Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL088376

Key Points:
- We present two physics-aware machine learning algorithms for ocean mesoscale parameterizations
- We discover closed-form equations for eddy momentum, temperature, and energy parameterizations that obey conservation properties
- The physics-constrained deep learning closure is more stable than closed-form equations when implemented in an ocean model

Supporting Information:
- Supporting Information S1

Correspondence to:
L. Zanna,
laure.zanna@nyu.edu

Citation:
Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*, e2020GL088376. https://doi.org/10.1029/2020GL088376

Received 20 APR 2020
Accepted 1 AUG 2020
Accepted article online 6 AUG 2020

Data-Driven Equation Discovery of Ocean Mesoscale Closures

Laure Zanna[1,2] **and Thomas Bolton**[2]

[1]Courant Institute of Mathematical Sciences, New York University, New York, NY, USA, [2]Department of Physics, University of Oxford, Oxford, UK

Abstract The resolution of climate models is limited by computational cost. Therefore, we must rely on parameterizations to represent processes occurring below the scale resolved by the models. Here, we focus on parameterizations of ocean mesoscale eddies and employ machine learning (ML), namely, relevance vector machines (RVMs) and convolutional neural networks (CNNs), to derive computationally efficient parameterizations from data, which are interpretable and/or encapsulate physics. In particular, we demonstrate the usefulness of the RVM algorithm to reveal closed-form equations for eddy parameterizations with embedded conservation laws. When implemented in an idealized ocean model, all parameterizations improve the statistics of the coarse-resolution simulation. The CNN is more stable than the RVM such that its skill in reproducing the high-resolution simulation is higher than the other schemes; however, the RVM scheme is interpretable. This work shows the potential for new physics-aware interpretable ML turbulence parameterizations for use in ocean climate models.

Plain Language Summary The complexity of numerical models used for future climate projections is limited by their computational cost. Many key processes, such as ocean eddies, are not adequately resolved and must be approximated using parameterizations. However, parameterizations are often imperfect and reduce the accuracy of the simulations. Machine learning is now opening new avenues to improve climate simulations by extracting such parameterizations directly from data, rather than using idealized theories as typically done. We show that efficient modern machine learning algorithms can accurately represent the physics of ocean eddies, can be constrained by physical laws, and can be interpretable. Our results simultaneously open the door to the discovery of new physics from data and the improvement of climate simulations.

1. Introduction

Turbulent processes are critical components of the climate system and influence the circulation of both the ocean and atmosphere. For example, ocean mesoscale eddies, which are turbulent features of scale 10–100 km, dominate the oceanic kinetic energy reservoir (Ferrari & Wunsch, 2009) and are key for the lateral and vertical transport of tracers, such as heat, carbon, and oxygen. These turbulent processes occur on scales that are below the resolution of typical global climate models, which is roughly 25–100 km (IPCC, 2013). Therefore, the effects of these turbulent processes on the large-scale must be approximated.

These approximations, called parameterizations or closures, are often developed using idealized theories of the bulk effect of the subgrid process on the large scale (Warner, 2010). This approach has been used for many decades but is not necessarily optimal as it neglects certain physical effects. Imperfections in current parameterizations and missing physics in climate models introduce significant biases in simulations and considerable uncertainty in anthropogenic climate change projections (IPCC, 2013). For example, current parameterizations of ocean eddies target the effect of (i) buoyancy fluxes by removing large-scale available potential energy (Gent & Mcwilliams, 1990) and (ii) momentum fluxes using viscous closures which dissipate momentum (Zanna et al., 2020).

While improving certain properties of the flow (Danabasoglu et al., 1994), eddy parameterizations are missing key energy pathways such as the conversion of available potential energy into subgrid kinetic energy, or the backscatter of kinetic energy to the large-scale flow (Bachman, 2019; Jansen et al., 2015; Zanna et al., 2017). In addition, these parameterizations spuriously dissipate kinetic energy (Jansen & Held, 2014; Kjellsson & Zanna, 2017). These imperfect representations of ocean eddy physics in models can affect the

strength and variability of large-scale ocean currents and ocean heat uptake (Kuhlbrodt & Gregory, 2012; Zanna et al., 2017). Increasing resolution can reduce some of these biases; however, due to the computational expense of running turbulence-resolving simulations, subgrid parameterizations will be in demand for several decades.

Recently, the advent of machine learning (ML) has given rise to a new class of data-driven parameterizations. Studies rely on ML to optimally tune parameters of existing closures (Ling et al., 2016; Schneider et al., 2017). This approach, while useful, still neglects the missing physics not encapsulated in the current parameterizations. Instead, several studies have shown the promise of new ML parameterizations of subgrid processes in the atmosphere (Brenowitz & Bretherton, 2018; Gentine et al., 2018; O'Gorman & Dwyer, 2018; Rasp et al., 2018) and ocean (Bolton & Zanna, 2019). However, this new class of ML parameterizations often uses black box algorithms (e.g., neural networks) such that the laws of physics are not necessarily respected unless imposed (Beucler et al., 2019; Ling et al., 2016), and interpreting the data-driven parameterization becomes intractable.

Here, we propose a complementary or alternative route to both the traditional physics-driven bulk approach and the ML black box approach of deep learning. We use ML to discover closed-form equations for mesoscale eddy parameterizations for coarse-resolution ocean models using high-resolution model data. Given some spatiotemporal data set of the subgrid eddy forcing, we uncover an equation that could have produced that data set (Rudy et al., 2017; Zhang & Lin, 2018). This approach has the following advantages over more complex methods such as convolutional neural networks (CNNs): the end result is significantly easier to interpret physically, the computational cost of implementation is lower, and training time of the algorithm is also lower. Data-driven discovery of equations has been extensively used to reveal known equations, such as Burger's or Navier-Stokes' equations (Kutz, 2017). However, unlike in these studies, we use the algorithm to discover unknown equations for the subgrid eddy forcings.

2. Data and Methods

2.1. Training Data and Coarse Graining

We use a primitive equation model, MITgcm (Marshall et al., 1997), to generate high-resolution data and construct new eddy momentum, temperature, and energy parameterizations. We run highly idealized double-gyre eddy-resolving barotropic and baroclinic simulations in a square domain. The simulations use a beta-plane approximation and a constant surface zonal wind forcing. These simulations are designed to create highly turbulent flow regimes, with mesoscale eddies shedding from the jet extension.

The barotropic model has a single layer of depth 500 m and length 3,840 km, similar to Cooper and Zanna (2015). We spin-up the model from rest for 10 years, at a spatial resolution of 3.75 km. The baroclinic model comprises of 15 vertical levels, with a total depth of 3,600 m. Due to the increased computational cost of running the baroclinic simulation compared to the barotropic model, we decreased the domain size from 3,840 km in length to 1,920 km, with a spatial resolution of 7.5 km. The meridional temperature gradient is imposed via surface restoring to a linear profile. We spin-up the baroclinic model for 100 years and then run for a further 10 years for data collection. Further details about the simulations are given in the Supplementary Information S1.

