Supplementary Information for " High-resolution probabilistic estimation of three-dimensional regional ocean dynamics from sparse surface observations"

Niloofar Asefi[2] Tianning Wu[3] Ruoying He[3]
Ashesh Chattopadhyay[1]

[1]Department of Applied Mathematics, University of California, Santa Cruz, CA 95064, USA
[2]Department of Electrical and Computer Engineering, University of California, Santa Cruz, CA 95064, USA
[3]Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA

aschatto@ucsc.edu

1 Analysis at different depths and comparison with baselines

Here, we provide additional analysis of our proposed models as compared with baselines for different depths.

1.1 Reconstruction of subsurface ocean states from sparse surface observations

Figure S1 and Figure S2 illustrate the depth-aware DDPM reconstruction of subsurface states, specifically temperature (T), salinity (S), and zonal/meridional velocities (U, V). These snapshots correspond to representative depths of 25 m, 109 m, and 155 m in Figure S1, and 222 m, 453.9 m, and 643 m in Figure S2. Despite the extremely sparse conditioning data, the DDPM reconstruction preserves the primary spatial patterns and depth-dependent structures of the subsurface fields with just small deviations from the GLORYS reanalysis reference. These findings are consistent across all evaluated depths; notably, even at levels as deep as 643 m, the DDPM model successfully reconstructs the full-resolution of variables with high fidelity.

1.2 Quantitative and physical evaluation of reconstruction performance

Figure S3 and Figure S4 show the mean Fourier power spectra of temperature, salinity, and zonal and meridional velocities (U, V) computed across 100 test samples at corresponding depths of 25 m, 109 m, and 155 m in Figure S3, and 222 m, 453.9 m, and 643 m in Figure S4. The depth-aware DDPM captures precise temperature and salinity spectra across all wavenumbers, showing successful reconstruction of both large-scale and small-scale features, even at the deepest level (643 m). In contrast, velocity components are more sensitive to small-scale variability, resulting in slight variations at higher wavenumbers. Overall, these findings show that the depth-aware DDPM can effectively recover multi-scale dynamical structures over a variety of subsurface variables and depths.

The depth-aware DDPM accurately captures the temperature and salinity spectra across the entire wavenumber range, demonstrating a robust reconstruction of both large-scale and small-scale features—even at the deepest level of 643 m. While the velocity components exhibit greater sensitivity to small-scale variability, leading to minor deviations at higher wavenumbers, the model consistently recovers the fundamental dynamical structures. Overall, these results are consistent with the main paper's results and confirm that the depth-aware DDPM effectively reconstructs multi-scale features across diverse subsurface variables and depth levels.

Figure S5 and Figure S6 demonstrate the reconstruction performance for temperature, salinity, zonal velocity (U), and meridional velocity (V) across 100 test samples. Columns correspond to three evaluation metrics—normalized RMSE (NRMSE), correlation coefficient (CC), and structural similarity index (SSIM)—evaluated at corresponding depths of 25 m, 109 m, and 155 m in Figure S5, and 222 m, 453.9 m, and 643 m in Figure S6. NRMSE was specifically selected to facilitate a standardized comparison across variables with differing physical units and magnitudes. Regarding NRMSE, the hybrid-conditioned variants, i.e., UNet+DDPM and FNO+DDPM, exhibit slightly lower error than the standalone depth-aware DDPM, though this improvement is marginal. For the CC and SSIM metrics, all three models perform with nearly identical magnitude. In summary, incorporating pre-trained deterministic models (UNet, FNO) as conditioning components does not yield a substantial performance gain, suggesting that the depth-aware DDPM alone is capable of achieving highly competitive results.

Figure S7, Figure S8, and Figure S9 provide a comparative analysis of the reconstructions generated by UNet+DDPM, FNO+DDPM, and the depth-aware DDPM at depths of 55 m, 318 m, and 1062 m, respectively. While the hybrid-conditioned variants—UNet+DDPM and FNO+DDPM—recover a greater density of fine-scale features, a portion of these structures appear to be artifacts. Overall, however, all three models demonstrate a robust ability to reconstruct subsurface states accurately across the entire vertical water column.

Figure S10 demonstrates the reconstructions of subsurface ocean variables produced by the depth-aware FNO and UNET base models at depths of 55 m and 1062 m. Overall, the depth-aware FNO model provides more accurate reconstructions than the depth-aware UNET model, particularly for temperature and salinity at 55 m depth. For the zonal and meridional velocity components, the depth-aware UNET model performs poorly and exhibits strong off-scale artifacts in some regions. Overall, these results indicate that deterministic base models are limited in their ability to reconstruct complex three-dimensional ocean structures, whereas the probabilistic DDPM framework provides more accurate and physically consistent reconstructions

Figure S11 shows the averaged Fourier power spectra of subsurface ocean variables for the base models at 55 m and 1062 m depths. The mean spectra of temperature, salinity, zonal velocity (U), and meridional velocity (V), computed over 100 test samples, are shown for the depth-aware FNO and UNET base models and compared with the corresponding ground-truth fields from the GLORYS reanalysis dataset. Each column corresponds to a depth level (55 m and 1062 m), and each row corresponds to one variable. For temperature and salinity, both base models closely match the ground-truth spectra across most wavenumbers. For the zonal and meridional velocity components (U, V), the depth-aware FNO model more accurately captures the spectral distribution than the depth-aware UNET model.

