IMPERIAL

Duo-AttnOPNets: An End-to-End ML–4D-Var Framework for Atmospheric Composition Forecasting and Data Assimilation

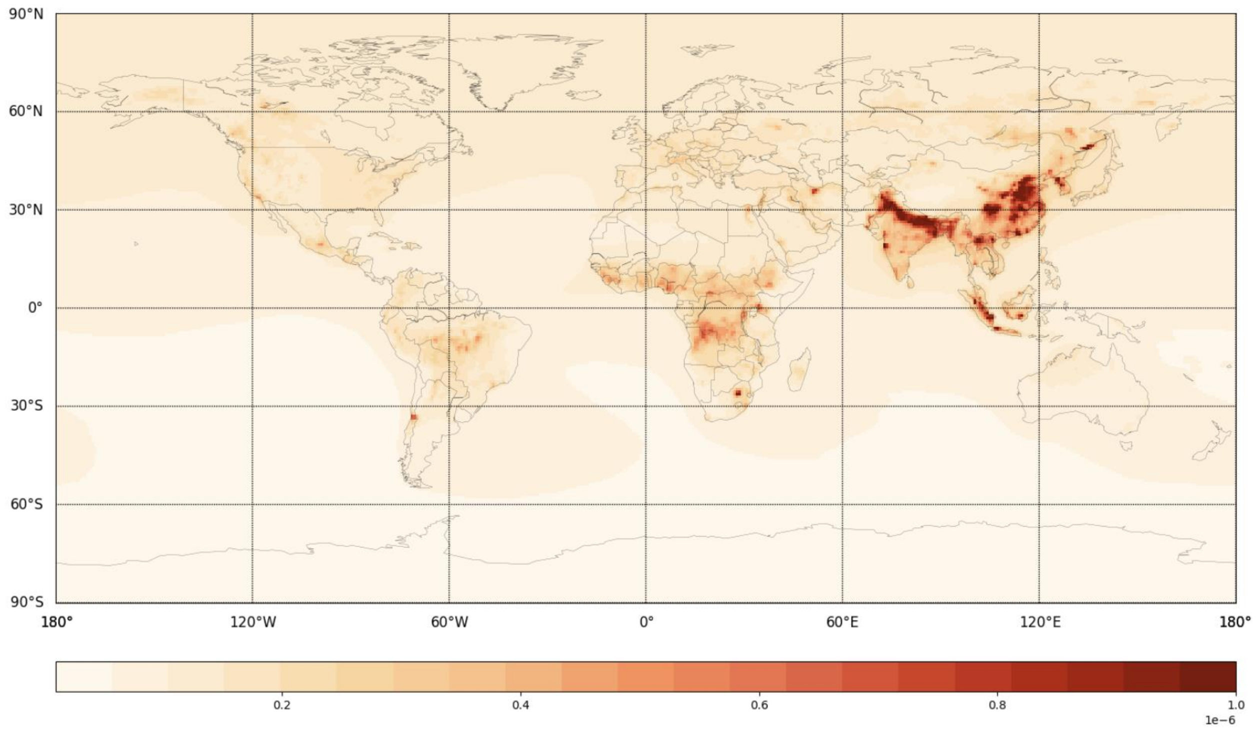
Judongyang (Judy) Zhou
Department of Earth Science and Engineering
Imperial College London

Part 1

Background

Why CO as a Testbed?

- Strongly influenced by **emissions and transport**
- Large **spatial and temporal variability**
- Sensitive to both **model dynamics and observations**
- Routinely monitored in **operational systems (e.g., C-IFS by CAMS)**



Averaged CO Concentration at surface level in 2018

Current Solutions & Limitations

01 Chemistry Transport Models

Examples: C-IFS (by CAMS); GEOS-Chem; WRF-Chem

Strengths:
- Physically grounded
- Mature operational systems

Challenges:
- PDE integration cost
- Sensitive parameter tuning
- Adjoint development burden
- Maintenance complexity

02 Purely AI-based Models

Typical Types:
- Forecast surrogates (reanalysis -> reanalysis)
- Learned DA models (observation -> reanalysis)

Limited scalability in DA:
- Changing assimilation window
- Adding new observation types
- Modifying the forward model

⇒ retraining required

Part 2

Methodology

Conceptual Framework

Goal: Develop a scalable framework for **AI-empowered variational data assimilation and forecasting**

Core Idea: To combine

- A **data-driven surrogate forecast model**
- A **training-free variational assimilation module**

