Sea surface salinity downscaling using deep generative diffusion models

Enzo Forestier[1], Luther Ollier[2], Roy El Hourany[2], Jacqueline Boutin[1], Carlos Mejia[1], and Sylvie Thiria[1]

[1]Laboratoire d'Océanographie et du Climat : Expérimentations et Approches Numériques (LOCEAN), UMR 7159, Sorbonne Université / CNRS / IRD / Muséum national d'Histoire naturelle (MNHN), 4 place Jussieu, 75005 Paris, France

[2]Laboratoire d'Océanologie et de Géosciences (LOG), UMR 8187, Université du Littoral Côte d'Opale (ULCO) / Université de Lille / CNRS, Wimereux, France

Correspondence: Enzo Forestier (enzo.forestier@locean.ipsl.fr)

Abstract. High-resolution satellite observations are essential for studying fine-scale ocean processes. We investigate diffusion models, a class of deep generative models, for improving the resolution of Yet, present satellite sea surface salinity (SSS) from coarse inputs and for reconstruction under noisy and incomplete observations. We train an unconditional prior products remain too coarse to resolve many fine-scale structures. We investigate denoising diffusion models as a generative framework for SSS downscaling in a controlled proof-of-concept experiment based on GLORYS reanalysis fields. A multichannel diffusion prior is trained on $1/12°$ reanalysis fieldsand condition the model SSS, sea surface temperature (SST), and sea surface height (SSH) fields, and is then conditioned at inference time on coarse SSS a synthetically degraded coarse SSS observation ($1/3°$) together with high-resolution ($1/12°$) sea surface temperature (SST) andsea surface height (SSH) as auxiliary variablesauxiliary SST and/or SSH. Conditioning is performed via a through pseudo-inverse guidanceapproach, which steers sampling toward solutions that are both statistically consistent with the learned prior and the generated samples toward states that are compatible with the observationscoarse observation while remaining within the learned GLORYS-consistent multivariate distribution. We also introduce a simple test a gradient-enhancement procedure applied during inference designed to increase contrast while maintaining consistency with the conditioning constraintsduring inference. Experiments in the Gulf Stream region compare models conditioned on SST only, on SSH only, and on both variables. Validation over the year 2020 uses root-mean-square error (RMSE), structural similarity (SSIM), gradient distributions, and temporal Fourier spectra. Conditioning In the present GLORYS configuration, conditioning on SST substantially improves accuracy relative to SSH alone; combining SST and SSH yields further gainsand slightly outperforms a convolutional baseline, comparable to a strong convolutional baseline under RMSE/SSIM, while additionally providing an ensemble of plausible reconstructions. The gradient-enhanced sampler restores sharper fronts and increased weekly-daily variance at a small cost in increases structural contrast but can risk amplifying part of the variability, illustrating a trade-off between pixel-wise scoresaccuracy and structural realism. Overall, the results show that guided diffusion models can downscale SSS while recovering fine-scale structure, with SST providing the dominant small-scale constraint and SSH adding complementary mesoscale context. The framework is designed to extend naturally to support guided diffusion as a promising framework for SSS downscaling and uncertainty-aware reconstruction, while showing that transfer to real satellite SSS products and future higher-resolution missionswill require product-aware observation operators, uncertainty weighting, and independent in-situ validation.

1 Introduction

Sea surface salinity (SSS) is a fundamental variable for understanding the ocean-climate system. Together with temperature, it controls seawater density and thereby influences thermohaline circulation and the redistribution of heat and carbon. Variations in surface salinity also modulate density stratification of the upper ocean and the solubility of carbon dioxide (CO_2), linking ocean properties and biogeochemical fluxes (Vinogradova et al., 2019; Reul et al., 2020). Accurate high-resolution SSS observations are therefore essential for constraining air-sea exchanges and characterizing freshwater pathways.

Satellite L-band radiometry missions such as ESA's Soil Moisture and Ocean Salinity (SMOS; (Kerr et al., 2010)), NASA's Aquarius (Lagerloef et al., 2008) and Soil Moisture Active Passive (SMAP; (Piepmeier et al., 2017)) provide quasi-global SSS coverage with revisit times of a few days, complementing sparse in situ networks such as Argo. However, their effective spatial resolution (typically 40-100 km) remains too coarse to resolve fine-scale processes such as river plumes, sharp frontal gradients, and mesoscale /submesoscale features frontal structures that drive upper-ocean variability (Reverdin et al., 2024). Retrieval uncertainties related to surface roughnessand sea stateintroduce biases , sea state, land contamination, rain effects, radio-frequency interference, and auxiliary geophysical corrections introduce errors that are difficult to correct in dynamic coastal and frontal regions (Boutin et al., 2014)(Boutin et al., 2014, 2021), while temporal averaging and filtering further smooth the fields.

To mitigate these limitations, multi-mission and data-fusion products have been developed, combining satellite and model information via optimal interpolation or variational approaches (Olmedo et al., 2016). While these products improve spatio-temporal continuity, they often suppress gradients and small-scale variability. As a consequence, key processes such as eddy dynamics and freshwater advection remain under-resolved, especially in energetic regions such as the Gulf Stream or the Amazon plume (Reul et al., 2020; Reverdin et al., 2024).

Machine learning methods have recently emerged as effective tools to enhance the resolution and quality of oceanographic fields by exploiting nonlinear relationships between variables. Convolutional neural networks (CNNs) have been used to downscale satellite images and reanalysis products, including approaches such as RESAC (Thiria et al., 2023) and related work (Fablet et al., 2021). These approaches are typically deterministic and often produce overly smooth reconstructions, partly because mean-square training objectives favour averaged solutions. They also generally require retraining for each task (e.g. super-resolution, denoising, gap-filling). In this study, RESAC is used as a state-of-the-art baseline.

Diffusion models provide a complementary generative perspective. Originally developed for image generation (Ho et al., 2020; Rombach et al., 2022), diffusion models learn a distribution over high-resolution fields and can generate multiple plausible realizations conditional on observations. This probabilistic formulation naturally represents uncertainty and can better preserve physically plausible variability. Diffusion models are increasingly used in Earth-system science applications such as downscaling, reconstruction, and forecasting (Mardani et al., 2023; Watt and Mansfield, 2024; Tomasi et al., 2025; Finn et al., 2024), highlighting their ability to integrate auxiliary variables and constraints.

Here, we explore denoising diffusion implicit models (DDIM) for downscaling SSS in a controlled setting based on a high-resolution the GLORYS global ocean reanalysis. High-resolution reanalysis fields provide dynamically consistent GLORYS provides complete, collocated gridded SSS, SST, and SSH and fields, which allow us to synthetically degrade salinity to satellite-like resolution a coarse observation while retaining access to a well-defined reference field high-resolution reference for evaluation. This avoids, at this stage, complications due to sensor-specific biases, radio-frequency interference, and coastal contaminationTherefore, GLORYS is used here as a controlled gridded reference, not as an error-free truth, and allows us to focus on methodological aspects of downscalingthe downscaling problem.

Within this framework, we condition the generative process on coarse SSS together with high-resolution auxiliary variables. SST can be coupled to SSS at meso- and submesoscale through mesoscale and fine scales represented in GLORYS via mixed-layer dynamics and lateral advection, while SSH encodes mesoscale circulation and front/eddy organization. By conditioning on SST and SSH, we aim to reconstruct the fine-scale SSS patterns frontal and filamentary SSS variability present in the $1/12°$ GLORYS reference from degraded observations and assess when to assess what diffusion-based downscaling provides an advantage over deterministic CNNs. Ultimately, the goal is to recover submesoscale features that are ubiquitous in energetic regions such as the Gulf Stream but remain unresolved by current satellite SSS products, thereby bridging the gap between adds relative to deterministic CNN baselines. The objective is therefore not to generate unresolved fine-scale dynamics absent from the training data, but to recover the structures represented by the high-resolution numerical models and remote sensing observationsreference while maintaining consistency with the coarse SSS constraint.

2 Materials

2.1 Study area, data and variables of interest

The study area is located in the North Atlantic Ocean, in the Gulf Stream region (see Fig. 1), bounded to the north and south by 33°N and 43°N, and to the west and east by 63°W and 52°W, respectively. This sector is characterized by one of the most energetic western boundary currents of the global ocean, with strong jets, meanders and eddies that generate intense temperature and salinity fronts over a wide range of spatial and temporal scales (Chassignet and Xu, 2017; Wu et al., 2022). Such conditions make it an ideal natural laboratory for testing downscaling methods that aim to reconstruct fine-scale structures from coarser information.

Within this domain, SSS, SST and SSH exhibit distinct patterns of spatial variability. As illustrated in Figure 2, SSS and SST share sharp frontal features and filamentary structures associated with mesoscale and submesoscale frontal activity. In the present GLORYS-based setup, SSH appears smoother than expected at the fine scale. The close resemblance between SSS and SST gradients suggests that SST is a particularly informative auxiliary variable for constraining the small-scale structure of SSS, whereas SSH primarily encodes the broader mesoscale circulation that organizes the positions and strengths of fronts and eddies.

This leads to the working hypothesis that underpins our modeling strategy: at the mesoscale and submesoscaleat the fine scales represented by GLORYS, SSS variability is tightly coupled to SST through processes such as mixed-layer dynamics and lateral advection, whereas SSH provides a complementary large-scale dynamical constraint. We therefore expect high-resolution SST to provide strong local guidance for reconstructing SSS gradients, with SSH acting as an additional constraint on the overall dynamical context. The diffusion model is designed to exploit these relationships by conditioning the generative process on both variables.

2.2 Data preparation and processing

Our models were trained on GLORYS12V1 CMEMS global ocean reanalysis daily data available at $1/12°$ horizontal resolution on 50 vertical levels and produced in the framework of the Copernicus Marine Environment Monitoring Service (Lellouche et al., 2021). In this study, we only use the surface layer. The dataset used in this study consists of daily gridded fields of SSS, SST, and SSH over the Gulf Stream region for the period 1993-2020.

To emulate satellite-based salinity observations, we explicitly distinguish between a high-resolution "truth" reference field and a coarse-resolution "observation " observation field. The native GLORYS SSS at $1/12°$ is taken as the reference high-resolution SSS, while a corresponding coarse SSS field at $1/3°$ is obtained by applying a 3×3 4×4 spatial averaging and subsampling operator. This correction makes the degradation factor consistent with the ratio between $1/12°$ and $1/3°$. The resulting grid spacing of about 35-40 km is intermediate between the effective footprint of present L-band radiometer missions (typically 40-100 km) and the nominal resolution sampling of commonly used Level-4 multi-mission products(25-50 km), and thus provides a reasonable useful idealized proxy for current satellite SSS products (Vinogradova et al., 2019; Reul et al., 2020). Only SSS is degraded in this way; SST and SSH are kept at their native $1/12°$ resolution and are used as idealized high-resolution auxiliary fields in downscaling experiments, as they can be observed at a finer resolution of 10 and 5 km from satellite operating missionsthe controlled experiments. The choice of $1/3°$ is not intrinsic to the method: in principle, the same workflow could be applied to any coarser another target resolution by adapting the degradation operator.

This idealized treatment of SSS, SST, and SSH should be kept in mind when considering satellite applications. The present SSS/SST/SSH conditioning experiments should therefore be interpreted as controlled tests rather than as direct operational satellite configurations.

