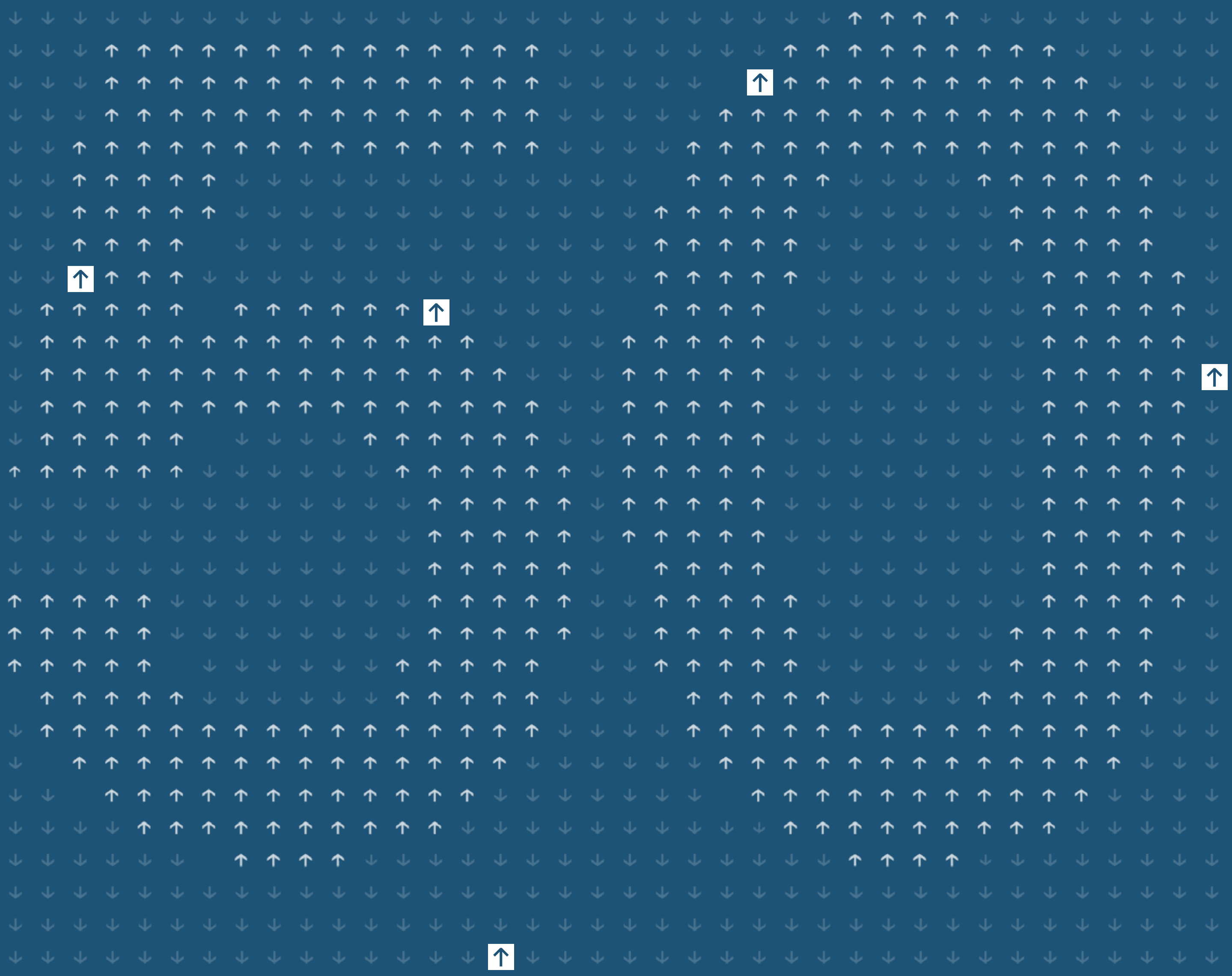
Machine learning for numerical weather prediction



ADVANCING WEATHER SCIENCE
THROUGH COLLABORATION

ECMWF | 50

Machine learning for numerical weather prediction

ABSTRACT

↘

Artificial intelligence (AI), and particularly machine learning (ML), is a mainstay of newspaper headlines, coffee conversations and everyday life across the world. Many fields and disciplines find themselves in the middle of a revolution, where the combination of data, algorithms and compute can provide low-cost solutions for a wide range of tasks. Weather forecasting is no exception to this, with this 50th ECMWF anniversary year seeing the operationalisation of the Artificial Intelligence Forecasting System (AIFS) at ECMWF. We will take a stroll through the history of machine learning at ECMWF, which starts earlier than one may expect, talk about the current state of play and gaze into a crystal ball in discussing the role of machine learning at ECMWF in the years to come.

AWAKENING: 2018–2022 →

Machine learning (ML) is the use of statistical algorithms that can learn from data and generalise to unseen data without explicit instructions. It is a sub-class of artificial intelligence (AI), which is the capability of computer systems to perform tasks typically associated with human intelligence. Many of the statistical methods that have been used for decades in Earth sciences can be counted into the wider class of ML. Examples are multi-dimensional linear regression, or dimensionality reduction via principal components. Even deep learning – the use of neural networks to perform ML – was already applied at ECMWF for the emulation of the radiation scheme more than two decades ago (Chevallier et al., 1998). It may therefore surprise that many claim that we have seen an ML "revolution" during the last couple of years. What happened?

