Best Initial Condition Estimation Using a Data-Driven Ocean Emulator

Gradient Descent-Based Variational Data Assimilation Framework

Jungwon
LEE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

RÉPUBLIQUE FRANÇAISE

Co-funded by the European Union

MERCATOR OCEAN INTERNATIONAL

This project has received co-funding from the European Union's Horizon Europe Research and Innovation Program under Grant Agreement No. 101126644. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting

Structure

I. Introduction
- Motivation
- Research question

II. Strong-Constraint 4D-Var
- Best Initial Condition Estimation
- Experiment

III. Strong 4D-Var with learned optimizer
- Meta Learner

IV. Conclusions and perspectives

Structure

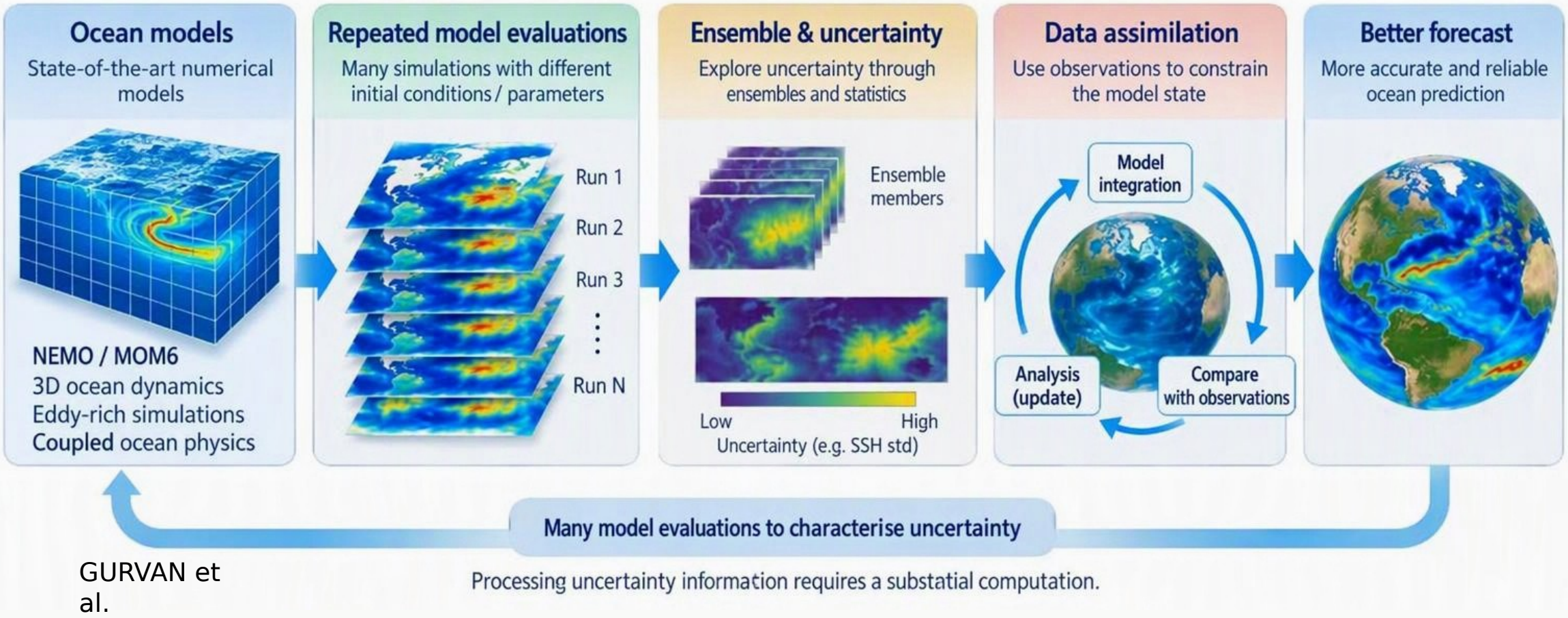
I.Introduction
Motivation
Research question

II. Strong-Constraint 4D-Var
Best Initial Condition Estimation
Experiment

III. Strong 4D-Var with learned optimizer
Meta Learner

IV. Conclusions and perspectives

NEED OF OCEAN SURROGATE MODEL

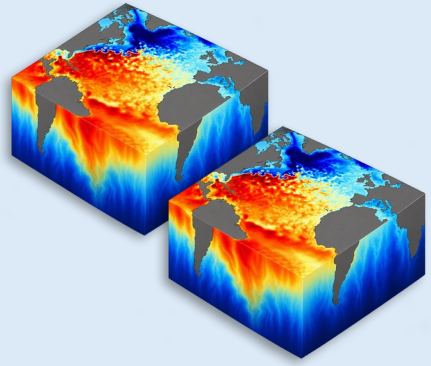


GURVAN et al.

Better State Estimation Requires Repeated Differentiable Model Evaluations

GLONET

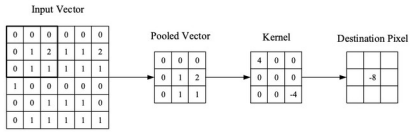
Input



Ψ_{t-1}, Ψ_t

GLONET EL AOUNI et al.

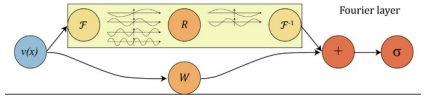
CNN



Capturing local features

O'SHEA et al.

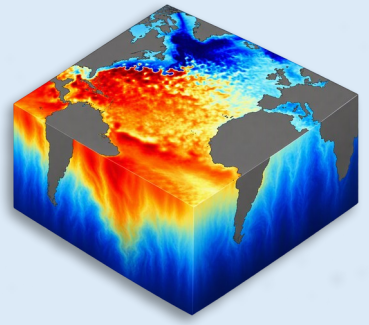
FNO



Capturing long-range dependencies

LI et al.

Output



Ψ_{t+1}

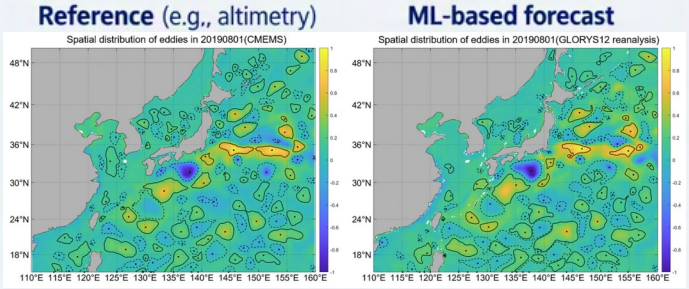
Keywords | **global** | **3D multivariate** | **Autoregressive** | **differentiable**

GLONET as a Differentiable Global 3D Ocean Emulator

METHODOLOGICAL GAP – RESEARCH QUESTION

1 Existing evaluations focus mainly on forecasting skill

Most ML-based ocean models are assessed in free-running forecast experiments.

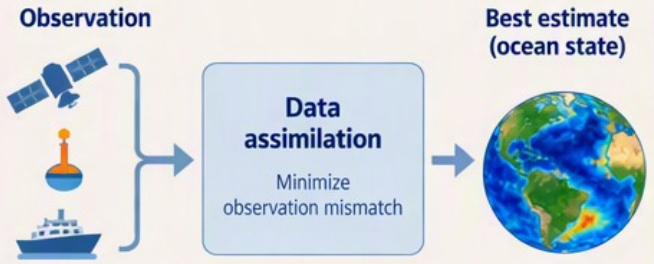


- Evaluation metrics: RMSE, ACC, etc.
- Focus on short- to medium-term forecasts
- Demonstrated promising skill for many variables (e.g., SSH, SST, currents)

Wang et al.

2 Use in data assimilation remains largely unexplored

Operational and research DA systems mainly rely on numerical models with established adjoint/gradient capabilities.

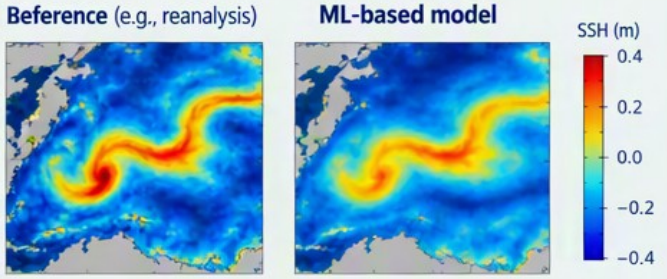


- End-to-end, real-time ML-ODA integration remains rare.
- Realistic multivariate and 3D scenarios remain largely untested.
- Full-cycle evaluation is limited, with cycling robustness largely untested.

GRANDE et al.

3 ML-based models can have biases due to training data

Emulators learn from data and may inherit systematic biases or miss rare/extreme dynamics.



- Training data limitations (coverage, resolution, errors)
- Systematic biases and smoothing of variability
- Difficulty in representing rare/extreme events
- Impact of model biases on DA performance is unclear

Research Question

Can we exploit the **differentiability of an ML-based ocean emulator** to directly **optimize its initial condition** using future **observational information**?

