

15th UNESCO/IOC-ODC Training Course

on Ocean Dynamics and Ocean Models

From Traditional Data Assimilation to AI Data Assimilation

New Representations for Earth System State Estimation

Yanfei Xiang, Ph.D.

Lecturer

Chengdu University of Information Technology (CUIT)

Organized by
UNESCO / IOC-ODC

September 8, 2026

About Me

Education and research experience in AI, Earth system modeling, and data assimilation



Yanfei Xiang, Ph.D.
CUIT Lecturer
xiangyanfei212@outlook.com

Research interests

AI-based Data Assimilation

Intelligent Weather Forecasting

Earth System State Estimation

Education

- Ph.D. · Atmospheric Science** 2022.08 – 2026.06
Tsinghua University · Department of Earth System Science
Intelligent Earth system modeling · AI forecasting · AI data assimilation
- M.Eng. · Computer Science and Technology** 2019.09 – 2022.06
Chengdu University of Information Technology (CUIT)
AI-assisted weather forecasting
- B.Eng. · Digital Media Technology** 2015.09 – 2019.06
Chengdu University of Information Technology (CUIT)
Computer science, graphics, algorithms

Internship Experience

Microsoft Asia
2023.10 – 2025.04
AI-based data assimilation

Shanghai Meteorological Bureau
2023.06 – 2023.07
AI precipitation nowcasting and
model optimization

National Supercomputing Center in Wuxi
2020.07 – 2022.01
Global ocean reconstruction and coastal fog /
precipitation forecasting

A cross-disciplinary background connecting computer science, atmospheric science, ocean modeling, and AI-based data assimilation.

Where I Come From?

CHENGDU

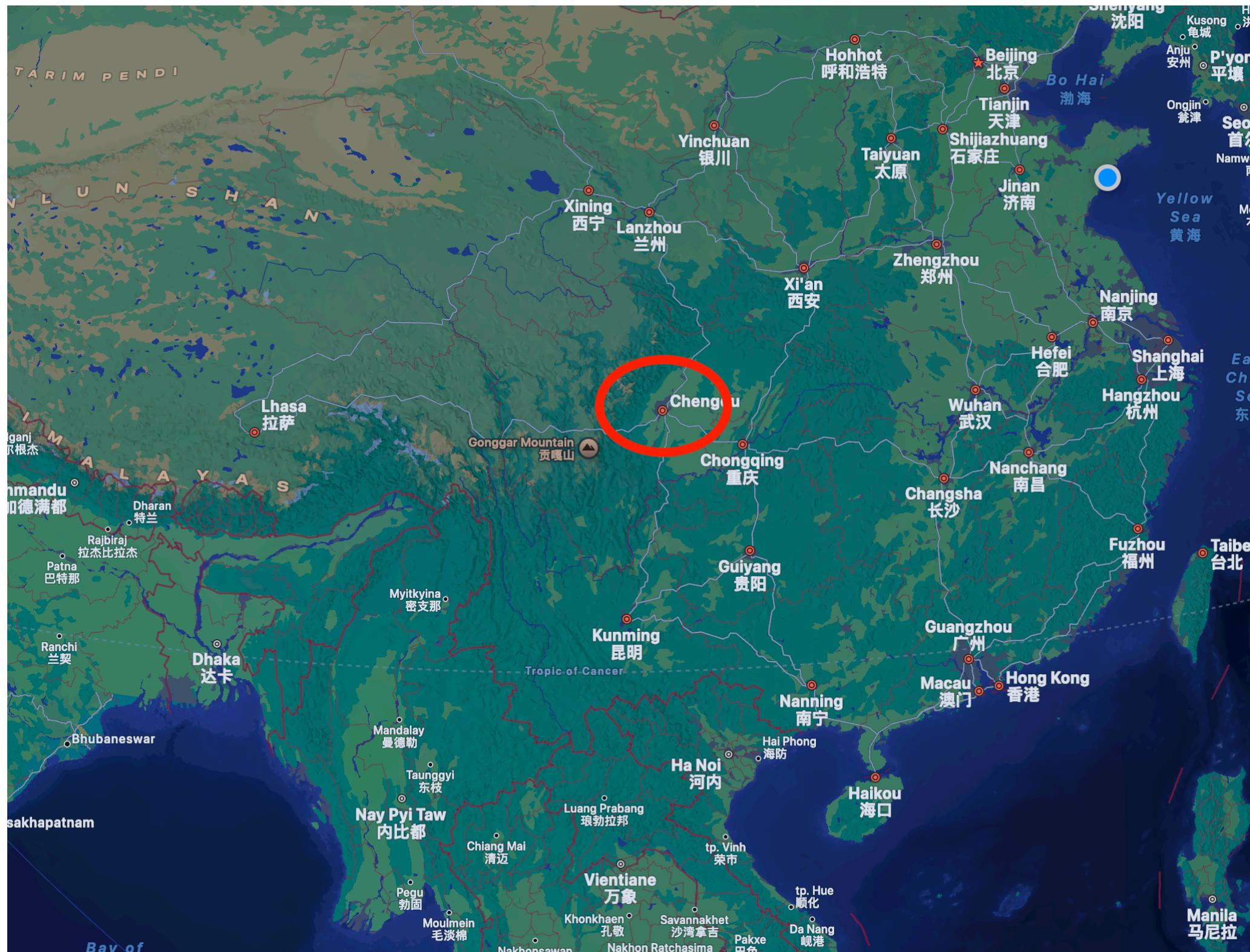
Home of Giant Pandas 🐼

🌶️ **Sichuan Food**

Famous for spicy cuisine and hotpot

🏔️ **Gateway to Western China**

Close to mountains and the Tibetan Plateau

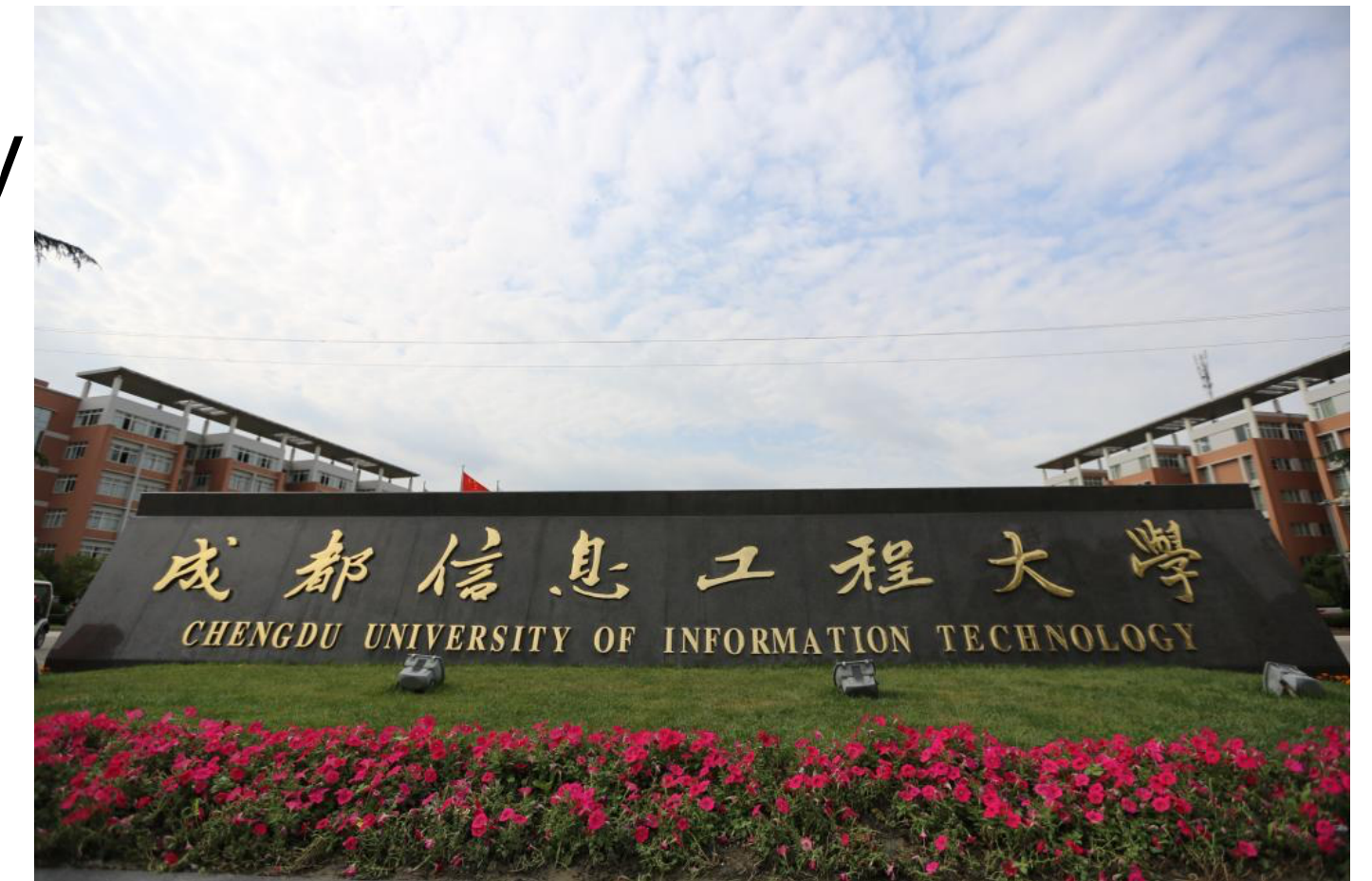


Chengdu University of Information Technology (CUIT)

☁️ Meteorology

💻 Information Technology

🤖 AI + Earth Science

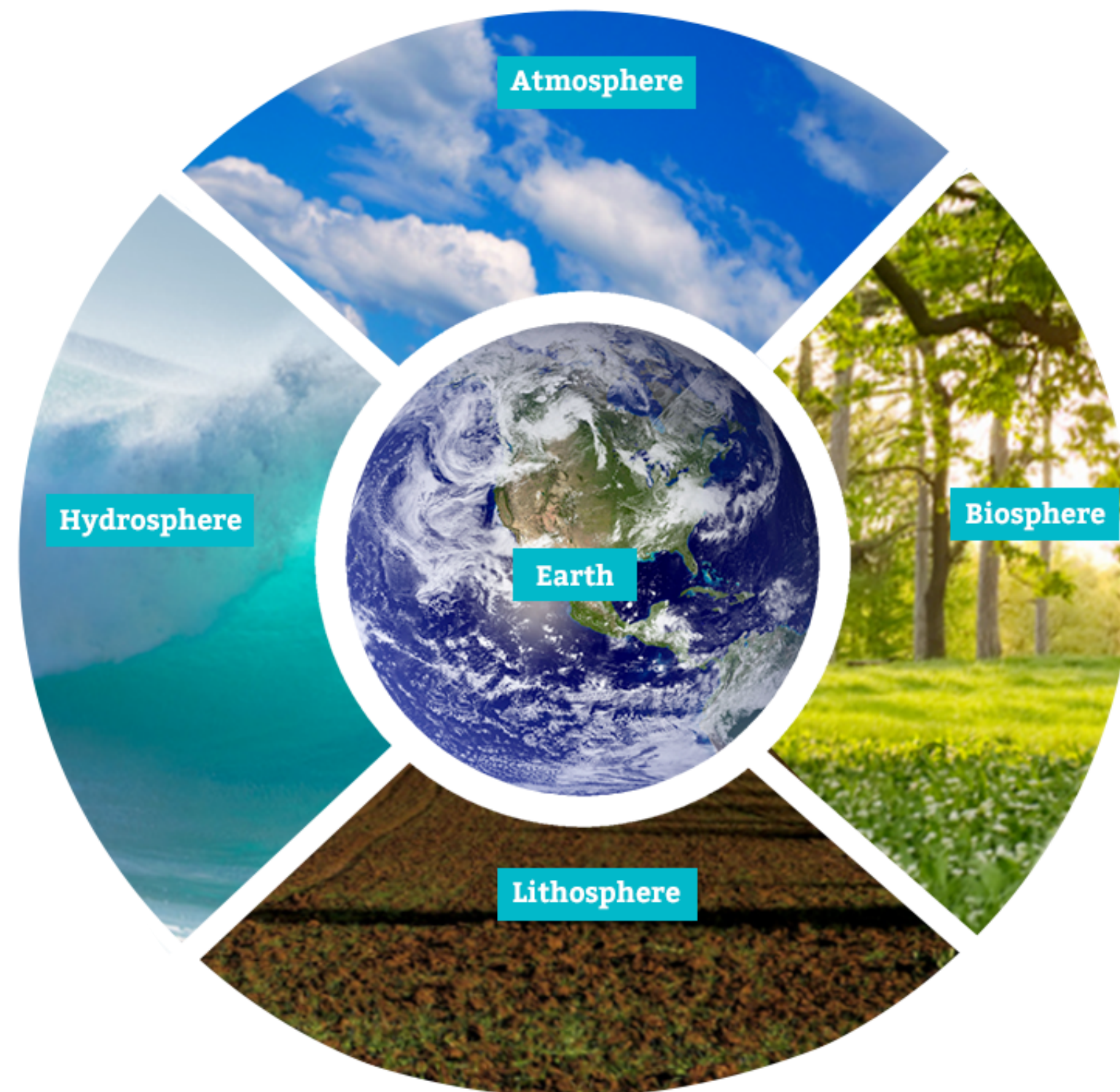


Part 1

Why Do We Need Data Assimilation (DA)?

From Observations to Earth-System Prediction

Prediction starts from the current state of the Earth system.



Earth system

We need the current state

The **initial condition** influence the quality of prediction.



But we cannot observe every point

Observations are samples of the state — sparse in space and time.



How do we infer the complete state?

Observations + Prior knowledge → Complete state

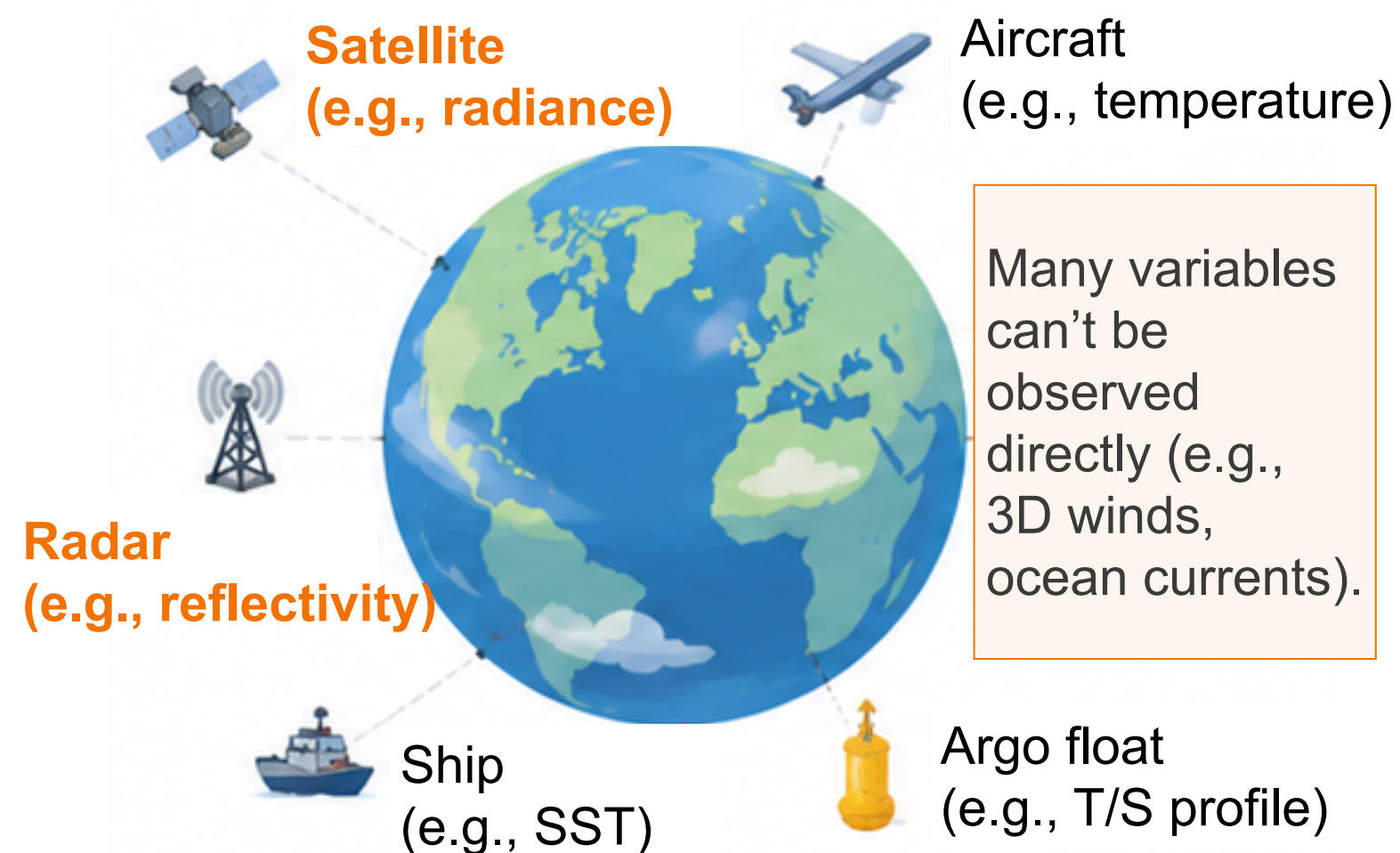
Data assimilation (DA) try to solve this state-estimation problem.

Key Challenges in State Estimation

Real-world systems contain hidden variables, noisy sensors, and complex physical behaviors that cannot be measured directly or perfectly.

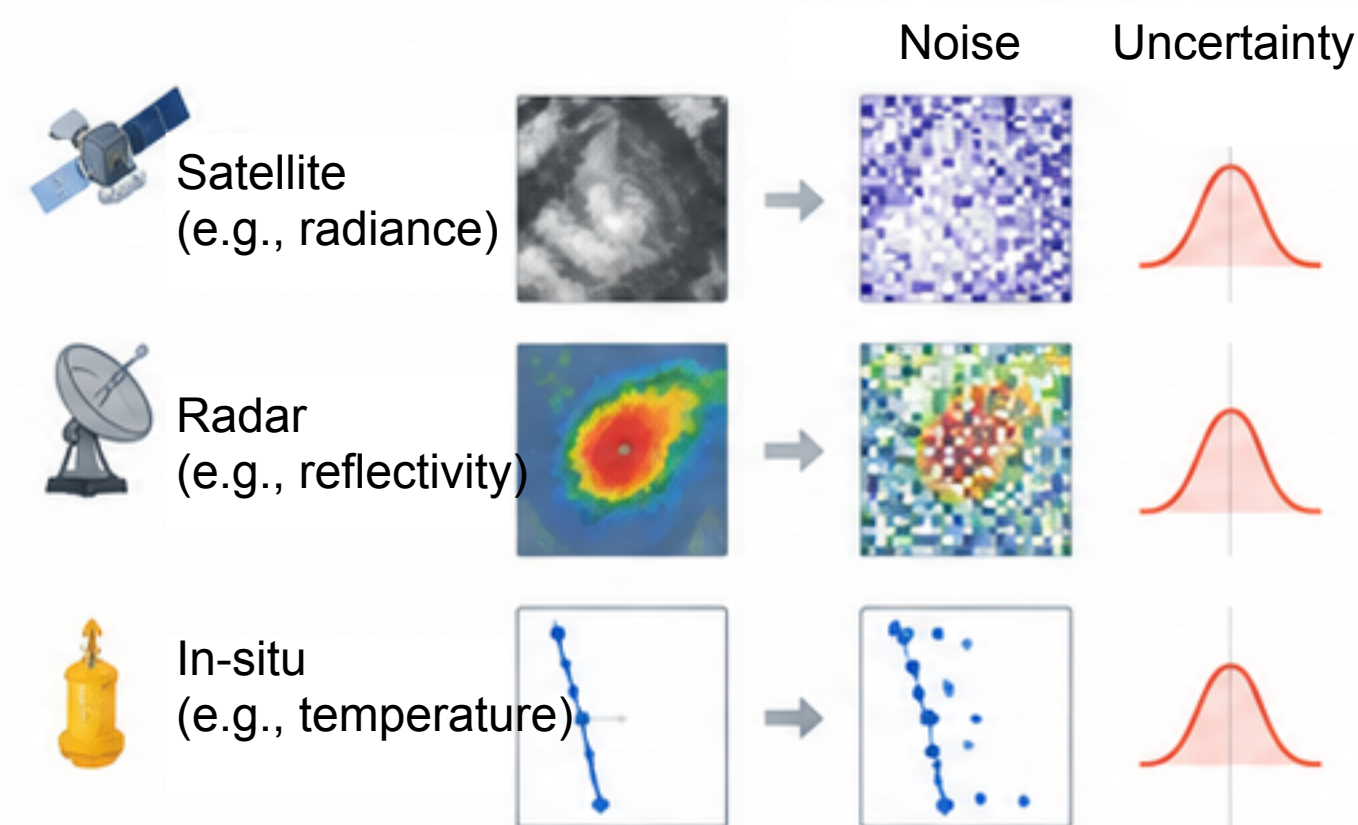
1 Indirect and Incomplete Measurements

We can only observe a few variables at limited locations and times.



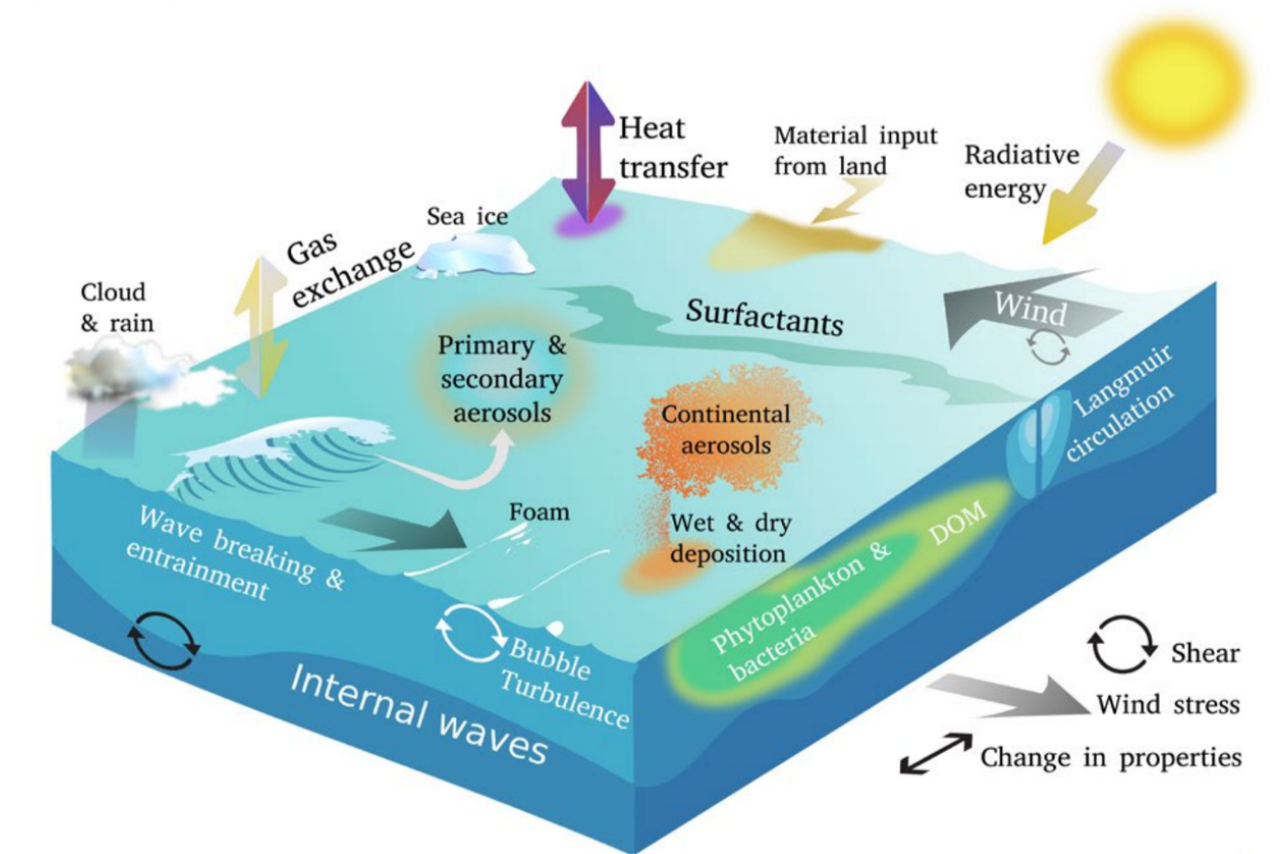
2 Sensor Noise and Uncertainty

Observations are not valuable, but they are not perfect.



3 Coupled Nonlinear Interactions

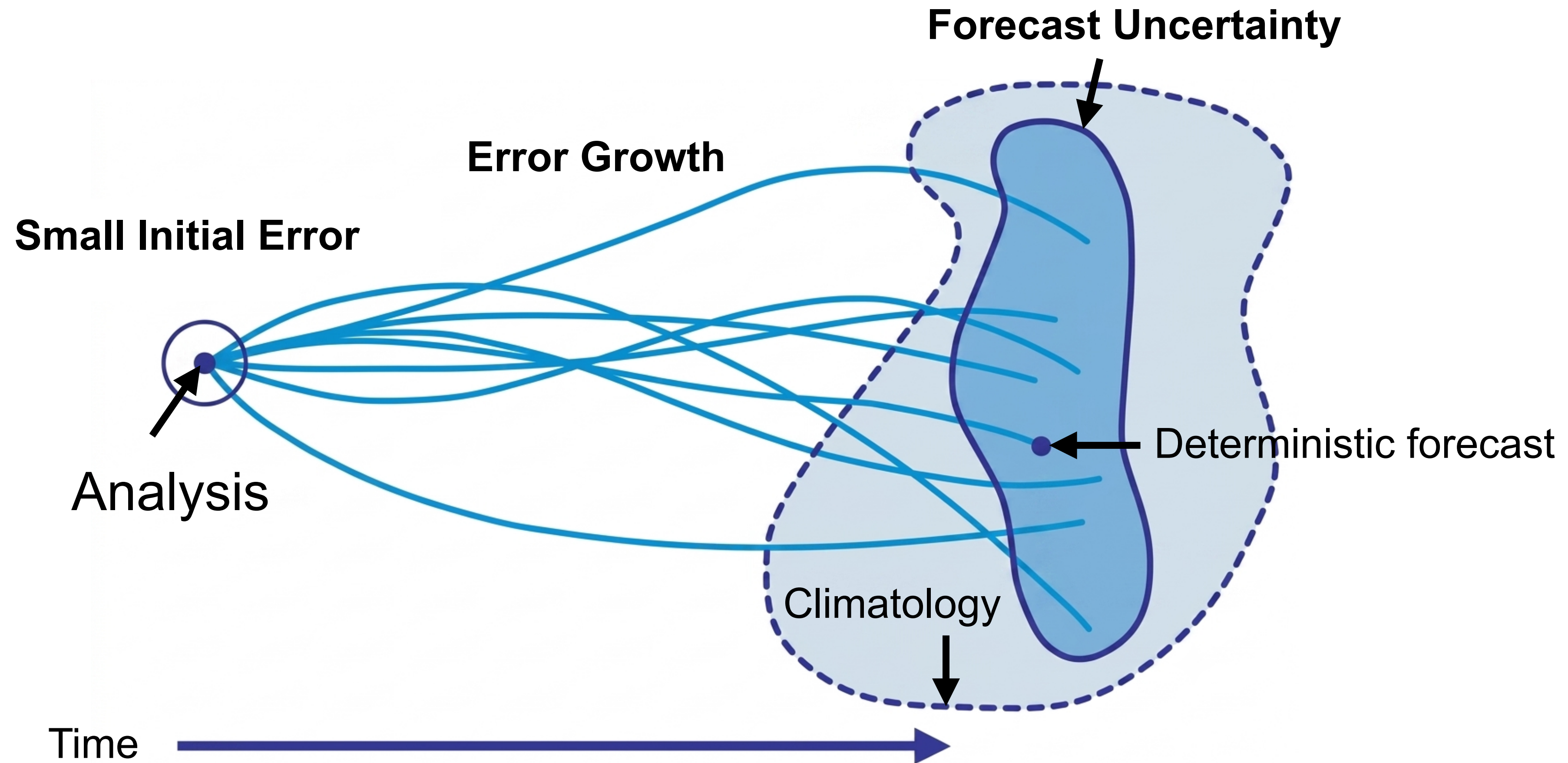
Atmosphere, Ocean, Land ...
Energy, momentum and material are exchanged continuously.



(Neukermans et al. 2018)

Why Does the Initial Condition Matter?

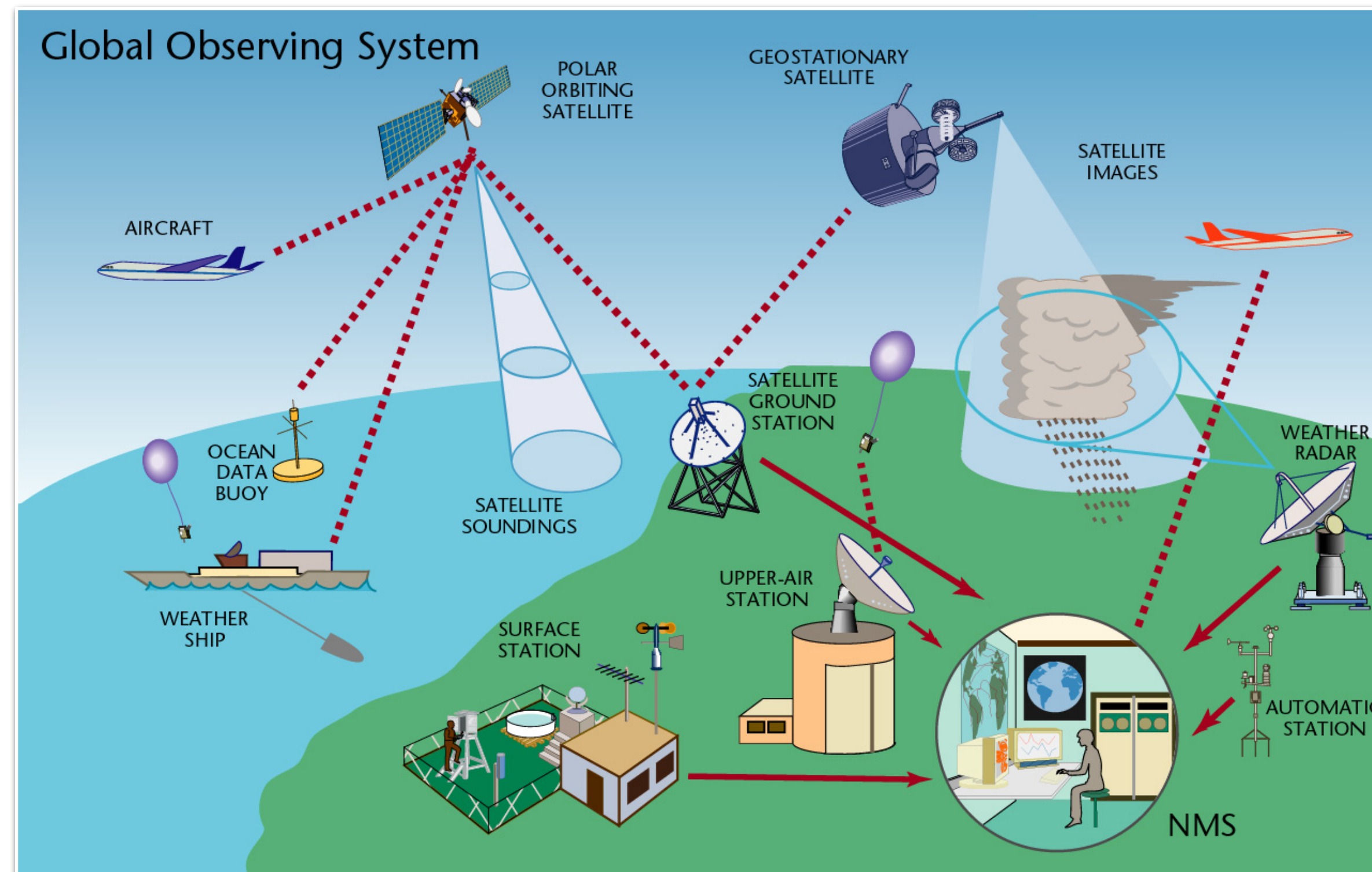
Better prediction begins with a better estimate of the current state.



**Better initial condition → slower error growth → more skillful prediction
(if model is perfect)**

How Do We Observe the Earth?

Earth observations come from many platforms — each samples different time, variables, and locations.



Global observing system schematic (Source: WMO)

Different measurements

Satellite

Broad coverage

SST, SSH, SSW
Atmospheric temperature, Cloud, Humidity

Ground & In-situ

Local & direct

stations, radar
aircraft observations

Ocean

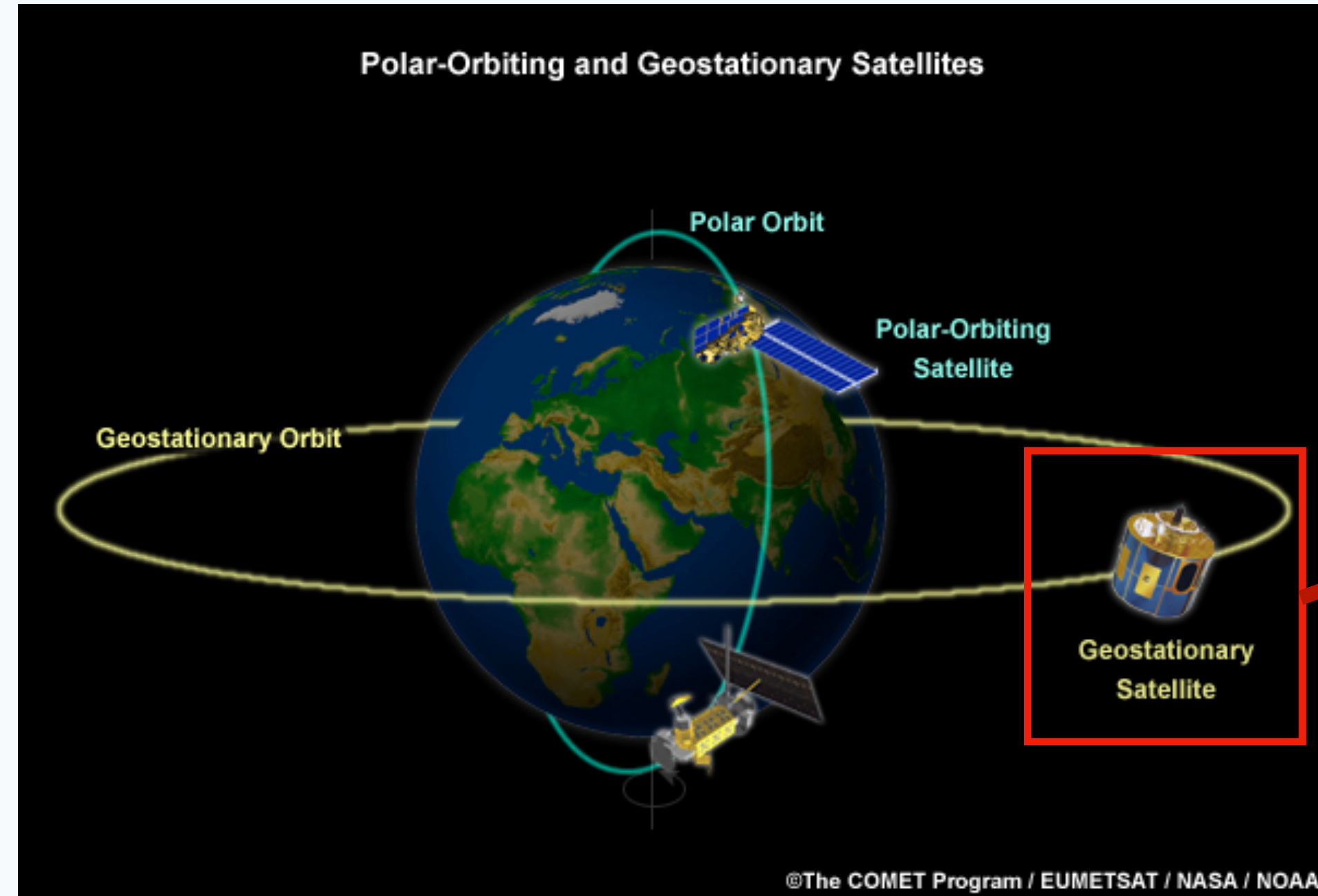
Sparse & depth-aware

Argo floats, buoys, ships
surface + subsurface

They observe the Earth in different way.