After spin-up, we select 1,000 time slices of model output, with 4 days between each time slice. We remove information at small scales by applying a horizontal Gaussian filter of width 30 km and then coarse-grain to a 30-km grid, which is denoted by Bolton and Zanna (2019) (Supporting Information S2). The subgrid eddy momentum and temperature forcing terms, for each vertical level, are then defined by

$$\mathbf{S_u} = \begin{pmatrix} S_x \\ S_y \end{pmatrix} = (\overline{\mathbf{u}} \cdot \overline{\nabla})\overline{\mathbf{u}} - \overline{(\mathbf{u} \cdot \nabla)\mathbf{u}}, \tag{1}$$

$$S_T = (\overline{\mathbf{u}} \cdot \overline{\nabla})T - \overline{(\mathbf{u} \cdot \nabla)T}, \tag{2}$$

respectively. Here, ∇ is the horizontal 2D gradient operator, T is the temperature, and the horizontal velocity $\overline{\mathbf{u}} = (\overline{u}, \overline{v})$. These terms reflect the turbulent nonlinear terms truncated in coarse-resolution models which need to be parameterized (Berloff, 2005; Mana & Zanna, 2014). At every grid point for every

time slice, we both (i) calculate the target eddy forcing, i.e, Equations 1 and 2, and (ii) construct a library of diverse functions which are necessary for the relevance vector machine (RVM) method described below and are relevant to the process being parameterized.

2.2. ML Algorithms

We use two physics-aware ML algorithms to learn subgrid ocean closures: (1) RVM to learn closed-form equations and (2) CNNs, with momentum conservation embedded in the architecture.

2.2.1. Relevance Vector Machine

Here, we employ the sparse Bayesian regression method used in Zhang and Lin (2018) based on RVMs (Tipping, 2001) to reveal new eddy parameterizations. RVM is a regression technique that assumes Gaussian prior distributions for each regression weight (Bishop, 2006). The width of the Gaussian prior of each regression weight provides a measure of uncertainty of that regression weight. The method relies on a library of functions, which can comprise of any function such as products or derivatives of relevant quantities defined as basis functions (e.g., velocities, velocity shears, and temperature shears). The sparse regression is applied iteratively to the library of functions, and then a pruning of the library of functions is carried out by discarding the functions with an uncertainty higher than a prespecified threshold (Zhang & Lin, 2018). This uncertainty threshold, δ, is the only parameter that requires setting in the Zhang and Lin (2018) method. The algorithm finishes when the uncertainty measures of each regression weight stop changing from iteration to iteration. We found the Zhang and Lin (2018) method to be more robust than the sequential threshold ridge regression (STRidge) of Rudy et al. (2017). For example, using data to discover the known 2D advection-diffusion equations, we found that STRidge required substantially more data for training than the RVM method. In addition, STRidge has a large number of tunable hyperparameters which substantially influenced the form of the discovered equation compared to the RVM method, which has only one hyperparameter. Finally, unlike STRidge, the Zhang and Lin (2018) method provides an error associated with the weights discovered. Given these tests were performed on known equations in which we knew the answers, we opted for the use of Zhang and Lin (2018) RVM method to discover unknown ocean parameterizations.

At every grid point for every time slice from the MITgcm coarse-grained output (described above), we construct a library of diverse functions, ϕ_i, which are derived from a set of basis functions relevant to the process being parameterized. We build the library from the filtered velocities $\overline{u}$, $\overline{v}$, and $\overline{T}$ using up to second order for both spatial derivatives and polynomial products, mainly due to memory limitations. The basis of functions used for the momentum and temperature eddy parameterizations differ and will be discussed in the next section. We normalized each function individually such that they have zero mean and unit variance. We use 50% of the 1,000 time slices for training and the other 50% for validation. For both the eddy momentum and temperature forcing, we impose a physical constraint for global conservation. To do so, we only specify library functions that can be written as the divergence of a flux (or as the divergence of a tensor $\mathbf{T}$ for the eddy momentum forcing, i.e., $\overline{\nabla} \cdot \mathbf{T}$), such that with the appropriate boundary conditions, there is no net input of momentum or temperature.

We then apply the iterative RVM algorithm to prune the library of functions and construct the final equation for the subgrid forcing (independently for S_x, S_y, and S_T) as a linear sum of the functions, ϕ_i, each weighted by the regression coefficient, w_i. We estimate the performance of the final equation by calculating the R^2 coefficient of determination using the withheld validation data. The full process of discovery with RVM is illustrated in Figure 1a. Further details can be found in the Supporting Information S1.

2.2.2. Convolutional Neural Network

We are using a fully CNN (FCNN) on the high-resolution validation data (the truth). The authors have already shown that CNNs are powerful at parameterizing mesoscale eddy momentum forcing and can generalize very well to different regimes, in particular to different dynamical regions and different turbulent regimes (Bolton & Zanna, 2019). Other studies have shown the success of neural network in representing turbulent closures from large-eddy simulations (Ling et al., 2016; Maulik & San, 2017; Wang et al., 2020), though none have been implemented in a forced-dissipative model as of yet. The FCNN used here (Figure 1b and Supporting Information S1) is trained using the same barotropic model data as for the RVM expression, with the velocity components, $\overline{u}$ and $\overline{v}$ as inputs. There are four convolution layers

