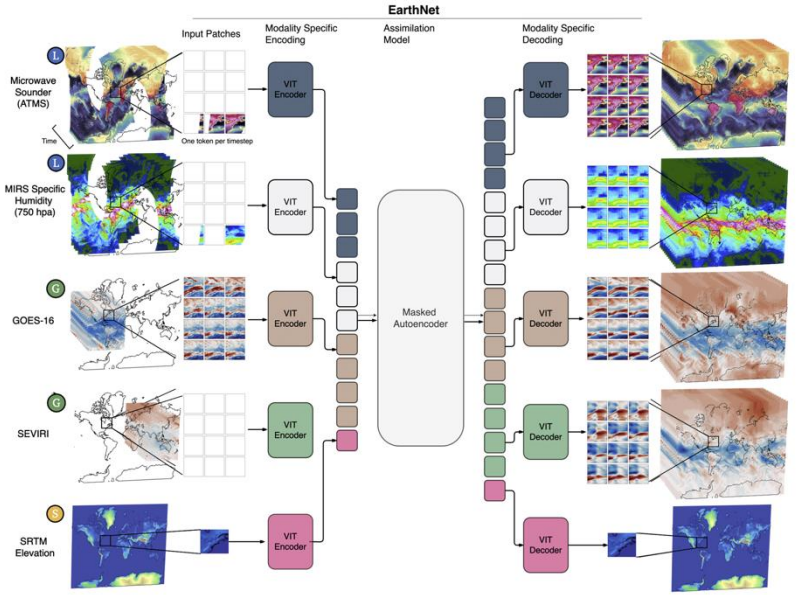
Atmos ML Journal Club

August Posch 13 Feb 2026

zeus·AI

Global atmospheric data assimilation with multi-modal masked autoencoders

Thomas J. Vandal[1], Kate Duffy[1], Daniel McDuff[1], Yoni Nachmany[1] and Chris Hartshorn[1]
[1]Zeus AI



Global atmospheric data assimilation with multi-modal masked autoencoders

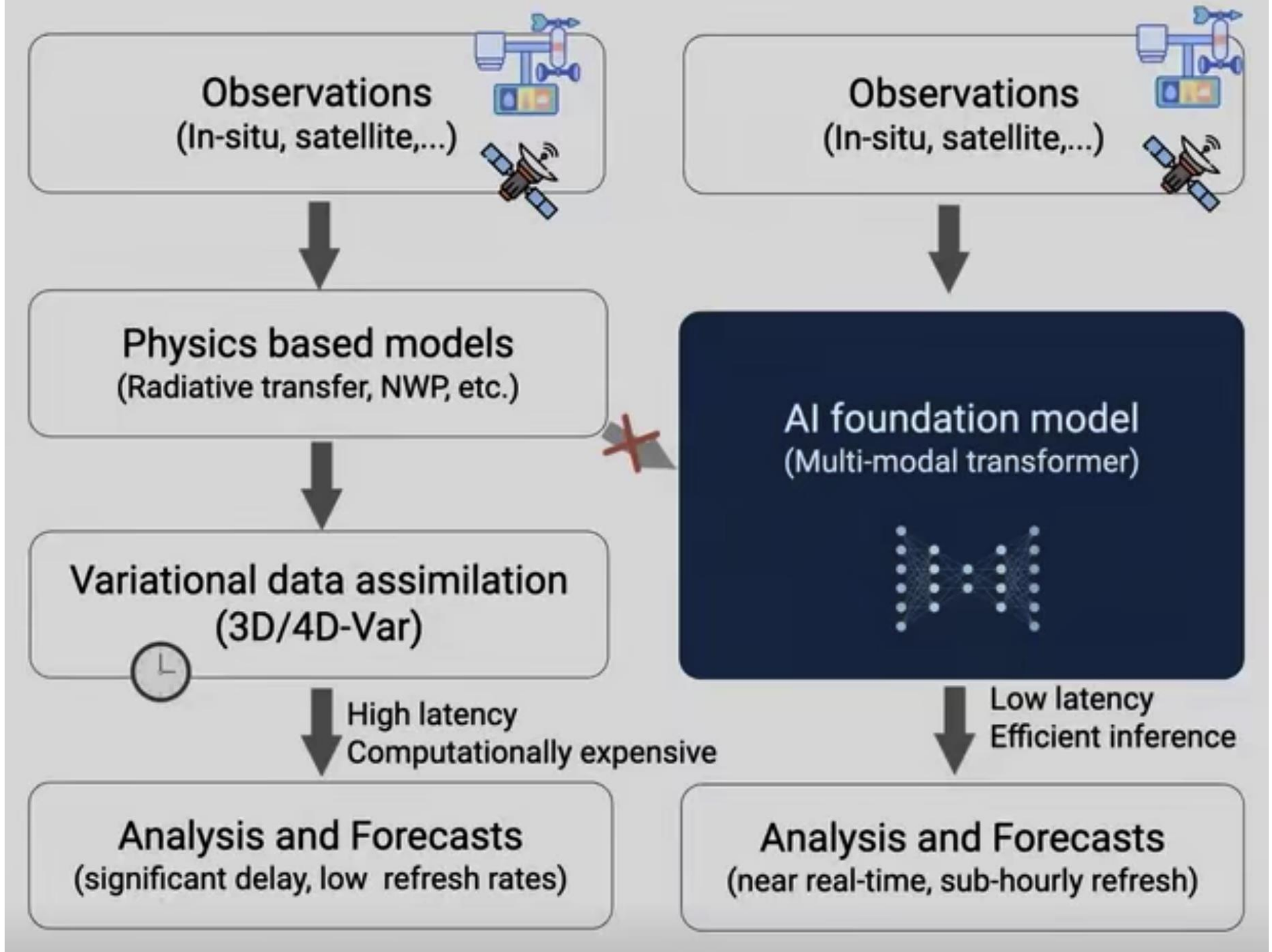
Thomas J. Vandal[1], **Kate Duffy**[1], **Daniel McDuff**[1], **Yoni Nachmany**[1] **and Chris Hartshorn**[1]

[1]Zeus AI

Transparency statement: A month ago I started asking Kate and TJ questions about this paper and that turned into a job conversation.

Introduction

- There are *so many* observations (e.g., from satellites) these days and the NWPs can't assimilate them
- NWPs do not incorporate the *most recent* observations
- Operational AI forecasts depend on DA
- This paper presents a pure-AI approach to ingest observations and produce profiles of temperature and humidity.


Observations
(In-situ, satellite,...)
Physics based models
(Radiative transfer, NWP, etc.)
Variational data assimilation
(3D/4D-Var)
High latency
Computationally expensive
Analysis and Forecasts
(significant delay, low refresh rates)
Observations
(In-situ, satellite,...)
AI foundation model
(Multi-modal transformer)
Low latency
Efficient inference
Analysis and Forecasts
(near real-time, sub-hourly refresh)

