# Generative AI for Weather Data Assimilation

Ruizhe Huang, Qidong Yang,
Jonathan Giezendanner, Sherrie Wang

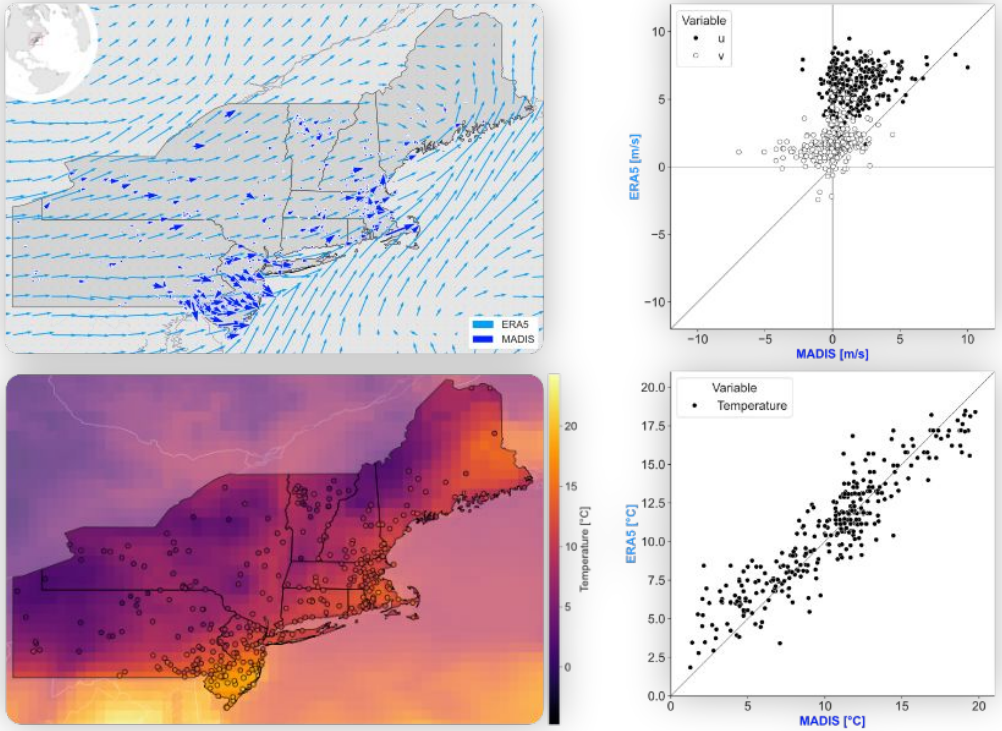
# Global reanalysis product doesn’t always match station observations

Wind

- Turbulences ignored by global weather products (ERA5)
- Obstacles such as buildings and forests are smoothed out

