

MEETING REPORT OPEN



ESA-ECMWF Report on recent progress and research directions in machine learning for Earth System observation and prediction

Rochelle Schneider^{1,2✉}, Massimo Bonavita^{1,2}, Alan Geer^{1,2}, Rossella Arcucci³, Peter Dueben², Claudia Vitolo¹, Bertrand Le Saux¹, Begüm Demir⁴ and Pierre-Philippe Mathieu¹

This paper provides a short summary of the outcomes of the workshop on Machine Learning (ML) for Earth System Observation and Prediction (ESOP / ML4ESOP) organised by the European Space Agency (ESA) and the European Centre for Medium-Range Weather Forecasts (ECMWF) between 15 and 18 November 2021. The 4-days workshop had more than 30 speakers and 30 poster-presenters, attracting over 1100 registrations from 85 countries around the world. The workshop aimed to demonstrate where and how the fusion between traditional ESOP applications and ML methods has shown limitations, outstanding opportunities, and challenges based on the participant's feedback. Future directions were also highlighted from all thematic areas that comprise the ML4ESOP domain.

npj Climate and Atmospheric Science (2022)5:51 ; <https://doi.org/10.1038/s41612-022-00269-z>

INTRODUCTION

The purpose of this workshop, hosted by the European Space Agency (ESA) in collaboration with the European Centre for Medium-Range Weather Forecasts (ECMWF), was to bring together a diverse community interested in the fusion between traditional Earth System Observation and Prediction (ESOP) applications and Machine Learning (ML) methods. This report summarises the range and depth of the discussions captured during the workshop and highlights the current limitations, challenges, and opportunities in ML4ESOP.

This second workshop edition¹ was divided into 4 days (15–18 November 2021). The first 3 days were mainly devoted to 33 oral presentations from experts across four Thematic Areas (TA): (1) Enhancing Satellite Observation with ML, (2) Hybrid Data Assimilation—ML approaches, (3) Geophysical Forecasting with ML and Hybrid Models, and (4) ML for Post-Processing and Dissemination. The workshop also hosted a live e-poster session with 30 poster presentations hosted in separate virtual meeting rooms to foster networking and nurture new collaborations. On the last day, it was the time to reverse the order and listen to the participants, coming from both academic and industry backgrounds with rich experiences and expertise on current ML methods for ESOP applications. The working groups split into TAs brought a diverse community to discuss the advantages and limitations of ML in comparison with more traditional methods and outline future directions.

The workshop opened with Pierre-Philippe Mathieu (Head of ESA Φ -lab Explore Office) and Andy Brown (ECMWF Director of Research) providing the vision of developments to enable both ESA and ECMWF's Member and Co-operating States to benefit from ML advances in satellite observations, weather, and climate modelling. Following the steps outlined in the introductory remarks, Devis Tuia (Associate Prof at Swiss Federal Institute of Technology Lausanne and visiting Prof at ESA Φ -lab) and Alan Geer (Principal Scientist at ECMWF) kicked off the scientific talks

by giving complementary overviews on how ML capabilities are currently being investigated and applied at ESA and ECMWF, respectively, both as internal projects and in collaboration with external ML experts.

The importance of advancing on explainable ML tools was highlighted by Dr Tuia, which referred to ML methods where humans can make interpretations beyond predictions (since ML tools are often perceived as 'black-boxes') and understand the inner-functioning of the model, the internal decisions from a given parameter definition to the final model output. It was also stressed that the distribution of Deep Learning (DL) methods in remote sensing applications is very skewed towards a well-resolved problem: feature detection (e.g. building mapping, road extraction, etc). According to Tuia and colleagues², a new ML in Earth Science agenda has arrived to revolutionise the value extracted from Earth Observation (EO) satellite images/videos, for example: event recognition (cultural events vs manifestations), human feeling detection from their landscape perception, and building permission control based on text mining from urban planning regulations.