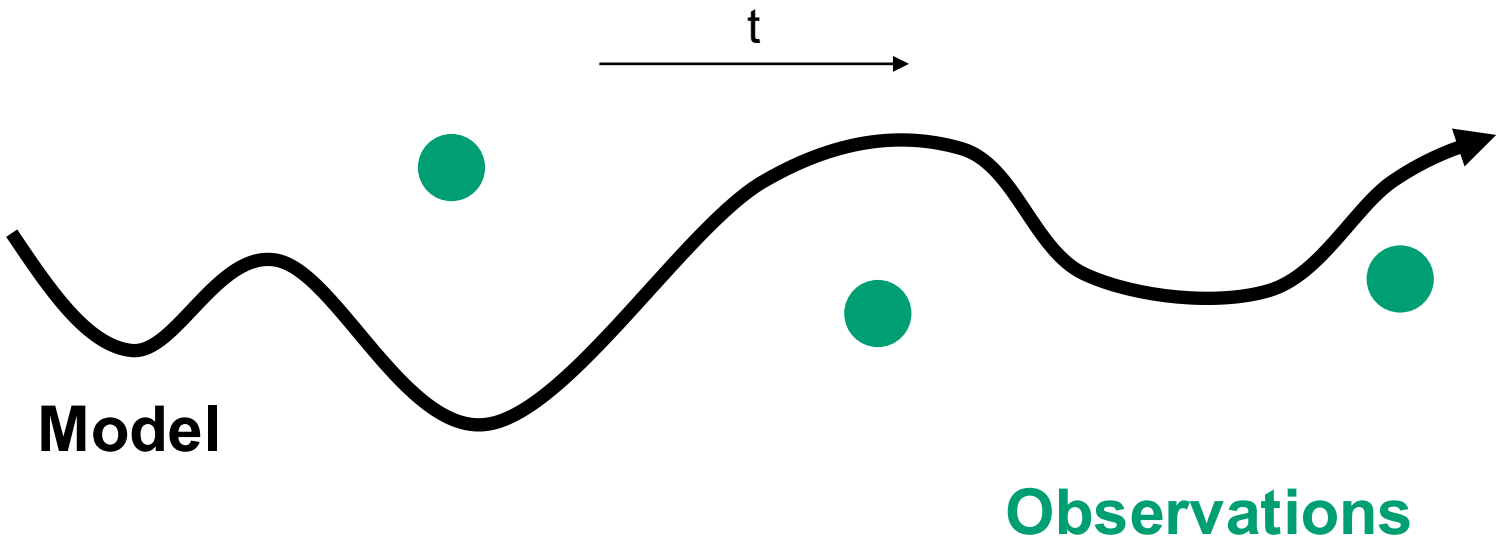
Potential advantages

- Reduce latency
- Do DA every hour instead of every 6 hours
- Better initial state for other ML weather models
- Ingest larger volumes of observations (from 1% to 100%)
- Can create ensembles of initial states

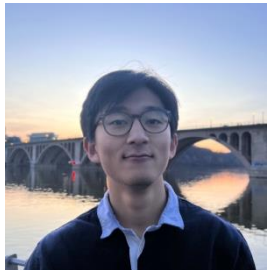
What is Data Assimilation?

- A way of keeping your physics-based model aligned with observations
  - Nudge the state every 6 hours (IFS)
- Maulik et al. description
  - Prior guess about atmospheric state (“background”)
  - Imperfect computational model
  - Noisy observations
  - Posterior guess about the state (“analysis”)

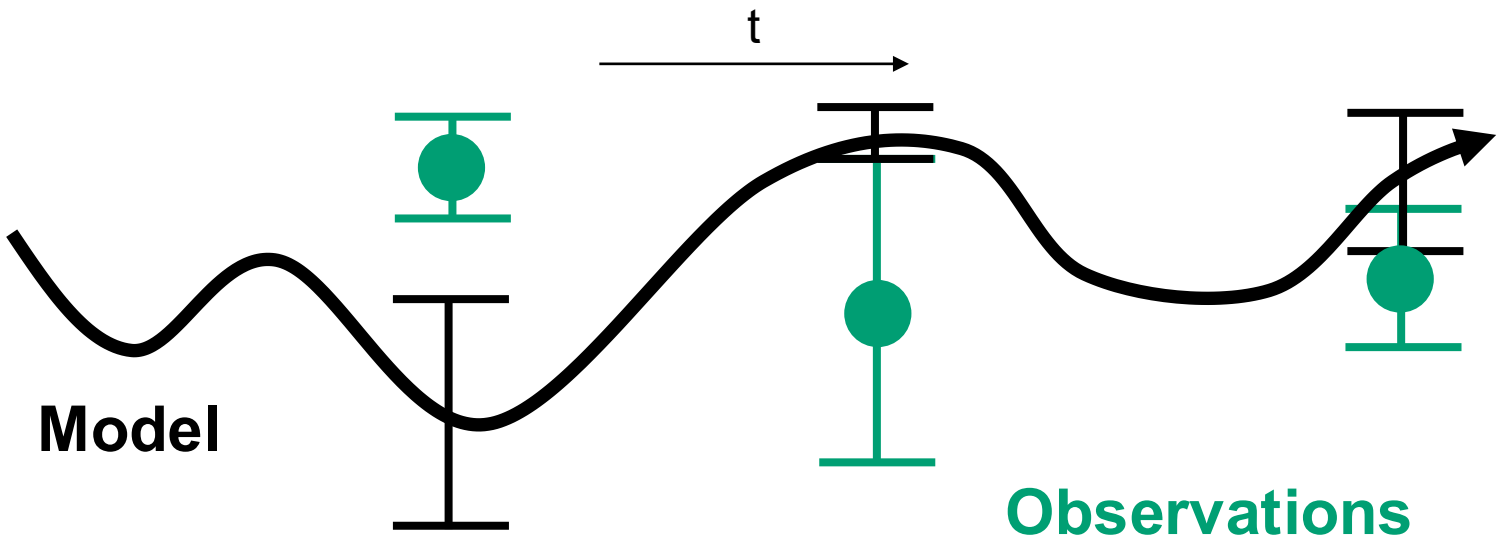
What is “data assimilation”?



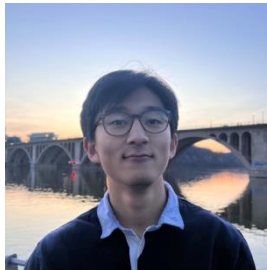
Given **model estimates** and **observations**, what is the most likely atmospheric state?



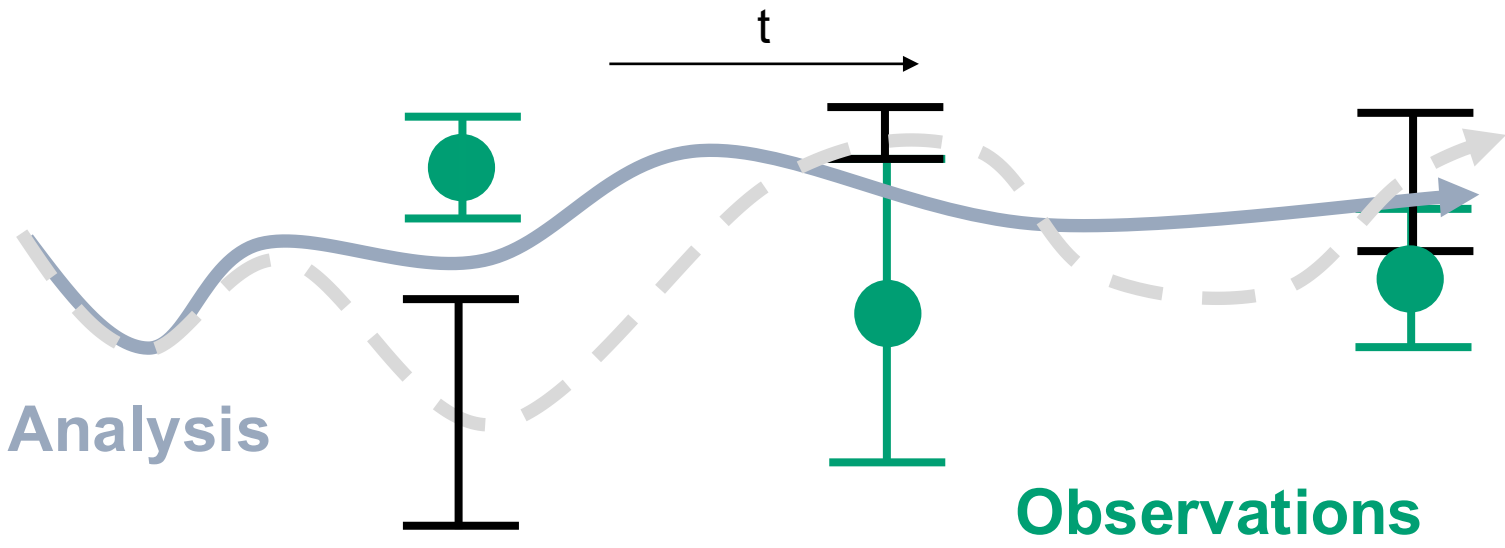
What is “data assimilation”?



