SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2024……………………………………………………
Project Title:	High-resolution data-driven weather forecasts for Western Europe
Computer Project Account:	……………………………………..
Principal Investigator(s):	Dr. M.J. Schmeits
Affiliation:	KNMI
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Dr. K. Whan (KNMI), Dr. Y. Shapovalova (deep-learning expert, Radboud University Nijmegen)
Start date of the project:	2024
Expected end date:	2026

Computer resources allocated/used for the current year and the previous one

(if applicable)

Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)				
Data storage capacity	(Gbytes)				

Summary of project objectives (10 lines max)

The main objective is to train a deep learning weather prediction (DLWP) model over Western Europe, with a focus on the Netherlands, using the 10-year DOWA reanalysis. We focus on predicting surface variables that are most important to national meteorological institutes for general forecasts and warnings, such as wind speed/gusts, precipitation, air pressure, and temperature. The project contains several steps. We plan to first develop data-driven (preferably ensemble/probabilistic) forecasts out to +48 hours using deep learning methods. Given the focus on warnings, we will extend previous work by exploring how the skill for extremes can improve with different loss functions (i.e., balanced or weighted loss). We further plan to compare these forecasts with the operational Harmonie-Arome (ensemble) forecasts that are used at KNMI. Finally, we plan to compare the data-driven (ensemble) forecasts with post-processed (probabilistic) forecasts from the operational Harmonie-Arome model.

Summary of problems encountered (10 lines max)

The retrieval of the DOWA dataset has been non-trivial. The data is saved on tapes of the European Centre File Store (ECFS), managed by the Data Handling System (DHS) of ECMWF in Bologna. As of this date, we have not been able to gain access to the full dataset due to the large size of the date combined with complications in the SSH transfer to the EWC. For the little data received, the conversion to the ECMWF Zarr format used by AIFS posed greater challenges than expected. Furthermore, the original AIFS model was not suited to run in the EWC framework and the adaptation to this framework in terms of data-loading and model code adaptation was a time-consuming process. Lastly, the combination of datasets caused the computation time of the model to increase significantly.

Summary of plans for the continuation of the project (10 lines max)

We are in process of transferring the DOWA reanalysis data to an s3 bucket on the EWC. In all other respects the data loading pipeline is prepared for training at a higher resolution. We plan to first train on a down-sampled version of the data, possibly using transfer learning, in collaboration with the Norwegian Meteorological Institute (MET Norway). Transfer learning could also allow us to train more efficiently by using a model that is pre-trained on ERA5 and finetuned on the combined DOWA/ERA5 data. The AIFS ensemble model is far in development, plans are to adapt this model to forecast at high-resolution over Western Europe. The performance of the model on extremes remains to be explored, as well as the comparison with the HARMONIE-AROME operational and post-processed forecasts.

List of publications/reports from the project with complete references

N.A.

Summary of results

During the first 6 months of this project, the Artificial Intelligence/Integrated Forecast System (AIFS), as developed by ECMWF (Lang et al., 2024), was adapted for a stretched-grid approach in collaboration with Met Norway and ECMWF.

1. Stretched-grid adaptation of AIFS

The initial AIFS model is based on GraphCast (Lam et al., 2023) and has an encoder-processor-decoder structure. To include higher-resolution data over a limited domain a stretched-grid approach was used in the processor, where the triangles overlapping with the area of interest are refined multiple times to keep the number of neighbours constant for the finest level of the multi-mesh (see Figure 1). The local data source was then integrated with the global data source and used as input for the encoder and output of the decoder. For testing purposes, the local and global data source were simulated by ERA5 reanalysis data, with resolution o96 on the global domain and resolution n320 on the local domain. The global data points are excluded from the limited area domain. The model was trained on 1 NVIDIA A100 GPU with training data ranging from 1979-2020. The model was validated on the year 2021 and tested on the year 2022.