Structure

I. Introduction
   Motivation
   Research question

II. Strong-Constraint 4D-Var
Best Initial Condition Estimation
Experiment

III. Strong 4D-Var with learned optimizer
   Meta Learner

IV. Conclusions and perspectives

OPTIMAL INITIAL CONDITION

Strong 4D-Var Best Initial Condition Estimation (BICE)

Objective

$$\underset{\psi_0}{argmin}\, J_{obs}(\psi_0)$$

Cost function

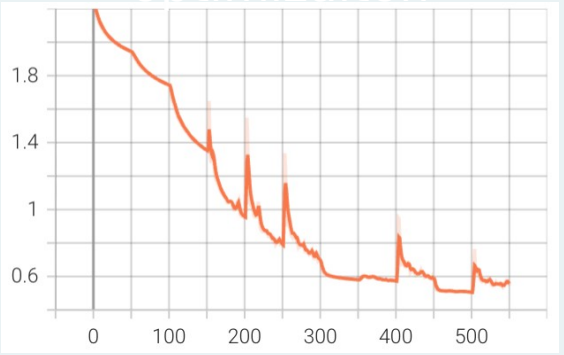
$$J_{obs}(\psi_0) = \sum_{t=0}^{\tau} ¿¿¿¿$$

From varational cost funciton, only the obsevation term is used to build an objective function of the strong 4D-Var BICE. For simplicity obsevation operator H was implemented as a selection matrix and the observation error covariance matrix is negliged

$$\nabla_{\psi_0} J(\psi_0^{(k)})$$

backpropagation

Gradient descent optimizaiton



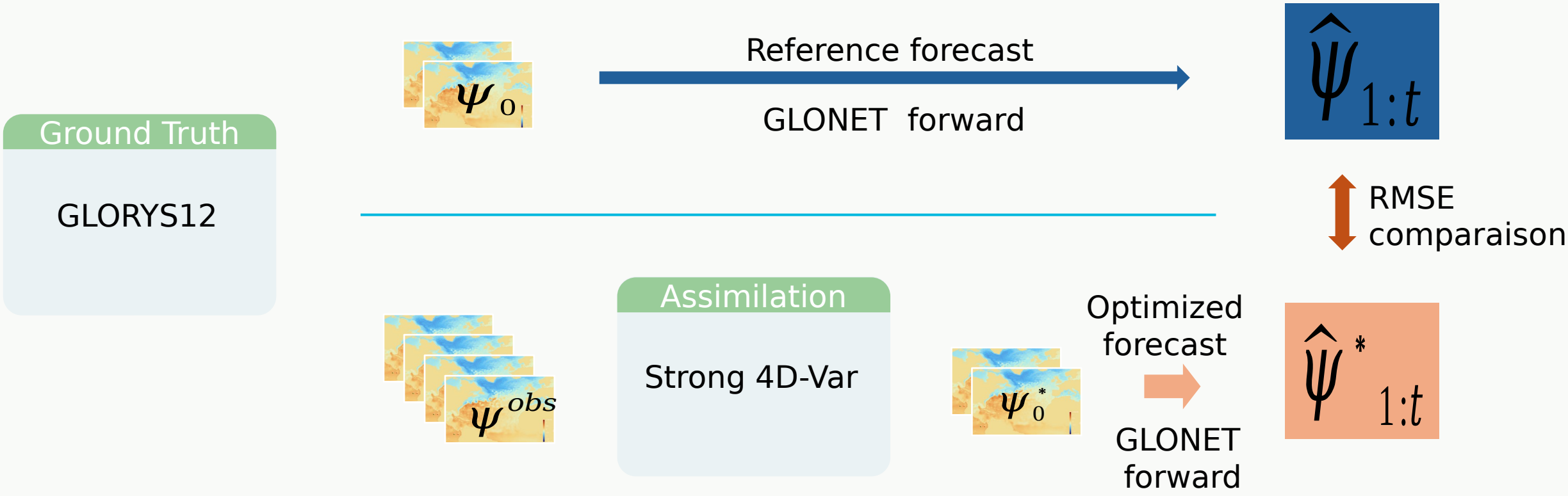
$$\psi_0^{(k+1)} = \psi_0^{(k)} - f(\nabla)$$

The initial condition can be updated by a funciton of the gradient of the cost function with regards to the initial condition. The function can help to numerical convergece of the optimization

iteration until converge

EXPERIEMENT DESIGN: IMPACT OF OPTIMAL IC

Objective: Assess whether 4D-Var initial-condition optimization through the differentiable surrogate improves **state estimation and subsequent forecast skill**



EXPERIEMENT DESIGN: IMPACT OF OPTIMAL IC

Can an optimized initial condition improve the subsequent forecast?

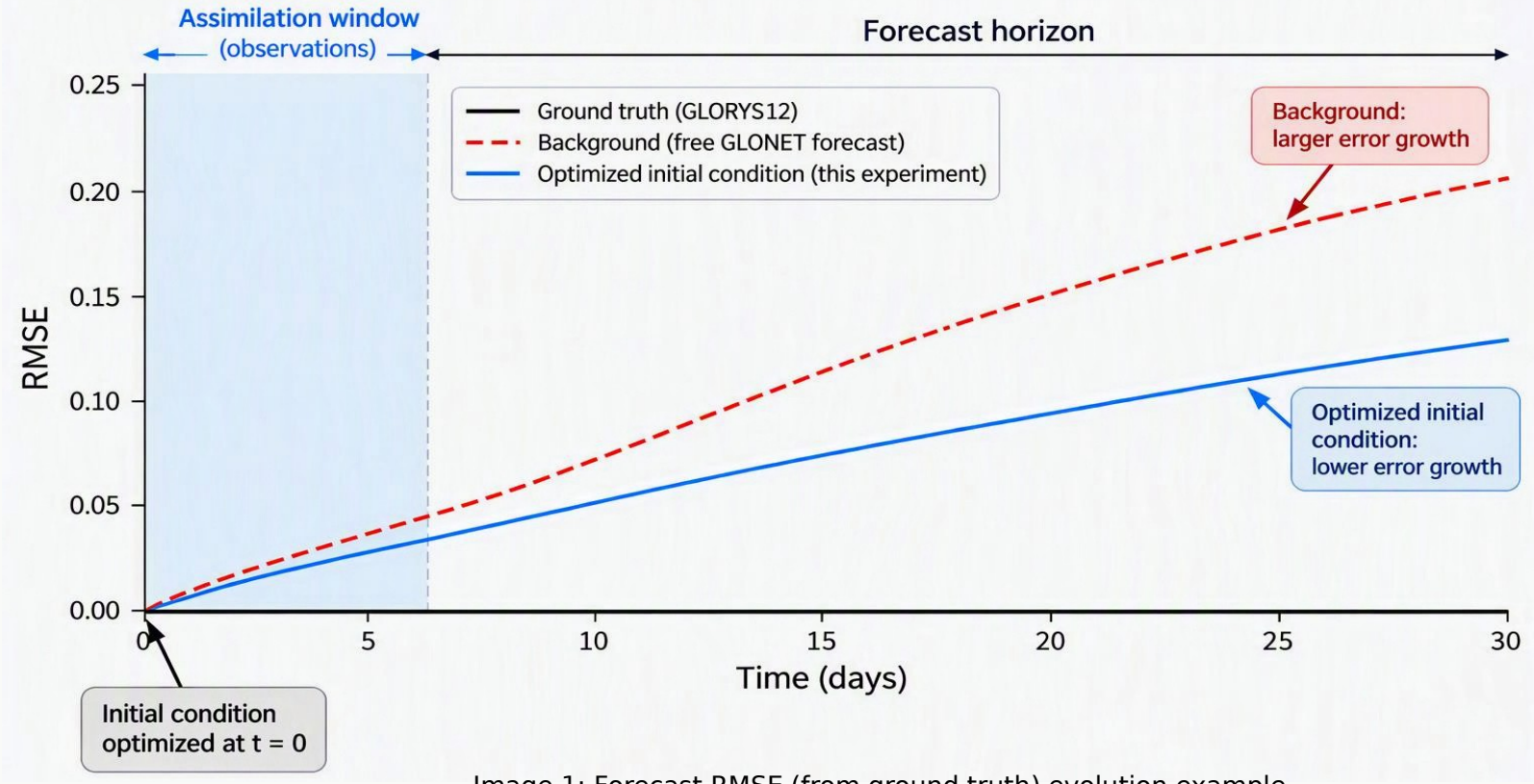
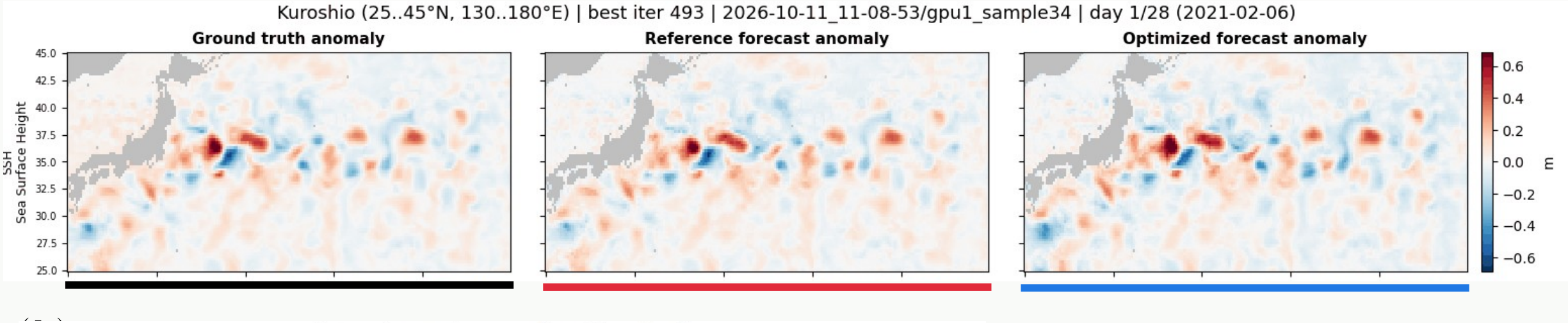


