbel 분포형을 최적 분포형으로 선정하였다. 또한 최우도법을 적용한 경우에도 동일한 비정상성 Gumbel 분포형을 최적 분포형으로 선정하였으므로, 두 방법 모두 1차(선형) 경향성을 잘 반영함을 알 수 있었다.
- 3) 평균에 2차(곡선) 경향성이 존재하는 자료에 대해서 비정상성 빈도해석을 수행한 경우, 최우도법을 통해 선정된 최적 분포형은 앙상블 경험적 모드분해법을 통해 선정된 최적 분포형보다 작은 AIC값을 가짐에도 불구하고 2차(곡선) 경향성을 반영하지 못한 반면, 앙상블 경험적 모드분해법의 경우 이러한 경향성을 반영한 비정상성 Gumbel 분포형을 최적 분포형으로 선정하였다. 따라서 앙상블 경험적 모드분해법을 통해 분해된 자료의 잔여값에 대한 경향

을 확인하여 비정상성 빈도해석을 수행한다면 보다 정확한 자료의 경향성을 반영할 수 있을 것으로 판단된다.

감사의 글

본 연구는 국토교통부 물관리사업의 연구비지원(14AWMP-B082564-01)에 의해 수행되었습니다.

References

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle." *Second International Symposium on Information Theory*, edited by B. N. Petrov and F. Csaki, pp. 267-281, Acad. Kiado', Budapest.

Breaker, L. C., and Ruzmaikim, A. (2011). "The 154-year record of sea level at San Francisco: extracting the long-term trend, recent changes, and other tidbits." *Climate Dynamics*, Vol. 36, No. 3-4, pp. 545-559.

Chen, P. C., Wang, Y. H., You, G. J. Y., and Wei, C. C. (2017). "Comparison of methods for non-stationary hydrologic frequency analysis: Case study using annual maximum daily precipitation in Taiwan." *Journal of Hydrology*, Vol. 545, pp. 197-211.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Di, C., Yang, X., and Wang, X. (2014). "A four-stage hybrid model for hydrologic time series forecasting." *PLoS ONE* Vol. 9, Issue 8, e104663. doi: 10.1371/journal.pone.0104663

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C., and Liu, H. H. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." *Proceedings of the Royal Society of London. Series A*, Vol. 454, pp. 903-995.