Beyond the domain of Earth system science, AI and ML have seen an enormous rise that was mainly fuelled by:

- A massive increase in computational power with computer hardware customised towards the needs of deep learning, and deep learning being the ideal application for state-of-the-art supercomputers that excel for simple arithmetic and linear algebra.
- The exponential increase of data in many domains – including weather and climate – and the ability of deep learning to learn systems of arbitrary complexity if enough data and compute capacity are available.
- The availability of software libraries such as TensorFlow and PyTorch that allow a user to create complex deep learning architectures with very minimal Python code.
- The massive amount of experience that was collected on how to design efficient deep learning methods with new neural network architectures and training procedures being invented, including convolutional neural networks, recurrent neural networks, generative adversarial networks, attention and transformers, and diffusion networks.

It became more obvious around 2018 that the developments in general machine learning would also impact data assimilation and Earth system modelling. Early success stories across ECMWF's workflow included the use of neural networks for SMOS soil moisture data assimilation for the land surface (Rodríguez-Fernández et al., 2019) and the use of neural networks for bias correction learned within the 4D-Var data assimilation framework (Bonavita and Laloyaux, 2020). Deep learning has been used successfully for the emulation of the gravity wave drag parametrization schemes (Chantry et al., 2021a), and the deep learning emulators could be used to generate tangent linear and adjoint model code for 4D-Var data assimilation (Hatfield et al., 2021). Furthermore, decision trees have been used for the post-processing of ensemble predictions for precipitation (Hewson and Pillosu, 2020), and there have been plenty of links and similarities between data assimilation and deep learning (Geer, 2021). As ECMWF hosts more than one exabyte of weather and climate data, there were plenty of possible application areas for versatile, scalable tools that allow the extraction of complex information from data – such as deep learning. The potential applications were distributed across all parts of the numerical weather prediction (NWP) pipeline, from observation processing (Dahoui, 2023) to data assimilation, to the forecast model and the post-processing and dissemination of the forecasts. These included *methods to improve our understanding of the Earth system* such as unsupervised learning and causal discovery, uncertainty quantification, and AI powered visualisation; *methods to speed-up conventional models* such as emulators for parametrization schemes including with low numerical precision, the optimisation of the high-performance computing (HPC) and data workflow, and data compression; *methods to improve the models* such as bias correction tools, tools for quality control of observations, feature detection algorithms, and the learning of model components from observations; and *methods that linked different datasets to weather and climate datasets* that have interesting applications for health, energy, transport and pollution applications, as well as for extremes such as wildfires or flooding.

To bring a bit more structure into the multitude of applications that could be explored and to quickly develop the infrastructure and know-how that was needed to move quickly in the developments, ECMWF published a Machine Learning Roadmap for the next ten years in 2021 (Dueben et al., 2021). This roadmap outlined five objectives for the developments in machine learning at ECMWF (see Figure 1).

Figure 1: Objectives for the ML activities at ECMWF as defined in the ML Roadmap in 2021. As described below, ECMWF has been very successful in following the objectives.

Objective 1
Explore machine learning applications across the weather and climate prediction workflow and apply them to improve model efficiency and prediction quality.

Objective 2
Expand software and hardware infrastructure for machine learning.

Objective 3
Foster collaborations between domain and machine learning experts with the vision of merging the two communities.

Objective 4
Develop customised machine learning solutions for Earth system sciences that can be applied to various applications and at scale on current and future supercomputing infrastructure.

Objective 5
Train staff and Member and Co-operating State users and organise scientific meetings and workshops.

For the community to fully embrace machine learning required the opportunity for existing domain scientists to learn about the methodologies. Since its inception ECMWF has had a rich history of providing training. In 2022 a first course on machine learning for weather forecasting was run at ECMWF's Reading site. This has since been repeated most years, with each course being significantly oversubscribed. The topics covered in this course have evolved each year, to cover fresh communities in the field. In 2023 ECMWF introduced a Massive Open Online Course (MOOC) for machine learning in weather and climate, designed to help en-masse the weather forecasting community engage with the topics. The MOOC had 40 hours of content and was a great success attracting more than 9,000 registered participants from 159 countries. It featured contributions from 60 experts around the globe, covering three tiers.

In the first years after the awakening of deep learning for weather and climate, the community mostly focused on "Soft AI" to allow for improvements in computing efficiency via emulation, or "Medium AI" that incorporated machine learning into physics-based models e.g. via learning within data assimilation (Chantry et al., 2021b). However, ECMWF has also written the first-ever paper for "Hard AI" for medium-range NWP. This work aimed to replace the entire forecast model based on a pure deep learning tool trained from reanalysis data (Dueben & Bauer, 2018). ECMWF has also contributed significantly to WeatherBench, the first benchmark dataset to train global weather prediction models from reanalysis data (Rasp et al., 2020). However, at ECMWF and other meteorological centres, "Hard AI" approaches were treated rather as a testbed and scientific playground for new neutral network architectures and training mechanisms than as a serious alternative to physical models for operational predictions. This changed in 2022.

