End-to-end and physics-informed learning for dynamical systems

R. Fablet, Prof. IMT Atlantique
Lab-STICC, INRIA Team Odyssey

Joint work with B. Chapron, L. Drumetz, F. Rousseau, E. Mémin, S. Ouala, M. Beauchamp, Q. Febvre, A. Pascual …

Email: ronan.fablet@imt-atlantique.fr
Web: cia-oceanix.github.io

March 2022

Overview of the talk

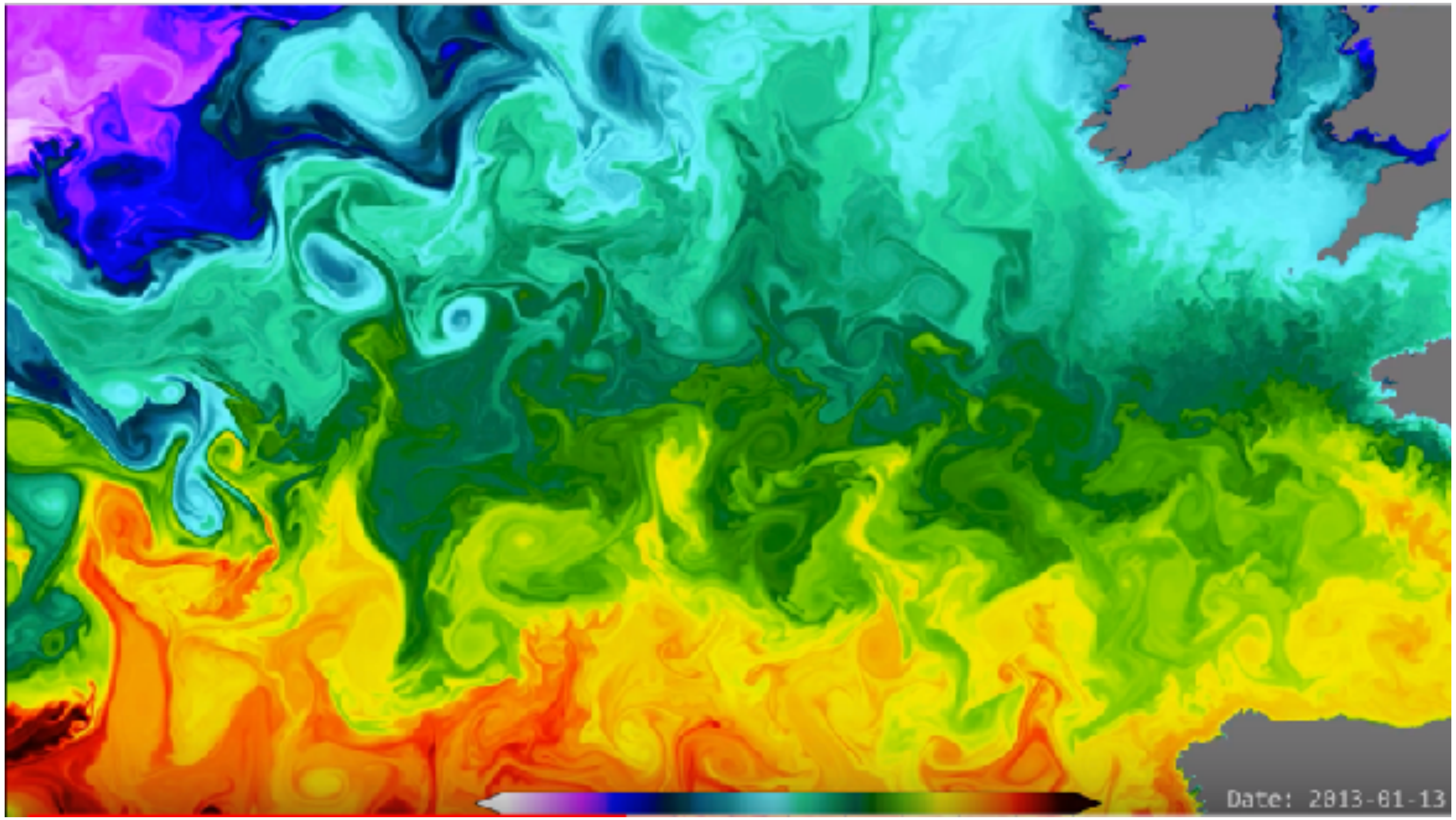
- ***Challenges in ocean observation, modeling and forecasting***

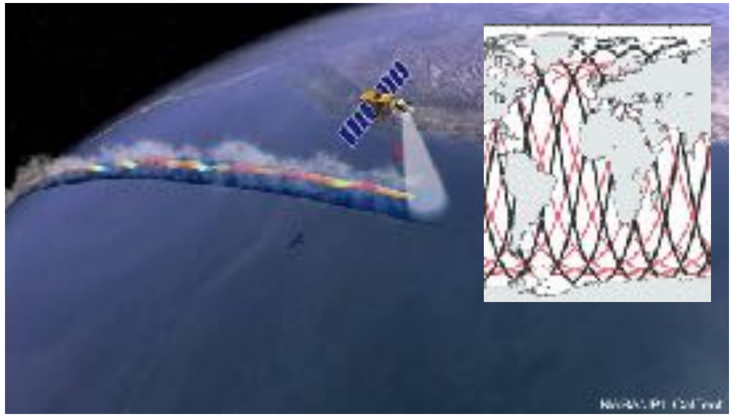
- ***Beyond Neural Networks regarded as black boxes***

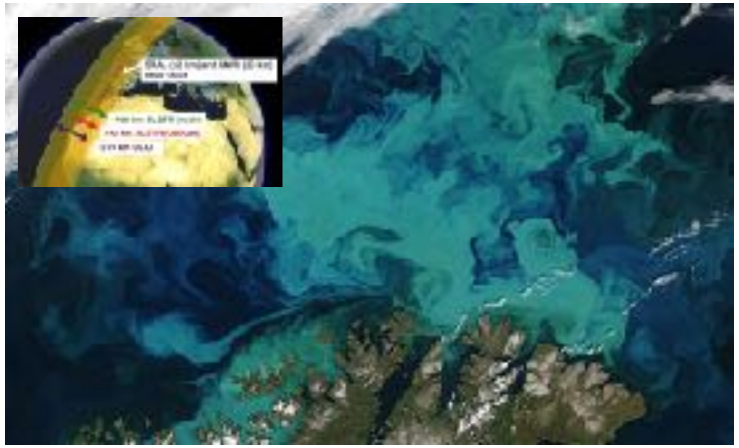
- ***End-to-end learning can make a difference***

Challenges in Ocean Modeling and Forecasting

Context: No observation / simulation system to resolve all scales and processes simultaneously

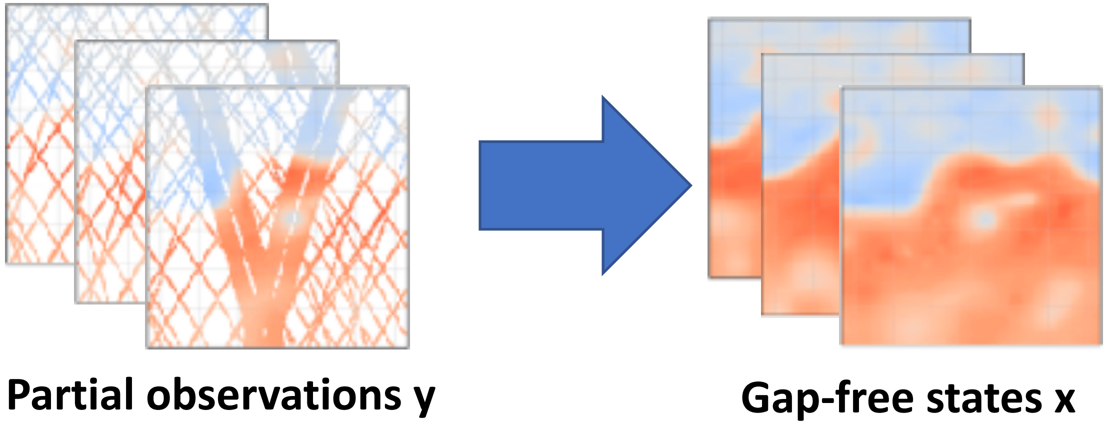




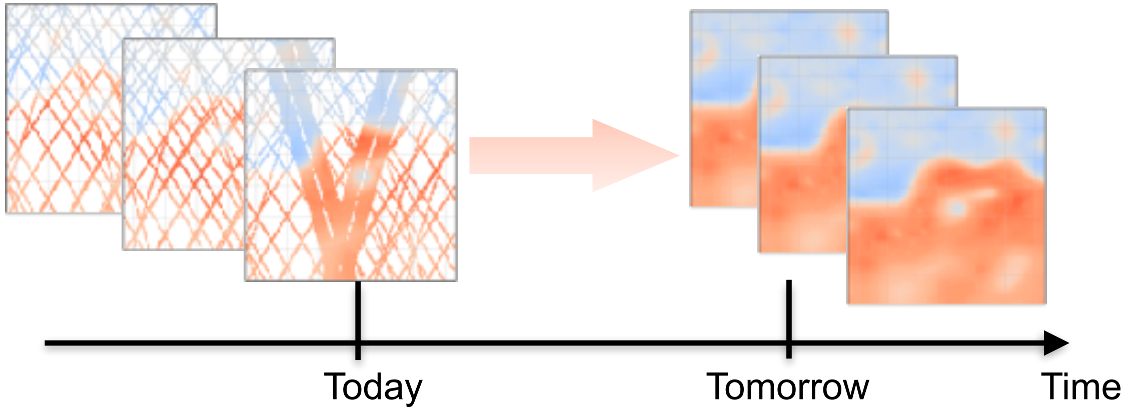


Challenges in space oceanography

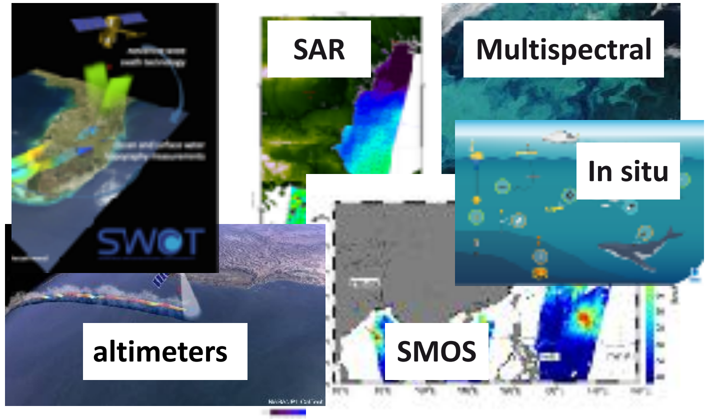
How to reconstruct from observations ?



