

2020

X-AI 기술을 이용한 기후변화예측  
자료기반모형 시범개발  
- 한반도 해역 고수온  
현상 연구 -

[www.kiost.ac.kr](http://www.kiost.ac.kr)

A pilot development of the data driven model for ocean climate  
change simulation using X-AI technology  
- Study on the marine heat waves in the vicinity of Korean  
Peninsula -

# 제 출 문

한국해양과학기술원장 귀하

본 보고서를 “X-AI 기술을 이용한 기후변화예측 자료기반모형 시범개발 - 한반도 해역 고수온 현상 연구-”과제의 최종보고서로 제출합니다.

2021. 02. 28

총괄연구책임자 : 권민호

참 여 연 구 원 : 김성대

“ : 김진아

“ : 박혁민

“ : 이강진

“ : 최유미

## 보고서 요약서

|  |           |  |                                |             |   |
|--|-----------|--|--------------------------------|-------------|---|
| 과제고유<br>번호   | PE        | 해당단계<br>연구기간   | 2020. 02. 01 -<br>2020. 12. 31 | 단계 구분       | 단년도   |
| 연구사업명  | 중사업명      | 한국해양과학기술원 주요사업   |                                |             |   |
|  | 세부사업명     | 미래선도사업   |                                |             |   |
| 연구과제명  | 대과제명      | X-AI 기술을 이용한 기후변화예측 자료기반모형 시범개발<br>- 한반도 해역 고수온 현상 연구-         |                                |             |   |
|  | 세부과제명     |  |                                |             |   |
| 연구책임자  | 권민호       | 해당단계<br>참여연구원수   | 총 : 7 명<br>내부: 7 명<br>외부:    명 | 해당단계<br>연구비 | 정부: 243,000 천원<br>기업:        천원<br>계 : 243,000 천원 |
|  |           | 총연구기간<br>참여연구원수  | 총 : 7 명<br>내부: 7 명<br>외부:    명 | 총<br>연구비    | 정부: 243,000 천원<br>기업:        천원<br>계 : 243,000 천원 |
| 연구기관명<br>및<br>소속부서명  | 한국해양과학기술원 |  | 참여기업명                          |             |   |
| 국제공동연구<br>위탁연구   |           |  |                                |             |   |
| 요약   |           |  |                                | 보고서<br>면수   | 108   |
| <p>설명가능 인공지능(X-AI) 기술을 이용하여 한반도 주변해역 고수온 이벤트의 예단인자를 찾고, 자료기반모형 개발을 통해 예측성을 진단하고 평가함</p> <ul style="list-style-type: none"> <li>○ 해양기후변화 관련 자료수집 및 전처리 <ul style="list-style-type: none"> <li>- 현장관측자료와 인공위성자료 수집</li> <li>- 기상, 해양 재분석자료 수집</li> <li>- 머신러닝을 위한 자료 추출 및 전처리</li> </ul> </li> <li>○ 한반도 주변해역 고수온 이벤트 분석 <ul style="list-style-type: none"> <li>- 관측자료 기반 한반도 주변 고수온 이벤트 정의</li> <li>- 머신러닝/딥러닝 기술을 이용한 고수온 이벤트의 객관적 분류</li> <li>- 머신러닝/딥러닝 기술을 이용한 고수온 현상의 선행인자 분석</li> </ul> </li> <li>○ 설명가능 인공지능(X-AI) 기술 이용 한반도 고수온 현상 모델링 <ul style="list-style-type: none"> <li>- 분류된 고수온 이벤트와 선행인자를 이용한 기계학습기반 예측모형 개발</li> <li>- 설명 가능한 인과관계 학습 추론 기술을 이용한 한반도 주변해역 고수온 현상 예측성 진단 및 평가</li> </ul> </li> </ul> |           |  |                                |             |   |
| 색인어<br>(각 5개 이상)   | 한 글       | 고수온, 기후변화, 설명가능 인공지능, 딥러닝, 자료기반 모형                             |                                |             |   |
|  | 영 어       | Marine heat waves, climate change, X-AI, DL, data-driven model |                                |             |   |

# 요 약 문

## I. 제 목

X-AI 기술을 이용한 기후변화예측 자료기반모형 시범개발  
- 한반도 해역 고수온 현상 연구-

## II. 연구개발의 목적

설명가능 인공지능(X-AI) 기술을 이용하여 한반도 주변해역 고수온 이벤트의 예단인자를 찾고, 자료기반모델 개발을 통해 예측성을 진단하고 평가

## III. 연구개발의 내용 및 범위

- 해양기후변화 관련 자료수집 및 전처리
  - 현장관측자료와 인공위성자료 수집
  - 기상, 해양 재분석자료 수집
  - 머신러닝을 위한 자료 추출 및 전처리
- 한반도 주변해역 고수온 이벤트 분석
  - 관측자료 기반 한반도 주변 고수온 이벤트 정의
  - 머신러닝/딥러닝 기술을 이용한 고수온 이벤트의 객관적 분류
  - 머신러닝/딥러닝 기술을 이용한 고수온 현상의 선행인자 분석
- 설명가능 인공지능(X-AI) 기술 이용 한반도 고수온 현상 모델링
  - 분류된 고수온 이벤트와 선행인자를 이용한 기계학습기반 예측모형 개발
  - 설명 가능한 인과관계 학습 추론 기술을 이용한 한반도 주변해역 고수온 현상 예측성 진단 및 평가

## IV. 연구개발결과

- 해양기후변화 관련 자료수집 및 전처리
  - OISTIA, G1 fndSST 등 11 종 위성자료 등 관측자료 수집 완료
  - NOAA 재분석자료 등 수집 완료
  - 머신러닝 훈련자료인 SST, MLD, OLR 등 자료 전처리 완료

- 한반도 주변해역 고수온 이벤트 분석
  - 선행연구에 따라 퍼센타일 기준으로 고수온 이벤트 정의 채택
  - EOF 분석 등을 이용한 고수온 이벤트 모드의 분류
  - 고수온 현상의 선행인자로서 열대태평양 OLR과 SST 인자 추출 완료
- 설명가능 인공지능(X-AI) 기술 이용 한반도 고수온 현상 모델링
  - SST 장기변동 학습을 위한 딥러닝 기반 베이스라인 모델 구축
  - 한반도 고수온 예측을 위한 딥러닝 모델 개발
  - Feature attribution 기반 설명가능 인공지능 분석방법 적용을 통한 한반도 고수온 예측성 평가

#### V. 연구개발결과의 활용계획

- 고수온 이벤트의 예단인자 발굴
- 고수온 이벤트의 머신러닝 기반 예측모형 개발
- 발굴된 고수온 예단인자는 고수온 현상의 역학적 이해 및 모형 개발에 활용
- 고수온 예측모형은 고수온 재해 저감 방안 마련에 활용
- 「AI+해양순환·기후변화」 융합연구 활성화 및 우수 학술성과 도출

# SUMMARY

## I. Title

A pilot development of the data driven model for ocean climate change simulation using X-AI technology

- Study on the marine heat waves in the vicinity of Korean Peninsula -

## II. Objectives of this Study

Diagnosis and evaluation of predictability of the data-driven model, which could forecast marine heatwaves in the vicinity of Korea using X-AI technology

## III. Contents of this Study

○ Data compiling and preprocessing associated with Ocean climate variability

- Obtaining of in-situ observation and satellite data

- Downloading of atmosphere and ocean reanalysis datasets

- Preprocessing of selected data for machine learning model

○ Analysis of marine heatwaves in the vicinity of Korea

- Definition of marine heatwaves near South Korea

- Objective classification of marine heatwaves using machine learning/deep learning

- Analysis of predictors for marine heat waves using machine learning technique

○ Modeling of marine heat waves in the vicinity of Korea using X-AI technology

- Development of marine learning-based prediction model using predictors of marine heatwaves

- Diagnosis and evaluation of predictability for marine heatwaves in the vicinity of Korea using explainable causality learning model

#### IV. Results

- Data compiling and preprocessing associated with Ocean climate variability
  - Procuring of 11 satellite-based datasets such as OISTIA, G1 fndSST
  - Downloading complete of NOAA reanalysis
  - Preprocessing complete of SST, MLD, OLR for machine learning model
- Analysis of marine heatwaves in the vicinity of Korea
  - Definition of marine heatwaves using percentile criteria
  - Classification of marine heatwaves using EOF analysis
  - Excavation of predictors for marine heatwaves such as tropical Pacific OLR and SST anomalies
- Modeling of marine heatwaves in the vicinity of Korea using X-AI technology
  - Construction of deep-learning baseline model for long-term SST variability
  - Development of deep learning model for prediction of marine heatwaves near South Korea
  - Evaluation of predictability of the prediction model using feature attribution-based X-AI technology

#### V. Application plan of research results

- Finding of predictors for marine heatwave prediction
- Development of prediction model based on machine learning for marine heatwaves
- Contribution to understanding of marine heatwaves and improvement of climate model by newly discovered predictors for marine heatwaves
- Contribution to preventing from damages for natural disasters
- Encouraging of AI+Climate multi-disciplinary study and outstanding research achievement

## KEYWORDS

고수온, 기후변화, 설명가능 인공지능, 딥러닝, 자료기반 모형  
Marine heat waves, climate change, X-AI, DL, data-driven model





## < Contents >

|   |    |
|---|----|
| 1. Overview of the Study .....                                    | 1  |
| 2. Current Status of Technological Advances .....                 | 14 |
| 3. Contents and Results .....                                     | 17 |
| 4. Achievement of Goals and Contributions to Relevant Field ..... | 91 |
| 5. Plans for Utilization of Results .....                         | 94 |
| 6. References .....   | 95 |

## < 목 차 >

|                             |    |
|-----------------------------|----|
| 1. 연구개발과제의 개요 .....         | 1  |
| 2. 국내외 기술 개발 현황 .....       | 14 |
| 3. 연구 수행 내용 및 결과 .....      | 17 |
| 4. 연구개발목표 달성도 및 대외기여도 ..... | 91 |
| 5. 연구개발성과의 활용계획 .....       | 94 |
| 6. 참고 문헌 .....              | 95 |

# 제 1 장 연구개발과제의 개요

## 1. 연구개발 목적

2016년 세계경제포럼인 다보스포럼에서 Klaus Schwab 회장의 ‘4차 산업혁명 (Industry 4.0)’ 사전 선언과 AI 기술의 폭발적 성장으로 AI 기술은 자체 기술발전과 혁신을 넘어 과학기술, 산업기술 전반의 성장을 가속화하는 단계에 이르렀으며, 이는 경제·사회 변혁, 산업구조 변화의 핵심동력으로 각국 경제성장에 미치는 파급효과가 비약적이고 이를 통해 국민 삶의 질과 국가 경쟁력이 제고되는 시점에 이르렀다.

이에 정부는 ‘I-Korea 4.0’이라는 정책 브랜드로 AI 기술 혁신 인프라 구축과 [AI+X]라는 AI기술 융합을 통한 타 산업분야 응용활성화를 통한 혁신 시너지 확보로 세계 수준의 AI 기술력을 확보하고, 지능형 DNA (Data·Network·AI) 인프라 구축 및 AI 기술의 확산을 통해 각 분야 혁신성장을 가속화하는 정책을 수행하고 있다.

기존 해양기후변화 연구에서는 관측자료, 위성자료와 역학모델을 이용하여 현황분석과 예측을 수행하여 왔으나, 자연현상에 대한 완전한 이해에는 도달하지 못하고 있다. 최근 개발, 확장되고 있는 머신러닝기술은 해양기후 연구를 위한 새로운 도구로 활용 가능하며, 전 세계적으로도 다양한 연구들이 시도되고 있다. 현존하는 자료를 최대한 수집하고 이를 이용한 자료기반모형을 개발함으로써, 기후변화와 관련된 많은 현상에 대한 이해를 증진시킬 수 있을 것으로 기대된다.

폭염, 고수온 현상 등과 같은 고영향(高影響) 극한 자연현상은 사회·경제적 파급효과가 크기 때문에 현상의 이해와 예측에 대한 노력이 필요하나 고수온 현상은 지역적 규모에서 많이 일어나는 특징 때문에 전세계적인 관심에도 불구하고 유럽, 호주 등 일부 국가에서만 현상에 대한 이해나 예측에 대한 연구가 수행되고 있다. 현재 지속적으로 개발되고 적용되고 있는 머신러닝 기술을 활용하여, 지역적 고수온 현상을 이해하고 예측할 수 있는 자료기반모형 (data driven model)을 개발하

고자 한다.

이를 위해 ▲ 고수온과 관련된 전세계 관측자료, 인공위성자료, 재분석자료를 수집하고, ▲ 인공지능 시스템의 동작과 결과를 이해하고 올바르게 해석하여 결과물이 생성되는 과정을 설명 가능하도록 해주는 기술인 X-AI 기술을 이용하여, ▲ 한반도 해역의 고수온 현상의 메카니즘 이해를 통해 이를 모델링하고 예측하는 기술을 개발하고자 한다.

## 2. 연구개발의 필요성

### 가. 기술적 측면

4차 산업혁명시대, AI 기술 폭발적 성장이 시작되었다. 데이터 축적·컴퓨팅 파워 진전·알고리즘 (딥러닝) 진화 등으로 AI 부흥기에 진입, AI 발전 속도는 점차 가속화되어 실험실 연구수준을 넘어 상용화 단계로 발전하면서, 폭발적 시장 성장<sup>1)</sup>과 신산업의 지능형 융합을 견인하고 있다.

또한 인공지능 기술은 자체의 기술혁신을 넘어 과학기술과 산업기술 영역의 성장을 가속하는 혁신의 조력자 (Enabler) 역할을 수행할 것으로 기대된다. 나아가 AI 기술이 급속도로 발전하면서 머신러닝/딥러닝의 심층신경망은 수많은 연결 노드의 복잡한 구조를 가지게 되고 이는 AI가 결정을 추론하는 방법과 과정에 불투명성으로, 입력자료와 결과와의 인과관계 규명에 어려움이 있다.

전문분야 (법률, 군사, 의료, 지식관리 등) 융합연구를 위해 의사결정과 결론, 예측할 하는 AI가 인공지능의 행위와 판단을 사람이 이해하는 형태로 설명할 수 있는 설명가능 인공지능(X-AI: eXplainable AI) 기술 개발이 2017년 DARPA<sup>2)</sup>를 중심으로 활발하게 이루어지고 있다.

해양연구 분야에서도 머신러닝기술의 적용 연구가 대폭 증가하고 있다. 미국 NOAA 산하의 NESDIS, ORNL 등의 연구기관에서는 NDVIA 등과의 협력 작업을 통해 기후변화와 해양 연구에 머신러닝 기술을 적용하고 있으며, 태풍 조기탐지, 극한기후 예측, 자료동화, 예측결과 검증, 결측자료 보간, 모수화 등에 대한 연구가 활발히 진행되고 있음. 중국, 유럽 등의 국가에서도 관측자료 품질 검증, 강수량 예측 등 많은 연구가 시도되고 있다.

---

1) 인공지능 세계 시장 규모(IDC, '17.11) : ('16) 78.1억\$ → ('21) 522억\$, CAGR 34%

2) 미 국방성 산하 국방고등연구계획국

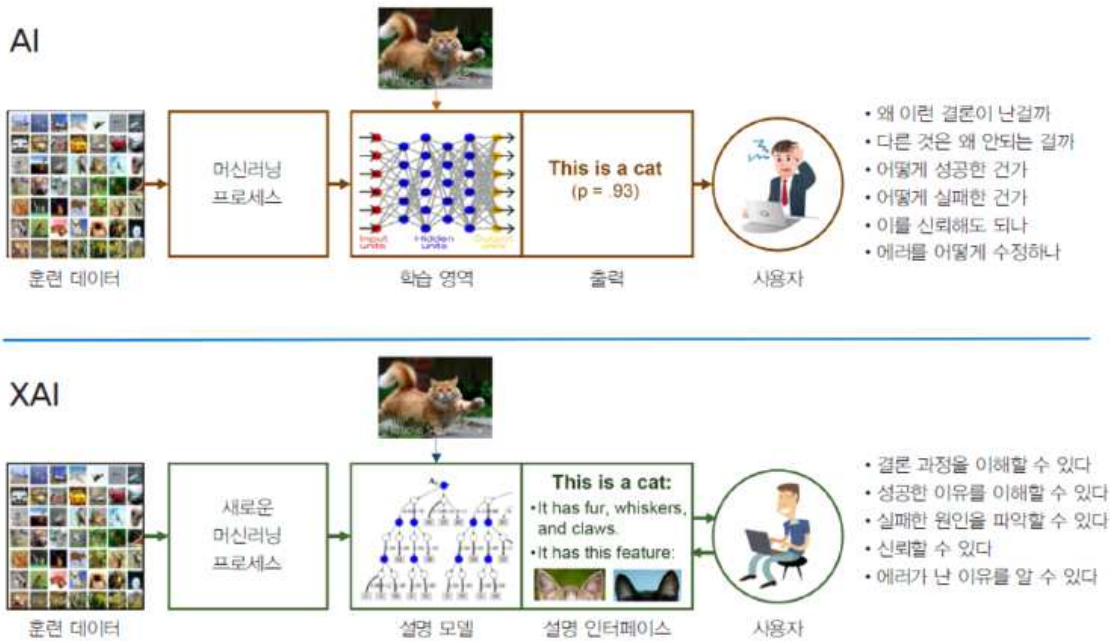


그림 2.1 설명가능 인공지능(X-AI: eXplainable AI) 개념

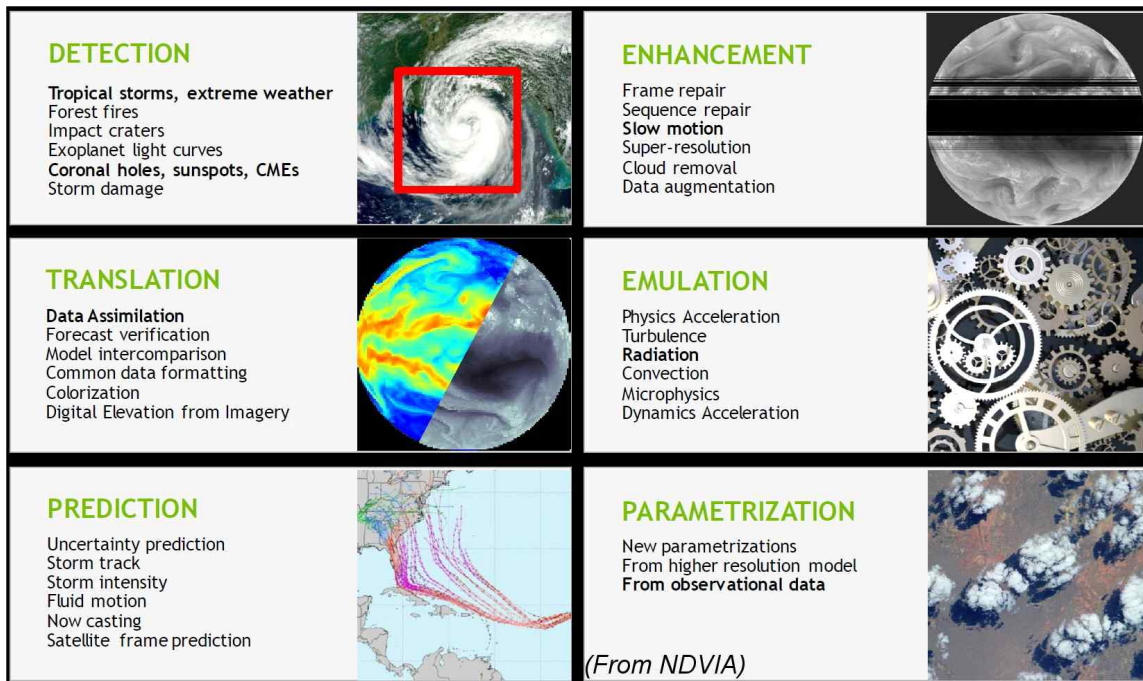


그림 2.2 기후변화, 해양연구 분야에서의 머신러닝 활용



# Deep learning for multi-year ENSO forecasts

Yoo-Geun Ham<sup>1\*</sup>, Jeong-Hwan Kim<sup>1</sup> & Jing-Jia Luo<sup>2,3</sup>

Variations in the El Niño/Southern Oscillation (ENSO) are associated with a wide array of regional climate extremes and ecosystem impacts<sup>1</sup>. Robust, long-lead forecasts would therefore be valuable for managing policy responses. But despite decades of effort, forecasting ENSO events at lead times of more than one year remains problematic<sup>2</sup>. Here we show that a statistical forecast model employing a deep-learning approach produces skilful ENSO forecasts for lead times of up to one and a half years. To circumvent the limited amount of observation data, we use transfer learning to train a convolutional neural network (CNN) first on historical simulations<sup>3</sup> and subsequently on reanalysis from 1871 to 1973. During the validation period from 1984 to 2017, the all-season correlation skill of the Nino3.4 index of the CNN model is much higher than those of current state-of-the-art dynamical forecast systems. The CNN model is also better at predicting the detailed zonal distribution of sea surface temperatures, overcoming a weakness of dynamical forecast models. A heat map analysis indicates that the CNN model predicts ENSO events using physically reasonable precursors. The CNN model is thus a powerful tool for both the prediction of ENSO events and for the analysis of their associated complex mechanisms.

The ability to forecast large-scale climate variability, with its effects on global social and environmental systems, is highly dependent on

the quality of ENSO predictions<sup>4</sup>. Although ENSO forecasts using atmosphere–ocean coupled models<sup>5,6</sup> generally outperform those of current statistical models<sup>1</sup>, state-of-the-art dynamical forecast systems nevertheless do not provide a skilful prediction of ENSO for lead times longer than one year. The multi-year prediction of ENSO events therefore remains a major challenge<sup>1</sup>.

However, the presence of an oscillating element in ENSO, linked to slowly varying oceanic variations and their coupling to the atmosphere, suggests that multi-year forecasts are possible<sup>5,7</sup>. Interestingly, equatorial Pacific anomalies during several La Niña events lingered for several years<sup>8</sup>. The high-frequency equatorial winds are less predictable<sup>9</sup>, but the slowly varying component of the equatorial winds coupled with underlying sea surface temperatures (SST)<sup>10</sup> is predictable to some extent. SST anomalies outside the equatorial Pacific can lead to an ENSO event with a time-lag longer than a year<sup>11,12</sup>. These studies imply that there is still room for improvement in ENSO prediction, although current methodologies might not be appropriate for this.

With the advent of the big data era, deep learning has had a dramatic impact on many domains by discovering intricate structures within large datasets<sup>13</sup>. In particular, CNNs have produced outstanding results in the processing of data with multiple-dimensional arrays with spatial structure (for example, used in the identification of objects within

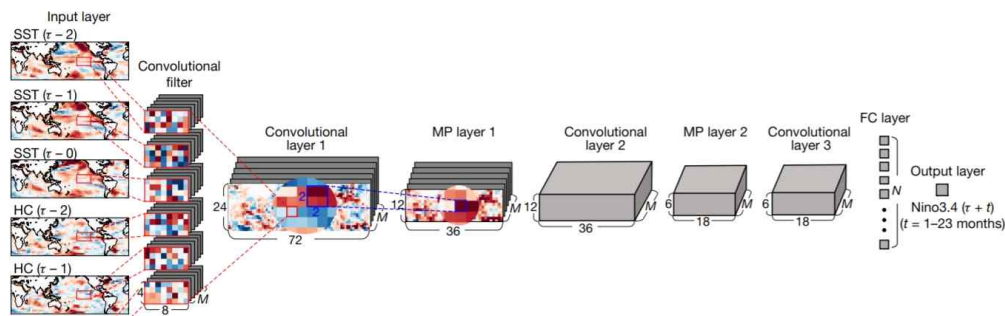


그림 2.3 머신러닝을 이용한 ENSO 예측 논문 - Ham et al. (2019), Nature

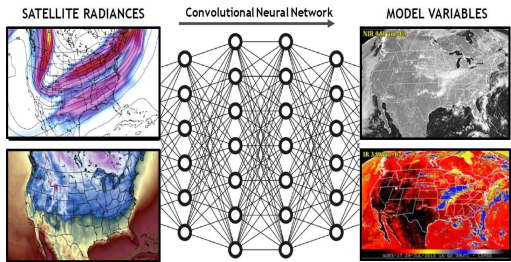
우리나라 연구진이 2019년 9월에 Nature에 머신러닝을 이용한 ENSO 예측에 관한 논문을 수록하는 등, 우리나라 대부분의 연구기관에서도 머신러닝을 연구에 적용하기 위해 노력하고 있다.

딥러닝 초기에 개발된 CNN, RNN, LSTM 알고리즘들이 해양자료의 처리, 해양현상 분석, 시계열자료 예측에 다양하게 활용되고 있으며, 최근에는 GAN (Generative Adversarial Network)이 위성자료와 모델자료의 결합 분석과 결측 위성자료 보간 등에 사용되고 있다.