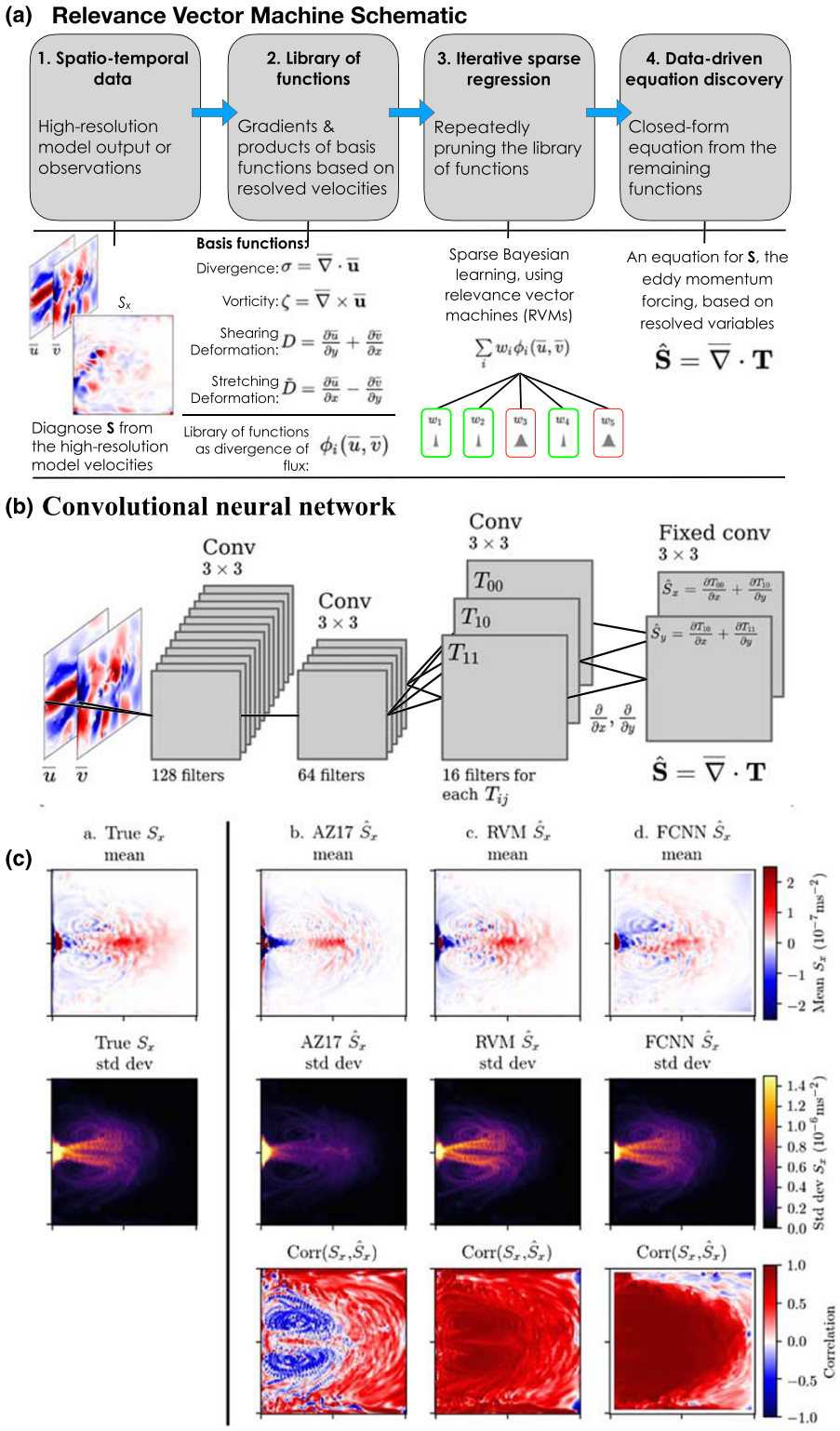

Figure 1. (a) Illustration of the RVM procedure; (b) schematic of the architecture of the physics-constrained fully convolutional neural network (FCNN); (c) offline validation of the subgrid momentum forcing from the barotropic simulations for three parameterizations, denoted as $\hat{\mathbf{S}}$ —the physics-driven $\hat{\mathbf{S}}^{AZ}$, $\hat{\mathbf{S}}^{BT}$ revealed by the RVM algorithm (Equation 5), and the FCNN—against the diagnosed forcing from high-resolution data, $\mathbf{S}$. Top row shows the mean (ms^{-2}), middle row the standard deviation (ms^{-2}), and the bottom row the Pearson correlation of the zonal component of the eddy momentum forcing, S_x and $\hat{S}_x$ (the meridional component is shown in Supporting Information S1). The x- and y-axes are longitude and latitude, respectively; the extent is 3,840 km in each direction.

simultaneously predicting both components of the eddy momentum forcing. The architecture of the FCNN is physically constrained (Beucler et al., 2019) such that the activation maps (i.e., the results) of the third convolution layer represent the elements of a symmetric eddy stress tensor **T**. The final convolution layer then takes the spatial derivatives of the eddy stress tensor elements, using fixed filters representing central-difference stencils, forming predictions S_x and S_y. By physically constraining the architecture to form the elements of a symmetric eddy stress tensor, global momentum and vorticity conservation can be achieved. The hyperparameters of the architecture, such as the number of convolution layers and the number of filters, were chosen by experimenting with numerous configurations and examining the impact of the R^2 coefficient on the validation data, as commonly done. We do not use bias parameters in any of the convolution layers. The details of the FCNN architecture are in the Supporting Information S1 for full reproducibility of the results.

2.3. Numerical Model for Implementation

The RVM and FCNN parameterizations are implemented in an idealized ocean model. Implementation of the FCNN into a Fortran code (e.g., MITgcm) is nontrivial; therefore, we opt to implement the parameterizations using Python since it was used to train and save the FCNN. The Python-based idealized ocean model is a shallow water model, which bears many resemblances to the MITgcm primitive equation barotropic model, including the horizontal velocities and sea surface height as prognostic variables, a double-gyre configuration with a constant wind forcing, and an idealized bathymetry. The parameterizations are implemented into a 30-km resolution version of the idealized shallow-water Python model, which was span-up from rest for 10 years, and then run for an additional 10 years for analysis. Further details are available in the Supporting Information S1.

3. Data-Driven Equation Discovery for Mesoscale Eddies

Improved parameterizations of mesoscale eddy momentum, temperature, and energy are crucial to improving the transport of tracers, as well as countering the energy deficit caused by scale truncation, and viscous and diffusion parameterizations within coarser resolution models. To derive new data-driven closures, we use the data generated from idealized eddy-resolving barotropic and baroclinic simulations, with horizontal resolutions of 3.5 and 7.5 km, respectively, which emulate western boundary currents and their jet extensions at midlatitudes (section 2). Our target is to parameterize eddy momentum (section 3.1) and temperature fluxes and an eddy prognostic equation (section 3.2) for coarser-resolution models, here chosen to be of 30-km horizontal resolution (eddy-permitting), similar to CMIP class eddy-permitting models. We will extract the subgrid forcing using the RVM algorithm.

3.1. Discovering Eddy Momentum Parameterizations

For constructing the library of functions to reveal expressions for the eddy momentum forcing, we write the spatial derivatives of the velocity field using the following basis functions

$$\zeta = \overline{v}_x - \overline{u}_y, \qquad \sigma = \overline{u}_x + \overline{v}_y, \tag{3a}$$

$$D = \overline{u}_y + \overline{v}_x, \qquad \tilde{D} = \overline{u}_x - \overline{v}_y, \tag{3b}$$

where the short hands $()_{x,y} \equiv \frac{\partial}{\partial x,\, y}$ are used for spatial derivatives, ζ is the relative vorticity, σ is the divergence, and D and $\tilde{D}$ are the shearing and stretching deformation of the flow field, respectively. We chose to write the library of functions using this basis because (i) initially our data-driven discovery method was automatically forming Equations 3a and 3b, when given only velocity components and their derivatives, without a priori knowledge, and (ii) the dynamical quantities defined by Equation 3 are relevant to turbulent eddy parameterizations (Pope, 1975; Smagorinsky, 1963). The RVM algorithm, therefore, revealed an improved basis in which to write the library of functions. Finding the optimal physical basis is important to identify the key dynamical components from which to construct parameterizations in general, as well as helping with physics discovery from data.

