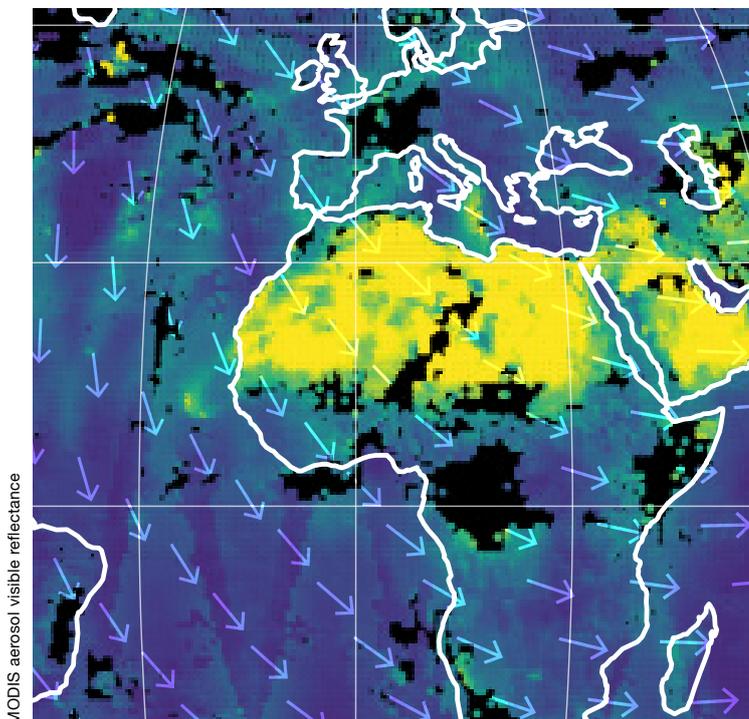




EARTH SYSTEM SCIENCE

AI-DOP: an update on medium-range forecast scores



This article appeared in the Earth system science section of ECMWF Newsletter No. 186 – Winter 2025/26, pp. 25–29

AI-DOP: an update on medium-range forecast scores

Eulalie Boucher, Mihai Alexe, Peter Lean, Ewan Pinnington, Patrick Laloyaux, Simon Lang, Tony McNally

Over the past year, ECMWF has continued its research on an end-to-end Artificial Intelligence (AI)-based Direct Observation Prediction (AI-DOP) model. In previous Newsletter articles (McNally et al., 2024 and McNally et al., 2025), we reported the first demonstration that medium-range forecasts using an AI-DOP concept could be produced. This update summarises the progress made during 2025 and outlines the next steps for this rapidly evolving research track.

Machine-learned (ML) weather prediction models typically derive their skill from reanalysis datasets such as ERA5. These datasets, however, rely on traditional approaches of incorporating observations into numerical weather prediction (NWP). Although modern data assimilation (DA) systems are remarkably sophisticated, they still struggle to extract the full information content from the global observing system. Simplified forward operators between measured radiances and geophysical variables, limitations in representing uncertainties, and the volume of satellite observations mean that some of the potential value in the observations remains under-utilised.

As a result, ML models trained solely on reanalysis inherit the constraints of the upstream DA system, rather than the richness of the observations themselves. AI-DOP was initiated to explore whether AI models could learn directly from observations, either alongside reanalysis, in place of parts of the DA pipeline, or in an end-to-end system and thereby unlocking information that current systems cannot fully exploit. Over the past year, it has become increasingly clear that integrating observations more directly into AI-based forecasting systems may be central to advancing forecast skill. This makes the AI-DOP research track an important step toward the next generation of observation-aware weather prediction models.

