



**Subject Areas:**

xxxxx, xxxxx, xxxx

**Keywords:**

xxxx, xxxx, xxxx

**Author for correspondence:**

Matthew Chantry

e-mail:

[matthew.chantry@physics.ox.ac.uk](mailto:matthew.chantry@physics.ox.ac.uk)

# Opportunities and Challenges for Machine Learning in Weather and Climate Modelling: Hard, Medium and Soft AI

Matthew Chantry<sup>1</sup>, Hannah Christensen<sup>1</sup>,  
Peter Düben<sup>2</sup>, Tim Palmer<sup>1</sup>

<sup>1</sup>Atmospheric, Oceanic and Planetary Physics,  
University of Oxford, Oxford, United Kingdom

<sup>2</sup>European Centre for Medium Range Weather  
Forecasts, Reading, United Kingdom

In September 2019 a workshop was held to highlight the growing area of applying machine learning techniques to improve weather and climate prediction. In this introductory piece we outline the motivations, opportunities and challenges ahead in this exciting avenue of research.

Ever since the development of the first fully programmable electronic digital computers at the end of World War 2, the cutting-edge of weather prediction science has lain in the development of accurate numerical representations of the governing dynamical equations of meteorology. As begun in the 1950s [Harper et al., 2007], these numerical representations have been further developed to allow first-principle simulations of climate. Since then, numerical modelling of weather and climate have dominated meteorological forecasting, to the extent that only 10 years ago prediction using statistical empirical models seemed rather antiquated.

With the recent advances in deep neural networks and other related machine learning techniques, statistical empirical prediction is firmly back in vogue. This raises a number of questions. Will the methods of artificial intelligence replace numerical models? Or, failing that, can they be used to supplement or enhance numerical models?

These were the overriding themes of a workshop held at Corpus Christi College Oxford in September 2019, papers from which are published in this Special Issue.

In order to address the questions posed above, it is worth breaking weather and climate prediction into different timescales.

Now-casting describes the evaluation of the weather and in particular precipitation right now or with a forecast lead time of a couple of hours. Here, conventional predictions are mostly based on observations and advection models with only limited representation of physical processes such as cloud physics or turbulence. Prudden et al. [2020] summarises the current state of the field, particularly in the context of machine learning.

Short-range weather prediction describes forecasts for a day or two ahead. On these timescales one may be interested in the timing of a cold front coming in or how individual thunderstorm complexes propagate and develop. This is a challenging problem as such cloud systems require very high-resolution modelling using spatial grids that have horizontal spacings on the order of 1 kilometre. Inevitably, given current computational power, this means short-range prediction models are regional and not global in their domain. The development of such convective cloud systems depends critically on the condensation of moisture, these frequently occur in updrafts too narrow to be properly resolved even with a 1 kilometre grid [Yang et al., 2016]. As a result, modelling on these space and timescales remains challenging.

Medium-range prediction, on timescales of a couple of days out to a couple of weeks, has perhaps more than any other timescale, seen the most dramatic increases in skill over past decades Bauer et al. [2015]. The archetypal phenomenon on the medium-range timescale is the baroclinic wave, whose characteristic scales are well resolved by contemporary global numerical weather prediction models with grid spacings on the order of 10 kilometres. Accurate prediction on these timescales over two generations of baroclinic waves is often possible.

Subseasonal forecasting spans timescales from two weeks to one season. A key source of predictability on this timescale is the Madden Julian Oscillation [Madden and Julian, 1971, Waliser et al., 2009]. This is an eastward-propagating equatorially trapped oscillation in tropical winds with a timescale of 30 – 80 days, associated with suppressed or enhanced organised convection. This timescale has seen much attention in recent years with the birth of the Subseasonal to Seasonal (S2S) Prediction Project [Vitart et al., 2017]. Associated with this project are several large databases containing S2S forecasts from operational and research models [Vitart et al., 2017, Pegion et al., 2019], invaluable for potential ML applications.

Seasonal forecasting depends critically on the coupled dynamics of the atmosphere and ocean and an archetypal phenomenon is the El Nino event in the tropical Pacific Ocean [Neelin et al., 1998, Timmermann et al., 2018]. Coupled ocean-atmosphere models show skill in prediction for El Nino on seasonal timescales [Johnson et al., 2019]. However, at this range, models must use grid spacings of many tens of kilometres. At these resolutions, models show substantial biases in key fields such as precipitation and wind. These biases arise from the need to parametrise key physical processes, notably associated with deep convective cloud systems.

