scientific reports

OPEN

OceanNet: a principled neural operator-based digital twin for regional oceans

Ashesh Chattopadhyay[1✉], Michael Gray[2], Tianning Wu[2], Anna B. Lowe[2] & Ruoying He[2✉]

While data-driven approaches demonstrate great potential in atmospheric modeling and weather forecasting, ocean modeling poses distinct challenges due to complex bathymetry, land, vertical structure, and flow non-linearity. This study introduces OceanNet, a principled neural operator-based digital twin for regional sea-suface height emulation. OceanNet uses a Fourier neural operator and predictor-evaluate-corrector integration scheme to mitigate autoregressive error growth and enhance stability over extended time scales. A spectral regularizer counteracts spectral bias at smaller scales. OceanNet is applied to the northwest Atlantic Ocean western boundary current (the Gulf Stream), focusing on the task of seasonal prediction for Loop Current eddies and the Gulf Stream meander. Trained using historical sea surface height (SSH) data, OceanNet demonstrates competitive forecast skill compared to a state-of-the-art dynamical ocean model forecast, reducing computation by 500,000 times. These accomplishments demonstrate initial steps for physics-inspired deep neural operators as cost-effective alternatives to high-resolution numerical ocean models.

Keywords Data-driven model, Ocean forecasting, Neural operator, Spectral bias

An essential component of the Earth's ocean circulation is the northern Atlantic Ocean's western boundary current (WBC) system, which includes the Loop Current (LC), Loop Current eddies (LCE), the Gulf Stream (GS), and Gulf Stream meander (GSM). The WBC plays a critical role in the transport of heat, salt, and nutrients, and strongly influences global weather, climate, and marine ecosystems. Recent studies have shown the promise of using data-driven approaches to model and predict atmospheric circulation, but complexities in the ocean system present additional challenges when building data-driven models[1–3]. One challenge is the complex land boundaries, such as in the Gulf of Mexico (GoM). Equally difficult to model, the GS frequently sheds eddies as a result of nonlinear flow and the interactions of cool, subpolar circulation from the north and warm, subtropical circulation from the south. Predicting LCEs in the GoM and GS meander and eddy separation have been long-standing challenges for numerical ocean models[4–6]. The separation mechanism involves nonlinear barotropic and baroclinic instabilities of ocean currents, complex flow-topography interactions, and, in the case of the GS, interaction with the Deep Western Boundary Current[7–10]. These challenging features require carefully calibrated, high-resolution numerical models and tremendous computing resources to resolve[11–13].

Recent efforts in emulating ocean dynamics with deep learning-based approaches have primarily focused on predicting large-scale circulation features, such as those resolved by empirical orthogonal functions (EOFs)[14] or on constructing low-dimensional representations[15]. However, there has yet to be a comprehensive data-driven model for the ocean akin to those used in global atmospheric modeling.

Data-driven modeling in Earth sciences has focused on predicting global atmospheric dynamics[1–3,16,17]; weather prediction has been used as a benchmark. The primary advantage of data-driven models lies in their computational efficiency, enabling the execution of a large number of ensembles. This facilitates seamless and efficient data assimilation leading to improved estimates of extreme weather events. Several recent studies[1–3] have demonstrated that these data-driven models can achieve accuracy comparable to, and often surpassing, state-of-the-art numerical weather prediction models such as the Integrated Forecasting System (IFS), at a significantly reduced computational cost. This achievement holds substantial promise for incorporating deep learning-based data-driven models into the suite of operational weather forecasting models.

A significant limitation of these models arises when they are integrated over longer time scales (≥ 2 weeks), leading to instability and the emergence of nonphysical features (see Fig. 1 in Chattopadhyay et al.[18]). The cause

[1]Applied Mathematics, University of California, Santa Cruz, Santa Cruz, CA 95060, USA. [2]Marine, Earth & Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA. ✉email: aschatto@ucsc.edu; rhe@ncsu.edu

of this instability was identified as "spectral bias"[18], an inductive bias in *all* deep neural networks that hinders their ability to capture small-scale features in turbulent flows. A potential solution was proposed in[18] in the form of a framework to construct long-term stable digital twins for atmospheric dynamics.

Research into data-driven digital twins for both global and regional ocean modeling is still in its infancy and has not been explored to the same extent as in atmospheric modeling. The development of long-term stable, accurate data-driven ocean models that maintain computational efficiency is essential for realizing data-driven digital twins for the entire Earth system.

We present here a principled neural operator-based digital twin to predict the sea-surface height (SSH) of the northwest Atlantic Ocean's western boundary current, built upon the same principles as the FouRK model introduced in Chattopadhyay et al.[18]. The proposed model, OceanNet, relies on a Fourier neural operator (FNO), which incorporates a predictor-evaluate-corrector (PEC) integration scheme to suppress autoregressive error growth. Additionally, a spectral regularizer is employed to mitigate spectral bias at small scales (Fig. 1).

OceanNet is trained on historical SSH data from a high-resolution northwest Atlantic Ocean reanalysis and demonstrates long-term stability and competitive skills. It remains competitive with SSH prediction made by a regional ocean dynamical prediction that is based on the state-of-the-art Regional Ocean Modeling System (ROMS) while maintaining a computational cost that is 500,000 cheaper. While OceanNet only emulates high-resolution SSH fields for the regional ocean, future work incorporating other key variables in ocean dynamics would underscores the potential of structured, physics-inspired deep neural operators as a cost-effective alternative to computationally expensive high-resolution numerical models.