Dr Geer also presented how essential EO products are to data assimilation systems, providing the initial conditions and parameter estimates of the geophysical atmospheric state to describe complex physical dynamics that are needed to make geophysical forecasts. The incorporation of ML methods into data assimilation can attempt to emulate the whole or part of the dynamical system; therefore, bringing new capabilities to the process.

The following sections will describe in more detail the discussions on each TA covering three key-topics: (1) Current ML applications, (2) Limitations, opportunities, and challenges (3) Future directions. The workshop content (recordings, slides, and e-posters) is available on ECMWF webpage: <https://events.ecmwf.int/event/291/timetable/>.

¹European Space Agency, Frascati 00044, Italy. ²European Centre for Medium-Range Weather Forecast, Reading RG2 9AX, UK. ³Imperial College London, London SW7 2AZ, UK. ⁴Technische Universität Berlin, 10587 Berlin, Germany. ✉email: rochelle.schneider@esa.int

TA1: ENHANCING SATELLITE OBSERVATION WITH ML

Current ML applications

The working group team, chaired by Begüm Demir (Prof. at Technische Universität Berlin) and Bertrand Le Saux (Senior Scientist at ESA) outlined the use of ML/DL methods (e.g. ensemble methods, random forests, convolutional neural networks, etc) for Earth surface monitoring, such as forest and biomass estimations³. Radar backscatter (e.g. ESA's Sentinel 1 C-band) and optical images (e.g. ESA's Sentinel 2 Multispectral Instrument) were described as the relevant EO products to estimate the forest ecosystem inventory. Volcanic plume monitoring was also mentioned, it can be either observed by geostationary and low-earth orbit satellites (e.g. MSG-SEVIRI and Sentinel 5P) or simulated/tracked (i.e. its ash and gases dispersion) by chemical transport models (e.g. Copernicus Atmosphere Monitoring Service—CAM5). The exploration of ML can be linked to several volcanic-related topics, for example, retrieving ash components and Sulphur Dioxide layer height using Full-Physics Inverse neural networks⁴. Additionally, it also discussed the power of ML-regression approaches to estimate ice sheet mass balance by using EO data. Therefore, open access to EO data (e.g. Copernicus Services⁵) brings a lot of benefits, advancing ML techniques to EO application.

Limitations, opportunities, and challenges

When discussing limitations, the participants mentioned ground-truth reliability issues during the ML training and validation processes, which might require the use of weakly-supervised learning⁶ or semi-supervised strategies⁷. Dealing with big datasets⁸, quite often required to satisfy the learning process of some ML algorithms, has been pinpointed together with storage and processing power as constraints.

Transfer learning (TL), a relevant approach across ML/DL applications, has been classified by the participants as a limitation and an opportunity. TL means applying a trained-ML model to different geographical regions or temporal periods to the same or a similar problem. Some experts believe that this extrapolation to unknown regions or timestep could generate highly unreliable estimates. At the same time, this knowledge transfer is claimed to save time and resources by not having to train many ML models from the beginning to perform similar tasks elsewhere. Additionally, it can also fill geographical data gaps caused by a lack of training data.

Many challenges were discussed during the working group session. Starting with uncertainty in ML outputs which comes from many sources (noisy data, deficient sample size for training, and model imperfection) that need to be properly quantified. The known 'black-box' challenge in ML was also highlighted by some participants. The lack of human interpretability has triggered researchers to develop tools to comprehend better and trust decisions made by a ML algorithm to specific parameter values, model design, and estimates. These tools are known as explainable Artificial Intelligence (AI)⁹.

Finally, the combination of ML with physics-based models of the sensors¹⁰ and the observed systems was considered a priority: ML methods are very task-oriented and may have difficulties making predictions about physical processes (e.g. volcanic activity) since they lack prior knowledge about the system they want to describe. The latter will be covered in more detail in the TA3.