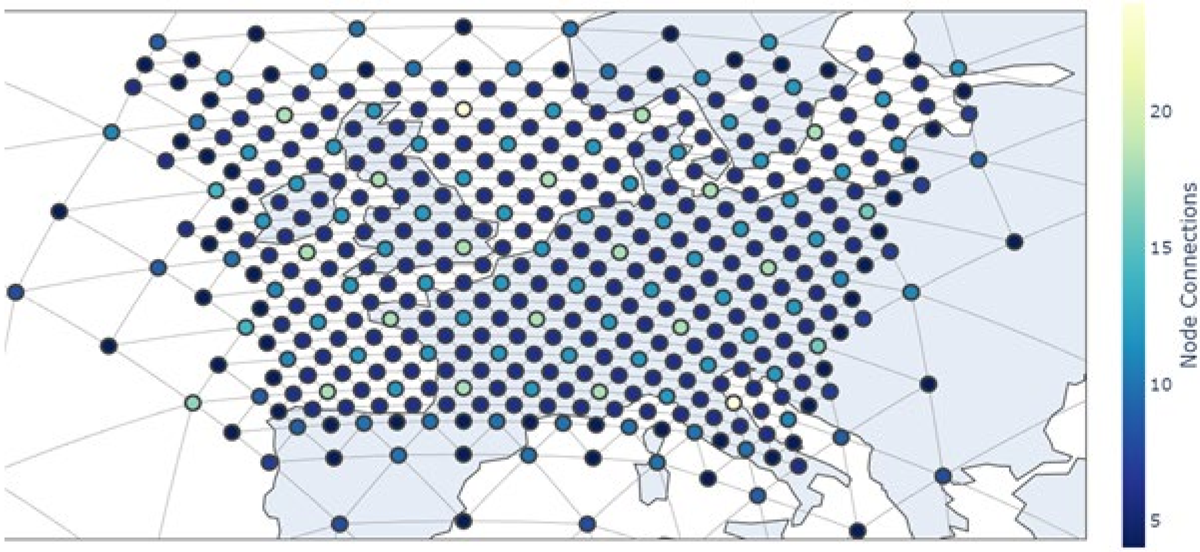

Figure 1: Visualization of the stretched hidden grid, with the refined area corresponding to the HARMONIE-AROME/DOWA full model domain. The triangles of the icosahedron are iteratively refined and projected to the unit sphere.

Qualitative analysis reveals reasonable predictions for short lead times. In general, we note that surface variables and variables at higher pressure levels show better performance than upper air variables. This is comparable to the GraphCast and AIFS model results (Lam et al., 2023, Lang et al., 2024). For longer lead times the model tends towards the mean, resulting in blurred forecasts. This issue is prevalent in most deterministic DLWP models, since the mean squared error (MSE) loss tends to favour predictions towards the mean of the distribution (Xu et al., 2024). On the other hand, probabilistic DLWP models (e.g. GenCast (Price et al., 2023)) have shown to produce spatially sharp ensemble forecasts.

