GloFM: a GLORYS Flow-Matching emulator for spatio-temporal ocean data assimilation

Pierre Garcia, Théo Archambault, Dominique Béréziat, Anastase Charantonis

GloFM: a GLORYS Flow-Matching emulator for spatio-temporal ocean data assimilation

Pierre GARCIA[1,2], Théo ARCHAMBAULT[2], Dominique BÉRÉZIAT[1] and Anastase CHARANTONIS[3]
[1]*LIP6, Sorbonne Université, (CNRS), 4 place Jussieu 75005, Paris, France,* [2]*Amphitrite, 149 Avenue du Maine, 75014 Paris, France,* [3]*Arches, Inria, 48 rue Barrault 75013, Paris, France*

Keywords: Generative Models, Data assimilation, Oceanography, Observing System Experiment, Flow matching, Denoising Diffusion Probabilistic Model

Abstract: Providing regular and physically consistent predictions of the ocean state is critical for numerous scientific, operational, and societal needs. Observations of the ocean surface are gathered through various remote sensing and *in situ* instruments, and are typically assimilated into numerical models to reconstruct the ocean state. However, this often involves millions of data points, making it computationally intensive, which suggests deep learning may be a cheaper alternative. Deterministic data-driven approaches typically learn about ocean dynamics from numerical simulations or sparse observational data. However, such methods often lack physical realism in uncertain settings. Due to mode averaging, they produce non-physical or overly simplified states. Generative models offer a promising approach to generating physically realistic ocean states. We present GloFM: a Glorys Flow-Matching emulator for spatio-temporal ocean data assimilation. Our generative model produces coherent estimates of ocean surface fields. GloFM uses flow matching to assimilate observational data for nowcasting of surface currents, sea surface height (SSH), and sea surface temperature (SST). Compared to deterministic regression-based approaches, GloFM demonstrates improved realism metrics, capturing finer-scale variability and more physically plausible ocean states.

1 Introduction

The state of the ocean surface both reflects and influences climate change. Due to its massive thermal capacity, the ocean absorbs and redistributes heat across the Earth, making it one of the first climate regulators. Beyond its climatic role, it is central to the world's economy, ecology, and geopolitics, supporting activities ranging from energy production to fisheries. Due to its significance, understanding ocean dynamics has long been an involved scientific challenge, motivating decades of research and sustained observation of the ocean surface.

A central challenge lies in estimating ocean surface currents, with numerous applications, including navigation, meteorology, and the propagation of pollutants. Surface velocity is observed directly through sparse in situ sensors, known as drifters, as they are advected by ocean circulation. Diverse surface variables, indirectly linked to currents, are also observed by satellite remote sensing, such as the Sea Surface Height (SSH) and the Sea Surface Temperature (SST). An accurate estimation of the ocean state often involves combining information from these diverse observations, using physics-based priors.

1.1 Data Assimilation

The realism of numerical simulations depends on the quality of the functional representation of physical laws, as well as the boundary conditions. Data assimilation is the process by which observations and physical models are combined, allowing for an accurate and dynamically consistent representation of the ocean state.

To produce realistic fields, simulation models such as NEMO (Madec et al., 2017) employ a 3D approach to represent the ocean state. However, these 3D numerical simulations are limited by their grid resolution and some unknown physical quantities. Furthermore, assimilation using these methods requires nudging, Kalman filtering, or variational assimilation, for instance, which significantly increases the product computational budget. The GLObal physical ReanalYsiS (GLORYS) (CMEMS, 2020) is an assimilation of *in situ* and remote-sensing observations using NEMO as a backbone for the simulation, and an Ensemble Kalman filter for assimilation. GLORYS has many hyperparameters related to both assimilation and the NEMO simulation itself. These parameters are challenging to tune, especially since

NEMO and assimilation runs are computationally expensive and temporally sequential, running on central processing units (CPUs). Thanks to numerous sea-surface observations, multiple studies investigated sea surface reconstruction using expert-designed priors.

For instance, Optimal Interpolation (OI) is an early data assimilation method that can be used to interpolate sparse observations. DUACS (CMEMS) is a Level 4 (full 2D field) map of SSH, produced using OI of sparse along-track satellite observations, based on expert-tuned covariance matrices. It is linear and univariate, providing highly interpretable outputs, but it misrepresents non-linear interactions and does not exploit observations of other oceanic variables.

