

서울지점 강우자료의 정량적 동질성 분석

The Quantative Homogeneity Analysis of Seoul Rainfall

황석환* · 김종훈** · 유철상*** · 유도근****

Hwang, Seok Hwan · Kim, Joong Hoon · Yoo, Chulsang · Yoo, Doguen

Abstract

In this study, quantitative homogeneity analysis was performed between rainfall observation data set of Chukwooki(CWK) and rainfall observation data set of modern rain gage(MRG) using statistical methods such as basic statistics, K-S test and Boxplots. To analyze the homogeneities of CWK and MRG four rainfall characteristic series such as monthly rainfall, the ratio of maximum daily rainfall to monthly rainfall, number of rainy days for each month, and the ratio of monthly rainfall to numbers of rainy days are made, and the homogeneity tests using two sample K-S test and quantitative comparisons were performed. The test results showed that observation precisions between CWK and MRG of original data set(M00) were differed because M00 clearly showed the statistical significances on differences of numbers of monthly rainy days of CWK and MRG. But, rainfall showed a little differences which were not significant between CWK and MRG.

Key words : Homogeneity analysis, Rainfall characteristics, Chukwooki

요 지

본 연구에서는 기본 통계치 비교, K-S 검정 및 상자그림과 같은 통계적 기법을 이용하여 측우기 강우량 관측계열(CWK)과 근대우량계 강우량 관측계열(MRG)에 대해 정량적 동질성 분석을 실시하였다. 측우기 관측계열과 근대우량계 관측계열간의 월별 동질성을 분석하기 위해 월강우량, 월강우일에 대한 해당월 최대 일강우량의 비, 월강우일수, 일강우강도의 4개 강우특성자료 계열을 산정하였고, 이표본 K-S 검정을 통한 분포에 대한 동질성 검정과 상자그림을 이용한 정량적 비교를 수행하였다. 분석 결과 각 분석과정에서 M00은 전체적으로 CWK와 MRG의 월강우일수간 차이에 명확한 통계적 유의성을 보이고 있어 CWK와 MRG 간의 관측정밀도에 차이가 있다고 판단된다. 그러나, CWK와 MRG의 강우량은 상대적으로 유의성이 크지 않은 미소한 차이를 보이고 있는 것으로 나타났다.

핵심용어 : 동질성 분석, 강우특성, 측우기

1. 서 론

장기간의 강우량기록은 기후 변동성 연구를 위한 귀중한 자료이다. 한반도에서는 정량적인 측우사업이 측우기의 발명과 더불어 15세기 중반부터 시작되었다. 이는 세계최장의 기록이나 측우기 자료가 현대의 일정한 계측시간 중심의 정량적 자료가 아닌 호우사상 중심의 강우지속기간과 지속기간내 강우량형태로 기록되어 근대의 강우기록 기준과 다소 상이하기 때문에 우리나라 기후변화에 관련한 연구에서 강수량 관련 분석들은 1908년 이후 근대 관측자료가 대부분 이용되었다. 측우기 강우량과 관련된 기존 주요 연구성과를 요약하면 다음과 같다. 조선시대 강우량까지 포함한 서울의 강우량을

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 1770-1907년의 측우기 강수량 자료와 서울 측후소에서 측정된 1908-1944년 동안의 강수량 자료를 비교 분석하여 측우기 자료가 상당한 신뢰성을 가진다고 하였다. Lim과 Jung(1992)은 서울 지역의 근대우량계 관측 연강수량은 측우기로 관측된 과거 조선시대의 연강수량과 동일한 수준임을 통계분석을 통해 제시한 바 있다. 측우기 자료와 근대 강우량 자료는 연속 시계열자료로서 장기간의 시간적 차이를 보이기 때문에 경년변동을 고려하기 힘든 평균이나 표준편차와

* 한국건설기술연구원 수자원연구실 연구원(E-mail: sukany@kict.re.kr)
** 정회원 · 고려대학교 공과대학 건축 · 사회환경공학부 교수(교신저자)
*** 정회원 · 고려대학교 공과대학 건축 · 사회환경공학부 교수
**** 고려대학교 공과대학 건축 · 사회환경공학부 박사과정

같은 정량적인 기본통계특성의 크기 비교만으로는 정확히 동질성을 판단하는 것이 매우 힘들다. 이렇듯 정량적인 통계치에 의한 동질성 검정방법이 연속적인 경년변화(trend)를 고려하지 못하는 한계를 가지고 있지만, 본 분석의 목적은 상대적으로 관측기록에 불신을 받고 있는 측우기 관측 강우량자료에 대한 정량적 통계치의 비교를 통해 신뢰수준을 가늠해 보는데 있기 때문에 이러한 분석을 통하여 측우기 관측 강우량자료의 기초 통계량에 대한 신뢰도가 확보된다면 보다 다양한 통계분석이나 추계학적 분석을 수행하는 데 보다 도움이 될 것이다.

