Physically consistent sampling for ocean model initialization

Blandine Gorce - Luther Ollier
David Kamm - Etienne Meunier

NeurIPS Workshop 2025
Tackling Climate Change with Artificial Intelligence

Ocean modeling & ‘spin-up’ - Overview

Challenge: To initialise an ocean model, it is needed to perform a **spin-up** (initialization phase reach stable phase). This implies a high initial computational costs for ocean modeling.

Spin-up at IPSL
=
2.5M cpu-hours/year

Use generative models to **learn the distribution of initial ocean states** and accelerate the initialization phase.

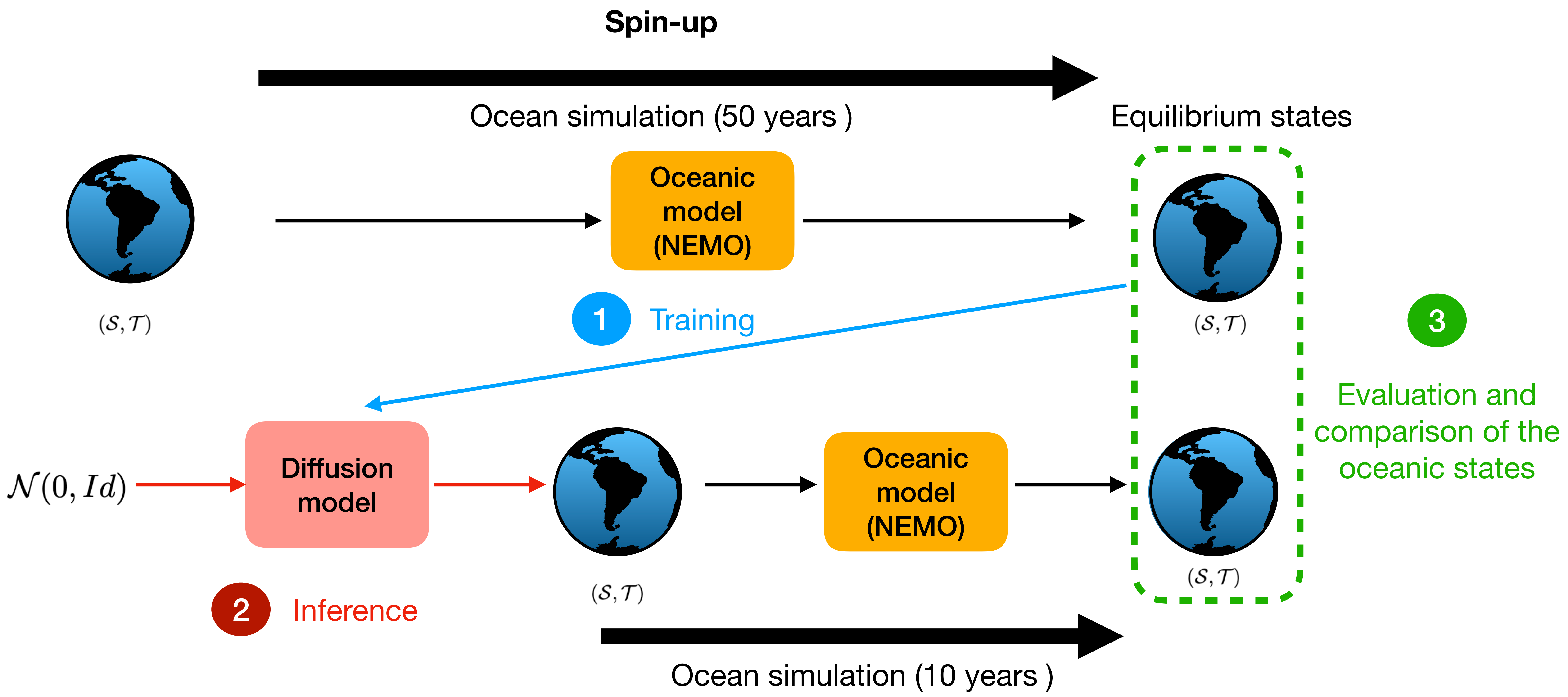


Figure: Pipeline of the training and evaluation protocol.

