TOPICAL REVIEW • OPEN ACCESS

Bridging observations, theory and numerical simulation of the ocean using machine learning

View the article online for updates and enhancements.

ENVIRONMENTAL RESEARCH
LETTERS

TOPICAL REVIEW

Bridging observations, theory and numerical simulation of the ocean using machine learning

Maike Sonnewald[1,2,3,9,*] (ID), Redouane Lguensat[4,5] (ID), Daniel C Jones[6] (ID), Peter D Dueben[7] (ID), Julien Brajard[5,8] (ID) and V Balaji[1,2,4] (ID)

1   Princeton University, Program in Atmospheric and Oceanic Sciences, Princeton, NJ 08540, United States of America
2   NOAA/OAR Geophysical Fluid Dynamics Laboratory, Ocean and Cryosphere Division, Princeton, NJ 08540, United States of America
3   University of Washington, School of Oceanography, Seattle, WA, United States of America
4   Laboratoire des Sciences du Climat et de l'Environnement (LSCE-IPSL), CEA Saclay, Gif Sur Yvette, France
5   LOCEAN-IPSL, Sorbonne Université, Paris, France
6   British Antarctic Survey, NERC, UKRI, Cambridge, United Kingdom
7   European Centre for Medium Range Weather Forecasts, Reading, United Kingdom
8   Nansen Center (NERSC), Bergen, Norway
9   Present address: Princeton University, Program in Atmospheric and Oceanic Sciences, 300 Forrestal Rd., Princeton, NJ 08540, United States of America
*   Author to whom any correspondence should be addressed.

E-mail: maikes@princeton.edu

Abstract

Progress within physical oceanography has been concurrent with the increasing sophistication of tools available for its study. The incorporation of machine learning (ML) techniques offers exciting possibilities for advancing the capacity and speed of established methods and for making substantial and serendipitous discoveries. Beyond vast amounts of complex data ubiquitous in many modern scientific fields, the study of the ocean poses a combination of unique challenges that ML can help address. The observational data available is largely spatially sparse, limited to the surface, and with few time series spanning more than a handful of decades. Important timescales span seconds to millennia, with strong scale interactions and numerical modelling efforts complicated by details such as coastlines. This review covers the current scientific insight offered by applying ML and points to where there is imminent potential. We cover the main three branches of the field: observations, theory, and numerical modelling. Highlighting both challenges and opportunities, we discuss both the historical context and salient ML tools. We focus on the use of ML *in situ* sampling and satellite observations, and the extent to which ML applications can advance theoretical oceanographic exploration, as well as aid numerical simulations. Applications that are also covered include model error and bias correction and current and potential use within data assimilation. While not without risk, there is great interest in the potential benefits of oceanographic ML applications; this review caters to this interest within the research community.

1. Introduction

1.1. Oceanography: observations, theory, and numerical simulation

The physics of the oceans have been of crucial importance, curiosity and interest since prehistoric times, and today remain an essential element in our understanding of weather and climate, and a key driver of biogeochemistry and overall marine resources. The eras of progress within oceanography have gone hand in hand with the tools available for its study. Here, the current progress and potential future role of machine learning (ML) techniques is reviewed and briefly put into historical context. ML adoption is not without risk, but is here put forward as having the potential to accelerate scientific insight, performing tasks better and faster, along with allowing avenues of serendipitous discovery. This review focuses on physical oceanography, but concepts discussed are applicable across oceanography and beyond.

Perhaps the principal interest in oceanography was originally that of navigation, for exploration, commercial and military purposes. Knowledge of the ocean as a dynamical entity with predictable features—the regularity of its currents and tides—must have been known for millennia. Knowledge of oceanography likely helped the successful colonization of Oceania [180], and similarly Viking and Inuit navigation [119], the oldest known dock was constructed in Lothal with knowledge of the tides dating back to 2500–1500 BCE [50], and Abu Ma'shar of Baghdad in the 8th century CE correctly attributed the existence of tides to the Moon's pull.

The ocean measurement era, determining temperature and salinity at depth from ships, starts in the late 18th century CE. While the tools for a theory of the ocean circulations started to become available in the early 19th century CE with the Navier–Stokes equation, observations remained at the core of oceanographic discovery. The first modern oceanographic textbook was published in 1855 by M. Mauri, whose work in oceanography and politics served the slave trade across the Atlantic, around the same time CO_2's role in climate was recognized [96, 248]. The first major global observational synthesis of the ocean can be traced to the Challenger expeditions of 1873–75 CE [69], where observational data from various areas was brought together to gain insight into the global ocean. The observational synthesis from the Challenger expeditions gave a first look at the global distribution of temperature and salinity including at depth, revealing the 3-dimensional structure of the ocean.

Quantifying the time mean ocean circulation remains challenging, as ocean circulation features strong local and instantaneous fluctuations. Improvements in measurement techniques allowed the Swedish oceanographer Ekman to elucidate the nature of the wind-driven boundary layer [87]. Ekman used observations taken on an expedition led by the Norwegian oceanographer and explorer Nansen, where the Fram was intentionally frozen into the Arctic ice. The 'dynamic method' was introduced by Swedish oceanographer Sandström and the Norwegian oceanographer Helland-Hansen [217], allowing the indirect computation of ocean currents from density estimates under the assumption of a largely laminar flow. This theory was developed further by Norwegian meteorologist Bjerknes into the concept of *geostrophy*, from the Greek geo for earth and strophe for turning. This theory was put to the test in the extensive Meteor expedition in the Atlantic from 1925 to 1927 CE; they uncovered a view of the horizontal and vertical ocean structure and circulation that is strikingly similar to our present view of the Atlantic meridional overturning circulation [177, 210].

While the origins of geophysical fluid dynamics (GFD) can be traced back to Laplace or Archimedes, the era of modern GFD can be seen to stem from linearizing the Navier–Stokes equations, which enabled progress in understanding meteorology and atmospheric circulation. For the ocean, pioneering dynamicists include Sverdrup, Stommel, and Munk, whose theoretical work still has relevance today [182, 232]. As compared to the atmosphere, the ocean circulation exhibits variability over a much larger range of timescales, as noted by [183], likely spanning thousands of years rather than the few decades of detailed ocean observations available at the time. Yet, there are phenomena at intermediate timescales (that is, months to years) which seemed to involve both atmosphere and ocean, e.g. [186], and indeed Sverdrup suggests the importance of the coupled atmosphere-ocean system in [234]. In the 1940s much progress within GFD was also driven by the second world war (WWII). The introduction of accurate navigation through radar introduced with WWII worked a revolution for observational oceanography together with bathythermographs intensively used for submarine detection. Beyond *in situ* observations, the launch of Sputnik, the first artificial satellite, in 1957 heralded the era of ocean observations from satellites. Seasat, launched on the 27th of June 1978, was the first satellite dedicated to ocean observation.

Oceanography remains a subject that must be understood with an appreciation of available tools, both observational and theoretical, but also numerical. While numerical GFD can be traced back to the early 1900s [2, 31, 209], it became practical with the advent of numerical computing in the late 1940s, complementing that of the elegant deduction and more heuristic methods that one could call 'pattern recognition' that had prevailed before [11]. The first ocean general circulation model with specified global geometry were developed by Bryan and Cox [44, 45] using finite-difference methods. This work paved the way for what now is a major component of contemporary oceanography. The first coupled ocean-atmosphere model of [167] eventually led to their use for studies of the coupled Earth system, including its changing climate. The low-power integrated circuit that gave rise to computers in the 1970s also revolutionized observational oceanography, enabling instruments to reliably record autonomously. This has enabled instruments such as moored current meters and profilers, drifters, and floats through to hydrographic and velocity profiling devices that gave rise to microstructure measurements. Of note is the fleet of free-drifting Argo floats, beginning in 2002, which give an extraordinary global dataset of profiles [212]. Data assimilation (DA) is the important branch of modern oceanography combining what is often sparse observational data with either numerical or statistical ocean models to produce observationally-constrained estimates with no gaps. Such estimates are referred to as an 'ocean state', which is especially important for

understanding locations and times with no available observations.

Together the innovations within observations, theory, and numerical models have produced distinctly different pictures of the ocean as a dynamical system, revealing it as an intrinsically turbulent and topographically influenced circulation [101, 266]. Key large scale features of the circulation depend on very small scale phenomena, which for a typical model resolution remain parameterized rather than explicitly calculated. For instance, fully accounting for the subtropical wind-driven gyre circulation and associated western boundary currents relies on an understanding of the vertical transport of vorticity input by the wind and output at the sea floor, which is intimately linked to mesoscale (ca. 100 km) flow interactions with topography [85, 133]. It has become apparent that localized small-scale turbulence (0–100 km) can also impact the larger-scale, time-mean overturning and lateral circulation by affecting how the upper ocean interacts with the atmosphere [95, 124, 242]. The prominent role of the small scales on the large scale circulation has important implications for understanding the ocean in a climate context, and its representation still hinges on the further development of our fundamental understanding, observational capacity, and advances in numerical approaches.

The development of both modern oceanography and ML techniques have happened concurrently, as illustrated in figure 1. This review summarizes the current state of the art in ML applications for physical oceanography and points towards exciting future avenues. We wish to highlight certain areas where the emerging techniques emanating from the domain of ML demonstrate potential to be transformative. ML methods are also being used in closely-related fields such as atmospheric science. However, within oceanography one is faced with a unique set of challenges rooted in the lack of long-term and spatially dense data coverage. While in recent years the surface of the ocean is becoming well observed, there is still a considerable problem due to sparse data, particularly in the deep ocean. Temporally, the ocean operates on timescales from seconds to millennia, and very few long term time series exist. There is also considerable scale-interaction, which also necessitates more comprehensive observations.

There remains a healthy scepticism towards some ML applications, and calls for 'trustworthy' ML are also coming forth from both the European Union and the United States government (Assessment List for Trustworthy Artificial Intelligence [ALTAI], and mandate E.O. 13 960 of 3 December 2020). Within the physical sciences and beyond, trust can be fostered through transparency. For ML, this means moving beyond the 'black box' approach for certain applications. Moving away from this black box approach and adopting a more transparent approach involves

gaining insight into the learned mechanisms that gave rise to ML predictive skill. This is facilitated by either building *a priori interpretable* ML applications or by retrospectively *explaining* the source of predictive skill, coined interpretable and explainable artificial intelligence (IAI and XAI, respectively [26, 134, 214, 228]). An example of interpretability could be looking for coherent structures (or 'clusters') within a closed budget where all terms are accounted for. Explainability comes from, for example, tracing the weights within a Neural Network (NN) to determine what input features gave rise to its prediction. With such insights from transparent ML, a synthesis between theoretical and observational branches of oceanography could be possible. Traditionally, theoretical models tend towards oversimplification, while data can be overwhelmingly complicated. For advancement in the fundamental understanding of ocean physics, ML is ideally placed to identify salient features in the data that are comprehensible to the human brain. With this approach, ML could significantly facilitate a generalization beyond the limits of data, letting data reveal possible structural errors in theory. With such insight, a hierarchy of conceptual models of ocean structure and circulation could be developed, signifying an important advance in our understanding of the ocean.

In this review, we introduce ML concepts (section 1.2), and some of its current roles in the atmospheric and Earth System Sciences (section 1.3), highlighting particular areas of note for ocean applications. The review follows the structure outline illustrated in figure 2, with the ample overlap noted through cross referencing the text. We review ocean observations (section 2), sparsely observed for much history, but now yielding increasingly clear insight into the ocean and its 3D structure. In section 3 we examine a potential synergy between ML and theory, with the intent to distil expressions of theoretical understanding by dataset analysis from both numerical and observational efforts. We then progress from theory to models, and the encoding of theory and observations in numerical models (section 4). We highlight some issues involved with ML-based prediction efforts (section 5), and end with a discussion of challenges and opportunities for ML in the ocean sciences (section 6). These challenges and opportunities include the need for transparent ML, ways to support decision makers and a general outlook. Appendix has a list of acronyms.

1.2. Concepts in ML

Throughout this article, we will mention some concepts from the ML literature. We find it then natural to start this paper with a brief introduction to some of the main ideas that shaped the field of ML.

ML, a sub-domain of artificial intelligence (AI), is the science of providing mathematical algorithms and computational tools to machines, allowing them
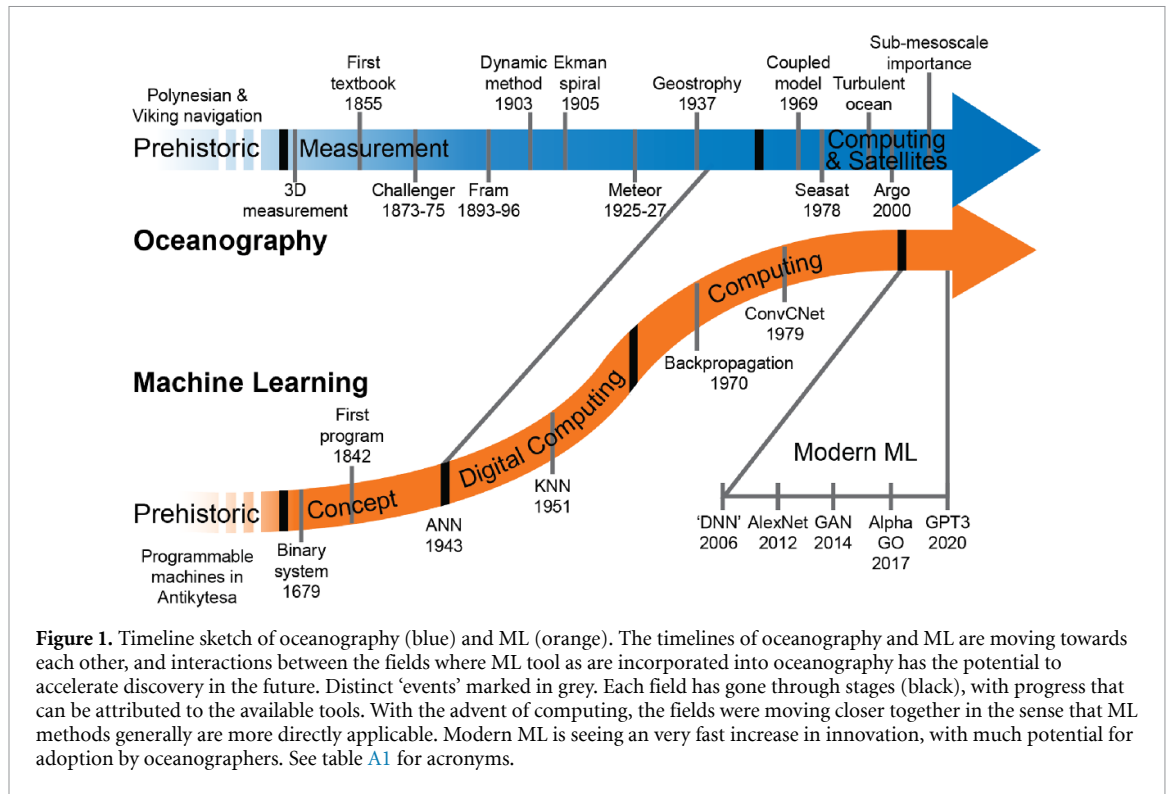
Figure 1. Timeline sketch of oceanography (blue) and ML (orange). The timelines of oceanography and ML are moving towards each other, and interactions between the fields where ML tool as are incorporated into oceanography has the potential to accelerate discovery in the future. Distinct 'events' marked in grey. Each field has gone through stages (black), with progress that can be attributed to the available tools. With the advent of computing, the fields were moving closer together in the sense that ML methods generally are more directly applicable. Modern ML is seeing an very fast increase in innovation, with much potential for adoption by oceanographers. See table A1 for acronyms.
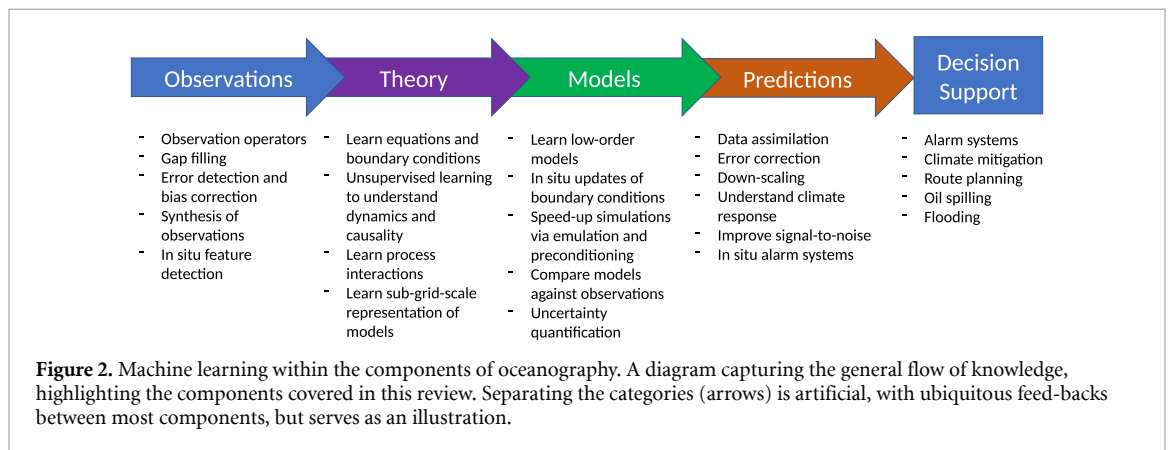

Figure 2. Machine learning within the components of oceanography. A diagram capturing the general flow of knowledge, highlighting the components covered in this review. Separating the categories (arrows) is artificial, with ubiquitous feed-backs between most components, but serves as an illustration.

to perform selected tasks by 'learning' from data. This field has undergone a series of impressive breakthroughs over the last years thanks to the increasing availability of data and the recent developments in computational and data storage capabilities. Several classes of algorithms are associated with the different applications of ML. They can be categorized into three main classes: supervised learning, unsupervised learning, and reinforcement learning (RL). In this review, we focus on the first two classes which are the most commonly used to date in the ocean sciences.

