

이상기후 예측을 위한 인공지능 기술 개발(I)

부제: 동아시아 이상기후 계절내 예측을 위한 인공지능 모델 원형 개발

Exploring Artificial Intelligence Techniques for Better Forecast of Climate Extremes (I)

김미애, 이운영, 정유란

기후사업본부 예측기술개발과

2025. 12. 12.

APEC기후센터

Executive Summary

The development of intraseasonal extreme climate prediction technologies has become increasingly urgent in light of the growing frequency of climate-related hazards under global warming. This project aims to establish prototype AI models capable of predicting East Asian extremes—anomalous high temperature, heavy rainfall, and marine heatwaves—with lead times of three to four weeks. For anomalous high temperature, we adopt a bias-correction approach based on ECMWF (European Centre for Medium-Range Weather Forecasts) ensemble forecasts, while heavy rainfall and marine heatwaves are addressed through direct prediction using the latest available observations. The ECMWF dynamical model serves as the baseline, against which we identify AI model setups that demonstrate statistically significant improvements.

First, this study investigates the sensitivity of anomalous high-temperature detection within the ECMWF Subseasonal-to-Seasonal (S2S) prediction system, focusing on how different labeling strategies influence model performance. Using 3-week averaged maximum temperature (TMAX) data from ECMWF-S2S (version 2024) and ERA5 reanalysis, binary and multi-class labeling schemes were applied to classify extreme heat defined by percentile thresholds (≥ 90 th percentile for binary; ≥ 75 th and < 90 th percentile, and ≥ 90 th percentile for two classes of multi-class). These labels were used to train two convolutional neural-network architectures—U-Net and Attention U-Net—designed to detect spatial patterns linked to anomalous heat events. The models were trained using 18 years of data (2004–2021) and tested using two years (2022–2023). An evaluation based on the Brier Skill Score (BSS) indicates that while the raw ECMWF forecasts exhibit modest skill (~ 0.2), the deep-learning models generally yield lower absolute scores but still provide meaningful spatial improvements relative to ECMWF. The binary-label Attention U-Net configuration consistently yields the most realistic probability patterns and the highest fraction of positive BSS improvements, demonstrating its relative advantage in extreme-heat signal detection. The results suggest that label structure plays a decisive role in model sensitivity and that attention mechanisms enhance the identification of high-impact spatial features.

Next, this study developed an initial deep learning model for forecasting heavy rainfall events at lead times of 3–4 weeks on a grid-point basis. The research focused on three aspects: the effectiveness of multi-task learning, the combination and scale decomposition of input variables, and the impact of model architecture on forecasting performance. The multi-task learning approach, which jointly predicts precipitation amounts and the number of heavy rainfall days in a week, effectively captured the temporal variability of precipitation and achieved improved performance metrics (e.g., CSI, TCC). Including 10–60-day and long-term filtered components further enhanced forecast skill, indicating the importance of separating and incorporating intraseasonal variability and background fields for extreme

rainfall prediction. The combined ResNet-LSTM architecture demonstrated synergistic effects in learning spatiotemporal patterns. Future work will incorporate probabilistic forecasting techniques to quantify prediction uncertainty improving predictability and apply explainable AI (XAI) methods to identify and analyze the key predictors and mechanisms underlying extreme rainfall forecasts.

As the final climate extreme component, a prototype prediction model for marine heatwaves in the East Asia Marginal Sea(EAMS) is developed. Predictive signals are extracted from large-scale oceanic and atmospheric reanalysis fields—including heat content, salinity, winds, temperature, and precipitation—and projected onto East Asian marine heatwave patterns three weeks ahead using a UNet architecture. Sensitivity experiments on predictors, loss functions, and model design identified an optimal setup: a relatively narrow 2.5° regional domain, inclusion of fluxes, static variables, and marine heatwave indicators alongside the basic predictors, and training a swinUNet with a 2:1 train/validation split. Detection skill was further improved by applying a weighted loss function emphasizing extreme SST grids and integrating river discharge information as an independent multi-modal layer. Across 52 weeks in 2023, this setup consistently outperformed ECMWF in marine heatwave detection. Ensemble means of high-performance models under similar conditions also showed stable and significant skill, highlighting the promise of multi-model AI ensembles for operational prediction in East Asia.

In order to systematize and disseminate data and statistical analyses derived from prototype model development, a repository (EastAsiaClimateExtremes, <https://github.com/yyalexlee/EastAsiaClimateExtremes>) was employed. This repository provides historical data on climate extremes across the East Asia region, basically archiving the events of regular grid-based climate extremes, and weekly/monthly extremeness metrics can be utilized as labels for AI-based(or -assisted) models to predict anomalous climate events in East Asia. Additionally, it offers code for long-term, grid-based analysis and visualization of major climate extremes with functions for data storage, time-series graphs, heatmaps, and two-dimensional long-term statistics. Currently, the repository focuses on extreme phenomena such as anomalously high temperatures(AHT), heavy rainfall(HR), and marine heatwaves(MHW), with the potential for further expansion. This work is intended to serve as a meaningful and practical foundation for researchers engaged in the investigation of the dynamics and prediction of extreme phenomena, offering a structured entry point that facilitates subsequent methodological development and interdisciplinary collaboration.

국문 요약

기후변화로 인해 날로 빈번해지는 기후 관련 위험·재난에 대비하기 위한 계절 내 시간 규모의 이상기후 예측 기술 개발은 매우 시의적절하다. 본 과제는 동아시아 영역의 이상고온(Anomalously High Temperature), 폭우(Heavy Rainfall), 해양열파(Marine HeatWave)에 대한 선행 3-4주 주별 예측 인공지능 기반 원형 모델 개발을 목표로 한다. 이상고온은 유럽중기예보센터(ECMWF, European Centre for Medium-Range Weather Forecasts) 예측 앙상블의 오차 보정 방식을 적용하고, 폭우·해양열파는 최신 가용 관측 자료를 활용한 직접 예측 방식으로 접근하였다. ECMWF 역학 모델 성능을 기초선으로 삼아, 유의미한 성능을 보이는 인공지능 모델 setup을 탐색하고 그 결과를 제시한다.

이상고온의 경우 ECMWF S2S 예측 앙상블 오차 보정을 통한 이상치 탐지 성능이 라벨링 방식에 따라 어떻게 달라지는지를 정량적으로 평가하였다. ECMWF-S2S(2024 버전)의 3주 평균 최고기온(TMAX)과 ERA5 재분석 자료를 대상으로, 90백분위수 기반의 이진 라벨과 75·90백분위수 기반의 다중 클래스 라벨을 각각 적용하였다. 이후 U-Net과 Attention U-Net 두 가지 딥러닝 구조에 훈련하여, 고온 패턴의 공간 확률 분포를 예측하도록 설계하였다. 18년(2004~2021)의 학습 기간과 2년(2022~2023)의 테스트 기간을 설정하여 모델을 평가한 결과, ECMWF 자체 Brier Skill Score(BSS)는 약 0.2 수준으로 낮았으나, 딥러닝 모델은 전체적으로 낮은 BSS에도 불구하고 ECMWF 대비 양(+)의 개선이 나타나는 격자들이 존재하였다. 특히 이진 라벨 + Attention U-Net 조합이 관측 패턴과 가장 유사한 확률 분포를 재현하며, BSS도 상대적으로 높은 개선 폭을 보였다. 이는 고온 탐지 문제에서 라벨 정의와 Attention 메커니즘이 매우 중요한 역할을 함을 시사한다.

다음으로, 격자점 단위의 선행 3-4주 폭우를 예측하기 위한 딥러닝 기반 초기 모형을 개발하였다. 연구는 Multi-task learning의 효과, 변수 조합과 스케일 분해 방식, 모델 구조가 예측 성능에 미치는 영향을 중심으로 수행되었다. Multi-task learning 기법은 강수량과 극한 강수일수를 함께 학습함으로써 강수의 시간적 변동성을 효과적으로 포착하고 주요 예측 성능 지표(CSI, TCC 등)에서 개선된 결과를 보였다. 또한 10-60일 대역 및 장주기 필터링 변수를 포함할 때 예측력이 향상되어 계절내 변동성과 배경장 변수를 분리하여 활용하는 것이 극한 강수 예측에 중요함을 확인하였다. ResNet에 LSTM을 결합한 구조는 시공간 패턴 인식에 강점을 보여 두 모델의 결합 효과가 검증되었다. 향후에는 확률 예측과 설명가능한 인공지능(XAI) 기법을 도입하여 예측의 불확실성을 정량화하고 폭우 주요 예측 요인을 분석하고자 한다.

마지막으로, 동아시아 해역 고수온 예측을 위해 이 지역을 포함하는 광역 지역의 해양 열용량·염분, 대기 표면 바람·온도, 강수 등 해양·대기 기본 관측 정보를 학습하여 주요 시그널을 추출하고, 이를 3주 이후의 동아시아 고수온 패턴으로 확장하는 UNet 구조의 원형 모델을 고안하였다. 개발 과정에서는 훈련 변수뿐 아니라 손실 함수와 모델 구조에 대한 다양한 민감도 조건을 적용하고, 교차 조건이 반영된 모든 모델을 사후에 종합 비교하여 선행 3주 동아시아 해역 고수온 예측의 최적 setup을 도출하였다. 결과적으로, 2.5도 해상도의 비교적 좁은 지역 도메인에서 기본 변수 외에 플릭스·정적 변수·고수온 해양 정보를 적층하고, 이를 2:1의 훈련/검증 샘플 비율로 swinUNet 모델에 학습·검증하는 방식이 가장 우수한 고수온 탐지

력을 보였다. 또한 고수온 현상이 나타난 격자에 가중치를 부여하는 손실 함수 적용과, 하천유량 정보를 독립된 레이어로 추가하는 멀티모달 방식이 성능 개선에 기여하였다. 제시한 setup의 모델들은 2023년 52주 테스트 기간 동안 평균적으로 ECMWF를 상회하는 고수온 탐지 성능을 보였다. 또한, 비슷한 조건의 고성능 개별 모델들의 앙상블 평균이 안정적이고 유의미한 성능을 보여 향후 다중 모델 인공지능 앙상블 예측의 가능성을 확인할 수 있었다.

원형 모델 개발 과정에서 얻어진 동아시아 기후 극한현상에 대한 자료와 통계 분석 결과를 체계화하고 공유하기 위해 저장소(GitHub)를 활용하였다. 이 저장소는 동아시아 지역(21-48° N, 114-141° E)을 대상으로 한 기후 극한현상 저장소(EastAsiaClimateExtremes, <https://github.com/yyalexlee/EastAsiaClimateExtremes>)를 구축하고, 이상고온(AHT), 폭우(HR), 해양열파(MHW) 등의 주요 극한 현상을 장기간 및 격자 기반으로 정량화한 자료와 시각화 코드를 제공한다. ERA5, OISST, ECMWF-hindcast(버전 2016 및 2024)를 활용하여 일별 및 주별 단위의 기온, 강수, 해수면온도 자료의 원시 시계열, 기후 평년값, 90/95백분위 임계값을 산출하였다. 이를 기반으로 지속일수(D), 허용간극(G) 규칙(D3G5 등)을 적용해 극한 이벤트의 발생 빈도, 지속기간, 평균 및 최대 강도, 영향지수 등 세부 프로파일과 주 단위 extremeness 지수(Extreme Days, Mean Intensity, Max Intensity, Impact Factor)를 생성하였다. 또한, 제공된 분석 코드는 시계열 그래프, 열지도(heatmap) 및 장기 평균 2차원 분포 등 다양한 형태의 시각화를 지원한다. 데이터(DATA) 및 코드(CODES) 인벤토리는 동아시아 기후 극한의 기작 분석, 장단기 예측, 그리고 AI 기반 기후 예측 모델의 학습 및 라벨링 자료로 활용될 수 있는 기반 자료로서 구축되었다. 해당 자료와 도구의 통합 제공을 통해 기후 극한현상의 특성 진단과 예측 정확도 향상 연구에 실질적인 기여가 가능할 것으로 기대한다.

목차

1. 서론	1
2. 방법 및 자료	6
2.1. 이상기후 예측 타깃 선정	6
2.1.1. 예측 대상 이상기후 현상 정의	6
2.1.2. 이상기후 시·공간 예측 타깃	7
2.2. 훈련 자료 및 전처리	8
2.2.1. 관측 및 역학 모델 원자료	8
2.2.2. 이상기후 예측 모델 입력 변수 및 라벨 선정	9
2.3. 인공지능모델 및 실험 구성	19
2.3.1. 동아시아 선행 3/4주 이상고온 역학 예측 오차 보정 모델 개발	19
2.3.2. 격자별 주 단위 폭우 강도 및 빈도 예측을 위한 multi-task learning 기반 모델 개발	24
2.3.3. 선행 3주 동아시아 해역 고수온 예측 모델 개발	31
2.4. 이상기후 예측 성능 평가 방법	37
2.4.1. 주요 검증 메트릭	37
2.4.2. 인공지능 모델 성능 기초선	39
2.5. 동아시아 이상기후 인벤토리 자료 공유 방식	39
2.5.1. GitHub 및 Git Repository 소개	39
2.5.2. 클라우드를 통한 대용량 자료 연결 및 colab 연동	40
2.5.3. Repository 설계	40
3. 결과	42
3.1. 동아시아 선행 3/4주 이상고온 역학 예측의 오차 보정 모델 개발	42
3.1.1. 라벨 기법에 따른 이상고온 분류 평가	42
3.1.2. 오차 보정 모델 적용	45
3.2. 격자별 주 단위 폭우 강도 및 빈도 예측 모델	57
3.2.1. 최종 모델 상세 성능	57
3.2.2. Multi-task learning 적용 실험	64
3.2.3. 입력 변수 조합 실험	65
3.2.4. 모델 구조 변형 실험	67
3.2.5. 기존 역학 모델 성능과 비교 평가	68

3.3. 선행 3주 동아시아 해역 고수온 예측 모델 개발	75
3.3.1. 이상치 예측에서 있어 가중 손실 함수 적용성 평가	75
3.3.2. UNet 구조 변형 실험을 통한 모델 민감도 평가	77
3.3.3. 입력 데이터 민감도 평가	78
3.3.4. 고수온 예측 원형 모형 도출 및 성능 평가	82
3.4. 동아시아 이상기후 주간 자료 온라인 공유 시스템 구축	89
3.4.1. 웹페이지 소개	89
3.4.2. 이상기후 인벤토리 및 시각화 자료	90
3.4.3. 활용법 소개	94
4. 요약 및 결론	97
References	100

표 차례

Table 2.1 This table describes the configuration of the prediction targets and domains for a subseasonal TMAX forecasting experiment over East Asia. It summarizes the spatial and temporal domain, resolution, target variable, and lead-time settings for daily maximum temperature forecasts centered on the Korean Peninsula.	10
Table 2.2 A set of atmospheric variables used for detecting and characterizing anomalously high-temperature. The table summarizes each variable's category and explains its specific role in identifying circulation, moisture, and pressure patterns associated with extreme high temperature.	10
Table 2.3 Classification Criteria for Binary and Multi-Class Heat Extremes Based on 3rd Week Mean TMAX. The labeling rules used to categorize temperature conditions during the 3rd week (days 15–21) based on mean maximum temperature percentiles.	11
Table 2.4 Classification Criteria for Binary and Multi-Class Heat Extremes Based on 4th Week Mean TMAX. The labeling rules used to categorize temperature conditions during the 4th week (days 22–28) based on mean maximum temperature percentiles.	12
Table 2.5. List of basic input variables used in this study.	14
Table 2.6. List of multi-scale decomposed input variables used in this study. A total of 48 input variables were used and tested in this study. The variables shown in bold are the final input variables selected in this study.	16
Table 2.7. Categorical labels assigned to each count of extreme rainfall events.	17
Table 2.8. Data split strategy for extreme rainfall prediction model.	18
Table 2.9 Training input: variables and temporal/spatial range	19
Table 2.10 Hyper-parameter Settings of the Study. The table summarizes the key hyper-parameters used in this study, including training configuration and model architecture settings. It lists values such as the number of epochs, batch size, learning rate, optimizer, activation function, and kernel sizes applied.	23
Table 2.11 Activation Functions and Loss Functions for U-Net and Attention U-Net According to Label Type. This table summarizes which activation functions and loss functions should be used when training U-Net or Attention U-Net models for binary and multi-class segmentation tasks. It outlines appropriate combinations such as Sigmoid with Binary Cross Entropy for	

binary labels and Softmax with Categorical Cross Entropy for multi-class labels.	24
Table 2.12. Hyperparameter settings.	31
Table 2.13 Prediction Target: variables and temporal/spatial range of label	31
Table 2.14 Key methods: learning rate, loss function setting, input data standardization and evaluation metrics applied.	33
Table 2.15 Climate indices utilized as external inputs in UNet model	36
Table 2.16 Directory Structure and Example Files in the repository “EastAsiaClimateExtremes“. The table provides an overview of the main directories used in the repository along with representative example files. It outlines folders for code scripts, datasets, and visualization outputs.	41
Table 3.1 This table contrasts U-Net with Attention U-Net across key dimensions such as probability strength, spatial patterns, BSS, extreme-region behavior, and overfitting tendencies. It highlights the clear performance gains and improved stability achieved by incorporating attention mechanisms.	56
Table 3.2 This table contrasts binary and multi-class modeling approaches in terms of learning stability, sensitivity to class imbalance, probability prediction behavior, BSS outcomes, and extreme-heat detection. It shows that binary models generally exhibit higher stability and accuracy, especially for extreme events, while multi-class models suffer from greater noise and imbalance effects.	57
Table 3.3. Performance metrics for rainfall amount prediction at a 3-week lead.	59
Table 3.4. Performance metrics for classification of extreme rainfall weeks at a 3-week lead.	59
Table 3.5. Performance metrics for rainfall amount prediction at a 4-week lead.	61
Table 3.6. Performance metrics for classification of extreme rainfall weeks at a 4-week lead.	61
Table 3.7. Performance metrics for rainfall amount prediction for different grid points.	63
Table 3.8. Performance metrics for classification of extreme rainfall weeks for different grid points.	64
Table 3.9. List of input variable sets.	66
Table 3.10. Example codes for data processing and visualization (file name, description and location)	90
Table 3.11 List of output statistics generated and stored by the example codes.	94

그림 차례

Figure 1.1 Comparison of annual mean MHW statistics-total days (top), frequency (middle), and mean intensity (bottom)-derived from OISST for 1982-2024(1st column), P1: 1982-2002 (2nd column), P2: 2004-2024 (3rd column), and the difference between the two periods (4th column).	5
Figure 2.1. Study domain for extreme rainfall. The entire study domain commonly defined (left column) and enlarged view of the selected grid points (right column).	7
Figure 2.2. Diagram of temporal variation of input variables by influence category.	14
Figure 2.3. Schematic of multi-scale temporal components of daily variability.	16
Figure 2.4. Example of how weekly rolling sum on daily rainfall works. Rainfall amount (mm/day) shown in bold is the extreme event that exceeds 90th percentile threshold. Figures shown in red are the extreme cases, and those in blue are the non-extreme cases.	17
Figure 2.5. Extreme rainfall counts (red bars) and extreme rainfall amount (blue bars) at a grid point in 2024.	18
Figure 2.6 Attention-Enhanced U-Net Architecture Overview. The diagram illustrates an encoder-decoder U-Net architecture augmented with attention gates that refine skip-connection features. It shows how multi-scale encoded features pass through attention modules before being merged with decoder layers to produce the final output.	21
Figure 2.7. Diagram of single-task learning.	25
Figure 2.8. Diagram of multi-task learning. The structure in the colored box is the one chosen for this study.	26
Figure 2.9. Overview of the model architecture used in this study.	28
Figure 2.10. The model architecture used in this study with the details on the modified ResNet-18 model.	28
Figure 2.11. List of architectures for residual neural networks with different layers.	29
Figure 2.12. Residual network with 34 parameter layers (He et al., 2015)	30
Figure 2.13 Diagram illustrating the training process, including input data stacking, model architecture, and the two-dimensional label distribution with EAMS.	32
Figure 2.14 Seven ablation tests to isolate the optimal AI model setup for EAMS MHW forecast.	34

Figure 2.15 The five Earth System components (Flux, Ocean, Static, Wave, and Upper-Atmosphere) evaluated as potential input features for the MHW forecast model. Component variables are listed in Table 2.9.35

Figure 3.1 ERA5 3-week mean of daily maximum temperature binary labels for ensemble members on October 10, 2023.42

Figure 3.2 Spatial distribution of 3-week ECMWF mean daily maximum temperature binary labels across ensemble members.42

Figure 3.3 Multi-Class ERA5 3-week mean of daily maximum temperature label ensemble distribution on October 10, 2023.43

Figure 3.4 This figure shows the multi-class classification results of ECMWF ensemble members for the 3-week mean daily maximum temperature on 10 October 2023.43

Figure 3.5 ERA5 4-week mean of daily maximum temperature binary labels for ensemble members on October 10, 2023.44

Figure 3.6 Spatial distribution of 4-week ECMWF mean daily maximum temperature binary labels across ensemble members.44

Figure 3.7 Multi-Class ERA5 4-week mean of daily maximum temperature label ensemble distribution on October 10, 2023.45

Figure 3.8 Spatial distribution of ECMWF 4-week mean daily maximum temperature multi-class labels on October 10, 2023.45

Figure 3.9 Ensemble U-Net binary probability prediction maps for October 10, 2023. The figure illustrates 11 ensemble probability distributions predicted from a U-Net model trained with a 3-week mean daily maximum-temperature label.46

Figure 3.10 Ensemble Attention U-Net binary probability prediction maps for October 10, 2023. The figure illustrates 11 ensemble probability distributions predicted from a Attention U-Net model trained with a 3-week mean daily maximum-temperature label.46

Figure 3.11 Ensemble U-Net binary probability prediction maps for October 10, 2023. The figure illustrates 11 ensemble probability distributions predicted from a U-Net model trained with a 4-week mean daily maximum-temperature label.47

Figure 3.12 Ensemble Attention U-Net binary probability prediction maps for October 10, 2023. The figure displays binary probability predictions generated by a Attention U-Net model with a 4-week mean daily maximum-temperature for 11 ensemble members on October 10, 2023.47

Figure 3.13 Ensemble based U-Net multi-class probability prediction maps on October 10, 2023. These maps show ensemble probability distributions predicted using labels based on the 3-week mean maximum temperature on October 10, 2023.48

Figure 3.14 Ensemble based Attention U-Net probability prediction maps on October 10, 2023. This figure presents probability maps generated by a multi-class U-Net model with the 3-week mean maximum temperature for 11 ensemble members on October 10, 2023.48

Figure 3.15 Ensemble based U-Net multi-class probability prediction maps on October 10, 2023. These maps show ensemble probability distributions predicted using labels based on the 4-week mean maximum temperature on October 10, 2023.49

Figure 3.16 This figure presents probability maps generated by a multi-class U-Net model with the 4-week mean maximum temperature for 11 ensemble members on October 10, 2023. 49

Figure 3.17 The left panel shows the binary label of 3-week mean daily maximum temperature derived from ERA5 for 10 October 2023. The right panel presents the average of eleven binary labels from the ECMWF-S2S ensemble for the same 3-week mean temperature period.50

Figure 3.18 This figure shows the averaged probabilities from 11 ensemble predictions produced by deep learning models trained on binary labels of 3-week mean maximum temperature. The left panel displays results from the U-Net model, while the right panel shows corresponding outputs from the Attention U-Net model for 10 October 2023.50

Figure 3.19 The maps show the mean probabilities obtained by averaging 11 predictions for the multi-class labels of 3-week mean maximum temperature on 10 October 2023. The left panel corresponds to the U-Net model, while the right panel shows the results from the Attention U-Net model.50

Figure 3.20 The left panel shows the binary label of 4-week mean daily maximum temperature derived from ERA5 for 10 October 2023. The right panel presents the average of eleven binary labels from the ECMWF-S2S ensemble for the same 4-week mean temperature period.51

Figure 3.21 This figure shows the averaged probabilities from 11 ensemble predictions produced by deep learning models trained on binary labels of 4-week mean maximum temperature. The left panel displays results from the U-Net model, while the right panel shows corresponding outputs from the Attention U-Net model for 10 October 2023.51

Figure 3.22 The maps show the mean probabilities obtained by averaging 11 predictions for the multi-class labels of 4-week mean maximum temperature on 10 October 2023. The left panel corresponds to the U-Net model, while the right panel shows the results from the Attention U-Net model.52

Figure 3.23 The figure compares brier skill scores (BSS) for binary 3-week mean maximum-temperature over 2022-2023, contrasting ECMWF-S2S (left) predictions with probabilities generated by U-Net (center) and Attention U-Net (right) models.52

Figure 3.24 The figure compares brier skill scores (BSS) for binary 4-week mean maximum-temperature over 2022-2023, contrasting ECMWF-S2S (left) predictions with probabilities generated by U-Net (center) and Attention U-Net (right) models.53

Figure 3.25 The figure compares the multi-class BSS derived from raw ECMWF-S2S 3-week mean maximum temperature with those from U-Net and Attention U-Net models trained on multi-class labels. The middle (U-Net) and right (Attention U-Net) panels show how the deep-learning models improve or modify regional skill relative to the ECMWF-S2S predictions (left).53

Figure 3.26 The figure compares the multi-class BSS derived from raw ECMWF-S2S 4-week mean maximum temperature with those from U-Net and Attention U-Net models trained on multi-class labels. The middle (U-Net) and right (Attention U-Net) panels show how the deep-learning models improve or modify regional skill relative to the ECMWF-S2S predictions (left).54

Figure 3.27 The left panel shows violin plots comparing the spatial BSS distributions of ECMWF-S2S and the U-Net, Attention-U-Net-based models (binary and multi-class labels) for 3-week mean maximum temperature. The right panel displays the BSS differences obtained by subtracting the ECMWF-S2S based BSS from each model' s BSS at every grid point. ...55

Figure 3.28 The left panel shows violin plots comparing the spatial BSS distributions of ECMWF-S2S and the U-Net, Attention-U-Net-based models (binary and multi-class labels) for 4-week mean maximum temperature. The right panel displays the BSS differences obtained by subtracting the ECMWF-S2S based BSS from each model' s BSS at every grid point. ...55

Figure 3.29 Comparison of BSS differences between 3-Week and 4-Week.56

Figure 3.30. Timeseries of rainfall amount at a 3-week lead predicted by the final model for a grid point near Seoul on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in

black dashed line. Extreme cases that exceed the 90th percentile threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.58

Figure 3.31. Scatter plot of rainfall amount at a 3-week lead predicted by the final model for Seoul grid point compared to actual amount.59

Figure 3.32. Timeseries of rainfall amount at a 4-week lead predicted by the final model for a grid point near Seoul on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th percentile threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.60

Figure 3.33. Scatter plot of rainfall amount at a 4-week lead predicted by the final model for Seoul grid point compared to actual amount.61

Figure 3.34. Timeseries of rainfall amount predicted by the final model for a grid point near Daejeon on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.62

Figure 3.35. Timeseries of rainfall amount predicted by the final model for a grid point near Busan on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.62

Figure 3.36. Timeseries of rainfall amount predicted by the final model for a grid point near Gwangju on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.63

Figure 3.37. Timeseries of rainfall amount predicted by the final model for a grid point near

Pyungchang on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.63

Figure 3.38. Comparison between single-task learning-based results (green solid line) and multi-task learning-based results (red solid line). Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded.64

Figure 3.39. Model performance of critical success index (CSI) values for each bar on the left column and temporal anomaly correlation coefficients (TCC) values for each bar on the right column for each combination of variable set (refer to Table X.X for list of variable sets). Best combination in pink circle is composed of ISO, low frequency, and climatology components.66

Figure 3.40. Summary of experiments of combinations of variable set with critical success index (CSI) values for each bar on the left column and temporal anomaly correlation coefficients (TCC) values for each bar on the right column. Best combination in pink box is composed of ISO, low frequency, and climatology components.67

Figure 3.41. Comparison between ResNet-18 (green solid line) and ResNet-18+LSTM (red solid line). Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded.68

Figure 3.42. Bar plots of ECMWF (1.5°) forecast model performance for total precipitation compared to ERA5 reanalysis data for week 3 (radish colored bars) and 4 (blueish colored bars) for different grid points for all seasons and individual season.70

Figure 3.43. Timeseries of weekly accumulated rainfall amount at grid point 1 for week 3 and 4 in the top two panels and the corresponding ECMWF forecast results. The 90th percentile thresholds are shown in black dashed line for ERA5 and ECMWF hindcast data. Hit cases that overlap each other are shown in red circle in both ERA5 and ECMWF panels.71

Figure 3.44. Timeseries of weekly accumulated rainfall amount at grid point 2 for week 3 and 4 in the top two panels and the corresponding ECMWF forecast results. The 90th percentile thresholds are shown in black dashed line for ERA5 and ECMWF hindcast data. Hit cases that

overlap each other are shown in red circle in both ERA5 and ECMWF panels.72

Figure 3.45. Timeseries of weekly accumulated rainfall amount at grid point 3 for week 3 and 4 in the top two panels and the corresponding ECMWF forecast results. The 90th percentile thresholds are shown in black dashed line for ERA5 and ECMWF hindcast data. Hit cases that overlap each other are shown in red circle in both ERA5 and ECMWF panels.73

Figure 3.46. ECMWF forecast model performance for extreme precipitation for week 3 and 4 with performance metrics including accuracy, precision, recall, F1-score, and CSI.74

Figure 3.47. A case study with results for the Seoul grid point.75

Figure 3.48. Boxplots showing the distribution of (left) ACC and (right) CSI across three groups with different input domain size/resolution (G5: global 5deg, L5: large 5deg, S2.5: small 2.5deg) on the x-axis, with colors indicating the scale of loss function weighting (0,2,4, and 6) for the 104 test samples across 2023–2024. The central line (white circle) in each box denotes the median (mean), box edges indicate the first and third quartiles, whiskers extend to 1.5 times the interquartile range, and points beyond the whiskers represent outliers.76

Figure 3.49. Similar as in Figure 3.48 but for two groups with different train/validation sample fraction (3:1 vs 2:1) for x-axis. Blue horizontal lines indicate the overall mean across all samples, independent of the loss function weight scale.77

Figure 3.50. Boxplots showing the distribution of (a) ACC and (b) CSI across three models (Base UNet, Attention UNet, and Swin-Transformer UNet) along the x-axis. Colors indicate the temporal sequence of input data (white boxes for one-week inputs and navy boxes for three-week inputs). The green horizontal line represents the ECMWF marine heatwave forecast skill at a 3-week lead time.78

Figure 3.51. Boxplots showing the distribution of (a, c) ACC and (b, d) CSI across six input-variable combinations (all, woF, woO, woS, woW, and woU; see Table X.X). The red horizontal line indicates the mean skill of models trained with all input variables. Reddish and blue arrows indicate relative performance of models trained with each reduced input set (without Flux, Ocean, Static, Wave, or upper-atmosphere variables); upward arrows mark lower skill and downward arrows higher skill relative to the all-input baseline, highlighting the positive or negative contribution of each component to model performance.80

Figure 3.52. Similar to Figure 3.51, but focusing only on the CSI results for (a) UNet and (b) swinUNet across five different input-variable combinations. Opaque and transparent red horizontal lines represent the mean and median skill of models trained with the minimum

(baseline) input variables. Opaque and transparent arrows indicate the mean and median skill improvements relative to this baseline.81

Figure 3.53 Boxplots showing the distribution of ACC (upper panels) and CSI (lower panels) across four additional-input conditions (None, Both, Climate Modes only, River Discharge only; see Table X.X). Panels (a, c) show UNet models, while (b, d) show swinUNet models (latter tested only with/without River Discharge). The green horizontal line denotes ECMWF marine heatwave forecast skill at a 3-week lead time (the same as in Fig. 3.50b).82

Figure 3.54 (a) Scatter plots of ACC versus CSI for all trained models, organized by model architecture, input domain size/resolution, validation sample proportion, loss scale, multi-modal inputs, and temporal input sequence (from left to right). (b) The fraction of good models, defined as those falling within the top-right gray box, relative to all ensembles in each category mentioned in (a).84

Figure 3.55 Weekly evaluation of ECMWF 3-week lead marine heatwave predictions across 2023. The top-left three panels present the temporal ACC (left), RMSE (center), and CSI (right) at each grid point. The bottom-left line plots show the spatial ACC (black) and CSI (magenta) over EAMS for each week. The top-right line plots illustrate the temporal variation of the weekly Impact Factor (weekly extremeness) for the grid point nearest to Dok-do (37.5° N, 130.5° E), based on observations (OISST, yellow) and ECMWF forecasts (gray for individual ensemble members and thick black for their ensemble mean). The bottom-right two panels depict the spatial distribution of the Impact Factor for the marine heatwave event during the week beginning 30 July 2023, comparing observations (OISST, left) with the ECMWF ensemble mean forecast (right).86

Figure 3.56 Performance of the two best AI models: weekly evaluation of 3-week lead marine heatwave predictions throughout 2023. (a) shows a swinUNet model with the 3Wseq/S2.5/wFOS/RD_added/2:1 train/validation ratio setup, while (b) presents a model with the same configuration but using a 3:1 train/validation ratio. Details are the same as in Figure 3.55.87

Figure 3.57 Same as in Figure 3.56, but for AI model ensemble forecasts. Ensemble members in (a) are swinUNet models with the S2.5/2:1 ratio/weight scale=2/RD_added setup, while ensemble members in (b) are swinUNet models with the S2.5/1W sequence/weight scale=2/RD_added setup.88

Figure 3.58. EastAsiaClimateExtremes GitHub main page screen shot 1.89

Figure 3.59. EastAsiaClimateExtremes GitHub main page screen shot 2.90

Figure 3.60. Timeseries of daily 2m air temperature at 36.0° N, 127.5° E in 2024, showing the original data (black), climatological mean (green), 90th percentile threshold (red), and anomalous hot days exceeding threshold line (red shaded) and extreme days identified as event periods (blue shaded).91

Figure 3.61 Spatial distribution of long-term mean and trend of marine heatwave event statistics: (left) frequency, (middle) total days, (right) mean intensity derived from daily SST timeseries over East Asia Marginal Sea.92

Figure 3.62 Spatial distribution of extreme weeks: annual mean statistics of anomalously hot weeks derived from weekly T2m timeseries over East Asia: (Left) frequency (times per year), (middle) mean intensity (° C per year), and (right) maximum intensity (° C per year).93

Figure 3.63 Spatial distribution of 90th-Percentile extreme temperature frequency based on ECMWF-S2S Versions 2016 and 2024.93

Figure 3.64. Screen shot of visualization code 1: Timeseries and Heatmap Visualization for Extreme Events.95

Figure 3.65. Screen shot of visualization code 2: Weekly Extreme Analysis: Frequency, Mean, and Max Intensity.96

1. 서론

전지구적 기후 변화로 인해 폭염 및 이상고온 발생 빈도와 강도가 증가하고 폭우·가뭄의 극단성과 변동성이 증대되고 있다. 이러한 이상기후 관련 재난에 대비해 기후 위험 적응 정보 통합플랫폼 구축 등 정부 차원의 대응이 구체화 되고 있으며 기후 재난에 대비가 가능한 중기에서 장기 사이의 계절내(2-6주) 시간 규모 예측의 중요성이 크게 높아지고 있다.

이상기후로 인한 피해가 날로 늘어감에 따라 수 주 단위(subseasonal to seasonal, S2S)의 이상 기상·기후 발생을 신뢰성 있게 예측하는 기술에 대한 수요가 급격히 증가하고 있고, 인공지능 기술 적용의 기대도 높아지고 있다. 최근 인공지능 기술이 빠르게 발전함에 따라 전 세계 주요 기상 기관과 대형 기업들은 전 지구 규모의 딥러닝 기반 대기 예측 모델을 활발하게 개발하고 있다. 이러한 모델들은 특히 단기 및 중기 예보에서 뚜렷한 성능 향상을 보이며 실제 운영 예보 과정에 인공지능 기반 예측 결과를 결합 또는 활용하는 시도가 점차 확대되고 있다. 하지만 이른바 파운데이션 모델(Foundation model)이라 불리는 대형 모델 대부분은 1-2주 이내의 예보에 초점이 맞춰져 있어 2주를 넘어가는 계절 내 규모의 예측에는 여전히 뚜렷한 한계가 존재한다. 하지만, Vonich and Hakim (2025)에 따르면 전구 순환 인공지능 모형 GraphCast (Lam et al. 2023)의 역전파 알고리즘의 경사하강법을 이용해 최적의 초기 조건을 추출하고 이를 초기장으로 처방했을때 미국 북서부 폭염에 대한 GraphCast 예측력이 30일까지 연장되었다는 결과를 보인 바 있다. 이는 초기 조건이 아무리 정확해도 자연계의 비선형성으로 인한 모델의 불확실성이 급격히 증가하고 2주의 예측 한계를 넘기 어렵다는 기존의 고정관념을 깨는 연구 결과이다. 수치들이 GraphCast 모델에 편향되어 있고 현업적 활용 방안이 마련되지 않은 분석적 연구 결과지만, 이상기후 계절내 예측 개선을 위한 인공지능 기술 적용의 근거가 될 수 있겠다.

계절내 시간 규모 이상 고온 예측은 농업, 에너지, 보건, 재난 대응 등 다양한 분야에서 필수적인 정보로 필요하다. 특히 동아시아와 한반도 지역은 여름철 아열대 고기압의 변동성, 몬순 순환, 원격 영향(텔레커넥션), 대기파동의 전달 등 다양한 대규모 기후 요인이 복합적으로 작용하여 여름철 이상고온 뿐 아니라 장기간 고온이 자주 발생하며 환절기 이상고온의 발생 가능성이 증가하여 그 파급 영향이 매우 증가하고 있어서 이에 따른 전력 수급, 보건·복지, 농업 및 수자원 관리 측면의 사회·경제적 피해가 크다. 기존 수치예보체계와 통계·기계학습 기반 연구들은 주로 일 단위 또는 1-2주 단기 리드를 대상으로 고온 발생 여부, 극값 크기, 지속 기간을 예측해 왔으나, 3주 이상 S2S 범위에서는 예측 기술이 급격히 저하되고, 특히 확률적 이상고온 탐지 능력이 충분하지 않은 것으로 보고되고 있다.

최근 S2S 예측 연구에서는 전지구 예측모델의 기본 예측 성능에 딥러닝 기반 후처리(post-processing)을 결합하여, 예측 패턴의 구조적 오류를 보완하는 방법이 주목받고 있다. Convolutional Neural Network(CNN), U-Net 계열 모델, Attention 기반 구조 등이 대표적이며, 이들 모델은 강수·기온·순환 패턴의 공간적 비선형성을 효과적으로 포착할 수 있다는 장점이 있다. 특히 U-Net은 의료 영상 분야에서 시작되었지만, 공간 격자 기반 기후 예측에도 높은 성능을 보여 다수의 최신 기후 연구에서 활용되고 있다(예: Vandal et al., 2019; Harilal et al., 2021). 이와 같이 딥러닝 구조를 활용하여 폭염 및 극한고온을 예측하는 연구가 활발히 진행되

고 있으며, 관측·재분석자료와 수치모델 예측장을 결합하여 특정 리드 타임에서의 온도, 열지수, 복합지수 등을 예측하는 시도가 이루어지고 있다. 예를 들어, 일부 연구에서는 ERA5와 S2S 예측장을 활용한 딥러닝 기반 월 혹은 계절 기온 이상 예측에서 전통적 통계모형 대비 향상된 기술을 제시하고(Lin, H., et al., 2023; Xu, L., et al., 2023), 다른 연구에서는 XAI(Explainable AI)를 결합하여 극한고온 예측에 기여하는 주요 변수(예: TMAX, 습도, 기압)를 해석적으로 제시하기도 하였다(Shafiq, F., et al., 2025; Lee, Y., et al., 2024).

그러나, 이러한 연구들은 대체로 단일 값 예측(회귀) 또는 단순한 이진 분류(발생/비발생)에 초점이 맞추어져 있는데 이러한 모델들의 성능은 라벨링(labeling) 방식에 크게 의존된다. 기존 연구 대부분은 단순 이진 분류(극한/비극한)를 적용해 왔지만, 최근에는 극한 강도 범위를 더 세분화한 다중 클래스 분류(multi-class)를 적용하려는 시도가 늘고 있다. 다중 클래스는 정보량이 많다는 장점이 있지만, 클래스 경계가 모호하거나 데이터 편중(class imbalance)이 큰 경우 오히려 성능을 저해할 수도 있다. 따라서 라벨링 방식이 기후 후처리 딥러닝 모델의 성능에 어떤 영향을 주는지에 대한 연구 등이 미흡하다. 또한 S2S 범위의 앙상블 확률예측에서 “라벨 정의 방식(이진 vs 다중 클래스)” 과 “딥러닝 구조(U-Net, Attention U-Net 등)” 의 조합이 이상고온 확률 패턴 및 검증 지표(예, Brier Skill Score 등)에 미치는 영향 등의 연구는 상대적으로 덜 다루어져 왔다.

S2S 예측에서 이상 기온의 발생 확률을 정확하게 산정하는 것은 위험 기반 의사결정(risk-based decision making)에 핵심이라고 할 수 있다. 기존의 단일 값에 대한 예측보다 확률 기반 예측(probabilistic forecasting)이 사용자(정책 결정자, 산업체 등)에게 더 유용할 수 있음에도 딥러닝 기반 후처리에서 확률 예측의 정량 비교 연구는 부족하다. 딥러닝 모델 중 Attention 메커니즘이 공간적 특징의 선택(feature selection)을 향상하는 것으로 알려져 있는데 S2S 예측에서 이상 기온 탐지에 적용된 연구 또한 드물다.

본 연구는 이러한 배경에서, 유럽중기예보센터(ECMWF, European Centre for Medium-Range Weather Forecasts) S2S 3주 및 4주 예측 평균 최고기온(TMAX)과 ERA5 재분석 TMAX를 대상으로 이상고온 라벨 설계(이진 vs 다중 클래스)와 U-Net·Attention U-Net 기반 딥러닝 모델의 조합이 이상고온 확률 예측의 재현성과 Brier Skill Score(BSS)에 미치는 영향을 체계적으로 평가하는 것을 목적으로 한다. 연구 영역은 한반도를 중심으로 한 동아시아 도메인(약 24-46.5° N, 118.5-141° E)으로 설정하고, ECMWF S2S Version 2024의 2004-2023년 자료와 ERA5 재분석을 활용하여 3주 및 4주 평균 TMAX에 대한 이상고온(상위 10% 구간)을 정의한다. 이를 위해 90백분위수 이상을 기준으로 한 이진 라벨(0/1)과, 75-90백분위수 및 90백분위수 이상 구간을 포함하는 다중 클래스 라벨(0/1/2)을 각각 ERA5와 ECMWF S2S TMAX에 공통 적용하고, ECMWF S2S의 TMAX·TMIN, OLR, MSLP, Z500, 850hPa 풍 및 비습(Q850) 등 8개 변수를 입력장으로 사용하는 확률 분류 프레임워크를 구축하였다.

딥러닝 모델은 공간 패턴 인식에 강점이 있는 U-Net과 중요 영역에 대한 가중을 부여할 수 있는 Attention gate를 결합한 Attention U-Net(Att-U-Net)을 선택하여, 각 라벨링 체계에 대해 독립적으로 학습 및 예측을 수행한다. 학습 기간은 2004-2021년(18년), 검증 기간은 2022-2023년(2년)으로 설정하고, 출력은 라벨별 발생 확률장이며, 성능 평가는 Brier Score 및 Brier

Skill Score(BSS)를 통해 ECMWF 원시 앙상블 예측 대비 개선 정도를 파악한다. 특히, 이진 및 다중 클래스 라벨에 따라 BSS의 정의와 해석이 달라지는 점을 고려하여, 각 라벨 구조에 적합한 다중 범주 BSS를 사용하고, ECMWF 기후값 혹은 기후 평년(climatology)을 기준으로 한 BSS를 병행함으로써, 라벨 복잡도가 예측 기술에 미치는 영향을 정량적으로 비교한다.

이와 같은 접근을 통해 본 연구는 ECMWF S2S 예측의 3주 평균 최고기온을 활용하여, 이진 라벨과 다중 클래스 라벨의 구조적 차이가 U-Net 및 Attention U-Net 모델의 이상고온 탐지 능력에 어떤 영향을 미치는지를 정량적으로 비교하고자 한다. 특히, 확률 예측을 기반으로 하는 Brier Score 및 Brier Skill Score를 활용해, 라벨 방식이 예측 성능·공간 패턴·신호 재현성에 미치는 실질적인 차이를 평가한다. 따라서, ERA5와 S2S 자료를 활용하여 라벨링 기법이 예측 결과에 미치는 영향을 직접적으로 비교하고, U-Net과 Attention U-Net 구조의 특성과 장단점을 극한 고온 탐지(extreme-heat detection)의 맥락에서 체계적으로 분석하였다.

또한, 이 연구는 Binary와 Multi-class 라벨이 갖는 구조적 특성이 모델 성능에 미치는 영향을 정량적으로 평가하였으며, Brier Skill Score(BSS)를 기반으로 한 공간적 정확도 개선 패턴을 제시함으로써 기계학습 기반 극한기온 예측의 신뢰도 향상에 기여한다. 요약하면 (1) S2S 범위 이상고온 예측에서 라벨 설계(이진 vs 다중 클래스)가 확률 패턴과 검증 지표에 미치는 역할을 규명하고, (2) Attention U-Net과 같은 공간적 주의 메커니즘이 고온 신호 탐지에 어떤 이점을 제공하는지 평가하며, (3) ECMWF S2S-ERA5 시스템과 딥러닝 모델을 결합한 이상고온 확률 예측 프레임워크의 가능성과 한계를 제시하고자 한다. 이는 향후 한반도 및 동아시아 지역의 열파 조기경보 시스템 고도화와, 기후 변화 환경에서 증가하는 이상고온 위험 관리 전략 수립에 기초 자료로 활용될 수 있을 것이다.

폭우는 인명 및 재산 피해를 포함한 막대한 사회 경제적 손실을 초래하기 때문에 정확한 예측이 무엇보다 중요하다. 최근 지구 온난화가 가속화되면서 폭우의 발생 빈도와 강도가 전 지구적으로 증가하는 경향을 보이는 것으로 보고되고 있다(Westra et al., 2014; Fowler et al., 2021). 이러한 변화는 지역적 극한기후 발생 양상에도 영향을 미치며 재해 위험 관리 측면에서 폭우 예측의 중요성을 더욱 부각한다. 하지만 강수는 대기, 해양, 지표 등 다양한 요인 간의 비선형적인 상호작용을 통해 시공간적으로 복잡한 기작으로 발생하여 예측이 매우 어렵다. 특히 폭우와 같은 극한 강수는 발생 빈도가 낮아 분석 가능한 사례가 제한적이고 그 발생 기작 또한 지역과 시기에 따라 다양해 예측 불확실성이 크고 분석이 어렵다. 더 나아가 2주를 넘는 계절 내 예측 구간에서는 대기 해양 변수들의 예측 가능성이 급격히 저하되고 주요 영향 인자의 시그널 지속성이 약화되기 때문에 기존의 통계 모델이나 전통적인 물리 기반 수치예보 모델만으로는 예측 성능 향상에 한계가 있다.

기존의 딥러닝 기반 계절 내 강수 예측 연구는 주로 역학 모델 결과를 활용한 후처리(bias correction)에 초점을 두고 있다. Chung et al. (2024)는 U-Net 구조의 모델을 활용하여 여러 기후역학 모델의 강수 예측을 ERA5 자료에 맞게 보정함으로써 예측 성능을 향상시켰다. Nie and Sun (2024)도 ECMWF 모델이 예측한 대규모 대기 순환장을 딥러닝으로 먼저 보정한 뒤, 이렇게 보정된 순환장을 기존 통계 모델에 넣어 중국 남서부 지역의 계절 내 강수를 예측하는 방식을 제안했다. 이처럼 계절 내 폭우 예측을 직접 다룬 인공지능 연구는 아직 많지 않고 이상

기후나 폭우와 같은 극한 사건에 대한 예측 정확도는 상대적으로 낮으며, 이에 특화된 연구 사례도 충분하지 않아 추가적인 방법론 개발과 성능 향상이 필요한 실정이다. 이에 본 연구에서는 인공지능 기반의 방법론을 활용하여 계절 내 선행 3, 4주의 이상 강수를 직접 예측하는 모델을 개발하고자 한다. 궁극적으로는 계절 내 이상 강수를 예측하는 초기 모델을 개발하고 계절 내 규모에서 이상 강수의 조기 예측 가능성을 탐색하는 것이다.