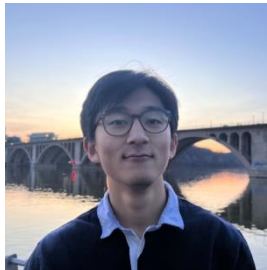
Both model estimates and observations have **errors**



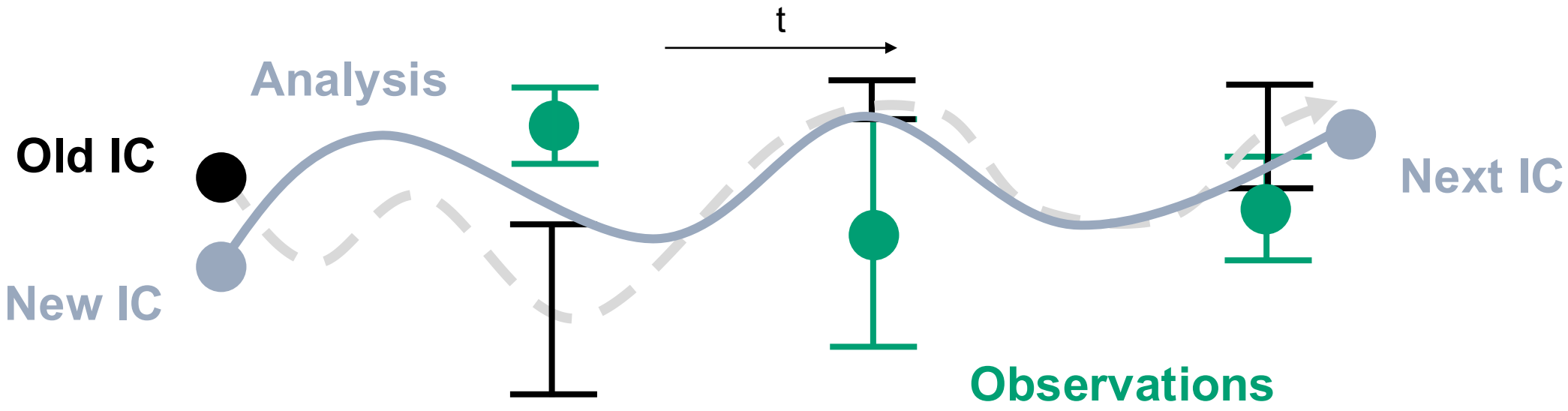
What is “data assimilation”?



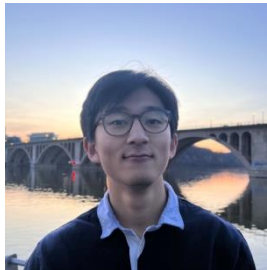
Final **analysis** is determined by error-weighted contributions from observations and model states



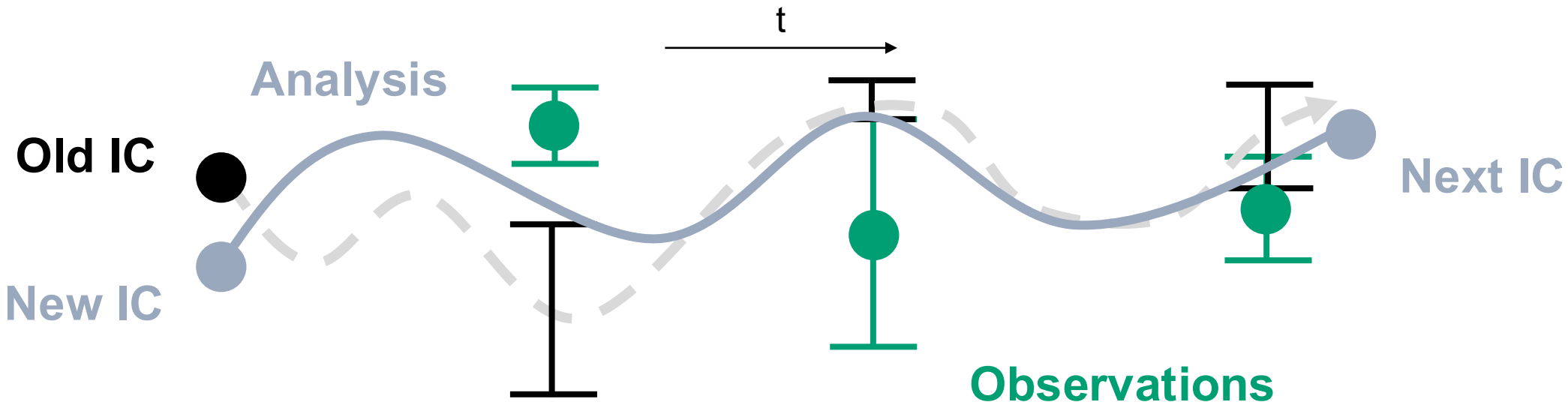
Applications of Data Assimilation



- Real-time Forecasting: Optimizing initial conditions
  - Ex: Operational weather forecasts @ ECMWF, Met Office, NCEP, etc.



Applications of Data Assimilation

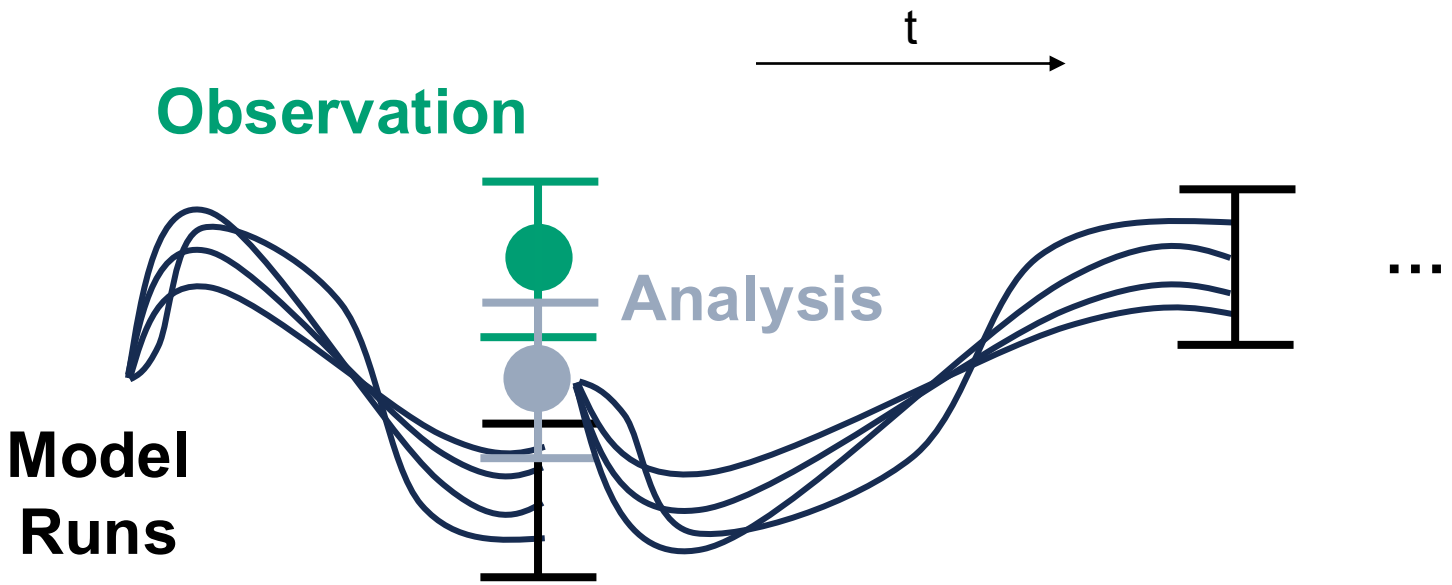


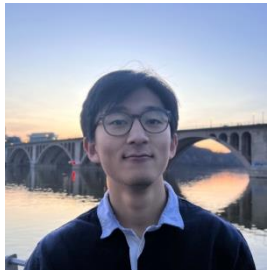
- Real-time Forecasting: Optimizing initial conditions
  - Ex: Operational weather forecasts @ ECMWF, Met Office, NCEP, etc.
- Re-analyses: Re-constructing best estimate of prior atmospheric state
  - Ex: **ERA5** for weather and climate
  - Ex: **CAMS** for atmospheric composition

