

한반도 연안 기온자료의 분포함수 추정 Estimation on the Distribution Function for Coastal Air Temperature Data in Korean Coasts

정신태* · 조홍연** · 고동희* · 황재동***

Shin Taek Jeong*, Hongyeon Cho**, Dong Hui Ko* and Jae Dong Hwang***

요 지 : 연안 생태환경변화의 주요 지표로 이용되는 수온은 기온과 밀접한 상관관계가 있기 때문에 기온자료를 이용하여 기후변화에 따른 수온변화와 연안 생태환경변화를 추정할 수 있다. 기후변화는 기온 및 수온의 발생빈도 변화와 연안의 생태환경변화를 유발하기 때문에 기온과 수온의 발생 빈도분포 함수를 추정하는 연구가 필요하다. 본 연구에서는 우리나라 연안의 장기 기온자료를 이용하여 빈도분포함수를 추정하였다. 빈도분포 함수는 조위 분포함수 추정(Cho et al., 2003), 수온 분포함수 추정(Jeong et al., 2013) 등에 사용된 Bi-modal 형태의 분포함수를 이용하였으며, 분포함수의 매개변수는 최소자승법으로 최적 추정하였다. 최적 추정된 매개변수는 기온자료의 기본적인 통계정보에 해당하는 평균, 표준편차, 왜도계수와 강한 상관관계를 보이고 있는 것으로 파악되었으며, 통계정보를 이용한 매개변수 추정공식의 RMS 오차는 5% 정도로 파악되었다. 한편 본 연구에서 제시하는 Bi-modal 분포함수는 전체적인 기온 분포양상을 적절하게 표현하고 있으나, 고온 영역의 급격한 빈도감소 양상 및 관측 자료의 분포에서 보이는 작은 첨두 재현에는 한계가 있는 것으로 파악되었다.

핵심용어 : Bi-modal 분포, 발생빈도 분포, 기온, 수온, 기후변화, 환경-생태변화인자

Abstract : Water temperature due to climate change can be estimated using the air temperature because the air and water temperatures are closely related and the water temperatures have been widely used as the indicators of the environmental and ecological changes. It is highly necessary to estimate the frequency distribution of the air and water temperatures, for the climate change derives the change of the coastal water temperatures. In this study, the distribution function of the air temperatures is estimated by using the long-term coastal air temperature data sets in Korea. The candidate distribution function is the bi-modal distribution function used in the previous studies, such as Cho et al.(2003) on tidal elevation data and Jeong et al.(2013) on the coastal water temperature data. The parameters of the function are optimally estimated based on the least square method. It shows that the optimal parameters are highly correlated to the basic statistical informations, such as mean, standard deviation, and skewness coefficient. The RMS error of the parameter estimation using statistical information ranges is about 5 %. In addition, the bi-modal distribution fits good to the overall frequency pattern of the air temperature. However, it can be regarded as the limitations that the distribution shows some mismatch with the rapid decreasing pattern in the high-temperature region and the some small peaks.

Keywords : Bi-modal distribution, frequency distribution, air temperature, water temperature, climate change, environmental and ecological change indicators

1. 서 론

온실가스 배출 증가에 의한 영향으로 가까운 미래의 기후변화가 예상되는 상황에서 기후변화의 지표로 기온이 가장 활발하게 사용되고 있다. 더불어 기후변화로 인한 해양 환경변화도 예상되는 상황에서 해양 환경변화의 지표로 해수면 변

화가 가장 활발하게 사용되고 있으나, 해양 생태환경의 관점에서는 수온이 해양 생태환경을 결정하는 중요한 지표이기 때문에 보다 적절한 지표로 판단된다(IPCC, 2007; Cho & Lee, 2012). 특히 연안의 수온변화는 기온변화와 직접적인 관계가 있기 때문에 수온변화에 의한 연안의 환경변화는 기온변화와 관련지을 수도 있다. 연안의 기온 및 수온변화는 연

*원광대학교 토목환경공학과, 원광대학교 부설 공업기술개발연구소 연구위원(Department of Civil and Environmental Engineering, Wonkwang University, 460, Iksandae-ro, Iksan, Jeonbuk, 570-749, Korea, stjeong@wku.ac.kr)

**한국해양과학기술원, 해양환경보전연구부(Corresponding author, Marine Environments and Conservation Research Division, Korea Institute of Ocean Science and Technology, 787 Hae-an-ro, Sangrok-gu, Ansan, Gyeonggi, 426-744, Korea, Tel: +82-31-400-6318, Fax: +82-31-400-7868, hycho@kiost.ac)

***두산중공업, 기술연구원(Doosan Heavy Industries & Construction, Technology Strategy & Planning Team(Daejeon) Corporate R & D institute)

안에서의 열에너지 수치분석을 통하여 예측이 가능하지만 열 에너지 분석에 사용되는 입력 자료의 불확실성에 따른 오차 전달도 예상되기 때문에 연안의 대기와 해양환경을 대표하는 기온과 수온을 변화의 지표로 이용하는 것도 적절할 것으로 판단된다. 기상자료에 대한 통계적인 분석, 특히 관측 자료에 근거한 발생빈도 분포분석은 기온변화 양상과 수온변화 양상을 파악할 수 있기 때문에 매우 중요한 내용으로 판단되고 있다. McCloy and Dolan(1973)은 미국 대서양 북부 연안의 수온과 기온자료의 관계를 연구하였으나 전반적으로 이 분야의 연구는 미진한 실정이다(Storch & Zwiers, 1999).