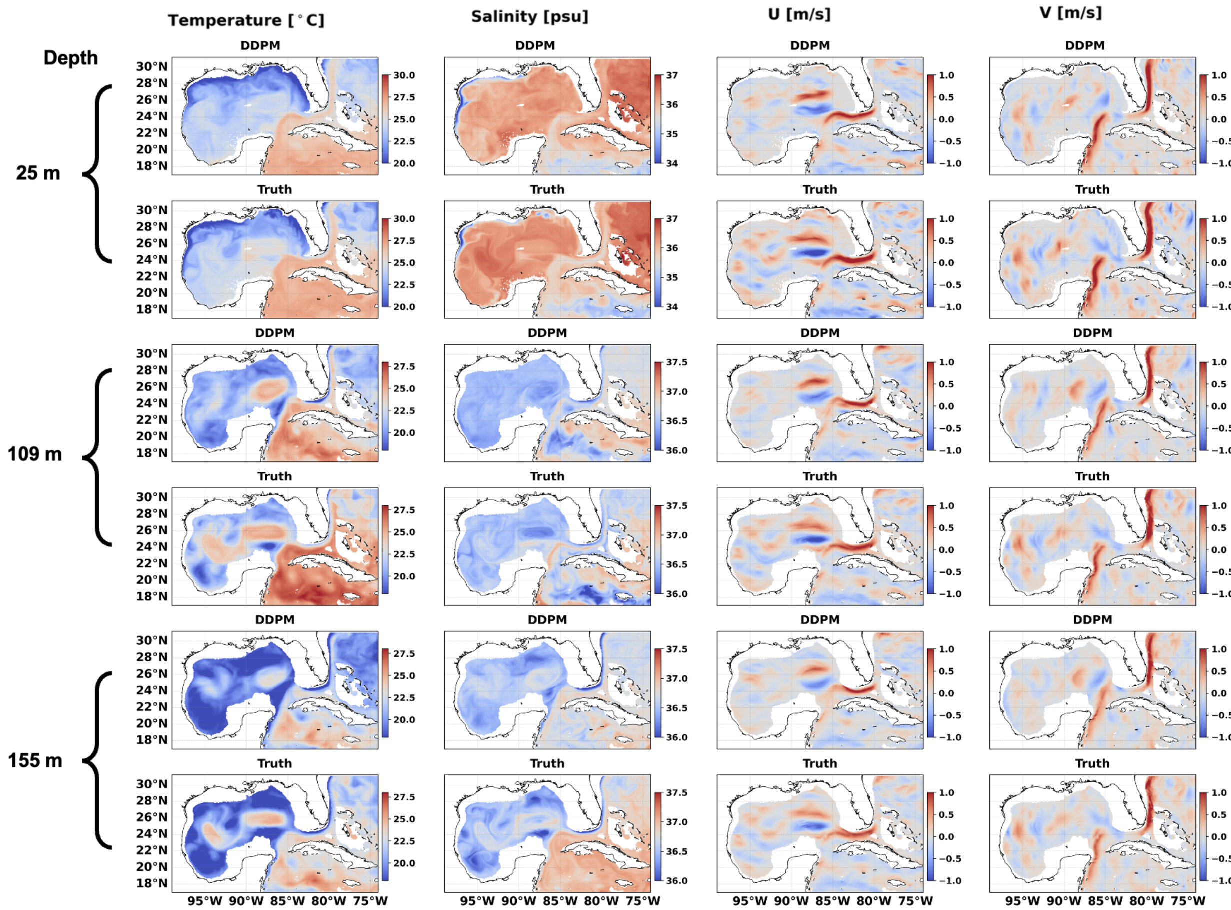

Figure S1: Depth-aware DDPM reconstruction of subsurface ocean variables at three depths. Subsurface ocean variables, including temperature, salinity, zonal velocity (U), and meridional velocity (V), are shown over the Gulf of Mexico at three different depths: 25 m, 109 m, and 155 m. The DDPM reconstructions (top row) are compared with the corresponding ground-truth fields from the GLORYS reanalysis (bottom row) for each depth. White areas indicate masked or land areas. As depth increases, the valid ocean domain decreases due to bathymetric constraints, resulting in a larger land mask at deeper levels.