Geostationary Satellite

Satellite observations have different sampling pattern.



Fixed viewpoint → frequent, image-like sampling

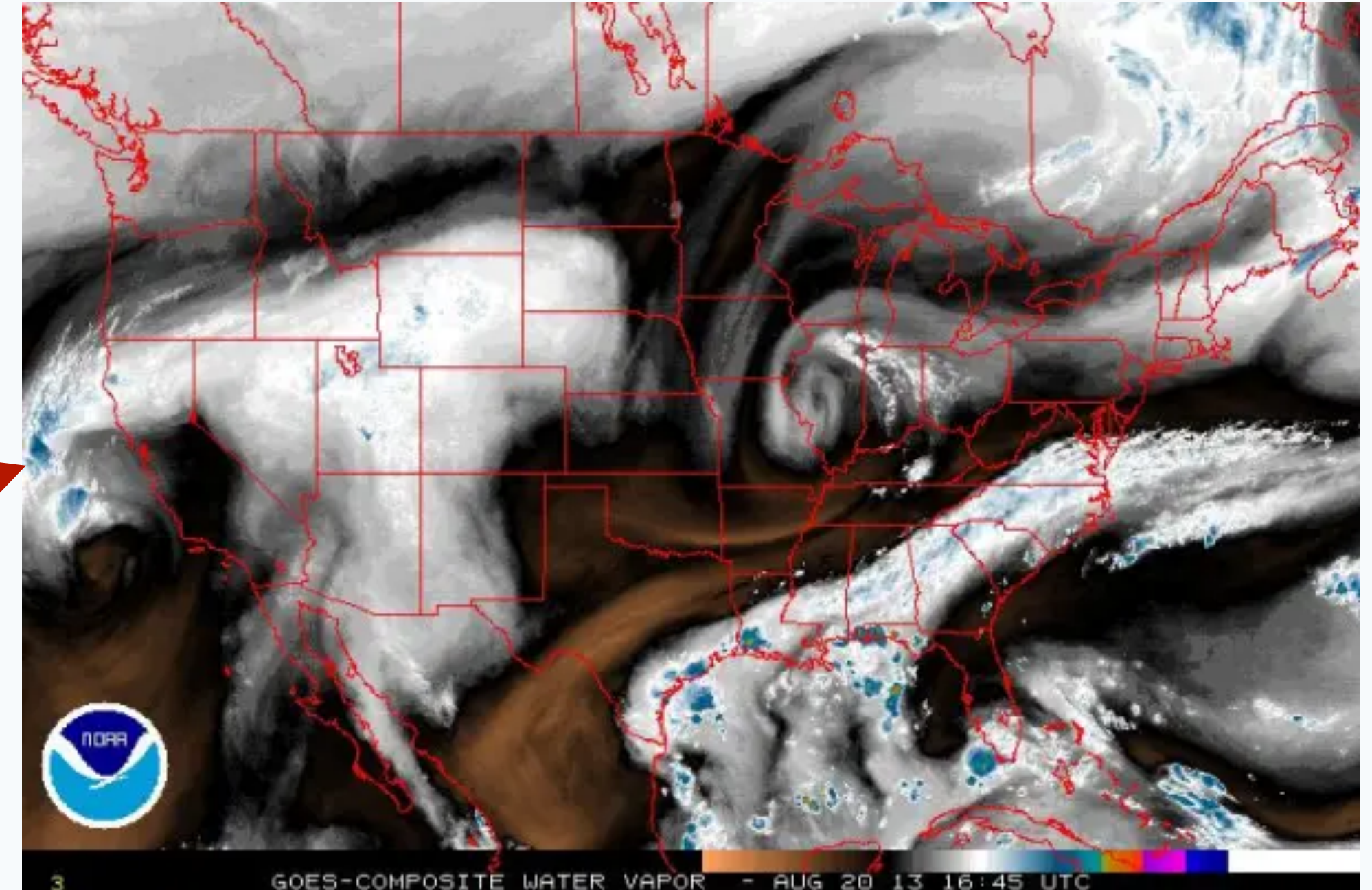


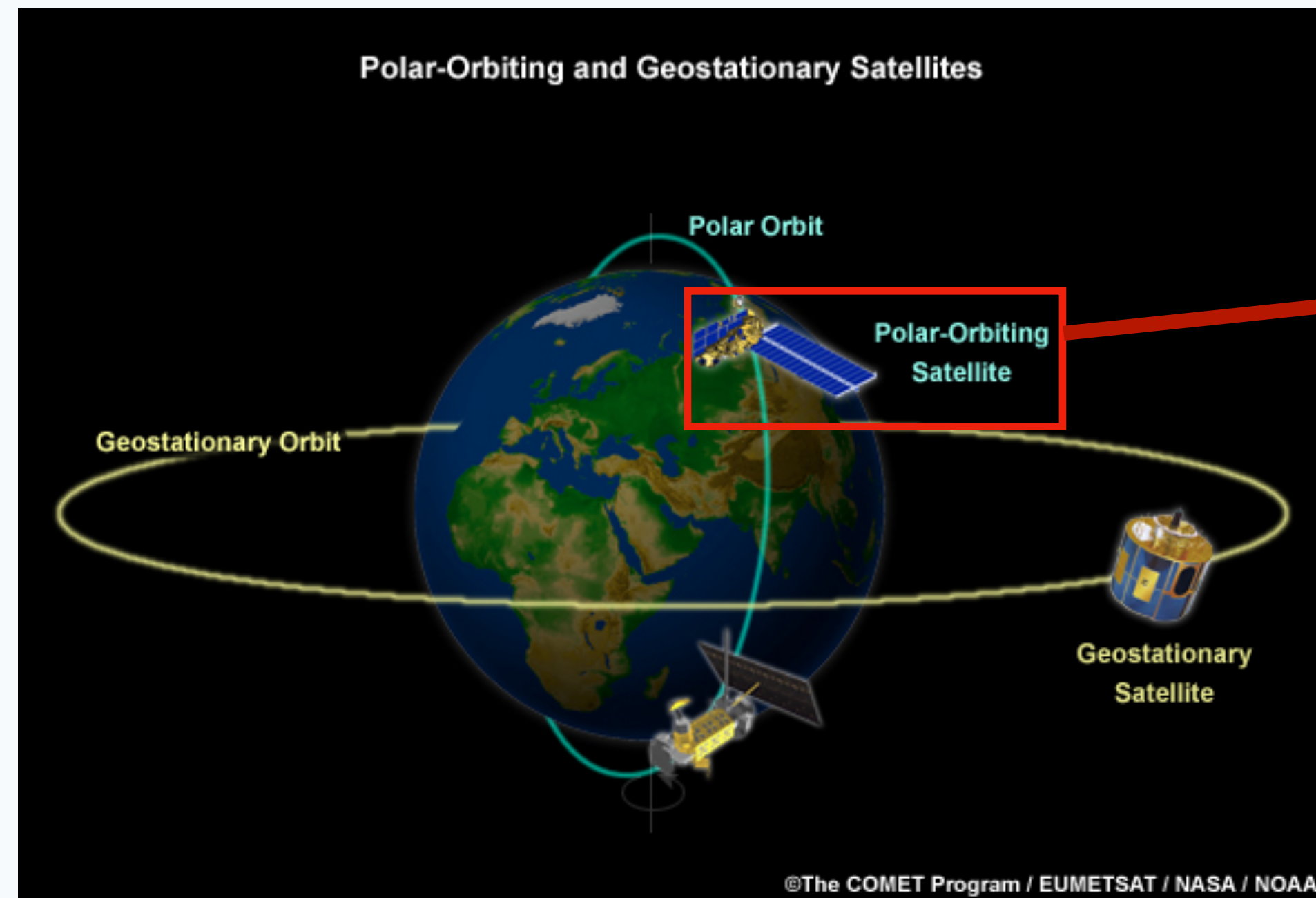
Image of water vapor (Source: NOAA)

**Dense spatial coverage over one region
Tracking the evolution of weather systems**

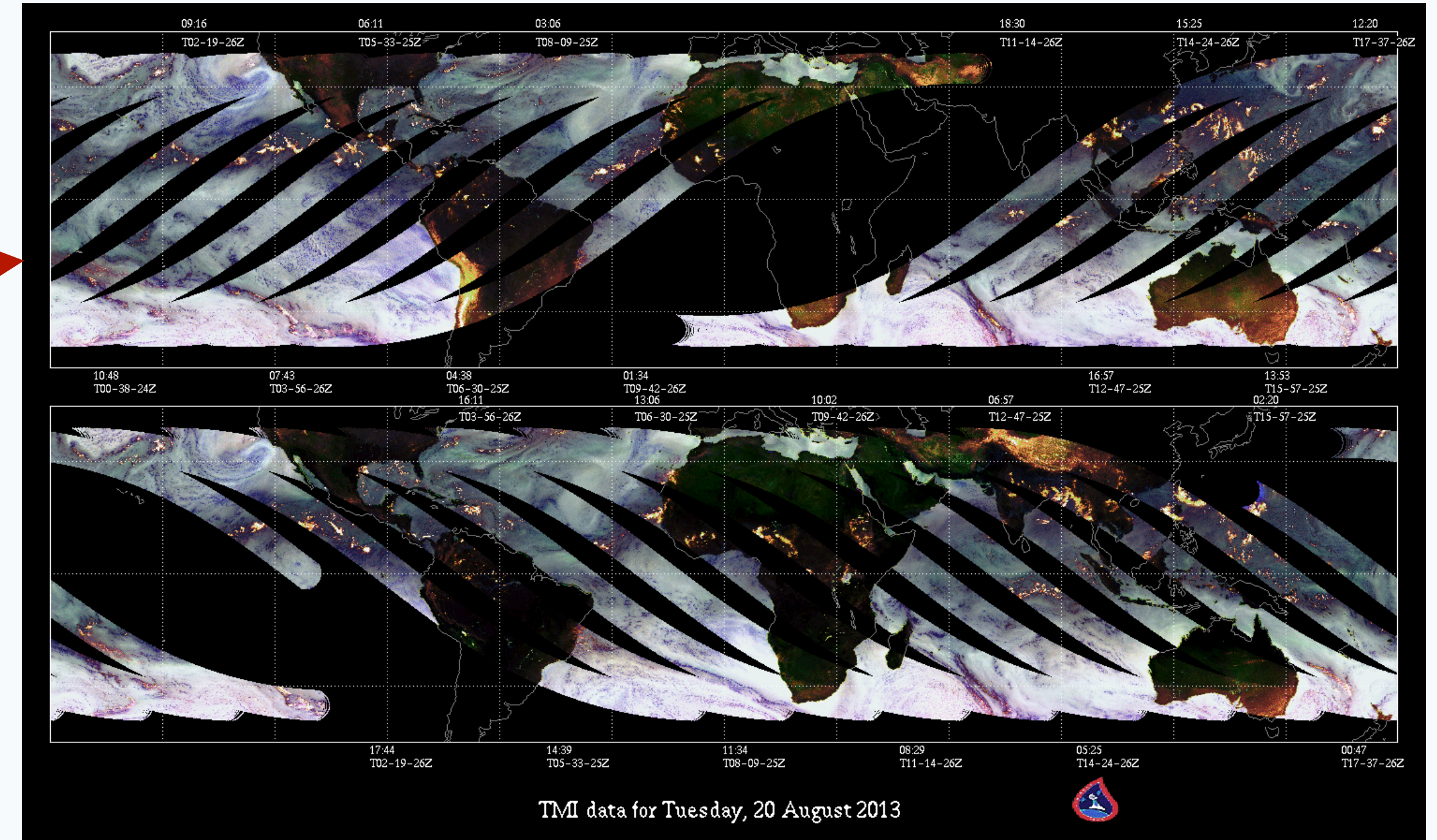
Dense 2D imagery with high temporal frequency

Polar-Orbiting Satellites

Satellite observations have different sampling pattern.



Moving orbit → swath / track sampling

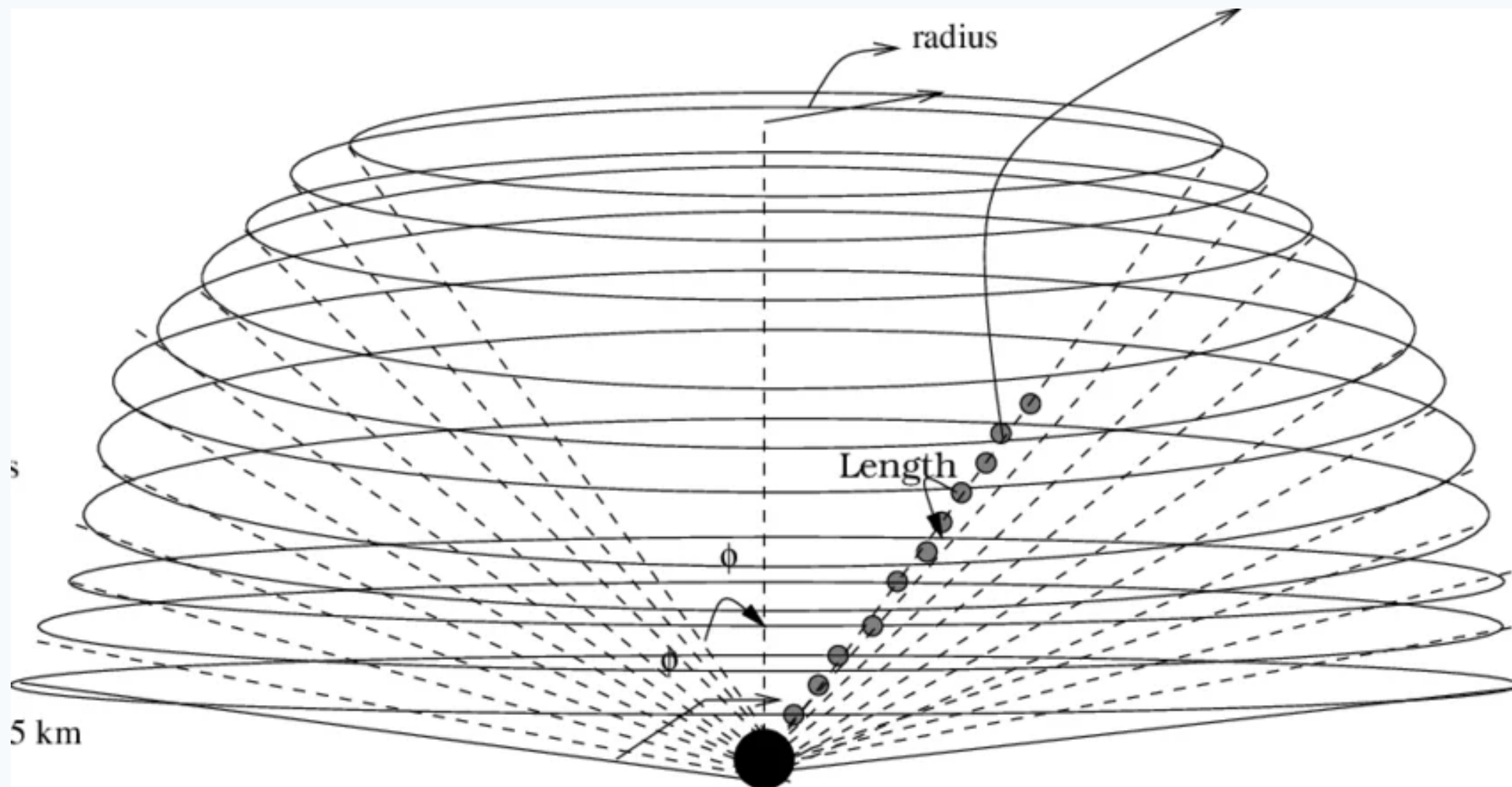


Images from the TMI imager (Source: NASA)
**Global coverage over time,
with gaps between swaths**

Swaths / trajectories over time

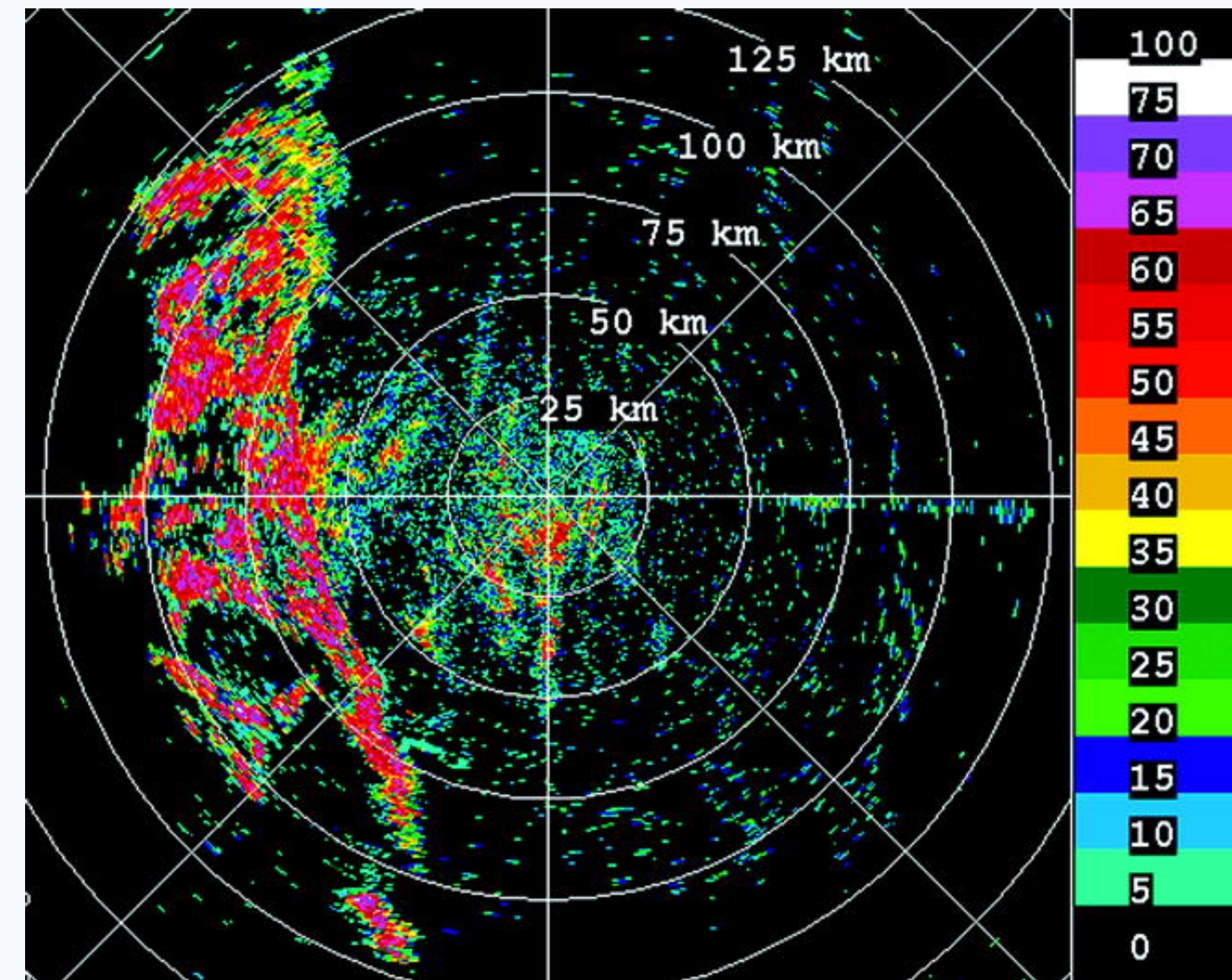
Weather Radar

Weather radar samples the atmosphere in range–azimuth–elevation coordinates.



The structure of 3D Doppler radar data. (Qiu et al., 2001)

Conical scans build a 3D polar volume

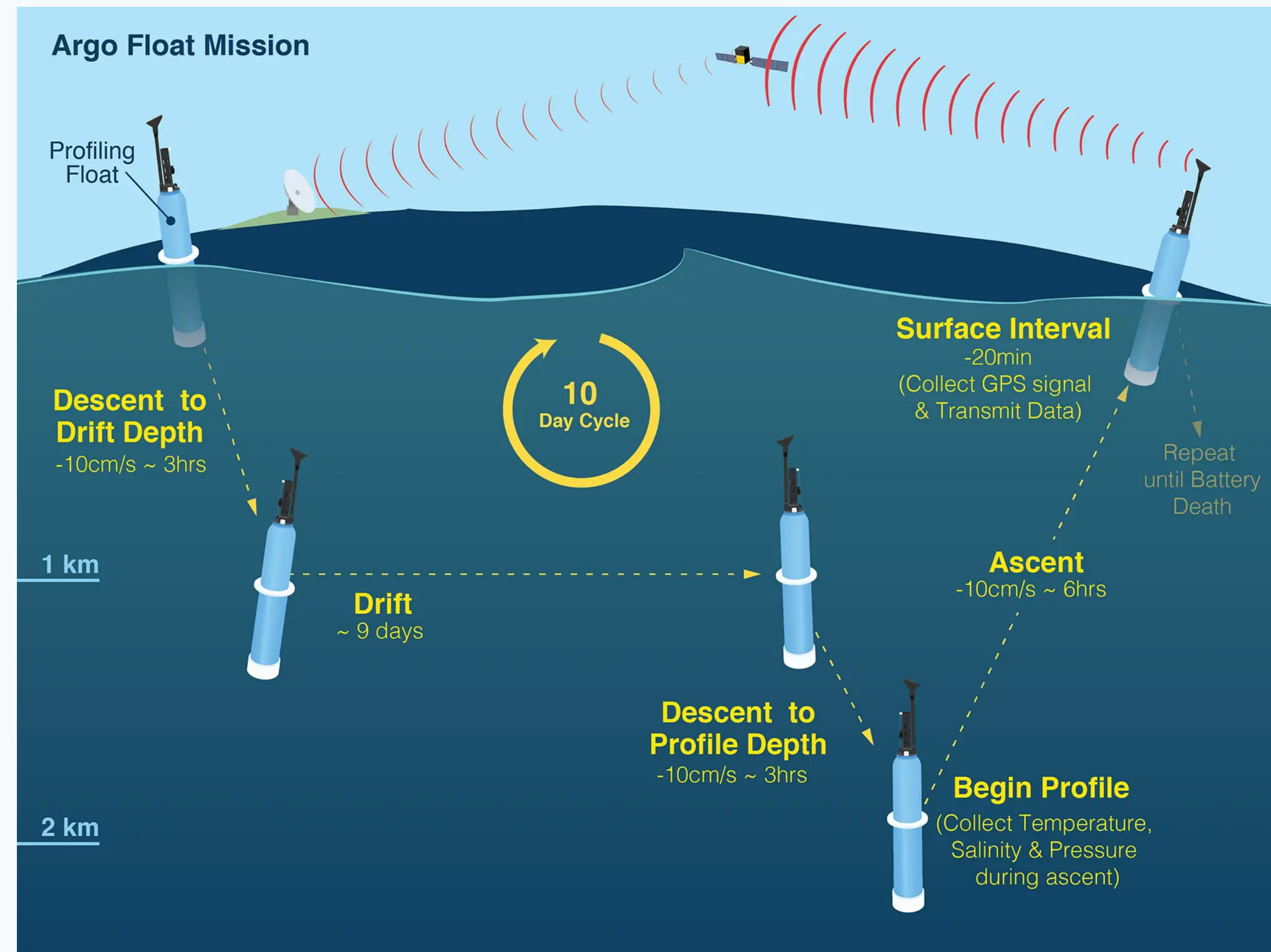


Sampling density and height vary with range from the radar

3D polar volume

Argo

Argo observes the ocean as sparse vertical profiles that move through space and time.



- **Points at surface**
Float locations are sparse
- **Vertical Profiles**
Temperature / salinity sampled with depth
- **Time-Varying**
Floats drift; sampling locations change

Sparse samples of a 3D ocean

Sparse, moving vertical profiles in 3D ocean.

Q&A

What kind of observations have you used?

What are their characteristics?

What are they used for?

.....

Observation $\neq$ Complete State

Observations are informative where and when they exist, but the Earth-system state is continuous in space, time, and variables.

- **Sparse**

Only some locations are sampled

- **Irregular**

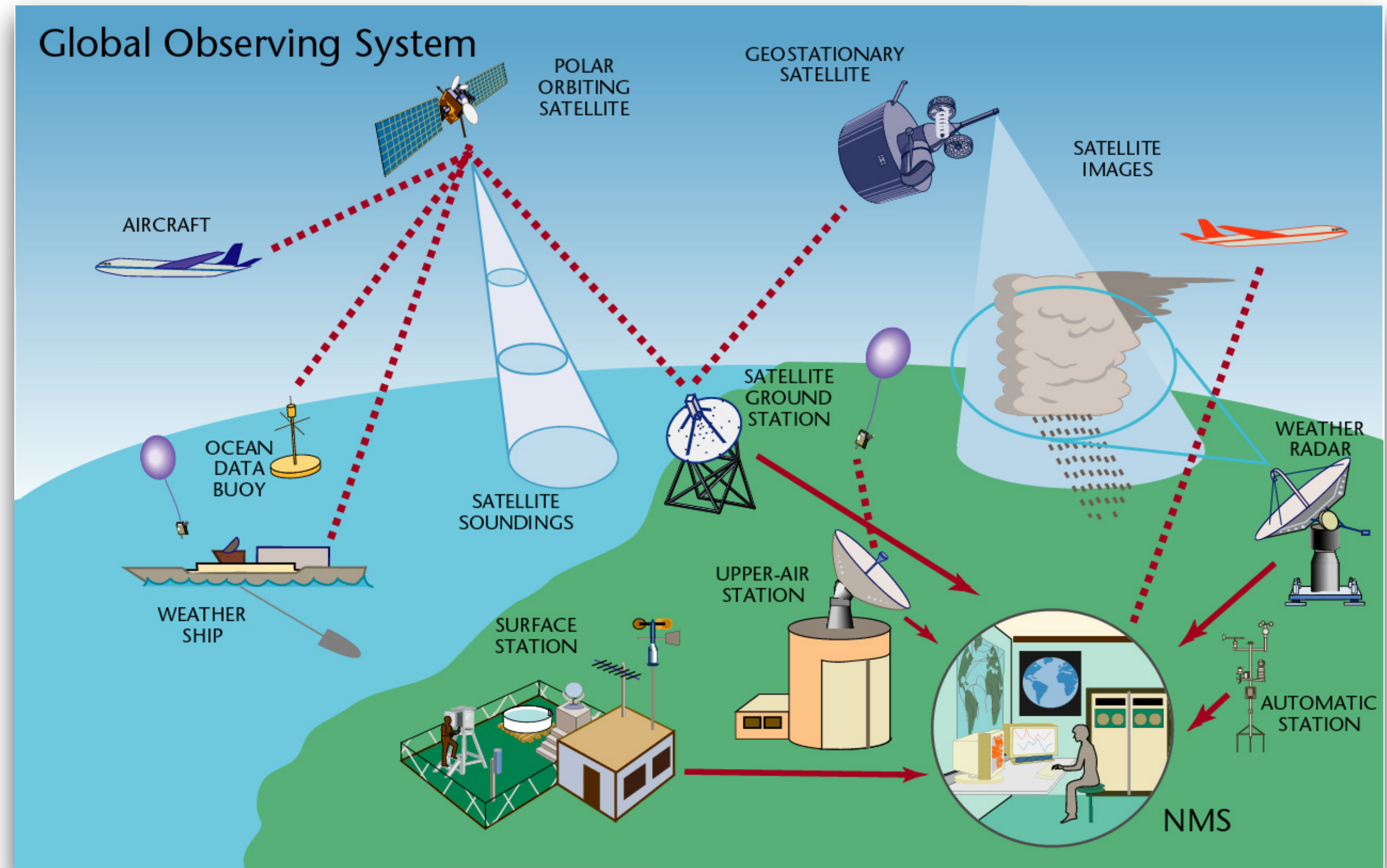
Sampling locations and times are nonuniform

- **Heterogeneous**

Sensors observe different variables and geometries

- **Non-fixed**

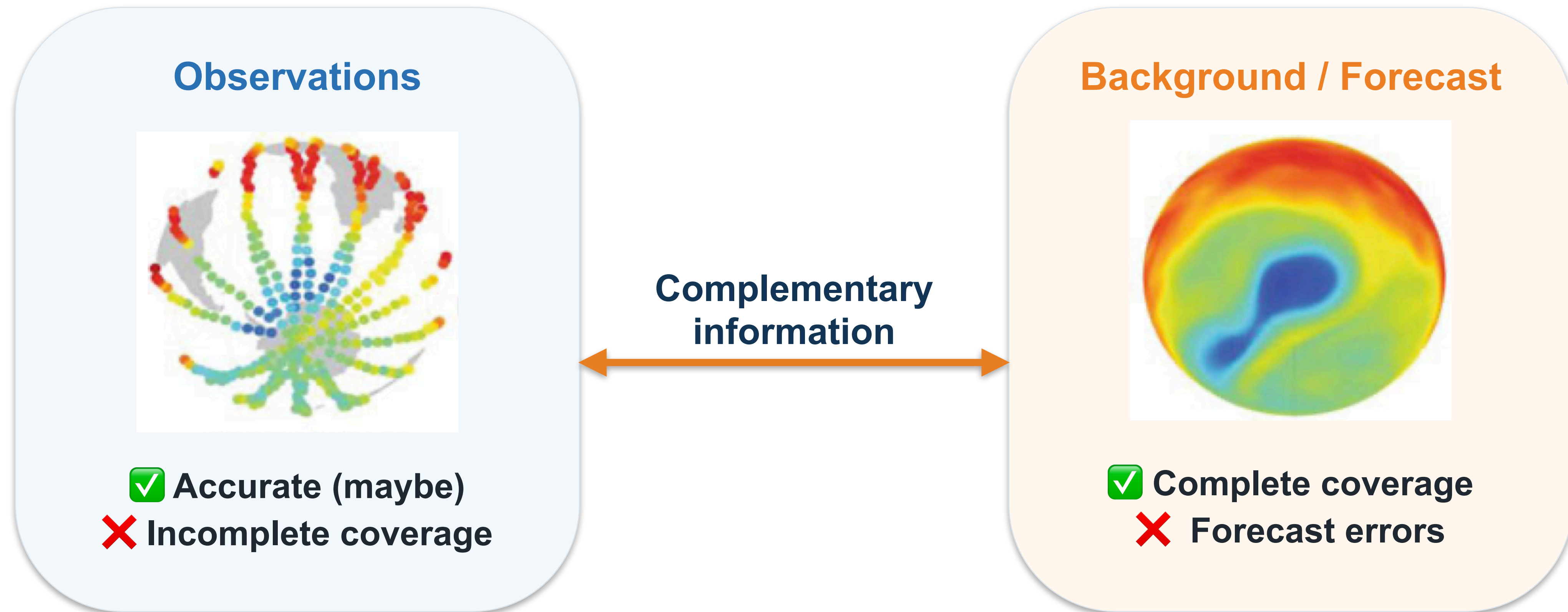
Changing locations, moving platforms



How can we combine multi-sources observations into one coherent state estimate?

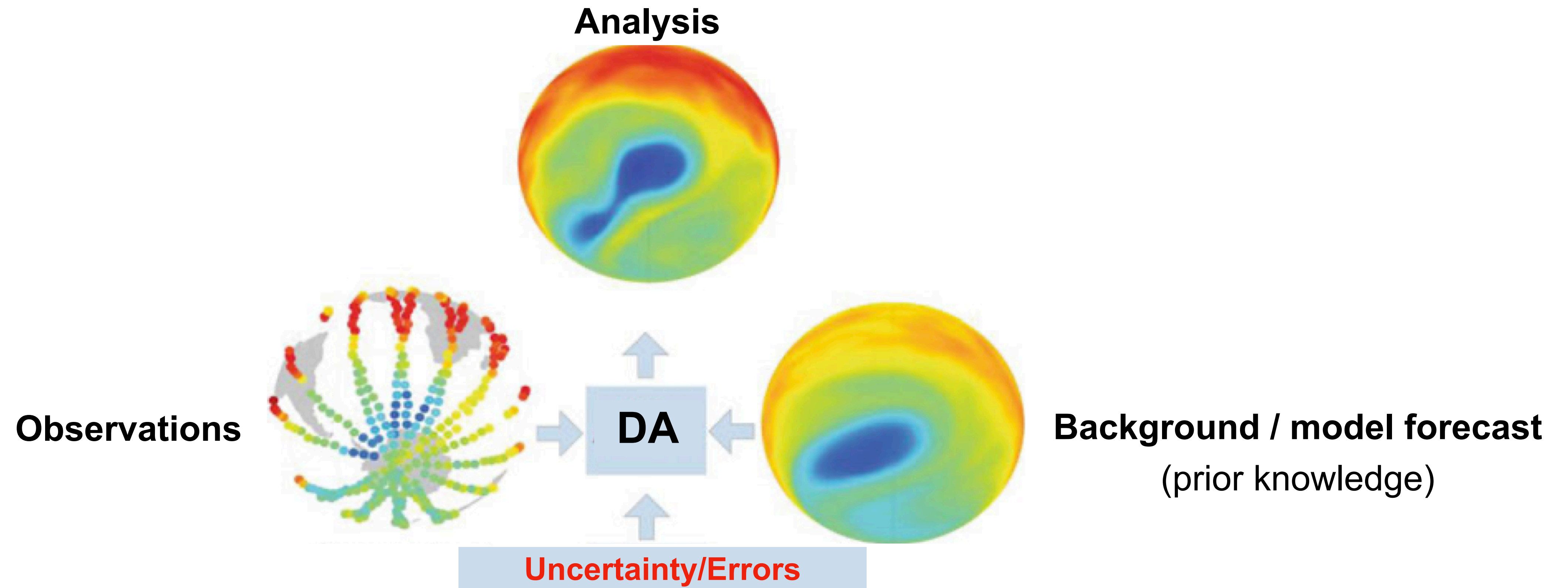
The Background Provides a Complete State — But Errors Exist.

