

OceaniX-Melody Sandbox: Deep Learning and Data Assimilation (Melody WP3)

OceaniX-Melody Sandbox, April 2021

OceaniX-Melody sandbox, April 9th, 10am-12am

Program

- 4DVarNN@OceaniX: current status and new applications (**Ronan F.**), 10am-10.30am
- Data Assimilation for SSH mapping (**Emmanuel C.**), 10.30am-10.50am
- SSH mapping Data Challenge (**Quentin F.**), 10.50am-11.00am
- Learning physical parameterizations using DA and EM (**Pierre T.**), 11.00am-11.20am
- Discussion on end-to-end learning, DA and uncertainty propagation (**Said O.**), 11.20am-11.40am
- Discussion on future work, 11.40am-12.00am

4DVarNN@OceaniX: current status and new applications

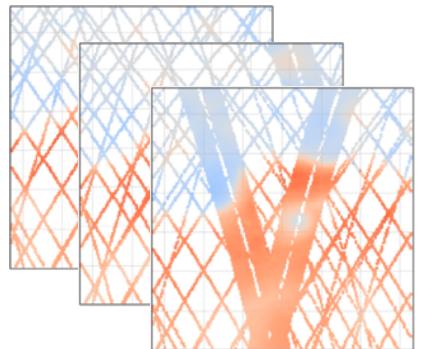
OceaniX team

Slides available at https://cia-oceanix.github.io/oceanix_sandbox

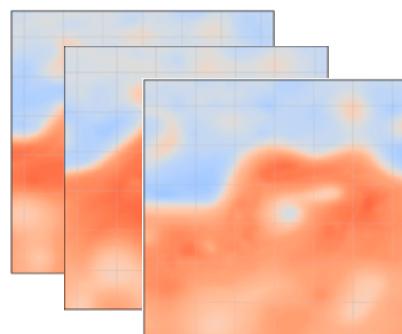
Code: https://github.com/CIA-Oceanix/DinAE_4DVarNN_torch

OceaniX-Melody Sandbox, April 2021

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



Partial observations y



True states x

State-space formulation:

$$\begin{cases} \frac{\partial x(t)}{\partial t} = \mathcal{M}(x(t)) \\ y(t) = x(t) + \epsilon(t), \forall t \in \{t_0, t_0 + \Delta t, \dots, t_0 + N\Delta t\} \end{cases}$$

Associated variational formulation:

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

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

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

4DVarNN: Learning 4DVar models and solvers

Trainable Variational DA formulation

$$\hat{x} = \arg \min_x \|x - y\|_{\Omega}^2 + \lambda \|x - \Phi(x)\|^2$$

Predefined

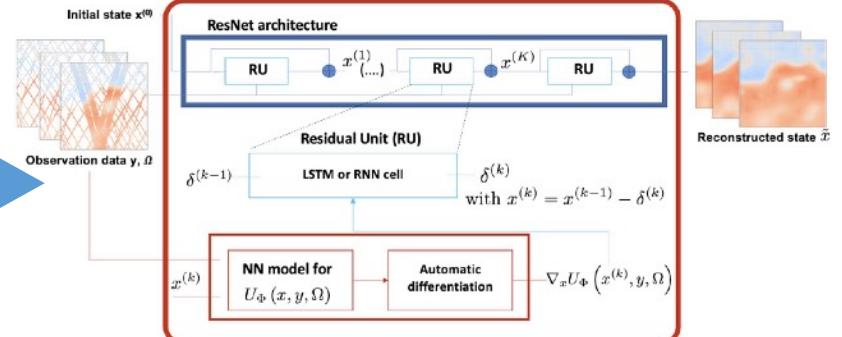
Observation model

Variational model implemented in a differentiable framework (eg, pytorch)

Trainable prior

ODE-based
U-Net
AE

End-to-end architecture



Trainable gradient-based solver

Learning criterion

Variational cost (non-supervised)
Reconstruction error (supervised)