Current model architecture

GraphDOP (Alexe et al., 2024) is an AI-DOP model based on a graph neural network with an encoder-processor-decoder architecture (Figure 1). While its design shares many aspects with ECMWF’s Artificial Intelligence Forecasting System (AIFS) (Lang et al., 2024), GraphDOP’s architecture is adapted specifically

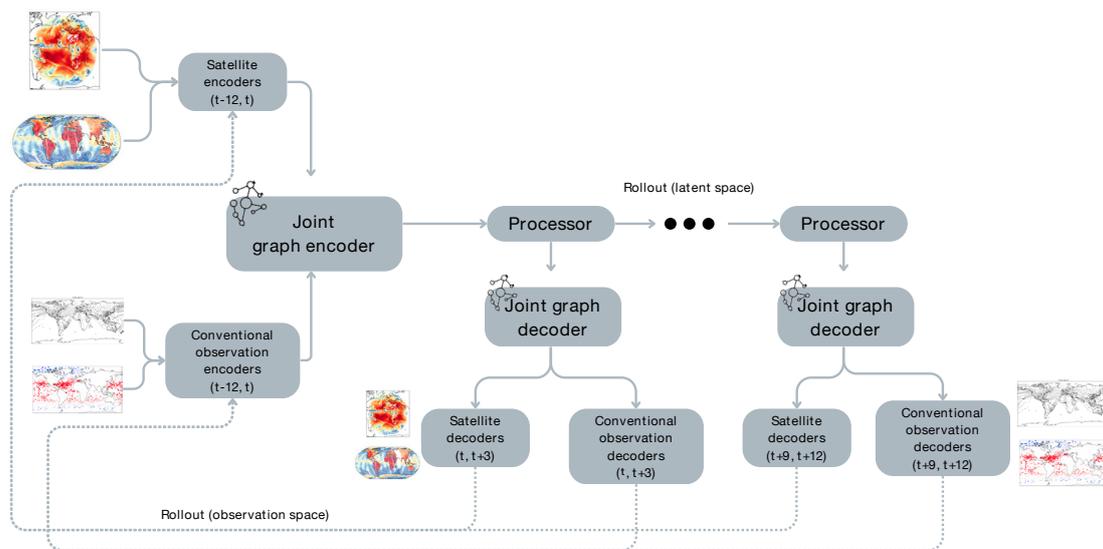


FIGURE 1 GraphDOP architecture with three-hour autoregressive time-stepping (rollout) in the latent space. During finetuning, the forecasted observations are fed back as inputs to produce the next forecast (rollout in observation space).

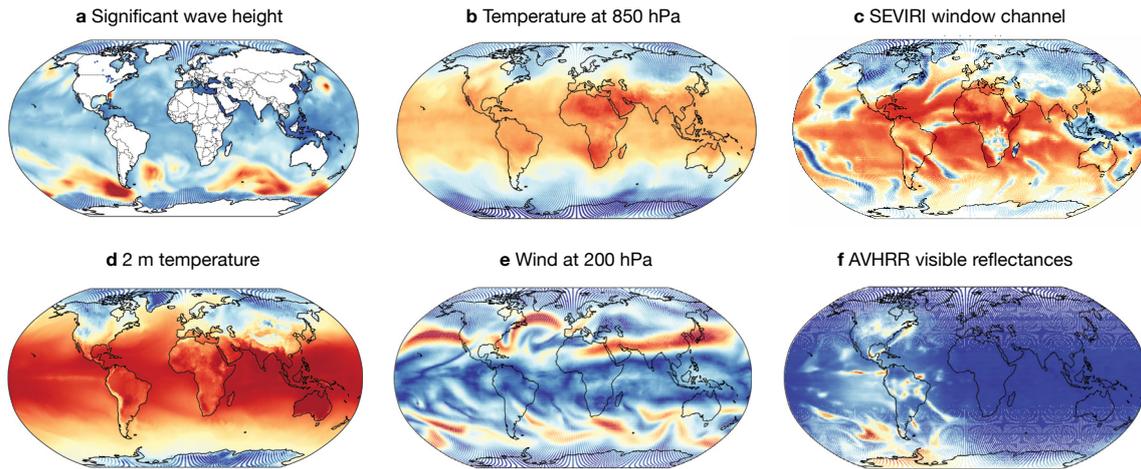


FIGURE 2 Example of gridded forecasts at 12-hour lead time of different parameters and satellite channels produced from the GraphDOP model.

for observation prediction, operating solely in observation space, without gridded climatology or NWP (re) analysis inputs or feedback.