We separately apply the RVM algorithm to data from the barotropic and baroclinic model. The predicted subgrid momentum forcing is denoted by $\hat{\mathbf{S}}_{\mathbf{u}} = \left(\hat{S}_x, \hat{S}_y\right)$. We performed an extensive sensitivity analysis to

the sole hyperparameter, the threshold δ, of the method (Supporting Information S5). At low threshold values, the RVM algorithm selects a single function, namely, the gradients of enstrophy $(\zeta^2)_x$ and $(\zeta^2)_y$ for predictions of S_x and S_y, respectively, which captures ~20% of the variance. As the pruning threshold increases, there is a large increase from ~20% to ~50% variance captured, with the number of functions only increasing from 1 to 3 for both S_x and S_y. The expression revealed by the RVM is then given by

$$\hat{\mathbf{S}}_{\mathbf{u}}^{BT} = \begin{pmatrix} w_0(\zeta^2)_x - w_1(\zeta D)_x + w_2(\zeta\tilde{D})_y \\ w_3(\zeta^2)_y + w_4(\zeta D)_y + w_5(\zeta\tilde{D})_x \end{pmatrix}, \tag{4}$$

where $w_0 = -4.096 \times 10^8$, $w_1 = -5.483 \times 10^8$, $w_2 = -4.384 \times 10^8$, $w_3 = -4.100 \times 10^8$, $w_4 = -6.332 \times 10^8$, and $w_5 = -4.815 \times 10^8$, with units of m^2. Each coefficient has an uncertainty estimate which is on the order of a few percent and never exceeds 10%. The uncertainty associated with each weight is not listed as it is always smaller than the coefficient of variation used (see below) for each parameterization discovered. The zonal and meridional components of the predicted RVM expression capture 55.6% and 50.6% of the variance, respectively. Adding six more functions would increase the R^2 value to up to 80% but increasing the complexity of the expression (Equations S12 and S13).

To quantify the differences between the regression coefficients, we use the coefficient of variation (i.e., relative standard deviation), which provides a standardized measure of the dispersion of a probability distribution. For the regression coefficients w_i above, the coefficient of variation is 14.2%. We therefore decide to write the regression coefficients as approximately equal, i.e., $w_i \approx \kappa_{BT} = -4.87 \times 10^8\ m^2$, with an average error of 14.2%. Using this approximation, we can then rewrite Equation 4 as

$$\hat{\mathbf{S}}_{\mathbf{u}}^{BT} \approx \kappa_{BT}\overline{\nabla} \cdot \begin{pmatrix} \zeta^2 - \zeta D & \zeta\tilde{D} \\ \zeta\tilde{D} & \zeta^2 + \zeta D \end{pmatrix}. \tag{5}$$

The expression now has a single scalar as a tunable parameter, κ_{BT}, which determines the "strength" of the parameterization. The expression depends only on the spatial derivatives of the vorticity and deformation terms and is similar to the parameterization developed by Anstey and Zanna (2017) (see below). In addition, the tensor found is symmetric, despite separately applying the RVM algorithm to the zonal and meridional components of the eddy momentum forcing and without imposing symmetry as a constraint (unlike for the FCNN).

We perform the same procedure using data from the baroclinic model. We provide the RVM algorithm with data from multiple vertical layers at once. As for the barotropic model, a significant increase in the R^2 occurs when three functions are retained, capturing over 40% of the variance. Here, the RVM algorithm constructs the same eddy momentum forcing from the barotropic model (Equation 4), albeit with different values for the regression coefficients. A second increase in the R^2 occurs for larger values of the threshold parameter where five functions are retained, capturing approximately 70% of the variance (Equation S5). We proceed to calculate the average of the regression coefficients and found a mean value $\kappa_{BC} = -8.723 \times 10^8\ m^2$, with a coefficient of variation of 9.8%. Due to the relatively low coefficient of variation, we again assume that all regression coefficients are approximately equal to κ_{BC}, such that the RVM expression can be approximated as

$$\hat{\mathbf{S}}_{\mathbf{u}}^{BC} \approx \kappa_{BC}\overline{\nabla} \cdot \begin{pmatrix} -\zeta D & \zeta\tilde{D} \\ \zeta\tilde{D} & \zeta D \end{pmatrix} + \mathbf{I}\,\frac{1}{2}\,\kappa_{BC}\overline{\nabla}(\zeta^2 + D^2 + \tilde{D}^2), \tag{6}$$

for each vertical layer. The baroclinic expression depends only on the spatial derivatives of the shearing deformation, the stretching deformation, and the vorticity. Like the barotropic expression, the tensor is symmetric. The baroclinic expression can be written as the barotropic expression plus the gradient of the squared deformation terms: $\hat{\mathbf{S}}_{\mathbf{u}}^{BC} \approx 2\hat{\mathbf{S}}_{\mathbf{u}}^{BT} + \mathbf{I}\,\frac{1}{2}\,\kappa_{BC}\overline{\nabla}(D^2 + \tilde{D}^2)$.

For the physical interpretation of the discovered parameterizations, we rely on previous studies (Anstey & Zanna, 2017; Mana & Zanna, 2014; Meneveau & Katz, 2000; Nadiga, 2008; Pope, 1975). The RVM expressions discovered encapsulate the tensor form that a Reynolds stress could take assuming frame invariance

and symmetry in a 2D flow based on Pope (1975). However, not surprisingly, the RVM did not discover the standard viscous stress tensor (also proposed in the framework of Pope, 1975), given that we are mainly learning quasi-geostrophic effects rather than 3D turbulence. Both the expressions for $\hat{\mathbf{S}}_{\mathbf{u}}^{BT}$ and for $\hat{\mathbf{S}}_{\mathbf{u}}^{BC}$ contain within them the recently proposed deformation-based momentum parameterization of Anstey and Zanna (2017), referred to as AZ17, and defined by

$$\hat{\mathbf{S}}_{\mathbf{u}}^{AZ17} = \kappa_{AZ17}\overline{\nabla} \cdot \begin{pmatrix} -\zeta D & \zeta \tilde{D} \\ \zeta \tilde{D} & \zeta D \end{pmatrix}, \tag{7}$$

therefore, $\hat{\mathbf{S}}_{\mathbf{u}}^{BT} = \mathbf{S}_{\mathbf{u}}^{AZ17} + \kappa_{BT}\overline{\nabla}\zeta^2$. AZ17 is also related to the Pope (1975) tensors; see further discussion in AZ17. Using data from a MITgcm baroclinic simulation, AZ17 diagnosed a value of κ_{AZ17} and found it to be on the order of -5×10^8 m^2, similar to the value of κ_{BC}. AZ17 is known to capture up-gradient momentum fluxes and to conserve kinetic energy. The parameterizations, $\mathbf{S}^{AZ17}$, $\hat{\mathbf{S}}_{\mathbf{u}}^{BT}$, and $\hat{\mathbf{S}}_{\mathbf{u}}^{BC}$ can be related the nonlinear gradient model (Meneveau & Katz, 2000; Nadiga, 2008), though comprising of additional terms (AZ17). The nonlinear gradient model, which is derived as a Taylor expansion of the filtered nonlinear stresses, has shown promise in a range turbulent flows applications. This class of parameterizations, based on the deformation tensor of Pope (1975), has also been shown to generalize to different dynamical regimes and scales within a range of eddying resolution (AZ17; Mana & Zanna, 2014; Zanna et al., 2017). The vorticity contribution of each $\hat{\mathbf{S}}_{\mathbf{u}}^{BT}$ and $\hat{\mathbf{S}}_{\mathbf{u}}^{BC}$ is identical to that of $\hat{\mathbf{S}}_{\mathbf{u}}^{AZ17}$ (Supporting Information S4). However, $\hat{\mathbf{S}}_{\mathbf{u}}^{BT}$ and $\hat{\mathbf{S}}_{\mathbf{u}}^{BC}$ lead to a net source or sink of kinetic energy, which depends on the divergence of the flow (or the potential energy of the system; Equation S11). Therefore, the RVM expressions capture processes not included in currently implemented eddy parameterizations and have revealed new parameterizations for energy pathways between reservoirs.

