

층화추출법에 의한 해양기상환경의 표본추출 타당성 연구

한송이* · 조용진*

*동의대학교 조선해양공학과

Feasibility Study on Sampling Ocean Meteorological Data using Stratified Method

Song-I Han*, Yong-Jin Cho*

*Dept. of Naval Architecture and Ocean Engineering, Dong-Eui University, Busan, Korea

KEY WORDS: Ship infrared signature 함정적외선신호, Ocean Meteorological 해양기상환경, Simple Random Sampling 단순임의추출방법, Stratified Sampling 층화추출방법

ABSTRACT: The infrared signature of a ship is largely influenced by the ocean environment of the operating area, which has been known to cause large changes in the signature. As a result, the weather condition has to be clearly set for an analysis of the infrared signatures. It is necessary to analyze meteorological data for all the oceans where the ship is supposed to be operated. This is impossibly costly and time consuming because of the huge size of the data. Therefore, the creation of a standard environmental variable for an infrared signature research is necessary. In this study, we compared and analyzed sampling methods to represent ocean data close to the Korean peninsula. In order to perform this research, we collected ocean meteorological records from KMA (Korea Meteorological Administration), and sampled these in numerous ways considering five variables that are known to affect the infrared signature. Specifically, a simple random sampling method for all the data and 1-D, 2-D, and 3-D stratified sampling methods were compared and analyzed by considering the mean square errors for each method.

1. 서 론

스텔스기술은 함정이나 항공기 등과 같은 군수장비들에 상대의 레이더나 적외선 탐지기, 음향 탐지기 및 육안으로부터 모든 탐지 가능한 신호를 제거 또는 감소시키는 기술이다. 이는 장비에 따라서 레이더(Radar cross section, RCS), 적외선(Infra red, IR), 수중방사소음(Underwater radiation noise, URN) 등과 같은 다양한 분야가 있으며, 그 중 함정의 적외선 스텔스 성능은 함의 생존성을 높이기 위해 필수적으로 갖추어야 할 특수성능 중 하나이다.

함정의 적외선신호는 환경으로부터 입사된 신호가 표면에 반사되어 나가는 신호와 표면의 특성과 온도에 의한 표면 자체의 방사신호의 합으로 나타나며, 표적으로부터 방사된 적외선신호는 관측기 위치까지 대기환경 조건에 의한 감쇠과정을 거쳐 관측기로 최종신호(Radiosity)가 도달하게 된다(Cho, 2009). 이와 같이 적외선신호는 표면온도를 결정하는 열전달 과정과 배경신

호와 함정표면으로부터의 반사와 열방사신호의 대기감쇠 과정의 영향을 받는다. 이로 인해 태양이 있는 주간의 함정 방사신호는 환경에 많은 지배를 받게 되어 적절한 해양기상환경이 주어지지 않으면 올바른 함정의 적외선 신호를 얻지 못하게 된다. 함정의 적외선신호는 적외선의 특성 때문에 운용하는 해양기상환경에 매우 영향을 많이 받기에 신호 값의 변화가 크다. 그렇기 때문에 함정 적외선신호 또는 미사일에 의한 피탐 거리에 대한 해석을 위해서는 함정이 운용될 해양의 기상조건이 명확하게 설정이 되어야 한다(Cho, 2009).

현재 함정의 적외선신호해석에 사용되는 기상환경조건은 현재 국내에서는 최상 주간/야간, 최악 주간/야간 조건의 4가지에 대해서 연구가 수행되어져 왔다. 그러나 이 조건들은 급격하게 변화하는 해양환경들을 대표할 수가 없다. 그러나 모든 실제 관측환경으로부터 해석이 이루어지기에는 시간과 비용 문제로 인하여 어려움이 있으므로, 본 연구로부터 실제 해양기상환경으로부터 표본을 추출하여 해양환경을 설정할 필요가 있다.

Received 13 February 2014, revised 27 May 2014, accepted 11 June 2014

Corresponding author Yong-Jin Cho: +82-51-890-2593, cjydeu@deu.ac.kr

© 2013, The Korean Society of Ocean Engineers

It is noted that this paper is revised edition based on proceedings of KAOST 2013 in Jeju.