DA State-of-the-art

Ensemble Kalman Filters (EnKF)

Run a large ensemble of models to compute model error at each time-step

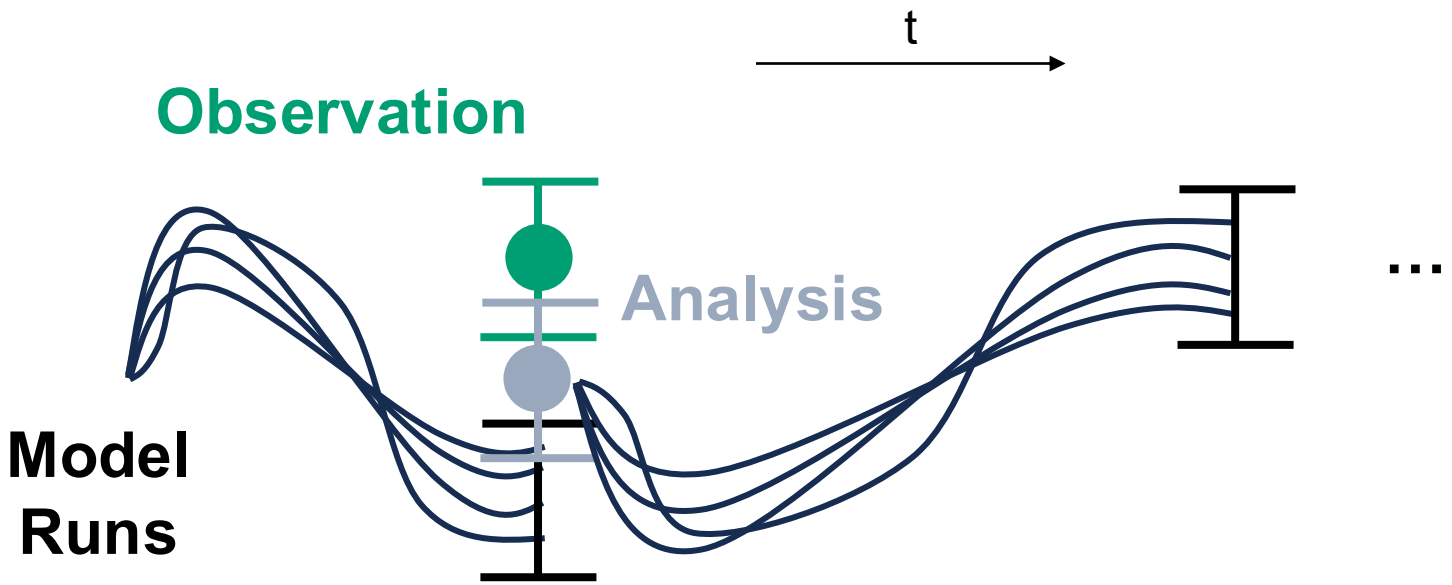


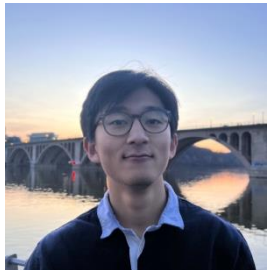


DA State-of-the-art

Ensemble Kalman Filters (EnKF)
Run a large ensemble of models to compute model error at each time-step

+ Model error updated during assimilation
+ Easy to implement
– Requires 10-100s of expensive model runs
– Outputs discrete analysis trajectory





DA State-of-the-art

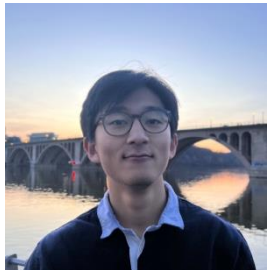
Ensemble Kalman Filters (EnKF)

Run a large ensemble of models to compute model error at each time-step

4D Variational DA (4D-Var)

Determine initial model state which gives best fits to observations in **assim. window***

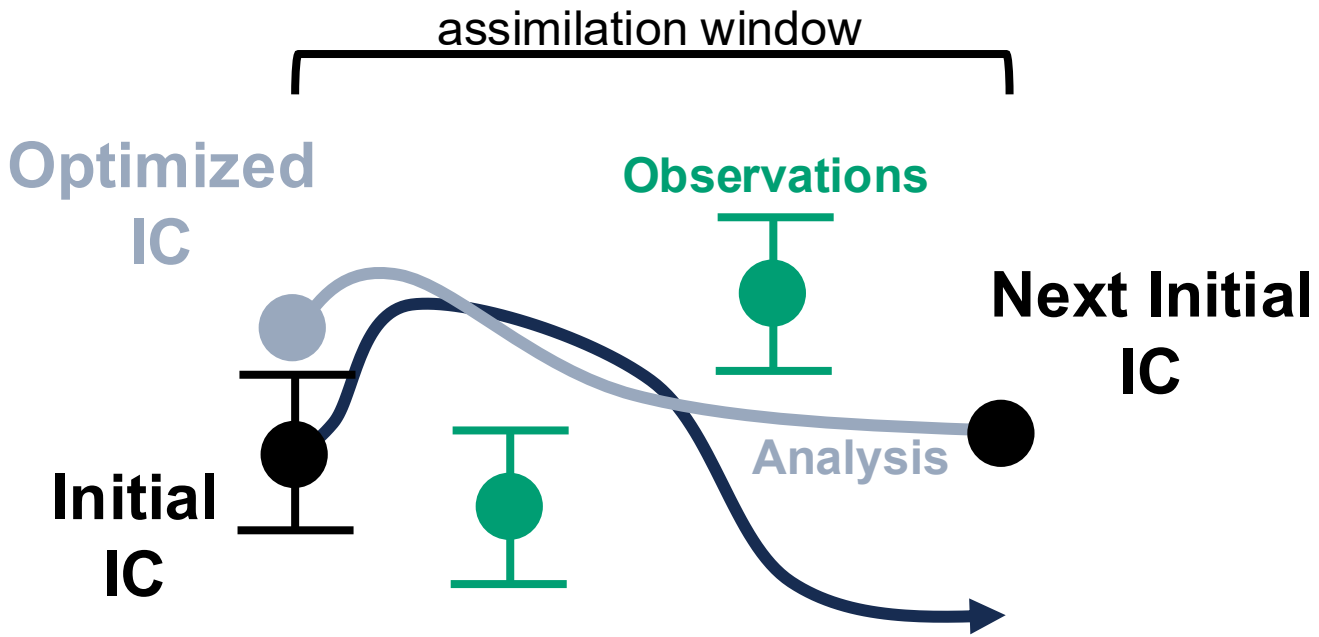
*3D-Var collapses observations to time of IC

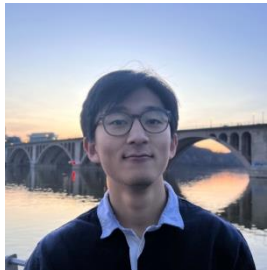


DA State-of-the-art

Ensemble Kalman Filters (EnKF)
Run a large ensemble of models to compute model error at each time-step

4D Variational DA (4D-Var)
Determine initial model state which gives best fits to observations in **assim. window***
*3D-Var collapses observations to time of IC



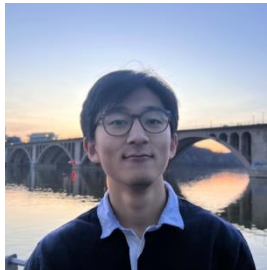


DA State-of-the-art

Ensemble Kalman Filters (EnKF)
Run a large ensemble of models to compute model error at each time-step

4D Variational DA (4D-Var)
Determine initial model state which gives best fits to observations in assim. window

+ Outputs continuous analysis trajectory which obeys model physics
– Requires backwards model gradients during optimization (expensive!)



