

POSITION PAPER

Evolution of machine learning in environmental science—A perspective

William W. Hsieh 

Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Corresponding author: E-mail: whsieh@eos.ubc.ca

Current address: 4028 Hopesmore Drive, Victoria, British Columbia V8N 5S9, Canada.

Received: 08 June 2021; **Revised:** 29 December 2021; **Accepted:** 24 January 2022

Keywords: Data assimilation; encoder–decoder; generative-adversarial network; machine learning; neural network

Abstract

The growth of machine learning (ML) in environmental science can be divided into a slow phase lasting till the mid-2010s and a fast phase thereafter. The rapid transition was brought about by the emergence of powerful new ML methods, allowing ML to successfully tackle many problems where numerical models and statistical models have been hampered. Deep convolutional neural network models greatly advanced the use of ML on 2D or 3D data. Transfer learning has allowed ML to progress in climate science, where data records are generally short for ML. ML and physics are also merging in new areas, for example: (a) using ML for general circulation model parametrization, (b) adding physics constraints in ML models, and (c) using ML in data assimilation.

Impact Statement

This perspective paper reviews the evolution and growth of machine learning (ML) models in environmental science. The opaque nature of ML models led to decades of slow growth, but exponential growth commenced around the mid-2010s. Novel ML models which have contributed to this exponential growth (e.g., deep convolutional neural networks, encoder–decoder networks, and generative-adversarial networks) are reviewed, as well as approaches to merging ML models with physics-based models.

1. Introduction

Thirty years ago, a typical environmental scientist would know some statistics but would not have heard of “machine learning” (ML) and would know artificial intelligence (AI) only through science fiction. After World War II, the great popularity of AI in science fiction led to very unrealistic expectations on how fast AI research would progress. The inevitable disappointment led to negative reviews and two major “AI winters,” periods of poor funding around 1974–1980 and 1987–1993 (Crevier, 1993; Nilsson, 2009). Partly to focus on a more specific aspect and partly to avoid the stigma associated with AI, many researchers started to refer to their work using other names, for example, ML, where the goal of ML is to have computers learn from data without being explicitly programmed.

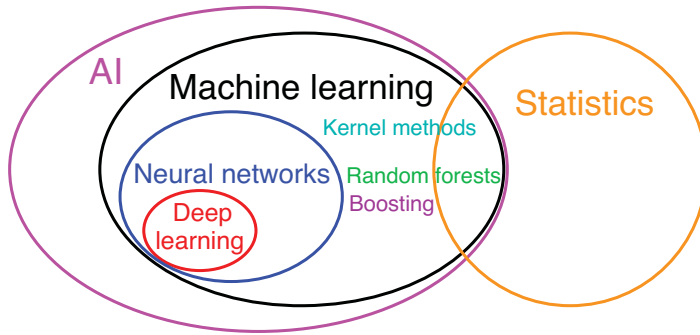


Figure 1. Venn diagram illustrating the relation between artificial intelligence, statistics, machine learning, neural networks, and deep learning, as well as kernel methods, random forests, and boosting.

While there is overlap between the data methods developed in statistics and in the younger field of ML (Figure 1), ML germinated mainly in computer science, psychology, engineering, and commerce, while statistics had largely been rooted in mathematics, leading to two fairly distinct cultures (Breiman, 2001b). When fitting a curve to a dataset, a statistician would ensure the number of adjustable model parameters is small compared to the sample size (i.e., number of observations) to avoid overfitting (i.e., the model fitting to the noise in the data). This prudent practice in statistics is not strictly followed in ML, as the number of parameters (called “weights” in ML) can be greater, sometimes much greater, than the sample size (Krizhevsky et al., 2012), as ML has developed ways to avoid overfitting while using a large number of parameters. The relatively large number of parameters renders ML models much more difficult to interpret than statistical models; hence, ML models are often regarded somewhat dismissively as “black boxes.” In ML, the artificial neural network (NN) model called the multilayer perceptron (MLP; Rumelhart et al., 1986; Goodfellow et al., 2016) has become widely used since the late 1980s.

