

기상청 기후자료의 균질성 문제 (II): 통계지침의 변경

류상범¹ · 김연희^{2,*}

¹기상청 수원기상대

²기상청 기상연구소 응용기상연구실

(2006년 12월 13일 접수; 2007년 1월 8일 승인)

Inhomogeneities in Korean Climate Data (II): Due to the Change of the Computing Procedure of Daily Mean

Sang-Boom Ryoo¹ and Yeon-Hee Kim^{2,*}

¹Suwon Weather Station, Korea Meteorological Administration

²Applied Meteorology Research Laboratory, Meteorological Research Institute, KMA

(Manuscript received 13 December 2006; in final form 8 January 2007)

Abstract

The station relocations, the replacement of instruments, and the change of a procedure for calculating derived climatic quantities from observations are well-known nonclimatic factors that seriously contaminate the worthwhile results in climate study. Prior to embarking on the climatological analysis, therefore, the quality and homogeneity of the utilized data sets should be properly evaluated with metadata. According to the metadata of the Korea Meteorological Administration (KMA), there have been plenty of changes in the procedure computing the daily mean values of temperature, humidity, etc, since 1904. For routine climatological work, it is customary to compute approximate daily mean values for individual days from values observed at fixed hours. In the KMA, fixed hours were totally 5 times changed: at four-hourly, four-hourly interval with additional 12 hour, eight-hourly, six-hourly, three-hourly intervals. In this paper, the homogeneity in the daily mean temperature dataset of the KMA was assessed with the consistency and efficiency of point estimators. We used the daily mean calculated from the 24 hourly readings as a potential true value. Approximate daily means computed from temperatures observed at different fixed hours have statistically different properties. So this inhomogeneity in KMA climate data should be kept in mind if you want to analysis secular aspects of Korea climate using this data set.

Key words: Inhomogeneity in climate record, Metadata, Daily mean temperature, KMA

1. 서 론

지구온난화와 연관된 기후변화 문제는 대기과학자들의 과학적 탐구심을 넘어, 국가경제와 일상생활에 까지 그 영향력을 확장하고 있다 (IPCC, 2001). 따라서 지구온난화 문제와 관련된 과학적 평가의 기반이

되는 기후자료의 품질에 대한 엄밀한 평가 역시 간과할 수 없다. 관측지점의 이동, 관측기기 및 통계방법의 변경과 같은 기후정보 수집방법의 변경, 관측소 주변 환경의 변화, 자료수집 방법의 변경, 대규모 순환의 변화에 의한 자연적인 기후변동성, 산업화에 따른 대기 중 온실기체 농도의 증가 등은 지구온난화현상과 같이 오랜 시간에 걸쳐 점진적으로 발생하고 있는 기후변화를 분석, 평가하는데 있어 심각한 영향을 미칠 수 있는 외적 요인들이다 (Mitchell, 1953). 이러한 외적 요인 중 관측지점의 이동 및 기후정보 수집방법의 변경은 기후자료의 연속성에 인위적인 불연속을 야기하여 기후변화의 원인과 세기에 대한 과학적 추론 및 평가를 방해한다. 여기에서 기후자료의 연속성은 지구온난화와 같

*Corresponding Author: Yeon-Hee Kim, Applied Meteorology Research Lab., Meteorological Research Institute/KMA, 460-18, Shindaebang-dong, Dongjak-ku, Seoul 156-720 KOREA.
Phone : +82-2-846-2850, Fax : +82-2-846-2851
E-mail: kyh@metri.re.kr

은 장기 기후변화를 분석, 평가하는데 사용되는 각종 기후정보의 시계열분석에 앞서, 사용되는 시계열이 같은 장소에서 같은 방법에 의해 규칙적으로 생산되어 왔음을 보증하는 특성이며, 분석 결과에 대한 신뢰성과 정확성을 담보하는 시계열 분석의 1차적인 자료 구성 조건이다.

우리나라에서 기상현상의 관측은 삼국시대를 그 기원으로 하나, 체계적인 관측은 1441년 측우기 발명 이후로 볼 수 있다. 하지만 대부분의 기후 연구에 있어 활용하고 있는 자료는 대한제국 시절 일본중앙기상대의 임시 관측소 형태로 설치된 1904년 이후에 생산된 강수량을 포함한 8종(기압, 온도, 바람, 운량, 구름의 방향, 수증기 장력, 습도, 우설량)의 기상요소 관측값들이다(기상청, 2001; 2004). 하지만, 1904년 이래 생산된 기후자료에는 술한 관측 지점의 이전, 기상측기의 교체 혹은 변경, 일평균기온과 같은 일평균값을 산출하는 방법의 변경과 같은 기후자료의 연속성을 단절시키는 인위적인 외적 요인들이 많이 포함되어 있다.

본 논문은 기상청에서 1975년 이후 전산화하여 데이터베이스를 구축한 기후자료의 연속성을 파악하여 기후자료 메타데이터와 연계하여 관리하기 위해 수행한 기상청 기후자료의 균질성 분석의 일부이다. 본 논문은 이중 위상 회귀모형(two-phase regression model)을 사용하여 관측지점의 이전으로 인한 기후자료의 연속성 단절을 분석한 류상범 등(2006)의 연속 논문이다. 본 논문에서는 일평균기온을 예로 하여 1904년 이후 일평균값을 산출하기 위해 사용하였던 관측 횟수와 시각들이 어떻게 변해왔으며, 이러한 변화가 기후자료 연속성에 끼친 영향을 평가하고자 한다.