Before training, all variables were linearly normalized to zero mean and unit variance over the training period. This normalization places SSS, SST and SSH on comparable numerical scales and stabilizes optimization. During training, square patches of 64×64 pixels (approximately $600\ \text{km} \times 600\ \text{km}$) were randomly extracted from the domain and provided to the model (see Fig. 1). Working with patches rather than full-domain images considerably increases the data set and reduces computational costreduces the memory required per training sample and increases the number of training examples, which is critical given the high dimensionality of diffusion models. At the same time, the energetic nature of the Gulf Stream ensures that strong contrasts and eddy structures are frequently present, even within such smaller spatial windows, given that our goal is to study mesoscale reconstruction. Therefore it is . These patches therefore provide a rich set of examples for learning the multivariate relationships between SSS, SST and SSH that are exploited in the generative downscaling framework. In doing so, our methodology aims to obtain an SSS patch of 64×64 pixels at $1/12°$ resolution, independently of any geolocation information within the region of interest.

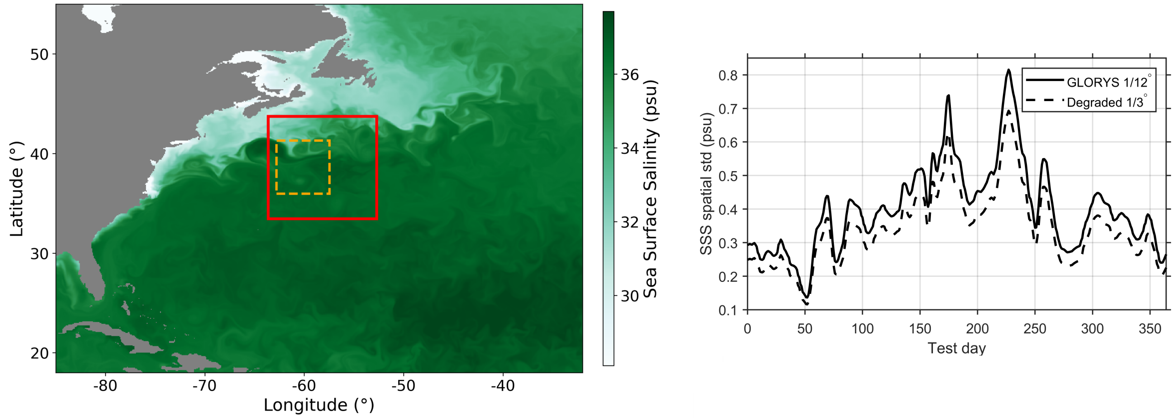

Figure 1. Left: SSS $1/12°$ map from GLORYS reanalysis data, in the North Atlantic on 1 January 2021. Our study focuses on the Gulf Stream, in the region highlighted in red. A random 64×64 pixels yellow square inside the red domain indicates the size of patches used during learning phase. Right: daily standard deviation of SSS in the study area for the year 2020.

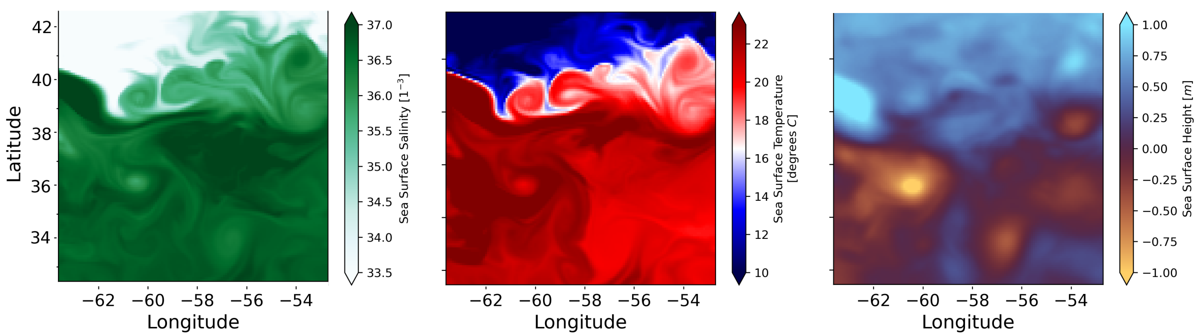

Figure 2. Left to right: Images of SSS, SST, and SSH maps for June 2020 provided by GLORYS for the studied area. SSS and SST present several correlated fine-scale structures, while SSH is much smoother.

3 Method

3.1 The downscaling problem

In this study, downscaling refers to the reconstruction of a high-resolution SSS field from a coarser SSS observation and high-resolution auxiliary variables. Let We follow the notation commonly used in inverse problems and data assimilation: X denote the unknown denotes the unknown high-resolution state, Y denotes the available observation, and H is the observation operator mapping the state space to the observation space. In the present case, X is the SSS field on the $1/12°$ grid (the "true" field) and Y is the corresponding coarse SSS field on a observation on the $1/3°$ grid. The two are linked by an observation operator

H,

$$Y = HX + \eta, \tag{1}$$

where H represents spatial averaging and sub-sampling from $1/12°$ to $1/3°$, and η is an observation noise term. In our experiments based on GLORYS, Y is synthetically generated from X by applying H, with or without added noise and masking to mimic satellite-like conditions.

In addition to the coarse SSS Y, we assume that high-resolution auxiliary variables $Z = (\mathrm{SST}, \mathrm{SSH})$ are available on the same $1/12°$ grid as X. They can be used alone or both in order to help estimating Ytogether to help estimate the high-resolution state X. These fields share many of the fronts and filaments present in SSS and provide information on the underlying dynamics. From a learning perspective, the downscaling problem consists in estimating a mapping:

$$\hat{X} = f_\theta(Y, Z), \tag{2}$$

parametrized by the model weights θ, such that the prediction $\hat{X}$ is a high-resolution SSS field on the $1/12°$ grid, consistent with Y through the observation operator H and statistically similar to the reference high-resolution SSS fields. Thus, X and $\hat{X}$ are always defined on the high-resolution grid, whereas Y is defined on the coarse observation grid. In the next subsection, we describe the diffusion-based generative model used to implement this mapping.

The backbone of our diffusion model is a U-Net architecture, widely used in image generation and reconstruction tasks (Ronneberger et al., 2015; Ho et al., 2020; Nichol and Dhariwal, 2021; Rombach et al., 2022). It follows the usual encoder-decoder structure: a sequence of convolutional and attention blocks gradually reduces the spatial resolution while increasing the number of feature channels, then a symmetric decoding path progressively upsamples the representation back to the original $1/12°$ grid. In practice, we use a multi-channel U-Net that takes as input the stacked SSS, SST and SSH fields (when available) together with a sinusoidal time embedding of the diffusion step, as in standard DDIM implementations (Ho et al., 2020; Nichol and Dhariwal, 2021). Self-attention layers are included at the coarsest resolutions to capture long-range spatial dependencies, while convolutional residual blocks handle local interactions. Skip connections link encoder and decoder blocks at matching scales, allowing the network to combine local fine-scale information with broader contextual features and to better preserve sharp fronts and filaments (details about the architecture in annex). This design provides a good balance between expressiveness and computational cost for high-resolution oceanographic images.

We present hereinafter the two phases of the classical diffusion model and we propose a new algorithm, the Gradient-Enhancing Gradient-Enhancing (GE) algorithm that allows to increase the contrast in the generated SSS $1/12°$ samples.

3.2 Learning phase: Forward Diffusion model

Diffusion models are a class of probabilistic generative models that learn the probability distribution of a set of training samples in order to generate new, statistically consistent realizations (Ho et al., 2020). In our case, a training sample x_0 corresponds to a high-resolution image on the $1/12°$ grid (e.g. including SSS, and in some configurations also SST and SSH channels). Training consists in two phases: a "forward" noising process where gaussian noise is added to the target sample, and a "backward" process where the network learns to gradually remove the noise to retrieve the original sample.

In the forward diffusion process, Gaussian noise is therefore progressively added to an image x_0 over a fixed number T of steps (typically $T = 1000$). At each step t, a coefficient $\alpha_t \in (0,1)$ controls the variance of the added noise. After sufficiently many steps, the original information is effectively lost and the sample becomes close to pure isotropic Gaussian noise. The forward process can be written in closed form as

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \tag{3}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ is the cumulative product of the noise schedule. This formulation allows x_t to be sampled directly from x_0 without explicitly simulating all intermediate steps.

The model then learns to reverse this process during the denoising phase, recovering x_0 step by step by predicting (and removing) the noise ϵ added at each step. In the ϵ-prediction formulation (Ho et al., 2020), the reverse dynamics can be written as

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\, \epsilon_\theta(x_t, t) \right) + \sigma_t \epsilon, \tag{4}$$

where $\epsilon \sim \mathcal{N}(0, I)$ is fresh Gaussian noise injected to maintain stochasticity in the reverse process, σ_t controls its variance, and $\epsilon_\theta(x_t, t)$ is the neural network prediction of the noise at step t.

The network parameters θ are trained by minimizing the mean-square error between the true noise and its prediction:

$$\mathcal{L}_\theta = \mathbb{E}_{x_0, \epsilon, t} \left[\left\| \epsilon - \epsilon_\theta(x_t, t) \right\|^2 \right], \tag{5}$$

where t is typically sampled uniformly from $\{1, \ldots, T\}$ and x_t is obtained from x_0 using the forward process above. At this stage, the training is unconditional with respect to the coarse observation: the model learns a prior distribution over high-resolution multichannel images drawn from the GLORYSdatasetGLORYS, without yet involving the coarse SSS observations Y or the observation operator H. This prior represents the joint GLORYS distribution of SSS with SST and/or SSH, including their spatial organization and cross-variable relationships. Time is not explicitly modeled in the present architecture but relies on the joint information between SSS and auxiliary variables at the daily situation scale. Extending the framework to a spatio-temporal diffusion model may allow temporal coherence to be learned more directly. Conditioning on (Y, Z) to perform downscaling is introduced at inference time and described in the next subsection.

3.3 Inference phase: Conditioning

Once training is complete, the model is able to generate high resolution samples matching the distribution of the correlated SST, SSH, SSS fieldshigh-resolution samples that are consistent with the learned joint distribution of SSS, SST and/or SSH as represented in GLORYS. The conditioning on a given observation can now be performed during the inferencephaseis then performed during inference. Conceptually, the learned diffusion prior plays a role similar to a statistical background, while the coarse SSS observation provides the constraint imposed through the observation operator.

Several conditioning strategies have been proposed in the literature (Song et al., 2023; Chung et al., 2024); here we follow the pseudo-inverse guidance method of (Song et al., 2023), which offers a good compromise between performance, simplicity and computational cost and can be applied to a wide range of inverse problems beyond super-resolution (see Section 5).

We rely on the observation model given in Eq. (1), where X denotes the high-resolution SSS field on the $1/12°$ grid, Y the coarse SSS field on the $1/3°$ grid, and H the observation operator (spatial averaging and subsampling). We denote by $H^\dagger$ a pseudo-inverse of H that lifts residuals from observation space back to state space (in practice, an upsampling operator from $1/3°$ to $1/12°$).

During inference, let x_t be the current latent state in diffusion step t, and $\epsilon_\theta(x_t, t)$ the predicted noise. As in the standard ϵ-prediction formulation, we first form a one-step estimate of the clean sample,

$$\hat{X}_t = \frac{x_t - \sqrt{1-\bar{\alpha}_t}\,\epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}. \tag{6}$$

We then compute the mismatch between this estimate and the observation in the coarse space,

$$r_t = Y - H\hat{X}_t, \tag{7}$$

and project this residual back onto the high-resolution grid using the pseudo-inverse $H^\dagger$. This defines a guidance term

$$g_t \propto H^\dagger r_t = H^\dagger\big(Y - H\hat{X}_t\big), \tag{8}$$

which nudges the reverse diffusion towards the subset of the learned distribution that is compatible with the observations. In practice, g_t is added to the standard reverse update for x_{t-1} (see Algorithm A1 in the annex for a complete description of the guided inference procedure).