Finally, climate-change prediction asks the question of how the statistics of weather changes as the greenhouse gas concentration of the atmosphere increases, on timescales of decades and

longer. Again, on these timescales, the numerical models (with grid spacings of 100 kilometres) exhibit quite substantial biases whose magnitudes are comparable with the climate-change signal one wishes to predict [Palmer and Stevens, 2019].

With this as background, let us ask the question: could prediction on any or all of these different timescales simply be replaced by AI? One might refer to this as “Hard AI”.

Now-casting applications appear to be a good starting point for Hard AI. On these timescales, physical constraints, such as conservation laws, can be ignored as errors will not accumulate significantly over a couple of hours. Furthermore, the wealth of Internet-of-Things (IoT) data, for example in the form of mobile phone data, is increasingly available and usable for weather predictions. Assimilating this data using conventional methods will be very difficult due to the high number of measurements and large errors. In promising recent work, machine learning methods are becoming competitive for short timescales (see for example S nderby et al. [2020]). These approaches combine the data assimilation and forecasting problems, directly using observations (e.g. satellites) as inputs to the prediction system.

For short and medium range predictions, an approach using Hard AI becomes more challenging. Skilful forecasts on these timescales would implicitly require the representation of the equations of motion and the interactions between features of the system (such as clusters of deep-convection, gravity waves and mid-latitude jets). It is in principle possible to learn the equations of motion from data using machine learning methods [Dueben and Bauer, 2018, Scher and Messori, 2019, Weyn et al., 2019], however challenges with this approach include satisfying conservation principles and stability of the simulations. Forecast skill is found to decrease after a few days, such that it is currently difficult to envisage such Hard AI outperforming and hence potentially replacing numerical models on medium-range timescales. However for short-range predictions this approach can produce skilful forecasts [Weyn et al., 2019] indicating potential for “Hard-AI” to compete with numerical models. An alternative approach which has shown some promise on short timescales is to combine analog forecasting (prediction using a database of past evolving weather patterns) with deep learning [Chattopadhyay et al., 2020].

One of the biggest challenges for “Hard-AI” on short and medium range timescales is the lack of high-quality data that is available for training. To train machine learning systems from conventional models would potentially generate improvements in high performance computing efficiency. This would, however, not improve prediction quality. If the model data could be generated at a higher spatial resolution than can be currently run within an operational forecasting window then there is scope for prediction improvements, however the computation power required to generate this training dataset would be large. To train from observations requires consistent and high-quality data which is only available since satellite measurements have been made, beginning with Spunik 1 in 1959 [Kuznetsov et al., 2015]. This restricts the training datasets to  $\mathcal{O}(10^4)$  days of training data. There is perhaps scope to pre-train on existing model data before final training on some observational datasets, but it is unclear if model and observational datasets are close enough for pre-training to be a valuable step.

For seasonal predictions, machine learning methods are more likely to beat conventional prediction systems. Here, the systematic errors of models are sufficiently large that AI may soon be competitive [Nooteboom et al., 2018, Ham et al., 2019]. However the same challenges of training datasets size still exist for the seasonal prediction problem.

Adopting the Hard AI approach on the climate change timescale is particularly difficult. The problem is inherently one of extrapolation, predicting climates not yet observed, a major challenge for machine learning techniques. One key question for climate change is whether cloud feedbacks are positive or negative. There is no indication from existing data that global cloud cover is increasing or decreasing [Norris et al., 2016]. Hence a single black-box AI scheme would struggle in giving us meaningful projections of the future. Our only hope are schemes which fully embrace the laws of physics. This could incorporate AI but likely as a sum of many elements rather than a single entity.

The potential for “Hard-AI” was discussed extensively at the workshop, and no consensus was reached as to whether it will be likely to outperform conventional tools for medium-range and climate predictions. In particular, participants with a focus on high-performance computing were optimistic regarding machine learning tools, since their computational power will jump by a large factor in the coming years as chips and computing systems are co-designed towards the need of deep learning. Furthermore, the speed of developments in machine learning techniques is breathtaking with  $O(100)$  papers published on artificial intelligence every day.

