HAL
open science

Simultaneous Assimilation and Downscaling of Simulated Sea Surface Heights with Deep Image Prior

Arthur Filoche, Dominique Béréziat

To cite this version:

Arthur Filoche, Dominique Béréziat. Simultaneous Assimilation and Downscaling of Simulated Sea Surface Heights with Deep Image Prior. RFIAP (Congrès Reconnaissance des Formes, Image, Apprentissage et Perception), Jul 2022, Vannes, France. hal-03682250

HAL Id: hal-03682250

https://hal.sorbonne-universite.fr/hal-03682250v1

Submitted on 30 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Simultaneous Assimilation and Downscaling of Simulated Sea Surface Heights with Deep Image Prior

Arthur Filoche[1] Dominique Béréziat[1]

[1] Sorbonne Université, CNRS, LIP6

arthur.filoche@lip6.fr

Résumé

Les observations océanographiques existent à différentes résolutions spatio-temporelles et peuvent être assimilées à résolutions variables. La disponibilité de simulations numériques, telles que des réanalyses, rendent pertinent l'apprentissage supervisé pour traiter des problèmes inverses liés aux échelles. Toutefois, l'assimilation de données aux résolutions les plus fines de ces modèles océanographiques est très coûteux en ressource de calcul et construire ainsi une base d'apprentissage n'est difficilement qu'accessible. Dans ce travail, nous nous intéressons à l'utilisation d'un réseau génératif de type "deep image prior" à l'intérieur même d'un algorithme d'assimilation variationnelle de données et cela pour assimiler à échelle grossière des observations de la hauteur d'eau de la surface océanographique. L'algorithme reconstruit alors les hauteurs d'eau à échelle plus fine et cela de façon non supervisée. Nous illustrons les performances de l'algorithme sur des expériences jumelles en utilisant un modèle shallow-water et en les comparant avec des méthodes classiques. Nous constatons que l'utilisation d'un réseau "deep image prior" a des effets régularisants forts.

Mots Clef

Descente d'échelle dans les images de hauteur d'eau, assimilation de données, Deep Image prior, Shallow-water

Abstract

Oceanographic observations exist with different spatio-temporal resolutions and can be assimilated at various resolutions. The availability of numerous numerical simulations like ocean re-analysis makes supervised machine learning appealing to deal with scale-related inverse problems. However, data assimilation at finest resolutions using detailed oceanographic models is computationally intensive, and building an exhaustive database may not be practical. In this work, we investigate the use of a deep image prior within the variational data assimilation framework to simultaneously assimilate and downscale sea surface height observations. The algorithm then estimates the whole high-resolution system state trajectory in a fully-unsupervised manner. We set a twin assimilation experiment using a shallow-water model and compare the results with traditional methods to demonstrate its use. We observe that using a deep image has a strong regularizing effect.

Keywords

Sea surface height downscaling, Data assimilation, Deep Image Prior, Shallow water.

1 Introduction

Monitoring and modeling the Ocean is a constant scientific preoccupation whether for global climate understanding or numerical weather prediction. To do so, information from various sensors are combined with physics-based dynamical models in a data assimilation scheme, in order to estimate the state of the ocean. However, available data and known physics came with numerous spatio-temporal resolutions [14, 30] which leads to scale-matching problems [15, 28].

Data assimilation [8] provides a strong framework to handle various observations and physical models at different scales. Approaches are being developed to assimilate observations directly in high-resolution [3]. Such problems being often ill-posed, regularization is needed. It is usually done parametrizing explicit prior knowledge to enforce satisfying solution. Even though variational data assimilation has a long standing experience in model-constrained optimization, deep learning techniques have revolutionized inverse problems solving [26]. Leveraging an exhaustive database, it is possible to directly learn such parametrization [6]. But available satellite observations are usually sparse, noisy and indirect, so that a supervised learning set up is not realistic. Regarding the super-resolution (or downscaling), it has been shown in this survey [34] that an unsupervised method like Deep Image Prior (DIP) [32] can provide

performances close to supervised approaches.
In this work we propose to use a DIP in a data assimilation framework to simultaneously assimilate and downscale observations. The architecture is trained in a fully-unsupervised manner using the variational assimilation loss so that gradients are backpropagated through the observation operator, the dynamical model, and then the neural network.
Finally we test the proposed method in a twin experiment based on a shallow water dynamical model and compare it with classical methods, still used in operational numerical weather forecasting. We show that using a deep prior provides interesting regularizing properties for the studied inverse problem. The code is available at the Github address [1].

2 Related work

2.1 Unsupervised Inversion

Using the inverse problem in imaging classification provided in [26], variational data assimilation and deep image prior [32] both belong to the unlearned methods class. In both cases the forward model is partially known and the training done only using observation, so they perform unsupervised inversion with the particularity to be optimized on a single observation. Using multiple series of observations, adversarial learning could be considered as depicted in [7].

2.2 Error in Data Assimilation

The partially known forward operator is usually the result of many years of physical and sensor modeling, back-tested and tuned on observations. Even best models are not able to resolve physics at all scale so that representation errors persist [16] which lead to bias in the estimation [10]. Correcting the forward model with offline or online machine learning methods is still investigated [12].

2.3 Physics-constrained Learning

Variational data assimilation has a pioneering expertise in PDE-constrained optimization [19], making use of automatic differentiation to retro-propagate gradients through the dynamical system. Several physics-constrained deep learning architectures already exist and [5] describes a general framework. In [9] and [24] the output of a neural network is used as input in a dynamical model and architectures are trained in supervised and adversarial manner, respectively. We proceed similarly but using a deep image prior.