Future directions

In the TA1 context, one research direction recognised as a sustainable solution is TL approaches. TL still carries some limitations and challenges, but its generalisation capacity is a real game-changer for locations with not enough data for training,

providing a feasible solution to several applications (e.g. food security, climate change mitigation).

TA2: HYBRID DATA ASSIMILATION—ML APPROACHES

Current ML applications

The working group team, chaired by Rossella Arcucci (Assistant Prof at Imperial College London) and Alan Geer (Principal Scientist at ECMWF) outlined the benefit of using ML, especially neural networks, with data assimilation in terms of both accuracy and efficiency. Neural networks show great capability in approximating nonlinear systems and extracting meaningful features from high-dimensional data. These properties could be very useful in data assimilation applications. Autoencoders are of interest for dimensionality reduction, and an emerging application is to convert non-Gaussian problems into Gaussian problems consistent with using traditional data assimilation methods. For building surrogate models, the neural networks can learn the dynamics behind the data, and they also can be used in the prediction-correction data assimilation cycle¹¹, or for example either to replace a physically based model entirely or to apply error corrections^{12–14}, in four-dimensional data assimilation models. The hybridisation of data assimilation and neural networks is expected to produce both faster and more accurate assimilation-prediction systems. Helping to merge data assimilation and ML is the fact that both fields are inverse methods that can be united under a Bayesian framework¹⁵.

Limitations, opportunities, and challenges

Within data assimilation, the observation operator links the geophysical variables of interest (e.g. sea-ice fraction, or winds) to the observed quantities (e.g. satellite-observed radiance). Often there is no adequate physically based operator so ML could help create empirical observation operators. A challenge is that full training datasets do not exist, since observations are often sensitive to variables that no model can completely simulate. Likely a generative approach is needed in which the unknown physical variables are represented in a latent space.

When creating empirical forecast models, the need to use physical constraints is clear, and the techniques are increasingly available, for example, additional physical layers or terms in the loss function. Other issues concern extrapolation in a multiple regime chaotic system—can a ML model trained in one regime extrapolate to another? But we could also start using ML models to understand and predict the regime-dependent predictability of the system.

Future directions

In the near future, the use of ML for model error estimation and bias correction will be important. An imperfect physical model can be retained, but augmented by a neural network that learns to apply a state-dependent error correction at every timestep of the model¹⁶.

Another interesting application is causality, even though ML is a nonlinear extension to linear methods like correlation, with the same issue that it can learn associations or patterns rather than causality. However, some ML techniques attempt to learn causal relationships such as between climate variables.

Neural networks can require large amounts of data before they begin to produce reliable results, and the larger the architecture, the more data are needed. In addition, if the data available are too noisy, too scarce or there is a lack of salient features to represent the problem, the network will not perform well. Physics-informed neural networks (PINN) are neural networks that solve supervised learning problems with the constraints of given laws of physics

described by general nonlinear partial differential equations¹⁷, also helping to reduce the amount of training data required.

Standard data assimilation and neural network can have problems with unstructured or even adaptive meshes. But graph neural networks are expected to start including unstructured and adaptive meshes with varying numbers of nodes.

TA3: GEOPHYSICAL FORECASTING WITH ML AND HYBRID MODELS

Current ML applications

The working group team, chaired by Claudia Vitolo (Senior Scientist at ESA) and Peter Dueben (Coordinator of ML and AI activities at ECMWF), outlined that there are various types of ML and hybrid model applications that range from applications to speed up complex and time-consuming processing (e.g. model emulators¹⁸) to applications that are focused on diagnostics (e.g. unsupervised learning and causal discovery¹⁹) and model improvements (e.g. learning from observations and high-resolution data²⁰), or for uncertainty quantification and error analysis). ML also allows learning the entire model from data. This application is the most controversial as ML scientists are confident that they will soon be able to beat operational weather models²¹, while most domain scientists are convinced that pure ML models will not beat conventional models any time soon.