해양 열파는 어류 산란장 및 회유 경로의 변화, 산호 및 해조류 군집 붕괴, 저서생태계 장기 교란 등을 유발하여 동아시아 수산 자원과 생태계를 구조적으로 변화시키며, 연안 해역의 이례적인 고수온이 인접 육상의 폭염, 관련 대기 순환 결합할 경우 농업·보건·에너지 수요 등에까지 영향을 확산시킨다(Gu et al. 2025, Im et al 2017).

안타깝게도 1980년대 이후 동아시아 주변 해역에서 해양열파(MHW)의 발생 빈도, 총 지속 일수, 평균 강도가 모두 통계적으로 유의하게 증가하고 있음이 최근 관측 및 재분석 기반 연구에서 반복적으로 보고되고 있다(Lee et al., 2023). 이러한 경향은 동아시아를 넘어 전 지구적으로 나타나는 흐름이다(Oliver et al., 2018). Figure 1.1은 1982년부터 2024년까지의 동아시아 해역 해양열파 발생일수·빈도·평균 강도를 1982-2002와 2004-2024 두 기간으로 나누어 평균한 뒤 그 차이를 제시한 것이다. 2003년 이전에는 연간 해양열파 발생 빈도가 2회 미만, 발생일수도 20일 미만으로 나타난 반면, 2003년 이후에는 이 전 기간에 비해 평균적으로 2회와 20일 정도 증가하는 경향을 보였다. 특히 이러한 증가폭은 동해에서 매우 뚜렷하다. 평균 강도의 증가는 일본 동쪽 쿠로시오 확장 지역에서 두드러지게 나타나며, 황해와 동해에서도 약 0.2~0.3 정도의 증가가 관찰된다.

해양열파가 농업·보건·에너지 분야에 미치는 막대한 파급력을 고려할 때, 동아시아 해양 열파의 정의를 명확히 하고 장기 추세와 공간 패턴을 정량화하며, 부문별 영향을 최소화하고 이에 대비할 수 있도록 예측 능력을 고도화하는 일은 기후위기 시대 동아시아 연안의 재난·위험 관리, 수산 및 연안 개발 정책 수립에 매우 중요하다. 이에 본 연구에서는 인공지능 기술을 활용하여 2주 이상의 선행시간에 대해 동아시아 해역, 고수온 현상 자체에 특화된 모델을 고안함으로써 기존 역학 모델과 차별화된 데이터 드리븐 방식의 예측 잠재력을 확인하고자 한다.

마지막으로 최근 인공지능을 활용한 기후 예측 연구가 활발히 이루어지면서 공통된 자료 환경에서 모델을 개발하고 성능을 비교할 수 있는 벤치마크 데이터셋의 중요성이 커지고 있다. 예를 들어, WeatherBench는 ERA5 재분석 자료를 전 지구의 변수/기간별 전처리해 제공하는 중기(1-14일) 예측용 벤치마크 데이터셋이다(Rasp et al. 2020). 연구자들이 공동의 입력자료와 평가 지표를 사용해 다양한 딥러닝/통계/물리 기반 모델의 성능을 비교할 수 있도록 설계되었다. WeatherBench2는 이 개념을 확장하여 ERA5 기반 고해상도 장기 데이터를 제공하고 평가 코드를 포함해 데이터 기반 예측 모델을 공동 환경에서 비교 평가하는 것을 목표로 한다(Rasp et al. 2024). 하지만 이러한 대부분의 기존 벤치마크 자료는 일반적인 기후 예측에 초점을 맞추고 있어 이상기후 현상을 체계적으로 다루는 데이터셋은 부족한 실정이다. 본 연구에서는 동아시아 지역을 중심으로 한 주요 이상기후 요소에 대한 인벤토리를 구축하고, 이를 GitHub 페이지를 통해 공개함으로써 다양한 AI 기반 이상기후 연구를 위한 공통 벤치마크 자료로 활용될 수 있도록 하고자 한다.

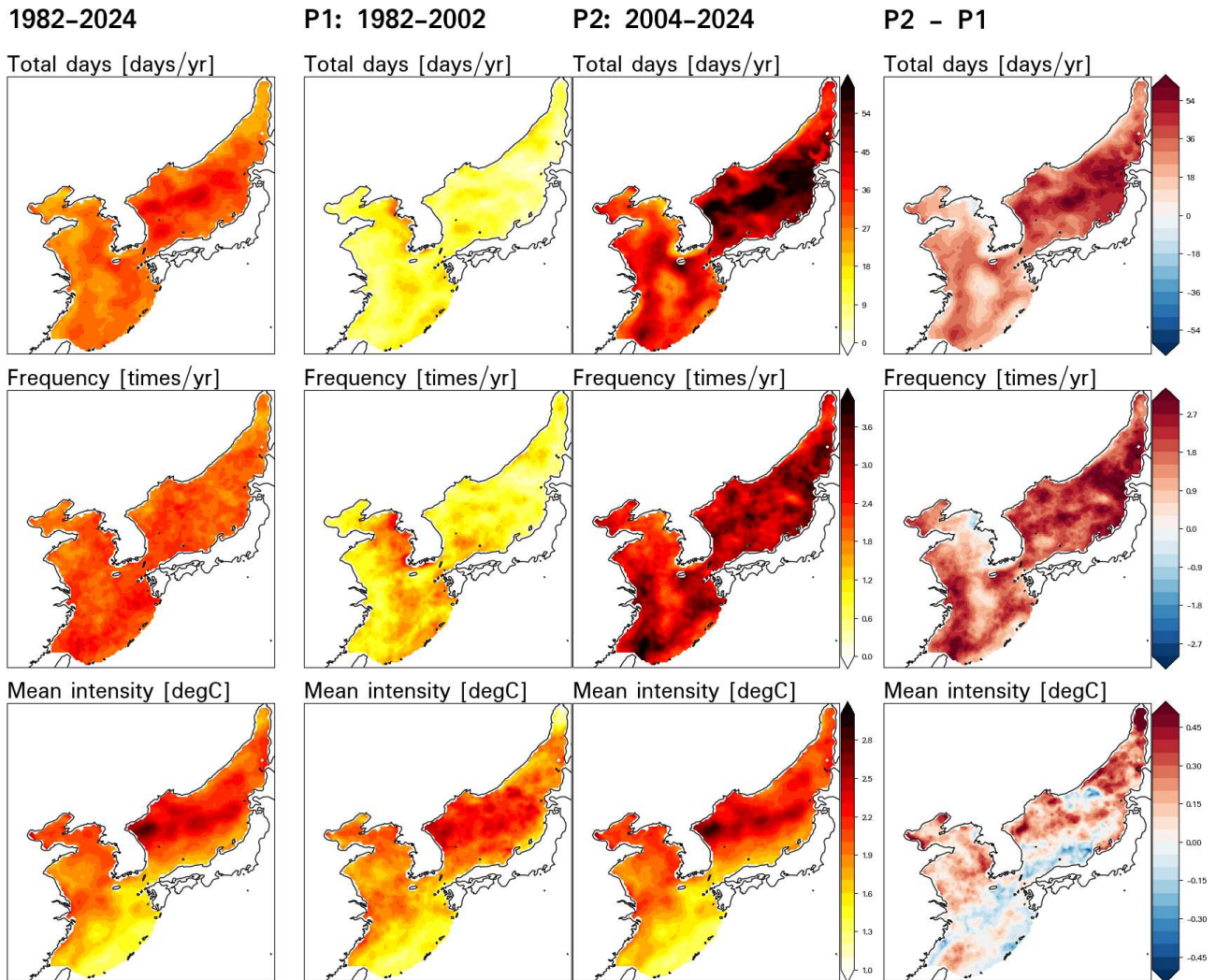


Figure 1.1 Comparison of annual mean MHW statistics—total days (top), frequency (middle), and mean intensity (bottom)—derived from OISST for 1982-2024(1st column), P1: 1982-2002 (2nd column), P2: 2004-2024 (3rd column), and the difference between the two periods (4th column).

2. 방법 및 자료

2.1. 이상기후 예측 타깃 선정

2.1.1. 예측 대상 이상기후 현상 정의

2.1.1.1 이상고온(AHT)이란 무엇인가?

이상고온(anomalous high temperature)은 기준 기간(예, 30년 기후 평년: 1991-2020년) 동안의 기온 분포에서 상위 10%(90th percentile)를 초과하는 일(또는 시공간 격자)의 기온 상태로 정의한다(IPCC, 2021). 기상청의 기후정보포털에 의한 이상고온 정의는 일 최저기온 또는 최고기온의 “평년 편차값”이 기후통계에서 상위 10%에 해당하는 값(90th percentile)을 초과할 경우 “이상고온”으로 정의한다. 즉, 기준 기간에 대해 산출한 최저·최고기온의 편차 분포에서, 편차가 상위 10%를 넘는 고온 사건(현상)을 “이상고온”이라고 부르는 것이 기상청에서 사용하는 공식적인 정의이다.

2.1.1.2 폭우(HR)란 무엇인가?

강수는 시간을 기준으로 하면 4가지로 구분하고 있다. 시간당 3mm 미만을 약한 비, 3~15mm 미만을 보통 비, 15~30mm 미만을 강한 비, 30mm 이상을 매우 강한 비로 정의된다(기상청 예보용어해석). 호우(豪雨)는 기상청에서 공식적으로 사용하는 용어로, 짧은 시간에 매우 많은 비가 내려 홍수나 침수 등의 피해를 일으킬 수 있는 정도의 강수를 의미한다. 기상청의 호우 관련 기상특보 기준에 따르면, 3시간 누적강우량이 60mm 이상 예상되거나 12시간 누적강우량이 110mm 이상 예상될 때 ‘호우주의보’가, 3시간 누적강우량이 90mm 이상 예상되거나 12시간 누적강우량이 180mm 이상 예상될 때 ‘호우경보’가 발표된다. 폭우는 일상적으로 호우와 자주 혼용되어 사용되는 용어로, 단기간 특정 지역에 많은 비가 집중되는 현상, 즉 집중호우나 이상 강수를 의미한다. 폭우를 정의하는 명확한 공식 기준은 존재하지 않으나, 선행 연구에서는 시간당 강수량의 경우 평년의 99 또는 99.95th percentile 이상 등, 일일 강수량의 경우 평년의 90 또는 95th percentile 이상, 혹은 일일 총 강수량이 80 혹은 100mm 이상 등으로 다양한 기준으로 폭우를 정의하고 있다 (Park et al., 2021; Do et al., 2023). 본 연구에서 폭우는 평년 대비 일일 강수량의 90th percentile 임계값 이상인 사건으로 정의하였다.

2.1.1.3 해양열파(MHW)란 무엇인가?

동아시아 해양 열파(Marine Heatwave, MHW)는 동아시아 연안 및 인접 해역에서 해수면 온도(SST)가 해당 일자의 특정 임계값 초과하는 상태가 수일 이상 연속으로 지속될 때 발생하는 극한 고수온 사건으로 정의된다. 임계값 이하로 떨어지는 기간이 최대 허용 기준 일수를 넘어설 때는 기간을 나누어 독립 사건(event)으로 구분한다. 전지구 규모에서 일반적으로 용인되는 해양열파 사건 정의, 즉 기후학적 분포 상위 90퍼센타일 이상의 고수온이 최소 연속 5일 유

지되고 임계값 이하로 떨어지는 날이 최대 2일을 넘지 않을 경우(Hobday et al. 2016)의 조건을 동아시아에서도 동일하게 적용할 수 있다. 다만 임계값, 최소 유지/최대 간격 기준일을 조정함에 따라 발생 통계 및 분포는 달라진다.

2.1.2. 이상기후 시·공간 예측 타깃

본 과제는 계절내 시간 규모에서의 동아시아 이상기후 예측 모델 개발을 목표로 하기에 시간적으로는 선행 3-4주 주별 이상치, 공간적으로는 한반도 중심의 동아시아 영역을 공통 예측 대상으로 하고, 컴포넌트별 세부적인 설정은 아래와 같다.

이상고온 예측 타깃은 24° N ~ 46.5° N, 118.5° E ~ 141° E 도메인을 대상으로 1.5° × 1.5° 격자 해상도에서 3주차(15~21일) 및 4주차(22일~28일)의 주간 평균 최고기온이다.

폭우 예측 모델의 시간 타깃은 선행 3주 및 4주의 폭우 발생이며, 주별 누적 강수량 및 극한강수일수를 목적 변수로 설정하였다. 공간 예측 타깃 지역은 Figure 2.1과 같다. 수도권 및 주요 도시, 지리적인 특성과 극한 강수 빈도가 높은 지역을 고려하여 총 5곳(서울, 부산, 대전, 광주, 평창)의 격자점을 선정하였다. 폭우의 경우 발생 공간 규모 및 발생 역학이 다른 현상에 비해 국지적일 수 있어 해상도 0.25도에 해당하는 격자점에서의 값을 라벨로 정의하였고, 각 격자점별로 개별 예측 모델을 개발하였다.

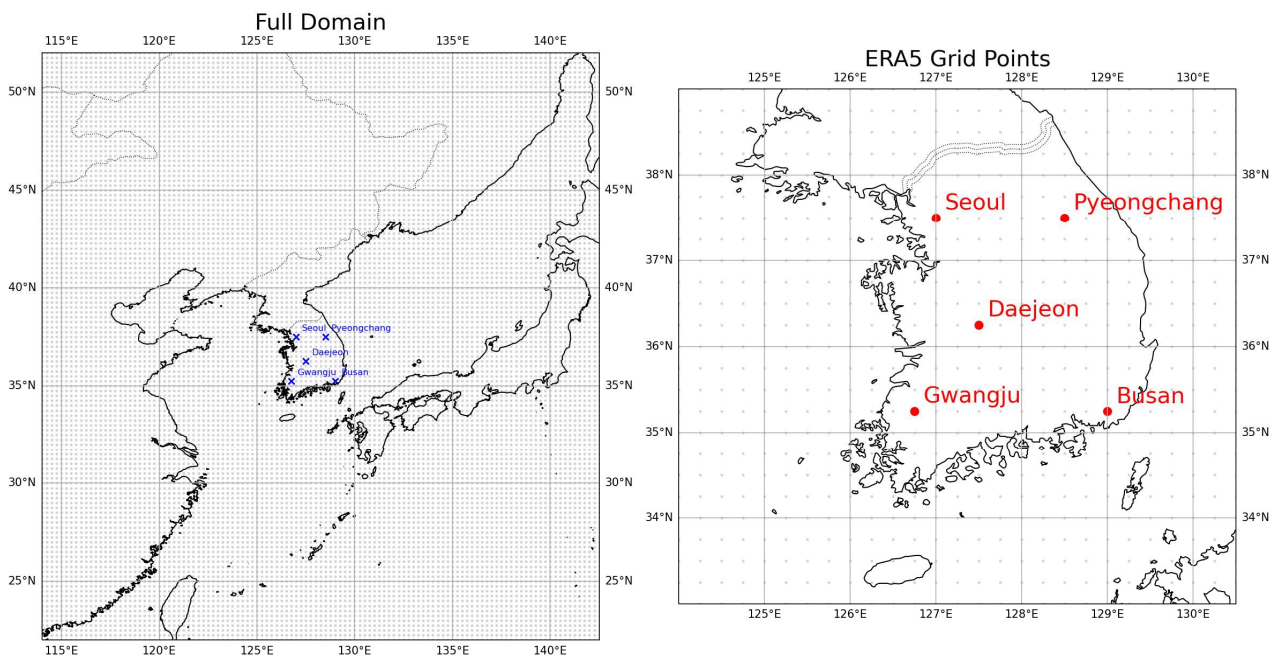


Figure 2.1. Study domain for extreme rainfall. The entire study domain commonly defined (left column) and enlarged view of the selected grid points (right column).

해양열파의 경우 이상고온과 마찬가지로 24° N ~ 46.5° N, 118.5° E ~ 141° E 도메인 1.5° 격자 해상도를 타깃으로 하되, 동해/황해/동중국해에 포함된 격자에 대해 선행 3주차 예측만을 수행하였다.

2.2. 훈련 자료 및 전처리

2.2.1. 관측 및 역학 모델 원자료

본 과제 of 이상기후 예측 인공지능 모델 훈련장은 ERA5(ECMWF Reanalysis v5) 재분석 자료(Copernicus Climate Change Service (C3S), 2017)에 기반해 구축되었다. ERA5 재분석 자료는 전 세계 기후와 날씨에 대한 유럽 중기예보센터(ECMWF, European Centre for Medium-Range Weather Forecasts) 5세대 재분석 자료이다. 재분석 자료(Reanalysis Data)란 전 세계에서 수집된 관측 데이터를 모델 데이터와 결합해 물리법칙에 따라 전 지구적으로 완전하고 일관된 자료로 만든 것이다. 이는 자료 동화(data assimilation)라는 방법을 통해 만들어지는데, 이는 일정 시간(ECMWF 기준 12시간)마다 이전 예보와 새로 들어온 관측자료를 최적으로 결합해 대기 상태의 최상의 추정값(즉, 분석, analysis)을 만들고 이를 바탕으로 최신 예보를 생산하는 방법이다. 이를 통해 관측자료를 기반으로 예전의 자료를 개선하여 재처리함으로써 과거 자료의 품질을 향상시킨다. ERA5는 대기, 해양, 육상에 대해 다양한 변수의 시간별 추정값을 제공한다. 불확실성 추정도 10개 멤버로 이뤄진 내부 앙상블로 샘플링해 3시간 간격으로 산출되며, 앙상블 평균 및 분산 자료도 함께 제공된다. 자료는 코페르니쿠스 기후변화 서비스(Copernicus Climate Change Service, C3S)에서 CDS(Climate Data Store)를 통해 온라인으로 쉽게 접근할 수 있다. 자료는 0.25도 공간해상도로 제공되며 전 지구 격자화된 등간격 위경도 표준격자로 매핑되어 있으며, 1940년부터 자료가 제공되고 있다. 시간별 자료는 ERA5 hourly data on single levels from 1940 to present, ERA5 hourly data on pressure levels from 1940 to present 자료셋에서 제공되고 있으며, pressure levels는 1000hPa부터 1hPa까지 37개 층으로 구성되어 있다. 시간별 ERA5 자료는 매일 업데이트되며 약 5일의 지연이 있다. 최근에 ERA5 post-processed daily statistics on single levels from 1940 to present와 ERA5 post-processed daily statistics on pressure levels from 1940 to present의 일일 자료셋이 출시되었다. 일평균, 일일 최대 및 최소 통계치로 제공되며, 원자료에서의 일일 하위 빈도 샘플링 간격은 1, 3, 6시간 간격으로 제공된다. 일일 통계치 자료는 매일 업데이트되며 약 6일의 지연이 있다.

선행 3/4주 이상고온·폭우·해양열파의 예측 원천으로서 잠재력이 있는 변수들을 조사하고, ERA5 재분석 변수 리스트 중에서 해당/유사 변수 추출·수집·전처리·정제의 과정을 거쳐 훈련장을 구축하게 된다. 이상기후 라벨 역시 ERA5에서 제공하는 최고 기온(Tmax), 총강수량(TP), 해수면 온도(SST)자료를 관측, 즉 참값으로 두고 입력장을 구축하였다. 동아시아 고수온 현상 관측 대응으로는 OISST (Optimum Interpolation Sea Surface Temperature) V2 일자료를 1982-2024 기간에 대해 추가적으로 활용하였다(Reynold et al. 2007; Huang et al. 2021).

계절 내 이상고온 오차 보정 및 폭우·해양열파 인공지능 예측 성능 평가를 위해 ECMWF 역학 모델 예측(Forecast) 자료를 활용하였다. ECMWF 예측 시스템은 매주 두 번 초기

화되어 모델 결과를 생산하며, 각 초기화 날짜(initialization date)별로 최대 46일까지 예측값을 제공한다. 예측 자료는 실시간 예측(real-time forecast)인 Forecast와 재예측(reforecast)인 Hindcast로 구성된다. Forecast는 1개의 제어 실행(control run)과 여러 개의 섭동 실행(perturbation run)을 포함한 101개의 앙상블 멤버로 이뤄져 있고, Hindcast는 동일한 구조의 11개의 앙상블 멤버로 구성된다. 여기서 control run은 초기 조건이 변하지 않은 기준 모델 실행을 의미하고, perturbation run은 관측 및 초기 조건에 작은 변동을 추가하여 불확실성을 반영한 모델 실행을 의미한다. Hindcast 자료는 특정 버전의 모델로 과거 기간을 재계산한 자료로 주로 모델 자체의 예측 성능 평가에 활용된다. Forecast 자료는 실시간 운영 예측으로 실제 예측 가능성을 검증하는 데 사용된다. 예를 들어 2024년 최신 모델 버전이 있을 때 2003-2022년(20년) 기간에 대해서 재예측된 Hindcast 자료가 생산되고 2024년에 대해서는 실시간 Forecast 결과가 제공되는 방식이다. ECMWF 자료는 수평 해상도 1.5도로 제공되며, 본 연구에서는 control run만을 사용하였다. 각 초기화 날짜에 대해 선행 3, 4주차에 해당하는 총 강수량을 추출하여 분석에 활용하였고, 2003-2022년 Hindcast 자료로부터 일일 강수량 기후값(climatology)과 90번째 분위수(90th percentile) 임계값을 계산하여 연구 전반에 사용하였다.

2.2.2 이상기후 예측 모델 입력 변수 및 라벨 선정

2.2.2.1 동아시아 이상고온(AHT) 오차 보정(Bias Correction)을 위한 훈련 자료

본 연구는 유럽중기예보센터(ECMWF, European Centre for Medium-Range Weather Forecasts)가 운영하는 Subseasonal-to-Seasonal Prediction System(이하 ECMWF-S2S)의 version 2024 Hindcast 앙상블 예측의 오차 보정을 목표로 하였다. ECMWF-S2S 예측은 46일 길이의 예측 기간을 제공하며, 초기(edge-start) 날짜 기반의 앙상블 예측을 포함한다. 이상고온 평가에 중요한 변수인 일 최고기온(TMAX)을 비롯하여 대기 순환 및 열수지와 관련된 총 8개 예측 변수를 훈련자료로 활용했다(Table 2.2). 공간 및 시간 해상도는 Table 2.1과 같다.

ECMWF ver2024 hindcast 세트는 2004~2023(20년) 기간 자료이고, 이에 대한 24° N ~ 46.5° N, 118.5° E ~ 141° E 도메인을 대상으로 1.5° × 1.5° 격자 해상도에서 선행 3주차(15~21일)의 주간 평균 최고기온과 기타 변수(Table 2.2)를 추출하였다. 관측 기준으로는 동일한 공간 영역과 시간 해상도를 갖는 ERA5 재분석의 일 최고기온을 사용하여 S2S 예측 리드 타임(lead-time)과 정합되도록 3주 평균장으로 구성하였다. 예측 변수 집합은 TMAX, TMIN, OLR, MSLP, Z500, 850hPa, 바람(u/v), 850hPa의 비습으로 구성되며, 이는 열역학, 대기역학, 수문장을 모두 반영해 이상 고온 발생의 물리적 요인을 충분히 설명하고자 한 선택이었다. 전체 기간 2004년에서 2023년 중에서 2004년에서 2021년을 학습 구간으로, 2022년에서 2023년을 테스트 구간으로 정의하였다. ECMWF S2S 버전 2024의 초기 날짜(initial date)를 기반으로 비교 평가를 위한 초기 날짜는 최근 지속적으로 따랐던 10월의 특성에 기초에서 10월 10일로 선정하였다.

이상고온의 오차 보정(bias correction) 모델 개발에서는 ERA5를 관측(true answer) 라벨 생성용 자료로 사용하였다. ERA5의 TMAX는 ECMWF-S2S TMAX의 예측값과 동일한 방식으로 3주 및 4주 평균을 산출하여 S2S 예측의 타겟(참값)으로 사용하였다. Table 2.1에 연구 도메인

의 시공간 해상도 및 대상 시기를 정리하였다.

Table 2.1 This table describes the configuration of the prediction targets and domains for a subseasonal TMAX forecasting experiment over East Asia. It summarizes the spatial and temporal domain, resolution, target variable, and lead-time settings for daily maximum temperature forecasts centered on the Korean Peninsula.

Description	Details
Spatial Domain	24° N ~ 46.5° N, 118.5° E ~ 141° E Includes East Asia's mid-latitude region, centered around the Korean Peninsula
Spatial Resolution	1.5 ° × 1.5 °
Temporal Resolution	Daily (per day)
Target Variable	Daily Maximum Temperature (TMAX)
Target Lead-time	7-day average corresponding to days 15-21 of 3-week daily forecast and 7-day average corresponding to days 22-28 of the 4-week daily forecast

Table 2.2 A set of atmospheric variables used for detecting and characterizing anomalously high-temperature. The table summarizes each variable's category and explains its specific role in identifying circulation, moisture, and pressure patterns associated with extreme high temperature.

Category	Variable Name	Purpose
Temperature	TMAX, TMIN	Key variables for labeling and detecting high temperature events
Radiation	OLR	Analysis of upper-atmosphere heat signal
Pressure	MSLP	Examination of spatial structure of low/high pressure patterns
Dynamics	Z500	Identification of mid-level atmospheric circulation patterns
	U850, V850	Reflection of lower-level jet and westerly wind variability
Moisture	Q850	Early signals of heat mass transport and heat dome formation

본 연구에서 이진(Binary) 라벨과 다중 클래스(Multi-Class) 라벨, 두 가지 라벨 기법(Labeling)을 적용하였다. 이진 라벨은 극한 고온의 단순 판정 여부에 초점을 둔 방식이다. 임계값은 상위 10% 초과하는 최고기온을 1로 정의하고 그 외 나머지는 0으로 정의하는 방법이다. 즉, 3주 평균 TMAX가 90 백분위수 이상이면 1로 변환하고, 그 외는 0으로 변환한다(Table

2.3). 이진 라벨의 장점은 극한 기온 비중이 작아서 데이터 불균형에 강하고 해석이 단순하며 예측 확률도 0 혹은 1이기 때문에 직관적이다.

다중 클래스 라벨은 극한 정도를 더 세분화하여 3개 구간으로 나눈다. 3주 평균 TMAX가 90 백분위수 이상이면 2로 분류하고 75 백분위수 이상에서 90 백분위수 미만이면 1로 분류하고 나머지는 0으로 변환한다(Table 2.3). 다중 클래스의 단점은 클래스 불균형이 심해지거나 softmax 기반의 다중 범주 확률 예측에서 경계 모호성이 증가할 수 있는 위험 요인이 있다. 또한 모델이 복잡해져서 연산량이 증가할 수 있지만 극한 강도를 더 정교하게 학습이 가능하다는 장점이 있다. 이진 및 다중 클래스 라벨의 임계값 적용을 Table 2.3에 요약하였다. ERA5-TMAX의 기준 기간은 ECMWF 버전 2024와 동일하게 설정하였다.

ERA5와 ECMWF-S2S의 차원이 달라서 차원 정렬(Dimension Alignment)이 필요하다. 또한 라벨 확장(Label Replication)을 ERA5-TMAX에 적용했는데, ERA5-TMAX의 라벨에는 앙상블이 없어 차원이 time, lat, lon이다. 앙상블 차원(ensemble dimension)이 있는 ECMWF-S2S의 TMAX는 time, number=11, lat, lon이다. 그래서 앙상블이 없는 ERA5-TMAX의 라벨의 차원을 앙상블이 있는 ECMWF-S2S의 TMAX 차원과 맞추기 위해 단순 반복, 즉 파이썬에서 “expand_dims”를 사용하여 ERA5에 앙상블을 생성하였다. 또한 4주 평균 TMAX 라벨도 3주 평균 TMAX와 동일한 방법으로 작성하였다. 즉, 4주 평균 TMAX가 90 백분위수 이상이면 1로 변환하고, 그 외는 0으로 변환한다(Table 2.4). 다중 클래스 라벨도 4주 평균 TMAX가 90 백분위수 이상이면 2로 분류하고 75 백분위수 이상에서 90 백분위수 미만이면 1로 분류하고 나머지는 0으로 변환한다(Table 2.4).

Table 2.3 Classification Criteria for Binary and Multi-Class Heat Extremes Based on 3rd Week Mean TMAX. The labeling rules used to categorize temperature conditions during the 3rd week (days 15-21) based on mean maximum temperature percentiles.

Labeling	Category	Criteria
Binary	Class1	3rd week (15-21 days) mean TMAX \geq 90th percentile
	Class0	3rd week (15-21 days) mean TMAX $<$ 90th percentile
Multi-Class	Class2	3rd week (15-21 days) mean TMAX \geq 90th percentile
	Class1	$75 \leq$ 3rd week (15-21 days) mean TMAX $<$ 90 percentile
	Class0	3rd week (15-21 days) mean TMAX $<$ 75 percentile

ECMWF-S2S 예측 자료에 대한 타깃, 즉 3주 평균 최고기온(이하 TMAX) 라벨 적용 방법은 ERA5 3주 평균 TMAX 라벨과 동일하다. ECMWF-S2S의 버전 2024의 과거재현 기간은 2004년부터 2023년(20년)이다. 일 최고기온 기준 기간에서 초기 날짜 기반으로 ECMWF-S2S 예측 자료를 활용하여 각 격자점별 3주 평균 TMAX에 대해 Table 2.3과 같은 분위수를 기반으로 이상고온을 정의하고, 이에 따라 라벨링 전략을 구성하였다. 먼저, 이진 라벨링은 TMAX가 90백분위수 이상이면 1, 그렇지 않으면 0으로 구분하는 방식으로, 단순성과 학습 안정성, 직관적 해

석이 빠른 장점이 있으나 강도 범위 표현의 한계와 불균형 문제를 동반한다.

반면, 다중 클래스 라벨링은 TMAX가 75백분위수 미만이면 0, 75백분위수에서 90백분위수 사이면 1, 90 백분위수 이상이면 2로 구분하여 보다 풍성한 범위를 제공하지만, 복잡성 증가와 경계 모호성이라는 단점이 존재한다. 또한 직접 라벨링과 계층적 라벨링 방식이 가능하나, 본 연구는 모든 샘플에 대해 직접 다중 클래스 분류를 적용하고, 특히 상위 10%에 해당하는 클래스 2의 발생 확률과 BSS(Brier Skill Score)를 주요 평가 대상으로 삼았다.

ECMWF-S2S의 4주 평균 TMAX 라벨도 3주 평균 TMAX와 동일한 방법으로 작성하였다 (Table 2.4). 즉, 4주 평균 TMAX가 90 백분위수 이상이면 1로 변환하고, 그 외는 0으로 변환한다. 다중 클래스 라벨도 4주 평균 TMAX가 90 백분위수 이상이면 2로 분류하고 75 백분위수 이상에서 90 백분위수 미만이면 1로 분류하고 나머지는 0으로 변환한다.

Table 2.4 Classification Criteria for Binary and Multi-Class Heat Extremes Based on 4th Week Mean TMAX. The labeling rules used to categorize temperature conditions during the 4th week (days 22-28) based on mean maximum temperature percentiles.

Labeling	Category	Criteria
Binary	Class1	4th week (22-28 days) mean TMAX \geq 90th percentile
	Class0	4th week (22-28 days) mean TMAX $<$ 90th percentile
Multi-Class	Class2	4th week (22-28 days) mean TMAX \geq 90th percentile
	Class1	$75 \leq$ 4th week (22-28 days) mean TMAX $<$ 90 percentile
	Class0	4th week (22-28 days) mean TMAX $<$ 75 percentile

모델에 훈련 자료로 입력하기 위해서는 ECMWF-S2S의 훈련 자료로 준비한 변수들의 단위가 모두 다르므로 전처리가 필요하다. 신경망 모델에 훈련 자료를 입력하기 전 훈련 자료에 대한 가공이 필요한데 이 중 전처리 과정은 매우 중요하다. 흔히 전처리는 자료 전처리(data preprocessing)를 의미한다. 자료를 모델이 이해할 수 있는 형태로 변환하거나 자료의 품질을 올리는 일련의 과정으로, feature tuning이라고도 한다. 전처리는 여러 단계로 이루어지는데, 단계별로 처리된 자료는 3가지 형태로 명명할 수 있다. 첫 번째가 원시자료(raw data)로 초기에 수집된 원본 데이터로 분석이나, 기계학습 혹은 심화학습의 학습 용도로 전혀 전처리가 되지 않은 데이터로 데이터 저장소 등에 저장된 데이터가 이에 해당된다. 두 번째는 정제자료(prepared data)는 Data engineering, 즉 전처리 기법에 따라서 학습을 위한 데이터로만 추출한 데이터(예, subset)를 의미한다. 예를 들면, 앞서 언급한 바와 같이 S2S 예측 자료는 다루기 어려운데, 이것을 (시간, 위도, 경도, 변수) 등으로 바꾸고, 모형 자료라서 -값이나 0이 없겠지만 nan 데이터가 있는지도 확인해서 준비된 데이터를 훈련 자료라고 할 수 있다. 따라서 본 연구에서는 원시자료가 정제자료까지의 단계를 말한다. 세 번째가 훈련 자료(예, engineered feature 혹은 engineered data)이다. 정제된 데이터가 딥러닝 학습에 적절한 형태로 재가공 되

어야 하는데, 이를 Feature Engineering이라고 한다. 예를 들어, 숫자와 같은 값을 0에서 1 사이로 변환시키거나 범주형 자료로 예를 들어 남자 혹은 여자를 0 혹은 1과 같은 값으로 변환시키는 과정이다. 본 연구에서 전처리 방법으로 표준화(Standardization)를 선정하였다. 표준화는 변수 스케일 차이를 큰 복수의 기상 변수를 U-Net 모델에 투입하기 때문에 각 변수를 다음과 같이 standard score로 정규화하였다(Eq. [2.1]). 각 변수별로 2004년에서 2023년 전체 학습 기간의 평균과 표준편차를 산출하여 모드 시점에 동일한 파라미터로 적용하였다(Casella et al., 2002).

$$X' = \frac{X - \mu}{\sigma} \quad [2.1]$$

여기서 X 는 원래 값(관측값 혹은 입력변수)이고, μ 는 그 변수의 평균(mean)이며, σ 는 표준편차(standard deviation)을 의미한다.

2.2.2.2. 한반도 주요 격자점 폭우(HR) 예측용 훈련자료

선행 3, 4주 폭우 예측 인공지능 모델 개발에 활용된 기본 입력 변수 목록은 Table 2.5과 같다. 입력 변수는 강수량 예측과 관련된 여러 영향 요소를 고려하였다. 우리나라 강수에 영향을 미치는 요인은 크게 대기, 해양, 지표에 의한 영향으로 나뉠 수 있다 (Figure 2.2). 극한 강수를 예측하기 위해서는 서로 다른 시간 규모에서 영향을 미치는 다양한 물리 변수를 함께 고려할 필요가 있다. 먼저 대기 영향 변수로는 상하층 바람(U200/850, V200/850)과 지위고도(Z500), 해면기압(MSLP) 등이 포함되며, 이는 저기압, 장마전선, 기압골과 같은 대기 순환의 강도와 위치를 나타내 극한 강수의 동적 역학적 조건을 설명한다. 수증기 관련 변수인 850hPa 상대습도(Q850)와 총 대기 수증기량(Total Column Water Vapor, TCWV)은 강수에 필요한 수증기 공급량과 수송 경로를 진단하는데 도움을 줄 수 있으며 장기간의 습윤, 건조 패턴을 통해 강수 잠재력을 파악할 수 있을 것이다. 복사 관련 변수인 OLR은 적운활동과 대류 강도를 나타내며 열대 대류와 중위도 지역의 대기 파동과의 연결을 통해 폭우를 유발하는 요인을 나타낼 수 있다. 표면 영향 변수로는 지면온도(Skin temperature)와 토양수분을 활용했으며, 이는 지표 가열과 잠열 공급, 경계층 안정도 변화 등을 통해 대기 대류를 강화하거나 억제하는 역할을 한다. 마지막으로 해양 영향 변수인 해수면온도(SST)는 해수와 대기 간 열, 수증기 교환을 통해 장주기의 수증기 공급 조건을 결정하며, 원격 상관(teleconnection) 패턴을 통해 폭우 발생에 간접적인 배경장 정보를 제공할 수 있다.

0.25도의 공간해상도를 가지는 ERA5 입력 변수들은 모두 2.5도로 양선형 보간법(Bilinear Interpolation)을 적용하여 업스케일링(upsampling) 하였다. 저해상도로의 변환은 입력 변수의 다변화에 의한 전산량 증가를 완화하여, 다양한 입력 변수 조합, 모델 구조 및 하이퍼파라미터 최적화 등의 실험을 보다 효율적으로 수행할 수 있는 장점이 있었다.

총 세 단계를 거치는데 첫째, 천천히 변화하는, 즉 주기가 90일 이상인 기후학적 연주기 성분을 제거한다. 이는 원시자료 (X)에서 90일 저주파 통과(low-pass) 필터로 걸러낸 기후학적 성분(\bar{X})을 빼는 방식으로 구해진다 (Eq. [2.2]).

$$X' = X - \bar{X} \quad [2.2]$$

여기서 아노말리 장(X')은 10~60일 구간의 계절 내 성분, 고주파(synoptic-scale) 성분, 장기(연간 혹은 수십 년 규모)의 저주파 성분이 여전히 포함되어 있다. 여기서 10~60일의 계절 내 성분만을 남기기 위해 두 번째 단계로 마지막 30일(즉, day -30일부터 day 0까지)의 이동평균(running mean)을 통해 저주파 신호를 제거한다 (Eq. [2.3]).

$$X'' = X' - \overline{X'}^{30d} \quad [2.3]$$

마지막 단계로 이 잔차(X'')에 5일 이동평균으로 추가적인 synoptic-scale 변동성을 제거한다 (Eq. [2.4]).

$$X^* = \overline{X''}^{5d} \quad [2.4]$$

Hsu 외 (2015)에서는 이렇게 얻어진 10~60일 성분을 Butterworth band-pass 필터링 결과와도 비교하여 높은 상관관계를 보임으로써 방법론이 유효함을 입증하였다.

본 연구에서는 추가로 $X'' - X^*$ 를 계산하여 이를 Synoptic component로 정의하였다. 따라서 입력 변수에 대한 High frequency component (Synoptic-scale component), ISO component (X^*), Low frequency component ($\overline{X'}^{30d}$), Background component (Climatology, \bar{X})의 4가지 성분들을 활용하였다. Figure 2.3에서 제시된 바와 같이 각 기본 입력 변수에 대해 이 4가지 성분을 개별적으로 추출함으로써, 다중 규모 분해를 통해 생성된 추가 입력 변수를 모델에 활용하였다. 이러한 접근은 기존의 단일 규모 분석에 비해 다양한 시간 척도의 변동성 정보를 제공하여, 기상 현상의 복잡한 변동성을 포착하는데 유리할 수 있다. 따라서 본 방법론을 통해 입력 자료의 다양성을 확대하여 폭우 예측 모델의 성능 개선을 도모하고, 이를 중심으로 본 연구의 모델 개발 및 평가를 수행하였다.

전체 입력 변수들은 Table 2.6와 같다. 각 입력 변수의 특성에 따라 시간 스케일별 기여도가 상이할 수 있으므로, 변수별로 적합한 성분만을 선택적으로 추출하는 것이 이상적일 수 있

다. 그러나 입력 변수의 각 성분이 3-4주 후 강수량 예측에 구체적으로 어떻게 기여하는지에 대한 기존 지식으로는 명확히 파악하기 어렵기 때문에, 본 연구에서는 우선 모든 입력 변수에 대해 모든 성분을 추가 입력 변수로 활용하였다. 최종적으로 다중 규모 분해를 적용한 총 입력 변수 목록은 Table 2.6와 같다. 모델 학습 및 평가 과정에서 최적의 입력 변수 조합을 탐색한 결과, 고주파 성분 변수들(*_hfreq)을 제외한 ISO, 저주파, 배경장 입력 변수 조합이 최종 모델의 입력 변수로 확인되었다 (Table 2.6의 굵은 글씨).

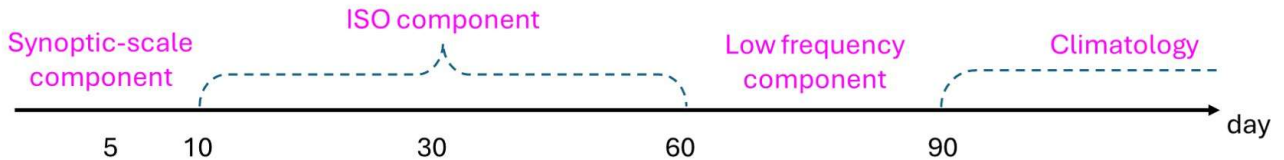


Figure 2.3. Schematic of multi-scale temporal components of daily variability.

Table 2.6. List of multi-scale decomposed input variables used in this study. A total of 48 input variables were used and tested in this study. The variables shown in bold are the final input variables selected in this study.

High frequency	ISO component	Low frequency	Climatology
U200_hfreq	U200_iso	U200_lfreq	U200_bgd
U850_hfreq	U850_iso	U850_lfreq	U850_bgd
V200_hfreq	V200_iso	V200_lfreq	V200_bgd
V850_hfreq	V850_iso	V850_lfreq	V850_bgd
Q850_hfreq	Q850_iso	Q850_lfreq	Q850_bgd
TCWV_hfreq	TCWV_iso	TCWV_lfreq	TCWV_bgd
OLR_hfreq	OLR_iso	OLR_lfreq	OLR_bgd
Z500_hfreq	Z500_iso	Z500_lfreq	Z500_bgd
MSLP_hfreq	MSLP_iso	MSLP_lfreq	MSLP_bgd
ST_hfreq	ST_iso	ST_lfreq	ST_bgd
Soil water_hfreq	Soil water_iso	Soil water_lfreq	Soil water_bgd
SST_hfreq	SST_iso	SST_lfreq	SST_bgd

3, 4주 선행 강수량 예측을 위한 목적 변수 라벨링 방법에 대한 모식도는 Figure 2.4와 같다. 그림은 약 3주간의 일일 강수량 시계열 일부를 예시로 보여준다. 이 중 90th percentile 임계값을 초과하는 사례를 이상 강수 케이스로 정의하고 굵은 글씨로 표시하였다. 7일 이동합(rolling sum)을 적용할 때 해당 window 내에 이상 강수 케이스가 포함되면 이상 강수일수를 카운트하여 window의 가장 왼쪽 지점에 값을 부여하였다 (left anchored). 마찬가지로 주간 누적 강수량 역시 7일 이동합을 통해 window 왼쪽에 할당하였다. 빨간색 숫자는 극한 강수(extreme events)에 해당하며, 파란색 숫자는 비극한 강수(non-extreme events) 사례를 나타낸다.

더 나아가 극한강수일은 Table 2.7에 제시된 바와 같이 4가지 카테고리로 세분화하여 목적 변수를 구축하였다. 이는 Figure 2.5에서 확인할 수 있듯이, 극한강수일(빨간 막대)이 적더라도 누적 강수량(파란 막대)이 많은 경우와 반대로 극한강수일이 많아도 강수량이 상대적으로 적은 경우가 공존하는 특성을 반영한 것이다. 이러한 패턴의 다양성을 학습하기 위해 극한강수일을 카테고리화함으로써, 각 클래스 별로 다른 강수량 분포를 모델이 효과적으로 포착할 수 있도록 설계하였다.

전체 훈련자료는 1950년부터 2024년까지 일일 자료에 대해서 구축되었다. Table 2.8에 보여주는 것처럼 전체 자료는 훈련(training), 검증(validation), 테스트(test)용 자료로 구분하였다. 훈련은 1950년부터 2008년 11월까지, 검증은 2009년부터 2022년 11월까지, 테스트 자료는 2023, 2024년으로 설정하였다. 훈련 및 검증을 11월까지 설정한 이유는 3, 4주 선행시간을 고려해서 자료 간에 겹치는 기간이 없도록 하기 위함이다. 자료 간에 기간이 겹치게 되면 학습이 제대로 되지 않거나 테스트 평가 시 공정한 평가가 불가하다. 이를 방지하기 위해서 한 달의 간격을 두었다. 훈련 및 검증 자료는 모델 훈련에만 활용된다. 모델은 훈련자료에 대해서 학습하고 검증자료에 대해서 내부적으로 평가하여 검증자료에서의 성능이 일정 기간 향상되지 않을 때 학습을 멈추도록 설계하였다. 최종 모델이 만들어지면 그 이후 테스트 자료가 사용되어 최종 모델이 테스트 자료에 대한 예측 결과를 산출하고 이를 바탕으로 최종 모델의 성능을 평가하게 된다. 모델 훈련 과정에서는 일별 7일 이동합이 적용된 주간 극한 강수일수와 누적 강수량을 동시에 목적 변수로 활용하여 학습을 진행하였다. 반면 테스트 자료에 대한 모델 성능 검증은 실제 운영 환경을 반영하여 매주 월요일 예측만을 대상으로 수행하였다.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Date	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Rainfall (mm/day)	7	13	7	13	83	13	11	1	2	2	3	5	11	13	15	30	107	95	50	89	17	10
Count (Rollsum)	1	1	1	1	1	0	0	0	0	0	1	2	2	3	3	3	3	2	1	1	1	1
Amount (Rollsum)	147	141	130	125	115	37	35	37	51	79	184	276	321	399	403	398	378	273	185	148	142	138

Figure 2.4. Example of how weekly rolling sum on daily rainfall works. Rainfall amount (mm/day) shown in bold is the extreme event that exceeds 90th percentile threshold. Figures shown in red are the extreme cases, and those in blue are the non-extreme cases.

Table 2.7. Categorical labels assigned to each count of extreme rainfall events.

Extreme rainfall counts	Categorical label
0 (No extreme)	Class 0
1 extreme day/week	Class 1
2-3 extreme days/week	Class 2
4+ extreme days/week	Class 3

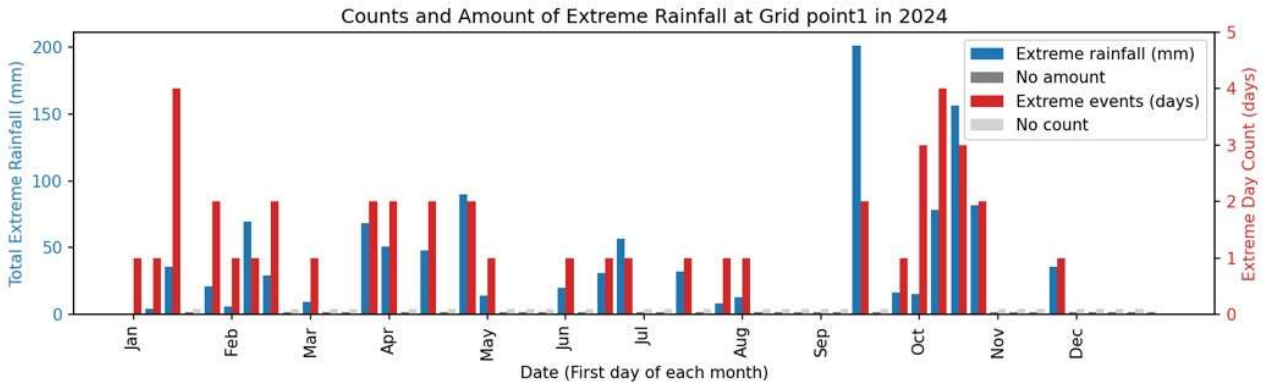


Figure 2.5. Extreme rainfall counts (red bars) and extreme rainfall amount (blue bars) at a grid point in 2024.

Table 2.8. Data split strategy for extreme rainfall prediction model.