2.4 Downscaling Ocean simulation

Increasing the resolution of numerical simulations allows to represent finer physical dynamics, instead of representing it with parametrization [6], and then better explain spatio-temporal variability of eddy fields. But when coupled with various satellite observations in a data assimilation scheme, the interpolation tends to smooth small-scale processes [15] and results in diminishing forecast skills [28]. By simultaneously assimilating and increasing the resolution of observations, we hope to participate in developing methods softening this issue.

[1] https://github.com/EarthVision22id44/cvpr_workshop

3 Method

3.1 Data assimilation framework

A dynamical system is considered where a state $\mathbf{X}$ evolves over time following a perfectly-known dynamics $\mathbb{M}$, see Eq. (1). Partial and noisy observations $\mathbf{Y}$ are available through an observation operator $\mathbb{H}$, Eq. (2). A background $\mathbf{X}_b$ gives prior information about the initial system state, Eq. (3).

$$\text{Evolution:} \quad \mathbf{X}_{t+1} = \mathbb{M}_t(\mathbf{X}_t) \tag{1}$$

$$\text{Observation:} \quad \mathbf{Y}_t = \mathbb{H}_t(\mathbf{X}_t) + \varepsilon_{R_t} \tag{2}$$

$$\text{Background:} \quad \mathbf{X}_0 = \mathbf{X}_b + \varepsilon_B \tag{3}$$

Additive noise ε_B and ε_R represent uncertainties about the observations and the background, respectively. These noises are quantified by their assumed known covariance matrices $\mathbf{B}$ and $\mathbf{R}$, respectively. The dynamics is here considered perfect, but the framework could easily be extended to an imperfect dynamics. For any given matrix $\mathbf{A}$, we note $\|x-y\|_A^2 = \langle (x-y) | \mathbf{A}^{-1}(x-y) \rangle$ the associated Mahalanobis square distance.

3.2 Strong-constraints 4DVar

The objective of data assimilation is to provide an estimation of the system state $\mathbf{X}$ by optimally combining available data, $\mathbf{X}_b$ and $\mathbf{Y}$, and the dynamical model $\mathbb{M}$. In the variational formalism [19], this is done via the minimization of a cost function which is the sum of background errors and observational errors, $\mathcal{J}_{\text{4DVar}} = \frac{1}{2}\|\varepsilon_b\|_{\mathbf{B}}^2 + \frac{1}{2}\sum_{t=0}^{T}\|\varepsilon_{R_t}\|_{\mathbf{R}_t}^2$. The optimization problem is model-constrained and the dynamics being assumed perfect, estimating the full system state trajectory is equivalent to estimate the initial condition, as described in Eq. (4). What motivates this cost function is that minimizing it leads to the maximum *a posteriori* estimation of the state under independent Gaussian errors, linear observation operator and linear model hypothesis. The corresponding optimization algorithm is named 4DVar.

$$\begin{aligned} \min_{\mathbf{X}_0} \quad & \mathcal{J}_{\text{4DVar}}(\mathbf{X}_0) = \frac{1}{2}\|\mathbf{X}_0 - \mathbf{X}_b\|_{\mathbf{B}}^2 + \frac{1}{2}\sum_{t=0}^{T}\|\mathbf{Y}_t - \mathbb{H}_t(\mathbf{X}_t)\|_{\mathbf{R}_t}^2 \\ \text{s.t.} \quad & \mathbf{X}_{t+1} = \mathbb{M}_t(\mathbf{X}_t) \end{aligned} \tag{4}$$

The link between variational assimilation and Tikhonov regularization is well described in [18]. For example,

choosing a particular matrix $\mathbf{B}$ will promote a particular set of solutions. Therefore, making alike choices can be seen as a handcrafted regularization to take advantage of expert prior knowledge.

3.3 Super-resolution

The single image super-resolution task consists in recovering a particular high resolution image denoted $\mathbf{X}^h$ from a low-resolution observation $\mathbf{X}^l$, modeled as the output of a decimation operator denoted d, such that $\mathbf{X}^l = d(\mathbf{X}^h)$. Estimating super-resolution from an observation can be seen as an optimization problem minimizing an energy function of the general form $\|d(\mathbf{X}^h) - \mathbf{X}^l\|$. By rewriting the observation operator $\mathbb{H} = d \circ p$ where p depends on the application, we highlight the fact that the variational data assimilation framework is suited for simultaneous system state estimation and super-resolution. However, as d is non-injective, it increases the ill-posedness of the inverse problem. For the sake of simplicity, we won't use the notation $\mathbf{X}^h$ so that when $\mathbf{X}$, it is implicitly at high-resolution.

3.4 Deep Image Prior

The idea behind DIP is that using a well-suited neural architecture to generate a solution of the variational problem can act as a handcrafted regularization. This means that the control parameters are shifted from the system state space to the neural network parameters space. From a practical standpoint, a generator network $g_{\boldsymbol{\theta}}$ outputs the solution from a latent variable z such that $g_{\boldsymbol{\theta}}(z) = \hat{\mathbf{X}}_0^h$. In our case, we ask the network to output a super-resolution estimation of a full system state trajectory from sparse, noisy and low-resolved observations. An overview of the performed forward integration is drawn in Figure 1.

Cost function. The generator is then trained with the variational assimilation cost $\mathcal{J}(\boldsymbol{\theta})$. To emphasize the regularizing effect of the deep prior method, we choose to fix $\mathbf{B}^{-1} = 0$ so that no background information is used. It means that $\mathcal{J}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{t=0}^{T} \|\varepsilon_{R_t}\|^2_{R_t}$ and by denoting multiple integration between two times $\mathbb{M}_{t_1 \to t_2}$, the cost can be developed as in Eq. (5).

$$\mathcal{J}(\boldsymbol{\theta}) = \frac{1}{2}\sum_{t=0}^{T} \|\mathbf{Y}_t - \mathbb{H}_t(\mathbb{M}_{0\to t}(g_{\boldsymbol{\theta}}(z)))\|^2_{R_t} \tag{5}$$

It is important to note that this approach is unsupervised, the architecture being trained from scratch on one assimilation window with no pre-training. All the prior information should be contained in the architecture choice.