Preprint: <https://arxiv.org/abs/2006.03653>

Code: https://github.com/CIA-Oceanix/DinAE_4DVarNN_torch

4DVarNN: Learning 4DVar models and solvers

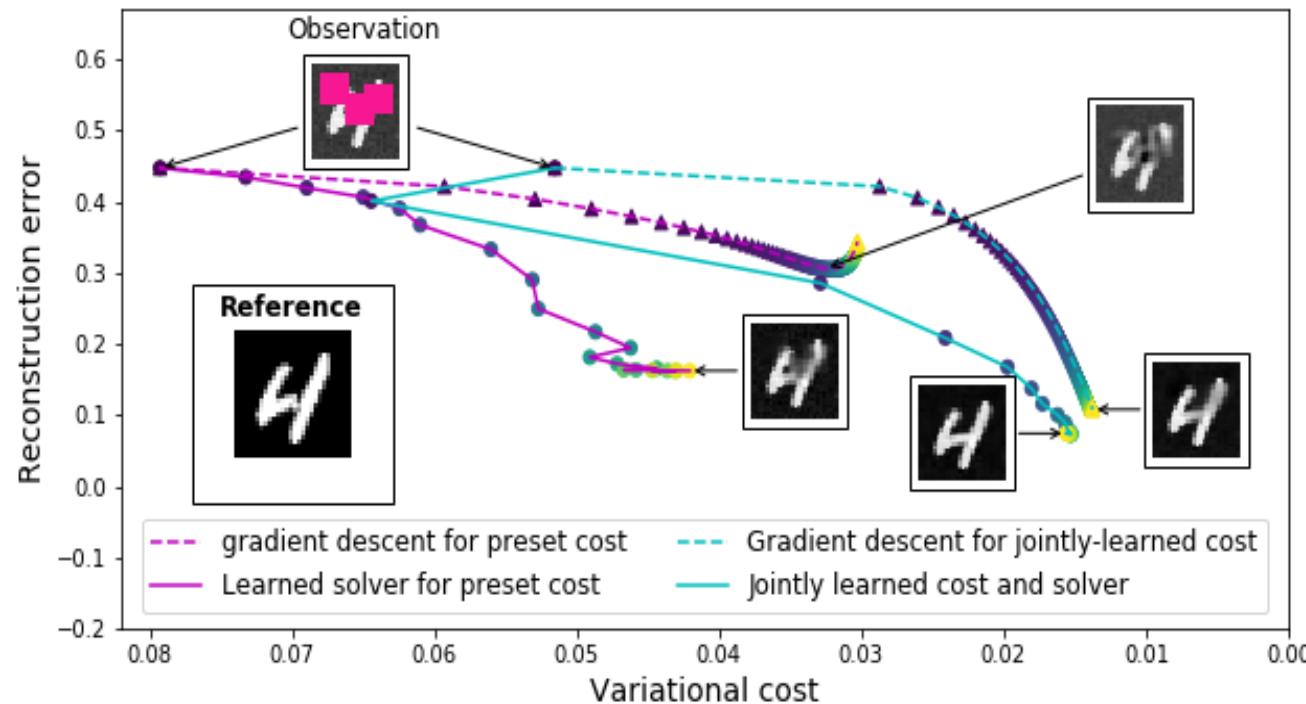
Trainable

$$\hat{x} = \arg \min_x$$

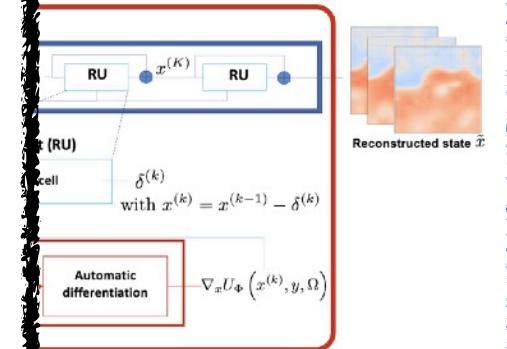
Pred

Observati

Variatio
differentiable framework (eg, pytorch)



End-to-end architecture



client-based solver

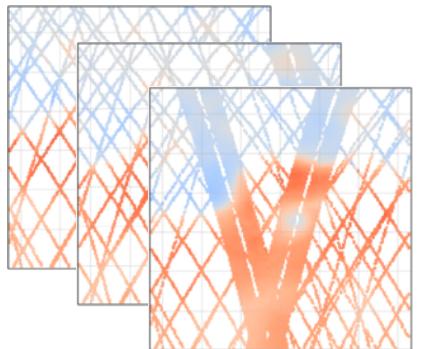
criterion

(non-supervised)
error (supervised)

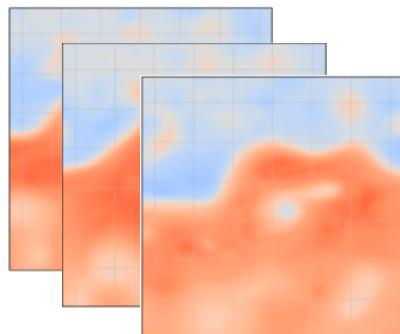
Preprint: <https://arxiv.org/abs/2006.03653>

Code: https://github.com/CIA-Oceanix/DinAE_4DVarNN_torch

4DVarNN: Application to satellite-derived SSH mapping



Partial observations y



True states x

Considered formulation:

$$x = (\bar{x}, \delta x)$$

Coarse-scale Fine-scale anomaly

$$y = (\bar{y}, y_{alt})$$

OI field Altimeter data

Variational model:

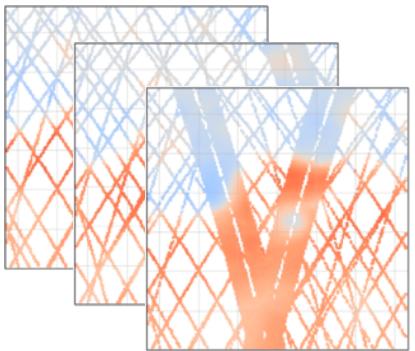
$$\sum_t \lambda_{1,t} \|\bar{y}(t) - \bar{x}(t)\|^{\alpha_1} + \lambda_{2,t} \|y_{alt}(t) - \bar{y}(t) - \delta x(t)\|_{\Omega_t}^{\alpha_2} \\ + \lambda_3 \|x - \Phi(x)\|^{\alpha_3}$$

Trainable prior: U-Net prior (not ODE-based, cf. L63/L96 experiments)

Solver: iterative gradient-based LSTM solver

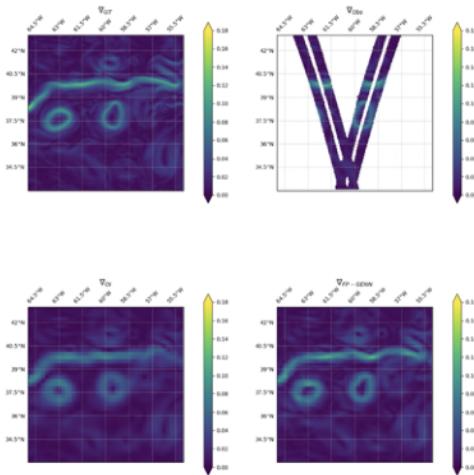
Training criterion: MSE for the SSH and its gradient norm (supervised learning)

4DVarNN: Application to satellite-derived SSH mapping



OSSE:

- NATL60 data / Gulf Stream region ($10^\circ \times 10^\circ$)
- Noise-free SWOT+4-nadir observation dataset
- Test period on 20 days (remaining data used for training)



Partial observations y

Results

Models	4DVarNN-L2-LSTM	4DVarNN-L1-LSTM	4DVarNN-L2-FP
relative MSE gain vs. OI for SSH	53 %	52 %	37 %
Relative MSE gain vs. OI for gSSH	45 %	45 %	30 %

