

Symbiotic Ocean Modeling using Physics-Controlled Echo State Networks

T.E. Mulder^{1,2}, S. Baars¹, F.W. Wubs¹, F.I. Pelupessy³, M. Verstraaten³, and
H.A. Dijkstra^{4,5}

¹Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen,
the Netherlands.

²Swedish Meteorological and Hydrological Institute, Norrköping, Sweden.

³Netherlands eScience Center, Amsterdam, the Netherlands.

⁴Institute for Marine and Atmospheric research Utrecht, Department of Physics, Utrecht University,
Utrecht, the Netherlands.

⁵Center for Complex Systems Studies, Utrecht University, Utrecht, the Netherlands.

Key Points:

- We demonstrate part of a symbiotic ocean modeling framework where models of different complexities benefit from each other.
- Unresolved processes are represented through hybrid machine learning methods using data from the symbiotic framework.
- Hybrid correction strategies with imperfect physics as control input improve the representation of key long-term flow properties.

Corresponding author: T.E. Mulder, erik.mulder@smhi.se

19 **Abstract**

20 We introduce the concept of ‘symbiotic’ ocean modeling where high- and low-resolution
 21 dynamical models coexist and benefit from each other through data-driven improvements.
 22 In this work we specifically focus on how a low-resolution model may benefit from such
 23 a symbiotic setup. The broader aim is to improve the efficiency of high-resolution mod-
 24 els, while simultaneously enhancing the representation of unresolved processes in low-
 25 resolution models. To achieve a symbiosis we use a grid-switching approach together with
 26 hybrid modeling techniques that combine linear regression-based methods with nonlin-
 27 ear echo state networks (ESNs). The approach is applied to both the Kuramoto–Sivashinsky
 28 equation and a single-layer quasi-geostrophic ocean model, and shown to simulate short-
 29 term and long-term behavior better than either purely data-based methods or low-resolution
 30 models.

31 **Plain Language Summary**

32 Models of the ocean vary in complexity. Some are very detailed and manage to show
 33 oceanic vortices, whereas others are very efficient but coarse, and unable to compute such
 34 vortices. The idea in this paper is to let these different model types work together, as
 35 if in a symbiosis. With knowledge of differences between the detailed and coarse model
 36 we can use machine learning techniques to improve the coarse model, while a coarse model
 37 can be used to aid a detailed model computationally. Here we focus on the former part
 38 and perform numerous experiments to test different kinds of coarse model improvements.
 39 We apply our ideas to the Kuramoto–Sivashinsky (KS) model and a quasi-geostrophic
 40 (QG) ocean model, where we show that promising short-term KS results may general-
 41 ize to models of the ocean. Long-term equilibrium experiments with QG show in addi-
 42 tion how the correction strategies let a coarse model produce correct flow properties, where
 43 standalone physics- or data-based approaches fail.

44 **1 Introduction**

45 One of the most important spatial scales in the ocean circulation is the internal Rossby
 46 radius of deformation L_D ; it ranges from 50-100 km at midlatitudes to a few km in the
 47 polar regions (Hallberg, 2013). At this scale, perturbations are amplified on mean flows
 48 through mixed barotropic/baroclinic instability, giving rise to ocean eddies. Interactions
 49 between these eddies and the mean flow can lead to upgradient momentum transport
 50 affecting the strength and separation of ocean western boundary currents such as the Kuroshio
 51 and Agulhas (Chassignet et al., 2020).

52 Most climate models, in particularly those used in CMIP5 and CMIP6, do not re-
53 solve ocean processes at the scale L_D as the spatial grid size used is too large, e.g. typ-
54 ically 1° (Eyring et al., 2016). The main reason is computational costs, as doubling the
55 horizontal resolution increases these costs roughly by a factor 10. Effects of subgrid-scale
56 processes are hence parameterized in these models. For example, the effect of ocean ed-
57 dies on tracer transport is represented by the Gent–McWilliams (Gent et al., 1995) scheme,
58 but such a scheme cannot capture, for example, the upgradient momentum transport.
59 Hence, western boundary flows are too weak and diffuse, and do not separate at the cor-
60 rect location (Chassignet et al., 2020).

61 Over the last few years, first simulations have been performed with global climate
62 models, where the ocean model component has a resolution of 0.1° , which is smaller than
63 L_D for many locations on the globe (Chang et al., 2020; Jüling et al., 2021). We will re-
64 fer to those models as high-resolution (HR) models to contrast them with the 1° mod-
65 els which we will call low-resolution (LR) models. But also the high-resolution models
66 are not completely eddy-resolving as this requires an even higher spatial resolution. There
67 is now a substantial amount of model data available to compare results on ocean-climate
68 variability and climate change for both types of models. Clearly, high-resolution mod-
69 els reduce biases compared to observations particularly in western boundary currents,
70 sea surface temperature variability patterns and Southern Ocean mean flows (Chang et
71 al., 2020; Jüling et al., 2021).

72 However, HR model simulations form a great drain on computational resources and
73 hence there are still many efforts to represent the effects of unresolved processes in LR
74 models. This parameterization process has been around for decades and approaches can
75 be grouped into three types. First, semi-empirical parameterizations are used, where ob-
76 servation motivated schemes are implemented (Gargett, 1989; Viebahn et al., 2019). Sec-
77 ond, theoretically derived schemes, where specific approximations are made in the un-
78 derlying equations (Gent et al., 1995) have been used. Third, stochastic schemes derived
79 from sample high-resolution model simulations (Berloff, 2005; Mana & Zanna, 2014) have
80 shown potential in representing unresolved processes in LR models (Hewitt et al., 2020).

81 To this, recently a new approach has been added, where the subgrid-scale model
82 is derived from a machine learning (ML) model, such as a neural network. In Bolton &
83 Zanna (2019), a convolutional neural network (CNN) was trained with data from a high-
84 resolution model of the midlatitude gyres. This CNN was shown to successfully capture
85 the small-scale processes and the effects of those on the mean flow in the low-resolution
86 version of the same model. Traditional feedforward neural network models (FFNN) have
87 also been used as subgrid-scale representations in both ocean and atmospheric models

88 (Irrgang et al., 2021; Rasp et al., 2018). Another ML technique that shows promise in
 89 the modeling of climate physics is the reservoir computing approach, often referred to
 90 as an echo state network (ESN). An ESN is a type of recurrent neural network (RNN)
 91 that is especially suited to simulate chaotic dynamics (Jaeger & Haas, 2004; Pathak et
 92 al., 2017) and is shown to be capable of emulating EOF interactions (Nadiga, 2021). Where
 93 FFNNs are generally regarded as functions, RNNs can be seen as artificial dynamical
 94 systems (Lukoševičius & Jaeger, 2009). Computationally there are close relations be-
 95 tween ESN-based methods, linear regression and models based on a dynamic mode de-
 96 composition (DMD) (Schmid, 2010; Kutz et al., 2016). Theoretical connections between
 97 the ESN approach, DMD and also vector autoregression (VAR) have been explored in
 98 Bollt (2021).

99 Recent ‘hybrid’ (or physics-controlled) ESN advances (Pathak et al., 2017, 2018)
 100 provide an elegant approach to correct known model imperfections, such as those due
 101 to the lack of eddies in LR ocean models. With training data based on ground truths
 102 and imperfect model predictions, model tendencies and nonlinear model mismatches are
 103 encoded in an ESN. The result is an artificial dynamical system that can be controlled
 104 using an imperfect model. Combining an imperfect model with corrections from a trained
 105 ESN creates a hybrid dynamical system that greatly outperforms both the network and
 106 the imperfect model (Wikner et al., 2020). This approach was recently applied to an at-
 107 mospheric model (SPEEDY) and shown to be able to improve the simulations of mean
 108 flow and variability considerably on short time scales (Arcomano et al., 2022).

109 In this paper we use the hybrid modeling framework as key ingredient for a ‘sym-
 110 biotic’ ocean modeling approach. The idea is to couple models of different complexities
 111 and configure them to solve the same problem, where we distinguish between perfect and
 112 imperfect models in terms of differing resolution and parameterizations. This model co-
 113 existence can be made mutually beneficial using data-driven techniques. With the sym-
 114 biotic approach we aim to improve the computational efficiency of HR models, while si-
 115 multaneously enhancing the parameterizations of unresolved processes in LR models. We
 116 will focus on the latter part and use the hybrid modeling strategy with data generated
 117 from both LR and HR models to correct imperfect model transients. Model corrections
 118 made can then be seen as modeling subgrid effects. To this end, we employ a grid-switching
 119 approach and introduce a correction framework that includes models based on linear re-
 120 gression, DMD, ESN and hybrid variants (Section 2). We apply the correction strategy
 121 to coupled LR and HR versions of the Kuramoto–Sivashinsky equations (Section 3) and
 122 a single-layer quasi-geostrophic ocean model (Section 4). Both short-term predictions
 123 and long-term equilibrium runs are performed with the ocean model to compare the avail-

124 able corrective models. A summary and discussion with the main conclusions is provided
 125 in Section 5.

126 2 Methodology

127 In a general framework, the HR model is defined on a fine grid Ω^f and is regarded
 128 as a *perfect* model. An LR model is considered as an *imperfect* model, and is defined on
 129 a coarse grid Ω^c . The grids Ω^f and Ω^c have dimensions N_f and N_c , respectively, and
 130 cover the same domain. Both models attempt to solve the same problem, but apart from
 131 different grids we also allow differences in key parameters and forcings between the per-
 132 fect and imperfect model. The physics resolved by the perfect model is then used as ground
 133 truth and the imperfect model results are considered to be in need of correction.

134 The perfect model is a system of coupled partial differential equations (PDEs), spa-
 135 tially discretized on Ω^f , which leads to a large system of differential-algebraic equations
 136 (DAEs):

$$137 \quad M_P \dot{\boldsymbol{\xi}} = F_P(\boldsymbol{\xi}), \quad \text{with } \boldsymbol{\xi} \in \mathbb{R}^{N_f}. \quad (1)$$

138 Here, $\boldsymbol{\xi} = \boldsymbol{\xi}(t)$ is a time dependent state vector and $M_P \in \mathbb{R}^{N_f \times N_f}$ is a mass matrix
 139 that determines the dependence on temporal derivatives. The nonlinear operator $F_P : \mathbb{R}^{N_f} \rightarrow \mathbb{R}^{N_f}$
 140 is a spatial discretization of the perfect model physics. Similarly, the semi-
 141 discretized imperfect model has a coarse state $\mathbf{x} = \mathbf{x}(t)$ that evolves according to

$$142 \quad M_I \dot{\mathbf{x}} = F_I(\mathbf{x}), \quad \text{with } \mathbf{x} \in \mathbb{R}^{N_c}, \quad (2)$$

143 where $M_I \in \mathbb{R}^{N_c \times N_c}$ and $F_I : \mathbb{R}^{N_c} \rightarrow \mathbb{R}^{N_c}$ are again the mass matrix and spatial dis-
 144 cretization operator. For simplicity we only consider models in this form (equations (1)
 145 and (2)), but the methodology explained here is not restricted to this formulation.