A graph-based encoder projects the 12 hours of input observations onto a latent space. Subsequently, the processor (a sliding-window transformer) evolves the latent space representation throughout a 12-hour target observation window. Finally, a decoder projects the latent representation back into observation space to produce the model forecasts. The latent representation is defined on an O96 reduced Gaussian grid (a spatial resolution of ca. 1 degree), with 1,024 features. The encoder graph mapping is constructed on the fly from the input observations available in each training and validation batch to their nearest grid points in the latent space representation, a k -nearest neighbour graph, with $k=1$. Conversely, the decoder graphs connect each target observation to its three nearest neighbour nodes on the latent mesh.

The latest version of the model forecasts 12-hour observation windows by auto-regressively processing and decoding target observations across four consecutive three-hour chunks, a procedure referred to in Figure 1 as latent-space rollout. This approach has improved the sharpness and accuracy of the GraphDOP forecasts compared to the original model that decoded all the observations inside the 12-hour target window in a single step. Currently, the training objective is the weighted mean square error.

Once trained, GraphDOP is able to forecast any of the seen observation types, either at their true location or at any point in time and space, producing global forecasts of any observed parameter or satellite channel (Figure 2).

A series of experiments has been conducted to provide evidence that the GraphDOP model develops internal representations of the Earth system state, structure and dynamics, as well as the characteristics of different observing systems (Lean et al., 2025). GraphDOP simultaneously embeds information from diverse observation sources spanning the full Earth system into a shared latent space. This enables predictions that implicitly capture cross-domain interactions in a single model, without the need for any explicit coupling (Boucher et al., 2025).

Growing set of observations

A key advantage of a purely observation-driven system such as GraphDOP is its ability to ingest a wide range of observational data, including many that are not currently used in traditional DA systems. In parallel with model development, we have been steadily expanding our collection of ML-ready observational datasets for training, spanning both in-situ (conventional) measurements and satellite observations (Figure 3).

Not all curated datasets have been used thus far in training, and the performance results presented in the next section reflect only a subset of them. Nevertheless, we are gradually evaluating and incorporating additional datasets into our standard configuration as the system matures. Quality control tools initially

developed for 4D-Var have been implemented in GraphDOP to better quantify the importance of the different instruments (Laloyaux et al., 2025).

In addition, work has begun on developing a real-time workflow built on top of existing observation-receiving infrastructure, supporting a potential future operational end-to-end observation-to-forecast system.