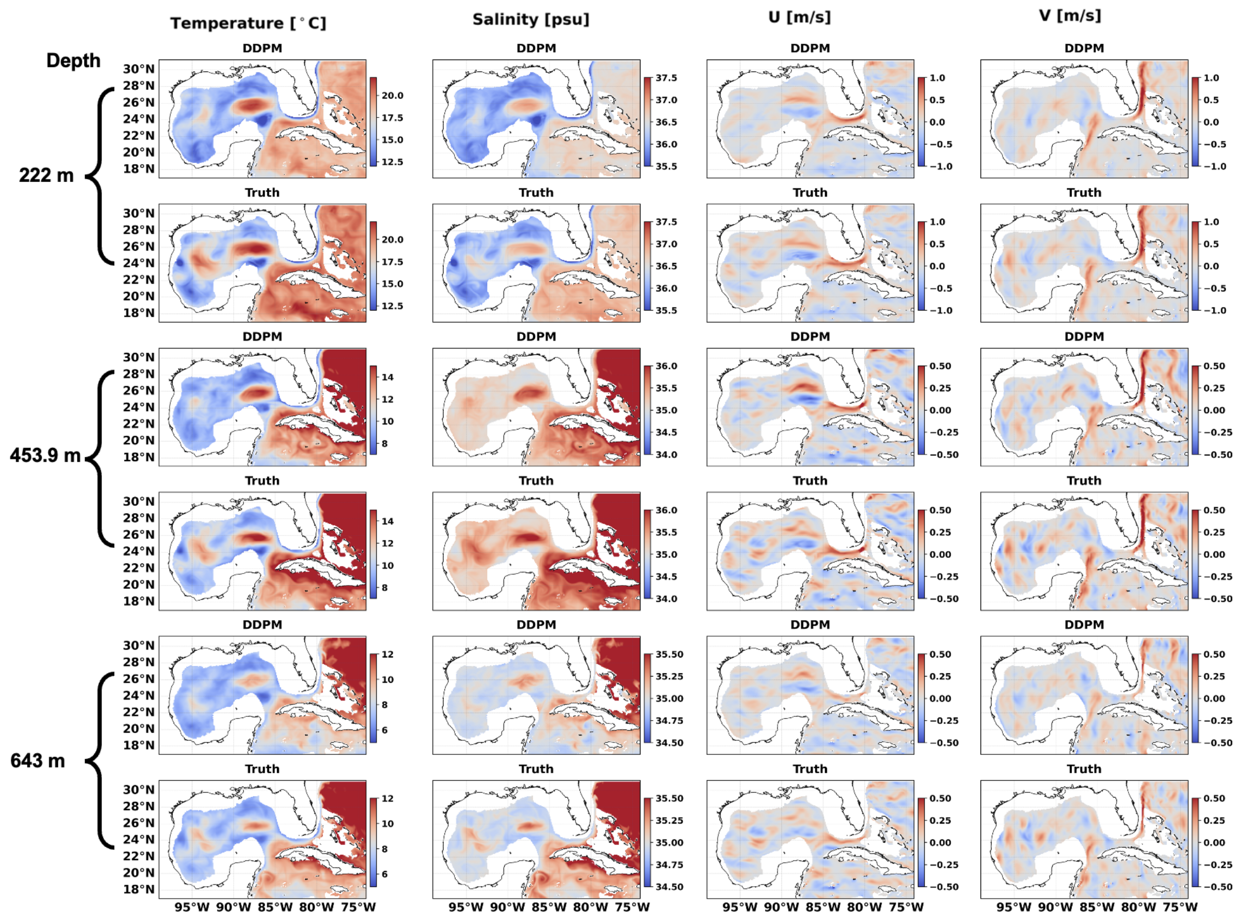

Figure S2: Depth-aware DDPM reconstruction of subsurface ocean variables at three depths. Subsurface ocean variables, including temperature, salinity, zonal velocity (U), and meridional velocity (V), are shown over the Gulf of Mexico at three different depths: 222 m, 453.9 m, and 643 m. The DDPM reconstructions (top row) are compared with the corresponding ground-truth fields from the GLORYS reanalysis (bottom row) for each depth. White areas indicate masked or land areas. As depth increases, the valid ocean domain decreases due to bathymetric constraints, resulting in a larger land mask at deeper levels.

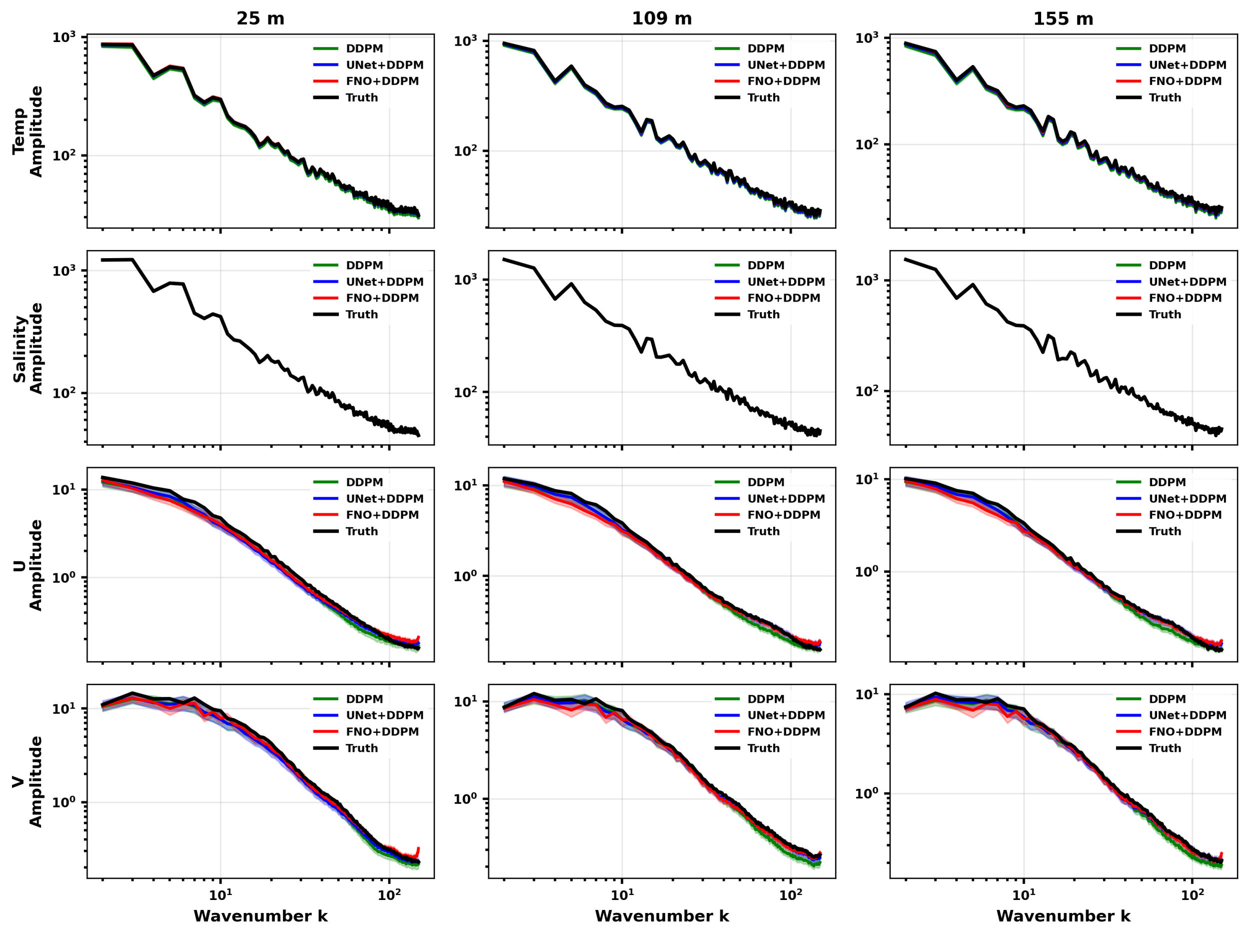

Figure S3: Averaged fourier power spectra for subsurface ocean variables at three depths. The power spectra of temperature, salinity, zonal velocity (U), and meridional velocity (V) is displayed for 100 test samples at three different depths: 25 m, 109 m, and 155 m. Power spectra from the DDPM reconstructions are compared with the corresponding GLORYS reanalysis reference across spatial wavenumbers.