해양의 기상환경을 관측하기 위해 기상청에서는 등표기상관측장비와 해양기상부이를 이용하여 총 18개의 지점에서 측정을 하고 있다. 함정의 적외선신호에서는 열전달과 대기감쇠에 직접적으로 영향을 미치는 해양과 대기환경변수가 중요하다. 따라서 관측되는 자료 중에서 특히 적외선신호에 영향을 미치는 변수는 수온, 기온, 습도, 풍향 및 풍속으로 5가지 변수로 들 수가 있으며, 변수 간에는 계절에 따라 종속적인 관계에 있다 (Kim, 2012).

따라서 본 연구에서는 함정개발과정에서 올바른 적외선신호 연구를 위하여 실제 관측된 해양기상 자료로부터 함정의 적외선 신호 연구에 활용 가능하도록 해양기상환경의 표본추출 방법에 대한 연구를 하였다. 기상청으로부터 축적된 한반도 연근해역의 해양기상환경에 대한 자료를 수집하였고, 측정되는 자료에는 적외선신호에 영향을 미치는 5가지 변수에 대해서 고려하여 여러 방법의 표본추출을 수행을 하였다. 그 방법으로는 전체 데이터로부터 단순임의추출, 전체 데이터를 변수에 따라 층을 나누는 1차원적, 2차원적, 3차원적인 층화추출을 수행하였다. 그리고 추출된 표본의 신뢰성은 모집단(Population)과 표본(Sample)의 평균계급오차를 이용하여 비교하여 검증하였다.

2. 연구방법

2.1 기상청 자료의 특성

함정의 적외선신호는 수온, 기온, 습도, 풍향 및 풍속과 같은 환경변수에 영향을 받는다. 또한 이와 같은 변수들에는 계절(또는 월간)에 따라 변화하는 서로 종속적인 특성을 지니고 있다. 본 연구를 수행하기 위해 기상청으로부터 바다관측 자료를 확인하였다. 기상청에서는 해양기상현상을 관측하기 위해서 등표기상관측장비와 해양기상부이를 운용하고 있으며, 주기적으로 해양기상 및 해양환경의 계측은 기상청이 주관되어 국내해안지역 기상관측소에서 자료를 관측하여 관측데이터들은 관련연구 등을 위한 자료로 일반에 제공되고 있다. 두 관측 장비 모두 기본요소인 파고(유의, 최대), 파주기, 조위, 풍향, 풍속, 기압, 기온, 수온 등이 관측이 된다. 그러나 등표기상관측장비에서는 함정의 적외선신호 환경에 필요한 다섯 가지 변수 중 한가지인 습도에 대한 관측이 이루어지지 않을 뿐만 아니라, 해안가의 모래사장의 유무, 해안에 인접한 숲의 유무, 해안절벽 등의 해안 위치의 형태는 주변의 해상환경 및 기상환경에 많은 영향을 미치므로 육상에서의 기상자료의 넓은 해양의 기상환경을 대표하기에는 신뢰성에 문제가 있다고 판단하였다.

그렇기 때문에 본 연구에서는 우리나라 연근해의 해양기상환경을 정의하기 위해서 국내 기상청으로부터 수집된 부이자료를 활용하였다. 기상청에서는 해양기상현상을 관측하기 위해서 8개의 부이를 국내 연안에 설치 운영 중이다. 해양기상부이는 해수면에서 해양기상현상을 각종 관측센서로 측정하고, 측정값을 일정한 물리량으로 변환, 처리한 후에 사용가능한 통신장비를 이용하여 관측 자료를 전송하는 장비이다. 해양기상부이는 관측 자료가 없는 해상과 악천후로 관측이 어려운 지역에서도 관측이 가능하며, 기상분석과 예보에 매우 유용한 정보를 생

산하는 첨단 해양관측 장비이며, 장기간 누적된 자료는 매우 유용한 분석 자료로 활용가능하다(Kim, 2012).

이 연구에서는 '동해' 부이(동해시 동쪽 80km 해상, 37.53°N/130.0°E)의 데이터를 활용하였으며, 동해 부이에서는 2001년 5월부터 해양환경에 대한 관측이 이루어졌다. 10년 치 이상의 데이터를 고려하여 적외선신호 연구를 수행하면 다양한 환경조건에 대한 연구가 수행이 되겠지만 그 방대한 관측데이터로 인해 직접 해석하는데 걸리는 시간을 고려하면 시간과 비용적인 측면에서는 불가능하다. 그렇기 때문에 표본추출이 중요하게 작용을 한다. 그러나 이 연구에서는 표본추출 방법에 대한 연구이므로 약 1년 치 데이터인 7,900여개의 모집단을 형성하였고 함정의 적외선신호 해석 시간을 고려하였을 때, 100개의 표본을 추출하는 것을 수행하였다.