Data Split	Period
Training	1950 - 30-Nov-2008
Validation	2009 - 30-Nov-2022
Test	2023 - 2024

2.2.2.3 동아시아 해역 고수온(MHW) 예측용 훈련자료

동아시아 해역 고수온 현상 예측을 위해 다양한 입력 변수를 고려하였다(Table 2.9). 기본 변수로는 해수면 온도를 직접적으로 조절하는 해양열수지(hc: ocean heat content), 염도(sl: salinity), 지표 기온(t2m: 2m temperature above surface), 지상 바람(u10/v10: 10m zonal and meridional wind), 그리고 강수(tp: total precipitation)를 설정하였다.

상층 대기 수송과 관련된 변수로는 u200/v200 (zonal/meridional wind at 200 hPa), u850/v850 (zonal/meridional wind at 850 hPa)를 포함하였다. 또한 tcc (total cloud cover), ssrd (surface solar radiation downwards), strd (surface thermal radiation downwards)등 해수면으로 입사하는 복사와 관련된 플럭스 변수들도 고려하였다.

선행 3주 예측을 위한 가용 메모리는 주로 해양 자체의 에너지와 그 수송에 기반한다. 이에 따라 sst_ep (90%tile 기준치를 초과하는 날의 주별 누적 편차값)을 고수온 자체 변수로 활용하였으며, mwd (mean wave direction), shww (significant wave height of wind waves)등 웨이브와 관련 변수도 기본 입력으로 설정하였다. 해류 수송 변수를 입력으로 포함하는 방식을 검토하였으나, 대기 재분석 자료에 비해 자료 기간이 현저히 짧아 본 연구에서는 훈련에 포함하지 않았다. 다만 향후 가용성이 확보될 경우 추가 활용을 고려할 예정이다.

대기와 해양을 구분하기 위한 정적 변수로는 lsm (land-sea mask)와 위도(latitude)를 포함

하였다. 또한 지구 온난화로 인한 고수온 현상의 빈도 및 강도가 증가하는 점을 반영하여, 날짜(time step)정보를 추가하여 time-aware input으로 활용하였다.

3주 후 고수온 현상 예측을 위한 현재 상태 입력 변수의 공간적 범위와 해상도는 세 가지 옵션으로 설정하였다: G5(Global domain, $5^\circ \times 5^\circ$ 해상도), L5(Large domain, $5^\circ \times 5^\circ$ 해상도), S2.5(East Asia subregional small domain, $2.5^\circ \times 2.5^\circ$ 해상도). 또한 입력 시퀀스 구성 방식으로는 가장 최근 1주 자료를 활용하거나 최근 3주 자료를 적층(stacking)하는 두 가지 옵션을 적용하였다(Table 2.9).

Table 2.9 Training input: variables and temporal/spatial range

Training INPUT	VARIABLES	
Comp.	Base (Ocean & Atmos)	hc & /sl and u10/v10/t2m/tp
	Upper Atm.	u200/v200/u850/v850
	Flux	tcc/ssrd/strd
	Wave & Swell	mwd/shws
	Ocean	sst_ep
	Time invariant (Static) / Time	land-sea mask, latitude / Date (step)
T/P range	3(1)W: R(-3)/R(-2)/R(-1) or R(-1) G5: [0~0~82.5W, 65S~65N], 128x27, 5° res. L5: [160W~157.5E, 65S~65N], 64x27, 5° res. S2.5: [70~147.5E, 20S~57.5N], 32x32, 2.5° res.	

주별 고수온 현상을 나타내는 지표로 90%tile 기후학적 임계값에 대한 SST 편차 7일 누적 값을 극한도(Impact Factor)를 새롭게 정의하여 고수온현상을 대표할 격자별 라벨로 사용하였다

2.3. 인공지능모델 및 실험 구성

2.3.1. 동아시아 선행 3/4주 이상고온 역학 예측 오차 보정 모델 개발

2.3.1.1. 오차 보정 모델 구성

이진 및 다중 클래스 라벨을 두 가지 딥러닝 모델, U-Net과 Attention U-Net에서 훈련하였다(Ronneberger et al., 2015; Oktay et al., 2018; Weyn et al., 2021). 본 연구에서 사용한 U-Net은 encoder-decoder 구조를 기반으로 하는 기본 U-Net으로, skip connection을 통해 공간 해상도 손실을 보완하는 구조이다(Figure 2.6). U-Net은 영상 기반의 세그멘테이션(segmentation)에 최적화된 모델로 지역적 특징(local features)과 전역적 구조(global pattern)를

모두 학습할 수 있어서 이상 기온의 공간 패턴 추정에 적합할 것이다. Attention U-Net은 기본 U-Net에 Attention Gate를 삽입하여 skip connection의 정보 중 중요한 공간 영역만 선택적으로 전달하도록 개선한 구조이다(Figure 2.6). Attention Gate는 불필요한 주변 잡음을 억제하여 모델의 신호-잡음비(signal to noise ratio)를 개선할 수 있다. 즉, 관련 없는 영역의 활성화를 억제하고 관련 영역에 집중하도록 학습하는 것인데 Attention Gate가 고온 anomaly의 핵심 위치를 강조하는데 기여할 수 있다. 이에 따라 실험 결과에서 Attention U-Net이 U-Net보다 BSS 개선률이 높게 나타난 것으로 판단된다.

U-Net 구조는 인코더-디코더(encoder-decoder) 형태에 skip connection을 결합하고 32-64-128 채널로 구성된 합성곱 레이어를 통해 지역적 이상고온에 관련된 다중 규모 공간 패턴을 효과적으로 학습하도록 설계되었다. 기본 U-Net은 2015년 Ronneberger et al. (2015)에 의해 제안되었으며 6단계의 1차원(22년) 연구의 기본 모델이다(APEC기후센터, 2022, 2023). U-Net은 이미지의 전반적인 특징 정보를 얻기 위한 수축 형태 (Contracting Path)의 네트워크와 정확한 지역화를 위한 팽창 형태 (Expanding Path)의 네트워크를 U자형의 대칭 형태로 갖는데, 기후 분야에서 활용도가 높은 모델이다 (Weyn et al., 2021). 이미지 세그멘테이션(image segmentation)이란 이미지의 모든 픽셀이 어떤 카테고리(예를 들면 자동차, 사람, 도로 등)에 속하는지 분류하는 것을 말한다. 이미지 전체에 대해 단일 카테고리를 예측하는 이미지 분류(image classification)와는 달리, 이미지 세그멘테이션은 픽셀 단위의 분류를 수행하므로 일반적으로 더 어려운 문제로 인식되고 있다. U-Net은 이미지의 전반적인 특징 정보를 얻기 위한 수축 형태(Contracting Path)의 네트워크와 정확한 지역화를 위한 팽창 형태(Expanding Path)의 네트워크를 U자형의 대칭 형태로 갖는데(Ronneberger et al., 2015; Weyn et al., 2021), 선례 연구에서는 TimeDistributed 레이어를 U-Net의 Convolution 레이어에 추가하여 시계열 효과를 주고자 했다. TimeDistributed는 차원 변경 없이 간단하게 2D convolution을 사용할 수 있으므로 5차원 배열의 훈련자료가 바로 입력된다.

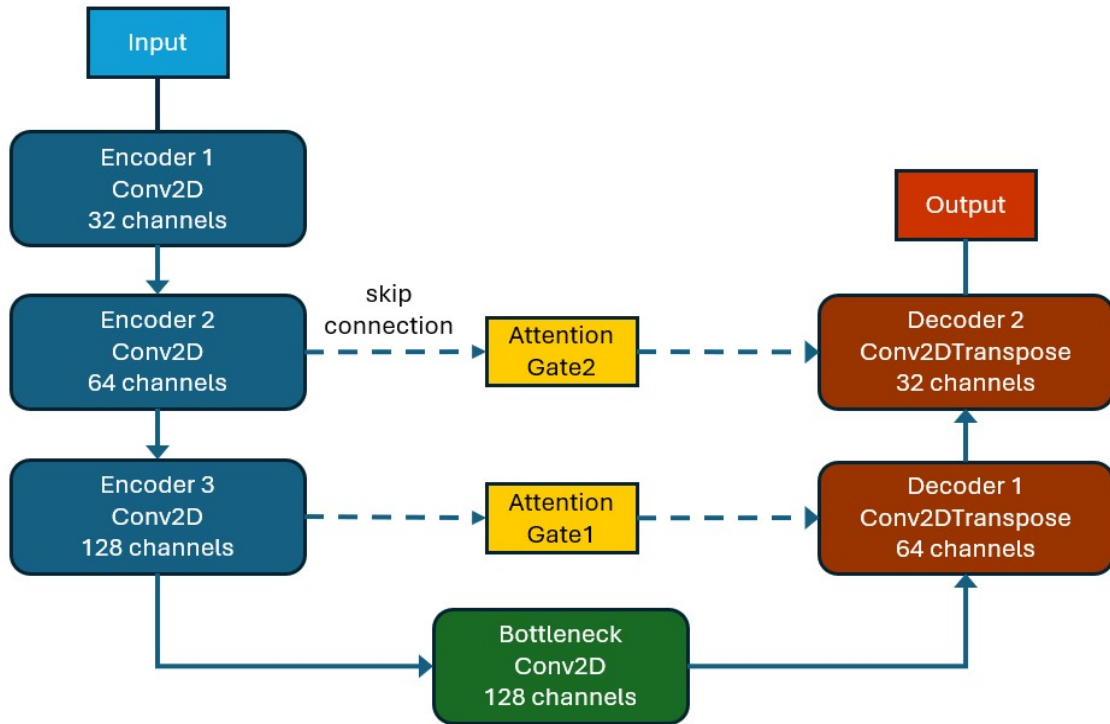


Figure 2.6 Attention-Enhanced U-Net Architecture Overview. The diagram illustrates an encoder-decoder U-Net architecture augmented with attention gates that refine skip-connection features. It shows how multi-scale encoded features pass through attention modules before being merged with decoder layers to produce the final output.

또한 Attention U-Net에서는 skip connection 경로에 Attention Gate(이하 AG)를 통합하여 고은 이상이 두드러지는 공간 영역의 중요도를 증폭시킴으로써 특징 선택을 강화하고 노이즈를 억제하는 효과를 기대하였다. 앞서 언급한 바와 같이 U-Net 모델은 image segmentation을 목적으로 제안된 fully-convolutional 기반 모델 형태가 U자형으로 context 포착과 세밀한 localization을 위해서 대칭으로 구성되었다. 이에 반해 CNN은 인간의 시신경을 통해서 모습이 보이는 것에서 착안하여 만든 딥러닝 모델이다 (Oktay et al., 2018). 우리의 눈은 전체를 보지만 일부를 집중해서 보는 것과 같이 겹치는 부분이 있는데, 이처럼 CNN은 이미지 처리에 특히 특화되어 있으며 convolutional layer, pooling layer가 있다. U-Net, U-Net++, VB-Net 등이 다양하게 사용되다가, 기존 U-Net의 학습 정확도를 개선하기 위해 Attention Gate(이하, AG) 장치가 장착된 Attention U-Net을 제안하였다(APEC기후센터, 2023, 2024). Attention이란 이미지 세그멘테이션에서 Attention은 학습할 때 관심 있는 영역 혹은 관련된 영역을 강조하는 방법이다. 예측이나 분류할 정답이 관련 없는 영역들을 무시하거나 관심을 덜 두기 때문에 계산 비용이 줄어든다는 특징이 있는데, 이미지의 특정 영역에만 집중(Attention)한다는 것이다. 여기에 U-Net의 주요 특징은 Skip Connection이 있다는 것이다. 업샘플링이 진행되는 expanding path에서 소실되었던 공간 정보(다운 샘플링=contracting path에서 공간정보가 소실)를 다시 만드는 데, 문제는 이 정보가 부정확하다는 것이다. 그래서 U-Net에서는 Skip Connection을 이용해서 다운 샘플링하는 contracting path의 공간정보를 업샘플링 path에 더해 이 문제를 어느 정도 해

결했다. 하지만 Skip Connection을 사용함으로써 추출된 특징들이 중복된다는 부가적인 문제가 발생했다. 신경망의 초기 layer들의 특징 표현이 조잡(poor) 하기 때문이다. Attention 메커니즘은 2016년에 computer vision에서 사용되었다(Bahdanau et al., 2016). Neural Network에서 prediction을 만드는 것이 무엇인지 이해하기 위해서 사용되었다. Attention 메커니즘은 입력에 따른 다음 결과를 예측하는 작업에서 특정 정보에 더 집중해야 하기에 encoder에서의 어떤 값에 집중해야 하는지를 찾는 방법이다.

AG를 다시 설명하면 CNN 아키텍처와 쉽게 통합할 수 있고, 최소한의 계산량 증가와 함께 모델의 민감도와 예측 정확도를 향상될 수 있다. 이것은 U-Net도 Convolutional layer에 ReLU 함수가 결합한 layer가 한 쌍으로 CNN에서 파생되었으므로 CNN과 잘 맞을 것이며, dimension이 줄어들기 때문에 비용(예, 계산 속도 등)이 감소한다는 것은 즉, 계산량이 작은 상태에서도 정확도를 높일 수 있으니 효율적이다. 이것은 layer가 깊어지면 모델 매개변수(parameter)가 과도하게 중복되고 유사한 저수준의 피처(feature)가 반복적으로 추출되므로 계산 리소스도 많이 들기 때문이다. 이것을 개선하려고 시도한 것이 AG이다. 그러나, Attention은 길이 다시 말하면 time-sequence가 길어질수록, 내적 연산에 따른 속도 저하가 매우 크다고 알려져 있다. Beam-Search 알고리즘 등을 사용하면 속도 저하 문제를 조금 타파할 수는 있겠지만, 분류 문제에서 선택이 가능하고, 회귀 문제에서는 선택 불가능하다. 예를 들면, Encoder와 Decoder의 time-series가 무척 긴 경우(예를 들면, 약 200개 이상씩 될 때), 매 과정에서 찾아 나가서 결정하다 보니 느려지는 경향이 있다고 한다. Attention은 정해진 Time-Sequence에서만 집중할 수 있다.

그래서 Attention 시계열 데이터 예측에 관한 연구가 비교적 최근부터 시작되었다. 하지만 현실의 문제는 다르다. 실제 데이터가 100일 범위로만 결정될 수 있을 것으로 생각했지만, 특정 어떤 데이터는 200일 전 데이터가 Attention이 더 잘 될 수도 있기 때문이라고 Qin et al. (2017)은 언급하였다. 이런 데이터 구성의 경우 Attention이 잘 동작하지 않는다. 또한, 대부분의 문제 해결 방식이 CNN을 이용하고 Attention을 진행하기 때문에, CNN을 적용하는 부분이 데이터셋과 적합한 모델 아키텍처 구조인지도 판단해 봐야 한다고도 언급하였다(Shih et al., 2019). 그 데이터의 카테고리가 매우 제한적이며 실제 사례에 적용된 레퍼런스를 찾아볼 수 없다. Attention은 어떤 한 시점을 기준으로 유한한 단어를 학습시키기 위해 최적화 되어있고, 현실 세계에서 주가와 같은 무한한 실수 및 정수 예측 문제에서는 활용할 수 없다는 것이다. 대부분 %와 같은 상대적인 유한한 값으로 치환하여 사용한다든지 하는 형태로 활용만 하고 안정성이 떨어진다고 한다. 그래서, seq2seq나 Encoder-Decoder 기반의 BI-LSTM 형태로 구성하는 것이 현재로서는 가장 합리적인 Time-Series Forecasting 방법이라고 제안하였다.

2.3.1.2 모델 성능 평가를 위한 실험 설계

모델 학습 구성은 입력 특성에 따라 2종의 라벨과 2종의 훈련모델로 구성된다. X는 입력, Y는 출력이라고 정의한다. 출력층과 손실함수는 라벨 형식에 따라 이진 분류의 경우 시그모이드 활성화(sigmoid activation) 함수와 BCE(binary cross-entropy)를 적용하고, 다중 클래스의 경우는 소프트맥스 활성화 (softmax activation) 함수와 CCE(categorical cross-entropy)을 적용하였

다(Table 2.11). 이를 통해 출력값을 클래스별 확률로 해석한다. 학습 설정 측면에서는 변수 전반에 걸친 입력 표준화(normalization), 배치 구성과 최적화 방식, 그리고 early-stopping 전략을 포함하여 모델 훈련 중 검증 성능이 더 이상 개선되지 않을 때 학습을 조기에 중단시켜 과적합을 방지하고 일반화 성능을 최적화하기 위해 사용된다. 초매개변수(hyperparameter)는 모델을 생성할 때 사용자가 직접 설정하는 변수이고, 파라미터(parameter, 매개변수)는 모델의 학습 과정에서 생성되는 변수이다. 신경망 모델에서 에포크(epoch), 배치 사이즈(batch size), 학습률(learning rate), 활성화 함수(activation) 등이 초매개변수에 속한다(Table 2.10).

Table 2.10 Hyper-parameter Settings of the Study. The table summarizes the key hyper-parameters used in this study, including training configuration and model architecture settings. It lists values such as the number of epochs, batch size, learning rate, optimizer, activation function, and kernel sizes applied.

Hyper-parameter	Setting of the study
epoch	100
batch size	64
learning rate	0.0001(1e-04)
Optimizer	Adam
Activation function	ReLu
Filter size (Kernel size)	16, 32

이진 분류의 훈련을 위한 입력은 7개 ECMWF-S2S 예측 변수를 ECMWF의 TMAX에 이진 분류를 적용한 자료와 함께 결합한 8개 ECMWF-S2S 예측 자료와 ERA5의 TMAX에 이진 분류를 적용한 훈련 자료로 구성하였다. 다중 클래스 훈련을 위한 입력은 역시 7개 ECMWF-S2S 예측 변수에 ECMWF의 TMAX에 다중 클래스 분류를 적용한 자료와 함께 결합한 ECMWF-S2S 예측 자료와 ERA5의 TMAX에 다중 클래스 분류를 적용한 훈련 자료로 구성하였다. 이렇게 각각의 훈련 자료를 훈련 모델에 입력하여 훈련 후 예측하면, 이진 라벨의 경우 활성화(activation) 함수로 시그모이드(sigmoid)를 적용하므로 확률값으로 출력된다.

다중 클래스의 경우는 마지막 레이어의 활성화 함수를 소프트맥스(softmax)를 적용하기 때문에 클래스별 확률 벡터로 출력된다. 훈련 모델의 훈련 환경에서 손실 함수(loss function)로 이진 라벨의 경우는 Binary Cross Entropy(BCE)를 다중 클래스 라벨의 경우는 Categorical Cross Entropy(CCE)를 적용하였다(Table 2.11). 옵티마이저(optimizer)로는 두 훈련 모델에서 공통적으로 Adam을 적용하고 학습률(learning rate)은 $1e-4(0.0001)$ 로 ReduceLROnPlateau를 적용하여 검증 손실 등의 지표가 일정 에포크 동안 개선되지 않을 때 학습률을 자동으로 줄여 모델 학습의 정체를 극복하고 수렴을 촉진하도록 설정하였다(Kingma et al., 2017; Mehmood et al., 2023). 이를 통해 학습률이 너무 크거나 작아 발생하는 발산이나 느린 수렴 문제를 효과적으로 방지할 수 있습니다. 검증 손실 감소를 멈춤 할 때 early stopping이 되도록 설정하였다. 훈련 기간은 전체 기간 중 90%에 해당하는 2004년에서 2021년(18년)으로 설정하고 나머지 10%인 2022년에서 2023년을 테스트 즉 예측 기간으로 설정하였다. 또한, 훈련 기간 내 20%를 검증

기간으로 분리하도록 설정하였는데 이때 훈련 자료를 섞는 기능인 셔플은 사용하지 않았다.

Table 2.11 Activation Functions and Loss Functions for U-Net and Attention U-Net According to Label Type. This table summarizes which activation functions and loss functions should be used when training U-Net or Attention U-Net models for binary and multi-class segmentation tasks. It outlines appropriate combinations such as Sigmoid with Binary Cross Entropy for binary labels and Softmax with Categorical Cross Entropy for multi-class labels.

Type of label	Type of training model	
	U-Net	Attention U-Net
Binary	Activation: Sigmoid	
	Loss function: Binary Cross Entropy	
Multi-Class	Activation: SoftMax	
	Loss Function: Categorical Cross Entropy	

표본 규모와 예측 독립성을 고려하여 2004년에서 2021년을 학습 구간으로, 202-2023년의 2년을 테스트 구간으로 구분하였다. 실험 구성은 2종 라벨과 2종 훈련 모델 조합, 즉 이진 라벨과 U-Net 훈련(Binary+U-Net), 이진 라벨과 Attention U-Net 훈련(Binary+Attention U-Net), 다중 클래스 라벨과 U-Net 훈련(Multi-Class+U-Net), 마지막으로 다중 클래스 라벨과 Attention U-Net(Multi-Class+Attention U-Net)으로 구성하였다. 예측 출력은 격자점별 사건 발생 확률로서 이진 라벨의 경우 해당 사건의 확률을, 다중 클래스 라벨의 경우 클래스 2(이상고온)의 확률을 주요 평가 대상으로 선정하였다. 비교 기준선으로는 동일 임계값을 적용하여 산출한 ECMWF-S2S 원시 앙상블 확률, 즉 훈련 전 ECMWF-S2S와 ERA5와의 앙상블 확률값을 사용하였으며, 원시 ECMWF-S2S 확률 대비 딥러닝 기반 보정 확률의 성능 향상을 정량적으로 평가하였다.

2.3.2. 격자별 주 단위 폭우 강도 및 빈도 예측을 위한 multi-task learning 기반 모델 개발

2.3.2.1. Single-task learning과 Multi-task learning

본 연구에서는 폭우 예측을 위해서 멀티태스크 러닝(MTL, multi-task learning) 방법을 도입하였다. 일반적인 딥러닝 개발 방법은 한 과제(task)에 대해서 하나의 입력 자료를 하나의 모델이 학습한다. 또 다른 과제에 대해서도 이 과제에 대한 개별적인 모델의 개발된다. 이를 single-task learning이라고 한다 (Figure 2.7). 반면 MTL은 여러 과제들을 동시에 수행하는 방법이다(Ruder, 2017). Figure 2.8에서 보이는 것처럼 목적 변수가 하나 이상인 다중 과제를 모델이 학습하는데 이때 모델이 각각 따로 학습하거나, 개별적인 모델 간의 약한 파라미터 공유(soft parameter sharing)를 삽입하거나, 하나의 모델로 강한 파라미터 공유(hard parameter sharing) 방식으로 구성하거나, 두 파라미터 공유를 혼합하는 형태가 가능하다. 이 외에도 입력 데이터를 각 과제별로 따로 활용할 수도 있고 MTL은 다양한 형태로 변형되어 개발되어져 오고 있다. 본 연구에서는 가장 흔히 활용되는 hard parameter sharing(Eq. [2.5]) 을 하는 단일

모델과 멀티 헤드(multi-head)를 가지는 MTL 구조를 활용하였다.

$$\min_{\theta} \sum_{i=1}^T L_i(\{\theta_{sh}, \theta_i\}, D) \quad [2.5]$$

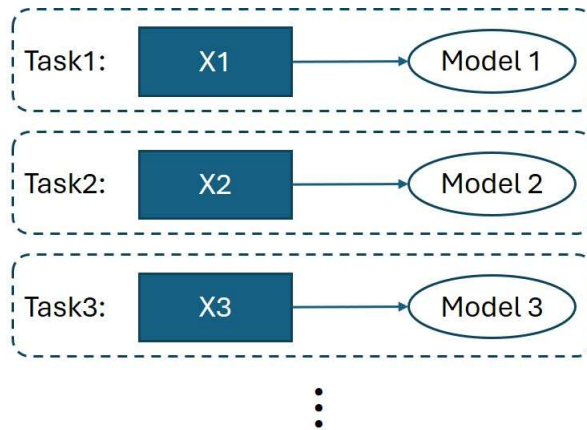


Figure 2.7. Diagram of single-task learning.

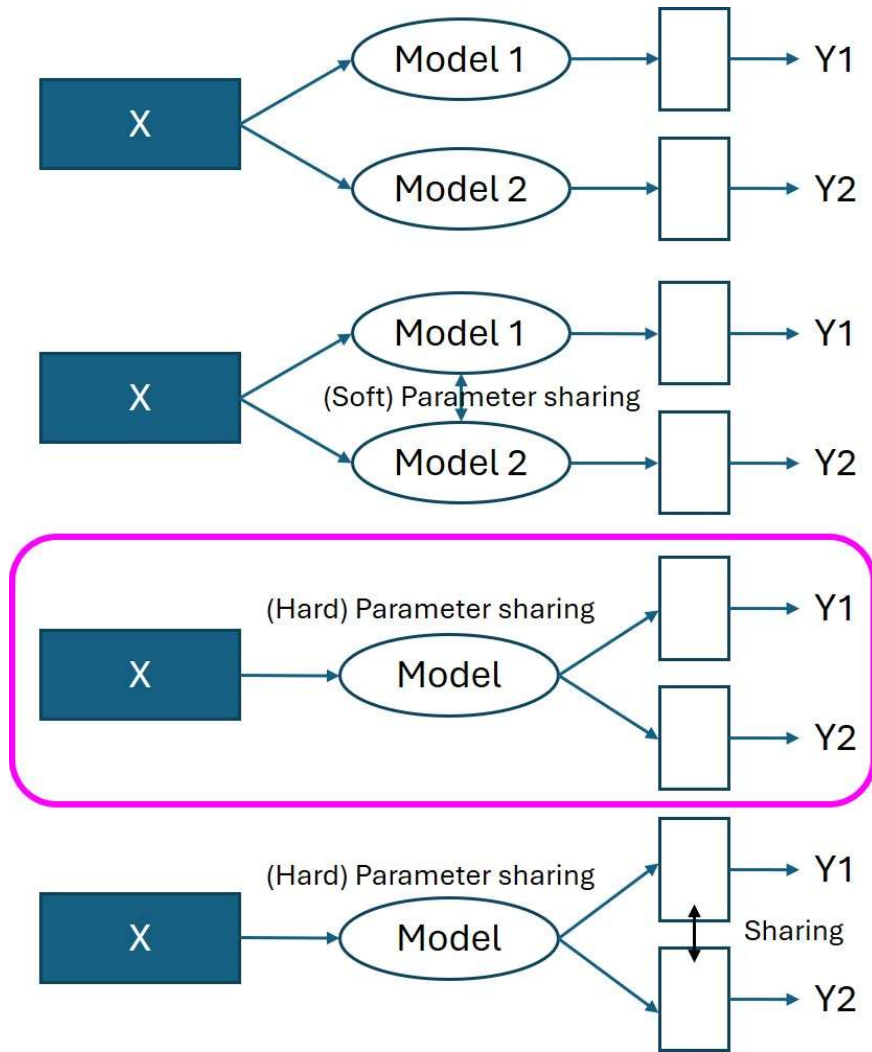


Figure 2.8. Diagram of multi-task learning. The structure in the colored box is the one chosen for this study.

2.3.2.2. 지점별 폭우 강도/빈도 multi-task learning 기반 최종 모델 구조

본 연구에서 활용한 모델 구조의 전체적인 흐름도는 Figure 2.9과 같다. ResNet-18과 양방향 LSTM(BiLSTM, Bidirectional Long Short-Term Memory) 기반의 멀티태스크 학습 모델로, 하나는 주간 누적 강수량 회귀(regression)와 다른 하나는 4개 카테고리의 극한 강수일수 분류(classification)을 동시에 예측하도록 설계되었다. 다양한 대기 및 해양 변수들을 입력 자료로 넣으면 공간적인 특징을 추출하는 ResNet-18 모델이 조금 변형되어 학습하고, 이어서 BiLSTM 구조를 붙여 시간적인 특징을 추출하게 하였다. 그 뒤 2가지 목적 변수인 주별 강수량과 주별 극한 강수일수를 예측하는 멀티 헤드 레이어를 연결해서 두 가지 다른 목적 변수를 동시에 학습하도록 하였다. 좀 더 상세한 구조는 Figure 2.10와 같다.

ResNet-18은 18개 계층으로 구성된 합성곱 신경망으로, 잔차 연결(residual connection)을 도입해 깊은 네트워크 학습 시 발생하는 기울기 소실과 성능 저하 문제를 완화한 모델이다(He et al., 2016). 기본적으로 여러 개의 잔차 블록을 통해 특성을 단계적으로 추출하며, 최종적으로

로 전결합층을 거쳐 이미지 분류 등 시각 인식 작업에 활용된다. ImageNet과 같은 대규모 데이터셋에서 널리 사용되며 비교적 가벼운 구조라 전이 학습과 Backbone 모델로 자주 선택된다. Figure 2.11은 ResNet의 18개 계층 이외의 34, 50, 101, 152개의 계층으로 구성된 모델 구조를 보여준다. Figure 2.12는 계층 사이의 잔차 연결을 보여주고 있다. 본 연구에서는 기존의 ResNet-18 구조에서 마지막 ResNet Block 3과 4를 삭제하는 것이 여러 모델 구조 실험에서 밝혀져 최종 모델은 마지막 두 블록을 삭제한 구조로 결정되었다. 이어서 ResNet 모델에서 추출된 7일 연속의 공간적인 특성들은 BiLSTM 모델에서 시간적인 정보를 추출한다.

LSTM(Long Short-Term Memory)은 RNN(Recurrent Neural Networks)의 한 종류로, 입력, 망각, 출력 게이트와 셀 상태를 이용해 긴 시퀀스에서도 과거 정보를 비교적 안정적으로 보존하며 시점 간 의존성을 학습하는 모델이다(Hochreiter and Schmidhuber, 1997). BiLSTM(Bidirectional LSTM)은 여기에 역방향 LSTM을 하나 더 추가해 시계열 정보를 과거에서 미래로 또 미래에서 과거로 모두 처리함으로써 양방향 문맥 정보를 동시에 활용할 수 있게 한 구조로 자연어 처리 및 음성 인식 등에서 성능 향상을 위해 제안되었다.

두 가지 목적 변수를 학습하는 MTL 학습을 위해서 멀티태스크 손실함수(Multi-task loss)를 아래와 같이 구성하였다(Eq. [2.6]).

$$Total\ Loss = w_{regression} \times Regression_{loss} + w_{classification} \times Classification_{loss} \quad [2.6]$$

모델 학습 시 조절이 필요한 다양한 하이퍼파라미터가 존재한다. Table 2.12는 본 연구에서 테스트한 하이퍼파라미터 목록을 보여준다.

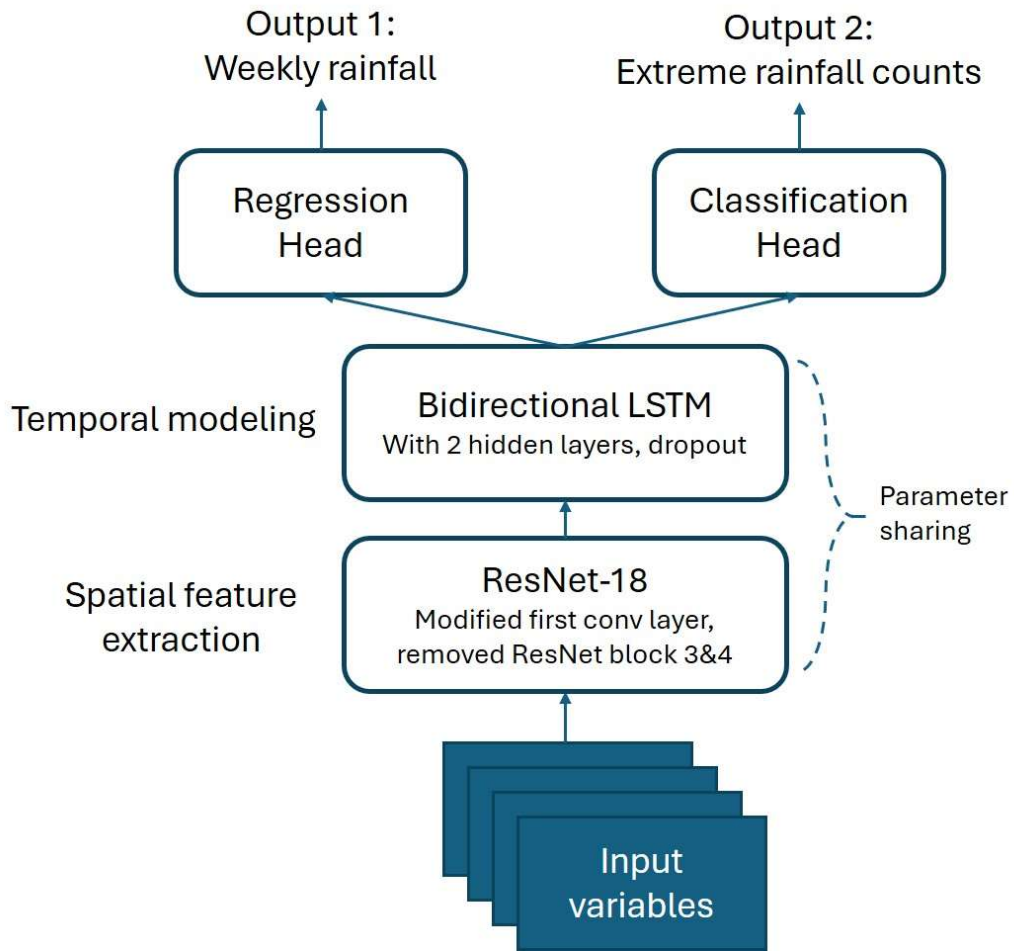


Figure 2.9. Overview of the model architecture used in this study.

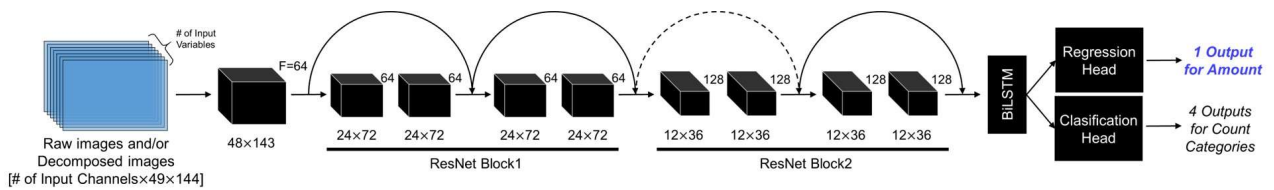


Figure 2.10. The model architecture used in this study with the details on the modified ResNet-18 model.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
		3×3 max pool, stride 2				
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Figure 2.11. List of architectures for residual neural networks with different layers.

34-layer residual

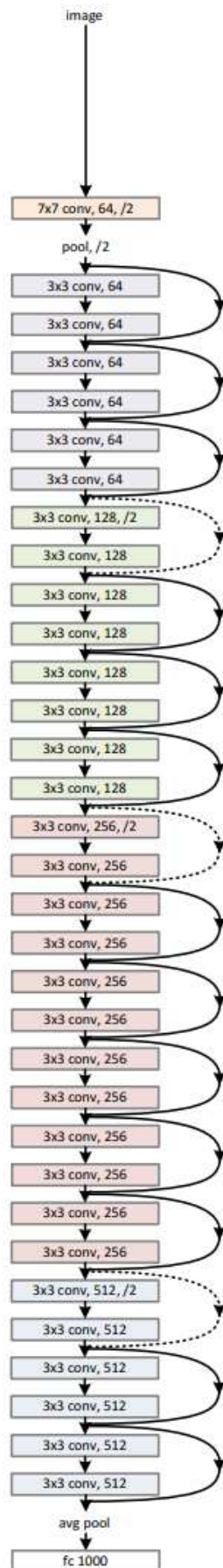


Figure 2.12. Residual network with 34 parameter layers (He et al., 2015)

Table 2.12. Hyperparameter settings.

Parameter	Set Value / Method
Dropout	0.1
Weight Initialization	Xavier Initialization
Optimization Algorithm	Adam
Initial Learning Rate	0.001
Learning Rate Scheduler	ReduceLRonPlateau (Patience=3)
Early Stopping	Patience = 5

2.3.3. 선행 3주 동아시아 해역 고수온 예측 모델 개발

2.3.3.1. 예측 목표 및 인공지능모델 기본 구조

본 연구는 기후 재난 대비를 위한 실용적 해양열파(MHW) 예측 모델 개발을 목표로, 타깃 시점을 선행 3주차 주별 예측으로 설정하였다(Table 2.13). 전통적인 MHW 이벤트의 경우 월/계절/연별 통계 추출은 용이하나, 빈도·지속기간 등을 주별 예측을 위한 AI 모델 목적 변수로 활용하는 데 한계가 있다. 이에 주별 SST 극한도(Impact Factor)를 새롭게 정의하여 격자별 라벨로 사용하였다.

Impact Factor 산출 과정은 다음과 같다: 1) 일별 SST에서 날짜 주변 1991-2020년 30년간 11일 윈도우(330 샘플)의 90th percentile을 31일 스무딩으로 기후 기준값(climatology) 도출; 2) 일별 SST에서 기준값을 차감하여 극한 여부 판별; 3) 7일 이동윈도우에서 극한값 합산으로 주별 누적 극한도 계산. 이 과정에서 daily moving 방식으로 자료를 7배 확장할 수 있었는데, 연속적인 날짜들의 경우 서로 독립적이지는 않지만 훈련 가능 샘플을 효과적으로 늘리는 방식이라 이러한 방식을 취함.

최종 Impact Factor(7일 누적 극한도)를 동아시아 연안 해역(EAMS: East Asia Marginal Sea) 격자별 라벨로 추출, 이를 예측하는 AI 모델을 설계하였다. EAMS는 global ocean shapefile 경계 기준 동해·황해·동중국해 격자만 선별하고 육지·기타 해양은 mask 처리하였다.

Table 2.13 Prediction Target: variables and temporal/spatial range of label

TARGET	DETAILS
Variable	Impact Factor (SST based)
Temporal Scale	Weekly (7 days expectation) with 3W lead time
Domain Coverage	EAMS (East Sea, Yellow Sea, East China Sea only): [24° -46.5° N, 118.5° -141° E], 1.5° res.

동아시아 연안 선행 3주차 고수온 예측을 위한 인공지능 모델은 아래 Figure 2.13와 같이 설계되었다. 본 모델은 최신 대기·해양 재분석 자료를 충분히 활용하여, 3주 이후 동아시아 연안의 고수온 발생 여부를 2차원 지도 형태로 예측하는 것을 목표로 한다. 이를 위해 최대 17개 변수의 최근 1주 또는 3주 관측 데이터를 선택·집적하여 채널별로 쌓아 lat×lon 2차원 훈련 입력을 구성한다. 이 입력과 3주 후 Impact Factor 2D 지도 라벨을 짝지어, 15,000개 이상의 샘플(365일×42년, 7일 이동 윈도우 극한 샘플)을 생성하고 이를 훈련 및 검증 세트로 분할하게 된다.

기본 모델은 UNet 구조를 기반으로 하며, 인코더 경로에서 2차원 신호를 압축(축소)한 뒤 디코더 경로에서 이를 2차원 고수온 패턴으로 확장하여 샘플에 적합시키는 방식으로 모수를 추정한다. 최종 레이어는 활성 함수를 사용하지 않는 fully connected layer로 구성된 회귀(regression) 모델로 설정하였다. 모델은 에폭별 손실 함수 값을 최소화하는 방식으로 최적화되며, 별도의 독립 검증 세트에서 손실을 확인함으로써 일반화 성능을 평가하고 과적합을 방지한다.

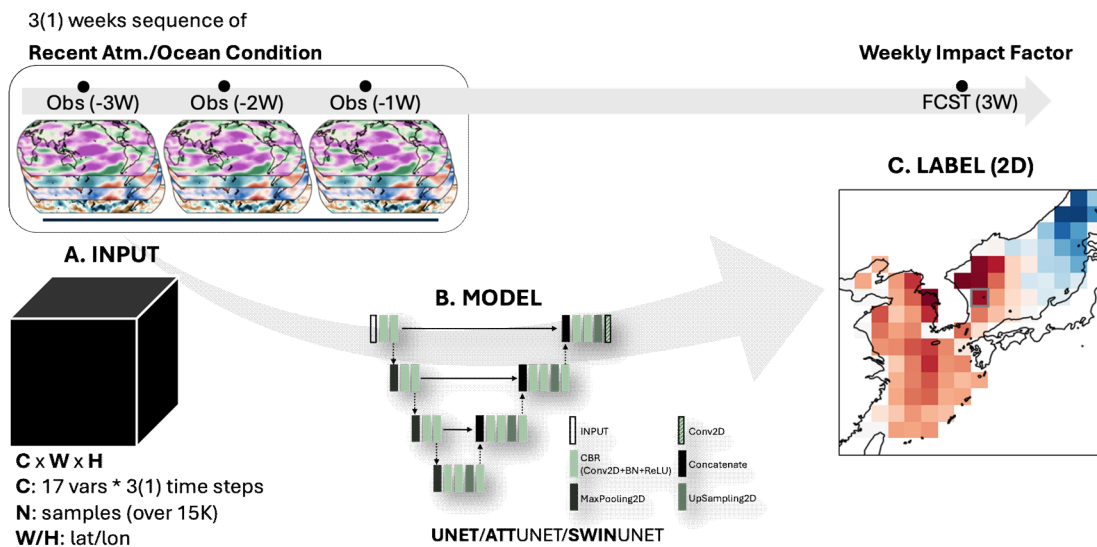


Figure 2.13 Diagram illustrating the training process, including input data stacking, model architecture, and the two-dimensional label distribution with EAMS.

옵티마이저로 Adam을 사용하였으며, 학습률은 기본값인 10⁻³으로 설정하였다. 검증 손실 값이 patient 설정값을 초과한 이후에도 감소하지 않을 경우, 학습률을 절반으로 줄이도록 스케줄러를 구성하였다(Table 2.14). 손실 함수는 널리 활용되는 평균제곱오차(MSE, mean square error)와 이상치 기준을 초과하는 경우에 가중치를 주는 방식의 가중평균제곱오차(WMSE, weighted mean square error)를 적용하였고, 수식은 아래와 같다(Eq. [2.7], [2.8]).

$$MSE = (1/n) \sum (y_i - \hat{y}_i)^2 \quad [2.7]$$

$$WMSE = (1/\sum w_i) \sum w_i (y_i - \hat{y}_i)^2 \quad [2.8]$$

여기서 n : 샘플 개수, y_i : 실제 값, \hat{y}_i : 예측 값, w_i : 각 샘플에 부여된 가중치를 의미한다.

훈련에 사용되는 변수들은 Table 2.9에서 확인할 수 있듯이, 여러 지구 시스템에 속하는 상이한 형태의 자료들로서 그 물리적 의미와 단위가 서로 다르다. 이러한 변수 간 극단적인 규모 차이를 보완하기 위해 적절한 형태의 표준화가 필요하다. 본 연구에서는 각 변수의 원자료에서 기후값을 제외한 편차를 산출한 뒤, 이를 기후값 기간 동안 경도 방향으로 평균된 표준편차로 나누는 방식을 적용하였다(Table 2.14).

테스트 자료에 대한 예측 결과 검증 지표로 ACC, RMSE, CSI를 활용하였으며(자세한 설명은 2.4.1절 참고), 이 중 격자별 이상치 탐지 성능을 측정하는 CSI를 주요 지표로 삼아 해양 열과 예측을 위한 최적 인공지능 모델을 도출하는 과정을 수행하였다.

Table 2.14 Key methods: learning rate, loss function setting, input data standardization and evaluation metrics applied.

METHODS	DETAILS
Learning Rate	Basically 1×10^{-3} with a scheduler making it $\frac{1}{2}$ when no loss reduction
Loss Function	MSE, customized(positive-inflated) MSE
Standardizing	Normalized Anomalies (subtracting daily climatology and deviding it by longitudinally averaged standard deviation at the corresponding latitude)
Evaluation Metric	ACC (Ano. Corr. Coeff.), RMSE (Root Mean Squared Error), CSI (Critical Success Index)

2.3.3.2. 최적 인공지능모델 도출을 위한 민감도 실험

본 연구에서는 고수온 예측을 위한 최적 모델 설정을 도출하기 위해 입력장, 모델 구조, 손실 함수 등 다양한 실험 조건을 설정하고, 모든 교차 조건에 대해 학습 모델을 훈련한 뒤 그 결과를 종합적으로 비교·분석하는 방식을 적용하였다. Figure 2.14에 제시된 총 7개의 민감도 실험 중 4개는 입력장 자체에 관한 것으로,

- 1) 훈련장 도메인의 크기와 해상도를 어떻게 조정할 것인지,
- 2) 입력 변수의 조합을 어떤 방식으로 구성할 것인지,
- 3) 훈련 자료를 과거 여러 시점에서 활용할 것인지, 혹은 직전 시점 자료만 사용할 것인지,
- 4) 2차원 재분석 격자 자료 외에 이질 자료를 추가할 경우의 영향은 어떠한지에 대한 질문에 답하고자 하였다.

나머지 3개의 민감도 실험은 입력 자료 외적 요소에 대한 것으로,

5) 제한된 샘플에서 훈련·검증 세트의 비율을 조정하여, 무작위 분리 방식에도 불구하고 고수온 현상이 검증 세트에 포함될 가능성을 높이는 방식을 시험하였다.

6) 또한 기본 UNet에 attention 및 swin-transformer 블록을 추가한 모델 구조 변형과

7) MSE 손실 함수에 가중치를 부여했을 때 이상치 예측 성능 개선에 실제로 기여할 수 있는지를 확인하였다.

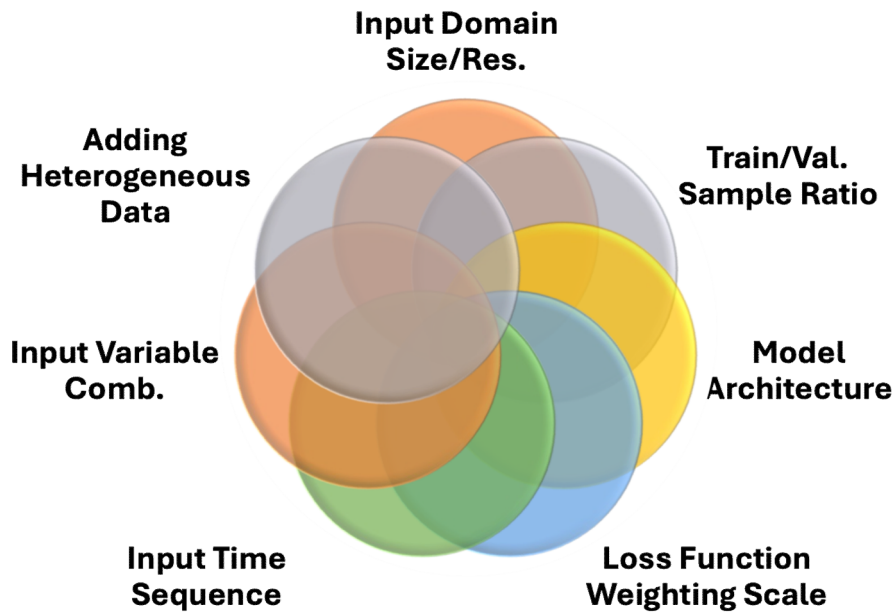


Figure 2.14 Seven ablation tests to isolate the optimal AI model setup for EAMS MHW forecast.

앞서 제시한 민감도 실험 종류와 구체적인 설명은 아래와 같다.

- i. 훈련 도메인 크기/해상도 조정 실험: 훈련에 사용된 17개 변수 자료를 예측 대상인 동아시아 해역을 기준으로 어디까지 확장해서 훈련에 활용될 공간 규모를 정하는 문제에 대한 답을 얻고자 아래의 3가지 조건에 대한 입력자료를 만들어 실험을 수행하였다. G5/L5의 경우는 128x128, 64x64 정사각형 격자를 맞추기 위해 위도 방향 부족한 격자에 대해 zero padding 처리후 훈련을 진행하게 된다.