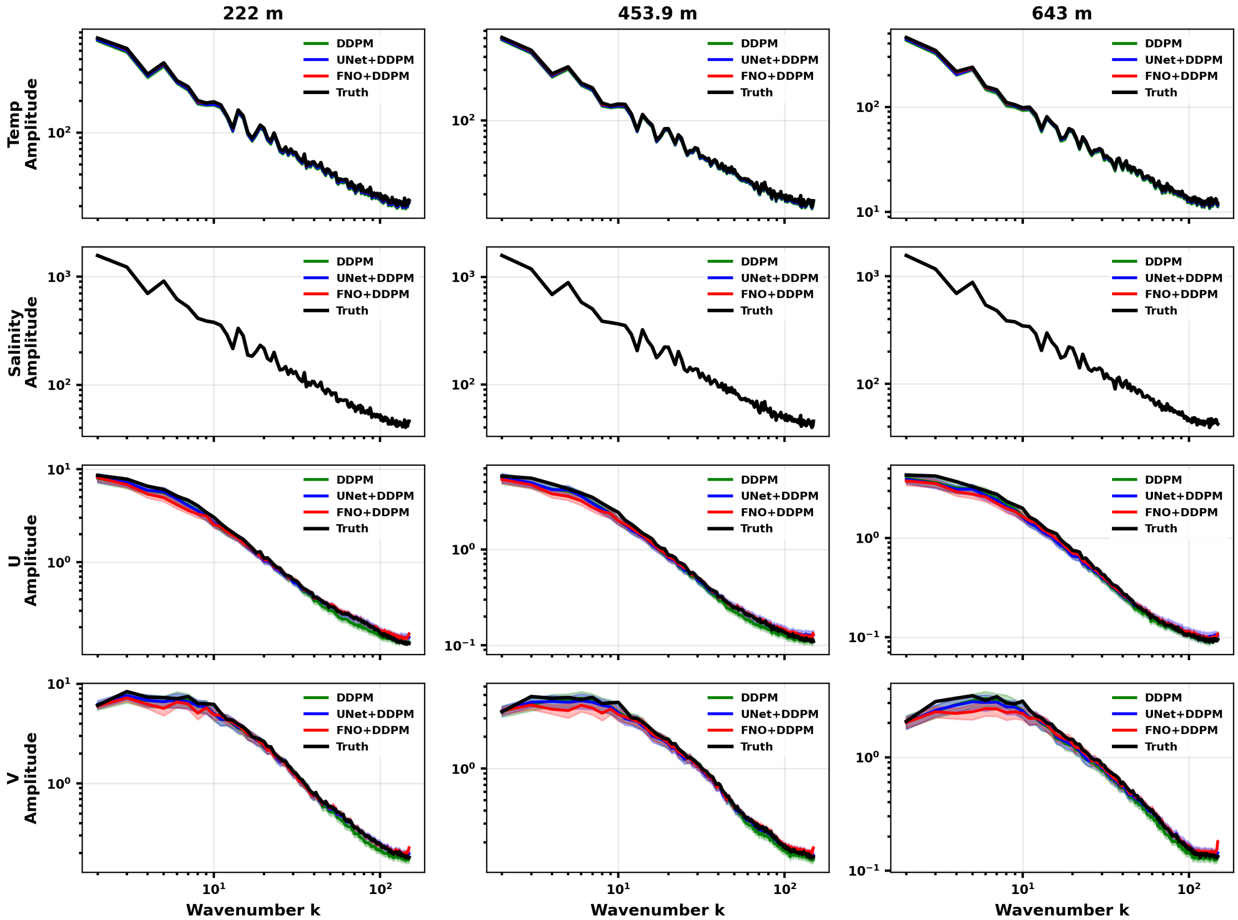

Figure S4: Averaged fourier power spectra for subsurface ocean variables at three depths. The power spectra of temperature, salinity, zonal velocity (U), and meridional velocity (V) is displayed for 100 test samples at three different depths: 222 m, 453.9 m, and 643 m. Power spectra from the DDPM reconstructions are compared with the corresponding GLORYS reanalysis reference across spatial wavenumbers.