Resolved horizontal scale ~70km
(To be checked for these results)

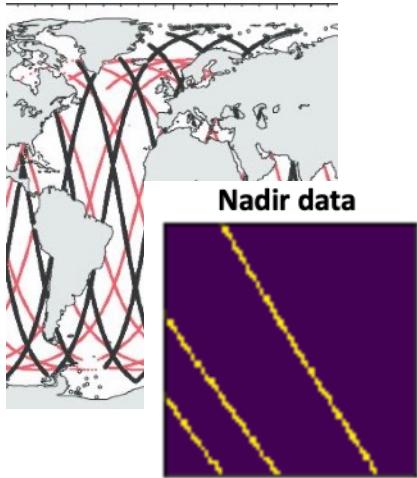
Lower to no improvement if the prior model to a ConvAE

Trained models also apply to nadir-only observation data (with a lower improvement)

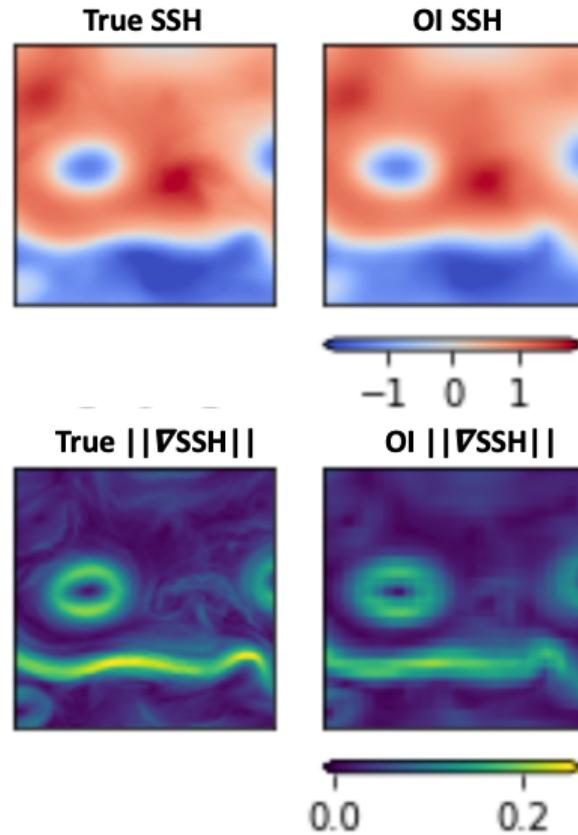
True states x

4DVarNN #2: New applications

Learning where to sample ?



Available sampling pattern (nadir satellite altimeter)

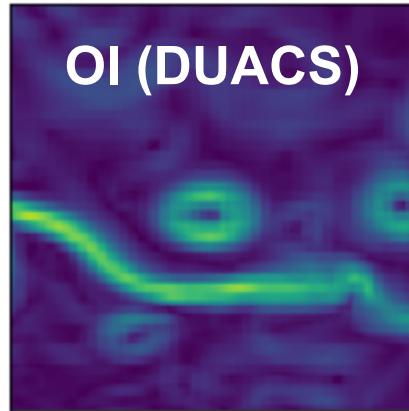
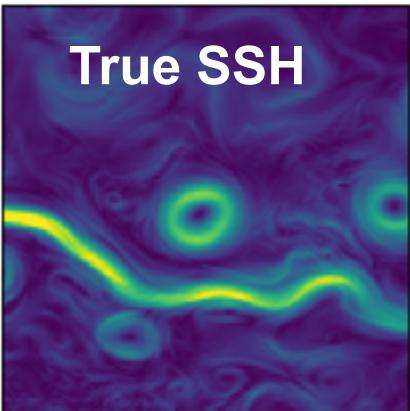


Where to collect new altimetry/current data to improve the reconstruction of the SSH field and of its gradient ?



4DVarNN #2: New applications

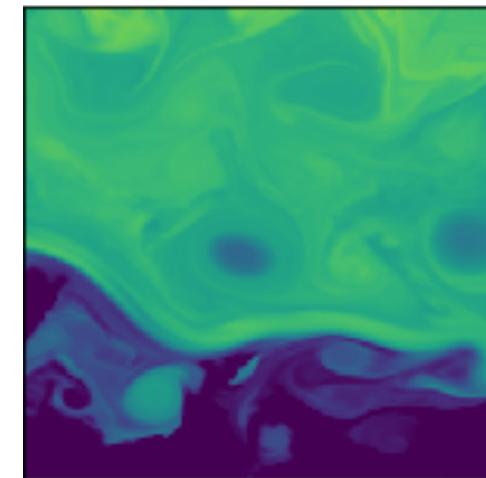
Learning what to measure or extract?



Nadir and SWOT
altimeter data
mask



What to extract from sea surface fields
to improve the reconstruction of the
SSH field and of its gradient ?