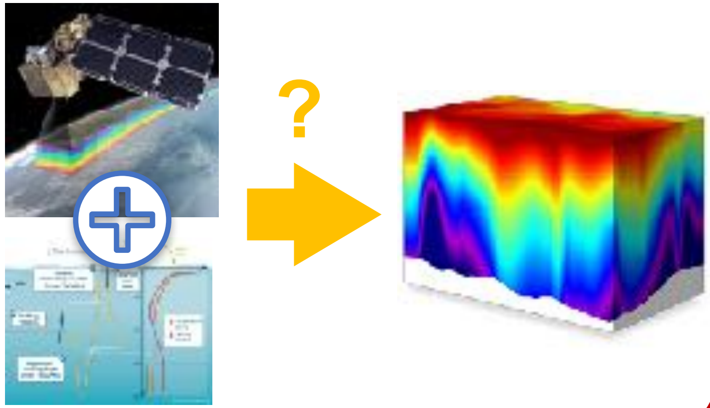
How to simulate/forecast from observations ?



Using multi-tracer observations ?



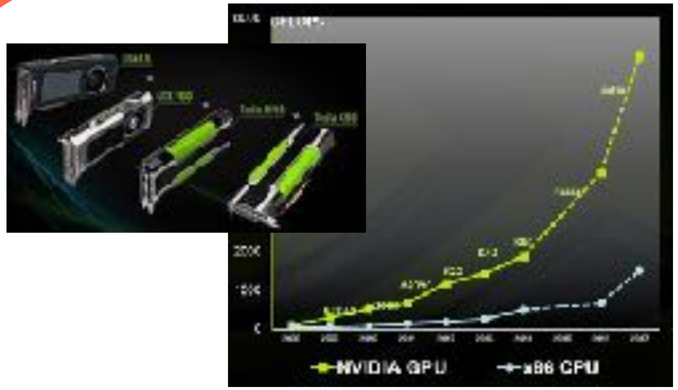
From surface to interior ?



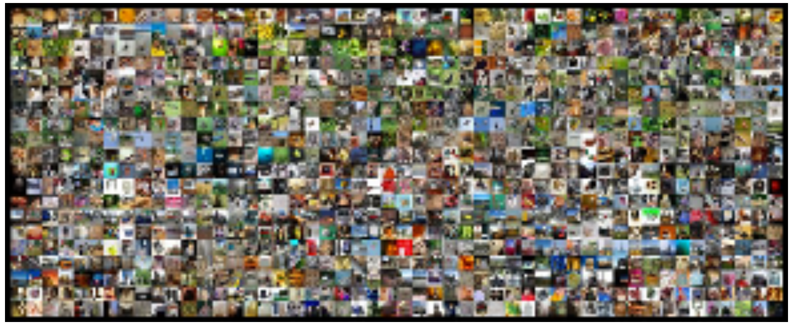
Where and what to sample ?

Deep Learning and Earth Sciences

Key reasons for considering DL



High-performance computing (GPU)



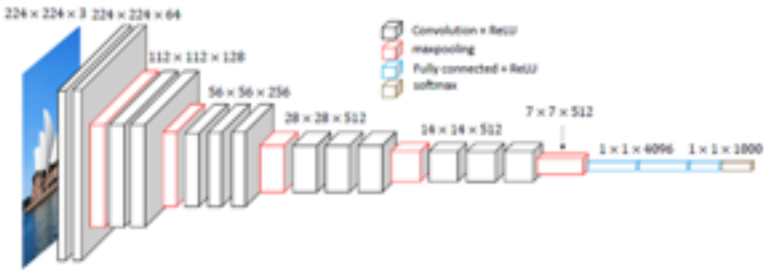
Large annotated dataset (> 1M)

TensorFlow

Keras

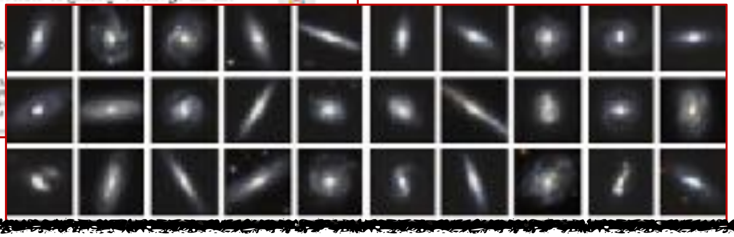
PYTORCH

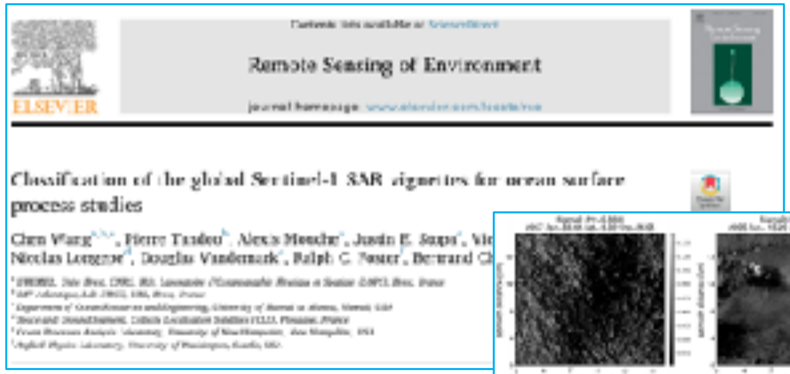
Efficient & easy-to-use frameworks

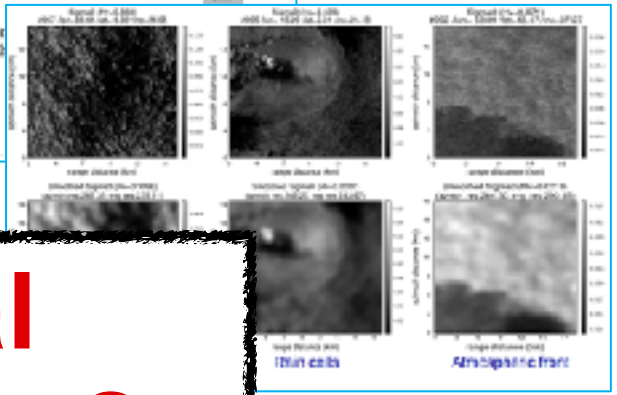


End-to-end learning

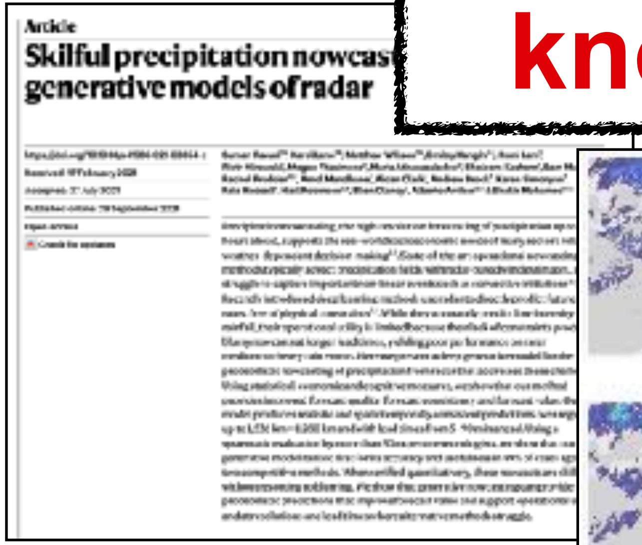
“State-of-the-art” DL schemes applied to physics-related problems