A key advantage of this strategy is that the conditioning is entirely decoupled from the training: the diffusion model is first trained as an unconditional generator on high-resolution GLORYS fields, and different inverse problems (e.g. super-resolution, noisy or masked observations) are handled at inference time simply by changing the operator H and the corresponding pseudo-inverse $H^\dagger$. Because the reverse process remains stochastic, starting from different realizations of the initial Gaussian noise yields different high-resolution samples that all satisfy the same coarse SSS observation Y, allowing us to explore a range of plausible fine-scale fields instead of a single averaged solution. A schematic representation of the conditional inference phase is shown in Figure 3.

Different random initialization gives rise to a different denoised patch, one question arisesBecause the reverse diffusion process remains stochastic, different random initializations give rise to different plausible denoised patches for the same coarse SSS constraint. This raises two related questions: how to select a unique down-scaled SSS field? The method we chose is the following: the mean image of the generated samples distribution is computed, and then the sample single displayed reconstruction, and how to interpret the spread across generated samples. In this study, we compute the ensemble mean image and select the generated member that is closest to this mean image is chosen. By doing so, we make sure to remain close to what we can approximately call the "center" of the samples distribution. Such an assumption remains valid only because, as is the case in our problem, the distribution is well constrained by additional SST and SSH observations (for a more diverse distribution, the mean of samples that are too different from one another does not make sense anymore). This method has the advantage of being rather general for the broad evaluation of the model's performances that is conducted in this article, and more task-specific methods could be developed depending on the applications. Choosing a number N of different initializations provides a plausible set of N $1/12°$ SSS patchesin ℓ_2 norm. This closest-to-mean member acts as a medoid-like representative of the conditional ensemble: it remains close to the ensemble central tendency while preserving the spatial texture of an individual realization. By contrast, the

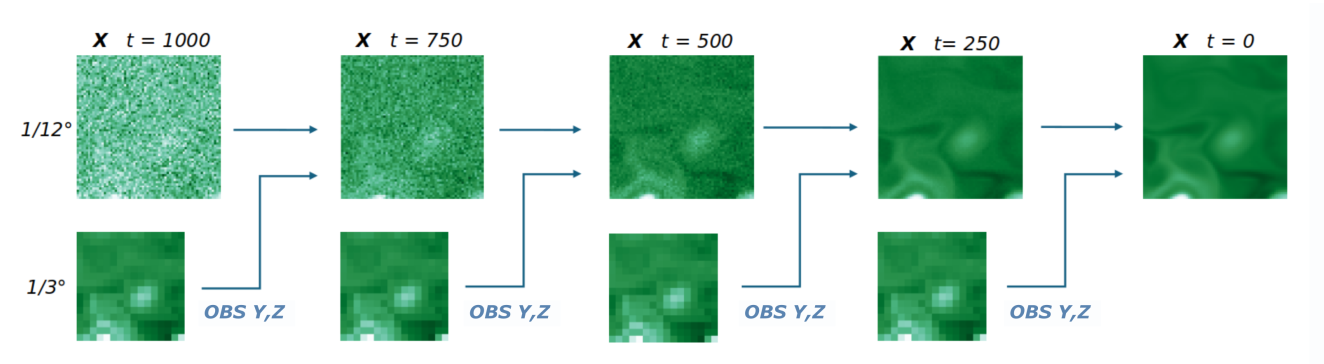

Figure 3. Schematic representation of the conditional inference phase. First The first line , represents shows the reverse diffusion trajectory from left, a random initialization , to right, toward a retrieved $1/12°$ denoised SSS patch at different steps t. Second The second line represents shows the $1/3°$ SSS observations observation used to guide the denoising procedureat each step. For the clarity of the figure only the coarse SSS field constraint is shown, but Z $1/12°$ images are inferred at the same time where Z, depending of on the input of configuration the diffusion model , can be generates a multichannel high-resolution state including SSS and the associated SST , SSH and/or bothSSH channels. The evaluated downscaling target remains SSS.

ensemble mean may reduce pixel-wise error but can smooth fronts and filaments when plausible structures are slightly displaced across members. The choice between an ensemble mean, a representative member, or another criterion is therefore application dependent. The ensemble variance can also be interpreted as a first diagnostic of reconstruction ambiguity.

In the following, we propose the Gradient-Enhancing (GE) algorithm to select the SSS $1/12°$ field with the highest contrast, enhancing small structures.

3.4 Gradient-Enhancing algorithm,

The image gradient refers to the directional change in the pixel intensity in both the x-axis and the y-axis of an image. It is a key ingredient in image processing and computer vision, particularly in detecting edges or in processing images with edges. The problem of too-low image gradients is inherent to many image-generation tasks. When using an MSE-type cost function during training, the model is not explicitly encouraged to produce high-contrast predictionscommon in image reconstruction tasks. Deterministic models optimized with mean-square-error objectives tend to approximate the conditional mean of possible high-resolution solutions, which can smooth fronts and filaments when several plausible fine-scale configurations are compatible with the same coarse observation. During inference, the **GE** procedure guides the generative process towards higher-contrast samples in the learned distribution. The idea is to favor, at each denoising step, the realizations that exhibit stronger local gradients while still remaining consistent with the distribution learned during traininglearned distribution and the coarse SSS constraint. The method is built on top of the pseudo-inverse guidance described above, by adding an additional contrast-based guidance term.

The complete GE algorithm is detailed in Annex A1 and is applied each time step during the inference phase. In the diffusion timestep t, we consider the one-shot prediction of the network, that is, the current estimate of the final fully denoised image available in DDIM. This estimate gets better and better as the denoising process goes on. On kernels of 2×2 pixels of the image, we enhance the local contrast by subtracting the kernel's mean to the 4 pixels, multiplying by a scaling factor, then adding the mean back. In regions where pixel values deviate strongly from the local kernel mean, indicating the presence of a front or other fine-scale structure, the contrast is amplified. In smoother regions, the effect remains weaker so that no spurious small-scale features are created. The resulting contrast-enhanced image is then fed back into the update for $t-1$, thereby nudging the network towards higher-contrast samples, which still lie within the learned distribution. The process is sensitive to instabilities, especially for high gradients that can explode over numerous time-steps of multiplicative scaling. The following scaling term provides a damping effect on the highest gradients of the image to prevent those instabilities:

$s = \gamma - \delta \sqrt{\frac{|\hat{X}_t|}{\max|\hat{X}_t|}}$ where max is computed over all kernels.

The scaling factor s is equivalent to $\gamma - \delta$ for high gradients and to γ for small gradients; both parameters are kept close to one and tuned so that the final gradient distribution of the generated images matches as closely as possible that of the test set. In summary, we want to reduce the overall proportion of low gradients, enhance the overall proportion of high gradients, and do so at the right places in the generated samples. In the following, we refer to the diffusion model combined with this gradient-enhancing (GE) procedure as DIFF-SST-SSH-GE.

4 Experiments design

We perform 6 different experiments in order six experiments to understand how diffusion models can deal with a set of correlated variables and compare to a state of the art handle correlated ocean variables and to compare them with a state-of-the-art deterministic approach. The 5 first are diffusion models using the same architecture but taking different auxiliary variables as input, and the last experiment is used as a benchmark for the methodological approachfirst five experiments use the same diffusion architecture but differ in the auxiliary variables and inference strategy. The first two add a single input either SST or SSH to the SSS (they SSS and are denoted DIFF-SST and DIFF-SSHrespectively), respectively. The third is the complete diffusion model as presented in section 3 conditioned on both SST and SSH, denoted DIFF-SST-SSH. As will be shown below, the performances of DIFF-SST and DIFF-SSH are inferior to those of DIFF-SST-SSH, so the last three experiments focus on comparisons with DIFF-SST-SSH only. Then, we . We then apply the **GE** algorithm during the Conditioning phase enhancing the contrasts of the SSH output images (conditioning phase to enhance the contrast of the generated SSS fields, denoted DIFF-SST-SSH-GE). Finallywe compute the "Best" theoretical performances by comparing the estimated outputs of DIFF-SST-SSH with the actual desired output images and choose the nearest to the actual desired output (denoted . Finally, we compute an oracle diagnostic, DIFF-SST-SSH-BESTrespectively), thus we get a lower bound of the performances: among multiple samples generated by DIFF-SST-SSH, we select the member closest to the known high-resolution GLORYS reference SSS. This selection requires access to the truth and is therefore not operational; it is used only to estimate the headroom of the generated ensemble.

In a last experiment we apply the RESAC methodology adapted to the problem of downscaling of the SSS with the same GLORYS data (SSS and SST). The RESAC model (REconstruction by Scale-Adaptive Convolution) is based on the idea that structures observed in the ocean have approximate fractal properties, and that in incompressible fluids eddies interact most strongly with vortices of comparable size (Thiria et al., 2023). Consequently, two variables are expected to be most strongly correlated at similar spatial scales. These considerations motivated a convolutional (CNN) architecture split into several blocks, in which the resolution of the target variable, in our case SSS, is progressively refined by successively adding SST observations at matching resolutions. The network is thus guided by the scale-dependent physical characteristics of the fields under study, and RESAC provides a strong baseline for assessing the added value of diffusion-based downscaling.

To define our training and testing strategy, the temporal variability of surface salinity must be taken into account. For the Gulf Stream region, the dominant signal is a marked seasonal cycle, with shifts in the position and intensity of fronts and eddies that modulate the distribution of SSS throughout the year, in connection with changes in mixed-layer depth and air–sea fluxes (Chassignet and Xu, 2017). These seasonal shifts are consistent with documented variations in eddy kinetic energy and Gulf Stream path/front dynamics (Zhai et al., 2008; Joyce et al., 2019b, a). We exploit this temporal variability when defining the test configuration. The entire year 2020 is reserved as the independent test set, while the remaining years (1993-2018) are used for training and 2019 for validation and model parameterization. This setup allows us to evaluate not only the reconstruction skill of the diffusion model under typical conditions, but also its ability to generalize across the full seasonal cycle.

In summary, we assess the super-resolution of a coarse $1/3°$ SSS observation to a target $1/12°$ field, given additional SSS field, using high-resolution SSS and SST auxiliary SST and/or SSH fields, under the observation model of EqEqs. (1) and (2). The domain, period, and preprocessing follow (Section 2.1). The forward diffusion . The diffusion prior is trained unconditionally on $1/12°$ GLORYS images (Section 3.2); observational information is injected at each inference step through the pseudo-inverse guidance of Section 3.3. For DIFF-SST-SSH-GE, the gradient-enhancing algorithm (Section 3.4) is used. For all diffusion models (except for except DIFF-SST-SSH-BEST), we compute a 'Median' realization , we select the closest-to-mean representative member over 15 samples generated from different random seeds. The main characteristics of the experiments are summarized in Table 1. We note that in diffusion experiments, SST and SSH are treated as generated channels because the diffusion model samples a multichannel ocean surface state. They are not evaluated as final downscaled products in this study; the target variable remains SSS. This formulation keeps a common observation-operator framework for future applications in which SST or SSH may also be partially observed, degraded, or associated with spatially varying uncertainty.

During training , we use the root-mean-square error (RMSE) between predicted For the deterministic RESAC baseline, RMSE is used as the training loss. For diffusion models, training uses the standard noise-prediction mean-square objective defined in Eq. (5); RMSE between reconstructed and reference SSS as loss function and validation is then used as a validation and evaluation metric. Although RMSE is standard in super-resolution and data assimilation applications, it tends to favour overly smoothed solutions that minimize the mean-squared deviation from the reference field. To better assess the quality of small-scale structures, we complement RMSE with three additional indicators: the Structural Similarity Index (SSIM) (Wang et al., 2004), the distribution of spatial gradients, and the temporal Fourier spectrum (FFT) of SSS.