➡ To enable efficient DA–forecast cycling

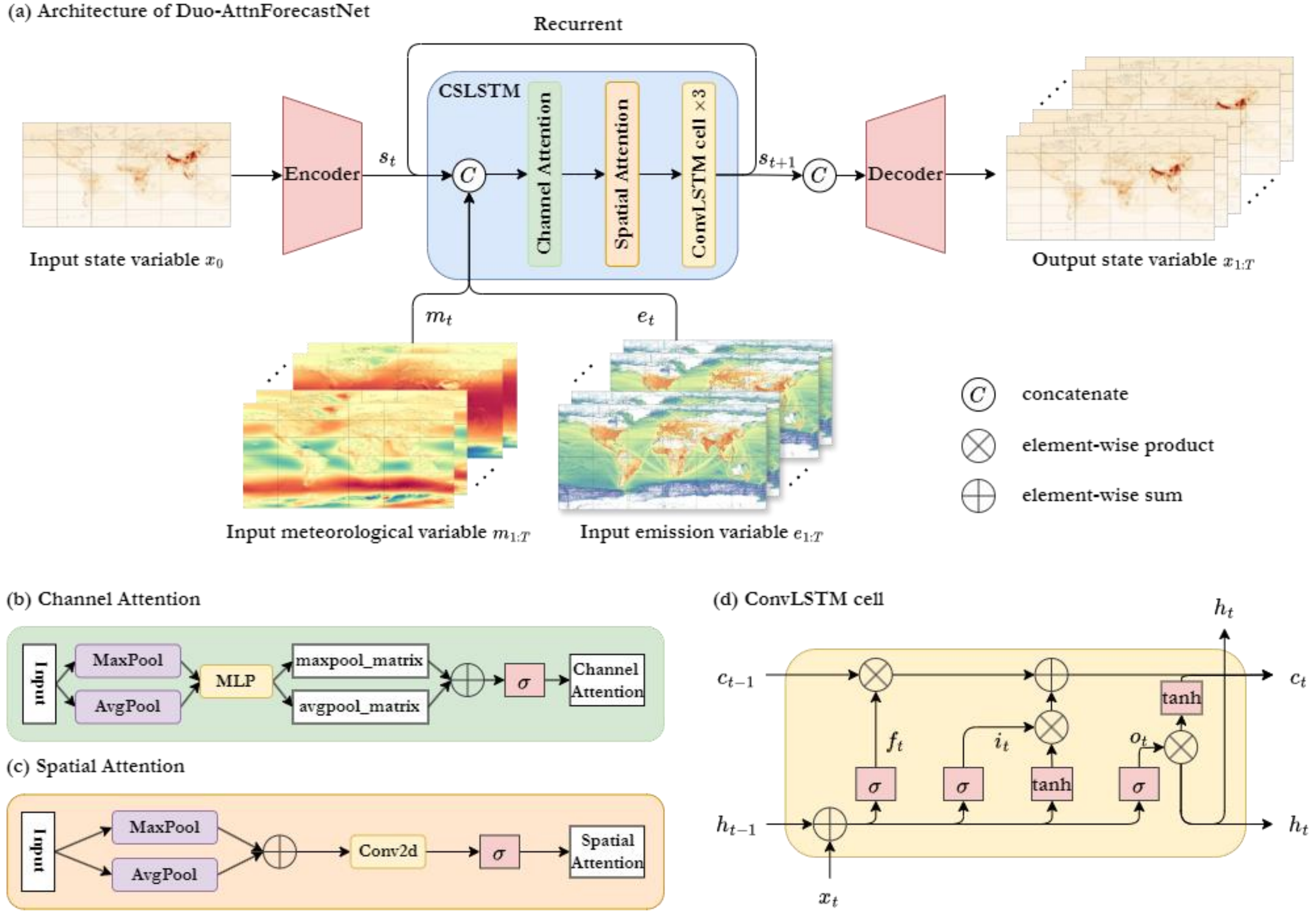
Operational Logic:

Forecast → Assimilation → Updated forecast → Cycling

➡ AI-enabled implementation of the DA-forecast cycle

IMPERIAL

Forecast Module – Model Architecture



Data-driven Surrogate of Transport-Chemistry Dynamics

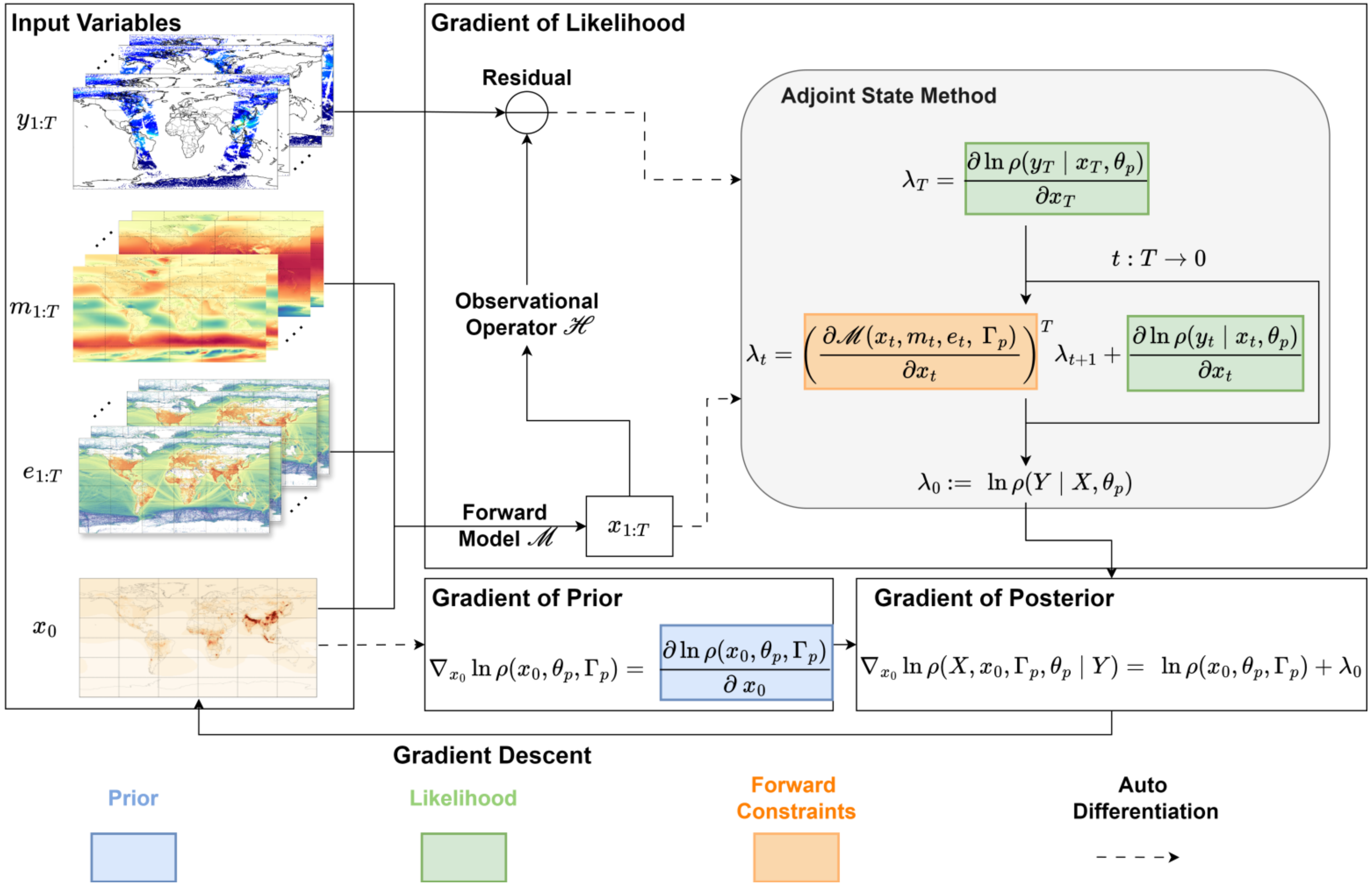
Inputs:

- CO state
- Meteorology
- Emissions

Architecture:

- ConvLSTM backbone
- Dual attention
- Encoder–decoder

AI-Empowered 4D-Var



Core idea:

➢ Replace manual adjoint with auto-diff

Advantages:

- Exact gradients
- GPU acceleration
- No tangent-linear and adjoint coding
- No retraining needed

Same Philosophy, Different Implementation

Part 3

Experiments

Experimental Design

Global Experiments (2018)

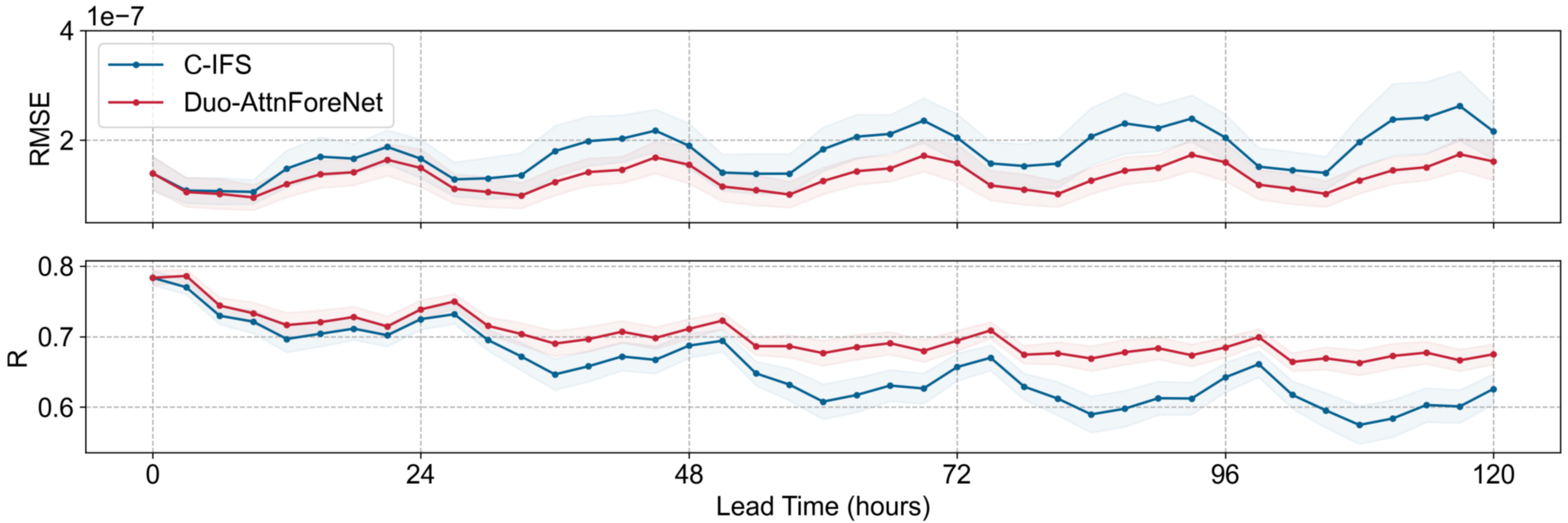
Designed to ensure a like-for-like comparison with C-IFS.

Fair Comparison Protocol

- Same drivers as C-IFS
- Same satellite observations
  - MOPITT and IASI
- Identical assimilation configuration where applicable
- Error covariance matrices:
  - Univariate and time-invariant
- Satellite data quality control

