Refining temperature profile from GEO-KOMPSAT-2A using ERA5 reanalysis data and a light gradient boosting machine over East Asia

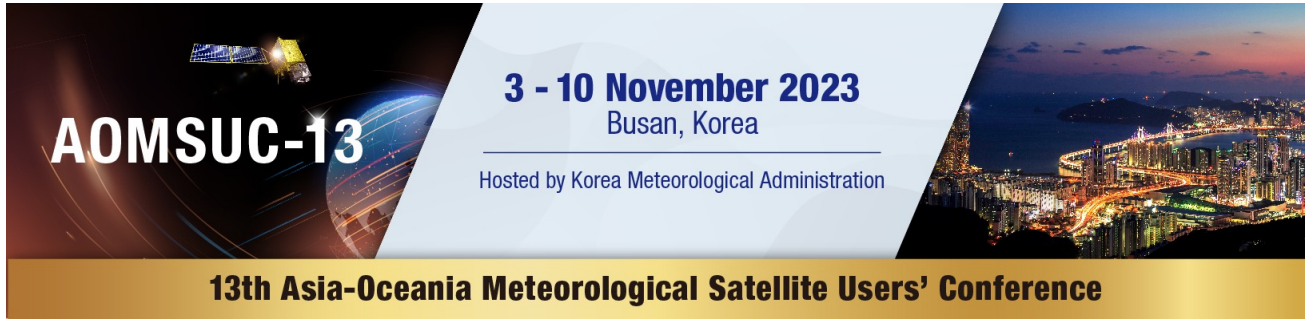
Daehyeon Han, Sihun Jung, Minki Choo, Juhyun Lee, Dongjin Cho, Jungho Im