Huang, N. E., and Wu, Z. (2008). "A review on Hilbert-Huang

- transform: method and its applications to geophysical studies.” *Review of Geophysics*, Vol. 46, RG2006.
- Intergovernmental Panel on Climate Change (2014). “Climate Change 2014: Synthesis Report.” Contribution of Working Groups I, II, and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri, and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.
- 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.
- Jang, H., Kim, S., and Heo, J. H. (2015). “Comparison study on the various forms of scale parameter for the nonstationary Gumbel model.” *Journal of Korea Water Resources Association*, Vol. 48, No. 5, pp. 331-343.
- Katz, R. W., (2013). “Statistical Methods for Nonstationary Extremes.” In: *Extremes in a Changing Climate*, Edited by AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., and Sorrooshian, S., Chapter 2, Springer, London.
- Kim, Y., and Cho, K. (2016). “Sea level rise around Korea: Analysis of tide gauge station data with the ensemble empirical mode decomposition method.” *Journal of Hydro-environment Research*, Vol. 11, pp. 138-145.
- Kim, T., Shin, H., Nam, W., and Heo, J. H. (2015). “A study on the predictive power improvement of time series model using empirical mode decomposition method.” *Journal of Korea Water Resources Association*, Vol. 48, No. 12, pp. 981-993.
- Korea Meteorological Administration (2014). “Korean Climate Change Assessment Report 2014.”
- Kullback, S., and Leibler, R. A. (1951). “On information and sufficiency.” *Annals of Mathematical Statistics*, Vol. 22, No. 1, pp. 79-86.
- Kwon, H. H., and Lee, J. J. (2011). “Seasonal rainfall outlook of Nakdong river basin using nonstationary frequency analysis model and climate information.” *Journal of Korea Water Resources Association*, Vol. 44, No. 5, pp. 339-350.
- Laio, F., Baldassarre, G. D., and Montanari, A. (2009). “Model selection techniques for the frequency analysis of hydrological extremes.” *Water Resources Research*, Vol. 45, No. 7, W07416.
- Leadbetter, M. R., Lindren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.
- Lee, J. J., Kwon, H. H., and Hwang, K. N. (2010). “Concept of seasonality analysis of hydrologic extreme variables and design rainfall estimation using nonstationary frequency analysis.” *Journal of Korea Water Resources Association*, Vol. 43, No. 8, pp. 733-745.
- Lee, T., Ouarda, T. B. M. J., and Li, J. (2013). “An orchestrated climate song from the Pacific Atlantic Oceans and its implication on climatological process.” *International Journal of Climatology*, Vol. 33, pp. 1015-1020.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., and Stouffer, R. J. (2008). “Stationarity Is Dead: Whither Water Management?” *Science*, Vol. 319, pp. 573-574.
- O’Brien, N. L., and Burn, D. H. (2014). “A nonstationary index-flood technique for estimating extreme quantiles for annual maximum streamflow.” *Journal of Hydrology*, Vol. 519, pp. 2040-2048.
- Olsen, J. R., Stedinger, J. R., Matalas, N. C., and Stakhiv, E. Z. (1999). “Climate variability and flood frequency estimation for the upper Mississippi and lower Missouri Rivers.” *Journal of the American Water Resources Association*, Vol. 36, No. 6, pp. 1509-1523.
- Sang, Y. F., Wang, Z., and Liu, C. (2014). “Comparison of the MK test and EMD method for trend identification in hydrological time series.” *Journal of Hydrology*, Vol. 510, pp. 293-298.
- Shin, H., Ahn, H., and Heo, J. H. (2014). “A study on the changes of return period considering nonstationarity of rainfall data.” *Journal of Korea Water Resources Association*, Vol. 47, No. 5, pp. 447-457.
- Tramblay, Y., Neppel, L., Carreau, J., and Najib, K. (2013). “Non-stationary frequency analysis of heavy rainfall events in southern France.” *Hydrological Sciences Journal*, Vol. 58, No. 2, pp. 280-294.
- Vasiliades, L., Galiatsatou, P., and Loukas, A. (2015). “Nonstationary frequency analysis of annual maximum rainfall using climate covariates.” *Water Resources Management*, Vol. 29, No. 2, pp. 339-358.
- Villarini, G., Smith, J. A., and Napolitano, F. (2010). “Nonstationary modeling of a long record of rainfall and temperature over Rome.” *Advanced in Water Resources*, Vol. 33, No. 10, pp. 1256-1267.
- Villarini, G., Smith, J. A., Serinaldi, F., Bales, J., Bates, P. D., and Krajewski, W. F. (2009). “Flood frequency analysis for non-stationary annual peak records in an urban drainage basin.” *Advanced in Water Resources*, Vol. 32, No. 8, pp. 1255-1266.
- Wang, W. C., Chau, K. W., Xu, D. M., and Chen, X. Y. (2015). “Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition.” *Water Resources Management*, Vol. 29, pp. 2655-2675.
- Wi, S., Valdés, J. B., Steinschneider, S., and Kim, T. W. (2016). “Non-stationary frequency analysis of extreme precipitation in South Korea using peaks-over-threshold and annual maxima.” *Stochastic Environmental Research and Risk Assessment*, Vol. 30, No. 2, pp. 583-606.
- Wu, Z., and Huang, N. E. (2004). “A study of the characteristics of white noise using the empirical mode decomposition method.” *Proceedings of the Royal Society of London. Series A*, Vol. 460, pp. 1597-1611.
- Wu, Z., and Huang, N. E. (2009). “Ensemble empirical mode decomposition: a noise-assisted data analysis method.” *Advances in Adaptive Data Analysis*, Vol. 1, No. 1, pp.1-41.
- Wu, Z., Huang, N. E., Long, S. R., and Peng, C. K. (2007). “On the trend, detrending, and variability of nonlinear and nonstationary time series.” *Proceedings of the National Academy of Sciences*, Vol. 104, pp. 14889-14894.