A model forecast supplies a dynamically coherent field, but forecast and model errors remain.



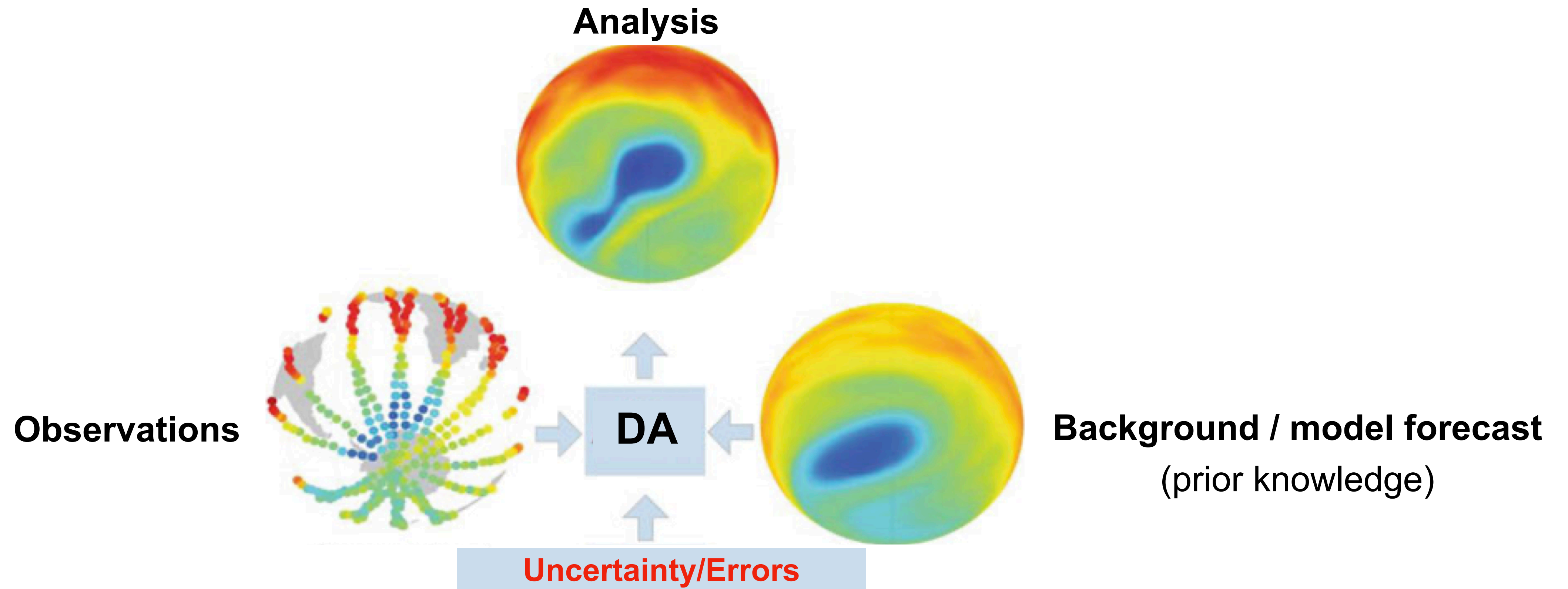
Neither source alone is sufficient → combine them.

The Goal of Data Assimilation



DA estimates the state by combining prior knowledge with observations.

The Goal of Data Assimilation



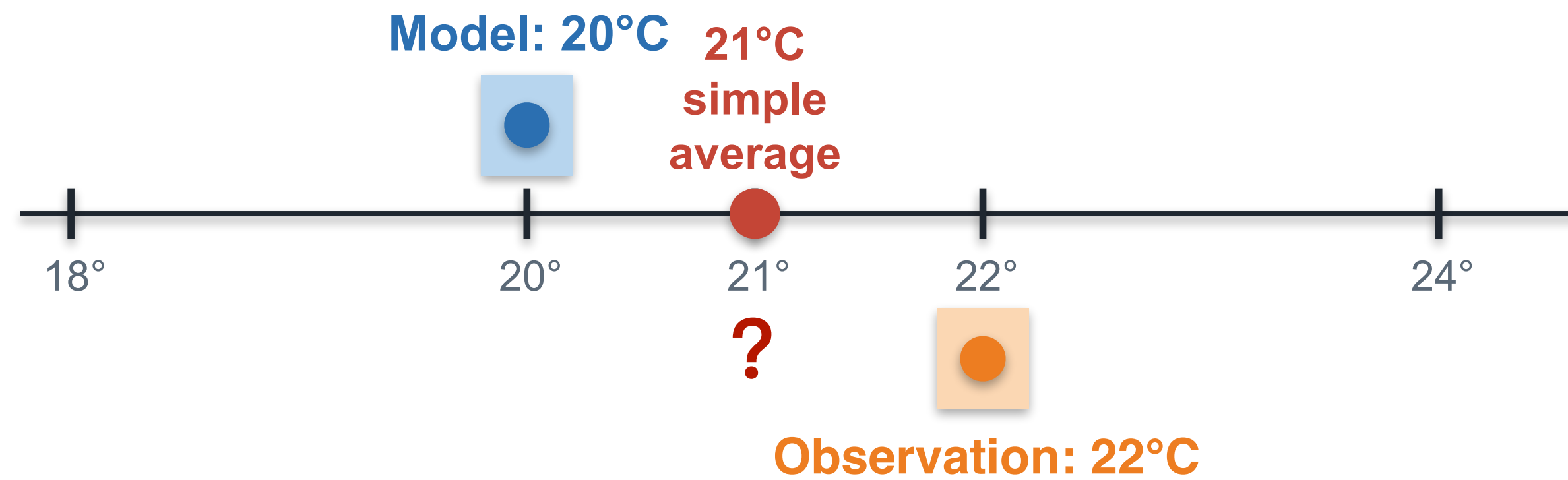
DA estimates the state by combining **prior knowledge** with **observations**.

Can we simply take average?

Why Not Simply Average the Model and Observations?

Because the two sources have different uncertainty.

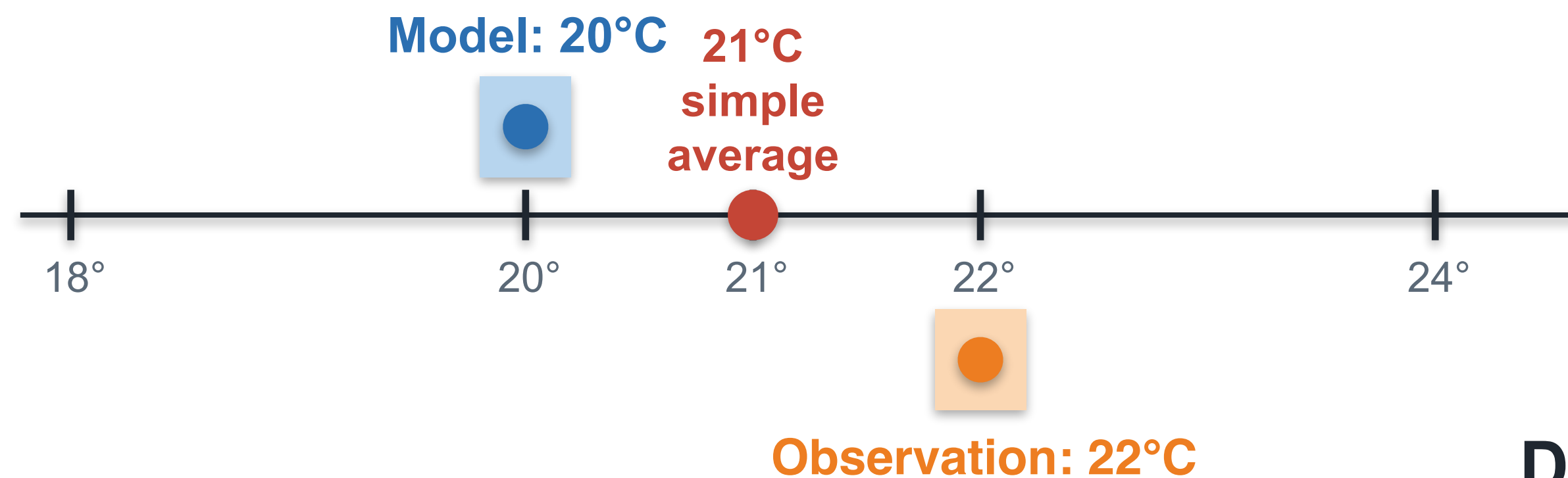
A simple example



Why Not Simply Average the Model and Observations?

Because the two sources have different uncertainty.

A simple example



Reliability should determine the weight

Smaller
uncertainty

→ more weight

Larger
uncertainty

→ less weight

DA performs an uncertainty-weighted correction.

Trust each source according to its uncertainty.

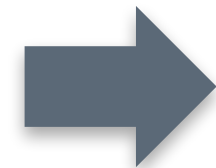
Analysis = Background + Observation-Informed Correction

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K} [\mathbf{y} - \mathbf{H}(\mathbf{x}^b)]$$

Background

$\mathbf{x}^b$

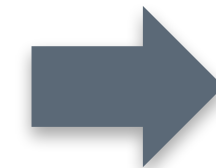
Prior estimate



Innovation

$\mathbf{y} - \mathbf{H}(\mathbf{x}^b)$

What do observations
say is wrong?



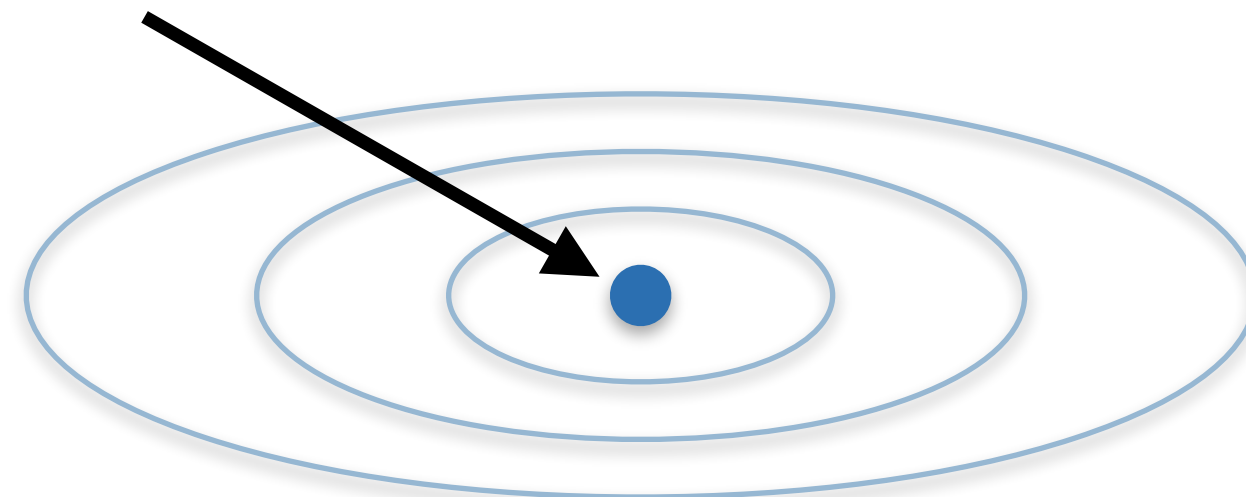
Gain

$\mathbf{K}$

how strongly /
where to correct

The analysis is not an average — it is a structured, uncertainty-aware update of the background.

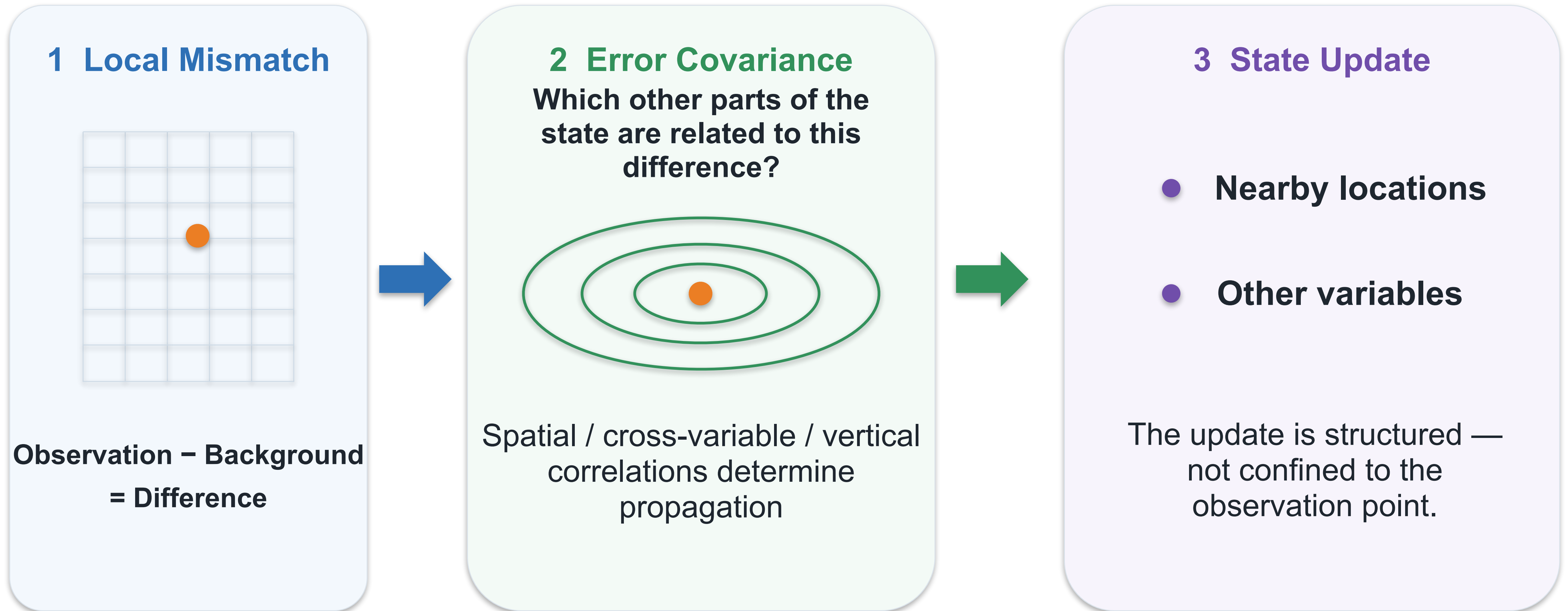
Observation



How large of an area can we correct?

Where Does the Correction Go?

A local observation can influence an area through error covariance.



DA not only tells us how much to correct, but also how to correct.

What Controls the Correction?

The Kalman gain combines background uncertainty, observation uncertainty, and the observation operator.

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K} [\mathbf{y} - \mathbf{H}(\mathbf{x}^b)]$$



Kalman gain

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$$

K determines both the magnitude and the spatial / multivariate spread of the correction.

B

**Background-error
covariance**

How uncertain is the background?
How are errors correlated?

R

**Observation-error
covariance**

How uncertain are the observations?
How are their errors correlated?

H

Observation operator

Connect the model state with what
the sensor actually observes.

The observation tells us what is wrong.

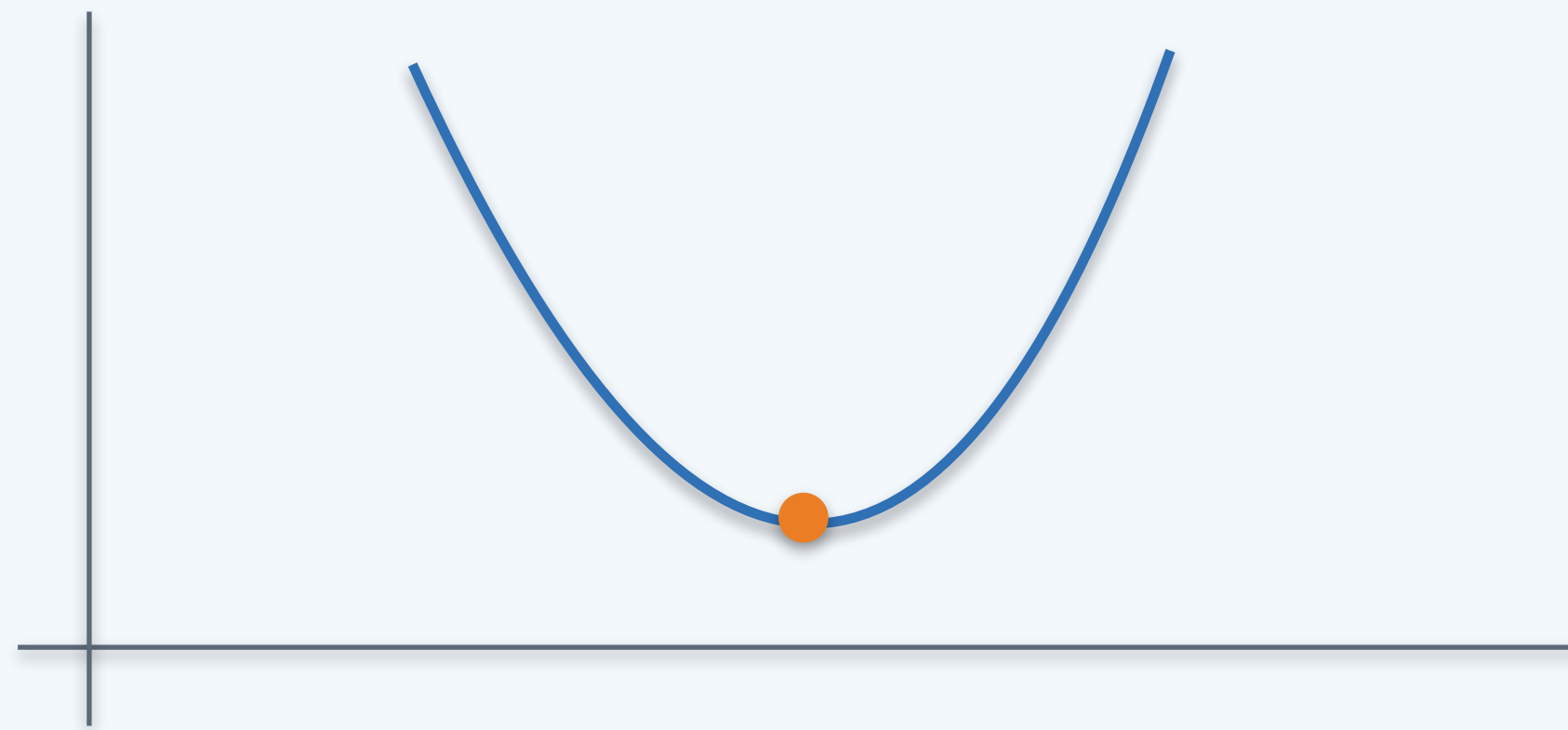
The covariance tells us how the correction should spread.

Two Classical Ways to Estimate the State

Same goal, different computational view: optimize one state or estimate a distribution.

Variational DA

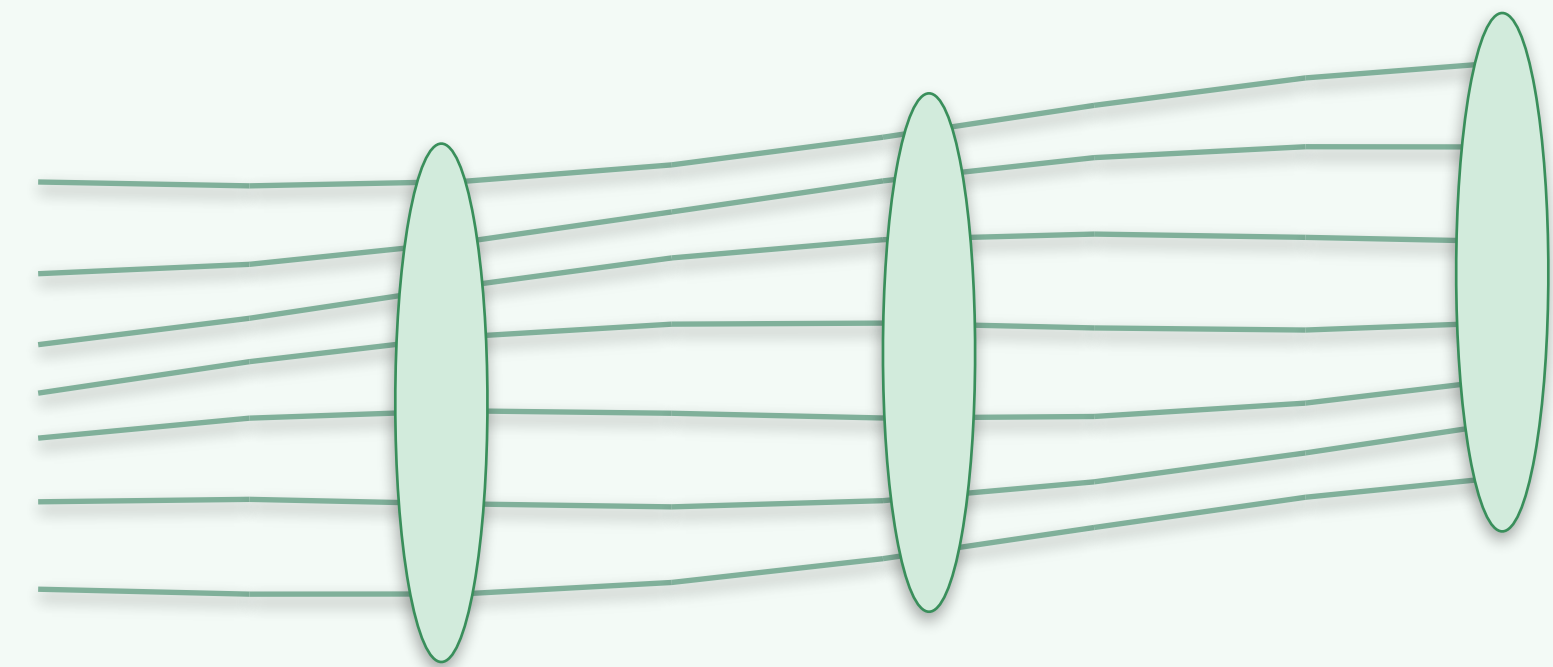
Optimization



Find the state that minimizes
background + observation mismatch

Ensemble DA

Probabilistic estimation



Use multiple forecasts to estimate
uncertainty + flow-dependent covariance

Combine observations and model information according to their uncertainty.

Traditional DA: Ensemble Kalman Filter (EnKF)

Represent uncertainty with an ensemble — then use its spread and covariance to update the state.

Core idea

One forecast is not enough.

Run multiple plausible forecasts (ensemble members).

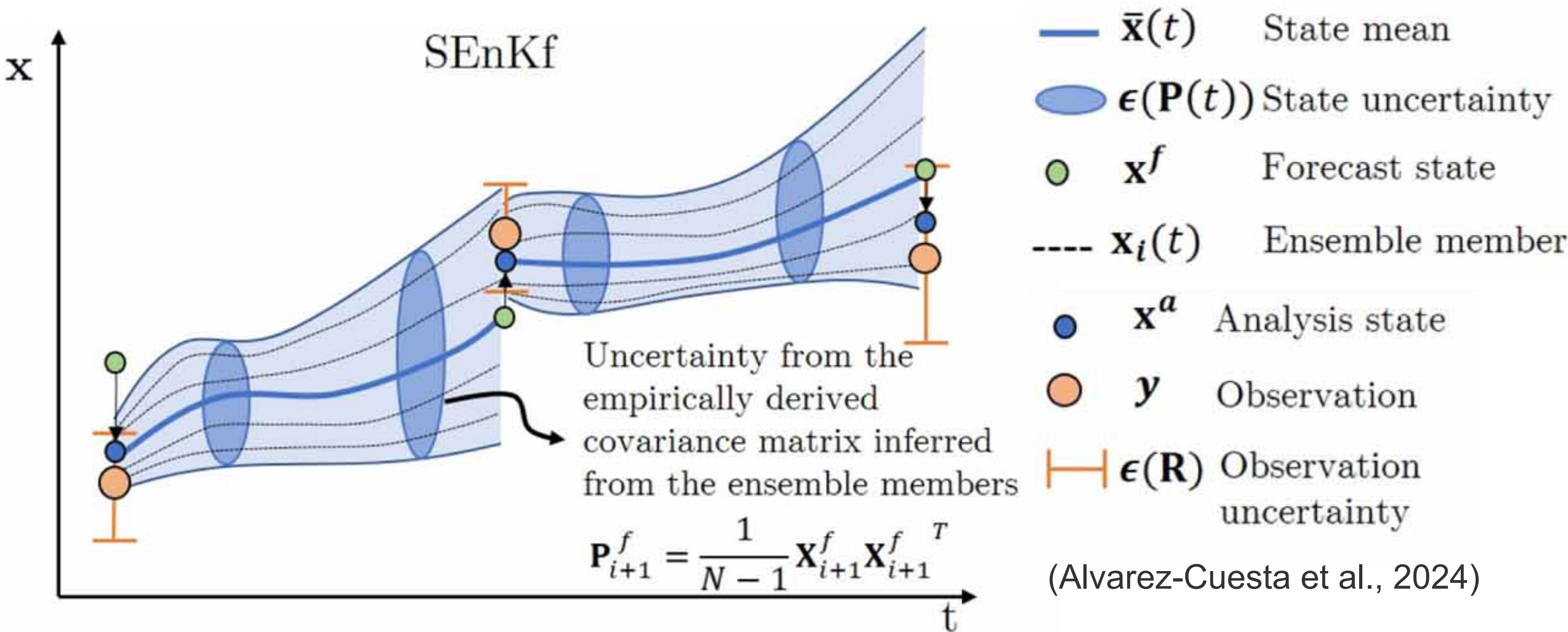


Ensemble spread → uncertainty

Ensemble covariance → how errors co-vary

Flow-dependent covariance

Forecast ensemble evolves with time



EnKF treats DA as probabilistic state estimation.

Traditional DA: Variational Data Assimilation

Find the state that best balances the background and observations over an assimilation window.

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}[\mathbf{H}(\mathbf{x}) - \mathbf{y}]^T \mathbf{R}^{-1}[\mathbf{H}(\mathbf{x}) - \mathbf{y}]$$

Minimize: Distance to background + Distance to observation

Optimization view

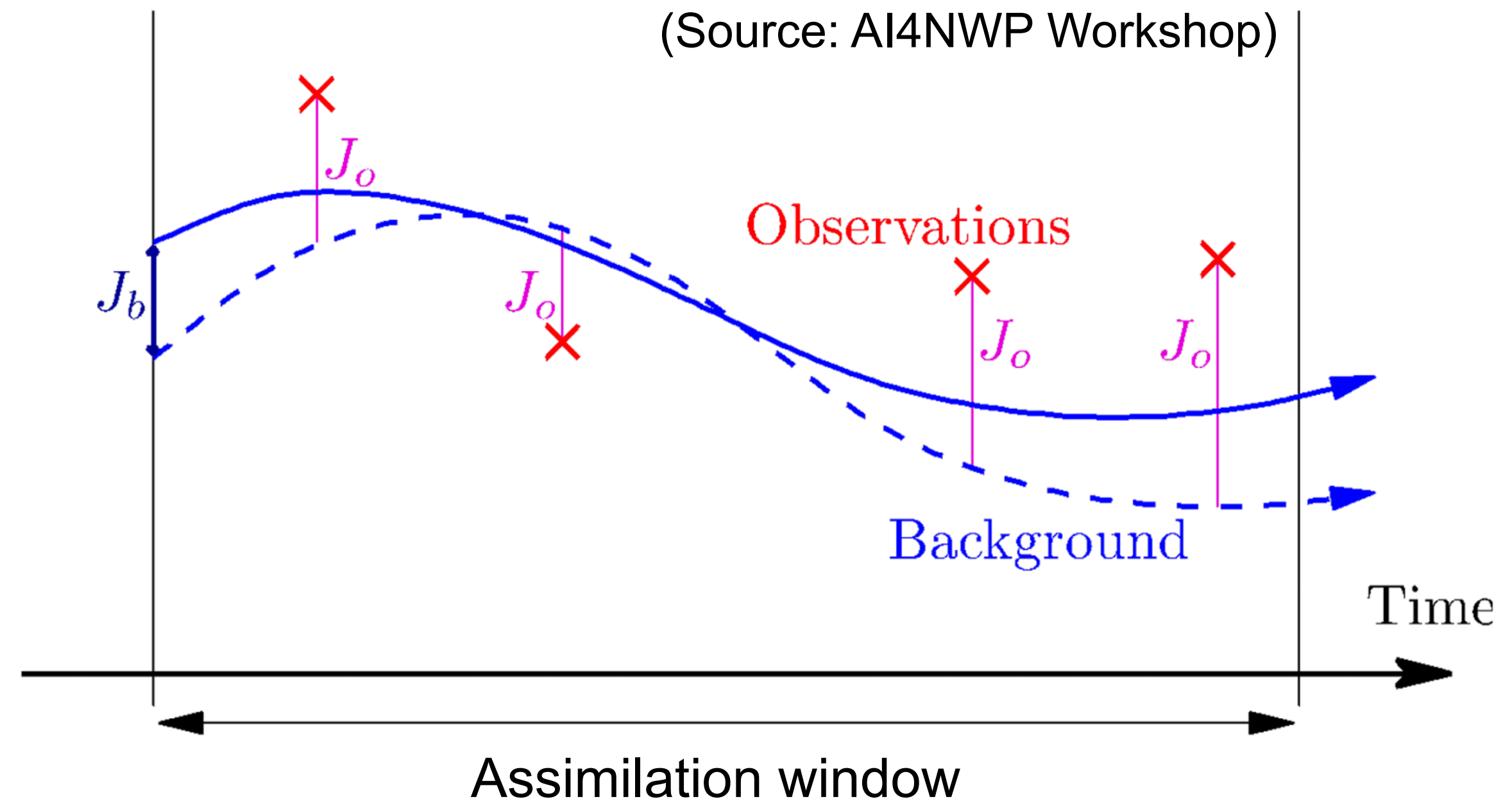
Start from the background



Compare model-equivalent values with observations



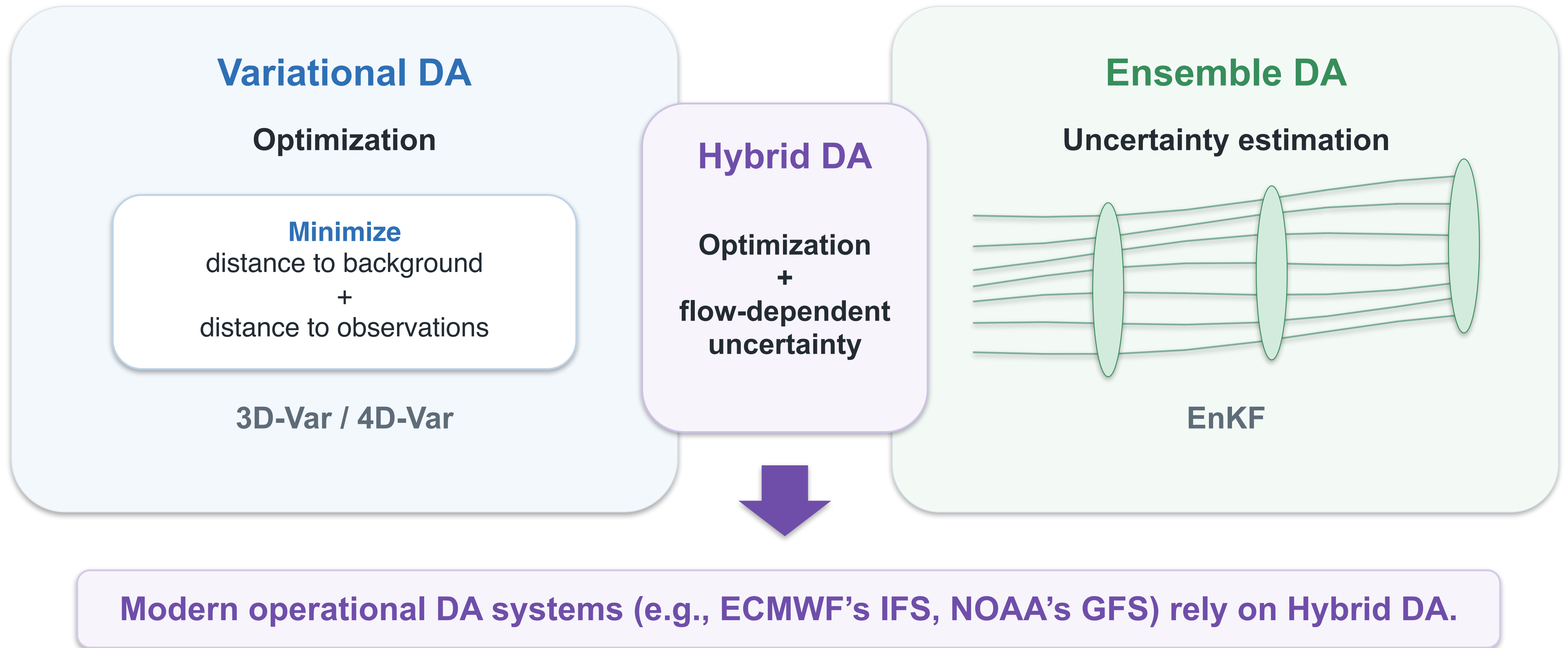
Iteratively minimize the mismatch



Variational DA treats state estimation as an optimization problem.

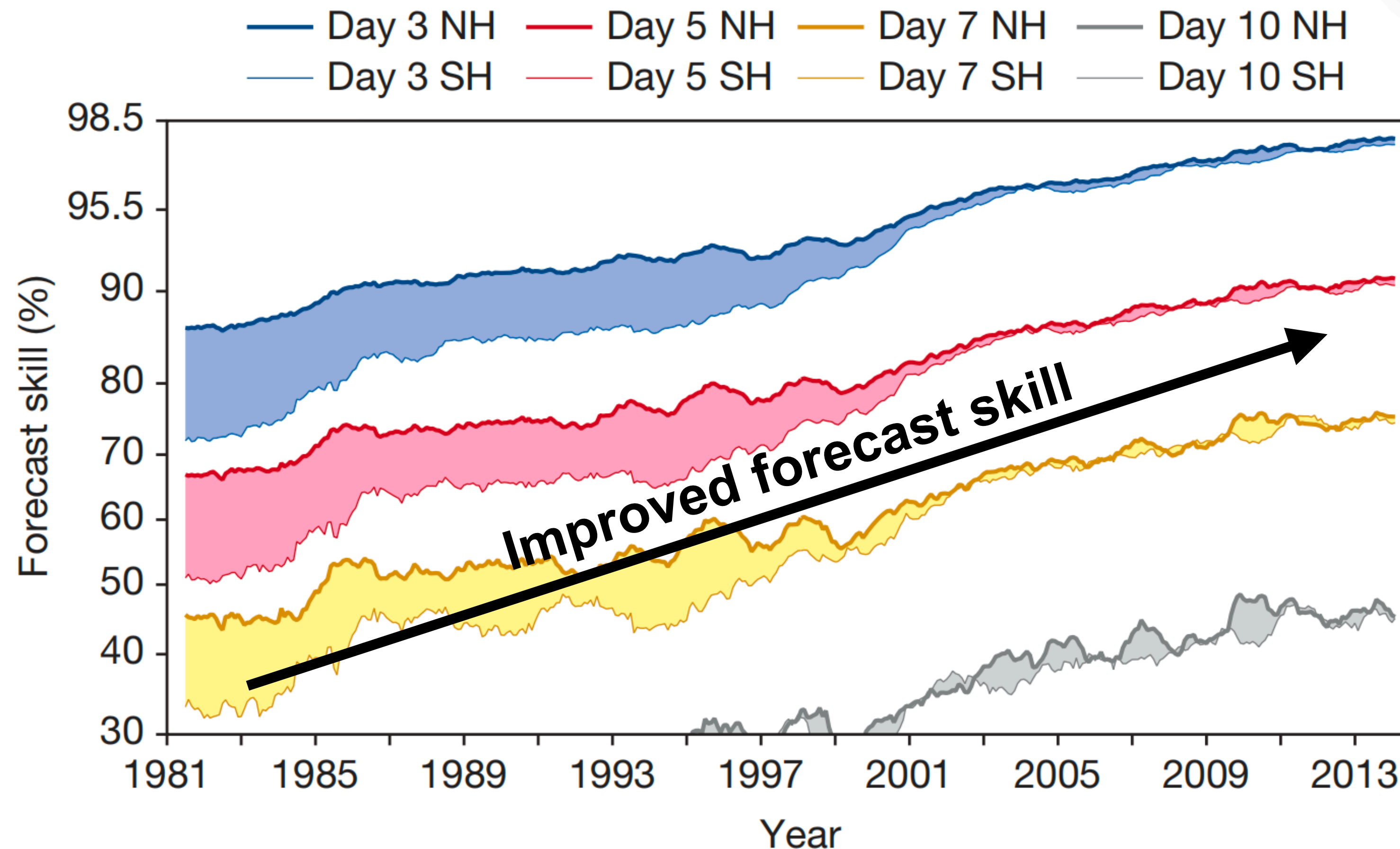
Modern DA Combines Both Ideas

Operational systems increasingly combine optimization with flow-dependent uncertainty.