Intelligent Remote sensing and geospatial information Science (IRIS) Lab

Ulsan Natioinal Institute of Science and Technology (UNIST), South Korea

https://iris.unist.ac.kr

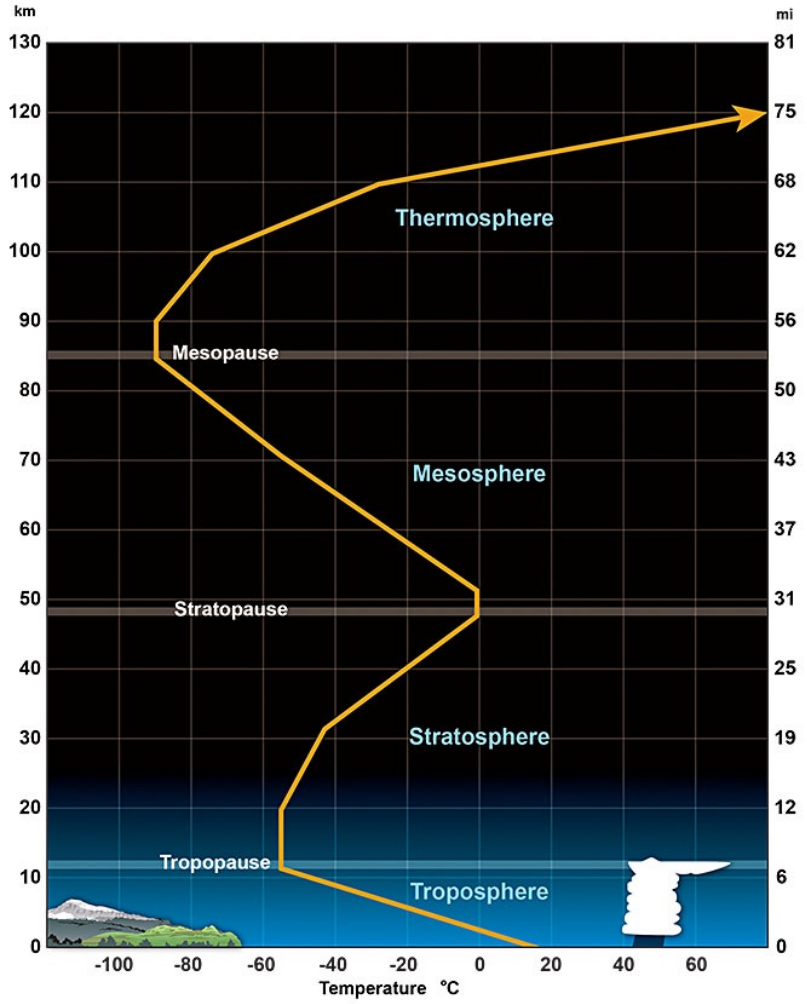
November 8, 2023



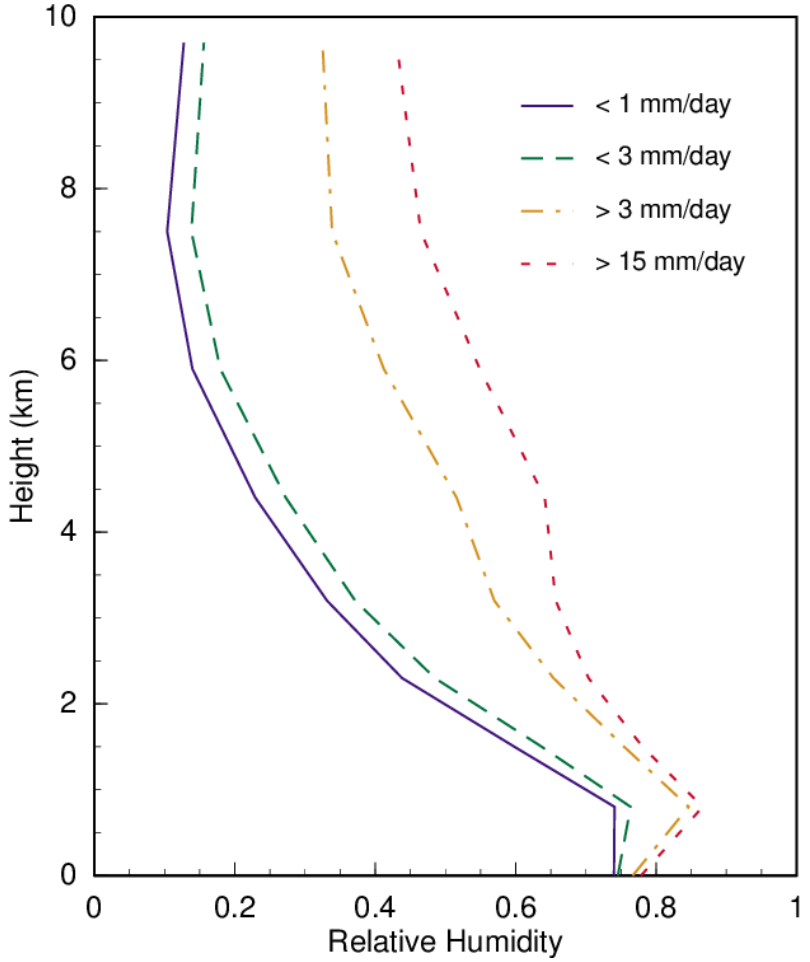
daehyeon.han@unist.ac.kr

Introduction

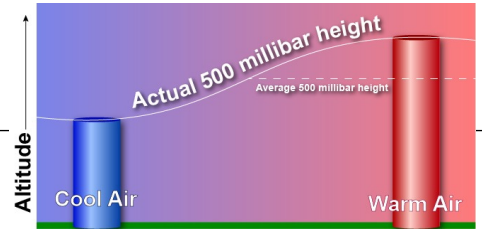
Atmospheric profile

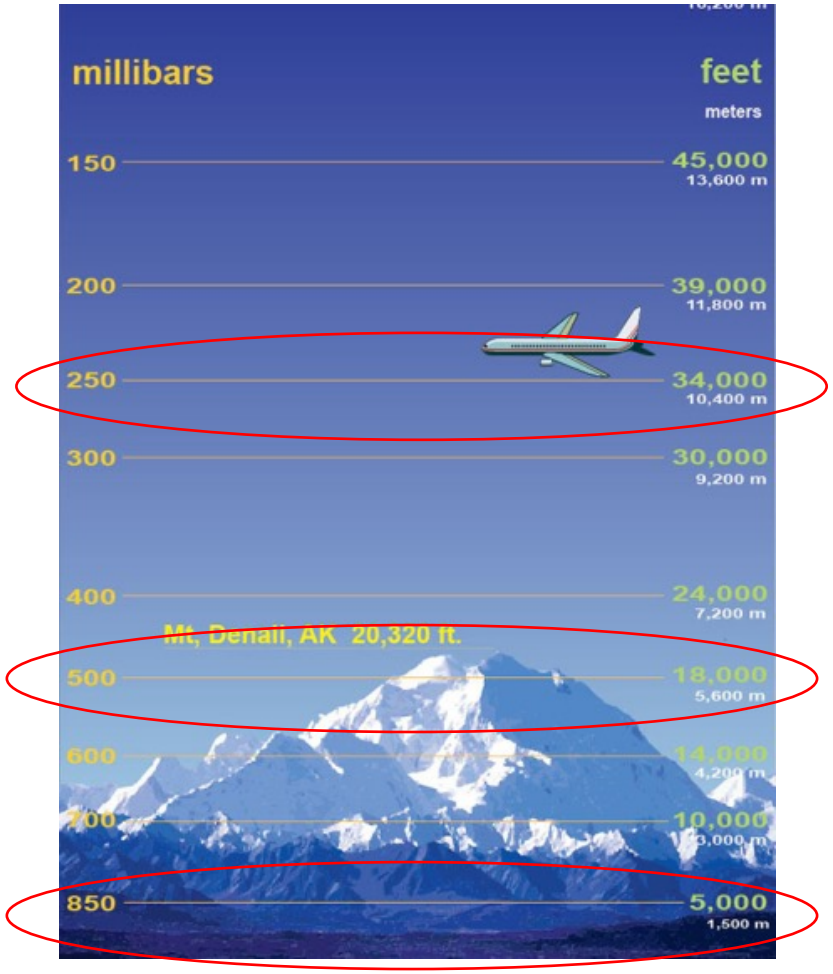


NOAA



Baker et al., 2002





NOAA

daehyeon.han@unist.ac.kr

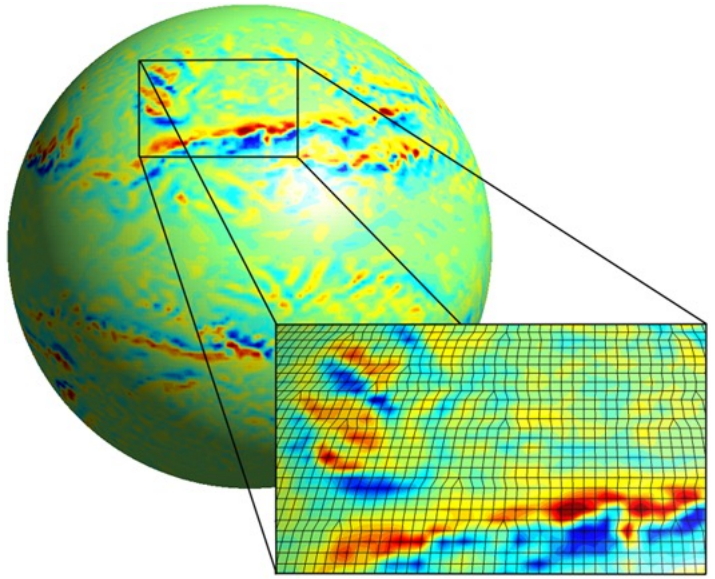
Introduction

Atmospheric profile from different sources

Radiosonde

Most accurate, point data
(ground truth)