2. 서울지점 월강수량 및 일강수량 자료의 특성

2.1 기본 통계특성

본 연구에서 사용한 측우기 관측자료는 조희구와 나일성(1979)과 전종갑과 문병권(1997)의 연구결과로부터 얻어졌다

(김기욱 등에서 일부 재인용, 2007). 측우기의 관측 단위는 관측최소 단위가 1분(1分=2 mm)이므로 0.1 mm 단위로 측정하는 근대 우량계 관측과 비교하면 소량의 강수는 측정되지 않았을 가능성이 크다(정현숙 등, 1994). 그러나 이것이 겨울철을 제외한 4월에서 10월까지의 강우량에 미치는 영향은 상대적으로 크지 않을 것으로 판단된다. 이러한 이유에서 본 연구에서는 4월에서 10월까지의 자료만을 분석하였고, 그 중 강우량이 상대적으로 많은 6월에서 9월까지의 결과에 대해서만 비교하여 보았다. 그리고 측우기 최소 관측단위의 영향도 고려해 보고자 원자료계열을 그대로 사용한 그룹(M00)과 측우기 관측정도와 유사하게 근대 우량계 자료 중 2 mm 이하 자료를 제외한 그룹(M20)으로 나누었고, 각각을 관측 종류(CWK와 MRG)와 기후변화가 예상되는 시점(BCC와 ACC)을 기준으로 나누었다. CWK(1778년-1907년)는 측우기 관측 강우량계열이고 MRG(1908년-2006년)는 근대우량계 관측 강우량계열이다. 그리고 BCC(1778년-1960년)는 기후변화가 예

Table 1. Basic statistics of observed daily rainfall data (M00)

Groups	Data sets year	Statistics	Month						
			APR	MAY	JUN	JUL	AUG	SEP	OCT
All	TOT 1778-2006	MEAN	2.4	2.8	4.5	11.5	8.8	4.3	1.5
		MAX	115.1	130.3	219.9	392.0	354.7	281.2	110.5
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	63.1	83.0	199.3	704.6	531.9	228.4	34.8
		SKEW	5.7	5.5	5.5	4.3	5.3	7.1	7.5
		KURT	44.5	39.8	42.1	28.3	43.9	76.7	79.8
G1	CWK 1778-1907	MEAN	2.2	2.6	4.3	11.0	8.4	4.0	1.4
		MAX	115.1	130.3	157.5	392.0	345.4	281.2	110.5
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	55.9	77.9	187.6	651.5	464.5	189.9	36.2
		SKEW	5.8	5.5	5.3	4.5	5.1	7.2	8.0
		KURT	49.0	41.4	35.5	33.7	43.2	82.4	93.6
	MRG 1908-2006	MEAN	2.6	3.0	4.8	12.2	9.2	4.8	1.5
		MAX	104.2	122.3	219.9	294.6	354.7	268.2	79.4
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	72.9	90.0	215.3	776.9	624.0	280.7	33.0
		SKEW	5.4	5.4	5.8	4.1	5.3	6.9	6.6
		KURT	39.4	37.8	48.5	22.7	42.7	68.2	56.9
G2	BCC 1778-1960	MEAN	2.2	2.7	4.5	11.3	8.2	4.1	1.4
		MAX	115.1	130.3	219.9	392.0	354.7	281.2	110.5
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	56.9	74.6	200.1	679.9	472.9	198.6	33.6
		SKEW	5.7	5.3	5.4	4.3	5.4	7.1	7.7
		KURT	46.2	38.1	40.1	28.9	48.0	76.6	86.3
	ACC 1961-2006	MEAN	2.9	3.2	4.5	12.2	10.8	5.3	1.6
		MAX	104.2	122.3	194.6	294.6	332.8	268.2	79.4
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	87.2	116.0	196.4	800.7	756.7	343.1	39.3
		SKEW	5.3	5.5	5.9	4.3	4.7	6.8	6.9
		KURT	37.0	38.1	50.3	26.1	32.5	65.3	60.0

Note that

MEAN: mean, MAX: maximum, MIN: minimum, VAR: Variance, SKEW: skewness, KURT: kurtosis, G1: data group 1, G2: data group 2, TOT: data group of 1778-2006, CWK: data group of 1778-1907, MRG: data group of 1908-2006, BCC: data group of 1778-1960, ACC: data group of 1961-2006

상되는 시점 이전 강우량 계열이고 ACC(1961년-2006년)는 기후변화가 예상되는 시점 이후 강우량계열이다.

