

Bootstrap 기법을 이용한 서울지점 강우자료의 통계적 동질성 분석

A Statistical Homogeneity Analysis of Seoul Rainfall using Bootstrap

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

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

Abstract

In this study, homogeneity analysis was performed between rainfall observation data set of Chukwooki (CWK) and rainfall observation data set of modern rain gage (MRG) using Bootstrap method. Since traditional statistical homogeneity test method are validated only when distribution of their population is known, meteorological data which their statistical distributions of population are complicated were difficult to verify the homogeneity and there were plenty of room for doubt for their statistical significance using historical method. In this reason, in this study homogeneity test was evaluated between two data sets using bootstrap method which is not necessary to infer distribution of population. The test results show that there was an statistical homogeneity between CWK and MRG except for slight impact of climatical trend.

Keywords : Bootstrap, climate change, Chukwooki, homogeneity

요 지

본 연구에서는 부트스트랩(Bootstrap) 기법을 이용하여 측우기 관측계열(CWK)과 근대우량계 강우량 관측계열(MRG)에 대해 동질성 분석을 실시하였다. 서로 다른 두 자료계열에 대한 전통적인 통계적 동질성 검정 방법은 모집단의 분포형을 알고 있어야 검정결과가 유효하였기 때문에 모집단의 분포가 복잡한 기상자료들은 이러한 전통적 방법을 사용하여 동질성을 파악하는 것이 매우 어려웠고 결과로 제시된 통계적 유의성에 대해서도 의심의 여지가 있었다. 이러한 이유로 본 논문에서는 모집단을 가정하지 않아도 되는 비모수적 모의 방법인 부트스트랩 기법을 이용하여 모집단을 직접 추정한 후 경험누가확률분포를 산정하여 두 자료계열간 통계적 동질성 검정을 실시하였다. 분석 결과 CWK와 MRG는 미소한 기후의 경년변화(trend)의 영향을 제외하면 동질성을 가진 자료로 볼 수 있었다.

핵심용어 : 부트스트랩, 기후변화, 측우기, 동질성

* 정희원, 한국건설기술연구원 수자원연구실 연구원

Researcher, Korea Institute of Construction Technology, GyeongGi-Do 411-712, Korea

** 교신저자, 정희원, 고려대학교 공과대학 건축·사회환경공학부 교수

Corresponding Author, Prof., School of Architecture, Civil & Environmental Eng., Korea University, Seoul, 136-713, Korea (e-mail: jaykim@korea.ac.kr)

*** 정희원, 고려대학교 공과대학 건축·사회환경공학부 교수

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

**** 정희원, 한국건설기술연구원 수자원연구실 책임연구원

Chief Researcher, Korea Institute of Construction Technology, GyeongGi-Do 411-712, Korea

***** 정희원, 고려대학교 공과대학 건축, 사회환경공학부 박사과정

Doctoral Student, School of Architecture, Civil, and Environmental Eng., Korea University, Seoul, 136-713, Korea

1. 서론

장기간의 강수량기록은 기후 변동성 연구를 위한 귀중한 자료이다. 한반도에서는 정량적인 측우사업이 측우기의 발명과 더불어 15세기 중반부터 시작되었다. 이는 세계최장의 기록이나 우리나라 기후변화에 관련한 연구에서 강수량 관련 분석들은 1908년 이후 근대 관측자료가 대부분 이용되었다. 이는 측우기 자료가 현대의 일정한 계측시간 중심의 정량적 자료가 아닌 호우사상 중심의 강우지속기간과 지속기간내 강수량형태로 기록되어, 근대의 강수기록 기준과 동일하지 않아 근대 강수기록으로 변환시 다소의 불확실성이 포함된다는 이유에서이다. 조선시대 강수량까지 포함한 서울의 강수량을 분석한 이병설(1970)은 1770년부터 1907년까지의 138년 동안의 연평균 강수량이 1908년에서 1960년까지의 연평균 강수량과 큰 차이를 보이지만 여름철만을 분석한 경우는 두 기간의 강수량 차이가 작음을 지적하였다. Arakawa (1956)는 1770년-1907년의 측우기 강수량자료와 서울 측후소에서 측정한 1908년-1944년 동안의 강수량 자료를 비교 분석하여 측우기 자료가 상당한 신뢰성을 가진다고 하였다. Lim and Jung (1992)은 서울 지역의 근대우량계 관측 연강수량은 측우기로 관측된 과거 조선시대의 연강수량과 동일한 수준임을 통계 분석을 통해 제시한 바 있다. 그러나 기존의 통계적인 유의수준에서의 측우기 관측 강우량 자료계열과 근대 강우량 관측 자료계열간의 동질성 분석은 해당 강우량 계열의 모집단 분포형에 대한 정확한 추정이 수반되어야만 가설검정에 대한 신뢰도가 높아진다는 단점이 있다. 따라서 본 연구에서는 모집단의 분포형을 가정하지 않는 재표본에 의한 모의 방법인 Bootstrap 기법을 이용하여 측우기 관측 강우량 자료계열과 근대우량계 강우량 관측계열의 평균과 분산에 대한 동질성 분석을 수행하였다.

Pereira *et al.* (1984)은 연유출량 다지점 분해모형의 잔차항을 발생시키기 위해 부트스트랩 기법을 이용하였다. Cover and Unny (1986)는 유출량 ARMA 모형의 매개변수 불확실도를 추정하기 위해 부트스트랩 방법을 이용하였다. Sharma *et al.* (1977)은 장기간의 유출량 자료를 모의발생시키고자 부트스트랩 기법을 이용한 바 있다. Tasker and Dunne (1997)은 가뭄빈도해석을 실시하고자 다지점 PAR(1) 모형에 부트스트랩 기법을 적용하였다. 김병식 등(2002)은 추계학적 모형의 잔차값을 부트스트랩 기법을 이용하여 연 및 월 하천 유출량자료를 모의 복원 추출하였고 이를 Monte Carlo 모형, AR(1), PAR(1)과 같은 추계학적 방법과 비교하였을 때

기존의 방법과 부트스트랩 방법 모두 평균, 표준편차, 자기상관성은 잘 재현하였으나 왜곡도 계수의 경우는 부트스트랩이 기존의 방법보다 더 잘 재현한다고 하였다. 성기원(2003)은 모수적 부트스트랩 방법을 이용하여 단위도의 신뢰구간을 산정한 바 있다. 유철상 등(2007)은 부트스트랩 기법을 이용하여 유역 및 기상상태를 고려한 Clark 단위도의 매개변수 신뢰도 평가를 실시하여 집중시간 및 저류상수의 신뢰구간의 폭이 기존에 비해 좁다고 결론지었고 신뢰도 있는 집중시간 및 저류상수의 추정을 위해서는 최소 20개 이상의 독립된 호우사상이 필요하다고 하였다. 이러한 연구동향을 살펴보면 부트스트랩은 주로 실측자료에 근거한 장기간의 모의자료를 확보하기 위해 기존의 추계학적 방법의 대안으로 사용되었거나, 모형의 매개변수의 불확실도(혹은 신뢰구간)를 추정하는 연구에 주로 사용되어 왔음을 알 수 있다.

본 연구에서는 부트스트랩 방법을 이용하여 측우기 자료계열과 근대우량계 자료계열의 월별 동질성을 분석하였다. 이를 위해 부트스트랩 방법의 적용을 위해 필요한 통계적 독립성을 자기상관함수 분석, 연검정, 전환점 검정, 1차 자기상관계수 검정과 같은 기존의 통계적 방법을 이용하여 검정하였고, 이를 통해 측우기 자료계열과 근대우량계 자료계열의 통계적 독립성을 확인한 후 부트스트랩 방법을 이용하여 관측자료로부터 모집단을 추정하고 ECDF (경험누가확률분포)를 산정하여 실측자료에 기반한 두 자료계열의 정량적 동질성을 분석하여 보았다.