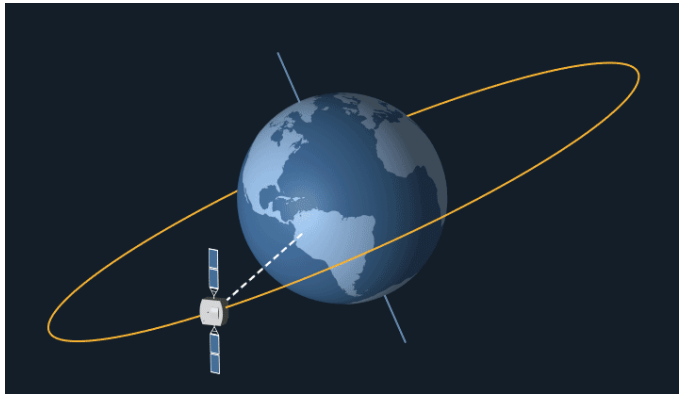
Numerical weather prediction (NWP)
Reanalysis data (ERA5, NCEP/NCAR)



-ECMWF

Spatial information
Accurate historical data

Geostationary satellites (GEO)
(GK2A, Himawari, GOES, MeteoSat)



-NASA

Real-time observation

daehyeon.han@unist.ac.kr

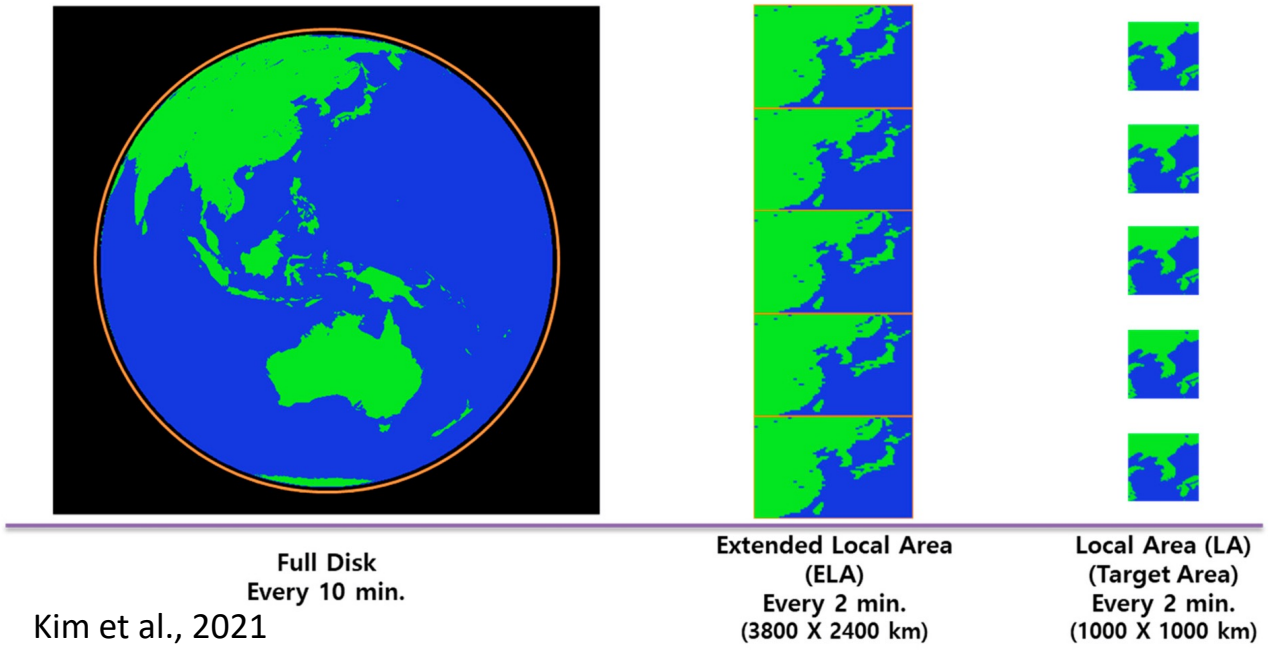
Data and method

GK2A AMI

- Geo-KOMPSAT-2A (GK2A)
- Advanced Meteorological Imager (AMI)
- Korean 2nd geostationary satellite (launched in 2019)
- 16 channels (visible, NIR, water vapor, IR)
- 0.5-2 km spatial resolution
- Every 2-10 min. temporal resolution

Category	Channel No.	Center of Wave-length (μm)	Bandwidth (μm)	Resolution (km)
Visible	1	0.47	0.43–0.48	1
	2	0.51	0.50–0.52	1
	3	0.64	0.63–0.66	0.5
Near Infrared	4	0.86	0.85–0.87	1
	5	1.37	1.37–1.38	2
	6	1.61	1.60–1.62	2
Water Vapor	7	3.83	3.74–3.96	2
	8	6.20	6.06–6.42	2
	9	6.90	6.89–7.01	2
Infrared	10	7.30	7.26–7.43	2
	11	8.60	8.44–8.76	2
	12	9.60	9.54–9.72	2
	13	10.40	10.25–10.61	2
	14	11.20	11.08–11.32	2
	15	12.30	12.15–12.45	2
	16	13.30	13.21–13.39	2