We learn the distribution of post spin-up ocean state simulations and sample from this learned distribution.
How can we ensure the the generated states respect the laws of physics?

How to condition a generative model?

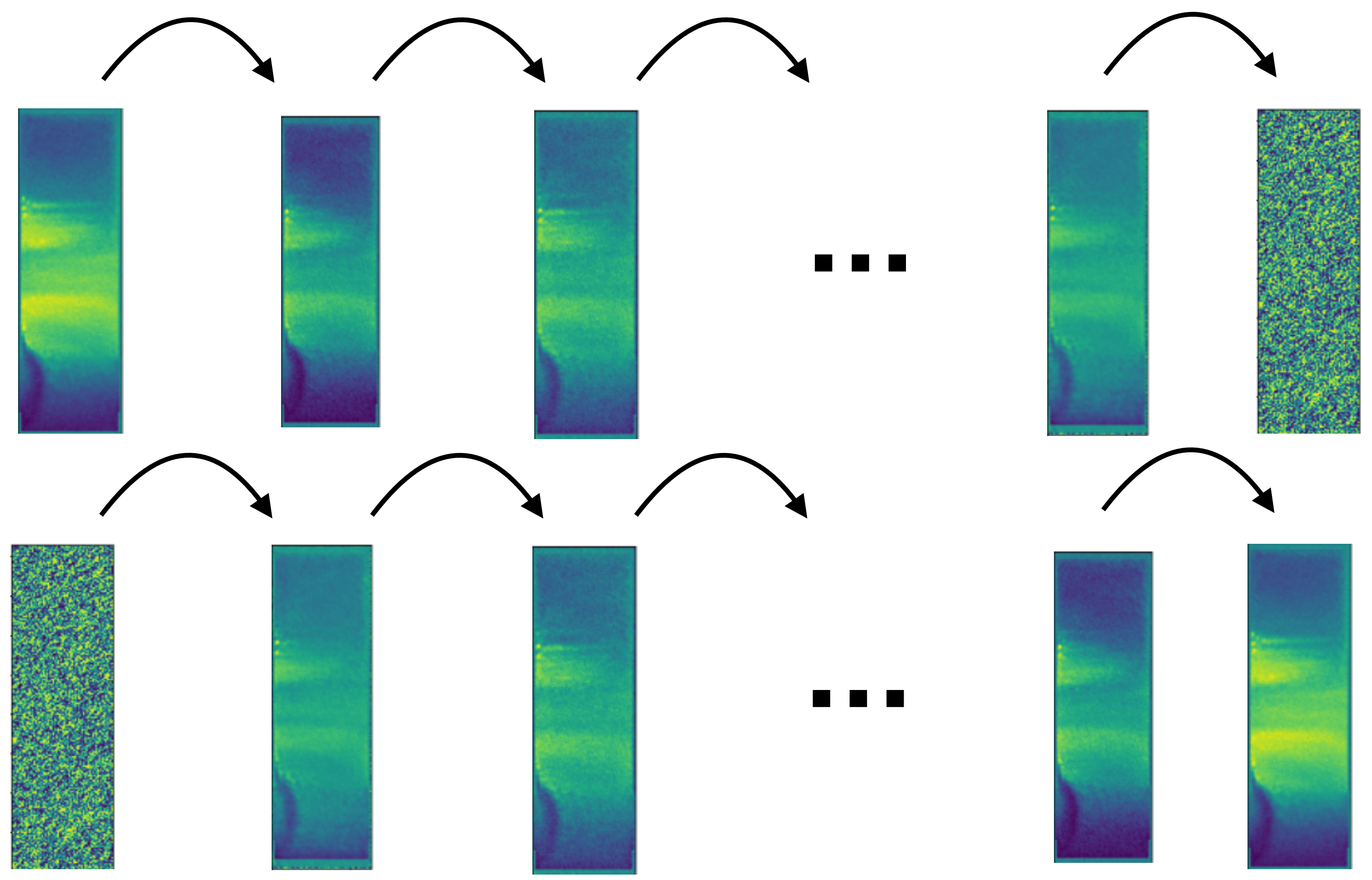
We want to enforce a constraint on vertical stratification