SST data

4DVarNN #2: New applications

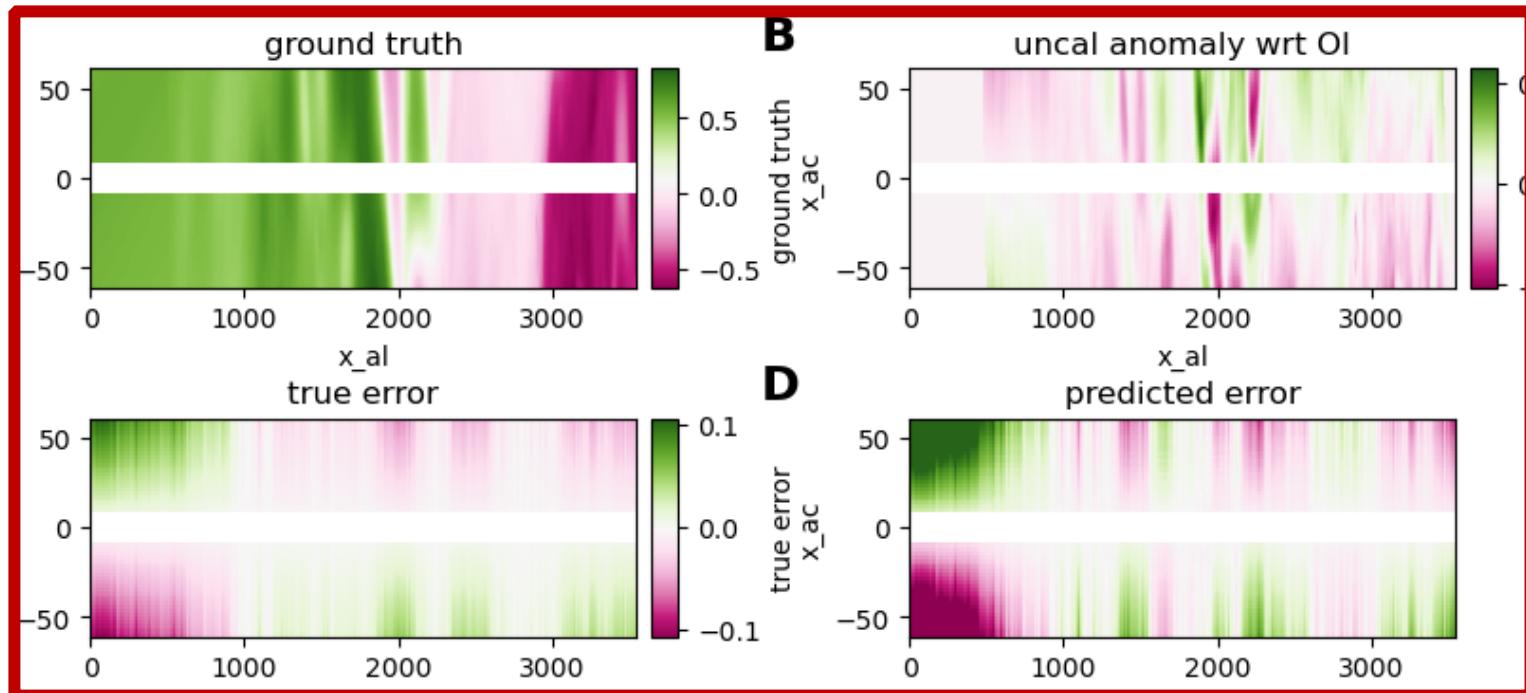
Unsupervised SWOT Calibration

OSSE with NATL60 data +
Roll Error signal from SWOT
simulator

4-nadir-altimeter + SWOT +
DUACS baseline

Gulf Stream area ($10^\circ \times 10^\circ$)

In the 4DVarNN framework we can learn
the calibration operator as an observation
model jointly with the prior operator

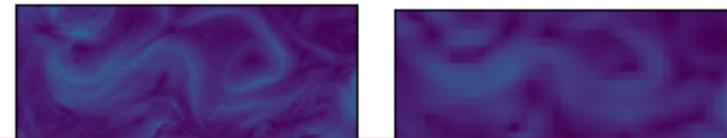


4DVarNN #2: New applications

Learning where to sample ?



Learning what to measure ?



4DVarNet models with trainable observation models

$$\hat{x} = \arg \min_x \|x - y\|^2 + \lambda \|x - \phi(x)\|^2$$

Sparse sampling operator

$$\|H(z) * (x - y)\|^2$$

$$\text{s.t. } \forall z, \|H(z)\|_1 < \epsilon$$

Multimodal observation

$$\begin{aligned} & \|x - y\|^2 \\ & + \alpha \|G * x - F * z\|^2 \end{aligned}$$

SWOT calibration

$$\|H(y) - x\|_{\Omega}^2$$

4DVarNN #2: New applications

4DVarNet models with trainable observation models

Spase sampling operator

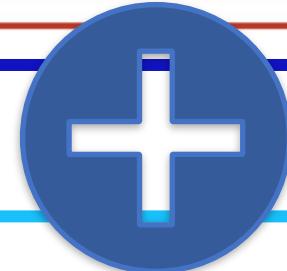
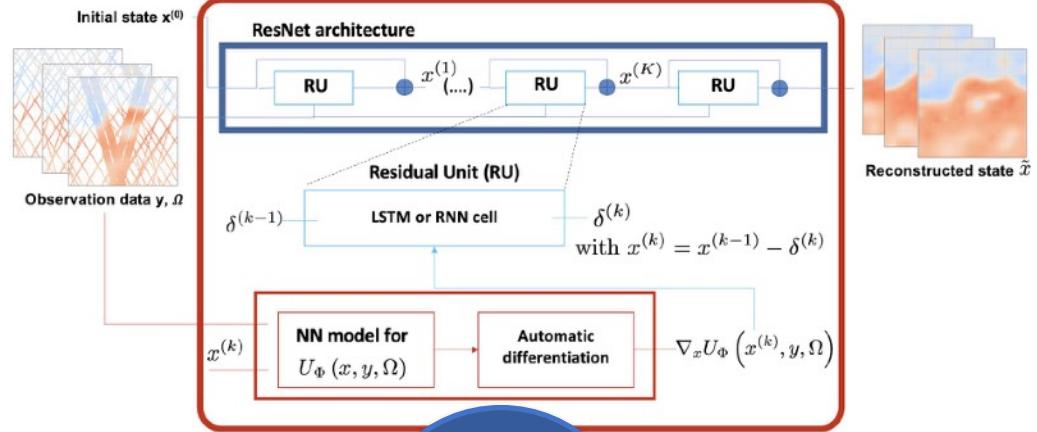
$$\|H(z) * (x - y)\|^2$$

$$\text{s.t. } \forall z, \|H(z)\|_1 < \epsilon$$

Multimodal observation

$$\begin{aligned} &\|x - y\|^2 \\ &+ \\ &\alpha \|G * x - F * z\|^2 \end{aligned}$$