계절 변화가 뚜렷한 연안의 기온변화를 정규분포함수로는 표현하는 것은 한계가 있으며(Nese, 1994), 수온과 같이 Bi-modal 분포함수 형태를 가지는 것으로 추정할 수 있으나, 그 추정 모수에서는 차이를 보일 것으로 예상할 수 있다. Cho et al.(2004)과 Jeong et al.(2008)은 조위 분포함수 추정에 Bi-modal 분포함수를 이용하였으며, Jeong et al.(2013)은 수온 분포함수 추정 등에 Bi-modal 형태의 분포함수를 이용하였으며, Grace & Curran(1993)은 일 최대 기온자료의 월별 분포함수로 bi-normal 분포함수를 제시한 바 있다. 다수의 첨두를 가진 이러한 분포함수는 Gaussian 혼합모형(Gaussian mixture model, GMM)의 특정한 형태로 다양한 자료의 분포함수 추정에 이용되고 있다(Huang & Chau, 2008; Grace & Curran, 1993). 이 모형은 혼합하는 함수의 개수 또는 표현하고자 하는 첨두의 개수에 따라 차수를 표현할 수 있기 때문에 Bi-modal 분포함수는 GMM(2) 함수로 간략하게 표현할 수 있다.

본 연구에서는 우리나라 연안에 위치한 주요 8개 지점(인천, 군산, 목포, 여수, 부산, 제주, 포항, 강릉) 기상대의 기온 자료를 이용하여 발생 빈도 분포함수를 추정하였다. 기후분석에 이용되는 적절한 기간에 해당하는 30년 정도의 장기 기온 자료를 이용하여 빈도분포함수를 Bi-modal 분포함수를 가정하여 최적 모수를 추정하고, 그 매개변수와 기온자료의 관계분석을 수행하였다. 또한 Jeong et al.(2013) 등에 의하여 수행된 수온 분포함수와 비교 분석을 수행하였다. Bi-modal 분포함수는 꼬리부분에서 다소 큰 오차를 보이는 한계가 있

으나 정규분포보다 적합한 매개변수를 제시하여 매개변수 분석을 통한 전반적인 발생빈도 변화 양상을 추정할 수 있을 것으로 판단된다.

2. 기온 관측자료

우리나라 연안의 기온자료는 가장 기본적이고 측정이 간단하기 때문에 다양한 기관에서 보유하고 있으나, 품질검정과정을 거쳐 제공되는 장기간의 자료는 기상청 자료이다. 기온 자료의 발생빈도 분포를 추정하기 위해서는 단기간의 자료보다는 장기간의 자료가 필요하기 때문에, 본 연구에서는 장기간의 기온자료가 제공되는 기상청 일평균(daily-mean) 기온 자료를 이용하였다(KMA, 2014). 연구에 사용된 자료는 우리나라 연안 총 8개 지점의 자료로, 각각의 지점은 인천, 군산, 목포(이하 서해안 3개 지점), 여수, 부산, 제주(이하 남해안 3개 지점), 포항, 강릉(이하 동해안 2개 지점)이다. 관측지점의 위치 및 관측기간 등의 기본적인 정보는 Table 1에 제시하였으며, 관측기간에서 결측자료가 아주 적은 비중을 차지하는 1981년부터 2013년까지 33년 동안의 자료를 선택하여 연구에 이용하였다.

3. 분석방법

일단위(daily) 기온자료는 자료의 자기상관영향이 크게 나타나고 있는 시계열 자료(time-series data)에 해당한다. 따라서 어떤 확률분포함수를 추정하는 기본조건에 해당하는 자료의 독립성이 결여되어 있다. 그러나 자연에서 관측되는 대부분의 자료는 어느 정도의 자기상관(auto-correlation)을 가지기 때문에 발생빈도 분포측면에서 분포함수를 추정하게 된다. 분포 함수를 추정하기에 앞서 개략적인 자료 변화 양상 분석을 위하여 자료의 시계열도(time-series plot)를 Fig. 1에 제시하였다. 기온자료가 아주 장기간의 자료이기 때문에 일 또는 주 단위의 작은 시간변동 범위를 보기에 적합하지 않으나 전체적인 변화양상 및 큰 규모의 결측 양상 등은 간단하게 판단할 수 있다. Fig. 1에서 볼 수 있는 바와 같이 전체적으

Table 1. Basic information of the meteorological stations of the air temperature data

Location	Geographical coordinate (N°, E°)	Recording Period		No of data	Data coverage
		Start	End		
Incheon	37° 29' 126° 37'	1904.04.08	2014.03.18	40,155	98.78%
Gunsan	36° 00' 126° 45'	1968.01.01	2014.07.13	16,996	100.0%
Mokpo	34° 49' 126° 23'	1906.01.01	2014.03.18	39,524	99.99%
Yeosu	34° 44' 127° 44'	1942.03.01	2014.07.13	26,068	100.0%
Busan	35° 06' 129° 02'	1904.04.08	2014.03.18	40,156	100.0%
Pohang	36° 02' 129° 23'	1949.01.01	2014.03.18	23,818	99.35%
Gangneung	37° 45' 128° 53'	1912.01.01	2014.07.13	36,355	100.0%
Jeju	33° 30' 126° 31'	1923.05.01	2014.03.18	33,195	100.0%

