Symmetry Constraints Enhance Long-term Stability and Accuracy in Unsupervised Learning of Geophysical Fluid Flows

Yunfei Huang[1, 2] and David S. Greenberg[1, 2]
[1)]*Institute of Coastal Systems, Helmholtz-Zentrum Hereon, Geesthacht, Germany*
[2)]*Helmholtz AI*

(*Electronic mail: david.greenberg@hereon.de)

(*Electronic mail: Yunfei.Huang@hereon.de)

(Dated: 23 January 2024)

Fluid dynamical systems are well described by discretized partial differential equations (PDEs), but computational costs limit accuracy, duration and/or resolution in numerical integrations. Recent studies showed that deep neural networks trained on simulations or PDE-derived losses can improve cost-accuracy tradeoffs, but purely data-centric approaches discard physical and mathematical insights and require computationally costly training data. Here we draw on advances in geometric deep learning to design solver networks that respect PDE symmetries as hard constraints. We construct equivariant convolutional layers for mixed scalar-vector input fields in order to capture the symmetries inherent to specific PDEs. We demonstrate our approach on a challenging one-dimensional semi-implicit shallow water scheme with closed boundaries, applying unsupervised learning with a physics-derived loss function. We report strong improvements in accuracy and stability of equivariant solvers compared to standard convolutional networks with the same architectures and parameter counts. Solver equivariance also improves performance on new initial conditions not encountered during training, and suppresses error accumulation in global momentum and energy. Strikingly, these benefits do not reduce loss values during training, but appear later during machine learning (ML)-assisted rollouts over time steps. Our results suggest that symmetry constraints could improve deep learning performance across a wide range of fluid dynamical tasks, learning algorithms and neural architectures.

I. INTRODUCTION

Partial differential equations (PDEs) are essential for understanding and simulating complex fluid dynamics. Examples include convection-diffusion[1], Euler[2] and Navier–Stokes equations (NS)[3,4]. The shallow water equations (SWEs)[5], derived by depth integration of NS, are mathematically simpler but widely employed as test cases to evaluate solution techniques for ocean, weather and climate applications[6–11].

PDEs describing geophysical fluid flows require numerical methods, for example, finite difference[12,13], finite element[14,15], finite volume[16], boundary element[17], and spectral element methods[18]. While small spatial domains admit direct numerical simulation, geophysical applications require coarse grids with Reynolds-averaging[19] or large eddy simulation (LES)[20,21] to approximate unresolved scales. Explicit time stepping simplifies computations but requires small steps for stability, while (semi)implicit schemes take larger time steps but must iteratively solve a system of equations [22–25]. However, these classical approaches incur heavy computational costs at high spatial and temporal resolutions.

Recent machine learning approaches aim to transcend these cost-accuracy tradeoffs by training a model to accurately and efficiently solve PDEs on modern computational hardware. *Supervised learning* uses simulations from a classical solver to train a machine learning (ML) model that uses larger space and time steps or skips the iterative computations of an implicit scheme. This approach has shown success in accelerating PDE solutions while maintaining accuracy, obeying conservation laws and preserving high frequency features[26–32], and has also been applied to mesh-free particle-based solvers[33].

Unsupervised learning trains the model to satisfy the PDE without requiring training data. It is most effective for implicit schemes, since solving their equations iteratively is complex and costly but verifying a solution is simple and fast. Unsupervised learning avoids overfitting by training on its own outputs, but cannot avoid discretization errors for large the space or time steps. It has been used to solve several fluid dynamical PDEs[34–43].

Hybrid models replace only part of a classical PDE solver with an ML model, leaving other components unchanged. Early work applied this to computer graphics[44], while a later study demonstrated an approach combining a fluid solver with ML techniques to approximate NS in a Lagrangian framework using regression forests[45]. More recently, authors[46] used supervised learning to compute an additive correction to low-resolution incompressible NS, so that its evolution mimics a high-resolution model coarsened at each time step. An LSTM-based hybrid approach[47] with significant practical speed-ups has been presented for predicting pressure changes for incompressible flow, while ref.[48] proposed an ML-based approach for replacing the linear projection in the Eulerian fluid implicit simulation, authors[27] combined two well-established turbulent flow simulation techniques with deep learning and the paper[49] developed an accelerated integrative ML solverto aid convergence of Reynolds Averaged Navier-Stokes simulations. Overall, hybrid methods allow us to effectively incorporate physical knowledge while simplifying the learning task, and can improve accuracy and generalization capabilities.

Major challenges remain for ML-based PDE solvers: long-term stability and accuracy are not guaranteed even for low loss values on training and testing data[46,50–52], and general-

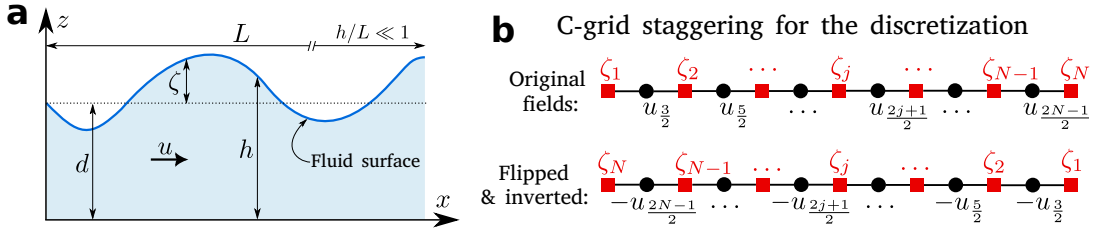

FIG. 1. Schematic representation of a one-dimensional shallow water problem with C-grid staggering of discretized fields. (a) Shallow water system with domain length L. d and h are un-disturbed- and disturbed- water depth, ζ is fluid surface elevation and $h = \zeta + d$. u is the velocity along the space coordinate x. (b) Staggered grid for elevation and velocity. Top: fluid surface elevation ζ_j is represented at red squares while velocity $u_{j+1/2}$ is represented at black circles. Bottom: A flipped fluid surface elevation ζ_i, as well as a flipped-and-inverted velocity. These transformed fields solve the SWE with transformed initial and boundary conditions.

ization to new scenarios remains problematic[53–56]. A partial explanation is that training neural networks means choosing from a large, high-dimensional family of functions, many of which are physically or mathematically implausible. Narrowing the search by constraining the learned function has shown great promise: for example, conservation laws improve process representations in climate, weather, and ocean models[57,58], while symmetry constraints aid image classification[59] and segmentation[60,61]. However, the potential benefit for fluid dynamics remains mostly unclear.

In this work we construct hybrid PDE solvers using equivariant neural networks that obey PDE symmetry constraints. We draw on previous work in geometric deep learning[59,62], but extend group equivariant convolutions to handle mixed scalar/vector inputs with the correct, PDE-specific transformation rules. We demonstrate the benefit of equivariant solver networks using an unsupervised learning task, in which the network is trained to integrate a semi-implicit scheme for one-dimensional (1-D) shallow water equations. These equations exhibit challenging stiff dynamics due to closed boundaries and reflecting waves. Our experiments show significant improvements in long term accuracy and stability compared to standard convolutional neural networks (CNNs), despite similar loss values during training. We also observe that symmetry constraints improve performance on initial conditions not encountered during training, as well as representations of global mass, momentum and energy.

II. NUMERICAL INTEGRATION OF FLUID DYNAMICS

In this section we establish concepts and notation for PDE integration with classical numerical techniques, allowing us to describe our task and approach in the following section. We consider a general governing partial differential equation for fluid dynamics:

$$\frac{\partial q(t,x)}{\partial t} = \mathcal{F}[q] = f(t,x,q(t,x),\frac{dq}{dx},\frac{d^2q}{dx^2},\ldots), \quad x \in \Omega \tag{1}$$

$$q(x,t) = q_\Omega(x), \quad x \in \partial\Omega \tag{2}$$

$$q(x,0) = q^0(x) \tag{3}$$

x and t are space and time coordinates and $q(t,x)$ is the vector of modeled variable fields at one place and time, such as velocity and pressure in NS or velocity and height in SWE. $\mathcal{F}$ is a nonlinear operator computing time derivatives as nonlinear functions of functions of the fields and their spatial derivatives. The Dirichlet boundary conditions (BCs) $q_\Omega(x)$ on the boundary $\partial\Omega$ and initial conditions (ICs) $q^0(x)$ are given while $q(x,t)$ is the unknown quantity for which we solve the PDE. Eqs. 1-3 are a common form for governing a fluid flow, though other types of BCs and constraints (such as incompressibility) can also be used.

A. Spatial Discretization

We solve our PDEs with the classical finite difference methods with uniform time step Δt and all prognostic variables defined on a regular grid with space step Δx. For 1-D fields, and denoting the k-th variable field in q by z, we use $z^n_j = q(j\Delta x, n\Delta t)_k$ to denote the value of z at the j-th location on the n-th time step. We employ staggered representations of scalar fields and velocities using Arakawa C-grids[63], and the notation $z^n_{1/2}, z^n_{3/2}, \ldots$ for shifted variables in Fig. 1b.

B. Time Stepping

Given the discretized variable fields q^n at time $t = n\Delta t$, a time stepping scheme is used to compute the next fields q^{n+1}. Here we consider the broad range of schemes in which each occurence of q in the definition of $\mathcal{F}$ is replaced by a weighted average of q^n and q^{n+1}, and the weighting may be different for each field and each term of the PDE. Thus when q^n is used in every case we have an explicit method, while using $(q^n + q^{n+1})/2$ in every case gives a Crank-Nicholson method[64]. Denoting the discretized version of $\mathcal{F}$ by $\widetilde{\mathcal{F}}$, the scheme can be written as a system of equations

$$q^{n+1}_j = q^n_j + \Delta t \widetilde{\mathcal{F}}[q^n, q^{n+1}] \tag{4}$$

Because q^{n+1} appears on both sides of the equation, we must solve the equations to obtain it, for example by using iterative

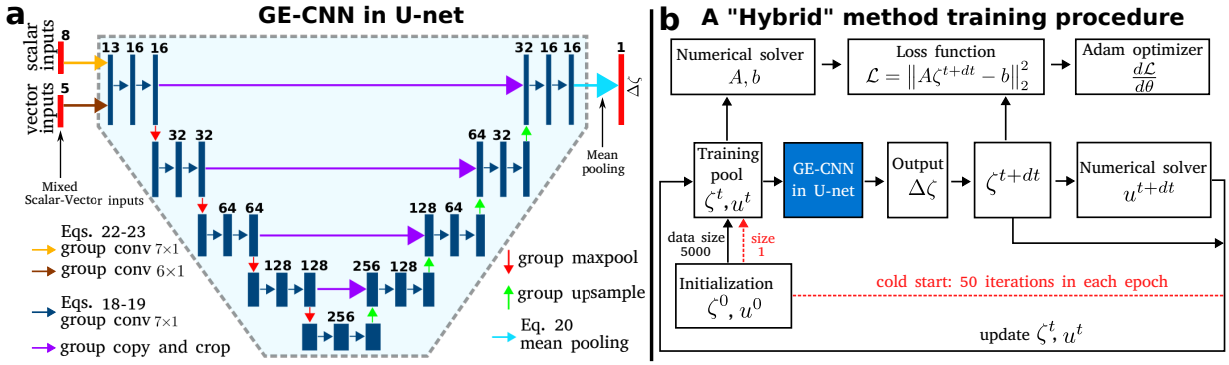

FIG. 2. Schematic representation of equivariant U-net architecture and 'hybrid' training procedure for a deep SWE solver. (a) Equivariant U-Net architecture with an example of channels. Grid staggering leads to input channels with different sizes, so we use equivariant input layers with different kernel sizes to obtain a uniform size across channels in the first hidden layer. All activations except for inputs and outputs are defined as real-valued functions on the infinite discrete group H containing reflections and translations (light blue area). Network outputs provide a fluid surface update $\Delta\zeta$. (b) A 'hybrid' training loop adapted from[34]. A pool of system states is first filled with the initial conditions ζ^0, u^0. For a randomly selected batch of system states, the U-net then generates ζ^{t+dt} at the next time step while velocity is calculated as in a numerical solver. The unsupervised loss function summed over the batch, its gradients are used to update network parameters and the new states overwrite their own inputs in the pool.

methods. Calculating $\widetilde{\mathcal{F}}$ for the discretized variable fields requires discretized versions of all spatial derivatives, which in general must be designed and tested for each PDE to ensure accuracy and stability. This class of time stepping schemes is widely used for fluid dynamical PDEs: examples include incompressible Navier Stokes[65–67] and certain shallow water solvers with land-water boundaries[68] (see below).

III. UNSUPERVISED LEARNING OF PDE INTEGRATION

A. Problem Statement

We aim to replace an expensive semi-implicit time scheme (Eq. 4) with a faster, neural-network based solver. Critically, we do not assume that we have access to simulation data for training purposes, but must train the network using only our knowledge of the PDE, spatial discretization and time stepping scheme.

Concretely, we wish to train (that is, optimize) the parameters ϕ of a flexible function approximator $\widehat{S}_\phi$, such that $\widehat{S}_\phi(q^n) \approx S(q^n) = q^{n+1}$. Here S denotes a single step of time integration using a classical numerical solver that acts as our target reference solution. We aim to achieve a close approximation between $\widehat{S}_\phi$ and S on PDE integrations with initial distributions drawn from a specified probability distribution $\Pi(q)$:

$$q^0 \sim \Pi(q) \tag{5}$$

$$q^n = S^{(n)}(q^0) \tag{6}$$

$$\widehat{q}^n = \widehat{S}_\phi^{(n)}(q^0) \tag{7}$$

$$\widehat{q}^n \approx q^n \tag{8}$$

Without access to simulation data, we cannot carry out supervised training of $\widehat{S}_\phi$ using input-output pairs (q^n, q^{n+1}). The motivation behind this problem formulation without access to training data is that it avoids expensive simulations, and does not require us to commit to a fixed set of simulated system states at the onset of training.