G5: global domain [0~0~82.5W, 65S~65N] / 5deg resolution, 128x27
 L5: large domain [160W~157.5E, 65S~65N] / 5deg resolution, 64x27
 S2.5: small domain [70~147.5E, 20S~57.5N] /2.5deg resolution, 32by32

- ii. 개별 시스템 관련 변수 제외 실험: 3주 이후 해양열과 발생에 기여할 수 있는 성분들

을 꼽아보면 현재의 고수온 해양 상태 (Ocean), 파랑/파고 등 웨이브 에너지 (Wave), 대기에서 오는 하방 플럭스 성분들 (Flux), 대기 상층 시스템 (Upper atmosphere), 그리고 대륙/해양 구분 정보나 위도 등 정적 정보 (Static)이 있을 수 있다 3주 이후 해양열과 발생에 기여할 수 있는 주요 성분으로는 현재의 고수온 해양 상태(Ocean), 파랑 및 파고 등 웨이브 에너지(Wave), 대기에서 유입되는 하방 복사 플럭스(Flux), 대기 상층 시스템(Upper atmosphere), 그리고 대륙/해양 구분 정보 및 위도와 같은 정적 정보(Static)가 있다(Figure 2.15). 기본 해양 대기 변수 외에 추가 가능한 다양한 변수들을 에너지 출처별로 분류한 결과는 Table 2.9에 제시하였다. 각 에너지 성분이 해양열과 예측에 미치는 상대적 기여도를 평가하기 위해, 해당 성분에 속하는 변수들을 제거한 모델을 구성하고 이를 모든 성분을 포함한 모델과 비교하는 방식으로 실험을 수행하였다. 이때, 모든 성분을 포함한 실험을 all, 해양 성분 제외 실험을 woO, 파랑/파고 성분 제외 실험을 woW, 플럭스 성분 제외 실험을 woF, 대기 상층 성분 제외 실험을 woU, 정적 정보 제외 실험을 woS로 명명하였다. 각 실험의 성능이 all 실험 대비 저하된 정도는 해당 성분의 중요도를 나타내며, 반대로 성능이 향상된 경우 해당 성분이 훈련 과정에서 잉여 정보였을 가능성을 시사한다.

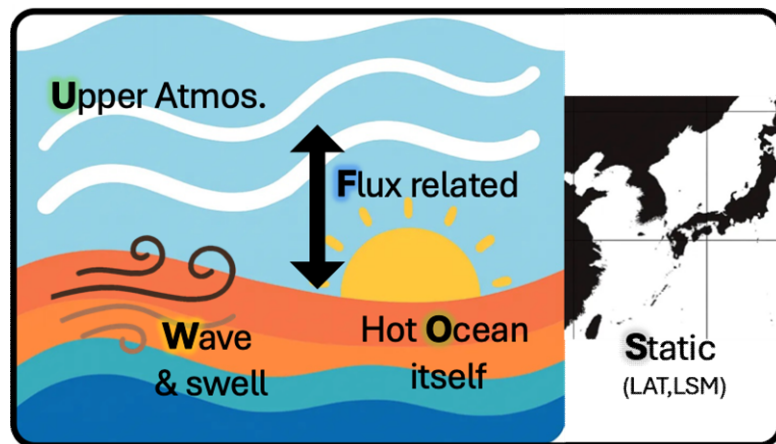


Figure 2.15 The five Earth System components (Flux, Ocean, Static, Wave, and Upper-Atmosphere) evaluated as potential input features for the MHW forecast model. Component variables are listed in Table 2.9.

- iii. 최근 3주/1주 훈련자료 비교 실험: 선행 3주 고수온 예측을 위해 가장 최근 1주 평균 관측 자료(T-1)를 기본 입력으로 사용하고, 과거 3주간의 주평균 자료(T-3, T-2, T-1)를 채널 단위로 적층하여 가용 정보를 확장하는 방식의 영향력을 비교하였다.
- iv. 기후지수와 하천유량 자료 추가 실험: 기본 통계 분석 결과, 주요 대기·해양 기후 지수들은 3주 이후 동아시아 해역에서 발생하는 해양 열파와 유의한 상관성을 보였다. 또한 기존 연구에 따르면, 양쯔강 하천유량 증가로 인한 담수 유입이 동중국해 해양

열파의 지속 및 강화에 영향을 미치는 것으로 보고된 바 있다 (Oh et al., 2023). 따라서, UNet 중간 층에 이질 자료 추가하는 아래 2가지 형태 실험을 주로 수행하였다.

- 동아시아 연안 고수온 현상과 유의한 상관성을 갖는 8개 기후 지수(Niño3.4, Niño4, Atlantic Niño, IOBM, IOD, Southern IOD, Western North Pacific Subtropical High, South Asian Monsoon, sea Table 2.15)를 별도의 Conv2D 연산으로 시그널을 추출한 뒤, 이를 concat하여 UNet encoder에 추가하는 방식
- 별도의 Conv2D 연산으로 추출한 하천유량 정보를 UNet decoder의 마지막 layer에 concat하는 방식

Table 2.15 Climate indices utilized as external inputs in UNet model

약어	Full Name (English)	정의 (Definition)	주요 참고문헌
Niño 3.4	Niño 3.4 Sea Surface Temperature Index	5° N-5° S, 170° W-120° W 지역 SST 편차값 평균	NOAA Climate Data Guide
Niño 4	Niño 4 Sea Surface Temperature Index	5° N-5° S, 150° W-160° E 지역 SST 편차값 평균 (중앙 태평양 El Niño)	NOAA PSL
Atlantic Niño	Atlantic Niño Index	대서양 적도 동부 SST 편차값 평균 (5° N-5° S, 20° W-0° E)	Zebiak (1993)
IOBM	Indian Ocean Basin-wide Mode	인도양 전역 SST 이상 (20° S-20° N, 40° -120° E)	Du & Evans (2010)
IOD	Indian Ocean Dipole	서인도양(10° N-10° S, 50° -70° E) vs 동인도양(10° S-0° , 90° -110° E)	Saji et al. (1999),
sIOD	Southern Indian Ocean Dipole	남인도양 IOD 변이 (10° S 이남 중심)	Izumo et al. (2010)
WNPSH	Western North Pacific Subtropical High	북서태평양 고기압 서쪽 가장자리 (1000-1500gpm contour)	Chen et al. (2020)
SAM	South Asian Monsoon Index	남아시아 몬순 (5° -22.5° N, 35° -97.5° E 또는 55° -100° E, 3° -15° N)	Wang et al. (2001); Li & Zeng (2002)

- v. 검증 샘플 비율 조정 실험: 90%tile 기준값을 초과하는 경우를 고수온으로 정의하였기 때문에, 훈련/검증 비율을 3:1로 나누었을 때 실제 고수온 사례가 검증 샘플에 충분히 분포하지 않을 가능성이 있다. 훈련/검증 샘플을 랜덤하게 분류하는 방식의 한계를 보완하기 위해, 검증 샘플의 크기를 2:1 비율로 확대하여 고수온 사례가 보다 적절히 포함되도록 하는 비교 실험을 수행하였다.
- vi. 동아시아 해역 2차원 고수온 패턴 예측을 위해 합성곱 기반의 UNet 모형을 기본 구조로 설계하였다. 이상치 탐지 성능을 향상시키기 위해, 기본 UNet의 합성곱 블록을 attention 또는 swin-transformer 블록으로 대체·추가한 변형 모델(ATTUNet, wwinUNet)을 구성하여 비교 실험을 수행하였다.

- vii. 손실 함수 가중치 조정 실험: 기본 손실 함수로 널리 사용되는 MSE를 적용하되, 이상치 탐지에 집중할 수 있도록 이상치에 해당하는 경우 손실 값이 커지도록 인위적 가중치를 부여하는 방식을 고안하였다. 특히 가중치 크기를 scale = 0, 2, 4, 6으로 단계적으로 증가시키며 민감도 실험을 수행하였다.

앞서 제시된 민감도 실험의 가능한 모든 교차 조건을 적용하여 모델을 훈련한 뒤, 훈련된 모델들을 모두 모아 같은 선상에서 종합 비교·분석함으로써 선행 3주 동아시아 고수온 현상 예측 및 탐지를 위한 최적의 모델 구조와 조건을 최종 도출하였다.

2.4. 이상기후 예측 성능 평가 방법

2.4.1. 주요 검증 메트릭

본 과제의 이상기후 예측 모델은 강도 회귀와 이상치 여부 탐지/분류를 동시에 수행하기 때문에 각각의 평가 메트릭이 다르다. 회귀(regression) 예측 평가 지표로는 R2(Coefficient of Determination), RMSE(Root Mean Square Error), ACC(Anomaly Correlation Coefficient) 등이 활용된다. 분류(classification) 및 탐지 평가는 기후학적 90th percentile 임계값을 적용하여 이상치 여부를 산출하고 이를 기반으로 성능을 평가하였다. 그 지표로는 CSI(Critical Success Index), POD(Probability of Detection), Accuracy 등이 활용되며, 자세한 설명은 아래와 같다.

R2(결정계수)는 회귀 모델이 종속변수의 변동을 얼마나 잘 설명하는지를 0과 1 사이의 값으로 나타내는 지표이다. 총 변동 중에서 모델이 설명하는 비율을 의미하며, 1에 가까울수록 데이터에 잘 맞는 모델, 0에 가까울수록 설명력이 낮은 모델로 해석한다. 일반적으로 선형 회귀의 적합도 평가나 회귀 기반 인공지능 모델의 성능을 평가할 때 자주 사용되지만, 값이 높다고 해서 항상 좋은 모델이라고 단정할 수는 없으므로 다른 지표와 함께 해석하는 것이 권장된다.

ACC (Anomaly Correlation Coefficient)는 예측된 이상 패턴과 실제 이상 패턴의 상관성 평가하기 위한 것으로 시·공간 패턴 일치도를 측정, 기후·기상 모델 검증에 적합하다. centered의 경우는 1에 가까울수록, uncentered의 경우는 양의 값의 클수록 예측과 관측 패턴이 잘 일치함을 의미한다(Eq. [2.9]).

$$ACC = \frac{\Sigma[(f_i - \bar{f}_i)(o_i - \bar{o}_i)]}{\sqrt{[\Sigma(f_i - \bar{f}_i)^2 \times \Sigma(o_i - \bar{o}_i)^2]}} \quad [2.9]$$

여기서 f_i : 예측값, \bar{f}_i : 예측 기후값, o_i : 관측값, \bar{o}_i : 관측 기후값을 의미한다.

RMSE (Root Mean Squared Error)는 예측값과 실제값의 평균적 차이 평가하는 대표적인 지표로서 데이터와 동일 단위, 직관적 해석 가능하다는 장점이 있다. 0 ~ ∞ 범위를 가지며, 값이 작을수록 평균 오차가 적다는 것을 의미한다(Eq. [2.10]).

$$RMSE = \sqrt{\frac{\sum (f_i - o_i)^2}{n}} \quad [2.10]$$

여기서 f_i : 예측값, o_i : 예측 기후값, n : 샘플 개수를 의미한다.

Accuracy는 전체 예측 정확도 평가하기 위한 직관적 지표로서 전체 대비 맞춘 비율을 보여주지만 불균형 데이터에서는 한계가 있어 이상치 탐지에는 취약하다는 특징이 있다. 0 ~ 1 범위를 가지며, 값이 클수록 전체적인 정확도가 높다는 것을 의미한다(Eq. [2.11]).

$$\frac{(TP + TN)}{(TP + TN + FP + FN)} \quad [2.11]$$

TP : True Positive, TN : True Negative, FP : False Positive, FN : False Negative

CSI (Critical Success Index, 성공지수)는 False Alarm과 Missed Event 모두 반영하는 이상치 탐지에 특화된 지표로서 해양 열파 예측의 주요 지표로 사용된다. 0 ~ 1 범위를 가지며, 값이 클수록 이상치 탐지율이 높다는 것을 의미한다(Eq. [2.12]).

$$\frac{TP}{(TP + FP + FN)} \quad [2.12]$$

TP : True Positive, FP : False Positive, FN : False Negative

분류의 문제에 있어 결정론이 아닌 확률론적 예측에서의 오차를 매저하는 방식으로 Brier Score(BS)와 Brier Skill Score(BSS)를 활용하였다. 이들은 이진 라벨링의 예측 정확도를 평가하기 위한 지표로서, BSS는 기후 평균 또는 기준 모델 대비 성능 향상을 정량화한다는 점에서 중요한 역할을 한다. 이러한 BS 및 BSS 개념은 원-핫 인코딩(one-hot encoding)을 적용하여 클래스별 오차를 합산하는 방식으로 다중 범주 예측에도 확장할 수 있으며, 본 연구에서는 다중 클래스 라벨 중 특히 이상고온을 의미하는 클래스 2에 대한 BSS에 초점을 두었다. BSS 값은 음수일 경우 기준 모델보다 성능이 저하되었음을 의미하고, 0이면 동등함을, 양수이면 성능 향상을 의미하며, 본 연구에서는 그 공간 분포뿐 아니라 바이올린 플롯(violin plot)과 차이 지도(difference map)를 함께 활용하여 ECMWF 대비 성능 개선의 정량적·시각적 진단을 수행하였다.

먼저, 예측 확률과 실제 관측값의 차이를 제곱 평균한 값을 계산하는 BS 계산이 필요하다(Eq. [2.13]). 그런데, 다중 클래스 라벨의 확률의 BS는 Eq. [2.14]과 같이 계산된다. 이렇게 계산된 관측의 BS와 예측 BS의 비를 1에서 빼서 BSS를 계산하였다(Wilks, D. S., 2011). BSS가 양수이면 모델 예측이 향상되었다는 의미이다(Eq. [2.15]).

$$BS = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2 \quad [2.13]$$

여기서 N 은 사례의 총 개수, i 는 각 사례를 나타내는 인덱스, f_i 는 i 번째 사례에 대한 예측 확률이며, o_i 는 i 번째 사례에 대한 관측값(사건 발생 시 1, 미발생 시 0)을 의미한다.

$$BS = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K (f_{i,k} - o_{i,k})^2 \quad [2.14]$$

여기서 N 은 사례의 총 개수, K 는 클래스(범부)의 총 개수, i 는 각 사례를 나타내는 인덱스, k 는 클래스 인덱스, $f_{i,k}$ 는 i 번째 사례에 대한 모델이 k 번째 클래스에 할당한 예측 확률이며, $o_{i,k}$ 는 i 번째 사례에 대한 k 번째 클래스의 관측값으로 관측 클래스일 경우 1, 나머지 클래스는 0인 one-hot 형태를 의미한다.

$$BSS = 1 - \frac{BS_{\text{model}}}{BS_{\text{ref}}} \quad [2.15]$$

여기서 BS_{ref} 는 관측 BS이며, BS_{model} 은 각각의 훈련 모델의 BS이다.

2.4.2. 인공지능 모델 성능 기초선

본 과제에서 개발된 인공지능 모델의 효용성/활용성은 기존 예측 기술 대비 성능 개선 여부에 크게 관계한다. 직접적으로 선행 3/4주 주별 이상기후 예측용 현업 모델은 존재하지 않으므로, 현재 계절내 예측에 있어 가장 우수한 성능을 보이는 ECMWF 앙상블 예측을 주 단위 결정론적/확률론적 이상치 예측으로 변환해 활용하였다. ECMWF 예측 성능을 평가하고 이를 기초선으로 이를 능가하는 방향으로 인공지능 모델 원형 모델과 최적 setup을 도출하고자 하였다.

2.5. 동아시아 이상기후 인벤토리 자료 공유 방식

2.5.1. GitHub 및 Git Repository 소개

인공지능 원형 모델 개발과정에서 얻어진 동아시아 이상기후 핵심 자료를 체계화하고 공유하기 위한 저장소를 설계하는 데 GitHub와 git을 활용하였다.

GitHub는 소프트웨어 개발자와 연구자들이 소스 코드를 온라인에서 저장, 관리하고 협업할 수 있는 웹 기반 플랫폼으로, Git의 분산 버전 관리 시스템을 기반으로 동작한다. GitHub는 코드 변경 이력 추적, 브랜치 기반 개발, Pull Request를 통한 코드 리뷰, 이슈 추적 등 개발 전 과정을 지원하며, Microsoft 산하 서비스로 전 세계 오픈소스 프로젝트와 기업 프로젝트의 중심 역할을 한다. 퍼블릭 저장소(public repository)는 누구나 접근 가능하며, 프라이빗 저장소(private repository)는 팀 내부 협업에 적합하다.

시작 방법은 GitHub 계정 생성 후 새 저장소(repository)를 만들고, 로컬에서 git init, git add, git commit 등으로 변경을 기록한 뒤 git push로 업로드하는 흐름이다. 협업 시 브랜치(branch)를 생성해 기능 개발 후, pull request 등으로 리뷰하고 메인 브랜치에 병합한다. 연구자나 데이터 분석가는 Jupyter 노트북, 스크립트, 데이터를 공유해 재현성을 높일 수 있으며, GitHub 웹페이지로 README나 결과를 정적 웹사이트로 호스팅 가능하다. 본 연구는 GitHub를 활용해 “EastAsiaClimateExtremes” 저장소를 통해 코드, 데이터, 결과를 공유한다.

2.5.2. 클라우드를 통한 대용량 자료 연결 및 colab 연동

“EastAsiaClimateExtremes” 저장소(GitHub)는 이상기후 사례의 통계적 분석과 시각화를 위한 기반 자료들과 코드들을 공유하기 위해 구축되었다. ERA5나 ECMWF 재분석 자료와 같은 대용량 원시 자료는 GitHub의 LSF(Large File Storage) 제한으로 직접 업로드해 활용하기 어렵기 때문에, 해당 자료들은 Dropbox 클라우드에 저장되어 있다. 사용자는 GitHub 레포지토리 내의 안내 문서를 참고하여 wget 명령어를 이용해 필요한 데이터를 Dropbox에서 직접 다운로드할 수 있다. 이를 통해 데이터 용량의 제약 없이 연구자가 동일한 자료를 활용할 수 있도록 하였다. 또한 분석 가시화 코드와 통계적 결과 저장 부분을 담은 Jupyter Notebook이 GitHub에 포함되어 있으며, 이를 사용자가 다운로드 받아 Google Colab 환경에서 바로 실행이 가능하다. Colab은 로컬 환경에 별도의 패키지 설치나 설정 없이 인터넷 브라우저에서 파이썬 코드를 실행하거나 GPU를 활용할 수 있는 등 호스팅(hosted, 구글 서버에 실행되는) Jupyter Notebook 서비스이다. 이는 연구자나 학생이 손쉽게 분석을 재현하거나 확장 연구를 수행하는데 용이하게 활용된다.

2.5.3. Repository 설계

새 Git의 저장소(repository)는 프로젝트 단위로 명확한 이름을 부여해야 하며, 최상위에 README.md 파일을 배치해 프로젝트 개요, 설치 방법, 사용법(Usage), 사사(기여), 가이드라인 등을 마크다운(markdown) 형식으로 작성한다. 마크다운은 간단한 기호로 제목, 목록, 링크 등을 표현하며, GitHub가 이를 렌더링해 가독성 높은 문서로 표시한다.

본 연구의 “EastAsiaClimateExtremes” 저장소는 CODES(분석 스크립트), DATA(기후 데이터셋), IMAGES(시각화 결과) 디렉토리로 구성되며(Table 2.16), 각 디렉토리에 별도 README.md를 두어 내용 설명과 사용 예시를 제공한다. “EastAsiaClimateExtremes”의 최상위 README.md에는 전체 구조도 및 자료 다운로드 등을 포함하고 테이블로 디렉토리 설명 등을 추가하였다.

Table 2.16 Directory Structure and Example Files in the repository “EastAsiaClimateExtremes“. The table provides an overview of the main directories used in the repository along with representative example files. It outlines folders for code scripts, datasets, and visualization outputs.

Directory	Description	Example files
CODES	Python/Jupyter Notebook analysis scripts	TART_visualization1_output.ipynb
DATA	pickle/NetCDF datasets	event.AHT.era5.1940-2024.EA1.5.pkl
OUTPUTS	NetCDF/Numpy datasets	AHT_extreme_stats_1991_2020_p90.nc
IMAGES	Visualization results	TP_extreme_maps_1991_2020_p90.png

3. 결과

3.1. 동아시아 선행 3/4주 이상고온 역학 예측의 오차 보정 모델 개발

3.1.1. 라벨 기법에 따른 이상고온 분류 평가

3.1.1.1. ERA5와 ECMWF-S2S의 라벨링 차이

ERA5와 ECMWF-S2S의 3주 평균 TMAX에 각각 이진(Binary) 및 다중 클래스(Multi-class) 라벨을 적용하여 비교한 결과, 두 자료 간 라벨 일치도에서 흥미로운 특징이 나타났다. 먼저, 이진 라벨의 경우 ERA5와 ECMWF 간 일치도는 전반적으로 높은 편이었으나, 90백분위수를 초과하는지 여부(라벨 1)의 판정이 양상불 간 불규칙적으로 나타났다. 특히 2023년 10월 10일 사례에서는 ECMWF에서 공간적으로 불균일한 극한 판정이 나타나 ERA5 판측과 뚜렷한 패턴 차이를 보였다(Figure 3.1 및 Figure 3.2).

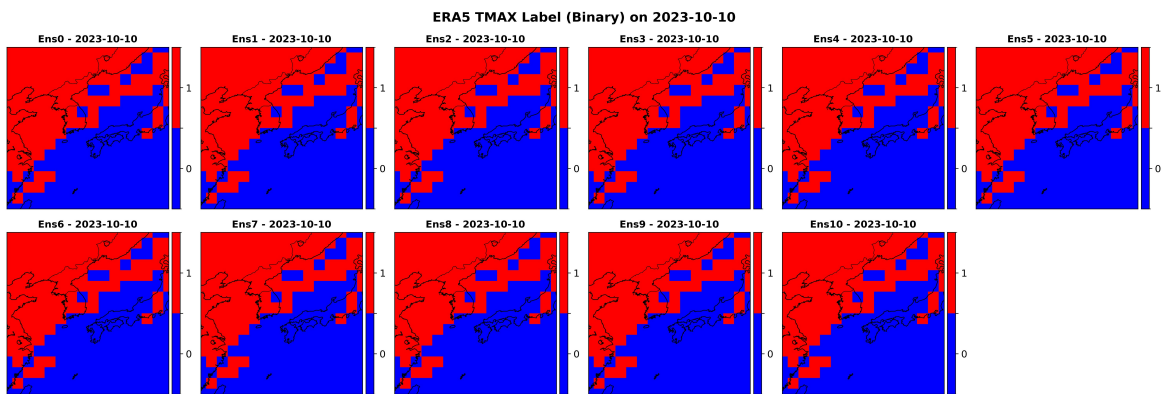


Figure 3.1 ERA5 3-week mean of daily maximum temperature binary labels for ensemble members on October 10, 2023.

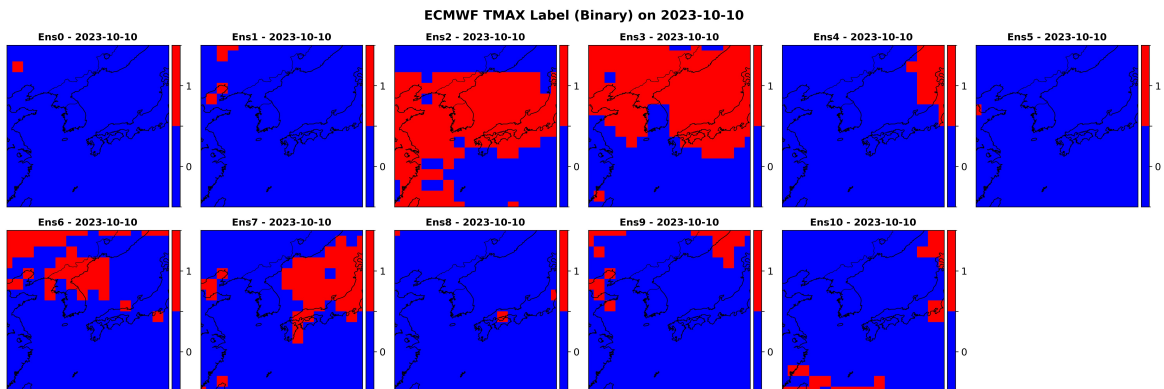


Figure 3.2 Spatial distribution of 3-week ECMWF mean daily maximum temperature binary labels across ensemble members.

반면 다중 클래스 라벨의 경우, 클래스 2(≥ 90 th)에 해당하는 비율이 전체 격자의 약 10% 내외로 극히 적은 비중의 희귀 사건 형태를 유지하였으며, 클래스 1(75~90th) 영역은 더욱 넓게 분포하였다(Figure 3.3 및 Figure 3.4). 이러한 분포 특성은 ECMWF 예측의 편차에 의해 ERA5 대비 공간적 차이를 더욱 확대시키는 요인으로 작용하였다. 결과적으로 다중 클래스 라벨을 적용할 때 클래스 경계의 미세한 차이가 예측 불확실성을 많이 증가시키는 구조적 한계가 존재함을 확인할 수 있었다.

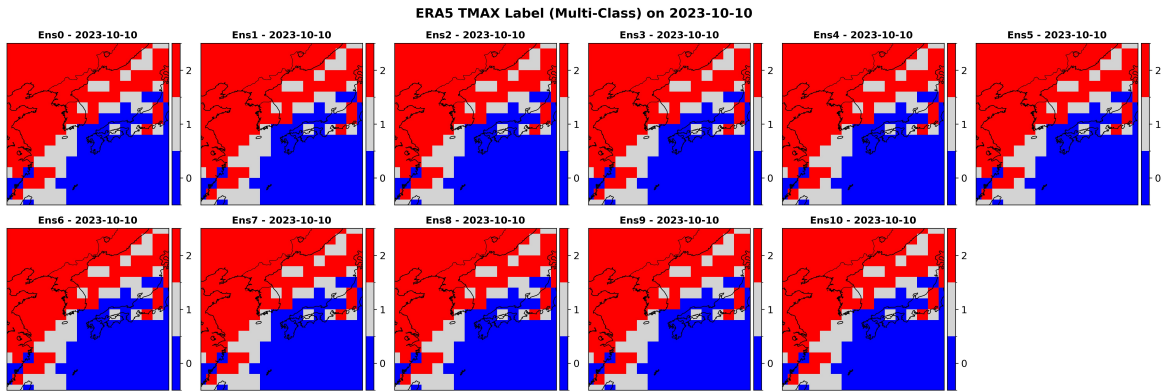


Figure 3.3 Multi-Class ERA5 3-week mean of daily maximum temperature label ensemble distribution on October 10, 2023.

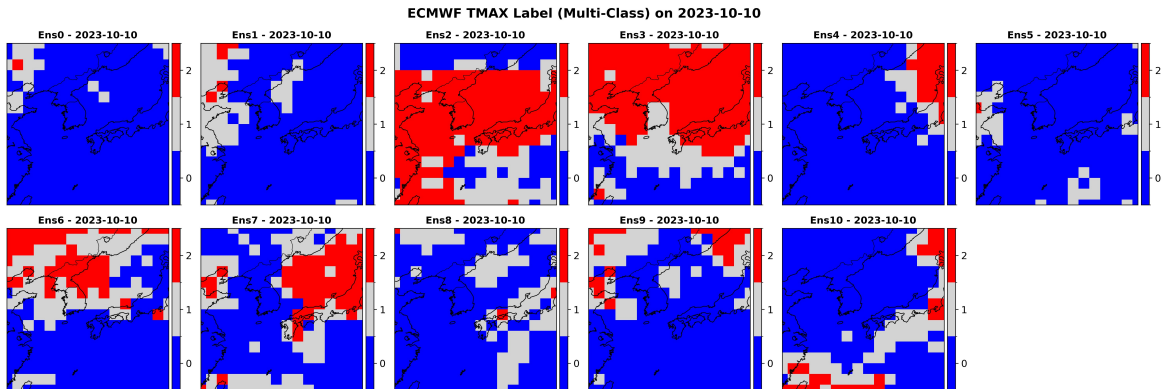


Figure 3.4 This figure shows the multi-class classification results of ECMWF ensemble members for the 3-week mean daily maximum temperature on 10 October 2023.

ERA5와 ECMWF-S2S의 4주 평균 TMAX에 각각 이진(Binary) 및 다중 클래스(Multi-class) 라벨을 적용하여 비교한 결과, 3주와 유사한 결과로 나타났다(Figure 3.5 및 Figure 3.6). 3주와 마찬가지로 이진 라벨의 경우 ERA5와 ECMWF 간 일치도는 전반적으로 높은 편이었으나, 90백

분위수를 초과하는지에 대한 판단은 더욱 양상불 간 불규칙적으로 나타났다. 4주 평균 TMAX의 다중 클래스 라벨의 경우도 3주와 유사한 결과를 보였는데, 클래스 2(≥ 90 th)에 해당하는 비율이 적게 나타났으며 클래스 1(75~90th) 영역은 더욱 넓게 분포하였다.

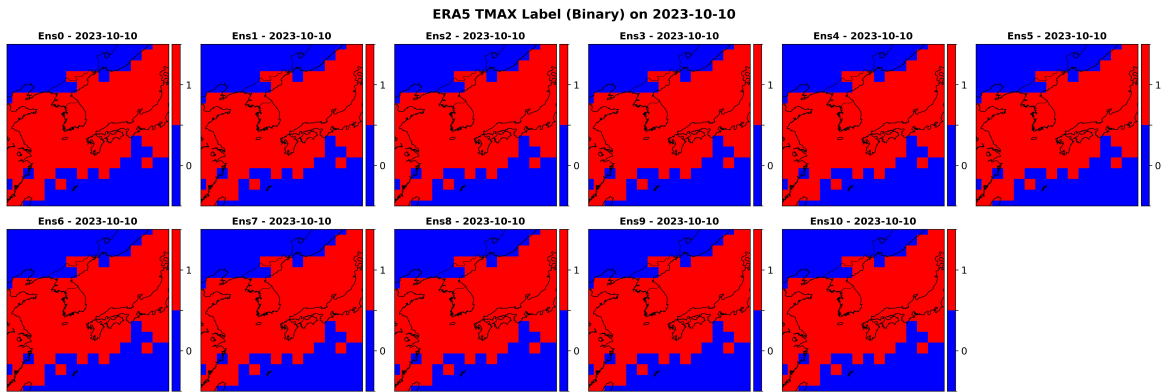


Figure 3.5 ERA5 4-week mean of daily maximum temperature binary labels for ensemble members on October 10, 2023.

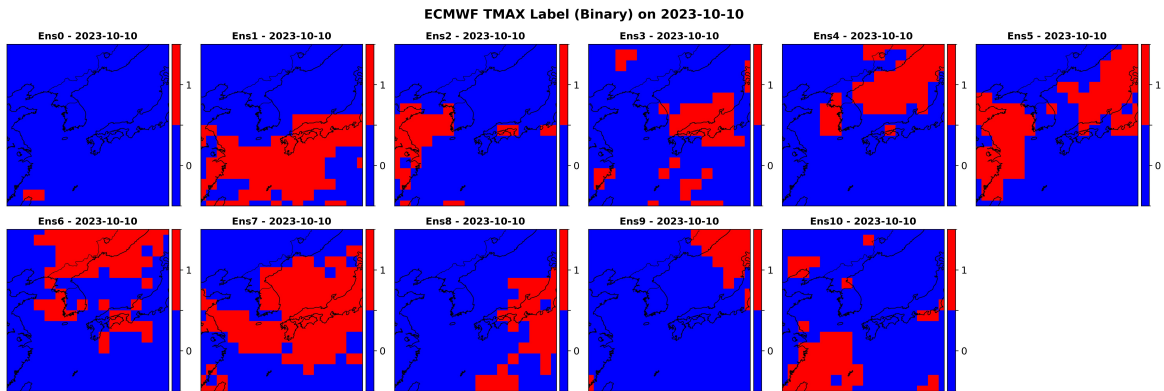


Figure 3.6 Spatial distribution of 4-week ECMWF mean daily maximum temperature binary labels across ensemble members.

이러한 분포 특성은 ECMWF 예측의 편차에 의해 ERA5 대비 공간적 차이를 더욱 확대시키는 요인으로 작용한 것이 3주 결과와 유사하겠다. 마찬가지로 다중 클래스 라벨을 적용할 때 클래스 경계의 미세한 차이가 4주에서도 예측 불확실성을 많이 증가시키는 구조적 한계가 존재함을 확인할 수 있었다(Figure 3.7 및 Figure 3.8).

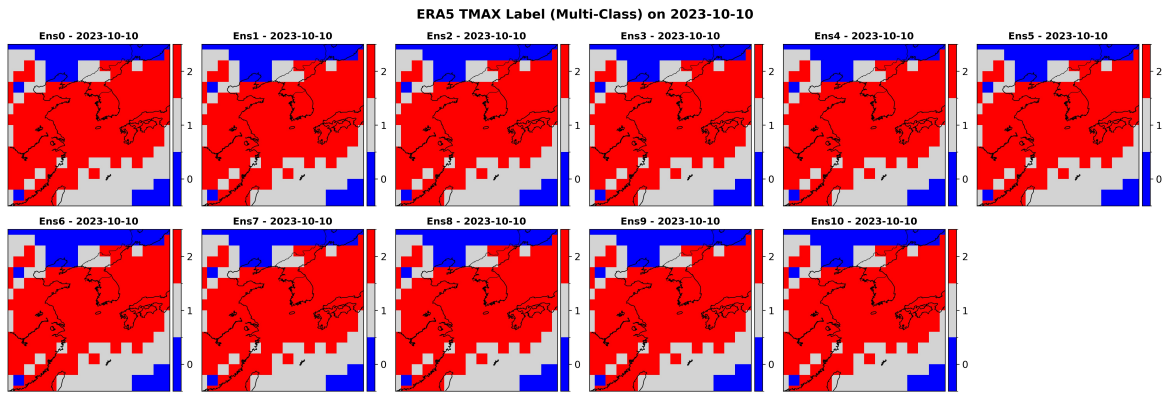


Figure 3.7 Multi-Class ERA5 4-week mean of daily maximum temperature label ensemble distribution on October 10, 2023.

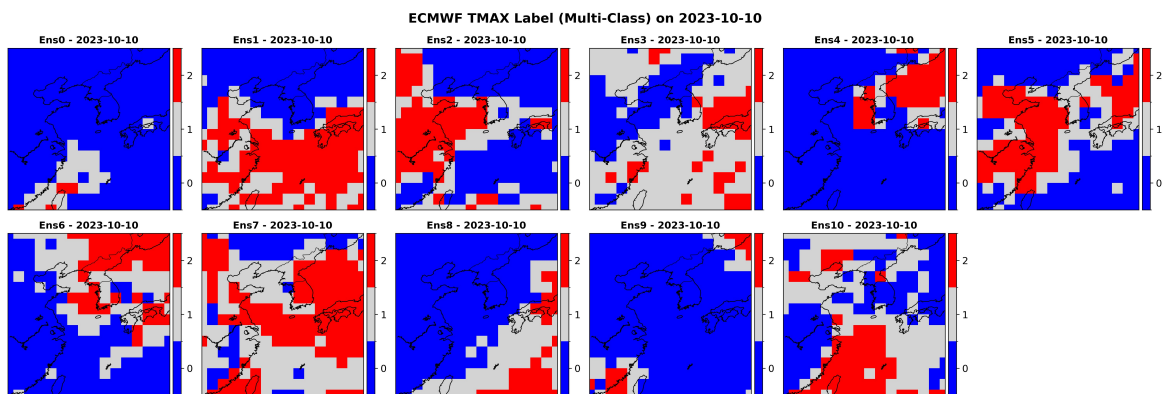


Figure 3.8 Spatial distribution of ECMWF 4-week mean daily maximum temperature multi-class labels on October 10, 2023.

3.1.2. 오차 보정 모델 적용

3.1.2.1. 라벨 및 모델간 조합에 따른 확률 예측 분석

2023년 10월 10일 사례는 2022년과 2023년 10월에 한반도에서 이례적으로 높은 일 최고기온(TMAX)이 발생했던 시기의 대표적인 사례로, 이상고온 탐지 성능 비교에 적합한 시험 케이스로 선정되었다. 이 사례를 이용해 3주 평균 TMAX 라벨 기반의 훈련을 통해 총 11개의 앙상블 모델 출력을 비교한 결과, U-Net의 확률 예측은 전반적으로 Att-U-Net보다 강도 측면에서 더 높게 나타나는 경향을 보였다(Figure 3.9 및 Figure 3.10). 반면 Att-U-Net은 관측된 이상고온의 공간적 분포를 보다 유사하게 재현하였는데, 이는 Attention 메커니즘이 공간적 특징을 선택적으로 반영하는 능력에 기인한 것으로 해석된다.

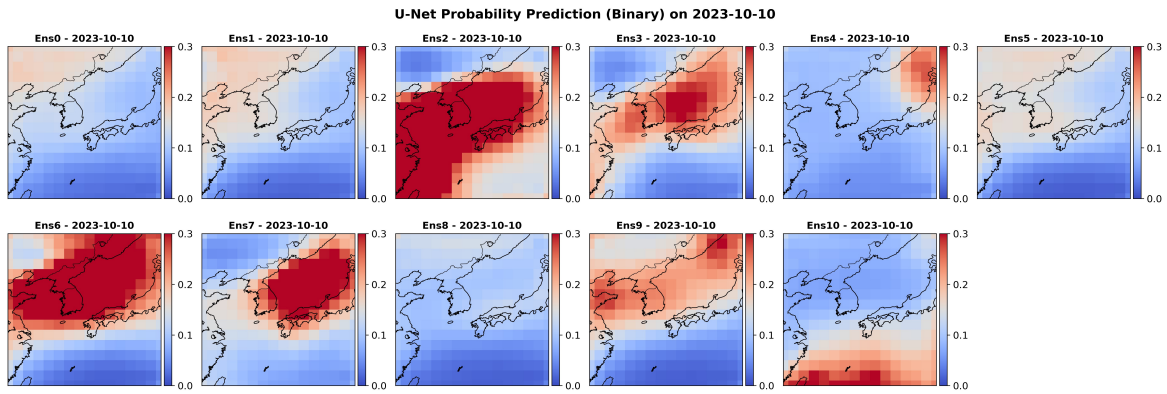


Figure 3.9 Ensemble U-Net binary probability prediction maps for October 10, 2023. The figure illustrates 11 ensemble probability distributions predicted from a U-Net model trained with a 3-week mean daily maximum-temperature label.

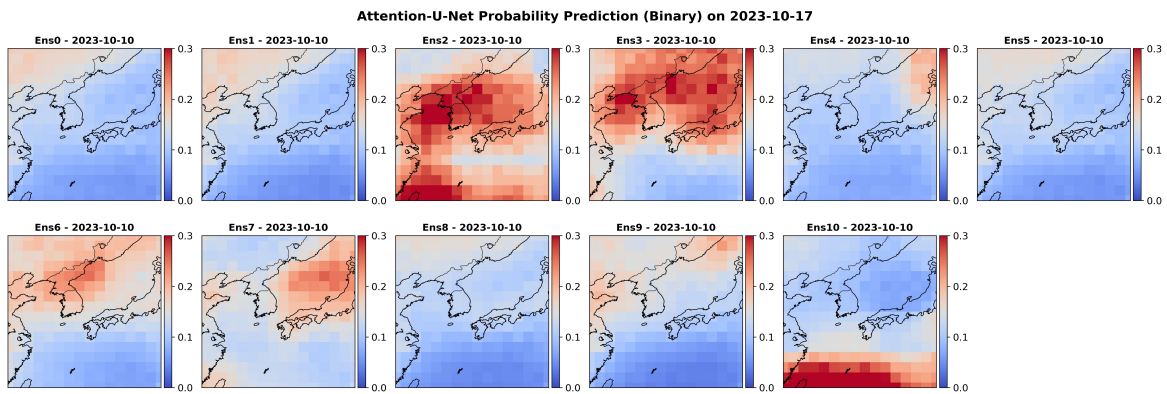


Figure 3.10 Ensemble Attention U-Net binary probability prediction maps for October 10, 2023. The figure illustrates 11 ensemble probability distributions predicted from a Attention U-Net model trained with a 3-week mean daily maximum-temperature label.

Figure 3.11과 Figure 3.12은 4주 평균 TMAX 라벨 기반의 훈련 모델을 통해 출력된 확률 예측 결과이다. 3주 결과와 다소 다르게 U-Net의 확률 예측은 전반적으로 Att-U-Net보다 강도 측면에서 더 높게 나타나는 경향을 보였다. 반면 Att-U-Net은 관측된 이상고온의 공간적 분포를 보다 유사하게 재현하였다.

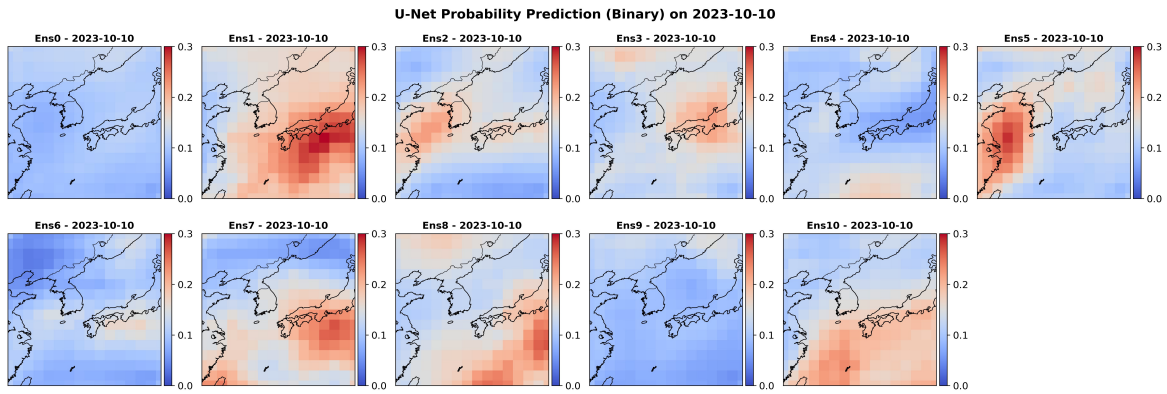


Figure 3.11 Ensemble U-Net binary probability prediction maps for October 10, 2023. The figure illustrates 11 ensemble probability distributions predicted from a U-Net model trained with a 4-week mean daily maximum-temperature label.

2023년 10월 10일은 테스트 기간(2022~2023) 중 대표적인 고온 사례로 선정된 날짜로, U-Net과 Attention U-Net 모델의 11개 앙상블 확률 예측 결과를 비교하였다. 먼저 이진 라벨 기반 확률 예측을 살펴보면, U-Net은 전반적으로 확률 강도가 높게 나타나며 동일 지역에서 예측 확률이 과도하게 집중되는 현상이 확인되었다. ERA5에서 고온으로 판정된 지역과 유사한 영역도 일부 존재했으나, 확률 분포가 매끄럽지 않고 hotspot이 과도하게 형성되는 등 공간적 연속성이 부족한 모습을 보였다(Figure 3.13).

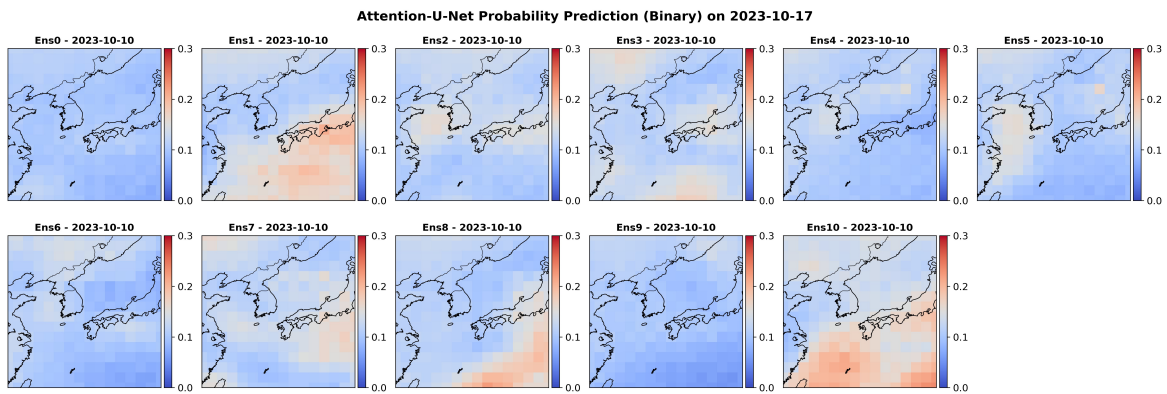


Figure 3.12 Ensemble Attention U-Net binary probability prediction maps for October 10, 2023. The figure displays binary probability predictions generated by a Attention U-Net model with a 4-week mean daily maximum-temperature for 11 ensemble members on October 10, 2023.

반면 Attention U-Net은 전반적인 확률 크기가 다소 낮게 산출되었지만, ERA5 관측 패턴과 공간 구조 면에서 더 높은 일치를 보였다(Figure 3.14). 이는 모델 내 Attention gate가 불필요한 주변 신호를 억제하여 시그널-잡음을 향상시킨 결과로 해석된다. 이러한 특성으로 인해 Binary 라벨 기반 Attention U-Net 조합이 ERA5가 제시한 극한 고온 지역을 가장 정확하고 구조적으로 일관되게 재현한 것으로 평가되었다.

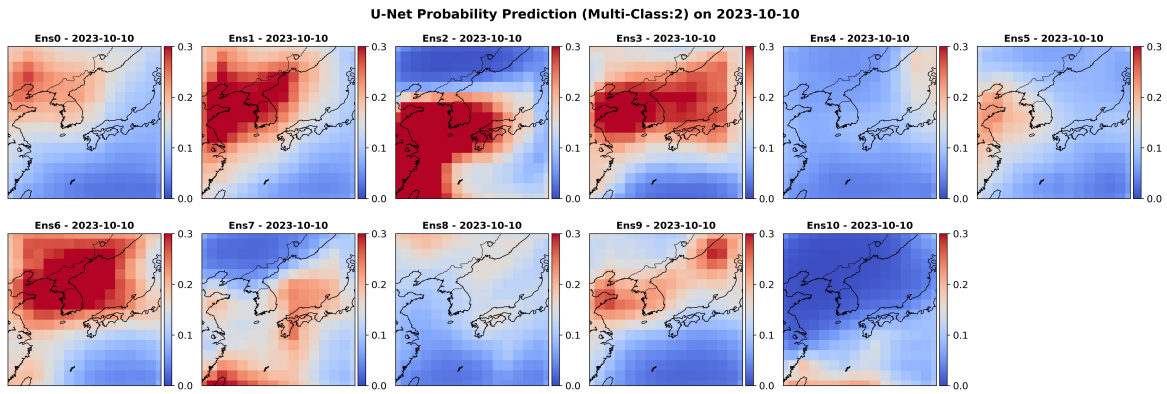


Figure 3.13 Ensemble based U-Net multi-class probability prediction maps on October 10, 2023. These maps show ensemble probability distributions predicted using labels based on the 3-week mean maximum temperature on October 10, 2023.

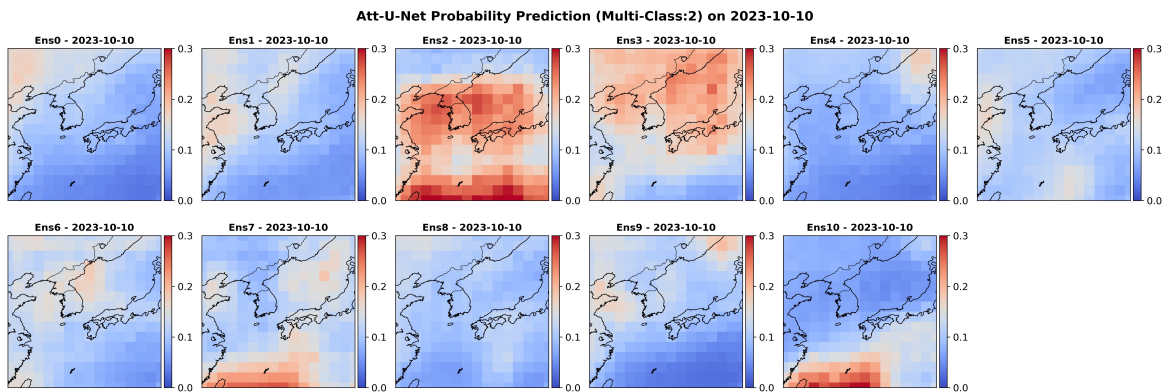


Figure 3.14 Ensemble based Attention U-Net probability prediction maps on October 10, 2023. This figure presents probability maps generated by a multi-class U-Net model with the 3-week mean maximum temperature for 11 ensemble members on October 10, 2023.

