SOFAR

14–17 November 2022 | #AIforEOWS

ECMWF-ESA Workshop on Machine Learning for Earth Observation and Prediction

ECMWF esa

2022

An Overview of AI/ML and Data Assimilation:

current state of the art, challenges, and opportunities

SOFAR

14–17 November 2022 | #AIforEOWS

ECMWF-ESA Workshop on Machine Learning for Earth Observation and Prediction

ECMWF esa

An Overview of AI/ML and Data Assimilation

Current state of the art, challenges, and opportunities

STEPHEN G. PENNY

DATA ASSIMILATION LEAD
SOFAR OCEAN TECHNOLOGIES

RESEARCH AFFILIATE
CIRES, UNIVERSITY OF COLORADO BOULDER

VISITING ASST. PROFESSOR
UNIVERSITY OF MARYLAND

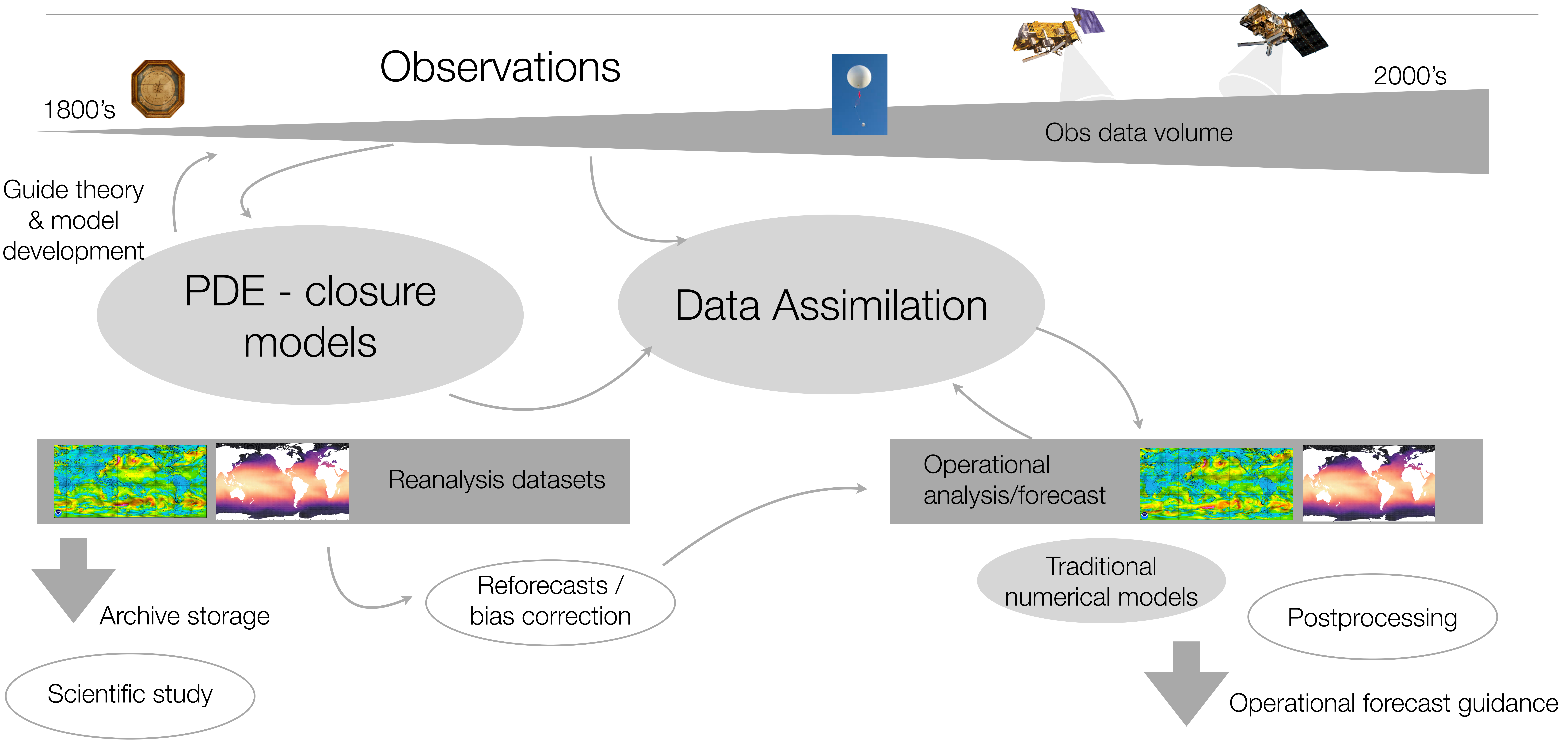
14 NOVEMBER 2022

CIRES

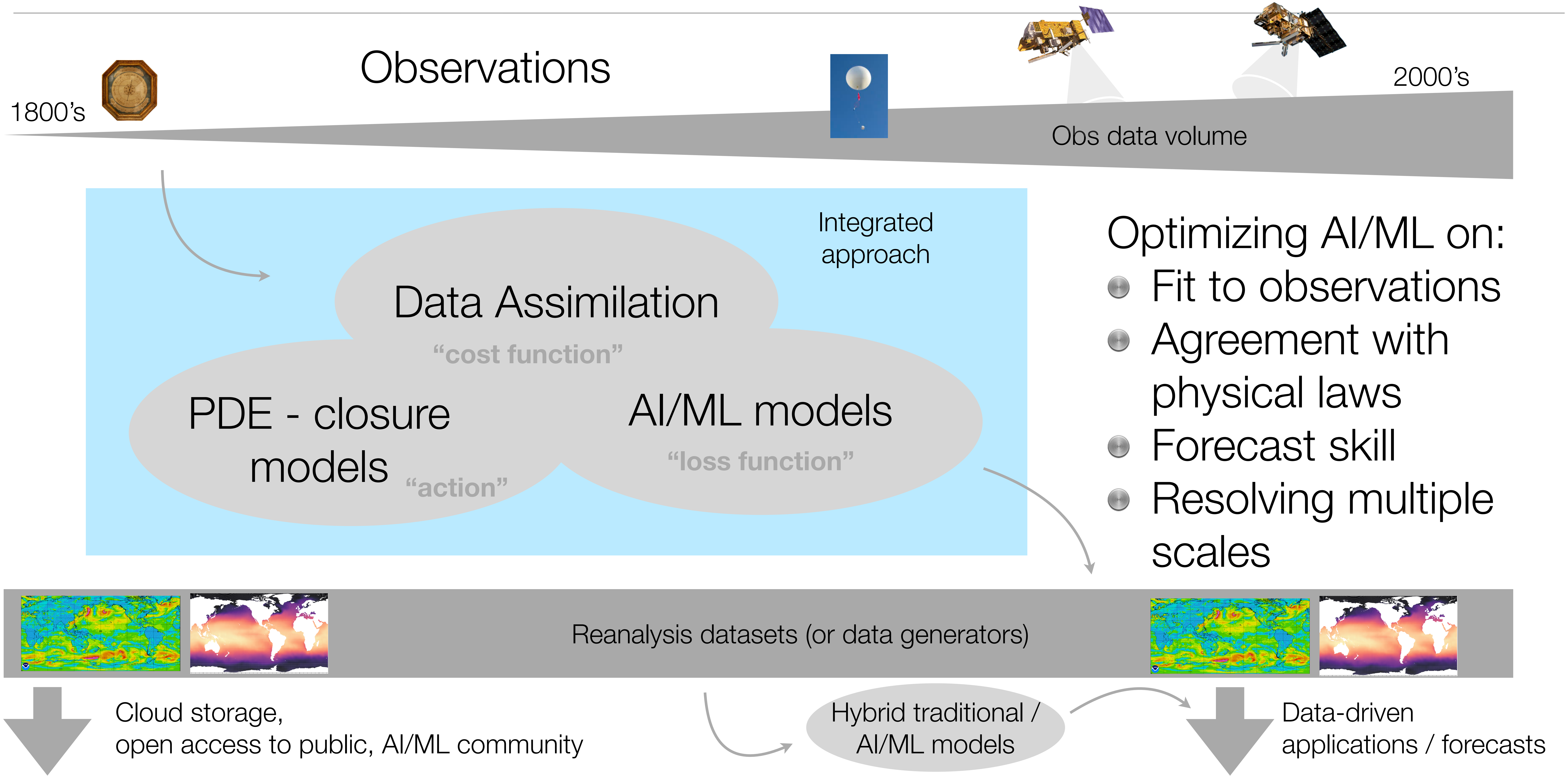
University of Colorado Boulder

University of Maryland 1856

Current path from observations to operational forecast



Transition focus to:
High quality, scientifically validated, curated data products that will drive training and execution of AI/ML-based methods



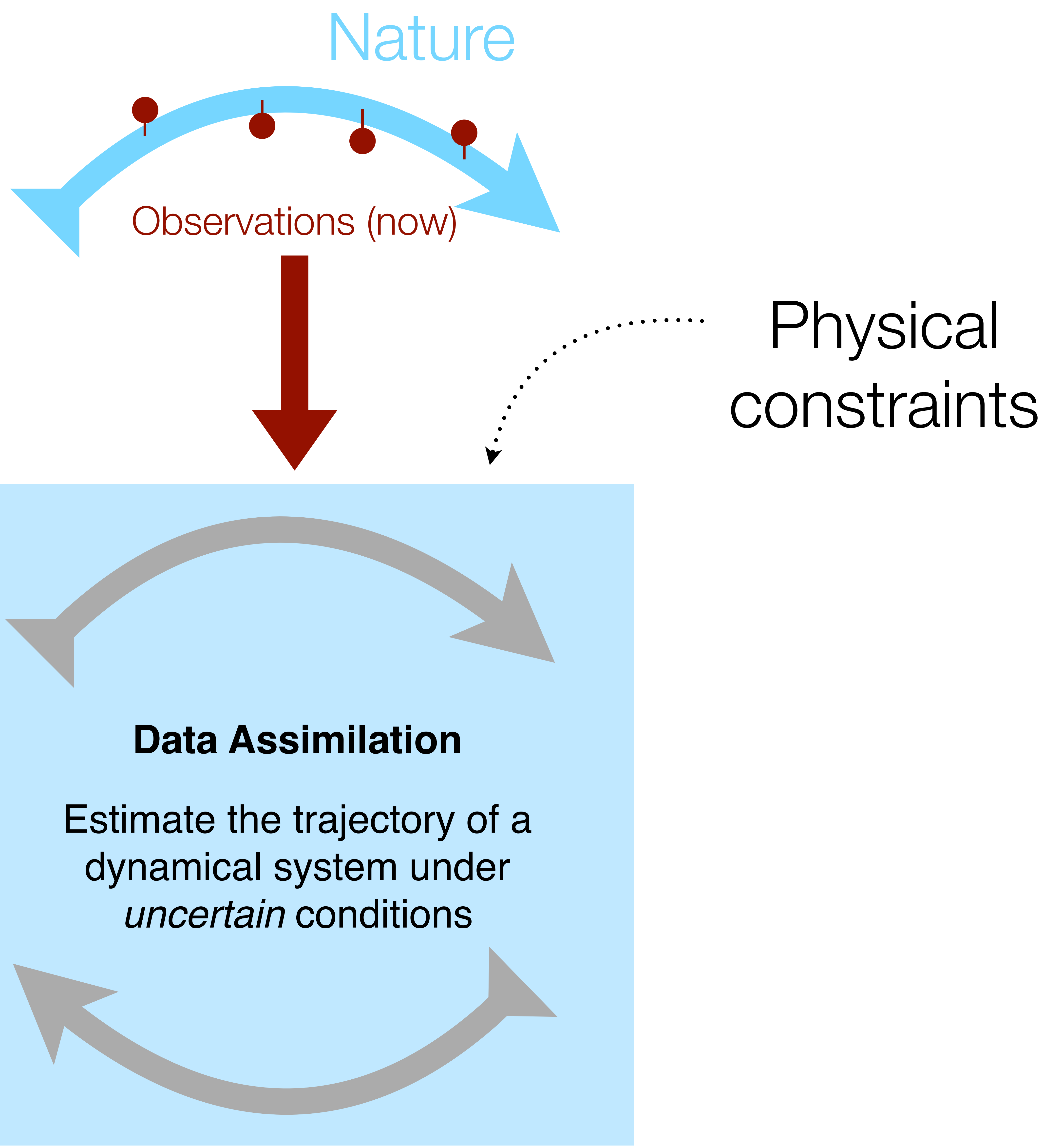
Optimizing AI/ML on:

- Fit to observations
- Agreement with physical laws
- Forecast skill
- Resolving multiple scales

SECTION

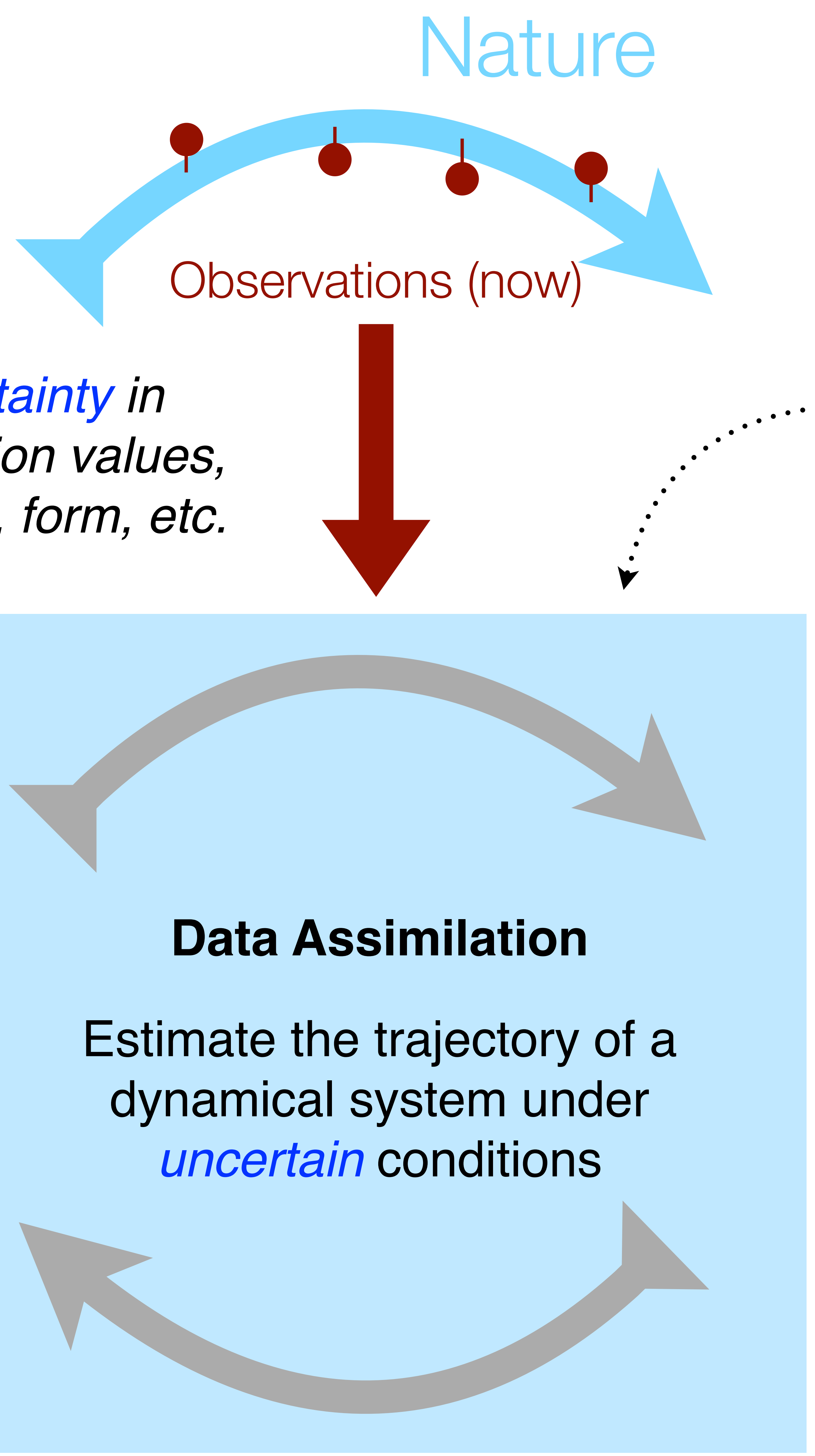
Data Assimilation: what is the fundamental problem?

Inputs:


Nature
Observations (now)
Physical constraints
Data Assimilation
Estimate the trajectory of a dynamical system under *uncertain* conditions

Inputs:

Uncertainty in observation values, locations, form, etc.

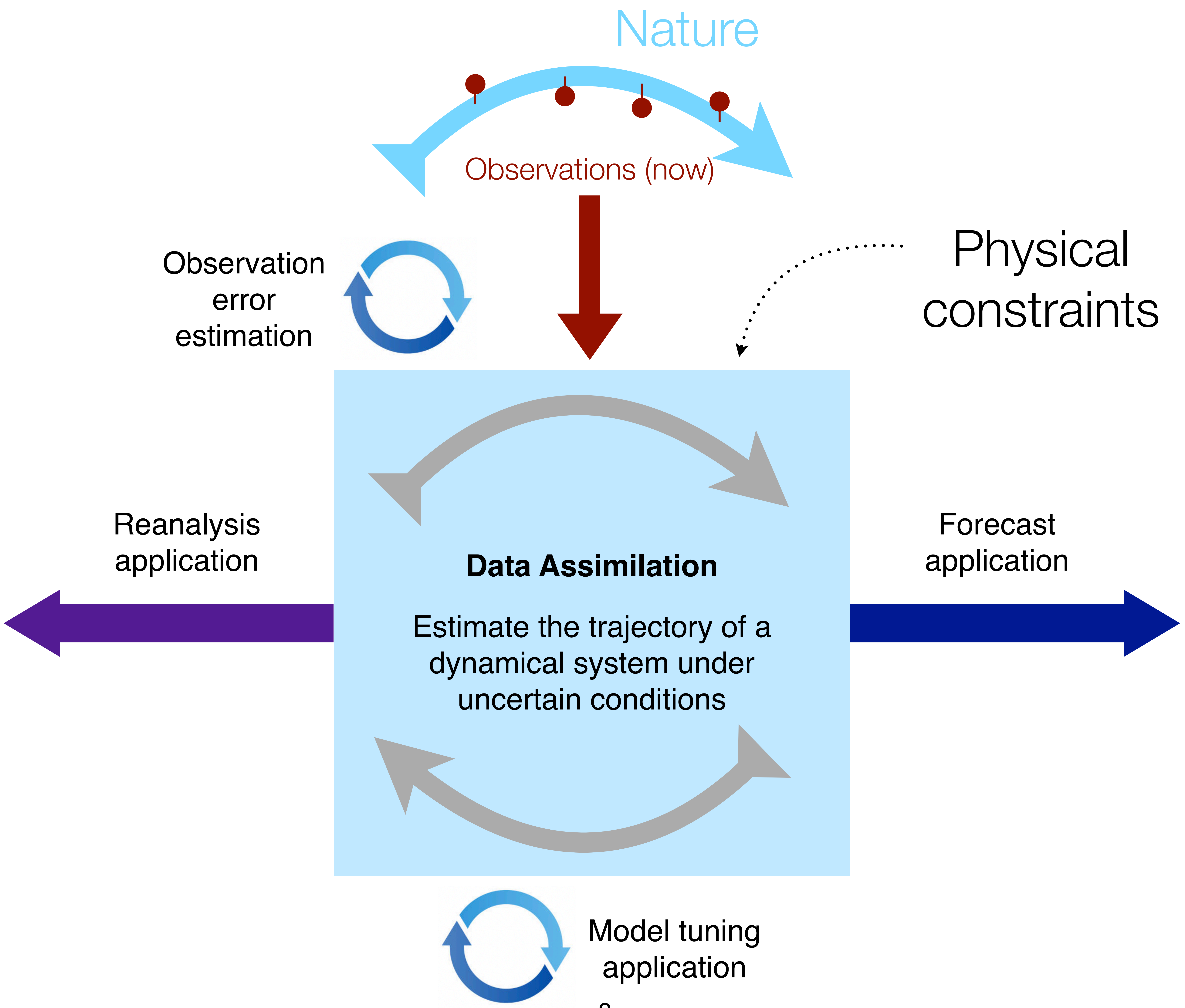


Physical constraints

Uncertainty in how physical laws apply after formulating a tractable representation of the problem (e.g. discretization, grid resolution, numerical solvers, parameterized processes, etc.)

We assume we are working with very large systems - e.g. $O(10^9)$ - i.e. we are only interested in scalable methods

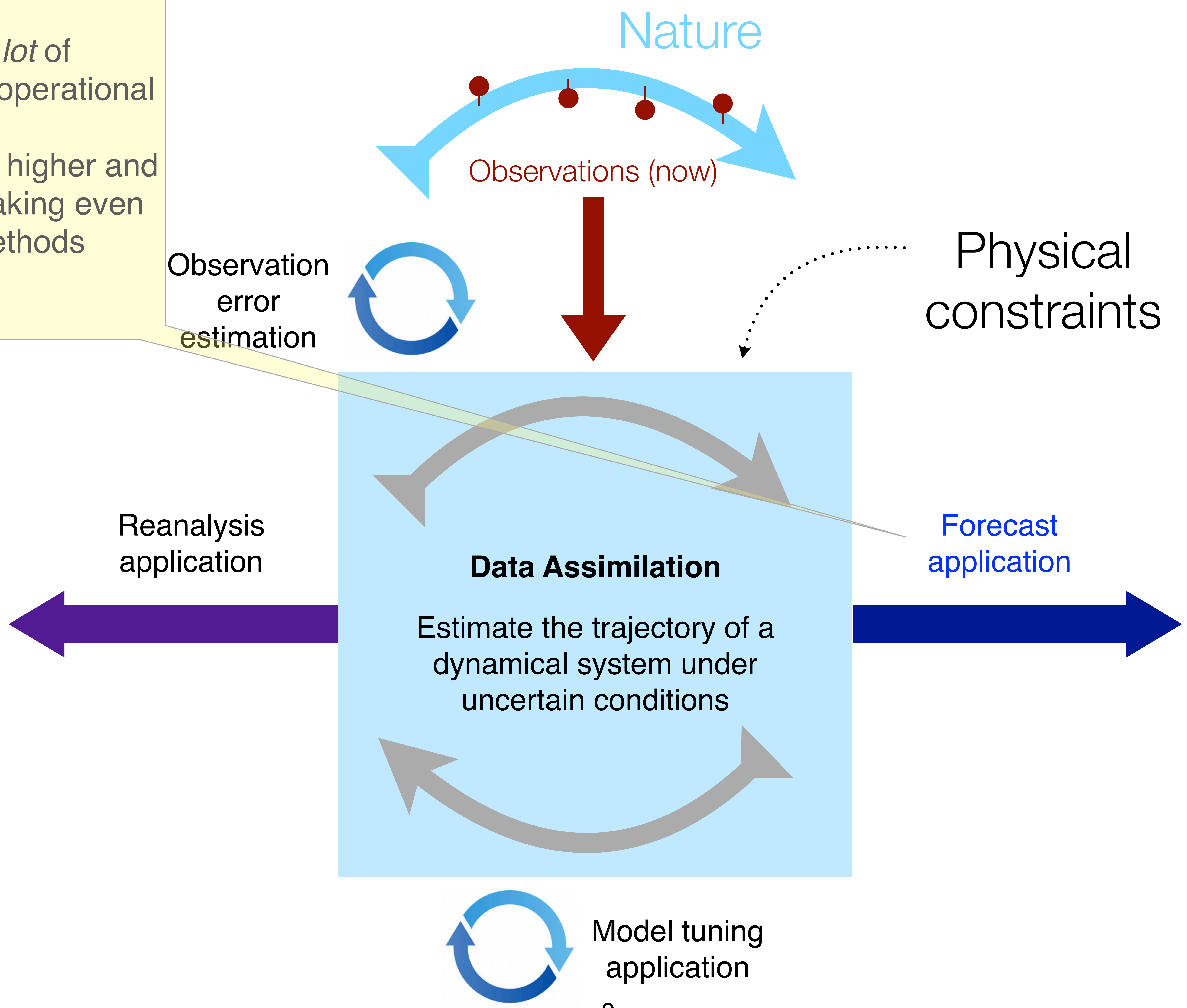
Applications:



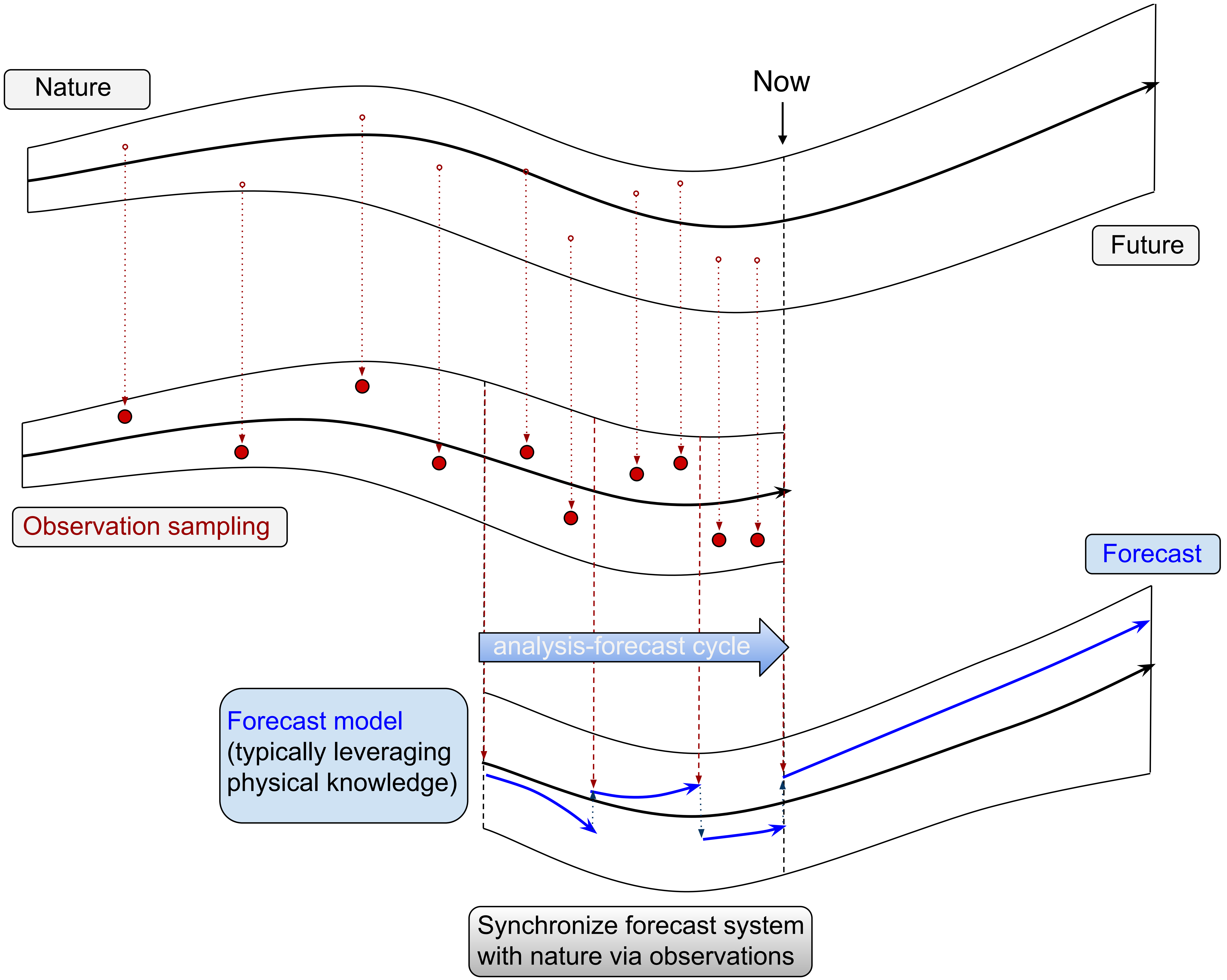
Growing problems:

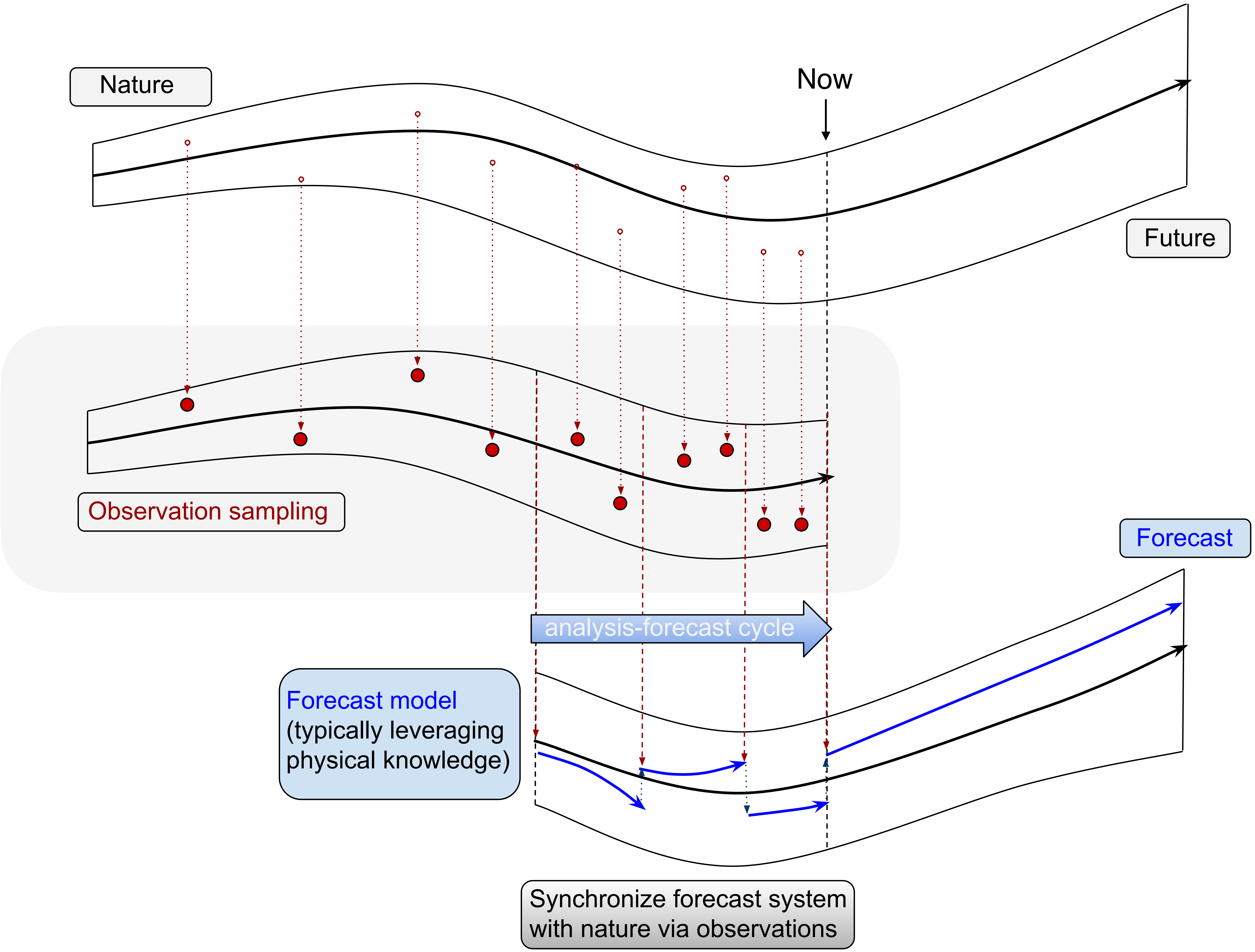
- Modern DA ignores *a lot* of observational data in operational forecast applications
- Models are moving to higher and higher resolutions, making even state-of-the-art DA methods sometimes too costly

Applications:



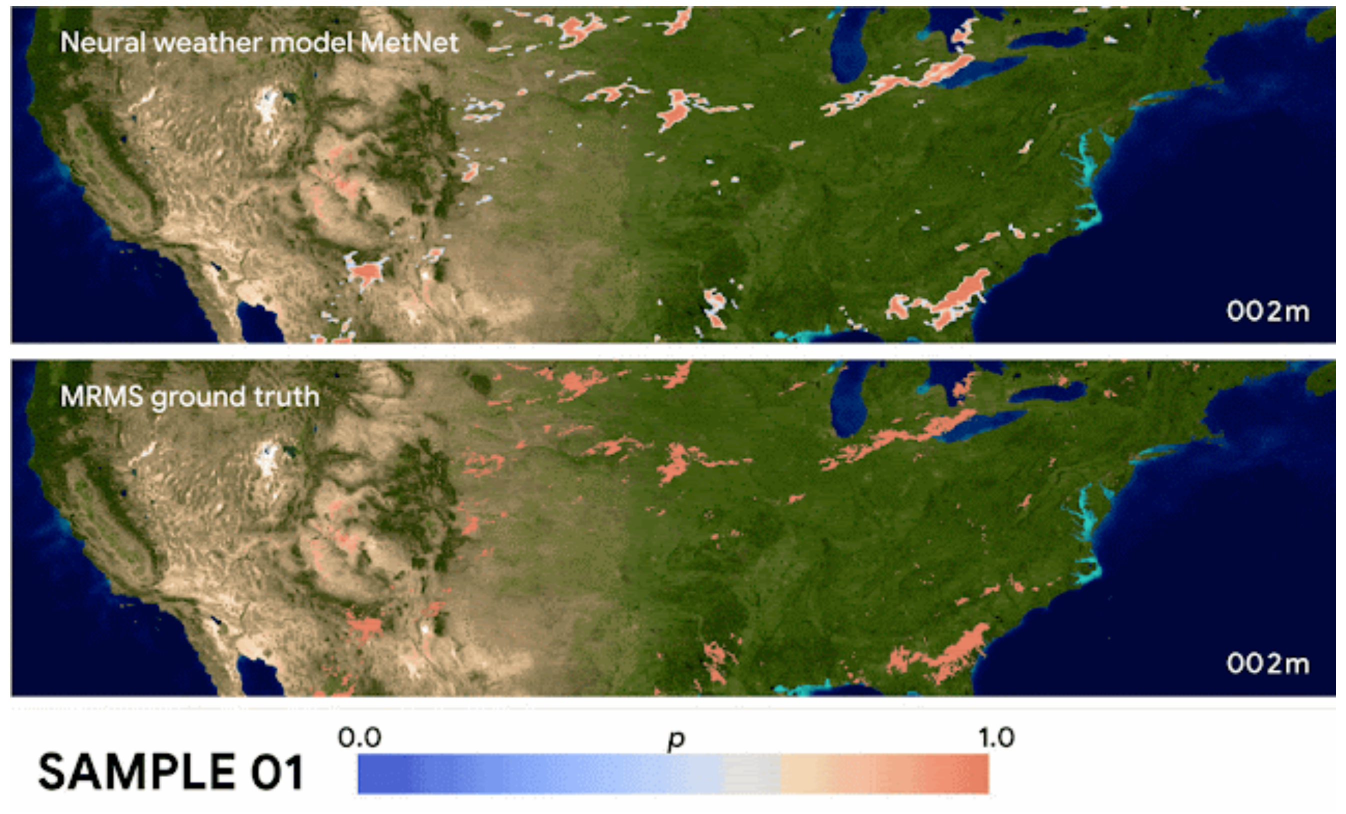
What is the path towards an end-to-end AI/ML DA solution?