1.2.1. Supervised learning
Supervised learning refers to the task of inferring a relationship between a set of inputs and their corresponding outputs. In order to establish this relationship, a 'labelled' dataset is used to constrain

the learning process and assess the performance of the ML algorithm. Given a dataset of N pairs of input-output training examples $\{(x^{(i)}, y^{(i)})\}_{i \in 1 \ldots N}$ and a loss function $\mathcal{L}$ that represents the discrepancy between the ML model prediction and the actual outputs, the parameters $\boldsymbol{\theta}$ of the ML model f are found by solving the following optimization problem:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}\left(f\left(x^{(i)}; \boldsymbol{\theta}\right), y^{(i)}\right). \quad (1)$$

If the loss function is differentiable, then gradient descent based algorithms can be used to solve equation (1). These methods rely on an iterative tuning of the models' parameters in the direction of the negative gradient of the loss function. At each iteration k, the parameters are updated as follows:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \mu \nabla \mathcal{L}(\boldsymbol{\theta}_k), \quad (2)$$

where μ is the rate associated with the descent and is called the learning rate and ∇ the gradient operator.

Two important applications of supervised learning are regression and classification. Popular statistical techniques such as least squares or ridge regression, which have been around for a long time, are special cases of a popular supervised learning technique called linear regression (in a sense, we may consider a large number of oceanographers to be early ML practitioners.) For regression problems, we aim to infer continuous outputs and usually use the mean squared error (MSE) or the mean absolute error (MAE) to assess the performance of the regression. In contrast, for supervised classification problems we sort the inputs to a number of classes or categories that have been pre-defined. In practice, we often transform the categories into probability values of belonging to some class and use distribution-based distances such as the cross-entropy to evaluate the performance of the classification algorithm.

Numerous types of supervised ML algorithms have been used in the context of ocean research, as detailed in the following sections. Notable methods include:

- *Linear univariate (or multivariate) regression (LR)*, where the output is a linear combination of some explanatory input variables. LR is one of the first ML algorithms to be studied extensively and used for its ease of optimization and its simple statistical properties [181].
- *k-nearest neighbours (KNN)*, where we consider an input vector, find its k closest points with regard to a specified metric, then classify it by a plurality vote of these k points. For regression, we usually take the average of the values of the k neighbours. KNN is also known as 'analog methods' in the numerical weather prediction community [163].
- *Support vector machines (SVM)* [61], where the classification is done by finding a linear separating hyperplane with the maximal margin between two classes (the term 'margin' here denotes the space between the hyperplane and the nearest points in either class.) In case of data which cannot be separated linearly, the use of the *kernel trick* projects the data into a higher dimension where the linear separation can be done. *Support vector regression (SVR)* are an adaption of SVMs for regression problems.
- *Random forests (RF)* that are a composition of a multitude of decision trees (DT). DTs are constructed as a tree-like composition of simple decision rules [29].
- *Gaussian process regression (GPR)* [264], also called kriging, is a general form of the optimal interpolation algorithm, which has been used in the oceanographic community for a number of years
- *Neural networks (NN)*, a powerful class of universal approximators that are based on compositions of

interconnected nodes applying geometric transformations (called affine transformations) to inputs and a nonlinearity function called an 'activation function' [66]

The recent ML revolution, i.e. the so-called deep learning (DL) era that began in the early 2010s, sparked off thanks to the scientific and engineering breakthroughs in training neural networks (NN), combined with the proliferation of data sources and the increasing computational power and storage capacities. The simplest example of this advancement is the efficient use of the algorithm of backpropagation (known in the geoçience community as the adjoint method) combined with stochastic gradient descent for the training of multi-layer NNs, i.e. NNs with multiple layers, where each layer takes the result of the previous layer as an input, applies the mathematical transformations and then yields an input for the next layer [25]. DL research is a field receiving intense focus and fast progress through its use both commercially and scientifically, resulting in new types of 'architectures' of NNs, each adapted to particular classes of data (text, images, time series, etc) [155, 219]. We briefly introduce the most popular architectures used in deep learning research and highlight some applications:

- Multilayer perceptrons (MLP): when used without qualification, this term refers to fully connected feed forward multilayered neural networks. They are composed of an input layer that takes the input data, multiple hidden layers that convey the information in a 'feed forward' way (i.e. from input to output with no exchange backwards), and finally an output layer that yields the predictions. Any neuron in a MLP is connected to all the neurons in the previous and to those of next layer, thus the use of the term 'fully connected'. MLPs are mostly used for tabular data.
- Convolutional neural networks (ConvNet): contrarily to MLPs, ConvNets are designed to take into account the local structure of particular type of data such as text in 1D, images in 2D, volumetric images in 3D, and also hyperspectral data such as that used in remote sensing. Inspired by the animal visual cortex, neurons in ConvNets are not fully connected, instead they receive information from a subarea spanned by the previous layer called the 'receptive field'. In general, a ConvNet is a feed forward architecture composed of a series of convolutional layers and pooling layers and might also be combined with MLPs. A convolution is the application of a filter to an input that results in an activation. One convolutional layer consist of a group of 'filters' that perform mathematical discrete convolution operations, the result of these convolutions are called 'feature maps'. The filters along with biases are the parameters of the

ConvNet that are learned through backpropagation and stochastic gradient descent. Pooling layers serve to reduce the resolution of feature maps which lead to compressing the information and speeding up the training of the ConvNet, they also help the ConvNet become invariant to small shift in input images [155]. ConvNets benefited much from the advancements in GPU computing and showed great success in the computer vision community.

- Recurrent neural networks (RNN): with an aim to model sequential data such as temporal signals or text data, RNNs were developed with a hidden state that stores information about the history of the sequences presented to its inputs. While theoretically attractive, RNNs were practically found to be hard to train due to the exploding/vanishing gradient problems, i.e. backpropagated gradients tend to either increase too much or shrink too much at each time step [127]. Long short term memory (LSTM) architecture provided a solution to this problem through the use of special hidden units [219]. LSTMs are to date the most popular RNN architectures and are used in several applications such as translation, text generation, time series forecasting, etc. Note that a variant for spatiotemporal data was developed to integrate the use of convolutional layers, this is called ConvLSTM [224].

1.2.2. Unsupervised learning
Unsupervised learning is another major class of ML. In these applications, the datasets are typically unlabelled. The goal is then to discover patterns in the data that can be used to solve particular problems. One way to say this is that unsupervised classification algorithms identify sub-populations in data distributions, allowing users to identify structures and potential relationships among a set of inputs (which are sometimes called 'features' in ML language). Unsupervised learning is somewhat closer to what humans expect from an intelligent algorithm, as it aims to identify latent *representations* in the structure of the data while filtering out unstructured noise. At the NeurIPS 2016 conference, Yann LeCun, a DL pioneer researcher, highlighted the importance of unsupervised learning using his *cake analogy*: 'If machine learning is a cake, then unsupervised learning is the actual cake, supervised learning is the icing, and RL is the cherry on the top.'

Unsupervised learning is achieving considerable success in both clustering and dimensionality reduction applications. Some of the unsupervised techniques that are mentioned throughout this review are:

- *k-means*, a popular and simple space-partitioning clustering algorithm that finds classes in a dataset by minimizing within-cluster variances [230]. Gaussian Mixture Models (GMMs) can be seen as a generalization of the k-means algorithm that assumes the data can be represented by a mixture (i.e. linear combination) of a number of multidimensional Gaussian distributions [176].

- Kohonen maps [also called *Self Organizing Maps* (SOM)] is a NN based clustering algorithm that leverages topology of the data; nearby locations in a learned map are placed in the same class [147]. K-means can be seen as a special case of SOM with no information about the neighbourhood of clusters.

- *t-SNE* and *UMAP* are two other clustering algorithms which are often used for not only finding clusters but also because of their data visualization properties which enables a two or three dimensional graphical rendition of the data [175, 250]. These methods are useful for representing the structure of a high-dimensional dataset in a small number of dimensions that can be plotted. For the projection, they use a measure of the 'distance' or 'metric' between points, which is a sub-field of mathematics where methods are increasingly implemented for t-SNE or UMAP.

- *Principal Component Analysis* (PCA) [191], the simplest and most popular dimensionality reduction algorithm. Another term for PCA is Empirical Orthogonal Function analysis (EOF), which has been used by physical oceanographers for many years, also called Proper Orthogonal Decomposition (POD) in computational fluids literature.

- *Autoencoders* (AE) are NN-based dimensionality reduction algorithms, consisting of a bottleneck-like architecture that learns to reconstruct the input by minimizing the error between the output and the input (i.e. ideally the data given as input and output of the autoencoder should be interchangeable). A central layer with a lower dimension than the original inputs' dimension is called a 'code' and represents a compressed representation of the input [149].

- *Generative modelling*: a powerful paradigm that learns the latent features and distributions of a dataset and then proceeds to generate new samples that are plausible enough to belong to the initial dataset. Variational auto-encoders (VAEs) and generative adversarial networks (GANS) are two popular techniques of generative modelling that benefited much from the DL revolution [111, 144].

Between supervised and unsupervised learning lies *semi-supervised learning*. It is a special case where one has access to both labelled and unlabelled data. A classical example is when labelling is expensive, leading to a small percentage of labelled data and a high percentage of unlabelled data.

Reinforcement learning is the third paradigm of ML; it is based on the idea of creating algorithms where an *agent* explores an *environment* with the aim of reaching some goal. The agent learns through a trial and error mechanism, where it performs an

action and receives a response (reward or punishment), the agent learns by maximizing the expected sum of rewards [238]. The DL revolution did also affect this field and led to the creation of a new field called deep reinforcement learning (Deep RL) [233]. A popular example of Deep RL that got huge media attention is the algorithm AlphaGo developed by DeepMind which beat human champions in the game of Go [225].

The importance of understanding why an ML method arrived at a result is not confined to oceanographic applications. Unsupervised ML lends itself more readily to being interpreted (IAI), but for example for methods building on DL or NN in general, a growing family of methods collectively referred to as additive feature attribution (AFA) is becoming popular, largely applied for XAI. AFA methods aim to explain predictive skill retrospectively. These methods include connection weight approaches, local interpretable model-agnostic explanations (LIME), Shapley additive explanation (SHAP) and layer-wise relevance propagation (LRP) [26, 153, 165, 179, 193, 208, 228, 246]. Non-AFA methods rooted in 'saliency' mapping also exist [174].

The goal of this review paper is not to delve into the definitions of ML techniques but only to briefly introduce them to the reader and recommend references for further investigation. The textbook by Bishop [30] covers essentials of the fields of pattern recognition and Hsieh's book [131] is probably one of earliest attempts at writing a comprehensive review of ML methods targeted at earth scientists. Another notable review of statistical methods for physical oceanography is the paper by Wikle *et al* [262]. We also refer the interested reader to the book of Goodfellow *et al* [25] to learn more about the theoretical foundations of DL and some of its applications in science and engineering.

1.3. ML in atmospheric and the wider earth system sciences

Precursors to modern ML methods, such as regression and principal component analysis, have of course been used in many fields of Earth system science for decades. The use of PCA, for example, was popularized in meteorology in [162], as a method of dimensionality reduction of large geospatial datasets, where Lorenz also speculates here on the possibility of purely statistical methods of long-term weather prediction based on a representation of data using PCA. Methods for discovering correlations and links, including possible causal links, between dataset features using formal methods have seen much use in Earth system science. e.g. [18]. For example, Walker [256] was tasked with discovering the cause for the interannual fluctuation of the Indian monsoon, whose failure meant widespread drought in India, and in colonial times also famine [68]. To find possible correlations, Walker put to work an army of Indian clerks to carry

out a vast computation by hand across all available data. This led to the discovery of the Southern Oscillation, the seesaw in the West-East temperature gradient in the Pacific, which we know now by its modern name, El Niño Southern Oscillation (ENSO). Beyond observed correlations, theories of ENSO and its emergence from coupled atmosphere-ocean dynamics appeared decades later [270]. Walker speaks of statistical methods of discovering 'weather connections in distant parts of the earth', or teleconnections. The ENSO-monsoon teleconnection remains a key element in diagnosis and prediction of the Indian monsoon [236, 237]. These and other data-driven methods of the pre-ML era are surveyed in [42]. ML-based predictive methods targeted at ENSO are also being established [120]. Here, the learning is not directly from observations but from models and reanalysis data, and outperform some dynamical models in forecasting ENSO.

There is an interplay between data-driven methods and physics-driven methods that both strive to create insight into many complex systems, where the ocean and the wider Earth system science are examples. As an example of physics-driven methods [11], Bjerknes and other pioneers discussed in section 1.1 formulated accurate theories of the general circulation that were put into practice for forecasting with the advent of digital computing. Advances in numerical methods led to the first practical physics-based atmospheric forecast [199]. Until that time, forecasting often used data-driven methods 'that were neither algorithmic nor based on the laws of physics' [187]. ML offers avenues to a synthesis of data-driven and physics-driven methods. In recent years, as outlined below in section 4.3, new processors and architectures within computing have allowed much progress within forecasting and numerical modelling overall. ML methods are poised to allow Earth system science modellers to increase the efficient use of modern hardware even further. It should be noted however that 'classical' methods of forecasting such as analogues also have become more computationally feasible, and demonstrate equivalent skill, e.g. [73]. The search for analogues has become more computationally tractable as well, although there may also be limits here [76].

Advances in numerical modelling brought in additional understanding of elements in Earth system science which are difficult to derive, or represent from first principles. Examples include cloud microphysics or interactions with the land surface and biosphere. For capturing cloud processes within models, the actual processes governing clouds take place at scales too fine to model and will remain out of reach of computing for the foreseeable future [221]. A practical solution to this is finding a representation of the aggregate behaviour of clouds at the resolution of a model grid cell. This has proved quite difficult and progress over many decades has been halting

[36]. The use of ML in deriving representations of clouds is now an entire field of its own. Early results include the results of [105], using NNs to emulate a 'super-parameterized' model. In the super-parameterized model, there is a clear (albeit artificial) separation of scales between the 'cloud scale' and the large scale flow. When this scale separation assumption is relaxed, some of the stability problems associated with ML re-emerge [41]. There is also a fundamental issue of whether learned relationships respect basic physical constraints, such as conservation laws [160]. Recent advances [27, 268] focus on formulating the problem in a basis where invariances are automatically maintained. But this still remains a challenge in cases where the physics is not fully understood.

There are at least two major efforts for the systematic use of ML methods to constrain the cloud model representations in GCMs. First, the calibrate-emulate-sample (CES [58, 81]) approach uses a more conventional model for a broad calibration of parameters also referred to as 'tuning'[129]. This is followed by an emulator, that calibrates further and quantifies uncertainties. The emulator is an ML-based model that reproduces most of the variability of the reference model, but at a lower computational cost. The low computational cost enables the emulator to be used to produce a large ensemble of simulations, that would have been too computationally expensive to produce using the model that the emulator is based on. It is important to retain the uncertainty quantification aspect (represented by the emulated ensemble) in the ML context, as it is likely that the data in a chaotic system only imperfectly constrain the loss function. Second, emulators can be used to eliminate implausible parameters from a calibration process, demonstrated by the HighTune project [63, 130]. This process can also identify 'structural error', indicating that the model formulation itself is incorrect, when no parameter choices can yield a plausible solution. Model errors are discussed in section 5.1. In an ocean context, the methods discussed here can be a challenge due to the necessary forwards model component. Note also, that ML algorithms such as GPR are ubiquitous in emulating problems thanks to their built-in uncertainty quantification. GPR methods are also popular because their application involves a low number of training samples, and function as inexpensive substitutes for a forward model.