DISRUPTION: 2022 AND 2023 →

In 2022 came new players and accelerated progress. The arXiv, a home for preprints before peer review, saw a succession of papers broadly following the problem as described in Dueben & Bauer (2018), but introducing new methodologies. First came work in February 2022 by Ryan Keisler, an individual without affiliation intrigued by the topic. He built on a literature of graph neural networks for science and trained a message-passing graph neural network (see Figure 2). This approach significantly increased the skill relative to previous efforts and outperformed the GFS model (Global Forecast System of the US National Centers for Environmental Prediction, NCEP). Just one week later came work from a team at NVIDIA, who were the first to train at ERA5's full resolution, represented via 0.25 degree latitude-longitude grid (approximately 28 km), using a spectral approach which enabled the model to learn dynamics through a mix of neural networks operating in spectral and grid-point spaces. NVIDIA's model marked the first of a series of large technology companies entering the domain of ML weather forecasting. November 2022 saw Pangu-Weather, a preprint by Huawei (Bi et al., 2023). Pangu introduced a novel timestepping approach, creating models optimised to make timesteps between 1 and 24 hours, to be used in combination when delivering medium-range forecasts. Pangu was the first model to make claims of outperforming ECMWF's Integrated Forecasting System (IFS) across the majority of variables. Perhaps even more eye-catching were results evaluating the skill of tropical cyclone tracks, where the authors claimed a significant skill gain. A month later, gifted to the community on Christmas Eve, came GraphCast, a preprint submission by Google Deepmind (Lam et al., 2023). It adopted a similar graph-based approach to Keisler but introduced multi-scale connections in its graph-based approach and worked on the 0.25 degree grid. GraphCast provided more in-depth evaluation, inspired by the ECMWF scorecard, and argued for supremacy over the IFS across over 99% of variables and timescales. 2023 saw more papers, each claiming further increases in skill or utility, for example running at 9 km as seen in Aurora. A new sub-domain had emerged.

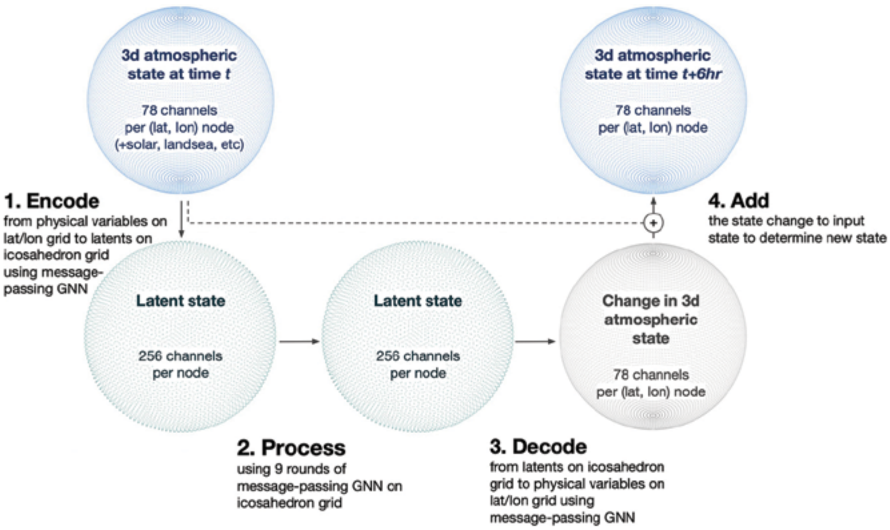


Figure 2: Reproduced from Keisler (2022). The figure shows a schematic for building a weather forecasting neural network. **1**. The 3D state is encoded into a latent state on a coarser grid. **2**. Information is then passed between nodes along edges in successive layers to calculate a latent estimate of the change in the atmospheric data. **3**. This information is then decoded onto the state of atmospheric variables. **4**. The information is added to the starting time state to give a prediction 6 hours into the future.

Perhaps even more arresting than the forecast skill were the energy costs to make a forecast. Once trained, the above systems could finish a ten-day forecast in a couple of minutes, on a single commercial-grade graphics processing unit (GPU) – to be contrasted with approximately 30 minutes on around 50 nodes for a forecast using the IFS. Roughly, these systems could reduce the energy costs of the forecasting piece of the chain by a factor of 1,000.

Common to all of these works were two aspects which draw sharp contrast with physics-based models and which are part of the explanation for the computational efficiency. One was the relatively minimal prognostic state of the atmosphere required for accurate forecasting. This featured approximately 13 pressure levels up to 50 hPa and only the basic atmospheric variables of humidity, winds, temperature and geopotential. This meant accurate forecasting without an explicit representation of clouds, and with a far coarser representation in the vertical dimension than the 137 levels used by the IFS. The second was the huge timestep, typically 6 hours, made by the models, in contrast to 9 minutes for the IFS. A timestep this large would not be numerically stable for 30 km resolution physics-based models using conventional timestepping schemes.

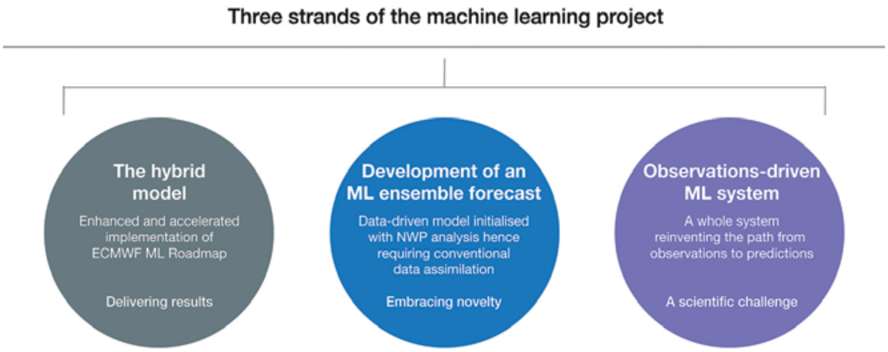
Due to the preprint nature of contributions, and the work stemming from outside of the meteorological community, there was natural uncertainty to these results. Several features threatened to undercut the validity of the results and warranted closer investigation. Two examples were the spatial smoothing of forecasts induced by optimising root-mean-square errors, and the use of ERA5 as initial conditions, meaning incorporation of fresher observations. Both features are known to artificially inflate skill. To explore this, in early 2023, ECMWF became the first centre to start running these models in real time from the operational ECMWF analysis and it also showcased plots of live forecasts to users on ECMWF's open charts. The goal was to help the whole community explore and understand these systems. Through in-house scoring, and verification of case studies, it became quickly clear that the results broadly held up to this further scrutiny (Ben Bouallègue et al., 2024). ERA5 initial conditions and forecast smoothing play a part in the skill gains but did not explain away significant improvements in forecasting skill. The open charts were popular on social media, with experienced meteorologists exploring live case studies and generally finding favourable results.