Traditional DA Improves Weather Forecast Skill

Decades of advances in observations, models, and data assimilation have steadily increased forecast skill.



Forecast skill at 3-, 5-, 7-, and 10-day ranges (Bauer et al., 2015)

Why does the forecast skill improve?

More observations
→ especially satellite data

Better assimilation
→ more effective use of observations

DA converts observations into better initial conditions.

Challenge in Traditional DA

Traditional DA works well, but it becomes increasingly expensive and complex as the system grows.

1 High computational cost

3D-Var/4D-Var

Iterative optimization

Complex tangents

EnKF

Many ensemble members

High-resolution system

2 Curse of Dimensionality

Atmosphere state

($\sim 10^8$)



Massive covariance matrix
($\sim 10^{16}$)

3. Complex Observation Operators

Radiance from satellites

Observation operators
(nonlinear and high computational cost)



Temperature, humidity,...

Challenge in Traditional DA

Traditional DA works well, but it becomes increasingly expensive and complex as the system grows.

1 High computational cost

3D-Var/4D-Var

Iterative optimization

Complex tangents

EnKF

Many ensemble members

High-resolution system

2 Curse of Dimensionality

Atmosphere state

($\sim 10^8$)



Massive covariance matrix
($\sim 10^{16}$)

3. Complex Observation Operators

Radiance from satellites

Observation operators
(nonlinear and high computational cost)



Temperature, humidity,...

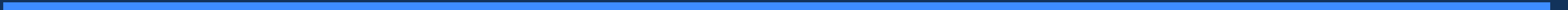
Can we use AI to overcome these challenges?

Part 2

AI-based Data Assimilation (DA)

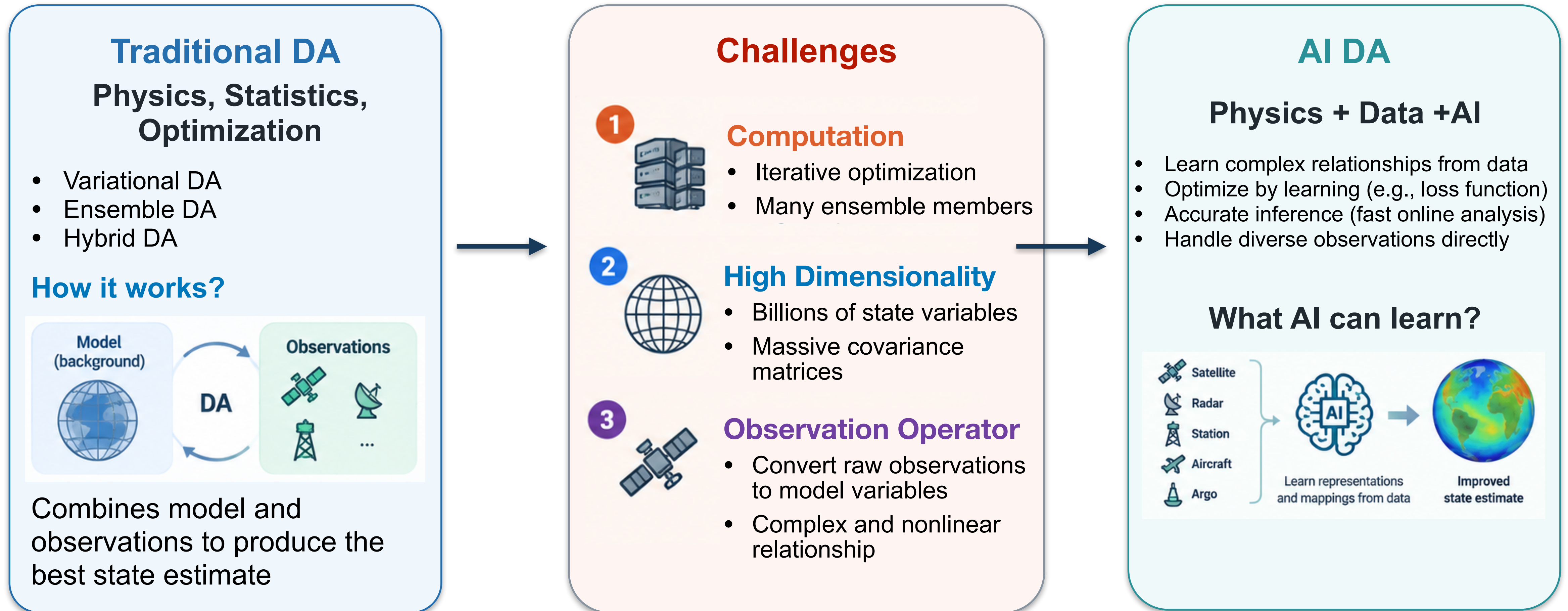
Why AI is useful for DA?

How AI can be integrated into DA?



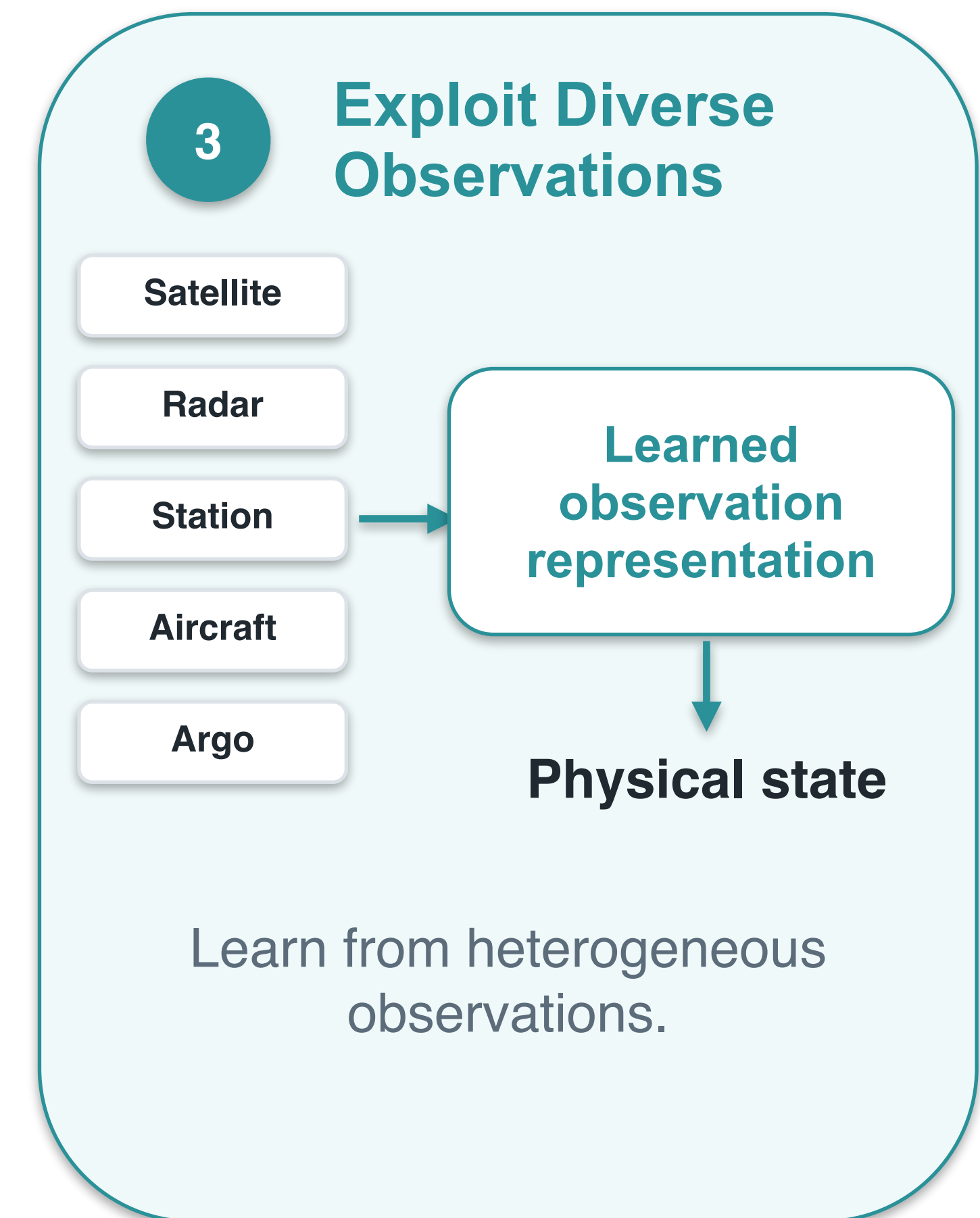
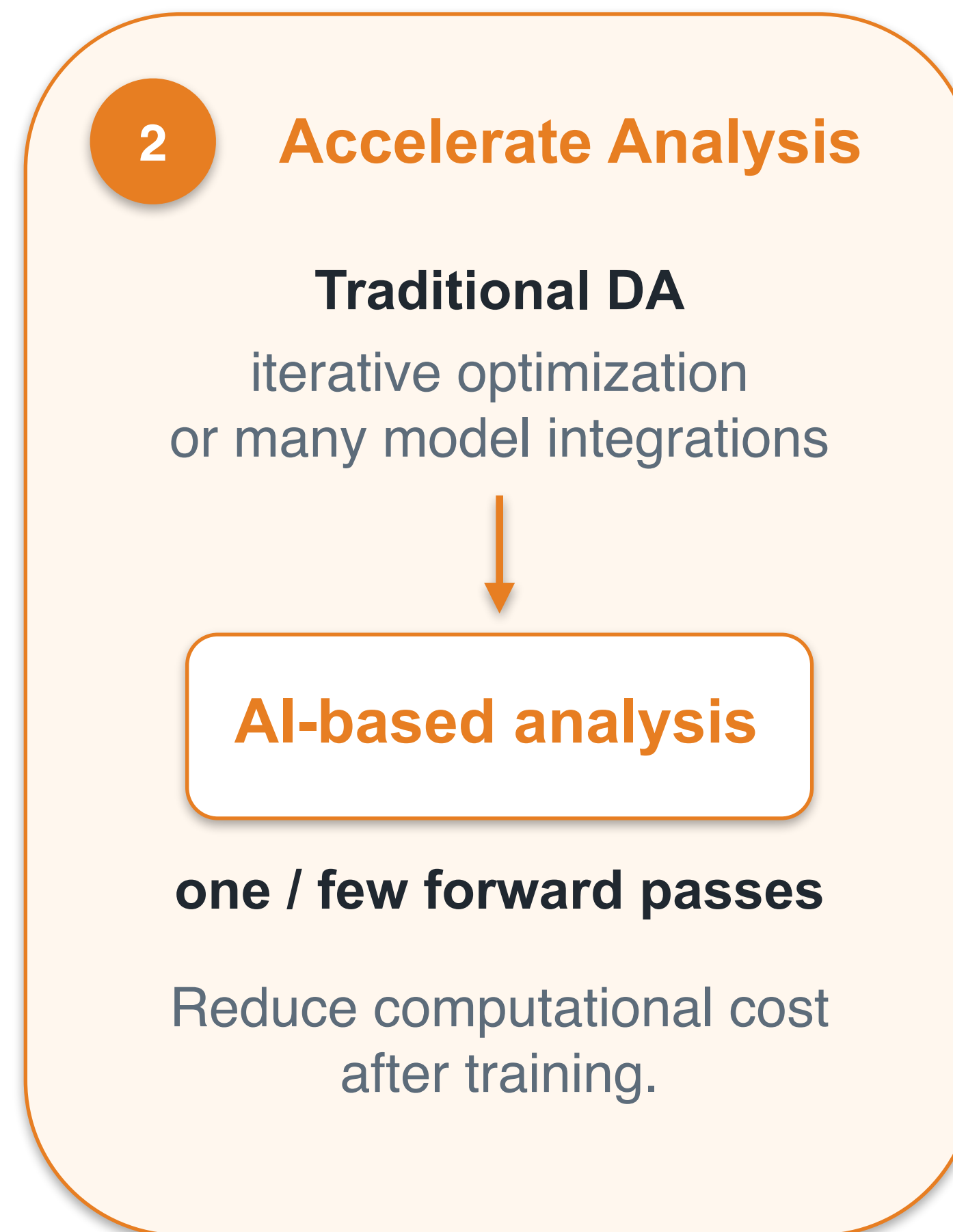
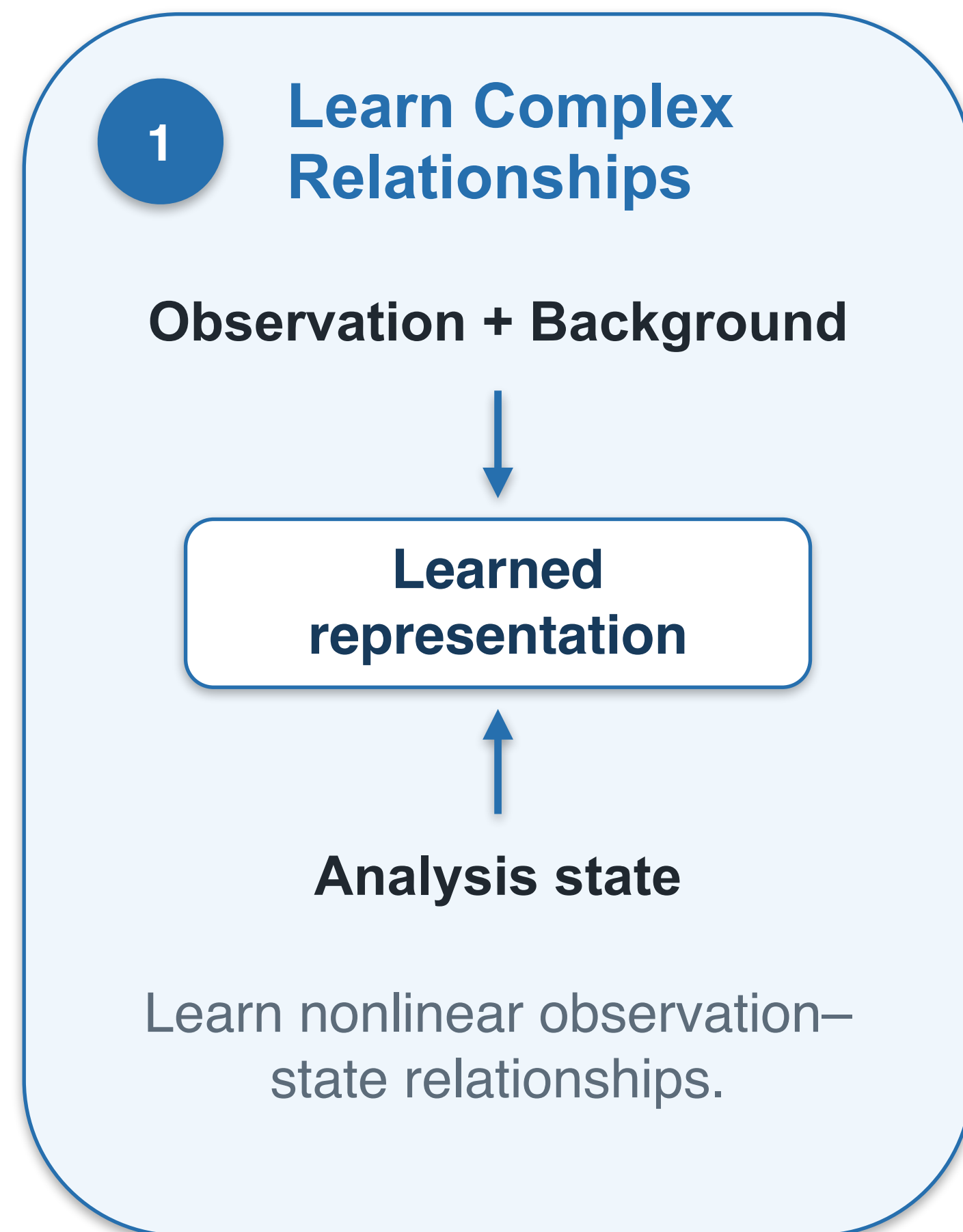
From Traditional DA to AI DA

The objective stays the same. Better estimate of the Earth system state by combining models and observations.

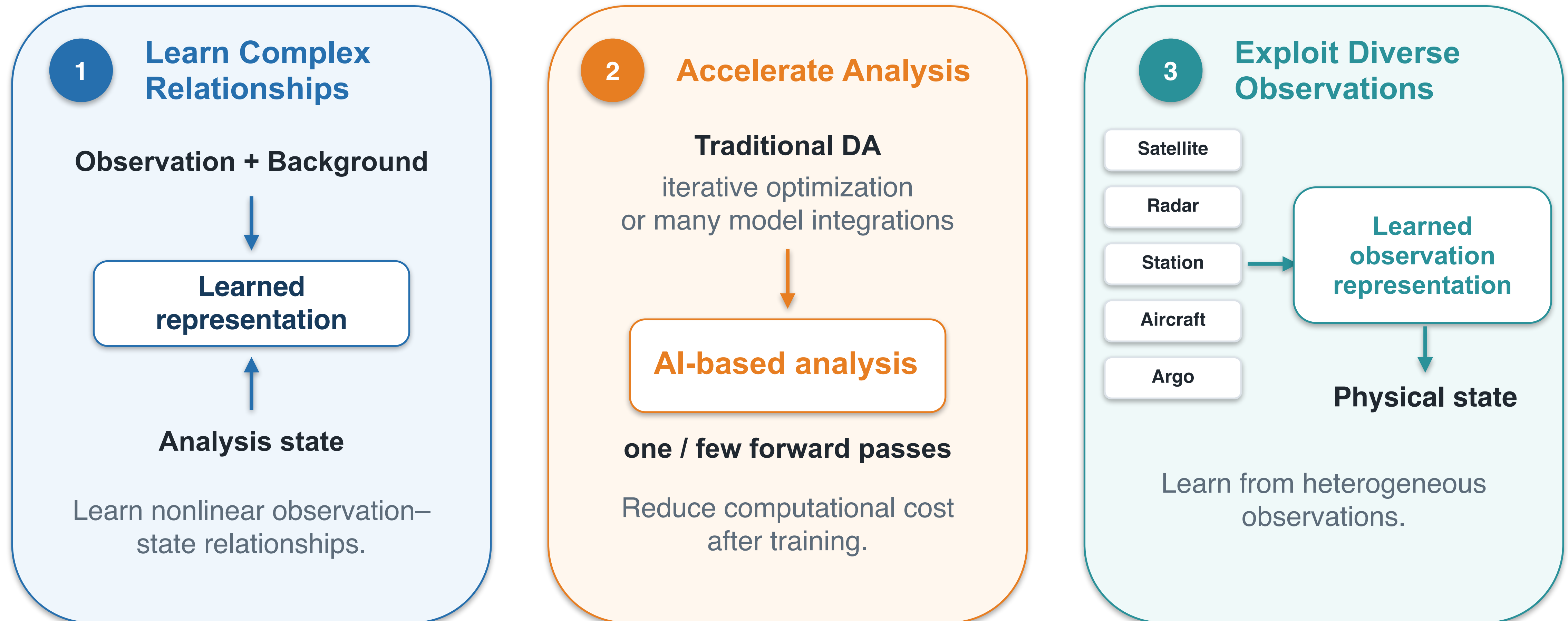


AI provides a new way to represent multi-sources information and accelerate the DA process.

Why AI for Data Assimilation?



Why AI for Data Assimilation?



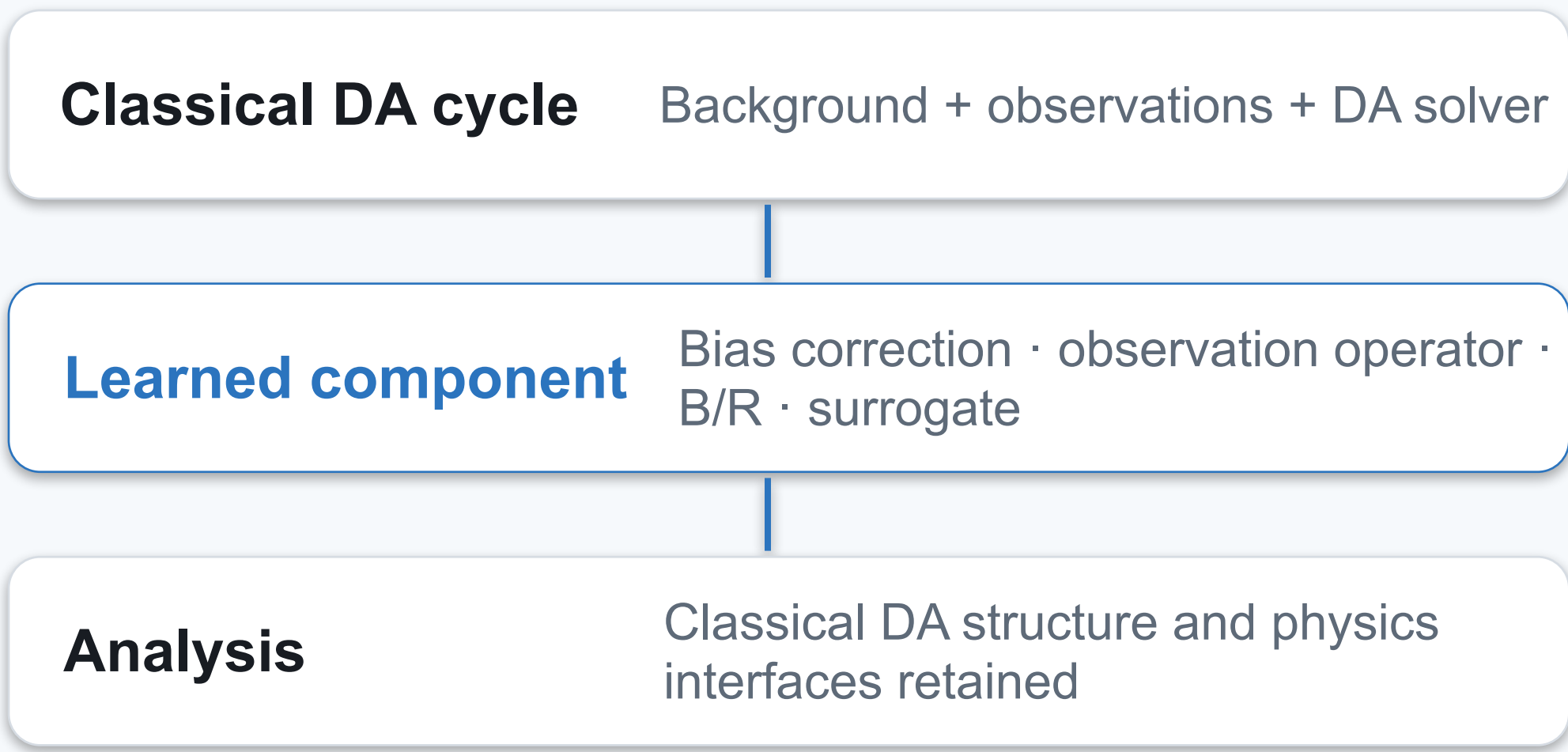
If you could improve only one of these three aspects, which one would be most important for your research?

Two Routes to AI-based Data Assimilation

AI can augment selected components of a classical DA system, or learn the observation-informed update more directly.

AI-assisted DA

Keep the DA framework; replace some components.

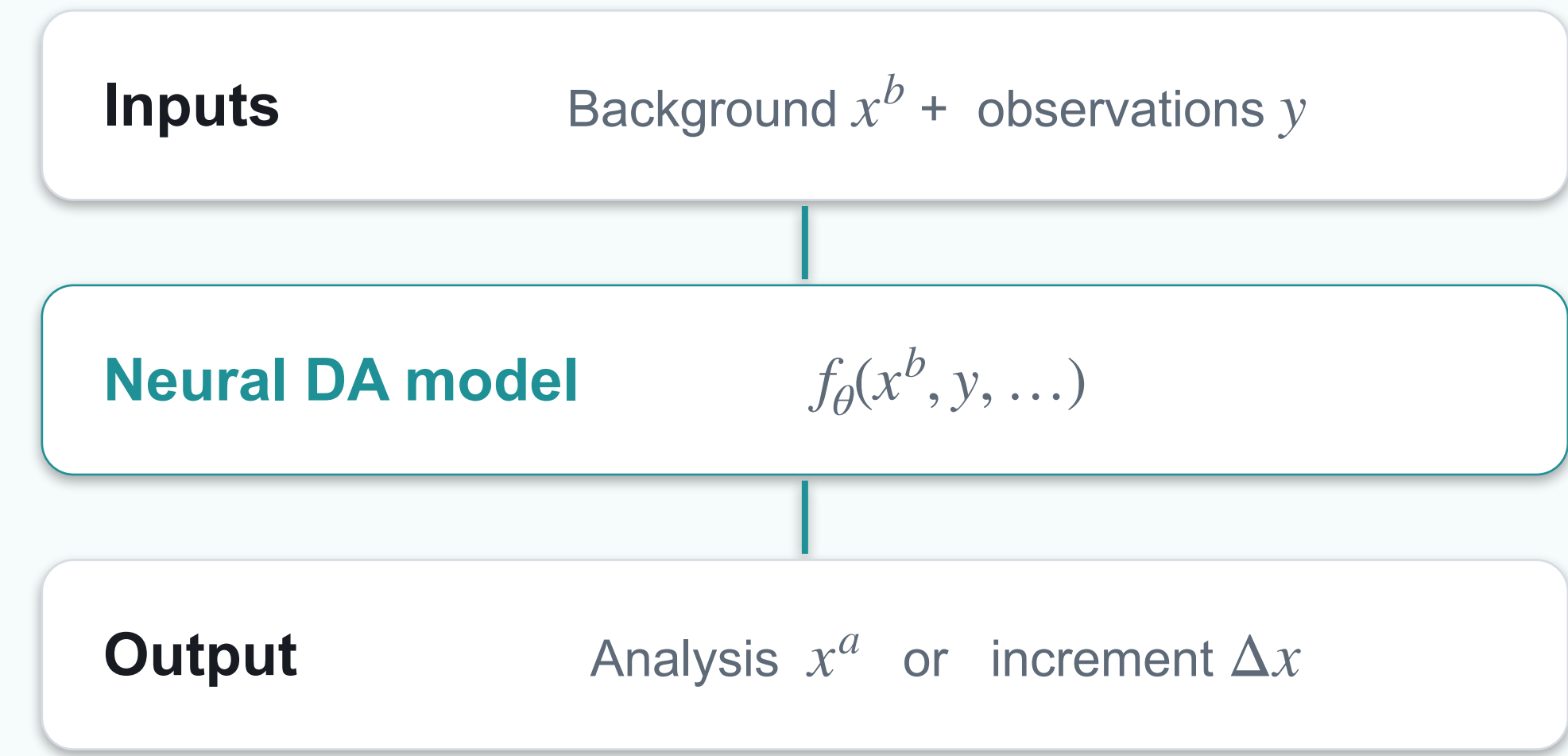


Advantage

Modularity · interpretability · operational compatibility

End-to-end DA

Jointly learn the analysis state or analysis increment.



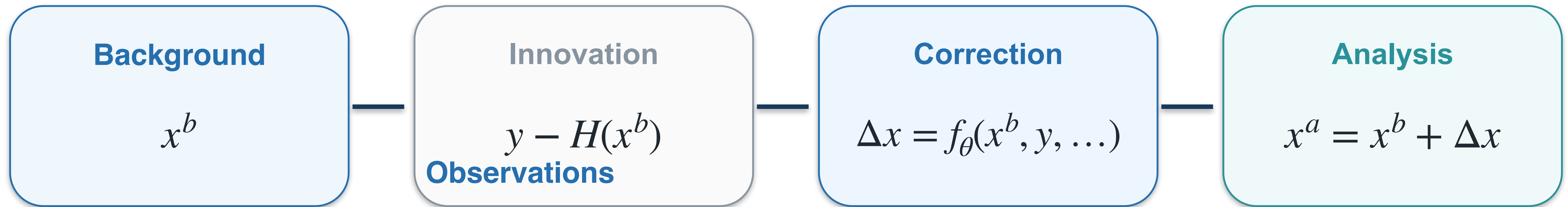
Advantage

Joint representation learning · fast online inference

AI Changes the way to estimate the Update — Not the Objective

The analysis is still an observation-informed correction of background.

What changes is how we estimate this correction.



AI can learn to predict this correction

Traditional DA $\Delta x = K(y - H(x^b))$

How Do We Know AI-DA Actually Works?

1 Analysis quality

RMSE / MAE vs. reference
Spatial / vertical / regime-wise errors

2 Forecast impact

Forecast skill initialized from analysis
Error growth with lead time
Extreme-event performance

3 Physical consistency

Balance and dynamical compatibility
Conservation / stability diagnostics
No spurious small-scale structures

4 Efficiency & robustness

Latency, memory, throughput
Scaling with observations / resolution
Sensitivity, ablation, failure regimes

**Low analysis RMSE alone is not enough —
the analysis must also support better forecasts.**

AI is an Opportunity — Not a Free Pass

Goal: Build a reliable state-estimation system in real-world applications.

Issues to Consider

Generalization	New seasons & climates	Extreme events	New sensors / missing sensors	Model or observing-system changes
Physical consistency	Balance	Conservation	Dynamical compatibility	Stable forecast initialization
Uncertainty	Observation & background errors	Ambiguity / non-uniqueness	Probabilistic analysis	Calibration
Interpretability & trust	Which observations matter?	Where does information propagate?	Sensitivity / ablation	Failure regimes

Part 3

Our research Two cases of AI-DA

1. **ADAF (Regional Near-Surface Atmospheric DA)**
 2. **ADAF-Ocean (Sea Surface DA)**
-

Regional Near-Surface Atmospheric DA

Learning observation-informed corrections on a regular grid

Gridded Inputs

Background

1-h forecast

Surface obs.

station fields

Satellite

radiance fields

Terrain

topography

Use a neural network to learn analysis increment Δx

Encoder

feature
extraction

Fusion

multi-source

Reconstruction

predict
increment

Δx

analysis
increment

$$x^a = x^b + \Delta x$$

Background + learned correction (Analysis increment)

Setup

Domain

CONUS

Resolution

5 km
Hourly

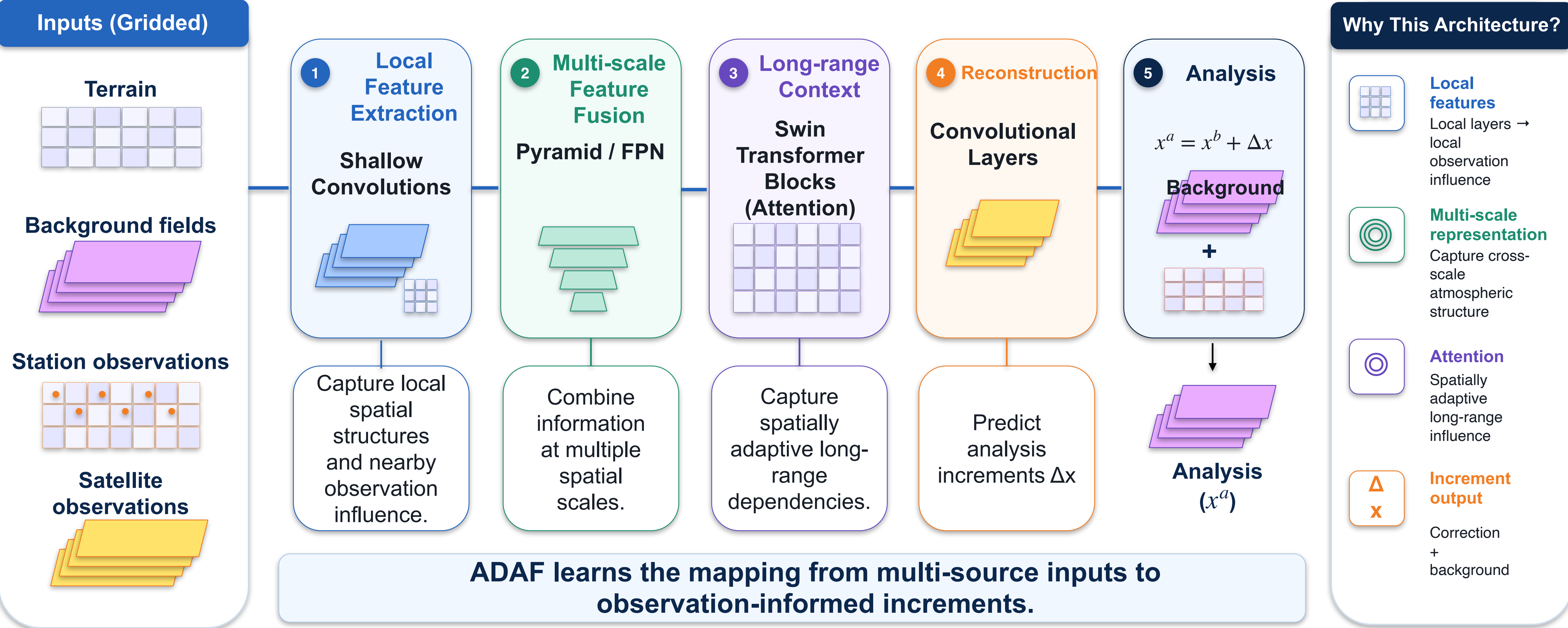
Variables

T2m Q2m
U10m V10m

Can a **neural network** learn accurate, high-resolution **analysis increment** without explicitly constructing a huge covariance matrix?