2.2 표본추출방법

현실적인 여건상 모집단에 속한 전체 데이터를 적용하기 위해서는 일부 데이터를 뽑으면 된다. 이 과정을 표본추출이라고 한다. 표본추출의 기본은 랜덤화이기에 잘 설계된 추출은 난수표 또는 컴퓨터를 통한 무작위의 수(Random number)를 이용하는 등 항상 우연의 모습을 지닌다. 그러나 이러한 확률기법의 활용만으로는 표본의 대표성이 충분히 보장되지는 않는다. 그리고 표본추출은 아무리 잘 설계되고 잘 수행되더라도 모수(Population parameter)의 한 추정치인 표본통계량(Sample statistics)을 제공해줄 뿐이다. 같은 모수에 대한 추정치 일지라도 추출된 표본이 다르면 표본통계량도 달라지며, 표본오차 또는 표본추출오차(Sampling error)라 불리는 차이가 당연히 존재하게 된다. 일반적으로 표본의 크기가 커지면 오차발생의 기회는 적어진다. 그러나 규모가 큰 표본이더라도 빈약한 설계나 결함 있는 자료 수집을 보완해 줄 수는 없다.

표본추출의 기본은 랜덤화(Randomize)이다. 이는 단순임의추출방법이 가장 기본적인 추출방법이라는 것을 설명을 한다. 해양기상환경 데이터는 변수 간에 종속적인 특성을 지니는 특징이 있어 어떠한 추출방법이 적합한가에 대해서 여러 가지 방법의 표본추출을 고려하였다.

전체 해양기상환경에 대해서 적외선신호를 수행하는 것이 필요로 하지만 방대한 데이터로 인해 이를 수행하는 것은 시간과 비용적인 측면에서 불가능하다. 그러므로 본 연구에서는 전체 해양기상환경의 데이터를 포함하고 있는 모집단으로부터 적외선신호 연구가 가능한 범위의 표본을 추출하고자 하였다.

먼저 단순임의추출방법(Simple random sampling)을 통하여 표본을 추출하였다. 단순임의추출방법은 표본추출법 중 가장 간단한 방법이다. 이는 모집단의 각각의 요소들이 표본으로 선택될 가능성이 같게 되는 표본추출 방법이며, 크기가 N 인 모집단에서 n 개의 표본을 추출할 경우 표본들의 추출확률은 모두 동일하게 해주는 추출방법을 말한다(Kim et al., 2006). 단순임의추출방법은 가장 단순하고 상대적으로 자료를 분석하기 쉽지만 모집단을 대표할 수 있는 표본이 누락될 가능성이 있어 모집단을 대표하기 힘들다는 단점이 있다. 이 방법으로부터 시간(연월일시)을 기준으로 7900여개의 모집단으로부터 100개의 표

본을 추출하였다.

두 번째 방법은 층화추출법을 이용하였다. 층화추출방법(Stratified sampling)은 크기가 N 인 모집단을 중복되지 않는 여러 개의 층으로 나눈 후에 각 층에서 단순임의추출방법을 통하여 n 개의 표본을 얻는 방법이다(Lee and Lee, 2002; Son and Min, 2011). 층화에 앞서 층을 나눌 때 누적분포함수(Cumulative distribution function, CDF)를 이용하였다. 누적분포함수는 어떤 확률분포에 대해서 확률변수가 특정 값보다 작거나 같은 확률을 나타낸다. 그리고 누적이라는 단어가 붙은 이유는 이 함수가 확률밀도함수의 적분 값의 의미를 가지기 때문이다. 누적분포함수는 그 특성상 순증가 함수 형태를 띠고 있기 때문에, 역함수를 통해 확률분포함수를 따르는 샘플링(Sampling)을 하는 것이 용이하다(Jeon et al., 2012). 먼저 1차원적인 층화추출을 하였는데, 함정의 적외선신호에 가장 영향을 많이 미치는 기온을 중심으로 누적분포함수에서 Fig. 1과 같이 25%, 50%, 75%를 기준으로 4등분하여 4개의 층으로 나눈 층화추출을 수행하였다(David and Kim, 2013). 이때, 각 층 내에 있는 데이터의 개수는 같기 때문에 각 층마다 동일한 개수로 표본을 추출하였다.