SSIM is widely used in image processing to evaluate the perceptual quality of generated images (Wang et al., 2004; Ledig et al., 2017). It combines luminance, contrast and structural information into a single similarity index, typically between -1 and 1, with 1 indicating locally identical images. Although it has no direct physical meaning for salinity, it provides a convenient way to compare local structures and contrasts between predicted and reference fields. The SSIM between two images x and y is defined locally on a kernel by

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{9}$$

where

- μ_x and μ_y are the local means of x and y,
- σ_x^2 and σ_y^2 are the local variances,
- σ_{xy} is the local covariance between x and y,
- $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ are stabilization constants,
- L is the dynamic range of the pixel values (e.g. 255 for 8-bit images),
- K_1 and K_2 are set to 0.01 and 0.03, respectively.

Contrast is also directly related to spatial and temporal gradients and to temporal fourier spectra. For each model, we therefore generate SSS maps over the entire 365-day test period and analyze them in term of gradient distribution and Fast Fourier Transform. These two diagnostics provide complementary information on the ability of the different methods to reproduce realistic levels of small-scale variability and sharp fronts, beyond what can be inferred from RMSE and SSIM alone.

In the following we present the results of the six experiments trained with the same training set, computing the same indices on the year 2020.

5 Results

We focus our results on the mesoscale reconstruction , for that purposefine-scale reconstruction at the GLORYS reference resolution. For that purpose, all the following performances are estimated on the central patch of the Gulf Stream area using the year 2020.

5.1 Influence of auxiliary variables

For the first three experiments (DIFF-SST, DIFF-SSH, DIFF-SST-SSH): from , we generate $K{=}15$ samples generated with independent noise seeds, we compute the ensemble meanimage , and select the single realization that is closest to that mean in ℓ_2 norm. This "closest-to-mean " medoid member provides a pragmatic proxy for the center of the conditional distribution and is appropriate here because auxiliary SST/SSH strongly constrain fine-scale structure, yielding tightly clustered sampleswhile preserving the texture of an individual realization. An example of different plausible SSS outputs of DIFF-SST-SSH is illustrated in Figure 4. Note that inference provides patches of Inference also generates the associated $1/12°$ SST and$1/12°$ SSH associated with the SSS output recovered simultaneously during the inference phase. This could be interesting when dealing with observations. The Variance Map /or SSH channels, depending on the configuration, because the diffusion prior represents a multichannel GLORYS-consistent state. The evaluated downscaling target, however, remains SSS. The variance map displays the pixel-wise ensemble variance across the total collection of thirty SSS $1/3°$ generated patchesgenerated high-resolution SSS patches, considered as a first diagnostic of reconstruction ambiguity. As expected, the variance appears to be higher in turbulent spread is higher in dynamically active regions where fine-scale structures are more contrasted or slightly displaced across samples.

For DIFF-SST-SSH-BEST: from $K=40$ samples, we pick the realization that minimizes RMSE to the reference; this is not operational (it requires ground truth) but indicates the headroom available from the same trained prior.

For DIFF-SST-SSH-GE: we apply the gradient-enhancing guidance during inference and use the same central selection (closest-to-mean over $K=15$ guided samples), yielding higher-contrast outputs while remaining distribution-consistent.

In Table 2 we give for the six experiments the mean RMSE (psu) and SSIM during the test year 2020 over the test domain. These performances are computed from a daily basis. Conditioning on high-resolution auxiliaries at $1/12°$ improves SSS reconstructions from coarse SSS at $1/3°$. DIFF-SST outperforms DIFF-SSH and adding together SST and SSH improves the global RMSE. Indeed, as mentioned in Introduction, SST adds more information than SSH; SST provides the most precise guidance for the reconstruction of SSS. SST exhibits correlated structures at a very fine scale, unlike SSH, which is much smoother in the current GLORYS-based setup. A yearly average quantify these gains: DIFF-SST improves RMSE by $\sim 21\%$ relative to DIFF-SSH (0.045 vs. 0.057 psu); adding SSH on top of SST yields a further $\sim 11\%$ improvement (0.040 vs. 0.045 psu for DIFF-SST-SSH vs. DIFF-SST), indicating complementary mesoscale constraints from SSH.

To better understand the behavior of the six models, we present daily time series of RMSE over the year 2020. The y-axis is the mean RMSE computed over the Gulf Stream domain. As the order of magnitude is different for the six experiments, we compare DIFF-SST and DIFF-SSH in Figure 5 and the 4 other models in Figs. 6. For the six experiments, across the seasonal cycle, errors peak in late spring–summer (day $\sim$170-200) and decrease towards winter. This timing coincides with seasonal maxima of eddy kinetic energy and frontal activity in the Gulf Stream region (Zhai et al., 2008; Joyce et al., 2019b, a).

As expected, in Figure 5 DIFF-SST (magenta line) always systematically outperforms DIFF-SSH (green line). When looking at the daily time series of In Figure 6, DIFF-SST-SSH (magenta line) and RESAC (yellow line) give similar time series very close to the lower bound of DIFF-SSH-SST-BEST display very similar daily RMSE time series, while DIFF-SST-SSH-BEST (blue line) provides a baseline/lower envelope because it uses the known reference for sample selection. DIFF-SST-SSH-GE (green line) is always above the two others. Using the GE algorithm deteriorates both performances a little, but differences are quite small for all experiments. This is visible in table 2 where has a larger RMSE than the standard DIFF-SST-SSH configuration. Thus, GE does not improve pixel-wise scores; rather, it is designed to test whether contrast-oriented inference can recover sharper structures at the cost of a moderate degradation in RMSE/SSIM. This distinction is important because RMSE and SSIM are quite similar. The differences are often small but we expect that DIFF-SST-SSH-GE has provided more contrasted imagesalone do not fully characterize frontal contrast and high-frequency variability.

For that purpose we compare different snapshots and verify that the images always show an advantage in contrast for DIFF-SST-SSH-GE. When compared with the REFERENCE field, visual inspection (then quantify the structural differences using gradient and spectral diagnostics. Visual differences in Fig. 7) shows that DIFF-SST-SSH-GE sharpens fronts and filaments relative to DIFF-SST-SSH and RESAC; its effects are subtle at the scale of the full patch, and the figure should be interpreted as a qualitative illustration rather than as direct proof of sharpening.

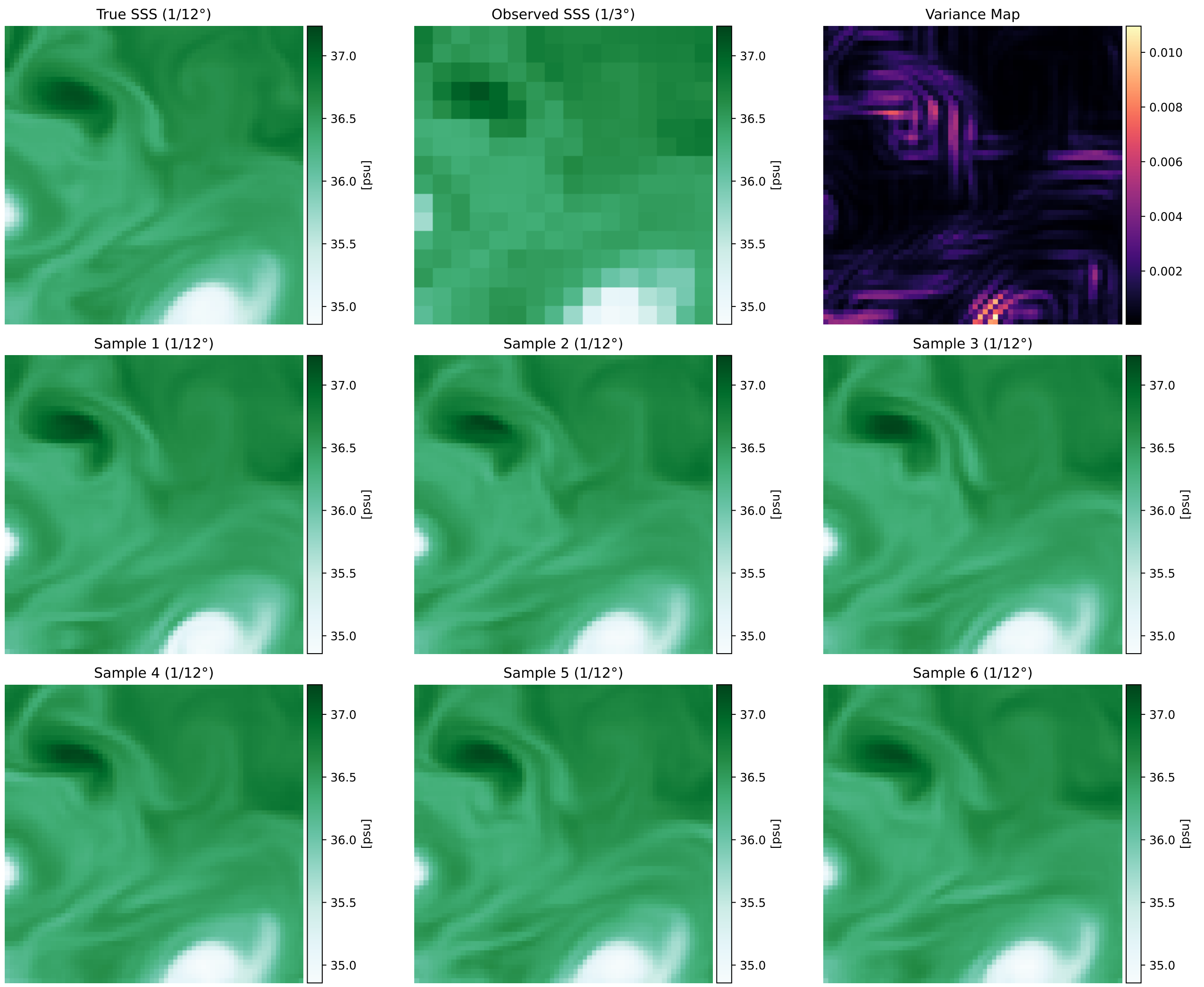

Figure 4. Collection of six $1/12^\circ$ SSS patches generated on for the same day of 2020 with 6 six different inference noisesnoise initializations. First line presents The first row shows, from the left to the right: , the Reference reference high-resolution SSS patchto be retrieved (True SSS), the $1/3^\circ$ coarse SSS patch used for calibrationconstraint, and the variance map displays the pixel-wise ensemble variance computed across the total collection of thirty generated high-resolution SSS patches. As expected, the The variance values appear to be higher is larger in turbulent dynamically active regions where several fine-scale structures are more contrastedconfigurations remain compatible with the same coarse constraint as seen in the second and third rows. This spread is interpreted as a reconstruction-ambiguity diagnostic.

The effect of GE on small-scale structure are analysed quantitatively is therefore assessed more directly in the next subsection using the gradient distribution and temporal Fourier spectrum. For all indices (RMSE, SSIM) computed yearly or daily the ranking of the models remains stable (Fig. 6). Thus, as using both auxiliary variables gives the best performances, DIFF-SST-SSH only will be used for the last comparisons.

5.2 Spatio-temporal structure and contrast enhancement

To assess fine-scale structure beyond pixelwise error, we analyse spatial gradient-magnitude distributions and temporal Fourier spectra over the 2020 test year. The reference gradient distribution (Fig. 8, left) is strongly peaked at low gradient values, corresponding to smooth regions, with a long tail representing fronts and energetic structures. The right panel shows the difference between model and reference gradient distributions for DIFF-SST-SSH, DIFF-SST-SSH-GE, and RESAC. All methods overestimate the proportion of weak gradients and underestimate intermediate-to-strong gradients, consistent with residual smoothing. DIFF-SST-SSH-GE reduces this bias over the low-to-intermediate range, indicating improved contrast and sharper reconstruction of fine-scale structure.