2. 분석자료

2.1 월별 동질성 분석을 위한 통계치의 설정

다음 Table 1은 측우기 관측계열과 근대우량계 관측계열간의 월별 동질성 분석을 하기 위해 선정된 4개의 통계치로 강우변화 특성을 효과적으로 나타내 줄 수 있는 특성들이다. 강우는 강우량과 같은 양적 특성과 발생횟수와 같은 빈도 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. Table 1의 *Monthly rainfall*은 4월에서 10월까지 각 월의 월강우량이고 D_{max} ratio는 월강우량에 대한 해당월 일최대 강우량이 차지하는 비율의 변화를 분석하기 위한 통계치이다. $N_{rainy\ days}$ 는 각 월별 평균적인 강우일수를 산정한 통계치이고 $I_{rainy\ days}$ 는 각 월별 강우일수에

Table 1. Statistics which are used to test the homogeneity between CWK and MRG

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

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 단기 강우사상의 규모 및 특성 변화를 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.

2.2 각 통계치 자료계열의 통계적 독립성 검정

부트스트랩을 사용하여 자료계열의 통계적 신뢰도를 추정하기 위해서는 기본적으로 자료계열의 독립성이 보장되어야 한다. 전형적인 부트스트랩은 자료계열의 재배치를 통해 새로운 자료계열을 재생성하는 방법이 기 때문에 자료계열에 시간적 상관성이 존재한다면 재생성 과정에서 이 시간적 상관성을 고려하기 어려워 적용성에 문제가 생긴다. 따라서 본 논문에서는 부트스트랩을 이용하여 두 자료계열간의 동질성을 검정하기 위해 앞에서 제시한 4개의 통계치 자료계열별로 각각의 시간적 독립성을 통계적으로 검정하여 보았다. 자료계열이 독립성을 가진다는 의미는 자료계열에 연속된 시간적 상관성이 존재하지 않는다는 의미로 해석할 수 있다. 따라서 가장 쉽게 사용하는 방법이 연속된 시간축 위에서 상관도를 나타낼 수 있는 자기상관함수(ACF, Autocorrelation function)를 이용하여 시차별 자기상관계수의 경향성을 도식적으로 분석하는 것이다. 일반적으로 시간축에 대해 통계적 자기상관을 가지는 계열의 경우는 시차에 따라 연속성을 가지고 증감하는 경향을 보이고, 또한 주기성이 있는 경우는 이 증감양상이 일정 시차를 두고 반복하게 된다. 시간축에 연속적으로 ACF를 도식하여 시계열의 독립성을 분석하는

방법은 직관적이면서도 객관적이기 때문에 많이 사용된다. 그리고 도식적인 방법외에 시계열 자료의 독립성을 판단하기 위한 유의수준을 설정하여 가설을 검정하는 검정통계량 방법들이 많이 연구되어 왔는데 그 대표적인 방법으로 Table 2와 같은 연검정(run test), 전환점 검정(turning point test), 1차 자기상관계수 검정(lag-1 autocorrelation test)등이 있다. 시계열 자료의 독립성은 무작위성과 매우 밀접한 관련을 가지고 있는데(동일한 의미로도 사용) 무작위성을 가진 자료는 독립성이 있다고 볼 수 있다. 더불어 추세를 가지고 있는 계열도 통계적 독립성이 유지되기 어렵기 때문에, 이 방법들은 대부분 시계열 자료의 무작위성 검정이나 추세 검정을 통해 자료의 독립성 여부를 통계적으로 판단하는 방법들이다. 일반적으로 자기상관함수의 유용한 추정값을 얻기 위해서는 적어도 50개의 관측값이 필요하며, 추정된 자기상관함수의 시차는 관측값 개수의 1/4보다 크지 않아야 한다. 본 논문에서 적용한 각 방법별 검정을 위한 귀무가설(H_0)과 검정통계량은 다음 Table 2와 같다.

2.3 각 통계치 자료계열의 통계적 독립성 검정 결과

본 연구에서 제시한 4개의 강우 통계치 자료계열별로 월별 자기상관함수를 도식하여 시간적 상관여부를 분석하여 보았으나 특별한 상관성을 보이지는 않았다. 그리고 연검정, 전환점 검정, 1차 자기상관계수 검정의 세 가지 검정통계량을 이용하여 독립성을 검정한 결과는 Tables 3~5와 같다. 독립성 검정은 크게 측우기 자료계열과 근대우량계 자료계열을 전부 포함하는 전체자료계열(1778년-2006년), 측우기 자료계열(1778년-1907년) 그리고 근대우량계 자료계열(1908년-2006년)의 세 가지 시간자료계열로 구분하여 시행하였다. 그리고 측우기 관측자료계열의 관측최소 정밀도를 고려하여 근대우량계 관측자료계열에서 2 mm보다 작은 강우기록

Table 2. Independence hypothesis test methods using test statistics

Test method		Null hypothesis	Test statistics
Name	Type		
Run test	Non-parametric	No trend(μ) and Independent(ϵ_t)	$Z = \frac{ R - \mu_R }{\sigma_R}$, rejection: $ Z > Z_{\alpha/2}$, α : significance level R is the number of runs, $\mu_R = m + 1, \sigma_R = \left(\frac{m(m-1)}{2m-1} \right)^{1/2}$, m is the number of data above(or below) median
Turning points test	Non-parametric	No trend and Independent	$Z = \left \frac{U - \mu_U}{\sigma_U} \right $, rejection: $ Z > Z_{\alpha/2}$, α : significance level U is the number of turning points, $\mu_U = \frac{2(m-2)}{3}, \sigma_U = \left(\frac{16m-29}{90} \right)^{1/2}$, m is the number of data
Lag-1 autocorrelation test	Parametric	Independent	$Z = \frac{ r_1 - \mu_{r_1} }{\sigma_{r_1}}$, rejection: $ Z > Z_{\alpha/2}$, α : significance level r_1 is lag-1 autocorrelation coefficient, $\mu_{r_1} = -1/m$, $\sigma_{r_1} = [(m^3 - 3m^2 + 4) / \{m^2(m^2 - 1)\}]^{1/2}$, m is the number of data

을 제외한 M20과 원자료계열을 그대로 사용한 M00으로 나누어 비교하였다.

전체자료계열의 경우, 95 % 신뢰수준에서 연검정 결과 *Monthly rainfall* 과 $D_{max} ratio$ 는 모두 경향성이 없고 독립성이 있다고 판정되었다. $N_{rainy days}$ 는 M00의 8월과 M20의 6월과 8월만이 경향성이 없고 독립성이 있다고 판정되었다. $I_{rainy days}$ 는 M00의 6월만이 경향성이 있거나 자기상관에 유의성이 있다고 판정되었다. 전환점 검정 결과 *Monthly rainfall*의 M00과 M20의 9월만이 경향성이나 자기상관에 유의성이 있다고 판정되었다. $D_{max} ratio$ 는 M00과 M20 모두 경향성이 없고 독립성이 있다고 판정되었다. $N_{rainy days}$ 는 M00과 M20의 8월만이 경향성이 없고 독립성이 있다고 판정되었다. $I_{rainy days}$ 는 M00과 M20 모두 경향성이 없고 독립성이 있다고 판정되었다. 1차 자기상관계수 검정 결과 *Monthly rainfall* 과 $D_{max} ratio$ 는 M00과 M20 모두 독립성이 있다고 판정되었다. $N_{rainy days}$ 는 M00의 4월, 5월, 6월, 7월, 9월, 10월에서 자기상관에 유의성이 있다고 판정되었다. $I_{rainy days}$ 는 M00과 M20의 8월에서만 자기상관에 유의성을 보였다.

