

# Low-GloSea6 기상 예측 소프트웨어의 머신러닝 기법 적용 연구

박혜성, 조예린, 신대영, 윤은옥, 정성욱\*

## A Study of the Application of Machine Learning Methods in the Low-GloSea6 Weather Prediction Solution

Hye-Sung Park, Ye-Rin Cho, Dae-Yeong Shin, Eun-Ok Yun, Sung-Wook Chung\*

**요약** 슈퍼컴퓨팅 기술 및 하드웨어 기술이 발전함에 따라 기후 예측 모델도 고도화되고 있다. 한국 기상청 역시 영국 기상청으로부터 GloSea5을 도입하였고 한국 기상 환경에 맞추어 업데이트된 GloSea6를 운용 중이다. 각 대학 및 연구 기관에서는 슈퍼컴퓨터보다는 사양이 낮은 중소규모 서버에서 활용하기 위해 저해상도 결합모델인 Low-GloSea6를 구축하여 사용하고 있다. 본 논문에서는 중소규모 서버에서의 기상 연구의 효율성을 위한 Low-GloSea6 소프트웨어를 분석하여 가장 많은 CPU Time을 점유하는 대기 모델의 tri\_sor.F90 모듈의 tri\_sor\_dp\_dp 서브루틴을 Hotspot으로 검출하였다. 해당 함수에 머신러닝의 한 종류인 선형 회귀 모델을 적용하여 해당 기법의 가능성을 확인한다. 이상치 데이터를 제거 후 선형 회귀 모델을 학습한 결과 RMSE는 2.7665e-08, MAE는 1.4958e-08으로 Lasso 회귀, ElasticNet 회귀보다 더욱 좋은 성능을 보였다. 이는 Low-GloSea6 수행 과정 중 Hotspot으로 검출된 tri\_sor.F90 모듈에 머신러닝 기법 적용 가능성을 확인하였다.

**Abstract** As supercomputing and hardware technology advances, climate prediction models are improving. The Korean Meteorological Administration adopted GloSea5 from the UK Met Office and now operates an updated GloSea6 tailored to Korean weather. Universities and research institutions use Low-GloSea6 on smaller servers, improving accessibility and research efficiency. In this paper, profiling Low-GloSea6 on smaller servers identified the tri\_sor\_dp\_dp subroutine in the tri\_sor.F90 atmospheric model as a CPU-intensive hotspot. Applying linear regression, a type of machine learning, to this function showed promise. After removing outliers, the linear regression model achieved an RMSE of 2.7665e-08 and an MAE of 1.4958e-08, outperforming Lasso and ElasticNet regression methods. This suggests the potential for machine learning in optimizing identified hotspots during Low-GloSea6 execution.

**Key Words** : Global Climate Model, Low-GloSea6, Machine Learning, Linear Regression, IQR Outlier Detection

### 1. 서론

컴퓨터 및 하드웨어 기술이 발전함에 따라 각국의 기후 예측 모델 또한 고도화 되고 있다. 그 중 널리 사용되고 있는 수치적 기상 예보(Numerical Weather Prediction, NWP) 방식은 역학 및 물리적 원리를 기

반으로 기상 현상에 대한 지배 방정식을 통해 계산하는 방식이다[1]. 수치적 기상 예보를 수행하기 위해서는 관측데이터에서 도출된 초기 조건(Input Data)을 제공하여 미래의 기상 예측을 진행하게 되며, 이때 많은 양의 데이터를 방대하고 다량의 연산이 필요하여 고성능 슈퍼컴퓨터를 활용한 기술이 연구 및 개발 되

This research was funded by the Korea Meteorological Administration Research and Development Program under Grant K MI 2021-01310.

\*Corresponding Author : Department of Computer Engineering, Changwon National University (swchung@changwon.ac.kr)  
 Received October 10, 2023                      Revised October 13, 2023                      Accepted October 19, 2023

고 있다.

지구 기상 예보 모델(Global Climate Model, GCM)은 수치적 기상 예보 모델 중 하나이며, 이는 지구 전체를 대상으로 기상 분석 및 예측 하는 모델이다 [2]. 지구 기상 예보 모델은 지구 전체를 격자로 나누어 각 격자에 대한 대기, 해양, 해빙, 육지에 관련된 데이터를 수집 후, 이를 토대로 모든 격자를 적분하여 출력 데이터를 생성 하여 예측 데이터를 제공한다.