1.2 Deterministic deep learning methods

Machine learning is emerging as a promising tool in ocean modeling by offering flexible, data-driven alternatives or complements to traditional assimilation methods. Unlike classical approaches, which rely heavily on physical models, machine learning models learn complex nonlinear relationships directly from data. For ocean applications, machine learning models can be used to emulate the assimilation process. Neural networks, for instance, are trained to forecast the ocean state given incoming satellite or *in situ* observations, dramatically reducing computational costs compared to variational methods.

Models from (Archambault et al., 2024a,b; Ciani et al., 2024; Garcia et al., 2025; Lenain et al., 2025; Martin et al., 2024) are examples of deterministic regression-based models. These models are typically trained to estimate the mean of the conditional output distribution given a set of inputs. While this approach yields a reliable estimation of the expectation, it often produces unrealistic or non-physical fields in regions where uncertainty is high, a phenomenon known as mode averaging. This occurs because the model effectively estimates an average of multiple possible outcomes, often smoothing the prediction – which is caused by the convolutional architecture prior – collapsing several possible physical scenarios. This makes the physical interpretation of the estimated mean difficult.

1.3 Generative modeling

On the other hand, Generative Models (GMs) are trained to minimize the Kullback-Leibler Divergence (DKL) between the modeled distribution and the dataset distribution. DKL is the reference measure of similarity between two distributions for GM. As such, once trained, they effectively **learned a prior** allowing one to **sample** a distribution of **realistic states**. GMs are natively unconditional; they model a distribution by sampling a random variable from a learned distribution, using noise. Numerous methods exist to enforce the conditionality of Deep Generative Model inference, some modifying the training scheme, others modifying the sampling algorithm.

We investigated Denoising-Diffusion-Probabilistic-Model (DDPM) models because they are flexible and the best performing GMs for high dimensions (Dhariwal and Nichol, 2021). Diffusion models were the first class of iterative sampling models, expressing the DKL minimization objective as a Markovian regression objective (Ho et al., 2020). In 2023, (Lipman et al., 2023) defined a new class of deep generative models that uses the Continuous Normalizing Flow formalism (Chen et al., 2018), generalizing DDPM models.

1.4 Generative modeling in Oceanography

Modeling observation-conditioned distribution using a DDPM is explored for oceanography in (Martin, 2025). In this paper, the authors train an unconditional DDPM on GLORYS snapshot patches. Then, the authors demonstrate how this model samples conditionally on observations using a score-based data assimilation algorithm. They successfully assimilate sea surface wind fields, SSH, SST, and Sea Surface Salinity (SSS), and show promising results highlighting DDPM's capability to sample realistic states in an Observing System Experiment (OSE) when compared to GLORYS assimilation.

Following (Archambault et al., 2024a; Martin, 2025), we use **GLORYS** (CMEMS, 2020), a global ocean reanalysis product, mapped onto a 1/12° resolution longitude-latitude grid with daily resolution for each variable. Building upon this work, we assimilate observations from seven consecutive days, whereas (Martin, 2025) used daily snapshots, and we train and sample from a Flow Matching (FM) model instead of a DDPM. Additionally, we use another score-based data assimilation algorithm for conditioning of sampled state, Moment Matching Posterior Sampling (MMPS) (Rozet et al., 2024). Intuitively, at each sampling step, MMPS steers the sampled state towards a likely state, minimizing the reconstruction error of the state's simulated observation with respect to the ground-truth observation, while enforcing that the sampled state still belongs to the prior distribution.

Thus, given some partial observations, using a learned prior distribution of sea-states, stored in a FM

model's weights, allows us to sample from the full distribution of plausible realizations.

Training generative models typically requires access to complete ground-truth fields, often only available from numerical simulations. However, the iterative structure of DDPM makes it particularly suitable for incorporating observation-based constraints when generating physical realizations of the ocean state learned from the numerical simulation.

1.5 Contributions

- We train an unconditional FM model on GLORYS video patches, with 7 days by 128 by 128 $1/12°$ pixels.
- We condition and evaluate our FM model's generated ensemble using multivariate real-world observations on several days, using the MMPS algorithm, allowing for Bayesian estimation of uncertainty and mean.
- We compare our assimilated ensemble realism with GLORYS and ensemble prediction skill, and uncertainty estimates using withheld remote sensing observations against various hindcast 1-member regression classical and DL products.

2 Data

2.1 Prior training dataset

We train our FM model to unconditionally sample 4 coherent variables from GLORYS spatio-temporal patches: SSH, SST, components U and V of the sea surface velocity, using the native resolution of GLORYS gridding, 1/12°, with a snapshot at noon every day. We learn GLORYS in the North Atlantic region defined by the 25°N to 45°N by 40°W to 70°W square. We use data from January 1, 1995, to December 15, 2019, withholding 2017 data (from December 15, 2016, to January 15, 2018) for validation.