축우기 자료계열의 경우 근대우량계 자료계열이 포함되지 않았으므로 M00과 M20이 동일하기 때문에 M00과 M20의 검정결과는 같다, 95 % 신뢰수준에서 연

검정 결과 *Monthly rainfall* 과 $D_{max} ratio$ 는 모두 경향성이 없고 독립성이 있다고 판정되었다. $N_{rainy days}$ 는 M00과 M20 모두 6월과 8월만이 경향성이 없고 독립성이 있다고 판정되었다. $I_{rainy days}$ 는 M00과 M20 모두 경향성이 없고 독립성이 있다고 판정되었다. 전환점 검정 결과 *Monthly rainfall* 과 $D_{max} ratio$ 은 모두 경향성이 없고 독립성이 있다고 판정되었다. $N_{rainy days}$ 는 M00과 M20의 7월, 9월, 10월에서 경향성이 있거나 자기상관에 유의성이 있다고 판정되었다. $I_{rainy days}$ 는 M00과 M20 모두 경향성이 없고 독립성이 있다고 판정되었다. 1차 자기상관계수 검정 결과 *Monthly rainfall*, $D_{max} ratio$, $N_{rainy days}$ 그리고 $I_{rainy days}$ 모두 독립성이 있다고 판정되었다.

근대우량계 자료계열의 경우, 95 % 신뢰수준에서 연검정 결과 *Monthly rainfall*은 M00의 10월과 M20의 5월만이 경향성이 있거나 자기상관에 유의성이 있다고 판정되었다. $D_{max} ratio$ 는 M20의 5월만이 경향성이 있거나 자기상관에 유의성이 있다고 판정되었다. $N_{rainy days}$ 는 M00의 6월, 7월, 8월과 M20의 7월, 8월이 경향성이 없고 독립성이 있다고 판정되었다. $I_{rainy days}$ 는 M00과 M20 모두 경향성이 없고 독립성이 있다고 판정되었다. 전환점 검정 결과 *Monthly rainfall*은 M00과 M20의 9월에서 경향성이 있거나 자기상관에 유의성이

Table 3. Results of statistical independence hypothesis test for each data sets (A.D. 1778 to A.D. 2006)

Statistics	Data sets	Month	RT		TP		AC		$Z_{\alpha/2}$ $\alpha = 5\%$	
			Z	H_0	Z	H_0	Z	H_0		
<i>Monthly rainfall</i>	M00	APR	0.468	A	1.640	A	0.209	A	1.960	
		MAY	1.670	A	0.423	A	0.027	A	1.960	
		JUN	1.002	A	1.693	A	0.459	A	1.960	
		JUL	0.200	A	1.005	A	0.602	A	1.960	
		AUG	0.601	A	0.688	A	0.515	A	1.960	
		SEP	1.537	A	2.646	R	0.683	A	1.960	
	M20	APR	1.804	A	0.053	A	0.875	A	1.960	
		MAY	0.200	A	1.640	A	0.322	A	1.960	
		JUN	1.670	A	0.423	A	0.017	A	1.960	
		JUL	1.002	A	1.376	A	0.436	A	1.960	
		AUG	0.334	A	1.005	A	0.620	A	1.960	
		SEP	0.601	A	1.005	A	0.523	A	1.960	
	<i>D_{max} ratio</i>	M00	APR	1.136	A	2.646	R	0.646	A	1.960
			MAY	1.670	A	0.053	A	0.884	A	1.960
JUN			1.537	A	0.741	A	0.548	A	1.960	
JUL			0.468	A	0.423	A	0.959	A	1.960	
AUG			0.334	A	0.688	A	0.275	A	1.960	
SEP			0.468	A	0.370	A	0.334	A	1.960	
M20		APR	0.200	A	0.053	A	1.056	A	1.960	
		MAY	1.403	A	0.370	A	0.048	A	1.960	
		JUN	0.735	A	0.053	A	0.545	A	1.960	
		JUL	1.670	A	1.376	A	0.382	A	1.960	
		AUG	1.537	A	0.212	A	0.679	A	1.960	
		SEP	0.869	A	0.370	A	0.369	A	1.960	
<i>N_{rainy days}</i>		M00	JUL	0.601	A	0.370	A	0.335	A	1.960
			AUG	0.200	A	0.053	A	1.019	A	1.960
	SEP		0.869	A	0.370	A	0.126	A	1.960	
	OCT		0.468	A	0.370	A	0.870	A	1.960	
	APR		4.076	R	2.011	R	3.310	R	1.960	
	MAY		6.615	R	3.440	R	4.909	R	1.960	
	M20	JUN	4.610	R	2.963	R	2.917	R	1.960	
		JUL	5.278	R	2.805	R	3.336	R	1.960	
		AUG	1.670	A	0.900	A	0.175	A	1.960	
		SEP	6.080	R	4.233	R	3.338	R	1.960	
		OCT	5.011	R	3.440	R	2.585	R	1.960	
		APR	4.076	R	3.916	R	0.070	A	1.960	
	<i>I_{rainy days}</i>	M00	MAY	5.679	R	4.233	R	1.719	A	1.960
			JUN	1.804	A	3.122	R	0.770	A	1.960
JUL			4.209	R	2.170	R	0.728	A	1.960	
AUG			1.136	A	0.900	A	1.887	A	1.960	
SEP			4.610	R	4.868	R	1.054	A	1.960	
OCT			4.878	R	3.916	R	0.071	A	1.960	
M20		APR	0.468	A	0.688	A	0.164	A	1.960	
		MAY	1.670	A	0.741	A	0.209	A	1.960	
		JUN	2.873	R	1.005	A	0.767	A	1.960	
		JUL	0.601	A	1.640	A	0.512	A	1.960	
		AUG	0.735	A	0.741	A	2.259	R	1.960	
		SEP	1.136	A	0.265	A	0.227	A	1.960	
M20		OCT	0.200	A	0.688	A	0.283	A	1.960	
		APR	1.270	A	1.058	A	1.683	A	1.960	
	MAY	0.735	A	0.582	A	0.972	A	1.960		
	JUN	0.067	A	0.265	A	0.164	A	1.960		
	JUL	1.002	A	0.847	A	1.099	A	1.960		
	AUG	0.735	A	1.058	A	2.162	R	1.960		
M20	SEP	1.002	A	1.005	A	0.435	A	1.960		
	OCT	0.200	A	0.106	A	0.094	A	1.960		

Note that Z : calculated Z-value of PDF, H_0 : Hypothesis test result at 95 % confidence level, RT: Run test, TP: Turning points test, AC: Lag-1 autocorrelation test

Table 4. Results of statistical independence hypothesis test for each data sets of CWK