146 Transfers between the solutions on the two grids Ω^f and Ω^c are made through a
 147 fully weighted restriction $R \in \mathbb{R}^{N_c \times N_f}$ and a prolongation operator $P \in \mathbb{R}^{N_f \times N_c}$. We
 148 choose these operators for their convenient (variational) property that they are each other's
 149 transpose up to a constant factor: $R = cP^\top$ (Briggs et al., 2000). The perfect model
 150 evolves according to $\phi_P : \mathbb{R} \times \mathbb{R}^{N_f} \rightarrow \mathbb{R}^{N_f}$. Similarly, the evolution of the imperfect
 151 model is given by $\phi_I : \mathbb{R} \times \mathbb{R}^{N_c} \rightarrow \mathbb{R}^{N_c}$. The evolution operator ϕ_I solves for the tran-
 152 sient state $\tilde{\mathbf{x}}^{k+1}$ according to a certain time-discretization and hence $\tilde{\mathbf{x}}^{k+1} = \phi_I(\mathbf{x}^k)$.

153 The imperfect spatial discretization F_I is incapable of capturing the physics resolved
 154 by the perfect model and we therefore attempt to improve the imperfect evolution ϕ_I
 155 with a combination of linear and non-linear corrections. We employ an auxiliary (sur-
 156rogate) model f with auxiliary state $\mathbf{s} \in \mathbb{R}^{N_r}$ of size N_r , that is forced by imperfect and

157 improved predictions. As these corrections are data-driven we divide our approach into
 158 a data gathering and a prediction phase.

159 2.1 Data gathering

160 We gather data from a trajectory of $\boldsymbol{\xi}(t)$ on Ω^f . From this transient, associated
 161 restricted states, imperfect predictions and auxiliary states are computed. Starting at
 162 time t_0 , we collect $N_T + 1$ snapshots of the evolving state $\boldsymbol{\xi}(t)$:

$$163 \quad \left\{ \boldsymbol{\xi}^0, \boldsymbol{\xi}^1, \dots, \boldsymbol{\xi}^{N_T} \right\}, \quad \boldsymbol{\xi}^k = \boldsymbol{\xi}(t_0 + k\Delta t), \quad (3)$$

164 at fixed time intervals Δt such that we cover the model time $T = N_T\Delta t$. The snap-
 165 shots are restricted to the coarse grid and combined into two data matrices:

$$166 \quad X = [\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^{N_T-1}] = [R\xi^0, R\xi^1, \dots, R\xi^{N_T-1}], \quad (4)$$

$$167 \quad X' = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{N_T}] = [R\xi^1, R\xi^2, \dots, R\xi^{N_T}]. \quad (5)$$

168 Apart from the restricted data matrix $X \in \mathbb{R}^{N_c \times N_T}$ and its shifted version $X' \in \mathbb{R}^{N_c \times N_T}$,
 169 we also create a collection of imperfect predictions $\Phi(X)$:

$$170 \quad \Phi(X) = [\phi_I(\mathbf{x}^0), \phi_I(\mathbf{x}^1), \dots, \phi_I(\mathbf{x}^{N_T-1})] \in \mathbb{R}^{N_c \times N_T}. \quad (6)$$

171 The elements of X and $\Phi(X)$ serve as forcing to the auxiliary model f , which we evolve
 172 and gather snapshots from. We iterate according to

$$173 \quad \mathbf{u}^k = h(\mathbf{x}^k, \phi_I(\mathbf{x}^k)), \quad (7)$$

$$174 \quad \mathbf{s}^{k+1} = f(\mathbf{s}^k, \mathbf{u}^k), \quad (8)$$

175 with combined input \mathbf{u} given by a mapping h and with initialization $\mathbf{s}^0 = \mathbf{s}_0$ at $t =$
 176 t_0 . For h we either use a selection, e.g. $h(\mathbf{x}^k, \tilde{\mathbf{x}}^{k+1}) = \mathbf{x}^k$ or combine the forcing such
 177 that $h(\mathbf{x}^k, \tilde{\mathbf{x}}^{k+1}) = (\mathbf{x}^k; \tilde{\mathbf{x}}^{k+1})$, where $(\ ;)$ denotes vertical stacking. These are the
 178 most straightforward choices and of course other options are possible here. The surro-
 179 gate model f comes in the form of an Echo State Network (ESN) and is described in Sec-
 180 tion 2.3. From the evolution of f we gather N_T+1 state snapshots $\mathbf{s}^0, \dots, \mathbf{s}^{N_T}$ and com-
 181 bine them into a data matrix, with the exception of the initialization \mathbf{s}^0 :

$$182 \quad S = [\mathbf{s}^1, \mathbf{s}^2, \dots, \mathbf{s}^{N_T}] \in \mathbb{R}^{N_r \times N_T}. \quad (9)$$

183 2.2 Prediction

184 The data gathered up until time $t = t_0 + T$ is used to obtain linear best fit oper-
 185 ators. Given data $X, X', \Phi(X)$ and S , these operators optimally combine $\mathbf{x}, \phi_I(\mathbf{x})$ and

186 **s** to improve the imperfect evolution given by ϕ_I alone. Here we provide a general tran-
 187 sient strategy that covers a number of different corrective methods.

188 A corrected imperfect transient is started at $t_0 + T$. Now, the models ϕ_I and f
 189 operate in isolation from any perfect model data and f augments ϕ_I . Using starting states
 190 \mathbf{x}^{N_T} and \mathbf{s}^{N_T} , the transient proceeds as follows:

191 $\tilde{\mathbf{x}}^{k+1} = \phi_I(\mathbf{x}^k)$ create an imperfect model prediction, (10)

192 $\mathbf{u}^k = h(\mathbf{x}^k, \tilde{\mathbf{x}}^{k+1})$ construct a forcing, (11)

193 $\mathbf{s}^{k+1} = f(\mathbf{s}^k, \mathbf{u}^k)$ evolve the auxiliary state, (12)

194 $\mathbf{x}^{k+1} = A\mathbf{x}^k + B\tilde{\mathbf{x}}^{k+1} + C\mathbf{s}^{k+1}$ create an improved prediction, (13)
 195

196 for $k = N_T, N_T+1, \dots$. Hence the trajectory of \mathbf{x} is initialized with a restricted truth
 197 ($\mathbf{x}^{N_T} = R\xi^{N_T}$) but continues independently of the perfect model ($\mathbf{x}^{N_T+1} \neq R\xi^{N_T+1}$).

198 With the general formulation in (10)-(13) we aim to include several methods and
 199 their combinations in the same framework. The operators A, B, C have separate inter-
 200 pretations. On its own, A is obtained as a linear best fit of the propagation from X to
 201 X' . Its eigendecomposition is known as a dynamic mode decomposition (DMD) (Schmid,
 202 2010; Kutz et al., 2016) and A is often called a DMD-operator. The matrix B is the best
 203 direct correction of $\Phi(X)$ to X' in the least squares sense. Lastly, as f is a neural net,
 204 the operator C is the optimal output layer, i.e., the linear best fit translation of S to X' .

205 Hence these different methods can be seen as special cases in (10)-(13).

206 Combinations of the operators A, B and C are fitted at $t = t_0 + T$ using regu-
 207 larized linear regressions with the data matrices $X, X', \Phi(X)$ and S . Choices for the ar-
 208 chitecture of f and h and the use of operators A, B, C lead to a variety of predictive meth-
 209 ods (Table 1). A *model only* approach uses $B = I$ and ignores A and C . The transient
 210 (10)-(13) is reduced to only the imperfect model evolution. In an *ESN* prediction we trans-
 211 late from states of the neural net (ESN) to predictions using a best fit C . Here, f is forced
 212 with restricted states only: \mathbf{x}^k . A *DMD* prediction is based on the best linear approx-
 213 imation of the propagation from X to X' . When the operators B and C are combined
 214 and $\mathbf{u}^k = (\mathbf{x}^k, \tilde{\mathbf{x}}^{k+1})$, the auxiliary model f is subjected to a physics-based control $\phi_I(\mathbf{x})$,
 215 both internally through \mathbf{u} and externally through B . With f an ESN this is referred to
 216 as *ESNc*, which is equivalent to the hybrid scheme in Pathak et al. (2018). *DMDc* de-
 217 notes DMD with control (Proctor et al., 2016) and is obtained by combining operators
 218 A and B . In DMDc the imperfect physics assist the DMD model which, on its own, gen-
 219 eralizes poorly outside the training data. A basic *correction-only* approach follows from
 220 using only B , whereas additional combinations lead to the varieties *ESN+DMD* and *ESN+DMDc*.

Table 1: Overview of corrective methods based on operator configurations in (13). The associated minimizations are linear regression problems for which we do not include the regularization here; $\|\cdot\|_F$ is the Frobenius norm. Additional variations on these methods rely on the specific architecture chosen for h and f . The choices we make for h are added as a separate column to this table.

Method	$h(\mathbf{x}^k, \tilde{\mathbf{x}}^{k+1})$	Operator choices	Minimization to compute operators
Model only		$A = 0, B = I, C = 0$	No minimization necessary
ESN	\mathbf{x}^k	$A = 0, B = 0, C = ?$	$\min_C \left\ CS - X' \right\ _F$
DMD		$A = ?, B = 0, C = 0$	$\min_A \left\ AX - X' \right\ _F$
ESNc	$\begin{bmatrix} \mathbf{x}^k \\ \tilde{\mathbf{x}}^{k+1} \end{bmatrix}$	$A = 0, B = ?, C = ?$	$\min_{[B \ C]} \left\ \begin{bmatrix} B & C \end{bmatrix} \begin{bmatrix} \Phi(X) \\ S \end{bmatrix} - X' \right\ _F$
DMDc		$A = ?, B = ?, C = 0$	$\min_{[A \ B]} \left\ \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X \\ \Phi(X) \end{bmatrix} - X' \right\ _F$
Correction-only		$A = 0, B = ?, C = 0$	$\min_B \left\ B\Phi(X) - X' \right\ _F$
ESN+DMD	\mathbf{x}^k	$A = ?, B = 0, C = ?$	$\min_{[A \ C]} \left\ \begin{bmatrix} A & C \end{bmatrix} \begin{bmatrix} X \\ S \end{bmatrix} - X' \right\ _F$
ESN+DMDc	$\begin{bmatrix} \mathbf{x}^k \\ \tilde{\mathbf{x}}^{k+1} \end{bmatrix}$	$A = ?, B = ?, C = ?$	$\min_{[A \ B \ C]} \left\ \begin{bmatrix} A & B & C \end{bmatrix} \begin{bmatrix} X \\ \Phi(X) \\ S \end{bmatrix} - X' \right\ _F$

221 Connections between ESNs and DMD exist (Boltt, 2021) and within this framework it
 222 is straightforward to combine (and consequently isolate) both approaches.