Data-driven models were not a panacea. At that time, current challenges were estimation of small-scall extreme values, e.g. wind speeds in tropical cyclones, or intense small-scale precipitation. These challenges all had roots in the training approach. By minimising the mean-squared-error, models were not rewarded for making bold predictions for harder to predict events. The tool developed at ECMWF to enable this easy running of Pangu-Weather, FourCastNet, GraphCast and more from a single interface, named ai-models, was created as an open-source repository enabling the wider scientific community to run these models more easily and better understand their dynamics.

ADOPTION: 2023–2025 →

In the summer of 2023, with strong support from the ECMWF Council, ECMWF started the Machine Learning Project, a four-year project to embrace the disruption of machine learning and develop operational systems. This project has three strands (see Figure 3). The first focuses on hybrid combinations of machine learning and physics, as championed in the machine learning roadmap. The hybrid strand, already mature with elements in the operational pipeline, promises value to further improve the IFS. The second focuses on data-driven weather forecasting, trained on reanalysis and analysis data. Here the goal was to build on the scientific publications by first creating a system that matched the skill of these systems, before then aiming to be world-leading in this domain. One of the major targets was an operational ensemble system two years after the start of the project. The third strand is the use of observations to build forecasting systems, including forecasting directly from observations, i.e. building a system inclusive of data-assimilation and forecasting.

Figure 3: The three strands of the machine learning project, which started at ECMWF in the summer of 2023.



The first strand, hybridising the IFS with machine learning, explores many approaches to improve the skill and efficiency of the IFS. This includes learning properties of sea-ice state using deliberately minimal learning systems (Geer, 2024); bias correction of 4D-Var data assimilation (Bonavita and Laloyaux, 2020), which will feature in ERA6; and nudging the IFS towards a data-driven model, following Husain et al. (2025), aiming to combine the large-scale improvements of data-driven modelling with the enhanced small-scale details and expanded product set of the IFS.

Prototype work for the Artificial Intelligence Forecasting System (AIFS), began immediately. The design choice of graph neural networks was made for the first implementation. This was chosen due the flexibility of data grid choices and natural encoding of the spherical geometry of the Earth. Building on the previous works by Keiser and GraphCast, rapid progress was made, with a first real-time running system in place by October 2023. After several further experimental model cycles (Lang et al., 2024a), the AIFS Single – named to capture a system designed to produce a single trajectory – was implemented as an operational system in February 2025. AIFS Single 1 outperforms the IFS across the vast majority of scores, with tropical cyclone track accuracy a notable place of significant improvements in forecasting ability (see Figures 4 and 5 for examples). Tropical cyclone intensity estimates by contrast are a current weak point, with AIFS Single significantly underestimating intensity, comparable with other data-driven models. AIFS Single 1 was made fully open source, including both data and model, enabling anyone to easily run the forecasting system themselves. The ability of users to easily run models themselves without HPC systems or extensive HPC knowledge is another advantage of these data-driven systems.

Figure 4: Skill comparison of AIFS Single 1 and the IFS for temperature at 850 hPa in the northern hemisphere extratropics for the spring period MAM 2025.

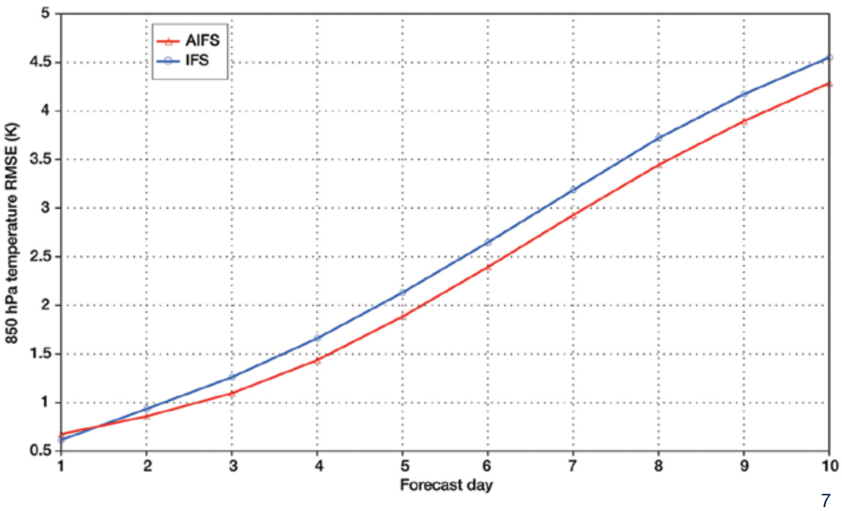
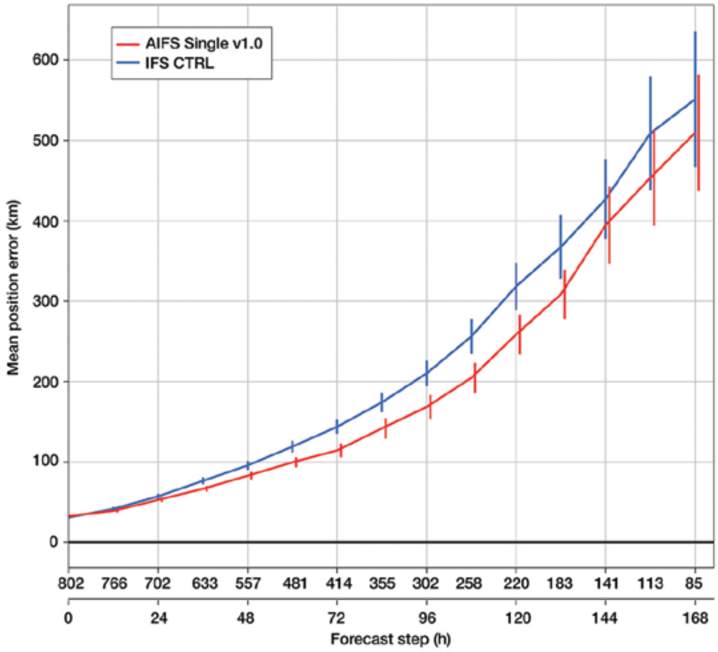


Figure 5: Skill comparison of AIFS Single 1 and the IFS for tropical cyclone track accuracy from July 2024 to June 2025.