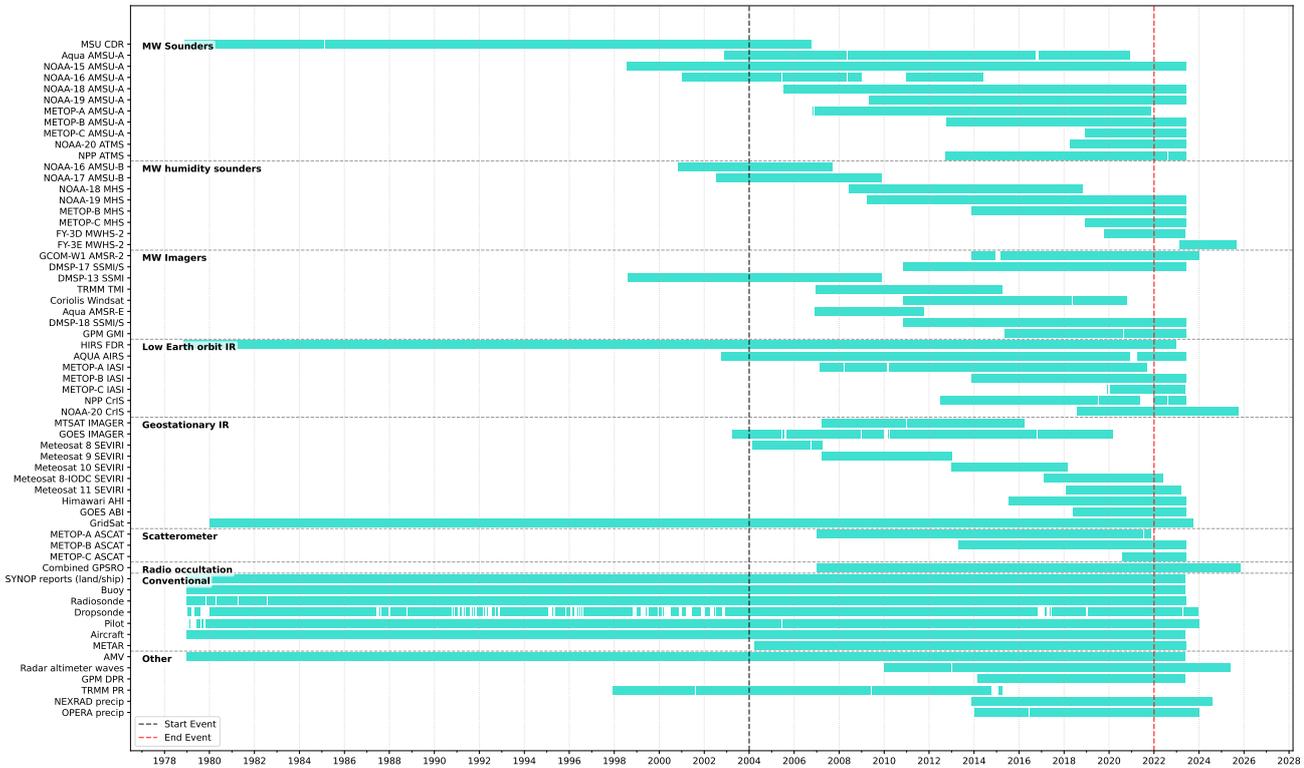


FIGURE 3 Observation types that have so far been converted into ML-friendly training datasets.

Medium-range forecast scores

The model updates combined with an enhanced set of observations have led to large improvements in the medium-range forecast scores. GraphDOP is especially skilful at forecasting surface parameters, and its performance is steadily improving for upper-air parameters, now matching the physics-based Integrated Forecasting System (IFS) at 12-hour lead time. In Figures 4 and 5, we show root-mean-square errors (RMSE) of gridded forecasts against in-situ observations for a summer month (June 2022) and compare those of GraphDOP to the IFS at the same resolution. Similar evaluations for winter months show higher errors, which we believe can be attributed to a harder predictability but also to a lack of observations representing winter conditions, such as snow observations.

Figure 4 illustrates these improvements for upper-air parameters at 850 hPa, 500 hPa, and 200 hPa (forecasted on an O96 grid) in the northern hemisphere extratropics (top row) and in the tropics (bottom row). In the extratropics and for all three levels, GraphDOP and the IFS exhibit very similar skill in the short range, with differences generally within the sampling uncertainty during the first 24–36 hours. Beyond this point, the IFS retains an advantage, particularly for the dynamically sensitive 500 hPa geopotential height and 200 hPa wind speed, where its errors grow more slowly through the medium range. Nevertheless, GraphDOP tracks the IFS closely throughout the ten-day period. To provide additional context, the figure also includes a dashed line indicating the performance of an older version of GraphDOP. Over the tropics, GraphDOP does especially well for 850 hPa temperature and 500 hPa geopotential in the first five days. This comparison highlights that while upper-air forecast skill continues to be led by the physical model, GraphDOP is making steady progress, supported, for example, by the improved training configuration introduced recently.