Training

$$dx = f(x,t)dt + g(t)dw$$

Forward

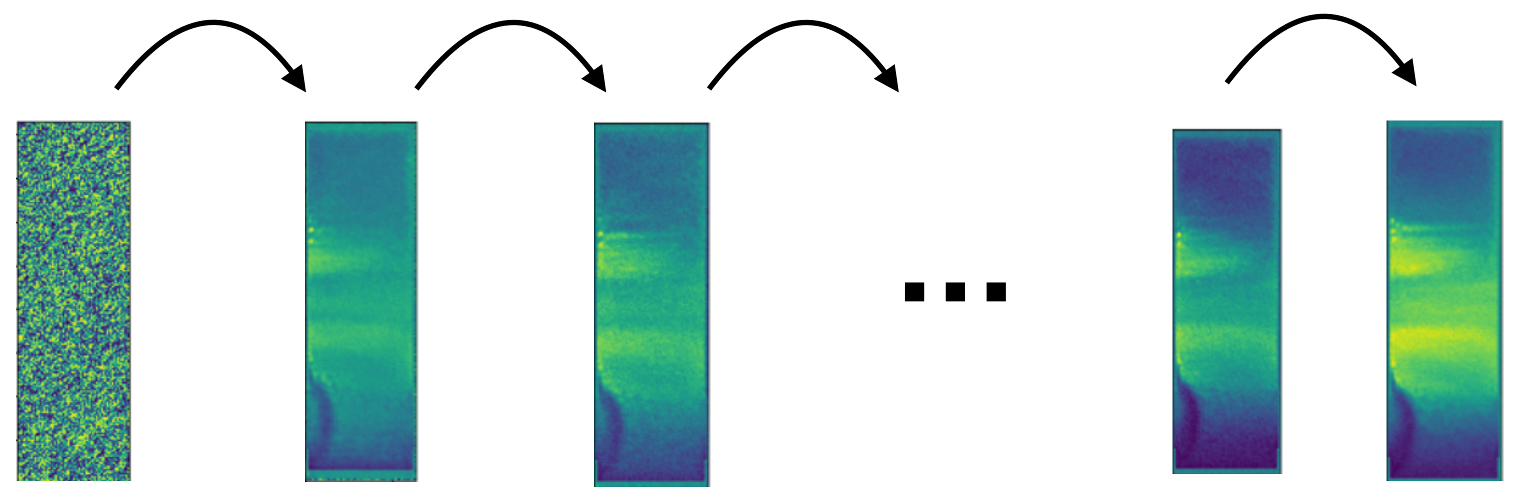
Backward



$$dx = [f(x,t) - g^2(t)\nabla_x \log p(x)]dt + g(t)dw$$

The model learns the distribution of the ocean initialization states.