세 번째 방법 또한 층화추출을 이용하였다. 이 표본은 상관분석을 수행하여 상관도가 높은 두 개의 변수를 중심으로 각각 4등분씩 하여 총 16개의 층으로 분할한 2차원적 층화추출을 수행하였으며, 상관분석의 결과는 Table 1과 같다.

Table 1의 상관분석은 피어슨상관분석으로부터 나온 결과 값이다. 피어슨상관분석결과는 -1~1사이의 값을 가지게 되며 절대값 0.7 이하면 거의 상관이 없다고 본다고 알려져 있다.

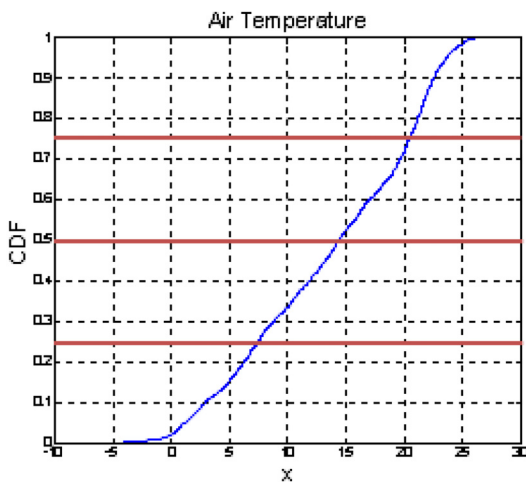


Fig. 1 Division of Air Temperature's CDF

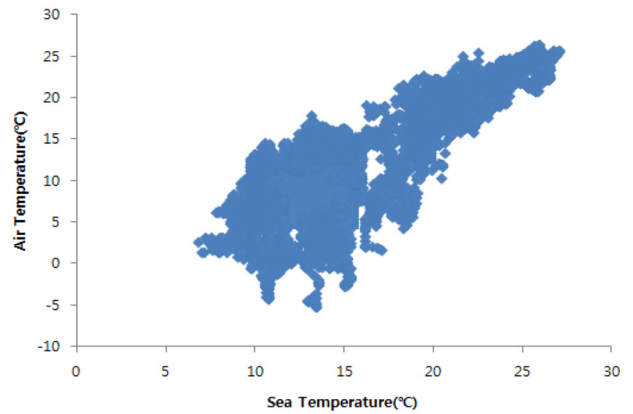


Fig. 2 Scatter Diagram of Air Temperature and Sea Temperature

상관분석의 결과로부터 가장 상관도가 높다고 보이는 수온과 기온의 산점도는 Fig. 2와 같으며, 산점도를 확인하면 정확하게 양의 기울기를 가지고 있어 두 변수가 상관이 있다고 볼 수 있다.

Table 1에서 상관도가 가장 높은 두 변수는 기온과 수온이며, 기온과 수온을 중심으로 각각 25%, 50%, 75%로 분할하여 4등분씩 하였다. 이렇게 분할하는 경우에 기온과 수온의 1차원적인 측면에서 보면 각 층 내에서는 같은 양의 데이터를 갖게 되지만 2차원적인 측면에서 보면 층 내의 데이터의 양은 달라진다. 그렇기 때문에 각 층별로 동일한 개수의 표본을 추출하면 데이터가 한쪽으로 치우치게 되는 경우가 발생하기 때문에 각 층별로 비율에 맞추어 표본 100개를 추출하였다.

마지막으로 네 번째 표본추출방법은 3차원적 층화추출이다. Table 1에서 보면 0.7 이하의 값을 갖고 있어 상관도가 거의 없다고 보이지만 그나마 높은 값을 가진 변수인 습도까지 포함하여 수온, 기온, 습도를 중심으로 4등분씩 나누어 총 64개의 층으로 분할한 3차원적인 층화추출이 이루어졌다. 이때 또한 각 층 내에 있는 데이터의 개수는 다르기에 각 층별로 비율에 맞추어 표본 100개를 추출하였다.

좀 더 고차원적인 층화추출을 수행하려 하였으나, 데이터의 정확성에 대한 보장이 없고, 시간이 많이 소요되기 때문에 3차원적인 층화추출까지만 수행하였다.