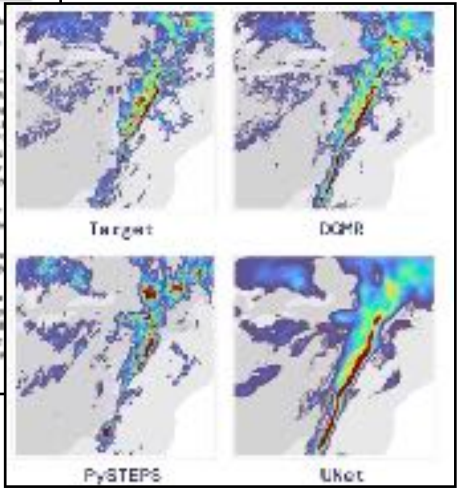


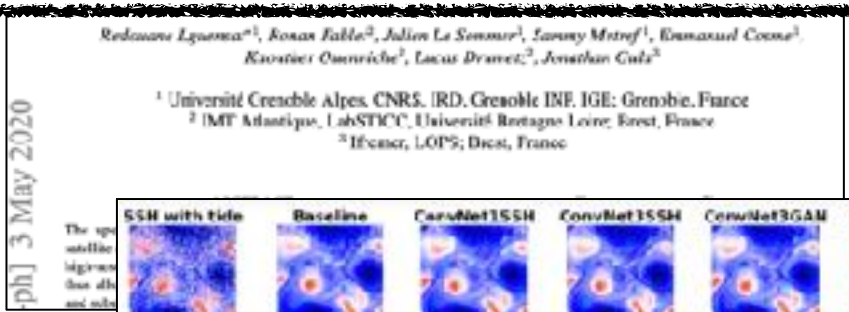


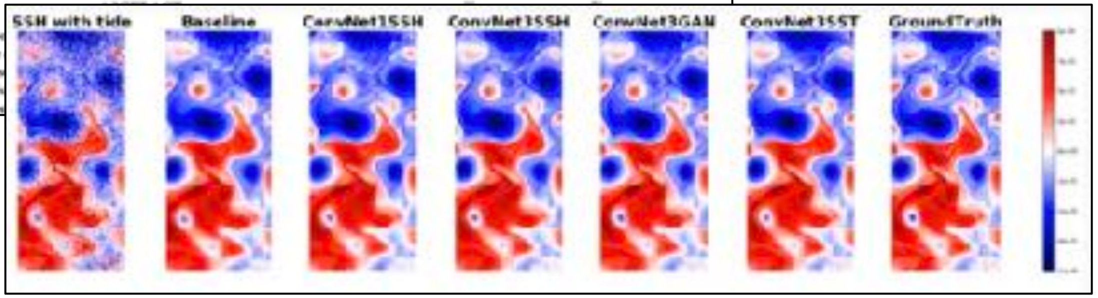


How to embed physical knowledge in DL schemes ?



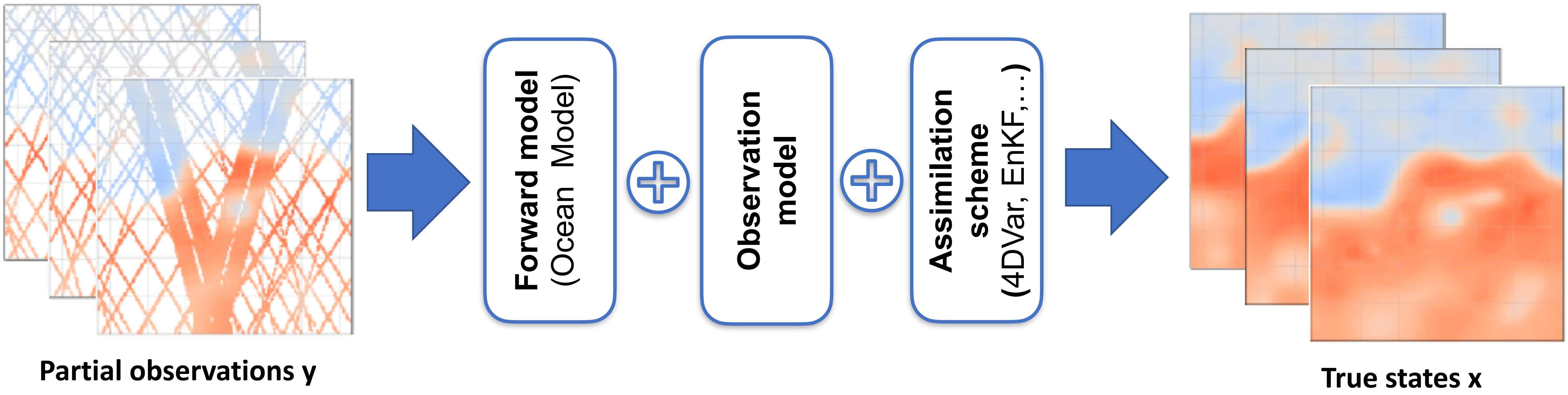




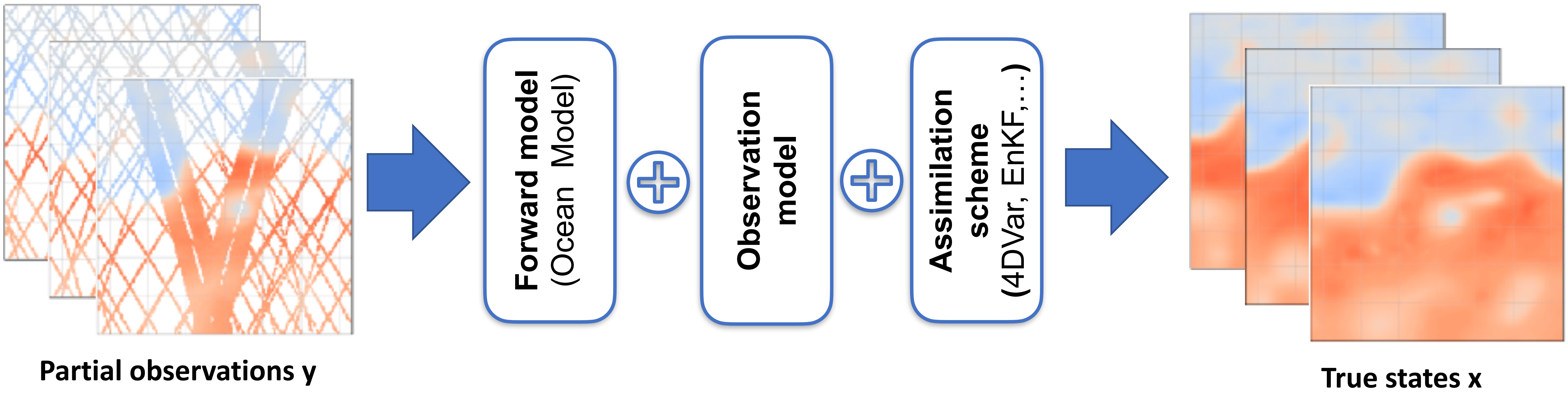


End-to-end physics-informed Learning for reconstruction and forecasting problems

Data assimilation in earth sciences [Evensen, 2000]

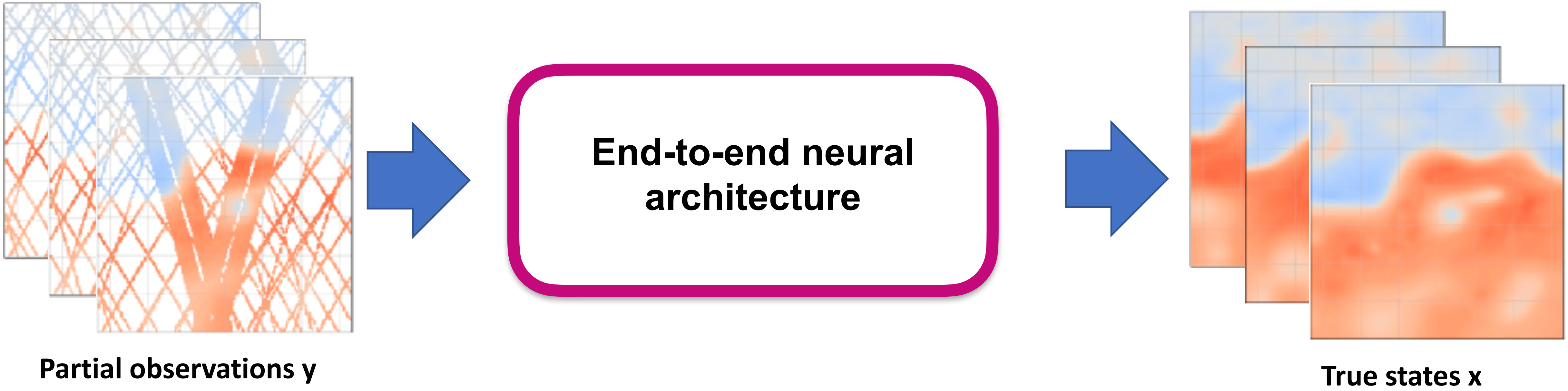


Data assimilation in earth sciences [Evensen, 2000]



Each component designed using model-driven principles and mostly independently…. But somewhat ability to fully exploit observation datasets.

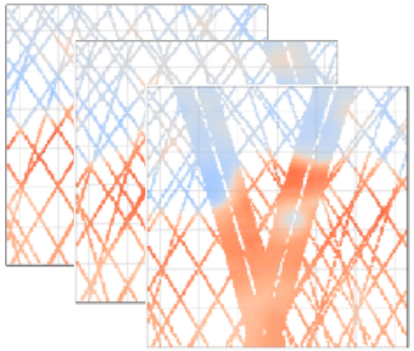
What about end-to-end learning for data assimilation?



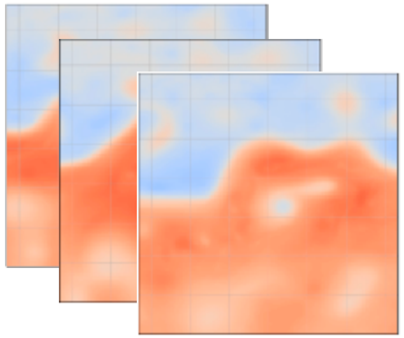
How to exploit prior knowledge for DA ?