2.2 Observational data

Satellite and *in situ* observations are used either for assimilation or for model evaluation, and are never seen during training. The observational datasets include:

SST L3: Level-3 image SST data (NOAA/STAR, 2023), providing high-resolution infrared SST observation fields. These data provide high-resolution conditioning during assimilation and can be assimilated for current or SSH reconstruction as sea currents advect the temperature. Cloud occultation of infrared waves sparsens these remote-sensing observations.

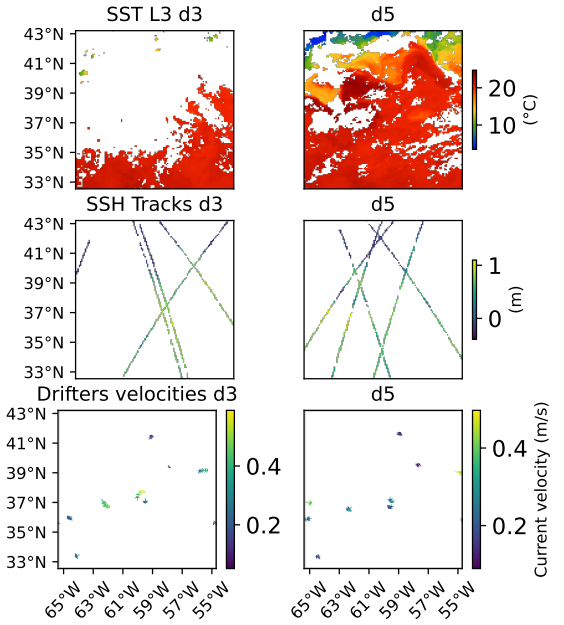


Figure 1: Observational data used for assimilation. Here, two timesteps are displayed in the two columns: the third (d3) and fifth (d5) day of the assimilation, being January 3 and 5, 2017, respectively. We see data sparsity takes a different form for each variable and timestep.

L3 SSH from nadir satellites: Level-3 SSH measurements from nadir-pointing altimetry (CMEMS, 2021). These point-based observations are assimilated into reference models. Nadir-pointing altimeters observe the SSH directly beneath them; their measurements are corrected to only observe mesoscale phenomena. NADIR measurements are also used for validation.

Drifters: Surface current data derived from Lagrangian drifters (SEANOE, 2024) are often used to evaluate the accuracy of inferred surface velocity fields. Drifters are the gold standard of current measurement, tracking their advection by the sea using the Global Positioning System.

We display two spatio-temporal patches of assimilated observations in Figure 1 and provide a data summary in Table 1.

3 Proposed Methods

3.1 Flow matching models

Flow matching models aim to sample from any distribution by mapping a known distribution to it. Given

Table 1: Observational data summary

Data	Reference	Use	Resolution	Description
SST L3	CMEMS (2023)	Assimilation	1/40° 1 day	Cloud occulted supercolated product
SSH tracks	CMEMS (2021)	Assimilation Validation	7km, 1Hz	Ellipse averaged along track measurement
Drifter buoys	SEANOE (2024)	Assimilation Validation	1h	GPS derived mesoscale velocities
GLORYS SSH	CMEMS (2020)	Training	1 day 1/12°	Numerical simulation height
GLORYS SST	CMEMS (2020)	Training	1 day 1/12°	Numerical simulation temperature
GLORYS currents	CMEMS (2020)	Training	1 day 1/12°	Numerical simulation currents

$p_0(x)$ and $p_1(x)$, (Lipman et al., 2023) defines $p_s = [\phi_\tau]_* p_0$, where p_1 is the distribution to sample, p_0 the distribution used for sample initialization, and $\phi : [0,1] \times \mathbb{R}^d \to \mathbb{R}^d$ the flow verifying $\frac{d}{ds}\phi_s(x) = v_s(\phi_s(x))$. The flow ϕ_s and the optimal velocity v_s are intractable, but the authors demonstrate that optimizing the Conditional Flow Matching Loss $\mathcal{L}_{\text{CFM}}(\theta)$

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{s,q(x_1),p_s(x|x_1)} \left(u_s(x|x_1) - v_{s,\theta}(x)\right)^2 \quad (1)$$

is equivalent in expectation to optimizing the Flow matching objective $\mathcal{L}_{\text{FM}}(\theta)$.