이러한 방법으로부터 얻은 표본들에 각각 Sample 1~4까지 명칭을 부여하였으며, 다음의 표와 같다.

2.3 표본의 오차분석 방법

표본과 모집단의 비교를 위해서 먼저 누적분포함수(CDF)와 같은 확률분포를 고려하였다. 이 누적분포함수를 이용하면 어

Table 1 Result of Correlation Analysis among the Five Variable

	Wind Direction	Wind Speed	Humidity	Air Temperature	Sea Temperature
Wind Direction	1				
Wind Speed	0.167	1			
Humidity	-0.334	-0.080	1		
Air Temperature	-0.326	-0.356	0.469	1	
Sea Temperature	-0.234	-0.206	0.296	0.851	1

Table 2 Method of the Samples

Sample No.	method
Sample 1	Simple Random Sampling
Sample 2	1-D Stratified Sampling
Sample 3	2-D Stratified Sampling
Sample 4	3-D Stratified Sampling

떠한 방법이 모집단과 가장 유사한 결과를 가지는지에 대해서 그래프로부터 육안으로 어느 정도 판단이 가능하지만 그 방법이 신뢰성이 있다고 하기는 어렵다. 그렇기 때문에 정량적 방법을 사용하고자 표본의 검증방법을 위해서 표본오차의 개념을 이용하였다. 표본오차란 표본추출 과정에서 어쩔 수 없이 발생하는 통계적 오차를 말하며 이는 모집단과 표본 사이의 불일치 정도를 나타낸다(Son and Min, 2011).

본 연구에서는 모집단을 평균으로 보고 표본을 데이터라고 가정하여 분산 식을 적용하여 새로운 식을 만들었다.

$$Mean\ Square\ Error = \frac{\sum_{i=1}^n (CDF_{pi} - CDF_{si})^2}{n} \quad (1)$$

표본의 신뢰도를 검증하기 위하여 위와 같은 식을 만들었으며 평균제곱오차(Mean square error)라고 명명하였다. 위의 식을 살펴보면 n 은 표본의 개수, $(CDF_{pi} - CDF_{si})$ 는 각각의 변수 내에서 누적분포함수 기준으로의 모집단과 표본의 차이를 나타

낸다. 제곱을 하지 않았을 때 (+)값과 (-)값의 상쇄작용으로 정확한 값이 나오지 않을 것을 고려하여 제곱을 하였으며, 값이 너무 커지는 것을 고려하여 표본의 개수만큼 나누어 주었다. 이러한 분석방법은 분산이 작을수록 평균 근처에 밀집해 있다는 것처럼 평균제곱오차의 값이 작을수록 모집단과 가깝다는 것을 알 수 있다.

3. 표본추출 결과 및 분석

3.1 표본추출 결과

4가지방법의 표본추출의 결과는 Fig. 3의 그림과 같다. 순서대로 풍향과 풍속, 습도, 기온 그리고 수온의 누적분포곡선이며 다음의 결과 그래프는 MathWorks사에서 개발한 프로그램인 MATLAB을 사용하였다. 그리고 누적분포함수(CDF)가 모집단과 표본의 차이를 한눈에 잘 보여주어 그래프로 비교하였다.

이상의 그래프로부터 단순임의추출의 결과는 단순임의추출의 단점처럼 모집단을 대표할 수 있는 표본이 누락되어 모집단의 특성을 비교적 대표하지 못한다는 것을 알 수 있었다. 그로인해 단순임의추출보다는 층화추출의 결과가 더 모집단과 유사하다고 판단하였다. 하지만 육안으로 확인하기에는 층화추출 방법 중에 어떠한 방법이 더 효율적인지는 판단이 어려워 각 표본의 평균제곱오차를 계산하여 수치적으로 검증을 하였다.