How readily an environmental science (ES) adopted NN or other ML models depended on whether successful physics-based models were available. Meteorology, where dynamical models have been routinely used for weather forecasting, has been slower to embrace NN models than hydrology, where by year 2000 there were already 43 hydrological papers using NN models (Maier and Dandy, 2000), as physics-based models were not very skillful in forecasting streamflow from precipitation data.

Relative to linear statistical models, the nonlinear ML models also need relatively large sample size to excel. Hence, oceanography, a field with far fewer in situ observations than hydrology or meteorology, and climate science, where the longtime scales preclude large effective sample size, are fields where the application of ML models has been hampered. Furthermore, averaging daily data to produce climate data linearizes the relation between the predictors and the response variables due to the central limit theorem, thereby reducing the nonlinear modelling advantage of ML models (Yuval and Hsieh, 2002). Another disadvantage of many ML models (e.g., NN) relative to linear statistical models is that they can extrapolate much worse when given new predictor data lying outside the original training domain (Hsieh, 2020), as nonlinear extrapolation is an ill-posed problem. This is not ML-specific, since any nonlinear statistical model faces the same ill-posed problem when used for extrapolation.

Overall, around year 2010, ML models were fairly well accepted in hydrology and remote sensing, but remained fringe in meteorology and even less developed in oceanography and climate science. Nevertheless, a number of books were written on the application of ML methods to ES in this earlier development phase (Abrahart et al., 2004; Blackwell and Chen, 2009; Haupt et al., 2009; Hsieh, 2009; Krasnopolsky, 2013). After this relatively flat phase, rapid growth of ML in ES commenced in the mid-2010s.

Section 2 looks at the emergence of new and more powerful ML methods in the last decade, whereas Section 3 reviews their applications to ES. The merging of ML and physical/dynamical models is examined in Section 4.

2. Evolution of ML Methods

The MLP NN model maps the input variables through layers of hidden neurons/nodes (i.e., intermediate variables) to the output variables. Given some training data for the input and output variables, the model weights can be solved by minimizing an objective function using a back-propagation algorithm. The traditional MLP NN is mostly limited to one or two hidden layers, because the gradients (error signals) in the back-propagation algorithm become vanishingly small after propagating through many layers. Without a solution for the vanishing gradient problem, NN research stalled while newer methods—kernel methods (e.g., support vector machines (Cortes and Vapnik, 1995)) emerging from the mid-1990s and random forests from 2001 (Breiman, 2001a)—seriously challenged NN’s dominant position in ML.

The traditional MLP NN has each neuron in one layer connected to all the neurons in the preceding layer. When working with image data, the MLP uses a huge number of weights—for example, mapping an input 100×100 image to just one neuron in the first hidden layer requires 10,000 weights! Since, in nature, biological neurons are connected only to neighboring neurons, to have every neuron in one layer of an NN model connected to all the neurons in the preceding layer is unnatural and very wasteful of computing resources. With inspiration from the animal visual cortex, the convolutional layer has been developed, where a neuron is only connected to a small patch of neurons in the preceding layer, thereby resulting in a drastic reduction of weights compared to the traditional fully connected layers and giving rise to *convolutional neural networks* (CNNs; LeCun et al., 1989).

Eventually, the vanishing gradient problem was overcome, and *deep NN* or *deep learning* (DL) models, that is, NN having ≥ 5 layers of mapping functions with adjustable weights (LeCun et al., 2015), emerged. In 2012, a deep NN model won the ImageNet Large Scale Visual Recognition Challenge (Krizhevsky et al., 2012). The huge reduction in the number of weights in convolutional layers made deep NN feasible.

With all the impressive breakthroughs in DL since 2012 (Goodfellow et al., 2016), there is a popular misconception that DL models are superior to all other ML models. Actually, the best ML model is very problem-dependent. There are two main types of datasets, structured and unstructured. *Structured* datasets have a tabular format, for example, like an Excel spreadsheet, with the variables listed in columns. In contrast, *unstructured* datasets include images, videos, audio, text, and so forth. For unstructured datasets, DL has indeed been dominant. For structured datasets, however, ML models with shallow depth structure, for example, *gradient boosting* models such as XGBoost (Chen and Guestrin, 2016), have often beaten deep NN in competitions, for example, those organized by Kaggle (www.kaggle.com).