Statistics	Data sets	Month	RT		TP		AC		$Z_{\alpha/2}$
			Z	H_0	Z	H_0	Z	H_0	$\alpha = 5\%$
<i>Monthly rainfall</i>	M00 M20	APR	0.442	A	0.978	A	0.622	A	1.960
		MAY	0.265	A	0.349	A	0.073	A	1.960
		JUN	0.442	A	1.117	A	0.743	A	1.960
		JUL	0.088	A	1.606	A	0.311	A	1.960
		AUG	0.796	A	0.070	A	0.907	A	1.960
		SEP	0.088	A	1.536	A	0.628	A	1.960
		OCT	0.442	A	1.117	A	0.134	A	1.960
$D_{max} ratio$	M00 M20	APR	0.972	A	0.489	A	0.861	A	1.960
		MAY	0.619	A	0.698	A	1.243	A	1.960
		JUN	0.796	A	1.397	A	0.648	A	1.960
		JUL	0.265	A	0.489	A	0.213	A	1.960
		AUG	0.972	A	0.070	A	0.970	A	1.960
		SEP	0.265	A	0.559	A	0.241	A	1.960
		OCT	0.088	A	0.279	A	0.348	A	1.960
$N_{rainy days}$	M00 M20	APR	3.094	R	1.746	A	0.388	A	1.960
		MAY	2.563	R	1.536	A	1.221	A	1.960
		JUN	0.088	A	1.955	A	1.607	A	1.960
		JUL	3.978	R	3.003	R	1.117	A	1.960
		AUG	1.679	A	0.908	A	1.503	A	1.960
		SEP	3.978	R	4.259	R	0.757	A	1.960
		OCT	3.801	R	4.259	R	1.311	A	1.960
$I_{rainy days}$	M00 M20	APR	1.326	A	1.117	A	0.250	A	1.960
		MAY	1.149	A	0.768	A	1.387	A	1.960
		JUN	0.442	A	0.559	A	0.445	A	1.960
		JUL	0.619	A	0.349	A	0.718	A	1.960
		AUG	0.796	A	0.698	A	1.113	A	1.960
		SEP	0.972	A	0.349	A	0.245	A	1.960
		OCT	1.149	A	0.279	A	0.019	A	1.960

Note that Z : calculated Z-value of PDF, H_0 : Hypothesis test result at 95 % confidence level, RT: Run test, TP: Turning points test, AC: Lag-1 autocorrelation test

있다고 판정되었다. $D_{max} ratio$ 는 모두 경향성이 없고 독립성이 있다고 판정되었다. $N_{rainy days}$ 는 M00의 5월, 6월과 M20의 4월, 5월, 6월, 9월이 경향성이 있거나 자기상관에 유의성이 있다고 판정되었다. $I_{rainy days}$ 는 M00의 4월과 7월에서 경향성이 있거나 자기상관에 유의성이 있다고 판정되었다. 1차 자기상관계수 검정 결과 *Monthly rainfall*과 $D_{max} ratio$ 는 모두 독립성이 있다고 판정되었다. $N_{rainy days}$ 는 M20의 10월만이 자기상관에 유의성을 보였다. $I_{rainy days}$ 는 M00과 M20의 4월에서 자기상관에 유의성이 있다고 판정되었다.

연검정, 전환점 검정, 1차 자기상관계수 검정 결과를 전체적으로 종합해 보면, *Monthly rainfall*, $D_{max} ratio$ 및 $I_{rainy days}$ 는 경향성이 없고 독립성이 있다고 판단되었다. $N_{rainy days}$ 는 연검정과 전환점 검정에서 전체적으로 경향성이나 자기상관에 유의성을 보이고 있는 반면 1차 자기상관계수 검정에서는 전체자료계열의 M00을 제외하면 전체적으로 자기상관에 유의성을 보이지 않고

있으므로 $N_{rainy days}$ 는 자기상관성보다는 경향성이 크다고 판단된다. 단, 전체자료계열의 M00에서 1차 자기상관계수에 유의성을 보이고 있으나 다른 계열에서는 유의성을 보이지 않고 있어 자료계열에 일관된 경향성은 아닌 것으로 판단되고, 이는 잔차계열의 자기상관성이라기 보다는 평균의 추세성에서 기인한 것으로 보인다. 이는 연검정 결과와 전환점 결과에서 경향성이 있는 것으로 판정된 것도 하나의 근거이다. 즉, 측우기 자료계열에 비해 근대우량계 자료계열의 관측 정밀도가 더 높기 때문에 측우기 자료계열에 비해 상대적으로 근대우량계 자료계열의 $N_{rainy days}$ 는 평균적으로 증가한 경향을 보이고 있고 이로 인한 평균의 증가추세가 발생하여 자기상관성이 나타나는 것으로 추정된다. 그러나 그 통계적 유의성은 크지 않은 것으로 판단된다. 본 분석을 통해 측우기 관측자료계열과 근대우량계 관측자료계열은 통계적인 독립성이 있다고 보여지며, 따라서 부트스트랩 기법을 적용하기 위한 통계적 독립성 전제조건을 만족한다고 판단된다.

Table 5. Results of statistical independence hypothesis test for each data sets of MRG

Statistics	Data sets	Month	RT		TP		AC		$Z_{\alpha/2}$
			Z	H_0	Z	H_0	Z	H_0	$\alpha = 5\%$
<i>Monthly rainfall</i>	M00	APR	0.309	A	1.228	A	0.002	A	1.960
		MAY	1.754	A	0.737	A	0.076	A	1.960
		JUN	1.135	A	1.228	A	1.232	A	1.960
		JUL	0.103	A	0.246	A	0.522	A	1.960
		AUG	0.516	A	1.228	A	0.072	A	1.960
		SEP	0.928	A	2.211	R	0.017	A	1.960
		OCT	2.166	R	1.474	A	1.683	A	1.960
	M20	APR	0.722	A	1.228	A	0.027	A	1.960
		MAY	2.166	R	0.737	A	0.145	A	1.960
		JUN	1.135	A	0.737	A	1.229	A	1.960
		JUL	0.103	A	0.246	A	0.522	A	1.960
		AUG	0.516	A	1.720	A	0.084	A	1.960
		SEP	0.722	A	2.211	R	0.012	A	1.960
		OCT	1.547	A	1.474	A	1.654	A	1.960
<i>D_{max} ratio</i>	M00	APR	1.547	A	0.491	A	0.032	A	1.960
		MAY	1.547	A	0.000	A	0.590	A	1.960
		JUN	0.103	A	0.737	A	1.542	A	1.960
		JUL	0.722	A	0.983	A	1.164	A	1.960
		AUG	0.516	A	0.246	A	0.283	A	1.960
		SEP	1.135	A	0.000	A	0.485	A	1.960
		OCT	0.928	A	0.246	A	0.835	A	1.960
	M20	APR	1.547	A	1.474	A	0.567	A	1.960
		MAY	2.166	R	0.983	A	0.694	A	1.960
		JUN	0.309	A	1.228	A	1.508	A	1.960
		JUL	0.309	A	0.983	A	1.116	A	1.960
		AUG	0.103	A	0.246	A	0.243	A	1.960
		SEP	1.135	A	0.000	A	0.400	A	1.960
		OCT	0.516	A	0.737	A	0.935	A	1.960
<i>$N_{rainy\ days}$</i>	M00	APR	3.198	R	1.228	A	0.202	A	1.960
		MAY	3.816	R	3.440	R	1.117	A	1.960
		JUN	1.341	A	1.965	R	0.452	A	1.960
		JUL	0.516	A	0.491	A	0.422	A	1.960
		AUG	0.928	A	0.000	A	1.277	A	1.960
		SEP	2.372	R	1.474	A	0.029	A	1.960
		OCT	2.166	R	0.491	A	1.885	A	1.960
	M20	APR	2.785	R	4.177	R	0.522	A	1.960
		MAY	4.229	R	4.668	R	1.206	A	1.960
		JUN	2.785	R	2.457	R	1.240	A	1.960
		JUL	1.960	A	0.246	A	0.614	A	1.960
		AUG	0.309	A	0.000	A	1.389	A	1.960
		SEP	2.372	R	2.457	R	0.786	A	1.960
		OCT	3.198	R	1.228	A	2.151	R	1.960
<i>$I_{rainy\ days}$</i>	M00	APR	0.722	A	2.211	R	2.551	R	1.960
		MAY	1.547	A	1.720	A	0.720	A	1.960
		JUN	0.309	A	0.983	A	0.573	A	1.960
		JUL	1.754	A	2.211	R	1.094	A	1.960
		AUG	0.309	A	0.000	A	1.680	A	1.960
		SEP	0.722	A	0.983	A	0.608	A	1.960
		OCT	0.516	A	1.474	A	1.126	A	1.960
	M20	APR	0.103	A	0.491	A	2.218	R	1.960
		MAY	0.309	A	1.720	A	0.550	A	1.960
		JUN	0.309	A	0.737	A	0.220	A	1.960
		JUL	0.722	A	0.983	A	1.253	A	1.960
		AUG	0.722	A	0.737	A	1.355	A	1.960
		SEP	0.309	A	0.983	A	0.814	A	1.960
		OCT	1.341	A	0.246	A	0.168	A	1.960