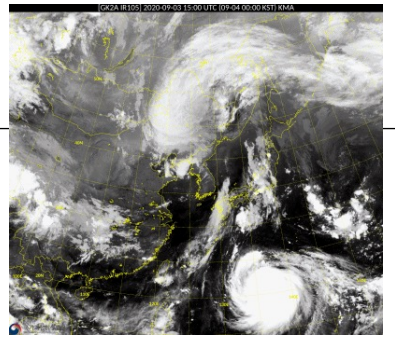
https://nmsc.kma.go.kr



Kim et al., 2021

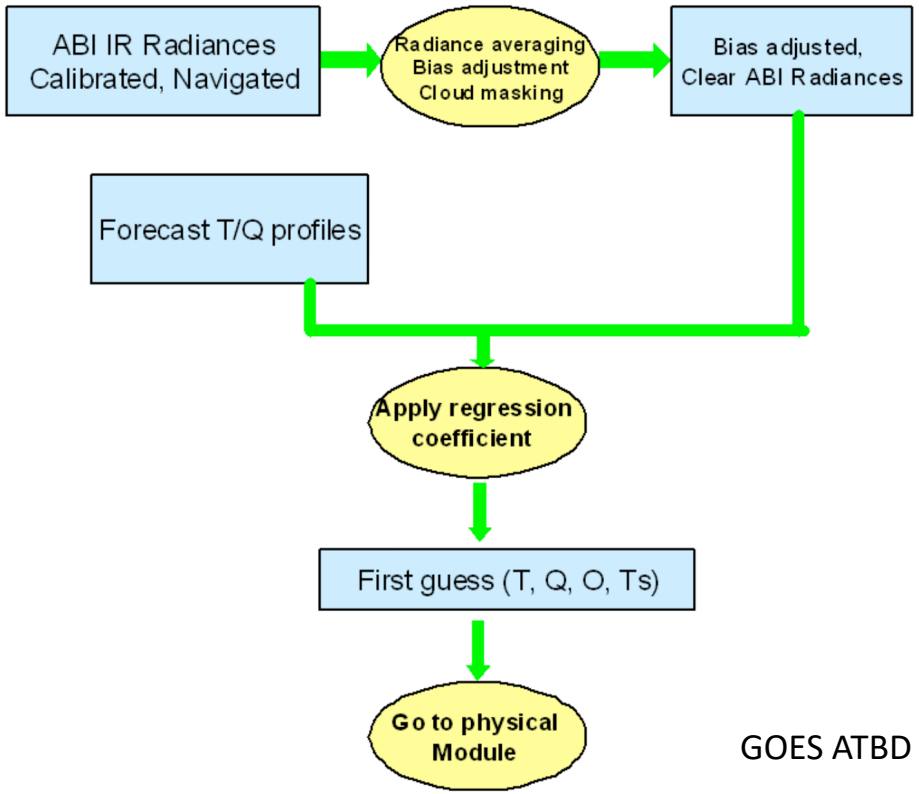
daehyeon.han@unist.ac.kr

Introduction



Satellite-based atmospheric profile

- NWP forecasting as a first guess -> correction in real time over cloud-free area.
- Based on radiative transfer simulation.



GOES ATBD

GOES legacy atmospheric profile algorithm

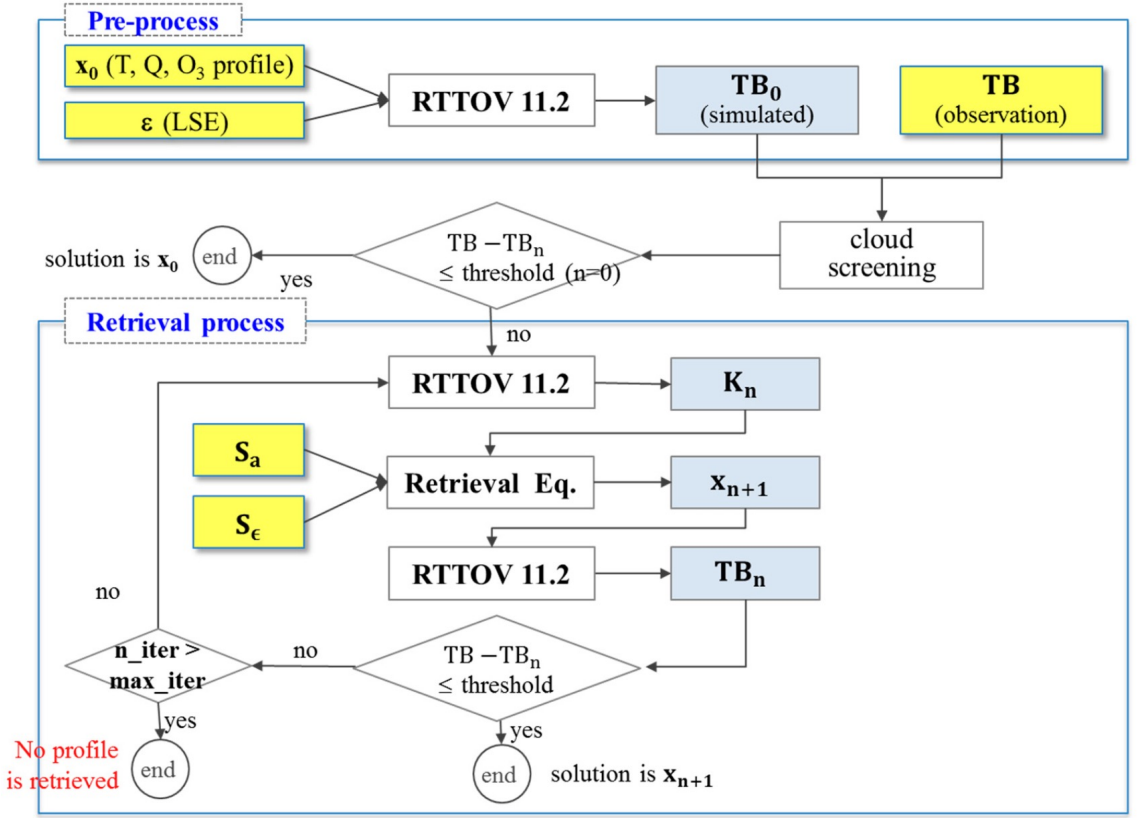


Figure 1. Algorithm flowchart for the retrieval of clear-sky atmospheric profiles.