End-to-end 4DVarNet



Supervised training loss

(under sparsity constraint for the optimal sampling case)

4DVarNN #2: Preliminary results

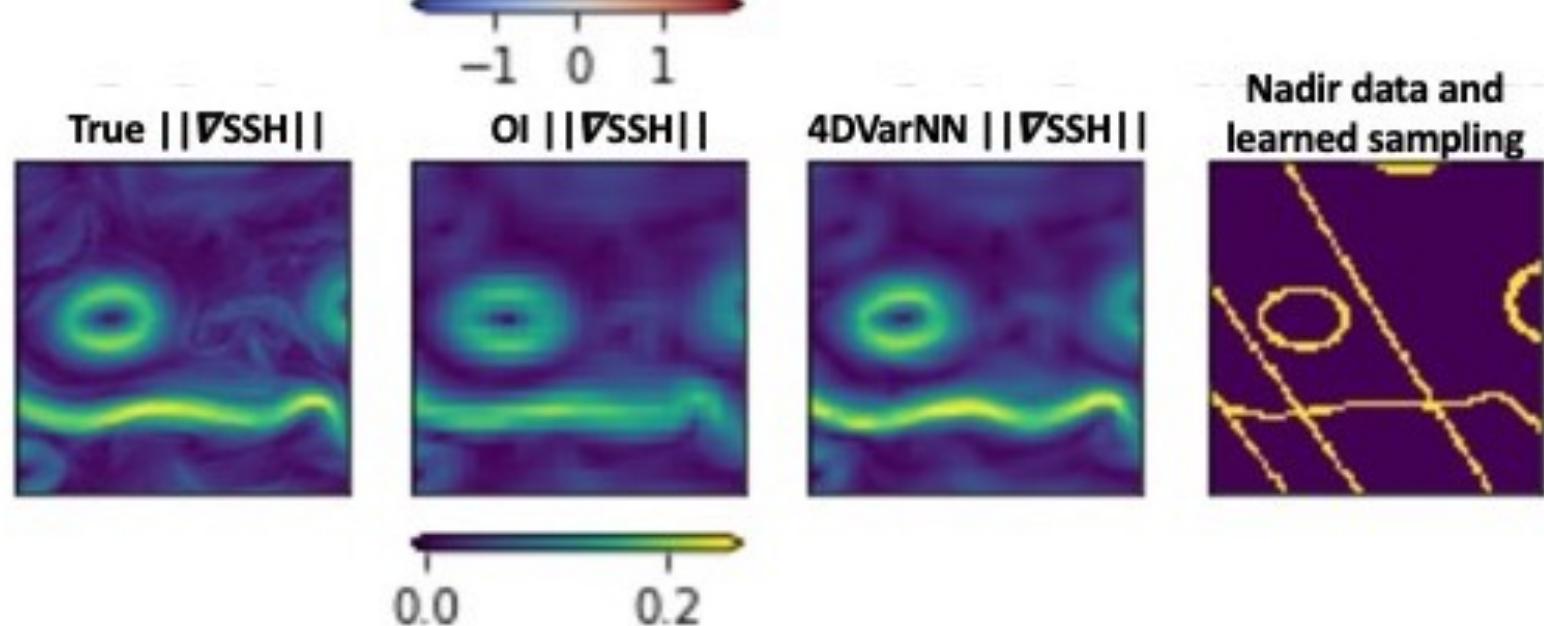
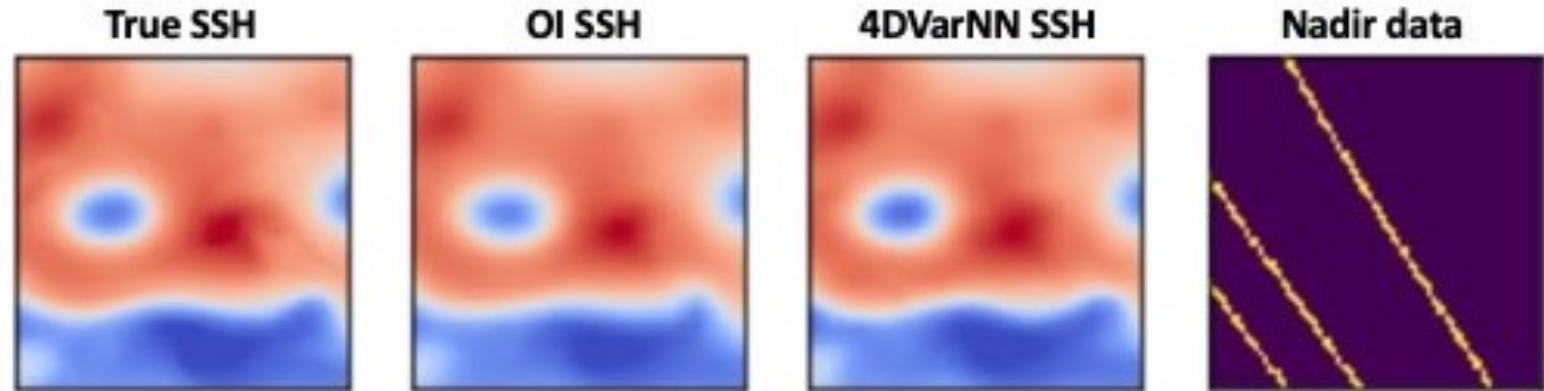
Optimal sampling

OSSE with NATL60 data

4-nadir-altimeter + DUACS baseline

Gulf Stream area ($10^\circ \times 10^\circ$)

Mean relative gain of 60% in the reconstruction of the SSH using the learned sampling (~6% of the pixels vs. 1.3% for nadir altimeters)