Results

Mesoscale ocean circulation dynamics can be well represented by the spatio-temporal evolution of SSH fields. In this section we conduct a comprehensive comparison of SSH predictions generated by OceanNet and the regional ocean dynamical model forecast with independent reanalysis data. To assess the performance rigorously, we employ both qualitative and quantitative measures. The metrics for evaluating predictive accuracy of SSH include root mean squared error (RMSE) and correlation coefficient (CC), which are widely recognized and employed in forecasting[1–3,18]. In addition, we incorporate a specialized object-tracking metric to evaluate the prediction of major ocean features delineated by SSH contours: the modified Hausdorff distance (MHD,[19]). MHD quantifies the comparison of predicted eddy/meander structures to their counterparts in the reanalysis data; identical *shapes* at the same time instant have an MHD of zero. To provide a comprehensive assessment, we
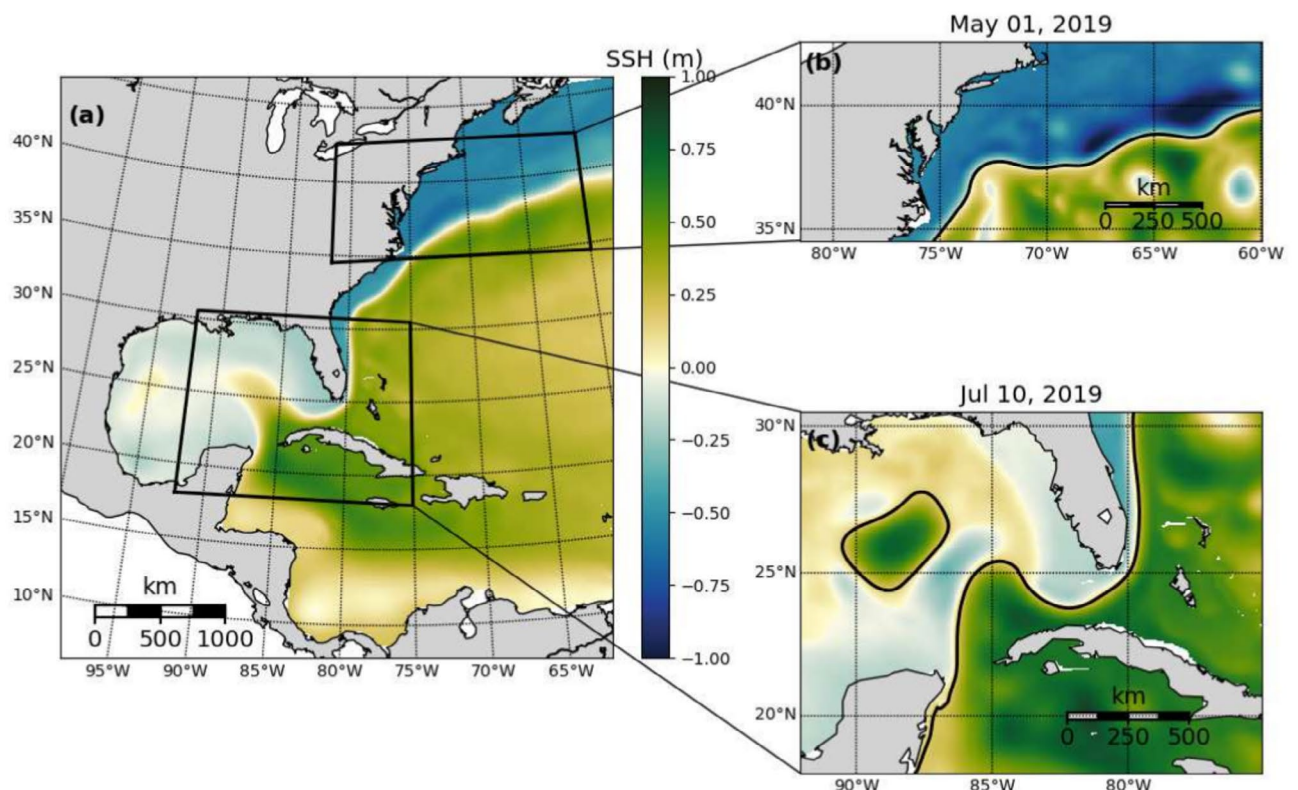
Fig. 1. The domain for the reanalysis data covering the northwestern Atlantic (**a**). The two subdomains used to develop OceanNet, specifically (**b**) the GS separation from the central US east coast to 60 °W and (**c**) the LC eddy-shedding region in the eastern GoM, extending from 92°W into the Atlantic 75°W. The mean SSH from from 1993 to 2020 in the reanalysis data is shown in (**a**), while (**b**) and (**c**) depict daily mean SSH on May 1, 2019, and July 10, 2019, respectively. All three domains share the same color scale.

also present qualitative snapshots of the predicted SSH fields generated by OceanNet, regional ocean dynamical forecast, and the independent SSH field derived from the reanalysis.

Performance of OceanNet in the GoM

We present an evaluation of OceanNet's seasonal (up to 120 days) forecasting performance compared to that of a regional ocean dynamical model forecast for two major GoM LCE shedding events: Eddy Sverdrup (Jul 2019–Jan 2020) and Eddy Thor (Jan 2020–Sep 2020). The target LC dynamics, represented by spatio-temporal evolutions of SSH, are from the reanalysis dataset. Specifics regarding the generation and structure of the reanalysis dataset are covered in section "Reanalysis data".

A qualitative assessment (Fig. 2) reveals that OceanNet performs relatively well in terms of maintaining the general physical consistency of the emulation of the ocean dynamics. However, there are certain cases, where ROMS does signficantly better than OceanNet. For example, the eddy configuration in September 29, 2019, OceanNet predictions of the eddy diffuses quickly while ROMs maintains the strength of the eddies. However, e.g., the eddy cofiguration of Aug 30, 2019, is better captured by OceanNet, while ROMs predicts an unphysical low-pressure region just off the coast. Generally for the GoM region, both RMSE and CC for OceanNet and ROMs is roughly the same, while the eddy shape is better captured by OceanNet as revealed by the MHD. OceanNet, however, maintains both stability and physical consistency for at least 120 days. For this domain, OceanNet is trained with a 5-day lead, consistent with the LC and LCE evolution timescale. ROMS based regional dynamical model predictions were initialized five days apart to allow for fair comparisons between the two models. OceanNet predictions have a "wall-clock" execution time of microseconds, approximately 500,000 times faster than regional dynamical model forecasts. So, while ROMS, in this case, is just as accurate or even better for some of the eddy shedding cases, the computational advantage of OceanNet is undeniable.