GK2A atmospheric profile algorithm Lee et al., 2017

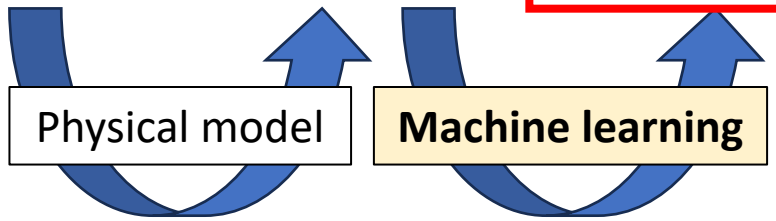
daehyeon.han@unist.ac.kr

Introduction

Motivation

- **The goal of this study is to generate reanalysis-like temperature profile using GEO data and machine learning.**

	Radiosonde	Reanalysis (ERA5)	NWP (KIM)	GEO (GK2A)	Reanalysis-like GEO
Accuracy	Ground truth	High	Medium	Medium-High	High
Spatial resolution	Point	Coarse	Medium	High	High
Temporal resolution	00, 12 UTC	High (1 h)	Medium-High (1-6 h)	High (1 h)	High (1 h)
Real time	Nearly	X	O	O	O



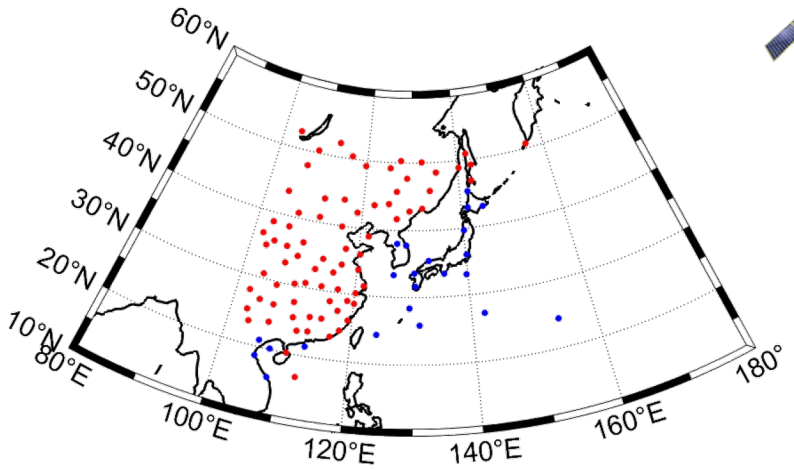
daehyeon.han@unist.ac.kr

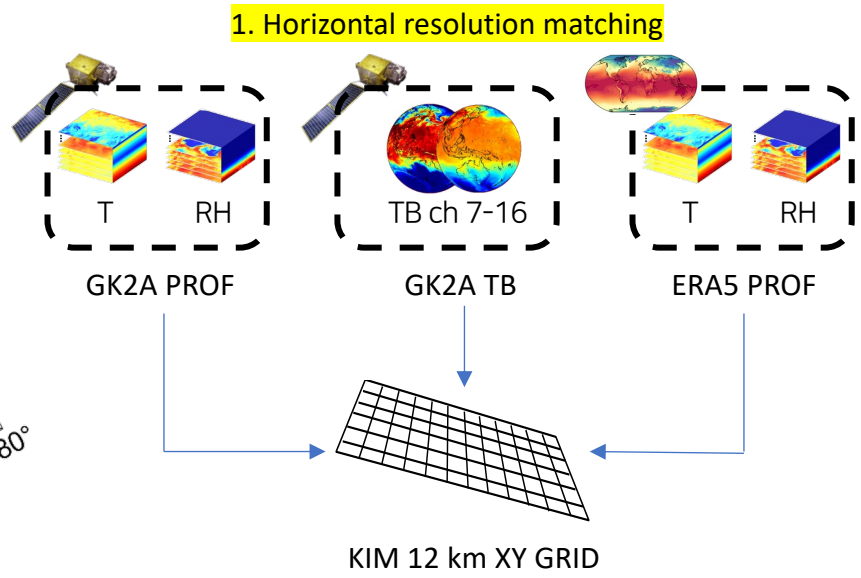
Data and method

Data and preprocessing

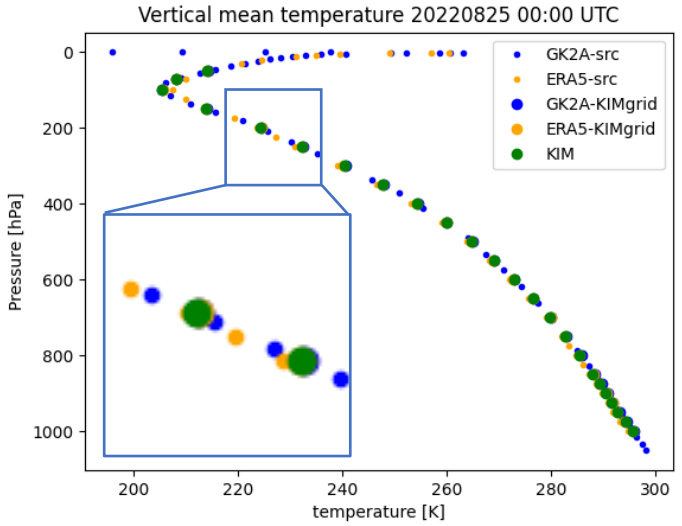
AAP: AMI Atmospheric Profile Retrieval
NMSC: National Meteorological Satellite Center
KMA: Korea Meteorological Administration