Temporal spectra provide a complementary view (Fig. 9). DIFF-SST-SSH and RESAC under-represent high-frequency (weekly-daily) variance, while DIFF-SST-SSH-GE remains closer to the reference across most frequencies and becomes slightly positive at the highest frequencies, consistent with recovered small-scale temporal energy. DIFF-SST-SSH-BEST does not uniformly improve all spectral ranges, illustrating that selecting a sample solely by minimizing RMSE does not guarantee optimal spatio-temporal consistency. Since BEST selection is not applicable operationally, it is interpreted only as a performance headroom indicator.

Overall, when pixelwise accuracy is the primary objective, DIFF-SST-SSH provides a strong operational choice with performance comparable to RESAC. When the fidelity of fronts and filaments is critical (e.g. frontal detection, feature tracking, or Lagrangian applications), DIFF-SST-SSH-GE offers a sharper representation of fine scales at a small cost in RMSE/SSIM. Across experiments, diffusion-based reconstructions are visually sharper than RESAC, although their gradient magnitudes remain, on average, weaker than those of the reference fields by roughly 10%.

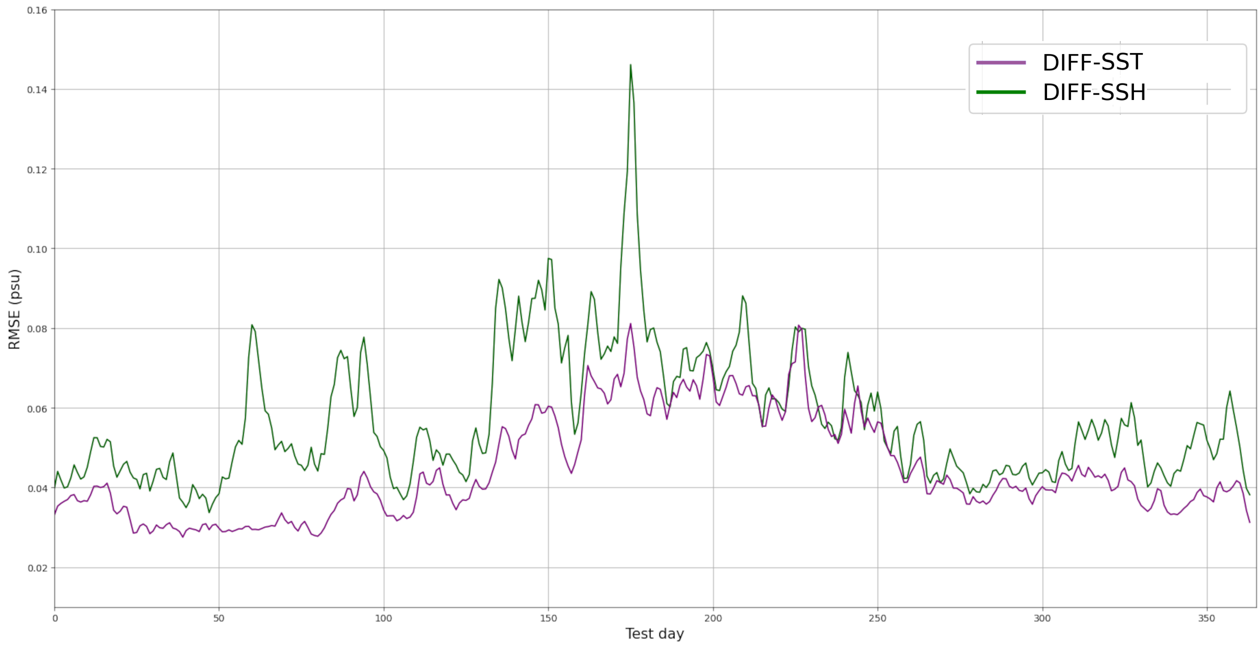

Figure 5. Daily RMSE of SSS $1/12^{\circ}$ reconstructions over the year 2020 using DIFF-SST (purple) or DIFF-SSH (green). RMSE on the y-axis is the mean RMSE computed over the Gulf Stream domain. DIFF-SST provides systematically stronger guidance than SSH, with the largest gaps during the most energetic period (day $\sim$170-200).

Table 1. Main characteristics of the six experiments.

Model	**Observation / conditioning**	**Generated high-resolution channels**	**Evaluated target**
RESAC	SSS $1/3^{\circ}$, SST $1/12^{\circ}$	SSS $1/12^{\circ}$	SSS $1/12^{\circ}$
DIFF-SST	SSS $1/3^{\circ}$, SST $1/12^{\circ}$	SSS, SST $1/12^{\circ}$	SSS $1/12^{\circ}$
DIFF-SSH	SSS $1/3^{\circ}$, SSH $1/12^{\circ}$	SSS, SSH $1/12^{\circ}$	SSS $1/12^{\circ}$
DIFF-SST-SSH	SSS $1/3^{\circ}$, SST and SSH $1/12^{\circ}$	SSS, SST and SSH $1/12^{\circ}$	SSS $1/12^{\circ}$
DIFF-SST-SSH-BEST	SSS $1/3^{\circ}$, SST and SSH $1/12^{\circ}$	SSS, SST and SSH $1/12^{\circ}$	SSS $1/12^{\circ}$
DIFF-SST-SSH-GE	SSS $1/3^{\circ}$, SST and SSH $1/12^{\circ}$	SSS, SST and SSH $1/12^{\circ}$	SSS $1/12^{\circ}$

Table 2. Average metrics (RMSE and SSIM) for the year 2020 for the six experiments.

Metric	**DIFF-SSH**	**DIFF-SST**	**DIFF-SST-SSH**	**DIFF-SST-SSH-BEST**	**DIFF-SST-SSH-GE**	**RESAC**
RMSE (psu)	0.057	0.045	0.040	**0.038**	0.046	0.041
SSIM	0.92	0.93	0.94	**0.95**	0.93	0.95

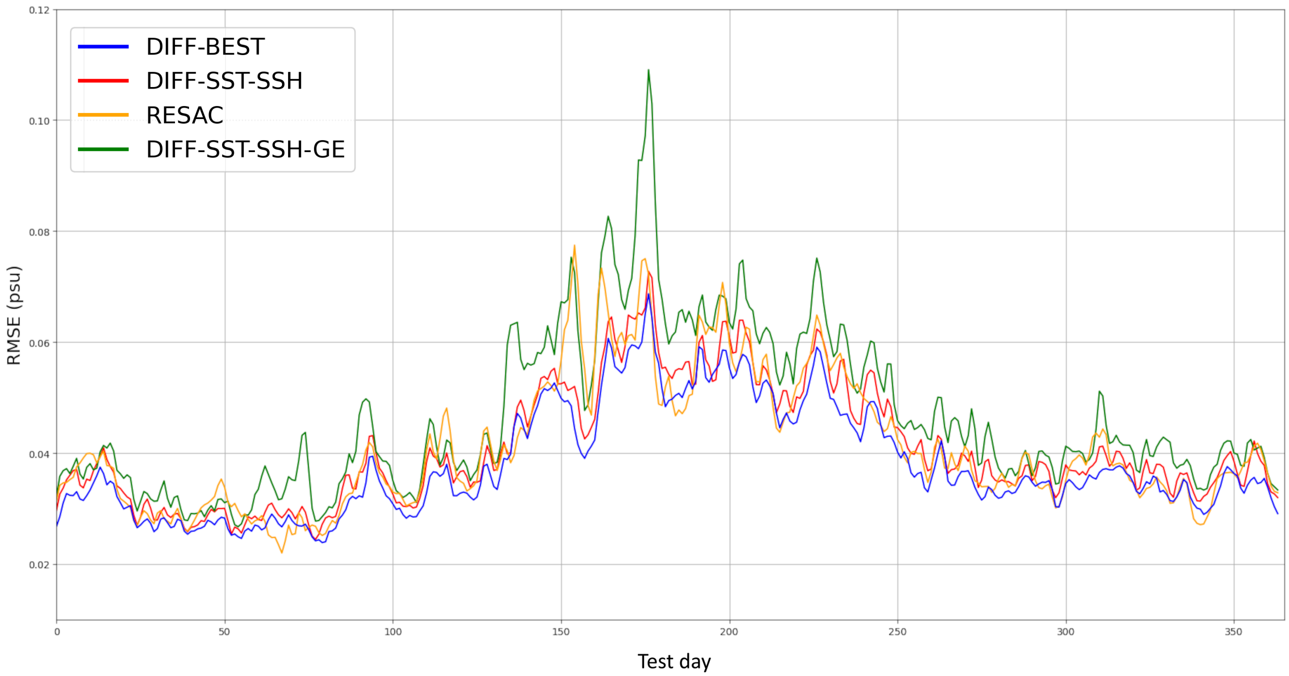

Figure 6. Daily mean RMSE of $1/12°$ SSS reconstructions over the year 2020 using DIFF-SST-SSH (red line), DIFF-SST-SSH-BEST (Blue), DIFF-SST-SSH-GE (green), RESAC (yellow). The x-axis represents the days of year 2020; RMSE, on the y-axis, is the mean RMSE computed over the Gulf Stream domain. DIFF-SST-SSH-BEST gives the lowest error envelope; the ranking of the models remains stable with respect to Table 2.

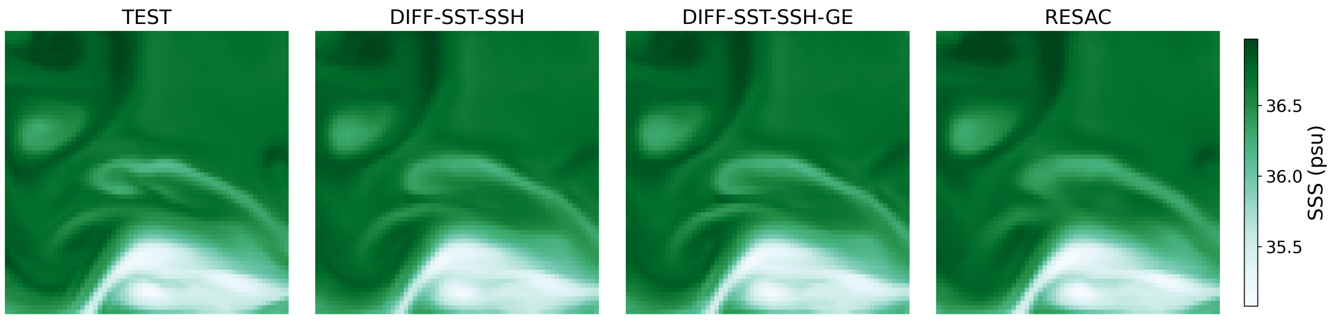

Figure 7. Example of SSS reconstruction at $1/12°$ for a Gulf Stream patch (day 120 of 2020). Left to right: REFERENCE (TEST), DIFF-SST-SSH, DIFF-SST-SSH-GE, and RESAC. All panels share identical color limits (psu).