B. Physics-derived Loss Function

To train $g_\phi \approx S$ without simulation data, we construct a physics-derived loss [34,41] that is zero if and only if the discretized PDE is precisely solved:

$$\mathcal{L}_{\text{PDE}}(q,\phi) = \left\| q + \Delta t \widetilde{\mathcal{F}}[q, g_\phi(q)] - \widehat{S}_\phi(q) \right\|_2^2 + \mathbb{1}_\Omega \cdot \left\| \widehat{S}_\phi(q) - q_\Omega \right\|_2^2 \tag{9}$$

The first loss term measures deviations from the prescribed PDE, and is clearly zero when $\widehat{S}_\phi = S$. The second term measures violation of the Dirichlet BCs, with $\mathbb{1}_\Omega$ an indicator function for the boundary. If $\widehat{S}_\phi$ is constructed to satisfy the BCs for any ϕ the second term can be dropped.

C. Training Algorithm

In principle, we could minimize $\mathcal{L}_{\text{PDE}}$ using any input fields q, but to obtain optimal results when ICs are drawn from $\Pi(q)$, we should train on fields likely to occur when time-integrating from those ICs. We therefore train $\widehat{S}_\phi$ on fields it has itself integrated, following the strategy[34] in Fig. 2b.

We first initialize a pool of 5000 simulations with ICs drawn randomly from $\Pi(q)$. For each gradient step, a batch of simulations is randomly selected from the pool, and stepped forward using $\widehat{S}_\phi$. The fields at the old and new time steps for this batch are used to compute $\mathcal{L}_{\text{PDE}}$, and the resulting gradients are used to update ϕ. The updated simulations are then

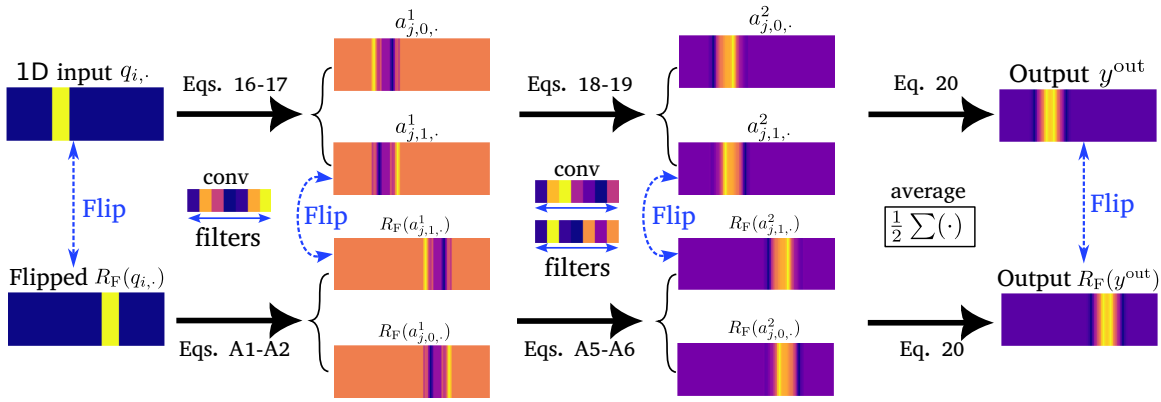

FIG. 3. Reflection-equivariant 1-D convolutional neural network (GE-CNNs). For illustrative purposes, a network with only one channel in each layers is shown. Blue arrows indicate a pair of fields that are reflections of each other, while black arrows indicate equivariant convolution layers. The 1-D input and its flipped duplicate are shown on the left side. Activations in each layer are computed by applying both a standard and flipped version of the convolutional filter to the previous layer. The final output is obtained by averaging over reflected and non-reflected version of each channel pooling on the geometrical two features.

stored in the simulation pool, where they overwrite their own previous values. After every 50 gradient steps, a randomly selected simulation is overwritten with a new initialization from $\Pi(q)$. We use a default batch size of $b = 100$, 60 epochs, and the Adam optimizer[69] with an initial learning rate of 0.001.

D. Hybrid Solvers

For a fully data-driven architecture, the time stepping function $\widehat{S}_\phi$ can be fully specified by a deep neural network or other general function approximator. However, several studies have shown that combining deep learning and numerical physics in a single model can provide better results than either approach alone[35,48,49,70,71]. A particular focus of this hybrid approach has been semi-implicit numerical schemes that require a system of equations to be solved at each time step. Classical numerical solvers for these schemes often use variable substitution to reduce the number of equations and unknowns: examples include elimination of velocity when solving for fluid height in the SWEs[68] and the pressure projection step for incompressible NS. Hybrid approaches therefore train a deep neural network to efficiently solve the reduced set of equations, after which the remaining output variables are calculated using formulae from the original numerical solver. We describe this approach in detail for a SWE scheme in Fig. 2b.

This hybrid approach can offer several benefits compared to a learning a fully data-driven time stepping scheme. By replacing only expensive computations, it retains some inductive biases of the original scheme, and ensures that the full set of updated fields are accurate when the learned computations are correct. Fewer input and output channels for the trained model also reduce parameter counts and improve optimization and data efficiencies.

E. Neural Architectures

To learn (hybrid) time stepping for spatially structured fields, we employ the U-net architecture[72]. The U-net is a convolutional encoder-decoder network. In the encoder, the number of channels increases with depth while spatial resolution decreases, while the decoder enacts the opposite transformations while receiving skip connections from the encoder at each resolution (Fig. 2a). At each resolution, the encoder and decoder employ two convolution layers with kernel size 7. For input fields with C-grid staggering, different kernel sizes are used to achieve a uniform spatial extent for output fields in the encoder's first convolution layer. The final outputs of the U-net are interpreted as updates Δq, and added to the corresponding input fields q^n to produce the time-step fields q^{n+1}. The default resolution of input is 200 for mass grids and 199 for velocity grids. The number of parameters for the network is changed by using a multiplier for the channel counts of all hidden layers (Fig. 2 shows a multiplier of 16).

IV. GEOMETRIC CONSTRAINTS

Many PDEs have symmetries: certain spatial transformations of initial and boundary conditions lead to a corresponding transformation of the system state at future time points. We aim to improve ML-based PDE solvers by endowing them with these properties as a hard constraint built into the neural architecture. This effectively narrows the class of functions through which we are searching for an effective and efficient solver, by filtering out functions inconsistent with the symmetry.

A. Equivariance

Suppose we have a finite group of symmetries $g \in G$ acting

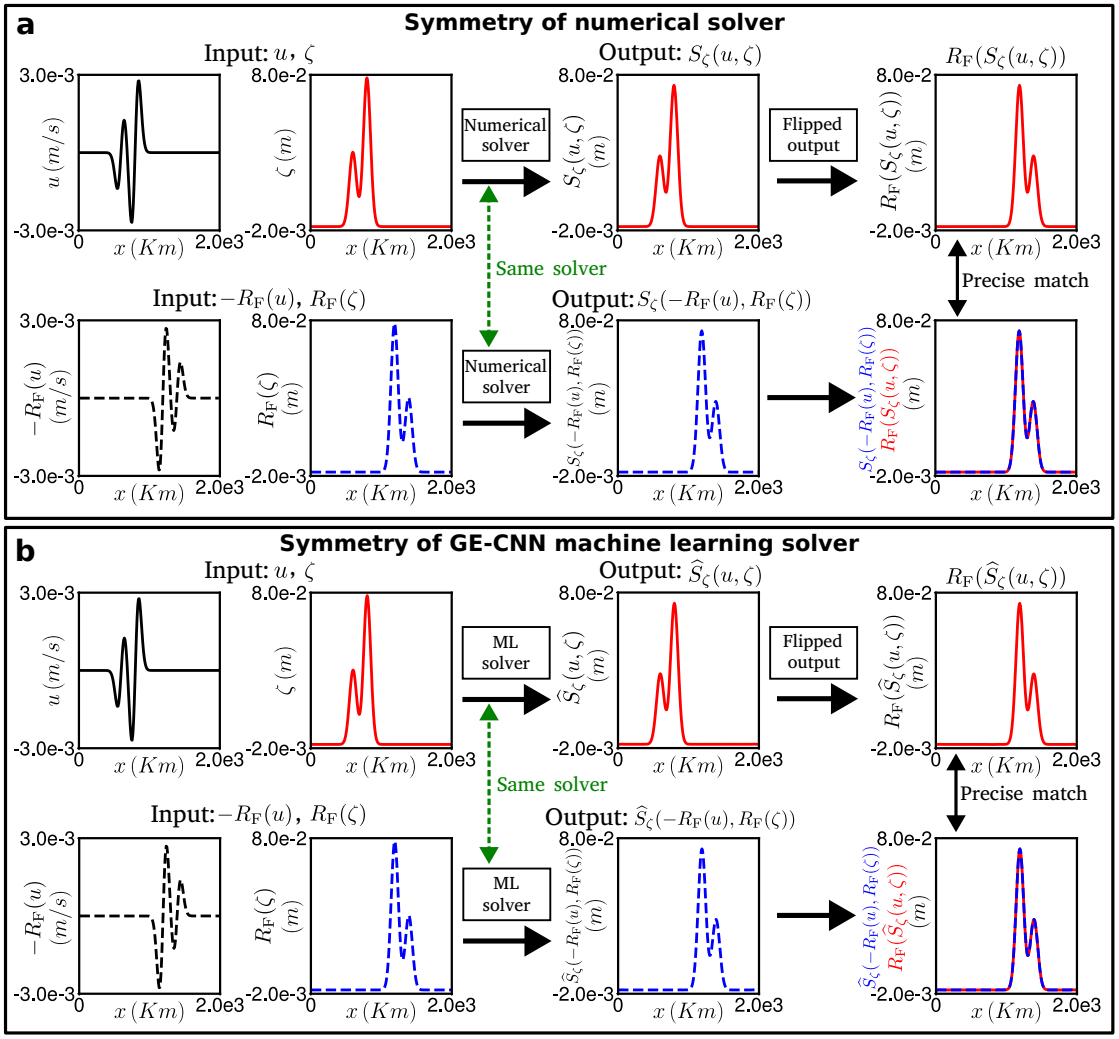

FIG. 4. Empirical verification of identical PDE and GE-CNN symmetries for the shallow water equations. (a) ICs, ζ, and u for the numerical solver (top left). A flipped version is shown with inversion of u (bottom left). One step of numerical integration produces a new ζ (top, third column), from which we compute a flipped version. Integration from the flipped/inverted ICs produces a different ζ (bottom, third column), which precisely matches the flipped output from the non-flipped ICs (bottom right). (b) As in 'a,' but for a trained U-net based solver. Note however that the equivariance of the network is not learned, but exists as a hard constraint throughout the training process.

on a set of spatially extended fields q by transformations $q \longrightarrow \mathcal{T}_g(q)$, with $\mathcal{T}_{g_1 g_2} = \mathcal{T}_{g_1} \circ \mathcal{T}_{g_2}$. A function $\Psi(q)$ is equivariant when transforming its inputs is equivalent to transforming its outputs. Concretely, for each $g \in G$ the transformations $\mathcal{T}_g, \mathcal{T}'_g$ act on Ψ's inputs and outputs respectively, and

$$\forall g, q : \Psi(\mathcal{T}_g q) = \mathcal{T}'_g \Psi(q) \tag{10}$$

For example, let Ψ denote time integration of the 2D heat equation $\frac{\partial q}{\partial t} = \kappa\left(\frac{\partial^2 q}{\partial x_1^2} + \frac{\partial^2 q}{\partial x_2^2}\right)$. This Ψ is equivariant to rotations, reflections, and translations of the heat field q. In this case both the input and output transformations are simply the same point-to-point mappings of the scalar heat fields, but for more complex PDEs involving vector fields the transformations can be more involved (see below).

B. Equivariant Convolutions

We now describe the construction of convolutional networks with equivariance as a hard constraint, meaning that the $\widehat{S}_\phi$ is equivariant for any ϕ. Throughout this section we follow[59], but simplify notation by describing a single input and output channel, both of which are n-D fields of the same size. We denote by $\mathbb{Y} \subset \mathbb{Z}^n$ the regular grid of integer valued coordinates on which the input and output channels are defined.

A standard convolutional layer applies an n-D convolutional filter W to a spatially extended scalar input field q to produce a scalar output field $q \star W$:

$$[q \star W](x) = \sum_{y \in \mathbb{Y}} q(y) W(y - x) + b \tag{11}$$

where $W(y) = 0$ for y outside the spatial extent of the filter. q, W, and $q \star W$ are simply real-valued functions on $\mathbb{Y}$, while b is a scalar. Standard convolutions are equivariant with respect to translations, but not other symmetries.