한국에서도 기후 예측 분야의 많은 연구가 이루어지고 있으며, 슈퍼컴퓨터 도입 후 더욱 정밀한 기상 예보가 가능하게 되었다. 한국 기상청(KMA, Korea Meteorological Administration) 또한 영국 기상청으로부터 2014년에 GloSea5(Global Seasonal Forecasting System 5)을 도입하였다[3]. 그 이후 2018년 8월부터 현재까지 국립기상과학원에서 한국의 기상환경에 적합하도록 소스코드를 수정하여 GloSea6(Global Seasonal Forecasting System 6)로 업데이트 하여 사용 중이다[4]. 현재 대기모델의 역학 코어에서 이류 계산을 위하여 semi-implicit semi\_lagrangian 수치 기법을 사용한다. 이 기법에서 GloSea5의 경우 수분 예단변수(moist prognostics)의 수평 방향으로 bi-cubic 내삽법이 수직 방향으로 quantic 내삽법으로 다르며, 건조 예단변수(dry prognostics)에서 cubic Hermite 내삽법이 사용된다. GloSea6로 업데이트되면서 모두 동일하게 cubic Hermite 내삽법만을 사용한다. 이를 통해 열대 대류권의 온난 편차를 감소하였다.

Low-GloSea6는 GloSea6의 사용량이 증가하여 슈퍼컴퓨터의 부하가 증가함에 따라 성능이 저하되는 문제를 막기 위해 격자의 해상도를 고해상도에서 저해상도로 변환한 저해상도 결합모델이다. Low-GloSea6를 통해 기존의 슈퍼컴퓨터에서만 진행되어온 기상예보 모델 연구 및 개발을 각 연구기관, 기업 및 대학에서 연구 및 활용이 가능하다. 이를 통해 중소규모 서버를 기반으로한 실험 환경에서도 수월하게 개발 및 연구가 이루어 질 수 있다.

최근 들어 머신 러닝 기술의 발전으로 데이터 기반의 머신 러닝 모델을 사용하여 기상 예보를 진행하는 연구가 있으며, 기존의 수치적 기상 예보 모델에 비해

매우 빠른 속도 및 낮은 컴퓨팅 파워로 예측이 가능하다는 장점이 있다[5].

이를 통해 해당 논문에서는 각 대학 및 연구기관에서 원활한 기상 연구 환경을 위해 한국의 기상환경에 최적화 되어있는 Low-GloSea6를 Vtune 프로파일링 툴을 사용하여 분석한다. 프로파일링에서 검출된 Hotspot인 tri\_sor.F90 모듈을 머신 러닝의 한 종류인 선형 회귀(Linear Regression) 모델을 사용하여 데이터 기반 학습을 진행 한다. 해당 모델에 선형 회귀 모델을 적용 함으로써 Low-GloSea6 수행 시간 단축을 위한 머신 러닝 적용 방안의 가능성을 확인한다.

## 2. Low-GloSea6

### 2.1 Low-GloSea6 구성

Low-GloSea6는 영국 기상청으로부터 도입한 GloSea5를 한국의 기상환경에 맞게 업데이트한 버전인 GloSea6의 저해상도 결합모델이다. Low-GloSea6는 GloSea6와 같은 구조를 가지고 있다. GloSea6는 표 1과 같이 대기, 지면, 해양, 해빙 총 4가지의 모델이 서로 결합되어 구성된다.

표 1. GloSea6 모델  
Table 1. Model of GloSea6

model	name	purpose
UM 11.2	Met Office Unified Model	Atmosphere
CICE 5.1.2	Los Alamos Sea Ice Model	Sea Ice
NEMO 3.6	Nucleus of European Modeling of the Ocean	Ocean
JULES 5.0	Joint UK Land Environment Simulator	Land

각 4가지 모델의 상세 명칭과 결합자는 그림 1과 같다. 대기 모델(Met Office Unified Model, UM)과 지면 모델(Joint UK Land Environment Simulator, JULES)은 영국 기상청, 해양 모델(Nucleus for European Modeling of the Ocean, NEMO)은 유럽 컨소시움, 해빙 모델(Los Alamos Sea Ice model, CICE)

은 미국 로스앨러모스 국립연구소에서 개발되었다.