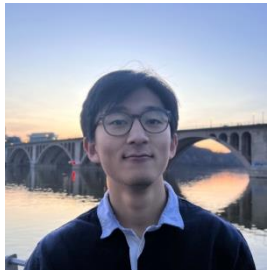
DA State-of-the-art

Ensemble Kalman Filters (EnKF)
Run a large ensemble of models to compute model error at each time-step

4D Variational DA (4D-Var)
Determine initial model state which gives best fits to observations in assim. window

+ Outputs continuous analysis trajectory which obeys model physics
– Requires backwards model gradients during optimization (expensive!)

**** 4D-Var is SOTA, used by ECMWF for IFS and ERA5! ****



State of the Art DA takes time

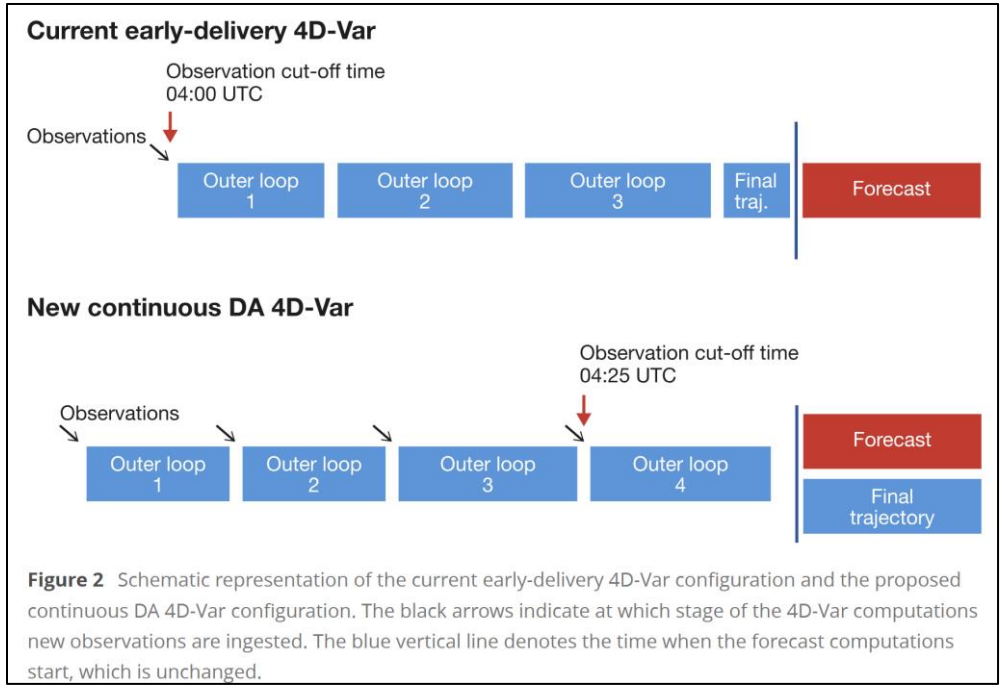


Figure 2 Schematic representation of the current early-delivery 4D-Var configuration and the proposed continuous DA 4D-Var configuration. The black arrows indicate at which stage of the 4D-Var computations new observations are ingested. The blue vertical line denotes the time when the forecast computations start, which is unchanged.

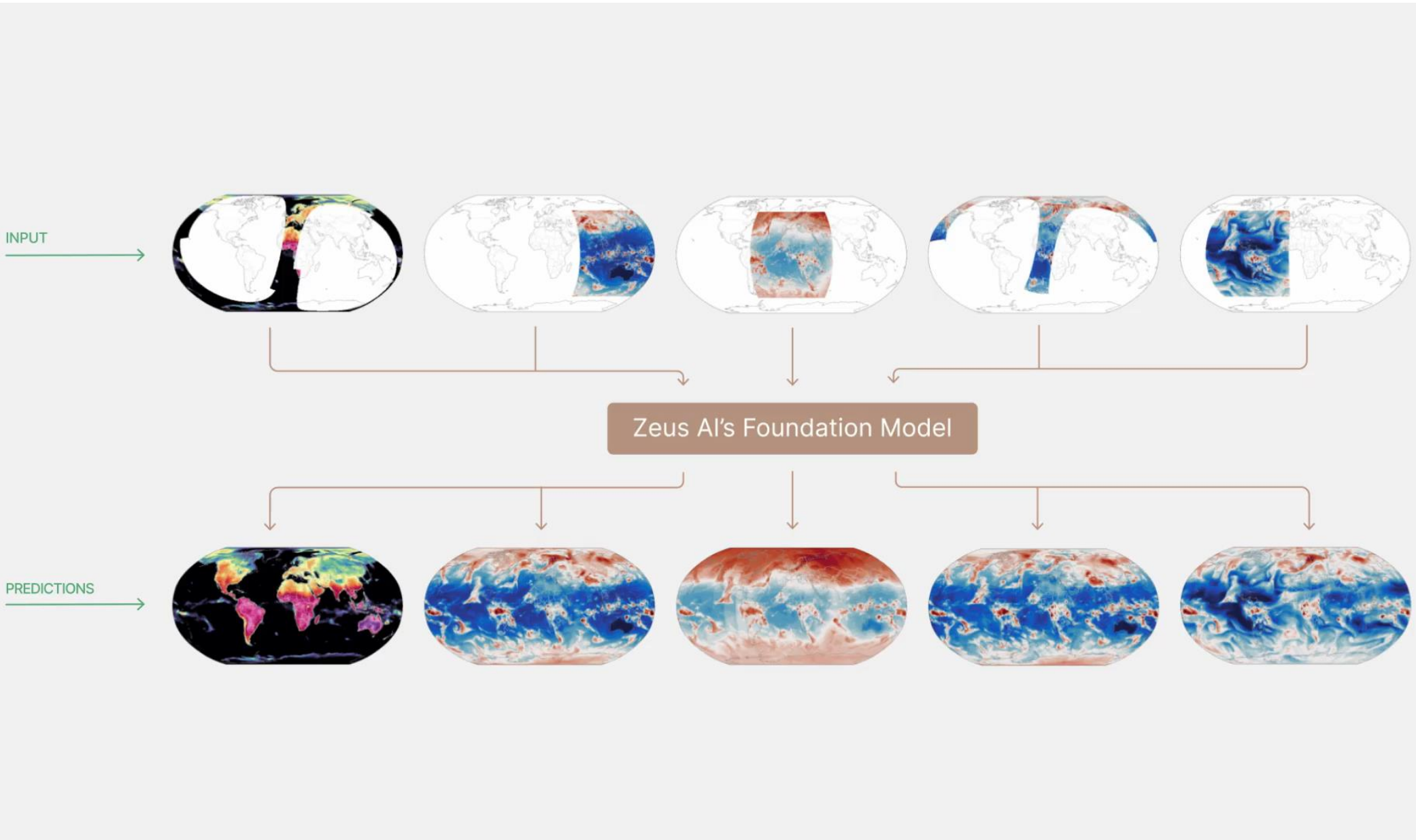
- 4DVar requires large team and compute
- Continuous data assimilation for the IFS | ECMWF
- "For example, computing the 00 UTC analysis with a 21 to 03 UTC assimilation window starts at around 04 UTC. The observation quality control and DA computations take about an hour to complete. This means that by the time the analysis has been produced, the most recent observations that went into producing it are about two hours old."
- Depending on how you count, 4D-Var takes 2 hours or 5 hours to complete.

Back to Vandal et al.