223 The minimizations shown in Table 1 are computed using Tikhonov regularization,
 224 which introduces an additional penalty on the size of the fitted operator. Regularization
 225 is crucial as it reduces overfitting and improves the stability of a long-term transient (Luko-
 226 sevicus, 2012). For instance, the DMD-operator actually minimizes

$$227 \min_A \left(\left\| AX - X' \right\|_F + \lambda^2 \left\| A \right\|_F \right), \quad (14)$$

228 with $\lambda > 0$ a regularization parameter.

229

2.3 Echo State Network

230

231

232

233

234

235

236

237

238

239

240

241

An echo state network (Jaeger, 2001; Jaeger & Haas, 2004) will act as the auxiliary predictive model f . Here we will roughly outline the organization of an ESN. For a detailed explanation we refer to Pathak et al. (2018), which we follow closely. An ESN is a recurrent neural network that can be viewed as an artificial nonlinear dynamical system with a state $\mathbf{s} \in \mathbb{R}^{N_r}$ of sufficient dimension N_r . The components of \mathbf{s} interact through a sparse, random linear operator $W : \mathbb{R}^{N_r} \rightarrow \mathbb{R}^{N_r}$ that is not altered after initialization. The average degree of the adjacency graph associated with W is denoted with \bar{d} . Input data $\mathbf{u} \in \mathbb{R}^{N_u}$ is standardized (every unknown has zero mean and unit variance) and is fed as forcing to the system, where it is combined with the state using a fixed linear operator $W_{in} : \mathbb{R}^{N_u} \rightarrow \mathbb{R}^{N_r}$. The input operator W_{in} is random and sparse, with only a single element per row that is drawn from a uniform distribution on $[-1, 1]$. The internal state evolves according to

242

$$\mathbf{s}^{k+1} = f(\mathbf{s}^k, \mathbf{u}^k) = (1 - \alpha)\mathbf{s}^k + \alpha \tanh(W\mathbf{s}^k + W_{in}\mathbf{u}^k), \quad \mathbf{s}^0 = \mathbf{s}_0 \quad (15)$$

243

244

245

246

247

248

249

250

251

252

with initialization \mathbf{s}_0 and a relaxation parameter $\alpha \in (0, 1]$ (also known as the leaking rate) that controls the ‘speed’ of the artificial dynamics (Lukoševičius & Jaeger, 2009). Hence the state \mathbf{s} evolves according to a deterministic iteration with internal interactions given by a random (but fixed) W and forcing provided by the input data. The $\tanh(\cdot)$ activation function introduces a nonlinearity that is controlled by the weights in W_{in} . The spectral radius $\rho(W)$ determines the damping (or memory) of the system. From (15) it is apparent that α allows a matching of time-scales between the network and the variability in the training data, which is beneficial to the network’s predictive performance (Lukoševičius & Jaeger, 2009). The addition of the relaxation parameter α is the only significant difference between our formulation of f and that in Pathak et al. (2018).

253

254

255

256

257

258

259

260

Starting at $t = t_0$ with \mathbf{s}_0 , the recursion (15) generates N_T new states that are combined into a data matrix S , as described in Section 2.1. A linear operator C provides output predictions by translating the auxiliary state to a prediction. In the standard ESN approach the output operator $C : \mathbb{R}^{N_r} \rightarrow \mathbb{R}^{N_c}$ is computed from a regularized minimization problem using S^* and X' , see Section 2.2. Here S^* is an adapted version of S . As in Pathak et al. (2018), we take the square of the even elements in each state $\mathbf{s}^k \in S$. The motivation for this is largely empirical but related to problems that may originate with capturing symmetry in the model equations (Lu et al., 2017).

3 Results: Kuramoto–Sivashinsky model

In Pathak et al. (2018) a hybrid ESN was applied to the Kuramoto–Sivashinsky (KS) equation. Here we will begin with a test of our framework by replicating these results. We will first consider equal grids ($N_f = N_c$) and introduce an imperfection through a perturbation in one of the parameters. Later we explore a perfect/imperfect model setup with $N_f = 2N_c$ and no parameter perturbation in the KS-equation.

The KS-equation is capable of displaying rich spatiotemporal dynamics and is used to study a variety of phenomena such as flame front dynamics (Sivashinsky, 1977) and reaction-diffusion dynamics (Kuramoto, 1984). In one dimension it is given by

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + (1 + \epsilon) \frac{\partial^2 u}{\partial x^2} + \frac{\partial^4 u}{\partial x^4} = 0, \quad (16)$$

with $u \in [0, L]$, initial value $u(x, 0) = u_0(x)$ and periodic boundaries $u(x, t) = u(x + L, t)$. The domain size L is also the bifurcation parameter of the problem. In Pathak et al. (2018), the domain size is chosen at $L = 35$, for which the KS-equation has a positive maximum Lyapunov exponent λ_{max} and produces chaotic behavior (Hyman & Nicolaenko, 1986). A perturbation $\epsilon \geq 0$ is introduced to create an imperfection. With $\epsilon = 0$ we obtain the true, ‘perfect’ evolution whereas our ‘imperfect’ model will have $\epsilon > 0$.

The KS-equation is discretized on an equidistant grid: $x_i = i/N_f$ with $i = 1, 2, \dots, N_f = N_c = 64$. We use a fully-implicit time stepping scheme with $\Delta t = 0.25$ and initialize with

$$u_0(x_i) = \begin{cases} 1, & i = 1, \\ 0, & i > 1. \end{cases}$$

Starting at $t = t_0$, a transient is computed up to $T = 6000$ from which we select a large number of training and testing intervals. In the remaining experiments we also use long transients to sample training periods from. This approach is efficient from a data-management perspective but does not guarantee uncorrelated data.

The ESN used closely follows that in Pathak et al. (2018). The spectral radius is set at $\rho(W) = 0.4$, the average degree is $\bar{d} = 3$, we use training intervals of size $T = 5000$ and ignore any relaxation with $\alpha = 1$. The KS-equation and its discretization are also equivalent to Pathak et al. (2018) so, for a coherent interpretation of the predictions, we scale the obtained timings with the same Lyapunov exponent $\lambda_{max} = 0.07$.

The methods summarized in Table 1 are compared in a scaling experiment where the auxiliary state size N_r is doubled several times (see Fig. 1). Only those methods based on an ESN depend on this parameter which leads to constant results for the other pre-

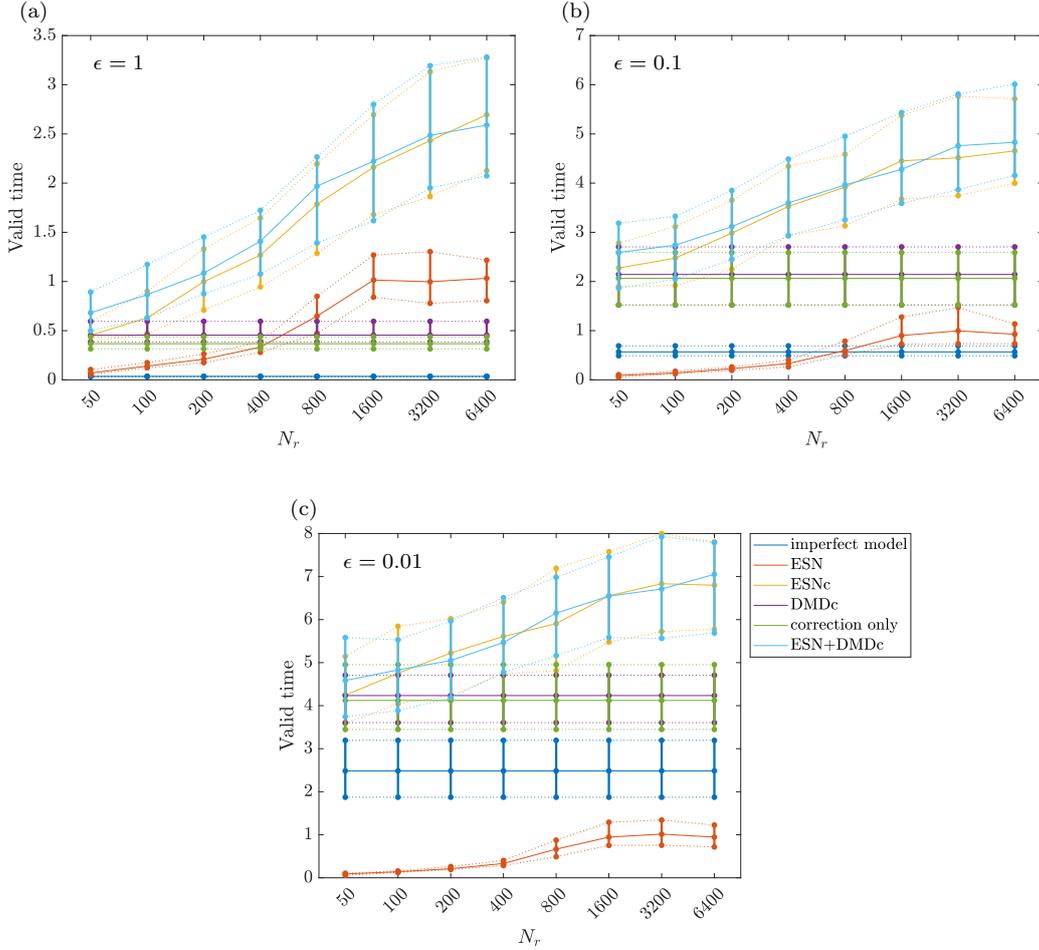


Figure 1: Results for the replication of the experiments in Pathak et al. (2018) where the imperfect model is a perturbed version of the KS-equation with perturbation parameter ϵ . ‘Valid time’ is the time it takes until the error threshold is passed: $E(\mathbf{x}^k, \mathbf{y}^k) > 0.4$. These timings are in Lyapunov units ($\lambda_{max}t$). The experiment is repeated for 100 different training intervals and network realizations. For each N_r a box plot is depicted showing the first, second and third quartile.