대기 모델은 Dynamics 과정에서 ENDGame (Semi-Lagrangian advection)을[6], Microphysics 과정에서 Warm rain scheme 알고리즘을 사용하며 [7], 지면 모델은 Dynamics 과정에서 Centered 2nd-order finite difference approximation을, Data Assim 과정에서 3-d variational method 알고리즘을 사용하였다[8]. 해양 모델은 비선형 자유수면파와 가변 체적층(variable volume layers, VVL)을 도입하여 해수면 함수 플럭스를 더 정확하게 표현할 수 있다. 해빙 모델은 ice fraction 고려하여 재격자화해주는 semi-implicit coupling 알고리즘을 사용한다[9].

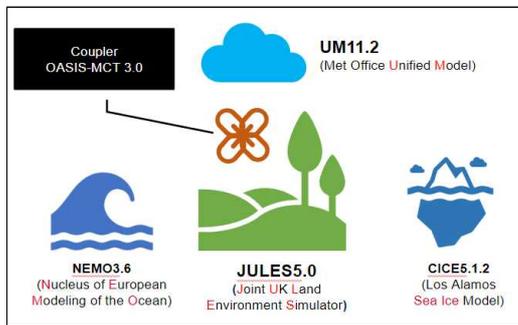


그림 1. GloSea6 결합 모델  
Fig. 1. Copuled Model of GloSea6

해당 4가지 모델은 GloSea6에서 상호 의존적으로 영향을 주고 받으며 유기적으로 동작하며, 서로 다른 4 가지 모델을 결합하기 위해서 OASIS-MCT를 사용하여 결합한다[10].

## 2.2 Low-GloSea6 프로파일링

본 논문에서는 Low-GloSea6의 Hotspot 검출을 위하여 Intel에서 개발 및 제공하는 VTune Profiler를 사용하여 프로파일링 진행을 하였다[11]. Vtune Profiler는 어플리케이션 성능, 시스템 성능 및 최적화를 위한 도구를 제공한다. 프로파일링을 통해 어플리케이션 동작 시코드 내에서 가장 오랜 시간 소요되는 Hotspot을 탐색 가능하다. CPU Usage 기반 Hotspot 분석에는 두가지 샘플링 기반 탐색 모드가 존재한다. 오버헤드가 존재하지만 수집을 위해 샘플링 드라이버가 필요하지 않은 사용자 모드 샘플링과, 최소 수집 오버헤드를 샘플링 드라이버를 설치해야 하는 하드웨어 이벤트 기반 샘플링이 존재한다. 본 논문에서는 샘플링 드라이버를 설치하지 않는 사용자 모드 샘플링 Hostpot 분석을 시도하였다. 프로파일링을 진행한 서버의 사양은 표 2와 같다.

CPU Time	Instructions Retired	Microarchitecture Usage	Module	Function (Full)
13032.006s	120,687,327,400,000	79.8%	libmpi.so.12.0.5	MPIDI_CH3I_Progress
3097.394s	21,479,359,400,000	36.6%	libmpi.so.12.0.5	MPID_nem_tcp_connpoll
1408.185s	4,498,531,200,000	34.6%		[Outside any known module]
1283.903s	17,827,578,400,000	100.0%	libmpi.so.12.0.5	MPIDU_Sched_are_pending
1125.987s	1,706,568,600,000	13.3%	um-atmos.exe	tri_sor_mod_MOD_tri_sor_dp_omp_fn.7
888.238s	4,721,956,000,000	48.4%	libm-2.17.so	__ieee754_pow_sse2
617.219s	4,255,281,800,000	63.0%	libm-2.17.so	__ieee754_exp_avx
450.796s	2,156,222,600,000	41.3%	libm-2.17.so	__exp1
437.765s	212,013,400,000	23.0%	libgfortran.so.4.0.0	func@0x1c270
347.046s	986,077,200,000	22.4%	um-atmos.exe	__mod_cosp_MOD_cosp_iter
334.226s	1,264,324,600,000	43.5%	libmpi.so.12.0.5	MPID_nem_network_poll
315.775s	2,268,923,000,000	64.5%	um-atmos.exe	__eg_cubic_lagrange_mod_MOD_eg_cubic_lagrange_.
315.586s	1,213,792,200,000	43.6%	libmpi.so.12.0.5	func@0x3da00
241.898s	245,096,200,000	8.3%	um-atmos.exe	__glue_conv_6a_mod_MOD_glue_conv_6a
168.967s	1,118,036,000,000	49.7%	libc-2.17.so	__int_malloc
143.053s	496,743,600,000	29.9%	um-atmos.exe	__eg_calc_ax_mod_MOD_eg_calc_ax_dp_dp_omp_fn.
137.651s	868,364,600,000	57.4%	um-atmos.exe	bdy_impl3_omp_fn.0
128.050s	123,070,600,000	14.9%	libgfortran.so.4.0.0	func@0x1c230
115.389s	53,329,200,000	18.5%	libgfortran.so.4.0.0	func@0x1c4b0
115.055s	849,501,400,000	54.2%	libc-2.17.so	__int_free
114.831s	294,956,000,000	16.7%	libmpi.so.12.0.5	MPIDI_CH3U_Recvq_FU
93.374s	176,665,800,000	15.9%	um-atmos.exe	__ls_ppnc_mod_MOD_ls_ppnc_gather