In contrast, equivariant convolutional layers produce outputs that are real-valued functions of an extended discrete group H generated by the symmetry group G of interest as well as translations in $\mathbb{R}^n$. The first such layer takes standard scalar fields as input:

$$[q \star W](h) = \sum_{y \in \mathbb{Y}} q(y) W(h^{-1} y) + b \tag{12}$$

Subsequent layers use functions on H as both inputs and outputs:

$$[\gamma \star W](h) = \sum_{h' \in \mathbb{H}} \gamma(h') W(h^{-1} h') + b \tag{13}$$

γ, W, and $\gamma \star W$ are real-valued functions of H. As shown in[59], Eqs. 12-13 satsify equivariance (Eq. 10). The input transformation in Eq. 12 is simply G's action on $\mathbb{Z}^n$ described by G, while other transformations act on real-valued functions of H. For any such function $\alpha(h)$ we have

$$[T_h \alpha](h') = \alpha(h^{-1} h') \tag{14}$$

We visualize functions on H as collections of maps over $\mathbb{Y}$, with one map (i.e. real-valued function on $\mathbb{Y}$) for each $g \in G$ (Fig. 3). Since the nonzero regions of q, γ and W are bounded, the outputs' nonzero regions are as well. In Eq. 12, W is defined on a patch of $\mathbb{Z}^n$, but in Eq. 13 W is a function on H. To include multiple input output and channels, we simply sum over inputs for each output in Eqs. 11-13, and note that the arrays storing W acquire two additional dimensions. The bias b is then indexed by the output channel, but not by location or group element.

To build an equivariant convolutional network, a sequence of equivariant convolutional layers is interspersed with pointwise nonlinearities. To obtain an equivariant final output on $\mathbb{Y}$ instead of H, a pooling operation (e.g. a mean or maximum) operates 'along the G-axis' of the G-indexed collection of maps on $\mathbb{Y}$. Concretely, we take the mean or maximum over all elements of H that share the same translational component.

For a symmetry group of size $|G|$, a standard convolution with c input and output channels has as many parameters as an equivariant layer with $c/\sqrt{|G|}$ channels. When $\mathbb{Y}$ is a D-dimensional grid with N points per axis, the computational complexity of the forward and backward passes is $O(N^D K^D c^2)$ in both cases, since the equivariant network has $|G|$ times fewer input-output channel pairs but each input channel must be convolved with a transformed slice of the filter bank $|G|$ times.

1. Reflection-equivariant 1-D Convolutions

Having introduced equivariant convolutions for any finite symmetry group G acting on a discrete grid $\mathbb{Y}$, we next focus on the concrete example of 1-D reflective symmetries. This two-element group is the only nontrivial symmetry group and the main focus of this paper. Here $\mathbb{Y} = \{-N, \ldots, N\}$, G contains the identity and a reflection, and H consists of either element of G followed by any translation. In this case, a convolutional time stepping network $\widehat{S}_\phi : q^n \to \hat{q}^{n+1}$ is equivariant if $R_\mathrm{F}(\widehat{S}_\phi(q)) = \widehat{S}_\phi(R_\mathrm{F}(q))$, for any input field(s) q and where the 'mirroring' operator R_F reflects the fields on the spatial axis, i.e. $[R_\mathrm{F}(q)](x) = q(-x)$.

A standard 1-D convolutional layer in Eq. 11 with c_in inputs q and c_out outputs a is defined as:

$$a_{j,\cdot} = \sum_{i=1}^{c_\mathrm{in}} W_{j,i,\cdot} \star q_{i,\cdot} + b_j \tag{15}$$

where the $\cdot$ symbol denotes all values along a given axis. W is an $c_\mathrm{in} \times c_0 \times K$ array for filter size K, while b is a c_out-element vector.

In 1-D reflection-equivariant networks, the first layer is a special case with c_in input channels q defined on $\mathbb{Y}$ and c_1 outputs a^1 defined on H. Eq. 12 thus becomes:

$$a^1_{j,0,\cdot} = \sum_{i=1}^{c_\mathrm{in}} W^1_{j,i,\cdot} \star q_{i,\cdot} \qquad + b^1_j \tag{16}$$

$$a^1_{j,1,\cdot} = \sum_{i=1}^{c_\mathrm{in}} R_\mathrm{F}\left(W^1_{j,i,\cdot}\right) \star q_{i,\cdot} + b^1_j \tag{17}$$

While q, W^1, b^1 have the same size here as in standard convolutional layers, a^1 gains a third dimension that indexes the elements of G.

For subsequent layers, both the $c_{\ell-1}$ input channels $a^{\ell-1}$ and c_ℓ outputs a^ℓ are defined on H and are stored in 3-D arrays. Eq. 13 becomes:

$$a^\ell_{j,0,\cdot} = \sum_{i=1}^{c_{\ell-1}} W_{j,i,0,\cdot} \star a^{\ell-1}_{i,0,\cdot} \qquad + W_{j,i,1,\cdot} \star a^{\ell-1}_{i,1,\cdot} \qquad + b^\ell_j \tag{18}$$

$$a^\ell_{j,1,\cdot} = \sum_{i=1}^{c_{\ell-1}} R_\mathrm{F}\left(W_{j,i,1,\cdot}\right) \star a^{\ell-1}_{i,0,\cdot} + R_\mathrm{F}\left(W_{j,i,0,\cdot}\right) \star a^{\ell-1}_{i,1,\cdot} + b^\ell_j \tag{19}$$

The filter bank W now has four dimensions, the third of which indexes G. When computing results at the second index along the second, G-indexing dimension of each output channel (Eq. 19), the filters are flipped on the spatial axis and permuted on the G axis (Fig. 3). While the equivariance of these layers follows as a special case of the results in[59], we include simple proofs for the case of 1-D reflections in the Appendices A 1-A 2.

Finally, to produce a network output that is defined on a simple 1-D grid (not as a function on H), we use a mean pooling operation over the symmetry dimension

$$y^\mathrm{out}_{j,\cdot} = \left(a^L_{j,0,\cdot} + a^L_{j,1,\cdot}\right)/2 \tag{20}$$

which also obviously has the desired equivariance property. Thus, by chaining together these input, internal, and output layers, our entire network Ψ is reflection equivariant.

2. Extension to Mixed Scalar-Vector Inputs

Unfortunately, the equivariance defined for convolutions above does not match the reflection symmetries of many PDEs, since it fails to account for differences in how vector and scalar fields are affected by rotation and reflection. For a scalar field, the value of the transformed field (e.g. heat) is simply the value of the original field at a different point. But for vector fields (e.g. velocity) both a location change and a reflection/rotation of the vector at the corrected location are required. Simply transforming each component of the velocity field in a PDE solution as a separate scalar would yield a new field that does not solve the PDE.

For reflections of 1-D vector field u, the necessary transformation is

$$[R_{\mathrm{F}}u](x) = -u[-x] \tag{21}$$

To implement the proper transformation when q contains both scalar fields ζ and vector fields u, we define the following input layer (compare to Eq. 19):

$$a^1_{j,0,\cdot} = \sum_{i=1}^{c^\zeta_{in}} W^\zeta_{j,i,\cdot} \star \zeta_{i,\cdot} \qquad + \sum_{i=1}^{c^u_{in}} W^u_{j,i,\cdot} \star u_{i,\cdot} \qquad + b^\ell_j \tag{22}$$

$$a^1_{j,1,\cdot} = \sum_{i=1}^{c^\zeta_{in}} R_{\mathrm{F}}(W^\zeta_{j,i,\cdot}) \star \zeta_{i,\cdot} + \sum_{i=1}^{c^u_{in}} -R_{\mathrm{F}}(W^u_{j,i,\cdot}) \star u_{i,\cdot} + b^\ell_j \tag{23}$$

Since the output of this layer is a real-valued function on H, subsequent equivariant layers can be used without modification in Eq. 13. Defining an equivariant output layer to produce vector fields is straightforward, but because we construct hybrid solvers (see below) scalar outputs are sufficient for our purposes.

We prove the equivariance of our input layer in A 3. For an empirical confirmation of this, Fig. 4 shows the equivariance of a classical PDE solver S and a trained equivariant convolutional network $\widehat{S}_\phi$ for the shallow water equations, which govern scalar height and vector velocity fields.

C. One-dimensional shallow water equations

We evaluated our learning strategies using the 1-D SWEs, composed of momentum and continuity equations:

$$\frac{\partial u}{\partial t} = -C_D \frac{1}{h} u|u| - g\frac{\partial \zeta}{\partial x} \tag{24}$$

$$\frac{\partial \zeta}{\partial t} = -\frac{\partial (hu)}{\partial x} \tag{25}$$

with spatial coordinate $x \in [0, L]$, time t, velocity u, surface disturbance ζ, total depth $h = d + \zeta$, bottom drag C_D, and gravitational acceleration g. SWEs are commonly used to describe large-scale flows in coasts, oceans, estuaries, and rivers, based on the assumption that fluid depth is well below the length scale of horizontal motion, as illustrated in Fig. 1a. Our dry BCs, common in riverine and coastal models[68,73,74], mean that no fluid enters or escapes:

$$u(x = 0) = u(x = L) = 0 \tag{26}$$

$$\zeta(x = 0) = \zeta(x = L) = 0 \tag{27}$$

By default we use 'Gaussian bell' ICs:

$$u(x, 0) = 0 \tag{28}$$

$$\zeta(x, 0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/\sigma^2} \tag{29}$$

with μ and σ uniformly distributed on $[100\,Km, 1900\,Km]$ and $[10\,Km, 100\,Km]$ respectively.

Substituting $\zeta \leftarrow R_{\mathrm{F}}(\zeta), u \leftarrow -R_{\mathrm{F}}(u)$ into Eqs. 24-27 demonstrates reflection equivariance of the SWEs, which we confirmed empirically in Fig. 4a.

Since closed, wave-reflecting boundaries tend to require minuscule time steps for explicit schemes, we used a semi-implicit scheme[68] to generate reference simulations and to construct loss functions (Eq. 9) for unsupervised learning. This finite difference method stores velocities and surface elevations on staggered grids (details in the B, simulation parameters in Table I). Its computational cost is dominated by solving a tridiagonal linear system

$$A\zeta^{n+1} = b \tag{30}$$

Where A, b are functions of ζ^n and u^n. The relative costs of calculating the coefficients of A and b or computing u^{n+1} given ζ^{n+1} are negligible.

TABLE I. Simulation parameters used for SWEs.

Parameters	Explanation	Value
L	simulation domain	$2000\,(Km)$
d	undisturbed water depth	$100\,(m)$
C_D	bottom drag coefficient	$1.0e-3$
g	acceleration due to gravity	$9.81\,(m/s^2)$
Δx	space step	$10\,(Km)$
Δt	time step	$300\,(s)$
$w_{\mathbf{imp}}$	implicit weighting	0.5

V. EVALUATION METRICS

We compare trained solvers $\widehat{S}_\phi$ to a reference numerical method S with four error measures described previously[75]. We calculate these metrics for each individual field $\widehat{z}^n$ estimating $z^n \in q^n$.

- *Normalized Root Mean Square Error* (NRMSE) describes a relative difference between estimated and reference solutions:

$$\mathrm{NRMSE} = \frac{\|\widehat{z}^n - z^n\|_2}{\|z^n\|_2} \tag{31}$$

  Note that all estimated fields $\widehat{z}^n$ are integrated n time steps from the ICs of the reference solution.

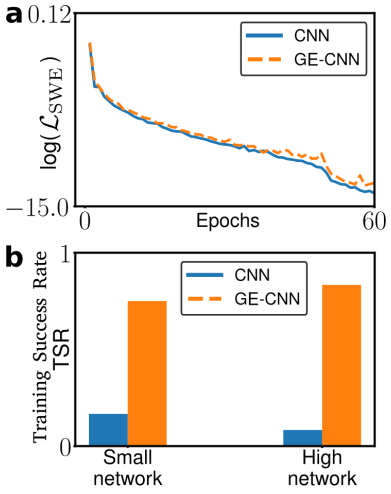

FIG. 5. Progress and success rate of training for standard and reflection-equivariant convolutional networks. (a) Training loss curves for standard CNN and GE-CNN solvers. During training the loss of GE-CNN solver is a little larger than one of CNN solver. (b) Training success rate standard and GE-CNNs. Here, high- and small network respectively have about 148M and 1.6M parameters.

- *Time-averaged Normalized Root Mean Square Error* ($\mathbb{E}_t$NRMSE) averages the NRMSE over the full duration of a simulation (in this work, 1200 time steps spanning 100 simulated hours).
- *Pearson's correlation* $\rho(\widetilde{z}^n, z^n)$ of reference and learned solutions.
- *Training Success Rate* (TSR) is the probability that training will converge to a $\mathbb{E}_t\mathrm{NRMSE} < 10$. This measure allows us to incorporate the stochastic aspect of deep learning in evaluating performance (both the initial weights and the order of ICs differ across runs). In this work, we quantify TSR by using $\mathbb{E}_t\mathrm{NRMSE} - \zeta$ for SWEs.

VI. SOFTWARE IMPLEMENTATION

Classical numerical and machine learning solvers are implemented in Pytorch and Numpy. Code for equivariant convolutions is partly adapted from GrouPy at `https://github.com/jornpeters/GrouPy/tree/pytorch\_p4\_p4m\_gconv/groupy/gconv`, while code for training on evolving simulations is based partly on code published in[34], `https://github.com/wandeln/Unsupervised_Deep_Learning_of_Incompressible_Fluid_Dynamics`. The code for both numerical- and ML solver are publicly at `https://github.com/m-dml/GE-CNN_learning_SWEs`.

VII. EXPERIMENTS

We carried out unsupervised training of a hybrid PDE integration scheme for the SWEs, to determine whether hard symmetry constraints improve long-term accuracy and stability. Our reference simulations (see section IV C) used a staggered grid with 200 height and 199 velocity points shown in Fig. 1b.