자연과학적 연구에 머신러닝을 적용할 때, 입력변수에 따른 분류 등 추론 결정에 영향을 미치는 구조 설명과 예측결과로부터 역전파(back-propagation) 형태로 신경망의 계층별 기여도 측정이 가능한 X-AI 기술은, 해양분야 연구에도 널리 활용할 수 있는 기술로, 기후변화와 관련된 관측자료, 재분석자료, 위성관측자료를 수집하고 최신 X-AI기술을 이용하여 기후변화예측 자료기반모델을 개발하는 것은 선도적인 연구수행이 될 것이다.

**MAP: MODEL TO SATELLITE (FORWARD OPERATOR)**

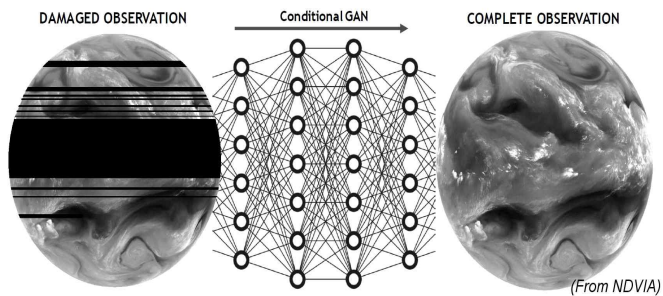
Model analyses to satellite observations



Maps from 3d fields to 3d fields, rather than one column at a time  
Can use spatial patterns to guide predictions (From NDVIA)

**IMPUTE MISSING DATA**

Train a conditional GAN to reconstruct missing pixels



(From NDVIA)

그림 2.4 CNN, GAN을 이용한 위성자료와 모델자료의 비교, 예측, 보간 연구

고수온 현상의 발생 및 특성 분석 등의 연구는 현재 전 세계에서 이슈화되어 있는 주제로서 국내에서는 체계적인 연구가 현저하게 부족하고, 여전히 개념 정립 이전 단계에 머물러 있는 실정이다. 고수온 현상의 대기인자인 한반도 주변 폭염에 대한 연구는 비교적 많이 알려져 있으며, 기상청은 2017년 폭염 센터를 설립하여 폭염에 대한 메커니즘 분석과 예측에 대한 연구가 수행중이며, 기상청에서는 폭염계절예측을 위한 연구를 지원하고 있으며, 해양과기원에서 수주를 받은 사업을 통하여 폭염계절예측을 위한 확률통계모형을 개발하여 현업화되었다.

고수온 현상에 영향을 주는 대기인자는 대규모 원격상관패턴에 의해 영향을 받는다. 한반도를 포함하는 북동아시아 지역은 대규모적인 관점에서 북서태평양 원격상관, CGT(CircumGlobal Teleconnection) 패턴, 그리고 EU(EUrasian) 패턴에 의하여 영향을 받음 크게 세 가지 원격상관에 의하여 영향을 받으나, 최근 엘니뇨 형태의 변화에 따라 이러한 원격상관 패턴이 변화되고 있다. 최근 연구에 따르면 한반도 폭염은 동서모드와 남북모드 두 가지에 의해 주로 영향을 받음이 알려졌다(Yeo et al., 2019).

해수부사업 등을 통하여 고수온 현상에 대한 이해와 예측을 위한 관측 및 연구가 수행되고 있으나, 머신러닝/딥러닝의 인공지능 기술을 적용한 예측 연구는 아직 수행되고 있지 않고 있다.

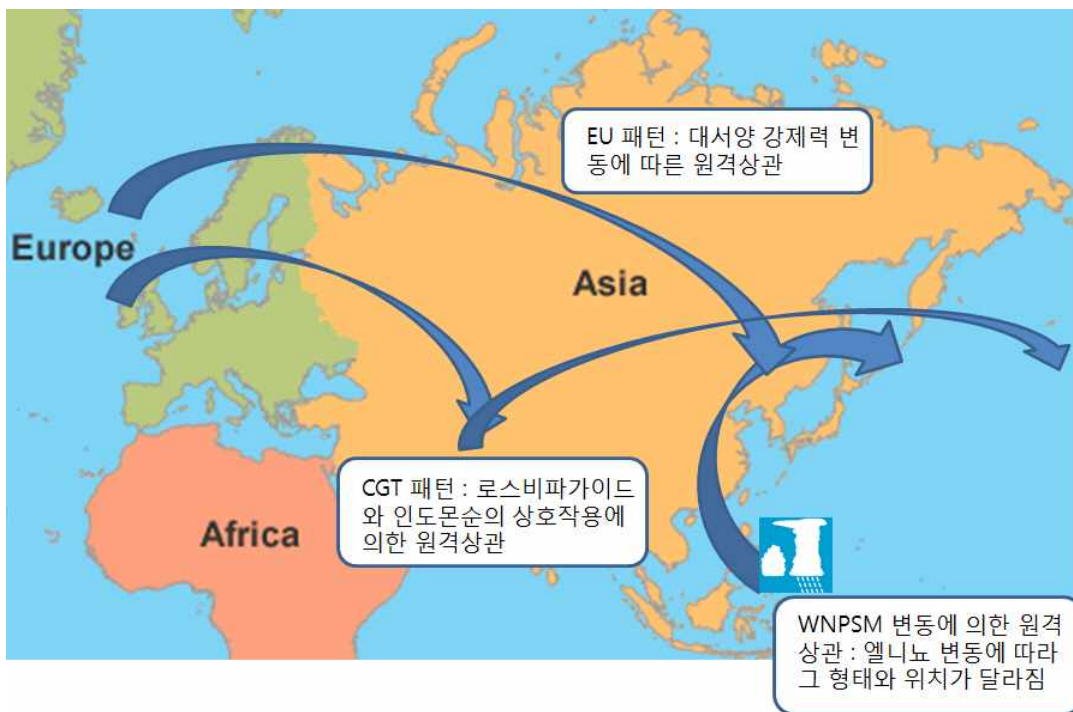


그림 2.5 한반도를 포함한 북동아시아 여름철에 영향을 주는 원격상관 모식도

## 나. 경제·산업적 측면

사회·경제적으로 막대한 영향을 미치는 고수온 현상 예측을 위하여 인공지능 기술을 도입함으로써, 해양분야와 데이터기술 분야의 융합연구체계 구성을 통해 기존에 찾지 못했던 예단 인자나 현상에 대한 새로운 이해가 가능하다.

AI는 단순한 신기술이 아닌 4차 산업혁명<sup>3)</sup>을 촉발하는 경제·사회 대변혁의 핵심 동력으로, 산업구조의 변화를 야기하고 향후 경제·사회 전반을 혁신할 근본기술로 각국 경제 성장에 비약적인 과급효과<sup>4)</sup>를 실현할 전망으로, 국민 삶의 질과 국가 경쟁력 제고를 위해 AI 기술력 확보가 필수적이다.

특히 국가 AI R&D 방향을 AI 기술혁신 인프라(데이터, 컴퓨팅 파워, AI 기술 등) 즉, 지능형 DNA(Data·Netowrk·AI) 인프라 구축 및 ‘AI + 응용산업’ 활성화에 초점을 정부 I-Korea 4.0 및 AI 기술개발 전략에 부합한다.

정부는 공공분야 (국방, 의료, 안전 등)를 대상으로 머신러닝, 시각지능 등 범용기술 연구를 중심으로 하되 안전예방시스템 등 향후 정부가 직접 사용할 분야에서 독자적인 AI기술력 확보를 위하여 고위험·차세대 기술 분야에 대한 중장기 투자 지을 병행하고 있다. 이에 해양분야 AI융합기술 발전은 해양관측/분석 연구에 있어 효율화를 증대시킬 것이며, 이는 국가·사회적 비용절감 효과로 이어질 것이다.

## 다. 사회·문화적 측면

본 연구는 사회·경제적으로 막대한 영향을 미치는 고수온 현상 예측을 통해 고수온 재해 저감 방안 마련에 활용될 것이다. 본 연구는 다음과 같은 국가적 아젠다(정부 국정과제, 제4차 과학기술기본계획 등)와 연계되어 있다.

### ○ 문재인 정부 100대 국정과제

- 전략 4 : 과학기술 발전이 선도하는 4차 산업혁명
  - (33)소프트웨어 강국, ICT 르네상스로 4차 산업혁명 선도 기반 구축
  - (34)고부가가치 창출 미래형 신산업 발굴·육성

### ○ 제4차 과학기술기본계획 중 안심하고 살 수 있는 안전한 사회 구현

- 해양/해난 사고 예방·대응 및 신속 구난 시스템 개발
- 재난현장 지원·대응을 위한 스마트 재난안전관리시스템 확보

또한 국가 AI R&D 방향을 AI 기술혁신 인프라 구축 및 「AI+산업응용」 활성화 및 관계부처 합동·4차산업혁명위원회, 세계적 수준 AI 기술력 확보 및 기술 확산을 통한 각 분야 혁신 성장 가속화를 위한 국가 R&D 발굴 및 투자 (2018.05.16.) 등 정부 ‘I-KOREA 4.0’ 정책 및 AI 기술개발 전략과도 부합한다.

나아가 해양과학기술의 기관 고유미션인 해양과학기술 및 해양산업 발전에 필요한 원천연구, 응용 및 실용화 연구 및 기관 경영목표\* 달성을 위한 4가지 중심 전략 중 ‘4차 산업혁명 대응 해양산업 허브 조성’이라는 기관 경영목표에도 부합하며, 국가사회 현안해결을 위한 체계적 해양과학기술 지원을 위한 기술로 활용 가능하다.

(\* 도전적 융합연구로 선도적 해양과학기술 역량강화·해양 신산업 발굴, 육성 및 산업화 기반 기술 개발)

### 3. 연구개발 범위

본 연구의 세부 성과목표 및 연구개발 내용과 범위는 다음과 같다.

#### 1) 해양기후변화 관련 자료수집 및 전처리

- 현장관측자료 및 위성자료 수집
- 기상, 해양 재분석자료 수집
- 머신러닝 수행을 위한 자료 전처리

#### 2) 한반도 주변해역 고수온 이벤트 분석

- 관측자료 기반 한반도 주변 고수온 이벤트 정의
- 머신러닝/딥러닝 기술을 이용한 고수온 이벤트의 객관적 분류
- 머신러닝/딥러닝 기술을 이용한 고수온 현상의 선행인자 분석

#### 3) X-AI 기술 이용 한반도 고수온 현상 모델링

- 분류된 고수온 이벤트와 선행인자를 이용한 기계학습기반 예측모형 개발
- 설명 가능한 인과관계 학습 추론 기술을 이용한 한반도 주변해역 고수온 현상 예측성 진단 및 평가



## 제 2 장 국내외 기술개발 현황

동아시아 여름철 기후의 중장기 예측에 대한 연구는 꾸준히 이루어져 있고, 현재 대기, 해양, 해빙 등이 접합된 기후모형(climate model) 또는 지구시스템모형(Earth System Model; ESM)과 같은 역학모형(dynamic model)과 통계적 방법을 이용하는 통계모형(statistical model) 등이 현업적 예측에 이용되고 있다. 그러나 기후모형은 자료동화를 포함한 초기화 과정이 잘 이루어졌다고 하더라도 중위도 지역의 예측성이 현저히 떨어지며, 지구시스템모형 역시 모형의 복잡성은 증가하였지만, 예측성능은 크게 차이가 나지 않고, 통계모형의 경우 그 예측성이 불안정하여 실제적인 예측성은 높지 않은 문제점을 가지고 있다. 또한, 열대야, 폭염, 고수온 등 극한 현상에 대한 중장기 예측도 현 수준의 예측기술은 크게 진보하지 못한 실정이다.

기후모형은 전지구 기상 및 기후를 예측하는 유용한 도구이지만, 대다수의 인류가 거주하는 중위도 지역 기후의 예측성은 현저히 떨어진다. 현재 수준의 기후모형의 계절규모 예측성은 변수마다 차이가 있지만, 중위도 지역에서 거의 없다. 그러므로 계절규모를 포함하는 중장기 기후에 대한 예측성을 확보하기 위해서는 현업적 장기전망을 할 때, 기후모형의 결과를 참조하기는 하지만 통계적 모형의 결과도 상보적으로 활용해야 한다.

기상청을 중심으로 서울대, 부산대 등은 통계적 기술을 바탕으로 다양한 기후 변수에 대하여 계절예측기술을 개발하였다. 현업적으로는 유사년분석과 같은 비정량적 모델은 물론, 예측인자와 예측변수의 선형적 상관관계에 기반한 다중선형회귀모형(Multiple Linear Regression Model; MLR model) 등이 많이 활용되고 있으며, 정준상관분석모형(Canonical Correlation Analysis Model; CCA model)과 같이 복잡한 선형 통계모형도 활용되고 있다. 그러나 CCA 모형의 경우 예측인자 영역과 예측변수 영역의 선택에 따라 그 예측성이 크게 차이가 나므로 현업적으로는 그 유용성이 불확실하다.

1990년대 말부터 기후예측에 인공신경망(artificial neural network; ANN)기법

이 도입되기 시작했다. Hsieh and Tangang (1998)은 인공신경망을 이용하여 ENSO(El Nino and Southern Oscillation)를 예측하였으며, 선행시간이 작을 때는 선형통계모형과 그 예측성이 큰 차이가 없지만, 선행시간이 클 때는 그 예측성이 현저히 개선됨을 보였다. 그러나 기후예측모형을 구성할 때, 입력 뉴런의 차원이 커질수록 예측성은 낮거나 불안정하여 입력 뉴런의 차원을 축소하는 문제는 여전히 남아 있었다.

최근 눈부신 발전을 이뤄낸 인공지능 기술은 단순히 기술적으로 우리에게 큰 편의를 주고 있을 뿐만 아니라 의료, 교육, 치안관리 등 사회경제적으로 큰 영향을 미칠 수 있는 분야에 많이 활용될 전망이다. 이에 인공지능 기술이 올바른 판단을 내렸는지, 결정의 근거가 무엇인지 등 신뢰할 수 있도록 인공지능을 ‘설명’할 수 있어야 한다는 요구가 증가하고 있다. 특히 이미지 분류, 자연어 처리 등 많은 분야에서 활동되고 있는 인공신경망 기반 인공지능의 복잡도 및 규모가 점차 증가함에 따라 설명가능한 인공지능 기술의 필요성이 더욱 더 강조되고 있다.

인공지능의 결정에 대한 설명의 필요성은 보안적 측면에서 더욱 중요하며, 예를 들어 자율주행차와 같은 안전 필수 시스템에서 특히 치명적인데, 만약 공격자가 자율주행 자동차의 카메라에 특수한 노이즈를 섞어 정지 표지판을 속도제한 표지판으로 잘못 인식하게 하는 경우 자동차가 멈추어야할 지점에서 멈추지 않아 큰 사고로 이어질 수 있음. 또한 해양과학 연구에 인공지능 기술을 융합한 예측 또는 예보 정보 제공 시 인공지능의 판단 근거를 올바르게 이해하지 못한다면 잘못된 예측 또는 예보 정보 서비스로 인해 사용자, 나아가 국가 정책에 피해를 주는 등의 부작용이 뒤따를 수 있다.

이와 같이 복잡한 인공지능 모델이 내린 판단에 대해 사람이 이해하고 납득할 수 있도록 설명을 생성하는 방법론을 설명가능한 인공지능 (eXplainable AI, X-AI)라고 한다. 통상적으로 X-AI 기술은 합성곱 신경망, 순환 신경망 등 복잡한 구조를 가지고 있는 심층신경망 기반 인공지능 모델들이 내린 결정과정을 알기 쉽게 더 자세히 분석하는 기술들을 말한다.

이에 본 연구에서는 인공지능 기술의 안전성 및 신뢰성, 과학적 도구로의 활용을 위해 예측 또는 판단에 대한 설명을 위한 ‘설명가능 인공지능’ 기술 개발의 핵심 기술 중 하나인 특성기여도 분석 방법의 적용을 통해 딥러닝 모델을 통해 예측된 한반도 주변해역의 수온 예측 결과에 대해 예측에 영향을 미친 시공간적 요인들을 분석함으로써 예측에 대한 근거를 함께 제시하는데 활용하고자 한다.

## 제 3 장 연구개발수행내용 및 결과

### 1. 해양기후변화 관련 자료수집 및 전처리

#### 가. 현장관측자료 및 위성자료 수집

해수면온도(Sea Surface Temperature, SST) 자료는 해양기후변화와 관련된 가장 기본적인 자료항목으로, 기후변화 연구에서 많이 사용되고 있다. 기후변화가 지구 규모의 현상이므로, 전 지구의 동시성 있는 자료 확보가 필요하지만, 보통의 해양관측 방법으로 자료 확보가 용이치 않으므로, 인공위성으로 측정된 자료를 많이 사용하고 있다.

전 세계의 다수 연구기관과 연구그룹에서 자체 방법에 따라 다양한 글로벌 SST 자료를 생산하여 공유하고 있다. SST 자료를 생산하는 연구진과 방법에 따라 SST 자료의 품질, 포맷, 오차표현 등에 차이가 많아서 사용자들이 어려움을 겪곤 했다. 이에 글로벌 SST 자료를 생산하는 연구자들이 GHRSSST (Group for High Resolution Sea Surface Temperature)을 구성하고, 자료생산 방법, 품질표시 방법 등을 통일시키는 규격으로 GDS (GHRSSST Data Specification)을 만들어 사용하고 있다. GHRSSST 그룹에서 생산하는 SST 자료세트는 총 20종이 넘으며, 대서양, 지중해 등 일부 해역을 대상으로 하는 자료세트와 전 지구를 대상으로 하는 자료세트가 있다. 기후변화에 대한 AI 모델 개발을 위해 GHRSSST 자료 중에서 글로벌 자료세트 16종을 수집하였다. 각 자료세트별 격자크기, 자료 기간, 자료 생산 기관 등을 정리하면, 표 3.1.1과 같다. 최근에 생산되는 자료 세트들은 위성자료 뿐 아니라, 관측자료를 같이 사용하여 자료의 품질을 높이고 있다. 각 자료세트의 생산에 사용한 원시자료의 내용을 정리하면 표 3.1.2와 같다.

표 3.1.1 수집한 GHRSSST 자료 목록

| No. | Name               | Grid  | Version | Depth | Start      | End        | Organ.                                     |
|-----|--------------------|-------|---------|-------|------------|------------|--|
| 1   | AVHRR_AMSR_OI SST  | 0.25  |         | 0.3m  | 2002.06.01 | 2011.10.05 | NOAA NCDC                                  |
| 2   | AVHRR_OI SST       | 0.25  | 2.0     |       | 1981.09.01 | Present    | NCEI (NCDC)                                |
| 3   | CMC SST (0.1)      | 0.1   | GDS V2  | fnd   | 2016.01.01 | Present    | Canadian Meteo. Center                     |
| 4   | CMC SST (0.2)      | 0.2   | GDS V2  | fnd   | 1991.09.01 | 2017.03.18 | Canadian Meteo. Center                     |
| 5   | DMI OI SST         | 0.05  | GDS V2  | fnd   | 2013.04.30 | Present    | Danish Meteo. Institute                    |
| 6   | GAMSSA SST         | 0.25  |         | fnd   | 2008.08.24 | Present    | Australian Bureau of Meteorology           |
| 7   | GAMSSA SST (GDS)   | 0.25  | GDS V2  | fnd   | 2008.07.23 | Present    | Australian Bureau of Meteorology           |
| 8   | NAVO K10_SST       | 0.1   |         | 1m    | 2008.04.01 | Present    | Naval Oceanographic Office                 |
| 9   | NAVO K10_SST (GDS) | 0.1   | GDS V2  | 1m    | 2008.10.01 | Present    | Naval Oceanographic Office                 |
| 10  | MUR SST (0.25)     | 0.25  | 4.2     | fnd   | 2002.06.01 | 2019.09.30 | JPL PO DAAC                                |
| 11  | MW_IR_OI SST       | 0.09  | 5.0     | fnd   | 2002.06.01 | Present    | Remote Sensing Systems                     |
| 12  | MW_OI SST          | 0.25  | 5.0     | fnd   | 1997.12.31 | Present    | Remote Sensing Systems                     |
| 13  | OSPO SST           | 0.05  | GDS V2  | fnd   | 2014.06.02 | Present    | Office of Satellite and Product Operations |
| 14  | OSPO Nighttime SST | 0.05  | GDS V2  | fnd   | 2014.06.02 | Present    | Office of Satellite and Product Operations |
| 15  | OSTIA SST          | 0.054 |         | fnd   | 2006.04.01 | Present    | UK Met Office                              |
| 16  | OSTIA SST (GDS)    | 0.054 | GDS V2  | fnd   | 2013.04.25 | Present    | UK Met Office                              |

표 3.1.2 수집한 GHRSSST 자료세트 생산에 사용한 원시자료 내역

| No. | Platform           | InSitu | NOAA  | Aqua   | Aqua  | Terra | METOP | SUOMI-NPP | GCOM-W1 | ENVISAT | TRMM | ERS  | COROLIS | MSG    | GOES-16 |
|-----|--------------------|--------|-------|--------|-------|-------|-------|-----------|---------|---------|------|------|---------|--------|---------|
|     | Sensor             | InSitu | AVHRR | AMSR-E | MODIS | MODIS | AVHRR | VIIRS     | AMSR2   | AATSR   | TMI  | ATSR | WINDSAT | SEVIRI | ABI     |
| 1   | AVHRR_AMSR_OI SST  | ○      | ○     | ○      |       |       |       |           |         |         |      |      |         |        |         |
| 2   | AVHRR_OI SST       | ○      | ○     |        |       |       |       |           |         |         |      |      |         |        |         |
| 3   | CMC SST (0.1)      | ○      | ○     |        |       |       | ○     | ○         | ○       |         |      |      |         |        |         |
| 4   | CMC SST (0.2)      | ○      | ○     | ○      |       |       | ○     |           |         | ○       | ○    | ○    | ○       |        |         |
| 5   | DMI OI SST         |        | ○     |        | ○     | ○     | ○     | ○         | ○       |         |      |      |         | ○      |         |
| 6   | GAMSSA SST         | ○      | ○     |        |       |       | ○     |           | ○       |         |      |      |         |        |         |
| 7   | GAMSSA SST (GDS)   | ○      |       |        |       |       | ○     |           | ○       |         |      |      |         |        |         |
| 8   | NAVO K10_SST       |        | ○     | ○      |       |       | ○     |           |         |         |      |      |         |        |         |
| 9   | NAVO K10_SST (GDS) |        |       |        |       |       | ○     | ○         |         |         |      |      |         | ○      |         |
| 10  | MUR SST (0.25)     | ○      | ○     | ○      | ○     | ○     |       |           | ○       |         |      |      | ○       |        |         |
| 11  | MW_IR_OI SST       |        |       | ○      | ○     | ○     |       | ○         | ○       |         | ○    |      | ○       |        |         |
| 12  | MW_OI SST          |        |       | ○      |       |       |       |           | ○       |         | ○    |      | ○       |        |         |
| 13  | OSPO SST           |        |       |        |       |       | ○     | ○         |         |         |      |      |         |        | ○       |
| 14  | OSPO Nighttime SST |        |       |        |       |       | ○     | ○         |         |         |      |      |         |        | ○       |
| 15  | OSTIA SST          | ○      | ○     | ○      |       |       |       |           |         |         | ○    |      |         | ○      |         |
| 16  | OSTIA SST (GDS)    | ○      | ○     |        |       |       | ○     |           |         |         | ○    |      |         | ○      | ○       |

대기에서 우주로 방출되는 총 복사에너지를 나타내는 OLR (Outgoing longwave radiation)은 대기 가스, 에어로졸, 구름, 표층에서의 흡수, 산란, 방출과정을 포함하는 지표로, 기후변화 연구에서 매우 중요한 자료항목이다. 고수온 현상 연구를 위한 자료로 미국 NCAR (National Center for Atmospheric Research)에서는 인공위성 AVHRR 자료를 이용하여 생산한 OLR (Outgoing longwave radiation) 데이터세트를 수집하였다. NCAR Archive에서 제공하는 OLR 자료의 종류는 표 3.1.3과 같으며, 월별 자료세트의 예는 그림 3.1.1과 같다.