3.2 표본의 비교분석

표본들의 비교분석을 위해서 모집단을 기준으로 표본들이 모집단과 얼마나 떨어져 있는가를 확인하였다. 이를 위해서 평균

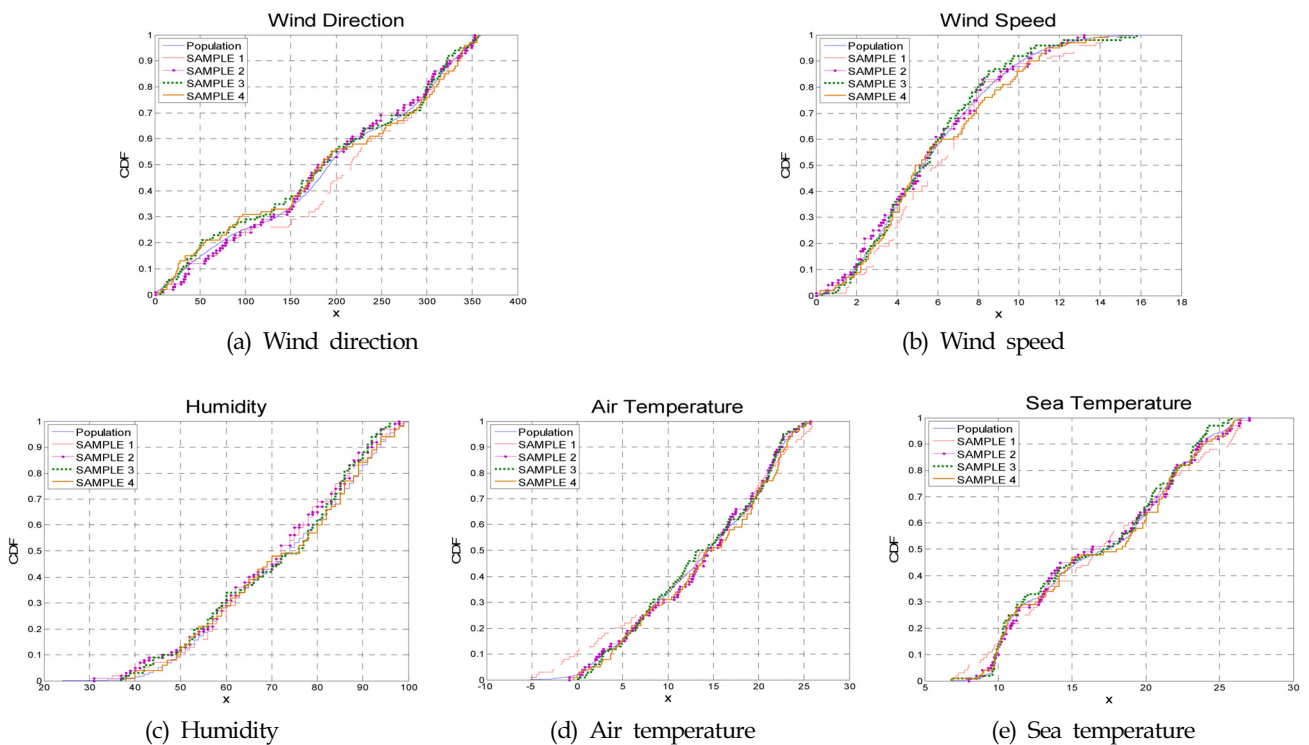


Fig. 3 Comparison of all CDF results

Table 3 Results of the Comparisons of Samples

Mean Square Error	Sample 1	Sample 2	Sample 3	Sample 4
	Random Sampling	1-D Stratified Sampling	2-D Stratified Sampling	3-D Stratified Sampling
Wind Direction	215.78	99.37	86.87	156.35
Wind Speed	0.667	0.167	0.076	0.218
Humidity	4.22	6.61	2.76	3
Air Temperature	1.594	0.307	0.149	0.244
Sea Temperature	0.473	0.246	0.139	0.156

Table 4 Results of the Sample by 2-D Stratified Sampling

Number	Wind Direction (deg)	Wind Speed (m/s)	Humidity (%)	Air Temperature (°C)	Sea Temperature (°C)
1	328	3	53	10.1	18.1
2	299	8.1	50	4.8	9.6
3	297	9.7	77	8.1	9.9
4	41	3.6	77	18.8	19.9
5	321	8.3	64	1.7	11.2
⋮	⋮	⋮	⋮	⋮	⋮
97	347	6.6	66	0.1	10.4
98	193	6.5	86	11.5	9.9
99	129	0.3	42	4.9	9.7
100	345	15	68	5.6	14.2

제공오차로부터 비교분석을 하였으며, 평균제공오차를 이용한 표본의 결과는 Table 3와 같다. 이는 각각의 단위에 맞추어 계산된 값이며, 풍향의 값이 비교적 큰 이유는 풍향의 값의 범위가 크기 때문이다.