표 1에서 보듯이 M00에 대하여 각 자료계열의 월별 일강우량 평균은, 6월의 경우 4.3 mm에서 4.8 mm의 값을 보이고 있으며 CWK가 4.3 mm로 가장 작았고 MRG가 4.8 mm로 가장 컸다. 특히, ACC의 월별 일강우량 평균이 4.5 mm이므로 평균이 최대인 시기는 1908년에서 1960년 기간이다. 7월의 경우는 11.0 mm에서 12.2 mm의 값을 보였으며 CWK가 11.0 mm로 가장 작았고 MRG와 ACC가 12.2 mm로 가장 큰 값을 보였다. 8월의 경우는 8.2 mm에서 10.8 mm의 값을 보였으며 BCC가 8.2 mm로 가장 작았고 ACC가 10.8 mm로 가장 커서 1960년 전후로 확연한 강우 특성의 변화를 보이고 있다. 9월의 경우는 4.0 mm에서 5.3 mm으로 CWK가 4.0 mm로 가장 작았고, ACC가 5.3 mm로 가장 컸다. 전체적으로 여름철 강우 자료의 평균값은 CWK가 가장 작았고, ACC가 가장 컸다.

M00에 대하여 각 자료계열의 월별 일강우량 분산은, 6월의 경우 187.6 mm²에서 215.3 mm²의 값을 보였으며, CWK가 187.6 mm²로 가장 작았고, MRG가 215.3 mm²로 가장 컸다. 특히, ACC의 월별 일강우량 분산이 196.4 mm²로 MRG보다 작아 분산이 최대인 시기는 평균과 마찬가지로 1908년에서 1960년 기간이었다. 7월의 경우 651.5 mm²에서 800.7 mm²의 값을 보였으며, CWK가 651.5 mm²로 가장 작았고, ACC가 800.7 mm²로 가장 컸다. 8월의 경우 464.5 mm²에서 756.7 mm²의 값을 보였으며, CWK가 464.5 mm²로 가장 작았고, ACC가 756.7 mm²로 가장 컸다. 9월의 경우 189.9 mm²에서 343.1 mm²의 값을 보였으며, CWK가 189.9 mm²로 가장 작았고, ACC가 343.1 mm²로 가장 컸다. 전체적으로 여름철 일강우 자료의 분산은 평균과 마찬가지로 CWK가 가장 작았고, ACC가 가장 컸다.

M00의 월별 일강우량 최대치는 6월과 8월의 경우는 각각 219.9 mm와 354.7 mm로 1908년에서 1960년 사이(MRG와

Table 2. Basic statistics of observed daily rainfall data (M20)

Groups	Data sets	Statistics	Month						
			APR	MAY	JUN	JUL	AUG	SEP	OCT
All	TOT 1778-2006	MEAN	2.4	2.7	4.5	11.5	8.7	4.3	1.5
		MAX	115.1	130.3	219.9	392.0	354.7	281.2	110.5
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	63.2	83.2	199.6	705.6	532.5	228.6	34.9
		SKEW	5.7	5.5	5.5	4.3	5.3	7.1	7.5
		KURT	44.4	39.8	42.0	28.2	43.9	76.6	79.7
G1	CWK 1778-1907	MEAN	2.2	2.6	4.3	11.0	8.4	4.0	1.4
		MAX	115.1	130.3	157.5	392.0	345.4	281.2	110.5
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	55.9	77.9	187.6	651.5	464.5	189.9	36.2
		SKEW	5.8	5.5	5.3	4.5	5.1	7.2	8.0
		KURT	49.0	41.4	35.5	33.7	43.2	82.4	93.6
	MRG 1908-2006	MEAN	2.6	2.9	4.7	12.1	9.1	4.8	1.5
		MAX	104.2	122.3	219.9	294.6	354.7	268.2	79.4
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	73.2	90.4	216.0	779.3	625.5	281.2	33.1
		SKEW	5.4	5.4	5.8	4.1	5.3	6.9	6.6
		KURT	39.3	37.7	48.3	22.6	42.6	68.0	56.7
G2	BCC 1778-1960	MEAN	2.2	2.6	4.5	11.3	8.2	4.1	1.4
		MAX	115.1	130.3	219.9	392.0	354.7	281.2	110.5
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	56.9	74.6	200.3	680.5	473.3	198.8	33.7
		SKEW	5.7	5.3	5.4	4.3	5.4	7.1	7.7
		KURT	46.2	38.0	40.0	28.8	47.9	76.6	86.3
	ACC 1961-2006	MEAN	2.8	3.1	4.4	12.1	10.7	5.2	1.6
		MAX	104.2	122.3	194.6	294.6	332.8	268.2	79.4
		MIN	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		VAR	87.5	116.3	197.1	803.4	758.3	343.7	39.5
		SKEW	5.3	5.5	5.9	4.3	4.7	6.8	6.9
		KURT	36.8	38.0	50.1	26.0	32.4	65.1	59.9

Note that

MEAN: mean, MAX: maximum, MIN: minimum, VAR: Variance, SKEW: skewness, KURT: kurtosis

BCC가 일치하는 기간)에서 발생하였고, 7월, 9월의 경우는 각각 392.0 mm, 281.2 mm로 1778년에서 1907년 사이(CWK와 BCC가 일치하는 기간으로 결국 CWK)에서 발생하였다.