다중 클래스 기반 확률 예측에서는 클래스 2(90백분위수 이상) 구간의 데이터 비율이 극히 낮아, Softmax를 통해 산출된 확률값이 전반적으로 작게 나타났다(Figure 3.15 및 Figure 3.16). U-Net은 클래스 2의 확률을 과소평가하는 경향을 보였으며, Attention U-Net은 공간적 초점은 비교적 잘 포착했으나 전체 신호 세기가 약한 한계를 드러냈다. 이러한 결과는 다중 클래스 구조의 희소한 고온 라벨 분포로 인해 확률 예측의 안정성이 저하되고, 모델이 클래스 0과 클래스 1 중심으로 학습되는 특성에 기인한다. 그 결과 이상고온(클래스 2) 패턴의 재현성은 이진 분류 접근에 비해 현저히 떨어져, 다중 클래스 학습에서는 고온 극단 패턴 포착력이 상대적으로 약화됨을 확인할 수 있었다.

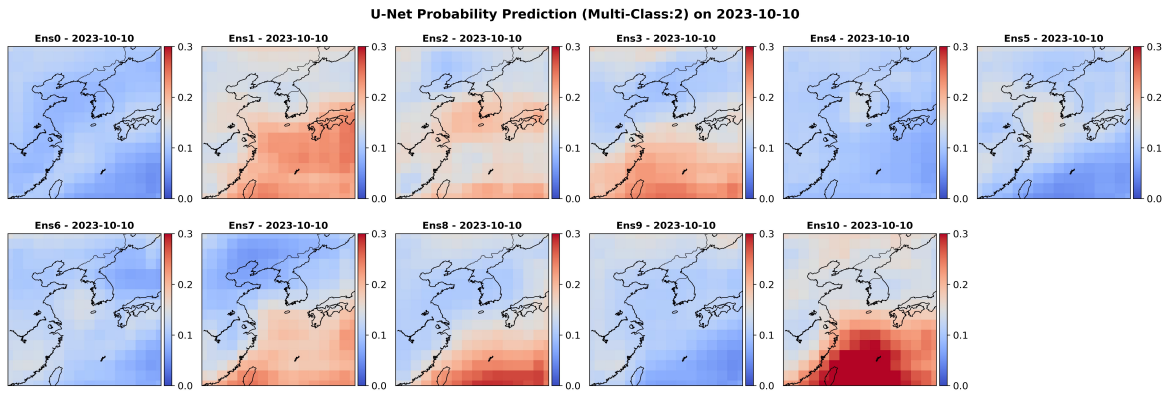


Figure 3.15 Ensemble based U-Net multi-class probability prediction maps on October 10, 2023. These maps show ensemble probability distributions predicted using labels based on the 4-week mean maximum temperature on October 10, 2023.

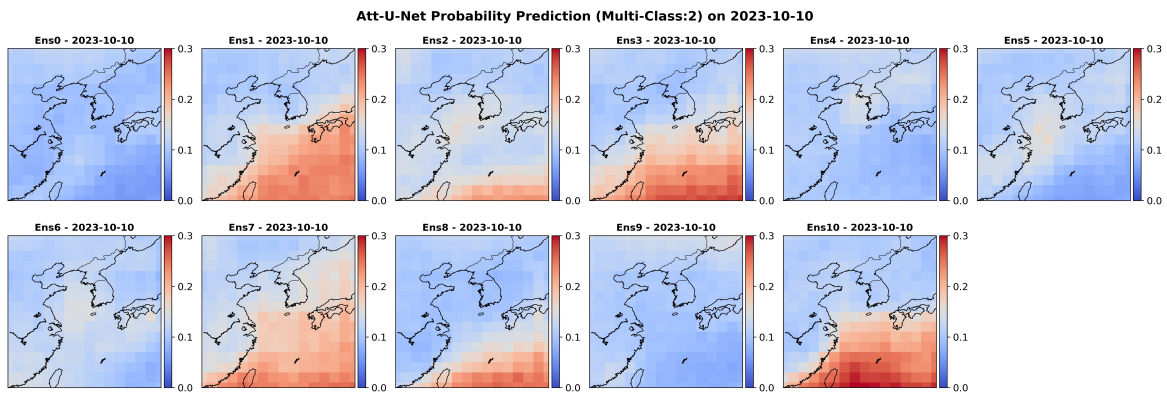


Figure 3.16 This figure presents probability maps generated by a multi-class U-Net model with the 4-week mean maximum temperature for 11 ensemble members on October 10, 2023.

앙상블 평균 확률 비교에서는 각각 3주 및 4주 평균 TMAX에 대해 초기일별 11개 앙상블 구성원의 확률을 평균하여 ERA5 라벨과 비교하였다(Figure 3.17부터 Figure 3.22). 이진 분류에서 U-Net은 확률 예측의 공간적 대비가 지나치게 크게 나타났으며, 특히 hotspot 주변에서 확률이 급격하게 변하는 패치(patchiness) 양상이 관찰되었다. 이에 반해 Attention U-Net은 공간적 매끄러움(spatial smoothing)이 자연스럽게 이루어졌고, ERA5의 1(클래스=1) 지역과 유사한 확률 분포를 보였다. 이러한 구조는 가장 현실적인 공간 일관성(space-consistent structure)을 제시하며, 이진 라벨 환경에서 Attention U-Net이 기후 후처리(post-processing) 모델로서 가장 적합함을 확인할 수 있었다(Figure 3.18 및 Figure 3.19).

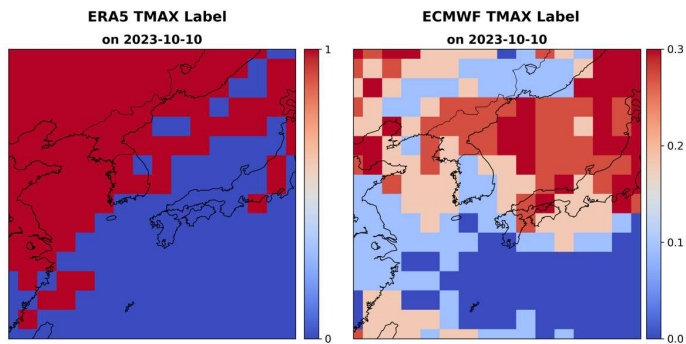


Figure 3.17 The left panel shows the binary label of 3-week mean daily maximum temperature derived from ERA5 for 10 October 2023. The right panel presents the average of eleven binary labels from the ECMWF-S2S ensemble for the same 3-week mean temperature period.

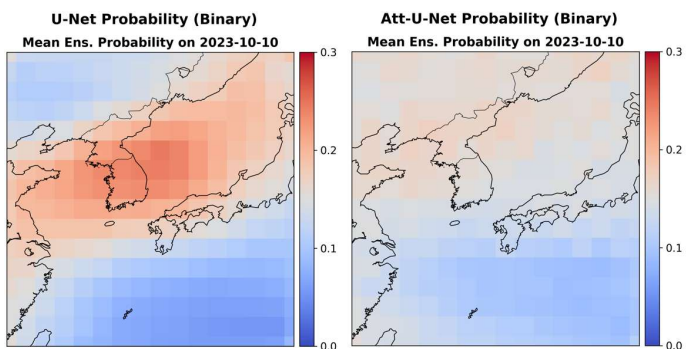


Figure 3.18 This figure shows the averaged probabilities from 11 ensemble predictions produced by deep learning models trained on binary labels of 3-week mean maximum temperature. The left panel displays results from the U-Net model, while the right panel shows corresponding outputs from the Attention U-Net model for 10 October 2023.

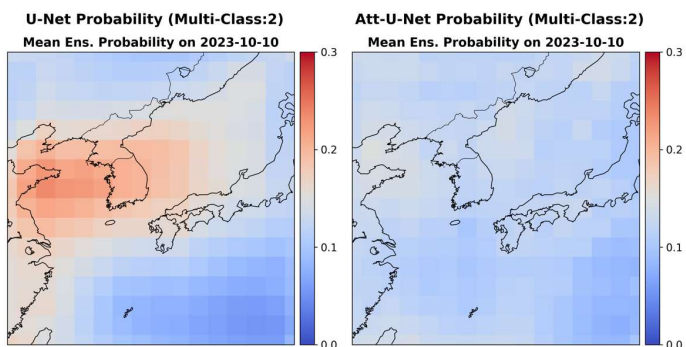


Figure 3.19 The maps show the mean probabilities obtained by averaging 11 predictions for the multi-class labels of 3-week mean maximum temperature on 10 October 2023. The left panel corresponds to the U-Net model, while the right panel shows the results from the Attention U-Net model.

다중 클래스 분류에서는 전반적으로 클래스 2의 평균 확률이 매우 희박하게 나타나, 격자 단위의 노이즈에 민감하게 반응하는 경향을 보였다(Figure 3.21 및 Figure 3.22). U-Net과 Attention U-Net 모두 클래스 0과 클래스 1 중심의 확률 구조를 형성하였으며, 클래스 2 탐지 능력은 제한적이었다. 이러한 결과는 다중 클래스 라벨 기법이 이상고온(anomalously high temperature) 탐지 성능을 저하시킬 수 있음을 실증적으로 보여주며, 향후 예측 체계 설계 시 Binary 접근의 실효성을 시사한다.

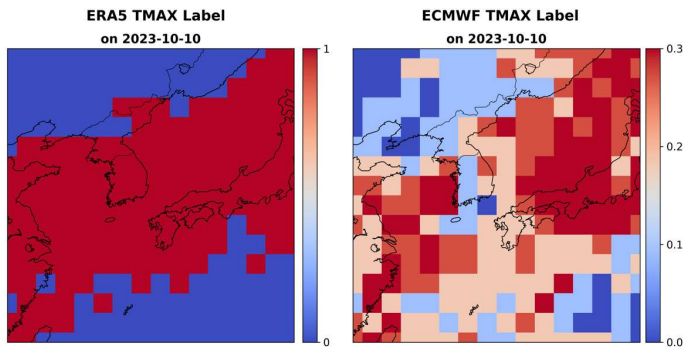


Figure 3.20 The left panel shows the binary label of 4-week mean daily maximum temperature derived from ERA5 for 10 October 2023. The right panel presents the average of eleven binary labels from the ECMWF-S2S ensemble for the same 4-week mean temperature period.

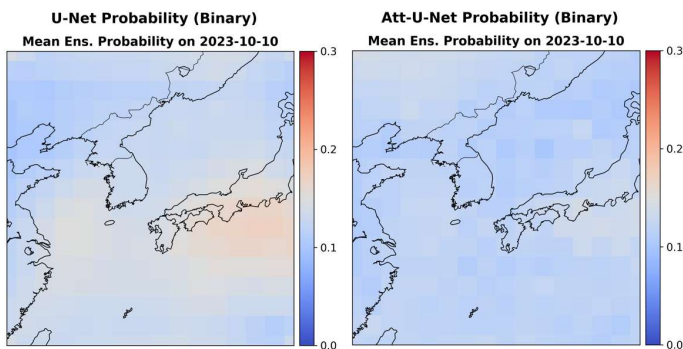


Figure 3.21 This figure shows the averaged probabilities from 11 ensemble predictions produced by deep learning models trained on binary labels of 4-week mean maximum temperature. The left panel displays results from the U-Net model, while the right panel shows corresponding outputs from the Attention U-Net model for 10 October 2023.

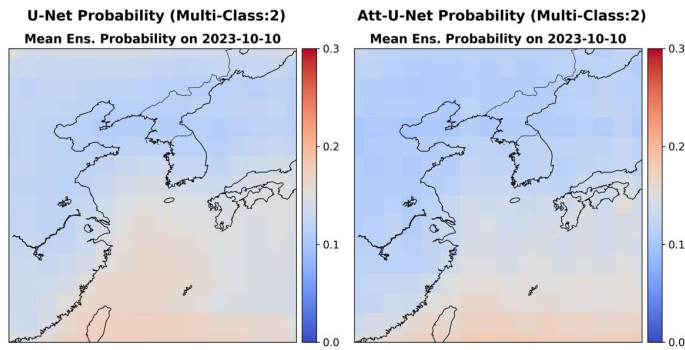


Figure 3.22 The maps show the mean probabilities obtained by averaging 11 predictions for the multi-class labels of 4-week mean maximum temperature on 10 October 2023. The left panel corresponds to the U-Net model, while the right panel shows the results from the Attention U-Net model.

3.1.2.2. 라벨 및 모델간 조합에 따른 성능 평가

2022-2023년 테스트 기간 전체에 대해 grid별 Brier Skill Score(BSS)를 산출하였다. 먼저 ECMWF 원시 예측 결과를 기준으로 삼았으며, ECMWF의 BSS는 전반적으로 약 0.2 내외의 값을 보였다(Figure 3.23 및 Figure 3.24). 이상고온은 본래 예측성이 낮은 이벤트로, S2S 예측 모델의 특성상 신호 대 잡음비가 낮아서 이러한 결과는 일반적인 수준이다. 따라서 ECMWF-S2S의 BSS는 딥러닝 후처리 모델의 기준(reference) 점수로 활용되었다.

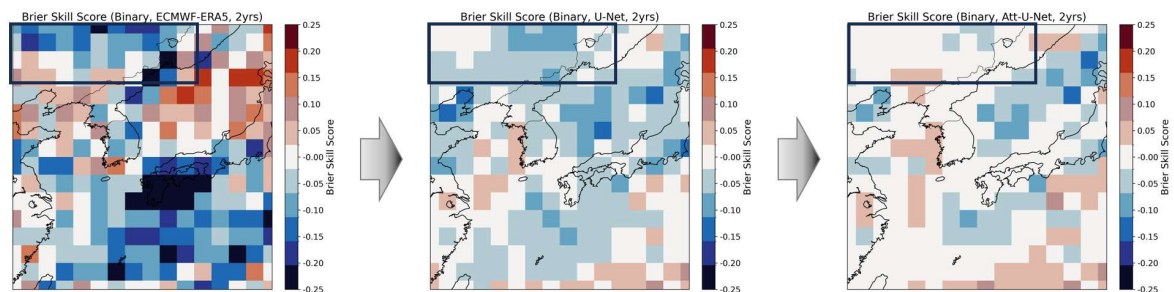


Figure 3.23 The figure compares brier skill scores (BSS) for binary 3-week mean maximum-temperature over 2022-2023, contrasting ECMWF-S2S (left) predictions with probabilities generated by U-Net (center) and Attention U-Net (right) models.

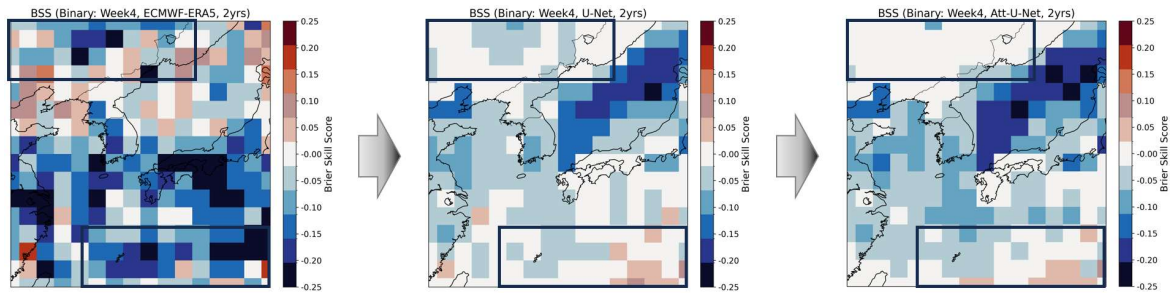


Figure 3.24 The figure compares brier skill scores (BSS) for binary 4-week mean maximum-temperature over 2022-2023, contrasting ECMWF-S2S (left) predictions with probabilities generated by U-Net (center) and Attention U-Net (right) models.

U-Net 기반의 딥러닝 예측에서는 BSS가 -0.1 에서 $+0.05$ 범위로 계산되었으며, ECMWF 대비 개선되는 격자의 비율은 한 자리수 수준으로 제한적이었다. 반면 Attention U-Net 모델은 공간적으로 더 넓은 영역에서 ECMWF 대비 양(+)의 BSS 개선을 보였다. 특히 한반도 남부, 중국 동부, 서해 상 지역에서 확률 예측의 견고성이 향상되는 경향이 뚜렷했다. 이러한 결과는 Attention 메커니즘이 기후 패턴 탐지 과정에서 노이즈를 줄이는 역할을 수행함을 실증적으로 보여준다.

다중 클래스 기반의 BSS 분석에서는 class 2에 대해서만 평가를 수행하였다. 전체적으로 BSS 값은 binary 모델에 비해 상대적으로 낮게 나타났으며, U-Net과 Attention U-Net 모두에서 ECMWF 대비 개선된 영역이 거의 관찰되지 않았다. 이는 극한 클래스의 희소성(class sparsity)으로 인해 BSS 계산 시 과도한 패널티(over-penalization)가 발생한 결과로 해석된다(Figure 3.25 및 Figure 3.26).

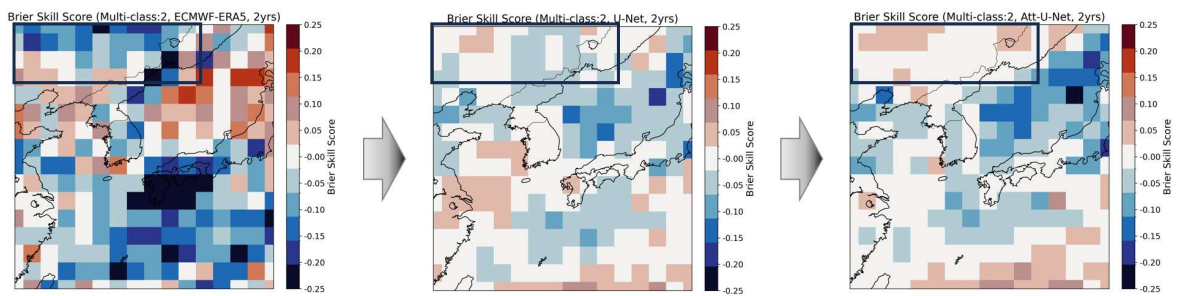


Figure 3.25 The figure compares the multi-class BSS derived from raw ECMWF-S2S 3-week mean maximum temperature with those from U-Net and Attention U-Net models trained on multi-class labels. The middle (U-Net) and right (Attention U-Net) panels show how the deep-learning models improve or modify regional skill relative to the ECMWF-S2S predictions (left).

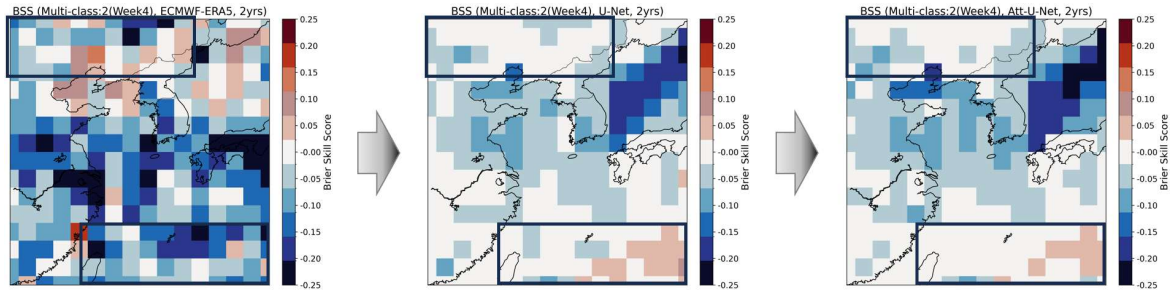


Figure 3.26 The figure compares the multi-class BSS derived from raw ECMWF-S2S 4-week mean maximum temperature with those from U-Net and Attention U-Net models trained on multi-class labels. The middle (U-Net) and right (Attention U-Net) panels show how the deep-learning models improve or modify regional skill relative to the ECMWF-S2S predictions (left).

BSS 분포(violin plot) 비교 결과를 살펴보면, PDF 분석과 동일한 세 가지 특징이 확인되었다(Figure 3.27 및 Figure 3.28). 첫째, ECMWF-S2S의 이진 라벨과 다중 클래스 라벨(클래스 2)의 BSS 분포는 거의 동일한 형태를 보였다. 둘째, 이진 라벨 기반 Attention U-Net의 violin 분포는 0 근처로 밀집되어 있어 예측의 중립성과 안정성이 향상되었음을 시사한다. 셋째, 다중 클래스 라벨 기반 모델은 후방 꼬리(long tail)가 길게 나타나 과적합 및 예측 불안정성이 증가했음을 보여준다.

U-Net과 Attention U-Net 모델을 비교한 결과, 두 모델은 예측 확률의 강도와 공간적 표현에서 뚜렷한 차이를 보였다. U-Net은 전반적으로 강한 확률 강도를 보이지만, 공간 패턴이 거칠고 패치 형태로 분절되는 경향이 나타났다. 반면 Attention U-Net은 확률 강도는 다소 완화되었으나, 공간적으로 매끄럽고 구조적 일관성이 높아 기상 패턴을 보다 안정적으로 표현했다. 또한 Brier Skill Score(BSS) 관점에서 Attention U-Net은 ECMWF-S2S 대비 개선 폭이 더 컸으며, 특히 극한 기온 영역에서의 초점이 명확하게 나타났다. U-Net이 과적합 위험이 높은 반면, Attention U-Net은 이를 효과적으로 억제하는 것으로 평가되어, 극한 기온 탐지에 더욱 적합한 모델로 확인되었다(Table 3.1 및 Table 3.2).

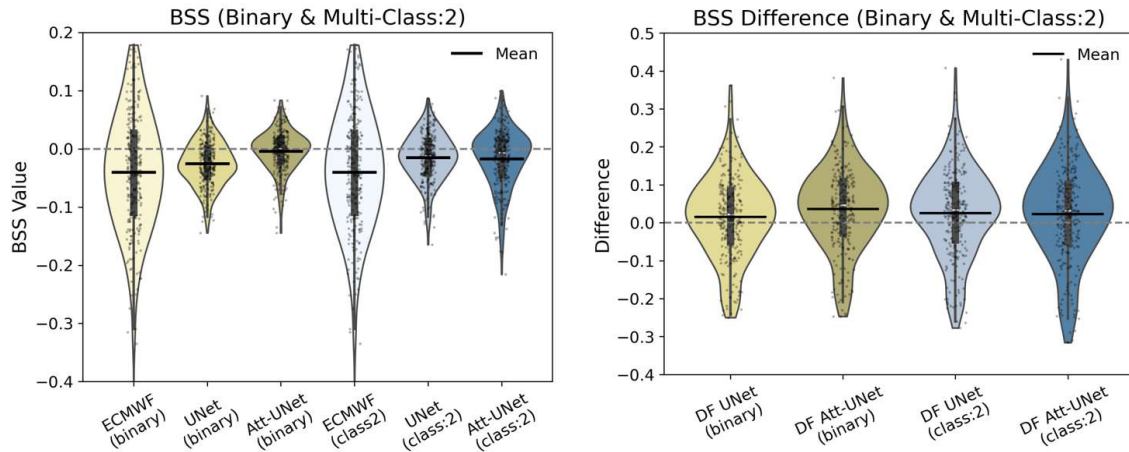


Figure 3.27 The left panel shows violin plots comparing the spatial BSS distributions of ECMWF-S2S and the U-Net, Attention-U-Net-based models (binary and multi-class labels) for 3-week mean maximum temperature. The right panel displays the BSS differences obtained by subtracting the ECMWF-S2S based BSS from each model's BSS at every grid point.

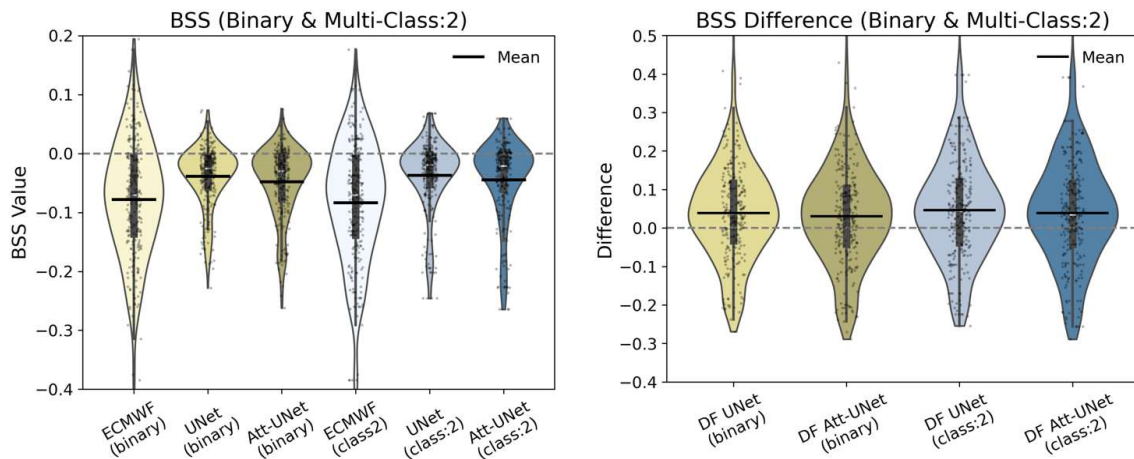


Figure 3.28 The left panel shows violin plots comparing the spatial BSS distributions of ECMWF-S2S and the U-Net, Attention-U-Net-based models (binary and multi-class labels) for 4-week mean maximum temperature. The right panel displays the BSS differences obtained by subtracting the ECMWF-S2S based BSS from each model's BSS at every grid point.

라벨링 방식에 방식에 따른 성능 비교에서도 이진 라벨과 다중 클래스 라벨 구조 간 차이가 뚜렷하게 나타났다. 이진 라벨링의 경우 학습 안정성이 매우 높고, 클래스 불균형의 영향을 상대적으로 잘 완화하였다. 이에 따라 확률 예측 구조가 선형적이고 안정적으로 유지되며, BSS 또한 향상 가능성이 높았다. 반면 다중 클래스 라벨링은 학습 과정에서 잡음이 증가하고, 클래스 불균형이 심각하게 작용하여 예측 성능이 저하되는 경향을 보였다. 특히 극한 고온 탐지 측면에서 이진 라벨링이 훨씬 우수한 결과를 보여, 본 연구에서는 이진 라벨 방식이 명확한 우세를 나타냈다(Figure 3.29).

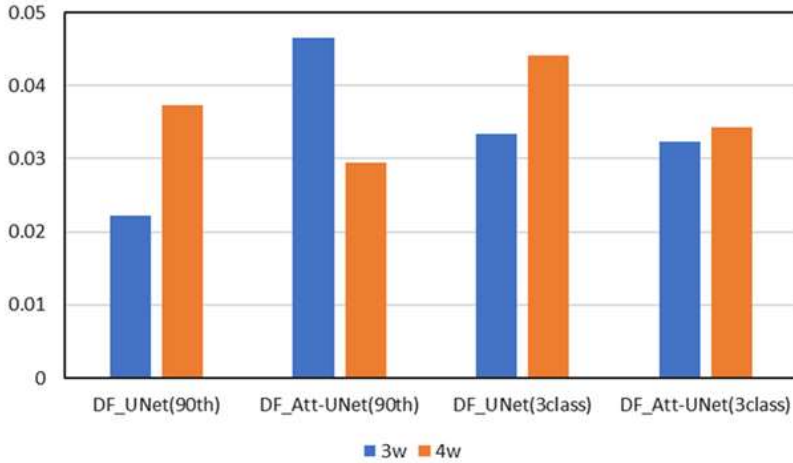


Figure 3.29 Comparison of BSS differences between 3-Week and 4-Week.

연구 결과, 이진 라벨링과 Attention U-Net 모델의 조합이 모든 검증 지표에서 가장 우수한 성능을 보였다(Table 3.1). 이 조합은 ERA5 관측 패턴과 가장 유사한 공간 확률 구조를 재현하며, ECMWF-S2S 대비 양의 BSS 개선 비율이 가장 높게 나타났다. 또한 극한 고온 지점(hotspot) 탐지 성공률이 향상되었고, 모델의 안정성과 일반화 능력에서도 우수한 특성을 보였다. 특히 다중 클래스 분류 구조에 비해 노이즈에 대한 민감도가 크게 낮아, 단순한 라벨 구조에 공간적 주의(attention) 메커니즘을 강화하는 접근이 이상고온 탐지 문제에서 가장 효율적인 전략임을 시사한다.

Table 3.1 This table contrasts U-Net with Attention U-Net across key dimensions such as probability strength, spatial patterns, BSS, extreme-region behavior, and overfitting tendencies. It highlights the clear performance gains and improved stability achieved by incorporating attention mechanisms.

Comparison Item	U-Net	Attention U-Net
Probability strength	Strong	Moderate
Spatial pattern	Rough and blocky	Smooth with higher structural coherence
BSS	Low	High, with clear improvement over ECMWF
Extreme region focus	Unstable	Clear
Risk of over-fitting	High	Low

이진 라벨이 다중 클래스 라벨보다 ECMWF-S2S 예측에서 이상고온 탐지 성능이 우수한 이유는 클래스 경계 불확실성 감소, 학습 안정성 향상, 그리고 극한 사례 표본의 부족 문제를 완화한다는 점에서 설명될 수 있다(Table 3.2). Attention U-Net에 포함된 Attention Gate는 고

기압성 순환 구조나 저층의 따뜻한 기류 수송과 같이 극한 고온과 연관된 핵심 대기 특성에 공간적 초점을 강화하며, 이는 예측 패턴과 BSS 개선에 확인된 바와 같이 모델의 공간적 민감도 향상에 기여한 것으로 해석된다. 본 연구의 결과는 양쯔강 유역 또는 전 지구적 극한 고온 예측을 대상으로 한 최근 딥러닝 기반 연구들과 비교해도 경쟁력 있는 성과로 사료된다(Zuo, J., et al., 2025; Mu, Z., et al., 2024; Fan, Y., et al., 2024). 딥러닝을 활용한 S2S 이상고온 예측 분야에서 동아시아 지역을 대상으로 한 확률 기반 예측 보정 연구로서 그 가치가 있다고 사료된다.

모델 성능의 결정 요인은 라벨 구조에 크게 좌우되는 것으로 판단된다. 이진 라벨 구조는 단순하지만 이상값 신호를 탐지하는데 적합한 반면, 다중 클래스 라벨 구조는 기후 데이터의 특성상 발생 빈도가 낮은 희귀 이벤트에서 오분류 가능성을 높였다. 또한 Attention 메커니즘은 기후 예측 후처리 단계에서 핵심적인 역할을 할 가능성이 높다. 특히 동아시아 지역의 공간 구조적 패턴, 예를 들어 북태평양 고기압의 확장 혹은 티벳 고원의 열적 상승 현상 등은 attention gate를 통해 더욱 명확하게 포착될 수 있다. 마지막으로, ECMWF-S2S의 원시 예측이 가지는 한계를 딥러닝 기반 접근법이 보완할 수 있음을 확인하였다. 비록 BSS 개선 폭이 크지 않더라도 ECMWF 대비 예측의 확률 구조가 향상된 점은 의미가 있으며 이를 통해 S2S 범위(scale)의 이상고온 예측력이 한층 보완될 가능성이 있다.

Table 3.2 This table contrasts binary and multi-class modeling approaches in terms of learning stability, sensitivity to class imbalance, probability prediction behavior, BSS outcomes, and extreme-heat detection. It shows that binary models generally exhibit higher stability and accuracy, especially for extreme events, while multi-class models suffer from greater noise and imbalance effects.

Comparison Item	Binary	Multi-Class
Learning stability	Very high	Low
Effect of class imbalance	Mitigated	Severe
Probability prediction structure	Linear, stable	Increased noise
BSS level	Can be improved	Little to no improvement
Extreme heat detection	Excellent	Inaccurate

3.2. 격자별 주 단위 폭우 강도 및 빈도 예측 모델

3.2.1. 최종 모델 상세 성능

3.2.1.1. 선행 3주 폭우 예측 상세 성능

Figure 3.30은 서울 지역의 격자점에 대한 최종 모델의 2023년부터 2024년까지의 테스트

자료에 대한 선행 3주의 예측된 강수량의 시계열 결과를 보여준다. ERA5 재분석 자료 관측값(검은색 실선)이 90th percentile 임계값(검은색 점선)을 넘는 극한 강수 주들(분홍색 음영 지역)이 여러 번 발생하였다. 모델(노란색 실선)은 많은 경우를 정확하게 포착하고 있다(빨간색 원). 하지만 모델은 때때로 강수가 약할 때 극한으로 예측하는 경향을 보인다(보라색 원). 모델이 극한 강수의 발생 여부와 시기는 어느 정도 잘 잡지만 정량적 강수량 예측은 아직 제한적인 것으로 보인다. Figure 3.31는 관측값과 모델의 예측값에 대한 산점도 그래프를 보여준다. Table 3.3에서 보여주고 있는 강수량 회귀 예측 지표에서 R2 0.266은 실제 변동성의 약 27%만 설명한다는 의미로, 전체 강수량 분포를 정밀하게 예측하기에는 설명력이 낮은 편이다. RMSE 32.338, MAE 19.729 역시 평균 오차가 실제 값의 규모에 비해 큰 편이기 때문에 강수량이 클수록 과소예측, 강수량이 작을수록 과대예측이 크게 발생함을 시사한다. 반면 TCC 0.483은 시계열 상의 상관관계가 중간 수준으로, 비가 많은 구역과 적은 구간의 패턴 자체는 어느 정도 잘 예측하고 있음을 보여준다.

극한강수 분류 예측 지표에서 CSI 0.286, POD 0.375는 90th percentile 임계값 이상의 폭우에 해당하는 주들의 약 30~40% 정도를 검출하고 있다는 의미로, 극한 사례를 놓치는 비율이 아직 적지 않다(Table 3.4). 그럼에도 불구하고 Accuracy가 0.842로 높은 것은 강수량이 많지 않은 구간이 대부분을 차지해 이를 잘 맞추는 것만으로도 전체 정확도가 높게 나왔기 때문으로 해석된다. 따라서 이 모델은 추가적인 특성 반영 및 모델 구조 개선 등으로 성능을 더욱 향상시킬 수 있는 가능성을 시사한다.

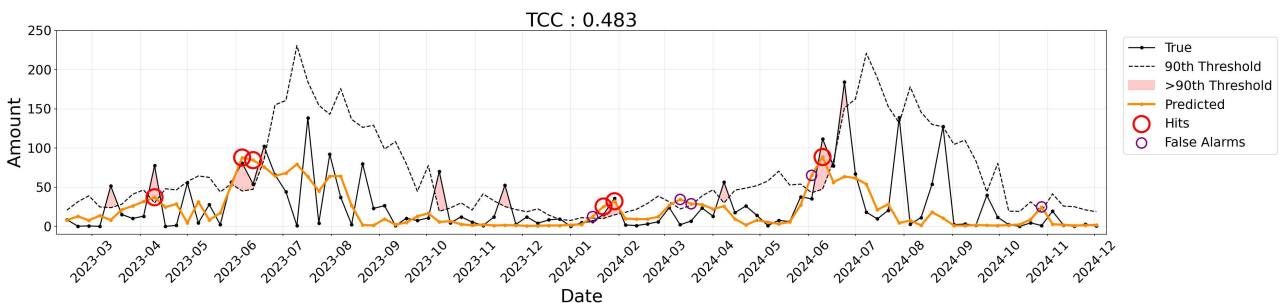


Figure 3.30. Timeseries of rainfall amount at a 3-week lead predicted by the final model for a grid point near Seoul on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th percentile threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.

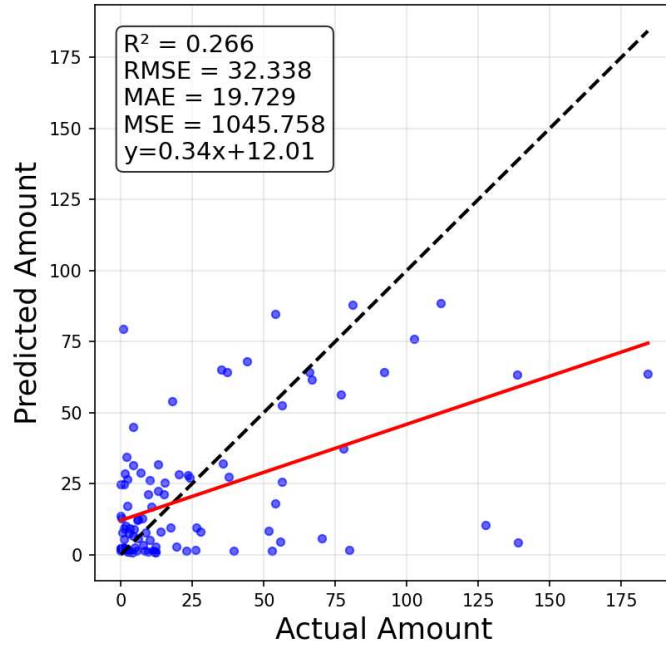


Figure 3.31. Scatter plot of rainfall amount at a 3-week lead predicted by the final model for Seoul grid point compared to actual amount.

Table 3.3. Performance metrics for rainfall amount prediction at a 3-week lead.

R2	RMSE	TCC
0.266	32.338	0.483

Table 3.4. Performance metrics for classification of extreme rainfall weeks at a 3-week lead.

CSI	POD	Accuracy
0.286	0.375	0.842

3.2.1.2. 선행 4주 폭우 예측 상세 성능

Figure 3.32은 서울 지역의 격자점에 대한 최종 모델의 2023년부터 2024년까지의 테스트 자료에 대한 선행 4주 예측 강수량 시계열을 보여준다. ERA5 재분석 자료 관측값(검은색 실선)이 90th percentile 임계값(검은색 점선)을 넘는 극한 강수 주들(분홍색 음영 영역)이 여러 번 발생하였다. 모델(노란색 실선)은 많은 경우를 정확하게 포착하고 있다(hits, 빨간색 원). 하지만 모델은 때때로 강수가 약할 때 극한으로 예측하는 경향을 보인다(보라색 원). 선행 3주 예측 결과와 유사하게 모델이 극한 강수의 발생 여부와 시기는 어느 정도 잘 잡지만 정량적

강수량 예측은 여전히 제한적인 것으로 보인다.

Figure 3.33는 동일 격자점에 대해 관측값과 모델 예측값의 산점도 그래프를 보여주며, 선행 4주 강수량 회귀 성능은 R2 0.193, RMSE 34.84로 실제 변동성 중 약 19%만 설명하고 평균 오차 규모도 큰 편이다. 선행 3주차와 같이 강수량이 클수록 과소예측, 작을수록 과대예측이 발생하는 경향이 있다. 한편 시계열 상관계수 TCC는 0.434로 중간 수준의 상관을 보여 강수량이 많은 구간과 적은 구간의 상대적 패턴은 어느 정도 재현되고 있음을 알 수 있다(Table 3.5).

극한강수 분류 성능에서 CSI 0.294, POD 0.357은 90th percentile 임계값 이상의 폭우에 해당하는 주들의 약 30~40%를 탐지하고 있음을 의미하며, 여전히 상당수 극한 사례를 놓치고 있음을 보여준다(Table 3.6). FAR 0.375와 Accuracy 0.872는 극한으로 잘못 예측하는 비율이 존재하지만, 전체 기간에서 비극한 주가 다수를 차지하고 이를 비교적 잘 맞춘 결과 전체 정확도는 높게 나타난다. 이러한 결과는 선행 4주 예측에서도 극한 강수의 발생 여부와 패턴은 일정 수준 포착되지만, 정량적 예측 정밀도 향상과 미탐지 및 과대탐지의 감소를 위한 추가적인 변수 특성 반영 혹은 모델 구조 개선 여지가 남아 있음을 시사한다.

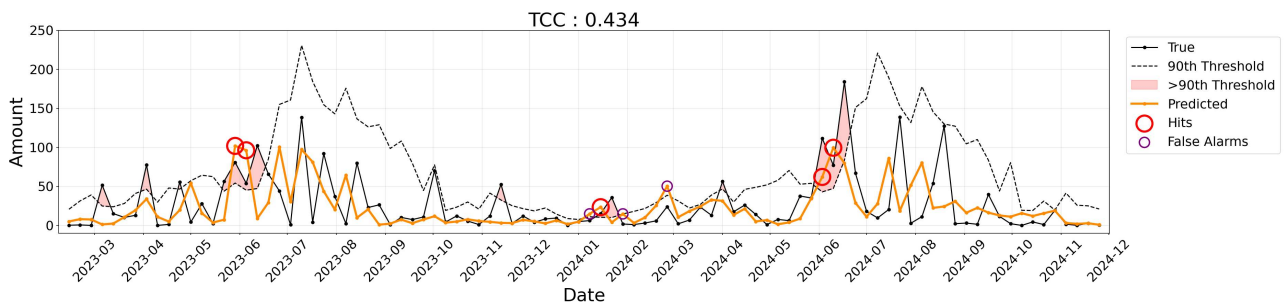


Figure 3.32. Timeseries of rainfall amount at a 4-week lead predicted by the final model for a grid point near Seoul on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th percentile threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.

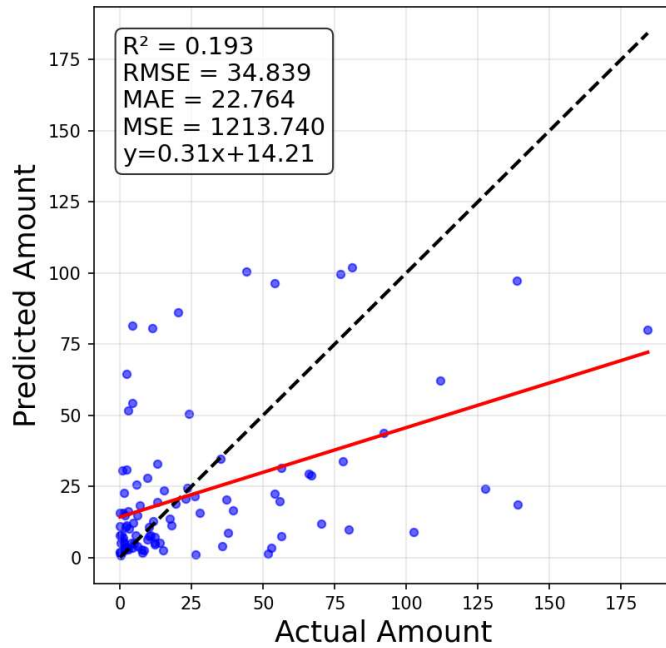


Figure 3.33. Scatter plot of rainfall amount at a 4-week lead predicted by the final model for Seoul grid point compared to actual amount.

Table 3.5. Performance metrics for rainfall amount prediction at a 4-week lead.

R2	RMSE	TCC
0.193	34.839	0.434

Table 3.6. Performance metrics for classification of extreme rainfall weeks at a 4-week lead.

CSI	POD	Accuracy
0.294	0.357	0.872

3.2.1.3. 다른 격자점 최종 모델 상세 성능

Figures 3.34-37은 서울 지역 이외에 대전, 부산, 광주, 평창 격자점에 해당하는 최종 모델의 결과를 보여준다. Tables 3.7-8은 각각 강수량 회귀 성능 평가 결과와 극한 강수 주간의 분류 결과를 보여준다. 다른 격자점들도 서울 격자점과 마찬가지로 2023년과 2024년 여름철의 폭우 주간을 잘 탐지하고 있는 것을 보여주지만 모델의 성능이 지역별 차이가 존재한다.

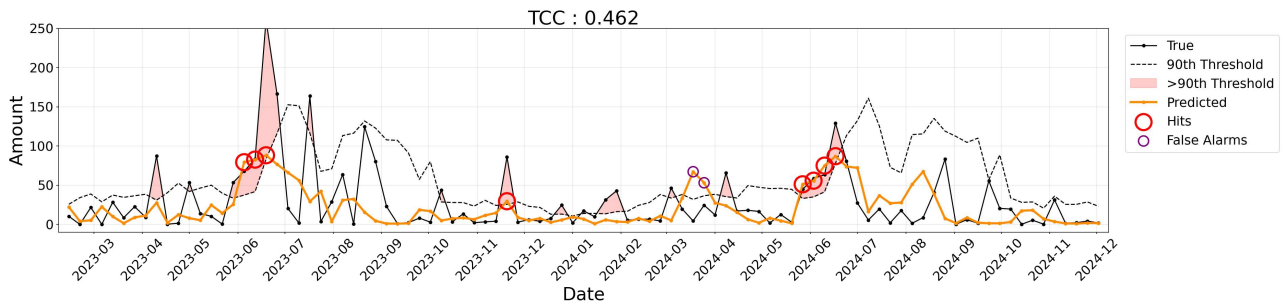


Figure 3.34. Timeseries of rainfall amount predicted by the final model for a grid point near Daejeon on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.

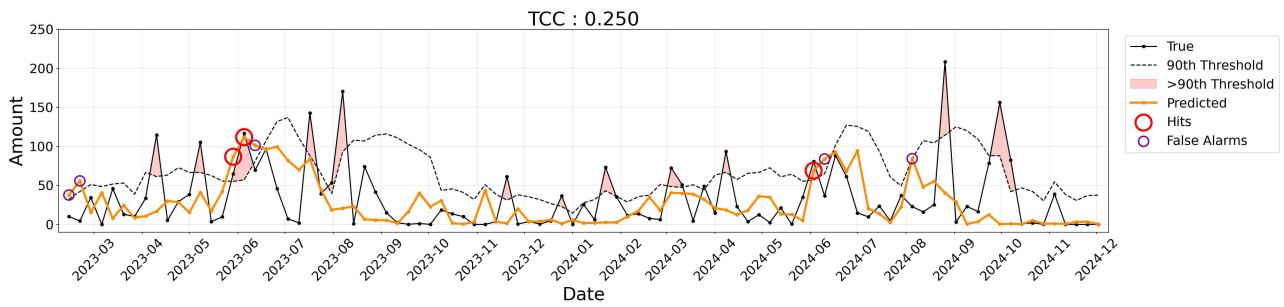


Figure 3.35. Timeseries of rainfall amount predicted by the final model for a grid point near Busan on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.

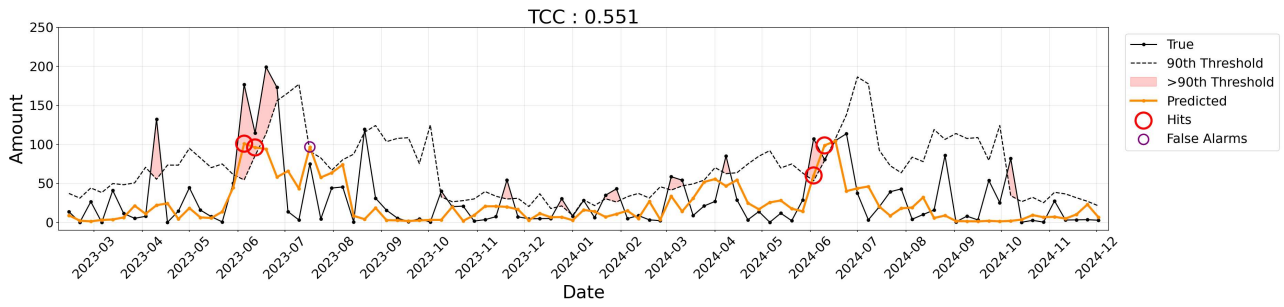


Figure 3.36. Timeseries of rainfall amount predicted by the final model for a grid point near Gwangju on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.

