Autoregressive Atmospheric Profile Data Assimilation using Diffusion Model

Minki Choo and Jungho Im

[1]Department of Civil, Urban, Earth, and Environmental Engineering, Ulsan National Institute of Science and Technology, Ulsan , S. Korea

(mazera1995@unist.ac.kr)

105TH ANNUAL MEETING 12–16 JANUARY NEW ORLEANS & ONLINE AMERICAN METEOROLOGICAL SOCIETY

ID# 458804

Introduction

- Weather forecasting is a critical component in numerous scientific and practical domains, such as aviation, logistics, agriculture, and disaster management.
- Existing deep learning-based studies often treat weather forecasting and data assimilation as separate problems. Integrated approaches that combine both forecasting and assimilation into a unified framework remain rare.
- In response to these challenges, we propose an autoregressive data assimilation framework that integrates deep learning-based diffusion models to assimilate diverse observational data sources in real time.

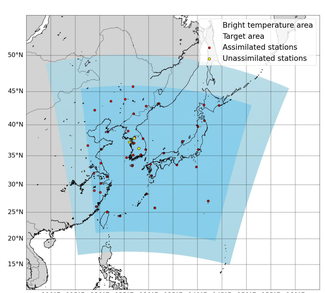
Data & Study Area

Data Product

Data	Variables	Temperal Resolution	Spatial Resolution	Source	Note
GK2A AAP	Temperature	10 minutes	0.06 degree	NMSC	For model input
GK2A Brightness Temperature	WV063, WV069, WV073, IR087, IR096, IR105, IR112, IR123, IR133	10 minutes	0.02 degree	KMA	For model input
ERA5	Temperature	1 hour	0.25 degree	ECMWF	For model training
Radiosonde	Temperature	12 hours	Point	NOAA ESRL	For conditional input and validation

- The clear sky part of the GK2A AAP is retrieved with brightness temperature data and the cloud region is filled with UM forecast data. Only clear sky data is used for forecasting and assimilation in this study.

Study Area

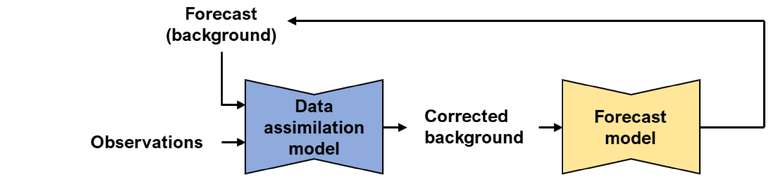


- **Study area**: East Asia
- **Study period**: 2020–2024
- **Test period**: Last 7 days of each month
- **Test variable**: Temperature (T)
- **Test layer**: 100, 150, 200, 250, 400, 500, 700, 850, 925, 1000 hPa

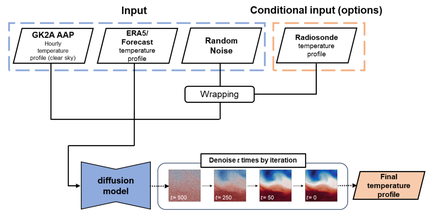
- GK2A AAP was resized to 12 km spatial resolution due to computational cost.
- Brightness temperature uses a larger area as input than GK2A AAP temperature.
- All data are interpolated to match pressure levels and geo-coordinates.