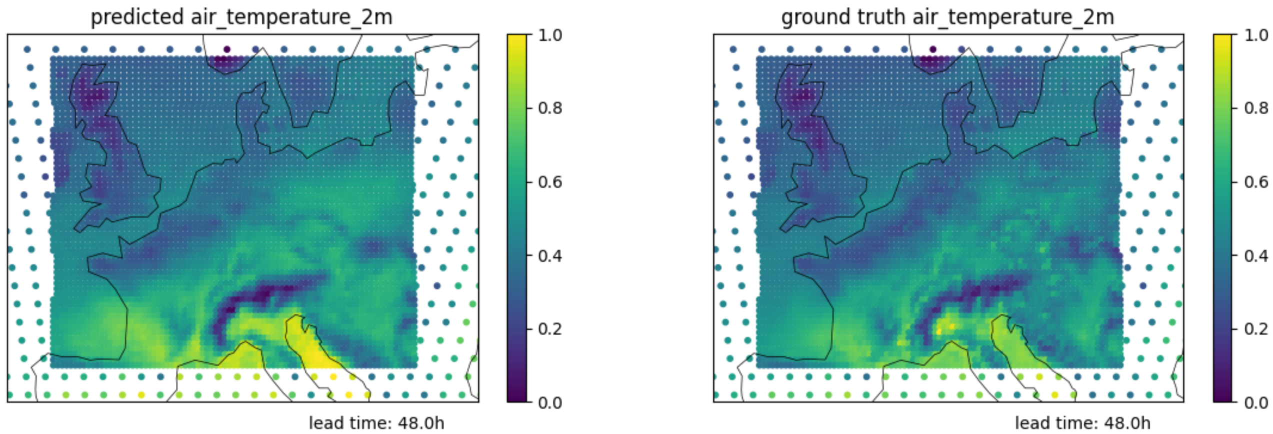

Figure 2: Stretched-grid AIFS +42h autoregressive prediction of the air temperature at 2 meter above sea level.. Note that this model has resolution 5 on the hidden grid with two refinements, and was finetuned using rollout training up to 3 days. The model predicts the future state the atmosphere adequately, although blurring effects occur at longer lead times.

Predictions at longer lead times can be improved by finetuning with a loss evaluated over longer lead times (rollout training). We tested the influence of rollout training on the performance of the model. In our experiments, the hidden grid icosahedron was refined four times or five times, with two extra refinements on the limited area domain. The predicted Root Mean Squared Error (RMSE) of the windspeed at 850 hPa as a function of lead time can be seen in Figure 3.

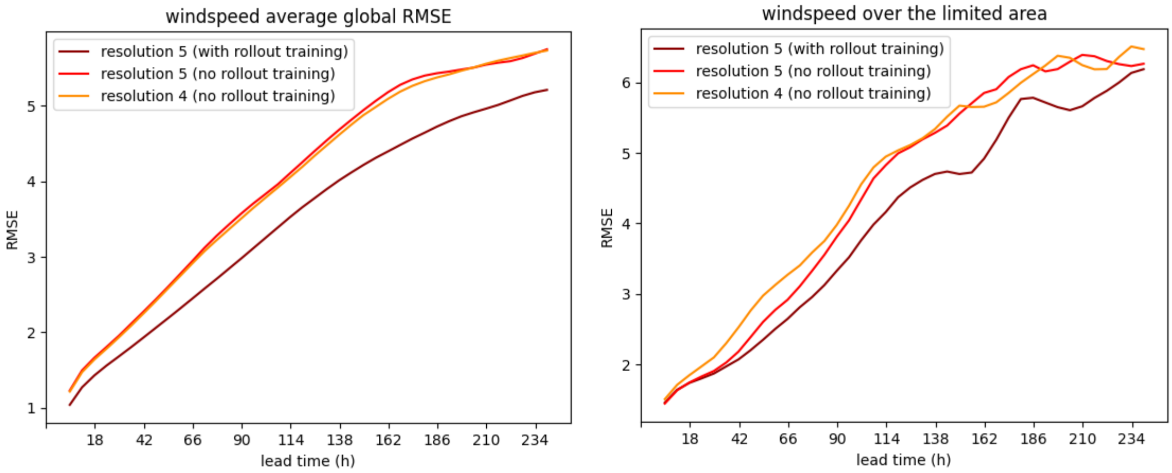

Figure 3: Stretched-grid AIFS evaluation on the limited area and global area: RMSE values for the windspeed at 850 hPa evaluated on the test set (2022) for lead times up to 10 days. Increasing the hidden grid refinement to resolution 5 produces improved RMSE values for short lead times. However, performance is similar or worse for longer lead times. Rollout training substantially lowers RMSE for all lead times.

The difference in performance between the hidden grid resolutions is minimal, indicating that four refinements suffice for this data resolution. Computational resources (in particular GPU memory) can be saved by using a lower resolution hidden grid. However, it remains to be shown how these results transfer to different data resolutions.

On the other hand, we notice that rollout training significantly improves the performance for both short and long lead times for both models. This may indicate that the model has not been trained to convergence yet, this remains to be investigated however.

