Machine Learning for Earth System Observation and Prediction

Massimo Bonavita,[a] Rossella Arcucci,[d] Alberto Carrassi,[b,c] Peter Dueben,[a] Alan J. Geer,[a]

Bertrand Le Saux,[e] Nicolas Longépé,[e] Pierre-Philippe Mathieu,[e] and Laure Raynaud,[f]

[a] ECMWF, *Reading (UK)*

[b] *Dept of Meteorology and National Centre for Earth Observation, University of Reading, UK*

[c] *Mathematical Institute, University of Utrecht, The Netherland*

[d] *Data Science Institute, Imperial College London, UK*

[e] *European Space Agency, European Space Research Institute, Φ-lab, Italy*

[f] *Météo-France*

Corresponding author: Massimo Bonavita, massimo.bonavita@ecmwf.int

First ECMWF–ESA Workshop on Machine Learning for Earth System Observation and Prediction

What: ECMWF and ESA convened a workshop to explore the current status, prospects and opportunities in the application of Machine Learning/Deep Learning for Earth System Observation and Prediction.

When: 5-8 October 2020

Where: Online

Website: https://events.ecmwf.int/event/172/

1. Motivation

Almost 400 researchers from across the world joined the first ECMWF–ESA Workshop on Machine Learning for Earth System Observation and Prediction (ESOP), which was hosted by ECMWF and held online from 5 to 8 October 2020.

The workshop brought together experts from a variety of backgrounds to survey the status of the uptake of Machine and Deep learning (ML/DL) methodologies in the ESOP communities, evaluate their expected impacts and build a consensus view on the best way to realize the untapped potential of ML/DL in Earth System science.

Machine and Deep Learning techniques have made possible remarkable advances in an ever-growing number of disparate application areas, e.g. natural language processing, computer vision, autonomous vehicles, healthcare, finance and many others. These advances have been driven by the huge increase in available data and computing power as well as the emergence of more effective and efficient algorithms to extract relevant information from these data pools. Earth System Sciences have benefited greatly from well-known laws governing system behavior, such as the Navier-Stokes equations for fluid flow or the laws governing radiative transfer, and the success of the field has in many ways been based on its ability to translate these physical/dynamical laws into accurate numerical models. However, the field is expanding into areas that, due to the scales involved, are not easy to model from fundamental principles - this includes areas such as cloud and precipitation physics and biological processes; further there are areas where system parameters are heterogeneous on fine scales and hence are not fully-known - e.g. the Earth's surface, its vegetation, soil and hydrological properties. Earth System Sciences have arguably been latecomers to the ML/DL party, but interest is rapidly growing, as borne out by the increasing number of publications in the field which make use of ML/DL techniques and confirmed by the large number of

2

File generated with AMS Word template 1.0

participants to this workshop and to other similar events. Innovative applications of ML/DL tools are becoming increasingly common and the workshop was structured to reflect their increasing relevance for all aspects of the Numerical Weather Prediction (NWP) and Climate Prediction workflow (Fig. 1). The following report focuses on the main discussion points and outcomes in the areas covered by the Workshop.

2. The Workshop

The Workshop started with two introductory talks by David Gagne (NCAR) and Alan Geer (ECMWF) who provided high-level overviews of the current status of ML/DL in the Earth sciences and their connections to more traditional methodologies. David Gagne showed examples of how ML/DL models can be used as computationally efficient emulators of more complex models that would otherwise be unaffordable in standard weather/climate simulations. Alan Geer highlighted the similarities between ML/DL and data assimilation (DA), which underpins the creation of initial conditions for weather forecasting, geophysical retrievals from satellites, and many other applications in the Earth sciences. Both ML/DL and DA can be seen as inverse methods based on Bayes' theorem, with a clear potential for cross-fertilisation, which was one of the key aspects explored in the workshop (see section b). The Earth sciences have for many years been developing DA to infer not just geophysical states, but also uncertain model parameters - this is known as parameter estimation. ML/DL shows ways to extend this towards full inference of models, although the many practical difficulties of parameter estimation would affect ML/DL too. Future developments in the use of observations that might benefit from ML/DL were also covered, including internet of things (IoT) observations such as mobile phone pressures, and the growing data volumes from space platforms, as sensors and satellites can increasingly be miniaturized and deployed in ever greater numbers

3

File generated with AMS Word template 1.0

The main body of Workshop was structured in four sessions with talks covering the building blocks of the NWP/Climate prediction workflow (observations, data assimilation, models, ensembles and product development) and a Poster Session where presenters and participants could meet in digital rooms to discuss their research. Oral presentations recordings and slides, and poster presentations are available on the Workshop website (https://events.ecmwf.int/event/172/).

a. Machine Learning for Earth System Observation

The first session focused on the potential of ML/DL to improve the exploitation of available and future observations of the Earth system.