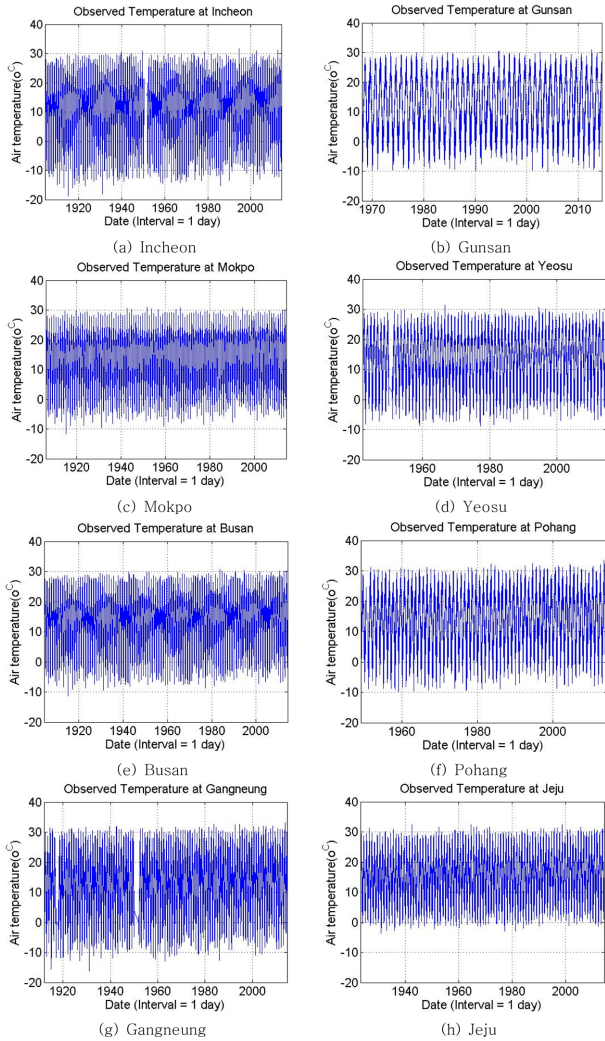


Fig. 1. Time-series plot of the daily air-temperature data at each station.

로 1950년도의 한국전쟁 기간을 제외하고는 특별한 결측 구간이 발생하지 않고 있음을 알 수 있다. 또한 계절변화가 뚜렷하기 때문에 매년 기온의 최대-최소가 반복되는 양상을 보이고 있음을 알 수 있다.

또한 년 최저기온은 인천과 강릉이 상대적으로 다른 지점에 비하여 낮고, 제주는 높게 나타나고 있으며, 년 최고기온은 지점별로 큰 차이를 보이지 않는다.

이러한 기온자료의 시계열 도시는 결측 구간(missing interval), 이상자료(outliers) 등의 정보는 쉽게 파악할 수 있으나, 시간에 따른 자료의 변화추이 및 기본적인 통계정보, 분포 양상 등은 파악할 수 없다.

3.1 도수분포

어떤 확률변수의 분포함수는 다양한 통계적인 추론을 위한 가장 기본적이고 중요한 정보로 자료로부터 추정하는 것이 일반적이다. 이러한 확률밀도함수의 추정에 관한 연구는 통계적으로 모수적(parametric) 방법과 비모수적(non-parametric) 방법으로 구분할 수 있다(Raccine, 2008; Silverman, 1986).

본 연구에서는 대표적인 비모수적 방법에 해당하는 히스토그램을 이용하여 기온자료의 개략적인 분포형태를 분석하고, 빈도분포함수의 빈도자료를 분석에 사용된 자료의 개수로 나누어 확률분포 함수(밀도분포 함수, density histogram)로 계산하였다. 밀도 도수분포는 곡선 아래 면적으로 도수분포를 나눈 도수분포는 면적이 1(one)이 된다.

히스토그램을 이용하여 분포를 추정하는 경우, 분포 형태는 2개의 모수 즉, 계급의 원점(시작점, t_0)과 계급 폭(bin width, h)에 의하여 결정된다. 계급 폭은 ‘smoothing parameter’라고도 부르며, 도수분포의 유연도를 결정하게 된다. 즉, 계급 폭이 작으면, 계급의 높이의 변화가 심하고, 계급 폭이 크면, 계급의 높이의 변화가 작다. 동일 자료인 경우에도 계급 폭의 크기에 따라 상이한 형태의 도수분포도가 생성되고, 적절한 계급 폭을 사용하여야 자료의 분포형태를 쉽게 파악할 수 있다(Martinez et al., 2011). 본 연구에서는 자료 개수와 표본 표준편차를 사용한 관계식 (1)을 이용하여 계급 폭을 결정하였다.

$$h = 3.5 \times s \times n^{-1/3} \quad (1)$$

여기서, n 은 분석 자료의 개수, s 는 표본 표준편차이다.