Can we calibrate all the components of a DA scheme at once?

(Weak constraint) **4DVar Data Assimilation (DA) formulation**



Partial observations y



True states x

State-space formulation:

$$\begin{cases} \dfrac{\partial x(t)}{\partial t} & = & \mathcal{M}\left(x(t)\right) \\ \\ y(t) & = & x(t) + \epsilon(t), \forall t \in \{t_0, t_0 + \Delta t, \ldots, , t_0 + N\Delta t\} \end{cases}$$

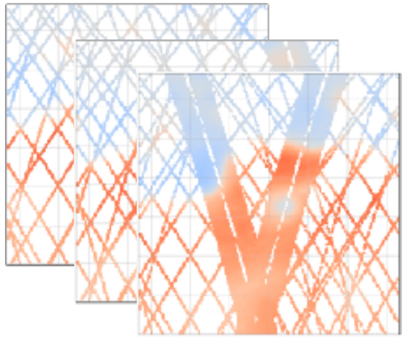
Associated variational formulation:

$$\arg\min_x \lambda_1 \sum_i \|x(t_i) - y(t_i)\|^2_{\Omega_{t_i}} + \lambda_2 \sum_n \|x(t_i) - \Phi(x)(t_i)\|^2$$

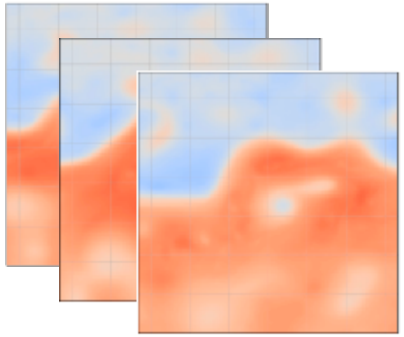
with $$\Phi(x)(t) = x(t-\Delta) + \int_{t-\Delta}^{t} \mathcal{M}\left(x(u)\right) du$$

$$\boxed{\arg\min_x \lambda_1 \|x - y\|^2_\Omega + \lambda_2 \|x - \Phi(x)\|^2}$$

Bridging 4DVar DA and Deep Learning (Fablet et al., 2020)



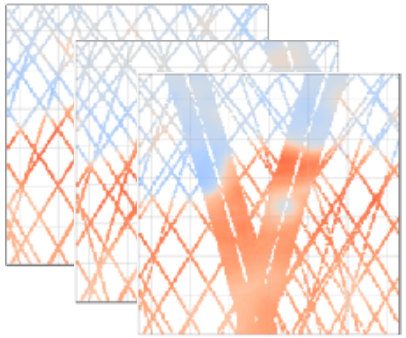
Partial observations y



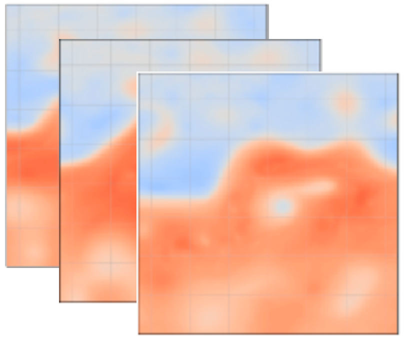
True states x

Model-driven schemes: $\hat{x}_n = \arg\min_x \|x - y_n\|^2_{\Omega_n} + \lambda \|x - \Phi(x)\|^2$

Bridging 4DVar DA and Deep Learning (Fablet et al., 2020)



Partial observations y

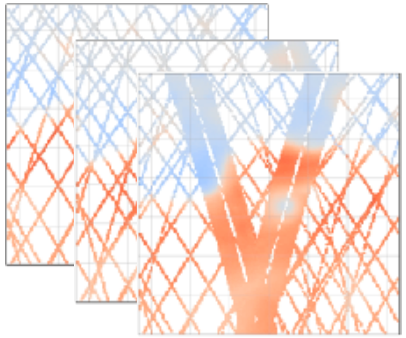


True states x

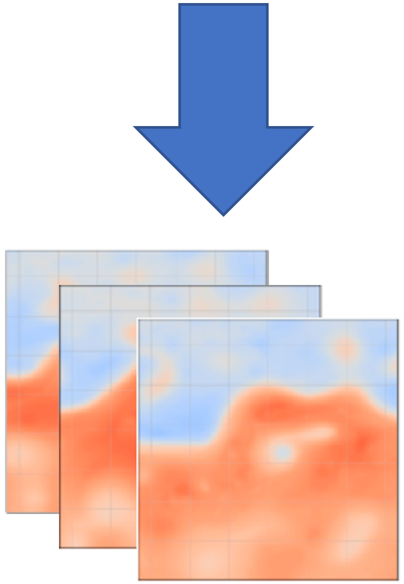
Model-driven schemes: $\hat{x}_n = \arg\min_x \|x - y_n\|^2_{\Omega_n} + \lambda \|x - \Phi(x)\|^2$

Direct learning for inverse problems: $\widehat{x} = \Psi(y)$ $\quad y \longrightarrow \boxed{\text{CNN}} \longrightarrow x$

Bridging 4DVar DA and Deep Learning (Fablet et al., 2020)



Partial observations y



True states x

Model-driven schemes: $\hat{x}_n = \arg\min_x \|x - y_n\|^2_{\Omega_n} + \lambda \|x - \Phi(x)\|^2$

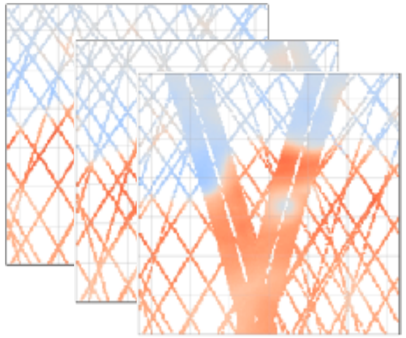
Direct learning for inverse problems: $\widehat{x} = \Psi(y)$ $\qquad y \rightarrow \boxed{\text{CNN}} \rightarrow x$

Proposed scheme: joint learning of the variational model and solver

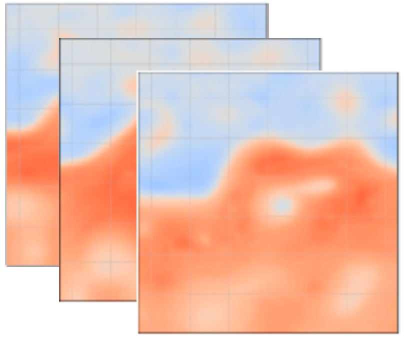
- **Bi-level optimization formulation**

$$\arg\min_\Phi \sum_n \underbrace{\mathcal{L}\left(x_n, \tilde{x}_n\right)}_{\text{Training loss}} \;\text{ with }\; \tilde{x}_n = \arg\min_x \|x - y_n\|^2_{\Omega_n} + \lambda \underbrace{\|x - \Phi(x)\|^2}_{\text{Neural Network implementation}}$$

Bridging 4DVar DA and Deep Learning (Fablet et al., 2020)



Partial observations y



True states x

Model-driven schemes: $\widehat{x} = \arg\min_{x} \lambda_1 \|x - y\|^2_{\Omega} + \lambda_2 \|x - \Phi(x)\|^2$

Direct learning for inverse problems: $\widehat{x} = \Psi(y)$ $\quad y \rightarrow \boxed{\text{CNN}} \rightarrow x$

Proposed scheme: joint learning of the variational model and solver

- **Bi-level optimization formulation**

$$\arg\min_{\Phi} \sum_{n} \mathcal{L}\left(x_n, \tilde{x}_n\right) \ \text{with} \ \tilde{x}_n = \arg\min_{x} \|x - y_n\|^2_{\Omega_n} + \lambda \|x - \Phi(x)\|^2$$

- **Restated with a gradient-based NN solver for inner minimization**

$$\arg\min_{\Phi} \sum_{n} \mathcal{L}\left(x_n, \tilde{x}_n\right) \ \text{with} \ \tilde{x}_n = \Psi_{\Phi,\Gamma}\left(x_n^{(0)}, y_n, \Omega_n\right)$$

Trainable iterative gradient-based solver using automatic differentiation to compute gradient of $\|x - y_n\|^2_{\Omega_n} + \lambda \|x - \Phi(x)\|^2$

4DVarNet: Learning 4DVar models and solvers

Trainable Variational DA formulation

$$\widehat{x} = \arg\min_{x} \|y - H(x)\|^2 + \lambda \|x - \Phi(x)\|^2$$

Trainable or pre-defined observation model

Trainable or pre-defined prior

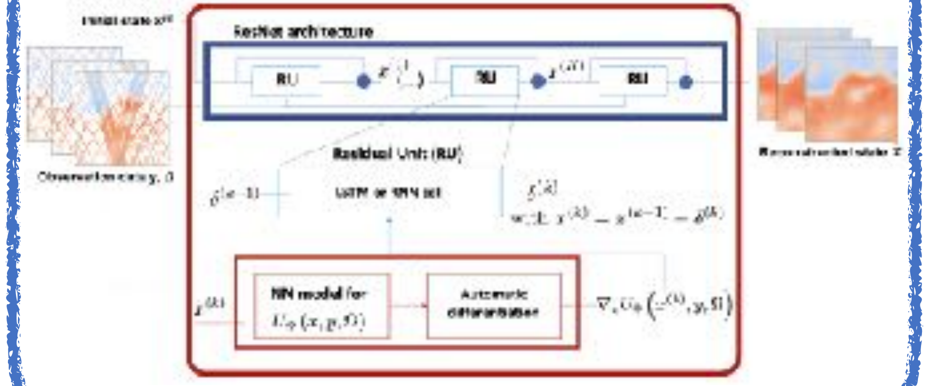
Trainable solver

$$x^{(k+1)} = x^{(k)} - \mathcal{H}\left[\nabla_x U_\Phi\left(x^{(k)}, y\right)\right]$$