그림 2. Low-GloSea6 분석  
Fig. 2. Low-GloSea6 Profiling

표 2. 프로파일링에 수행된 시스템 스펙  
Table 2. System specs used for profiling

Name	Hardware Specification
CPU	Intel® Core 10 <sup>th</sup> GEN i7-10700K
RAM	64GB
SSD	1TB
GPU	GeForce RTX 3080

Low-GloSea6 1일차 기준 전체 Elapsed Time은 4278.722초, CPU Time은 29877.390초 소요되었다. 그 중에서 그림 2와 같이 Low-GloSea6를 구성하는 4가지 모델 중에서 대기(UM) 모델의 tri\_sor.F90 모듈의 tri\_sor\_dp\_dp 함수가 가장 많은 CPU Time이 소요된 Hotspot으로 검출되었다. 해당 함수는 1125.987초 소요되었으며 전체 CPU Time 중 13.3%의 비중을 차지하는 것으로 분석되었다.

### 2.3 tri\_sor.F90

앞서 Low-GloSea6의 가장 오래 CPU Time을 소모하는 Hotspot으로 검출된 tri\_sor.F90 모듈은 Low-GloSea6를 구성하는 4가지 모델중 대기(UM) 모델에 존재하며 dynamics solver 기능을 하는 모듈이다. Low-GloSea6 연산 과정에서 선형 시스템의 해를 근사적으로 계산하고 병렬화를 개선하는 역할을 하며 tri\_sor.F90 모듈 내에는 tri\_sor\_dp\_dp, tri\_sor\_dp\_sp, tri\_sor\_sp\_dp, tri\_sor\_sp\_sp 총 4가지 서브 루틴이 존재하며 각 서브 루틴의 수행 과정은 tri\_sor.h 헤더파일에 존재한다. 프로파일링을 통해 Hotspot으로 검출된 함수는 tri\_sor\_dp\_dp 함수이며, 해당 함수는 수치적 계산을 통해 동작하며 사용되는 변수로 Input data 9가지와 Output data 1가지를 사용한다.

Hotspot으로 검출된 tri\_sor.F90 모듈은 그림 3과 같이 mg\_solver.F90, eg\_precon.F90, mg\_solve.F90 3개의 요소에 종속되어 있다. 또한 tri\_sor.F90은 헬름홀츠 방정식 모듈인 helmholtz\_const\_matrix\_mod.F90를 포함하여 총 8개의 모듈이 tri\_sor.F90 모듈을 의존한다.

본 논문에서는 tri\_sor\_dp\_dp 함수에서 사용되는 데이터와 동일한 feature들을 이용하여 수치적 계산이

아닌 머신러닝 모델 중 하나인 선형 회귀 모델을 적용하여 가능성을 확인하였다.