How can “shallow” gradient boosting beat deep NN in structured data? Typically, in a structured dataset, the predictors are quite inhomogeneous (e.g., pressure, temperature, and humidity), whereas in an unstructured dataset, the predictors are more homogeneous (e.g., temperature at various pixels in a satellite image or at various grid points in a numerical model). Boosting is based on decision trees, where the effects of the predictors x are treated independently of each other, as the path through a decision tree is controlled by questions like: is “ $x_1 > a?$,” “ $x_2 > b?$,” and so forth (Breiman et al., 1984). In contrast, in NN, the predictors are combined by a linear combination ($\sum_i w_i x_i$) before being passed through an activation/transfer function onto the next layer. With inhomogeneous predictors, for example, temperature and pressure, treating the two separately as in decision trees intuitively makes more sense than trying to add the two together by a linear combination.

Another type of NN architecture is the *encoder–decoder* model, where the encoder part first maps from a high-dimensional input space to a low-dimensional space, then the decoder part maps back to a high-dimensional output space (Figure 2). If the target output data are the same as the input data, the model becomes an *autoencoder*, which has been used for nonlinear principal component analysis (PCA), as the low-dimensional space can be interpreted as nonlinear principal components (Kramer, 1991; Monahan, 2000; Hsieh, 2001). The popular U-net (Ronneberger et al., 2015) is a deep CNN model with an encoder–decoder architecture.

In many games, having two individuals playing against each other enhances the skill level of both, for example, a soccer goal scorer practicing against a goal keeper. The *generative adversarial network* (GAN)

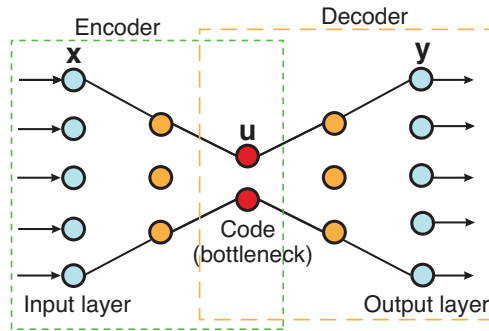


Figure 2. The encoder–decoder is an NN model with the first part (the encoder) mapping from the input x to u , the “code” or bottleneck, and the second part (the decoder) mapping from u to the output y . Dimensional compression is achieved by forcing the signal through the bottleneck. The encoder and the decoder are each illustrated with only one hidden layer for simplicity.

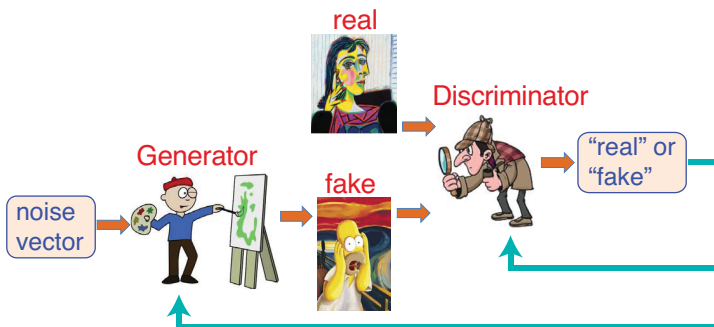


Figure 3. Generative adversarial network with the generator creating a fake image (e.g., a fake Picasso painting) from random noise input, and the discriminator classifying images as either real or fake. Whether the discriminator classifies a fake image rightly or wrongly leads, respectively, to further training for the generator or for the discriminator.

has two submodels, the generator and the discriminator, playing as adversaries, with the goal of producing realistic fake data (Goodfellow et al., 2014). Given a random input vector, the generator outputs a set of fake data. The discriminator receives either real data or fake data as input and classifies them as either “real” or “fake” (Figure 3). If a fake is correctly classified, the generator’s model weights are updated, whereas if the fake is mistaken to be real, the discriminator’s model weights are updated. The skill levels of both players improve until at the end the discriminator can only identify fake data from the generator about 50% of the time. After training is done, the discriminator is discarded while the generator is retained to produce new fake data. The variational autoencoder model provides an alternative to GAN (Kingma and Welling, 2014).

Conditional generative adversarial network (CGAN), introduced by Mirza and Osindero (2014), supplies additional input x to both the generator and the discriminator. If x is an image, CGAN can be used for image-to-image translation tasks (Isola et al., 2017), for example, translating a line drawing to a photo image, a map to a satellite map, or vice versa (Figure 4).