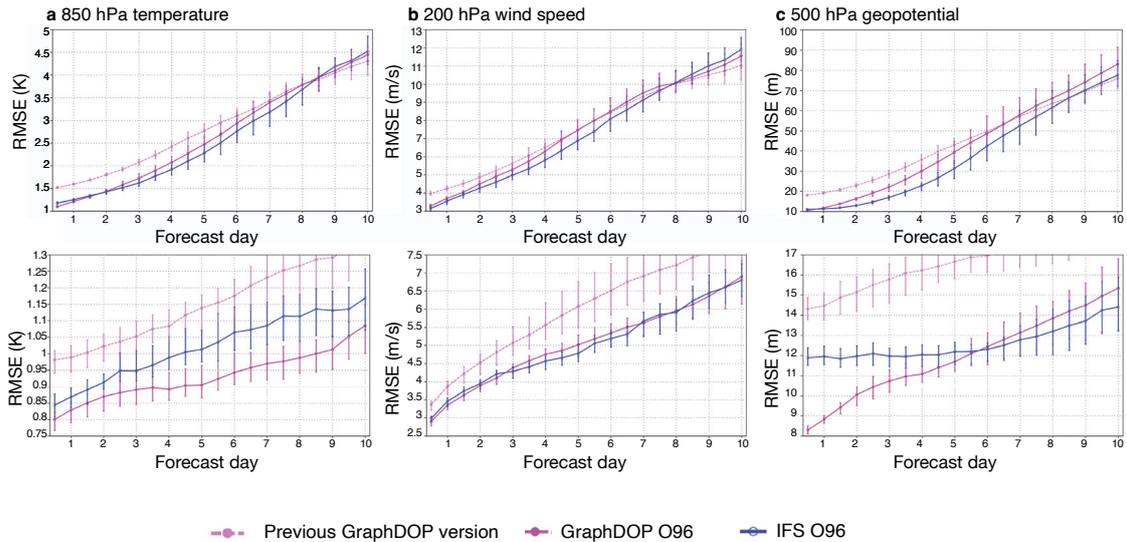


FIGURE 4 Root-mean-square error (RMSE) of (a) 850 hPa temperature, (b) 200 hPa wind speed and (c) 500 hPa geopotential forecasts by GraphDOP (magenta) and the IFS (blue) over the northern hemisphere extratropics (top row) and the tropics (bottom row) for June 2022. Scores are computed against observations.

Figure 5 presents corresponding results for key surface parameters, showing scores for 2 m temperature and 10 m winds forecasted on an N320 grid, over the northern hemisphere extratropics. For both parameters, GraphDOP delivers substantially improved short-range performance. In the case of 2 m temperature, GraphDOP begins with a markedly lower error (1.8 K compared with 2.4 K for the IFS at the same resolution), and preserves this advantage through to day 4, at which point the two systems