4DVarNN #2: Preliminary results

Multimodal DA

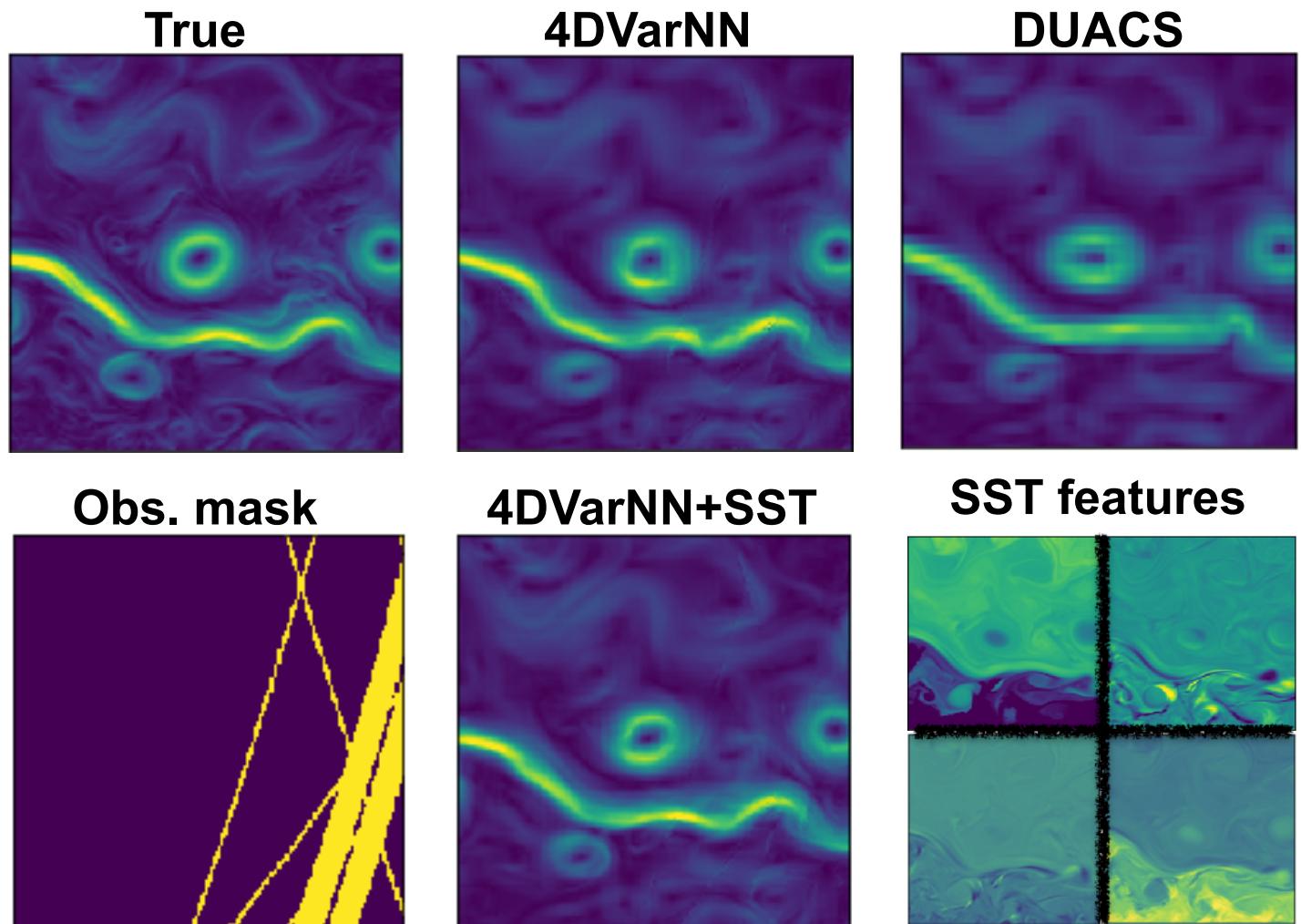
OSSE with NATL60 data

4-nadir-altimeter + SWOT +
DUACS baseline

Gulf Stream area ($10^\circ \times 10^\circ$)

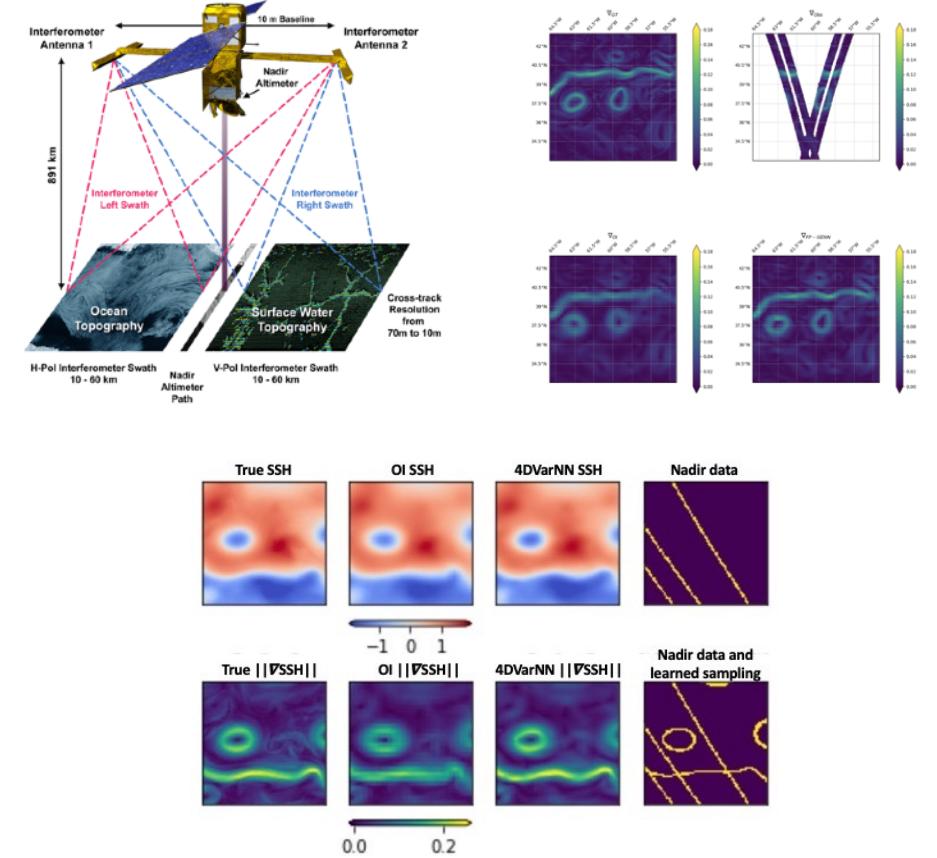
63% vs. 53% gain in SSH
MSE w.r.t. DUACS with/
without SST (Winter period)