The neural network inferred surface height ζ_n from the system state at step $t-1$. Following[34], we provide 13 input channels describing u_{n-1}, z_{n-1} and the BCs:

$$\begin{aligned}\text{input} = \big(&\zeta, h, m^b_\zeta, m_\zeta, m^b_\zeta \cdot \zeta, m_\zeta \cdot \zeta, m^b_\zeta \cdot h,\\ &m_\zeta \cdot h, u, m^b_u, m_u, m^b_u \cdot u, m_u \cdot u\big)\end{aligned} \tag{32}$$

Here, m_ζ is a mask for ζ. It is zero for boundary values and one for interior values while $m^b_\zeta = 1 - m_\zeta$, and m_u, m^b_u are the same for velocities. To deal with different spatial dimensions across input channels, we used kernel size 6 for ζ-sized inputs and kernel size 7 for u-sized inputs, and added the results in Fig. 2a.

In our hybrid scheme, the neural network replaced the expensive tridiagonal solve in Eq. 30 to compute ζ^{n+1}, while the numerical scheme computes coefficients of the tridiagonal system and updates u^{n+1} while imposing BCs. We can therefore drop the second term in Eq. 9 and replace the first with:

$$\mathcal{L}_{\mathrm{SWE}} = \left\| A\zeta^{n+1} - b \right\|_2^2 \tag{33}$$

We compared equivariant networks (GE-CNN) to standard convolutional U-nets with the same architecture, loss and training procedure. For both network types, we adjusted the number of trainable parameters by scaling the number of output channels for all convolutions except the final layer, and for our default configuration as shown in Fig. 2a. This resulted in 148M parameters for the equivariant convolutions and 149M for standard convolutions.

A. Equivariance Improves Accuracy and Convergence

We compared default configurations of our standard convolutional neural networks (CNN) to reflection-equivariant versions (GE-CNN), see Fig. 2a. We trained on a library of Gaussian bells ICs with occasional resets, as shown in section III C. Both standard and equivariant convolutional networks achieved low loss values and accurately predicted how the next time step for the SWEs (Fig. 5a). Since the standard CNN architecture describes a less restricted function class than the GE-CNN, it achieves a slightly lower loss value during training, at the cost of failing to respect symmetry in autoregressive predictions.

To test whether respecting symmetry would improve long term-accuracy, we therefore evaluated we compared the CNN and GE-CNN after training in autoregressive rollouts. Network outputs were used to define inputs for the next time step, and the results were compared to reference numerical solutions over 1200 time steps (100 simulated hours, Fig. 6a). The reference solutions describe waves propagating outward from the initial Gaussian bell before reflecting off the domain boundary five times (Fig. 6a, Fig. 6d black).

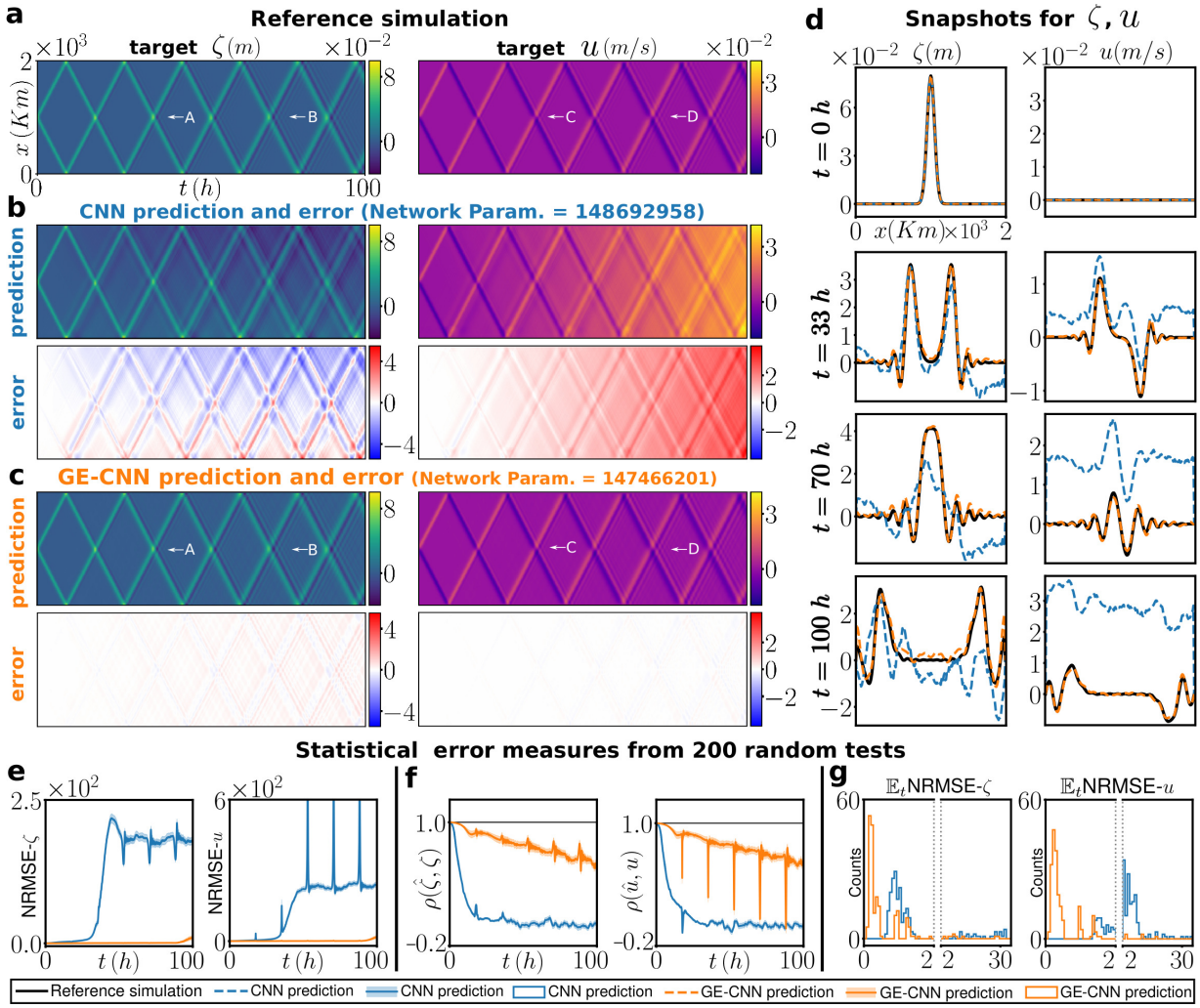

FIG. 6. GE-CNNs produce accurate SWE rollouts from Gaussian bell ICs. (a) Reference simulation of surface elevation ζ and velocity u. (b) Rollouts from CNN solver for ζ and u, with errors (prediction − reference). (c) As in 'b,' but for GE-CNNs. (d) Snapshots of ζ and u obtained from the reference solver (black), CNN (blue) and GE-CNN orange. (e) Mean over ICs of NRMSE for ζ and u as a function of time from the start of the simulation, for CNNs (blue) and GE-CNNs. (f) As in 'e,' but for correlation. (g) Histograms of time averaged NRMSE in ζ and u for CNNs (blue) and GE-CNNs (orange).

Individual CNN rollouts successfully reproduced the propagation and reflection of waves, but exhibited gradually increasing errors that increased when the waves were reflected by the closed boundaries (Fig. 6b). Over the CNN rollout the waves broadened, developed additional peaks in ζ not present in the reference simulation. Compared to to reference simulations, CNN rollouts exhibited higher spatial frequencies, a positive velocity bias and spatially asymmetric errors (Fig. 6d, blue). By the end of the rollout the magnitude of errors reached the amplitude of the simulated wave heights and velocities.

The GE-CNN followed the reference solution more closely, with errors at least one order of magnitude smaller than the simulated signals, and difficult to discern visually (Fig. 6c). The shape and width of the propagating and reflecting waves and surrounding smaller ripples closely matched the reference simulation. By the end of the simulation, errors appear as additional undulations in ζ, while u continues to to follow the reference simulation closely (Fig. 6d, orange).

The training success rate, defined as the probablity over multiple training runs and ICs of achieving low time-averaged error (in Section V), was 10/12 for the GE-CNN but only 2/12 for the CNN (Fig. 5b). When using networks $\sim$ 100 times smaller, we observed 9/12 successes for the GE-CNN (1.6M parameters) and 1/12 for the CNN (1.7M).

We further measured how accuracy of CNN and GE-CNN rollouts varied over time and ICs by computing rollouts for each over 200 ICs. NRMSE (see Section V) in both ζ and u increased more quickly for the CNN, reaching average values 1-2 order of magnitude higher (Fig. 6e). Correlation coefficients between rollouts and the reference simulation followed a similar trend, with a decrease over time but clear superiority for the GE-CNN (Fig. 6f). We further examined the distribution of time-averaged NRMSE across ICs, computing histograms

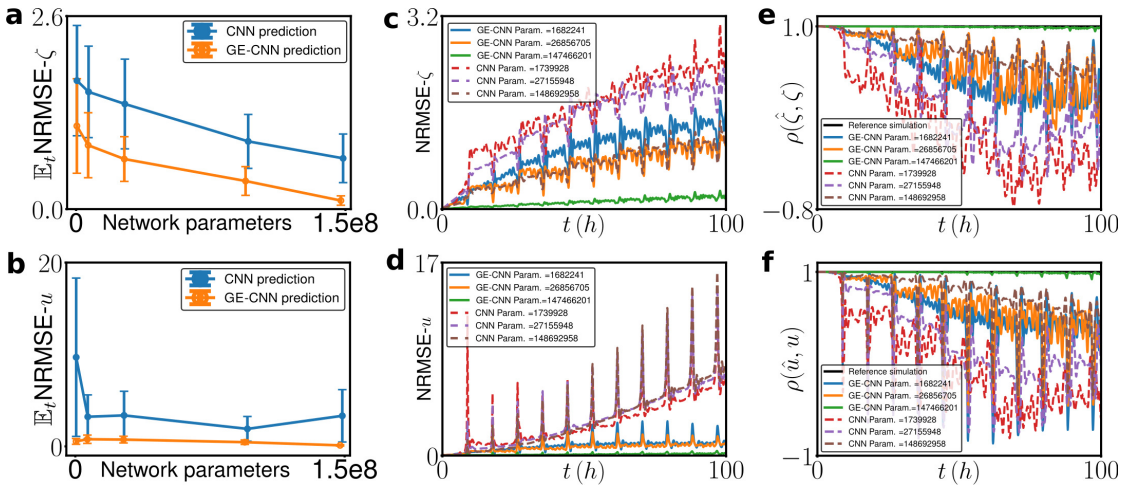

FIG. 7. Reflection equivariance improves accuracy for all network sizes. (a-b) $\mathbb{E}_t$NRMSE-ζ and $\mathbb{E}_t$NRMSE-u as functions of network parameter counts in CNNs and GE-CNNs. Error bars show standard deviations which are obtained by using the predictions in $100\,h$. (c-f) Plots of NRMSE-ζ, NRMSE-u, $\rho(\hat{\zeta},\zeta)$, and $\rho(\hat{u},u)$ as functions of integration time for several sizes of CNN and GE-CNN.

for u- and ζ-errors on logarithmic scales (Fig. 6g). The GE-CNN error distributions peaked near zero, while CNN errors peaked around the targeted signals' amplitude, with a long-tailed distribution.

For both CNN and GE-CNN rollouts, $\mathbb{E}_t$NRMSE-ζ decreases with parameter count of the trained networks, and is lower for GE-CNNs (Fig. 7a-b). The largest CNNs computed ζ as accurately as the smallest GE-CNNs, but for u even the smallest GE-CNNs outperformed CNNs of all sizes tested. The same trend was observed at individual time points for NRMSEs (Fig. 7c-d) and correlation values (Fig. 7e-f). Overall, these results show that the long-rollout accuracy improvement provided by equivariance is robust to the choice of network size and accuracy metric.

B. Generalization Capabilities After Training

We next examined how well PDEs solvers trained using Gaussian bell ICs would perform on conditions beyond their training data.

We first measured rollout accuracy for an ICs described by an isosceles triangle in ζ a the domain center with 400 km base and 0.12 m height (Fig. 8a). This system state was never encountered during training, as it contains a discontinuous first spatial derivative in contrast to the smooth Gaussian bell. Rollout errors we were considerably higher than for a novel Gaussian bell scenario, though the GE-CNN was again more accurate, especially for u (Fig. 8b-c). In the CNN rollout the propagating wave dissipated into many high-frequency ripples, while for the GE-CNN waves propagated with the correct shape but too slowly, leading to a position mismatch with reference simulations (Fig. 8d). CNN and GE-CNN rollout performance for triangular ICs with random height, width and position (uniform on 0.09-0.36 m, 200-300 km and 0-200 km respectively) showed similar trends to Gaussian bells. GE-CNN was uniformly superior, its errors grew more slowly over time, and its error distribution peaked near zero while the CNN's peaked above 4 times the estimated signals (Fig. 8g).

For a more challenging generalization task, we used a sum of 3 Gaussian bells as an initial condition. The reference simulation (Fig. 9a) shows 6 propagating and reflecting waves that form a complex interference pattern. As previously, in CNN rollouts the waves were distorted and dissipated over time with a positive bias emerging for u, while the GE-CNN maintained the correct shapes but introduced timing errors (Fig. 9c), and was more accurate at every time point (Fig. 9d). We also computed accuracy measures for triple-bell ICs (Fig. 9e-g) with random means (uniform on 100-1900 km) and widths (10-100 km). The GE-CNN yielded better NRMSE and correlation values for all time delays, and a distribution of time-averagd NRMSE that peaked near zero and showed little overlap with CNN results.