3.2 분포 함수

우리나라는 여름의 고온과 겨울의 저온이 반복되는 지역으로 도수분포 형태도 저온과 고온의 특정 기온 부근에 빈도수가 집중되는 쌍봉(double-peak) 형태로 표현될 수 있다. 따라서, 본 연구에서는 도수분포도가 쌍봉(double peak) 형태임을 고려하여, 기온자료의 빈도분포 함수($p_T(x)$)로는 평균과 분산이 서로 다른 2개의 정규분포함수를 선형 결합한 형태, 즉 6개의 매개변수를 가진 Gaussian Mixture Distribution(GMD) 함수 형태를 제안하였다. 이 식은 다음 식 (2)와 같이 표현된다.

$$p_T(x) = \alpha_1 f(x; \mu_1, \sigma_1) + \alpha_2 f(x; \mu_2, \sigma_2) \\ = \alpha_1 \left\{ \frac{1}{\sigma_1 \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu_1}{\sigma_1} \right)^2 \right] \right\} + \alpha_2 \left\{ \frac{1}{\sigma_2 \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \mu_2}{\sigma_2} \right)^2 \right] \right\} \quad (2)$$

여기서, α_1, α_2 는 scale parameter로 $\alpha_1 + \alpha_2 = 1$ 의 관계가 있으며, μ_1, μ_2 는 각각 평균으로 첨두의 위치를 결정하는 매개변수이며, σ_1, σ_2 는 각각 표준편차로 각각의 첨두 부근의 자료에 대한 분산정보와 관련이 있는 매개변수로 간주할 수 있다. 이 값이 작을수록 자료의 분포는 집중되기 때문에 꼬리 부분은 급격하게 감소하는 양상을 보이게 된다.

3.3 분포함수의 매개변수 추정

전술한 확률변수의 주 특성값들, 즉, 평균 및 표준편차 등은 가장 일반적인 자료로 관측 정점별로 잘 알려져 있거나, 쉽게 산정할 수 있다. 본 연구에서는 MATLAB 프로그램을

이용하여 8개 연안 정점의 관측자료로부터 평균 및 표준편차 등 주 특성값을 산정하였다. 이들 주 특성값과 기온 분포 매개변수의 관계는 다음과 같은 최소자승법을 이용하여 도출하였다. 즉 기온자료가 가용한 정점(본 연구의 경우, 8개 정점)의 주 특성값(본 연구에서는 평균, 표준편차, 왜곡계수 사용) 행렬을 H_{ij} (i = 정점 인자, j = 주 특성값 인자; known), 기온 분포함수의 매개변수 행렬을 P_{ik} (i = 정점 인자, k = 매개변수 인자; known)라고 정의하고, 분포함수의 매개변수가 주 특성값의 선형조합으로 표현된다고 가정하는 경우 다음과 같은 관계 식 (3)이 성립된다.

$$HC = P \quad (3)$$

여기서, C_{jk} 는 선형조합 계수로 구성된 미지의 행렬(unknown)이다. 위 식을 정리하면, 식 (3)을 이용하여 식 (4)와 같이 최적 선형조합 계수 행렬(C_{jk})을 계산할 수 있다.

$$(H^T H)C = H^T P \rightarrow C = (H^T H)^{-1} H^T P \quad (4)$$

여기서, H^T 는 행렬 H 의 전치행렬이고, $(H^T H)^{-1}$ 는 $(H^T H)$ 의 역행렬이다.

4. 결과 및 토의

연안 8개 지점의 기온 자료를 분석하는 과정은 Fig. 2와 같다.

4.1 자료의 기본적인 통계정보

자료의 분포양상과 관련되어 있는 기본적인 통계정보는 평균, 표준편차, 왜곡도 계수(skewness coefficient)로 제한하여 산정하였으며, Table 2에 제시하였다. 전체적으로 평균은 12~15°C 내외이며, 표준편차는 7.7~9.9°C 범위를 보이고 있으며, 모든 점에서 왜곡도 계수가 음(-)의 값으로 나타나 분포가 우측으로 왜곡된 형태를 보이고 있음을 알 수 있다.

4.2 도수분포도

도수분포도를 작성하기 위해서는 먼저, 서로 인접한 자료값을 집단화하여 전체 자료 집합을 몇 개의 그룹으로 나눈다. 이를 위하여 전체 자료의 범위를 몇 개의 계급(class, bin)으로 나누는 것이 필요하다. 본 연구에서는 식 (1)을 이용하여 최적 계급 폭(bin width, h)을 산정하였으며, 한국 연안 주요 8개 정점의 기온 도수분포도는 Fig. 3과 같다.

Fig. 3에서 볼 수 있는 바와 같이, 빈도분포는 전반적으로 Bi-modal 분포형태를 보이고 전반적으로 2개의 첨두가 보이고 있음을 알 수 있다. 목포 및 제주지점에서는 뚜렷하지는 않으나 3개 정도의 첨두가 보이는 것으로 추정할 수 있다. 한편 저온 영역의 완만한 분포형태에 비하여 고온 영역은 상대적으로 급격하게 빈도가 감소하고 있음을 알 수 있다. 이

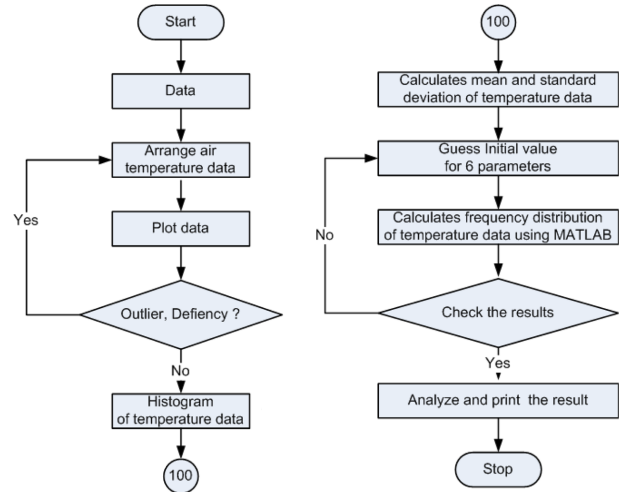


Fig. 2. Flowchart of the air temperature data analysis.