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{s,p_s(x)} \left\| v_{s,\theta}(x) - v_s(x) \right\|^2 \quad (2)$$

This provides a first-order estimate of the flow point-wise velocity used in conjunction with an ODE solver for sampling. They later define various $p_s(x|x_1)$ transport paths. The Optimal-Transport (OT) path is the most efficient.

3.2 Flow matching training

Let $x_1 \in \mathbb{R}^{\text{C}\times\text{T}\times\text{H}\times\text{W}}$ be the targeted GLORYS variables; SST, SSH, and components U and V of the currents, at T timesteps. We train a UNet (Ronneberger et al., 2015) to minimize (1), with $x_1 \in \mathbb{R}^{\text{C}\times\text{T}\times\text{H}\times\text{W}}$ being composed of a timeseries of spatial patches with several channels. We set $\text{C} = 4, \text{T} = 7, \text{H} = 128, \text{W} = 128$.

For training, we follow the OT path formula for the model's target and input with the $\mathcal{L}_{\text{CFM}}$ loss (1). The OT path inputs and target distributions are defined by $p_s(x \mid x_1)$ and $u(x \mid x_1)$, respectively.

$$p_s(x \mid x_1) = sx_1 + (1-s)x_0 \quad (3)$$

$$u_{\text{target}}(x \mid x_1) = x_1 - x_0 \quad (4)$$

We found sampling $z \sim \mathcal{N}(0,1)$ and then computing $s = \frac{1}{1+e^z}$ gives a qualitatively superior result when comparing with other s sampling; we define this algorithmic timestep sampling distribution as ψ. This allows us to use our trained prior for unconditional sampling of GLORYS-like sea states, as well as assimilation. Thus our network weights θ are optimized to minimize:

$$\mathbb{E} \left\| v_{s,\theta}\left(sx_1 + (1-s)x_0\right) - (x_1 - x_0) \right\|^2 \quad (5)$$

$$\text{with } s \sim \psi, x_1 \sim q_{\text{GLORYS}}, x_0 \sim \mathcal{N}(0, I) \quad (6)$$

.

3.3 Sampling with a flow matching model

In the following, we detail how this trained network can sample from the learned GLORYS distribution. First, we explain unconditional sampling algorithms; then how we condition the iterative sampling to assimilate observations.

This optimization allows us to sample the dynamics of the sea on a defined spatio-temporal window. OT-path flow matching models are trained to predict the velocity of the OT ODE Equation 8, which gives x_1 when integrated from x_0. Formally, we integrate the ODE defined by:

$$x_0 \sim \mathcal{N}(0, I) \quad (7)$$

$$\frac{dx_s}{ds} = v_\theta(x_s, s), \text{with } s \in [0;1] \quad (8)$$

We train a 6.1M-parameter UNet for 200 epochs, which represents 163k steps with a batch size of 8. We use Adam optimizer with a learning rate of 10^{-5}, linear decreasing learning rate schedule with warmup. The training takes 1.6 days on a single V100 GPU.

3.3.1 Posterior sampling

We implement Rozet et al. (2024)'s Moment-Matching-Posterior-Sampling (MMPS) algorithm. We selected this posterior sampling algorithm for its demonstrated stability. Posterior sampling for diffusion models often relies on iterative score steering of the estimated state x_s using a gradient-based optimization of expected realization at each step, computed against the observations. They define the observation and final sample distribution as normal distributions parametrized by the learnt prior, the current sample

state x_s, the observation y, with estimated covariances. They then maximize the log-likelihood of the resulting distribution by gradient descent on x_s. Implementing MMPS for flow-matching, at each matching step, we add a second-order Tweedie approximation of the posterior score on x_s, by gradient descent through the network:

$$\nabla_{x_s} \log p(y \mid x_s) \simeq \nabla_{x_s} \log \mathcal{N}(y|\hat{y}, \Sigma_y + \mathbb{V}[\hat{y}]) \quad (9)$$

$$\simeq \left(\nabla_{x_s} \hat{y}^\top\right) \left(\Sigma_y + A\mathbb{V}[x \mid x_s]A^\top\right)^{-1} \delta_{y,x_s} \quad (10)$$

with:

- x_s the state at algorithmic timestep s,
- $x \in \mathbb{R}^{\mathrm{C}\times\mathrm{T}\times\mathrm{H}\times\mathrm{W}}$ the final state of the sampling process,
- $y \in \mathbb{R}^k$ the observation,
- A the linear observation operator, it can be expressed as $A \in \mathbb{R}^{(\mathrm{C}\times\mathrm{T}\times\mathrm{H}\times\mathrm{W})\times k}$ matrix
- $\mathbb{V}[x \mid x_s]$ the covariance matrix of our sampling prior for the realization at timestep s,
- $\hat{y}$ the observation estimated by the FM predicted final step multiplied by the observation matrix: $A\mathbb{E}[x_1 \mid x_s]$
- with $\mathbb{E}[x \mid x_s] = x_s + (1-s)v_\theta(x_s)$ the expectancy of the realization at algorithmic timestep s,
- $\delta_{y,x_s} = (y - \hat{y})$ the expected error, and
- Σ_y the instrumental uncertainty covariance matrix.