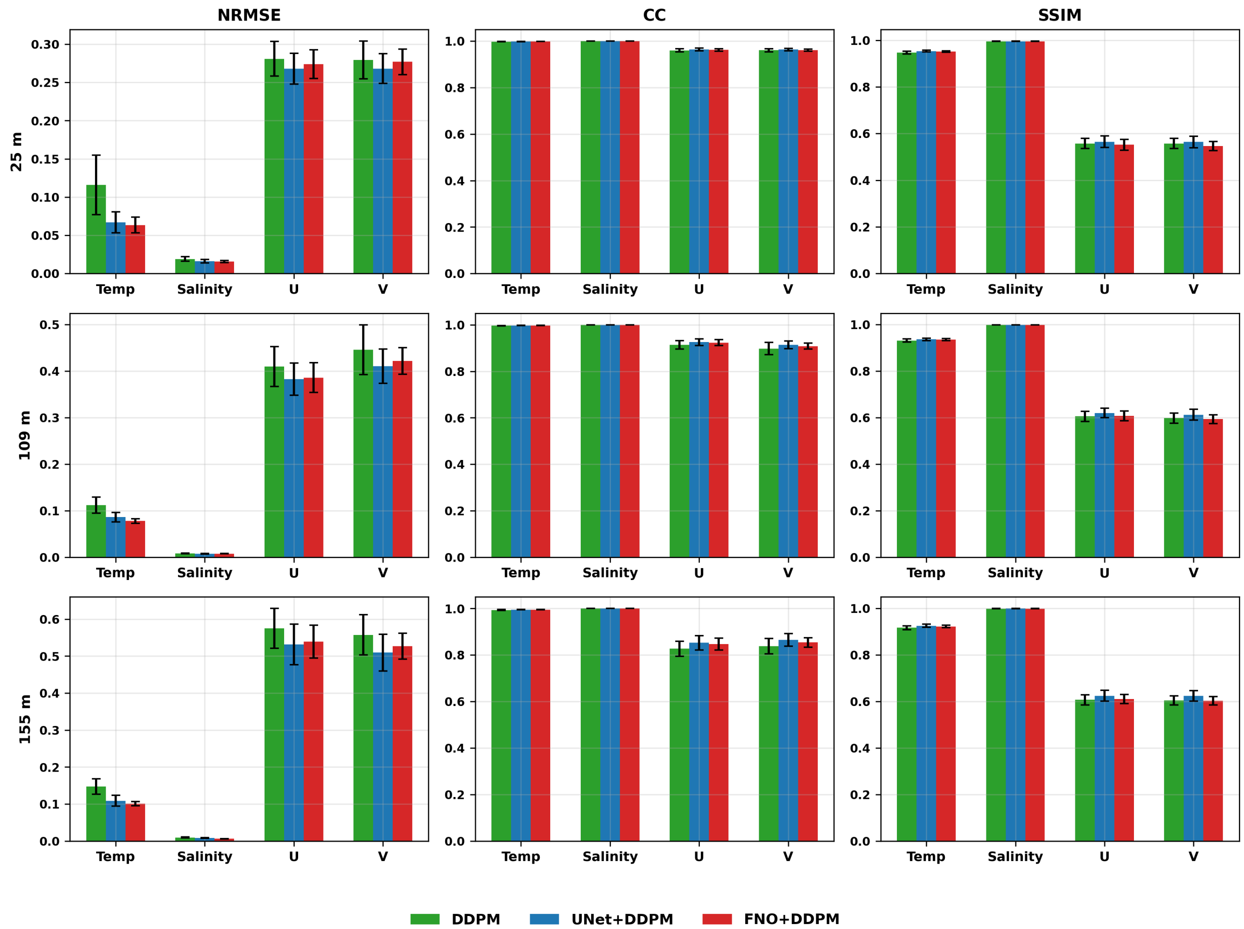

Figure S5: Quantitative evaluation of reconstructed subsurface variables across depths and models. Reconstruction performance for temperature, salinity, zonal velocity (U), and meridional velocity (V) across 100 test samples. Columns correspond to various evaluation metrics, including normalized RMSE (NRMSE), correlation coefficient (CC), and SSIM. Results are compared across three models: DDPM, UNet+DDPM, and FNO+DDPM. Rows show results at three depths: 25 m, 109 m, and 155 m.

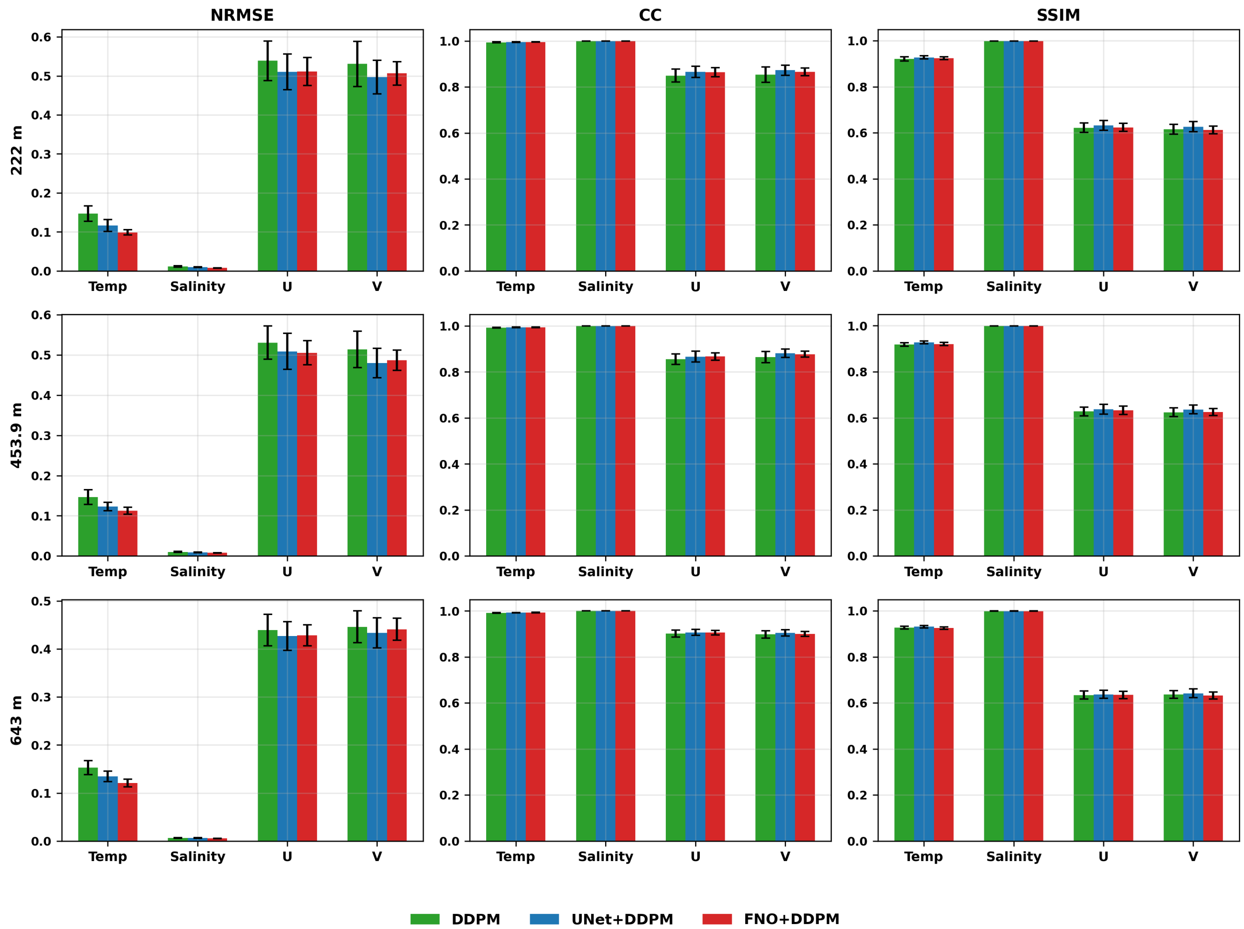

Figure S6: Quantitative evaluation of reconstructed subsurface variables across depths and models. Reconstruction performance for temperature, salinity, zonal velocity (U), and meridional velocity (V) across 100 test samples. Columns correspond to various evaluation metrics, including normalized RMSE (NRMSE), correlation coefficient (CC), and SSIM. Results are compared across three models: DDPM, UNet+DDPM, and FNO+DDPM. Rows show results at three depths: 222 m, 453.9 m, and 643 m.