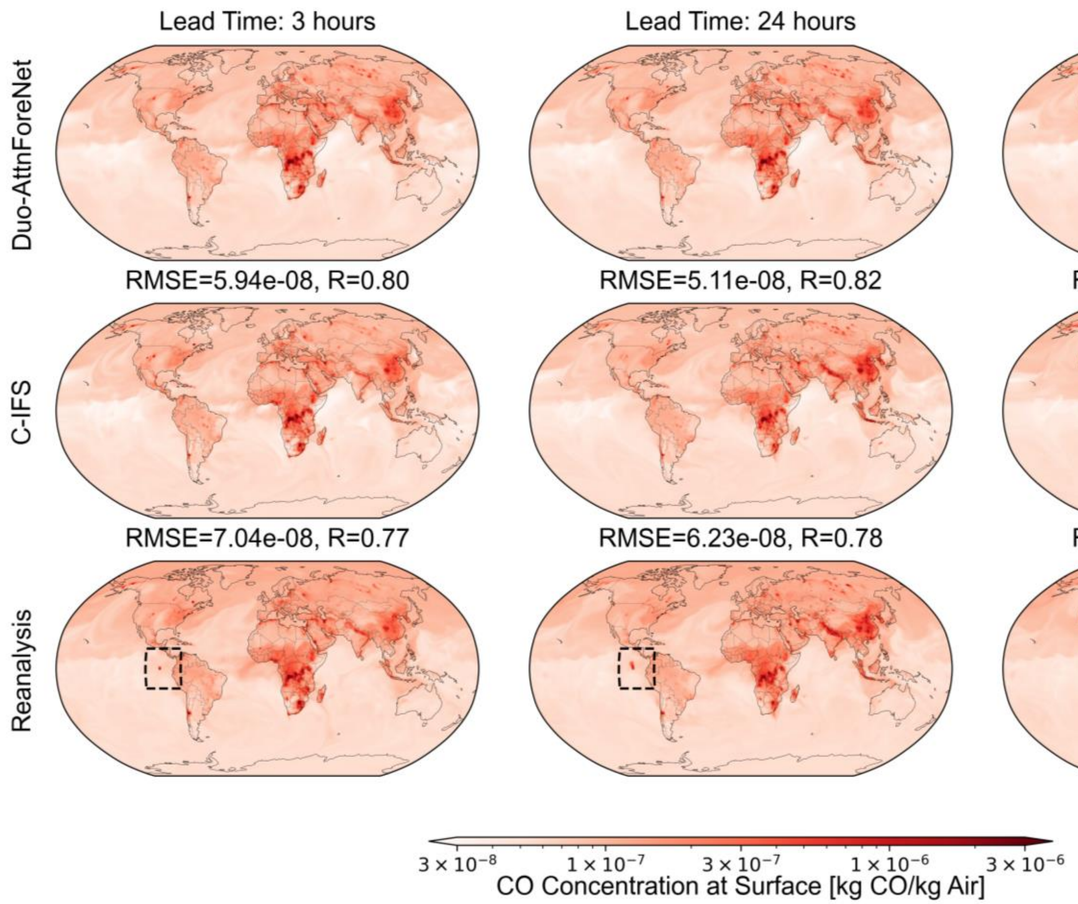
Data Type	Source	Details	Resolution
Emission	GFAS v1.2	CO fire emissions	0.1°×0.1°, daily
Emission	CAMS-GLOB-BIO v3.1	CO biogenic fluxes	0.25°×0.25°, monthly
Emission	CAMS-GLOB-ANT v2.1	CO anthropogenic emissions	0.1°×0.1°, monthly
Reanalysis	CAMS global reanalysis (EAC4)	1. 10m u-wind 2. 10m v-wind 3. 2m temperature 4. Surface geopotential 5. Total column water vapour 6. CO (model level=60)	0.75°×0.75°, 3-hourly
Forecast	CAMS forecast	Same above	0.4°x0.4°, 3-hourly

Forecast Skill vs C-IFS



The two plots present RMSE (top) and Pearson Coefficient (R) (bottom) in different lead times. For each metric, the mean value (solid lines) and the 95% confidence interval (shaded regions) are computed at each lead time across all samples from 2018.

Case Studies



An illustration of the surface-level CO concentration forecasted by Duo-AttnForeNet and C-IFS, and the corresponding EAC4 reanalysis. The forecasts are initialized at 0:00 UTC on 17 June 2018.

What We Observe

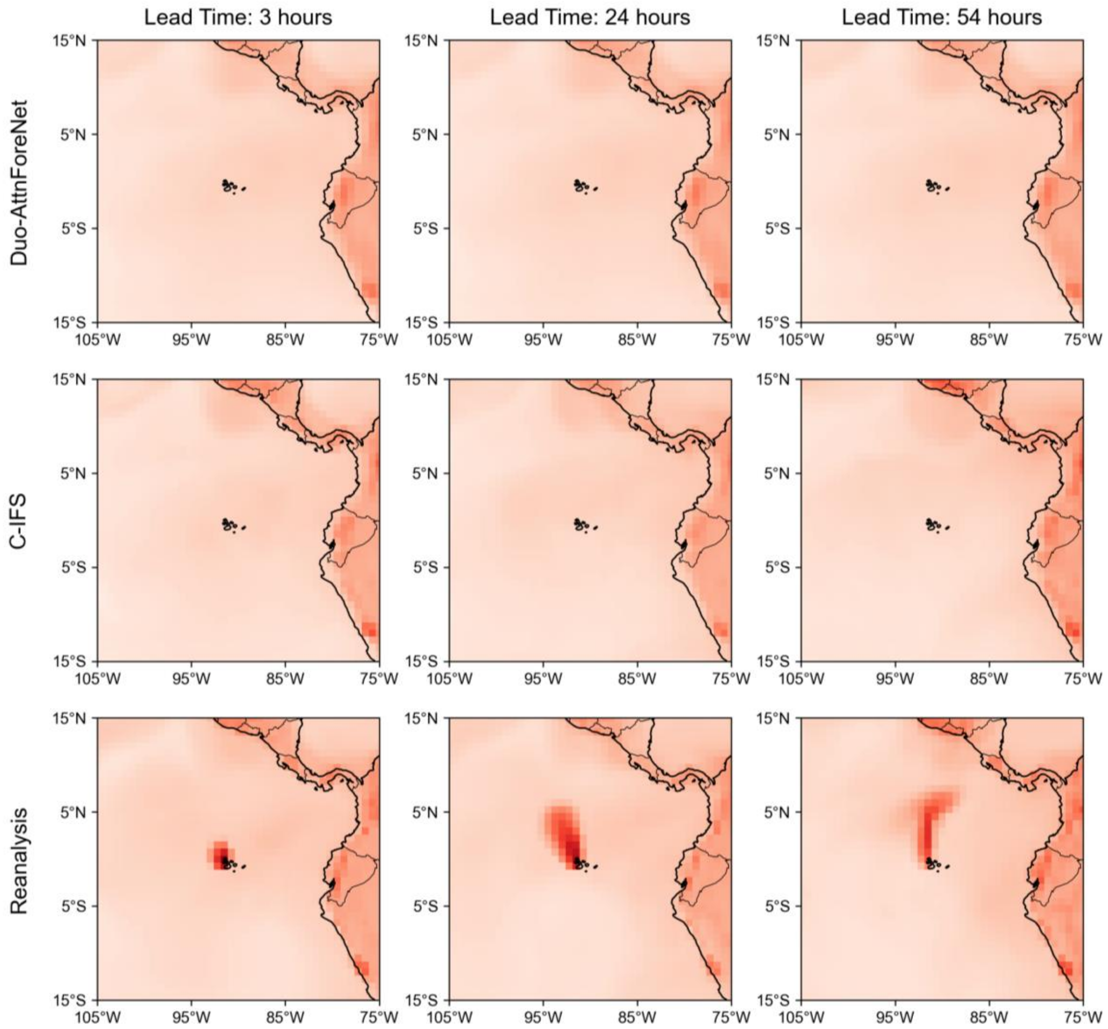
- ✓ Captures meteorologically driven transport
- ✓ Reproduces high-CO regions linked to emissions
- ✓ Spatial coherence comparable to C-IFS