기상청 지상기상통계지침에 의하면, 일평균값이란 ‘하루 중 시간별 합계값을 하루 관측 횟수로 나눈 것을 말하며,.....정시관측값을 사용하는 것을 표준으로 한다’. 정시관측이란 정해진 시각(fixed hour)에 기상관측을 실시하는 것이다. 초창기에는 06시를 기준으로 4시간이나 8시간 간격으로 주로 수행되었으나, 1950년 3월에 세계기상기구(World Meteorological Organization, WMO)가 창설된 이후 03시를 기준으로 3시간이나 6시간 간격으로 기상관측이 이루어지고 있다. 우리나라는 1956년 WMO에 가입한 후 기상관측방법을 WMO 규정에 따라 점차적으로 개정하였다(기상청, 2004). 기상청에서 발간한 *기상관측환경편람*(기상청, 1995)과 *기상통계자료집*(기상청, 2001), *근대기상 100년사*(기상청, 2004) 등에 따르면, 1904년 이후 우리나라에서 일

평균값을 산출하는 방법이 시대에 따라 계속 변해 왔으며, 관측에 사용된 측기의 형태 및 종류 역시 달라져 왔다. 예를 들어 1904년의 일평균기온은 유리제 건습구 온도계와 자기온도계로 관측된 1일 6회(02시, 06시, 10시, 14시, 18시, 22시)의 기온을 더해 6으로 나눈 값이지만, 2006년 현재의 일평균기온은 종관기상관측시스템(Automated Surface Observing System, ASOS)의 차광통풍기 내에 설치된 온도센서에 의해 관측된 1일 8회(03시, 06시, 09시, 12시, 15시, 18시, 21시, 24시)의 기온을 더해 8로 나누어 구한 값으로 서로 차이가 많다. 본 논문에서는 관측기기의 변경과 교체에 대한 메타데이터의 확보가 쉽지 않아 이에 대한 평가를 수행하지 못하고, 오로지 일평균 산출에 사용된 횟수와 시각의 변경으로 인한 기후자료 연속성 단절에 대해서만 논하였다. 특히, 이러한 요인의 영향은 기후변화의 과학적인 해석을 방해하므로, 이용자가 기후자료를 사용할 때 주의해야 한다.

2. 일평균기온의 산출 방법과 자료, 분석방법

기상청 기후자료관리시스템의 기상월보에서 제공하는 일별 기상 자료에는 평균해면기압, 평균기온, 평균이슬점온도, 평균상대습도, 평균풍속, 평균운량 등 일평균값으로 나타내고 있는 기후정보들이 포함되어 있다. 이중 일평균기온은 오늘날 기후변화 연구에서 가장 많이 사용되고 있는 가장 중요한 기후정보 중 하나이다. 특히, 기온에는 뚜렷한 일일주기의 변화 특성이 나타나므로 관측시각의 변경에 따른 일평균값의 차이가 다른 요소들보다 더 잘 나타날 것으로 생각된다. 이에 본 연구의 분석 요소로 선택하였다.

한반도에서 체계적인 기상관측이 처음 시작된 대한제국 시절 이후의 기상관측에 대한 메타데이터들이 *기상통계자료집*(기상청, 2001)과 *근대기상 100년사*(기상청, 2004)에 서술되어 있지만, 서로 일치하지 않아 본 논문에서는 *기상통계자료집*의 기록을 따랐다. 대한제국 시절, 당시 한반도의 기상업무를 총괄하던 인천관측소에서는 관측자에 의해 실시되었던 1일 6회(02시, 06시, 10시, 14시, 18시, 22시)의 정시관측값으로, 부산 등 나머지 부속 측후소에서는 관측자에 의해 실시된 1일 3회(06, 14시, 22시)의 정시관측값과 1일 3회(02시, 10시, 18시)의 자기기록계 성적에서 취한 값으로 일평균기온을 산출하였다. 1917년부터는 12시의 값을 더해 인천의 조선총독부관측소에서는 1일 7회

의 정시관측값으로, 나머지 측후소에서는 1일 3회의 정시관측과 1일 4회의 자기기록계의 값으로 일평균기온을 산출하였다. 1940년부터는 모든 관측소에서 1일 3회의 정시관측값으로 일평균기온을 산출하였다. 한국전쟁이 끝난 1953년부터 1일 4회 (03시, 09시, 15시, 21시)의 정시관측값으로 일평균기온을 산출하다가, 1997년부터 현행의 1일 8회 (03시, 06시, 09시, 12시, 15시, 18시, 21시, 24시)의 정시관측값으로 일평균기온을 산출하고 있다. Table 1은 1904년 이후 우리나라 기상업무에서 일평균기온을 산출하기 위해 선택된 시각과, 그리고 사용된 측기를 유리제 온도계, 자기온도계, 디지털방식의 Thermocouple로 구분하여 정리한 것이다.

Table 1에 정리된 것처럼 현재 기상청 기후자료 데이터베이스에 전산화되어 ‘일평균기온’이라는 같은 이름으로 저장되어 있는 정보들은 실은 시대에 따라 다른 산출 방법으로 생산된 값들이다. 본 논문에서는 기상청 기후자료 데이터베이스에서 서울과 부산 지역의 매시 기온관측 자료 (1999년~2005년, 7년간)를 추출하여, 일평균 산출에 사용된 횟수와 시각의 변경에 의한 일평균값의 차이를 분석하였다. 본 논문에서 사용한 분석방법은 단순하다. 24개의 매시 관측값을 단순산술평균하여 산출한 일평균기온 (T_{24})을 ‘잠재적 참값’으로 정하고, 1904년 이래 일평균기온을 산출하기 위해 사용했던 시각의 값들로 구한 각종 일평균기온 (T_6, T_7, T_3, T_4, T_8)과의 차이를 구하여 그 통계적 특성을 분석하였다. WMO 규정집에 일평균기온의 유용한 근사값으로 명시되어 있는 최고기온과 최저기온의 평균으로 구한 $T_{extreme}$ 을 계산하여 잠재적 참값과 차이를 비교하였다. 또한 1983년 1월 1일부터 1999년 12월 31일까지 17년간의 서울지역 1일 8회 관측값을 사용하여 1일 8회 관측값을 사용한 T_8 과 1일 4회 관측

값을 사용한 T_4 의 차이를 심층 분석하였다. 2000년 1월 1일 이후에도 ASOS의 차광통풍기 내 설치된 온도 센서에 의한 1일 8회의 관측값으로 일평균기온을 산출하고 있지만, 측기의 변경으로 인한 영향을 미연에 방지하기 위해 백엽상 내 설치된 유리제 온도계로 1일 8회 관측이 이루어졌던 1999년 12월 31일까지의 1일 8회 관측값만 가지고 재생산한 T_8 와 T_4 의 차이를 비교하였다.

