Physically consistent sampling for ocean model initialization

Blandine Gorce, Luther Ollier,
David Kamm, Etienne Meunier

1. Introduction

- OGCMs are critical for climate science but are computationally expensive, particularly during spin-up phases requiring millions of CPU hours. This high computational costs limit parameter exploration and ensemble simulations crucial for uncertainty quantification.
- Diffusion models are a class of generative AI models that learn to produce realistic data by reversing a noise-adding process, enabling fast and accurate emulation of complex physical systems. However, we show that diffusion models, even when trained on valid physical data, don't necessarily respect physical constraints when generating new states, introducing the need for physical regularization.

2. Our proposed method for the generation of ocean states…

- Diffusion models can **generate physically consistent oceanic states** that are usable as initial conditions for numerical integration.
- **Computational costs:**
  (a) Numerical integration : $\sim 10^3$ CPU-hours
  (b) Diffusion model: 48 h for training + 30s / state generation (on V100 GPU)

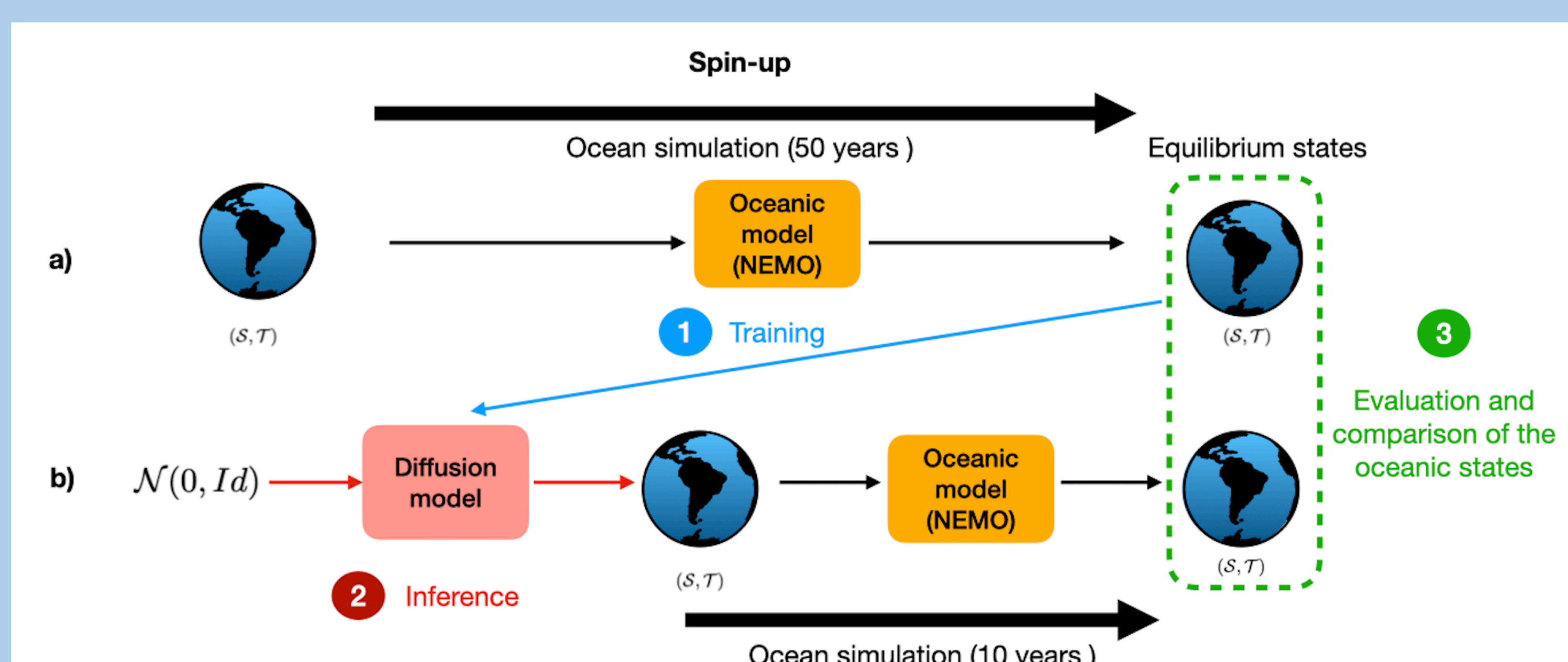

Figure: Pipeline of the training and evaluation protocol.