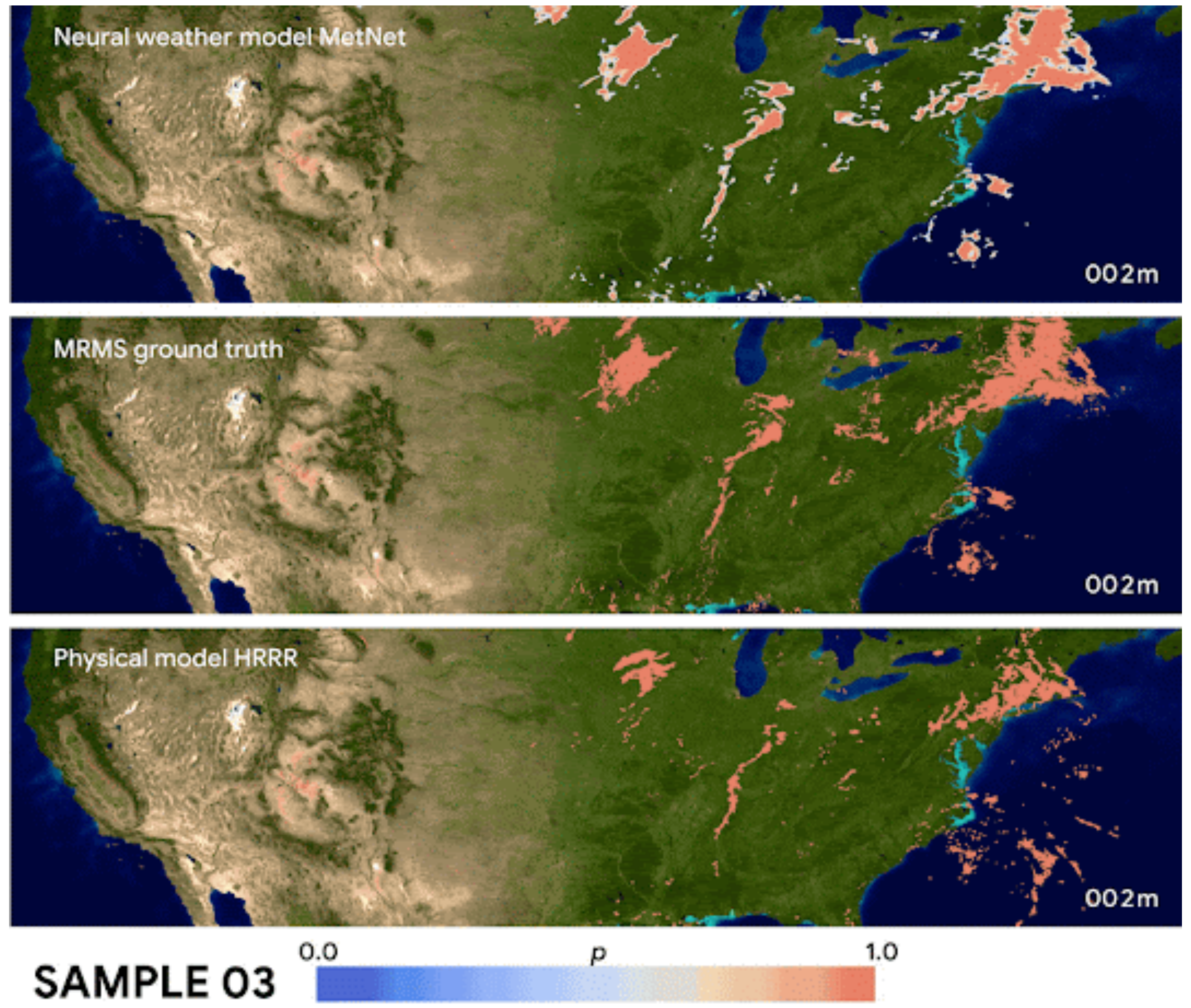

Nature
Now
Future
Observation sampling
Forecast
analysis-forecast cycle
Forecast model
(typically leveraging
physical knowledge)
Synchronize forecast system
with nature via observations

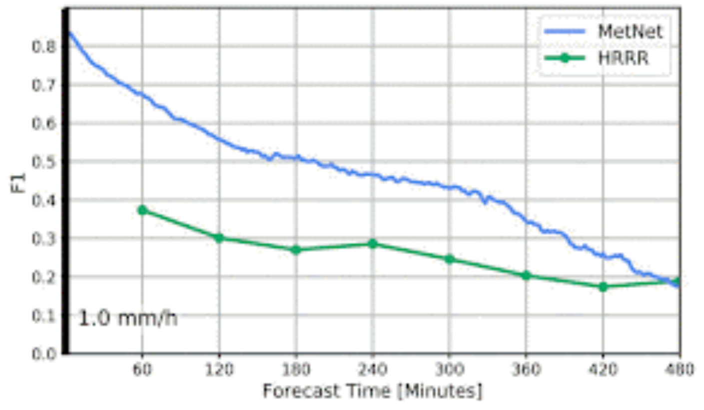


Nature
Now
Future
Observation sampling
Forecast
analysis-forecast cycle
Forecast model
(typically leveraging
physical knowledge)
Synchronize forecast system
with nature via observations

MetNet *(Sønderby et al. 2020)*
- Taking inputs from ground radar and satellite measurements
- Predicting point-wise precipitation probabilities







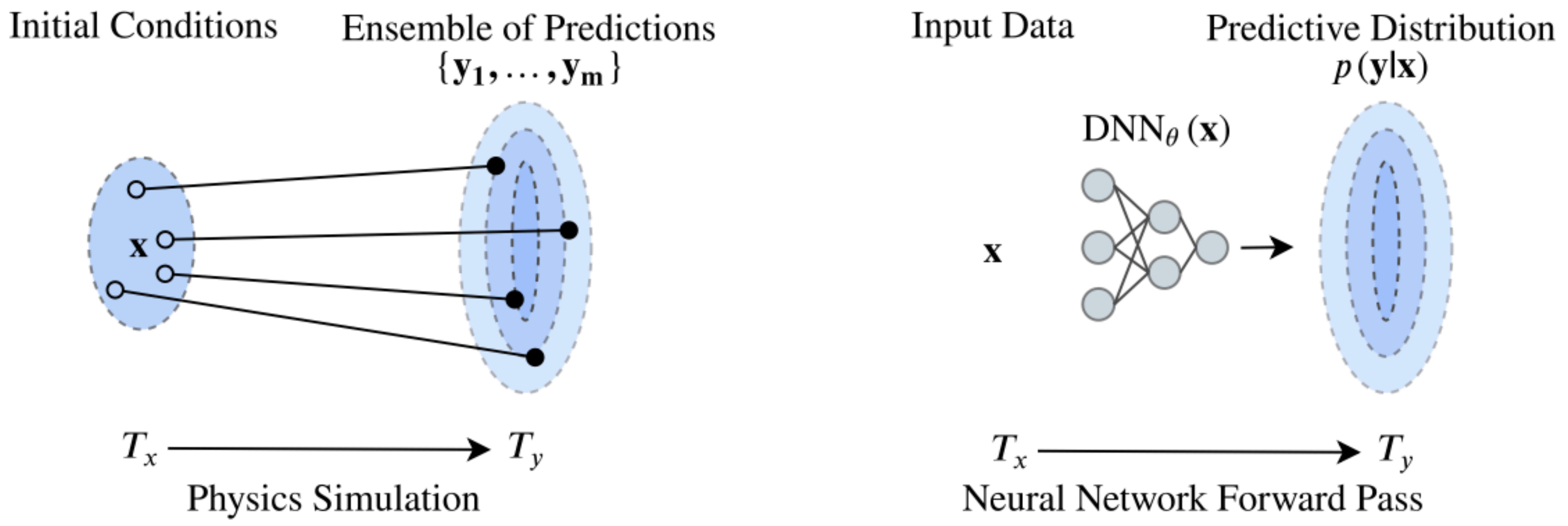


Figure 4: Properties of NWP and NWMs. Left: NWP performs a deterministic physical simulation starting from the initial conditions. The predictive uncertainty is estimated from an ensemble of predictions each run with slightly different initial conditions. Right: The NWM treats the current observations as direct inputs to a DNN, directly estimating the distribution over future conditions $p(\boldsymbol{y}|\boldsymbol{x})$.

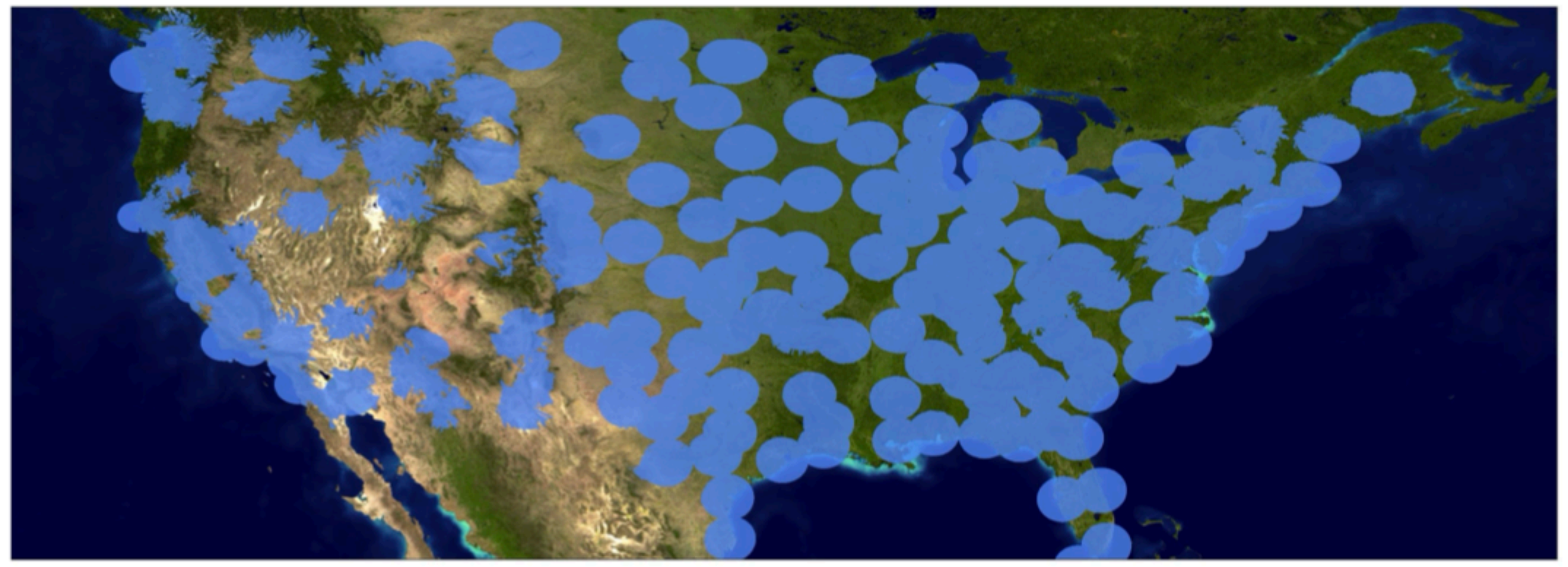
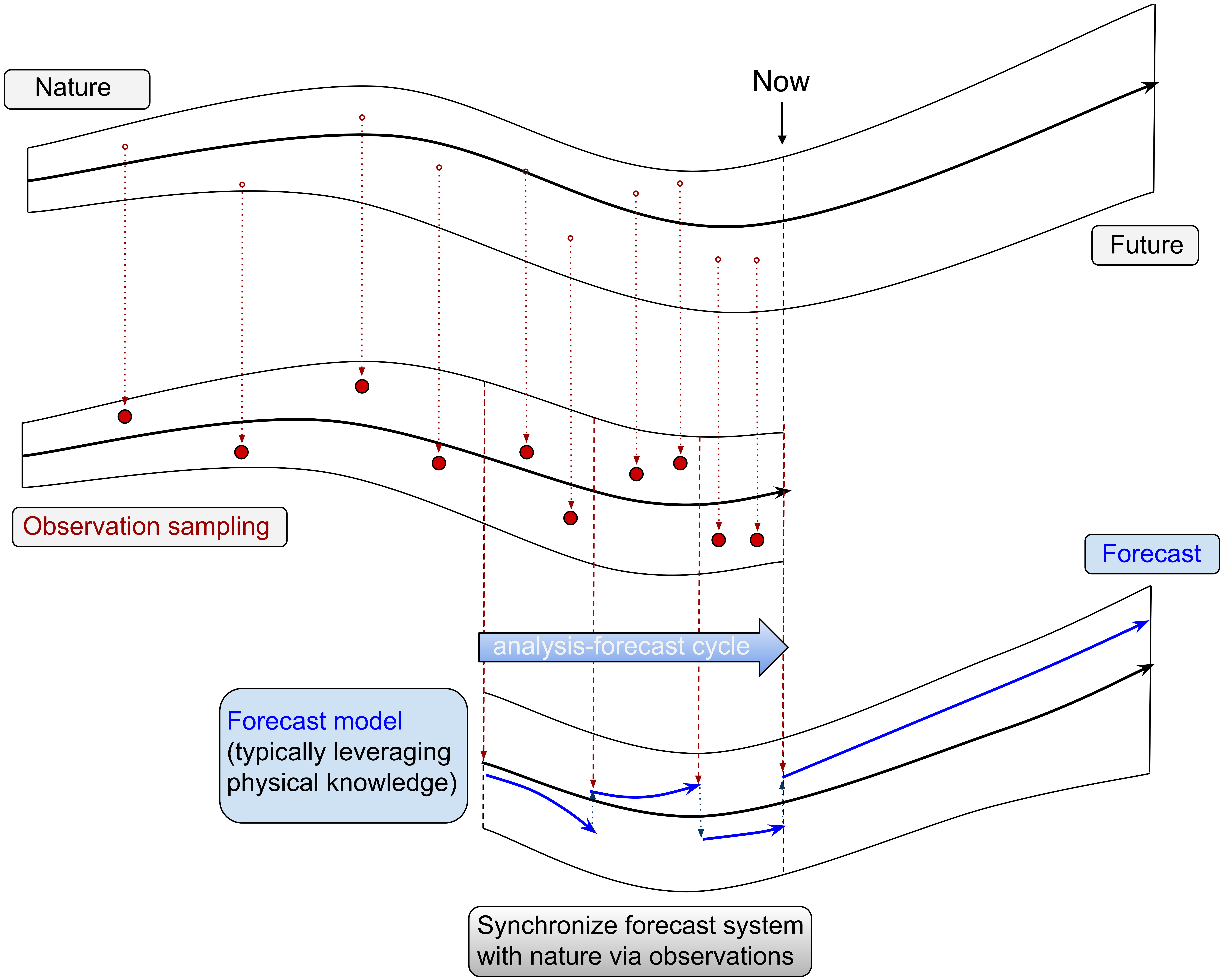
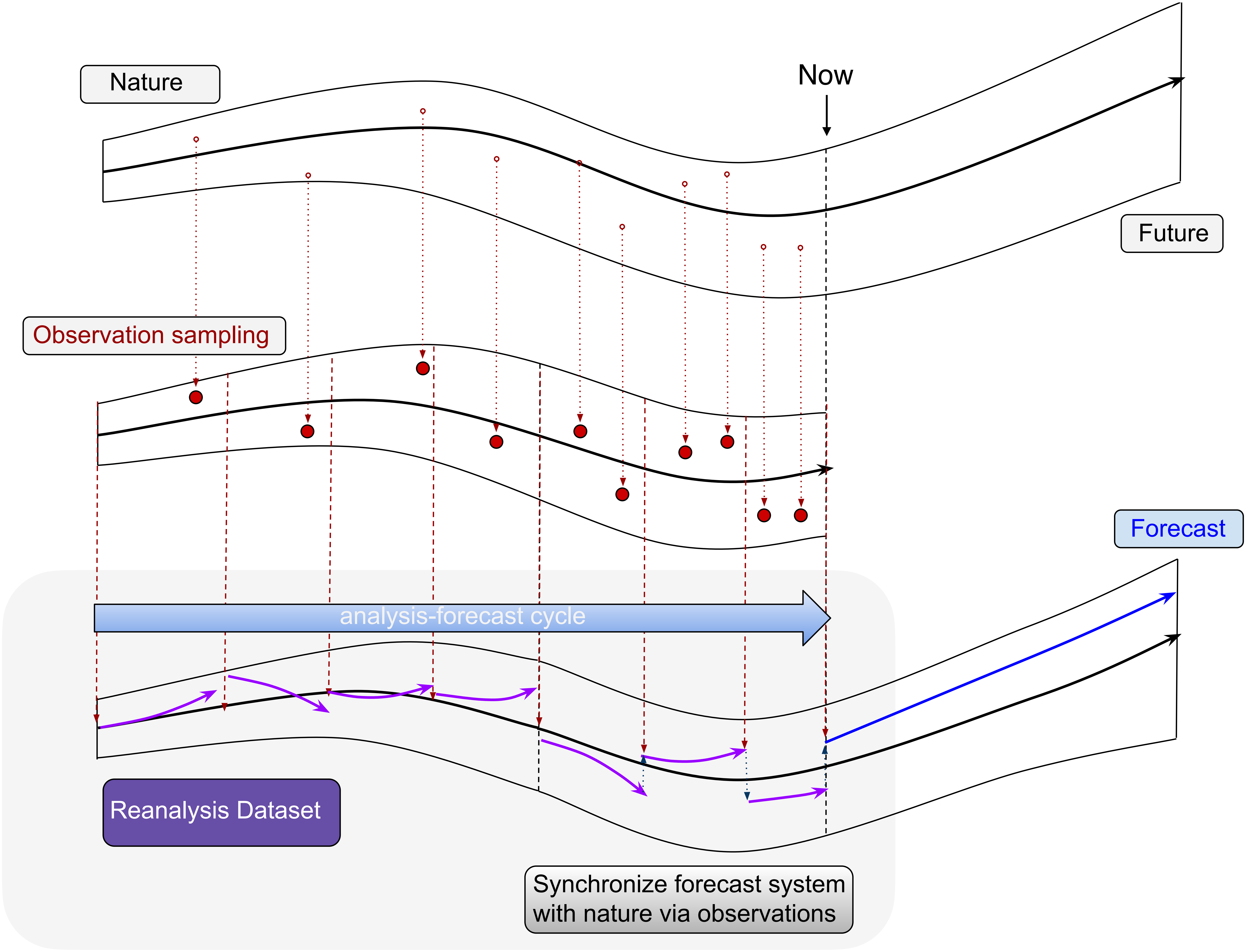


Figure 8: NOAA better quality masks. The loss is only calculated for targets covered by the mask to minimize issues with wrongly labelled tagets.

<u>Challenges</u> -
- Only observed quantities are forecasted, known physical relationships between observed and unobserved processes are not leveraged
- Observations are sparse and noisy, but are treated as ‘ground truth’, observation uncertainty is not characterized
- Dynamic uncertainty is not characterized (i.e. forecasts are ‘probabilistic’, but not dynamically so)


Nature
Now
Future
Observation sampling
Forecast
analysis-forecast cycle
Forecast model
(typically leveraging
physical knowledge)
Synchronize forecast system
with nature via observations



Nature
Now
Future
Observation sampling
Forecast
analysis-forecast cycle
Reanalysis Dataset
Synchronize forecast system with nature via observations

- DA is ‘implicit’ - it is only used to generate the training data.
- There is a chain of efforts aimed at increasing efficiency and performance, to allow increasing resolution in the ML forecast model

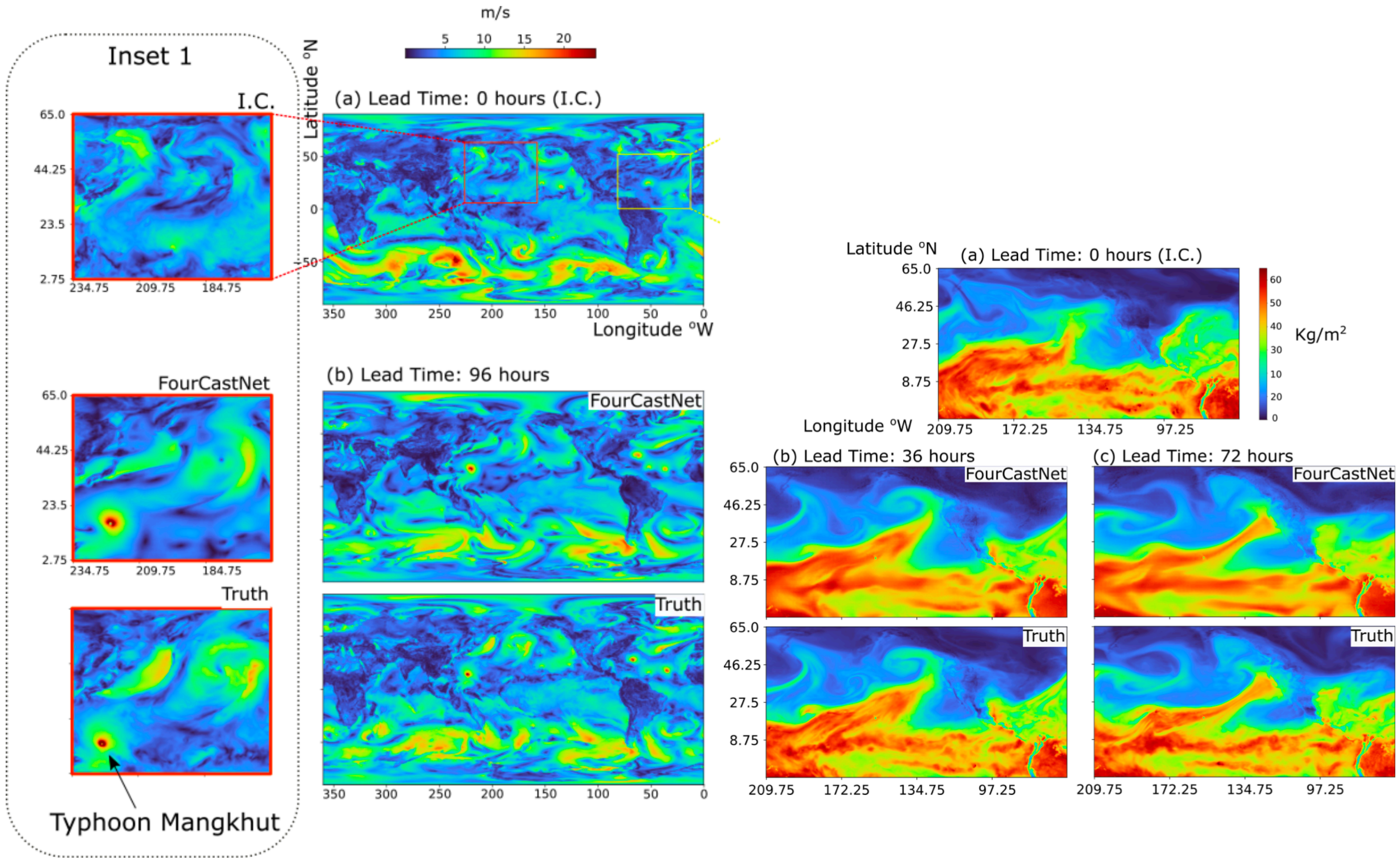
WeatherBench

Leaderboard

Model	Z500 RMSE (3 / 5 days) [m^2/s^2]	T850 RMSE (3 / 5 days) [K]	Notes	Reference
Operational IFS	154 / 334	1.36 / 2.03	ECWMF physical model (10 km)	Rasp et al. 2020
Rasp and Thuerey 2020 (direct/continuous)	**268 / 499**	**1.65 / 2.41**	Resnet with CMIP pretraining (5.625 deg)	Rasp and Thuerey 2020
IFS T63	268 / 463	1.85 / 2.52	Lower resolution physical model (approx. 1.9 deg)	Rasp et al. 2020
Weyn et al. 2020 (iterative)	**373 / 611**	**1.98 / 2.87**	UNet with cube-sphere mapping (2 deg)	Weyn et al. 2020
Clare et al. 2021 (direct)	375 / 627	2.11 / 2.91	Stacked ResNets with probabilistic output (5.625 deg)	Clare et al. 2021
IFS T42	489 / 743	3.09 / 3.83	Lower resolution physical model (approx. 2.8 deg)	Rasp et al. 2020
Weekly climatology	816	3.50	Climatology for each calendar week	Rasp et al. 2020
Persistence	936 / 1033	4.23 / 4.56		Rasp et al. 2020
Climatology	1075	5.51		Rasp et al. 2020