Limitations, opportunities, and challenges

Advantages of DL are the high computational efficiency (due to dense linear algebra and reduced numerical precision), the simple arithmetic that makes tools very flexible (for example regarding automatic differentiation), and the flexibility to combine information from various data sources, including Internet of Things (IoT) data. It can also be an advantage that DL tools can be used as black-box when the underlying physical processes are not understood. Advantages of statistical approaches beyond DL are the sound mathematical formulation and (for some of the approaches) the ability to extract physical understanding. A number of successful examples of the above applications were presented during the dedicated oral session.

The main limitation is the limited access to datasets. When available, licence restrictions may limit their use. From a modelling perspective, some Earth System processes are rare (e.g. extreme events such as tropical cyclones) and difficult to characterise statistically. In classification algorithms, for instance, such imbalance can be misleading and cause over-confidence. There is a need for more AI-ready datasets (so-called benchmark datasets), improvements to make data handling easier, and access to more pre-trained-ML models that only need to be customised for a specific application.

Challenges are the lack of trust in ML tools due to the inability to extrapolate into unseen weather regimes and limited generalisation. It requires more work to facilitate the use of ML tools in operational weather and climate predictions. Furthermore, data ethics are relevant for IoT data²².

Future directions

The field will grow further and benefit from the availability of customised ML hardware, improvements in explainable AI and physics-informed ML, established benchmark datasets for weather and climate applications, more open datasets, more transfer and re-inforcement learning, and advancements in software—for example for the coupling of conventional and ML tools and the use of ML at scale. ML will therefore improve all relevant components across the workflow of weather and climate modelling. There will also be more scientists who can do both, ML and Earth system science, who will help to link the two

communities. Further opportunities could be made available under the European Digital Strategy, as in the case of the *Destination Earth* initiative (coordinated by the European Commission and jointly implemented by ESA, ECMWF and EUMETSAT) and the application of transfer learning.

TA4: ML FOR POST-PROCESSING AND DISSEMINATION

Current ML applications

The working group team, chaired by Rochelle Schneider (Researcher at ESA) and Massimo Bonavita (Senior Scientist at ECMWF) discussed with participants from a widely diverse professional background, ML and ESOP experience, and point-of-view for the use of ML in the TA4 area. Participants from different fields such as operational weather services, energy (e.g. wind and solar forecasting) and hydrology (e.g. flood forecasting²³) sectors shared a common interest in using ML for post-processing to optimise their forecast modelling system and provide predictions from the short-range (hours-to-days) to the extended range (months) up to a seasonal outlook. Other participants made an interesting observation about the use of ML methods to downscale products from global/local climate and atmosphere models to improve the meteorological information provided to on/offshore wind farms. This super-resolution ML application was acknowledged also to benefit research studies on environmental health^{24,25} and sustainable cities which need weather and air quality data at high spatio-temporal resolution at surface level^{26,27}.

Limitations, opportunities, and challenges

Not surprisingly, and similarly to the findings of other working groups, heterogeneous ground-truth distribution and computational costs for model training were identified as potential challenges. A rich discussion started when one of the participants brought the “benchmark” concept into the conversation. The interesting situation here is that this topic was reported with three different meanings. Currently, there is not a consensus view in the literature on the proper way to score different ML models designed to do the same (or similar) tasks². This situation suggests the need to develop a standard benchmark that would help future researchers to progress their model designs based on the performance found in previous studies. Finally, the implementation of a one-fits-all benchmark method is also challenging to validate different ML model designs. The discussion also raised the issue of the lack of published information available to interpret the model architecture, and the sample size used to generate the reported performance of the trained-ML models.

The chairs asked the participants working in the private sector why ML methods are not yet commonly implemented into their services. The answer was that many ML frameworks published in the literature are difficult to replicate due to the lack of basic information (e.g. dataset and code publicly available). Additionally, this reluctance on the operational/user services to explore ML approaches was explained by the strong interpretability and trustworthiness of (benchmark) statistical methods, and the concern about possible service disruptions due to unforeseen ML model issues.