Data	Variables	Temporal Resolution	Spatial Resolution	Study area	Source
ERA5 Reanalysis Data	Temperature	1 hour	0.25 deg ~31 km	East Asia	ECMWF
GK2A L2 profile (AAP algorithm)	Temperature	1 hour	6 km		NMSC/KMA
GK2A L1B Brightness temperature (TB)	TB channel 7-16 (10 channels)	10 min	2 km		NMSC/KMA





2. Vertical resolution matching
KIM 16 levels, (spline interpolation)



daehyeon.han@unist.ac.kr

Data and method

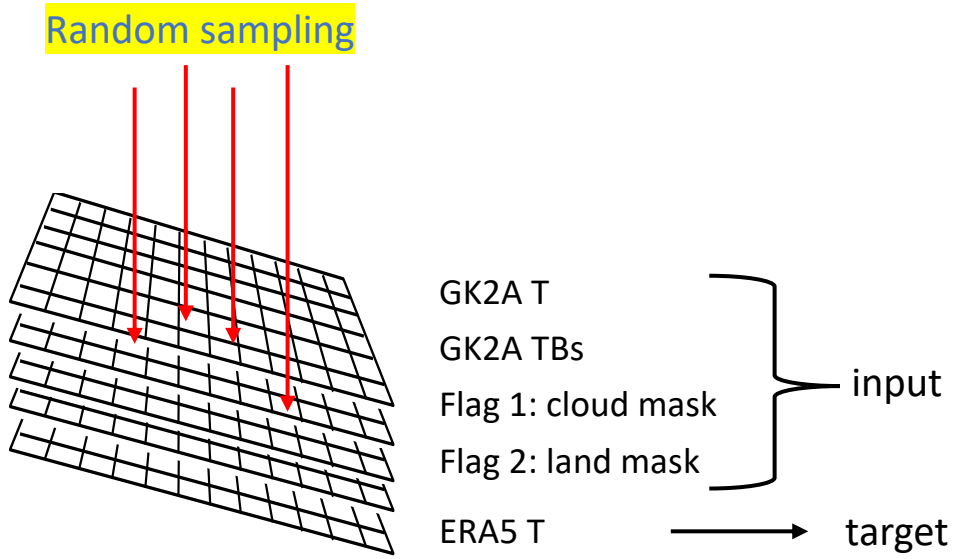
Sampling and modeling

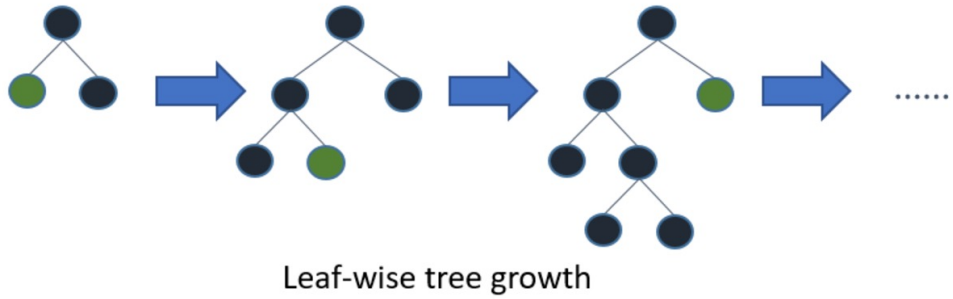
- Data period: 2021-2022, every month
- Training period : 1st – 20th, every hour 00-23
- Validation period: 21st – end of month, every hour 00-23
- 5000 points random sample per each scene

- Input variables
  - GK2A L2 profile (temperature)
  - GK2A L1B ch7-16 brightness temperature (TB)
  - Flag 1: Cloud mask (cloud, clear-AAP, clear-firstguess)
  - Flag 2: Land mask (land, water)

- Target: ERA5 profile (temperature)

- Model: Light gradient boosting machine (LGBM)
  - A boosting model that learns from mistakes.
  - Faster than other models, good for large datasets.
  - More prone to overfitting, so careful tuning is needed.