Presentation: Stephan Rasp,
Wednesday 12:10 to 12:40

FourCastNet at 0.25° *(Pathak et al., 2022)*



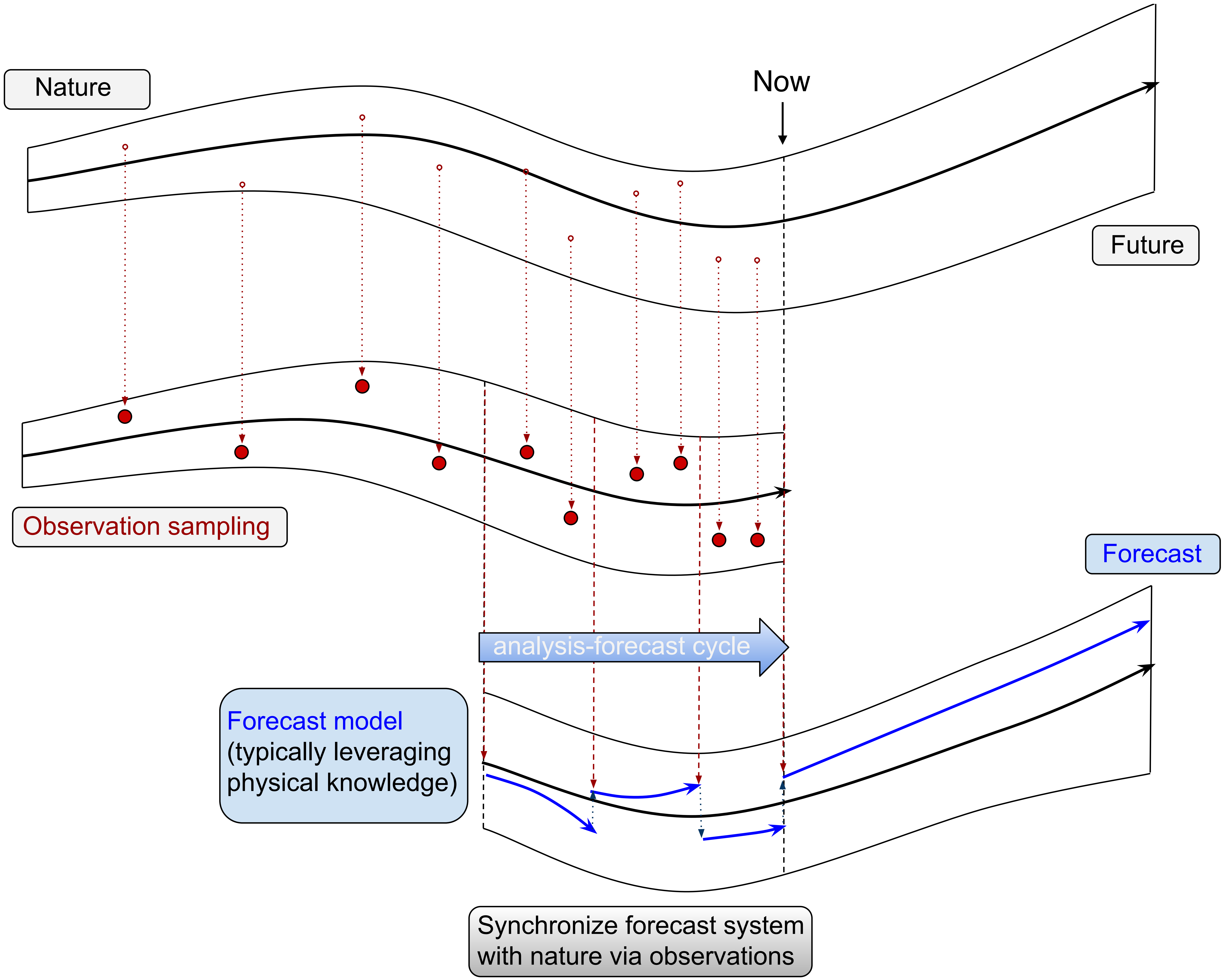
<u>Challenges</u> -

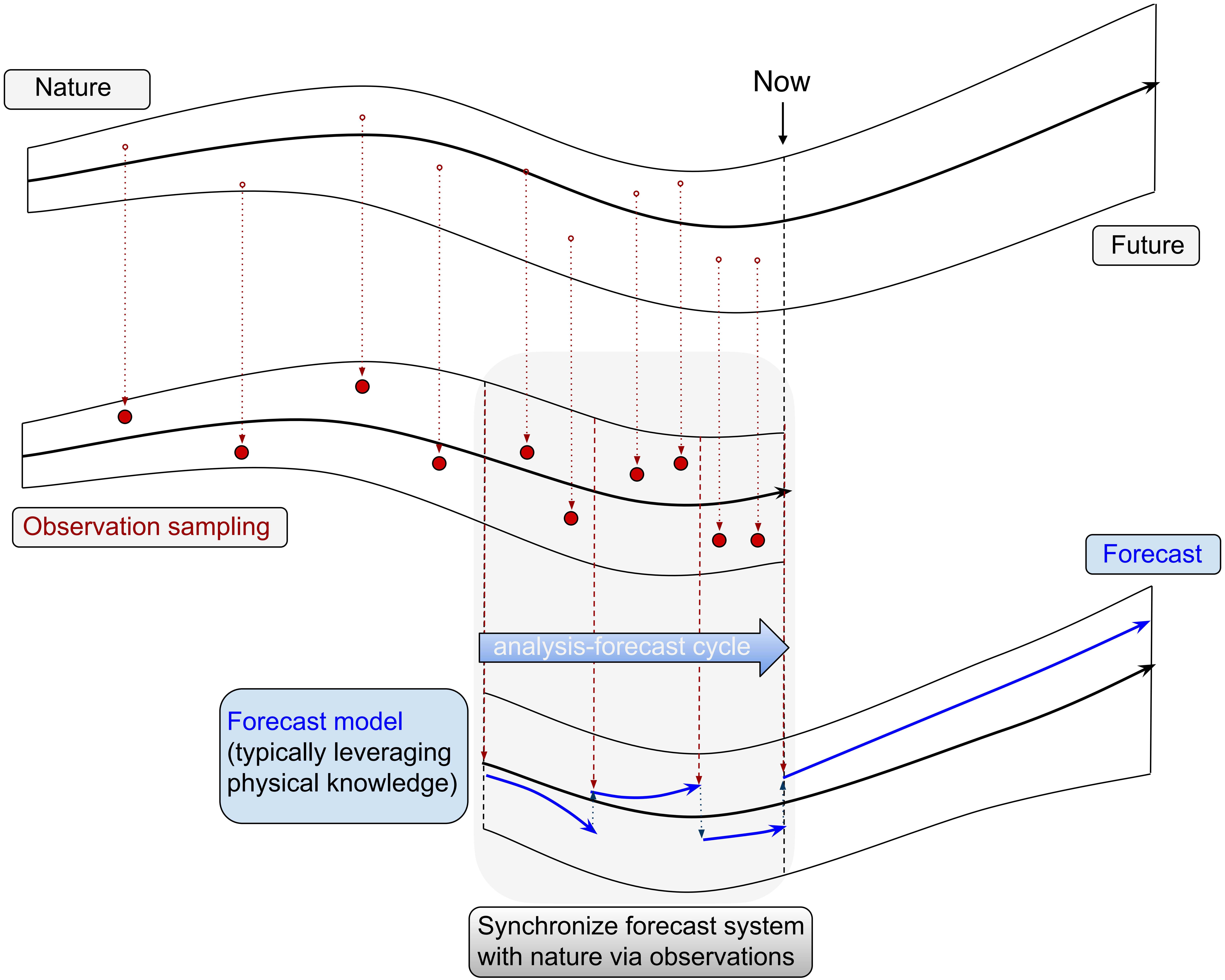
- Even 0.25° is fairly low resolution compared to operational forecast system (e.g. 9km global IFS, or 1-3km TC applications, or 1/25° ocean model run by US Navy). **ML models at these low resolutions still showing significant numerical diffusion**.
- **Limitation using reanalysis as ‘truth’**: While training on reanalysis data may alleviate some needs for model bias correction, it introduces a whole new range of problems (unphysical discontinuities, limits on temporal and spatial resolution, unknown error characteristics)

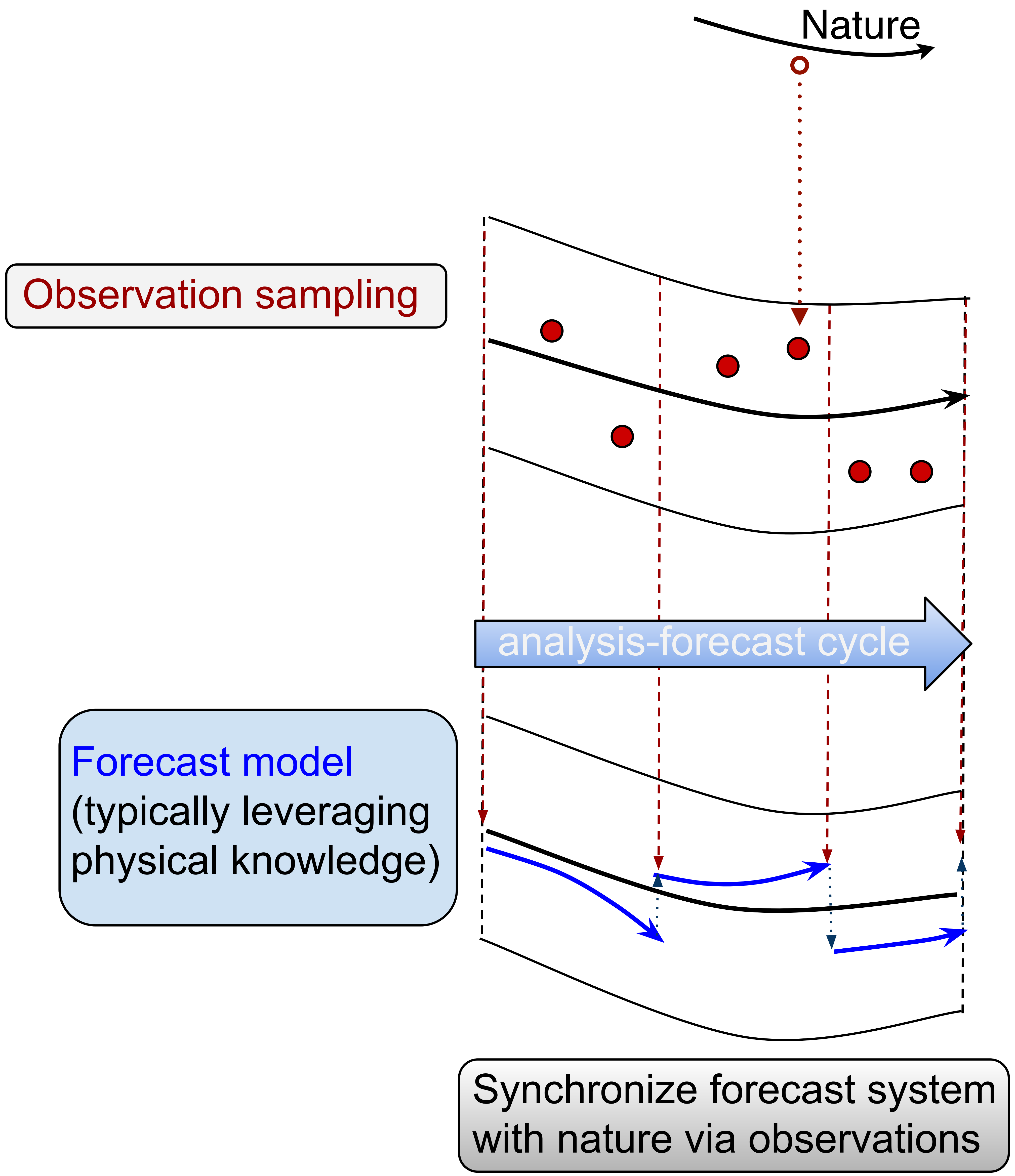
If we no longer pretend that observations and reanalyses are perfect
(i.e. we begin to think about the problem from a data assimilation perspective),
then…

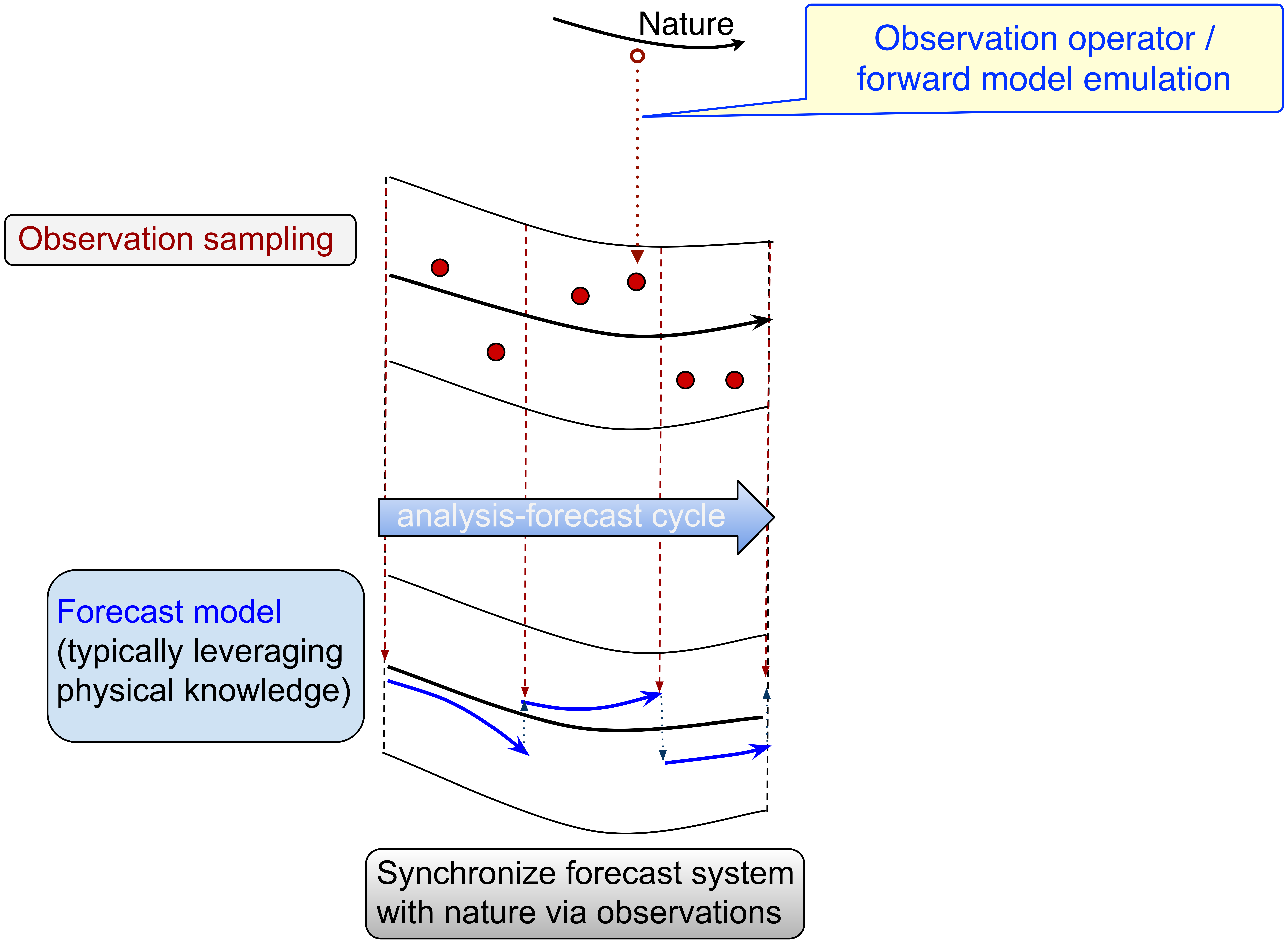
Challenge:
Fundamental questions -

Are reanalysis datasets an adequate source of training data for ML?
Or, are pure simulation datasets more effective?
How then will biases & systematic errors be handled?
Do we need these at all - or can we learn directly from observations plus basic physics constraints?

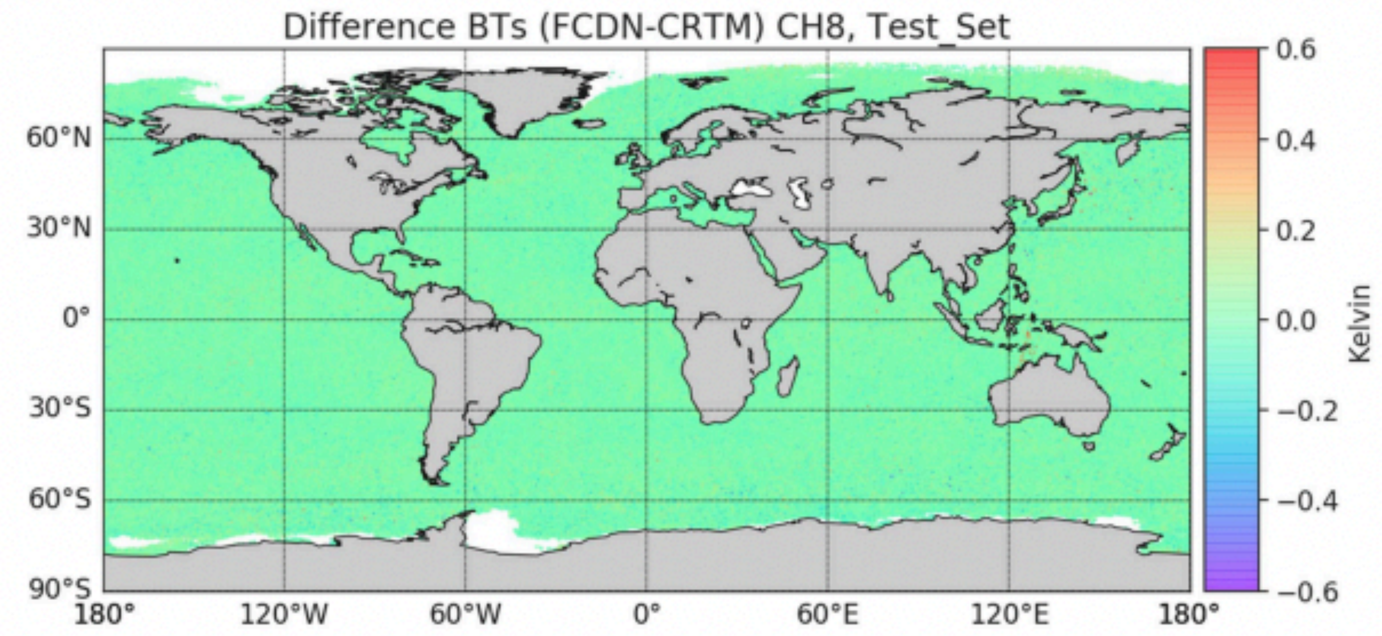

Nature
Now
Future
Observation sampling
Forecast
analysis-forecast cycle
Forecast model
(typically leveraging
physical knowledge)
Synchronize forecast system
with nature via observations

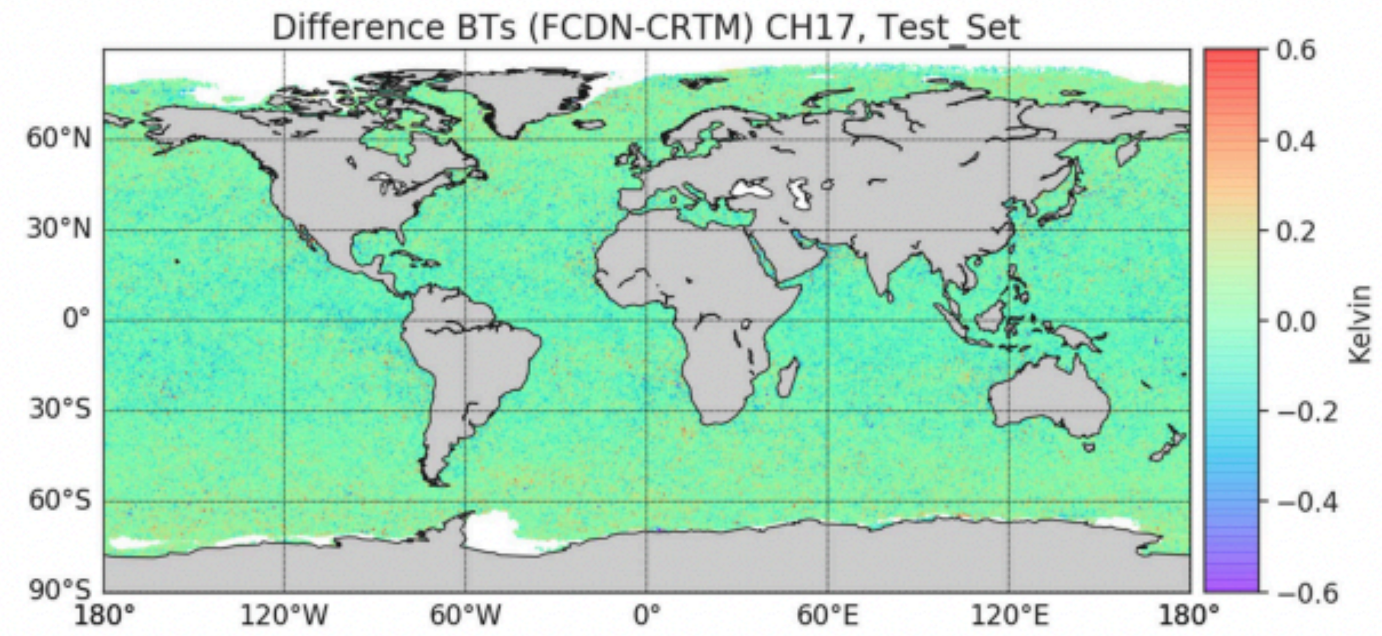

Nature
Now
Future
Observation sampling
Forecast
analysis-forecast cycle
Forecast model
(typically leveraging
physical knowledge)
Synchronize forecast system
with nature via observations


Nature
Observation sampling
analysis-forecast cycle
Forecast model
(typically leveraging
physical knowledge)
Synchronize forecast system
with nature via observations


Nature
Observation operator /
forward model emulation
Observation sampling
analysis-forecast cycle
Forecast model
(typically leveraging
physical knowledge)
Synchronize forecast system
with nature via observations

“A Deep-Learning-Based Microwave Radiative Transfer Emulator for DA and Remote Sensing” *Liang et al. (2022)*





Partial derivatives with respect to air temperature profile for band 1, 8, and 22

Partial derivatives with respect to water vapor content profile for band 1, 8, and 22

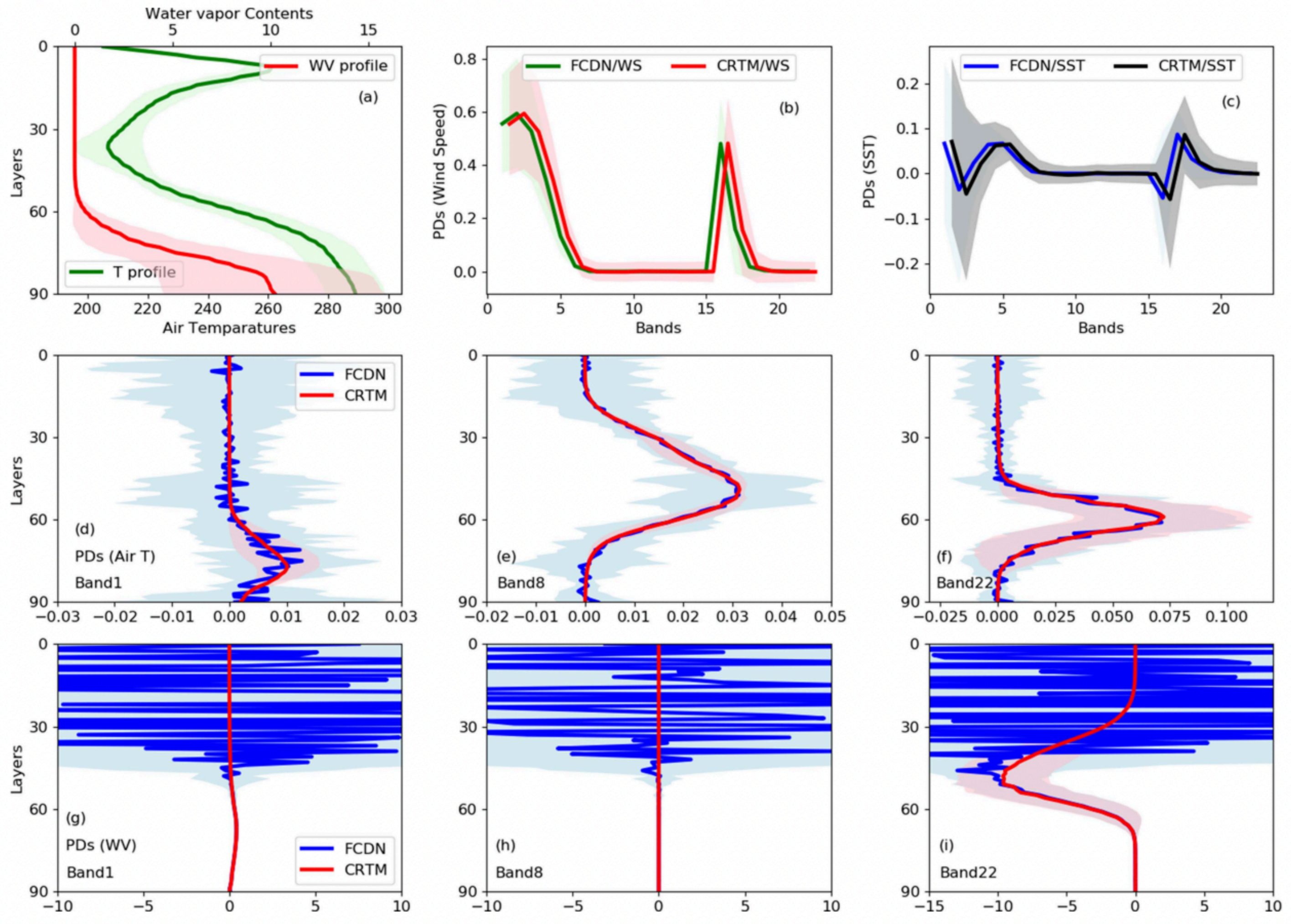


Fig. 11. Summary of the FCDN_CRTM Jacobian compared to CRTM. (a) Profiles of air temperature and water vapor contents. (b) Partial derivatives with respect to wind speed. (c) Partial derivatives with respect to sea surface temperature. (d–f) Partial derivatives with respect to air temperature profile for band 1, 8, and 22. (g–i) Partial derivatives with respect to water vapor content profile for band 1, 8, and 22. The curves are the mean of all samples with $\pm\sigma$ in the filled areas. The curves (b) and (c) are slightly shifted in *x*-axis to clearly distinguish partial derivatives between CRTM and FCDN_CRTM.

<u>Advantages</u>:
Speed

<u>Challenges</u>:
While 1st order estimates appear accurate, accurate estimation of the Jacobians is still challenging, but essential for DA applications

“A deep learning approach to fast radiative transfer” *Stegmann et al. (2022)*

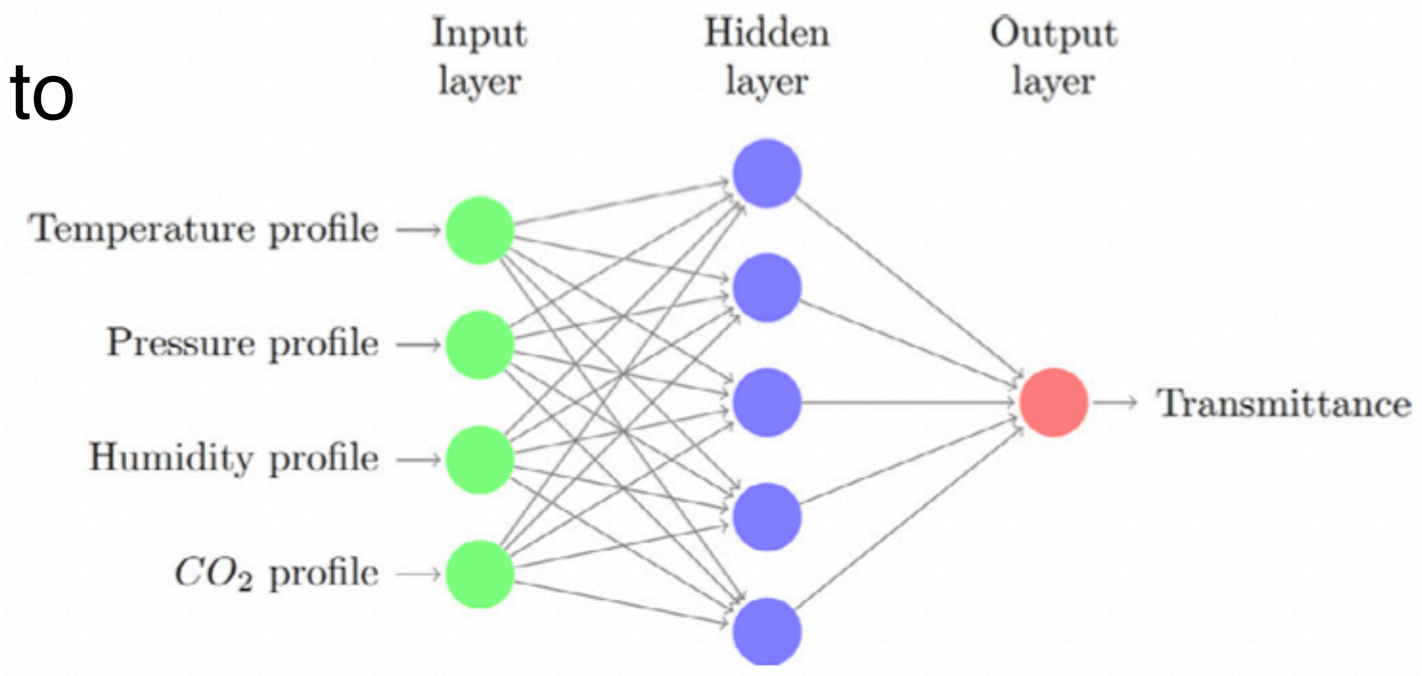


Fig. 3. Sketch of the atmospheric transmittance regression using a hidden layer neural network.

Fig. 1. Flowchart of the transmittance regression approach.

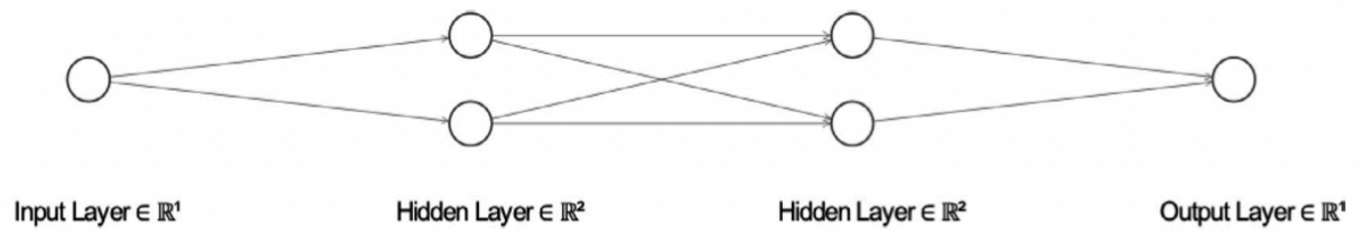


Fig. 2. Illustration of a simple hidden-layer neural network with two hidden layers.