Note that Z : calculated Z-value of PDF, H_0 : Hypothesis test result at 95 % confidence level, RT: Run test, TP: Turning points test, AC: Lag-1 autocorrelation test

3. 부트스트랩 기본이론

3.1 부트스트랩(Bootstrap) 기본 원리

일반적으로 강우를 포함한 기상인자의 정량화를 위하여 모수적인 방법을 사용하고 있으나 이를 극복할 필요가 있다. 이 모수적 방법은 기존의 통계적인 방법으로 익히 알려진 확률밀도함수에 특정 기상인자의 특성이 완벽하게 부합한다는 가정에 근거한다. 다시 말해 표준오차의 신뢰한계를 명확히 산정할 수 있어야 한다. 그러나 기상요소와 같이 변동성이 크거나 불규칙한 장기적인 거동특성의 경우에 그러한 가정이 적합하지 않을 가능성이 매우 크다. 그리고 기존의 방법으로 통계적 유의성을 검정할 때 모집단의 분포를 모르기 때문에 모집단을 정규분포로 가정하고 표본의 크기가 작은 경우 t -분포를 사용한다. 이는 모분산 대신 표본분산을 사용하는 방법으로 정규분포에 비해 t -분포는 더 납작한데 이는 모분산대신 표본분산을 사용하여 더 큰 불확실성을 갖게 되어 꼬리가 긴 분포를 만들기 때문이다. 그러나 여기서 t -분포 역시 표본의 크기에 대한 정규분포인 모집단과의 관계를 나타낼 뿐 실제로 모집단이 정규분포인지에 대한 의문은 여전히 남아있다. 따라서 관측자료가 가지고 있는 특성을 그대로 보전하여 재표본 방법에 의해 직접 분포형을 추정하는 자료모의 방법이 제시되었고, 이러한 재표본 방법 중 가장 간단하고 보편적인 방법으로 Efron (1979)의 부트스트랩(Bootstrap)기법이 있다. Efron (1979)에 의해 제안된 부트스트랩기법은 주어진 표본에 근거하여 재표본(resampling)을 취하여 연구대상이 되는 통계량의 성질을 파악한다. 부트스트랩기법은 기존의 방법들과 달리 주어진 검정통계량의 분포를 가정하지 않고, 주어진 검정통계량의 분포를 직접 근사한다. 또한 부트스트랩기법은 근본적으로 모형을 설정하지 않는 자료에 기반한 모의방법이기 때문에 모형에 의존하지 않는 이점을 가지고 있다(전명식, 1996). 따라서 표준오차의 신뢰한계를 구할 수 없는 경우, 그 표본을 마치 모집단인 것처럼 다루는 재표본추출 방법인 부트스트랩 분석의 기본 원리 및 절차는 Fig. 1과 같이 매우 간단하다. 원래의 표본에서 재표본을 추출한 후 재표본에 대한 추정치를 산정하고 이 두 단계를 여러 번 반복하여 재표본 추정치의 산포를 파악하는 것이다. 표준 비모수적 부트스트랩의 기본 절차는, 주어진 확률표본 x_1, x_2, \dots, x_n 으로부터 복원 추출 방법으로 부트스트랩 표본 $x_1^*, x_2^*, \dots, x_n^*$ 을 얻는다. 이때 무

작위 추출이므로 추출 가능한 부트스트랩 표본의 총수는 n^n 임을 알 수 있다. 다음으로 이 n^n 개(대부분 이보다 작음)의 부트스트랩 표본에 대해 구하고자 하는 통계량 $\hat{\theta}^*$ 를 얻을 수 있으며 이를 통해 추정 통계량 $\hat{\theta}^*$ 의 부트스트랩 분포를 추정하게 된다. Efron and Tibshirani (1986)은 정확한 신뢰구간 추정값을 계산하기 위해서는 최소한 1,000개의 부트스트랩 표본이 적절하다고 제안하였다. 따라서 B 개(본 연구에서는 1000개)의 부트스트랩 표본으로부터 통계량 $\hat{\theta}^*$ 을 구하고 오름차순으로 정렬하여 θ 의 추정량($\hat{\theta}$)과 순번이 정해진 부트스트랩 추정값($\hat{\theta}_i^*$)을 산정한다. 이러한 부트스트랩 추정값을 이용하여 θ 에 대한 $100(1-\alpha)\%$ 부트스트랩 신뢰구간을 구할 수 있다. B 개의 부트스트랩 추정값 $\hat{\theta}_i^*$ 로부터 표본 평균과 표본 분산은 다음과 같다.

$$\bar{\theta}^* = \frac{1}{B} \sum_{i=1}^B \hat{\theta}_i^* \quad (1)$$

$$v_{\theta}^{*2} = \frac{1}{B-1} \sum_{i=1}^B (\hat{\theta}_i^* - \bar{\theta}^*)^2 \quad (2)$$

여기서 v_{θ}^* 은 $\hat{\theta}$ 표준편차의 추정량이고 그 분포가 근사적으로 정규분포이면 θ 에 대한 $100(1-\alpha)\%$ 신뢰구간은 다음과 같다.

$$\hat{\theta} \pm z_{\alpha/2} v_{\theta}^* \quad (3)$$

여기서 z_{α} 는 표준정규분포의 상위 $100 \times \alpha\%$ 백분위수이다.