표 3.1.3 수집한 OLR 자료세트 목록

| Statistic              | Temporal Coverage  | Spatial Coverage           | Grid Size             |
|------------------------|--------------------|----------------------------|-----------------------|
| Daily Mean             | 1974.06-2020.12.31 | 90.0N-90.0S<br>0.0E-357.5E | 2.5°x2.5°<br>(144x73) |
| Monthly Mean           | 1974.06-2020.12    |                            |                       |
| Daily Long Term Mean   | 1981-2010          |                            |                       |
| Monthly Long Term Mean | 1981-2010          |                            |                       |



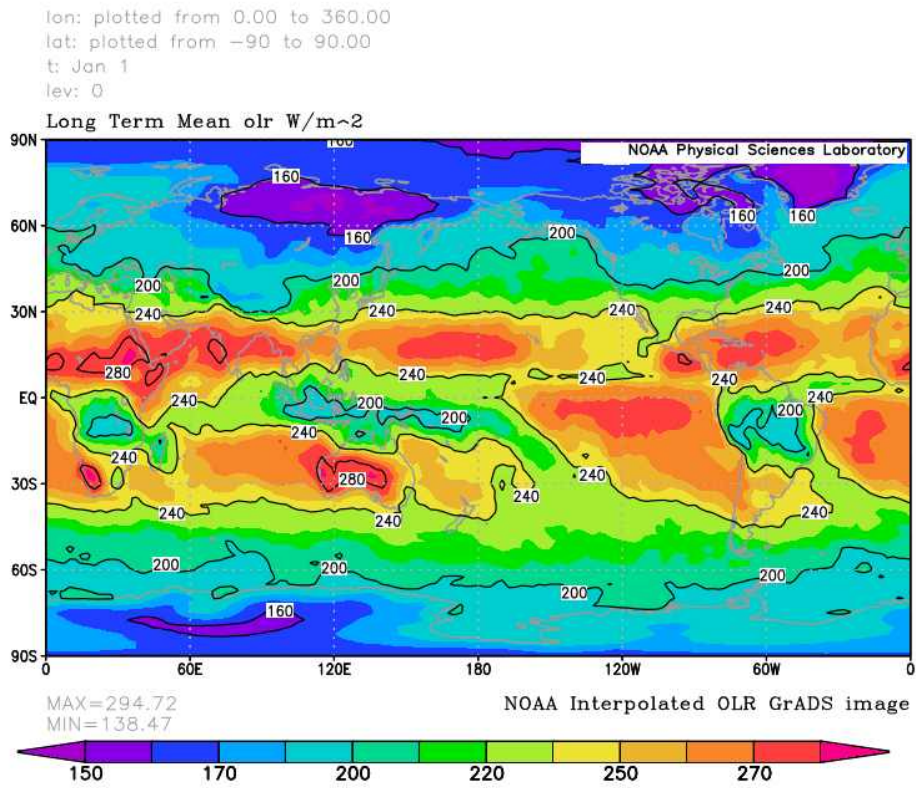


그림 3.1.1 월별 OLR 자료의 예 (NOAA OLR, 2020)

해양이 지구표면의 70%를 차지하는 만큼, 해양이 보유하고 있는 열량 (Ocean Heat Content, OHC)는 지구 규모의 기후변화와 밀접한 관계가 있으며, 다음과 같이 수온의 수직분포를 알면 계산할 수 있다.

$$OHC = \iiint_{h_2}^{h_1} \rho C_p \Theta dx dy dz$$

Here, (following TEOS-10; IOC et al., 2010)  $\rho$  is seawater density,  $C_p$  is the (constant) specific heat capacity,  $\Theta$  is conservative temperature (derived from *in situ* temperature, absolute salinity, and pressure), and  $h_1$  and  $h_2$  are the depth range over which the heat content is computed.

해양의 Heat Content 계산식 (Meyssigan et al., 2019)

미국 NOAA 산하 NCEI (National Centers for Environmental Information)에서는 전 세계 수온/염분 자료를 수집하여 World Ocean Database (WOD)라는 이름으로 자료세트를 생산하고 있다. 가장 최근의 자료세트는 WOD 2018이며, 이 중에서 북서태평양의 수직프로파일 자료를 240만 건 이상 수집하였으며, 이를 이용하여 OHC를 계산할 수 있는지 검토하였다. WOD 자료는 전 세계에서 다양한 관측방법 (Bottle, BT, CTD, PFL 등)으로 수집한 수직 프로파일 자료로 자료의 시공간 분포가 매우 불규칙적이므로, 이를 이용하여 OHC를 직접 계산하는 것은 많은 시간과 노력이 투입되어야 한다. 관측자료를 이용한 OHC 산출이 어렵기 때문에, 재분석자료로 산출한 장기 글로벌 OHC 자료들이 기후변화 연구에 많이 이용되고 있다. 관측자료를 이용하여 계산한 OHC 자료로는 NCEI에서 WOD 자료를 이용하여 Global heat contents 자료가 있다. 또한, 인공위성 자료를 이용하여 산출한 OHC 자료가 있으나, 고위도의 자료가 포함되어 있지 않아서, 사용하는데 제한이 있다. 관측자료를 이용하여 생산한 OHC 자료세트의 목록은 표 3.1.4와 같다. 본 연구에서는 NCEI OHC 자료 (seasonal, yearly, pentad)를 수집하여, 학습용 데이터로 사용하였으며, 0-700m 데이터의 예는 그림 3.1.3과 같다.

표 3.1.4 관측자료를 이용하여 생산한 OHC 자료세트 목록

| Name                                    | Grid | Time Interval | Depth                      | Period                       | Etc  |
|---|------|---------------|----------------------------|------------------------------|--|
| NCEI Global Ocean Heat and Salt Content | 1°   | 3-Month       | 0-700<br>0-2000            | 1955-present<br>2005-present | WOD 사용<br>10 <sup>18</sup> J                           |
|   |      | Yearly        | 0-700<br>0-2000            | 1955-2019<br>2005-2019       |  |
|   |      | Pentadal      | 0-700<br>0-2000            | 1955-2019                    |  |
| Satellite Ocean Heat Content Suite      | 0.5° | Daily         | 0-MLD<br>20-26<br>isotherm | 2012/08/27- Pre.             | Lat: -50°~60°<br>Lon: -180°~180°<br>kJ/cm <sup>2</sup> |

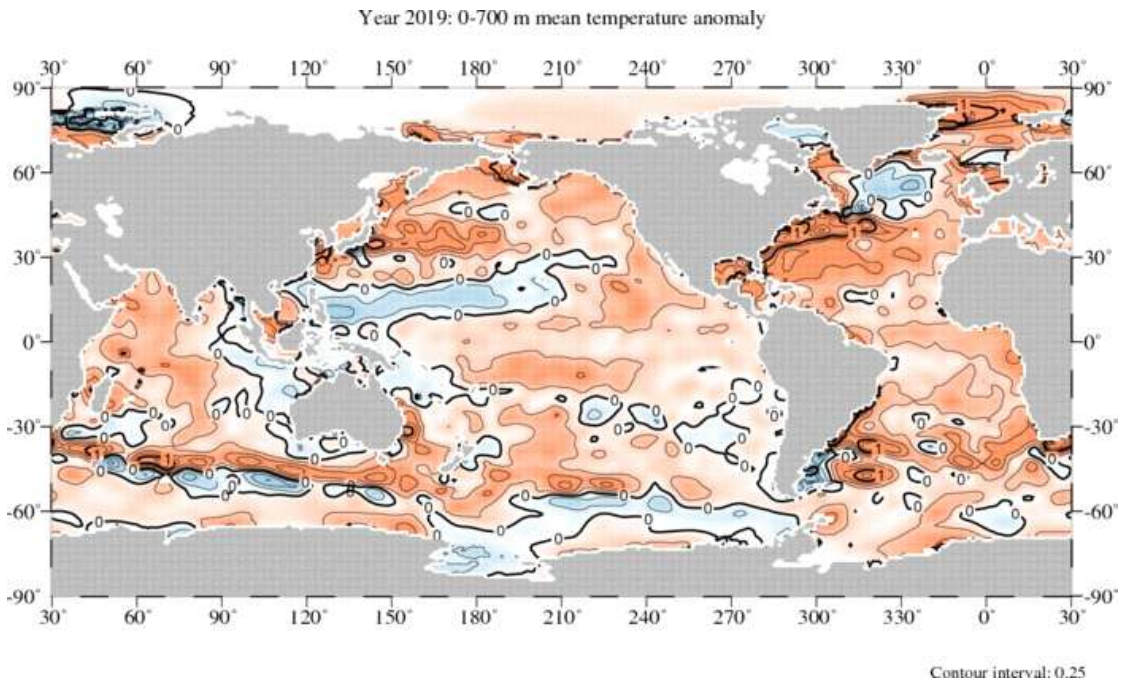


그림 3.1.3 NCEI OHC 자료의 예 (NCEI OHC, 2020).

## 나. 해양/기후 재분석자료 수집

해양에서의 장기 관측자료 확보가 어려우므로, 해양/기후변화 분석과 연구에는 수치모델을 이용하여 산출한 장기 시뮬레이션 자료를 많이 활용하고 있다. 본 연구에서도 국제적으로 많이 유통되는 해양재분석 자료를 수집하여 AI 모델 구축에 활용하였다. 기본적인 AI 모델을 수립을 위하여 CMIP5 (Coupled Model Intercomparison Project Phase 5) 과거 시뮬레이션 자료 중에서 1000년 동안의 OHC와 SST 자료를 수집하였다(표 3.1.5).

전 세계의 여러 해양/기상 관련 기관에서는 자체 운영 중인 수치모델을 이용하여 해양재분석자료를 생산하여 국제 해양학계에 제공하고 있다. AI 모델의 세부 튜닝에는 1950년 이후의 자료를 사용하였으며, 이를 위해 CMCC (Canadian Memorial Chiropractic College)의 C-GLORS (CMCC Global Ocean Reanalysis System), ECMWF (European Centre for Medium-Range Weather Forecasts)의 ORAS5 (Ocean Reanalysis System 5), CLIVAR GSOP (Climate Variability and Predictability Global Synthesis and Observations Panel)과 GODAE (The Global Ocean Data Assimilation Experiment) OceanView의 ORA-IP (Ocean Reanalyses Intercomparison Project)의 OHC 자료세트를 수집하였다. 각 자료별 시공간 격자, 중심, 기간 등을 정리하며 표 3.1.6과 같으며, C-GLORS OHC 자료의 예는 그림 3.1.4와 같다.

표 3.1.5 AI 모델 수립을 위해 수집한 CMIP5 자료세트 목록

| Data                     | Time Interval | Temporal Coverage | Spatial Coverage             | Grid Size          |
|--------------------------|---------------|-------------------|------------------------------|--------------------|
| CMIP5 Historical Run SST | Monthly       | 1001-2000         | 90.0°N-90.0°S<br>0.0°E-360°E | 1°x1°<br>(360x180) |
| CMIP5 Historical Run OHC | Monthly       | 1001-2000         | 90.0°N-90.0°S<br>0.0°E-360°E | 1°x1°<br>(360x180) |

표 3.1.6 재분석자료를 이용하여 생산한 OHC 자료세트 목록

| Name   | Grid | Time Interval | Depth   | Temporal Coverage              | Etc   |
|--|------|---------------|---|--------------------------------|---|
| CMCC G-GLORS<br>(Global Ocean Physical Reanalysis System)      | 0.5° | Monthly       | 0-300<br>0-700<br>0-2000<br>0-bottom                                      | 1980-2015                      | V5<br>J/m <sup>2</sup>                                      |
| ECMWF ORAS5<br>(Ocean Reanalysis System)                       | 1°   | Monthly       | 0-300<br>0-700<br>0-bottom  | 1979-2018                      | ensemble : 0-4<br>J/m <sup>2</sup>                          |
| ECMWF ORAS5<br>backward extension<br>(Ocean Reanalysis System) | 1°   | Monthly       | 0-300<br>0-700<br>0-bottom  | 1958-1978                      | ensemble : 0-4<br>J/m <sup>2</sup>                          |
| ORA-IP<br>(Ocean Reanalyses Intercomparison Project)           | 1°   | Monthly       | 0-10<br>0-100<br>0-300<br>0-700<br>0-1500<br>0-3000<br>0-4000<br>0-bottom | 1992-2014(V3)<br>1993-2015(V4) | CLIVAR GSOP and<br>GODAE OceanView<br>(Pseudo Heat Content) |

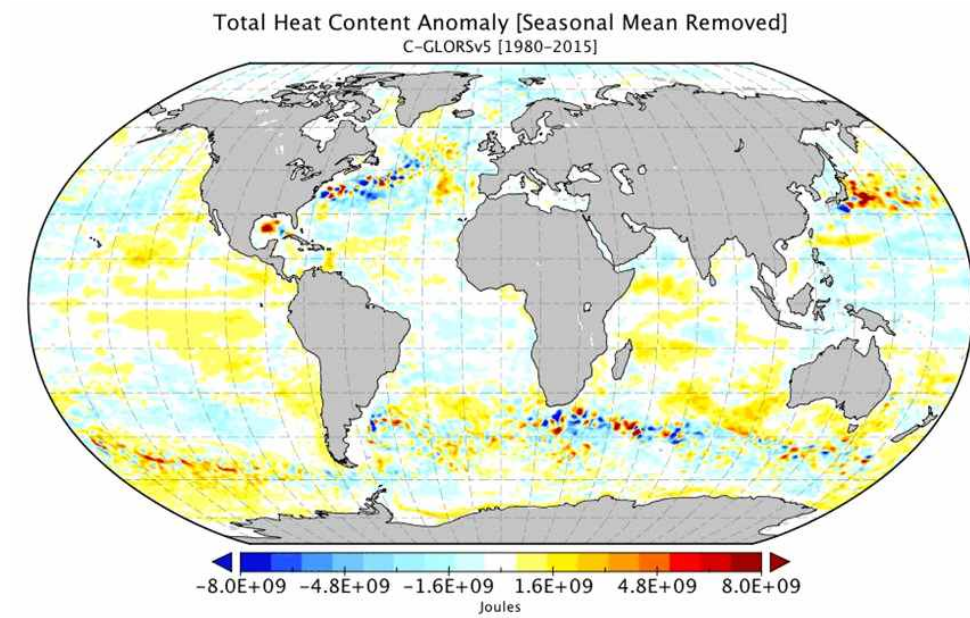


그림 3.1.4 C-GLORS OHC 자료의 예 (C-GLORS, 2020).



표 3.1.7 OHC 자료 계산을 위해 사용한 해양재분석자료 목록

| Data                      | Time Interval | Temporal Coverage | Spatial Coverage           | Grid Size              |
|---------------------------|---------------|-------------------|----------------------------|------------------------|
| SODA 2.2.4<br>Temperature | Monthly       | 1871–2010         | 89.5°N–75.5°S<br>0°E–360°E | 0.5°x0.5°<br>(720x330) |
| NCEP GODAS<br>Temperature | Monthly       | 1980–2019         | 65°N–75°S<br>0°E–360°E     | 1°x1°<br>(360x180)     |

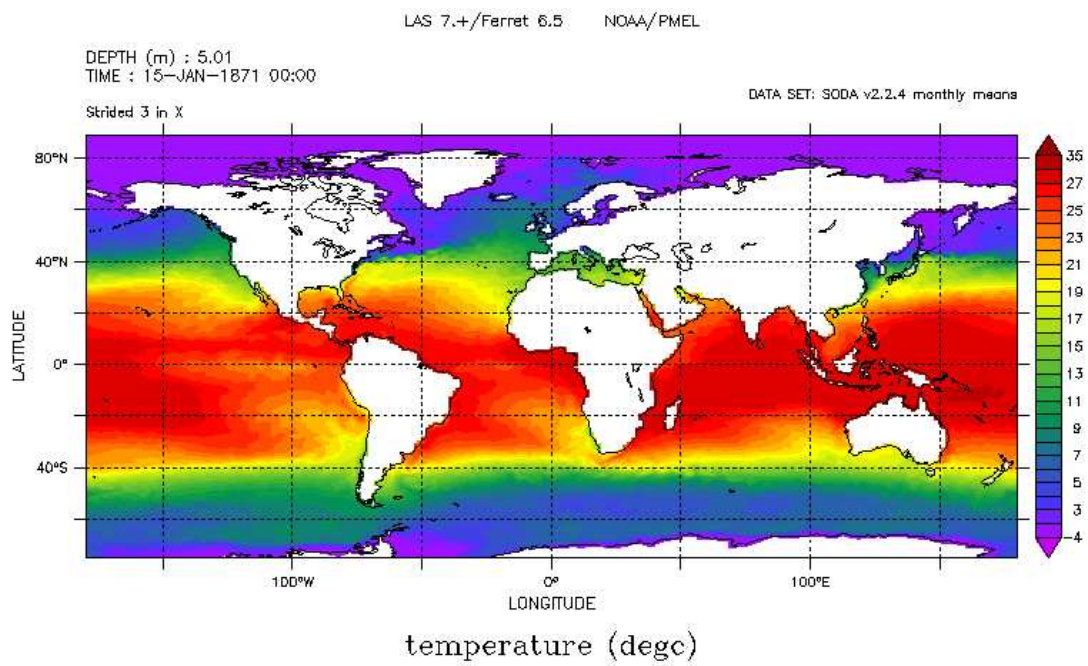


그림 3.1.5 SODA 수온자료의 예 (ADPRC, 2020).

AI 모델 튜닝을 위한 OHC 자료 확보를 위해, 해양기후변화 연구에 많이 사용하는 해양재분석자료 중에서 SODA (Simple Ocean Data Assimilation) 자료세트와 NCEP (National Centers for Environmental Prediction) GODAS (Global Ocean Data Assimilation System) 자료세트의 수온자료를 수집하였다(표 3.1.7, 그림 3.15). 수집한 SODA와 GODAS 수온자료 중에서 0-300m 자료를 이용하여 OHC를 계산하였다.

Nino 예측을 위한 AI 모델에서는 Nino 인덱스를 훈련데이터로 사용하므로, NOAA NWS (National Weather Service)의 Climate Prediction center에서 제공하는 Nion 12, 3, 4, 34 해역의 Nino 인덱스 자료를 수집하였다 (그림 3.1.6). 또한, 장기 Nino 인덱스를 계산을 위해 NCEI에서 제공하는 ERSST (Extended Reconstructed Sea Surface Temperature) v5 자료를 수집하였다(표 3.1.8).

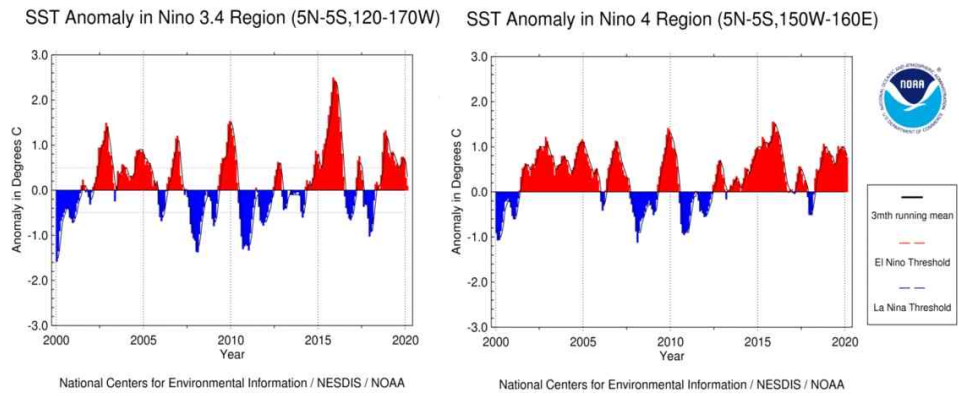


그림 3.1.6 SODA 수온자료의 예 (NCEI ENSO, 2020).

표 3.1.8 Nino 인덱스 계산을 위해 수집한 ERSST 자료 구성

| Data  | Time Interval | Temporal Coverage | Spatial Coverage        | Grid Size         |
|-------|---------------|-------------------|-------------------------|-------------------|
| ERSST | Monthly       | 1854-2020         | 89°N-89°S<br>-1°E-359°E | 2°x2°<br>(180x89) |

## 다. 머신러닝 수행을 위한 자료 전처리

수집한 해양관측자료, 위성자료, 재분석자료는 다수의 해양기관/조직에서 다양한 방법으로 생산한 자료이므로, 자료 종류, 시공간 분포, 육지 마스크, 자료 단위 등이 제각기 달라서, 여러 자료를 동시에 사용하는 AI 모델 개발에는 수집자료를 그대로 사용할 수 없다. 따라서, 개발하고자 하는 AI 모델의 목적에 맞춰 수집자료를 학습데이터와 검증데이터로 가공하여야 한다. 본 연구에서는 고수온 현상과 엘니뇨 예측을 위한 AI 모델을 개발을 추진하였으며, 이를 위한 학습데이터 확보를 위해 인공위성 SST 자료, OHC 자료, OLR 자료에 대한 전처리 작업을 수행하였다.

고수온 현상 머신러닝에는 적도( $20^{\circ}\text{S}$ - $30^{\circ}\text{N}$ ,  $0^{\circ}$ - $360^{\circ}\text{E}$ ) 해역 자료와 우리나라 남해( $30^{\circ}$ - $35^{\circ}\text{N}$ ,  $122.5^{\circ}$ - $130^{\circ}\text{E}$ ) 해역 자료를 사용하므로, 수집한 SST 자료와 OLR 자료를 학습데이터로 가공하였다(표 3.1.9). 적도해역 자료는 OI SST와 OSTIA SST 자료를  $0.25^{\circ}$  공간격자와 1일 시간 간격 자료로 구성하였다. OSTIA SST는 원시자료의 공간격자 크기가  $0.05^{\circ}$ 이므로,  $0.25^{\circ}$  자료를 만들기 위하여 평균값을 계산하였으며, 육지 마스크의 차이로 인한 오류를 줄이기 위하여 OI SST의 육지마스크와 일치 시키는 작업을 수행하였다. OLR 자료는 원시자료가  $2.5^{\circ}$ 이므로, 그대로 사용하였다. 남해 지역의 자료는 OI SST와 OSTIA SST 자료를 사용하였으며, 남해 해역을 3개의 해역으로 구분하여 자료를 재정리하였다. 남해의 3개 해역의 각 격자점에서의 SST 자료와 각 해역의 전체 SST 자료를 평균한 시계열 자료를 준비하였다. 학습데이터로 준비한 데이터세트는 통일된 형태의 netcdf 포맷으로 변환하였으며, 적도 SST 자료 저장을 위한 netcdf 파일의 구성은 그림 3.1.7과 같다.

표 3.1.9 고수온 현상 머신러닝용 자료세트 목록

| 자료                    | 해역 | 기간        | 시간간격 | 공간간격  | 시공간격자          |
|-----------------------|----|-----------|------|-------|----------------|
| OLRA                  | 적도 | 1982-2019 | 1일   | 2.5°  | 144x21x13870   |
| OI SST                | 적도 | 1982-2019 | 1일   | 0.25° | 1440x200x13870 |
| OSTIA SST             | 남해 | 1982-2019 | 1일   | 0.25° | 30x20x13870    |
| OI SST time series    | 남해 | 1982-2019 | 1일   |       | 3x13870        |
| OSTIA SST             | 적도 | 2007-2019 | 1일   | 0.25° | 1440x200x4745  |
| OSTIA SST             | 남해 | 2007-2019 | 1일   | 0.05° | 150x100x4745   |
| OSTIA SST time series | 남해 | 2007-2019 | 1일   |       | 3x4745         |

```

Format:
  netcdf4_classic
Global Attributes:
  title = 'SST anomaly at the Equator (From OSTIA SST)'
  institution = 'KIOST'
  name_creator = 'Sung-Dae Kim'
  date_creation = '07-Aug-2020 09:39:47'
  conventions = 'CF-1.6'
  comments = 'Average SST calculation : 2007-2019'
Dimensions:
  ntime = 4745
  nlat = 200
  nlon = 1440
  nday = 365
Variables:
  time
    Size: 4745x1
    Dimensions: ntime
    Datatype: int32
    Attributes:
      standard_name = 'time'
      units = 'days since 0000-01-01 00:00:00'
      axis = 'T'
      calendar_type = 'NOLEAP'
      _FillValue = 999999
  lat
    Size: 200x1
    Dimensions: nlat
    Datatype: single
    Attributes:
      standard_name = 'latitude'
      units = 'degrees_N'
      axis = 'Y'
      _FillValue = 999999
  lon
    Size: 1440x1
    Dimensions: nlon
    Datatype: single
    Attributes:
      standard_name = 'longitude'
      units = 'degrees_E'
      axis = 'X'
      _FillValue = 999999
  ssta
    Size: 1440x200x4745
    Dimensions: nlon,nlat,ntime
    Datatype: int16
    Attributes:
      standard_name = 'SST Anomaly'
      units = 'degrees C'
      add_offset = 0
      scale_factor = 0.01
      _FillValue = 9999
  avgsst
    Size: 1440x200x365
    Dimensions: nlon,nlat,nday
    Datatype: int16
    Attributes:
      standard_name = 'Average SST'
      units = 'degrees C'
      add_offset = 0
      scale_factor = 0.01
      _FillValue = 9999

```

그림 3.1.7 적도 OSTIA SST 자료의 netCDF 구성



엘니뇨 예측을 위한 머신러닝 학습에는 전지구 SST 자료와 OHC 자료를 사용하며, 인공위성 자료와 기후/해양 재분석 자료를 이용하였다. 인공위성 자료 중에서는 해양예보모델의 자료동화에 가장 많이 사용하는 OI SST 자료와 OSTIA SST 자료를 학습용 데이터로 가공하였다. 재분석 자료 중에서는 CM의 OHC와 SST 자료, CMCC G-GLORS의 OHC 자료, ECMWF ORAS5의 OHC 자료, SODA와 GODAS의 수온자료로 부터 산출한 OHC 자료를 학습용 데이터로 준비하였다(표 3.1.10). 엘니뇨 머신러닝에 사용하는 자료는 시간간격을 1개월, 공간격자 크기를 1°로 통일하였으며, 이 과정에서 육지 마스크는 CM 자료의 마스크를 이용하여 육지/해양 격자를 동일하게 구성하였다. 각 자료별로 전체 기간에 대한 평균값을 구한 후 anomaly 값을 계산하였으며, 동일한 형식의 netcdf 파일에 저장하였다.