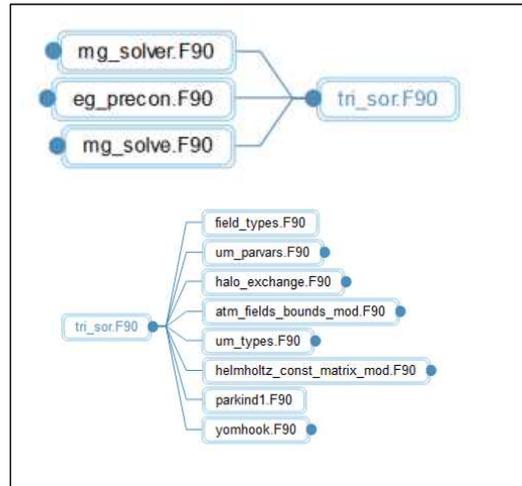


그림 3. tri\_sor.F90 플로우 차트  
Fig. 3. tri\_sor.F90 Flowchart

## 3. 머신러닝 기법 적용

머신 러닝의 기법 중 하나인 선형 회귀(Linear Regression)는 독립 변수와 종속 변수 간의 관계를 분석하는 기법이다. 선형 회귀는 단일 선형 회귀와 다중 선형 회귀로 나뉠 수 있으며, 회귀 모델은 종속 변수 y와 y의 영향을 미치는 독립 변수 xi(i=1,2,3,...)로 구성된 모델을 분석한 뒤 가장 y의 정확도가 높은 가중치를 탐색하는 방식이다[12, 13]. 본 논문에서는 Lasso Regression, ElasticNet Regression, 이상치 데이터를 제거 후 학습 진행한 Linear Regression 3가지 모델을 각각 구성 하여 평가 및 비교한다.

### 3.1 선형 회귀

선형 회귀는 크게 단일 선형 회귀와 다중 선형 회귀로 구분할 수 있다. 본 논문에서 적용할 데이터는 여러 개의 독립 변수로 구성되어 있기 때문에 식 1과 같은 다중 선형 회귀로 구성 하였다. 데이터를 학습 할 때 선형 함수로 나타내기 위해서는 실제 데이터를 사용하여 반복적인 계산을 시도한다. 이때 손실을 기준으로

경사하강법을 사용하여 계산한다. 경사하강법은 임의의 파라미터를 지정한 다음 조금씩 변화 시키가며 실제 값과의 손실을 점차 줄여나가며 최적의 파라미터를 탐색하는 방식이다.

$$y = w_1x_1 + w_2x_2 + \dots + w_nx_n + b \quad (1)$$

### 3.2 Lasso 회귀

Lasso 회귀는 기존 선형 회귀의 단점을 보완하여 범용성을 부여하기 위해 고안된 모델이다. 기존의 선형 회귀는 최적의 가중치와 편향을 찾아내는 것이 관건이었다면 Lasso 회귀는 기존의 선형 회귀에서 더 나아가 가중치들의 절댓값의 합이 최소가 되는 조건을 추가한다. 가중치의 모든 원소가 0이 되거나 0에 가깝게 되도록 해야 한다. 즉 해당 모델은 수식 2와 같이 L1 norm 패널티를 가진 선형 회귀 모델이다.

$$J(\theta) = MSE(\theta) + \alpha \sum_{i=1}^n |\theta_i| \quad (2)$$

### 3.3 ElasticNet 회귀

ElasticNet 회귀는 Lasso 회귀와는 다르게 L1, L2 두가지 norm을 둘 다 사용하는 모델이다. 두 가지 norm의 장점을 모두 가지고 있어 ElasticNet 회귀 모델은 큰 데이터 셋에서 좋은 성능을 보인다는 특징이 있다. 또한 수식 3과 같이 변수의 수를 줄이고 가중치들의 절댓값의 합이 최소가 되게한다.

$$J(\theta) = MSE(\theta) + \alpha \sum_{i=1}^i \theta_i^2 + \alpha \sum_{i=1}^i |\theta_i| \quad (3)$$

### 3.4 이상치 탐색

본 논문에서는 IQR(Inter Quantile Range) 방식 이상치 데이터 제거를 진행 한 다음, 기본적인 선형 회귀 모델에 학습하였다. IQR이란 전체 데이터들을 오름차순으로 정렬한 뒤, 정렬된 데이터를 사분위로 나누어 75%의 지점과 25%의 지점의 값의 차이를 뜻한다.

tri\_sor\_dp\_dp 함수의 주어진 데이터 셋으로 얻은 IQR 값은 5.3959e-09 이다. 이때 IQR값을 1.5배 후

학습하였을 때 MAE는 1.495788e-08으로 가장 낮게 측정 되었으며 IQR에 1.25배 하였을 때 MAE가 1.495799e-08, IQR에 1.75배 하였을 때 MAE가 1.495832e-08, IQR에 2배 하였을 때 MAE가 1.495848e-08, IQR에 1배 하였을 때 MAE가 1.834965e-08 순서로 측정되었다. 따라서 IQR 값으로 이상치 데이터를 검출 하기 위하여 본 논문에서는 가장 우수한 성능을 보이는 IQR값에 1.5배를 하여 생성 된 범위를 벗어나는 데이터를 이상치 데이터로 판단하여 선형 회귀 학습을 진행하였다.