An examination of RMSE, CC, and MHD for SSH, over a 120-day period (Fig. 3) demonstrates that OceanNet is competitive with regional dynamical model forecasts, and outperforms random predictions (indicated by the horizontal dashed lines). These random predictions are derived from 1000 random pairs of samples from the training data covering the period of 1993–2018 to calculate the saturation values, which quantify the intrinsic predictability of the ocean system[20–22]. RMSE values above random predictions indicate that the model has lost all short-term forecasting skill, as is obvious for chaotic dynamical systems. Beyond 100 days, ROMs clearly show an
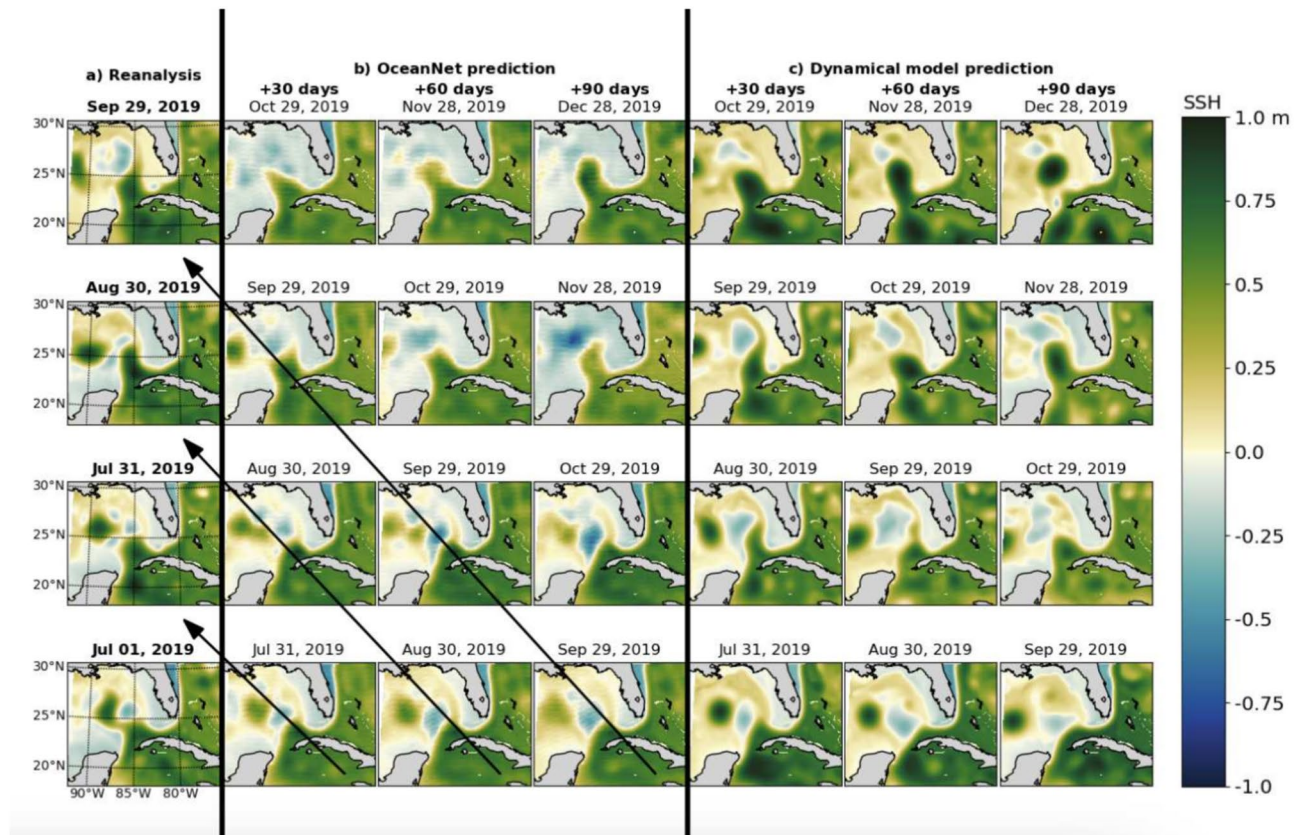
Fig. 2. Prediction performance of OceanNet. (**a**) SSH fields from the ocean reanalysis. (**b**) Predicted SSH generated by the OceanNet. (**c**) regional ocean dynamical model forecasts. Each row in the OceanNet and dynamical model predictions was initialized with the corresponding reanalysis data in the left column. The predictions display SSH forecasts for 30, 60, and 90 days. To evaluate the predictions, we can examine a diagonal comparison with the reanalysis SSH, as indicated by the black arrows in (**b**). The same diagonal comparison can be conducted with the ocean reanalysis data for (**c**).
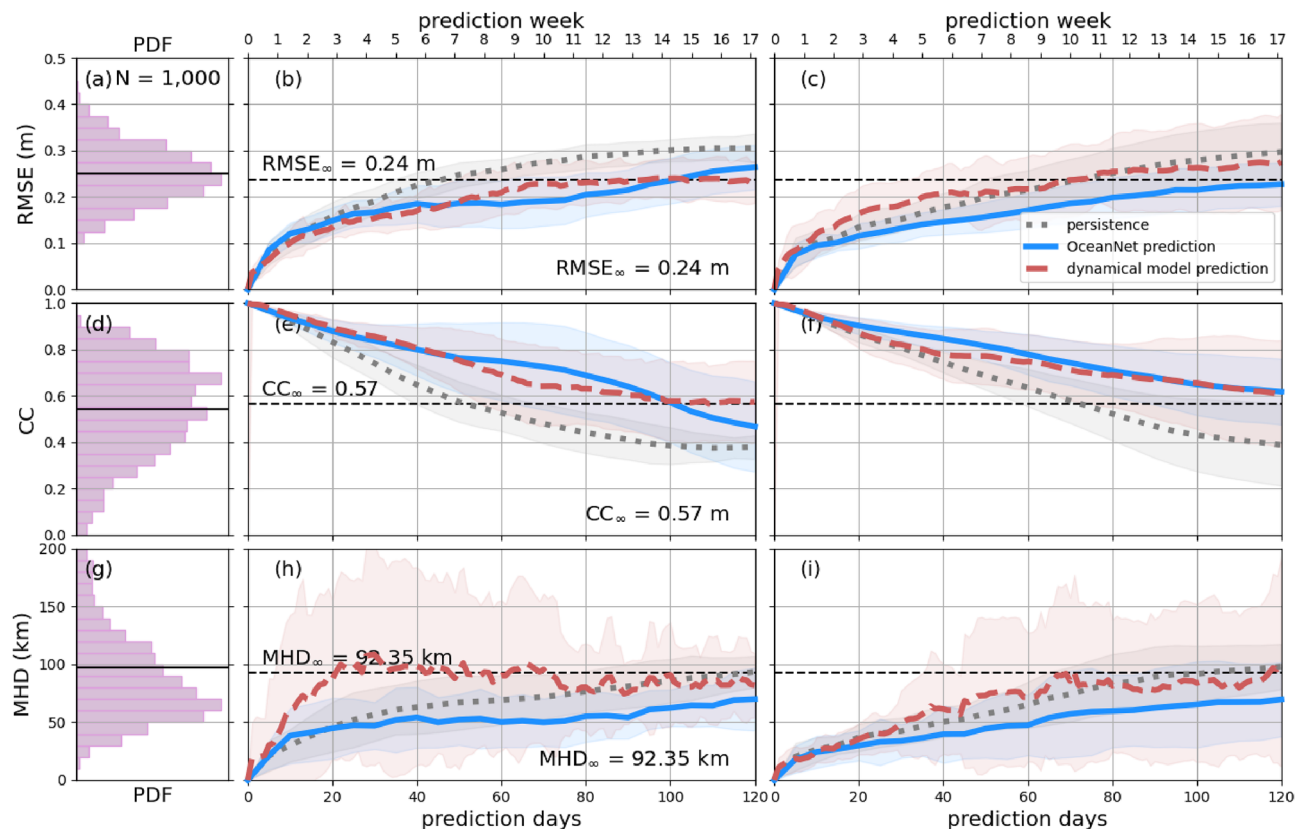
Fig. 3. OceanNet performance metrics in the GoM: RMSE (**a–c**), CC (**d–f**), and MHD (**g–i**). In the left column, probability density functions (PDFs) are presented for 1000 random pairs sourced from the training data spanning 1993–2018 (means shown by black horizontal lines). The performance statistics, calculated based on forecasts of 0–120 days for eddies Sverdrup (middle column) and Thor (right column), are displayed as mean values (lines) with standard deviations (shading). The black horizontal dashed lines in these columns denote saturation values determined as 95% of the means derived from the random pairs. In the middle and right columns, solid blue lines are OceanNet, dashed red lines are the regional ocean dynamical forecast model, and gray dots are persistence. These representations illustrate how each method's statistics compare with the target SSH from the reanalysis dataset. The shading is obtained from the standard deviation due to the choice of different initial conditions to evluate the forecast skill..