표 2는 측우기의 관측정도와 유사하게 근대 우량계 관측기록에서 2 mm이하를 제외하고 분석한 경우(M20)이다. 각 자료계열의 월별 일강우량 평균은 6월의 경우 4.3 mm에서 4.7 mm의 값을 보였으며, CWK가 4.3 mm로 가장 작았고, MRG가 4.7 mm로 가장 컸다. ACC의 경우는 4.4 mm로 상대적으로 작은 값을 보였다. 7월의 경우는 11.0 mm에서 12.1 mm의 값을 보였으며 CWK가 11.0 mm로 가장 작았고, MRG와 ACC가 12.1 mm로 가장 컸다. 8월의 경우는 8.2 mm에서 10.7 mm의 값을 보였으며, BCC가 8.2 mm로 가장 작았고 ACC가 10.7 mm로 가장 커서 1960년 전후로 확연한 강우 특성의 변화를 보였다. 9월의 경우는 4.0 mm에서 5.2 mm으로 CWK가 4.0 mm로 가장 작았고, ACC가 5.2 mm로 가장 컸다. 전체적으로 여름철 일강우자료의 평균 값은 CWK가 가장 작았고, ACC가 가장 컸다. 여기서 주목할 점은 월별 일강우량 평균에 있어 2 mm 이하 일강우량의 제외 여부는 큰 차이를 보이지 않아 2 mm 이하 일강우의 정량적인 영향은 크지 않음을 알 수 있었다. 그리고 월별 특성을 보면 전체적으로 CWK가 가장 작은 평균값을 나타냈다. 이는 기후변화가 원인일 수도 있으나 일부 관측 자료의 정확도 문제일 여지도 있다. 월별 일강우량의 평균이 가장 큰 기간은, 6월은 1908년에서 1960년 이었고 8월, 9월은 1961년에서 2006년이었다. 단, 7월은 MRG와 ACC값이 같아 평균에 있어 1908년 이후 큰 변화가 없는 것으로 나타났다. 따라서 대략 1960년 이후 8월, 9월의 일강우량이 증가한 것으로 보인다.

M20의 경우 각 자료계열의 월별 일강우량 분산은 6월의 경우 187.6 mm²에서 216.0 mm²의 값을 보였으며 CWK가 187.6 mm²로 가장 작았고 MRG가 216.0 mm²으로 가장 컸다. 특히, ACC가 197.1 mm²로 MRG보다 작기 때문에 분산이 최대인 시기는 1908년에서 1960년 기간으로 판단된다. 7월의 경우 651.5 mm²에서 803.4 mm²의 값을 보였으며 CWK가 651.5 mm²로 가장 작았고 ACC가 803.4 mm²로 가장 컸다. 8월의 경우 464.5 mm²에서 758.3 mm²의 값을 보였으며 CWK가 464.5 mm²로 가장 작았고 ACC가 758.3 mm²로 가장 컸다. 9월의 경우 189.9 mm²에서 343.7 mm²의 값을 보였으며 CWK가 189.9 mm²로 가장 작았고 ACC가 343.7 mm²로 가장 컸다. 전체적으로 여름철 일강우량의 분산은 평균과 마찬가지로 CWK가 가장 작고, ACC가 가장 컸다. 결과를 종합하면 월별 일강우량 자료의 분산도 2 mm이하의 제외여부에 따라 큰 차이를 보이지 않고 있어 2 mm이하 일강우의 정량적인 영향이 크지 않음을 확인할 수 있었다. 그리고 월별 특성은 전체적으로 CWK의 분산이 가장 작았다. 월별 일강우량의 분산이 가장 큰 기간은 6월의 경우 1908년에서 1960년 이었고, 7월, 8월, 9월의 경우는 1961년에서 2006년이었다. 따라서 대략 1960년 이후 7월, 8월, 9월의 일강우량 변동폭이 증가한 것으로 보인다.

M20의 경우 최대치는 6월과 8월의 경우는 각각 219.9 mm

Table 3. 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	$D_{max\ ratio} = \frac{Maximum\ daily\ rainfall}{Monthly\ rainfall}$
Number of rainy days for each month	$N_{rainy\ days} = Numbers\ of\ rainy\ days$
The ratio of monthly rainfall to numbers of rainy days	$I_{rainy\ days} = \frac{Monthly\ rainfall}{Numbers\ of\ rainy\ days}$

와 354.7 mm로 1908년에서 1960년 사이(MRG와 BCC가 일치하는 기간)에서 발생하였고, 7월, 9월의 경우에 각각 392.0 mm, 281.2 mm로, 1778년에서 1907년 사이(CWK와 BCC가 일치하는 기간으로 결국 CWK)에서 발생하였다.