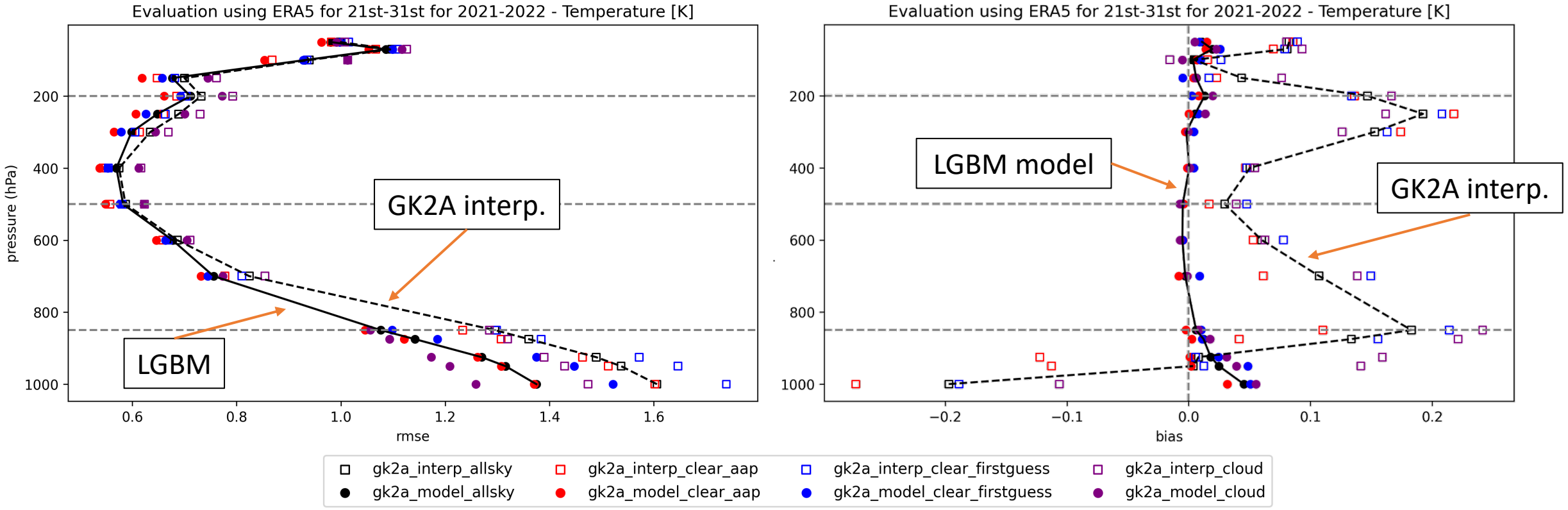


daehyeon.han@unist.ac.kr

Results

Temperature (with Flag1: clear_aap, clear_firstguess, cloud)

- Comparison of vertical temperature (T) distribution across KIM levels (Average of validation data from all 21-31 days in 2021-2022)



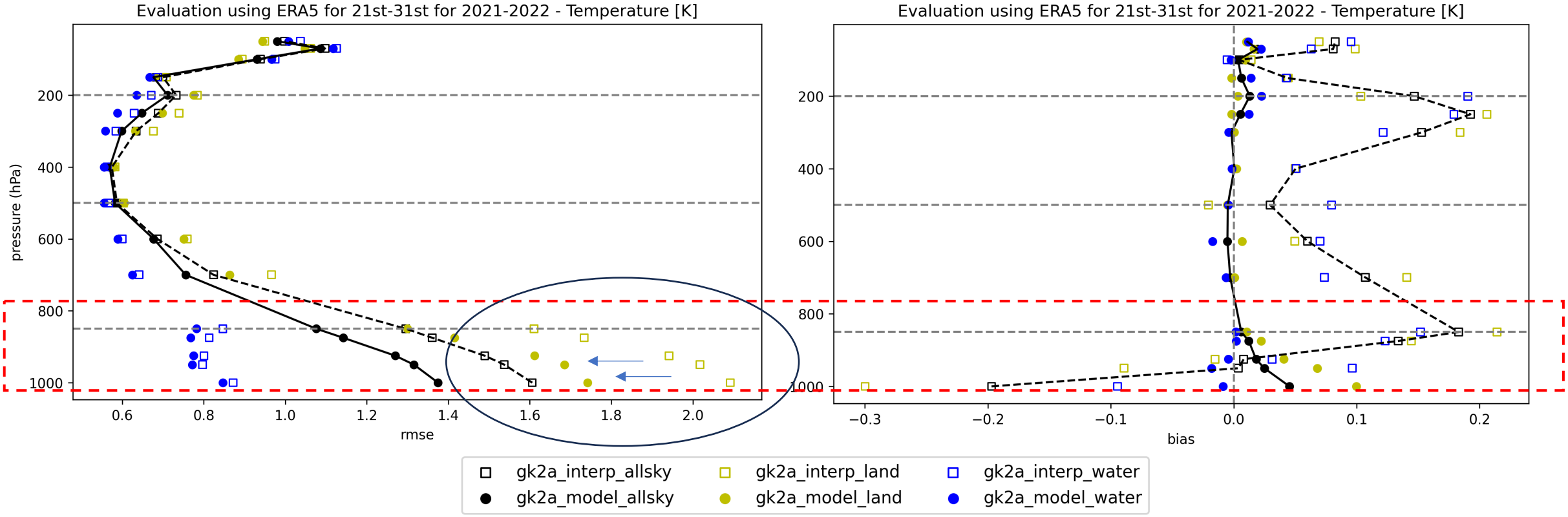
- The corrected model (gk2a_model) is more aligned with ERA5 in temperature than the simple interpolation (gk2a_interp).
- RMSE improves significantly in the lower atmospheric layers (800-1000 hPa) when using the corrected model.
- The corrected model reduces vertical average bias with ERA5.

daehyeon.han@unist.ac.kr

Results

Temperature (with Flag2: land/water)

- Comparison of vertical temperature (T) distribution across KIM levels (Average of validation data from all 21-31 days in 2021-2022)