If at least some of the weather and climate applications cannot currently be replaced in total, how can machine learning methods help to still improve their predictions? Via the use of what we call “Medium AI” and “Soft AI”. Here, the intention is not to replace the entire model. It is rather the aim to improve forecasts using tools that have been learned from observations – Medium AI – or via the emulation of existing kernels of weather forecasting models – Soft AI – to allow for improvements in computing efficiency. Several papers in this Special Issue discuss such an application of AI techniques.

“Medium AI” should have a sufficient wealth of independent data points for each application. One approach uses machine learning as a postprocessing tool. This includes downscaling grid scale output to some point-specific location within a gridbox, bias correction, improving the representation of forecast uncertainty, and automating model post-processing such as feature detection for early warning of extreme events. These applications are fairly uncontroversial as they have a long-standing history – not necessarily for deep learning but certainly for the wider toolbox of machine learning [Vannitsem et al., 2020, Hewson and Pilloso, 2020]. One promising technique is the use of Generative Adversarial Networks (GANs) [Goodfellow et al., 2014] to increase spatial resolution of model or satellite data [Leinonen et al., 2020].

A second approach with “Medium-AI” is to replace components of the models with new tools that have been learned from observations or other data sources (such as high-resolution models). This is an attractive proposal, which could transform the representation of poorly understood, complex physical processes in our models. However this is hindered by the lack of knowledge of the coupling between grid-scale and sub-grid-scale processes. There is a view that parametrisation – reducing sub-grid variability to some sort of bulk formula – is an inherently ill-posed concept, which would set an upper-bound on the improvements possible from AI [Palmer, 2019]. There are also difficulties associated with using observational data (for example from satellites) for training of parametrisation schemes as data on all vertical levels of the model are required. This is typically only possible to achieve within data assimilation systems but the signal-to-noise ratio will be challenging when extracting information from observations. Nevertheless, there are some model components which are amenable to a “Medium-AI” approach, particularly those which are based on empirical relationships. An example is the scaling function used within Monin-Obukhov similarity theory to determine interactions between the atmosphere and the surface layer [Monin and Obukhov, 1954].

The Soft AI approach was first used by Krasnopolsky [1997], Chevallier et al. [1998] over twenty years ago. In particular, by training a neural net on some existing parametrisation scheme (e.g. long-wave radiation) we can replace the parametrisation with an AI scheme which will not be more accurate than the original scheme, but will arrive at the parametrised tendency with significantly reduced computational cost. The saved computational cost can then be reinvested to improve the resolution or complexity of the numerical model, which in principle should improve the accuracy of the model. This approach is, in our view, among the lowest hanging fruit in the applications of AI to weather and climate modelling. It should be noted that this approach still has challenges, for example Brenowitz and Bretherton [2018] found that multi-timestep training was necessary to produce a stable simulation.

Putting all this together, we believe that the real way forward is through a careful assessment of how AI-based schemes and numerical models can be synthesised to allow the production of next-generation high-resolution global weather and climate models running on exascale supercomputers. It can be expected that such models will, as far as possible, be used across all

timescales, from days to decades. With climate change affecting the nature and frequency of weather extremes, making many in the world even more vulnerable to natural hazards, the need for such enhanced prediction models has never been greater.

The weather and climate modelling community has a long history in the use of machine learning techniques, such as parameter estimations or the extraction of the leading dynamic modes. However, the recent boost in capabilities of machine learning, and in particular deep learning, and the recent advance in HPC resources that enable the use of machine learning on much more complicated problems, is yet to be fully exploited. Doing so requires the development of additional skill and infrastructure within the community.

## Workshop overview

In addition to presentations and posters, our workshop also included breakout groups. These discussed:

- How to strengthen the link between physical understanding and machine learning
- How to improve physical parametrisation schemes within models
- How to improve HPC efficiency of weather and climate models using machine learning
- How to improve the useability of model output using machine learning.

It was evident that there is a need for the community to develop machine learning solutions that are customised to the needs of weather and climate modelling, rather than relying on machine learning solutions that are mainly developed for other application areas such as image recognition. Such customised solutions would need to be able to impose physical understanding within the solution architecture. This is necessary to build machine learning tools of minimal complexity to optimise efficiency during training and inference. Furthermore, machine learning solutions for weather and climate applications would need to cope with changes of dynamic regimes due to climate change and would therefore need to be able to be trained outside of training regimes that are available in past weather. In an ideal but unlikely scenario, solutions would be able to accurately extrapolate in addition to interpolation, where standard deep learning thrives. Some promising machine learning approaches that were presented at the workshop focused on the correction of a signal rather than the prediction of the signal itself. This provides a natural connection between conventional tools (that are used to generate the signal) and machine learning tools (that are used to correct the signal) and makes solutions more resilient against errors as the machine learning signal becomes smaller. Another promising approach is the use of machine learning in uncertainty quantification. This includes both the use of approaches to represent uncertainty within models, for example via the use of multi-bin Markov chains, Generative Adversarial Networks (GANs) [Gagne et al., 2020], or the use of machine learning techniques to post-process the output of ensemble simulations [Rasp and Lerch, 2018, Grönquist et al., 2020].