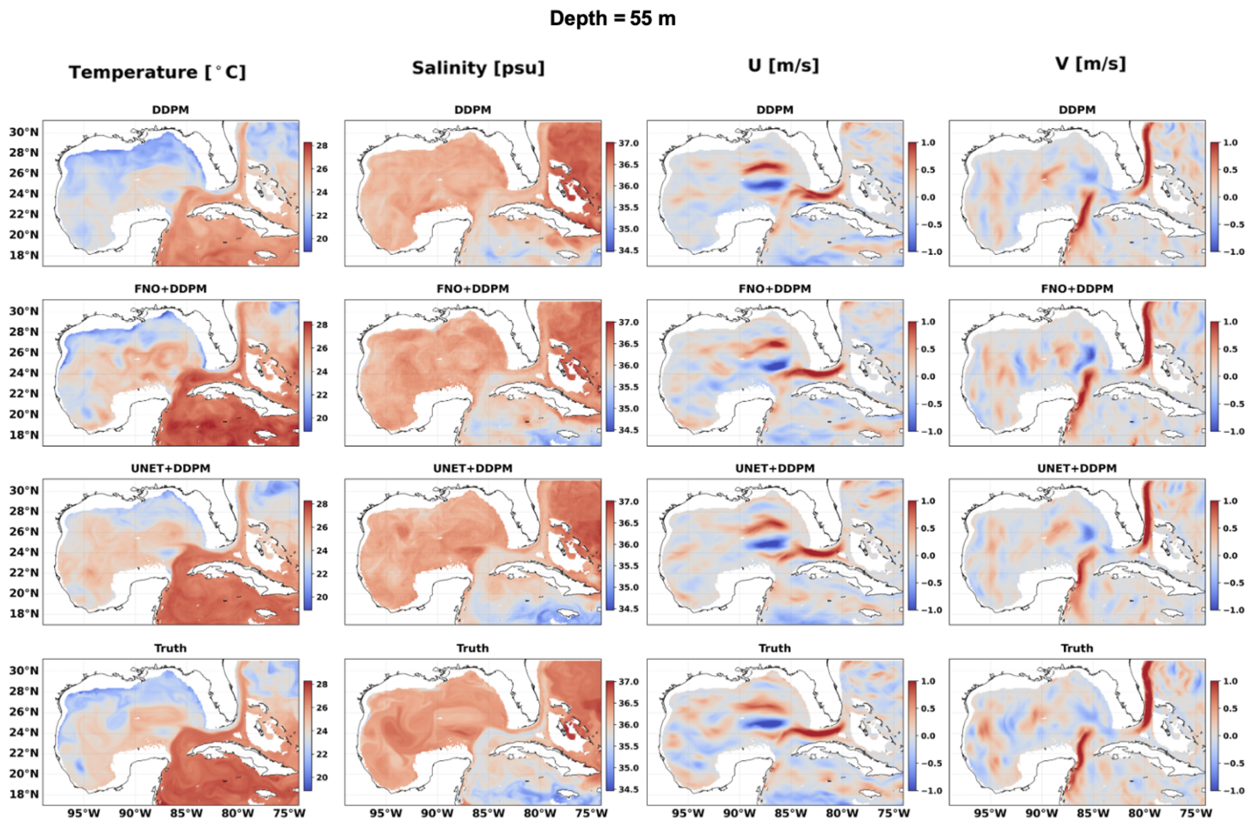

Figure S7: Comparative analysis of model reconstructions at 55 m depth. Subsurface ocean variables, including temperature, salinity, zonal velocity (U), and meridional velocity (V), are shown over the Gulf of Mexico at depth 55 m. Reconstructions from the DDPM, FNO+DDPM, and UNET+DDPM models are compared with the corresponding ground-truth fields from the GLORYS reanalysis dataset. White regions denote land or masked areas.

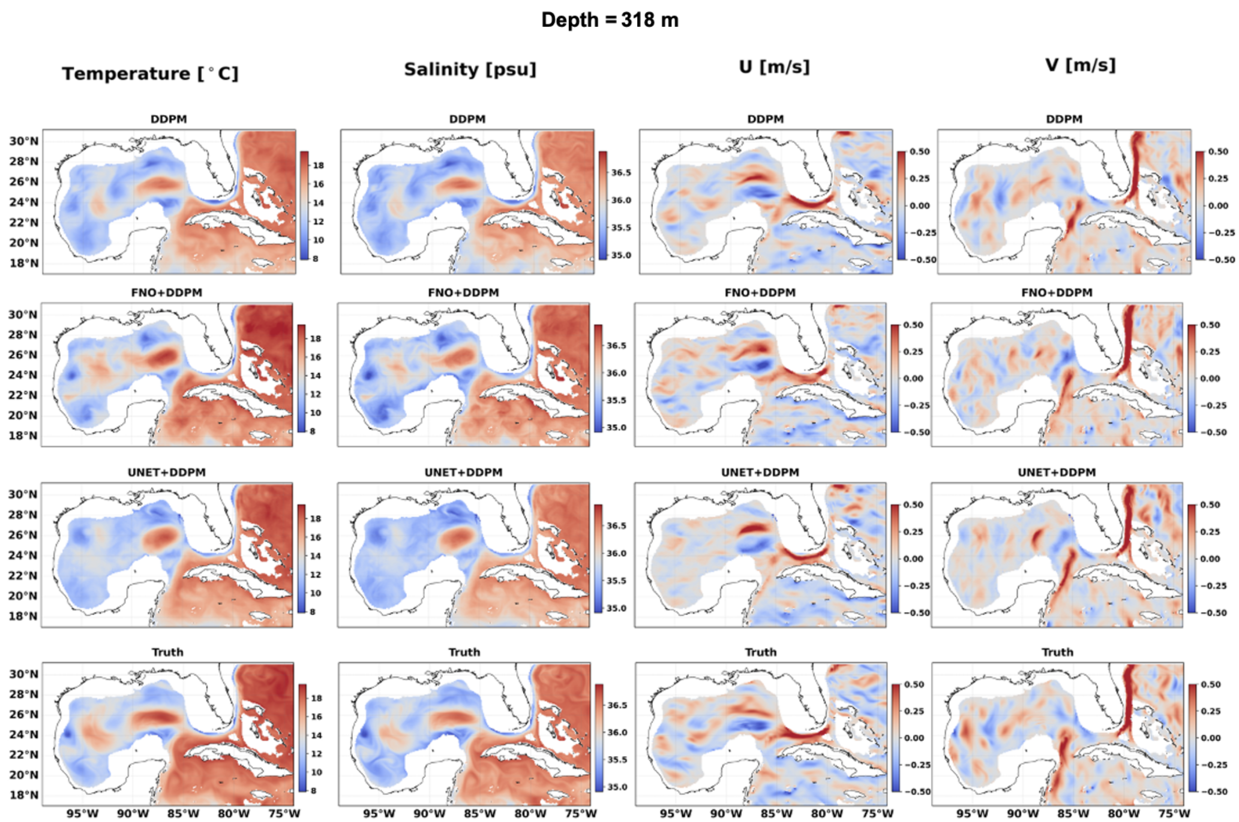

Figure S8: Comparative analysis of model reconstructions at 318 m depth. Subsurface ocean variables, including temperature, salinity, zonal velocity (U), and meridional velocity (V), are shown over the Gulf of Mexico at depth 318 m. Reconstructions from the DDPM, FNO+DDPM, and UNET+DDPM models are compared with the corresponding ground-truth fields from the GLORYS reanalysis dataset. White regions denote land or masked areas.