improvement in the RMSE (CC) values as compared to OceanNet, however those values are higher (lower) than the saturation values and hence do not provide any insights into predictive performance. Comparing OceanNet and regional dynamical model with persistence forecasting, we see that regional dynamical model suffers from inaccuracies in predicting the timing and configuration of the eddy-shedding and reattachment process, and thus underperforms persistence forecast on several occasions.

We conducted a further comparison (Supplementary Information) of OceanNet with previous deep learning-based models that have been successfully used in weather prediction, such as the model introduced by Weyn et al.[23]. The spectral characteristics of these models exhibit deviations in the high wavenumbers, rendering their predictions nonphysical beyond a 2-week forecasting horizon (Fig. S1). A geostrophic constraint within the loss function was also tested (Fig. S2), but OceanNet in its current design performed better.

Performance of OceanNet in the North Atlantic GS region
We present qualitative and quantitative evaluations of OceanNet's forecasting performance compared to the regional ocean dynamical model forecast for GSM in the northwest Atlantic. OceanNet was specifically trained for this domain with a lead time of 4 days to resolve the evolution of the northern boundary of the GS. To ensure alignment with the initialization dates for regional ocean model (which were separated by 5 days), we show the 20th, 40th, and 60th-day forecasts, as opposed to the 30th, 60th, and 90th-day forecasts used for the LC.

A qualitative assessment reveals that OceanNet effectively captures the SSH propagation of undulations in the northern boundary of the GS However, in certain instances it fails to skillfully captures large-scale eddies and traveling into and out of the domain; this can be attributed to the absence of lateral boundary conditions. (Fig. 4). The predictions of OceanNet are also signficantly more diffusive than reanalysis which suggests that OceanNet cannot with just a single variable capture the correct climatology and variability for decadal simulations. The dynamical model forecast also tends to overpredict extreme values of SSH and the meridional amplitude of the northern boundary. While it is sensitive to initial conditions, OceanNet remains physically consistent over
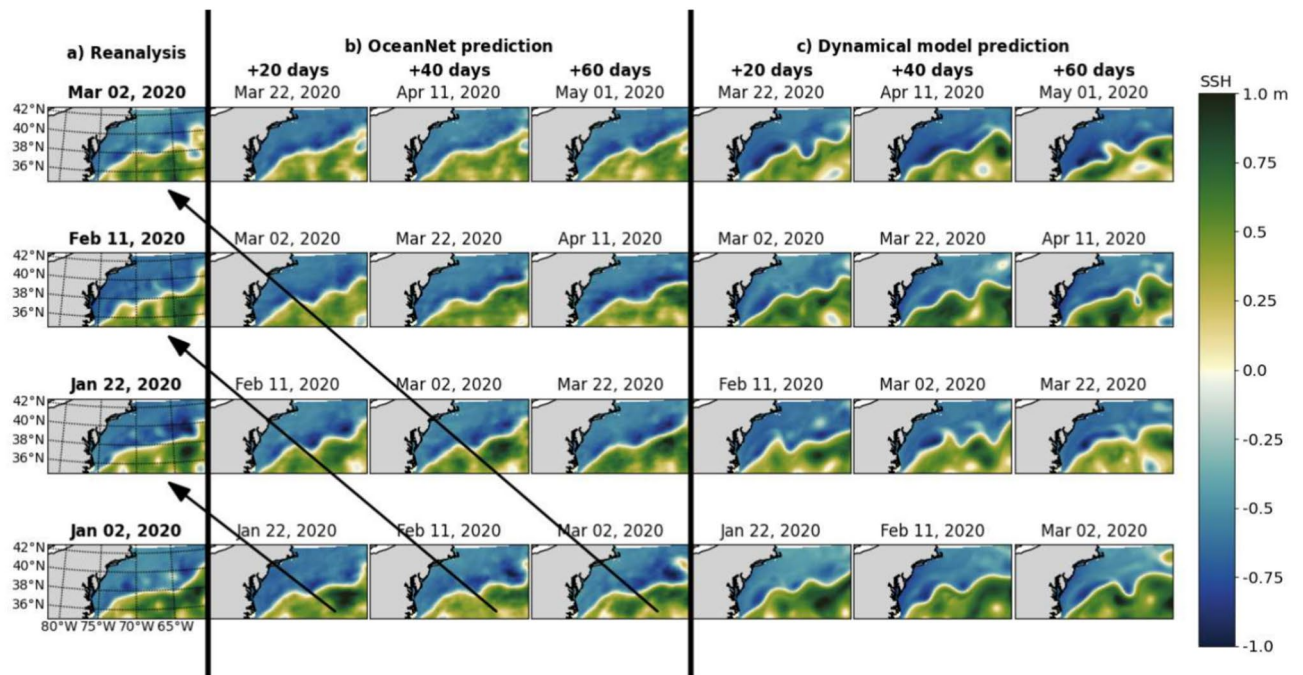
Fig. 4. Performance of OceanNet for GS prediction. (**a**) SSH fields from the ocean reanalysis. (**b**) Predicted SSH generated by OceanNet. (**c**) Regional ocean dynamical model forecasts. In both OceanNet and dynamical model predictions, each row was initialized with the corresponding reanalysis data in the left column. SSH forecasts are provided for 20, 40, and 60 days. To evaluate the predictions, we can perform a diagonal comparison with the reanalysis SSH, as indicated by the black arrows in (**b**). The same diagonal comparison can also be conducted with the ocean reanalysis data for (**c**).