Table 2. Statistical information of the air temperature data in Korean coastal areas

Location	Mean, μ (°C)	S.D., σ (°C)	Skewness, γ
Incheon	12.17	9.90	-0.24
Gunsan	12.86	9.54	-0.13
Mokpo	13.88	8.86	-0.13
Yeosu	14.34	8.42	-0.22
Busan	14.71	8.17	-0.24
Pohang	14.27	8.83	-0.16
Gangneung	13.15	9.13	-0.15
Jeju	15.84	7.72	-0.02

러한 경향은 Jeong et al.,(2013)이 분석한 수온의 빈도분포 양상, 즉 고온 영역의 완만한 감소와 저온 영역의 0°C 지점에서의 급격한 감소와는 상이한 양상을 보이고 있음을 알 수 있다.

4.3 분포 함수의 매개변수 추정

연안 8개 정점의 기온 상대도수분포도를 분석한 결과, Fig. 3에 도시된 바와 같이 대부분 쌍봉(double peak) 형태이다. 따라서 기존의 Single-modal 분포함수를 사용하기에는 한계가 있기 때문에 기온자료의 빈도분포 함수로는 평균과 분산이 서로 다른 2개의 정규분포함수의 합인 형태, 즉 6개의 매개변수를 가진 Gaussian Mixture Distribution(GMD) 함수형태를 제안하였다(식 (2) 참조).

제안된 후보 분포함수의 최적 매개변수는 MATLAB Optimization 도구상자에서 제공하고 있는 함수인 'lsqnonlin'를 이용하여 추정하였다. 비선형 최적화 문제는 모수의 초기값 지정이 중요하기 때문에 본 연구에서는 위치모수는 Histogram 형태에서 보여지는 첨두 영역을 기준으로 다른 영역을 침범하지 않는 범위로 지정하였으며, 분산모수에 해당하는 모수는 간단하게 양수로 지정하였다. 연안 8개 정점의 기온 자료에 대하여 식 (2)의 매개변수를 추정한 결과는 Table 3에 제

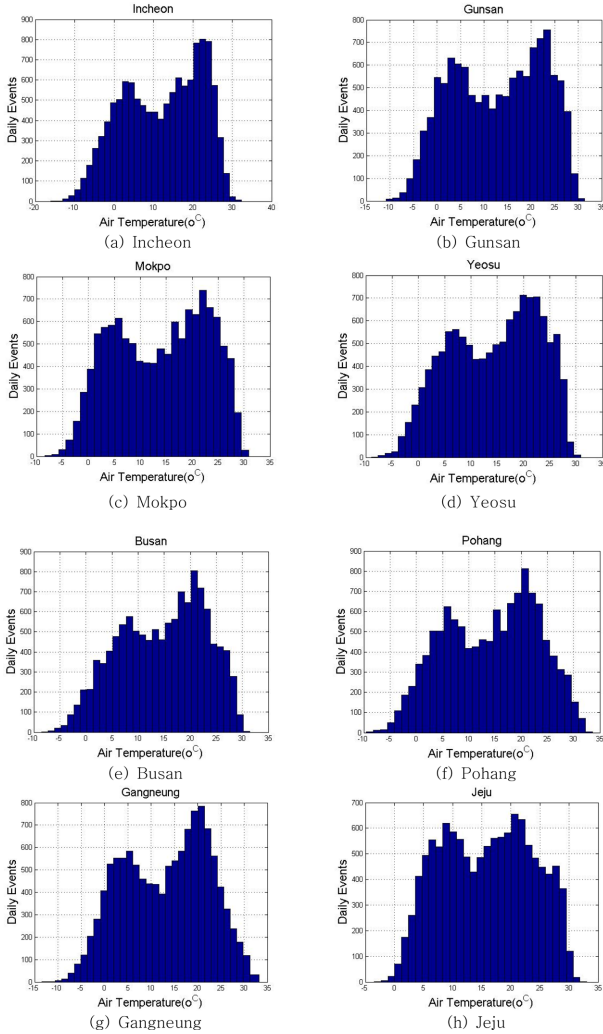


Fig. 3. Histograms of the air temperature data in Korean coastal areas.

시하였다. 히스토그램 분포에서 이미 예상한 바와 같이, 목포 및 제주지점의 표준편차 매개변수를 다른 지점의 매개변수 수치와 비교한 결과 σ_1 은 확연하게 작고, σ_2 는 크게 나타나고 있다. 한편, 관측 자료를 이용한 분포함수(밀도 도수분포도, density histogram), 추정된 매개변수를 사용하여 산정한 GMD 함수형태의 곡선을 Fig. 4에 동시에 도시하였다.