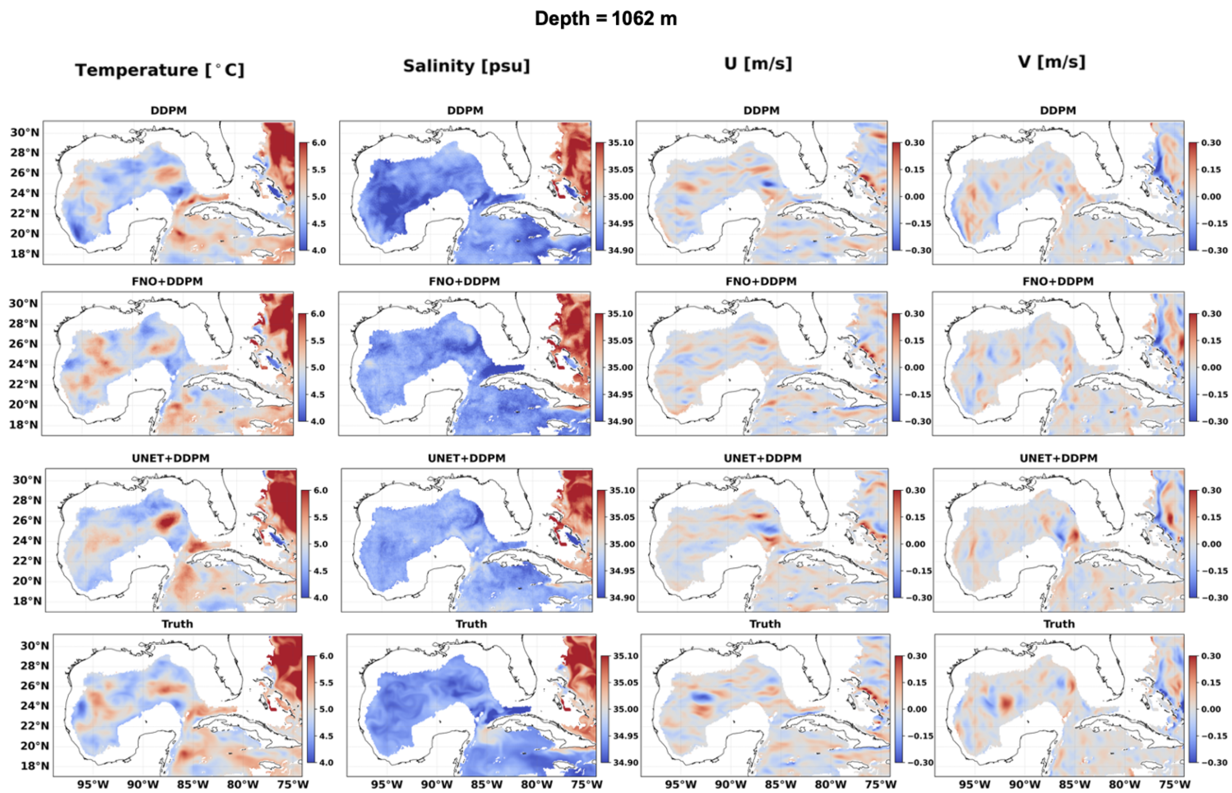

Figure S9: Comparative analysis of model reconstructions at 1062 m depth. Subsurface ocean variables, including temperature, salinity, zonal velocity (U), and meridional velocity (V), are shown over the Gulf of Mexico at depth 1062 m. Reconstructions from the DDPM, FNO+DDPM, and UNET+DDPM models are compared with the corresponding ground-truth fields from the GLORYS reanalysis dataset. White regions denote land or masked areas.