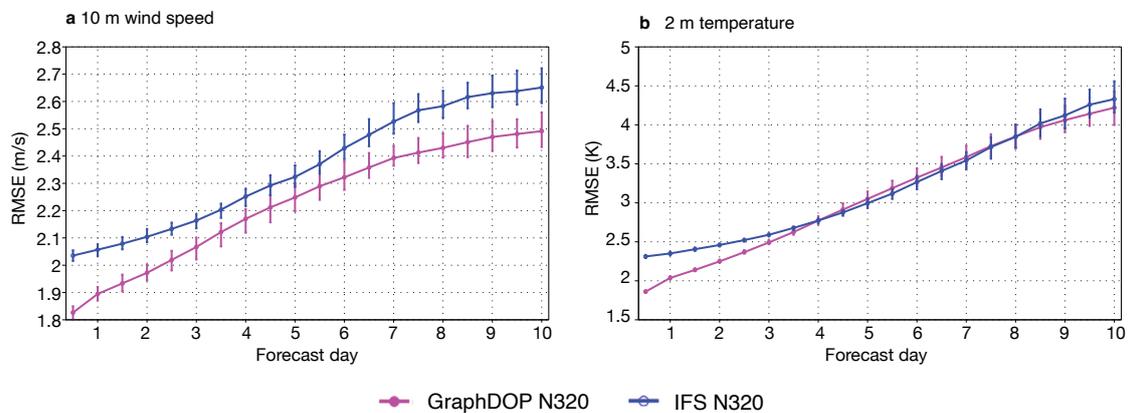


FIGURE 5 Root-mean-square error (RMSE) of (a) 10 m wind speed and (b) 2 m temperature forecasts by GraphDOP (magenta) and the IFS (blue) over the northern hemisphere extratropics for June 2022. Scores are computed against observations.

converge. Ongoing work focuses on improving the rollout strategy to curb error growth into the medium range, with the aim of extending GraphDOP’s short-range advantage. The largest gains appear for 10 m winds, where GraphDOP maintains significantly lower RMSE out to day 10. Together, these results suggest that the implicit, observation-driven, coupling embodied in GraphDOP – where near-surface conditions are constrained directly rather than through parametrized land–atmosphere exchange pathways – yields particularly strong benefits for surface variables.

Conclusions and outlook

The progress achieved over the past year demonstrates that the AI-DOP research track is maturing rapidly, both in terms of model capability and in the breadth of the observational information it can exploit. As we continue to expand the range of ML-ready observations, several key observing systems stand out as high-priority candidates.

These include GNSS radio-occultation (GNSS-RO) data, in which the bending angles have been reprocessed for ML use. This includes more sophisticated normalisation and quality control of the bending angles, allowing the model to pick up a useful signal. This is showing to be particularly valuable in current prototype experiments. At the modelling level, several promising directions of research are being explored to allow the network to benefit from longer sequences of observations and to reduce error growth in the rollout.

Overall, the results presented here highlight the growing potential of observation-driven AI models to complement the capabilities of traditional NWP.

Further reading

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. *arXiv* 2412.15687. <https://doi.org/10.48550/arXiv.2412.15687>

Boucher, E., M. Alexe, P. Lean, E. Pinnington, S. Lang, P. Laloyaux et al., 2025: Learning Coupled Earth System Dynamics with GraphDOP. *arXiv* 2510.20416. <https://doi.org/10.48550/arXiv.2510.20416>

Laloyaux, P., M. Alexe, E. Boucher, P. Lean, E. Pinnington, S. Lang et al., 2025: Using data assimilation tools to dissect GraphDOP. *arXiv* 2510.27388. <https://doi.org/10.48550/arXiv.2510.27388>

Lang, S., M. Alexe, M. Chantry, J. Drams, F. Pinault, B. Raoult et al., 2024: AIFS – ECMWF’s data-driven forecasting system. *arXiv* 2406.01465. <https://doi.org/10.48550/arXiv.2406.01465>

Lean, P., M. Alexe, E. Boucher, E. Pinnington, S. Lang, P. Laloyaux et al., 2025: Learning from nature: insights into GraphDOP’s representations of the Earth System. *arXiv* 2406.01465. <https://doi.org/10.48550/arXiv.2508.18018>

McNally, T., C. Lessig, P. Lean, M. Chantry, M. Alexe & S. Lang, 2024: Red Sky at night... producing weather forecasts directly from observations. *ECMWF Newsletter No. 178*, 30–34. <https://doi.org/10.21957/tmc81jo4c7>

McNally, A., C. Lessig, P. Lean, E. Boucher, M. Alexe, E. Pinnington et al., 2025: An update on AI-DOP: skilful weather forecasts produced directly from observations. *ECMWF Newsletter No. 182*, 15–19. <https://doi.org/10.21957/tmi6y913dc>

© Copyright 2026

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

The content of this document, excluding images representing individuals, is available for use under a Creative Commons Attribution 4.0 International Public License. See the terms at <https://creativecommons.org/licenses/by/4.0/>. To request permission to use images representing individuals, please contact pressoffice@ecmwf.int.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.