3. 결과

Figure 1은 1999년에서 2005년까지 7년간 서울에서 매시 관측한 기온값을 사용하여 일평균기온의 잠재적 참값 (T_{24})과 그동안 일평균기온을 산출하기 위해 표본 구성을 달리한 $T_6, T_7, T_3, T_4, T_8, T_{extreme}$ 의 차이를 시계열과 돛수분포도로 나타낸 것이다. Table 2에 일평균기온 산출방법에 따른 차이들의 평균과 표준편차를 정리하였다. 여기서 차이들의 평균은 편이 (bias)로서, 잠재적 참값과의 불일치성을 의미한다. 그리고 차이들의 표준편차는 추정치들이 참값 주변에 집중된 정도를 나타내므로 산출방법의 유효성을 상대적으로 평가하는데 사용된다. 즉 표준편차가 작을수록 추정치들이 참값 주변에 집중됨을 나타내므로 유효성이 좋은 것으로 평가된다 (Wilks, 1995). 가장 초기에 사용했던 T_6 의 편이가 0.004°C 로 다른 산출 방법보다 잠재적 참값에 가장 일치하며, 변화도 가장 적다. 현재 일평균기온을 구하는데 사용하고 있는 T_8 역시 잠재적 참값과 차이를 작으나, 2000년 여름 부근에 다른 해에 비해 약간 더 큰 양의 차이를 보여준다. 20세기 후반 대부분 기간동안 사용됐던 T_4 는 T_8 와 비슷한 특성을 보여주지만 차이의 변화가 심하다. T_3 과 잠재적 참값의 차이는 가장 뚜렷한 계절변화를 한다. 즉 겨울에

Table 1. Fixed hours for computing of KMA daily mean temperature, and instruments used since 1904.

Period	Fixed hours chosen for daily mean	Instrument
1904-1916	6 fixed hours (02, 06, 10, 14, 18, 22)	Liquid-in-glass thermometer and Thermograph
1917-1939	7 fixed hours (02, 06, 10, 12, 14, 18, 22)	Liquid-in-glass thermometer and Thermograph
1940-1952	3 fixed hours (06, 14, 22)	Liquid-in-glass thermometer
1953-1996	4 fixed hours (03, 09, 15, 21)	Liquid-in-glass thermometer
1997-1999	8 fixed hours (03, 06, 09, 12, 15, 18, 21, 24)	Liquid-in-glass thermometer
2000-	8 fixed hours (03, 06, 09, 12, 15, 18, 21, 24)	Thermocouple

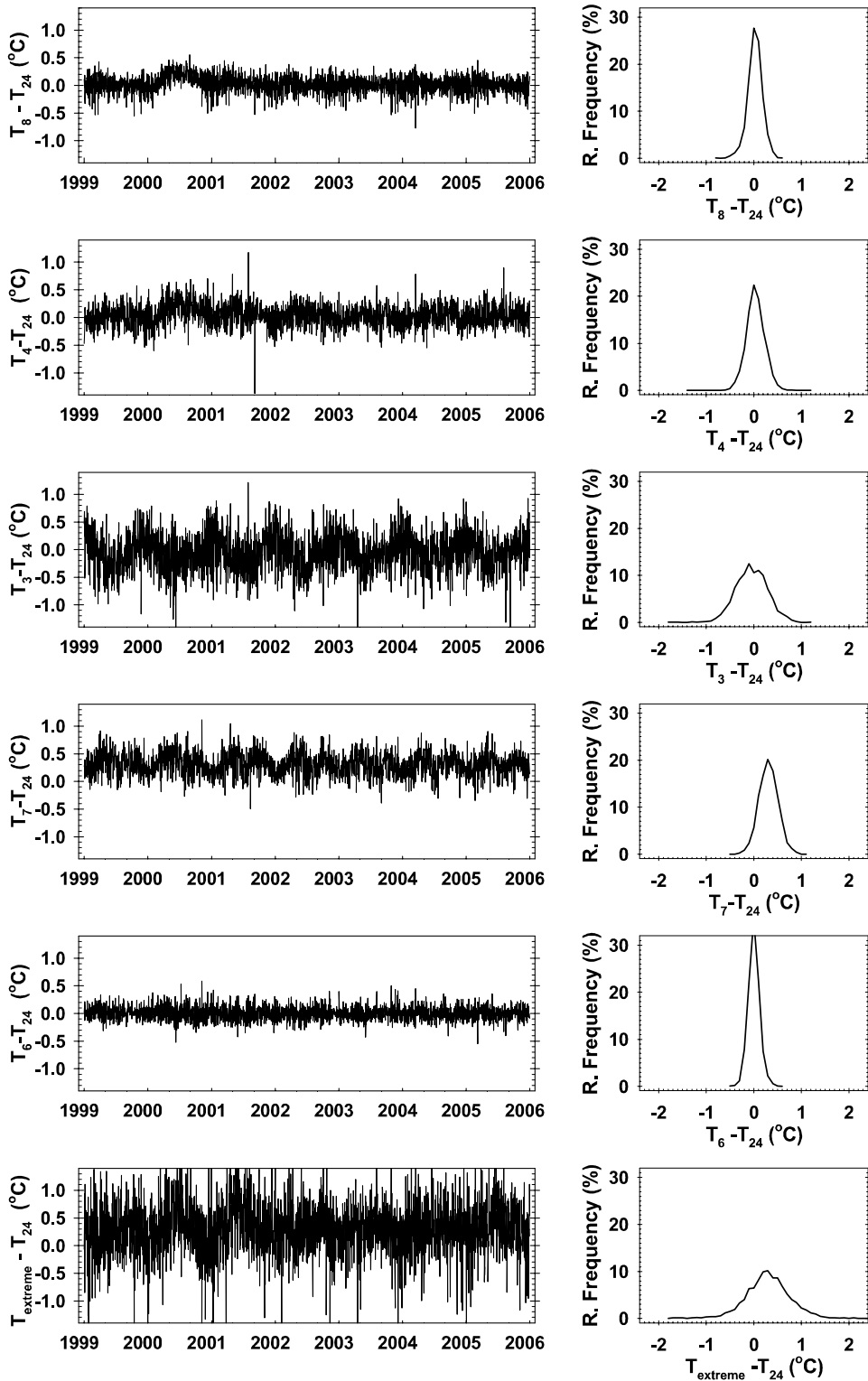


Fig. 1. Time series and distribution of the differences between the daily mean temperature computed from 24 hourly recordings and the daily mean temperatures computed from values observed at different fixed hours at Seoul.