Next came the ensemble. Two methodological strands were explored in development towards the first AIFS ensemble system. One used the diffusion training methodology, state-of-the-art in image and video generation ML systems. The second directly targeted optimising the Continuous Ranked Probability Score (CRPS, Lang et al., 2024b). This latter approach proved the more accurate and computationally cheaper. A first ensemble system, AIFS ENS 1, was implemented in July 2025, two years after the start of the ML project.

The CRPS-optimisation versions of the AIFS demonstrate skill not only on the medium-range timescale, but also at sub-seasonal timescales, outperforming the IFS sub-seasonal system across a number of key metrics. To engage the community in this developing field of sub-seasonal forecasting, ECMWF is organising the AI Weather Quest, a competition for sub-seasonal forecasting.[1] At the beginning of the event, 55 models had been entered across 33 teams and 14 countries. Three AIFS variants are being submitted to this competition.

The third strand investigates whether it is possible to encompass the full forecasting system, from observations to predictions, with machine learning. Following the success of machine learning for forecasting from analysis, this is a natural question to ask, one with wide reaching impact if true. A number of groups have engaged with research in this fascinating topic, with different problem framings being tested. Some, like the work of Allen et al. (2025), seek to utilise ERA5 in training, but still produce an end-to-end system without real-time dependencies on ERA5 or similar products. At ECMWF, a novel approach – AI-DOP (Direct Observation Prediction) – was proposed by McNally et al. (2024), seeking to only use observations in building an end-to-end system capable of forecasting future observations from current ones. GraphDOP (Alexe et al., 2024), a prototype of this approach, was created (see Figure 6), building on some of the work for the AIFS. This model showed that accurate forecasting was possible. However, currently the forecast skill of this work lags behind that of the IFS but is continually improving. Whether these works are the equivalent to the work of Keisler, showing promise without yet being state of the art, or whether machine learning fails to surpass physics-based data assimilation, we will learn in the coming years.

1 https://aiweatherquest.ecmwf.int/

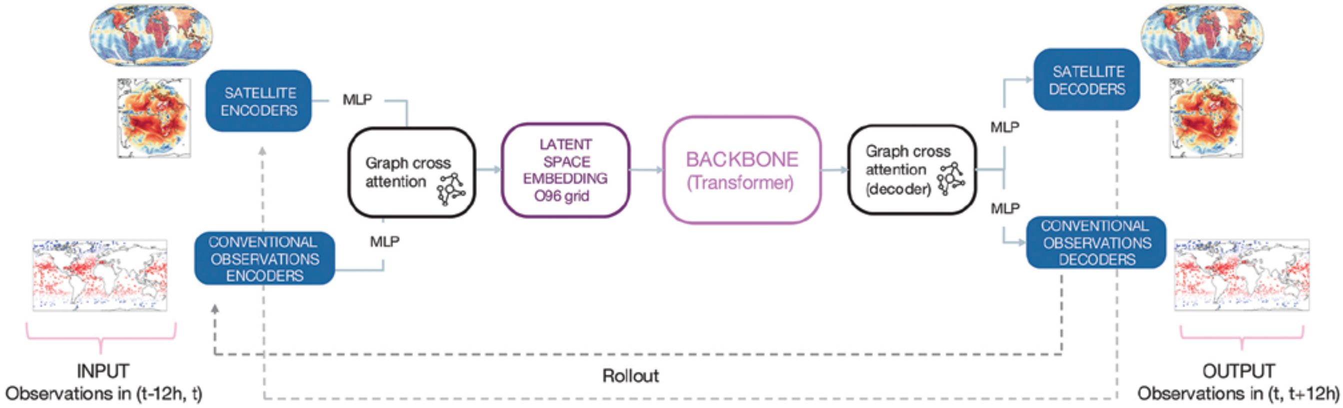


Figure 6: Schematic of the Graph-DOP approach developed in Alexe et al. (2024) which learns to predict future observations from current observations.

At the founding of the ECMWF ML project, opportunity was seen for ECMWF and its Member States to collaborate closely on the topic of data-driven weather forecasting. The ECMWF Member State ML Pilot Project was set up as a vehicle for organising this collaboration, featuring 14 partners across Europe at the time of writing, engaging in five work packages of data-driven modelling. ECMWF refactored the code underlying the AIFS to create a new open-source framework for data-driven modelling, dubbed Anemoi. This identified that the same code underlying a specific data-driven forecast system could also be used to develop global and regional forecasting systems for organisations across Europe, and this code could be co-developed by the European meteorological community, who could make rapid progress together. The first demonstration of this was Nipen et al. (2024), which built the first stretched-grid models, a forecasting system that featured higher spatial resolution over the Nordics but learnt from data around the globe. Anemoi was introduced in 2024 and in 2025 it won the European Meteorological Society (EMS) technology achievement award. The Anemoi community has grown to more than 12 Member States, who use and contribute to Anemoi.