C. Learned Representations of Global Mass, Momentum and Energy

A challenge for ML-based PDE solvers is that their predictions do not always satisfy conservation laws, even when these laws are manifested in their unsupervised loss or training data[71,76]. Our reference SWE solver conserves mass, conserves energy except for bottom drag, and conserves momentum except for bottom drag and boundary effects. Each reflection of a propagating wave from the closed boundaries involves a temporary conversion of kinetic to potential energy.

We investigated how well trained networks represent 4 quantities: mass $\sum h_i$, momentum $\sum h_i u_i$, kinetic energy $\frac{1}{2}\sum h_i u_i^2$, and potential energy $\frac{1}{2}\sum g h_i$. Note that we employed the disturbed water depth h to represent mass. We computed these for CNN and GE-CNN rollouts and compared to the reference solver. In individual held-out Gaussian bell ICs (Fig. 10a, upper row), CNN rollouts show a rapid error accumulation in all 4 quantities, while GE-CNNs exhibited a

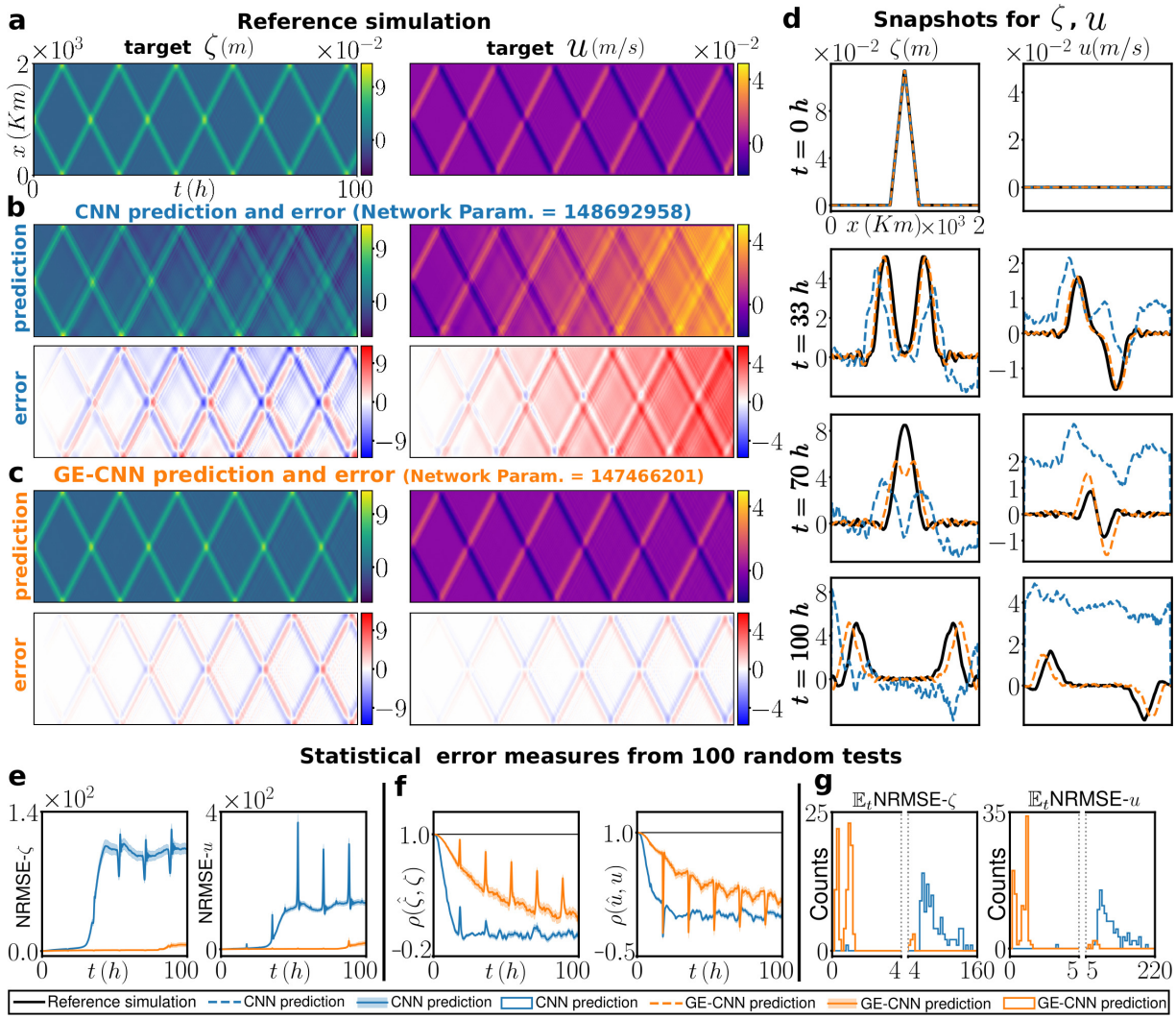

FIG. 8. GE-CNN solver trained on Gaussian bell ICs generalizes to triangular ICs. (a) Reference simulation of surface elevation ζ and velocity u from a triangular IC. (b) Rollouts from CNN solver for ζ and u, with errors (prediction − reference). (c) As in 'b,' but for GE-CNNs. (d) Snapshots of ζ and u obtained from the reference solver (black), CNN (blue), and GE-CNN (orange). (e) Mean over ICs of NRMSE for ζ and u as a function of time from the start of the simulation, for CNNs (blue) and GE-CNNs. (f) As in 'e,' but for correlation. (g) Histograms of time averaged NRMSE in ζ and u for CNNs (blue) and GE-CNNs (orange).

slow drift in total mass that produced a drift in potential energy, and negligible errors in momentum and kinetic energy. When averaging over many ICs, we found that average values of the conserved quantities matched closely for GE-CNN and the reference simulation until almost 100 hours, while the CNN showed clear differences after 50 h (Fig. 10a, lower row). Similar results were observed for triangular (Fig. 10b) and multi-bell ICs (Fig. 10c); for these ICs errors grew more quickly but the GE-CNN matched the reference simulation more closely.

VIII. DISCUSSION

We developed reflection-equivariant 1-D convolutional networks for mixed vector-scalar inputs, and trained them to solve the SWEs with an unsupervised loss. We showed how these networks can be endowed with the same symmetries and the targeted PDEs, and our experiments showed how they improve accuracy and stability over standard CNNs with similar parameter counts, over a broad range of scenarios and tests. GE-CNNs matched reference simulations more closely at all time points, performed on new IC types and more faithfully represented mass, momentum, and energy. A remarkable aspect of these equivariant networks is that their advantages first become apparent when generating and evaluating longer rollouts, with no differences from standard CNNs apparent during training. Our results show that equivariant architectures offer significant benefits for long-term accuracy and physical consistency, with no modifications to the loss function or training procedures.

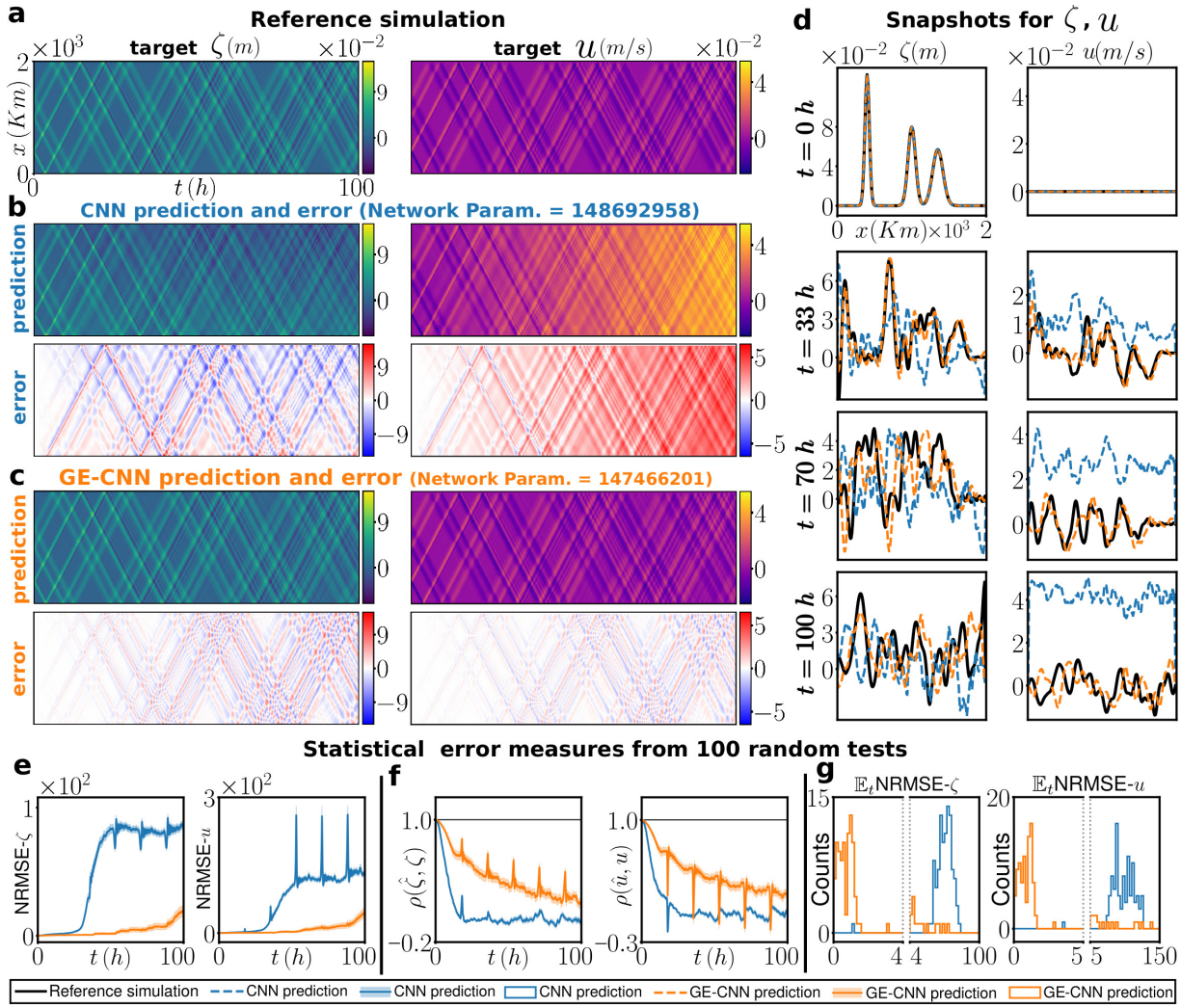

FIG. 9. GE-CNN solver trained on Gaussian bell ICs generalizes to multi-bell ICs. (a) Reference simulation of surface elevation ζ and velocity u from a multi-bell IC (centers 400, 1000, 1350 km; widths 3, 5, 7 km). (b) Rollouts from CNN solver for ζ and u, with errors (prediction − reference). (c) As in 'b,' but for GE-CNNs. (d) Snapshots of ζ and u obtained from the reference solver (black), CNN (blue) and GE-CNN orange. (e) Mean over ICs of NRMSE for ζ and u as a function of time from the start of the simulation, for CNNs (blue) and GE-CNNs. (f) As in 'e,' but for correlation. (g) Histograms of time averaged NRMSE in ζ and u for CNNs (blue) and GE-CNNs (orange).

A. Related Work

Our work builds on and complements existing studies seeking to exploit symmetries for solving PDEs. Authors[77] use PDE symmetries to design data augmentations for use during training, instead of making their networks equivariant. Ref.[78] built steerable CNNs and the paper[79]demonstrated their utility predicting the evolution of incompressible NS and an advected temperature field, but do not consider mixed scalar/vector inputs and examine their predictions only 10 time steps into the future.

In Ref.[80], the authors use network layers that solve a specific PDE to build convolutional networks, instead of constructing layers to match the symmetry groups of a specific PDE as we do here. Authors[81] use rotation equivariant convolutional layers to operate on vector fields, but do not consider mixed input types or solve PDEs.

The equivariant convolution layers we have developed for mixed scalar/vector inputs could also be realized using steerable convolutions[79] with the correct combination of scalar and vector capsules. Instead of transforming filter banks, steerable convolutions are based on optimizing convolution weights within a pre-computed linear subspace that satisfies the desired constraints. While this approach is flexible and efficient, we believe our layers can provide considerable utility through their mathematical simplicity. Some studies have also reported successful implementation of equivariant network through filter bank transformation, but could not achieve the same results through steerable convolutions[82].