But Challenges Remain

Difficult-to-predict events:

- Volcanic eruptions
- Wildfires
- Emission uncertainties

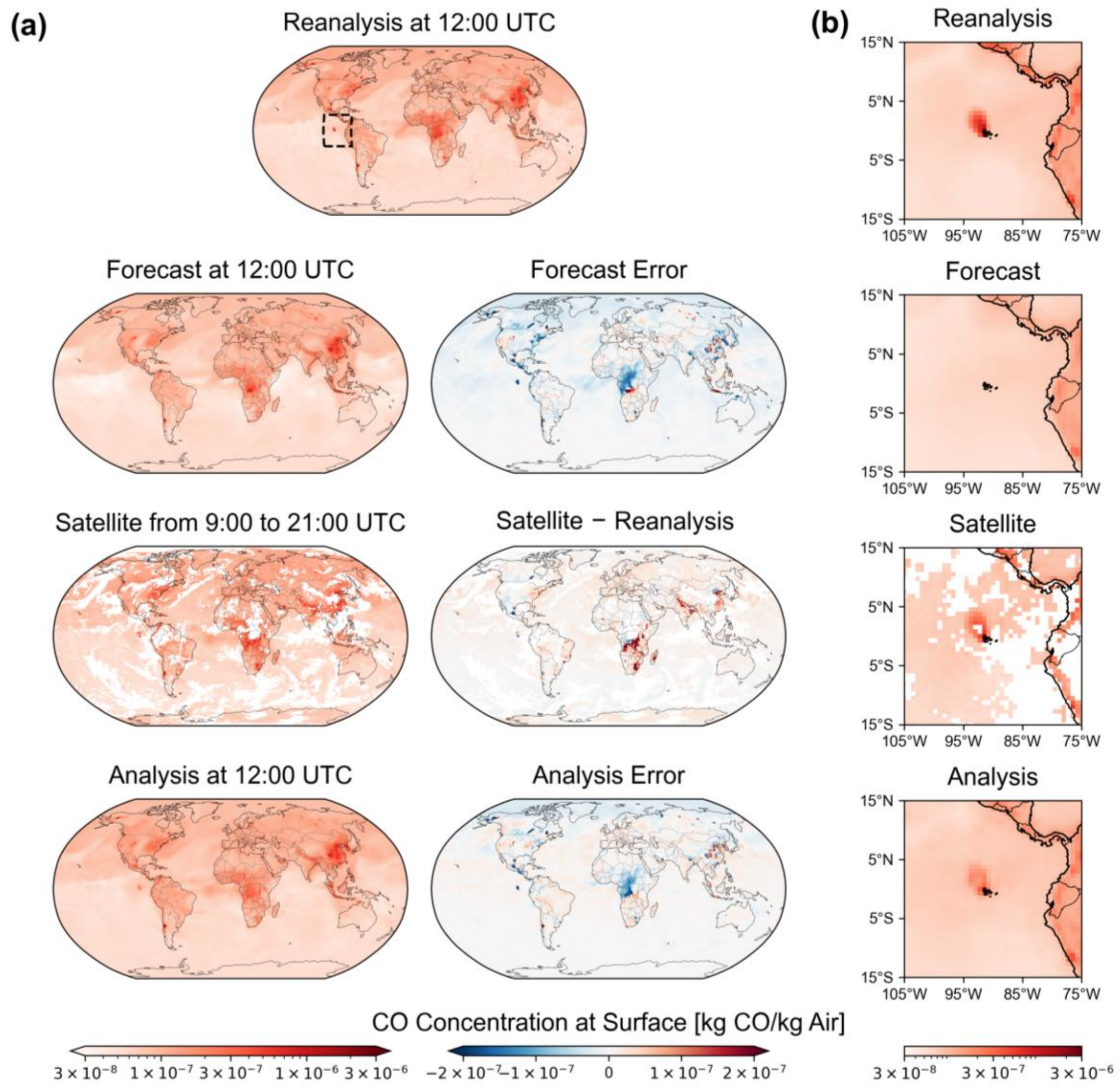
Data Assimilation is Needed



Zoom in around the **Galápagos Islands**, where a sudden CO plume associated with the **La Cumbre Volcano eruption occurred**.

This is exactly where data assimilation becomes critical, as it corrects the state using real-time observations.