3. 월별 동질성 분석을 위한 통계치의 설정

다음 표 3은 측우기 관측계열과 근대우량계 관측계열간의 월별 동질성 분석을 하기 위해 선정한 4개의 통계치로 강우 변화 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. 표 3의 *Monthly rainfall*은 4월에서 10월까지 각 월의 월강우량이고 $D_{max\ ratio}$ 는 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 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.

4. 정량적 동질성 분석 기법

4.1 CWK와 MRG의 강우특성별 통계적 분포에 대한 동질성 검증

앞서 제시한 CWK와 MRG의 강우특성들이 통계적으로 동질한 분포를 가지고 있는지에 대한 검정을 하기 위해 이표본 K-S(Two sample Kolmogorov-Smirnov) 검정을 통해 두 자료계열이 통계적으로 같은 특성을 가지고 있는지를 검정해

보았다. 일반적으로 일표본 K-S 검정은 표본자료계열이 특정 이론적 확률분포형에 적합하기를 판단하는 방법으로, 표본자료계열의 누가확률분포와 가정된 이론적 확률분포의 누가확률분포를 비교하는 것으로서, 양자의 최대편차가 표본의 크기와 유의수준에 따라 결정되는 한계편차보다 크면, 분포는 기각된다. 2개의 독립적인 표본자료계열이 동일한 모집단 또는 동일한 분포를 가지는 모집단으로부터 추출되었는지의 여부를 결정하기 위해서는 이표본 K-S 검정을 사용한다. 일표본 K-S 검정에서와 마찬가지로 이표본 K-S 검정도 2개의 누가확률분포 사이의 일치여부를 다루는데, 이 경우에 있어서는 양자가 모두 표본자료계열이다. 만일 두 표본자료계열이 동일한 모집단으로부터 추출되었다면 표본자료계열의 누가확률분포는 모집단 누가확률분포로부터 무작위 오차만을 보이기 때문에 서로 매우 유사하고 누가확률분포 사이의 최대편차가 상당히 크다면 귀무가설을 기각한다. 최대편차를 구하기 위해서는 최대누적편차를 구하는 일이 되지 않도록 가능한 한 간격을 조밀하게 잡아야 한다. 이표본 K-S 검정의 검증통계량 D_2 는 다음 식에 의하여 계산한다.

$$D_2 = \max\{|F_1(x) - F_2(x)|\} \quad (1)$$

$$= \max\{|F_1(x_i) - F_2(x_i)|\}_{i=1}^n$$

여기서, $F_1(x)$ 와 $F_2(x)$ 는 각각 표본1과 표본2의 누가도수이고 D_2 를 주어진 유의수준 α 로 적합성을 검정하고자 할 때 D_2 는 다음 식으로 정의되는 한계값 D_2^α 와 비교하게 된다.

$$P(D_2 \leq D_2^\alpha) = 1 - \alpha \quad (2)$$

식(2)와 같이 검정통계량 D_2 는 일표본 K-S 검정의 경우와 마찬가지로 계산되지만 한계값이 달라지는데 적어도 한 표본이 40을 넘는 경우엔 두 표본의 크기가 같을 필요가 없다. 그리고 K-S 검정의 경우, χ^2 검정에서와는 달리 표본자료를 이론분포에 맞출 필요가 없어 자료를 계급구간으로 나누어 이론분포에 근사하기 위한 매개변수를 계산할 필요가 없는 비모수적 방법으로 이 점이 χ^2 검정보다 유리한 점이다.

4.2 CWK와 MRG의 강우특성별 정량적 비교

그림 1에서 보는 바와 같이 상자그림(Boxplot)은 자료계열의 산포를 측정하는 방법의 하나로, 최소한계($Q_1 - 1.5IQR$), 제1사분위수(Q_1), 중위수(Median), 제3사분위수(Q_3), 최대한계($Q_3 + 1.5IQR$)의 5개의 중요 통계특성을 상자와 선으로 나타내어 자료의 중심과 산포, 모양을 개략적으로 파악할 수 있고, 이상치의 존재여부를 파악할 때 사용한다. 또한 그룹간 평균과 산포비교에 효과적이다. 제1사분위수와 제3사분위수는 크기순으로 배열된 자료에서 중위수를 기준으로 두 집단으로 나눈 뒤 다시 두 집단의 중위수를 찾았을 때 낮은 집단의 중위수가 제1사분위수이고 높은집단의 중위수가 제3사분위수이다. 사분위수범위(IQR)는 제3사분위수와 제1사분위수의 차이를 의미하며, 이상치(outlier)는 최소한계보다 작거나 최대한계보다 큰 값을 의미한다(wikipedia.org). 그림 1은 사분위

