Variational assimilation of ocean observation with deep prior

Arthur Filoche[1], Théo Archambault[1,2], Anastase Charantonis[2] and Dominique Béréziat[1]

[1]LIP6, Sorbonne Université, Paris

[2]LOCEAN, Sorbonne Université, Paris

July 6, 2022

MACLEAN Workshop

Data Assimilation - reminder

Classical Geosciences context:

- **Imperfect observations**: sparse, noisy, incomplete, indirect..
- Physics-based, PDE-driven **dynamical model**
- Forecast? Estimate **initial conditions**
- **Data Assimilation**: optimally combine model and observations

Data Assimilation framework:

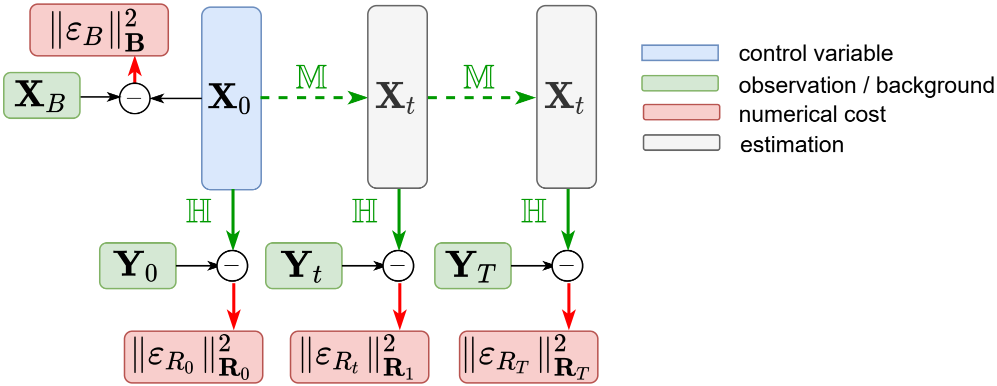
- System state: $\mathbf{X}_t$
- Dynamics: $\mathbf{X}_{t+1} = \mathbb{M}(\mathbf{X}_t)$
- Observations: $\mathbf{Y}_t = \mathbb{H}(\mathbf{X}_t) + \varepsilon_{R_t}$
- Background: $\mathbf{X}_0 = \mathbf{X}_B + \varepsilon_B, \quad \mathbf{B} = \mathbb{E}[\varepsilon_B^T \varepsilon_B]$

Variational inverse problem:

- Bayesian inversion: $\mathbb{P}(\mathbf{X}_0 | \mathbf{Y}_{0:T}, \mathbf{X}_B)$
- Under **linear / Gaussian hypothesis**: $\log \mathbb{P}(\mathbf{X}_0 | \mathbf{Y}_{0:T}, \mathbf{X}_B) \propto -J(\mathbf{X}_0)$
- **Optimization** problem:

$$\min_{\mathbf{X}_0} \quad J(\mathbf{X}_0) = \underbrace{\sum_t \|\mathbf{Y}_t - \mathbb{H}(\mathbf{X}_t)\|^2_{R_t}}_{\text{fit-to-data}} + \underbrace{\|\mathbf{X}_0 - \mathbf{X}_B\|^2_B}_{\text{Regularizer}} \quad \text{s.t.} \quad \mathbf{X}_{t+1} = \mathbb{M}(\mathbf{X}_t)$$

4DVar forward operator:



Gradients: adjoint state method ≈ backpropagation algorithm

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

DL-like design: Only the forward operator is coded, avoid explicit adjoint modeling

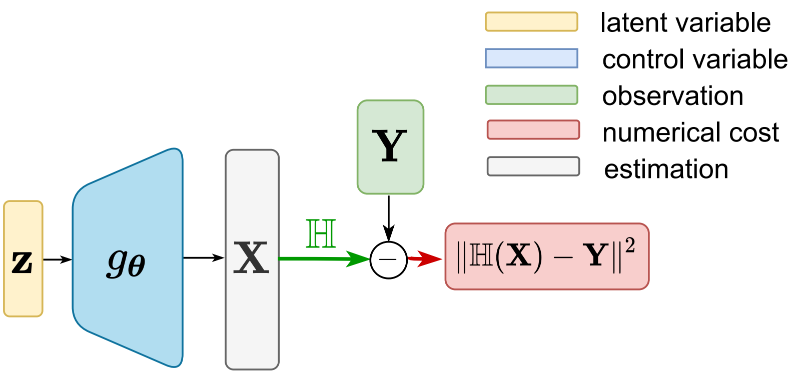
Deep Image Prior

Deep Image Prior [Ulyanov et al. 2018]:

- **Unlearned method** = "Unsupervised with one sample"
- **Solve various imaging inverse problem**: denoising, super-resolution, inpainting
- **Prior**: the solution has to be generated by a CNN

Schematic view:

- Data observed through a **known observation operator**: $\mathbb{H}(\mathbf{X}) = \mathbf{Y}$
- **Only the "fit-to-data" term is optimized**: no additional cost/regularizer
- g_θ is **optimized on one image only!**



DIP example: Inpainting of natural image

Inpainting example: [Ulyanov et al. 2018]

- ▷ $\mathbb{H}$ is a masking operator

Deep Image Prior 13

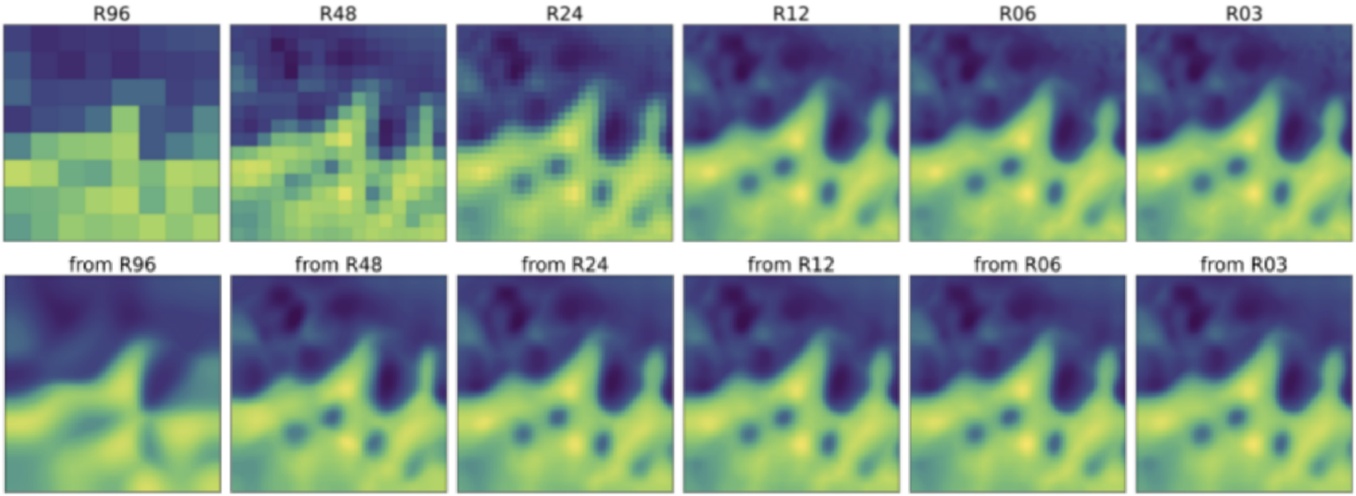
(a) Input (white=masked) (b) Encoder-decoder, depth=6

DIP example: Super-resolution of SSH

[Archambault et al., 2022]:

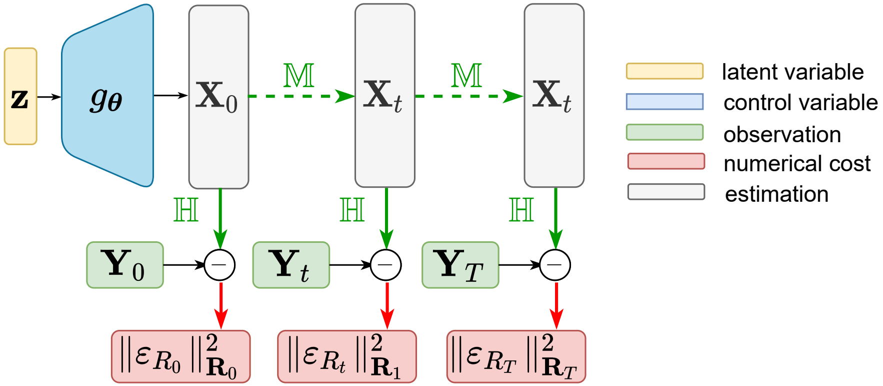
Unsupervised Downscaling of Sea Surface Height with Deep Image Prior