In areas of Earth observation relying on image processing, ML/DL technologies are well established as existing network designs and strategies from the wider community can be applied with often minimal adaptations. Several talks outlined the importance of ML/DL in providing insights from an increasingly vast stream of data from satellites. Pierre-Philippe Mathieu addressed "the rise of AI for Earth Observation (EO)" by presenting examples such as physics-aware AI, i.e. integration of physical domain knowledge in the statistical formulation of ML/DL models, or trustworthy and explainable AI which goes beyond the "black-box" vision often associated with these models. Begüm Demir explained how data mining techniques allow for leveraging the large archives of EO data through image retrieval with query by example or automatic description ("captioning") of scenes and landscapes seen from above (Hoxha et al., 2020).

Earth system observations can be difficult to use – they are typically sparse, affected by uncertainties and made using indirect methods, often with substantial nonlinearities. Beyond image processing, e.g. when applications require geophysical retrievals or aim to produce initial conditions for NWP forecasts, these issues have driven the development of inverse

4

File generated with AMS Word template 1.0

methods like DA. ML may need to adapt substantially to deal with sparse, uncertain and indirect earth system observations, taking inspiration from DA. However, machine learning still has numerous potential applications in this area. For example, deep learning tools could emulate satellite observation operators and retrieval methods; ML/DL could also be used to learn these components where physical models are not yet available. Examples include the retrievals from the Soil Moisture Ocean Salinity (SMOS) satellite that are assimilated operationally at ECMWF, as outlined in Peter Weston's talk. Here, in a method akin to transfer learning, an existing Neural Network has been re-trained with the observations and the ECMWF model soil moisture, reducing the possibility of assimilating biased observations. Imme Ebert-Uphoff and David J. Gagne focused on the need for interpretation of neural network prediction in the meteorological domain, or "making the black box more transparent". This relies on understanding the physical implications of machine learning on the one hand (McGovern et al., 2019), and understanding the neural network behaviour and the correlations found between the data and the output on the other hand (Ebert-Uphoff and Hilburn, 2020). There are also applications in quality control, data monitoring and observation bias correction; these featured heavily in the discussions of the observations working group showing that research in all these areas is very active.

b. Machine Learning for Data Assimilation

Data assimilation and machine learning (ML) have much in common: they both ultimately try to describe and model the system of interest by using data. Data assimilation focuses on time-evolving systems, it has been primarily used to estimate the system's state and its uncertainty and is known for its ability to combine effectively available data with a given, inevitably imperfect and incomplete, model. Machine learning is generally purely data-driven, it does not rely on any prior knowledge of the underlying process and is not

5

File generated with AMS Word template 1.0

exclusively applied to time-evolving systems. Can DA and ML be mutually helpful? What are their respective strengths and weaknesses?

These were some of the general questions that the talks and posters tried to give answers to, and which were then further discussed in the dedicated Working Group. Three main research areas exploring the connections between DA and ML were identified: (i) Integrate ML and DA, (ii) Combine ML and DA, and, (iii) Unify ML and DA.

The first area refers to the pragmatic approach of exploiting ML as an alternative way to provide some key components of the DA system and/or vice-versa. In this context one research direction is to preserve the DA system as much as possible but to incorporate ML tools to improve or make some critical components of the DA system faster/more efficient. Examples include using different ML architectures to improve/estimate the observation operator or to provide a parameterization of the background errors operator, and even to try to improve the solver itself by, e.g. devising data-driven tangent linear and adjoint models, or to "learn" the Kalman gain in an ensemble Kalman filter framework. Another development direction is to employ DA in the training process of a ML model to improve the accuracy of the resulting data-driven models. The use of DA in the training process of Generative Adversarial Networks (GANs) to improve the accuracy of a resulting surrogate model is an example. The talks from Rossella Arcucci, Takemasa Miyoshi, Ronan Fablet, and the posters of Shigenori Otsuka and Andrea Storto provided examples of these developments.