Gradient. The gradient of this cost function can be determined analytically. First, the chain rule gives Eq. (6). Then using the adjoint state method we can develop $\nabla_{\mathbf{X}_0}\mathcal{J}(\mathbf{X}_0)$ as in Eq. (7). In the differentiable programming paradigm, such analytical expression is not needed to obtain gradients, adjoint modeling is implicitly performed

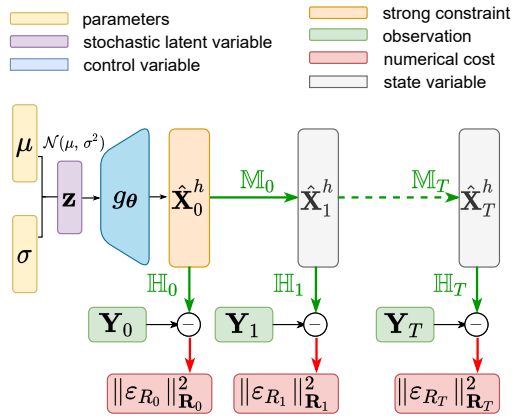

Figure 1: Overview of the forward operator to be optimized in the super-resolution 4DVar with deep image prior

as gradients are backpropagated automatically.

$$\nabla_{\boldsymbol{\theta}}\mathcal{J}(\boldsymbol{\theta}) = \nabla_{\mathbf{X}_0}\mathcal{J}(\mathbf{X}_0)\nabla_{\boldsymbol{\theta}}\mathbf{X}_0 = \nabla_{\mathbf{X}_0}\mathcal{J}(\mathbf{X}_0)\nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}}(z) \tag{6}$$

$$\nabla_{\mathbf{X}_0}\mathcal{J}(\mathbf{X}_0) = \sum_{t=0}^{T}\left[\frac{\partial(\mathbb{H}_t\mathbb{M}_{0\to t})}{\partial \mathbf{X}_0}\right]^{\top} \mathbf{R}_t^{-1}\varepsilon_{R_t} \tag{7}$$

4 Experiments

4.1 Twin experiment

The proposed methodology is tested within a twin experiment where data are generated from a numerical dynamical model. Observations are then created by sub-sampling, decimating (down-sampling) and adding noise. The aim of this experiment is to highlight the regularizing effect of DIP in an excessively ill-posed, ocean-like data assimilation problem. To do so, we compare several super-resolution 4DVar algorithms, one using a deep image prior. We repeat the same experiments with different downscaling ratios and different levels of noise. In all the assimilation experiments, the dynamical model $\mathbb{M}$ and the decimation operator d are assumed perfectly known.

4.2 Dynamical system

State. State variables of the considered system are η, the height deviation of the horizontal pressure surface from its mean height, and $\mathbf{w}$, the associated velocity field. $\mathbf{w}$ can be decomposed in u and v, the zonal and meridional velocity, respectively. At each time t, the system state is then $\mathbf{X}_t = \begin{pmatrix}\eta_t & \mathbf{w}_t^{\top}\end{pmatrix}^{\top}$.

Shallow water model. The dynamical model used here corresponds to a discretization of the shallow water equations system in Eq. (8) with first order explicit numerical schemes. H represents the mean height of the horizontal pressure surface and g the acceleration due to gravity. After reaching an equilibrium starting from

Gaussian random initial conditions, system trajectories are simulated.

$$\begin{cases} \dfrac{\partial \eta}{\partial t} + \dfrac{\partial (\eta + H)u}{\partial x} + \dfrac{\partial (\eta + H)v}{\partial y} = 0 \\ \dfrac{\partial u}{\partial t} + g\dfrac{\partial \eta}{\partial x} = 0 \\ \dfrac{\partial v}{\partial t} + g\dfrac{\partial \eta}{\partial y} = 0 \end{cases} \tag{8}$$

Observations. At regularly-spaced observational dates, a low-resolution of η is observed, up to an additive white noise. The velocity field $\mathbf{w}$ is never observed. The chosen decimation operator d is a 2-dimensional average pooling convolution of kernel size $r \times r$, r being the upscaling factor. This means that at observational date t, the observation operator $\mathbb{H}$ is the composition of a linear projector and d so that $\mathbb{H}\mathbf{X}_t = d(\eta_t)$. White noise is then added. Examples of generated observations are displayed in Figure 2.

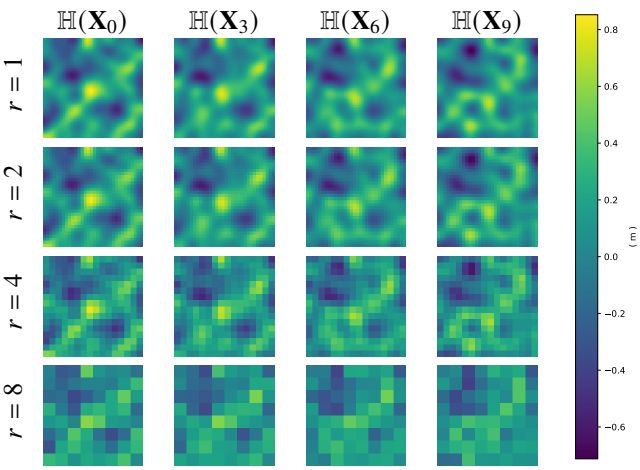

Figure 2: Examples of one noiseless generated observation series, using different down-sampling ratio r

4.3 Algorithms

The objective of each 4DVar algorithm is to estimate the high-resolution true system state $\mathbf{X}$ from low-resolution observation of η. The downscaling factor of the algorithm then corresponds to the upscaling factor of the decimation operator.