Constrain with improved Langevin sampling:

Given a constraint function $C(x) : \mathbb{R}^{Z\times W\times H} \rightarrow \mathbb{R}$ defined over our fields, we formulate our problem as a **minimization task**:

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

With $D_{KL}(q\|p)$ is the Kullback-Leibler divergence between q (sampling distribution subject to constraint C) and p (target distribution learned from data).

Using a primal–dual decomposition, the optimization is performed through a **gradient descent–ascent scheme** [1] at each step s of the sampling process:

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

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

with the resulting Lagrangian potential defined as $\mathcal{L}(x,\lambda) = \log p(x) + \lambda^T C(x)$.

Vertical stratification constraint:

We constrain the hydrostatic balance of our fields through two constraints on ocean density $\rho(S,T)$ to fit the expected mean ocean density profile μ:

$$C_1(x) = ||\mu - \frac{1}{N}\sum_{i,j}\rho(x_{i,j})||_2^2 \quad ; \quad C_2(x) = ||\nabla_k \mu - \frac{1}{N}\sum_{i,j}\nabla_k \rho(x_{i,j})||_2^2$$

3. Results

Dataset: Generated from the DINO configuration. [2] It uses a mercator grid with 1/4 horizontal resolution and 36 vertical levels. The domain spans 60° longitude and 70° latitude from equator to both poles. For this study, we generated a dataset by running DINO for 50 years, saving 1800 snapshots of temperature and salinity fields. The resulting dataset consists of 1800 states, each containing three fields (SSH, T,S).