LSTM

Automatic differentiation

End-to-end architecture



Preprint: https://arxiv.org/abs/2006.03653
Code: https://github.com/CIA-Oceanix/4dvarnet-core

4DVarNet: Learning 4DVar models and solvers

Trainable Variational DA formulation

$$\widehat{x} = \arg\min_{x} \|y - H(x)\|^2 + \lambda \|x - \Phi(x)\|^2$$

Trainable or pre-defined observation model

Trainable or pre-defined prior

Learning criterion

Variational cost (non-supervised)

Reconstruction error (supervised)

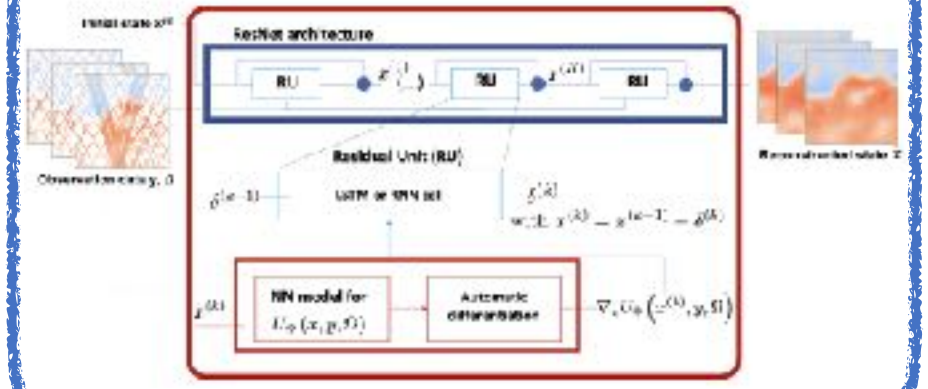
Trainable solver

$$x^{(k+1)} = x^{(k)} - \mathcal{H}\left[\nabla_x U_\Phi\left(x^{(k)}, y\right)\right]$$

LSTM

Automatic differentiation

End-to-end architecture

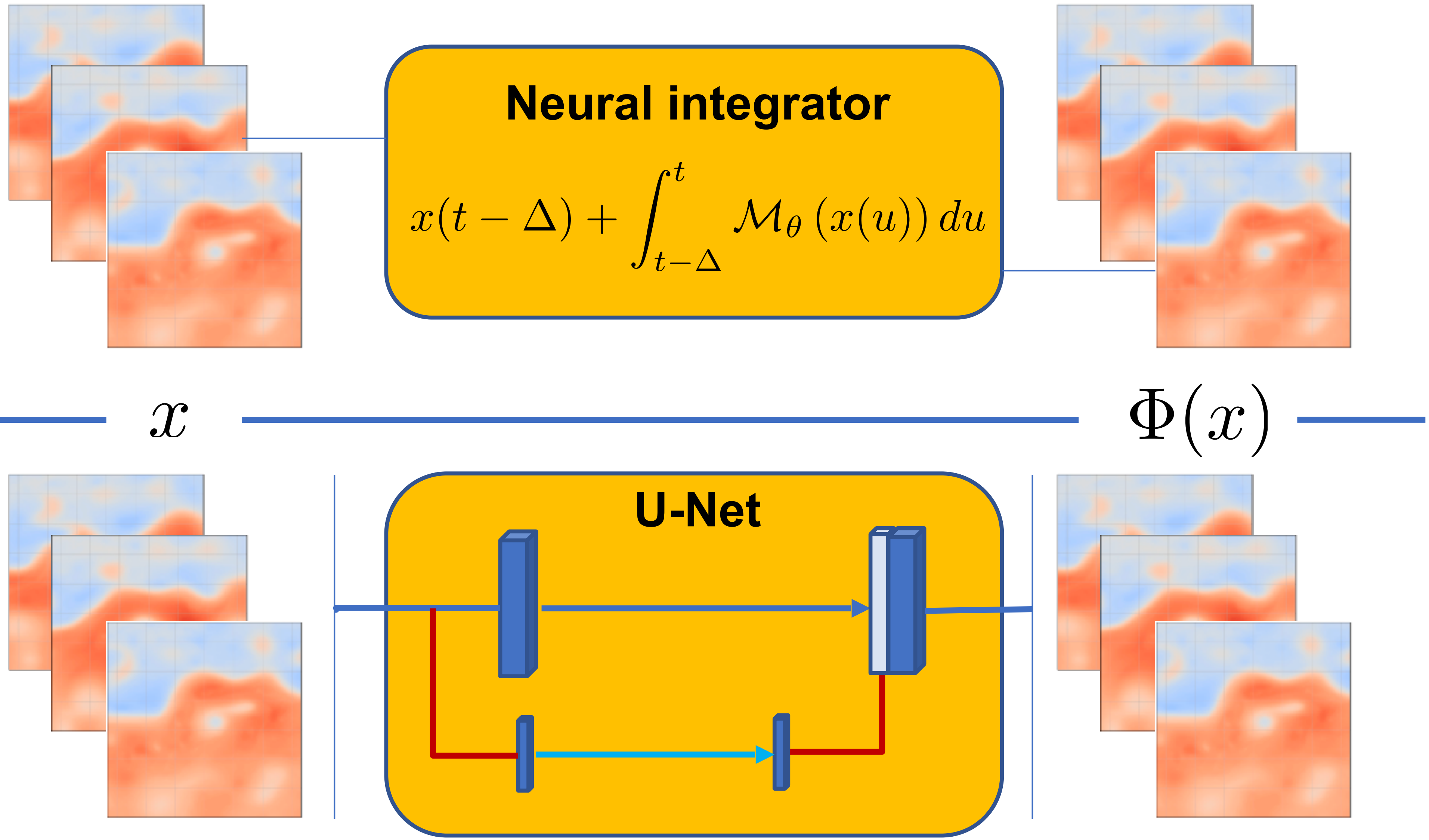


Preprint: https://arxiv.org/abs/2006.03653
Code: https://github.com/CIA-Oceanix/4dvarnet-core

4DVarNet: which projection operator Φ?

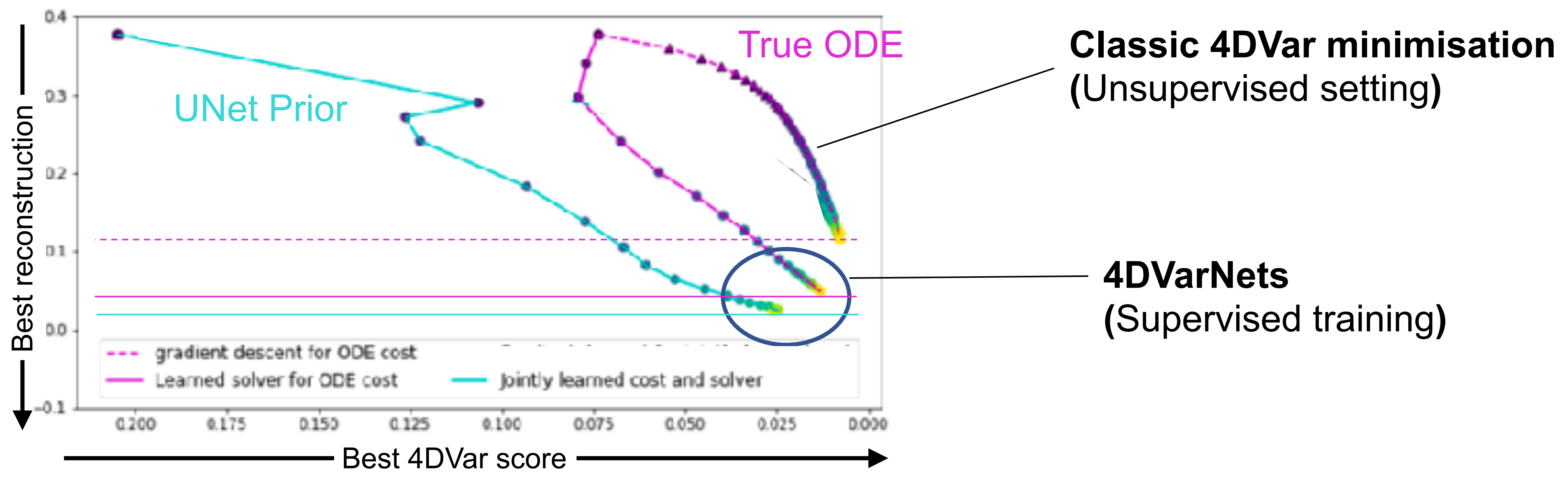
Parameterization using (trainable) ODE operator

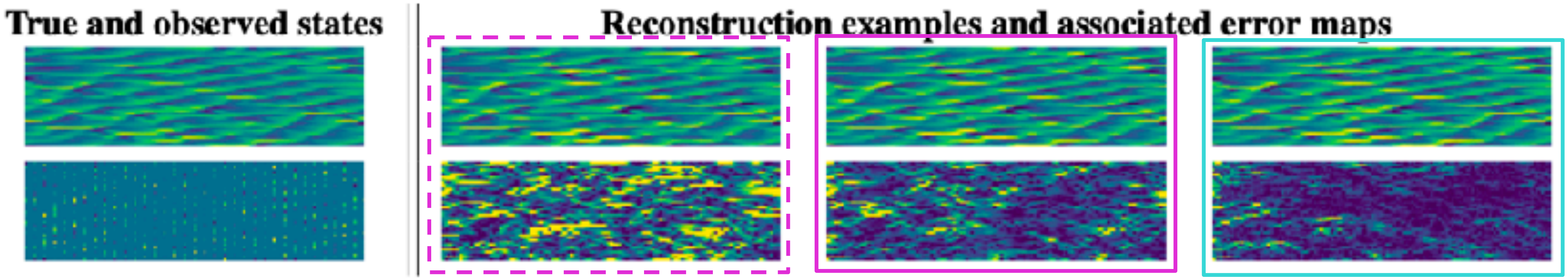
$$\frac{\partial x(t)}{\partial t} = \mathcal{M}_\theta\left(x(t)\right)$$



Two-scale U-Net-like Parameterization (similar to Gibbs/ Markov Random Field)

Application to Lorenz-96 system





The true dynamical model may not be the best prior for DA ?