3. Applications in Environmental Science

Applications of ML to ES have come in roughly two groups. In the first group, ML methods are used largely as nonlinear generalizations of traditional statistical methods. For instance, MLP NN models are



Figure 4. Conditional generative adversarial network where the generator G receives an image x and a random noise vector z as input. The discriminator D receives x plus either a fake image from G (left) or a real image y (right) as input. Here, a line drawing is converted to a photo image; similarly, a photo image can be converted to a line drawing. Adapted from Figure 2 of Isola et al. (2017).

used for nonlinear regression, classification, PCA, and so forth (Hsieh, 2009). In the second group, the ML methods do not have counterparts in statistics, for example, GAN or CGAN.

CNN models have appeared in ES in the last few years, for example, to estimate the posterior probability of three types of extreme events (tropical cyclones, atmospheric rivers, and weather fronts) from 2D images of atmospheric variables (Liu et al., 2016), to detect synoptic-scale weather fronts (cold front, warm front, and no front; Lagerquist et al., 2019) and for next-hour tornado prediction from radar images (Lagerquist et al., 2020).

Performing classification on each pixel of an image is called *semantic segmentation* in ML. Most CNN models for semantic segmentation use an encoder–decoder architecture, including the popular U-net deep CNN model (Ronneberger et al., 2015), giving classification (or regression) on individual output pixels. U-net has been applied to rain-type classification (no-rain, stratiform, convective, and others) using microwave satellite images (Choi and Kim, 2020), to precipitation estimation using satellite infrared images (Sadeghi et al., 2020) and to cloud cover nowcasting using visible and infrared images (Berthomier et al., 2020).

As traditional MLP NN models require a very large number of weights when working with 2D images or 3D spatial fields, limited sample size often necessitates the compression of input variables to a modest number of principal components by PCA, a linear technique (Jolliffe, 2002). The introduction of the CNN model has drastically reduced the number of weights, so compression by PCA is not needed, and deep CNN models have noticeably improved the performance of ML methods when working with 2D or 3D spatial fields in ES since the mid-2010s.

PCA has been used to impute missing values in datasets (Jolliffe, 2002, Section 13.6) and U-net, with its encoder–decoder architecture, can replace PCA in this task. Using sea surface temperature (SST) data from two sources (NOAA Twentieth-Century Reanalysis [20CR] and Coupled Model Intercomparison Project Phase 5 [CMIP5]), Kadow et al. (2020) trained U-net models, which outperformed PCA and kriging methods in imputing missing SST values.

The CGAN (Figure 4) has also been used in ES: In atmospheric remote sensing, CGAN generated cloud structures in a 2D vertical plane in the satellite’s along-track direction (Leinonen et al., 2019). As an alternative to U-net in super-resolution applications, CGAN has been used to convert low-resolution unmanned aircraft system images to high-resolution images (Pashaei et al., 2020).

The application of ML methods to climate problems has been impeded by the relatively small effective sample size in observational records. In *transfer learning*, ML models trained on a dataset with large sample size can transfer their learning to a different problem hampered by a relatively small sample size. Using a CNN to learn the El Niño–Southern Oscillation (ENSO) behavior from the dynamical models (CMIP5 climate model data for 2,961 months), then transferring the learning to observed data (by further training with 103 months of reanalysis data), Ham et al. (2019) developed a CNN model with better accuracy in ENSO prediction than the dynamical models.

4. Merging of Machine Learning and Physics

As some components of a physical model can be computationally expensive, ML methods have been developed to substitute for the physics: For atmospheric radiation in atmospheric general circulation models (GCMs), MLP NN models have been used to replace the equations of physics (Chevallier et al., 2000; Krasnopolsky et al., 2008). For a simple coupled atmosphere–ocean model of the tropical Pacific, the atmospheric component has been replaced by an NN model (Tang and Hsieh, 2002). Resolving clouds in a GCM would require high spatial resolution and prohibitive costs; hence, NN models, trained by a cloud-resolving model, have been used to supply convection parametrization in a GCM (Krasnopolsky et al., 2013; Brenowitz and Bretherton, 2018; Rasp et al., 2018). Increasingly, ML methods are used to learn from high-resolution numerical models, then implemented as inexpensive parametrization schemes in GCMs.