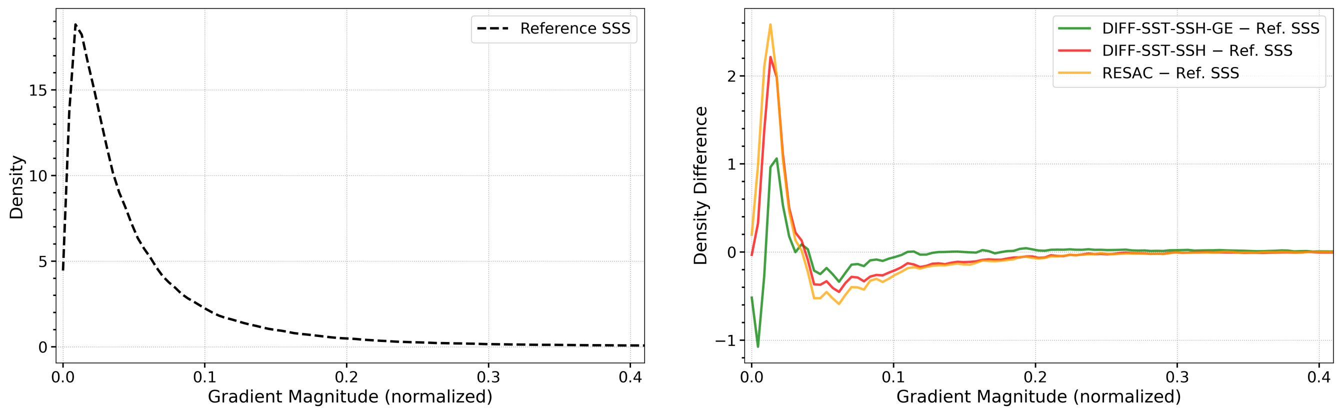

Figure 8. Left: Test Gradient Distribution performed for each pixel gradient-magnitude distribution computed over all pixels and each image all images from the 2020 TEST test dataset; . Right: difference of Test Gradient Distribution with those of between each model distribution and the different models reference distribution for DIFF-SST-SSH (redline), DIFF-SST-SSH-GE (greenline) and RESAC (orangeline). For the TEST Gradient Distribution the The high proportion of low gradients corresponds to flat smoother regionsin the images, and while the fewer higher gradients correspond tail corresponds to turbulent fronts and energetic regionsstructures. Although all models overestimate low gradients to the detriment of higher ones, DIFF-SST-SSH-GE performs reduces part of the best weak-gradient excess and effectively recovers some high more intermediate-to-strong gradients, likely resulting in slightly better contrast and front reconstructionbut the negative difference at the weakest gradients also indicates that the procedure may over-sharpen some regions.

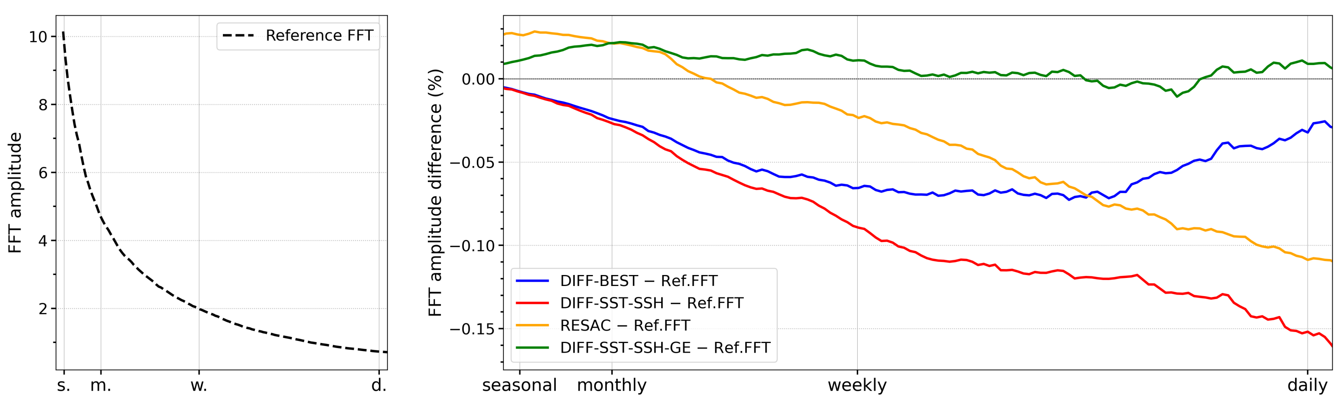

Figure 9. Left: Reference's Mean Temporal reference mean temporal Fast Fourier Transform (RMT FFT) amplitude performed computed for each pixel over the 2020 test year 2020, and averaged over all pixels in the studied study area; . Right: Models' Mean Temporal Fast Fourier Transform (RMT$_{model}$ FFT) amplitudes percentage difference with the Reference in percent: RMT$_{model}$ FFT - RMT between each model mean temporal FFT over 2020, for all modelsamplitude and the reference. DIFF-SST-SSH-GE (greenline) is closest increases high-frequency temporal energy relative to the truth across most of standard diffusion configuration, but the spectrum and turns slightly positive at difference over parts of the highest frequencies, indicating recovered spectrum also indicates a possible over-amplification of small-scale temporal energyvariability.

6 Discussion

The experiments were designed to isolate methodological behavior in a controlled setting, using GLORYS as a dynamically consistent constrained high-resolution reference and synthetically degraded SSS as coarse observationsat coarse resolution. Within this framework, the results support four main conclusions and clarify several limitations that must be addressed before transfer to satellite SSS products.

6.1 Main findings and interpretation

High-resolution SST is the dominant auxiliary constraint for SSS downscaling in the Gulf Stream. Across the full 2020 test year, conditioning on SST yields substantially lower error than conditioning on SSH alone (Table 2; Fig. 5). This ranking is robust throughout the seasonal cycle, including the most energetic period (day ∼170-200) when reconstruction is most challenging (Figs. 5 and 6). This is consistent with the visual comparison of the three variables (Fig. 2): SST shares filamentary and frontal structures with SSS at finer scales, whereas SSH fields used here are smoother and mostly express mesoscale organization. In other words, SST appears to supply the *local* small-scale information that the model can "lock onto", while SSH provides a weaker but complementary *contextual* constraint on the position and intensity of mesoscale features.

Adding SSH on top of SST provides a smaller but consistent gain. DIFF-SST-SSH improves on DIFF-SST in yearly-mean RMSE (Table 2) and remains competitive day-by-day (Fig. 6). The improvement is modest compared to the DIFF-SSH/DIFF-SST gain, but it is physically plausible. Yet this result should be interpreted with caution. It does not necessarily imply that SSH is intrinsically less relevant for constraining salinity fine scales; rather, it may reflect the comparatively smooth character of the SSH fields used in the present GLORYS-based setup. This point is particularly important in light of recent SWOT observations, which reveal a much richer fine-scale sea-surface height variability than previously accessible from conventional altimetry (Morrow et al., 2019; Archer et al., 2025). Recent studies also suggest that part of this small-scale and high-frequency SSH variability is still underestimated or only partially represented in present-generation 1/12° model-based products (Le Traon et al., 2019). Therefore, the weaker performance of SSH in our framework may be tied to the effective resolution and information content of the SSH product used here, rather than to the physical usefulness of SSH itself. Revisiting this question with SWOT-based or SWOT-constrained SSH products is a natural and important next step.

Diffusion with pseudo-inverse guidance is competitive with comparable to a strong CNN baseline under pixelwise metrics, while providing a conditional ensemble. With the "closest-to-mean " selection strategy (Section 5), DIFF-SST-SSH and RESAC yield very similar yearly-mean RMSE/SSIM (Table 2) and similar temporal evolution (Fig. 6). This is an important outcome: The diffusion approach should therefore not be interpreted as providing a large RMSE improvement over RESAC in this controlled super-resolution setting, the diffusion approach is not merely visually appealing but quantitatively competitive. Its added value lies in the fact that the experiment. RESAC remains a strong deterministic baseline, specifically designed for scale-adaptive downscaling. The added value of the diffusion framework is different: inference naturally produces an *ensemble* ensemble of plausible high-resolution fields conditioned on the same coarse observation (Fig. 4), which can be used for uncertainty-aware diagnostics and downstream applications. This generative aspect is a key advantage over deterministic CNNs, provided the ensemble spread remains calibratedand provides a diagnostic of reconstruction ambiguity.

Contrast-oriented guidance shifts the solution toward more realistic stronger small-scale structurecontrast, but trades off against RMSE/SSIM and can over-amplify variability. The GE procedure produces sharper fronts/filaments (Fig. 7) and improves fine-scale diagnostics: the gradient-magnitude distribution bias is reduced in the low-to-intermediate range (Figincreases gradient contrast relative to the standard DIFF-SST-SSH configuration (Figs. 8), andthe temporal spectrum is closer to the reference at weekly–daily frequencies (Fig.and 9). At the same time, DIFF-SST-SSH-GE slightly degrades RMSE/SSIM relative to DIFF-SST-SSH (Table 2), which is expected : because RMSE favors smoother reconstructions and penalizes phase/position errors strongly, even if gradients are more realistic. This confirms that model choice should . The gradient histogram also shows that GE reduces the proportion of weak gradients relative to the reference, and the temporal spectrum can exceed the reference over part of the frequency range. These diagnostics indicate that GE should be interpreted as a contrast-enhancing variant with a structural trade-off, not as a uniformly superior reconstruction. Model choice should therefore depend on the scientific objective: for applications prioritizing pixelwise accuracy(e.g. direct assimilation with strong observation-space penalties), DIFF-SST-SSH is preferable; for applications where *front and filament fidelity* matters(front detection, feature tracking, Lagrangian diagnostics), DIFF-SST-SSH-GE may be advantageous despite slightly worse RMSE/SSIM, consistent with the well-known tension between distortion metrics and perceptual/structural realism (Wang et al., 2004; Ledig et al., 2017), GE-like guidance may be useful but should be tuned carefully.

6.2 What the diagnostics reveal beyond RMSE

The additional diagnostics clarify *how* errors differ between approaches. Gradient histograms show that all methods overpopulate weak gradients and under-represent the mid–high part of the mid-high tail (Fig. 8), indicating residual over-smoothing relative to GLORYS. The GE variant partially corrects this by sharpening enhancing low/intermediate gradients, which aligns with the visual impression of a better-defined front (Fig. 7)but the negative difference at the weakest gradients suggests that it may also over-sharpen the field by converting too many weak gradients into stronger ones. The temporal FFT analysis then provides an independent view: DIFF-SST-SSH and RESAC under-represent high-frequency (weekly–daily) variance, while DIFF-SST-SSH-GE is closest to the reference across most restores more energy but can exceed the reference over part of the spectrum (Fig. 9). Taken together, these results support a nuanced conclusion: *the diffusion prior with standard selection is already competitive under pixelwise scores, but its default samples remain slightly too smooth; targeted inference-time guidance can recover part of the missing variability*, at the risk of over-amplification if the contrast guidance is too strong.

A related implication concerns sample selection. The "BEST" choosing strategy indicates headroom from the same trained prior (Table 2), but it is not operational and should not be over-interpreted as achievable performance. More importantly, the fact that "BEST" does not uniformly improve all diagnostics (e.g. FFT spectral behavior can differ) highlights that selecting samples solely to minimize RMSE does not guarantee spatio-temporal realism. This motivates future work on *task-aware* or *physics-aware* selection criteria that can be computed without access to truth.

6.3 Limitations and perspectives for transfer to satellite products

The present study was designed as a controlled methodological experiment, using GLORYS as a dynamically consistent constrained high-resolution reference and synthetically degraded SSS as coarse observations. This framework is well suited to isolate the behavior of the diffusion-based downscaling method, but it also defines the main limitations of the present work and the conditions required for transfer to real satellite SSS products.

First, GLORYS is not assumed to be a perfect truth. It combines an ocean model with heterogeneous assimilated observations, and local inconsistencies between SSS, SST and SSH may remain. In this paper, GLORYS provides a complete multivariate gridded reference from which the diffusion model learns a GLORYS-consistent joint prior over SSS, SST and SSH. The generated reconstructions should therefore be interpreted as physically plausible within the GLORYS framework, not as guaranteed representations of the real ocean at all scales.