P.G. Stegmann, B. Johnson, I. Moradi et al. *Journal of Quantitative Spectroscopy & Radiative Transfer 280 (2022) 108088*

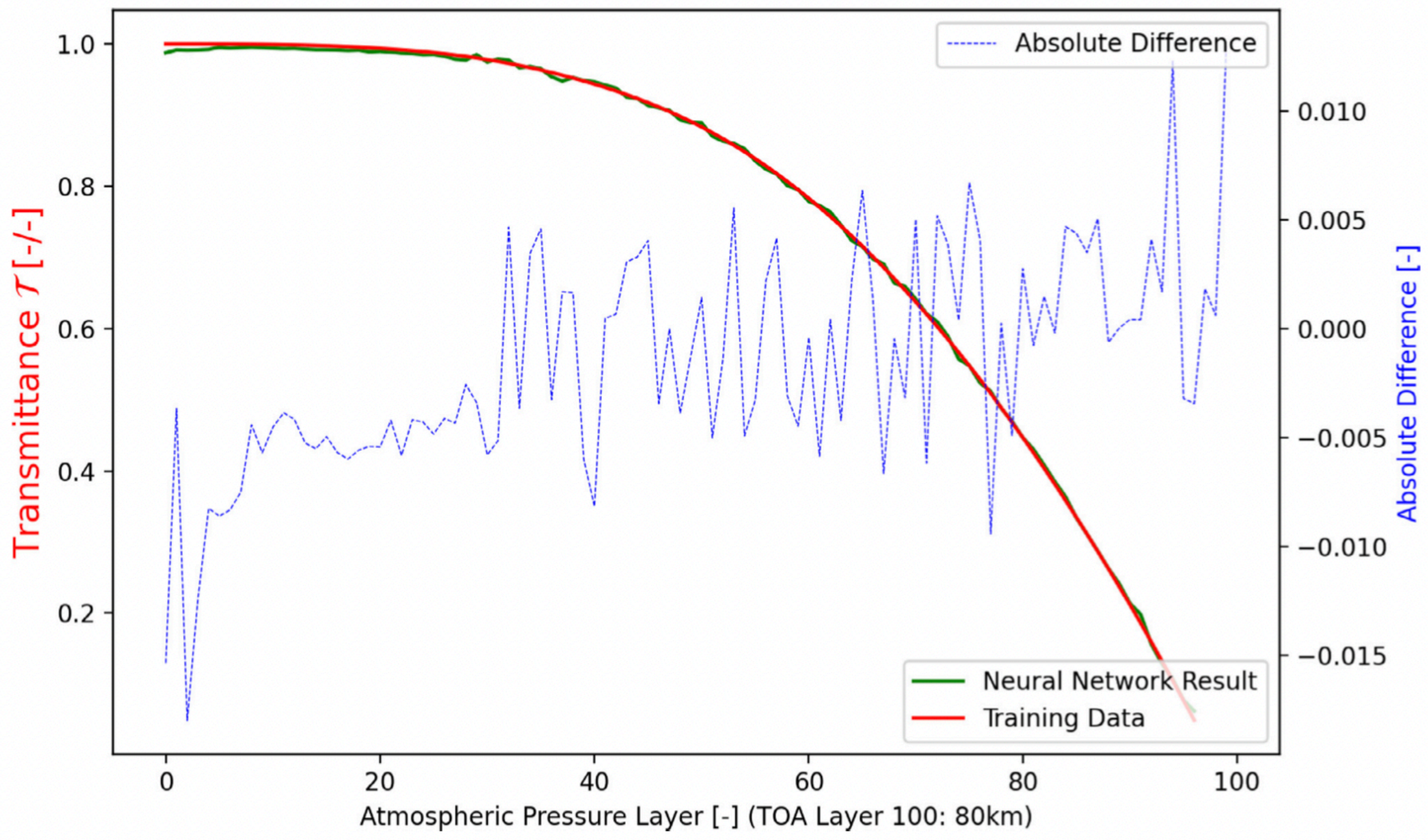
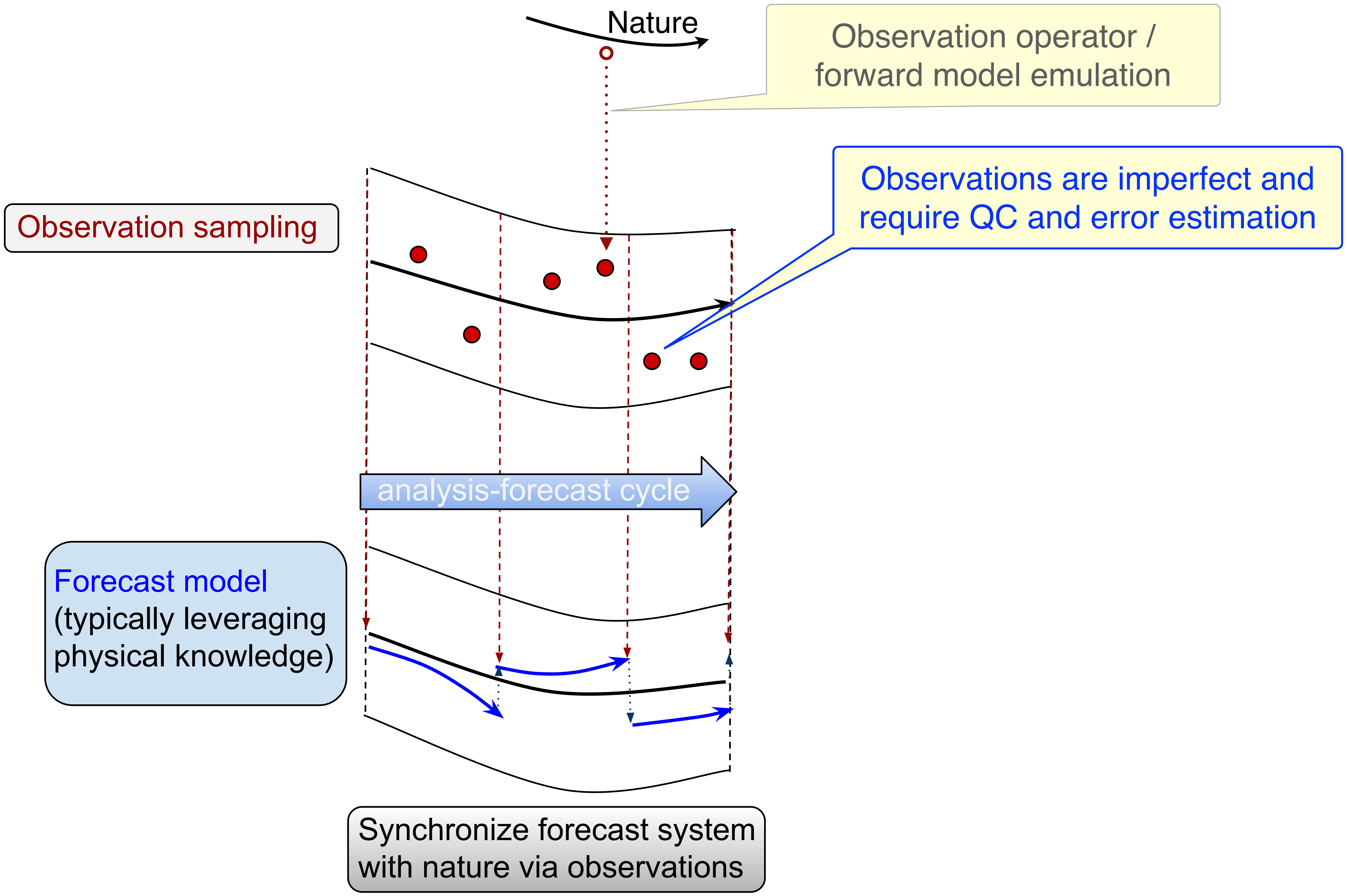
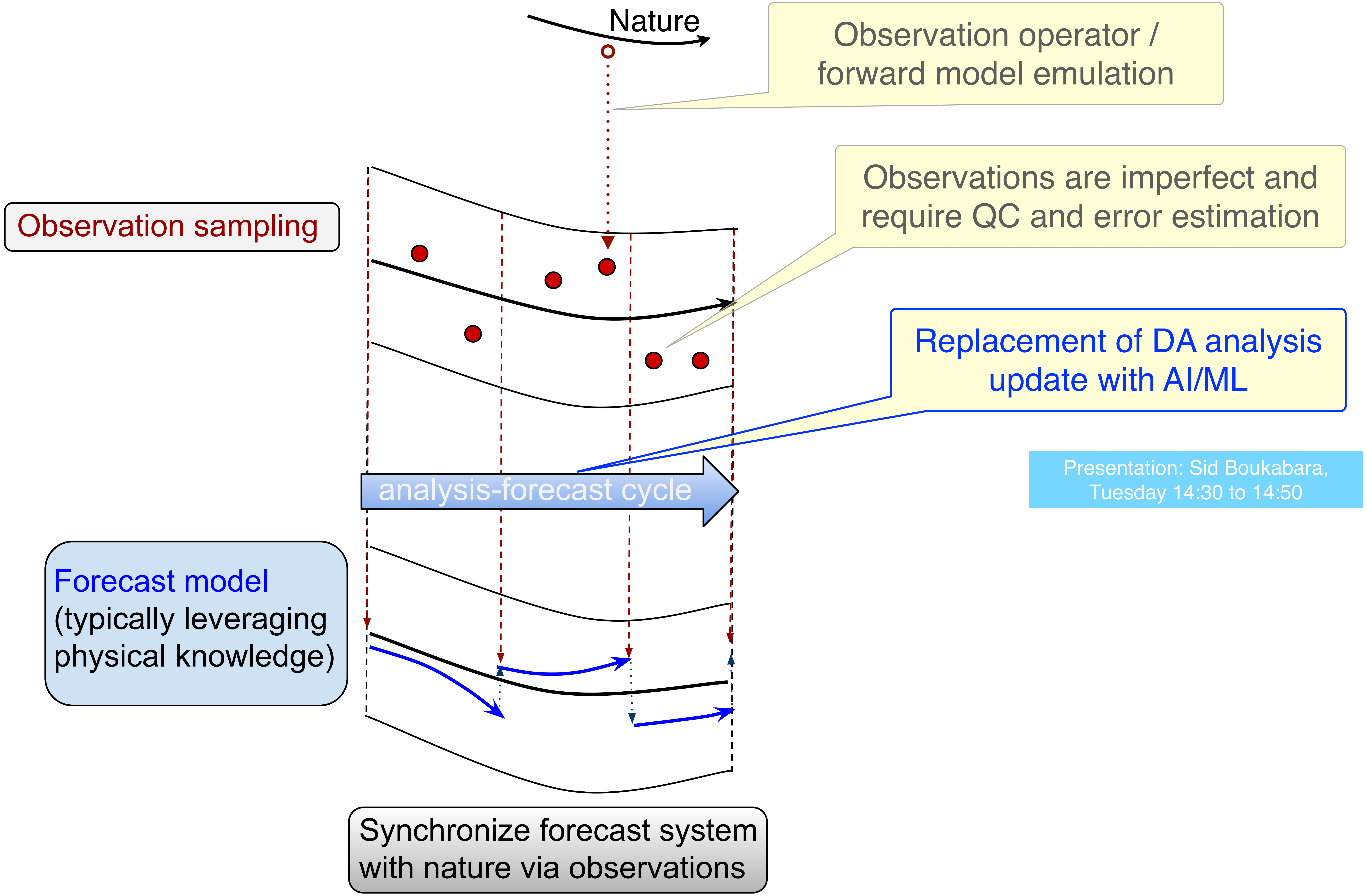
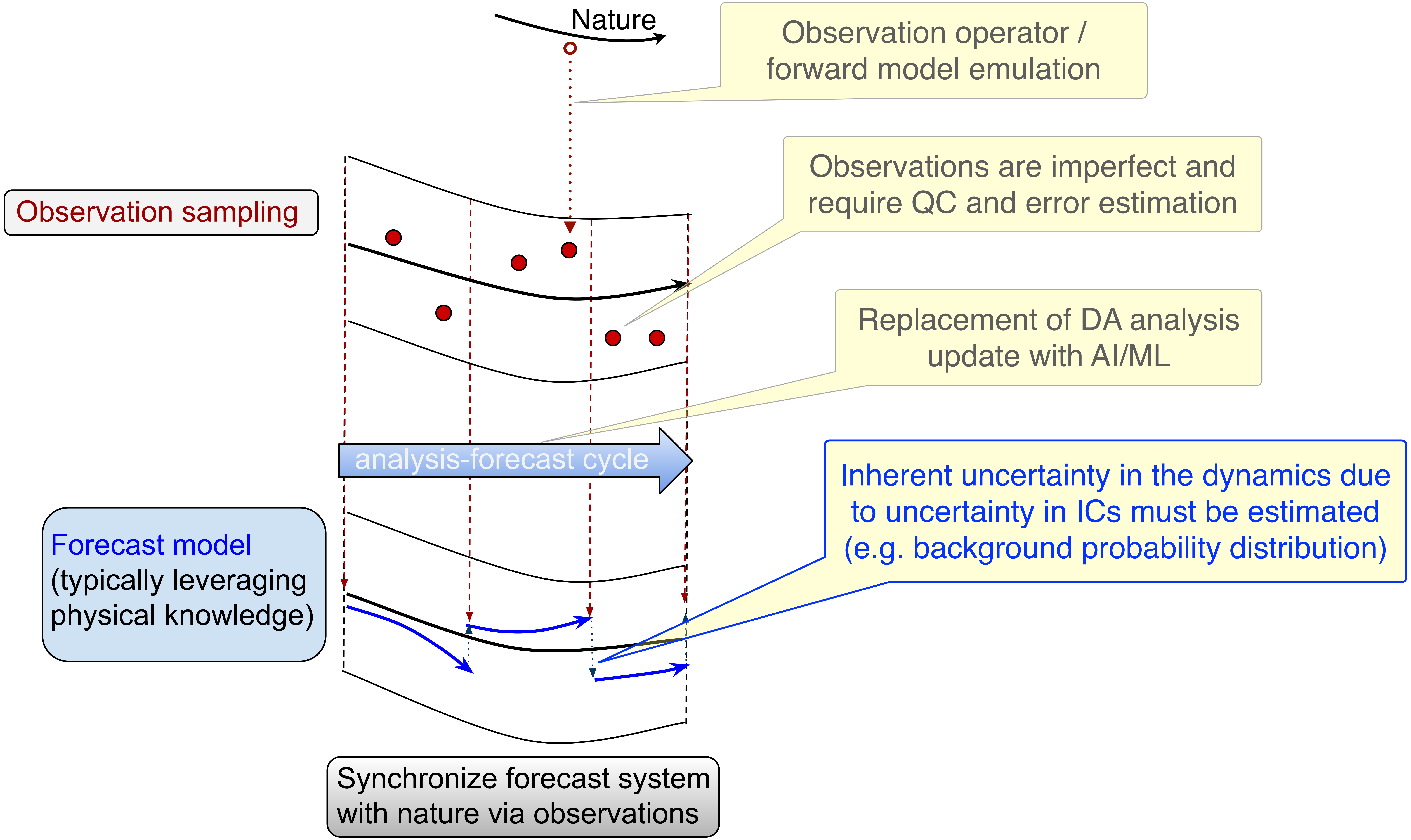


Fig. 16. Normalized total layer transmittance as a function of atmospheric pressure layers for ECMWF83 training profile number 3 and approach 1.


Nature
Observation operator /
forward model emulation
Observations are imperfect and
require QC and error estimation
Observation sampling
analysis-forecast cycle
Forecast model
(typically leveraging
physical knowledge)
Synchronize forecast system
with nature via observations

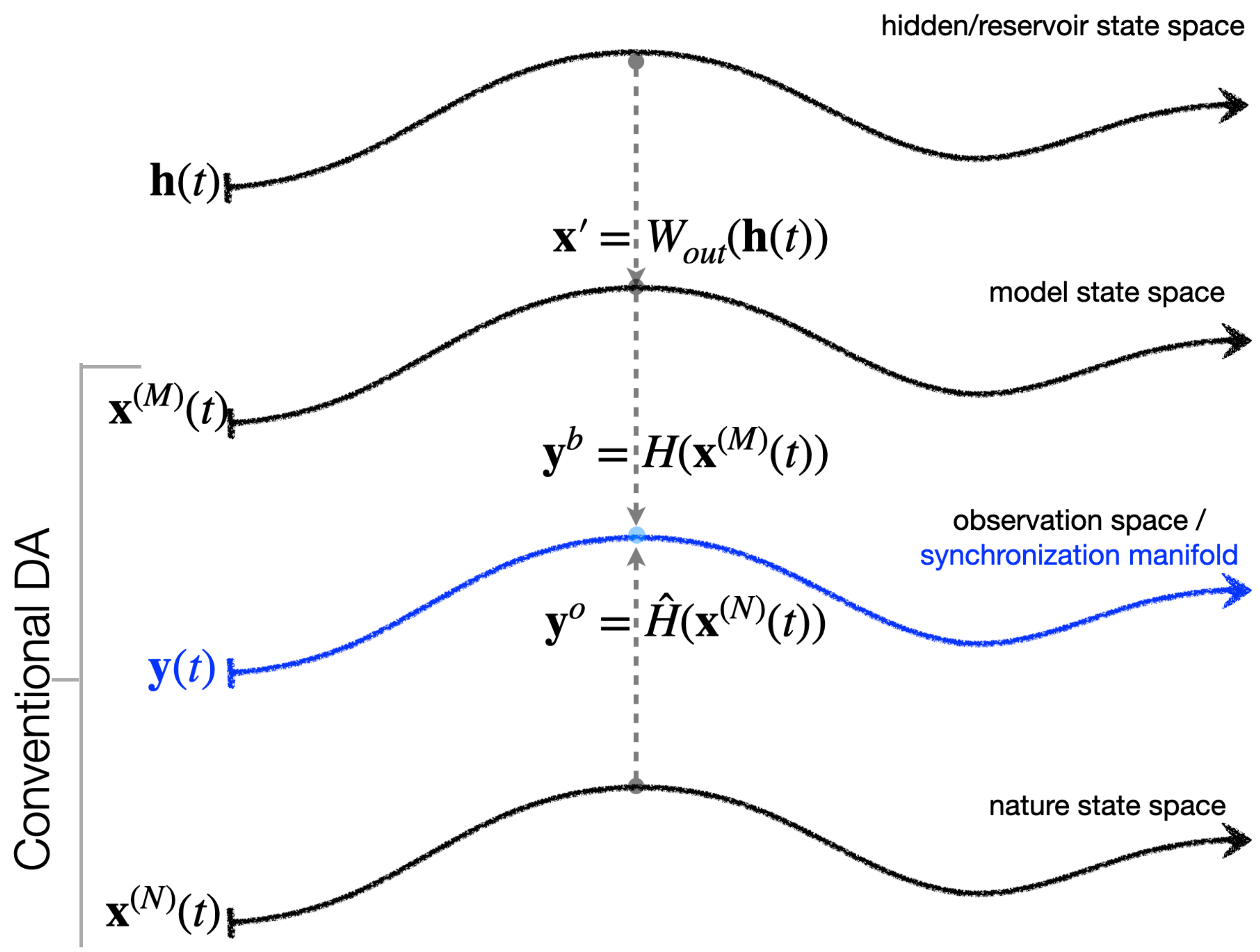


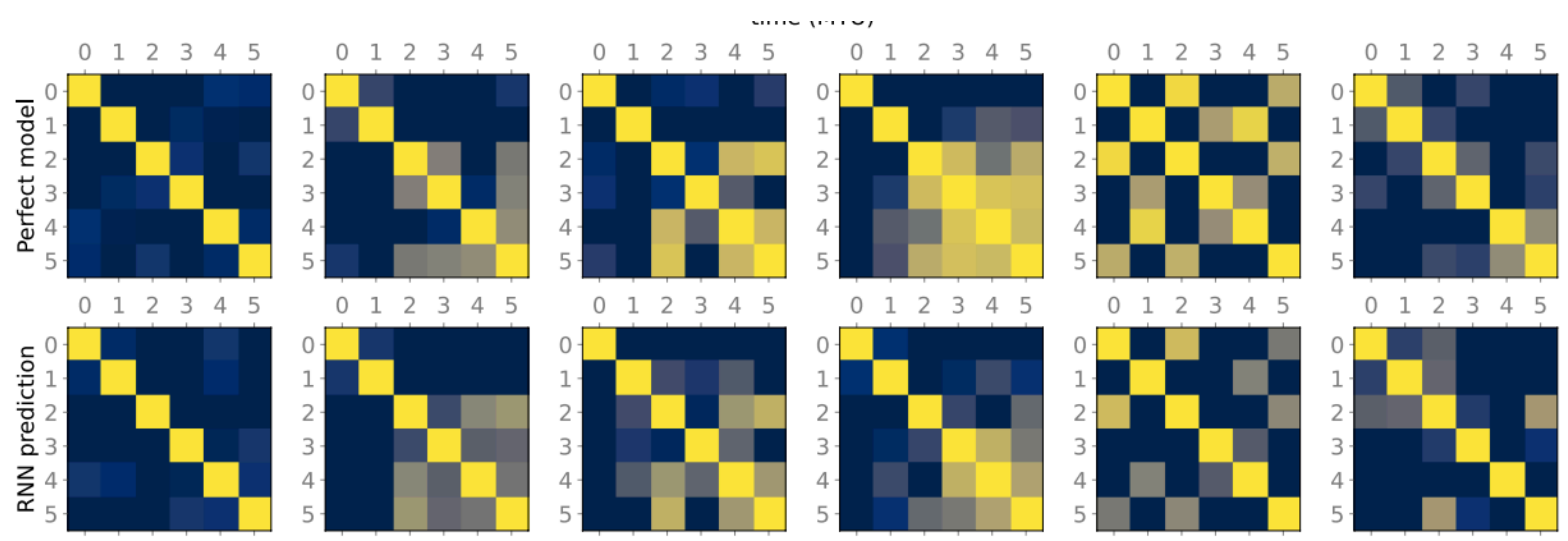
Presentation: Sid Boukabara,
Tuesday 14:30 to 14:50

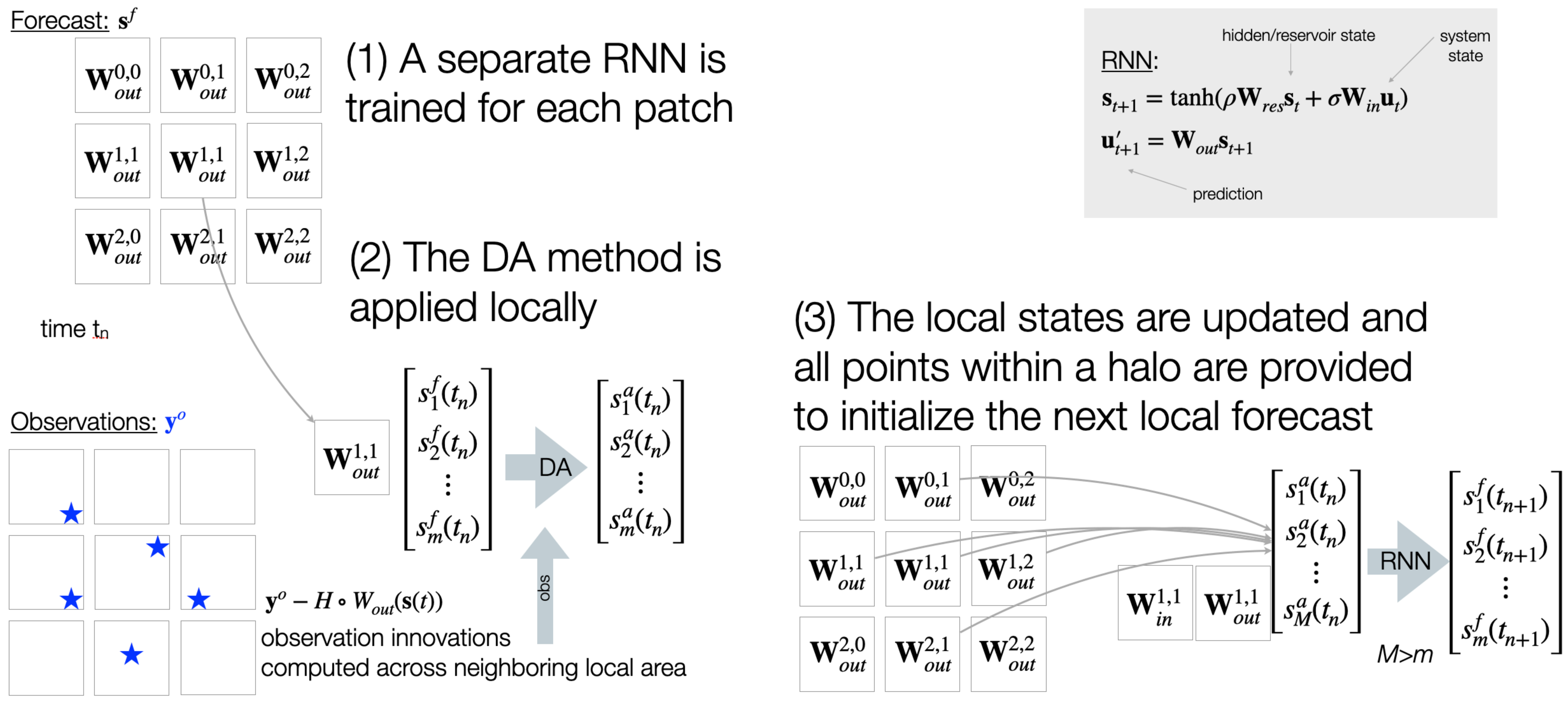


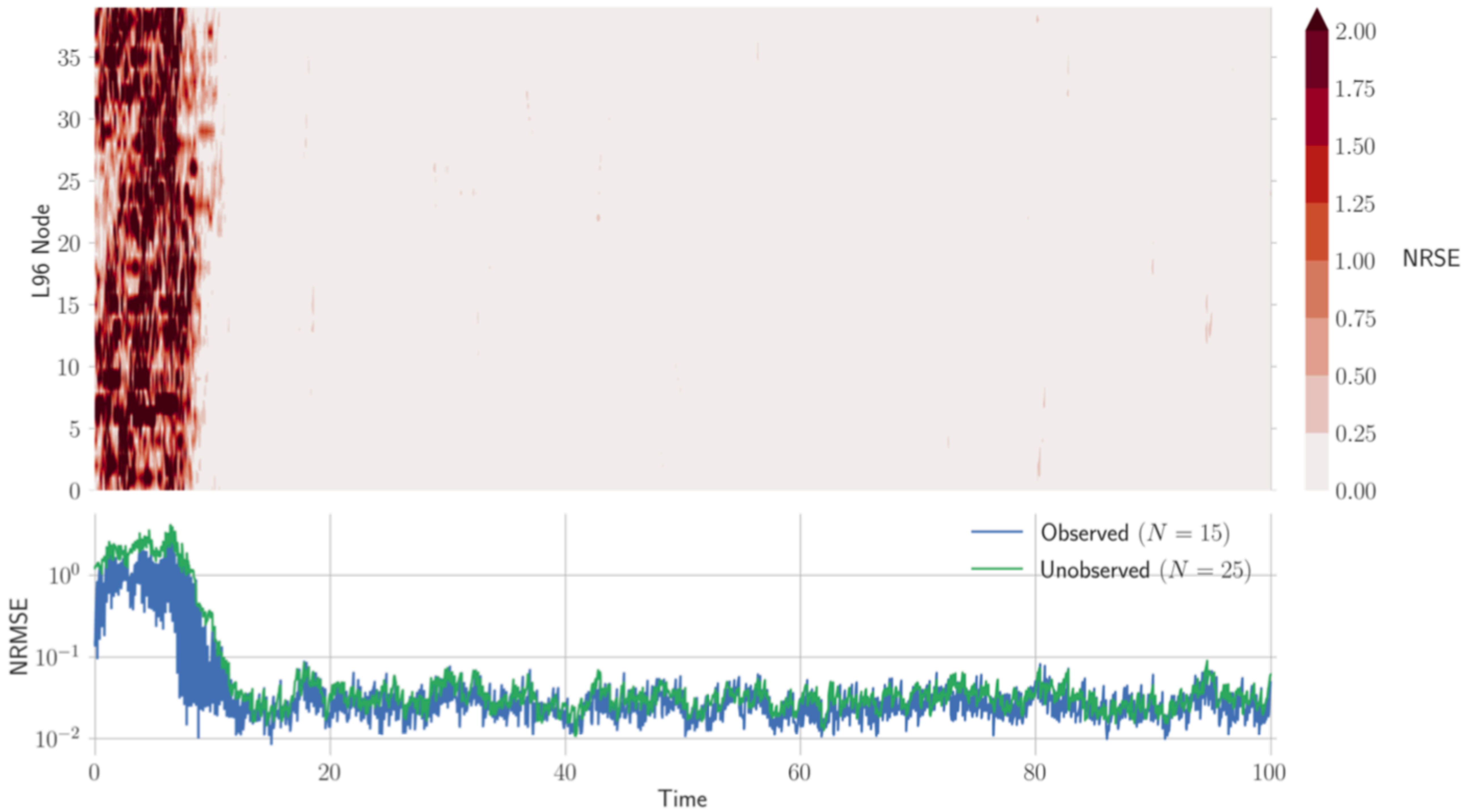
Nature
Observation operator / forward model emulation
Observations are imperfect and require QC and error estimation
Observation sampling
Replacement of DA analysis update with AI/ML
analysis-forecast cycle
Inherent uncertainty in the dynamics due to uncertainty in ICs must be estimated (e.g. background probability distribution)
Forecast model
(typically leveraging physical knowledge)
Synchronize forecast system with nature via observations

Penny et al. (2022), integration of RNN (RC) surrogate models with conventional DA methods

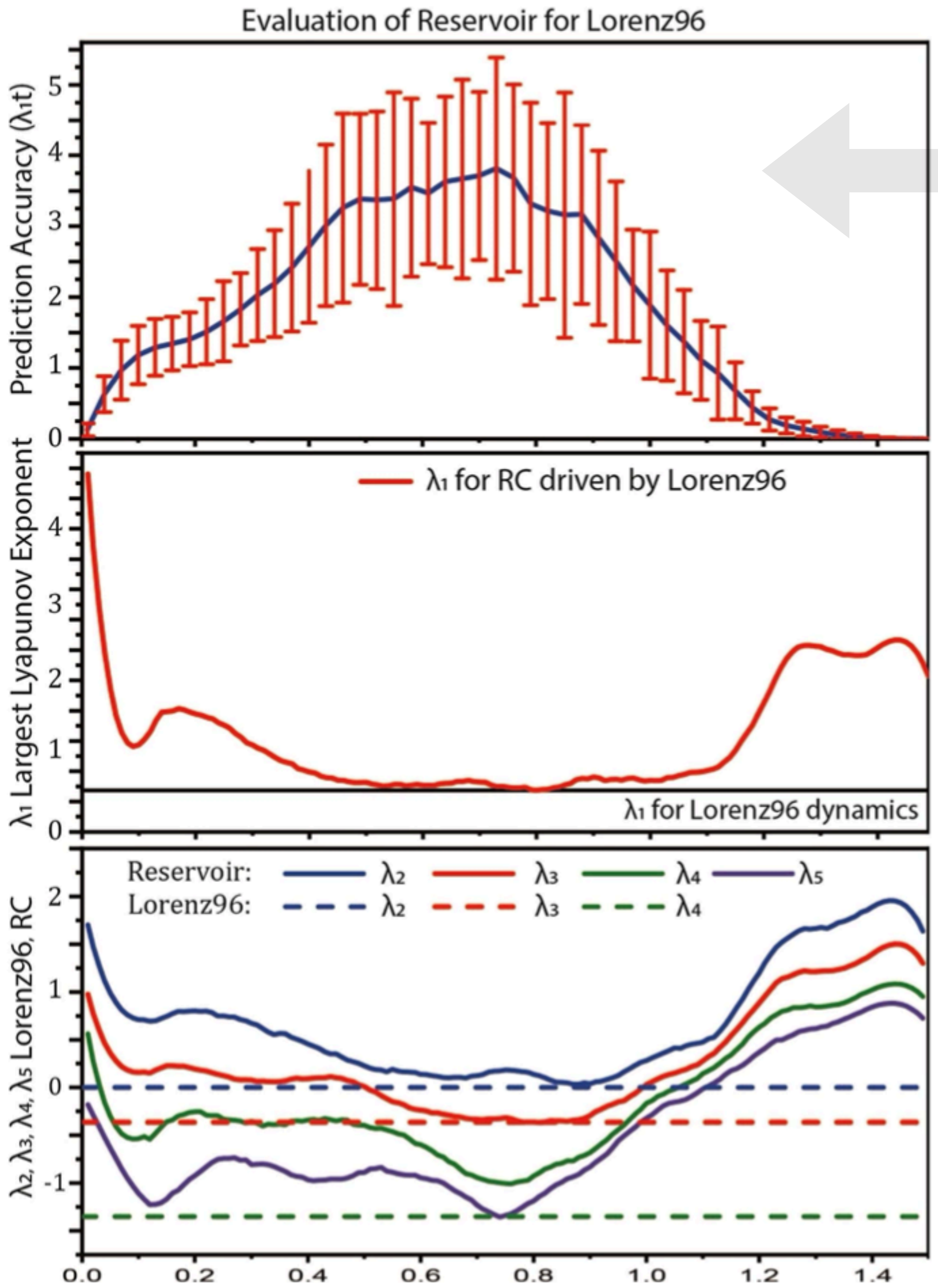






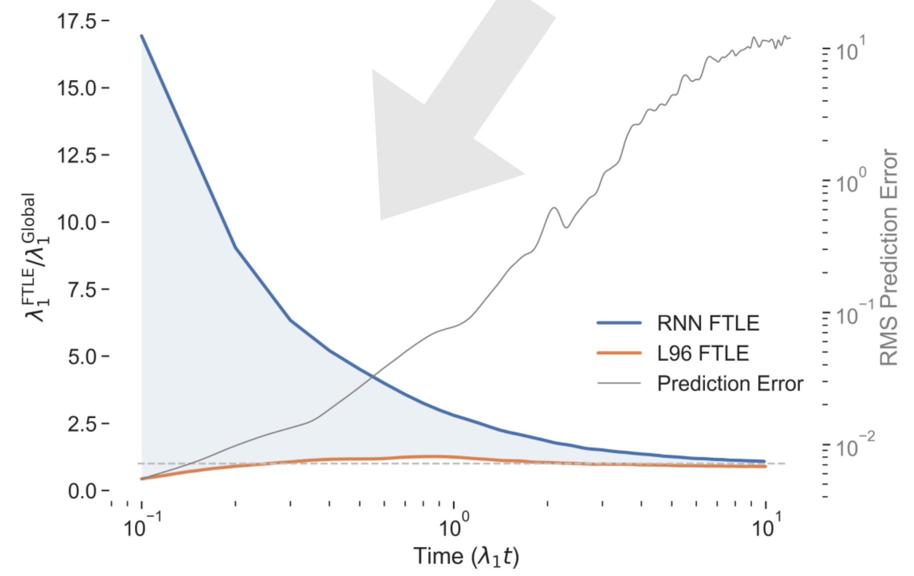


Penny et al. (2022) https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021MS002843