In *physics-informed machine learning*, ML models can be solved satisfying the laws of physics, for example, conservation of energy, mass, and so forth. In the soft constraint approach, the physics constraints are satisfied approximately by adding an extra regularization term in the objective function of an NN model (Karniadakis et al., 2021). Alternatively, in the hard constraint approach, the physics constraints are satisfied exactly by the NN architecture (Beucler et al., 2021).

Initially developed in numerical weather prediction (NWP), *data assimilation* (DA) aims to optimally merge theory (typically a numerical model based on physics) and observations. The most common DA method used in NWP is variational DA—4D-Var (three spatial dimensions plus time) and 3D-Var (spatial dimensions only; Kalnay, 2003). Hsieh and Tang (1998) noted that back propagation used in finding the optimal NN solution is actually the same technique as the backward integration of the adjoint model used in variational DA; hence, one could combine numerical and NN models in DA by solving a single optimization problem. For the three-component dynamical system of Lorenz (1963), Tang and Hsieh (2001) replaced one of the dynamical equations with an NN equation, and tested using variational assimilation to estimate the parameters of the dynamical and NN equations and the initial conditions. In recent years, there has been increasing interest in merging ML and DA using 4D-Var in a Bayesian framework (Bocquet et al., 2020; Geer, 2021) or using the ensemble Kalman filter (Brajard et al., 2020). NN models have the potential to greatly facilitate the building/maintaining of the tangent linear and adjoint models of the model physics in 4D-Var (Hatfield et al., 2021).

5. Summary and Conclusion

The recent growth of ML in ES has been fueled mainly by deep NN models (Camps-Valls et al., 2021), especially deep CNN models, which have greatly advanced the application of ML to 2D or 3D spatial data, gradually replacing many standard techniques like multiple linear regression, PCA, and so forth. Furthermore, some of the new ML models (e.g., GAN or CGAN) are no longer merely nonlinear generalization of a traditional statistical method. Progress has also been made in rendering ML methods less opaque and more interpretable (McGovern et al., 2019; Ebert-Uphoff and Hilburn, 2020). In climate science, where observational records are usually short, transfer learning has been able to utilize long simulations by numerical models for pretraining ML models. ML and physics have also been merging, for example, in (a) the increasing use of ML for parametrization in GCMs, with the ML model trained using data from high-resolution numerical models, (b) the implementation of physics constraints in ML models, and (c) the increasing interest in using ML in DA.

My overall perspective is that the evolution of ML in ES has two distinct phases, slow initial acceptance followed by exponential growth starting in the mid-2010s. If an assessment were made on the progress of ML in ES in 2010, one would have come to the conclusion that, with the exception of hydrology and remote sensing, ML was not considered mainstream in ES, as numerical models (or even statistical models) were far more transparent and interpretable than “black box” models from ML. It is encouraging to see the resistance to ML models waning, as new ML approaches have been increasingly successful in tackling areas of ES where numerical models and statistical models have been hampered.

Data Availability Statement. No data were used in this perspective paper.

Author Contributions. Conceptualization: W.W.H.; Methodology: W.W.H.; Visualization: W.W.H.; Writing—original draft: W.W.H.; Writing—review & editing: W.W.H.

Funding Statement. This work received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing Interests. The author declares no competing interests exist.