Model resolution that is inadequate for many practical purposes has led to the development of data-driven methods of 'downscaling'. For example climate change adaptation decision-making at the local level based on climate simulations too coarse to feature enough detail. Most often, a coarse-resolution model output is mapped onto a high-resolution reference truth, for example given by observations [4, 251]. Empirical-statistical downscaling (ESD, [24]) is an example of such methods. While ESD emphasized the downscaling aspect, all of these downscaling methods include a substantial element of bias correction. This is highlighted in the name of some of the popular methods such as Bias Correction and Spatial Downscaling [265] and Bias Corrected Constructed Analogue [171]. These are trend-preserving statistical downscaling algorithms, that combine bias correction with the analogue method of Lorenz [164]. ML methods are rapidly coming to dominate the field as discussed in section 5.1, with examples ranging from precipitation (e.g. [252]), surface winds and solar outputs [231], as well as to unresolved river transport [108]. Downscaling methods continue to make the assumption that transfer functions learned from present-day climate continue to hold in the future. This stationarity assumption is a potential weakness of data-driven methods [74, 192], that requires a synthesis of data-driven and physics-based methods as well.

2. Ocean observations

Observations continue to be key to oceanographic progress, with ML increasingly being recognized as a tool that can enable and enhance what can be learned from observational data, performing conventional tasks better/faster, as well as bring together different forms of observations, facilitating comparison with model results. ML offers many exciting opportunities for use with observations, some of which are covered in this section and in section 5 as supporting predictions and decision support.

The onset of the satellite observation era brought with it the availability of a large volume of effectively global data, challenging the research community to use and analyse this unprecedented data stream. Applications of ML intended to develop more accurate satellite-driven products go back to the 90s [241]. These early developments were driven by the data availability, distributed in normative format by the space agencies, and also by the fact that models describing the data were either empirical (e.g. marine biogeochemistry [218]) or too computationally costly and complex (e.g. radiative transfer [143]). More recently, ML algorithms have been used to fuse several satellite products [116] and also satellite and in-situ data [52, 70, 142, 170, 185]. For the processing of satellite data, ML has proven to be a valuable tool for extracting geophysical information from remotely sensed data (e.g. [51, 82]), whereas a risk of using only conventional tools is to exploit only a more limited subset of the mass of data available. These applications are based mostly on instantaneous or very short-term relationships and do not address the problem of how these products can be used to improve our ability to understand and forecast the oceanic system. Further use for current reconstruction using ML [169], heat fluxes [106], the

3-dimensional circulation [228], and ocean heat content [135] are also being explored.

There is also an increasingly rich body of literature mining ocean in-situ observations. These leverage a range of data, including Argo data, to study a range of ocean phenomena. Examples include assessing North Atlantic mixed layers [172], describing spatial variability in the Southern Ocean [138], detecting El Niño events [128], assessing how North Atlantic circulation shifts impacting heat content [71], and finding mixing hot spots [213]. ML has also been successfully applied to ocean biogeochemistry. While not covered in detail here, examples include mapping oxygen [110] and CO_2 fluxes [46, 152, 259].

Modern in-situ classification efforts are often property-driven, carrying on long traditions within physical oceanography. For example, characteristic groups or 'clusters' of salinity, temperature, density or potential vorticity have typically been used to delineate important water masses and to assess their spatial extent, movement, and mixing [121, 126]. However, conventional identification/classification techniques assume that these properties stay fixed over time. The techniques largely do not take interannual and longer timescale variability into account. The prescribed ranges used to define water masses are often somewhat ad-hoc and specific (e.g. mode waters are often tied to very restrictive density ranges) and do not generalize well between basins or across longer timescales [9]. Although conventional identification/classification techniques will continue to be useful well into the future, unsupervised ML offers a robust, alternative approach for objectively identifying structures in oceanographic observations [33, 138, 197, 213].

To analyse data, dimensionality and noise reduction methods have a long history within oceanography. PCA is one such method, which has had a profound influence on oceanography since Lorenz first introduced it to the geosciences in 1956 [162]. Despite the method's shortcomings related to strong statistical assumptions and misleading applications, it remains a popular approach [178]. PCA can be seen as a super sparse rendition of k-means clustering [72] with the assumption of an underlying normal distribution in its commonly used form. Overall, different forms of ML can offer excellent advantages over more commonly used techniques. For example, many clustering algorithms can be used to reduce dimensionality according to how many significant clusters are identifiable in the data. In fact, unsupervised ML can sidestep statistical assumptions entirely, for example by employing density-based methods such as DBSCAN [227]. Advances within ML are making it increasingly possible and convenient to take advantage of methods such as t-SNE [227] and UMAP, where the original topology of the data can be conserved in a low-dimensional rendition.
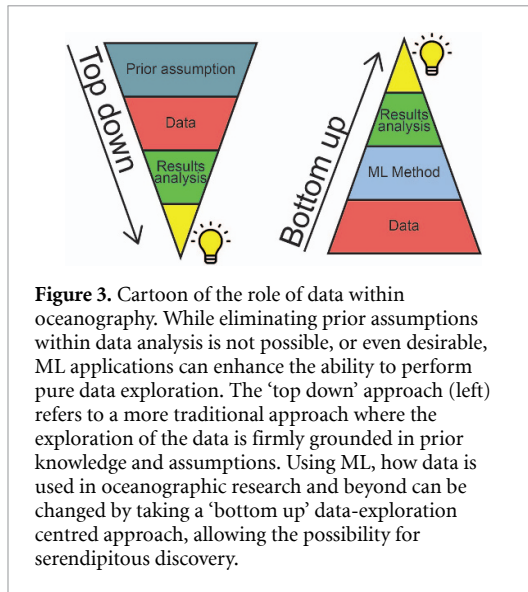
Interpolation of missing data in oceanic fields is another application where ML techniques have been used, yielding products used in operational contexts. For example, Kriging is a popular technique that was successfully applied to altimetry [154], as it can account for observation from multiple satellites with different spatio-temporal sampling. In its simplest form, kriging estimates the value of an unobserved location as the linear combination of available observations. Kriging also yields the uncertainty of this estimate, which has made it popular in geostatistics. EOF-based techniques are also attracting increasing attention with the proliferation of data. For example, the DINEOF algorithm [6] leverages the availability of historical datasets, to fill in spatial gaps within new observations. This is done via projection onto the space spanned by dominant EOFs of the historical data. The use of advanced supervised learning, such as DL, for this problem in an oceanographic contexts is still in its infancy. Attempts exist in the literature, including deriving a DL equivalent of DINEOF for interpolating SST [19].

3. Exchanges between observations and theory

Progress within observations, modelling, and theory go hand in hand, and ML offers a novel method for bridging the gaps between the branches of oceanography. When describing the ocean, theoretical descriptions of circulation tend to be oversimplified, but interpreting basic physics from numerical simulations or observations alone is prohibitively difficult. Progress in theoretical work has often come from the discovery or inference of regions where terms in an equation may be negligible, allowing theoretical developments to be focused with the hope of observational verification. Indeed, progress in identifying negligible terms in fluid dynamics could be said to underpin GFD as a whole [249]. For example, Sverdrup's theory [235] of ocean regions where the wind stress curl is balanced by the Coriolis term inspired a search for a predicted 'level of no motion' within the ocean interior.

The conceptual and numerical models that underlie modern oceanography would be less valuable if not backed by observational evidence, and similarly, findings in data from both observations and numerical models can reshape theoretical models [101]. ML algorithms are becoming heavily used to determine patterns and structures in the increasing volumes of observational and modelled data [33, 47, 71, 128, 138, 139, 172, 197, 213, 229, 240]. For example, ML is poised to help the research community reframe the concept of ocean fronts in ways that are tailored to specific domains instead of ways that are tied to somewhat ad-hoc and overgeneralized property definitions [54]. Broadly speaking, this area of work largely utilizes unsupervised ML and is thus well-positioned to discover underlying structures and patterns in data that can help identify negligible

Figure 3. Cartoon of the role of data within oceanography. While eliminating prior assumptions within data analysis is not possible, or even desirable, ML applications can enhance the ability to perform pure data exploration. The 'top down' approach (left) refers to a more traditional approach where the exploration of the data is firmly grounded in prior knowledge and assumptions. Using ML, how data is used in oceanographic research and beyond can be changed by taking a 'bottom up' data-exploration centred approach, allowing the possibility for serendipitous discovery.

terms or improve a conceptual model that was previously empirical. In this sense, ML methods are well-placed to help guide and reshape established theoretical treatments, for example by highlighting overlooked features. A historical analogy can be drawn to d'Alembert's paradox from 1752 (or the hydrodynamic paradox), where the drag force is zero on a body moving with constant velocity relative to the fluid. Observations demonstrated that there *should* be a drag force, but the paradox remained unsolved until Prandtl's 1904 discovery of a thin boundary layer that remained as a result of viscous forces. Discoveries like Prandtl's can be difficult, for example because the importance of small distinctions that here form the boundary layer regime can be overlooked. ML has the ability to be both objective, and also to highlight key distinctions like a boundary layer regime. ML is ideally poised to make discoveries possible through its ability to objectively analyse the increasingly large and complicated data available. Using conventional analysis tools, finding patterns inadvertently rely on subjective 'standards' e.g. how the depth of the mixed layer or a Southern Ocean front is defined [54, 75, 243]. Such standards leave room for bias and confusion, potentially perpetuating unhelpful narratives such as those leading to d'Alembert's paradox.

With an exploration of a dataset that moves beyond preconceived notions comes the potential for making entirely new discoveries. It can been argued that much of the progress within physical oceanography has been rooted in generalizations of ideas put forward over 30 years ago [101, 137, 184]. This foundation can be tested using data to gain insight in a 'top-down' manner (figure 3). ML presents a possible opportunity for serendipitous discovery outside of this framework, effectively using data as the foundation and achieving insight purely through its objective analysis in a 'bottom up' fashion. This can also be

achieved using conventional methods but is significantly facilitated by ML, as modern data in its often complicated, high dimensional, and voluminous form complicates objective analysis. ML, through its ability to let structures within data emerge, allows the structures to be systematically analysed. Such structures can emerge as regions of coherent covariance (e.g. using clustering algorithms from unsupervised ML), even in the presence of highly non-linear and intricate covariance [227]. Such structures can then be investigated in their own right and may potentially form the basis of new theories. Such exploration is facilitated by using an ML approach in combination with IAI and XAI methods as appropriate. Unsupervised ML lends itself more readily to IAI and to many works discussed above. Objective analysis that can be understood as IAI can also be applied to explore theoretical branches of oceanography, revealing novel structures [47, 229, 240]. Examples where ML and theoretical exploration have been used in synergy by allowing interpretability, explainability, or both within oceanography include [228, 269], and the concepts are discussed further in section 6.

As an increasingly operational endeavour, physical oceanography faces pressures apart from fundamental understanding due to the increasing complexity associated with enhanced resolution or the complicated nature of data from both observations and numerical models. For advancement in the fundamental understanding of ocean physics, ML is ideally placed to break this data down to let salient features emerge that are comprehensible to the human brain.

3.1. ML and hierarchical statistical modelling

The concept of a model hierarchy is described by [125] as a way to fill the 'gap between simulation and understanding' of the Earth system. A hierarchy consists of a set of models spanning a range of complexities. One can potentially gain insights by examining how the system changes when moving between levels of the hierarchy, i.e. when various sources of complexity are added or subtracted, such as new physical processes, smaller-scale features, or degrees of freedom in a statistical description. The hierarchical approach can help sharpen hypotheses about the oceanographic system and inspire new insights. While perhaps conceptually simple, the practical application of a model hierarchy is non-trivial, usually requiring expert judgement and creativity. ML may provide some guidance here, for example by drawing attention to latent structures in the data. In this review, we distinguish between statistical and numerical ML models used for this purpose. For ML-mediated models, a goal could be discovering other levels in the model hierarchy from complex models [11]. The models discussed in sections 2 and 3 constitute largely statistical models,

such as ones constructed using a k-means application, GANs, or otherwise. This section discusses the concept of hierarchical models in a statistical sense, and section 4.2 explores the concept of numerical hierarchical models. A hierarchical statistical model can be described as a series of model descriptions of the same system from very low complexity (e.g. a simple linear regression) to arbitrarily high. In theory, any statistical model constructed with any data from the ocean could constitute a part of this hierarchy, but here we restrict our discussion to models constructed from the same or very similar data.

The concept of exploring a hierarchy of models, either statistical or otherwise, using data could also be expressed as searching for an underlying manifold [161]. The notion of identifying the 'slow manifold' postulates that the noisy landscape of a loss function for one level of the hierarchy, conceals a smoother landscape in another level. As such, it should be plausible to identify a continuum of system descriptions. ML has the potential to assist in revealing such an underlying slow manifold, as described above. For example, equation discovery methods shown promise as they aim to find closed form solutions to the relations within datasets representing terms in a parsimonious representation (e.g. [100, 220, 269] are examples in line with [11]). Similarly, unsupervised equation exploration could hold promise for utilizing formal ideas of hypothesis forming and testing within equation space [140].

In oceanographic ML applications, there are tunable parameters that are often only weakly constrained. A particular example is the total number of classes K in unsupervised classification problems [138, 139, 172, 227, 229]. Although one can estimate the optimal value K^* for the statistical model, for example by using metrics that reward increased likelihood and penalize overfitting (e.g. the Bayesian information criteria (BIC) or the Akaike information criterion (AIC)), in practice it is rare to find a clear value of K^* in oceanographic applications. Often, tests like BIC or AIC return either a range of possible K^* values, or they only indicate a lower bound for K. This is perhaps because oceanographic data is highly correlated across many different spatial and temporal scales, making the task of separating the data into clear sub-populations a challenging one. That being said, the parameter K can also be interpreted as the complexity of the statistical model. A model with a smaller value of K will potentially be easier to interpret because it only captures the dominant sub-populations in the data distribution. In contrast, a model with a larger value of K will likely be harder to interpret because it captures more subtle features in the data distribution. For example, when applied to Southern Ocean temperature profile data, a simple two-class profile classification model will tend to separate the profiles into those north and south of the Antarctic Circumpolar Current, which

is a well-understood approximate boundary between polar and subtropical waters. By contrast, more complex models capture more structure but are harder to interpret using our current conceptual understanding of ocean structure and dynamics [138]. In this way, a collection of statistical models with different values of K constitutes a model hierarchy, in which one builds understanding by observing how the representation of the system changes when sources of complexity are added or subtracted [125]. Note that for the example of k-means, while a range of K values may be reasonable, this does not largely refer to merely adjusting the value of K and re-interpreting the result. This is because, for example, if one moves from $K = 2$ to $K = 3$ using k-means, there is no *a priori* reason to assume they would both give physically meaningful results. What is meant instead is similar to the type of hierarchical clustering that is able to identify different sub-groups and organize them into larger overarching groups according to how similar they are to one another. This is a distinct approach within ML that relies on the ability to measure a 'distance' between data points. This rationale reinforces the view that ML can be used to build our conceptual understanding of physical systems, and does not need to be used simply as a 'black box'. It is worth noting that the axiom that is being relied on here is that there exists an underlying system that the ML application can approximate using the available data. With incomplete and messy data, the tools available to assess the fit of a statistical model only provide an estimate of how wrong it is certain to be. To create a statistically rigorous hierarchy, not only does the overall co-variance structure/topology need to be approximated, but also the finer structures that would be found within these overarching structures. If this identification process is successful, then the structures can be grouped with accuracy as defined by statistical significance. This can pose a formidable challenge that ML in isolation cannot address; it requires guidance from domain experts. For example, within ocean ecology, [227] derived a hierarchical model by grouping identified clusters according to ecological similarity. In physical oceanography, [213] grouped some identified classes together into zones using established oceanographic knowledge, in a step from a more complex statistical model to a more simplified one that is easier to interpret. When performing such groupings, one has to pay attention to a balance of statistical rigour and domain knowledge. Discovering rigorous and useful hierarchical models should hypothetically be possible, as demonstrated by the self-similarity found in many natural systems including fluids, but limited and imperfect data greatly complicate the search, meaning that checking for statistical rigour is important.

As a possible future direction, assessing models using IAI and XAI and known physical relationships will likely make finding hierarchical models that are

meaningful much easier. Unsupervised ML is more intuitively interpretable than supervised ML and may prove more useful for identifying such hierarchical models. Moving away from using ML in a 'black box' sense, with IAI and XAI or otherwise, may yield a synthesis of observations and theory, allowing the field to go beyond the limitations of either; theory may allow one to generalize beyond the limits of data, and data may reveal possible structural errors in theory.