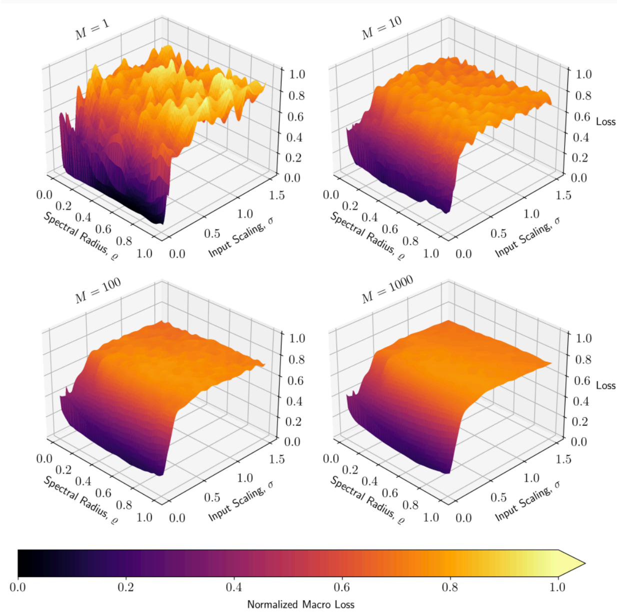


Recovering the Lyapunov spectrum produces better forecasts

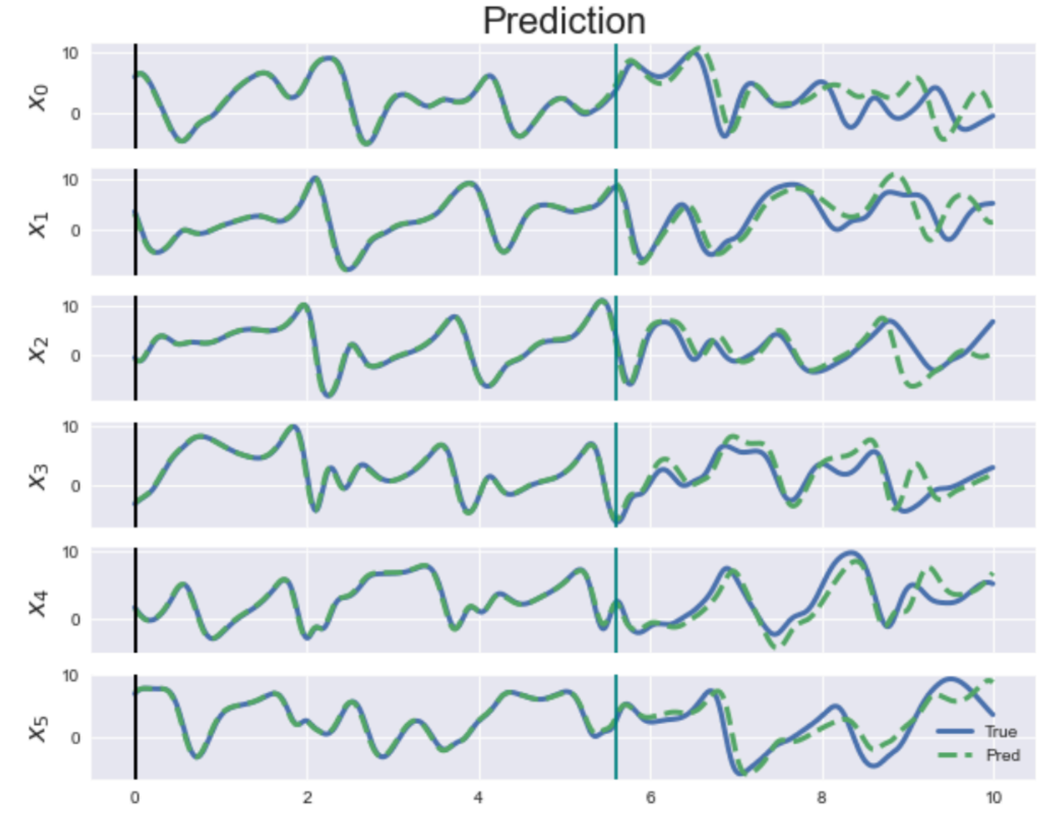
Recovering the Finite Time Lyapunov Exponents produces better error covariance statistics



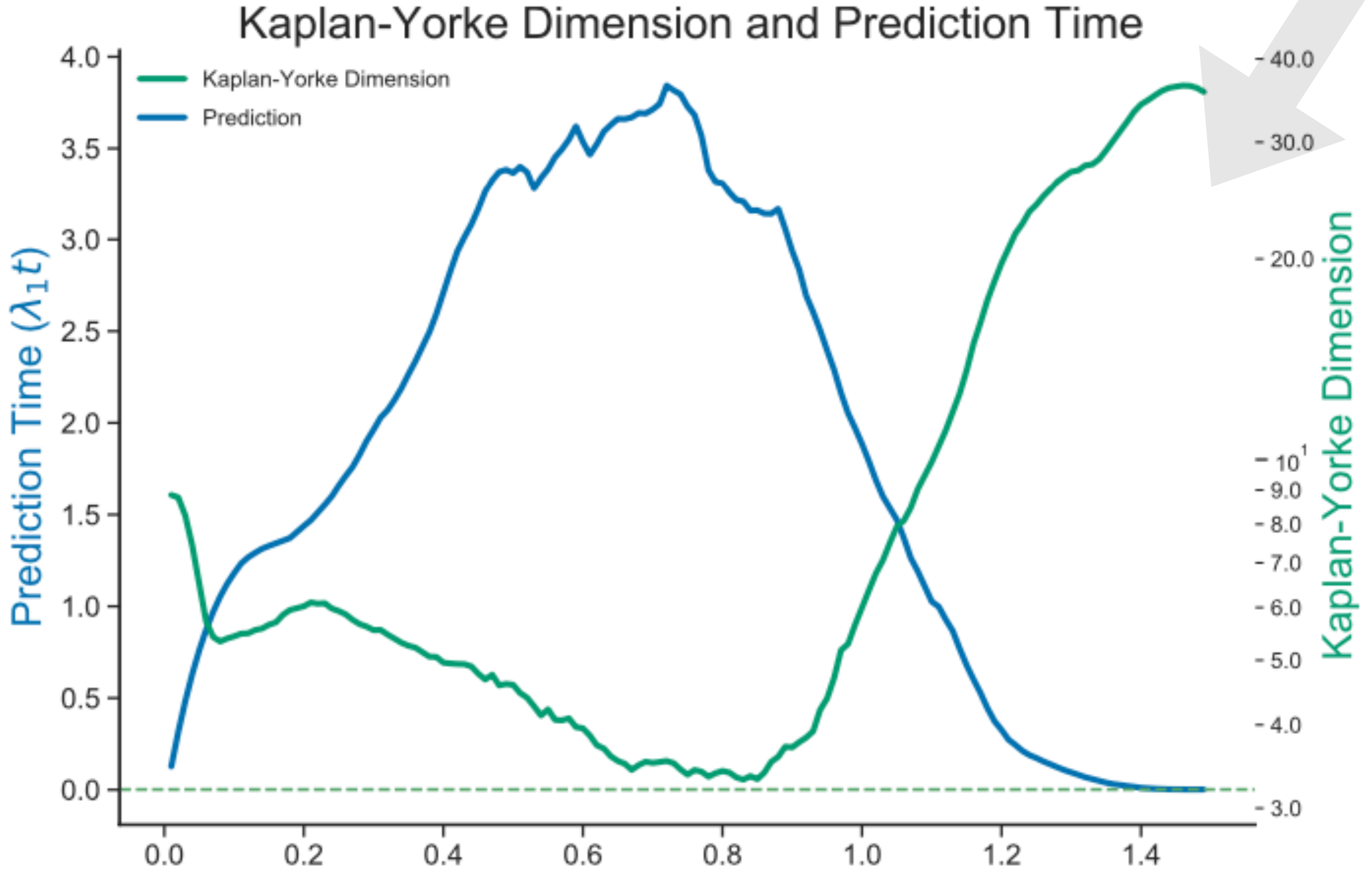
Bayesian optimization of macro-scale parameters improves consistency



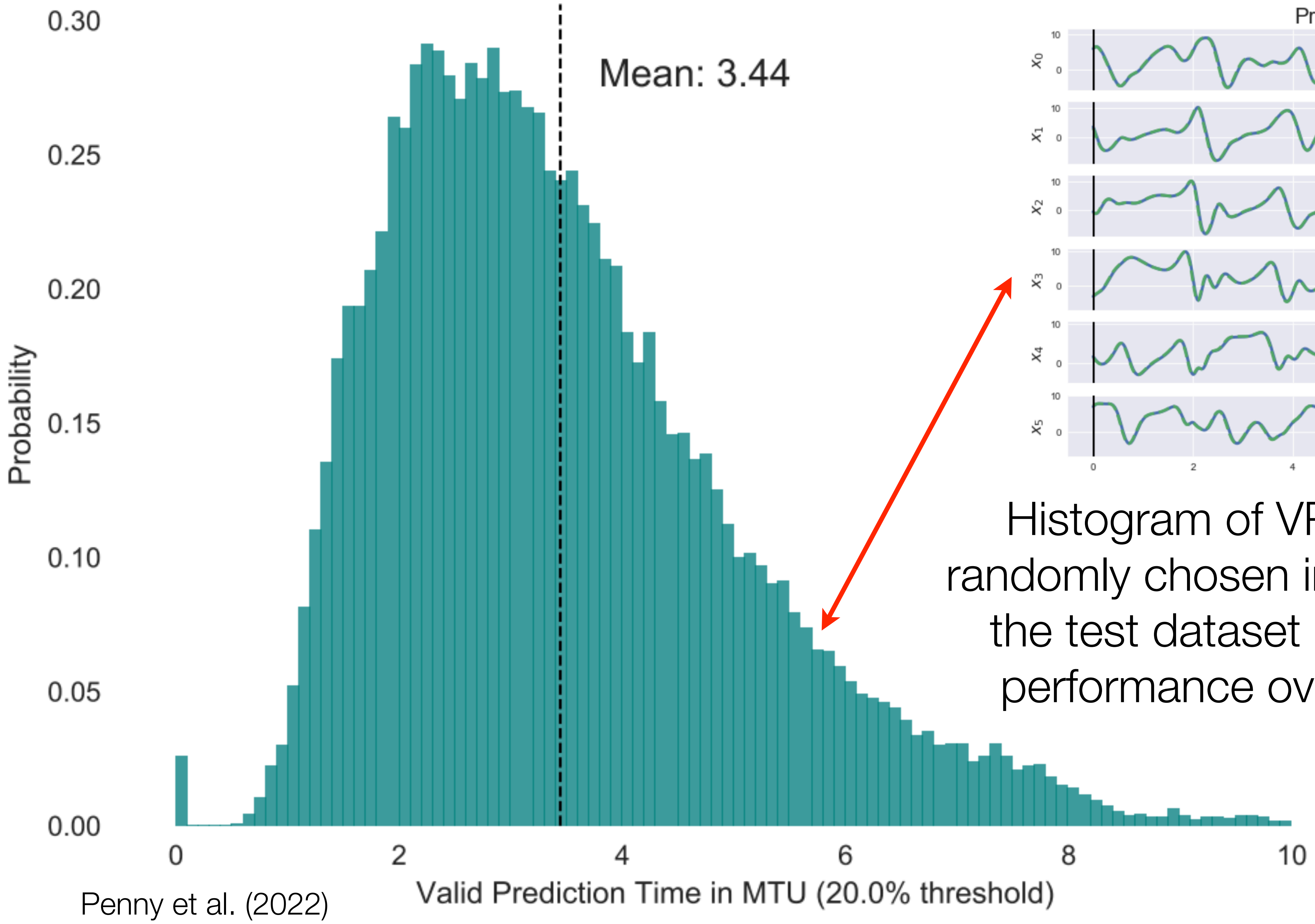
Example single forecast with a **'valid prediction time' (VPT)** marked



Reproducing the Kaplan-Yorke dimension produces better forecasts



Platt et al. (2021)



Histogram of VPT for 100,000 randomly chosen initial conditions in the test dataset shows range of performance over the attractor

Penny et al. (2022)

Neural Networks 153 (2022) 530–552

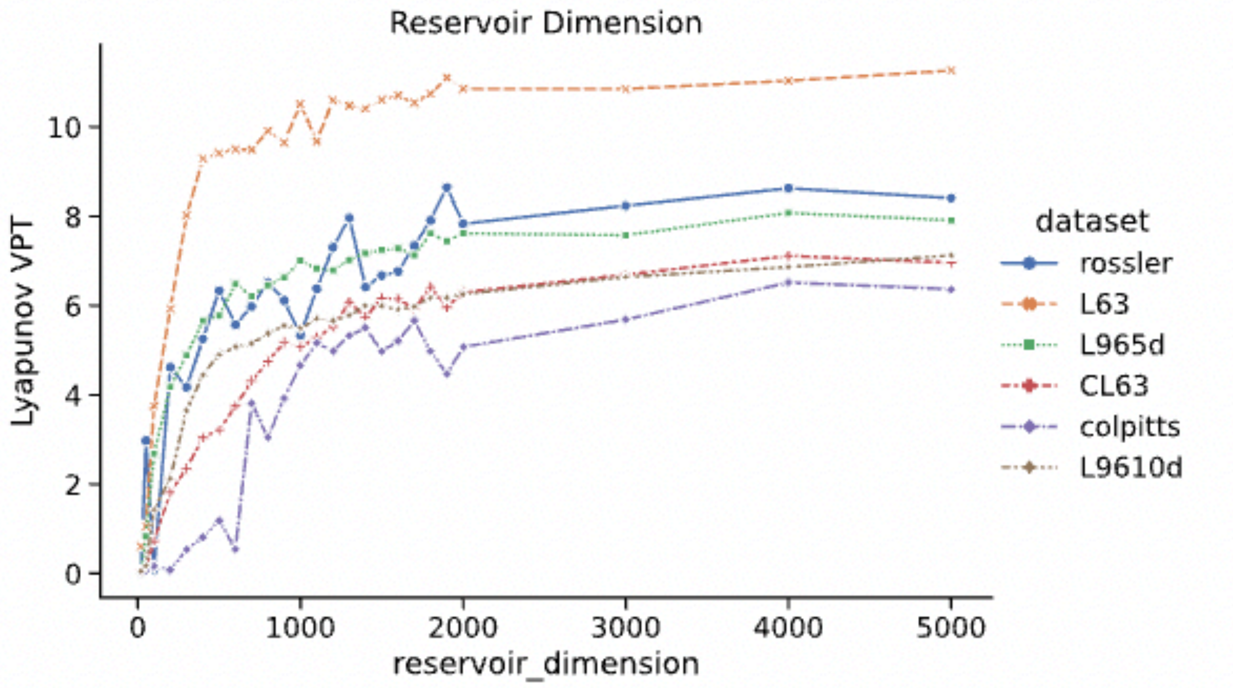
Contents lists available at ScienceDirect

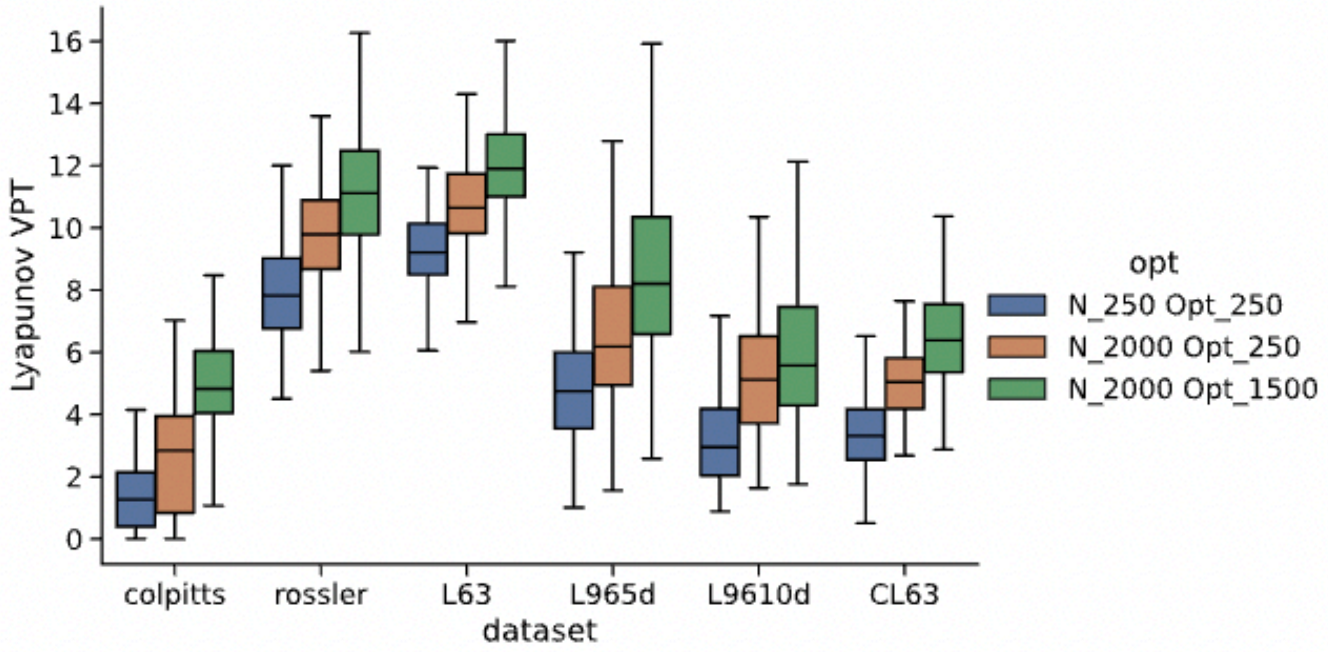
Neural Networks

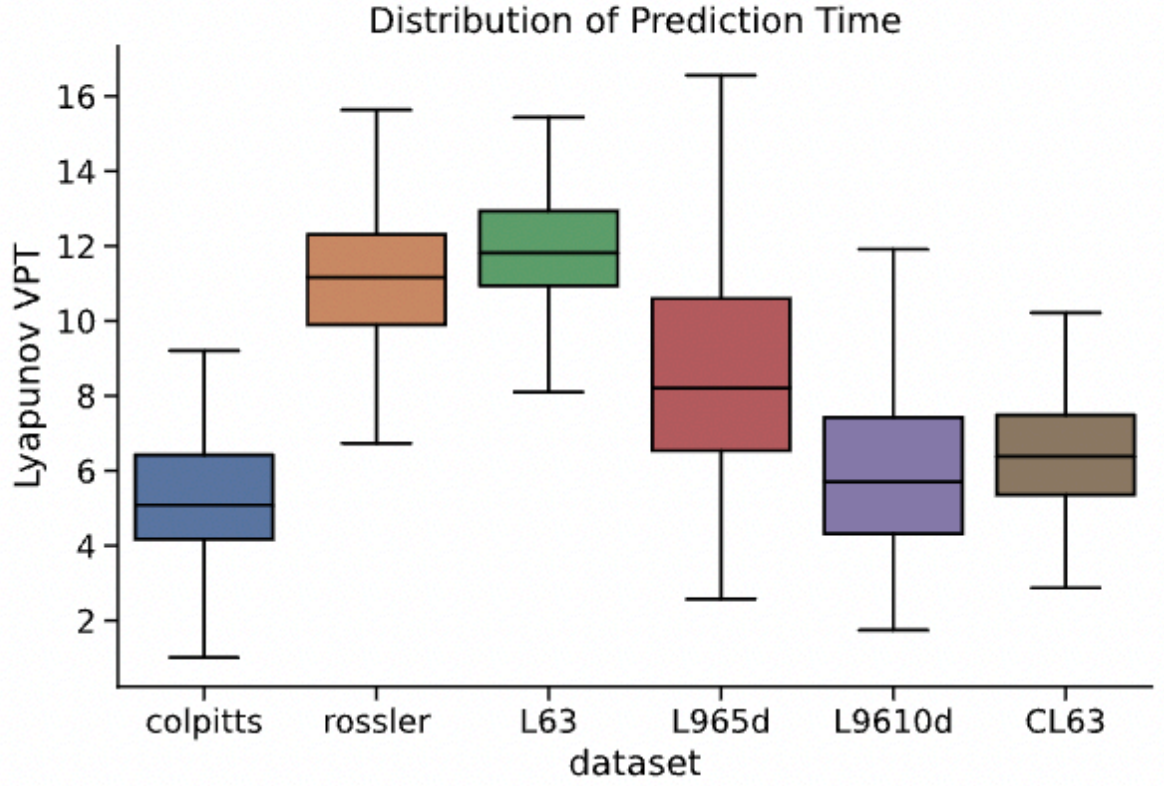
journal homepage: www.elsevier.com/locate/neunet

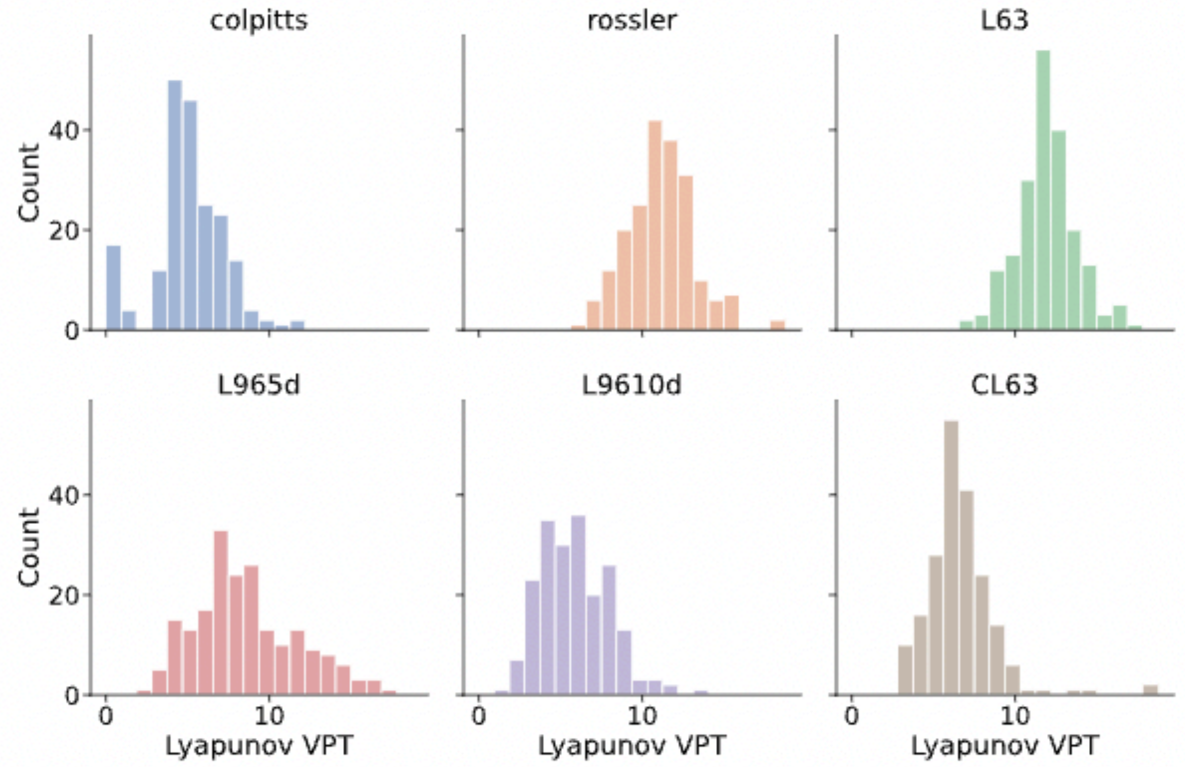
A systematic exploration of reservoir computing for forecasting complex spatiotemporal dynamics

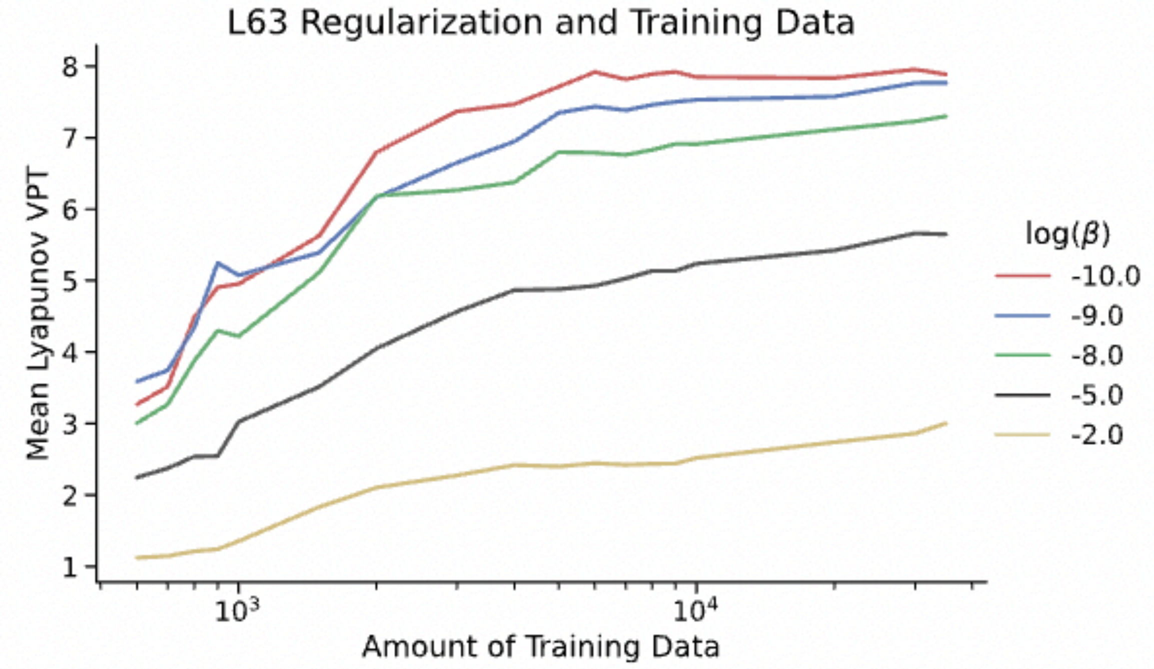
Jason A. Platt [a,*], Stephen G. Penny [b,c], Timothy A. Smith [b,c], Tse-Chun Chen [b,c], Henry D.I. Abarbanel [a,d]