Temperature

- ERA5 is good, but too smooth compared to local observations

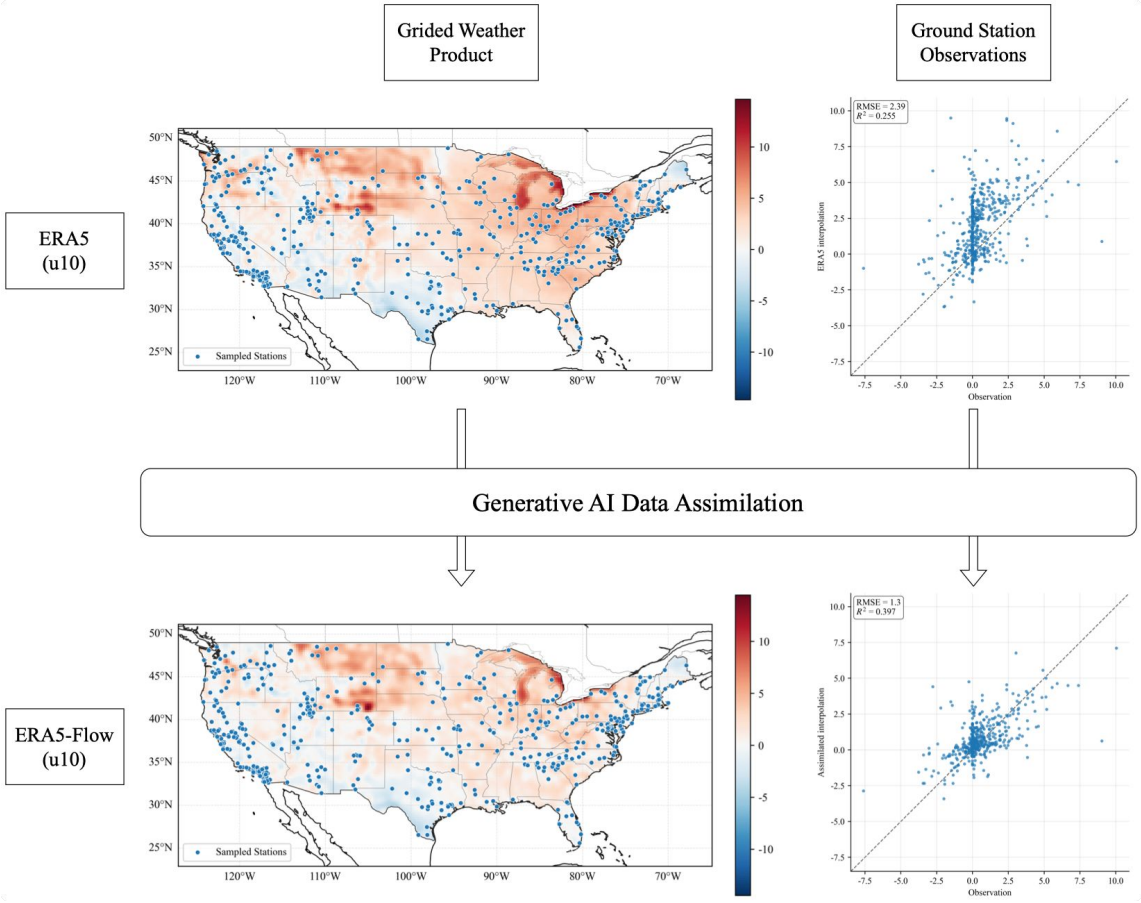



Global Model vs Local Weather Stations

# **Our Work:** Generative AI for Weather Data Assimilation

GenAI data assimilation doesn't have assumptions that traditional data assimilation has:

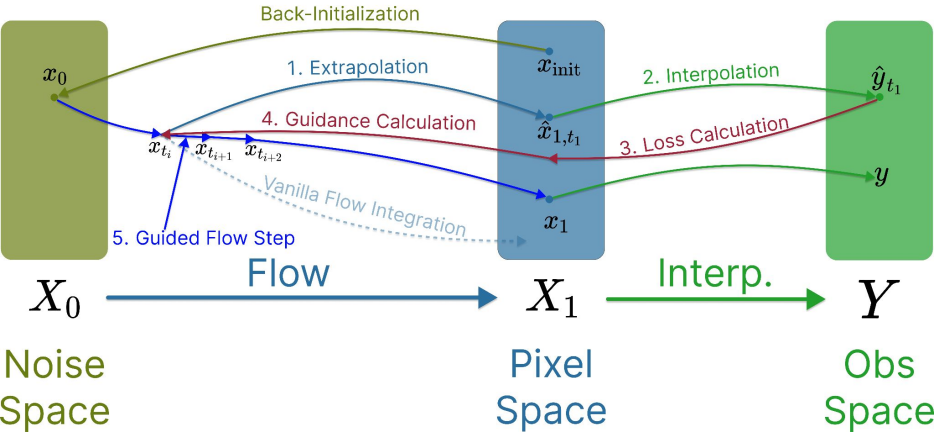
- Near-Gaussian errors
- Linearized dynamics



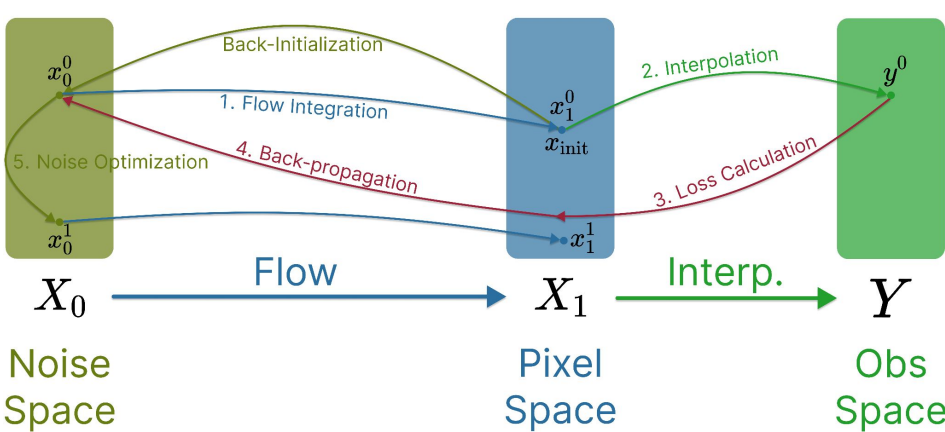
# We benchmarked 6 generative AI methods