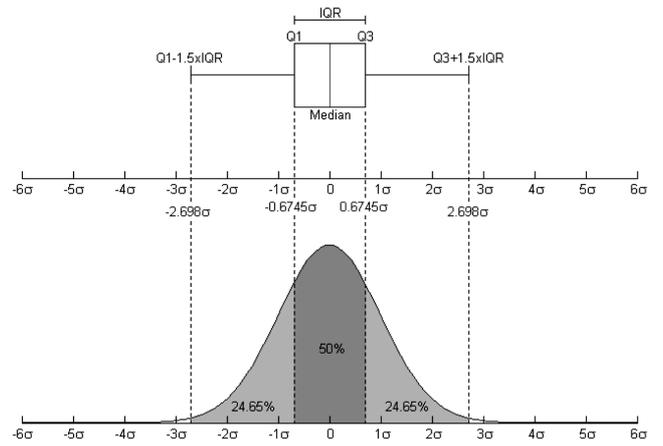


Fig. 1. Boxplot with an interquartile range(IQR); Source: wikipedia.org

범위를 가진 상자그림과 정규모집단의 확률밀도함수와의 관계를 보여주고 있다.

5. 정량적 동질성 분석 결과

5.1 CWK와 MRG의 강우특성별 통계적 분포에 대한 동질성 검정 결과

각 통계특성치 자료계열별로 유의수준 5%에서 이표본 K-S 검정을 실시한 결과는 다음 표 4와 같다. Monthly rainfall이나 D_{max} ratio는 M00의 5월을 제외하면 M00과 M20의 CWK와 MRG가 월별로 통계적 분포에 큰 차이가 없는 것으로 판정되었다. M00과 M20이 큰 차이를 보이지 않는 것은 월강우량에서 2mm이하의 강우가 차지하는 정량적인 비중이 크지 않기 때문이다. 그러나 $N_{rainy\ days}$ 는 M00의 경우 전체적으로 CWK와 MRG 간에 통계적으로 분포에 차이가 있다고 판정되었고 M20의 경우는 통계적으로 분포에 큰 차이가 없다고 판정되었다. 이러한 결과를 통해 측우기 관측계열과 근대우량계 관측계열 사이의 강우일수에 차이가 있음을 알 수 있고 이 차이는 시간적 변화로 해석될 수도 있으나 관측정밀도의 차이로 해석하는 것이 보다 적절하다고 판단된다. $I_{rainy\ days}$ 는 M00의 경우 9월을 제외하면 모두 CWK와 MRG 간에 통계적 분포에 차이가 있다고 판정되었고 M20의 경우는 4월, 7월, 9월, 10월이 CWK와 MRG 간에 통계적 분포에 차이가 있다고 판정되었다. M20의 CWK와 MRG간에 Monthly rainfall과 $N_{rainy\ days}$ 가 각각 통계적으로 차이가 없는데 반해 $I_{rainy\ days}$ 가 차이를 보이는 것은, Monthly rainfall과 $N_{rainy\ days}$ 각각에 통계적 유의수준 이내의 경년변화가 있다고 보여지며 두 요소가 결합되면서 통계적으로 유의한 경년변화 양상으로 나타났다고 판단된다.

5.2 CWK와 MRG의 강우특성별 정량적 비교 결과

다음 그림 2는 4월에서 10월까지의 측우기 관측자료계열(1778년-1907년, CWK)과 근대우량계 관측자료계열(1908년-2006년, MRG)에 대한 월별 강우량의 통계치 상자그림(Boxplot) 비교결과이다. 측우기 관측 최소 정밀도인 약

Table 4. Two-sample Kolmogorov-Smirnov test for CWK and MRG

Statistics	Month	H_0	
		M00	M20
<i>Monthly rainfall</i>	APR	A	A
	MAY	A	A
	JUN	A	A
	JUL	A	A
	AUG	A	A
	SEP	A	A
	OCT	A	A
$D_{max} ratio$	APR	A	A
	MAY	R	A
	JUN	A	A
	JUL	A	A
	AUG	A	A
	SEP	A	A
	OCT	A	A
$N_{rainy day}$	APR	R	A
	MAY	R	A
	JUN	R	A
	JUL	R	A
	AUG	R	A
	SEP	R	A
	OCT	R	A
$I_{rainy days}$	APR	R	R
	MAY	R	A
	JUN	R	A
	JUL	R	R
	AUG	R	A
	OCT	R	R

Note that A: Accepted, R: Rejected

2 mm를 고려하여 근대우량계 관측계열에서 2 mm 이하의 일강우량을 제외한 계열인 M20과 원시계열 M00으로 나누었다. 그림의 왼쪽이 M00이고, 오른쪽이 M20이다. 그리고 각각의 상자그림에서 왼쪽의 노란상자계열이 CWK이고, 오른쪽의 오렌지상자계열이 MRG이다.

*Monthly rainfall*에 대한 상자그림 결과를 보면 M00과 M20사이에는 물론, CWK와 MRG 사이에도 중위수와 사분위수범위에 큰 차이가 없음을 알 수 있다. 단, 7월의 최댓값이 M00은 CWK가 MRG보다 크나, M20의 경우는 MRG가 CWK보다 커졌음을 볼 수 있다.