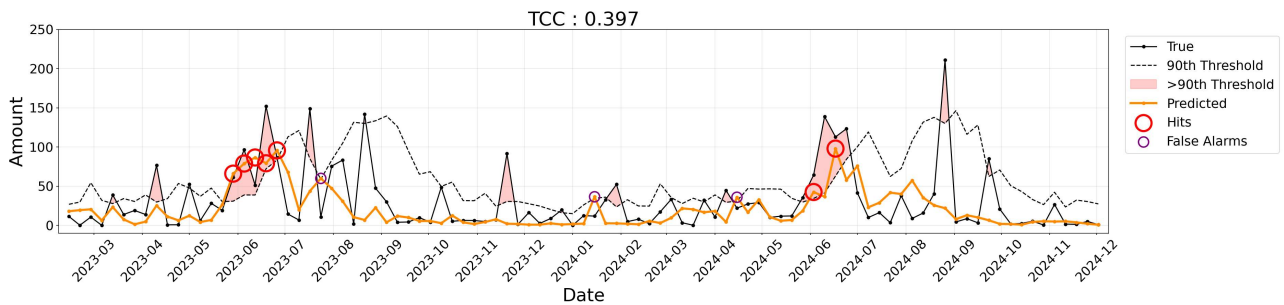


Figure 3.37. Timeseries of rainfall amount predicted by the final model for a grid point near Pyungchang on test data from 2023 to 2024. Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded. Predicted rainfall amount is shown in yellow solid line with hits cases in red circle and false alarms in purple circle.

Table 3.7. Performance metrics for rainfall amount prediction for different grid points.

	R2	RMSE	TCC
Daejeon	0.262	37.159	0.462
Busan	0.081	43.766	0.250
Gwangju	0.352	34.531	0.551
Pyungchang	0.227	37.894	0.397

Table 3.8. Performance metrics for classification of extreme rainfall weeks for different grid points.

	CSI	POD	Accuracy
Daejeon	0.364	0.4	0.853
Busan	0.130	0.167	0.789
Gwangju	0.222	0.235	0.853
Pyungchang	0.28	0.318	0.811

3.2.2. Multi-task learning 적용 실험

Figure 3.38은 single-task learning과 multi-task learning 기반 인공지능 모델이 주간 누적 강수량과 극한 강수 이벤트를 얼마나 잘 재현하는지 비교한 결과이다. ERA5 실제 강수량 (검은 실선), 90th percentile 임계값 (검은 점선), 임계값을 초과하는 극한 강수(분홍색 음영 지역)를 기준으로, single-task learning 기반 모델(초록색 실선)은 전반적으로 강수량 수준을 과소추정하며, 특히 여름철과 같이 실제 강수 변동성이 큰 구간에서 피크를 거의 따라가지 못하는 것을 보였다. 반면 multi-task learning 기반 모델(빨간색 실선)은 계절별 강수량 변화와 극값의 크기를 더 잘 예측하며, 극한 강수 주차에서 실제 피크 시점과 규모를 상대적으로 유사하게 예측하고 있어 시계열 패턴 적합도가 더 높은 것을 보인다. 또한 multi-task learning 기반 모델은 강수량이 거의 없는 기간에는 낮은 값을 유지하면서 급격히 상승하는 구간에서는 빠르게 반응해 급증하는 형태를 보여 임계값 주변의 민감도가 개선된 것으로 판단된다. 이러한 차이는 강수량 회귀와 극한강수일수를 함께 학습함으로써 multi-task learning 기반 모델이 극한 강수의 특성을 더 풍부하게 학습하고 단순 강수량 추정에만 집중한 single-task learning 기반 모델보다 실제 극한 사례 예측 능력이 우수해졌음을 시사한다.

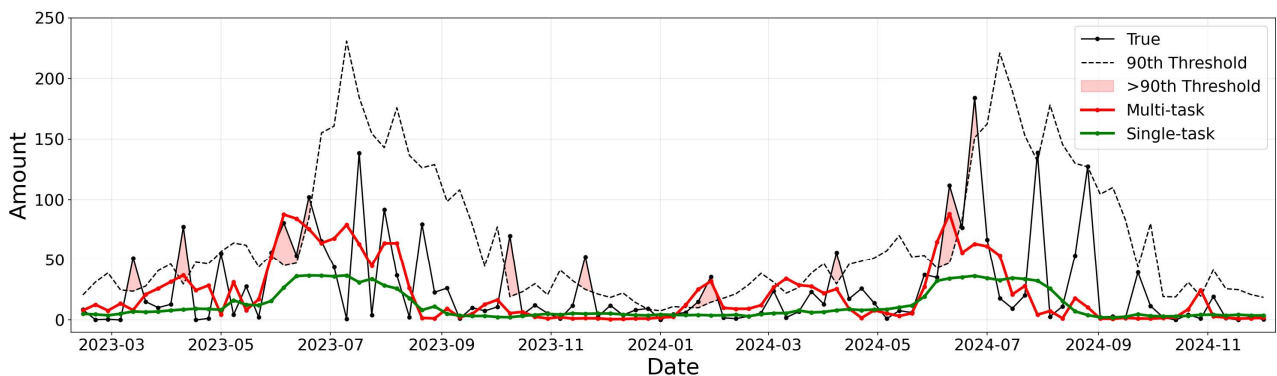


Figure 3.38. Comparison between single-task learning-based results (green solid line) and multi-task learning-based results (red solid line). Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded.

3.2.3. 입력 변수 조합 실험

Figure 3.39은 전체 변수 세트에 따른 CSI와 TCC 값을 비교한 결과이다. 실험한 전체 변수 세트는 Table 3.9에 보여준다. 전반적으로 세트가 복잡해질수록 성능이 향상되는 경향이 나타났다. 특히 분홍색으로 표시된 Set9가 두 지표 모두에서 가장 우수한 값을 보여 최적 조합임을 시사한다. CSI는 관측과 예측이 동시에 사건을 맞힌 비율을 나타내며 값이 클수록 극한 강수 예측이 잘 되었다는 것을 의미한다. 단일 변수 세트(Set0, Set1 등)보다는 여러 시간 스케일의 정보를 함께 사용할 때 CSI가 뚜렷하게 증가하며, 그중 ISO, 저주파, 기후값을 모두 포함한 Set9에서 최대값이 나타났다. 이는 다양한 시간 스케일의 변동성을 함께 활용하는 것이 극한 강수 발생 여부를 맞히는 데 가장 효과적임을 의미한다. TCC는 예측과 관측의 시계열 이상값 상관을 나타내며, 1에 가까울수록 시기별 변동성을 잘 재현한다. 그림에서 대부분의 변수 세트가 양의 상관을 보이지만, Set9에서 상관계수가 가장 커져 시간 변화 패턴을 가장 잘 따라간다. 이는 ISO, 저주파, 기후 성분을 동시에 고려하면 주간 이상 강수의 연도 간 변동성을 더욱 안정적으로 예측할 수 있음을 보여준다. CSI와 TCC를 함께 고려했을 때, 극한 강수의 발생 여부와 그 시간적 변동성을 동시에 잘 설명하는 조합은 Set9 이며, 다른 세트들은 한 지표에서만 상대적으로 좋아도 둘 다 높은 값을 얻지는 못했다. 따라서 ISO, 저주파, 기후 성분을 모두 포함하는 변수 구성이 본 연구의 예측 모형에서 가장 균형 잡힌 예측 성능을 제공하며 이후 실험이나 운영 적용에서도 기본 세팅으로 사용하는 것이 타당한 것으로 판단하였다.

각 성분별 입력 변수 조합 실험 결과는 Figure 3.40에서 좀 더 명확하게 보여준다. 왼쪽 패널은 CSI, 오른쪽은 TCC를 보여준다. ISO, ISO 성분과 저주파수 성분, 기후평균 성분으로 구성된 Best Combination(분홍색 박스)이 CSI 0.286와 TCC 0.483로 다른 실험 대비 가장 높게 나타났다. 반면 Raw Only, ISO Only, Synoptic Only와 같이 단일 유형의 변수를 사용할 경우 CSI와 TCC가 전반적으로 낮아져 개별 성분만으로는 극한 발생 여부와 시기, 강도 변화를 충분히 예측하지 못하는 것으로 보인다. 특히 Synoptic Only 실험의 CSI와 TCC가 가장 낮은 것은 순환 규모의 정보만으로는 주기성이 강한 ISO나 저주파수 배경 상태, 기후 평균에서 오는 신호를 포착하기 어렵다는 점을 반영한다. 따라서 극한 강수 예측에서는 단일 정보에 의존하기보다 시간 스케일과 배경 상태가 다른 여러 성분을 결합해 다중 스케일 정보를 통합하는 것이 필수적이며 본 연구의 Best Combination 결과가 그 효과를 정량적으로 보여주고 있다.

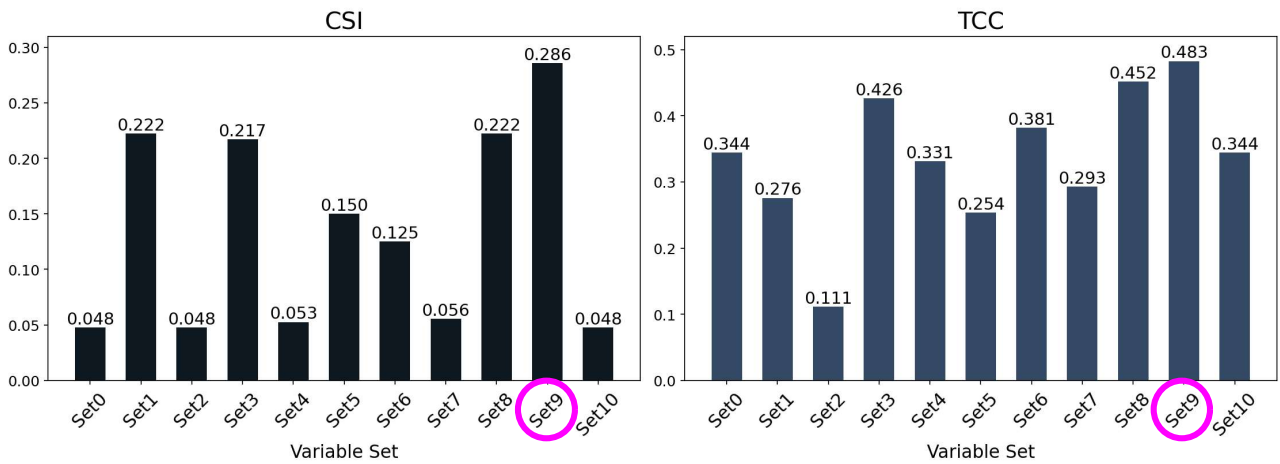


Figure 3.39. Model performance of critical success index (CSI) values for each bar on the left column and temporal anomaly correlation coefficients (TCC) values for each bar on the right column for each combination of variable set (refer to Table X.X for list of variable sets). Best combination in pink circle is composed of ISO, low frequency, and climatology components.

Table 3.9. List of input variable sets.

Experiment set	Variable combination
Set0	All Variables
Set1	Masks, ISO
Set2	Masks, ISO, Synoptic
Set3	Masks, ISO, Low Freq.
Set4	Masks, ISO, Synoptic, Low Freq.
Set5	Masks, Synoptic
Set6	Masks, Synoptic, Low Freq.
Set7	Masks, ISO, Clim.
Set8	Masks, ISO, Synoptic, Clim.
Set9	Masks, ISO, Low Freq., Clim.
Set10	Masks, Synoptic, Low Freq., Clim.

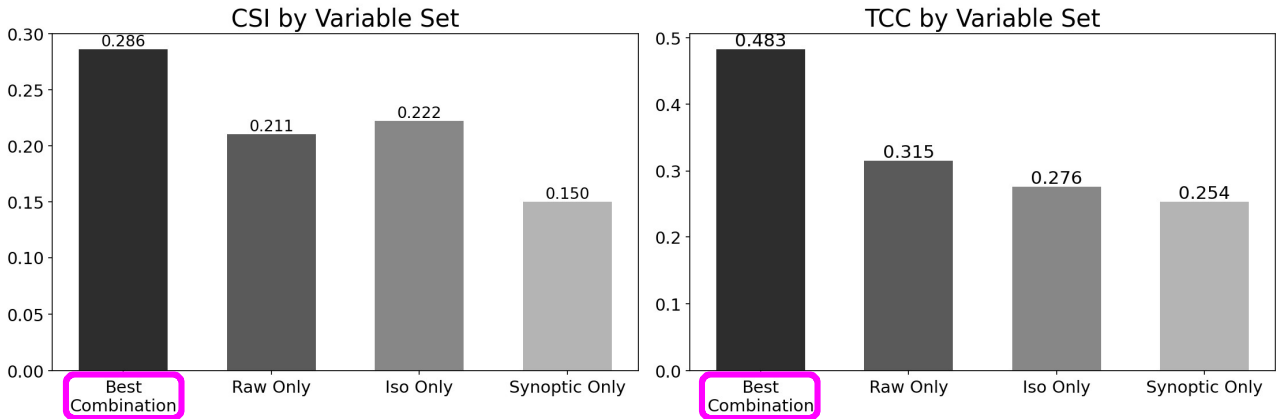


Figure 3.40. Summary of experiments of combinations of variable set with critical success index (CSI) values for each bar on the left column and temporal anomaly correlation coefficients (TCC) values for each bar on the right column. Best combination in pink box is composed of ISO, low frequency, and climatology components.

3.2.4. 모델 구조 변형 실험

Figure 3.41은 주간 누적 강수량 시계열에서 ResNet18 단독 모형과 ResNet18+LSTM 하이브리드 모형의 차이를 보여준다. 두 모형 모두 전체적인 변동성은 ERA5 관측(검은 실선)을 따라 가지만 극값에 대한 재현 능력에서 차이가 나타난다. 우선 90th percentile 임계값(검은 점선) 주변을 보면 ResNet18+LSTM(빨간선)은 여름철과 장마철과 같이 강수 피크가 크게 증가하는 구간에서 관측 피크의 시점과 크기를 상대적으로 더 잘 따라간다. 반면 ResNet18 단독(초록선)은 몇몇 강한 호우 구간에서 피크의 크기를 과소 모의하거나 시점을 약간 놓치는 경우가 더 자주 나타나 시간적 종속성을 충분히 활용하지 못하는 모습을 보인다. 또한 비극한 구간에서는 두 모형 모두 관측에 비해 평활화된 패턴을 보이지만, LSTM이 결합된 경우 계절적 증감과 저빈도 변동을 좀 더 부드럽게 재현하여 전체 RMSE와 상관계수를 개선하는 데 기여하는 것으로 해석된다. 따라서 CNN 기반 ResNet18에 LSTM을 결합한 구조가 주간 강수의 시계열 구조와 특히 90th percentile 임계값 이상 극한 강수의 크기와 시점을 포착하는 데 더 적합한 모형인 것으로 확인된다.

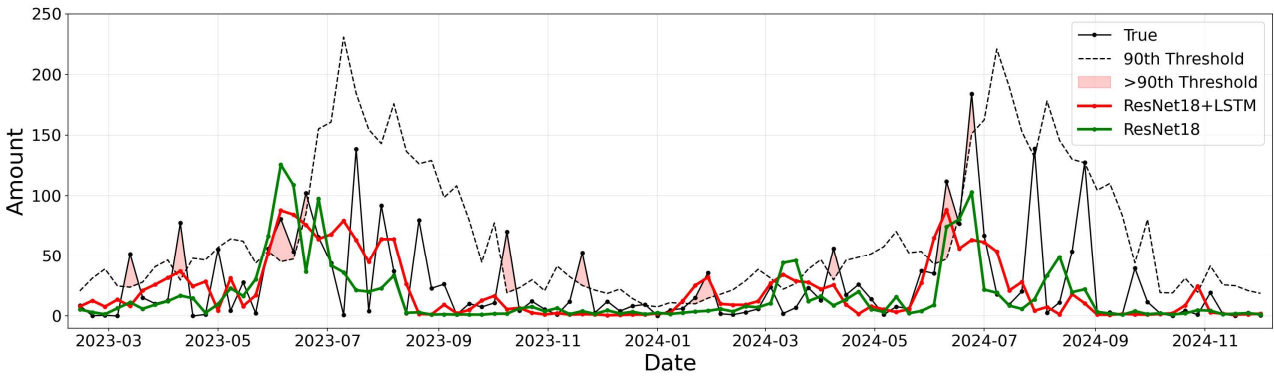


Figure 3.41. Comparison between ResNet-18 (green solid line) and ResNet-18+LSTM (red solid line). Weekly accumulated rainfall amount from ERA5 reanalysis data are shown in black solid line. The 90th percentile threshold are in black dashed line. Extreme cases that exceed the 90th threshold are shaded.

3.2.5. 기존 역학 모델 성능과 비교 평가

3.2.5.1. 기존 역학 모델 성능 평가

Figure 3.42은 ECMWF 1.5° 격자 Forecast 자료의 7일 누적 강수량 예측 성능을 ERA5 재분석 강수량과의 상관계수로 나타낸 막대그래프이다. 각 패널은 서로 다른 세 개의 격자점 (Grid point1: 부산, Grid point2: 청주, Grid point3: 서울)에 대해 전체 계절(All)과 봄, 여름, 가을, 겨울별로 검증한 결과를 보여준다. 빨간색 계열 막대는 3주차(lead week 3), 파란색 계열 막대는 4주차(lead week 4) 예측의 상관계수를 의미하며, 값이 클수록 ECMWF 예측이 관측(ERA5)과 서로 더 잘 일치함을 뜻한다. 전반적으로 특정 격자와 계절, 특히 여름철에서 상관계수가 상대적으로 높아 예측 가능성이 크지만, 일부 계절과 격자에서는 상관계수가 0에 가깝거나 음수로 나타나 계절 내 강수 예측이 매우 어려움을 시사한다.

Figure 3.43는 격자점 1에서의 ECMWF 초기화 날짜에 따른 주간 누적 강수량 시계열을 보여준다. 상단의 두 패널은 ERA5 재분석 자료의 3, 4주차 주간 누적 강수량을 나타낸다. 각 패널에서 검은 실선은 실제 주간 누적 강수량, 검은 점선은 ERA5 자료로부터 계산한 90th percentile 임계값으로, 이 값을 초과하면 극한 강수 사건으로 정의된다. 세 번째와 네 번째 패널은 동일 격자점에 대한 ECMWF 3, 4주차 Forecast 자료의 주간 누적 강수량과 ECMWF Hindcast로부터 계산한 90th percentile 임계값(점선)을 보여준다. ERA5와 ECMWF에서 모두 임계값을 초과한 경우, 즉 예측과 관측이 동시에 극한 강수로 나타난 “적중(hit)” 사례는 빨간 원으로 표시하여 주별 초기일자에 따른 극한 강수 예측 성공 여부를 나타내었다. 격자점 1에서는 ECMWF Forecast 모델이 극한 강수의 발생 시점을 잘 탐지하지 못하는 것을 보여준다. Figures 3.44-45은 격자점 2와 3에 대한 ECMWF Forecast 및 ERA5 재분석 자료를 비교한 그림이다. 두 격자점에 대해서는 ECMWF는 극한 강수 사건을 예측하지 못하고 있다.

Figure 3.46는 ECMWF Forecast 모델의 3, 4주차 극한 강수 예측 성능을 여러 검증 지표로 요약한 막대그래프이다. 가로축에는 Accuracy, Precision, Recall, F1-score, CSI(Critical

Success Index)가 나열되어 있으며, 각 지표마다 짙은 색 막대는 선행 3주, 연한 색 막대는 선행 4주 결과를 나타낸다. 정확도는 두 리드타임 모두 약 0.8 수준으로 비교적 높지만, 극한 강수는 발생 빈도가 낮기 때문에 Precision, Recall, F1, CSI 값은 0.1 안팎으로 낮게 나타나 사례가 드문 극한 강수 사건을 포착하는 데 한계가 있음을 시사한다.

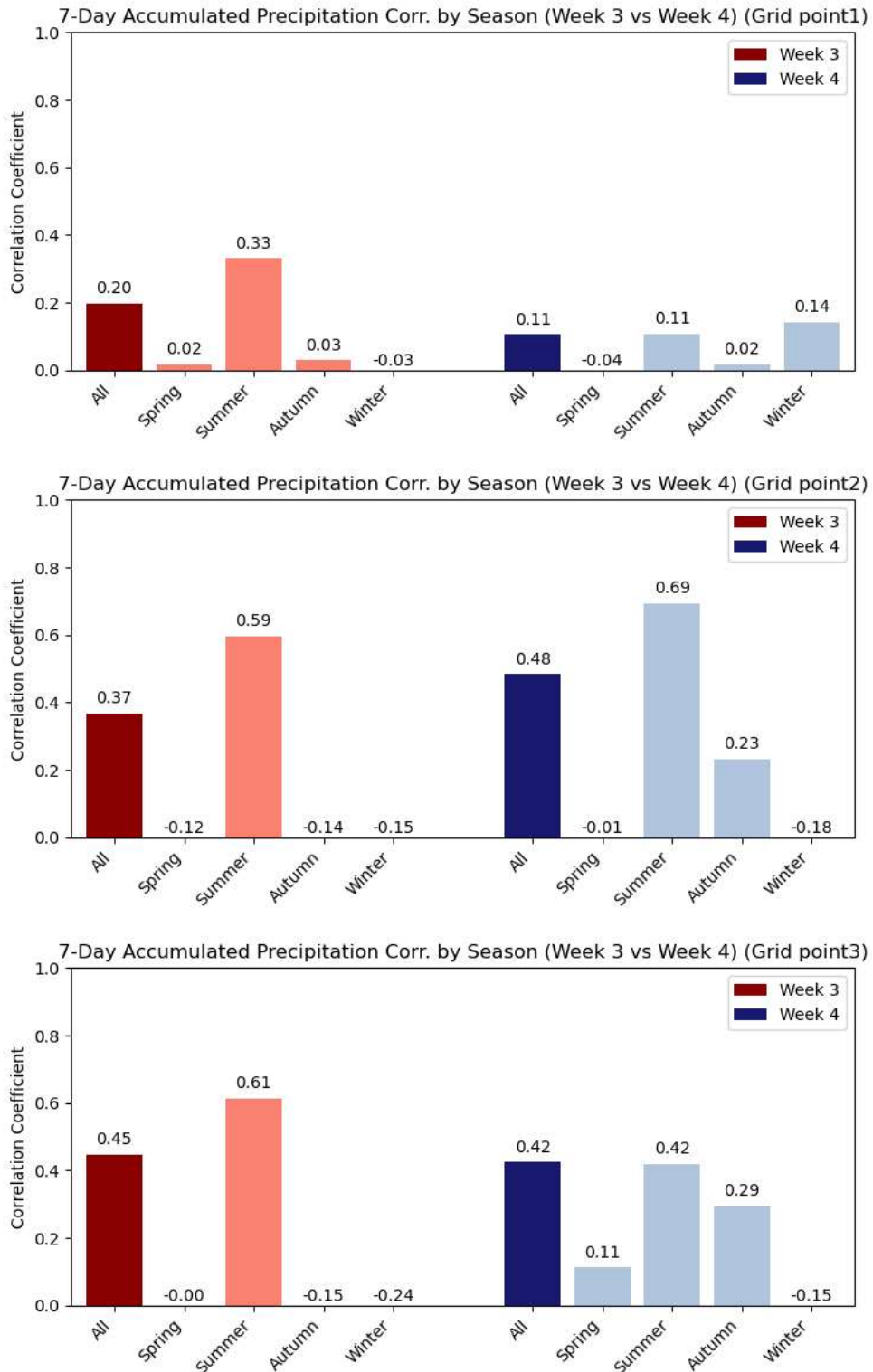


Figure 3.42. Bar plots of ECMWF (1.5°) forecast model performance for total precipitation compared to ERA5 reanalysis data for week 3 (radish colored bars) and 4 (blueish colored bars) for different grid points for all seasons and individual season.

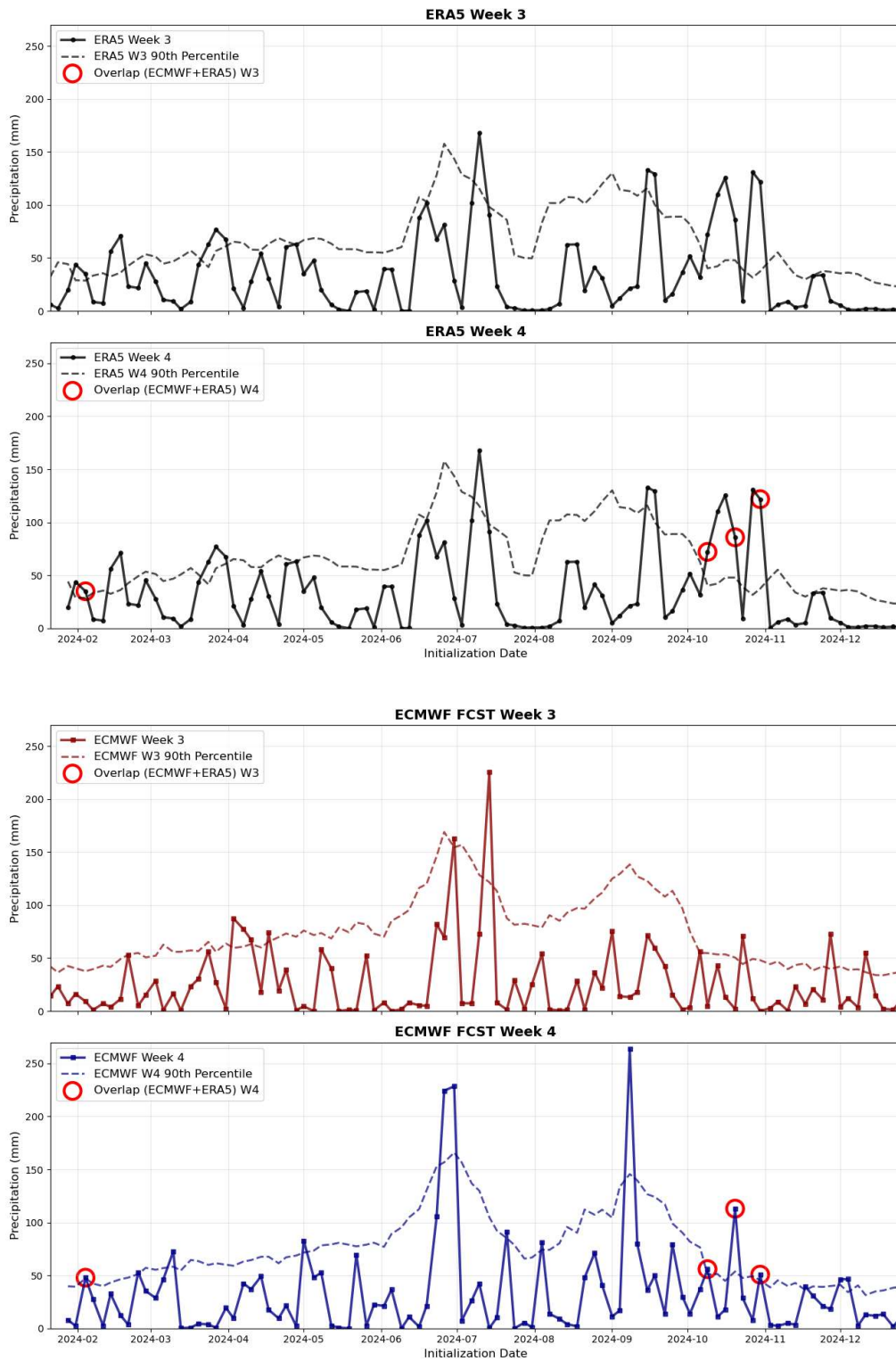


Figure 3.43. Timeseries of weekly accumulated rainfall amount at grid point 1 for week 3 and 4 in the top two panels and the corresponding ECMWF forecast results. The 90th percentile thresholds are shown in black dashed line for ERA5 and ECMWF hindcast data. Hit cases that overlap each other are shown in red circle in both ERA5 and ECMWF panels.

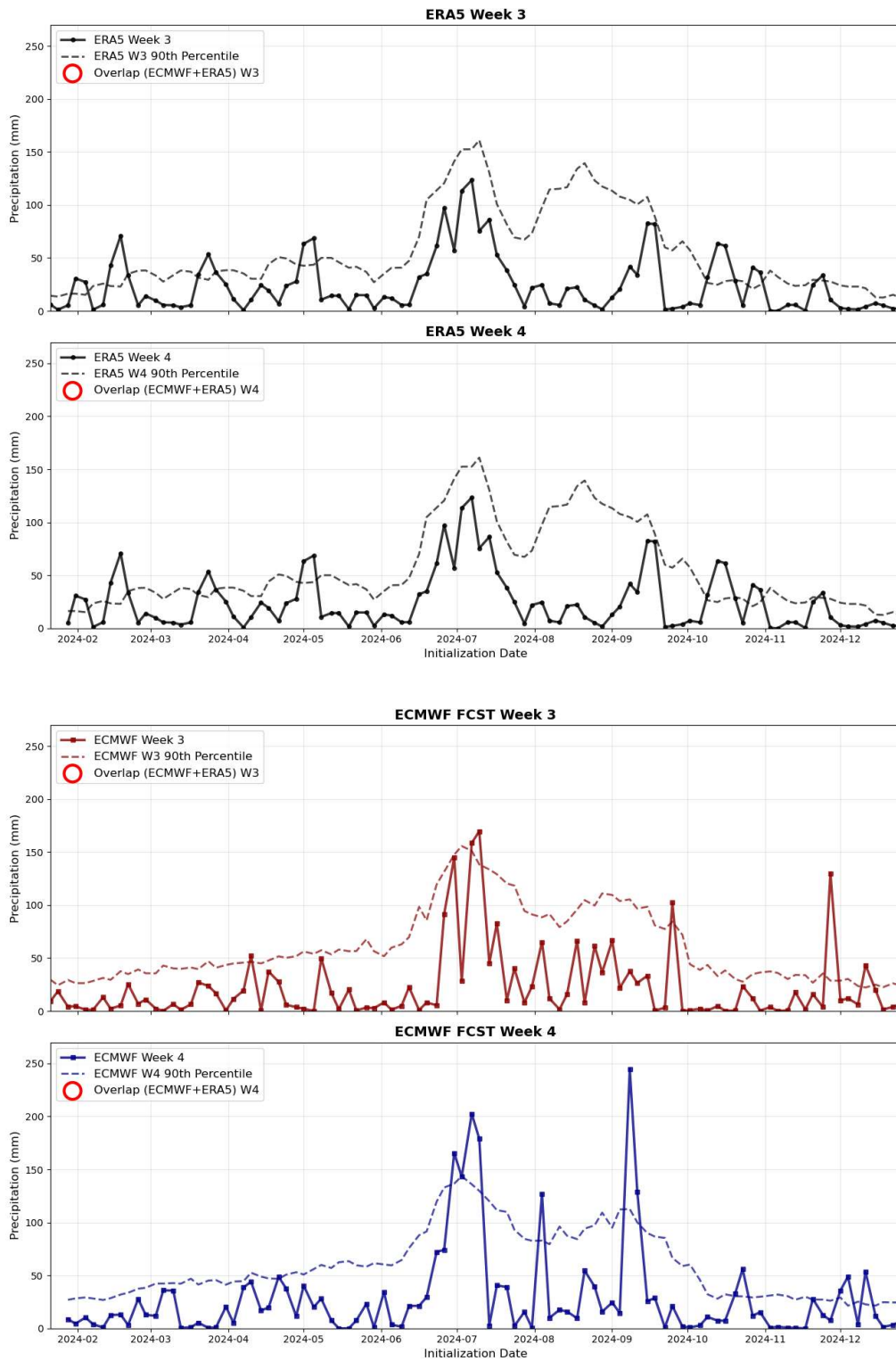


Figure 3.44. Timeseries of weekly accumulated rainfall amount at grid point 2 for week 3 and 4 in the top two panels and the corresponding ECMWF forecast results. The 90th percentile thresholds are shown in black dashed line for ERA5 and ECMWF hindcast data. Hit cases that overlap each other are shown in red circle in both ERA5 and ECMWF panels.

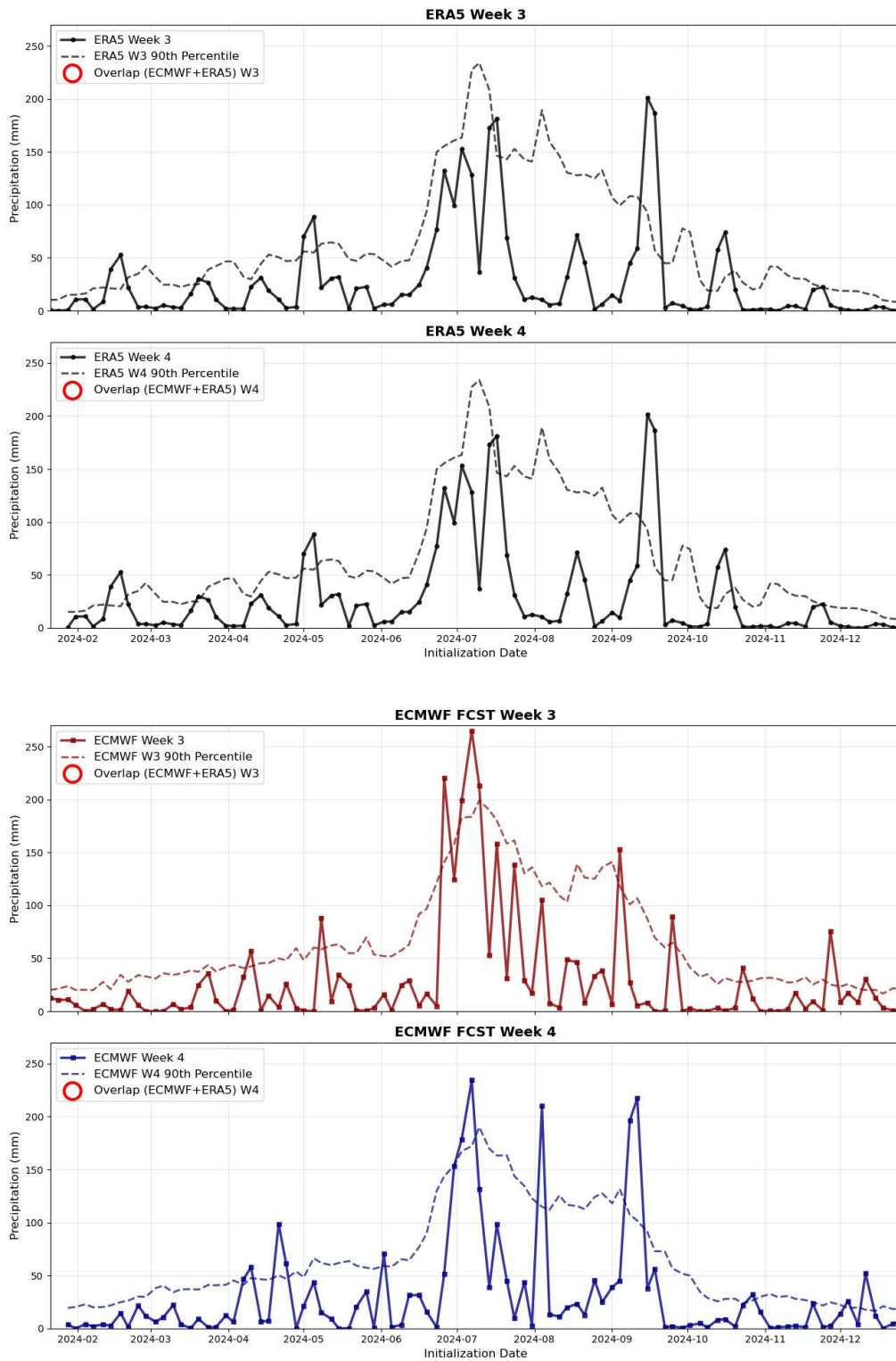


Figure 3.45. Timeseries of weekly accumulated rainfall amount at grid point 3 for week 3 and 4 in the top two panels and the corresponding ECMWF forecast results. The 90th percentile thresholds are shown in black dashed line for ERA5 and ECMWF hindcast data. Hit cases that overlap each other are shown in red circle in both ERA5 and ECMWF panels.

ECMWF Extreme prec. Week3 & Week4

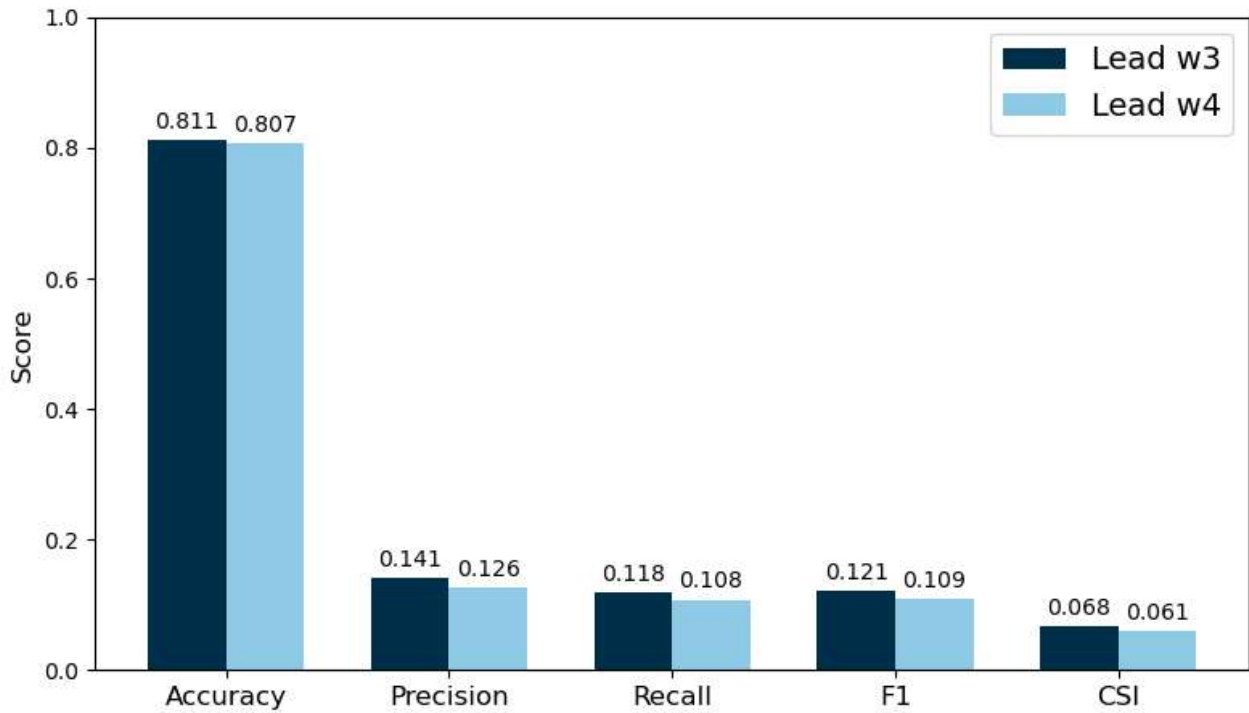


Figure 3.46. ECMWF forecast model performance for extreme precipitation for week 3 and 4 with performance metrics including accuracy, precision, recall, F1-score, and CSI.

3.2.5.2. 사례 대비 선행 3, 4주 폭우 예측 정성적 평가

Figure 3.47은 서울 격자점에서 발생한 한 극한 강수 사례를 분석한 사례 연구 결과를 보여준다. 왼쪽 패널의 시계열 그래프에서 특정 주간에 ERA5 관측과 예측 강수량이 90th percentile을 초과한 지점이 분홍색 타원으로 강조되어 있으며, 이 시기에 극한 강수가 발생했음을 나타낸다. 가운데 지도는 해당 주(2023년 6월 5-11일) 동안 우리나라 전역의 주간 누적 강수량의 백분위 분포를 나타내며, 서울 인근에 두드러진 강수 발생이 형성된 모습을 보여준다. 오른쪽 확대 패널은 서울 격자 주변을 확대한 것으로, 예측된 강수량이 60-80% 이상 수준으로 집중되어 있어 시계열에서 확인된 극한 강수 신호와 공간적으로도 잘 부합함을 시각적으로 제시한다.

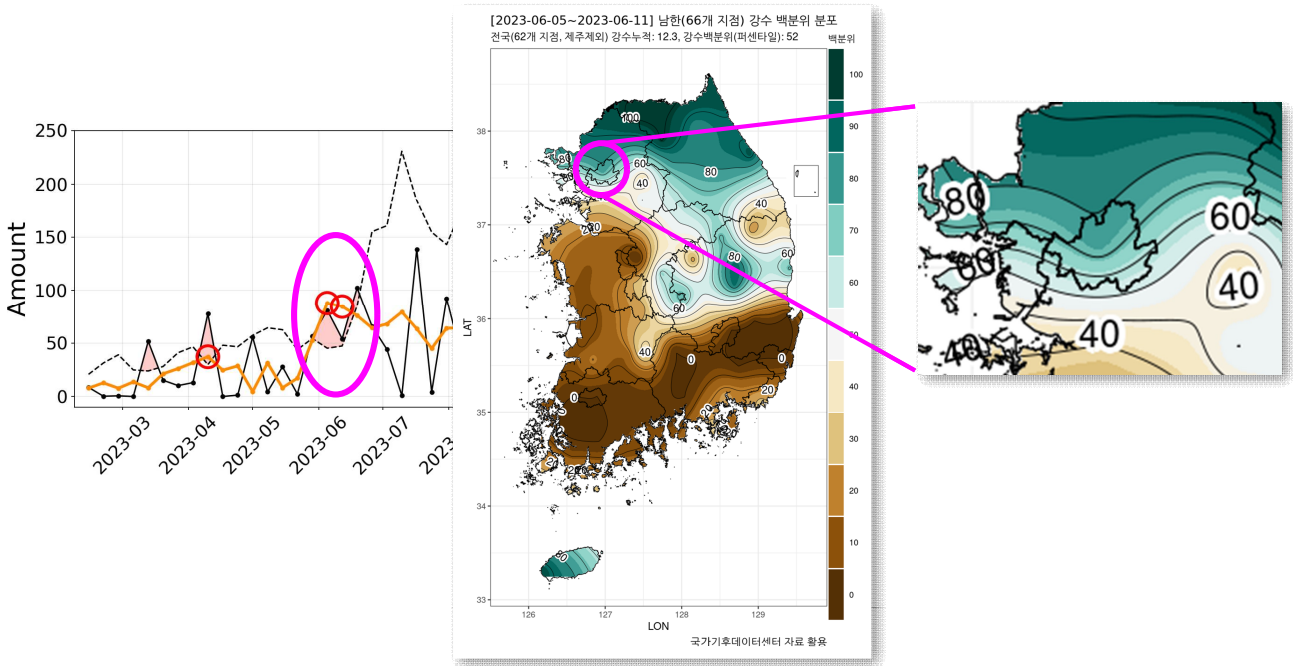


Figure 3.47. A case study with results for the Seoul grid point.

3.3. 선행 3주 동아시아 해역 고수온 예측 모델 개발

3.3.1. 이상치 예측에서 있어 가중 손실 함수 적용성 평가

본 절에서는 앞서 2.3.3.2절에서 제시한 7가지 민감도 실험 모델 전체를 대상으로 손실 함수 가중치 규모(0, 2, 4, 6)에 따른 비교 결과를 제시한다.

3.3.1.1. 입력 도메인 사이즈/해상도 변형 실험 평가

가중 손실 함수의 규모 외에도 입력 도메인 크기와 해상도에 대한 민감도 분석 결과는 아래 Figure 3.48에 제시하였다. 5° 간격의 정사각 격자를 사용하는 큰 도메인(G5/L5) 실험은, 상세 격자($2.5^\circ \times 2.5^\circ$)를 적용한 작은 도메인(S2.5) 실험에 비해 ACC와 CSI 성능이 모두 낮게 나타났다. 이는 격자 해상도의 영향일 수도 있으나, 선행 3주 동아시아 해역 고수온 현상 예측에 필요한 영향 영역을 전구 수준의 큰 도메인까지 확장할 필요가 없음을 시사한다. 즉, 동아시아 해역을 적절히 커버하는 범위 내에서 높은 해상도의 시그널을 입력하는 방식이 더 유리하다고 해석할 수 있다. 또 다른 가능성은 G5와 L5 모두 정사각형 격자가 아니므로 위도 방향의 빈 격자를 zero padding 처리하였는데, 이러한 방식으로 생성된 정보는 2D convolution을 통해 시그널을 추출한 뒤 최종적으로 32×32 정사각형 격자로 재확장하는 UNet 모형 실험의 입력으로 적절하지 않을 수 있다. 따라서 훈련 자료와 라벨 간의 구조적 일치 여부 역시 중요한 요소로 판단된다.

가중 손실 함수의 규모에 따른 예측 성능은 도메인 크기와 해상도와 크게 관계없이 ACC는 감소하고 CSI는 증가하는 경향을 보였다. 다만 CSI의 경우, 가중치 규모가 6일 때 오히려 성능이 다소 저하되는 양상이 나타났다. 종합적으로 판단할 때, 가중치 규모 2 또는 4가 적절한

값으로 선택될 수 있었다.

이상치 공간 패턴의 유사성을 평가하는 ACC와 격자별 이상치 탐지율의 도메인 평균을 평가하는 CSI가 다소 상반된 경향을 보인다는 점은, 원형 모델을 조정하는 과정에서 어느 지표에 주안점을 둘 것인지에 대한 선택의 문제를 야기한다. 본 연구에서는 이상치 탐지 여부에 보다 중점을 두었기에 CSI 개선 정도를 모델 개발의 주요 목표로 설정하였으며, 이를 모델 성능 판단의 기준으로 삼았다.

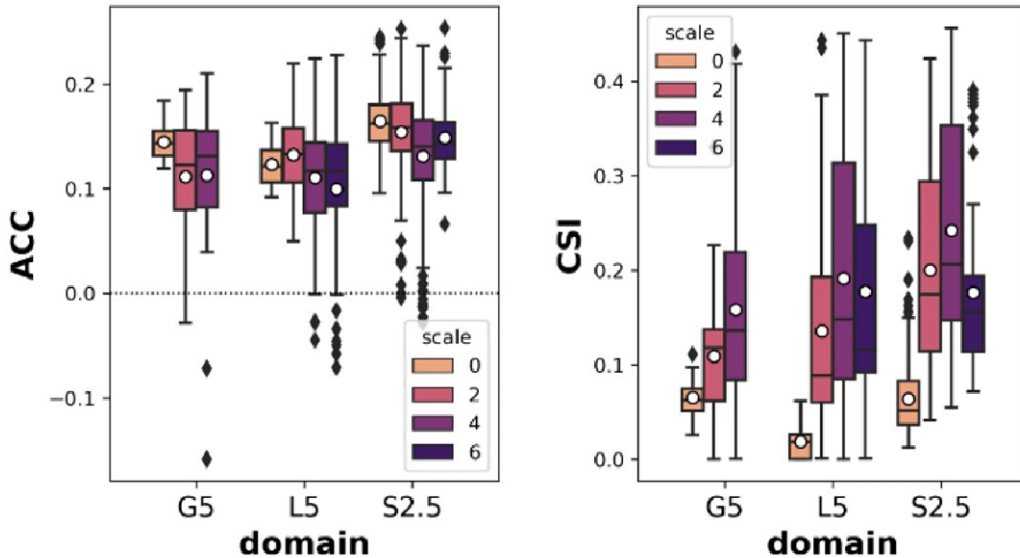


Figure 3.48 Boxplots showing the distribution of (left) ACC and (right) CSI across three groups with different input domain size/resolution (G5: global 5deg, L5: large 5deg, S2.5: small 2.5deg) on the x-axis, with colors indicating the scale of loss function weighting (0,2,4, and 6) for the 104 test samples across 2023–2024. The central line (white circle) in each box denotes the median (mean), box edges indicate the first and third quartiles, whiskers extend to 1.5 times the interquartile range, and points beyond the whiskers represent outliers.

3.3.1.2. 학습/검증 데이터 비율 변경 실험 결과 해석

15K가 넘는 데이터를 훈련/검증 샘플로 분할하는 비율에 따른 민감도를 평가하였다 (Figure 3.49). 일반적으로 훈련용 샘플을 충분히 확보하기 위해 5:1 혹은 4:1 비율로 나누는 경우가 많다. 그러나 본 연구에서는 고수온 현상을 90%tile 기준으로 정의하였기에, 7일 누적 이상치 값은 매우 드문 사례로 나타날 수밖에 없으며, 검증 샘플에 고수온 사례가 충분히 포함되지 못할 가능성이 존재한다. 이러한 우려를 반영하여 검증 데이터 비율을 높여 3:1과 2:1 옵션을 비교하였다.