Image 1: Forecast RMSE (from ground truth) evolution example

BASELINE CONFIGURATION

Parameter/hyper-parameter	Reference Value
Emulator	GLONET ¼
Assimilation window	7 days
Variables	SSH, SST, SSS, SSU and SSV
Observation type	Fully observed Ground truth by day
Optimizer	Adagrad
Learning rate	10
Iterations	550
Objective function	Variational observation mismatch
Evaluation region	Global

OPTIMIZED FORECAST VISUALIZATION



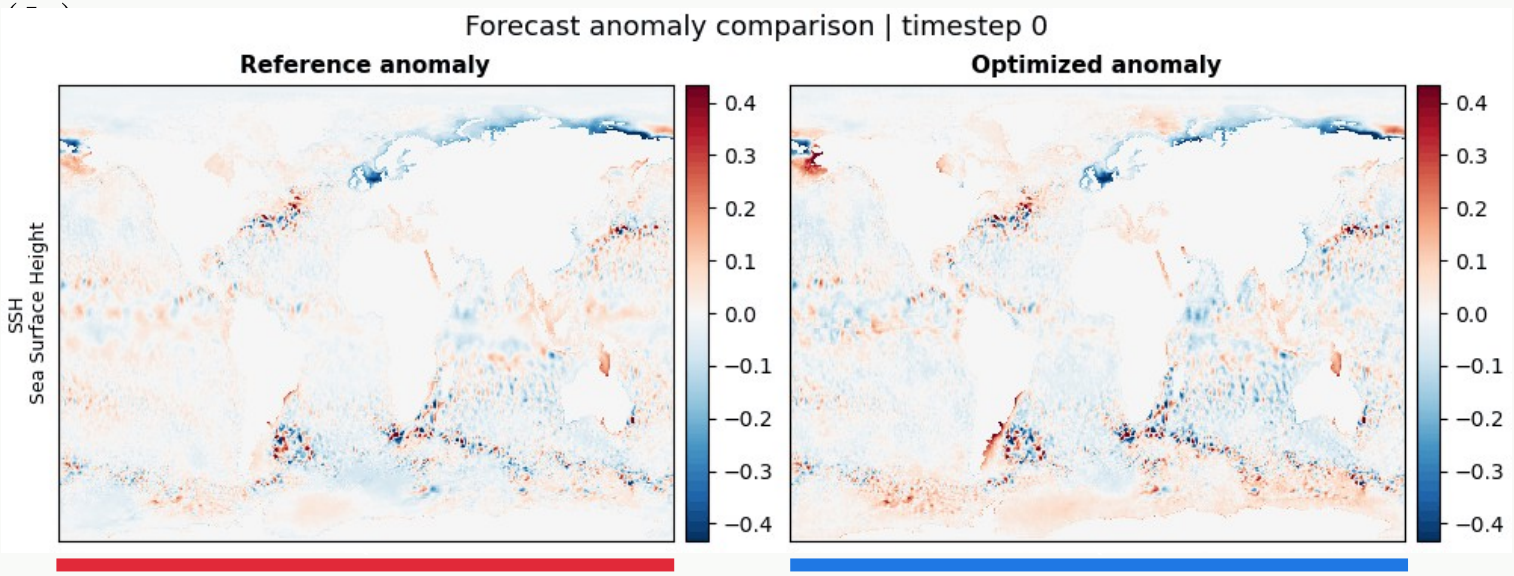


Image 2: SSH anomaly in forecast lead time (a) Kuroshio region and (b) global

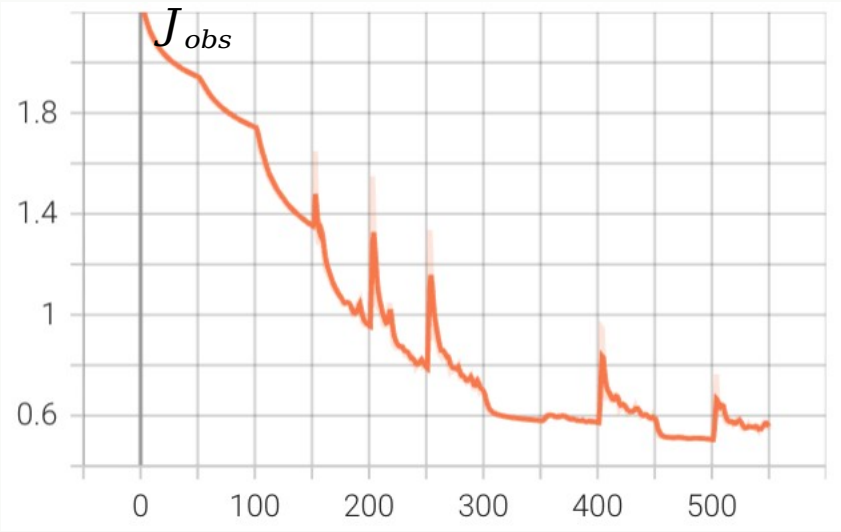


Image 3: cost Function J optimization curve

Q1. Can BICE Bring Better Ocean Forecast?

The Optimized Initial Condition Improves the Subsequent Forecast

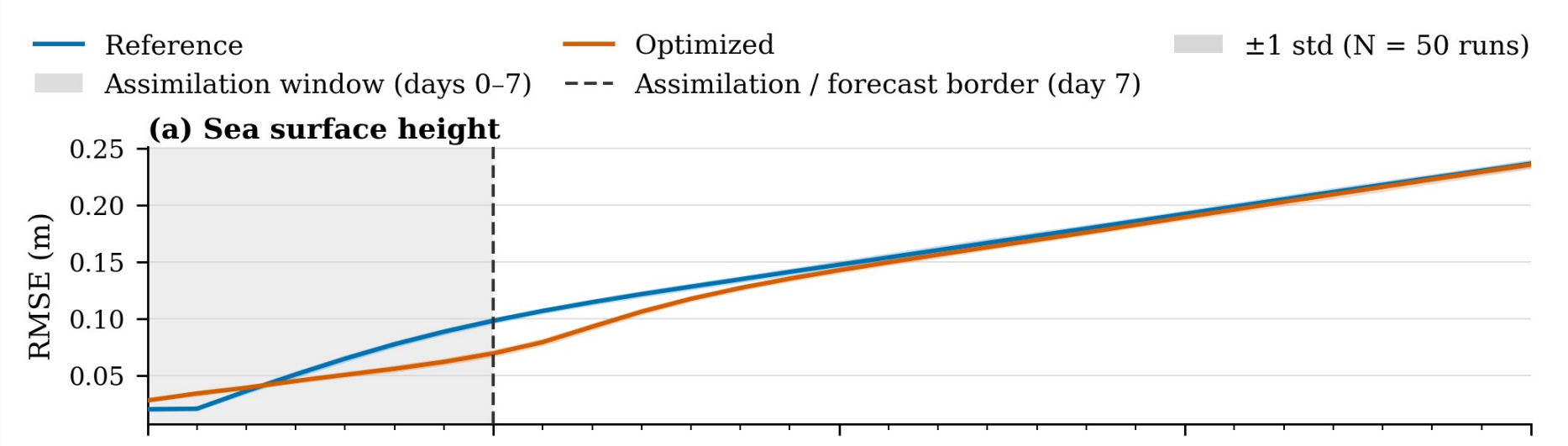


Image 4: SSH forecast RMSE evolution with strong-4DVar BICE

	RMSE [m]	Cost
Reference forecast	0.098	1 min
BICE analysis	**0.067 (31.40%)**	5 hours

Table 1: Forecast RMSE (SSH) at t=7 and Computational Cost

Q2. What Determines Optimization Performance

Summary of the three-sensitivity investigation

$$J_{target}(\psi_0) = background_{term} + \sum_{t=0}^{\tau} ¿¿¿¿$$

Observation information	Assimilation window	Loss formulation
Sparse observations strongly degrade reconstruction	Longer window provides stronger information about the I.C. and its trajectory with higher computation cost	Loss formulation is critical. Multivariate weighting strongly re-distribute estimations skill over the variables

Table 2: strong 4DVar BICE sensitivity summary

Q3. Are The Gradients Physically Meaningful?