This allows us to maximize the log-likelihood of the conditional distribution $p_\theta(x|y)$, effectively adjusting x_s to partial observations.

Formally, we integrate the ODE defined by:

$$x_0 \sim \mathcal{N}(0, I), \text{ with } s \in [0;1] \quad (11)$$

$$\frac{dx_s}{ds} = v_\theta(x_s, s) + \nabla_{x_s} \log p(y \mid x_s) \quad (12)$$

Remark that we use the posterior score while still integrating an ODE. This allows us to use the DOPRI5 (Dormand and Prince, 1980) solver for ODE posterior sampling. We experimented with integrating an SDE using: the FM model's velocity-derived prior score, noise injection, and Equation (10) for the posterior score. SDE samples produced worse results (biased and underspread ensemble), probably because integrating an SDE prevents the use of ODE solvers.

4 Experimental Results

Additional results are available in our GitHub repository[1], namely the realism evaluation of assimilation members, Power Spectral Density analysis, a comparison between two forecasting approaches, and the evaluation of reconstructed current velocity against drifter measurements.

4.1 Evaluation Metrics

We assess the realism and predictive skill of our model using a combination of probabilistic, spectral, and deterministic metrics. We evaluate our reconstructions in the Gulf Stream area, in the square defined by 55°W to 65°W by 33°N to 43°N.

4.1.1 Physical consistency evaluation

We assess the statistical realism of model output in both assimilation and forecast settings. We compute the empirical Probability Density Functions (PDFs) and Power spectral densities (PSD) of selected channels (e.g., SSH, SST, U, V, and current magnitude) and compare them to reference distributions from GLORYS.

4.1.2 Deterministic evaluation

We compute RMSE errors with respect to observed data (SSH tracks, drifters) for various assimilation or forecast configurations.

We also report metrics on the 2017 Ocean Data Challenge dataset (Ballarotta and Le Guillou, 2021).

4.1.3 Ensemble reconstruction evaluation

We evaluate the ability of our ensemble to capture uncertainty and variability:

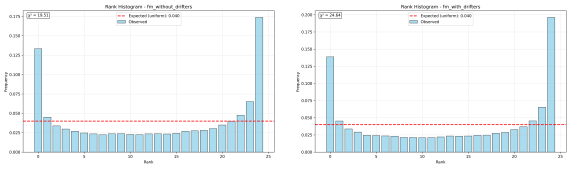
- **Rank Histogram**: given a sorted ensemble reconstruction of a scalar variable, we compute the rank of the target value, then report the frequency of that rank. This metric allows for visualization of the calibration of the generated ensemble. A flat rank-histogram suggests that marginally, the ensemble is estimating the right univariate distribution, but not necessarily that the model is estimating spatially coherent fields.
- **Cumulative Ranked Probability Score (CRPS)**: $E_i[|X_i - y|] - \frac{1}{2}E_{i,j}[|X_i - X_j|]$, given an ensemble of reconstruction $X_{0\ldots,n}$ and a non-assimilated observation y, CRPS estimates the squared univariate CDF L_2 distance between observation and ensemble reconstruction. As such, it provides an estimation of the mean of each member univariate reconstruction absolute bias to observation, balanced by the whole ensemble uncertainty. It is a proper scoring rule.

[1] https://github.com/PierreGarciaRafael/GloFM

Table 2: Ocean data challenge SSH mapping RMSE and effective resolution λ_x.

Model	RMSE (cm) ↓	λ_x (km) ↓
FM without drifters	8.74 ± 3.06	126
FM with drifters	8.61 ± 2.91	127
DUACS	7.15 ± 2.32	149
MIOST	6.42 ± 2.06	137
4DVarNet	6.21 ± 1.67	109
ConvLSTM SSH-SST	5.58 ± 1.41	99
ABED-SSH-SST	5.56 ± 1.42	108

4.2 SSH Bayesian interpolation



(a) Rank histograms of SSH without drifter assimilation. (b) Rank histograms of SSH with drifter assimilation.