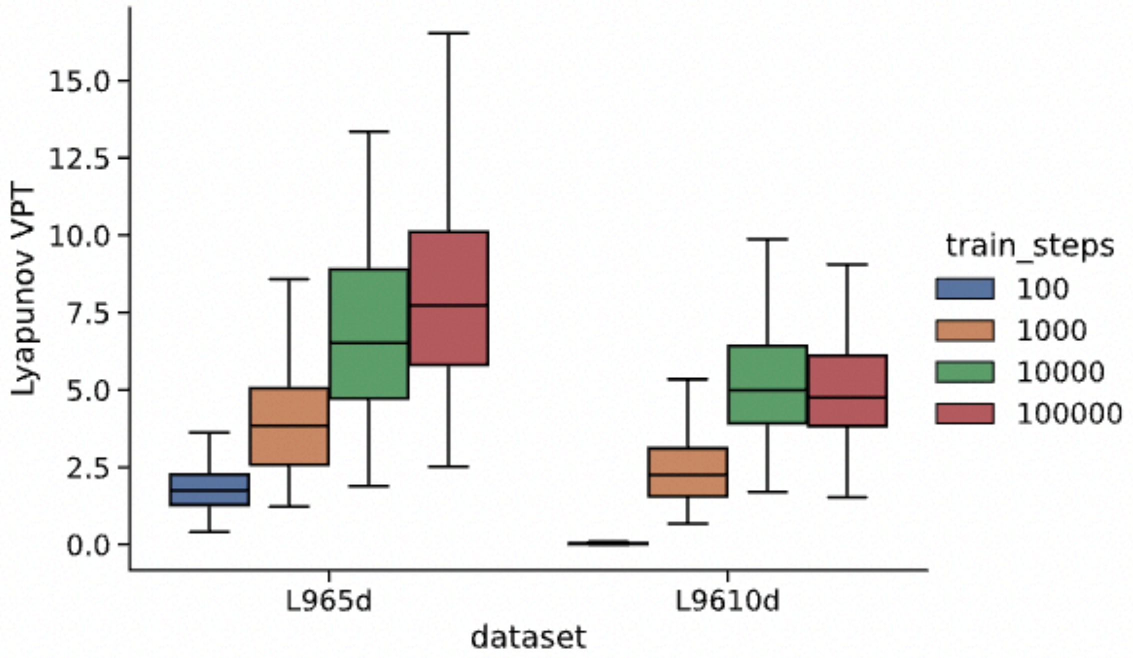


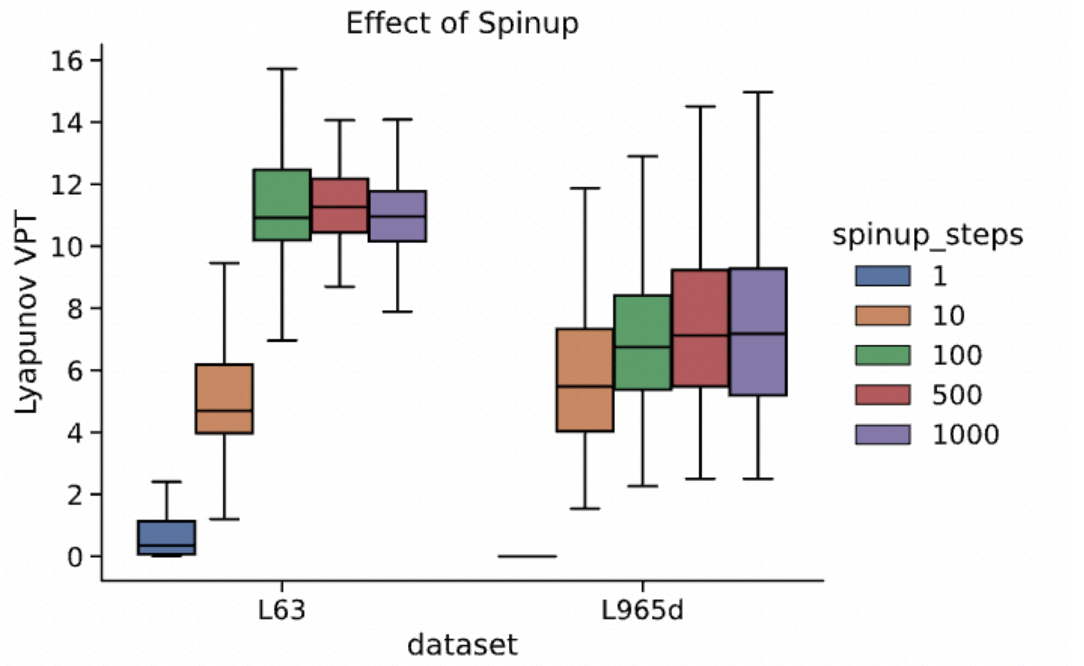


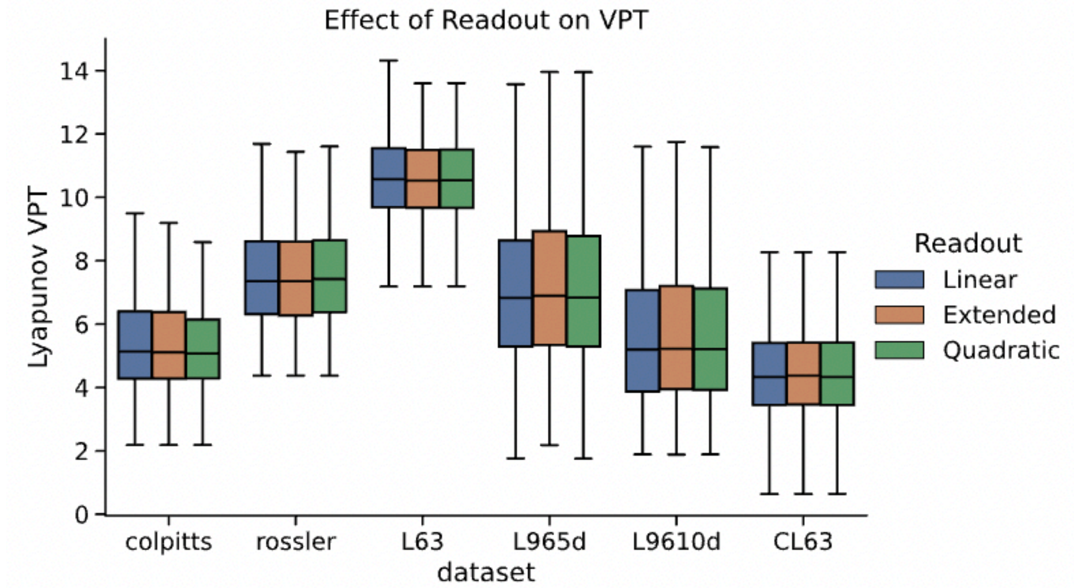


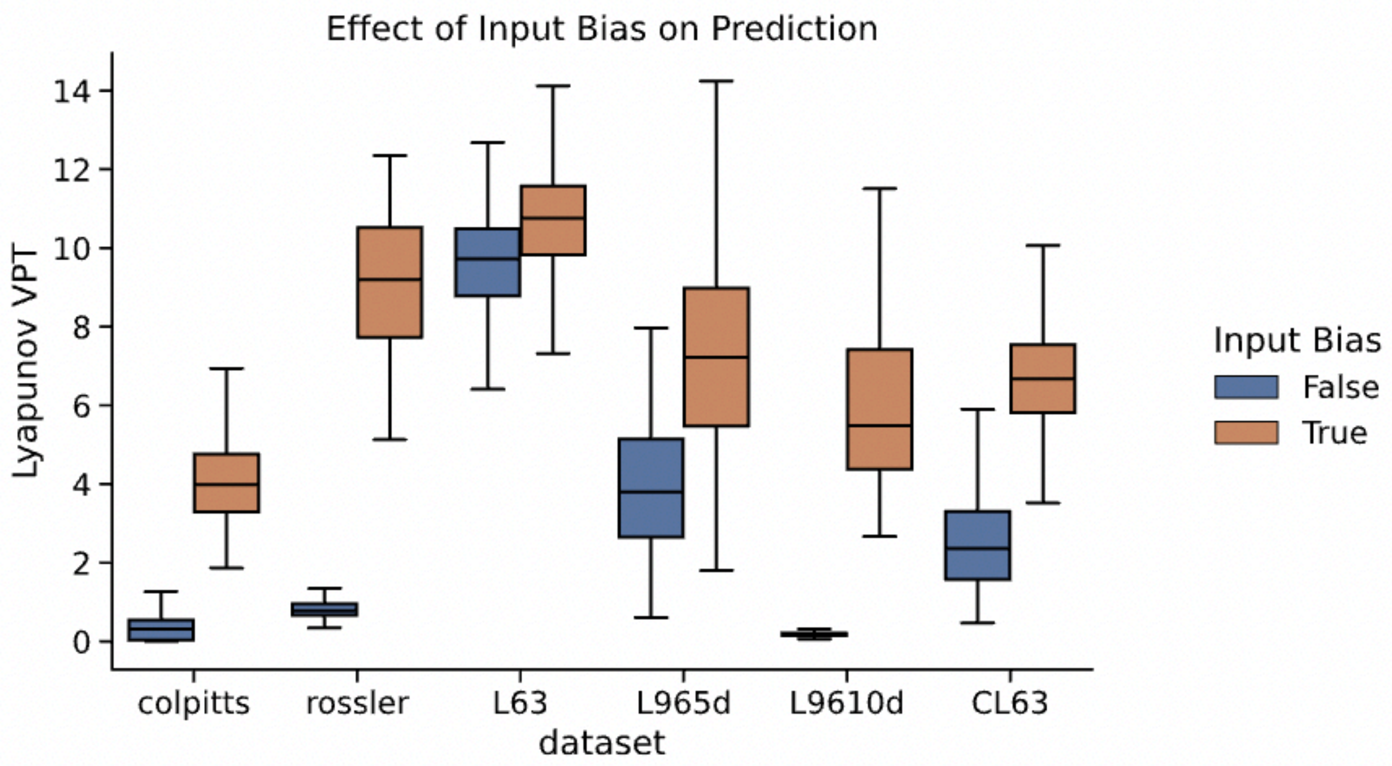


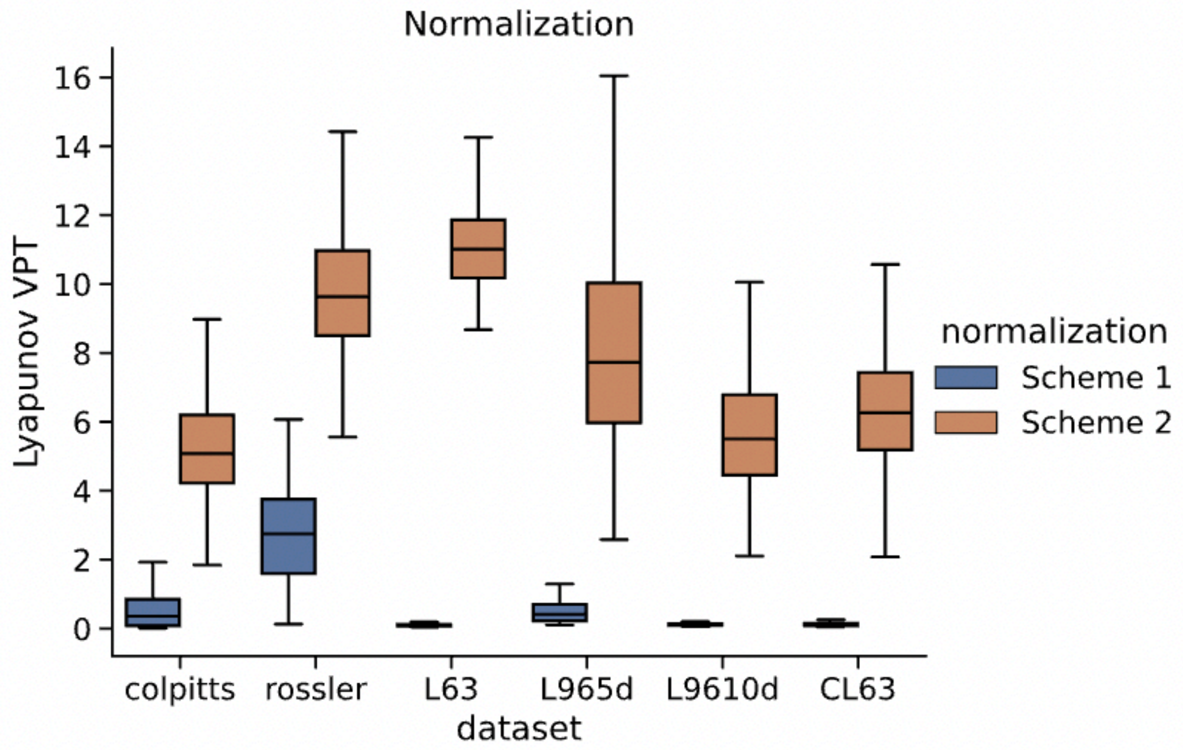


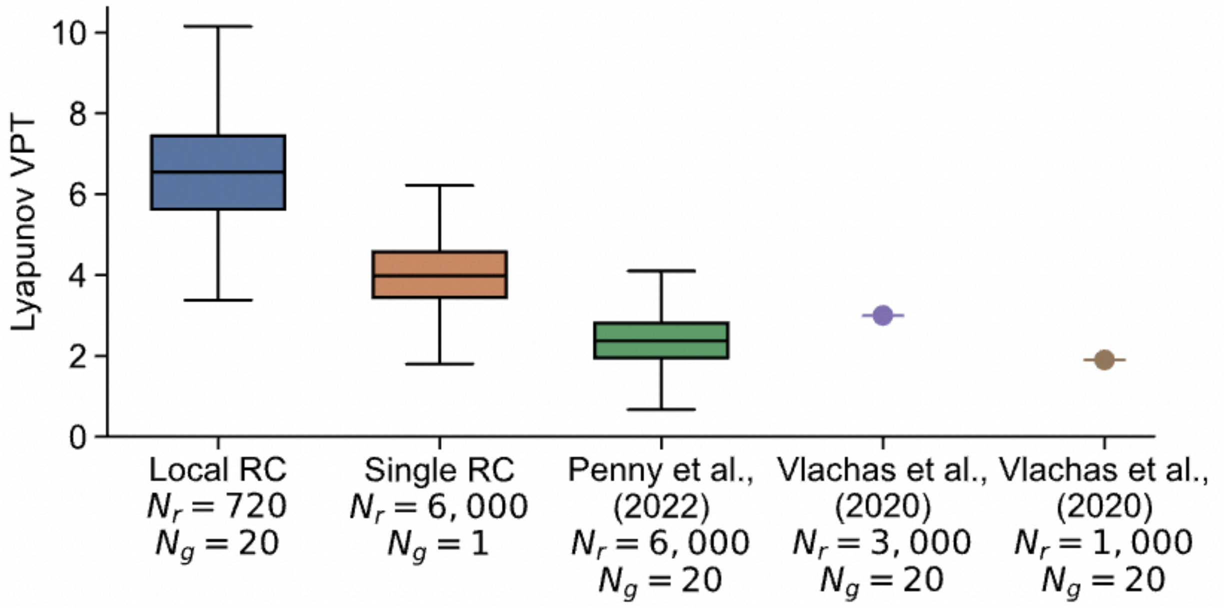










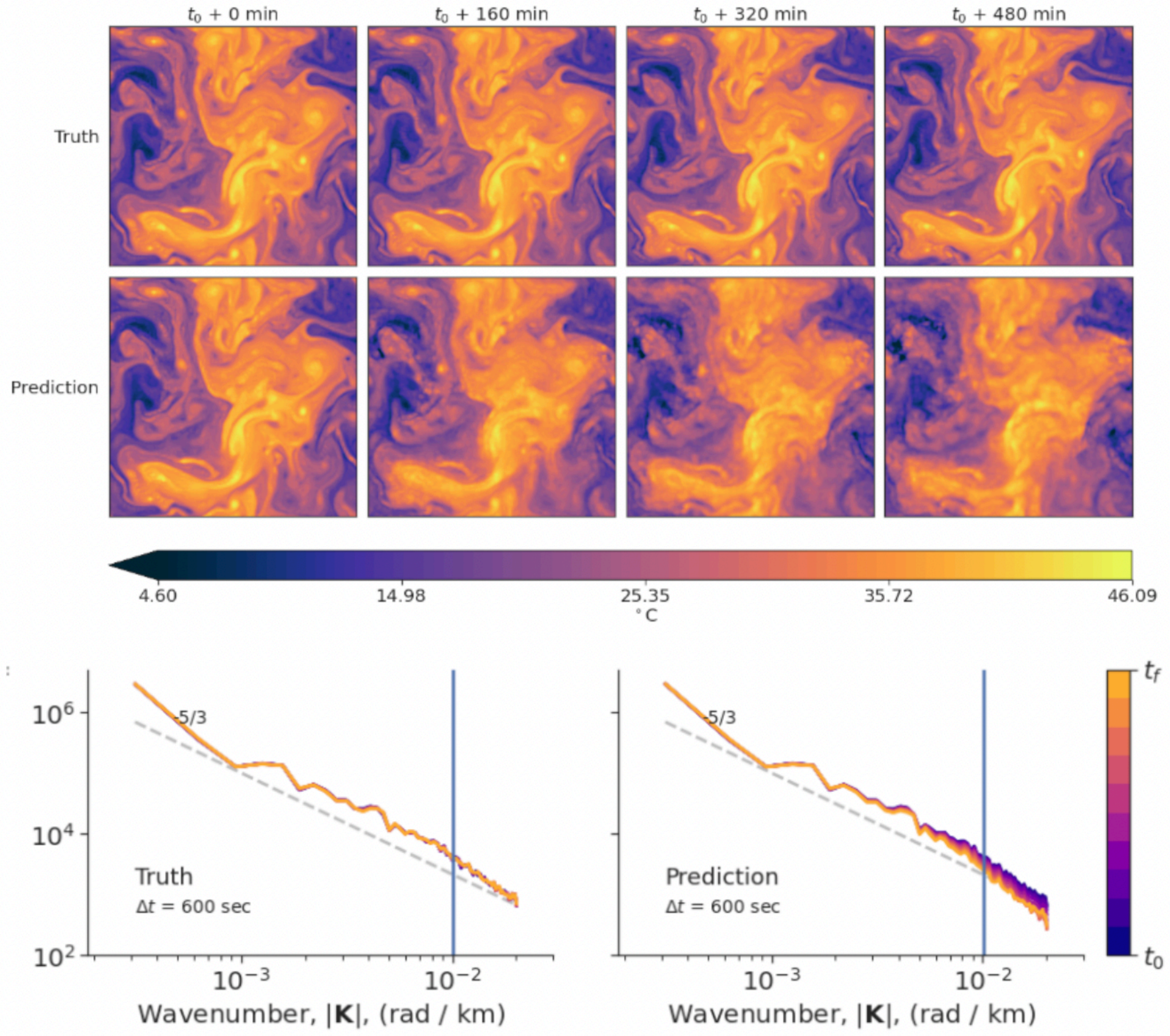


Challenge:

Visual comparisons are not adequate tests for geophysical AI/ML applications.
RMSE is not sufficient to measure the behavior of an AI/ML *forecast* model.
Can we find better metrics for geophysical AI/ML applications?

Challenge:

Can we develop AI/ML models that respond correctly to perturbations in initial conditions? What is the best way to measure this?



- Parallel NVAR prediction of Surface Quasi-Geostrophic Turbulence

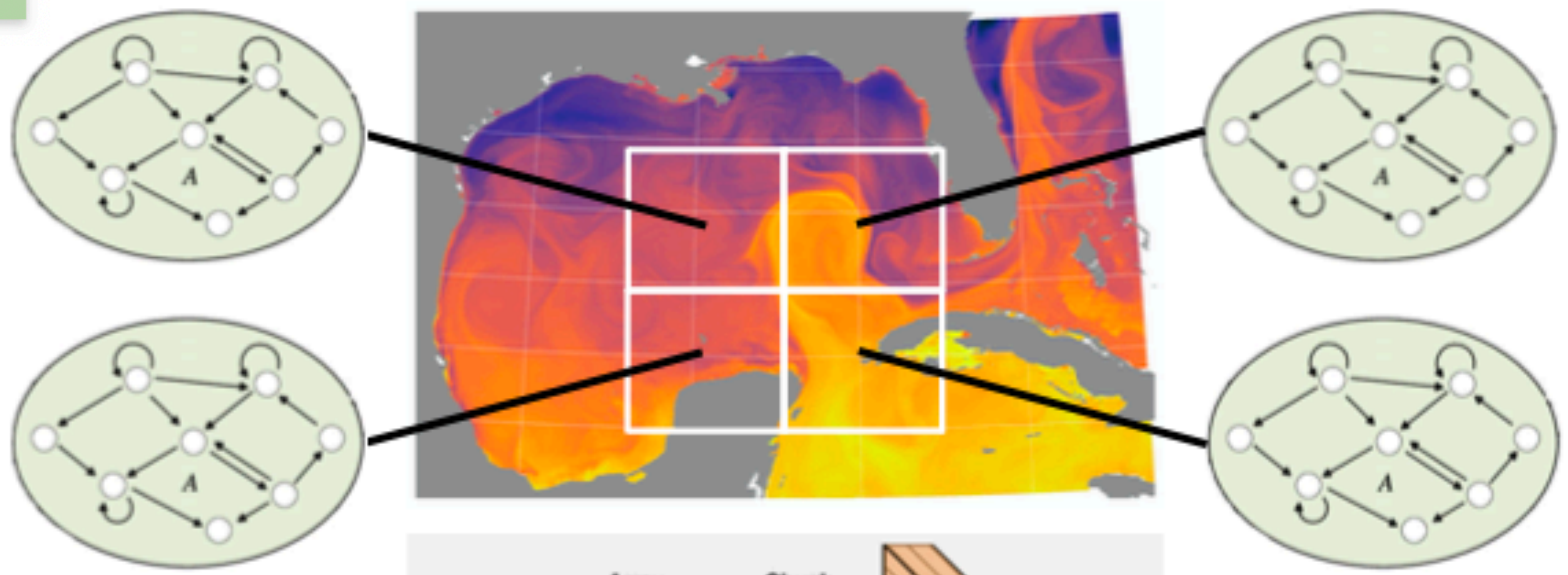
- Synoptic & into mesoscales well captured

- Scales below this unconstrained

Result: overly smooth prediction & spectrum that decays too rapidly

Result: smoothing effect becomes more dramatic as temporal sampling frequency decreases

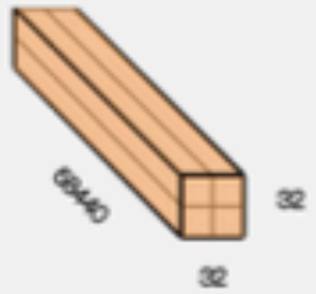
Takeaway: Training on subsampled output (e.g. reanalysis) reduces "effective" emulator resolution



Presentation: Timothy Smith, Wednesday 11:50 to 12:10

DASK

	Array	Chunk
Bytes	228.28 MiB	57.07 MiB
Shape	(58440, 32, 32)	(58440, 16, 16)
Count	719 Tasks	4 Chunks
Type	float32	numpy.ndarray



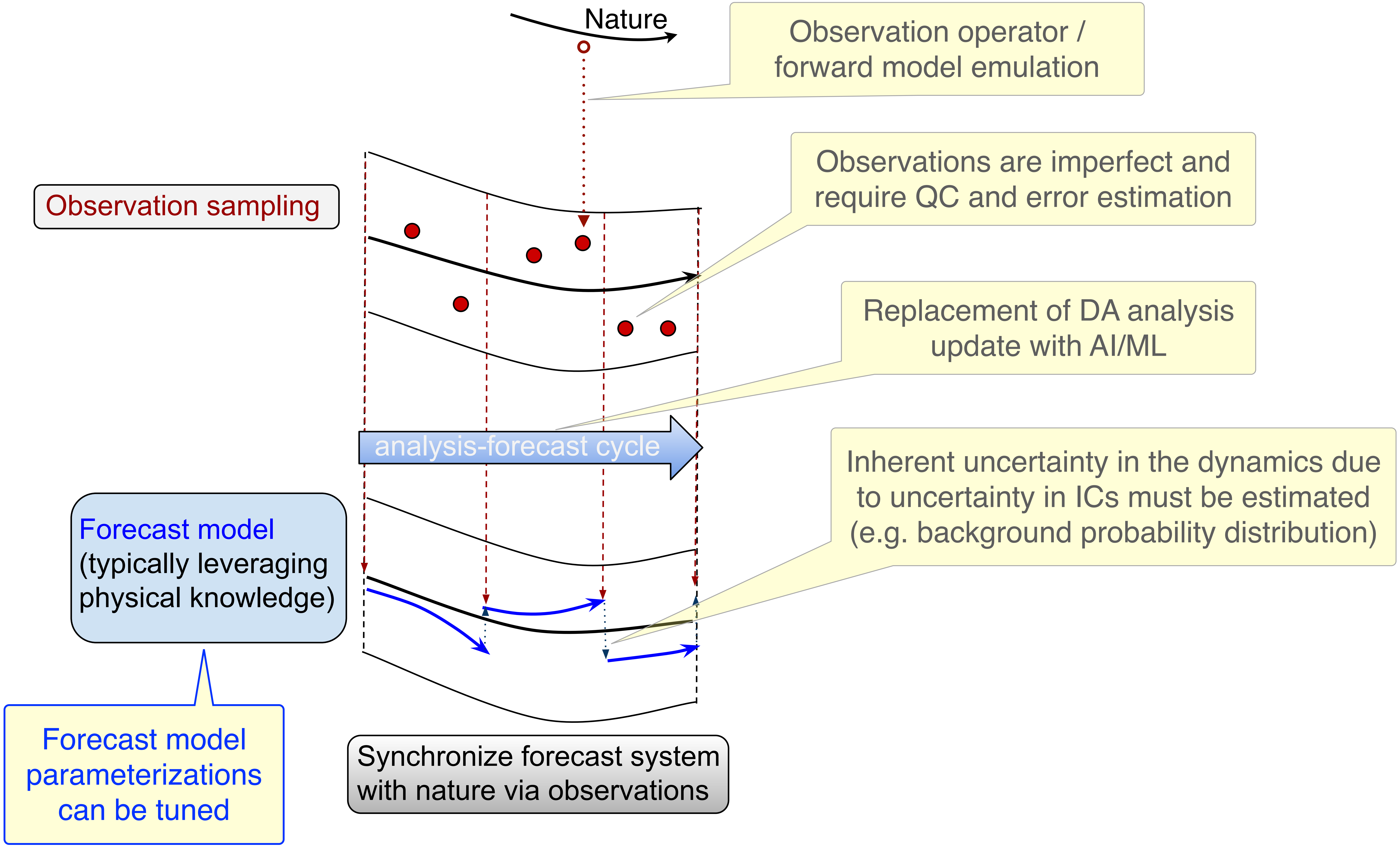
Zarr

Courtesy: Timothy A. Smith, CIRES / NOAA PSL

Geophysical AI/ML methods need some degree of generalizability to apply to unseen data, but they must also retain the desired error characteristics and detail at finer scales.

Challenge:

How do we push to resolutions finer than large-scale synoptic flow while controlling numerical diffusion to acceptable levels?
What are the requirements for a next generation of 'reanalysis' products whose primary purpose may be to support AI/ML applications?


Nature
Observation operator / forward model emulation
Observation sampling
Observations are imperfect and require QC and error estimation
Replacement of DA analysis update with AI/ML
analysis-forecast cycle
Inherent uncertainty in the dynamics due to uncertainty in ICs must be estimated (e.g. background probability distribution)
Forecast model (typically leveraging physical knowledge)
Forecast model parameterizations can be tuned
Synchronize forecast system with nature via observations

Full replacement of parameterizations in NWP models - this has a long history e.g. (Chevallier et al. 1998, 2000; Krasnopolsky et al. 2005, 2008; Krasnopolsky and Fox-Rabinovitz, 2006)

Belochitski and Krasnopolsky (2020)

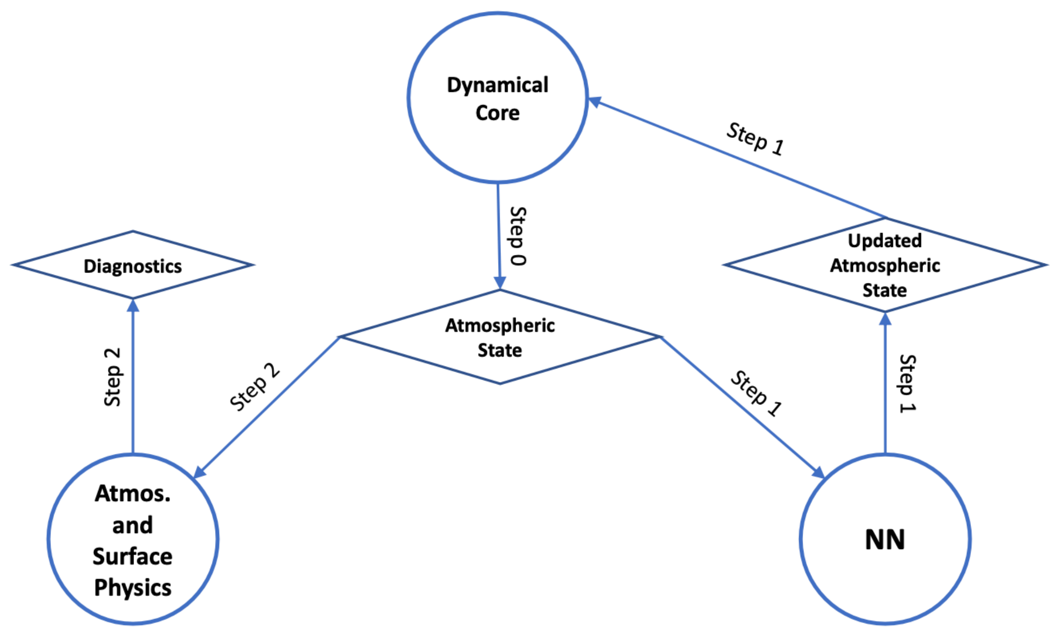


Figure 1. Coupling of the NN model physics to GFS. Step 2 may be performed significantly less frequently than Step 1 that is performed at each time step.

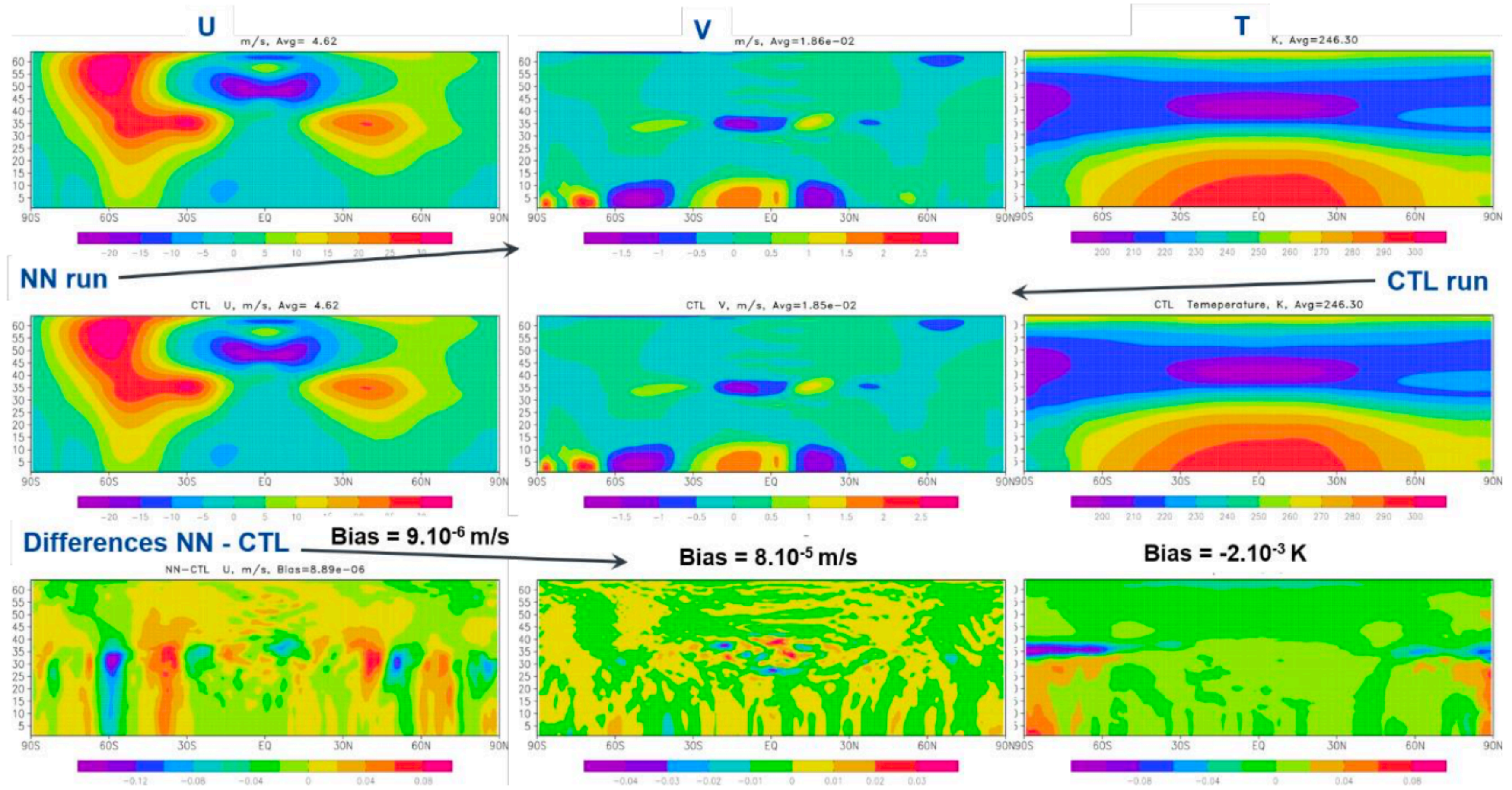


Figure 2. Zonal and time means of an overage over 24 10-day forecasts for U (left column), V (central column), and temperature (right column). Upper row – results produced by HGFS, medium – by GFS, and the lower row the difference (HGFS – GFS). Vertical coordinate shows model level number.