는 참값보다 크고 여름에는 참값보다 작음, 계절에 따라 잠재적 참값과의 차이의 방향이 달라지는 일평균기온을 기록한다. 잠재적 참값과 가장 잘 일치하는 T_6 에 12시의 기온값을 더해 일평균기온을 구한 T_7 의 경우, 0.308°C 에 해당하는 뚜렷한 양의 편이를 가지면서 2중 모드의 계절변화를 한다. T_7 의 일평균기온과 잠재적 참값의 차이에 2중 모드의 계절변화를 만들고 T_6 에 비해 0.304°C 더 큰 편이를 만든 원인은 단순히 더해진 12시 기온 때문이다. Figure 2는 서울에서 매시 관측한 기온값으로 잠재적 참값 T_{24} 을 계산한 후 각 시각별 기온과의 차이를 7년간 (1999년~2005년) 평균한 값으

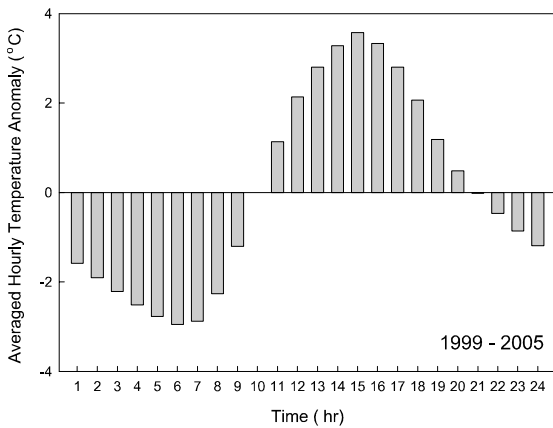


Fig. 2. Average of the hourly temperature anomalies from the daily mean temperature which was computed from 24 hourly temperatures at Seoul (1999~2005).

로, 잠재적 참값 T_{24} 에 대한 각 시각별 기온의 기여도를 의미한다. 12시 기온의 경우 2.1337°C 로서, T_6 에 이 값을 더해 T_7 을 계산하면 2.1337°C 의 1/7에 해당하는 0.304°C 만큼 일평균기온이 높아진다. 그리고 T_7 에 포함된 12시 기온의 기여도가 계절에 따라 달라진다. 4월과 9월에는 상대적으로 높고 1월과 7월에 상대적으로 낮은 2중 모드의 변화를 한다. 최고기온과 최저기온을 평균하여 구한 $T_{extreme}$ 과 잠재적 참값의 차이는 뚜렷한 양의 편이 (0.298°C)와 가장 큰 변화의 폭 (표준편차 = 0.482°C)을 보여준다.

잠재적 참값과의 차이를 월별로 box plot를 작성하여, 시계열에서 나타난 계절별 변화 특성을 확인하였다 (Figure 3). 그림의 box plot에서 가장 위와 아래에 위치한 점은 95분위수와 5분위수, 상자 밖 수평 띠는 90분위수와 10분위수, 상자 외곽은 75분위수와 25분위수, 상자 안 수평 띠는 평균을 각각 나타낸다. 상자 길이와 상자 밖 수평 띠 사이의 간격이 짧을수록 집중도가 좋다. T_3 과 $T_{extreme}$ 의 변화가 가장 심하며 T_7 에서는 이중모드의 계절변화가 나타난다. 월별 box plot 분석에서도 T_6 이 월에 관계없이 가장 작은 차이를 보여준다. T_4 , T_8 모두 뚜렷한 계절변화 특성은 보이지 않으나, T_4 가 T_8 보다 변화의 폭이 크다.

이러한 특성의 지역성을 알아보기 위해, 같은 기간 (1999년~2005년) 부산에서 매시관측한 기온값을 사용하여 잠재적 참값 (T_{24})과 T_6 , T_7 , T_3 , T_4 , T_8 , $T_{extreme}$ 의 차이를 Figure 4에 시계열과 dots분포도로 나타내었다. Figure 1에서 나타난 특성 가운데 다른 해

Table 2. Mean and standard deviation of differences between the daily mean temperature computed from 24 hourly recordings and the daily mean temperatures computed from values observed at different fixed hours presented in Table 1.

	Station	Mean	Standard deviation
$T_8 - T_{24}$	Seoul	0.026	0.156
	Busan	-0.044	0.158
$T_4 - T_{24}$	Seoul	0.037	0.195
	Busan	-0.022	0.234
$T_3 - T_{24}$	Seoul	-0.046	0.338
	Busan	-0.054	0.356
$T_7 - T_{24}$	Seoul	0.308	0.203
	Busan	0.343	0.223
$T_6 - T_{24}$	Seoul	0.004	0.124
	Busan	-0.020	0.144
$T_{extreme} - T_{24}$	Seoul	0.298	0.482
	Busan	0.469	0.460