ADAF Architecture: Local Features and Long-Range Context

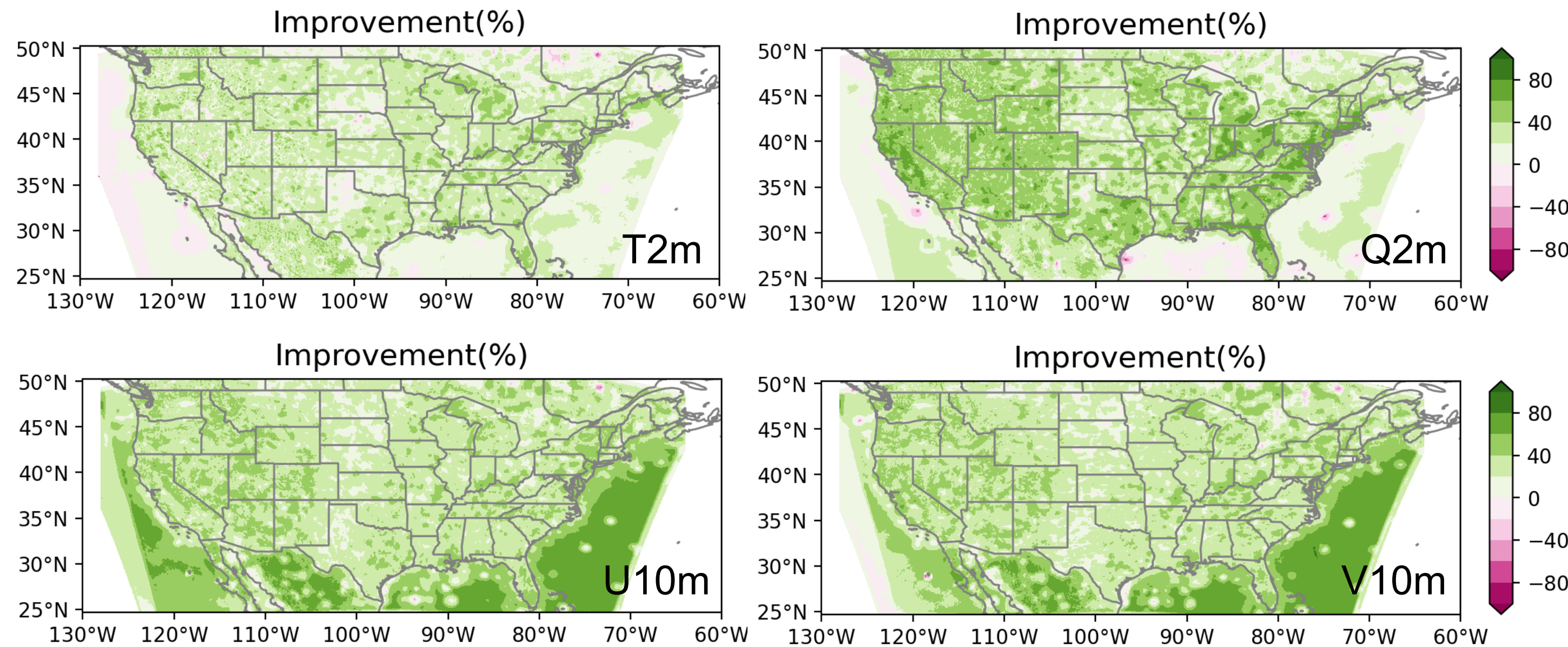
ADAF learns observation-informed increments on a grid by combining local, multi-scale information, and adaptive long-range influence.



ADAF Improves Analysis Accuracy Across Most of the Domain

MAE reduction relative to HRRRDAS across four near-surface variables

Spatial distribution of MAE reduction (%)



16 ~ 35% lower regional-mean MAE

≈ 85% of the domain improved

≈ 2 s per assimilation

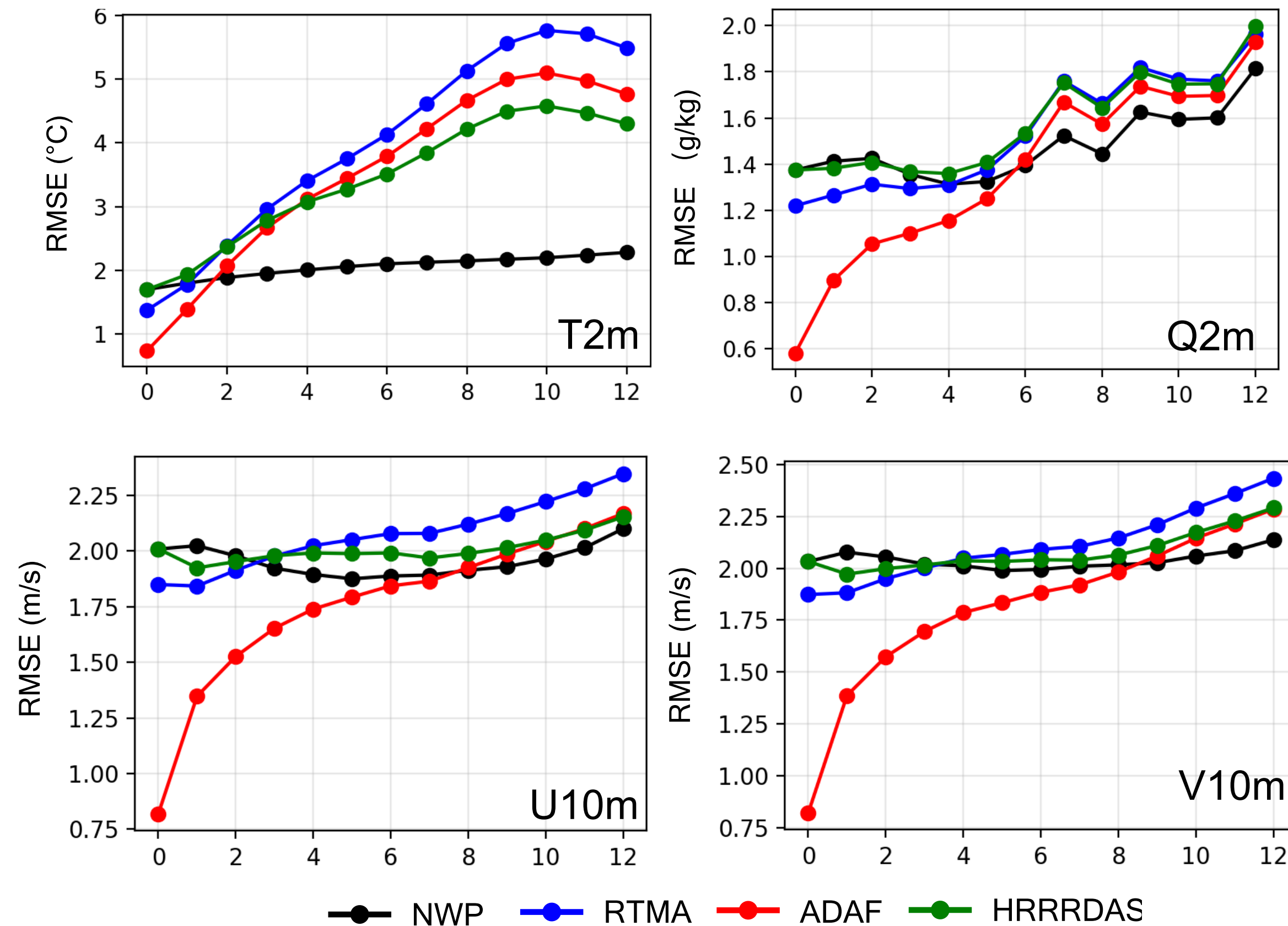
$$\text{MAE reduction (\%)} = \frac{MAE_{HRRRDAS} - MAE_{ADAF}}{MAE_{HRRRDAS}} \times 100\%$$

> 0 (Green): ADAF improves

Does a Better Analysis Lead to a Better Forecast?

Forecast improvement is the stronger test that ADAF produces a useful initial state.

Forecast Error (use an AI forecasting model)



0–6 h

Strongest gains

Q2m & winds

Largest benefits

Longer leads

Benefits diminish

Better analysis

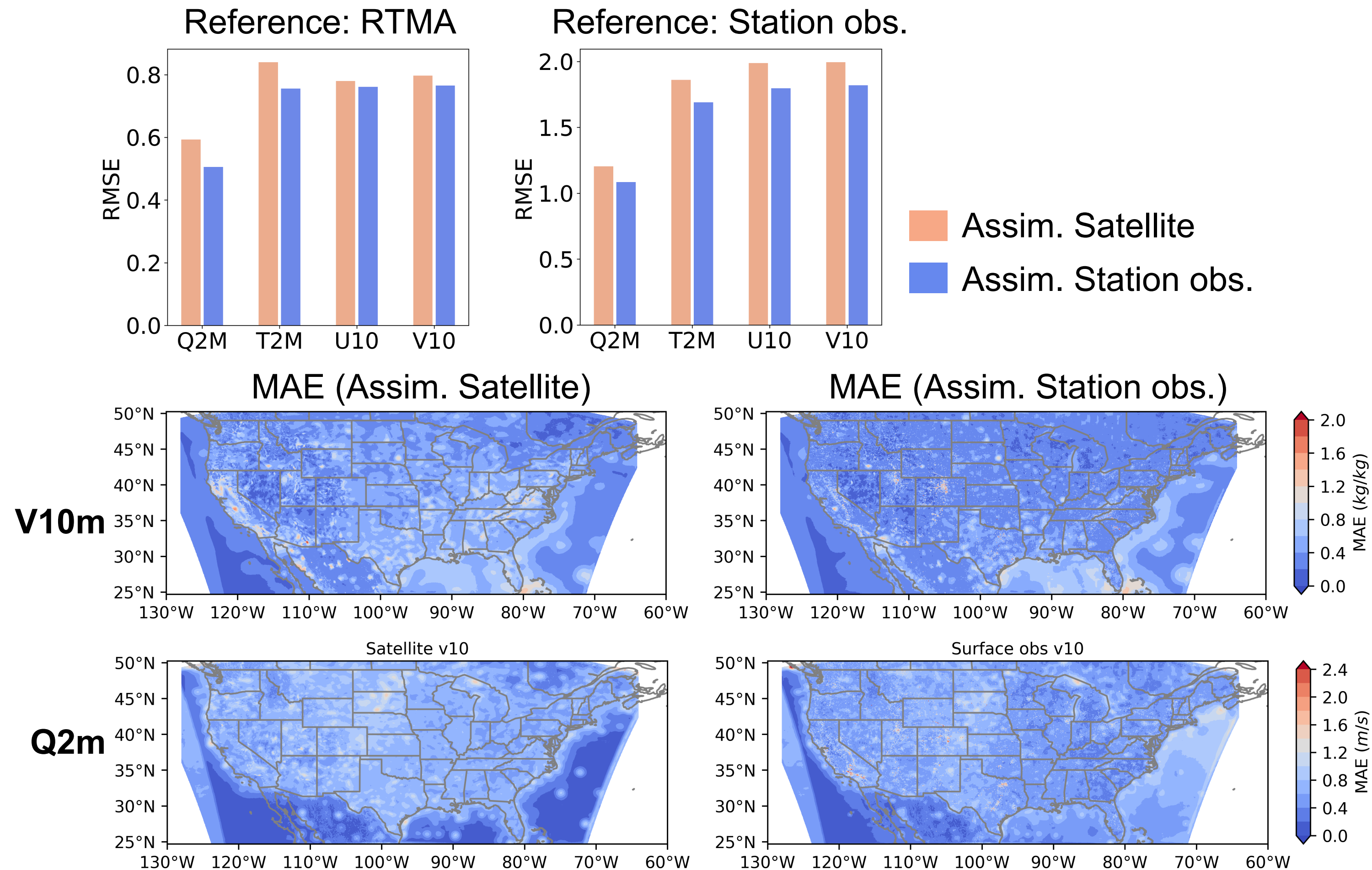


Better short-range forecast

Analysis produced by ADAF
is useful for forecasting

What Has ADAF Learned from Different Observations?

Ablation experiments reveal complementary roles of surface stations and satellite observations.



What the model learned?

**Surface
stations**

**Strong local
constraints**

Satellite

**Broad
complementary
information**

Complementary

ADAF learns source-dependent observation influence.

- **Surface stations provide strong local constraints.**
- **Satellites add spatially broad information in poorly observed regions.**

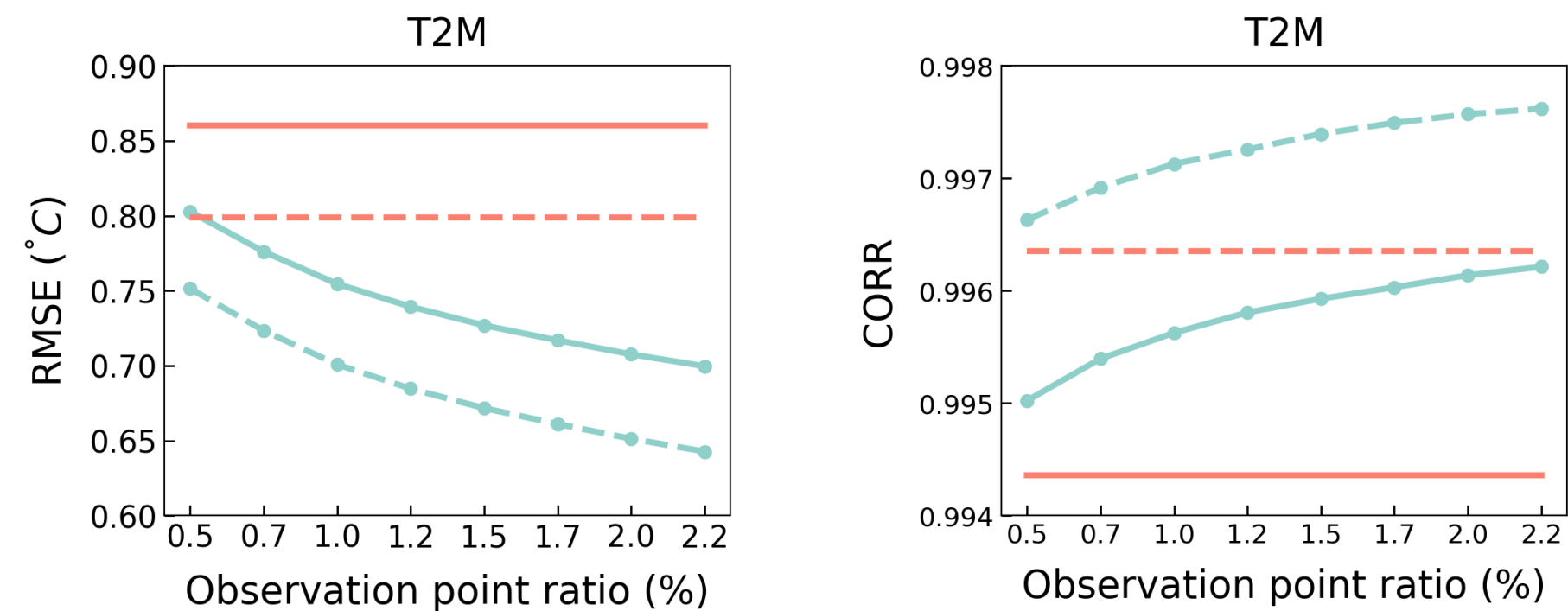
How Robust Is Grid-based AI-DA?

Does the model still work when the input information become worse?

1 Fewer Observations

Observation density ↓ → RMSE ↑

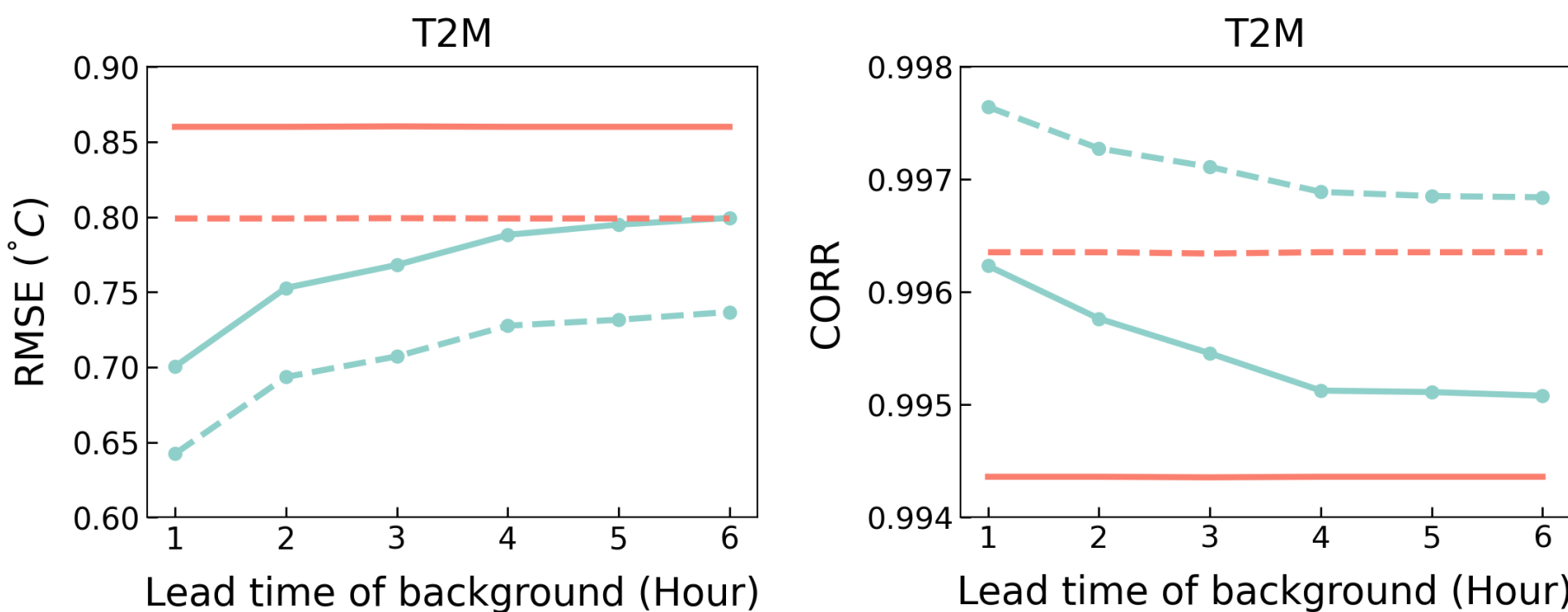
✓ **ADAF remains competitive**
even under sparse observations.



2 Worse Background

Background lead time ↑ / Background quality ↓
→ Analysis accuracy ↓

✓ **Analysis degrades**
but maintains clear advantages.



— ADAF - - - ADAF (land) — HRRRDAS - - - HRRRDAS (land)

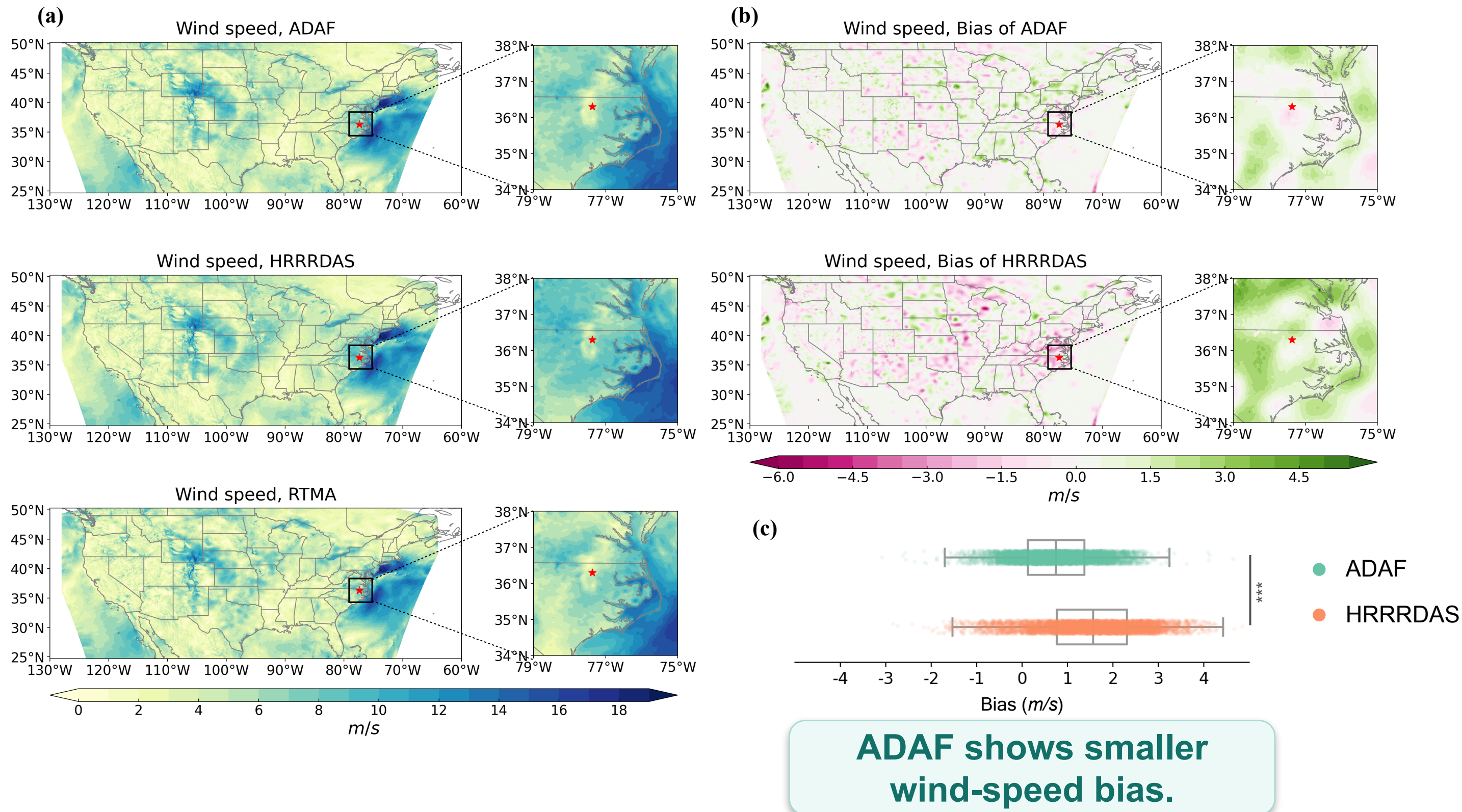
The learned update remains effective across changes in observation density and background quality.

Can Grid-based AI-DA Handle High-impact Weather?

Case study: Tropical Storm Ophelia · 23 September 2023 · 18:00

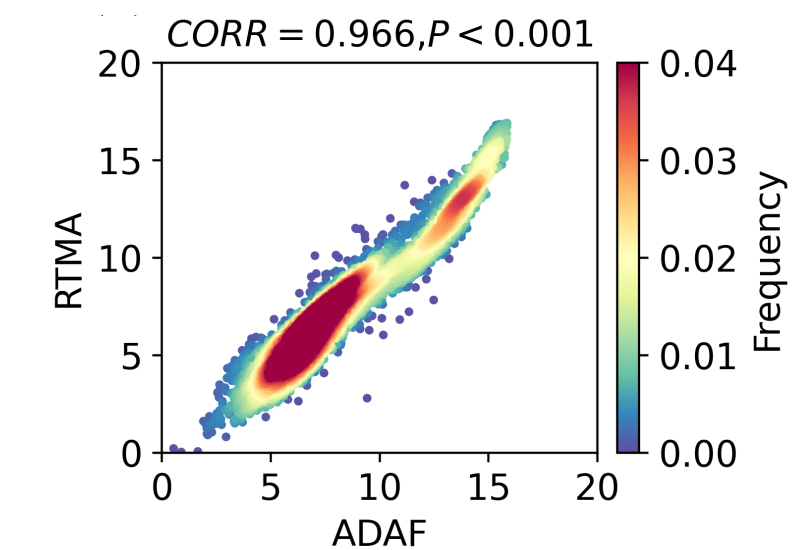
1 Can ADAF recover the storm wind structure?

Compare ADAF and HRRRDAS against RTMA

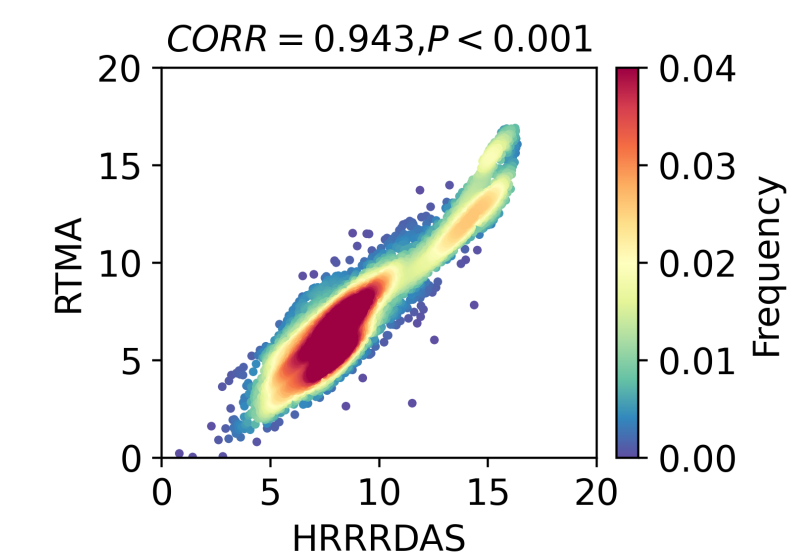


2 Quantitative verification

Agreement with the RTMA



ADAF
 $r = 0.966$



HRRRDAS
 $r = 0.943$

- ✓ Storm-scale wind pattern is preserved
- ✓ Bias is reduced relative to HRRRDAS
- ✓ Higher agreement with RTMA

High-impact case: ADAF retains the observed storm structure while reducing wind-speed bias and improving agreement with RTMA.

What Have We Learned from Grid-based AI-DA?

Predict the analysis increment successfully — but the observation representation is grid-constrained.

✓ What grid-based AI-DA does well

It learns the update efficiently on a regular grid.

✓ Nonlinear mapping

Learn observation–state relationships

✓ Multi-source fusion

Combine background, stations, satellite, terrain

✓ Fast inference

Generate analysis increments in a few forward passes

Are we finished?

What Have We Learned from Grid-based AI-DA?

Predict the analysis increment successfully — but the observation representation is grid-constrained.

✓ What grid-based AI-DA does well

It learns the update efficiently on a regular grid.

✓ Nonlinear mapping

Learn observation–state relationships

✓ Multi-source fusion

Combine background, stations, satellite, terrain

✓ Fast inference

Generate analysis increments in a few forward passes

What remains constrained

Irregular observations are converted into regular grid.

Native observations are irregular



Regridding / interpolation

May introduce errors



Regular grid

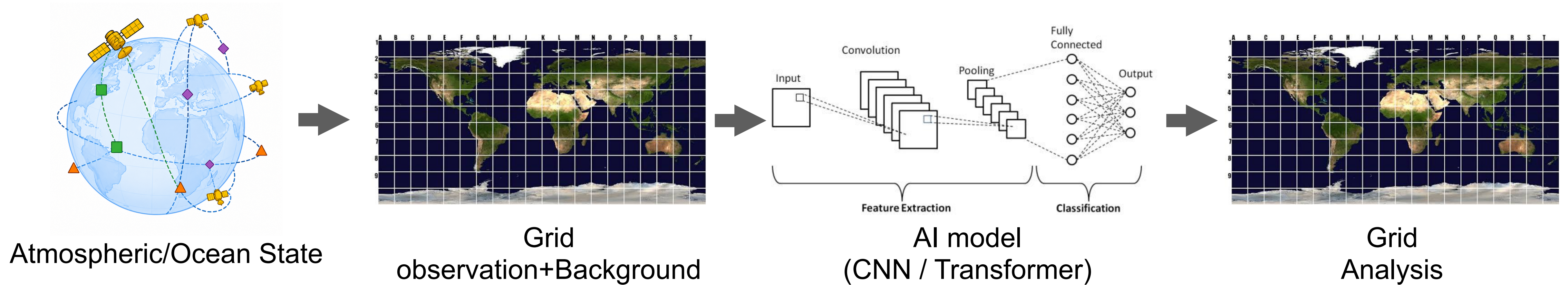


Grid-based neural network

Can we learn directly from observations in their native coordinates without forcing them onto a grid first?

Part I — Let's rethink grid-based AI-DA

Grid-based AI-DA Paradigm

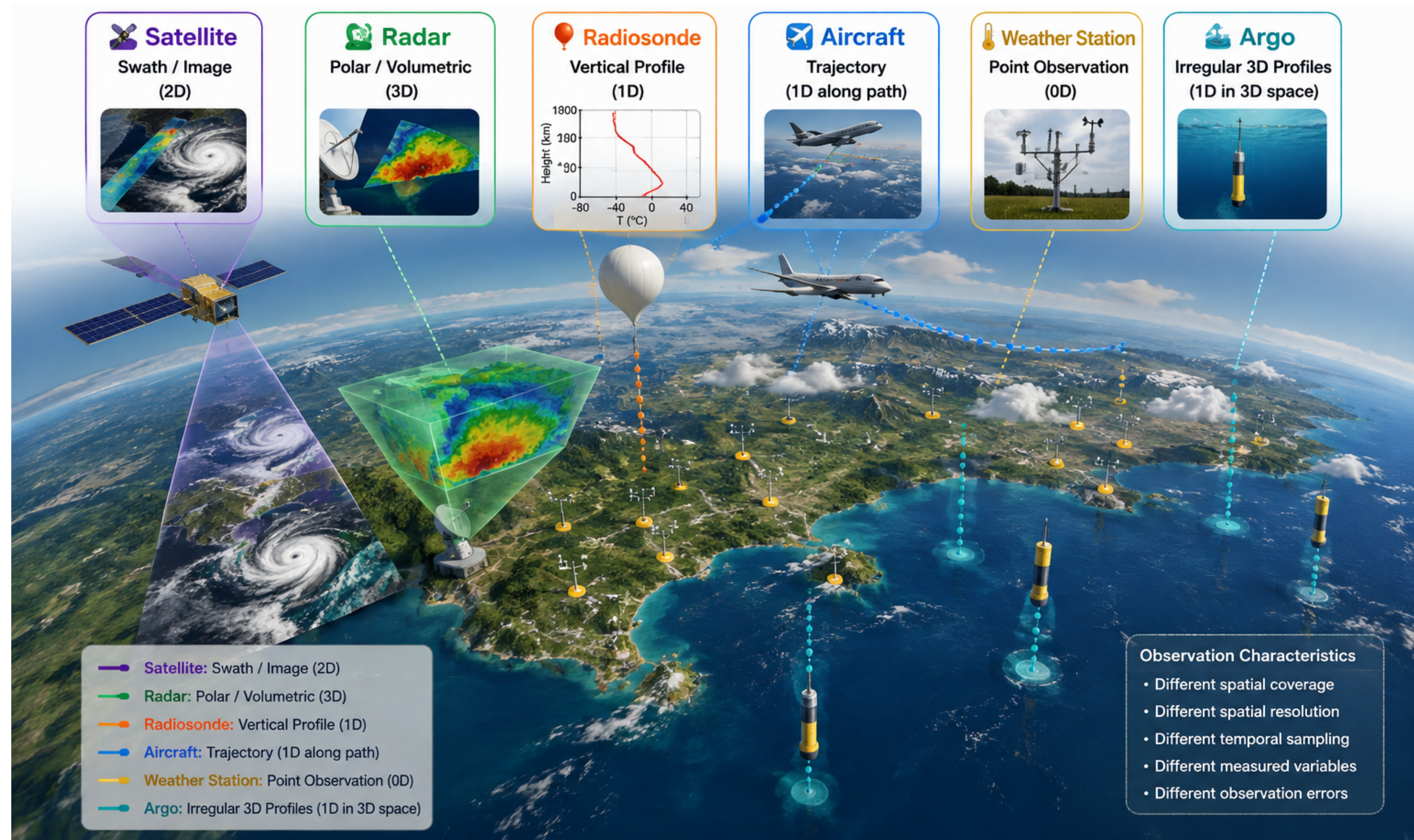


Grids provide a convenient representation of the Earth system.

But are observations naturally grid-based?

Earth Observations Have Different Geometries

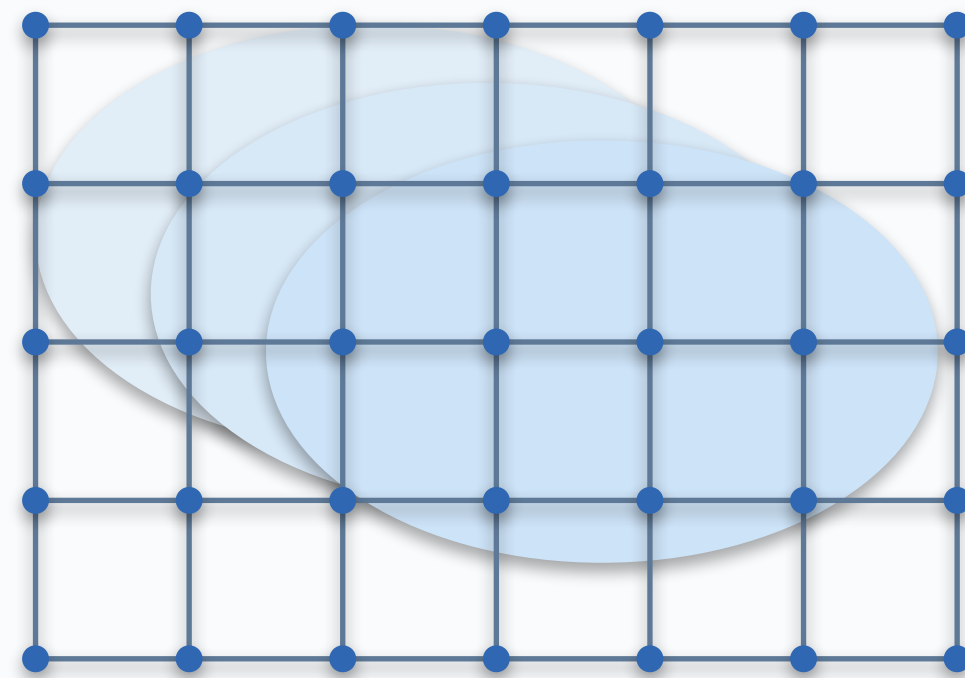
Different observing systems sample the Earth system in fundamentally different ways.