## 4. 실험 및 실험 결과

### 4.1 실험 환경

본 논문에서 머신러닝 수행에 사용된 시스템 스펙은 표 3과 같다. 머신러닝 학습 및 데이터 분석에 사용된 라이브러리로 python3.6, pandas, numpy, scikit-learn, matplotlib을 사용하였다. Lasso 회귀, ElasticNet 회귀, 이상치 데이터를 제거 후 선형 회귀 학습한 모델로 총 3가지 모델을 각각 구성하여 학습 후 성능 평가 및 비교를 진행하였다.

Lasso 회귀 모델과 ElasticNet 회귀 모델은 각각 Alpha 값을 0.001로 설정하였으며 최대 학습 횟수는 10만번으로 설정 하였다. 선형 회귀 모델은 IQR 값으로 측정된 값인 5.3959e-09을 사용하여 IQR 값의 1.5배를 벗어나는 범위의 데이터를 이상치 데이터로 판단하여 제외 후 학습을 진행 하였다.

표 3. 실험에 수행된 시스템 스펙  
Table 3. System specs used by test

Name	Hardware Specification
CPU	Intel® Core i3 <sup>th</sup> GEN i9-13900F
RAM	128GB
SSD	2TB
GPU	GeForce RTX 4070TI

### 4.2 성능 평가

기존의 함수 호출에 적용하기 용이하도록 함수 호출 1회 데이터를 기준으로 예측 후 평가 및 비교를 진행

하였다.

tri\_sor\_dp\_dp 함수는 input data로 x, ltHlm\_Ln, ltHlm\_Ls, ltHlm\_Le, ltHlm\_Lw, ltHlm\_Ld, ltHlm\_lu, ltHu\_k, ltHd\_k 총 9개의 feature가 존재하며 output data로 Px 1개의 feature가 존재한다. 이때 input feature인 x에서 나머지 변수들을 이용하여 solver의 역할을 하는 tri\_sor\_dp\_dp 함수를 통해 Px의 값이 수치적으로 연산된다.

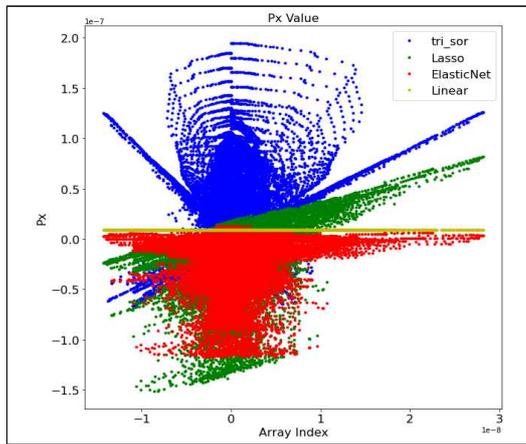


그림 4. Px 예측값  
Fig. 4. Px prediction value

첫 번째로 그림 4과 같이 예측을 진행한 Output feature로 변수 “Px”를 예측 하였고 Input feature와의 상관관계를 통해 예측 유사도를 그래프로 확인하기 위하여 변수 “x”를 Input feature로 사용하여 그래프를 작성하였다. 그래프를 살펴 보았을 때 선형 회귀 모델은 아주 작은 범위로 편향되었으며 ElasticNet 회귀 모델은 0보다 낮은 값으로 Lasso 회귀 모델은 비교군 중 가장 실제값과 유사한 산점도를 예측한 것으로 나타난다.

두 번째로 그림 5과 같이 각 모델별 RMSE를 비교 하였을 때 선형 회귀 모델이 2.7665e-08으로 가장 낮게 측정 되었으며 그 다음으로 ElasticNet 회귀 모델이 4.2826e-08, Lasso 회귀 모델이 4.4334e-08 순서로 측정 되었다. 그래프 상에서 가장 유사하지 않았던 선형 회귀 모델이 가장 낮은 RMSE로 측정 된 이유로는 작은 범위에서 측정이 이루어 졌기 때문에 기존 실제 데이터의 단위가 매우 작은 특성 상 이러한 결

과가 도출되었다.