4DVar. We denote 4DVar the simple version only optimizing the fit-to-data term in the cost function. No prior knowledge on the solution is used, the background covariance matrix $\mathbf{B}^{-1}$ vanishes. It is optimized with the L-BFGS algorithm.

Regularized 4DVar. We denote Reg. 4DVar the same algorithm additionally optimizing a background penalty term. The estimated motion field is forced to be smooth by constraining $\|\nabla \mathbf{w}_0\|_2^2$ and $\|\nabla . \mathbf{w}_0\|_2^2$ to be small. As proved in [21], these terms can be directly included in the background error using a particular matrix $\mathbf{B}$ such that $\alpha\|\nabla \mathbf{w}_0\|_2^2 + \beta\|\nabla . \mathbf{w}_0\|_2^2 = \|\mathbf{X}_0 - \mathbf{X}_b\|_{B_{\alpha,\beta}}^2$ where α and β are hyper-parameters to be tuned. Such regularization is a classical optical flow regularization and fits in ocean motion estimation [4, 29].

Deep image prior 4DVar. As depicted in the previous section, the only choice to be made is the architecture of the network g_θ. We use the generative architecture with the associated Adam optimizer and momentum parameters presented in [27]. We just replaced deconvolution operations by upsampling followed by convolution to avoid checkerboard artifacts, as described in [25]. Finding hyper-parameters, and particularly the number or epoch, for DIP is still an active research field as described in [33]. Investigated early-stopping methods are beyond the scope of our study so we made the choice to fix the number of epoch to 1000 and only having the learning rate as hyper-parameter.

Numerical details.

Shallow water simulation
The high-resolution ground truth for each physical quantity at each time step ground truth is a 64×64 image. It represents a square area of $10^5 \times 10^5\,\mathrm{m}^2$ so that each pixel corresponds to a square of side $dx = dy \approx 1500\,\mathrm{m}$ mimicking the scale of high-resolution Ocean simulation [22]. H and g are fixed to $100\,\mathrm{m}$ and $9.81\,\mathrm{m.s}^{-2}$, respectively. The integration time step dt is defined for numerical stability as $dt = \min(dx, dy)/2\sqrt{gH}$. 100 ground truth trajectories are simulated.

Observations
The considered temporal window has a fixed length of $10 \times dt$ so that $T = 9$. Observations are sampled at date $t = 0, 3, 6, 9$. Several upscaling factors, $r = 1, 2, 4, 8$, are investigated, $r = 4, 8$ roughly corresponding to the factor between high resolution numerical simulations and sea surface height satellite resolution [30]. Several levels of white noise are investigated. The standard deviation of the noise is expressed as a percentage of the low-resolution observation dynamic range. We investigate 0%, 1%, 2%, and 3% which is the level of noise to be expected in most recent altimeter [17]. For each upscaling factor and each level of noise, 100 series of observations are generated.

Hyper-parameters tuning
For each downscaling factor, hyper-parameters are tuned using Bayesian optimization on low-resolution observations forecasts so that ground truth is never used. The noise level is fixed to 1%. Regarding DIP, we experimentally found similar optimal learning rate for each downscaling factor.

5 Results

For each downscaling factor and each level of noise, we optimized the three different version of 4DVar on 100 series of observations. As the dynamical

model is considered perfect, the initial condition entirely characterize the estimated state so that $\hat{\eta}_0$ and $\hat{\mathbf{w}}_0$ are the quantities to look at to assess the quality of the estimation. Once such condition estimated, it is possible to produce a forecast integrating the dynamics, we will then compute $\hat{\eta}_{T+1}$ and $\hat{\eta}_{T+5}$ to evaluate forecasts performances. All these estimations are in high-resolution and are then compared quantitatively and visually to the ground truth.

5.1 Quantitative results

To quantify the quality of the estimation, we use several metrics the root mean square error (RMSE) and the structural similarity (SSIM) for the the height deviation η. For velocity map we use optical flow metrics, the endpoint error (EPE) and the average angular error (AAE). They calculate the Euclidean distance and the average angular deviation between the estimation and the ground truth, respectively. Quantitative results of the main experiment are displayed in Table 1. Such table is available for each level of noise. We decided to display the one for 1% level of noise as it is the closest to satellite noise range and because we tuned hyper-parameters at this level.

The first interesting result to note is that when no downscaling task is performed ($r = 1$), classical 4DVar algorithms perform better, which makes sense as the optimized cost function has been designed assuming Gaussian errors. However, augmenting r, the DIP 4DVar algorithm seems to be performing better, at least in forecast performances. Regarding metrics quantifying the quality of the motion fields estimation, DIP 4DVar systematically underperforms 4DVar and more particularly Reg. 4DVar.

5.2 Qualitative results

To better understand error sources, we displayed in Figures 3, 4, 5 error maps for $\hat{\eta}_0$, $\hat{\mathbf{w}}_0$ and $\hat{\eta}_{T+5}$. First looking at assimilation results, we see that square patterns tend to appear in estimated states with 4DVar and Reg. 4DVar while DIP 4DVar estimation stay smooth. The harder the downscaling task the bigger the squares. When combining, square degradation propagate through the dynamical model which explains the better forecasts performances of DIP 4DVar. However, DIP 4DVar seems less precise and this is confirmed in Figure 6 in a experiment with no noise, we see that the used deep prior can't perfectly fit observations, contrarily to the classical prior.