Figure 2: Rank histogram shows that the ensemble reconstructed SSH underestimates the spread. Even more when assimilating drifter observations. Our method is slightly biased towards low values, as more values fall in the last bin.

We assimilate SSH and SST satellite observations of the year 2017 in the Gulf Stream (from 54.67° to 65.33°W and 32.67° to 43.33°N). For validation purposes, we withhold the SSH observations of one satellite, Cryosat-2, and compute regression scores on its observations. We define SSH and SST observation covariance matrices as diagonal matrices, with observational uncertainties set to 0.03 m and 0.4 °C for SSH and SST, respectively.

Optionally, we assimilate drifter observations, setting their uncertainty to 0.05 $\mathrm{m\,s^{-1}}$. Assimilated observations for the third and fifth days of the first week are displayed in Figure 1. We set all observation matrices as selection matrices.

In Table 2, we report the results of the mean of the ensemble assimilation, with drifters being assimilated for the first line, but not in the second. We benchmark against the state-of-the-art of SSH interpolation: DUACS (Taburet et al., 2019) and MIOST (Ubelmann et al., 2022), two variants of the optimal interpolation method, and 4DVarNet (Fablet et al., 2024), ABED-SSH-SST (Archambault et al., 2024b), ConvLSTM SSH-SST (Martin et al., 2024), three deep learning models. For RMSE, our model is the least performing one, although it has a finer effective resolution than DUACS and MIOST. As the model wasn't explicitly trained to predict the mean of the distribution, it produces higher RMSE. Although these seemingly conflicting results on the RMSE and effective resolution are counterintuitive. Drifter assimilation provides a slight improvement in RMSE with our method, but a decrease in effective resolution.

We also compare our ensemble members' with GLORYS' Power Spectral Density (PSD), and Probability Density Function (PDF). The PSD similarities between the two assimilation setups demonstrate the flexibility of the posterior sampling algorithm; however, our model produces excessively high frequencies for SSH, U, and V as members of the assimilation remain noisy at $s = 1$.

We display assimilation results in Figures 3a and 3b, without and with drifters assimilation, respectively. Note that Figure 1's first column matches the ensemble results of Figures 3a and 3b, resulting in the ensemble standard deviation of the associated variables, at the observation location, being close to 0, while higher in the unobserved region. For the visualization of ensemble members and animations of the whole-year assimilated ensemble means and standard deviations, we refer the reader to our GitHub repository.

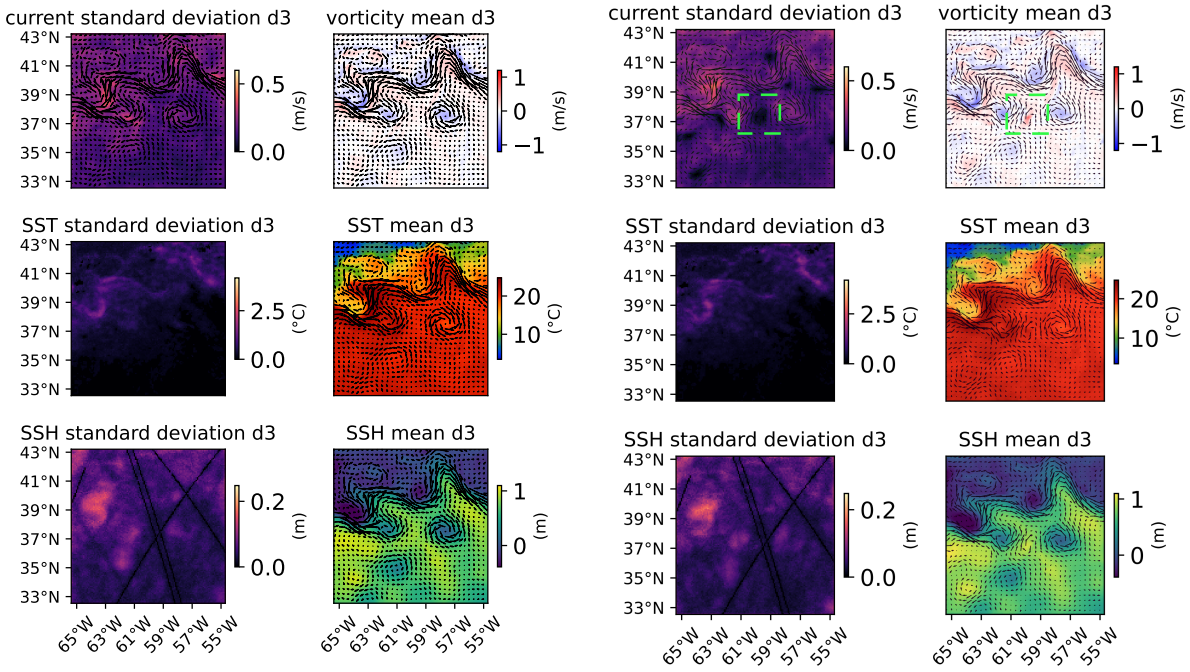
Our assimilation with drifters produces a CRPS of 3.2 ± 2.6cm, compared to CryoSat-2 observations, without drifter assimilation, the CRPS evaluates to 3.3 ± 2.7 cm, $\pm$ denotes the standard deviation of the CRPS along day examples. These CRPS scores are coherent with the assimilated drifters ensemble decrease in RMSE (Table 2). Figure 2 shows that our ensemble reconstruction is underspread compared to CryoSat-2 observations. Nadir altimeter measurements are naturally noisy; this may explain the underspread reconstruction ensemble. We can make the hypothesis that the learnt GLORYS prior underestimates variability.