Fig. 4에서 볼 수 있는 바와 같이 정규분포 함수 가정은 자료의 중앙 부분과 꼬리 부분에서 매우 큰 차이를 보이고 있음을 알 수 있으며, 본 연구에서 제안한 Bi-modal(Gaussian Mixture(2) Model) 형태의 분포함수는 뚜렷하게 보이는 대략 2개의 첨두 형태 재현은 우수한 것으로 판단되나, 2개 이상의 첨두가 존재하는 제주 및 목포지점의 분포양상 재현에는 한계가 있는 것으로 판단된다. 또한 꼬리 영역의 급격한 감소 형태 재현에도 한계가 있는 것으로 판단된다. 이러한 문제는 3개 다수의 첨두를 표현할 수 있는 Multi-modal 분포함수를 이용하거나 정규분포만의 Mixture 형태가 아닌 가용한 다른 다양한 분포함수(Log-Normal 분포함수, Rayleigh 분포함수, Weibull 분포함수 등)의 조합을 통하여 최적 분포함수 형태를 도출할 필요가 있을 것으로 판단된다.

Table 3. Optimal parameters of the GMD(2) distribution functions for the daily air temperature data

No	Location	α_1	$\mu_1(^{\circ}\text{C})$	$\sigma_1(^{\circ}\text{C})$	α_2	$\mu_2(^{\circ}\text{C})$	$\sigma_2(^{\circ}\text{C})$
1	Incheon	0.572	5.063	7.526	0.428	21.747	4.173
2	Gunsan	0.487	4.373	5.724	0.513	21.343	5.103
3	Mokpo	0.429	5.105	4.723	0.571	21.044	5.348
4	Yeosu	0.466	6.831	5.389	0.534	21.287	4.620
5	Busan	0.470	7.744	5.495	0.530	21.146	4.572
6	Pohang	0.419	5.654	4.934	0.581	20.599	5.245
7	Gangneung	0.427	4.207	4.983	0.573	19.842	5.194
8	Jeju	0.370	7.837	3.845	0.630	20.888	5.789

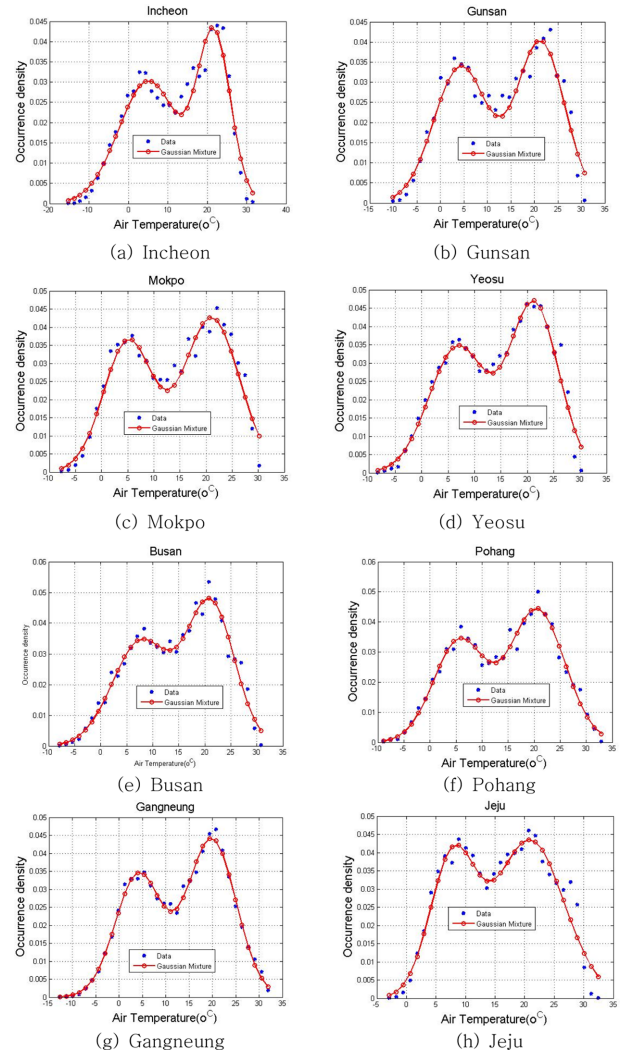


Fig. 4. Comparisons of measured and predicted distribution for air temperature data.

4.4 주 특성 값과 GMD 분포 매개변수의 선형조합 계수 추정

식 (4)를 이용하면 평균 및 분산과 GMD 분포함수의 매개변수의 관계를 표현하는 최적 계수를 추정할 수 있다. 여기서, 주 특성 행렬 H_{ij} 은 (8×3) 행렬(여기서, 8 = 정점의 개수, 3 = 주 특성 값, 평균, 표준 편차 그리고 왜곡도)이며, 매개변수 행렬 P_{ik} 도 (8×6) 행렬(여기서, 6 = 매개변수의 개수)이 되

기 때문에 최적 선형조합 계수행렬 $C_{j,k}$ 은 (3×6) 행렬로 다음 식(5)과 같다.

$$C = \begin{bmatrix} -0.0002 & 0.8141 & -0.1130 & 0.0362 & 0.7235 & 0.2451 \\ 0.0436 & -0.7567 & 0.6425 & 0.0119 & 1.1691 & 0.2875 \\ -0.4553 & -7.4999 & -7.6327 & 0.3914 & -38174 & 5.8136 \end{bmatrix} \quad (5)$$