$D_{max} ratio$ 는 M00의 경우 7월과 8월은 CWK와 MRG가 평균과 분산에 있어 유사한 값을 보이고 있다. 그러나 4월, 5월, 6월, 10월은 CWK가 MRG보다 중위수와 사분위수범위가 크게 나타나고 있다. 단, 9월은 중위수는 MRG가 크고, 사분위수범위는 CWK가 크게 나타났다. M20의 경우는 6월에서 9월을 제외하고 M00에 비해 CWK와 MRG의 중위수

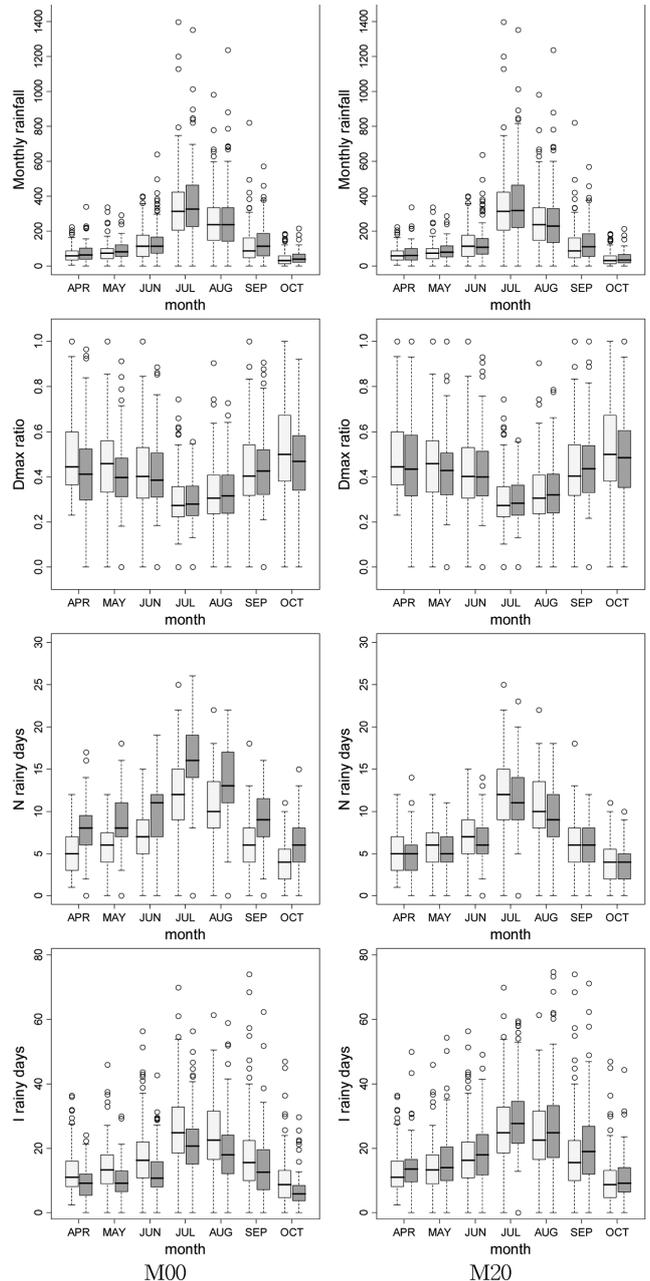


Fig. 2. Boxplots for each statistical rainfall characteristics (left box: CWK, right box: MRG)

차이가 작아졌다. M20의 6월에서 9월은 M00의 6월과 9월에 비해 CWK와 MRG 중위수 사이에 상대적으로 큰 변화를 보이지 않고 있다. 이는 M20이 M00에 비해 MRG의 월강우량이 작아져 발생한 결과로, 4월, 5월, 10월의 경우 갈수기로 *Monthly rainfall*의 작은 변화에도 $D_{max} ratio$ 이 민감하게 증가할 만큼 상대적으로 최댓일강우량이 작기 때문이다. M20 사분위수범위는 7월과 8월은 CWK와 MRG가 유사한 크기를 보이고 있고, 5월, 6월, 9월, 10월은 CWK가 MRG보다 크게 나타났다.

$N_{rainy days}$ 의 경우는 M00의 경우 CWK가 MRG에 비해 현저히 중위수가 작은 것을 알 수 있다. 특히 M00의 CWK의 경우 최솟값이 4월을 제외하고, 모두 0이나 MRG의 경우는 9월 이외에는 0이상입(MRG의 경우 0은 이상치 정도임)

을 알 수 있다. 그러나 M20의 경우는 CWK와 MRG의 중위수 차이가 현저히 감소하였다. 4월, 9월, 10월은 중위수가 거의 일치하고 5월, 6월, 7월, 8월은 CWK가 MRG보다 오히려 약간 크게 나타났다. 이를 통해 볼 때 과거에 측우기 관측밀도인 2mm이하의 강우는 기록이 안 되었다는 설이 타당성이 있어 보인다. 그러나 M20의 5월, 6월, 7월, 8월에서 MRG가 CWK보다 작은 값을 보이는 것으로 보아 과거의 강우일수가 현재와 크게 다르지 않다면, CWK가 2mm이상만 계속하였을 것이라는 가정에 대한 반론의 여지는 여전히 남게된다.