293 For each method we use 100 different training intervals and hence network re-
 294 alizations, as we do not reuse W . We fix the regularization parameter at $\lambda = 1 \cdot 10^{-5}$.
 295 The pure DMD-based methods (DMD and ESN+DMD) are not shown as they did not
 296 produce meaningful results. This is likely caused by DMD generalizing poorly and show-
 297 ing only valid predictions for a short period after $t_0 + T$.

298 The short-term prediction accuracy is measured using the normalized error used
 299 in Pathak et al. (2018). We compare the k -th prediction \mathbf{x}^k with the restricted truth $\mathbf{y}^k =$

300 $R\xi^k$ through

$$301 \quad E(\mathbf{x}^k, \mathbf{y}^k) = \frac{\|\mathbf{x}^k - \mathbf{y}^k\|}{\sqrt{\langle \|\mathbf{y}^k\|^2 \rangle}}, \quad (17)$$

302 with $\langle \cdot \rangle$ the mean over a time window up until k .

303 In Fig. 1 we see a strong resemblance with the results in Pathak et al. (2018). The
 304 imperfect model performs poorly on its own and the ESN-based methods improve the
 305 prediction as expected. A standalone ESN is able to achieve decent predictions for $\epsilon =$
 306 1 and $\epsilon = 0.1$. For $\epsilon = 0.01$, however, it appears impossible for a standalone ESN to
 307 perform better than the imperfect model. In all studied cases it is remarkable how the
 308 hybrid variant ESNc stands out. By combining the imperfect model physics with the ESN
 309 a significant improved is achieved.

310 The three additional models in Table 1 further explain the advantage of the hy-
 311 brid ESNc over the standalone ESN. The correction-only and DMDc predictions do not
 312 depend on N_r and show up here as constant solutions. These two regression-based cor-
 313 rections outperform the standalone ESN for $\epsilon = 0.1$ and $\epsilon = 0.01$. The third approach,
 314 ESN+DMDc, follows the ESNc performance but with an overall slight advantage for the
 315 two largest perturbations ϵ . This advantage is explained by the performance of DMDc
 316 and correction-only, as these are the linear components of ESN+DMDc and ESNc, re-
 317 spectively. In experiments where DMDc outperforms the correction we find a similar over-
 318 all gain between ESN+DMDc and ESNc. From the experiments in Fig. 1 it is apparent
 319 that ESN+DMDc and ESNc reduce to their linear components for low N_r , which is what
 320 would be expected from the correction equation (13). Hence the performance of the lin-
 321 ear models can be seen as a departure point for hybrid variants that add a nonlinear ESN.
 322 This largely explains the performance gain of, e.g. ESNc over the standalone ESN.

323 In a different perfect/imperfect model setup, illustrating the symbiotic modeling
 324 approach, the models both use $\epsilon = 0$ and have different spatial resolutions instead. The
 325 perfect model is discretized on a grid with twice the resolution, $N_f = 2N_c$. The domain
 326 size, ESN parameters and regularization remain unchanged. As explained in Section 2.1,
 327 fine grid information is restricted to the coarse grid and any data-driven corrections are
 328 made to the imperfect, coarse model evolution. Hence, instead of a model perturbation,
 329 it is now the difference in truncation errors and resolved scales between two resolutions
 330 that causes a model mismatch. With this setup the approach given by Equations (10)-
 331 (13) can be seen as a subgrid modeling technique.

332 The coarse model is capable of a good prediction in this setup (Fig. 2). DMDc, the
 333 correction-only and the standalone ESN are all unable to improve the coarse model. How-

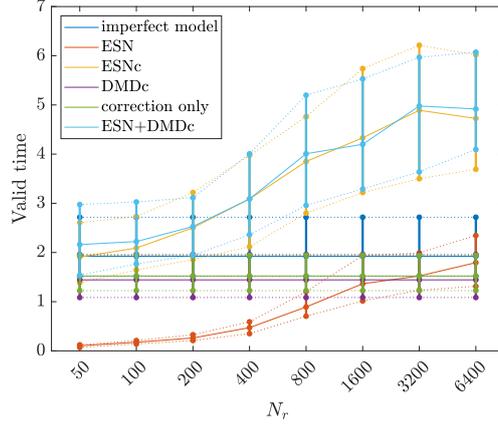


Figure 2: Grid experiment with the KS-equation. The imperfect model consists of the same equations but discretized on a grid half the resolution of the perfect model. Solutions are valid until $E(\mathbf{x}^k, \mathbf{y}^k) > 0.4$. As in Fig. 1, we repeat the experiment for 100 different training sets and network realizations.

334 ever, the hybrid variants ESNc and ESN+DMDc do show an overall improvement and
 335 an increase in predictive skill for larger N_r , similar to the parameter perturbation results
 336 (Fig. 1). For large values of N_r the hybrid methods double the predictive performance.
 337 This, again, shows the benefit of introducing the imperfect physical predictions to both
 338 force and control the artificial ESN. Hence the hybrid approach in Pathak et al. (2018)
 339 shows promise as a nonlinear subgrid modeling technique.

340 4 Results: quasi-geostrophic model

341 The barotropic quasi-geostrophic (QG) vorticity equation for a square (length L ,
 342 constant depth D) ocean basin is solved on a β -plane. The ocean flow is driven by an
 343 idealized zonal wind-stress forcing τ^x . Typical horizontal length and velocity scales are
 344 denoted L and U , from which the time scale follows as L/U . Using $L = 10^6$ m and $U =$
 345 $3.17 \cdot 10^{-2} \text{ ms}^{-1}$, we obtain a time scale of approximately one year. The equations are
 346 solved on a square domain, $x \in [0, 1]$, $y \in [0, 1]$, with periodic boundaries in both di-
 347 rections.

348 The QG equations in non-dimensional form are given by

$$349 \left[\frac{\partial}{\partial t} - \frac{\partial \psi}{\partial y} \frac{\partial}{\partial x} + \frac{\partial \psi}{\partial x} \frac{\partial}{\partial y} \right] (\omega + \beta y) = \frac{1}{\text{Re}} \nabla^2 \omega + \alpha_\tau C_\tau(x, y), \quad (18)$$

$$350 \omega = \nabla^2 \psi, \quad (19)$$

351 with ω the vertical component of the vorticity vector and streamfunction ψ . The Reynolds
 352 number is $\text{Re} = LU/A_H$, where A_H is the horizontal mixing coefficient and $\beta = \beta_0 L^2/U$,
 353 with $\beta_0 = 1.6 \cdot 10^{-11} \text{ (ms)}^{-1}$. Wind forcing enters through the nondimensional param-
 354 eter $\alpha_\tau = \tau_0 L/(\rho D U^2)$, with forcing amplitude $\tau_0 = 0.3 \text{ Nm}^{-2}$, density $\rho = 1 \cdot 10^3$
 355 kg m^{-3} and layer depth $D = 6 \cdot 10^2 \text{ m}$. We use a constant idealized wind-stress curl
 356 forcing in the form of a stirring pattern with stirring wavenumber $k_f = 5$ in both di-
 357 rections:

$$358 \quad C_\tau(x, y) = \cos(2k_f \pi x) \cos(2k_f \pi y). \quad (20)$$

359 This problem setup is a variant of the approach in Edeling & Crommelin (2019), but here
 360 we add a rotating frame.

361 4.1 Approach

362 Following the perfect/imperfect model approach we discretize the QG equations
 363 on two different grids. The perfect model uses a fine discretization on Ω^f with $N_f =$
 364 $2 \cdot 256^2$ unknowns and the imperfect variant is discretized on Ω^c with $N_c = 2 \cdot 32^2$ un-
 365 knowns ($N_f = 64N_c$). Furthermore, for both grids we model a flow with a Reynolds
 366 number that does not cause any numerical artifacts. With the perfect model we can run
 367 with $\text{Re}_f = 2000$. A stable flow for this Reynolds number and forcing amplitude τ_0 is
 368 impossible to achieve on the coarse grid and we therefore choose to use $\text{Re}_c = 500$ for
 the imperfect model.

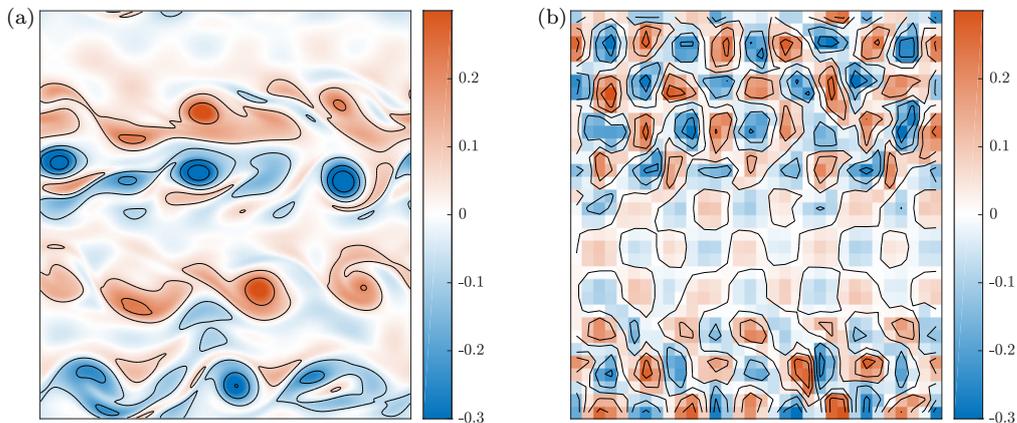


Figure 3: Snapshots of the vorticity fields (in day^{-1}) at the end of the transient depicted in Fig. 5. (a) Perfect model vorticity snapshot from a statistical equilibrium with $N_f = 2 \cdot 256^2$ and $\text{Re}_f = 2000$. (b) Imperfect model vorticity, also in a statistical equilibrium, $N_c = 2 \cdot 32^2$ and $\text{Re}_c = 500$.

370 For the discretization in time we use a fully implicit time stepping scheme that al-
 371 lows the use of the same time step for both models. In our experiments we will use $\Delta t =$
 372 1 day. The perfect QG solution (ω, ψ) is randomly initialized and run into a statistical
 373 steady state. From the steady state we select training periods of size $T = N_T = 10,000$
 374 days and follow the data gathering process described in Section 2.1. To get an idea of
 375 the perfect and imperfect flows we restart the imperfect model from a restricted fine state
 376 and run it into a steady regime. Snapshots from the two different statistical steady states
 377 are shown in Fig. 3. The imperfect model solution in Fig. 3b is highly diffusion domi-
 378 nated and shows a flow that strongly reflects the forcing pattern. The ‘perfect’ solution
 379 in Fig. 3a is — with 256^2 grid points — a moderately high-resolution flow and the dif-
 380 ference in resolved features with the imperfect model is substantial, which makes it an
 381 ideal testing ground for the corrective approaches in Table 1.