To enable machine learning solutions to take up information from physical knowledge, or to at least obey conservation properties, a number of research projects were presented such as an adjustment of architecture designs, constraint terms in loss functions during training, genetic programming or regression techniques that learn coefficients for terms that were guided by physical understanding (equation discovery), or the more conventional approach to perform parameter optimisation, for example via the use of Approximate Bayesian Computing.

A number of infrastructure requirements to support machine learning applications within weather and climate modelling have been identified during the workshop. These include:

- To design standard methods that allow to link Python and Fortran programmes as most machine learning methods are trained and used within Python code (based on machine learning libraries such as TensorFlow [Abadi et al., 2015] or Keras [Chollet et al., 2015]) while weather and climate models are typically based on Fortran code. Such a link

between Python and Fortran would in particular be important for machine learning approaches that try out online learning within a running model. Some progress on this task has been made by [Ott et al. \[2020\]](#).

- To develop benchmark training datasets to allow for a qualitative comparison between machine learning approaches that are developed within the different scientific groups (see also [Rasp et al. \[2020\]](#)). The benchmark datasets would need to be tiered in terms of complexity including small datasets for conceptual testing up to the terabyte range to allow for tests of high performance computing efficiency during both training and inference. For most datasets that are available currently, the time frequency is not good enough to allow for training of machine learning solutions.
- To make better use of heterogeneous hardware when running weather and climate models. Most current weather forecasting models run on CPU hardware only, whereas machine learning solutions are typically faster on GPU hardware. This may generate tension for weather and climate computing centres with their own supercomputing facilities. Better adoption of heterogeneous hardware for weather and climate models would have benefits outside of machine-learned kernels as these models feature dense linear algebra which can be further accelerated with low numerical precision [[Hatfield et al., 2019](#)].
- To prepare data centres for larger data requests by machine learning users and to provide computing infrastructure that is optimal for training and inference close to the data.
- To develop vanilla solutions for machine learning in weather and climate applications that reduce the need for excessive screening of all architecture and training options within deep learning. It may even be possible to use and share pre-trained machine learning solutions between groups. In particular, machine learning solutions that are able to work on unstructured grids and with spherical geometry as well as the three dimensional structure of atmosphere and ocean models with changing grid spacings in the vertical are required.

It is still unclear how the workflow of weather and climate modelling will change in the coming 5–10 years. However, it is very likely that machine learning will be used within many components within the workflow. This will not necessarily make the workflow – which is already very complicated – easier and it will require machine learning resources such as special hardware and staff with experience in the design of machine learning tools. However, the use of machine learning within the workflow will likely allow a number of improvements in results.

This special issue contains a mix of new research and review articles. Most of the articles directly attack the problem of weather forecasting, with many in the community believing that this is a good testing ground before considering the problem of climate prediction. The new research articles mostly focus on what we have dubbed Medium or Soft AI, looking to improve parameterisation kernels of weather and climate models or optimise existing kernels. There are several papers concerning the problem of data assimilation, an important and expensive part of weather prediction. Contributions also take on the subject of probabilistic forecasting, a key approach to forecasting on all timescales, as well as the exciting topic of machine learning for post-processing (please see TOCITEHAUPT in this issue for an overview on post-processing approaches).

## References

- Kristine Harper, Louis W Uccellini, Eugenia Kalnay, Kenneth Carey, and Lauren Morone. 50th anniversary of operational numerical weather prediction. *Bulletin of the American Meteorological Society*, 88(5):639–650, 2007.
- Rachel Prudden, Samantha Adams, Dmitry Kangin, Niall Robinson, Suman Ravuri, Shakir Mohamed, and Alberto Arribas. A review of radar-based nowcasting of precipitation and applicable machine learning techniques. *arXiv preprint arXiv:2005.04988*, 2020.