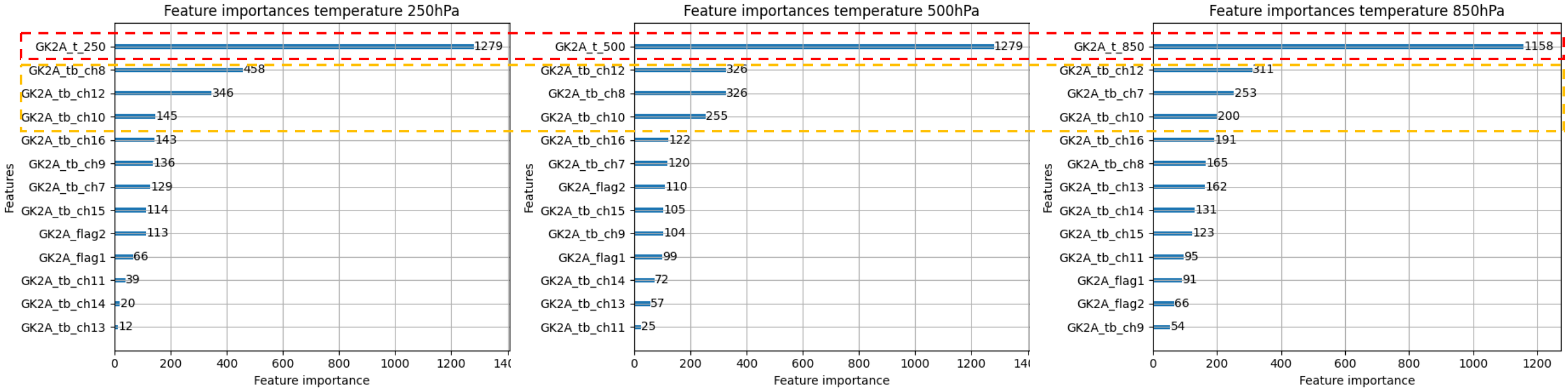


- RMSE improves more clearly over land than over the water in the lower atmosphere.
- RMSE is significantly larger over land compared to the water.
  - Using DEM data as input could lead to further improvements.

daehyeon.han@unist.ac.kr

Results

Feature importance for temperature

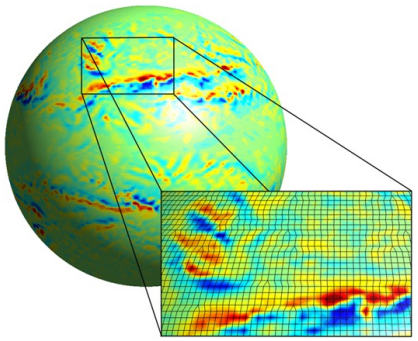


- Input GK2A temperature scored significantly high feature importance in LGBM model over all three levels.
- TB of Ch. 8, 10 (water vapor) and 12 (IR) showed relatively higher feature importance.
- Flag 1 & 2 showed low importance.

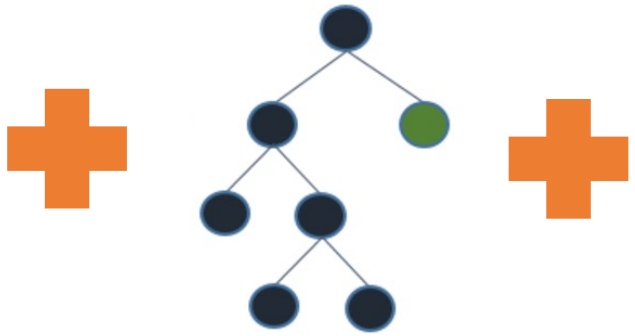
daehyeon.han@unist.ac.kr

Summary

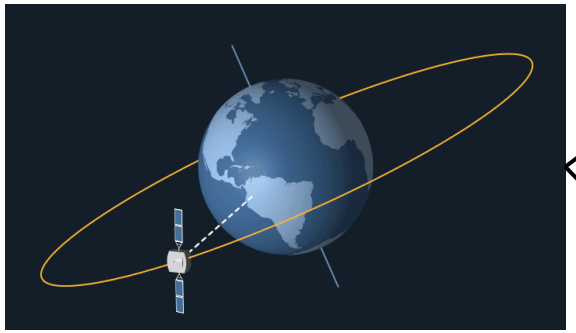
- A real-time correction of temperature profile using GK2A, ERA5 and machine learning.
- In terms of RMSE and mean bias, the suggested model showed improvement.
- More analysis will be conducted.
  - Seasonal, diurnal cycle, adding DEM, sonde comparison…
- When use the original spatial resolution, it is expected to conduct downscaling at the same time.



Historical ERA5 Reanalysis data

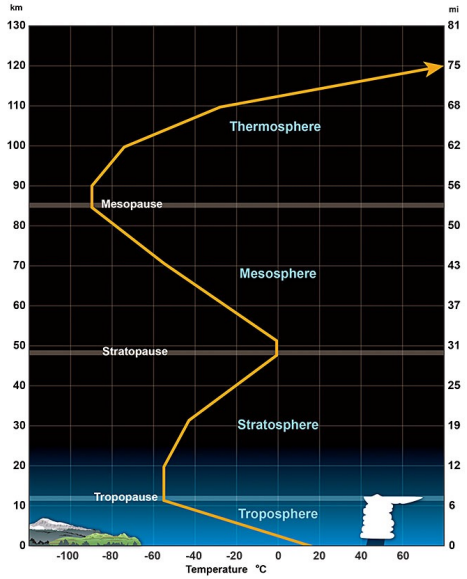


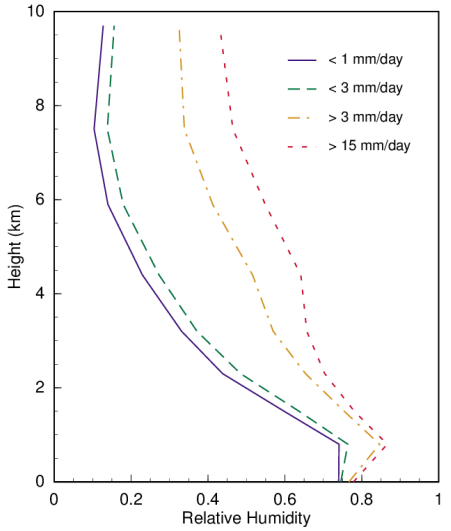
Machine learning



GEO data – GK2A

Real-time corrected Temp. profile (ERA5-like)





daehyeon.han@unist.ac.kr

Thank you!

daehyeon.han@unist.ac.kr