Method

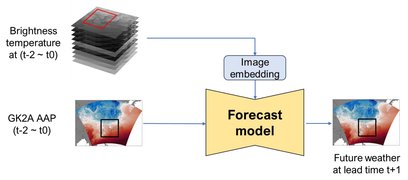
Proposed Flow



Assimilation model

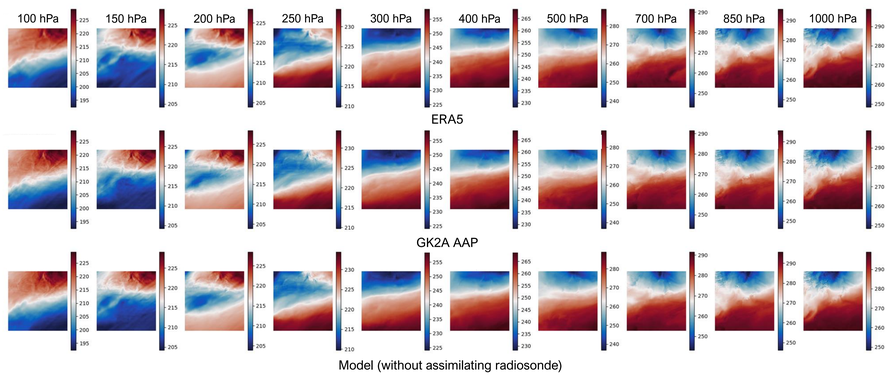


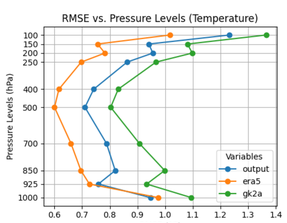
Forecast model



Result

Analysis results using ERA5 as a background





- The output of the model demonstrates spatial resolution comparable to that of GK2A. To evaluate the impact of the assimilation model, validation was conducted using radiosonde observations during the test period. Samples with a temperature difference exceeding 4°C were excluded from the analysis.
- The proposed model exhibited higher overall accuracy compared to the GK2A AAP across the entire pressure levels. Notably, the improvements were more pronounced in the lower atmospheric layers than in the upper layers.

Result

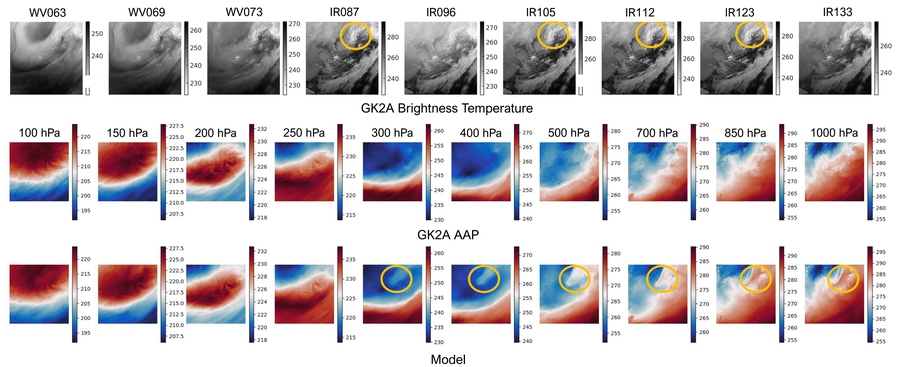
Validation on Intensive Observation Periods (2021–2022)

RMSE	Level (hPa)	Number of samples	GK2A AAP	ERA5	Model
T(°C)	100	253	1.14	**0.97**	1.05
	150	253	0.83	**0.72**	0.77
	200	255	0.79	**0.67**	0.75
	250	254	0.68	**0.56**	0.62
	400	255	0.70	0.70	**0.69**
	500	253	0.69	**0.65**	**0.65**
	700	254	0.82	**0.70**	0.75
	850	255	0.74	**0.71**	0.77
	925	254	0.77	0.75	**0.74**
	1000	112	1.38	**1.27**	1.29

- Results generated using ERA5 as a background and validated with radiosondes from stations not used in the assimilation during the 2021 and 2022 intense observation periods.
- The model outperformed the GK2A AAP and was more accurate than ERA5 in some pressure levels.

Autoregressive generation result

- Temperature data were generated autoregressively for the period from 25 to 31 January 2020. Only the initial state utilized GK2A AAP, while subsequent steps relied on GK2A AAP clear sky data and GK2A brightness temperature.
- At the 120th generation step, the middle and upper pressure levels resembled GK2A AAP, but observational influences triggered a butterfly effect, propagating errors across pressure levels. The brightness temperature also contributed to the inaccuracies, which will be addressed in future improvements.



Conclusion

- This study tried to introduce a deep learning-based data assimilation framework that does not require numerical products except for the initial states.
- In our assimilation model, the incorporation of ERA5 as a background field was able to improve the sonde accuracy with minimal resolution degradation.
- However, the autoregressive results still need improvement. As the forecast process gets older, the so-called butterfly effect can cause the model to produce erratic forecasts. Further improvements are needed.
- In the future, we will utilize additional microwave-based sensor data (ATMS, AMSU-A) to improve these errors.