4. From theory to numerical models

The observation of patterns in data is the precursor to a theory, which can lead to predictive models, provided theory can be converted to practical computation. In this section, we discuss how ML could change the way theory is represented within ocean modelling. To represent the ocean using numerical models is to help filling in missing information between observations. In addition, models act as virtual laboratories in which we can work to understand physical relationships. For example, how the separation of boundary currents such as the Gulf Stream depends on local topography or boundary conditions. The focus of this discussion will be on models that represent the three-dimensional ocean circulation, but most of these ideas can also be used in the context of modelling sea-ice, tides, waves, or biogeochemistry. We also discuss a recurring issue within ocean modelling: the presence of coastlines that complicate the application of methods that are convolutional or spectral.

4.1. Timescales and space scales
When building numerical models, the ocean is largely treated like a typical fluid that follows the Navier–Stokes equations, and the challenges faced therein are similar to those presented by general computational fluid dynamics. The filamentation of the flow results in scale interactions that make it necessary to represent all spatial scales within the model, while the model resolution needs to be truncated due to the finite nature of computational power. The dynamics at different scales can either be represented via the explicit, resolved representation within the model or via the parametrization of sub-grid-scales as a turbulent closure.

Much research has gone into the formulation of parametrizations to represent the sub-grid-scales. Such representations range from classical closures for turbulent fluids, using formulations such as Gent-McWilliams [104] that take the dynamics of sub-grid ocean eddies into account, to empirical closure schemes that are determined by comparing simulations at a target resolution to simulations at higher resolution [60, 215]. Lately, ML has also been used to learn the sub-grid-scale, either via the direct learning of the terms using NNs [34] or via the learning

of the underlying equations [269]. Similar and promising DA applications are also emerging, discussed in section 5.2.

Next to the representation of the sub-grid-scale, numerical ocean models are also prone to errors due to the necessary discretization of the differential equations on a numerical grid. A number of methods are used to discretize the equations [83], including finite difference, finite volume, and finite element methods. In comparison to the atmosphere, spectral discretization methods cannot easily be applied to the ocean due to the presence of coastlines, as creating a representation using global basis functions is not straightforward.

In the presence of perfect data and adequate computational power with which to train a DL application, it would be theoretically possible to learn the dynamics of the ocean with no knowledge of the equations of motion. This is because DL can learn the update of the physical fields based on time-series of observations or model data. This has been done successfully for certain atmospheric applications [78, 204, 261] and for an idealized ocean model [99]. However, DL representations of the ocean are more difficult than for the atmosphere. This is because there is much less reliable three-dimensional training data available for the ocean spatially, and because relevant time-scales of the ocean are much longer together with the shorter time scales that together make up the ocean state. This is because the ocean has important low-frequency variability, resulting in a need for longer training data sets. Furthermore, coastlines form lateral boundaries that may reduce the quality of NN solutions. This is because, NN models often require a certain stencil of local information to update the physical fields at a given gridpoint. For example, CNNs perform best if the underlying system is invariant by translation. While grid-points on land could be incorporated into local stencils with pre-defined values, the presence of coastlines may reduce the amount of training data for specific pattern of the coast line. Also, having sharp discontinuities from ocean to land pixels results in a more challenging problem for NN in general. For example, a CNN could spend a considerable amount of its parameters learning the coastline boundary patterns, which may not be of interest if the user is focusing on ocean-only patterns.

ML tools could also serve as a method to represent the ocean with fewer degrees of freedom than a full conventional numerical model. Such use cases for ML include being used (1) as part of a coupled Earth system model that is either used for short-term weather forecasts, or (2) in long climate simulations. For example, if a model is only trying to represent the surface fields that are most important for the coupling to the atmosphere, the model could focus on the use of the leading principal components (if these can be derived in the presence of coastlines),

and learn the interactions between the different components using data from a time-series extracted from long model (or observational) trajectories. Here, a first approach towards building low-order ML models using a barotropic model showed that results from high-dimensional ML tools from DL, such as complex RNNs, may not necessarily provide better results when compared to ML techniques that are based on regression techniques and stochastic forcing [5].

The surface of the ocean has strong and direct impact on society, and ML tools may help to better understand and characterize its dynamics. For coupled models, the importance of representing the interactions between the atmosphere and sea ice components is well established. However, allowing this interaction involves major computational challenges and uncertainties, associated with understanding of the complex feedback and coupling processes between system components. Many surface processes, for example ocean waves and storm surges, are typically excluded from long-term ocean simulations and modelled separately, although exceptions exist [226]. This exclusion is largely because of the very large differences in phase speeds, for example between surface waves and the deep ocean that result in a barotropic (fast) component being treated separately from a baroclinic (slow) component [112, 245]. However, there are now approaches to improve wave and sea-ice modelling or predictions using ML, that either try to improve on the computational efficiency or the accuracy of conventional methods [8, 56, 195, 200, 271].

4.2. Concepts of ML and hierarchical numerical modelling

This section discusses hierarchical modelling in a numerical sense, complementing section 3.1 that discusses hierarchical modelling in a statistical sense. Within oceanography, observations and theory are more meaningful when viewed together. Observational scientists (see section 2) make choices of what to sample based on some prior conceptions of relevance, and of course theory is ungrounded without data. In epistemology, this is often summarized in Duhem's formulation, 'theory is data-laden, and data is theory-laden' [80]. In talking about climate and weather modelling, Edwards made the corollary, 'models are data-laden, and data is model-laden' [86]. For example, the concept of a reanalysis dataset comes from a model. The sequence from observations to theory to models and predictions shows this interplay. This is a key sequence where we expect ML to display its strengths, e.g. where IAI and XAI methods may yield a synthesis of observations and theory, allowing one to go beyond the limitations of either: theory allowing one to generalize beyond the limits of data, and data revealing possible structural errors in theory as detailed in section 3. Ideally, we would like to go beyond these and use ML to discover the underlying equations (e.g. [43]), and deliver

a model hierarchy that can then be implemented numerically [11, 125]. While simple in principle, in practice this concept is less straightforward to implement. An example of a form of equation discovery can be seen in Zanna and Bolton's [269] reduction of resolved turbulent dynamics into a representation suitable for use in coarse-grained models. The coarse-grained models represent a different level of the hierarchy, if tiers are set by horizontal resolution. This ML model was arrived at applying an RVM, with the different equation terms serving as the input. This is an example where the results of an ML application yield a parsimonious closed-form representation, and are therefore interpretable. Using XAI, it would also be possible to infer what gave the ML application its predictive skill, which could eliminate e.g. contamination from numerical issues that are model resolution specific. Methods constituting equation discovery are an exciting, and potentially powerful, way ML could impact numerical modelling, particularly if IAI/XAI can be applied to ensure the ML application predictive skill is grounded in physics.

4.3. Computational challenges

Since the first ocean general circulation model [44, 45], available computational power has grown exponentially, following Moore's law. The realization that the ocean is fundamentally turbulent and topographically influenced [101, 266] resulted in numerical model development focused on increasing model complexity and refining the model discretization. Numerical model performance is often measured in simulated years per day (SYPD). Computational challenges largely manifest as a balance between preserving the significant legacy present in current ocean modelling codes and harnessing the significant advances within the field of high performance computing, which is often tailored to ML. ML is a trillion dollar industry which is based on high-performance computing power [57]. It is therefore driving developments in modern supercomputing.

The growth of processing speed in supercomputers is no longer exponential, but improvements in the computational efficiency of ocean models are still possible through customization of the computing hardware. ML may likely have a place within a revision of ocean models to improve their computational efficiency. Even within Earth system models as a whole, a 'digital revolution' has been called for [20], where harnessing efficiency in modern hardware is central. Computers can increasingly be customized as hardware is becoming more heterogeneous, meaning that different components for data movement and processing can be combined [21]. Examples of such heterogeneous hardware include the so-called Graphical Processing Units (GPU), Tensor Processing Units, Field-Programmable Gate Arrays, and Application Specific Integrates Circuits, which largely are highly compatible with ML. To take advantage of this

heterogeneous hardware, making current ocean models 'portable', a significant effort would be necessary [21]. Current ocean models use the Fortran programming language and are parallelized to run on many processors via interfaces such as MPI and OpenMP. This parallelization approach is not compatible with hardware accelerators such as GPUs. Compatibility could be achieved via re-writing or enhancement by programming interfaces such as OpenACC or Cuda. Some model groups are investigating a move to newer computing languages, such as Julia (such as the Oceananigans model as part of the CliMA project [203]). Languages like Julia can hide technical details in high-level descriptions of the model code making it more portable. So-called domain-specific languages can be used to facilitate portability [118]. Here, the main algorithmic motives are formulated into library functions that can be ported to different systems with no need to change the model code that is used by the model developer.

ML is expected to play a role in issues associated with the purely computational approach to ocean modelling, beyond devising portability to different hardware accelerators such as GPUs. Hardware accelerators are best suited to problems of high operational intensity (floating-point operations per memory operation). The discretized differential equations governing fluid flow typically result in sparse operations resulting from near-neighbour dependencies ('stencils'). Stencil codes remain notorious for their low operational intensity [15] resulting in poor computational performance, and despite substantial efforts in recent years there has been little progress [12, 188]. This problem is accentuated in oceans, whose long timescales often require $\mathcal{O}(1000\ \mathrm{SYPD})$ for the basic dynamics to emerge. The role of ML in emulating turbulent ocean dynamics is likely to be critical in achieving the level of performance required. This is because resolving key phenomena such as mesoscale eddies remain computationally out of reach, and the current parameterizations such as from Gent and Mc Willians [104] discussed in section 4.1 continue to exhibit deficiencies in simulating meridional eddy transport [98].

ML, and in particular DL, could play a significant role in improve computational efficiency of ocean models due to its ability to work with low numerical precision. Many operations are memory bandwidth bound, and as DL is based on dense linear algebra it is capable of working with very low numerical precision, such as IEEE half precision with 16 bits per variable [150]. The trend towards ML hardware that is optimized for dense linear algebra and low numerical precision may have an impact on future ocean modelling. The use of low numerical precision has been discussed for weather and climate models [79]. The NEMO model [117] was run in single precision with 32 bits per variable instead of the default of double precision with 64 bits per variable [244],

and half precision with 16 bits per variable is being explored for weather and climate models [145] and hardware that is customized for ML has been tested to speed-up expensive components of conventional models [122]. However, in particular for the long-term simulations needed in the ocean, care needs to be taken to make sure that rounding errors do not impact on conservation laws. Certain specific aspects of ocean dynamics require a large dynamic range. For instance, sea level rise, which is a secular change measured in cm/century, must be simulated against a backdrop where surface waves have an amplitude measured in $\mathcal{O}(\mathrm{m})$ and a phase speed of $\mathcal{O}(100\ \mathrm{m\,s^{-1}})$, at least 8 orders of magnitude larger over a typical ocean timestep. For subsequent analysis, it is worth noting that using lower numerical precision would also impact the ability of doing analysis on budgets, as closing these can be complicated when rounding errors are biased.

ML is being explored as a method to emulate computationally costly components of ocean models. This was done successfully in a number of studies [41, 206] for physical parametrization schemes of atmospheric models. For ocean modelling, NN emulators could for example speed-up biogeochemical components [190], which often form a large cost-fraction for ocean models in climate predictions, or sea-ice models, which are often a computational bottleneck as they are difficult to parallelize. ML could also be useful for improving advection schemes and learning local corrections and limiters of fluxes between grid-cells [146]. Furthermore, it may also be possible to improve efficiency of ocean models with semi-implicit timestepping schemes. Here, ML could be used to precondition solver for the large linear problem that needs to be solved in every timestep by estimating the results [3].

The exponential growth of computing power has been accompanied by an exponential growth in data volume. This growth represents a big challenge for operational weather and climate predictions [13]. As data movement is very expensive and a bottleneck in performance, ocean models need to be 'data-centric' and the workflow of the model should be designed in a way that would reduce data movement to a minimum. For example, data is conventionally simply written to discs or tapes after a model simulation, to be retrieved by users afterwards for analysis. A data-centric workflow would process data on-the-fly before it is stored. ML, and in particular unsupervised ML, would be essential in enabling domain scientists to extract the relevant information in such a data-centric workflow. However, such a workflow would also results in additional requirements in terms of the training of staff and the software and hardware infrastructure of weather and climate centres [77]. For example, it is not trivial to incorporate ML tools that are commonly developed using the programming language Python, into atmosphere or ocean

models that are typically written in Fortran. Efforts towards such integration are taking shape, with libraries in development that link NNs into Fortran code [65, 194].

Also of note is the increased difficulty in extracting scientifically interesting information from the vast amounts of data produced by numerical models that *is* stored. The complexity and sheer size of many of these data hinder data dissemination and analysis and severely hamper efforts to analyse the data and address research goals. This emerging class of problems can be illustrated by the Coupled Model Intercomparison Project (CMIP) ensemble now in its sixth phase, which is expected to generate an estimated 40 000 TB of climate model data, a 20-fold increase in data volume from the previous phase [90, 92]. Many variables needed for analysis are effectively unavailable due the difficulty in saving or sharing the data. ML has the capacity to efficiently analyse large datasets as shown in section 2 and 3, but it has also been used to infer, for example, information about subsurface currents [53, 169], eddy heat fluxes [106] and full 3-dimensional dynamics in CMIP6 [228]. ML in many forms has the potential to be highly valuable for researchers interested in the analysis of data that is increasingly large, potentially sparse, and partially unavailable for logistical reasons [91].

4.4. Enforcing physical priors in ML methods

When physical constraints are enforced within ML techniques, this is equivalent to incorporating physical understanding into the applications. Using statistical language, we can describe this process as 'enforcing physical priors'. ML techniques backed by massive datasets have achieved groundbreaking results in vision, speech, and natural language processing, but they have yet to reach the physical oceanography community or largely the physical sciences in general. The ocean is governed by complex phenomena that have been studied by oceanographers for centuries, and taking advantage of this scientific heritage is one way of helping ML techniques reduce the search space of solutions, i.e. by guiding them using physical theories. This research direction is increasingly attracting attention as it helps constrain ML algorithms to be physically plausible and facilitates the interpretation of the results by domain experts. There is a broad spectrum of techniques to supplement ML with physical constraints [263], of which only the most directly relevant are discussed here.

The simplest way to enforce physical priors is through the loss function used to train the ML model. Concretely, this is done by adding an error term related to the physical constraint that needs to be respected, such as a conservation law. For example, if the output field F in a regression problem need to be divergence-free, the term $\|\nabla F\|$ is added to the total loss function to ensure that the divergence of F is close to zero. This approach has its mathematical roots in

the theory of Lagrange multipliers. It can also be seen as a way of doing *regularization*, meaning that finding solutions that generalize well to unseen data is more likely. However, adding physical priors as terms in loss functions comes with a price, which is the problem of weighting different loss terms to impose which ones are most important. The problem of weighting can be solved using cross-validation techniques. With cross-validation, a holdout dataset called a validation dataset is left apart, and the weights of the losses are tuned to achieve the best performance by comparison to this validation dataset. However, such cross-validation techniques can be difficult when the number of constraints is high.

A second strategy that has gained much attention in the recent years is enforcing the constraints directly in the mapping function used for learning. This strategy is best suited to NNs given their flexibility and the rich design choices that enable them to be tailored to specific data. The NN architecture is designed with the physical priors in mind. For example, if we already know that the quantity we want to find is a multiplication of two quantities, then we can encode this inside the neural net by creating two sub-networks whose outputs are multiplied in the last layer [38, 94].

While enforcing physical priors has been a very active area of research in the atmospheric community (see section 1.3), few papers investigating the potential of combining ML and physics can be found in the ocean science literature. In the following we cite some of these examples. Authors in [34] reconstruct subgrid eddy momentum forcing using ConvNets and found that enforcing a constraint on global momentum conservation can best be done by either postprocessing the ConvNet's output or hardcoding a last layer in the ConvNet that removes the spatial mean of the data. [269] proposes to use an equation discovery algorithm, namely Relevance Vector Machines (RVM), for ocean eddy parameterizations. Few attempts have been made to forecast ocean variables using a mix of physical models and DL tools, notably in [28] where authors model an advection diffusion equation in a DL architecture used to forecast SST, while [88] tackle the same problem by combining an autoencoder with ideas from Lyapunov analysis, and [158], where a NN is embedded inside a one-layer quasi-geostrophic numerical model to reduce its bias towards a 3D ocean model.

Enforcing physical priors by solving differential equations with ML techniques is an active research direction that features the development of interesting tools for the ocean community, which are still under-exploited. Physically informed neural networks (PINNs) [202] is a notable example of a technique that leverages the power of NNs to solve differential equations such as the incompressible Navier–Stokes equation [136] without a need for mesh generation, which could accelerate model development. Other recent techniques for learning

ordinary differential equations using either NNs [55] or a combination of NNs and physical-based components [201] are a promising line of research at the interface of NNs and differential equations, which to the best of the authors' knowledge has not yet been applied to ocean modelling.