Assimilation Impact: Correcting the Current State



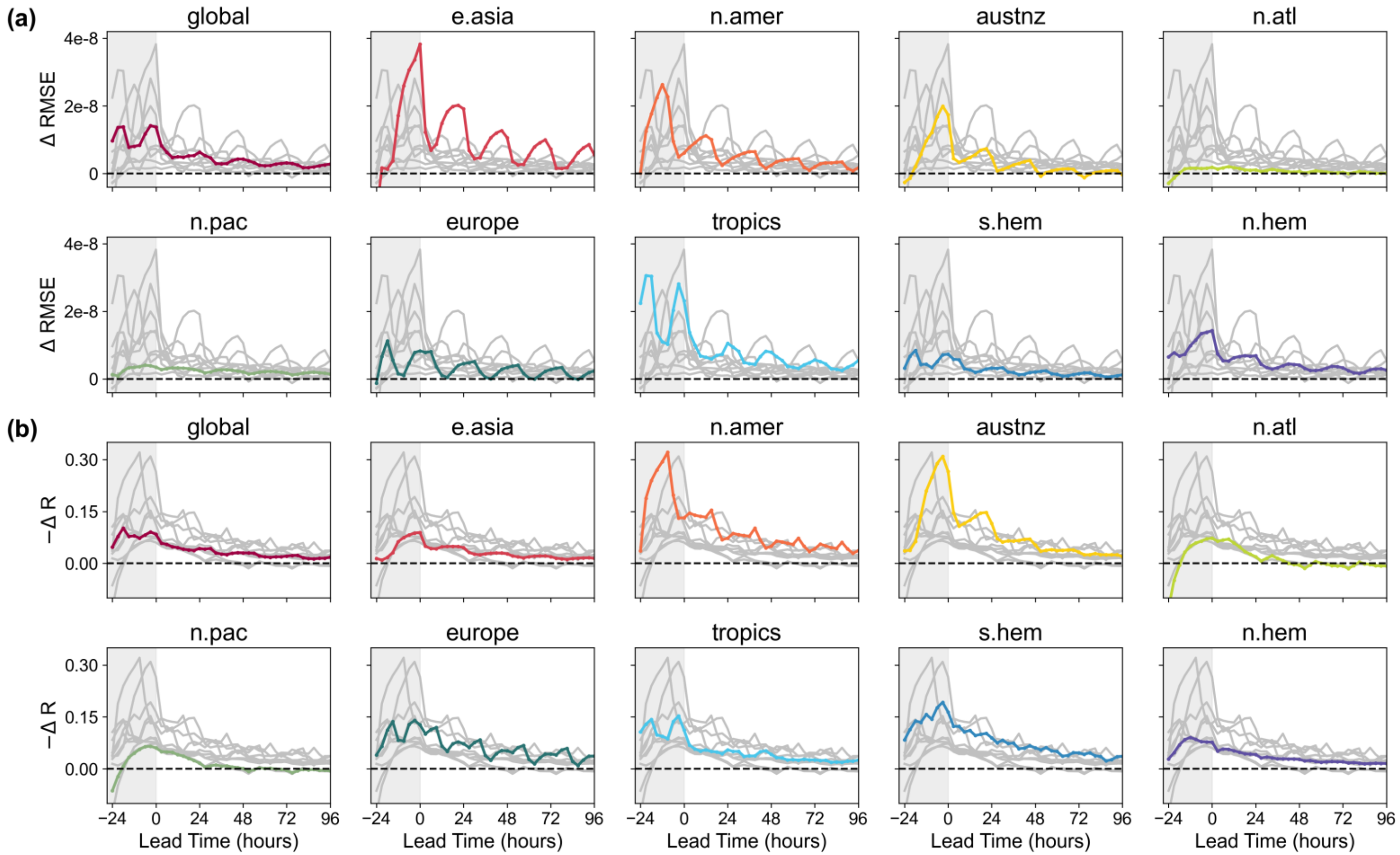
Impact of Duo-AttnVarNet:

- Reduce large regional biases
- Reduce errors over emission-sensitive regions
  - e.g., U.S. East Coast
- Improves consistency over oceans and tropics

Example: Volcanic CO Plume

- Satellite detects eruption signal
- DA reconstructs high-CO plume
- Works despite partial cloud contamination

Regional & Temporal DA Impact



The grey shaded areas indicate the assimilation window.

Temporal Behavior:

- Strongest impact within first 24 h
- Benefits gradually decrease with lead time
- By ~48 h, differences become smaller

Regional Behavior:

DA impact is spatially heterogeneous

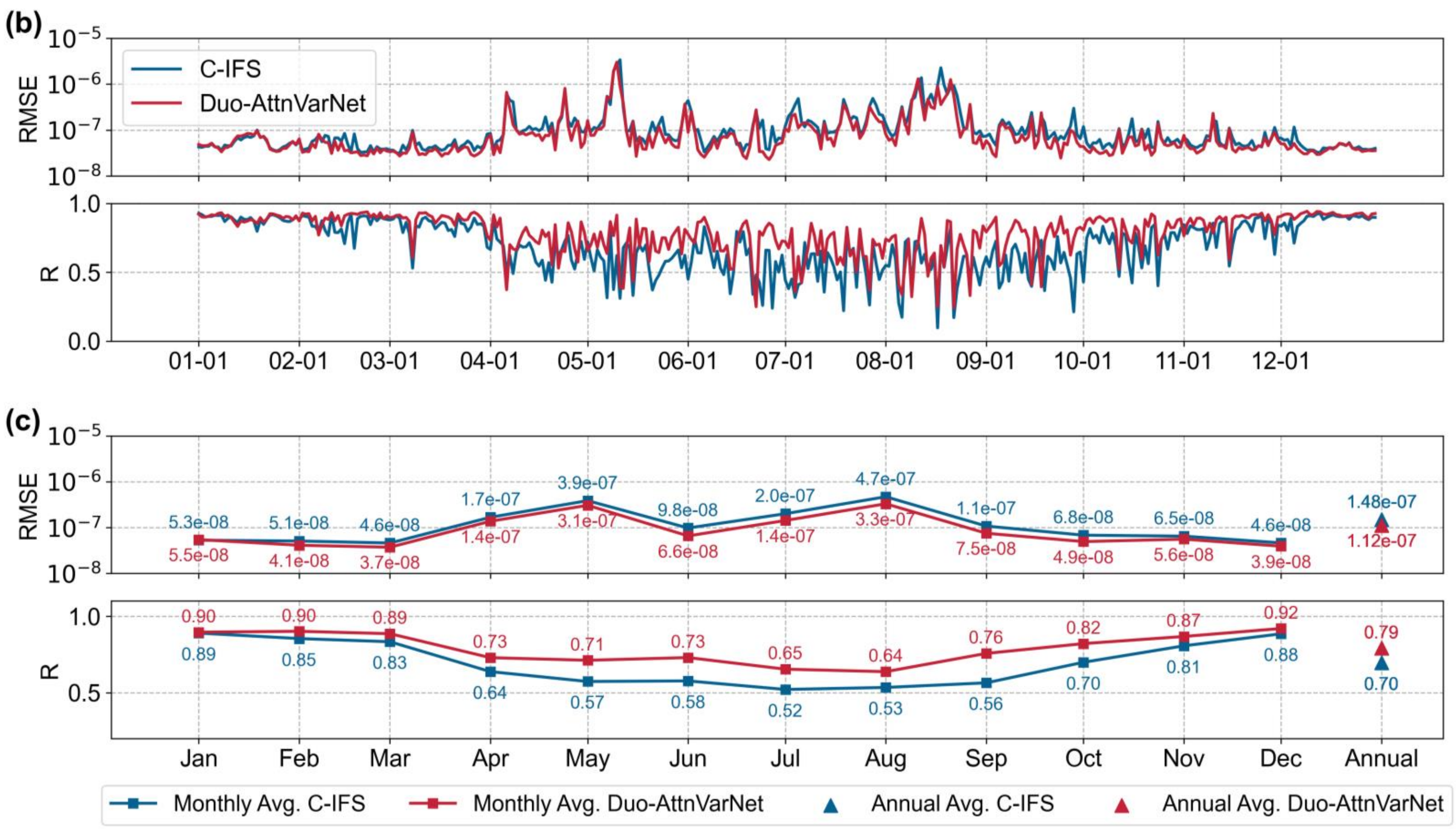
Largest improvements:

- East Asia
- North America

Other regions have smaller but positive impacts

IMPERIAL

Assimilation skill vs C-IFS



Like-for-like Comparison

- Daily analyses at 12:00 UTC
  Same assimilation window (09–21 UTC)
- Same initial conditions
- Same meteorological sources (not reanalysis)
- Same satellite observations

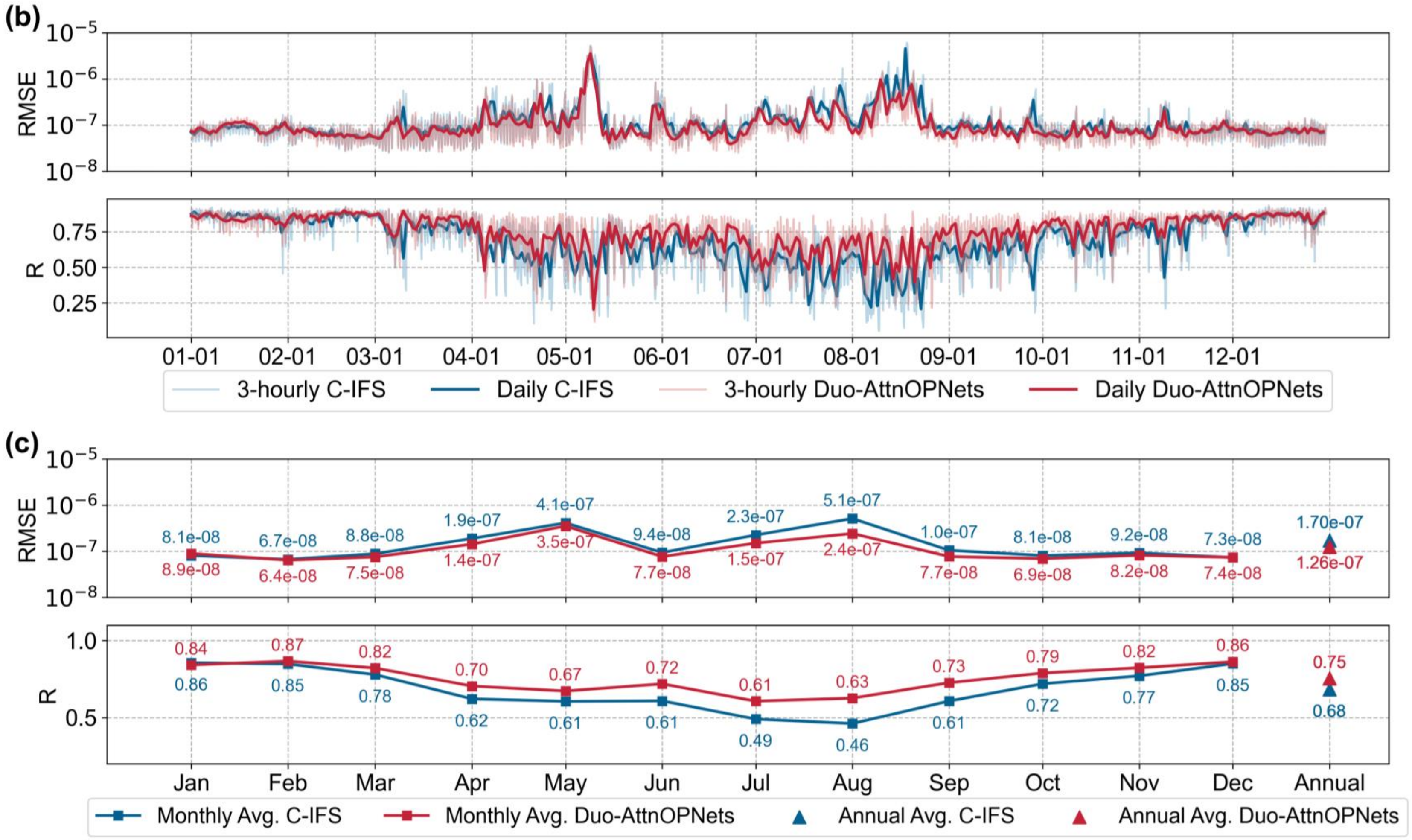
Daily & Monthly Performance

- Lower RMSE and higher correlation on most days and months
- Improvements not limited to isolated events