382 For the QG flow problem we will investigate the performance of the corrected tran-
 383 sients, following (10)-(13), in two different ways. We will make short-term predictions
 384 with the methods in Table 1 and compare with the truth using a normalized error, sim-
 385 ilar to the KS results in the Figs. 1 and 2. Then we let the different corrective mod-
 386 els run into a statistical equilibrium regime and compare the probability density func-
 387 tions (PDFs) of key flow properties with the perfect model equilibrium using their Kullback-
 388 Leibler divergence D_{KL} .

389 4.2 Short-term predictions

390 In Fig. 4 we present a short-term prediction experiment using the methods in Ta-
 391 ble 1. Only the standalone DMD and DMD+ESN corrections are excluded for their lack
 392 of meaningful results. For the ESN operators we again use $\rho(W) = 0.4$ and $\bar{d} = 3$, but
 393 with $T = 10,000$ days and $\Delta t = 1$ day we use half the amount of training data. For
 394 this problem we find that the optimal relaxation parameter lies around $\alpha = 0.2$ and
 395 the regularization is increased to $\lambda = 1 \cdot 10^{-4}$. The number of accurate days is mea-
 396 sured using a stricter tolerance $E(\mathbf{x}^k, \mathbf{y}^k) < 0.2$, allowing only a small departure from
 397 the true trajectory.

398 The poor performance of the imperfect QG model shown in Fig. 4 is improved by
 399 all studied methods. The standalone ESN needs at least $N_r = 1600$, while the other
 400 methods show a significant improvement for all chosen N_r . From Equation (13) and Ta-
 401 ble 1 it is evident that ESNc is an ESN combined with the correction-only approach. In
 402 the short-term QG predictions we find that these methods coincide for small N_r . A sim-
 403 ilar observation can be made for DMDc and the combination ESN+DMDc, which also
 404 coincide for low N_r . Controlled DMDc has better short-term predictive power than the

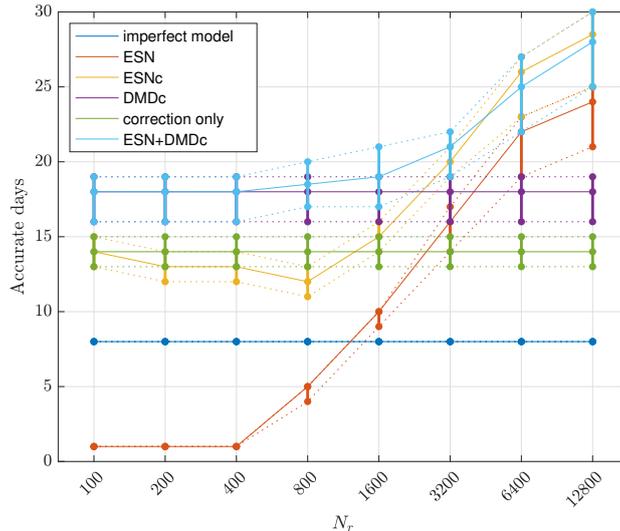


Figure 4: Short-term prediction experiments with the imperfect QG equations in a setup similar to Fig. 2. The experiments are repeated for 50 different network realizations and training sets. ‘Accurate days’ marks the time steps ($\Delta t = 1$ day) it takes until the error threshold is passed: $E(\mathbf{x}^k, \mathbf{y}^k) > 0.2$.

405 correction-only variant, which is also reflected in the behavior of ESN+DMDc and ESNc
 406 at low N_r . With this domain setup we expect DMDc to perform reasonably well on short
 407 time scales and it can therefore be viewed as a linear benchmark. The nonlinear ESN+DMDc
 408 hybrid improves on it immediately and it takes at least $N_r = 3200$ for the other ESN-
 409 based methods to take over. For large N_r , both hybrid methods (ESN+DMDc and ESNc)
 410 almost coincide and any positive influence of the DMD component is negligible.

411 The N_r doubling results are reminiscent of the findings with the KS-equation here
 412 and in Pathak et al. (2018). Similar to the KS scaling results, increasing N_r improves
 413 the short-term predictions of ESN-based methods for the QG problem. Based on the ex-
 414 periments with the KS-equation we expect that also here a plateau or a maximum will
 415 be reached for $N_r > 12800$. For ESN state sizes ranging between 200 and 1600 the ESN+DMDc
 416 combination gives the best results, where ESNc shows a slight decrease in performance.
 417 After $N_r = 1600$, the ESN component begins to dominate the results and ESNc becomes
 418 comparable to ESN+DMDc. Note, however, that also the standalone ESN is doing re-
 419 markably well for large state sizes.

420

4.3 Long-term dynamical regime

421

422

423

424

425

426

427

For the short-term results in the previous subsection, we used a normalized error based on the full fields (ω, ψ) for a comparison of the ‘hybrid’ model results with the (restricted) perfect model truth. Failure in terms of this measure does not imply the predictions are invalid, only that the exact truth is not reproduced. We are therefore also interested in reproducing ergodic properties of long-term time series as in Pathak et al. (2017). In this fashion we will continue here and investigate three flow properties for long-term transient runs: mean kinetic energy K_m , eddy kinetic energy K_e and enstrophy Z .

428

429

430

431

Horizontal velocities u, v follow from the streamfunction ψ , with $u = -\partial\psi/\partial y$, $v = \partial\psi/\partial x$, and are decomposed into a (time) mean and transient component: $u = \langle u \rangle + u'$, $v = \langle v \rangle + v'$ with the mean $\langle \cdot \rangle$ taken over a window of 50 days. The quantities K_m , K_e and Z are then given by

432

$$K_m = \int_{\Omega} (\langle u \rangle^2 + \langle v \rangle^2) d\Omega, \quad (21)$$

433

$$K_e = \int_{\Omega} (\langle u'^2 \rangle + \langle v'^2 \rangle) d\Omega = \int_{\Omega} (\langle u^2 \rangle - \langle u \rangle^2 + \langle v^2 \rangle - \langle v \rangle^2) d\Omega, \quad (22)$$

434

$$Z = \int_{\Omega} \omega^2 d\Omega, \quad (23)$$

435

where the integral is approximated with a Riemann sum over the coarse domain Ω^c .

436

437

438

439

440

441

442

443

444

445

A switch from the perfect ($N_f = 2 \cdot 256^2$, $\text{Re}_f = 2000$) to the imperfect ($N_c = 2 \cdot 32^2$, $\text{Re}_f = 500$) QG model solution will inevitably lead to a different statistical steady state. An example of this process is presented in Fig. 5. The perfect QG model is randomly initialized and runs into a statistical equilibrium. Predictions using imperfect QG, a standalone ESN and the hybrid ESNc then start from a restricted perfect QG state and run for 100 years. For stable long-term transients with the ESN-based methods we need a significantly larger regularization parameter ($\lambda = 1$) compared to the short-term experiments. Vorticity snapshots of the perfect and imperfect model depicted in Fig. 3 are taken at the end of the trajectories in Fig. 5. In Fig. 6 we present vorticity snapshots at the end of the ESN and ESNc trajectories.

446

447

448

449

450

The imperfect model reaches a very different statistical equilibrium after a transition period of approximately 10 years. A corrected transient based on (10)-(13) should stay closer to the perfect model’s dynamical regime and the presented ESN and ESNc trajectories show that this is feasible. Especially the hybrid ESNc shows a significantly better reproduction of the perfect model’s K_m PDF, compared to imperfect model (Fig. 5b).

451

452

In Fig. 7 the average energy spectrum over the final 80 years in Fig. 5a is shown. The spectrum provides another demonstration of the improved dynamics given by the

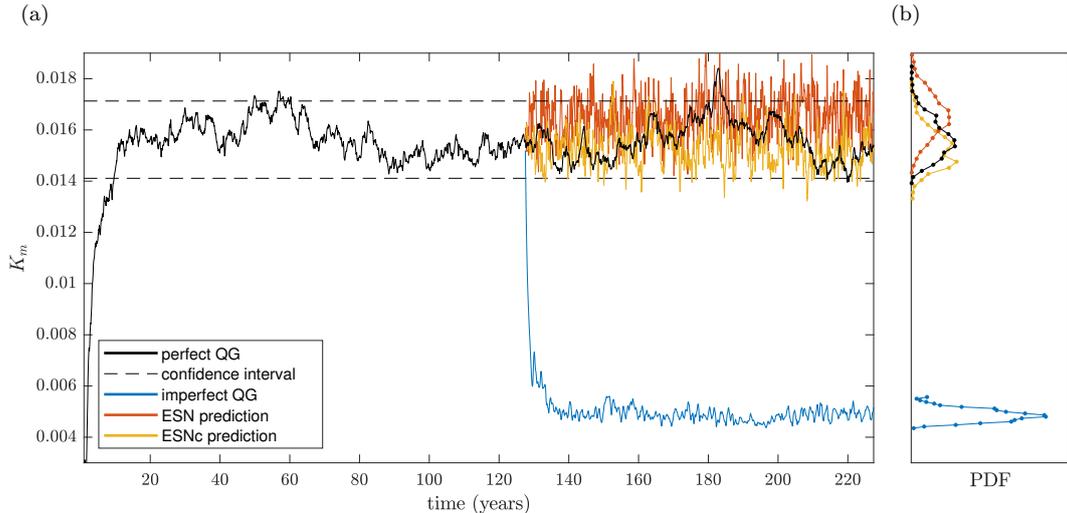


Figure 5: Spinup and long-term transient dynamics indicated by mean kinetic energy K_m . (a) A 100 year spinup with the perfect QG equations using a time step $\Delta t = 1$ day is followed by a training period of $N_T \Delta t = T = 10,000$ days. After the training period, 100 year predictions with imperfect QG, ESN and ESNc are shown, using $N_r = 3200$ and $\lambda = 1$. (b) Approximations of the probability density functions (PDFs) associated with the equilibrium transients, using 20 bins and excluding spinup/transition periods. Vorticity snapshots at the end of the depicted trajectories are shown in Figs. 3 and 6. The ESN results are with single realizations and serve as a demonstration of the corrected dynamics.