따라서 GMD 분포함수의 매개변수는 각각 다음과 같은 식 (6)으로 계산할 수 있다. 식 (6)으로 계산되는 매개변수 수치는 자료의 분포함수에서 수학적으로 추정된 최적 매개변수와 자료의 기본적인 통계정보의 선형관계로부터 추정되기 때문에 최적 매개변수를 이용한 경우보다 분포 재현수준이 감소하게 된다. 그러나 분포 매개변수의 물리적인 의미부여가 가능하게 된다.

$$\begin{aligned} \alpha_1 &= -0.0002\mu + 0.0436\sigma - 0.4553\gamma \\ \mu_1 &= 0.8141\mu - 0.7567\sigma - 7.4999\gamma \\ \sigma_1 &= -0.1130\mu + 0.6425\sigma - 7.6327\gamma \\ \alpha_2 &= 0.0362\mu + 0.0119\sigma + 0.3914\gamma \\ \mu_2 &= 0.7325\mu + 1.1691\sigma - 3.8174\gamma \\ \sigma_2 &= 0.2451\mu + 0.2875\sigma + 5.8136\gamma \end{aligned} \quad (6)$$

한편 식 (6)에서 제시되는 가중계수(α_1, α_2)는 다른 매개변수를 이용하여 추정할 수 있는 식이 제시(Ko et al., 2013)되어 있으므로, 그 식을 이용하여 추정할 수도 있다. Ko et al.(2013)이 제시한 식을 이용하여 가중계수를 계산하는 경우, 가중계수 추정정도가 크게 향상되는 것으로 파악되었다.

추정된 확률밀도 함수는 기상대의 기온자료의 통계적 정보를 이용하여 추정된 확률밀도 함수와 비교하여 오차를 분석하였다(Fig. 5 참조). 관측 자료를 이용한 확률밀도 함수와 본 연구에서 개발한 공식(식 (6))을 이용한 확률밀도 함수를 비교한 결과 전반적으로 매우 우수하게 일치하는 양상을 보이

고 있으나, 고온 영역의 꼬리 부분에서는 오차가 크게 증가하고 그 변동범위도 크게 나타나고 있음을 알 수 있다. 특히 목포 및 제주지점의 고온영역 분포양상은 그 재현에 한계가 뚜렷하게 나타나고 있다.

5. 결론 및 제언

본 연구에서 우리나라 연안의 기온자료 빈도분포 함수로 제안한 Gaussian 혼합분포(Gaussian Mixture Distribution, GMD) 함수는 정규분포에 비하여 기온자료의 확률밀도함수로 매우 적합한 것으로 파악되었다. 또한 자료의 주 특성값인 평균, 표준편차, 왜곡도로부터 GMD(2) 분포함수의 매개변수를 산정하여 기온 자료의 빈도분포를 추정하는 공식을 개발하였다. 개발된 추정공식은 RMS 오차가 5% 정도로 관측 자료를 이용한 결과와 우수한 일치를 보이고 있다. 본 공식은 장기간의 기온 자료가 없는 경우, 특정 정점의 기온 자료의 평균, 표준편차 그리고 왜곡도만을 이용하여 기온 분포함수를 간단하고 정확하게 추정하는 데 활용할 수 있다. 또한 장기간의 기온 분포함수 변화 양상을 파악하여 평균 기온만을 이용한 기후변화 양상분석이 아니라 빈도분포를 이용한 기후변화 양상분석을 통하여 보다 다양하고 새로운 결과도출도 가능할 것으로 판단된다.

한편, 본 연구에서 제안한 Gaussian 혼합분포 함수 형태가 기온자료의 빈도분포 함수에서 침투 형태 재현에는 우수한 결과를 보이고 있으나, 분포의 꼬리에 해당하는 부분에서는 재현성이 떨어지고 있다. 이 부분은 꼬리부분의 표현이 다양한 함수형태에 대하여 다각적으로 검토하는 연구가 필요 있을 것으로 사료된다. 특히 제주의 경우, 다른 관측지점에 비하여 고온 자료가 많기 때문에 기온자료의 통계적 정보를 이용하여 확률밀도 함수를 추정하는 경우 특별한 주의가 필요하다.