5. From models to predictions

A basic goal and a test of the understanding of a physical process is the ability to predict its behaviour. Predictions of the weather for several days are a major geoscientific success. Such forecasts have improved with the increasing availability of computational power and observational networks, as well as better algorithms and process understanding [22]. However, predictions of the Earth system on longer timescales are still a major challenge. This is problematic, as predictions often form the basis of decision making. Understanding model error, and combining models with observations, is also at the core of supporting decision makers discussed in section 6.

5.1. Model bias and model error

Bias and error in models are addressed through a systematic process of improvements in our understanding, but the needs of decision support can be immediate. Constraining simulations using observations is the process of data assimilation, covered below in section 5.2. But where errors are recalcitrant, oceanographers and applied scientists in general use methods of 'artificial' error reduction, driven by comparisons against data. An early example of correcting for a bias related to the ocean's role in climate is the use of 'q-flux adjustments', or simply flux adjustments. Here, the issue was a persistent error in the evaporative flux from the ocean surface. The adjustment to ameliorate this bias was a correction to restore energy balance to the coupled system by artificially adding a compensation term [168]. This adjustment method fell into disfavour owing to its blatant 'fudge factor' nature [223], although recent studies indicate that 'flux adjusted' models continue to exhibit greater predictive skill [253].

When assessing a prediction from a model, the accuracy of the output can be assessed by comparing to a 'truth' benchmark. Such a benchmark can for example be from observations or a target model representation of the system. Observations, although mostly not complete, constitute a best guess. This process can also identify 'structural error', also mentioned in section 1.3, indicating that the model formulation itself is incorrect. Compared to observations, model outputs can show differences that cannot only be attributed to differences in initial conditions, but instead reflect errors within the model itself discussed in section 5.2 below. Some of these errors can be explained by unresolved scales in the discretized version of numerical models, but model errors can also originate from incomplete physical knowledge. For example, within a sub-gridscale parameterization the exact physics that need to be represented may be unclear, as discussed in section 4. Incomplete physical knowledge also impacts uncertainties in the parameters used, for example in the coupling terms between model components. Within model error as a whole, there may be a systematic component, which is referred to as model bias.

For post-processing of model output, statistical methods, related to ML, have been used to correct biases (for example [151, 222] or flux adjustments). Bias correction methods are used frequently in operational weather predictions with DL playing an ever-increasing role [14, 113, 205]. However, using downscaling as described in section 1.3, ML can also be used to relate model output with local information, such as the local topography at very high resolution or observations that are available, to improve predictions when model simulations have already finished. Called up-scaling within the ML community, some of the mapping procedures used for downscaling, such as GANs, even allow for uncertainty quantification [156]. Within climate models, the LRP XAI method have successfully been used to identify key model biases for certain prediction tasks [16], with potential for application to the ocean. However, the LRP method application is still in its infancy.

5.2. Ocean data assimilation

5.2.1. Data assimilation methods: a brief history and main assumptions

Data assimilation (DA) is the process of constraining a theoretical representation of a system, usually using a numerical or statistical model, using a collection of observations. The results of this process typically include optimized estimates of (1) the time-evolving state of the system (sometimes called the 'trajectory'), (2) initial conditions, (3) boundary conditions, and (4) other intrinsic model parameters (e.g. mixing coefficients). The optimization process typically consists of correcting the values of the initial conditions, boundary conditions, and model parameters in order to minimize a selected model-data misfit metric. To use the language of the theory of differential equations, one may think of DA as a set of methods for rigorously identifying which solution, among the family of solutions to a system of differential equations, best satisfies the given constraints.

Although there is a long history of DA in numerical weather prediction stretching across much of the 20th century, oceanographic DA only began in the late 1980s. The first experiments were regional [211], followed a few years later by the ambitious World Ocean Circulation Experiment (WOCE, [267]), and a community was subsequently assembled under the Global Ocean Data Assimilation Experiment (GODAE, [23]). These first DA approaches used in weather and ocean prediction were directly derived

using optimal interpolation [107] and were based on strong assumptions, namely that the evolution model is linear and perfect and that the data error distribution is unbiased and well-represented by a Gaussian. In time, DA algorithms evolved to relax some of these assumptions, extending the scope of DA applications to the ocean.

The developments within DA have led to two main sets of techniques. These are ensemble approaches, of which the ensemble Kalman filter (EnKF) is a standard example, and variational approaches such as four dimensional variational assimilation (4DVar). Both classes of methods conceptually represent the abstract trajectory of the target system as a probability distribution across possible trajectories. EnKF constructs an ensemble of forecast states such that the ensemble mean and the sample covariance are expected to be the best estimates. A core assumption is that the ensemble probability distribution can be well-represented by a Gaussian function [89]. The 4DVar method uses a linear model to calculate which perturbations to the initial conditions, boundary conditions, and parameters tend to increase the agreement between the time-evolving state of the model and the observational constraints [62].

Each of the DA classes of methods are used in their various flavours for both global or regional studies [97, 157, 173, 189, 216, 254]. DA is used routinely both in operational forecast and reanalysis mode. DA is used in the framework of several national and international projects. In no particular order, examples include the ECCO[10] project, ECMWF[11] or the NOAA NCEP[12] Global Ocean Data Assimilation System (GODAS) in the USA.

In idealized comparisons between the two classes of methods, EnKF produces more accurate estimates for shorter assimilation windows, whereas 4DVar produces more accurate estimates when data constraints are sparse. For ocean applications, data is often sparse, making 4DVar attractive [141]. In practice, different DA approaches derived from optimal interpolation, 3DVar, the EnKF, or 4DVar are used [64]. The type depends on the application (e.g. short-term forecast or climate application), the available computing resources, the type of observations that are assimilated, and the historical expertise in each group.

5.2.2. Model errors and ML within data assimilation
Historically, DA techniques mainly focus on the estimation of the state of the system, but the estimation of model error in the DA process is increasingly

important [49]. Several approaches that are used to handle model error apply DA frameworks that can be considered ML approaches [59, 239]. The estimation of model errors is particularly important if DA is being used to calculate forecasts over long timescales, i.e. from sub-seasonal to decadal scales. This is of particular importance for ocean forecasts, where timescales are longer than in the atmosphere; DA has been shown to be effective in this context [257].

5.2.3. Data assimilation and ML
Several studies have highlighted the connection between DA and ML [1, 32, 37, 102]. The connection is more direct with 4DVar, in which a function that quantifies model-data disagreement (i.e. a 'cost function') is minimized using a gradient descent algorithm, wherein an adjoint model calculates the gradient. In this perspective, 4DVar is approximately equivalent to the process of training of a neural network for regression. This is because the adjoint model can be seen as equivalent to the gradient backpropagation process [132, 148].

There are several ways ML can be used in combination with a DA framework. First, a data-driven model can be used to emulate a numerical model, partially or totally to provide the forecast. The objective is then to correct the model error, or to decrease the computational cost [159]. Note that emulation could become instrumental, since DA methods increasingly rely on ensemble runs, which are costly [48]. As DA allows one to bring the model and observations close enough together to represent the same physical situation, DA can in principle be used to extend the learning of parametrization to the learning of improved models directly from observations [35, 39], described further in the section 4. It is still unclear whether observations are too sparse for this approach to be successful within ocean modelling. This is particularly the case, because the time period where dense observations are available is relatively short, compared to the long timescales that are known to be important for ocean dynamics. Another benefit of using an ML emulators arises because most ML tools, such as NNs, are easy to differentiate. Given the structure of NN (interconnected simple operators), and the libraries used to implement them, the computation of the gradient of the NN model is straightforward and efficient. This means that the computation can be used to efficiently develop tangent linear and adjoint model code, which is required for DA methods such as 4DVar [123]. This is noteworthy, because traditionally the development of tangent linear and adjoint models has required major efforts from the research community, either by manually coding an adjoint or by the semi-automatic process of algorithmic differentiation (e.g. [109]).

---

[10] Estimating the Circulation and Climate of the Ocean.
[11] The European Centre for Medium-Range Weather Forecasts.
[12] National Oceanic and Atmospheric Administration, National Centers for Environmental Prediction.

Second, ML can be instrumental in strongly coupled DA. Strongly coupled DA consists of correcting a coupled system (e.g. ocean-atmosphere) in a unified way. This allows, for example, atmospheric observations to constrain the ocean state and vice versa, which is not the case in uncoupled DA, where only ocean observations are used to constrain the ocean system. Strongly coupled DA is expected to be efficient but challenging due to the high variety of temporal and spatial scales [198]. In this sense, ML can be used to relax some strong assumptions of the DA algorithm (e.g. the assumption that the errors follow a Gaussian distribution), or to isolate relevant scales in observational and model states. Traditional DA algorithms correct the state of the model either directly, or by using a linear transformation. ML could extend this approach by applying a non-linear transformation to the space before computing the DA correction. Some examples of this approach have been developed [7, 93, 114, 166], but so far none of them have been applied to realistic ocean DA setups.

Finally, ML can help deal with the mass of available observations. In section 2, we discussed how ML can help derive new type of products from observations. These new products are good candidates for inclusion in a DA system. ML can also be used to provide more accurate and/or faster observation operators, for example to emulate satellite observations [35].

6. Discussion: towards a new synthesis of observations, theory, modelling, and prediction in ocean sciences using ML

6.1. The need for transparent ML

To increase confidence in the use of ML, stepping out of the 'black box' is advisable. Towards this, having ML methods be transparent is very important. A transparent ML application is one where source of skill is known, or put differently why the ML came to its conclusion. Possibly the largest hurdles for ML adoption are a lack of trust and the difficulty of generalization. These two are linked, and if generalization is not reached trust is certainly not merited for ML applications within oceanography. Generalization refers to a model's ability to properly adapt to new, previously unseen data. Within oceanography and beyond, the ideal generalization would come from the ML application learning the underlying physics. With a lack of good data coverage, the possible underspecification [67] and shortcut learning [103] are important to keep in mind, where a model can seemingly perform well for example in the current climate but will fail in a future scenario as something physical was not learned. These issues are ubiquitous and not unique to oceanography or to Earth science, with a call for 'physics informed' ML

[207]. Accordingly, the field of ML already has, and is, developing methods to address these issues such as 'few shot learning' and 'transfer learning'. Here, for 'few shot learning' the ML models are developed to deal with very small amounts of training data [258]. Using 'transfer learning' refers to where NN are trained on general tasks, but are used and adapted to specific applications without the need to retrain from scratch. If it is possible to reliably quantify and account for the uncertainties associated with an ML application during training, this could also increase confidence in the model, but assessing the reliability may face challenges similar to underspecification. Recent works show promising progress toward this by learning a probability distribution of outcomes that can be stochastically sampled [81, 115]. For uncertainty quantification, the uncertainty could be determined during training with ML, likely increasing the reliability of the results. Other methods such as regularization, invariances, dimensionality reductions are also a powerful tool to increase the generalization skill. For climate applications, a key issue when training ML applications is that the system they are being trained on is largely non-stationary. This complicates the problem of generalization even further, but ML methods have demonstrated that having good generalization skills in a non-stationary context is possible [196]. Increasingly, the ML community is suggesting a focus on using IAI [214], driven among other things by the consistent racial and gender bias revealed in DL applications. With the ability to interpret the ML model itself, and intuitively discern if it is meaningful, the danger of introducing such bias is likely reduced dramatically. Similarly, XAI methods for example for NN, that retrospectively explain the source of ML predictive skill, can also help inspire confidence [84, 174, 246]. XAI methods such as layerwise relevance propagation (LRP [10, 193]) have been gaining traction within the atmosphere [16, 17, 40, 246], and ocean, but making their application explicitly appropriate to oceanography, and indeed the physical sciences in general, may require targeted method development.

6.2. Decision support

There is a need for accurate and actionable information about the ocean for a wide range of decision making. As noted above in section 5.1, the need for actionable predictions and decision support can short-circuit the scientific process of error elimination. This is because the information may have 'customers'/users with an immediate need: for example decisions on shorelines ranging from building seawalls, issuing housing permits, to setting insurance premiums. ML may play a role in bridging the gap between what model-based predictions are able to provide, and what users wish to know. The role of data-driven methods could be particularly important

for filling in the gaps where theory and models underspecify the system, potentially leaving considerable uncertainty as noted in section 5.2.

The reliable quantification of uncertainties is often essential to support decision making. However, uncertainty quantification is often difficult for conventional approaches used in ocean science. This is because model errors cannot be described by physical equations or physical reasoning in most applications, and errors are often noisy and non-linear. On the other hand, model error can often be diagnosed against a reference truth, such as observations or target model simulations. Therefore, ML can be useful for the quantification of uncertainties. In particular as datasets from different data sources, on different reference grids and for different variables can be fused and compared using ML techniques. For example, ML can be used to post-process ensemble simulations [222], and Bayesian ML techniques can also be used to learn the uncertainty quantification together with the ML tool. In addition, targeted loss function design could help target, and lessen areas of uncertainty important for specific decisions. In an ML context, the models can be calibrated or tuned toward a particular loss function [129]. A loss function could be designed to capture physical constraints, such as the closure of the energy budget in climate simulations. However, loss functions can also be specifically designed for use cases in decision support. An open area of research remains in relating the results that may be obtained from different calibrations (loss functions) of the same system trained on the same data.

ML can help to map model data and observations to predict or detect events for which we cannot provide a useful physical representation of the interactions. This could, for example, be a mapping from observational data of a time series in a specific location or observations from a buoy, to large-scale model data with the goal of making customized predictions of surface waves and local wind. Such data could for example be used for a sailing competition. Such tools based on ML could become essential for decision support, for example when used to predict sea levels [255]. ML based mapping tools could also be useful to inform where more observational data is needed, for example when deciding where to sample on a cruise or where to send autonomous platforms. To date, satellite images are largely used, but added guidance from ML techniques could be very valuable, particularly if sub-surface observations are the target [52, 169, 228]. ML may eventually be used to support observational campaigns in near-real time by interactively connecting networks of non-autonomous and autonomous observing platforms (e.g. gliders) to decision planning systems. These systems can take environmental conditions,

target observations, and task scheduling into account. The vision is to have a 'cyberinfrastructure' that can maximize the spatiotemporal coverage of the observations without a specific need for human intervention. The potential use of such observational planning and adjustment systems is being explored by international initiatives such as the Southern Ocean Observing System (SOOS, www.soos.aq/). Similarly, for planning legislation, having knowledge of what is within a nations marine area and how this may connect to the surrounding ocean can be very valuable. Here ML has been used to provide actionable information [227], as the ocean does not adhere to borders drawn by humans.

Next to DL methods, the calibration of parameters is very important as many within atmosphere and ocean models cannot be validated within their physical uncertainty range. Because they cannot be validated, the parameters need to be tuned [58, 247]. Given this physical uncertainty, using ML and DL in particular will likely be very valuable as noted in section 4. If successful, such breakthroughs could help inform a wide range of decisions including those based on climate models such as CMIP, or in a more general sense. This is particularly the case for longer timescale integrations from seasonal and onward, due to the longer timescale active within the ocean.

An important component of supporting decision makers is communication. The ability to communicate effectively between the people that are making decisions and oceanographers can pose a problem. Oceanographers would need to be aware of what *is* useful information, and how to provide this. Decision makers largely may not have intimate knowledge of what available tools are capable of addressing, but mainly knowledge of the problem at hand. While seeming trivial, improving this line of communication is an important component of increasing the utility of oceanographic work.

6.3. Challenges and opportunities

In this review, we have highlighted some of the many challenges within observational, computational, and theoretical oceanography where ML offers an exciting opportunity to improve the speed and efficiency of conventional work and also to explore completely new avenues. As a merger of two distinct fields, there is ample opportunity to incorporate powerful, established ML methods that are largely new to oceanography as a field. While not without risk, the potential benefits of ML methods is creating increasing interest in the field. This review has presented some of the challenges and opportunities involved when leveraging ML techniques to improve the modelling, observing, fundamental understanding, and prediction of the ocean system.

ML applications fundamentally rely on the data available for learning, and here the ocean presents a unique challenge for ML applications. The important timescales in the ocean range from seconds to millennia, with strong interactions between processes across those scales. For example, a wind gust can trigger a phytoplankton bloom. Observations are largely sparse, noisy, and unbalanced. Temporally, very few long-timescale observations exist that span more than a few decades. A general problem with models of the ocean, either ML derived or more conventional, is that the system is highly non-stationary. With climate change, the mean state and its variance are liable to change, and a model that is trained from today's data may not be general enough to accurately represent an ocean in a warmer climate. Other components of the Earth system such as land or atmospheric models, or GFD in general, also face similar challenges, but they are exacerbated within oceanography due to the lack of spatial and temporal observational coverage.