Future directions

ML methods are known to be affected by generalisation issues, i.e. having problems in dealing with outliers from the training distribution. Still, in a changing climate, one of the most important issues is predicting extreme events, in which case the predicted state will be located in the tails (or even outside) of the training distribution. For these reasons, the working group emphasised the need to adjust the ML models to widen the magnitude of their

prediction range to capture severe events due to their catastrophic impact on society and the economy. Finally, some participants were sceptical about the use of the digital twin-engine to deliver high spatio-temporal resolution data since post-processing and downscaling approaches (linked to ML methods) could perform these tasks more efficiently.

Final remarks

The ML4ESOP attracted over 1100 registrations from 85 countries around the world, with a large number of participants from Germany, Italy, and the United Kingdom. The opening and closing sessions were broadcast live by ESA Web TV to an audience not registered at the workshop. They captured more than 1200 views during the opening session, followed by many requests on social media and emails to the organisers to join the event.

These numbers indicate the success of the workshop and confirm that there is interest to run another edition in 2022. This report provided evidence of the valuable exchange of ideas across ML and ESOP communities. More significantly, it has reinforced the call to produce replicable, explainable, and sustainable ML methods. ML4ESOP plays a fundamental role also in terms of end-users, those who will use the modelling outputs to drive economic, political, and health decisions.

Received: 14 April 2022; Accepted: 13 May 2022;