표 3.1.10 엘니뇨 머신러닝용 자료세트 목록

| 자료        | 해역  | 기간        | 시간간격 | 공간간격 | 시공간격자         |
|-----------|-----|-----------|------|------|---------------|
| CM HC     | 전지구 | 1001-2000 | 1 개월 | 1°   | 360x180x12000 |
| CM SST    | 전지구 | 1001-2000 | 1 개월 | 1°   | 360x180x12000 |
| OI SST    | 전지구 | 1982-2015 | 1 개월 | 1°   | 360x180x408   |
| OSTIA SST | 전지구 | 2007-2015 | 1 개월 | 1°   | 360x180x108   |
| CMCC HC   | 전지구 | 1982-2015 | 1 개월 | 1°   | 360x180x408   |
| ECMWF HC  | 전지구 | 1982-2015 | 1 개월 | 1°   | 360x180x408   |
| SODA SST  | 전지구 | 1871-2010 | 1 개월 | 1°   | 360x180x1680  |
| SODA HC   | 전지구 | 1871-2010 | 1 개월 | 1°   | 360x180x1680  |
| GODAS SST | 전지구 | 1980-2019 | 1 개월 | 1°   | 360x180x480   |
| GODAS HC  | 전지구 | 1980-2019 | 1 개월 | 1°   | 360x180x480   |

## 2. 한반도 주변해역 고수온 현상 분석

일최고기온 특정 기준을 넘으며, 일정 기간 유지되는 폭염은 대기에서 일어나는 기상현상이다. 이러한 폭염과 더불어 해양에서도 열파(heatwaves)가 존재할 수 있다. 이러한 해양의 극한 현상(extreme phenomena)을 해양열파(marine heatwaves; MHW)라고 한다. 이러한 해양열파, 즉 고수온 현상은 일반적으로 폭염을 동반하지만 항상 그렇지는 않으며, 해양 자체 역학을 통해서도 형성 및 유지되기도 한다. 해양열파는 여름 및 겨울에도 일어날 수 있다.

해양 열파, 즉 해양의 고수온 현상은 지구온난화에 따라 그 빈도수가 증가하고 그 강도 역시 커지는 경향이 있다. 특히, 2003년 지중해의 고수온은 한 달 동안 평균에 비하여 섭씨 4도 이상을 보였으며, 해양 생태계에 막대한 피해를 끼쳤다. 2011년에는 서호주에서 두 달 동안 섭씨 약 3도를 넘는 기록적인 수온을 보였으며, 2012년에는 북대서양에서 두 달 동안 섭씨 약 2.5도를 상회하는 수온을 보였다. 이러한 기록적인 수온은 해양생물을 장거리 이동을 야기하거나 수산업 등 주변국의 막대한 경제적 파급효과를 미친다. 또한, 2013년에서 2015년 동안 북동태평양에서 200일 이상 섭씨 2.5도 이상의 역사상 가장 긴 온난 이벤트가 발생하였으며, 이는 북아메리카의 비정상적인 온난기후를 야기하였다.

해양의 고수온 현상은 대기의 열파에 의해서도 형성되지만, 따뜻한 수괴의 이류에 의해서도 나타날 수 있다. 고수온 현상의 발생역학은 지역적인 차이도 크고 아직 명확히 밝혀지지 않았지만, 일반적으로 대기의 고기압 지역에서 태양의 단파복사 가열에 의한 수온 상승에 의해서 나타나고, 해양의 혼합층이 얇거나 밀도가 낮은 조건에서 그 가열이 더 커지기도 한다. 또한, 아열대 지역에서 중위도로 향하는 따뜻한 해류에 의해서도 형성될 수 있다. 2011년 서호주와 2012년 북대서양 고수온 현상은 이러한 해류를 동반한다.

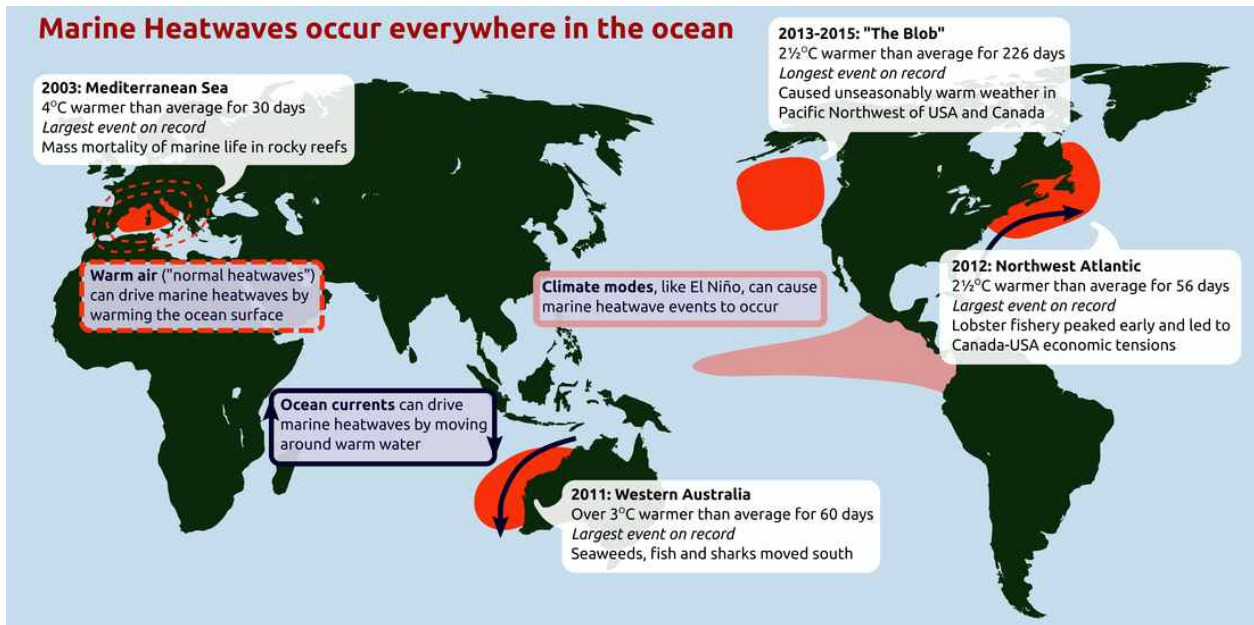


그림 3.2.1 전세계적인 고수온 현상(Hobday et al., 2016)

해양의 고수온 현상과는 달리 대기의 폭염은 많은 연구자들에 의해 그 특성들이 연구되어 왔다. 특히, 한반도의 폭염은 기상청 기준 일최고기온 섭씨 33도가 2일 이상 지속되는 것(폭염주의보 기준)으로 정의되며, 한반도의 관측소가 비약적으로 늘어난 1973년부터 현재까지 그 빈도수가 꾸준히 증가되고 있다(그림 3.2.2). 일반적으로 한반도에 나타나는 폭염은 기후적으로 7월 하순에서 8월까지 높은 빈도수로 나타나며, 이는 한반도에 나타나는 장마와 관련이 있다. 한반도는 동아시아의 몬순이 나타나는 지역으로서 겨울철에는 강수량이 적고, 여름철에는 강수량이 많은 뚜렷한 몬순 기후의 특성을 갖는다. 한편, 기후적으로 한반도에 나타나는 우기(rainy period)는 세 번의 정점을 갖는다. 첫 번째 정점은 그 강도는 약하지만 5월에 나타나며, 두 번째 정점은 그 강도가 기후적으로 가장 크며, 6월 하순에서 7월 하순에 나타난다. 흔히 이 시기를 전통적으로 장마라고 한다. 세 번째 정점은 8월 하순에서 9월 중순까지 나타나며, 이 시기를 가을장마라고 하며, 이 시기에 내리는 강수를 장마후 강수라고 부른다. 이러한 변동성은 북서태평양 지역에서 나타나는 기후적인 현상인 북진하는 계절내진동과 관련이 있다. 북서태평양 지역은 대기-해양 상호작용이 활발한 지역으로 해양 지역이지만 여름철에 강수가 많고, 겨울철에 강수가 적은 여름 몬순이 나타나는 지역이다. 이 북서태평양 지역의 여름 몬순은 세 번의 정점을 가지는 시간적 변동성을 갖는데, 이 북서태평양 지역의 대기 대류활동은 우리나라를 포함하는 경도 영역에서 북진하여 우리나라 여름철에 영향을 준다. 이러한 우리나라 지역의 우기와 우기 사이에는 휴지기 혹은 건조기(dry spell)가 존재하는데, 이러한 건조기 시기에 폭염이 많이 나타난다. 우리나라 지역에서 나타나는 폭염의 변동성은 관측소에 따라 상당한 공변동(coherency)이 있으며, 이는 폭염의 발생 및 유지 역학이 대기 혹은 해양의 대규모 역학에 의해서 나타난다는 것을 의미한다.

Lee and Lee (2016)은 한반도에 나타나는 폭염의 경년변동이 최근 시기 즉, 1990년대 중반 이후 북극진동(Artic Oscillation)과 밀접한 연관성이 있음을 보였다(그림 3.2.3). 그림 3.2.3은 북극진동과 상층 동서바람과의 관련성을 보이는데, 상층동서바람(zonal wind at 200 hPa)에 의한 상층 고기압성 순환은 한반도 지역에서 폭염과 아주 높은 상관성을 갖는다. 한편, 한반도의 폭염의 강도는 1980년대 이후 유의미한 증가경향을 갖지 않지만(그림 3.2.4), 그 지속일수는 증가하는 경향을 갖는다(Yeo et al., 2018). 이러한 대기의 온난 강제력은 해양의 가열을 생성 혹은 유지시킬 수 있는 필요조건을 제공한다.

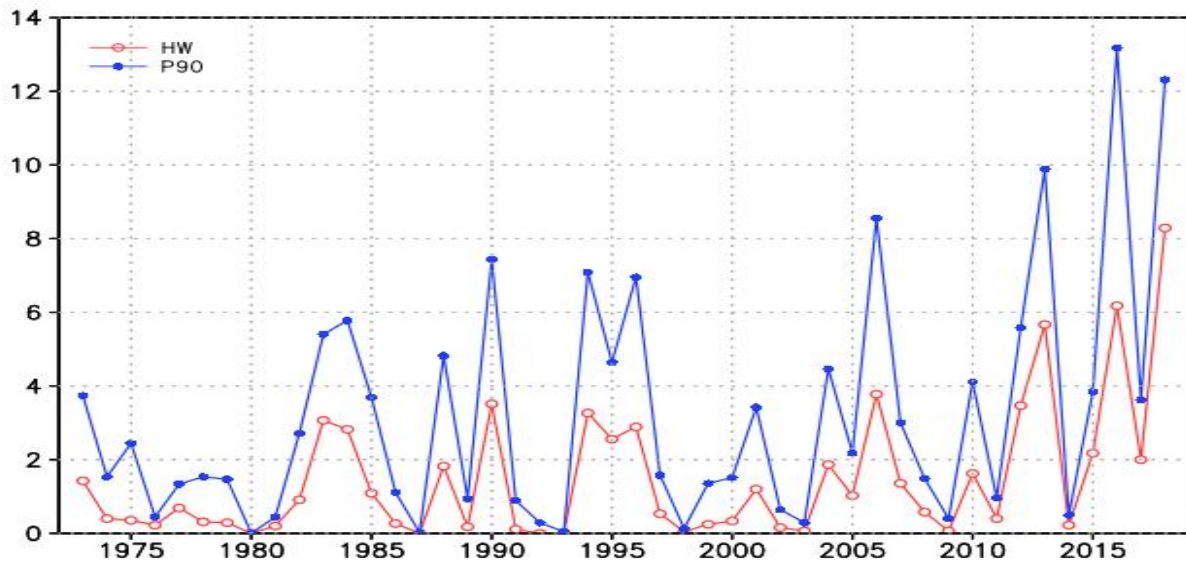


그림 3.2.2 한반도 59개 관측소 일최고기온에 따른 8월 폭염일수와 90퍼센타일 일수

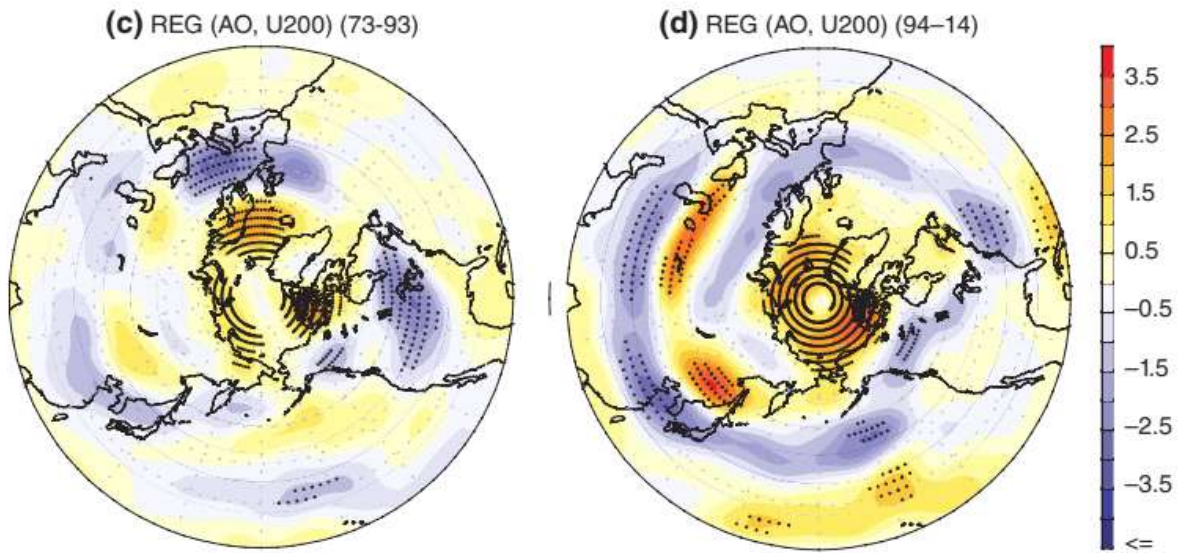


그림 3.2.3 북극진동과 상층 동서바람의 상관관계(Lee and Lee, 2016)

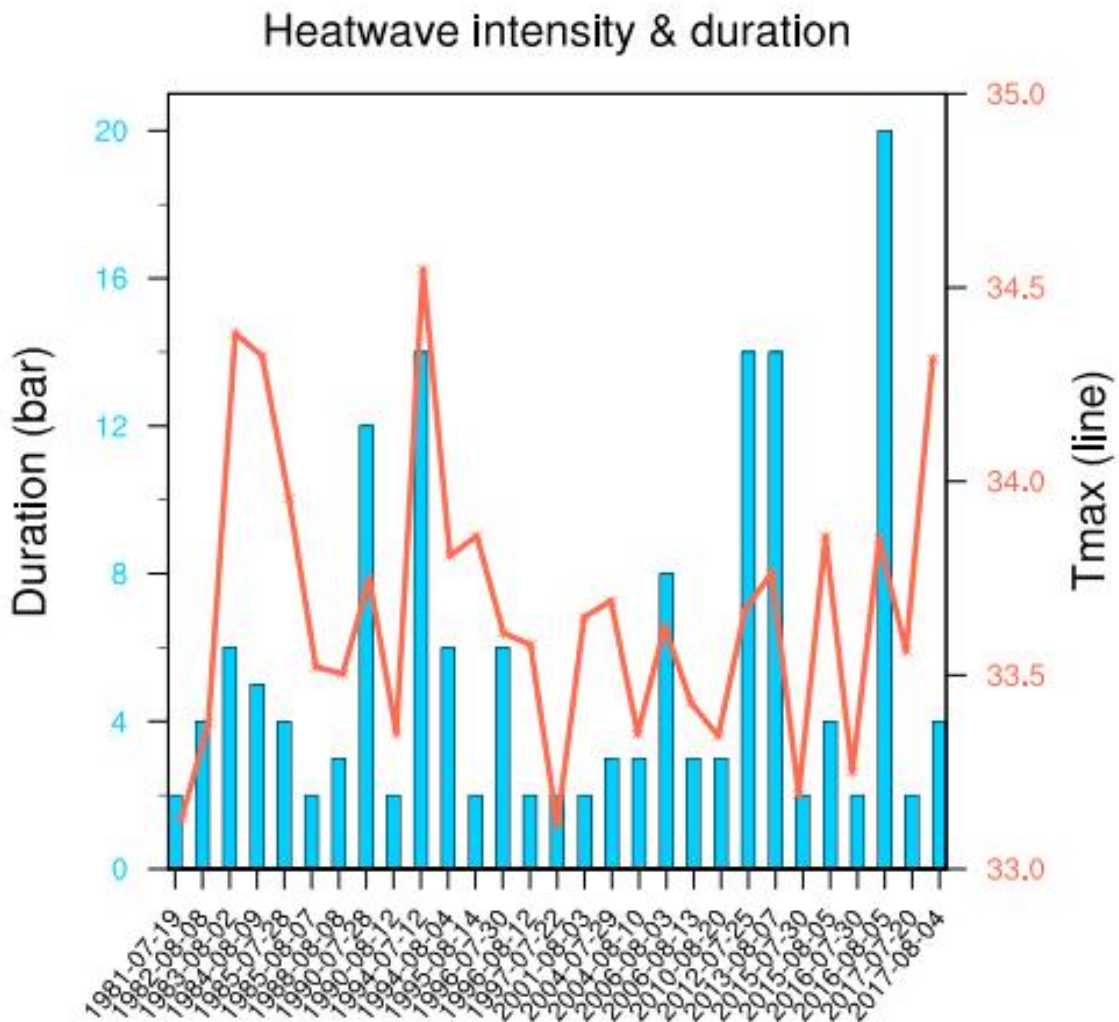


그림 3.2.4 여름철 한반도 열파의 강도와 지속시간의 경년변동(Yeo et al., 2018)



고수온이 수산업에 미치는 경제적 영향은 막대하다. 특히, 연안에 집중되는 양식장은 고수온에 취약하여 고수온이 발생했을 때, 양식어들이 집단폐사하는 경우가 빈번히 일어난다. 2013년 고수온의 경우 가두리양식 집단폐사 수가 약 500만 마리,, 2016년의 경우 약 377만 마리에 이르며, 경제적 피해는 각 해 50억원의 피해가 발생하였다(KBS 보도자료). 이에 따라 연안지역의 고수온의 특보는 경제적으로 막대한 영향력을 갖는다(그림 3.2.5).

기상청 폭염주의보는 일최고기온 33도이상 2일 이상 지속되는 경우에 발령된다. 즉, 폭염의 기준은 일최고기온의 절대값으로 정의된다. 그러나, 지역적으로 기후적 특성이 다르고 기후값도 다르므로 과학적으로 특정 기온 기준값(threshold temperature)으로 폭염을 정의하는 것은 바람직하지 못하다. 그러므로 다른 방식의 폭염 기준이 필요한데, 극한현상을 정의하는 기준으로 보통 퍼센타일(percentile) 방법이 많이 이용된다. 극값의 기준으로 90 퍼센타일 혹은 99 퍼센타일이 많이 인용된다. 그림 3.2.2는 한반도 59개 관측소 평균 8월 일최고 기온의 폭염일수와 90 퍼센타일 일수를 보인다. 폭염일수와 90 퍼센타일일수는 유사한 경년변동성을 보이지만, 90 퍼센타일 일수가 극값을 더 잘 대표한다고 할 수 있다. 90 퍼센타일 일수를 보면 여름철 전체의 폭염은 큰 증가가 보이지 않는 반면 8월의 폭염이 최근 큰 폭으로 증가하고 있음을 확인할 수 있다.

한반도 폭염이 최근 증가하는 경향과 유사하게 고수온의 발생빈도도 남해의 경우 꾸준히 증가하는 경향이 있다(그림 3.2.6). 2001년과 2003년에 강한 고수온이 발생하였는데, 2016년과 2018년에는 그림에서는 보이지 않지만, 더 강하고 지속적인 고수온이 발생하였다. 이러한 경향성은 지구온난화에 장기기후변동성이 포함된 것으로 보이며, 그 정확한 이유는 밝혀지지 않았다. 한반도를 포함하는 북동아시아 지역은 강한 지구온난화 경향을 갖는 동시에 강한 십년변동성 포함 장기 변동성을 갖는다. 이는 몬순기후가 다른 기후변동성과 상호작용하여 장기변동성을 갖기 때문이다. 특히, 1990년대 중반 이후 한반도 여름철 8월 강수가 현저히 증가하였으며, 2000년대 중반 이후에는 다시 감소하였다. 이는 폭염빈도 뿐 아니라 고수온 발생의 빈도에도 큰 영향을 미친다.



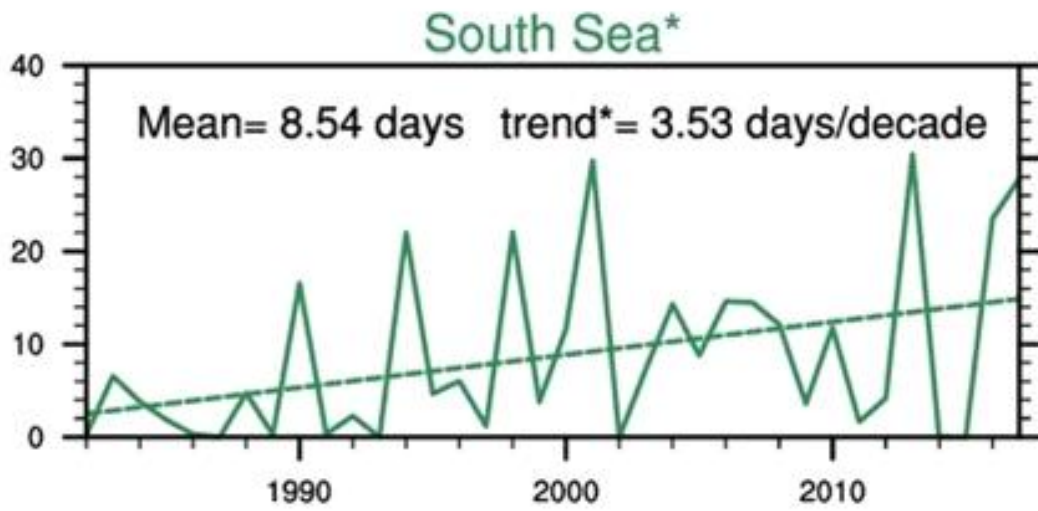


그림 3.2.6 한반도 남해 수온 90 퍼센타일 일수(KBS 보도자료)

고수온의 정의는 현재 통일되어 있지 않으며, 기관마다 다른 기준을 사용한다. 과학적으로 고수온의 정의는 문헌마다 다소 차이가 있다. 해양수산부는 섭씨 28도 이상의 수온을 고수온으로 정의하며, 국립수산과학원은 섭씨 25도 이상의 수온이 20일 이상 지속되었을 때 고수온으로 정의한다. Berkelmans et al. (2004)는 수온 최대값이 섭씨 30도를 초과한 경우에 고수온으로 정의하며, Wernberg et al. (2013)은 고수온 현상을 월평균 수온 편차가 약 3.5도 이상 높을 때로 정의한다. 이러한 고수온 정의에 대한 선행연구는 특정수온(threshold ocean temperature)에 기반하기 때문에 기후적 특성이 차이 나는 지역적 상황을 고려하지 못한다. 예를 들어, 남해의 기후적 평균 수온과 중위도 동해의 기후적 평균 수온은 큰 차이를 보이고 있고, 기후적 변동성이 작은 지역과 큰 지역인 경우에도 위의 특정수온 기반 정의는 적절한 정의로 보기 어렵다. 그러므로 새로운 방식의 고수온 정의가 필요한데, Hobday et al. (2016)은 폭염의 퍼센타일 정의법과 유사하게 90퍼센타일법으로 고수온 이벤트를 정의하였다. 즉, 특정 지역의 고수온을 순위통계량인 90 퍼센타일 값을 구하고 이 기준을 넘은 수온이 5일 이상 지속된 경우를 고수온으로 정의하는 것이다(그림 3.2.7). 5일 이상의 지속 기간이 필요한 이유는 특정일에 수온이 올라가는 열스파이크(heat spike)를 제거하기 위함이다. 특정 수온이 아닌 기후값 편차나 퍼센타일로 고수온을 정의하는 것이 과학적 연구에 적합하므로 본 연구에서는 Hobday et al. (2016) 연구를 인용하여 퍼센타일로 고수온 이벤트를 정의하였다.