ML offers many avenues with which the challenges listed above could be tackled. For example, with instantaneous processes (such as radiative transfer) or small spatial scale problems (for example eddy detection), a cross-validation approach with an associated independent test dataset could be fruitful. Indeed, cross-validation is widely advisable. On longer timescales, methods related to physical constraints would likely offer better results. Hybrid approaches for combining physics-driven models and ML models are becoming increasingly useful to aid the development of ocean models and to increase their computational efficiency on HPC platforms. Such 'Neural Earth System Models' (NESYM [134]) can, for example, use ML for parameterization of sub-gridscale processes. Pairings of ML and conventional methods also show great promise for improving signal-to-noise ratios during training while also anchoring ML learning to a stronger physical foundation [260].

Both the field of oceanography and ML are quickly evolving, and the computational tools available to implement ML techniques are also becoming increasingly accessible. With ample enthusiasm for ML applications to address oceanographic problems, it is also important to keep in mind that ML as a field is largely not concerned with the physical sciences. Approaching ML applications with caution and care is necessary to ensure meaningful results. The importance of increasing trust in ML methods also highlights a need for collaboration between oceanographers and ML domain experts. ML as a field is developing very swiftly, and promoting collaboration can help develop methods that are tailored to also suit the needs of oceanographers.

This review has outlined the recent advances and some remaining challenges associated with ML adoption within oceanography. As with any promising new set of methods, while there is ample opportunity, it is also worth noting that ML adoption also comes with risk. However, exploring the full potential and charting the limits of ML within oceanography is crucial and deserves considerable attention from the research community.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

Appendix

Table A1. List of acronyms.

Abbreviation	Description
4DVar	4-dimensional variational assimilation
AE	autoencoder
AI	Artificial intelligence
AIC	Akaike information criterion
BIC	Bayesian information criterion
ConvNet	convolutional neural network
DBSCAN	density-based spatial clustering of applications with noise
DA	data assimilation
DL	deep learning
DNN	deep neural network
ECCO	estimating the circulation and climate of the ocean
EnKF	ensemble Kalman filter
EOF	empirical orthogonal functions
GAN	generative adversarial network
GFD	geophysical fluid dynamics
GMM	Gaussian mixture model
GODAE	global ocean data assimilation experiment
GODAS	global ocean data assimilation system
GPR	Gaussian process regression
GPU	graphical processing units (GPU)
HPC	high performance computing
IAI	interpretable artificial intelligence
KNN	K nearest neighbours
LR	linear regression
MAE	mean absolute error
ML	machine learning
MLP	multi-layer perceptron
MSE	mean square error
NESYM	neural earth system models
NN	neural networks
PCA	principal component analysis
PINN	physics informed neural networks
RF	random forest
RL	reinforcement learning
RNN	recurrent neural network
RVM	relevance vector machines
SGD	stochastic gradient descent
SOM	self organizing maps
SVM	support vector machines
SVR	support vector regression
t-SNE	t-distributed stochastic neighbour embedding
UMAP	uniform manifold approximation and projection
VAE	variational autoencoder
XAI	explainable artificial intelligence
WOCE	world ocean circulation experiment
WWII	World War II

ORCID iDs

Maike Sonnewald ⓘ https://orcid.org/0000-0002-5136-9604

Redouane Lguensat ⓘ https://orcid.org/0000-0003-0226-9057

Daniel C Jones ⓘ https://orcid.org/0000-0002-7408-6690

Peter D Dueben ⓘ https://orcid.org/0000-0002-4610-3326

Julien Brajard ⓘ https://orcid.org/0000-0003-0634-1482

V Balaji ⓘ https://orcid.org/0000-0001-7561-5438

References

[1] Abarbanel H D I, Rozdeba P J and Shirman S 2018 Machine learning: deepest learning as statistical data assimilation problems *Neural Comput.* **30** 2025–55

[2] Abbe C 1901 The physical basis of long-range weather forecasts *Mon. Weather Rev.* **29** 551–61

[3] Ackmann J, Dueben P D, Palmer T N and Smolarkiewicz P K 2020 Machine-learned preconditioners for linear solvers in geophysical fluid flows (arXiv:2010.02866)

[4] Adewoyin R, Dueben P, Watson P, He Y and Dutta R 2021 TRU-NET: a deep learning approach to high resolution prediction of rainfall (arXiv:2008.09090)

[5] Agarwal N, Kondrashov D, Dueben P, Ryzhov E and Berloff P 2021 A comparison of data-driven approaches to build low-dimensional ocean models *J. Adv. Modeling Earth Syst.* submitted

[6] Alvera-Azcárate A, Barth A, Parard Gelle and Beckers J-M 2016 Analysis of SMOS sea surface salinity data using DINEOF *Remote Sens. Environ.* **180** 137–45

[7] Amendola M, Arcucci R, Mottet L, Casas C Q, Fan S, Pain C, Linden P and Guo Y-K 2020 Data assimilation in the latent space of a neural network (arXiv:2012.12056)

[8] Andersson T R *et al* Seasonal arctic sea ice forecasting with probabilistic deep learning 2021 *Earth Arxiv*

[9] Aoki S, Yamazaki K, Hirano D, Katsumata K, Shimada K, Kitade Y, Sasaki H and Murase H 2020 Reversal of freshening trend of Antarctic bottom water in the Australian-Antarctic basin during 2010s *Sci. Rep.* **10** 14415

[10] Bach S, Binder A, Montavon G, Klauschen F, Müller K-R and Samek W 2015 On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation *PLoS One* **10** 1–46

[11] Balaji V 2021 Climbing down Charney's ladder: machine learning and the post-Dennard era of computational climate science *Phil. Trans. R. Soc.* A **379** 20200085

[12] Balaji V *et al* 2017 CPMIP: measurements of real computational performance of Earth system models in CMIP6 *Geosci. Model Dev.* **10** 19–34

[13] Balaji V *et al* 2018 Requirements for a global data infrastructure in support of CMIP6 *Geosci. Model Dev.* **11** 3659–80

[14] Baran A, Lerch S, Ayari M E and Baran S 2020 Machine learning for total cloud cover prediction *Neural Comput. Appl.* **33** 2605–20

[15] Barba L A and Yokota R 2013 How will the fast multipole method fare in the exascale era *SIAM News* **46** 1–3

[16] Barnes E A, Mayer K, Toms B, Martin Z and Gordon E 2020 Identifying opportunities for skillful weather prediction with interpretable neural networks (arXiv:2012.07830)

[17] Barnes E A, Toms B, Hurrell J W, Ebert-Uphoff I, Anderson C and Anderson D 2020 Indicator patterns of forced change learned by an artificial neural network *J. Adv. Modeling Earth Syst.* **12** e2020MS002195

[18] Barnett T P and Preisendorfer R 1987 Origins and levels of monthly and seasonal forecast skill for united states surface air temperatures determined by canonical correlation analysis *Mon. Weather Rev.* **115** 1825–50

[19] Barth A, Alvera-Azcárate A, Licer M and Beckers J-M 2020 DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations *Geosci. Model Dev.* **13** 1609–22

[20] Bauer P, Dueben P D, Hoefler T, Quintino T, Schulthess T C and Wedi N P 2021 The digital revolution of earth-system science *Nat. Comput. Sci.* **1** 104–13

[21] Bauer P *et al* 2020 The ECMWF scalability programme: progress and plans (available at: www.ecmwf.int/node/19380)

[22] Bauer P, Thorpe A and Brunet G 2015 The quiet revolution of numerical weather prediction *Nature* **525** 47–55

[23] Bell M J, Lefèbvre M, Le Traon P-Y, Smith N and Wilmer-Becker K 2009 GODAE: The global ocean data assimilation experiment *Oceanography* **22** 14–21

[24] Benestad R E, Hanssen-Bauer I and Chen D 2008 *Empirical-Statistical Downscaling* (Singapore: World Scientific)

[25] Bengio Y, Goodfellow I and Courville A 2017 *Deep Learning* vol 1 (Cambridge, MA: MIT Press)

[26] Beucler T, Ebert-Uphoff I, Rasp S, Pritchard M and Gentine P 2021 Machine learning for clouds and climate (invited chapter for the agu geophysical monograph series "clouds and climate") *Earth Space Sci. Open Arch.* 27

[27] Beucler T, Pritchard M, Rasp S, Ott J, Baldi P and Gentine P 2021 Enforcing analytic constraints in neural networks emulating physical systems *Phys. Rev. Lett.* **126** 098302

[28] de Bezenac E, Pajot A and Gallinari P 2019 Deep learning for physical processes: incorporating prior scientific knowledge *J. Stat. Mech.: Theory Exp.* **2019** 124009

[29] Biau Gerard and Scornet E 2016 A random forest guided tour *Test* **25** 197–227

[30] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Berlin: Springer)

[31] Bjerknes V 1904 Das problem der wettervorhers-age, betrachtet vom standpunkte der mechanik und der physik *Meteor. Z.* **21** 1–7

[32] Bocquet M, Brajard J, Carrassi A and Bertino L 2019 Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models *Nonlinear Process. Geophys.* **26** 143–62

[33] Boehme L and Rosso I 2021 Classifying oceanographic structures in the Amundsen sea, Antarctica *Geophys. Res. Lett.* **48** e2020GL089412

[34] Bolton T and Zanna L 2019 Applications of deep learning to ocean data inference and subgrid parameterization *J. Adv. Modeling Earth Syst.* **11** 376–99

[35] Bonavita M and Laloyaux P 2020 Machine learning for model error inference and correction *J. Adv. Modeling Earth Syst.* **12** e2020MS002232

[36] Bony S, Stevens B, Held I H, Mitchell J F, Dufresne J-L, Emanuel K A, Friedlingstein P, Griffies S and Senior C 2013 Carbon dioxide and climate: perspectives on a scientific assessment *Climate Science for Serving Society* (Berlin: Springer) pp 391–413

[37] Brajard J, Carrassi A, Bocquet M and Bertino L 2019 Connections between data assimilation and machine learning to emulate a numerical model *Proc. 9th Int. Workshop on Climate Informatics* (NCAR)

[38] Brajard J, Carrassi A, Bocquet M and Bertino L 2020 Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model *J. Comput. Sci.* **44** 101171

[39] Brajard J, Carrassi A, Bocquet M and Bertino L 2021 Combining data assimilation and machine learning to infer unresolved scale parametrisation *Phil. Trans. R. Soc.* A **379** 20200086

[40] Brenowitz N D, Beucler T, Pritchard M and Bretherton C S 2020 Interpreting and stabilizing machine-learning parametrizations of convection *J. Atmos. Sci.* **77** 4357–75

[41] Brenowitz N D and Bretherton C S 2018 Prognostic validation of a neural network unified physics parameterization *Geophys. Res. Lett.* **45** 6289–98

[42] Bretherton C S, Smith C and Wallace J M 1992 An intercomparison of methods for finding coupled patterns in climate data *J. Clim.* **5** 541–60

[43] Brunton S L, Proctor J L and Kutz J N 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems *Proc. Natl Acad. Sci.* **113** 3932–7

[44] Bryan K 1997 A numerical method for the study of the circulation of the world ocean *J. Comput. Phys.* **135** 154–69

[45] Bryan K and Cox M D 1968 A nonlinear model of an ocean driven by wind and differential heating. Part I. Description of the three-dimensional velocity and density fields *J. Atmos. Sci.* **25** 945–67

[46] Bushinsky S, Landschutzer P, Rödenbeck C, Gray A, Baker D, Mazloff M, Resplandy L, Johnson K and Sarmiento J 2019 Reassessing southern ocean air-sea CO_2 flux estimates with the addition of biogeochemical float observations *Glob. Biogeochem. Cycles* **33** 1370–88

[47] Callaham J L, Koch J V, Brunton B W, Kutz J N and Brunton S L 2021 Learning dominant physical processes with data-driven balance models *Nat. Commun.* **12** 1016

[48] Carrassi A, Bocquet M, Bertino L and Evensen G 2018 Data assimilation in the geosciences: an overview of methods, issues and perspectives *Wiley Interdiscip. Rev.: Clim. Change* **9** e535

[49] Carrassi A and Vannitsem Sephane 2010 Accounting for model error in variational data assimilation: a deterministic formulation *Mon. Weather Rev.* **138** 3369–86

[50] Cartwright D 2007 On the origins of knowledge of the sea tides from antiquity to the thirteenth century *Earth Sci. Hist.* **20** 105–26

[51] Castellani M 2006 Identification of eddies from sea surface temperature maps with neural networks *Int. J. Remote Sens.* **27** 1601–18

[52] Chapman C and Charantonis A A 2017 Reconstruction of subsurface velocities from satellite observations using iterative self-organizing maps *IEEE Geosci. Remote Sens. Lett.* **14** 617–20

[53] Chapman C and Charantonis A A 2017 Reconstruction of subsurface velocities from satellite observations using iterative self-organizing maps *IEEE Geosci. Remote Sens. Lett.* **14** 617–20

[54] Chapman C, Lea M-A, Meyer A, Sallée J-baptiste and Hindell M 2020 Defining southern ocean fronts and their influence on biological and physical processes in a changing climate *Nat. Clim. Change* **10** 1–11

[55] Chen R T Q, Rubanova Y, Bettencourt J and Duvenaud D 2018 Neural ordinary differential equations (arXiv:1806.07366)

[56] Choi M, De Silva L and Yamaguchi H 2019 Artificial neural network for the short-term prediction of arctic sea ice concentration *Remote Sens.* **11** 1071

[57] Chui M, Manyika J, Miremadi M, Henke N, Chung R, Nel P and Malhotra S 2018 Notes from the ai frontier insights from hundreds of use cases *Discussion Paper* (available at: www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning)

[58] Cleary E, Garbuno-Inigo A, Lan S, Schneider T and Stuart A M 2021 Calibrate, emulate, sample *J. Comput. Phys.* **424** 109716

[59] Cocucci T J, Pulido M, Lucini M and Tandeo P 2021 Model error covariance estimation in particle and ensemble Kalman filters using an online expectation–maximization algorithm *Q. J. R. Meteorol. Soc.* **147** 526–43

[60] Cooper F C and Zanna L 2015 Optimisation of an idealised ocean model, stochastic parameterisation of sub-grid eddies *Ocean Modelling* **88** 38–53

[61] Cortes C and Vapnik V 1995 Support-vector networks *Mach. Learn.* **20** 273–97

[62] Courtier P, Thépaut J-N and Hollingsworth A 1994 A strategy for operational implementation of 4D-Var, using an incremental approach *Q. J. R. Meteorol. Soc.* **120** 1367–87

[63] Couvreux F *et al* 2020 Process-based climate model development harnessing machine learning. I. A calibration tool for parameterization improvement *J. Adv. Modeling Earth Syst.* **13** e2020MS002217

[64] Cummings J *et al* 2009 Ocean data assimilation systems for godae *Oceanography* **22** 96–109

[65] Curcic M 2019 A parallel fortran framework for neural networks and deep learning *ACM SIGPLAN Fortran Forum* **38** 4–21

[66] Cybenko G 1989 Approximation by superpositions of a sigmoidal function *Math. Control Signals Syst.* **2** 303–14

[67] D'Amour A *et al* 2020 Underspecification presents challenges for credibility in modern machine learning (arXiv:2011.03395)

[68] Davis M 2001 *Late Victorian Holocausts* (United Kingdom: Verso)

[69] Deacon M 2018 *Scientists and the Sea, 1650–1900: A Study of Marine Science* (Routledge)

[70] Denvil-Sommer A, Gehlen M, Vrac M and Mejia C 2019 LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean pCO$_2$ over the global ocean *Geosci. Model Dev.* **12** 2091–105

[71] Desbruyères D, Chafik L and Maze G 2021 A shift in the ocean circulation has warmed the subpolar north atlantic ocean since 2016 *Commun. Earth Environ.* **2** 48

[72] Ding C and He X 2004 k-means clustering via principal component analysis *Proc. Twenty-First Int. Conf. Machine Learning, ICML'04 (New York, NY)* (Association for Computing Machinery) p 29

[73] Ding H, Newman M, Alexander M A and Wittenberg A T 2019 Diagnosing secular variations in retrospective ENSO seasonal forecast skill using CMIP5 model-analogs *Geophys. Res. Lett.* **46** 1721–30

[74] Dixon K W, Lanzante J R, Nath M J, Hayhoe K, Stoner A, Radhakrishnan A, Balaji V and Gaitán C F 2016 Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Clim. Change* **135** 395–408

[75] Dong S, Sprintall J, Gille S T and Talley L 2008 Southern ocean mixed-layer depth from argo float profiles *J. Geophys. Res.: Oceans* **113** C6

[76] Dool H M V D 1994 Searching for analogues, how long must we wait? *Tellus* A **46** 314–24