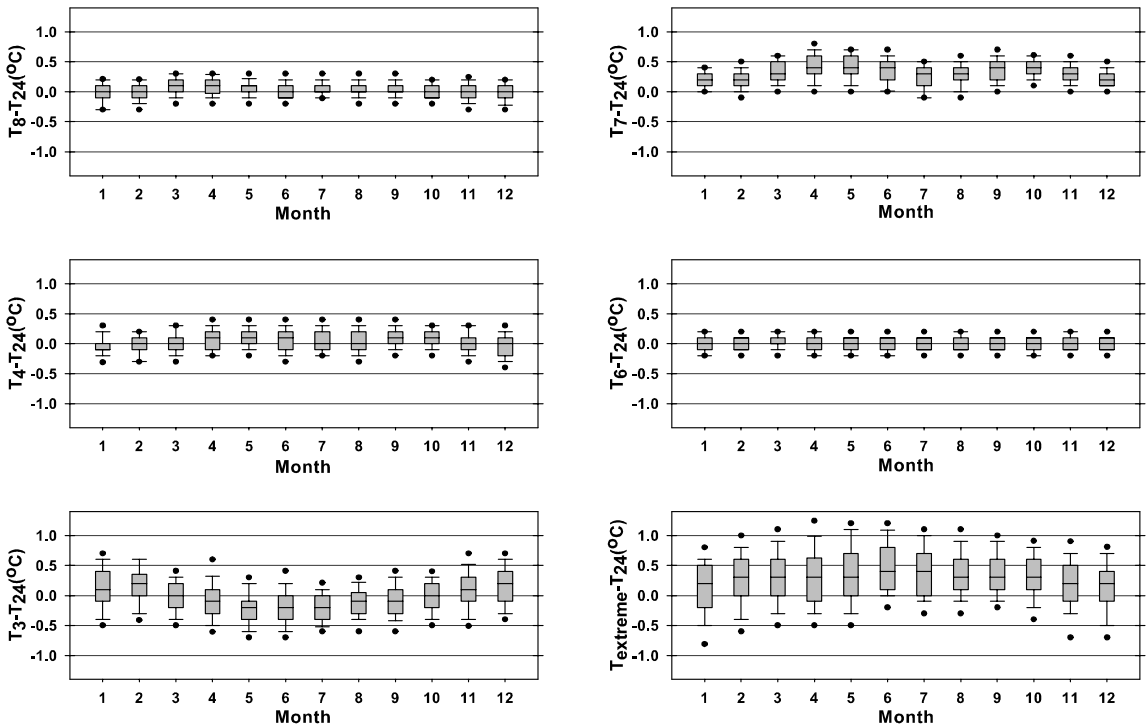


Fig. 3. Box plot of differences between the daily mean temperature computed from 24 hourly recordings and the daily mean temperatures computed from values observed at different fixed hour.

에 비해 2000년 여름 부근에 약간 더 큰 양의 차이를 나타낸 T_4 와 T_8 의 특이성을 제외한 대부분의 특성들이 Figure 4에 비슷하게 나타났다. Figure 4의 T_4 와 T_8 시계열에서는 2001년 여름 부근에 다른 해에 비해 약간 더 큰 음의 차이가 나타났다. 돛수분포도의 형태나 Table 2에 정리된 통계량들의 크기가 서울의 경우와 수치적으로 약간의 차이가 있으나 의미하는 바는 거의 유사하다. 본 논문에서는 제시하지 않았지만 box plot를 사용한 월별 변화 분석 결과도 서울의 경우와 같은 특성을 보여 주었다. 따라서 그동안 일평균기온을 산출하기 위해 표본 구성을 달리한 T_6 , T_7 , T_3 , T_4 , T_8 , $T_{extreme}$ 과 잠재적 참값 (T_{24})의 차이의 특성이 지역에 관계없이 거의 일정하게 나타나고 있으며, 가장 초기에 사용했던 T_6 이 잠재적 참값과 가장 일치했다. 현재 일평균기온을 산출하기 위해 사용하고 있는 T_8 의 경우 잠재적 참값과 차이가 크지 않으나, 간혹(서울에서는 2000년 여름, 부산에서는 2001년 여름과 같이) 불규칙적인 특이한 차이를 나타내기도 한다.

참고로 본 논문에 사용한 7년간의 매시 관측값을 생

산한 측기의 종류는 2000년 1월 1일을 기점으로 백엽상 내 유리제 온도계에서 ASOS의 차광통풍기 내 설치된 온도센서로 바뀌었다. 실제 분석에서는 유리제 온도계에 의한 2년간 매시 관측값과 ASOA에 의한 5년간의 매시 관측값을 분리하여 비교하였지만, 그 차이가 통계적으로 유의하지 않고 또한 T_8 에서 나타나는 서울의 2000년 여름과 부산의 2001년 여름의 특이성을 보이기 위해 본 논문에서는 7년간 자료를 사용하여 분석한 결과를 제시했다.

우리나라 기상관측 조직과 체계가 오늘날과 비슷하게 구축된 20세기 후반 대부분 일평균기온을 산출하기 위해 사용한 T_4 와 현재 사용하고 있는 T_8 의 차이를 1983년 1월 1일부터 1999년 12월 31일까지 17년간 서울에서 1일 8회 관측한 기온값을 사용하여 분석하였다. Figure 5의 산포도에 나타났듯이 T_8 과 T_4 의 상관계수는 99.971%로 선형 상관성이 아주 강하나 약간의 불일치가 두 값 사이에서 감지된다. 이러한 T_8 과 T_4 를 각각 월평균하여 구한 월평균기온의 상관계수는 99.999%로 거의 완전 상관에 가까운 일치성을 외형적