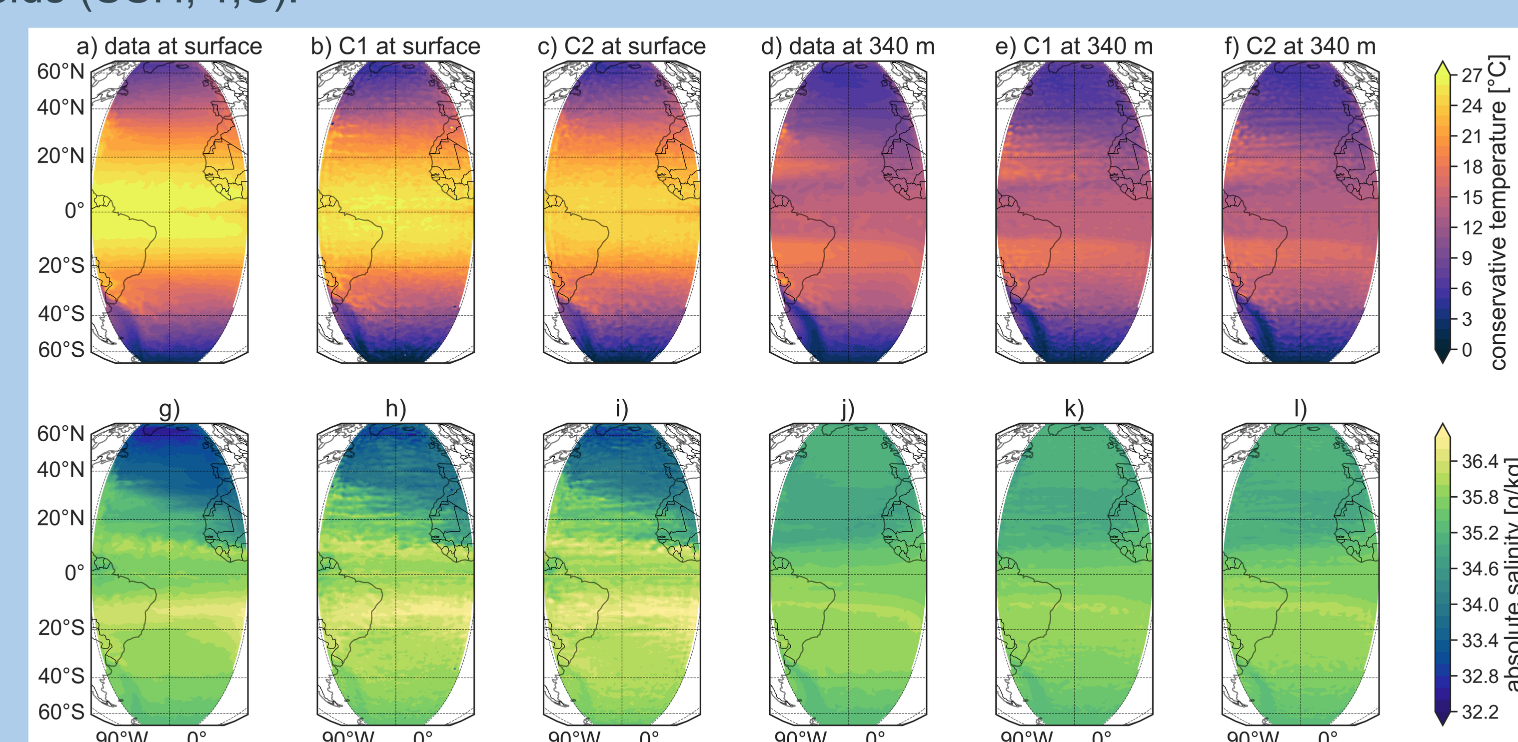

Figure: Temperature and salinity fields at surface and at 340m deep

Quality of the generated states:

- The generated temperature and salinity fields successfully capture large-scale patterns, including warm, saline waters in the tropics and cooler, fresher waters at higher latitudes, while maintaining coherent vertical relationships between different depth levels.

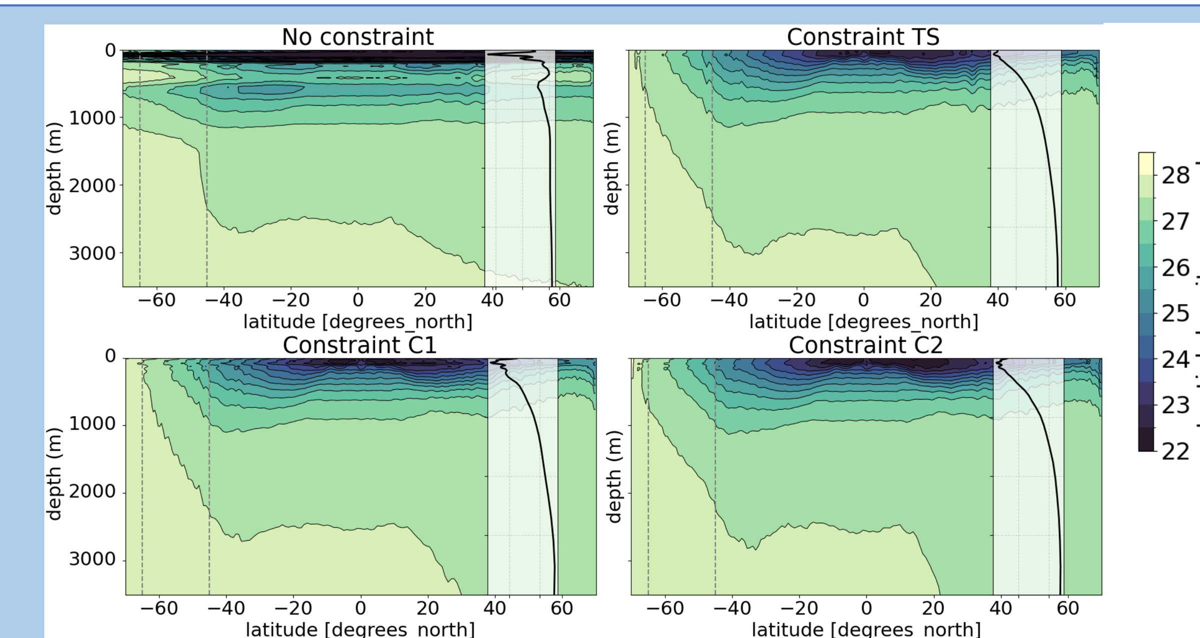

Figure: zonal average representation of the density vs depth

- The density profiles above further validate that both constraints allow the model to captures the overall stratification structure of the global ocean in particular near the surface unlike the unconstrained case.
- The proposed constraint allows for more variability in the generated states than previous constraint implemented in Meunier et al. [3]