453 standalone ESN and the hybrid ESNc. The imperfect QG solution strongly reflects the
 454 forcing, which is also noticeable in the vorticity snapshot (Fig. 3b). In an incompress-
 455 ible 2D flow we expect energy to be transferred from the stirring wavelength to the larger
 456 scales, whereas enstrophy is transferred to the smaller scales and dissipated (Vallis, 2019).
 457 Both the energy and the enstrophy transfer are poorly represented in the imperfect model.
 458 The ESN-based methods are a lot better at producing the correct energy transfer and
 459 achieve a good correspondence for the lowest frequencies. Around the stirring frequency
 460 ESNc still performs well, whereas the standalone ESN is overestimating. The enstrophy
 461 transfer appears even more difficult to capture correctly but still the hybrid ESNc shows
 462 a great improvement over the standalone ESN at these scales.

463 The transients shown in Fig. 5 are specific examples and provide only information
 464 for a single realization of the ESN and a single training range. For a more rigorous ap-
 465 proach we compute transients for 50 training periods (and hence network realizations).

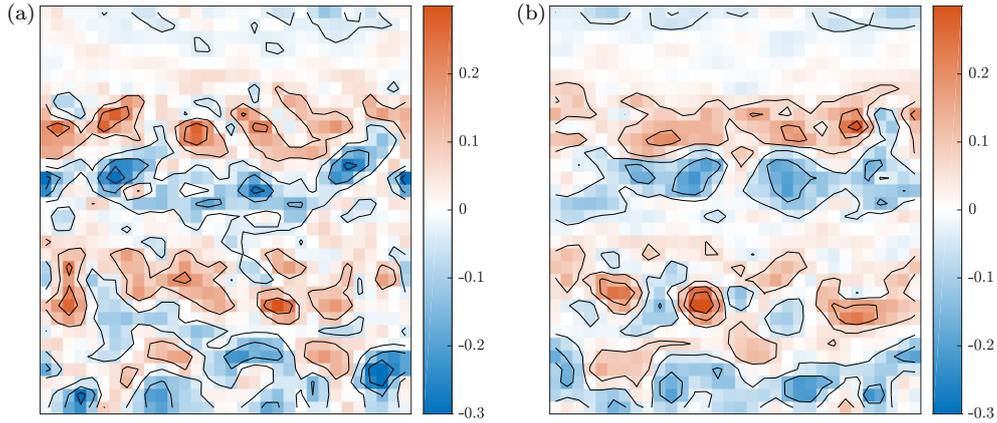


Figure 6: Snapshots of the vorticity fields (in day^{-1}) at the end of the transient in Fig. 5. (a) standalone ESN prediction with $N_r = 3200$ and $\lambda = 1$, (b) hybrid ESNc prediction with $N_r = 3200$ and $\lambda = 1$.

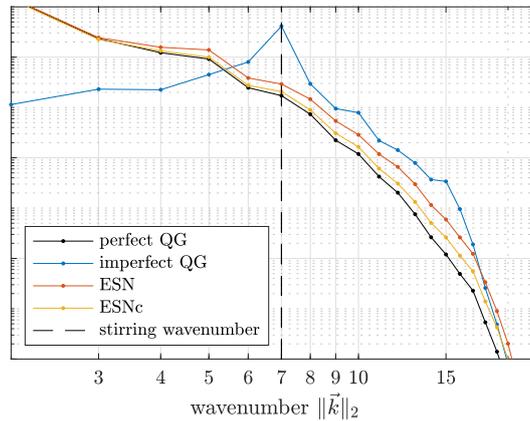


Figure 7: Average equilibrium energy spectrum based on the final 80 years of the trajectories in Fig. 5a. A dashed line is added to mark the frequency of the forcing.

466 We turn to all models studied in the short-term experiment (Fig. 4) and, to maintain
 467 a stable iteration, need to increase the regularization parameter λ . For the ESN-based
 468 methods we use $\lambda = 1$, for correction-only we will use $\lambda = 5$ and with DMDC we use
 469 $\lambda = 10$ to compute stable evolutions. Later in this section we explore how these meth-
 470 ods perform for various other λ choices.

471 From the trajectories we compute flow properties (K_m, K_e, Z) and compare their
 472 PDFs to the perfect model using their Kullback–Leibler (KL) divergence (Cover & Thomas,

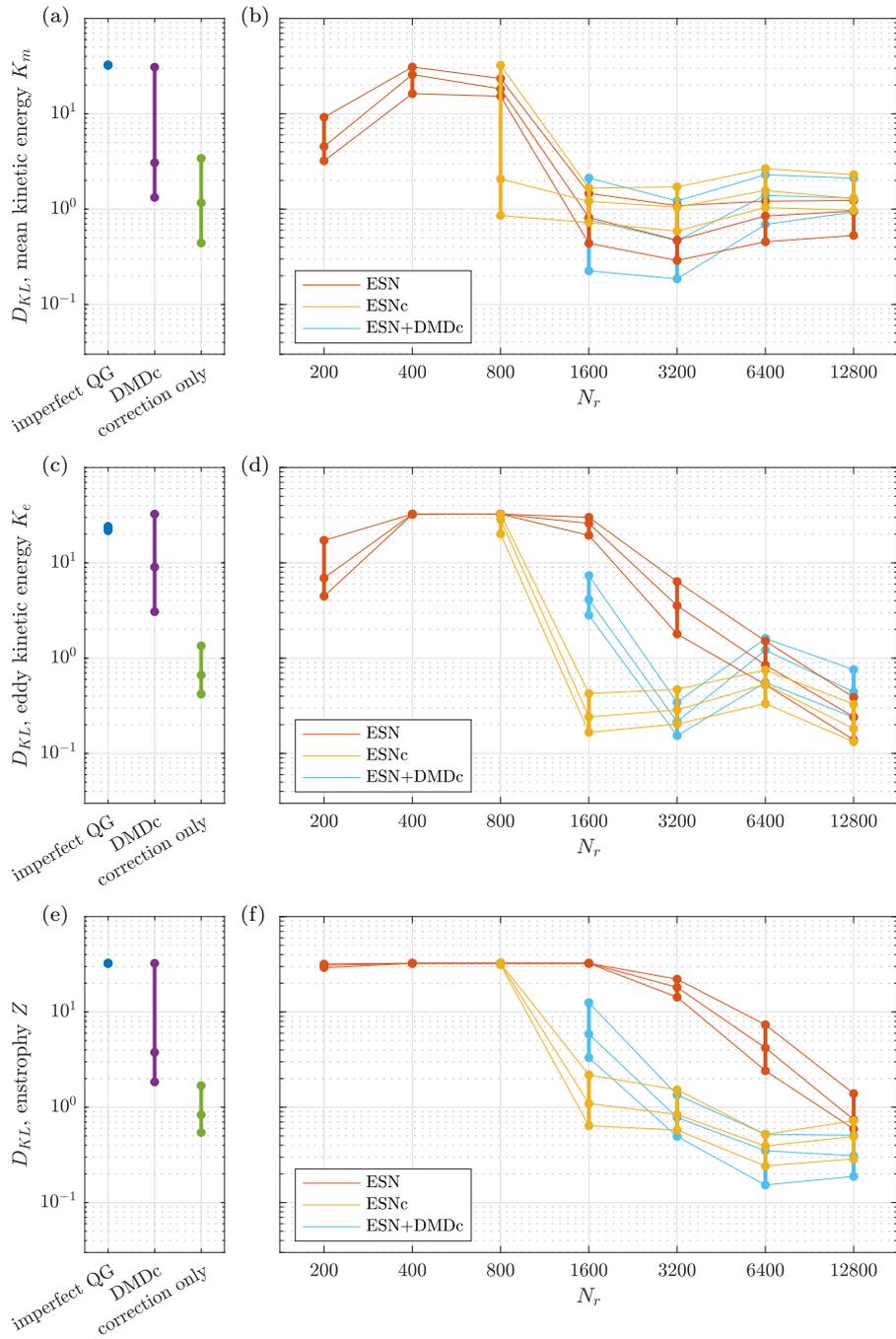


Figure 8: Long-term (100 year) transient results for 50 different training intervals. D_{KL} results from methods that do not depend on an ESN are shown in (a), (c) and (e), for K_m , K_e and Z , respectively. In (b), (d) and (f) the respective scalings with N_r are depicted for models with an ESN-dependence. Missing values in the plots are caused by unstable configurations.

473 2006): for two discrete distributions P and Q , the divergence of Q from P is given by

$$474 \quad D_{KL}(P, Q) = \sum_i P_i \ln \left(\frac{P_i}{Q_i} \right). \quad (24)$$

475 The PDFs are approximated using a domain that ranges beyond the perfect model’s PDF
 476 with twice the standard deviation. This domain is divided into 100 bins and every tran-
 477 sient is truncated to exclude initial spinup effects. For each flow property the divergence
 478 of its PDF from the ‘truth’ is computed and combined into boxplots for different ESN
 479 state sizes N_r (Fig. 8). We avoid division by zero in (24) by substituting zero-values with
 480 machine precision. This leads to large but finite divergences for non-overlapping distri-
 481 butions (~ 32). The imperfect model shows a poor representation of the variability, which
 482 should be expected from the transient example in Fig. 5. The PDFs for all flow prop-
 483 erties show no resemblance with the true PDF, giving D_{KL} results that remain at the
 484 maximum divergence value. Controlled DMDc and correction-only methods are better
 485 at capturing the variability, although this is highly dependent on the stabilizing regular-
 486 ization. Especially for the correction-only approach it is possible to find a configuration
 487 such that PDFs give a reasonable correspondence.

488 The KL-divergences for ESN-based methods in Fig. 8 are partly missing. For low
 489 N_r , ESNc and ESN+DMDc are unstable when $\lambda = 1$. The remaining results show an
 490 overall improvement for increasing ESN state size N_r (cf. Fig. 1), although not very clear
 491 for all flow properties. Both mean and eddy kinetic energy KL-divergences are some-
 492 what irregular with optima at moderate N_r values. For enstrophy, the ESN-based meth-
 493 ods gradually improve with ESN state size. From the energy spectrum in Fig. 7 we know
 494 that the enstrophy transfer is difficult to capture and here a similar effect is visible in
 495 the correspondence between PDFs. ESNc requires at least $N_r = 1600$ to obtain small
 496 KL-divergences from the enstrophy PDF, further improving for larger N_r .

497 Diverged trajectories show up as non-overlapping with either a maximal KL-divergence
 498 or a missing value in the D_{KL} results. Poor performing methods are hence indistinguish-
 499 able from unstable ones. Especially the combination ESN+DMDc appears to suffer from
 500 stability issues for small N_r , leading to missing D_{KL} values. We find that the ESN sta-
 501 bilizes regression-based corrective methods, as already noted in Arcomano et al. (2022).
 502 When the regression-based methods run on their own we choose a regularization that
 503 stabilizes sufficiently such that divergent trajectories are rare.