Before implementing the parameterizations in an ocean model, we test their performance offline with the validation data within the barotropic model (Figure 1c). We compare $\hat{\mathbf{S}}_{\mathbf{u}}^{BT}$, $\hat{\mathbf{S}}_{\mathbf{u}}^{AZ}$, and the FCNN trained using the same barotropic model data as for the RVM expression, with the velocity components, $\overline{u}$ and $\overline{v}$ as inputs (Figure 1b and Supporting Information S1).

In the time mean maps of $\hat{S}_x$ (Figure 1c, top row), the RVM expression most accurately captures the spatial patterns of the high-resolution model. The FCNN also captures the majority of the spatial patterns of the true time mean but exhibits a negative bias in the eastern part of the domain. The AZ17 parameterization loosely captures the negative values near the western boundary and positive values in the interior but struggles to capture the finer small-scale patterns of the true time mean. Similar results hold for the standard deviation (middle row): the RVM expression and FCNN reproduce the true standard deviation almost exactly, with differences only visible close to the western boundary, whereas the AZ17 standard deviation underestimates the true standard deviation by 50% in the ocean interior. The higher order moments, skewness, and kurtosis (Supporting Information S8) are also best captured by the RVM expression and FCNN, which outperform the AZ17 expression. In terms of predictive skill, measured by the correlation between the parameterized term and the true subgrid forcing (bottom row), the FCNN captures almost all of the variance in the vicinity of the jet, but this high skill is not consistent across the domain, particularly near the eastern boundary. The predictive skill of the RVM expression is not as high as the FCNN within the jet region but is significantly more consistent across the domain, with fewer patches of zero or negative correlation. AZ17 performs poorly in a significant part of the domain. The amount of data for training the RVM could be reduced by half without deteriorating the results; this is not the case for the FCNN. Performance of the baroclinic momentum expression from RVM can be found in the Supporting Information S1. Overall, the ML parameterizations perform well in offline validation, compared to a physics-based scheme.

3.2. Discovering Eddy Temperature and Energy Forcing

We apply the same procedure to find the eddy temperature forcing, defined by Equation 2 as a flux, using data from the baroclinic model. The basis functions for the eddy temperature forcing are based on derivatives of momentum and temperature. For a given threshold parameter, the R^2 reaches 54.3% with only four functions, resulting in the following expression for the predicted subgrid temperature forcing:

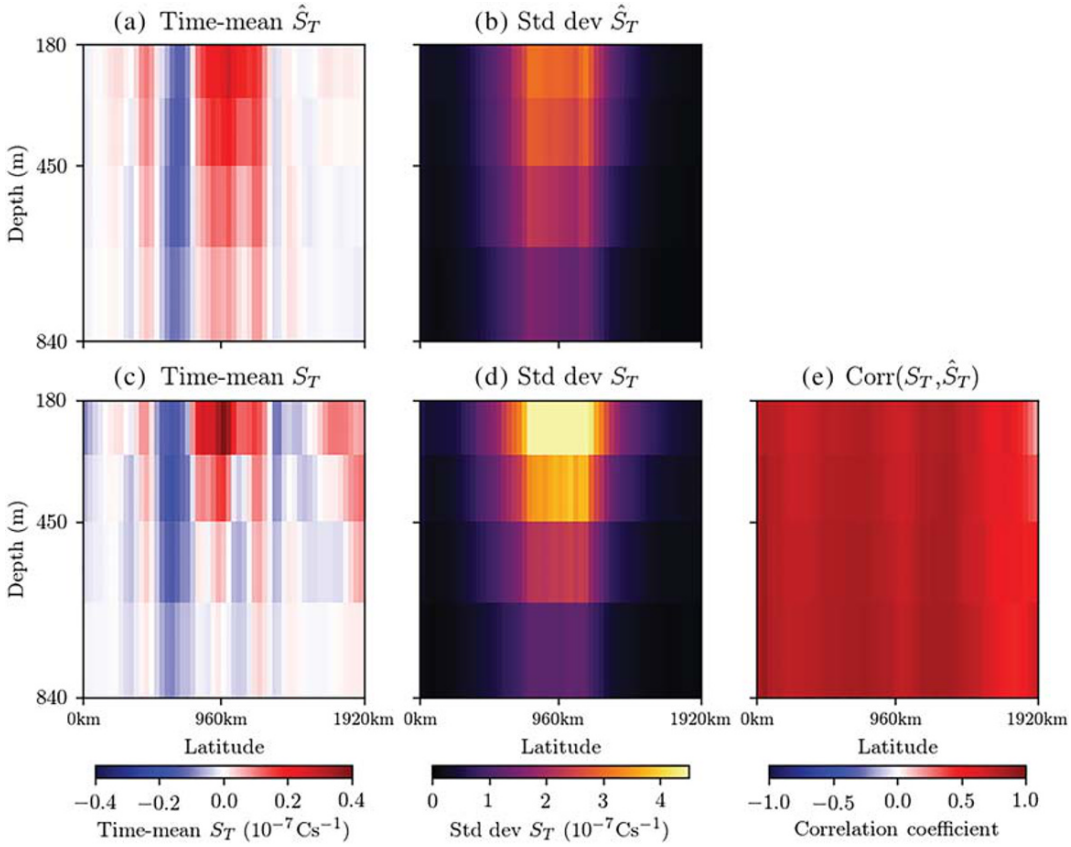

Figure 2. Validation, using the baroclinic model data, of the zonally averaged predicted, $\hat{S}_T$ (Equation 9; panels c and d); against the diagnosed eddy temperature forcing, S_T (Equation 2; panels a and b), as a function of latitude and depth for the mean and standard deviation. Correlation between the prediction and the diagnosed forcing (panel e).

$$\hat{S}_T = w_0(\overline{u}_x\overline{u}_z)_y + w_1(\overline{v}_x\overline{v}_z)_y - w_2(\overline{u}_y\overline{u}_z)_x - w_3(\overline{v}_y\overline{v}_z)_x, \tag{8}$$

with the following values for the regression coefficients $w_0 = 1.573$, $w_1 = 1.495$, $w_2 = 1.518$, and $w_3 = 1.504$, which have units of 10^8 Cms. The mean coefficient value is $\kappa_T = 1.523 \times 10^8$ Cms with a coefficient of variation of 1.7%. Approximating all the regression coefficients as being equal to the mean, with an average error of 1.7%, yields the following expression:

$$\hat{S}_T = \kappa_T \overline{\nabla} \cdot \begin{pmatrix} -\overline{u}_y\overline{u}_z - \overline{v}_y\overline{v}_z \\ \overline{u}_x\overline{u}_z + \overline{v}_x\overline{v}_z \end{pmatrix}. \tag{9}$$

The zonally averaged offline diagnostics for the upper ocean, below the mixed-layer, show that the RVM expression, $\hat{S}_T$, captures the pattern of the mean and standard deviation of the true S_T, however, it underestimates the variance by approximately 50% (Figure 2). The correlation between $\hat{S}_T$ (the prediction) and S_T (the true forcing) is vertically uniform with a value of 0.6. However, near the northern boundary of the domain, the RVM does not capture the pattern nor the amplitude of the true S_T.