120-day forecasts in this region as well. As observed in the GoM, OceanNet provides stable and physically reasonable SSH predictions for the GSM for at least 120 days (not shown for brevity).

Similar to the GoM, OceanNet consistently outperforms regional ocean dynamical model in RMSE, CC, and MHD computed between the predicted SSH values and the reanalysis SSH over 120 days (Fig. 5). Persistence forecasting also fares reasonably well in this region due to the background state of the GSM; however, OceanNet still outperforms persistence in all three metrics over 120 days.

Discussion

In this study, we introduce OceanNet, a deep neural operator, with principled design structures within the architecture, to function as a digital twin for long-term (up to 120 days) mesoscale ocean circulation forecasting. Our major technical contributions encompass the development of an FNO and the application of a PEC-based convergent integration scheme, along with a spectral regularizer. Together, these techniques mitigate autoregressive error growth and the small-scale spectral bias frequently observed in data-driven models of multi-scale dynamical systems.

By training the model on 26 years (1993–2018) of high-resolution regional ocean reanalysis data for the northwest Atlantic, we evaluate OceanNet's ability to predict mesoscale ocean processes such as the GoM LC eddy shedding and GSM over a 90–120 day forecasting horizon. One major shortcoming of OceanNet is the absence of lateral boundary conditions from the open ocean. However, since OceanNet is trained using more than two decades of ocean reanalysis data, which considers open ocean processes and conditions, including Rossby waves and ocean eddies propagating into the regional domain, information about ocean boundaries may already embedded in the training dataset. Furthermore, we focus on sub-seasonal to seasonal time scales (30–120 days) in the mid-latitude Atlantic Ocean. The propagation speed of Rossby waves at 40°N is approximately 1 cm/s. As a result, meanders and eddies within our time scale of interest only travel 30–120 km from the open boundary. With this limited travel distance, open ocean condition has a relative small impact on the interior circulation predictions made by the machine learning model. We acknowledge that for longer-term forecasting problems or regions in lower latitudes (e.g., Rossby wave speed is 25 cm/s at 5°N), having accurate open boundary conditions will be crucial for model predictions, whether it's a dynamical model or a machine learning model.

Nevertheless, OceanNet demonstrates remarkable stability and consistently competes with the regional ocean dynamical model forecast while being 500, 000 times faster. These results demonstrate the potential of utilizing scientific machine learning to develop long-term stable, and accurate data-driven ocean models of great computational efficiency, paving the way for realizing a data-driven digital twin encompassing the entire climate system.

Despite the promising forecast accuracy observed with reanalysis data, our algorithm possesses several limitations. First, this study exclusively trained and tested OceanNet using reanalysis data pertaining to mesoscale ocean eddies and meanders. Real-world ocean forecasting systems operate with dynamical ocean models and real-time ocean observational data, covering dynamic processes across diverse spatial scales. The disparities
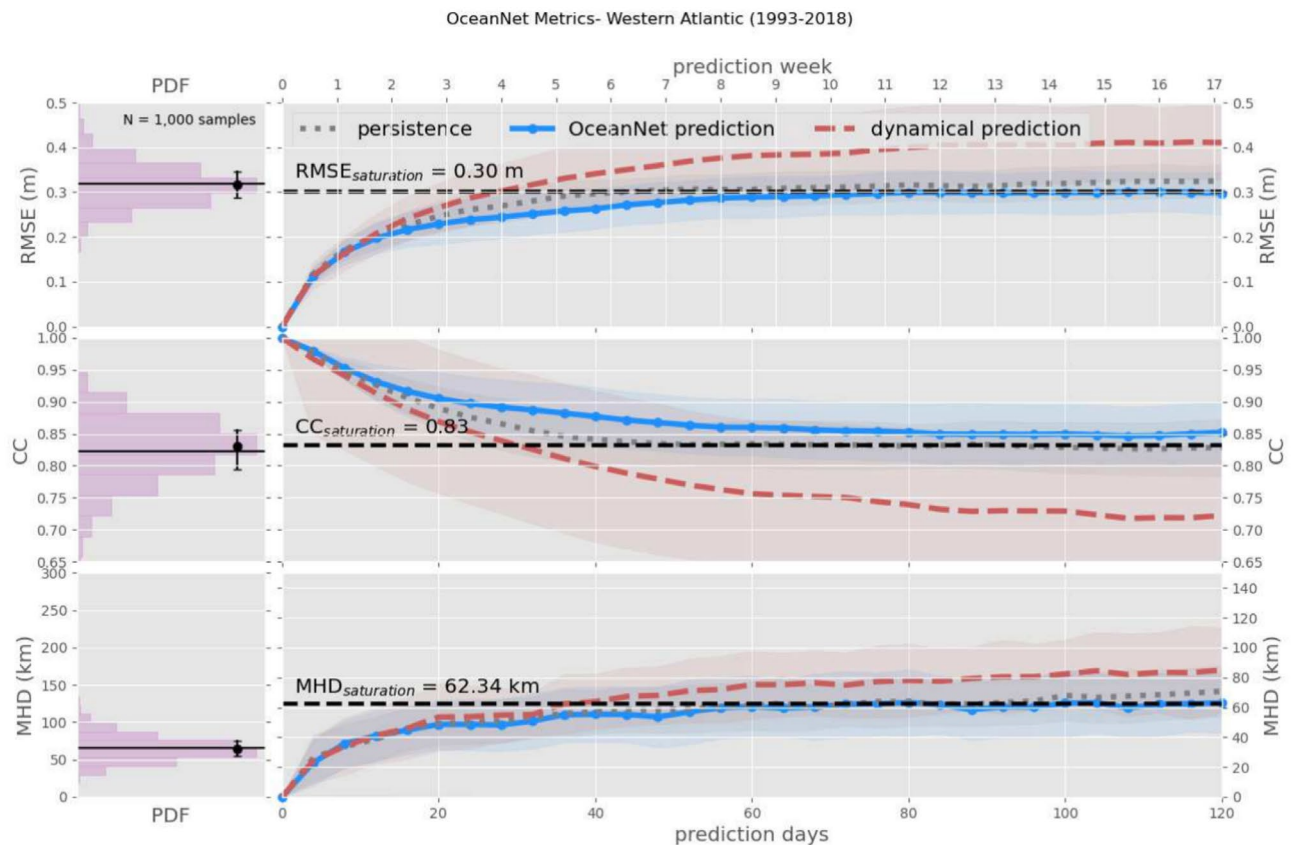
Fig. 5. OceanNet's performance metrics in the northwest Atlantic: RMSE (top), CC (middle), and MHD (bottom), compared to the persistence forecast and regional ocean dynamical model forecast. In the left column, probability density functions (PDFs) are presented, derived from 1000 random pairs sourced from the training data spanning 1993–2018 (means shown by black horizontal lines). The performance statistics, calculated based on forecasts of 0–120 days, are displayed as mean values (lines) with standard deviations (shading). The black horizontal dashed lines denote saturation values, which are determined as 95% of the means derived from the random pairs. Solid blue lines are OceanNet, dashed red lines are the regional ocean dynamical forecast model, and gray dots are persistence. These representations illustrate how each method's statistics compare with the target SSH from the reanalysis dataset. The shading is obtained from the standard deviation due to the choice of different initial conditions to evluate the forecast skill..