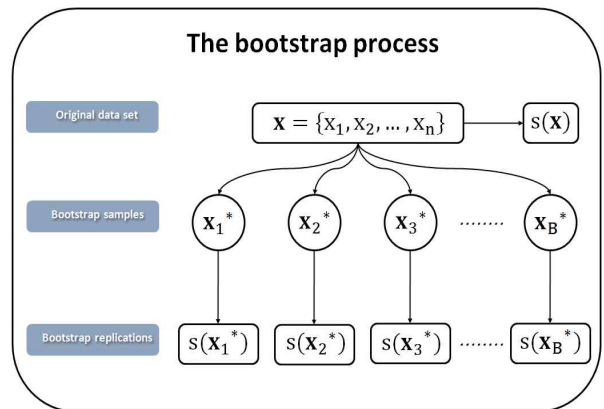


Fig. 1. Basic ideas of Bootstrap and its conceptualized process

3.2 평균과 분산변화에 대한 부트스트랩 유의성 검정

앞서 언급한 바와 같이 측우기의 관측최소 단위가 1 분(1分=2 mm)이므로 0.1 mm단위로 측정하는 근대 우량계 관측과 비교하면 소량의 강수는 측정되지 않았을 가능성이 높다(정현숙과 임규호, 1994) 라고 보고 있다. 상식적으로 보아도 당시의 관측수준이 현대보다 높다고 보긴 힘들다. 그러나 이러한 연구결과는 미소한 수준의 관측 정확도가 높고 낮음에 대한 언급이지, 본 논문의 분석 초점인 통계학적으로 유의성을 가질 만큼 정량적인 차이가 있고 없음에 대한 판단 결과는 아니다. 즉, 호우사상의 크기나 빈도 등의 변화량에 따라 판단하는 경년변동, 기후변화 등을 파악하는 목적에서는 미소한 수준의 정량적인 차이가 중요하긴 하지만, 여름철을 기준으로, 분석하고자 하는 목적에 비추어 비교대상간의 차이에 유의성이 없다면 두 자료를 활용하는 데는 큰 문제가 되지 않을 것이다. 본 논문에서는 부트스트랩 방법을 이용하여 실측자료에 근거한 측우기 자료와 근대 우량계 자료의 *Monthly rainfall*, D_{max} ratio, $N_{rainy\ days}$, $I_{rainy\ days}$ 에 대한 평균과 분산의 차이에 대한 유의성 검정을 실시하였다. 측우기 자료의 경우 강설이 고려되지 않았기 때문에 4월에서 10월까지의 연단위 강수량 자료와 각 월별 자료에 대한 검정을 실시하였다. 더불어 이러한 미소수준의 차이를 줄여 측우기 관측단위와 동일한 조건에서 유의성 검정을 실시하기 위하여 근대 우량계 자료중 2 mm이하의 경우는 무강우로 간주하고 동일한 유의성 검정을 실시하였다. 이는 두 가지 목적에서 실시되었는데 첫째는 측우기 자

료의 정확도를 동일한 수준에서 가늠해 보는 것이며, 두 번째는 연단위 자료의 특성이 유사한 경우 월별 강우의 경년변동(trend) 특성을 파악해 보고자 하는 목적이다. 평균에 대한 귀무가설(null hypothesis)은 측우기 자료와 근대 우량계 자료의 “평균은 같다”이고, 대립가설의 경우는 측우기 자료와 근대 우량계 자료는 “평균이 같지 않다”라고 정의하였다. 즉, 양측검정에 해당한다. Table 6의 STEP 1에서 분산에 대한 귀무가설도 측우기 자료와 근대 우량계 자료의 “분산은 같다”이고, 대립가설의 경우는 측우기 자료가 근대 우량계 자료와 “분산이 같지 않다”라고 정의하였고, 이경우도 양측검정에 해당한다. 다음 Table 6은 Bootstrap 방법을 이용하여 측우기 관측계열의 강우량자료와 근대우량계 관측계열의 4가지 검정통계량 별 평균과 분산의 변화에 대한 유의성 검정을 위한 초기 가설이다. STEP 2에서 4가지 통계특성 중 *Monthly rainfall*에 대해 별도로 단측검정을 실시한 이유는 일반적으로 관측의 정확도(특히 최소 관측 단위) 면에서 측우기 관측 강우량 자료가 근대우량계 관측 강우량 자료보다 낮다고 보여지며 최근 평균적으로 강우량이 증가하고 있다는 연구결과를 토대로 측우기 강우량 평균이 근대우량계 강우량 평균보다 작을 것으로 보아 단측검정을 실시하였다. 그러나 *Monthly rainfall*의 변동량(분산)은 평균과는 달리 개별 발생호우 사상의 크기에 좌우되므로 근대우량계보다 “작다”라고 단정지을 수 없기 때문에 양측검정만을 실시하였다. 그리고 부트스트랩을 이용하여 자료를 재생성하기 위해 비모수적인 표준 부트스트랩 방법을 이용하였고 각 자료계열별로 표본추출회수 B 는 1,000회를 실시하였다.

Table 6. Initial hypotheses for Differences of means and variances between CWK and MRG using bootstrap method

Step	Hypothesis test for same distribution at 95 % CL		
	Test statistics	Hypothesis	Formula
Step 1: two-side test	Difference of mean	H_0	$T = \mu_1 - \mu_2 = 0$
		H_1	$T = \mu_1 - \mu_2 \neq 0$
	Difference of variance	H_0	$T = v_1 - v_2 = 0$
		H_1	$T = v_1 - v_2 \neq 0$
Step 2: one-side test	Difference of mean (for <i>monthly rainfall</i> only)	H_0	$T = \mu_1 - \mu_2 = 0$
		H_1	$T = \mu_1 - \mu_2 < 0$

Note that μ_1 and v_1 are the mean and variance of CWK statistics, μ_2 and v_2 are the mean and variance of MRG statistics

Table 7. Results of hypothesis test using bootstrap methods (two-sides test)

Statistics	Groups	Month	Mean					Variance				
			p	α	t_{c_L}	t_{c_U}	H_0	p	α	t_{c_L}	t_{c_U}	H_0
<i>Monthly rainfall</i>	M00	APR	0.022	0.05	-28.23	-1.40	R	0.112	0.05	-3269	194	A
		MAY	0.146	0.05	-24.13	3.95	A	0.768	0.05	-1613	1979	A
		JUN	0.312	0.05	-42.08	11.56	A	0.300	0.05	-10072	2460	A
		JUL	0.176	0.05	-93.03	19.34	A	0.804	0.05	-29605	34035	A
		AUG	0.344	0.05	-70.94	23.41	A	0.396	0.05	-36766	10103	A
		SEP	0.062	0.05	-53.43	1.95	A	0.996	0.05	-7266	9952	A
	M20	OCT	0.442	0.05	-14.48	5.87	A	0.456	0.05	-669	1222	A
		APR	0.080	0.05	-26.01	1.40	A	0.112	0.05	-3150	216	A
		MAY	0.284	0.05	-22.27	7.52	A	0.734	0.05	-1540	2105	A
		JUN	0.436	0.05	-37.53	16.30	A	0.364	0.05	-9613	2690	A
		JUL	0.206	0.05	-92.42	19.33	A	0.766	0.05	-28047	34849	A
		AUG	0.394	0.05	-76.75	25.03	A	0.346	0.05	-35830	9915	A
<i>D_{max} ratio</i>	M00	SEP	0.086	0.05	-51.07	4.31	A	0.998	0.05	-8182	10299	A
		OCT	0.730	0.05	-12.07	8.41	A	0.516	0.05	-620	1245	A
		APR	0.024	0.05	0.007	0.108	R	0.122	0.05	-0.003	0.025	A
		MAY	0.038	0.05	0.002	0.082	R	0.018	0.05	0.002	0.023	R
		JUN	0.888	0.05	-0.039	0.043	A	0.506	0.05	-0.007	0.015	A
		JUL	0.964	0.05	-0.027	0.029	A	0.002	0.05	0.003	0.014	R
	M20	AUG	0.762	0.05	-0.031	0.036	A	0.080	0.05	-0.001	0.015	A
		SEP	0.830	0.05	-0.055	0.043	A	0.020	0.05	0.003	0.033	R
		OCT	0.116	0.05	-0.007	0.100	A	0.002	0.05	0.011	0.045	R
		APR	0.326	0.05	-0.026	0.081	A	0.536	0.05	-0.011	0.020	A
		MAY	0.160	0.05	-0.012	0.071	A	0.162	0.05	-0.004	0.020	A
		JUN	0.590	0.05	-0.053	0.029	A	0.838	0.05	-0.011	0.013	A
<i>N_{rainy days}</i>	M00	JUL	0.850	0.05	-0.032	0.025	A	0.002	0.05	0.003	0.013	R
		AUG	0.964	0.05	-0.038	0.031	A	0.246	0.05	-0.003	0.014	A
		SEP	0.264	0.05	-0.082	0.022	A	0.332	0.05	-0.009	0.026	A
		OCT	0.702	0.05	-0.052	0.074	A	0.262	0.05	-0.009	0.031	A
		APR	0.000	0.05	-3.76	-2.32	R	0.112	0.05	-5.56	0.38	A
		MAY	0.000	0.05	-3.93	-2.47	R	0.126	0.05	-5.32	0.57	A
	M20	JUN	0.000	0.05	-4.40	-2.77	R	0.712	0.05	-3.74	2.62	A
		JUL	0.000	0.05	-5.45	-3.22	R	0.008	0.05	2.34	15.20	R
		AUG	0.000	0.05	-4.09	-1.99	R	0.702	0.05	-4.50	6.90	A
		SEP	0.000	0.05	-3.96	-2.38	R	0.940	0.05	-3.83	3.63	A
		OCT	0.000	0.05	-3.33	-1.96	R	0.124	0.05	-4.65	0.46	A
		APR	0.932	0.05	-0.73	0.61	A	0.902	0.05	-2.39	2.27	A
<i>I_{rainy days}</i>	M00	MAY	0.340	0.05	-0.32	0.93	A	0.102	0.05	-0.29	3.51	A
		JUN	0.688	0.05	-0.61	0.91	A	0.106	0.05	-0.52	5.13	A
		JUL	0.632	0.05	-0.82	1.28	A	0.002	0.05	3.96	16.96	R
		AUG	0.080	0.05	-0.12	1.80	A	0.040	0.05	0.22	10.12	R
		SEP	0.984	0.05	-0.69	0.76	A	0.066	0.05	-0.17	6.28	A
		OCT	0.820	0.05	-0.50	0.62	A	0.084	0.05	-0.37	3.62	A
	M20	APR	0.000	0.05	2.09	4.93	R	0.002	0.05	9.06	40.89	R
		MAY	0.000	0.05	2.27	5.57	R	0.008	0.05	6.63	53.01	R
		JUN	0.002	0.05	2.51	7.27	R	0.008	0.05	13.81	97.10	R
		JUL	0.002	0.05	1.62	6.87	R	0.104	0.05	-9.45	99.09	A
		AUG	0.014	0.05	0.66	6.31	R	0.782	0.05	-68.25	44.37	A
		SEP	0.036	0.05	0.13	5.92	R	0.184	0.05	-33.34	142.70	A
M20	OCT	0.004	0.05	1.23	4.84	R	0.012	0.05	6.85	71.05	R	
	APR	0.118	0.05	-3.66	0.25	A	0.602	0.05	-52.39	20.50	A	
	MAY	0.028	0.05	-4.91	-0.28	R	0.120	0.05	-81.77	6.82	A	
	JUN	0.170	0.05	-4.34	0.97	A	0.284	0.05	-18.50	69.77	A	
	JUL	0.014	0.05	-6.99	-1.09	R	0.882	0.05	-48.74	61.86	A	
	AUG	0.004	0.05	-7.94	-1.33	R	0.012	0.05	-191.27	-13.79	R	
SEP	0.012	0.05	-8.14	-0.72	R	0.422	0.05	-257.83	72.99	A		
OCT	0.384	0.05	-2.85	1.04	A	0.330	0.05	-23.48	57.47	A		