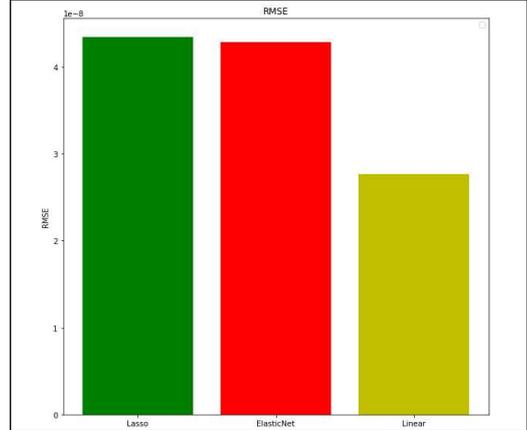


그림 5. 각 모델별 RMSE  
Fig. 5. RMSE for each Model

세 번째로 그림 6과 같이 각 모델별 MAE를 비교하였을 때 선형 회귀 모델이 1.4958e-08로 가장 낮게 측정 되었으며, 그 다음으로 ElasticNet 회귀 모델이 3.3823e-08, Lasso 회귀 모델이 3.4317e-08으로 가장 높게 측정 되었다. 가장 유사하지 않게 그래프상 확인되었던 선형 회귀 모델의 예측이 가장 좋은 성능을 보이는 이유로는 마찬가지로 낮은 범위내에서만 예측이 이루어 졌기 때문에 데이터의 단위가 매우 작은 특성상 가장 좋은 성능이 측정되었다.

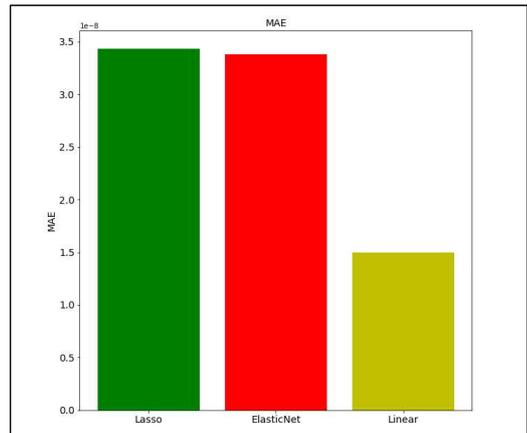


그림 6. 각 모델별 MAE  
Fig. 6. MAE for each Model

## 5. 결론

본 논문에서는 GloSea6의 저해상도 결합모델인 Low-GloSea6의 프로파일링을 진행하였으며 해당 프로파일링에서 가장 많은 CPU Time으로 측정된 Hotspot인 대기(UM) 모델 tri\_sor.F90 모듈의 tri\_sor\_dp\_dp 함수를 검출하였다. 해당 모듈은 기존의 수치적 연산을 기반으로 수행되며 본 논문에서는 더욱 빠른 수행 시간을 위한 머신러닝 적용 가능성을 확인하기 위해 선형 회귀, Lasso 회귀, ElasticNet 회귀로 총 3가지 모델을 구성하여 학습 후 평가 및 비교 하였다. 그 결과 기존 수치적 연산과 비교하였을 때 이상치 데이터를 제거 후 선형 회귀 모델에 학습하였을 때 RMSE가 2.7665e-08, MAE가 1.4958e-08로 다른 모델보다 가장 좋은 성능을 나타내는 것을 보여주고 있다. 따라서 본 논문을 통하여 Low-GloSea6 과정 중 CPU Time이 가장 오래 소요하는 tri\_sor.F90 모듈을 검출하였고 해당 모듈에 대한 머신러닝의 적용 가능성을 확인하였다.