- Jing Yang, Zhien Wang, Andrew J Heymsfield, and Jeffrey R French. Characteristics of vertical air motion in convective clouds. *Atmos. Chem. Phys. Discuss.*, 2016:1–48, 2016.
- Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.
- Roland A. Madden and Paul R. Julian. Detection of a 40–50 day oscillation in the zonal wind in the Tropical Pacific. *Journal of the Atmospheric Sciences*, 28:702–708, 1971.
- Duane Waliser, K. Sperber, H. Hendon, D. Kim, E. Maloney, M. Wheeler, K. Weickmann, C. Zhang, L. Donner, J. Gottschalck, W. Higgins, I. S. Kang, D. Legler, M. Moncrieff, S. Schubert, W. Stern, F. Vitart, B. Wang, W. Wang, and S. Woolnough. MJO simulation diagnostics. *Journal of Climate*, 22(11):3006–3030, 2009. doi: 10.1175/2008JCLI2731.1.
- F. Vitart, C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E. Fucile, M. Fuentes, H. Hendon, J. Hodgson, H. S. Kang, A. Kumar, H. Lin, G. Liu, X. Liu, P. Malguzzi, I. Mallas, M. Manoussakis, D. Mastrangelo, C. MacLachlan, P. McLean, A. Minami, R. Mladek, T. Nakazawa, S. Najm, Y. Nie, M. Rixen, A. W. Robertson, P. Ruti, C. Sun, Y. Takaya, M. Tolstykh, F. Venuti, D. Waliser, S. Woolnough, T. Wu, D. J. Won, H. Xiao, R. Zaripov, and L. Zhang. The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98(1):163–173, 2017. ISSN 00030007. doi: 10.1175/BAMS-D-16-0017.1.
- Kathy Pegion, Ben P. Kirtman, Emily Becker, Dan C. Collins, Emerson Lajoie, Robert Burgman, Ray Bell, Timothy Delsole, Dughong Min, Yuejian Zhu, Wei Li, Eric Sinsky, Hong Guan, Jon Gottschalck, E. Joseph Metzger, Neil P. Barton, Deepthi Achuthavarier, Jelena Marshak, Randal D. Koster, Hai Lin, Normand Gagnon, Michael Bell, Michael K. Tippett, Andrew W. Robertson, Shan Sun, Stanley G. Benjamin, Benjamin W. Green, Rainer Bleck, and Hyemi Kim. The subseasonal experiment (SUBX). *Bulletin of the American Meteorological Society*, 100(10):2043–2060, 2019. ISSN 00030007. doi: 10.1175/BAMS-D-18-0270.1.
- J David Neelin, David S Battisti, Anthony C Hirst, Fei-Fei Jin, Yoshinobu Wakata, Toshio Yamagata, and Stephen E Zebiak. ENSO theory. *Journal of Geophysical Research: Oceans*, 103(C7):14261–14290, 1998.
- Axel Timmermann, Soon-Il An, Jong-Seong Kug, Fei-Fei Jin, Wenju Cai, Antonietta Capotondi, Kim M Cobb, Matthieu Lengaigne, Michael J McPhaden, Malte F Stuecker, et al. El Niño–southern oscillation complexity. *Nature*, 559(7715):535–545, 2018.
- S. J. Johnson, T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni, L. Magnusson, S. Tietsche, D. Decremier, A. Weisheimer, G. Balsamo, S. P. E. Keeley, K. Mogensen, H. Zuo, and B. M. Monge-Sanz. Seas5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3):1087–1117, 2019. doi: 10.5194/gmd-12-1087-2019. URL <https://www.geosci-model-dev.net/12/1087/2019/>.
- Tim Palmer and Bjorn Stevens. The scientific challenge of understanding and estimating climate change. *Proceedings of the National Academy of Sciences*, 116(49):24390–24395, 2019.
- Casper Kaae Sørderby, Lasse Espenholt, Jonathan Heek, Mostafa Dehghani, Avital Oliver, Tim Salimans, Shreya Agrawal, Jason Hickey, and Nal Kalchbrenner. Metnet: A neural weather model for precipitation forecasting. *arXiv preprint arXiv:2003.12140*, 2020.
- Peter D Dueben and Peter Bauer. Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10):3999–4009, 2018.
- Sebastian Scher and Gabriele Messori. Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geoscientific Model Development*, 12(7):2797–2809, 2019.
- Jonathan A Weyn, Dale R Durran, and Rich Caruana. Can machines learn to predict weather? using deep learning to predict gridded 500-hPa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, 11(8):2680–2693, 2019.
- Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. Analog forecasting of extreme-causing weather patterns using deep learning. *Journal of Advances in Modeling Earth Systems*, 12(2):e2019MS001958, 2020.
- VD Kuznetsov, VM Sinelnikov, and SN Alpert. Yakov Alpert: Sputnik-1 and the first satellite ionospheric experiment. *Advances in Space Research*, 55(12):2833–2839, 2015.