Gradient Smoothing Improves Spatial Coherence Without Suppressing the Useful Signal

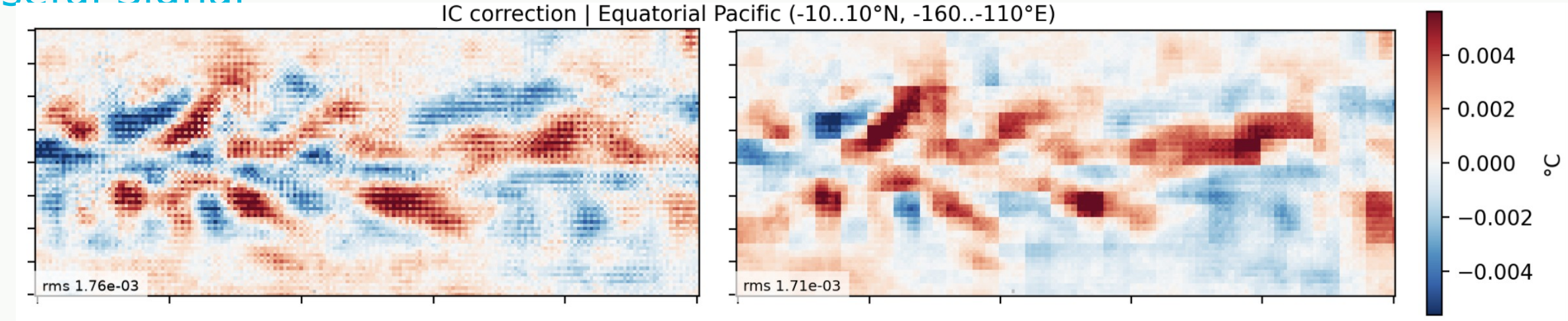


Image 5: Gradient Smoothing effect before(left) and after(right) on IC correction of SST

Methode	RMSE (gain w.r.t ref)	P-score★
BICE analysis	0.067 **(30.85%)**	0.12
Smoothed BICE	0.069 **(29.20%)**	**0.03**

Table 3: Forecast RMSE (SSH) at t=7 and Pixelization Score

★0.04: ground truth field
0.07: threshold for pixelization

CURRENT LIMMITATION FROM BEING PRACTICAL

A. Non-phyciscal optimization
   - Grid-scale artefacts detected on initial condition correction
   - Cost function sensitivity & learned ocean dynamic sensitivity

B. Computational cost

   Iterative forward and backward model evaluation

   5 hours for each optimization on a GPU NVIDA H100

   50 samples ~ 10 days

Structure

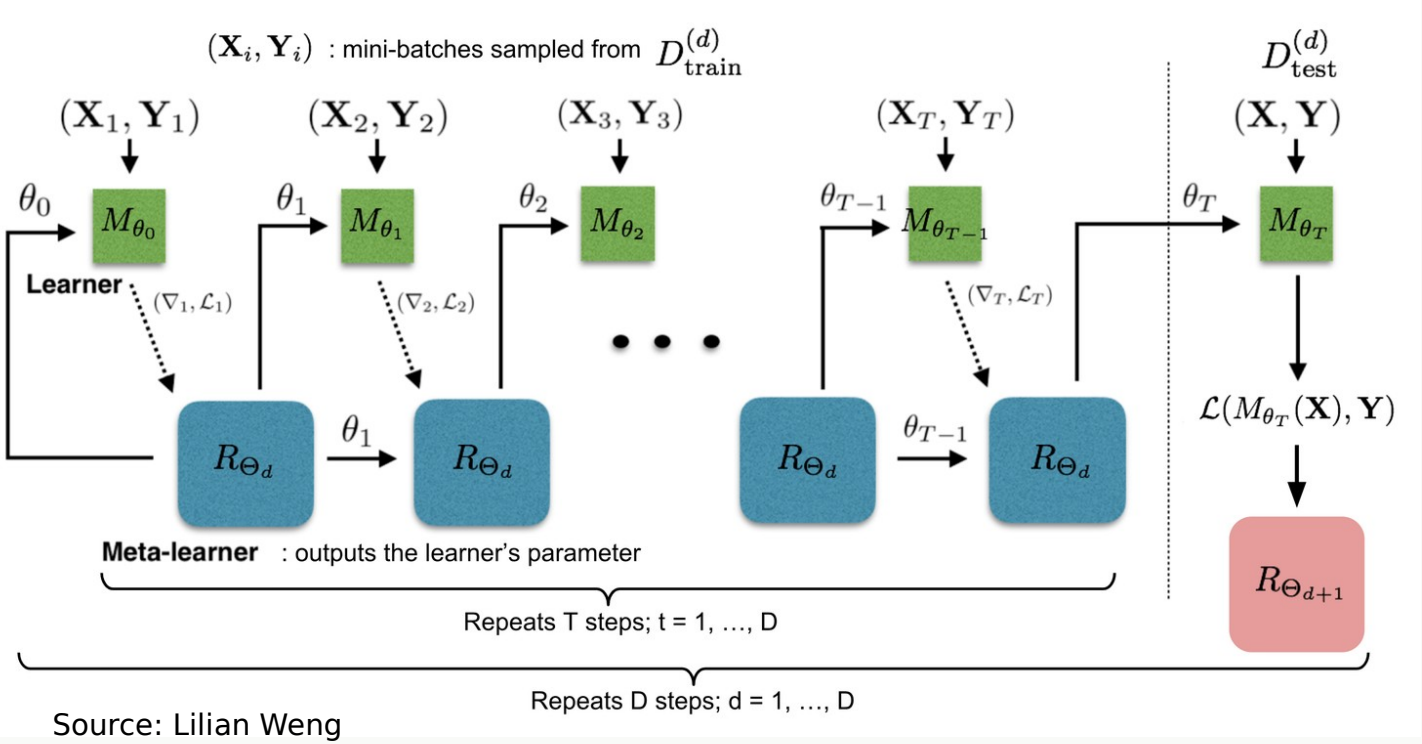
I. Introduction
- Motivation
- Research question

II. Strong-Constraint 4D-Var
- Best Initial Condition Estimation
- Experiment

III. Strong 4D-Var with learned optimizer
- **Meta Learner**

IV. Conclusions and perspectives

META LEARNING



Source: Lilian Weng

Image 6: Meta learning scheme

Meta learning

Optimization algorithm can be a learning problem

Learning a neural solver enchances physics-informed methods in time

ANDRYCHOWICZ et al.

LE BOUDEC et al.

Question

Can part of the variational optimization rule itself be learned too?

LEARNING UPDATE RULES

From Strong 4D-Var Best Initial Condition Estimation (BICE)

Objective

$$\underset{\psi_0}{argmin}\, J_{obs}(\psi_0)$$

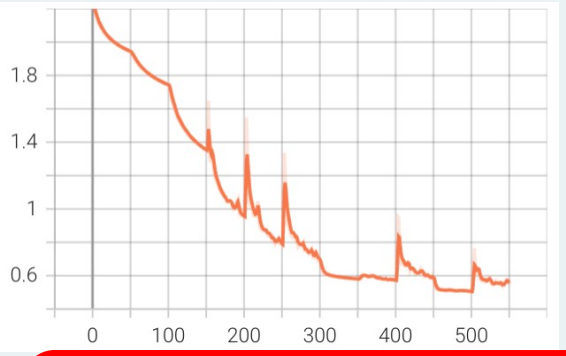
Cost function

$$J_{obs}(\psi_0) + \sum_{t=0}^{\tau} ¿¿¿¿$$

$$\nabla_{\psi_0} J(\psi_0^{(k)})$$

backpropagation

Gradient descent optimizaiton



$$\psi_0^{(k+1)} = \psi_0^{(k)} - \mathcal{M}(\psi, \nabla, k; \theta)$$

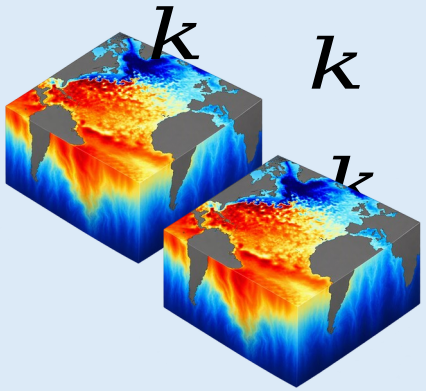
iteration until converge

VONICH et al.