References

- Abrahart RJ, Kneale PE and See LM (eds.) (2004) *Neural Networks for Hydrological Modelling*. London: CRC Press.
- Berthomier L, Pradel B and Perez L (2020) Cloud cover nowcasting with deep learning. In *Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. Paris: IEEE, pp. 1–6. <https://doi.org/10.1109/IPTA50016.2020.9286606>
- Beucler T, Pritchard M, Rasp S, Ott J, Baldi P and Gentile P (2021) Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters* 126, 098302.
- Blackwell WJ and Chen FW (2009) *Neural Networks in Atmospheric Remote Sensing*. Boston: Artech House.
- Bocquet M, Brajard J, Carrassi A and Bertino L (2020) Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. *Foundations of Data Science* 2, 55–80.
- Brajard J, Carrassi A, Bocquet M and Bertino L (2020) Combining data assimilation and machine learning to infer unresolved scale parametrization. *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences* 379, 20200086.
- Breiman L (2001a) Random forests. *Machine Learning* 45, 5–32.
- Breiman L (2001b) Statistical modeling: The two cultures. *Statistical Science* 16, 199–215.
- Breiman L, Friedman J, Olshen RA and Stone C (1984) *Classification and Regression Trees*. New York: Chapman and Hall.
- Brenowitz ND and Bretherton CS (2018) Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters* 45, 6289–6298.
- Camps-Valls G, Tuia D, Zhu XX and Reichstein M (eds.) (2021) *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences*. Hoboken, NJ: Wiley.
- Chen T and Guestrin C (2016) XGBoost: A scalable tree boosting system. In *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. New York: Association for Computing Machinery.
- Chevallier F, Morcrette JJ, Cheruy F and Scott NA (2000) Use of a neural-network-based long-wave radiative-transfer scheme in the ECMWF atmospheric model. *Quarterly Journal of the Royal Meteorological Society* 126, 761–776.
- Choi Y and Kim S (2020) Rain-type classification from microwave satellite observations using deep neural network segmentation. *IEEE Geoscience and Remote Sensing Letters* 18, 2137–2141.
- Cortes C and Vapnik V (1995) Support vector networks. *Machine Learning* 20, 273–297.
- Crevier D (1993) *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books.
- Ebert-Uphoff I and Hilburn K (2020) Evaluation, tuning, and interpretation of neural networks for working with images in meteorological applications. *Bulletin of the American Meteorological Society* 101, E2149–E2170.
- Geer AJ (2021) Learning earth system models from observations: Machine learning or data assimilation? *Philosophical Transactions of the Royal Society A: Mathematical, Physical, and Engineering Sciences* 379, 20200089.
- Goodfellow I, Bengio Y and Courville A (2016) *Deep Learning*. Cambridge, MA: MIT Press.
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y (2014) Generative adversarial nets. In *Advances in Neural Information Processing Systems*, Vol. 27. Red Hook, NY: Curran Associates, Inc., pp. 2672–2680.
- Ham Y-G, Kim J-H and Luo J-J (2019) Deep learning for multi-year ENSO forecasts. *Nature* 573, 568–572.
- Hatfield S, Chantry M, Dueben P, Lopez P, Geer A and Palmer T (2021) Building tangent-linear and adjoint models for data assimilation with neural networks. *Journal of Advances in Modeling Earth Systems* 13, e2021MS002521.
- Haupt SE, Pasini A and Marzban C (eds.) (2009) *Artificial Intelligence Methods in the Environmental Sciences*. Dordrecht: Springer.
- Hsieh WW (2001) Nonlinear principal component analysis by neural networks. *Tellus A: Dynamic Meteorology and Oceanography* 53, 599–615.
- Hsieh WW (2009) *Machine Learning Methods in the Environmental Sciences*. Cambridge: Cambridge University Press.
- Hsieh WW (2020) Improving predictions by nonlinear regression models from outlying input data. *Preprint*, arXiv:2003.07926.
- Hsieh WW and Tang B (1998) Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society* 79, 1855–1870.
- Isola P, Zhu J-Y, Zhou T and Efros AA (2017) Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1125–1134. Available at https://openaccess.thecvf.com/content_cvpr_2017/html/Isola_Image-To-Image_Translation_With_CVPR_2017_paper.html?ref=https://githubhelp.com.