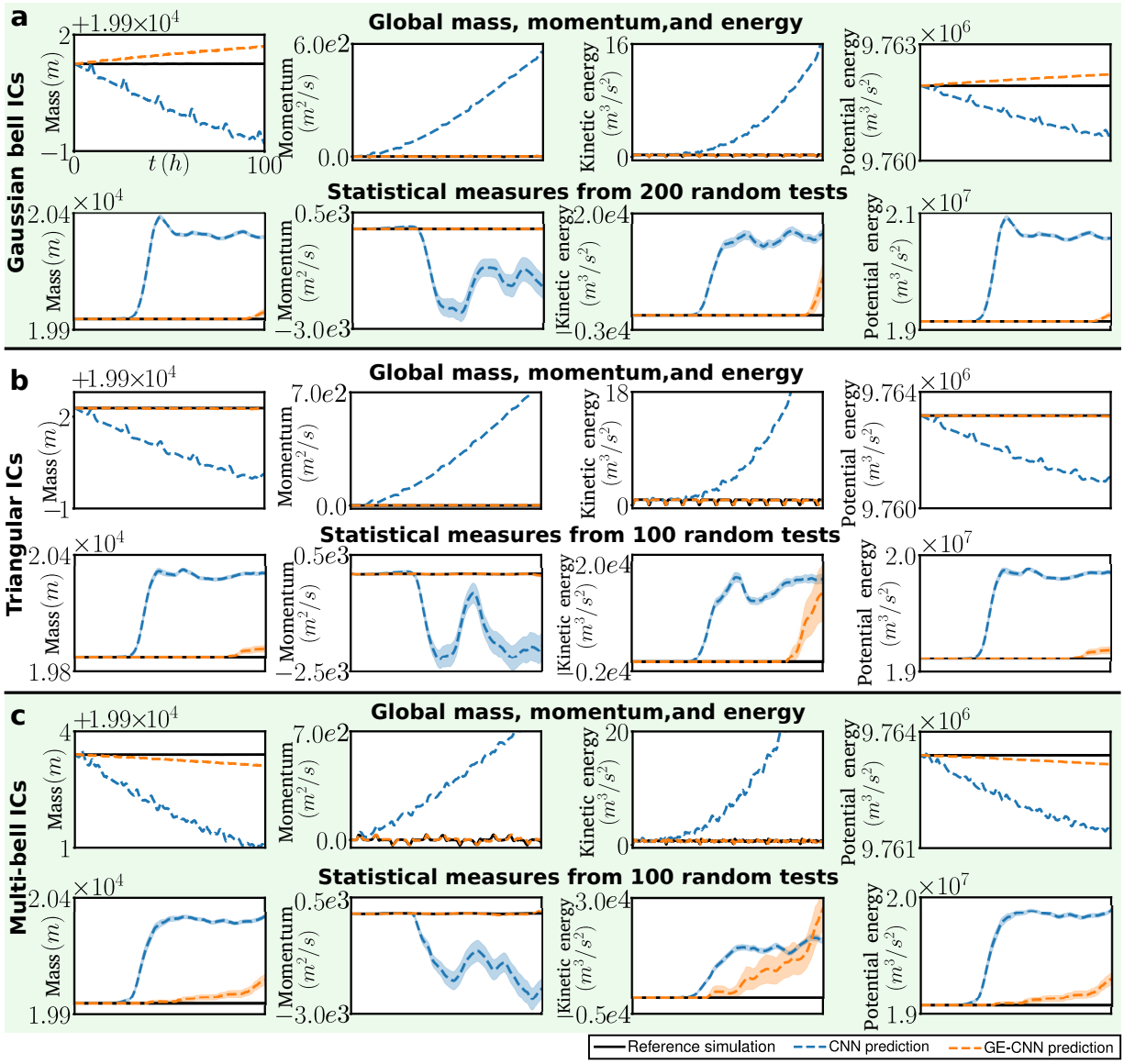

FIG. 10. GE-CNN solver robustly predicts mass, momentum, and energy. (a) Global mass, momentum, kinetic- and potential energy from the reference solver, CNN- and GE-CNN rollouts (upper; same IC as Fig. 6). Mean and standard error of the mean for global mass, momentum, kinetic- and potential energy over 200 Gaussian bell ICs. (b) As in 'a,' but for triangular ICs. (c) As in 'a,' but for multi-bell ICs.

B. Future Outlook

In future work, we intend to extend our results to higher dimensional and more complex systems, and to combine geometric and physical constraints[83,84]. We also anticipate that by offering a combination of long-rollout performance and automatic differentiability, equivariant deep PDE solvers could prove useful for solving inverse problems[85,86]. The observed performance gains for long rollouts could also find useful applications in climate, weather and ocean modeling, which require stability and accuracy over far longer time intervals than commonly evaluated scenarios for deep PDE solvers.

ACKNOWLEDGMENTS

We thank Kai Logemann for providing code and valuable insights into the semi-implicit shallow water solver. We also thank Vien Minh Nguyen-Thanh, Ali Can Bekar, Tobias Schanz, Shivani Sharma, Vadim Zinchenko, and Andrey Vlasenko for useful discussions and comments on the manuscript.

Appendix A: Proof of Equivariance for Convolution Layers

For completeness, we first provide proofs for the equivariance of the original scalar-field-only convolution layers[59]

in our notation. We then prove equivariance for our mixed scalar/vector input layers.

1. Proof of Equivariance for Scalar-input Convolutional Input Layer

The outputs of the first (input) layer of an equivariant convolutional network $a^1_{j,0,\cdot}$ and $a^1_{j,1,\cdot}$ are defined in Eq. (16) and Eq. (17). We flip the input $R_{\mathrm{F}}(q_{i,\cdot})$. Applying this layer to a fliped input using the same weights and biases gives:

$$\tilde{a}^1_{j,0,\cdot} = \sum_{i=1}^{c_{\mathrm{in}}} W^1_{j,i,\cdot} \star R_{\mathrm{F}}\left(q_{i,\cdot}\right) + b^1_j \tag{A1}$$

$$\tilde{a}^1_{j,1,\cdot} = \sum_{i=1}^{c_{\mathrm{in}}} R_{\mathrm{F}}\left(W^1_{j,i,\cdot}\right) \star R_{\mathrm{F}}\left(q_{i,\cdot}\right) + b^1_j \tag{A2}$$

To prove the equivariance in this layer, we flip the first layer of output. This flipping operator can be moved into the convolution. Then, we obtain

$$\begin{aligned} R_{\mathrm{F}}(\tilde{a}^1_{j,0,\cdot}) &= R_{\mathrm{F}}\left(\sum_{i=1}^{c_{\mathrm{in}}} W^1_{j,i,\cdot} \star R_{\mathrm{F}}\left(q_{i,\cdot}\right) + b^1_j\right) \\ &= \sum_{i=1}^{c_{\mathrm{in}}} R_{\mathrm{F}}\left(W^1_{j,i,\cdot}\right) \star q_{i,\cdot} + b^1_j = a^1_{j,1,\cdot} \end{aligned} \tag{A3}$$

$$\begin{aligned} R_{\mathrm{F}}(\tilde{a}^1_{j,1,\cdot}) &= R_{\mathrm{F}}\left(\sum_{i=1}^{c_{\mathrm{in}}} R_{\mathrm{F}}\left(W^1_{j,i,\cdot}\right) \star R_{\mathrm{F}}\left(q_{i,\cdot}\right) + b^1_j\right) \\ &= \sum_{i=1}^{c_{\mathrm{in}}} W^1_{j,i,\cdot} \star q_{i,\cdot} + b^1_j = a^1_{j,0,\cdot} \end{aligned} \tag{A4}$$

Now, these two equations satisfy the definition of the group equivariance in Eq. (10). Thus, we finish the proof. An example plot for the group equivariance of this layer is shown in Fig. (3).

2. Proof of Equivariance for Non-Input Convolution Layers

The output of subsequent layers, for which inputs and output channels are both real-valued functions on H, is given in eq. 19. A flipped input $R_{\mathrm{F}}(x_{i,\cdot})$ gives the output

$$\tilde{a}^\ell_{j,0,\cdot} = \sum_{i=1}^{c_{\ell-1}} W_{j,i,0,\cdot} \star \tilde{a}^{\ell-1}_{i,0,\cdot} + W_{j,i,1,\cdot} \star \tilde{a}^{\ell-1}_{i,1,\cdot} + b^\ell_j \tag{A5}$$

$$\tilde{a}^\ell_{j,1,\cdot} = \sum_{i=1}^{c_{\ell-1}} R_{\mathrm{F}}(W_{j,i,0,\cdot}) \star \tilde{a}^{\ell-1}_{i,1,\cdot} + R_{\mathrm{F}}(W_{j,i,1,\cdot}) \star \tilde{a}^{\ell-1}_{i,0,\cdot} + b^\ell_j \tag{A6}$$

Flipping outputs gives

$$R_{\mathrm{F}}(\tilde{a}^\ell_{j,0,\cdot}) = R_{\mathrm{F}}\left(\sum_{i=1}^{c_{\ell-1}} W_{j,i,0,\cdot} \star \tilde{a}^{\ell-1}_{i,0,\cdot} + W_{j,i,1,\cdot} \star \tilde{a}^{\ell-1}_{i,1,\cdot} + b^\ell_j\right) \tag{A7}$$

$$R_{\mathrm{F}}(\tilde{a}^\ell_{j,1,\cdot}) = R_{\mathrm{F}}\left(\sum_{i=1}^{c_{\ell-1}} R_{\mathrm{F}}(W_{j,i,0,\cdot}) \star \tilde{a}^{\ell-1}_{i,1,\cdot} + R_{\mathrm{F}}(W_{j,i,1,\cdot}) \star \tilde{a}^{\ell-1}_{i,0,\cdot} + b^\ell_j\right) \tag{A8}$$

Then, we move the flipping operator R_{F} into the conversation features. We not only need to flip the weights but also switch the non-flipped input. Thus, Eqs. (A7-A8) can be written as

$$\begin{aligned} R_{\mathrm{F}}(\tilde{a}^\ell_{j,0,\cdot}) &= \sum_{i=1}^{c_{\ell-1}} R_{\mathrm{F}}\left(W_{j,i,1,\cdot}\right) \star a^{\ell-1}_{i,0,\cdot} + R_{\mathrm{F}}\left(W_{j,i,0,\cdot}\right) \star a^{\ell-1}_{i,1,\cdot} + b^\ell_j \\ &= a^\ell_{j,1,\cdot} \end{aligned} \tag{A9}$$

$$R_{\mathrm{F}}(\tilde{a}^\ell_{j,1,\cdot}) = \sum_{i=1}^{c_{\ell-1}} W_{j,i,0,\cdot} \star a^{\ell-1}_{i,0,\cdot} + W_{j,i,1,\cdot} \star a^{\ell-1}_{i,1,\cdot} + b^\ell_j = a^\ell_{j,0,\cdot} \tag{A10}$$

Therefore, according to the definition of equivariance, we have proven the equivariance convolution in subsequent layers. The example plot is also illustrated in Fig. (3).

3. Proof of Equivariance for Mixed Scalar-Vector Convolution Layers

The first layer's outputs for mixed scalar-vector inputs are shown in Eqs. (22-23). Here, we prove the equivariance in this layers. According to the symmetry of the vector field shown in Eq. (21), we transform the input as $R_{\mathrm{F}}(\zeta_{i,\cdot})$ and $-R_{\mathrm{F}}(u_{i,\cdot})$. Thus, the first layer of output using the flipping input is written as

$$\tilde{a}^1_{j,0,\cdot} = \sum_{i=1}^{c^\zeta_{in}} W^\zeta_{j,i,\cdot} \star R_{\mathrm{F}}(\zeta_{i,\cdot}) + \sum_{i=1}^{c^u_{in}} W^u_{j,i,\cdot} \star -R_{\mathrm{F}}(u_{i,\cdot}) + b^\ell_j \tag{A11}$$

$$\tilde{a}^1_{j,1,\cdot} = \sum_{i=1}^{c^\zeta_{in}} R_{\mathrm{F}}(W^\zeta_{j,i,\cdot}) \star R_{\mathrm{F}}(\zeta_{i,\cdot}) + \sum_{i=1}^{c^u_{in}} -R_{\mathrm{F}}(W^u_{j,i,\cdot}) \star -R_{\mathrm{F}}(u_{i,\cdot}) + b^\ell_j \tag{A12}$$

Next, we flip these outputs

$$R_{\mathrm{F}}(\tilde{a}^1_{j,0,\cdot}) = R_{\mathrm{F}}\left(\sum_{i=1}^{c^\zeta_{in}} W^\zeta_{j,i,\cdot} \star R_{\mathrm{F}}(\zeta_{i,\cdot}) + \sum_{i=1}^{c^u_{in}} -W^u_{j,i,\cdot} \star R_{\mathrm{F}}(u_{i,\cdot}) + b^\ell_j\right) \tag{A13}$$

$$R_{\mathrm{F}}(\tilde{a}^1_{j,1,\cdot}) = R_{\mathrm{F}}\left(\sum_{i=1}^{c^\zeta_{in}} R_{\mathrm{F}}(W^\zeta_{j,i,\cdot}) \star R_{\mathrm{F}}(\zeta_{i,\cdot}) + \sum_{i=1}^{c^u_{in}} R_{\mathrm{F}}(W^u_{j,i,\cdot}) \star R_{\mathrm{F}}(u_{i,\cdot}) + b^\ell_j\right) \tag{A14}$$

Now, we move the flipping operator into the convolution operator. The flipping weight and input feature can be changed as the following equations,

$$R_{\mathrm{F}}(\tilde{a}^1_{j,0,\cdot}) = \sum_{i=1}^{c^{\zeta}_{in}} R_{\mathrm{F}}(W^{\zeta}_{j,i,\cdot}) \star \zeta_{i,\cdot} + \sum_{i=1}^{c^{u}_{in}} -R_{\mathrm{F}}(W^{u}_{j,i,\cdot}) \star u_{i,\cdot} + b^{\ell}_j = a^1_{j,1,\cdot} \tag{A15}$$

$$R_{\mathrm{F}}(\tilde{a}^1_{j,1,\cdot}) = \sum_{i=1}^{c^{\zeta}_{in}} W^{\zeta}_{j,i,\cdot} \star \zeta_{i,\cdot} + \sum_{i=1}^{c^{u}_{in}} W^{u}_{j,i,\cdot} \star u_{i,\cdot} + b^{\ell}_j = a^1_{j,0,\cdot} \tag{A16}$$

Thus, according to the definition of equivariance of convolution, we have proven the equivariance for mixed scalar-vector convolution layers.

Appendix B: Numerical Discretization of SWEs

Here we describe how the space- and time-discretized variable fields u^n and ζ^n of the SWE at the n-th time step are used to compute the $(n+1)$-th time step. We describe the procedures used for the semiiplict non-deep-learning-based classical numerical solver, which is a biased upwind scheme[68].