The second area describes attempts to combine DA and ML in a hybrid configuration. While DA provides efficient tools to handle noisy and sparse observations in conjunction with a (usually imperfect) model, ML does not need a model but it requires the data to be highly accurate as well as spatially and temporally sufficient to describe all the relevant degrees of freedom of the system. Alberto Carrassi proposed combining DA and ML in a

6

File generated with AMS Word template 1.0

hybrid, iterative framework where DA is used to assimilate sparse and noisy data to obtain the mean and covariance of the analysis pdf. These estimates of the state are then fed to the ML step where a model of the dynamical system is retrieved. The same approach can also be used to infer model parameterizations or to estimate the model error as discussed in the posters of Alban Farchi and Arata Amemiya and in the talk by Massimo Bonavita. Another venue of cross fertilization between DA and ML was presented by Manuel Pulido where expectation maximization algorithms common in ML were used to estimate hyper-parameters, e.g. the error covariances, needed in DA, and by Pieter Houtekamer who showed how to embed a genetic ML algorithm in an operational ensemble data assimilation system to obtain improved estimates of the prognostic model parameters.

The third area includes contributions that have identified methodological analogies between DA and ML, or that have proposed unifying the two families of methods under a common theoretical formalism. For instance, Massimo Bonavita pointed out that, from a methodological perspective, ML/DL can be viewed as a particularization of DA to the task of estimating the underlying model dynamics and that some existing DA methods, e.g. weak-constraint 4D-Var, already perform a form of online ML of the model errors. The contribution from Marc Bocquet discussed combining DA and ML under a more comprehensive and general Bayesian formalism. Peter Jan van Leeuwen's talk showed how to embed Deep Learning in a Bayesian framework so as to equip it with the ability to provide uncertainty estimates of the learned model. Anthony Fillion's poster presentation proposed a fully data driven DL architecture based on generalizing recurrent Elman networks.

The discussions in the Working Group dedicated to these topics highlighted several concrete challenges. These include data format compatibility, ML for model reduction (including

7

preconditioning), and the implementation/incorporation of ML approaches within existing DA systems.

Available data of the Earth systems are sparse, noisy and/or not fully representative. It was discussed how to equip DA with ML-derived observation operators. A possible avenue is to improve, or to fully estimate, ML-based observation operators, e.g. using AutoEncoders (AE), thanks to their capability to represent nonlinearities. AEs are also very promising for dimensionality reduction, as has been already shown in various applications, also compared to more traditional linear approaches such as Truncated Singular Value Decomposition (TSVD), Principal Component Analysis (PCAs), etc.. Nevertheless, AEs are not yet very efficient for problems with the typically very large dimensions encountered in DA applications in the geosciences and may require the prior identification of a reduced space with physical interpretability.

Another important concern is the ensemble construction, relevant for both ensemble-based DA and ensemble prediction systems. Here the main idea is to use AEs to shift the ensemble generation process to a much smaller subspace which is able to capture the main modes of variability of the geophysical system.

Finally, the discussion centred on exploring the potential of "online learning" such that the result of the training process can be continuously updated as new data become available without the need for offline retraining of the whole ML-model.

c. Machine Learning for Weather and Climate Models

This section of the Workshop was devoted to understanding what ML/DL can do to improve current geophysical models, with a specific focus on weather and climate models. A promising application is in the emulation of specific model components, such as the radiation, gravity wave-drag and boundary layer turbulence parameterization schemes, in

8

File generated with AMS Word template 1.0

order to exploit the computational efficiency of ML/DL technologies and obtain speed-up and energy efficiency gains. Examples of this type of application were provided in the contributions from Leyi Wang, Peter Dueben, Peter Ukkonen and Matthew Chantry. An extension of this idea is to train a NN to emulate the full set of physical parameterizations, as shown e.g. by Alexei Belochitski in an application aimed at reproducing the physics parameterizations of the NCEP GFS atmospheric model, and by Christiane Jablonowski in a hierarchy of GCMs of increasing realism and complexity.