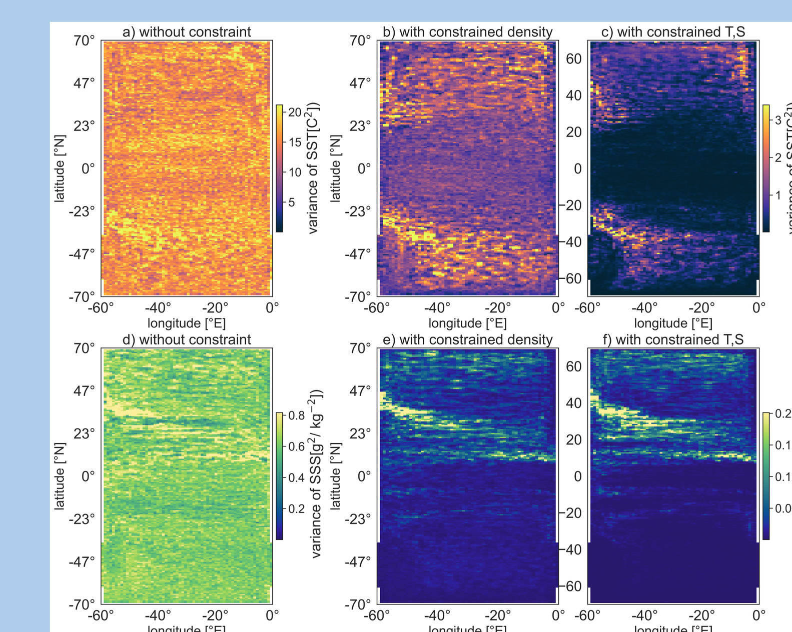

Figure: Mean point-by-point variability for T and S fields

Physical consistency of the generated fields:

- The density profiles below are computed through 10 years of NEMO integration. The results show minimal changes over the period of integration both on the zonal average and on the mean density profile at low latitudes.
- The generation without constraint still underestimates the uplifting of Antarctic dense water around -60/-70°, a key aspect in the formation of this water mass, whereas the constrained generation is able to resolve this physical property.

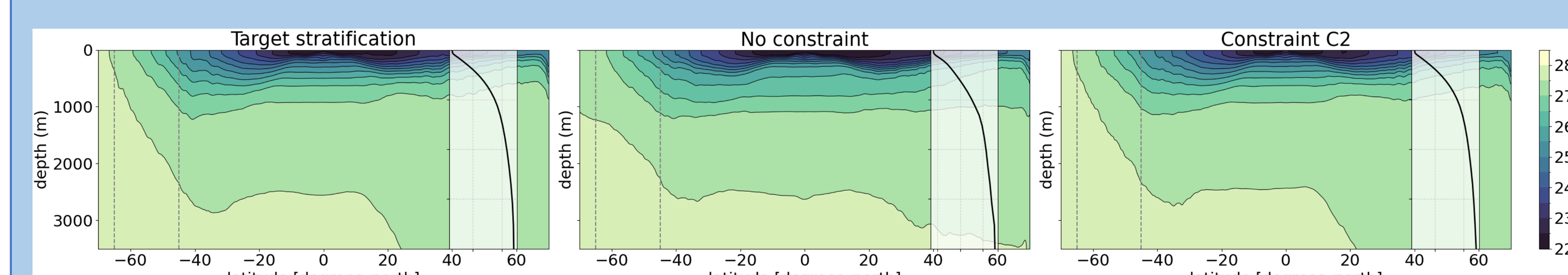

Figure: zonal average representation of the density vs depth after integration

References

[1]. Luiz F.O. Chamon et al. (2013). "Constrained sampling with primal-dual langevin monte carlo." https://arxiv.org/abs/2411.00568.

[2]. David Kamm, DINO. 2025. https://github.com/vopikamm/DINO.

[3]. Etienne Meunier et al. 2025. https://arxiv.org/abs/2502.02499.