Annual Mean

- RMSE reduced by ~24%
- R increases from 0.70 → 0.79

Operational Forecasting vs C-IFS



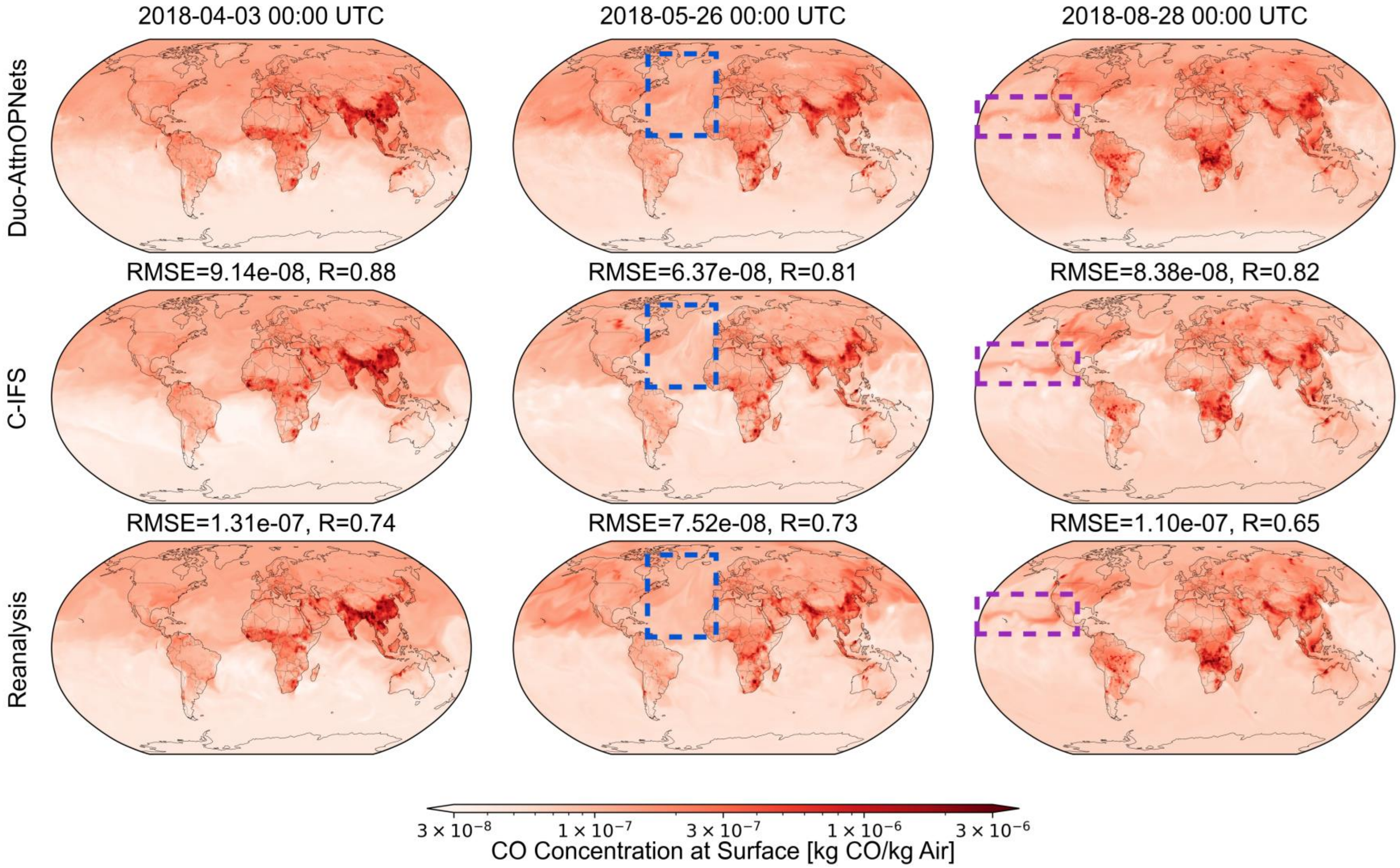
Fully Self-Cycling AI System (daily cycle):

- Assimilate near-real-time observations
- Produce analysis
- Generate next forecast
- Forecast feeds next cycle

Configurations:

- Assimilation window: 00–24 UTC
- 24-h forecasts compared with C-IFS

Operational Forecasting with Duo-AttnOPNets



Computational Cost

Hardware Setup: a single Nvidia A4500 GPU

Forecast Cost

Offline Training (one-time)
- 2015~2017 data
- 400 epochs
- 66 hours on one GPU

Once Trained
- 0.3s per forward step
- 1.6s for a 5-day forecast

Assimilation Cost

Depends on optimization settings:
- Step size/Learning rate
- Early stopping threshold

Trade-off:
Higher accuracy <-> longer runtime
Typical range:
- Fast mode: ~6 s per cycle
- Balanced mode: ~1 min per cycle

Part 4

Implications

Implications and Outlook

Key Contributions:

- First demonstration of an **AI-empowered CO DA–forecast system at global operational scale**
- Training-free AI-based assimilation with improved scalability
- Demonstrated real-time feasibility on modest hardware

Implications for Variational DA:

- AI can *augment* variational DA, not replace it
- Auto-diff offers an alternative to manual adjoint development
- AI-empowered framework reduces computational burden

Where This is Most Useful:

- Resource-limited environments
- Potential ensemble applications

Thank you

Questions & Discussion Welcome