- Guidance++ (improved version of guidance)
- CD-Flow (improved version of dflow)
- Guided CD-Flow (CD-Flow with guidance)
- Latent Guidance++
- Latent CD-Flow
- Guided Latent CD-Flow

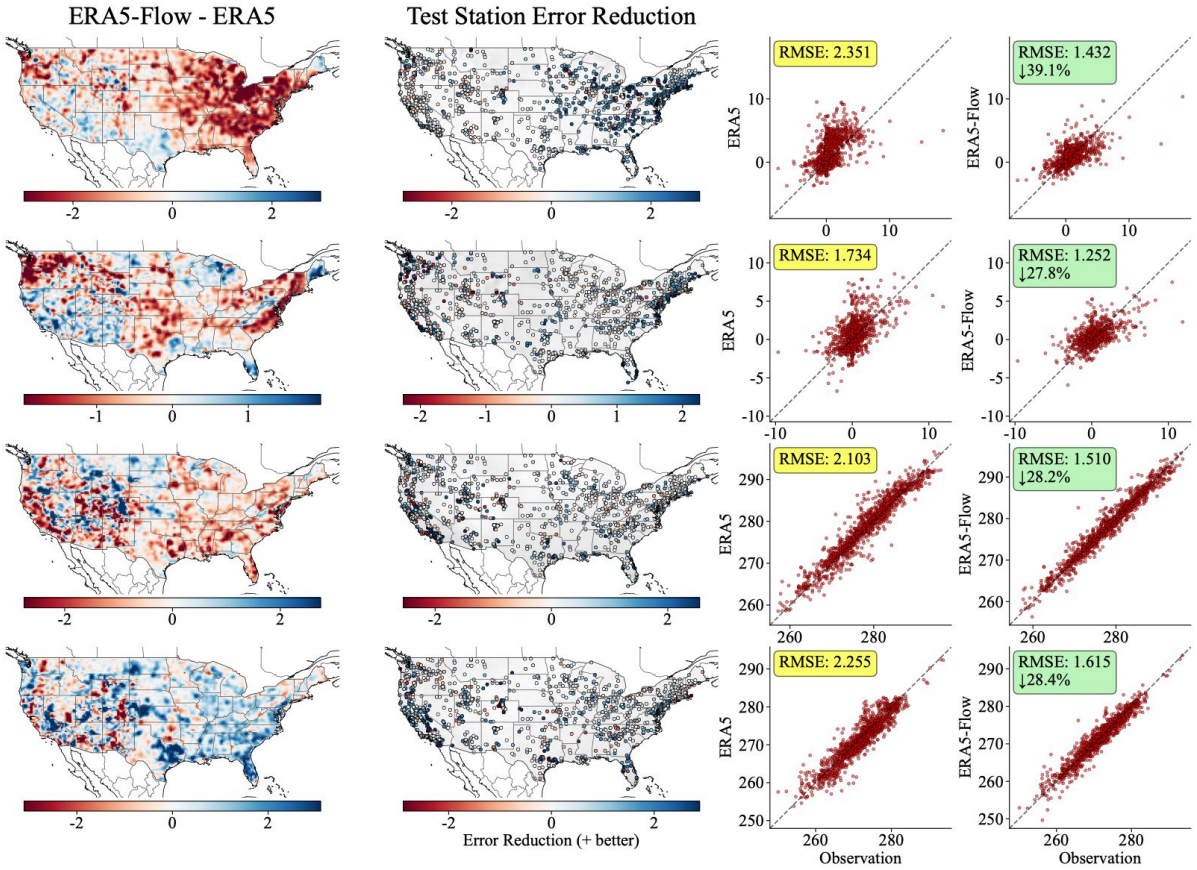