between these data sources and scales necessitate further investigation into OceanNet's performance across various ocean applications. Second, OceanNet's focus is solely on SSH. This study did not explore other ocean state variables, such as currents and temperature. The omission of these variables may restrict the model's capacity to accurately predict small-scale ocean circulation phenomena, such as shelf break jets and frontal currents. Third, AI-based methods frequently produce smoother forecast results than dynamical models, which can potentially lead to an underestimation of the magnitudes of extreme ocean events. Therefore, additional research is imperative to assess OceanNet's performance under extreme ocean conditions, e.g., during severe storms. Finally, while control simulations with OceanNet has been performed in this study, the behavior of OceanNet to perturbations or forcings has not been considered. A physically-consistent AI-based ocean model, akin to physics-based numerical models must produce accurate response to forcings. In future studies, with several variables incorporated into OceanNet, its response to external forcings would also be compared to numerical models to establish physical consistency.

Significant opportunities exist for improvement in both AI-based methods and dynamical model-based ocean forecasting. In the AI domain, potential advancements involve the integration of subsurface ocean states and additional ocean variables, the incorporation of temporal dimensions through the training of four-dimensional deep networks, and the exploration of more complex network architectures with increased depth and breadth. In the realm of numerical ocean forecast modeling, the development of pre- and post-processing techniques can help mitigate the inherent biases found in ocean models. We expect that a hybrid approach, combining data-driven and dynamical numerical models will play a pivotal role in pushing the boundaries of excellence in ocean prediction.

Data and methods

Reanalysis data

We utilized high-resolution northwest Atlantic regional ocean reanalysis data to train OceanNet. This reanalysis was generated using a regional implementation of ROMS with the ensemble Kalman filter data assimilation method (EnKF). It features a horizontal resolution of 4 km with 50 vertical layers. For surface atmospheric forcing, we employed data from the European Center for Medium-Range Weather Forecasting (ECMWF) Reanalysis v5 (ERA5), while open boundary conditions were derived from the Copernicus Global Ocean Physics Reanalysis (GLORYS). Ten major tidal constituents from the OSU TPXO tide database were used. The model incorporated 120 river inputs, sourced from the National Water Model and climatology datasets. The temporal scope of the reanalysis spans from January 1, 1993 to December 31, 2020, with daily output.

The assimilated observations encompass a variety of sources, including AVHRR and MODIS Terra sea surface temperature, AVISO along-track sea surface height anomaly, glider temperature/salinity observations from the Integrated Ocean Observing System (IOOS), and the EN4 dataset which aggregates data from Argo floats, shipboard surveys, drifters, moorings, and other sources. EnKF assimilates data with no influence or retainment of previous timesteps, as opposed to 3D/4D-var methods, which include forward and backward passes over the results to ensure continuity. The resulting data assimilative ocean reanalysis allows us to train OceanNet with a time-space continuous SSH dataset with no external knowledge (as would be the case for observations).

OceanNet is trained on 5-day mean SSH data with a 5-day lead time for the GoM domain and 4-day lead time for the GSM domain. Both training datasets consist of 26 years of reanalysis data from 1993 through 2018. SSH Forecasting was conducted for 2019–2020. Further details on OceanNet's training process are provided in section "OceanNet: configuration and design".

OceanNet: configuration and design

The OceanNet model (Fig. 6) has three major components. the Fourier Neural Operator (FNO), which is used to learn the residual, similar to Krishnapriyan et al.[24] and Chattopadhyay et al.[18]; the PEC integrator; and the spectral regularizer, which alleviates spectral bias.