The revealed expression is tied to vertical variations in velocity, which is a reflection of the eddy heat fluxes impacting the density field. The dependence of Equation 9 on vertical variability can be examined by assuming that thermal wind balance holds for the mesoscale variability. Using a linear equation of state, we can rewrite Equation 9 as

$$\hat{S}_T = -\frac{\kappa_T g\alpha}{f}\,\overline{\nabla} \cdot \left[\begin{pmatrix} \overline{v}_y & -\overline{u}_y \\ -\overline{v}_x & \overline{u}_x \end{pmatrix} \overline{\nabla}\,\overline{T} \right], \tag{10}$$

where α is the thermal expansion coefficient, g is gravity, and f is the Coriolis parameter. The coefficient

$\kappa_T g\alpha/f$ has units of m^2, similarly to the coefficient for the momentum parameterization. The eddy temperature flux is now dependent on the lateral temperature gradient, modulated by lateral velocity gradients. We can further reformulate the predicted eddy temperature forcing, using the residual-mean formulation (Ferrari & Plumb, 2003; Greatbatch & Lamb, 1990; Marshall et al., 2012), into a vertical flux of horizontal momentum with a magnitude that depends on the velocity gradient (Equation S15)—the flux can be up- or down-gradient.

To further improve the energetics of the model, an additional prognostic equation for the eddy energy can be solved to account for all sources and sinks of energy within the system. However, the prognostic eddy energy equation is unknown and must therefore be constructed (Cessi, 2008; Eden & Greatbatch, 2008; Jansen et al., 2015; Mak et al., 2016; Marshall & Adcroft, 2010). For both the barotropic and baroclinic models, the RVM algorithm constructs a prognostic equation which is the advection of eddy kinetic energy (EKE) and captures 50–60% of the variance in the validation data (Supporting Information S7). Changing the pruning threshold, the target equation, or the spatial scale of the Gaussian filter for defining the eddy scale, did not modify the equation revealed by the algorithm.

4. Implementation Into a Coarse-Resolution Ocean Model

Online performance, meaning when the parameterizations are coupled to a coarser-resolution model, is an important test for future implementation in global climate models. A key issue of any parameterizations is that diagnostic (offline) performance does not translate into prognostic (online) performance due to both the underlying model structure to be integrated forward (e.g., subgrid parameters and numerics) and the nonlinear nature of the equation of motions, in which the parameterizations continuously interact with the resolved scales. Here, the physics-driven parameterization from AZ17, $\hat{\mathbf{S}}^{AZ17}$, and the data-driven barotropic momentum expression (Equation 5) revealed by the RVM, and the data-driven FCNN are implemented into a 30-km resolution version of a very idealized shallow water model (section 2 and Supporting Information S1).It is the first time that a CNN parameterization for ocean turbulence is implemented into an ocean model; therefore, for easy implementation and testing, we chose a model coded in Python. Choosing a model that is different than the model used for learning provides also a stronger (and more difficult) test for the success of the parameterizations. For all three parameterizations, conservation of global momentum and vorticity is satisfied. The goal of the parameterizations is to reduce model biases and in particular energize the flow, to replace the energy lost due to truncation of small scales and large viscosity coefficients at coarse resolution.

We compare the 30-km parameterized simulations, with the 30-km simulation without parameterization and a 3.75-km high resolution (the truth). We initially set the same parameter for both the RVM and AZ17 expressions to $\kappa = -4.87 \times 10^8\ m^2$. However, this implementation led to issues of numerical stability for both the RVM and AZ17 parameterizations, while to the implementation of the FCNN led to overenergized flow, with an efficient inverse cascade and velocities reaching large values of $O(10\ m\ s^{-1})$. To alleviate these issues, we attenuate the strength of each parameterization; i.e., at each time step, we simply multiply $\hat{\mathbf{S}}_\mathbf{u}$ by a coefficient τ between 0 and 1. Through trial and error, we use values of τ of 0.5, 0.5, and 0.7 for the RVM, AZ17, and FCNN parameterizations, respectively (Supporting Information S8).

All three parameterizations increase the amount of kinetic energy in the model (Figure 3a). Both the RVM and AZ17 expressions increase the kinetic energy to approximately halfway between the 30- and 3.75-km models, at a value of $0.038\ m^2\ s^{-2}$. It is not surprising that the RVM and AZ17 parameterizations lead to similar results in a shallow-water barotropic model, as their contributions to the vorticity budget are identical (Supporting Information S4). The FCNN parameterizations increase the kinetic energy of the model to within approximately 5% of the high-resolution model at $0.056\ m^2\ s^{-2}$.

The kinetic energy power spectrum (Figure 3b) shows evidence of increased kinetic energy for the parameterized simulations at all spatial scales, compared to the low-resolution unparameterized simulation. At spatial scales larger than 300–400 km, all parameterizations increase the kinetic energy to approximately the same level as the high-resolution simulation, therefore implying a more efficient backscatter or inverse energy cascade. The FCNN parameterization increases the kinetic energy to above that of the