4DVarNet: Application to sea surface dynamics

Method

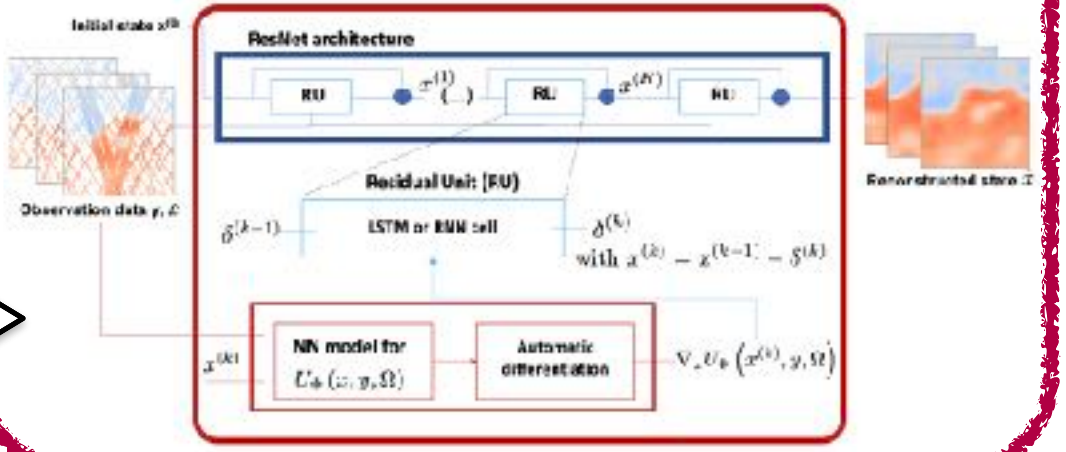
From a Variational DA formulation

$$\widehat{x} = \arg\min_{x} \|y - H(x)\|^2 + \lambda \|x - \Phi(x)\|^2$$

Trainable variational model

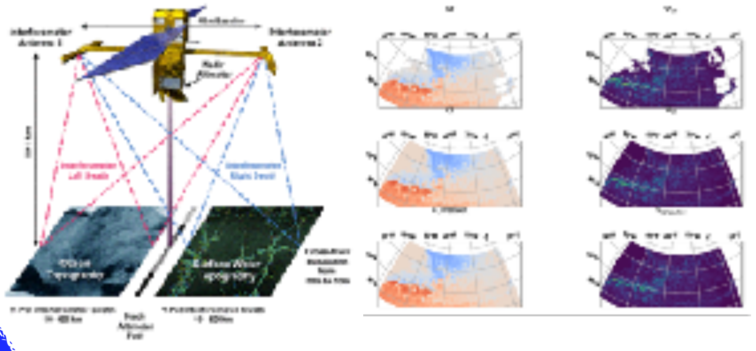
Trainable gradient-based solver

Associated end-to-end scheme

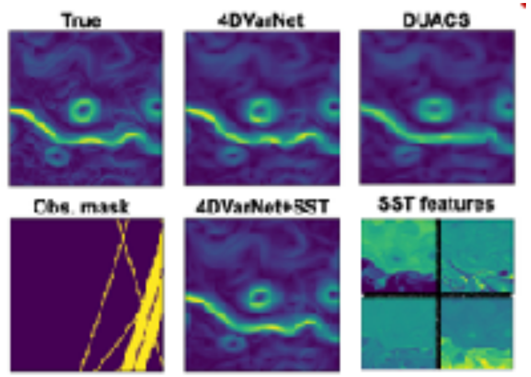


Applications

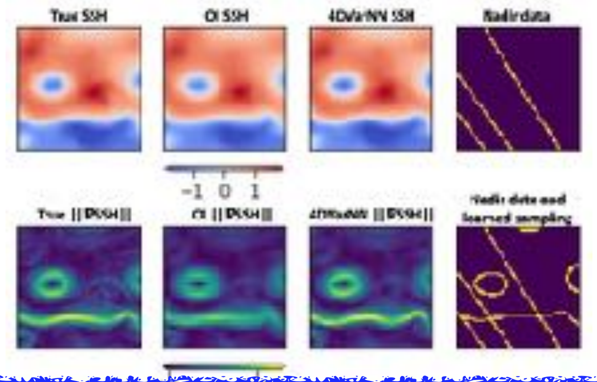
Interpolation & Forecasting



Multimodal DA

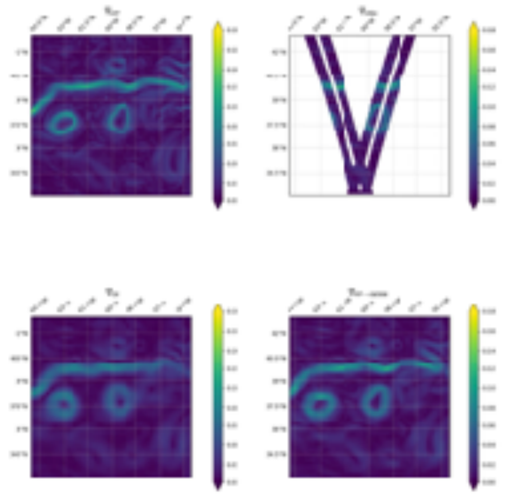


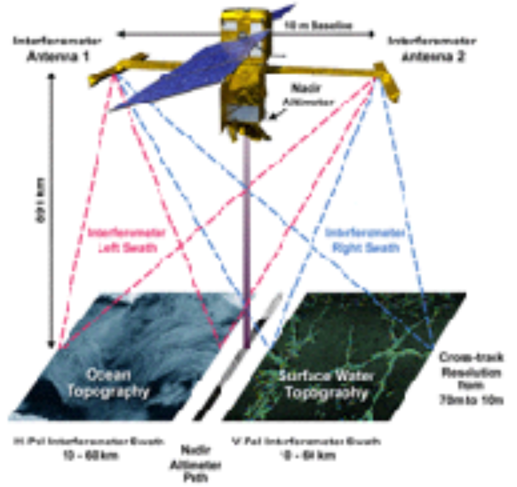
Learning where to sample ?



Space-time interpolation of sea surface geophysical fields

Satellite altimetry





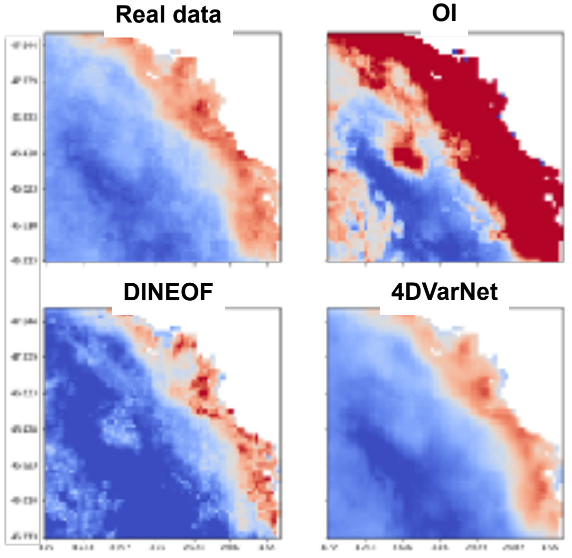
Best score for BOOST-SWOT SLA Data Challenge

duacs 1 swot + 4 nadirs	0.92	0.02	1.22	11.15	Covariances DUACS	eval_duacs.ipynb
bfn 1 swot + 4 nadirs	0.93	0.02	0.8	10.09	QG Nudging	eval_bfn.ipynb
dymost 1 swot + 4 nadirs	0.93	0.02	1.2	10.07	Dynamic mapping	eval_dymost.ipynb
miost 1 swot + 4 nadirs	0.94	0.01	1.18	10.14	Multiscale mapping	eval_miost.ipynb
4DVarNet 1 swot + 4 nadirs 🏆	0.96	0.01	0.82	6.57	4DVarNet mapping	eval_4dvarnet.ipynb

https://github.com/ocean-data-challenges/2020a_SSH_mapping_NATL60

Sea surface suspended sediments

Metric	Dataset	Unit	Samp. Strat	OI	DinEOF	4DVarNet
RMSE	OSSE	$\log_{10}[g/L]/m$	-	0.176	0.167	**0.104**
	MODIS	$\log_{10}[g/L]/m$	Random	0.304	0.237	**0.156**
	MODIS	$\log_{10}[g/L]/m$	Patch	0.346	0.253	**0.168**
R-score	OSSE	%	-	90.4	91.3	**96.6**
	MODIS	%	Random	60.5	76.4	**89.5**
	MODIS	%	Patch	56.5	73.8	**87.3**