Fourier neural operator (FNO)[25]
OceanNet is built upon the FNO[25]. Training utilizes labeled pairs of historical 5-day mean SSH data in the GoM (section "Reanalysis data"), $X(t)$ (input), $X(t + 5\Delta t)$ (label), and $X(t + 10\Delta t)$ (label), where $\Delta t = 1$ day. For training OceanNet in the northwest Atlantic region, the labels are $X(t + 4\Delta t)$ and $X(t + 8\Delta t)$. The training assumes the governing partial differential equation for the reduced ocean system involves ocean states $X(t)$:
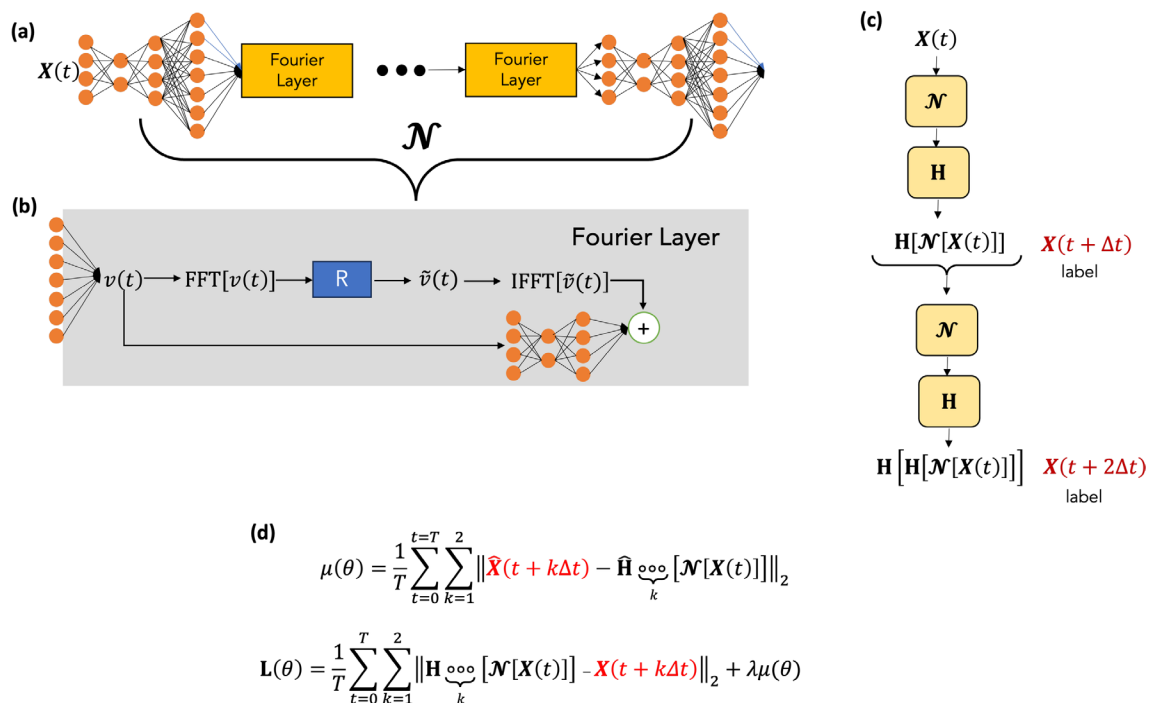
Fig. 6. (**a**) A schematic of the OceanNet model. (**b**) The Fourier Neural Operator, depicted as $\mathcal{N}$. The neural operator uplifts the input state, $X(t)$ to a high-dimensional space using two convolutional layers. The uplifted state then undergoes a Fourier transform, subsequent coarse-graining by removing the high-wavenumber modes, and then undergoes an inverse Fourier transform. Finally a bias layer is also added to account for aperiodicity in the data. Finally, two more convolutional layers are added to preserve the dimension of the final output, $X(t + k\Delta t)$. (**c**) 2-time-step training scheme. (**d**) The loss function used. Here, X is the state of the system that is predicted, and **H** is the PEC-based convergent integration scheme used.

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}(t)). \tag{1}$$

To integrate the system from the initial condition, $X(t)$, we represent Eq. (1) in its discrete form:

$$\mathbf{X}(t + 5\Delta t) = \mathbf{X}(t) + \underbrace{\int_t^{t+5\Delta t} \underbrace{\mathbf{F}(\mathbf{X}(t))dt}_{\mathcal{N}[\circ,\theta]}}_{\mathbf{H}[\circ]}. \tag{2}$$

Here, $\mathcal{N}$ is an FNO that parameterizes $\mathbf{F}$ with four Fourier layers, similar to Li et al.[25], each layer retaining 64 Fourier modes. θ represents the $\approx 80 \times 10^6$ the trainable parameters of the FNO. $\mathbf{H}$ is a higher-order predictor–evaluate–corrector (PEC) integrator (section "OceanNet: configuration and design").

Predictor-Evaluate-Corrector integration scheme
Similar to the higher-order integration scheme in the form of fourth-order Runge-Kutta (RK4) (Chattopadhyay et al.[18]), we implement a PEC scheme in OceanNet, represented by the operator, $\mathbf{H}$. The operations in $\mathbf{H}$ are given by:

$$i_1 = \mathcal{N}[\mathbf{X}(t), \theta], \tag{3a}$$

$$z = \mathbf{X}(t) + \mathcal{N}\left[\mathbf{X}(t) + \frac{1}{2}i_1, \theta\right]. \tag{3b}$$

The predicted state is given by $z = \mathbf{H}[\mathcal{N}(\mathbf{X}(t)), \theta]$.

Although most of the higher-order integration schemes, including RK4, demonstrate good performance for this problem, we have identified PEC as the most effective choice. A theoretical study of the effect of each integration scheme on the inductive bias of the trained $\mathcal{N}$ is an active area of research[24], especially for understanding the role it plays on the subsequent spectral bias (section "OceanNet: configuration and design").

Fourier regularizer and 2-time-step loss function
In OceanNet's loss function, we incorporate a spectral regularizer based on Fourier transforms introduced in Chattopadhyay et al.[18]. This is in addition to the standard mean squared error loss (MSE) function, which is computed exclusively for grid points located over the ocean.

The spectral regularizer penalizes deviations in the absolute value of the latitude-averaged zonal Fourier spectrum (Fourier spectrum from now on) of the SSH field at small wavenumbers. Such deviations arise due to spectral bias[18,26], which represents an inherent inductive bias in deep neural networks. This bias is responsible for their limitations in learning the fine-scale dynamics of turbulent flow.

Furthermore, the combined loss function, which integrates both MSE and the spectral regularizer is evaluated over two time steps, namely $t + 5\Delta t$ ($t + 4\Delta t$ for GS) and $t + 10\Delta t$ ($t + 8\Delta t$ for GS), to extend the stability horizon. This two-time-step loss function had remarkable success in the weather prediction model, FourCastNet[1]. The spectral regularizers for time $t + 5\Delta t$ and $t + 10\Delta t$ are defined as follows:

$$\mu(t + 5\Delta t, \theta) = \sum_{t=0}^{t=T} \left\| \hat{\mathbf{X}}(t + 5\Delta t)\Big|_{k_x \geq k_T} - \hat{\mathbf{H}}[\mathcal{N}(\mathbf{X}(t), \theta)]\Big|_{k_x \geq k_T} \right\|_2^2. \tag{4}$$

$$\mu(t + 10\Delta t, \theta) = \sum_{t=0}^{t=T} \left\| \hat{\mathbf{X}}(t + 10\Delta t)\Big|_{k_x \geq k_T} - \hat{\mathbf{H}}[\mathbf{H}[\mathcal{N}(\mathbf{X}(t), \theta)]]\Big|_{k_x \geq k_T} \right\|_2^2. \tag{5}$$

Here, the $[\hat{\circ}]$ operator is the latitude-averaged zonal Fourier transform of both the predicted and the target SSH fields. k_T is a cutoff wavenumber chosen to be 100 after significant trial and error. The loss functions for $t + 5\Delta t$ and $t + 10\Delta t$ are given by $\mathbf{L}(t + 5\Delta t, \theta)$:

$$\mathbf{L}(t + 5\Delta t, \theta) = \sum_{t=0}^{t=T} \|\mathbf{X}(t + 5\Delta t) - \mathbf{H}[\mathcal{N}(\mathbf{X}(t), \theta)]\|_2^2 + \lambda\mu(t + 5\Delta t, \theta). \tag{6}$$

$$\mathbf{L}(t + 10\Delta t, \theta) = \sum_{t=0}^{t=T} \|\mathbf{X}(t + 10\Delta t) - \mathbf{H}[\mathbf{H}[\mathcal{N}(\mathbf{X}(t), \theta)]]\|_2^2 + \lambda\mu(t + 10\Delta t, \theta). \tag{7}$$