We discretize the momentum equation (eq. 24) as follows:

$$u^{n+1} = u^n - \Delta t C_D \frac{1}{h} u^n |u^n| - \Delta t g (1 - w_{\mathbf{imp}}) \frac{\partial \zeta^n}{\partial x} - \Delta t g w_{\mathbf{imp}} \frac{\partial \zeta^{n+1}}{\partial x} \tag{B1}$$

where $w_{\mathbf{imp}}$ is a fixed parameter controlling weighting between implicit and explicit time stepping. The mass equation (eq. 25) is discretized as:

$$\zeta^{n+1} = \zeta^n - \Delta t (1 - w_{\mathbf{imp}}) \frac{\partial h^n u^n}{\partial x} - \Delta t w_{\mathbf{imp}} \frac{\partial h^n u^{n+1}}{\partial x} \tag{B2}$$

Recall that $h = d + \zeta$ and d is the undisturbed water depth. Eq. (B1) is inserted into eq. (B2) to obtain

$$\zeta^{n+1} = \zeta^n - \Delta t (1 - w_{\mathbf{imp}}) \frac{\partial h^n u^n}{\partial x} - \Delta t w_{\mathbf{imp}} \frac{\partial h^n u^*}{\partial x} + \Delta t^2 w^2_{\mathbf{imp}} g \frac{\partial^2 h^n \zeta^{n+1}}{\partial x^2} \tag{B3}$$

where u^* is an 'interim solution' defined by

$$u^* = u^n - \Delta t c_D \frac{1}{h} u^n |u^n| - \Delta t g (1 - w_{\mathbf{imp}}) \frac{\partial \zeta^n}{\partial x} \tag{B4}$$

When calculating the product of two variables defined on the velocity and mass points of the Arakawa C-grid (Fig. 1b), we interpolate the mass variable to velocity grid points by averaging adjacent values. For a quantity α defined on the velocity or mass gird, the first spatial derivative is discretized as $\frac{\partial \alpha}{\partial x}_i = \frac{\alpha_{i+1/2} - \alpha_{i-1/2}}{\Delta x}$, with outputs staggered by $\Delta x/2$ from inputs. The second spatial derivative is discretized using the second order finite difference $\frac{\partial^2 \alpha}{\partial x^2}_i = \frac{\alpha_{i+1} - 2\alpha_i + \alpha_{i-1}}{\Delta x^2}$, with outputs on the same grid as inputs. Therefore, eq. (B3) can be written as

$$\zeta^{n+1}_i = \frac{1}{1 + c_E + c_W} \left[\zeta^n + \mathrm{div} + c_E \zeta^{n+1}_{i+1} + c_W \zeta^{n+1}_{i-1} \right] \tag{B5}$$

where $\mathrm{div} = -\Delta t (1 - w_{\mathbf{imp}}) \frac{\partial h^n u^n}{\partial x} - \Delta t w_{\mathbf{imp}} \frac{\partial h^n u^*}{\partial x}$, while c_E and c_W are defined as

$$c_E = \begin{cases} \frac{0.5 \Delta t^2 w^2_{\mathbf{imp}} g (h^n_i + h^n_{i+1})}{\Delta x^2} & if\ h^n_{i+1} > 0 \\ 0 \text{ otherwise.} \end{cases}$$

$$c_W = \begin{cases} \frac{0.5 \Delta t^2 w^2_{\mathbf{imp}} g (h^n_i + h^n_{i-1})}{\Delta x^2} & if\ h^n_{i-1} > 0 \\ 0 \text{ otherwise.} \end{cases}$$

Eq. (B5) describes a linear system of equations in ζ^{n+1} that can be written in matrix-vector form

$$A \zeta^{n+1} = b \tag{B6}$$

where A is a $N \times N$ tridiagonal matrix ($N = L/\Delta x$) with $A_{k,k} = 1$, $A_{k,k-1} = -\frac{c_W}{1 + c_E + c_W}$, $A_{k,k+1} = -\frac{c_E}{1 + c_E + c_W}$ and all other elements zero. $b \in \mathbb{R}^N$ with $b = \frac{\zeta^n + div}{1 + c_E + c_W}$. Following[68], we employ Gauss-Seidel iterations to solve eq. B6. Having obtained ζ^{n+1}, the new velocity u^{n+1} is calculated as

$$u^{n+1} = u^* - \Delta t g w_{\mathbf{imp}} \frac{\partial \zeta^{n+1}}{\partial x} \tag{B7}$$

[1] H. Egger and J. Schöberl, "A hybrid mixed discontinuous galerkin finite-element method for convection–diffusion problems," SIAM J. Numer. Anal. **30**, 1206–1234 (2010).

[2] L. Euler, "Principes généraux du mouvement des fluides," Mémoires de l'académie des sciences de Berlin , 274–315 (1757).

[3] R. Temam, *Navier-Stokes equations: theory and numerical analysis*, Vol. 343 (American Mathematical Soc., 2001).

[4] P. Constantin and C. Foias, *Navier-stokes equations* (University of Chicago Press, 2020).

[5] B. De St Venant, "Theorie du mouvement non-permanent des eaux avec application aux crues des rivers et a l'introduntion des marees dans leur lit," Academic de Sci. Comptes Redus **73**, 148–154 (1871).

[6] P. J. Dellar and R. Salmon, "Shallow water equations with a complete coriolis force and topography," Phys. Fluids **17**, 106601 (2005).

[7] S. Bunya, E. J. Kubatko, J. J. Westerink, and C. Dawson, "A wetting and drying treatment for the runge–kutta discontinuous galerkin solution to the shallow water equations," Comput. Methods Appl. Mech. Eng. **198**, 1548–1562 (2009).

[8] T. Kärnä, B. De Brye, O. Gourgue, J. Lambrechts, R. Comblen, V. Legat, and E. Deleersnijder, "A fully implicit wetting–drying method for dg-fem shallow water models, with an application to the scheldt estuary," Comput. Methods Appl. Mech. Eng. **200**, 509–524 (2011).

[9] G. Zängl, D. Reinert, P. Rípodas, and M. Baldauf, "The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core," Quarterly Journal of the Royal Meteorological Society **141**, 563–579 (2015).

[10] M. Klöwer, P. Düben, and T. Palmer, "Number formats, error mitigation, and scope for 16-bit arithmetics in weather and climate modeling analyzed with a shallow water model," J. Adv. Model. Earth Syst. **12**, e2020MS002246 (2020).

[11] P. Korn, N. Brüggemann, J. H. Jungclaus, S. Lorenz, O. Gutjahr, H. Haak, L. Linardakis, C. Mehlmann, U. Mikolajewicz, D. Notz, *et al.*, "Icon-o: The ocean component of the icon earth system model—global simulation characteristics and local telescoping capability," J. Adv. Model. Earth Syst. **14**, e2021MS002952 (2022).

[12] R. Sadourny, "The dynamics of finite-difference models of the shallow-water equations," J. Atmos. Sci. **32**, 680–689 (1975).

[13] V. Casulli, "Semi-implicit finite difference methods for the two-dimensional shallow water equations," J. Comput. Phys. **86**, 56–74 (1990).

[14] O. Zienkiewicz and P. Ortiz, "A split-characteristic based finite element model for the shallow water equations," Int. J. Numer. Methods Fluids. **20**, 1061–1080 (1995).

[15] F. Bassi and S. Rebay, "A high-order accurate discontinuous finite element method for the numerical solution of the compressible navier–stokes equations," J. Comput. Phys. **131**, 267–279 (1997).

[16] R. Eymard, T. Gallouët, and R. Herbin, "Finite volume methods," Handbook of numerical analysis **7**, 713–1018 (2000).

[17] S. Grilli, T. Pedersen, and P. Stepanishen, "A hybrid boundary element method for shallow water acoustic propagation over an irregular bottom," Eng. Anal. Bound. Elem. **21**, 131–145 (1998).

[18] M. Taylor, J. Tribbia, and M. Iskandarani, "The spectral element method for the shallow water equations on the sphere," J. Comput. Phys. **130**, 92–108 (1997).

[19] O. Reynolds, "Iv. on the dynamical theory of incompressible viscous fluids and the determination of the criterion," Philos. Trans. Royal Soc. A , 123–164 (1895).

[20] J. Smagorinsky, "General circulation experiments with the primitive equations: I. the basic experiment," Mon. Weather Rev. **91**, 99–164 (1963).

[21] J. W. Deardorff, "A numerical study of three-dimensional turbulent channel flow at large reynolds numbers," J. Fluid Mech. **41**, 453–480 (1970).

[22] Y. Saad, *Iterative methods for sparse linear systems* (SIAM, 2003).

[23] B. T. Polyak, "The conjugate gradient method in extremal problems," USSR Comput. Math. & Math. Phys. **9**, 94–112 (1969).

[24] Y. Saad and M. H. Schultz, "Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems," SIAM J. Sci. Comput. **7**, 856–869 (1986).

[25] H. A. Van der Vorst, "Bi-cgstab: A fast and smoothly converging variant of bi-cg for the solution of nonsymmetric linear systems," SIAM J. Sci. Comput. **13**, 631–644 (1992).

[26] B. Kim, V. C. Azevedo, N. Thuerey, T. Kim, M. Gross, and B. Solenthaler, "Deep fluids: A generative network for parameterized fluid simulations," in *Computer graphics forum*, Vol. 38 (Wiley Online Library, 2019) pp. 59–70.

[27] R. Wang, K. Kashinath, M. Mustafa, A. Albert, and R. Yu, "Towards physics-informed deep learning for turbulent flow prediction," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2020) pp. 1457–1466.

[28] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," arXiv preprint arXiv:2010.08895 (2020).

[29] G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, and S. M. Benson, "U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow," Adv. Water Resour. **163**, 104180 (2022).

[30] G. Gupta, X. Xiao, and P. Bogdan, "Multiwavelet-based operator learning for differential equations," Adv. Neural Inf. Process Syst. **34**, 24048–24062 (2021).

[31] Y. Li, L. Xu, and S. Ying, "Dwnn: Deep wavelet neural network for solving partial differential equations," Mathematics **10**, 1976 (2022).

[32] G. Kohl, L.-W. Chen, and N. Thuerey, "Turbulent Flow Simulation using Autoregressive Conditional Diffusion Models," (2023), arXiv:2309.01745 [physics].

[33] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Neural operator: Graph kernel network for partial differential equations," arXiv preprint arXiv:2003.03485 (2020).

[34] N. Wandel, M. Weinmann, and R. Klein, "Learning incompressible fluid dynamics from scratch–towards fast, differentiable fluid models that generalize," arXiv preprint arXiv:2006.08762 (2020).

[35] N. Wandel, M. Weinmann, and R. Klein, "Teaching the incompressible navier–stokes equations to fast neural surrogate models in three dimensions," Phys. Fluids **33**, 047117 (2021).

[36] M. Raissi, P. Perdikaris, and G. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys. **378**, 686–707 (2019).

[37] L. Bar and N. Sochen, "Unsupervised deep learning algorithm for pde-based forward and inverse problems," arXiv preprint arXiv:1904.05417 (2019).

[38] Z. Cai, J. Chen, M. Liu, and X. Liu, "Deep least-squares methods: An unsupervised learning-based numerical method for solving elliptic pdes," J. Comput. Phys. **420**, 109707 (2020).

[39] A. Stanziola, S. R. Arridge, B. T. Cox, and B. E. Treeby, "A helmholtz equation solver using unsupervised learning: Application to transcranial ultrasound," J. Comput. Phys. **441**, 110430 (2021).

[40] M. Y. Michelis and R. K. Katzschmann, "Physics-constrained unsupervised learning of partial differential equations using meshes," arXiv preprint arXiv:2203.16628 (2022).

[41] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris, "Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data," J. Comput. Phys. **394**, 56 – 81 (2019).

[42] L. Sun, H. Gao, S. Pan, and J.-X. Wang, "Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data," Comput. Methods Appl. Mech. Eng. **361**, 112732 (2020).

[43] N. Geneva and N. Zabaras, "Modeling the dynamics of pde systems with physics-constrained deep auto-regressive networks," J. Comput. Phys. **403**, 109056 (2020).

[44] R. Grzeszczuk, D. Terzopoulos, and G. Hinton, "NeuroAnimator: fast neural network emulation and control of physics-based models," in *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '98 (Association for Computing Machinery, New York, NY, USA, 1998) pp. 9–20.

[45] L. Ladickỳ, S. Jeong, B. Solenthaler, M. Pollefeys, and M. Gross, "Data-driven fluid simulations using regression forests," ACM Trans. Graph. **34**, 1–9 (2015).

[46] K. Um, R. Brand, Y. R. Fei, P. Holl, and N. Thuerey, "Solver-in-the-loop: Learning from differentiable physics to interact with iterative pde-solvers," Adv. Neural Inf. Process. Syst. **33**, 6111–6122 (2020).

[47] S. Wiewel, M. Becher, and N. Thuerey, "Latent space physics: Towards learning the temporal evolution of fluid flow," in *Computer graphics forum*, Vol. 38 (Wiley Online Library, 2019) pp. 71–82.

[48] J. Tompson, K. Schlachter, P. Sprechmann, and K. Perlin, "Accelerating eulerian fluid simulation with convolutional networks," in *International Conference on Machine Learning* (PMLR, 2017) pp. 3424–3433.

[49] O. Obiols-Sales, A. Vishnu, N. Malaya, and A. Chandramowliswharan, "Cfdnet: A deep learning-based accelerator for fluid simulations," in *Proceedings of the 34th ACM international conference on supercomputing* (2020) pp. 1–12.