The logical next step is to use NN to emulate the full model. This idea is enticing but it comes with its own caveats, as ML models are statistical models and they are exposed to the curse of dimensionality (see, for example, the recent discussion in Bonavita and Laloyaux, 2020). For this reason, applications have mainly focussed on reduced order models (Yang Liu, Maha Mdini), or at forecasting a small set of weather parameters at low resolution (Ashesh Chattopadhyay, Dale Durran, Jonathan Weyn) for medium-range to seasonal forecasting ranges. Another, related idea, is to use ML to forecast weather parameters where standard NWP products show more room for improvement. A case in point is the forecasting of precipitation, which a number of contributors showed it can be improved with the use of NN models (Shigenori Otsuka, Duncan Watson-Parris, Mingming Zhu, Carlos Gonzalez, Jie Xiahou).

A hot topic in the application of ML/DL in NWP and Climate Prediction is the use of physics-informed, interpretable and trustworthy ML methods. Imme Ebert-Uphoff and Vipin Kumar explained that such methods would allow our understanding of the Earth system to be leveraged to develop customized ML tools that fit the needs of the community, instead of using black/grey-box ML tools whose results are difficult to interpret and for which it will be difficult to convince domain scientists about their trustworthiness. .

9

File generated with AMS Word template 1.0

d. Machine Learning for Product Development and ensemble processing

ML/DL can be used for extracting and enhancing information from the NWP/Climate prediction workflow through the development of tailored, custom-oriented products. Laure Raynaud presented the potential of Convolutional Neural Networks (CNN) to perform the automatic detection of weather "objects" such as atmospheric fronts and tropical cyclones in NWP model outputs. Ryan Lagerquist showed that CNN can also be used for the prediction of next-hour tornado probability, and different diagnostics and tools for ML interpretation were presented. Sue Ellen Haupt discussed the use of ML for the production of weather products tailored for the renewable energy industry. In the contribution of Claire Monteleoni, ML/DL was shown to be a promising way to perform the statistical downscaling (called Super Resolution in the Computer Vision community) of low-resolution temperature and precipitation forecasts.

The post-processing of ensemble forecasts is another promising and well-advanced application area which was widely covered in both the oral and the poster sessions (e.g., contributions from Nikoli Dryden, Daniele Nerini, Sebastian Lerch). ML/DL was shown to be able to provide a flexible, data-driven modelling of nonlinear relations between predictors and the distribution parameters of the predictands, with results already competitive with state-of-the-art approaches.

Other potential DL-based applications have been mentioned, such as the use of Natural Language Processing (NLP) for the automatic generation of weather bulletins or chatbot to interact with end-users.

The current uptake of ML/DL methodology for post-processing is still limited in operational settings. In the short term, we would expect ML/DL to have the largest impact where nonlinear methods are needed. Post processing of forecasts (in the sense of calibration)

10

File generated with AMS Word template 1.0

is likely to be the first application to benefit from these methodologies and reach operational maturity (we note that benchmarks already exist in this domain). In the longer term, ML/DLmethods may help in the design of Ensemble Prediction System (EPS) where a tradeoff is required between increased horizontal resolution and increased ensemble size.

Several key challenges have also been identified. In particular, standard ML frameworks need to be adapted to the specific needs of the community. For instance, a concern was raised about the blurry effect of CNN-based downscaling which should be avoided for an operational application. The specification of a dedicated loss function might be required to solve this problem. Beyond the lack of labeled datasets for pattern detection, the need for homogeneous data over a long forecast period is important. As post-processing ML/DL algorithms rely on weather model outputs, the necessity to retrain the ML/DL model whenever the weather models are updated may pose a potential issue. Similarly to the other application areas, the need for interpretability and explainability of ML/DL-based solutions in post-processing applications was deemed a crucial factor for their wider adoption.