The total loss is given by:

$$\mathbf{L}(\theta) = \mathbf{L}(t + 5\Delta t, \theta) + \mathbf{L}(t + 10\Delta t, \theta). \tag{8}$$

The spectral regularizer ensures that the high wavenumbers in the Fourier spectrum of SSH remain consistent with the target Fourier spectrum. This measure is effective in reducing spectral bias throughout the training process. The two-time-step loss function enhances the stability horizon of the model. The number of time steps

over which the loss is calculated can potentially be extended. However, with each increase in the number of time steps, the memory requirement for the subsequent backpropagation process during training grows exponentially.

References

1. Pathak, J. *et al*. FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv preprint arXiv:2202.11214 (2022).
2. Lam, R. *et al*. Graphcast: Learning skillful medium-range global weather forecasting. arXiv preprint arXiv:2212.12794 (2022).
3. Bi, K. *et al.* Accurate medium-range global weather forecasting with 3d neural networks. *Nature* **2023**, 1–6 (2023).
4. Dengo, J. The problem of gulf stream separation: A barotropic approach. *J. Phys. Oceanogr.* **23**(10), 2182–2200 (1993).
5. Chassignet, E. & Marshall, D. *Gulf Stream Separation in Numerical Ocean Models* (American Geophysical Union, 2008).
6. Ezer, T. Revisiting the problem of the gulf stream separation: On the representation of topography in ocean models with different types of vertical grids. *Ocean Model.* **104**, 15–27 (2016).
7. Spall, M. A. Dynamics of the gulf stream/deep western boundary current crossover. Part i: Entrainment and recirculation. *J. Phys. Oceanogr.* **26**(10), 2152–2168 (1996).
8. Zhang, R. & Vallis, G. K. The role of bottom vortex stretching on the path of the North Atlantic western boundary current and on the northern recirculation gyre. *J. Phys. Oceanogr.* **37**(8), 2053–2080 (2007).
9. Hurlburt, H. E. & Hogan, P. J. The gulf stream pathway and the impacts of the eddy-driven abyssal circulation and the deep western boundary current. *Dyn. Atmos. Oceans* **45**(3–4), 71–101 (2008).
10. Schoonover, J., Dewar, W. K., Wienders, N. & Deremble, B. Local sensitivities of the gulf stream separation. *J. Phys. Oceanogr.* **47**(2), 353–373 (2017).
11. Andres, M. On the recent destabilization of the gulf stream path downstream of cape hatteras. *Geophys. Res. Lett.* **43**(18), 9836–9842 (2016).
12. Gangopadhyay, A., Gawarkiewicz, G., Silva, E. N. S., Monim, M. & Clark, J. An observed regime shift in the formation of warm core rings from the gulf stream. *Sci. Rep.* **9**(1), 12319 (2019).
13. Silver, A., Gangopadhyay, A., Gawarkiewicz, G., Fratantoni, P. & Clark, J. Increased gulf stream warm core ring formations contributes to an observed increase in salinity maximum intrusions on the northeast shelf. *Sci. Rep.* **13**(1), 1–9 (2023).
14. Wang, J. L., Zhuang, H., Chérubin, L. M., Ibrahim, A. K. & Muhamed Ali, A. Medium-term forecasting of loop current eddy cameron and eddy darwin formation in the gulf of mexico with a divide-and-conquer machine learning approach. *J. Geophys. Res.: Oceans* **124**(8), 5586–5606 (2019).
15. Agarwal, N., Kondrashov, D., Dueben, P., Ryzhov, E. & Berloff, P. A comparison of data-driven approaches to build low-dimensional ocean models. *J. Adv. Model. Earth Syst.* **13**(9), 2021–002537 (2021).
16. Chattopadhyay, A., Nabizadeh, E. & Hassanzadeh, P. Analog forecasting of extreme-causing weather patterns using deep learning. *J. Adv. Model. Earth Syst.* **12**(2), 00195 (2020).
17. Chattopadhyay, A., Mustafa, M., Hassanzadeh, P. & Kashinath, K. Deep spatial transformers for autoregressive data-driven forecasting of geophysical turbulence. In *Proceedings of the 10th International Conference on Climate Informatics* 106–112 (2020).
18. Chattopadhyay, A., & Hassanzadeh, P. Long-term instabilities of deep learning-based digital twins of the climate system: The cause and a solution. arXiv preprint arXiv:2304.07029 (2023).
19. Dukhovskoy, D. S., Ubnoske, J., Blanchard-Wrigglesworth, E., Hiester, H. R. & Proshutinsky, A. Skill metrics for evaluation and comparison of sea ice models. *J. Geophys. Res.: Oceans* **120**(9), 5910–5931 (2015).
20. Dalcher, A. & Kalnay, E. Error growth and predictability in operational ecmwf forecasts. *Tellus A* **39**(5), 474–491 (1987).
21. DelSole, T. Predictability and information theory. Part i: Measures of predictability. *J. Atmos. Sci.* **61**(20), 2425–2440 (2004).
22. Dukhovskoy, D. S., Chassignet, E. P., Bozec, A. & Morey, S. L. Assessment of predictability of the loop current in the gulf of mexico from observing system experiments and observing system simulation experiments. *Front. Mar. Sci.* **10**, 1153824 (2023).
23. Weyn, J. A., Durran, D. R. & Caruana, R. Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J. Adv. Model. Earth Syst.* **12**(9), 2020–002109 (2020).
24. Krishnapriyan, A. S., Queiruga, A. F., Erichson, N. B. & Mahoney, M. W. Learning continuous models for continuous physics. arXiv preprint arXiv:2202.08494 (2022).
25. Li, Z. *et al.*. Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895 (2020).
26. Xu, Z.-Q. J., Zhang, Y., Luo, T., Xiao, Y. & Ma, Z. Frequency principle: Fourier analysis sheds light on deep neural networks. arXiv preprint arXiv:1901.06523 (2019).

Acknowledgements

Author contributions

All authors provided equal contributions to ideation, result generation, and discussions in this manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-72145-0.

Correspondence and requests for materials should be addressed to A.C. or R.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.