4. 분석결과

본 논문에서는 측우기와 근대우량계 자료계열간의 평균과 분산의 차이에 대한 양측검정과 평균의 차이에 대한 단측검정을 위해 부트스트랩 방법을 이용하여 관측자료로부터 모집단을 추정하고 ECDF(경험누가확률분포)를 산정하여 각각 95 % 신뢰수준에서 유의성을 검토하였다. Table 7과 Table 8에서 α 는 유의수준이고 t_c , t_{c_L} , t_{c_U} 는 신뢰한계로 t_{c_L} 과 t_{c_U} 은 양측검정에서 신뢰구간의 하한계와 상한계를 나타낸다. 측우기와 근대우량계 자료계열간의 변동특성을 파악하기 위해 제시한 4가지 통계특성인 *Monthly rainfall*, D_{max} ratio, $N_{rainy\ days}$, $I_{rainy\ days}$ 에 대한 월별 평균과 분산의 차이에 대한 표준 부트스트랩 유의성 검정 결과는 다음 Table 7과 같다. 자료의 분석대상 기간은 4월에서 10월까지로 월별로 검정을 실시하였다.

*Monthly rainfall*의 경우 M00은 4월을 제외하면 평균 차이에 대한 유의성은 보이지 않았고 분산은 분석대상 기간 전체에서 차이에 대한 유의성을 보이지 않았다. M20은 분석대상기간 전체에서 평균과 분산의 차이에 대한 유의성을 보이지 않았다. 이를 통해 볼 때 두계열간의 강우의 양적 차이는 크지 않으며 특히 2 mm 이하의 강우가 차지하는 비중이 크지 않음을 알 수 있다.

D_{max} ratio의 경우 M00은 4월과 5월이 평균의 차이에 유의성을 보였고 5월, 7월, 9월, 10월에서 분산의 차이에 유의성을 보였다. 평균의 차이는 강우량이 적은 봄철에서 두드러 졌다. M20은 평균의 차이에는 전기간에서 유의성을 보이지 않았고 분산의 차이는 7월에서만 유의성을 보였다. 이를 통해 강우량이 가장 많은 7월의 단기호우 변동폭은 과거와 비교해 근대이후에 변화가 있었음을 추정할 수 있다.

$N_{rainy\ days}$ 의 경우 M00은 전체기간에서 평균의 차이에 유의성을 보였고 분산은 7월에서만 유의성을 보였다. 이를 통해 볼 때 두자료계열간의 강우일수는 확연한 차이가 있음을 알 수 있고 이는 강우량 변화가 크지 않음을 고려해 볼 때 관측 정밀도가 가장 큰 요인일 것으로 판단된다. M20은 평균의 차이에는 전기간에서 유의성을 보이지 않았고 분산의 차이는 7월과 8월에서만 유의성을 보였다. 특히 7월의 p 값이 매우 작아 연별 강우일수의 변동폭에 변화가 있음을 알 수 있다.

$I_{rainy\ days}$ 의 경우 M00은 전기간에서 평균의 차이에 유의성을 보였고 분산은 4월, 5월, 6월과 10월에서 유의성을 보였다. 특히 분산은 집중호우가 자주 발생하는 7월, 8월, 9월을 제외한 기간에서 차이에 대한 유의성을 보이고 있는데 이는 강우량에 대한 민감도가 우기에 비해 상대적으로 같수기인 4월, 5월, 6월, 10월이 크기 때문으로 판단된다. M20은 5월, 7월, 8월, 9월의 평균 차이에서 유의성을 보였고 분산은 8월의 경우만 그 차이에 유의성을 보였다. 이를 통해 볼 때 여름철 강우강도는 과거에 비해 근대이후 변화가 있음을 알 수 있다.

지금까지의 각 검정대상 통계치의 양측검정 결과는 증감의 여부는 알 수 있었으나 증가인지 감소인지에 대한 정보는 주지 않았다. 따라서 월강우량 평균의 차이에 대한 95 % 신뢰수준에서 단측검정(대립가설: “측우기 자료의 평균이 근대우량계 자료보다 작다”) 결과는 다음 Table 8과 같다.