Datasets (modality/sensor)

Modality / Sensor	Orbit	Channels / Variables
GOES-16 Advanced Baseline Imager (ABI) [11]	G	Thermal infrared 10 bands
GOES-18 Advanced Baseline Imager (ABI) [11]	G	Thermal infrared 10 bands
Geostationary Korea Multi-Purpose Satellite - 2A (GK2A) [12]	G	Thermal infrared 10 bands
Spinning Enhanced Visible Infra-Red Imager (SEVIRI) [13]	G	Thermal infrared 8 bands
Advanced Technology Microwave Sounder (ATMS) [14]	L	Brightness temperature 22 bands
Visible Infrared Imaging Radiometer Suite (VIIRS) [15]	L	Thermal infrared 7 bands
Shuttle Radar Topography Mission (SRTM) [16]	S	Elevation, land-sea mask
Microwave integrated Retrieval System temperature [17]	L	3D temperature (37 levels)
Microwave integrated Retrieval System humidity [17]	L	3D specific humidity (37 levels)

Table 1 | **Sensor modalities assimilated in EarthNet** comprise a diversity of spectra and orbital perspectives. Along with static topographical data, these sources provide complementary views of atmospheric and surface states and enable the model to learn complex relationships across space, time, and modality.
G =Geostationary (GEO), L =Low Earth Orbit (LEO), S =Static.


INPUT
Zeus AI's Foundation Model
PREDICTIONS

Model architecture

- Multi-modal Masked Autoencoders
- Subset of input features predicts the remaining features
- Gap-fill across space, time, and sensor

Data processing

- 9 datasets
- Interpolate and reproject to hourly 0.16 degree resolution
- Represent each modality as a tensor with time, space, and channel

Modality / Sensor	Orbit	Channels / Variables
GOES-16 Advanced Baseline Imager (ABI) [11]	G	Thermal infrared 10 bands
GOES-18 Advanced Baseline Imager (ABI) [11]	G	Thermal infrared 10 bands
Geostationary Korea Multi-Purpose Satellite - 2A (GK2A) [12]	G	Thermal infrared 10 bands
Spinning Enhanced Visible Infra-Red Imager (SEVIRI) [13]	G	Thermal infrared 8 bands
Advanced Technology Microwave Sounder (ATMS) [14]	L	Brightness temperature 22 bands
Visible Infrared Imaging Radiometer Suite (VIIRS) [15]	L	Thermal infrared 7 bands
Shuttle Radar Topography Mission (SRTM) [16]	S	Elevation, land-sea mask
Microwave integrated Retrieval System temperature [17]	L	3D temperature (37 levels)
Microwave integrated Retrieval System humidity [17]	L	3D specific humidity (37 levels)

Table 1 | **Sensor modalities assimilated in EarthNet** comprise a diversity of spectra and orbital perspectives. Along with static topographical data, these sources provide complementary views of atmospheric and surface states and enable the model to learn complex relationships across space, time, and modality.
G =Geostationary (GEO), L =Low Earth Orbit (LEO), S =Static.

Training big picture

- Input sample: 12 frames (hours), 144x144 gridcells, many channels from different modalities.
- Tokens per modality: 1 frame, 16x16 gridcells, all channels of this modality.

- Modality specific encoders: Vision transformer and variational autoencoder
- Multimodal masked autoencoder: Transformer (Missing tokens are masked)
- Modality-specific decoders: Variational autoencoder decoding

- Output is gap-filled sample with all tokens present. 12 frames (hours), 144x144 gridcells, many channels from different modalities.
  - Use it as DA: Input the first 11 frames in time, and predict the 12^{th} frame.

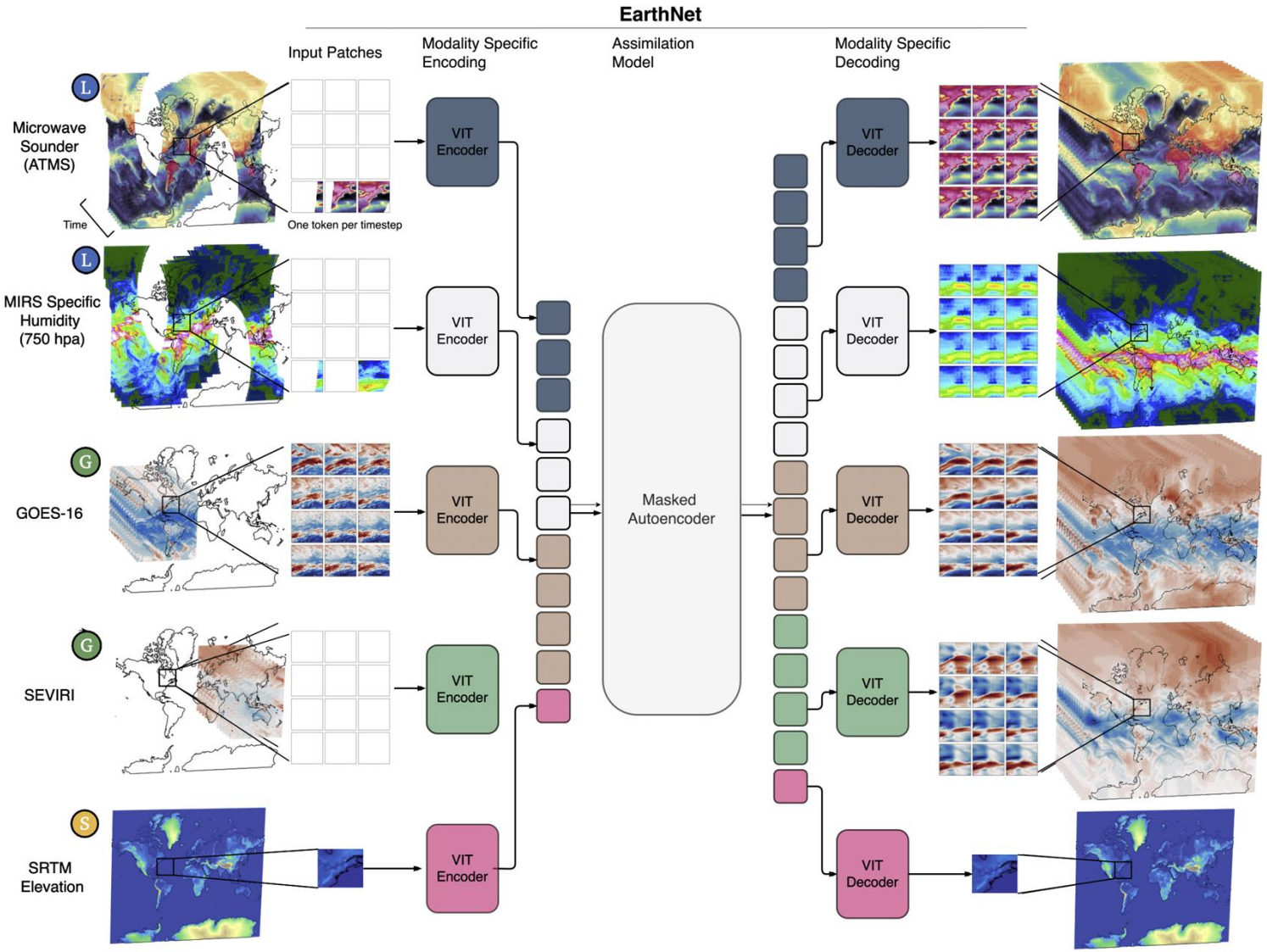


Figure 1 | **EarthNet ingests multi-dimensional Earth observations** from varying orbits and spectra. Sensor modalities in the first four rows have 12 hours of input sequence with a number of channels. The last row is a static elevation variable defining the topography. Sub-images are extracted spatially of size $(144, 144)$ with a token size of $(16, 16)$. Tokens are encoded with a vision transformer and embedded per sensor modality. After tokens pass through the backbone transformer, each decoder sees all context tokens. This process is applied as a moving window across the image and reassembled using Hann windows.