Sampling



$$x_{s+1} = x_s + \tau_s \underbrace{\nabla_x \log p(x_s)}dt + \sqrt{2\tau_s}\epsilon_s \qquad \epsilon_s \sim \mathcal{N}(0,\mathbf{I})$$

$+$

$$\boxed{\beta \nabla_x C(x_s)}$$

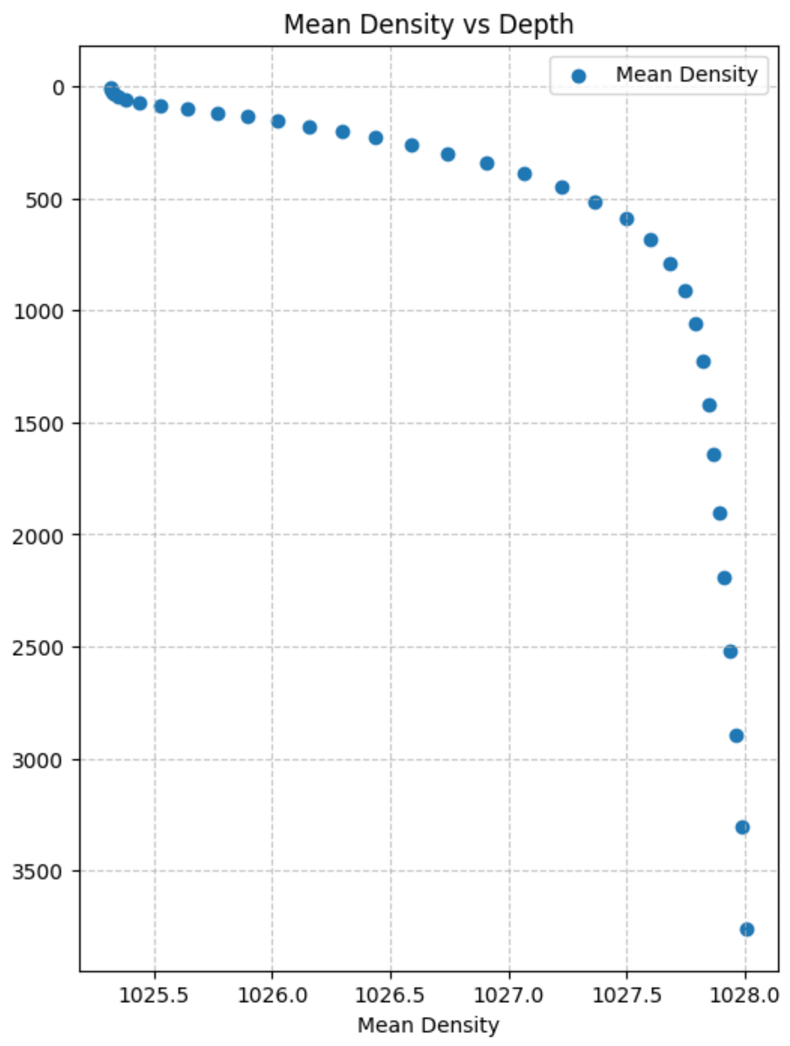
Constraint enforced during sampling process of the model through Langevin sampling algorithm.

How to define the constraint and how to sample?

Constraint selection:

Density: Interpretable by oceanographers and linked to ocean physical properties

For $x = (\mathcal{T}, \mathcal{S})$:



$$\mathcal{C}_1 = \left\{ x\,;\ C_1(x) = \sum_{k=0}^{K} \left(\mu_k - \frac{1}{N} \sum_{i,j} \rho(x_{i,j,k}) \right)^2 = 0 \right\}$$

$$\mathcal{C}_2 = \left\{ x\,;\ C_2(x) = \sum_{k=0}^{K-1} \left(\nabla_k \mu_k - \frac{1}{N} \sum_{i,j} \nabla_k \rho(x_{i,j,k}) \right)^2 = 0 \right\}$$

Density is computed with the approximation from Roquet et al. [4] used for DINO.

Sampling algorithm:

From algorithm proposed in Chamon et al. [1], we rewrite the constrained sampling problem as an optimisation problem :

Problem

$$\underset{q \in \mathcal{P}_2(\mathbb{R}^d)}{\text{minimize}}\ D_{KL}(q \| p) \quad \text{subject to } \mathbb{E}_{x\sim q}[C(x)] = 0$$

With D_{KL} divergence de Kullback-Leibler, p the learned distribution, C our constraint function

The is performed with a primal gradient descent alternated by a dual ascent.