5.3 Sensitivity to noise

We repeated similar experiments with different levels of noise and constated in Figure 6 that the used deep prior is robust against noise increases. It is to be reminded that hyper-parameters has been tuned at a single level of noise so thay may be less relevant in other cases.

We believe that the inductive bias from the architecture choice brings a desirable regularity to our problem. For instance the qualitative results show that the estimated motion fields have interesting characteristics that are not

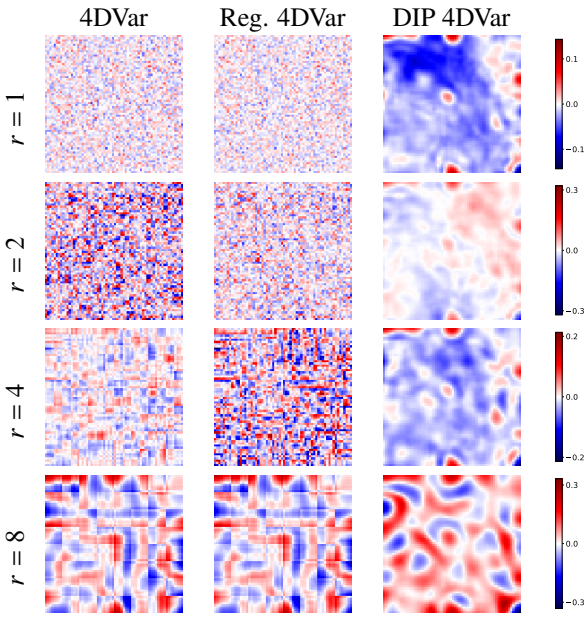

Figure 3: Assimilated height $\hat{\eta}_0$ error maps for each downscaling factors and 4DVar algorithms, noise level of 1%

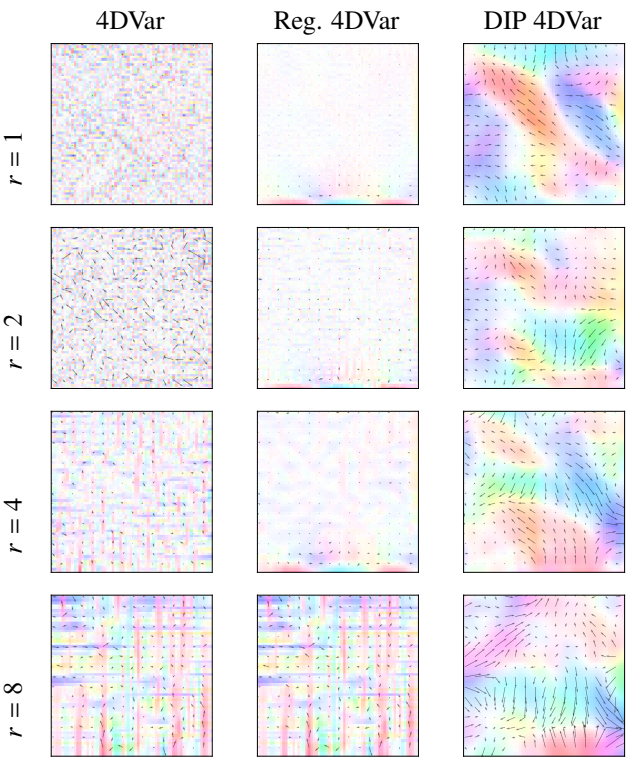

Figure 4: Assimilated $\hat{\mathbf{w}}_0$ error maps for each downscaling factors and 4DVar algorithms, noise level of 1%. Color correspond to the motion field orientation but intensity is normalized, quiver arrow quantify intensity of the motion field

Table 1: Metrics quantifying the quality of the assimilation and the following forecast for various downscaling factors and a fixed level of noise of 1%