Update Initial Condition with Learned Update Rules

IMPLEMTED META LEARNER

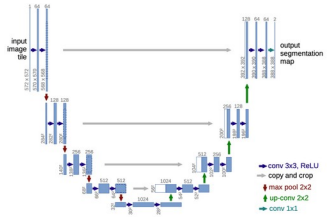
Input



$\psi^{(k)}, \nabla J^{(k)}, k$

Meta Model (31M parameters)

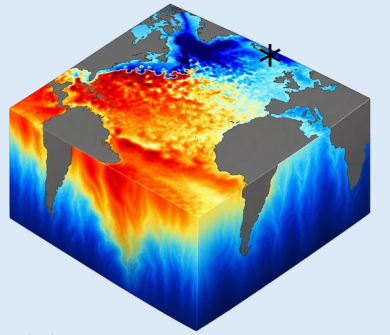
U-Net



Capturing different-scale features with embedded

RONNEBERGER et al.

Output



$\nabla J^{(k)*}$ = optimal update

Where is current optimization problem ?
How far update moved from original I.C. ?
How steep the current optimization problem ?

META LEARNER OFFLINE TRAINING

Meta learner is trained to make further I.C. update in nested inner loop

1 iteration of outer optimization

20 inner iteration (glonet forward+backward)

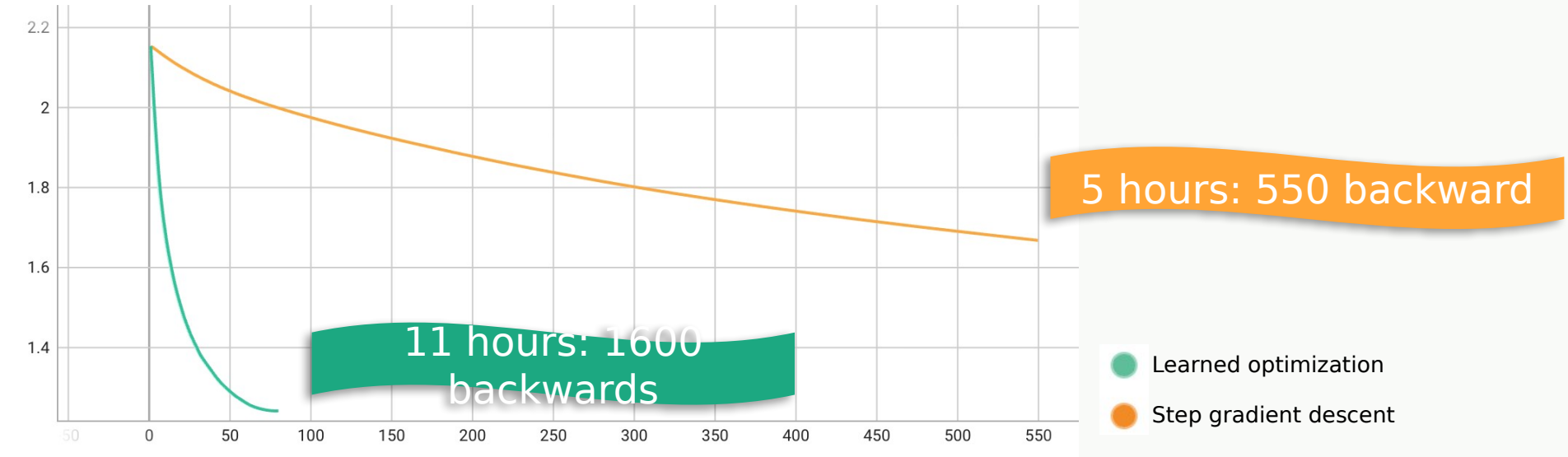


Image 7: Meta learner training curve

Image 7: Meta Learner Training Curve

META PERFORMANCE

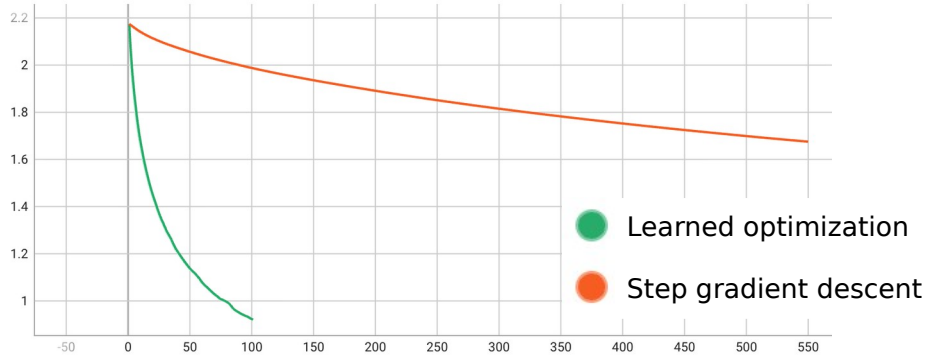


Image 8: Optimization comparaison between learned optimizer and step gradient descent method.

Methode	RMSE gain	P-score	C-cost/ run
reference BICE ★	31.40%	0.12	6 hours
Meta inference	25.44%	**0.04**	**1 hour**

Table 4: Meta Learner Forecast RMSE (SSH) at t=7 and Computational Cost

★Trained on only one optimization sample, then 50 inference after 50 sequential days optimization