Second, the use of GLORYS avoids many of the complexities of L-band satellite retrievals, including sensor-specific biases, correlated retrieval errors, and irregular observation patterns (Boutin et al., 2014; Reul et al., 2020; Reverdin et al., 2024)(Boutin et al., 2014, 2021; Reul et al., 2020; Reverdin et al., 2024). While this is appropriate for method development, it also means that the reported metrics should not be interpreted as direct estimates of the performance expected for SMOSor SMAPobservations. In particular, the synthetic degradation applied here remains much simpler than the effective observation process of present-day satellite SSS products, SMAP, or multi-satellite products such as CCI products (Boutin et al., 2021). Real Level-4 SSS products are not clean low-pass versions of the true SSS field. For example, ESA CCI SSS is a multi-sensor Level-4 product derived from L-band radiometry; although it is sampled on a $0.25°$ grid, its effective spatial resolution is about 50 km and it provides uncertainty and quality information such as random error, unexplained variability, number of observations, outliers, and land/ice/SSS quality flags (Boutin et al., 2021). Such information should be used in future work to weight the observation guidance spatially and to mask unreliable pixels.

SecondThird, the observation operator H used in this study mainly reflects spatial averaging and sub-sampling (Eq. 1). For real applications, H should mimic the targeted satellite product much more faithfully, including its effective spatial footprintfootprint, grid, masks, and spatially heterogeneous noise characteristics (Boutin et al., 2021; Reul et al., 2020; Reverdin et al., 2024)uncertainty. Because pseudo-inverse guidance injects observational information at each denoising step (Song et al., 2023; Chung et al., 2024), the realism of H is a central requirement for domain transfer. This also argues for enforcing consistency in observation space, i.e. In practical applications, the guidance term should therefore be uncertainty aware, matching $H(X)$ to the satellite observation Y on the satellite grid while being uncertainty aware, especially in the presence of gaps and variable retrieval qualitydown-weighting uncertain, missing, or contaminated pixels.

Within this perspective, the present results indicate that high-resolution SST is currently the most effective auxiliary variable for constraining fine-scale SSS. HoweverFourth, the auxiliary variables also have product-dependent limitations. High-resolution SST is an effective constraint in the present experiment, but observed infrared SST is unavailable under clouds. Daily gap-free Level-4 SST products may be distributed on fine grids, yet their local information content can degrade under persistent cloud cover because they rely on interpolation, microwave SST, previous clear-sky observations, in-situ data, and the analysis method (Embury et al., 2024; Donlon et al., 2012; Chin et al., 2017). In such regions, SST should be treated as an uncertain auxiliary constraint rather than as a uniformly reliable high-resolution observation. Similarly, the relatively weaker contribution of SSH in our experiments should not be regarded as definitive, as it may partly reflect the smooth character of the SSH fields in the present GLORYS-based setup. This balance may evolve as higher-resolution SSH observations become available, particularly through SWOT, which may provide a richer fine-scale dynamical constraint than that represented in current-generation reanalyses.

Fifth, the role of in-situ observations should be stated, such as Argo profiles within the proposed transfer framework. The high-resolution prior would still be learned from gridded reanalysis or simulation fields, while the observational constraint would be provided by a satellite SSS product through a product-aware observation operator. Argo, thermosalinograph data, moorings, drifters, and regional campaigns would then provide independent validation, calibration, and bias assessment.

Finally, operational deployment requires sample selection and validation uncertainty-calibration strategies that do not rely on access to the true high-resolution field. Validation should rely on independent references such as Argo near-surface salinity, moorings, drifters, and regional analyses (Vinogradova et al., 2019; Reul et al., 2020; Reverdin et al., 2024)The closest-to-mean member used here provides one representative realization, while the ensemble spread shown in Fig. 4 provides a first diagnostic of reconstruction ambiguity. A calibrated uncertainty product would require testing the relationship between ensemble spread and actual reconstruction error using independent validation data. In that context, the character of the diffusion framework is not only a source of multiple plausible reconstructions, but also a means way to diagnose when the solution is weakly constrained, for example near coasts, under cloud-degraded auxiliary SST, or under strong retrieval noisesatellite SSS retrieval uncertainty.

6.4 Concluding remarks

In a controlled Gulf Stream experiment, diffusion models guided by pseudo-inverse conditioning can downscale coarse SSS to $1/12°$ with accuracy comparable to a strong convolutional baseline, while naturally providing a conditional ensemble of plausible high-resolution fields. In the current GLORYS-based context, high-resolution SST is the primary source of fine-scale guidance; SSH adds complementary mesoscale context. A contrast-oriented guidance term can partially recover missing gradients and high-frequency temporal variance, at a small cost in RMSE/SSIM, highlighting an application-dependent trade-off between distortion metrics and structural realism (Wang et al., 2004; Ledig et al., 2017).

The next methodological frontier is domain transfer transferability to satellite SSS. This requires (i) realistic sensor-aware product-aware observation operators and noise/mask models in H, (ii) uncertainty-aware weighting of satellite SSS and auxiliary SST/SSH constraints, (iii) strategies to mitigate reanalysis-to-satellite distribution shift, and (iiiiv) operational sample selection and uncertainty calibration using independent in situ in-situ validation. With these steps, diffusion-based downscaling has the potential to enrich present SMOS/SMAPproducts with physically consistent /CCI products with GLORYS-consistent fine-scale structure, and to provide uncertainty-aware ensemble-based reconstructions that are directly usable for process studies and, ultimately, data assimilationdata-assimilation-oriented applications.

Code and data availability. GLORYS12V1 product accessed via https://data.marine.copernicus.eu/product/GLOBAL_MULTIYEAR_PHY_001_030/description. The diffusion implementation and code will be made available through the GitHub git@github.com:enzoforestier/Diff-Down-SSS.git, or upon direct request to the authors.

Appendix A: Denoising diffusion implicit models' architecture and gradient enhancement algorithm

We used a multi-channel U-Net that takes as input the stacked SSS, SST, and SSH fields and implements a standard Denoising diffusion implicit model (DDIM) procedure (Ho et al., 2020; Nichol and Dhariwal, 2021). Self-attention layers are included at the coarsest resolutions to capture long-range spatial dependencies, while convolutional residual blocks handle local interactions. This design provides a good balance between expressiveness and computational cost for high-resolution oceanographic images.

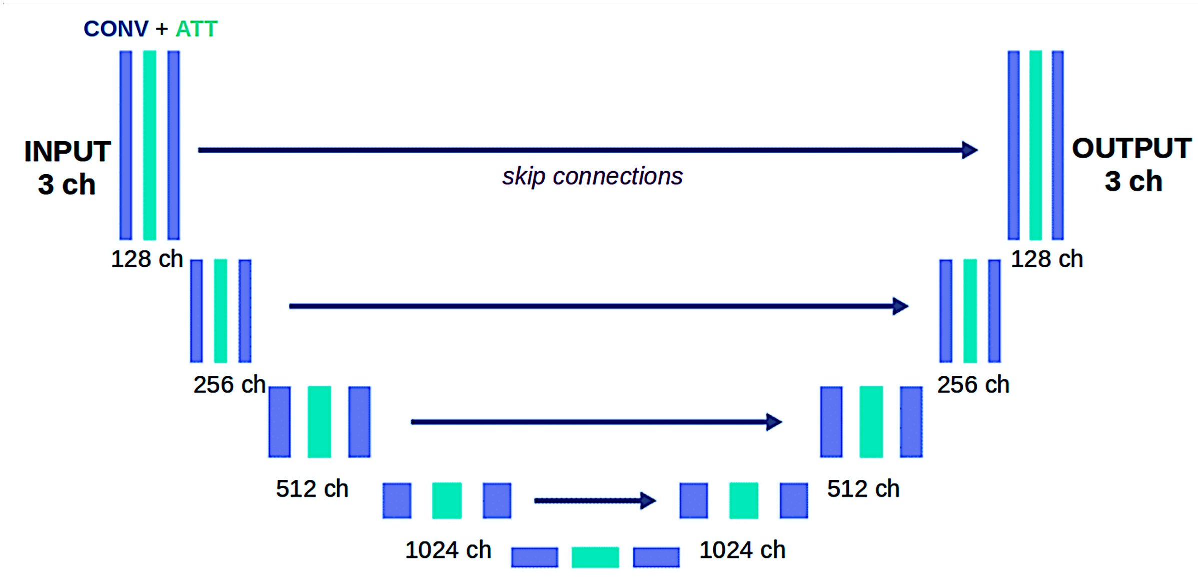

Figure A1. Architecture of the diffusion model, which uses a U-Net with skip connections, attention blocks (in green), and convolution blocks (in blue). The input and output generated images are 64×64 in size and have up to three channels for SSS, SST, and SSH depending of on the model configuration; the evaluated downscaling target is SSS.

Algorithm A1. Gradient-Enhancing (GE) Algorithm

Inputs: coarse SSS Y, observation operator H, pseudo-inverse $H^\dagger$, contrast parameters γ, δ (with $\gamma - \delta > 1$), ϵ-prediction diffusion model, time schedule $(v_i)_{i=0}^{N}$ from 0 to T.

Initialize $X \sim \mathcal{N}(0, I)$

for $i = T$ **to** 1 **do**

$t \leftarrow v_i, \quad s \leftarrow v_{i-1}$

$\alpha_t \leftarrow \dfrac{1}{1+\sigma_t^2}, \quad \alpha_s \leftarrow \dfrac{1}{1+\sigma_s^2}$

$c_1 \leftarrow \eta \sqrt{\left(1 - \dfrac{\alpha_t}{\alpha_s}\right) \dfrac{1-\alpha_s}{1-\alpha_t}}$

$c_2 \leftarrow \sqrt{1-\alpha_s - c_1^2}$

$\epsilon_\theta \leftarrow \epsilon\text{-prediction}(X,t)$

$\hat{X}_t \leftarrow \dfrac{X - \sqrt{1-\alpha_t}\,\epsilon_\theta}{\sqrt{\alpha_t}}$

if $t < N/4$ **then**

$\quad \mu \leftarrow \mathrm{MeanPool}_{2\times 2,2}(\hat{X}_t)$

$\quad g \leftarrow \left(\left(H^\dagger(Y) - H^\dagger(H(\hat{X}_t)) \right)^\top \dfrac{\partial \hat{X}_t}{\partial X} \right)^\top$

$\quad \hat{X}_t \leftarrow \hat{X}_t - \mu$

$\quad \hat{X}_t \leftarrow \left(\gamma - \delta \sqrt{\dfrac{|\hat{X}_t|}{\max|\hat{X}_t|}} \right) \hat{X}_t$

$\quad \hat{X}_t \leftarrow \hat{X}_t + \mu$

else

$\quad g \leftarrow 0$

end if

$\epsilon \sim \mathcal{N}(0,I)$

$X \leftarrow \sqrt{\alpha_s}\,\hat{X}_t + c_1\epsilon + c_2\epsilon_\theta + \sqrt{\alpha_t}\,g$

end for

Author contributions. S. T., E. F., L. O., R. EH., C. M., Conceptualization and methodology; E. F., S. T., R. EH., C. M., L. O., J. B., Formal analysis; E. F., Investigation; C. M., Resources; E. F., Data curation; E. F., S. T., R. EH., Writing - original draft; S. T., C. M., J. B., L. O., Writing - review & editing.