For the ensemble where drifters are not assimilated, we evaluate our model's synthesized current on drifters. See rank histograms displayed in our Github repository. The rank histogram reveals an underspread ensemble for velocity predictions. CRPS reveals no clear trend among days of assimilation; this trend varies between U and V components. The number of drifters limits the analysis.

5 Discussion

5.1 Limitations

Although retraining the model is avoided, posterior sampling is expensive even with FM models renowned for unconditional sampling efficiency. In this paper, the design choices of the observation operator and uncertainty favor simplicity due to the high



(a) Generated ensemble mean and standard deviation. With SSH and SST observation assimilated.

(b) Generated ensemble mean and standard deviation. With SSH, SST, and current observations assimilated.

Figure 3: Results of the 3rd day (out of 7) of assimilation for the SSH Bayesian interpolation, January 3rd 2017. Notice SSH is significantly sharper in the mean of the drifter assimilating ensemble b). Notably, the RS only assimilating ensemble a) doesn't produce the small eddy at $\sim 60°$W 38°N, whereas it is present for b). This qualitative difference is also visible in the vorticity fields.

computational costs of the overall approach. We assume that this hindered the model's reconstruction realism. This has a further cascading effect when applied to forecasts, which will accumulate the blurriness of previous steps. Finally, the current approach is computationally prohibitive for operational use.

5.2 Perspective and Synthesis

The implementation of the FM training and posterior sampling loops presented in this work is simple. The efficiency of FM unconditional sampling can be increased using rectified flow matching regularization; maybe it could hold in the case of posterior sampling, reducing the costs of the conditional ensemble generation. Taking into account the blurring caused by the altimeter's measurement by modifying observation matrices should also improve realism metrics. To address the over-estimation of the high scales of the FM samples with respect to GLORYS, we could train and validate on an ensemble of experts for the flow matching ODE integration, like in (Brenowitz et al., 2025). While the results at present are worse than the state of the art, both in metrics and computational cost, we demonstrated that the FM model's flexibility enables the assimilation of a wide range of observations. While the model was trained only on numerical simulation, sampling from a conditional distribution on multimodal partial observation is possible thanks to MMPS. This is a very valuable attribute in a field where multiple sources of varying quality exist, whose information is usually difficult to incorporate.

The effective resolution of the ensemble mean points at the model's capacity to resolve smaller scales. For operational interpolation, using a sampling model in combination with a regression model should provide better results, using the best of both worlds, the SOTA RMSE from ABED (Archambault et al., 2024a), and the Bayesian uncertainty prediction of an FM model trained on its residuals to ground truths. To enhance the assimilation and sampling efficiency, one could hope to reduce the computational budget of the FM model using latent matching (Dao et al., 2023). Forecasting using a flow-matching model would probably benefit from Masked Conditional Video Diffusion, from (Voleti et al., 2022).

REFERENCES

T. Archambault, A. Filoche, A. Charantonis, and D. Béréziat. Pre-training and Fine-tuning Attention Based Encoder Decoder Improves Sea Surface Height Multi-variate Inpainting. In *VISAPP*, February 2024a.

T. Archambault, A. Filoche, A. Charantonis, D. Béréziat, and S. Thiria. Learning Sea Surface Height Interpolation from Multi-variate Simulated Satellite Observations. *JAMES*, 16(6), 2024b.

M. Ballarotta and F. Le Guillou. ocean-data-challenges/2021a_SSH_mapping_OSE: Material for SSH mapping OSE data challenge, September 2021. URL https://doi.org/10.5281/zenodo.5511905.