Training details

- VAE pretraining per modality
- *For a 3D modality, a "sample" consists of 12 hourly timesteps, 144x144 cells horizontally, and all channels.*
- *We tokenize these into tokens we call size (1,16,16) meaning 1 timestep, 16x16 cells horizontally, and all channels.*
- *Per sample, about 1750 tokens out of 8752 are complete (no gaps).*
- *For each sample, we select 128 of the complete-data tokens as inputs, and predict (gap-fill) the other ~8500 tokens, ~1500 of which are complete-data tokens we can use for validation.*

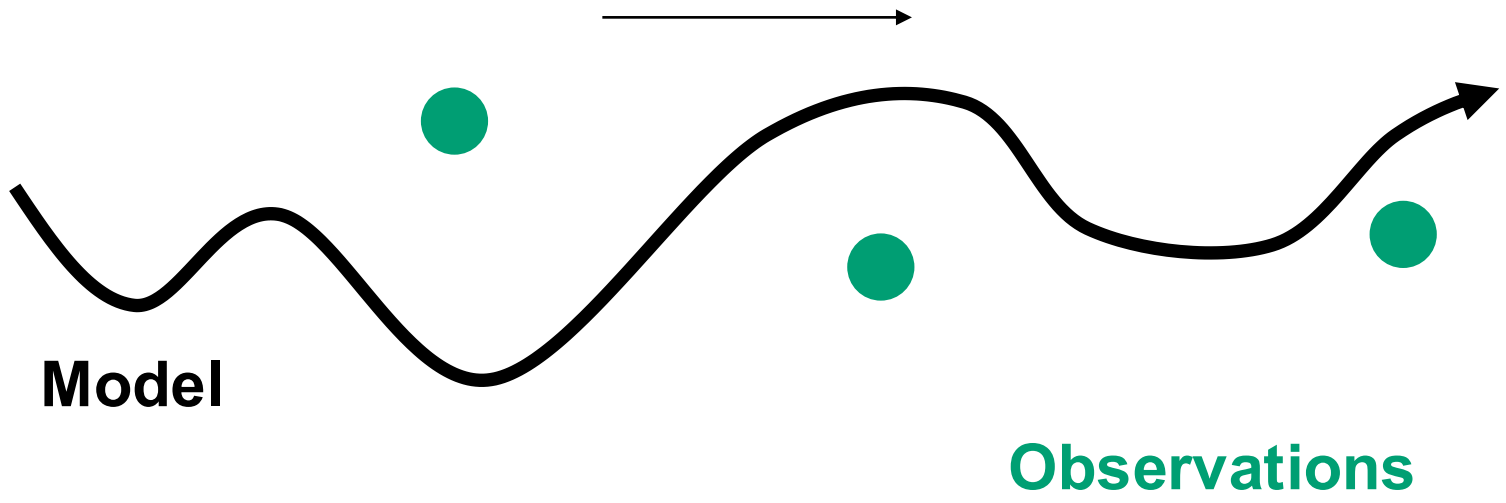
Verification

Verification

- Test set: Febrary and March 2024
- Background departures
- Sensor importance
- Radiosonde verification and comparison with ERA5 and MERRA-2

Verification based on background departures

- Traditionally “background departures” means residuals between the prior model guess and the observations



Verification based on background departures

- In this paper: *We can use EarthNet to make 1-hour forecasts by inputting 11 frames and having it predict the 12th frame. We validated the 1-hour forecasts against observations.*
- Horizontal and vertical distributions of temperature and humidity look reasonable.

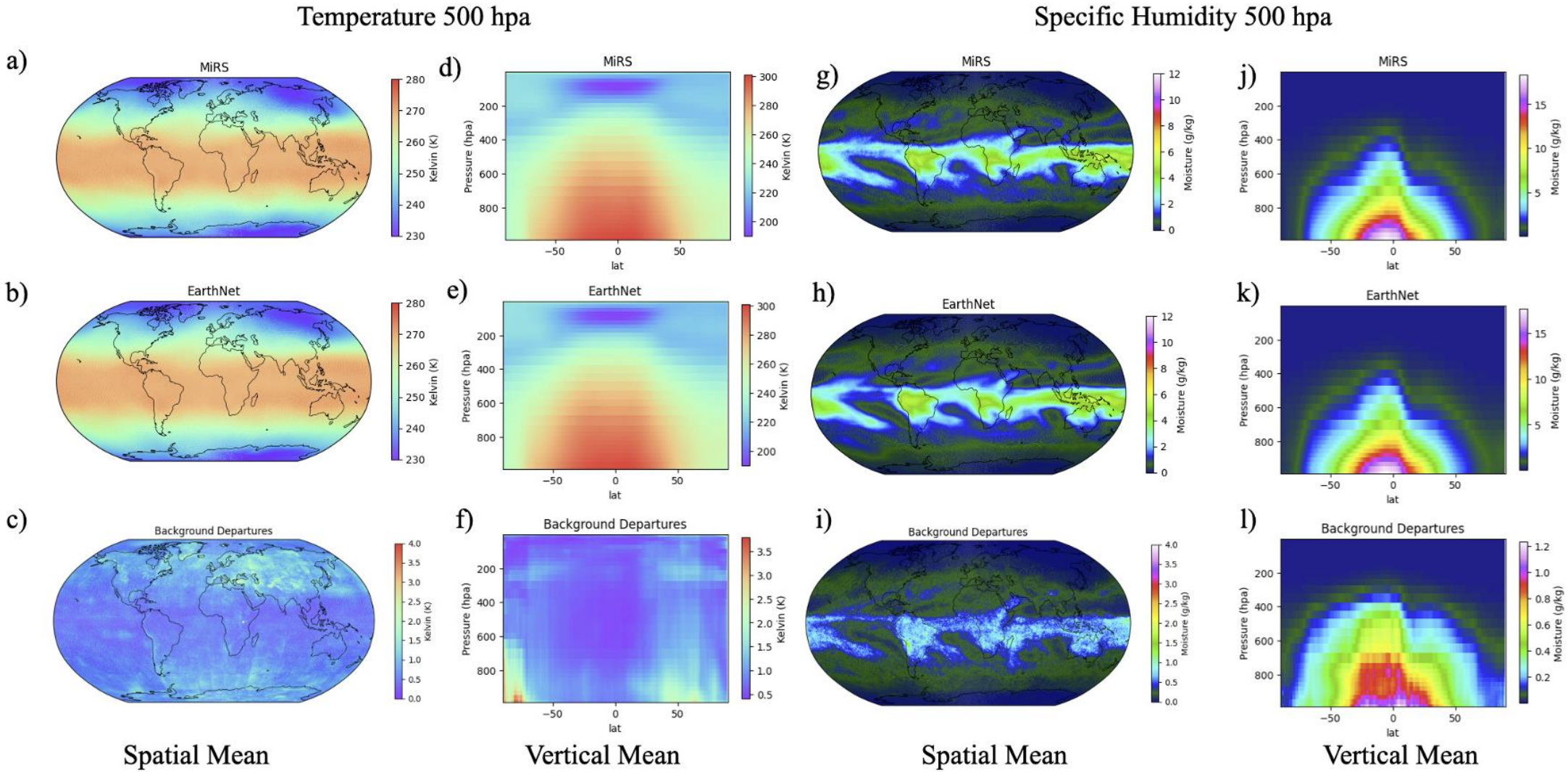


Figure 2 | **Background departures** of EarthNet's temperature (a-f) and specific humidity (g-l) against observations temporally averaged across the spatial and vertical dimensions. The top row shows MiRS average temperature and humidity values across February and March 2024. EarthNet's 1 hour background state average temperature and humidity are shown in the second row. The third row shows error departures as computed from 1 hour background state predictions minus the MiRS observation.