- Jolliffe IT** (2002) *Principal Component Analysis*, 2nd Edn. New York: Springer.
- Kadow C, Hall DM and Ulbrich U** (2020) Artificial intelligence reconstructs missing climate information. *Nature Geoscience* 13, 408–413.
- Kalnay E** (2003) *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge: Cambridge University Press.
- Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S and Yang L** (2021) Physics-informed machine learning. *Nature Reviews Physics* 3, 422–440.
- Kingma DP and Welling M** (2014) Auto-encoding variational Bayes. *Preprint*, arXiv:1312.6114.
- Kramer MA** (1991) Nonlinear principal component analysis using autoassociative neural networks. *AIChE Journal* 37, 233–243.
- Krasnopolsky VM** (2013) *The Application of Neural Networks in the Earth System Sciences: Neural Network Emulations for Complex Multidimensional Mappings*. Dordrecht: Springer.
- Krasnopolsky VM, Fox-Rabinovitz MS and Belochitski AA** (2008) Decadal climate simulations using accurate and fast neural network emulation of full, longwave and shortwave, radiation. *Monthly Weather Review* 136, 3683–3695.
- Krasnopolsky VM, Fox-Rabinovitz MS and Belochitski AA** (2013). Using ensemble of neural networks to learn stochastic convection parameterizations for climate and numerical weather prediction models from data simulated by a cloud resolving model. *Advances in Artificial Neural Systems 2013*, 485913.
- Krizhevsky A, Sutskever I and Hinton GE** (2012) ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, Vol. 25, pp. 1090–1098. RedHook, NY: Curran Associates, Inc.
- Lagerquist R, McGovern A and Gagne II DJ** (2019) Deep learning for spatially explicit prediction of synoptic-scale fronts. *Weather and Forecasting* 34, 1137–1160.
- Lagerquist R, McGovern A, Homeyer CR, Gagne II DJ and Smith T** (2020) Deep learning on three-dimensional multiscale data for next-hour tornado prediction. *Monthly Weather Review* 148, 2837–2861.
- LeCun Y, Bengio Y and Hinton G** (2015) Deep learning. *Nature* 521, 436–444.
- LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W and Jackel LD** (1989) Backpropagation applied to handwritten zip code recognition. *Neural Computation* 1, 541–551.
- Leinonen J, Guillaume A and Yuan T** (2019) Reconstruction of cloud vertical structure with a generative adversarial network. *Geophysical Research Letters* 46, 7035–7044.
- Liu Y, Racah E, Prabhat, Correa J, Khosrowshahi A, Lavers D, Kunkel K, Wehner M and Collins W** (2016) Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *Preprint*, arXiv:1605.01156.
- Lorenz EN** (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 20, 130–141.
- Maier HR and Dandy GC** (2000) Neural networks for the prediction and forecasting of water resources variables: A review of modelling issues and applications. *Environmental Modelling and Software* 15, 101–124.
- McGovern A, Lagerquist R, Gagne II DJ, Jergensen GE, Elmore KL, Homeyer CR and Smith T** (2019) Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society* 100, 2175–2199.
- Mirza M and Osindero S** (2014) Conditional generative adversarial nets. *Preprint*, arXiv:1411.1784.
- Monahan AH** (2000) Nonlinear principal component analysis by neural networks: Theory and application to the Lorenz system. *Journal of Climate* 13, 821–835.
- Nilsson NJ** (2009) *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press.
- Pashaei M, Starek MJ, Kamangir H and Berryhill J** (2020) Deep learning-based single image super-resolution: An investigation for dense scene reconstruction with UAS photogrammetry. *Remote Sensing* 12, 1757.
- Rasp S, Pritchard MS and Gentine P** (2018) Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America* 115, 9684–9689.
- Ronneberger O, Fischer P and Brox T** (2015) U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015)*, Lecture Notes in Computer Science, Vol. 9351. Cham, Switzerland: Springer, pp. 234–241.
- Rumelhart DE, Hinton GE and Williams RJ** (1986) Learning internal representations by error propagation. In Rumelhart D, McClelland J and PDP Research Group (eds.), *Parallel Distributed Processing*, Vol. 1. Cambridge, MA: MIT Press, pp. 318–362.
- Sadeghi M, Phu N, Hsu K and Sorooshian S** (2020) Improving near real-time precipitation estimation using a U-net convolutional neural network and geographical information. *Environmental Modelling and Software* 134, 104856.
- Tang Y and Hsieh WW** (2001) Coupling neural networks to incomplete dynamical systems via variational data assimilation. *Monthly Weather Review* 129, 818–834.
- Tang Y and Hsieh WW** (2002) Hybrid coupled models of the tropical Pacific: II ENSO prediction. *Climate Dynamics* 19, 343–353.
- Yuval and Hsieh WW** (2002) The impact of time-averaging on the detectability of nonlinear empirical relations. *Quarterly Journal of the Royal Meteorological Society* 128, 1609–1622.