Model Grid vs. Observation Geometry

Model

Regular Grid



Dense · Structured

Representation Gap

Grid Space

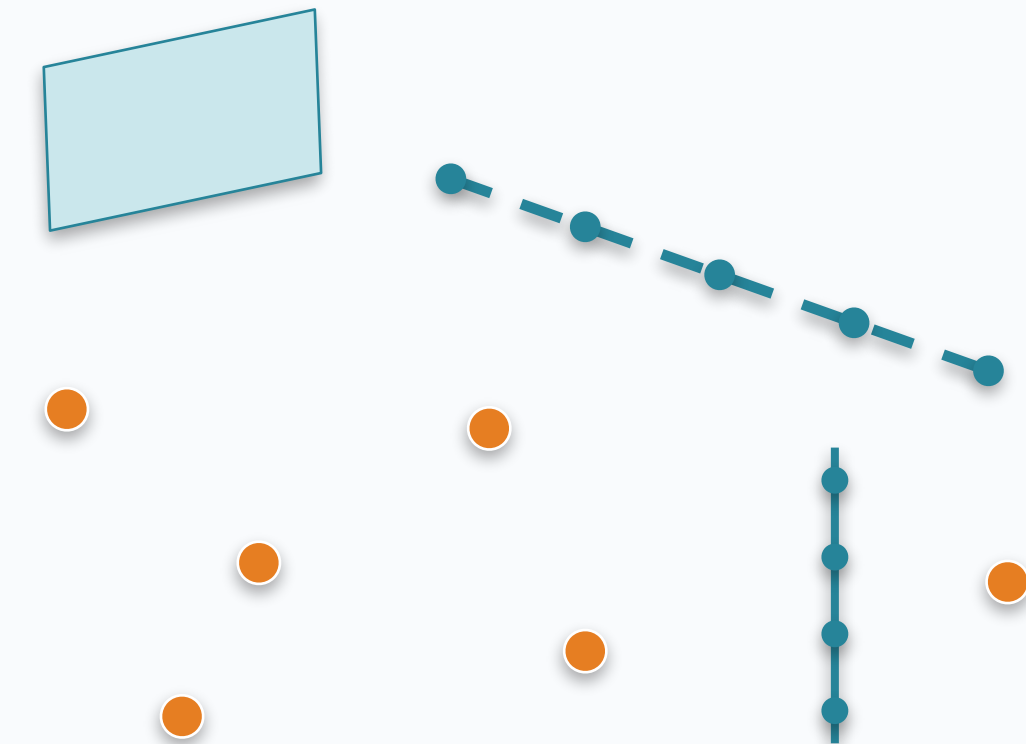
≠

Observation Space

Different sampling geometries
Different spatial resolutions

Observations

Different Geometry



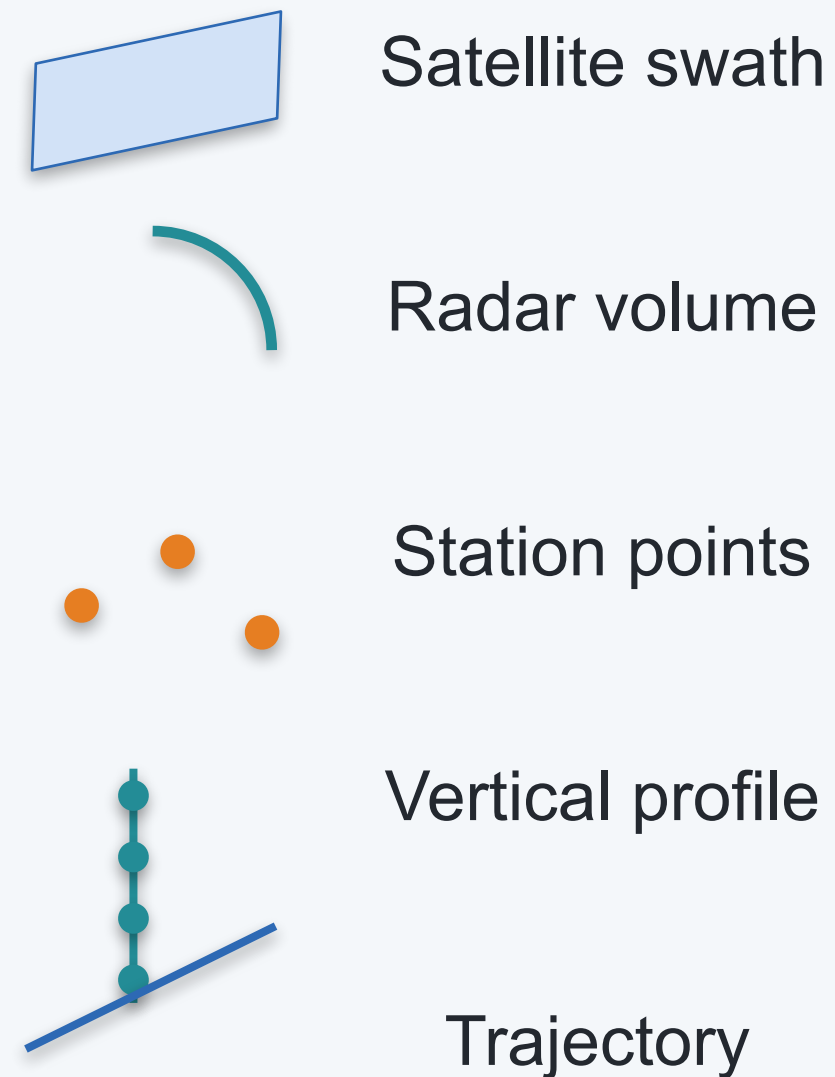
Irregular · Sparse · Heterogeneous

Part II — What does regridding actually do?

How Does Grid-based DA Bridge the Representation Gap?

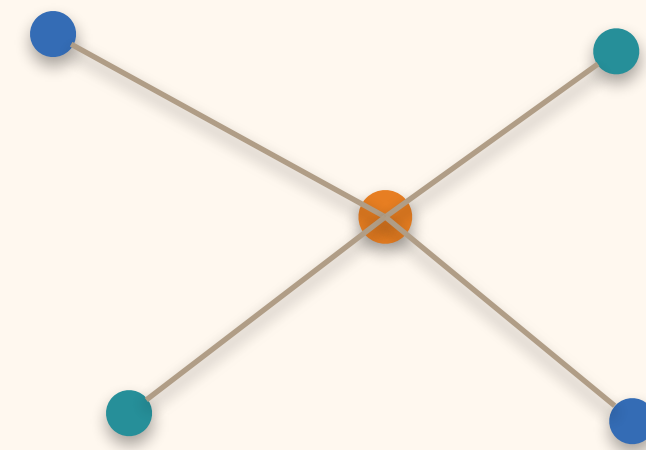
1 Native observations

Heterogeneous Geometry



2 Grid alignment

Interpolation & Re-gridding

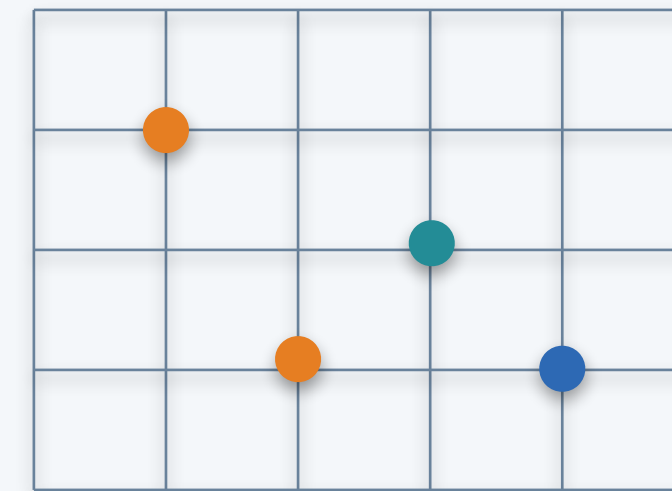


$$v_g = \sum_{i=1}^N w_i v_i$$

Map native observations to predefined grid points

3 Grid-based fusion

Common Regular Grid

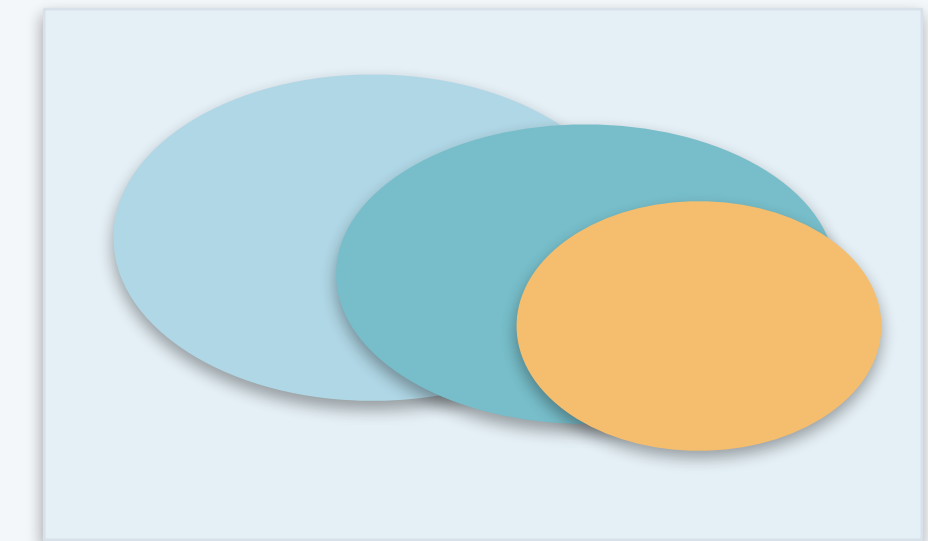


Grid-based AI Model
CNN / ViT

Grid-based information fusion

4 Analysis

Analysis Field



Output aligned with
the model grid

Regridding acts as a bridge between irregular observation space and regular model space.

Why Has Grid-based DA Been Successful?

A regular-grid representation aligns naturally with numerical weather prediction.

01

✓ Efficient Computation

GPU-friendly tensors

02

✓ Mature Architectures

CNN, U-Net, ViT

03

✓ Easy Integration with NWP

Analysis output directly matches model grids

04

✓ Effective for Dense Observations

Radar and satellite imagery are grid-formatted data.

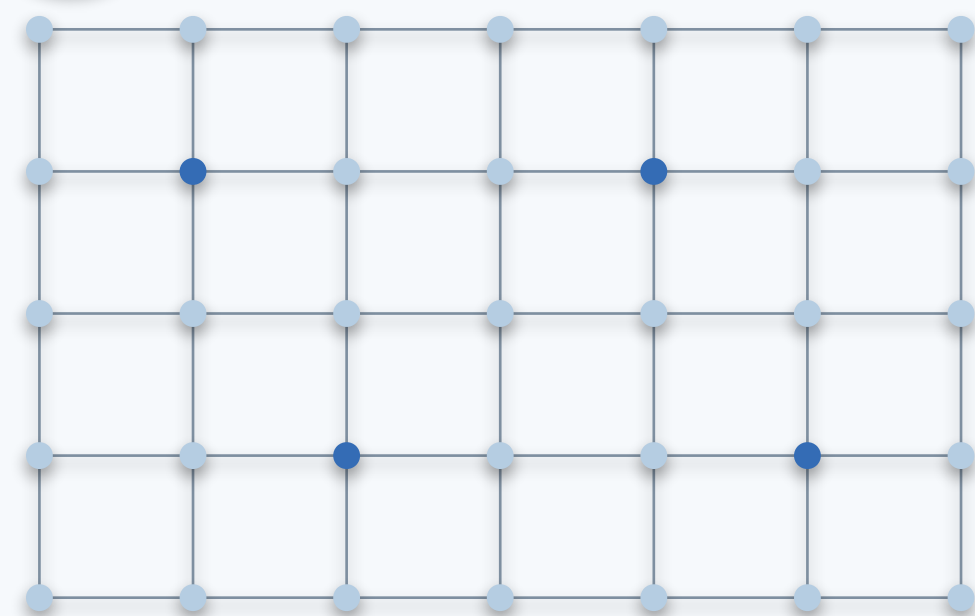
Grid-based representation is highly effective when the data are grid-formatted.

Any limitations?

What Do We Lose During Re-gridding?

Re-gridding simplifies data fusion, but it may loss the original observation information.

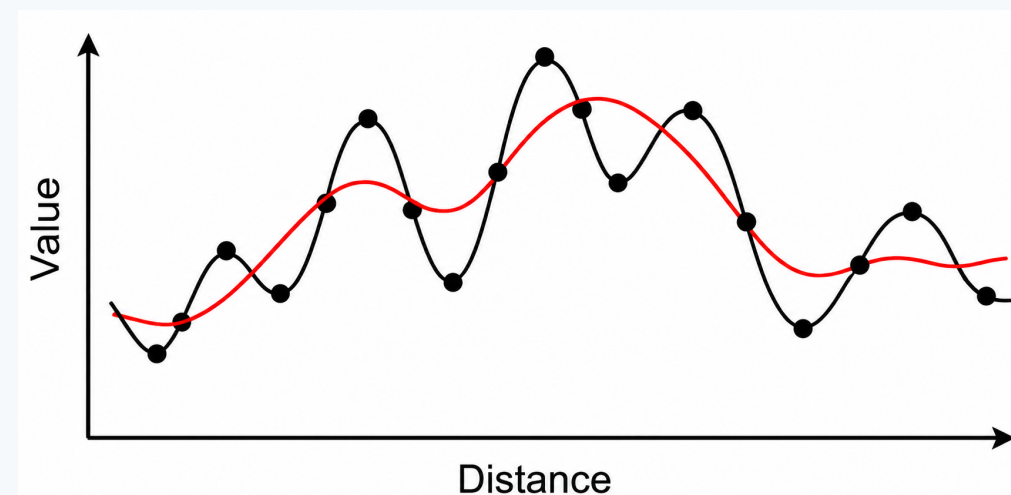
01 Artificial Values



- Original Observation
- Interpolated Value

Problem:
Values introduced where no observation exists.

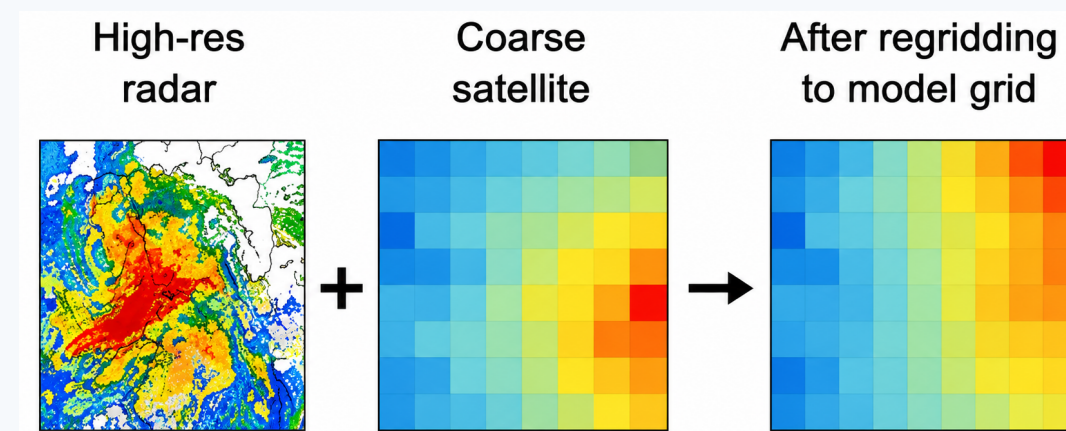
02 Smoothing



- Observations
- True value
- Interpolated

Problem:
Extremes and fine-scale structures are weakened or lost.

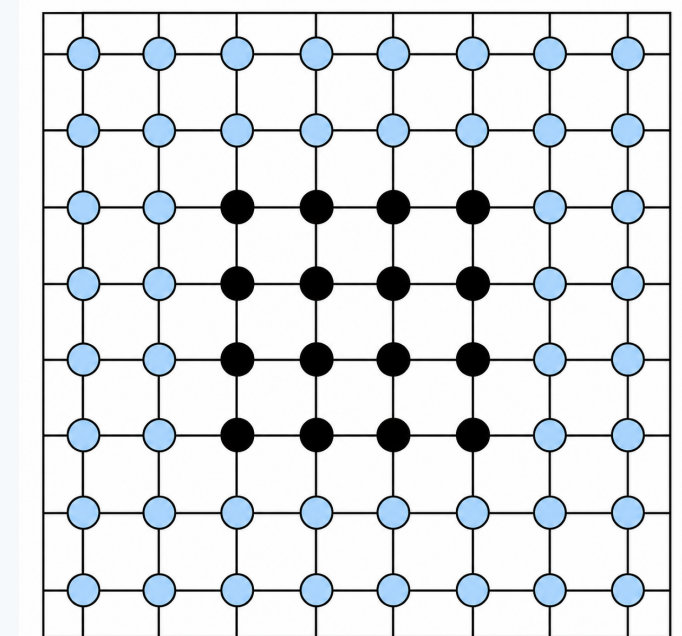
03 Resolution Mismatch



Different sensors are forced onto one grid.

Problem:
Important scale information may be lost or misrepresented.

04 Computational Redundancy



- With observation
- Without Observation (But computed)

Problem:
Increased memory, computation.

The network learns from the **interpolated representation** — not directly from the **original observations**.

Can we learn directly from observations?

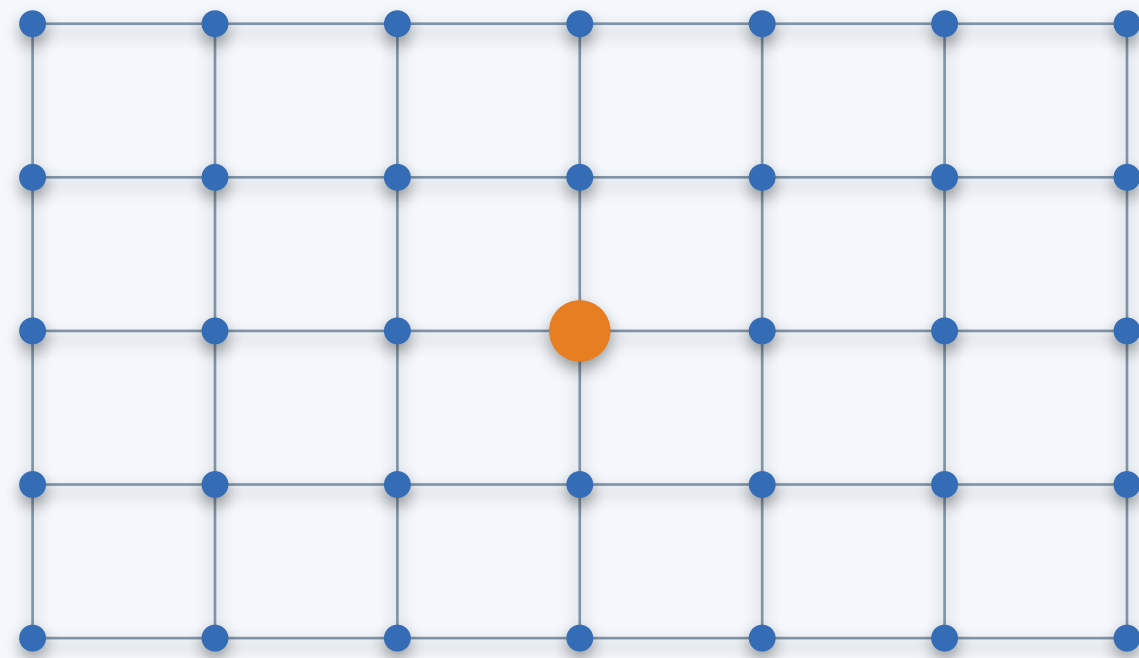
Part III — Can we learn directly from observations?

Key Idea 1: Observation = Coordinate + Value

A grid cell is no longer the basic unit. An observation is.

Grid-based

Values attached to predefined grid



$$x[i, j, k]$$

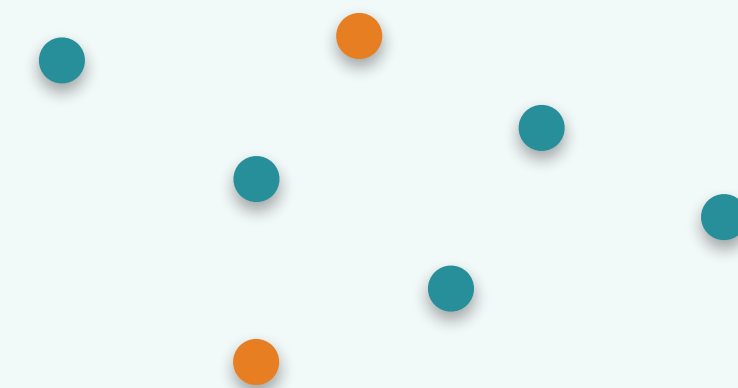
$(i, j, k) \rightarrow$ grid value

Presentation Shift

Grid $\rightarrow$ Coordinate

Grid-free

Observations represented in native coordinates



$$o_i = (c_i, v_i)$$

Where? What?

$$c_i = (\text{lat}_i, \text{lon}_i, z_i)$$

$$v_i = (T_i, Q_i, U_i, V_i, Z_i)$$

Represent observations by where they are — not by where they fall on a predefined grid.

Key Idea 2: From Observations to Latent Representation

Different sensors keep their native geometry.
The model learns a latent representation instead of first building a common observation grid.

$$\{o_i\}_{i=1}^N \rightarrow \mathbf{z}$$

Heterogenous Observations

Satellite

swath / pixels

Buoy

irregular points

Ship

trajectory

Radar

polar sampling

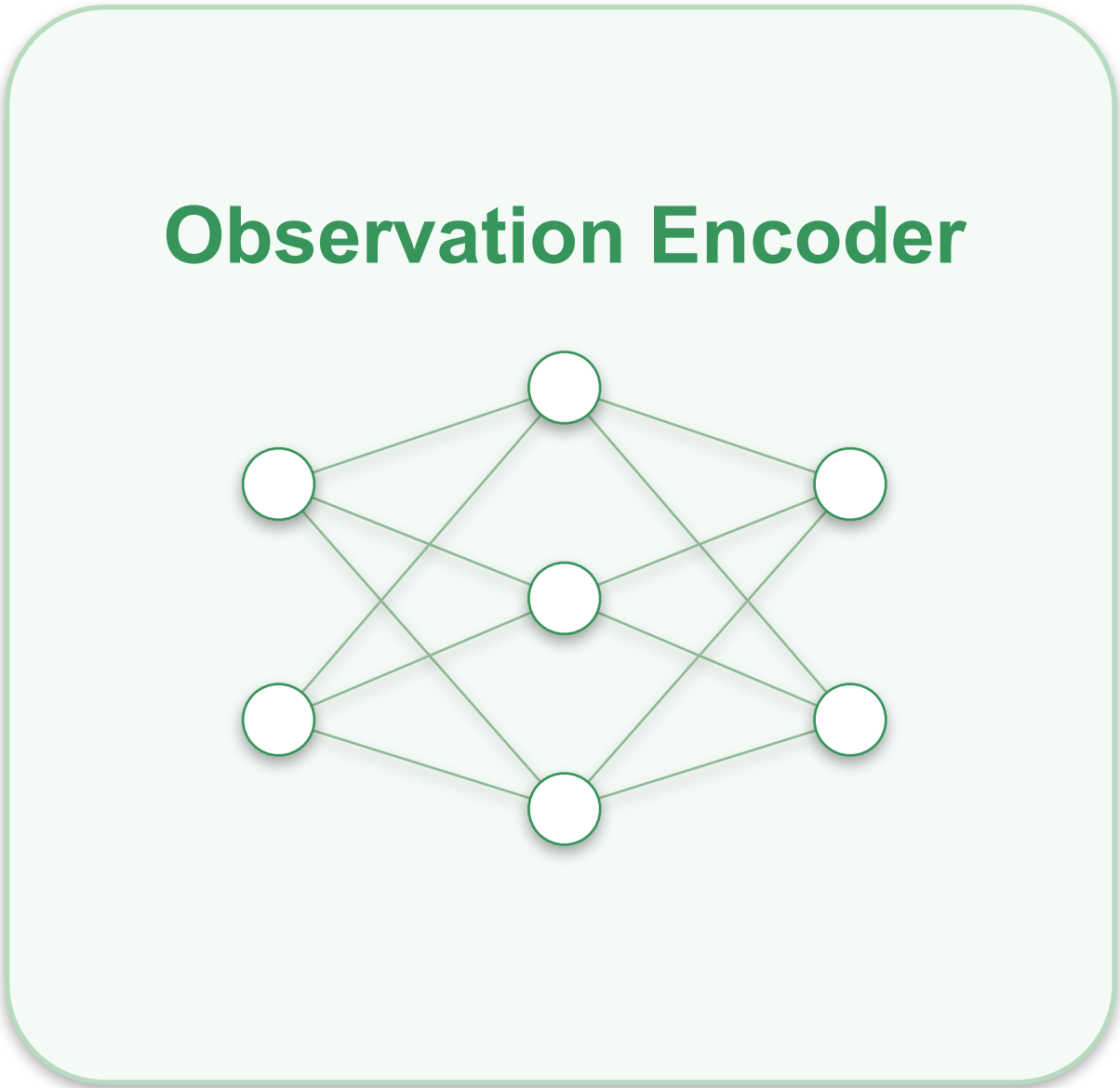
Station

point observation

$o_i = (c_i, v_i)$

coordinate value

Representation Learning



Latent Representation

Satellite

Buoy

Ship

Radar

Station

$\mathbf{z}$

Latent Representation

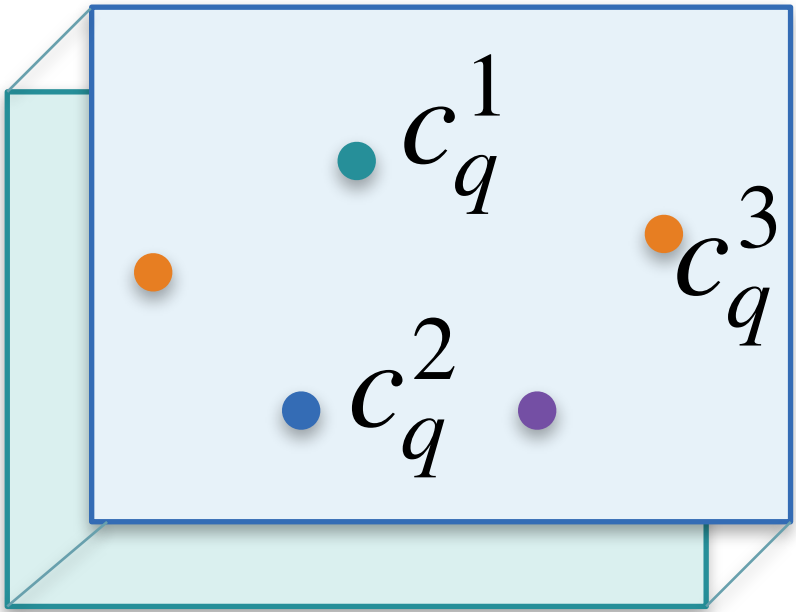
Key Idea 3: Latent State + Coordinate → State

A latent representation lets us ask for the ocean state at any coordinate.

$$\mathbf{x}(\mathbf{c}_q) = f_{\theta}(\mathbf{c}_q, \mathbf{z})$$

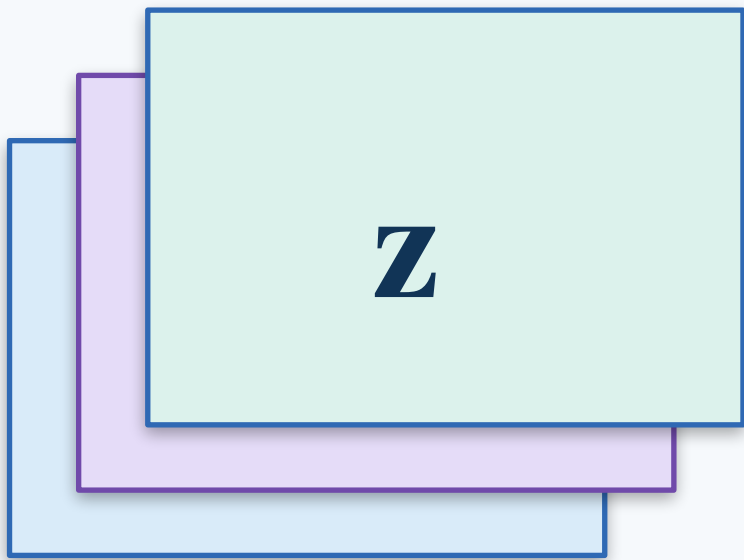
Query coordinate Latent state

Query Anywhere



$$\mathbf{c}_q = (\text{lat}, \text{lon}, z)$$

Latent Representation



Fused information from
observations and background

Query 1 $\mathbf{c}_q = (30^{\circ}N, 150^{\circ}E, 0\text{m})$

What is SST here?

State at that location

$$\text{SST} = \mathbf{x}(\mathbf{c}_q)$$

Query 2 $\mathbf{c}_q = (40^{\circ}N, 130^{\circ}E, 500\text{m})$

What is temperature at 500 m?

State at that location

$$T_{500\text{m}} = \mathbf{x}(\mathbf{c}_q)$$

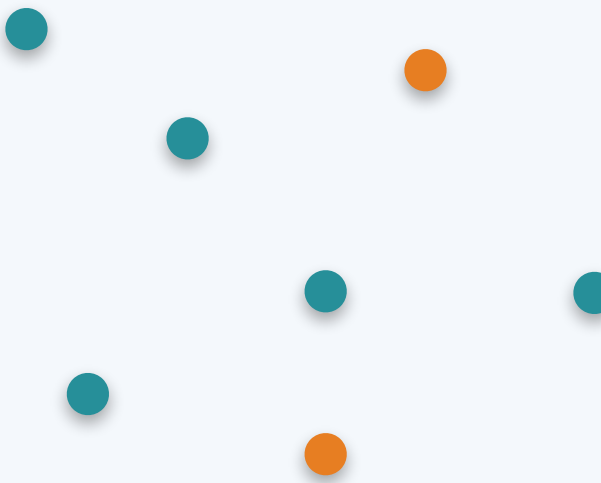
Query = provide a coordinate and ask the latent representation for the state there.

Grid-free Data Assimilation

Assimilate observations in their native coordinates — without first mapping them onto a common grid.

1 Native Observations

Coordinate–Value Pairs

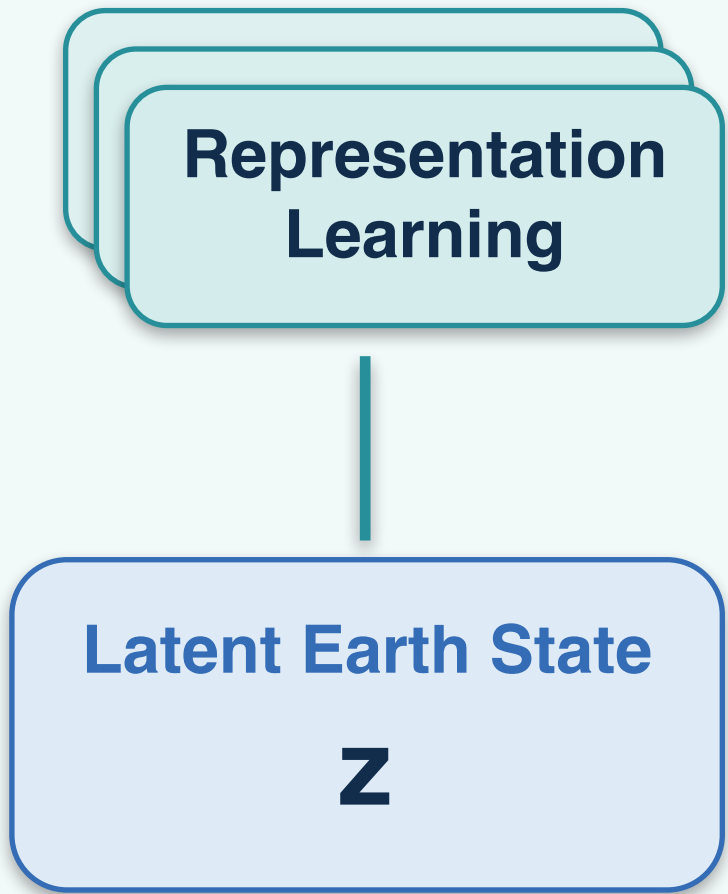


$(\text{lat}_1, \text{lon}_1, z_1, v_1)$
 $(\text{lat}_2, \text{lon}_2, z_2, v_2)$
 $(\text{lat}_3, \text{lon}_3, z_3, v_3)$
...

No regridding required

2 Latent Representation

Observation Encoder



Common latent space

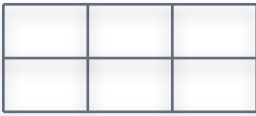
3 Query Coordinates

Query Anywhere

$$c_q = (\text{lat}, \text{lon}, z)$$

Query Point
arbitrary location

Query Grid

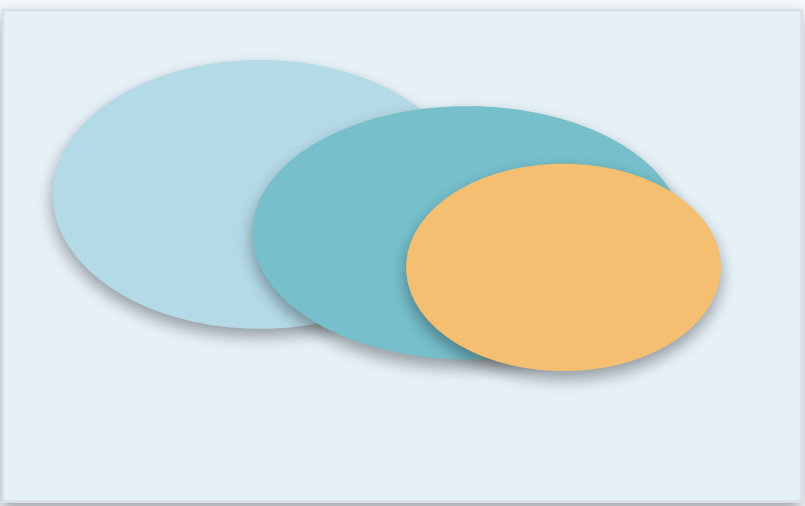


NWP / user-defined grid

4 Analysis

Continuous Analysis

$$x(c_q) = f_\theta(c_q, z)$$



T, Q, U, V, Z

At arbitrary coordinates or grids

Assimilate observations where they are; query the analysis where we need it.

Grid-based vs. Grid-free Data Assimilation

Two complementary representation paradigms for different observation geometries and assimilation tasks.