[77] Düben P *et al* 2021 Machine learning at ECMWF: a roadmap for the next 10 years (available at: www.ecmwf.int/node/19877)

[78] Dueben P D and Bauer P 2018 Challenges and design choices for global weather and climate models based on machine learning *Geosci. Model Dev.* **11** 3999–4009

[79] Dueben P D and Palmer T N 2014 Benchmark tests for numerical weather forecasts on inexact hardware *Mon. Weather Rev.* **142** 3809–29

[80] Duhem P M M 1906 *La ThÉorie Physique: Son Objet Et Sa Structure* (Chevalier & Rivière)

[81] Dunbar O R A, Garbuno-Inigo A, Schneider T and Stuart A M 2020 Calibration and uncertainty quantification of convective parameters in an idealized GCM (arXiv:2012.13262 [math, stat])

[82] Duncan D I, Eriksson P, Pfreundschuh S, Klepp C and Jones D C 2019 On the distinctiveness of observed oceanic raindrop distributions *Atmos. Chem. Phys.* **19** 6969–84

[83] Durran D 2010 *Numerical Methods for Fluid Dynamics with Applications to Geophysics* vol 32 (Berlin: Springer)

[84] Ebert-Uphoff I and Hilburn K 2020 Evaluation, tuning and interpretation of neural networks for working with images in meteorological applications *Bull. Am. Meteorol. Soc.* **101** 1–47

[85] Eden C and Olbers D 2010 Why western boundary currents are diffusive: a link between bottom pressure torque and bolus velocity *Ocean Modelling* **32** 14–24

[86] Edwards P 2010 *A Vast Machine: Computer Models, Climate Data and the Politics of Global Warming* (Cambridge, MA: MIT Press)

[87] Ekman V W 1905 On the influence of the earth's rotation on ocean currents *Arch. Math. Astron. Phys.* **2**

[88] Benjamin Erichson N, Muehlebach M and Mahoney M W 2019 Physics-informed autoencoders for Lyapunov-stable fluid flow prediction (arXiv:1905.10866)

[89] Evensen G 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics *J. Geophys. Res.: Oceans* **99** 10143–62

[90] Eyring V, Bony S, Meehl G, Senior C, Stevens B, Ronald S and Taylor K 2015 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organisation *Geosci. Model Dev.* **8** 10539–83

[91] Eyring V *et al* 2019 Taking climate model evaluation to the next level *Nat. Clim. Change* **9** 102–10

[92] Eyring V *et al* 2016 Towards improved and more routine Earth system model evaluation in CMIP *Earth Syst. Dyn.* **7** 813–30

[93] Fablet R, Chapron B, Drumetz L, Etienne Mémin, Pannekoucke O and Rousseau Fçois 2020 Learning

variational data assimilation models and solvers (arXiv:2007.12941)

[94] Fablet R, Ouala S and Herzet C 2017 Bilinear residual neural network for the identification and forecasting of dynamical systems (arXiv:1712.07003)

[95] Ferrari R, McWilliams J C, Canuto V M and Dubovikov M 2008 Parameterization of eddy fluxes near oceanic boundaries *J. Clim.* **21** 2770–89

[96] Foote E 1856 Circumstances affecting the heat of the sun's rays *Am. J. Sci. Arts* **XXII** 357–9

[97] Forget G, Campin J M, Heimbach P, Hill C N, Ponte R M and Wunsch C 2015 ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation *Geosci. Model Dev.* **8** 3071–104

[98] Fox-Kemper B, Bachman S, Pearson B and Reckinger S 2014 Principles and advances in subgrid modelling for eddy-rich simulations *CLIVAR Exchanges* **19** 42–46

[99] Furner R, Haynes P, Munday D, Paige B, Jones D C and Shuckburgh E 2021 Sensitivity analysis of a data-driven model of ocean temperature *Geosci. Model Dev.*

[100] Gaitan C F, Balaji V and Moore B III 2016 Can we obtain viable alternatives to Manning's equation using genetic programming? *Artif. Intell. Res.* **5** 92

[101] Garabato A 2012 A perspective on the future of physical oceanography *Phil. Trans. R. Soc.* A **370** 5480–511

[102] Geer A J 2021 Learning earth system models from observations: machine learning or data assimilation? *Phil. Trans. R. Soc.* A **379** 20200089

[103] Geirhos R, Jacobsen J-H, Michaelis C, Zemel R, Brendel W, Bethge M and Wichmann F A 2020 Shortcut learning in deep neural networks *Nat. Mach. Intell.* **2** 665–73

[104] Gent P R and Mcwilliams J C 1990 Isopycnal mixing in ocean circulation models *J. Phys. Oceanogr.* **20** 150–5

[105] Gentine P, Pritchard M, Rasp S, Reinaudi G and Yacalis G 2018 Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.* **45** 5742–51

[106] George T, Manucharyan G and Thompson A 2019 Deep learning to infer eddy heat fluxes from sea surface height patterns of mesoscale turbulence *Earth ArXiv*

[107] Ghil M and Malanotte-Rizzoli P 1991 Data assimilation in meteorology and oceanography *Adv. Geophys.* **33** 141–266

[108] Ghosh S and Mujumdar P P 2008 Statistical downscaling of GCM simulations to streamflow using relevance vector machine *Adv. Water Resour.* **31** 132–46

[109] Giering R and Kaminski T 1998 Recipes for adjoint code construction *ACM Trans. Math. Softw.* **24** 437–74

[110] Giglio D, Lyubchich V and Mazloff M R 2018 Estimating oxygen in the southern ocean using argo temperature and salinity *J. Geophys. Res.: Oceans* **123** 4280–97

[111] Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial networks (arXiv:1406.2661)

[112] Griffies S, Pacanowski R, Schmidt M and Balaji V 2001 Tracer conservation with an explicit free surface method for z-coordinate ocean models *Mon. Weather Rev.* **129**

[113] Groenquist P, Yao C, Ben-Nun T, Dryden N, Dueben P, Li S and Hoefler T 2021 Deep learning for post-processing ensemble weather forecasts *Phil. Trans. R. Soc.* A **379** 20200092

[114] Grooms I 2021 Analog ensemble data assimilation and a method for constructing analogs with variational autoencoders *Q. J. R. Meteorol. Soc.* **147** 139–49

[115] Guillaumin A and Zanna L 2021 Stochastic deep learning parameterization of ocean momentum forcing *Earth Space Sci. Open Arch.* **31**

[116] Guimbard Sebastien, Gourrion Jerôme, Portabella M, Turiel A, Gabarró C and Font J 2012 SMOS semi-empirical ocean forward model adjustment *IEEE Trans. Geosci. Remote Sens.* **50** 1676–87

[117] Gurvan M *et al* 2019 Nemo ocean engine *Zenodo* (https://doi.org/10.5281/zenodo.3878122)

[118] Gysi T, Osuna C, Fuhrer O, Bianco M and Schulthess T C 2015 STELLA: a domain-specific tool for structured grid methods in weather and climate models *SC'15: Proc. Int. Conf. for High Performance Computing, Networking, Storage and Analysis* pp 1–12

[119] Haine T 2008 What did the Viking discoverers of America know of the north Atlantic environment? *Weather* **63** 60–65

[120] Ham Y-G, Kim J-H and Luo J-J 2019 Deep learning for multi-year enso forecasts *Nature* **573** 568–72

[121] Hanawa K and Talley L 2001 *Ocean Circulation and Climate* (International Geophysics Series) (New York: Academic) pp 373–86

[122] Hatfield S, Chantry M, Düben P and Palmer T 2019 Accelerating high-resolution weather models with deep-learning hardware *Proc. Platform for Advanced Scientific Conf., PASC'19 (New York, NY)* (Association for Computing Machinery)

[123] Hatfield S E, Chantry M, Dueben P D, Lopez P, Geer A J and Palmer T N 2021 Building tangent-linear and adjoint models for data assimilation with neural networks *Earth Space Sci. Open Arch.* **34**

[124] Hazeleger W and Drijfhout S S 2000 Eddy subduction in a model of the subtropical gyre *J. Phys. Oceanogr.* **30** 677–95

[125] Held I 2005 The gap between simulation and understanding in climate modeling *Bull. Am. Meteorol. Soc.* **86** 1609–14

[126] Helland-Hansen B 1916 Nogen hydrografiske metoder. Forhandlinger ved de 16 Skandinaviske Naturforsherermøte *Kristiania* **39** 357–9

[127] Hochreiter S 1991 Untersuchungen zu dynamischen neuronalen Netzen Diploma Thesis Technische Universität München

[128] Houghton I A and Wilson J D 2020 El niño detection via unsupervised clustering of argo temperature profiles *J. Geophys. Res.: Oceans* **125** e2019JC015947

[129] Hourdin F *et al* 2017 The art and science of climate model tuning *Bull. Am. Meteorol. Soc.* **98** 589–602

[130] Hourdin F *et al* 2020 Process-based climate model development harnessing machine learning. II. Model calibration from single column to global *Earth Space Sci. Open Arch.* **75**

[131] Hsieh W W 2009 *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels* (Cambridge: Cambridge University Press)

[132] Hsieh W W and Tang B 1998 Applying neural network models to prediction and data analysis in meteorology and oceanography *Bull. Am. Meteorol. Soc.* **79** 1855–70

[133] Hughes C W and De Cuevas B A 2001 Why western boundary currents in realistic oceans are inviscid: a link between form stress and bottom pressure torques *J. Phys. Oceanogr.* **31** 2871–85

[134] Irrgang C, Boers N, Sonnewald M, Barnes E A, Kadow C, Staneva J and Saynisch-Wagner J 2021 Will artificial intelligence supersede earth system and climate models? (arXiv:2101.09126)

[135] Irrgang C, Saynisch J and Thomas M 2019 Estimating global ocean heat content from tidal magnetic satellite observations *Sci. Rep.* **9** 7893

[136] Jin X, Cai S, Li H and Karniadakis G E 2021 NSFnets (Navier–Stokes flow nets): physics-informed neural networks for the incompressible Navier–Stokes equations *J. Comput. Phys.* **426** 109951

[137] Johnson G C and Bryden H L 1989 On the size of the Antarctic circumpolar current *Deep Sea Res.* A **36** 39–53

[138] Jones D C, Holt H J, Meijers A J S and Shuckburgh E 2019 Unsupervised clustering of southern ocean argo float temperature profiles *J. Geophys. Res.: Oceans* **124** 390–402

[139] Jones D C and Ito T 2019 Gaussian mixture modeling describes the geography of the surface ocean carbon budget *Proc. 9th Int. Workshop on Climate Informatics: CI 2019* ed J Brajard, A Charantonis, C Chen and J Runge (University Corporation for Atmospheric Research (UCAR)) pp 108–13

[140] Kaiser B E, Saenz J A, Sonnewald M and Livescu D 2021 Objective discovery of dominant dynamical processes with intelligible machine learning (arXiv:2106.12963)

[141] Kalnay E, Li H, Miyoshi T, Yang S-C and Ballabrera-Poy J 2007 4-D-Var or ensemble kalman filter? *Tellus* A **59** 758–73

[142] Kavanaugh M T, Oliver M J, Chavez F P, Letelier R M, Muller-Karger F E and Doney S C 2016 Seascapes as a new vernacular for pelagic ocean monitoring, management and conservation *ICES J. Mar. Sci.* **73** 1839–50

[143] Key J R and Schweiger A J 1998 Tools for atmospheric radiative transfer: streamer and fluxnet *Comput. Geosci.* **24** 443–51

[144] Kingma D P and Welling M 2013 Auto-encoding variational Bayes (arXiv:1312.6114)

[145] Kloewer M, Dueben P D and Palmer T N 2020 Number formats, error mitigation and scope for 16-bit arithmetics in weather and climate modeling analyzed with a shallow water model *J. Adv. Modeling Earth Syst.* **12** e2020MS002246

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

[147] Kohonen T 1982 Self-organized formation of topologically correct feature maps *Biol. Cybern.* **43** 59–69

[148] Kovachki N B and Stuart A M 2019 Ensemble Kalman inversion: a derivative-free technique for machine learning tasks *Inverse Problems* (https://doi.org/10.1088/1361-6420/ab1c3a)

[149] Kramer M A 1991 Nonlinear principal component analysis using autoassociative neural networks *AIChE J.* **37** 233–43

[150] Kurth T *et al* 2018 Exascale deep learning for climate analytics *SC18: Int. Conf. for High Performance Computing, Networking, Storage and Analysis* pp 649–60

[151] Laloyaux P, Bonavita M, Chrust M and Gürol S 2020 Exploring the potential and limitations of weak-constraint 4D-Var *Q. J. R. Meteorol. Soc.* **146** 4067–82

[152] Landschutzer P, Gruber N, Bakker D and Schuster U 2014 Recent variability of the global ocean carbon sink *Glob. Biogeochem. Cycles* **28** 1–23

[153] Lapuschkin S, Binder A, Montavon G, Klauschen F, Müller K-R and Samek W 2015 On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation *PLoS One* **10** e0130140

[154] Le Traon P Y, Nadal F and Ducet N 1998 An improved mapping method of multisatellite altimeter data *J. Atmos. Ocean. Technol.* **15** 522–34

[155] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44

[156] Leinonen J, Nerini D and Berne A 2020 Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network *IEEE Trans. Geosci. Remote Sens.* 1–13

[157] Lellouche J-M *et al* 2018 Mercator ocean global high-resolution monitoring and forecasting system *New Front. Oper. Oceanogr.* 563–92

[158] Lguensat R, Le Sommer J, Metref S, Cosme E and Fablet R 2019 Learning generalized quasi-geostrophic models using deep neural numerical models (arXiv:1911.08856)

[159] Lguensat R, Tandeo P, Ailliot P, Pulido M and Fablet R 2017 The analog data assimilation *Mon. Weather Rev.* **145** 4093–107

[160] Ling J, Jones R and Templeton J 2016 Machine learning strategies for systems with invariance properties *J. Comput. Phys* **318** 22–35

[161] Lorenz E N 1992 The slow manifold—what is it? *J. Atmos. Sci.* **49** 2449–51

[162] Lorenz E N 1956 Empirical orthogonal functions and statistical weather prediction

[163] Lorenz E N 1969 Atmospheric predictability as revealed by naturally occurring analogues *J. Atmos. Sci.* **26** 636–46

[164] Lorenz E N 1969 Atmospheric predictability as revealed by naturally occurring analogues *J. Atmos. Sci.* **26** 636–46

[165] Lundberg S M and Lee S-I 2017 A unified approach to interpreting model predictions *Advances in Neural Information Processing Systems 30* ed I Guyon, U V Luxburg, S Bengio, H Wallach, R Fergus, S Vishwanathan and R Garnett (Curran Associates, Inc.) pp 4765–74

[166] Mack J, Arcucci R, Molina-Solana M and Guo Y-K 2020 Attention-based convolutional autoencoders for 3d-variational data assimilation *Comput. Methods Appl. Mech. Eng.* **372** 113291

[167] Manabe S and Bryan K 1969 Climate calculations with a combined ocean-atmosphere model *J. Atmos. Sci* **26** 786–9

[168] Manabe S, Stouffer R J, Spelman M J and Bryan K 1991 Transient responses of a coupled ocean–atmosphere model to gradual changes of atmospheric CO_2. Part I. Annual mean response *J. Clim.* **4** 785–818

[169] Manucharyan G E, Siegelman L and Klein P 2021 A deep learning approach to spatiotemporal sea surface height interpolation and estimation of deep currents in geostrophic ocean turbulence *J. Adv. Modeling Earth Syst.* **13** e2019MS001965

[170] Martinez E, Gorgues T, Lengaigne M, Fontana C, Sauzède R, Menkes C, Uitz J, Di Lorenzo E and Fablet R 2020 Reconstructing global chlorophyll-a variations using a non-linear statistical approach *Front. Mar. Sci.* **7** 464

[171] Maurer E P, Hidalgo H G, Das T, Dettinger M D and Cayan D R 2010 The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California *Hydrol. Earth Syst. Sci.* **14** 1125–38

[172] Maze G, Mercier H, Fablet R, Tandeo P, Radcenco M L, Lenca P, Feucher Cène and Clement L G 2017 Coherent heat patterns revealed by unsupervised classification of argo temperature profiles in the north Atlantic ocean *Prog. Oceanogr.* **151** 275–92

[173] Mazloff M R, Heimbach P and Wunsch C 2010 An eddy-permitting southern ocean state estimate *J. Phys. Oceanogr.* **40** 880–99

[174] McGovern A, Lagerquist R, Gagne D J, Jergensen G E, Elmore K L, Homeyer C R and Smith T 2019 Making the black box more transparent: understanding the physical implications of machine learning *Bull. Am. Meteorol. Soc.* **100** 2175–99

[175] McInnes L, Healy J and Melville J 2018 UMAP: uniform manifold approximation and projection for dimension reduction (arXiv:1802.03426)