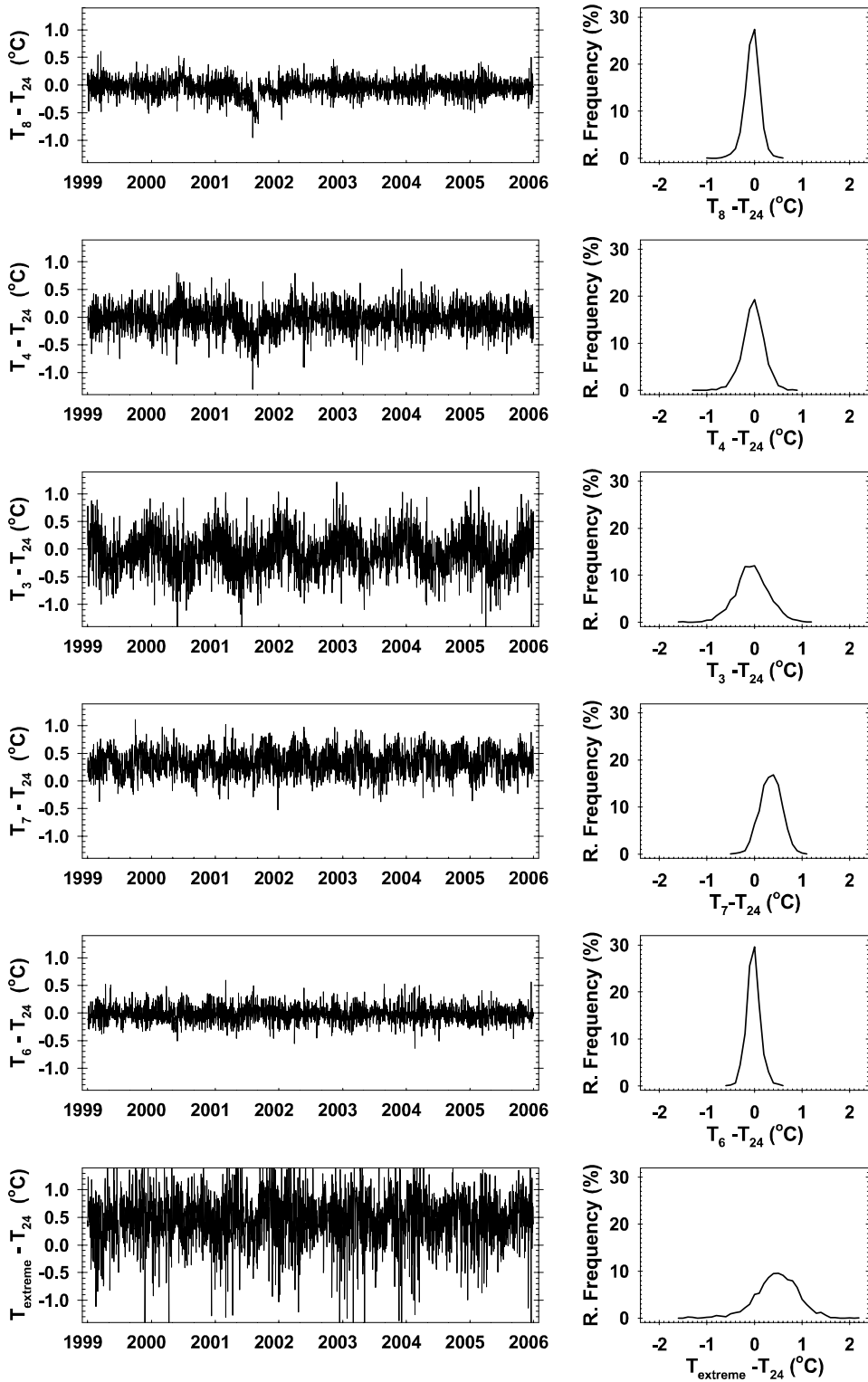


Fig. 4. Same as Figure 1 but at Busan.

으로 나타내고 있다 (Figure 6). 하지만 두 월평균기온 사이의 차이를 box plot로 표현한 Figure 7에서 최대 -0.1°C 에서 $+0.1^{\circ}\text{C}$ (평균 -0.05°C 에서 $+0.05^{\circ}\text{C}$) 범위에서 이중 모드 계절변화를 하고 있는 두 월평균기온 사이의 불일치성을 뚜렷하게 볼 수 있다. 여기서 box

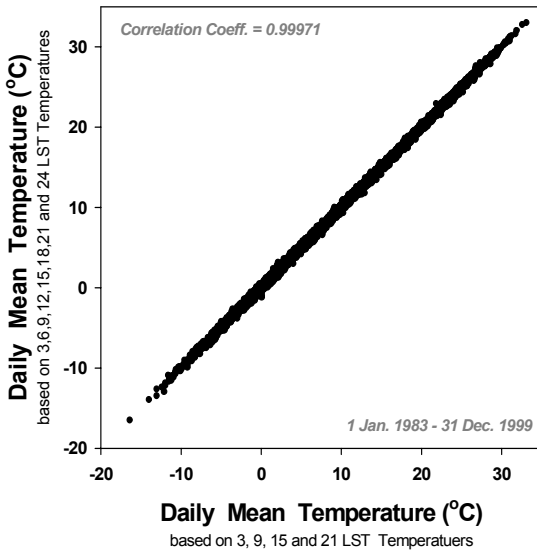


Fig. 5. Scatted diagram of the daily mean temperature computed from values observed at 6-hourly intervals and daily mean temperature at 3-hourly intervals at Seoul (1983-1999).

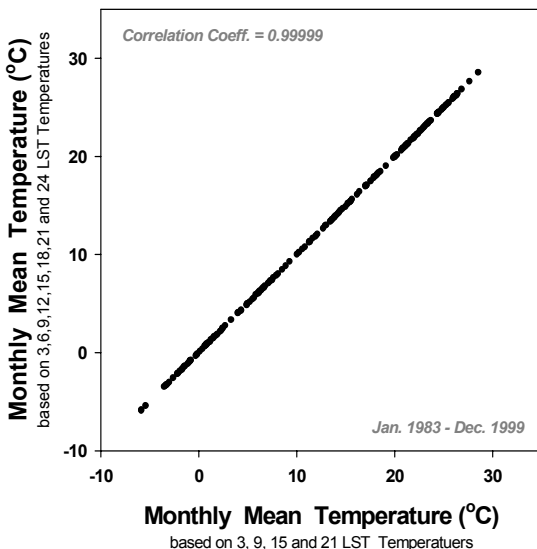


Fig. 6. Same as Figure 5 but monthly mean temperature.

plot의 세부 설명은 Figure 3과 같다. 단순히 일평균기온 산출 방법의 차이로 인해, 오늘날 2월의 월평균기온 값은 20세기 후반부의 2월 월평균기온값들보다 약 0.07°C 정도 높게 추정될 수 있다. 이러한 차이는 지난 세기 1.5°C 의 평균기온 상승 (월 0.00125°C 의 점진적인 기온 상승과 같은 크기) (권원태, 2005)과 비교하면, 결코 작지 않은 크기이다.