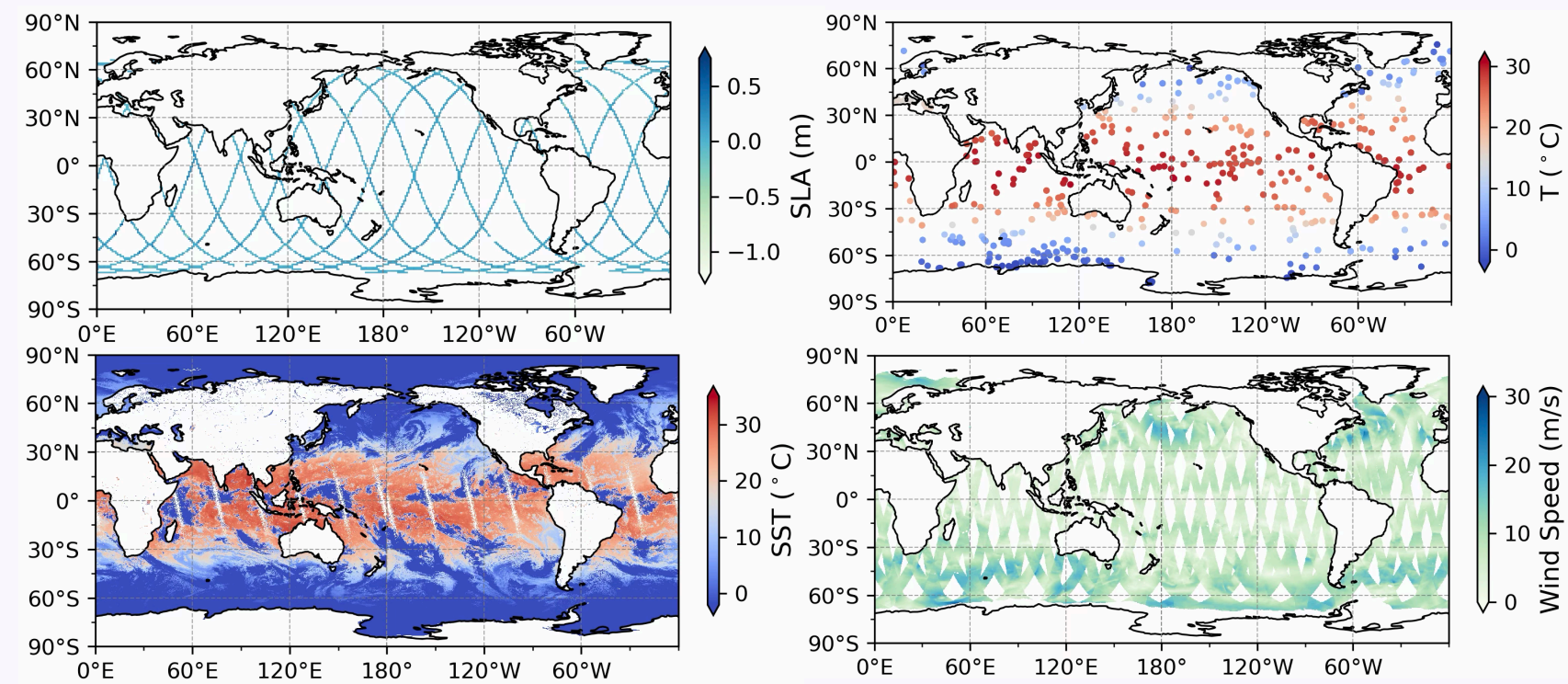
Comparison	GRID-BASED AI-DA	GRID-FREE AI-DA
Observation representation	Regular grid	Coordinate–value pairs
Irregular observations	Regridding needed	Directly represented
Spatial resolution	Usually fixed	Flexible / query-dependent
Query locations	Grid points	Arbitrary coordinates
Multi-resolution data	Requires alignment	Naturally supported
Typical AI models	CNN / ViT	INR / coordinate networks / set-based models
Computation	Highly optimized	Can be challenging at large scale

Part IV — Does Grid-Free DA Actually Work?

Can We Assimilate Irregular Ocean Observations Directly?

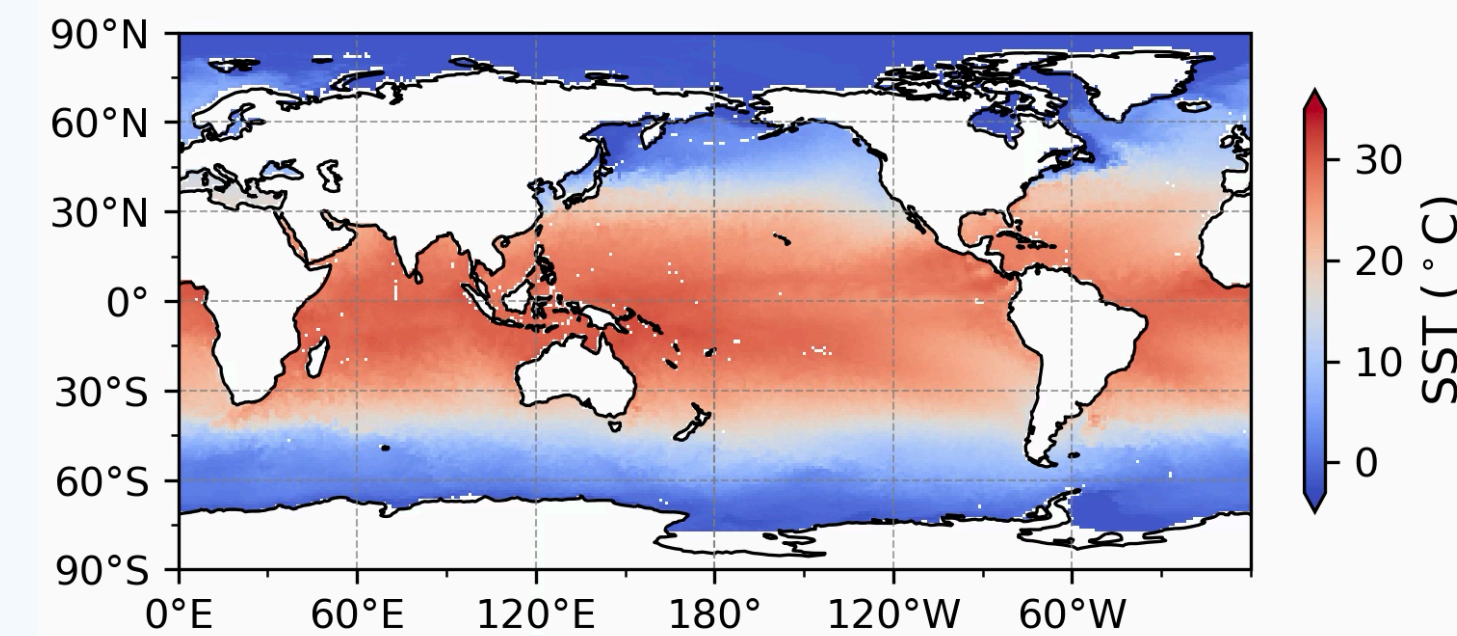
Sea-surface assimilation is a natural testbed for grid-free DA.

Irregular Observations



- Satellite observations → dense but incomplete
- Drifters / buoys → sparse points
- Ship observations → trajectories
- Missing regions → clouds / orbital sampling

Ocean State



T

S

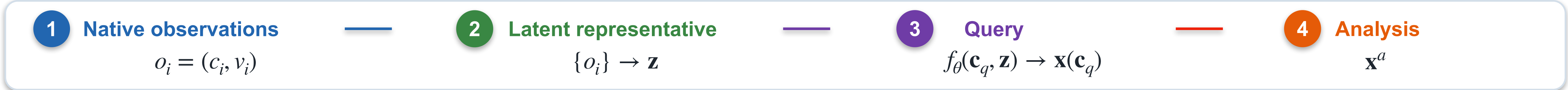
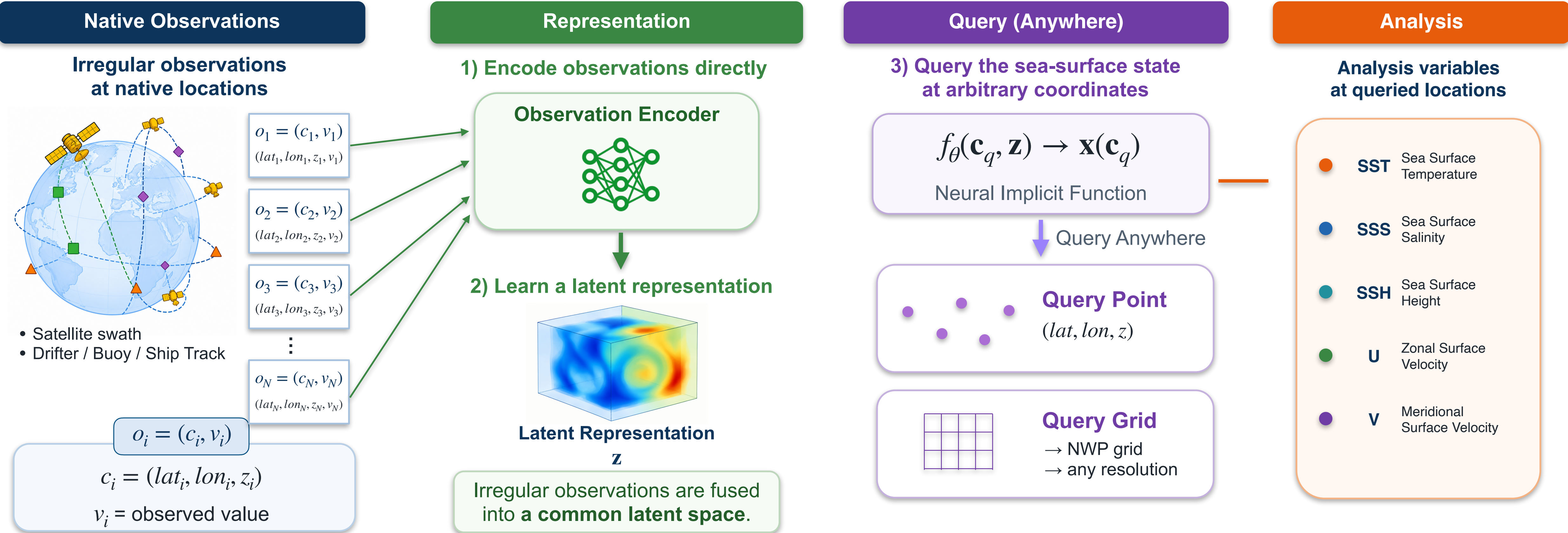
SSH

U

V

Grid-free Sea Surface Data Assimilation

Assimilate irregular sea-surface observations in their native coordinates — without first mapping them onto a common grid.

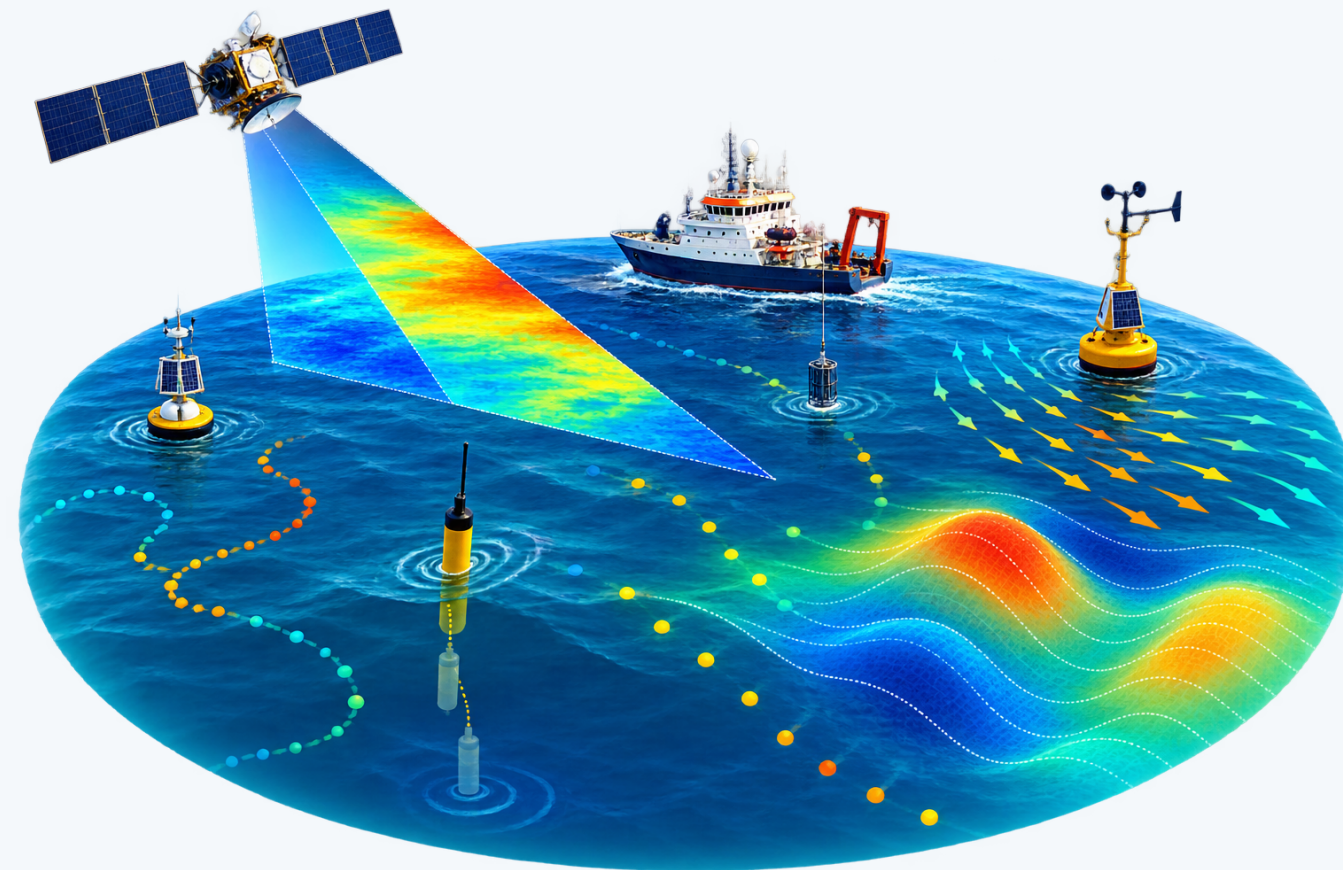


Observations are assimilated where they are; the sea-surface analysis is queried where we need it.

Experimental Design

Learn an analysis state from an AI-ocean background and heterogeneous observations, supervised by GLORYS.

1 Inputs



Background x^b
AI-based ocean model / forecast

Observations O
Satellite: SST · SSW · SSS · SLA
In-situ observations: ship · buoy · float

2 Supervised Learning

What does the model learn?

$(x^b, O) \rightarrow x^a$
background + observations $\rightarrow$ analysis

Supervision Target
GLORYS reference ocean state
 x^*

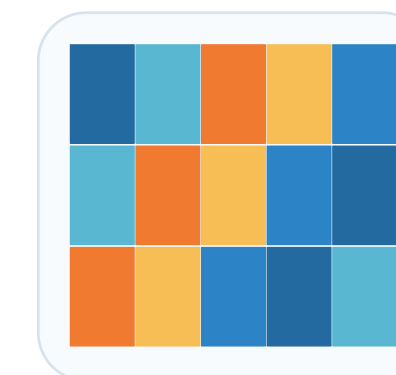
The network learns to correct the background toward the reference using available observations.

3 Evaluation

Does assimilation improve the ocean state?



Background
 x^b



Analysis
 x^a



Reference
 x^*

Evaluation Goal
Analysis should be closer to GLORYS than the background.

Evaluate: Background vs. Analysis vs. Reference
Metrics: RMSE, ACC

AI forecast x^b +

Observations O $\rightarrow$

Analysis x^a $\rightarrow$

GLORYS comparison

Q1 Can we conduct the analysis using a new, grid-free method?

Across all five sea-surface variables, assimilation consistently reduces error.

5 / 5

variables improved

RMSE ↓

5 / 5

variables improved

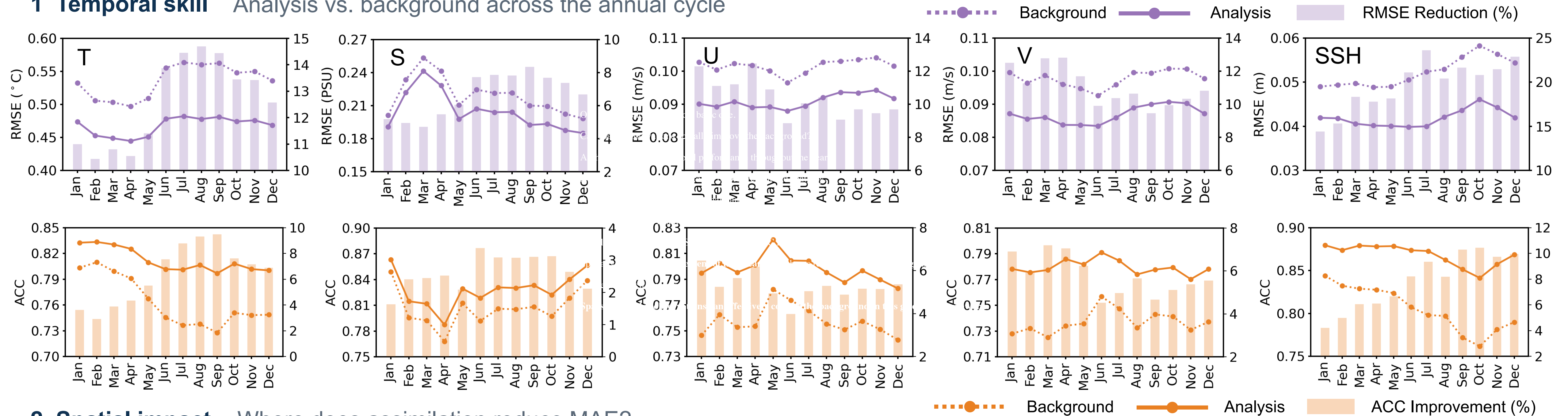
ACC ↑

67–95%

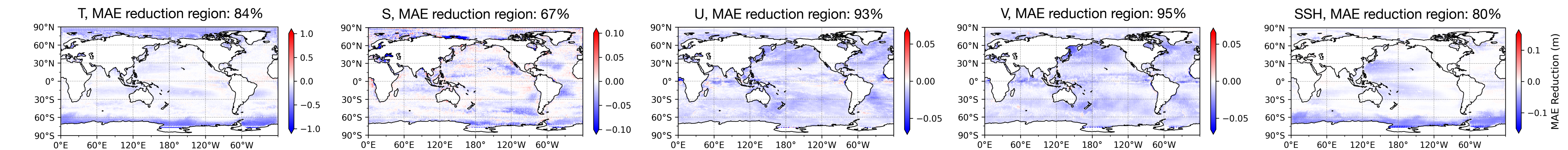
ocean area shows

MAE reduction

1 Temporal skill Analysis vs. background across the annual cycle



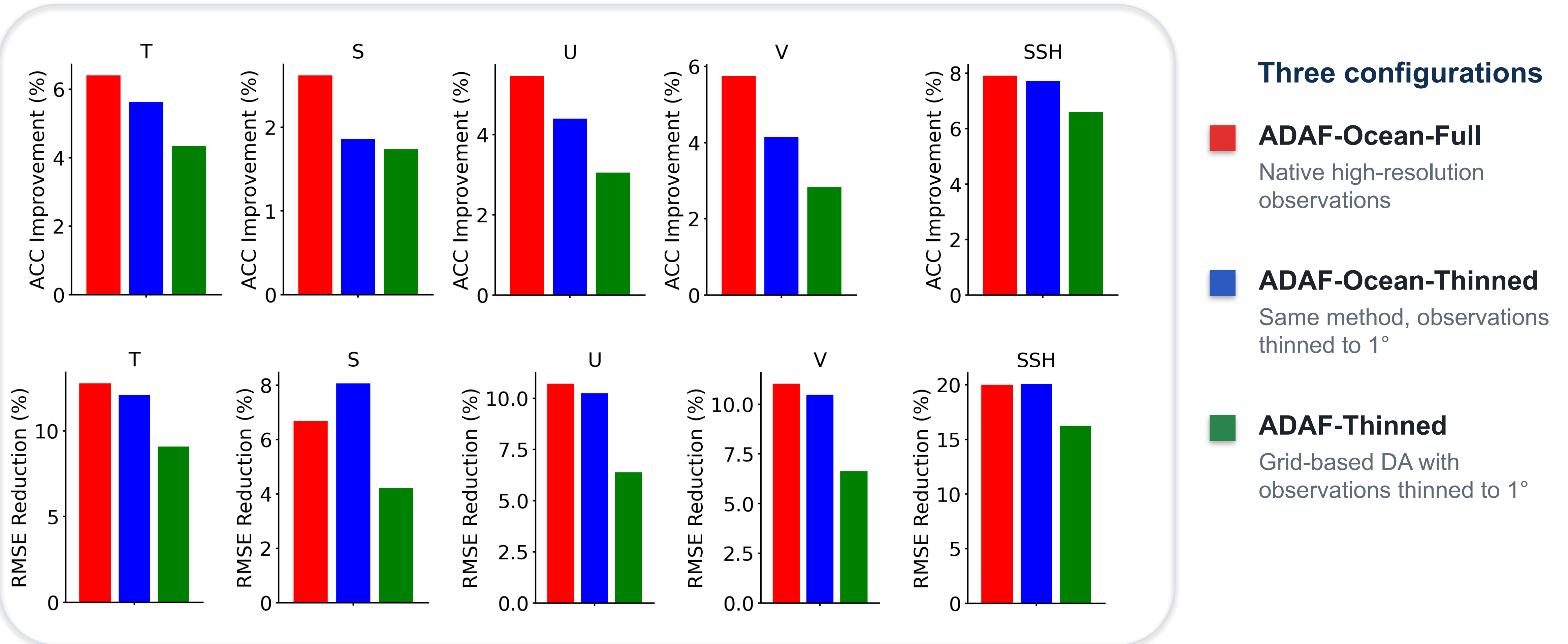
2 Spatial impact Where does assimilation reduce MAE?



Analysis is generally closer to the reference (GLORYS) than the background.

Q2 Does Native-resolution Observations Matter?

Do we benefit from fine-scale observations?



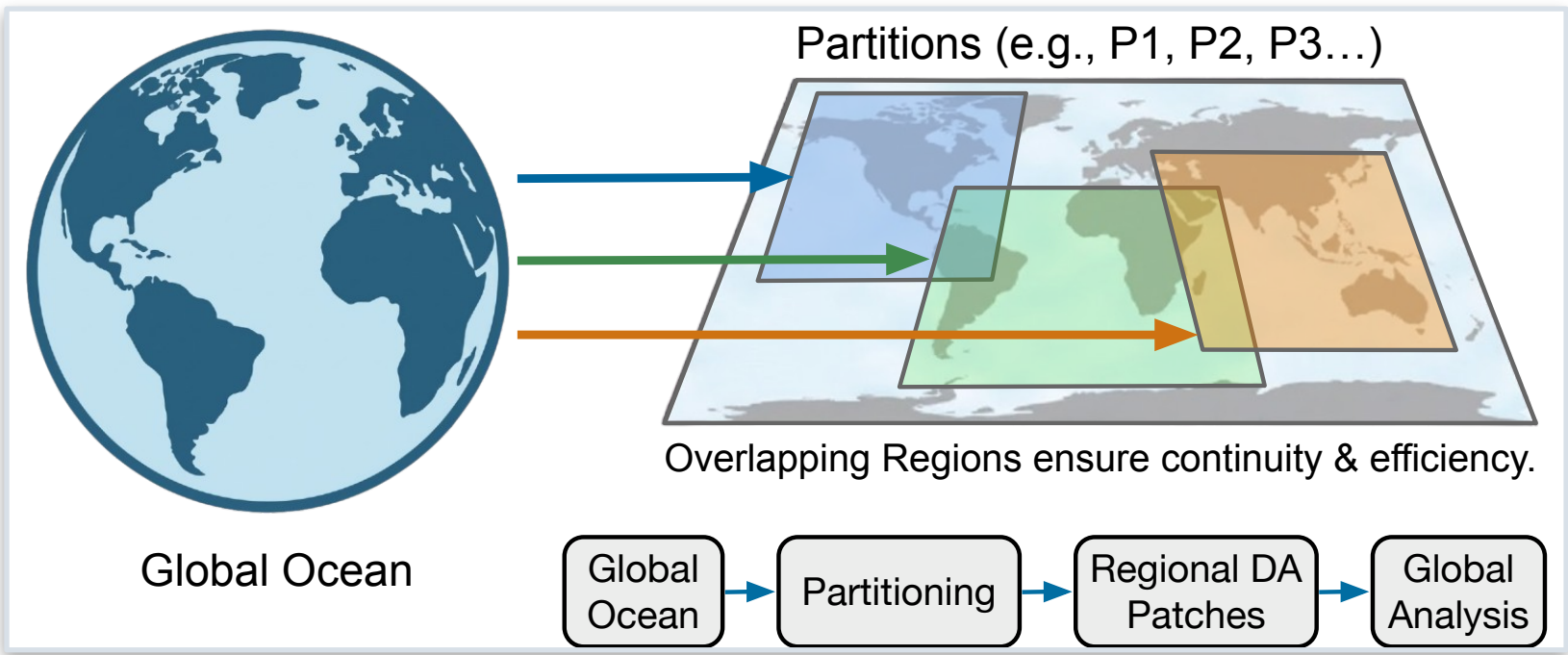
High-resolution observations contain more information, which can be used to improve analysis.

Q3 Can Grid-free DA Scale to High-Resolution Reconstruction?

ADAF-Ocean reconstructs 0.25° sea-surface fields from a 1° analysis while preserving global scalability.

Global domain decomposition

Split the globe into overlapping sub-domains, assimilate them in parallel, and blend the local analyses.



Scaling from 1° → 0.25°

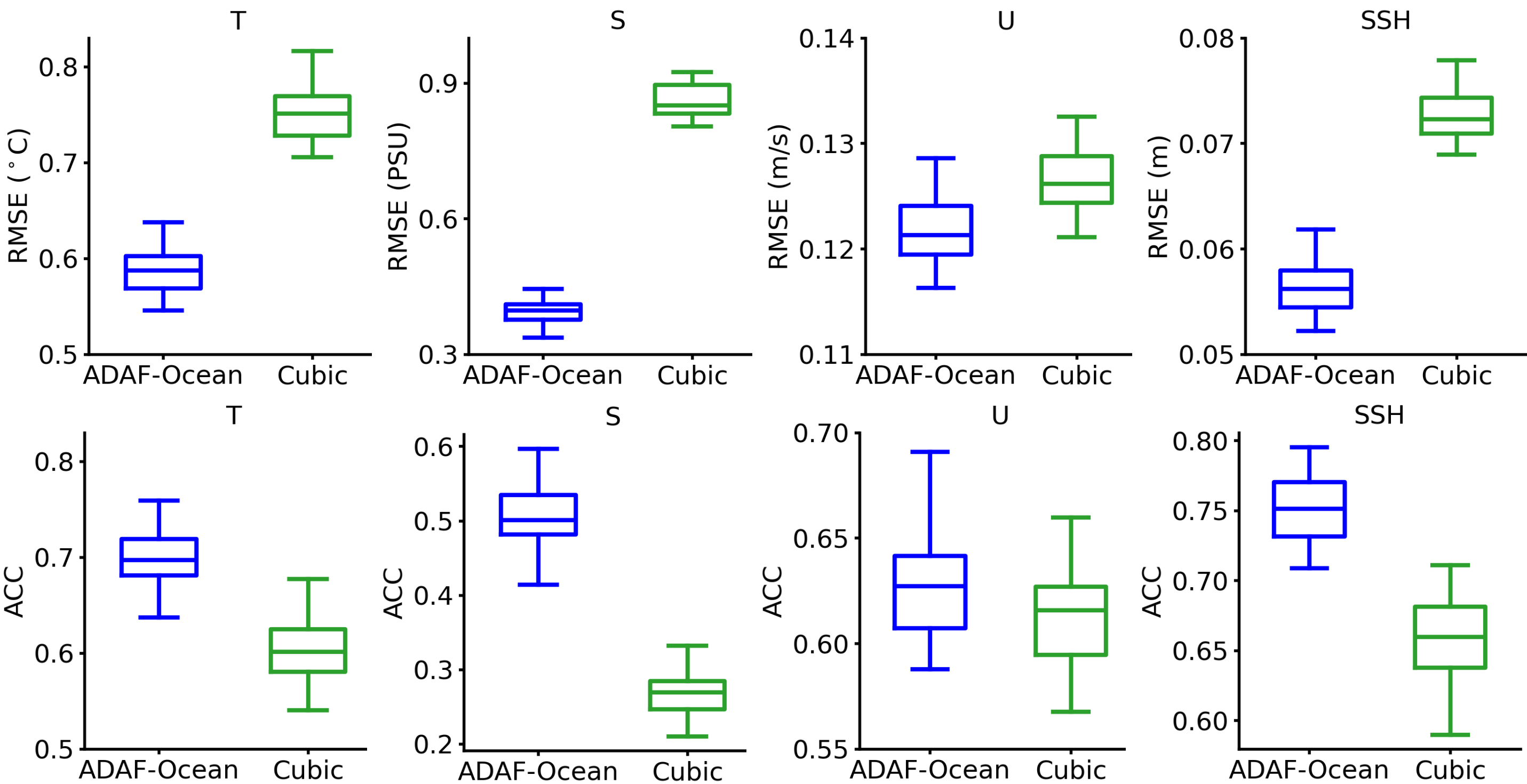
Grid size 180×360 → 720×1440 16×

Runtime 12 s → 120 s 10×

Compute grows more slowly than grid size.

Downscaling skill: 1° → 0.25°

Compared with cubic interpolation, ADAF-Ocean consistently improves reconstruction skill across variables.



Independent SST validation

10.4%

lower RMSE vs. baseline

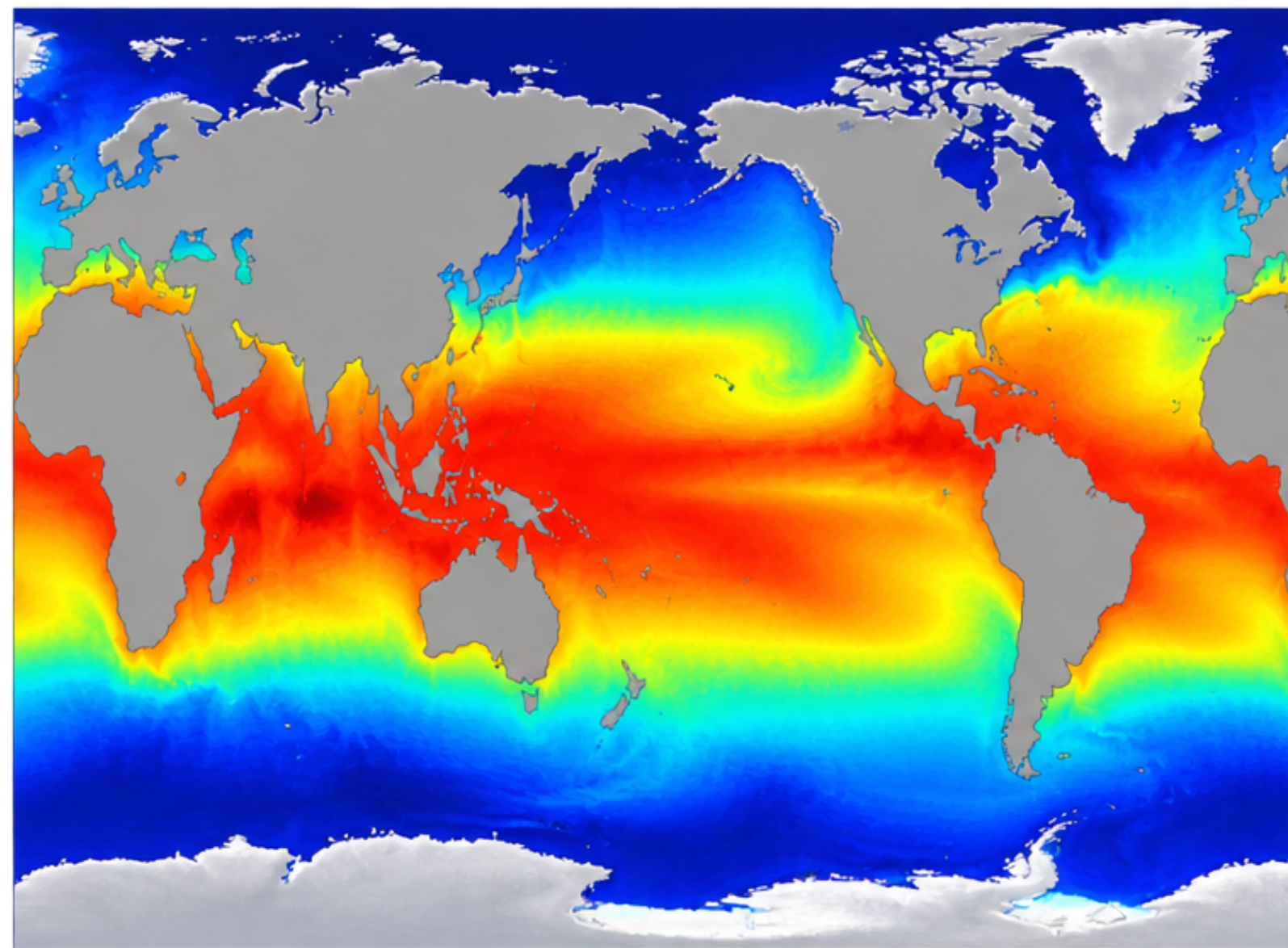
- Lower RMSE and higher ACC than cubic interpolation.
- Independent OISST validation confirms the improvement.
- Accuracy approaches the reference reanalysis at high resolution.

From Surface Reconstruction to 3D Ocean State

Moving from horizontal reconstruction to volumetric state estimation introduces a new challenge: physics.

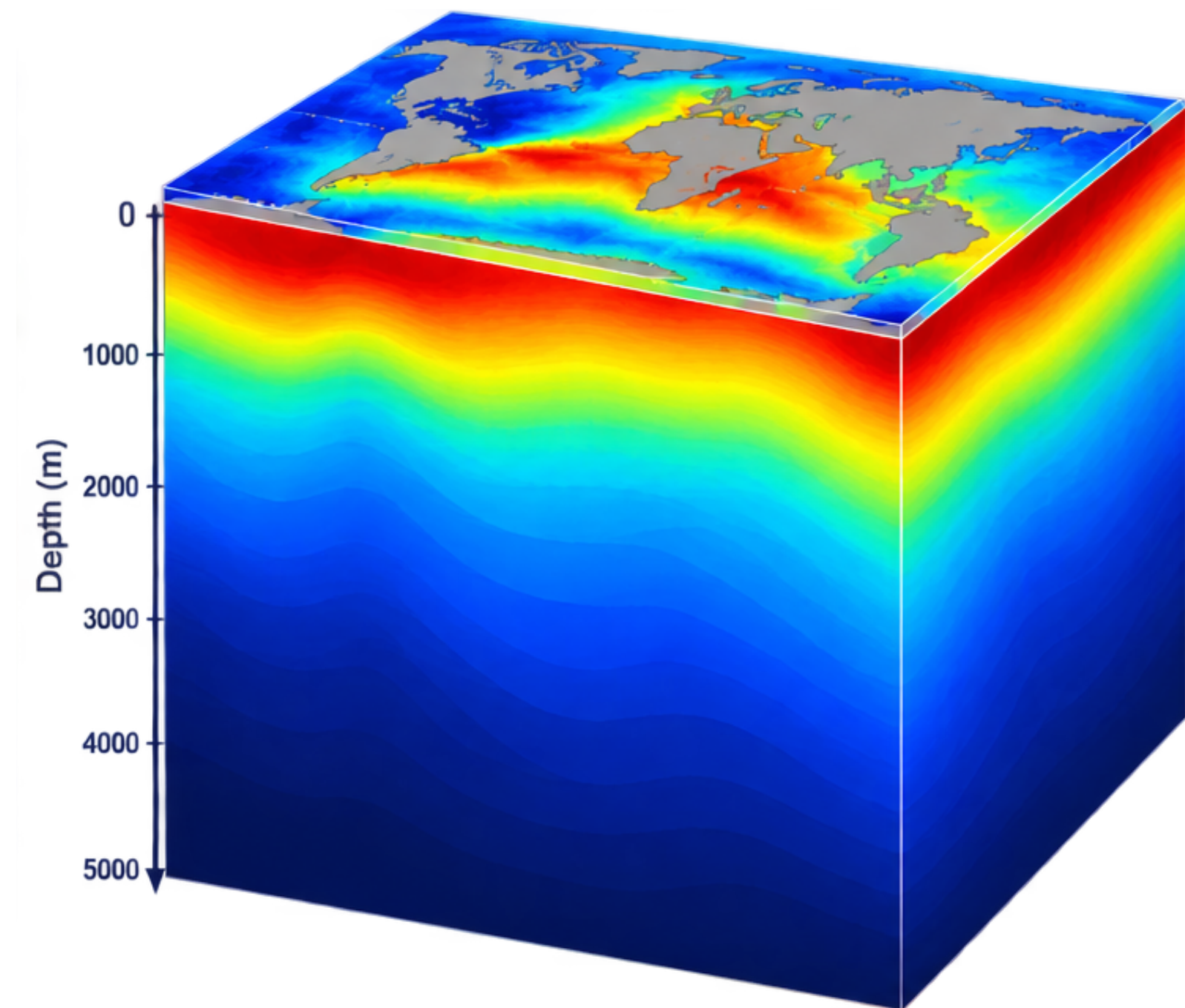
Sea Surface

2D reconstruction



3D Ocean

3D reconstruction with vertical structure



How do we recover the interior without violating ocean physics?