50% vs. 13% gain in SSH
MSE w.r.t. DUACS using
nadir altimeter data only



4DVarNN #2: Key messages

- Learning jointly variational priors, observation models and solvers
- Optimal sampling as a learning issue under sparsity constraint
- Multimodal DA as a learning issue with a trainable feature extraction operator
- Generic framework beyond the considered testbed on SSH mapping



Preprint: <https://arxiv.org/abs/2006.03653>
Code: https://github.com/CIA-Oceanix/DinAE_4DVarNN_torch

Synthesis: current status for 4DVarNN framework

Trainable Variational DA formulation

$$\|H * y - G * x\|_{\Omega}^2 + \lambda \|x - \Phi(x)\|^2$$

(Trainable) obs. model

Identity model

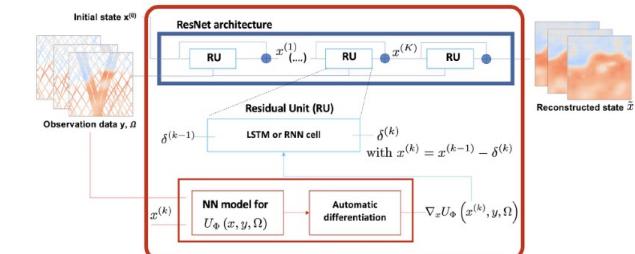
Sampling operator

Linear operator

Trainable prior

ODE-based
U-Net

Trainable solver



Fixed-point
gradient-based LSTM

Learning scheme

Variational cost (non-supervised)
Reconstruction error (supervised)

Synthesis: ongoing/future developments

New applications / case-studies

Uncertainty/stochasticity
in 4DVarNN

Dealing with extremes
(PhD N. Lafon)

Application to (unobserved)
states (eg, vertical velocities)

Conditional
forecasting

Trainable Variational DA formulation

$$\|H(z, y) - G * x\|_{\Omega}^{\alpha} + \lambda \|x - \Phi(x)\|^{\beta}$$

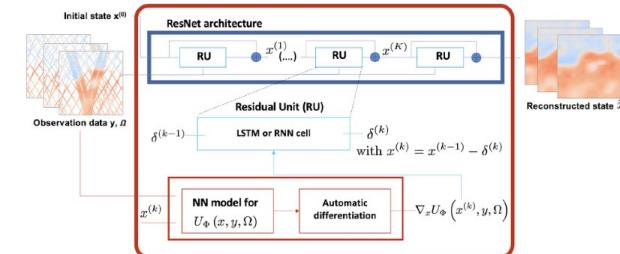
(Trainable) obs. model

Non-linear model
without identity term ?

Trainable prior

Stochastic priors ?
“Functional” priors ?

Trainable solver



Multigrid/multiscale solver ?
Stochastic/Ensemble solver ?

Learning scheme

Beyond MSE losses ?
Adversarial learning ?

Synthesis: SSH mapping case-study

Current status

Benchmarking of 4DVarNN frameworks vs. AnDA/DinEOF on a Gulf Stream region

Promising results of 4DVarNN models vs. OI

New version of benchmarking experiments (To be validated/Melody WP4) (cf. Quentin's slides)

Data challenge framework available on git (cf. Quentin's slides)

Planned/targeted developments

Proposal for a SSH mapping data challenge for SWOT ST DA working group

Scaling up to the ocean basin scale

From OSSEs to real observations: which (transfer) learning strategies ?

Embedding uncertainties (cf. Said's talk)

Some references

Generic 4DVarNN framework

Arxiv preprint on 4DVarNN with applications on L63/L96: <https://arxiv.org/abs/2007.12941>

Short version in Proc. ICASSP'2021, <https://hal-imt-atlantique.archives-ouvertes.fr/hal-03139133>

Applications/case-studies

Benchmarking experiments for SSH mapping (fixed-point solver), Rem. Sens., <https://www.mdpi.com/2072-4292/12/22/3806>