504 To provide an idea of how regularization affects the long-term performance of var-
 505 ious methods we perform numerous equilibrium runs for different λ . In Fig. 9 we present
 506 the results for enstrophy Z . The correction-only approach gives remarkably good results
 507 within a narrow optimal region for λ . It is also only slightly enhanced by the combina-

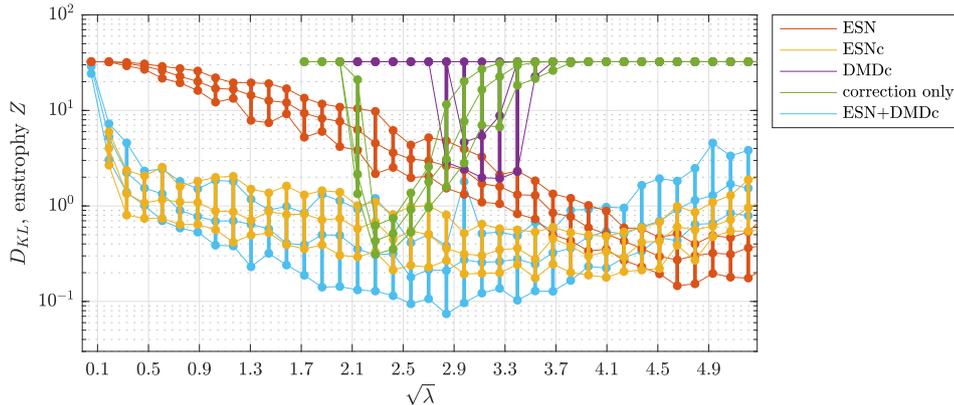


Figure 9: Reproduction of the enstrophy Z variability for different regularization parameters using an equidistant spacing in $\sqrt{\lambda}$. Long-term (100 year) equilibrium runs are performed for 50 different (but partially overlapping) training sets and network realizations. Boxplots show the first, second and third quartile of the resulting spread of divergences D_{KL} . The ESN-based methods have dimension $N_r = 3200$.

508 tion with an ESN (i.e. ESNc). The hybrid ESNc and ESN+DMDc are, however, much
 509 more robust and overall better at reproducing the correct enstrophy variability. From
 510 the regularization parameter study it is clear that DMDc needs a stronger regulariza-
 511 tion than the correction-only approach. The KL-divergences in Fig. 8 show a related prob-
 512 lem for the models that incorporate an ESN, where the ESN that combines with DMDc
 513 needs a much larger state size N_r to achieve sufficient stabilization. Hence stabilization
 514 is achieved through both regularization λ and ESN complexity N_r .

515 5 Summary and discussion

516 In this paper we demonstrated part of a symbiotic ocean modeling approach, i.e.,
 517 a framework in which models with different complexities are coupled in order to bene-
 518 fit from each other. We distinguish between perfect and imperfect models in terms of
 519 differing spatial resolutions and key parameterizations, and focus on how an imperfect
 520 model can benefit from a symbiotic setup. With data generated from both model types
 521 we seek to correct imperfect model transients. To this end, we make use of hybrid mod-
 522 eling techniques that combine linear regression-based methods with nonlinear echo state
 523 networks (ESNs). Currently, efforts are under way to demonstrate the second part of the
 524 symbiotic framework, i.e., HR models that benefit computationally from LR models.

525 We establish that our hybrid (or physics-controlled) ESNc implementation repro-
526 duces short-term predictions for the Kuramoto–Sivashinsky equation (KS) that are con-
527 sistent with earlier work in Pathak et al. (2017). Our framework furthermore allows a
528 straightforward comparison with purely regression-based methods. We show how cor-
529 rections based on linear regression contribute to the success of the hybrid machine learn-
530 ing combinations and serve as a departure point for hybrid methods. When we apply
531 these techniques to a subgrid modeling version of the KS problem, we observe a simi-
532 lar scaling behavior with ESN complexity and departure points rooted in the linear re-
533 gression techniques. For the subgrid modeling problem of the single-layer quasi-geostrophic
534 potential vorticity equations (QG), short-term predictions give results that are compa-
535 rable to the findings with the KS-equation. A scaling behavior is found with the size of
536 the ESN. When the ESN complexity is negligible, the hybrid methods reduce to their
537 linear regression components. For the long-term flow development, our comparison of
538 statistical steady states shows that the hybrid combinations are robust and perform well
539 for various flow parameters. The parameter study with long-term statistics also shows
540 how the ESN-based methods improve with ESN state size, reminiscent of the short-term
541 full-field reproductions. For our purposes, however, the comparison of long-term flow char-
542 acteristics is more informative than an error norm on state differences.

543 The parameter studies with equilibrium simulations show that subgrid models based
544 on only an ESN or regression are often inaccurate or difficult to stabilize. We control the
545 stabilization through regularizing the regression-problem, which is another (hyper) pa-
546 rameter to tweak. For purely regression-based methods, regularization is the only tun-
547 able apart from data choices. A benefit of combining regression with an ESN is appar-
548 ent from our regularization experiments. Here we observe that the ESN stabilizes its regression-
549 based component, which was also mentioned in Arcomano et al. (2022). We conclude
550 that stabilization is achieved through both regularization λ and ESN state size N_r . How-
551 ever, other parameters such as the spectral radius $\rho(W)$ (controlling damping in the ESN)
552 were not studied in this context and may have similar stabilizing effects. Combinations
553 with a DMD model are available within our framework, which yields interesting com-
554 parisons, especially in the short-term QG experiments. Benefits of adding a DMD model
555 are visible for moderate ESN state sizes. For long-term transient runs the advantage of
556 hybrid DMD-ESN models is less pronounced, which is possibly due to the DMD model
557 being valid for only a short period and hence it should be (partially) rebuilt in an on-
558 line fashion (Pendergrass et al., 2016).

559 Obviously, the QG ocean model used here is highly idealized compared to state-
560 of-the-art ocean models. However, we think that these ideas are applicable to the gen-

561 eral problem of correcting large scale flows, i.e., improving a coarse and more viscous ver-
 562 sion of the flow problem at hand. For models with a higher dimension than studied here
 563 a reduced order version of the corrective transient framework, as defined by (10)-(13),
 564 is worth investigating. Here the best choice of reduced coordinates (POD, Fourier, wavelets)
 565 in combination with an ESN remains uncertain. Projecting with global POD modes, for
 566 instance, greatly reduces the ESN’s predictive skill (Vlachas et al., 2020). A localized
 567 representation as used in Wan et al. (2021) shows more promise. Another way to tackle
 568 high-dimensional problems is through parallelization. A parallel hybrid ESNc based on
 569 a local domain decomposition is used in Wikner et al. (2020) and Arcomano et al. (2022).
 570 It would be interesting to apply this approach as a subgrid model and reproduce long-
 571 term flow characteristics, comparing especially its ability to correctly capture energy and
 572 enstrophy transfer at low wave numbers.

573 Open Research

574 The software developed for this paper is archived at Zenodo and available through
 575 <https://doi.org/10.5281/zenodo.7572246>.

576 Acknowledgments

577 This work was supported by funding from the SMCM project of the Netherlands eScience
 578 Center (NLeSC) with project number 027.017.G02. We thank the Center for Informa-
 579 tion Technology of the University of Groningen for their support and for providing ac-
 580 cess to the Peregrine high performance computing cluster. In addition we thank T.M. Hep-
 581 kema for helpful discussions and two anonymous reviewers for their thorough and use-
 582 ful comments which greatly improved this work.

583 References

- 584 Arcomano, T., Szunyogh, I., Wikner, A., Pathak, J., Hunt, B. R., & Edward,
 585 O. (2022). A Hybrid Approach to Atmospheric Modeling that Combines
 586 Machine Learning with a Physics-Based Numerical Model. *Preprint*. doi:
 587 10.1002/essoar.10507548.2
- 588 Berloff, P. S. (2005). Random-forcing model of the mesoscale oceanic eddies. *J.*
 589 *Fluid Mech.*, *529*, 71–95. doi: 10.1017/S0022112005003393
- 590 Bolt, E. (2021). On explaining the surprising success of reservoir computing fore-
 591 caster of chaos? The universal machine learning dynamical system with contrast
 592 to VAR and DMD. *Chaos*, *31*(1). doi: 10.1063/5.0024890
- 593 Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data in-

- 594 ference and subgrid parameterization. *Journal of Advances in Modeling Earth*
595 *Systems*, 11(1), 376-399. Retrieved from <https://agupubs.onlinelibrary>
596 [.wiley.com/doi/abs/10.1029/2018MS001472](https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001472) doi: [https://doi.org/10.1029/](https://doi.org/10.1029/2018MS001472)
597 2018MS001472
- 598 Briggs, W. L., Henson, V. E., & McCormick, S. F. (2000). *A multigrid tutorial* (Sec-
599 ond ed.). Society for Industrial and Applied Mathematics (SIAM), Philadelphia,
600 PA. Retrieved from <https://doi.org/10.1137/1.9780898719505> doi: 10.1137/
601 1.9780898719505
- 602 Chang, P., Zhang, S., Danabasoglu, G., Yeager, S. G., Fu, H., Wang, H., . . . Wu, L.
603 (2020). An Unprecedented Set of High-Resolution Earth System Simulations for
604 Understanding Multiscale Interactions in Climate Variability and Change. *Journal*
605 *of Advances in Modeling Earth Systems*, 12(12). doi: 10.1029/2020ms002298
- 606 Chassignet, E. P., Yeager, S. G., Fox-Kemper, B., Bozec, A., Castruccio, F., Dan-
607 abasoglu, G., . . . Xu, X. (2020). Impact of horizontal resolution on global ocean-
608 sea ice model simulations based on the experimental protocols of the ocean model
609 intercomparison project phase 2 (omip-2). *Geoscientific Model Development*,
610 13(9), 4595-4637. Retrieved from [https://gmd.copernicus.org/articles/13/](https://gmd.copernicus.org/articles/13/4595/2020/)
611 4595/2020/ doi: 10.5194/gmd-13-4595-2020
- 612 Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (Second ed.).
613 Wiley-Interscience [John Wiley & Sons], Hoboken, NJ.
- 614 Edeling, W., & Crommelin, D. (2019). Towards data-driven dynamic surrogate mod-
615 els for ocean flow. *Proceedings of the Platform for Advanced Scientific Computing*
616 *Conference, PASC 2019*. doi: 10.1145/3324989.3325713
- 617 Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., &
618 Taylor, K. E. (2016). Overview of the coupled model intercomparison project
619 phase 6 (cmip6) experimental design and organization. *Geoscientific Model De-*
620 *velopment*, 9(5), 1937-1958. Retrieved from [https://gmd.copernicus.org/](https://gmd.copernicus.org/articles/9/1937/2016/)
621 [articles/9/1937/2016/](https://gmd.copernicus.org/articles/9/1937/2016/) doi: 10.5194/gmd-9-1937-2016
- 622 Gargett, A. E. (1989). Ocean turbulence. *Annual Review of Fluid Mechan-*
623 *ics*. Retrieved from [http://www.annualreviews.org/doi/pdf/10.1146/annurev](http://www.annualreviews.org/doi/pdf/10.1146/annurev.fl.21.010189.002223)
624 [.fl.21.010189.002223](http://www.annualreviews.org/doi/pdf/10.1146/annurev.fl.21.010189.002223)
- 625 Gent, P. R., Willebrand, J., McDougall, T. J., & McWilliams, J. C. (1995). Param-
626 eterizing eddy-induced tracer transports in ocean circulation models. *Journal Of*
627 *Physical Oceanography*, 25, 463 - 474.
- 628 Hallberg, R. (2013). Using a resolution function to regulate parameterizations of
629 oceanic mesoscale eddy effects. *Ocean Modelling*, 72, 92-103.
- 630 Hewitt, H. T., Roberts, M., Mathiot, P., Biastoch, A., Blockley, E., Chassignet,