N. Brenowitz, T. Ge, A. Subramaniam, A. Gupta, D. Hall, M. Mardani, A. Vahdat, K. Kashinath, and M. Pritchard. Climate in a Bottle: Towards a Generative Foundation Model for the Kilometer-Scale Global Atmosphere, May 2025. arXiv:2505.06474 [physics].

R. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations. In *NeuRIPS*, volume 31, 2018.

D. Ciani, C. Fanelli, and B. B. Nardelli. Estimating ocean currents from the joint reconstruction of absolute dynamic topography and sea surface temperature through deep learning algorithms. *EGUsphere*, pages 1–25, April 2024.

CMEMS. Global Ocean Gridded L 4 Sea Surface Heights And Derived Variables Reprocessed 1993 Ongoing. URL https://doi.org/10.48670/moi-00148.

CMEMS. Global ocean physics reanalysis [dataset], 2020. https://doi.org/10.48670/moi-00021.

CMEMS. Global ocean along-track L3 sea surface heights reprocessed (1993-ongoing) tailored for data assimilation [dataset], 2021. https://doi.org/10.48670/MOI-00146.

CMEMS. Global oceans sea surface temperature multi-sensor L3 observations [dataset], 2023. https://doi.org/10.48670/MOI-00164.

Q. Dao, H. Phung, B. Nguyen, and A. Tran. Flow Matching in Latent Space, July 2023. arXiv:2307.08698 [cs].

P. Dhariwal and A. Nichol. Diffusion Models Beat GANs on Image Synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pages 8780–8794. Curran Associates, Inc., 2021.

J. R. Dormand and P. J. Prince. A family of embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics*, 6(1):19–26, March 1980.

R. Fablet, B. Chapron, J. Le Sommer, and F. Sévellec. Inversion of Sea Surface Currents From Satellite-Derived SST-SSH Synergies With 4DVarNets. *JAMES*, 16(6), 2024.

P. Garcia, I. Larroche, A. Pesnec, H. Bull, T. Archambault, E. Moschos, A. Stegner, A. Charantonis, and D. Béréziat. ORCAst: Operational High-Resolution Current Forecasts. *Artificial Intelligence for the Earth Systems*, 4(4), 2025.

J. Ho, A. Jain, and P. Abbeel. Denoising Diffusion Probabilistic Models. In *NIPS*, volume 33, pages 6840–6851, December 2020.

L. Lenain, K. Srinivasan, R. Barkan, and N. Pizzo. An unprecedented view of ocean currents from geostationary satellites, August 2025.

Y. Lipman, R. Chen, H. Ben-Hamu, M. Nickel, and M. Le. Flow Matching for Generative Modeling, February 2023.

G. Madec, R. Bourdallé-Badie, P.-A. Bouttier, C. Bricaud, D. Bruciaferri, D. Calvert, JSSH. Chanut, E. Clementi, A. Coward, D. Delrosso, et al. NEMO ocean engine. Scientific Notes of Climate Modelling Center, 2017.

S. Martin. Generative Data Assimilation for Surface Ocean State Estimation from Multi-Modal Satellite Observations, 2025. URL https://eartharxiv.org/repository/view/8688/.

S. Martin, G. Manucharyan, and P. Klein. Deep Learning Improves Global Satellite Observations of Ocean Eddy Dynamics. *Geophysical Research Letters*, 51(17), 2024.

NOAA/STAR. GHRSST NOAA/STAR ACSPO v2.81 0.02 degree L3S Dataset from Daily LEO Satellites, 2023. URL https://doi.org/10.5067/GHLDY-3S281.

O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *MICCAI*, pages 234–241, 2015.

F. Rozet, G. Andry, F. Lanusse, and G. Louppe. Learning Diffusion Priors from Observations by Expectation Maximization. In *NeuRIPS*, volume 37, pages 87647–87682, December 2024.

SEANOE. Copernicus marine in situ - global ocean-delayed mode in situ observations of surface (drifters, HFR) and sub-surface (vessel-mounted ADCPs) water velocity, 2024. https://doi.org/10.17882/86236.

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, September 2019.

C. Ubelmann, L. Carrere, C. Durand, G. Dibarboure, Y. Faugère, M. Ballarotta, F. Briol, and F. Lyard. Simultaneous estimation of ocean mesoscale and coherent internal tide sea surface height signatures from the global altimetry record. *Ocean Science*, 18(2):469–481, April 2022.

V. Voleti, A. Jolicoeur-Martineau, and C. Pal. MCVD: Masked Conditional Video Diffusion for Prediction, Generation, and Interpolation, 2022.