Primal-dual resolution

For $s \in \{0,..,S\}$:

Primal update: $x_{s+1} = x_s - \tau_s \nabla_x \mathcal{L}(x_s, \lambda_s) + \sqrt{2\tau_s}\,\epsilon, \quad \epsilon \sim \mathcal{N}(0,1)$

Dual update: $\lambda_{s+1} = \lambda_s + \eta \nabla_\lambda \mathcal{L}(x_s, \lambda_s)$

Results: Generated ocean states

<u>Qualitative evaluation of the generated states</u>: We compare the generated states with a dataset extracted from DINO ocean model simulations [2].

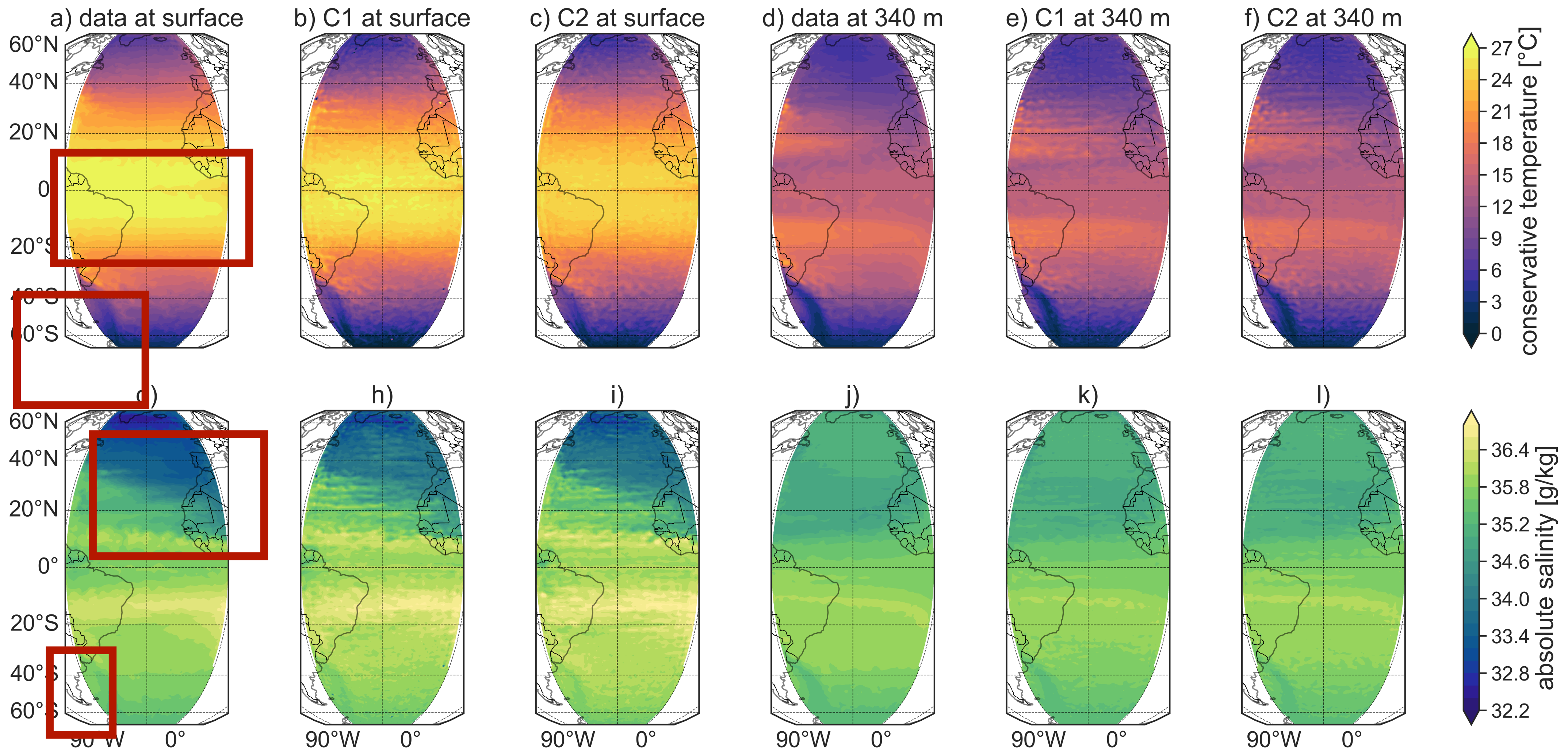


Figure: Temperature and salinity fields at surface and at 340m deep

- Presence of large scale characteristic features in the temperature and salinity fields.
- Validation of the vertical density profiles for constrained generation.
- Uplifting of Antarctic dense water around -60/-70° only present in constrained case.