Published online: 21 June 2022

REFERENCES

- Bonavita, M. et al. Machine learning for earth system observation and prediction. *Bull. Am. Meteorol. Soc.* **102**, E710–E716 (2021).
- Tuia, D. et al. Toward a collective agenda on AI for earth science data analysis. *IEEE Geosci. Remote Sens. Mag.* **9**, 88–104 (2021).
- Li, Y., Li, M., Li, C. & Liu, Z. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Sci. Rep.* **10**, 9952 (2020).
- Hedelt, P., Efremenko, D. S., Loyola, D. G., Spurr, R. & Clarisse, L. Sulfur dioxide layer height retrieval from Sentinel-5 Precursor/TROPOMI using FP_ILM. *Atmos. Meas. Tech.* **12**, 5503–5517 (2019).
- Copernicus Services. <https://www.copernicus.eu/en>.
- Daudt, R. C., Le Saux, B., Boulch, A. & Gousseau Y. Weakly supervised change detection using guided anisotropic diffusion. *Mach. Learn.* <https://doi.org/10.1007/s10994-021-06008-4> (2021).
- Castillo-Navarro, J., Le Saux, B., Boulch, A., Audebert, N. & Lefèvre, S. Semi-supervised semantic segmentation in Earth Observation: the MiniFrance suite, dataset analysis and multi-task network study. *Mach. Learn.* <https://doi.org/10.1007/s10994-020-05943-y> (2021).
- Sumbul, G., Ravanbakhsh, M. & Demir, B. BigEarthNet-MM: a large-scale, multi-modal, multilabel benchmark archive for remote sensing image classification and retrieval [software and data sets]. *IEEE Geosci. Remote Sens. Mag.* **9**, 174–180 (2021).
- Roscher, R., Bohn, B., Duarte, M. F. & Garcke, J. Explainable machine learning for scientific insights and discoveries. *IEEE Access* **8**, 42200–42216 (2020).
- Kang, J. et al. Learning convolutional sparse coding on complex domain for interferometric phase restoration. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 826–840 (2021).
- Arcucci, R., Zhu, J., Hu, S. & Guo, Y. K. Deep data assimilation: integrating deep learning with data assimilation. *Appl. Sci.* **11**, 1114 (2021).
- Buizza, C. et al. Data learning: integrating data assimilation and machine learning. *J. Comput. Sci.* **58**, 101525 (2022).
- Mack, J., Arcucci, R., Molina-Solana, M. & Guo, Y. K. Attention-based convolutional autoencoders for 3d-variational data assimilation. *Computer Methods Appl. Mech. Eng.* **372**, 113291 (2020).
- Brajard, J., Carrassi, A., Bocquet, M. & Bertino, L. Combining data assimilation and machine learning to infer unresolved scale parametrization. *Phil. Trans. R. Soc. A.* <https://doi.org/10.1098/rsta.2020.0086> (2021).
- Geer, A. J. Learning earth system models from observations: machine learning or data assimilation?. *Phil. Trans. R. Soc. A.* <https://doi.org/10.1098/rsta.2020.0089> (2021).
- Farchi, A., Laloyaux, P., Bonavita, M. & Bocquet, M. Using machine learning to correct model error in data assimilation and forecast applications. *Quart. J. R. Meteorol. Soc.* <https://doi.org/10.1002/qj.4116> (2021).
- Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
- Beucler, T. et al. Enforcing analytic constraints in neural networks emulating physical systems. *Phys. Rev. Lett.* **126**, 098302 (2021).
- Nowack, P., Runge, J., Eyring, V. & Haigh, J. D. Causal networks for climate model evaluation and constrained projections. *Nat Commun.* <https://doi.org/10.1038/s41467-020-15195-y> (2020).
- Watt-Meyer, O. et al. Correcting weather and climate models by machine learning nudged historical simulations. *Geophys. Res. Lett.* **48**, e2021GL092555 (2021).
- Keisler, R. Forecasting Global Weather with Graph Neural Networks. Preprint at <https://doi.org/10.48550/arXiv.2202.07575> (2022).
- McGovern, A., Ebert-Uphoff, I., Gagne II, D. J. & Bostrom, A. The Need for Ethical, Responsible, and Trustworthy Artificial Intelligence for Environmental Sciences. Preprint at <https://arxiv.org/abs/2112.08453> (2021).
- Paul, S. & Ganju, S. Flood Segmentation on Sentinel-1 SAR Imagery with Semi-Supervised Learning. Preprint at <https://doi.org/10.48550/arXiv.2107.08369> (2021).
- Schneider, R. et al. A satellite-based spatio-temporal machine learning model to reconstruct daily PM2.5 concentrations across Great Britain. *Remote Sens.* **12**, 3803 (2020).
- Stafoggia, M. et al. Estimation of daily PM10 and PM2.5 concentrations in Italy, 2013–2015, using a spatiotemporal land-use random-forest model. *Environ. Int.* **124**, 170–179 (2019).
- Schneider, R. D. S. Estimating spatio-temporal air temperature in London (UK) using machine learning and earth observation satellite data. *Int. J. Appl. Earth Obs. Geoinf.* **88**, 1–10 (2020).
- Kloog, I., Nordio, F., Coull, B. A. & Schwartz, J. Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. *Remote Sens. Environ.* **150**, 132–139 (2014).

ACKNOWLEDGEMENTS

The organising committee wishes to thank all the speakers and participants at the workshop. The session chairs would also like to thank the participants who attended the working groups. We also appreciated the contribution from the scientific committee and the hard work from the communication teams from both organisations. A final special thanks to ESA and ECMWF for providing the technical and logistical resources for the workshop. The organising and scientific committee are delighted that the workshop brought speakers from different research backgrounds, countries, institutions, and professional levels. However, they acknowledged that the gender balance must be improved in the next edition.

AUTHOR CONTRIBUTIONS

R.S. and P.P.M. defined the questions that structured the basis of the manuscript. R.S. took the lead in writing the manuscript. R.S., B.D., and B.L.S. wrote TA1 section; R.A. and A.G. wrote TA2 section; C.V. and P.D. wrote TA3 section; and R.S. and M.B. wrote TA4 section. All authors contributed writing and providing critical feedback to the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Rochelle Schneider.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022