앞에서 말한 듯이 분산의 크기가 작을수록 평균에 근접해 있기에 평균제공오차의 값이 작을수록 모집단과 근접하다고 볼 수 있다. 위의 표에서 보면 Sample 3번의 결과 값이 가장 적은 것을 알 수 있다. 그러므로 상관분석을 통하여 상관도가 높은 두 가지의 변수를 중심으로 나는 2차원적 층화추출의 결과가 가장 우수한 것을 알 수 있었으며, 2차원적 층화추출인 Sample 3의 데이터는 Table 4와 같다.

4. 결 론

본 논문에서는 층화추출법으로부터 해양기상환경의 표본추출 방법의 타당성에 대한 연구로서 다양한 표본추출 방법에 대하여 살펴보았다. 연구에 필요한 자료의 경우 기상청으로부터 관측된 동해부이의 해상기상환경자료를 받았으며, 동해부이의 자료는 2001년부터 관측된 데이터이기에 그 데이터의 양은 매우 방대하다. 적외선신호 연구를 위해서 모든 해양환경의 조건을 대상으로 해석을 수행하는 것이 가장 좋은 방법이지만, 적외선신호해석을 하는데 걸리는 시간과 비용을 고려하면 불가능하므로 본 연구는 실제 해양환경에 가장 근접한 표본을 도출해 내 고자 표본추출 방법에 대한 연구를 하였으며 다음과 같은 결과

를 도출하였다.

(1) 표본의 비교분석모집단에 대한 단순임의추출 결과는 대표성을 가지지 못하는 표본의 추출 가능성에 의해 모집단의 특성을 따르지 못하는 것을 확인하였다. 이는 해양기상환경데이터와 같은 변수간의 독립성이 없는 데이터에는 이 방법이 적합하지 않다는 결과를 얻었다.

(2) 모집단의 특성을 충실히 따르도록 하기 위한 층화추출 방법은 적외선신호에 가장 큰 영향을 주는 것으로 알려진 대기온도를 층화한 1차원적 층화방법과 상관분석 결과를 통해 상관도가 높은 2가지 변수들에 의한 2차원 층화방법 및 3가지 변수에 의한 3차원 층화추출을 수행하였다. 이에 대한 비교분석 결과, 상관성이 있는 2개의 변수를 고려한 2차원적 층화추출 결과가 모집단과 가장 유사한 결과를 보여주었다.

(3) 위의 결과를 토대로 볼 때, 변수 간에 종속적인 특성을 갖는 데이터의 경우 상관분석 후에 상관성이 있는 변수들의 수에 맞는 차원적인 층화추출을 하는 방법이 가장 우수한 방법으로 판단된다.

이러한 표본추출 방법론에 대한 연구는 차후의 함정의 적외선신호 해석에서 환경기준을 제안하는 기초자료로 충분히 활용할 수 있을 것으로 보인다. 또한 본 연구에서는 동해지점의 데이터를 이용하여 분석을 수행하였지만 더 나아가 한반도 전 해역의 데이터를 이용하여 해양기상환경의 분석이 필요할 것으로 사료된다.

후 기

본 연구는 동의대학교 교내일반연구과제(2012AA196) “적외선 신호해석을 위한 한반도 해양 기상환경의 통계분석” 결과의 일부임을 밝힙니다.

References

- Cho, Y.J., 2009. A Study on Measuring Procedure and Analysis Technique of Ship Infrared Signature at Sea. The Korea Society of Ocean Engineers, 23(3), 53-58.
- David, A.V., Kim, Y.S., 2013. Climatic data analysis for input to ShipIR. the SPIE Defence, Security, and Sensing, Baltimore, Maryland USA, 8706.
- Jeon, C.H., Jung, M.G., Lee, H.S., 2012. Applied Statistics for Engineers. Hong-Reung. Republic of Korea.
- Kim, W.C., Kim, J.J., Park, B.W., Park, S.H., Song, M.S., 2006. General Statistics. YoungjiMunhwa, Republic of Korea.
- Kim, Y.S., 2012. A Study on the Infrared Signature of a Naval Ship Under the Marine Climate. Journal on the Society of Naval Architects of Korea, 49(3), 264-272.
- Lee, H.Y., Lee, P.Y., 2002. Sampling Survey Introduction. Kyowoosa. Republic of Korea.
- Son, B.Y., Min. M.S., 2011. Probability and Statistics. Hantimedia. Republic of Korea.