	Assimilation				Forecast			
Quantity	$\hat{\eta}_0$		$\hat{w}_0$		$\hat{\eta}_{T+1}$		$\hat{\eta}_{T+5}$	
Metric	RMSE ($\times 10^2$)	SSIM	EPE ($\times 10^2$)	AAE	RMSE ($\times 10^2$)	SSIM	RMSE ($\times 10^2$)	SSIM
				$r = 1$				
4DVar	1.4 ± 0.2	$\mathbf{0.98 \pm 0.01}$	$\mathbf{1.1 \pm 0.2}$	7 ± 1	3.9 ± 0.8	0.87 ± 0.05	3.8 ± 0.7	0.89 ± 0.04
Reg. 4DVar	$\mathbf{1.3 \pm 0.2}$	$\mathbf{0.98 \pm 0.01}$	1.1 ± 0.3	$\mathbf{6 \pm 1}$	$\mathbf{1.8 \pm 0.3}$	$\mathbf{0.96 \pm 0.01}$	$\mathbf{1.8 \pm 0.3}$	$\mathbf{0.97 \pm 0.01}$
DIP 4DVar	2.8 ± 0.9	0.94 ± 0.04	5.6 ± 1.3	31 ± 5	4.2 ± 1.3	0.92 ± 0.04	4.8 ± 1.5	0.91 ± 0.04
				$r = 2$				
4DVar	7.2 ± 2.7	0.74 ± 0.12	5.4 ± 1.8	29 ± 6	16.7 ± 5.7	0.40 ± 0.07	17.7 ± 5.6	0.39 ± 0.12
Reg. 4DVar	7.0 ± 5.8	0.77 ± 0.19	$\mathbf{1.9 \pm 0.6}$	$\mathbf{11 \pm 2}$	6.6 ± 4.7	0.76 ± 0.18	6.9 ± 4.5	0.76 ± 0.16
DIP 4DVar	$\mathbf{2.9 \pm 0.9}$	$\mathbf{0.93 \pm 0.05}$	5.5 ± 1.0	30 ± 5	$\mathbf{4.4 \pm 1.4}$	$\mathbf{0.92 \pm 0.05}$	$\mathbf{5.0 \pm 1.6}$	$\mathbf{0.90 \pm 0.05}$
				$r = 4$				
4DVar	$\mathbf{2.4 \pm 0.3}$	$\mathbf{0.95 \pm 0.01}$	2.1 ± 0.4	13 ± 2	6.0 ± 1.1	0.76 ± 0.07	6.1 ± 1.1	0.77 ± 0.05
Reg. 4DVar	4.7 ± 1.3	0.85 ± 0.06	$\mathbf{1.6 \pm 0.3}$	$\mathbf{9 \pm 1}$	$\mathbf{4.5 \pm 0.9}$	0.85 ± 0.05	$\mathbf{4.6 \pm 1.0}$	0.85 ± 0.04
DIP 4DVar	3.3 ± 0.8	0.92 ± 0.04	6.1 ± 1.3	33 ± 6	5.3 ± 1.3	$\mathbf{0.89 \pm 0.04}$	5.7 ± 1.6	$\mathbf{0.88 \pm 0.04}$
				$r = 8$				
4DVar	8.1 ± 1.6	0.71 ± 0.11	$\mathbf{3.9 \pm 0.3}$	$\mathbf{25 \pm 5}$	11.8 ± 0.9	0.51 ± 0.06	11.9 ± 1.0	0.51 ± 0.06
Reg. 4DVar	8.1 ± 1.6	0.71 ± 0.11	$\mathbf{3.9 \pm 0.3}$	$\mathbf{25 \pm 5}$	11.7 ± 1.0	0.51 ± 0.06	11.8 ± 1.0	0.51 ± 0.07
DIP 4DVar	$\mathbf{7.9 \pm 1.7}$	$\mathbf{0.74 \pm 0.12}$	9.7 ± 3.0	47 ± 9	$\mathbf{11.2 \pm 1.3}$	$\mathbf{0.66 \pm 0.08}$	$\mathbf{10.9 \pm 1.4}$	$\mathbf{0.67 \pm 0.09}$

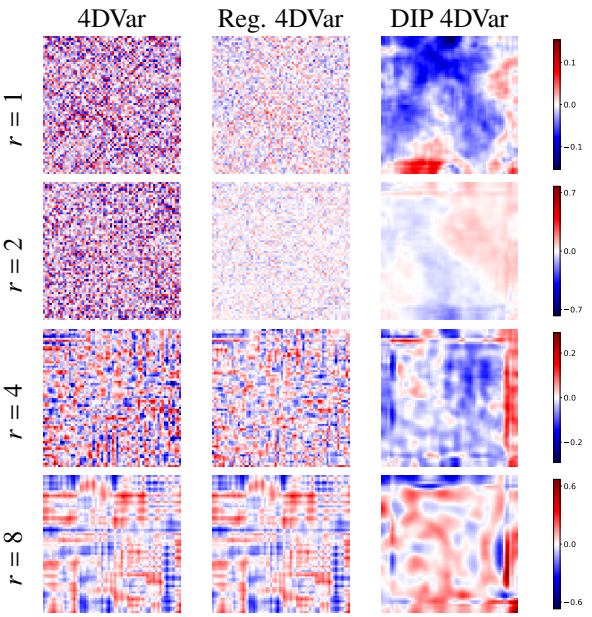

Figure 5: Forecasted $\hat{\eta}_{T+5}$ error maps for each downscaling factors and 4DVar algorithms, noise level of 1%

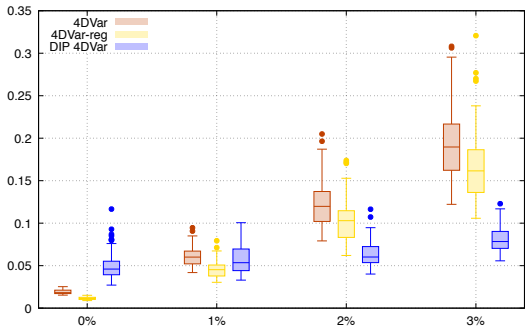

Figure 6: Evolution of RMSE of $\hat{\eta}_{T+5}$ forecasts regarding the level of noise in the observation, with $r = 4$, for each 4DVar algorithm

translated in the chosen optical flow metrics. So we looked at $\|\nabla \mathbf{w}_0\|_2$, $\|\nabla . \mathbf{w}_0\|_2$ and $\|\Delta \mathbf{w}_0\|_2$, statistics characterizing smoothness and observed in Figure 7 that are naturally close the desired ones and not very sensitive to noise increase.

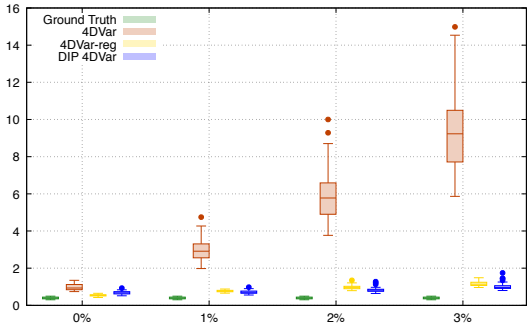

Figure 7: Evolution of $\|\Delta \hat{\mathbf{w}}_0\|$ regarding the level of noise in the observation, with $r = 4$, for each 4DVar algorithm and for the ground truth

5.4 Ensemble of DIP

By training multiple times the same network in the same context, we noticed significant differences in performances. We verified that stochasticity in the latent space stabilizes the optimization as described in the original paper [32]. But we identified that theses differences were mainly explained by weights initialization. It is now well documented that ensemble of neural networks can produced significantly better results [2]. Even though our approach is unsupervised we observed similar behaviors. Training multiple networks with different weights initialization and averaging their outputs, we obtain better estimation than the best one performed by an individual network.