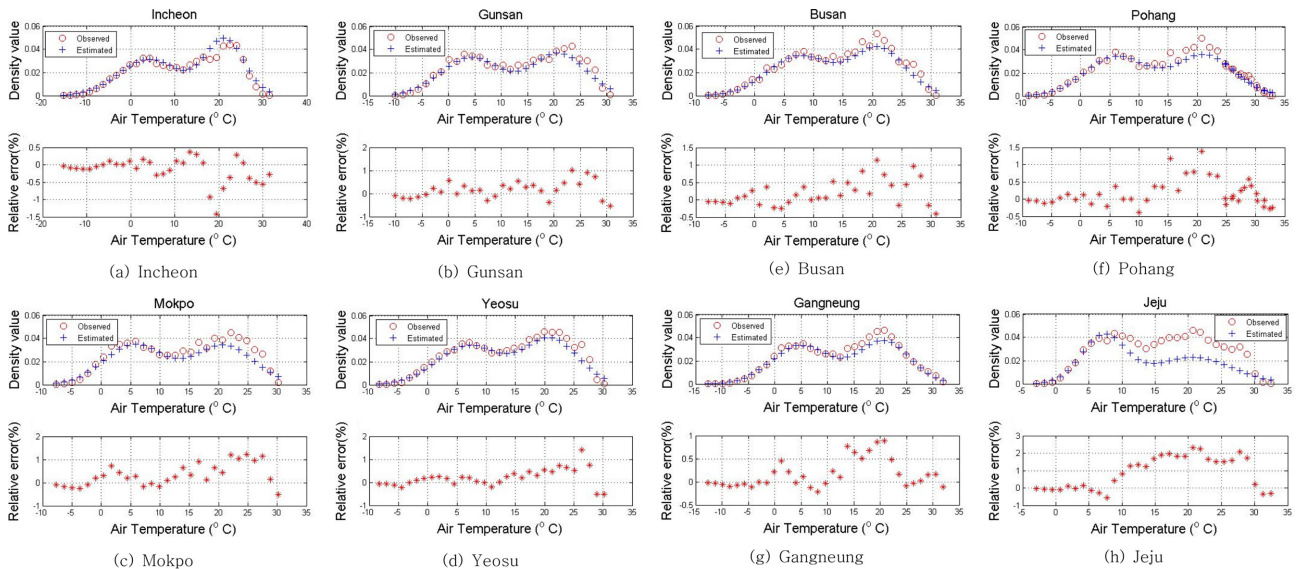


Fig. 5. Comparison plots of the frequency distribution and difference error.

감사의 글

이 논문은 2013학년도 원광대학교의 교비 지원에 의해서 수행되었습니다. 연구비 지원에 감사드립니다.

References

- Cho, H.Y. and Lee, K. (2012). Development of an air-water temperature relationship model to predict climate-induced future water temperature in estuaries, *J. of Environmental Engineering - ASCE*, 138(5), pp.570-577.
- Cho, H.Y., Jeong, S.T., and Oh, Y.M., (2004). Estimation of Probability Density Function of Tidal Elevation Data, *Journal of Korean Society of Coastal and Ocean Engineers*, Vol.16, No.3, pp. 152-161(in Korean).
- Cho, H.Y., Jeong, S.T., Lee K.-H., and Kim, T.H., (2010). Analysis of the Mean and Standard Deviation due to the Change of the Probability Density Function on Tidal Elevation Data, *Journal of Korean Society of Coastal and Ocean Engineers*, Vol.22, No.4, pp. 279-285(in Korean).
- Grace, W. and Curran, E. (1993). A binormal model of frequency distributions of daily maximum temperature, *Aust. Met. Mag.*, 42, 151-161.
- Haldar, A. and Mahadevan, S., (2000). *Probability, Reliability and Statistical Methods in Engineering Design*, John-Wiley & Sons.
- Huang, Z-K. and Chau, K-W. (2008). A new Image Thresholding Method Based on the Gaussian Mixture Model, *Applied Mathematics and Computation*, 205(2), 899-907.
- Intergovernmental Panel on Climate Change (IPCC). (2007). *Climate change 2007: The physical science basis*, Cambridge University Press, Cambridge, UK.
- Jeong, S.T., Cho, H.Y., and Ko D.H., (2012). Development of the Inter-tidal Exposure Duration Formulae Using Tidal Harmonic Constants, *Journal of Korean Society of Coastal and Ocean Engineers*, Vol.24, No.5, pp. 319-325(in Korean).
- Jeong, S.T., Cho, H.Y., Kim, J.D., and Ko D.H., (2008). Estimation of Probability Density Function of Tidal Elevation Data using the Double Truncation Method, *Journal of Korean Society of Coastal and Ocean Engineers*, Vol.20, No.3, pp. 247-254 (in Korean).
- Jeong, S.T., Cho, H.Y., Ko D.H., Oh N.S. and Son K.-P., (2013). Estimation of Probability Distribution Functions for Water Temperature Data in Korean Coasts, *Journal of Korean Society of Coastal and Ocean Engineers*, Vol.25, No.1, pp. 11-19(in Korean).
- Ko D.H., Jeong, S.T., and Cho, H.Y., (2013). Statistical Characteristics of Hourly Tidal Levels around the Korean Peninsula, *Journal of Korean Society of Coastal and Ocean Engineers*, Vol.25, No.6, pp. 365-373.
- Korea Meteorological Administration (2014), Homepage <http://www.kma.go.kr> (in Korean).
- Martinez, W.L., Martinez, A.R. and Solka, J.L.(2011). *Exploratory Data Analysis with MATLAB*, Chapter 9, Chapman & Hall/CRC.
- McCloy, J.M. and Dolan, R. (1973). Water-Air Temperature relationship along Coastal North Carolina, U.S.A., *Geografiska Annaler: Series A, Physical Oceanography*, 55(3), 117-121.
- Ministry of Maritime Affairs and Fisheries (2005). *Harbor and fishery design criteria* (in Korean).
- Nese, J.M. (1994). Systematic biases in manual observations of daily maximum and minimum temperature, *J. of Climate*, 7, 834-842.
- Racine, S.R. (2008). *Nonparametric Econometrics: A Primer*, *Foundations and Trends in Econometrics*, 3(1), 1-88. (2011.5.27. <http://socserv.mcmaster.ca/~racine/ECO0301.pdf>)
- Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*, *Monographs on Statistics and Applied Probability*, Chapman & Hall.
- Storch, H.v. and Zwiers, F. W. (1999). *Statistical Analysis in Climate Research*, Cambridge Univ. Press.

원고접수일: 2014년 9월 25일

게재확정일: 2014년 10월 16일