기후 실무를 위한 WMO규정집에 의하면 (WMO, 1983), 정시에 관측된 값으로 일평균값을 계산하는 것이 관습적이며 만약 매시간 관측값을 활용할 수 있다면 24시간 관측값의 평균으로 일평균값을 계산하며 종관관측소에서는 3시간 혹은 6시간 간격의 관측값을 사용한다고 명시되어 있다. 또한 최고, 최저기온만 관측하는 대부분의 기후관측소에서는 최고기온과 최저기온의 평균으로 구한 일평균기온이 유용한 근사값이지만 다른 방법으로 구한 일평균기온과 서로 비교할 때에는 한계를 고려해야 한다고 설명하고 있다. Figure 8은 서울에서 1908년 이래로 관측한 최고기온과 최저기온을 평균하여 구한 일평균기온과 그동안 여러 가지 방법으로 구한 일평균기온의 차이를 시계열과 각 방법별 도수분포도로 나타낸 것이다. 한국전쟁으로 1950년에서 1952년 일부 자료가 없다. Figure 8은 일평균기온을 구하는 방법의 차이에 관계없이 약 $0.3\sim 0.4^{\circ}\text{C}$ 의 편이와 약 $0.5\sim 0.7^{\circ}\text{C}$ 의 표준편차를 보여주고 있다. Table 2에 제시된 잠재적 참값과 $T_{extreme}$ 의 차이들의 평균 0.298°C , 표준편차 0.482°C 에 비해 더 나쁜 불편성과 유효성이다. 이는 그동안 사용된 각 방법의 불일치성에 비교한 $T_{extreme}$ 의 불일치성이 합쳐진 것으로

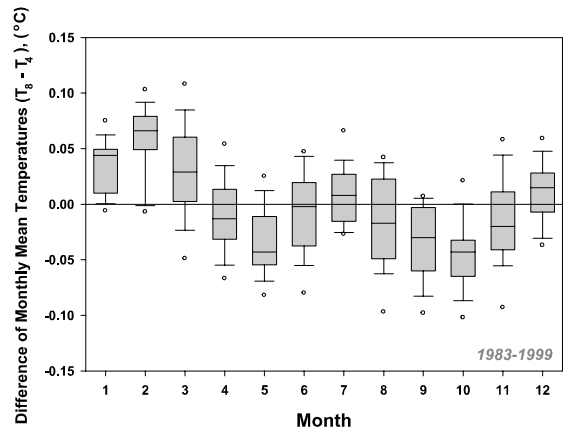


Fig. 7. Box plot of differences between the monthly mean temperatures based on values observed at 6-hourly intervals and those at 3-hourly intervals at Seoul (1983-1999).

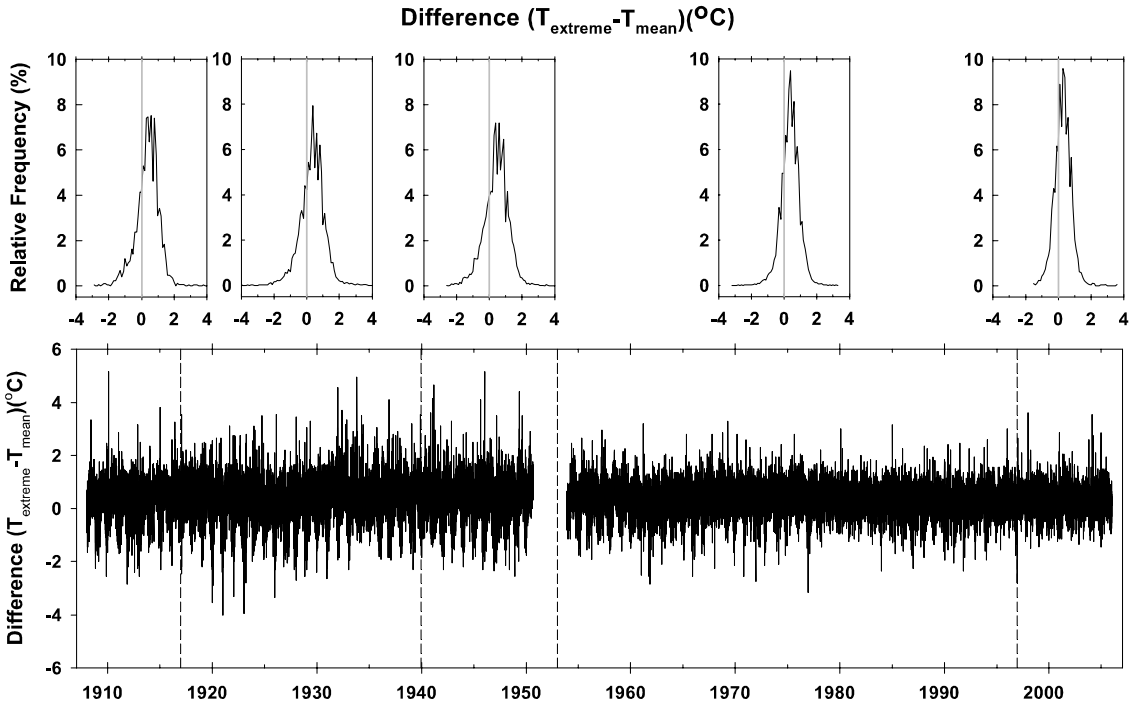


Fig. 8. Time series and distributions of differences between the daily mean temperatures recording at KMA climate database and alternative daily mean temperatures given by the average of the maximum and minimum temperatures at Seoul (1908-2005).

해석된다. 따라서 WMO (1983)의 권고처럼, 기후자료로서 영속성이 단절된 일평균기온 자료의 유용한 대체물로 최고기온과 최저기온을 평균한 값을 기후변화 연구 등에 사용하는 것은 그렇게 좋지 않다.

4. 결론 및 토의

기온은 개념적으로 일정범위의 값을 연속적으로 가지는 연속확률변수이므로, 일평균기온이라는 정보를 정확하게 얻기 위해서는 현실적으로는 전체조사가 불가능한 무한대의 기온 관측값들이 필요로 한다. 따라서 실제적으로는 편의를 위해, 관측오차 등을 고려한 최소 의미의 관측 단위 (기온의 경우 보통 0.1°C)를 사용하고, 정시관측이라는 일정시간의 관측을 통해 구한 값들을 표본으로 활용해 일평균기온을 구하고 있다. 이러한 표본조사에는 필연적으로 표본오차가 수반되므로 표본조사에 의해 구한 값은 구하고자 하는 참값과 정확히 같지 않고, 참값을 추정하기 위한 하나의 추정치로서 일종의 확률변수이며 확률분포를 가진다. 표

본오차 이외에 표본의 불완전한 선택 등으로 발생하는 비표본오차가 표본조사에 의한 추정치에 포함되어 있을 수 있다. 이러한 비표본오차는 이론적으로 통제하기 곤란하므로 표본을 추출할 때 각별히 주의를 해야 한다.