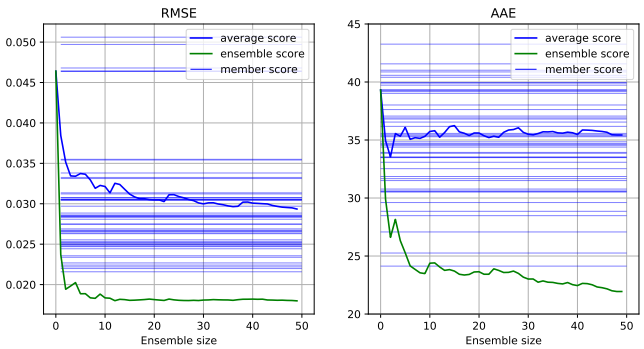

Figure 8: Evolution of deep prior ensembles RMSE and AAE scores, realized on 1 series of observations, downscaling factor $r = 4$, noise level 1%

In Figure 8 we focus on the ×4 super-resolution assimilation one series of observations images. We then optimize 50 deep image prior, and note that scores of averaged estimation are superior to the average score. We arbitrarily displayed RMSE and AAE assimilation scores but same behaviors are obtained with all the metrics. A similar plot with various downscaling factor is presented in Figure 9 and show that the enhanced performances appears every time for every measured metrics.

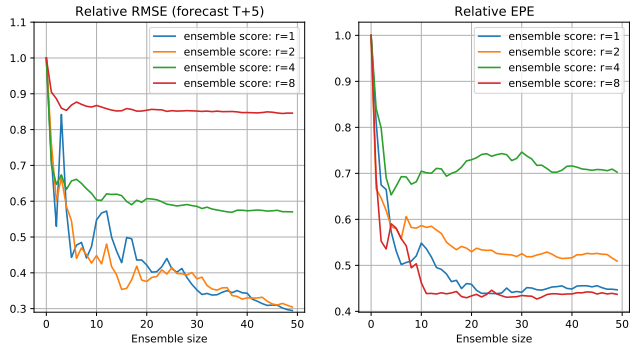

Figure 9: Evolution of deep prior ensembles RMSE (forecast) and EPE scores, realized on 1 series of observations, downscaling factor $r = 1, 2, 4, 8$, noise level 1%, scores presented here are relative so that the whole curve is normalized by the performance of the first member

Looking at errors for different members in Figure 9, we see that they tend to compensate. We conclude that ensemble of deep image prior enhances performances. To be fair, ensemble of 4DVar already exist and could provide similar improvement. However, as the optimization is deterministic others strategies have to be found like adding noise to the observations.

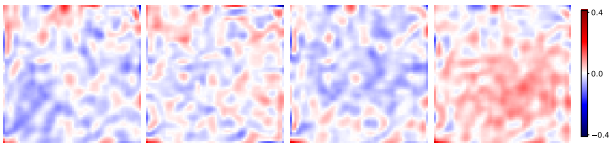

Figure 10: Evolution of deep prior ensembles RMSE (forecast) and AAE scores, realized on 1 series of obsevrations, downscaling factor $r = 1, 2, 4, 8$, noise level 1%

6 Perspectives

6.1 Large-scale ocean data

The presented preliminary study aims to be deployed on large-scala data delivered by the ocean physics-based model NATL60 [1], based on NEMO 3.6 [22] and using initial conditions from MERCATOR [20]. The finality is to propose better unsupervised methods to simultaneously interpolate data from various satellites [14]. Even though scaling the method involves engineering work, the main issue remains the non-availability of the ocean model in a differentiable framework. A solution would be to learn it

from re-analysis with a neural network, differentiable by nature.

6.2 Weak-constraints assimilation

From a data assimilation standpoint, an other possibility would be to relax strong constraints and allow model errors as introduced in [31]. This raises further questions as model error statistics are usually not known. Tackling this problem with machine learning is an active research field [13] and after the presented experiment, we believe DIP may help to capture important statistics. Also, from a numerical optimization perspective, allowing error model in the architecture could be done as ResNet-like shortcut, avoiding complication from backpropagation through time with chaotic models [23].

6.3 Ensemble methods

Quantifying uncertainty is a constant preoccupation in numerical weather forecasting. And this is naturally done by ensemble data assimilation methods, like the ensemble Kalman filter, but not by variational ones, even if hybrid methods already exist [11]. By sampling the latent space of our model, it is possible to obtain probabilistic forecasts in the same fashion. However, we did not observe significant correlation between errors and estimated uncertainty and do not fully understand the role in the estimation, besides from numerical optimization, of the stochasticity introduced in the latent space.

7 Conclusion

We presented a framework bridging unsupervised inversions image methods with physics-constrained problems inspired by geosciences, where ground truth is rarely known. We show in a twin experiment that using a deep image prior in a 4DVar algorithm brings interesting regularity in very ill-posed case. Finally, we drew perspectives towards further developments.

References

[1] A. Ajayi, J. Le Sommer, E. Chassignet, J.-M. Molines, X. Xu, A. Albert, and E. Cosme. Spatial and temporal variability of north atlantic eddy field at scale less than 100 km. *Earth and Space Science Open Archive*, page 28, 2019.

[2] Z. Allen-Zhu, Y. Li, and Z. Allen-Zhu. Towards understanding ensemble, knowledge distillation and self-distillation in deep learning. arXiv preprint, December 2020.