Competing interests. Authors declare no competing interests are present

Acknowledgements. We acknowledge support from the IPSL SAMA initiative (Statistics for Analysis, Modelling and Assimilation), the Sorbonne Center for Artificial Intelligence (SCAI), the Centre National Agence Nationale de la Recherche (CPJ – ANR-22-CPJ1-0003-01, Graduate school IFSEA – ANR-21-EXES-0011), and the Centre national d'Études Spatiales études spatiales (CNES)through the TOSCA APR-funded , France (ROR: https://ror.org/04h1h0y33), within the framework of the SMOS and SMOS-HR space missions (Soil Moisture and Ocean Salinity High Resolution Mission), and the funded SUMO project (Salinity Understanding through Integrated Models and Observations, 2025-2027), and the Agence Nationale de la Recherche (CPJ – ANR-22-CPJ1-0003-01, Graduate school IFSEA – ANR-21-EXES-0011).

References

Archer, M., Wang, J., Klein, P., Dibarboure, G., and Fu, L.-L.: Wide-swath satellite altimetry unveils global submesoscale ocean dynamics, Nature, 640, 691–696, 2025.

Boutin, J., Martin, N., Reverdin, G., Morisset, S., Yin, X., Centurioni, L., and Reul, N.: Sea surface salinity under rain cells: SMOS satellite and in situ drifters observations, Journal of Geophysical Research: Oceans, 119, 5533–5545, https://doi.org/10.1002/2014JC010070, 2014.

Boutin, J., Reul, N., Koehler, J., Martin, A., Catany, R., Guimbard, S., Rouffi, F., Vergely, J.-L., Arias, M., Chakroun, M., Corato, G., Gabia, G., Hasson, A., Josey, S. A., Supply, A., Thouvenin-Masson, C., Turiel, A., Vialard, J., and Waldteufel, P.: Satellite-Based Sea Surface Salinity Designed for Ocean and Climate Studies, Journal of Geophysical Research: Oceans, 126, e2021JC017 676, https://doi.org/10.1029/2021JC017676, 2021.

Chassignet, E. P. and Xu, X.: Impact of Horizontal Resolution ($1/12°$ to $1/50°$) on Gulf Stream Separation, Penetration, and Variability, Journal of Physical Oceanography, 47, 1999–2021, https://doi.org/10.1175/JPO-D-17-0031.1, 2017.

Chin, T. M., Vazquez-Cuervo, J., and Armstrong, E. M.: A multi-scale high-resolution analysis of global sea surface temperature, Remote Sensing of Environment, https://doi.org/10.1016/j.rse.2017.07.029, 2017.

Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C.: Diffusion Posterior Sampling for General Noisy Inverse Problems, https://arxiv.org/abs/2209.14687, 2024.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sensing of Environment, 116, 140–158, https://doi.org/10.1016/j.rse.2010.10.017, 2012.

Embury, O., Good, S., Merchant, C. J., et al.: Satellite-based time-series of sea-surface temperature since 1980 for climate applications, Scientific Data, https://doi.org/10.1038/s41597-024-03147-w, 2024.

Fablet, R., Chapron, B., Drumetz, L., Pannekoucke, O., Raynaud, R., and Ménard, R.: Joint interpolation and representation learning for irregularly sampled satellite-derived geophysical fields, Frontiers in Applied Mathematics and Statistics, 7, 655 224, https://doi.org/10.3389/fams.2021.655224, 2021.

Finn, T. S., Durand, C., Farchi, A., Bocquet, M., Rampal, P., and Carrassi, A.: Generative diffusion for regional surrogate models from sea-ice simulations, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 395, https://doi.org/10.1029/2024MS004395, 2024.

Ho, J., Jain, A., and Abbeel, P.: Denoising diffusion probabilistic models, in: Advances in Neural Information Processing Systems, vol. 33, pp. 6840–6851, https://doi.org/10.48550/arXiv.2006.11239, 2020.

Joyce, T. M., Kwon, Y.-O., Seo, H., and Ummenhofer, C. C.: Meridional Gulf Stream Shifts Can Influence Wintertime Variability in the North Atlantic Storm Track and Greenland Blocking, Geophysical Research Letters, 46, 1702–1708, https://doi.org/10.1029/2018GL081087, 2019a.

Joyce, T. M., Kwon, Y.-O., Seo, H., and Ummenhofer, C. C.: Meridional Gulf Stream Shifts Can Influence Wintertime Variability in the North Atlantic Storm Track and Greenland Blocking, Geophysical Research Letters, 46, 1702–1708, https://doi.org/10.1029/2018GL081087, 2019b.

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., et al.: The SMOS mission: New tool for monitoring key elements ofthe global water cycle, Proceedings of the IEEE, 98, 666–687, 2010.

Lagerloef, G., Colomb, F. R., Le Vine, D., Wentz, F., Yueh, S., Ruf, C., et al.: The Aquarius/SAC-D mission: Designed to meet the salinity remote-sensing challenge, Oceanography, 21, 68–81, https://doi.org/10.5670/oceanog.2008.68, 2008.

Le Traon, P. Y., Reppucci, A., Alvarez Fanjul, E., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando, V. E., Kreiner, M. B., Benkiran, M., Carval, T., Ciliberti, S. A., Claustre, H., Clementi, E., Coppini, G., Cossarini, G., De Alfonso Alonso-Muñoyerro, M., Delamarche, A., Dibarboure, G., Dinessen, F., Drevillon, M., Drillet, Y., Faugere, Y., Fernández, V., Fleming, A., Garcia-Hermosa, M. I., Sotillo, M. G., Garric, G., Gasparin, F., Giordan, C., Gehlen, M., Gregoire, M. L., Guinehut, S., Hamon, M., Harris, C., Hernandez, F., Hinkler, J. B., Hoyer, J., Karvonen, J., Kay, S., King, R., Lavergne, T., Lemieux-Dudon, B., Lima, L., Mao, C., Martin, M. J., Masina, S., Melet, A., Buongiorno Nardelli, B., Nolan, G., Pascual, A., Pistoia, J., Palazov, A., Piolle, J. F., Pujol, M. I., Pequignet, A. C., Peneva, E., Pérez Gómez, B., Petit de la Villeon, L., Pinardi, N., Pisano, A., Pouliquen, S., Reid, R., Remy, E., Santoleri, R., Siddorn, J., She, J., Staneva, J., Stoffelen, A., Tonani, M., Vandenbulcke, L., von Schuckmann, K., Volpe, G., Wettre, C., and Zacharioudaki, A.: From Observation to Information and Users: The Copernicus Marine Service Perspective, Frontiers in Marine Science, Volume 6 - 2019, https://doi.org/10.3389/fmars.2019.00234, 2019.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 105–114, https://doi.org/10.1109/CVPR.2017.19, 2017.

Lellouche, J., Greiner, E., Bourdalle-Badie, R., Garric, G., Melet, A., Drevillon, M., Bricaud, C., Hamon, M., Levier, B., Regnier, C., et al.: The Copernicus Global $1/12°$ Oceanic and Sea Ice GLORYS12 Reanalysis, Frontiers in Earth Science, 9, 698 876, https://doi.org/10.3389/feart.2021.698876, 2021.

Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C., et al.: Generative residual diffusion modeling for km-scale atmospheric downscaling, arXiv preprint, arXiv:2309.15214, https://doi.org/10.48550/arXiv.2309.15214, 2023.

Morrow, R., Fu, L.-L., Ardhuin, F., Benkiran, M., Chapron, B., Cosme, E., d'Ovidio, F., Farrar, J. T., Gille, S. T., Lapeyre, G., Le Traon, P.-Y., Pascual, A., Ponte, A., Qiu, B., Rascle, N., Ubelmann, C., Wang, J., and Zaron, E. D.: Global Observations of Fine-Scale Ocean Surface Topography With the Surface Water and Ocean Topography (SWOT) Mission, Frontiers in Marine Science, Volume 6 - 2019, https://doi.org/10.3389/fmars.2019.00232, 2019.

Nichol, A. Q. and Dhariwal, P.: Improved Denoising Diffusion Probabilistic Models, arXiv preprint, arXiv:2102.09672, https://doi.org/10.48550/arXiv.2102.09672, 2021.

Olmedo, E., Martínez, J., Umbert, M., Hoareau, N., Portabella, M., Ballabrera-Poy, J., and Turiel, A.: Improving time and space resolution of SMOS salinity maps using multifractal fusion, Remote Sensing of Environment, 180, 246–263, https://doi.org/10.1016/j.rse.2016.02.038, special Issue: ESA's Soil Moisture and Ocean Salinity Mission - Achievements and Applications, 2016.

Piepmeier, J. R., Focardi, P., Horgan, K. A., Knuble, J., Ehsan, N., Lucey, J., Brambora, C., Brown, P. R., Hoffman, P. J., French, R. T., et al.: SMAP L-band microwave radiometer: Instrument design and first year on orbit, IEEE Transactions on Geoscience and Remote Sensing, 55, 1954–1966, 2017.

Reul, N., Grodsky, S., Arias, M., Boutin, J., Catany, R., Chapron, B., et al.: Sea surface salinity estimates from spaceborne L-band radiometers: An overview of the first decade of observation (2010–2019), Remote Sensing of Environment, 242, 111 769, https://doi.org/10.1016/j.rse.2020.111769, 2020.

Reverdin, G., Supply, A., Boutin, J., and Yin, X.: Missing Argo float profiles in highly stratified waters of the Amazon River plume, Journal of Atmospheric and Oceanic Technology, 41, 221–233, https://doi.org/10.1175/JTECH-D-23-0072.1, 2024.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B.: High-resolution image synthesis with latent diffusion models, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10 684–10 695, https://doi.org/10.1109/CVPR52688.2022.01042, 2022.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by Navab, N., Hornegger, J., Wells, W. M., and Frangi, A., vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241, Springer, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Song, J., Vahdat, A., Mardani, M., and Kautz, J.: Pseudoinverse-Guided Diffusion Models for Inverse Problems, in: International Conference on Learning Representations, https://openreview.net/forum?id=9_gsMA8MRKQ, 2023.

Thiria, S., Sorror, C., Archambault, T., Charantonis, A., Béreziat, D., Mejia, C., et al.: Downscaling of ocean fields by fusion of heterogeneous observations using deep learning algorithms, Ocean Modelling, 182, 102 174, https://doi.org/10.1016/j.ocemod.2023.102174, 2023.

Tomasi, E., Franch, G., and Cristoforetti, M.: Can AI be enabled to perform dynamical downscaling? A latent diffusion model to mimic kilometer-scale COSMO5.0_CLM9 simulations, Geoscientific Model Development, 18, 2051–2078, https://doi.org/10.5194/gmd-18-2051-2025, 2025.

Vinogradova, N., Lee, T., Boutin, J., Drushka, K., Fournier, S., Sabia, R., et al.: Satellite salinity observing system: Recent discoveries and the way forward, Frontiers in Marine Science, 6, 243, https://doi.org/10.3389/fmars.2019.00243, 2019.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P.: Image Quality Assessment: From Error Visibility to Structural Similarity, IEEE Transactions on Image Processing, 13, 600–612, https://doi.org/10.1109/TIP.2003.819861, 2004.

Watt, R. A. and Mansfield, L. A.: Generative diffusion-based downscaling for climate, arXiv preprint, arXiv:2404.17752, https://doi.org/10.48550/arXiv.2404.17752, 2024.

Wu, R., Jia, L., Li, C., Liu, Y., Han, B., and Chen, D.: Impact of Horizontal Resolution (Submesoscale Permitting vs. Mesoscale Resolving) on Ocean Dynamic Features in the South China Sea, Earth and Space Science, 9, e2022EA002 448, https://doi.org/10.1029/2022EA002448, 2022.

Zhai, X., Johnson, H. L., and Marshall, D. P.: On the Seasonal Variability of Eddy Kinetic Energy in the Gulf Stream Region, Geophysical Research Letters, 35, L24 609, https://doi.org/10.1029/2008GL036412, 2008.