- 631 E. P., ... Zhang, Q. (2020). Resolving and Parameterising the Ocean Mesoscale
 632 in Earth System Models. *Current Climate Change Reports*, 6(4), 137–152. doi:
 633 10.1007/s40641-020-00164-w
- 634 Hyman, J. M., & Nicolaenko, B. (1986). The kuramoto-sivashinsky equation:
 635 A bridge between pde’s and dynamical systems. *Physica D: Nonlinear Phe-*
 636 *nomena*, 18(1), 113-126. Retrieved from [https://www.sciencedirect.com/](https://www.sciencedirect.com/science/article/pii/0167278986901661)
 637 [science/article/pii/0167278986901661](https://www.sciencedirect.com/science/article/pii/0167278986901661) doi: [https://doi.org/10.1016/](https://doi.org/10.1016/0167-2789(86)90166-1)
 638 [0167-2789\(86\)90166-1](https://doi.org/10.1016/0167-2789(86)90166-1)
- 639 Irrgang, C., Boers, N., Sonnewald, M., Barnes, E. A., Kadow, C., Staneva, J., &
 640 Saynisch-Wagner, J. (2021). Towards neural Earth system modelling by integrat-
 641 ing artificial intelligence in Earth system science. *Nature Machine Intelligence*,
 642 3(8), 667–674. doi: 10.1038/s42256-021-00374-3
- 643 Jaeger, H. (2001). *The "echo state" approach to analysing and training recurrent*
 644 *neural networks* (GMD Report No. 148). GMD - German National Research
 645 Institute for Computer Science. Retrieved from [http://www.faculty.jacobs-](http://www.faculty.jacobs-university.de/hjaeger/pubs/EchoStatesTechRep.pdf)
 646 [university.de/hjaeger/pubs/EchoStatesTechRep.pdf](http://www.faculty.jacobs-university.de/hjaeger/pubs/EchoStatesTechRep.pdf)
- 647 Jaeger, H., & Haas, H. (2004). Harnessing Nonlinearity: Predicting Chaotic Systems
 648 and Saving Energy in Wireless Communication. *Science*, 304(5667), 78–80. doi:
 649 10.1126/science.1091277
- 650 Jüling, A., Zhang, X., Castellana, D., Heydt, A. S. v. d., & Dijkstra, H. A. (2021).
 651 The Atlantic’s freshwater budget under climate change in the Community Earth
 652 System Model with strongly eddying oceans. *Ocean Science*, 17(3), 729–754. doi:
 653 10.5194/os-17-729-2021
- 654 Kuramoto, Y. (1984). *Chemical oscillations, waves, and turbulence* (Vol. 19).
 655 Springer-Verlag, Berlin. Retrieved from [https://doi.org/10.1007/978-3-642-](https://doi.org/10.1007/978-3-642-69689-3)
 656 [69689-3](https://doi.org/10.1007/978-3-642-69689-3) doi: 10.1007/978-3-642-69689-3
- 657 Kutz, J. N., Brunton, S. L., Brunton, B. W., & Proctor, J. L. (2016). *Dynamic mode*
 658 *decomposition : Data-driven modeling of complex systems*. SIAM. Retrieved from
 659 <http://www.dmdbook.com/>
- 660 Lu, Z., Pathak, J., Hunt, B., Girvan, M., Brouckett, R., & Ott, E. (2017). Reser-
 661 voir observers: Model-free inference of unmeasured variables in chaotic systems.
 662 *Chaos*, 27(4). Retrieved from <http://dx.doi.org/10.1063/1.4979665> doi:
 663 10.1063/1.4979665
- 664 Lukosevicius, M. (2012). A practical guide to applying echo state networks. In
 665 G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the*
 666 *trade (2nd ed.)* (Vol. 7700, p. 659-686). Springer.
- 667 Lukoševičius, M., & Jaeger, H. (2009). Reservoir computing approaches to recurrent

- 668 neural network training. *Computer Science Review*, 3(3), 127–149. doi: 10.1016/j
 669 .cosrev.2009.03.005
- 670 Mana, P. P., & Zanna, L. (2014, 07). Toward a stochastic parameterization of ocean
 671 mesoscale eddies. *Ocean Modelling*, 79(C), 1 – 20. doi: 10.1016/j.ocemod.2014.04
 672 .002
- 673 Nadiga, B. T. (2021). Reservoir Computing as a Tool for Climate Predictability
 674 Studies. *Journal of Advances in Modeling Earth Systems*, e2020MS002290. doi: 10
 675 .1029/2020ms002290
- 676 Pathak, J., Lu, Z., Hunt, B. R., Girvan, M., & Ott, E. (2017). Using machine learn-
 677 ing to replicate chaotic attractors and calculate Lyapunov exponents from data.
 678 *Chaos*, 27(12). doi: 10.1063/1.5010300
- 679 Pathak, J., Wikner, A., Fussell, R., Chandra, S., Hunt, B. R., Girvan, M., & Ott,
 680 E. (2018). Hybrid forecasting of chaotic processes: Using machine learning
 681 in conjunction with a knowledge-based model. *Chaos*, 28(4). Retrieved from
 682 <https://doi.org/10.1063/1.5028373> doi: 10.1063/1.5028373
- 683 Pendergrass, S. D., Kutz, J. N., & Brunton, S. L. (2016). *Streaming gpu singular*
 684 *value and dynamic mode decompositions*. arXiv. Retrieved from [https://arxiv](https://arxiv.org/abs/1612.07875)
 685 [.org/abs/1612.07875](https://arxiv.org/abs/1612.07875) doi: 10.48550/ARXIV.1612.07875
- 686 Proctor, J. L., Brunton, S. L., & Kutz, J. N. (2016). Dynamic mode decomposition
 687 with control. *SIAM Journal on Applied Dynamical Systems*, 15(1), 142-161. Re-
 688 trieved from <https://doi.org/10.1137/15M1013857> doi: 10.1137/15M1013857
- 689 Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid
 690 processes in climate models. *Proceedings of the National Academy of Sciences*,
 691 115(39), 9684-9689. Retrieved from [https://www.pnas.org/doi/abs/10.1073/](https://www.pnas.org/doi/abs/10.1073/pnas.1810286115)
 692 [pnas.1810286115](https://www.pnas.org/doi/abs/10.1073/pnas.1810286115) doi: 10.1073/pnas.1810286115
- 693 Schmid, P. J. (2010). Dynamic mode decomposition of numerical and experimental
 694 data. *Journal of Fluid Mechanics*, 656, 5–28. doi: 10.1017/S0022112010001217
- 695 Sivashinsky, G. I. (1977). Nonlinear analysis of hydrodynamic instability in laminar
 696 flames. I. Derivation of basic equations. *Acta Astronaut.*, 4(11-12), 1177–1206. Re-
 697 trieved from [https://doi.org/10.1016/0094-5765\(77\)90096-0](https://doi.org/10.1016/0094-5765(77)90096-0) doi: 10.1016/
 698 0094-5765(77)90096-0
- 699 Vallis, G. K. (2019). *Essentials of atmospheric and oceanic dynamics*. Cambridge
 700 University Press. doi: 10.1017/9781107588431
- 701 Viebahn, J., Crommelin, D., & Dijkstra, H. (2019). Toward a turbulence closure
 702 based on energy modes. *Journal of Physical Oceanography*, 49(4), 1075 - 1097.
 703 Retrieved from <https://journals.ametsoc.org/view/journals/phoc/49/4/>

704 jpo-d-18-0117.1.xml doi: 10.1175/JPO-D-18-0117.1

705 Vlachas, P. R., Pathak, J., Hunt, B. R., Sapsis, T. P., Girvan, M., Ott, E., &
706 Koumoutsakos, P. (2020). Backpropagation algorithms and Reservoir Comput-
707 ing in Recurrent Neural Networks for the forecasting of complex spatiotemporal
708 dynamics. *Neural Networks*, *126*, 191–217. Retrieved from [https://doi.org/](https://doi.org/10.1016/j.neunet.2020.02.016)
709 [10.1016/j.neunet.2020.02.016](https://doi.org/10.1016/j.neunet.2020.02.016) doi: 10.1016/j.neunet.2020.02.016

710 Wan, Z. Y., Dodov, B., Lessig, C., Dijkstra, H., & Sapsis, T. P. (2021). A data-
711 driven framework for the stochastic reconstruction of small-scale features with
712 application to climate data sets. *Journal of Computational Physics*, *442*,
713 110484. Retrieved from <https://doi.org/10.1016/j.jcp.2021.110484> doi:
714 [10.1016/j.jcp.2021.110484](https://doi.org/10.1016/j.jcp.2021.110484)

715 Wikner, A., Pathak, J., Hunt, B., Girvan, M., Arcomano, T., Szunyogh, I., . . . Ott,
716 E. (2020). Combining machine learning with knowledge-based modeling for
717 scalable forecasting and subgrid-scale closure of large, complex, spatiotemporal
718 systems. *Chaos (Woodbury, N.Y.)*, *30*(5), 053111. doi: 10.1063/5.0005541