Mean gradient norm

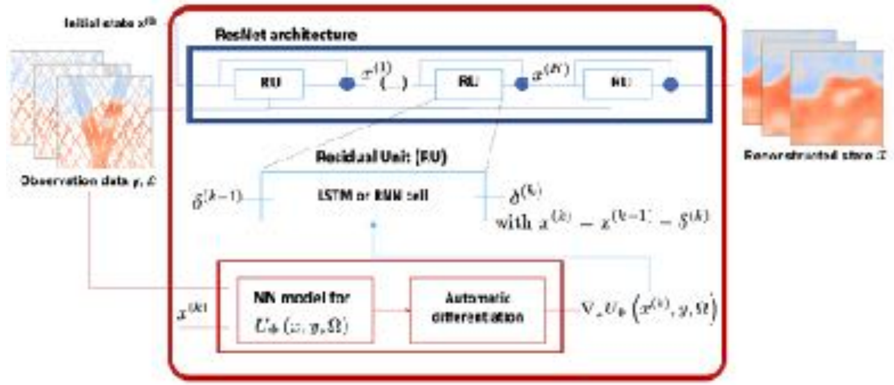
Simulation and real data

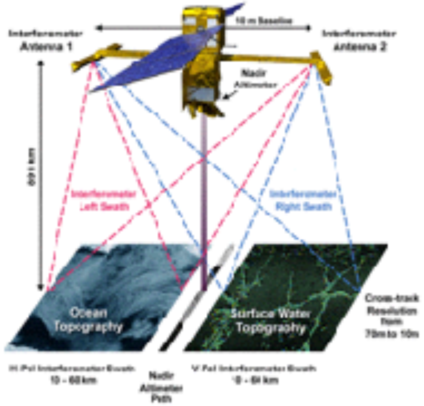
Learning from real gappy data only

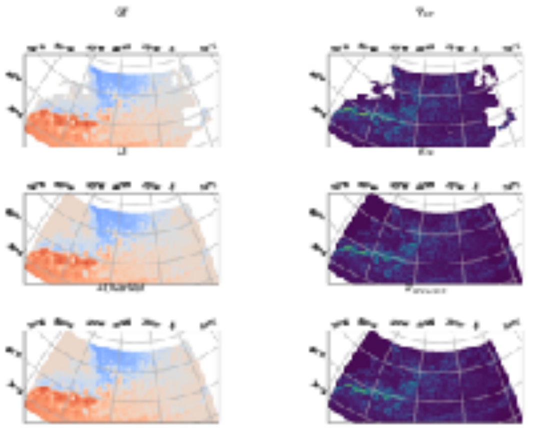
Collab. IMT Atl./SHOM/ LGO

Key messages

- **Beyond NNs as data-driven-only black bloxes**
- **Trainable variational DA models** (observation model, prior, solver)
- **Application to interpolation, forecasting sampling and multimodal synergies**
- **End-to-end learning makes it easier**
- **Scaling up to the global scale using**

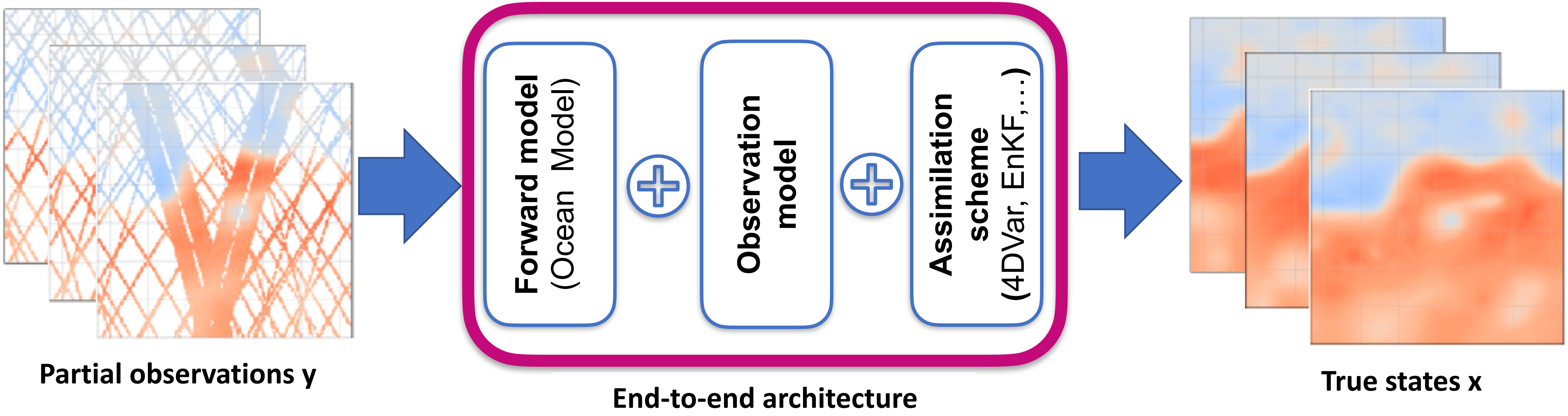






Paper (doi): [10.1029/2021MS002572](https://doi.org/10.1029/2021MS002572)
Code: https://github.com/CIA-Oceanix/4dvarnet-core

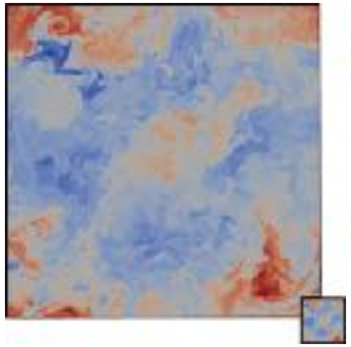
Deep Learning and DA: insights and future work



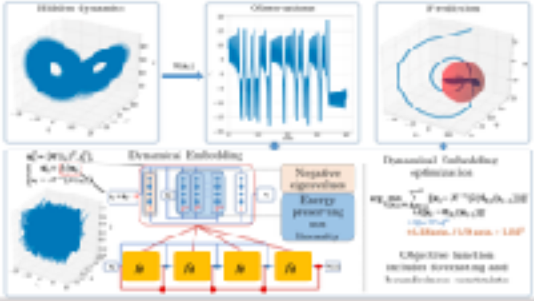
- 4DVarNet = Differentiable Emulator for Data Assimilation
- Is the true forward model always the "optimal" choice? Which (differentiable) priors for which observing systems and/or variables?
- How to embed uncertainty?

End-to-end deep and physics-informed learning and dynamical systems (cia-oceanix.github.io)

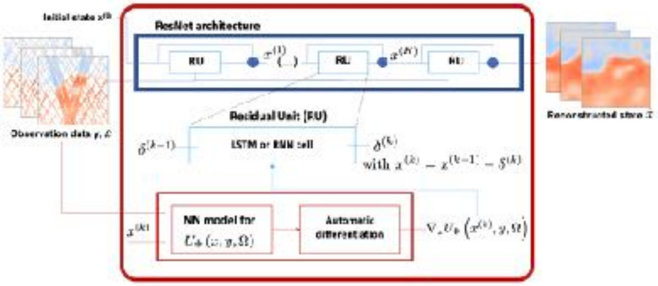
Learning & Simulation



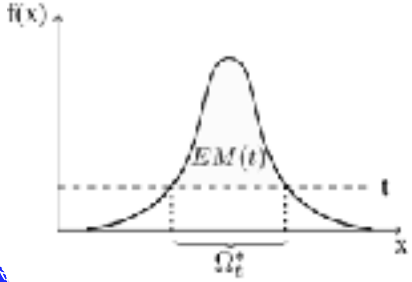
Observation-driven forecasting



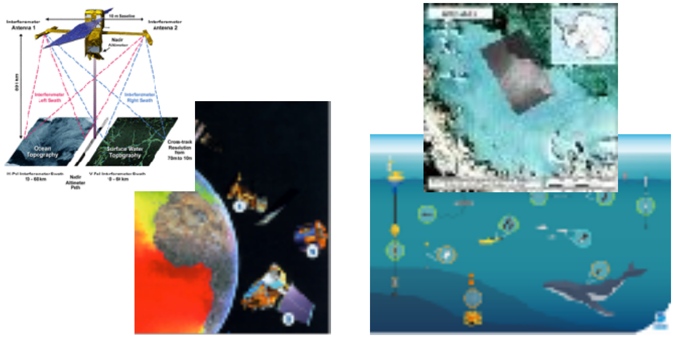
Learning & Data Assimilation



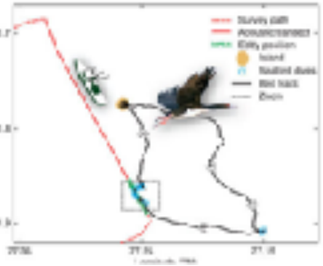
Learning for geophysical extremes



Multimodal observations



Trajectory data modelling and analysis



End-to-end learning & Geophysical Dynamics

End-to-end (a posteriori) learning to better train stable neural closures (Frezat et al., ML4PS 2021)