3. Key outcomes and next steps

The oral and poster presentations sparked interest and lively discussions among the participants during the four working groups covering the main areas of the application of ML/DL to NWP and climate. The main findings of the working groups were discussed in the final plenary session and are available on the Workshop website. The overarching conclusion was that the field is in rapid development and there is huge interest in the ESOP communities. It was also felt that progress will be fast as domain scientists increase collaboration with ML/DL specialists and at the same time ML/DL expertise becomes more widespread. This second aspect is particularly important. The evolution of ML/DL has been

File generated with AMS Word template 1.0

historically characterised by an evolution from general-purpose Dense Neural Networks to more tailored Neural Networks architectures which exploit prior knowledge about the problem at hand to build more effective and efficient ML models (Goodfellow et al., 2016). In a similar way ML/DL technologies need to be adapted and tailored to the specific ESOP applications and this is where in-depth domain knowledge from ESOP scientists is deemed to be crucial for the success of the ML/DL application.

To check on the pace of progress and further strengthen the community effort, ECMWF and ESA have provisionally scheduled a second, follow-on workshop for Q4 2021. Details will be announced in due course.

REFERENCES

Bonavita, M., & Laloyaux, P. (2020). Machine Learning for Model Error Inference and Correction. *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002232. Accepted Author Manuscript. https://doi.org/10.1029/2020MS002232

Ebert-Uphoff, I., and K. Hilburn, 2020: Evaluation, Tuning and Interpretation of Neural Networks for Working with Images in Meteorological Applications. *Bull. Amer. Meteor. Soc.*, doi: https://doi.org/10.1175/BAMS-D-20-0097.1.

Goodfellow, I., Bengio, Y. and A. Courville, 2016: *Deep Learning*, MIT Press.

Hoxha G., Melgani F., and Demir B., 2020: Toward Remote Sensing Image Retrieval under a Deep Image Captioning Perspective, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, https://doi.org/10.1109/JSTARS.2020.3013818

McGovern, A., R. Lagerquist, D. John Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the Black Box More Transparent: Understanding the

12

File generated with AMS Word template 1.0

Physical Implications of Machine Learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199,

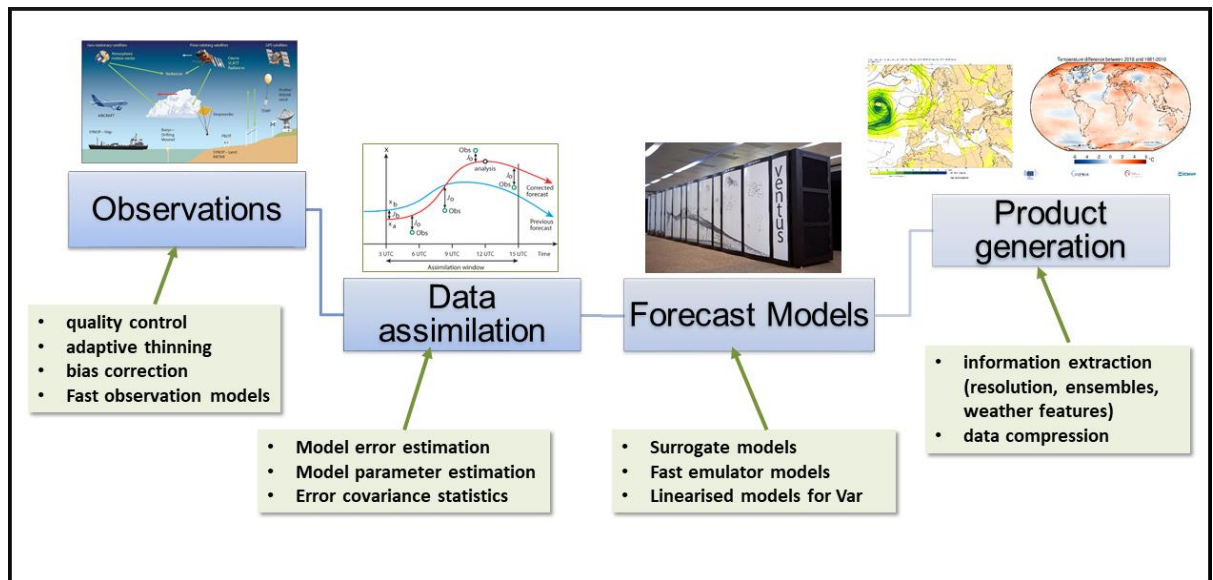https://doi.org/10.1175/BAMS-D-18-0195.1.

FIGURES

Figure 1: Examples of possible Machine Learning applications in the various components of

a standard NWP workflow.