- P. D. Nooteboom, Q. Y. Feng, C. López, E. Hernández-García, and H. A. Dijkstra. Using network theory and machine learning to predict el niño. *Earth System Dynamics*, 9(3):969–983, 2018. doi: 10.5194/esd-9-969-2018. URL <https://www.earth-syst-dynam.net/9/969/2018/>.
- Yoo-Geun Ham, Jeong-Hwan Kim, and Jing-Jia Luo. Deep learning for multi-year enso forecasts. *Nature*, 573(7775):568–572, 2019.
- Joel R Norris, Robert J Allen, Amato T Evan, Mark D Zelinka, Christopher W O'Dell, and Stephen A Klein. Evidence for climate change in the satellite cloud record. *Nature*, 536(7614): 72–75, 2016.
- Stéphane Vannitsem, John Bjørnar Bremnes, Jonathan Demaeyer, Gavin R Evans, Jonathan Flowerdew, Stephan Hemri, Sebastian Lerch, Nigel Roberts, Susanne Theis, Aitor Atencia, et al. Statistical postprocessing for weather forecasts—review, challenges and avenues in a big data world. *arXiv preprint arXiv:2004.06582*, 2020.
- Tim D Hewson and Fatima M Pillosu. A new low-cost technique improves weather forecasts across the world. *arXiv preprint arXiv:2003.14397*, 2020.
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- Jussi Leinonen, Daniele Nerini, and Alexis Berne. Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *arXiv preprint arXiv:2005.10374*, 2020.
- TN Palmer. Stochastic weather and climate models. *Nature Reviews Physics*, 1(7):463–471, 2019.
- Andrei Sergeevich Monin and Aleksandr Mikhailovich Obukhov. Basic laws of turbulent mixing in the surface layer of the atmosphere. *Contrib. Geophys. Inst. Acad. Sci. USSR*, 151(163):e187, 1954.
- V Krasnopolsky. A neural network forward model for direct assimilation of ssm/i brightness temperatures into atmospheric models. *Research activities in atmospheric and oceanic modeling*, 1997.
- F Chevallier, F Chérut, NA Scott, and A Chédin. A neural network approach for a fast and accurate computation of a longwave radiative budget. *Journal of Applied Meteorology*, 37(11): 1385–1397, 1998.
- Noah D Brenowitz and Christopher S Bretherton. Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, 45(12):6289–6298, 2018.
- David John Gagne, Hannah M Christensen, Aneesh C Subramanian, and Adam H Monahan. Machine learning for stochastic parameterization: Generative adversarial networks in the lorenz'96 model. *Journal of Advances in Modeling Earth Systems*, 12(3):e2019MS001896, 2020.
- Stephan Rasp and Sebastian Lerch. Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review*, 146(11):3885–3900, 2018.
- Peter Grönquist, Chengyuan Yao, Tal Ben-Nun, Nikoli Dryden, Peter Dueben, Shigang Li, and Torsten Hoefler. Deep learning for post-processing ensemble weather forecasts, 2020.
- Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <http://tensorflow.org/>. Software available from tensorflow.org.
- François Chollet et al. Keras. <https://keras.io>, 2015.
- Jordan Ott, Mike Pritchard, Natalie Best, Erik Linstead, Milan Curcic, and Pierre Baldi. A fortran-keras deep learning bridge for scientific computing. *arXiv preprint arXiv:2004.10652*, 2020.
- Stephan Rasp, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukayna Mouatadid, and Nils Thuerey. Weatherbench: A benchmark dataset for data-driven weather forecasting, 2020.
- Sam Hatfield, Matthew Chantry, Peter Düben, and Tim Palmer. Accelerating high-resolution weather models with deep-learning hardware. In *Proceedings of the Platform for Advanced Scientific Computing Conference*, pages 1–11, 2019.