▷ $\mathbb{H}$ is a subsampling operator



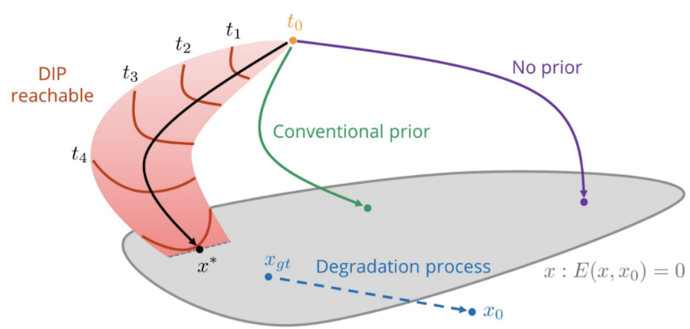
Deep Prior 4DVAR

Forward operator: control parameters are moved into the generator g_θ



Gradients: $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) = \nabla_{\mathbf{X}_0} J(\mathbf{X}_0) \nabla_{\boldsymbol{\theta}} \mathbf{X}_0 = \nabla_{\mathbf{X}_0} J(\mathbf{X}_0) \nabla_{\boldsymbol{\theta}} g_{\boldsymbol{\theta}}(z)$

Mind visualisation [Ulyanov et al. 2018]: Exploring the solution space differently



System:

State

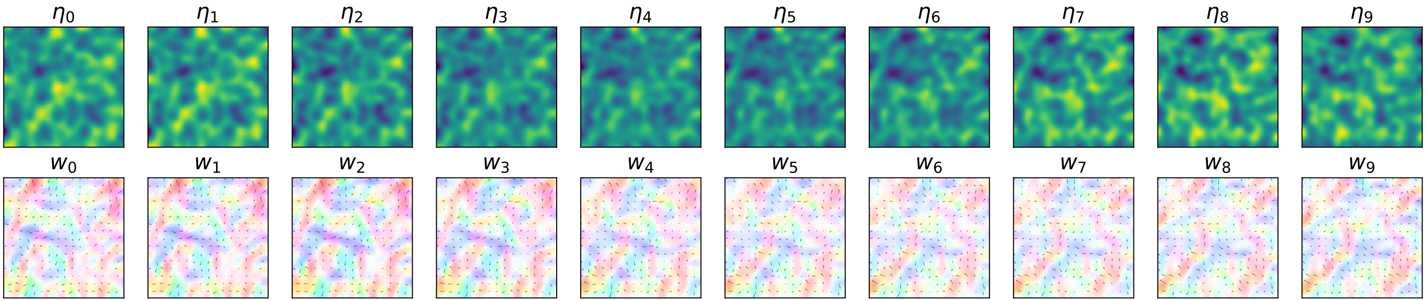
$$\mathbf{X}_t = \begin{bmatrix} \eta_t \\ \mathbf{w}_t \end{bmatrix} \in \mathbb{R}^{3\times 64\times 64}$$

η: SSH-like
$\mathbf{w}$: velocity field

Shallow water dynamics

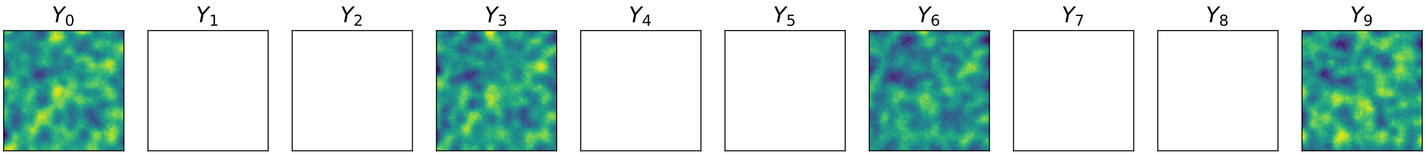
$$\begin{cases} \frac{\partial \eta}{\partial t} + \frac{\partial(\eta + H)u}{\partial x} + \frac{\partial(\eta + H)v}{\partial y} = 0 \\ \frac{\partial u}{\partial t} + g\frac{\partial \eta}{\partial x} = 0 \\ \frac{\partial v}{\partial t} + g\frac{\partial \eta}{\partial y} = 0 \end{cases}$$

Examples of simulated trajectory:



Simulated observations:

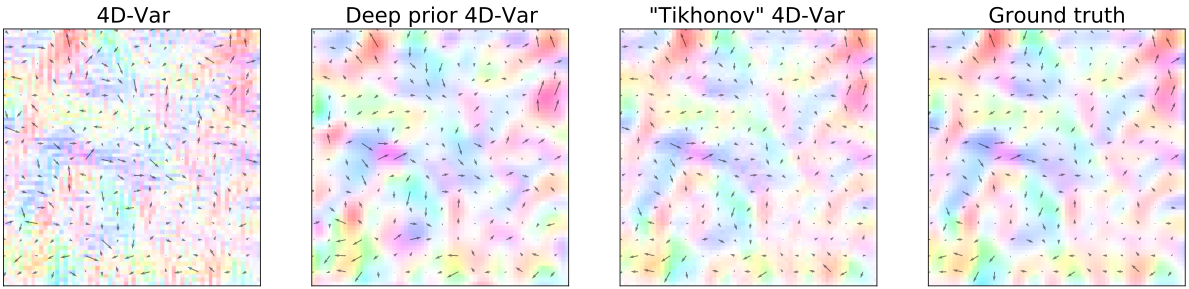
- ▷ Only η_t can be observed, up to an **additive white noise**
- ▷ **Temporal subsampling**: $\mathbb{H}_t$ depends on time and is a linear projector
- ▷ Goal: **estimate $\mathbf{w}_0$**



Tested algorithms:

- ▷ **4D-Var without regularization**: only optimize the "fit-to-data" term
- ▷ **4D-Var with background regularization**:
  - additionaly optimize $\|\mathbf{X}_0 - \mathbf{X}_b\|^2_{B_{\alpha,\beta}} = \alpha\|\nabla\mathbf{w}_0\|_2^2 + \beta\|\nabla.\mathbf{w}_0\|_2^2$
  - optical flow based regularization, enforcing smooth solution (expert design)
- ▷ **Deep Prior 4D-Var**: only optimize the "fit-to-data" term

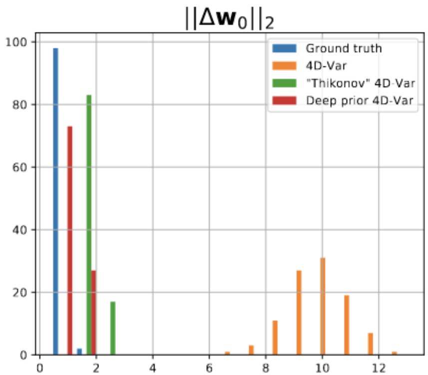
Estimated $\hat{\mathbf{w}}_0$:



Assimilation score

	Assimilation Score	
Metric[1]	Endpoint error ($\times 10^2$)	Angular error
4D-Var	4.2 ± 0.4	28.4 ± 9.8
Deep Prior 4D-Var	4.6 ± 2.0	26.7 ± 5.0
"Tikhonov" 4D-Var	1.6 ± 0.6	9.9 ± 9.8

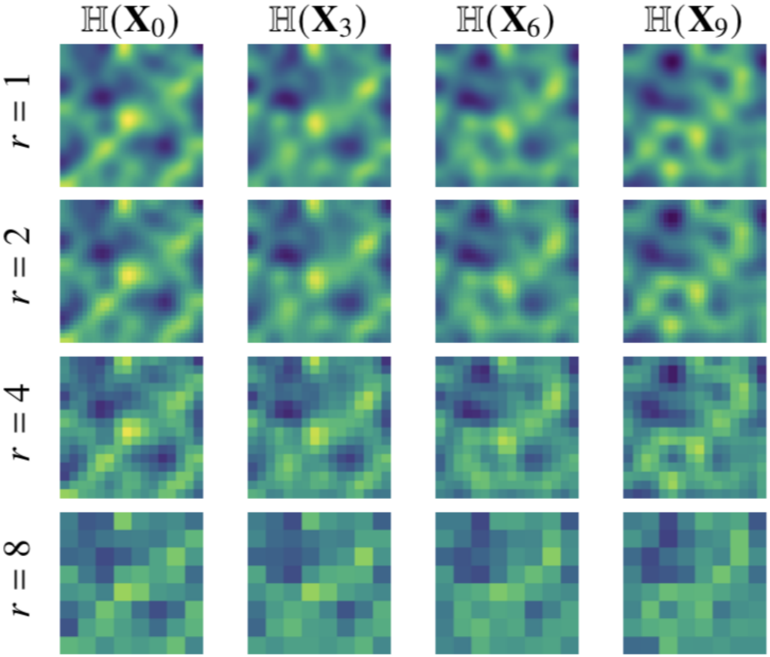
Smoothness statistic



Perspective - Simultaneous super-resolution

Simultaneous assimilation and downscaling:

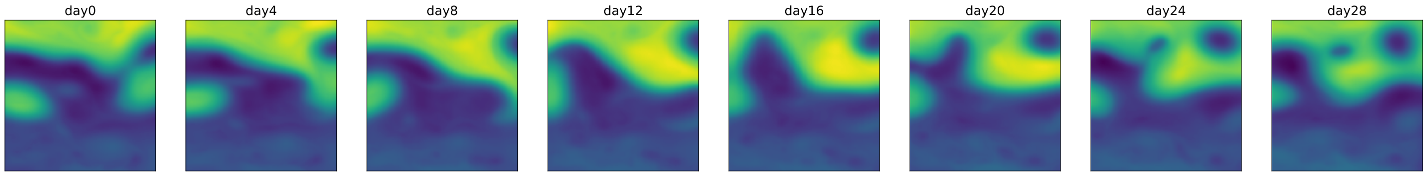
- $\mathbb{H}' = d \circ \mathbb{H}$, d being a **subsampling** operator (non injective)
- **Increasing the ill-posedness** of the inverse problem



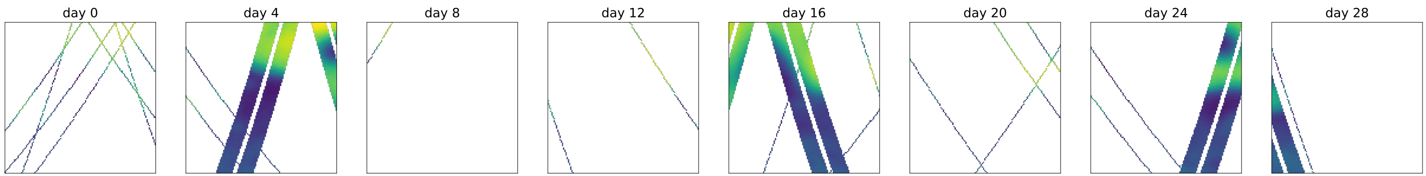
Results: strong regularizing effect of DIP

Case study - SSH altimeter interpolation

Ground truth Sea Surface Height:

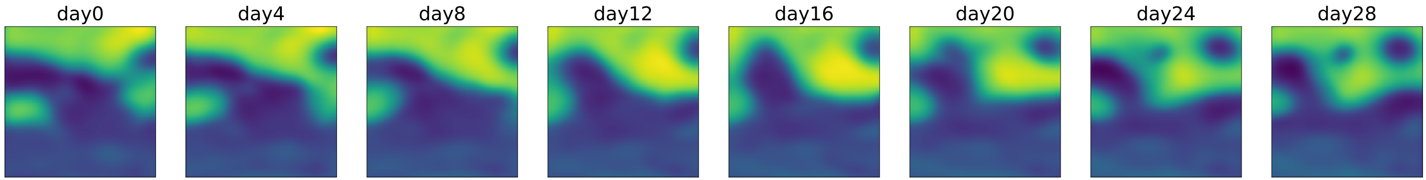


Simulated altimeter satellite tracks:



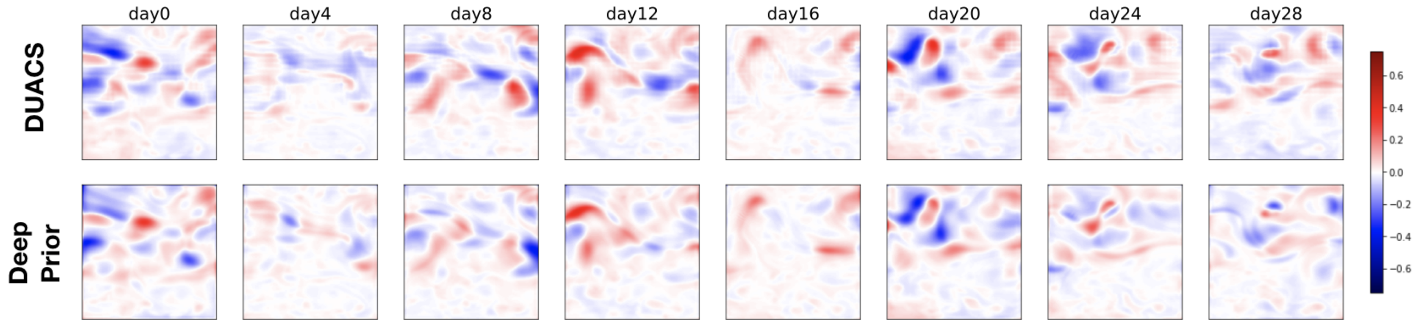
Spatio-temporal deep prior:

- ▷ no dynamical model, time is treated like space

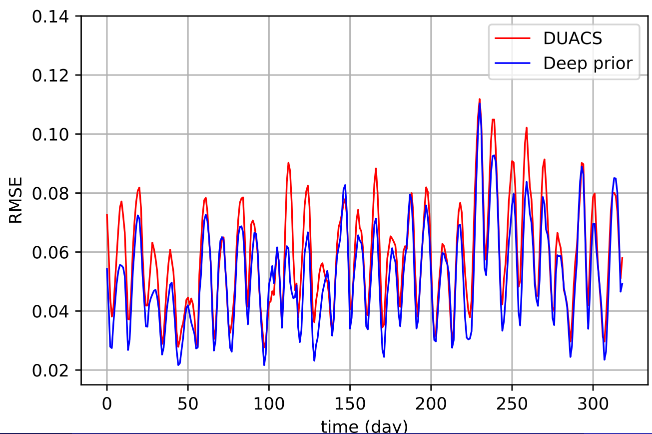


Comparison with optimal interpolation (3DVAR):

▷ OI use historical statistics to specify background regularization



Year long analysis:

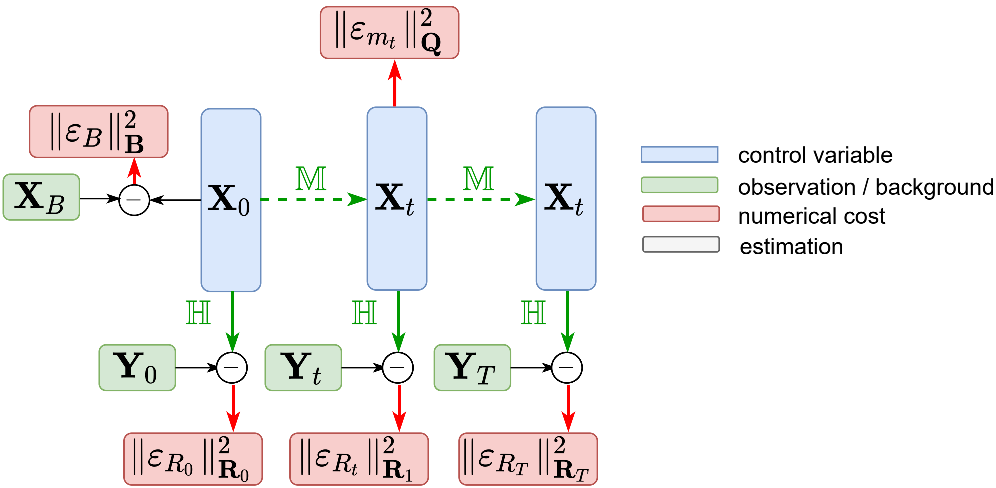


Accounting for model error:

- Dynamics: $\mathbf{X}_{t+1} = \mathbb{M}(\mathbf{X}_t) + \varepsilon_{m_t}, \quad \mathbf{Q}_t = \mathbb{E}[\varepsilon_{m_t}^T \varepsilon_{m_t}]$
- **Optimization** problem:

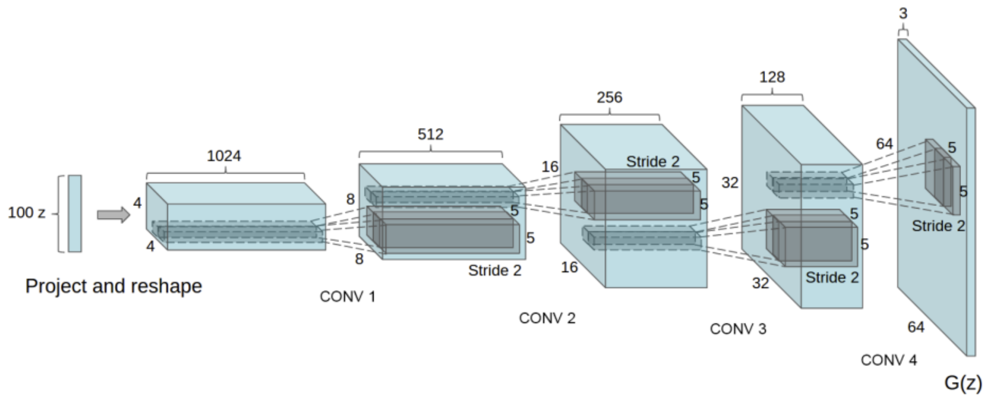
$$\min_{\mathbf{X}_0} \quad J(\mathbf{X}_0) = \underbrace{\sum_t \|\mathbf{Y}_t - \mathbb{H}(\mathbf{X}_t)\|^2_{R_t}}_{\text{fit-to-data}} + \underbrace{\|\mathbf{X}_0 - \mathbf{X}_B\|^2_B}_{\text{Background}} + \underbrace{\sum_{t>0} \|\mathbf{X}_{t+1} - \mathbb{M}(\mathbf{X}_t)\|^2_{Q_t}}_{\text{Model error}}$$

- **Control parameters are set all along the trajectory**



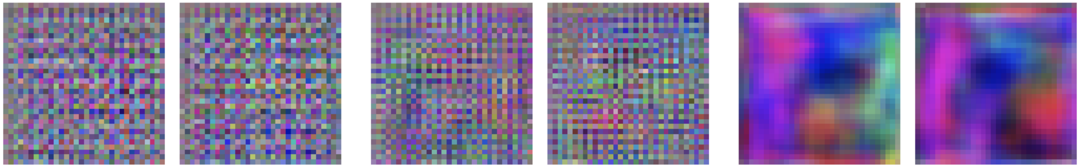
Design of deep prior ?: Model-errors covariance free assimilation?

"DCGAN" generator [Radford et al. 2016]:



Replace deconvolution [Odena et al. 2016]:

- ▷ Resize convolution ≈ upsampling+convolution
- ▷ Avoid checkerboard artifact

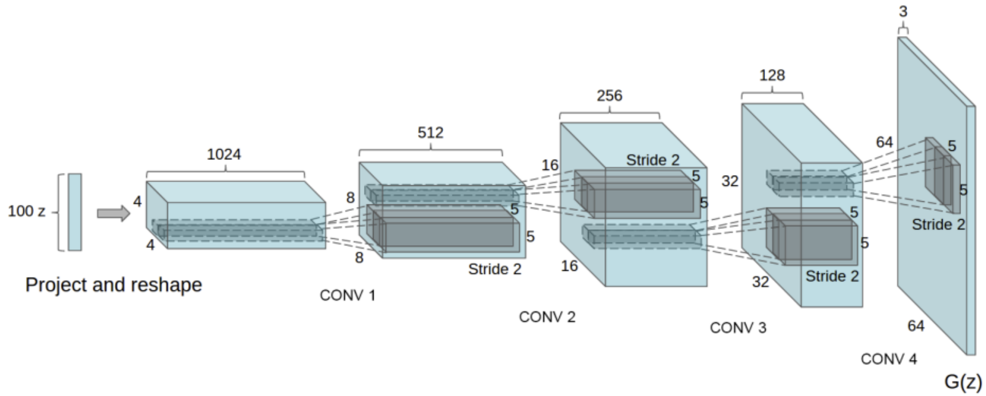


Deconvolution in last two layers.
Artifacts prior to any training.

Deconvolution only in last layer.
Artifacts prior to any training.

All layers use resize-convolution.
No artifacts before or after training.

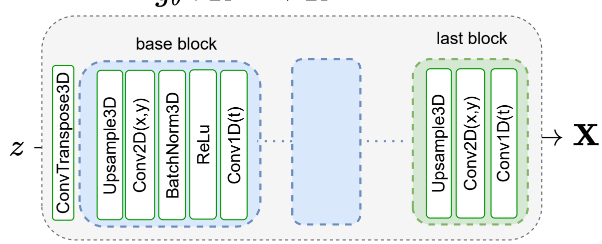
"DCGAN" generator [Radford et al. 2016]:



Spatio-temporal deep prior :