검증 데이터 비율이 더 높은 2:1의 경우 ACC 성능은 전체적으로 다소 낮아졌으나, CSI는 크게 개선되는 결과를 확인할 수 있었다. 특히 가중치 규모 2와 4에서 CSI 평균값이 3:1 대비

유의미하게 높게 나타나, 이상치 탐지 모델에서는 검증 데이터 비율을 높이는 방식이 효과적일 수 있다는 판단을 내릴 수 있었다. Figure 3.48에서와 마찬가지로 가중치 규모 4에서 가장 높은 CSI를 기록하였으나 ACC 감소를 고려하여, 가중치 규모 2를 최적값으로 선정하였다.

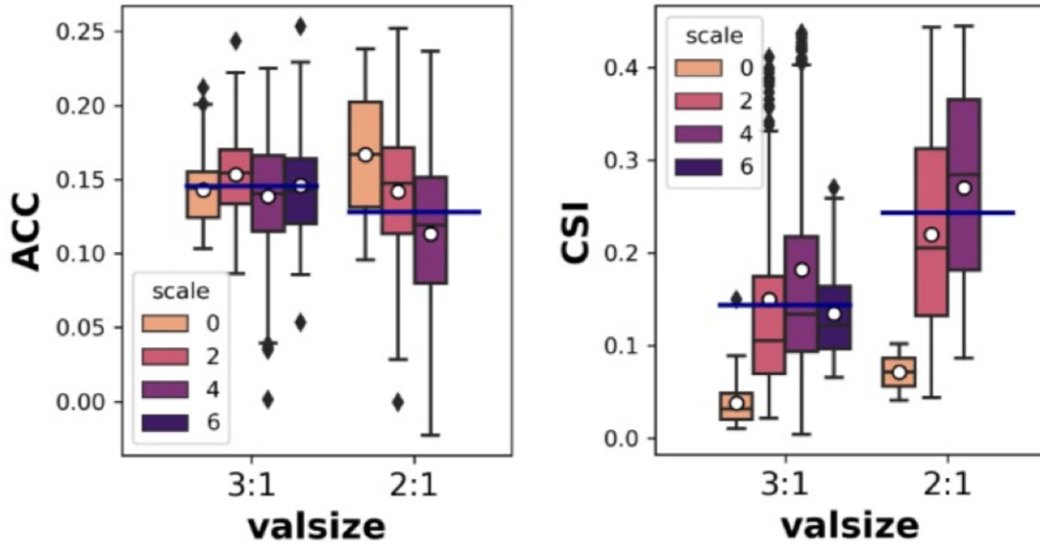


Figure 3.49 Similar as in Figure 3.48 but for two groups with different train/validation sample fraction (3:1 vs 2:1) for x-axis. Blue horizontal lines indicate the overall mean across all samples, independent of the loss function weight scale.

3.3.2. UNet 구조 변형 실험을 통한 모델 민감도 평가

본 연구에서는 훈련 자료의 공간적 시그널을 효과적으로 추출하여 동아시아 해역 고수온 현상으로 펼쳐내는 UNet 기반 모델을 고안하였다. 기본 UNet은 2차원 합성곱을 통해 지리적으로 인접한 격자들의 시그널을 유연하게 연결함으로써 전체적인 패턴을 학습하는 구조를 가진다. 그러나 UNet을 포함한 대부분의 기계학습·딥러닝 모형은 시·공간 이상치 탐지에서 blurring 현상이라는 한계를 지닌다.

이를 극복하기 위해 두 가지 아키텍처 변형을 적용하여 민감도를 평가하였다(Figure 3.50). 첫째, attention gate를 추가한 attUNet, 둘째, encoder path에서 2차원 합성곱을 swin-transformer 블록으로 대체한 swinUNet이다. 실험 결과, attention gate 추가는 성능에 큰 차이를 주지 않았으며 오히려 CSI가 감소하였다. 반면 swin-transformer 블록을 적용한 경우 CSI 평균이 유의미하게 개선되었다. 다만 ACC는 다소 감소하였는데, 이는 기존 UNet이 공간 패턴을 학습하는 방식에 강점을 지니고 있어 전체적인 패턴 재현 성능이 높게 나타난 것으로 해석할 수 있다. Figure 3.50ab에서 초록색 가로선은 ECMWF 역학 모형의 ACC와 CSI 수치를 각각 나타낸다. UNet 기반 모델을 활용한 동아시아 고수온 현상 예측에서 전체적인 패턴 재현 성능은 역학 모형에 크게 못 미쳐 아직 성능 격차가 존재한다. 그러나 swinUNet의 일부 훈련 모델은 고수온 탐지 성능에서 역학 모형을 초과하는 결과를 보여, 역학 모형을 극복할 가능성

이 가시권에 있음을 확인할 수 있었다.

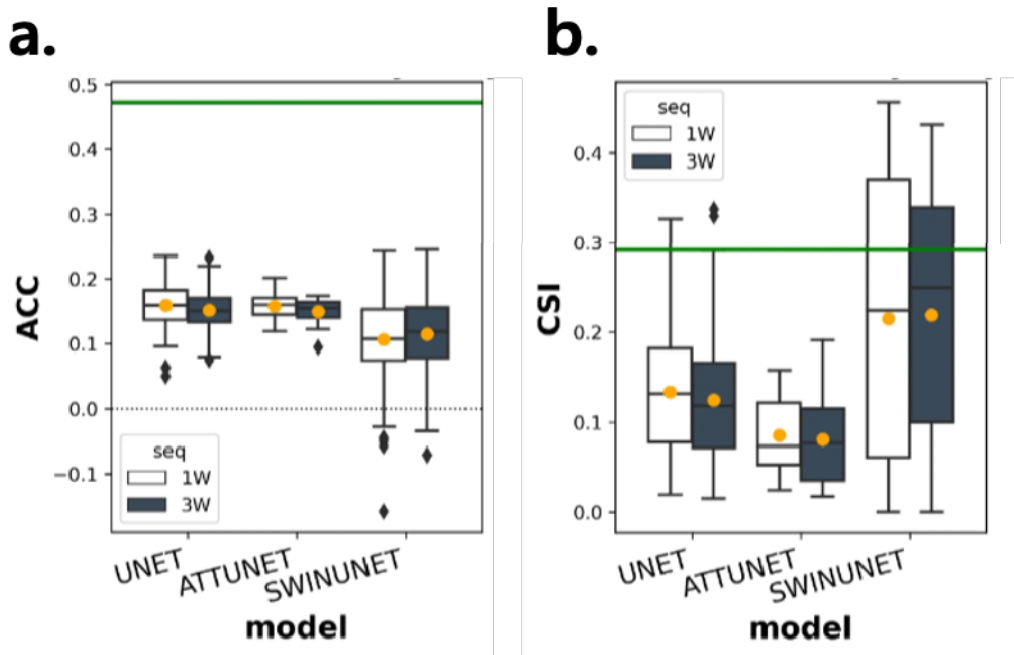


Figure 3.50 Boxplots showing the distribution of (a) ACC and (b) CSI across three models (Base UNet, Attention UNet, and Swin-Transformer UNet) along the x-axis. Colors indicate the temporal sequence of input data (white boxes for one-week inputs and navy boxes for three-week inputs). The green horizontal line represents the ECMWF marine heatwave forecast skill at a 3-week lead time.

3.3.3. 입력 데이터 민감도 평가

본 절에서는 앞서 2.3.3.2절에서 제시한 7가지 민감도 실험 모델 전체를 대상으로 훈련에 사용되는 입력 데이터 조합에 따른 비교 결과를 제시한다.

3.3.3.1. 개별 지구 시스템 훈련 정보에 대한 민감도 평가

앞서 Figure 2.15에서 제시한 각 지구 시스템의 관련 변수들(Table 2.9)을 제외하여 구성된 훈련장(woO, woF, woS, woU, woW)을 모델에 입력하고 그 결과를 해석함으로써, 각 시스템이 동아시아 해역 고수온 예측에 기여하는 정도를 평가하였다. 모든 변수를 포함해 학습한 모델(all)을 기초선으로 삼았으며, 성능이 기초선(평균값: 진한 붉은 실선, 중간값: 연한 붉은 실선)보다 낮아진 경우는 해당 지구 시스템의 정보가 중요한 기여를 한다고 판단하였다. 반대로 성능이 기초선보다 높아진 경우는 해당 지구 시스템의 정보가 오히려 성능을 저해하는 것으로 해석할 수 있으며, 이러한 경우 해당 변수를 최종 모델 훈련 자료에서 제외할 수 있다.

UNET 모형 실험에서는 ACC와 CSI 모두에서 woO, 그 다음으로 woS의 성능 저하가 뚜렷하게 나타나(Figure 3.51ab), 고수온 자체(hot ocean itself)와 정적 변수 관련 지구 시스템이 예측

성능에 크게 기여하는 것으로 판단된다. 반대로 woF, woW, woU 실험에서는 성능이 오히려 향상되어, 이들 지구 시스템을 포함할 경우 성능이 떨어질 수 있음을 시사한다.

swinUNet 모델 실험에서는 두 메트릭에서 결과가 다소 상이하게 나타났다(Figure 3.51ab). ACC에서는 woO, CSI에서는 woF의 성능 저하가 뚜렷하여 두 지구 시스템 정보가 고수온 예측에 기여하는 것으로 해석된다. 이 두 경우를 제외하면 변수 제외 결과가 all 케이스와 큰 차이를 보이지 않았으며, 전반적으로 소폭의 성능 기여가 확인되는 등 swinUNet 모델은 UNet에 비해 변수 민감도가 크지 않은 것으로 판단된다.

이러한 결과를 바탕으로, 고수온 자체(O), 플럭스(F), 정적 변수(S)를 조합한 새로운 훈련장을 테스트한 결과는 Figure 3.52에 제시되어 있다. Table 2.9에서 정의한 최소 기본 변수(min), 기본 변수에 시간 변수를 추가한 조합(min_wTime), 모든 변수를 포함한 조합(all), 기본 변수(min)에 플럭스와 고수온 자체 변수를 결합한 조합(wFO), 그리고 기본 변수(min)에 플럭스·고수온·정적 변수를 함께 결합한 조합(wFOS의 성능 분포를 비교하였다.

min 조합을 기초선으로 설정했을 때, 다른 변수 조합들의 CSI 평균 및 중간값은 UNet과 swinUNet 모두에서 점진적으로 향상되는 경향을 보였다. 특히 wFO와 wFOS 조합의 훈련 결과가 두 모델에서 가장 높은 수준의 성능을 나타냈다. 이는 앞서 개별 지구 시스템 변수를 제외한 실험들의 평가 결과와 일관되며, 플럭스·정적 변수·고수온 해양 정보를 함께 포함하는 것이 동아시아 해역 고수온 예측에 최적의 변수 조합임을 시사한다.

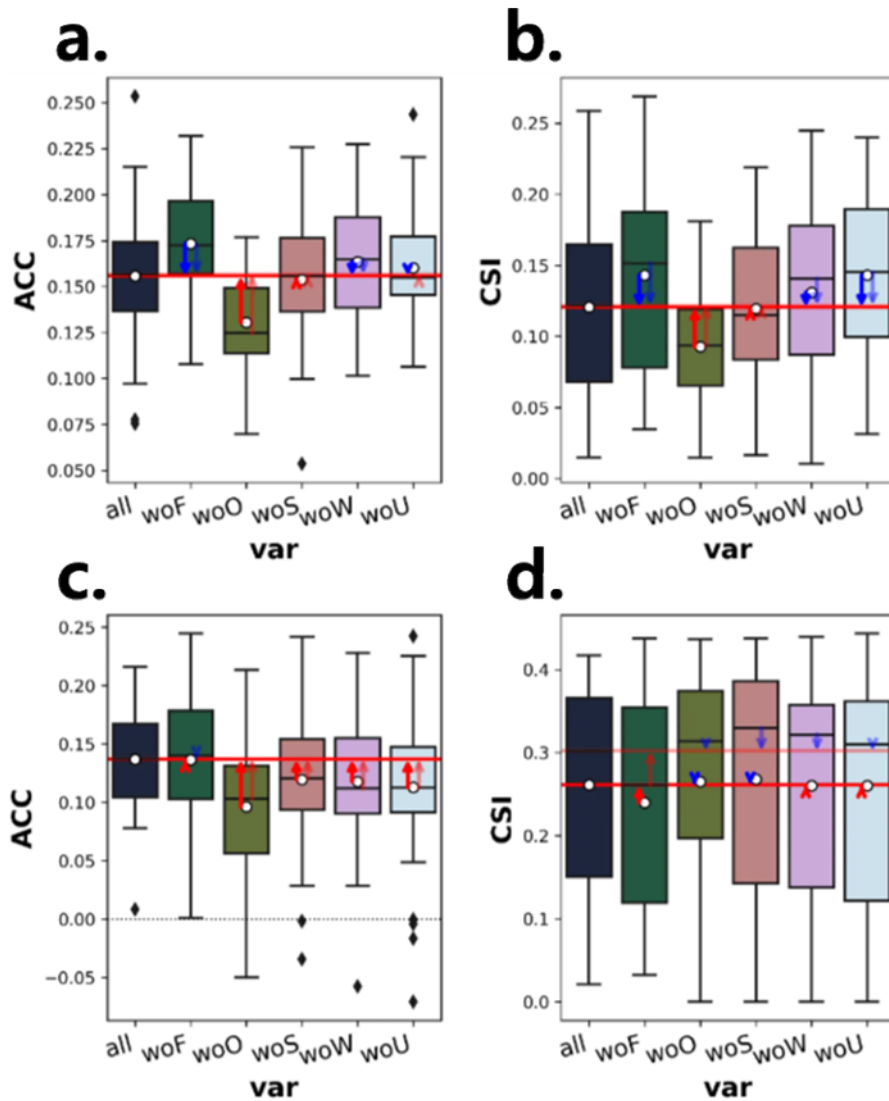


Figure 3.51 Boxplots showing the distribution of (a, c) ACC and (b, d) CSI across six input-variable combinations (all, woF, woO, woS, woW, and woU; see Table X.X). The red horizontal line indicates the mean skill of models trained with all input variables. Reddish and blue arrows indicate relative performance of models trained with each reduced input set (without Flux, Ocean, Static, Wave, or upper-atmosphere variables); upward arrows mark lower skill and downward arrows higher skill relative to the all-input baseline, highlighting the positive or negative contribution of each component to model performance.

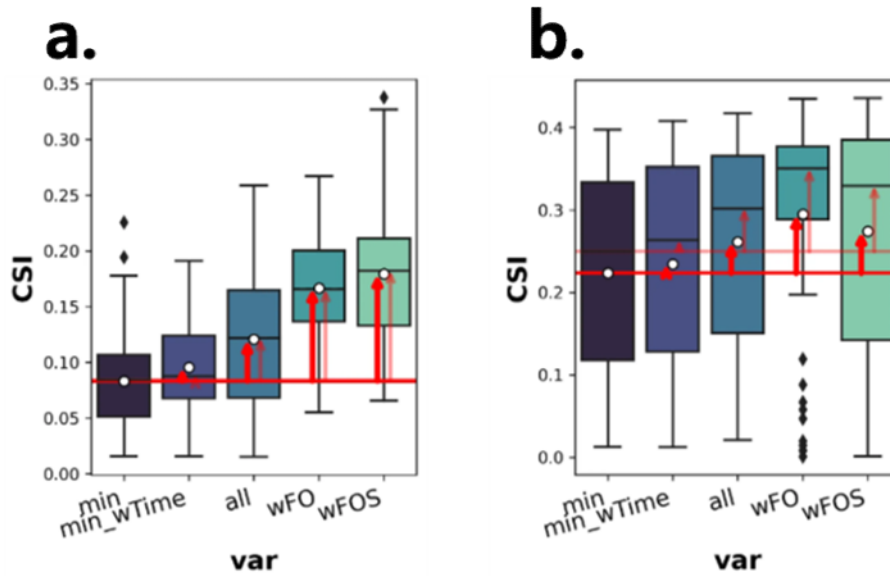


Figure 3.52 Similar to Figure 3.51, but focusing only on the CSI results for (a) UNet and (b) swinUNet across five different input-variable combinations. Opaque and transparent red horizontal lines represent the mean and median skill of models trained with the minimum (baseline) input variables. Opaque and transparent arrows indicate the mean and median skill improvements relative to this baseline.

3.3.3.2. multi-modal inputs에 대한 고수온 예측 개선 가능성 평가

앞서 2.3.3.2절에서 제시한 이질 자료 추가 실험 결과는 Figure 3.53에 제시하였다. 기후지수(CM only)와 하천유량(RD only)을 각각 추가한 경우, 두 정보를 동시에 추가한 경우(Both), 그리고 이질 자료를 추가하지 않은 모델(None)들을 구분하여 박스플롯을 통해 성능 분포를 시각적으로 검토하고 전반적인 성능을 비교하였다.

UNet 모델에서는(Figure 3.53ac) 두 이질 자료를 각각 추가한 경우 ACC와 CSI 성능이 모두 개선되었으며, ACC에서는 RD의 기여가, CSI에서는 CM의 기여가 상대적으로 더 크게 나타났다. 그러나 두 자료를 동시에 추가한 경우에는 오히려 성능이 저하되는 결과가 확인되었다.

이를 바탕으로 swinUNet 모델에 RD를 추가한 실험을 수행한 결과(Figure 3.53bd), ACC 성능은 다소 감소했으나 CSI 평균은 ECMWF 역학 예측을 유의미하게 상회하는 것으로 나타났다. 이러한 결과는 swinUNet에서 하천유량 정보를 처리하는 층을 추가하고, 추출된 정보를 마지막 레이어에서 효과적으로 통합함으로써 동아시아 고수온 탐지력을 향상시킬 수 있음을 시사한다. 특히 하천유량 정보의 대부분은 중국 동쪽을 흐르는 강들에 대한 것으로 황해와 동중국해로 흘러드는 담수 정보가 해당 해역의 고수온 탐지를 개선하는 데 기여하여 전체적인 성능을 높이는 것으로 판단된다.

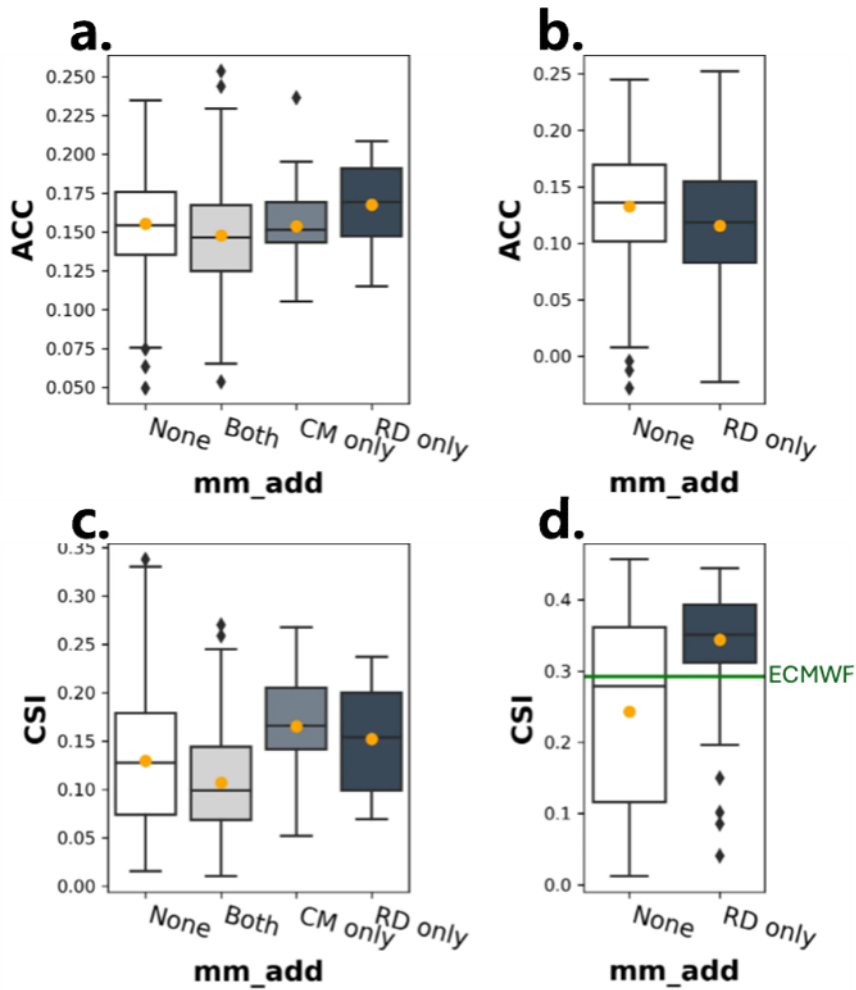


Figure 3.53 Boxplots showing the distribution of ACC (upper panels) and CSI (lower panels) across four additional-input conditions (None, Both, Climate Modes only, River Discharge only; see Table X.X). Panels (a, c) show UNet models, while (b, d) show swinUNet models (latter tested only with/without River Discharge). The green horizontal line denotes ECMWF marine heatwave forecast skill at a 3-week lead time (the same as in Fig. 3.50b).

3.3.4. 고수온 예측 원형 모형 도출 및 성능 평가

3.3.4.1. 민감도 실험 종합 평가 및 최적 모델 setup 도출

동아시아 해역 고수온 예측을 위한 최적 모델 setup을 도출하기 위해, 앞 절에서 설명한 여러 민감도 실험 조건의 모델들을 ACC와 CSI의 2차원 산포도 상에서 시각적으로 비교하였다 (Figure 3.54a). UNet과 swinUNet 모델 아키텍처를 비교한 결과, swinUNet은 전체적으로 우측에 분포하여 CSI가 전반적으로 높고 일부는 ACC 또한 높은 성능을 보였다. 반면 UNet은 대부분 좌측 상단에 분포하여 CSI 성능이 상대적으로 저조함을 확인할 수 있었다.

입력 도메인 크기/해상도 조건(L5 vs S2.5)을 비교했을 때, 두 조건은 유사한 분포를 보였으나 S2.5 모델들이 상대적으로 더 많이 우측 상단에 위치하였다. 훈련/검증 샘플 비율(3:1 vs 2:1) 비교에서는 검증 샘플 비율이 높을수록(2:1) 우측 상단에 위치하는 모델 수가 많았다. 가중 손실 함수 조건에서는 가중치 규모가 클수록 CSI가 높게 나타났으며, 특히 가중치 규모 2에서는 CSI와 ACC 모두 높은 성능을 보여 다수의 모델이 우측 상단에 분포하였다. 이질 자료 추가 여부와 관련해서는 하천유량 정보를 포함한 경우 CSI 값이 높게 나타났으며, 입력 자료 sequence(1W vs 3W) 조건에서는 시각적으로 큰 차이가 확인되지 않았다.

고수온 예측 개선 조건을 정량적으로 추론하기 위해, 산포도에서 ACC와 CSI가 모두 높은 우측 상단 구역에 분포하는 훈련 모델의 비율을 계산하였다(Figure 3.54b). 그 결과 swinUNet 아키텍처와 하천유량 이질 자료 추가 조건에서 월등한 성능 개선이 확인되어, 두 조건은 동아시아 해역 고수온 예측의 필수 요소로 판단된다. 그 외에 상대적으로 높은 해상도의 지역 도메인(S2.5), 검증 샘플 비율 2:1, 가중 손실 함수 가중치 규모 2에서 유의미한 성능 차이가 나타났다. 반면 입력 자료 sequence에서는 정량적으로도 차이가 없어, 메모리 사용이 증가하는 3W 입력의 가치는 크지 않은 것으로 판단된다.

종합하면, 2.5도 해상도의 지역 도메인 훈련 자료를 2:1의 훈련/검증 샘플로 나누고, 가중 손실 함수의 가중치 규모를 2로 설정하며, 초기 훈련 자료에 더해 하천유량 정보를 중간 층에서 추가적으로 추출하는 구조의 swinUNet 모델을 활용하는 것이 선행 3주 동아시아 해역 고수온 예측에 있어 본 연구 결과에 기반한 최적의 모델 setup이라고 판단된다. 훈련 변수 조합은 조건이 다양해 동일한 방식의 분석을 수행하지는 않았으나, Figure 3.52b에서 제시한 바와 같이 플릭스·정적 변수·고수온 해양 정보를 함께 포함하는 것이 최적의 조합으로 판단된다.

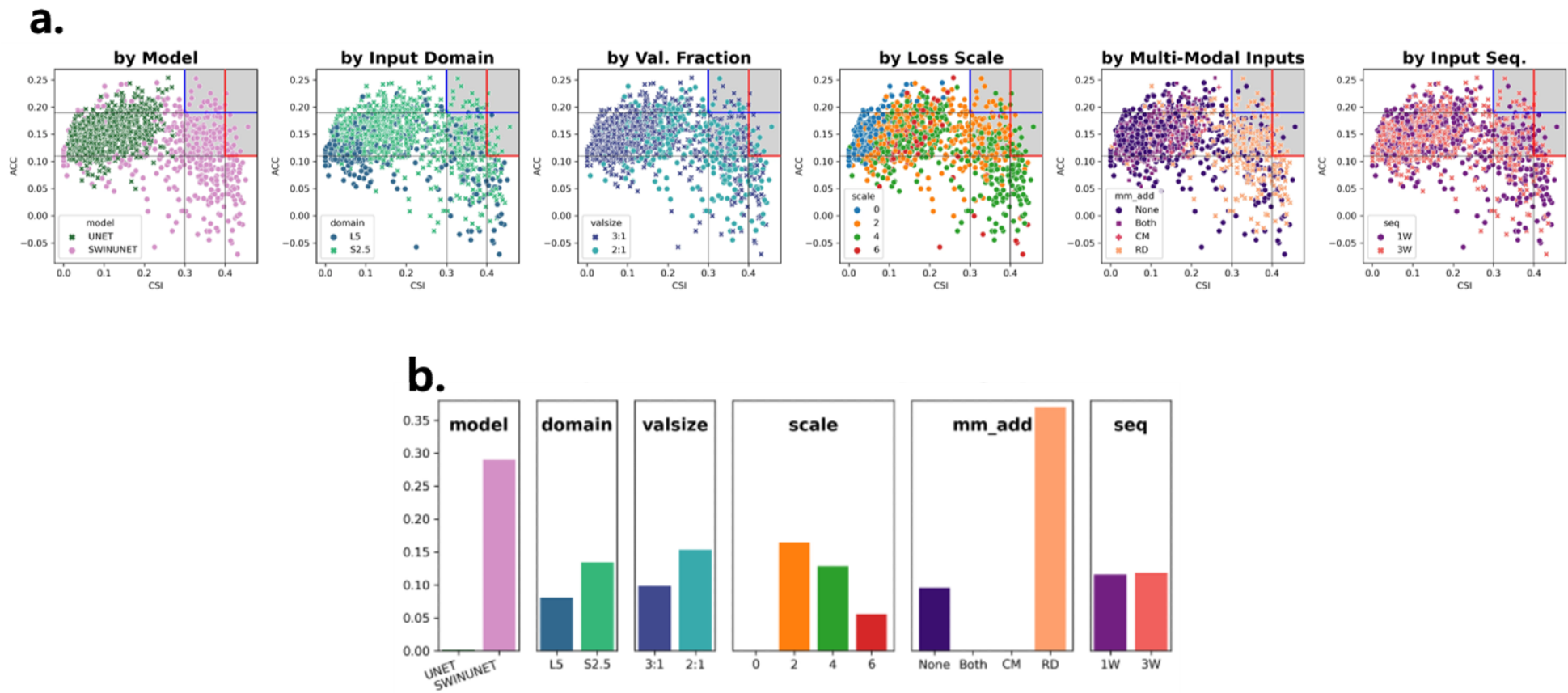


Figure 3.54 (a) Scatter plots of ACC versus CSI for all trained models, organized by model architecture, input domain size/resolution, validation sample proportion, loss scale, multi-modal inputs, and temporal input sequence (from left to right). (b) The fraction of good models, defined as those falling within the top-right gray box, relative to all ensembles in each category mentioned in (a).

3.3.4.2. ECMWF 선행 3주 예측과 비교

계절 내 시간 규모에서 가장 우수한 성능을 보이는 ECMWF 역학 예측과의 비교를 통해, 동아시아 해역 고수온 예측 모델에 인공지능 기술을 적용하는 정당성을 확인할 수 있다. ECMWF 실시간 예측은 101개 앙상블을 기반으로 일일 예측(daily forecast)을 제공한다. 그러나 역학 예측은 이상치 자체가 직접적인 목적 변수가 아니므로 후처리가 필요하다. 이를 위해 ① 선행 3주(15~21일) 구간에 대한 90% 분위 기대값을 hindcast set으로부터 산출하고, ② 이 기준을 초과하는 이상치 누적값을 각 앙상블에 대해 계산하며, ③ 마지막으로 개별 앙상블과 앙상블 평균에 대해 고수온 관측값 대비 ACC와 CSI를 평가하였다.

Figure 3.55은 2023년 52주 전체에 대한 ECMWF 선행 3주 예측 검증 결과를 도식화한 것이다. 좌측 상단의 세 패널은 격자별 시계열 기반 ACC, RMSE, CSI를 보여주며 평균값은 각각 0.41, 8.08, 0.35이다. 우측 상단은 독도 인근 격자(37.5° N, 130.5° E, 지도 상 pink open square)에 해당하는 Impact Factor 시계열이다. OISST 시계열에서는 봄철 고수온 현상이 뚜렷하게 나타나지만, ECMWF 앙상블 평균은 8월 이전까지 이를 전혀 예측하지 못한다. 8월 중반 이후에는 고수온 현상을 적절히 모의하나, 가을 이후에는 다시 음의 값을 예측한다. 주별 고수온 공간 패턴에 대한 예측 성능은 하단 좌측 시계열에서 확인할 수 있으며, 전체 평균은 ACC=0.41, CSI=0.24이고, 주별 변동성이 커 불안정하다. 특히, 2023년 7월 30일 주간의 2차원 맵에서는 황해와 동해의 고수온 현상을 ECMWF가 전혀 예측하지 못한 것으로 나타났다.

앞 절에서 제시한 동아시아 고수온 예측 최적 setup에 해당하는 두 모델의 2023년 52주 성능을 도식화하여(Figure 3.56) ECMWF와 비교하였다. 상단 모델은 3Wseq/S2.5/wFOS/RD_added/2:1 train/validation ratio 조건의 swinUNet 훈련 결과이다. 격자별 temporal ACC/CSI(0.33/0.30)는 ECMWF(0.41/0.35)보다 다소 낮지만, RMSE(7.0 vs 8.08)는 개선되었다. 특히 독도 인근 격자 시계열에서는 봄철 고수온 현상을 예측하였고, CSI도 개선되었다(0.47 vs 0.23). 공간적 CSI의 전체 52주 평균값(0.36) 역시 ECMWF(0.24) 대비 향상되었으며, 7월 30일 주간의 동해 고수온 현상도 적절히 예측되었다. 하단 모델은 train/validation ratio만 3:1로 변경한 결과로, 격자별 RMSE와 CSI가 ECMWF 대비 개선되었고, 공간 CSI의 52주 평균도 0.42로 크게 향상되었다. 특히 7월 30일 주간의 황해 고수온 현상도 적절히 모의된 것으로 확인되었다.

개별 인공지능 모델 결과는 ECMWF와 비교했을 때 앙상블이 없다는 점에서 차이가 두드러진다. ECMWF 앙상블 평균 시계열과 달리, 개별 모델은 zigzagging하며 불안정한 특성을 보인다. 가능한 동일한 조건에 대한 평가를 위해, 앞서 도출한 최적 setup에 해당하는 모델들을 추출해 앙상블 평균을 산출하고 ECMWF와 비교하였다(Figure 3.57). 상단은 S2.5/2:1 ratio/weight scale=2/RD_added 조건의 swinUNet 모델들, 하단은 S2.5/1W sequence/weight scale=2/RD_added 조건의 swinUNet 모델들을 활용한 결과이다. 독도 인근 격자의 앙상블 평균 시계열은 zigzagging이 줄어들어 안정된 양상을 보였으나, ECMWF 앙상블에 비해 spread가 커 평균값의 변동성이 크게 줄어들었다. 그럼에도 불구하고 봄철 고수온 현상을 예측하는 등 CSI는 두 경우 모두 ECMWF를 상회하였다. 또한 두 케이스 모두 2023년 7월 30일 주간의 동해 고수온 현상을 적절히 예측하였다. 다만 황해의 고수온 예측은 실패하였고, 앙상블들의 분산이

커서 앙상블 평균 예측의 절대값 크기가 작다는 한계가 있다.

종합하면, 동아시아 해역 고수온 예측에서 최적 setup을 적용한 개별 인공지능 모델은 2023년 테스트 기간 동안 ECMWF보다 높은 이상치 탐지 성능을 보여 원형 모델로서 적절하다고 판단된다. 또한 인공지능 모델 앙상블 예측은 2023년 52주 동안 공간 ACC와 CSI 지표에서 ECMWF 및 개별 인공지능 모델 대비 더 높은 시간적 안정성을 보여 강점으로 평가된다. 그러나 훈련 자료에서 상상할 수 없을 정도로 극단적인 고수온 현상이 발생한 2024년의 경우, ECMWF 성능을 능가하기는 어려웠으며 데이터 기반 인공지능 모델이 극복하기 힘든 한계를 분명히 드러냈다.

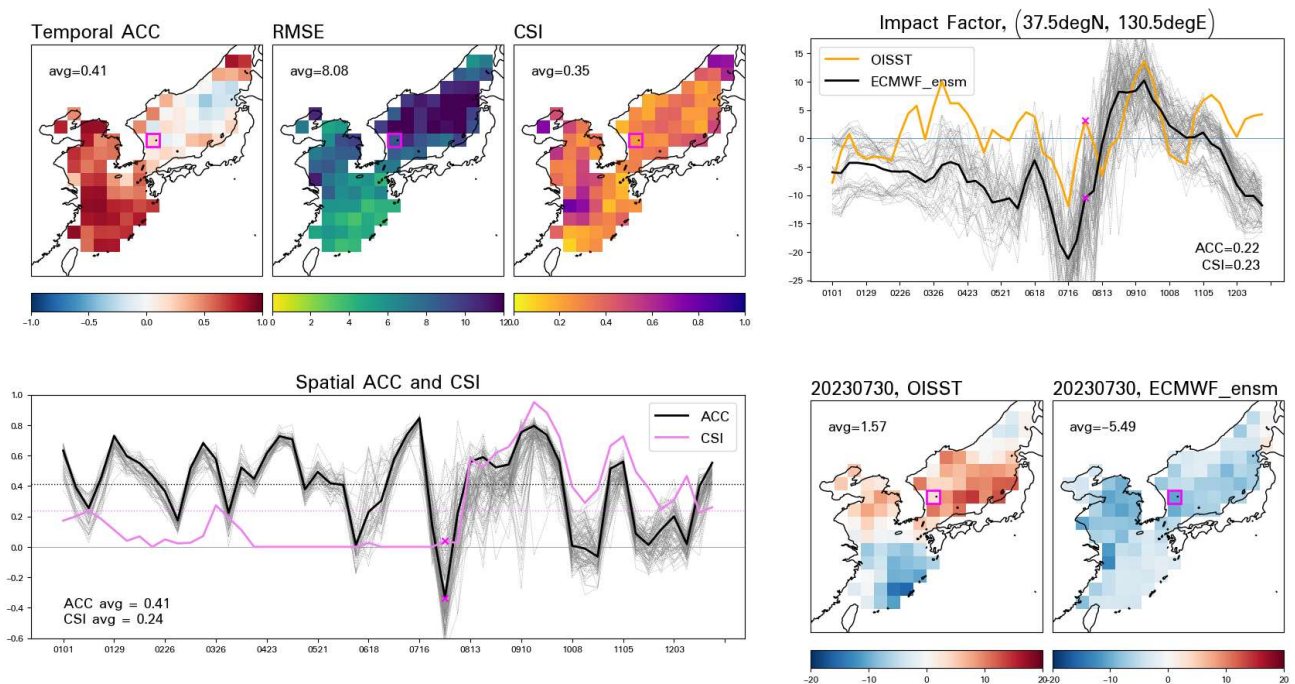


Figure 3.55 Weekly evaluation of ECMWF 3-week lead marine heatwave predictions across 2023. The top-left three panels present the temporal ACC (left), RMSE (center), and CSI (right) at each grid point. The bottom-left line plots show the spatial ACC (black) and CSI (magenta) over EAMS for each week. The top-right line plots illustrate the temporal variation of the weekly Impact Factor (weekly extremeness) for the grid point nearest to Dok-do (37.5° N, 130.5° E), based on observations (OISST, yellow) and ECMWF forecasts (gray for individual ensemble members and thick black for their ensemble mean). The bottom-right two panels depict the spatial distribution of the Impact Factor for the marine heatwave event during the week beginning 30 July 2023, comparing observations (OISST, left) with the ECMWF ensemble mean forecast (right).

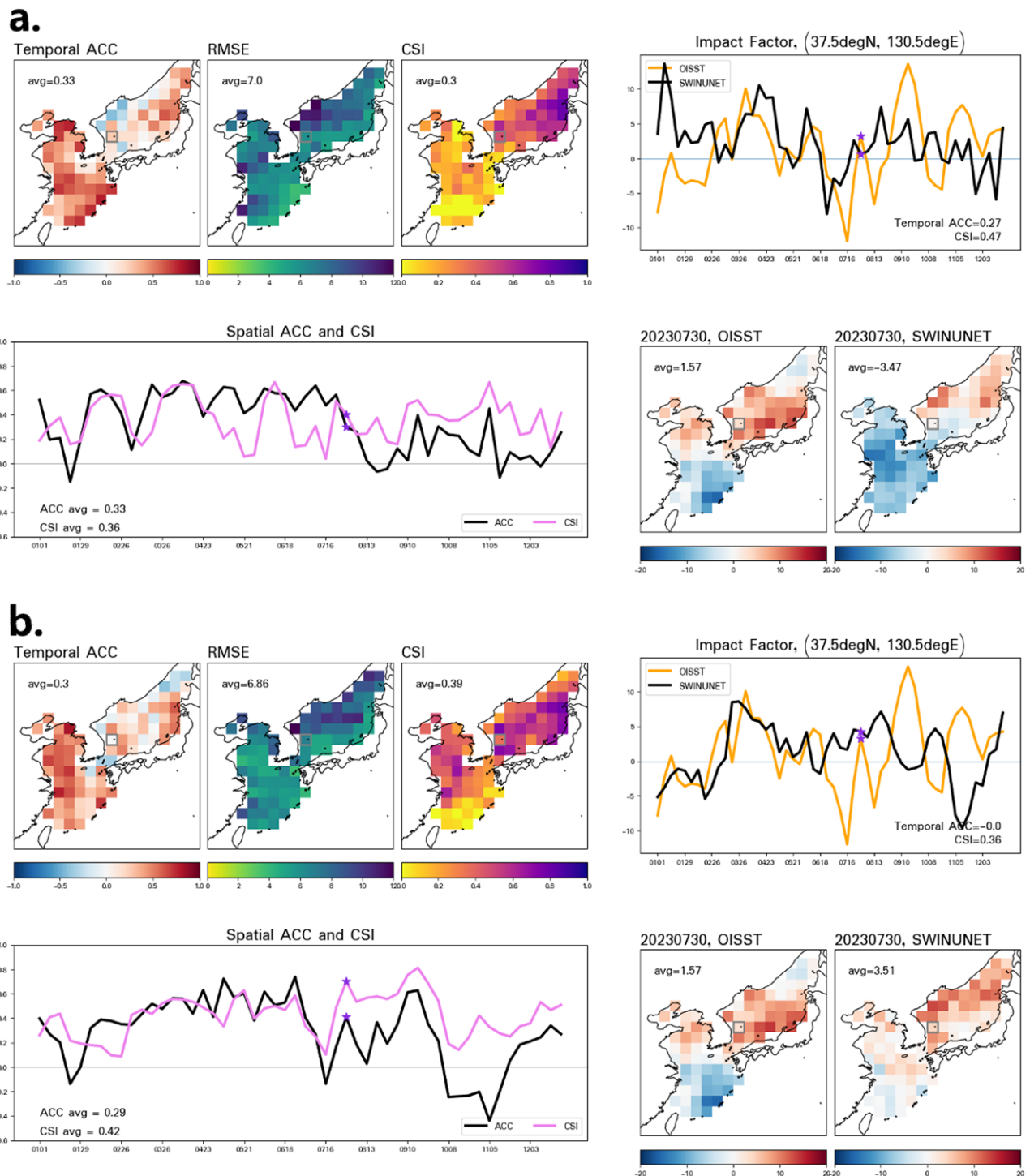
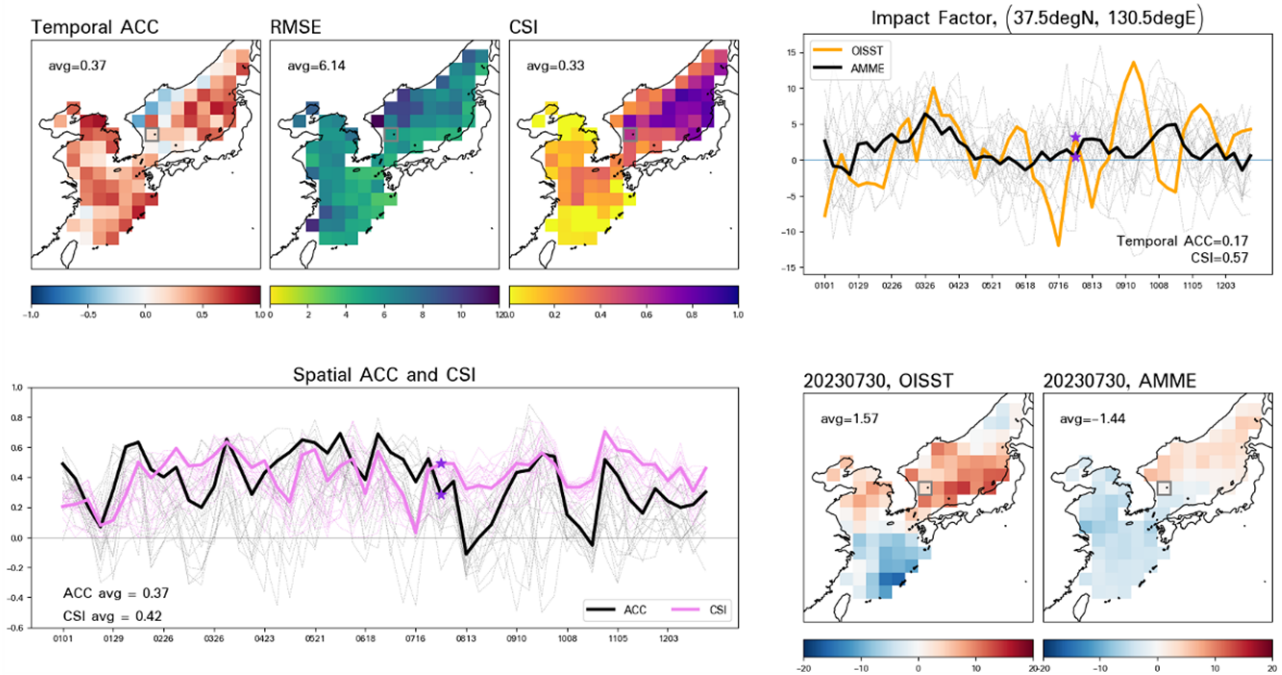


Figure 3.56 Performance of the two best AI models: weekly evaluation of 3-week lead marine heatwave predictions throughout 2023. (a) shows a swinUNet model with the 3Wseq/S2.5/wFOS/RD_added/2:1 train/validation ratio setup, while (b) presents a model with the same configuration but using a 3:1 train/validation ratio. Details are the same as in Figure 3.55.

a.



b.

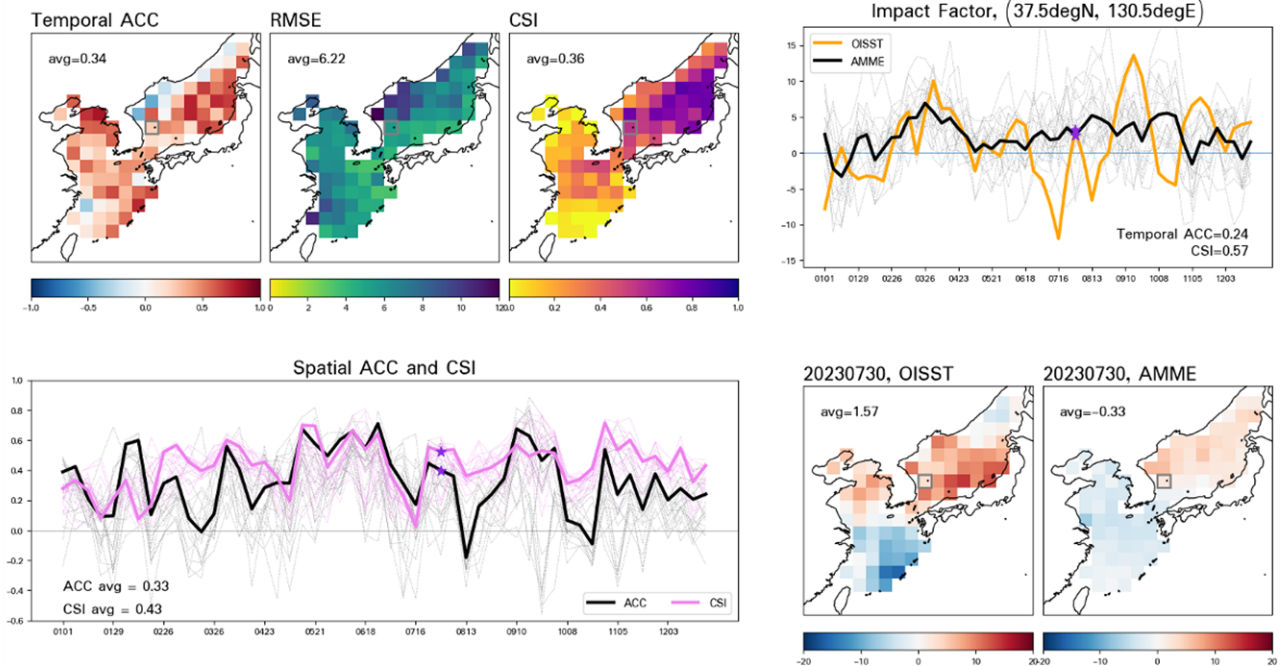


Figure 3.57 Same as in Figure 3.56, but for AI model ensemble forecasts. Ensemble members in (a) are swinUNet models with the S2.5/2:1 ratio/weight scale=2/RD_added setup, while ensemble members in (b) are swinUNet models with the S2.5/1W sequence/weight scale=2/RD_added setup.

3.4. 동아시아 이상기후 주간 자료 온라인 공유 시스템 구축

3.4.1. 웹페이지 소개

“EastAsiaClimateExtremes” 는 동아시아 지역의 기후 극한 현상 정보를 통합적으로 제공하기 위해 구축된 데이터 및 코드 기반의 GitHub 플랫폼이다(Figures 3.58-63). 이 웹 페이지는 ERA5, OISST, ECMWF-hindcast 등 다양한 재분석 및 기후 역학모델의 예측 자료를 활용해 구축된 이상고온, 폭우, 해양열파 등의 이상기후 현상의 장기 시계열과 통계 정보를 체계적으로 정리한다. 페이지 구조는 README.md를 중심으로 구성되며, 데이터 설명과 함께 시각화 예제 코드를 포함하여 자료의 차원과 변수(예, t2m, sst 등)를 상세히 문서화함으로써 재현성을 강조하였다. 사용자는 공개된 폴더 구조를 따라 일/주 단위의 시계열, 극한 임계값, 이상 기후 이벤트 등 필요한 정보를 손쉽게 내려받을 수 있다. 또한 제공된 여러 형태의 분석 및 시각화 코드를 통해 기후 극한의 시공간적 변동성을 직관적으로 탐색할 수 있다(Table 3.10). 향후 단순한 자료 구축의 형태를 넘어 지속적인 업데이트를 통해 다양한 연구 분야로의 확장과 응용을 지향하는 플랫폼이다.