한편, 한반도 전해역 평균을 이용하여 고수온 이벤트를 정의하는 경우에도 문제가 있다. 2001년 고수온의 경우 남해를 중심으로 서해와 동해의 고수온이 나타나는데, 2016년의 경우 남해에는 고수온이 나타나지 않고, 고수온의 발생이 중위도 즉, 서해와 동해에 집중된다. 즉, 한반도 전해역 평균을 이용하여 고수온 이벤트를 정의하는 경우 2016년의 경우 아주 강한 고수온 해임에도 불구하고, 약한 고수온 해 혹은 고수온이 아닌 해로 분류될 수 있다. 이러한 문제를 해결하기 위해 고수온 이벤트를 해역별로 나누는 것이 바람직하다. 본 연구에서는 남해, 서해, 그리고 동해 각각의 영역 평균한 수온에 대하여 고수온을 정의하였다. 또한, 남해를 평균하는 경우에도 영역이 넓은 경우 평균값이 평활화되어 고영향을 미치는 고수온의 극값을 반영하기 어렵기 때문에 본 연구에서는 비교적 좁은 영역에 대하여 고수온을 정의하였다. 특히, 본 연구에서는 생태적으로 고영향을 미치는 남해의 고수온에 초점을 맞추었다.

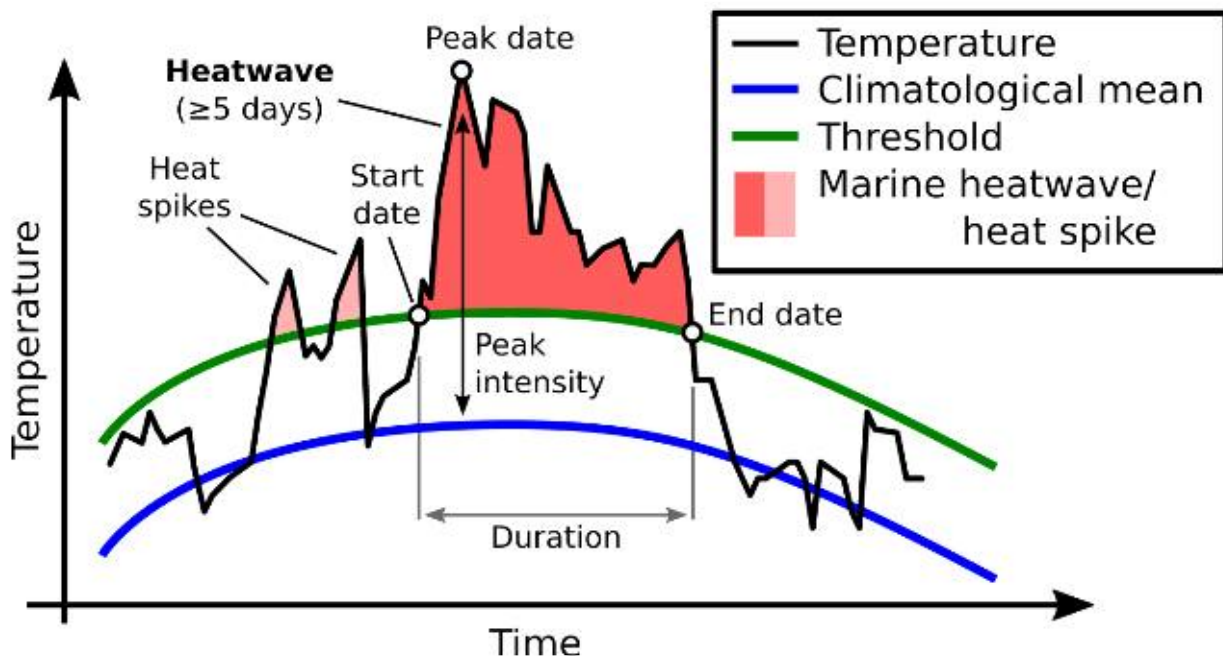


그림 3.2.7 고수온의 정의(Hobday et al., 2016)

한반도를 비롯한 북동아시아 여름철 기후는 원격상관(teleconnection)의 영향을 받는다. 원격상관은 원거리 기후편차의 상관성을 의미하며, 이 원격상관은 기후예측의 예측성이 나타나는 주요 이유이다. 한반도 여름철 기후변동성은 크게 세 가지 종류의 원격상관에 의해 영향을 받는 것으로 알려져 있다(그림 2.5). 첫 번째는 북서태평양 지역에서 북동아시아 지역에 영향을 주는 원격상관으로 일본-태평양 패턴(Pacific-Japan pattern; PJ pattern)으로 알려져 있다(그림 3.2.8). 태평양-일본 패턴은 필리핀 해와 일본 중부지역에 하층 대류권 순환 편차의 활동 중심을 갖는 여름철 원격상관 패턴으로, 1987년 Tsuyoshi Nitta에 의하여 제안되었다. 중국의 장마인 메이유와 관련하여 동아시아-태평양 패턴(East Asia Pacific Pattern)이라고도 불린다. 북서태평양-동아시아 여름몬순 평균 흐름인 연직취어가 포함된 하층 남서풍에서 필리핀 해 부근에 대기의 대류활동, 즉 대류권 중층에 대기의 가열이 있을 때, 북동쪽으로 향하는 로스비파가 형성된다. 이 중위도까지 영향을 주는 정상(stationary) 로스비파는 순압모드와 경압모드의 결합 반응으로서 열대지방에는 주로 첫 번째 경압모드가 큰 특징을, 그리고 중위도지방에는 순압모드의 특징을 보인다. 이와 같이 열대지방 대류활동의 반응으로 나타나는 로스비파의 형태가 태평양-일본 패턴이다. 태평양-일본 패턴은 선행하는 엘니뇨와 관련되어 있어서 강한 경년변동성을 갖는다. 즉, 겨울철 엘니뇨와 태평양-일본 패턴은 높은 상관성을 갖는다. 그러나 이러한 상관성은 1970년대 전에는 통계적으로 유의하지 않다. 엘니뇨-태평양-일본 패턴의 관련성은 북서태평양 지역 대류활동의 계절내진동과 높은 상관성이 있으며 그 관련성은 여름철 북서태평양 지역에 대한 계절예측에 이용될 수 있다. 그러나 1990년대 이후에도 엘니뇨와 태평양-일본 패턴의 상관성이 떨어지는 경향을 갖는다. 그 이유는 엘니뇨에 의한 태평양-일본 패턴과 북서태평양 여름몬순에 따른 태평양-일본 패턴이 중첩되어 나타나기 때문이다. 1990년대 중반 이후 북서태평양 여름몬순에 의한 원격상관 강도가 증가되었는데, 이는 엘니뇨 자체의 공간패턴 및 특성이 1990년대 중반 이후 변화하였기 때문이다. 북서태평양 여름몬순과 관련되어 나타나는 지위고도의 원격상관 패턴은 북서태평양 패턴(western North Pacific pattern)이라고 한다.

두 번째는 지구순환 원격상관(CircumGlobal Teleconnection; CGT) 패턴이다. 일반적으로 지구순환 원격상관은 북반구 여름철에 인도 몬순 혹은 제트기류의 변동성에 의해서 발현되어 유라시아 대륙부터 북미까지 연결되는 연속적인 로스비 파동 패턴으로 정의되며(그림 3.2.9), 줄임말은 CGT이다. 인도 몬순과 엘니뇨의 상호작용에 의한 열대 몬순 지역 잠열 방출이 주요한 발현 요인이다. 라니냐 (엘니뇨)가 시작되는

여름철에는 인도 몬순이 강해지는 (약해지는) 경향이 있으며 이에 따라 양의 (음의) CGT 패턴이 주로 나타나며 북반구의 여러 지역 기후 변동성에 영향을 미친다. CGT가 양(음)의 위상을 나타내는 여름철에 우리나라 지표기온은 평년보다 높고(낮고) 장마가 약화(강화)되는 경향이 있다(Ding and Wang, 2005).

넓은 의미에서 북반구 혹은 남반구 중위도에서 전 지구에 걸쳐 제트 기류를 따라 유도되는 연속적인 파동 형태(circumglobal waveguide pattern)를 뜻하며 제트기류가 강한 겨울철에 강하게 나타나는 경향이 있다. 일반적으로 파수(wave number)가 5이다. 즉, 전 지구적으로 5개의 파동을 나타내는 특성이 있다. 지구순환 원격상관에 의해서 열대와 중위도의 기후 변동성이 서로 연관성을 가질 수 있다. 또한 같은 중위도 내에서도 멀리 떨어져 있는 지역들의 기후가 연관된 변동성을 나타내게 된다. 중위도 제트기류 변동성에 의해 내재적으로 발생할 수 있는 자연적 변동 모드(internal variability mode)로 해석되기도 하며, 열대 지역 대류활동에 의한 잠열방출 (latent heat release)에 의해서 발현되는 모드로 설명되기도 한다. 전자의 측면에서는 지구순환 연속 파동 패턴 (circumglobal waveguide pattern)으로 주로 불리며 후자의 측면에서 지구순환 원격상관 (circumglobal teleconnection)으로 불리는 경향이 있다(그림 3.2.10). 최근에는 좀더 좁은 의미로 북반구 여름철 유라시아 대륙부터 북미까지 연결되는 원격상관 패턴으로 정의되는 경향이 있으며 줄임말로 CGT로 표현된다. 여름철에는 겨울철에 비해 제트기류가 약하기 때문에 일반적으로 중위도 파동이 약화된다. 즉, 제트기류 변동에 의한 내재된 자연적 변동 모드가 현저히 약화된다. 하지만 여름철 열대지역 혹은 아열대 몬순 지역에서의 강한 대류활동은 상당한 잠열 방출로 대기 파동을 발현시킬 수 있으며, 이에 따라 전 지구적으로 이동하는 로스비 파동 패턴, 즉 원격상관 패턴이 만들어 질 수 있다. 위의 그림은 7월과 8월에 전형적으로 나타나는 CGT 패턴이며 지배적으로 나타나는 파수는 5이다. 일반적으로 특정 지역, 즉 유럽 서부, 러시아 근방, 인도 북서부, 우리나라를 포함하는 동아시아, 북미에서 강한 변동성을 나타낸다. CGT가 양의 위상일 때는 주요 변동지역에서 대류권 상층 (200 hPa) 지위고도가 양의 편차를 나타내며 음의 위상일 때는 음의 편차를 나타낸다. 여름철 중위도 파동은 순압파의 특성을 보이기 때문에 하층에서도 같은 위상을 보이게 된다. 즉, CGT 양의 위상 시 주요 변동지역에서 대류권 전체에 걸쳐 양의 지위고도 편차가 나타나고 그에 따른 고기압성 순환 편차에 의해 지표 기온이 평년보다 높아지는 특징을 나타낸다(기상학백과, 한국기상학회).

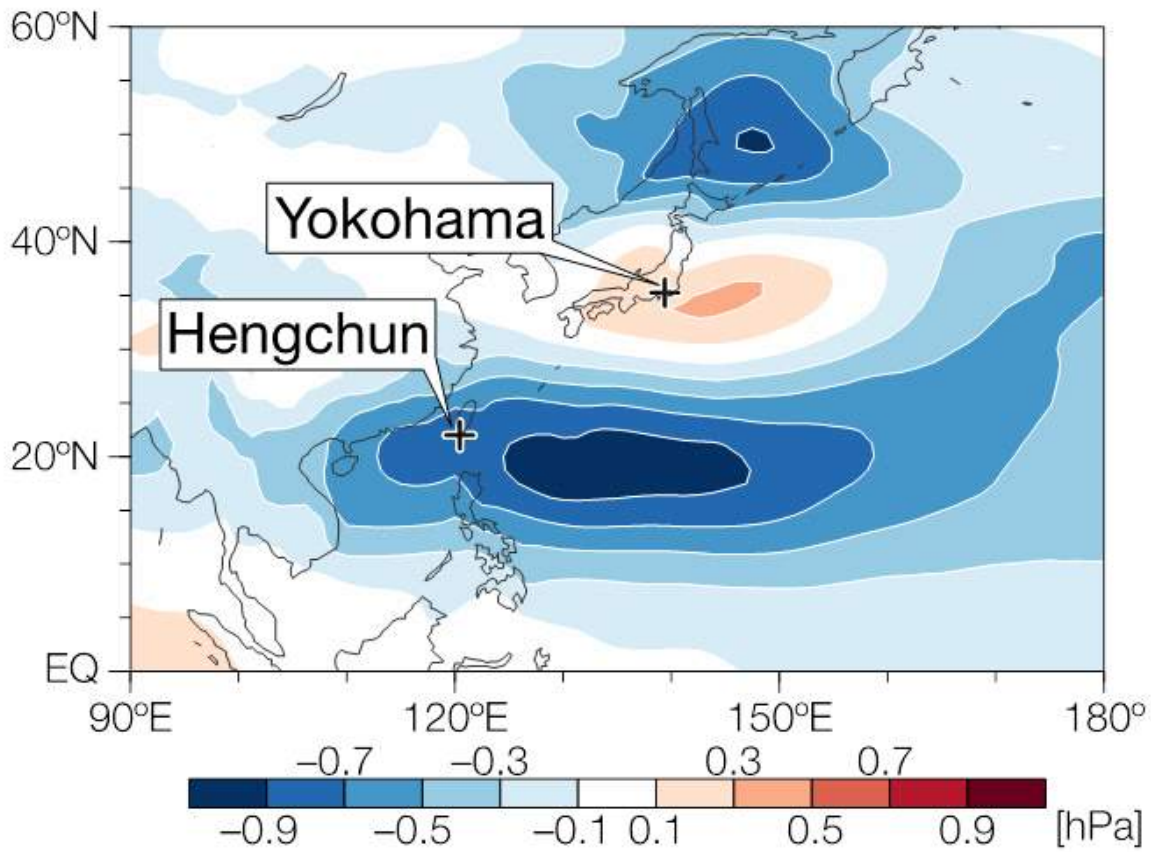


그림 3.2.8 태평양-일본 패턴(Kosaka and Nakamura, 2007)



### 인도 몬순-중위도 순환 상호작용 모식도



그림 3.2.9 지구순환 원격상관 모식도(네이버 기상학백과)

세 번째는 유라시안(Eurasian; EU) 패턴이다. 유라시안 패턴은 주로 북반구 겨울철에 나타나지만, 특정한 시기에는 여름철에도 나타난다. 유라시안 패턴은 크게 스칸디나비아(SCAND) 패턴, 동대서양/서러시아(EATL/WRUS) 패턴으로 나뉜다. 이러한 패턴은 일반적으로 REOF 분석을 통해 추출한다. 이러한 EU 패턴은 유라시아 지역의 준수압 파열 구조를 가지며, 로스비파 소스(source)의 위치는 일반적으로 다양하게 나타난다. 뿐만 아니라 평균장의 변화에 따라 그 파동의 전파방향과 위치는 비교적 큰 차이를 보인다. 그러므로 유라시안 패턴이 한반도를 포함한 동아시아 지역에 나타나는 것은 매우 제한적이다. 일반적으로 여름철 평균장에 대해 대서양 유라시안 패턴 강제력은 유라시아 지역을 관통하기 어려우며, 그림 3.2.9에서 보이듯이 대서양에서 시작된 대규모 로스비파동은 북인도 지역이나 티벳지역으로 향하는 경우가 많다. 그러나 특정한 시기에는 이 유라시안 패턴이 여름철 우리나라 지역에 뚜렷한 영향을 주기도 한다(그림 3.2.10). 한편, 여름철 유라시안 패턴은 2000년대 중반 이후 현재까지 북동아시아 지역에 영향을 거의 주지 않는다. 이러한 세 가지 원격상관 패턴은 그 시기에 따라 그 형태와 위치가 달라지는 특성을 가지므로 원격상관 패턴을 이용하여 동아시아 지역을 예측할 때, 이러한 원격상관의 장기변동성을 반드시 고려해야 한다.

고수온, 폭염과 같이 극한 현상도 기후편차에 크게 의존하므로 한반도 지역에 나타나는 폭염은 크게 원격상관과 연관되어 두 가지 역학에 의해 나타나는 것으로 알려져 있다. 첫 번째는 북서태평양 패턴에 의해 나타나는 폭염이고, 두 번째는 지구순환 원격상관 패턴에 의해 나타나는 폭염이다(Lee and Lee, 2016; Kim et al., 2018). 북서태평양 패턴과 관련된 폭염은 북서태평양 지역의 대기 대류활동에 의한 정상 로스비파 반응으로 이해할 수 있으며, 지역적 해들리순환(local Hadley circulation)으로도 불린다(그림 3.2.11). 즉, 북서태평양 지역의 대기 대류활동은 한반도를 포함하는 북동아시아 지역 남쪽에 대기의 고기압 패턴을 형성하며, 이 고기압 패턴은 하강 기류로 인해 구름이 없는 맑은 날씨를 동반하게 한다. 이는 태양 단파복사에 의한 지표면 가열이 상대적으로 강화되어 한반도 지역에 폭염을 야기한다. 이러한 형태의 폭염은 한반도 남해의 고수온을 일으키는 주요 역학이 된다.

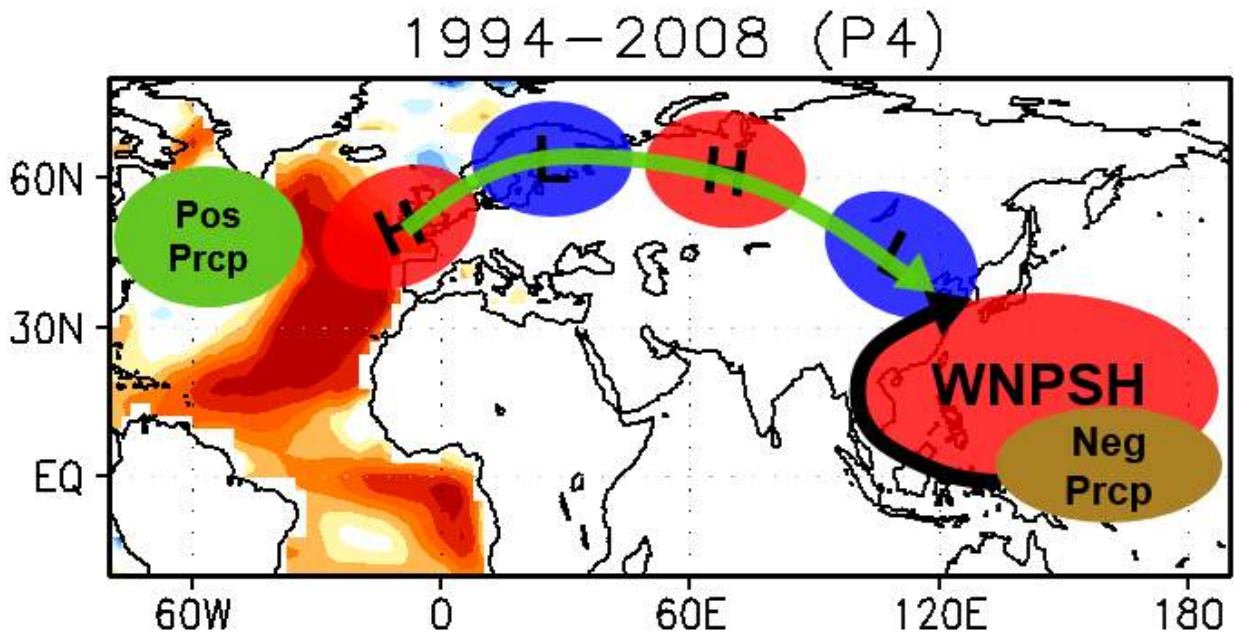


그림 3.2.10 1990년대 중반 이후 유라시안 패턴의 모식도(Ham et al., 2017)

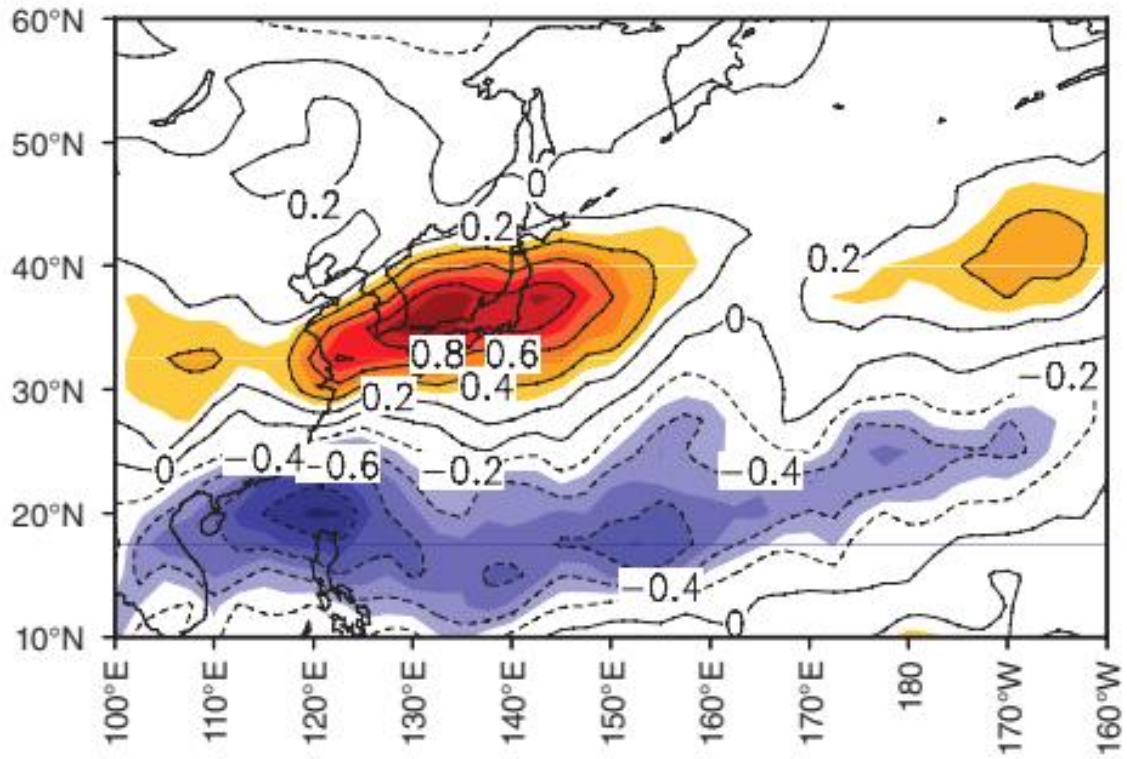


그림 3.2.11 한반도 폭염과 관련된 북서태평양 원격상관 패턴(Lee and Lee, 2016).