## Guidance++




## CD-Flow



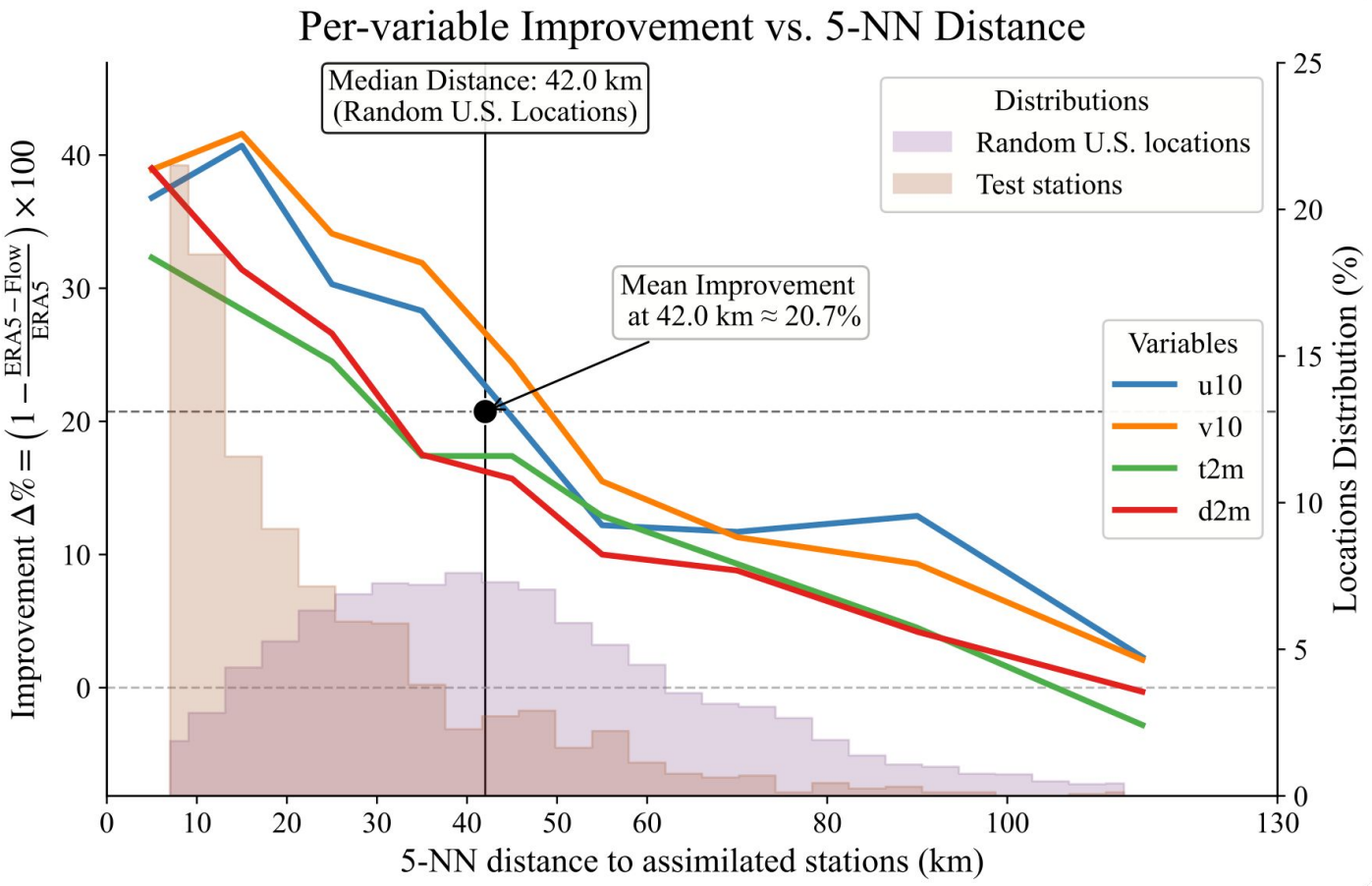
# ERA5-Flow reduces ERA5 error by >30% on test stations



# Guidance++ captures station variance well

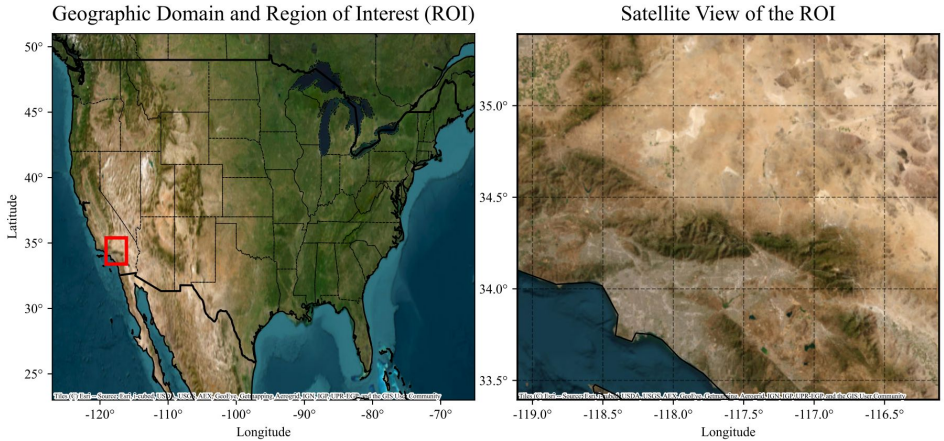
Computational cost:

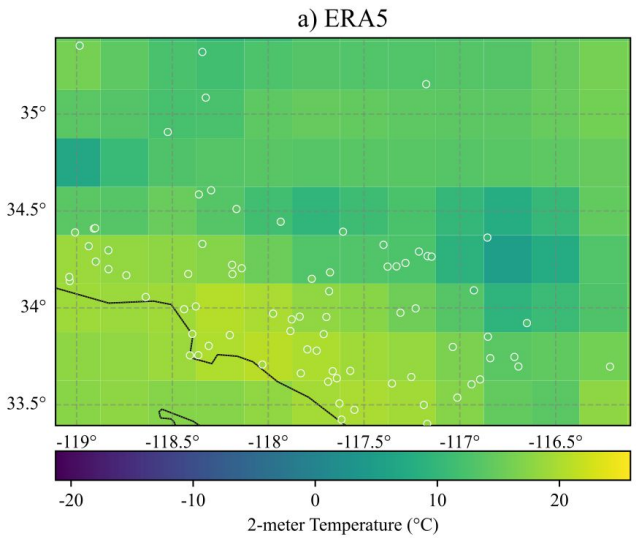
- 365 snapshots
- < 0.5 GB memory
- < 5 minutes

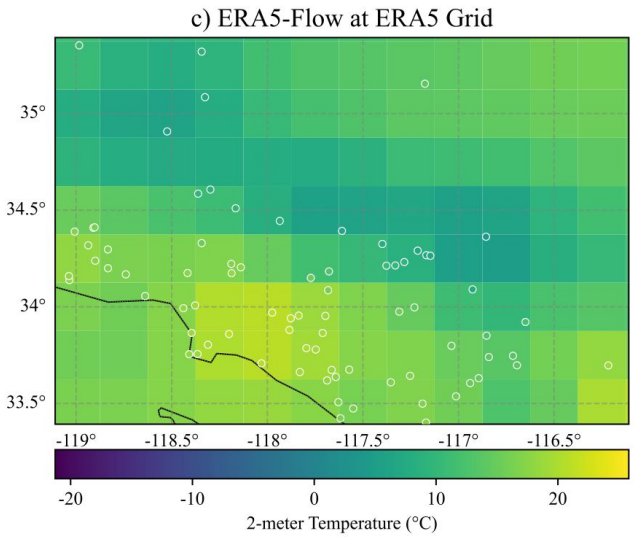


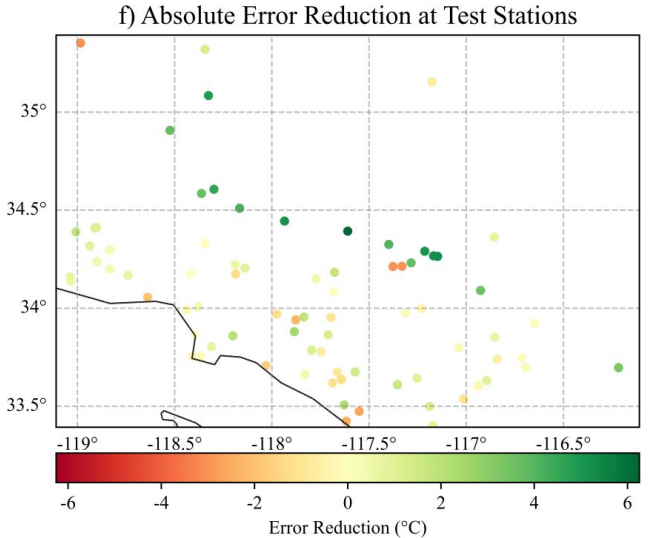
# Case Study

In the ERA5, the mountains are not cold enough, and the cities are not hot enough. These biases are corrected by Guidance++.



Thanks!