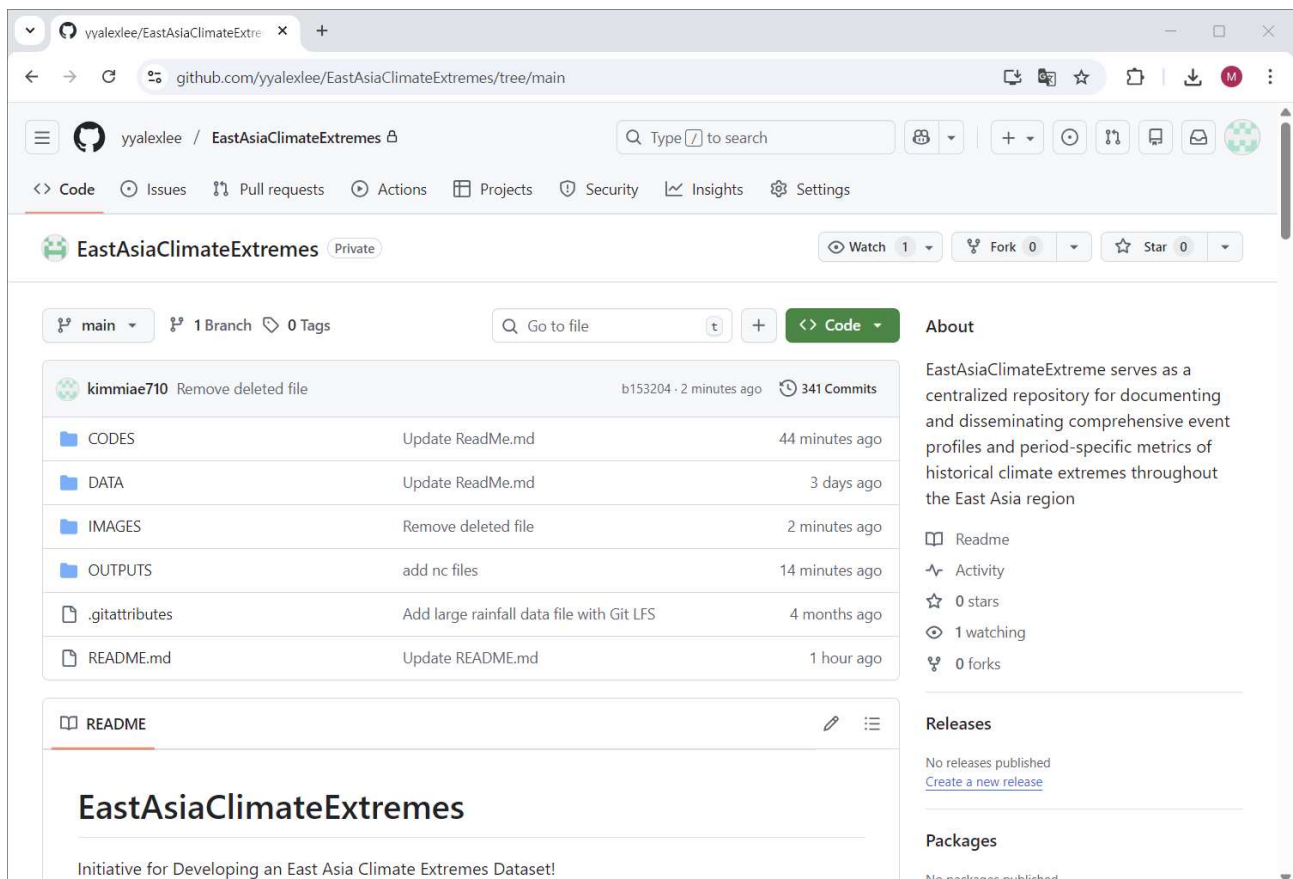


Figure 3.58. EastAsiaClimateExtremes GitHub main page screen shot 1.

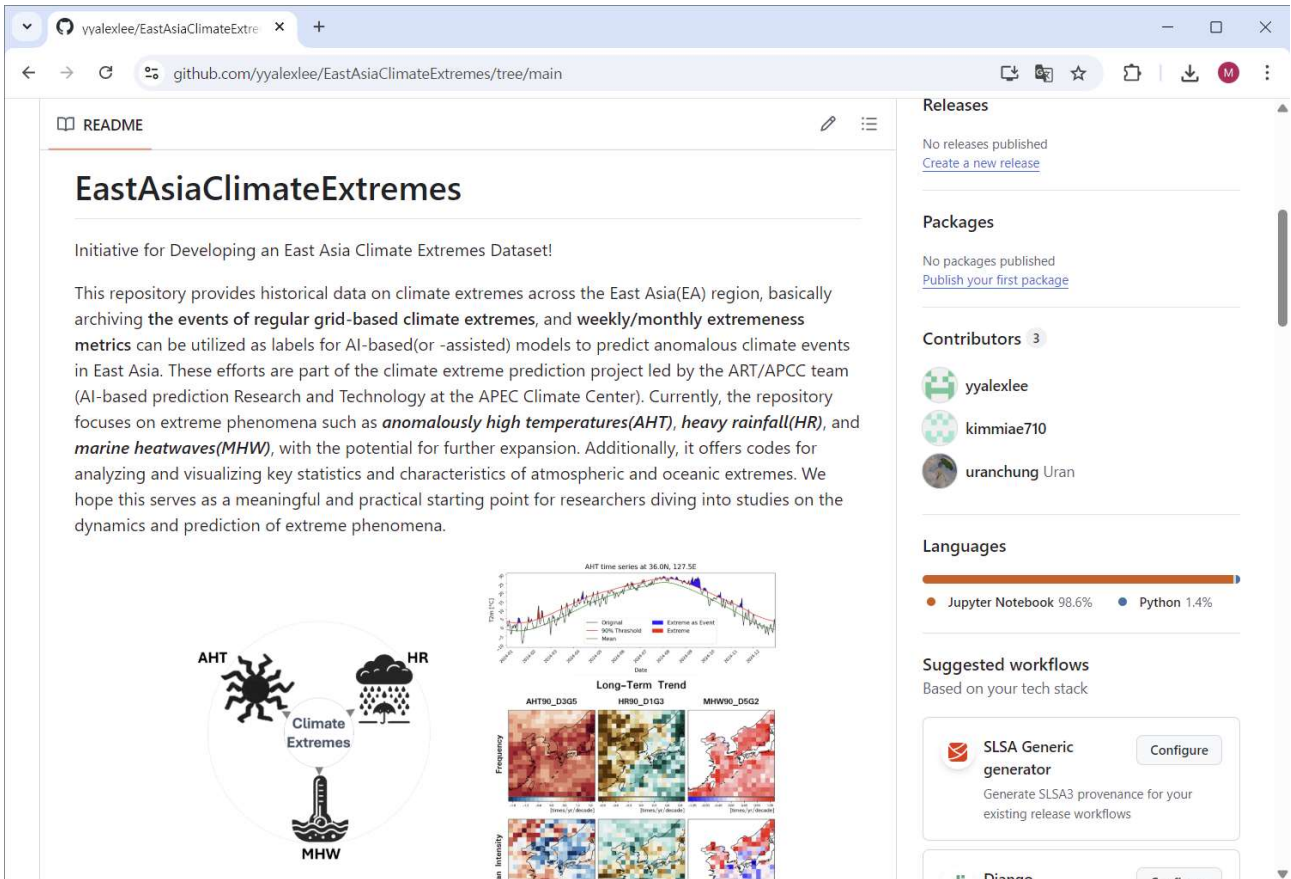


Figure 3.59. EastAsiaClimateExtremes GitHub main page screen shot 2.

Code File	Description	Location
<code>0.display_EAextremes_event_statistics.py</code>	Code for calculating and displaying long-term mean and trend of event statistics	EastAsiaClimateExtremes/CODES/
<code>TART_visualization1_output.ipynb</code>	Code for timeseries and heatmap visualization	EastAsiaClimateExtremes/CODES/
<code>TART_visualization2_output.ipynb</code>	Code for 2-D map visualization of long-term statistics of weekly climate extremes	EastAsiaClimateExtremes/CODES/

Table 3.10. Example codes for data processing and visualization (file name, description and location)

3.4.2. 이상기후 인벤토리 및 시각화 자료

“EastAsiaClimateExtremes” 는 극한기후 인벤토리를 구축하기 위해 각 이상기후 요소가 90백분위수를 초과하는 날을 사건(event)으로 정의하고 이를 지역 및 시간별로 정량화하였다. 평균기온(T2m)의 경우 각 연중 날짜별(day of year) 90백분위 기준값을 먼저 산정해 평년 기후

를 계산하고, 이후 특정 연도나 기간의 값이 이 기준을 초과하는지에 따라 극한 발생 여부를 판정함으로써 계절성과 지역 특성을 반영한 극한 지표를 제공한다. 레포지토리에는 이렇게 정의된 극한 발생 빈도와 강도를 지도 형태로 시각화하는 예제 코드가 포함되어 있으며 특정 연도의 기온 변동을 평년과 비교하여 이상고온 기간을 강조하는 시계열 그래프도 함께 제공된다 (Table 3.10). 예시 그림에서와 같이 사용자는 특정 연도와 격자 지점을 설정하여 이상 기후 이벤트 발생 경향이 시간에 따라 어떻게 나타나는지(Figures 3.60), 동아시아의 각 격자에서 90백분위를 초과하는 이상기후 이벤트(Figures 3.61) 및 극한 주(week)(Figures 3.62)가 어떠한 패턴으로 분포하는지를 직관적으로 확인할 수 있다. 이러한 이상 기후 통계치는 NetCDF/Numpy 파일로 저장 활용할 수 있다(Table 3.11).

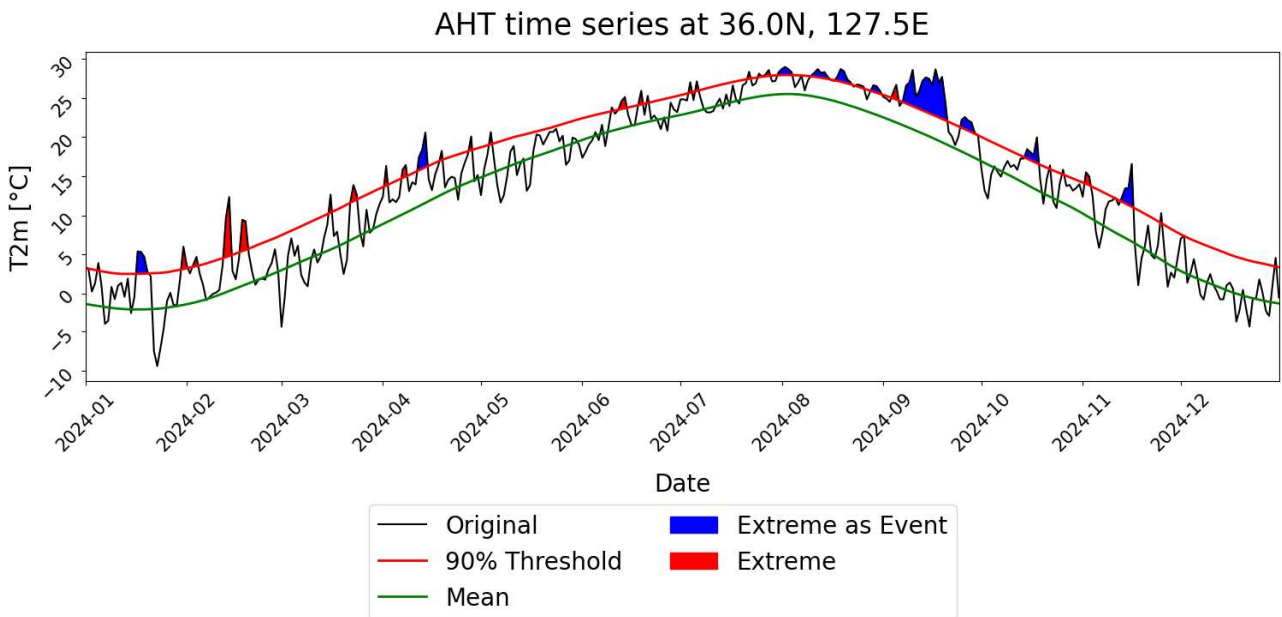


Figure 3.60. Timeseries of daily 2m air temperature at 36.0° N, 127.5° E in 2024, showing the original data (black), climatological mean (green), 90th percentile threshold (red), and anomalous hot days exceeding threshold line (red shaded) and extreme days identified as event periods (blue shaded).

MHW event90_D5G2

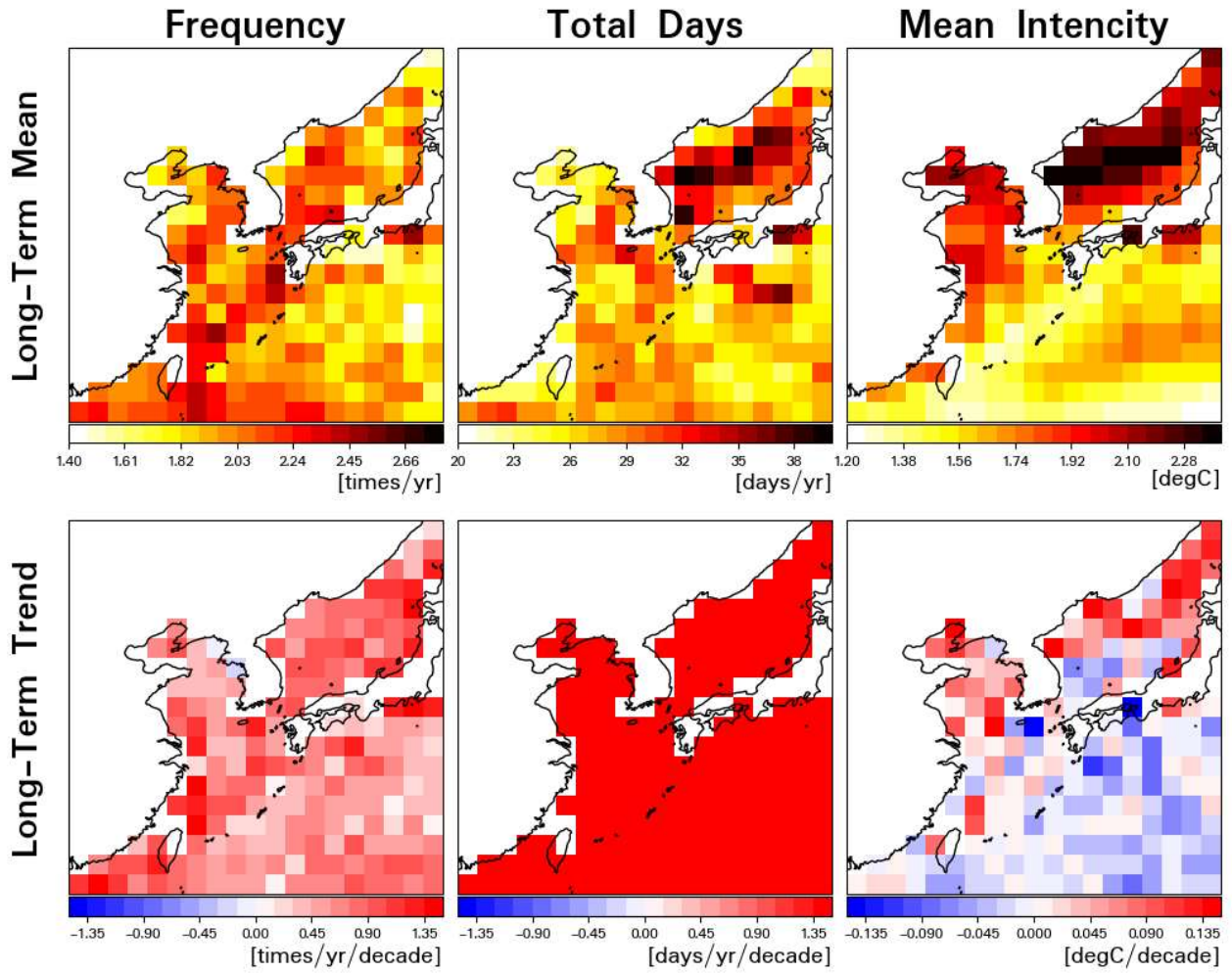


Figure 3.61 Spatial distribution of long-term mean and trend of marine heatwave event statistics: (left) frequency, (middle) total days, (right) mean intensity derived from daily SST timeseries over East Asia Marginal Sea.

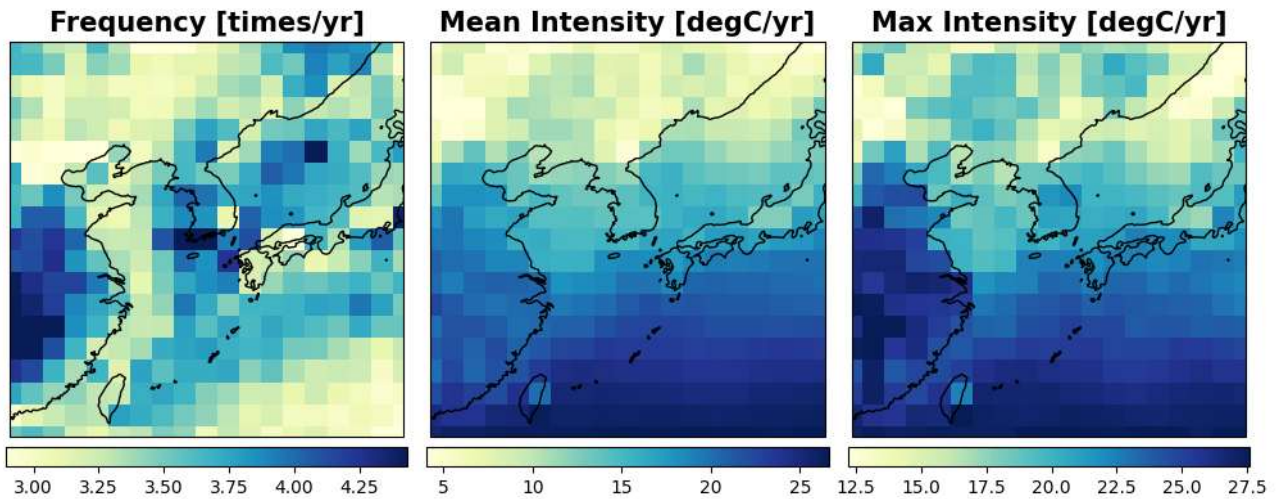


Figure 3.62 Spatial distribution of extreme weeks: annual mean statistics of anomalously hot weeks derived from weekly T2m timeseries over East Asia: (Left) frequency (times per year), (middle) mean intensity ($^{\circ}$ C per year), and (right) maximum intensity ($^{\circ}$ C per year).

ECMWF-S2S의 Hindcast 경우, 자료 1(주간 T2M, SST, PREC 예측)과 자료 2(90th percentile 평년값)를 결합한 자료로부터 동아시아 격자에서 90퍼센타일 초과 발생(예, 빈도: 초과일수, 강도: 초과 편차)을 계산하는데, 예를 들어 T2M_clim90th(doy, lat, lon)는 doy(1-365) 기준 percentile을 제공해 이상고온(예, TX90p) 탐지를 지원한다. 시각화 자료는 matplotlib 기반 지도 플롯으로, 버전 2016과 버전 2024간 예측 성능 차이를 빈도 및 강도 맵으로 표현 및 비교할 수 있다(Figure 3.63). NetCDF 파일(148MB t2m 변수 등)을 xarray로 로드하여 Cartopy 투영으로 동아시아 도메인을 렌더링하며 colormap 최적화로 극한 패턴을 직관적으로 보여준다. 인벤토리는 드롭박스(Dropbox) 다운로드도 가능하며 향후 메트릭(예, Brier Skill Score 등) 적용 예시를 통해 S2S 예측 향상 효과도 공유 예정이다.

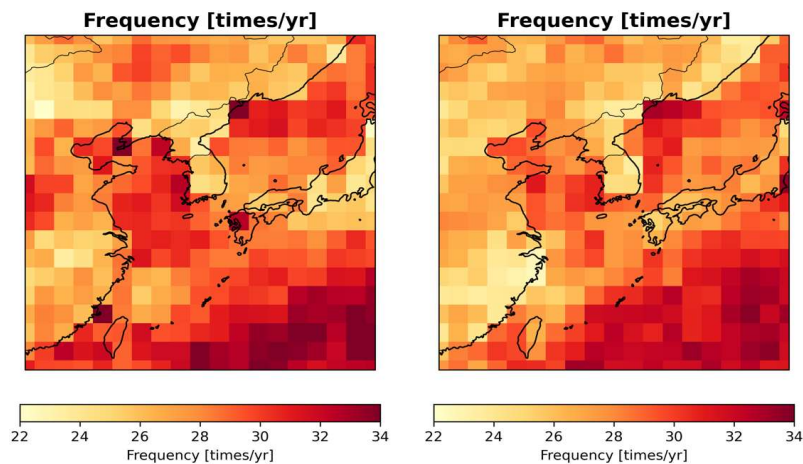


Figure 3.63 Spatial distribution of 90th-Percentile extreme temperature frequency based on ECMWF-S2S Versions 2016 and 2024.

Outputs	Visualization	Data Format
Daily based extreme events, Yearly event counts	Timeseries	NetCDF
Monthly Event Counts+Trend(+/-)	Heatmaps	NetCDF
Long-term Mean/Trend of Event Statistics	2D maps	Numpy
Count of Extreme Weeks, Mean/Max Intensity	2D maps	NetCDF

Table 3.11 List of output statistics generated and stored by the example codes.

3.4.3. 활용법 소개

“EastAsiaClimateExtremes”가 제공하는 자료와 코드는 기후 변화 연구, 극한 현상 모니터링, 예측 모델 검증 등 다양한 분야에서 활용될 수 있다. 연구자는 장기간에 걸친 이상기온 또는 폭우의 경향성을 분석하여 극단 사건의 발생 빈도 변화나 계절별 패턴을 파악할 수 있으며, 기후변동성 연구나 도시 및 농업 분야의 리스크 평가에도 적용할 수 있다. 더불어 S2S(subseasonal to seasonal) 예측 시스템 또는 기계학습 기반 예측 모델을 사용하는 경우, 이 레포지토리에 정리된 극한 인벤토리를 기준자료(reference)로 삼아 예측의 정확도와 일관성을 평가할 수 있다. 제공된 시각화 예제를 활용하면 다양한 문서에 사용할 수 있는 품질의 그래프를 쉽게 만들 수 있으며, 향후 분석 목적에 따라 극한 기준을 새롭게 정의하거나 다른 변수를 추가하는 확장도 가능하다. 이러한 점에서 EastAsiaClimateExtremes는 동아시아 극한기후 연구를 위한 강력한 출발점이자 응용할 수 있는 분석 플랫폼으로 기능할 수 있을 것이다. 이 레포지토리를 활용하려면 먼저 DATA 디렉토리의 README에서 제공하는 링크 또는 wget 명령을 통해 극한 이벤트 인벤토리, ERA5·OISST 기반 일별 및 주별 단위 데이터, 그리고 ECMWF S2S hindcast의 재구성된 예측 자료를 다운받을 수 있다. 데이터는 Pickle과 NetCDF 형식으로 저장되어 있어 Python 환경에서 즉시 활용할 수 있으며, 특히 xarray 기반의 처리에 최적화되어 있다. 다운로드한 자료를 바탕으로 사용자는 특정 지역 또는 격자에서의 극한 발생 빈도를 계산하거나, 연 혹은 계절별 강도와 지속성의 변화를 분석하고, p90 및 p95 임계값 대비 편차 기반의 extremeness 지수도 손쉽게 산출할 수 있다. 제공되는 시각화 코드를 실행하면 시계열 그래프, 열지도(heatmap), 그리고 장기 통계의 공간 분포가 자동으로 생성되며(Figures 3.64–65), 이 과정에서 극한 이벤트 특성을 요약한 NetCDF 출력 파일도 함께 저장된다(Table 3.11). 이러한 기능은 단순한 분석을 넘어, AI 기반 극한 예측 연구에서도 유용한데, 일별 및 주별 단위 extremeness 지수는 지도학습 레이블로 직접 활용할 수 있고, 예측 모델 평가 지표나 편차 보정 실험에도 적용할 수 있다.

kimmiae710 Fix code

e5b2a77 · 3 weeks ago History

Preview

Code

Blame

Raw



Timeseries and Heatmap Visualization for Extreme Events

Updated 23-Nov-2025

Team ART(AI-based prediction Research and Technology)/APCC(APEC Climate Center)

Contact Miae Kim (miaekim@apcc21.org)

In this notebook, you will display timeseries of extreme phenomena such as anomalously high temperatures(AHT), heavy rainfall(HR), and marine heatwaves(MHW) for a specific grid point in East Asia.

Requirements

"You can mount /content/drive/ to your Google Drive, where you will download data from the GitHub page(<https://github.com/yyalexlee/EastAsiaClimateExtremes/tree/main>)."

```
In [ ]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

Download data

Create FIRST directories if needed with the following line OR by yourself.

```
In [ ]: #!mkdir -p /content/drive/MyDrive/APCC/Data/
```

Figure 3.64. Screen shot of visualization code 1: Timeseries and Heatmap Visualization for Extreme Events.

kimmiae710 Fix code

21f5824 · 2 days ago History

Preview

Code

Blame

Raw



Weekly Extreme Analysis: Frequency, Mean, and Max Intensity

Updated 03-Dec-2025**Team** ART(AI-based prediction Research and Technology)/APCC(APEC Climate Center)**Contact** Miae Kim (miaekim@apcc21.org)

In this notebook, you will analyze weekly averaged or accumulated climate extremes from ERA5 over East Asia. You will calculate climatological weekly extreme statistics (frequency, mean intensity, max intensity) and visualize them spatially.

NOTE: The codes for calculating extreme statistics were written to handle the limited RAM constraints of Google Colab through use of Dask chunking.

Requirements

"You can mount /content/drive/ to your Google Drive, where you will download data from the GitHub page(<https://github.com/yyalexlee/EastAsiaClimateExtremes/tree/main>)."

```
In [ ]: from google.colab import drive
        drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

Set paths

CHANGE the following paths if needed!

Figure 3.65. Screen shot of visualization code 2: Weekly Extreme Analysis: Frequency, Mean, and Max Intensity.

4. 요약 및 결론

이 보고서는 세 가지 동아시아 이상기후 컴포넌트별 계절내 시간규모 인공지능 원형 모델 개발 과정을 소개하고 도출된 유의미한 결과를 제시한다. 본 연구의 초기 결과는 딥러닝 기반 계절내 예측의 가능성을 보여주며, 이상기후 조기 경보 시스템 개발의 기초 자료로 활용될 수 있겠다.

A. 첫 번째는 동아시아 이상고온 탐지에 있어 라벨링-모델 선택의 중요성을 강조한 연구로, ECMWF-S2S 기반 이상고온 예측에서 이진 라벨 구조가 다중 클래스보다 성능이 우수한 이유를 라벨 경계 불확실성 감소, 학습 안정성 향상, 희귀 극한 사례의 표본 부족 완화에서 찾고 있으며, Attention U-Net의 Attention Gate가 공간 구조적 패턴에 초점을 강화해 예측 패턴 및 BSS 향상에 기여함을 보여준다. 이러한 결과는 양쯔강 유역 및 전 지구적 극한고온을 다룬 최신 딥러닝 연구들과 비교해도 경쟁력 있는 성과로 평가되며, 동아시아 지역에서 S2S 확률 기반 이상고온 예측 보정 연구의 의의를 제시한다. 이는 향후 Sub-season 기후서비스 개선, 폭염 조기경보 향상, 산업별 기후 위험 관리 등에 활용될 수 있을 것으로 기대된다.

- (확률 비교) 이진 라벨 기반의 Att-U-Net 모델에서 관측(ERA5)와 유사한 패턴을 보임으로 고온 신호 탐지에 다중 클래스 라벨보다 안정적일 것으로 확인됨
- (모델 비교) Attention U-Net이 U-Net보다 0보다 큰 BSS의 비율이 높게 나타난 것은 Attention Gate가 공간적 중요 영역에 대한 탐지(feature selection)에 기여한 것으로 판단됨
- (BSS 평가) ECMWF-S2S 이상고온 예측에서 BSS가 약 0.2 수준으로 비교적 낮은 기술을 보였지만 딥러닝 기반 모델은 전반적 기술이 낮은 상황에서도 ECMWF 대비 특정 지역에서 BSS를 실질적으로 향상(+)-시킴
- 종합하면, 확률 공간 패턴 및 BSS 분석에서 이진 라벨과 Attention U-Net 조합이 가장 우수한 성능을 보임

B. 두 번째 이상기후 컴포넌트로 격자점 단위의 선행 3-4주 폭우를 예측하는 딥러닝 모델의 초기 모형을 개발하였다. 연구는 다음의 세 가지의 연구 질문을 중심으로 수행되었다: 1) Multi-task learning 적용이 폭우 예측 성능 향상에 미치는 영향, 2) 대기-해양 변동성을 반영하는 최적 변수 조합 탐색, 3) 모델 구조 선택이 예측력에 미치는 영향 분석이다.

- 먼저, Multi-task learning 기법은 강수량의 시간적 변동성을 보다 효과적으로 포착하여 극한 강수 탐지의 민감도를 향상시킴을 확인함. 단일 태스크 모델 대비 주요 검증 지표(CSI, TCC 등)에서 일관된 개선을 보였으며, 이는 모델이 강수량을 학습하는 과정에서 강수량 뿐만 아니라 극한강수일수를 함께 학습하는 것이 강수의 변동 특성을 더욱 정

교하게 반영한 결과로 해석됨.

- 두 번째로 예측 변수 조합 및 입력 변수의 스케일 분해 방식을 평가한 결과, 10-60일 대역의 변수를 포함한 경우 예측에 효과적이었으며, synoptic 규모의 성분은 계절내 극한 강수 예측에는 큰 기여를 하지 못함. 따라서 계절 내 예측에서는 ISO 대역과 장주기 대역 변수를 포함한 다중 스케일 입력 구성이 최적으로 판단됨.
- 마지막으로 기본 ResNet 구조에 LSTM 모듈을 결합한 형태가 시공간적 패턴의 연속성을 효과적으로 학습하는 것을 확인함. ResNet은 공간적 특징 추출에, LSTM은 시간적 특성 인식에 강점을 보여 두 구조의 결합이 시너지 효과를 낸 것으로 보임.

향후 연구에서는 Transformer 등 추가적인 구조 실험을 통해 보다 효율적이고 일반화 가능한 모델 탐색을 계획하고 있다. 또한, 확률 예측(probabilistic forecasting) 기법을 도입하여 예측의 불확실성을 정량화하고, 설명가능한 인공지능(XAI) 기법을 활용해 모델이 학습한 주요 원천 변수 및 공간 패턴을 체계적으로 분석할 예정이다.

C. 마지막 이상기후 컴포넌트로 해양 열파 예측 기술 개발을 진행하였다. 구체적으로, 인공지능 기반 선형 3주 동아시아 해역 고수온 예측에 적합한 인공지능 원형 모델 개발을 목표로 하였다. 이를 위해 동아시아를 넘어서는 광범위한 최신 해양·대기 관측 정보로부터 시그널을 추출하고, 이를 다시 동아시아 해역 도메인으로 확장하는 UNet 구조의 모델을 설계하였다. 고수온 탐지 성능, 즉 CSI를 모델 개선의 지표로 삼았으며, 이 과정에서 해양 열용량·염분, 대기 표면 바람·온도, 강수 등 해양·대기 기본 변수를 학습 자료로 활용하도록 구성하였다. 또한 훈련 변수뿐만 아니라 손실 함수와 모델 구조에 대한 다양한 민감도 실험을 수행하였으며, 도출된 선형 3주 동아시아 해역 고수온 예측 최적 setup은 다음과 같다.

- 2.5도 해상도의 동아시아 지역 도메인(70° - 147.5° E, 20° - 57.5° N)에 대해 기본 변수에 플릭스·정적 변수·고수온 해양 정보를 추가하여 훈련장을 구성하고, 훈련/검증 샘플 비율을 2:1로 설정했을 때 최적 성능을 확보함.
- 기본 손실 함수(MSE) 대신 실제 고수온 격자에 가중치를 부여하는 가중 손실 함수(weighted MSE)를 적용했을 때 성능이 개선됨. 특히 가중치 스케일을 2로 설정했을 때 고수온 탐지 성능이 유의미하게 향상되고, 고수온 패턴 재현성(ACC)의 감소도 크지 않음.
- 기본 UNet 모델의 encoder 경로에서 2차원 합성곱을 swin-transformer block으로 대체했을 때 고수온 탐지 성능이 현저히 향상됨.
- 2차원 재분석 격자 자료 외에 해양 열파의 강화 및 유지에 영향을 미치는 하천유량 정보를 독립된 레이어로 추가하여 최종적으로 합산하는 방식을 테스트한 결과, 이러한 이질 자료 추가가 성능 개선에 기여함을 확인함.
- 위의 조건하에 훈련된 모델들은 ECMWF를 넘어서거나 상응하는 고수온 탐지력을 보임.

다만 고수온 현상의 패턴 모의에 있어서는 인공지능 모델이 ECWMF를 넘어서지 못하는데 이는 해류나 해역간 상호작용 같은 특성을 학습하기 어려운 입력 조건으로 때문으로 해석할 수 있음. 올해 과제에서는 훈련 샘플을 늘리기 위해서 자료 길이가 다소 짧은 해양 정보보다는 대기 정보의 비중이 높여서 모델을 구성하였음. 향후에는 다소 자료 길이가 짧더라도 해양 깊이에 따른 수온/바람 입력 정보 비중을 늘림으로서 해류나 용승과 같은 해양정보를 좀더 포함할 수 있는 방안을 모색하고자 함.

보통의 인공지능 모델 개발 과정은 구조 테스트와 변수 조합 테스트를 분리해, 영향이 적은 쪽을 먼저 고정한 뒤 나머지를 세밀하게 튜닝하는 단계적 최적화 흐름을 겪는다. 즉, 다양한 민감도 테스트 결과를 비교해 최적 조건으로 수렴하는 방식을 택하게 된다. 여러 이상기후 컴포넌트 모델을 종합해야하는 과제 특성상 명확한 기준·tone·자동화 방식이 존재한다면 좀더 개발 과정을 객관화하고 목적에 부합하도록 효율성을 높일 수 있을 것이다. 향후 확률 예측 모델 전환 과정에서 올해 도출된 결과를 바탕으로 모델 최적화 절차/기준에 대해 팀내 논의가 필요할 것으로 보인다.

EastAsiaClimateExtremes는 동아시아 기후 극한현상을 체계적으로 정리한 데이터 및 코드 기반의 공개 저장소로, 극한 이벤트의 공간 및 시간적 변동성을 분석하고, 예측 모델 연구에 직접 활용할 수 있는 토대를 제공한다. ERA5, OISST, ECMWF-hindcast 자료를 통합하여 일별 및 주별 단위의 임계값(p90 및 p95)과 극한 이벤트 통계(빈도, 지속기간, 강도) 등을 구축하였으며, 일별 및 주별 단위 extremeness 지수는 AI 기반의 극한 예측 연구에서 레이블링 자료로 활용할 수 있도록 표준화하였다. 이와 같이 EastAsiaClimateExtremes는 동아시아 기후 극한현상을 연구 및 예측하는 데 필요한 표준화된 데이터 기반을 제공함으로써 다음과 같은 의의를 가진다.

- 극한현상의 장기적 변화 진단에 활용 가능한 종합 데이터베이스 구축: 동일한 기준으로 일관되게 산출된 이벤트 통계 등을 제공할 수 있으며, 기후변화에 따른 극한 빈도 및 지속성, 강도 평가 등에 활용 가능함
- AI 기반 기후 예측 연구의 라벨 제공 체계 확립: 일 밀 주 단위 extremeness 지수는 기계학습 및 딥러닝 모델의 출력 라벨로 최적화하여 제공하여 극한 예측 성능 개선 및 새로운 모델 개발에 활용할 수 있음
- 재현성 및 확장성 높은 분석 및 시각화 도구 제공: 코드 공개를 통해 누구나 동일한 결과를 재현할 수 있어 향후 다양한 극한 이벤트(폭염, 한파 등)로 확장 가능함

따라서 본 인벤토리는 동아시아 기후 극한 연구의 공공 인프라를 강화하고 기후변화 대응 기술 및 예측 기술 개발의 핵심 자원으로 활용될 것으로 기대된다.

References

- APEC기후센터, 2024: 2024년도 연차보고서: 객관적 기후예측에 적합한 인공지능활용 기반기술 개발(III). p38-43.
- Bahdanau, D., K. Cho, and Y. Bengio, 2016: Neural Machine Translation by Jointly Learning to Align and Translate. arXiv:1409.0473, <https://doi.org/10.48550/arXiv.1409.0473>.
- Casella, G., R. L. Berger, 2002: Statistical Inference (2nd ed.). Duxbury Mendenhall, W., & Sincich, T. Statistics for Engineering and the Sciences.
- Chen, J., Zhang, X., Wang, Y., & Li, T. (2020). Emergent constraints on future projections of the western North Pacific subtropical high. *Nature Communications*, 11, 2802. <https://doi.org/10.1038/s41467-020-16631-9>
- Chung, U., J. Rhee, M. Kim, S.-J. Sohn, 2023: Investigating data preprocessing algorithms of a deep learning postprocessing model for improvement of sub-seasonal to seasonal climate predictions. *Korean Journal of Agricultural and Forest Meteorology*, 25, 80-98, <https://doi.org/10.5532/KJAFM.2023.25.2.80>.
- Copernicus Climate Change Service (C3S), 2017: “ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate.” Copernicus Climate Change Service Climate Data Store (CDS)
- Do, H.-S., Kim, J., Cha, E.-J., Chang, E.-C., Son, S.-W., & Lee, G., 2023: Long-term change of summer mean and extreme precipitations in Korea and East Asia. *Int. J. Climatol.* 43(7), 3476-3492. <https://doi.org/10.1002/joc.8039>
- Du, Y., & Evans, K. J. (2010). Role of the Indian Ocean basin mode in driving the interdecadal variability of the Indian summer monsoon. *Geophysical Research Letters*, 37, L17702. <https://doi.org/10.1029/2010GL044415>
- Fan, Y., Y. Lyu, S. Zhu, Z. Yin, M. Duan, X. Zhi, B. Zhou, 2024: Monthly prediction on summer extreme precipitation with a deep learning approach: Experiments over the mid-to-lower reaches of the Yangtze River. *Earth Space Sci.*, 11, e2024EA003926. <https://doi.org/10.1029/2024EA003926>.
- Fowler, Hayley J., et al. “Anthropogenic intensification of short-duration rainfall extremes.” *Nature Reviews Earth & Environment* 2.2 (2021): 107-122.
- Gu, X., Huang, S., Togtokh, C., Jing, Z., Wu, M., Chen, Z., Matsuo, K., & Ren, G. (2025). Frequent land-ocean transboundary migration of tropical heatwaves. *Nature Communications*, 16, 2608. <https://doi.org/10.1038/s41467-025-58586-9>

- He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- Hobday, A.J. et al. (2016), A hierarchical approach to defining marine heatwaves, Progress in Oceanography, 141, pp. 227-238, doi: 10.1016/j.pocean.2015.12.014
- Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.
- Hsu, PC., Li, T., You, L. et al. A spatial-temporal projection model for 10-30 day rainfall forecast in South China. Clim Dyn 44, 1227-1244 (2015). <https://doi.org/10.1007/s00382-014-2215-4>
- Huang, B., Liu, C., Banzon, V., Freeman, E., Graham, G., Hankins, B., Smith, T. M., & Zhang, H.-M. (2021). Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) version 2.1. Journal of Climate, 34(8), 2923-2939. <https://doi.org/10.1175/JCLI-D-20-0166.1>
- Im, E. S., Pal, J. S., & Eltahir, E. A. B. (2017). Deadly heat waves projected in the densely populated agricultural regions of South Asia. Science Advances, 3(8), e1603322. <https://doi.org/10.1126/sciadv.1603322>
- IPCC, 2021: Chapter 11, "Weather and Climate Extreme Events in a Changing Climate." In: Climate Change 2021: The Physical Science Basis (AR6 WG1).
- Izumo, T., Yamagata, T., Behera, S. K., Luo, J. J., Masson, S., & Masumoto, Y. (2010). Influence of the state of the Indian Ocean Dipole on the following year's El Niño. Nature Geoscience, 3(3), 168-172. <https://doi.org/10.1038/ngeo760>
- Kingma D.P., Adam J. Ba, 2017: a Method for Stochastic Optimization. arXiv 1412.6980, <https://doi.org/10.48550/arXiv.1412.6980>.
- Lam, R., Sanchez-Gonzalez, A., Willson, M., Kufner, N., Hollingsworth, J., Lang, N., Godwin, J., Pfaff, T., Tomašev, N., Vinyals, O., Mohamed, S., & Battaglia, P. (2023). Learning skillful medium-range global weather forecasting. Science, 382(6677), 1416-1421. <https://doi.org/10.1126/science.adi2336>
- Lee, S., Park, M. S., Kwon, M., Park, Y. G., Kim, Y. H., & Choi, N. (2023). Rapidly changing East Asian marine heatwaves under a warming climate. Journal of Geophysical Research: Oceans, 128(12), e2023JC019761. <https://doi.org/10.1029/2023JC019761>
- Lee, Y., D. Cho, J. Im, C. Yoo, J. Lee, Y.-G. Ham, M.-I. Lee, 2024: Unveiling teleconnection drivers for heatwave prediction in South Korea using explainable artificial intelligence. npj clim. atmos. sci., 7, <https://doi.org/10.1038/s41612-024-00722-1>.

- Li, J., & Zeng, Q. (2002). A new application for diagnosing the dynamical and thermodynamical anomalies of the western North Pacific subtropical high. *Journal of the Meteorological Society of Japan*, 80(2), 279-299. <https://doi.org/10.2151/jmsj.80.279>
- Lin, H., J. Tang, S. Wang, 2023: Deep learning downscaled high-resolution daily near surface meteorological datasets over East Asia. *Sci Data* 10, <https://doi.org/10.1038/s41597-023-02805-9>.
- Mehmood, F., S. Ahmad, T.K. Whangbo, 2023: An efficient optimization technique for training deep neural networks. *Mathematics*, 11, <https://doi.org/10.3390/math11061360>
- Mu Z, J. Xia, 2024: Predicting the influence of extreme temperatures on grain production in the Middle-Lower Yangtze Plains using a spatially-aware deep learning model. *Peer J.*, <https://doi.org/10.7717/peerj.18234>.
- National Centers for Environmental Information. (n.d.). *Nino SST indices (Nino 1+2, 3, 3.4, 4; ONI and TNI)*. NOAA Climate Data Guide. <https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni>
- Nie, Yanbo, and Jianqi Sun. "Improving dynamical-statistical subseasonal precipitation forecasts using deep learning: A case study in Southwest China." *Environmental Research Letters* 19.7 (2024): 074013.
- Oh, H., Kim, G.-U., Chu, J.-E., Lee, K., & Jeong, J.-Y. (2023). The record-breaking 2022 long-lasting marine heatwaves in the East China Sea. *Environmental Research Letters*, 18(6), Article 064015. <https://doi.org/10.1088/1748-9326/acd267>
- Oktay, O., and Coauthors, 2018: Attention U-Net: Learning Where to Look for the Pancreas. arXiv:1804.03999, <https://doi.org/10.48550/arXiv.1804.03999>.
- Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., Benthuisen, J. A., Feng, M., Gupta, A. S., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Straub, S. C., & Wernberg, T. (2018). Longer and more frequent marine heatwaves over the past century. *Nature Communications*, 9, 1324. <https://doi.org/10.1038/s41467-018-03732-9>
- Physical Sciences Laboratory. (n.d.). *Niño 3.4 SST index from the HadISST1.1*. NOAA Physical Sciences Laboratory. <https://psl.noaa.gov/data/timeseries/month/Nino34/>
- Qin, Y., D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, 2017: A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. arXiv:1704.02971, <https://doi.org/10.48550/arXiv.1704.02971>.
- Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., & Thuerey, N. (2020).

WeatherBench: A benchmark dataset for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11), e2020MS002203. <https://doi.org/10.1029/2020MS002203>

Rasp, S., Hoyer, S., Merose, A., Langmore, I., Battaglia, P., Russell, T., Sanchez-Gonzalez, A., Yang, V., Carver, R., Agrawal, S., Chantry, M., Ben Bouallegue, Z., Dueben, P., Bromberg, C., Sisk, J., Barrington, L., Bell, A., & Sha, F. (2024). WeatherBench 2: A benchmark for the next generation of data-driven global weather models. *Journal of Advances in Modeling Earth Systems*, 16(1), e2023MS004019. <https://doi.org/10.1029/2023MS004019>

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution blended analyses for sea surface temperature. *Journal of Climate*, 20(22), 5473–5496. <https://doi.org/10.1175/2007JCLI1824.1>

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.0459, <https://doi.org/10.48550/arXiv.1505.04597>.

Ruder, Sebastian. “An overview of multi-task learning in deep neural networks.” arXiv preprint arXiv:1706.05098 (2017).

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., & Yamagata, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751), 360–363. <https://doi.org/10.1038/43854>

Shafiq F, A. Zafar, M.U. Ghani Khan, S. Iqbal, A.S. Albeshar, M.N. Asghar, 2025: Extreme heat prediction through deep learning and explainable AI. *PLoS ONE*, 20, <https://doi.org/10.1371/journal.pone.0316367>.

Shih, S. Y., F. K. Sun, and H. yi Lee, 2019: Temporal pattern attention for multivariate time series forecasting. *Machine Learning*, 108, 1421–1441, <https://doi.org/10.1007/s10994-019-05815-0>.

Vandal, T., E. Kodra, A. R. Ganguly, 2019: Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation. *Theor. Appl. Climatol.*, 137, <https://doi.org/10.1007/s00704-018-2613-3>.

Vonich, P. T., & Hakim, G. J. (2025). Testing the Limit of Atmospheric Predictability with a Machine Learning Weather Model. arXiv preprint arXiv:2504.20238.

Wang, B., Ding, Q., & Li, J. (2001). A dynamical global model for Asian summer monsoon precipitation. *Journal of Climate*, 14(18), 3971–3987. [https://doi.org/10.1175/1520-0442\(2001\)014<3971:ADGMFA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<3971:ADGMFA>2.0.CO;2)

Westra, Seth, et al. “Future changes to the intensity and frequency of short-duration extreme rainfall.” *Reviews of geophysics* 52.3 (2014): 522–555.

- Weyn, J. A., D. R. Durran, R. Caruana, and N. Cresswell-Clay, 2021: Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, 13, <https://doi.org/10.1029/2021MS002502>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences* (3rd ed.). Academic Press.
- Xu, L., H. Yu, Z. Chen, W. Du, N. Chen, M. Huang, 2023: Hybrid Deep Learning and S2S Model for Improved Sub-Seasonal Surface and Root-Zone Soil Moisture Forecasting. *Remote Sens.*, 15, <https://doi.org/10.3390/rs15133410>.
- Zebiak, S. E. (1993). Air-sea interaction in the equatorial Atlantic: A comparison of two coupled models. *Journal of Climate*, 6(12), 1567–1590. [https://doi.org/10.1175/1520-0442\(1993\)006<1567:ASIITE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1567:ASIITE>2.0.CO;2)
- Zuo, J., J. Cao, L. Chen, 2025: Predictability of the Summer 2022 Yangtze River Valley Heatwave in Multiple Seasonal Forecast Systems. *Adv. Atmos. Sci.* 42, <https://doi.org/10.1007/s00376-024-4304-6>.