한반도에 폭염이 나타나는 가장 중요한 역학은 지구순환 원격상관에 의해 발생하는 것이다(그림 3.2.12). 그림 3.2.12는 한반도 폭염 지수에 회귀된 상층 지오폠펌셜 고도를 나타낸다. 지구순환 원격상관은 순압로스비파의 정상상태로 이해할 수 있으며, 그 열적 근원은 북인도의 강수편차에 의한 대기 중층의 비단열가열(diabatic heating)이다(그림 3.2.13). 이 비단열가열 패턴과 지구순환 원격상관 패턴의 관계는 다음의 파동활동속(Wave Activity Flux; WAF)을 이용하여 확인할 수 있다(그림 3.2.14; Kim et al., 2020).

$$\mathbf{W} = \frac{p \cos \phi}{2|\mathbf{U}|} \left\{ \begin{array}{l} \frac{U}{a^2 \cos^2 \phi} \left[ \left( \frac{\partial \psi'}{\partial \lambda} \right)^2 - \psi' \frac{\partial^2 \psi'}{\partial \lambda^2} \right] + \frac{V}{a^2 \cos^2 \phi} \left[ \frac{\partial \psi'}{\partial \lambda} \frac{\partial \psi'}{\partial \phi} - \psi' \frac{\partial^2 \psi'}{\partial \lambda \partial \phi} \right] \\ + \frac{U}{a^2 \cos^2 \phi} \left[ \frac{\partial \psi'}{\partial \lambda} \frac{\partial \psi'}{\partial \phi} - \psi' \frac{\partial^2 \psi'}{\partial \lambda \partial \phi} \right] + \frac{V}{a^2} \left[ \left( \frac{\partial \psi'}{\partial \lambda} \right)^2 - \psi' \frac{\partial^2 \psi'}{\partial \lambda^2} \right] \end{array} \right\}$$

Yeo et al. (2018)의 연구에 따르면, 한반도에 나타나는 폭염의 경년변동성을 크게 Z(zonal)-wave 패턴, M(meridional)-wave 패턴으로 나눌 수 있으며, 평균적으로는 Z-wave 패턴이 한반도 폭염을 발생시키는 데 주요 역학으로 작용하지만, 최근 기간 즉, 2000년대 중반 이후 M-wave 패턴이 특히 8월에 비약적으로 증가하였다는 것을 보였다. 이는 최근 기간에 한반도의 폭염이나 고수온이 M-wave 패턴에 더 영향을 많이 받는 것과 동시에 그 특성이 바뀌었다는 것을 의미한다. 이러한 장기 변동성의 변화는 아직 명확히 밝혀지지 않았지만, 가장 유력한 가설로서 2000년대 중반 이후 북극 빙하의 평균적인 감소에 따라 북태평양의 원격상관 패턴이 계절적으로 지연되어 나타나고, 이러한 기후편차의 변화는 엘니뇨 혹은 라니냐와 같은 적도 태평양의 기후변동성과 아열대 기후변동성의 상호작용을 강화시켜 북서태평양 원격상관 패턴의 구조가 양극자 패턴에서 삼중극자 패턴로 변화하고, 그 강도도 강화되었기 때문이라고 해석하고 있다.

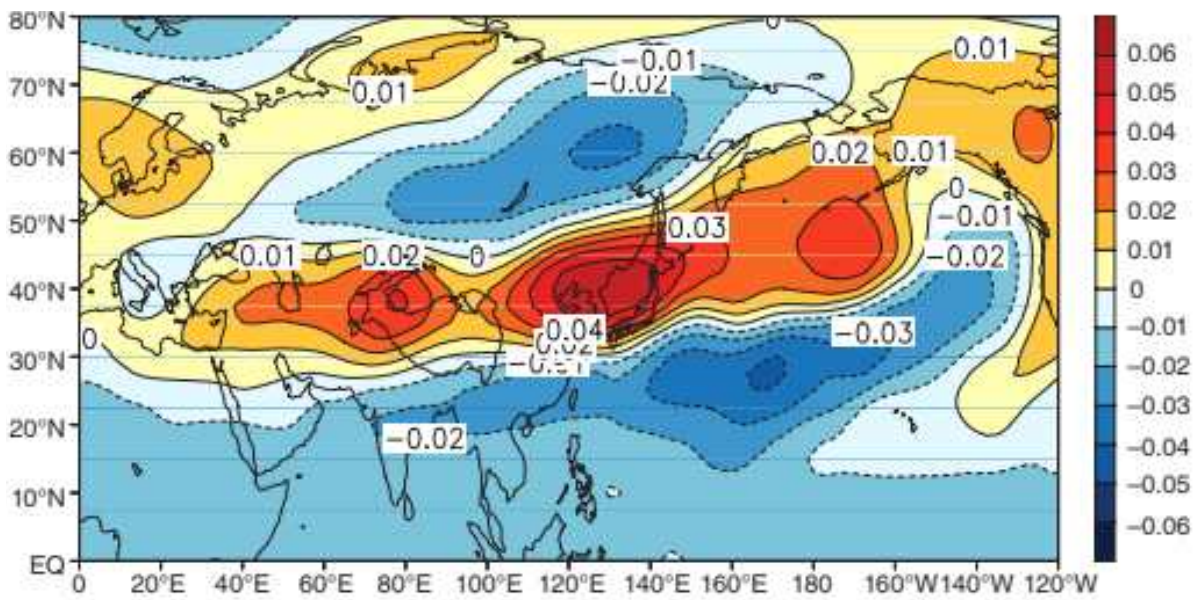


그림 3.2.12 한반도 폭염과 관련된 지구순환원격상관(Kim et al., 2018)

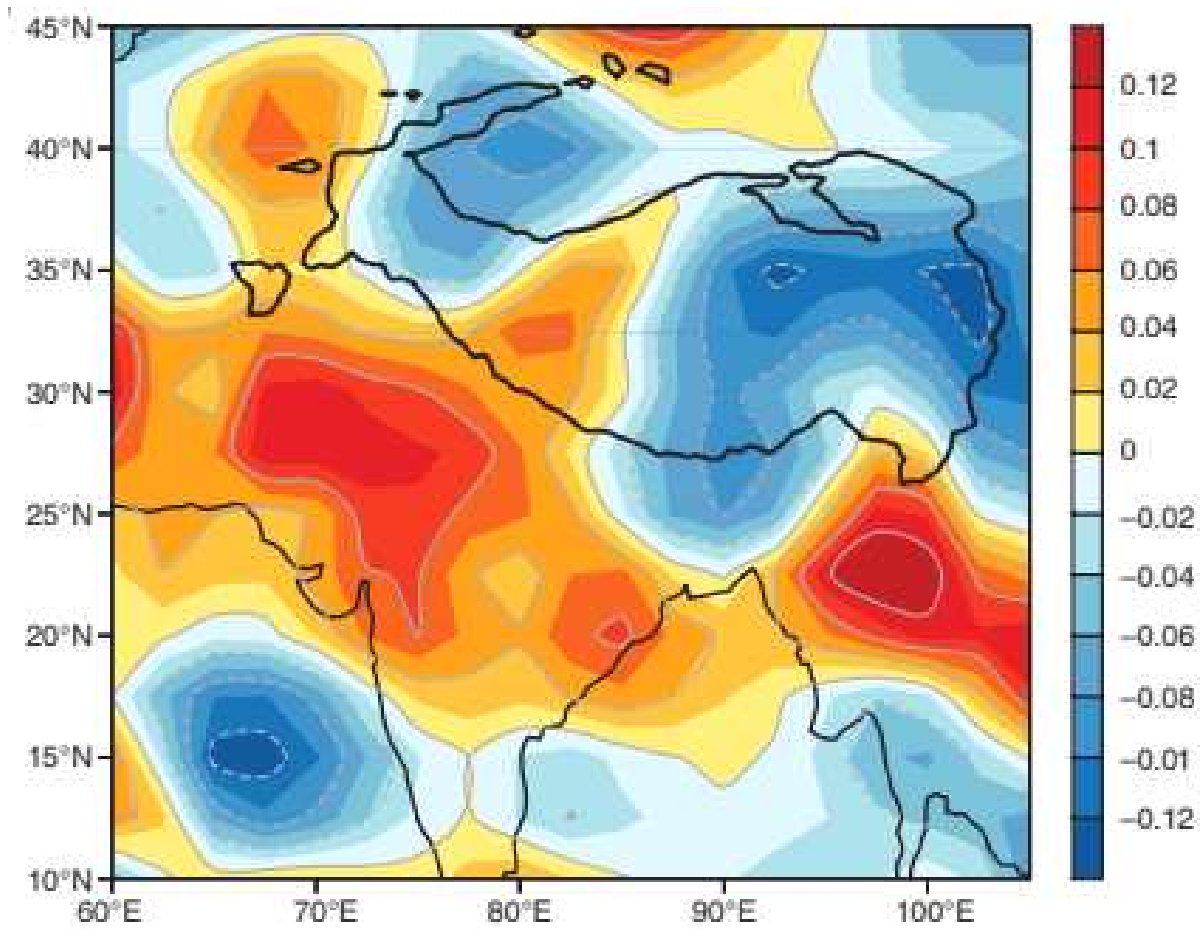


그림 3.2.13 한반도 폭염과 관련된 비단열 가열 패턴(Kim et al., 2018)

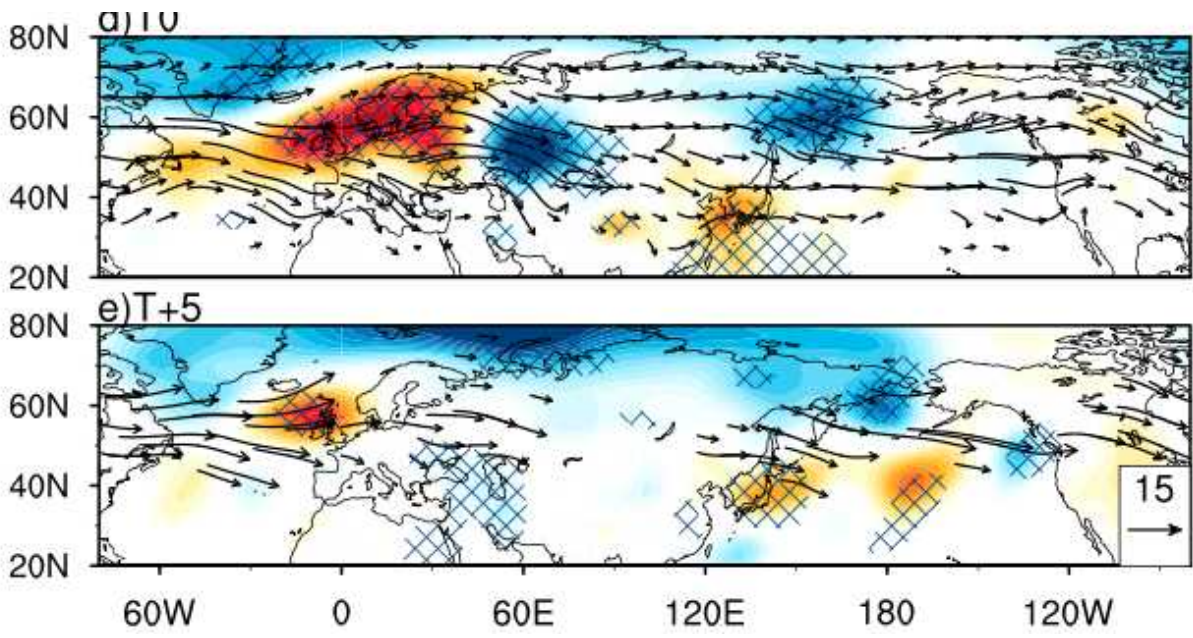


그림 3.2.14 지구순환 원격상관과 관련된 파동활동속(Yeo et al., 2018)



한반도 주변에 나타나는 고수온은 일반적으로 폭염을 동반하지만 항상 그렇지는 않다. 그림 3.2.15는 한반도에 나타나는 폭염과 고수온의 경년변동성을 보인다. 그림에서 보이듯이 한반도에 나타나는 폭염과 고수온 발생일수의 경년변동성은 일치하지 않는다. 여기서 폭염일수는 한반도 45개 관측소 일최고기온 중 상위 90% 이상의 값이 나타나는 날로 정의하였고, 고수온일수는 NOAA NCEI 일별 OISST 자료를 바탕으로 일별 해수면온도 중 상위 90% 이상의 값이 나타나는 날로 정의하였다. 고수온일수는 한반도 남해 즉, 경도 126도-129도, 위도 33도-35도를 평균한 값으로 정의하였다. 너무 넓은 영역으로 정의하거나 서해와 동해와 같이 다른 해역을 평균값에 넣게 되면, 고수온현상의 특성상 극값을 반영하기 어렵기 때문에 비교적 좁은 영역으로 고수온 현상을 정의하였다. 한반도에 폭염이 나타나는 경우 태양의 단파복사에 의한 지표면 가열을 동반하지만, 한반도에 폭염이 나타나지 않고, 남해에 고수온만 나타나는 경우 지표면 근처의 하향 단파복사를 동반하지는 않는다(그림 3.2.16). 즉, 해양의 열파는 항상 대기에 의해서 유도되고 유지되지 않을 수 있다. 그러나 이는 해양의 역학도 일부 작용한다는 것이지 여전히 해양 열파의 발생 및 유지 역학의 가장 중요한 것은 하향 단파복사이다.

한반도 해역에 나타나는 고수온 현상의 대규모 패턴은 앞서 정의한 고수온 이벤트에 대한 EOF 분석을 통해 확인할 수 있다(그림 3.2.17). 첫 번째 모드는 전체 변동성의 약 38.7% 정도를 차지하며, 한반도 전 해역에 걸쳐 같은 편차를 갖는다. 두 번째 모드는 북위 약 30도를 기준으로 남북의 쌍극자 형태의 패턴이며, 서해와 동해를 포함하는 중위도 해역과 남해를 포함하는 북서태평양 지역의 해수면온도 편차는 반대 위상을 갖는다. 이 모드는 앞서서 소개한 북서태평양 패턴과 높은 상관성을 가지며, 짧은 시간 규모로 대기에 의해 큰 영향을 받는 모드이다. 한편, 이 모드는 북서태평양에서 발생해서 북진하는 계절내진동도 유사한 패턴을 갖는다. 즉, 한반도 남해의 고수온 변동성은 북서태평양의 기후변동성과 높은 상관성을 가지며, 북서태평양 기후변동성은 잠재적 예측성을 가지고 있으므로, 한반도 남해 고수온을 예측하기 위해 북서태평양 기후변동성을 예측인자로 활용할 수 있다. 한편, 첫 번째 모드는 북서태평양의 높은 해수면온도 편차가 여름철까지 유지되는 특성을 보이며, 겨울철 엘니뇨가 여름철 라니냐로 전이되는 시기에 많이 나타난다. 엘니뇨가 라니냐로 빠르게 전이되는 시기에는 한반도에 폭염도 많이 나타나는 것으로 알려져 있다.

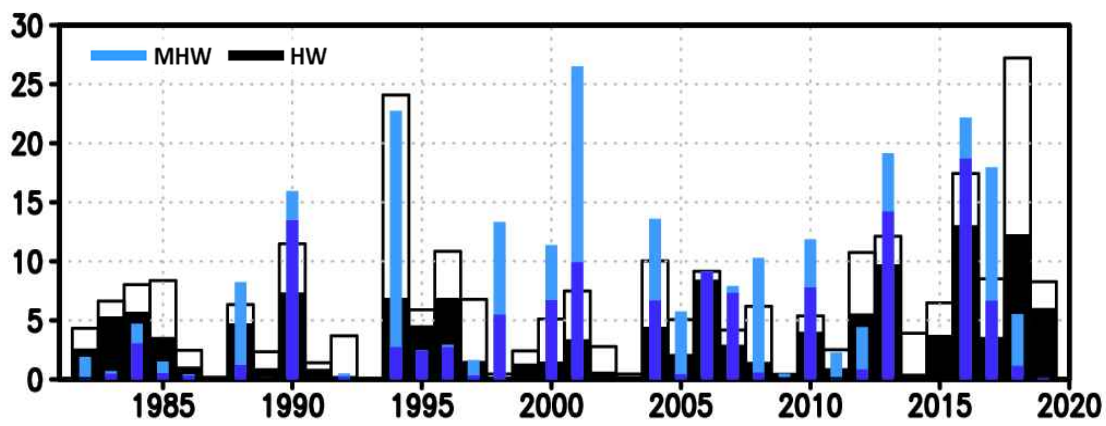


그림 3.2.15 한반도 폭염과 한반도 남해 고수온 일수

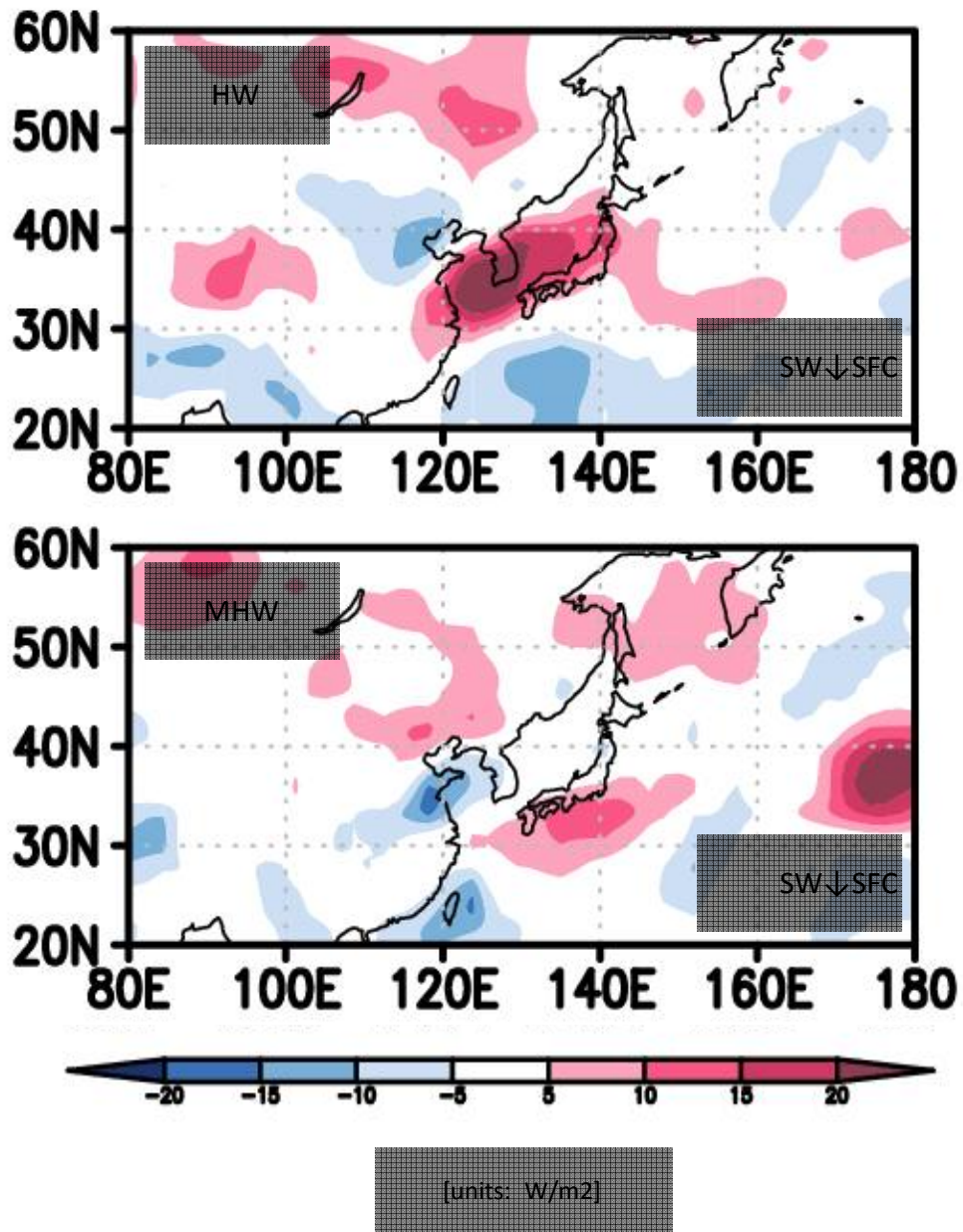


그림 3.2.16 한반도 폭염과 남해 고수온 시기 해양 단파복사

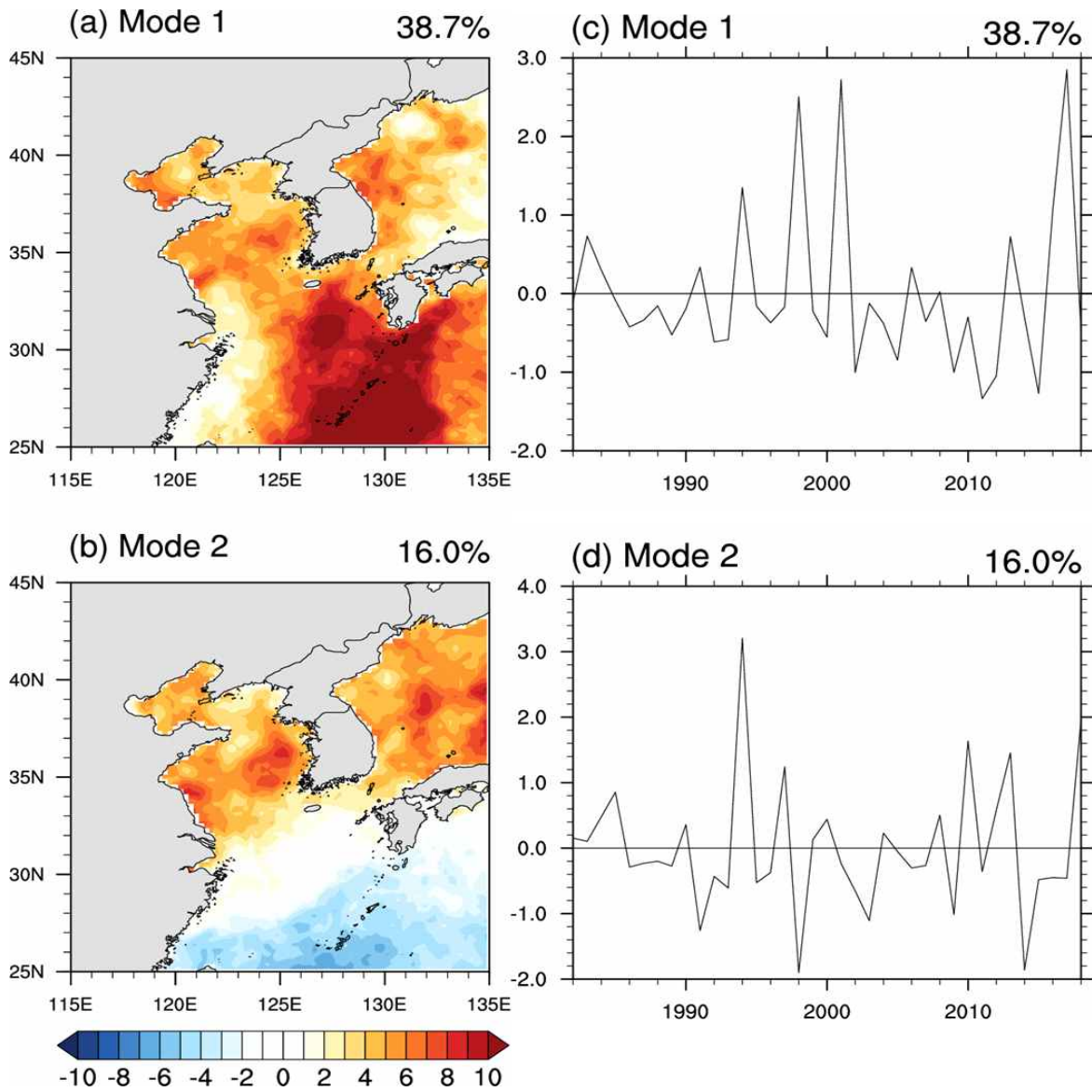


그림 3.2.17 해양 고수온의 주요 모드(Lee et al., 2020)

그림 3.2.17에서 나타나는 두 번째 모드는 30일에서 90일까지 주기를 갖는 대기-해양 기후모드와 높은 관련성을 갖는다. 열대/아열대 지역에서 그런 계절내 주기를 갖는 두 가지 기후 모드가 알려져 있다. 하나는 열대 인도양에서 대기 대류 군집이 동진하여 중태평양에서 소멸하는 특징을 갖는 Madden-Julian Oscillation(MJO)이고, 다른 하나는 아열대 인도양 및 서태평양에서 대기 대류 군집이 북진하여 아열대 및 중위도 기후에 영향을 주는 Boreal Summer IntraSeasonal Oscillation(BSISO)이다. MJO는 주로 북반구 겨울철에 그 변동성이 뚜렷하며, BSISO는 북반구 여름철에 그 변동성이 뚜렷한 특징을 갖는다. MJO는 적도의 습윤 켈빈파(Kelvin waves)에 의해서 많은 부분이 설명되며, BSISO는 동풍쉬어 메커니즘(Easterly shear mechanism)에 의해 많은 역학이 설명된다. 특히, BSISO는 북서태평양에서 북진하여 한반도를 비롯한 북동아시아 지역에 영향을 주는데, 이러한 기후모드는 중장기적으로 예측성을 가지며, 북동아시아 지역에 일주일 이상의 기간 동안 일정한 기후편차를 유지시키는 주요 기작으로 작용한다. 앞서 소개한 북서태평양의 남북모드의 중장기적 기후변동성을 만드는 주요 계절내 변동성이 BSISO이며, 이 현상은 한반도에 폭염이나 고수온이 나타나는데 중요한 역할을 할 수 있다. 한편, 인도양에서 북진하는 BSISO는 인도의 지역기후 편차와 인도 여름문순의 특징을 변화시키는데, 앞서 소개했듯이 인도 북서부지역의 대기 비단열가열은 지구순환 패턴을 통해 한반도의 폭염을 야기하므로, BSISO는 한반도 기온이나 수온편차에 매우 큰 역할을 한다. 그러므로, 본 연구에서는 고수온 예측모형을 구성하기 위하여 다음과 같이 예측인자와 예측변수를 설정하였다.