OUTLOOK →

Due to their superiority in deterministic and ensemble forecast scores, the ML models will become the default tool for most applications in NWP. However, it is also unlikely that physical models will disappear from the operational portfolio in the foreseeable future. Physics-based systems currently provide a much wider range of products for users. They can also serve as backup model configurations if unprecedented events are happening (how would an ML forecast model represent the impact of a volcanic eruption on NWP?). Physical models are the prime tool to generate training data when observational datasets are sparse or inconsistent.

While ML currently lags behind physics-based systems for data assimilation, we view that it is likely that there will be an end-to-end ML forecast suite in the future, covering observation ingestion, data assimilation, the forecast model, post-processing and product generation.

One open question is how many different ML models will be trained and used in parallel for operational NWP – one seamless model for all global predictions, or many specialised models for specific predictions, for example for tropical cyclones?

As part of the EU Destination Earth initiative, ECMWF is already developing ML model components for the ocean, ocean waves, sea ice, land surface, and hydrology. ECMWF will, therefore, soon have a full machine-learned Earth system model. There are interesting questions about the coupling of model components currently being explored – shall we couple ML models for the various components that have each been trained individually, or shall we train all components as a single model? A full machine-learned Earth system model that covers the coupled system will be much more useable for long-term and potentially even climate simulations when compared to the current AIFS NWP model.

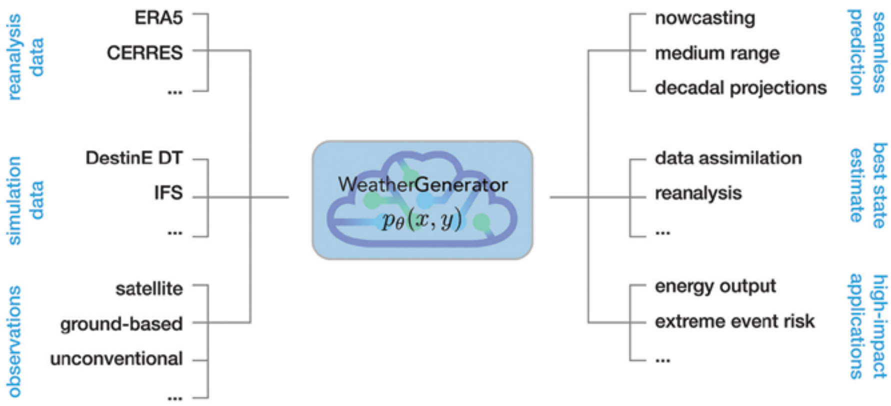
A topic of active discussion is the role for ML models on seasonal to climate timescales, a problem fundamentally about extrapolation into unseen climate regimes. Training on present day alone teaches a model to predict the current climatology, as shown in a recent study by ECMWF and Alfred Wegener Institute (AWI) scientists (Rackow et al., 2024). Work outside of ECMWF has shown that indefinitely stable and conservative data-driven models can be built (Watt-Meyer et al., 2024) and can produce skilful seasonal forecasts (Kent et al., 2025). The ability to create huge ensembles at low cost is an enticing one, particularly on the seasonal timescale. However, building ML models for long-term predictions is challenging due to the finite dataset length that needs to be split into a training and a sufficiently large validation dataset to construct statistically robust results. ECMWF is actively exploring this topic of data-driven seasonal forecasting.

Across the full spectrum of ECMWF activities, close thought is being given to consider the opportunities offered by machine learning. For atmospheric composition, interesting reanalysis datasets and preliminary results from Aurora (Bodnar et al., 2025) have prompted the exploratory development of an AIFS system for atmospheric composition. For reanalysis, the AI-DOP approach is being explored for the construction of reanalysis products.

The ability to make a forecast using a single GPU in just a couple of minutes opens new opportunities for the democratisation of weather forecasting. Within Destination Earth, ECMWF is developing a “forecast-in-a-box” prototype, which packages initial condition retrieval, forecasting, product generation and visualisation into a single portable unit, capable of running locally or on cloud facilities. Alongside MET Norway, ECMWF is working with national forecasters in Malawi to test and further develop this prototype.

Another question about the future of ML for Earth system modelling is whether the domain will follow the developments of large language models (LLMs) towards larger and more generic ML tools that can then be used for multiple application areas – so-called foundation models. LLMs are trained to fill in gaps in huge amounts of text, rather than for a specific task such as the translation from language A to language B. The resulting tools can be used for diverse tasks beyond their training objective, which include translations but also the almost instantaneous creation of a Shakespearean poem about TikTok cat videos. Along this line, it may be possible that a foundation model trained from various Earth system datasets and with a huge latent space with many billions of trainable parameters may perform better in certain tasks when compared with a task-specific model. To explore foundation models for weather and climate applications, ECMWF and a number of Member States have started the WeatherGenerator EU Horizon project that will build such a foundation model and serve as an additional digital twin for Destination Earth (see Figure 7).