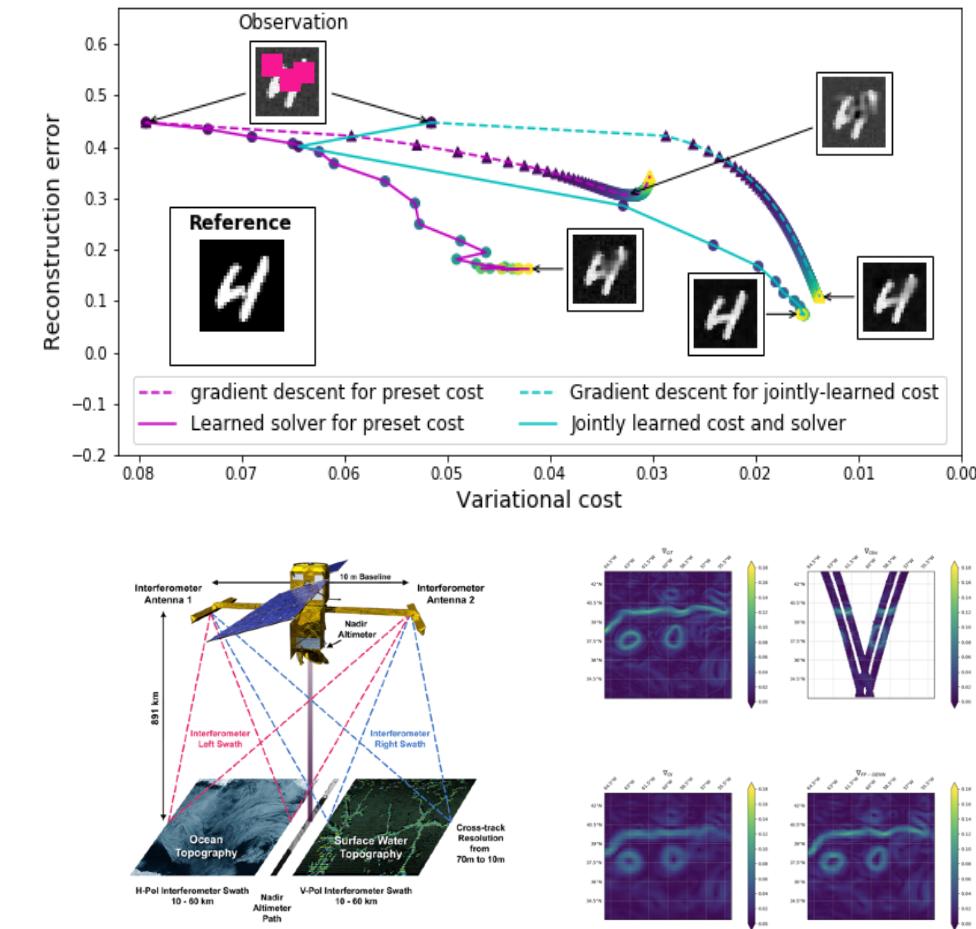
ISPRS'2021 paper (first illustration of forecasting/optimal sampling for SSH mapping): https://www.researchgate.net/publication/350592046_END-TO-END_PHYSICS-INFORMED REPRESENTATION LEARNING FOR SATELLITE OCEAN REMOTE SENSING DATA APPLICATIONS TO SATELLITE ALTIMETRY AND SEA SURFACE CURRENTS

Application of satellite-derived sea surface turbidity, https://www.researchgate.net/publication/350739763_DATA-DRIVEN SPATIO-TEMPORAL INTERPOLATION OF SEA SURFACE SEDIMENT CONCENTRATION FROM SATELLITE-DERIVED DATA AN OSSE CASE-STUDY IN THE BAY OF BISCAY

4DVarNN #1: end-to-end-learning of 4DVar models

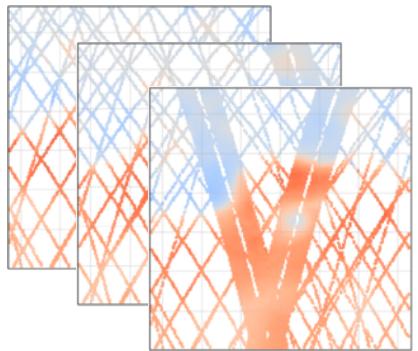
Key messages

- We can bridge DNN and variational models to solve inverse problems
- Learning both variational priors and solvers
- Learning “optimal” parameterisations w.r.t. a predefined performance metrics (e.g., reconstruction score)

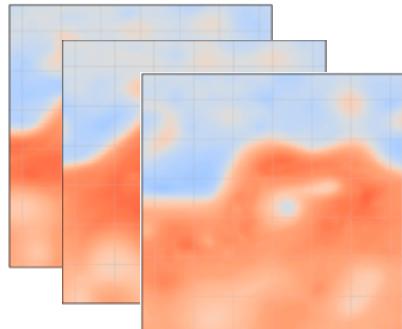
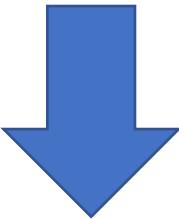


Preprint: <https://arxiv.org/abs/2006.03653>
Code: <https://github.com/CIA-Oceanix>

4DVarNN: Application to satellite-derived SSH mapping



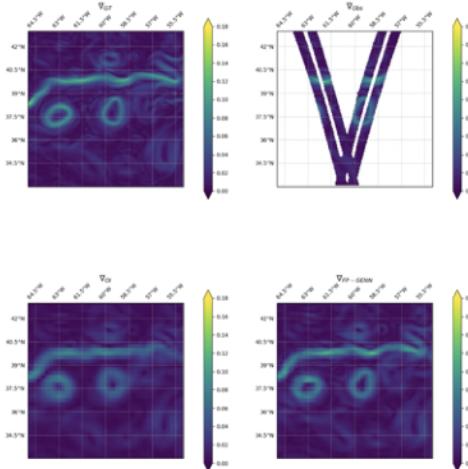
Partial observations y



True states x

OSSE:

- NATL60 data
- Noise-free SWOT+4-nadir/4-nadir observation dataset
- Gap-free/noise-free SST
- Test period on 20 days (remaining data used for training)



Results

Models	4DVarNN-L2-LSTM	4DVarNN-L2-LSTM+SST
relative MSE gain vs. OI for SSH	53 %	64 %
Relative MSE gain vs. OI for gSSH	45 %	58 %

SWOT+4-nadir

Models	4DVarNN-L1-LSTM	4DVarNN-L2-LSTM+SST
relative MSE gain vs. OI for SSH	12 %	52 %
Relative MSE gain vs. OI for gSSH	18 %	52 %

4-nadir

4DVarNN: end-to-end-learning of 4DVar models

From a Variational DA formulation

$$\hat{x} = \arg \min_x \|x - y\|_{\Omega}^2 + \lambda \|x - \Phi(x)\|^2$$

Trainable prior operator

Trainable gradient-based solver
using AD

To a End-to-end architecture