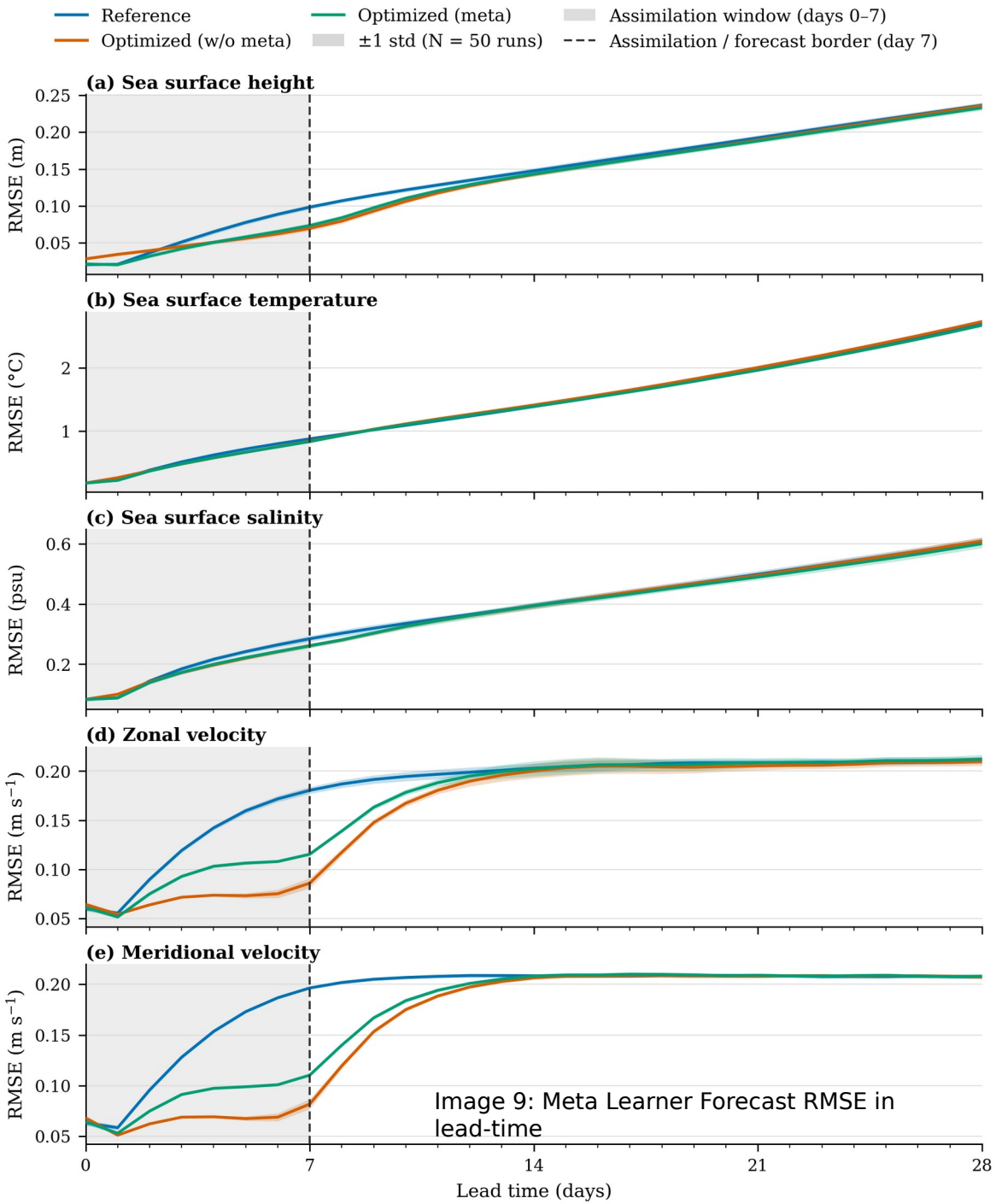


Image 9: Meta Learner Forecast RMSE in lead-time

META LEARNER EFFECT ON GRADIENT FIDELITY

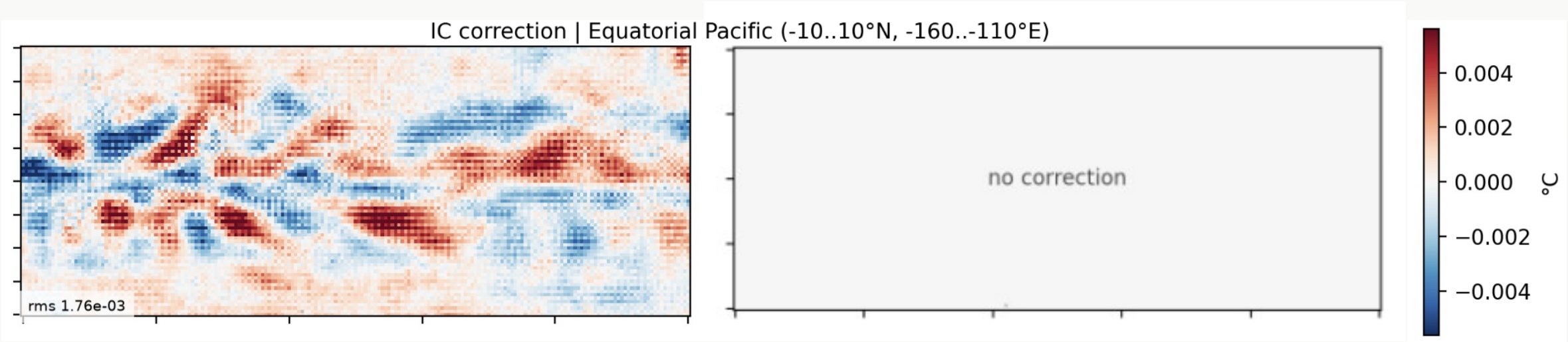


Image 10: Learned optimization I.C. solution on SST(right) and hand-designed gradient descent solution (left) on equatorial zone

LEARNED INITIAL CONDITION CORRECTION

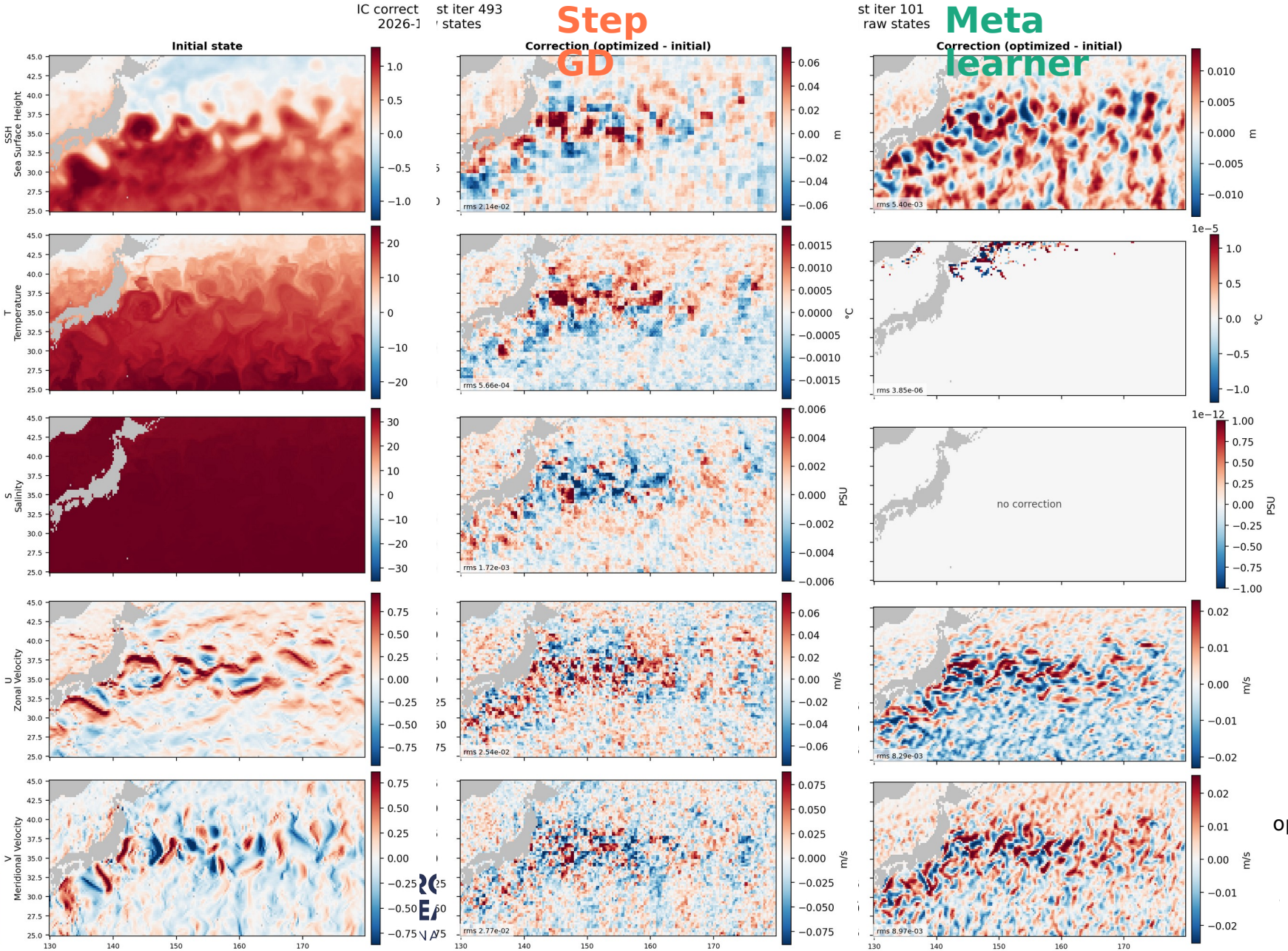


Image 11: Learned optimization I.C. solution on Kuroshio region

Structure

I. Introduction
- Motivation
- Research question

II. Strong-Constraint 4D-Var
- Best Initial Condition Estimation
- Experiment

III. Strong-Constraint 4D-Var with learned optimizer
- Meta Learner

IV. Conclusions and perspectives

TAKE-HOME MESSAGES

I. Differentiable ML-based emulastor can be used in variational IC estimation

II. Forecast RMSE based improvement is not enough; physical coherent gradient is needed

III. Learned update shows promising framework to reduce computational cost, although its behaviour need to be investigated further.

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Lab-STICC

Co-funded by the European Union

SEED
Societal, Energy, Environmental, industrial & Digital transitions

MERCATOR OCEAN INTERNATIONAL

THANK YOU

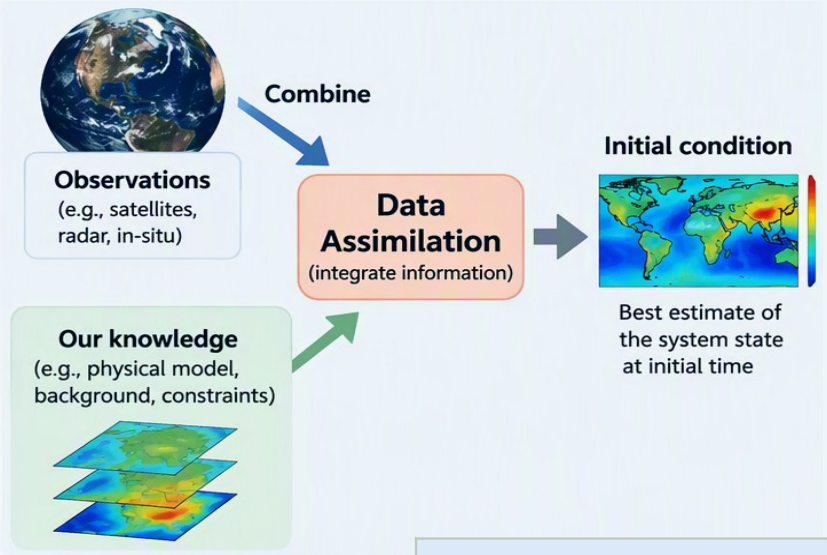
Jungwon LEE

junwon.lee@imt-atlantique.fr

+33 07 600 700 67

Training Initial Condition

Variational DA



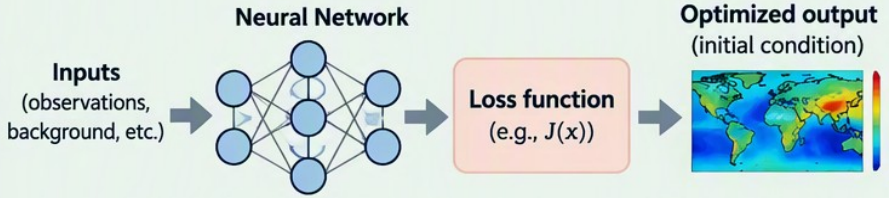
EVENSEN

NN training

Find the state x (initial condition) that minimizes

$$J(x) = \underbrace{\|x - x_b\|^2_{B^{-1}}}_{\text{Fit to our knowledge (background, model)}} + \underbrace{\|H(x) - y\|^2_{R^{-1}}}_{\text{Fit to observations}}$$

Viewed as training a NN:



FABLET et al.