- *We didn’t need to nudge the model based on observation space because we were directly learning in observation space.*
- *The 1-hour forecasts ARE the model guess.*

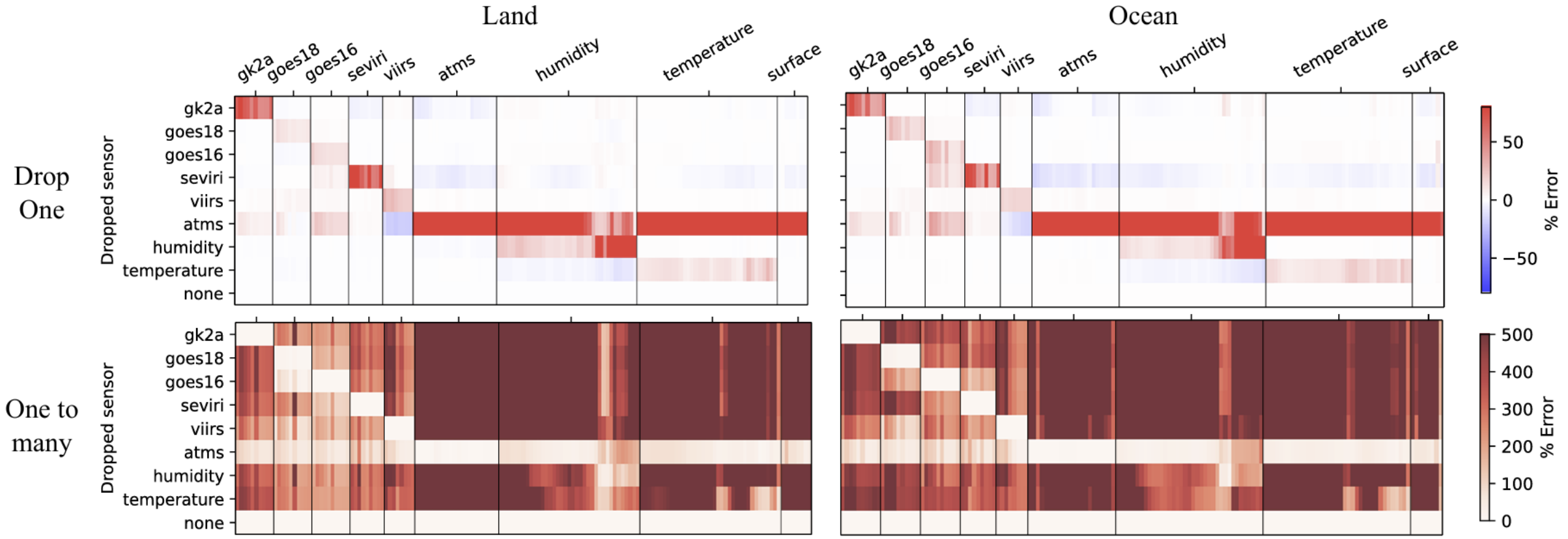


Figure 3 | **Sensitivity analysis and sensor importance** measured by relative mean absolute errors over the land and ocean. (Top row) Each sensor is dropped individually while reconstructing all modalities and comparing errors to the baseline of all modalities included. (Bottom row) One sensor is taken as input to reconstruct all with errors computed as above. The analysis is split into the land (left) and ocean (right) regions to delineate surface types.

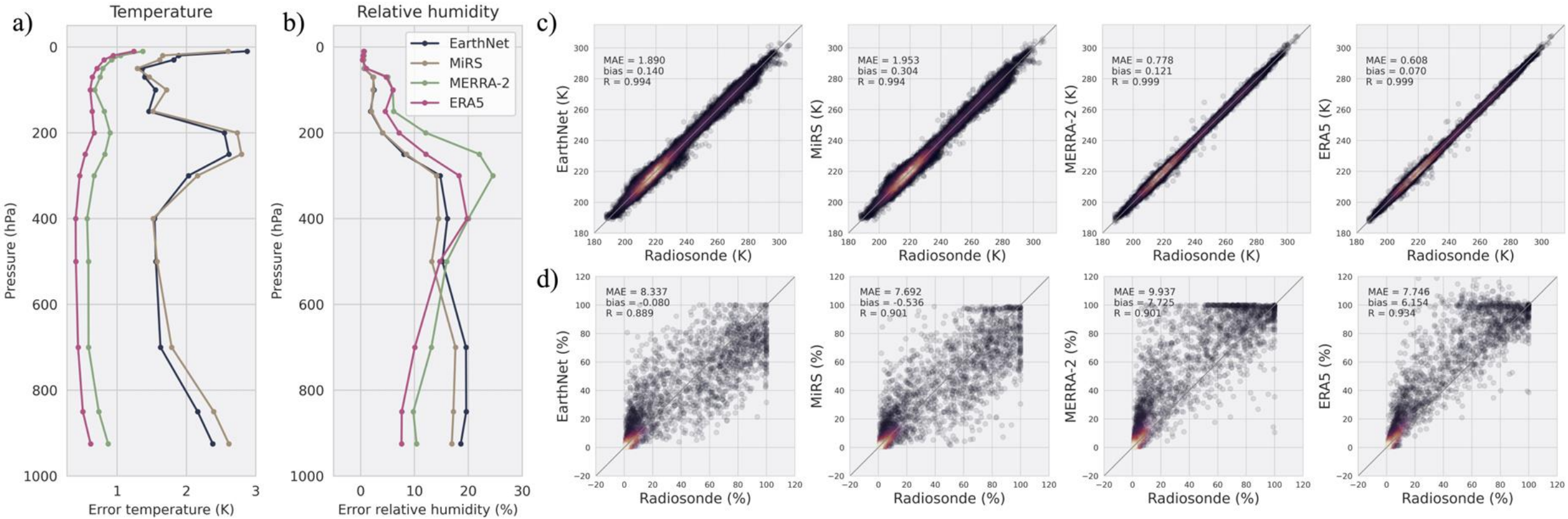


Figure 4 | **Comparison with radiosonde soundings** shows that EarthNet's performance is similar to MiRS observations in matching radiosonde observations of temperature (a) and humidity (b). EarthNet outperforms MERRA-2 and ERA5 reanalyses for humidity predictions between 50 and 500 hPa, without the benefit of having assimilated radiosonde data. Scatter plots show a strong linear relationship between ERA5/MERRA-2 reanalysis and radiosonde temperature observations (c). Relative humidity scatter plots show more normally distributed errors for EarthNet and MiRS at high humidity values (d).

Pressure (hPa)	Temperature Error (K)				Relative Humidity Error (%)			
	EarthNet	**MiRS**	**MERRA-2**	**ERA5**	**EarthNet**	**MiRS**	**MERRA-2**	**ERA5**
925	2.38	2.61	0.87	0.62	18.58*	16.91	10.39	7.59
850	2.16	2.40	0.74	0.51	19.60*	17.21	9.77	7.63
700	1.63	1.79	0.59	0.44	19.55*	17.59	13.12	10.05
500	1.56*	1.58	0.59	0.40	15.17*	13.21	15.97	14.69
400	1.54*	1.52	0.57	0.40	16.10*	14.44	20.02	19.75
300	2.03*	2.16	0.67	0.46	14.77*	14.11	24.53	18.27
250	2.61*	2.80	0.82	0.54	8.04	8.49	22.03	12.13
200	2.55*	2.73	0.90	0.67	4.06*	4.10	12.07	7.13
150	1.46*	1.52	0.81	0.64	1.86*	2.01	6.13	4.58
100	1.55*	1.71	0.68	0.61	2.47*	2.30	5.96	6.00
70	1.40	1.46	0.76	0.64	2.32*	2.40	5.03	4.73
50	1.37*	1.29	0.79	0.71	0.66*	0.70	1.01	1.06
30	1.82*	1.61	0.92	0.81	0.55*	0.54	0.34	0.34
20	1.89*	1.66	1.05	0.94	0.60*	0.60	0.46	0.46
10	2.88*	2.60	1.37	1.24	0.66*	0.66	0.57	0.58
Average	**1.92**	**1.96**	**0.81**	**0.64**	**8.33**	**7.68**	**9.83**	**7.67**

Table 6 | **3D data verification** for EarthNet, MiRS, MERRA-2 and ERA5. Temperature and relative humidity are compared to radiosondes at different pressure levels of the atmosphere. Starred (*) EarthNet values indicate no statistically significant difference from MiRS errors at the $p < 0.05$ level.

Radiosonde verification

- *If we gap-fill MiRS, we get essentially a global reanalysis dataset, which we can compare with other reanalysis datasets.*
- Compare with ERA5 and MERRA-2
- Compare with MiRS

Discussion

- Strictly observational data
- Can gap-fill across modalities, across time, and across space.
- Call it a 1-hour forecast
- Call it a new DA technique

MLJC: Is this DA?

- DA for initial state of my NWP?
- DA for initial state in (NWP) model-space?
- DA for initial state in observation space?
- Assimilation-free?