2. Evaluation of Transformer processing layers results

Since the start of this project, ECMWF has deviated from using a GNN in the processing layers of the AIFS and opted for using transformer layers instead. This has the potential to increase model speed and eliminate the need for predetermined edges. We experimented with adapting this novel framework for a stretched-grid approach. Since AIFS implements the transformer in a one-dimensional sliding window attention operation, complications arise when adjusting this operation to a combined hidden mesh. The decision was made for a simple data appendment, nevertheless resulting in a reasonable model due to the large window size (512). However, for future implementation we suspect that a two-dimensional transformer operation would be more suitable.

A comparison of RMSE values for the windspeed at 850 hPa as a function of lead time can be found in Figure 4. Without rollout training, the transformer shows comparable performance to the GNN models. However, the transformer shows faster convergence than the GNN.

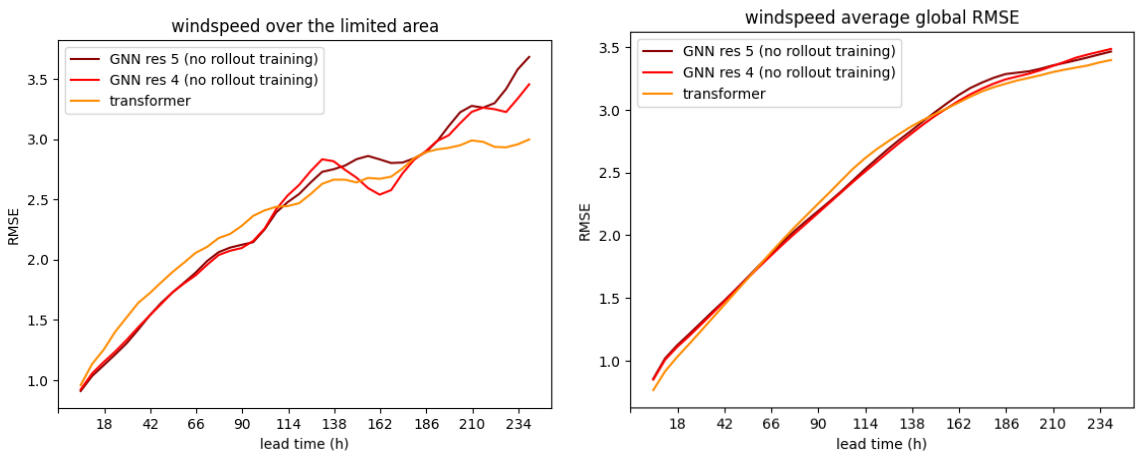

Figure 4: Transformer AIFS evaluation on the limited area and global area: RMSE values for the windspeed at 850 hPa evaluated on the test set (2022) for lead times up to 10 days. The transformer model shows comparable performance to the GNN models.

References:

R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, and P. Battaglia, "Learning skillful medium-range globalweather forecasting," Science, vol. 382, no. 6677, pp. 1416–1421, 2023

S. Lang, M. Alexe, M. Chantry, J. Dramsch, F. Pinault, B. Raoult, M. C. A. Clare, C. Lessig, M. Maier-Gerber, L. Magnusson, Z. B. Boual`ı, A. Prieto, N. Peter, D. Dueben, A. Brown, F. Pappenberger, and F. Rabier, "Aifs - ecmwf's data-driven forecasting system," 6 2024.

I. Price, A. Sanchez-Gonzalez, F. Alet, T. R. Andersson, A. El-Kadi, D. Masters, T. Ewalds, J. Stott, S. Mohamed, P. Battaglia, R. Lam, M. Willson, and G. Deepmind, "Gencast: Diffusion-based ensemble forecasting for medium-range weather," 12 2023.

W. Xu, K. Chen, T. Han, H. Chen, W. Ouyang, and L. Bai, "Extremecast: Boosting extreme value prediction for global weather forecast," 2 2024.