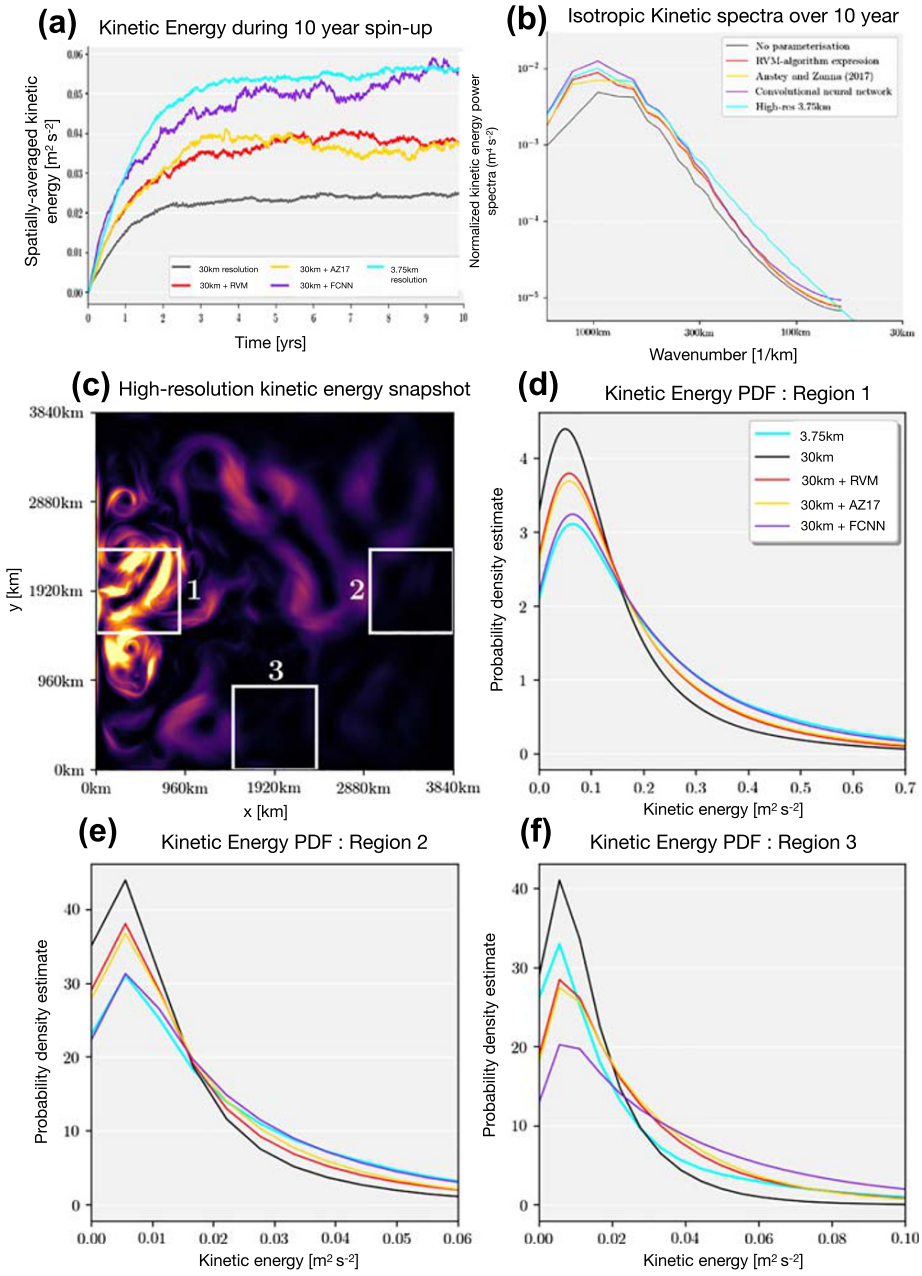

Figure 3. Kinetic energy diagnostics in the following idealized ocean simulation: high-resolution, 3.75 km (cyan), coarse-resolution 30 km without parameterizations (gray), coarse-resolution 30 km with FCNN (purple), coarse-resolution 30 km with RVM (red), and coarse-resolution 30 km with AZ17 (yellow). (a) Time series of globally averaged kinetic energy as a function of time; (b) kinetic energy spectrum as a function of wavenumber; (c) snapshot of kinetic energy in the high-resolution simulations, indicated three regions of interest (1–3) for extreme event diagnostics using probability distribution functions (PDF). (e and d) PDF of kinetic energy for Regions 1–3.

high-resolution model. At length scales smaller than 300 km, while all parameterizations increase the kinetic energy, it remains lower than that of the high-resolution simulation, likely due to viscosity.

In addition to the global mean kinetic energy, we consider the impact of the parameterizations on the statistics and extremes in kinetic energy. In the three representative regions selected (Figure 3c), the high-resolution probability density function (PDF) has more probability in the tails compared to the 30-km model without parameterization (Figures 3d–3f). The effect of all parameterizations is to increase

the probability in the tails, with little shift in the position of the peak. Therefore, the primary effect of the parameterizations is increasing the frequency of extreme kinetic energy values, as opposed to solely increasing the mean kinetic energy. In Regions 1 and 2 (Figures 3d and 3e), the FCNN is the best performing, with the kinetic energy PDF of the FCNN parameterization almost indistinguishable from the high-resolution model. AZ17 and the RVM expressions are almost indistinguishable from each other. However, in Region 3 (Figure 3f), all three parameterizations cause too much probability to be redistributed in the tails, as evident by the peaks of the RVM, AZ17, and FCNN kinetic energy PDFs all being below the high-resolution peak.

5. Summary

ML algorithms can facilitate the discovery of physical processes, embedded within data from high-resolution simulations or observations. However, physical intuition remains critical to explain the physics discovered by these algorithms. We have introduced the data-driven equation discovery method of Zhang and Lin (2018), namely, the RVM algorithm, for ocean eddy parameterizations, rather than for discovering fundamental equations of motions already known (Rudy et al., 2017). The mathematical expressions discovered by the RVM algorithm show that eddy momentum parameterizations should include up-gradient momentum fluxes and potentially a transfer between potential energy and kinetic energy. In addition, the RVM revealed that eddy temperature fluxes can act on vertical gradients of horizontal momentum with a magnitude that depends on the velocity gradient and that eddy energy advection accounts for half of the time tendency of EKE. A CNN, constrained with physical conservation laws, appears to be an excellent representation of the eddy momentum forcing, leading to vastly improved coarser-resolution simulations which, under certain metrics, are indistinguishable from the high-resolution target, confirming results from Bolton and Zanna (2019). Yet the reasons for the success of the CNN parameterization are difficult to extract. All parameterizations presented here have been shown to generalize well to other regimes (e.g., dynamical regions and Reynolds numbers or resolution; Anstey & Zanna, 2017; Bolton & Zanna, 2019; Pope, 1975; Mana & Zanna, 2014; Holm & Wingate, 2005). Unfortunately, the parameterizations presented here are also subject to tuning when implemented in an ocean model, as all parameterizations in use in current climate models are. The parameterizations, when implemented in a very idealized model, did not vastly improve the mean state (Figures S10 and S11), but tests in more complex models have showed that they have the ability to do so (Zanna et al., 2017). Here, the RVM (and the physics-based) expression, which performs well offline, does not show as good performance as the FCNN online due to numerical instabilities developing during the implementation. This result suggests that the complexity of a deep neural network may be more numerically stable compared to implementing a closed-form equation (Rasp et al., 2018), yet it is subject to heavy tuning (Brenowitz & Bretherton, 2019). However, we cannot rule out that the RVM results could be improved by adding more functions or by adding memory or stochasticity, which have been shown to drastically improve stability (Zanna et al., 2017), or finally by coupling the momentum parameterization to an eddy energy equation (Jansen & Held, 2014).

While the implementation of eddy forcing remains to be properly tested in more complex models, our results suggest that progress can be made using ML for physics discovery and interpretable parameterizations, which are more computationally efficient than running high-resolution simulations (Figure S12). We hope that this manuscript provide a new road map for data-driven parameterizations to be developed, tested, interpreted, and implemented in ocean climate models in the future. A new strategy, which combines the interpretability of equation discovery with the predictive skill of complex neural networks, could be an effective approach to improving ocean models and perhaps climate models in general.

Data Availability Statement

The code for the relevance vector machine algorithm can be found at github.com/TomBolton/rvm-find, https://doi.org/10.5281/zenodo.3758659.