[3] S. Barthélémy, J. Julien Brajard, L. Bertino, and F. Counillon. Super-resolution data assimilation. arXiv:2109.08017, 2021.

[4] D. Béréziat and I. Herlin. Image-based modelling of ocean surface circulation from satellite acquisitions. In *International Conference on Computer Vision Theory and Application (VISAPP)*, pages 1–8, January 2014.

[5] T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine. Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, 126(9):1079–7114, Mar 2021.

[6] T. Bolton and L. Zanna. Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, 11(1):376–399, 2019.

[7] A. Bora, E. Price, and A. Dimakis. AmbientGAN: Generative models from lossy measurements. In *International Conference on Learning Representations*, 2018.

[8] A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):e535, 2018.

[9] Emmanuel de Bézenac, Arthur Pajot, and Patrick Gallinari. Deep learning for physical processes: Incorporating prior scientific knowledge. *CoRR*, abs/1711.07970, 2017.

[10] D. P. Dee. Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343, 2005.

[11] G. Desroziers, J.-T. Camino, and L. Berre. 4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations. *Quarterly Journal of the Royal Meteorological Society*, 140(684):2097–2110, 2014.

[12] A. Farchi, M. Bocquet, P. Laloyaux, M. Bonavita, and Q. Malartic. A comparison of combined data assimilation and machine learning methods for offline and online model error correction. arXiv 2107.11114, 2021.

[13] A. Farchi, P. Laloyaux, M. Bonavita, and M. Bocquet. Using machine learning to correct model error in data assimilation and forecast applications. *Quarterly Journal of the Royal Meteorological Society*, 147(739):3067–3084, 2021.

[14] L. Gaultier, C. Ubelmann, and L.-L. Fu. The challenge of using future swot data for oceanic field reconstruction. *Journal of Atmospheric and Oceanic Technology*, 33(1):119 – 126, 2016.

[15] H. Hewitt, M. Bell, E. Chassignet, A. Czaja, D. Ferreira, S. Griffies, P. Hyder, J. McClean, A. New, and M. Roberts. Will high-resolution global

ocean models benefit coupled predictions on short-range to climate timescales? *Ocean Modelling*, 120:120–136, December 2017.

[16] T. Janjić, N. Bormann, M. Bocquet, J. A Carton, S. E. Cohn, S. L. Dance, S. N. Losa, N. K. Nichols, R. Potthast, J. A. Waller, and P. Weston. On the representation error in data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 144(713):1257–1278, 2018.

[17] M. Jiang, K. Xu, X. Xu, L. Shi, X. Yu, and P. Liu. Range noise level estimation of hy-2b radar altimeter and its comparison with jason-2 and jason-3 altimeters. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 8312–8315, 2019.

[18] C. Johnson, B. Hoskins, and N. Nichols. A singular vector perspective of 4D-Var: Filtering and interpolation. *Quarterly Journal of the Royal Meteorological Society*, 131:1–19, 2005.

[19] F.-X. Le Dimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, 38A(10):97, 1986.

[20] J.-M. Lellouche, E. Greiner, O. Le Galloudec, C. Regnier, M. Benkiran, C.-E. Testut, R. Bourdalle-Badie, M. Drevillon, G. Garric, and Y. Drillet. Mercator ocean global high-resolution monitoring and forecasting system. *New Frontiers in Operational Oceanography*, pages 563–592, 2018.

[21] Y. Lepoittevin and I. Herlin. Regularization terms for motion estimation. Links with spatial correlations. In *International Conference on Computer Vision Theory and Applications VISAPP*, volume 3, pages 458–466, 2016.

[22] G. Madec, R. Bourdallé-Badie, P.-A. Bouttier, C. Bricaud, D. Bruciaferri, D. Calvert, J. Chanut, E. Clementi, A. Coward, D. Delrosso, et al. Nemo ocean engine, 2017.

[23] L. Metz, C. Daniel Freeman, S. Schoenholz, and T. Kachman. Gradients are not all you need. *CoRR*, abs/2111.05803, 2021.

[24] L. Mosser, O. Dubrule, and M. Blunt. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. *Mathematical Geoscience*, pages 53–79, 2018.

[25] A. Odena, V. Dumoulin, and C. Olah. Deconvolution and checkerboard artifacts. *Distill*, 2016.

[26] G. Ongie, A. Jalal, C. Metzler, R. Baraniuk, A. Dimakis, and R. Willett. Deep learning techniques for inverse problems in imaging. *IEEE Journal on Selected Areas in Information Theory*, 1(1):39–56, 2020.

[27] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *International Conference on Learning Representation (ICLR)*, 2016.

[28] P. Sandery and P. Sakov. Ocean forecasting of mesoscale features can deteriorate by increasing model resolution towards the submesoscale. *Nature Communications*, 8(1):1566, Nov 2017.

[29] D. Suter. Motion estimation and vector splines. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 1994.

[30] G. Taburet, A. Sanchez-Roman, M. Ballarotta, M.-I. Pujol, J.-F. Legeais, F. Fournier, Y. Faugere, and G. Dibarboure. Duacs dt2018: 25 years of reprocessed sea level altimetry products. *Ocean Science*, 15(5):1207–1224, 2019.

[31] Y. Trémolet. Accounting for an imperfect model in 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 132:2483–2504, 2006.

[32] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[33] H. Wang, T. Li, Z. Zhuang, T. Chen, H. Liang, and J. Sun. Early stopping for deep image prior. *CoRR*, abs/2112.06074, 2021.

[34] Z. Wang, J. Chen, and S. C. H. Hoi. Deep learning for image super-resolution: A survey. *Transactions on Pattern Analysis and Machine Intelligence*, pages 3365–3387, 2021.