Presentation: Vladimir Krasnopolsky, Wednesday 09:00 to 09:30

Belochitski and Krasnopolsky (2021)

A. Belochitski and V. Krasnopolsky: Robustness of NN emulations of radiative transfer parameterizations 7433

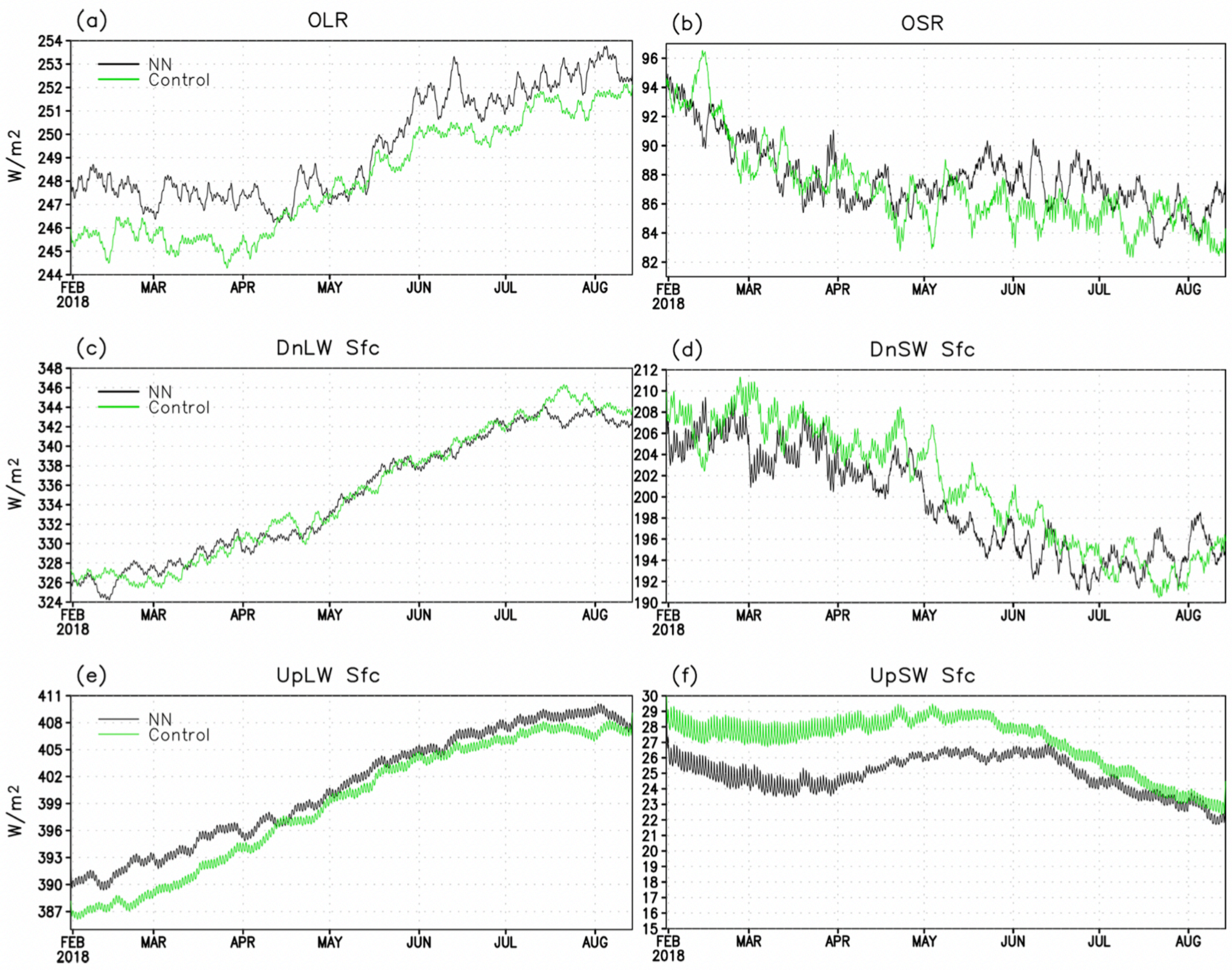


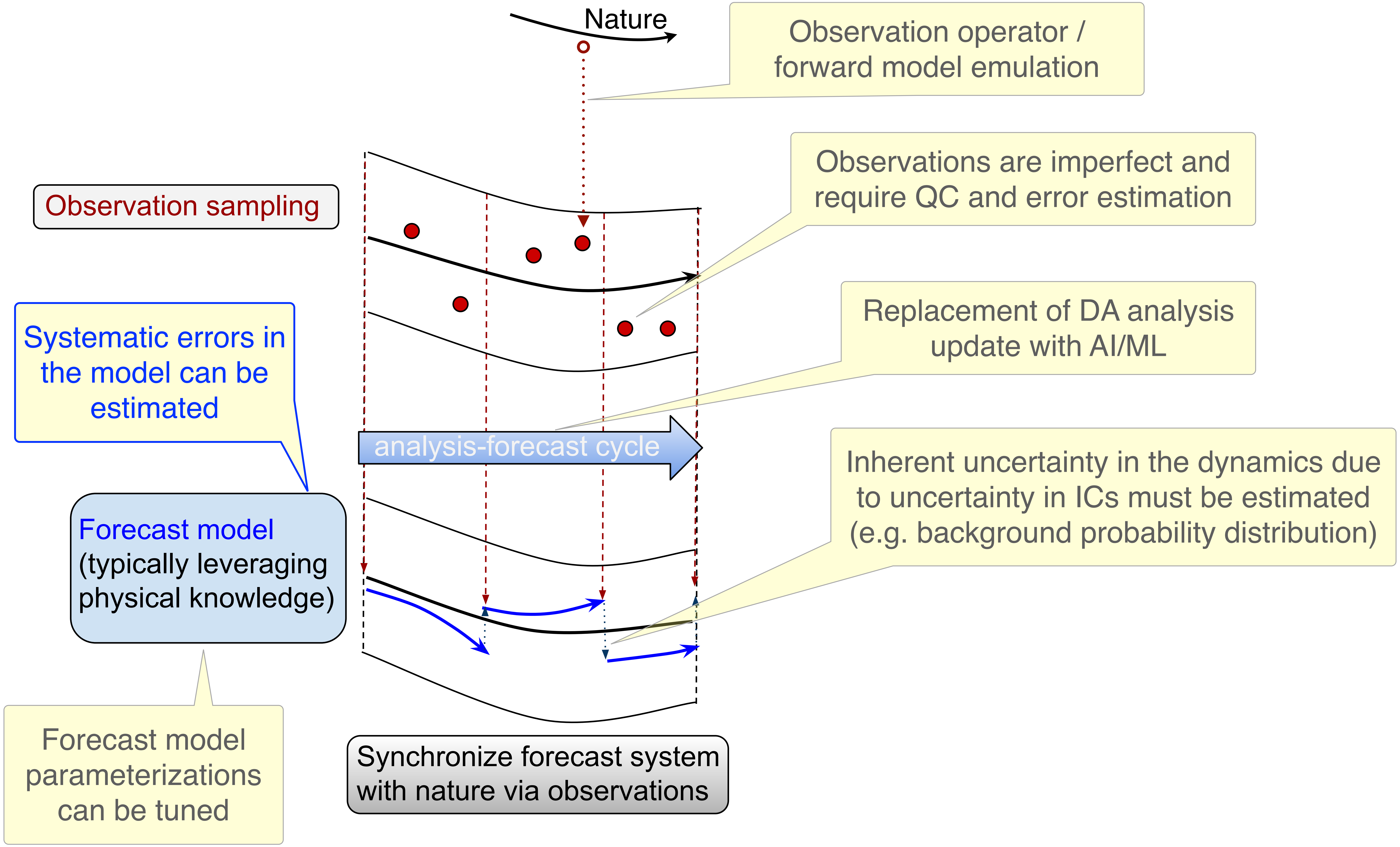
Figure 4. Time series of a running 10 d mean covering 1 February–1 August 2018 for **(a)** outgoing LW at TOA, **(b)** outgoing SW at TOA, **(c)** downwelling LW radiation at the surface, **(d)** downwelling SW radiation at the surface, **(e)** upwelling LW radiation at the surface, and **(f)** upwelling SW radiation at the surface. Black curves – results produced by HGFS; green – results by GFS.

<u>Opportunity</u>: Ultimately, some form of hybrid combination of conventional numerical modeling and data-driven/ML parameterization seems like a promising approach for trade-offs in accuracy and efficiency in the long-run

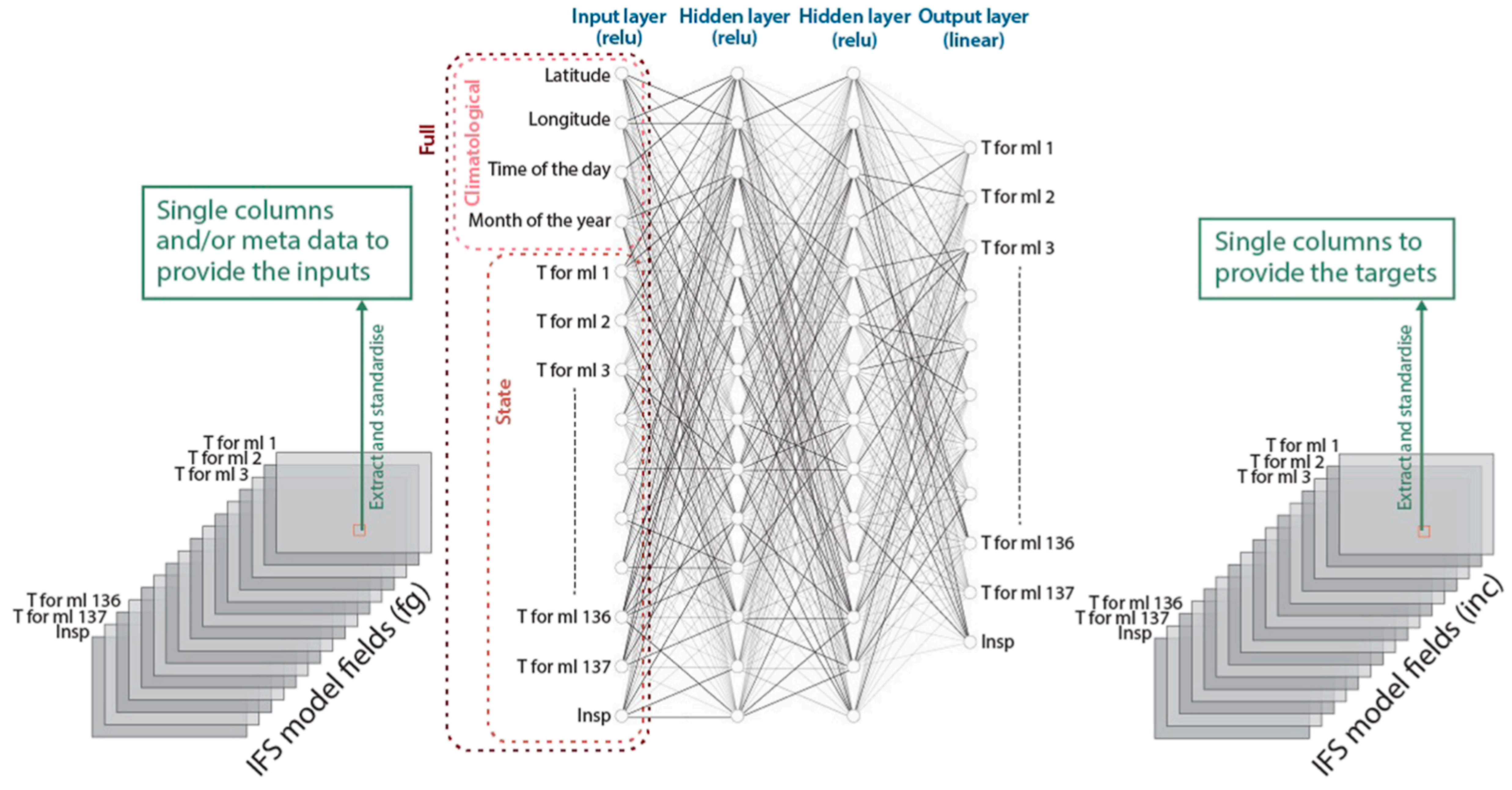
Considering Hybrid models:

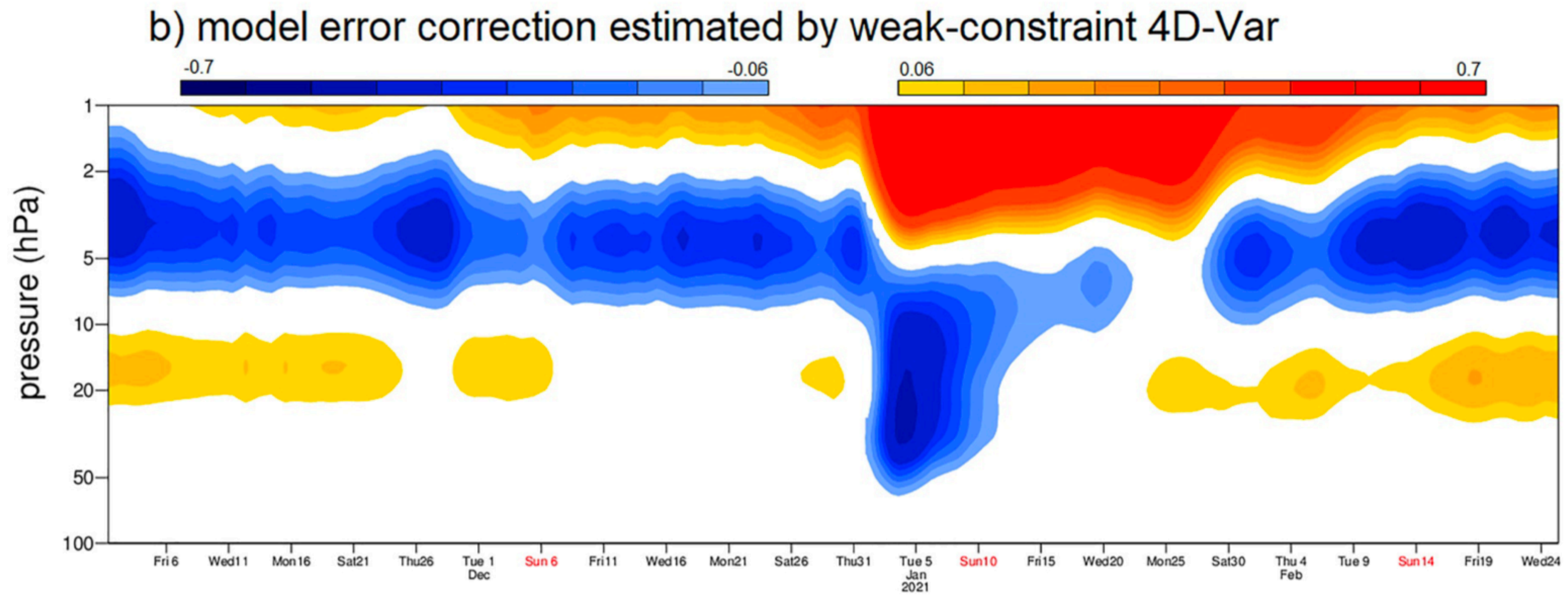
Challenge:

As numerical forecast models are modernized (e.g. written in new languages that support differentiation, and designed to take advantage of GPUs), can AI/ML solutions maintain a competitive edge (in terms of computational cost) over conventional modeling?


Nature
Observation operator / forward model emulation
Observation sampling
Observations are imperfect and require QC and error estimation
Systematic errors in the model can be estimated
Replacement of DA analysis update with AI/ML
analysis-forecast cycle
Inherent uncertainty in the dynamics due to uncertainty in ICs must be estimated (e.g. background probability distribution)
Forecast model (typically leveraging physical knowledge)
Forecast model parameterizations can be tuned
Synchronize forecast system with nature via observations

Bonavita and Laloyaux (2020); Laloyaux et al. (2022)





Chen et al., (2022) correction of systematic errors

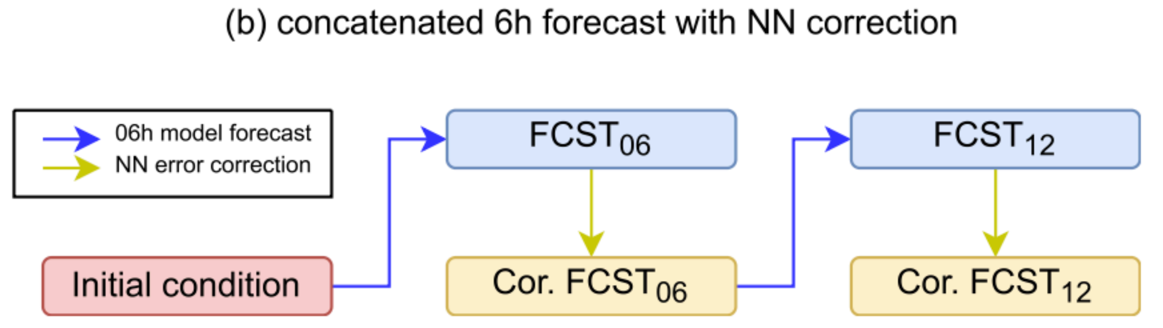
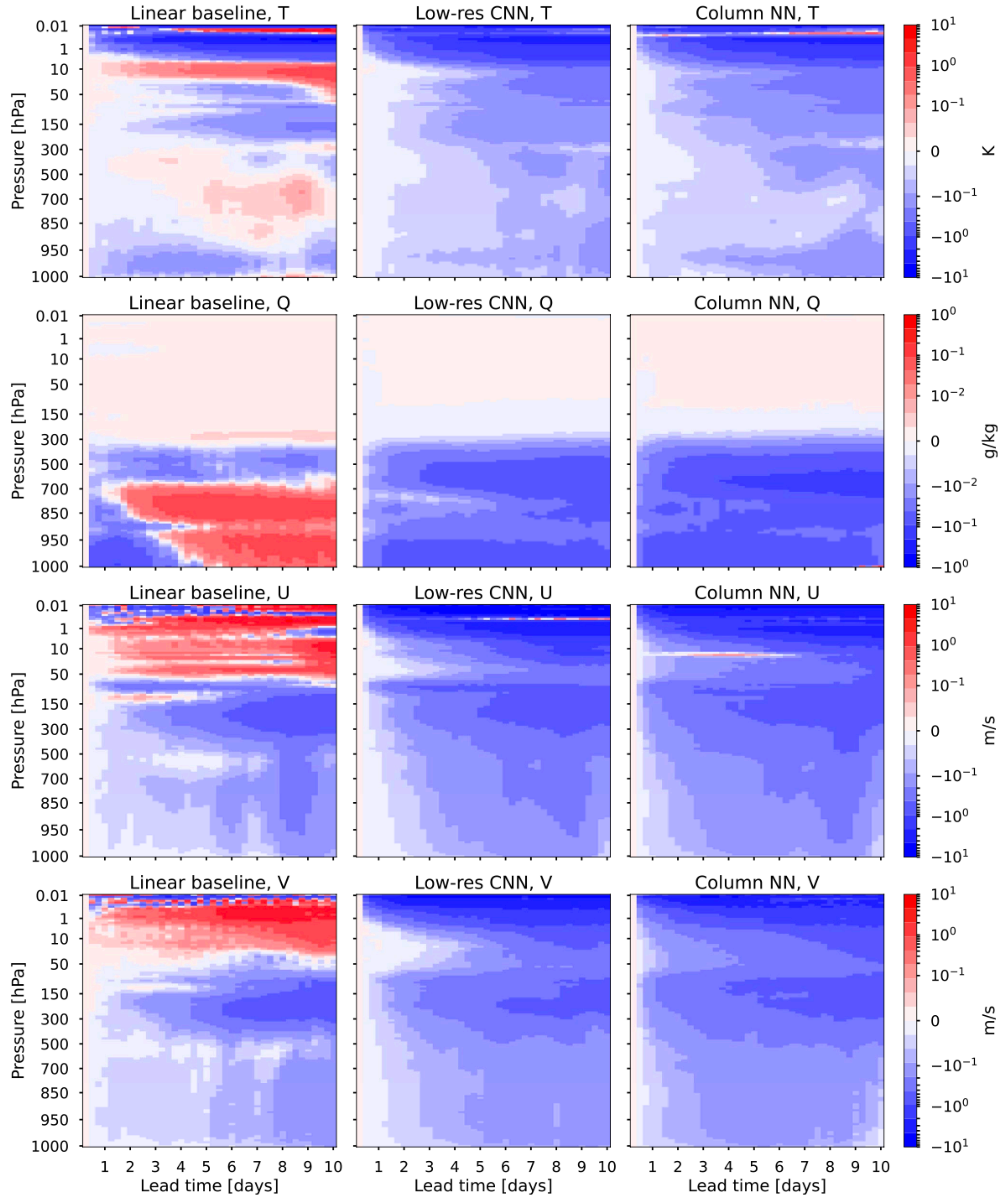
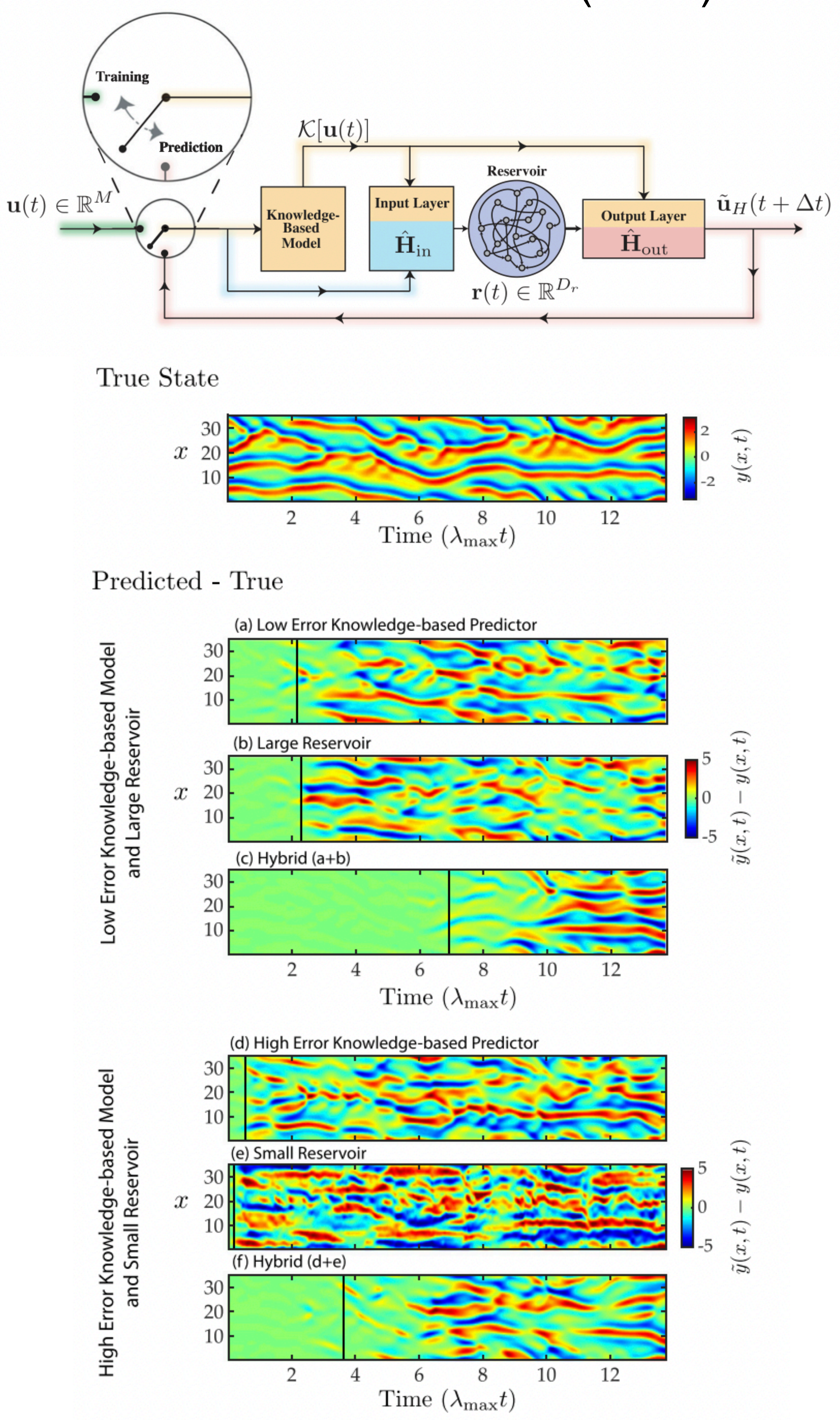


Figure 4. Schematic illustration of the integration of the error corrections with the workflow of (a) sequential data assimilation and (b) concatenated 6h free forecasts.

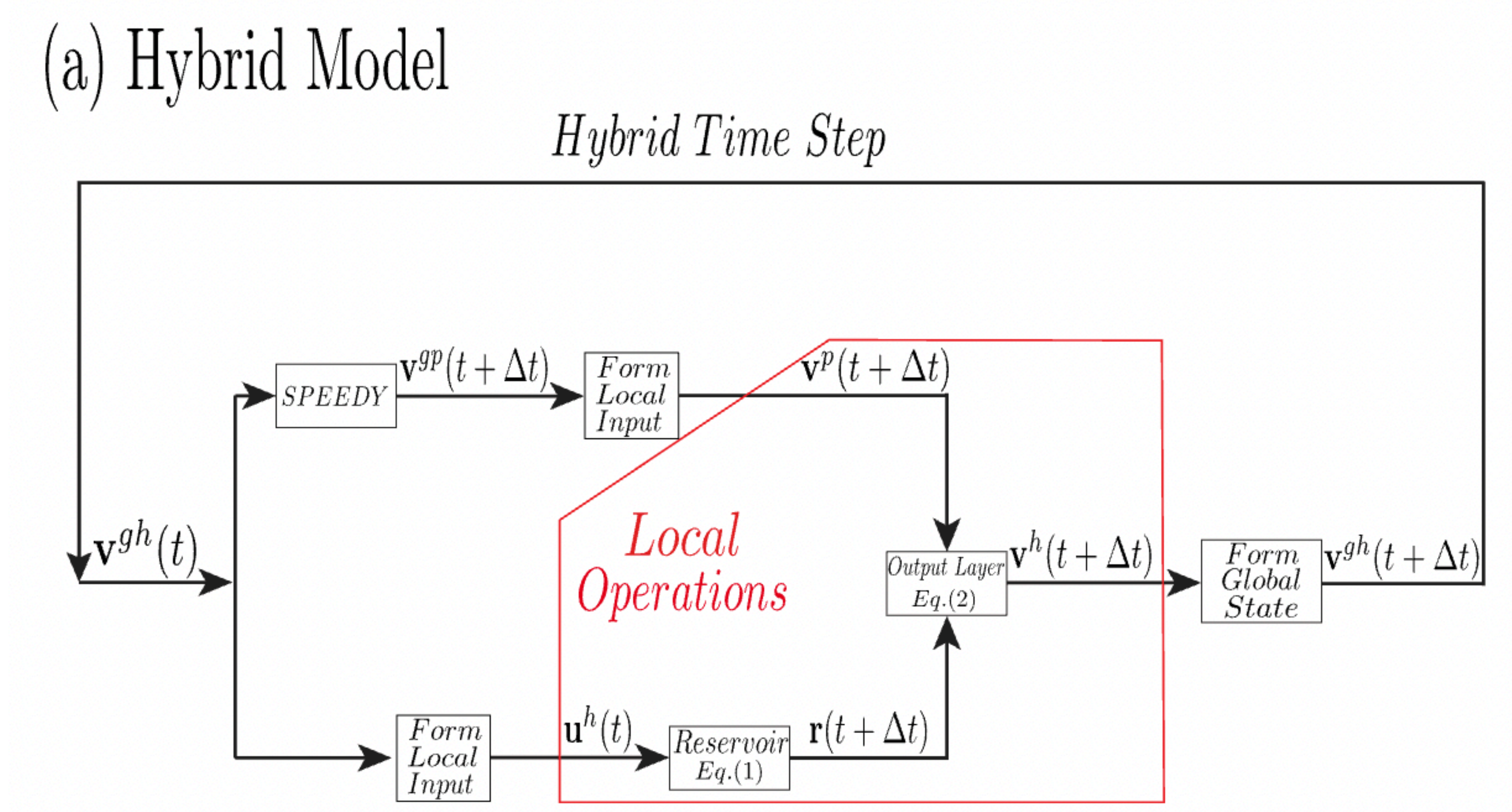


Pathak et al. (2018)



Hybrid models that combine primitive equation models with AI/ML models are also being developed to reduce overall biases.