- ▷ (2+1)D Convolutions [Tran et al. 2017]

$$g_\theta : \mathbb{R}^{n_z} \to \mathbb{R}^{n_x \times n_y \times T}$$



Hyper parameter tuning

Optimized metric:

- **Forecast performances** on observations (RMSE)
- So that **ground truth is never used**

4DVar:

- α and β are tuned using Bayesian optimization

DIP-4DVar:

- [Wang et al. 2021] Early-stopping methods
- [Heckel et al. 2018] "Deep Decoder" architecture, circumvent early stopping
- Our choice: **fix the number of epoch and tuned the learning rate** (no big improvement)

Simultaneous assimilation and downscaling

Qualitative assimilation results: $\hat{\mathbf{w}}_0$

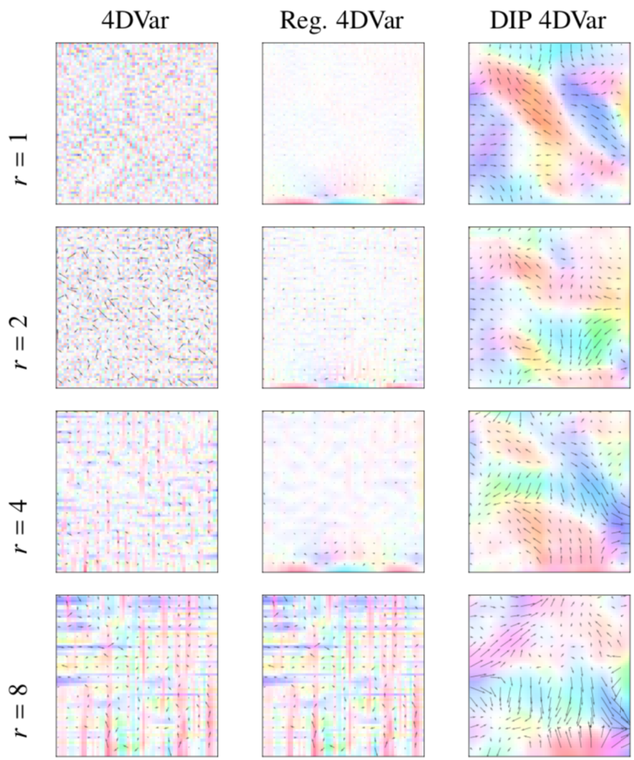


Figure 4. Assimilated $\hat{\mathbf{w}}_0$ error maps for each downscaling factors and 4DVar algorithms, noise level of 1%. Color correspond to the motion field orientation but intensity is normalized, quiver arrow quantify intensity of the motion field

Simultaneous assimilation and downscaling

Qualitative assimilation results: $\hat{\eta}_0$

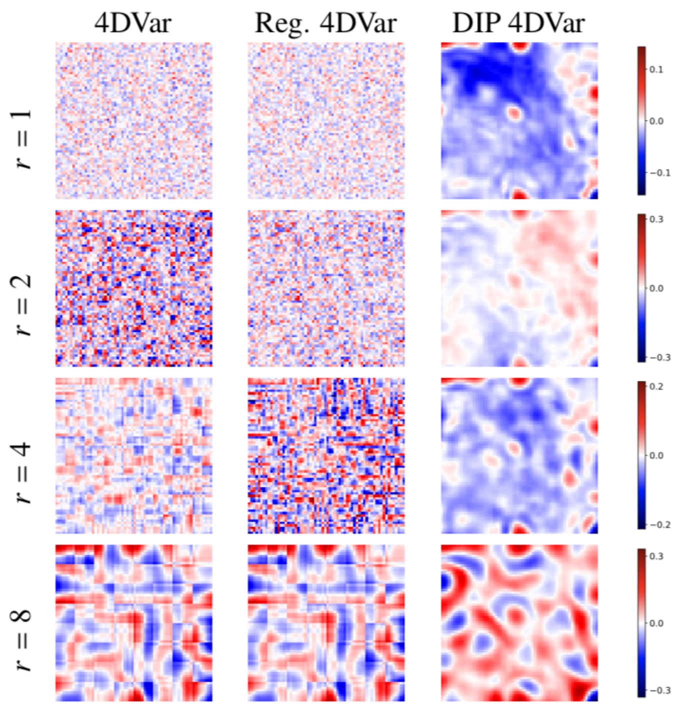


Figure 3. Assimilated height $\hat{\eta}_0$ error maps for each downscaling factors and 4DVar algorithms, noise level of 1%