Acknowledgments
We thank Alistair Adcroft for insightful discussions and Malte Jansen, Redouane Lguensat, and Pedram Hassanzadeh for their helpful reviews which helped improve the manuscript. The work was partially supported by NERC, NSF-GEO 1912357 and NOAA CVP NA19OAR4310364.

References

Anstey, J. A., & Zanna, L. (2017). A deformation-based parametrization of ocean mesoscale eddy reynolds stresses. *Ocean Modelling*, *112*, 99–111.
Bachman, S. D. (2019). The GM+ E closure: A framework for coupling backscatter with the gent and McWilliams parameterization. *Ocean Modelling*, *136*, 85–106.

Berloff, P. S. (2005). Random-forcing model of the mesoscale oceanic eddies. *Journal of Fluid Mechanics*, *529*, 71–95.
Beucler, T., Pritchard, M., Rasp, S., Gentine, P., Ott, J., & Baldi, P. (2019). Enforcing analytic constraints in Neural-Networks emulating physical systems. arXiv preprint arXiv:1909.00912.
Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer. ISBN: 9780387310732.
Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, *11*, 376–399. https://doi.org/10.1029/2018MS001472
Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*, 6289–6298. https://doi.org/10.1029/2018GL078510
Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network parametrization trained by Coarse-Graining. *Journal of Advances in Modeling Earth Systems*, *11*, 2728–2744. https://doi.org/10.1029/2019MS001711
Cessi, P. (2008). An energy-constrained parameterization of eddy buoyancy flux. *Journal of Physical Oceanography*, *38*(8), 1807–1819.
Cooper, F. C., & Zanna, L. (2015). Optimisation of an idealised ocean model, stochastic parameterisation of sub-grid eddies. *Ocean Modelling*, *88*, 38–53.
Danabasoglu, G., McWilliams, J. C., & Gent, P. R. (1994). The role of mesoscale tracer transports in the global ocean circulation. *Science*, *264*(5162), 1123–1126.
Eden, C., & Greatbatch, R. J. (2008). Towards a mesoscale eddy closure. *Ocean Modelling*, *20*(3), 223–239.
Ferrari, R., & Plumb, R. A. (2003). Residual circulation in the ocean. In *Near-boundary processes and their parameterization: Proc.Aha Hulikoaa Hawaiian Winter Workshop*, Citeseer (pp. 219–228). https://pdfs.semanticscholar.org/4a99/7fb28d0aeba7b68-ca1279e898d043fbdbf57.pdf?_ga=2.169950691.255983052.1597535352-1774259145.1597535352
Ferrari, R., & Wunsch, C. (2009). Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics*, *41*, 253–282.
Gent, P. R., & Mcwilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, *20*(1), 150–155.
Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, *45*, 5742–5751. https://doi.org/10.1029/2018GL078202
Greatbatch, R. J., & Lamb, K. G. (1990). On parameterizing vertical mixing of momentum in non-eddy resolving ocean models. *Journal of Physical Oceanography*, *20*(10), 1634–1637.
Holm, D. D., & Wingate, B. A. (2005). Baroclinic instabilities of the two-layer quasigeostrophic alpha model. *Journal of Physical Oceanography*, *35*(7), 1287–1296.
IPCC (2013). Summary for policymakers. In Stocker, T. F. et al., (Eds.), *Climate change 2013: The physical science basis. Contribution of working group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1–30). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004
Jansen, M. F., & Held, I. M. (2014). Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Modelling*, *80*, 36–48.
Jansen, M. F., Held, I. M., Adcroft, A., & Hallberg, R. (2015). Energy budget-based backscatter in an eddy permitting primitive equation model. *Ocean Modelling*, *94*, 15–26.
Kjellsson, J., & Zanna, L. (2017). The impact of horizontal resolution on energy transfers in global ocean models. *Fluids*, *2*(3), 45.
Kuhlbrodt, T., & Gregory, J. M. (2012). Ocean heat uptake and its consequences for the magnitude of sea level rise and climate change. *Geophysical Research Letters*, *39*, L18608. https://doi.org/10.1029/2012GL052952
Kutz, J. N. (2017). Deep learning in fluid dynamics. *Journal of Fluid Mechanics*, *814*, 1–4.
Ling, J., Kurzawski, A., & Templeton, J. (2016). Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics*, *807*, 155–166.
Mak, J., Maddison, J. R., & Marshall, D. P. (2016). A new gauge-invariant method for diagnosing eddy diffusivities. *Ocean Modelling*, *104*, 252–268.
Mana, P. P., & Zanna, L. (2014). Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Modelling*, *79*, 1–20.
Marshall, D. P., & Adcroft, A. (2010). Parameterization of ocean eddies: Potential vorticity mixing, energetics and arnolds first stability theorem. *Ocean Modelling*, *32*(3-4), 188–204.
Marshall, J., Adcroft, A., Hill, C., Perelman, L., & Heisey, C. (1997). A finite-volume, incompressible navier stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research*, *102*(C3), 5753–5766.
Marshall, D. P., Maddison, J. R., & Berloff, P. S. (2012). A framework for parameterizing eddy potential vorticity fluxes. *Journal of Physical Oceanography*, *42*(4), 539–557.
Maulik, R., & San, O. (2017). A neural network approach for the blind deconvolution of turbulent flows. *Journal of Fluid Mechanics*, *831*, 151–181.
Meneveau, C., & Katz, J. (2000). Scale-Invariance and turbulence models for large-eddy simulation. *Annual Review of Fluid Mechanics*, *32*(1), 1–32. https://doi.org/10.1146/annurev.fluid.32.1.1
Nadiga, B. T. (2008). Orientation of eddy fluxes in geostrophic turbulence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1875), 2489–2508.
O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*, 2548–2563. https://doi.org/10.1029/2018MS001351
Pope, S. B. (1975). A more general effective-viscosity hypothesis. *Journal of Fluid Mechanics*, *72*(2), 331–340.
Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689.
Rudy, S. H., Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2017). Data-driven discovery of partial differential equations. *Science Advances*, *3*(4), e1602614.
Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*, 12,396–12,417. https://doi.org/10.1002/2017GL076101
Smagorinsky, J. (1963). General circulation experiments with the primitive equations: I. The basic experiment. *Monthly Weather Review*, *91*(3), 99–164.
Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, *1*(Jun), 211–244.
Wang, R., Kashinath, K., Mustafa, M., Albert, A., & Yu, R. (2020). Towards physics-informed deep learning for turbulent flow prediction. Proceedings of the 26th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2020.
Warner, T. T. (2010). *Numerical weather and climate prediction*. Cambridge University Press. ISBN: 978-0-521-51389-0. Hardback, 526 PP. https://doi.org/10.1017/CBO9780511763243
Zanna, L., Bachman, S., & Jansen, M. (2020). Energizing turbulence closures in ocean models. *CLIVAR Exchanges/US CLIVAR Variations*, *18*(1), 3–8. https://doi.org/10.5065/g8w0-fy32

Zanna, L., Mana, P. P., Anstey, J., David, T., & Bolton, T. (2017). Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction. *Ocean Modelling*, *111*, 66–80.

Zhang, S., & Lin, G. (2018). Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *474*(2217), 20180305.