[50] M. Nonnenmacher and D. S. Greenberg, "Learning Implicit PDE Integration with Linear Implicit Layers," in *The Symbiosis of Deep Learning and Differential Equations* (2021).

[51] B. List, L.-W. Chen, and N. Thuerey, "Learned turbulence modelling with differentiable fluid solvers," arXiv preprint arXiv:2202.06988 (2022).

[52] T. Wu, T. Maruyama, and J. Leskovec, "Learning to accelerate partial differential equations via latent global evolution," arXiv preprint arXiv:2206.07681 (2022).

[53] N. Thuerey, K. Weißenow, L. Prantl, and X. Hu, "Deep learning methods for reynolds-averaged navier–stokes simulations of airfoil flows," AIAA Journal **58**, 25–36 (2020).

[54] K. Kashinath, M. Mustafa, A. Albert, J. Wu, C. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, R. Wang, A. Chattopadhyay, A. Singh, *et al.*, "Physics-informed machine learning: case studies for weather and climate modelling," Philos. Trans. R. Soc. A **379**, 20200093 (2021).

[55] K. O. Lye, S. Mishra, and D. Ray, "Deep learning observables in computational fluid dynamics," J. Comput. Phys. **410**, 109339 (2020).

[56] S. Fresca and A. Manzoni, "Pod-dl-rom: Enhancing deep learning-based reduced order models for nonlinear parametrized pdes by proper orthogonal decomposition," Comput. Methods Appl. Mech. Eng. **388**, 114181 (2022).

[57] J. Yuval, P. A. O'Gorman, and C. N. Hill, "Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision," Geophys. Res. Lett. **48**, e2020GL091363 (2021).

[58] J. Yuval and P. A. O'Gorman, "Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions," Nat. Commun. **11**, 3295 (2020).

[59] T. Cohen and M. Welling, "Group equivariant convolutional networks," in *International conference on machine learning* (PMLR, 2016) pp. 2990–2999.

[60] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, "Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation," Med. Image Anal. **63**, 101693 (2020).

[61] J. Han, J. Ding, N. Xue, and G.-S. Xia, "Redet: A rotation-equivariant detector for aerial object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021) pp. 2786–2795.

[62] J. E. Gerken, J. Aronsson, O. Carlsson, H. Linander, F. Ohlsson, C. Petersson, and D. Persson, "Geometric deep learning and equivariant neural networks," Artif. Intell. Rev. , 1–58 (2023).

[63] A. Akio and R. L. Vivian, "Computational design of the basic dynamical processes of the ucla general circulation model," in *General Circulation Models of the Atmosphere*, Methods in Computational Physics: Advances in Research and Applications, Vol. 17, edited by J. Chang (Elsevier, 1977) pp. 173–265.

[64] J. Crank and P. Nicolson, "A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type," in *Mathematical proceedings of the Cambridge philosophical society*, Vol. 43 (Cambridge University Press, 1947) pp. 50–67.

[65] S. Turek, "A comparative study of time-stepping techniques for the incompressible navier-stokes equations: from fully implicit non-linear schemes to semi-implicit projection methods," Int. J. Numer. Methods Fluids. **22**, 987–1011 (1996).

[66] P. F. Fischer, F. Hecht, and Y. Maday, "A parareal in time semi-implicit approximation of the navier-stokes equations," Lect. Notes Comput. Sci. Eng. **40**, 433–440 (2005).

[67] D. Forti and L. Dedè, "Semi-implicit bdf time discretization of the navier-stokes equations with vms-les modeling in a high performance computing framework," Comput. Fluids **117**, 168–182 (2015).

[68] J. O. Backhaus, "A semi-implicit scheme for the shallow water equations for application to shelf sea modelling," Cont. Shelf Res. **2**, 243–254 (1983).

[69] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980 (2014).

[70] Z. Long, Y. Lu, and B. Dong, "Pde-net 2.0: Learning pdes from data with a numeric-symbolic hybrid deep network," J. Comput. Phys. **399**, 108925 (2019).

[71] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, "Machine learning–accelerated computational fluid dynamics," Proc. Natl. Acad. Sci. U. S. A. **118**, e2101784118 (2021).

[72] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Springer, 2015) pp. 234–241.

[73] V. Casulli and R. A. Walters, "An unstructured grid, three-dimensional model based on the shallow water equations," Int. J. Numer. Methods Fluids **32**, 331–348 (2000).

[74] M. Zijlema and G. Stelling, "Efficient computation of surf zone waves using the nonlinear shallow water equations with non-hydrostatic pressure," Coast. Eng. **55**, 780–790 (2008).

[75] Y. Huang, G. Gompper, and B. Sabass, "A bayesian traction force microscopy method with automated denoising in a user-friendly software package," Comput. Phys. Commun. **256**, 107313 (2020).

[76] A. T. Mohan, N. Lubbers, M. Chertkov, and D. Livescu, "Embedding hard physical constraints in neural network coarse-graining of three-dimensional turbulence," Phys. Rev. Fluids **8**, 014604 (2023).

[77] J. Brandstetter, M. Welling, and D. E. Worrall, "Lie point symmetry data augmentation for neural pde solvers," in *International Conference on Machine Learning* (PMLR, 2022) pp. 2241–2256.

[78] R. Wang, R. Walters, and R. Yu, "Incorporating symmetry into deep dynamics models for improved generalization," arXiv preprint arXiv:2002.03061 (2020).

[79] T. S. Cohen and M. Welling, "Steerable cnns," arXiv preprint arXiv:1612.08498 (2016).

[80] B. M. Smets, J. Portegies, E. J. Bekkers, and R. Duits, "Pde-based group equivariant convolutional neural networks," J. Math. Imaging Vis. , 1–31 (2022).

[81] D. Marcos, M. Volpi, N. Komodakis, and D. Tuia, "Rotation equivariant vector field networks," in *Proceedings of the IEEE International Conference on Computer Vision* (2017) pp. 5048–5057.

[82] J. Helwig, X. Zhang, C. Fu, J. Kurtin, S. Wojtowytsch, and S. Ji, "Group equivariant fourier neural operators for partial differential equations," in *Proceedings of the 40th International Conference on Machine Learning*, ICML'23, Vol. 202 (JMLR.org, Honolulu, Hawaii, USA, 2023) pp. 12907–12930.

[83] Y. Guan, A. Subel, A. Chattopadhyay, and P. Hassanzadeh, "Learning physics-constrained subgrid-scale closures in the small-data regime for stable and accurate les," Phys. D: Nonlinear Phenom. , 133568 (2022).

[84] A. Ross, Z. Li, P. Perezhogin, C. Fernandez-Granda, and L. Zanna, "Benchmarking of machine learning ocean subgrid parameterizations in an idealized model," J. Adv. Model. Earth Syst. **15**, e2022MS003258 (2023).

[85] M. Nonnenmacher and D. S. Greenberg, "Deep emulators for differentiation, forecasting, and parametrization in earth science simulators," J. Adv. Model. Earth Syst. **13**, e2021MS002554 (2021).

[86] B. J. Holzschuh, S. Vegetti, and N. Thuerey, "Score matching via differentiable physics," (2023), arXiv:2301.10250 [cs.LG].

[87] K. Hu, C. G. Mingham, and D. M. Causon, "Numerical simulation of wave overtopping of coastal structures using the non-linear shallow water equations," Coast. Eng. **41**, 433–465 (2000).

[88] H. Salman, L. Kuznetsov, C. Jones, and K. Ide, "A method for assimilating lagrangian data into a shallow-water-equation ocean model," Mon. Weather Rev. **134**, 1081–1101 (2006).

[89] M. Heniche, Y. Secretan, P. Boudreau, and M. Leclerc, "A two-dimensional finite element drying-wetting shallow water model for rivers and estuaries," Adv. Water Resour. **23**, 359–372 (2000).

[90] N. Gouta and F. Maurel, "A finite volume solver for 1d shallow-water equations applied to an actual river," Int. J. Numer. Methods Fluids. **38**, 1–19 (2002).

[91] E. Hanert, D. Y. Le Roux, V. Legat, and E. Deleersnijder, "An efficient eulerian finite element method for the shallow water equations," Ocean Model. **10**, 115–136 (2005).

[92] B. CUSHMAN-ROISIN, "Exact analytical solutions for elliptical vortices of the shallow-water equations," Tellus A **39**, 235–244 (1987).

[93] M. Läuter, D. Handorf, and K. Dethloff, "Unsteady analytical solutions of the spherical shallow water equations," J. Comput. Phys. **210**, 535–553 (2005).

[94] Y. Huang, C. Schell, T. B. Huber, A. N. Şimşek, N. Hersch, R. Merkel, G. Gompper, and B. Sabass, "Traction force microscopy with optimized regularization and automated bayesian parameter selection for comparing cells," Sci. Rep. **9**, 1–16 (2019).

[95] J. Nordström and M. H. Carpenter, "Boundary and interface conditions for high-order finite-difference methods applied to the euler and navier–stokes equations," J. Comput. Phys. **148**, 621–645 (1999).

[96] P. L. Roe, "Characteristic-based schemes for the euler equations," Annu. Rev. Fluid Mech. **18**, 337–365 (1986).

[97] P. Constantin, "On the euler equations of incompressible fluids," Society. Bull. Am. Math. **44**, 603–621 (2007).

[98] K. W. Morton, *Numerical solution of convection-diffusion problems* (CRC Press, 2019).

[99] N. Thuerey, P. Holl, M. Mueller, P. Schnell, F. Trost, and K. Um, *Physics-based Deep Learning* (WWW, 2021).

[100] Z. Janjic, "Finite difference methods for the shallow water equations on various horizontal grids," in *Seminar, ECMWF, 1983* (1984) pp. 29–101.

[101] S. Wiewel, B. Kim, V. C. Azevedo, B. Solenthaler, and N. Thuerey, "Latent space subdivision: stable and controllable time predictions for fluid flow," in *Computer Graphics Forum*, Vol. 39 (Wiley Online Library, 2020) pp. 15–25.

[102] J. Yuval, P. A. O'Gorman, and C. N. Hill, "Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision," Geophys. Res. Lett. **48**, e2020GL091363 (2021).

[103] Y. Kaneda, T. Ishihara, M. Yokokawa, K. Itakura, and A. Uno, "Energy dissipation rate and energy spectrum in high resolution direct numerical simulations of turbulence in a periodic box," Phys. Fluids **15**, L21–L24 (2003).

[104] C. J. Greenshields and H. G. Weller, *Note on Computational Fluid Dynamics: General Principles* (CFD Direct https://cfd. direct., 2022).

[105] K. Atz, F. Grisoni, and G. Schneider, "Geometric deep learning on molec-

ular representations," Nat. Mach. Intell **3**, 1023–1032 (2021).
[106]P. J. Olver, *Applications of Lie groups to differential equations*, Vol. 107 (Springer Science & Business Media, 1993).
[107]G. W. Bluman and J. D. Cole, "The general similarity solution of the heat equation," J. math. mech. **18**, 1025–1042 (1969).
[108]J. Perez, M. Menendez, P. Camus, F. J. Mendez, and I. J. Losada, "Statistical multi-model climate projections of surface ocean waves in europe," Ocean Modelling **96**, 161–170 (2015).
[109]A. Fournier and W. T. Reeves, "A simple model of ocean waves," in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques* (1986) pp. 75–84.
[110]V. A. Squire, "Of ocean waves and sea-ice revisited," Cold Reg. Sci. Technol. **49**, 110–133 (2007).
[111]L. Yang, X. Meng, and G. E. Karniadakis, "B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data," J. Comput. Phys. **425**, 109913 (2021).
[112]R. Khodayi-Mehr and M. Zavlanos, "Varnet: Variational neural networks for the solution of partial differential equations," in *Learning for Dynamics and Control* (PMLR, 2020) pp. 298–307.
[113]X. Zhang and K. Garikipati, "Bayesian neural networks for weak solution of pdes with uncertainty quantification," arXiv preprint arXiv:2101.04879 (2021).
[114]M. Eliasof, E. Haber, and E. Treister, "Pde-gcn: novel architectures for graph neural networks motivated by partial differential equations," Adv. Neural Inf. Process Syst. **34**, 3836–3849 (2021).
[115]V. G. Satorras, E. Hoogeboom, and M. Welling, "E (n) equivariant graph neural networks," in *International conference on machine learning* (PMLR, 2021) pp. 9323–9332.
[116]Y. Huang, Y. Mabrouk, G. Gompper, and B. Sabass, "Sparse inference and active learning of stochastic differential equations from data," Sci. Rep. **12**, 21691 (2022).
[117]J. Magiera, D. Ray, J. S. Hesthaven, and C. Rohde, "Constraint-aware neural networks for riemann problems," J. Comput. Phys. **409**, 109345 (2020).
[118]S. Qi, X. Ning, G. Yang, L. Zhang, P. Long, W. Cai, and W. Li, "Review of multi-view 3d object recognition methods based on deep learning," Displays **69**, 102053 (2021).
[119]P. Rípodas, A. Gassmann, J. Förstner, D. Majewski, M. Giorgetta, P. Korn, L. Kornblueh, H. Wan, G. Zängl, L. Bonaventura, *et al.*, "Icosahedral shallow water model (icoswm): results of shallow water test cases and sensitivity to model parameters," Geosci. Model Dev. **2**, 231–251 (2009).
[120]H. Weller, H. G. Weller, and A. Fournier, "Voronoi, delaunay, and block-structured mesh refinement for solution of the shallow-water equations on the sphere," Mon. Weather Rev. **137**, 4208–4224 (2009).