본 논문은 1904년 이래 생산되어 기상청 기후자료 데이터베이스에 일평균기온이라는 이름으로 전산화되어 있는 자료들에 대한 영속성을 평가하였다. 그동안 수차에 걸쳐 일평균기온을 산출하는데 사용된 기온관측값의 횟수와 관측 시각이 변경되었다. 본 논문에서는 그동안 사용된 방법에 의해 추정된 일평균기온들의 불편성과 유효성 등을 상대 비교하여, 방법 차이에 의한 일평균기온 추정이 달라질 수 있음을 보였다. 즉 기상청 기후자료 데이터베이스에 전산화되어 있는 일평균기온 자료의 영속성 단절을 보였다. 기온과 같이 뚜렷한 일변화를 하는 기상요소의 일평균값을 산출하는데 있어 관측 시각의 구성은 단순한 횟수의 다소보다 관측 간격의 일정성 (3시간 간격의 8회나, 4시간 간격의 6회, 6시간 간격의 4회, 8시간 간격의 3회)이 더 중

요하며, 일정한 관측 간격에서도 횡수의 다소는 그다지 중요하지 않았다. 관측 간격이 일정하면, 간격폭에 관계없이 불편성 (unbiasedness)이 거의 일치하며 06시를 기점으로 한 1일 6회의 관측 시각이 가장 유효성 (efficiency)이 좋았다. 임의의 관측시각 (예: 1일 7회 관측의 12시)을 더했을 때 참값에서 가장 편이된 결과를 얻었다. 따라서 조사한 여러 가지 일평균기온 산출 방법 가운데 가장 초기에 사용했던 T_6 이 편이가 가장 작고 분산이 최소인, 참값의 ‘최소 분산 불편 추정량’ (minimum variance unbiased estimator)에 가깝다. 즉 참값과 일치성이 가장 좋으며 유효성이 좋은 일평균기온의 추정치이다. 현재 사용하고 있는 T_8 은 그 다음으로 일치성과 유효성이 모두 좋은 추정치이다.

현재 가지고 있는 정보 모두를 사용하여 복원할 수 있는 가장 좋은 일평균기온의 충족 추정량 (sufficient estimator)은 전산화되어 있는 1999년 이후의 매시 기온 관측값을 이용하는 한편, 그 이전에는 자기기록계에 기록된 매시 기온값들을 디지털화하여 계산한 T_{24} 이다. 하지만 자기기록계의 값들을 디지털화하는 것은 비용문제가 있으므로 현재 1일 8회 기온 관측값이 전산화되어 있는 1983년까지 일평균기온을 T_8 로 재계산하여 사용하는 것이 차선책이다. 이는 기후자료로서 연속성을 가지고, 일평균기온의 참값에 보다 가까운 ‘최소 분산 불편 추정량’의 사용을 현실적으로 가장 연장할 수 있는 것이다. 본 논문의 평가에서 나타난 T_6 은 표본을 구성하고 있는 정시관측 체계가 오늘날과 완전히 달라 사용이 어렵다.

본 논문은 기상청 기후자료관리시스템에 데이터베이스된 누년 기후자료가 가지는 있는 문제점에 대해 조사한 것으로, 조사방법론에 있어 독창성이나 이론적 깊이는 없다. 조건부 확률과 같은 복잡한 방법을 사용할 경우 방법론의 타당성을 증명하기 위해 많은 노력이 필요하나, 본 논문에서처럼 단순한 방법을 사용할 경우 누구나 분석 방법의 정오를 쉽게 확인할 수 있고 결과에 공감할 수 있는 장점이 있다. 본 논문의 목적은 우리나라 기상업무에서 일평균값을 구하는 데 있어 사용된 정시관측시각의 변경되어 왔음을 알리고, 이로 인한 일평균값의 차이를 평가하는 것이다. 이는 지구 온난화 같은 기후변화 연구에서 많이 사용되고 있는 시계열 분석에 있어, 시계열을 구성하는 값들은 동일

한 정의와 방법에 의해 같은 시간 단위로 규칙적으로 구해져야 한다는 가장 기본적인 요건을 강조하기 위해서이다. 이러한 목적을 달성하는데 있어 본 논문에서 사용된 단순한 조사방법론이 충분하다고 저자들은 생각한다. 본 논문과, 그리고 연속논문인 류상범 등 (2006)에서 다른 기상청의 누년 기후자료의 연속성 단절은 이들 자료를 사용하는 혹은 사용할 연구자들이 반드시 유념해야 할 사안이라 생각된다.

감사의 글

이 연구는 기상연구소의 “국지 기상 특성 진단 및 기상환경영향평가 기술 개발 연구” 과제의 지원으로 수행되었습니다. 저자들은 연구를 수행하는 동안 메타자료수집에 도움을 준 기상청 기후자료팀의 정선애씨와 자료처리에 도움을 준 기상연구소 응용기상연구실의 권태현 연구원에게 감사드립니다. 그리고 논문의 질적 향상에 많은 조언을 주신 두 분의 심사위원들에게 감사드립니다.

참고문헌

- 권원태, 2005: 기후변화의 과학적 현황 및 전망. *한국기상학회지*, **41**, 325-336.
- 기상청, 1995: 기상관측환경편람. 09200-73220-67-9505, 192 pp.
- 기상청, 2001: 기상통계자료집. 11-1360000-000089-14, 192 pp.
- 기상청, 2004: 근대기상 100년사. 11-1360000-000143-01, 1032 pp.
- 류상범, 김연희, 권태현, 박일수, 2006: 기상청 기후자료의 균질성 문제 (I): 관측지침의 이전. *대기*, **16**, 215-223.
- IPCC, 2001: *Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge, United Kingdom, 881 pp.
- Mitchell, J. M., Jr., 1953: On the causes of instrumentally observed secular temperature trends. *J. Meteor.*, **10**, 244-261.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- WMO, 1983: *Guide to Climatological Practice*. World Meteorological Organization. WMO-No 100.