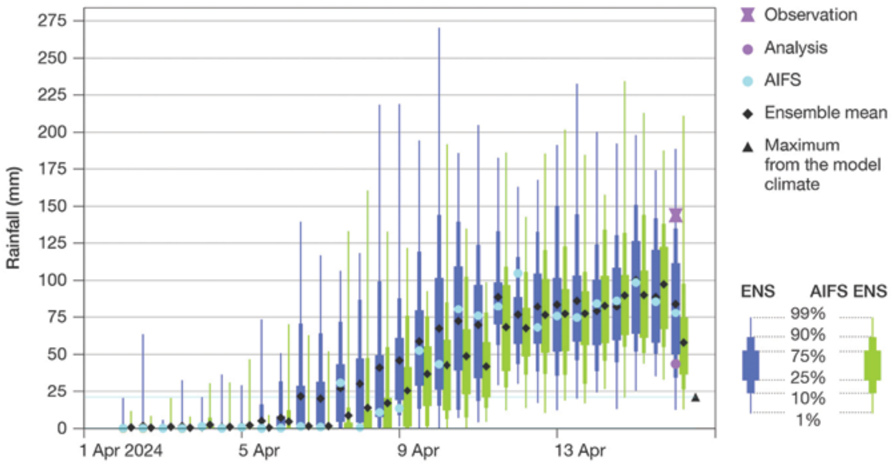
Figure 7: The WeatherGenerator will serve as foundation model for weather and climate applications. It will be able to digest datasets from various sources, including observations, local and global analysis products, and local and global models, and will be useful for many different application domains in weather and climate science. The input and output of the WeatherGenerator can be local or global, at a resolution between 100 m and 100 km, and can be in the past, present and future.



Data is obviously a key ingredient in accurate forecasting using physics-based approaches. We view that the role of data will grow even larger with the rise of data-driven forecasting. Curating large, calibrated datasets will be vital to feed these models. Direct incorporation of surface observations into model training and inference will increase the value of sharing these datasets. Novel data sources, e.g. cameras measuring visibility, could be directly included in model pipelines.

Representation of extreme, and particularly unprecedented, events remains a somewhat open question. Small-scale extreme events are typically not captured well in reanalysis and analysis datasets, and observation data sources for such events bring their own challenges. Unprecedented events can be categorised from a local or global perspective, i.e. events that have never been seen before in that region versus truly novel events that have never happened anywhere around the globe. Results so far suggest that for local extremes, machine learning models are able to surpass local climatologies significantly by learning to transfer lessons from other parts of the globe. Figure 8 shows an example of this for a case study in the UAE in April 2024. AIFS models confidently predicted record-breaking values well in advance of the event. For globally unprecedented events, researchers at the University of Chicago trained a data-driven forecasting system on ERA5 with all the strongest tropical cyclones removed from the training dataset (Sun et al., 2025). Without these events, the model was unable to produce these unseen strongest events. Further work is required to better understand the value for ML systems in extreme events. Live investigation through case studies will be vital to building trust.

Figure 8: Showing the evolution of 24-hour rainfall forecasts on 16 April 2024 over the grid box including Dubai. The model climate is about zero precipitation, with a maximum of less than 25 mm based on 1,800 forecasts (marked by the black triangle). Experimental versions of AIFS Single (blue dots) and AIFS Ensemble (green box and whisker) models both predicted precipitation values well-outside the model climatology and values in line with the IFS ensemble (blue box and whisker). All systems underestimated the observed value (purple hourglass).



Without the explicit underpinning of physics in data-driven systems, an increased emphasis is being placed on building trust. Across the wide user-base of ECMWF products, approaches for building trust will differ. For some, extensive verification will be the most important thing, for others case studies. Some users will prioritise physical consistency as an important facet. A holistic view across these dimensions will be important.

What is hopefully clear to all readers is that the world of numerical weather prediction, and ECMWF itself, is amid a revolution. The fundamentals of weather prediction are changing. Machine learning is bringing new opportunities and interesting scientific questions to be answered. Data-driven forecasting offers an opportunity for the meteorological community, particularly in Europe, to work even more closely and benefit from shared tooling without coalescing on a single model. The outside world, including large technology companies, will seize this opportunity, and if ECMWF wants to maintain its position in the community, continued agility and adaptability will be required. The next few years will be key for continued community building and answering the remaining fundamental questions captured above.

CONTRIBUTORS

ECMWF would like to thank the contributors to this paper:

Matthew Chantry and Peter Dueben.

Also the following for providing figures:

Ryan Keisler (Brightband), Michael Meyer Gerber, Zied Ben Bouallègue, Mihai Alexe, Eulalie Boucher, Linus Magnusson and Christian Lessig.

REFERENCES →

Alexe, M., E. Boucher, P. Lean, E. Pinnington, P. Laloyaux, A. McNally et al., 2024: GraphDOP: Towards skilful data-driven medium-range weather forecasts learnt and initialised directly from observations. https://arxiv.org/abs/2412.15687

Allen, A., S. Markou, W. Tebbutt, J. Requeima, W. P. Bruinsma, T. R. Andersson et al., 2025. End-to-end data-driven weather prediction. Nature, **641(8065)**, pp.1172–1179. https://doi.org/10.1038/s41586-025-08897-0

Ben Bouallègue, Z., M. C. Clare, L. Magnusson, E. Gascon, M. Maier-Gerber, M. Janoušek et al., 2024: The rise of data-driven weather forecasting: A first statistical assessment of machine learning–based weather forecasts in an operational-like context. *Bulletin of the American Meteorological Society*, **105(6)**, pp.E864-E883.

Bi, K., L. Xie, H. Zhang et al., 2023: Accurate medium-range global weather forecasting with 3D neural networks. *Nature* **619**, 533–538. https://doi.org/10.1038/s41586-023-06185-3

Bonavita, M. & **P. Laloyaux**, 2020: Machine learning for model error inference and correction. *Journal of Advances in Modeling Earth Systems*, **12**, e2020MS002232. https://doi.org/10.1029/2020MS002232

Bodnar, C., W. P. Bruinsma, A. Lucic et al., 2025: A foundation model for the Earth system. Nature, **641**, 1180–1187. https://doi.org/10.1038/s41586-025-09005-y

Chantry, M., S. Hatfield, P. Dueben, I. Polichtchouk & **T. Palmer**, 2021a: Machine learning emulation of gravity wave drag in numerical weather forecasting. *Journal of Advances in Modeling Earth Systems*, **13(7)**, p.e2021MS002477.