M00은 4월과 9월에서만 *Monthly rainfall* 평균의 차이에 유의성을 보여 4월과 9월은 근대우량계 자료계열의 *Monthly rainfall* 평균이 측우기 자료계열의 *Monthly rainfall* 평균보다 통계적으로 증가했다고 볼 수 있었다. 그 외의 기간에서는 유의성을 보이지 않았으나 5월과 7월의 p 값도 0.084와 0.083으로 상대적으로 작다는 점은 주목할 필요가 있다. M20은 4월에서만

Table 8. Results of hypothesis test using bootstrap methods (one-side test)

Data sets		M00				M20			
		p	α	t_c	H_0	p	α	t_c	H_0
Annual	APR-OCT	0.003	0.05	-50.82	R	0.013	0.05	-32.79	R
<i>Monthly rainfall</i>	APR	0.018	0.05	-2.97	R	0.040	0.05	-0.99	R
	MAY	0.084	0.05	2.57	A	0.147	0.05	5.14	A
	JUN	0.165	0.05	8.07	A	0.231	0.05	10.68	A
	JUL	0.083	0.05	6.83	A	0.113	0.05	10.40	A
	AUG	0.175	0.05	18.47	A	0.209	0.05	21.19	A
	SEP	0.044	0.05	-1.02	R	0.059	0.05	1.45	A
	OCT	0.243	0.05	5.23	A	0.375	0.05	7.32	A

Monthly rainfall 평균의 차이에 유의성을 보였다. 그러나 9월은 p 값이 기각수준(5 % 이내)에는 포함되지 않으나 0.059로 매우 작아 강우량의 증가 경향이 있을 가능성이 매우 높다.

5. 결 론

본 장에서는 부트스트랩 기법을 이용하여 측우기 강우량 관측계열(CWK)과 근대우량계 강우량 관측계열(MRG) 사이의 동질성을 분석해 보았다. 부트스트랩을 이용하여 두 집단의 동질성 분석을 한 결과를 정리하면 다음과 같다.

- 1) CWK와 MRG 자료계열의 독립성 검정을 위하여, 검정통계량을 이용하여 CWK와 MRG의 각 통계치 자료계열에 대한 독립성 검정 결과, M20의 경우 CWK와 MRG는 통계적인 독립성이 있다고 판단된다. 단, $N_{rainy\ days}$ 의 경우 연검정과 전환점 검정에서 유의성을 보이고 있고, 1차 자기상관계수 검정에서는 유의성을 보이지 않아 자기상관성 보다는 추세가 있다고 판단된다.
- 2) 부트스트랩 기법을 이용하여 95 % 신뢰수준에서 CWK와 MRG간의 동질성 검정결과(양측검정), M20의 경우 *Monthly rainfall*은 4월에서 9월 모두 평균과 분산의 차이에 유의성을 보이지 않았다. $D_{max\ ratio}$ 는, 모든 달에서 평균의 차이에 유의성을 보이지 않았고, 분산의 차이는 7월만이 유의성을 보였다. $N_{rainy\ days}$ 는 모든 달에서 평균의 차이에 유의성을 보이지 않았고, 분산의 차이는 7월과 8월만이 유의성을 보였다. $I_{rainy\ days}$ 는 5월, 7월, 8월, 9월이 평균의 차이에 유의성을 보였고, 분산은 8월의 경우만 차이에 유의성을 보였다.
- 3) 부트스트랩 기법을 이용하여 95 % 신뢰수준에서 *Monthly rainfall*의 평균의 증감을 검정하기 위해 평균의 증가에 대한 단측검정 결과, M20의 경우, 4월은 p 값이 0.040으로 CWK에 비해 MRG의 평균에 증가경향이 있는 것으로 판정되었다. 9월은 p 값이 0.059로 기각수준에 근접하여 통계적으로 MRG 평균의 증가경향이 있는 것으로 추정된다.

여기서, 양측검정과 단측검정의 결과는 모두 95 % 신뢰수준을 기준으로 판단하여 실제로 양측검정에 비해 단측검정의 기각수준이 넓다. 그러나 본 분석은 통상적인 기각수준에 의한 확정적인 검정 결과를 도출하고자 하는 목적이 아니고 확률적인 가능성의 상대적인 정도

를 가능해 보고자 하였으므로 결과도 이러한 측면에서 해석하였다.

본 연구를 통하여 *Monthly rainfall*의 경우는 상대적으로 갈수기의 변화가 두드러지고 있으며, $D_{max\ ratio}$ 나 $N_{rainy\ days}$ 의 경우는 7월(호우기)의 강우량의 분산에 변화를 보이고 있다. 이는 갈수기의 경우는 월강우량의 크기에 변화가 있으며, 호우기의 경우는 일강우량의 크기 및 호우의 형태에 변화가 있음을 나타낸다고 볼 수 있다. 이는 측우기 자료자체의 신뢰도 문제일 수 있으나 기존 연구 및 본 연구에서 나타난 월별 일관성을 볼 때 단편적이지만 기후변화 양상을 나타내는 것으로 판단된다.

참 고 문 헌

- 김병식, 김형수, 서병하 (2002). "Bootstrap 방법에 의한 하천유출량 모의와 왜곡도." **한국수자원학회 논문집**, 한국수자원학회, 제35권, 제3호, pp. 275-284.
- 김연형 (2002). 시계열 예측, 형설출판사, pp. 41-45.
- 성기원 (2003). "유역의 수문학적 상사성을 이용한 Nash 모형의 불확실성 평가." **한국수자원학회 논문집**, 한국수자원학회, 제36권, 제3호, pp. 399-411.
- 유철상, 이지호, 김기욱 (2007). "유역 및 기상상태를 고려한 Clark 단위도의 매개변수 평가: 2. 매개변수의 변동성 추정." **한국수자원학회논문집**, 한국수자원학회, 제40권, 제2호, pp. 171-182.
- 이병설 (1970). "서울의 연강수량 및 하기강수량의 Normality에 관한 연구." **한국기상학회지**, 한국기상학회, 제5권, pp. 11-14.
- 전명식 (1996). "부트스트랩방법의 실제적 활용-군집표본추출법에 근거한 분할표분석을 중심으로." **한국통계학회논문집**, 한국통계학회, 제3권, 제1호, pp. 179-188.
- 정현숙, 임규호 (1994). "서울 지역 월강수량 강수일의 관계, 1770-1907." **한국기상학회지**, 한국기상학회, 제30권, 제4호, pp. 487-505.
- Arakawa, H. (1956). "On the secular variation of annual total of rainfall at Seoul from 1770 to 1944." *Archives for Meteorology, Geophysics, and Bioclimatology*, Vol. 7, No. 2, pp. 205-211.
- Cover, K.A., and Unny, T.E. (1986). "Application of computer intensive statistics to parameter

- uncertainty in streamflow synthesis." *Water Resources Research Bull.*, Vol. 22, No. 3, pp. 495-507.
- Efron, B. (1979). "Bootstrap method: another look at the jackknife." *Journal of Statistics*, Vol. 7, pp. 1-26.
- Efron, B., and Tibshirani, R.J. (1986). "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy." *Statistical Science*, Vol. 1, pp. 54-77.
- Lim, G.H., and Jung, H.S. (1992). "Interannual variation of the annual precipitations at Seoul, 1771-1990." *Journal of the Korean Meteorological Society*, Vol. 28, pp. 125-132.
- Pereira, M.V.F., Oliveria, G.C., Costa, C.G., Kelman, J. (1984). "Stochastic stream flow models for hydroelectric system." *Water Resources Research* Vol. 20, No. 3, pp. 379-390.
- Sharma, A., Tarboton, D.G., and Lall, U. (1997). "Streamflow simulation: a non-parametric approach." *Water Resources Research*, Vol. 33, No. 2, pp. 291-308.
- Tasker, G.D., and Dunne, P. (1997). "Bootstrap Position Analysis for Forecasting Low Flow Frequency." *Journal of Water Resources Planning and Management*, Vol. 123, No. 6, pp. 359-367.

논문번호: 09-037	접수: 2009.04.01
수정일자: 2009.07.10/08.24	심사완료: 2009.08.24