Arcomano et al. (2022)



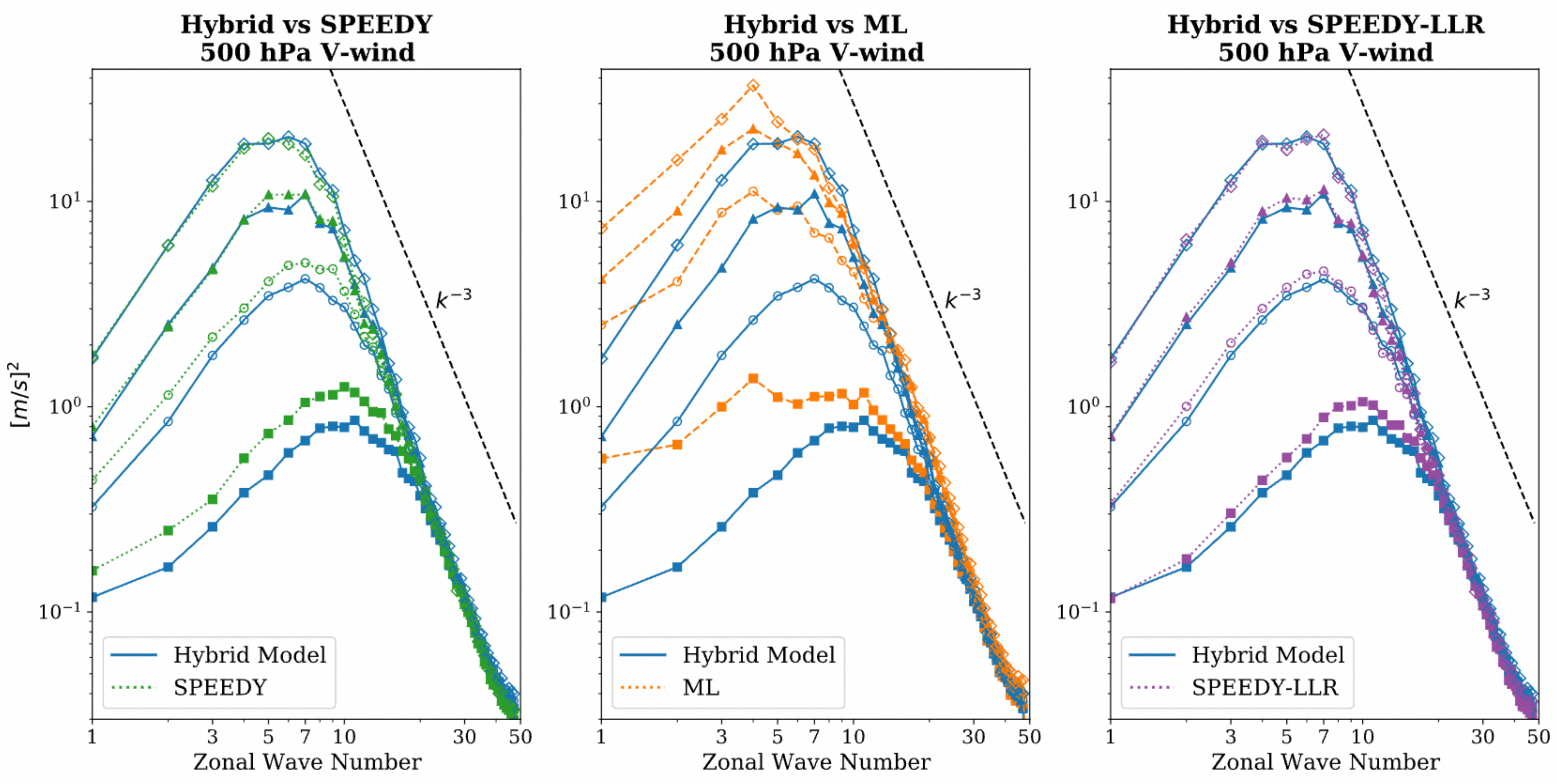


Figure 5. Spectral distribution of the 500 hPa meridional wind forecast error in the NH midlatitudes (between 30°N and 70°N) with respect to the zonal wave number. The power spectra of the forecast errors are shown (left) for the the hybrid model (blue) versus simplified parameterization, primitive-equation dynamics (SPEEDY) (green) (middle) the hybrid model (blue) versus the ML-only model (orange), and (right) hybrid model (blue) versus SPEEDY-LLR (purple) at day 1 (solid square), day 3 (open circle), day 5 (solid triangle), and day 10 (open diamond).

Challenge:

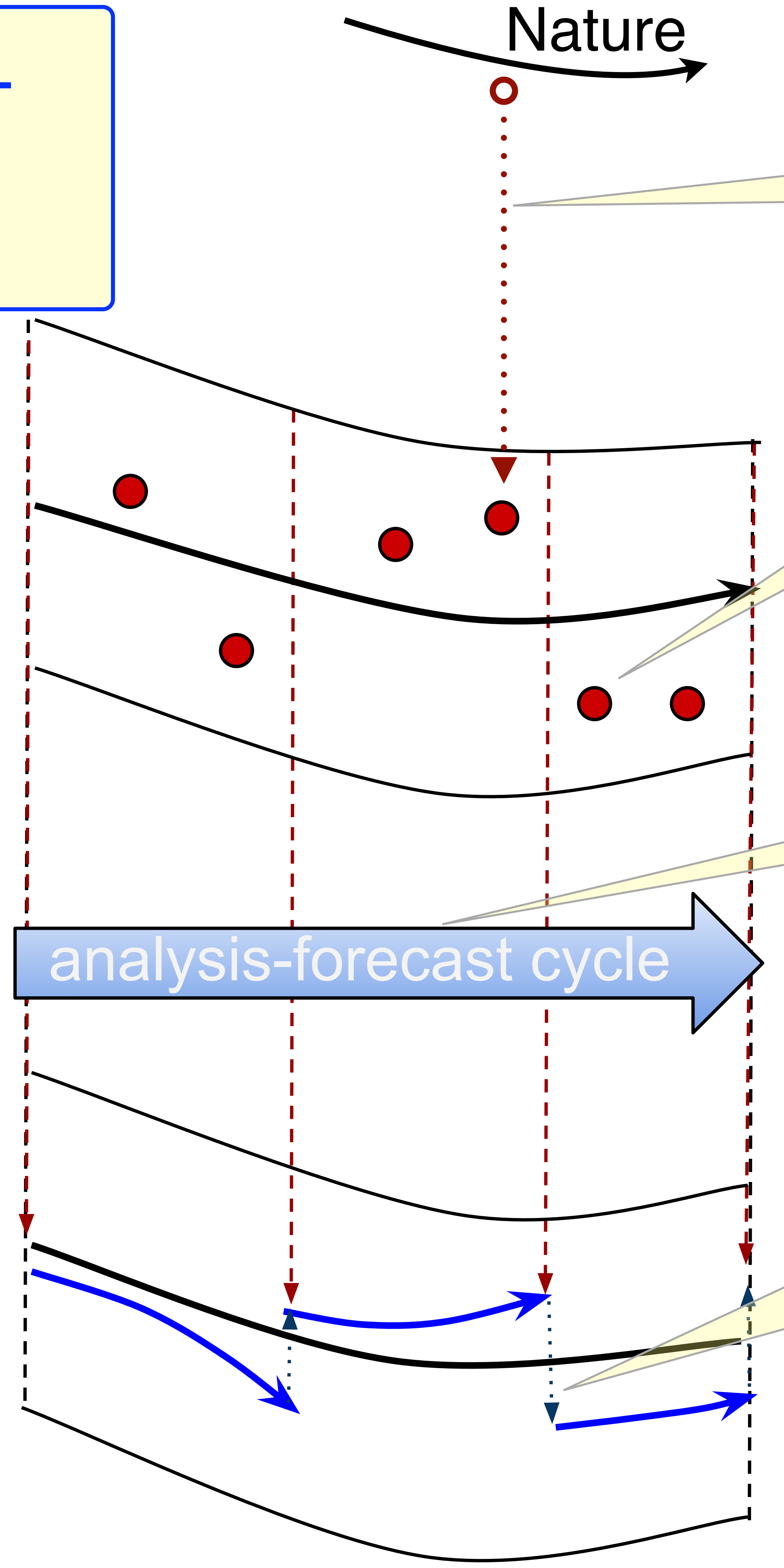
How much state-dependent (conventional) model error can we learn from comparison with observations?

How do we separate systematic observation errors from systematic model forecast errors?

Limitations due to under-sampling must be acknowledged

Observation sampling

Observation operator / forward model emulation



Observations are imperfect and require QC and error estimation

Systematic errors in the model can be estimated

Replacement of DA analysis update with AI/ML

Forecast model (typically leveraging physical knowledge)

Inherent uncertainty in the dynamics due to uncertainty in ICs must be estimated (e.g. background probability distribution)

Forecast model parameterizations can be tuned

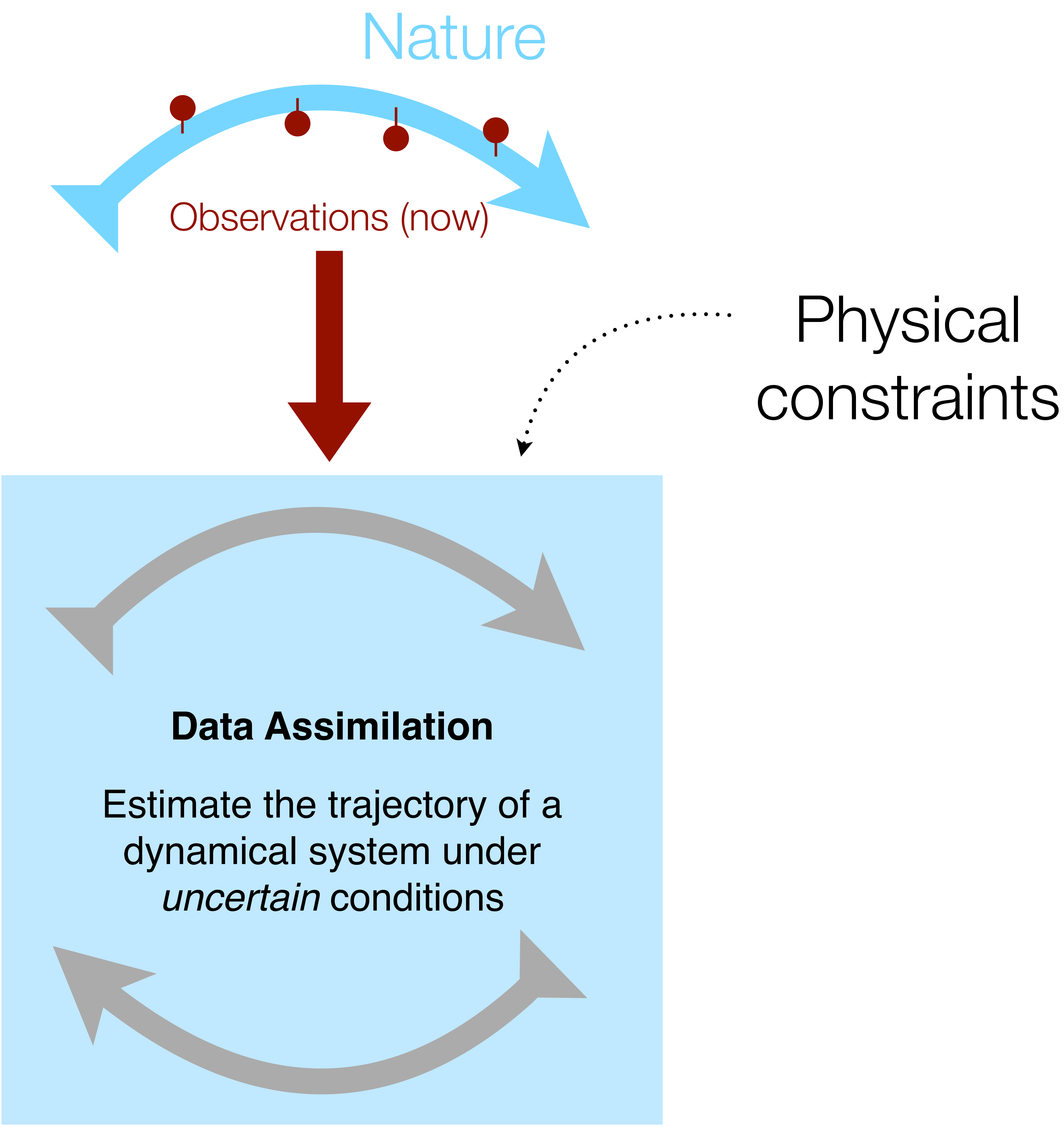
Synchronize forecast system with nature via observations

Limitations on how well the systems can synchronize must be acknowledged

SECTION

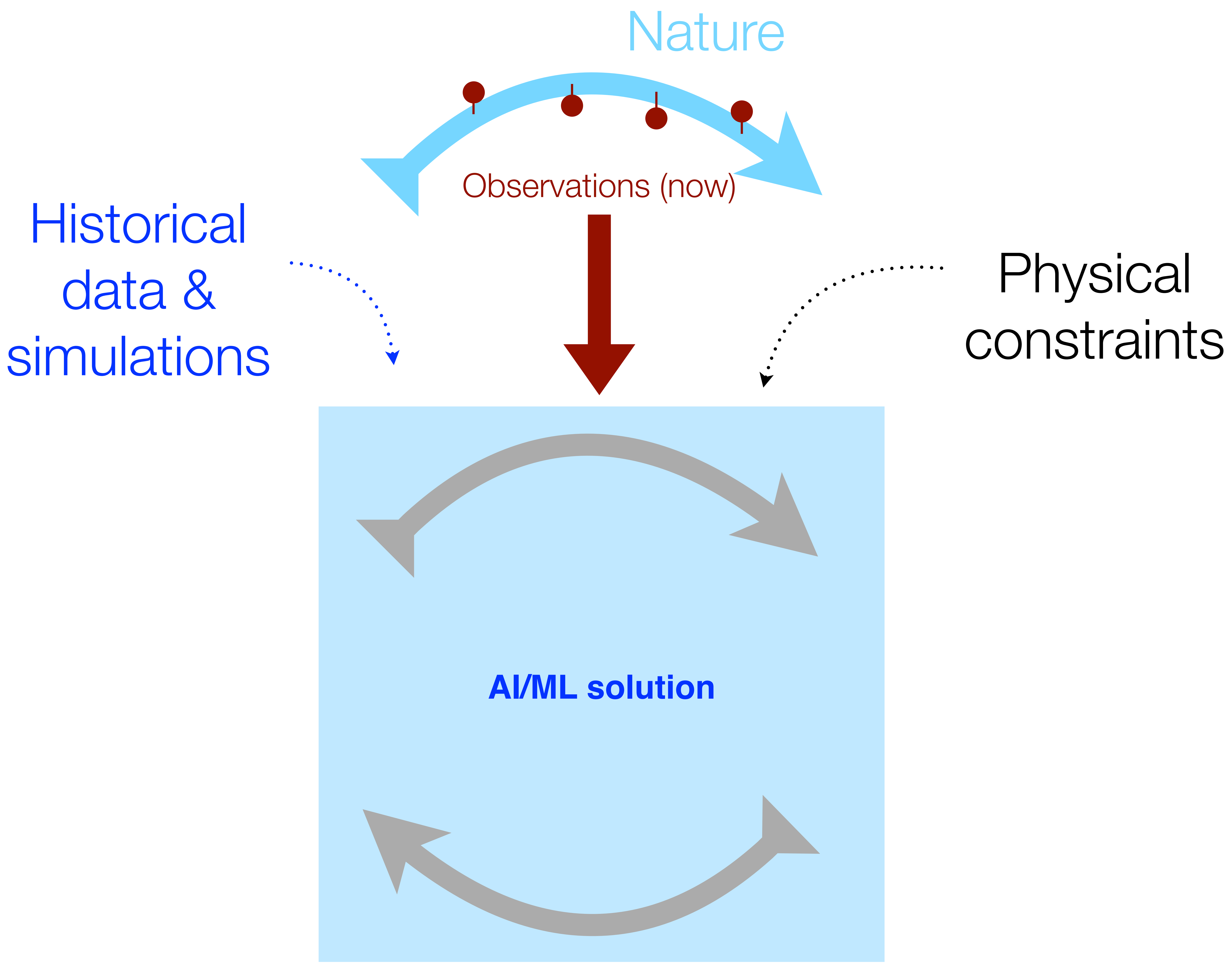
What is the grand vision for integrating AI/ML with data assimilation?

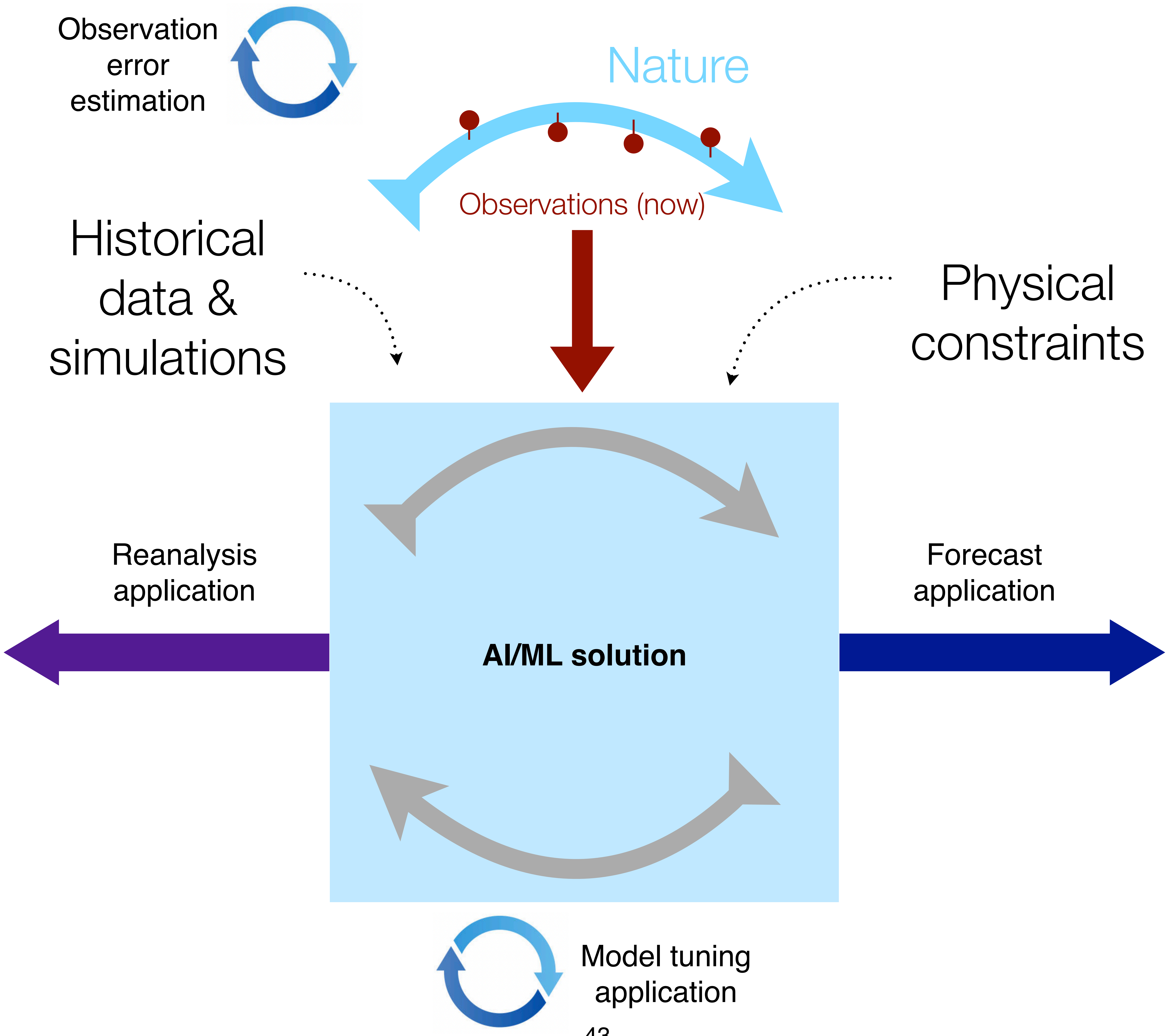
Inputs:



Nature
Observations (now)
Physical constraints
Data Assimilation
Estimate the trajectory of a dynamical system under *uncertain* conditions

Inputs:


Nature
Observations (now)
Historical data & simulations
Physical constraints
AI/ML solution


Observation error estimation
Nature
Observations (now)
Historical data & simulations
Physical constraints
Reanalysis application
AI/ML solution
Forecast application
Model tuning application

Early steps in this direction:
Bocquet, Brajard, Carrassi, and Bertino

Presentation: Marc Bocquet,
Tuesday 11:50 to 12:10

“Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models” *(Bocquet et al., 2019)*

“Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model” *(Brajard et al., 2020)*

“Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization” *(Bocquet et al., 2020)*

C. Buiza et al., (2022)

Presentation: Rossella Arcucci,
Tuesday 11:00 to 11:30

- Integration of Kalman Filters with Reduced Order Models and Neural Networks (oceanographic problem)
- DL model based on the integration of Variational Data Assimilation with Encoder-Decoder technologies (air-pollution forecasting)
- DL approach that integrates Convolutional Neural Networks with Kalman Filters and variational DA (pharmacokinetic modeling)
- DL approach that integrates variational Data Assimilation with neural networks for parameter estimation (Economic system)
- DL approach that integrates Gaussian Processes with Variational Data Assimilation (optimal sensor placement)

Opportunities:

Technical:
- AI/ML methods can lead to **improved understanding about what properties a skillful forecast model must have**
- There are **no strict guidelines to what an AI/ML solution should look like**, and a likely future is one in which we have hybrid combinations of conventional and AI/ML modeling
- **All future models should be software-differentiable**. The trends and tools for AI/ML make this easier to achieve.
- Conventional primitive equation models rely on emergent properties from a “bottom-up” design. **AI/ML approaches have much more flexibility and control over the scales of motion**, can separate them, allow or disallow interactions between them, constrain them differently and give them different levels of importance/priority.
- The basic concept of **producing and evolving error estimates for observations, model, and dynamics** is largely absent from AI/ML approaches right now - this is one of the biggest opportunities for DA to inform AI/ML development.

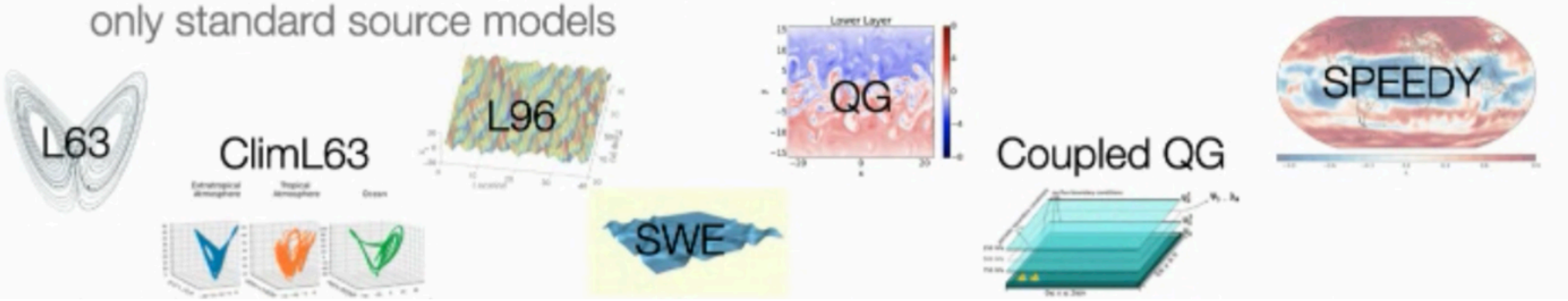
Data Assimilation
- Low-cost surrogate models provide the **opportunity for new DA methods** that were previously infeasible, e.g. large ensembles, high resolutions, non-Gaussian/nonlinear analysis methods, complex applications where conventional modeling is less mature
- Optimization framework for AI/ML methods may be able to be leveraged to create **new algorithmic approaches for DA**

Community
- Growing the **visibility of data assimilation** outside of weather forecasting community
- DA community helping to **shape the development of new ideas** for more general applications in AI/ML
- **Merging of communities** is inevitable - many different types of expertise are needed to solve these problems

DataAssimBench: Google-funded development of a benchmarking repo for integrating AI/ML with Data Assimilation (with Jax support)

Where are the benchmark training sets?

- We generate them from known dynamical systems, there is no standard set, only standard source models

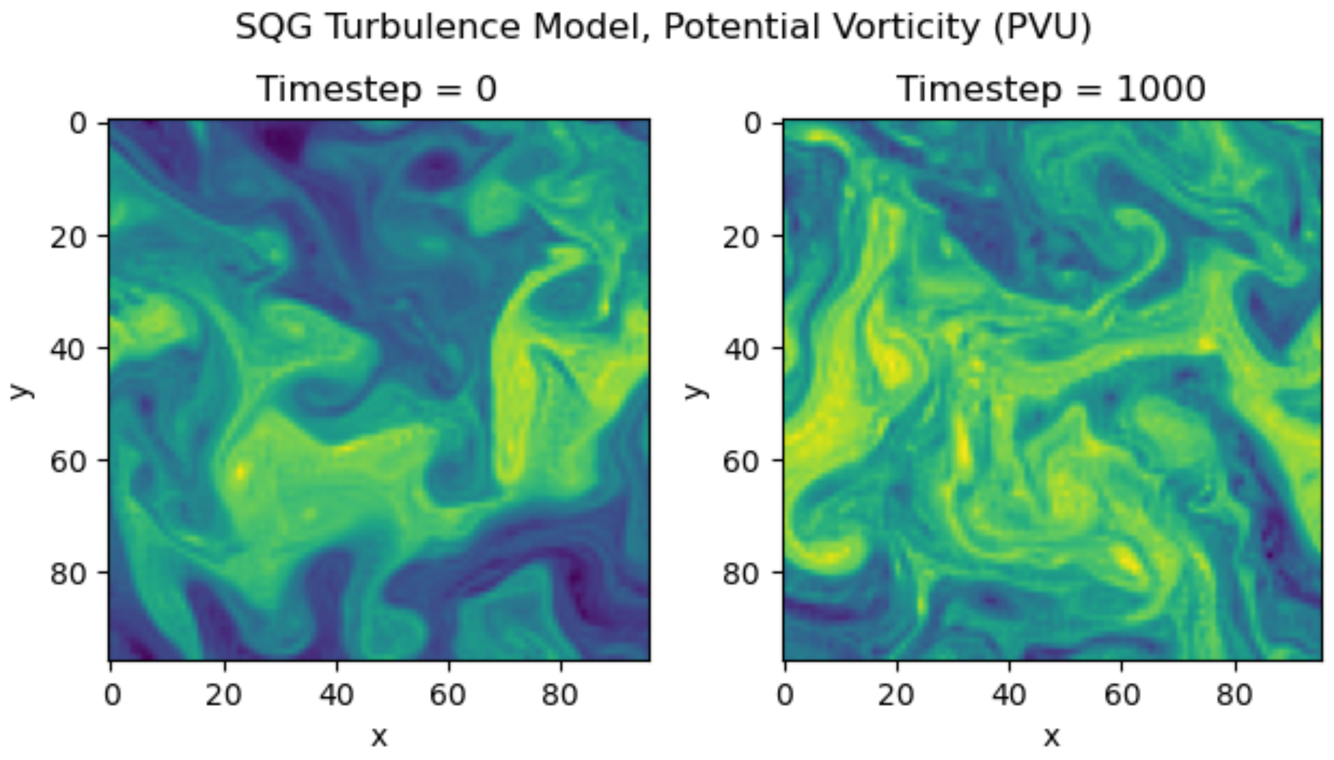


- This is a stepping stone toward the use of more realistic models, real real-world observational data

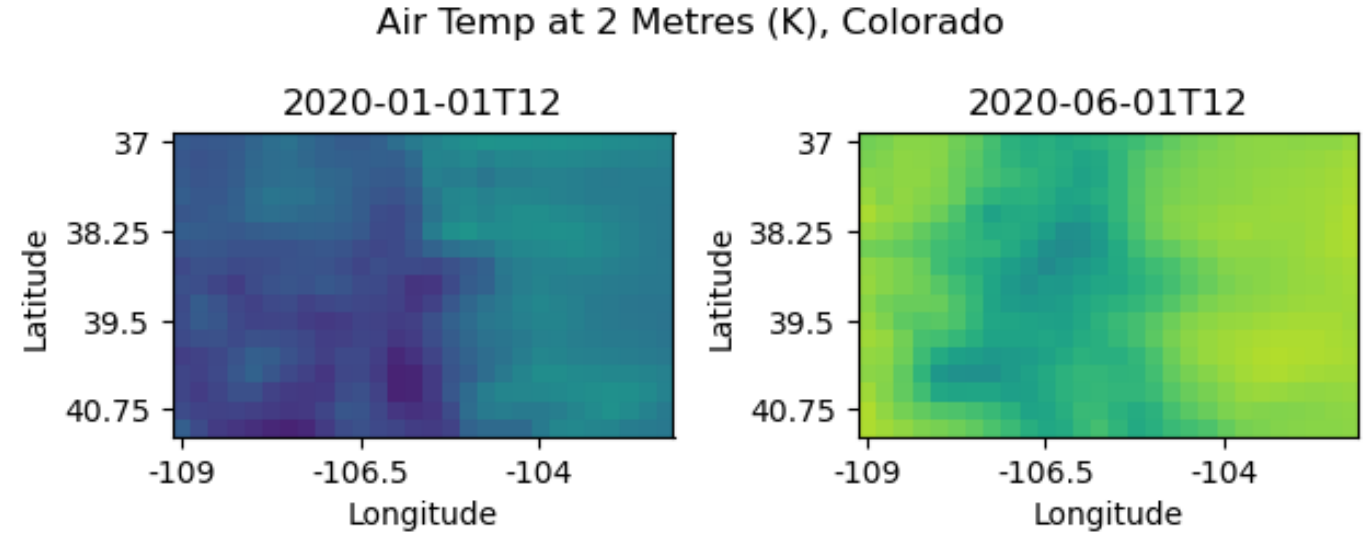
FV3-GFS IFS COAMPS Wavewatch3

NOAA UFS MOM6 Hycom NEMO Navgen

```
In [1]: from dabench.data import sqgturb
        import matplotlib.pyplot as plt

In [2]: model_obj = sqgturb.DataSQGturb()
        model_obj.generate(n_steps = 1000)
        gridded_vals = model_obj.to_original_dim()

In [3]: fig, ax = plt.subplots(1, 2)
        fig.suptitle('SQG Turbulence Model, Potential Vorticity (PVU)')
        ax[0].imshow(gridded_vals[0, 1])
        ax[0].set_title('Timestep = 0')
        ax[0].set_xlabel('x'); ax[0].set_ylabel('y')
        ax[1].imshow(gridded_vals[-1, 1])
        ax[1].set_title('Timestep = 1000')
        ax[1].set_xlabel('x'); ax[1].set_ylabel('y')
        fig.tight_layout()
        fig.subplots_adjust(top=1.2)
        plt.show()
```




```
In [1]: from dabench.data import aws
        import matplotlib.pyplot as plt
        import numpy as np

In [2]: data_obj = aws.DataAWS(variables = ['air_temperature_at_2_metres'],
                               years = [2020, 2021],
                               min_lat = 36.992426, max_lat = 41.003444,
                               min_lon = -109.060253, max_lon = -102.041524
                              )
        data_obj.load()
        gridded_values = data_obj.to_original_dim()

In [3]: fig, ax = plt.subplots(1, 2)
        fig.suptitle('Air Temp at 2 Metres (K), Colorado')
        ax[0].imshow(gridded_values[12], vmin=250, vmax=300)
        ax[0].set_title(np.datetime_as_string(data_obj.times[12], unit='h')); ax[0].set_xl
        ax[0].set_yticks(ticks=[0, 5, 10, 15], labels=[37, 38.25, 39.5, 40.75]); ax[0].set
        ax[1].imshow(gridded_values[3660], vmin=250, vmax=300)
        ax[1].set_title(np.datetime_as_string(data_obj.times[3660], unit='h')); ax[1].set_
        ax[1].set_yticks(ticks=[0, 5, 10, 15], labels=[37, 38.25, 39.5, 40.75]); ax[1].set
        fig.tight_layout()
        fig.subplots_adjust(top=1.4)
        plt.show()
```



Contact:
steve.penny@sofarocean.com
Steve Penny, Kylen Solvik (CU), Stephan Hoyer (Google), Tim Smith (CIRES/NOAA), Tse-Chun Chen (CIRES/NOAA), Sarah Balkissoon (CIRES)

Contact:
steve.penny@sofarocean.com