## REFERENCES

- [1] P. Bauer, A. Thorpe, G. Brunet, "The quiet revolution of numerical weather prediction", *Nature*, vol. 525, pp. 47-55, 2015.
- [2] D. W. Pierce, T. P. Barnett, B. D. Santer, P. J. Glecker, "Selecting global climate models for regional climate change studies", *National Acad Sciences*, vol. 106, no. 21, pp. 8441-8446, 2009
- [3] KMA - Exchange of Memorandum of Understanding between the National Center for Atmospheric Research and Korea Meteorological Administration, [online] <https://www.kma.go.kr/kma/news/press.jsp>
- [4] H. Kim, J. Lee, Y. Hyun, S. Hwang, "The KMA Global Seasonal Forecasting System (GloSea6) - Part 1: Operational System and Improvements", *Korean Meteorological Society*, vol. 31, no. 3, pp. 341-359, 2021.
- [5] L. Chen, F. Du, Y. Hu, Z. Wang, F. Wang, "SwinRDM: Integrate SwinRNN with Diffusion Model towards High-Resolution and High-Quality Weather Forecasting", *The Thirty-Seventh AAAI Conference on Artificial Intelligence*, vol. 37, no. 36, pp. 322-330, 2023.
- [6] N. Wood, A. Staniforth, A. White, T. Allen, M. Diamantakis, M. Gross, T. Melvin, C. Smith, S. Vosper, M. Zerroukat, J. Thurnburn, "An inherently massconserving semiimplicit semiLagrangian discretization of the deep atmosphere global nonhydrostatic equations", *Quarterly Journal of the Royal Meteorological Society*, vol. 140 no. 682, pp. 1505-1520, 2014.
- [7] I. A. Boutle, S. J. Abel, P. G. Hill, C. J. Morcrette, "Spatial variability of liquid cloud and rain: Observations and microphysical effects", *Quarterly Journal of the Royal Meteorological Society*, vol. 140 no. 679, pp. 583-594, 2014.
- [8] K. Mogensen, M. A. Balmaseda, A. T. Weaver, M. Martin, A. Vidard, "NEMOVAR: A variational data assimilation system for the NEMO ocean model", *ECMWF newsletter*, vol. 120, pp. 17-22, 2019.
- [9] J. K. Ridley, E. W. Blockley, A. B. Keen, J. G. Rae, A. E. West, D. Schroeder, "The sea ice model component of HadGEM3-GC3.1.", *Geoscientific Model Development*, vol. 11, no. 2, pp. 713-723, 2018.
- [10] S. Yukimoto, Y. Adachi, M. Hosaka, T. Sakami, H. Yoshimura, M. Hirabara, T. Y. Tanaka, E. Shindo, H. Tsujino, M. Deushi, R. Mizuta, S. Yabu, A. Obata, H. Nakano, T. Koshiro, T. Ose, A. Kitoh, "A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3-Model Description and Basic Performance-", *Meteorological Society of Japan*, vol. 90, pp. 23-64, 2012.
- [11] VTune-profiler [online] <https://www.intel.com/content/www/us/en/developer/tools/oneapi/vtune-profiler.html#gs.5xzf5j>
- [12] Su, X., Yan, X. and Tsai, C.L., "Linear regression", *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, No.3 pp.275-294, 2012.
- [13] R. Shen, B. Zhangd, "The research of

regression model in machine learning field”, *MATEC web of Conferences*, vol. 176, no. 3, pp. 01033, 2018.

---

저자약력

---

**박혜성 (Hye-Sung Park)** [학생회원]



- 2020년 3월 ~ 현재: 창원대학교 학부 과정

〈관심분야〉 AI, 머신러닝, 데이터 분석

**조예린 (Ye-Rin Cho)** [학생회원]



- 2020년 3월 ~ 현재: 창원대학교 학부 과정

〈관심분야〉 IoT, 실시간 분산 멀티미디어

**신대영 (Dae-Yeon Shin)** [학생회원]



- 2020년 3월 ~ 현재: 창원대학교 학부 과정

〈관심분야〉 IoT, 실시간 분산 멀티미디어

**윤은옥 (Eun-Ok Yun)** [학생회원]



- 2020년 3월 ~ 현재: 창원대학교 학부 과정

〈관심분야〉 유무선 네트워크, IoT, 스마트팩토리

**정성욱 (Sung-Wook Chung)** [종신회원]



- 2010년 8월: CISE dept. Univ. of Florida, USA, (Ph.D)
- 2010년 10월 ~ 2012년 2월: KT 종합기술원 중앙연구소 선임연구원
- 2012년 3월 ~ 현재: 창원대학교 컴퓨터공학과 정교수

〈관심분야〉 IoT, 스마트모빌리티, HPC, 실시간 분산 멀티미디어 시스템