DNS

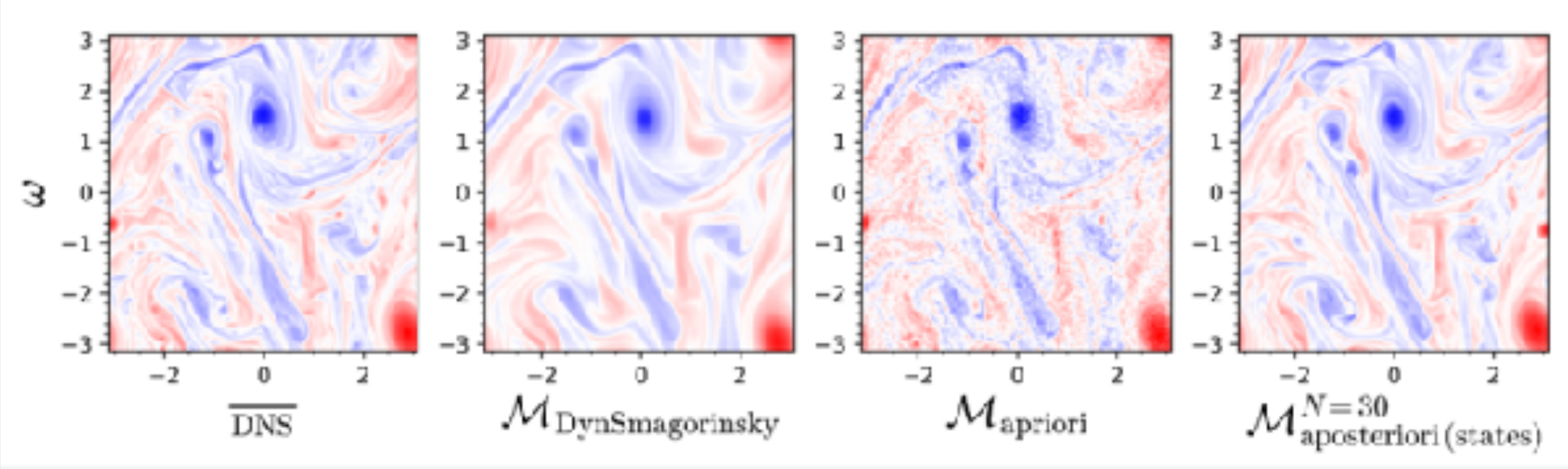
vs.

LES

$$\begin{cases} \dfrac{\partial X_t}{\partial t} = f\left(X_t\right) + g\left(Z_t\right) \\ \dfrac{\partial \overline{X}_t}{\partial t} = f\left(\overline{X}_t\right) + g\left(Z_t\right) + \boxed{h\left(\overline{X}_t\right)} \end{cases}$$

Known terms

unknown term
(Closure)



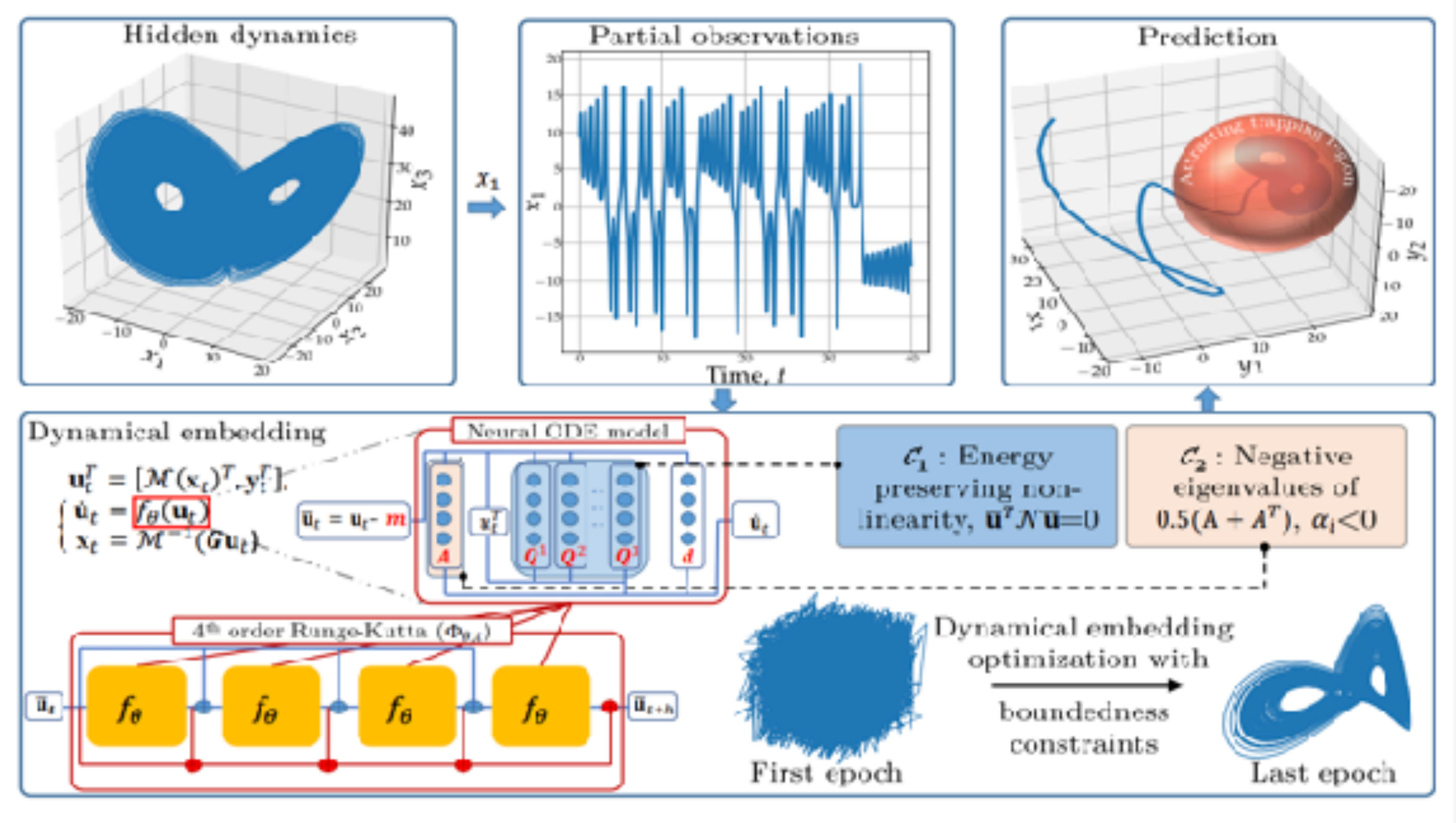
SQG case-study: example of simulated vorticity state using different closures

https://arxiv.org/pdf/2111.06841.pdf

End-to-end learning & Geophysical Dynamics

Learning bounded forward models for partially-observed systems

*(*Ouala et al., 2022*)*



Augmented Neural ODE

Bilinear ODE with boundedness constraint

End-to-end-learning to match observed states

https://arxiv.org/abs/2202.05750

AI Chair OceaniX 2020-2024

Physics-informed AI for Observation-Driven Ocean AnalytiX

Thank you.

R. Fablet, Prof. IMT Atlantique, Brest
Web: https://cia-oceanix.github.io/

References

- **Model calibration/identification**
  - Frezat et al. Physics-informed neural networks for sub-grid scale modeling in filtered turbulence. PRF, 2021. https://arxiv.org/abs/2010.04663
  - Pannekoucke et al. PDE-NetGen 1.0: from symbolic PDE representations of physical processes to trainable neural network representations. GMD, 2020. https://gmd.copernicus.org/articles/13/3373/2020/
  - Ouala et al. Learning Runge-Kutta Integration Schemes for ODE Simulation and Identification, aRxiv, 2021. https://arxiv.org/abs/2105.04999

- **Data assimilation and Forecasting**
  - Ouala et al. Learning Latent Dynamics for Partially-Observed Chaotic Systems. Chaos, 2020. https://arxiv.org/abs/1907.02452
  - Fablet et al. Joint interpolation and representation learning for irregularly-sampled satellite-derived geophysical fields. FAMS, 2021. https://doi.org/10.3389/fams.2021.655224
  - Fablet et al. Learning Variational Data Assimilation Models and Solvers. JAMES, 2021. https://arxiv.org/abs/2007.12941

- **Trajectory data analysis and modeling**
  - Nguyen et al. GeoTrackNet-A Maritime Anomaly Detector using Probabilistic NN Representation of AIS Tracks and A Contrario Detection. IEEE TITS, 2020. https://arxiv.org/abs/1912.00682
  - Roy et al. Deep Learning and Trajectory Representation for the Prediction of Seabird Diving Behaviour. biorRxiv, 2020. https://www.biorxiv.org/content/10.1101/2021.04.19.438554v1
  - A. Roy, S. Lanco Bertrand, **R. Fablet.** Generative Adversarial Networks (GAN) for the simulation of central-place foraging trajectories. Meth. Ecol. Ev., 2022. https://doi.org/10.1101/2021.09.27.461940

- **Applications to upper ocean dynamics**
  - Fablet et al. End-to-end physics-informed representation learning for satellite ocean remote sensing data: applications to satellite altimetry and sea surface currents. ISPRS APRSSI, 2021. https://doi.org/10.5194/isprs-annals-V-3-2021-295-2021
  - Vient et al. Data-Driven Interpolation of Sea Surface Suspended Concentrations Derived from Ocean Colour Remote Sensing Data. RS 2021. https://doi.org/10.3390/rs13173537
  - Beauchamp, et al. Intercomparison of data-driven and learning-based interpolations of along-track nadir and wide-swath altimetry observations. RS, 2020. https://doi.org/10.3390/rs12223806