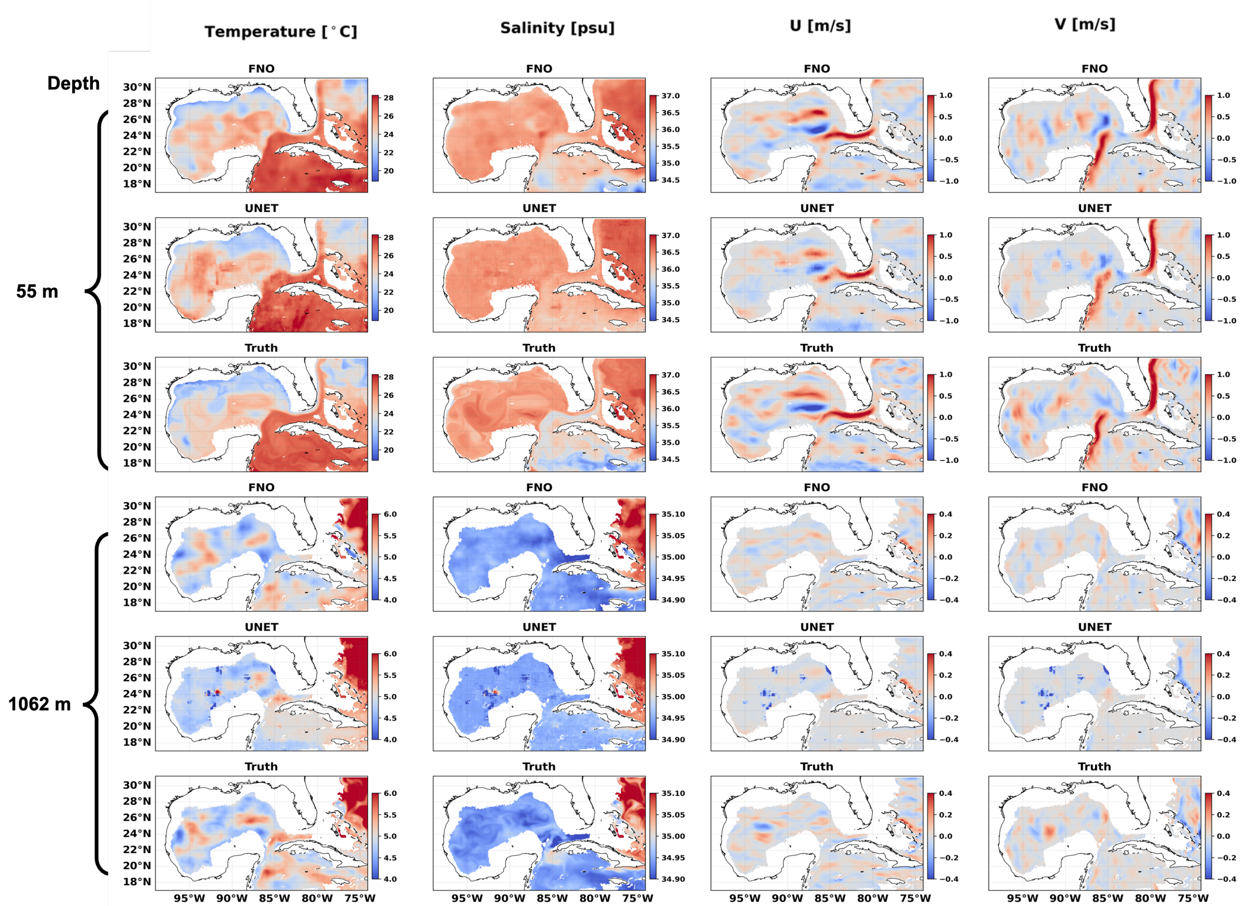

Figure S10: Comparison of depth-aware base-model reconstructions at 55 m and 1062 m depths. Subsurface ocean variables, including temperature, salinity, zonal velocity (U), and meridional velocity (V), are shown over the Gulf of Mexico at depths of 55 m and 1062 m. Reconstructions from the depth-aware FNO and UNET base models are compared with the corresponding ground-truth fields from the GLORYS reanalysis dataset. White regions indicate land or masked areas.

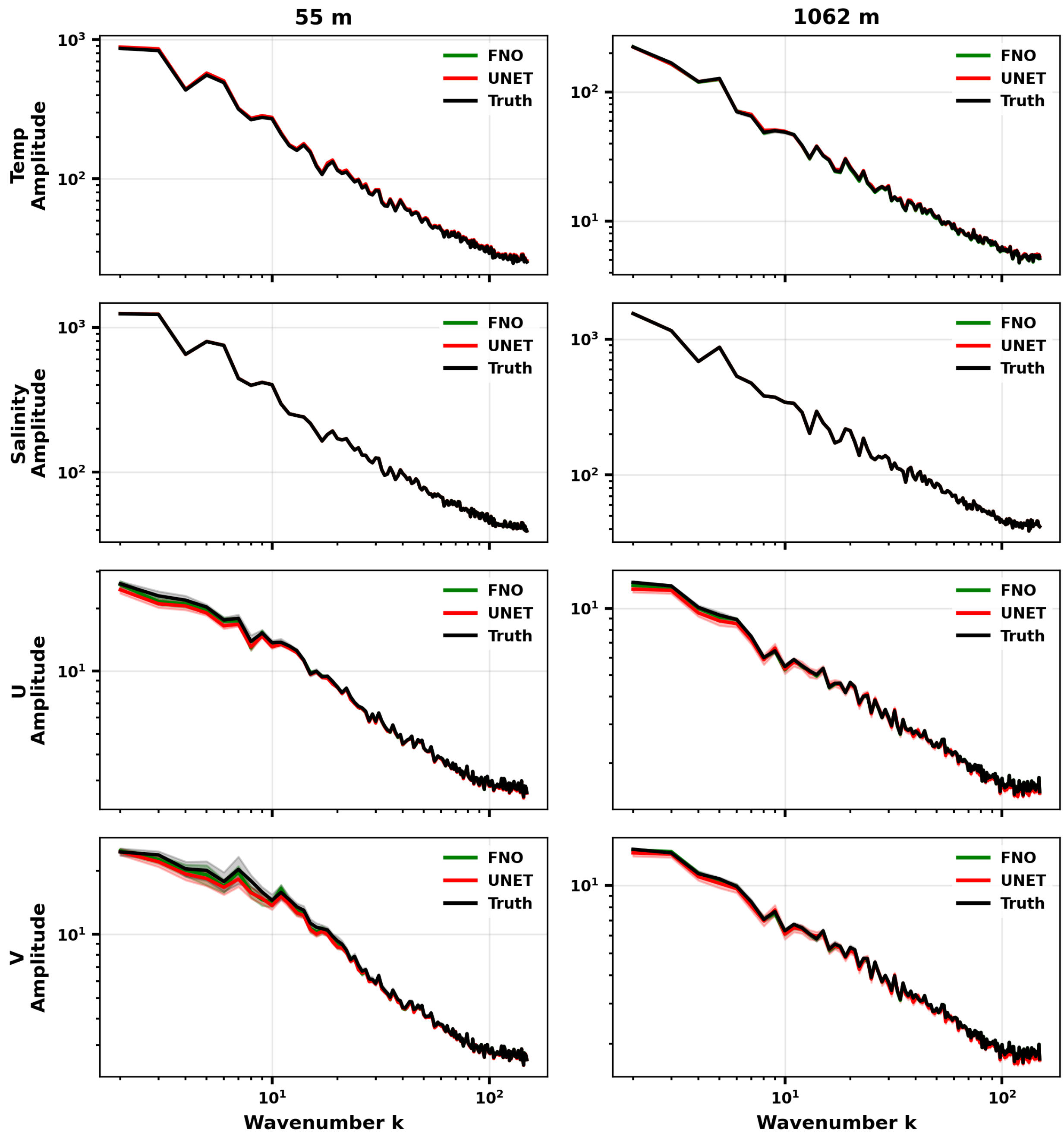

Figure S11: Averaged Fourier power spectra of subsurface ocean variables for the depth-aware base models at 55 m and 1062 m depths. Mean Fourier power spectra of temperature, salinity, zonal velocity (U), and meridional velocity (V) over 100 test samples are shown for the depth-aware FNO and UNET base models and compared against the corresponding ground-truth fields from the GLORYS reanalysis dataset. Each column corresponds to a depth level (55 m and 1062 m), and each row corresponds to one variable. White regions denote land or masked areas.