Data Assimilation	Machine Learning
Cost function	Loss function
Initial condition	Parameters
Adjoint	Backpropagation
Variational minimization	Gradient descent from loss function

A1. BICE performance table by variable

The Reconstructed Initial Condition Improves the Subsequent Forecast

Improvement at fixed time t=7

RMSE	Reference		BICE		Improvement
SSH [m]	0.097	0.002	0.067	0.003	**31.40%**
SST [°C]	0.881	0.012	0.848	0.014	4.27%
SSS [psu]	0.288	0.009	0.256	0.006	9.69%
SSU [m/s]	0.181	0.003	0.078	0.001	**56.32%**
SSV [m/s]	0.196	0.001	0.075	0.001	**61.56%**

Table 5: Forecast RMSE comparaison for all variable with standard deviation

A2a. Observation

Observation quality	RMSE	SSH	SST	SSS	SSU	SSV	P-score	C-cost/ run
Ideal BICE (ref)	absolute	0.067 m	0.848 ℃	0.256 psu	0.078m/s	0.075m/s		
	gain	30.38 %	4.38%	10.18%	56.30%	61.41%	0.04	5 hours
Fully observed only SSH	gain	37.59 %	-3.03%	1.76%	12.39%	14.30%	0.04	4 hours
10 % of ocean for all variables	gain	-2.84%	-14.14 %	-12.16%	0.55%	7.5%	0.03	4 hours
Noised I.C.	gain	28.69 %	4.03%	8.08%	51.30%	57.34%	0.03	4 hours

Table 6: Forecast RMSE gain with different observation configuration

A3b. Assimilation Window

Assimilation length	SSH	SST	SSS	SSU	SSV	P-score	C-cost/ run
7-day (ref)	30.38%	4.38%	10.18%	56.30%	61.41%	0.04	5 hours
3-day	10.34%	2.94%	4.33%	9.11%	10.64%	0.03	2 hours
1-day	2.28%	0.86%	0.92%	2.04%	2.33%	0.02	1 hour

Table 7: Forecast RMSE gain with different assimilation-window length

A3c. Loss Function

Spatial gradient mismatch

balancing contribution of the different variables on

	SSH	SST	SSS	SSU	SSV	P-score	C-cost/run
(ref)	30.38 %	4.38%	10.18 %	56.30 %	61.41 %	0.04	5 hours
+	29.09 %	4.43%	8.09%	51.90 %	58.04 %	0.04	5 hours
	67.77 %	54.87 %	46.14 %	52.89 %	55.63 %	0.06	5 hours

Table 8: Forecast RMSE gain with different loss function formulation

A3. Are The Gradients Physically Meaningful?

Optimization Converges, but the Reconstructed I.C. Contains Grid-Scale Artefacts

Although the optimization converged, we observed local artifacts

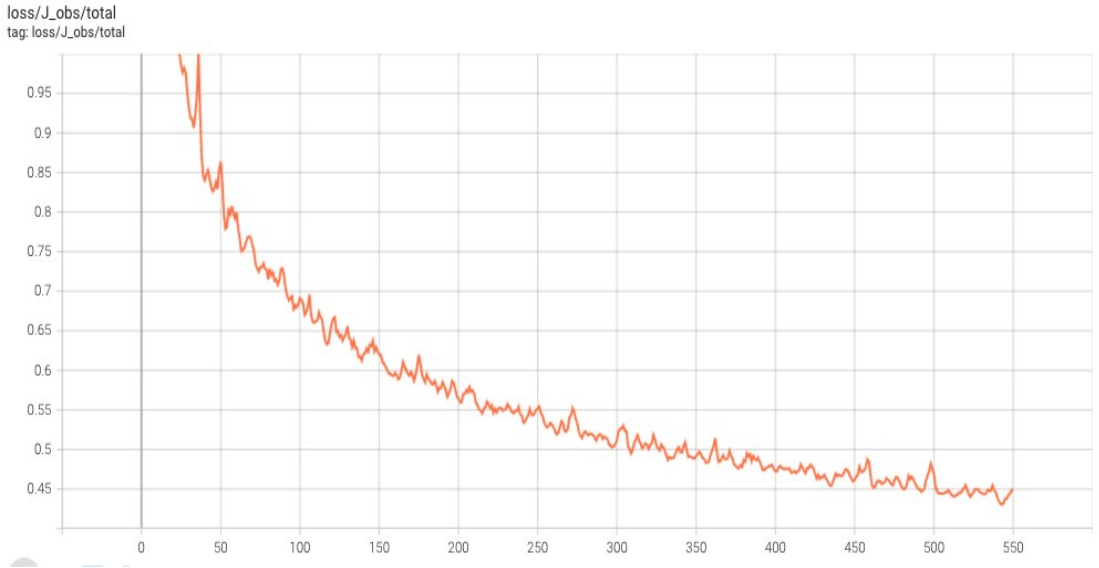


Image 12: optimization curve with pixelized case

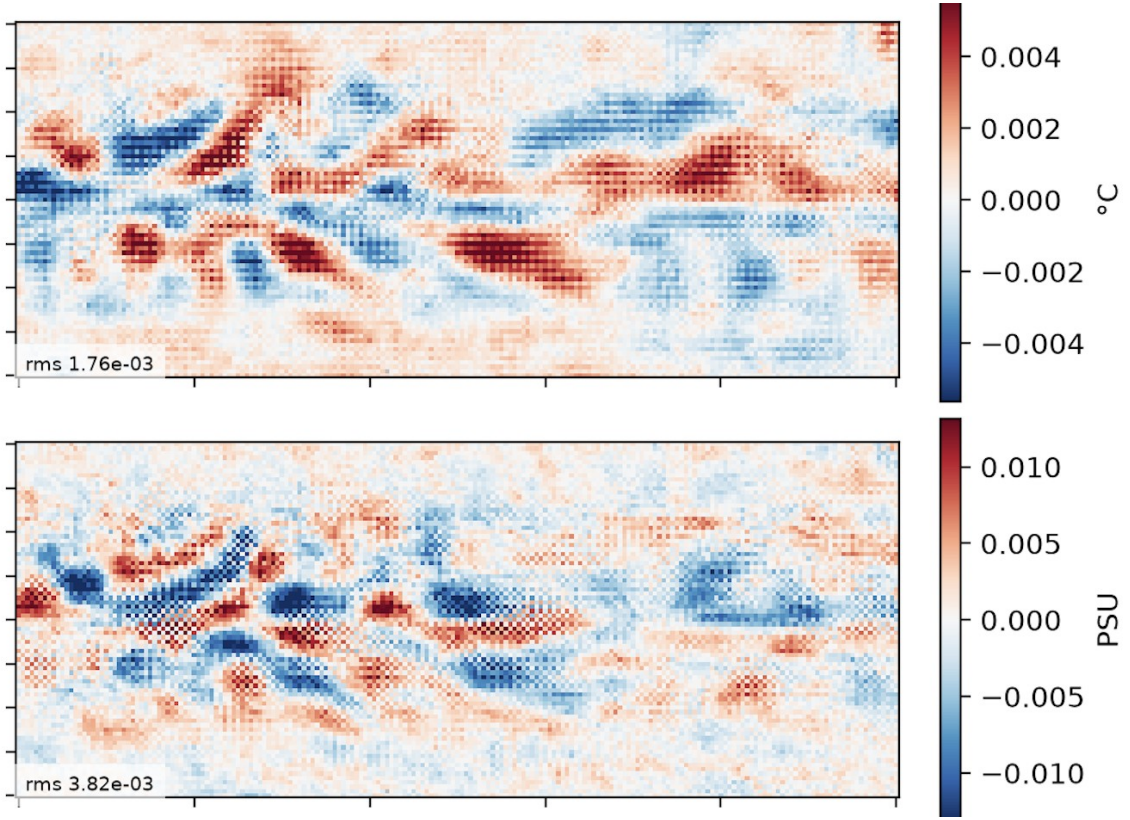
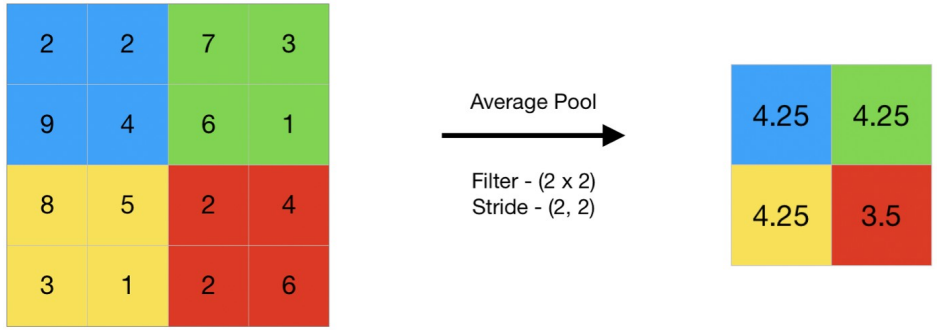


Image 13: Pixelized IC correction of SST(upper) SSS(bottom) on equatorial zone

A3. Are The Gradients Physically Meaningful?