$I_{rainy\ days}$ 는 $Monthly\ rainfall$ 과 $N_{rainy\ days}$ 의 관계이므로 이 두 통계치와 밀접한 관련을 가지고 있다. M00의 경우 CWK가 MRG에 비해 중위수나 시분위수범위가 전체적으로 크다. 이는 앞서 설명하였듯이, CWK와 MRG간에 $Monthly\ rainfall$ 의 차이는 크지 않으나, $N_{rainy\ days}$ 의 차이가 크기 때문에 $N_{rainy\ days}$ 에 반비례한 결과로 해석할 수 있다. 즉, CWK의 기록 $N_{rainy\ days}$ 가 MRG에 비해 작기 때문으로 판단된다. M20의 경우는 오히려 전체적으로 MRG가 CWK보다 약간 큰 값을 보이고 있다. 이는 M00과 반대로 CWK와 MRG간에 $Monthly\ rainfall$ 의 차이는 크지 않으나 M20의 MRG $N_{rainy\ days}$ 가 CWK에 비해 약간 작기 때문으로 해석된다.

6. 결 론

본 연구에서는 통계적 기법을 이용하여 측우기 강우량 관측계열(CWK)과 근대우량계 강우량 관측계열(MRG) 사이의 정량적 동질성을 분석하였다. 이를 위해 이표본 K-S 검정을 통한 분포에 대한 동질성 검정과 상자그림을 이용한 정량적 비교를 수행하였다. 두 집단의 정량적 동질성 분석 결과를 정리하면 다음과 같다.

- 1) 각 분석과정에서 M00은 전체적으로 CWK와 MRG의 월강우일수간 차이에 명확한 통계적 유의성을 보이고 있어 CWK와 MRG 간의 관측정밀도에 차이가 있다고 추정된다.
- 2) CWK와 MRG 통계적 분포형의 동질성 검정을 위한, 이표본 K-S 검정결과 CWK와 MRG의 $Monthly\ rainfall$, $D_{max\ ratio}$, $N_{rainy\ days}$ 는 M20의 경우 통계적 분포

의 차이에 유의성이 없었다. $I_{rainy\ days}$ 는 M20의 경우 4월, 7월, 9월, 10월에서 통계적 분포의 차이에 유의성을 보였다.

- 3) 상자그림을 이용한 각 통계치의 정량적 비교결과는, M20의 경우, $Monthly\ rainfall$ 은 전체적으로 CWK와 MRG간에 큰 차이를 보이지 않았다. $D_{max\ ratio}$ 는 중위수 기준으로 7월, 8월, 9월은 MRG가 CWK보다 약간 크게 나타났고, 그 외는 CWK가 MRG보다 약간 크게 나타났다. $N_{rainy\ days}$ 는 CWK가 MRG보다 약간 크게 나타났고, $I_{rainy\ days}$ 는 전체적으로 MRG가 CWK보다 약간 크게 나타났다.

분석 결과를 종합해 보면, M20의 경우 강우량은 CWK와 MRG간에 명확한 통계적인 차이를 보이지 않았다. 그러나 강우일수나 강우강도는 CWK와 MRG간에 차이를 보이고 있는 것으로 분석되었다.

참고문헌

- 김기욱, 유철상, 박민규, 김현준 (2007) 강우빈도해석에서의 측우기자료의 유용성 평가. **한국수자원학회논문집**, 제40권, 제11호, pp. 851-859.
- 이병설 (1970) 서울의 연강수량 및 하기강수량의 Normality에 관한 연구. **한국기상학회지**, 제5권, pp. 11-14.
- 전종갑, 문병권 (1997) 측우기 강우량 자료의 복원과 분석. **한국기상학회지**, 제33권, 제4호, pp. 691-707.
- 정현숙, 임규호 (1994). 서울 지역 월강수량 강수일의 관계, 1770-1907. **한국기상학회지**, 제30권, 제4호, pp. 487-505.
- 조희규, 나일성 (1979) 18세기 한국의 기후변동 : 강우량을 중심으로. **동방학지**, 제22권, pp. 83-103.
- Arakawa, H. (1956) On the secular variation of annual total of rainfall at Seoul from 1770 to 1944. **Journal of the Korean Meteorological Society**, Vol. 7, No. 2, pp. 205-211.
- [Http://en.wikipedia.org/wiki/File:Boxplot_vs_PDF.png](http://en.wikipedia.org/wiki/File:Boxplot_vs_PDF.png), Permission: CC-BY-SA-2.5.
- 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.

- © 논문접수일 : 09년 04월 01일
- © 심사의뢰일 : 09년 04월 03일
- © 심사완료일 : 09년 07월 02일