본 연구에서 선정된 예측변수(predictand)는 한반도 남해 평균(경도 126도-129도, 위도 33도-35도) 해수면온도 편차이다. 해수면온도 편차는 기후값을 제거하기 위하여 선형경향(linear trend)과 연진동(annual cycle)이 제거되었다. 예측 대상 변수로 한반도 남해 지역의 해수면온도가 선정된 이유는 서해와 동해의 고수온 특성이 다르고, 남해의 기후평균값이 다른 해역에 비하여 높아 고수온 발생시 남해 지역 생태계 및 수산업에 치명적으로 영향을 미칠 수 있기 때문이다.

본 연구에서 선정된 예측인자(predictors)는 앞서서 논의된 것과 같이 한반도 대기의 폭염과 한반도 해역 고수온 현상의 주요 역학이 유사하고, 관련된 두 가지 원격상관패턴 즉, 남북패턴과 동서패턴에 기반하여, 이 두 패턴의 기원이 되는 아열대 지역의 기후편차를 예측인자로 설정하는 것이 자연스럽다. 그러므로, 예측인자로 열대-아

열대 지역에 대하여 해수면온도 편차와 강수 및 대기 대류활동의 대체자료(proxy data)인 상향장파복사(Outgoing Longwave Radiation; OLR)를 이용하였다. 여름철 기후변동성이 대상이므로 열적도가 북반구에 있어 위도 영역을 남위 10도-북위 30도로 설정하였다. 한편, OLR 자료도 중요하지만, 중장기 메모리를 갖는 해수면온도 편차도 같이 활용하였다.

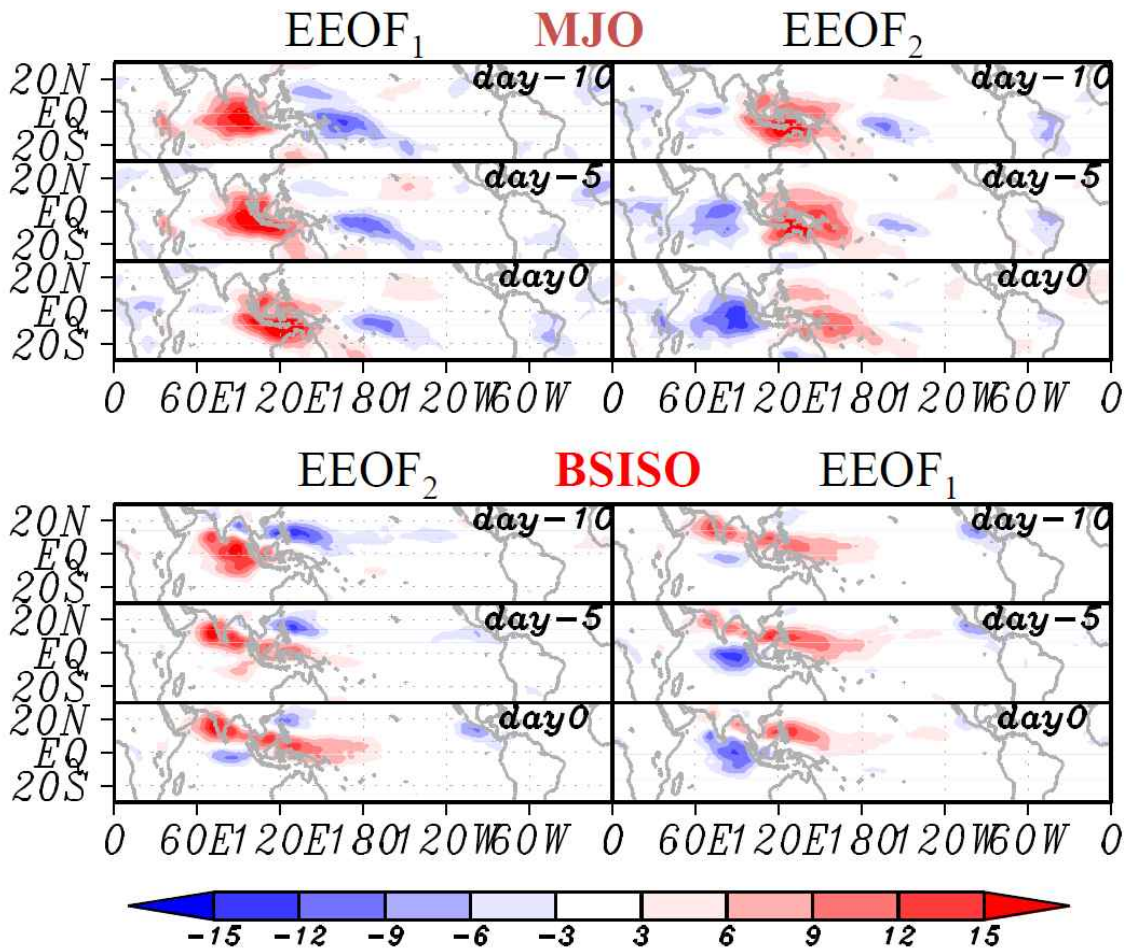


그림 3.2.18 상향장파복사의 EEOF 고유벡터 패턴(Lee et al., 2012)

### 3. X-AI 기술 이용 한반도 고수온 현상 모델링

#### 가. 분류된 고수온 이벤트와 선행인자를 이용한 기계학습기반 예측모형 개발

SST 장기변동 학습을 위한 4종류 딥러닝 베이스라인 모델 구축 및 예측 실험을 수행하였다. 모델 학습을 위한 단계 및 훈련 데이터는 다음과 같다. 우선 전이학습을 위한 사전훈련 (Pre-training)을 위해 CMIP5, monthly SST\_A, HC\_A (1000yr) 데이터를 사용하였으며, 사전학습 모델을 이용한 파인튜닝 (Fine-tuning)을 위한 재학습은 SODA, monthly SST\_A, HC\_A (140yr, 1871~2010→1980~2010), GODAS, monthly SST\_A, HC\_A (40yr, 1980-2019→1980~2010)를 사용하였다. 학습된 모델의 검증 및 테스트는 NOAA OI SST\_A (1982~2015→2011~2015) 및 ECMWF HC\_A (1982~2015→2011~2015) 자료를 사용하였다.



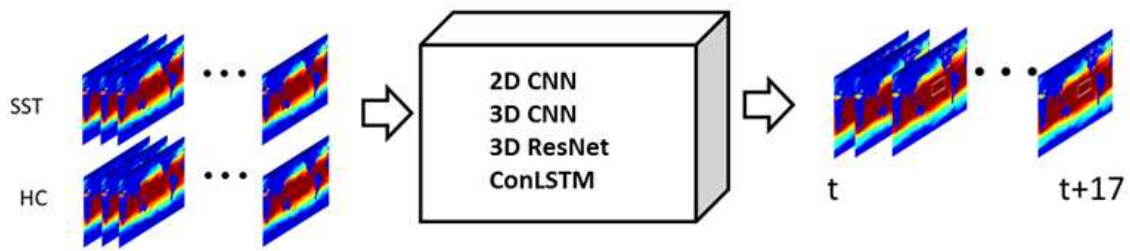


그림 3.3.1 4종류 딥러닝 베이스라인 모델을 이용한 SST 학습

다음은 상기 학습 프로세스를 따라 4가지 베이스라인 모델에 적용하고자 하며, 각 모델 구조의 특징은 다음과 같다.

[베이스라인모델-1] 심층 2D CNN (Convolutional Neural Network) 모델은 입력 데이터의 공간 정보를 유지하면서 인접 영역의 특징을 효과적으로 인식하며, 복수의 필터로 입력 데이터의 특징 추출을 위한 학습을 수행한다. 모델 구조는 다음 그림과 같다.

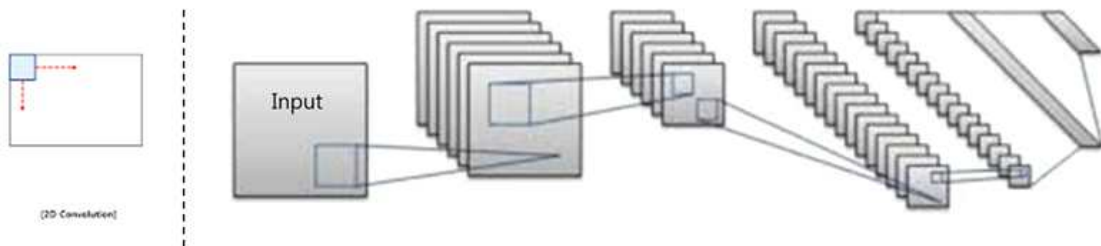


그림3.3.2 2D Convolution 연산 및 심층 2D CNN 모델 구조

[베이스라인모델-2] 심층 3D CNN (Convolutional Neural Network) 모델은 입력 데이터의 공간적 특징 뿐만 아니라 시공간 특징을 인식하는 모델 구조로 2D CNN과 비교하면 시간적 정보의 모델링이 가능한 아키텍처를 갖는다.

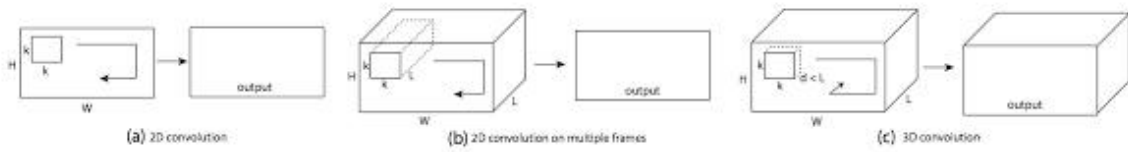


그림3.3.3 딥러닝 모델에서 2D 및 3D Convolution 연산 및 output의 차이

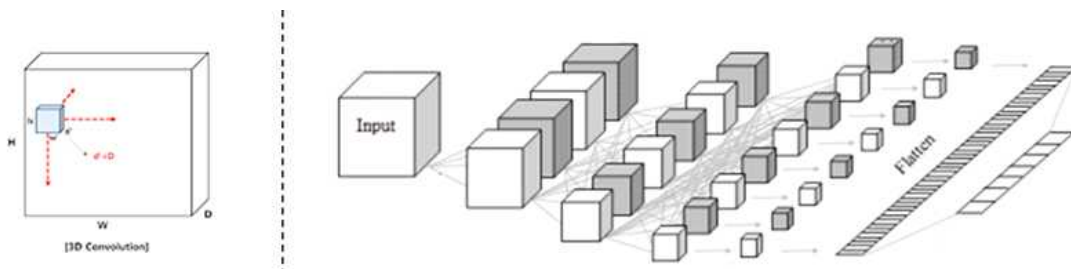


그림 3.3.4 3D Convolution 연산 및 심층 3D CNN 모델 구조

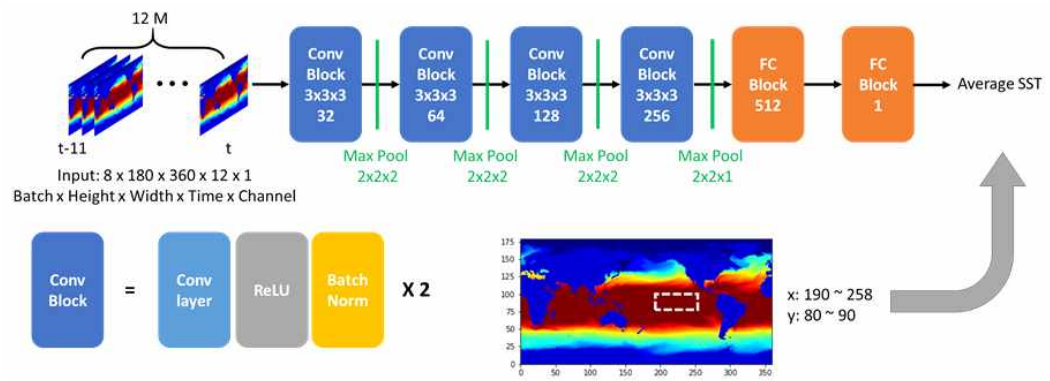


그림 3.3.5 3D SST 변동성 예측을 위한 CNN 모델 구조

[베이스라인모델-3] 심층 3D ResNet (Residual Network) 모델은 심층 신경망의 효과적인 학습을 위한 Residual learning 구조로 입력에서 출력으로 연결되는 연결을 통한 뎛셈 연산 추가로 연산량은 증가하지 않으면서 입력의 작은 변동을 쉽게 검출 가능하다는 장점을 갖고 있다.

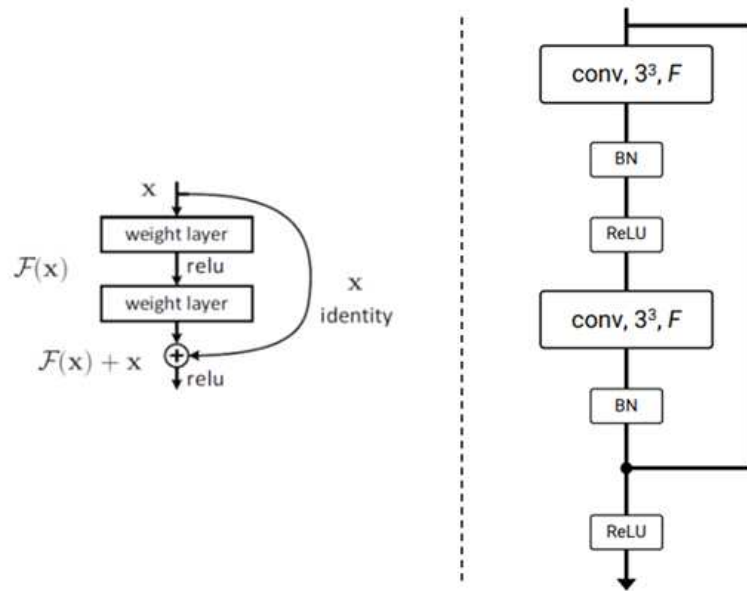


그림 3.3.6 Residual Learning 연산 및 3D ResNet 모델 기본 구조

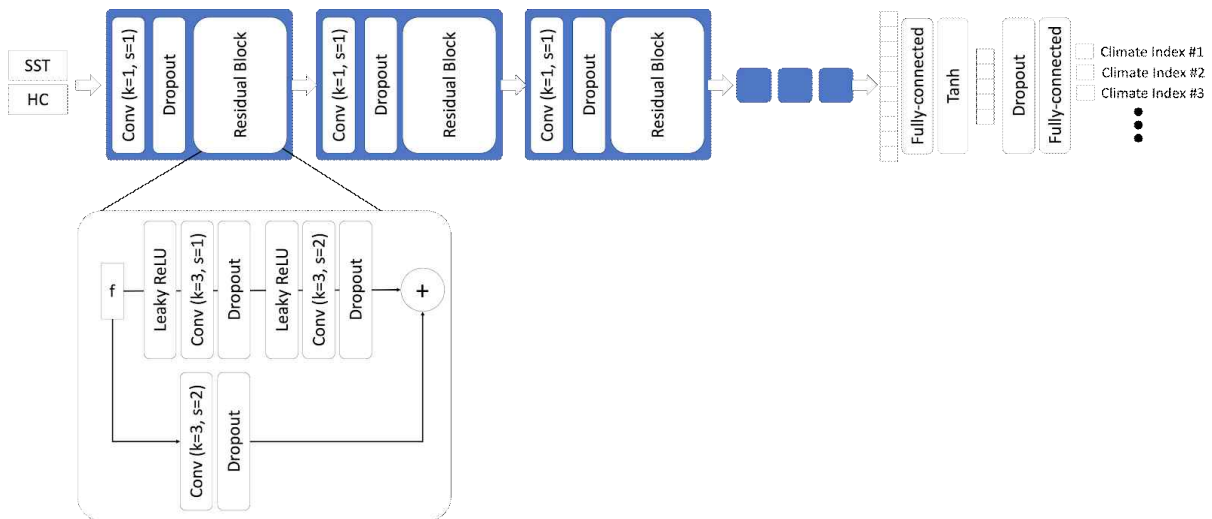


그림 3.3.6 SST 변동성 예측을 위한 3D ResNet 모델 구조

[베이스라인모델-4] 심층 ConvLSTM (Convolutional Long-term Short term Memory) 모델은 데이터의 시공간적 특징과 긴 시간 시계열 시퀀스 (Sequence) 특징을 동시에 학습하기 위한 모델 구조로 Visual feature에 대한 Sequence learning이 가능한 구조적 특징을 갖는다.

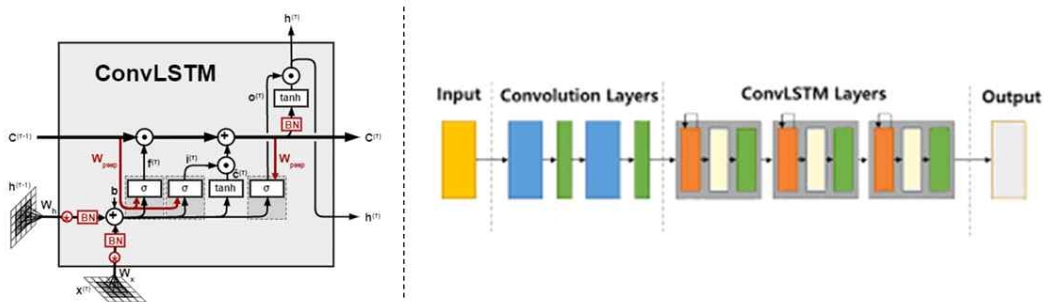


그림 3.3.7 ConvLSTM 기본 구조 및 심층 ConvLSTM 모델 구조



표 3.3.4 SST 변동성 예측을 위한 ConvLSTM 모델 구조

| Layers  | No. Filters | Size         |
|---|-------------|--------------|
| Conv Residual Block<br>(3D Conv + 2 3D Conv Block with Residual connection)                               | 8           | 37 x 36 x 12 |
| Conv Residual Block<br>(3D Conv (stride = 2, 2, 1) + 2 3D Conv Blocks with Residual connection)           | 16          | 19 x 18 x 12 |
| Conv Residual Block<br>(3D Conv (stride = 2, 2, 1) + 2 3D Conv Blocks with Residual connection)           | 32          | 10 x 9 x 12  |
| ConvLSTM Residual Block<br>(3D Conv (stride = 2, 2, 2) + 2 2D+t ConvLSTM Blocks with Residual connection) | 32          | 5 x 5 x 6    |
| Flatten   |             | 25 x 6       |
| Dense layer   |             | 3            |

이와 같은 4가지 베이스라인 모델에 대한 SST 장기변동을 학습 결과는 다음과 같다.

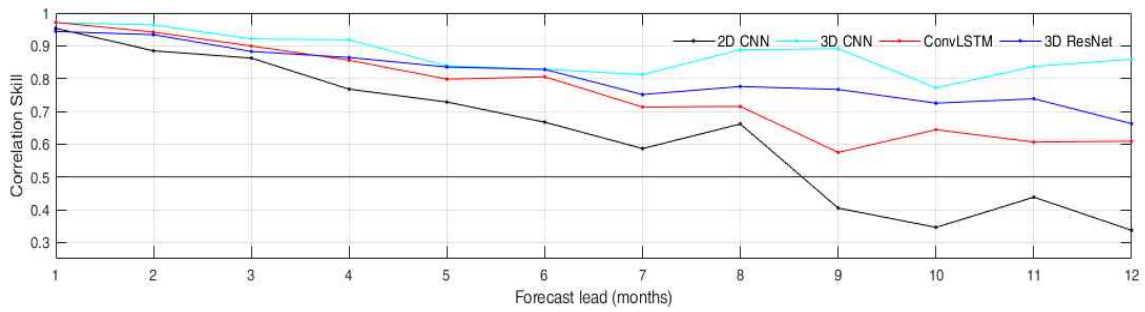


그림 3.3.8 12개월 예측에 대한 4종류 베이스라인 모델의 예측 성능

## 나. 구축된 베이스라인 딥러닝 모델과 선행인자 (OLR, SST)를 이용한 한반도 고수온 예측 모형 개발

구축된 4가지 베이스라인 모델 중 성능이 가장 좋은 Residual learning이 가능한 3D ResNet 모델을 이용하여 daily averaged OLRA (38yr, 1982~2019→1980~2011) 및 daily averaged OI SSTA (38 yr, 1982~2019→1980~2011)을 이용하여 모델을 재학습 (Fine-tuning)하고, 한반도 남해지역의 일평균 SSTA (38yr, 1982~2019→2012~2019) 자료를 이용하여 검증 및 테스트 성능을 평가하였으며, RMSE와 상관계수는 표 3.3.5와 같다.

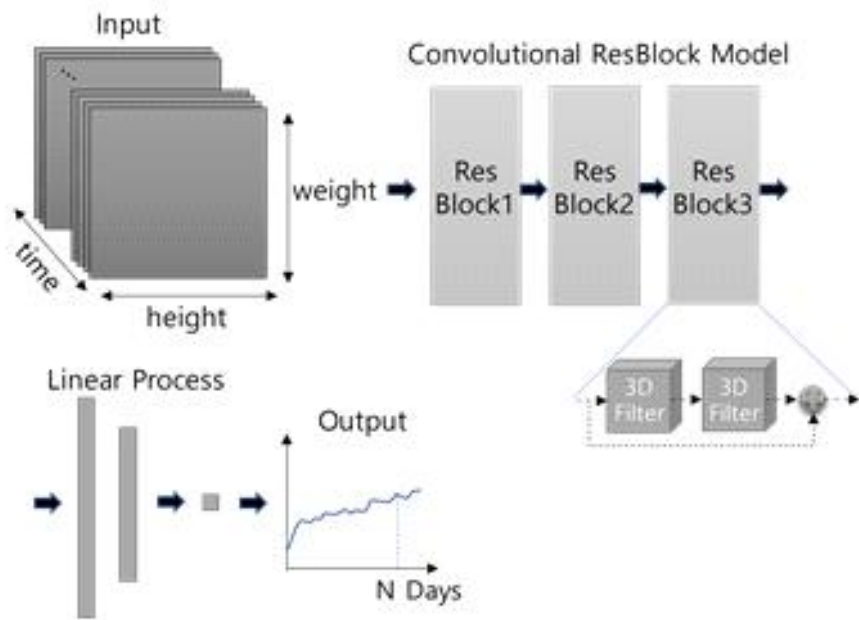


그림 3.3.8 한반도 수온 예측을 위한 3D ResNet기반 딥러닝 모델

표 3.3.5 한반도 수온 예측 성능 평가 (RMSE 및 상관계수)

| Metric            | Forecast Lead Time (day) |      |      |      |      |
|-------------------|--------------------------|------|------|------|------|
|                   | 1                        | 2    | 3    | 4    | 5    |
| RMSE<br>(제곱근평균오차) | 0.57                     | 0.59 | 0.61 | 0.64 | 0.67 |
| Corr.<br>(상관계수)   | 0.63                     | 0.58 | 0.53 | 0.50 | 0.50 |

## 다. 설명 가능한 인과관계 학습 추론 기술을 이용한 한반도 주변해역 고수온 현상 예측성 진단 및 평가

복잡한 심층신경망 기반 인공지능 모델을 설명하는데 다양한 방법들이 존재한다. 인공지능에 사용된 입력이 심층신경망 내부에서 어떤 잠재특성으로써 표현되는지를 보여줌으로써 어떤 표상들을 학습하는지 확인해 볼 수 있으며, 복잡한 인공지능을 단순화하여 내린 결정을 간단한 규칙의 집합으로서 환원하여 설명할 수도 있다.

본연구에서는 인공지능 모델에 사용된 입력의 기여도를 부여하여 중요한 특성을 추출하는데 사용되는 특성기여도 분석 방법론을 고수온 예측 인공지능 모델에 적용하고자 한다.

특성기여도 분석이란 단일 혹은 복수의 입력특성에 대해 그 특성이 출력에 대해 가지는 영향도 혹은 기여도를 정량화하는 분석 방법론으로, 인공지능에 사용되는 다양한 입력 특성들 중 해당 판단을 내리는데 긍정적 또는 부정적 영향을 준 특성들을 선별해냄으로써 인공지능이 판단을 내린 근거를 설명할 수 있다.

다양한 특성들로 이루어진 입력  $x$ 와 인공지능 모델  $f$ 가 주어졌을 때, 특성 기여도 분석 방법은  $i$ 번째 특성에 대해  $\psi(f, x, i)$ 로 나타낼 수 있는 기여도를 부여한다. 만약 이 기여도가 양수(음수)이고 그 절대값이 크다면 큰 기여를 했다는 것을 의미하며, 중요도가 0에 가까우면 해당 특성을 모델의 결과값에 큰 영향을 미치지 못했음을 의미한다.