Chantry, M., H. Christensen, P. Dueben & **T. Palmer**, 2021b: Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI. *Philosophical Transactions of the Royal Society A379*, **no. 2194**: 20200083.

Chevallier, F., F. Chéruy, N. A. Scott & **A. Chédin**, 1998: A Neural Network Approach for a Fast and Accurate Computation of a Longwave Radiative Budget. *J. Appl. Meteorol.*, **37**, 1385–1397. https://doi.org/10.1175/1520-0450(1998)037<1385:ANNAFA>2.0.CO;2

Dahoui, M., 2023: Use of machine learning for the detection and classification of observation anomalies. *ECMWF Newsletter* **No. 174**, 23–27. https://doi.org/10.21957/n64md0xa5d

Dueben, P. D. & **P. Bauer**, 2018: Challenges and design choices for global weather and climate models based on machine learning. *Geosci*. Model Dev., **11**, 3999–4009. https://doi.org/10.5194/gmd-11-3999-2018

Dueben, P., U. Modigliani, A. Geer, S. Siemen, F. Pappenberger, P. Bauer et al., 2021: Machine learning at ECMWF: A roadmap for the next 10 years. *ECMWF Technical Memorandum* **No 878**. https://doi.org/10.21957/ge7ckgm

Geer, A. J., 2021: Learning earth system models from observations: machine learning or data assimilation? *Philosophical Transactions of the Royal Society* A 379, **no. 2194**: 20200089. https://doi.org/10.1098/rsta.2020.0089

Geer, A. J., 2024: Simultaneous inference of sea ice state and surface emissivity model using machine learning and data assimilation. *Journal of Advances in Modeling Earth Systems*, **16(7)**, e2023MS004080. https://doi.org/10.1029/2023MS004080

Hatfield, S., M. Chantry, P. Dueben, P. Lopez, A. Geer & **T. Palmer**, 2021: Building tangent-linear and adjoint models for data assimilation with neural networks. *Journal of Advances in Modeling Earth Systems*, **13**, e2021MS002521. https://doi.org/10.1029/2021MS002521

Hewson, T. & **F. Pillosu**, 2020: A new low-cost technique improves weather forecasts across the world. arXiv:2003.14397v1

Husain, S. Z., L. Separovic, J. F. Caron, R. Aider, M. Buehner, S. Chamberland et al., 2025: Leveraging data-driven weather models for improving numerical weather prediction skill through large-scale spectral nudging. *Weather and Forecasting*, **40(9)**, pp.1749–1771.

Keisler, R., 2022: Forecasting global weather with graph neural networks. *arXiv preprint arXiv:2202.07575*

Lam, R. et al., 2023: Learning skillful medium-range global weather forecasting. *Science* **382**, 1416–1421. https://doi.org/10.1126/science.adi2336

Lang, S., M. Alexe, M. Chantry, J. Dramsch, F. Pinault, Raoult, B. et al., 2024a: AIFS-- ECMWF's data-driven forecasting system. *arXiv preprint arXiv:2406.01465*

Lang, S., M. Alexe, M. C. Clare, C. Roberts, R. Adewoyin, Z. B. Bouallègue et al., 2024b: AIFS-CRPS: ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score. *arXiv preprint arXiv:2412.15832*

McNally, A., C. Lessig, P. Lean, E. Boucher, M. Alexe, E. Pinnington et al., 2024: Data driven weather forecasts trained and initialised directly from observations. *arXiv preprint arXiv:2407.15586*

Nipen, T.N., H. H. Haugen, M. S. Ingstad, E. M. Nordhagen, A. F. S. Salihi, P. Tedescoet al., 2024: Regional data-driven weather modeling with a global stretched-grid. *arXiv preprint arXiv:2409.02891*

Kent, C., A. A. Scaife, N. J. Dunstone, D. Smith, S. C. Hardiman, T. Dunstan & **O. Watt-Meyer**, 2025: Skilful global seasonal predictions from a machine learning weather model trained on reanalysis data. *npj Climate and Atmospheric Science*, **8(1)**, p.314.

Rackow, T., N. Koldunov, C. Lessig, I. Sandu, M. Alexe, M. Chantry & **T. Jung**, 2024: Robustness of AI-based weather forecasts in a changing climate. *arXiv preprint arXiv:2409.18529*

Rasp, S., P. D. Dueben, S. Scher, J. A. Weyn, S. Mouatadid & **N. Thuerey**, 2020: WeatherBench: A benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, **12**, e2020MS002203. https://doi.org/10.1029/2020MS002203

Rodríguez-Fernández, N., P. de Rosnay, C. Albergel, P. Richaume, F. Aires, C. Prigent & **Y. Kerr**, 2019: SMOS Neural Network Soil Moisture Data Assimilation in a Land Surface Model and Atmospheric Impact. *Remote Sens*, **11**, 1334.

Sun, Y.Q., P. Hassanzadeh, M. Zand, A. Chattopadhyay, J. Weare & **D. S. Abbot**, 2025: Can AI weather models predict out-of-distribution gray swan tropical cyclones? *Proceedings of the National Academy of Sciences*, **122(21)**, e2420914122. https://doi.org/10.1073/pnas.2420914122

Watt-Meyer, O., B. Henn, J. McGibbon, S. K. Clark, A. Kwa, W. A. Perkins et al., 2025: ACE2: accurately learning subseasonal to decadal atmospheric variability and forced responses. *npj Climate and Atmospheric Science*, **8(1)**, 205.

Find out more about ECMWF at 50

ADVANCING WEATHER SCIENCE
THROUGH COLLABORATION

ECMWF | 50