Gradient Smoothing Filter with Multi-Resolution As a Solution Against Pixelization

F = average pooling - upsampling



Multi-resolution concept while optimizatic

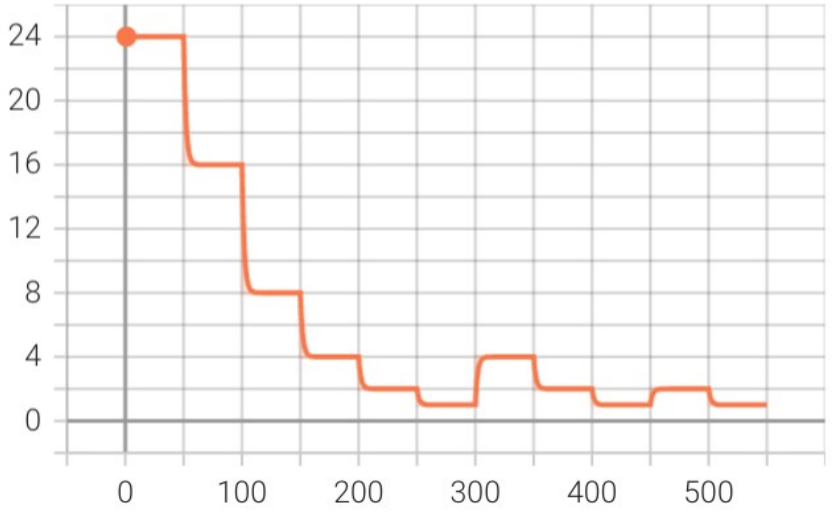


Image 14: Smoothing filter and multi-resolution

A2. Pixelization score

P_socre = checkerboard coefficient

For ocean window, apply checkerboard kernel k

Compute mean of over total energy of the window

A4: Surrogate fwd+bwd cost w.r.t. Window Length

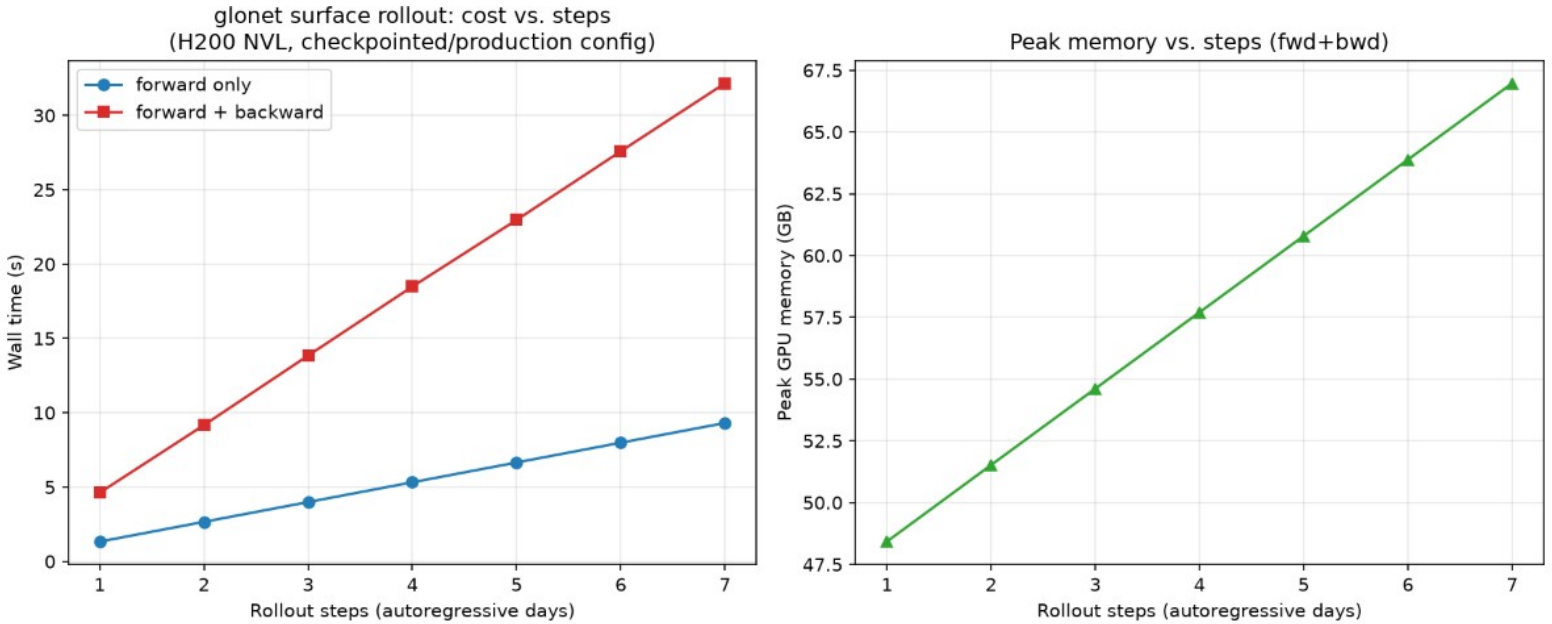


Image 15: GLONET computation cost by forecast rollout step

$37s * 550\ steps \approx 5.5\ hours$
forward ≈9.3s,
backward ≈22.8s,
fwd+bwd ≈32.1s,
peak mem ≈67GB

A5: Noise Added on I.C.

White noise

Std: 10% of standard deviation of the signal anomaly

	SSH [m]	T []	S [psu]	U [m/s]	V [m/s]
	0.08	1.73	0.53	0.148	0.136
Scale factor 0.1	0.008	0.17	0.053	0.0148	0.0136

Table 9: Added noise standard deviation.

References

VONICH, P. Trent et HAKIM, Gregory J. Atmospheric Predictability beyond 30 Days with Machine Learning. *Artificial Intelligence for the Earth Systems*, 2026, vol. 5, no 3, p. 260009.

MADEC, Gurvan, *et al.* NEMO ocean engine. 2015.

EL AOUNI, Anass, GAUDEL, Quentin, REGNIER, Charly, *et al.* GLONET: Mercator's end-to-end neural Global Ocean forecasting system. *Journal of Geophysical Research: Machine Learning and Computation*, 2025, vol. 2, no 3, p. e2025JH000686.

WANG, Xiang, WANG, Renzhi, HU, Ningzi, *et al.* Xihe: A data-driven model for global ocean eddy-resolving forecasting. *arXiv preprint arXiv:2402.02995*, 2024.

O'SHEA, Keiron et NASH, Ryan. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

LI, Zongyi, KOVACHKI, Nikola, AZIZZADENESHELI, Kamyar, *et al.* Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

GRANDE, Davide, BUIZZA, Roberto, et STORTO, Andrea. Machine learning in ocean data assimilation: Advances, gaps and the road to operations. *Ocean Modelling*, 2026, p. 102678.

LE BOUDEC, Lise, DE BÉZENAC, Emmanuel, SERRANO, Louis, *et al.* Learning a neural solver for parametric PDEs to enhance physics-informed methods. In : *International Conference on Learning Representations*. 2025. p. 46994-47045.

RONNEBERGER, Olaf, FISCHER, Philipp, et BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. In : *International Conference on Medical image computing and computer-assisted intervention*. Cham : Springer international publishing, 2015. p. 234-241.

ANDRYCHOWICZ, Marcin, DENIL, Misha, GOMEZ, Sergio, *et al.* Learning to learn by gradient descent by gradient descent. *Advances in neural information processing systems*, 2016, vol. 29.

FABLET, Ronan, CHAPRON, Bertrand, DRUMETZ, Lucas, *et al.* Learning variational data assimilation models and solvers. *Journal of Advances in Modeling Earth Systems*, 2021, vol. 13, no 10, p. e2021MS002572.

EVENSEN, Geir. *Data assimilation: the ensemble Kalman filter*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009.