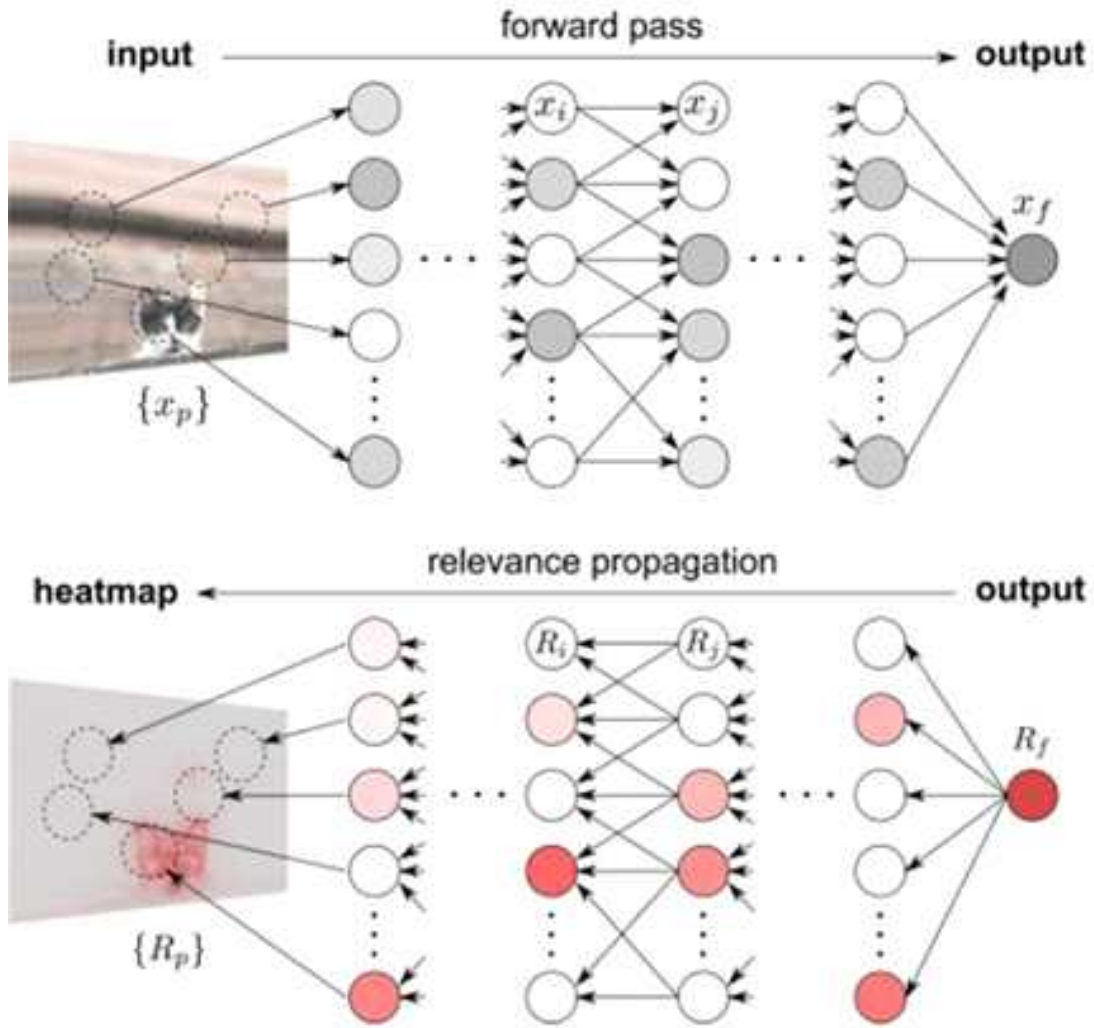


그림 3.3.9 각 계층별 관련도 전파 과정

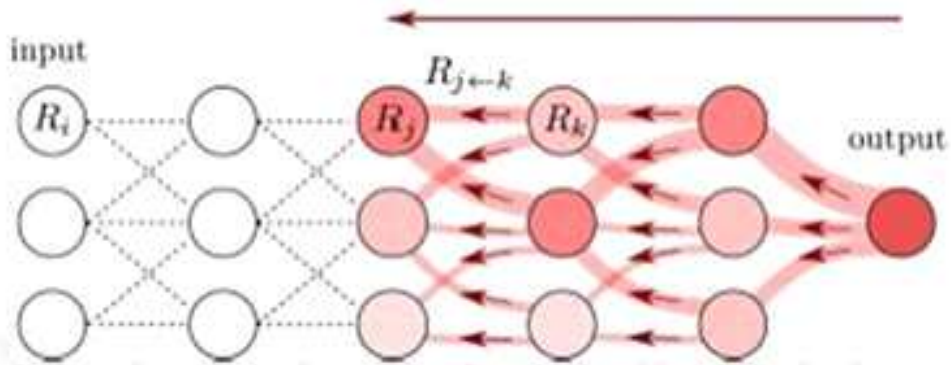
한반도 주변해역 고수온 예측에 대한 특성기여도 계산을 위해 계층별 관련도 전파 (Layer-wise Relevance Propagation, LRP)를 통해 심층신경망의 출력값을 분해, 각 계층별로 재분배함으로써 특성기여도를 계산하게 되며, LRP는 신경망에 존재하는 모든 뉴런에 대해 출력에 대한 관련도를 부여하는데, 이때 각 계층에 존재하는 뉴런들의 관련도의 합이 모델 출력값과 일치하도록 분배한다.

즉, 심층신경망  $f$ 에서  $l$ 번째 레이어의  $j$ 번째 뉴런에  $R_j^{(l)}$ 의 관련도가 부여된다고 하였을 때, 아래와 같은 식이 성립한다.

$$f(x) = \sum_j R_j^{(d)} = \sum_j R_j^{(d-1)} = \dots = \sum_j R_j^{(0)}$$

$$lrp(f, x, i) := R_i^{(0)}$$





전단계 뉴런의 활성화 함수 a와 가중치w의 곱

$$R_j = \sum_k \boxed{a_j w_{jk}^+} \boxed{\sum_j a_j w_{jk}^+} R_k$$

전단계에서 구한 관련성

출력에서 전단계까지의 모든 뉴런의 활성화 함수 a와 가중치w의 곱

그림 3.3.10 출력에서 입력방향으로 기여도의 재분배 과정

관련도 분배는 출력 레이어에서부터 시작하여 입력방향으로 한 레이어씩 역방향으로 전달되며, 위 등식에 따라 다양한 역전달 규칙이 있다. 테일러 근사법을 통해 심층신경망 상위 레이어를 하위 레이어의 일차식으로 근사함으로서 관련도를 그 기울기에 따라 재귀적으로 분배하는 방법이 있고, 이때 기울기의 근사 방식에 따라 다양한 역전달 규칙이 존재한다. 최종적으로 계산된 각 픽셀의 기여도는 그 기여도의 크기에 따라 히트맵으로 표현된다.

특성 기여도 기반 설명가능 인공지능 분석방법 적용을 통한 SST 장기변동 학습 모델 및 한반도 고수온 예측성 평가는 다음과 같이 수행하였다.

특성 기여도 분석이란 단일 혹은 복수의 입력 특성(input feature)에 대해 그 특성이 출력에 대해 가지는 영향도(influence) 혹은 기여도(contribution)을 정량화 하는 분석 방법론이다.

LRP (Layer-wise Relevance Propagation, LRP)는 계층별 관련도 전파 (LRP) 방법을 이용하여 심층신경망의 출력값을 분해하여 각 계층별로 재분배하므로써 특성 기여도(feature attribution)를 계산하게 된다. 레이어 단위로 결과에 영향을 주는 관련성을 구하는 역전파 기술로 입력 데이터 관점에서 분류 결과 뿐만 아니라 결정에 영향을 미치는 구조를 설명할 수 있다.

전지구 SST 장기변동 학습을 통한 열대 동태평양 지역 SSTA 예측결과 (Nino3.4 index)에 대한 모델별 LRP 적용과 계산된 관련도 맵(Relevance Map)을 통한 예측성 진단 및 평가결과는 다음 그림과 같으며 아래와 같은 정성적 해석이 가능하다.

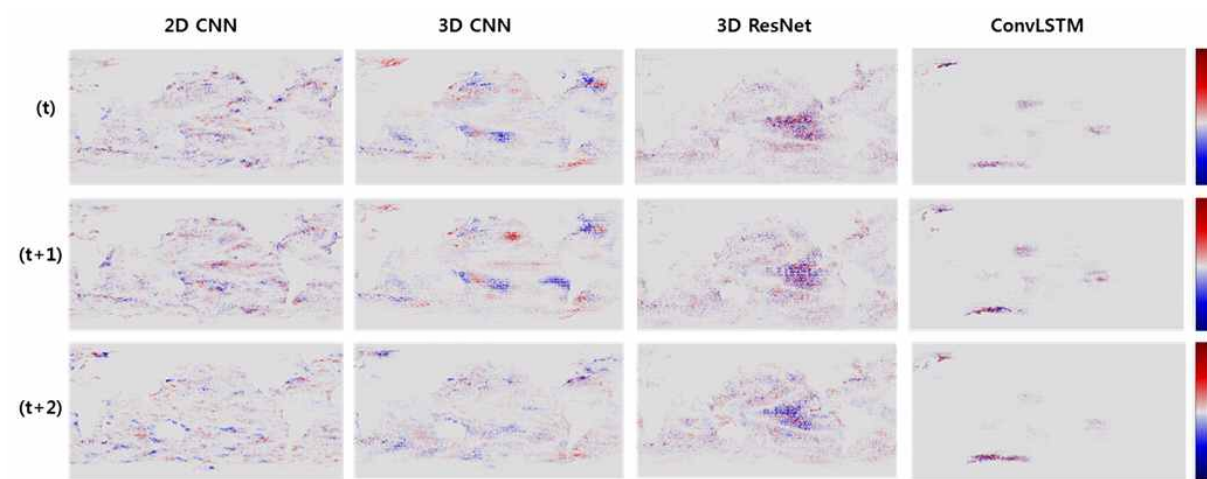


그림 3.3.11 4종류 모델의 열대동태평양 SSTA예측시 LRP 적용을 통한 관련도 맵 (Relevance Map)

2D CNN relevance map은 양과 음의 온도 편차가 섞여 전체 영역에 relevance가 분산되어 있다. 3D CNN relevance map에서는 양과 음의 온도 편차 분포가 지역적으로 구분된다. 3D ResNet relevance map, 양과 음의 온도 편차가 예측결과의 주변 영역에 주로 분포하고 있다. ConvLSTM relevance map, 양과 음의 온도 편차가 예측결과의 주변 영역을 포함한 다른 영역에도 구분되어 분포하고 있다.

3D CNN, 3D ResNet 및 ConvLSTM 모델이 예측성능이 우수하며, LRP 적용 relevance map에서 보여주듯 2011년~2015년 SSTA 예측시 열대 동태평양 지역뿐만 아니라 서태평양, 인도양 지역에서 유의미한 양의 강한 영향을 미쳤음을 설명 가능하다.

3D ResNet 이용 한반도 수온 예측 결과에 대한 LRP 적용과 계산된 관련도 맵 (Relevance Map)을 통한 예측성 진단 및 평가결과는 다음과 같다. 아래 그림은 한반도 수온 예측시 입력 데이터 중 적도 지역 SSTA의 relevance map을 보여주고 있다.

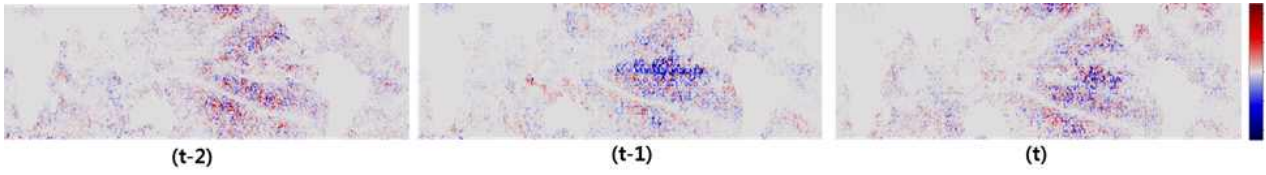


그림 3.3.12 3D ResNet 이용 한반도 수온 예측시 LRP 적용을 통한 관련도 맵 (Relevance Map)

## 제 4 장 연구개발 목표 달성도 및 대외기여도

### 1. 목표 달성도

| 성과목표                         | 달성내용   |  | 달성도 |
|------------------------------|--|--|-----|
|                              | 연구내용   | 달성실적   |     |
| 1. 해양기후변화 관련 자료 수집 및 전처리     | 1-1. 현장관측자료 및 위성자료 수집                                      | • OISTIA, G1 fndSST 등 11 종 위성자료 등 관측자료 수집 완료                       | 100 |
|                              | 1-2. 기상, 해양 재분석자료 수집                                       | • NOAA 재분석자료 등 수집 완료   |     |
|                              | 1-3. 머신러닝 수행을 위한 자료 전처리                                    | • 머신러닝 훈련자료인 SST, MLD, OLR 등 자료 전처리 완료                             |     |
| 2. 한반도 주변해역 고수온 이벤트 분석       | 2-1. 관측자료 기반 한반도 주변 고수온 이벤트 정의                             | • 선행연구에 따라 퍼센타일 기준으로 고수온 이벤트 정의 채택                                 | 100 |
|                              | 2-2. 머신러닝/딥러닝 기술을 이용한 고수온 이벤트의 객관적 분류                      | • EOF 분석 등을 이용한 고수온 이벤트 모드의 분류                                     |     |
|                              | 2-3. 머신러닝/딥러닝 기술을 이용한 고수온 현상의 선행인자 분석                      | • 고수온 현상의 선행인자로서 열대태평양 OLR과 SST 인자 추출 완료                           |     |
| 3. X-AI 기술 이용 한반도 고수온 현상 모델링 | 3-1. 분류된 고수온 이벤트와 선행인자를 이용한 기계학습기반 예측모형 개발                 | • SST 장기변동 학습을 위한 딥러닝 기반 베이스라인 모델 구축<br>• 한반도 고수온 예측을 위한 딥러닝 모델 개발 | 100 |
|                              | 3-2. 설명 가능한 인과관계 학습 추론 기술을 이용한 한반도 주변해역 고수온 현상 예측성 진단 및 평가 | • Feature attribution 기반 설명가능 인공지능 분석 방법 적용을 통한 한반도 고수온 예측성 평가     |     |

## 2. 대외기여도

지구온난화를 비롯한 기후변화에 따라 기상재해 등 자연재해에 의한 사회적·경제적 피해가 꾸준히 증가되고 있다. 그렇기 때문에 보다 정확한 중장기 예측기술의 개발은 피해저감 대책을 마련하거나 자연재해에 의한 피해 자체를 예방하는데 매우 큰 역할을 한다. 이러한 연구의 중요한 사회적 기여는 기상/해양 재해에 대한 대응에 보다 능동적으로 대처할 수 있는 기술력을 확보하는 것이다.

지구온난화와 기후변화, 혹은 장기 기후변동성이 중요한 사회적 문제로 대두되면서, 짧은 시간 규모에 대해서는 수십일 예측에서 긴 시간 규모는 100년의 기후변화 예측까지 기후예측에 대한 예측정보는 과학적 이해 뿐 아니라 사회·경제적 분야에 중요한 기반 정보로 활용될 것이다. 기후예측의 방법은 다음과 같이 크게 세 가지로 나눌 수 있다.

첫째, 초기화과정이 포함된 역학모형을 활용하는 방법

둘째, 관측자료만을 활용하는 자료기반 통계모형을 이용하는 방법

셋째, 역학모형의 계통적 편이(bias)를 통계적으로 보정하는 MOS(Model Output Statistic) 방법

역학모형의 중위도 기후의 예측성은 현저히 낮으며, 통계모형의 예측성은 그 예측성이 불안정하여, 현업에서는 이 둘의 결과를 상호보완적으로 활용하여야 한다. 본 과제에서 개발된 AI 모델은 자료기반 통계 예측모형 중 하나로서 불안정한 통계모형의 약점을 크게 보완하고, 그 예측성 자체도 크게 향상시킨 것이다. 이러한 모델의 개발은 비단 신뢰성 있고 활용도가 높은 예측기술 개발을 넘어, 보다 여름철 한반도 주변 기후변동에 대한 물리적 이해를 이끄는 핵심 기술력이 될 것이다.

한반도 뿐 아니라 중장기 기후예측기술은 지구온난화를 비롯한 기후변화와 관련하여 그 중요성이 점점 증가되고 있으며, 여름철 극한 현상 즉, 고수온, 폭염 발생과 같은 고영향 현상의 발생 역시 전 세계적으로 증가하는 추세에 있기 때문에, 예측 모델 자체를 활용하는 것 뿐 아니라 예측 모델 개발의 근간이 되는 이러한 기반 연구는 중요한 이슈이다.

동아시아 지역은 인구밀집 지역이기 때문에 동아시아 지역 기후 변동 연구에 대한 연구의 중요성 및 필요성은 매우 높으며, 향후에도 이와 관련된 연구는 계속 활성화 될 것이다. 이와 관련하여 최근 열대-중위도 상호작용에 대한 영향력 높은 논문들이 많이 출판되고 있다. 본 연구에서 얻은 과학적 성과는 열대-중위도 상호작용을 이해 하는 새로운 연구를 이끌 것으로 전망한다.

한반도 주변 고수온 예측을 위한 베이스 모델로서 엘니뇨 예측모형이 개발되었으며, ResNet 등으로 개선된 모형은 선행연구(Ham et al., 2019, Nature)에서 제시된 예측성보다 유의미하게 높은 예측 성능을 보여 기존의 모든 종류의 엘니뇨 예측모형 (수치모형, 통계모형) 보다 높은 예측성능을 가지는 것으로 나타났다.

엘니뇨 베이스 모델을 이용하여 개발된 한반도 주변 고수온 예측을 위한 딥러닝 기후 예측모형은 기후모형의 낮은 예측성능을 극복하는 자료기반 모형으로 평가함으로써 AI 기술을 이용하여 기후예측 융합연구를 활성화시키는 기틀을 만들었다.



## 제 5 장 연구개발결과의 활용계획

본 연구에서 수집한 해양관측자료, 인공위성자료, 재분석자료는 기후변화 관련 타 연구에도 활용 가능하다. 고수온 및 엘니뇨 예측에 사용한 머신러닝용 학습데이터는 다른 AI 모델 개발에서도 활용할 수 있다.

설명가능 인공지능(X-AI) 기술을 이용하여 한반도 주변해역 고수온 이벤트의 예단 인자를 찾고, 데이터기반 모델 개발을 통해 예측성을 진단하고 평가하며, 발굴된 고수온 예단인자는 고수온 현상의 역학적 이해 및 모형 개발에 활용가능하다.

또한 AI+해양순환·기후변화 융합연구 활성화를 위한 AI 기술력 확보하였고, 고수온 예측모형은 고수온 재해 저감 방안 마련에 활용될 것으로 기대된다. 나아가 차세대 인공지능기술과 KIOST 해양과학분야 융합연구 활성화 기여함으로써 기후변화 등 장기 해양변동성 분석과 연안 재해재난 예측·대응 등 해양 빅데이터 활용 극대화 및 차세대 인공지능기술 연구에 부응할 것으로 기대한다. 이는 해양과학 분야, 정부 'I-KOREA 4.0' AI 융합 기술개발 전략 제시 및 이를 통한 관련 정부부처, 지자체, 유관기관, 산업체 등 신규 연구과제 및 해양 지능정보산업 발굴에 기여할 것으로 기대하고 있다.

특히 본 연구 결과는 사회·경제적으로 막대한 영향을 미치는 고수온 현상 예측을 위하여 인공지능 기술을 도입함으로써, 해양분야와 데이터기술 분야의 융합연구체계 구성을 통해 기존에 찾지 못했던 예단 인자나 현상에 대한 새로운 이해가 가능하다. 이는 사회·경제적으로 막대한 영향을 미치는 고수온 현상 예측을 통해 (수온 예측 모형) 고수온 재해 저감 방안 마련에 활용될 것이며, 국가사회적 해양과학 기술 수요 예측 및 대응 연구에 대응할 것으로 보여진다.

## 제 6 장 참고문헌

김진용, 서경환, 예상욱, 김현경, 임소영, 이현수, 함유근, 2017: 2016년 한반도 장마 강수 특성 분석. *대기*, **27(3)**, 277-290.

ADPRC, 2020, [http://apdrc.soest.hawaii.edu/datadoc/soda\\_2.2.4.php](http://apdrc.soest.hawaii.edu/datadoc/soda_2.2.4.php)

Barnett, T.P. and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review*, 115, 1825-1850.

Berkelmans, R., and Van Oppen, M. J. H., 2006: The role of zooxanthellae in the thermal tolerance of corals: a ‘nugget of hope’ for coral reefs in an era of climate change. *Proc. R. Soc. B Biol. Sci.* 273, 2305-2312. doi: 10.1098/rspb.2006.3567.

C-GLORS, 2020, <http://c-glors.cmcc.it/index/index.html>

Ding, Qinghua, and Bin Wang, 2005: Circumglobal teleconnection in the northern Hemisphere summer. *Journal of Climate*, 18, 3483-3505.

Ham, Yoo-Geun, Jeong-Hwan Kim, and Jing-Jia Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature* 573.7775, 568-572.

Ham, Yoo-Geun, YeonJi Hwang, Young-Kwon Lim, MinHo Kwon, 2017: Inter-decadal variation of the tropical Atlantic-Korea teleconnection pattern during boreal summer season. *Climate Dynamics*, 51, 2609-2621.

HCEI OHC, 2020, <https://www.ncei.noaa.gov/access/global-ocean-heat-content>

Hobday, A., L. Alexander, S. Perkins-Kirkpatrick, and D. Smale, 2016: A hierarchical approach to defining marine heatwaves. *Progress in Oceanography*, 141, DOI:10.1016/j.pocean.2015.12.014.

IOC, SCOR, and IAPSO, 2010: The International Thermodynamic Equation of Seawater: Calculation and Use of Thermodynamic Properties. Intergovernmental Oceanographic Commission, Manuals and Guides No. 56. Paris: UNESCO, 196.

- Kim, Han-Kyoung, Byung-Kwon Moon, Maeng-Ki Kim, and MinHo Kwon, 2020: Dynamic mechanisms of summer Korea heat waves simulated in a long-term unforced Community Climate System Model version3, *Atmospheric Science Letters*, 2020;21:e973.
- Kim, Maeng-Ki, Ji-Seon Oh, Cheol-Kyu Park, Seung-Ki Min, Kyung-On Boo, Jeong-Hun Kim, 2018: Possible impact of the diabatic heating over the Indian subcontinent on heat waves in South Korea. *International Journal of Climatology*, 39, 1166-1180.
- Kosaka, Y., and H. Nakamura, 2007: Structure and dynamics of the summertime Pacific-Japan teleconnection pattern. *Quarterly Journal of the Royal meteorological Society*, 132, 2009-2030.
- Lee, June-Yi, Bin Wang, Matthew Wheeler, Xiuhua Fu, Duane Waliser, and In-Sik Kang, 2012: Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region. *Climate Dynamics*, 40, 493-509.
- Lee, Seonju, Myung-Sook Park, MinHo Kwon, Young Ho Kwon, Young-Gyu Park, 2020: Two major modes of East Asian marine heatwaves. *Environmental Research*, 15, 074008.
- Lee, Woo-Seop, and Myong-In Lee, 2016: Interannual variability of heat waves in South Korea and their connection with large-scale atmospheric circulation patterns. *International Journal of Climatology*, 36, 4815-4830.
- Meysigian, Benoit et al., 2019: Measuring Global Ocean Heat Content to Estimate the Earth Energy Imbalance, *Frontiers in Marine Science*, 6, 432
- Nam, W., S. Gur, J. Choi, L. Wolf, and S. Lee, “Relative attributing propagation: Interpreting the comparative contributions of individual units in deep neural networks,” in AAAI. AAAI Press, 2020.
- NCEI ENSO, 2020, <https://www.ncdc.noaa.gov/teleconnections/enso/indicators/sst/>
- NOAA OLR, 2020, [https://psl.noaa.gov/data/gridded/data.interp\\_OLR.html](https://psl.noaa.gov/data/gridded/data.interp_OLR.html)
- Wernberg, T., Bennett, S., Babcock, R. C., de Bettignies, T., Cure, K., Depczynski, M.,

et al., 2016. Climate-driven regime shift of a temperate marine ecosystem. *Science* 353, 169–172. doi: 10.1126/science.aad8745.

Yeo, Sae-Rim, Sang-Wook Yeh, and Woo-Seop Lee, 2018: Two types of heat wave in Korea associated with atmospheric circulation pattern. *Journal of Geophysical Research*, 124, 7498–7511.