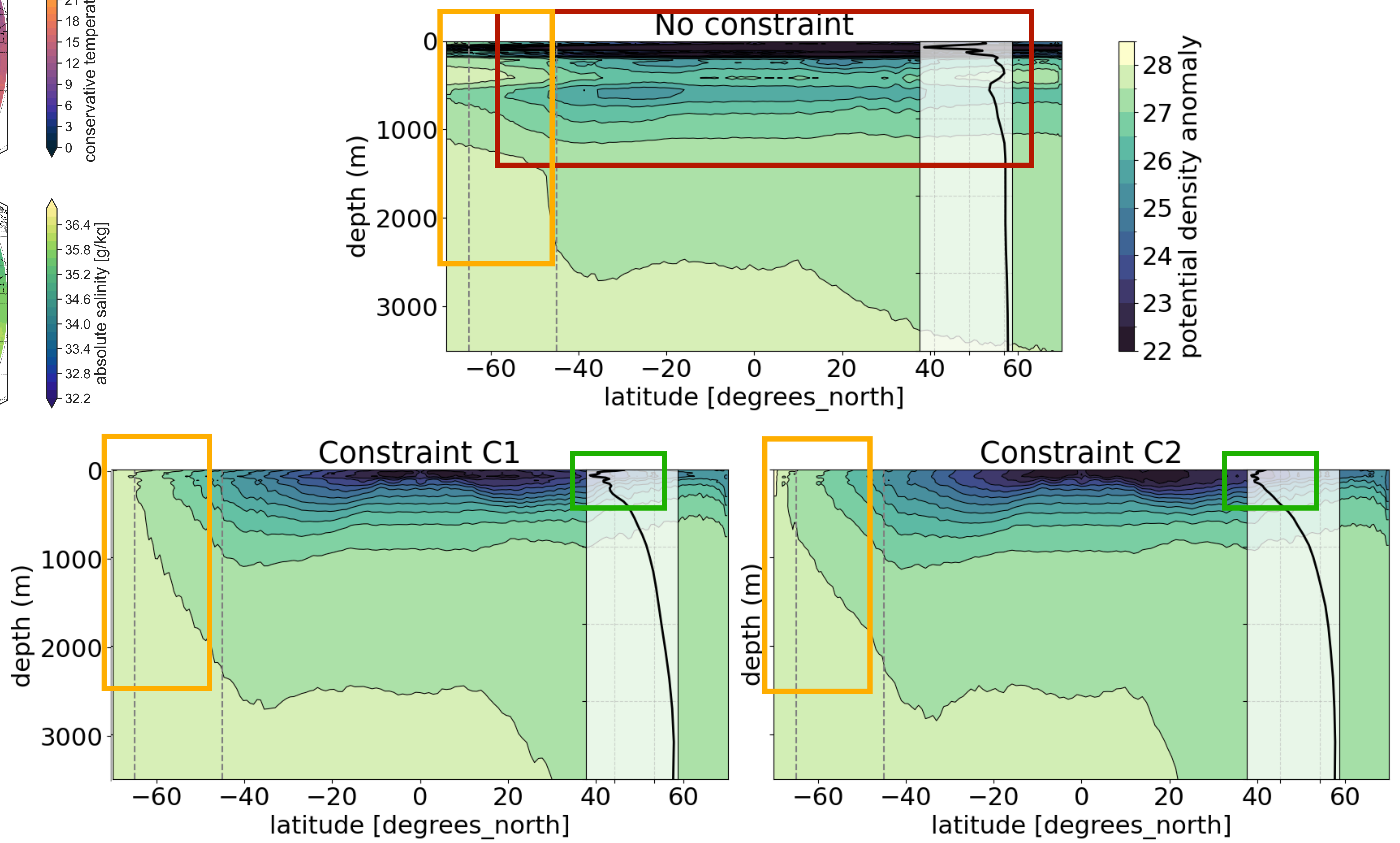


Figure: zonal average representation of the density vs depth

Results: Integration into ocean model

Results of a 10-years integration of the generated states into the oceanographic model (DINO)

Figure: zonal average representation of the density vs depth after integration

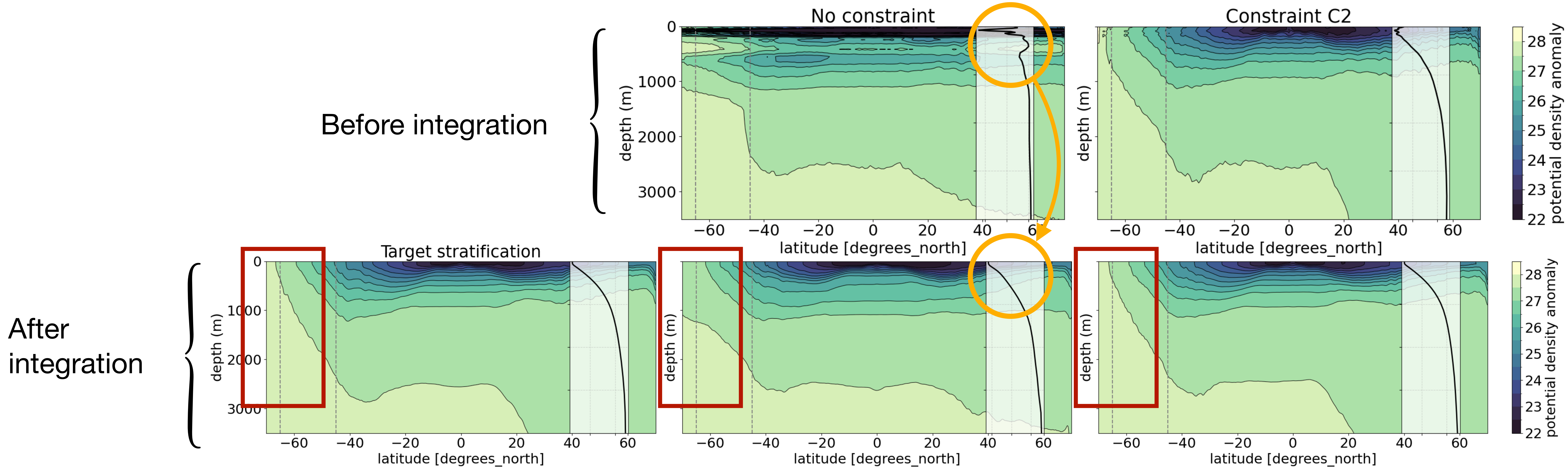


Table: mean point-by-point variability of the generated states

	No Constraint	Constraint TS	Constraint C1	Constraint C2
Salinity	0.2132	0.0095	0.0105	0.0126
Temperature	9.0859	0.3178	0.4156	0.6326

The diversity of the generated states is increased by the new constraint implemented.

Thank you !

blandine.gorce@locean.ipsl.fr