Q4 — Is an Accurate Reconstruction Physically Consistent?

Low reconstruction error does not necessarily guarantee a physically plausible ocean state.

Prediction Accuracy

RMSE / MAE / ACC

≠

Physical Consistency

T–S relation / density / stability

Can an accurate T–S field still be unstable?

From reconstructed variables to physical plausibility

Temperature + Salinity

Reconstructed state



Density $\rho(T, S, p)$

Derived physical state



Vertical Density Gradient

$\partial\rho/\partial z$



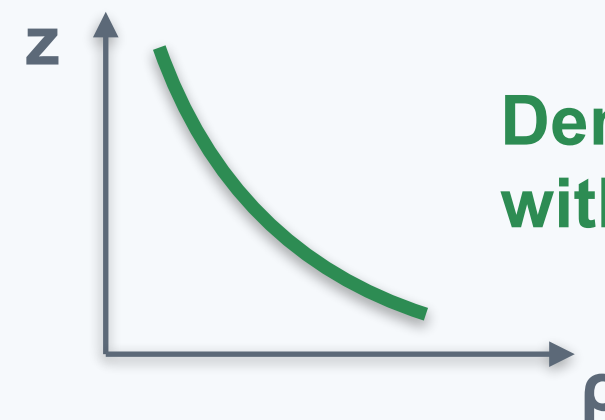
Vertical Stability

Stable or unstable?

What should we evaluate?

1. Accuracy of T and S
2. Consistency of derived density
3. Vertical static stability

A physically plausible profile



Density increases
with depth → stable

Accuracy is necessary,
but physics must be
checked explicitly.

Physics-Consistent Learning for 3D Ocean Reconstruction

Training objective

$$L_{total} = \lambda_{WMSE} L_{WMSE} + \boxed{\lambda_{den} L_{den}} + \lambda_{sta} L_{sta}$$

① Density consistency

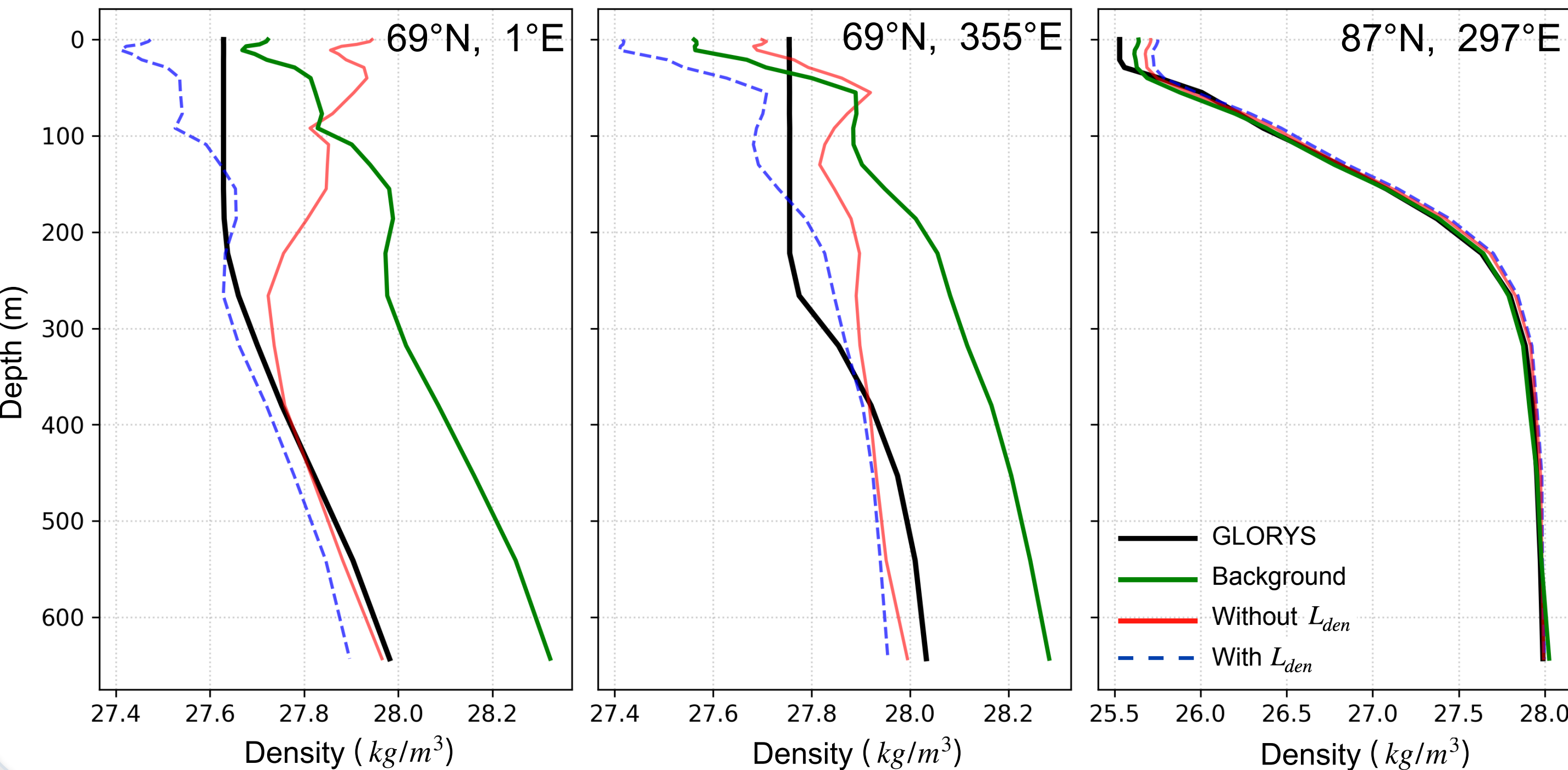
$$L_{sta} = \frac{1}{\sum_{j=1}^N M_j} \sum_{j=1}^N M_j \max \left(0, -\Delta\rho_{j,z} \right)$$

Physical meaning

Reconstructed temperature and salinity should imply a physically consistent density field.

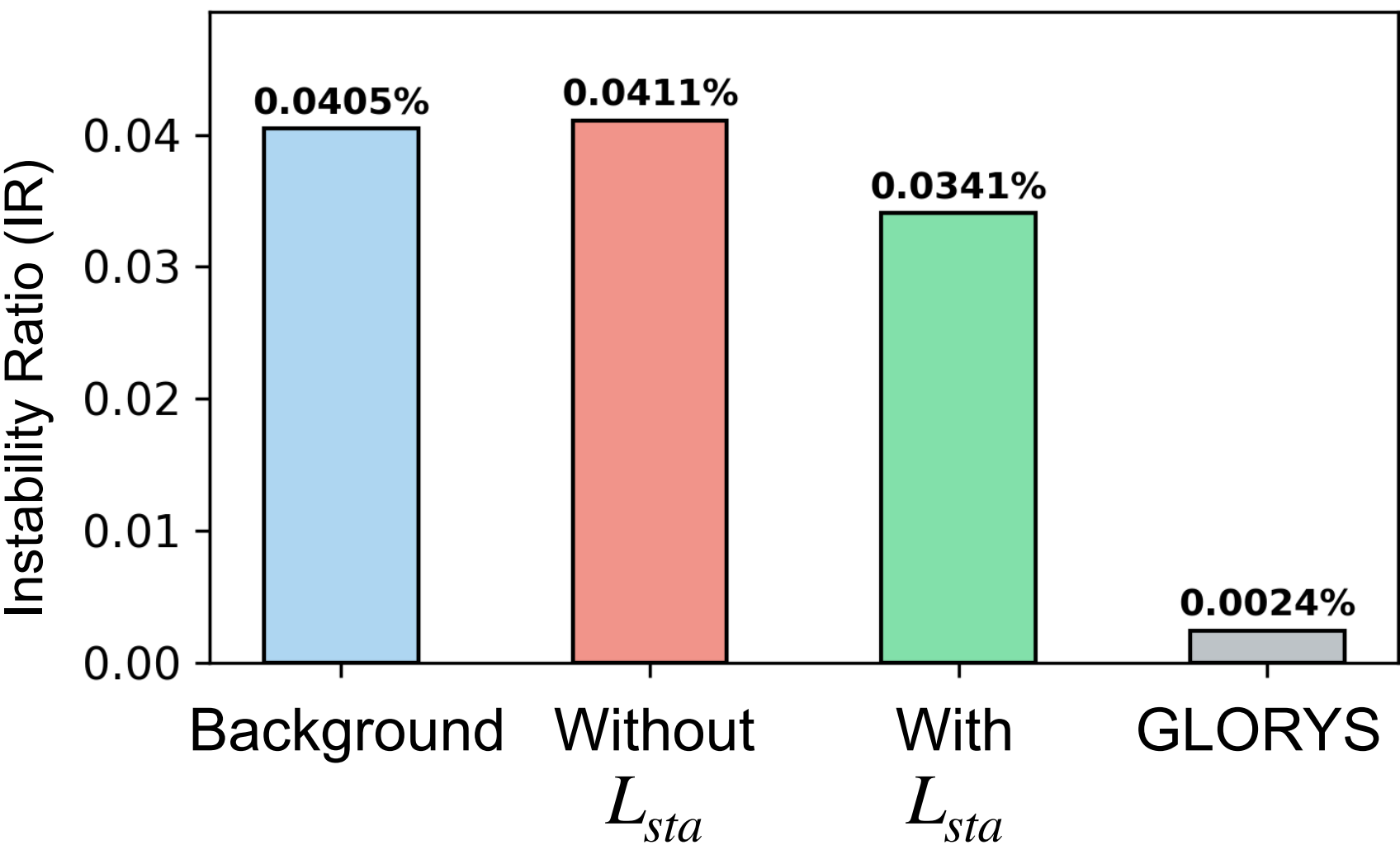
Does the constraint improve physics?

Density profiles at representative locations show how the constraint reshapes the vertical water-column structure.



Instability ratio (IR): fraction of vertically unstable ocean points — lower is better.

$$IR = \frac{N(\Delta\rho_k < -\epsilon)}{N_{total}} \times 100 \%$$



IR ↓ 17%

Regional T–S and Density Biases

Can density-consistent learning improve regional water-mass structure?

What does the constraint enforce?

Reconstructed temperature and salinity should imply a density field consistent with the reference ocean state.

Density consistency

$$L_{den} \propto [\rho(\hat{T}, \hat{S}) - \rho(T, S)]^2$$

Regional diagnostics

$dist_{TS} \downarrow$ closer T-S relationship

$|\Delta\rho| \downarrow$ smaller density bias

T, S $\rightarrow$ ρ $\rightarrow$ physics

Evaluate water-mass consistency

Ocean regime	Metric	Background	Without L_{den}	With L_{den}
Black Sea Extension	$dist_{TS}$	0.520	0.514	0.510
	$\Delta\rho$	0.037	0.003	-0.001
North Atlantic Mode Water	$dist_{TS}$	0.400	0.268	0.266
	$\Delta\rho$	0.102	0.009	-0.010
Antarctic Intermediate Water	$dist_{TS}$	0.293	0.263	0.263
	$\Delta\rho$	-0.004	0.001	0.001
Western Pacific Warm Pool	$dist_{TS}$	0.353	0.365	0.361
	$\Delta\rho$	0.008	-0.045	-0.034

$dist_{TS}$: deviation from the reference T–S relationship | $\Delta\rho$: density bias

L_{den} give us smaller T–S bias in all four regimes.

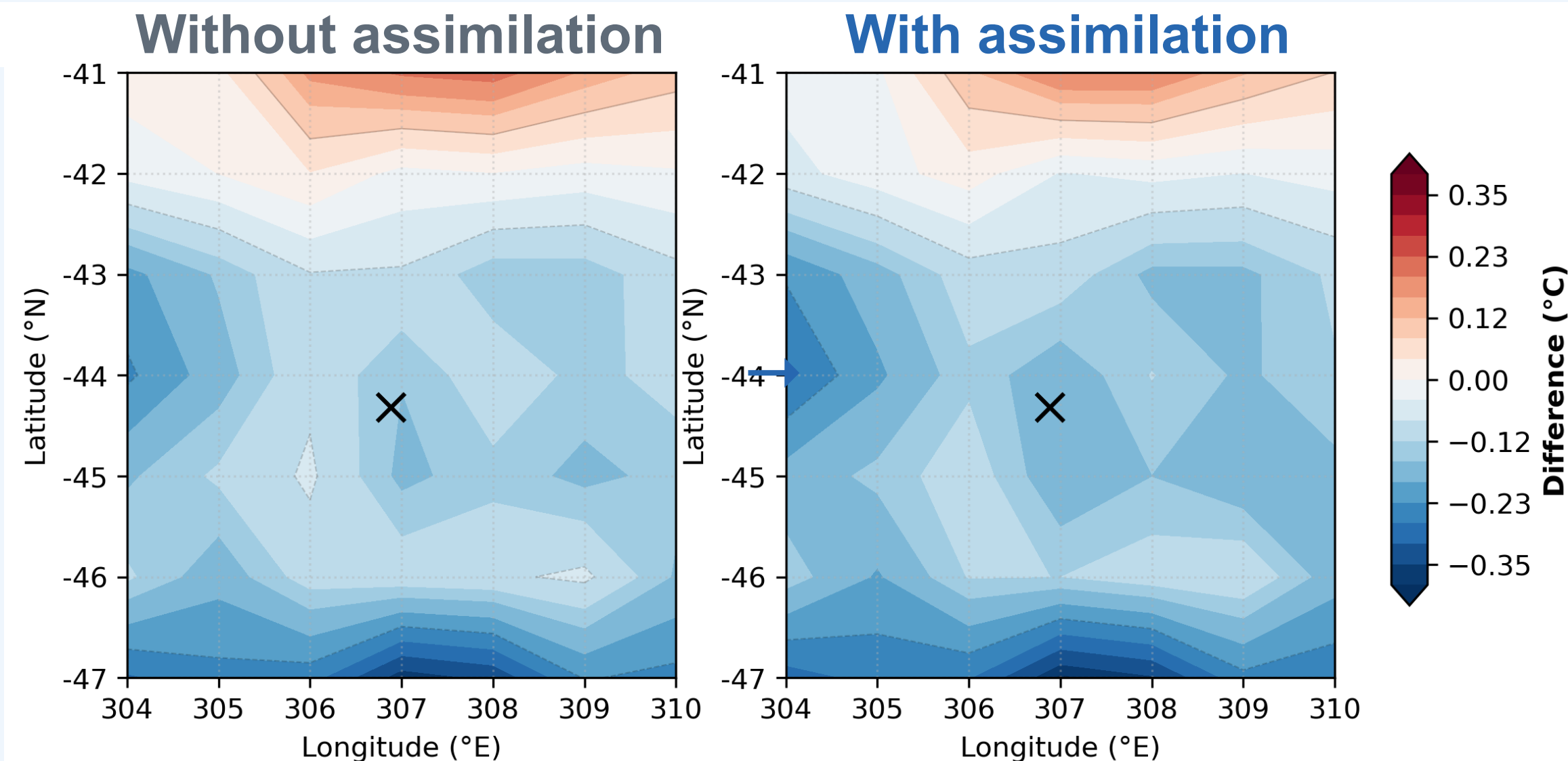
How Does a Single Observation Influence the Analysis?

A point observation produces a spatially coherent analysis increment, with the sign determined by the observation–background mismatch.

Case 1 Colder Observation

Observation 10.7°C < Background 14.3°C

Analysis correction: -3.44°C

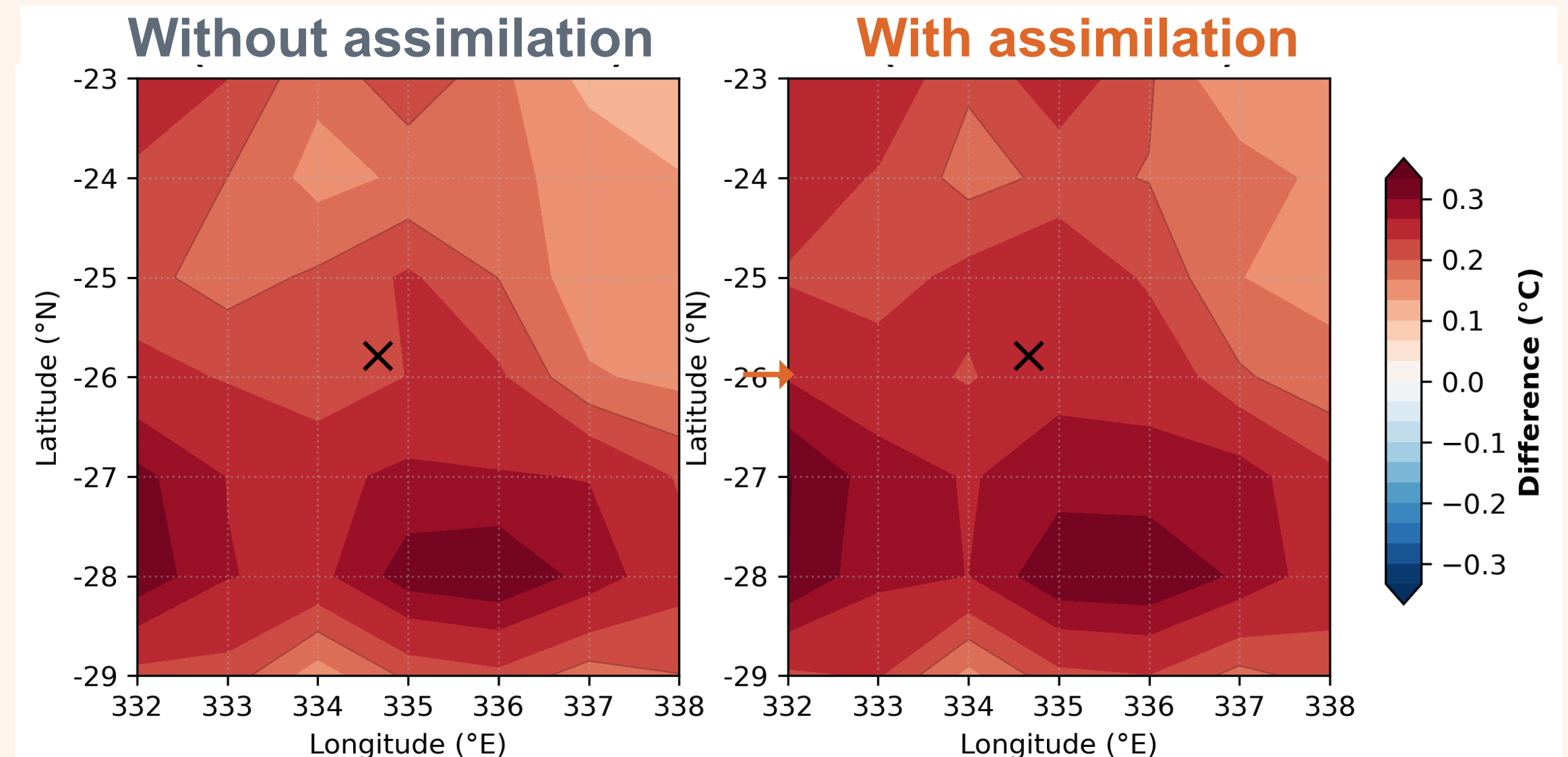


× observation location

Case 2 Warmer Observation

Observation 25.4°C > Background 24.7°C

Analysis correction: $+0.45^{\circ}\text{C}$



× observation location

The learned response is local but spatially structured.

The model propagates observational information beyond the observation point.

Papers related to this work:

- **Xiang Y**, Jin W, Dong H, et al. ADAF: An artificial intelligence data assimilation framework for weather forecasting[J]. Journal of Advances in Modeling Earth Systems, 2025, 17(9): e2024MS004839.
- **Xiang Y**, Gao Y, Wu H, et al. Advancing Ocean State Estimation with efficient and scalable AI[J]. arXiv preprint arXiv:2511.06041, 2025.
- Xiong W, **Xiang Y**, Wu H, et al. Ai-goms: Large AI-driven global ocean modeling system[J]. arXiv preprint arXiv:2308.03152, 2023.

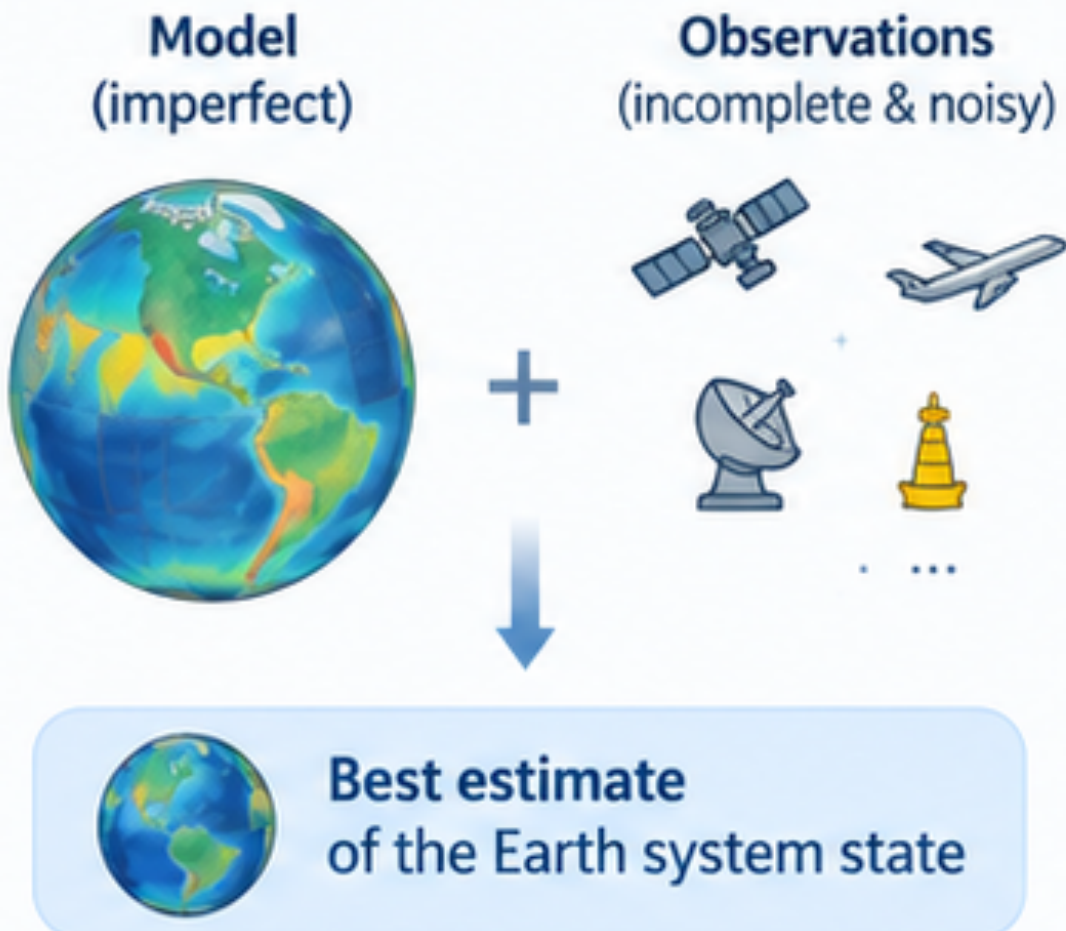
Take-Home Messages

From Data Assimilation to AI-enabled State Estimation

*Same goal
New tools
Greater potential*

1 DA is fundamentally a state estimation problem

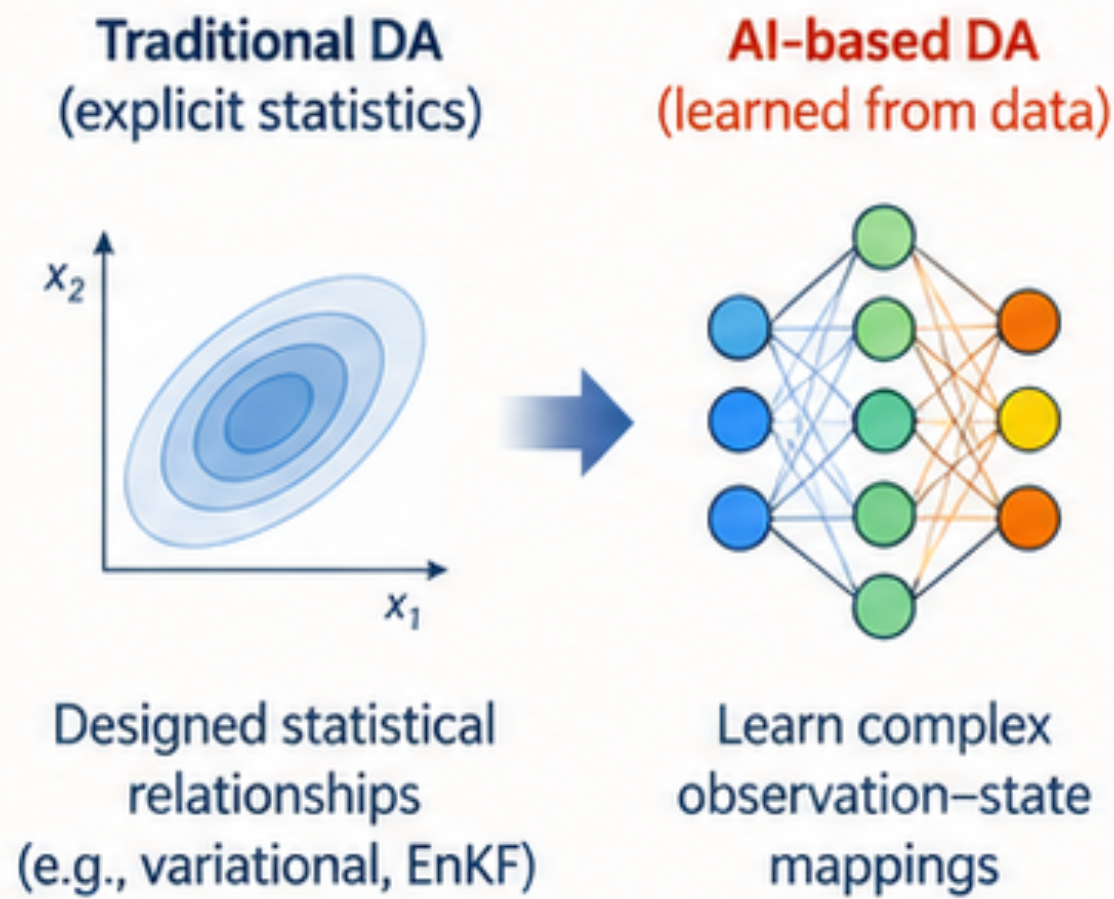
Model + Observations → Best estimate of the Earth system state



Observations are incomplete and noisy, while models are imperfect. DA combines their complementary information.

2 AI changes how we perform data assimilation

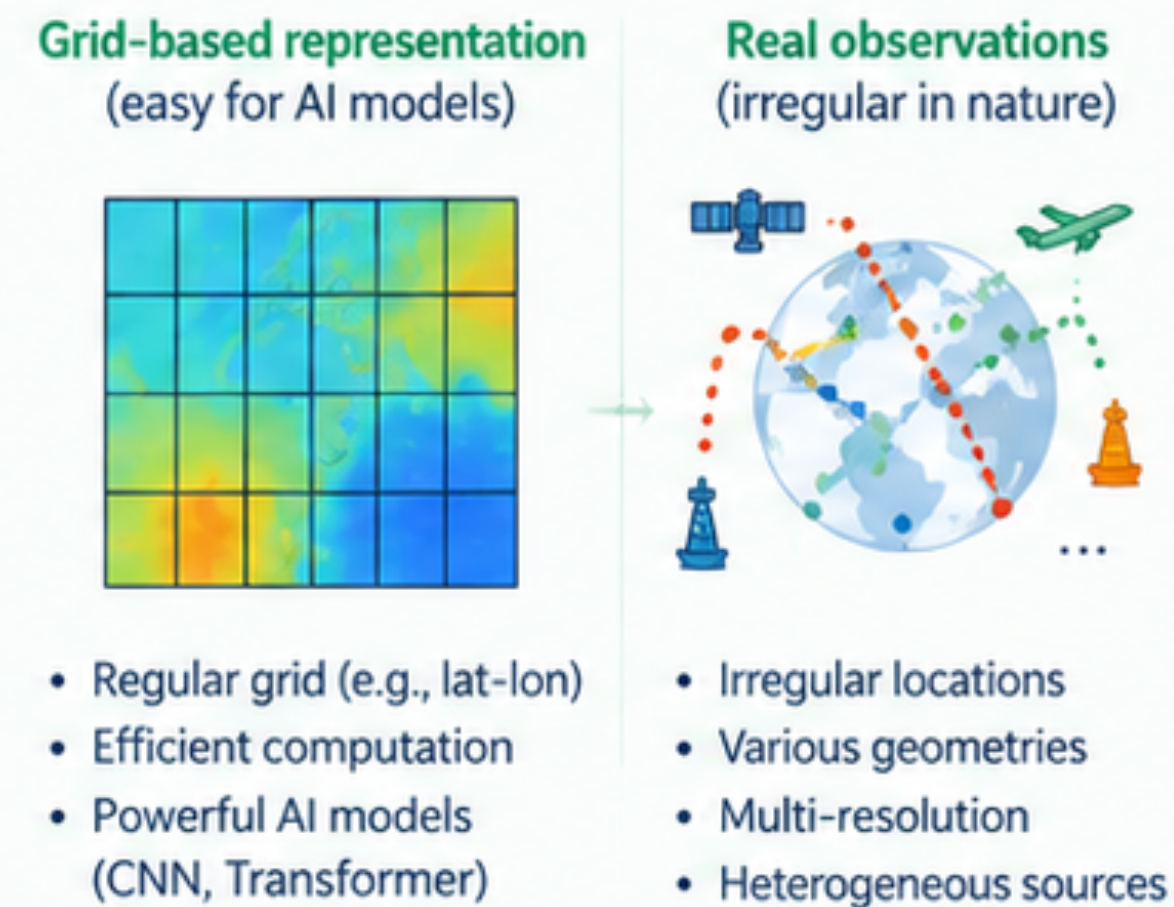
From explicit statistical updates → learned observation-state relationships



AI can learn complex mappings, fuse multi-source observations, and provide fast analysis updates.

3 Representation matters

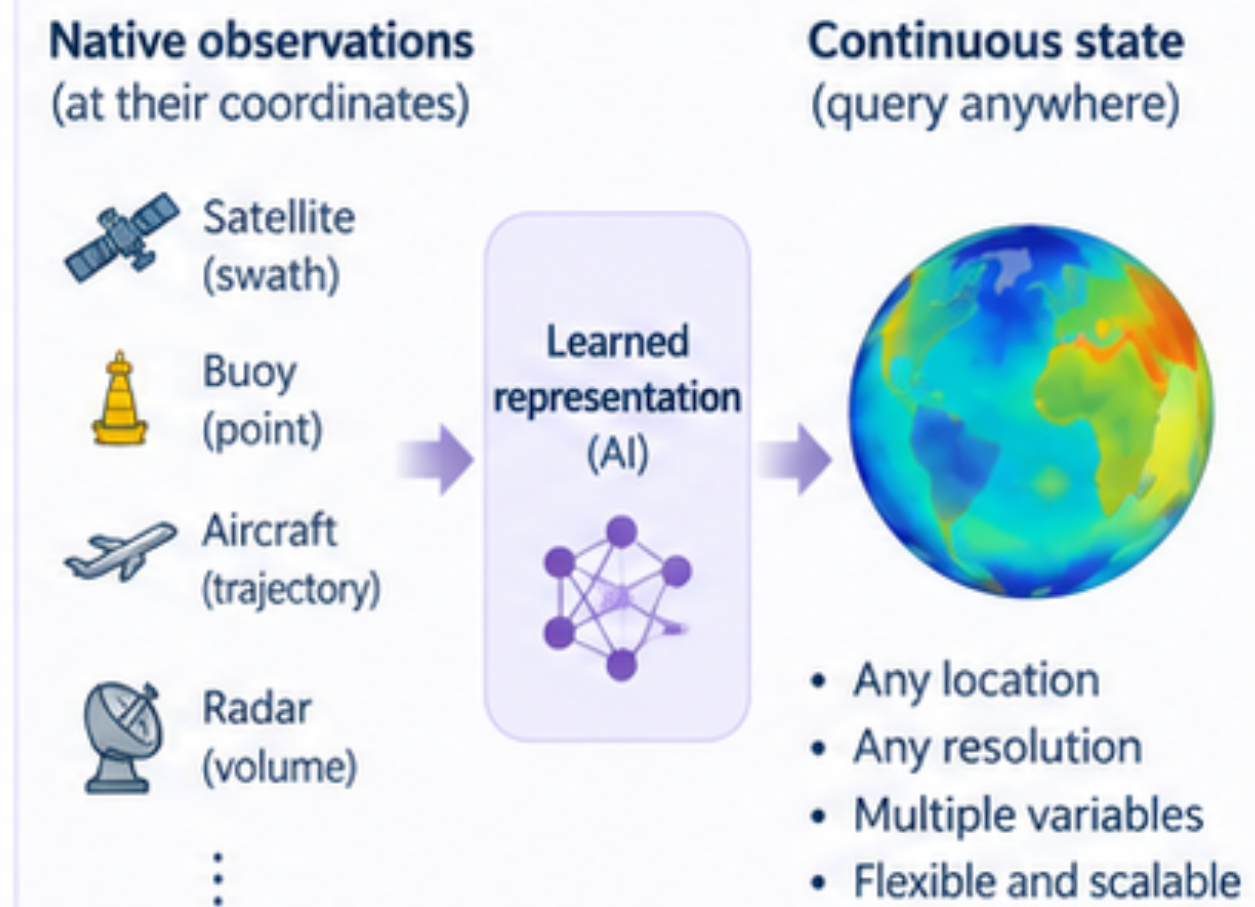
Grid-based AI-DA is powerful — but observations are naturally irregular



Regridding makes observations easy for neural networks to process, but may alter their native spatial information.

4 The next step is toward flexible, continuous representations

Native observations → Learned representation → Continuous state



Grid-free approaches provide a new way to assimilate observations directly at their native coordinates and query the state anywhere.



The goal remains the same: **use observations and models to better estimate the Earth system state.** What is changing is **how we represent information and learn the update.**