[176] McLachlan G J and Basford K E 1988 *Mixture Models: Inference and Applications to Clustering* vol 38 (New York: M. Dekker)

[177] Merz A and Wust G 1923 Die atlantische vertikal zirkulation *3 Beitrag. Zeitschr. D.G.F.E, Berlin*

[178] Monahan A, Fyfe J, Ambaum M, Stephenson D and North G 2009 Empirical orthogonal functions: the medium is the message *J. Clim.* **22** 6501–14

[179] Montavon G, Samek W and Müller K-R 2018 Methods for interpreting and understanding deep neural networks *Digit. Signal Process.* **73** 1–15

[180] Montenegro A, Callaghan R T and Fitzpatrick S M 2016 Using seafaring simulations and shortest-hop trajectories to model the prehistoric colonization of remote oceania *Proc. Natl Acad. Sci.* **113** 12685–90

[181] Montgomery D C, Peck E A and Vining G G 2021 *Introduction to Linear Regression Analysis* (New York: Wiley)

[182] Munk W H 1950 On the wind-driven ocean circulation *J. Atmos. Sci.* **7** 80–93

[183] Munk W H 1950 On the wind-driven ocean circulation *J. Meteorol.* **7** 80–93

[184] Munk W H and Palmén E 1951 Note on the dynamics of the Antarctic circumpolar current 1 *Tellus* **3** 53–5

[185] Mustapha Z B, Alvain S, Jamet C, Loisel H and Dessailly D 2014 Automatic classification of water-leaving radiance

anomalies from global seawifs imagery: application to the detection of phytoplankton groups in open ocean waters *Remote Sens. Environ.* **146** 97–112

[186] Namias J 1959 Recent seasonal interactions between north pacific waters and the overlying atmospheric circulation *J. Geophys. Res.* **64** 631–46

[187] Nebeker F 1995 *Calculating the Weather: Meteorology in the 20th Century* (Amsterdam: Elsevier)

[188] Neumann P *et al* 2019 Assessing the scales in numerical weather and climate predictions: will exascale be the rescue? *Phil. Trans. R. Soc.* A **377** 20180148

[189] Nguyen A T, Pillar H, Ocaña V, Bigdeli A, Smith T A and Heimbach P 2020 The arctic subpolar gyre state estimate (ASTE): description and assessment of a data-constrained, dynamically consistent; ocean-sea ice estimate for 2002–2017 *Earth Space Sci. Open Arch.* 62

[190] Nowack P, Braesicke P, Haigh J, Abraham N L, Pyle J and Voulgarakis A 2018 Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations *Environ. Res. Lett.* **13** 104016

[191] Obukhov A M 1947 Statistically homogeneous fields on a sphere *Usp. Mat. Nauk* **2** 196–8

[192] O'Gorman P A and Dwyer J G 2018 Using machine learning to parameterize moist convection: potential for modeling of climate, climate change and extreme events *J. Adv. Modeling Earth Syst.* **10** 2548–63

[193] Olden J D, Joy M K and Death R G 2004 An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data *Ecol. Modelling* **178** 389–97

[194] Ott J, Pritchard M, Best N, Linstead E, Curcic M and Baldi P 2020 A Fortran–Keras deep learning bridge for scientific computing *Scientific Programming* (arXiv:2004.10652)

[195] Palerme C and Müller M 2021 Calibration of sea ice drift forecasts using random forest algorithms *Cryosphere Discuss.* **2021** 1–22

[196] Patel D, Canaday D, Girvan M, Pomerance A and Ott E 2021 Using machine learning to predict statistical properties of non-stationary dynamical processes: system climate, regime transitions and the effect of stochasticity *Chaos* **31** 033149

[197] Pauthenet E, Roquet F, Madec G, Sallée J-B and Nerini D 2019 The thermohaline modes of the global ocean *J. Phys. Oceanogr.* **49** 2535–52

[198] Penny S G, Bach E, Bhargava K, Chang C-C, Da C, Sun L and Yoshida T 2019 Strongly coupled data assimilation in multiscale media: experiments using a quasi-geostrophic coupled model *J. Adv. Modeling Earth Syst.* **11** 1803–29

[199] Phillips N A 1956 The general circulation of the atmosphere: a numerical experiment *Q. J. R. Meteorol. Soc.* **82** 123–64

[200] Puscasu R M 2014 Integration of artificial neural networks into operational ocean wave prediction models for fast and accurate emulation of exact nonlinear interactions *Proc. Comput. Sci.* **29** 1156–70

[201] Rackauckas C, Ma Y, Martensen J, Warner C, Zubov K, Supekar R, Skinner D, Ramadhan A and Edelman A 2020 Universal differential equations for scientific machine learning (arXiv:2001.04385 [cs, math, q-bio, stat])

[202] Raissi M, Perdikaris P and Karniadakis G E 2017 Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations (arXiv:1711.10561)

[203] Ramadhan A *et al* 2020 Oceananigans.jl: fast and friendly geophysical fluid dynamics on gpus *J. Open Source Softw.* **5** 2018

[204] Rasp S, Dueben P D, Scher S, Weyn J A, Mouatadid S and Thuerey N 2020 Weatherbench: a benchmark data set for data-driven weather forecasting *J. Adv. Modeling Earth Syst.* **12** e2020MS002203

[205] Rasp S and Lerch S 2018 Neural networks for postprocessing ensemble weather forecasts *Mon. Weather Rev.* **146** 3885–900

[206] Rasp S, Pritchard M S and Gentine P 2018 Deep learning to represent subgrid processes in climate models *Proc. Natl Acad. Sci.* **115** 9684–9

[207] Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N and Prabhat M 2019 Deep learning and process understanding for data-driven earth system science *Nature* **566** 195–204

[208] Ribeiro M T, Singh S and Guestrin C 2016 "Why should I trust you?": explaining the predictions of any classifier *Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (San Francisco, CA, 13–17 August 2016)* pp 1135–44

[209] Richardson L F 2007 *Weather Prediction by Numerical Process* (Cambridge: Cambridge University Press)

[210] Richardson P L 2008 On the history of meridional overturning circulation schematic diagrams *Prog. Oceanogr.* **76** 466e486

[211] Robinson A R, Spall M A, Walstad L J and Leslie W G 1989 Data assimilation and dynamical interpolation in gulfcast experiments *Dyn. Atmos. Oceans* **13** 301–16

[212] Roemmich D, Johnson G C, Riser S, Davis R, Gilson J, Owens W B, Garzoli S L, Schmid C and Ignaszewski M 2009 The argo program: observing the global ocean with profiling floats *Oceanography* **22** 34–43

[213] Rosso I, Mazloff M R, Talley L D, Purkey S G, Freeman N M and Maze G 2020 Water mass and biogeochemical variability in the Kerguelen sector of the southern ocean: a machine learning approach for a mixing hot spot *J. Geophys. Res.: Oceans* **125** e2019JC015877

[214] Rudin C 2019 Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead *Nat. Mach. Intell.* **1** 206–15

[215] Ryzhov E A, Kondrashov D, Agarwal N, McWilliams J C and Berloff P 2020 On data-driven induction of the low-frequency variability in a coarse-resolution ocean model *Ocean Modelling* **153** 101664

[216] Sakov P, Counillon F, Bertino L, Lisæter K A, Oke P R and Korablev A 2012 TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic *Ocean Sci.* **8** 633–56

[217] Sandström J W and Helland-Hansen B 1903 Über die berechnung von meeresströmung *Report on Norwegian Fishery and Marine Investigations* (Bergen: J. Grieg) vol 2

[218] Schiller H and Doerffer R 1999 Neural network for emulation of an inverse model operational derivation of case II water properties from Meris data *Int. J. Remote Sens.* **20** 1735–46

[219] Schmidhuber J 2015 Deep learning in neural networks: an overview *Neural Netw.* **61** 85–117

[220] Schmidt M and Lipson H 2009 Distilling free-form natural laws from experimental data *Science* **324** 81–85

[221] Schneider T, Teixeira J, Bretherton C S, Brient F, Pressel K G, Schär C and Siebesma A P 2017 Climate goals and computing the future of clouds *Nat. Clim. Change* **7** 3–5

[222] Schuhen N, Thorarinsdottir T L and Gneiting T 2012 Ensemble model output statistics for wind vectors *Mon. Weather Rev.* **140** 3204–19

[223] Shackley S, Risbey J, Stone P and Wynne B 1999 Adjusting to policy expectations in climate change modeling *Clim. Change* **43** 413–54

[224] Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K and Woo W-chun 2015 Convolutional LSTM network: a machine learning approach for precipitation nowcasting (arXiv:1506.04214)

[225] Silver D *et al* 2017 Mastering the game of go without human knowledge *Nature* **550** 354–9

[226] Song Z, Bao Y, Zhang D, Shu Q, Song Y and Qiao F 2020 Centuries of monthly and 3-hourly global ocean wave data for past, present and future climate research *Sci. Data* **7** 226

[227] Sonnewald M, Dutkiewicz S, Hill C and Forget G 2020 Elucidating ecological complexity: unsupervised learning determines global marine eco-provinces *Sci. Adv.* **6** 1–12

[228] Sonnewald M and Lguensat R 2021 Revealing the impact of global heating on North Atlantic circulation using transparent machine learning *J. Adv. Model. Earth Syst.* e2021MS002496

[229] Sonnewald M, Wunsch C and Heimbach P 2019 Unsupervised learning reveals geography of global ocean dynamical regions *Earth Space Sci.* **6** 784–94

[230] Steinhaus H 1956 Sur la division des corps matériels en parties *Bull. Acad. Polon. Sci* **1** 801

[231] Stengel K, Glaws A, Hettinger D and King R N 2020 Adversarial super-resolution of climatological wind and solar data *Proc. Natl Acad. Sci.* **117** 16805–15

[232] Stommel H 1948 The westward intensification of wind-driven ocean currents *EOS Trans. Am. Geophys. Union* **29** 202–6

[233] Sutton R S and Barto A G 2018 *Reinforcement Learning: An Introduction* (Cambridge, MA: MIT Press)

[234] Sverdrup H U 1942 *Oceanography for Meteorologists* (Daya Books)

[235] Sverdrup H U 1947 Wind-driven currents in a baroclinic ocean; with application to the equatorial currents of the eastern pacific *Proc. Natl Acad. Sci.* **33** 318–26

[236] Swapna P, Krishnan R, Sandeep N, Prajeesh A G, Ayantika D C, Manmeet S and Vellore R 2018 Long-term climate simulations using the IITM earth system model (IITM-ESMv2) with focus on the South Asian Monsoon *J. Adv. Modeling Earth Syst.* **10** 1127–49

[237] Swapna P *et al* 2014 The IITM earth system model: transformation of a seasonal prediction model to a long term climate model *Bull. Am. Meteorol. Soc.* **96** 1351–67

[238] Szepesvári C 2010 Algorithms for reinforcement learning *Synth. Lect. Artif. Intell. Mach. Learn.* **4** 1–103

[239] Tandeo P, Ailliot P, Bocquet M, Carrassi A, Miyoshi T, Pulido M and Zhen Y 2020 A review of innovation-based methods to jointly estimate model and observation error covariance matrices in ensemble data assimilation *Mon. Weather Rev.* **148** 3973–94

[240] Tesdal J-E and Abernathey R P 2021 Drivers of local ocean heat content variability in ECCOv4 *J. Clim.* **34** 2941–56

[241] Thiria S, Mejia C, Badran F and Crepon M 1993 A neural network approach for modeling nonlinear transfer functions: application for wind retrieval from spaceborne scatterometer data *J. Geophys. Res.: Oceans* **98** 22827–41

[242] Thomas L N, Tandon A, Mahadevan A, Hecht M and Hasumi H 2008 *Ocean Modeling in an Eddying Regime* (*Geophysical Monograph Series*) vol 177 (Washington, DC: American Geophysical Union) pp 17–38

[243] Thomas S D A, Jones D C, Faul A, Mackie E and Pauthenet E 2021 Defining southern ocean fronts using unsupervised classification *Ocean Sci. Discuss.* **2021** 1–29

[244] Tintó Prims O, Acosta M C, Moore A M, Castrillo M, Serradell K, Cortés A and Doblas-Reyes F J 2019 How to use mixed precision in ocean models: exploring a potential reduction of numerical precision in NEMO 4.0 and ROMS 3.6 *Geosci. Model Dev.* **12** 3135–48

[245] Tolman H L 1991 A third-generation model for wind waves on slowly varying, unsteady and inhomogeneous depths and currents *J. Phys. Oceanogr.* **21** 782–97

[246] Toms B A, Barnes E A and Ebert-Uphoff I 2020 Physically interpretable neural networks for the geosciences: applications to earth system variability *J. Adv. Modeling Earth Syst.* **12** e2019MS002002

[247] Tuppi L, Ollinaho P, Ekblom M, Shemyakin V and Järvinen H 2020 Necessary conditions for algorithmic tuning of weather prediction models using openifs as an example *Geosci. Model Dev.* **13** 5799–812

[248] Tyndall J 1859 Note on the transmission of radiant heat through gaseous bodies *Proc. R. Soc.* A **10** 37–9

[249] Vallis G 2016 Geophysical fluid dynamics: whence, whither and why? *Proc. R. Soc.* A **472** 20160140

[250] Van der Maaten L and Hinton G 2008 Visualizing data using t-SNE *J. Mach. Learn. Res.* **9** 2579–605

[251] Vandal T, Kodra E, Dy J, Ganguly S, Nemani R and Ganguly A R 2018 Quantifying uncertainty in discrete-continuous and skewed data with Bayesian deep learning *Proc. 24th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining, KDD'18 (New York, NY)* (Association for Computing Machinery) pp 2377–86

[252] Vandal T, Kodra E and Ganguly A R 2019 Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation *Theor. Appl. Climatol.* **137** 557–70

[253] Vecchi G A *et al* 2014 On the seasonal forecasting of regional tropical cyclone activity *J. Clim.* **27** 7994–8016

[254] Verdy A and Mazloff M R 2017 A data assimilating model for estimating Southern Ocean biogeochemistry *J. Geophys. Res.: Oceans* **122** 6968–88

[255] Žust L, Fettich A, Kristan M and Ličer M 2020 HIDRA 1.0: deep-learning-based ensemble sea level forecasting in the northern adriatic *Geosci. Model Dev.* **2020** 1–25

[256] Walker G 1928 World weather *Q. J. R. Meteorol. Soc.* **54** 79–87

[257] Wang R, Kashinath K, Mustafa M, Albert A and Yu R 2020 Towards physics-informed deep learning for turbulent flow prediction (arXiv:1911.08655 [physics, stat])

[258] Wang Y, Yao Q, Kwok J and Ni L M 2020 Generalizing from a few examples: a survey on few-shot learning (arXiv:1904.05046)

[259] Watson A, Schuster U, Shutler J, Holding T, Ashton I, Landschützer P, Woolf D and Goddijn-Murphy L 2020 Revised estimates of ocean-atmosphere CO_2 flux are consistent with ocean carbon inventory *Nat. Commun.* **11** 4422

[260] Watson P 2019 Applying machine learning to improve simulations of a chaotic dynamical system using empirical error correction *J. Adv. Modeling Earth Syst.* **11** 1402–17

[261] Weyn J A, Durran D R and Caruana R 2020 Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere *J. Adv. Modeling Earth Syst.* **12** e2020MS002109

[262] Wikle C K, Milliff R F, Herbei R and Leeds W B 2013 Modern statistical methods in oceanography: a hierarchical perspective *Stat. Sci.* **28** 466–86

[263] Willard J, Jia X, Xu S, Steinbach M and Kumar V 2020 Integrating physics-based modeling with machine learning: a survey (arXiv:2003.04919 [physics, stat])

[264] Williams C K I and Rasmussen C E 1996 Gaussian processes for regression *Proc. 8th Int. Conf. on Neural Information Processing Systems*

[265] Wood A W, Leung L R, Sridhar V and Lettenmaier D P 2004 Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs *Clim. Change* **62** 189–216

[266] Wunsch C 2002 Ocean observations and the climate forecast problem *Int. Geophysics* **83** 233–45

[267] Wunsch C 2005 Towards the world ocean circulation experiment and a bit of aftermath *Physical Oceanography: Developments Since 1950* pp 181–201

[268] Yuval J, O'Gorman P A and Hill C N 2021 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

[269] Zanna L and Bolton T 2020 Data-driven equation discovery of ocean mesoscale closures *Geophys. Res. Lett.* **47** e2020GL088376

[270] Zebiak S E and Cane M A 1987 A model El Niñ–Southern Oscillation *Mon. Weather Rev.* **115** 2262–78

[271] Zhai J and Bitz C 2020 A machine learning model of Arctic sea ice motions *Earth Space Sci. Open Arch.* **11**